

Biophysical studies of selected intrinsically disordered proteins:
insect stress-response peptides and cancer-linked variants of human p53 transactivation domain

By

Lynn Goss Schrag

B.S., University of Kansas, 2013

AN ABSTRACT OF A DISSERTATION

Submitted in partial fulfillment of the requirements for the degree

DOCTOR OF PHILOSOPHY

Department of Biochemistry and Molecular Biophysics
College of Arts and Sciences

Kansas State University
Manhattan, Kansas

2020

Abstract

The intrinsically disordered proteins (IDPs) or regions (IDRs), which do not have a single well-defined structure, participate in critical roles of signal transduction in Eukaryotic cells. Using humans as an example, more than 1/3rd of all protein sequences contain an identifiable >30 residue stretch that can be defined as intrinsically disordered. Interestingly, this intrinsic structural flexibility often facilitates binding promiscuity of IDP or IDR and renders them ideal for participating in complex cell signaling interaction networks by acting either as a hub, an integrator of cell signaling, or as a ligand, an inducer of one or more pathways through receptors. Though modern biophysical approaches are optimized for well-folded proteins and segments but the understanding of the structural dynamics as it relates to critical aspects of biological activities for most IDPs is still relatively poor. In this dissertation, we will explore two distinct IDP structure activity relationship and discuss biophysical studies on two systems: The results of our first study involve the novel cytokine-like stress-response peptides of the insect *Manduca sexta* which contains highly dynamic termini that are critical for its activity. The second study, the transactivation domain of critical eukaryotic transcription factor and tumor suppressor p53 that acts as a critical integrator of intracellular signaling in regulating cell-fate. In both cases, studies carried out with selective mutagenesis reveal small changes in the preferred dynamics of the IDPs for each system that correspond to pronounced differences in physiological outcomes.

Biophysical studies of selected intrinsically disordered proteins:
insect stress-response peptides and cancer-linked variants of human p53 transactivation domain

By

Lynn Goss Schrag

B.S., University of Kansas, 2013

A DISSERTATION

Submitted in partial fulfillment of the requirements for the degree

DOCTOR OF PHILOSOPHY

Department of Biochemistry and Molecular Biophysics
College of Arts and Sciences

Kansas State University
Manhattan, Kansas

2020

Approved by:
Co-Major Prof.
Michal Zolkiewski

Approved by:
Co-Major Prof.
Om Prakash

Copyright

© Lynn Goss Schrag 2020

Abstract

The intrinsically disordered proteins (IDPs) or regions (IDRs), which do not have a single well-defined structure, participate in critical roles of signal transduction in Eukaryotic cells. Using humans as an example, more than 1/3rd of all protein sequences contain an identifiable >30 residue stretch that can be defined as intrinsically disordered. Interestingly, this intrinsic structural flexibility often facilitates binding promiscuity of IDP or IDR and renders them ideal for participating in complex cell signaling interaction networks by acting either as a hub, an integrator of cell signaling, or as a ligand, an inducer of one or more pathways through receptors. Though modern biophysical approaches are optimized for well-folded proteins and segments but the understanding of the structural dynamics as it relates to critical aspects of biological activities for most IDPs is still relatively poor. In this dissertation, we will explore two distinct IDP structure activity relationship and discuss biophysical studies on two systems: The results of our first study involve the novel cytokine-like stress-response peptides of the insect *Manduca sexta* which contains highly dynamic termini that are critical for its activity. The second study, the transactivation domain of critical eukaryotic transcription factor and tumor suppressor p53 that acts as a critical integrator of intracellular signaling in regulating cell-fate. In both cases, studies carried out with selective mutagenesis reveal small changes in the preferred dynamics of the IDPs for each system that correspond to pronounced differences in physiological outcomes.

Table of Contents

List of Figures	x
List of Tables.....	xii
Acknowledgements.....	xiii
Dedication	xiv
Project Goals	xv
Chapter 1- Introduction: Structure and Function of Stress-response Peptides in Insects ¹	1
1.1. Introduction.....	2
1.2. Discovery.....	3
1.3. Structure	5
1.3.1. Core Domain	7
1.3.2. N-terminal Extension	10
1.3.3. C-terminal Extension	11
1.4. Biological Function	13
1.4.1. Plasmatocyte Spreading and Attachment	13
1.4.2. Paralysis.....	14
1.4.3. Wound Healing	14
1.4.4. Phagocytosis.....	15
1.4.5. Reactive Oxygen/Nitrogen Species and Nitric Oxide Synthase Production.....	16
1.4.6. Induced AMP Synthesis.....	16
1.4.7. Growth-Blocking Activity	17
1.4.8. Mitogenic Activity.....	18
1.5. Conclusion	18
1.6. References.....	20
Chapter 2- NMR Spectroscopic and Other Biophysical Methods Used in the Study of Stress-response Peptide-1 and Stress-response Peptide-2.....	28
2.1. NMR Spectroscopy	28
2.2. Structure Calculations	29
2.3. Explicit Solvent Molecular Dynamic (MD) Simulations	30
2.4. MD Analysis.....	31
2.5. References.....	33
Chapter 3- Solution Structure and Expression Profile of an Insect Cytokine: <i>Manduca sexta</i> Stress Response Peptide-2 ¹	36
Abstract	36
3.1.1. Introduction.....	37
3.1.2. Materials and Methods.....	39

3.1.2.1	<i>Identification of SRP Genes in the Genome of M. sexta</i>	39
3.1.2.2	<i>Insect Rearing and Injection, Total RNA Preparation, and cDNA</i>	39
3.1.2.3	<i>Quantitative Polymerase Chain Reaction (qPCR)</i>	40
3.1.2.4	<i>Chemical Synthesis of Putative M. sexta SRP2</i>	42
3.1.2.5	<i>NMR Spectroscopy and Other Biophysical Analyses of SRP2</i>	42
3.1.3	Results	43
3.1.3.1	<i>Identification and Features of Putative SRPs of M. sexta</i>	43
3.1.3.2	<i>Immune Inducibility, Tissue Specificity, and Developmental Profiles of SRP2 Expression</i>	44
3.1.3.3	<i>Induction of AMP Gene Expression by SRP2 Injection</i>	46
3.1.3.4	<i>NMR Resonance Assignments and Secondary Structure</i>	47
3.1.3.5	<i>Tertiary Structure of M. sexta SRP2</i>	50
3.1.4	Discussion	53
3.1.5	Conclusion	55
3.2	Unpublished Results of MD Simulation of SRP2	56
3.2.1	Introduction	56
3.2.2	Results and Discussion	57
3.2.2.1	<i>Generation of WT Simulation and NOE Back-calculation Analysis</i>	57
3.2.2.2	<i>Impact of in silico Mutation of the Embedded Residue upon SRP2 Conformational Heterogeneity</i>	60
3.2.2.3	<i>Differences in Structural Motility and Conformation of SRP2</i>	60
3.2.2.4	<i>N-terminal Orientation of SRP2 is Influenced by Salt-bridges and Core region Stability</i>	64
3.2.3	Conclusions	66
3.3	References	68
3.4	Supplementary Material	75
Chapter 4- Expression and Characterization of <i>Manduca sexta</i> Stress Response Peptide-1; An Inducer of Antimicrobial Peptide Synthesis ¹		77
	Abstract	77
4.1	Introduction	78
4.2	Materials and Methods	80
4.2.1	<i>Rearing of Insects and Injection, Total RNA Preparation, and cDNA Synthesis</i>	80
4.2.2	<i>Quantitative Real-Time Polymerase Chain Reaction (qRT-PCR)</i>	81
4.2.3	<i>Synthesis of M. sexta Stress Response Peptide SRP1</i>	81
4.2.4	<i>NMR Spectroscopy and Other Biophysical Analyses of SRP1</i>	82
4.3	Results and Discussion	83
4.3.1	<i>Immune Inducibility, Tissue Specificity, Stress Responsiveness, and Developmental Profiles of SRP1 Expression</i>	83

4.3.2	<i>Induction of AMP Gene Expression by SRP1 Injection</i>	85
4.3.3	<i>NMR Resonance Assignments and Secondary Structure</i>	87
4.3.4	<i>Tertiary Structure of M. sexta SRP1</i>	89
4.4.	Conclusion.....	94
4.5.	References.....	96
4.6.	Supplementary Material.....	102
Chapter 5-	Introduction: Structure-activity Relationships in Human Tumor Suppressor p53	105
5.1.	Introduction.....	105
5.2.	Discovery.....	107
5.3.	Structure and Known Biological Activities of p53 Domains	108
5.3.1.	<i>The Transactivation Domain (TAD) (M1-D61)</i>	110
5.3.2.	<i>Proline-rich Domain (PRD) (D61-S94)</i>	115
5.3.3.	<i>The DNA-binding Domain (DBD) (T102-K292)</i>	116
5.3.4.	<i>C-terminal (N-terminally Regulated (NRD); P301-D393) and Tetramerization Domains (D324-A355)</i>	120
5.4.	References.....	124
Ch 6-	Materials, Methods, and Approaches in Analysis of Missense Mutations within the Transactivation Domain of p53.....	146
6.1.	<i>Expression, mutagenesis and purification of p53-TAD (performed in collaboration with Indhujah Thevarajan)</i>	146
6.2.	<i>Expression and purification of TAZ2 domain of CBP</i>	148
6.3.	<i>Expression and purification of TAZ1 domain of CBP</i>	149
6.4.	<i>Expression and purification of HDM2 (performed by Indhujah Thevarajan)</i>	149
6.5.	<i>Heteronuclear NMR experiments of p53TAD variants</i>	150
6.6.	<i>Relaxation data acquisition and analysis of p53TAD variants</i>	151
6.7.	<i>Biolayer interferometry assays of p53TAD variants to HDM2, CBP-TAZ1, and CBP-TAZ2 (performed in collaboration with Indhujah Thevarajan)</i>	153
6.8.	References.....	155
Chapter 7-	Biophysical and Functional Analysis of Cancer-linked Mutations within the Intrinsically Disordered Transactivation Domain of p53	157
Abstract	157
7.1.	Introduction.....	158
7.2.	Results	163
7.2.1.	<i>NMR resonance assignments of p53TAD variants</i>	163
7.2.2.	<i>Relaxation rate measurements and reduced spectral density approaches for p53TAD variants</i>	165
7.2.3.	<i>Impact upon binding kinetics between p53TAD variants and CBP-TAZ2, CBP-TAZ1, and HDM2</i>	172

7.3. Discussion.....	175
7.4. Conclusions.....	177
7.5. References.....	179

List of Figures

Figure 1.1. The Proteolytic Activation of a Stress-response Peptide (SRP) from Its Precursor.....	5
Figure 1.2. Sequence Alignment of Stress-response Peptides Characterized in Various Insects.....	6
Figure 1.3. An Alignment of the Core Region (residues 8-21) Consisting of Structurally Categorized Stress-response Peptides.....	8
Figure 3.1.1. Sequence Alignment of a Family of Proteins with Their Carboxyl-Termini Similar to SRPs.	43
Figure 3.1.2. qRT-PCR Measurement of <i>M. sexta</i> SRP2 mRNA Levels.....	45
Figure 3.1.3. Induction of AMP Expression by Injection of PBS (grey) or SRP2 (black).	47
Figure 3.1.4. Fingerprint Region (NH-C α H/Side chain protons) of 2D ^1H - ^1H TOCSY Spectrum of SRP2 in H $_2$ O.....	48
Figure 3.1.5. Summary of the NOE connectivity for SRP2 in water.	49
Figure 3.1.6. Structural Representations of SRP2.	52
Figure 3.2.1. Cluster Heterogeneity Analysis of Simulated SRP2 and Mutants.	59
Figure 3.2.2. Average β -strand Propensity per Residue: SRP2.....	61
Figure 3.2.3. Average Helical Propensity per Residue: SRP2.	62
Figure 3.2.4. Average Root-mean-square-fluctuation per Residue: SRP2.	63
Figure 3.2.5. Average End-to-End Distance: SRP2.....	64
Figure 3.2.6. Salt-bridge Formation Propensity: SRP2.....	65
Figure 3.4.1. Fingerprint Region (NH-NH) of 2D 1H-1H NOESY Spectrum of SRP2 in Water, Showing Downfield Connectivities.....	75
Figure 3.4.2. The C α Protons Chemical Shift Deviation ($\Delta\text{C}\alpha\text{H}$) for SRP2 as Compared with Random Coil Values.	76
Figure 4.1.1. Determination of <i>M. sexta</i> SRP1 mRNA Levels by Quantitative Real-time PCR.	84
Figure 4.1.2. Induction of AMP Expression by Injection of PBS (grey) or SRP1 (black).	85
Figure 4.1.3. Fingerprint Region (NH-C α H/Side Chain Protons) of 2D ^1H - ^1H TOCSY Spectrum of SRP1 in H $_2$ O.	86
Figure 4.1.4. Summary of the NOE Connectivity for SRP1 in Water.....	87
Figure 4.1.5. Structural Representations of SRP1.	90
Figure 4.1.6. Eight Most Populated Clusters Derived from Simulation of SRP1.	91
Figure 4.1.7. Summary of NOE Violation Analysis of the Eight Most Populated Clusters from the Simulated Ensemble.	92
Figure 4.6.1: Transcript Levels of <i>M. sexta</i> SRP1 Gene in the 52 Tissue Samples.....	102
Figure 4.6.2: Fingerprint Region (HN-HN) of 2D ^1H - ^1H NOESY Spectrum.....	103

Figure 4.6.3. C α H Chemical Shift Index (Δ C α H) of Individual Residues in SRP1.....	104
Figure 5.1. Linear Schematic of p53.	109
Figure 7.1. ¹⁵ N-HSQC of p53TAD.....	161
Figure 7.2. Chemical Shift Perturbation per Residue of p53TAD Variants Relative to WT.	162
Figure 7.3. Spin-lattice (R ₁) and Spin-spin (R ₂) Relaxation Rates of p53TAD Variants.	163
Figure 7.4. Difference in R ₂ /R ₁ Between p53TAD Variants and -WT.	165
Figure 7.5. NHNOE/NONOE Ratio for Each p53TAD Variant.....	166
Figure 7.6. Reduced Spectral Density Approximation of $J(0)$ for p53TAD Variants.	168
Figure 7.7. Reduced Spectral Density Approximation of $J(\omega_N)$ for p53TAD Variants.	169
Figure 7.8. Reduced Spectral Density Approximation of $J(\omega_H)$ for p53TAD Variants.	170
Figure 7.9. $J(\omega_N)/J(\omega_H)$ Spectral Density Ratio of p53TAD Variants.....	171
Figure 7.10. Binding of p53TAD Variants to Binding Partners CBP-TAZ2, CBP-TAZ1, and N-HDM2.....	173

List of Tables

Table 3.1.1. Structural Statistics of the 20 Lowest Energy Structures for SRP2.	51
Table 3.2.1. Summary of NOE Violations for SRP2.	58
Table 4.1.1. Structural Statistics of the 20 Lowest Energy Structures for SRP1.	89
Table 7.1. Statistics for Relaxation and Reduced Spectral Density Mapping for p53TAD Variants..	164

Acknowledgements

First, I need to express my overwhelming and sincerest appreciation to my advisors Dr. Om Prakash and Dr. Michal Zolkiewski for their incredible and constant support throughout my Ph.D. studies. Their collective knowledge, insights, and unending patience have been a source of security that have aided my growth as a research scientist. It has been a privilege to have them as my advisors.

I would like to extend my thanks to my doctoral committee members Dr. Jianhan Chen and Dr. Yong-cheng Shi for taking the time to serve on my committee and to my outside chairperson Dr. Jocelyn McDonald. I would also like to thank my outside collaborator Dr. Haobo Jiang for the privilege of working as part of his project and supporting my research efforts.

Extra special thanks to both current and previous members of the Dr. Prakash lab and Zolkiewski Lab: Dr. Alvaro Herrera, Dr. Daisuke Takahashi, Dr. Indrani Pal, Dr. Przemyslaw Glaza, Chathurange Ranaweera, and Dr. Sunitha Shiva for their emotional support, encouragement, and technical expertise. In addition, I would also like to extend an extra special thank you to my fellow project researchers Indhujah Thevarajan, Dr. Xiaorong Liu, and Hawa Dembele for their extensive assistance and hard work over these many years.

And finally, I want to express my sincerest gratitude to the staff and faculty of the entire Biochemistry and Molecular Biophysics Department for providing the facilities, education and support to pursue my degree. And an additional special thanks to Susan Whitaker and Dr. John Tomich at the Biotechnology/Proteomics Core Lab for allowing me unfettered access to and extensive education of their facilities.

Dedication

To my mother, Rita J. Goss, for always nurturing
my passions and pursuits wherever they led.

To my father, Allan D. Schrag, for your unending
patience and understanding.

To my sister, Leah Oswald, for always kindling
my ambitions by pushing me strive to be the best version of myself.

All the sacrifices that you have made for me,
all these years,
you are what made this possible,
and I can never show enough gratitude
or express my appreciation enough.

I love you all so much.

This is for you.

Sincerely, Lynn.

Project Goals

In the organism *Manduca sexta*, the stress response peptide-1 (SRP1) and -2 (SRP2) are highly conserved paralogs that are both predicted to be 25-residue peptides. Produced as part of unstructured inactive precursors, proSRP1 and proSRP2 are likely converted by endogenous serine proteases into their mature forms through site-specific cleavage. Studies conducted by Dr. Xiaolong Cao examined mRNA expression profiles of SRP1 and -2 via qPCR, which suggested both are subject to immune inducibility. However, each paralog retains distinct tissue-specific ontological profiles during the course of development leading to distinct biological roles. Furthermore, injection of chemically synthesized SRP1 and -2 into naïve larvae significantly up-regulated mRNA expression of distinct, non-overlapping antimicrobial peptide genes. To explore the structural basis of these nuanced activity difference, we determined the 3D solution structures of SRP1 and -2 by two-dimensional ^1H - ^1H nuclear magnetic resonance (NMR) spectroscopy and showed that both fold into highly analogous β -hairpin motifs stabilized by disulfide linkages, a common feature observed for SRP orthologs. A most peculiar structural feature for both SRP1 and SRP2 is the presence of a single embedded residue position which has been observed in a minority of structurally categorized orthologs. Analysis of molecular dynamic motion through simulations of WT and mutants for SRP2 suggest that the embedded residue plays a critical role in modulating the stability of the β -hairpin. In addition, this stability then guides the structural dynamics of the intrinsically disordered N- and C-termini, each of which imbues a distinct cytokine activity of the peptide, which may explain the observed differences in biological activities between these highly conserved paralogs.

The second system discussed in this dissertation is p53, is one of the most studied genes in the human genome largely due to its critical roles in regulating cellular genotoxic stress response,

DNA damage directed repair, and determining cell-fate. However, little is known concerning cancer-linked mutational impacts upon the intrinsically disordered N-terminal transactivation domain (TAD), a critical integrator of intracellular signaling and site of interactions with p53 partner proteins. Through this work, we analyzed four common p53TAD cancer-associated mutations, -K24N, -N29K/N30D, -D49Y, and -W53G. Using biolayer interferometric (BLI) techniques to assess thermodynamics and kinetics, we determined that p53TAD-N29K/N30D compromises binding with E3 ubiquitin ligase Human Double-minute-2 binding while interactions with histone acetyl transferase CREB Binding Protein are impacted by p53TAD-N29K/N30D, -D49Y and -W53G mutations. NMR spectroscopy with reduced spectral density approaches reveals that p53TAD-N29K/N30D, -D49Y, and -W53G display subtle changes in local internal motion on the picosecond timescales but -N29K/N30D affects distal region motion as well. When these non-native internal motions overlap with specific binding surfaces correlate with diminished binding overall. It is proposed that the p53TAD conformational ensemble natively adopts binding compliant conformers while in solution and that cancer-associated mutations, even when distal from a binding surface, have the potential to perturb interactions by subtly altering this conformational ‘landscape’ within the IDP.

Chapter 1- Introduction: Structure and Function of Stress- response Peptides in Insects¹

Abstract

This review focuses on providing background of current status of research, discovery, structural and biological properties of stress-response peptides in insects.

¹This chapter was published in *Peptide-based Drug Discovery: Challenges and New Therapeutics* (2017) 15, 438-451.

1.1. Introduction

Stress-response peptides are a class of cytokine-like peptides that have been identified in various insects. Their activities have substantial impact on insect immune responses, such as plasmatocyte spreading, hemocyte proliferation, differentiation, paralysis, wound healing, growth blocking, and many other activities. Cytokines are a diverse group of small proteins in vertebrates and invertebrates, which modulate immune response by binding to their receptors on the cell surface (Beschlin, 2012; Dembic, 2015). They differ from growth factors or peptide hormones in the physiological processes under regulations (i.e. immunity versus growth, development, and reproduction), but the distinctions between them are often blurred. Accumulating evidence indicates that insect cytokines, analogous to human ones, are involved in the regulation of various innate immune responses.

In contrast to >100 different human cytokines, fewer than a dozen cytokines are characterized in any single species of insects, largely due to lack of research (Duque, 2014). For instances, in *Drosophila* Spätzle, Eiger, growth-blocking peptide, Vago, dawdle, decapentaplegic, and sex-peptide elicit immune responses (Clark, 2011; Deddouche, 2008; Tsuzuki, 2012). A family of bioactive peptides, initially identified in lepidopteran species as growth-blocking (GBP), plasmatocyte-spreading (PSP), paralytic (PP), or stress-response peptides (SRP), is now known to exist in at least six orders of insects (Clark, 1997; Duressa, 2015; Hayakawa, 1990; Matsumoto, 2012; Skinner, 1991; Yamaguchi, 2012). As expected from their broad taxonomic distribution, these cytokines are structurally and functionally diverse. Typically, 22-32 residues long, these peptides follow a loose consensus or (R/K)-X₁₋₁₅-C-X₇₋₉-G-X_{1/2}-C-X₁₋₁₅, where R/K indicates the putative cleavage activation site and x represents residues other than cysteine (e.g. *M. separata* GBP: (R)-ENFSGGCVAGYMRTPDGRCKPTF). They alter hemocyte behaviors (e.g. spreading,

movement, aggregation, and proliferation), delay growth, induce muscle contraction, and control brain formation. Consequently, there is not a general name for these evolutionarily related peptides. After considering the existing terms, we propose to use “stress-response peptides” to emphasize their roles in regulating the immune response of insects. We are aware that brain development, for example, does not fit the strict definition of cytokine functions, although most of the known functions are related to biotic and abiotic stresses.

In this chapter, we describe the structure and function of SRPS in insects. By summarizing related papers published since 1988, we provide an overview of the field and suggest new directions for future research on SRPs. While there is no review on this subject, a few review papers are available for interested readers to get a general background of insect immunity research (Buchon, 2014; Jiang, 2010; Strand, 2008).

1.2. Discovery

GBP, as a prototype of SRPs, was first isolated from the armyworm *M. (P.) separata* parasitized with a wasp that arrests the host’s development (Hayakawa, 1990). Intrigued by the paralysis caused by injection of larval hemolymph from another individual of the same species, Skinner et al. purified a group of seven paralytic peptides from *M. sexta*, *S. exigua*, and *H. virescens*. PSP was isolated from the hemolymph of the moth *P. (C.) includens* (Skinner, 1991). These 23-residue peptides all start with **Glu-Asn-Phe** and, therefore, are collectively named as ENF peptides (Strand, 2000). For instance, the most potent paralytic peptide (*M. sexta* PP1) has the sequence **Glu-Asn-Phe-Ala-Gly-Gly-Cys-Ala-Thr-Gly-Tyr-Leu-Arg-Thr-Ala-Asp-Gly-Arg-Cys-Lys-Pro-Thr-Phe** (Skinner, 1991). The cDNA cloning, RNA-sequencing analysis, and genome sequencing suggested that ENF peptides widely exist in lepidopteran insects as a secreted

precursor to be activated by specific hemolymph proteases upon wounding or infection (Clark, 1998; Hayakawa, 1995; Kamimura, 2001; Srikanth, 2011; Wang, 1999).

Remarkably, the precursor of *M. separata* GBP is encoded not only by transcript with a single open reading frame (ORF), but also by a much longer transcript containing three ORFs (Kanamori, 2010). The first two ORFs encode homologous proteins with the same pre-pro-peptide structure and, thus, named uENF1 and uENF2 (u for upstream). The proteolytic activation site of their precursors has not yet been determined experimentally, but the uENFs do not start with Glu-Asn-Phe, based on their sequences. Tricistronic mRNAs of uENF1-uENF2-ENF are known to exist in at least five families of Lepidoptera.

A 25-residue cytokine (HGIVVGTCPGLGYTRRGGFCFQDDDY), designated SRP, was first identified in hemolymph of the cutworm *S. litura* (Yamaguchi, 2012). Produced as a pre-pro-peptide, this SRP is similar in structure and function to the ENFs. Its gene expression is highly induced under stress conditions, such as parasitization, wounding, heating, and vibration. Long before that discovery, Hdd23 cDNA was isolated from *H. cunea*, by polymerase chain reaction (PCR)-based differential display and subtractive cloning (Shin, 1998). The function of Hdd23, an SRP-related protein in the fall webworm (*H. cunea*), remains uncharacterized, but its expression is highly induced 3 h after injecting *Escherichia coli* into the larvae. SRP homologs, similar Hdd23, are widely distributed in at least six orders of insects (Matsumoto, 2012).

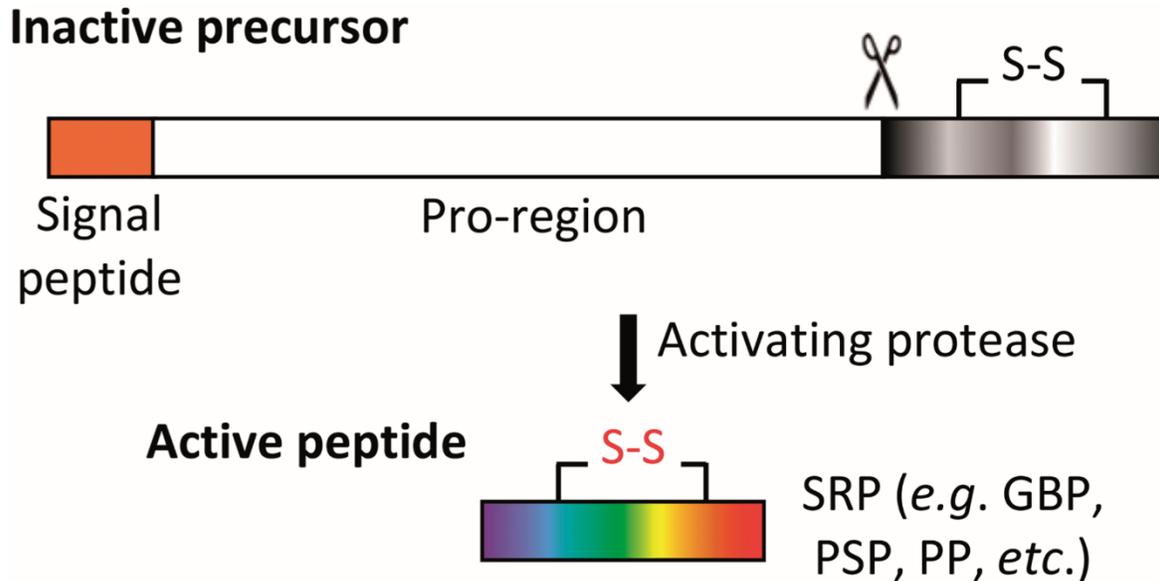


Figure 1.1. The Proteolytic Activation of a Stress-response Peptide (SRP) from its Precursor. GBP: growth-blocking peptide; PSP: plasmacyte-spreading peptide; PP: paralytic peptide.

1.3. Structure

SRP, as defined from a structural perspective, is the C-terminal domain of a much larger precursor. These pre-pro-peptides, typically 60-150 residues long, are disordered and poorly conserved for the most part (Matsumoto, 2012; Hayakawa, 1995; Wang, 1999; Volkman, 1999). Experimental results have suggested that these peptides are functionally inactive when they exist as a part of the pre-pro-peptides (Tsuzuki, 2012; Nakatogawa, 2009). The pre-region contains a signal for secretion and, after its removal, the mature proteins either go through the classical pathway of secretion to become extracellular precursors or are temporarily stored in secretory vesicles as active peptides waiting for the signal for release (Nakatogawa, 2009). Liberation of the C-terminal active peptide is accomplished through specific proteolytic cleavage by extracellular serine proteases or intracellular processing enzymes prior to storage (Hayakawa, 1995; Wang, 1999; Nakatogawa, 2009). This is shown in Figure 1.1. The active peptides span from 23 to 32

	1	5	10	15	20																			
MsGBP	E	N	F	S	G	G	C	V	A	G	Y	M	R	T	P	D	G	R	C	K	P	T	F	
MbGBP	E	N	F	A	G	G	C	L	T	G	F	M	R	T	P	D	G	R	C	K	P	T	F	
SlGBP	E	N	F	A	A	G	C	A	T	G	Y	Q	R	T	A	D	G	R	C	K	P	T	F	
TnPP1	E	N	F	S	G	G	C	L	A	G	Y	M	R	T	A	D	G	R	C	K	P	T	F	G
TnPP2	E	N	F	S	G	G	C	L	A	G	Y	M	R	T	A	D	G	R	C	K	P	T	F	
HvPP1	E	N	F	S	G	G	C	I	P	G	Y	M	R	T	A	D	G	R	C	K	P	T	Y	
HvPP2	E	N	F	A	G	G	C	I	P	G	Y	M	R	T	A	D	G	R	C	K	P	T	Y	
MsPP1	E	N	F	A	G	G	C	A	T	G	Y	L	R	T	A	D	G	R	C	K	P	T	F	
MsPP2	E	N	F	A	G	G	C	A	T	G	F	L	R	T	A	D	G	R	C	K	P	T	F	
SePP1	E	N	F	A	G	G	C	T	P	G	Y	Q	R	T	A	D	G	R	C	K	A	T	F	
SePP2	E	N	F	A	G	G	C	T	P	G	Y	Q	R	T	A	D	G	R	C	K	P	T	F	
SePP3	E	N	F	V	G	G	C	T	P	G	Y	Q	R	T	A	D	G	R	C	K	P	T	F	
AyPP	E	N	F	A	G	G	C	A	T	G	F	M	R	T	A	D	G	R	C	K	P	T	F	
BmPP1	E	N	F	V	G	G	C	A	T	G	F	K	R	T	A	D	G	R	C	K	P	T	F	
PiPSP	E	N	F	N	G	G	C	L	A	G	Y	M	R	T	A	D	G	R	C	K	P	T	F	
SeCAP	E	N	F	A	V	G	C	T	P	G	Y	Q	R	T	A	D	G	R	C	K	P	T	F	

Figure 1.2. Sequence Alignment of Stress-response Peptides Characterized in Various Insects. Growth-blocking peptides from *M. separata* (MsGBP), *M. brassicae* (MbGBP) and *S. litura* (SlGBP), paralytic peptides from *T. ni* (TnPP1 and TnPP2), *H. virescens* (HvPP1 and HvPP2), *M. sexta* (MsPP1 and MsPP2), *S. exigua* (SePP1, SePP2 and SePP3), *A. yamamai* (AYPP) and *B. mori* (bmpp1), plasmatocyte-spreading peptide from *P. includens* (PiPSP), and cardioactive peptide from *S. eridania* (SeCAP). Figure adapted from T. Aizawa, Y. Hayakawa, A. Ohnishi, N. Fujitani, K. D. Clark, M. R. Strand, K. Miura, N. Koganesawa, Y. Kumaki, M. Demura, K. Nitta and K. Kawano, *J. Biol. Chem.*, 2001, 276, 31813–31818. Copyright 2001, by the American Society for Biochemistry and Molecular Biology.

residues and demonstrate high sequence identity as shown in Figure 1.2. Structural cohesion of these peptides is maintained through a single disulfide loop and proper folding through a loosely defined β -hairpin sequence C-X(2)-G-X(4,6)-G-X(1,2)-C-(K/R) (Matsumoto, 2012).

Initial structural studies suggested a two-domain description (N-terminus and C-terminus) for these peptides. Each domain has a unique functional attribute(s) related to the structure of the peptide (Aizawa, 2001). However, more recent publications characterize ENF peptides that warrant bifurcation of C-terminus into a core region, which includes residues stabilized by the loop, and the disordered C-terminal residues (Nakatogawa, 2009; Umetsu, 2009). Characterization of each of these regions was accomplished by a combination of nuclear magnetic resonance (NMR), distance geometry, simulated annealing techniques, truncation, and mutation assays (Volkman, 1999; Nakatogawa, 2009; Aizawa, 1999; Yu, 1999; Clark, 2001a; Clark, 2001b; Miura, 2002; Watanabe, 2006). Conformations of these peptides was identified using coupling constants, chemical shift index, and nuclear Overhauser effect studies. Establishment of the positions of hydrogen bonds through either deuterium exchange resistance or temperature-gradient studies suggested regions of stability that exist within mature peptides (Volkman, 1999; Nakatogawa, 2009; Aizawa, 1999; Yu, 1999; Miura, 2002).

1.3.1. Core Domain

The core region is the most structurally well-defined and sterically consistent domain of the SRPs. Self-contained ensemble alignments of the core regions (residues 8-21) for each of these structurally analyzed SRP peptides yielded root-mean-square-deviations not exceeding 1.5Å for all heavy atoms, suggesting that this region is immobile (Volkman, 1999; Nakatogawa, 2009; Aizawa, 1999; Yu, 1999; Miura, 2002; Schrag, 2017). This can be visualized in an alignment of all structurally characterized peptides, as shown in Figure 1.3. The region spans from the first cysteine residue position to the $i + 2/3$ position relative to the second cysteine. Contained therein

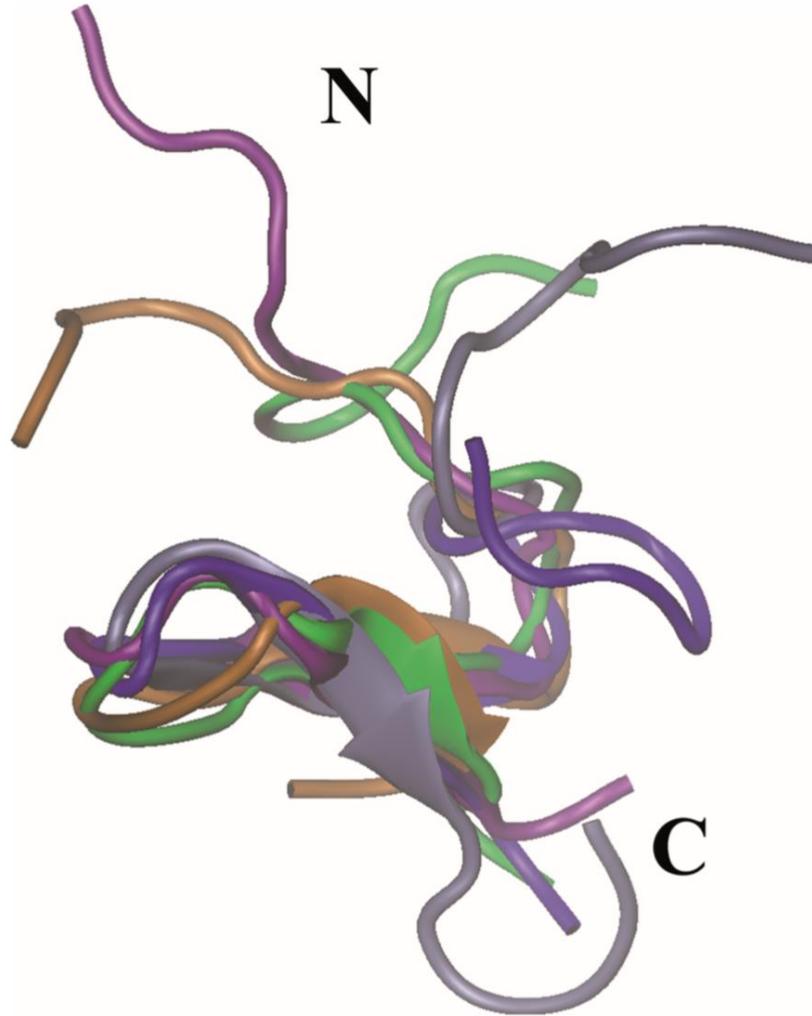


Figure 1.3. An Alignment of the Core Region (residues 8-21) Consisting of Structurally Categorized Stress-response Peptides. Cartoon representation with a color scheme: *M sexta* Paralytic peptide (violet), *P. includens* plasmatocyte spreading peptide (green), *M. separata* growth-blocking peptide (mauve), *B. mori* paralytic peptide (orange), and *M. sexta* stress-response peptide-2 (gray). Retention of the core region is apparent while the N- and C-termini are structural divergent due to observed disordered properties.

is the characteristic cysteine loop. Steric restriction of the loop in concert with the C-X(2)-G-X(4,6)-G-X(1,2)-C-[K/R] sequence tend to enforce a pair of anti-parallel β -strands that are only three or residues per strand (Matsumoto, 2012).

The β -strands are then flanked by two-turns. The first turn is located at the $i + 1$ position relative to the first cysteine and terminates at the $i + 4$ residue, typically constituting a four-member

turn. These have been recorded to occupy an array of hydrogen bond-stabilized β -turns (type I, I', II, and IV) as well as non-stabilized loops (Volkman, 1999; Nakatogawa, 2009; Aizawa, 1999; Yu, 1999; Miura, 2002). The second turn starts at the $i + 4/5$ residue position and terminates at the $i + 1/2$ position relative to the second cysteine. Due to the uneven number of amino acids that reside with these β -hairpins, an asymmetric stability arises, as shown in figure 1.3. Twisting is prevalent at turns opposite the stabilized region, often relying on the flexibility of residue G17 (Miura, 2002).

The establishment of a β -hairpin motif and primary sequences with alternating hydrophobic residues imbues the core region with amphipathic properties (Aizawa, 1999; Yu, 1999; Miura, 2002). When evaluated from a tertiary structural perspective, a combination of the disulfide linkage, hydrogen bonding with the β -hairpin structure, and hydrophobic side-chain packing stabilize the core region of these peptides. This amphipathic plane is thought to play a role in binding of these peptides with their receptors (Clark, 2001b). Mutation studies suggest that specific residues within the core region differentially affect the biological activities (Aizawa, 2001; Clark, 2001b).

So as to assess the impact of disrupting portions of the amphipathic plane, scanning alanine mutations of specific residues in the core region were utilized. Alanine replacement of the first and second cysteine profoundly affects the tertiary structure and eliminates all activity. Replacement of the charged residues surrounding the second cysteine diminishes activity to a lesser degree (Clark, 2001a). In another mutation study of *M. separata* GBP, the mutants D16E/N/L also had appreciable losses in plasmatocyte-spreading activity, but the introduction of these large side-chain residues does not result in a change in structure (Aizawa, 2001). This was confirmed through NMR structural studies. Antagonist binding studies of *P. includens* PSP's D16A, R18A, and K20A mutants revealed that none of these charged residues are critical for binding (Clark, 2004). This

implies that their role is crucial to varying degrees for proper receptor activation upon binding (Clark, 2001b). In contrast, R13 was the only charged residue in the core-region, which was essential for binding and activity, but not required for maintaining structure (Clark, 2004). The amino acid R13 is exposed on the β -strand opposite both N- and C-terminal extensions, and it is not incorporated into turns (Volkman, 1999; Nakatogawa, 2009; Aizawa, 1999; Yu, 1999; Miura, 2002). A positively charged residue (R/K) is highly conserved at this position among ENF factors and SRPs in the order of Lepidoptera (Matsumoto, 2012). Hydrophobic residues with the core are somewhat less understood than their hydrophilic counterparts, as mutation studies concerning them are sparse. A highly conserved motif with the published structures was the presence of a bulky residue at the $i + 4$ position relative to the first cysteine. NMR studies suggested that these residues are embedded into the structure and may be sterically confined (Volkman, 1999; Yu, 1999).

1.3.2. N-terminal Extension

This region is six to seven residues long and precedes the first cysteine. It is largely unstructured, as shown in structural studies on all SRPs, except for *M. sexta* PP1 (Volkman, 1999; Nakatogawa, 2009; Aizawa, 1999; Yu, 1999; Umetsu, 2009; Miura, 2002). NMR relaxation studies confirmed the lack of cohesive structural attributes (Umetsu, 2009). Nevertheless, it seems to have features for proteolytic activation and receptor-specific binding.

Although the definition of ENF factors has broadened since its discovery, the first three residues (typically ENF) may play a role in recognition by its activating protease that cleaves between K/R and (E/D)NF (Matsumoto, 2012; Strand, 2000). Precise cleavage at this bond is believed to be critical for yielding functional peptides (Umetsu, 2009; Clark, 2001a). Truncation

analysis of *M. separata* GBP revealed that removal of E1 results in complete loss of plasmatocyte-spreading activity. However, later *P. includens* PSP studies revealed that deletion of E1 had almost no effect on binding and the peptide still functions as a strong functional antagonist (Clark, 2004). Elongation of the PSP resulted in deleterious effects on spreading and binding (Clark, 2001a; Clark 2004). E1A and N2A mutants of *P. includens* PSP dramatically increased plasmatocyte spreading, indicating that association with its receptor is driven by the precise position of H₃N⁺-E1 via the backbone (Clark, 2001a; Clark, 2001b). The presence of E1 and N2 sidechains reduces the stimulatory effect on plasmatocyte spreading but may assist in efficient proteolysis of *P. includens* PSP precursor (Clark, 2001b; Strand, 2000). In addition to H₃N⁺-E1, F3 has a high stereospecificity requirement for inducing plasmatocyte spreading, but not much loss in binding behavior, as long as it retains its hydrophobic character, chirality, and the van der Waals radius (Clark, 2001a). It was then established that *P. includens* PSP must rely on crosslinking activation with its receptor, as deletion of the AGG linker between the ENF sequence and the core region results in inactivity with only moderate disruption to binding activity (Clark, 2001a; Clark, 2004).

