Theses and Dissertations   Graduate School

2014

Using structural analysis to investigate the function of Suppressor of IKK-epsilon (SIKE)

Sean W. McKinley
VCU

www.manaraa.com

USING STRUTURAL ANALYSIS TO INVESTIGATE THE FUNCTION OF

SUPPRESSOR OF IKK-ε (SIKE)

A thesis submitted in partial fulfillment of the requirements for the degree of Masters of

Science at Virginia Commonwealth University.

by

SEAN W. MCKINLEY

B.S., James Madison University, 2012

Director: JESSICA K. BELL, PH. D.

Virginia Commonwealth University

Richmond, VA

December 2014

Acknowledgements

I would like to briefly recognize everyone that aided in my academic success and completion of my thesis and degree. First, I would like to express my deepest gratitude to my advisor, Dr. Jessica Bell, for opening her lab to me and giving me the freedom to explore, while guiding me to develop as both a student and a scientist. Her patience and support have helped me overcome countless obstacles, teaching me invaluable knowledge and experience. Without her efforts, I would not have been able to produce this work; to her I extend my utmost gratitude.

I sincerely appreciate the time and help received from my committee members. I would like to thank Dr. Carlyon and Dr. Peterson for their time as professors (Microbiology and Biochemistry respectively) and their dedication to my experience at VCU. Both provided me with irreplaceable knowledge leading to my success as a master's student.

Recognition must be given to the Bell Laboratory at the University of Richmond for giving me the opportunity to continue experiments that were not possible at VCU. I would specifically like to thank Emily Bartle and Philip Varnes for their time and efforts spent contributing to my studies. Dr. Ellis Bell was instrumental in this process; to him and the whole lab, I would like to extend my sincere gratitude.

I greatly appreciate the help and support that I have received from VCU labmates and friends since my entry into the lab. I want to thank Kenny Lawrence for limitless

contributions to intellectual conversations, lab work, paper revisions, and study sessions that have all aided in my success in class and in the lab. I would also like to thank Dr. Sato-Bigbee and Logan Meyer for the help and support given to me from my very first day in the lab to the very last. Thanks to these individuals for their inspiration and involvement in my success as a student. Without them, my completion of this program would not have been possible.

Most importantly, I would like to express my appreciation to my family, to whom this thesis is dedicated. I would like to thank my parents, William and Linda McKinley, for their love and guidance, allowing me to mature and develop into who I am today. Their dedication to higher education and medical healthcare has subsequently instilled a great appreciation in myself as well. I must also thank my sisters, Kaitlin and Megan McKinley, for always believing in me. The everlasting encouragement, love, and support have been crucial in my achievements throughout life. I would also like to express my heart-felt gratitude to my extended family for the aid and encouragement throughout my endeavors. I sincerely appreciate the time, energy, and love that have been spent on my success and my future.

Table of Contents

List of Tables

List of Figures

Abbreviations

AP-1:	Activator Protein 1
BSA:	Bovine Serum Albumin
CV:	Column-Volume
FADD:	Fas-Associated with Death Domain protein
FL:	Full-Length
FT:	Flow-Through
HWI:	Hauptman Woodward Institute
IFN:	Interferon
IκBα:	Inhibitor of κBα
IKK:	IκB kinase complex
IL-1R:	Interleukin-1 Receptor
IMAC:	Immobilized Metal Ion Chromatography
IPTG:	Isopropyl-β-Galactopyranoside
IRF:	Interferon Regulatory Factor
JAK:	Janus Kinase
LB:	Luria Broth
LRR:	Leucine-Rich Repeat
MAL:	MyD88 Adaptor-Like
MALDI:	Matrix Assisted LASER Desorption/Ionization
MAPK:	Mitogen-Activated Protein Kinase
MyD88:	Myeloid Differentiation Factor 88
NEMO:	NF-κB essential modulator

NAK: NF-κB Activating Kinase

NAP1:   NAK-Associated Protein 1

NK cells: Natural Killer cells

PAMPs: Pathogen-Associated Molecular Patters

PEG: Polyethylene Glycol

PMN cells: Poly Morpho Nuclear cells

PRRs: Pattern Recognition Receptors

RHIM: RIP Homotypic Interaction Motif

RIP1:   Receptor-Interacting Protein 1

SDS-PAGE: Sodium Dodecyl Sulfate Polyacrylamide Gel Electrophoresis

SEC: Size Exclusion Chromatography

SIKE: Suppressor of IKK-ε

STAT: Signal Transducer and Activator of Transcription

TAB2   TAK Binding protein 2

TAK1:   Transforming growth factor β-Activating Kinase 1

TANK: TRAF-Associated NF-κB inhibitor

TBK1:   TANK-binding Kinase 1

TLR: Toll-Like Receptor

TNF: Tumor Necrosis Factor

TRAF: TNF Receptor-Associated Factor 1

TRAM: TRIF-Related Adaptor Molecule

TRIF: TIR domain-containing adaptor-inducing IFN-β

TFA: Trifluoroacetic acid

Abstract

USING STRUTURAL ANALYSIS TO INVESTIGATE THE FUNCTION OF

SUPPRESSOR OF IKK-ε (SIKE)

By Sean W. McKinley, M.S.

A thesis submitted in partial fulfillment of the requirements for the degree of Masters of

Science at Virginia Commonwealth University.

Virginia Commonwealth University, 2014

Major Director: Dr. Jessica K. Bell, Ph.D.

Assistant Professor, Department of Biochemistry & Molecular Biology

The innate immune system provides the body's first line of defense against

pathogenic challenge through pathogen recognition and initiation of the immune

response. Among the various cellular mechanisms of pathogen recognition in mammals,

Toll-like receptor 3 (TLR3) recognizes viral dsRNA. Stimulation of TLR3 signaling

pathway leads to transcription of pro-inflammatory cytokines and type-1 Interferons.

Suppressor of IKKε (SIKE) interacts with two kinases in the signaling pathway, IKKε

and TANK binding kinase 1 (TBK1), inhibiting the transcription of type I interferons.

Recently, the Bell Laboratory discovered that SIKE blocks TBK1-mediated activation of

type I interferons by acting as a high affinity, alternative substrate of TBK1.

To further characterize SIKE's function within the antiviral response, this study focused on defining the overall SIKE structure. Using recombinant protein expressed from *E. coli* and purified via immobilized metal affinity chromatography, SIKE crystals were obtained from a sample concentrated to 15 mg/ml under several crystallization conditions. Yet, reproducing these results has been difficult. In this study, we have modified the purification scheme to remove an *E. coli* contaminant, SlyD. Purification under denaturing conditions, removal of soluble proteins, incorporation of ion exchange and different IMAC (immobilized metal ion affinity chromatography) resins has been tested. For each scheme, size exclusion chromatography and SDS-PAGE/Coomassie/silver stain were used to assess purity. Crystallization trials for samples from each purification scheme were completed. In addition to crystallization trials, hydrogen-deuterium exchange (HDX) was investigated, accompanied with pepsin digests, in order to further characterize the dynamic structure of SIKE.

Chapter 1: Introduction

1.1 Immunology and the Immune Response

Immunology is a subdivision of biomedical science that explores all aspects of the immune systems found in different organisms.   The main focus of this field is the functioning of the immune system in both healthy and diseased states, as well as genetic or acquired immune disorders[1]. Immunology also involves studying other systems, where pathology and clinical symptoms are affected by immune reactions. The success of the immune system depends on its ability to distinguish host (self) cells from foreign (non-self) antigens. The immune system doesn't normally mount an immune response against itself; this is termed "tolerance". However, in rare cases tolerance can be lost and immune responses are initiated against the host's cells; this is called auto-immunity[1,2].

The Chinese are credited with making some of the first immunological observations, deliberately infecting people with mild forms of small pox to prevent infections with deadlier forms and even providing lifelong protection. After this discovery, knowledge and practices moved through Turkey in the late 18th century and ultimately to Britain in 1796. There, British physician, Edward Jenner, published his long-term observation that milkmaids exposed to cowpox became protected from smallpox[1,2]. In his experiments, he exposed a "healthy" 8-year-old boy to cowpox virus taken from a milkmaid cowpox pustule by inserting pus into an incision in the boys arm[2]. Cowpox shares antigens with smallpox, but doesn't cause the same disease; the vaccine

worked by initiating an immune response against cowpox antigen, subsequently immunizing the boy against smallpox antigen and establishing a smallpox immunity[2].

For the next fifty years, the prevailing theory was that chemical toxins were contracted from ill patients, causing the transmission of the disease; the bacteria observed were considered to be symptoms of the disease state, not the cause of it[2]. It wasn't until the mid-19th century that Friedrich Gustav Jakob Henle, following the ideas of Girolamo Fracastoro and Agostino Bassi, began to observe that growth and reproduction of some microorganisms within their hosts could be the cause of a disease. Gustav's essay, "On Miasma and Contagia," was an early report demonstrating the germ theory of disease. "Miasma" was considered to be a poisonous vapor or toxin emitted from infected organic matter that can cause a disease. This became the predominant theory of disease transmission until the germ theory of disease greatly changed immunology[3].

Louis Pasteur is traditionally considered as the "father of modern immunology" because of his extensive work in the late 19th century on fermentation, spontaneous generation, and other areas of microbiology[4]. Pasteur led the fields of epidemiology, bacteriology, and public health in that time, as well as being the first to establish the possibility of anaerobic organisms[5]. In his famous work, "Germ Theory of Disease," Louis Pasteur explained his theory that some infectious diseases, such as anthrax, cholera, tuberculosis, and smallpox, are caused by microorganisms. Pasteur is also renowned for inventing a rabies vaccine in 1885[4].

In 1876, Louis Pasteur's "Germ Theory of Disease" was proven when Robert Koch successfully isolated an infectious bacillus that he showed was the cause of both anthrax and cholera. Koch investigated pathogenicity, developing a set of postulates that

are still used in labs today to examine if a disease is caused by a specific virus, bacteria, fungi, or parasite. Koch also worked extensively to develop more efficient staining methods, culturing techniques, and was one of the first to demonstrate that unfavorable conditions can lead to sporulation in certain bacterial strains to survive[6].

After the breakthrough that diseases are caused by microbes, researchers began investigating how host organisms combat these diseases and symptoms. In 1890, Emil von Behring and Shibasaburo Kitasato immunized guinea pigs with heat-treated diphtheria toxin; by infecting the animals with a non-lethal form of the pathogen, they were able to develop a sufficient immune response to subsequent infections by those bacteria. After its success in the first human trial, it was known as the "first cure" and Von Behring was awarded the Nobel Prize in 1901 for this work[7].

Other scientists at this time were also researching different cellular processes. Russian zoologist, Elie Metchnikoff, discovered the process of phagocytosis while working with starfish larvae. He realized that this was practically the same process that white blood cells perform in the human body to destroy certain potentially harmful microbes. This belief was argued by some of the leading scientists such as Pasteur; however, Metchnikoff's work on phagocytosis won him the Nobel Prize in 1908[8].

Debates arose as to which theory regarding the cellular basis of immunity was correct: antibody-mediated immunity or phagocytosis. Today, they are both recognized as fundamental processes of cellular immunity. The antibody-mediated response is associated with the acquired (adaptive) immune response and requires the production of specific antibodies to target a particular pathogen. Germline-encoded gene segments have the ability to recombine to synthesize copious amounts of unique, antigen-

recognizing immunoglobulin[9]. This adaptability process requires 3-5 days to complete, following the initial infection; however, can result in lifelong immunity to subsequent infections by pathogens expressing the same antigen[2] (shown in Figure 1).

The other process of cellular immunity, phagocytosis, plays a primary role in, what is now known as, the innate (non-specific) immune response. This begins immediately after infection and provides the first line of defense against other organisms in a non-specific manner. These responses are mostly limited to defenses that are constitutively expressed and readily mobilized[1,2]. Germline encoded sensors and receptors recognize certain signals and stimulate phagocytic cells to engulf and kill the pathogens associated with these signals[2]. The innate immune system involves anatomical barriers such as epithelial surfaces, low pH levels, and normal flora that prevent infection and colonization by invading organisms. It also contains humoral barriers, for example the complement system, interferons, interleukins, and lysozyme, which work as part of the inflammatory process when invasive pathogens penetrate tissues. As proposed by Metchnikkoff, neutrophils, macrophages, Natural Killer (NK) cells, and eosinophils perform the cellular process of phagocytosis. These cells are recruited to the site of infection where they kill invading organisms extracellularly or phagocytose and kill them intracellularly[1,2]. The main focus of our lab, and the remainder of this document, is the role of the innate immune system and the function of a newly discovered component of the anti-viral immune response.

Figure 1: Main components of the innate and adaptive immune systems
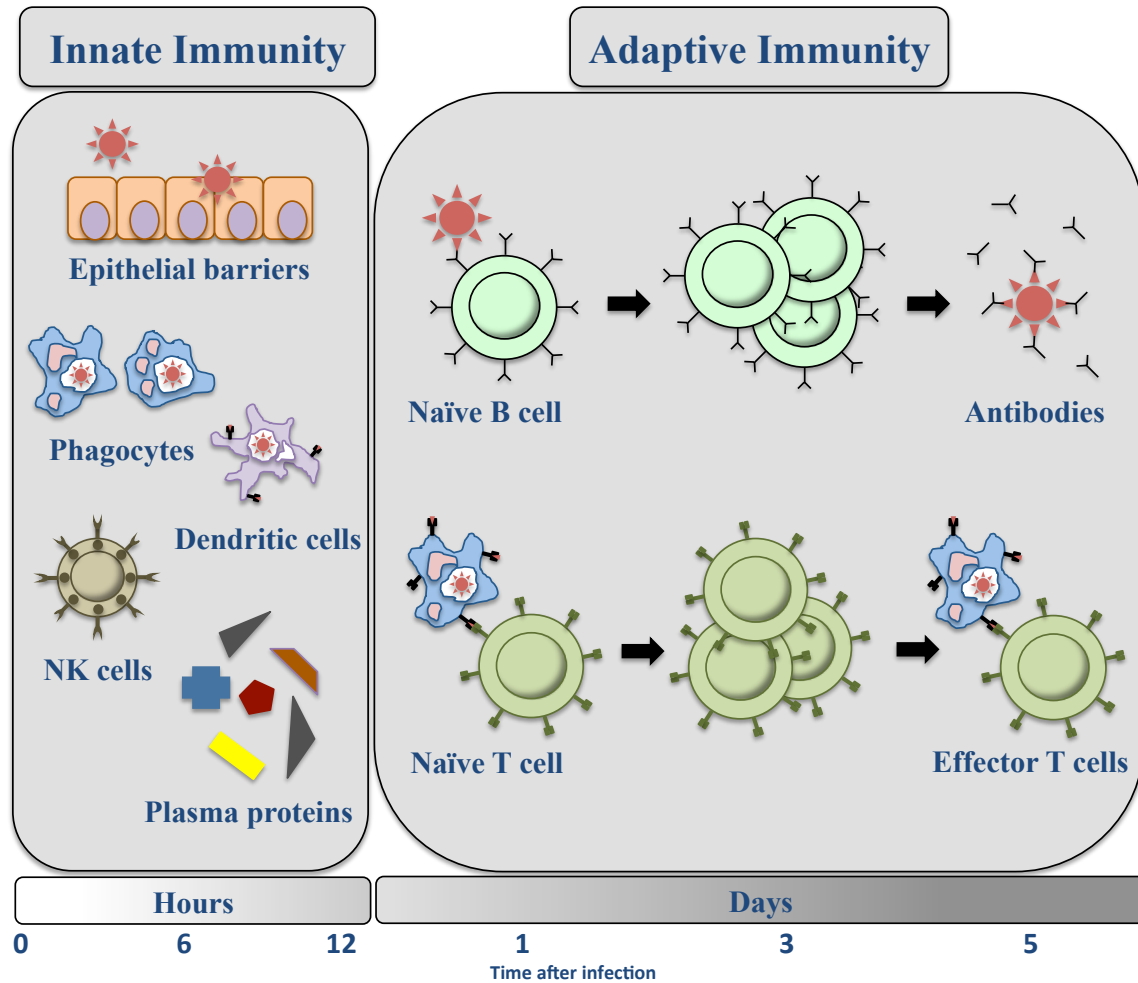
Figure 1: Main components of the innate and adaptive immune systems. Components of the immune system are divided into two categories: Innate immunity, (left) effective during initial stages of infection, and

1.2 The Innate Immune System

The human body encounters various pathogenic agents each day, yet infections are rarely established. These agents are destroyed within minutes to hours without the use of antigen-specific immunoglobulin for recognition[1]. The innate immune response employs a set of mechanisms to distinguish self from non-self that are thought to be older, evolutionarily[3]. These defense strategies are comprised of anatomical, humoral, and cellular barriers, all of which contribute to the first line of defense against pathogens[3].

Anatomical barriers provide the earliest protection of the innate defense systems[1,2]. These anatomical barriers are characterized by certain mechanical, chemical, and biological factors. Mechanical barriers refer to the epithelial surfaces that form a physical barrier that is impermeable to most infectious agents. These epithelial layers include the skin and lining of respiratory, gastrointestinal, and urinary tracts and are held together by tight junctions[1,2]. Mechanisms used by these barriers to remove bacteria and other infectious agents include: desquamation (or shedding) of the skin, oscillation of broncho-pulmonary cilia to clear respiratory tract, and the flushing of tears and saliva to protect the mouth and eyes[3]. Biological factors, such as normal microbial flora, function to prevent the colonization of pathogenic bacteria through various methods. Flora is commensal bacterium that can inhibit the establishment of bacterial infections by secreting toxic substances or by competing for nutrients or attachment to epithelial surfaces[1]. These commensal bacteria can also stimulate the nearby epithelial cells to secrete antimicrobial agents to ward off invasive organisms. Chemical factors complement the mechanical barriers in the innate immune response. Lysozyme and

phospholipase help break down the cell wall of bacteria and destabilize bacterial membranes. These are found in saliva, tears, and nasal secretions. Fatty acids and low pH levels inhibit the growth of bacteria and are found on the skin and in the gastrointestinal tract. Defensins are antimicrobial agents found in the lungs and gastrointestinal tract that destroy microbial cell membranes[1,2]. All together, these physical, chemical, and biological barriers provide the first stage of defense in the innate immune response.

Infections, however, can still occur when damage has been done to tissues and anatomical barriers are breached. When a microbe has established an infection, the next system of defense employed by the innate immune system is the anti-inflammatory response. This process involves humoral factors like the complement system, coagulation system, interferons, interleukins, lysozyme, and lactoferrin/transferrin[1,2]. These factors are constitutively present in serum and are also up-regulated at the site of infection, once the infection is recognized by host cells. The complement system is the major non-specific defense mechanism that, once activated, leads to vasodilation, recruitment of phagocytic cells, and opsonization and lysis of the organism[1]. The coagulation system is often activated in situations that involve more tissue damage. Products of the coagulation system increase vascular permeability, attract phagocytic cells to the site of infection, and can have direct anti-microbial properties. Lactoferrin and transferrin both function by binding iron in the environment. This is essential for bacterial growth and helps prevent infections. The release of interferons helps to limit viral replication in infected cells and lysozyme works to break down bacterial cell walls. Interleukins function to induce a fever and initiate the production of acute phase proteins

that can have antimicrobial properties. These factors are present in serum play an important role in the inflammation process, once infections have been established.

