2013

Mapping the YY1 and p65 binding sites on the transcription factor LSF

BOSTON UNIVERSITY

GRADUATE SCHOOL OF ARTS AND SCIENCES

Thesis

MAPPING THE YY1 AND P65 BINDING SITES ON THE TRANSCRIPTION FACTOR LSF

by

WILLIAM DAVID CHURCH

B.S., Northeastern University, 2010

Submitted in partial fulfillment of the

requirement for the degree of

Master of Arts

2013

Approved by

First Reader _____

Ulla Hansen, Ph.D.
Professor of Biology

Second Reader _____

Adrian Whitty, Ph.D.
Associate Professor of Chemistry

Third Reader _____

Cynthia Bradham, Ph.D.
Assistant Professor of Biology

MAPPING THE YY1 AND P65 BINDING SITES ON THE TRANSCRIPTION FACTOR LSF

WILLIAM DAVID CHURCH

Boston University Graduate School of Arts and Sciences, 2013

ABSTRACT

Late SV40 factor (LSF) is a CP2 family transcription factor involved in cell cycle regulation. In liver cancer, LSF is an oncogene, in part due to its role in upregulation of osteopontin leading to increase tumor size. As a result, LSF is a potential target for drug discovery. LSF binds the p65 subunit of the transcription factor NFκB and also the transcription factor ying yang 1 (YY1). In this thesis, I show that binding of both YY1 and p65 occurs at the ubiquitin-like domain of LSF in U2OS cell extracts. Interestingly, when phosphatase inhibitors are added during preparation of U2OS cell extracts, the binding of YY1 and p65 to LSF shifts from the ubiquitin-like domain of LSF to the DNA binding domain. The role of a yet unidentified docking protein may be responsible for this shift in binding. In an attempt to map the specific region of the LSF sequence that is involved in these interactions, I have developed a peptide identification assay which utilizes protease digestion, protein mediated peptide capture, and LC ESI-MS. Through the use of this assay, I'm confident that the sequence(s) involved in these LSF protein-protein interactions can be further defined.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

Abbreviation	Meaning
2xYT broth	2x Yeast Tryptone Broth
APP	Amyloid Precursor Protein
BamHI	*Bacillus amyloli* H1
cDNA	Complementary Deoxyribonucleic Acid
CO_2	Carbon Dioxide
COS7	CV-1 Origin, SV40 – 7
CP2	CCAAT-binding protein 2
DBD	DNA Binding Domain
DMEM	Dulbecco's Modified Eagle Medium
DNA	Deoxyribonucleic Acid
DTT	Dithiothreitol
EDTA	Ethylenediaminetetraacetic Acid
ELISA	Enzyme-linked Immunosorbent Assay
ESI	Electrospray Ionization
GE	General Electric
GST	Glutathione S-Transferase
H_2SO_4	Sulfuric Acid
HCC	Hepatocellular Carcinoma
HDAC1	Histone Deacetylase 1
HEK	Human Embryonic Kidney

HEPES	4-(2-hydroxyethyl)-1 piperazineethanesulfonic acid
HIV	Human Immunodeficiency Virus
HPLC	High-performance Liquid Chromatography
IGEPAL	octylphenoxypolyethoxyethanol
IκBα	nuclear factor of kappa light polypeptide gene enhancer in B-cells inhibitor, alpha
Ικκ	IκB kinase
Ικκβ	inhibitor of nuclear factor kappa-B kinase subunit beta
IL2	Interleukin 2
IL4	Interleukin 4
IPTG	Isopropyl β-D-1-thiogalactopyranoside
IRS	Insulin Receptor Substrates
LC	HPLC coupled
LSF	Late SV40 Factor
LTR	Long Terminal Repeat
NaCl	Sodium Chloride
NaF	Sodium Fluoride
NCBI	National Center for Biotechnological Information
NEB	New England Biolabs
NF-κB	Nuclear Factor Kappa-light-chain-enhance of activated B cells
NotI	*Nocardia otitidis-caviarum* 1
MS	Mass Spectrometry
PAX6	Paired box protein 6

PBS	Phosphate Buffered Saline
PBST	Phosphate Buffered Saline with Tween-20
PCR	Polymerase Chain Reaction
PRDII	Positive Regulatory Domain II
PVDF	polyvinylidene difluoride
RING1	Really Interesting New Gene 1
SAM	Sterile Alpha Motif
SDS	Sodium Dodecyl Sulfate
SDS-PAGE	Sodium Dodecyl Sulfate – Polyacrylamide Gel Electrophoresis
SOC	Super Optimal broth with Catabolite repression
SV40	Simian Virus 40
TBP	Tata Binding Protein
TBST	Tris Buffered Saline with Tween-20
TFIIB	Transcription Factor II B
TMB	3,3',5,5'-Tetramethylbenzidine
Tris-HCl	tris(hydroxymethyl)aminomethane hydrochloric acid
TS	Thymidylate Synthase
U2OS	U2 Human Osteosarcoma cell line
YY1	Ying Yang 1

Introduction

Late SV40 factor (LSF) is a CCAAT-binding protein 2 (CP2) family transcription factor responsible for stable progression through the G1/S transition of the cell cycle (Santhekadur *et al.* 2012). It is known to bind the SV40 promoter, for which it is named (Santhekadur *et al.* 2012), as well as to repress the interleukin-2 (IL-2) promoter (Traylor-Knowles *et al.*, 2010). LSF also binds the promoters of the α, β, and γ-globin genes, uroporphyrinogen III synthase, transferrin, IL4, serum amyloid A3, pax6, and osteopontin genes (Santhekadur *et al.*, 2012). LSF directly upregulates osteopontin in the development of hepatocellular carcinoma (HCC)(Yoo *et al.* 2010) and as a result, causes an increase in tumor size (Yoo *et al.* 2010). HCC is one of the most prevalent forms of liver cancer worldwide and kills more than 600,000 people each year (Ferenci *et al*. 2010) making LSF a potentially important target for drug discovery. Importantly, LSF binds the thymidylate synthase (TS) promoter allowing TS gene expression. TS is the rate limiting enzyme for deoxythymidine triphosphate synthesis (Santhekadur *et al*. 2012) and as such, it plays an essential role in the regulation of S-phase in the cell cycle (Santhekadur *et al*. 2012). Limited deoxyribonucleotide concentrations can trigger cell cycle arrest leading to apoptosis (Santhekadur *et al*. 2012). LSF cell cycle activity is regulated by phosphorylation by ERK and Cyclin C/ Cdk2 during G1 (Santhekadur *et al*. 2012). Phosphorylation at serine 291 by ERK and at serine 309 by Cyclin C/ Cdk2 inhibits the transcriptional activity of LSF while subsequent dephosphorylation reactivates it (Santhekadur *et al*. 2012).

Structurally, LSF is predicted to contain a sterile alpha motif (SAM) domain, a DNA binding domain, and an ubiquitin-like domain (Kokoszynska *et al*. 2008). The SAM and ubiquitin-like domains appear to be close together in sequence while the DNA binding domain is separated from the SAM domain by a 67 amino acid linker region (Kokoszynska *et al*. 2008) (Figure 1b). Though no crystal structure for LSF has yet been reported, the identification of the above domains in LSF was determined by homology modeling and fold recognition programs through the Structure Prediction Meta Server (Kokoszynska *et al*. 2008). Domain spacing was determined through GlobPlot (Kokoszynska *et al*. 2008). Furthermore, LSF's DNA binding domain was predicted to have a high similarity to the DNA binding domain of p53 (Kokoszynska *et al*. 2008) which contains a combination of α-helixes and β-sheets in a beta-sandwich fold (Kokoszynska *et al*. 2008). Though related to the Grainyhead family of transcription factors, LSF diverged from a common ancestor through a gene duplication event (Traylor-Knowles *et al*. 2010). LSF and Grainyhead are both utilized as developemental proteins (Traylor-Knowles *et al*. 2010). However, while LSF is only involved in chicken (Murata *et al.* 1998) and human eye development, through the regulation of PAX6 gene expression (Zheng *et al.* 2001), Grainyhead is required for the development of epithelial integrity in mice, *Xenopus, Drosophilia,* and *C. elegans* (Traylor-Knowles *et al*. 2010). Grainyhead is also necessary in the development of the central nervous system in mice and *Drosophilia* (Traylor-Knowles *et al*. 2010). LSF differs structurally from Grainyhead in that it contains the SAM domain and the ability to dimerize in solution yet binds DNA as a tetramer (Traylor-Knowles *et al*. 2010). In contrast, Grainyhead can only dimerize

and does not bind DNA as a tetramer. Although both LSF and Grainyhead dimerize individually and are structurally similar, they don't dimerize cross-family (Shirra and Hansen, 1998). Previous experimental attempts to map the LSF oligomerization domain have found that the region responsible for LSF dimerization and tetramerization lies between residues 266 and 501 (Shirra and Hansen, 1998). Furthermore, substitution mutants between residues 211 and 213 and residues 235 and 237 show decreased dimerization compared to wild-type LSF (Shirra *et al*, 1994. Shirra and Hansen, 1998).

LSF has been shown to participate in a number of protein-protein interactions with partners including YY1, p65, the major transcriptional proteins TBP and TFIIB, dinG a homolog of RING1, and Fe65, as described in more detail below. RING1 is a Polycomb group protein found in humans and is responsible for binding multiple Polycomb group proteins through its RING finger domain (Satjin and Otte, 1999). Polycomb group complexes are known to repress transcription though interestingly, when RING1 is overexpressed in Rat1a embryonic fibroblast cells, it leads to the overexpression of *c-myc* and subsequent increase in cell proliferation and cell size (Satjin and Otte, 1999). LSF interacts with dinG, a homolog of RING1, at dinG's C-terminus (Tuckfield *et al*. 2002) and this interaction leads to specific repression of the LSF-dependent α-globulin promoter in human embryonic kidney (HEK 293) cells (Tuckfield *et al*. 2002). Fe65 is a neuronal adaptor protein that binds the integral membrane protein APP (β-amyloid precursor protein) which is the precursor for the plaque forming β-amyloid peptide that is implicated in the pathology of Alzheimer's disease. Fe65 has been shown to interact with LSF in COS7 cells through co-immunoprecipitation experiments

between Fe65 and full-length LSF (Bruni *et al*. 1998). LSF also binds to the transcription factor YY1. Binding of YY1 to LSF along with HDAC1 forms the LSF-YY1-HDAC1 complex. After formation, this complex represses expression from the long terminal repeat (LTR) sequence comprising the HIV promoter (Coull *et al*. 2000). This repression causes HIV to enter a latent state allowing the virus to evade the host immune system. Inhibition of the LSF-YY1-HDAC1 complex would therefore force existing latent virus to continue expression and prevent new virus from entering latency. This would potentially allow treatments to combat the viral infection while preventing resurgence of the viral population. LSF has also been shown to interact with p65 (RelA) an essential component of the NF-κB signaling pathway (Bing *et al*. 2000). The p65 subunit of the p50-p65 heterodimer associates with LSF during cytokine induced activation of the NF-κB pathway (Bing *et al*. 2000).

YY1 is a zinc-finger transcription factor that is found in all cell types and binds a wide range of promoters (Flanagan, 1995). The sequence of YY1 contains four zinc finger motifs along with two N-terminal transcriptional activation domains, a glycine rich domain, and a C-terminal transcriptional repression domain (Gordon *et al*. 2006). Because it contains both an N-terminal activation domain and a C-terminal repression domain, YY1 can repress or stimulate transcription of the promoters it regulates (Gordon *et al*. 2006). Whether YY1 acts as a transcriptional repressor or an activator depends on the proteins associating with it (Gordon *et al*. 2006). The co-crystal structure the C-terminus domain (Gordon *et al*. 2006) of YY1 (amino acids 295-408) with a 20 base pair oligonucleotide corresponding to the adeno-associated virus p5 initiator element shows

that the secondary structure of the protein is made up of four α-helices connected by looped regions which are responsible for holding the zinc ions in place forming four zinc finger domains (Figure 2) (Houbaviy *et al*. 1996). All four zinc finger domains bind the major groove of DNA and three of the four zinc finger domains make multiple contacts with the major groove through interactions with their side chains while the fourth makes only a single base contact (Houbaviy *et al*. 1996). YY1 is implicated in various cellular processes including cyclin expression, apoptosis, and histone modification (Gordon *et al*. 2006). It has also been shown to associate with p53 and the oncoprotein c-myc (Gordon *et al*. 2006).

p65 functions as a transcription factor in the NF-κB pathway and predominantly dimerizes with NF-κB protein p50 at the rel homology domain. p65-p50 heterodimers are held in an inactive state through complex formation with IκBα. Inactivation of p65-p50 heterodimer is important in order to prevent nuclear translocation prior to activation of the pathway by upstream effectors since constitutive association of the heterodimer with the cell's transcriptional machinery at the promoter site would lead to constant gene expression (Dolcet *et al*. 2005). Following activation of the cell by growth factors, cytokines, or other specific extracellular stimuli, IκBα is phosphorylated by the IKKβ subunit of the IκB kinase (IKK) complex (Dolcet *et al*. 2005). Once IκBα is phosphorylated, it dissociates from p65-p50 and is degraded by the proteasome, releasing the p65-p50 heterodimer to translocate to the nucleus where it alters transcription at specific promoter sequences (Dolcet *et al*. 2005). The structure of p65, as shown in the co-crystal structure of the p65-p50 heterodimer with DNA derived from the interferon-β

promoter PRDII element, is composed of an α-helix adjacent to a β-sandwich connected to a pair of antiparallel β-sheets by a looped linker region (Figure 2) (Escalante *et al*. 2002). p65 is organized into an N-terminal rel homology domain which contains (from N-terminus to C-terminus) the DNA binding domain, dimerization domain, and nuclear localization signal and a C-terminal transactivation domain (Perkins. 2007) (Figure 3). The NF-κB pathway plays important roles in the regulation of apoptosis, cell proliferation, cell migration, and in the immune response. Because of this, it has been associated with tumorigenesis and thus, it is a potential target for cancer therapy (Dolcet *et al*. 2005).

The goal of the current work is to map the YY1 and p65 protein-protein binding sites of LSF to enable later characterization of binding site properties, as well as binding studies and inhibitor design. The results show that binding occurs between p65 and the ubiquitin-like domain of LSF in U2OS cell extracts. When phosphatase inhibitors are added during preparation of U2OS cell extracts, however, binding is observed to shift from the ubiquitin-like domain to the DNA binding domain of LSF and the N-terminus of LSF linked to a portion of the DNA binding domain. Because of the overlap between these two regions, the interaction between LSF and p65 most likely occurs between residues 65 and 103. Similar results were observed with YY1 binding to LSF. Furthermore, binding of bacterially expressed, recombinant p65 occurs at the DNA binding domain of LSF. A peptide identification assay was developed to determine the specific amino acid sequence involved in LSF binding to p65. Three out of five of the assay steps were validated and are useable for future experiments.

Methods

GST-LSF DNA construct design. I used the pGEX-5x-1 GST expression plasmid from GE Health sciences (GE) for all construct vectors. Inserts corresponding to the LSF DNA binding domain and SAM domain were amplified using polymerase chain reaction (PCR) from a plasmid containing LSF cDNA (EF1α-LSF). The specific primers were engineered to contain restriction sites for BamHI and NotI oriented 5’ to 3’. The DNA Binding domain forward primer sequence is 5'-CCCGGGGGATCCCCCTGCCTTTTCAATATGTGCTTTGTGCT-3' and the reverse primer sequence is 5'-AATGCCGCGGCCGCAGGTGTTCGTTTCTCCATTTTTTCCCTATCCGT-3'. The SAM domain forward primer sequence is 5'-CCCGGGGGATCCCCCCAACCACACCTCAG-3' and the reverse primer sequence is 5'-AATGCCGCGGCCGCTGGACGCACCATCCGGCC-3'. Both vector DNA and amplified inserts were digested separately with 20 units of BamHI and NotI (NEB) in a 50µL reaction with NEB buffer 3 and according to New England Biolabs’ (NEB) restriction digest protocol. 1µL of Temperature Sensitive Alkaline Phosphatase (Promega) in a 20µL reaction with 1µg of DNA and 2µL of 10x MULTI-CORE buffer (Promega) incubated at 25°C for 5 minutes was used to prevent re-circularization of digested vector and was heat inactivated at 74°C for 15 minutes prior to ligation. The digested inserts were gel-purified according to the QIAgen Gel Purification Spin Column Kit and then ligated into the gel-purified digested vector using 400 units of T4 DNA

ligase (NEB) at room temperature for 10 minutes with an insert to vector ratio of 10:1. I used sequencing (MWG Operon) and restriction enzyme digestion followed by agarose gel separation to determine the inserts were present and in the correct orientation. The pGEX1, pGEX2, pGEX3, pGEX4, and full length LSF constructs were generously provided by Gene Chin (Hansen lab member).

Expression of GST-LSF protein. GST-LSF constructs were transformed into T7 Express *lysY/I*^{q} competent cells (NEB) according to the provided protocol. The competent cell aliquots were thawed on ice for 10 minutes prior to transformation. Then, 1 µL of ligation mixture was added directly to the competent cells and they were subsequently mixed lightly by flicking the base of the tube. The competent cells were incubated on ice for 30 minutes before being rapidly heat shocked at 42°C for 20 seconds and then incubated on ice for 5 minutes. The cells were then placed in 950 µL of Super Optimal broth with Catabolite repression (SOC) media and incubated for 1 hour at 37°C before being plated on ampicillin LB agar plates overnight. Cultures were grown from glycerol stocks in 2xYT broth with 100 µg/mL of ampicillin at 37°C for 4 hours. The culture was then diluted 1:1 and incubated at room temperature for 16 hours in the presence of 100 µg/mL ampicillin and 0.5 mM IPTG. Cells were centrifuged at 4250xg at 4°C for 10 minutes and resuspended in lysis buffer (50 mM Tris-HCl pH 8.0, 50 mM EDTA, 15% glucose, and 2 mM DTT) at a ratio of 33.5 mL of buffer for 500 mL of culture. The cell suspension was then incubated with 20 mg/mL of lysozyme for 20 minutes at 4°C before being mixed with lysing solution (10 mM Tris-HCl pH 8.0, 50 mM EDTA, and 1%

IGEPAL) at one fourth the volume of the lysis buffer for 20 minutes at 4°C. The solution was passed through a needle four times, centrifuged for 5 minutes at 18,000xg at 4°C, and the supernatant was aliquoted, flash frozen in liquid nitrogen, and stored at -80°C. For extended use, the protein can be mixed with 50% glycerol and 0.01% sodium azide and stored at -20°C to limit repeated freeze-thaw cycles. Protein amount was quantified by Bradford Assay and 10% SDS-PAGE separation. Purity was determined by 10% SDS-PAGE separation. Protocol yields for individual constructs per 2 mL culture volumes are as follows: GST-DBD yields 250 ng, GST-SAM yields 500 ng, pGEX4 yields 500 ng, pGEX1 yields 250 ng, pGEX2 yields 250 ng, pGEX3 yields 167 ng, and GST-LSF yields 50 ng.

GST-LSF binding assay. Studies were performed with the U2OS mammalian osteosarcoma cell lines. U2OS cells were grown in DMEM media with 10% fetal bovine serum at 37°C with 5% CO_2 to 90% confluency, washed with ice cold 1xPBS twice, and collected in a 1.5 mL microcentrifuge tube by scraping the dish with a cell scraper after the addition of 1mL ice cold 1x Phosphate Buffered Saline (PBS) per 100 mm diameter tissue culture dish. The cells were then pelleted by centrifugation at 5,000xg for 2 minutes. The supernatant was removed and the pellet resuspended in 500 µL of 50 mM HEPES pH 7.5, 250 mM NaCl, 5 mM EDTA, 50 mM NaF, and 1% Triton X-100; Phosphatase Inhibitor Cocktail IV (Calbiochem) was added here for phosphatase inhibitor experiments. The cells were then centrifuged at 20,000xg at 4°C for 10 minutes and the supernatant was transferred to a fresh 1.5 mL microcentrifuge tube. The

supernatant then was either directly used in the GST-LSF binding assay or stored at 80°C after flash freezing in liquid nitrogen. The GST-LSF bacterial lysate (corresponding to 500 ng of bacterial lysate) was incubated with 20 µL of glutathione Sepharose 4B bead resin (GE Healthcare) in a 50% slurry with 1xPBS for 30 minutes at room temperature under constant rotation. In order to remove non-specific binding proteins, the beads were centrifuged at 500xg for 5 minutes and the supernatant was removed and replaced with 100 µL of 1xPBS; this was repeated two additional times. The bacterial lysate immobilized on bead resin was then added to 300 µL of U2OS cell extract and incubated for 1 hour at 4°C under constant rotation. The beads were washed again with 1xPBS three times and centrifuged at 500xg for 5 minutes per each wash. The beads were incubated with 50 mM Tris-HCl pH 8.0, 10 mM reduced glutathione for 5 minutes at room temperature before being centrifuged at 500xg for 5 minutes. The used beads were discarded and the supernatant was kept. For phosphatase inhibitor studies, the phosphatase inhibitor cocktail used was Phosphatase Inhibitor Cocktail IV (Calbiochem). Phosphatase inhibitor cocktail working dilution was 1:100.