1.3.3. C-terminal Extension

The C-terminal region has the greatest variety in terms of amino acid composition, structure, and regulatory mechanisms observed (see below). The best-characterized ENF factors (e.g. *M. Separata* GBP, *P. includens* PSP, *B. mori* PP and *M. sexta* PP1) terminate with Pro-Thr-Phe, which is disordered (Volkman, 1999; Aizawa, 1999; Yu, 1999; Miura, 2002). Truncation analysis revealed the importance of T22 for maintaining *M. separata* GBP's native secondary structure of the core region as well as plasmatocyte spreading activity (Aizawa, 2001). T22A mutation of *P. includens* PSP partially supports this finding with less stimulation of plasmatocyte

spreading (Clark, 2001b). Interestingly, the deletions of F23 and T22-F23 from the C-terminus of *M. separata* GBP abolished its mitogenic activity, but induced plasmatocyte spreading to a slight less extent than the wild-type. These studies suggest that T22 in the C-terminal extension distinctively affects the biological activities. *M. sexta* and *S. litura* SRPs have an acidic C-terminal extension containing two to three Asp residues followed by Tyr (Yamaguchi, 2012). This disordered region is more guided towards making a loop as the electrostatic charges repel the neighboring aspartic acid residues from one another. Analysis of truncation or mutation of residues within these SRPs is not yet available.

Some SRPs have a C-terminal region substantially longer than those in *S. litura* SRP and *H. cunea* Hdd23, which seems to have a greater diversity in sequence and function. Isolated hemocyte chemotactic peptide from the integument of *M. Separata* shows that residues T22 and T23 were O-glycosylated with *N*-acetylhexosamine (Nakatogawa, 2009). In solution, *M. separata* hemocyte chemotactic peptide (HCP) C-terminal domain is disordered. Analogous to ENF-sequence, *N*-acetylhexosamine modifications might guide cleavage to the scissile bond, but this remains unexplored through mutation analysis. It has not been established whether this modification results in a similar structural modification of the core region, as seen in GBP (1-21) truncation (Nakatogawa, 2009; Aizawa, 2001). In addition, it is not well understood if the extended C-terminus aids in stability, degradation, or localization of the peptide.

M. separata GBP consisting of 28 residues is the other characterized longer variant; identified in *M. separata* parasitized by the wasp *A. kariyai*. NMR studies suggested that these peptides have basically the same tertiary structures and the extension of the C-terminal regions is disordered. However, the C-terminal region of this peptide undergoes a conformational transition

for a random coil to an α -helical state in the presence of a dodecylphosphocoline micelle (Umetsu, 2009).

1.4. Biological Function

In essence, SRPs are insect cytokines that regulate immune responses. Based on the evidence available so far, this definition does entirely cover their diverse functions. SRPs induce plasmacyte spreading (a cytokine effect) during cellular immune response, stimulate cell proliferation (*i.e.* mitogenic effect of growth factors), and participate in the developmental of fat body, integument, hematopoietic organ and the brain. With multiple SRPs found in a single insect species (*e.g.* 11 in *M. sexta*), their functions are anticipated to extend beyond immunity, growth and development. Here we focus on the role of SRPs in the immunity regulation and their mitogenic effect on immune cells.

1.4.1. Plasmacyte Spreading and Attachment

Plasmacyte spreading is a function shared by most SRPs studied thus far including *M. separata* GBP, *P. includens* PSP, *M. sexta* PP1, *S. litura* SRP and *L. migratoria* GBP (Clark, 1997; Duressa, 2015; Yamaguchi, 2012; Wang, 1999; Aizawa, 2001). This response occurs shortly after injury to stimulate wound healing and reduce hemolymph loss. Having an open circulatory system, insects run a great risk of blood loss upon wounding (Krautz, 2014). Spread plasmacytes may form a clot along with other hemocytes and plasm proteins. As discussed below, *M. separata* HCP attracts certain hemocytes and stimulated their aggregation (Nakatogawa, 2009). Immuno-depletion of HCP results in a substantial increase in blood loss.

Plasmatocyte spreading and attachment to tissues, body wall, or invading pathogens are likely responsible for the mysterious loss and reappearance of plasmatocytes in *M. sexta* larvae following an injection of bacteria or PP1 (Wang, 1999; Geng, 1989). Recognition of microbes by hemolymph proteins triggers a serine protease system which may lead to the proteolytic activation of SRP precursors within minutes, and binding of the active SRPs to their receptors on plasmatocytes alters the freely circulating cells to adherent cell aggregates (Jiang, 2010; Strand, 2008).

1.4.2. Paralysis

Injection of active ENF factors into hemocoel of lepidopteran insects results in rigid paralysis of the larvae (Skinner, 1991; Wang, 1999; Ha, 1999; Seino, 1998) This change needed several minutes to take effect in *B. mori* and its duration was found to be dose-dependent (Ishii, 2015). PP1 may have induced a slow muscle contraction that overrides insect motility in a neuron-independent manner. While body length did change after the PP treatment, evidence that *B. mori* PP directly affects muscular activities is weak (Ishii, 2008). Measurement of electrophysiological response of muscle is needed to validate the hypothesis. In addition, it seems important to explore a possible relationship between hemocyte response and larval paralysis, which is an artifact created by injecting the activate peptide at a high dose or injecting activate hemolymph from another larva of the same species.

1.4.3. Wound Healing

P. includens PSP applied to plasmatocyte suspension in agarose-coated wells stimulates their aggregation to the gel surface. Conversely, granulocytes spread uniformly in the absence of

the peptide, but their spreading is inhibited by PSP in a concentration-dependent manner (Strand, 1999). Granulocytes and plasmatocytes cooperate to encapsulate foreign targets in a three-step process, likely regulated by *P. includens* PSP (Pech, 1996). In addition to plasmatocyte spreading, *B. mori* PP, *M. separata* GBP and HCP, *M. sexta* PP1 have all been demonstrated to promote plasmatocyte aggregation dose-dependently (Wang, 1999; Nakatogawa, 2009; Aizawa, 2001; Miura, 2002). Interestingly, HCP is expressed not only in hemocytes, but also in the epidermis of *M. separata* larvae. After wounding, the active peptide from ruptured epidermal cells recruits hemocytes to the site of injury and stimulates the release of more HCP from the aggregated hemocytes by a positive feedback mechanism. It is suggested that HCP increases the rate of chemotactic aggregation in response to tissue damage, as it promotes plasmatocyte aggregation and hemocyte deposition proximal to the wound site (Nakatogawa, 2009).

1.4.4. Phagocytosis

Aseptic wounding seldom occurs in a natural environment. Insect are well prepared for fighting microbial infection via phagocytosis, a key immune mechanism. Evidence suggest that SRPs assist in the cellular response that clears the surface bacteria. Quantitative PCR showed that *B. mori* PP upregulates the expression of tetraspnain, a hemocyte surface protein involved in encapsulation (Zhuang, 2007). Co-injection of *B. mori* PP with *Staphylococcus aureus* largely enhanced hemocyte phagocytosis of the bacterium (Ishii, 2010). Serralyisin is a metalloprotease from *S. marcescens*, suggested to digest tetraspanin to disrupt the hemocyte responses. Injection of serralyisin into hemocoel significantly decreased phagocytosis and plasmatocyte spreading in *B. mori* (Ishii, 2015).

1.4.5. Reactive Oxygen/Nitrogen Species and Nitric Oxide Synthase Production

Upon wounding or other stresses, insects generate peroxides, hydroxyl radical, nitric oxide, and other reactive oxygen/nitrogen species to kill microbes at the site of injury or injection. This is accompanied by proteolytic activation of extracellular serine proteases that yield active SRPs, Spätzle, and phenoloxidase. Phenoloxidase catalyzes the formation of reactive compound (*e.g.* 5,6-dihydroxyindole) that are broad-spectrum anti-microbials against viruses, bacteria, fungi, and parasites (Zhao, 2011; Jiang, 2010). As cytokines, SRPs and Spätzle lead to the upregulation of expression of immunity-related genes in minutes to hours. In *B. mori*, PP treatment elevated the expression of inducible nitric oxide synthase gene in the fat body and anterior midgut (Ishii, 2013). Injection of L-N^G-nitroarginine methyl ester (a NOS inhibitor) repressed the phosphorylation of p38 mitogen-activated protein kinase (MAPK) and expression of tetraspanin E and antimicrobial peptides (AMPs). Co-injection of PP and live *S. aureus* led to delayed larval death. These data suggest that nitric oxide is a downstream effector of SRPs (Ishii, 2013).

1.4.6. Induced AMP Synthesis

A hallmark of insect immunity is the dramatic increase of Amp transcript and protein levels after immune challenge (Buchon, 2014). Although not as pronounced as in *S. marcescens* infection, *D. melanogaster* GBP did upregulate Metchnikowin (Mtk, an AMP) expression independently of the Toll pathway, since removal of Spätzle or dorsal-dif had no impact on GBP-induced Mtk expression (Tsuzuki, 2012). RNA interference (i) of *Drosophila* immune deficiency protein (Imd) or c-Jun n-terminal kinase (JNK) resulted in the loss of Mtk induction by GBP and, since RNAi of relish, peptidoglycan recognition protein-LC or -LE (PGRP-LC or -LE) had no effect on or even enhanced Mtk expression, the signal of DAP-PG sensed by PGRP-LC and -LE and Imd does not

seem to go through the classical relish-mediated Amp gene upregulation. Instead, an Imd-JNK branch is likely taken to induce Mtk expression. Under various stresses, the GBP may interact with its receptor dependent on Imd-JNK signaling for upregulating Mtk expression and other immune reactions (Tsuzuki, 2012). differing somewhat from the *Drosophila* GBP, *B. mori* PP induced expression of AMP and other immunity-related genes via a MAPK signaling pathway (2010). Inhibition of p38 MAPK severely diminished host resistance of *B. mori* against *S. aureus*. Upregulated transcription of cecropin A and moricin genes goes through p38 MAPK in *B. mori* (Ishii, 2015). SRPs have been shown to play roles critical for organism survival. The downregulated GBP genotype in *T. castaneum*, either under mechanical or thermal stress, has substantially less survivability than the upregulated genotype (Kiyotake, 2014).

1.4.7. Growth-Blocking Activity

Insect larval growth is regulated by SRPs, including *M. separata* GBP, *P. includens* PSP, *D. melanogaster* GBP, *S. litura* SRP, and *S. exigua* and *L. migratoria* GBP (Tsuzuki, 2012; Clark, 1997; Duressa, 2015; Hayakawa, 1990; Yamaguchi, 2012; Wan, 2013). In *M. separata*, GBP injection somehow led to a large activity loss of juvenile hormone esterase, a key enzyme involved in downregulation of the hormone level in preparation for pupation (Hayakawa, 1990). The activity loss is accompanied by a reduction in larval weight gain and delayed pupal formation. In a bioassay of *B. mori* larvae, artificial introduction of the *M. separata* GBP retarded larval body weight gain in a dose-dependent fashion (Miura, 2002; Hayakawa, 1993). The reduced weight gain was less dose-sensitive than that due to injection of *B. mori* PP. After heat treatment, transgenic *D. melanogaster* expressing pro-GBP under a heat-sensitive promoter, experienced a 10–15 h delay in pupation in comparison to the controls (Tsuzuki, 2012). Hence, the *Drosophila* GBP was

suggested to a contributor of growth regulation and innate immunity. Delays in pupation were also observed in *S. litura* injected with the *M. separata* GBP or native SRP (Yamaguchi, 2012; Hayakawa, 1993). The latter had a correlating decrease in feeding behavior and weight gain. Together, these studies suggested that SRP at an elevated level affects insulin-like peptide gene expression correlated with reduced feeding and growth-retardation (Yamaguchi, 2012).

1.4.8. Mitogenic Activity

Following the early discoveries on the mitogenic activity of *M. separata* GBP and its mutants, stress-response *L. cuprina* GBP stimulates cell growth in a concentration-dependent manner (Tsuzuki, 2012; Aizawa, 2001; Hayakawa, 1998; Ohnishi, 2001). Ectopic introduction of the fly's GBP into high Five cells increased thymidine incorporation. *M. separata* GBP caused a similar incorporation into human keratinocytes and *Sf9* cells (Hayakawa, 1998). This is paralleled in day 0, last instar larvae of *M. separata* injected with 0.1 or 1 pMol of the peptide, which gained more weight than the injected control. However, larvae injected with 10 pMol of GBP gained significantly less weight than the control (Hayakawa, 1998). These observations suggest that mitogenic activity of the stress-response GBP is more prevalent at low concentrations than at high ones.

1.5. Conclusion

SRPs are 19- to 32-residue peptides located at C-terminus of the precursors. The peptide is liberated by specific proteases that recognize and cleave between end of the pro-region (K/R typically) and first residue of the active peptide. SRPs have three distinct regions: disordered N-

and C-termini flanking a stable core domain. It is evident from structure, activity, and binding studies of *M. separata* GBP and *P. includens* PSP that the C-terminus contains residues important for mitogenic activity, the N-terminus is essential for plasmatocyte spreading, and the core region is critical for both activities and most of the binding behavior. The degree of consistency in structural features of ENF factors is fairly high, because the peptides are >60% identical in sequence. They also have a large overlap in causing plasmatocyte spreading. retention of this activity was observed in more divergent homologs (e.g. SRPs), suggesting that the peptides co-evolved with receptor(s) that retain analogous activities. As such, there is an urgent need for the identification of their receptor(s) in order to explain the discriminatory properties of SRPs. It is also important to identify their activating proteases and explain how wounding or infection triggers the cleavage activation of SRP precursors. Such research would stimulate the exploration of more divergent homologs, both structurally and functionally, their interactions with the co-evolving receptor(s), and possible links with other responses (e.g. phenoloxidase and Spätzle generation) in a unified immune system.

1.6. References

- Aizawa, T., Fujitani, N., Hayakawa, Y., Ohnishi, A., Ohkubo, T., Kumaki, Y., Kawano, K., Hikichi, K., & Nitta, K. (1999). Solution structure of an insect growth factor, growth-blocking peptide. *Journal of Biological Chemistry*, 274(4), 1887-1890.
- Aizawa, T., Hayakawa, Y., Ohnishi, A., Fujitani, N., Clark, K. D., Strand, M. R., Miura, K., Koganesawa, N., Kumaki, Y., Demura, M., & Nitta, K. (2001). Structure and Activity of the Insect Cytokine Growth-blocking Peptide Essential Regions for Mitogenic and Hemocyte-stimulating Activities Are Separate. *Journal of Biological Chemistry*, 276(34), 31813-31818.
- Beschin, A., & Müller, W. E. (Eds.). (2012). *Invertebrate cytokines and the phylogeny of immunity: facts and paradoxes* (Vol. 34). Springer Science & Business Media.
- Buchon, N., Silverman, N., & Cherry, S. (2014). Immunity in *Drosophila melanogaster*—from microbial recognition to whole-organism physiology. *Nature reviews immunology*, 14(12), 796-810.
- Clark, K. D., Pech, L. L., & Strand, M. R. (1997). Isolation and Identification of a Plasmacyte-spreading Peptide from the Hemolymph of the Lepidopteran Insect *Pseudaletia includens*. *Journal of Biological Chemistry*, 272(37), 23440-23447.
- Clark, K. D., Witherell, A., & Strand, M. R. (1998). Plasmacyte Spreading Peptide Is Encoded by an mRNA Differentially Expressed in Tissues of the Moth *Pseudaletia includens*. *Biochemical and biophysical research communications*, 250(2), 479-485.
- Clark, K. D., Volkman, B. F., Thoetkiattikul, H., Hayakawa, Y., & Strand, M. R. (2001a). N-terminal residues of plasmacyte-spreading peptide possess specific determinants required for biological activity. *Journal of Biological Chemistry*, 276(40), 37431-37435.

- Clark, K. D., Volkman, B. F., Thoetkiattikul, H., King, D., Hayakawa, Y., & Strand, M. R. (2001b). Alanine-scanning mutagenesis of plasmatocyte spreading peptide identifies critical residues for biological activity. *Journal of Biological Chemistry*, 276(21), 18491-18496.
- Clark, K. D., Garczynski, S. F., Arora, A., Crim, J. W., & Strand, M. R. (2004). Specific residues in plasmatocyte-spreading peptide are required for receptor binding and functional antagonism of insect immune cells. *Journal of Biological Chemistry*, 279(32), 33246-33252.
- Clark, R. I., Woodcock, K. J., Geissmann, F., Trouillet, C., & Dionne, M. S. (2011). Multiple TGF- β superfamily signals modulate the adult *Drosophila* immune response. *Current Biology*, 21(19), 1672-1677.
- Deddouche, S., Matt, N., Budd, A., Mueller, S., Kemp, C., Galiana-Arnoux, D., D., Dostert, C., Antoniewski, C., Hoffmann, J.A. & Imler, J. L. (2008). The DExD/H-box helicase Dicer-2 mediates the induction of antiviral activity in *drosophila*. *Nature immunology*, 9(12), 1425.
- Dembic, Z. (2015). *The cytokines of the immune system: The role of cytokines in disease related to immune response*. Academic Press.
- Duressa, T. F., Boonen, K., Hayakawa, Y., & Huybrechts, R. (2015). Identification and functional characterization of a novel locust peptide belonging to the family of insect growth blocking peptides. *Peptides*, 74, 23-32.
- Geng, C., & Dunn, P. E. (1989). Plasmatocyte depletion in larvae of *Manduca sexta* following injection of bacteria. *Developmental & Comparative Immunology*, 13(1), 17-23.

- Ha, S. D., Nagata, S., Suzuki, A., & Kataoka, H. (1999). Isolation and structure determination of a paralytic peptide from the hemolymph of the silkworm, *Bombyx mori*. *Peptides*, 20(5), 561-568.
- Hayakawa, Y. (1990). Juvenile hormone esterase activity repressive factor in the plasma of parasitized insect larvae. *Journal of Biological Chemistry*, 265(19), 10813-10816.
- Hayakawa, Y., & Yasuhara, Y. (1993). Growth-blocking peptide or polydnavirus effects on the last instar larvae of some insect species. *Insect biochemistry and molecular biology*, 23(2), 225-231.
- Hayakawa, Y., Ohnishi, A., Yamanaka, A., Izumi, S., & Tomino, S. (1995). Molecular cloning and characterization of cDNA for insect biogenic peptide, growth-blocking peptide. *FEBS letters*, 376(3), 185-189.
- Hayakawa, Y., & Ohnishi, A. (1998). Cell growth activity of growth-blocking peptide. *Biochemical and biophysical research communications*, 250(2), 194-199.
- Ishii, K., Hamamoto, H., Kamimura, M., & Sekimizu, K. (2008). Activation of the silkworm cytokine by bacterial and fungal cell wall components via a reactive oxygen species-triggered mechanism. *Journal of Biological Chemistry*, 283(4), 2185-2191.
- Ishii, K., Hamamoto, H., Kamimura, M., Nakamura, Y., Noda, H., Imamura, K., Mita, K., & Sekimizu, K. (2010). Insect cytokine paralytic peptide (PP) induces cellular and humoral immune responses in the silkworm *Bombyx mori*. *Journal of Biological Chemistry*, 285(37), 28635-28642.
- Ishii, K., Adachi, T., Hamamoto, H., Oonishi, T., Kamimura, M., Imamura, K., & Sekimizu, K. (2013). Insect cytokine paralytic peptide activates innate immunity via nitric oxide

- production in the silkworm *Bombyx mori*. *Developmental & Comparative Immunology*, 39(3), 147-153.
- Ishii, K., Hamamoto, H., & Sekimizu, K. (2015). Paralytic peptide: an insect cytokine that mediates innate immunity. *Archives of insect biochemistry and physiology*, 88(1), 18-30.
- Jiang, H., Vilcinskas, A., & Kanost, M. R. (2010). Immunity in lepidopteran insects. In *Invertebrate immunity* (pp. 181-204). Springer, Boston, MA.
- Kamimura, M., Nakahara, Y., Kanamori, Y., Tsuzuki, S., Hayakawa, Y., & Kiuchi, M. (2001). Molecular cloning of silkworm paralytic peptide and its developmental regulation. *Biochemical and biophysical research communications*, 286(1), 67-73.
- Kanamori, Y., Hayakawa, Y., Matsumoto, H., Yasukochi, Y., Shimura, S., Nakahara, Y., Kiuchi, M., & Kamimura, M. (2010). A eukaryotic (insect) tricistronic mRNA encodes three proteins selected by context-dependent scanning. *Journal of Biological Chemistry*, 285(47), 36933-36944.
- Kiyotake, H., Matsumoto, H., Nakayama, S., Sakai, M., Miyatake, T., Ryuda, M., & Hayakawa, Y. (2014). Gain of long tonic immobility behavioral trait causes the red flour beetle to reduce anti-stress capacity. *Journal of insect physiology*, 60, 92-97.
- Kodama Susumu, Inui Hideyuki, & Okawa Hideo. (2001). Applying the molecular function of cytochrome P450 to monitoring and reducing the load of persistent organic pollutants. *Pestic. Biochem. Physiol*, 71, 156.
- Krautz, R., Arefin, B., & Theopold, U. (2014). Damage signals in the insect immune response. *Frontiers in plant science*, 5, 342.
- Lacy, P. (2015). Secretion of cytokines and chemokines by innate immune cells. *Frontiers in immunology*, 6, 190.

- Matsumoto, H., Tsuzuki, S., Date-Ito, A., Ohnishi, A., & Hayakawa, Y. (2012). Characteristics common to a cytokine family spanning five orders of insects. *Insect biochemistry and molecular biology*, 42(6), 446-454.
- Miura, K., Kamimura, M., Aizawa, T., Kiuchi, M., Hayakawa, Y., Mizuguchi, M., & Kawano, K. (2002). Solution structure of paralytic peptide of silkworm, *Bombyx mori*. *Peptides*, 23(12), 2111-2116.
- Nakatogawa, S. I., Oda, Y., Kamiya, M., Kamijima, T., Aizawa, T., Clark, K. D., Demura, M., Kawano, K., Strand, M.R., & Hayakawa, Y. (2009). A novel peptide mediates aggregation and migration of hemocytes from an insect. *Current Biology*, 19(9), 779-785.
- Ohnishi, A., Oda, Y., & Hayakawa, Y. (2001). Characterization of receptors of insect cytokine, growth-blocking peptide, in human keratinocyte and insect Sf9 cells. *Journal of Biological Chemistry*, 276(41), 37974-37979.
- Pech, L. L., & Strand, M. R. (1996). Granular cells are required for encapsulation of foreign targets by insect haemocytes. *Journal of cell science*, 109(8), 2053-2060.
- Schrag, L.G., Cao, X., I Herrera, A., Wang, Y., Jiang, H., & Prakash, O. (2017). Solution structure and expression profile of an insect cytokine: *Manduca sexta* stress response peptide-2. *Protein and peptide letters*, 24(1), 3-11.
- Shin, S. W., Park, S. S., Park, D. S., Kim, M. G., Kim, S. C., Brey, P. T., & Park, H. Y. (1998). Isolation and characterization of immune-related genes from the fall webworm, *Hyphantria cunea*, using PCR-based differential display and subtractive cloning. *Insect biochemistry and molecular biology*, 28(11), 827-837.
- Skinner, W. S., Dennis, P. A., Li, J. P., Summerfelt, R. M., Carney, R. L., & Quistad, G. B. (1991). Isolation and identification of paralytic peptides from hemolymph of the lepidopteran

- insects *Manduca sexta*, *Spodoptera exigua*, and *Heliothis virescens*. *Journal of Biological Chemistry*, 266(20), 12873-12877.
- Srikanth, K., Park, J., Stanley, D. W., & Kim, Y. (2011). Plasmacyte-spreading peptide influences hemocyte behavior via eicosanoids. *Archives of insect biochemistry and physiology*, 78(3), 145-160.
- Strand, M. R., Hayakawa, Y., & Clark, K. D. (2000). Plasmacyte spreading peptide (PSP1) and growth blocking peptide (GBP) are multifunctional homologs. *Journal of Insect Physiology*, 46(5), 817-824.
- Strand, M. R., & Clark, K. D. (1999). Plasmacyte spreading peptide induces spreading of plasmacytes but represses spreading of granulocytes. *Archives of Insect Biochemistry and Physiology: Published in Collaboration with the Entomological Society of America*, 42(3), 213-223.
- Strand, M. R., Hayakawa, Y., & Clark, K. D. (2000). Plasmacyte spreading peptide (PSP1) and growth blocking peptide (GBP) are multifunctional homologs. *Journal of Insect Physiology*, 46(5), 817-824.
- Strand, M. R. (2008). The insect cellular immune response. *Insect Sci* 15 (1): 1–14.
- Tsuzuki, S., Ochiai, M., Matsumoto, H., Kurata, S., Ohnishi, A., & Hayakawa, Y. (2012). *Drosophila* growth-blocking peptide-like factor mediates acute immune reactions during infectious and non-infectious stress. *Scientific reports*, 2, 210.
- Umetsu, Y., Aizawa, T., Muto, K., Yamamoto, H., Kamiya, M., Kumaki, Y., Mizuguchi, M., Demura, M., Hayakawa, Y., & Kawano, K. (2009). C-terminal elongation of growth-blocking peptide enhances its biological activity and micelle binding affinity. *Journal of Biological Chemistry*, 284(43), 29625-29634.

- Volkman, B. F., Anderson, M. E., Clark, K. D., Hayakawa, Y., Strand, M. R., & Markley, J. L. (1999). Structure of the insect cytokine peptide plasmatocyte-spreading peptide 1 from *Pseudoplusia includens*. *Journal of Biological Chemistry*, *274*(8), 4493-4496.
- Wan, H., Lee, K. S., Kim, B. Y., Yuan, M., Zhan, S., Lu, Y., You, H., Li, J., & Jin, B. R. (2013). Developmental regulation and antifungal activity of a growth-blocking peptide from the beet armyworm *Spodoptera exigua*. *Developmental & Comparative Immunology*, *41*(2), 240-247.
- Wang, Y., Jiang, H., & Kanost, M. R. (1999). Biological activity of *Manduca sexta* paralytic and plasmatocyte spreading peptide and primary structure of its hemolymph precursor. *Insect biochemistry and molecular biology*, *29*(12), 1075-1086.
- Watanabe, S., Tada, M., Aizawa, T., Yoshida, M., Sugaya, T., Taguchi, M., Kouno, T., Nakamura, T., Mizuguchi, M., & Demura, M. (2006). N-terminal mutational analysis of the interaction between growth-blocking peptide (GBP) and receptor of insect immune cells. *Protein and peptide letters*, *13*(8), 815-822.
- Yamaguchi, K., Matsumoto, H., Ochiai, M., Tsuzuki, S., & Hayakawa, Y. (2012). Enhanced expression of stress-responsive cytokine-like gene retards insect larval growth. *Insect biochemistry and molecular biology*, *42*(3), 183-192.
- Yu, X. Q., Prakash, O., & Kanost, M. R. (1999). Structure of a paralytic peptide from an insect, *Manduca sexta*. *The Journal of peptide research*, *54*(3), 256-261.
- Zhao, P., Lu, Z., Strand, M. R., & Jiang, H. (2011). Antiviral, anti-parasitic, and cytotoxic effects of 5, 6-dihydroxyindole (DHI), a reactive compound generated by phenoloxidase during insect immune response. *Insect biochemistry and molecular biology*, *41*(9), 645-652.

Zhuang, S., Kelo, L., Nardi, J. B., & Kanost, M. R. (2007). An integrin-tetraspanin interaction required for cellular innate immune responses of an insect, *Manduca sexta*. *Journal of Biological Chemistry*, 282(31), 22563-22572.

Chapter 2- NMR Spectroscopic and Other Biophysical Methods Used in the Study of Stress-response Peptide-1 and Stress-response Peptide-2

2.1. NMR Spectroscopy

1D and 2D ^1H - ^1H NMR experiments were performed with the 11.75 Tesla Varian 500 MHz VNMRS system (Agilent Technologies Inc., Palo Alto, CA), operating at 499.84 MHz for ^1H frequency. The NMR data were acquired at Biomolecular NMR Core Facility in the Department of Biochemistry and Biophysics at Kansas State University, using a 5 mm cryogenic triple resonance inverse detection pulse field gradient probe at 25 °C. The purified Stress-response peptide1 (SRP1) or SRP2 was dissolved in 650 μL of 10% D_2O and 90% triple distilled H_2O to a final concentration of 1.8mM (SRP1) or 2.2 mM (SRP2) for NMR analysis. A total of 512 increments of 4K data points were recorded for the 2D ^1H - ^1H Double Quantum Filtered Correlation Spectroscopy (DQFCOSY) experiments whereas 2D ^1H - ^1H Total Correlation Spectroscopy (TOCSY) and Nuclear Overhauser Effect Spectroscopy (NOESY) experiments were performed using 2000 t_2 and 256 t_1 data points with spectral width of 12 ppm in each dimension, and 16 transient per increment (Rance, 1983; Bax, 1985; Kumar, 1995). Spin-lock time of 80 ms at B1 field strength of 7 kHz was used for 2D ^1H - ^1H TOCSY experiments using MLEV-17 pulse sequence, and mixing times of 100, 300 and 500 ms were used for 2D ^1H - ^1H NOESY experiments (Bax, 1985). All data sets were collected in hyper-complex phase-sensitive mode. Before processing the t_2 dimension of DQF-COSY data sets were zero filled to 8K and the t_1 dimension of DQF-COSY data sets to 2K. For all other experiments, the linear prediction and zero filling to 4K data points in t_2 dimension were used during NMR data processing. When necessary, spectral

resolution was enhanced by Lorentzian-Gaussian apodization. Suppression of the solvent peak (HOD) was achieved using the WATERGATE pulse scheme during acquisition and the residual water peak (4.78 ppm) was used as reference for chemical shift assignments (Piotto, 1992). NMR data processing was done using VnmrJ3.2a (Agilent Technologies Inc., Palo Alto, CA) and analyzed using CCPNMR 2.4 and Sparky, graphical NMR Assignment and Integration programs (Vranken, 2005; Goddard, 2008). Side chain proton resonances were assigned by comparing cross peaks in the TOCSY spectra with those in the NOESY spectra acquired under similar experimental conditions. Amide proton resonances either appearing after 12 h or showing slow exchange in the 1D and 2D ^1H - ^1H TOCSY spectra acquired in 100% D₂O were considered for hydrogen bonding. For distance constraints determination, NOE cross peaks were classified as strong, medium, or weak intensity based on the observed number of contour lines.

2.2. Structure Calculations

NOESY spectra acquired with the different mixing times (100, 300 and 500 ms) were used to obtain inter-proton distance constraints. Initially, a mixing time of 300 ms was used for distance constraint calculations and the assigned peaks were also checked with the spectrum acquired with the 100 ms mixing time. In order to rule out the peaks due to spin diffusion, only NOE peaks present in the NOESY spectra acquired with 100 ms mixing time were considered. NOE-derived distance constraints were classified into three distance ranges (1.8–2.5, 1.8–3.5, and 1.8–5.0 Å) corresponding to strong, medium and weak NOE intensities. Upper bound distance constraint limits for methyl or non-stereo-specifically assigned methylene protons were corrected by adding 0.5 Å appropriately for center averaging (Wüthrich, 1986). Structure calculations were performed using CNS program version 1.3 that uses a combination of simulated annealing in torsional space

with refinement using molecular dynamics in a Cartesian space (Brünger, 1998). Resulting in calculated structures with minimal restraint violation (Stein, 1997). Force constants were scaled throughout the procedure using default parameters of CNS. The refinement protocol involved 100 ps restrained MD simulation at 300 K followed by energy minimization. A force constant of 50 kcal/mol/ \AA^2 for the NOE derived distances constrains and 100 kcal/mol/ \AA^2 for dihedral angles constraints, were imposed during molecular dynamics calculations. The force constant of NOE distances was increased to 100 kcal/mol/ \AA^2 during final energy minimization. All NMR restraints, including NOE, dihedral angles and hydrogen bond derived distance constraints were used to calculate over 200 structures (SRP1) or 400 structures (SRP2). The 20 lowest energy structures with no restraint violation were selected to obtain an ensemble of three-dimensional NMR structures. The average structure was used to represent 3D solution NMR in each case. The quality of calculated structure was analyzed using VMD and PROCHECK-NMR programs (Humphrey, 1996; Laskowski, 1996). Figures of the structures were generated using PyMOL software (Schrodinger, 2010).

2.3. Explicit Solvent Molecular Dynamic (MD) Simulations

The five lowest energy NMR structures of each peptide (SRP1 and SRP2) were used as the initial coordinates to initiate five independent MD simulations. Each of the five replicas was solvated in a cubic TIP3P explicit solvent water-box with an average volume of $\sim 55 \times 55 \times 55 \text{\AA}^3$ in CHARMM36 (Brooks, 2009; Brooks, 1983; Mark, 2001). The system was then charge balanced with either K^+ or Cl^- , if necessary. Each of the solvated SRP1/2 structures was then energy minimized and equilibrated for 100 ps in the CHARMM36 force field (Huang, 2013). Five 600-ns production simulations at standard temperature and pressure were then performed using NAMD,

together yielding a total of 3.0 μ s effective sampling for SRP1/2 production simulations (Phillips, 2005). Particle mesh Ewald was used for long-range electrostatic interactions and van der Waals interactions were smoothly switched off from 12 to 13 \AA (Darden, 1993). Hydrogen mass was increased to 3 a.m.u. through bound heavy atom mass repartition (Hopkins, 2015). The integration time-step was 4 fs, keeping all bonds to hydrogens atoms of fixed length with SHAKE (Ryckaert, 1977). Snapshots were saved at an interval of 10 ps for the duration of simulation time to be used for later analysis. *In silico* mutation of the residue at position twelve for each peptide was accomplished with sidechain only coordinate replacement of the target residue with the corresponding mutant residue using CHARMM36 (Huang, 2013). All backbone atom coordinates were conserved. Production simulation procedures were identical to each peptide as outlined above.

2.4.MD Analysis

All structural and clustering analysis were performed using the GROMACS program suite (Hess, 2008). For clustering, snapshots at 1ns intervals were included, giving rise to 3005 unique snapshots together from all five replicas for each SRP1 or SRP2 variant. GROMOS clustering was then employed using $C\alpha$ RMSD with a cut-off of 3 \AA (Daura, 1999). Theoretical NOE distances for SRP1 and SRP2 WT were back calculated for each of the eight most populated clusters, which presumably represent key conformational states of SRP1 and SRP2 WT. Alternatively, theoretical NOE distances for SRP1 and SRP2 were also calculated as a function of atomistic pair-wise distance using RMSdist in Gromacs. Back-calculated NOE-like distances were then compared to experimentally derived medium- and long-range NOE upper bounds involving NH, C_α H, C_β H, and C_γ H protons used in NMR structure calculations (see Section 2.2).

Using Databank of Secondary Structure Propensities (DSSP) version 2.2.1 definitions, the total helical (sum of α -, 3_{10} -, and π -helices) and β -strand (and β -bridge) propensities were then calculated on a per residue basis for the full 3005 snapshots for each unique SRP1 and SRP2 variant. Similarly, simulation average root-mean-square-fluctuation (RMSF) for each SRP1 or SRP2 variant was calculated by selecting C_{α} atom positions in conjunction with GROMACS RMSF utility. The average end-to-end distance was approximated using the pair-wise distance between the N-terminal amine N atom and C-terminal carboxyl C atom for all 3005 snapshots. Likewise, pair-wise direct and pseudo-salt-bridge contact propensity was measured with the salt-bridge plug-in (Leonardo Trabuco and Elizabeth Villa, Theoretical and Computational Biophysics Group, University of Illinois at Urbana-Champaign) for VMD with a heavy-atom cut-off distance of 4.25Å for both arginine and lysine residue side-chains and oxygen atoms with no angle constraints.

2.5. References

- Bax, A. D., & Davis, D. G. (1969). MLEV-17-based two-dimensional homonuclear magnetization transfer spectroscopy. *Journal of Magnetic Resonance*, 65(2), 355-360.
- Brooks, B. R., Brooks III, C. L., Mackerell Jr., A. D., Nilsson, L., Petrella, R. J., Roux, B., Won, Y., Archontis, G., Bartels, C., Boresch, S., (2009). CHARMM: the biomolecular simulation program. *Journal of computational chemistry*, 30 (10), 1545-1614.
- Brooks, B. R., Bruccoleri, R. E., Olafson, B. D., States, D. J., Swaminathan, S. A., Karplus, M. (1983) CHARMM: a program for macromolecular energy, minimization, and dynamics calculations. *Journal of computational chemistry*, 4 (2), 187-217.
- Brünger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J.S., Kuszewski, J., Nilges, M., Pannu, N.S., & Read, R. J. (1998). Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallographica Section D: Biological Crystallography*, 54(5), 905-921.
- Darden, T., York, D., Pedersen, L. (1993). Particle mesh Ewald: An $N \cdot \log(N)$ method for Ewald sums in large systems. *The Journal of chemical physics*, 98 (12), 10089-10092.
- Daura, X., Gademann, K., Jaun, B., Seebach, D., Van Gunsteren, W. F., Mark, A. E. (1999) Peptide folding: when simulation meets experiment. *Angewandte Chemie International Edition*, 38 (1-2), 236-240.
- Goddard, T. D., & Kneller, D. G. (2008). Sparky 3. University of California, San Francisco, USA 15 (20).

- Hess, B., Kutzner, C., Van Der Spoel, D., Lindahl, E. (2008). GROMACS 4: algorithms for highly efficient, load-balanced, and scalable molecular simulation. *Journal of chemical theory and computation*, 4 (3), 435-447.
- Hopkins, C. W., Le Grand, S., Walker, R. C., Roitberg, A. E. (2015). Long-Time-Step Molecular Dynamics through Hydrogen Mass Repartitioning. *Journal of Chemical Theory and Computation*, 11 (4), 1864-1874.
- Huang, J., MacKerell, A. D., Jr. (2013). CHARMM36 all-atom additive protein force field: validation based on comparison to NMR data. *Journal of computational chemistry*, 34 (25), 2135-2145.
- Humphrey, W., Dalke, A., & Schulten, K. (1996). VMD: visual molecular dynamics. *Journal of molecular graphics*, 14(1), 33-38.
- Kumar, A., Ernst, R. R., & Wüthrich, K. (1995). A two-dimensional nuclear Overhauser enhancement (2D NOE) experiment for the elucidation of complete proton-proton cross-relaxation networks in biological macromolecules. In *Nmr In Structural Biology: A Collection of Papers by Kurt Wüthrich* (pp. 103-108).
- Laskowski, R. A., Rullmann, J. A. C., MacArthur, M. W., Kaptein, R., & Thornton, J. M. (1996). AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR. *Journal of biomolecular NMR*, 8(4), 477-486.
- Mark, P., Nilsson, L. (2001). Structure and Dynamics of the TIP3P, SPC, and SPC/E Water Models at 298 K. *The Journal of Physical Chemistry A*, 105 (43), 9954-9960.

- Phillips, J. C., Braun, R., Wang, W., Gumbart, J., Tajkhorshid, E., Villa, E., Chipot, C., Skeel, R. D., Kale, L., Schulten, K. (2005). Scalable molecular dynamics with NAMD. *Journal of computational chemistry*, 26 (16), 1781-1802.
- Piotto, M., Saudek, V., & Sklenář, V. (1992). Gradient-tailored excitation for single-quantum NMR spectroscopy of aqueous solutions. *Journal of biomolecular NMR*, 2(6), 661-665.
- Rance, M., Sorensen, O. W., Bodenhausen, G., Wagner, G., Ernst, R. R., & Wuthrich, K. (2012). Improved Spectral Resolution in COSY H-1 NMR Spectra of Proteins Via Double Quantum Filtering. *Biochemical and biophysical research communications*, 425(3), 527-533.
- Ryckaert, J.-P., Ciccotti, G., Berendsen, H. J. (1977). Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *Journal of computational physics*, 23 (3), 327-341.
- Schrodinger, L. L. C. (2010). The PyMOL molecular graphics system. *Version*, 1(5), 0.
- Stein, E. G., Rice, L. M., & Brünger, A. T. (1997). Torsion-angle molecular dynamics as a new efficient tool for NMR structure calculation. *Journal of Magnetic Resonance*, 124(1), 154-164.
- Vranken, W. F., Boucher, W., Stevens, T. J., Fogh, R. H., Pajon, A., Llinas, M., Ulrich, E.L., Markley, J.L., Ionides, J., & Laue, E. D. (2005). The CCPN data model for NMR spectroscopy: development of a software pipeline. *Proteins: structure, function, and bioinformatics*, 59(4), 687-696.
- Wüthrich, K. (1986). NMR with proteins and nucleic acids. *Europhysics News*, 17(1), 11-13.

Chapter 3- Solution Structure and Expression Profile of an Insect

Cytokine: *Manduca sexta* Stress Response Peptide-2¹

Abstract

Manduca sexta stress response peptide-2 (SRP2) is predicted to be a 25-residue peptide (FGVKDGGKCPGRVRRLLGICVPDDDY), which may function as an insect cytokine to regulate immune responses. Produced as an inactive precursor, endogenous proSRP2 is probably converted to active SRP2 by limited proteolysis in response to invading pathogens, along with prophenoloxidase and pro-Spätzle activation. In addition to immunity, SRP2 may control head morphogenesis or other developmental processes in the lepidopteran insect. We have examined the profiles of SRP2 gene expression in terms of immune induction capacity, tissue specificity, and developmental changes. To gain insights into its functions, we chemically synthesized SRP2, injected the peptide solution into naïve larvae, and detected significant up-regulation of several antimicrobial peptide genes. We determined the 3D molecular structure in solution of SRP2 by two-dimensional ¹H-¹H NMR spectroscopy. SRP2 has an ordered structure, which is composed of two short β -strands at regions R12 - R15 and I18 - V20, one type-I' β -turn at region R15 - I18, and a half turn at region C8 - S10 in its well-defined core stabilized by a covalent disulfide bond between C8 and C19. The secondary and tertiary structures are further stabilized by hydrogen bonds. Possible relationships between the structure and function are also discussed.