The cellular functions of the innate immune system include the recruitment of polymorphonuclear (PMN) eosinophils and macrophages to the site of infection. Neutrophils are a type of PMN cell that are recruited to the site of infection to phagocytose invading organisms and kill them intracellularly[1]. Macrophages function using phagocytosis and intracellular killing, as well as extracellular killing of infected or dysfunctional host cells. Macrophages also contribute to antigen presentation that is required for the induction of specific immune responses. Natural Killer (NK) cells contribute to the killing of virally infected and tumor cells. Eosinophils contain proteins that are effective in killing certain infectious parasites[3]. Along with mechanical and humoral factors, the cellular defense systems provide a short-lived, non-specific, early immune response to control or eliminate infection by pathogenic agents.

1.3 Toll-like receptors: Discovery and Structure

The innate immune system uses a unique strategy in order to distinguish self from non-self antigens using non-specific means. Repeating patterns on the pathogen surface structure or in nucleic acid sequences are recognized by pattern recognition receptors (PRRs). These receptors identify evolutionarily conserved motifs that are found mainly in microorganisms, called PAMPs (pathogen-associated molecular patterns) [1,2,11]. PRRs are expressed by various epithelial, endothelial, and immune cells and immediately begin mounting an immune response. Receptors are organized into categories defined by their location and function: membrane-bound signaling receptors, cytosolic signaling

receptors, membrane-bound phagocytic receptors, or free receptors secreted into the circulatory or lymphatic systems[1,2]. One class of signaling receptors, Toll-like receptors (TLRs), is the focus of research in the Bell Laboratory and will be the basis of the remainder of the document.

TLRs represent an evolutionarily conserved defense system. First identified in fruit fly embryos, Toll receptors functioned in dorso-ventral patterning during embryogenesis. Christine Nϋsslein-Volhard observed underdevelopment in the patterning when mutated; her reaction was "Das was ja toll!" meaning, "That was weird!" which subsequently gave Toll receptors their name[10]. The protein product of the Toll gene that causes ventral patterning in fruit fly embryos, known as Dorsal, is a fly homologue of the transcription factor NF-κB[1,2]. Jules Hoffman observed these mutations in adult fruit flies and observed decreased production of antimicrobial peptides and increased susceptibility to fungal infections[1,10]. This was the first time Toll receptors were associated with the host immune system. Homologues (Toll-like receptors) were then identified in mammals by Ruslan Medzhitov and Charles Janeway, and shown to function in several host defense mechanisms [1,2,10]. The group determined that TLR activation induces NF-κB in a similar manner as interleukin-1 receptors (IL-1Rs) and the fruit fly Toll receptors, leading to the expression of inflammatory cytokines[2,11].

To date, at least 10 members of the TLR family have been identified in mammals and are part of a larger superfamily of interleukin-1 receptors (IL-1Rs). The members of this superfamily all have a conserved region of about 200 amino acids in the cytoplasmic (C-terminal) region termed the TIR (Toll/IL-1R) domain[11]. This suggests that both TLRs and IL-1Rs signal via shared downstream molecules[10,11]. These TIR domains are also

observed in plants, which may be evidence that this is a conserved motif that may have served immune functions prior to the divergence of plants and animals.

In contrast, the extracellular (N-terminal) regions of these TLRs contain leucine-rich repeats (LRRs), while IL-1Rs have three immunoglobulin domains, which function to bind ligands. These extracellular domains are highly variable and are directly involved in the recognition of a variety of pathogens[1,11,12]. Scientists next sought to identify which type of ligands stimulate the receptors: a "danger signal" produced by host cells or a microbial molecule released by the pathogen[10]. Alexander Poltorak and Bruce Beutler observed that were unresponsive to a gram-negative endotoxin component, LPS, which they mapped back to a mutation in the TLR4 gene. To date, the structures of TLR1-6 and TLR8 have been solved using x-ray crystallography[13]. The variable, extracellular domains could now be analyzed to determine the amino acids underlying the receptor:ligand interactions. Three-dimensional structure showed that the LRR domains consist of a varying number of repeats, each 22-29 amino acids in length, containing the motif XXLXLXX along with other conserved leucines[2,11,13]. These domains also contain hydrophobic residues spaced at distinct, conserved distances, pointing inward and forming a solenoid structure. TLRs are able to form either homo- or heterodimers, depending on the type of ligand. Each of the ten members of the TLR family recognize a variety of evolutionarily conserved PAMPs, such as (but are not limited to) lipopeptides (TLR2:TLR1 or TLR2:TLR6 heterodimers), double-stranded RNA (TLR3 homodimer), LPS (TLR4 homodimer), flagellin (TLR5 homodimer), single-stranded RNA (TLR7 and TLR8 monomer), or CpG-rich unmethylated DNA (TLR9 homodimer) (Figure 2)[1,10,11,12].

Signal transduction through TLRs can occur both on the cell surface and in intracellular vesicles. TLRs 1, 2, 4, 5, and 6 are located mainly on the cell surface, while 3, 8, and 9 are found on the membranes of endosomes and lysosomes[13] (Figure 2). Signal transduction for both occurs in the cytosol, allowing their ligand-binding domains to be exposed to either vesicular contents or the extracellular environment. Surface TLRs interact largely with microbial membrane components while endosomal TLRs are able to interact with genetic material from pathogens that have been internalized and broken down[13]. Many different types of pathogens can be detected despite the mechanism used to establish infections.

TLR expression is complex and differs between cell types. Expression is most significant in different white blood cell types such as macrophages, mast cells, and dendritic cells. This allows macrophages and mast cells to initiate innate immune responses and dendritic cells to initiate adaptive immune responses[13]. Ultimately, TLR signaling leads to activation of interferon regulatory factors (IRFs), activator protein 1(AP-1), or NF-κB transcription factor family members. These transcription factors are responsible for producing the antimicrobial effects; NF-κB and AP-1 mainly induce expression of pro-inflammatory cytokines and chemotactic factors while IRFs induce expression of antiviral type1 interferons[1]. The location and activating-ligand of each TLR define what products are expressed, antiviral or antibacterial.

Figure 2: Human Toll-like receptors and antigens they recognize.
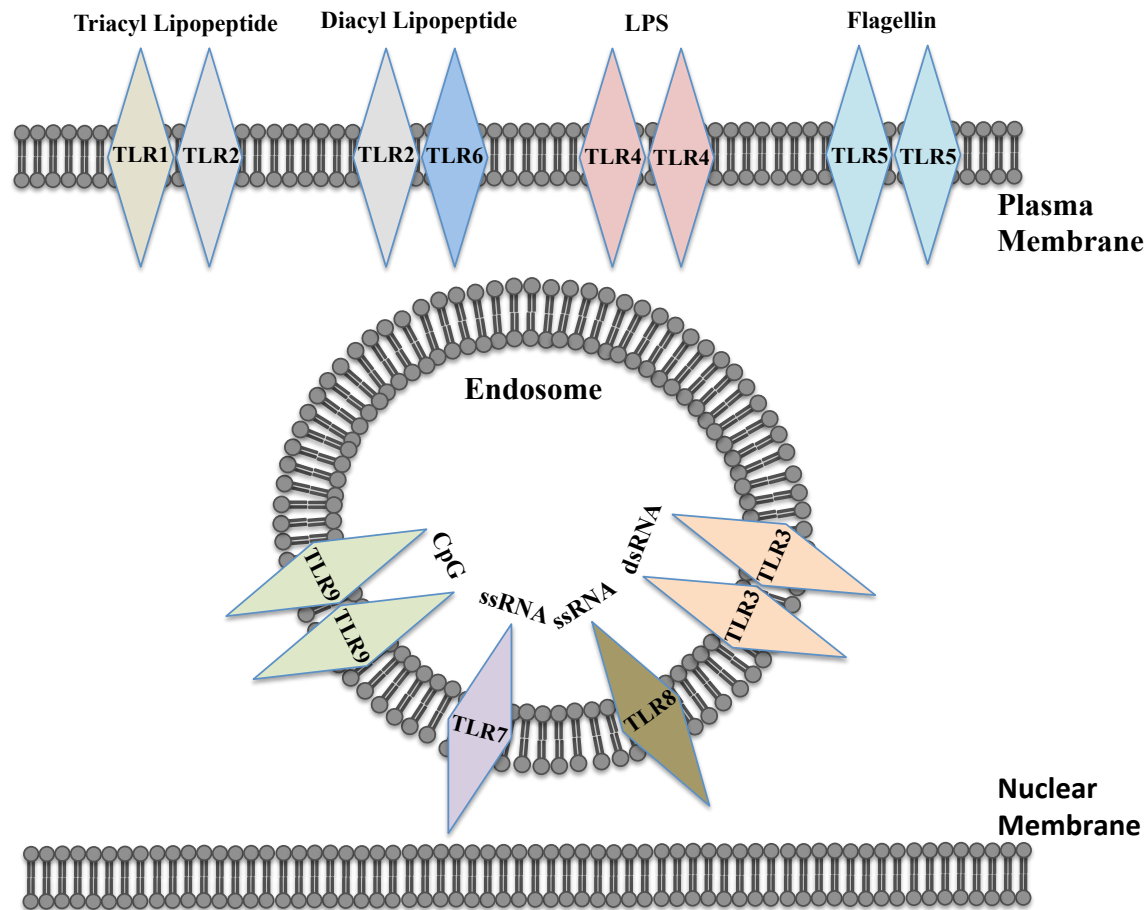
Figure 2: Human Toll-like receptors and antigens they recognize. Toll-like receptors are found on either the plasma or endosomal membrane, activated by various pathogen-associated molecular patterns.

1.4 Toll-Like Receptor 3: Role in Innate Immunity

The TLR of the pathway focused on for these studies, TLR-3, is expressed among different cell types[1,2]. As previously mentioned, TLR-3 is found on the membranes of endosomes and lysosomes and detects genetic material exposed after the pathogen has been internalized and digested. Among all of the Toll-like receptors, it is the only one that is capable of recognizing dsRNA[14]. Once the vesicular material is identified as foreign, a signal is generated to activate transcription factors to initiate an innate immune response is initiated.

Mammalian TLRs are active when ligand-binding stimulates formation of dimers or oligomers[3]. Ligand-induced dimerization of two TLR-3 ectodomains brings the two cytoplasmic TIR domains into close proximity, allowing them to interact with specific adaptor molecules that initiate the intracellular signaling[1,2,11]. Combinational use of different adaptor molecules allows for diverse and distinct biological effects of activating different toll-like receptors. Four adaptor molecules used by TLRs are: myeloid differentiation factor 88 (MyD88), MyD88 adaptor-like (MAL), TIR domain-containing adaptor-inducing IFN-β (TRIF), and TRIF-related adaptor molecule (TRAM)[1,11]. While other TLRs may use one or more of them, TLR-3 interacts with only one, TRIF, to induce activation of NF- κB and MAPK (mitogen-activated protein kinases) in MyD-88-deficient macrophages[11].

TRIF is the most crucial adaptor protein in the TLR-3 pathway and also provides a connection to a branch of TLR-4 signaling that converges downstream in the TLR-3 pathway[15]. Overexpression or deficiencies of TRIF alone can have adverse effects on the immune response generated by those TLR pathways[15]. TRIF has distinct protein-

interaction motifs that allow the recruitment of various effector proteins, which results in one of at least three possible signaling cascades for TLR-3 activation (Figure 3)[1,2].

Two of the three possible signal cascades initiated by TLR-3 activation involve the recruitment of TRIF, however, one results in NF-κB expression, while the other induces IRF-3[1,2,15]. TRIF recruits TRAF-6 to its N-terminal domain via a specific TRAF-6-binding sequence, which leads to the recruitment of RIP1 (receptor-interacting protein 1) kinase to the C-terminal domain of TRIF through interaction of their RIP homotypic interaction motifs (RHIMs)[1,15,16]. At this point, the two pathways diverge depending on the ubiquitination state of RIP1.

If ubiquitination of lysine 377 on RIP1 kinase does not occur, it will trigger a cell death-associated response. FADD (Fas-Associated with Death Domain protein) and Caspase 8 are recruited to complex with RIP1 to initiate downstream signaling. Activation of Caspase 8 leads to promotion of cell death by triggering the receptor-mediated apoptotic pathway[2,16]. This outcome suggests that this pathway is an important host defense for limiting the spread of viral infections[16].

Ubiquitination of Lysine residue 377 on RIP1 initiates recruitment and binding of ubiquitin acceptor proteins, TAK1 (transforming growth factor β-activating kinase 1) and TAB2 (TAK binding protein 2)[15,16]. Activation of TAK1 causes the complex to be released in the cytoplasm to phosphorylate, and activation IKK (IκB kinase complex), a heterodimer composed of IKKα and IKKβ subunits, as well as activate of NEMO (NF-κB essential modulator). In turn, the activated IKK complex phosphorylates two serine residues of IκBα (Inhibitor of κBα), which is normally sequestering NF-κB complex, RelA (p65) and NF-κB1, in a dormant state until this modification[15,16]. Phosphorylation,

along with further ubiquitination, targets IκBα for degradation by proteasome and allows translocation of the NF-κB complex into the nucleus. Once in the nucleus, it functions as a transcription factor, inducing the expression of genes involved in the immune response, cellular growth and survival responses, as well as forming an inhibitory feedback loop by up-regulating the expression of its own receptor, IκBα[2,16,17].

The third signaling cascade requires the recruitment of effector molecules TRAF1 (tumor necrosis factor receptor-associated factor 1), TRAF2, TRAF3, and TRAF6[1,2,15,16]. These proteins mediate the signal transduction pathway by functioning as molecular bridges that bring other effector proteins together. Once recruited to the activated receptors, these proteins mediate the activation of TBK1 (TANK-binding kinase 1)[1,2]. Exclusive interactions of three adaptors: TANK (TRAF-associated NK-κB inhibitor), NAP1 (NF-κB activating kinase (NAK)-associated protein 1), and SINTBAD (similar to NAP1 TBK1 adaptor), with two kinases: TBK1 and IKKε (IκB kinase epsilon), leads to activation of these kinases[2,15,16]. TBK1 and IKKε are responsible for phosphorylation and activation of certain transcription factors such as IRF-3 and IRF-7. Seven Ser/Thr sites can be phosphorylated near the C-terminus of IRF3, $^{385}SSX_9SXSXXXSXTS^{405}$. Phosphorylation of Ser-396 to Ser-405 induces interaction with CBP/p300, while phosphorylation of Ser-385 and Ser-386 activates IRF3 dimerization. Once it is activated and forms a dimer, IRF3 can transmit the signal into the nucleus to initiate the antiviral immune response through regulation of type 1 interferons[1,2,15,16]. These interferons (IFN-α and IFNβ) help stimulate macrophages and NK cells to either elicit an anti-viral response or a positive feedback loop by binding the IFNα/β receptors to activate the JAK

(Janus Kinase)/STAT (signal transducer and activator of transcription) pathway and expression of more anti-viral genes[2,15].

These three TLR-3-mediated signaling pathways initiate recruitment of immune cells, inhibition of pathogen replication, induction of apoptosis, and communication of danger signals to nearby cells. Because of its importance in innate immunity, our work has focused on the pathway involving IRF3 activation, the transcription factor that functions in production of type 1 interferons. In particular, we have examined the structure of a recently identified protein, which is known to play a role in this pathway. SIKE is found in phosphorylated and un-phosphorylated states; the functions of both remain unknown. As such, understanding the structure of SIKE may provide insight into its function in the TLR-3 pathway.

Figure 3: TRIF-dependent signaling cascades
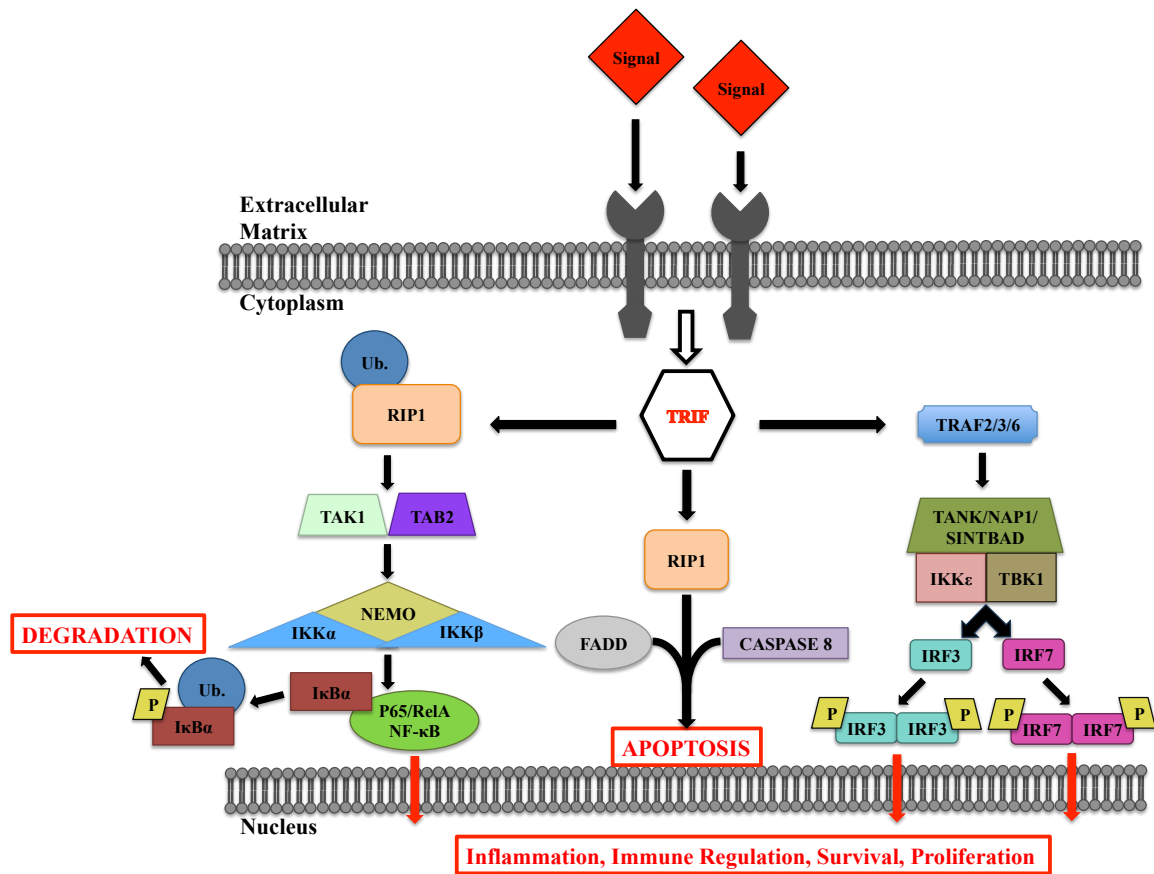
Figure 3: TRIF-dependent signaling cascades. Activation of TRIF leads to three possible signal cascades, initiating various cellular processes.

1.5 Suppressor of IKK-ε (SIKE): Discovery, Structure, and Function.

Several innate immune signaling pathways converge to activate TBK1, as previously mentioned, through receptor-mediated pathogen detection. Dysregulation of TBK1 has been linked to cancers and autoimmune disorders, intensifying the need to understand the regulatory mechanisms used by the proteins in these pathways[2,18]. Known mechanisms that modify the activity of TBK1 involve altering TBK1 ubiquitination or masking phosphorylation sites by interacting proteins in the cascade[18,19]. Investigating the functions of such interaction partners can reveal information about how the pathway is regulated in normal environmental conditions.