Western blot of GST-LSF binding proteins. 10 µL of the GST-LSF binding assay samples were electrophoresed through a discontinuous 10% SDS-PAGE mini-gel. The 4% acrylamide stacking gel and the 10% acrylamide running gel was made with Protogel premixed 37.5:1 acrylamide to bisacrylamide (National Diagnostics). For a single gel, the electrophoresis conditions were constant amps at 0.01 Amps for the stacking gel and 0.02 Amps for the running gel. The gel was electrophoresced until the dye front reached the

bottom of the mini-gel; this on average correlated into 30 minutes for the stacking gel and 1 hour for the running gel. The dye used was diluted from a 6x concentrated stock and contained dithiothreitol (DTT), 2-mercaptoethanol, bromophenol blue, sodium dodecyl sulfate (SDS), and 0.75 M Tris-Glycine buffer pH 6.8 (6x specification). The SDS-PAGE gel was then transferred to a methanol-activated PVDF membrane (GE Healthcare) in a wet sandwich containing the gel, filter paper, padding, and the membrane overnight at 25 V in an electrophoresis apparatus filled with Tris-glycine transfer buffer and 10% methanol at 4°C. The membrane was blocked with 5% milk in 1xTris buffered saline solution with 0.1% Tween-20 (TBST) for 1 hour at room temperature and then washed three times, ten minutes each, and under constant rotation with 1xTBST. The membrane was then blotted with either 1:100 of α-YY1 (Pierce; product number: PA5 - 12206) or 1:500 α-p65 (Pierce; product number: PA5 - 16545) for 1 hour in 5 mL of 5% milk 1xTBST in a heat sealable bag under constant rotation at room temperature. The membrane was then washed three times with 1xTBST for 10 minutes each before being blotted with 1:5000 of goat anti-rabbit conjugated IgG-HRP secondary (Santa Cruz Biotech; Product number: sc-2004) in 50 mL of 5% milk 1xTBST under constant rotation at room temperature. The membrane was then washed three times with 1xTBST for 10 minutes each and then incubated for 2 minutes with 10mL of the Millipore Chemilucent Plus Western Blot Enhancing Kit (Product number: 2650) in a small plastic container before being imaged on a Bio-Rad ChemiDoc system and Image Lab software (Bio-Rad).

Factor Xa Digests

0.01, 0.1, or 1 ng/µL of Factor Xa (NEB; product number: P8010S) was incubated with 1 µg of GST-DNA binding domain either immobilized on glutathione sepharose beads or in solution. For the digestions performed on the bead support, GST-DNA binding domain was first incubated with glutathione sepharose beads for 30 minutes at room temperature. The beads with the immobilized protein were then transferred by pipette to a solution containing 10 µL of 20 mM Tris-HCl pH 8.0 with 100 mM NaCl and 2 mM $CaCl_2$. For the digestion performed in solution, GST-DNA binding domain was first eluted from glutathione sepharose beads with 10 µL of 50 mM Tris-HCl pH 8.0 with reduced glutathione. Factor Xa digest was performed at 4 or 37˚C for 2 or 16 hours. The reaction conditions used for all experiments besides the Factor Xa pilot experiment are 1 ng/µL of Factor Xa incubated for 2 hours at 37˚C.

Endoproteinase AspN Digests

For native digestions, 1 µg/µL of Endoproteinase AspN (NEB; product number: P8104S) was incubated with 1 µg of LSF DNA binding domain for 16 hours at 37˚C in a 10 µL solution containing 50 mM Tris-HCl pH 8.0 with 2.5 mM $ZnSO_4$. For denatured digestions, the DNA binding domain was first denatured in a solution containing 6 M guanidinium-HCl and incubated for 5 minutes. Prior to digestion, this sample was then diluted to 2 M guanidinium-HCl with 50 mM Tris-HCl pH 8.0 with 2.5 mM $ZnSO_4$. The reaction was performed in a 10 µL reaction with 1 µg/µL Endoproteinase AspN for 16 hours at 37˚C. The denatured method was used for all LC ESI-MS experiments.

Peptide Purification on Reverse Phase C-18 Spin Columns

Pierce Biotechnology C-18 spin columns were used to prepare the peptide samples for the LC ESI-MS experiments. In order to prepare the peptide samples for binding to the C-18 resin, 10 µL of Endoproteinase AspN digest was mixed by pipette with 3.33 µL of 2% trifluoroacetic acid (TFA) in 20% acetonitrile (ACN). The column was activated with 200 µL 70% methanol and centrifuged at 1,500xg for 1 minute. The flow-through was removed and the activation step was repeated two additional times. 200 µL of 0.5% TFA in 5% ACN was used to equilibrate the column. The column was centrifuged at 1,500xg for 1 minute and the flow-through was removed. The equilibration step was repeated one additional time. All 13.33 µL of the activated sample was placed on top of the resin bed with a pipette. The column was placed into a new microcentrifuge tube and centrifuged at 1,500xg for one minute. The flow-through was taken by pipette and reapplied to the top of the resin bed to maximize absorption of the peptide sample. The column was centrifuged at 1,500xg for one minute and the flow-through was removed. The column was then washed with 200 µL of 0.5% TFA in 5% ACN and centrifuged at 1,500xg for one minute. The wash step was repeated three additional times. The column was placed into a new microcentrifuge tube and 25 µL of 50% ACN was added to elute the bound peptides. The column was centrifuged at 1,500xg and the flow-through was collected and transferred to a clean tube. The elution step was performed again in order to ensure all bound peptides were collected. The final sample volume was 50 µL.

LC ESI-MS Experiment

All LC ESI-MS experiments were performed by the Chemical Instrumentation Center of Boston University under the direction of Dr. Norman Lee. A 50 µL sample of 10 µg of peptides in 50% ACN was submitted to Dr. Lee for the experiment. This sample was purified using Pierce Biotechnology C-18 spin columns as previously described.

Biotinylation of p65

All biotinylation experiments were performed according to the guidelines in the Pierce EZ-Link Sulfo-NHS-Biotinylation Kit (product number: 21425). Biotinylation reagent was added at a 20 fold molar excess and the reaction proceeded for 1 hour at 4˚C in 1xPBS unless otherwise noted in the results section. Excess biotinylation reagent was removed with 2mL or 5mL Zebra Desalting Columns (Pierce) prior to sample use in downstream processes. For ELISA experiments where p65 (Trevor Siggers, Boston University) was bound to the LSF DNA binding domain during the biotinylation reaction, one molecule of DNA binding domain was added to one molecule of p65 in a high protein concentration 7 µL 1xPBS solution for 1 hour at 4˚C. The molecule to molecule ratio was calculated based on the molecular weight of the DNA binding domain and p65 as well as the mass of the two proteins in solution. Because the p65 existed in an impure solution, the amount of DNA binding domain was moderately higher. Since Sulfo-NHS-

Biotinylation is concentration dependent, when reactions had to be quenched prior to 1 hour, 800 µL of 1xPBS was used (a 30:800 dilution) in order to dilute the biotinylation reagent.

ELISAs

All ELISAs were performed on Nunc MaxiSorp flat bottom 96-well plates (Affymetrix). The coating buffer was 100 µL 1xPBS per well. The 1xPBS recipe used was 2.7 mM KCl, 10 mM KH_2PO_4, and 138 mM NaCl at pH 7.2. In some experiments ELISA plates were prepared by coating with 1 pmol avidin (Pierce) per well while for others 100 pmol per well was used; identified in the Results section. The assay plate was coated with avidin for 48 hours at 4 ˚C. Unless otherwise noted, 100 µL of 10 ng/µL biotinylated p65 was added to each experimental sample well. The biotinylated p65 was incubated for 16 hours at 4 ˚C. Two blocking buffers were used in different experiments. Most experiments used 200 µL of 1% bovine serum albumin (BSA) (Sigma) in 1xPBS with 0.05% Tween-20 (1xPBST) per well, with an incubation time of 16 hours at 4 ˚C. In some experiments, however, 300 µL per well Superblock (Pierce) was added to the plate and immediately removed - repeated two additional times. The wash buffer was 300 µL of 1xPBST per well. Three washes were performed following each incubation step for all ELISAs. All protein and antibody solutions were prepared in 1xPBS. GST-LSF DNA binding domain was added at 50 µL per well for 1 hour at room temperature. All antibodies were added at 100 µL per well and incubated for 30 minutes at room

temperature. All antibody dilutions are described in the results section. The detection substrate used was 50 µL 3,3',5,5'-Tetramethylbenzidine (TMB) (Pierce). The development reaction was quenched with 50 µL 0.2 M H_2SO_4. The signal was read at an absorbance of 450 nm by a Spectra Max M5 plate reader.

Results

Development of a set of protein domain constructs that collectively span the LSF Sequence

To elucidate the region of interaction between LSF and candidate binding proteins, it was necessary to divide the LSF sequence into separate regions for study. This was important because it allowed me to identify the domains within LSF that participate in a given protein-protein interaction, and eliminate non-interacting domains from future studies. I utilized a panel of seven LSF fragment constructs, two of which I generated and the remaining five which were a kind gift of Gene Chin. These constructs collectively spanned the entire LSF sequence, with individual constructs encoding the DNA binding domain, SAM domain, and ubiquitin-like domain, as well as the N-terminal region to half of the DNA binding domain, the linker regions between the DNA binding domain and SAM domain, and the C-terminal region (Figure 1a). A full-length LSF construct was also designed to function as a control in subsequent experiments. All constructs were tagged with a 26 kDa glutathione S-transferase (GST) protein tag for affinity purification

and enhanced stability in solution. The GST protein was encoded using a pGEX-5x-1 vector (GE Healthcare) which places the GST protein N-terminal to the fusion protein, connected by a 4 amino acid linker containing a cleavage site for the sequence specific protease, factor Xa.

GST-LSF DNA constructs were generated using sequence information from the NCBI database (Shirra *et al.* 1994) as well as domain sequence predictions from Kokoszynska *et al* (Kokoszynska *et al*, 2008). The GST-LSF constructs from pGEX1-4, kindly provided by Gene Chin, cover the full LSF sequence. The proteins expressed by the four truncated GST-LSF constructs pGEX1-4 overlap by 11 to 37 amino acids with their adjacent constructs. Not including the GST tag, the proteins expressed by the pGEX1, pGEX2, pGEX3, and pGEX4 constructs are 180, 151, 115, and 120 amino acids respectively (Figure 1a). Furthermore, the LSF DNA binding domain encoded by the DBD construct is 195 amino acids while the SAM domain encoded by the SAM construct is 63 amino acids (Figure 1a). Because the pGEX4 construct contains the full sequence of the ubiquitin-like domain of LSF, a separate ubiquitin-like domain construct was not generated. Sequence identity and orientation of all DNA constructs was confirmed using restriction enzyme digestion as well as by sequencing.

GST LSF constructs successfully express GST fusion proteins

The seven individual DNA constructs described above were expressed in bacteria for use in future binding studies. The main criteria for acceptable protein constructs were

that the protein should be soluble, have the correct size by SDS-PAGE, and be of high purity. To this end, I utilized BL21 T7 Express *lysY/I*^{q} competent *Escherichia coli* (*E. coli*) as an expression host. The benefits of this system are the absence of bacterial proteases Lon and OmpT, tight control of expression by *lacI*^{q}, use of T7 RNA polymerase for expression, and resistance to accidental lysis during expression due to *lysY*. By utilizing the bacterial *lac* operon for expression and the repressor *lacI*^{q}, toxic mammalian proteins can be expressed, and T7 RNA polymerase ensures leaky expression is minimized. Basal expression of T7 RNA polymerase is further controlled by *lysY* which encodes an amidase deficient T7 lysozyme that binds the T7 RNA polymerase in a 1:1 protein-protein complex (Zhang and Studier, 1997). T7 RNA polymerase-lysozyme complex formation is inhibited by an increased rate of T7 RNA polymerase elongation complex formation as T7 lysozyme lacks the ability to bind the elongation complex (Zhang and Studier, 1997). This allows full length mRNA transcripts to be produced following induction of the *lac* operon by IPTG. In addition, accidental lysis does not occur upon induction. Furthermore, the absence of bacterial proteases OmpT and Lon ensures that the expressed protein is less likely to be degraded after translation. Protein purity, and the size of each expressed construct, was determined by SDS-PAGE.

The predicted sizes of the DNA binding domain, SAM domain, and the expressed proteins from the pGEX1-4 constructs without the GST fusion partner are 21.3, 6.8, 19.8, 16.5, 12.5, and 13.2 kDa, respectively (Figure 3b). The GST fusion partner contributes an additional 26 kDa to each construct. SDS-PAGE analysis confirmed the sizes of the expected protein species (Figure 4). The pGEX1 expressed protein migrates slightly

faster than the expected size of 46 kDa in comparison to the expressed protein from pGEX2 (expected mw = 43kDa). The pGEX4 expressed protein also migrates slightly faster than the expected size of 39.2 kDa in comparison to the expressed protein from pGEX3 (expected mw = 38.5 kDa). These results were observed multiple times. The pGEX1-4 constructs were provided by Gene Chin and her SDS-PAGE analysis confirms that they run at their expected sizes. Her analysis of the pGEX1-4 DNA constructs by sequencing and restriction enzyme digestion also shows that the DNA constructs are the correct size. This means that an error most likely occurred in transfer to the BL21 host, protein expression, or in the preparation of the polyacrylamide gel. Since the pGEX1-4 DNA sequences were confirmed by sequencing and restriction enzyme digest post-transformation, the problem is most likely in protein expression or in preparation of the polyacrylamide gel. Because the sizes of the pGEX1-4 DNA constructs were confirmed to be accurate, I concluded that the inconsistencies in the migration of the expressed proteins were acceptable and the proteins were deemed appropriate for further study.

YY1 and p65 from mammalian cell extracts bind LSF's ubiquitin-like domain

Previous work from Coull *et al*. and Bing *et al*. shows that LSF binds both YY1 and p65, respectively. I therefore decided to use YY1 and p65 as the principle candidate proteins to identify protein-protein interaction sites on LSF. Interestingly, while Coull *et al.* showed that YY1 binds to LSF at the DNA binding domain, p65 was only shown to

bind with full-length LSF (Bing et al, 2000). Therefore, I was interested in determining the region on LSF where p65 binds.

To test for p65 and YY1 binding to LSF, I captured the expressed GST-LSF proteins to glutathione sepharose beads before incubating the beads with U2OS mammalian cell extracts containing p65 and YY1. The beads were then washed three times to remove any non-specific binding proteins. The bound proteins were eluted with reduced glutathione and analyzed by SDS-PAGE followed by western blot. The membrane was immunoblotted with anti-p65 or anti-YY1 polyclonal antibodies, as described in Methods. The results showed that YY1 binds to the LSF fragment encoded by the pGEX4 construct, corresponding to amino acids 383 - 502 of the LSF sequence, as well as to full length LSF as expected. Binding of YY1 to the LSF fragments encoded by the pGEX1 construct (amino acids 1 – 180) and the pGEX2 construct (amino acids 169 – 319) was also observed, although the bands corresponding to these binding interactions appeared to be much fainter than those observed for the pGEX4 construct suggesting that they were due to non-specific binding of the antibody (Figure 5). The faint band for full-length LSF is most likely due to poor expression of this construct since larger culture volumes were needed to express suitable amounts of protein for the experiments. When assaying for p65 similar results were observed, with a more intense band being displayed for pGEX4 and fainter bands for pGEX1 and pGEX2 (Figure 6). All protein binding experiments were performed in triplicate, however, the binding of YY1 and p65 to LSF was not investigated with cell extracts from different cell lines. The doublet band indicated in Figures 6 and 8 was determined to be due to a bacterial protease non-

specifically captured on the glutathione bead resin during incubation of the glutathione sepharose beads with the bacterially expressed GST-LSF proteins (Figure 9). As seen in Figure 9, the doublet band is not present in the input lane but when incubated with glutathione sepharose beads that were treated with lysed BL21 cells, the doublet becomes noticeable. This finding may indicate that a bacterial protease from the BL21 cells is non-specifically binding to the glutathione beads and is cleaving p65 into a cleavage product that is indicated by the smaller doublet band. This is significant because it presents a plausible argument that both the band for p65 and the doublet are representative of p65. Since the pGEX4 construct encodes a protein fragment containing the ubiquitin-like domain, these results of these binding experiments suggest that both p65 and YY1 bind to the ubiquitin-like domain of LSF.

Addition of phosphatase inhibitors shifts p65 and YY1 binding to the DNA binding domain of LSF

Because Coull *et al.* showed that YY1 binds to the DNA binding domain of LSF, while my data showed binding to the ubiquitin-like domain of LSF, it became necessary to determine the cause of the discrepancy. I hypothesized that perhaps the phosphorylation state of the p65 and the YY1 was different in our experiments compared to the earlier work of Coull *et al.*, and that this difference might account for the different binding properties we observed for these proteins with LSF. Post-translational modifications have been shown to affect protein-protein binding in protein interaction

domains (Pawson and Nash, 2003). We therefore repeated the binding experiments but included phosphatase inhibitors in the preparation of the U2OS cell extracts to maintain the phosphorylation state of the p65, YY1, and other proteins in these extracts. The U2OS cell extract samples were treated prior to binding GST-LSF with a phosphatase inhibitor cocktail containing

cantharidin, (-)-p-bromotetramisole oxalate, and calyculin A, responsible for inhibiting a range of threonine/serine and alkaline phosphatases. When I repeated the binding assays with the seven GST-LSF constructs, under these new conditions YY1 appeared to bind the DNA binding domain, full length LSF, and the pGEX1 expressed protein (Figure 7). These results differed dramatically from what was observed without inclusion of phosphatase inhibitors in the extract samples. Blotting with an anti-p65 antibody mirrored the results seen for YY1, showing binding with the DNA binding domain, full length LSF, and pGEX1 expressed proteins (Figure 8). As described above, the pGEX1 construct contains the N-terminal region of LSF along with the majority of the DNA binding domain (Figure 1b). The pGEX2 construct contains a fraction of the DNA binding domain corresponding to amino acids 104 – 259 of the LSF sequence (Figure 1a). Since binding of YY1 and p65 was observed for both the pGEX1 and DNA binding domain constructs (residues 65 to 259 of the LSF sequence) but not the pGEX2 construct, binding most likely involves regions of the LSF sequence located between residues 65 and 103 (Figure 1b). This 38 amino acid sequence, residues 1 – 38 of the DNA binding domain, corresponds to about one fifth of the full DNA binding domain sequence. My results therefore narrow the region of investigation for future experiments to residues 65

– 103 of the LSF sequence. The fact that binding of p65 and YY1 shifts when phosphatase inhibitors are present suggests that binding may involve a more elaborate mechanism than expected depending on the phosphorylation or dephosphorylation of one of more of the participating proteins.

Peptide Identification Assay Design and Optimization of Factor Xa Cleavage

Since an aim of the study was to determine the specific LSF peptide sequence or sequences responsible for binding p65 and YY1, I developed a new assay to investigate this question. The assay involves cleaving the DNA binding domain of LSF into peptide fragments with a site-specific protease, and then capturing these fragments on beads to which p65 had been immobilized. For the binding peptides to be identified, the bound p65 would have to be denatured thereby releasing the bound peptides for analysis. This release can be problematic for protein tags like GST because the conditions necessary for the denaturation of p65 would also denature the GST resulting in undesired elution from glutathione column. I therefore decided to use biotinylation and the streptavidin-biotin interaction for coupling of the p65 to the beads, since the high affinity of biotin for streptavidin ($K_d = 10^{-15}$ M) would likely withstand low to medium concentrations of denaturants allowing elution of the LSF peptide without washing off the p65. Identification of binding peptides would be performed by electrospray ionization mass spectroscopy (ESI-MS). For the ESI-MS identification, the eluted peptide sample would be compared to two control samples. The first control sample would consist of the flow-

through from the streptavidin-biotinylated p65 beads after the addition of protease digested DNA binding domain peptide fragments. The second would consist of a protease digestion of the DNA binding domain. The second control sample would not be applied to the beads and would contain all the peptide fragments capable of binding the biotinylated p65. Both control samples along with the sample eluted from the streptavidin-biotinylated p65 beads would be analyzed by LC ESI-MS. The mass-to-charge ratio of the samples would then be compared to the site-specific protease peptide fragment map for the DNA binding domain as well as the DNA binding domain amino acid sequence to identify the sequences of the binding peptides. This is necessary because the mass-to-charge ratio can increase or decrease depending on how the sample is ionized and a single peptide species can therefore resemble different peaks in the ESI-MS results. By comparing the peptides eluted from the streptavidin-biotinylated p65 beads to the two control samples, a consensus can be formed on the binding peptides mass from the mass-to-charge ratio. Furthermore, ESI-MS is not a direct sequencing method and sequence identity can only be determined by comparing the mass of the peptide from the mass-to-charge ratio to the predicted size of the peptide fragment map.