¹This chapter was published in *Protein and Peptide Letters* (2017) 24(1), 3-11.

Keywords

Cytokine; 2D nuclear magnetic resonance; epidermal growth factor; hemolymph protein; insect immunity; prophenoloxidase

Abbreviations

Anti-microbial peptide, AMP; growth-blocking peptide, GBP; hemolymph (serine) protease, HP; phenoloxidase and its precursor, PO and proPO; proPO activating protease, PAP; paralytic peptide, PP; plasmatocyte spreading peptide, PSP; stress response peptide, SRP, upstream Glu-Asn-Phe factor, uENF; serine protease and its non-catalytic homolog, SP and SPH.

3.1.1. Introduction

Analogous to innate immunity in vertebrates, insects rely on humoral and cellular responses to protect them from microbial infection (Jiang, 2010; Lemaitre, 2007). In many moth species, proteolytic activation of plasmatocyte spreading peptides (PSPs, also known as GBPs for growth blocking peptides or PPs for paralytic peptides) from their precursors stimulates hemocyte encapsulation (Strand, 2008). These active peptides have a conserved N-terminus of E-N-F and, therefore, are also called ENF factors. While ENF peptides are only found in lepidopterans, their evolution goes beyond moths and butterflies. Homologous peptides have been isolated from hemolymph of insects in six different orders (Matsumoto, 2012; Duressa, 2015). GBP homologs also include: upstream ENF peptides (uENFs) and stress response peptides (SRPs) (Kanamori, 2010; Yamaguchi, 2012). The latter are identified in many insects including mosquitoes, moths, beetles and locusts. Since most of these peptides are activated during stress conditions such as wounding, infection, ligation, and heat/cold shock, SRP appears to be a proper general name for

these phylogenetically related peptides. In addition to blocking growth, antimicrobial peptide induction, as well as hemocyte spreading and chemotaxis, SRPs play roles in brain morphogenesis and thanatosis (Tsuzuki, 2012; Ishii, 2015; Nakatogawa, 2009; Tsuzuki, 2005; Kiyotake, 2014). One of the five *Drosophila* SRPs induces dipteracin and metchnikowin expression in an Imd-dependent manner (Tsuzuki, 2012). These peptides essentially fulfill the role of cytokines for these insects, partly based on their expression in the brain, midgut, fat body, and hemocytes. As with cytokines in vertebrates, many of these short peptides' expression varies in different developmental stages of the organisms. In contrast to the knowledge on SRPs' diverse physiological functions, mechanisms for generating active SRPs from their inactive precursors are poorly understood. Evidence suggests that hemolymph proteases are responsible for the specific proteolysis. A system of serine proteases (SPs) exists in plasma of each one of the studied insects to mediate immune responses (Kanost, 2015; Park, 2010; Veillard, 2016). An SP inhibitor benzamidine blocks the processing of *M. sexta* proPSP in the larval hemolymph (Wang, 1999). Prophenoloxidase (proPO) activation generates PO which catalyzes melanization. Since this defense response often accompanies hemocyte encapsulation and nodule formation, we postulate that the SP system activated upon recognition of pathogens is also involved in the SRP activation. With the genome sequence of *M. sexta* available, we identified 105 non-digestive SPs and most of them are predicted to have trypsin-like specificity (Cao, 2012). To begin to study SRPs in this biochemical model insect, we performed a genome search and identified members of the SRP gene family, including PSP, uENF1, uENF2, and SRPs. While we do not yet know their exact physiological functions in *M. sexta*, the diverse functions observed in their homologs distributed in a wide range of taxonomic groups of insects have led us to undertake structural-functional studies of SRP2, which is one of the highly expressed SRPs in *M. sexta*. The putative SRP2 consists

of 25 amino acid residues (F-G-V-K-D-G-K-C-P-S-G-R-V-R-RL-G-I-C-V-P-D-D-D-Y) with a single disulfide bond between C8-C19. Here we report the three-dimensional solution structure of SRP2 solved by two dimensional ^1H - ^1H NMR spectroscopy. We found that the *M. sexta* SRP2 core region contains two short β -strands stabilized by a type-I' β -turn and a disulfide bond. The overall secondary and tertiary structure is further stabilized by hydrogen bonds. We also discuss the expression profile of the SRP2 gene and its possible functions in immune signal transduction.

3.1.2. Materials and Methods

3.1.2.1 Identification of SRP Genes in the Genome of M. sexta

The MCOT1.0 models were matched with K/R-X1-15-C-X3-10-G-X1/2-C-X1-15, where X represents any amino acid residue other than C (Cao, 2015). This motif is modified from the GBP consensus sequence of C-X2-G-X4/6-G-X1/2-C-K/R by adding K/R in the front, since 68 of the 105 potential activating SPs are predicted to cleave after K/R residue (Matsumoto, 2012; Cao, 2015). The initial search results were manually examined to identify putative SRPs that are located in the carboxyl-terminus of secreted proteins shorter than 250 residues.

3.1.2.2. Insect Rearing and Injection, Total RNA Preparation, and cDNA

M. sexta eggs, ordered from Carolina Biological Supply (Burlington, NC), were hatched at ambient temperature and larvae were reared on an artificial diet (Dunn, 1983). The 5th instar larvae (day 2) were separately injected with a mixture of bacteria (2×10^7 Escherichia coli cells and

20 µg *Micrococcus luteus*) (Sigma-Aldrich) and 20 µg curdian (insoluble β-1,3-glucan from *Alcaligenes faecalis*) (Sigma-Aldrich) suspended in H₂O (30 µL per larva). Hemolymph samples were collected from cut prolegs of three larvae 24 h later and centrifuged at 5,000 g for 5 min to harvest hemocytes prior to dissection of fat body tissue. Total RNA was prepared from induced hemocytes and fat body using TRIZOL Reagent (Thermo Fisher Scientific). Control fat body and hemocyte total RNA samples were isolated from three 5th instar naïve larvae (day 3) by the same method. Samples for SRP2 mRNA expression in other tissues (i.e., integument, muscle, midgut, trachea, Malpighian tubule, nerve tissue, and salivary gland) were dissected from 5– 10 5th instar naïve larvae (day 3) for RNA preparation. To examine the developmental changes, fat body and hemocytes were prepared from larvae and pupae at different stages (4th instar to adult, 3–6 insects per stage) for total RNA isolation. These total RNA samples (1 µg each) were incubated with 1× iScript Reverse Transcription Supermix (Bio-Rad) in a 10 µL reaction at 42 °C for 30 min to synthesize cDNA. The reaction mixtures were heated at 95 °C for 3 min to denature the enzyme. To examine its possible role in the expression regulation of antimicrobial peptide (AMP) genes, 30 or 40 µL of phosphate-buffered saline (PBS, pH 7.4, 137 mM NaCl, 2.7 mM KCl, 10 mM Na₂HPO₄, 1.8 mM KH₂PO₄) or SRP2 (0.1 µg/ µL) in PBS was injected into each of six day 2, 5th instar larvae (3 control and 3 test). Fat body tissues were separately collected 6 h later for RNA isolation and cDNA synthesis.

3.1.2.3. Quantitative Polymerase Chain Reaction (qPCR)

The cDNA samples were incubated with 1× iTaq Universal SYBR Green Supermix (BioRad) along with specific primers (0.5 µM each) in triplicate. In each 10 µL reaction, each cDNA is equivalent to starting with 50 ng total RNA. The primers were: j037 (5' CATGATC-

CACTCCGGTGACC) and j038 (5' CGGGAGCAT-GATTTTGACCTTAA) for rpS3; j1825 (5' TATTCCTGATCGTGCTGGTC) and j1826 (5' ACAATACATTCCGCCAGAGC) for diapausin-4; j1070 (5' GCAGGCGACGACAAGAAC) and j1071 (5' ATGCGTGTTGGTAAGAGTAGC) for attacin-1, 7, 8 and 10; j1072 (5' CCGTGTTTTATTCTTCGTCTTC) and j1073 (5' AATCCTTTGACCTGCACCC) for cecropin-6; j1827 (5' GCTGTTGATCTGCGTGACAT) and j1828 (5' TCCTCCTTTGAATCCACGTC) for defensin-2; j1074 (5' GCAAGTCGGCAACAATGG) and j1075 (5' ACCCTGTCCTGTCAGTTTG) for gloverin; j1829 (5' CCTGTACTGCCCTCGATCAT) and j1830 (5' TTGTATCCCGGGTGAGTAGC) for lebocin D; j1076 (5' GTGTGCCTCGTGGAGAATG) and j1077 (5' ATGCCTTGGTGATGTCGTC) for lysozyme-1; j1078 (TGCTTTCTTTAACCTTTGTCCTC) and j1079 (TATTC-TAACACAGCCTATAATGCG) for moricin-1. The PCR was carried out under the following thermal cycling conditions: denaturation at 95 °C for 3 min and 40 cycles of 95 °C for 10 seconds and 60 °C for 30 seconds. After the reactions were complete on a CFX Connect Real-Time PCR Detection System (Bio-Rad), melt curves of the products (106, 167, 135, 103, 151, 124, 200, 187 and 186 bp) were measured to ensure proper shape and T_M values. Amplification efficiencies (E) were measured by amplifying a cDNA sample diluted to 10⁻¹, 10⁻², 10⁻³, 10⁻⁴ and 10⁻⁵ using the specific primer pairs under the same conditions, which were 93.8% for rpS3, 63.0% for diapausin-4, 105.3% for attacin-1, 7, 8 and 10, 90.0% for cecropin-6, 104.7% for defensin-2, 71.2% for gloverin, 88.1% for lebocin D, 121% for lysozyme-1, and 88.9% for moricin-1. Relative mRNA levels were calculated as: $(1 + ErpS3) Ct, rpS3 / (1 + Ex) Ct, x$ (Rieu, 2009).

3.1.2.4. Chemical Synthesis of Putative *M. sexta* SRP2

The 25-residue SRP2 were each prepared by stepwise solid-phase synthesis using 9-fluorenylmethoxycarbonyl (Fmoc) chemistry on an ABI model 431A automatic peptide synthesizer (Applied Biosystems; Foster City, CA). For each peptide synthesis CLEAR amide resin (0.3 mmol g⁻¹; Peptides International, Louisville, KY) and N α -F-moc amino acids (Anaspec Inc., San Jose, CA) were used. Following de-protection and cleavage, the linear peptides were cyclized by formation of a disulfide bond between C8 and C19. The peptides were each purified to >95% purity by HPLC (System Gold HPLC; Beck-man Instruments, Inc., Fullerton, CA) with a Phenomenex reversed-phase C-18 column (Torrance, CA), and eluted using a linear gradient of 10–90% acetonitrile containing 0.1% trifluoroacetic acid at 1 mL min⁻¹. HPLC-purified peptides were characterized by matrix-assisted-laser desorption time-of-flight (MALDI-TOF) mass spectrometric analysis using a Bruker Ultraflex II spectrometer (Bruker Daltronics, Billerica, MA). After characterization, peptide was lyophilized and stored as dry powder until use.

3.1.2.5. NMR Spectroscopy and Other Biophysical Analyses of SRP2

NMR spectroscopic studies, structure calculation, and biophysical analysis were carried out using protocols and procedures described in chapter 2.

uENF1	...TKPPDNSAG K	MIDVPLNHSAC CSQ <u>G</u> TYLDSS <u>G</u> <u>V</u> <u>C</u> RQPWQ*	28	23	100
uENF2	...ENVQAGQI IR	VPELE C <u>P</u> <u>L</u> <u>G</u> QRRDSL <u>G</u> <u>N</u> <u>C</u> RQRF*	22	22	72
PSP	...AETTTT K EG R	ENFAGG C AT <u>G</u> FLRTAD <u>G</u> <u>R</u> <u>C</u> KPTF*	23	23	130
SRP1	...KPDVPSFEN R	FGVRVGT C PS <u>G</u> YVRR <u>G</u> TF <u>C</u> FPDDDY*	25	16	113
SRP2	...THPIVTLED R	FGVKDGK C PS <u>G</u> RVRR <u>L</u> <u>G</u> <u>I</u> <u>C</u> VPDDDY*	25	16	113
SRP3	...QKTKVTFPN R	FLIKSSG C PK <u>G</u> YVKR <u>G</u> TF <u>C</u> FPDEDYDY*	27	20	107
SRP4	...NDDEKV R ER K	SLKRGT C PD <u>G</u> RVLF <u>G</u> RV <u>C</u> VEIDSLDY*	26	16	144
SRP5	...DPSTSQ R V K R	DI IKAP C PAG <u>G</u> KVKV <u>M</u> <u>G</u> <u>K</u> <u>C</u> TDDDPDYGK*	27	20	74
SRP6	...AQPLPEVAL R	NMIVVPPN C PP <u>G</u> QQMGSD <u>G</u> <u>V</u> <u>C</u> RVVFN*	26	15	77
SRP7	...PSSPLL R V K R	HIAGDPK C PK <u>G</u> QEKI <u>N</u> <u>G</u> <u>I</u> <u>C</u> TEKSDDY*	26	21	75
SRP8	...VGLPLAR V K R	ALTVVGY C PE <u>G</u> QKKI <u>N</u> <u>G</u> <u>I</u> <u>C</u> SETDSY*	26	21	73

Figure 3.1.1. Sequence Alignment of a Family of Proteins with Their Carboxyl-Termini Similar to SRPs. The last ten residues of the predicted pro-region and the putative SRPs at the C-terminal end are aligned. The predicted activation cleavage site is shown as a gap after the Arg/Lys (in bold), some of which are located in the conserved recognition sites (R/K-X-R/K/X-R/K) of intracellular processing proteases. The two invariable C residues (bold, underlined) may form a disulfide bond to stabilize the peptide structures. The two highly conserved G are in bold and double underlined. The D/E (bold) residues after the 2nd C in most SRPs may provide negative charges to the C-terminal end (*). The three numbers following each sequence denote the sizes of active peptide, signal peptide, and entire pre-pro-peptide. (Performed by Dr. Xiaolong Cao)

3.1.3. Results

3.1.3.1. Identification and Features of Putative SRPs of *M. sexta*

Using a modified GBP consensus sequence, we searched the MCOT1.0 assembly of 31,666 protein sequences, because this dataset consists of more gene models (18,089) than *M. sexta* Official Gene Set 2.0 does (15,451) (Matsumoto, 2012; Cao, 2015). After initial sequence matching, the hits were manually examined to ensure they are <250 residues long, with a signal peptide at the N-terminus, and a GBP-like sequence at the C-terminus. Eleven sequences satisfied the specifications (Figure 3.1.1.): uENF1, uENF2, and PSP precursors are encoded by three open reading frames located in a single transcript (data not shown). A shorter transcript also encodes the 130-residue PSP precursor as reported previously, indicating the existence of an alternative promoter (Wang, 1999). Eight other proteins are designated as precursors of SRP1 through SRP8, ranging from 72 to 144 residues long.

Following a signal peptide (15 to 23-residue), a pro-region (28–102 residues) and a carboxyl-terminal fragment (22–32 residues) were predicted in the mature proteins. While similarity was observed among some of these pro-segments, this seemed to just reflect the closer relationships of recently duplicated genes. Of the eleven pro-regions ending with R or K (Figure 3.1.1), PSP, SRP4, SRP5, SRP7 and SRP8 contain the recognition sequence of intracellular processing enzymes (R/K-X-Y-R/K), where Y is often R/K (Veenstra, 2000). This suggests that most of the five processed cytokine-like molecules are stored inside the cells and released in response to relevant physiological cues. In comparison to their pro-regions, the predicted C-terminal peptides are conserved in sequence and, perhaps, 2°/3° structure also. The eleven peptides all match the GBP consensus sequence of C-X₂-G-X_{4/6}-G-X_{1/2}-C-R/K, except that the positively charged R/K was replaced by T/S/V/F in SRP1-5, 7 and 8 (Figure 3.1.1). There are 5–10 residues before the first C. Of the 4–9 residues after the second C, 3 to 4 were negatively charged (i.e. D/E).

3.1.3.2. Immune Inducibility, Tissue Specificity, and Developmental Profiles of SRP2 Expression

To begin to study the functions of these peptides, we examined their mRNA levels in different tissues at various life stages represented by 52 RNA-Seq datasets and in hemocytes and fat body from naïve and induced larvae represented by 4 other datasets (Cao, 2015; Zhang, 2011). Based on the results, we selected SRP2 to confirm their expression patterns. Quantitative PCR analysis demonstrated that SRP2 mRNA levels were significantly higher ($p < 0.05$) in the induced fat body

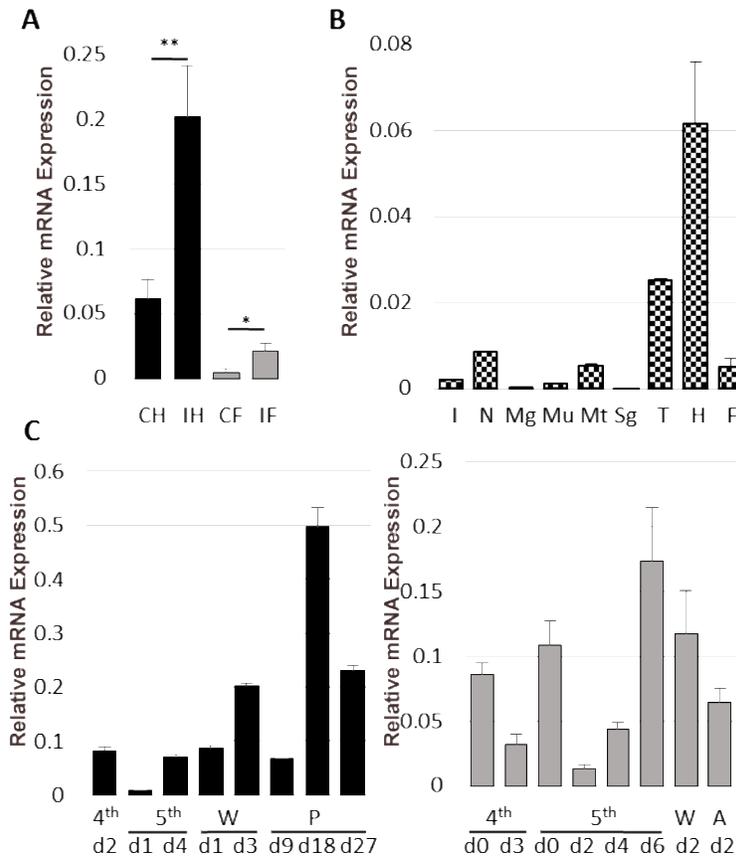


Figure 3.1.2. qRT-PCR Measurement of *M. sexta* SRP2 mRNA Levels. (A) **Immune inducibility.** As described in *Materials and methods*, the total RNA samples of control and induced hemocytes (CH and IH) and fat body (CF and IF) were reverse transcribed to cDNA and analyzed by quantitative real-time PCR using the SRP2-specific primers. The relative mRNA levels normalized with rpS3 based on their C_t values in three biological replicates (3 larvae per sample) and plotted as a bar graph (mean \pm SD, $n = 3$). The statistical significances were calculated using the Student's t-test and significant increases are shown as * ($p < 0.05$) and ** ($p < 0.01$). (B) Tissue specificity. RNA samples of integument (I), nerve tissue (N), midgut (Mg), muscle (Mu), Malpighian tubule (Mt), salivary gland (Sg), trachea (T), hemocytes (H), and fat body (F) from day 3, fifth instar naïve larvae were prepared and analyzed by qPCR under the same conditions. (C) Developmental profiles. The mRNA levels in hemocytes (*left*) and fat body (*right*) at different life stages were determined using the same method. W, wandering; P, pupal; A, adult. (Performed by Dr. Xiaolong Cao)

and hemocytes than their respective control samples (Figure 3.1.2.A). The mRNA levels in hemocytes were significantly higher than those in fat body. The higher expression in blood cells was confirmed when other tissues from day 3, 5th instar larvae were compared. The SRP2

transcripts in hemocytes were 2.5, 7.3, 11.5, 12.4-fold higher than those in trachea, nerve tissue, Malpighian tubules, and fat body (Figure 3.1.2.B), respectively. The mRNA levels in epithelial cells, midgut, muscles and salivary glands were negligible. We also determined the SRP2 expression in hemocytes and fat body at different life stages (Figure 3.1.2.C). In hemocytes, there was an up-regulation in the wandering stage before the larval pupal molt and another mRNA peak occurred in the middle to late pupal stage before the pupal-adult molt. In fat body, the transcript levels were higher in day 0, 4th and 5th instar larvae than day 3 larvae of the same stages. The SRP2 mRNA level in fat body peaked in the end of 5th instar feeding phase and slightly decreased in the wandering stage. A similar expression pattern was observed in some AMP genes that are upregulated in the wandering stage (He, 2015).

3.1.3.3 Induction of AMP Gene Expression by SRP2 Injection

To test whether SRP2 induces AMP synthesis, we chemically synthesized the 25-residue peptide and injected the peptide solution (3.0 µg) or PBS into day 2, 5th instar larvae. We analyzed the test and control RNA samples by qPCR and found five of the eight AMP genes were significantly induced ($p < 0.05$) at 6 h after the SRP2 injection (Figure 3.1.3). The cecropin-6, gloverin, lebocin D, lysozyme-1, and moricin-1 mRNA levels were 3.5, 3.1, 9.5, 4.2, and 14.0-fold higher than the control, respectively. While these increases were less dramatic than those caused by injection of a bacterial mixture it implicates SRP2's role in regulating the immune response in the tobacco hornworm (Zhang, 2011). This is consistent with the result from *Drosophila* and justifies a structural study of this cytokine (Tsuzuki, 2012).

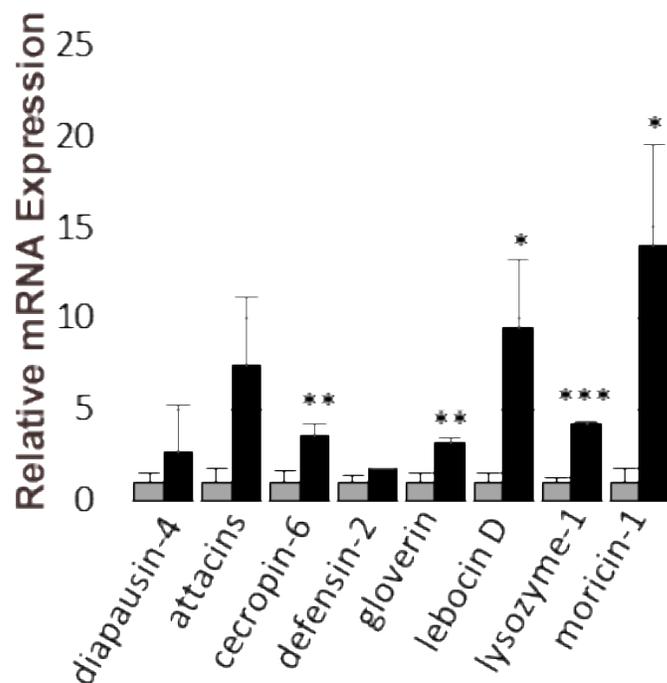


Figure 3.1.3. Induction of AMP Expression by Injection of PBS (grey) or SRP2 (black). As described in *Materials and methods*, the test and control total RNA samples of fat body from the larvae injected with buffer or peptide were analyzed by qPCR using the primers specific for eight AMP genes. After normalization against rpS3, the average SRP2 mRNA levels in the control samples were adjusted to one, and relative levels in the SRP2 injected group were calculated accordingly. The statistical significances were calculated using the Student's *t*-test and significant increases are shown as * ($p < 0.05$), ** ($p < 0.01$), or *** ($p < 0.001$). (Performed by Dr. Xiaolong Cao)

3.1.3.4. NMR Resonance Assignments and Secondary Structure

The amino acid spin system resonances and sequential proton chemical shift assignments for SRP2 were achieved through analyzing 2D ^1H - ^1H TOCSY NMR spectra for intra residue proton-proton correlations and 2D ^1H - ^1H NOESY spectra for inter residue proton-proton correlations as previously described (Bommineni, 2007; Herrera, 2010; Xiao, 2009). The TOCSY spectra acquired for SRP2 at 25 °C gave good chemical-shift dispersion with limited spectral overlap, enabling assignments of the spin systems for all residues and suggesting the presence of a well-defined secondary and tertiary structure for this peptide. In these assignments, $\text{C}_\alpha\text{H}(i) - \text{C}_\beta\text{H}(i+1)$:

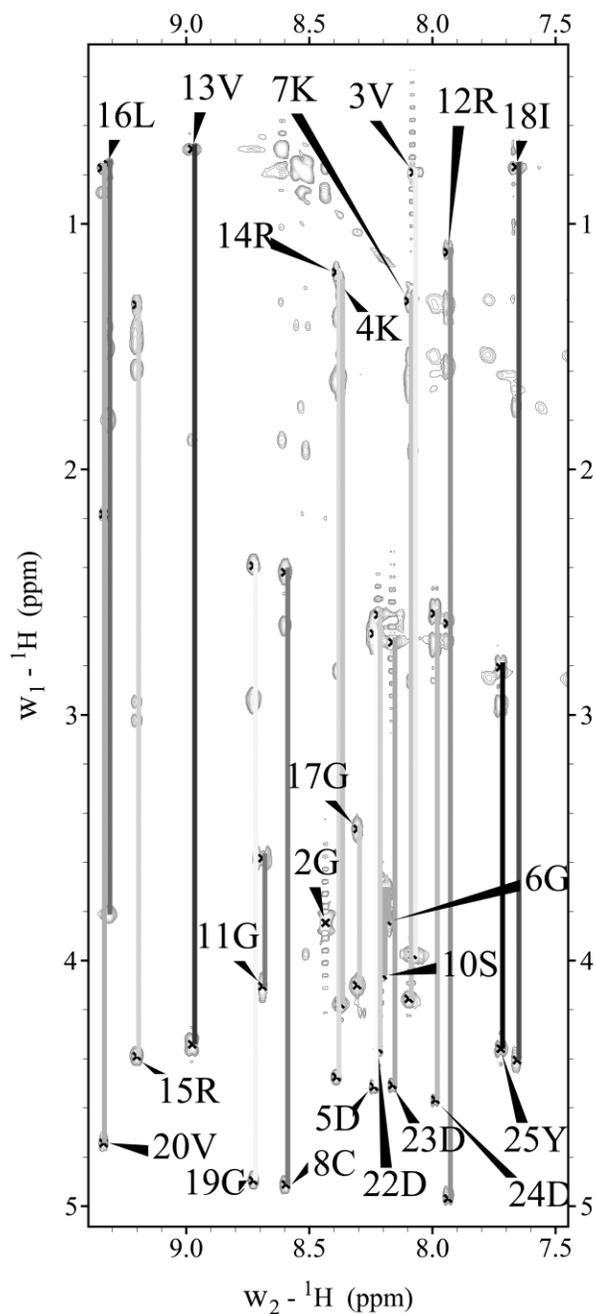


Figure 3.1.4. Fingerprint Region (NH-C α H/Side chain protons) of 2D ^1H - ^1H TOCSY Spectrum of SRP2 in H $_2$ O. Individual amino acid spin systems connected with vertical lines and labeled with residue name and number.

P) ($d_{\alpha\delta}$) or C α H (i) – C α H (i+1): P) ($d_{\alpha\alpha}$) NOEs instead of $\delta_{\delta\text{N}}$ (i, i+1) were used for P9 and P21.

Both prolines in SRP2 showed strong $d_{\alpha\delta}$ NOEs which were assigned to be trans configuration.

The presence of any Pro in cis configuration would not be detected due to absence of $d_{\alpha\alpha}$ NOE

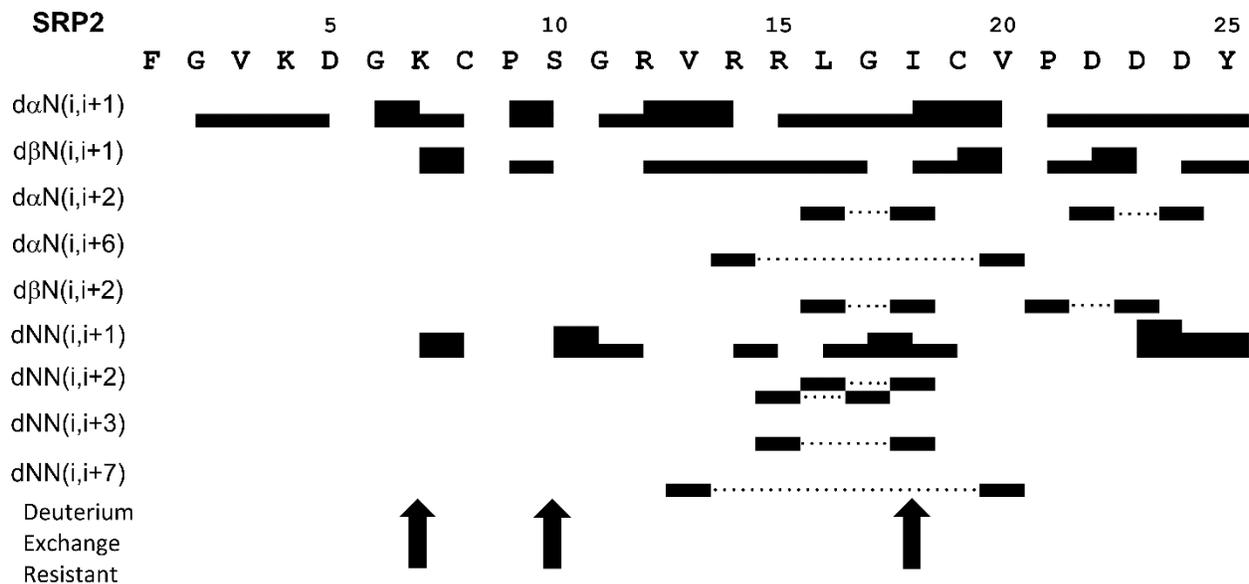


Figure 3.1.5. Summary of the NOE connectivity for SRP2 in water. The line thickness indicates the peak intensity, dotted lines indicate the non-sequential connections and arrows highlight residues that are resistant to deuterium exchange.

cross peaks in the 2D ^1H - ^1H NOESY spectra. The ^1H chemical shift assignments for *M. sexta* SRP2 have been deposited in BioMagResBank (<http://www.bmrb.wisc.edu>) under the accession number 26862. Figure 3.1.4. shows the NH- C_αH , NH- C_βH , NH- C_γH , and NH- C_δH cross peak assignments in the fingerprint region of the TOCSY spectrum used for residue identification. The 2D ^1H - ^1H NOESY spectra allowed the identification of $d_{\alpha N}(i, i+1)$ proton connectivity for sequential assignments and the disambiguation of repeated amino acid residues. Further analysis of the NOESY spectra obtained with different mixing times revealed short, medium and long-range connectivity including $d_{\alpha N}$, $d_{\beta N}$ and d_{NN} that provided structural information. An example of a NOESY spectrum showing d_{NN} connectivity is provided in Supporting Information (Figure 3.4.1.). All the sequential and medium range distance NOE connectivity as well as deuterium exchange resistant data for residues are illustrated in Figure 3.1.5. A number of non-sequential (weak $d_{\alpha N}(i, i+2)$ NOE between L16 and I18; weak $d_{NN}(i, i+3)$ NOE between R15 and I18) and long-range (weak $d_{\alpha N}(i, i+6)$ NOE between R14 and V20; weak $d_{NN}(i, i+7)$ NOE between V13

and V20) NOEs characteristic of twisted antiparallel β -sheets were observed for residues R12, V13, R14, R15, L16, I18, C19 and V20. This secondary structural observation was further supported by the down field shift of C_α proton resonances in chemical shift index ($\Delta C_\alpha H$) values (Figure 3.4.2.) and larger (9.5 -15 Hz) coupling constant ($3J_{NH-C_\alpha H}$) value for these residues. The random coil values for $C_\alpha H$ protons were taken from the published data (Wishart, 1994). The SRP2 structure consists two short β -strands at region R12-R15 (medium $d_{\alpha N}(i, i+1)$ NOEs from R12 to R14 and weak $d_{NN}(i, i+1)$ NOE between R14 and R15) and I18 – V20 (medium $d_{\alpha N}(i, i+1)$ NOEs from I18 to V20), a type-I' β -turn at region R15 – I18 (weak $d_{\alpha N}(i, i+1)$ NOE from L16 to G17 and medium $d_{NN}(i, i+1)$ NOE from G17 to I18) and a γ -turn at region C8 – S10 (carbonyl oxygen of i residue C8 forming a hydrogen bond with slowly exchanged amide proton of $i+2$ residue S10, a characteristic of γ -turn). The short stretch of twisted anti-parallel β -sheet was confirmed by observation of hydrogen bonds between V13 - V20 and R15 - I18. Low preponderance of NOE cross peaks for N-terminal residues F1 – K7 and C-terminal residues D22 – Y25 suggested extended conformation for these regions.

3.1.3.5. Tertiary Structure of *M. sexta* SRP2

For tertiary structure calculations, a total of 224 constraints involving 106 intra-residue, 62 sequential, 17 medium range, 9 long range NOE and 8 hydrogen bond distance constraints as well as 20 dihedral angle constraints were used. Two additional distance constraints were included between S and C_β atoms of C8 and C19 to produce normal disulfide bond geometry ($2.0 \text{ \AA} < d_{S,S} < 2.5 \text{ \AA}$, $2.2 \text{ \AA} < d_{C_\beta-S} < 3.5 \text{ \AA}$). The structural statistics of SRP2 in water are summarized in Table

Table 3.1.1. Structural Statistics of the 20 Lowest Energy Structures for SRP2.

NOE Constraints	Number
Total	224
Intraresidue	106
Sequential	62
Medium range (i-j ≤ 5)	17
Long range (i-j ≥ 5)	9
Disulfide bonds	2
Hydrogen bonds	8
Dihedral Constraints	20
Constraints/residue	8.96
Pairwise R.M.S.D. to Mean structure (residues 1-25)	
Backbone atoms (Å)	2.45±0.48
All nonhydrogen atoms (Å)	3.13±0.47
Pairwise R.M.S.D. to Mean structure (residues 8-21)	
Backbone atoms (Å)	0.87±0.21
All nonhydrogen atoms (Å)	1.88±0.62
Percentage of Residues in Regions of φ-ψ Space	
Allowed	72.1%
Additionally Allowed	23.2%
Generously Allowed	3.8%
Disallowed	0.9%

3.1.1. A simulated annealing molecular dynamics analysis was performed with these distance and dihedral constraints and found to a single family of conformers regardless of the starting structure. For these calculations, our starting structures were randomly chosen with extended conformation. From more than 400 conformers calculated, 20 structures with lowest energy were retained for further analysis. The torsion angle dynamics of all the 20 conformers were consistent with a pair of hydrogen bonds involving the backbone carbonyl of R15 and amide proton of I18. Evidence of

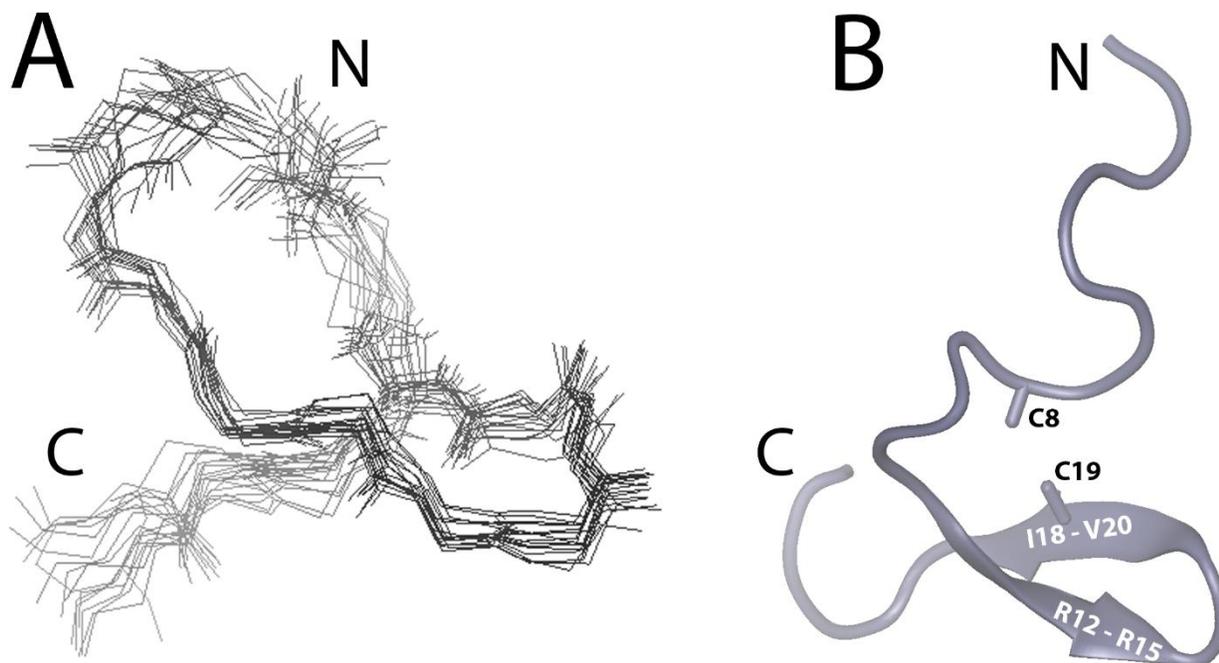


Figure 3.1.6. Structural Representations of SRP2. (A) Superimposition of the 20 lowest energy NMR structures of SRP2. (B) Ribbon diagram of the lowest energy structure of SRP2. The disulfide bond between C8 and C19 is indicated.

two additional hydrogen bonds (I18 backbone carbonyl to backbone amide proton of R15, and V20 backbone carbonyl to backbone amide proton of V13) were observed in a subset of the family of structures. The representative 20 superimposed structures (Figure 3.1.6.A) were in good agreement with experimental data with no distance violation larger than 0.3 Å and no dihedral angle violation larger than 5 degrees. A Ramachandran plot produced by ProcheckNMR showed that 72.1% of residues are in most favored stereo-chemically allowed regions and 23.2% are in additional allowed regions (Table 3.1.1) (Laskowski, 1996). A superimposition of the 20 lowest energy structures showed a considerable degree of flexibility, however, alignments along residues 8–21, forming the core region for SRP2, of the 20 lowest energy structures showed backbone root mean square deviation value of 0.87 Å. The schematic ribbon diagram of the energy minimized average tertiary structure of *M. sexta* SRP2 is shown in Figure 3.1.6.B. Three hydrogen bonds that

stabilized the tertiary structure of SRP2 were confirmed by the slowly exchanged amide protons from K7, S10, and I18.

3.1.4. Discussion

In this study, we have identified a family of eleven GBP-like genes in the genome of *M. sexta*, each coding for a protein with the pre-pro-structure. The signal peptide (i.e. prerregion) allows the pro-protein to be secreted into extracellular space or temporarily stored for later release through the secretory pathway, as these small proteins do not contain any transmembrane segment. In the former, the pro-proteins (precursors of uENF1, uENF2, PSP, SRP1-3, 6) are anticipated to be cleaved by extracellular SPs to become active. This mechanism is known for the proPSP activation (Wang, 1999). In the latter, the pro-proteins (precursors of SRP4, SRP5, SRP7 and SRP8) may be further processed by intracellular convertases (e.g. furin) to become active peptides. They are probably released in response to particular physiological or pathological stimuli because, otherwise, we must assume that these four proSRPs are constitutively processed and released as not regulated active cytokines. SRP2's intensity and unique mRNA expression pattern (Figure 3.1.2.) makes it an excellent candidate for function exploration of this protein family (Cao, 2015; Zhang, 2011). The qPCR results (Figure 3.1.2.) confirmed the data from the RNA-Seq analyses, suggesting that SRP2 takes part in antimicrobial defense of *M. sexta*. Fat body and hemocytes are major components of the immune system. After bacterial injection, the SRP2 transcript level increased 3.3-fold in hemocytes to 20.3% of rpS3's, which is highly abundant in cells. In the fat body, the 4.4-fold increase brought SRP2 transcript to 2.2% of the rpS3's level (Figure 3.1.2.A). As fat body has a much higher mass than hemocytes and acts as the major contributor of hemolymph proteome, it is unclear whether proSRP2 in plasma mainly comes from hemocytes or

not (He, 2016). The proSRP2 made in trachea may also participate in immune responses, whereas its production in nerve tissue suggests a possible role as a neuropeptide upon activation (Figure 2B). The upregulation of SRP2 expression in the fat body in the end of 5th instar feeding phase (Figure 3.1.2.C) resembles that of the AMP genes (He, 2015). Upregulation of AMP gene expression after SRP2 injection implicates an immunological role (Figure 3.1.3.). Based on these observations, we postulate that the SP system in the hemolymph is triggered by bacterial infection to activate proSRP2 along with other precursors of immune factors (e.g. PO, Spätzle, and PSP). Remarkably, the predicted cleavage site (R*F) of proSRP2 is identical to that of proPO1 and proPO2, suggesting that proPO activating proteases (PAP1-3) can also activate proSRP1-3 (Figure 3.1.1.). In *M. sexta*, PO-mediated melanization occurs along with proHP6 and proHP8 activation (An, 2010). Hemolymph protease-8 (HP8) generates active Spätzle-1 to induce AMP production via a putative Toll pathway (Cao, 2015; An, 2010). We propose SRP2, like Spätzle, may act as a cytokine to turn on another immune signaling pathways and induce AMP gene expression. In *Drosophila*, GBP (encoded by CG15917, one of the five SRP genes) induces an intracellular pathway via Imd to activate metchnikowin and diptericin (Tsuzuki, 2012). Instead of the classical Imd pathway including Relish, the GBP binding signal may go through the JNK branch. It would be interesting to identify proSRP2-activating protease(s), the SRP2 receptor, and downstream signal transducers in *M. sexta* using a biochemical approach. The analysis of the tertiary structure of SRP2 suggests that in aqueous solution this peptide possesses a well-defined structural core region with a short twisted anti-parallel β -sheet that is stabilized by a type-I' β -turn and disulfide bond while the N- and C-termini are largely unstructured and apparently with substantial flexibility. Overall tertiary structure is stabilized by a combination of the hydrogen bonds within the β -sheet structure and hydrophobic sidechain packing along with the covalent disulfide linkage. Within the

core region of the molecule, one side of the structure is mainly composed of disordered charged residues, including R12, R14, and R15. The side chain of R12 showed NOE contacts to the side chains of C8, P9, and C19. The opposite side of the structure is composed of hydrophobic residues, including V13, L16, I18 and V20. These structural characteristics are similar to *M. sexta* paralytic peptide structure indicate that the SRP2 structural core adopts an epidermal growth factor (EGF)-like fold very similar to the carboxyl-terminal region of the vertebrate EGF though they have very low sequence similarities (Yu, 1999). Furthermore, this structural similarity may suggest that SRP2 receptor interactions are analogous to EGF and its receptor to trigger the biological activities of this stress response peptide.