Using a yeast two-hybrid screen with a human B-cell cDNA library and IKKε as bait, a previously unknown interaction protein was identified and dubbed SIKE (for suppressor of IKKε)[19]. SIKE is a 207-amino acid protein encoded by a gene that maps to human chromosome 1p13.2[2,18]. GenBank database searches identified SIKE as an evolutionarily conserved protein that is part of an uncharacterized family of proteins like fibroblast growth factor receptor 1 oncogene partner 2 (FGFR10P2), which mapped to chromosome 12p12.1[2,18,20]. Although SIKE differs from FGFR10P2 by 40 amino acids, they were shown to be about 50% homologous. Additionally, studies showed that FGFR10P2 has the ability to dimerize, associate with coiled-coil structures, and interact with cytoskeleton networks of fibroblasts to enhance oral wound healing[18,19,20]. Since coiled-coil motifs are known to mediate protein-protein interactions, researchers questioned if SIKE could oligomerize. Analysis of SIKE revealed that it had three predicted coiled-coil domains, two of which were high probability predicted domains

(amino acids 72-102 and 163-197)[2,18,19]. Co-immunoprecipitation experiments showed that SIKE could interact with itself to form homodimers and oligomers (Figure 4)[19].

It was shown that SIKE formed interactions with IKKε, however, this kinase is not detectable under basal physiological conditions, but is induced by activated immune cells. Researchers then questioned whether SIKE could interact with TBK1, which is constitutively expressed in most cell types. Co-immunoprecipitation assays showed that SIKE not only interacted with TBK1 under physiological conditions, but it was revealed that the high probability coiled-coil domains of SIKE (amino acids 72-207) interacted with the coiled-coil domains of TBK1 (amino acids 601-729)[19]. This suggests that the coiled-coil motifs did mediate interactions between these two proteins. It also shows that the first 72 residues of SIKE are not necessary for interactions with TBK1, but failed to show any interactions between SIKE and the kinase domain of TBK1[2,19].

In mammalian overexpression systems, it was shown that SIKE associates with TBK1 in untransfected cells, but dissociates upon viral infection or TLR3 stimulation[18,19]. Data indicated that SIKE did not disrupt any interactions required by proteins for TLR3-mediated NF-κB activation, suggesting that SIKE specifically inhibited the IKKε- and TBK1-mediated IFN-β activation pathways[19]. Subsequent research performed by Andrei Medvedev's group indicated that LPS-induced pathway activation increased levels of SIKE mRNA, which suggests that transcription factors activated by these pathways control SIKE expression[21].

Aiming to characterize the competition between TBK1 substrates, IRF3 and SIKE, the Bell Laboratory showed that SIKE acts as a mixed type inhibitor with respect to IRF3 phosphorylation[18]. Data showed that both substrates had similar K_m values (~50

nM) at low concentrations, but SIKE displayed a K_m more than 8 times lower (IRF3 = 3.5 µM; SIKE = 0.4 µM) at higher substrate concentrations. In other words, at higher concentrations, TBK1 has a higher affinity for phosphorylating SIKE than IRF3, thus inhibiting IRF3 dimerization, translocation into the nucleus, and regulation of Type 1 interferon gene transcription. It was also shown that SIKE is phosphorylated at six sites that mimic the phosphorylation motif of IRF3. This phosphorylation modulates TBK1:SIKE interactions and correlates with the initiation of the antiviral response[18]. Once SIKE has been phosphorylated, the affinity of TBK1 for SIKE reduces drastically, allowing dissociation of the two proteins. Currently, the function of phosphorylated SIKE remains unknown. Using these findings, the Bell Lab proposed that SIKE functions, not as an inhibitor of TBK1-mediated phosphorylation of IRF3, but as a higher-affinity TBK1 substrate (Figure 5). Due to the fact that phosphorylation requires enormous amounts of energy and can cause immense changes in a proteins structure, the basis of this research is to investigate the function of phosphorylated SIKE once it has been released by the TBK1.

Figure 4 Hypothesized domain organization of SIKE and its kinase, TBK1.
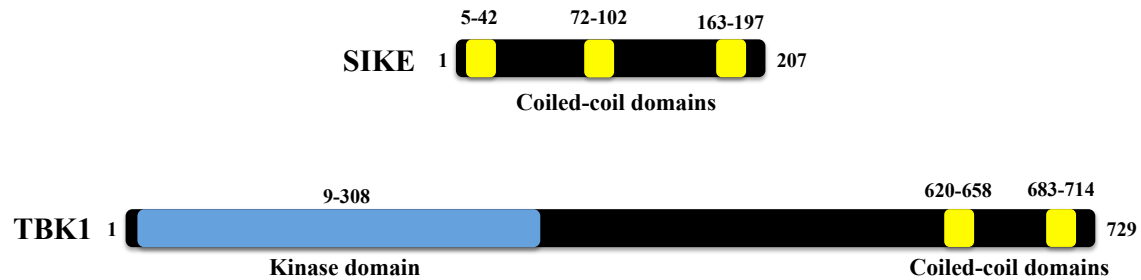
Figure 4 Hypothesized domain organization of SIKE and its kinase, TBK1. SIKE (top) is 207 amino acids in length and has three predicted coiled-coil domains. Its kinase, TBK1 (bottom), is 729 amino acids long and also contains two coiled-coil domains that are thought to mediate interactions with SIKE.

Figure 5: SIKE interactions in the TLR3 signaling pathway (Jessica K. Bell).
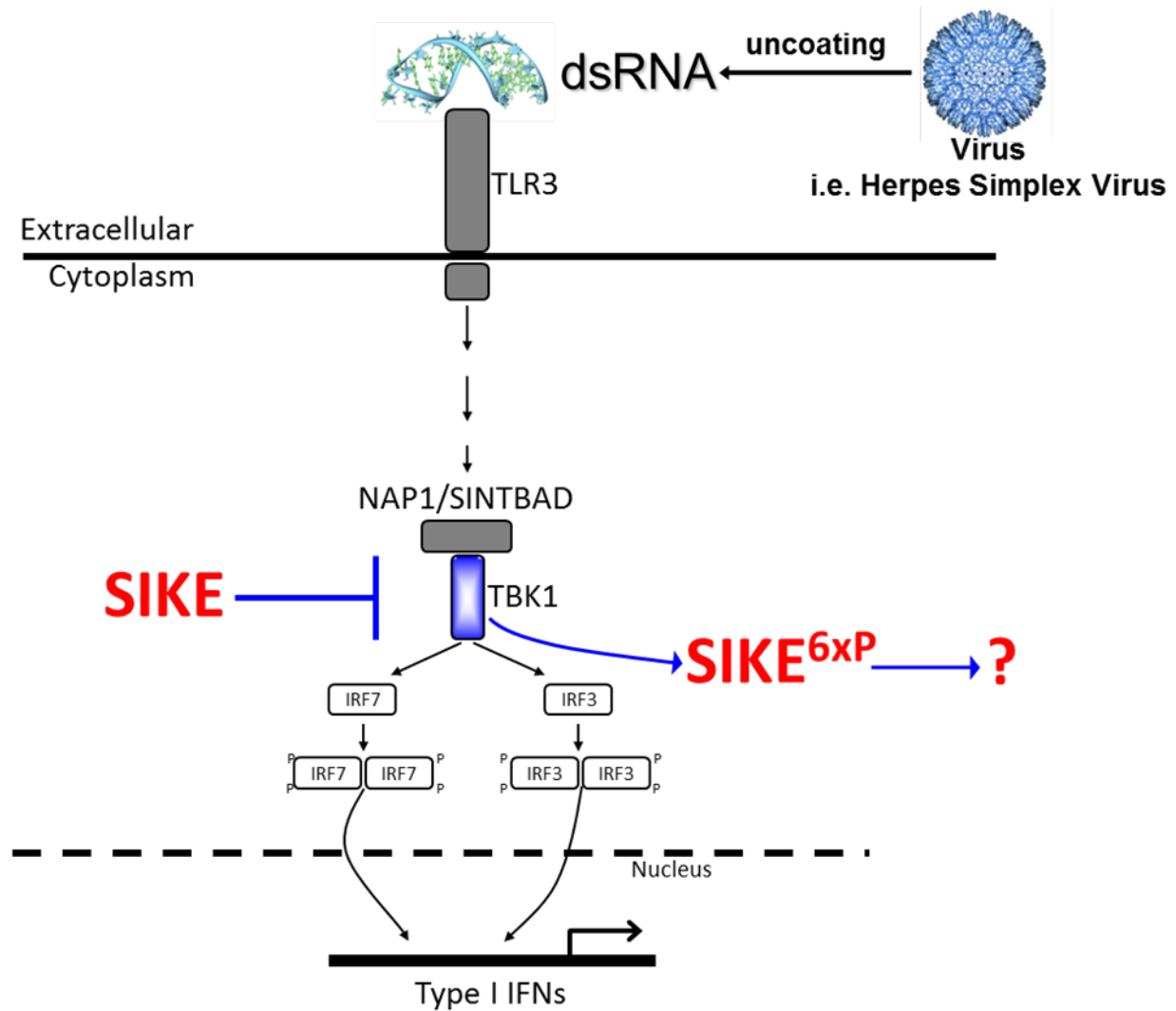
Figure 5: SIKE interactions in the TLR3 signaling pathway (Jessica K. Bell).

In the TLR3 signaling cascade, SIKE is known to function as a TBK1 substrate and subsequently inhibiting phosphorylation of IRF3. The function of SIKE once it has been phosphorylated remains unknown.

1.6 Current Study

To fully comprehend the function of proteins in a cellular mechanism or pathway requires in-depth knowledge of their overall structure and amino acid composition. Simply knowing pieces of this can only be suggestive of a protein's function[1,18,22]. Primary amino acid sequences can be used to predict secondary structural elements such as membrane-spanning α-helices. However, a tertiary structure cannot be reliably deduced unless the sequence is very similar to a protein whose three-dimensional structure is already known[22]. A protein's function can be predicted using other methods, such as cellular localization, identifying charged surface areas, or locating any active sites or catalytic pockets[1,22]. The main technique used to uncover three-dimensional structures at atomic level is X-ray crystallography.

To further characterize SIKE's function within the antiviral response, this study focused on defining the overall SIKE structure through X-ray crystallography. To accomplish this, the Bell lab created different constructs in hopes to increase chances of obtaining crystals. Two of the constructs were: a full-length SIKE construct (SIKE FL) and a construct that has had the first 71 amino acids removed (SIKE 72); these were the two used for this study. Mutations from serine to alanine were performed on TBK1 phosphorylation sites for each SIKE construct, known as S6A. This was shown to make SIKE more stable and reduced the number of species present in SEC (size exclusion chromatography). Using recombinant protein expressed from *E. coli* and purified via immobilized metal affinity chromatography, SIKE crystals were previously obtained from a sample concentrated to 15 mg/ml under several crystallization conditions. Yet, reproducing these results has been difficult.

In this study, we have modified the purification scheme to remove an *E. coli* contaminant, SlyD. Purification under denaturing conditions, removal of soluble proteins, incorporation of ion exchange and different IMAC resins were tested. For each scheme, size exclusion chromatography and SDS-PAGE/Coomassie/silver stain were used to assess purity. Crystallization trials for samples from each purification scheme were completed. In addition to crystallization trials, Hydrogen-Deuterium exchange was investigated, accompanied with pepsin digests, in order to further characterize the dynamic nature of SIKE structure. While these studies investigate the properties of recombinant SIKE using various methods, the findings as a whole establish a previously undefined protocol for SIKE purification that will improve future studies, as well as expanding our knowledge of the surface structure of SIKE, which will enhance our understanding of its function in the innate immune system.

Chapter 2: Materials and Methods

2.1 Bacterial Constructs and Recombinant SIKE expression

Constructs used in this study: Full-length SIKE, SIKE 72-207, and their corresponding S6A constructs were previously reported[2]. For construct expression, recombinant pET15b vectors were transformed into BL21-CodonPlus (DE3)-RIPL competent *E. coli* cells following the manufacturer's protocol (Agilent Technologies).

A culture of Luria broth (LB) and 100μg/mL ampicillin was inoculated with a single colony and left to incubate overnight, shaking at 200rmp at 37°C. The following day, sub-cultures (1:100 dilution) were made in 2-1L flasks containing LB and 100μg/mL ampicillin and incubated at 37°C, shaking at 200rpm until they reached an $A_{600} \approx 0.6$. At this point, each 1L culture was induced using 1mL of 1M isopropyl-β-galactopyranoside (IPTG) and allowed to grow for another 4 hours in the same conditions before harvesting the cells. In addition to those conditions, we investigated the effects of slowing down growth by icing for 30 minutes pre-induction, inducing with IPTG, and incubating at 16°C (200rpm) overnight before harvesting the cells.

Overnight Express™ Instant LB (auto-induction media) was also used to culture bacteria that had been transformed with SIKE constructs. Prepared using manufacturer's protocol (Novagen), 2-1L cultures of auto-induction media (100μg/mL ampicillin) were inoculated, 1:100, using the same regular LB overnight culture grown to an $A_{600} \approx 0.6$, as

previously mentioned for the normal LB prep. These 2L cultures were left to incubate

overnight at 37°C, shaking at 200rpm (no IPTG is needed to stimulate protein expression

using auto-induction media).

2.2 Harvesting Inclusion Bodies

2.2.1 Chemical Lysis

Cells were harvested by centrifugation at 7,000rpm for 20 minutes at 4°C. Pellet

was collected and solubilized in a guanidine hydrochloride (GudHCl) buffer (6M

guanidine hydrochloride, 100mM sodium phosphate (pH8.0), 500mM NaCl, and 1mM 2-

mercaptoethanol) and brought to a pH of 8.0. After stiffing to solubilize the pellet, cell

lysates were pelleted through centrifugation at 12,000rpm for 30 minutes at 4°C. The

supernatant was mixed with 5mL resin that had been pre-equilibrated in GudHCl buffer.

2.2.2 Freeze-thaw Lysis

Cells were lysed using a freeze-thaw technique. Cell pellet was suspended in

Buffer 1 (100mM sodium phosphate (pH8.0), 500mM NaCl, and 1mM 2-

mercaptoethanol) with added protease inhibitors (complete EDTA free, Roche). Once in

solution, tubes were placed in an ethanol/dry ice bath until frozen and then placed in 37°C

water bath until thawed. This process was repeated two additional times and then the cell

lysate was spun down at 14,000rpm for 15 minutes at 4°C. Pellet was suspended once

more in Buffer 1 and centrifugation was repeated. Pellet was collected and re-suspended

in GudHCl lysing buffer, stirring at room temperature for one hour. Cell lysates were

pelleted through centrifugation at 12,000rpm for 30 minutes at 4°C. Supernatant was

then mixed with 5mL resin that had been pre-equilibrated in GudHCl buffer.

2.2.3 Sonication Lysis

A third technique used to harvest SIKE from the transformed cells was sonication.

Pellet was suspended in ~75mL Buffer 1 with added protease inhibitors. While on ice,

cells were sonicated for 30 seconds (100% power). This was repeated 10-15 times,

waiting 1 minute between each sonication, and the resulting lysate was pelleted in a

centrifuge at 12,000rpm for 20 minutes at 4°C. Pellet was suspended once more in

Buffer 1 and centrifugation was repeated. Pellet was collected and re-suspended in

GudHCL buffer, stirring at room temperature for one hour. Cell lysates were pelleted

through centrifugation at 12,000rpm for 30 minutes at 4°C. Supernatant was then mixed

with 5mL resin that had been pre-equilibrated in GudHCl buffer.

2.3 Refolding Denatured SIKE

2.3.1 Ni-NTA Affinity Resin

Ni-NTA agarose resin (Qiagen) was used to bind recombinant protein construct

containing an N-terminal 6xHistadine tag (pET15b). 5mL of resin was pre-equilibrated

in the same lysing buffer as the cell lysate. The cell lysate was then added to the resin

and left on a rotator for at least one hour at room temperature to batch-bind. Resin-lysate

mixture was loaded into a BioRad Econo column and washed with 150mL GudHCl

buffer. Protein bound to the resin was refolded on the column using a 40-column-volume

(CV) reverse gradient of GudHCl buffer to Buffer 1 and eluted using 6 column volumes

of elution buffer (Buffer 1 with 500mM imidazole); 6-5mL tubes were collected on ice and stored at 4°C.

2.3.2 TALON® Affinity Resin (ClonTech)

IMAC using refolded protein was performed using one NiNTA elution fraction (E3) that had been dialyzed into TALON® Buffer 1 (50mM sodium phosphate, 300mM NaCl, and 1mM 2-mercaptoethanol) to remove imidazole. This was then combined with TALON® resin that had been pre-equilibrated in the same TALON® Buffer 1 and left to rock at room temperature for at least one hour at room temperature. The mixture was loaded into a BioRad Econo column and washed with 150mL TALON® Buffer 1. Protein bound to the resin was eluted using 6 column volumes of TALON® elution buffer (TALON® Buffer 1 with 500mM imidazole); 6-5mL tubes were collected on ice and stored at 4°C.

IMAC was performed on TALON® resin with denatured protein, using the same procedure used for the NiNTA resin, however, the lysing buffer (6M Guanidine Hydrochloride, 50mM sodium phosphate, 300mM NaCl, and 1mM 2-mercaptoethanol) and Buffer 1 deviated, slightly, from the Ni-NTA buffers. Once the cells had been harvested and resin had been pre-equilibrated in the lysing buffer, they were combined and left on a rotator at room temperature for at least one hour. Resin-lysate mixture was loaded into a BioRad Econo column and washed with 150mL TALON® lysing buffer. Protein bound to the resin was refolded on the column using a 40-CV reverse gradient of TALON® lysing buffer to TALON® Buffer 1 and eluted using 6 column volumes of

TALON® elution buffer (TALON® Buffer 1 with 500mM imidazole); 6-5mL tubes were collected on ice and stored at 4°C.

2.4 Purifying and Concentrating Proteins

2.4.1 Size Exclusion Chromatography

Each elution fraction was centrifuged at 14,000xG for 15 minutes to remove any precipitate before being loaded on a HiLoad 16/60 Superdex 200 column (GE Healthcare) to separate the remaining contents by size. The buffer used in filtration was: 50mM HEPES (4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid), 500mM NaCl, and 1mM 2-mercaptoethanol. Peak fractions were collected and protein was screened for using SDS-PAGE (sodium dodecyl sulfate polyacrylamide gel electrophoresis) analysis and gels were silver stained to confirm purity. Pure fractions of each SIKE species were combined and concentrated for crystallization trials.

2.4.2 Anion Exchange

Using dialysis tubing with a 10,000 Dalton molecular weight cut off, IMAC elution fractions were dialyzed overnight in a low-salt ethanolamine buffer (50mM ethanolamine (pH9.0), 50mM NaCl, and 1mM 2-mercaptoethanol) and then loaded on MonoQ XL ion chromatography column (GE healthcare) to separate remaining contents by anion exchange. Buffer salt concentration was gradually increased to 1M as fractions were collected; protein purity was screened using SDS-PAGE and silver staining techniques. Pure fractions from each SIKE species were combined and concentrated for crystallization trials.

2.4.3 Cation Exchange

Using dialysis tubing with a 10,000 Dalton molecular weight cut off, IMAC elution fractions were dialyzed overnight in a low-salt sodium acetate buffer (50mM sodium acetate (pH 4.5), 50mM NaCl, and 1mM 2-mercaptoethanol) and then loaded on HiTrap SP XL ion chromatography column (GE healthcare) to separate remaining contents by cation exchange. Buffer salt concentration was gradually increased to 1M as fractions were collected; protein purity was screened using SDS-PAGE and silver staining techniques. Pure fractions from each SIKE species were combined and concentrated for crystallization trials.

2.4.4 Concentration via Centrifugation

Purified SEC fractions containing the same species were combined into one tube and stored at 4°C. Using Amicon Ultra Centrifugal Units® from Millipore (10,000 Dalton molecular weight cut off), pooled fractions were concentrated to 15mg/mL (or as close as possible) using the manufacturers protocol. Protein concentration was quantified using absorbance at 260nm and 280nm. Concentrating was stopped when 15mg/mL was reached or when precipitate was visible. To minimize precipitate formation, 1mM 2-mercaptoethanol or 10% glycerol was added and mixed thoroughly before continuing centrifugation. When complete, the concentrated sample was transferred to a sterile, Eppendorf® LoBind protein micro-centrifuge tube and stored at 4°C.