Because my LSF DNA binding domain construct contains a GST tag that is susceptible to cleavage by proteases and chemical agents, the first step in designing the peptide identification assay was to show that the GST tag could be removed on the GST-DNA binding domain fusion protein. The pGEX-5x-1 vector contains a sequence coding for the recognition site for the protease Factor Xa proximal to the 5’ end of the vector’s cloning site. This places the Factor Xa recognition site in the linker region between the

GST tag and the fusion partner allowing the GST tag to be cleaved using Factor Xa. To determine the optimal cleavage conditions, I tested nine different reactions using the DNA binding domain, and analyzed the results by Coomassie Brilliant Blue staining on a polyacrylamide denaturing gel (Figure 10). The amount of Factor Xa, the temperature of the reaction, the length of the reaction, and whether the reaction was performed with the GST-DNA binding domain bound to the glutathione sepharose beads or in solution were the conditions that were altered. Based on this experiment, for future Factor Xa reactions I chose conditions of 1 ng/µL of Factor Xa incubated at 37°C for two hours with the GST-DNA binding domain fusion protein bound to the glutathione sepharose beads, because these conditions resulted a faint band at about 47 kDa and a bright band at about 13 kDa. The expected molecular weight of the GST protein is 26 kDa and the linker region between GST tag and fusion protein is 440 daltons. The DNA binding domain is calculated to be 21.5 kDa based on the LSF sequence and domain predictions by Kokoszynska *et al* however the band for the protein appears around 13 kDa on the polyacrylamide gel. When sequenced, the DNA insert for the GST-DBD construct was found to match the predicted DNA binding domain DNA sequence at 100% similarity (BLAST: bl2seq). Therefore, it is most likely that the 13 kDa band is the DNA binding domain and it is running at a lower molecular weight most likely due to the same issues affecting the expressed proteins from the pGEX1, pGEX2, pGEX3, and pGEX4 constructs as mentioned in the last paragraph of the “GST LSF constructs successfully express GST fusion proteins” section of the results. Furthermore, a Factor Xa cleavage site is not present in the DNA binding domain sequence. This suggests that additional

proteolysis of the DNA binding domain is unlikely and that the 13 kDa band is not representative of the DNA binding domain being cleaved multiple times.

Endoproteinase AspN Cleaves Denatured LSF DNA Binding Domain

Cleaving the DNA binding domain into peptides fragments is important for the identification of the protein-protein binding site on LSF. This is because it is easier to determine the identity of individual peptide fragments than it is to uncover the region of binding on a full length protein without a crystal structure. Digestion of protein into peptide fragments can be performed using proteases or chemical agents. I decided to use proteases because they are believed to be more reliable and safer than chemicals. To that end, I utilized ExPASy's peptide cutter tool (Hoogland *et al.* 2005) to predict where a select set of proteases would cleave the DNA binding domain of LSF. My criteria for selecting a protease for the cleavage experiments was that multiple peptide fragments would be generated, and that these fragments would be on average greater than six amino acids but less than one-hundred and twenty. These rules were selected because the binding motif would most likely be six amino acids or greater, while a one-hundred and twenty amino acid fragment would encompass half of the DNA binding domain and therefore provide only limited information on the location of the binding site. However, smaller peptide fragments were more disfavored because large peptides can be further digested into smaller fragments with subsequent proteases and the assay could be repeated again to localize the specific binding site. Using ExPASy's peptide cutter tool, I

found that the protease Endoproteinase AspN cleaved the DNA binding domain eleven times, giving fragments with an average size of 21 amino acids (Figure 11).

Endoproteinase AspN was used to cleave LSF DNA binding domain expressed from the GST-DBD construct, after removal of the GST tag using the previously described optimized conditions. 1 µg of protein was digested by 1 µg of Endoproteinase AspN in 50 mM Tris-HCl, 2.5 mM $ZnSO_4$ pH 8.0 for 16 hours at 37°C. The sample was then mixed with SDS-sample buffer and analyzed on a 20% SDS-PAGE gel with Coomassie Brilliant Blue staining alongside a control lane containing uncleaved DNA binding domain. After destaining, both the control band and the sample band were present at about 25 kDa suggesting that Endopretinase AspN digestion was unsuccessful. Since protein folding can inhibit protease digest and since native protein is not a requirement for downstream aspects of the assay, I denatured 1 µg of DNA binding domain in 6 M guanidine-HCl (Apffel *et al*, 1995). The sample was then diluted to 2 M guanidium-HCl prior to digestion with Endoproteinase AspN (Apffel *et al*, 1995). Lysozyme (molecular weight = 14.3 kDa) was used as a control substrate, and 70% formic acid as a control cleavage agent, because lysozyme can be digested by Endoproteinase AspN and formic acid fragments the DNA binding domain at the same sites as Endoproteinase AspN since it cleaves at the N-terminal bond of aspartic acid residues. The denatured samples were also compared against native samples digested as previously stated. All samples were analyzed on a 10-20% pre-made SDS-PAGE gradient gel (Thermo Fisher) with Coomassie Brilliant Blue staining (Figure 12). Both lysozyme and DNA binding domain samples were cleaved by Endoproteinase AspN as indicated by

the absence of a band around 25 kDa for the DNA binding domain and 14 kDa for lysozyme and the presence of a band at 25 kDa for the DNA binding domain and 14 kDa for lysozyme. Formic acid also cleaved native lysozyme and DNA binding domain samples. These results suggest that LSF DNA binding domain needs to be denatured prior to digestion by Endoproteinase AspN and that formic acid can also digest the DNA binding domain. However, it should be noted that if formic acid is used, the acid will need to be neutralized as it will most likely interfere with downstream assay processes. These results also suggest that the inconsistencies associated with the migration of the pGEX1-4 expressed proteins and the GST cleaved DNA binding domain in polyacrylamide gels is an issue involving the preparation of the polyacrylamide gels. This is because the DNA binding domain sample in Figure 12 is similar to the sample in Figure 10 with the exception that the sample in Figure 12 was electrophoresed on a pre-cast gradient polyacrylamide gel and the sample in Figure 10 was electrophoresed on a 15% polyacrylamide gel made in the lab. The DNA binding domain sample in Figure 12 migrates closer to its expected molecular weight of 21.5 kDa at 25 kDa while the DNA binding domain sample in Figure 10 migrates at 13 kDa.

ESI-MS Can Detect LSF DNA Binding Domain Peptide Fragments

Since peptide fragments of the DNA binding domain could be generated by Endoproteinase AspN digestion using the denaturing protocol, it became possible to validate downstream identification methods. Validation is important to determine if the

assay can produce the intended data. I decided to use HPLC-coupled ESI-MS for as the detection method (LC ESI-MS). For this method, samples are first concentrated on an HPLC column before being analyzed by ESI-MS. All LC ESI-MS studies were performed by Senior Instrument Specialist Dr. Norman Lee from the Chemical Instrumentation Center at Boston University.

10 µg of LSF DNA binding domain was denatured and digested by Endoproteinase AspN as previously described. I then captured the fragmented peptides on a C-18 spin column. This method is advantageous because it removes the guanidium-HCl as well as the Endoproteinase AspN as the C-18 resin is suitable for binding small peptides but not larger proteins. The peptides were eluted in 50 µL of 1:1 acetonitrile/water solution and given to Dr. Norman Lee for processing. The resulting chromatogram from the coupled HPLC column (Figure 13), along with the mass-to-charge readouts from the ESI-MS, were analyzed and the peptide fragments were assigned to the individual chromatogram peaks (Table 1). All peptide fragments except the 3,960 dalton fragment were assigned to peaks on the chromatogram. However, the 3,960 dalton fragment may have been too large to be captured by the MS detector. The chromatogram shows a peak at 1.45 minutes which was not assigned a peptide fragment, and so it is possible that the 3,960 dalton fragment corresponds to this peak. Regardless, these data suggest that the DNA binding domain peptide fragments generated by Endoproteinase AspN digestion can be detected by LC-MS and thus that this method of detection can be used for future studies.

Biotinylated p65 binds LSF's DNA Binding Domain

An important aspect of the binding peptide identification assay is the capture of natively structured, active p65 to the beads. Native p65 is needed because the protein will not bind peptide fragments if it is denatured. Furthermore, the active p65 must be placed on a bead support so it can be manipulated and carried through multiple assay steps. To this end, I decided to biotinylate of p65 and use streptavidin capture as a method to bind active p65 to a bead support. However, there are multiple biotinylation chemistries with different conditions to consider. I chose to use primary amine directed biotinylation through Pierce Biotechnology's Sulfo-NHS-Biotinylation Kit. This reagent specifically labels lysine residues and the N-terminal amine of proteins. However, because lysines may be involved in the protein-protein interaction site of p65, there is a possibility that biotinylation might render the protein inactive either through blocking the binding site or by inducing conformational changes in the protein that occlude the binding pocket. To test this possibility I used the GST-LSF bead binding assay to test the ability of the biotinylated p65 to bind to GST-DNA binding domain captured on glutathione sepharose beads. Recombinant p65 protein expressed in *E.coli* was generously provided by Dr. Trevor Siggers of Boston University.

The purity of the provided protein was assessed by 10% SDS-PAGE analysis with Coomassie Brilliant Blue staining (Figure 14). The result showed that protein sample was only partially purified with three major bands and multiple minor bands contaminating the sample. Because the contaminating proteins were expressed in bacteria, I did not

think they would interfere with downstream aspects of the assay since it is not likely that proteins expressed from a non-pathogenic bacterial strain (BL21 *E. coli*) would interact or inhibit a mammalian transcription factor. I therefore decided to use the protein preparation anyway. However, the use of this protein preparation could be problematic because the Coomassie Brilliant Blue staining does not give any indication of protein activity and so the ratio of active p65 to inactive p65 cannot be determined from the stain. Furthermore, because the concentration of p65 is lower than the concentration of the protein preparation, more total protein will have to be used for the assay. Biotinylation was performed according to the procedure recommended in the instructions for the Pierce Kit which recommends using a 20 fold molar excess of biotinylation reagent. The biotinylation reagent is unstable under the conditions used for labeling, and after 60 minutes at 4 °C the reactivity of the reagent decreases. Excess biotin was removed through the use of a provided Zebra Desalting Column. 100 ng of the biotinylated p65 was incubated with GST-DNA binding domain on glutathione sepharose beads for 16 hours at 4° C, before being eluting with reduced glutathione. The same procedure was performed with non-biotinylated p65 as a control. The two assay samples, along with 100 ng of non-biotinylated and biotinylated p65, were analyzed on a denaturing 10% polyacrylamide gel and then transferred to a PVDF membrane. The membrane was first incubated with an anti-p65 antibody before being stripped and then re-incubated with a streptavidin-HRP conjugate (EMD Millipore; product number: 18-152). The anti-p65 immunoblot shows the binding of p65 to GST-DNA binding domain (Figure 15a). When incubated with the streptavidin-HRP conjugate, binding of biotinylated-p65 to GST-DNA

binding domain is observed (Figure 15b). However, when the band corresponding to the biotinylated p65 bound to the DNA binding domain is compared to the input sample, the band is observed to be fainter for the bound biotinylated p65. This result suggests that only a fraction of the biotinylated p65 is active. Furthermore, this data also suggests that the DNA binding domain of LSF can bind recombinant p65.

ELISA Development to Quantify the Amount of Active Biotinylated p65

In order for the peptide binding assay to function as intended, a large percentage of the biotinylated p65 must be active. This issue is compounded by the fact that the donated p65 is not pure because in order to calculate the optimal amount of biotinylation reagent to use in the reaction, the exact concentration of p65 must be known. This cannot be determined easily from a partially purified sample. This presents a problem because excessive biotinylated can render the protein inactive, while insufficient biotinylation will not generate enough biotinylated p65 for use. Because it is difficult to determine how much of the p65 sample is biotinylated in the western blot, I began developing an ELISA to quantify the amount of biotinylated active p65.

The ELISA plate was coated for 48 hours at 4°C with 1 pmol of avidin per well in a 1xPBS coating buffer, to generate a surface that could capture biotinylated p65. This coating allows only biotinylated protein to be captured, while non-labeled p65 is washed off, ensuring that the signal produced from the assay is due solely to binding of biotinylated p65 and does not include any significant contribution from unlabeled p65.

The plate was washed three times with 1xPBST before blocking with 1% BSA in 1xPBST for 16 hours at 4 °C. The plate was washed three times and then incubated with 10 ng/µL of biotinylated p65 for 16 hours at 4°C. p65 was biotinylated according to the same conditions used for the previously described western blot. The plate was washed three times and then incubated with 50 µL of 10 nM or 100 nM GST-DNA binding domain for 1 hour at room temperature. GST-DNA binding domain was used because no antibody currently exists for the DNA binding domain of LSF. The GST-DNA binding domain solves this problem, however, because the DNA binding domain can bind p65 while the GST tag can be detected by a GST antibody. Two different GST-DNA binding domain concentrations were used to probe for the ends of a dose response curve if the assay succeeded. The plate was washed three times and then incubated with 100 µL of 0.4 ng/µL anti-GST antibody (Santa Cruz Biotech; product number: sc-459) for 30 minutes at room temperature. The anti-GST antibody recognizes the GST-DNA binding domain bound to the biotinylated p65 immobilized on the avidin coated plate surface, and can be detected by an anti-rabbit IgG-HRP antibody. The secondary antibody used to detect the rabbit anti-GST antibody was the same used for the previously performed western blots. The plate was washed three times and incubated with a 1:4000 dilution of anti-rabbit IgG-HRP antibody for 30 minutes at room temperature. After incubation, the plate wash again washed three times and then incubated with 50 µL of 3,3’,5,5’-Tetramethylbenzidine (TMB) for 15 minutes at room temperature. TMB is oxidized by horse radish peroxidase (HRP) and causes a color change that can be quenched and measured. The TMB reaction was quenched after 30 minutes with 50 µL of 0.2 M H_2SO_4

and the color change was measured at an optical density of 450 nm. A series of controls were included on the plate, each consisting of the reaction minus one assay component removed per control sample. When the experimental samples were compared to the controls, it was observed that the majority of the experimental sample signals were due to background from the GST-DNA Binding domain and the α-GST (Figure 16). This is because the signal for the 100 nM GST-DNA binding domain in the absence of p65 was equal to the 100 nM GST-DNA binding domain with p65, suggesting that the signal observed for the 100 mM GST-DNA binding domain experimental sample was really attributed to the non-specific binding of the GST-DNA binding domain and the anti-GST antibody. Furthermore, the anti-GST antibody contributes at least half of the background signal presented by the GST-DNA binding domain suggesting the background signal solely from the GST-DNA binding domain is not as high. This can be determined by the high signal for anti-GST antibody in the absence of the DNA binding domain. This means that part of the background signal attributed to the DNA binding domain is due to the non-specific binding of the anti-GST antibody.

For the next ELISA, I decided to test multiple biotinylation conditions as well as a range of anti-GST and anti-rabbit IgG-HRP antibody dilutions. This protocol allowed me to optimize the biotinylation reaction for different conditions using a single ELISA assay. The test biotinylation conditions were (i) a 40-fold M excess of biotinylation reagent, (ii) a 10-fold M excess of biotinylation reagent, and (iii) a 20-fold M excess of biotinylation reagent which was the same condition as used in the previous assay. A condition involving a ten minute reaction with a 20-fold M reagent concentration was also tested.

The assay procedure stayed the same as the last ELISA with the exception of a 1:15,000 dilution for the anti-rabbit IgG-HRP antibody. There was no observable difference in the signal obtained among the biotinylation conditions though their overall signal was not high and did not exceed that of the GST-DNA binding domain control (Figure 17). Based on this experiment, 0.4 ng/mL of anti-GST antibody and a 1:15,000 dilution of anti-rabbit IgG-HRP antibody were selected as optimal concentrations, due to their low level of background signal (Figure 17). The 1:40,000 dilution of anti-rabbit IgG-HRP antibody resulted in a lower background signal than the 1:15,000 dilution but it was a concern that this dilution would produce little to no signal from the experimental samples if used for future ELISAs.

Since little to no signal was observed for the experimental samples when compared to the background signal, I decided to use multiple amounts of p65 per condition. This would produce a dose response curve, and the biotinylation condition that would lead to the more active p65 could be selected. The p65 amounts selected for each condition were 5 µg, 1 µg, and 0.2 µg. The assay procedure was the same as in the previous experiment except that 10 nM GST-DNA binding domain was used instead of 100 nM. The control samples were designed to determine the background signal in the assay. They were (i) 10 nM GST-DNA binding domain with 0.4 ng/mL anti-GST antibody, and 1:15,000 anti-rabbit IgG-HRP, (ii) 0.4 ng/mL anti-GST antibody and 1:15,000 anti-rabbit IgG-HRP, and (iii) 1:15,000 anti-rabbit IgG-HRP. The change in the amount of p65 added to the plate did not significantly affect the signal (Figure 18). However, the signal did exceed the 10nM GST-DNA binding domain background

control. Thus, 5 µg of p65 gave the same signal as 0.2 µg of p65 which does not make sense since an increase in bound p65 should allow for more GST-DNA binding domain to bind conferring a higher signal. One possible explanation for this result is that 1 pmol per well of avidin is not sufficient enough to capture more than 0.2 µg of p65 and therefore, each well would be representative of the result if only 0.2 µg p65 was captured even if 5 µg was added to the well. 1 pmol of avidin in a 100 µL solution is 10^{-8} M avidin and 0.2 µg of p65 is 3.08 x 10^{-8} M p65. Since one molecule of avidin can bind four molecules of streptavidin (Livnah *et al.* 1993), the amount of p65 1 pmol per well of avidin can optimally bind is 4 x 10^{-8} M. However, since a fraction of the avidin incubated on the surface of the plate will not be in a conformation to bind biotin, the amount of biotin 1 pmol per well of avidin can bind is slightly less than 4 x 10^{-8} M. If 5ug of p65 in a 100 µL solution is equal to 7.69 x 10^{-7} M and the avidin coating in the well can only bind slightly less than 4 x 10^{-8} M biotin, we would only expect around 4 x 10^{-8} M biotinylated p65 to be captured in the well; the remaining p65 would be washed away in the wash step. Therefore, the 1 pmol per well avidin coating can capture 0.2 µg of biotinylated p65 but a greater amount of protein will be reduced to about 0.2ug or higher. This means that any amount of biotinylated p65 greater than 0.2 µg will give the same signal as about 0.2 µg of protein.

In order to address the problem experienced in the previous ELISA experiment, I increased the amount of avidin per well to 100 pmol. This amount of avidin (10^{-6} M) should be able to bind at most 4 x 10^{-6} M biotinylated protein which is greater than the amount needed to bind 5 µg of p65 (7.69 x 10^{-7} M). Furthermore, in an attempt to

increase the amount of active biotinylated p65, I bound the LSF DNA binding domain without the GST tag to p65 prior to biotinylation. This would help increase the amount of active p65 because the presence of the DNA binding domain in the protein-protein interaction site on p65 would prevent any important amino acid residues involved in the binding of p65 to LSF from being biotinylated. However, this would likely cause both the DNA binding domain and p65 to become biotinylated but it was thought that the DNA binding domain would be stripped away from the p65 on binding to the avidin-coated plate since the affinity of biotin for avidin is high. This would then expose the amino acid residues protected by the DNA binding domain in the protein-protein interaction site on p65 so they could bind GST-DNA binding domain. It was also thought that the DNA binding domain would not contribute to the sample's signal because it lacked the GST fusion protein detected by the GST antibody. Finally, I also utilized an anti-p65 antibody (Santa Cruz Biotech; product number: sc-372) raised to the C-terminal of p65. This antibody should detect the total amount of p65 present in the well and since non-biotinylated p65 is washed away, it should detect only the amount of p65 that is biotinylated. This is important because by determining the amount of biotinylated p65 in the sample I can then determine how much biotinylated p65 is active by comparing the total amount of biotinylated p65 to the amount of biotinylated p65 that binds GST-DNA binding domain. In order to determine how much of the p65 antibody to use, I performed a dilution curve on an untreated plate coated with 1xPBST and blocked with 1% BSA in 1xPBST. A 1:15,000 dilution of anti-rabbit IgG-HRP antibody was used to detect the amount of anti-p65 antibody nonspecifically bound to the assay plate. A buffer control

and an anti-rabbit IgG-HRP antibody control were also performed. A 1:3,000 dilution of anti-p65 antibdoy was chosen from the dilution curve because it was the lowest dilution that still showed a small degree of background signal (Figure 19). I felt this was important because I did not want to dilute the detection antibody too much and lose potential signal from future experimental samples. I then performed the previous ELISA again with three new samples containing the biotinylated p65, anti-p65 antibody, and anti-rabbit IgG-HRP antibody. Signal was observed for the experimental samples containing GST-DNA binding domain but the signal observed for the samples containing α-p65 did not exceed background (Figure 20). It was interpreted that the 1% BSA in 1xPBST block was inefficient at preventing α-p65 from binding to the plate because the samples containing GST-DNA binding and detected by anti-GST antibody had significantly increased signal over the samples detected with anti-p65 antibody. This comparative increase in signal was not due to background signal as the GST-DNA binding domain control showed only a slight increase in signal over the buffer control. Furthermore, the samples detected by the anti-p65 antibody did not generate a significant amount of signal to exceed the background signal provided by the anti-p65 antibody control suggesting that the signal observed in the samples detected by the anti-p65 antibody are due to background.