3.1.5. Conclusion

There is a family of cytokine-like genes in the genome of *M. sexta*, which encode eleven small proteins with the pre-pro-structure. Extracellular serine proteases and intracellular processing enzymes are likely responsible for their proteolytic activation and/or release in response to physiological stimuli. Expression profiling has shown the tissues and developmental stages in which SRP2 gene is expressed. More importantly, the SRP2 gene transcription is upregulated in hemocytes and fat body upon immune challenge and injection of the synthetic peptide significantly induced the expression of certain antimicrobial peptide genes. The latter suggest the existence of proSRP2 activating proteases, SRP2 receptor, and an immune signal transduction pathway. The elucidation of the SRP2 solution structure represents the first step in our endeavor aimed at understanding its interactions with a putative receptor in the future.

3.2. Unpublished Results of MD Simulation of SRP2

3.2.1. Introduction

Hydrophobic packing has long been established as a regulator of relative structural rigidity in proteins (Pace, 2011). Structural stability can play an important role in both binding kinetics and thermodynamics in the case of a protein or peptide that serves as a ligand. Subtle differences in internal dynamics of a peptide-based ligand can sometimes have a large impact on the biological activities, as changes in receptor(s) binding preference may result in varied biological effects down-stream. This may be especially important in the case of SRP2 in *M. sexta*. SRP2 and its highly conserved paralog, SRP1, are abundantly expressed in *M. sexta* but display rather distinct ontological and tissue specific mRNA expression profiles from one another (Figures 3.1.2, 4.1.1., and 4.3.1.). In addition, each peptide appears to exclusively induce mRNA expression of select AMP gene sets (Figures 3.1.3. and 4.1.2). Interestingly both peptides share 67.5% sequence identity, 75% conserved residues, yet display biologically distinct activities (Figure 3.1.1.). Between the two paralogs, only residue positions 5, 7, 12, 16, and 17 are non-conserved: Residues 5 and 7 are located within the unstructured N-terminus and are presumed to be non-participatory in receptor interactions, while residues 16 and 17 located within the β -turns of the respective SRPs and likely to participate directly in receptor interactions. The embedded residue at position 12 within *PiPSP*, *MsPP1*, *MsSRP1*, and *MsSRP2* is most intriguing. This embedded residue position is consistently conserved at the $i+1$ position corresponding to the first glycine within the core region of these peptides and are thought to be a relatively conserved structural feature for paralogs SRP3-5.

Although it is understood that the embedded residue cannot directly participate in receptor interactions, the structure-function mechanism by which this occurs remains unclear (Clark, 2001; Clark, 2004; Volkman, 1999; Yu, 1999). However, NMR analysis suggests that *M_s*SRP2, the only Arg embedded residue containing member analyzed, was reported to have the highest RMSD for its core region. Embedding of Arg as opposed to bulkier hydrophobic residues like Tyr or Phe may correlate lower structural stability and greater internal motion of the core region which in turn may inform receptor(s) binding preferences (thermodynamics and kinetics) that ultimately inform the distinct downstream AMP expression profiles of each peptide. To better understand the relevance of the embedded residue with respect to structure-activity relationship of certain SRPs, we undertook a series of *in silico* simulations of SRP2 and select embedded residue mutants to explore changes in conformation dynamics of the peptide in solution. The results represent the first atomistic molecular dynamic analysis of a stress-response peptide (SRP) in solution.

3.2.2. Results and Discussion

3.2.2.1. Generation of WT Simulation and NOE Back-calculation Analysis

In order to assess the impact of mutation *in silico* of the embedded residue at position 12 for SRP2, a preliminary set of WT simulations were prepared. The 5 lowest energy conformers derived from NMR structural studies were solvated and equilibrated in a CHARMM36 forcefield and in explicit solvent with TIP3P. The WT constructs were then subjected to 600ns of random walk molecular dynamics production simulation to enable robust conformational sampling of both

Table 3.2.1. Summary of NOE Violations for SRP2. The constraint list is derived from the original pair-wise constraint list employed for initial structure calculation in XPLOR format subdivided into long-range($i, j \geq i+5$) and medium-range ($i, i+2 \leq j \leq i+4$) lists. Constraint distance is the upper limit of the original constraint imposed for structure calculation. Simulation the weighted pair-wise distance observed from CHARMM36 simulation. Violation, if applicable, is the additional distance by which simulation exceeds the constraint distance.

Medium	Constraint Distance (Å)	Simulation (Å)	Violation (Å)
HG1VAL3-HNASP5	4.5	4.082	-
HG2VAL3-HNASP5	4.5	4.082	-
HG1VAL3-HNGLY6	4.5	5.867	1.367
HG2VAL3-HNGLY6	4.5	5.867	1.367
HG1VAL13-HNARG15	5	5.238	0.238
HNARG15-HNGLY17	4.5	4.36	-
HNLEU16-HNILE18	6	4.798	-
HALEU16-HNILE18	4.7	4.858	0.158
HB1LEU16-HNILE18	6	7	1
HNILE18-HG2VAL20	6	6.748	0.748
HBPRO21-HNASP23	4.5	4.877	0.377
HAASP22-HNASP24	6	4.8	-

LONG	Constraint Distance (Å)	Simulation (Å)	VIOLATION (Å)
HG1VAL13-HNCYS19	6	6.958	0.958
HG2VAL13-HNCYS19	6	6.958	0.958
HG1ARG12-HNVAL20	6	5.536	-
HG1ARG12-HNASP22	6	5.158	-
HG2ARG12-HNASP22	6	5.042	-
HD1ARG12-HNVAL20	6	5.768	-
HNVAL13-HNVAL20	4.5	4.202	-
HAARG14-HNVAL20	4.5	3.804	-
HBARG15-HNVAL20	6	6.135	0.135
HGARG14-HNVAL20	6	5.765	-

molecules. Validation of successful agreement between the NMR and MD simulation data was accomplished via NOE back-calculation using the weighted distance average observed for all replicas of simulation for SRP2. Then, pair-wise upper-bound distances imposed for medium- and long-range constraints in solution structure determination were compared with the weighted distance average values, see table 3.2.1. SRP2 displays a small number of violations. Overall

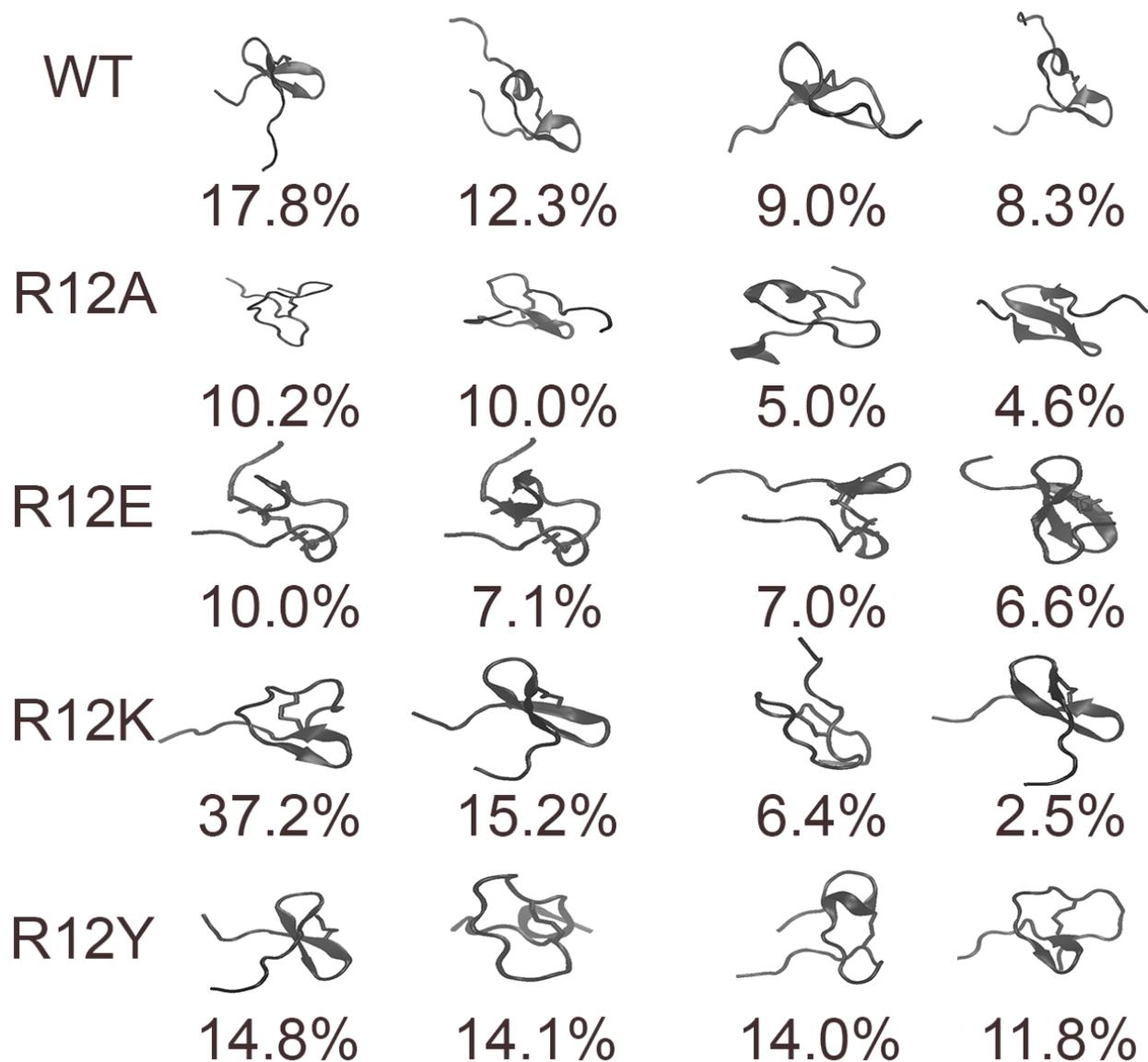


Figure 3.2.1. Cluster Heterogeneity Analysis of Simulated SRP2 and Mutants. Clustering was accomplished by the Gromos algorithm selecting for C_{α} coordinate positions with a 0.3nm cut-off for 3005 unique snapshots for WT and R12A/E/K/Y mutants. The largest four clusters for WT and mutants were selected and represented in cartoon. Percentages represent the relative cluster-size for that conformation relative to full simulation duration.

distance for each of the violations suggests relatively good agreement between simulation and imposed NOE constraints.

3.2.2.2. Impact of in silico Mutation of the Embedded Residue upon SRP2 Conformational Heterogeneity

As SRP2 displayed reasonable agreement in the CHARMM36 forcefield, it was concluded that the *in silico* system was satisfactory for exploring mutational analysis of SRP2 embedded residue positions. *In silico* sidechain coordinate only replacements were employed to create mutant facsimiles based upon the original 5 lowest energy conformers derived from NMR. Using this approach SRP2 R12A, R12E, R12K, and R12Y were generated and simulated under conditions analogous to their WT counterparts. Initial analysis of ensemble heterogeneity of all mutant variants against SRP2 WT was completed via visualization of preferred conformations and the corresponding propensities see figure 3.2.1. SRP2's four most populated clusters for each variant reveal that 47.4% of -WT, 29.7% of -R12A, 30.7% of -R12E, 61.3% of -R12K, and 54.7% of -R12Y simulations are represented. Both SRP2-R12A and -R12E mutants with their increase in ensemble heterogeneity as compared to SRP2-WT demonstrate low propensity of secondary structure. Conversely, SRP2 R12K and R12Y mutants display similar or less heterogeneity comparatively to SRP2-WT in simulation but fail to recapitulate WT ensemble conformers. This suggests that SRP2's R12 residue may have more complex role than simple stabilization of the molecule.

3.2.2.3. Differences in Structural Motility and Conformation of SRP2

To define and refine the structural impacts of the mutants upon native SRP structure, more quantitative approaches were utilized. Analysis of the per residue impact for each unique snapshot (3005 per SRP2 variant) for every SRP2 variant ensemble was analyzed using the Database of

Average Beta Strand Propensity per Residue: SRP2

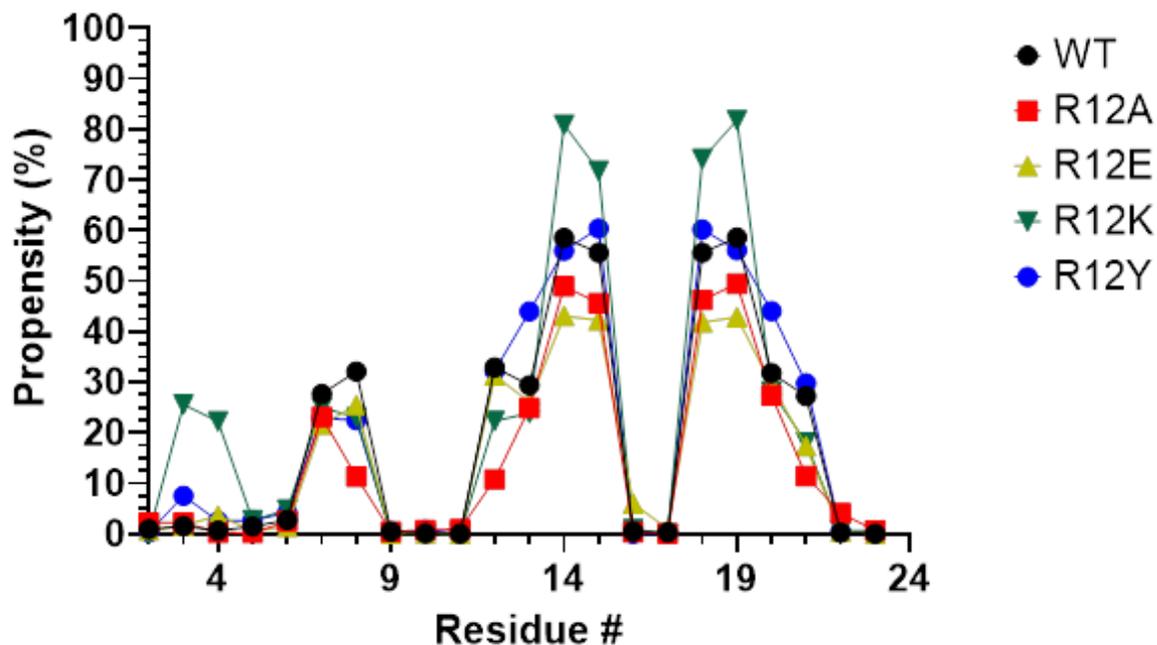


Figure 3.2.2. Average β -strand Propensity per Residue: SRP2. All 3005 snapshots of simulation for WT and R12A/E/K/Y mutants were categorized on a per residue by the Database of secondary structure prediction (DSSP2.2.1) defined by dihedral angles and hydrogen bond distance. The total propensity represents the percentage of the simulation in which a β -bridge, parallel β -strand, or anti-parallel β -strand was observed for residues 4 through 23.

Secondary Structural Prediction (DSSP) version 2.2.1 (Kabsch, 1983). Using these per frame definitions, the total secondary structural propensities for β -strand or helices were then calculated for all SRP2 variants, see figures 3.2.2 and 3.2.3. β -strand analysis of SRP2-WT displays a significant propensity for residues R12 through R15 and I18 through P21, consistent chemical shift values of β -strands for residues R12-R15 and I18-V20 observed from NMR. However, *in silico* SRP2-WT RMSD of $0.87 \pm 0.21\text{\AA}$ for residues 8-21 confers that SRP2's core region should be more mobile than a typical structure calculation. Average β -strand propensity is also consistent with heterogeneity analysis of SRP2 as R12K displays the greatest propensity specifically residues R14, R15, I18, and C19. SRP2-R12K and -R12A variants both display lower β -strand propensity within the core region, while SRP2-R12Y is roughly analogous to WT propensities. Expanding

Average Helical Propensity per Residue: SRP2

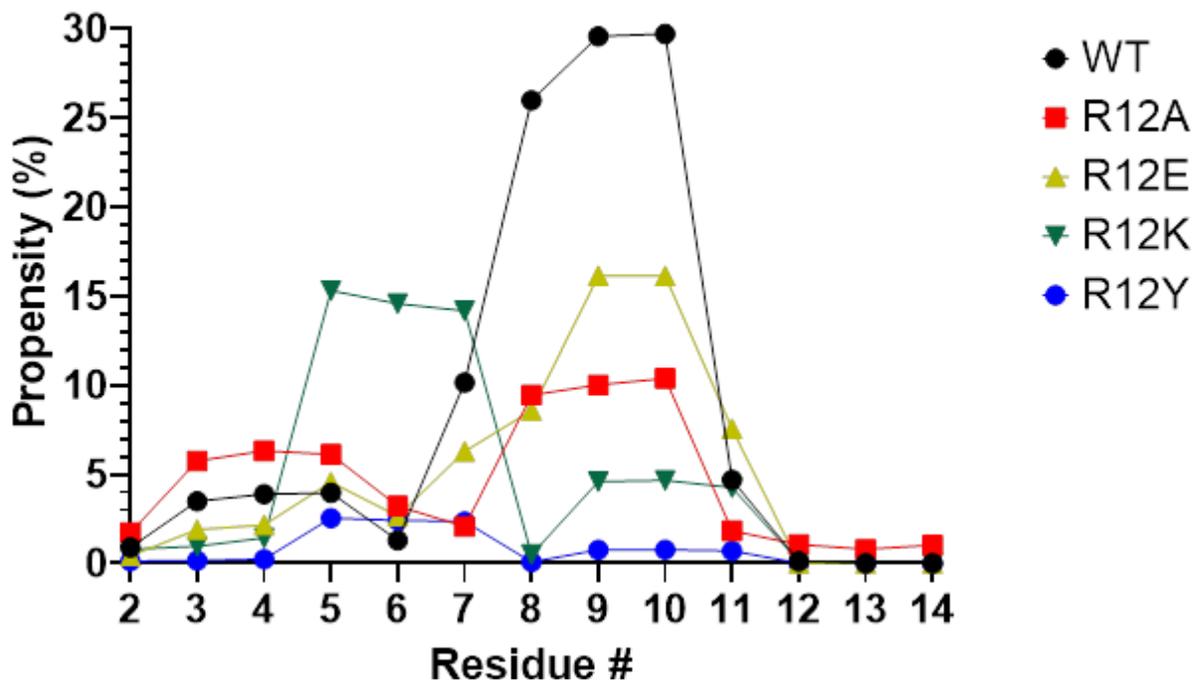


Figure 3.2.3. Average Helical Propensity per Residue: SRP2. All 3005 snapshots of simulation for WT and R12A/E/K/Y mutants were categorized on a per residue by the Database of secondary structure prediction (DSSP2.2.1) defined by dihedral angles and hydrogen bond distance. The total propensity represents the percentage of the simulation in which a 3-10-, α - or π -helix was observed.

secondary structural analysis to include residue positions K7 and C8 revealed the presence of a transient anti-parallel β -strand segment previously uncategorized via NMR studies, see figure 3.2.2. SRP2-WT displays the greatest propensity ~30% at these positions relative to all SRP2 mutants. Building upon this observation, it was hypothesized that perhaps other secondary structural features, such as helices, may have been overlooked from NMR due to the transience of their interaction. To verify the presence of helices, the per residue helical propensities for the N-terminal segment (residues 2-14) were then calculated and compared, as shown in figure 3.2.3. SRP2-WT shows a ~27.5% degree of helical propensity for residues C8, P9 and S10. All other SRP2 mutants did not conserve helical propensity and/or residue position as WT. This result is consistent with NMR observations as S10 was constrained as a possible γ' -turn based upon observations from

Average Root-mean-square-fluctuation per Residue: SRP2

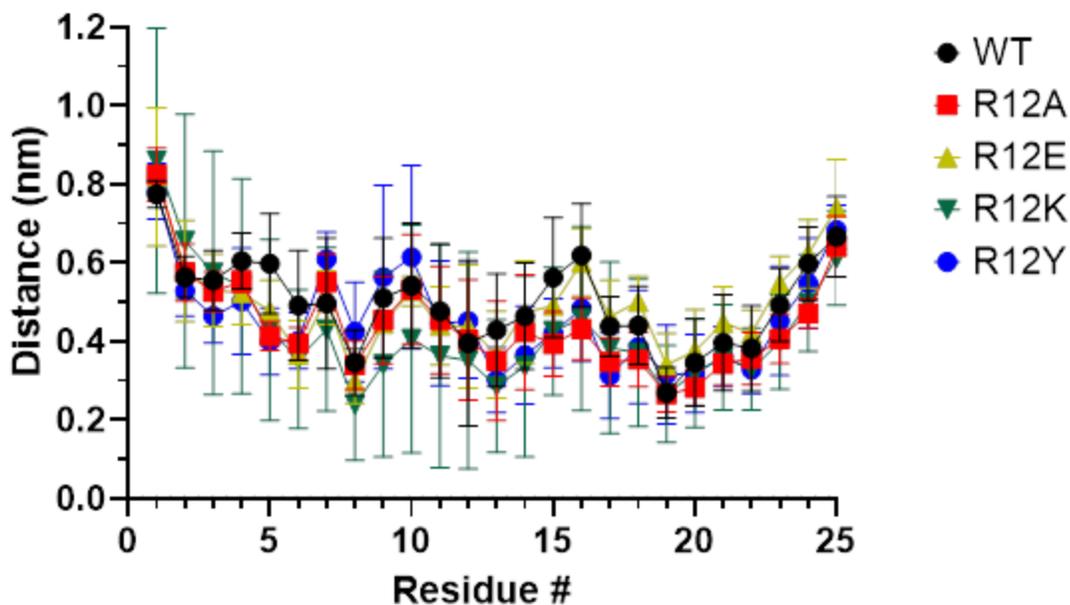


Figure 3.2.4. Average Root-mean-square-fluctuation per Residue: SRP2. RMSF was calculated by measuring variance in C_{α} coordinate position for all 601 snapshots per replica of WT and R12A/E/K/Y mutants. Per residue values for each variant are the average RMSF for all 5 replicas (3005snapshot) and the error bars reflects the standard deviation.

deuterium exchange and that it was simply mis-constrained. Taken together, secondary structure propensity would infer that there is substantive impact from embedded residue type upon preferred N-terminal structure of SRP2, which is reflected in heterogeneity analysis.

To ascertain whether secondary structure propensity is loosely correlated with increased residue motility, an analysis of the average root-mean-square-fluctuation (RMSF) per residue was employed for SRP2 and results are shown in figure 3.2.4. Interestingly SRP2 mutants show relatively similar RMSF values to the SRP2-WT across all residue types, suggesting that SRP2 mutants have similar overall rigidity to SRP2-WT.

Average End-to-End Distance: SRP2

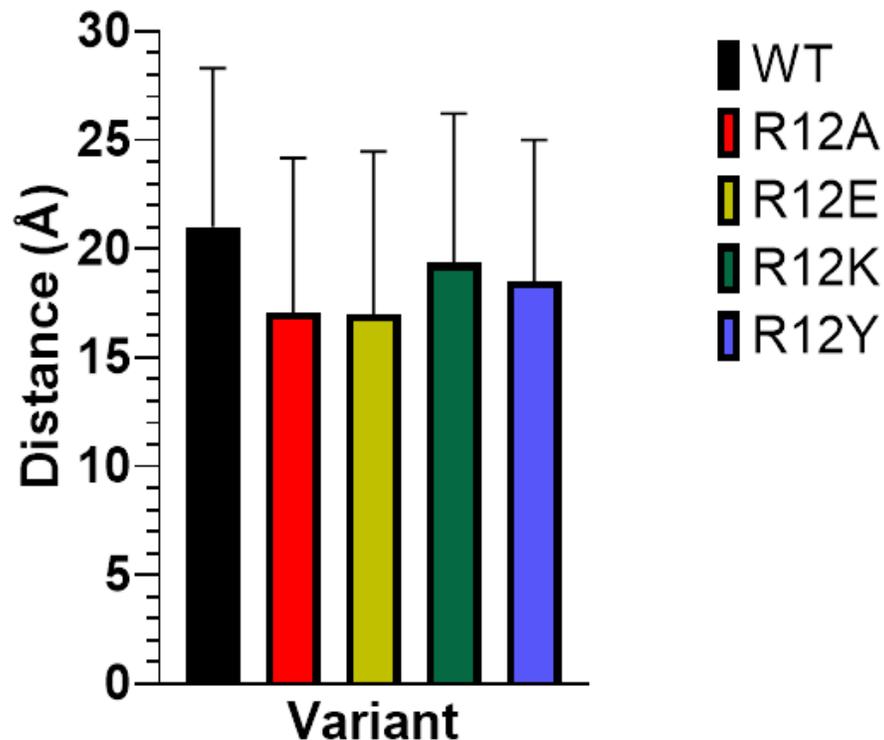


Figure 3.2.5. Average End-to-End Distance: SRP2. End-to-end was calculated by measuring pair-wise distance between n-terminal amine nitrogen and C-terminal carboxyl carbon coordinate positions for all 601 snapshots per replica of WT and R12A/E/K/Y mutants. The figure shows average End-to-End distance for all 5 replicas (3005 snapshots) and the error bars reflect the standard deviation measured on a per variant basis.

3.2.2.4. N-terminal Orientation of SRP2 is Influenced by Salt-bridges and Core region

Stability

Because of the relative location of residue K7 through G11 being at the beginning of the core region, it was hypothesized that the major impact of the residue embedding was not only for

Salt-bridge Propensity: SRP2

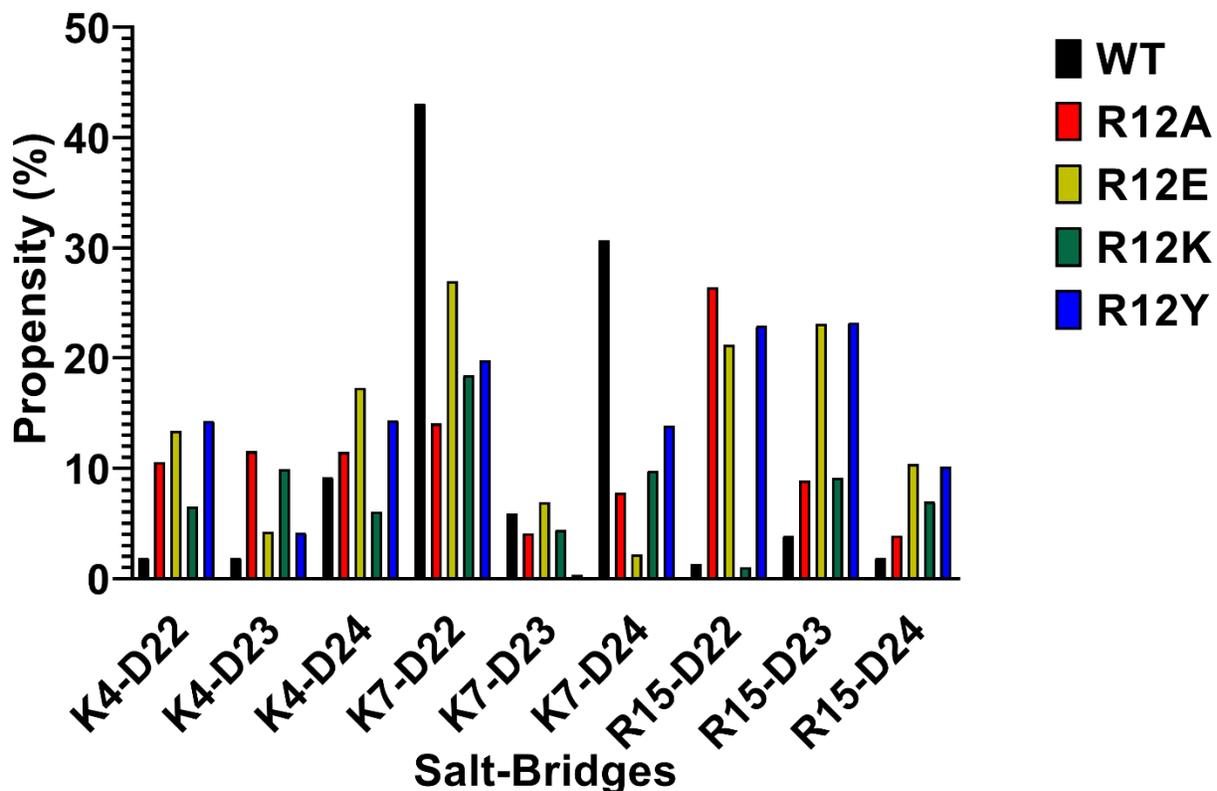


Figure 3.2.6. Salt-bridge Formation Propensity: SRP2. Propensity refers to the percentage of the simulation during which direct or pseudo salt-bridge contacts are observed for each SRP variant. Salt-bridge contacts are defined by residue contacts less than, or equal to 4.25Å. Heavy-atom distance between sidechains was measured on a pair-wise basis for possible contacts.

regulating core region stability, but for shepherding the orientation of the N-terminus relative to the C-terminus. Assessment of this global impact via average End-to-End distance analysis, as shown in figure 3.2.5., reveals that SRP2-WT displays the longest ($21.0 \pm 7.3\text{\AA}$) average end-to-end distance when compared to mutants. Further analysis of the end-to-end trajectories of SRP2 suggest that it makes little or no transient direct N- to C-terminus salt-bridge contacts and the two-domains should sparsely interact. However, SRP2 displays a series of well-defined distance regime transients suggesting sporadic restriction of the evolution of the molecule's N- and C-terminal conformations, results not shown. Because SRP2 contains a poly-aspartic C-termini and basic

residues within the N-terminus, an assay of both direct- and pseudo-contacts between salt-bridges was undertaken. Results of SRP2-WT salt-bridge analysis, as shown in figure 3.2.6., suggests that only WT makes contacts between K7-D22, and K7-D24 contacts while all other mutants fail to replicate these contact propensities. This observation indicates that R12 is the only residue type that facilitates this salt-bridge contact position. The presence of this contact position in SRP2 may provide an explanation of the observed secondary structure propensities for residues K7 through G11 as it likely has a cross-stabilizing effect. It is then possible that the presence of these secondary structures and/or N-terminal orientation may have significant implications upon selection of target receptor(s).

3.2.3. Conclusions

The results from *in silico* mutation studies on SRP2 demonstrate that NMR features of SRP2-WT are faithfully replicated *in silico* using CHARMM36. The subsequent heterogeneity, secondary structure analysis, and root-mean-square-fluctuation studies reveal that SRP2-WT is intermediate amongst its variants and flexible within the core region. However, SRP2 demonstrates a greater propensity for both helical and β -strand formation for residues K7 through G11. Salt-bridge analysis reveals SRP2 WT forms salt-bridges between K7 with D22 and D24, which cannot be replicated with any of its mutants. End-to-end distance analysis reveals that SRP2 WT is the longest among all SRP2 variants. As it is understood in the literature, the composition, orientation, and relative flexibility of the N-terminus is critical for imparting its biological activities. These results taken together suggest that relative structural dynamics of the core region as modulated by the embedded residue at position 12 in conjunction with N-terminal to -C-terminal salt-bridge contacts influence the effective end-to-end distance. These results might provide

important insights as to the differences in receptor(s) preference that result in the nuanced variation of biological expression of AMP mRNAs by SRP paralogs observed in *M. sexta*.

3.3. References

- An, C., & Kanost, M. R. (2010). Manduca sexta serpin-5 regulates prophenoloxidase activation and the Toll signaling pathway by inhibiting hemolymph proteinase HP6. *Insect biochemistry and molecular biology*, 40(9), 683-689.
- Bax, A. D., & Davis, D. G. (1985). MLEV-17-based two-dimensional homonuclear magnetization transfer spectroscopy. *Journal of Magnetic Resonance*, 65(2), 355-360.
- Bommineni, Y. R., Dai, H., Gong, Y. X., Soulages, J. L., Fernando, S. C., DeSilva, U., Prakash, O., & Zhang, G. (2007). Fowlicidin-3 is an α -helical cationic host defense peptide with potent antibacterial and lipopolysaccharide-neutralizing activities. *The FEBS journal*, 274(2), 418-428.
- Brünger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J.S., Kuszewski, J., Nilges, M., Pannu, N.S., & Read, R. J. (1998). Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallographica Section D: Biological Crystallography*, 54(5), 905-921.
- Cao, X., He, Y., Hu, Y., Zhang, X., Wang, Y., Zou, Z., Chen, Y., Blissard, G.W., Kanost, M.R., & Jiang, H. (2015). Sequence conservation, phylogenetic relationships, and expression profiles of nondigestive serine proteases and serine protease homologs in Manduca sexta. *Insect biochemistry and molecular biology*, 62, 51-63.
- Clark, K. D., Volkman, B. F., Thoetkiattikul, H., King, D., Hayakawa, Y., & Strand, M. R. (2001). Alanine-scanning mutagenesis of plasmatocyte spreading peptide identifies critical residues for biological activity. *Journal of Biological Chemistry*, 276(21), 18491-18496.

- Clark, K. D., Garczynski, S. F., Arora, A., Crim, J. W., & Strand, M. R. (2004). Specific residues in plasmatocyte-spreading peptide are required for receptor binding and functional antagonism of insect immune cells. *Journal of Biological Chemistry*, 279(32), 33246-33252.
- Dunn, P. E., & Drake, D. R. (1983). Fate of bacteria injected into naive and immunized larvae of the tobacco hornworm *Manduca sexta*. *Journal of Invertebrate Pathology*, 41(1), 77-85.
- Duressa, T. F., Vanlaer, R., & Huybrechts, R. (2015). Locust cellular defense against infections: sites of pathogen clearance and hemocyte proliferation. *Developmental & Comparative Immunology*, 48(1), 244-253.
- Goddard, T. D., & Kneller, D. G. (2008). Sparky 3. University of California, San Francisco, USA 15 (20).
- He, Y., Cao, X., Li, K., Hu, Y., Chen, Y. R., Blissard, G., Kanost, M.R., & Jiang, H. (2015). A genome-wide analysis of antimicrobial effector genes and their transcription patterns in *Manduca sexta*. *Insect biochemistry and molecular biology*, 62, 23-37.
- He, Y., Cao, X., Zhang, S., Rogers, J., Hartson, S., & Jiang, H. (2016). Changes in the plasma proteome of *Manduca sexta* larvae in relation to the transcriptome variations after an immune challenge: evidence for high molecular weight immune complex formation. *Molecular & Cellular Proteomics*, 15(4), 1176-1187.
- Herrera, A. I., Al-Rawi, A., Cook, G. A., Gao, J., Iwamoto, T., Prakash, O., Tomich, J.M., & Chen, J. (2010). Structural characterization of two pore-forming peptides: Consequences of introducing a C-terminal tryptophan. *Proteins: Structure, Function, and Bioinformatics*, 78(10), 2238-2250.

- Humphrey, W., Dalke, A., & Schulten, K. (1996). VMD: visual molecular dynamics. *Journal of molecular graphics*, 14(1), 33-38.
- Ishii, K., Hamamoto, H., & Sekimizu, K. (2015). Paralytic peptide: an insect cytokine that mediates innate immunity. *Archives of insect biochemistry and physiology*, 88(1), 18-30.
- Jiang, H., Vilcinskas, A., & Kanost, M. R. (2010). Immunity in lepidopteran insects. In *Invertebrate immunity* (pp. 181-204). Springer, Boston, MA.
- Kabsch, W., & Sander, C. (1983). Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers: Original Research on Biomolecules*, 22(12), 2577-2637.
- Kanamori, Y., Hayakawa, Y., Matsumoto, H., Yasukochi, Y., Shimura, S., Nakahara, Y., Kiuchi, M., & Kamimura, M. (2010). A eukaryotic (insect) tricistronic mRNA encodes three proteins selected by context-dependent scanning. *Journal of Biological Chemistry*, 285(47), 36933-36944.
- Kanost, M. R., & Jiang, H. (2015). Clip-domain serine proteases as immune factors in insect hemolymph. *Current opinion in insect science*, 11, 47-55.
- Kiyotake, H., Matsumoto, H., Nakayama, S., Sakai, M., Miyatake, T., Ryuda, M., & Hayakawa, Y. (2014). Gain of long tonic immobility behavioral trait causes the red flour beetle to reduce anti-stress capacity. *Journal of insect physiology*, 60, 92-97.
- Kumar, A., Ernst, R. R., & Wüthrich, K. (1995). A two-dimensional nuclear Overhauser enhancement (2D NOE) experiment for the elucidation of complete proton-proton cross-relaxation networks in biological macromolecules. In *Nmr In Structural Biology: A Collection of Papers by Kurt Wüthrich* (pp. 103-108).

- Lemaitre, B., & Hoffmann, J. (2007). The host defense of *Drosophila melanogaster*. *Annual Review of Immunology*, 25, 697-743.
- Laskowski, R. A., Rullmann, J. A. C., MacArthur, M. W., Kaptein, R., & Thornton, J. M. (1996). AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR. *Journal of biomolecular NMR*, 8(4), 477-486.
- Matsumoto, H., Tsuzuki, S., Date-Ito, A., Ohnishi, A., & Hayakawa, Y. (2012). Characteristics common to a cytokine family spanning five orders of insects. *Insect biochemistry and molecular biology*, 42(6), 446-454.
- Nakatogawa, S. I., Oda, Y., Kamiya, M., Kamijima, T., Aizawa, T., Clark, K. D., Demura, M., Kawano, K., Strand, M.R., & Hayakawa, Y. (2009). A novel peptide mediates aggregation and migration of hemocytes from an insect. *Current Biology*, 19(9), 779-785.
- Pace, C. N., Fu, H., Fryar, K. L., Landua, J., Trevino, S. R., Shirley, B. A., Hendricks, M.M., Iimura, S., Gajiwala, K., Scholtz, J.M., & Grimsley, G. R. (2011). Contribution of hydrophobic interactions to protein stability. *Journal of molecular biology*, 408(3), 514-528.
- Park, J. W., Kim, C. H., Rui, J., Park, K. H., Ryu, K. H., Chai, J. H., Hwang, H. O., Kurokawa, K., Ha, N. C., Soderhall, I., Soderhall, K., Lee, B. L. (2010). Beetle Immunity. In: Söderhäll K. (eds) *Invertebrate Immunity. Advances in Experimental Medicine and Biology*, vol 708. Springer, Boston, MA
- Piotto, M., Saudek, V., & Sklenář, V. (1992). Gradient-tailored excitation for single-quantum NMR spectroscopy of aqueous solutions. *Journal of biomolecular NMR*, 2(6), 661-665.

- Rance, M., Sorensen, O. W., Bodenhausen, G., Wagner, G., Ernst, R. R., & Wuthrich, K. (2012). Improved Spectral Resolution in COSY ¹H NMR Spectra of Protein Via Double Quantum Filtering. *Biochemical and biophysical research communications*, 425(3), 527-533.
- Rieu, I., & Powers, S. J. (2009). Real-time quantitative RT-PCR: design, calculations, and statistics. *The Plant Cell*, 21(4), 1031-1033.
- Schrodinger, L. L. C. (2010). The PyMOL molecular graphics system. *Version*, 1(5), 0.
- Stein, E. G., Rice, L. M., & Brünger, A. T. (1997). Torsion-angle molecular dynamics as a new efficient tool for NMR structure calculation. *Journal of Magnetic Resonance*, 124(1), 154-164.
- Strand, M. R. (2008). The insect cellular immune response. *Insect science*, 15(1), 1-14.
- Tsuzuki, S., Sekiguchi, S., Kamimura, M., Kiuchi, M., & Hayakawa, Y. (2005). A cytokine secreted from the suboesophageal body is essential for morphogenesis of the insect head. *Mechanisms of development*, 122(2), 189-197.
- Tsuzuki, S., Ochiai, M., Matsumoto, H., Kurata, S., Ohnishi, A., & Hayakawa, Y. (2012). Drosophila growth-blocking peptide-like factor mediates acute immune reactions during infectious and non-infectious stress. *Scientific reports*, 2, 210.
- Veenstra, J. A. (2000). Mono-and dibasic proteolytic cleavage sites in insect neuroendocrine peptide precursors. *Archives of Insect Biochemistry and Physiology: Published in Collaboration with the Entomological Society of America*, 43(2), 49-63.
- Veillard, F., Troxler, L., & Reichhart, J. M. (2016). Drosophila melanogaster clip-domain serine proteases: Structure, function and regulation. *Biochimie*, 122, 255-269.