2.4.5 Concentration via Lyophilization (or Cryodesiccation)

Purified SEC fractions containing the same species were combined into one tube and stored at 4°C. Using dialysis tubing with a 10,000 Dalton molecular weight cut off, the pooled fractions were dialyzed for 24 hours in buffer containing 20mM ammonium bicarbonate and 1mM 2-mercaptoethanol. Solution was then added to a round-bottom flask and attached to a lyophilizer to freeze-dry overnight (0.090 mBar and -46°C). The protein-containing powder was collected and solubilized in gel filtration buffer or 50mM phosphate buffer (for crystallization trials or MALDI-TOF analysis respectively) and brought up to a concentration of 15mg/mL. Protein concentration was quantified using absorbance at 260nm and 280nm.

2.5 Crystallization trials

2.5.1 Optimized Crystallization Conditions

Previously, the lab sent samples of protein to Hauptman Woodward Institute's (HWI) high-throughput screening laboratory to prepare crystal-growth screening experiments in 1536-well microassay plates. Conditions that resulted in the best crystal formation (Table 1) were recorded and optimized in the lab with the purpose of obtaining X-ray diffraction quality crystals. All three of the HWI screen hits were subjected to fine grid screens to optimize crystal growth, altering the pH and PEG (polyethylene glycol) concentration as shown in Table 2.

Table 1: HWI Protein Crystallization Screen hits

HWI Crystal Screen Hits	Buffer (0.1M)	Salt (0.1M)	Precipitant (40% w/v)	pH
337	Magnesium chloride hexahydrate	Sodium acetate	Polyethylene glycol 8,000	5.0
371	Magnesium nitrate hexahydrate	Sodium acetate	Polyethylene glycol 4,000	5.0
863	Calcium chloride dihydrate	Sodium acetate	Polyethylene glycol 4,000	5.0

Table 2: PEG percentage and pH for the 16 optimized crystallization conditions

30% PEG pH 5.1	35% PEG pH 5.1	40% PEG pH 5.1	42% PEG pH 5.1
30% PEG pH 5.0	35% PEG pH 5.0	40% PEG pH 5.0	42% PEG pH 5.0
30% PEG pH 4.8	35% PEG pH 4.8	40% PEG pH 4.8	42% PEG pH 4.8
30% PEG pH 4.6	35% PEG pH 4.6	40% PEG pH 4.6	42% PEG pH 4.6

2.5.2 Paraffin Oil Drop

Crystallization Trays were set up using 72-well Terasaki Plates (Greiner Bio-One). Wells were submerged using ~8mL of Paraffin oil before adding 2μL of each condition into its designated well. Concentrated protein sample was centrifuged at 14,000xG for 15 minutes to remove any precipitated protein and 2μL was carefully added to each well containing the optimized solutions. Plates were covered with a plastic lid and left untouched for ~3 days and then observed for crystal growth each day. Scores were made based on the amount and type of precipitate seen.

2.5.3 Sitting Drop

Crystallization screens were set up using a 96-well deep well plate. Selected conditions from various Qiagen and Hampton Research Crystallization Screen Kits (Table 3) were used to screen with SIKE 72 S6A (3.95mg/mL) for crystal formation. About 0.5mL of each selected condition was dispensed in the reservoir of its designated well. 0.2μL of the protein was loaded into the smaller sample well and topped with 0.2μL of the condition from the reservoir. The plate was carefully covered with optically clear sealing film and left untouched for ~3 days before being observed for crystal growth. Scores were made based on the amount and type of precipitate seen.

Table 3: Qiagen/Hampton Research screen kits and their corresponding Screen Hit/condition numbers used in sitting drop crystallization trials.

Source	Screen	Screen Hit/Condition
Qiagen	pHClear Suite	All that contain 0.1M HEPES or 0.1M Tris as buffer
Qiagen	PEGs Suite	13-24, 37-48, 51, 57, 60, 67, 71, 78, 83, 86, 91, 92, and 96
Qiagen	AmSO$_4$ Suite	1-11, 76, 77, 79, 82, 83, 88, and 94
Hampton Research	Crystal Screen-1	2-37 and 39-46
Hampton Research	PEGRx-2	1-47

2.5.4 HWI screen

A sample of SIKE 72 S6A (12.9 mg/mL) and one sample of plain filtration buffer were sent to HWI's High-Throughput Crystallization Lab to be screened with their available conditions and photographed daily for crystal formation. 0.2μL of the protein sample was combined with 0.2μL of each HWI condition; a control was also set up using 0.2μL of plain buffer instead of protein sample. These wells were photographed weekly and observed electronically for the appearance of crystals or changes in precipitate formation.

2.6 Hydrogen-Deuterium Exchange

Total volume of protein needed was determined by the amount of time points desired (~2 μL for each). Protein samples (between 4 and 12mg/mL) were diluted, 2:18, with deuterium oxide (^2H_2O or D_2O) at room temperature, and a timer was started. At each scheduled time point, 20μL of exchanged sample was removed and quenched with 180μL of TFA-d (deuterated trifluoroacetic acid) and immediately frozen on ice. Time points used for our testing were 0sec, 30sec, 1min, 1.5min, 2min, 3min, 5min, 10min, 15min, 30min, 45min, 1hr, 1.5hr, 2hr, 3hr, 4hr, 5hr, 6hr, 7hr, and 8hr. Exchanged samples were stored on dry ice until analyzed. D_2O exchange matrix buffer was made (50μL TFA-d, 200μL D_2O, 250μL acetonitrile, and 5-10mg sinapinic acid) and used to pre-spot MALDI plate for each sample including one for a control (phosphate buffer) and one for a calibration (bovine serum albumin (BSA)) sample. Plate was then vacuumed in a desiccator for ~20 minutes until pre-spots were dry. Results are best when exchange between samples and atmosphere is minimized following quenching, so tubes were kept

tightly sealed when not in use. 5μL of each exchanged sample was thawed and added to 5μL of matrix buffer. This was mixed quickly before plating 2μL on designated pre-spot on the plate. 2μL of BSA sample, phosphate buffer sample, and exchange samples from each time point were plated twice in case one could not be read properly. Once complete, plate was vacuumed in desiccator for another 20 minutes, until spots are dry. Time points were analyzed using MALDI-TOF (matrix assisted LASER desorption/ionization-time of flight), taking 1000 profiles for each spot.

2.7 Pepsin Digest

Total protein needed for all time points was calculated. Protein samples were diluted 1:100 with 50mM phosphate buffer to the same concentration as the deuterium exchange samples after being quenched with TFA-d. 50% immobilized pepsin mixture (Pierce) was re-suspended and 0.25mL was added to 4mL of digestion buffer (20mM sodium acetate, pH 3.5). Mixture was centrifuged at 1000xG for 5 minutes. Supernatant was discarded and pellet was re-suspended in another 4mL of digestion buffer before another 5 minutes of centrifugation at 1000xG. Supernatant was discarded and pellet was brought up in 0.5mL of digestion buffer. The pepsin mixture was then added to 0.5mL of diluted protein sample and left to incubate by shaking or inversion at 37°C. At each scheduled time point, 50μL was removed and pepsin was separated from the protein through centrifugation at 3,000xG for 3 minutes. Time points used for our testing were 0min, 1min, 5min, 10min, 15min, 30min, 1hr, 2hr, and 8hr. Supernatant, containing protein sample, was then removed and stored at room temperature until plated. Pepsin digest matrix buffer was made (5-10mg α-cyano-4-hydroxycinnamic acid dissolved in

500μL of 50% D$_2$O, 50% acetonitrile, and 0.1%TFA) and used to pre-spot MALDI plate

for each sample and low-molecular-weight calibration standards. The plate was then

vacuumed in the desiccator for ~20 minutes until pre-spots were dry. 5μL of each digest

time point sample was added to 5μL of matrix buffer. This was mixed before plating 2μL

on designated pre-spot on the plate, two spots for each sample were used. Standards were

plated for low-molecular-weight calibration, along with each time point sample. Once

complete, plate was vacuumed in desiccator for another 20 minutes, until spots were dry.

Time points were analyzed using MALDI-TOF, taking 1000 profiles for each spot. All

theoretical fragments that can result from pepsin digest of SIKE 72 S6A were acquired

(Protein Prospector) used to compare with experimental digest data.

Chapter 3: Results

3.1 Original SIKE Expression and Identification of Contamination

As indicated in the methods section, the original protocol included protein expression in *E. coli* grown from regular LB, chemical cell lysis in GudHCl, refolding on Ni-NTA resin, polishing through SEC, and centrifugal concentration. Expression of SIKE FL S6A was confirmed (after refolding) using SDS-PAGE, showing bands at ~29kDa (Figure 6a), which is standard for this protein weighing only 25,788 Daltons. All gels were silver stained to confirm purity of the sample. Expression of SIKE 72 S6A was confirmed using the same technique. Typical for this 17,964 Dalton protein, SIKE 72 S6A bands were seen at ~18kDa (Figure 6b). Also seen in the SIKE 72 S6A gel is an unknown band at ~29kDa (Figure 6b) that eluted in three of the six fractions. Due to the fact that this contaminant could be covered by the bands of interest for the full-length construct (also runs at ~29kDa), we chose to continue most of the study using SIKE 72 S6A to be sure that we collected pure samples only.

To further polish the samples by separating them based on size and shape, Ni-NTA elution fractions were subjected to SEC. Figure 7a shows an SEC chromatogram from one SIKE 72 S6A Ni-NTA elution fraction (E2 from Figure 6b). The protein that eluted first (peaks labeled "1") represents aggregate SIKE 72 S6A protein, while protein that eluted second (peaks labeled "2") was determined to be monomers/dimers of SIKE

72 S6A. However, SDS-PAGE revealed the existence of the same ~29kDa contaminant that was present in the original sample (Figure 7b). Previously, contamination at this level had not been observed and could possibly have been overlooked due to the full-length protein running at the same molecular weight in SDS-PAGE. In addition to using pure SIKE protein for crystallization trials, we elected to devote some time to modifying the existing purification procedure to maximize pure protein yield.

Figure 6: SDS-PAGE results of unmodified purification protocol and identification of contamination.
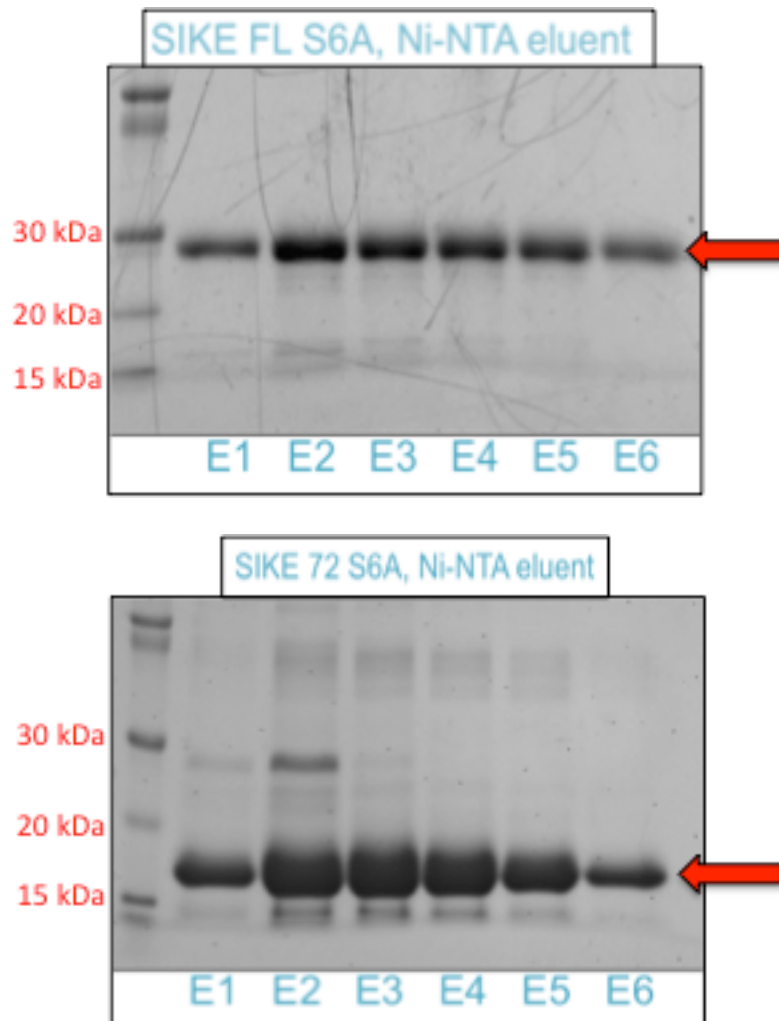
Figure 6: SDS-PAGE results of unmodified purification protocol and identification of contamination. (Top) SIKE FL S6A band separates to ~29 kDa. (Bottom) SIKE 72 S6A separates to 18 kDa, uncovering the contaminant that separates to 27 kDa.

Figure 7: SEC performed on SIKE 72 S6A (Figure 6b, fraction E2), purified using the original protocol
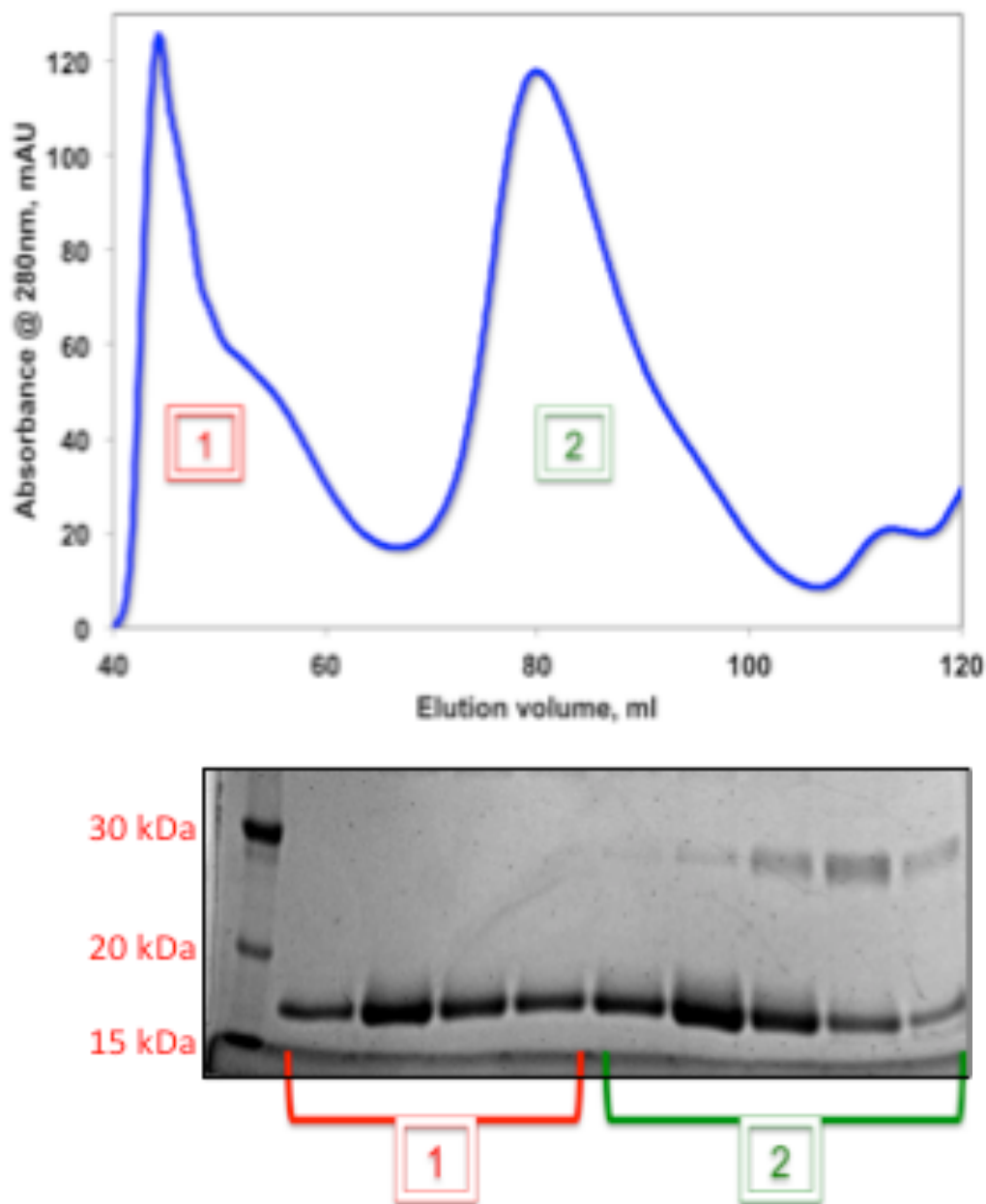
Figure 7: SEC performed on SIKE 72 S6A (Figure 6b, fraction E2), purified using the original protocol. (Top) Chromatogram showing Peak 1, aggregates of proteins, and Peak 2, monomer/dimer SIKE species. (Bottom) SDS-PAGE shows SIKE 72 S6A in all fractions and contamination found in peak 2 fractions.

3.2 Auto-induction media increased soluble protein yield; growth at 16°C did not

One way to amplify the protein yield was to use auto-induction media to induce expression of our protein more efficiently. Different metabolites in the media function to promote cell growth to a high density and automatically stimulate protein expression from IPTG-inducible *lac* promoters. Figure 8 shows SDS-PAGE analysis from expressions using the original LB or the new auto-induction media. The new auto-induction media successfully increased the overall yield of pure SIKE 72 S6A from an average of ~18 mg per 1L of cell culture to over 53 mg per L culture (~3 fold more protein).

Growth at 16°C is known to help reduce aggregation during folding of recombinant proteins[23]. Once cells reach an OD_{600}=0.6-1.0, cells are immediately put on ice to cool before being induced with 1mM IPTG and then grown overnight at 16°C. This helps induce cold shock proteins and chaperones that are active at lower temperatures. SIKE is expressed as insoluble protein in inclusion bodies- this approach was used to examine if lower temperatures could allow SIKE to be expressed as soluble protein. Nonetheless, this method was unsuccessful in increasing the amount of soluble protein and decreasing the amount of aggregate protein (data not shown). For the remainder of these studies, auto-induction media was used for induction of SIKE expression in *E. coli* grown at 37°C.