Since my previous block was shown to be ineffective, I performed a pilot experiment testing different blocking conditions. I tested the ability of 5% BSA in 1xPBST, 5% sucrose in 1xPBST, 5% non-fat dried milk in 1xPBST, and Superblock (Pierce Biotechnology) to prevent non-specific binding of the anti-p65 antibody to the

plate's surface. The candidate blocking conditions were also compared to the 1%BSA in 1xPBST blocking condition for reference and to act as a control since 1%BSA in 1xPBST was already shown to be ineffective at preventing non-specific binding of anti-p65 antibody. I utilized 1:150, 1:300, 1:1,500, and 1:15,000 dilutions of anti-p65 antibody. A buffer control was also included and the dilution of anti-rabbit IgG-HRP antibody used to detect the anti-p65 antibody was 1:15,000. The blocking conditions varied in their ability to prevent anti-p65 antibody from non-specifically binding the plate's surface but Superblock performed the best since it had smaller overall background signal and tighter error bars (Figure 21). Now that I had optimized the blocking conditions for the p65 antibody, I performed the previous ELISA with 1:150 of anti-p65 antibody and substituted Superblock for the 1% BSA in 1xPBST blocking condition. The biotinylation conditions of 20-fold M excess biotinylation reagent and 4-fold M excess biotinylation reagent resulted in a higher signal than 100-fold M excess biotinylation reagent the reagent (Figure 22). However, the control sample of anti-GST antibody with biotinylated p65 gave the same intensity in signal as the experimental samples. This would indicate that the anti-GST antibody is non-specifically binding the plate, except that the anti-GST antibody control shows minimal background signal. A potential explanation for this result is that some of the GST protein tag was removed from the glutathione sepharose beads during Factor Xa cleavage and contaminated the DNA binding domain preparation. During the biotinylation reaction, this GST protein would be biotinylated alongside the DNA binding domain and the p65 and would then stick to the

avidin-coated plate. Since the biotinylated GST would be captured on the avidin-coated plate, it would be easily detected by the α-GST leading to high background signal.

Discussion

YY1 and p65 Binding to LSF is Influenced by the Addition of Phosphatase Inhibitors

An interesting result of this thesis is that binding of both YY1 and p65 to regions on LSF shifts when phosphatase inhibitors are added during the preparation of U2OS cell extracts. Data has shown that in the presence of phosphatase inhibitors, binding occurs with LSF's DNA binding domain (Figures 7 and 8). When phosphatase inhibitors are absent from the U2OS cell extract preperation, however, binding occurs with the ubiquitin-like domain of LSF (Figures 5 and 6). Furthermore, bacterially-expressed biotinylated p65 has been shown to bind LSF's DNA binding domain in the absence of phosphatase inhibitors (Figure 15a). Since the GST-LSF constructs are expressed individually in bacteria and because binding is not seen at both the DNA binding and the ubiquitin-like domains in the absence of phosphatase inhibitors (in the event that the phosphate group was directing protein-protein interaction formation) the cause of the shift in binding is most likely occurring in the mammalian cell extract protein samples. This is further supported by the binding of biotinylated, bacterially expressed p65 to the DNA binding domain since these data were obtained in the absence of phosphatase inhibitors, whereas binding to the DNA binding domain required the presence of

phosphatase inhibitors for U2OS cell extract protein samples. However, it is unlikely that both YY1 and p65 are being phosphorylated at similar sites and that this phosphorylation is shifting the binding of both proteins to the same site on LSF because YY1 and p65 are different and unrelated transcription factors. This is further supported by the fact that phosphatase inhibitors weren't used during the binding of biotinylated bacterially expressed p65 to the DNA binding domain since binding still occurred suggesting that another protein in the cell extract, other than YY1 or p65, is being phosphorylated. This phosphorylated protein would then be mediating the interaction between LSF and p65 based on its phosphorylation state. The simplest explanation then is that the shift in binding from the ubiquitin-like domain to the DNA binding domain or *vice versa* is part of a more complex mechanism or pathway, as discussed below.

Scaffold proteins allow for the localization, compartmentalization, and organization of signaling cascades (Pan *et al*, 2012). By utilizing a complex network of protein-protein interactions, they are capable of ferrying signals from transmembrane receptors to targets in the nucleus through the association of multiple signaling kinases and interactions with other scaffold proteins (Pan *et al*, 2012). Some scaffolds can interact with transmembrane receptors as is the case with KSR, which is responsible for organizing the ERK pathway with the heterotrimeric G-protein gamma subunit, while other scaffolds can directly affect gene expression as is the case with scaffold JIP1 and JNK signaling (Pan *et al*, 2012). Furthermore, scaffold proteins can be regulated themselves through post-translational modification, the most common of which is phosphorylation and dephosphorylation (Pan *et al*, 2012).

Docking proteins are similar to scaffold proteins in that they are responsible for localizing and organizing signaling molecules and they can also be regulated by post-translational modification (Pan *et al*, 2012). However, they differ in that docking proteins tend to organize a small number of proteins at once and not large signaling cascades and they do not alter the properties of the proteins they bind (Pan *et al*, 2012). Insulin receptor signaling is mediated by phosphorylated docking proteins known as insulin receptor substrates (IRS) that are responsible for binding multiple insulin signaling kinases (White, 1998). IRS proteins contain phosphotyrosine binding and pleckstrin homology domains that are responsible for recognizing and binding the insulin receptor (White, 1998) as well as multiple tyrosine residues that are phosphorylated and recognized by downstream signaling proteins containing SH2 and SH3 domains (Virkamaki *et al*, 1999). Furthermore, multiple adaptor proteins, transforming proteins, structural proteins, and enzymes interact with IRS proteins but at unknown sites that are predicted to be outside the currently identified IRS protein-protein interaction domains (Virkamaki *et al*, 1999). Adaptor proteins also bind multiple protein targets (Pan *et al*, 2012) and their association with docking proteins suggests that a range of higher order protein complexes may be possible through the interaction of adaptor and docking proteins and the proteins that bind them.

Since YY1 and p65 are expressed from different pathways and the exact signaling pathway(s) of LSF have not yet been elucidated, I hypothesize that YY1 and LSF as well as p65 and LSF are organized and localized by a docking protein. A scaffold protein could also be responsible, however, scaffold proteins handle large signaling cascades and

the simplest explanation is that only three proteins are binding together. It should be mentioned, however, that the data do not rule out the binding of LSF to YY1 or p65 as being part of a more intricate cascade in which case, a scaffold protein may be more appropriate. To continue the explanation, in its unmodified state the docking protein allows for the binding of the ubiquitin-like domain of LSF as well as either YY1 or p65. The compartmentalization of the two transcription factors in close proximity induces a protein-protein interaction between YY1 or p65 and the DNA binding domain of LSF. The formation of this protein-protein interaction is viable in solution but would occur at a faster rate on the docking protein due to its localization function. Interactions between LSF or p65/YY1 in solution to its binding partner on the docking protein would not occur due to steric hindrance from the docking protein. When phosphorylated, the docking protein would undergo a conformational change thus releasing LSF bound to either p65 or YY1. When dephosphorylated by a phosphatase, the docking protein would return to its unmodified state and be able to bind substrates again. The LSF-p65 or LSF-YY1 complex could then translocate to the nucleus and interact with promoter targets. Such a mechanism might be advantageous for LSF as a nuclear localization sequence has not yet been found for the protein. YY1 contains a nuclear localization sequence in its C-terminal region (McNeil *et al,* 1998) while p65 contains a nuclear localization sequence in its Rel homology domain (Perkins *et al*, 2007). It may be that LSF translocates into the nucleus by binding to a protein that already has the ability to do so. An alternate but related hypothesis is that LSF and p65 or YY1 may be bound by a regulatory complex similar to IκB in the NFκB pathway. However, the IκB complex binds related Rel proteins whereas

p65, YY1, and LSF are all unrelated, making this hypothesis unlikely. A third hypothesis is that dissociation of p65 and LSF from the docking protein does not occur. In this hypothesis, the docking protein is capable of binding both the ubiquitin-like domain and the DNA binding domain of LSF depending on its phosphorylation state. This hypothesis is inherently simpler than the first described hypothesis however if this is the case, the protein would be carried along with p65/YY1 and LSF during transcriptional regulation unless it dissociates at a later point due to other interactions or post-translational modifications.

One way to determine the identity of the potential docking protein is to use the GST-LSF bead binding assay with HPLC tandem mass spectroscopy (LC-MS/MS). In this assay, GST-LSF is bound to glutathione sepharose beads and then incubated with recombinant p65. The beads are washed to remove non-specific binding proteins and then incubated with U2OS cell extracts. Since LSF is bound to p65, the only proteins that will bind the LSF-p65 complex are proteins that can bind to both LSF and p65 as well as proteins that can bind to the sites on LSF and p65 that are not occluded from the complex's formation. The beads are washed again to remove non-specific binding proteins and then the protein complexes are eluted with reduced glutathione. The eluted proteins are then prepared for mass spectroscopy and separated by HPLC (Chandramouli and Qian, 2009). After separation, the proteins are fractionated with trypsin and analyzed by MS/MS (Chandramouli and Qian, 2009). The individual fragments are compared to a MS trypsin fragment database for identification (Chandramouli and Qian, 2009). GST, LSF, and p65 can be removed from the analysis and the remaining proteins could either

be considered candidate docking proteins or be ruled out depending on their identity. The remaining proteins that weren't removed could then be characterized further.

Binding Peptide Identification Assay Design

The binding peptide identification assay consists of five parts designed to identify the region of the LSF DNA binding domain amino acid sequence responsible for binding p65. The individual stages are Factor Xa digestion of GST-DNA binding domain, Endoproteinase AspN digestion of DNA binding domain, biotinylation of p65, peptide capture by bound p65, and peptide detection with LC ESI-MS. At this time, only peptide capture by bound p65 remains to be tested while biotinylation of p65 needs to be further optimized. To summarize the results of this thesis in terms of the assay, Factor Xa cleaves GST from GST-DNA binding domain most efficiently when the reaction is performed at 37°C for two hours while the GST-DNA binding domain is bound to glutathione sepharose beads in the presence of 1 ng/µL of Factor Xa. Furthermore, Endoproteinase AspN can only digest denatured DNA binding domain. The peptide fragment detection method was also validated as it can detect all DNA binding domain peptide fragments with the exception of the 3,960 Dalton fragment though this fragment may appear on the HPLC chromatogram.

Future work should focus on optimizing the biotinylation of p65. Currently, three biotinylation reagent concentrations are being pursued; 4-fold molar excess, 20-fold

molar excess, and 100-fold molar excess of biotinylation reagent. In Figure 22, the amount of biotinylated p65 captured and detected with anti-p65 antibody is greater for 4-fold molar excess and 20-fold molar excess samples than the 100-fold molar excess sample. However, the signals overall are low and do not reflect the signal expected for 1 µg of protein. Two possible explanations exist for this result. The first is that the amount of p65 in the sample is dramatically lower than expected and the current signal is the best 1 µg of this protein preparation can perform. The second is that because the biotinylation reagent is highly reactive, even in solid form, the biotinylation reagent could have exceeded its shelf life and have lost reactivity as a result. A solution for both explanations is the acquisition of fresh biotinylation reagent coupled with the purchase or expression of a mostly pure p65 protein preparation.

The ELISA in its current form is already ready for biotinylation optimization. The amount of avidin per well, concentration of GST-DNA binding domain, amount of α-GST, dilution of α-rabbit IgG-HRP, and the dilution of α-p65 at 100 pmol per well, 10nM, 0.4 ng/mL, 1:15,000, and 1:150 respectively has been found to be sufficient. With pure p65 biotinylated with new reagent, it is probable that the p65 biotinylation reaction can be optimized with the existing assay.

Once optimized, the biotinylated p65 can be added to streptavidin agarose beads and the full binding peptide identification assay can be performed. Bead-bound p65 would be incubated in a solution with LSF DNA binding domain peptide fragments produced from the Endoproteinase AspN digestion of the DNA binding domain. After binding, the streptavidin agarose beads would be centrifuged and the supernatant would

be purified with a C-18 spin column and analyzed by LC ESI-MS. The associated peaks in the HPLC chromatogram corresponding to the mass-to-charge ratios from the peptides analyzed by ESI-MS would represent the peptide fragments not bound to p65. Then, the streptavidin bound p65 would be denatured under moderate conditions so that p65 is denatured without the biotin-streptavidin interaction being affected. Though improbable due to its high affinity for streptavidin, if some of the biotinylated p65 were stripped away, it would not bind the C-18 resin in the subsequent purification step due to its larger size. The denaturation of the p65 causes the bound peptide fragments to be released into the solution. The peptides would then be purified by a C-18 spin column and analyzed by LC ESI-MS. A control sample of a complete Endoproteinase AspN LSF DNA binding digestion would also be analyzed by LC ESI-MS. Together, the combined chromatograms and mass to charge ratios resulting from the three LC ESI-MS experiments would show a complete picture of the location of the individual peptide fragments throughout the assay. Based on this information, one would be able to determine the specific LSF DNA binding domain peptides responsible for binding p65.

Though the assay hasn't been performed, I would expect the results to be located between amino acids 65 and 103 in the DNA binding domain sequence based on the bead assay results in Figure 6. Furthermore, this amino acid region is also present in the pGEX1 construct's expressed protein which binds p65. As such, it would be interesting if the data from the binding peptide identification assay agreed or disagreed with this result. If they agreed, it would suggest that the peptide sequence within the region of 65 and 103 is most likely responsible for binding p65. If it differs, then more experiments would be

needed to determine the reason for the discrepancy. However, the results of both experiments can be confirmed with a modified version of the p65 biotinylation optimization ELISA. In this assay, the ELISA would be performed as previously but with one altercation. The peptides identified by the binding peptide identification assay could be synthesized and added to the assay at a higher concentration than the GST-DNA binding domain and made to compete for the binding site on the p65. Since this is a competition assay, the loss of binding of GST-DNA binding domain would be detected and compared to a sample where the peptides were absent. By using this assay, a binding curve of LSF's binding to p65 could be generated and the function of the binding peptides could be confirmed.

Drugability of LSF

The FTMAP server allows for protein-protein binding site predictions using the FTMAP algorithm and submitted protein data bank (PDB) files (Brenke *et al*. 2009). The FTMAP algorithm mimics X-ray crystallography and NMR binding site prediction methods by computationally generating a series of organic probe molecules and then docking these molecules on the submitted structure through the simulated binding of 2000 molecular positions using rigid body docking (Brenke *et al*. 2009). The position of the probe molecules is refined through energy minimization and the individual probe molecules are clustered together based on location (Brenke *et al*. 2009). The average free energy of the clusters is then ranked for comparison (Brenke *et al*. 2009). Clustering of

probe molecules is achieved by selecting the probe molecule with the lowest free energy and then clustering all probe molecules in a 3 angstrom radius (Brenke *et al*. 2009). The clusters with the lowest free energy of binding are considered to be more favorable protein-protein binding sites (Brenke *et al*. 2009).

I used the FTMAP server to analyze a LSF DNA binding domain homology structure created by Chetanya Pandya and based on the homology modeling and fold recognition data from Kokoszynska *et al*. (2008). Twelve clusters were generated from the FTMAP server and were numerically ranked based on their free energy of binding. Cluster 1 had the lowest free energy of binding and is located near amino acids 29, 42. 134, 135, 183, and 185 of the homology model's amino acid sequence. Cluster 2 is located near amino acids 17, 19, 21, 22, 23, 24, 33, 34, 36, 44, and 191. Cluster 3 is located near amino acids 134, 135, 136, 138, 183, and 184. Cluster 4 is located near amino acids 6, 8, 14, 16, 34, 39, and 179. Cluster 5 is located near amino acids 100, 121, 124, 130, and 131. Cluster 6 is located near amino acids 16, 19, 21, and 34. Cluster 7 is located near amino acids 5, 6, 67, 69, 73, 82, 156, and 178. Cluster 8 is located near 6, 7, 8, 156, 176, and 178. Cluster 9 is located near amino acids 54 and 161. Cluster 10 is located near amino acids 42, 44, 128, and 130. Cluster 11 is located near amino acids 26, 39, 71, 72, 139, 140, 150, 182, and 194. Cluster 12 is located near amino acids 5, 6, 71, 73, 178, and 180. It is important to consider whether any of the clusters exist in the regions of the sequence responsible for DNA binding since inhibitor design targeting the site may lead to a decrease in DNA binding. Dr. Barbara Ludeke, Christina Hao, and Steven Kim have mapped the proposed LSF DNA binding sites on the DNA binding

domain homology model based on the effects of point mutations on LSF DNA binding in electrophoretic mobility shift assays. Based on her analysis, the proposed DNA binding sites of LSF are amino acids 194, 199, 202, 205, 224, 236, 241, 244, 245, and 246. Only cluster 11 contains a proposed DNA binding site at amino acid 194. These results show that a range of potential target sites can be utilized for inhibitors thus increasing the drugability of LSF.

Conclusion

While the protein-protein binding site(s) of LSF wasn't definitively localized to a specific small stretch of the LSF sequence, significant progress was made. I found that both YY1 and p65 bind the DNA binding domain of LSF and that this interaction most likely takes place between amino acids 65 and 103. I also designed and provided preliminary optimization data for an assay to identify the specific binding sequence for the binding of LSF to p65. This assay makes use of protease digestions, p65 mediated peptide capture, and peptide detection using LC ESI-MS. This assay can easily be converted to use with YY1 as well. In this case, YY1 would need to be biotinylated and bound to streptavidin beads but the basic approach is the same. When this assay is completed, I'm confident that the specific LSF sequences required to bind p65 and YY1 will be identified.

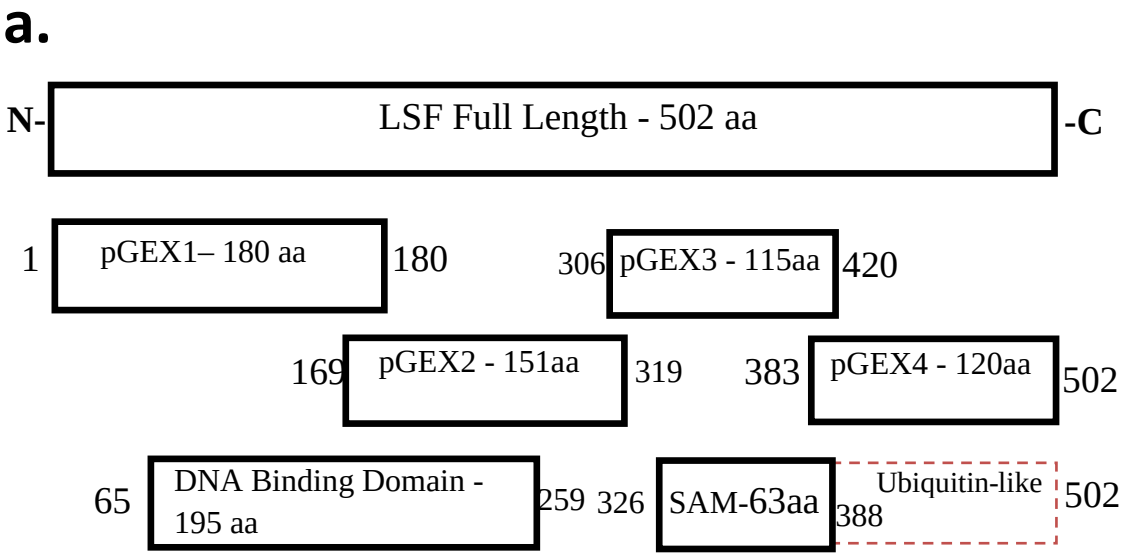

Figure 1. LSF construct information. a) Overlapping LSF DNA constructs and projected coded LSF sequences. The cDNA sizes are as follows: pGEX1 540bp; pGEX2 450bp; pGEX3 342bp; pGEX4 360bp; DNA binding domain 582bp; SAM domain 183bp. The pGEX1 and pGEX2 constructs overlap between the 169th amino acid and the 180th amino acid of the LSF sequence. This sequence range is also shared by the LSF DNA Binding Domain construct as well as parts of the pGEX1 and pGEX2 constructs between the 65th amino acid and the 259th amino acid of the LSF DNA sequence. The pGEX3 construct contains the SAM construct as well as regions on the 5' and 3' side of the construct. The ubiquitin-like domain was not independently made into a construct as the DNA Binding Domain and SAM domain but it is represented by the pGEX4 construct.