- Volkman, B. F., Anderson, M. E., Clark, K. D., Hayakawa, Y., Strand, M. R., & Markley, J. L. (1999). Structure of the insect cytokine peptide plasmatocyte-spreading peptide 1 from *Pseudoplusia includens*. *Journal of Biological Chemistry*, 274(8), 4493-4496.
- Vranken, W. F., Boucher, W., Stevens, T. J., Fogh, R. H., Pajon, A., Llinas, M., Ulrich, E.L., Markley, J.L., Ionides, J., & Laue, E. D. (2005). The CCPN data model for NMR spectroscopy: development of a software pipeline. *Proteins: structure, function, and bioinformatics*, 59(4), 687-696.
- Wang, Y., Jiang, H., & Kanost, M. R. (1999). Biological activity of *Manduca sexta* paralytic and plasmatocyte spreading peptide and primary structure of its hemolymph precursor. *Insect biochemistry and molecular biology*, 29(12), 1075-1086.
- Wishart, D. S., & Sykes, B. D. (1994). Chemical shifts as a tool for structure determination. In *Methods in enzymology*, 239, 363-392.
- Wüthrich, K. (1986). NMR with proteins and nucleic acids. *Europhysics News*, 17(1), 11-13.
- Xiao, Y., Herrera, A. I., Bommineni, Y. R., Soulages, J. L., Prakash, O., & Zhang, G. (2009). The central kink region of fowlicidin-2, an α -helical host defense peptide, is critically involved in bacterial killing and endotoxin neutralization. *Journal of Innate Immunity*, 1(3), 268-280.
- Yamaguchi, K., Matsumoto, H., Ochiai, M., Tsuzuki, S., & Hayakawa, Y. (2012). Enhanced expression of stress-responsive cytokine-like gene retards insect larval growth. *Insect biochemistry and molecular biology*, 42(3), 183-192.

- Yang, F., Wang, Y., He, Y., & Jiang, H. (2016). In search of a function of *Manduca sexta* hemolymph protease-1 in the innate immune system. *Insect biochemistry and molecular biology*, *76*, 1-10.
- Yu, X. Q., Prakash, O., & Kanost, M. R. (1999). Structure of a paralytic peptide from an insect, *Manduca sexta*. *The Journal of peptide research*, *54*(3), 256-261.
- Zhang, S., Gunaratna, R. T., Zhang, X., Najjar, F., Wang, Y., Roe, B., & Jiang, H. (2011). Pyrosequencing-based expression profiling and identification of differentially regulated genes from *Manduca sexta*, a lepidopteran model insect. *Insect biochemistry and molecular biology*, *41*(9), 733-746.

3.4. Supplementary Material

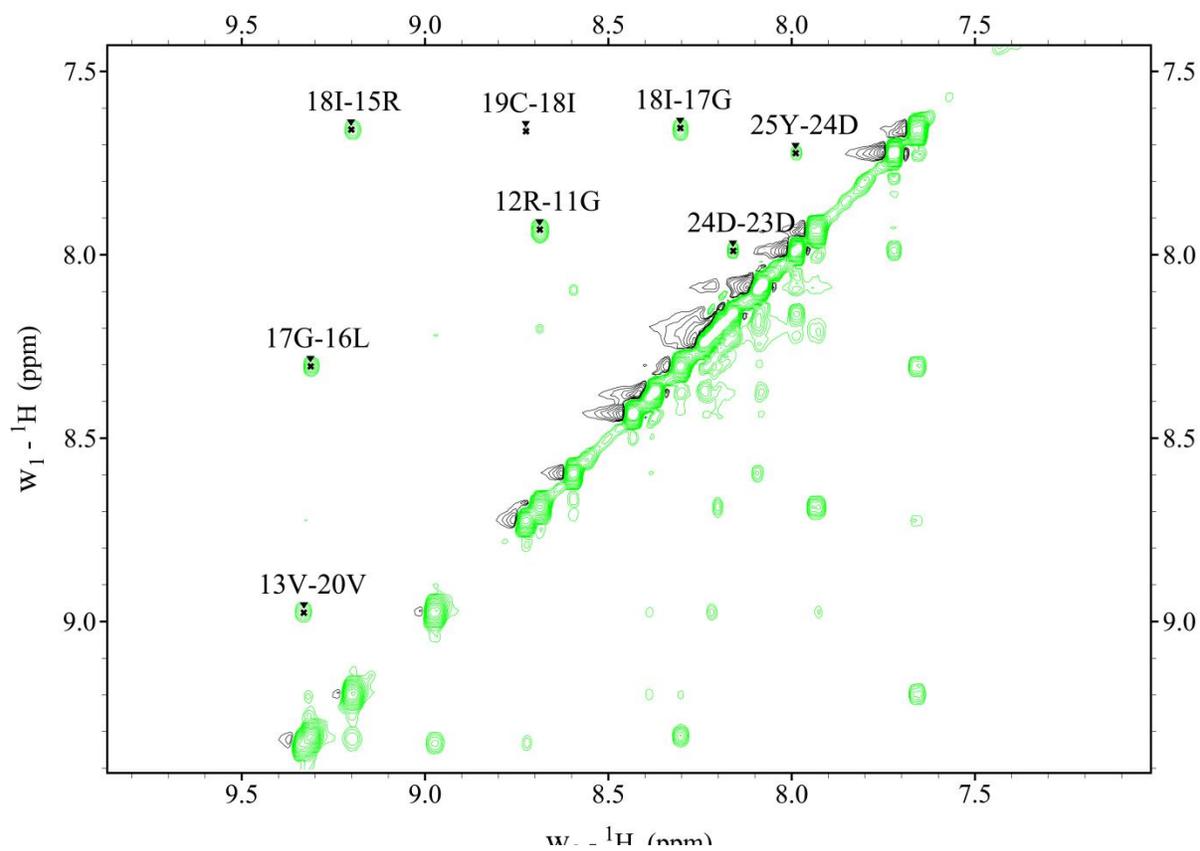


Figure 3.4.1. Fingerprint Region (NH-NH) of 2D ${}^1\text{H}$ - ${}^1\text{H}$ NOESY Spectrum of SRP2 in Water, showing downfield connectivities.

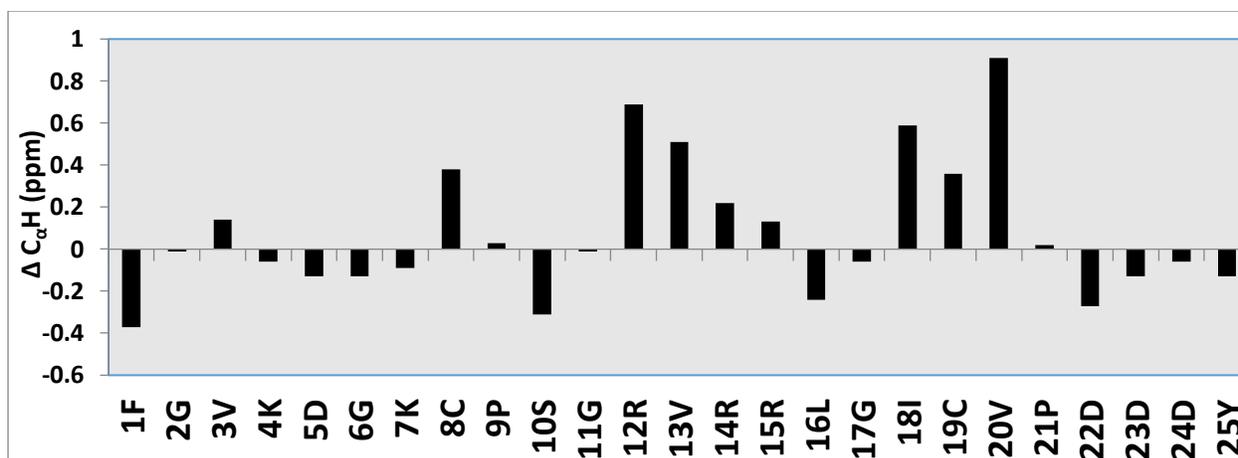


Figure 3.4.2. The C_{α} protons chemical shift deviation ($\Delta C_{\alpha}H$) for SRP2 as compared with random coil values. The down field shifts are characteristics of β -strand conformation.

Chapter 4- Expression and Characterization of *Manduca sexta*

Stress Response Peptide-1; An Inducer of Antimicrobial Peptide

Synthesis¹

Abstract

In response to stress conditions such as wounding or infections in insects, several short peptides are processed to act as cytokines that induce AMP gene expression. To study their structure-activity relationship, immune inducibility, tissue specificity, stress responsiveness, and development relatedness, we chemically synthesized *M. sexta* stress response peptide-1, a 25-residue peptide with one disulfide bond (SRP1: FGVRVGTCPSTGYVRRGTFCLPDDDY). Upon injection of the SRP1 into naïve larvae, several antimicrobial peptide genes were expressed at higher levels. The mRNA levels of SRP1 increased significantly in hemocytes and fat body after larvae were challenged with a mixture of bacteria and β -1,3-glucan. The expression patterns of SRP1 and its target genes are somewhat different from SRP2's, suggesting overlapping yet distinct functions. We elucidated the 3D structure of SRP1 in solution by two-dimensional ¹H-¹H NMR spectroscopy. The tertiary structure of SRP1 consists of two short β -strands at Y12–R15 and F18–F20, one type-II β -turn at R15–F18 in its well-defined core and is stabilized by a covalent disulfide bond between C8 and C19. The conformational ensemble of SRP1 from extensive atomistic simulation in explicit solvent (with 3.0 μ s total effective sampling) shows high consistency with experimental intramolecular NOEs of the core region. The SRP1 core adopts a fold similar to the carboxyl-terminal subdomain of epidermal growth factor (EGF),

¹This chapter was published in *Journal of Biochemistry and Molecular Biology* (2019) 4(3), 42-52.

suggesting that SRP1 may interact with EGF receptor-like molecules to trigger its biological function.

Keywords: Cytokine, Nuclear Magnetic Resonance, Molecular Dynamics, Epidermal Growth Factor, Hemolymph Protein, Insect Immunity, Phenoloxidase.

4.1. Introduction

Insects possess an effective innate immune system to fight against invading pathogens (Jiang, 2010; Strand, 2008). It is comprised of responses mediated by plasma factors made in fat body (pathogen recognition, melanization, antimicrobial peptides) and by blood cells known as hemocytes (phagocytosis, encapsulation). Extracellular serine protease pathways coordinate some of the responses, such as proteolytic activation of phenoloxidase, Spätzle, and plasmatocyte spreading peptide (PSP) precursors (Kanost, 2015). Phenoloxidase catalyzes the formation of reactive intermediates and melanin to kill and sequester pathogens, whereas Spätzle and PSP trigger the Toll pathway for antimicrobial peptide synthesis and cellular encapsulation, respectively (Strand, 2008; Lemaitre, 2007). In the holometabolous insects whose genomes are known, Spätzle and PSP homologs form two families of cytokines with as many as 7 and 11 members (Cao, 2015; Schrag, 2017a). We are just beginning to understand the functions of these related peptides.

Because PSP and its homologs are mostly activated under stress conditions such as wounding, infection, ligation, and heat or cold shock, we chose stress response peptide (SRP) as a general term to describe these structurally related molecules (Schrag, 2017b). These peptides

stimulate hemocyte spreading and chemotaxis, block larval growth, induce antimicrobial peptide (AMP) synthesis, and regulate other physiological processes. In the tobacco hornworm *M. sexta*, two paralytic peptides [PP1 and PP2: ENFAGGCATG(Y/F)LRTADGRCKPTF] were isolated from hemolymph collected through a cut proleg (Skinner, 1991). Injection of chemically synthesized PP1 into larval hemocoel induced rapid paralysis, reduced bleeding, and stimulated plasmatocyte spreading and aggregation (Wang, 1999). PP1 has an ordered structure comprised of a short antiparallel β -sheet at Y11–T14 and R18–P21, three β -turns at F3–G6, A8–Y11 and T14–G17, and a half turn at K20–F23 (Yu, 1999). Genome analysis shows a single PP/PSP gene and ten homologs (uENF1, uENF2, SRP1–8) in *M. sexta* (Schrag, 2017a). All of these genes encode predicted pre-pro-peptides, whose maturation is expected to require proteolytic cleavages. SRP2, which stimulates AMP expression, adopts a structure similar to PP1, and is composed of two short β -strands at R12–R15 and I18–V20, one type-I' β -turn at R15–I18, and a half turn at C8–S10 (Schrag, 2017a). The core structures of PP1 and SRP2 are each stabilized by a disulfide bond. To better understand structure-function relationships of this family of cytokines, we investigated the immune inducibility, stress responsiveness, expression pattern of *M. sexta* SRP1 and carried out a two-dimensional ^1H - ^1H NMR to examine its solution structure. We observed that SRP1 has an ordered structure, including a short twisted anti-parallel β -sheet, a type-II β -turn, and is stabilized by one disulfide bond.

4.2. Materials and Methods

4.2.1. Rearing of Insects and Injection, Total RNA Preparation, and cDNA Synthesis

M. sexta eggs were purchased from Carolina Biological Supply (Burlington, NC) and larvae were reared on an artificial diet at 25°C (Dunn, 1983). Day 2, 5th instar larvae were individually injected with a mixture of killed bacteria (2×10^7 *Escherichia coli* XL1-Blue and 20µg *Micrococcus luteus*, Sigma-Aldrich) and 20µg insoluble β-1,3-glucan from *Alcaligenes faecalis* (Sigma-Aldrich) suspended in water (30 µL per larva). Hemolymph was collected from cut prolegs of larvae (3 per group) after 24h and centrifuged at 5,000g for 5 min to harvest hemocytes. Fat body was then dissected from the bled larvae. Total RNA was prepared from pooled hemocytes and fat body using TRIZOL Reagent (Thermo Fisher Scientific) and was conducted in triplicate. Control hemocyte and fat body total RNA samples were prepared by the same method from day 3, 5th instar naïve larvae (3 per group, 3 groups as biological replicates).

For tissue specific expression analysis, tissues were dissected from 5–10 day 3, 5th instar naïve larvae (3–6 per group, 0 biological replicate) for RNA isolation. To study SRP1 expression after heat stress, day 2, 5th instar larvae (3 per group, 3 biological replicates) were incubated at 42°C for 1 h and then switched back to 25°C. Hemocytes, fat body and midgut tissues were collected 6 h later for total RNA preparation. The control insects (3 per group, 3 biological replicates) were day 2 naïve larvae continuously cultured at 25°C prior to hemocyte collection and tissue dissection. To examine developmental changes of SRP1 expression, hemocytes, fat body and midgut tissues were dissected from insects at stages ranging from day 2, 5th instar larvae to day 1, pupae (3–6 insects per stage, 0 biological replicate) for total RNA preparation. Each total RNA sample (1 µg each) was incubated with 1× iScript Reverse Transcription Supermix (Bio-

Rad) in a 10 μ L reaction at 42°C for 30min to synthesize cDNA. The reaction mixtures were then heated at 95°C for 3min to denature the enzyme.

To examine possible role of SRP1 in AMP expression regulation, 40 μ L of PBS (pH 7.4, 137mM NaCl, 2.7mM KCl, 10mM Na₂HPO₄, 1.8mM KH₂PO₄) or SRP1 (0.1 μ g/ μ L) in PBS were each injected into six day 2, 5th instar larvae (3 control and 3 test). Fat body tissues were separately collected 6h later for RNA isolation and cDNA synthesis as described previously. This experiment was repeated twice (total of 3 biological replicates).

4.2.2. Quantitative Real-Time Polymerase Chain Reaction (qRT-PCR)

To measure the mRNA levels of SRP1, rpS3, diapausin-4, attacin-1, 7, 8 and 10, cecropin-6, gloverin, lebocin D, lysozyme-1, and moricin-1 transcripts, qRT-PCR (3 technical repeats) was carried out under the conditions described previously [6]. The primers for SRP1 were: j1066 (5' GCCGAGGGTATCGTT) and j1067 (5' TCAGGCTTTGGCGTT). Amplification efficiencies (E) were 88.7% for SRP1, 95.8% for rpS3, 93.2% for diapausin-4, 97.9% for attacin-1, 7, 8 and 10, 83.9% for cecropin-6, 91.8% for gloverin, 90.7% for lebocin D, 102.0% for lysozyme-1, and 98.5% for moricin-1. mRNA levels relative to rpS3 were calculated as: $(1 + E_{\text{rpS3}})^{\text{Ct, rpS3}} / (1 + E_x)^{\text{Ct}}$ x (Rieu, 2009).

4.2.3. Synthesis of M. sexta Stress Response Peptide SRP1

M. sexta SRP1 was synthesized by automated protocols of stepwise solid-phase Fmoc chemistry on an automatic peptide synthesizer ABI model 431A (Applied Biosystems, Foster City, CA, USA). For peptide synthesis, CLEAR amide resin (0.3mmol/g; Peptides International,

Louisville, KY) and N^α-Fmoc amino acids (Anaspec Inc., San Jose, CA) were used. The linear peptides were cyclized by formation of a disulfide bond between C8 and C19 followed by purification on a high-pressure liquid chromatography (HPLC) system (Beckman Instruments, Inc., Fullerton, CA) using a Phenomenex reversed-phase C-18 column (Torrance, CA). The purified (>95%) peptide was eluted (1ml/min) using a linear gradient of 10-90% acetonitrile containing 0.1% trifluoroacetic acid. The characterization and mass determination of the peptide was performed by matrix-assisted-laser desorption time-of-flight (MALDI-TOF) mass spectrometric analysis (Bruker Ultraflex II spectrometer; Bruker Daltronics, Billerica, MA).

4.2.4. NMR Spectroscopy and Other Biophysical Analyses of SRP1

NMR spectroscopic studies and structure calculations were carried out using protocols and procedures described in chapter 2.

4.3. Results and Discussion

4.3.1. Immune Inducibility, Tissue Specificity, Stress Responsiveness, and Developmental Profiles of SRP1 Expression

SRP1 is homologous to *Hyphantria cunea* Hdd23, a gene known to be upregulated by immune challenge, but with unknown function (Shin, 1998). We examined the SRP1 mRNA levels in different tissues at various life stages represented by 52 previously obtained RNA-Seq datasets (Supplemental Figure 4.4.1.) and found that SRP1 is expressed at a lower level than SRP2 but substantially higher than SRP5, SRP7 and SRP8 (Cao and Jiang, unpublished data) (He, 2015). Our previous transcriptome analysis showed that *M. sexta* SRP1 (CIFH contig 06581) was up-regulated 6.6-fold in fat body and 11.9-fold in hemocytes after immune challenge (Zhang, 2011). We further examined SRP1 expression profiles by qRT-PCR analysis. The mRNA levels of SRP1 were very low in hemocytes and fat body from naïve larvae and became significantly higher ($p < 0.05$) in both tissues after larvae were injected with a mixture of bacteria and β -1,3-glucan (Figure 4.1.1.A). The SRP1 mRNA level in induced hemocytes was higher, relative to rpS3 mRNA, than that in induced fat body. In day 3, 5th instar naïve larvae the SRP1 transcripts in integument and trachea were 3–4 fold more abundant than in nerve tissue and muscle (Figure 4.1.1.B). The mRNA levels in midgut and fat body were negligible. We tested if SRP1 expression is induced after larvae were heat shocked at 42°C for 1 h. At 6h after the treatment, we detected significant up-regulation of SRP1 expression in midgut and hemocytes (Figure 4.1.1.C). The highest mRNA

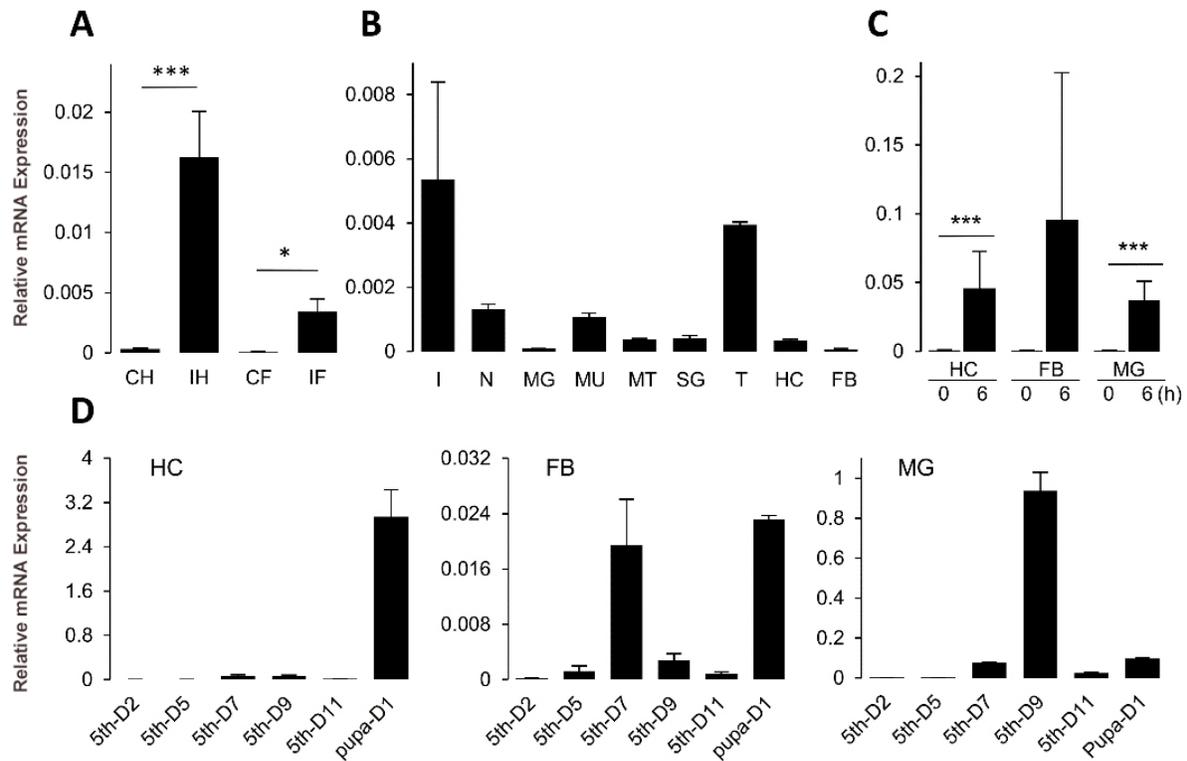


Figure 4.1.1. Determination of *M. sexta* SRP1 mRNA Levels by Quantitative Real-time PCR. (A) Immune inducibility. As described in *Materials and methods*, the total RNA samples of control and induced hemocytes (CH and IH) and fat body (CF and IF) were reverse transcribed to cDNA and analyzed by qRT-PCR using the SRP1-specific primers. The relative mRNA levels normalized with rpS3 based on their Ct values in three biological replicates (3 larvae per sample) and plotted as a bar graph (mean \pm SD, n = 3). The statistical significances were calculated using the Student's t-test and significant increases are shown as * ($p < 0.05$), ** ($p < 0.01$), and *** ($p < 0.001$). (B) Tissue specificity. RNA samples of integument (I), nerve tissue (N), midgut (MG), muscle (MU), Malpighian tubule (MT), salivary gland (SG), trachea (T), hemocytes (HC), and fat body (FB) from day 3, 5th instar naïve larvae were prepared and analyzed by qPCR under the same conditions. (C) Heat treatment. At 6 h after incubation at 42°C (or at 25°C for controls) for 1 h, the hemocyte, fat body and midgut RNA samples were isolated for the qRT-PCR analysis. (D) Developmental profiles. The mRNA levels in hemocytes (left) and fat body (middle) and midgut (right) at different life stages (5th instar, days 2, 5, 7, 9, 11 and pupa, day 1) were determined using the same method. (Performed by Dr. Xiaolong Cao)

level was detected in treated fat-body but the p-value was >0.05 when compared with the control.

We also measured the SRP1 mRNA levels in hemocytes, fat body and midgut during development from feeding stage 5th instar larvae to pupae (Figure 4.1.1.D). The SRP1 transcripts in hemocytes increased three-fold in day one pupae, compared with expression in larval hemocytes. Two peaks of expression were detected in fat body, one in the early wandering stage (day 7) and the other at

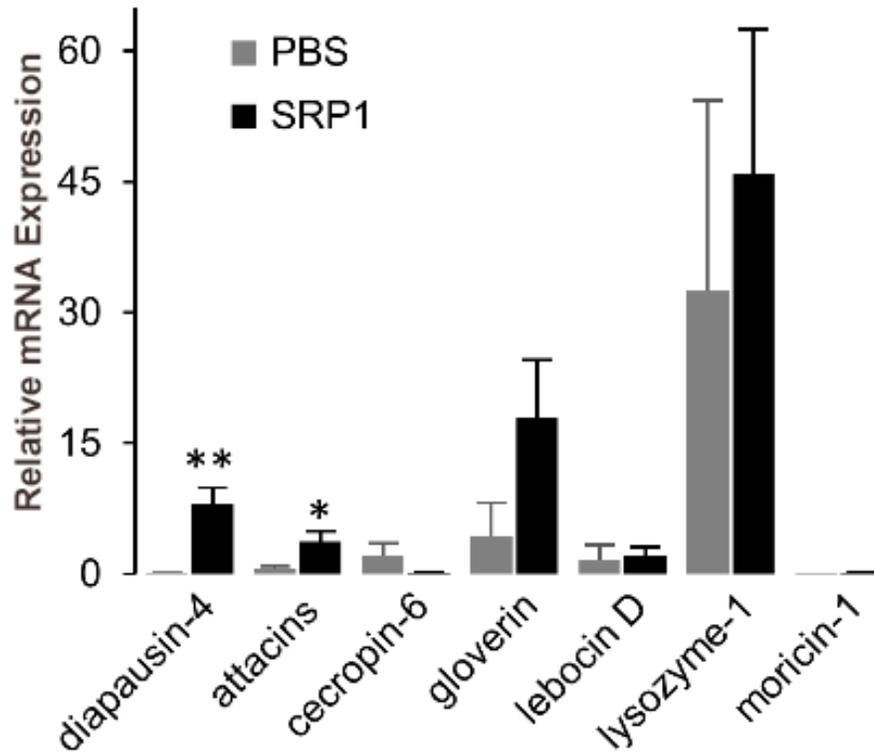


Figure 4.1.2. Induction of AMP Expression by Injection of PBS (grey) or SRP1 (black). As described in *Materials and methods*, the test and control total RNA samples of fat body from the larvae injected with buffer or peptide were analyzed by qRT-PCR using the primers specific for the seven AMP genes. After normalization against rpS3, the average SRP1 mRNA levels in the control samples were adjusted to one, and relative levels in the SRP1 injected group were calculated accordingly. The statistical significances were calculated using the Student's t-test and significant increases are shown as * ($p < 0.05$) or ** ($p < 0.01$). (Performed by Dr. Xiaolong Cao)

the early pupal stage (day 1). In similar stages, expression of AMP genes is also upregulated in response to developmental signals (He, 2015).

4.3.2 Induction of AMP Gene Expression by SRP1 Injection

In order to check whether SRP1 regulates AMP gene expression, we injected the SRP1 peptide solution (4.0 μg) or PBS as a control into day 2, 5th instar larvae. We analyzed the test and

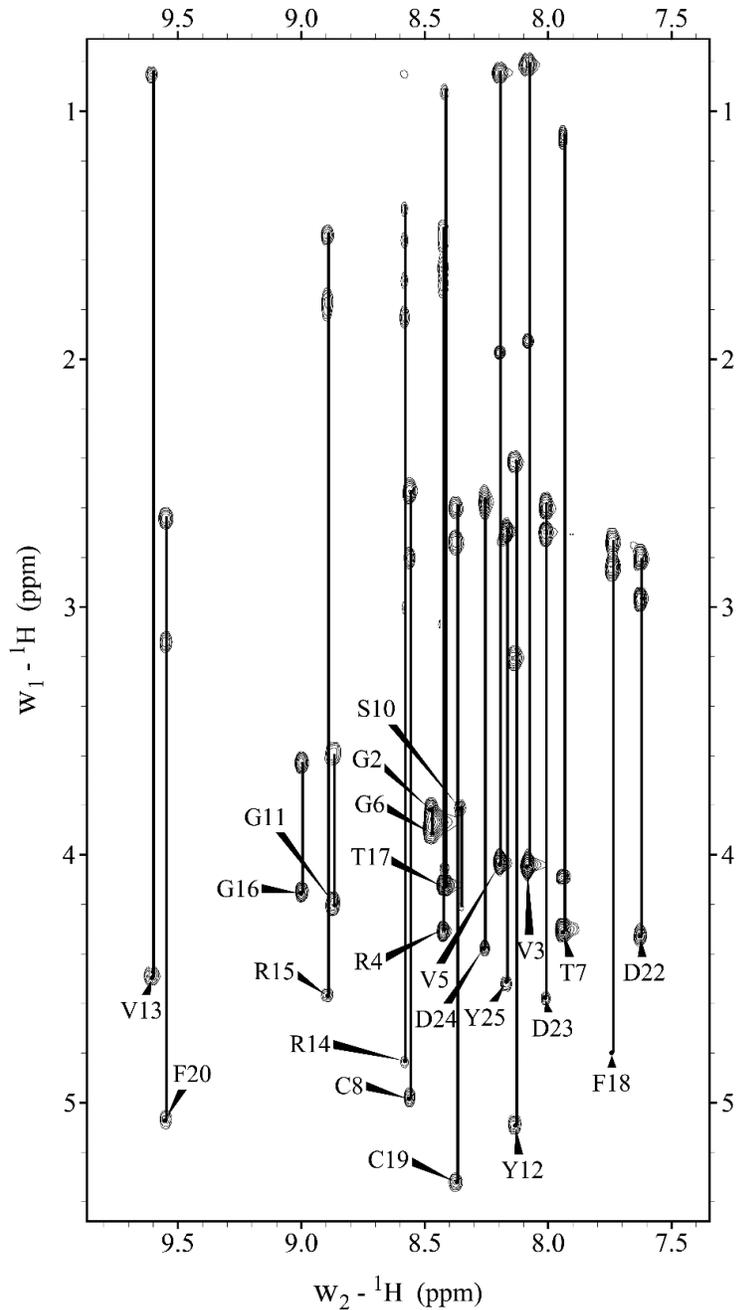


Figure 4.1.3. Fingerprint Region (NH-C α H/Side Chain Protons) of 2D ^1H - ^1H TOCSY Spectrum of SRP1 in H_2O . Individual amino acid spin systems connected with vertical lines and labeled with residue name and number.

control RNA samples by qRT-PCR and observed two of the seven AMP genes were significantly induced ($p < 0.05$) at 6h after SRP1 injection (Figure 4.1.2). The diapausin-4 and attacins mRNA levels were 358.8 and 7.6-fold higher respectively than the control ($p < 0.05$). These results suggest

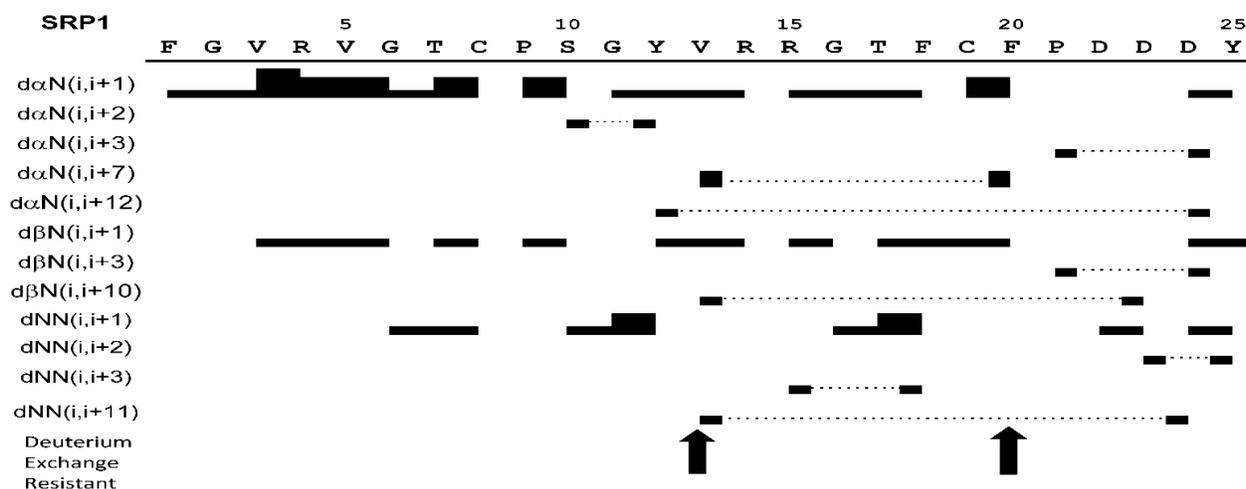


Figure 4.1.4. Summary of the NOE Connectivity for SRP1 in Water. The line thickness indicates the peak intensity, dotted lines indicate the non-sequential connections and arrows highlight residues that are resistant to deuterium exchange.

that SRP1 participates in regulation of immune responses in the tobacco hornworm and justifies a structural study of this cytokine (Zhang, 2011).

4.3.3. NMR Resonance Assignments and Secondary Structure

Following the protocols published earlier, the analyses of 2D ^1H - ^1H TOCSY and 2D ^1H - ^1H NOESY NMR spectra provided intra-residue and inter-residue proton-proton correlations on the amino acid spin system resonances and sequential proton chemical shift assignments respectively for SRP1 (Rance, 1983; Bax, 1985; Kumar, 1995; Schrag, 2017a). The TOCSY spectral data obtained at 25°C showed sufficient chemical shift dispersion for residues with limited spectral overlap. This enabled us to assign proton chemical shift of the spin systems resonances for all residues and indicated the presence of a well-defined secondary and tertiary structure for this peptide. In these assignments, $\text{C}_\alpha\text{H}(i) - \text{C}_\delta\text{H}(i+1: P)$ ($d_{\alpha\delta}$) or $\text{C}_\alpha\text{H}(i) - \text{C}_\alpha\text{H}(i+1: P)$ ($d_{\alpha\alpha}$) NOEs instead of $d_{\alpha\text{N}}(i, i+1)$ were used to assign trans or cis conformations for P9 and P21. We observed that both prolines (P9 and P21) in SRP1 showing strong $d_{\alpha\delta}$ NOEs suggesting trans configuration. The absence of $d_{\alpha\alpha}$ NOE cross peaks in the 2D ^1H - ^1H NOESY spectra indicated a

lack of any Pro in cis configuration. The proton chemical shifts assignments data for *M. sexta* SRP1 have been deposited in BioMagResBank (<http://www.bmrb.wisc.edu>) under the accession number 30507.

Figure 4.1.3. displays NH-C_αH, NH-C_βH, NH-C_γH, and NH-C_δH cross peak assignments of individual amino acid spin system in the fingerprint region of the TOCSY spectrum used for residue identification. The 2D ¹H-¹H NOESY spectral analysis allowed the identification of d_{αN}(i, i+1) cross peaks for sequential assignments and the disambiguation of repeated amino acid residues. Further analysis of the NOESY spectra acquired with different mixing times characterized short, medium and long-range connectivity including d_{αN}, d_{βN} and d_{NN} that provided structural information. An example of 2D ¹H-¹H NOESY spectrum fingerprint region showing d_{NN} connectivity is provided in Supporting Information (Figure 4.6.2.). Figure 4.1.4 illustrates all the sequential and medium range distance NOE connectivity as well as deuterium exchange resistant data for residues of SRP1. A number of non-sequential [weak d_{NN}(i, i+3) NOE between R15 and F18 and weak d_{NN}(i, i+7) NOE between V13 and F20] NOEs suggested an antiparallel β-sheet arrangement (Supplemental figure 4.6.2.). Residues Y12, V13, R14, R15, F18, C19 and F20 were likely to form two short β-strands. This secondary structural observation was further supported by the down field shift of C_α proton resonances in chemical shift index (ΔC_αH) values (Supplemental Figure 4.6.3.) and larger (9.5 -15Hz) coupling constant (3 J_{NH-C_αH}) values for these residues. The random coil values for the chemical shift of C_αH protons of residues were taken from the published data (Wishart, 1994). The SRP1 structure consisted of two short β-strands at region Y12-R15 [weak d_{αN}(i, i+1) NOEs from Y12 to R14] and F18– F20 [medium d_{αN}(i, i+1) NOE between C19 and F20]. These β-strands flanked a type-II β-turn between residues R15 – F18 [weak d_{NN}(i, i+3) NOE from R15 to F18; weak d_{αN}(i, i+1) NOE from residues R15 to F18; weak d_{NN}(i, i+1)

Table 4.1.1. Structural Statistics of the 20 Lowest Energy Structures for SRP1.

NOE Constraints	Number
Total	284
Intra-residue	139
Sequential	63
Medium range ($i-j \leq 5$)	29
Long range ($i-j \geq 5$)	21
Disulfide bonds	4
Hydrogen bonds	6
Dihedral constraints	22
Constraints/residue	11.36
Pairwise R.M.S.D. to mean structure (residues 1-25)	
Backbone atoms (Å)	2.99±0.88
All nonhydrogen atoms (Å)	3.74±0.82
Pairwise R.M.S.D. to mean structure (residues 8-23)	
Backbone atoms (Å)	0.28±0.08
All nonhydrogen atoms (Å)	0.85±0.10
Percentage of residues in regions of ϕ-ψ space	
Allowed	79.5%
Additionally allowed	18.4%
Generously allowed	1.9%
Disallowed	0.2%

NOE between G16 and T17; medium $d_{NN}(i, i+1)$ NOE between T17 and F18]. The short stretch of twisted anti-parallel β -sheet was confirmed by observation of hydrogen bonds between V13 - F20. Low number of NOE cross peaks for N-terminal residues F1 – T7 suggested an extended conformation for this region. However, the C-terminal residues P21 – Y25 displayed turn-like characteristics [weak $d_{\alpha N}(i, i+3)$ NOE between P21 and D24 and $d_{NN}(i, i+2)$ NOE between D23 and Y25].

4.3.4. Tertiary Structure of *M. sexta* SRP1

We determined the tertiary structure of SRP1 by hybrid distance geometry-simulated annealing molecular dynamics calculations using 284 constraints including 139 intra-residue, 63

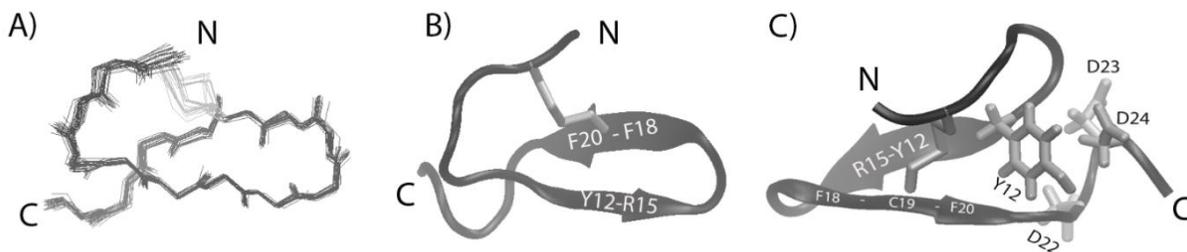


Figure 4.1.5. Structural Representations of SRP1. (A) Superimposition of the 20 lowest energy NMR structures of SRP1. (B) Cartoon of the backbone for the lowest energy structure of SRP1. The disulfide bond between C8 and C19 represented in licorice. (C) 180° rotation of B with sidechains of Y12, D22, D23, and D24 indicated in licorice.

sequential, 29 medium range, 21 long range NOE and 6 hydrogen bond distance constraints as well as 22 dihedral angle constraints. Two additional distance constraints were used between S_{γ} and C_{β} atoms of C8 and C19 to account for normal disulfide bond geometry ($2.0 \text{ \AA} < d_{S-S} < 2.5 \text{ \AA}$, $2.2 \text{ \AA} < d_{C_{\beta}-S} < 3.5 \text{ \AA}$). Table 4.1.1 summarized structural statistics of the 20 lowest energy structures for SRP1 in water. The calculated structures converged to a single family of conformers regardless of the starting structure. For these structure calculations, randomly chosen structures with extended conformation were selected as starting structures. From a total of 200 conformers calculated, 20 structures with lowest energy were selected for further analysis. The torsion angle dynamics of all the 20 selected conformers were consistent with a pair of hydrogen bonds involving the backbone carbonyl of V13 and amide proton of F20, and vice versa. Conformational analysis of these 20 structures suggested two additional hydrogen bonds (F18 backbone carbonyl to backbone amide proton of R15, and R15 backbone carbonyl to backbone amide proton of F18). Figure 4.1.5.A displays superimposition of the 20 lowest energy structure of SRP1. Alignments along residues C8-P21, forming the core region for SRP1, of the 20 lowest energy structures showed backbone root mean square deviation (RMSD) value of 0.28 \AA .