**Figure 8: Results of SIKE 72 S6A expression with original LB media and auto-induction media.**
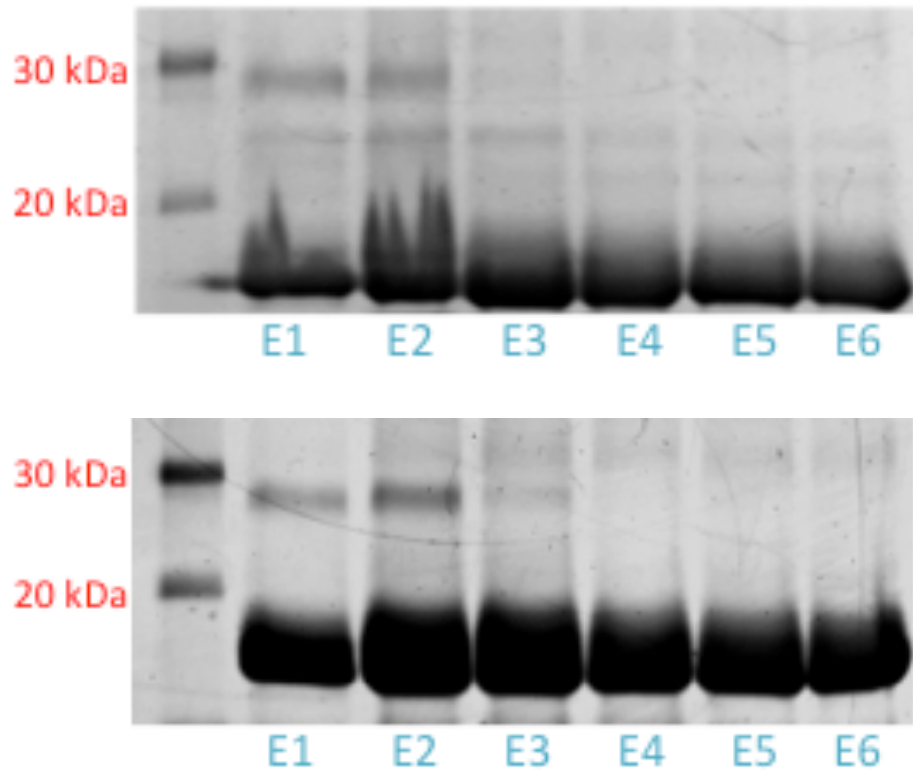
Figure 8: Results of SIKE 72 S6A expression with original LB media and auto-induction media. SDS-PAGE shows considerably more SIKE 72 S6A expression using auto-induction media (bottom) than regular Luria Broth (top).

3.3 Sonication, but not freeze-thaw lysis, decreased amount of soluble contaminant

Proteins that are being over-expressed, like SIKE 72 S6A, are often stored inside intracellular structures called inclusion bodies[24]. These inclusion bodies generally contain very little host protein, suggesting that harvesting SIKE protein from inclusion bodies could minimize contamination. Solubility test results showed most of the SIKE 72 S6A in the insoluble pellet sample, while the contaminant showed up solubilized in the supernatant (Figure 9). Figure 10 shows SDS-PAGE analysis from protein harvesting using the original chemical lysing protocol and the new sonication protocol. Although the first three elution fractions still contained the contaminant, the level of contamination was reduced. Freeze-thaw lysis was ineffective in reducing the amount of contaminant (data not shown). For the remainder of these studies, sonication was used to remove as much soluble contaminant as possible before harvesting inclusion bodies (insoluble portion) for chemical re-suspension.

Figure 9: Solubility test results.
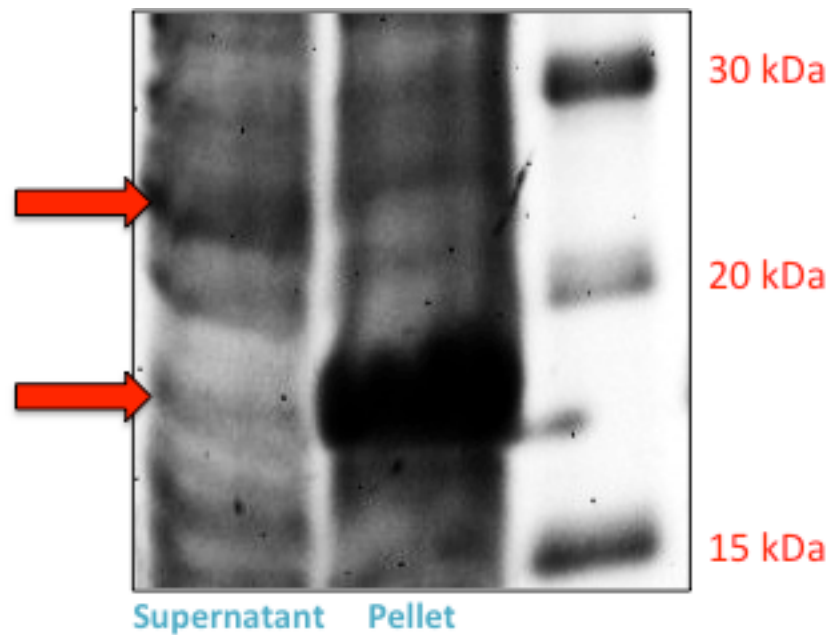
Figure 9: Solubility test results. Most of the SIKE 72 S6A is found in the insoluble pellet sample (bottom arrow), while the contaminant showed up solubilized in the supernatant (top arrow).

Figure 10: Results of SIKE 72 S6A harvesting using chemical or sonication cell lysis.
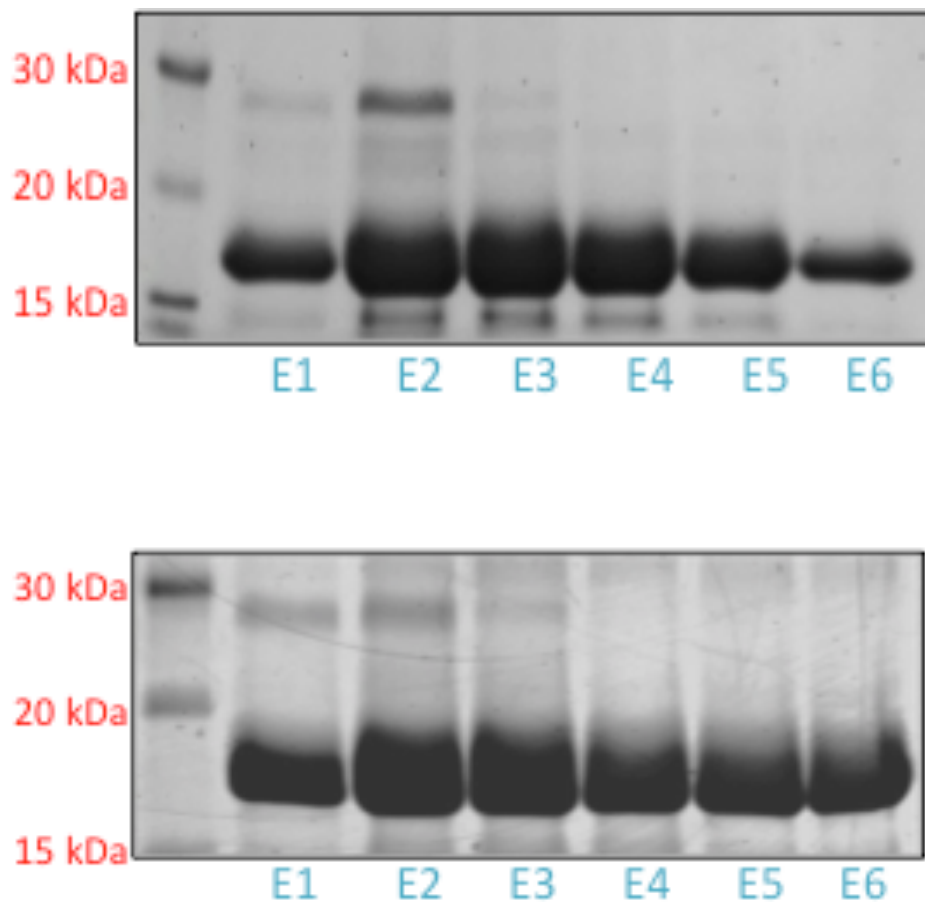
Figure 10: Results of SIKE 72 S6A harvesting using chemical or sonication cell lysis.

SDS-PAGE shows more soluble contaminant when chemical lysis was used (top) than

when sonication is used for cell lysis (botom).

3.4 Anion and cation exchange were unsuccessful in separating bacterial contaminant from SIKE.

Ion exchange chromatography is based on the reversible interaction between a charged protein and the oppositely charged chromatography medium used (http://www.gelifesciences.com). Based on the amino acid sequence, the estimated isoelectric point (pI) for SIKE 72 S6A is 6.2, so the anion buffer had a pH of 9.0, to ensure de-protonation of accessible acid groups; and the cation exchange buffer had a pH of 4.5, to ensure protonation of all basic groups. In this study, elution of proteins used an increasing gradient of NaCl to compete with the charge-charge interaction between protein and resin. It was hypothesized that SIKE and the contaminant would elute off of the column at different times as the NaCl concentration gradually increased from 50mM to 1M.

Figure 11 shows results from the first anion exchange attempt using IMAC elution fraction E1 from an auto-induction/sonication prep (E1 from Figure 10). SDS-PAGE and silver staining was used to confirm that the contaminant was successfully separated from the target SIKE protein; however, the flow-through was not collected so it could not be analyzed and all additional anion exchange trials were all unsuccessful in separating the two proteins (data not shown).

Figure 12 shows results from a cation exchange attempt using IMAC elution fraction E3 from an auto-induction/sonication prep (E3 from Figure 10). SDS-PAGE and silver staining was used to confirm that cation exchange was unsuccessful in separating the contaminant from target SIKE protein, as the ~29kDa protein is visible in both fractions labeled 1 and 2.

Figure 11: Anion exchange performed on SIKE 72 S6A (Figure 10, fraction E1).
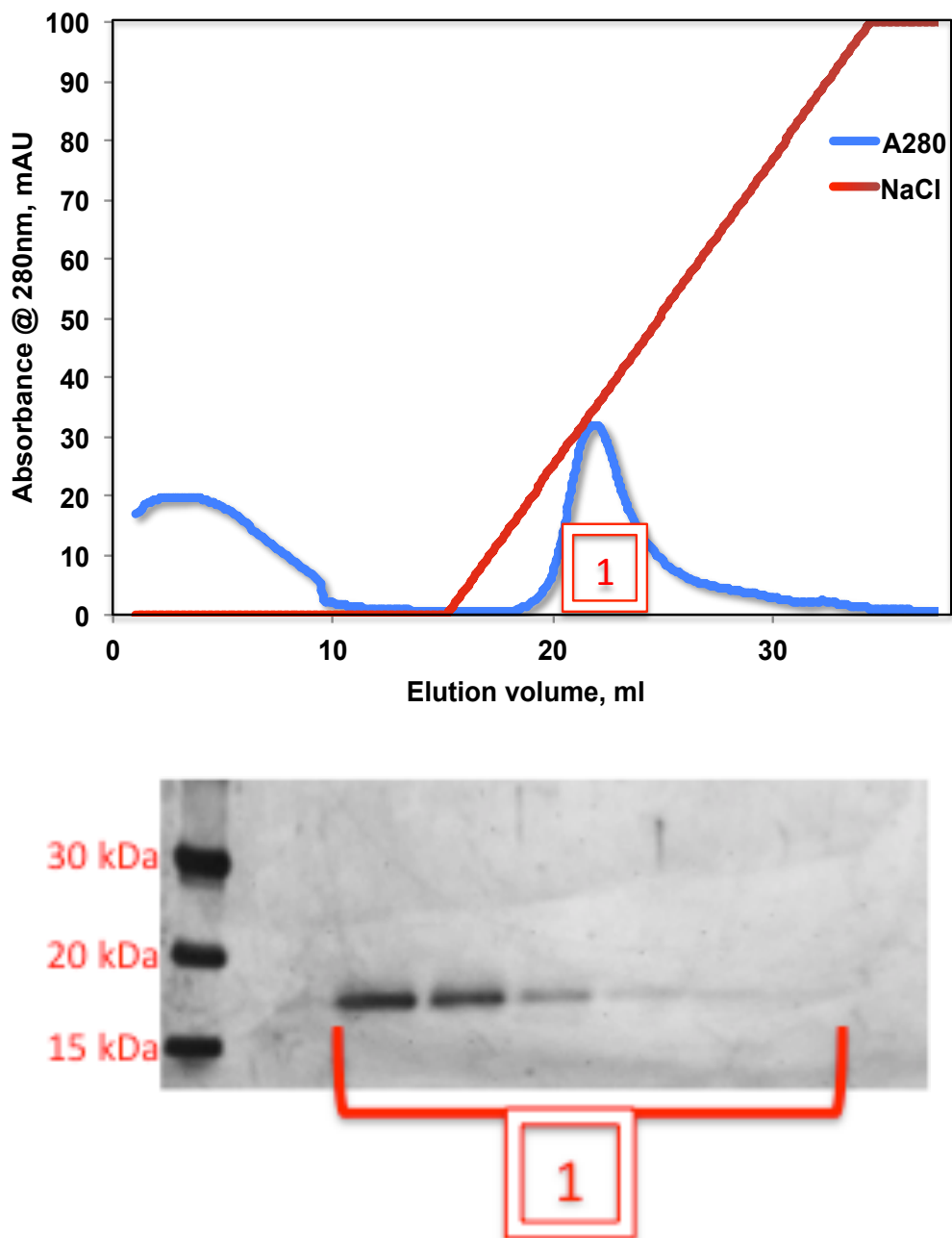
Figure 11: Anion exchange performed on SIKE 72 S6A (Figure 10, fraction E1).

Chromatogram shows separation of two species using a MonoQ XL column and a NaCl

gradient. Peak 1 (monomer/dimer SIKE species) proved to be pure using SDS-PAGE

and silver staining techniques.

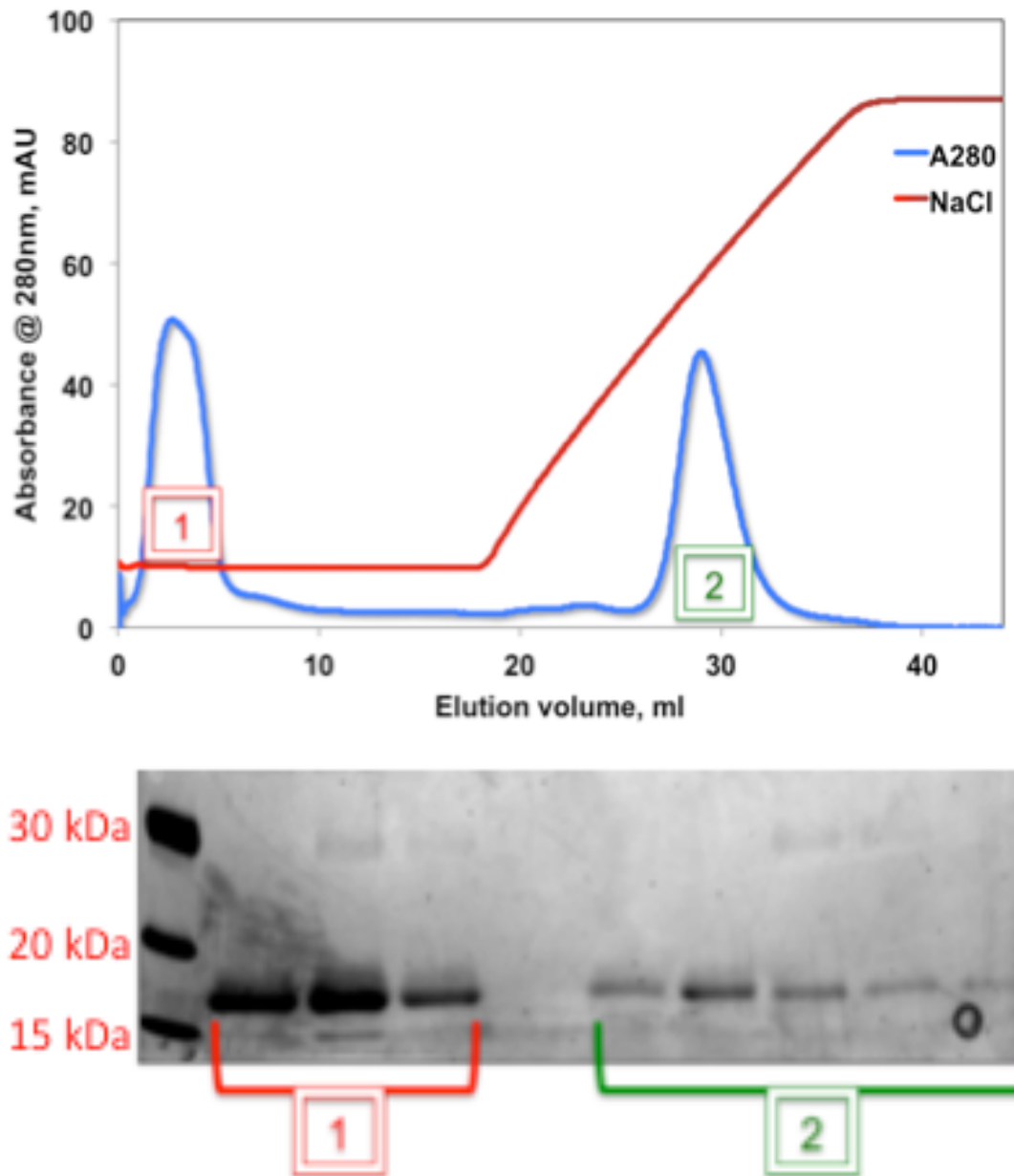**Figure 12: Cation exchange performed on SIKE 72 S6A (Figure 10, fraction E3).**

Figure 12: Cation exchange performed on SIKE 72 S6A (Figure 10, fraction E3).

Chromatogram shows separation of two groups using HiTrap SP XL column and a NaCl

gradient. Both peak 1 (aggregate protein) and peak 2 (monomer/dimer SIKE species)

proved to be contaminated using SDS-PAGE and silver staining techniques.

3.5 TALON® resin successfully binds SIKE and not contaminant

NiNTA elution fraction E2 (from figure 10) was used for IMAC with TALON® resin. SDS-PAGE shows positive binding of both (re-folded) SIKE and SlyD proteins (Figure 13). A significant amount of the SIKE did not bind the column and was found in the flow-through sample (FT), but was also visible in all six elution fractions. Additionally, contamination was only observed in the first elution fraction (E1). SEC analysis of the fractions indicated that the FT was primarily SIKE aggregate protein (data not shown), but elution fractions 2-6 were pure samples of SIKE monomer/dimer species (Figure 14). These were combined and concentrated for crystallization trials.

SDS-PAGE from IMAC shows positive binding of (denatured) SIKE protein, but no binding of (denatured) SlyD with TALON® resin (Figure 15). SlyD was seen in the flow-through portions, indicating that it was present in the sample originally, but failed to interact with the cobalt ions on the resin. SEC and SDS-PAGE analysis of elution fractions confirmed the purity of all SIKE samples (Figure 16; E4 results shown). In addition to eliminating contamination, the proportion of SIKE monomer/dimer species had increased substantially compared to the trials with re-folded protein samples. This allowed us to focus on concentrating the protein from our peak of interest and continue more crystallization trials.

Figure 13: SDS-PAGE results from IMAC with refolded proteins and TALON®

resin (Figure 10, fraction E2).
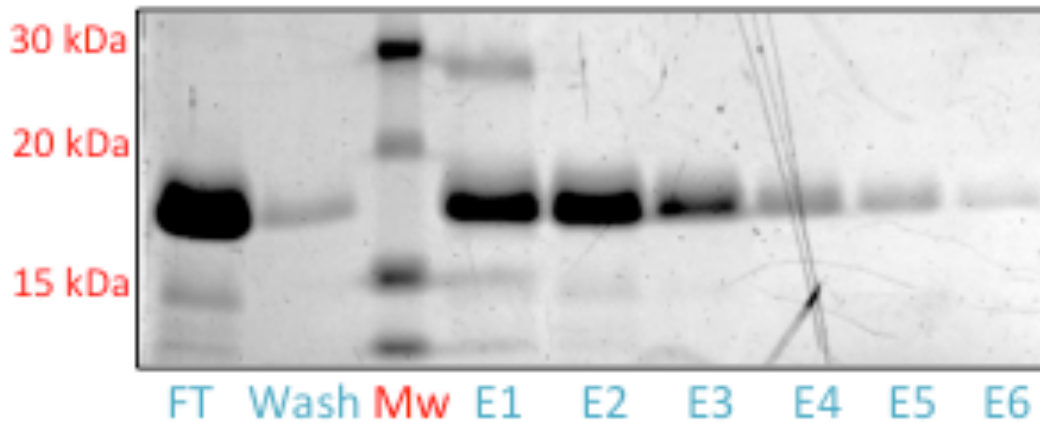
Figure 13: SDS-PAGE results from IMAC with refolded proteins and TALON®

resin (Figure 10, fraction E2). Resin bound both (re-folded) SIKE and SlyD proteins,

however, SlyD eluted mainly in the first elution fraction (E1).