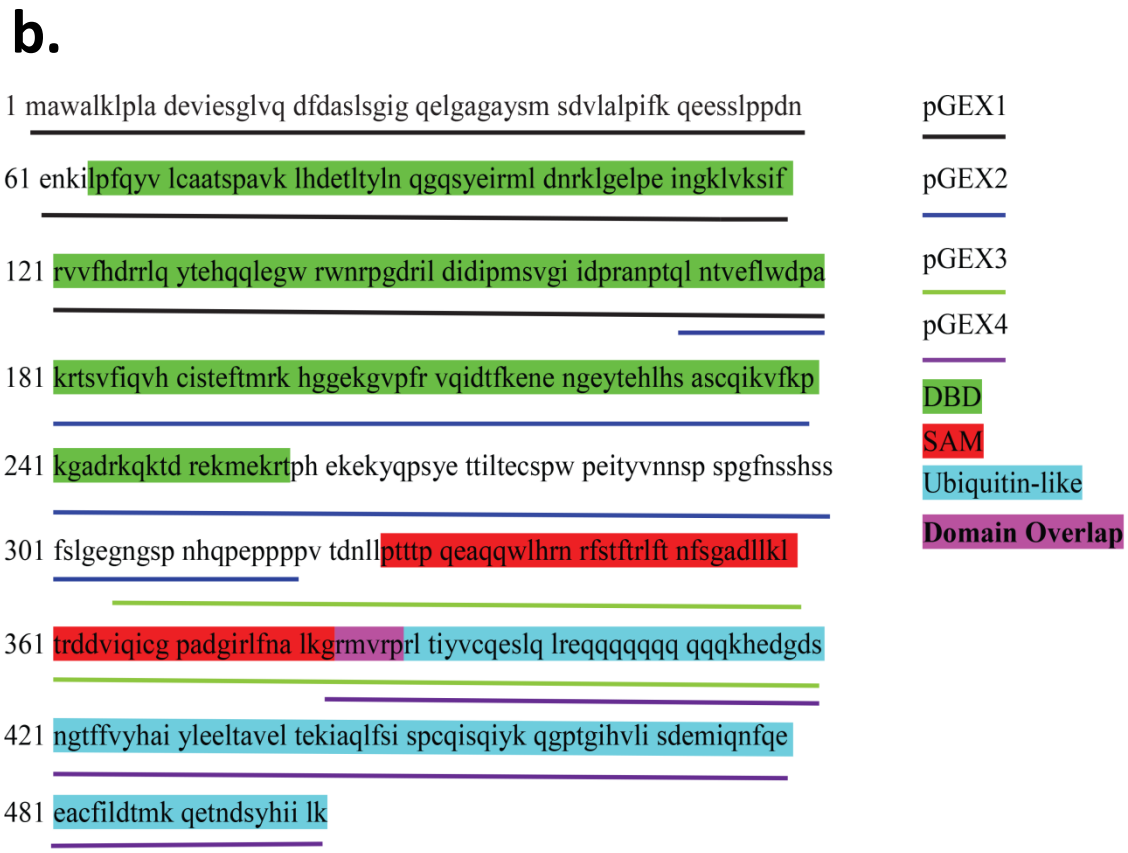

Figure 1. LSF construct information. b) Amino acid sequence of LSF with domain information and overlapping regions in FASTA format. Domain information is represented as predicted in Kokoszynska *et al*. The sequence is represented from the N-terminal region to the C-terminal.

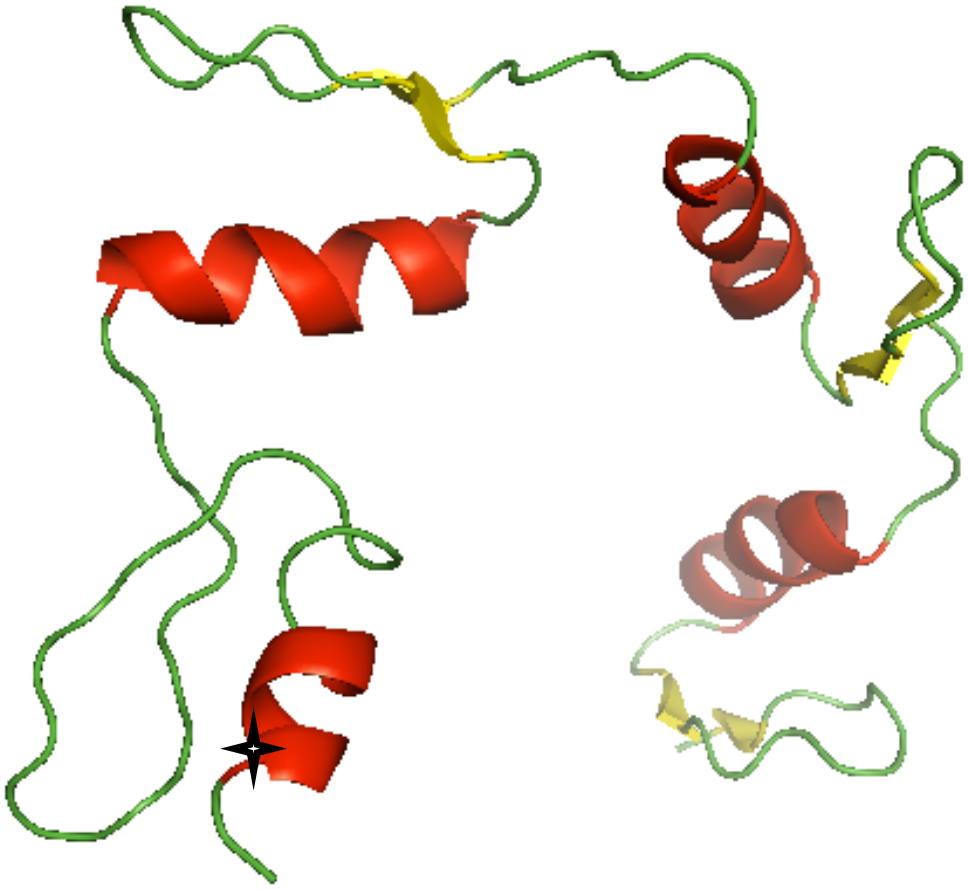

Figure 2. Secondary structure of YY1. Ribbon representation of YY1's secondary structure from Structure of YY1 Bound to the Adeno-associated Virus Initiator by Houbaviy *et al*, 1996 (PDB ID: 1UBD) binding a 20 base pair oligonucleotide (hidden in figure). Looped regions are in green, α-helices are in red, and β-turns are in yellow. The N-terminal of the protein is indicated by the black star.

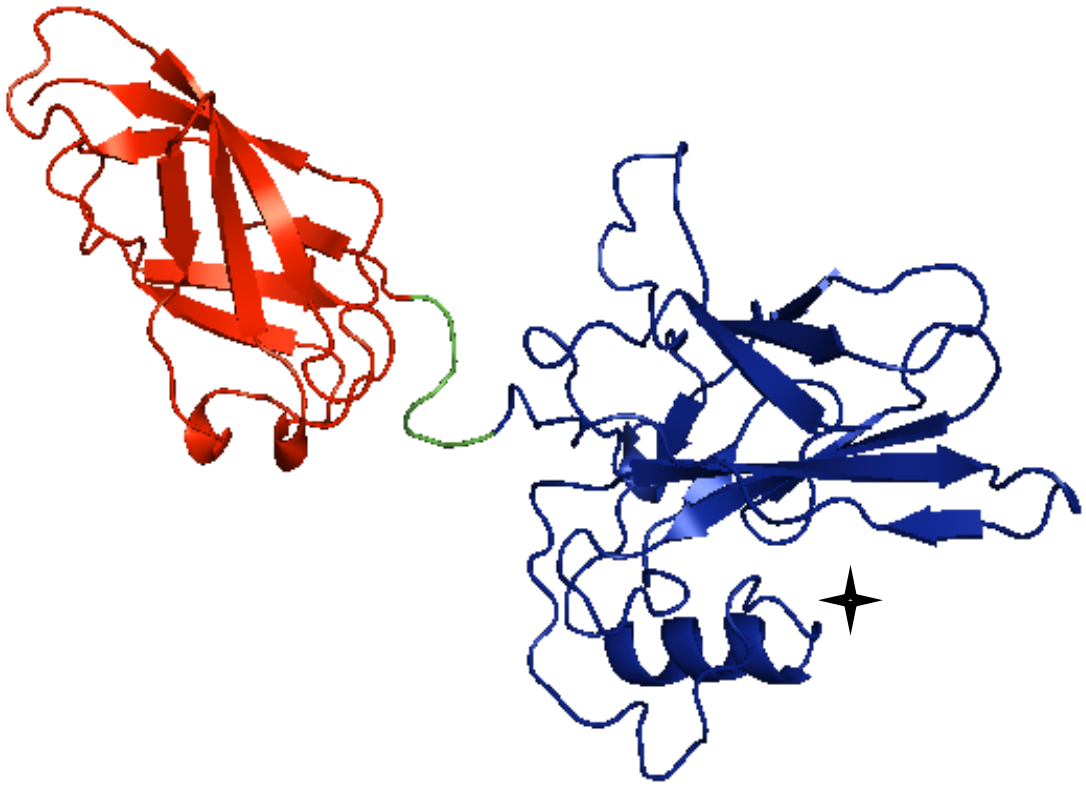

Figure 3. Crystal structure of p65's rel homology domain. Ribbon representation of the rel homology domain as shown in Structure of NF-kB 50/p65 Heterodimer Bound to the PRDII DNA Element from the Interferon-Beta Promoter by Escalante *et al*, 2002 (PDB ID: 2I9T). The black star indicates the N-terminus of the domain. The structure in blue is the DNA binding domain which is separated by a small linker region from the dimerization domain shown in red. The C-terminal transactivation domain is not shown as it was not included in the crystal structure.

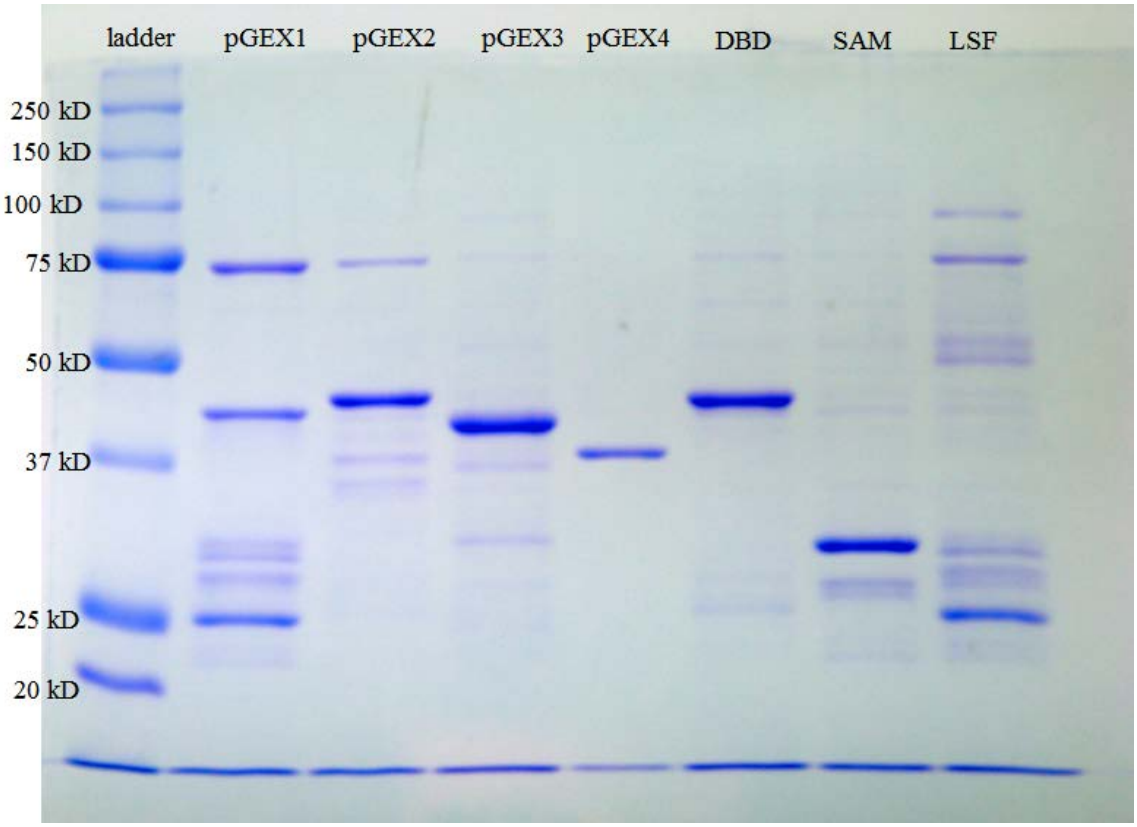

Figure 4. Characterization of GST-LSF fusion proteins by SDS-PAGE. Solubilized GST-fusion proteins were mixed with SDS sample buffer and loaded on a discontinuous 10% SDS-PAGE gel. The molecular weight ladder used was Precision Plus Prestained Ladder from Bio-Rad. Predicted molecular weights for protein species are GST-pGEX1: 45.8 kDa; GST-pGEX2: 42.5 kDa; GST-pGEX3: 38.5 kDa; GST-pGEX4: 39.2 kDa; GST-DBD: 47.3 kDa; GST-SAM: 32.8 kDa and GST-LSF: 81.4 kDa. DBD stands for LSF's DNA Binding Domain. 500ng of protein were loaded for each well.

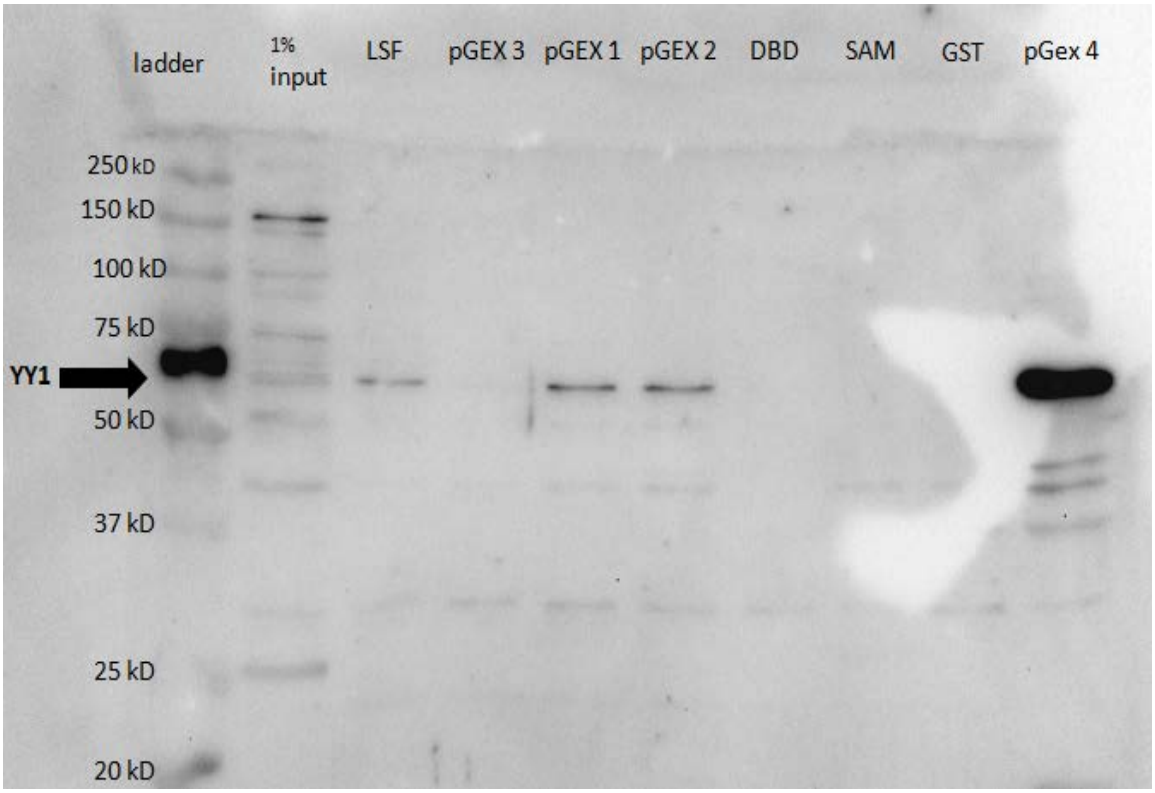

Figure 5. Binding assay of LSF with YY1. Mammalian U2OS cell extracts were incubated with GST-LSF fusion proteins on glutathione beads overnight at 4˚C. Resulting protein complexes were eluted from the binding assay and electrophoresed through on a 10% SDS-PAGE gel. The gel was then immunoblotted with α-YY1. The molecular weight of YY1 is approximately 58 kDa. YY1 binding occurs to the proteins encoded by the pGEX1 and pGEX2 constructs but primarily with the pGEX4 encoded protein. The pGEX4 construct codes for the ubiquitin-like domain of LSF. DBD stands for the DNA binding domain of LSF. GST corresponds to empty pGEX-5x-1 vector and encodes only GST protein. The amino acid range of the encoded proteins for the fusion constructs are as follows: LSF 1-502; pGEX1 1-180; pGEX2 169-319; pGEX3 306-420; pGEX4 383-502; DBD 65-259; SAM 326-388.

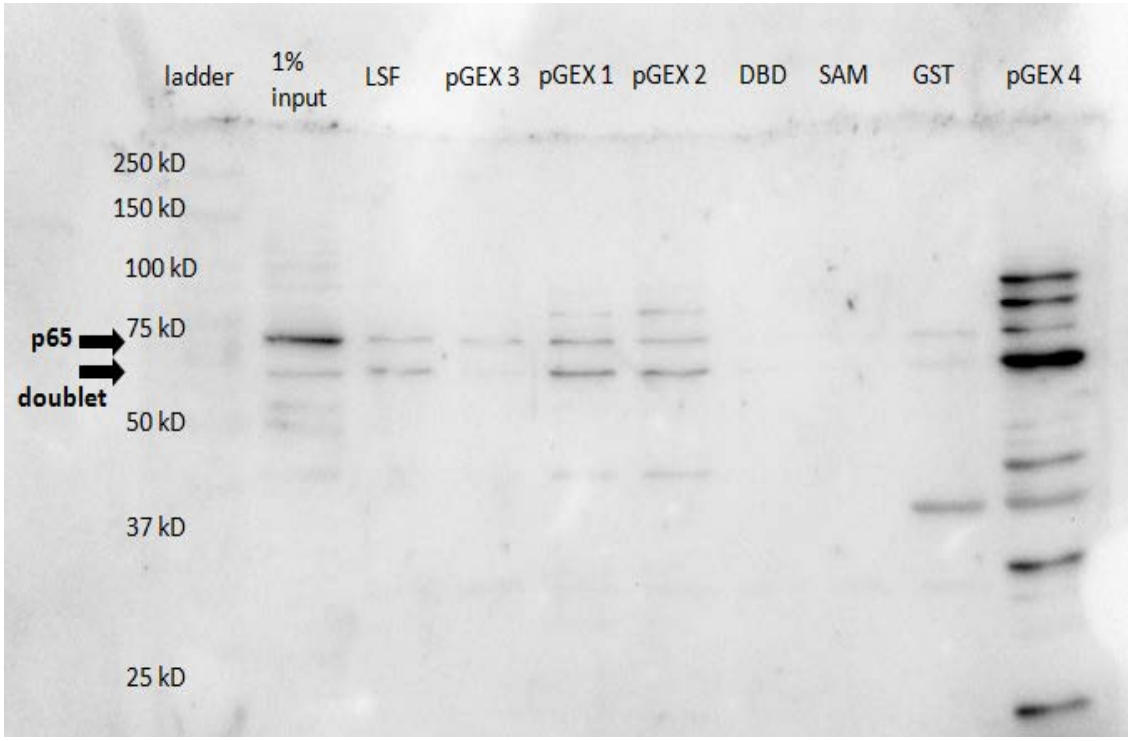

Figure 6. Binding assay of LSF with p65. Mammalian U2OS cell extracts were incubated with GST-LSF fusion proteins on glutathione beads overnight at 4˚C. Protein complexes were eluted from the binding assay and loaded on a 10% SDS-PAGE gel. The gel was then transferred to a PVDF membrane and blotted with α-p65. The molecular weight of p65 is 65 kDa. The binding assay results in a doublet for p65 where the lower band is a degradation product; both bands have been indicated on the blot with arrows. Binding occurs at the pGEX1 and pGEX2 constructs but primarily at the pGEX4 construct. The pGEX4 construct is responsible for expressing the ubiquitin-like domain of LSF. DBD stands for the DNA binding domain of LSF. GST corresponds to empty pGEX-5x-1 vector and encodes only GST protein. The amino acid range of the encoded proteins for the fusion constructs are as follows: LSF 1-503; pGEX1 1-180; pGEX2 169-319; pGEX3 306-420; pGEX4 383-503; DBD 65-259; SAM 326-388.