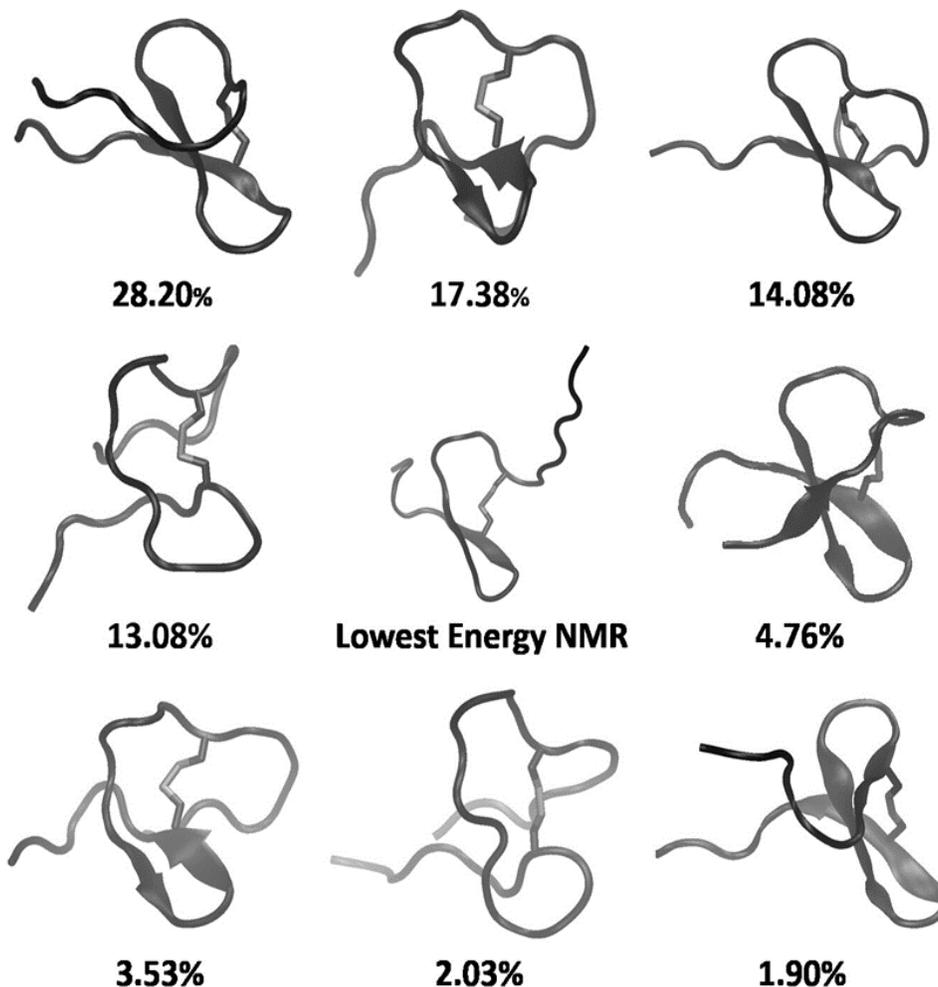


Figure 4.1.6. Eight Most Populated Clusters Derived from Simulation of SRP1. All conformations are aligned using all C α atoms. The percentages below each cluster denote the percentage of the simulation that it represents with lowest energy NMR-derived conformer is shown in the center. The disulfide bond between C8-C19 is represented in licorice with grey coloring for sulfur atoms.

These structures showed good agreement with experimental data with no distance violation larger than 0.3Å and no dihedral angle violation larger than 5°. A Ramachandran plot analysis using Procheck-NMR showed that 79.5% of residues were in most favored stereochemically allowed region and 18.4% were in additional allowed regions (Table 4.1.1) (Laskowski, 1996). The schematic cartoon diagram of the energy minimized average tertiary structure of *M. sexta* SRP1's core region (residues C8-P21) is shown in Figure 4.1.5.B & C. Two hydrogen bonds that

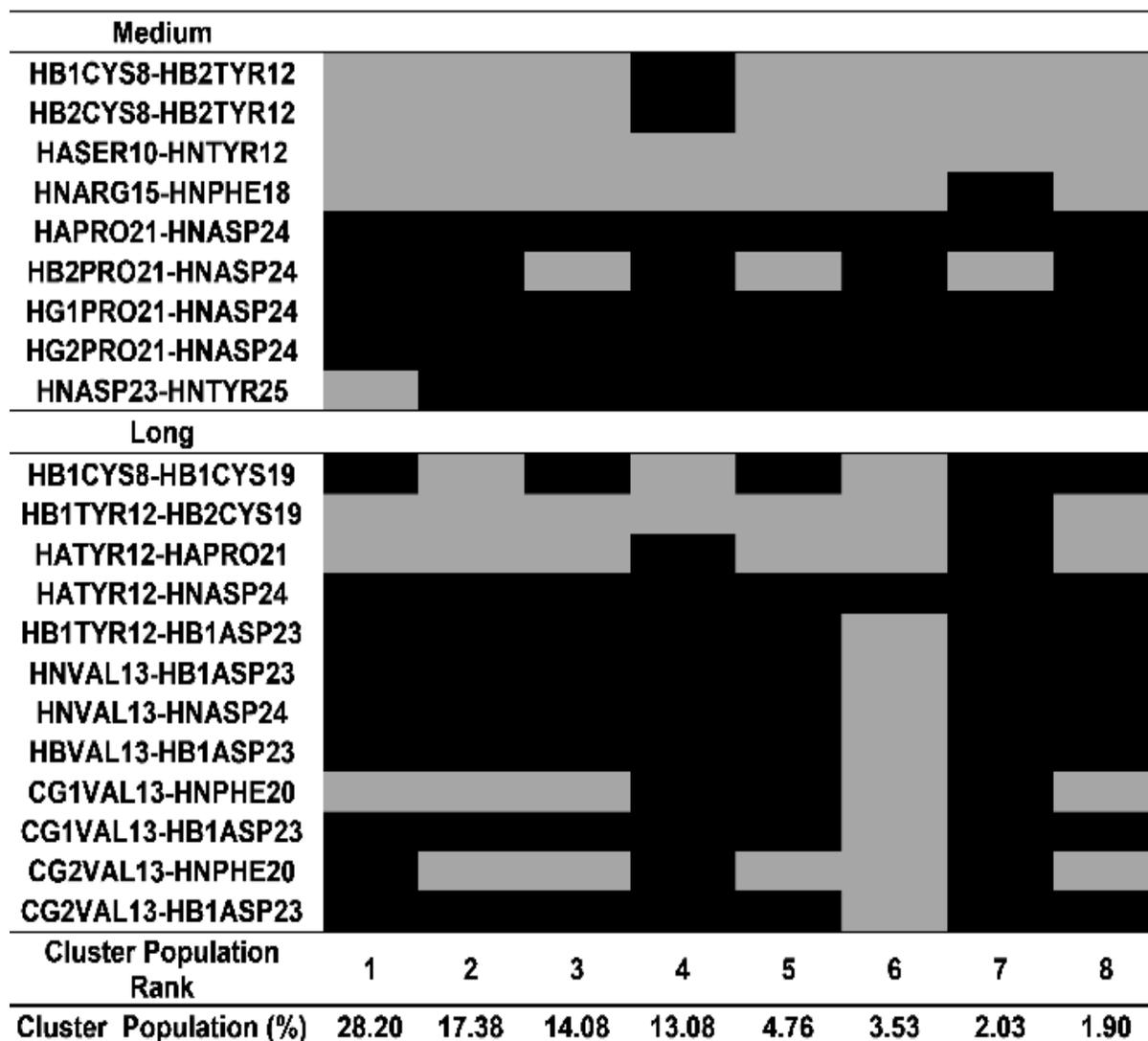


Figure 4.1.7. Summary of NOE Violation Analysis of the Eight Most Populated Clusters from the Simulated Ensemble. Each pairwise medium- and long-range NOE interactions are listed on the y-axis. Cluster population rank and percentage are denoted on the x-axis. Pseudo atom constraints were utilized for methyl group calculations. Black blocks mark NOE that are violated by at least 1.0Å by individual clusters. All instances of gray reflect satisfaction of a constraint for a given cluster.

stabilized the tertiary structure of SRP1 were confirmed by the slowly exchanged amide protons from V13 and F20.

As a means of structural refinement of the calculated NMR ensemble, 600-ns MD simulations seeded with the five lowest energy structures from NMR were conducted in explicit

solvent. A total of 3005 unique snapshots taken at 1-ns intervals were collected from these simulations. The average structures and populations of the eight largest clusters of the simulated ensemble are shown in Figure 4.1.6, which shows a strong degree of agreement in positions and relative integrity of the β -strands throughout simulation with respect to the NMR structure. It also elucidates some conformational heterogeneity present for both N- and C-termini. We have further performed NOE violation analysis on the sub-ensembles of structures that belong to the eight largest clusters. The NOE violation analysis results are summarized in Figure 4.1.7. These results suggest that constraints pertaining to the core region residues (C8-P21) are in strong agreement with NOE observations from NMR. However, there are NOEs that only satisfy in smaller subpopulations of the ensemble pertaining to residues D23 and D24. This likely reflects SRP1's conformational hypervariability in the C-terminus.

The analysis of the tertiary structure of SRP1 suggested that in aqueous solution this peptide possessed a very well-defined core region with a short twisted anti-parallel β -sheet that was stabilized by a type-II β -turn and a disulfide bond. The C-terminus displayed turn-like characteristics, possibly due to electrostatic repulsion of the aspartate residues, while the N-terminus was largely unstructured with substantial flexibility. Overall tertiary structure was stabilized by a combination of the hydrogen bonds within the β -sheet structure and the covalent disulfide linkage. Within the core region of the molecule, one side of the structure was mainly composed of charged residues R14 and R15. The opposite side of the structure was composed of hydrophobic residues, including Y12, F18 and F20. The side chain of Y12 showed NOE contacts to the side chains of C8, P9, and C19 (data not shown). This suggested residue Y12 is embedded, as observed for residues at a similar position in MsPp1 and MsSRP2 (Yu, 1999; Schrag, 2017a). These structural characteristics, similar to other ENF peptides, indicated that the SRP1 core region

adopted a fold very similar to the carboxyl-terminal region of the vertebrate EGF, though they had very low sequence similarities (Yu, 1999). Interestingly, plasmatocyte-spreading peptide 1 (PSP1) from insect *P. includens* showing very similar structural characteristics of its core region also adopted a fold with close similarity to the C-terminal subdomain of EGF-like module 5 of human thrombomodulin (hTM5) (Volkman, 1999). The sequence identity between PSP1 and SRP1 was limited to mainly cysteine, glycine, tyrosine residues and structures of human epidermal growth factor (hEGF) and the hTM5 shared the general structural features of the family of EGF-like domains (Hommel, 1992; Samoili-Benitez, 1997). These findings suggested that SRP1 interactions with its receptor may be analogous to EGF-receptor interactions to trigger the biological function of this stress response peptide.

4.4. Conclusion

In the *M. sexta* genome there are 11 encoded small prepropeptides in the SRP family that function as cytokines. Appropriate maturation of the peptide involves a combination of intracellular and extracellular serine proteases. The tissue expression profile of the SRP1 gene during development shows discrete tissue specific and temporal regulation of the mRNA transcripts. Induction of SRP1 gene expression by biotic stressors was observed in hemocytes and fat body, while heat shock induced gene expression in hemocytes and midgut. Injection of mature SRP1 demonstrates that it stimulates expression of specific AMP genes. These targeted genes are unique and somewhat different from AMP genes induced by SRP2, indicating a distinct function of SRP1. Our results suggest the existence of a proSRP1 activating protease(s), an SRP1 receptor, and a unique role of SRP1 within the immune signal transduction pathway. The structural

information of SRP1 will provide insight into our understanding of the interactions with its putative receptor.

4.5. References

- Bax, A. D., & Davis, D. G. (1969). MLEV-17-based two-dimensional homonuclear magnetization transfer spectroscopy. *Journal of Magnetic Resonance*, 65(2), 355-360.
- Brooks, B. R., Brooks III, C. L., Mackerell Jr, A. D., Nilsson, L., Petrella, R. J., Roux, B., Won, Y., Archontis, G., Bartels, C., Boresch, S. (2009). CHARMM: the biomolecular simulation program. *Journal of computational chemistry*, 30 (10), 1545-1614.
- Brooks, B. R., Bruccoleri, R. E., Olafson, B. D., States, D. J., Swaminathan, S. A., Karplus, M. (1983). CHARMM: a program for macromolecular energy, minimization, and dynamics calculations. *Journal of computational chemistry*, 4 (2), 187-217.
- Brünger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J.S., Kuszewski, J., Nilges, M., Pannu, N.S., & Read, R. J. (1998). Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallographica Section D: Biological Crystallography*, 54(5), 905-921.
- Cao, X., He, Y., Hu, Y., Zhang, X., Wang, Y., Zou, Z., Chen, Y., Blissard, G.W., Kanost, M.R., & Jiang, H. (2015). Sequence conservation, phylogenetic relationships, and expression profiles of nondigestive serine proteases and serine protease homologs in *Manduca sexta*. *Insect biochemistry and molecular biology*, 62, 51-63.
- Darden, T., York, D., Pedersen, L. (1993). Particle mesh Ewald: An $N \cdot \log(N)$ method for Ewald sums in large systems. *The Journal of chemical physics*, 98 (12), 10089-10092.
- Daura, X., Gademann, K., Jaun, B., Seebach, D., Van Gunsteren, W. F., Mark, A. E. (1999). Peptide folding: when simulation meets experiment. *Angewandte Chemie International Edition*, 38 (1-2), 236-240.

- Dunn, P. E., & Drake, D. R. (1983). Fate of bacteria injected into naive and immunized larvae of the tobacco hornworm *Manduca sexta*. *Journal of Invertebrate Pathology*, 41(1), 77-85.
- Goddard, T. D., & Kneller, D. G. (2008). Sparky 3. University of California, San Francisco, USA 15 (20).
- He, Y., Cao, X., Li, K., Hu, Y., Chen, Y. R., Blissard, G., Kanost, M.R., & Jiang, H. (2015). A genome-wide analysis of antimicrobial effector genes and their transcription patterns in *Manduca sexta*. *Insect biochemistry and molecular biology*, 62, 23-37.
- Hess, B., Kutzner, C., Van Der Spoel, D., Lindahl, E. (2008). GROMACS 4: algorithms for highly efficient, load-balanced, and scalable molecular simulation. *Journal of chemical theory and computation*, 4 (3), 435-447.
- Huang, J., MacKerell, A. D., Jr. (2013). CHARMM36 all-atom additive protein force field: validation based on comparison to NMR data. *Journal of computational chemistry*, 34 (25), 2135-2145.
- Humphrey, W., Dalke, A., & Schulten, K. (1996). VMD: visual molecular dynamics. *Journal of molecular graphics*, 14(1), 33-38.
- Hommel, U., Harvey, T. S., Driscoll, P. C., and Campbell, I. C. (1992) Human epidermal growth factor: High resolution solution structure and comparison with human transforming growth factor α . *Journal of Molecular Biology*, 227, 271-282.
- Hopkins, C. W., Le Grand, S., Walker, R. C., Roitberg, A. E. (2015). Long-Time-Step Molecular Dynamics through Hydrogen Mass Repartitioning. *Journal of Chemical Theory and Computation*, 11 (4), 1864-1874.

- Jiang, H., Vilcinskas, A., & Kanost, M. R. (2010). Immunity in lepidopteran insects. In *Invertebrate immunity* (pp. 181-204). Springer, Boston, MA.
- Kabsch, W., & Sander, C. (1983). Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers: Original Research on Biomolecules*, 22(12), 2577-2637.
- Kanost, M. R., & Jiang, H. (2015). Clip-domain serine proteases as immune factors in insect hemolymph. *Current opinion in insect science*, 11, 47-55.
- Kumar, A., Ernst, R. R., & Wüthrich, K. (1995). A two-dimensional nuclear Overhauser enhancement (2D NOE) experiment for the elucidation of complete proton-proton cross-relaxation networks in biological macromolecules. In *Nmr In Structural Biology: A Collection of Papers by Kurt Wüthrich* (pp. 103-108).
- Lemaitre, B., & Hoffmann, J. (2007). The host defense of *Drosophila melanogaster*. *Annual Review of Immunology*, 25, 697-743.
- Laskowski, R. A., Rullmann, J. A. C., MacArthur, M. W., Kaptein, R., & Thornton, J. M. (1996). AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR. *Journal of biomolecular NMR*, 8(4), 477-486.
- Mark, P., Nilsson, L. (2001). Structure and Dynamics of the TIP3P, SPC, and SPC/E Water Models at 298 K. *The Journal of Physical Chemistry A*, 105 (43), 9954-9960.
- Matsumoto, H., Tsuzuki, S., Date-Ito, A., Ohnishi, A., & Hayakawa, Y. (2012). Characteristics common to a cytokine family spanning five orders of insects. *Insect biochemistry and molecular biology*, 42(6), 446-454.

- Phillips, J. C., Braun, R., Wang, W., Gumbart, J., Tajkhorshid, E., Villa, E., Chipot, C., Skeel, R. D., Kale, L., Schulten, K. (2005). Scalable molecular dynamics with NAMD. *Journal of computational chemistry*, 26 (16), 1781-1802.
- Piotto, M., Saudek, V., & Sklenář, V. (1992). Gradient-tailored excitation for single-quantum NMR spectroscopy of aqueous solutions. *Journal of biomolecular NMR*, 2(6), 661-665.
- Rance, M., Sorensen, O. W., Bodenhausen, G., Wagner, G., Ernst, R. R., & Wüthrich, K. (1983). Improved Spectral Resolution in COSY H-1 NMR Spectra of Proteins Via Double Quantum Filtering. *Biochemical and biophysical research communications*, 425(3), 527-533.
- Rieu, I., & Powers, S. J. (2009). Real-time quantitative RT-PCR: design, calculations, and statistics. *The Plant Cell*, 21(4), 1031-1033.
- Ryckaert, J.-P., Ciccotti, G., Berendsen, H. J. (1977). Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *Journal of computational physics*, 23 (3), 327-341.
- Samoili-Benitez, B. A., Hunter, M. J., Meininger, D. P., and Komives, E. A. (1997). Structure of the fifth EGF-like domain of thrombomodulin: An EGF-like domain with a novel disulfidebonding pattern. *Journal of Molecular Biology*, 273, 913-920
- Schrag, L. G., Cao, X., Herrera, A. I., Wang, Y., Jiang, H., Prakash, O. (2017). NMR Structural Studies of Stress Response Peptide-2 from the Insect *Manduca sexta*. *Protein & Peptide Letters*, 24 (1), 3-11.

- Schrag, L. G., Herrera, A. I., Cao, X., Prakash, O., Jiang, H. (2017). Structure and Function of Stress-Responsive Peptides in Insects. *In Peptide-based Drug Discovery* (Srivastava, V., Ed.),; pp 438-451. The Royal Society of Chemistry, London, England.
- Schrodinger, L. L. C. (2010). The PyMOL molecular graphics system. *Version, 1(5), 0*.
- Shin, S. W., Park, S. S., Park, D. S., Kim, M. G., Kim, S. C., Brey, P. T., Park, H. Y. (1998). Isolation and characterization of immune-related genes from the fall webworm, *Hyphantria cunea*, using PCR-based differential display and subtractive cloning. *Insect Biochemistry and Molecular Biology*, 28 (11), 827-837.
- Skinner, W. S., Dennis, P. A., Li, J. P., Summerfelt, R. M., Carney, R. L., Quistad, G. B. (1991). Isolation and identification of paralytic peptides from hemolymph of the lepidopteran insects *Manduca sexta*, *Spodoptera exigua*, and *Heliothis virescens*. *Journal of Biological Chemistry*, 266 (20), 12873-7.
- Stein, E. G., Rice, L. M., & Brünger, A. T. (1997). Torsion-angle molecular dynamics as a new efficient tool for NMR structure calculation. *Journal of Magnetic Resonance*, 124(1), 154-164.
- Strand, M. R. (2008). The insect cellular immune response. *Insect science*, 15(1), 1-14.
- Vranken, W. F., Boucher, W., Stevens, T. J., Fogh, R. H., Pajon, A., Llinas, M., Ulrich, E.L., Markley, J.L., Ionides, J., & Laue, E. D. (2005). The CCPN data model for NMR spectroscopy: development of a software pipeline. *Proteins: structure, function, and bioinformatics*, 59(4), 687-696.

- Volkman, B. F., Anderson, M. E., Clark, K. D., Hayakawa, Y., Strand, M. R., Markley, J. L. (1999). Structure of the insect cytokine peptide plasmatocyte-spreading peptide-1 from *Pseudoplusia includens*. *Journal of Biological Chemistry*, 274, 4493-4496.
- Wang, Y., Jiang, H., & Kanost, M. R. (1999). Biological activity of *Manduca sexta* paralytic and plasmatocyte spreading peptide and primary structure of its hemolymph precursor. *Insect biochemistry and molecular biology*, 29(12), 1075-1086.
- Wishart, D. S., & Sykes, B. D. (1994). Chemical shifts as a tool for structure determination. In *Methods in enzymology* (Vol. 239, pp. 363-392). Academic Press.
- Wüthrich, K. (1986). NMR with proteins and nucleic acids. *Europhysics News*, 17(1), 11-13.
- Yu, X. Q., Prakash, O., & Kanost, M. R. (1999). Structure of a paralytic peptide from an insect, *Manduca sexta*. *The Journal of peptide research*, 54(3), 256-261.
- Zhang, S., Gunaratna, R. T., Zhang, X., Najar, F., Wang, Y., Roe, B., & Jiang, H. (2011). Pyrosequencing-based expression profiling and identification of differentially regulated genes from *Manduca sexta*, a lepidopteran model insect. *Insect biochemistry and molecular biology*, 41(9), 733-746.

4.6. Supplementary Material

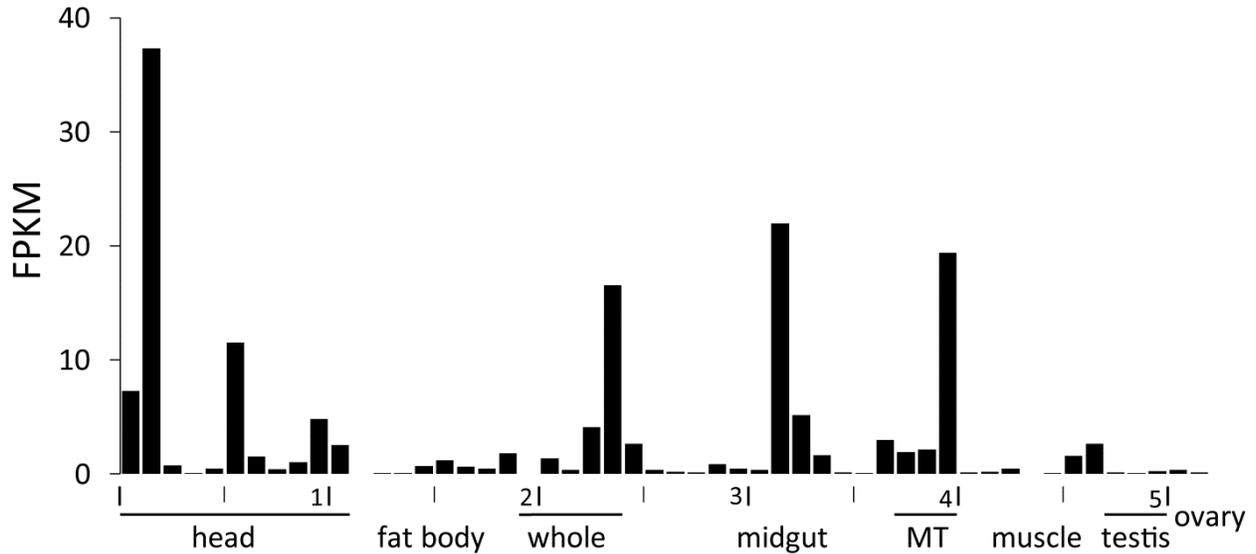


Figure 4.6.1: Transcript levels of *M. sexta* SRP1 gene in the 52 tissue samples. The relative mRNA levels, as represented by FPKM values, are shown in bar graphs. The 52 cDNA libraries (1 through 52) were constructed from the following tissues and stages: head [1. 2nd (instar) L (larvae), d1 (day 1); 2. 3rd L, d1; 3. 4th L, d0.5; 4. 4th L, late; 5. 5th L, d0.5; 6. 5th L, d2; 7. 5th L, pre-W (pre-wandering); 8. P (pupae), late; 9. A (adults), d1; 10. A, d3; 11. A, d7], fat body (12. 4th L, late; 13. 5th L, d1; 14. 5th L, pre-W; 15. 5th L, W; 16. P, d1-3; 17. P, d15-18; 18. A, d1-3; 19. A, d7-9), whole animals [20. E (embryos), 3h; 21. E, late; 22. 1st L; 23. 2nd L; 24. 3rd L], midgut (25. 2nd L; 26. 3rd L; 27. 4th L, 12h; 28. 4th L, late; 29. 5th L, 1-3h; 30. 5th L, 24h; 31. 5th L, pre-W; 32-33. 5th L, W; 34. P, d1; 35. P, d15-18; 36. A, d3-5; 37. 4th L, 0h), Malpighian tubules (MT) (38. 5th L, pre-W; 39. A, d1; 40. A, d3), muscle (41. 4th L, late; 42-43. 5th L, 12h; 44-45. 5th L, pre-W; 46-47. 5th L, W), testis (48. P, d3; 49. P, d15-18; 50. A, d1-3), and ovary (51. P, d15-18; 52. A, d1). Some libraries are from single-end sequencing; the others are from paired-end sequencing. Due to method differences, some synonymous libraries exhibit different FPKMs. (Performed by Dr. Xiaolong Cao)

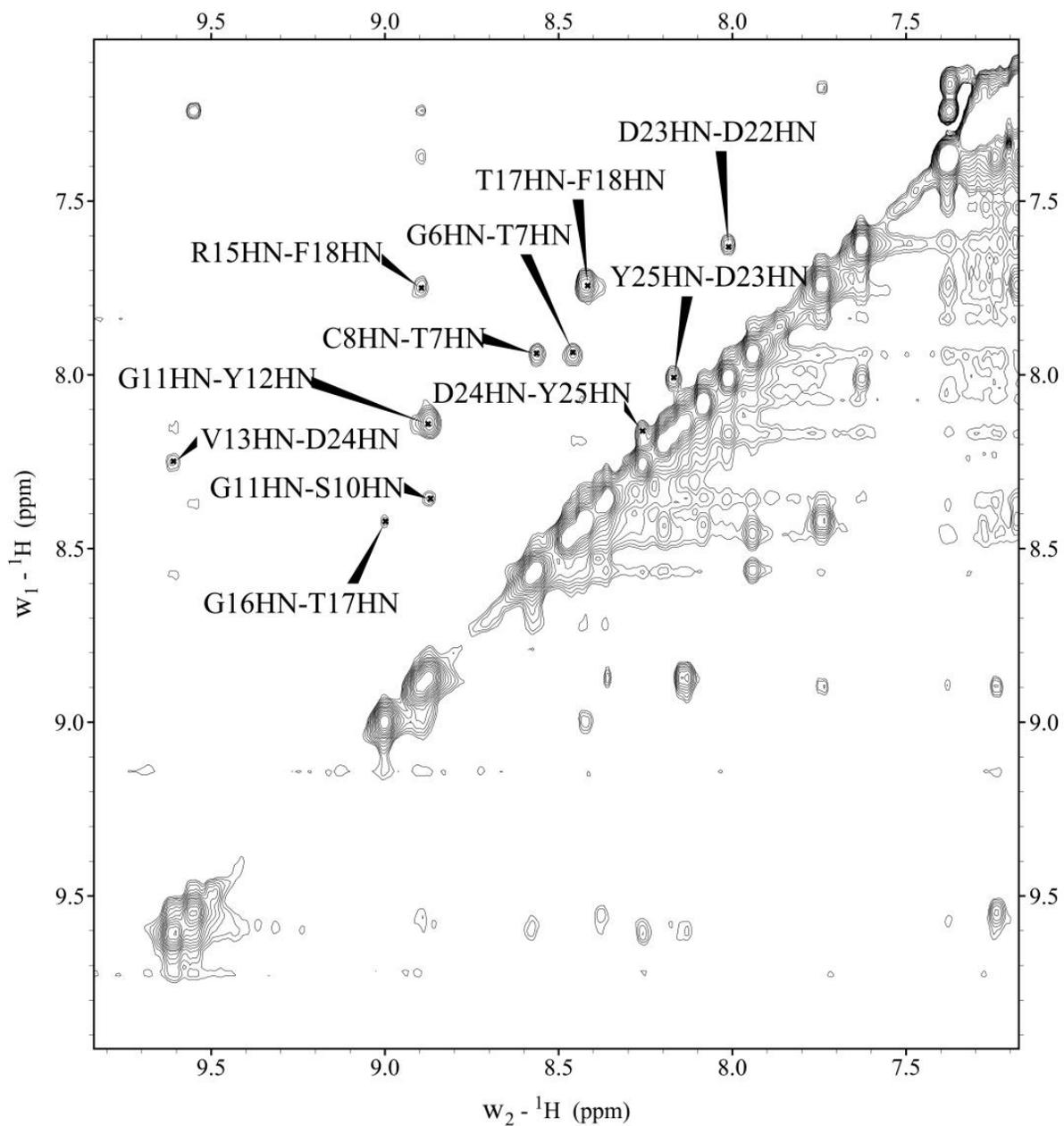


Figure 4.6.2: Fingerprint region (HN-HN) of 2D ^1H - ^1H NOESY spectrum. Uniquely assigned inter-residue NOEs are labeled by residue name, residue number, and xplor format atom name.

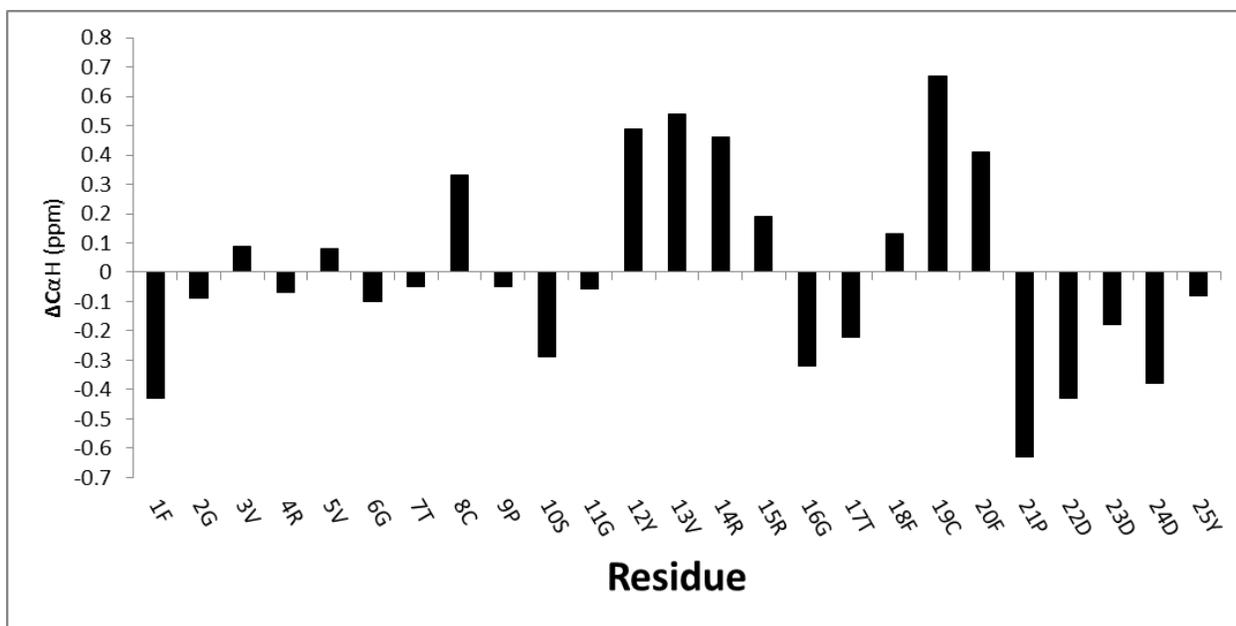


Figure 4.6.3. CaH chemical shift index (ΔCaH) of individual residues in SRP1. Reference chemical shift values are based upon published random coil values (Wishart, 1994). Bars represent difference between published values and values observed during NMR acquisition for SRP1.

Chapter 5- Introduction: Structure-activity Relationships in

Human Tumor Suppressor p53

5.1. Introduction

HsP53 (also known as TP53) is one of the most thoroughly studied molecules in molecular biology to date (Dolgin, 2017). P53 is the most commonly mutated gene in human cancer cells observed to date (Lawrence, 2014; Kandoth, 2013). This observation has led to p53 being heralded as the ‘guardian of the genome’ as it has been identified to have a key role in regulating the DNA damage response (DDR) in mammalian tissues, a role that is accomplished through p53’s activity as a transcription factor (TF).

Transcription factors are proteins that bind to DNA in a sequence specific manner to modulate mRNA transcription through interactions with other mediator proteins or RNA polymerase II directly (Latchman, 1993). Generally, TFs regulate cellular responses that include mitogenesis, differentiation, migration, senescence, and apoptosis, depending upon environmental stimuli. This is largely accomplished through regions of the protein independent from DNA binding, referred to as activation domains (Ptashne, 1988). These regions are often rich in aspartate, glutamate, glutamine, and proline (Mitchell and Tjian, 1989). More often than not, these regions are unstructured and facilitate critical protein-protein interactions that lead to either suppression or activation of the TF. These activation domains, as well as the TF as a whole, are often regulated additionally through post-translational modifications (PTMs).

The total number of experimentally determined TFs within the human genome was 611 as of 2009 (Vasquerizas, 2009). However, computational analyses of the human genome suggest that the total number is much larger, with conservative estimate of 1306 possible factors in total (Badis,

2009). Considering the multitude of different factors, it is often a point of contention on what specific criteria these factors are to be categorized: By the types of gene transcription that are promoted or suppressed, by the mechanisms that regulate transcription factor activation, or by the structural homology concerted to their DNA binding domains (DBD). Most often, TFs are classified according to the structural homology of their DBDs as it precludes the complexities of regulation and activity, which are often quite diverse depending upon the TF.

Using DBD structural homology rationalization, TFs have been classified into five distinct super-families (or super-classes) derived from a Hidden Markov Model based bioinformatics approach (Stegmaier, 2004): Superclass 1 contains the basic domains with a large number of lysine and arginine residues. This class often cannot fold in solution but in the presence of DNA will dimerize and form α -helical motifs such as leucine zipper (ZIP), helix-loop-helix (HLH), or a helix-span-helix (HSH) (Weiss, 1990). Superclass 2 contains the Zn-coordinating domains with varying types of cysteine coordinating elements. Often, these elements include C2H2 or C6 motifs and are typically located near turns and loops (Fedotova, 2017). Superclass 3 contains the Helix-turn-helix domains which have been structurally categorized to be either di-, tri-, and tetra-helical and winged (Religa, 2007; Ogata, 1992; Hinrichs, 1994). This class is exclusively monomeric in its interactions with DNA and the mechanism of DNA selectivity are perhaps best characterized. The C-terminal helix is generally considered as the recognition helix as it forms site-specific hydrogen bonds major groove contacts within the DNA, imparting sequence selectivity, while the immediately preceding helix forms electrostatic interactions with the phosphate backbone (Matthews, 1982; Wintjens, 1996). Superclass 4 contains the β -scaffold with minor groove contacts which is unique in that it uses β -strands and loops to create the contact surface with DNA. This deforms DNA geometry upon binding, regularly resulting in a sharp kink in the DNA helix

(Lebrun, 1999; Zacharias, 2006; Bewley, 1998; Schumacher, 1994; Kim, 1993a; Kim, 1993b; Kim, 1994; Nikolov, 1996; Love, 1995). And finally, superclass 0 includes TFs that fail to meet criteria within the DNA binding domain consistent with other known TF super-classes.

P53 belongs to superclass 4 as residues 102-292 have been shown by x-ray crystallography to form the iconic β -scaffold with minor groove contacts (Cho, 1994). The points of DNA contact are handled by two distinct loops for every monomer of p53 that are a part of a larger loop-sheet-helix motifs (Cho, 1994). These loops are then organized through a zinc coordinating center which is critical for DNA binding and prevention of aggregation by reducing loop motility (Joerger, 2004; Duan and Nilsson, 2006).

In this chapter, we will review how p53 was initially identified. We will also explore each of the structural components of p53 and relate these features to its known activities.

5.2. Discovery

P53 was initially categorized as the Middle T antigen of cells transformed by the Simian virus 40 (SV40). Viral studies focused on protein expression profiles of transformed cells identified three unique antigenic proteins: the large antigen (90kDa), the small antigen (17kDa) and the Middle T antigen (55kDa, p53). At the time, it was thought that p53 was not a host cell protein as it showed the capacity to bind to the T antigen of SV40, accumulated in the nucleus of transformed cells, and was often a target of humoral responses in murine cells (Kress, 1979; Lane and Crawford, 1979; Linzer and Levine, 1979; DeLeo, 1979; Crawford, 1983). Once identified as a host gene, cDNA cloning and initial characterization of p53 in the context of Ras signaling lead to early mis-categorization as an oncogene as clones were from dysfunctional p53 mutants

(Eliyahu, 1984; Jenkins, 1984; Parada, 1984). This mis-categorization of p53 continued until studies conducted in 1989 showed that most tumor isolates of p53 cDNA were mutants (Finlay, 1989). Further studies that year went on to show that point mutations and deletions within p53's chromosome 17p13 correlated its inactivation with incidences of colorectal carcinomas and lung cancers (Baker, 1989; Takahashi, 1989). This then firmly established p53 as the pro-oncogenic suppressor (tumor suppressor) and not an oncogene.

5.3. Structure and Known Biological Activities of p53 Domains

P53, as mentioned prior, operates as a critical regulator of the DDR (Lawrence, 2014; Kandoth, 2013). This is accomplished through a delicate balance of expression versus degradation. p53 itself is constitutively expressed in all somatic cells but under normal cellular conditions it will be efficiently degraded by the E3 ubiquitin ligase Mouse-double-minute-2 (MDM2/HDM2) (Haupt, 1997; Honda, 1997; Kubbutat, 1997). Ultimately, this keeps basal intracellular levels of p53 relatively low. This balance is only altered upon the introduction of genotoxic stressors that can range from DNA damage, hypoxia, oncogene activation, ribosomal stresses, presence of reactive oxygen species (ROSs), etc. (Hafner, 2019) These genotoxic stressors often lead to activation of DDR pathways. This can result in numerous regulatory outcomes, that ultimately detach p53 from its primary suppressor, HDM2 (Shieh, 1997). Once this interaction is interrupted,

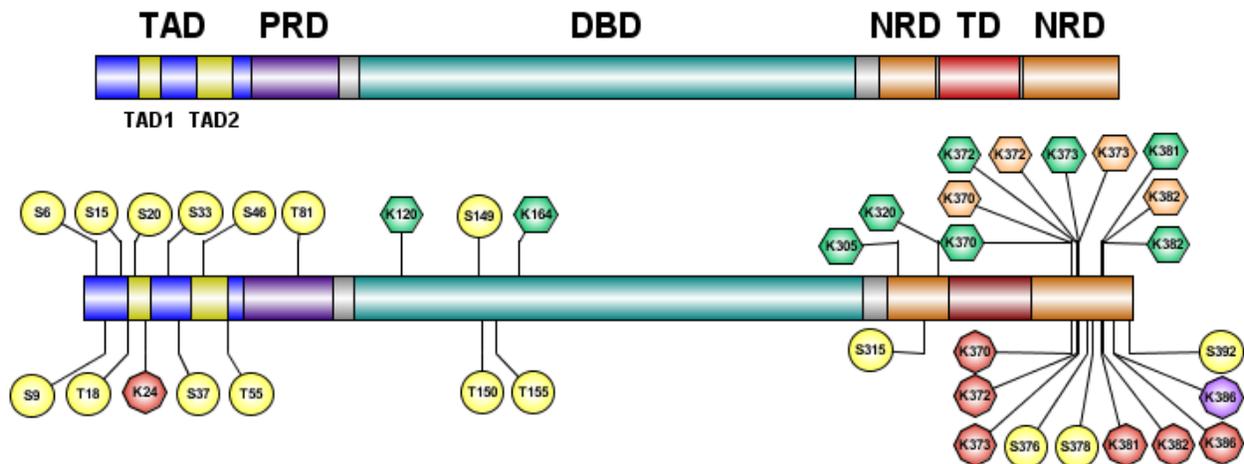


Figure 5.1. Linear Schematic of p53. (Upper) Simple linear schematic outlining the location of transactivation domain (TAD; blue) and its two pseudo-helices TAD1 and TAD2 (yellow), the proline-rich domain (PRD; purple), the DNA-binding domain (DBD; green), the N-terminally regulated domain (NRD; orange), and the tetramerization domain (TD; red). (Lower) Schematic of all known sites of phosphorylation (yellow circles), ubiquitination (red octagons), sumoylation (purple octagon), acetylation (green hexagons) and methylation (orange hexagons) with the anchoring PTM residue indicated for each. Representation generated using Illustrator of Biological Sequences (Liu, 2015).

p53 accumulates within the cell and upregulates responses such as cell cycle arrest, DDR, apoptosis, and metabolic modulation (Hager, 2014; Kang, 2019; Kruiswijk, 2015).

Because genotoxic stressors are so common in the cancer cell environment, it is unsurprising that dysfunction of p53 regulation, as well as p53 itself, is a necessary if not a critical step in oncogenesis and tumor development. As such, it has been viewed as necessary to study all possible forms of p53 regulatory dysfunction to ultimately aid in better diagnostic and prognostic outcomes for patients that are either at risk or currently undergoing cancer treatment. Making p53 one of the most studied biological molecules of all time (Dolgin, 2017).

In the following sections of this chapter, we will conduct an overview of the literature concerning major structural domains of p53: the Transactivation Domain, Proline-rich Domain, DNA-binding Domain, N-terminally Regulated Domain, and its embedded Tetramerization Domain, as shown in figure 5.1. Each domain will be addressed independently with references to

regulatory interactions from distal domains included when necessary. Each section will outline the major structural features of the domain independently and correlate these activities to known structural information.