Figure 14: SEC performed on purified SIKE sample (Figure 10, fraction E2).
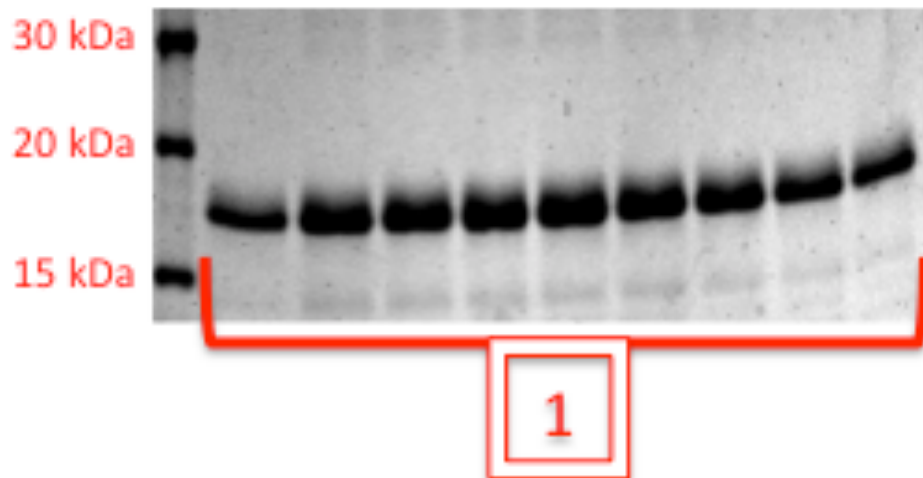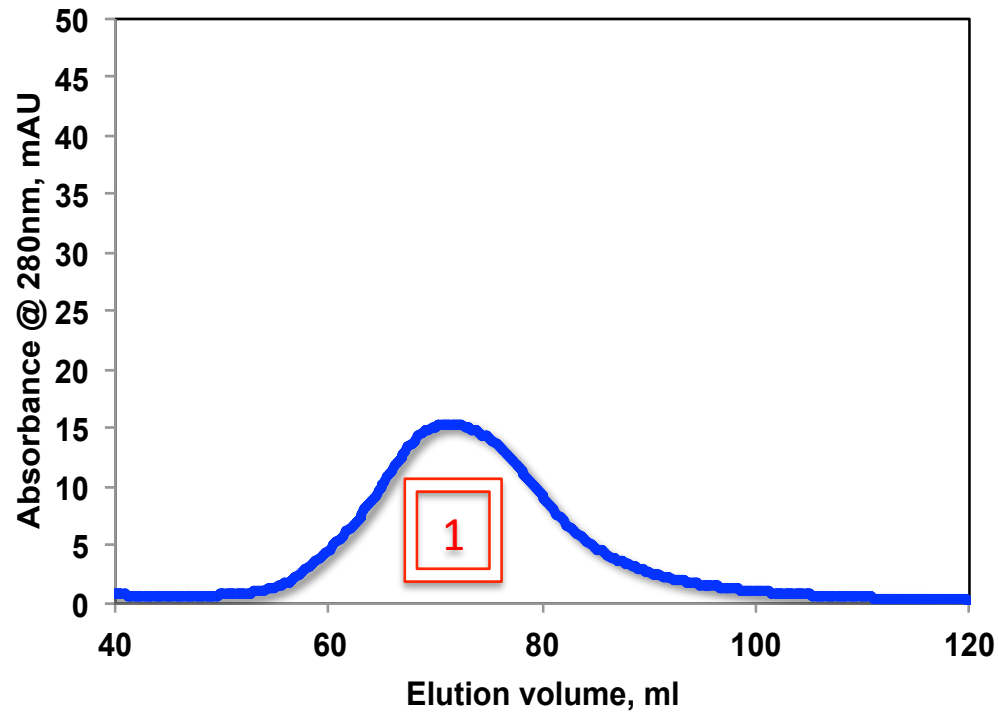
Figure 14: SEC performed on purified SIKE sample (Figure 10, fraction E2).

Chromatogram shows elution of SIKE oligomer species, labeled with a number 1. SDS-

PAGE/silver stain performed on Peak 1 fractions confirm that fractions E2-E6 were

purified SIKE 72 S6A protein

Figure 15: SDS-PAGE results from IMAC with denatured proteins and TALON®
resin.
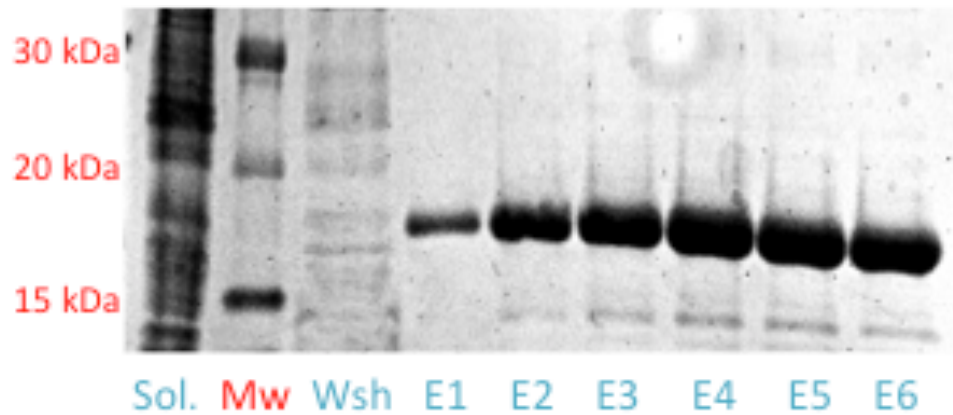
Figure 15: SDS-PAGE results from IMAC with denatured proteins and TALON®
resin. Resin bound denatured SIKE only; SlyD can be seen in the soluble supernatant
portion (Sol.) before IMAC, and flow-through portion (FT) that did not bind the resin.

Figure 16: SEC performed on purified SIKE sample (Figure 15, fraction E4).

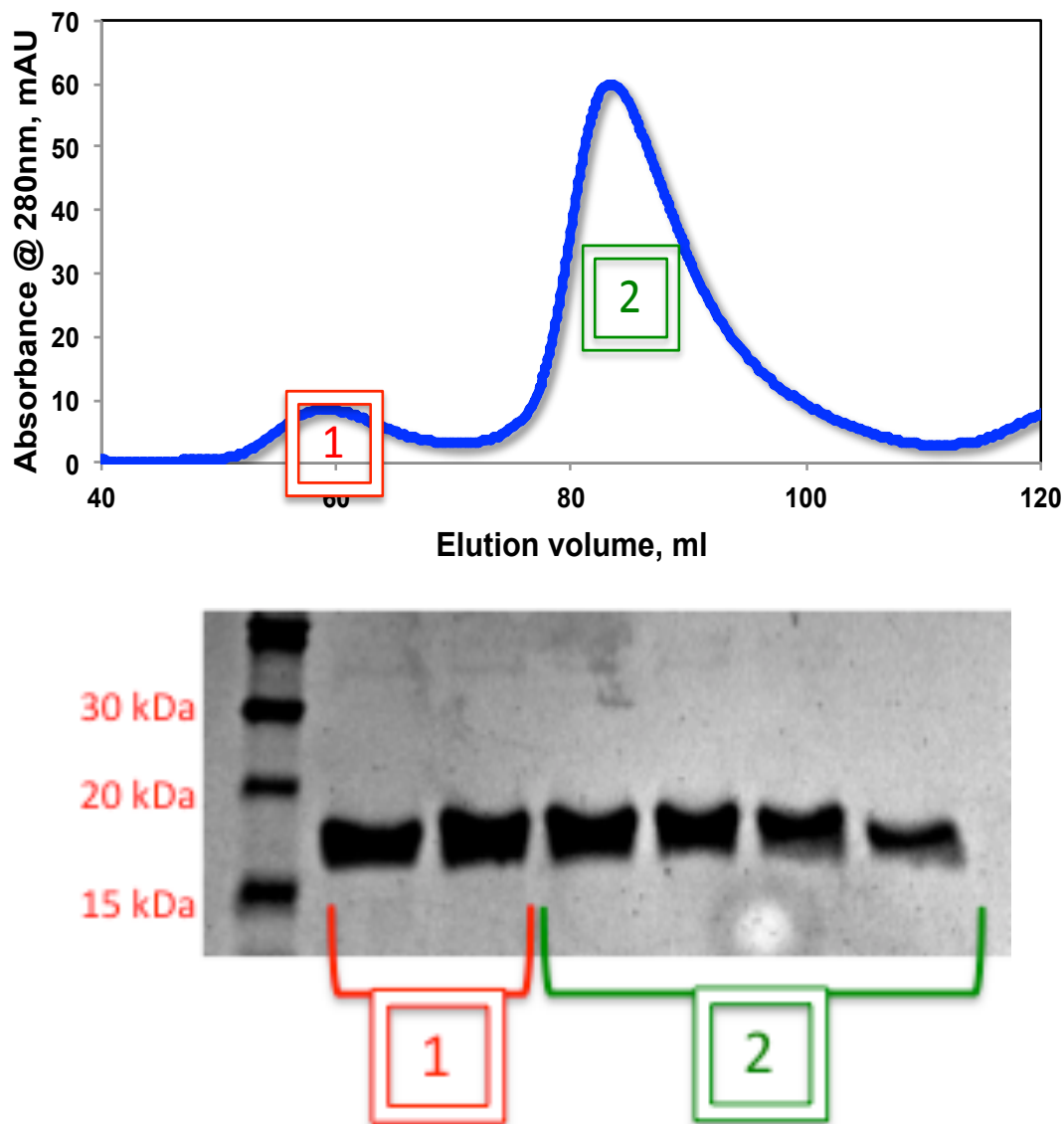**Figure 16: SEC performed on purified SIKE sample (Figure 15, fraction E4).**

Chromatogram shows Peak 1 (aggregates of SIKE protein) and Peak 2 (SIKE

monomer/dimer species). SDS-PAGE performed on fractions confirms the purity of our

samples.

3.6 Concentration of SIKE 72 S6A and crystallization trials

SEC fractions were pooled according to when they eluted during SEC in order to separate different SIKE species (monomer/dimer, aggregate, etc). Centrifugation concentration (with 10,000 Da molecular weight cut off filter) was used to remove buffer from the pooled fractions. Concentration was stopped immediately when any precipitate became visible in the sample, because further centrifugation would most likely precipitate more of the soluble protein. Through numerous expression and protein harvesting procedures, final concentrations ranged anywhere from less than 1mg/mL to over 20mg/mL, however, concentrations over 10mg/mL were rarely achieved for SIKE 72 S6A samples (data not shown). As soon as centrifugation was stopped, crystallization trays were set.

Prior to this current study, a sample of SIKE 72 S6A was sent to HWI to be screened using high-throughput crystallization methods and examined with 1,534 different conditions. Crystal formation was observed in various conditions, three of which were chosen for optimization and further crystallization trials: 337, 371, and 863 (Table 2.1). All three conditions contained sodium acetate salt (0.1M) dissolved in buffer (pH 5.0) and PEG (polyethylene glycol) 4,000 or PEG 8000 (40%w/v). These were optimized to find the most favorable condition for SIKE crystal formation by altering the pH and PEG concentrations, making 16 different conditions from each HWI condition (Table 2.2).

Vapor diffusion experiments were completed using sitting drop plates and various crystal screening kits to identify additional crystallization conditions. No single-crystal formation was observed in any of the sitting drop plates with SIKE 72 S6A concentrated

to 3.95mg/mL (data not shown). However, amorphous microcrystalline structures and needle-like structures were observed in numerous conditions for all five screens (data not shown).

Batch crystallization experiments were conducted to minimize any influence from the outside environment by covering the conditions with paraffin oil. Figure 17 shows examples of SIKE 72 S6A crystals that were obtained prior to identification of SlyD contamination (top images) and SIKE 72 S6A precipitate that was observed in most wells during the most recent protocol-modifying purification trials (bottom images). Original SIKE 72 S6A crystals were rod shaped and formed from areas of dense protein precipitation. For the majority of recent trials, protein precipitation was observed in almost every condition. The volume of precipitated protein observed in each well increased with higher concentrations of protein samples, yet X-ray diffraction quality crystals were rarely observed. Only two crystals were observed in batch crystallization trials during this study (Figure 18); however, they were not collected for any seeding or X ray diffraction experiments. Both crystals were seen using protein sample obtained from the protocol modified up to this point only.

To improve protein concentration, lyophilization was performed as a dehydration method to evaporate buffer from the sample, leaving only protein and any other solutes. Once completed, the remaining powder was brought up to ~15mg/mL in either 50mM phosphate buffer (for MALDI experiments) or gel filtration buffer (to set up crystallization plates). Not all of the protein was soluble, however this method greatly increased both the final concentrations and volumes of the pure protein samples. Final

purified SIKE 72 S6A samples ranged from 10-15mg/mL (data not shown) and were

used for further structural analysis.

Figure 17: Previous and current crystallization trials.

Previous SIKE 72 S6A crystals

337 A2   337 B2   371 C3   863 A1

Current SIKE 72 S6A precipitate
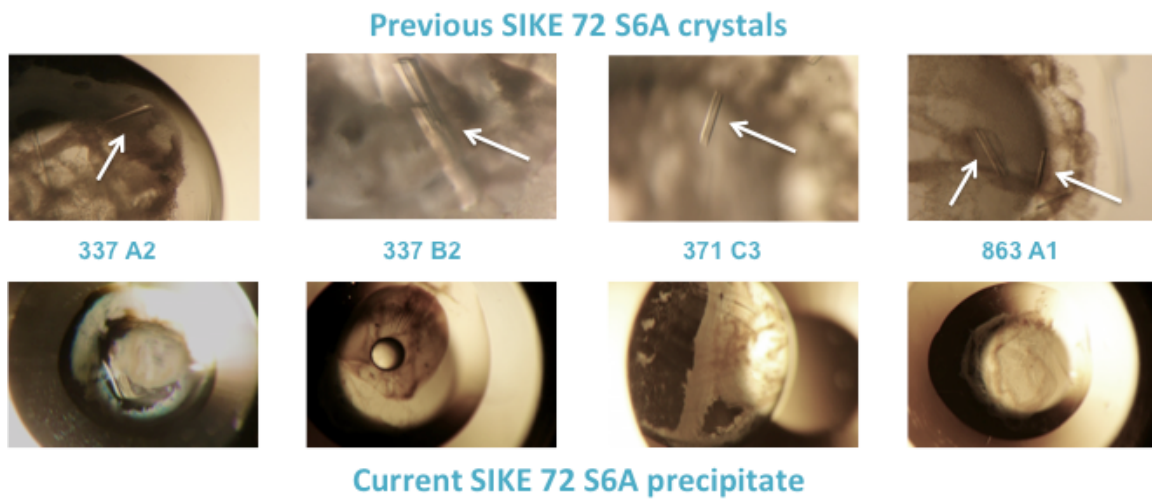
Figure 17: Previous and current crystallization trials. Top row, Crystals (indicated by white arrows) obtained prior to the identification of SlyD contamination. Bottom row, precipitate observed in current crystallization trials with the modified protocol.

Figure 18: Crystals from current SIKE 72 S6A trials.
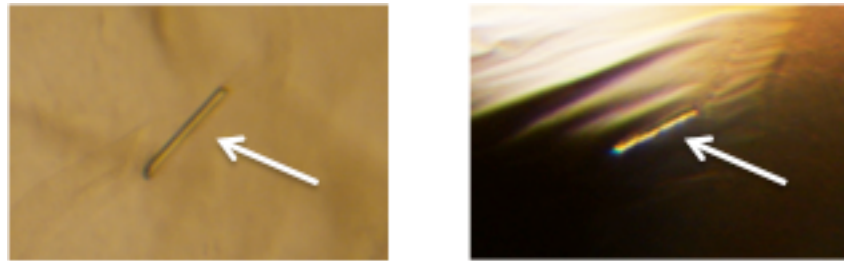
Current SIKE 72 S6A crystals

Figure 18: Crystals from current SIKE 72 S6A trials. Crystals (indicated by white arrows) obtained using the procedure with all protocol modifications up to this point except lyophilization.

3.7 Mass Spectrometry and Structural Analysis

In addition to crystallization trials, Hydrogen-Deuterium exchange was investigated, accompanied with pepsin digests, in order to further characterize SIKE structure. To confirm that this method could be useful, 0 and 24 hour HDX were performed for both SIKE FL S6A (Figure 19) and SIKE 72 S6A (Figure 20). After 24 hours, an increase in molecular weight was observed for both constructs, indicating that elucidation of structural information is possible through additional amide exchange, digest, and analysis experiments. The overall increase in molecular weight was observed to be 144 Daltons for the full-length construct (13,032 to 13,176 Daltons), and 293 Daltons for the truncated construct (17,886 to 18,179 Daltons).

To uncover structural information, fast, medium, and slow exchange of different amide protons in the protein's 3-demensional structure was investigated. Figure 21 shows the increase in molecular weight over 8 hours of HDX with SIKE 72 S6A. A large portion (~60%) of the total exchange occurred within the first 30 minutes, however, additional deuterium exchange was observed between 1-4 hours and 5-6 hours. The data suggest that a stable tertiary structure exists, protecting some portion of hydrogen from exchangingreadily.

In order to identify the location of HDX, pepsin digests must be completed with both protein that has undergone exchange and protein that has not. Un-exchanged SIKE 72 S6A protein was digested with pepsin and sampled at time points between 0-2 hours and at 8 hours. Spectra from each time point up to 2 hours are shown in Figure 22 and display successful protein digestion; at 8 hours, the protein was fully digested (data not

shown). This data will be used to determine a time point for reproducible digestion that
allows for identification of the largest percent of SIKE peptides.

Figure 19: Hydrogen-deuterium exchange for SIKE FL S6A (0 and 24hr).
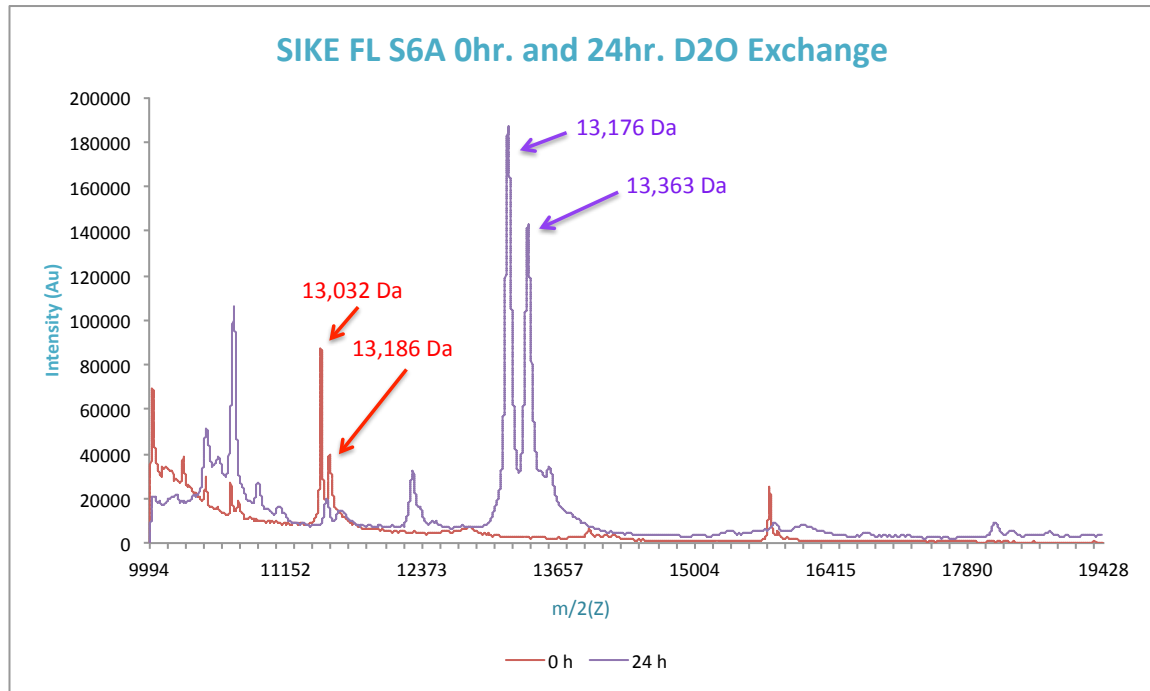
Figure 19: Hydrogen-deuterium exchange for SIKE FL S6A (0 and 24hr). Change in molecular weight of SIKE FL S6A due to hydrogen-deuterium exchange was observed using MALDI from 0 hours (red line) to 24 hours (purple line).