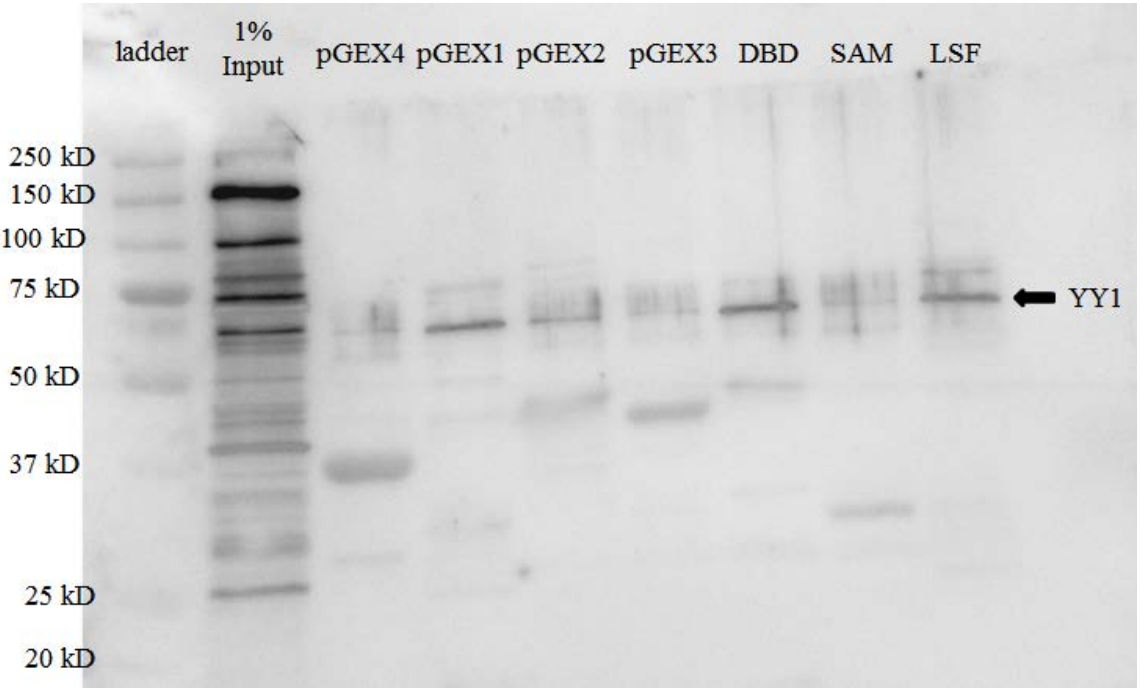

Figure 7. Binding assay of LSF with YY1 in the presence of phosphatase inhibitors. U2OS cell extracts were treated with phosphatase inhibitors prior to incubation with immobilized GST-fusion proteins. Protein complexes were eluted from the glutathione Sepharose beads with reduced glutathione and loaded on a SDS-PAGE gel. The gel was transferred to a PVDF membrane and probed with α-YY1. The molecular weight of YY1 is approximately 58 kDa. Blot shows binding of YY1 to the pGEX1 construct as well as the DNA Binding Domain of LSF.

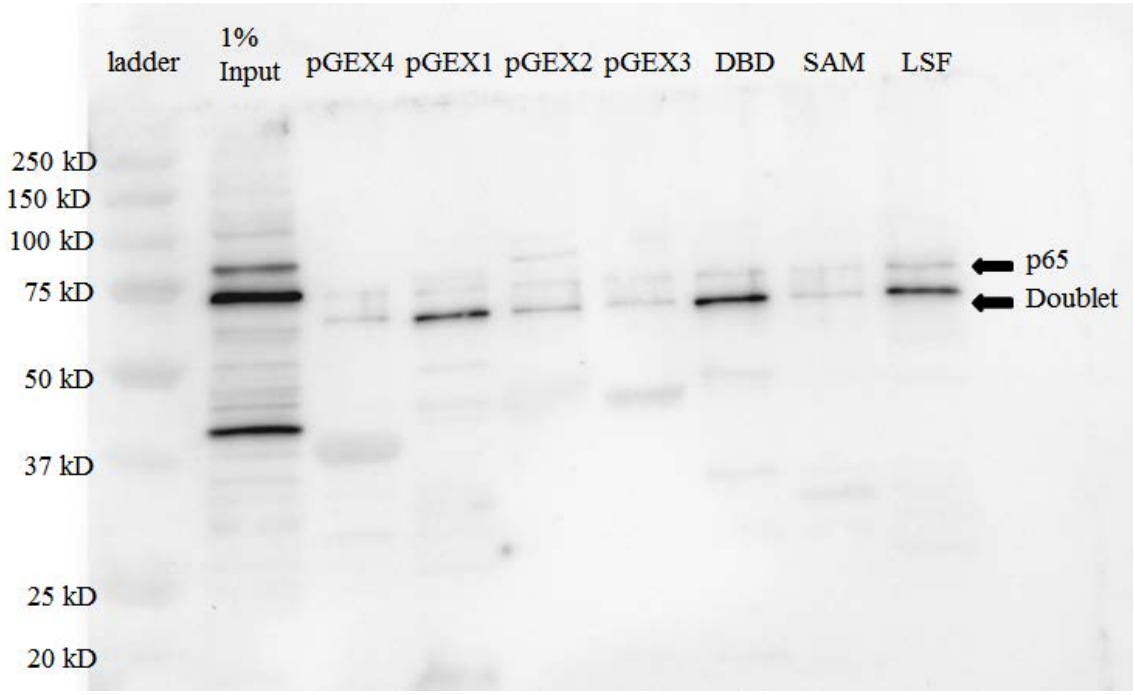

Figure 8. Binding assay of LSF with p65 in the presence of phosphatase inhibitors. U2OS cell extracts were treated with phosphatase inhibitors prior to incubation with immobilized GST-fusion proteins. Protein complexes were eluted from the glutathione Sepharose beads with reduced glutathione and loaded on a SDS-PAGE gel. The gel was transferred to a PVDF membrane and probed with α-p65. The molecular weight of p65 is 65 kDa. Blot shows binding of p65 to the pGEX1 construct as well as the DNA Binding Domain of LSF.

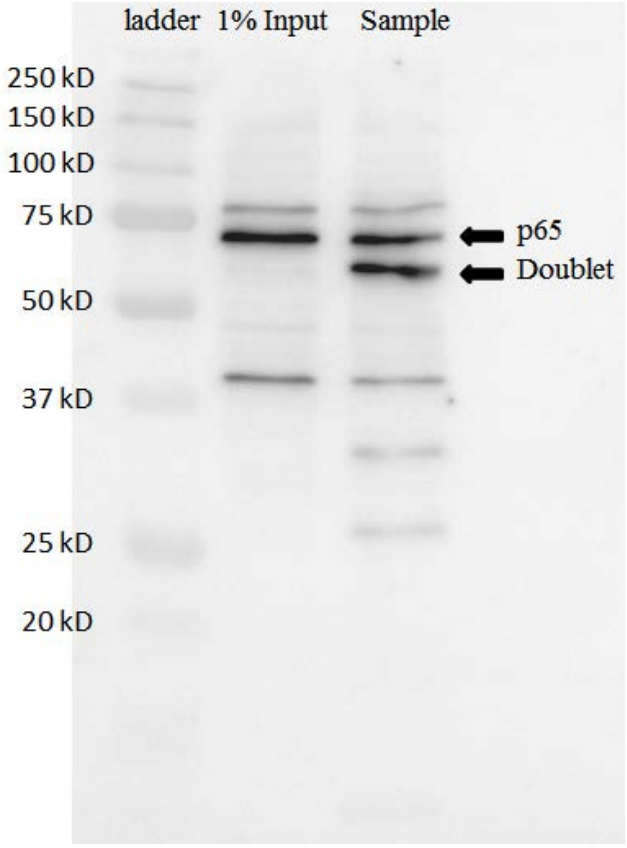

Figure 9. Two bands results from p65 due to bacterial protease bound to the glutathione resin. Lysed and solubilized BL21 T7 Express lysY/Iq competent *E. coli* extracts were incubated with glutathione Sepharose beads for 30 minutes at room temperature in order to mimic the conditions of the GST-LSF binding assay. The beads were then washed three times with sterile 1xPBS before being incubuated with 300 µL of U2OS cell extracts for 1 hour at 4°C. The beads were then centrifuged and the supernatant was analyzed on a 10% SDS-PAGE gel. A western blot was performed and the membrane was blotted with α-p65. The lower molecular weight band shows that the doublet is due to a potential bacterial protease non-specifically bound to the bead resin as the band is absent from the input control.

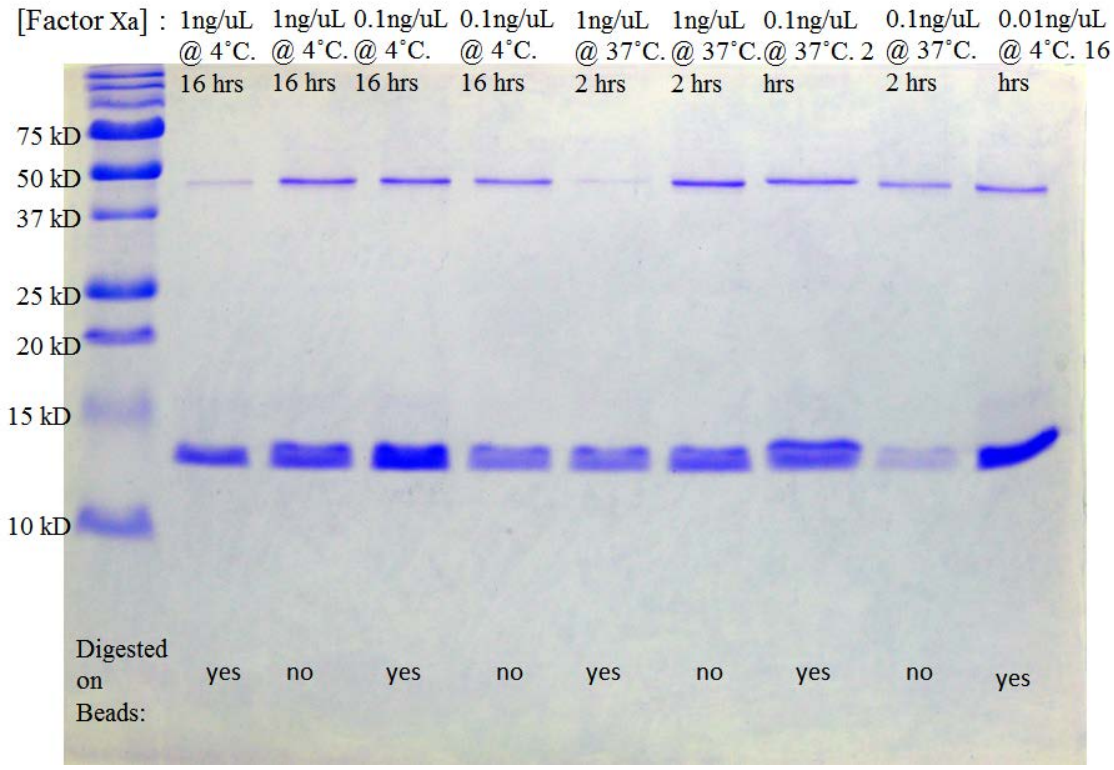

Figure 10. Factor Xa reaction optimization. GST-DNA binding domain fusion protein was cleaved with Factor Xa protease under various conditions. GST-DNA binding domain protein was captured on glutathione sepharose beads prior to digestion. For samples where digestion of the GST protein was performed in solution, the GST-DNA binding domain protein was first eluted with 50 mM Tris-HCl pH 8.0, 10 mM reduced glutathione for 5 minutes at room temperature. The samples were mixed with SDS-sample buffer, and analyzed on a 15% SDS-polyacrylamide gel with Coomassie Brilliant Blue. The protein band for GST-DNA binding domain is 47 kDa while the protein band for the DNA binding domain is about 13 kDa. Bead digestion with 1 ng/µL of Factor Xa for 2 hours at 37°C is the most optimal reaction condition.

Figure 11. ExPASy peptide cutter predicts cleavage sites for Endoproteinase AspN in the LSF DNA binding domain sequence. The LSF DNA binding domain amino acid sequence was submitted to the ExPASy peptide cutter tool. The program predicted 11 cleavage sites for Endoproteinase AspN with an average size of 21 amino acids for the peptide fragments. The peptide fragment map was used to predict the peptide fragments and Endoproteinase AspN cleavage sites for the binding peptide identification assay.

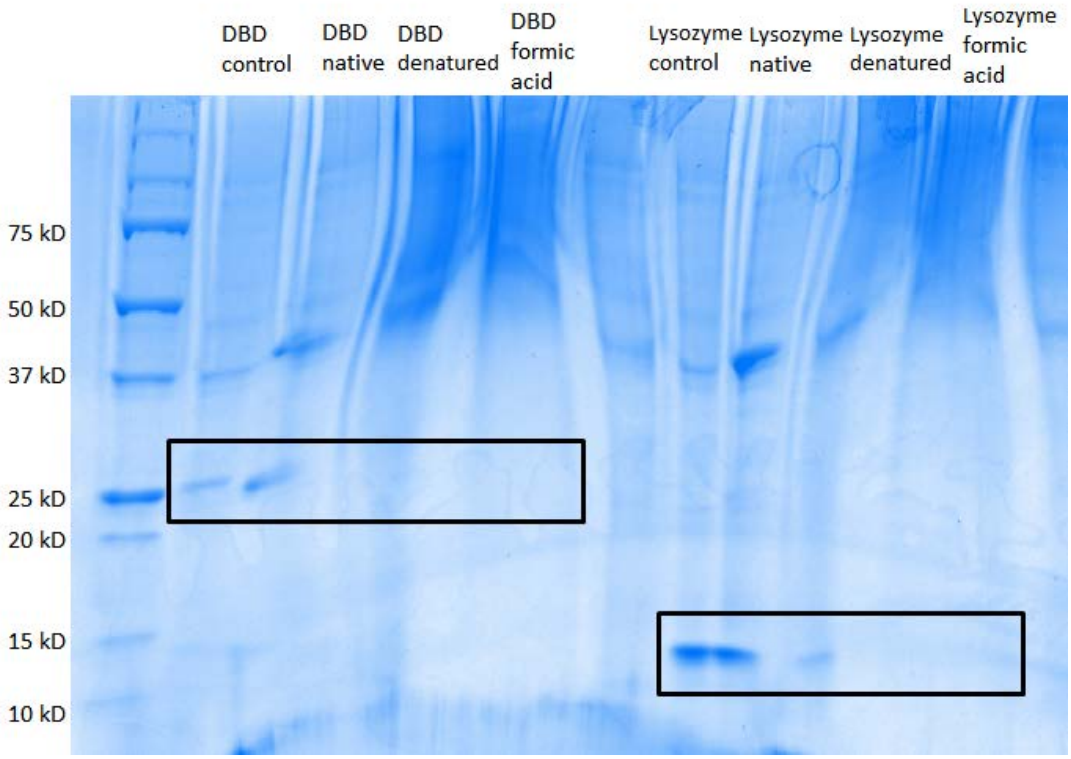

Figure 12. Endoproteinase AspN digests denatured DNA binding domain and lysozyme but not native protein. 1 µg/µL of Endoproteinase AspN was incubated in a 10uL solution containing 50 mM Tris-HCl pH 8.0 with 2.5 mM $ZnSO_4$ and 1 µg of LSF DNA binding domain or lysozyme for 16 hours at 37°C. For denatured samples, the DNA binding domain or lysozyme was first denatured in 6 M guanidium-HCl and then subsequently diluted to 2 M guanidium-HCl prior to digestion. The samples were analyzed on a denaturing polyacrylamide gel with Coomassie Brilliant Blue stain. The warping of the gel is due to the guanidium-HCl present in the denatured samples. 70% formic acid successfully cleaves lysozyme and DNA binding domain and Endoproteinase AspN cleaves denatured DNA binding domain and lysozyme as indicated by the absence of a protein band at 15 kDa.

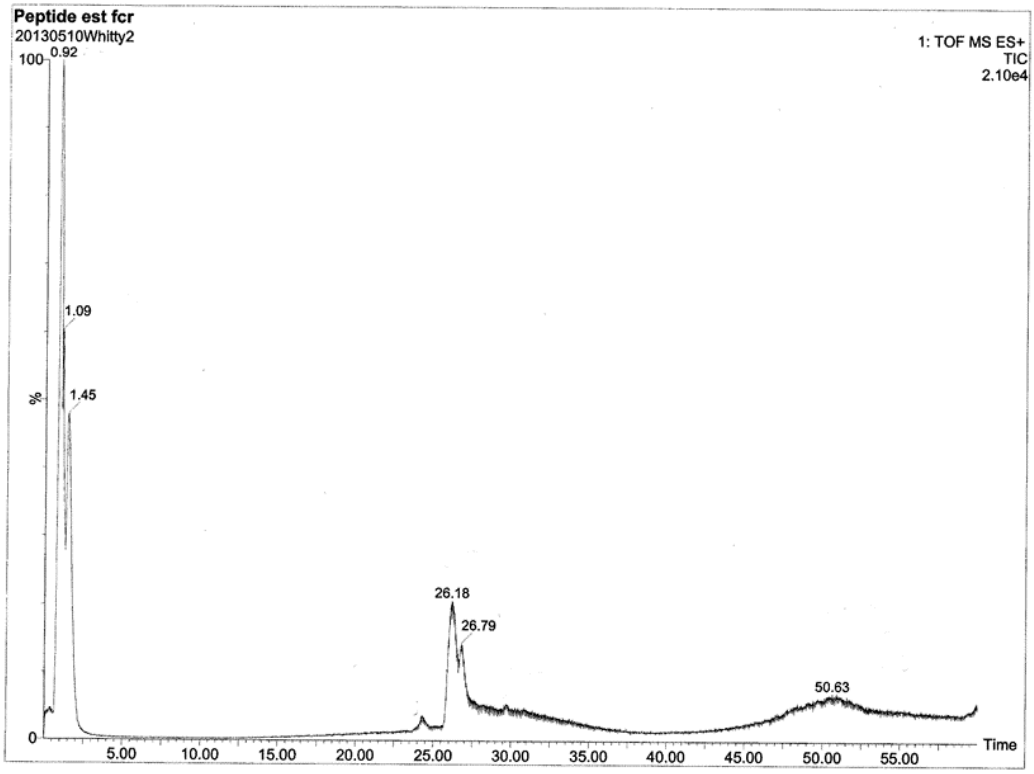

Figure 13. HPLC Chromatogram of LSF DNA binding domain peptide fragments. Chromatogram is representative of peptide peaks eluted from a reverse-phase HPLC column. The x-axis is in minutes. Denatured LSF DNA binding domain was digested by Endoproteinase AspN and purified on a C-18 spin column. The peptides were eluted from the spin column using a 1:1 solution of acetonitrile and water before being transferred to Dr. Norman Lee of the Chemical Instrumentation Center (Boston University) and loaded on a HPLC column. The figure shows that peptides produced from the binding peptide identification assay can be analyzed by HPLC coupled ESI-MS.

Peptide Size (daltons)	Chromatogram peak (minutes)	Position in LSF DBD sequence (aa)
3300	0.92	149-178
990	1.09, 50.63	185-194
2310	24.25	61-81
1980	24.25, 50.63	18-35
1650	50.63	82-96
1870	26.18, 26.79, 28.5	1-17
2750	50.63	36-60
1760	50.63	97-112

Table 1. Peptide fragments assigned to HPLC chromatogram peaks. LSF DNA binding domain peptide fragments obtained from mass to charge ratios from ESI-MS and their corresponding time signatures on the HPLC chromatogram displayed in Figure 13. Peptide fragment sizes were determined from the mass to charge ratios taken from the ESI-MS data in conjunction with the Endoproteinase AspN peptide fragment prediction map in Figure 11. The position in LSF DBD sequence refers to the peptide fragment positions in the sequence presented in Figure 11.