5.3.1. The Transactivation Domain (TAD) (M1-D61)

The transactivation domain of p53 is the primary region within which hetero-complex formation with a wide ensemble of regulatory partners occurs, including repressor and E3 ubiquitin ligases HDM2 (Kussie, 1996; Oliner, 1993; Chen, 1993; Wadgaonkar, 1999; Grossman, 1998) and MUL1 (Peng, 2016; Lee, 2019), SMADs (Cordenonsi, 2007), Prolyl-isomerase Pin1 (Berger, 2005; Zheng, 2002; Wulf, 2002), single strand binding protein hRPA70 (Vise, 2005; Miller, 1997; Li, 1993; Dutta, 1993; Abramova, 1997; He, 1993), and co-activators/Histone acetyltransferases CBP/p300 (Ferreon, 2009; Krois, 2016; Scolnick, 1997; Gu, 1997a; Lill, 1997; Avantaggiati, 1997; Livengood, 2002; van Orden, 1999; Wadgaonkar, 1999). This promiscuity of p53's transactivation domain is afforded to it by its nature as an intrinsically disordered region (Vise, 2005; Ferreon, 2009; Zhan, 2013; Lee, 2019; Kussie, 1996). This enables apo-p53 TAD to adopt unique structural conformations with each of its binding partners. Although incredibly dynamic, IDRs do not sample completely random conformations but often sample a preferred set of conformers. Thus, IDRs have a native ensemble of structures that likely influence their interaction profiles. As such, the high mobility of the p53 TAD, like other IDRs, limits the use of direct biophysical approaches such as nuclear magnetic resonance (NMR) spectroscopy, small angle X-ray scattering (SAXS), and Forster resonance energy transfer (FRET), etc. for characterizing structure. NMR is the most often employed technique as it is versatile at detecting changes in structure through nuclear Overhauser spectroscopy (NOESY) and chemical shift perturbation, as

well as in determining changes in molecular motion on pico-seconds to second time-scales through heteronuclear relaxation studies. In the absence of binding partners, the models generated from these techniques are low resolution as they measure the ensemble average structure and motion, not the distribution of discrete conformations contained within. To compensate for this apparent drawback, molecular dynamics (MD) simulations are regularly employed to characterize atomic-level detail of the IDP's conformations and approximation of the ensemble distribution. When MD is employed in conjunction with direct experimental approaches, such as NMR techniques, it has produced some of the most resolved mechanistic insights into the regulation of the p53 by its PTMs and dysregulation by mutations.

According to previous characterizations, the p53 TAD can be subdivided into two distinct regions based upon changes in their interactions with binding partners: TAD 1 (T18-L26) and TAD2 (M41-T55). Analysis of the free p53 TAD in both NMR and MD simulations has shown that both TAD1 and TAD2 segments have a propensity for formation of short helices in the unbound state, with the former showing a slightly greater disposition for such structures (Zhan, 2013; Ganguly, 2015). The TAD1 forms a conserved α -helical segment for residues T18-L26 upon binding with either HDM2 or CBP/p300 domains (Kussie, 1996; Dawson, 2003; Krois, 2016). MD simulation analysis suggests that π - π stacking interactions between F19 and W23 of p53 likely provides stability to the α -helical structure as mutation of either residue is detrimental to the helix formation (Lee, 2000). Further MD and NMR analyses have suggested that residues F19, L22, W23, and L26 are critical in hydrophobic contact formation with HDM2 and CBP/p300 domains (Joseph, 2010; Krois, 2016). TAD2 differs from TAD1 in that the position and structure of helix varies more widely depending upon binding partners. In the context of CBP/p300 binding, p53 TAD2 forms a coil for residues D42-S46 and a helix for residues P47-T55 in contact with both

TAZ1 and TAZ2 subdomains (Krois, 2016). Studies of p53 TAD bound to the MUL1 RING domain have revealed a helical contact surface for residues D48-T55 but with a possible rigidified extension for residues Q38-P47 (Lee, 2019). Additionally, studies of hRPA70 binding have suggested possible helix formation for residues D42-D49 with a rigidified extension from residues I50-T55 (Vise, 2005). Interestingly, a majority of reported p53 interaction partners specific to the TAD, save HDM2, interact primarily with the TAD2 segment and may be correlated to its greater IDP propensity to facilitate binding with a wider range of partners (Ferreon, 2009; Vise, 2005; Berger, 2005; Kussie, 1996; Lee, 2019). However, not all binding partners bind exclusively to a single TAD segment, as the transcription adaptor putative zinc finger (TAZ) domains of co-activators CBP/p300 have been demonstrated to bind both the TAD1 and TAD2 simultaneously (Krois, 2016; Ferreon, 2009).

TAD is the primary regulatory region of p53 in-terms of stability and targeting gene specificity. Like many other proteins, complex formation within this region is regulated through PTMs. These PTMs then change the effective interaction surface of the molecule by compromising formation of certain complexes while promoting others. Almost all known PTMs within p53's TAD are phosphorylation sites targeted by kinases (S6/9/15/20/33/37/46 and T18/55) and one ubiquitination site (K24) (Peng, 2016; Cordenonsi, 2007; Tibbetts, 1999; Craig, 1999; Sakaguchi, 1998; Oda, 2000; Waterman, 1998). These PTMs can integrate a variety of different pathways but most are shepherded towards p53 stability and co-activation. First, S15 is phosphorylated in response to DNA damage by DNA-PKcs, ATM, and ATR (Tibbetts, 1999; Kalid and Ben-Tal, 2009; Shiloh, 2013). Phosphorylation of this residue is a pre-requisite for successful phosphorylation of T18, by Casein Kinase 1 or DNA-PKcs, and S20, by Checkpoint Kinase-2 (CHK2), which is critical in interrupting HDM2 interaction (Kalid and Ben-Tal, 2009; Matsumoto,

2006; Dumaz, 1999). MD simulations suggests that this is accomplished through disruption of local helical dynamics of the p53TAD1 and introduction of anions in to HDM2's anionic patch at the p53 binding surface (Craig, 1999; Schnon, 2002; Kalid and Ben-Tal, 2009; Lee, 2007; Mavinahalli, 2010). Interruption of p53's interaction with HDM2 is critical for two reasons: HDM2 operates as E3 ubiquitin ligase and poly-ubiquitinates the NRD of p53 (Xu, 2003). HDM2 inhibits co-activation by CBP/p300 and others by occluding TAD1 segment from binding (Oliner, 1993; Chen, 1993). It is proposed that complete CBP/p300 complex formation with p53 TAD is a necessary step in further complex formation with RSF1. RSF1 promotes greater expression of p53 target genes by aiding nucleosome remodeling induced by CBP/p300 histone acetylation proximal to the target start site (TSS) and enables greater accessibility of the RNA polymerase II complex for faithful gene transcription (Su, 2015; Hafner, 2017; Min, 2018). S33 and S37 are subject to phosphorylation by DNA-PKcs and ataxia-telangiectasia mutant related (ATR) which promotes recruitment of PCAF and CBP/p300 to p53 TAD (Sakaguchi, 1998; Kalid and Ben-Tal, 2009). Phosphorylation of S46 and S37 in response to ionizing radiation demonstrate recruitment of Prolyl-isomerase 1 (Pin1) which is a necessary intermediate step in the phosphorylation of S20 by CHK2 (Oda, 2000; Wulf, 2002; Zhang, 2002; Berger, 2005; Buschmann, 2001). Phosphorylation of S6 and S9 both are accomplished through ataxia-telangiectasia mutated (ATM) and DNA-protein kinase catalytic sub-unit (DNA-PKcs) and results in SMAD recruitment by p53 (Cordenonsi, 2007). There is only a single known constitutive phosphorylation site in p53 TAD as T55. T55^P is thought to be suppressive phosphorylation CBP/p300 binding as it is dephosphorylated in response to ionizing radiation (Li, 2004; Gatti, 2000; Waterman, 1998). Finally, the only known site of ubiquitination within p53 TAD by mitochondrial Ubiquitin Ligase-

1 (MUL1) occurs at K24, ultimately leading to degradation of p53 similar to HDM2-mediated poly-ubiquitination (Peng, 2016; Lee, 2019).

An unusual feature of p53TAD, and in general of IDRs, is its low density of identified cancer-associated missense mutations, as only five percent of 2300 identified for the p53 protein map specifically to the TAD. This aberrantly low density, despite the key regulatory roles of p53 TAD segment, has led to a perception that missense mutations within the region do not have influential structural impacts with respect to cancer. Consequently, this apparent lack of preference bias of cancer-associated mutations towards TAD has led to a perceived lack of importance of studying such mutations within the TAD. As a result, most previously examined mutation sites have largely been restricted to either PTM sites or direct contact surfaces with regulatory partners. However, continuing studies of IDPs, as well as p53 TAD, have revealed that elements of their interactions with binding partners are influenced by the ensemble distribution and dynamics. Moreover, there exists a substantial population of cancer-associated missense mutations that do not correlate with either PTMs or direct interaction surfaces. These distal uncharacterized mutations, although proportionally few, may contain clues concerning ensemble dynamics and distribution that also influence not yet understood regulatory activities involving the p53 TF. As such, further analysis of these cancer-associated mutations utilizing both direct experimental approaches (e.g. NMR) and MD simulations will provide an adequate level of insight necessary for characterizing such changes in structure and dynamics.

5.3.2. Proline-rich Domain (PRD) (D61-S94)

The p53 Proline-rich domain (PRD) is unique in that it contains a series of PXXP amino acid repeats resulting in an unstructured domain akin to the p53's TAD, albeit with a lower structural flexibility. Similar to p53 TAD, the PRD is largely understudied due to its low cancer-associated mutation rates. More studies on the p53 PRD are required, as this domain has been shown to host several key activities relevant for p53 regulation and target selection. And although not nearly as dynamic, like all intrinsically dynamic regions this region may be involved in some of the uncategorized activities with p53 interactors.

The current knowledge of the PRD is that it performs a couple of essential roles with different co-activators in shifting HDM2 versus CBP/p300 binding. Phosphorylation of S37 and S46 of p53 TAD, as noted previously, recruits Pin1 to promote isomerization of the P82 residue and this enables phosphorylation of T81 (Wulf, 2002; Berger, 2005; Buschmann, 2001). Conversion of T81 to T81^P is a critical step in CHK2 recruitment which is a requisite phosphorylation event in disrupting HDM2 interactions within TAD1 (Berger, 2005; Buschmann, 2001). Studies of p53 Δ PRD have also demonstrated MDM2 mediated degradation and increase of exo-nuclear export, even under genotoxic stress conditions, suggesting diminished CBP/p300 activities (Berger, 2001). It was observed that CBP/p300 in the absence of PRD could not acetylate p53's C-terminus in a DNA-independent manner unless supplemented with PXXP repeat containing peptides (Dornan, 2003). These results suggested that p53's PRD also operates as an allosteric activator with CBP/p300, promoting acetylation of p53's C-terminus.

5.3.3. *The DNA-binding Domain (DBD) (T102-K292)*

The DBD is perhaps the most well studied domain of the entire p53 molecule. The DBD itself is the ‘work horse’ domain of the TF as it drives both non-specific and sequence specific interactions with DNA. Analysis of gene targets influenced by p53 using transcription profiling identified up to 1,500 unique genes in the human genome that are either directly or indirectly controlled by p53 (Kannan, 2001; Madden, 1997; Mirza, 2003). However, there is a fair amount of disagreement on total number as variance may be a result of method, chosen cell types, types of genotoxic stresses and modulation of genotoxic stress dosage. Evaluation of high confidence target genes from transcription profiling with chromatin immunoprecipitation (CHIP) revealed a majority bound p53 proximal to their TSSs (Cui, 2011; Smeenk, 2008; Wei, 2006; Botcheva, 2011; Nguyen, 2018). These p53 DBD binding sites often lie either within the promoter region or the first intronic sequence of the target gene (Fischer, 2017; Juven, 1993; Riley, 2008; Sullivan, 2018). Surprisingly, there is additional evidence that distal TSS p53 binding sites also exist and make up a non-trivial fraction of high confidence direct gene targets (>10,000bp) (Koeppel, 2011; Su, 2015). These distal sites are often found within Alu and other interspersed repeat types within human genome (Cui, 2011; Bao, 2017). Because of the limitations of existing CHIP approaches being restricted to 10,000bp from TSSs, it is often difficult to parse direct versus indirect regulation for many low confidence p53 modulated genes.

The p53 DNA binding site itself was identified to be composed of two 10bp half-sites (el-Deiry, 1992; Funk, 1992). Using *in vitro* assays with synthetic DNA oligomers, the canonical half-site consensus sequence of RRRCWWGYYY (R represents A or G; W represents A or T; and Y represents C or T) was independently confirmed by two separate groups (el-Deiry, 1992; Funk, 1992). Each quarter-site, comprised of 5bp, is recognized by an individual monomer of p53

tetramer with 20bp in total recognized by the whole tetramer (Kitayner, 2006; Kitayner, 2010). Each half-site is separated by a spacer sequence of 0-20bp (Cui, 2011; el-Deiry, 1992; Tonelli, 2015). These canonical half-sites have both quarter-sites that are often palindromic and arranged head-to-head (Hafner, 2019). *In vivo* assays showed that the majority of p53 binding sites within the human genome, especially those with high confidence of being under p53 regulation, contain this canonical orientation. However, non-canonical sites such as head-to-tail or tail-to-tail sequences within quarter-sites as well as long spacer sequences between half-sites have also been identified for genes strongly induced by p53 stabilization (Rouault, 1996). Moreover, sequences that contain these non-canonical sites are far less represented in the genome and suggest they may be limited to specific genes (Cui, 2011; Smeenk, 2008; Wei, 2006; Verfaillie, 2016; Chang, 2014). Bioinformatic approaches scoring transcription induced by binding of p53 implicated that most canonical sites were retained to known cell-cycle regulators and non-canonical sites to pro-apoptotic regulators (Weinberg, 2004; Schlereth, 2013; Inga, 2002; Noureddine, 2009; Qian, 2002).

Due to the demonstrated higher affinity of canonical sites, general cell-cycle regulators, over non-canonicals, pro-apoptotic effectors, the model for describing cell-fate dictated by p53 has been built around p53 concentration dependence (Chen, 1996). The implication is that low concentrations of p53, maintained by the HDM2 negative regulation, would result in cell-cycle arrest as upregulation of these gene targets are more sensitive to p53. Conversely, once high levels of p53 are achieved, after its stabilization induced by an extensive or irreparable genotoxic stress, then pro-apoptotic factors would reach expression levels significant enough to commit the cell to apoptosis. However, this model is incomplete. First, pro-apoptotic genes such as NOXA and PUMA contain canonical sites (Kaeser, 2002; Weinberg, 2004). Second, other pro-apoptotic effectors with non-canonical sites show binding to p53 at low concentrations while still not

inducing apoptosis (Kaesler, 2002; Kracikova, 2013; Smeenk, 2011). Even more astounding is that most genes identified with p53 bound to canonical and proximal sites fail to induce p53-dependent regulation (Smeenk, 2011; Wei, 2006; Schlereth, 2013). It can then be inferred that target gene selection and regulation by p53 is not only informed by p53 concentration, but also through PTMs of p53 itself in regulating complex formation with its various co-activators.

The structure of the p53 DBD is an anti-parallel β -sheet motif with interspersed loop-sheet-helix motifs (Joerger and Ferscht, 2007). Loop 1 (L1; F113-T1123), β -strand 2 (S2) and 2' (S2') and segments of the C-terminal domain form the non-specific contact surface with the DNA major groove (Cho, 1994). The L1 is highly flexible and it is posited that this facilitates translocation of the DBD along the strand in the absence of site-specific recognition. L2 (K164-C176 and C182-K194) along with L3 (M237-P250) are critical for formation of the contacts with the minor groove of the DNA double helix. The L2 and L3 are structurally stabilized by accommodating a single Zn^{2+} coordinating site with residues C176, H179, C238 and C242 (Joerger, 2004). Stabilized by Zn^{2+} , residues R175-S183, S183-R196, and Q198-N235 create the critical hydrogen-bond contacts between p53 DBD and the DNA minor groove which is thought to impart quarter-site selection (Lu, 2007). In addition to L3's role in DNA binding, it participates, along with a segment of L2, in a homodimer interface along the DNA half-site between residues K120, S241, R248, R249, R273, A276, C277, and R280 (Saha, 2014). This creates a clamp-like structure between L1 and L2 in both monomers which enables rigidification upon recognition of the quarter-site consensus sequence. Due to this clamp, p53 DBD and other β -scaffolds TFs often distort target DNA sequences upon recognition by introducing a sharp kink in the DNA double helix (Lebrun, 1999; Zacharias, 2006; Bewley, 1998; Schumacher, 1994; Kim, 1993a; Kim, 1993b; Kim, 1994; Nikolov, 1996; Love, 1995). It is then supposed that introduction of the kink locks p53 into place by

preventing lateral migration along the DNA strand. This model has gained popularity as tetrameric, dimeric, and monomeric states of p53 demonstrate binding so long as the DNA presents the necessary distortion in its geometry (Ma and Levine, 2007; Pan and Nussinov, 2007; Lubin, 2010).

As mentioned earlier, p53 is one of the most studied proteins due to its high incidence of indels and point mutations in clinical screenings (Dolgin, 2017; Lawrence, 2014; Kandoth, 2013). Even more important is that over 80% of the 2300 known missense mutations of p53, including those established in human cancers, are localized to the DBD (Canadillas, 2006; Khoo, 2009). This ‘hot-spot’ feature of p53’s DBD is largely driven by its low thermodynamic and kinetic stability in conjunction with its defining biological activity as a TF (Canadillas, 2006; Khoo, 2009). The availability of structural information and high incidence in point mutations within the DBD makes analysis of mutation outcomes more conclusive than for other domains of p53. A well-characterized set of mutations is R273H/C which compromises DNA contacts while retaining similar thermodynamic stability for the p53 monomer (Bullock, 2000; Joerger, 2006; Ang, 2006; Eldar, 2013). Subsequent analysis revealed that single monomer containing the R273H/C mutants, although capable of forming oligomers, compromises binding interaction for the entire p53 tetramer (Ma and Levine, 2007). Another well-studied point mutation is R175H, which compromised L2-L3 interactions by disrupting the Zn^{2+} coordination (Joerger and Fersht, 2007). Mutations S241F, R248W, and C288F also contribute to structural destabilization of L3 and consequently reduce DNA binding affinity (Joerger and Fersht, 2007). However, not all ‘hotspot’ mutations within DBD result in structural impacts that cannot be compensated for: For example, R248W/Q creates inappropriate interactions between L3 and L1 by increasing L1’s flexibility, but this flexibility has been observed to be compensated for by H115Q counter mutation in L1 (Chen, 1994; Merabet, 2010). This is intriguing as the number of identified cancer-associated mutations

in L1 or any intrinsically dynamic segment is relatively low in p53 (Lukman, 2013). This is consistent with general trend in cancer-associated mutations from other tumor suppressors, as mutations appear to be biased towards segments of rigid structure within critical contact interfaces rather than flexible segments.

Although p53's DBD is principally understood for its key role in the DNA binding and dimerization, it is also subject to PTM regulation observed in all other p53 domains. However, in comparison to TAD and the C-terminal domain, the number of categorized sites and their regulating enzymes is relatively sparse. Known phosphorylation sites within DBD are either structurally detrimental, such as S269^P, or signals for degradation, S149^P, T150^P, and T155^P (Craig, 1999; Schon, 2002; Bech-Otschir, 2001). In contrast, degradative role of acetylation appears to have more nuanced roles within the DBD. K164^{Ac}, which is acetylated by CBP/p300, likely promotes stabilization of p53 (Tang, 2008). Surprisingly, acetylation of K120^{Ac} by either Monocytic Leukemia Zinc finger protein (MOZ), MOF, or TIP60 appears to promote localization of p53 specifically to TSS of the Bax gene, possibly promoting expression of the pro-apoptotic factor (Vainer, 2016).

5.3.4. C-terminal (N-terminally Regulated (NRD); P301-D393) and Tetramerization Domains (D324-A355)

The structure and activities of the C-terminus of the p53 tetramer could be succinctly described as diverse. From a sequence perspective, the C-terminus, or NRD, is intrinsically unstructured within a p53 monomer and is populated with a complex assortment of PTM sites. These sites are often shared between different PTM types that are key to informing stability, intracellular localization, and target gene expression. These PTM regions are bridged by a single

hydrophobic patch devoid of PTM sites entirely. This hydrophobic patch constitutes the tetramerization domain from residues D324-A355. Principally, p53 forms dimers promoted by the self- and DNA-contact surfaces present within the DBD. Upon dimer formation, it is proposed that the p53 tetramerization domain creates a pseudo-stable helical structure within this segment that is driven by hydrophobic interactions (Mora, 2008). It is then posed that the p53 dimer pairs promote the formation of the p53 tetramer which creates a stable helical structure stabilized by 9 hydrophobic and 3 hydrophilic contacts between the helices of each monomer (Mora, 2008; Mateu and Fersht, 1998; McCoy, 1997).

In addition to participation in tetramer formation, other elements of the NRD, although unstructured, have been shown to participate in non-specific DNA-binding (Gu, 1997b; Hupp, 1992; Luo, 2004). This is likely driven by the NRDs numerous lysine residues interacting with phosphate backbone of DNA. Studies of recombinant p53 often resulted in weak DNA binding unless the NRD was occluded by an NRD specific antibody or was phosphorylated or acetylated (Gu, 1997a; Hupp, 1992; Luo, 2004). Subsequent structural analysis revealed that the NRD undergoes little change in structure upon DNA binding by maintaining a highly mobile state and would suggest that these DNA contacts are transient (Ayed, 2001). However, outright deletion of the NRD compromises DNA binding and transcription *in vitro* (McKinney, 2004). This has led to several competing theories concerning the NRD including the promotion of linear diffusion along dsDNA, nucleosome binding to promote remodeling, and recognition of non-linear dsDNA (McKinney, 2004; Tafvizi, 2008; Tafvizi, 2011; Gohler, 2002; McKinney, 2002; Espinosa, 2001). Unfortunately, whether any these activities are definitive is still a point of contention.

As mentioned previously, p53's NRD is subject to multiple PTMs. Interestingly, it is also affected by regulatory interactions of the TAD which is the other most post-translationally

modified region of p53. The best described set of PTMs within the NRD are the bulk acetylation and ubiquitination sites. The ubiquitination sites for residues K370/372/373/381/382/386 are all promoted by HDM2 and they all promote degradation of p53 (Xu, 2003). Sumoylation of K386, p53's sole identified site, increases this HDM2 mediated targeting and degradation with NRD ubiquitination sites (Chen, 2003). Ubiquitination by HDM2 of these sites is also negatively regulated by the phosphorylation status of TAD, as phosphorylation of key residues T18 and S20 interrupts HDM2's anchoring interaction to p53 (Lee, 2007; Craig, 1999; Schon, 2002). Once disentangled from HDM2 ubiquitination of NRD, CBP/p300 acetylates residues K305/320/370/372/373/381/382 or p300/CBP associated factor (PCAF) and TIP60 (alternatively known as histone acetyltransferase KAT5) acetylate residues K317/320 (Wang, 2003; Gu, 1997a; Liu, 1999; Chao, 2006). These sets of acetylation events promote stabilization of p53 but result in disparate outcomes for both intra-cellular localization and gene target activity. Acetylation by CBP/p300 is largely induced through interruption of HDM2 binding by phosphorylation of S15, T18 and S20 in the p53 TAD (Dumaz, 1999; Kalid and Ben-Tal, 2009). Once the NRD is poly-acetylated, p53 promotes complex formation with RSF1 and leads to upregulation of pro-apoptotic gene targets (Liu, 1999). Conversely, acetylation by either PCAF or TIP60 at K317/320 can occur independently of HDM2 regulation and promotes cell-cycle arrest through promotion of exo-nuclear transport of p53 (Tang, 2006; Knights, 2006). This leads to a depletion of nuclear p53 that ultimately prevents S15 phosphorylation ATM/ATR/DNAPKcs, the critical first step in interruption of HDM2 interaction, and sequesters p53 away from HDM2 regulation (Kalid and Ben-Tal, 2009; Dumaz, 1999; Knights, 2006; Loughery, 2014; Lambert, 1998). Beyond the sites of acetylation and ubiquitination, there are 4 serine phosphorylation sites within NRD at S315/376/378/392. S376/378 phosphorylation events appear to be constitutive and may participate

in regulating unnecessary interactions for proximal ubiquitination/acetylation sites (Gatti, 2000; Waterman, 1998; Li, 2004). Conversely, S315 phosphorylation ultimately promotes degradation by HDM2, possibly through interruption of PCAF/TIP60/CBP/p300 regulation (Katayama, 2004). The most unique phosphorylation site is S392 as it has been shown to aid in tetramer stabilization although the mechanism is unclear (Sakaguchi, 1998). Methylation of the NRD also plays a key role in regulating p53 as K370/372/373/382 are known sites of mono- and di-methylation (Huang, 2006; Chuikov, 2004; Huang, 2010; Shi, 2007; Kachirskaia, 2008). Although these sites are tightly clustered, unique methylation state result in wildly different outcomes for p53. Mono-methylation of K370 by SMYD2 inhibits localization to target pro-apoptotic genes promoters, while mono- or di-methylation of K372 by SETD7 promotes stability and apoptotic outcomes (Huang, 2006; Chuikov, 2004; Campaner, 2011; Lehnertz, 2011). In contrast, mono-methylation of K373 by either G9a or G9a-like Protein (GLP) represses the p53 TF activity uniformly (Huang, 2010). Methylation of K382 is perhaps the most unique in that the mono- and di-methylation states have two distinct effects. Mono-methylation of K382 inhibits localization to specific gene sets that include both cell-cycle arrest (CDKN1A) and pro-apoptosis (PUMA) regulators. Di-methylation of K382, in addition to its mono-methylation effects, promotes association with 53BP1 and is proposed to stabilize p53 by acting as a cap over the NRD to shield it from HDM2 ubiquitination (Kachirskaia, 2008).

5.4. References

- Abramova, N. A., Russell, J., Botchan, M., & Li, R. (1997). Interaction between replication protein A and p53 is disrupted after UV damage in a DNA repair-dependent manner. *Proceedings of the National Academy of Sciences*, *94*(14), 7186-7191.
- Ang, H. C., Joerger, A. C., Mayer, S., & Fersht, A. R. (2006). Effects of common cancer mutations on stability and DNA binding of full-length p53 compared with isolated core domains. *Journal of Biological Chemistry*, *281*(31), 21934-21941.
- Avantaggiati, M. L., Ogryzko, V., Gardner, K., Giordano, A., Levine, A. S., & Kelly, K. (1997). Recruitment of p300/CBP in p53-dependent signal pathways. *Cell*, *89*(7), 1175-1184.
- Ayed, A., Mulder, F. A., Yi, G. S., Lu, Y., Kay, L. E., & Arrowsmith, C. H. (2001). Latent and active p53 are identical in conformation. *Nature structural biology*, *8*(9), 756-760.
- Badis, G., Berger, M. F., Philippakis, A. A., Talukder, S., Gehrke, A. R., Jaeger, S. A., Chan, E.T., Metzler, G., Vedenko, A., Chen, X., & Kuznetsov, H. (2009). Diversity and complexity in DNA recognition by transcription factors. *Science*, *324*(5935), 1720-1723.
- Baker, S. J., Fearon, E. R., Nigro, J. M., Preisinger, A. C., Jessup, J. M., Ledbetter, D. H., Barker, D.F., Nakamura, Y., White, R., & Vogelstein, B. (1989). Chromosome 17 deletions and p53 gene mutations in colorectal carcinomas. *Science*, *244*(4901), 217-221.
- Bao, F., LoVerso, P. R., Fisk, J. N., Zhurkin, V. B., & Cui, F. (2017). p53 binding sites in normal and cancer cells are characterized by distinct chromatin context. *Cell Cycle*, *16*(21), 2073-2085.

- Bech-Otschir, D., Kraft, R., Huang, X., Henklein, P., Kapelari, B., Pollmann, C., & Dubiel, W. (2001). COP9 signalosome-specific phosphorylation targets p53 to degradation by the ubiquitin system. *The EMBO journal*, *20*(7), 1630-1639.
- Berger, M., Sionov, R. V., Levine, A. J., & Haupt, Y. (2001). A role for the polyproline domain of p53 in its regulation by Mdm2. *Journal of Biological Chemistry*, *276*(6), 3785-3790.
- Berger, M., Stahl, N., Del Sal, G., & Haupt, Y. (2005). Mutations in proline 82 of p53 impair its activation by Pin1 and Chk2 in response to DNA damage. *Molecular and cellular biology*, *25*(13), 5380-5388.
- Bewley, C. A., Gronenborn, A. M., & Clore, G. M. (1998). Minor groove-binding architectural proteins: structure, function, and DNA recognition. *Annual review of biophysics and biomolecular structure*, *27*(1), 105-131.
- Botcheva, K., McCorkle, S. R., McCombie, W. R., Dunn, J. J., & Anderson, C. W. (2011). Distinct p53 genomic binding patterns in normal and cancer-derived human cells. *Cell cycle*, *10*(24), 4237-4249.
- Bullock, A. N., Henckel, J., & Fersht, A. R. (2000). Quantitative analysis of residual folding and DNA binding in mutant p53 core domain: definition of mutant states for rescue in cancer therapy. *Oncogene*, *19*(10), 1245-1256.
- Buschmann, T., Potapova, O., Bar-Shira, A., Ivanov, V. N., Fuchs, S. Y., Henderson, S., Fried, V.A., Minamoto, T., Alarcon-Vargas, D., Pincus, M.R., & Gaarde, W. A. (2001). Jun NH2-terminal kinase phosphorylation of p53 on Thr-81 is important for p53 stabilization and transcriptional activities in response to stress. *Molecular and cellular biology*, *21*(8), 2743-2754.

- Cañadillas, J. M. P., Tidow, H., Freund, S. M., Rutherford, T. J., Ang, H. C., & Fersht, A. R. (2006). Solution structure of p53 core domain: structural basis for its instability. *Proceedings of the National Academy of Sciences*, *103*(7), 2109-2114.
- Campaner, S., Spreafico, F., Burgold, T., Doni, M., Rosato, U., Amati, B., & Testa, G. (2011). The methyltransferase Set7/9 (Setd7) is dispensable for the p53-mediated DNA damage response in vivo. *Molecular cell*, *43*(4), 681-688.
- Chao, C., Wu, Z., Mazur, S. J., Borges, H., Rossi, M., Lin, T., Wang, J.Y., Anderson, C.W., Appella, E., & Xu, Y. (2006). Acetylation of mouse p53 at lysine 317 negatively regulates p53 apoptotic activities after DNA damage. *Molecular and cellular biology*, *26*(18), 6859-6869.
- Chang, G. S., Chen, X. A., Park, B., Rhee, H. S., Li, P., Han, K. H., Mishra, T., Chan-Salis, K.Y., Li, Y., Hardison, R.C., & Wang, Y. (2014). A comprehensive and high-resolution genome-wide response of p53 to stress. *Cell reports*, *8*(2), 514-527.
- Chen, J., Marechal, V., & Levine, A. J. (1993). Mapping of the p53 and mdm-2 interaction domains. *Molecular and cellular biology*, *13*(7), 4107-4114.
- Chen, L., & Chen, J. (2003). MDM2-ARF complex regulates p53 sumoylation. *Oncogene*, *22*(34), 5348-5357.
- Chen, X., Ko, L. J., Jayaraman, L., & Prives, C. (1996). p53 levels, functional domains, and DNA damage determine the extent of the apoptotic response of tumor cells. *Genes & development*, *10*(19), 2438-2451.

- Cho, Y., Gorina, S., Jeffrey, P. D., & Pavletich, N. P. (1994). Crystal structure of a p53 tumor suppressor-DNA complex: understanding tumorigenic mutations. *Science*, 265(5170), 346-355.
- Chuikov, S., Kurash, J. K., Wilson, J. R., Xiao, B., Justin, N., Ivanov, G. S., McKinney, K., Tempst, P., Prives, C., Gamblin, S.J., & Barlev, N. A. (2004). Regulation of p53 activity through lysine methylation. *Nature*, 432(7015), 353-360.
- Cordenonsi, M., Montagner, M., Adorno, M., Zacchigna, L., Martello, G., Mamidi, A., Soligo, S., Dupont, S., & Piccolo, S. (2007). Integration of TGF- β and Ras/MAPK signaling through p53 phosphorylation. *Science*, 315(5813), 840-843.
- Craig, A. L., Burch, L., Vojtesek, B., Mikutowska, J., Thompson, A., & Hupp, T. R. (1999). Novel phosphorylation sites of human tumour suppressor protein p53 at Ser20 and Thr18 that disrupt the binding of mdm2 (mouse double minute 2) protein are modified in human cancers. *Biochemical Journal*, 342(1), 133-141.
- Crawford, L. (1983). The 53,000-dalton cellular protein and its role in transformation. *International review of experimental pathology*, 25, 1-50.
- Cui, F., Sirotin, M. V., & Zhurkin, V. B. (2011). Impact of Alu repeats on the evolution of human p53 binding sites. *Biology direct*, 6(1), 2.
- Dawson, R., Müller, L., Dehner, A., Klein, C., Kessler, H., & Buchner, J. (2003). The N-terminal domain of p53 is natively unfolded. *Journal of molecular biology*, 332(5), 1131-1141.
- DeLeo, A. B., Jay, G., Appella, E., Dubois, G. C., Law, L. W., & Old, L. J. (1979). Detection of a transformation-related antigen in chemically induced sarcomas and other transformed cells of the mouse. *Proceedings of the National Academy of Sciences*, 76(5), 2420-2424.

- Dolgin, E. (2017). The most popular genes in the human genome. *Nature*, *551*(7681), 427.
- Dornan, D., Shimizu, H., Burch, L., Smith, A. J., & Hupp, T. R. (2003). The proline repeat domain of p53 binds directly to the transcriptional coactivator p300 and allosterically controls DNA-dependent acetylation of p53. *Molecular and cellular biology*, *23*(23), 8846-8861.
- Duan, J., & Nilsson, L. (2006). Effect of Zn²⁺ on DNA recognition and stability of the p53 DNA-binding domain. *Biochemistry*, *45*(24), 7483-7492.
- Dumaz, N., Milne, D. M., & Meek, D. W. (1999). Protein kinase CK1 is a p53-threonine 18 kinase which requires prior phosphorylation of serine 15. *FEBS letters*, *463*(3), 312-316.
- Dutta, A., Ruppert, J. M., Aster, J. C., & Winchester, E. (1993). Inhibition of DNA replication factor RPA by p53. *Nature*, *365*(6441), 79-82.
- Eldar, A., Rozenberg, H., Diskin-Posner, Y., Rohs, R., & Shakked, Z. (2013). Structural studies of p53 inactivation by DNA-contact mutations and its rescue by suppressor mutations via alternative protein–DNA interactions. *Nucleic acids research*, *41*(18), 8748-8759.
- Eliyahu, D., Raz, A., Gruss, P., Givol, D., & Oren, M. (1984). Participation of p53 cellular tumour antigen in transformation of normal embryonic cells. *Nature*, *312*(5995), 646-649.
- El-Deiry, W. S., Kern, S. E., Pietenpol, J. A., Kinzler, K. W., & Vogelstein, B. (1992). Definition of a consensus binding site for p53. *Nature genetics*, *1*(1), 45-49.
- Espinosa, J. M., & Emerson, B. M. (2001). Transcriptional regulation by p53 through intrinsic DNA/chromatin binding and site-directed cofactor recruitment. *Molecular cell*, *8*(1), 57-69.

- Fedotova, A. A., Bonchuk, A. N., Mogila, V. A., & Georgiev, P. G. (2017). C2H2 zinc finger proteins: the largest but poorly explored family of higher eukaryotic transcription factors. *Acta Naturae (English version)*, 9(2 (33)).
- Fischer, M. (2017). Census and evaluation of p53 target genes. *Oncogene*, 36(28), 3943-3956.
- Funk, W. D., Pak, D. T., Karas, R. H., Wright, W. E., & Shay, J. W. (1992). A transcriptionally active DNA-binding site for human p53 protein complexes. *Molecular and cellular biology*, 12(6), 2866-2871.
- Ganguly, D., & Chen, J. (2015). Modulation of the disordered conformational ensembles of the p53 transactivation domain by cancer-associated mutations. *PLoS computational biology*, 11(4).
- Gatti, A., Li, H. H., Traugh, J. A., & Liu, X. (2000). Phosphorylation of human p53 on Thr-55. *Biochemistry*, 39(32), 9837-9842.
- Göhler, T., Reimann, M., Cherny, D., Walter, K., Warnecke, G., Kim, E., & Deppert, W. (2002). Specific interaction of p53 with target binding sites is determined by DNA conformation and is regulated by the C-terminal domain. *Journal of Biological Chemistry*, 277(43), 41192-41203.
- Grossman, S. R., Perez, M., Kung, A. L., Joseph, M., Mansur, C., Xiao, Z. X., Kumar, S., Howley, P.M., & Livingston, D. M. (1998). p300/MDM2 complexes participate in MDM2-mediated p53 degradation. *Molecular cell*, 2(4), 405-415.
- Gu, W., Shi, X. L., & Roeder, R. G. (1997a). Synergistic activation of transcription by CBP and p53. *Nature*, 387(6635), 819-823.

- Gu, W., & Roeder, R. G. (1997b). Activation of p53 sequence-specific DNA binding by acetylation of the p53 C-terminal domain. *Cell*, 90(4), 595-606.
- Hafner, A., Lahav, G., & Stewart-Ornstein, J. (2017). Stereotyped p53 binding tuned by chromatin accessibility. *bioRxiv*, 177667.
- Hafner, A., Bulyk, M. L., Jambhekar, A., & Lahav, G. (2019). The multiple mechanisms that regulate p53 activity and cell fate. *Nature Reviews Molecular Cell Biology*, 20(4), 199-210.
- Hager, K. M., & Gu, W. (2014). Understanding the non-canonical pathways involved in p53-mediated tumor suppression. *Carcinogenesis*, 35(4), 740-746.
- Hainaut, P., & Hollstein, M. (1999). p53 and human cancer: the first ten thousand mutations. In *Advances in cancer research* (Vol. 77, pp. 81-137). Academic Press.
- Haupt, Y., Maya, R., Kazaz, A., & Oren, M. (1997). Mdm2 promotes the rapid degradation of p53. *Nature*, 387(6630), 296-299.
- Hinrichs, W., Kisker, C., Duvel, M., Muller, A., Tovar, K., Hillen, W., & Saenger, W. (1994). Structure of the Tet repressor-tetracycline complex and regulation of antibiotic resistance. *Science*, 264(5157), 418-420.
- Honda, R., Tanaka, H., & Yasuda, H. (1997). Oncoprotein MDM2 is a ubiquitin ligase E3 for tumor suppressor p53. *FEBS letters*, 420(1), 25-27.
- Huang, J., Perez-Burgos, L., Placek, B. J., Sengupta, R., Richter, M., Dorsey, J. A., Kubicek, S., Opravil, S., Jenuwein, T., & Berger, S. L. (2006). Repression of p53 activity by Smyd2-mediated methylation. *Nature*, 444(7119), 629-632.

- Huang, J., Dorsey, J., Chuikov, S., Zhang, X., Jenuwein, T., Reinberg, D., & Berger, S. L. (2010). G9a and Glp methylate lysine 373 in the tumor suppressor p53. *Journal of Biological Chemistry*, 285(13), 9636-9641.
- Hupp, T. R., Meek, D. W., Midgley, C. A., & Lane, D. P. (1992). Regulation of the specific DNA binding function of p53. *Cell*, 71(5), 875-886.
- Inga, A., Storici, F., Darden, T. A., & Resnick, M. A. (2002). Differential transactivation by the p53 transcription factor is highly dependent on p53 level and promoter target sequence. *Molecular and cellular biology*, 22(24), 8612-8625.
- Jenkins, J. R., Rudge, K., & Currie, G. A. (1984). Cellular immortalization by a cDNA clone encoding the transformation-associated phosphoprotein p53. *Nature*, 312(5995), 651-654.
- Joerger, A. C., Allen, M. D., & Fersht, A. R. (2004). Crystal Structure of a Superstable Mutant of Human p53 Core Domain Insights Into The Mechanism Of Rescuing Oncogenic Mutations. *Journal of Biological Chemistry*, 279(2), 1291-1296.
- Joerger, A. C., Ang, H. C., & Fersht, A. R. (2006). Structural basis for understanding oncogenic p53 mutations and designing rescue drugs. *Proceedings of the National Academy of Sciences*, 103(41), 15056-15061.
- Joerger, A. C., & Fersht, A. R. (2007). Structure–function–rescue: the diverse nature of common p53 cancer mutants. *Oncogene*, 26(15), 2226-2242.
- Joseph, T. L., Madhumalar, A., Brown, C. J., Lane, D. P., & Verma, C. S. (2010). Differential binding of p53 and nutlin to MDM2 and MDMX: Computational studies. *Cell cycle*, 9(6), 1167-1181.

- Juven, T., Barak, Y., Zauberman, A., George, D. L., & Oren, M. (1993). Wild type p53 can mediate sequence-specific transactivation of an internal promoter within the mdm2 gene. *Oncogene*, 8(12), 3411-3416.
- Kachirskiaia, I., Shi, X., Yamaguchi, H., Tanoue, K., Wen, H., Wang, E. W., Appella, E., & Gozani, O. (2008). Role for 53BP1 Tudor domain recognition of p53 dimethylated at lysine 382 in DNA damage signaling. *Journal of Biological Chemistry*, 283(50), 34660-34666.
- Kaesler, M. D., & Iggo, R. D. (2002). Chromatin immunoprecipitation analysis fails to support the latency model for regulation of p53 DNA binding activity in vivo. *Proceedings of the National Academy of Sciences*, 99(1), 95-100.
- Kalid, O., & Ben-Tal, N. (2009). Study of MDM2 binding to p53-analogues: affinity, helicity, and applicability to drug design. *Journal of chemical information and modeling*, 49(4), 865-876.
- Kandoth, C., McLellan, M. D., Vandin, F., Ye, K., Niu, B., Lu, C., Xie, M., Zhang, Q., McMichael, J.F., Wyczalkowski, M.A., & Leiserson, M. D. (2013). Mutational landscape and significance across 12 major cancer types. *Nature*, 502(7471), 333-339.
- Kang, R., Kroemer, G., & Tang, D. (2019). The tumor suppressor protein p53 and the ferroptosis network. *Free Radical Biology and Medicine*, 133, 162-168.
- Kannan, K., Kaminski, N., Rechavi, G., Jakob-Hirsch, J., Amariglio, N., & Givol, D. (2001). DNA microarray analysis of genes involved in p53 mediated apoptosis: activation of Apaf-1. *Oncogene*, 20(26), 3449-3455.