Figure 20: Hydrogen-deuterium exchange for SIKE 72 S6A (0 and 24hr).
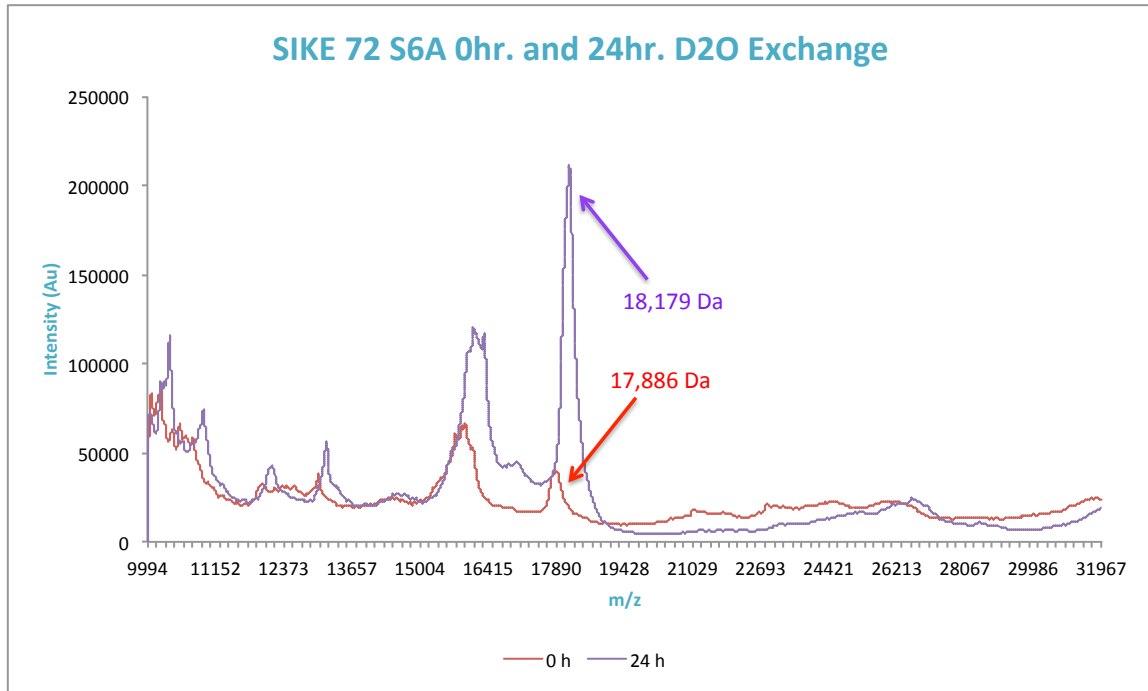
Figure 20: Hydrogen-deuterium exchange for SIKE 72 S6A (0 and 24hr). Change in molecular weight of SIKE 72 S6A due to hydrogen-deuterium exchange was observed using MALDI from 0 hours (red line) to 24 hours (purple line).

Figure 21: Deuterium exchange 8-hour time course for SIKE 72 S6A.
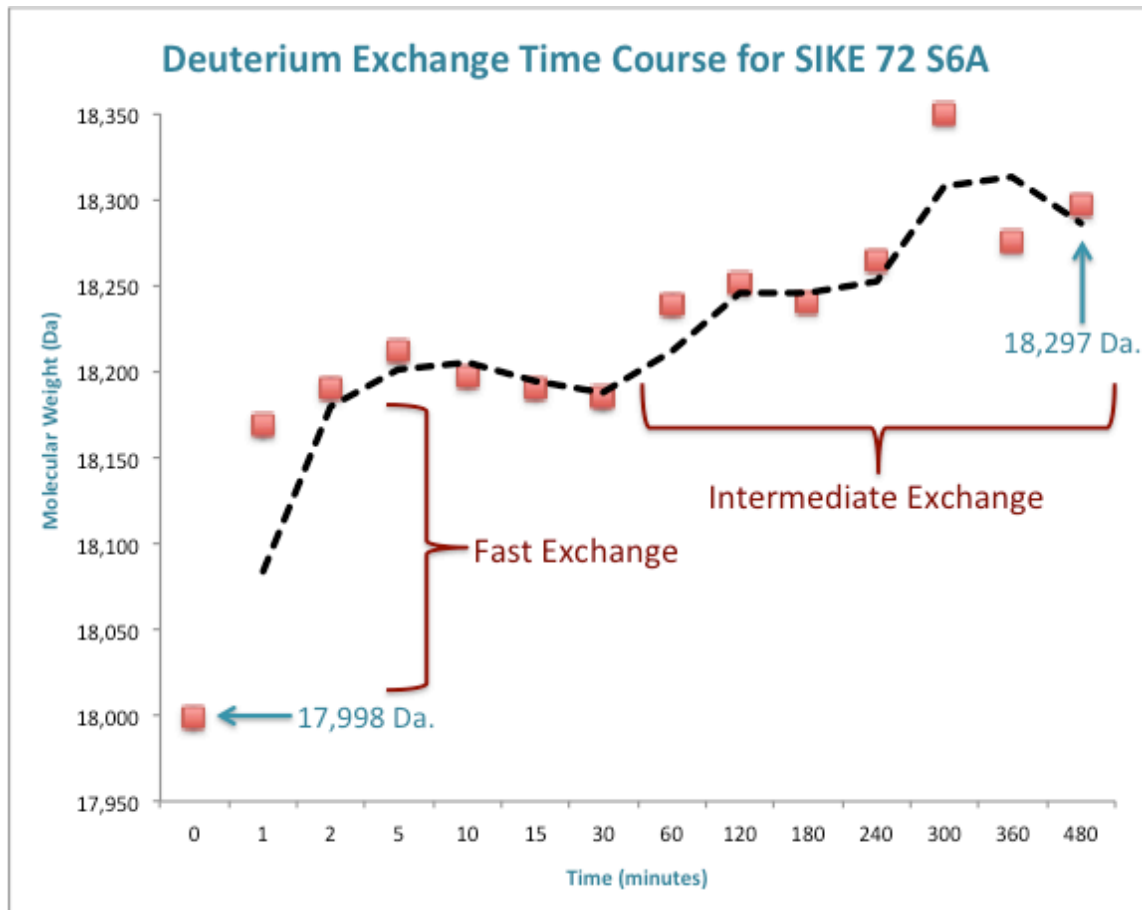
Figure 21: Deuterium exchange 8-hour time course for SIKE 72 S6A. Change in molecular weight of SIKE 72 S6A due to hydrogen-deuterium exchange was measured over an 8-hour time course.

Figure 22: Pepsin digest of SIKE 72 S6A (0-2 hours)
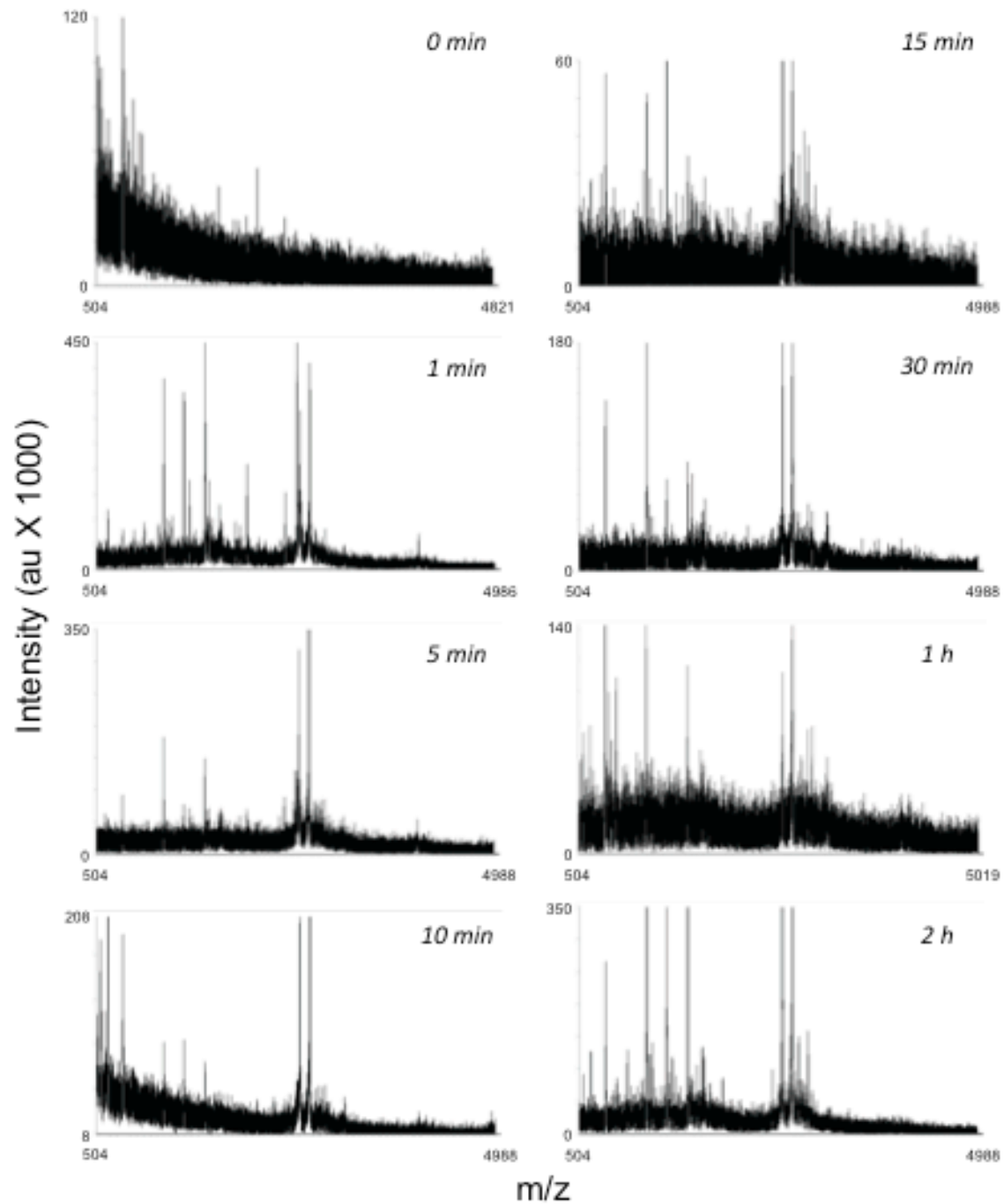
Figure 22: Pepsin digest of SIKE 72 S6A (0-2 hours). SIKE 72 was digested with

pepsin for the indicated time periods; peaks signify peptide fragments.

Chapter 4: Discussion

TLR-3-mediated signaling (through TBK1) initiates the innate immune response by recruiting immune cells to the site of infection, inhibiting pathogen replication, inducing apoptosis of infected cells, and communicating danger signals to other nearby cells [1]. Because of its importance in innate immunity, many studies concentrate on the pathway involving IRF3 activation, the transcription factor that functions in production of type 1 interferons. SIKE, a recently identified protein, is thought to play a role in this pathway and has been the focus of study.

SIKE is an evolutionarily conserved, 207-amino acid protein that is part of an uncharacterized family of proteins like fibroblast growth factor receptor 1 oncogene partner 2 (FGFR10P2)[2,18,20]. It was shown that SIKE associates with TBK1 under normal physiological conditions, but dissociates upon viral infection and TLR3 stimulation[18,19]. SIKE did not disrupt any interactions required by proteins for TLR3-mediated NF-κB activation, suggesting that SIKE specifically inhibited the IKKε- and TBK1-mediated IFN-β activation pathways[19]. In addition, previous fluorescent microscopy studies show SIKE (fused to GFP) accumulating in actin/cytoskeleton-like structures at the cell periphery[18]. It is hypothesized that SIKE, in response to pathogen challenge, functions in cytoskeletal rearrangement through phosphorylation by TBK1.

Seeking to characterize the competition between TBK1 substrates, IRF3 and SIKE, the Bell Laboratory previously determined that SIKE functions as a mixed type inhibitor with respect to IRF3 phosphorylation[18]. Study of its protein composition indicated that SIKE could be phosphorylated at six different sites, four of which resemble phosphorylation sites on IRF3[25-27]. Using these findings, the Bell Lab proposed that SIKE functions, not only as an inhibitor of TBK1-mediated phosphorylation of IRF3, but also as a TBK1 substrate (Figure 5). It is believed that SIKE, acting as a substrate, inhibits the ability of TBK1 to phosphorylate IRF3, thus inhibiting type 1 interferon production[18]. The function of phosphorylated SIKE once it has been released by the kinase remains unknown and is the basis of this research. As such, understanding the structure of SIKE through crystallographic and mass spectrometric methods may provide insight into its function in the TLR-3 pathway mediated by IRF3 activation.

In the current study, SIKE FL S6A and SIKE 72 S6A (~18 kDA) were successfully expressed using the original protein preparation protocol, however contamination by an unknown protein was identified in our samples (Figure 6b). Due to the fact that this contaminant ran at the same molecular weight in SDS-PAGE as the full-length construct (both ~29kDa, Figure 6), we chose to continue most of the study using SIKE 72 S6A to be sure that we collected pure samples only. The unknown protein was found to contaminate E1-E3 from SEC trials, leaving only three pure SIKE fractions (E4-E6) for crystallization trials and structural analysis. Protein that eluted first off of the SEC column (peaks labeled "1") represents aggregate SIKE 72 S6A protein, while protein that eluted second (peaks labeled "2") was determined to be monomers/dimers of SIKE 72 S6A. Supporting these results, SIKE's peptide sequence and predicted structure

suggest that it can interact with itself to form multimers such as dimers or tetramers. The contaminated fractions were used in various experiments with the goal of developing a more efficient protocol for the purification of SIKE 72 S6A.

To amplify yield for inducible protein expression in *E. coli*, auto-induction media was tested. Different metabolites in this media function to promote cell growth to a high density and automatically stimulate protein expression from IPTG-inducible T7 *lac* promoters[28,29]. This serves as an advantage by eliminating the need to monitor the growth of cell cultures after inoculation. Glucose in the media prevents uptake of lactose until the glucose is depleted, which is typically mid to late log phase. At that point, the lactose is taken up by the cells and converted into allolactose by β-galactosidase. Allolactose binds the *lac* repressor and causes it to release from the DNA. The T7 RNA polymerase can then bind and begin to induce genes controlled by the *lac* promoter. Such timely induction results in an increase in protein expression and overall yield. Once the concentration of lactose becomes low, glycerol in the media serves as the late energy source and the *lac* promoter is no longer induced[29]. In our experiments, auto-induction media successfully increased the yield of protein by almost 3 fold (Figure 8). Although the media efficiently increased protein expression, contamination was still observed in SEC elution fractions E1-E3.

One obstacle with the over-expression of recombinant protein in bacterial cells is formation of cytoplasmic aggregates in inclusion bodies[30]. At higher temperatures, aggregation of proteins is driven by hydrophobic interactions[23]. Growth at 16°C is known to help reduce aggregation during folding by increasing the activity of cold shock protein that function as chaperones to aid in protein folding[23,30]. One disadvantage is that

a sudden decrease in temperature slows replication, transcription, and translation. Ultimately, this method proved unsuccessful in increasing the amount of soluble protein and decreasing the amount of aggregate protein (data not shown). For the remainder of these studies, *E. coli* was grown at 37°C as noted in the original protocol.

Proteins being over-expressed are often stored inside intracellular structures called inclusion bodies[24]. These inclusion bodies generally contain very little host protein, suggesting that harvesting SIKE protein directly from inclusion bodies could minimize contamination. One method used to isolate inclusion bodies is to lyse the plasma membrane through a series of flash freezing and thawing[31]. Fragmentation of the membrane allows inclusion bodies to be easily harvested through centrifugation, separating them from any soluble contents. One possible disadvantage is that the procedure could result in protein degradation or precipitation[32]. In this study, freeze-thaw lysis proved ineffective in reducing the amount of contaminant in SIKE 72 S6A samples (data not shown). Furthermore, overall SIKE 72 S6A protein yield was less than that of the original protocol.

Another method recommended for harvesting protein from cell cultures, sonication, uses a metal probe that generates ultrasonic waves to shear the plasma membrane and release inclusion bodies[24,33]. The inclusion bodies can then be harvested through centrifugation and solubilized to release the enclosed protein. One disadvantage of sonication is excess heat generated from the high-energy waves can damage proteins, lowering the final protein yield. Solubility test results showed most of the SIKE 72 S6A in the insoluble pellet sample, while the contaminant separated to the soluble supernatant (Figure 9), suggesting that this method may be useful in isolating the protein from the

contaminant. Although fractions E1-E3 still contained the contamination, the overall level of contaminant was reduced in each of the three fractions (Figure 10). For the remainder of these studies, sonication was used to remove as much soluble contaminant as possible before harvesting inclusion bodies (insoluble portion) for chemical lysis.

Ion exchange chromatography is based on the reversible interaction between a charged protein and the oppositely charged chromatography medium used (http://www.gelifesciences.com). Small differences in the surface charges of two proteins causes them to elute at different points during the process. The surrounding buffer pH largely influences a protein's surface charge, as well as possible interactions that the protein can make. For anion exchange, the buffer pH should be at least one unit higher than the pI of the protein of interest to allow it to bind the positively charged solid support. For cation exchange, the buffer pH should be at least one unit below the pI of the protein of interest so it can bind the negatively charged solid support. Based on the amino acid sequence, the estimated pI for SIKE 72 S6A is 6.2, so the anion exchange buffer had a pH of 9.0 and the cation exchange buffer had a pH of 4.5. In this study, elution of proteins used an increasing gradient of NaCl to compete with the proteins binding to the column. As the NaCl concentration was gradually increased from 50mM to 1M, different elution patterns for SIKE and the contaminant were hypothesized.

Although positive separation of the two species was observed for the first anion exchange trial (Figure 11), all subsequent trials proved to be unsuccessful (data not shown). Additionally, all cation exchange trials were unsuccessful in separating the two proteins based on charge (Figure 12). Together, these results demonstrate that SIKE 72

S6A cannot be successfully separated from the contaminant by either anion or cation affinity chromatography.

Further research on the ~27kDa contaminant helped us identify it as a genome-encoded *E. coli* protein, named SlyD[34]. SlyD is a small (22,853 Daltons), cytosolic protein that has 14 histidine residues clustered at the C-terminal end of the peptide sequence (Figure 23). This amino acid composition results in contamination of purification procedures that use Ni-NTA resin to bind his-tagged proteins.