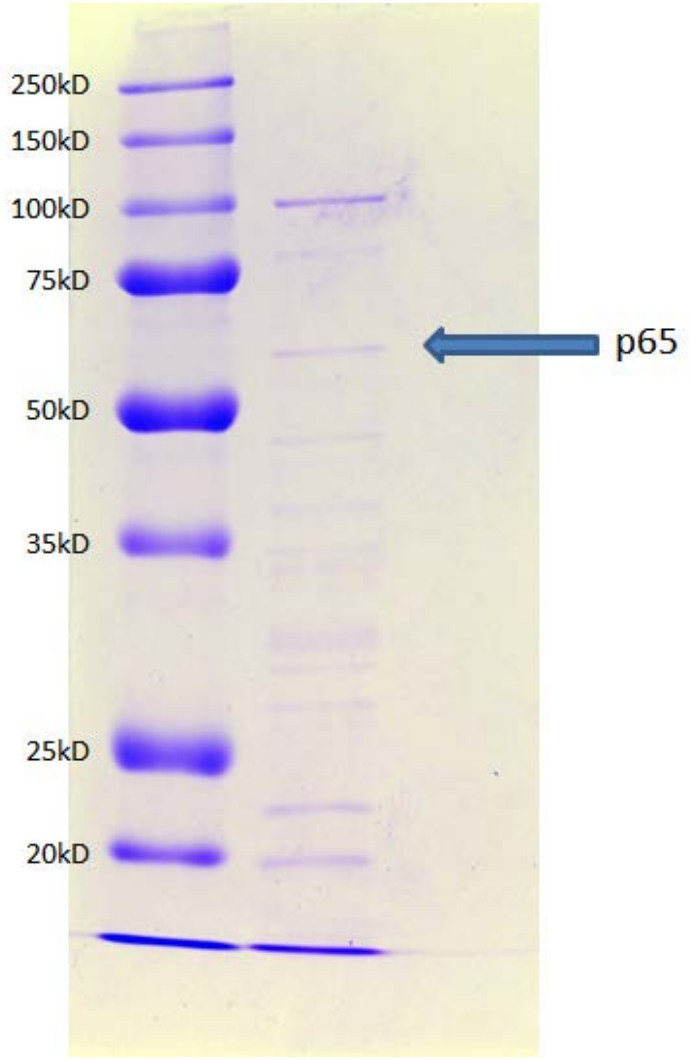

Figure 14. Coomasie blue analysis of p65 sample provided by Dr. Trevor Siggers on a 10% polyacrylamide denaturing gel. p65 protein obtained from Dr. Trevor Siggers of Boston University was mixed with SDS-sample buffer and analyzed by denaturing polyacrylamide gel electrophoresis with Coomassie Brilliant Blue staining. The protein preparation is only partially purified, with three other major protein bands besides the band correspond to p65; multiple minor bands are also present. This protein is used in all subsequent p65 biotinylation experiments.

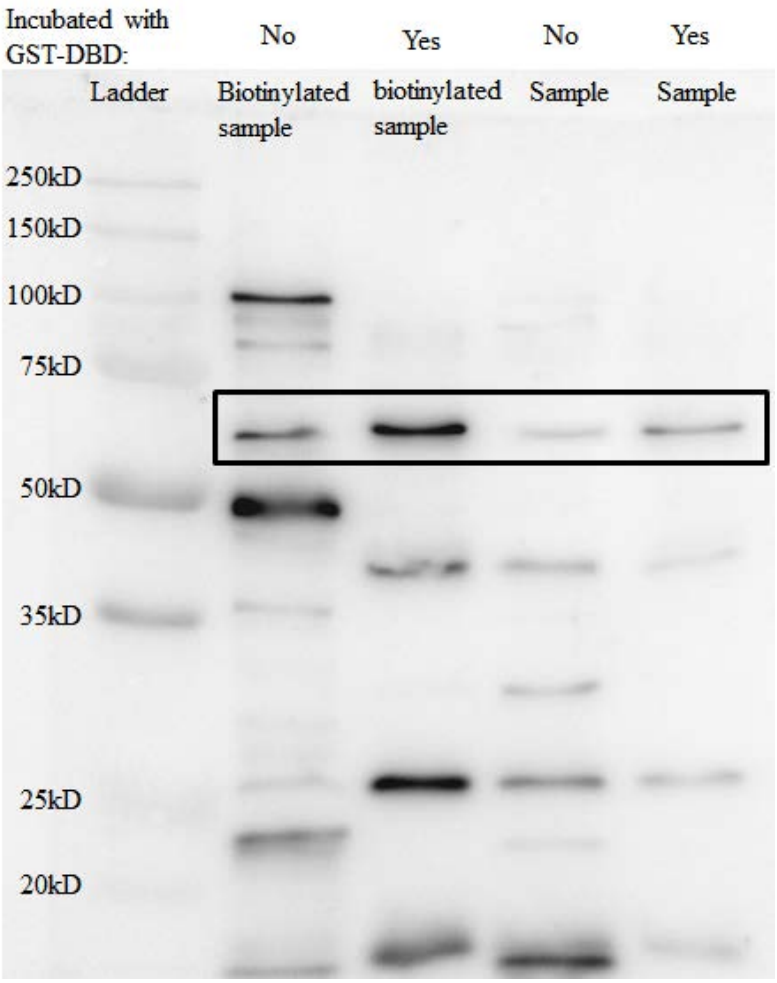

Figure 15. LSF-DNA binding domain binds p65. a). Bacterial expressed recombinant p65 protein was biotinylated and then incubated with GST-LSF DNA binding domain immobilized on glutathione sepharose beads. The bound proteins were eluted with reduced glutathione and separated on a denaturing polyacrylamide gel before being transferred to a PVDF membrane. The protein bands within the black rectangle belong to p65.The membrane was immunoblotted with α-p65. The figure shows that LSF DNA binding domain binds both biotinylated and non-biotinylated p65 and that the biotinylated p65 is active.

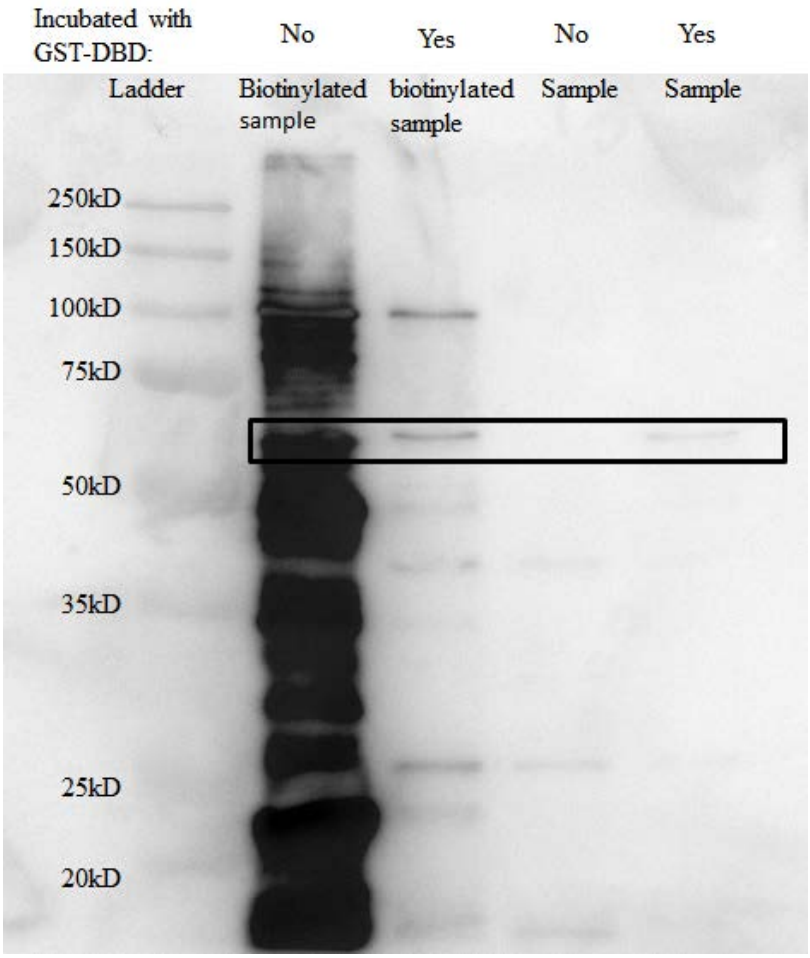

Figure 15. p65 bound to GST-LSF DNA binding domain is biotinylated. b) Bacterially expressed recombinant p65 protein was biotinylated and then incubated with GST-LSF DNA binding domain immobilized on glutathione sepharose beads. The bound proteins were eluted with reduced glutathione and separated on a denaturing polyacrylamide gel before being transferred to a PVDF membrane. The membrane was immunoblotted with streptavidin-HRP conjugate. The protein bands in the black rectangle represent biotinylated p65. This figure shows that the p65 bound to the GST-LSF DNA binding domain is biotinylated, however, when compared to Figure 15a, the amount of active p65 biotinylated is a small fraction of the overall p65 sample.

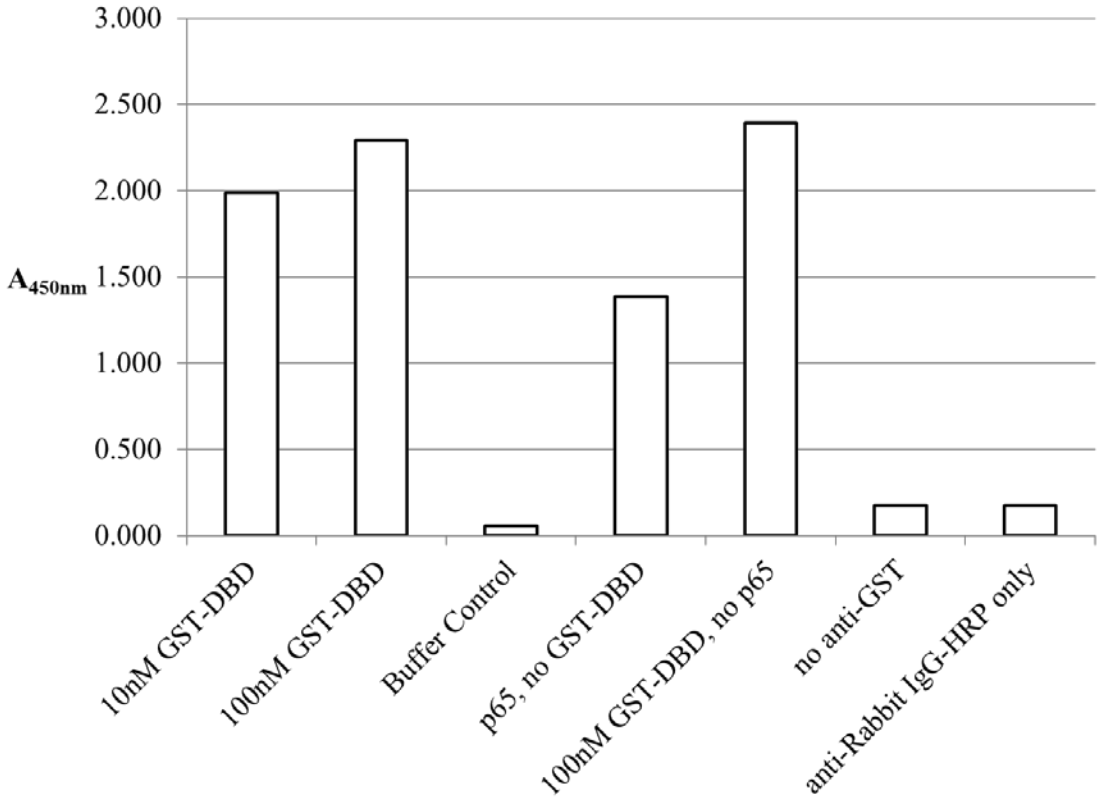

Figure 16. ELISA background signal is due to α-GST and GST-DNA binding domain. An ELISA was performed with 100 µL of 10 ng/µL biotinylated p65 captured on an immunoabsorbant 96-well plate coated with 1 pmol per well of avidin. 50 µL per well of 10 nM and 100 nM GST-DNA binding domain was bound to the p65 and detected with 100 µL of 0.4 ng/µL α-GST rabbit polyclonal. 100 µL of a 1:4,000 dilution of α-rabbit IgG-HRP was used to detect the α-GST. The color change of 50 µL TMB was quenched with 50 µL H_2SO_4 and measured at 450 nm on a plate reader. One out controls were used for each component. The figure shows that the majority of the signal is due to background from the GST-DNA binding domain and the α-GST components.

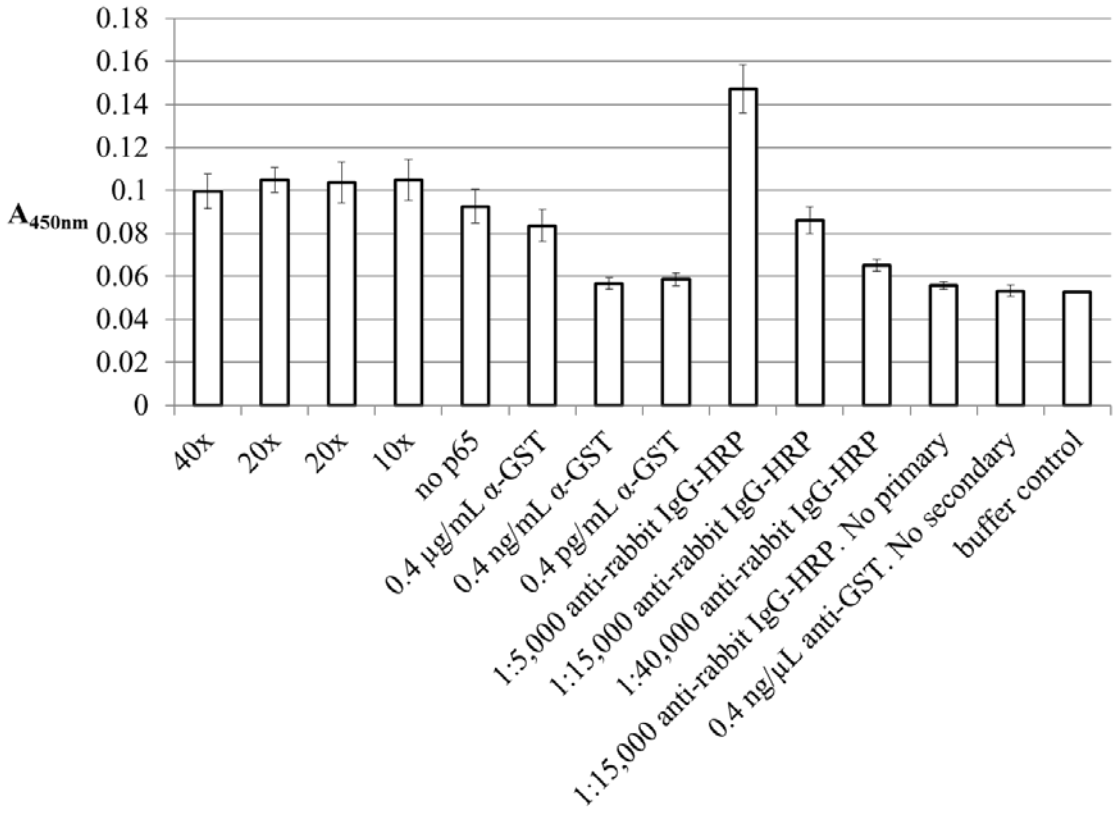

Figure 17. Multiple biotinylation conditions in ELISA. p65 was biotinylated at two times and one half of the suggested biotinylation reagent concentration of 20 fold M excess. 100 µL of 10 ng/µL biotinylated p65 was captured on an immunoabsorbant 96-well plate coated with 1 pmol per well of avidin. 50 µL per well of 100 nM GST-DNA binding domain was bound to the p65 and detected with 100 µL of 0.4 ng/mL α-GST rabbit polyclonal. Dilution samples of 0.4 µg/mL and 0.4 pg/mL α-GST were also used. 100 µL of a 1:15,000 α-rabbit IgG-HRP dilution was used to detect the α-GST. 1:5,000 and 1:40,000 dilutions of α-rabbit IgG-HRP were also tested as samples. The color change of 50 µL TMB was quenched with 50 µL H_2SO_4 and measured at 450 nm on a plate reader. The results show that 0.4 ng/mL of α-GST and 1:15,000 of α-rabbit IgG-HRP are suitable dilutions for future ELISAs.

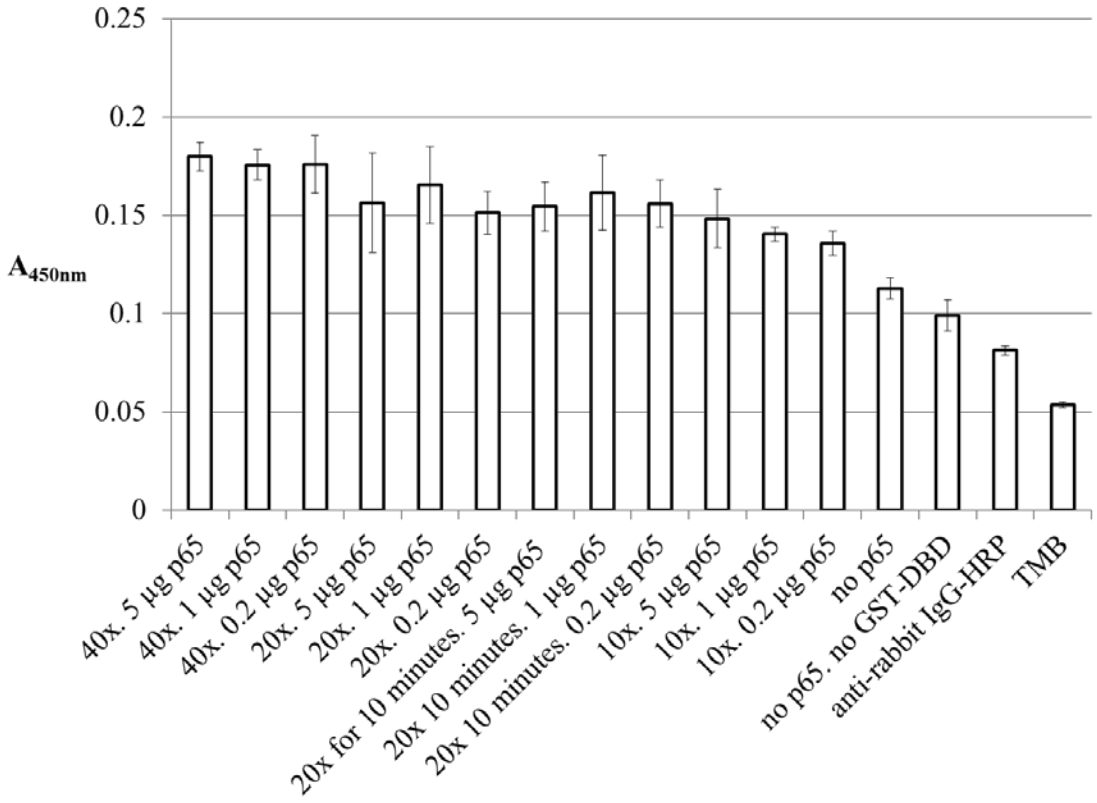

Figure 18. Biotinylated p65 dose response curve ELISA. 5 µg, 1 µg, and 0.2 µg of biotinylated p65 from different biotinylation reagent concentrations were added to plates coated with 1 pmol per well of avidin. 50 µL of 10 nM GST-DNA binding domain was bound to the biotinylated p65 and detected with 100 µL 0.4 ng/mL α-GST. 100 µL of a 1:15,000 α-rabbit IgG-HRP dilution was used to detect the α-GST. The color change of 50 µL TMB was quenched with 50 µL H_2SO_4 and measured at 450 nm on a plate reader. One out controls were used for each assay component. The figure suggests that the low signal is due to less than 0.2 µg of biotinylated p65 being captured on the plate surface.

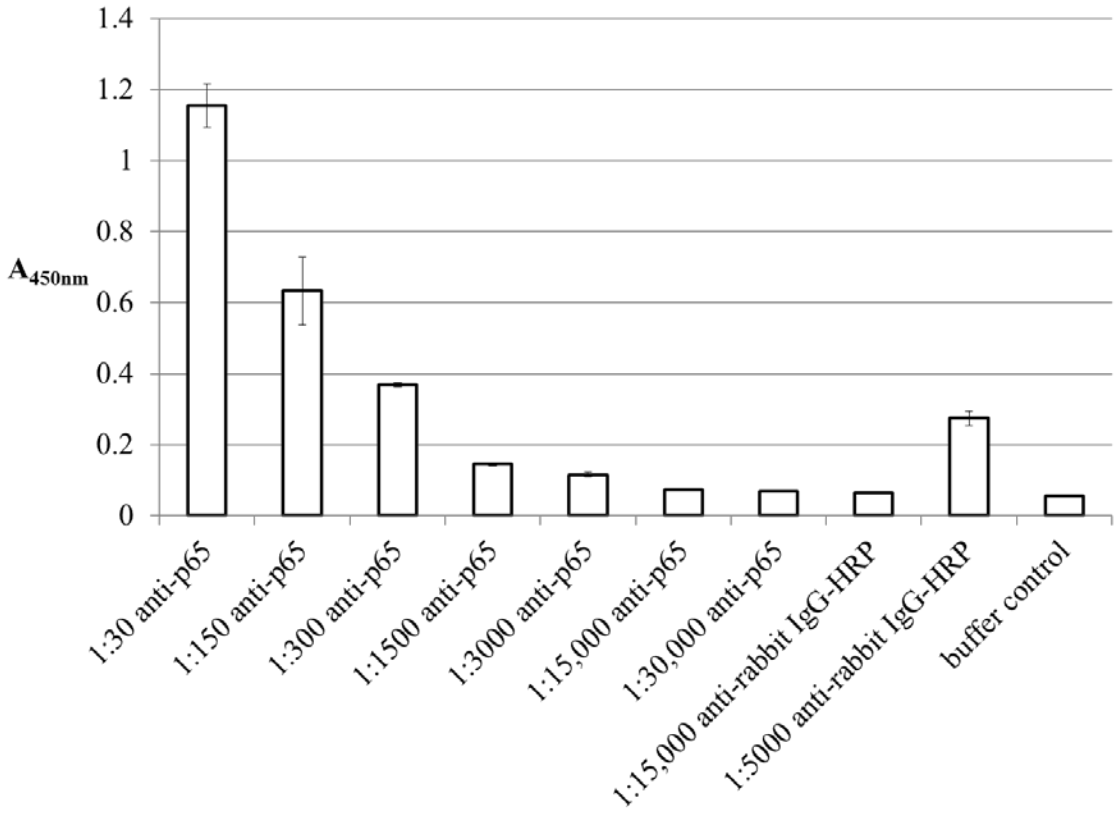

Figure 19. α-p65 dilution curve ELISA. 100 µL of different dilution of α-p65 were plated on an immunoabsorbant 96-well plate blocked with 1% BSA in 1xPBST. 100 µL of a 1:15,000 α-rabbit IgG-HRP dilution was used to detect the α-p65. A 1:5,000 dilution of α-rabbit IgG-HRP was used as a positive control. The non-specific binding of the p65 antibody was analyzed and it was determined that 1:3,000 was a suitable dilution for future studies.