Analysis of Ni-NTA elution fractions by SDS-PAGE showed positive binding (E1-E3) of both SIKE and SlyD proteins to the resin (Figure 6). Clontech Laboratories, Inc. advertises a resin, TALON®, which uses Co^{2+} metal ions instead of Ni^{2+} ions to bind histidine residues. Contaminated Ni-NTA fractions were used to assess the efficiency of the resin to bind (already refolded) SIKE 72 S6A. A significant amount of SIKE did not bind the resin and was found in the flow-through, however, SlyD contamination was limited to the first elution fraction (Figure 13). SEC analysis of the FT indicated that it was primarily SIKE aggregate protein (data not shown), but elution fractions 2-6 from these trials were determined to be pure SIKE monomer/dimer species (Figure 14). These were combined and concentrated for crystallization trials. Together, these results suggest that the SlyD residues are still interacting with the ions on the resin.

The amino acid sequence of SlyD (Figure 23) shows the histidine residues spread throughout the C-terminus, suggesting that separation using TALON® resin may be enhanced if the proteins were denatured and refolded on this resin, as opposed to using previously-refolded protein. SDS-PAGE from IMAC shows positive binding of denatured SIKE protein, but not denatured SlyD with TALON® resin (Figure 15). SlyD

was seen in the flow-through portions, indicating that it was present in the sample originally, but failed to interact with the cobalt ions on the resin. In addition to eliminating contamination, SEC showed the proportion of SIKE monomer/dimer species had increased substantially compared to the aggregate portion (Figure 16). This allowed us to focus on concentrating the protein from our peak of interest and continue crystallization trials.

Proteins that eluted at different times during SEC were pooled separately in order to separate SIKE species (monomer, dimer, etc). Following coentrifugation, final sample concentrations ranged from less than 1mg/mL to over 20mg/mL, however, SIKE 72 S6A samples over 10mg/mL were rarely attained (data not shown). When centrifugation was stopped, crystallization trays were set.

In supersaturated solutions, small protein particles (seeds) stimulate separation of the solute from the solvent; this is called precipitation. This process is influenced by the contents, concentration, and pH of a solution, as well as environmental conditions. The point when the concentration becomes ideal for crystal formation is known as the "nucleation point." Structure and charge of a protein determine the possible ways in which they can interact when they precipitate out of a solution[35]. Crystal formation requires the diffusion of protein from an aqueous liquid to a pure-solid crystalline phase. It is known that specific 3-dimensional repeating protein patterns (monomers/ multimers, complexes, subunits, etc) form crystals if they precipitate in well-defined size and shape[36]. Precipitation that occurs with a less-ordered arrangement of protein species produces amorphous structures. To help draw the protein out of solution, precipitants can be used in the crystallization experiments.

Prior to this current study, a sample of SIKE 72 S6A was sent to HWI to be screened using high-throughput crystallization methods and examined with 1,534 different conditions. Crystal formation was observed in various conditions, three of which were chosen for optimization and further crystallization trials: 337, 371, and 863 (Table 1). All three conditions contained sodium acetate salt (0.1M) dissolved in buffer (pH 5.0) and PEG 4,000 or PEG8000 (40%w/v). These were optimized to find the most favorable condition for SIKE crystal formation by altering the pH and PEG concentrations, making 16 different conditions from each HWI condition (Table 2.2).

Vapor diffusion experiments were completed using sitting drop plates and various crystal screening kits to identify additional crystallization conditions. Sitting drop plates use a reservoir well for the crystallization condition and a small well for the protein sample that is covered by the same condition in the reservoir well. This method is designed to allow diffusion of water vapor out of the sample, increasing the concentration of the protein (as well as the precipitant) until equilibration is achieved, ideally around the nucleation point[35]. No single-crystal formation was observed in any of the sitting drop plates with SIKE 72 S6A concentrated to 3.95mg/mL (data not shown). However, amorphous microcrystalline structures and needle-like structures were observed in numerous conditions for all five screens. Analysis of those conditions revealed certain common denominators: TRIS or HEPES as the buffer, PEG 4000 or 8000 (25-30% w/v) as the precipitant, and ammonium sulfate or sodium acetate as the salt. This was similar to the results from HWI, suggesting that these might be favorable conditions for obtaining SIKE 72 S6A crystals, but protein concentration may be too low.

Batch crystallization experiments were conducted to minimize any influence from the outside environment by covering the conditions with paraffin oil. Vapor diffusion/equilibration is minimized; therefore this method requires protein to be introduced directly to conditions around its nucleation point. Beginning 3 days after the trays were completed (to allow time for any crystal formation), plates were observed daily. Although crystals were obtained from crystallization trials before the identification of SlyD contamination (Figure 17, top row), current batch crystallization trials have mostly been unsuccessful in producing x-ray diffraction quality crystals with SIKE 72 S6A samples from the modified protocol (Figure 17, bottom row). Original SIKE 72 S6A crystals were rod shaped and formed from areas of dense protein precipitation. For the majority of recent trials, protein precipitation was observed in almost every single condition. The volume of precipitated protein observed in each well increased with higher concentrations of protein samples, yet no X-ray diffraction quality crystals were observed. This suggested that pure protein samples still might not be concentrated enough to allow any crystal formation.

To improve protein concentration, lyophilization was performed as a dehydration method to evaporate water from the sample, leaving only protein and any other solutes. Once completed, the remaining powder was brought up to ~15mg/mL in either 50mM phosphate buffer (for MALDI experiments) or gel filtration buffer (to set up crystallization plates). Not all of the protein was soluble, however this method greatly increased both the final concentrations and volumes of the pure protein samples. Final purified SIKE 72 S6A samples ranged from 10-15mg/mL (data not shown) and were used for more crystallization trials and additional structural analysis.

In addition to crystallization trials, Hydrogen-Deuterium exchange was investigated, accompanied with pepsin digests, in order to further characterize the location of HDX using Mass Spectrometry. Native proteins are typically found in tightly folded conformations; understanding these structures is essential to infer a protein's function. One method used to understand protein conformation is to use hydrogen-deuterium exchange followed by pepsin digestion and peptide analysis by mass spectrometry. Deuterium (D_2) is an isotope of hydrogen that freely exchanges with hydrogen atoms on proteins. The extent to which peptides in the protein are shielded from the solvent, and whether they participate in intra-molecular hydrogen bonding determine the rate at which the amide protons undergo exchange[37]. By identifying where HDX is possible in the SIKE sequence, we identify solvent-exposed or flexible structural regions.

To confirm that this method could be useful, 0 and 24 hour exchanges were performed for both SIKE FL S6A (Figure 19) and SIKE 72 S6A (Figure 20). After 24 hours, an increase in molecular weight was observed for both constructs, indicating that elucidation of structural information is possible through additional peptide exchange, digest, and analysis experiments. The overall increase in molecular weight was observed to be 144 Daltons for the full-length construct (13,032 to 13,176 Daltons), and 293 Daltons for the truncated construct (17,886 to 18,179 Daltons).

To uncover structural information, time points from 0 hours to 8 hours were used in deuterium exchange to determine when fast, medium, and slow exchange occurs for SIKE 72 S6A. Protons exposed to the solvent generally exchange with the environment faster than ones buried inside the structure. Identifying amino acids/ peptides located on

the surface requires uncovering which residues underwent fast exchange. Over the first 30 minutes, deuterium atoms had already replaced ~164/293 (56%) of the protons in SIKE 72S6A available for exchange (using 24hr. as "full exchange;" Figure 20). Analysis of SIKE 72 S6A peptide sequence reveals that ~179 amide protons are available for fast exchange, 17 of which are located on arginine/lysine side chains. It is known that amide protons of arginine/lysine are not as easily exchanged, leaving ~162 protons that (theoretically) can be readily exchanged. Together, these results support the hypothesis that different amino acids/peptides are exchanging hydrogen atoms at different rates. Furthermore, exchange for SIKE 72 S6A displayed a step-wise pattern, consistent with known (folded) proteins that undergo deuterium incorporation (Figure 21).

In addition to exchange, pepsin digests were performed to identify the ideal time for SIKE digestion to yield adequately sized peptides for mass spectrometry analysis. An ideal time is one that allows for the highest percent of peptide recovery for analysis, as well as one that is reproducible. Ultimately, the proteins that have undergone fast, medium, or slow exchange will be digested for the same time to yield the same set of peptides for each sample. Mass spectrometry analysis on the peptides allows comparison of exchanged and un-exchanged samples to identify which peptides exchanged more and when that exchange occurred. All together, this information will help us locate the peptides within the 3-dimensional protein structure that underwent exchange.

Pepsin digests of SIKE 72 S6A were performed for at time points between 0 and 2 hours, as well as one at 8 hours. Mass spectrometry analysis showed successful digestion of protein, revealing fragments of various sizes between 500 and 5,000 Daltons (Figure 22). A database search on the Protein Prospector website with the SIKE 72 S6A

sequence revealed every potential peptide fragment that could be created from pepsin digest of SIKE 72 (data not shown). Using this information, the ideal time for protein digestion will be determined that yields an appropriate amount of fragments sufficient for mass spectrometry analysis. With the ideal time point determined, SIKE 72 S6A protein that has already undergone exchange will be digested and analyzed with mass spectrometry. Molecular weight changes between exchanged and un-exchanged peptide fragments will indicate the amount of deuterium that has been exchanged on that specific peptide. Determining which peptides exchanged deuterium helps pinpoint the fragment within the overall 3-dimensional structure of the protein that is accessible for exchange. Peptides that undergo the fastest exchange are (most likely) positioned such that amide hydrogen atoms are more exposed to the exchange solvent. The peptides observed to exchange the least amount are (most likely) buried within the protein, making it difficult to exchange hydrogen atoms with deuterium atoms in the solvent. Currently, only digest of un-exchanged SIKE 72 S6A have been completed; subsequent digest for exchanged protein and peptide fragment analysis will be needed to infer any information about the proteins surface structure.

In this study, the purification of SIKE 72 S6A revealed an *E. coli* contaminant, later identified as SlyD. The purification scheme was modified to remove contamination and increase the quality of protein samples for crystallization trials. Auto-induction media was effective in amplifying protein yield by over 3 fold. Harvesting inclusion bodies by sonication (not freeze-thaw lysis) decreased the overall amount of SlyD in IMAC fractions, however, contamination was still observed in the same samples. Cation and anion affinity chromatography were both ineffective in separating the two proteins.

TALON® resin bound both denatured SIKE 72 S6A and SlyD, however, only SIKE 72 S6A bound the resin under denaturing conditions, allowing successful separation from the contaminant. Concentration issues were answered through the use of lyophilization techniques, in place of ultrafiltration. Crystallization trials for samples from each purification scheme were completed, but no X-ray diffraction quality crystals were observed. In addition to crystallization trials, Hydrogen-Deuterium exchange accompanied with pepsin digests, in order to further characterize the SIKE structure where initiated. Additional exchange and digest trials must be completed to conclude any structural information regarding the surface area of SIKE 72 S6A.

While these studies investigate the properties of recombinant SIKE using various methods, the findings as a whole establish a previously undefined protocol for SIKE purification that will further future structural studies.

Figure 23: Amino acid peptide sequence for *E*. coli protein, SlyD.

<div style="border:1px solid black">

SlyD peptide sequence

MKVAKDLVVSLAYQVRTEDGVLVDESPVSAPLDYLHGHGSLISGLETALEGHEVGDKFDVAVGANDAY
GQYDENLVQRVPKDVFMGVDELQVGMRFLAETDQGPVPVEITAVEDDHVVVDGNHMLAGQNLKFN
VEVVAIREATEEELA**H**G**H**V**H**GA**H**D**HHH**D**H**D**H**DGCCGG**H**G**H**D**H**G**H**E**H**GGEGCCGGKGNGGCGC**H**

</div>

Figure 23: Amino acid peptide sequence for *E*. coli protein, SlyD. Highlighted in red are 14 histidine residues located in the C-terminus of the protein sequence that cause it to contaminate purification procedures that use Ni^{2+} ions to bind his-tagged proteins.

Literature Cited

1. Janeway, C.J., Murphy, K., Travers, P., Walport, M. (2011) *Janeway's Immunobiology*. 8th edited ed. Garland Science Publishing

2. Marion, J.D (2013) The activation, receptior complexing and endogenous regulation of the type-1 interferon response as it pertains to innate immunity. Ph.D. dissertation. Virginia Commonwealth University, Richmond,Va.

3. Halliday, S. (2001) Death and miasma in Victorian London: an obstinate belief. *British Medical Journal*. 323/7327: 1469-1471

4. Smith KA. (2012) Louis Pasteur, the father of immunology? *Front Immunol.*

5. Bordenave G. (2003) Louis Pasteur (1822-1895). *Microbes Infect* 5:553-60

6. Robert Koch – Biographical. (1967) *Nobel Lecturers, Physiology or Medicine 1901-1921*. Elsevier Publishing Company, Amsterdam.

7. Emil von Behring – Biographical. (1967) *Nobel Lecturers, Physiology or Medicine 1901-1921*. Elsevier Publishing Company, Amsterdam.

8. Ilya Mechnikov – Biographical. (1967) *Nobel Lecturers, Physiology or Medicine 1901-1921*. Elsevier Publishing Company, Amsterdam.

9. Bayne, C. J. (2003) Origins and evolutionary relationships between the innate and adaptive arms of immune systems. *Integr. Comp. Biol.* 43(2):293-299

10. Hansson, G. K. (2005) Toll to obe paid a

11. Akira, S. (2003) Toll-like receptor signaling. *J. Biol. Chem.* 278:38105-38108

12. Kawai, T., Akira A., (2011) Toll-like receptors and their crosstalk with other innate receptors in infection and immunity. *Immunity* 34:637-650

13. Christmas, P. (2010) Toll-like receptors: sensors that detect infection. *Nature Education* 3(9):85

14. Edelmann, K.H., Richardson-Burns, S., Alexopoulou, L., Tyler, K.L., Flavell, R.A., Oldstone, M.B. (2004) Does toll-like receptor 3 play a biological role in virus infections? *Virology* 322:231-238

15. Sen, G.C., Sarkar, S.N (2005) Transcriptional signaling by double-stranded RNA: role of TLR3. *Cytokine Growth Factor Rev.* 16(1):1-14

16. Vercammen, E., Staal, J., Beyaert, R. (2008) Sensing of viral infection and activation of innate immunity by toll-like receptor 3. *Clin. Microbiol. Rev.* 21:13-25

17. Beg, A. A., Baltimore, D. (1996) An essential role for NF-kappa B in preventing TNFa-induced cell death. *Science.* 1;274(5288):782-784

18. Marion, J.D., Roberts, C.F., Call, R.J., Forbes, J.L., Nelson, K.T., Bell, J.E., Bell, J.K. (2013) Mechanism of Endogenous Regulation of the Type I Interferon Response by Suppressor of IκB Kinase ε (SIKE), a Novel Substrate of TANK-binding Kinase 1 (TBK1). *J. Biol. Chem.* 288(25):18612-18623

19. Huang, J., Ting, L., Xu, L.G., Chen, D., Zhai, Z., Shu, H.B. (2005) SIKE is an IKKε/TBK1-associated suppressor of TLR3- and virus-triggered IRF-3 activation pathways. *EMBO J.* 24:4018-4028

20. Lin, A., Hokugo, A., Choi, J., Nishimura, I. (2010) Small cytoskeleton-associated molecule, fibroblast growth factor receptor 1 oncogene partner 2/ wound inducible transcript-3.0 (FGFR1OP2/wit3.0), facilitates fibroblast-driven wound closure. *Am. J. Pathol.* 176(1):108-121

21. Piao, W., Song, C., Chen, H., Quevedo Diaz, M.A., Whal, L.M., Fitzgerald, K.A., Li, L., Medvedev, A.E. (2009) Endotoxin tolerance dysregulates MyD88- and Toll/IL-1R

domain-containing adapter inducing IFN-β-dependent pathways and increases expression of negative regulators of TLR signaling. *J. Leukoc. Biol.* 86(4):863-875

22. Alberts, B., Johnson, A., Lewis, J., et al. (2002) Analyzing Protein Structure and Function. Molecular Biology of the Cell. 4th edition. New York: Garland Science; 2002.

23. Sorensen, H.P., Mortensen, K.K. (2005) Soluble expression of recombinant proteins in the cytoplasm of *Escherichia coli*. *Microb. Cell Fact*. 4:1

24. Peternel, S., Radovan, K. (2010) Isolation of biologically active nanomaterial (inclusion bodies) from bacterial cells. *Microbial Cell Fact.* 9:66

25. Takahashi, K., Suzuki, N.N., Horiuchi, M., Mori, M., Suhara, W., Okabe, Y., Fukuhara, Y., Terasawa, H., Akira, S., Fujita, T., Fuyuhiko, I. (2003) X-ray crystal structure of IRF3 and its functional implications. *Nat. Struct. Biol.* 10,922-927

26. Servant, M.J., Grandvaux, N., tenOever, B.R., Duguay, D., Lin, R., Hiscott, J. (2003) Identification of the minimal phosphoacceptor site required for *in vivo* activation of interferon regulatory factor 3 in response to virus and double-stranded RNA. *J. Biol. Chem.* 287, 9441-9447

27. Panne, D., McWhirter, S.M., Maniatis, T., Harrison, S.C. (2007) Interferon regulatory factor 3 is regulated by a dual phosphorylation-dependent switch. *J. Biol. Chem.* 282, 22816-22822

28. Studier, F.W. (2005) Protein production by auto-induction in high-density shaking cultures. *Protein Expr. Purif.* 41:207-234

29. Blommel, P.G., Becker, K.J., Duvnjak, P., Fox, B.G. (2007) Enhanced bacterial protein expression during auto-induction obtained by alteration of lac repressor dosage and medium composition. *Biotechnol. Prog.* 23(3):585-598

30. Schein, C.H. (1989) Production of soluble recombinant proteins in bacteria. *BioTechnology*. 7:1141-1148

31. Rodriguez-Carmona, E., Cano-Garrido, O., Seras-Franzoso, J., Villaverde, A., Garcia-Fruitos, E. (2010) Isolation of cell-free bacterial inclusion bodies. *Microbial Cell Fact.* 9:71

32. Eilertsen, H.C., Huseby, S., Degerlund, M., Erlken, G.K., Ingbrigtsen, R.A., Hansen, E. (2014) The effect of freeze/thaw cycles on reproducibility of metabolic profiling of marine microalgal extracts using direct infusion high resolution mass mectrometry (HR-MS). *Molecules.* 19(10):16373-16380

33. Feliu, J.X., Cubarsi, R., Villaverde, A. (1998) Optimized release of recombinant proteins by ultrasonication of *E. coli* cells. *Biotechnol. Bioeng.* 58(5):536-40

34. McMurry, J., Macnab, R.M. (2004) BD TALON resin does not bind *E. coli* SlyD, a common contaminant in Ni-NTA IMAC. *Clontechniques.* Jan. 2004:2-4

35. Dessau, M.A., Modis, Y. (2011) Protein crystallization for x-ray crystallography. *J. Vis. Exp.* 47:2285

36. Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N., Bourne, P. E., (2000) *Nucleic Acids Research.* 28:235-242

37. Zhang, Z., Smith, D.L. (1993) Determination of amide hydrogen exchange by mass spectrometry: A new tool for protein structure elucidation. *Protein Science.* 2:522-531

Vita

Sean McKinley was born in Richmond, Virginia, U.S.A. in 1990. Following graduation from Deep Run High School, he attended James Madison University in Harrisonburg, Virginia where he graduated with a Bachelors of Science in Biology in May 2012. He received his Master of Science degree in Microbiology and Immunology from Virginia Commonwealth University in December 2014. He plans to pursue a career in microbiology, immunology, or biomedical research.