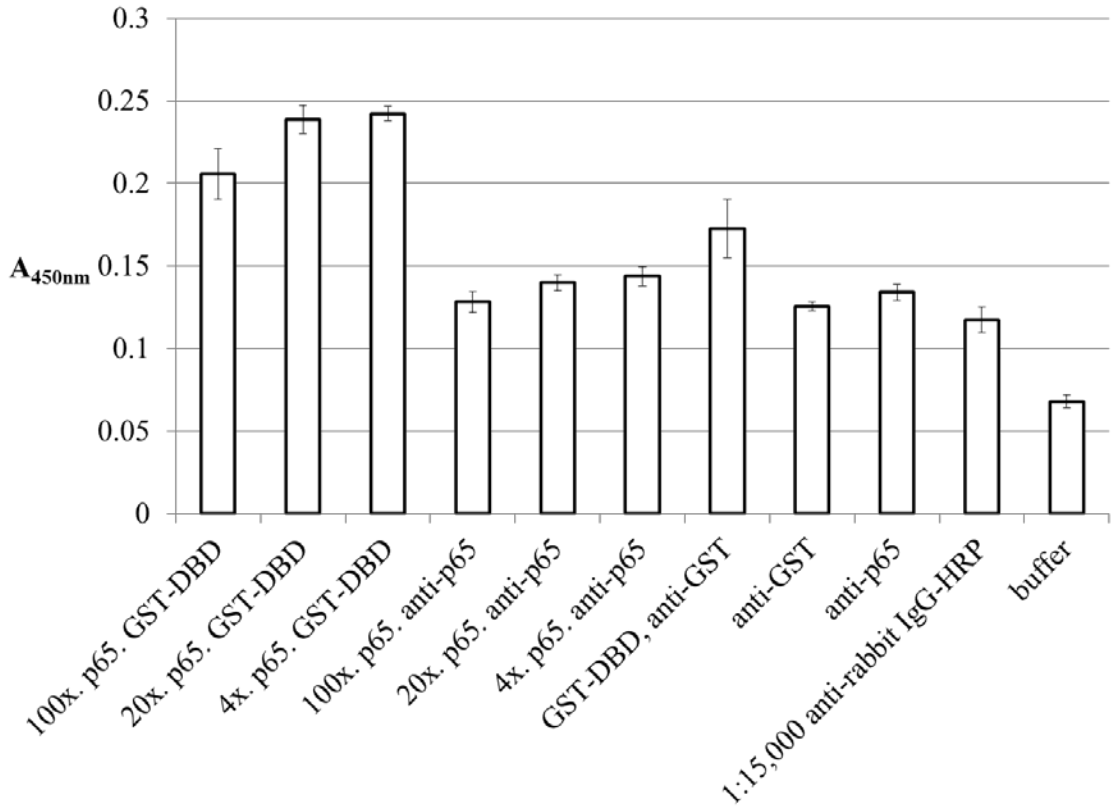

Figure 20. 100 pmol avidin-coating per well and DNA binding domain-p65 coupled biotinylation results in higher signal. Equal amounts of p65 and LSF DNA binding domain were bound together and biotinylated with a 20 fold, 100 fold, and 4 fold M excess biotinylation reagent concentration. 10 ng/µL biotinylated p65 was then captured on an immunoabsorbant 96-well plate coated with 100 pmol per well of avidin. 50 µL of 10 nM GST-DNA binding domain was bound to the biotinylated p65 and detected with 100 µL 0.4 ng/mL α-GST. A 1:3,000 dilution of α-p65 was used to detect the total amount of biotinylated p65. 100 µL of a 1:15,000 α-rabbit IgG-HRP dilution was used to detect the α-GST and the α-p65. One out controls were used for each assay component. The low signal for the α-p65 experimental samples suggests that not enough antibody is present to give a suitable signal over background.

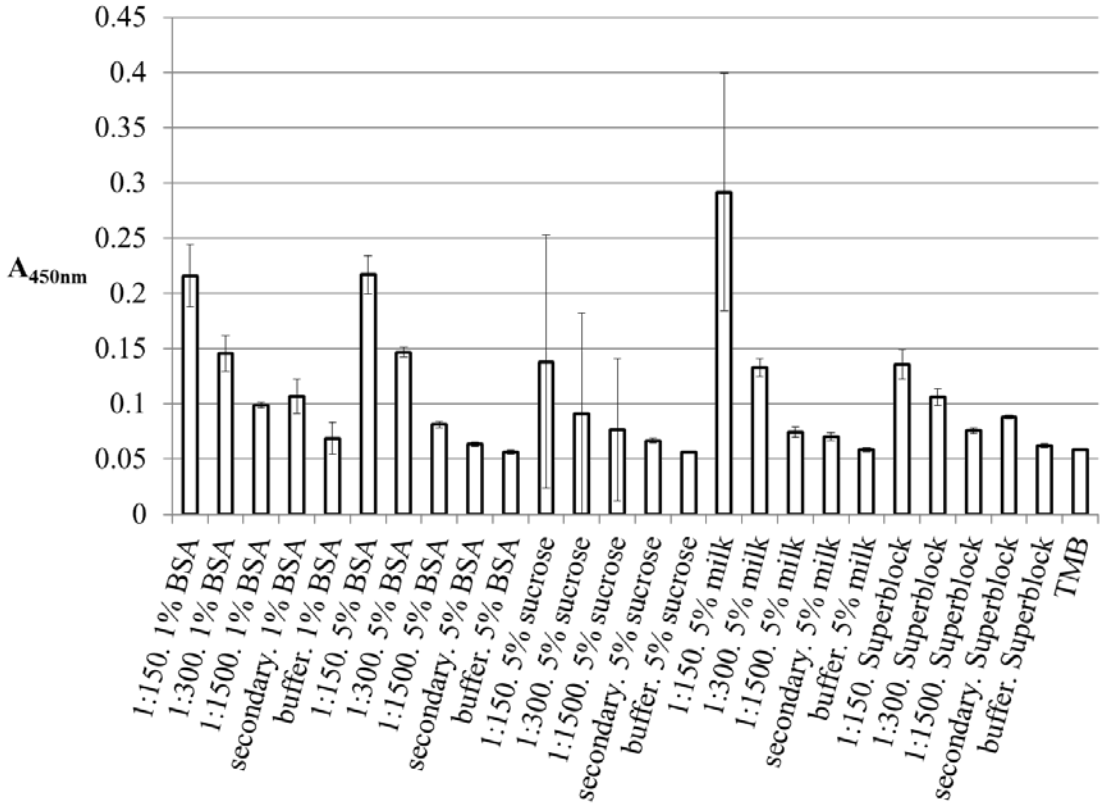

Figure 21. Pilot ELISA of different blocking conditions. 200 µL of 1% BSA, 5% BSA, 5% sucrose, or 5% milk in 1xPBST was added to an immunoabsorbant 96-well plate. Superblock from Pierce Biotechnology was also used as a candidate blocking solution. After a 16 hour incubation at 4°C, 100 µL of 1:150, 1:300, or 1:1500 α-p65 was added to plate. Non-specific binding of the p65 antibody was detected with 1:15,000 α-rabbit IgG-HRP. The color change of 50 µL TMB was quenched with 50 µL H_2SO_4 and measured at 450 nm on a plate reader. The results suggest that Superblock is the most suitable blocking solution for α-p65.

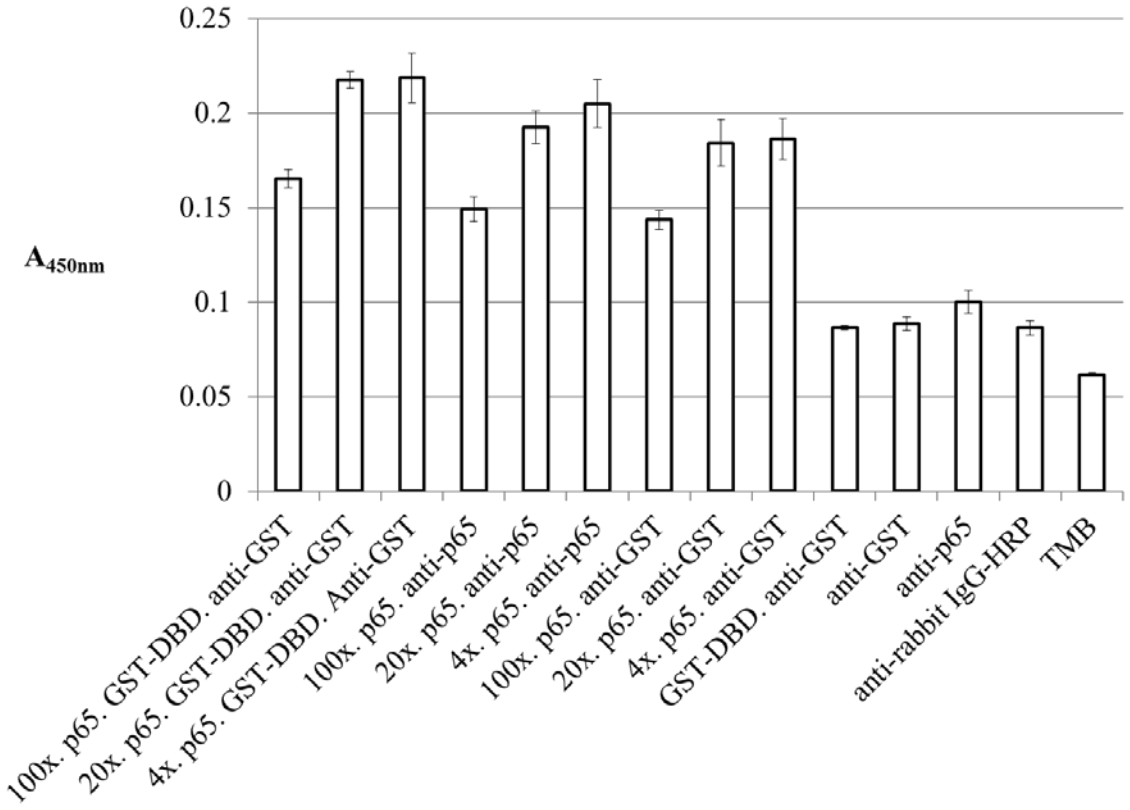

Figure 22. GST contamination from DNA binding domain-p65 coupled biotinylation results in high background signal. Equal amounts of p65 and LSF DNA binding domain were bound together and biotinylated with a 20 fold, 100 fold, and 4 fold M excess biotinylation reagent concentration. 10 ng/µL biotinylated p65 was then captured on an immunoabsorbant 96-well plate coated with 100 pmol per well of avidin. 50 µL of 10 nM GST-DNA binding domain was bound to the biotinylated p65 and detected with 100 µL 0.4 ng/mL α-GST. A 1:150 dilution of α-p65 was used to detect the total amount of biotinylated p65. 100 µL of a 1:15,000 α-rabbit IgG-HRP dilution was used to detect the α-GST and the α-p65. The high background signal in the biotinylated p65 α-GST control suggests that the observed signal for the GST-DNA binding domain bound to p65 is due to GST contamination of

Bibliography

Bing, Zhanyong. Huang, Jianyi H. Liao, Warren S.L. NFKB Interacts with Serum Amyloid A3 Enhancer Factor to Synergistically Activate Mouse Serum Amyloid A3 Gene Transcription. *The Journal of Biological Chemistry* 275, no. 41 (October 2000): 31616-23

Brenki, Ryan. Kozakov, Dima. Chuang, Gwo-Yu. Beglov, Dimitri. Hall, David. Landon, Melissa R. Mattos, Carla. Vajda, Sandor. Fragment-based identification of druggable 'hot spots' of proteins using Fourier domain correlation techniques. *Bioinformatics* 25, no. 5 (January 2009): 621-27

Bruni, Paola. Minopoli, Giuseppina. Brancaccio, Tiziana. Napolitano, Maria. Faraonio, Raffaella. Zambrano, Nicola. Hansen, Ulla. Russo, Tommaso. Mechanisms of Signal Transduction: Fe65, a Ligand of the Alzheimer's Beta-Amyloid Precursor Protein, Blocks Cell Cycle Progression by Down-Regulating Thymidylate Synthase Expression. *The Journal of Biological Chemistry* 277, no. 38 (September 2002): 35481-88

Chandramouli, Kondethimmanahalli. Qian, Pei-Yuan. Proteomics: Challenges, Techniques and Possibilities to Overcome Biological Sample Complexity. *Human Genomics and Proteomics* 1, no. 1 (December 2009): 1-22

Coull, Jason J. Romerio, Fabio. Sun, Jian-Min. Volker, Janet L. Galvin, Katherine M. Davie, James R. Shi, Yang. Hansen, Ulla. Margolis, David M. The Human Factors YY1 and LSF Repress the Human Immunodeficiency Virus Type 1 Long Terminal Repeat via Recruitment of Histone Deacetylase 1. *Journal of Virology* 74, no. 15 (August 2000): 6790-99

Dolcet, Xavier. Llobet, David. Pallares, Judit. Matias-Guiu, Xavier. NF-kB in Development and Progression of Human Cancer. *Virchows Archiv* 446, (April 2005): 475-82

Escalante, Carlos R. Shen, Leyi. Thanos, Dimitris, Aggarwal, Aneel K. Structure of NF kB p50/p65 Heterodimer Bound to the PRDII DNA Element from the Interferon-Beta Promoter. *Structure* 10, (March 2002): 383-91

Ferenci, Peter MD. Fried, Michael MD. Labrecque, Douglas MD. Bruix, Jordi MD. Sherman, Morris MD. Omata, Masao MD. Heathcote, Jenny MD. Piratsivuth, Teehra MD. Kew, Mike MD. Otegbayo, Jesse A MD. Zheng, S.S. MD. Sarin, S MD. Hamid, Saeed S MD. Modawi, Salma Barakat MD. Fleig, Wolfgang MD. Fedail, Suliman MD. Thomson, Alan MD. Aamir, Khan MD. Malfertheiner, Peter MD. Lau, George MD. Carillo, Flair J. Krabshuis, Justus Drs. Le Mair, Anton MD. Hepatocellular Carcinoma (HCC): A Global Perspective. *Journal of Clinical Gastroenterology* 44, no. 4 (April 2010): 239-45

Flanagan, JR. Autologous Stimulation of YY1 Transcription Factor Expression: Role of An Insulin-like Growth Factor. *Cell Growth and Differentiation* 6, (February 1995): 185-90

Gasteiger, Elisabeth. Hoogland, Christine. Gattiker, Alexandre. Duvuad, Severine. Wilkins, Marc. Appel, Ron. Bairoch, Amos. 2005. Protein Identification and Analysis Tools on the ExPASy Server. In The Proteomics Protocols Handbook, ed. John M. Walker, 571-606. Totowa, NJ: Humana Press Inc.

Gordon, S. Akopyan, G. Garban, H. Bonavida, B. Transcription Factor YY1: Structure, Funtion, and Therapeutic Implications in Cancer Biology. *Oncogene* 25, (November 2006): 1125-42

Houbaviy, Hristo B. Usheva, Anny. Shenk, Thomas. Burley, Stephen K. Cocrystal Structure of YY1 Bound to the Adeno-associated Virus Initiator. *Proceedings of the National Acadademy of Sciences USA* 93, (September 1996): 13577-82

Kokoszynska, Kararzyna. Ostrowski, Jerzy. Rychlewski, Leszek. Wyrwicz, Lucjan S. The Fold Recognition of CP2 Transcription Factors Gives New Insight Into the Function and Evolution of Tumor Suppressor Protein p53. *Cell Cycle* 7, no. 18 (September 2008): 2907-15

Livnah, Oded. Bayer, Edward A. Wilchek, Meir. Sussmen, Joel L. Three-dimensional structures of avidin and the avidinbiotin complex. *Proceedings of the National Academy of Sciences USA* 90, (June 1993): 5076-80

McNeil, Sandra. Guo, Bo. Stein, Janet L. Lian, Jane B. Bushmeyer, Sarah. Seto, Edward. Atchinson, Michael L. Penman, Sheldon. Van wijnen, Andre J. Stein, Gary. Targeting of the YY1 transcription factor to the nucleolus and the nuclear matrix in situ: The C terminus is a principal determinant for nuclear trafficking. *Journal of Cellular Biochemistry* 68, no. 4 (March 1998): 500-10

Pan, Catherine. Sudol, Marius. Sheetz, Michael. Low, Boon Chuan. Modularity and functional plasticity of scaffold proteins as p(l)acemakers in cell signaling. *Cellular Signaling* 24 (June 2012) 2143-65

Pawson, Tony. Nash, Piers. Assembly of Cell Regulatory Systems Through Protein Interaction Domains. *Science* 300, (April 2003): 445-52

Perkins, Neil D. Integrating Cell Signaling Pathways with NF-kB and IKK Function. *Nature Reviews Molecular Cell Biology* 8, (January 2007): 49-62

Santhekadur, Prasanna K. Rajasekaran, Devaraja, Siddiqm Ayesha. Gredler, Rachel. Chen, Dong. Schaus, Scott E. Hansen, Ulla. Fisher, Paul B. Sarkar, Devanand. The Transcription Factor LSF: A Novel Oncogene for Hepatocellular Carcinoma. *American Journal of Cancer Research.* 3, no. 2 (May 2012): 269-85

Satjin, David P.E. Otte, Arie P. RING1 Interacts with Multiple Polycomb-Group Proteins and Displays Tumorigenic Activity. *Molecular and Cellular Biology* 19, no. 1 (January 1999): 57- 68

Sheinerman, Felix B. Norel, Raquel, Honig, Barry. Electrostatic Aspects of Protein Protein Interactions. *Current Opinion in Structural Biology* 10, (2000): 153-159

Shirra, Margaret K. Zhu, Quan. Huang, Hui-Chuan. Pallas, David. Hansen, Ulla. One Exon of the Human LSF Gene Includes Conserved Regions Involved in Novel DNA Binding and Dimerization Motifs. *Molecular and Cellular Biology* 14 (August 1994): 5076– 5087.

Shirra, Margaret K. Hansen, Ulla. LSF and NTF-1 Share a Conserved DNA Recognition Motif yet Require Different Oligomerization States to Form a Stable Protein-DNA Complex. *The Journal of Biological Chemistry* 273, no. 30 (July 1998): 19260-68

Tuckfield, Annabel. Clouston, David R. Wilanowski, Tomasz, M. Zhao, Lin-Lin. Cunningham, John M. Jane, Stephen M. Binding of the RING Polycomb Proteins to Specific Target Genes in Complex with the *grainyhead*-Like Family of Developmental Transcription Factors.
Molecular and Cellular Biology 22, no. 6 (March 2002): 1936-46

Traylor-Knowles, Nikki. Hansen, Ulla. Dubuc, Timothy Q. Martindale, Mark Q. Kaufman, Les. Finnerty, John R. The Evolutionary Diversification of LSF and Granyhead Transcription Factors Preceded the Radiation of Basal Animal Lineages. *BMC Evolutionary Biology* 10, no. 101 (April 2010): 1-13

Virkamaki, Antti. Ueki, Khjiro. Khan, Ronald J. Protein–protein interaction in insulin signaling and the molecular mechanisms of insulin resistance. *The Journal of Clinical Investigation* 103, no. 7 (April 1999): 931-943

White, MF. The IRS-signalling system: a network of docking proteins that mediate insulin action. *Molecular Cell Biochemistry* (1998): 1823-11

Yoo, Byoung Kwon. Emdad, Luni. Gredler, Rachel. Fuller, Christine. Dumur, Catherine I. Jones, Kimberly H. Jackson-Cook, Colleen. Su, Zao-Zhong. Chen, Dong. Saxena Utsav H. Hansen, Ulla. Fisher, Paul B. Sarkar, Devanand. Transcription Factor Late SV40 Factor (LSF) Functions as an Oncogene in Hepatocellular Carcinoma. *Proceedings of the National Academy of Sciences* 107, no. 18 (May 2010): 8357-62

Zhang, Xing. Studier, William F. Mechanism of inhibition of bacteriophage T7 RNA polymerase by T7 lysozyme. *Journal of Molecular Biology*. 269, no. 1 (May 1997): 10-27