- Katayama, H., Sasai, K., Kawai, H., Yuan, Z. M., Bondaruk, J., Suzuki, F., Fujii, S., Arlinghaus, R.B., Czerniak, B.A., & Sen, S. (2004). Phosphorylation by aurora kinase A induces Mdm2-mediated destabilization and inhibition of p53. *Nature genetics*, 36(1), 55-62.
- Khoo, K. H., Andreeva, A., & Fersht, A. R. (2009). Adaptive evolution of p53 thermodynamic stability. *Journal of molecular biology*, 393(1), 161-175.
- Kim, J. L., Nikolov, D. B., & Burley, S. K. (1993a). Co-crystal structure of TBP recognizing the minor groove of a TATA element. *Nature*, 365(6446), 520-527.
- Kim, Y., Geiger, J. H., Hahn, S., & Sigler, P. B. (1993b). Crystal structure of a yeast TBP/TATA-box complex. *Nature*, 365(6446), 512-520.
- Kim, J. L., & Burley, S. K. (1994). 1.9 Å resolution refined structure of TBP recognizing the minor groove of TATAAAAG. *Nature structural biology*, 1(9), 638-653.
- Kitayner, M., Rozenberg, H., Kessler, N., Rabinovich, D., Shaulov, L., Haran, T. E., & Shakked, Z. (2006). Structural basis of DNA recognition by p53 tetramers. *Molecular cell*, 22(6), 741-753.
- Kitayner, M., Rozenberg, H., Rohs, R., Suad, O., Rabinovich, D., Honig, B., & Shakked, Z. (2010). Diversity in DNA recognition by p53 revealed by crystal structures with Hoogsteen base pairs. *Nature structural & molecular biology*, 17(4), 423.
- Knights, C. D., Catania, J., Giovanni, S. D., Muratoglu, S., Perez, R., Swartzbeck, A., Quong, A.A., Zhang, X., Beerman, T., Pestell, R.G., & Avantaggiati, M. L. (2006). Distinct p53 acetylation cassettes differentially influence gene-expression patterns and cell fate. *The Journal of cell biology*, 173(4), 533-544.

- Koepfel, M., van Heeringen, S. J., Kramer, D., Smeenk, L., Janssen-Megens, E., Hartmann, M., Stunnenberg, H.G., & Lohrum, M. (2011). Crosstalk between c-Jun and TAp73 α/β contributes to the apoptosis–survival balance. *Nucleic acids research*, 39(14), 6069-6085.
- Kracikova, M., Akiri, G., George, A., Sachidanandam, R., & Aaronson, S. A. (2013). A threshold mechanism mediates p53 cell fate decision between growth arrest and apoptosis. *Cell Death & Differentiation*, 20(4), 576-588.
- Kress, M., May, E., Cassingena, R., & May, P. (1979). Simian virus 40-transformed cells express new species of proteins precipitable by anti-simian virus 40 tumor serum. *Journal of virology*, 31(2), 472-483.
- Kruiswijk, F., Labuschagne, C. F., & Vousden, K. H. (2015). p53 in survival, death and metabolic health: a lifeguard with a licence to kill. *Nature reviews Molecular cell biology*, 16(7), 393-405.
- Kubbutat, M. H., Jones, S. N., & Vousden, K. H. (1997). Regulation of p53 stability by Mdm2. *Nature*, 387(6630), 299-303.
- Kussie, P. H., Gorina, S., Marechal, V., Elenbaas, B., Moreau, J., Levine, A. J., & Pavletich, N. P. (1996). Structure of the MDM2 oncoprotein bound to the p53 tumor suppressor transactivation domain. *Science*, 274(5289), 948-953.
- Lambert, P. F., Kashanchi, F., Radonovich, M. F., Shiekhattar, R., & Brady, J. N. (1998). Phosphorylation of p53 serine 15 increases interaction with CBP. *Journal of Biological Chemistry*, 273(49), 33048-33053.
- Lane, D. P., & Crawford, L. V. (1979). T antigen is bound to a host protein in SY40-transformed cells. *Nature*, 278(5701), 261-263.

- Latchman, D. S. (1993). Transcription factors: an overview. *International journal of experimental pathology*, 74(5), 417.
- Lawrence, M. S., Stojanov, P., Mermel, C. H., Robinson, J. T., Garraway, L. A., Golub, T. R., Meyerson, M., Gabriel, S.B., Lander, E.S., & Getz, G. (2014). Discovery and saturation analysis of cancer genes across 21 tumour types. *Nature*, 505(7484), 495-501.
- Lebrun, A., & Lavery, R. (1999). Modeling DNA deformations induced by minor groove binding proteins. *Biopolymers: Original Research on Biomolecules*, 49(5), 341-353.
- Lee, H., Mok, K. H., Muhandiram, R., Park, K. H., Suk, J. E., Kim, D. H., Chang, J., Sung, Y.C., Choi, K.Y., & Han, K. H. (2000). Local structural elements in the mostly unstructured transcriptional activation domain of human p53. *Journal of Biological Chemistry*, 275(38), 29426-29432.
- Lee, M. S., Lee, S. O., Lee, M. K., Yi, G. S., Lee, C. K., Ryu, K. S., & Chi, S. W. (2019). Solution structure of MUL1-RING domain and its interaction with p53 transactivation domain. *Biochemical and biophysical research communications*, 516(2), 533-539.
- Lehnertz, B., Rogalski, J. C., Schulze, F. M., Yi, L., Lin, S., Kast, J., & Rossi, F. M. (2011). p53-dependent transcription and tumor suppression are not affected in Set7/9-deficient mice. *Molecular cell*, 43(4), 673-680.
- Levine, A. J. (1989). The p53 proto-oncogene can act as a suppressor of transformation. *Cell*, 57(7), 1083-1093.
- Li, H. H., Li, A. G., Sheppard, H. M., & Liu, X. (2004). Phosphorylation on Thr-55 by TAF1 mediates degradation of p53: a role for TAF1 in cell G1 progression. *Molecular cell*, 13(6), 867-878.

- Li, R., & Botchan, M. R. (1993). The acidic transcriptional activation domains of VP16 and p53 bind the cellular replication protein A and stimulate in vitro BPV-1 DNA replication. *Cell*, 73(6), 1207-1221.
- Lill, N. L., Grossman, S. R., Ginsberg, D., DeCaprio, J., & Livingston, D. M. (1997). Binding and modulation of p53 by p300/CBP coactivators. *Nature*, 387(6635), 823-827.
- Linzer, D. I., & Levine, A. J. (1979). Characterization of a 54K dalton cellular SV40 tumor antigen present in SV40-transformed cells and uninfected embryonal carcinoma cells. *Cell*, 17(1), 43-52.
- Liu, L., Scolnick, D. M., Trievel, R. C., Zhang, H. B., Marmorstein, R., Halazonetis, T. D., & Berger, S. L. (1999). p53 sites acetylated in vitro by PCAF and p300 are acetylated in vivo in response to DNA damage. *Molecular and cellular biology*, 19(2), 1202-1209.
- Liu, W., Xie, Y., Ma, J., Luo, X., Nie, P., Zuo, Z., Lahrmann, U., Zhao, Q., Zheng, Y., Zhao, Y., & Xue, Y. (2015). IBS: an illustrator for the presentation and visualization of biological sequences. *Bioinformatics*, 31(20), 3359-3361.
- Livengood, J. A., Scoggin, K. E., Van Orden, K., McBryant, S. J., Edayathumangalam, R. S., Laybourn, P. J., & Nyborg, J. K. (2002). p53 Transcriptional activity is mediated through the SRC1-interacting domain of CBP/p300. *Journal of Biological Chemistry*, 277(11), 9054-9061.
- Loughery, J., Cox, M., Smith, L. M., & Meek, D. W. (2014). Critical role for p53-serine 15 phosphorylation in stimulating transactivation at p53-responsive promoters. *Nucleic acids research*, 42(12), 7666-7680.

- Love, J. J., Li, X., Case, D. A., Giese, K., Grosschedl, R., & Wright, P. E. (1995). Structural basis for DNA bending by the architectural transcription factor LEF-1. *Nature*, 376(6543), 791-795.
- Lu, Q., Tan, Y. H., & Luo, R. (2007). Molecular dynamics simulations of p53 DNA-binding domain. *The Journal of Physical Chemistry B*, 111(39), 11538-11545.
- Lubin, D. J., Butler, J. S., & Loh, S. N. (2010). Folding of tetrameric p53: oligomerization and tumorigenic mutations induce misfolding and loss of function. *Journal of molecular biology*, 395(4), 705-716.
- Lukman, S., Lane, D. P., & Verma, C. S. (2013). Mapping the structural and dynamical features of multiple p53 DNA binding domains: insights into loop 1 intrinsic dynamics. *PLoS one*, 8(11).
- Luo, J., Li, M., Tang, Y., Laszkowska, M., Roeder, R. G., & Gu, W. (2004). Acetylation of p53 augments its site-specific DNA binding both in vitro and in vivo. *Proceedings of the National Academy of Sciences*, 101(8), 2259-2264.
- Ma, B., & Levine, A. J. (2007). Probing potential binding modes of the p53 tetramer to DNA based on the symmetries encoded in p53 response elements. *Nucleic acids research*, 35(22), 7733-7747.
- Madden, S. L., Galella, E. A., Zhu, J., Bertelsen, A. H., & Beaudry, G. A. (1997). SAGE transcript profiles for p53-dependent growth regulation. *Oncogene*, 15(9), 1079-1085.
- Matthews, B. W., Ohlendorf, D. H., Anderson, W. F., & Takeda, Y. (1982). Structure of the DNA-binding region of lac repressor inferred from its homology with cro repressor. *Proceedings of the National Academy of Sciences*, 79(5), 1428-1432.

- McKinney, K., & Prives, C. (2002). Efficient specific DNA binding by p53 requires both its central and C-terminal domains as revealed by studies with high-mobility group 1 protein. *Molecular and cellular biology*, 22(19), 6797-6808.
- McKinney, K., Mattia, M., Gottifredi, V., & Prives, C. (2004). p53 linear diffusion along DNA requires its C terminus. *Molecular cell*, 16(3), 413-424.
- Merabet, A., Houlleberghs, H., Maclagan, K., Akanho, E., Bui, T. T., Pagano, B., Drake, A.F., Fraternali, F., & Nikolova, P. V. (2010). Mutants of the tumour suppressor p53 L1 loop as second-site suppressors for restoring DNA binding to oncogenic p53 mutations: structural and biochemical insights. *Biochemical Journal*, 427(2), 225-236.
- Miller, S. D., Moses, K., Jayaraman, L., & Prives, C. (1997). Complex formation between p53 and replication protein A inhibits the sequence-specific DNA binding of p53 and is regulated by single-stranded DNA. *Molecular and cellular biology*, 17(4), 2194-2201.
- Min, S., Kim, K., Kim, S. G., Cho, H., & Lee, Y. (2018). Chromatin-remodeling factor, RSF1, controls p53-mediated transcription in apoptosis upon DNA strand breaks. *Cell death & disease*, 9(11), 1-13.
- Mirza, A., Wu, Q., Wang, L., McClanahan, T., Bishop, W. R., Gheyas, F., Ding, W., Hutchins, B., Hockenberry, T., Kirschmeier, P. & Greene, J. R. (2003). Global transcriptional program of p53 target genes during the process of apoptosis and cell cycle progression. *Oncogene*, 22(23), 3645-3654.
- Mitchell, P. J., & Tjian, R. (1989). Transcriptional regulation in mammalian cells by sequence-specific DNA binding proteins. *Science*, 245(4916), 371-378.

- Mora, P., Carbajo, R. J., Pineda-Lucena, A., Sanchez del Pino, M. M., & Pérez-Payá, E. (2008). Solvent-exposed residues located in the β -sheet modulate the stability of the tetramerization domain of p53—A structural and combinatorial approach. *Proteins: Structure, Function, and Bioinformatics*, 71(4), 1670-1685.
- Nguyen, T. A. T., Grimm, S. A., Bushel, P. R., Li, J., Li, Y., Bennett, B. D., Lavender, C.A., Ward, J.M., Fargo, D.C., Anderson, C.W., & Li, L. (2018). Revealing a human p53 universe. *Nucleic acids research*, 46(16), 8153-8167.
- Noureddine, M. A., Menendez, D., Campbell, M. R., Bandele, O. J., Horvath, M. M., Wang, X., Pittman, G.S., Chorley, B.N., Resnick, M.A. & Bell, D. A. (2009). Probing the functional impact of sequence variation on p53-DNA interactions using a novel microsphere assay for protein-DNA binding with human cell extracts. *PLoS genetics*, 5(5).
- Nikolov, D. B., Chen, H., Halay, E. D., Hoffman, A., Roeder, R. G., & Burley, S. K. (1996). Crystal structure of a human TATA box-binding protein/TATA element complex. *Proceedings of the National Academy of Sciences*, 93(10), 4862-4867.
- Ogata, K., Hojo, H., Aimoto, S., Nakai, T., Nakamura, H., Sarai, A., Ishii, S., & Nishimura, Y. (1992). Solution structure of a DNA-binding unit of Myb: a helix-turn-helix-related motif with conserved tryptophans forming a hydrophobic core. *Proceedings of the National Academy of Sciences*, 89(14), 6428-6432.
- Oliner, J. D., Pietenpol, J. A., Thiagalingam, S., Gyuris, J., Kinzler, K. W., & Vogelstein, B. (1993). Oncoprotein MDM2 conceals the activation domain of tumour suppressor p53. *Nature*, 362(6423), 857-860.

- Pan, Y., & Nussinov, R. (2007). Structural basis for p53 binding-induced DNA bending. *Journal of Biological Chemistry*, 282(1), 691-699.
- Parada, L. F., Land, H., Weinberg, R. A., Wolf, D., & Rotter, V. (1984). Cooperation between gene encoding p53 tumour antigen and ras in cellular transformation. *Nature*, 312(5995), 649-651.
- Peng, J., Ren, K. D., Yang, J., & Luo, X. J. (2016). Mitochondrial E3 ubiquitin ligase 1: a key enzyme in regulation of mitochondrial dynamics and functions. *Mitochondrion*, 28, 49-53.
- Ptashne, M. (1988). How eukaryotic transcriptional activators work. *Nature*, 335(6192), 683-689.
- Qian, H., Wang, T., Naumovski, L., Lopez, C. D., & Brachmann, R. K. (2002). Groups of p53 target genes involved in specific p53 downstream effects cluster into different classes of DNA binding sites. *Oncogene*, 21(51), 7901-7911.
- Religa, T. L., Johnson, C. M., Vu, D. M., Brewer, S. H., Dyer, R. B., & Fersht, A. R. (2007). The helix–turn–helix motif as an ultrafast independently folding domain: The pathway of folding of Engrailed homeodomain. *Proceedings of the National Academy of Sciences*, 104(22), 9272-9277.
- Riley, T., Sontag, E., Chen, P., & Levine, A. (2008). Transcriptional control of human p53-regulated genes. *Nature reviews Molecular cell biology*, 9(5), 402-412.
- Rouault, J. P., Falette, N., Guéhenneux, F., Guillot, C., Rimokh, R., Wang, Q., Berthet, C., Moyret-Lalle, C., Savatier, P., Pain, B., & Shaw, P. (1996). Identification of BTG2, an

- antiproliferative p53-dependent component of the DNA damage cellular response pathway. *Nature genetics*, 14(4), 482-486.
- Sakaguchi, K., Herrera, J. E., Saito, S. I., Miki, T., Bustin, M., Vassilev, A., Anderson, C.W., & Appella, E. (1998). DNA damage activates p53 through a phosphorylation-acetylation cascade. *Genes & development*, 12(18), 2831-2841.
- Schlereth, K., Heyl, C., Krampitz, A. M., Mernberger, M., Finkernagel, F., Scharfe, M., Jarek, M., Leich, E., Rosenwald, A., & Stiewe, T. (2013). Characterization of the p53 cistrome-DNA binding cooperativity dissects p53's tumor suppressor functions. *PLoS genetics*, 9(8).
- Schon, O., Friedler, A., Bycroft, M., Freund, S. M., & Fersht, A. R. (2002). Molecular mechanism of the interaction between MDM2 and p53. *Journal of molecular biology*, 323(3), 491-501.
- Schumacher, M. A., Choi, K. Y., Zalkin, H., & Brennan, R. G. (1994). Crystal structure of LacI member, PurR, bound to DNA: minor groove binding by alpha helices. *Science*, 266(5186), 763-770.
- Scolnick, D. M., Chehab, N. H., Stavridi, E. S., Lien, M. C., Caruso, L., Moran, E., Berger, S.L., & Halazonetis, T. D. (1997). CREB-binding protein and p300/CBP-associated factor are transcriptional coactivators of the p53 tumor suppressor protein. *Cancer research*, 57(17), 3693-3696.
- Shi, X., Kachirskaja, I., Yamaguchi, H., West, L. E., Wen, H., Wang, E. W., Dutta, S., Appella, E., & Gozani, O. (2007). Modulation of p53 function by SET8-mediated methylation at lysine 382. *Molecular cell*, 27(4), 636-646.

- Shieh, S. Y., Ikeda, M., Taya, Y., & Prives, C. (1997). DNA damage-induced phosphorylation of p53 alleviates inhibition by MDM2. *Cell*, *91*(3), 325-334.
- Shiloh, Y., & Ziv, Y. (2013). The ATM protein kinase: regulating the cellular response to genotoxic stress, and more. *Nature reviews Molecular cell biology*, *14*(4), 197-210.
- Smeenk, L., van Heeringen, S. J., Koeppel, M., van Driel, M. A., Bartels, S. J., Akkers, R. C., Denissov, S., Stunnenberg, H.G., & Lohrum, M. (2008). Characterization of genome-wide p53-binding sites upon stress response. *Nucleic acids research*, *36*(11), 3639-3654.
- Smeenk, L., Van Heeringen, S. J., Koeppel, M., Gilbert, B., Janssen-Megens, E., Stunnenberg, H. G., & Lohrum, M. (2011). Role of p53 serine 46 in p53 target gene regulation. *PLoS one*, *6*(3).
- Stegmaier, P., Kel, A. E., & Wingender, E. (2004). Systematic DNA-binding domain classification of transcription factors. *Genome Informatics*, *15*(2), 276-286.
- Su, D., Wang, X., Campbell, M. R., Song, L., Safi, A., Crawford, G. E., & Bell, D. A. (2015). Interactions of chromatin context, binding site sequence content, and sequence evolution in stress-induced p53 occupancy and transactivation. *PLoS genetics*, *11*(1).
- Sullivan, K. D., Galbraith, M. D., Andrysik, Z., & Espinosa, J. M. (2018). Mechanisms of transcriptional regulation by p53. *Cell Death & Differentiation*, *25*(1), 133-143.
- Tafvizi, A., Huang, F., Leith, J. S., Fersht, A. R., Mirny, L. A., & Van Oijen, A. M. (2008). Tumor suppressor p53 slides on DNA with low friction and high stability. *Biophysical journal*, *95*(1), L01-L03.

- Tafvizi, A., Huang, F., Fersht, A. R., Mirny, L. A., & van Oijen, A. M. (2011). A single-molecule characterization of p53 search on DNA. *Proceedings of the National Academy of Sciences*, *108*(2), 563-568.
- Takahashi, T., Nau, M. M., Chiba, I., Birrer, M. J., Rosenberg, R. K., Vinocour, M., Levitt, M., Pass, H., Gazdar, A.F., & Minna, J. D. (1989). p53: a frequent target for genetic abnormalities in lung cancer. *Science*, *246*(4929), 491-494.
- Tang, Y., Luo, J., Zhang, W., & Gu, W. (2006). Tip60-dependent acetylation of p53 modulates the decision between cell-cycle arrest and apoptosis. *Molecular cell*, *24*(6), 827-839.
- Tang, Y., Zhao, W., Chen, Y., Zhao, Y., & Gu, W. (2008). Acetylation is indispensable for p53 activation. *Cell*, *133*(4), 612-626.
- Tonelli, C., Morelli, M. J., Bianchi, S., Rotta, L., Capra, T., Sabò, A., Campaner, S., & Amati, B. (2015). Genome-wide analysis of p53 transcriptional programs in B cells upon exposure to genotoxic stress in vivo. *Oncotarget*, *6*(28), 24611.
- Vainer, R., Cohen, S., Shahar, A., Zarivach, R., & Arbely, E. (2016). Structural basis for p53 Lys120-acetylation-dependent DNA-binding mode. *Journal of molecular biology*, *428*(15), 3013-3025.
- Van Orden, K., Giebler, H. A., Lemasson, I., Gonzales, M., & Nyborg, J. K. (1999). Binding of p53 to the KIX Domain of CREB Binding Protein A potential link to human t-cell leukemia virus, type i-associated leukemogenesis. *Journal of Biological Chemistry*, *274*(37), 26321-26328.

- Vaquerizas, J. M., Kummerfeld, S. K., Teichmann, S. A., & Luscombe, N. M. (2009). A census of human transcription factors: function, expression and evolution. *Nature Reviews Genetics*, *10*(4), 252-263.
- Verfaillie, A., Svetlichnyy, D., Imrichova, H., Davie, K., Fiers, M., Atak, Z. K., Hulselmans, G., Christiaens, V. & Aerts, S. (2016). Multiplex enhancer-reporter assays uncover unsophisticated TP53 enhancer logic. *Genome research*, *26*(7), 882-895.
- Vise, P. D., Baral, B., Latos, A. J., & Daughdrill, G. W. (2005). NMR chemical shift and relaxation measurements provide evidence for the coupled folding and binding of the p53 transactivation domain. *Nucleic acids research*, *33*(7), 2061-2077.
- Wadgaonkar, R., & Collins, T. (1999). Murine double minute (MDM2) blocks p53-coactivator interaction, a new mechanism for inhibition of p53-dependent gene expression. *Journal of Biological Chemistry*, *274*(20), 13760-13767.
- Wang, Y. H., Tsay, Y. G., Tan, B. C. M., Lo, W. Y., & Lee, S. C. (2003). Identification and characterization of a novel p300-mediated p53 acetylation site, lysine 305. *Journal of Biological Chemistry*, *278*(28), 25568-25576.
- Waterman, M. J., Stavridi, E. S., Waterman, J. L., & Halazonetis, T. D. (1998). ATM-dependent activation of p53 involves dephosphorylation and association with 14-3-3 proteins. *Nature genetics*, *19*(2), 175-178.
- Wei, C. L., Wu, Q., Vega, V. B., Chiu, K. P., Ng, P., Zhang, T., Shahab, A., Yong, H.C., Fu, Y., Weng, Z., & Liu, J. (2006). A global map of p53 transcription-factor binding sites in the human genome. *Cell*, *124*(1), 207-219.

- Weinberg, R. L., Veprintsev, D. B., & Fersht, A. R. (2004). Cooperative binding of tetrameric p53 to DNA. *Journal of molecular biology*, 341(5), 1145-1159.
- Weiss, M. A., Ellenberger, T., Wobbe, C. R., Lee, J. P., Harrison, S. C., & Struhl, K. (1990). Folding transition in the DNA-binding domain of GCN4 on specific binding to DNA. *Nature*, 347(6293), 575-578.
- Wintjens, R., & Rooman, M. (1996). Structural classification of HTH DNA-binding domains and protein–DNA interaction modes. *Journal of molecular biology*, 262(2), 294-313.
- Wulf, G. M., Liou, Y. C., Ryo, A., Lee, S. W., & Lu, K. P. (2002). Role of Pin1 in the regulation of p53 stability and p21 transactivation, and cell cycle checkpoints in response to DNA damage. *Journal of Biological Chemistry*, 277(50), 47976-47979.
- Xu, Y. (2003). Regulation of p53 responses by post-translational modifications.
- Zacharias, M. (2006). Minor groove deformability of DNA: a molecular dynamics free energy simulation study. *Biophysical journal*, 91(3), 882-891.
- Zhan, Y. A., Wu, H., Powell, A. T., Daughdrill, G. W., & Ytreberg, F. M. (2013). Impact of the K24N mutation on the transactivation domain of p53 and its binding to murine double-minute clone 2. *Proteins: Structure, Function, and Bioinformatics*, 81(10), 1738-1747.
- Zheng, H., You, H., Zhou, X. Z., Murray, S. A., Uchida, T., Wulf, G., Gu, L., Tang, X., Lu, K.P., & Xiao, Z. X. J. (2002). The prolyl isomerase Pin1 is a regulator of p53 in genotoxic response. *Nature*, 419(6909), 849-853.

Ch 6- Materials, Methods, and Approaches in Analysis of Missense Mutations within the Transactivation Domain of p53

6.1. Expression, mutagenesis and purification of p53-TAD (performed in collaboration with Indhujah Thevarajan)

A pET28a vector with sub-cloned cDNA of *Hsp53*-(1-73) (72R) was procured from Addgene (Addgene plasmid #62082; <http://n2t.net/addgene:62082>; RRID:Addgene_62082); deposition courtesy of Dr. Gary Daughdrill, University of South Florida. The p53-(1-73) plasmid was then transformed into DH10 β cells (New England Biolabs) to generate a plasmid stock in preparation of p53-(1-73) variants. p53/pET28a plasmid was then templated to create the desired mutations in p53-(1-73) through primer mismatch using Agilent Technologies QuickChange XL reactions: K24N (F: 5'-TCAGAVCTATGGAATCTACTTCCT-3'; R: 5'-AGGAAGTAGATTCCATAGGTCTGA), N29K/N30D (F: 5'-GGAACTACTTCCTGAAAAAGACGTTCTGTCCCCCTTGCC-3'; R: 5'-GGCAAGGGGGACAGAACGTCTTTTTTCAGGAAGTAGTTTCC-3'), D49Y (F: 5'-CTGTCCCCGGACTATATTGAA-3'; R: 5'-TTCAATATCGTCCGGGGACAG-3'), and W53G (F: 5'-CGGACGATATTGAACAAGGCTTCACTGAAGACCCAGGTCC-3'; R: 5'-GGACCTGGGTCTTCAGTGAAGCCTTGTTCAATATCGTCCG-3').

p53-(1-73) variants were transformed into BL21(DE3) low background strain (LOBSTR) chemically competent cells. The transformed cells were selected for using lysogeny broth (LB) agar plates with 50ug/mL Kanamycin and 1%(w/v) glucose. Expression of p53-(1-73) for Biolayer Interferometry (BLI) was accomplished using LB media. p53-(1-73) variants expressed using LB media were grown at 30°C in 4L cultures of LB with Kanamycin and 1% glucose. 1mM Isopropyl-

β -D-thiogalactoside (IPTG) was added when cultures reached an $OD_{600} = 0.5-0.7$. After IPTG induction, the cultures were grown at 37°C for an additional 6hrs. Purification of p53-(1-73) from the clarified cell lysate was accomplished using a Nickel Nitrilotriacetic Acid (Ni-NTA) column using 50mM TRIS-HCl with 300mM NaCl at pH=8.0. Further, TRIS-HCl buffers containing 10mM Imidazole for wash or 250mM Imidazole for elution were used. Eluent fractions of p53-(1-73) variants were then concentrated using Amicon Ultra-15 (3.5KDa MWCO) centrifugal filters. Concentrates were subjected to two rounds of G75 Superdex Gel-filtration column chromatography at 4°C using 50mM NaH_2PO_4 buffer with 300mM NaCl and 1mM EDTA at pH=7.0. Final concentrates of p53-(1-73) variants were then dialyzed overnight at 4°C in a Slide-A-Lyzer dialysis cassette (3.5KDa MWCO) into a 20mM Tris-HCl buffer with 50mM NaCl and 10% glycerol at pH=7.5 prior to storage at -20°C .

^{15}N labeling of p53-(1-73) variants for NMR studies was obtained by growing cells in 6L cultures of M9 media (2.5mg/L Thiamine, 2.5mg/L Choline Chloride, 2.5mg/L Ca-d-Pantothenate, 2.5mg/L Nicotinamide Adenine Dinucleotide, 1.25mg/L Pyridoxal Hydrochloride, 50mg/L EDTA, 8.3mg/L $\text{FeCl}_3 \cdot 6\text{H}_2\text{O}$, 840 $\mu\text{g/L}$ ZnCl_2 , 130 $\mu\text{g/L}$ $\text{CuCl}_2 \cdot 2\text{H}_2\text{O}$, 100 $\mu\text{g/L}$ $\text{CoCl}_2 \cdot 6\text{H}_2\text{O}$, 100 $\mu\text{g/L}$ H_3BO_3 , and 16 $\mu\text{g/L}$ $\text{MnCl}_2 \cdot 6\text{H}_2\text{O}$; 6g/L Na_2HPO_4 , 3g/L KH_2PO_4 , 500mg/L NaCl, 4g/L Glucose, 120mg/L MgSO_4 , 33mg/L CaCl_2 and 500mg/L $^{15}\text{NH}_4\text{Cl}$) at 30°C . 1mM Isopropyl- β -D-thiogalactoside (IPTG) was added when cultures reached an $OD_{600} = 0.5-0.7$. The cultures were incubated for an additional 6hrs at 37°C . Initial Ni-NTA purification of ^{15}N labeled p53-(1-73) variants was carried out identically to conditions outlined in LB expressions. Further, the eluents of labelled p53-(1-73) variants were concentrated and dialyzed into 50mM Tris-HCl buffer overnight. Cleavage of N-terminal 7His-tag was accomplished using Sigma-Aldrich Clean Cleave thrombin-agarose kit via recombinant thrombin cleavage site for 6hr at 23°C . Cleaved p53-(1-73)

variants were then concentrated and purified using the Gel-filtration chromatography. p53-(1-73) eluents were passed over Ni-NTA column. Fractions containing the cleaved p53-(1-73) variants were collected. Finally, purified ¹⁵N labeled p53-(1-73) variants were concentrated to a final volume of less than 1mL prior to dialysis into the NMR buffer (50mM NaH₂PO₄, 50mM NaCl, 1mM DTT, 1mM EDTA at pH=6.5). The NMR samples were prepared to a concentration of 300uM for each p53-(1-73) variant in 675μL NMR buffer mixed with 35μL volume of 99% D₂O.

6.2. Expression and purification of TAZ2 domain of CBP

A pET22b recombinant plasmid coding a bi-cistronic mRNA containing 6His-Gb1-STAT1 (710-750) and CBP (1764-1850) TAZ2 was provided as a gift from Dr. Peter Wright (Addgene plasmid # 99342; <http://n2t.net/addgene:99342>; RRID:Addgene_99342). The pET22b/STAT1/TAZ2 encodes a second ribosomal binding site at the start of the TAZ2 sequence. This enables concurrent expression of TAZ2 and binding partner STAT1. pET22b/STAT1/TAZ2 was transformed into BL21(DE3) Rosetta chemically competent cells. Transformed cells were selected for expression on LB agar plates with 1% glucose in the presence of 25ug/mL Chloramphenicol and 100ug/mL of Ampicillin.

STAT1/TAZ2 expression was accomplished using 2L volume of LB with 1% glucose, 25ug/mL Chloramphenicol and 100ug/mL Ampicillin via shaker at 37°C. Once the culture OD₆₀₀=0.1-0.3, approximately 250uM of ZnCl₂ was introduced to facilitate stability of expressed TAZ2. Induction of culture expression with 1mM IPTG was accomplished when OD₆₀₀=0.5-0.7. Cultures were then allowed to express for an additional 3hrs at 37°C. Following pelleting and cell lysis STAT1/TAZ2 was resuspended in 20mM Tris-HCl buffer with 300mM NaCl and 10mM Imidazole at pH=8.0. The clarified cell lysate was then passed over Ni-NTA column and elution

fractions containing STAT1/TAZ2 were collected. The STAT1/TAZ2 fractions were applied to a sulfopropyl (SP) sepharose cation exchange column with a flow rate of 0.3mL/min. TAZ2 was independently eluted using a linear salt gradient from 50mM to 1M NaCl in a 20mM Tris-HCl buffer with 10mM DTT at pH=6.5. Purified TAZ2 fractions were quickly concentrated and dialyzed into a 25mM Na-Acetate buffer with 25mM NaCl, 0.1mM DTT, and 0.1mM ZnCl₂ at pH=5.8. Appropriate folding of TAZ2 was confirmed through Circular Dichroism (CD) spectroscopy prior to storage at 4°C.

6.3. Expression and purification of TAZ1 domain of CBP

pET21a plasmid containing CBP-(340-439) TAZ1 was given as a gift by Dr. Peter Wright's lab. The transformation, selection, and expression of pET21a/TAZ1 was carried as outlined for STAT1/TAZ2. TAZ1 containing lysate was suspended in 20mM Tris-HCl buffer with 50mM NaCl and 10mM DTT at pH=6.9. Further isolation of TAZ1 was carried using SP-sepharose linear gradient procedure as outline for TAZ2. TAZ1 containing eluent was then concentrated and injected onto a G75 superdex column using a 20mM Tris-HCl buffer with 300mM NaCl and 10mM DTT at pH=6.8. Fractions containing TAZ1 were concentrated and dialyzed into 20mM Tris-HCl buffer with 50mM NaCl and 1mM DTT at pH=6.8. TAZ1 folding, as with TAZ2, was assessed using CD spectroscopy prior to storage at 4°C.

6.4. Expression and purification of HDM2 (performed by Indhujah Thevarajan)

A pGEX-6p-2 vector with subcloned cDNA of MDM2 (17-125) was obtained from Dr. Gary Daughdrill's lab (Addgene # 62063; <http://n2t.net/addgene:62063>; RRID:Addgene_62063). The pGEX-6p-2-MDM2(17-125) plasmid was transformed into DH10- β chemically competent cells (New England Biolabs) for plasmid stock preparation. The plasmid was then transformed into Rosetta BL21(DE3) chemically competent cells. Transformed cells were selected for expression on LB agar plates in the presence of 25ug/mL Chloramphenicol and 100ug/mL of Ampicillin. MDM2 was expressed using LB media and grown at 37°C with Chloramphenicol and Ampicillin to an OD₆₀₀ = 0.8. 1mM Isopropyl- β -D-thiogalactoside (IPTG) was added to the culture and cells were grown at 25°C for an additional 5hrs. The culture was centrifuged at 5000 rpm for 30 mins and pellets were resuspended in GST binding buffer (50 mM Tris-HCl, 300 mM NaCl, 2.5 mM EDTA, 0.02% NaN₃, 2mM DTT, pH=7.4). The cells were lysed using French press and centrifuged at 13,000 rpm for 40 mins. HDM2 containing supernatant was loaded onto a Glutathione Sepharose 4B resin (GE Healthcare) column that had been already equilibrated with GST binding buffer. The MDM2 protein fractions were eluted with GST binding buffer with 10 mM reduced glutathione. The GST tag was removed using HRV 3C protease (Thermo Scientific) with an enzyme to substrate ratio of 1:100 at 4°C for 16 hrs. After the removal of the tag the MDM2 protein was concentrated using Amicon Ultra-15 (3.5 KDa MWCO) centrifugal filters. Finally, the protein was buffer exchanged into 20 mM Tris-HCl buffer with 150 mM NaCl, 1mM DTT, and 10% glycerol at pH 8.0 prior to storage at -20°C.

6.5. Heteronuclear NMR experiments of p53TAD variants

All 2D heteronuclear ¹⁵N-¹H NMR experiments were performed on a 11.74 Tesla Varian 500MHz VNMRs system (Agilent Technologies Inc., Palo Alto, CA) with an operational ¹H

frequency of 499.84MHz. All NMR experiments were accomplished at the NMR Core Facility in the Department of Biochemistry and Molecular Biophysics at Kansas State University, utilizing a 5mm cryogenic triple resonance inverse detection pulse field gradient probe operating at a temperature of $25 \pm 0.1^\circ\text{C}$. The NMR sample of p53-(1-73) variants were prepared in NMR buffer containing 5% D_2O . The effective concentration of p53-(1-73) variants was $300\mu\text{M}$ for heteronuclear experiments (Bodenhausen, 1980). 2D ^{15}N -Heteronuclear Single Quantum Coherence (^{15}N -HSQC) experiments of p53-(1-73) variants were performed with a spectral width of 6009Hz and 256 complex points in t_2 dimension, a sweep width of 1944Hz and 256 complex points in t_1 dimension. All HSQC cross-peak resonance assignments for WT p53-(1-73) were assigned based upon reported chemical shift values from Vise et al, 2005. Assignments for K24N, N29K/N30D, D49Y, and W53G p53-(1-73) variants were achieved in the same manner as WT with missing cross-peaks being tentatively assigned to the most proximal orphan peaks. All NMR experiments were conducted with a pre-saturation pulse scheme for suppression of the solvent peak (HOD) at 4.78ppm. The residual water peak was then used as an internal reference for chemical shift assignments (Piotto, 1992). The processing of NMR data was accomplished using VNMRJ3.2a software (Agilent Technologies Inc., Palo Alto, CA). Spectral analysis was then carried out using graphical assignment utilities CCPNMR and Sparky-NMRFAM (Goodard, 2004; Vranken, 2005).

6.6. Relaxation data acquisition and analysis of p53TAD variants

Relaxation experiments for uniformly ^{15}N -labeled p53-(1-73) variants with a concentration of 0.5mM in NMR buffer were performed at 25°C . Spin-lattice relaxation (R_1), spin-spin relaxation (R_2) and ^1H - ^{15}N NOE experiments were carried out by inverse-detected 2D NMR methodologies

(Kay et al., 1989). Spin-lattice relaxation rates were measured by collecting ten 2D NMR spectra with relaxation delays of 10, 50, 110, 190, 310, 500, 650, 1000, 1500, and 1900ms. Spin-spin relaxation rates were also measured by collecting ten 2D NMR spectra with relaxation delays of 10, 30, 50, 90, 110, 150, 190, 210, 230 and 250ms. Peak heights of each series of relaxation experiments were then fitted to a single decaying exponential function to determine their respective relaxation times. Measurement of ^1H - ^{15}N NOEs was accomplished with 2 separate spectral acquisitions, one with a 3s mixing time for NOE buildup (NHNOE) and a second with a 3s recycle delay as reference (NONOE). In general, fitting errors were within 10% of calculated relaxation rates.

Analysis of relaxation rates was integrated using a reduced spectral density mapping approach (Farrow et al, 1995; Lefevre et al., 1996; Peng et al., 1992; Peng et al., 1995; Vise et al., 2005). The influences upon ^{15}N nuclei relaxation are largely contingent upon the ^{15}N chemical shift anisotropy (c) and the dipolar coupling interaction (d) between amide ^{15}N and bound ^1H (Abragam, 1961). The values for R_1 , R_2 , and NOE between amide proton and nitrogen nuclei can then be related in terms of spectral density functions:

$$R_1 = (d^2/4)[3J(\omega_N) + 6J(\omega_H + \omega_N) + J(\omega_H - \omega_N)] + c^2J(\omega_N)$$

$$R_2 = (d^2/8)[4J(0) + 3J(\omega_N) + 3J(\omega_H + \omega_N) + 6J(\omega_H) + J(\omega_H - \omega_N)] + (c^2/6)[J(0) + 6J(\omega_N)] + R_{ex}$$

$$\text{NOE} = 1 + (d^2/4R_1)(\gamma_H/\gamma_N)[6J(\omega_H + \omega_N) - J(\omega_H - \omega_N)]$$

$J(\omega)$ is the power spectral density function that defines the reorientation of the H-N bond in both stochastic (global) and intramolecular motions as a function of frequency. The reduced spectral density mapping approach assumes that linear combinations of $J(\omega_H + \omega_N)$, $J(\omega_H - \omega_N)$, and $J(\omega_H)$ are treated equally as an average value of $J(\omega_H)$ (or $J(0.87\omega_H)$). Using this

approximation, the aforementioned relationships can be rewritten in terms of $J(0)$, $J(\omega_H)$, and $J(\omega_N)$:

$$\sigma_{\text{NH}} = R_1(\text{NOE} - 1)(\gamma_{\text{N}}/\gamma_{\text{H}})$$

$$J(\omega_H) = 4\sigma_{\text{NH}}/5d^2$$

$$J(\omega_N) = (4R_1 - 5\sigma_{\text{NH}})/(3d^2 + 4c^2)$$

$$J(0) = (6R_2 - 3R_1 - 2.72\sigma_{\text{NH}})/(3d^2 + 4c^2)$$

Where σ_{NH} is the spectral density, γ_{N} and γ_{H} are the gyromagnetic ratios for ^{15}N and ^1H nuclei, $d = (\mu_0 h \gamma_{\text{H}} \gamma_{\text{N}} / 8\pi^2)(r_{\text{NH}}^{-3})$ is the dipole-dipole relaxation coefficient, μ_0 is the magnetic permeability of a vacuum, h is Plank's constant, r_{NH} is bond-length between amide nitrogen pairs ($\sim 1.02\text{\AA}$), and $c = \omega_{\text{N}} \Delta\sigma / \sqrt{3}$ is the chemical shift relaxation coefficient, and $\Delta\sigma$ is the chemical shift anisotropy for ^{15}N which is (-160ppm). Using this approach, a single field acquisition of R_1 , R_2 , and HETNOE rates enables effective approximation $J(0)$, $J(\omega_H)$, and $J(\omega_N)$ values.

6.7. Biolayer interferometry assays of p53TAD variants to HDM2, CBP-TAZ1, and CBP-TAZ2 (performed in collaboration with Indhujah Thevarajan)

Aliquots of TAZ2, TAZ1, and HDM2 from storage buffer were diluted into freshly generated 20mM Tris-HCl buffer with 50mM NaCl and 1mM DTT at pH=7.5 (6.8 for TAZ1) to appropriate target concentrations for analysis of binding kinetics. The buffer for binding kinetics studies was prepared with a composition of 20mM Tris-HCl, 50mM NaCl, 0.01% (W/V) BSA, and 0.002% (V/V) Tween-20. An aliquot of p53 and its variants were each prepared via dilution into binding kinetics buffer solution to generate a final concentration of 100ng/ μL of p53-(1-73) variants. All samples were stored on ice at 0C° until used. 200 μL aliquots of the control and p53-

(1-73) variant samples were transferred into 500 μ L black Eppendorf tubes. These samples were equilibrated to room temperature 10 minutes prior to their use. Ni-NTA conjugated BLItz tips were hydrated for ~10 minutes using binding kinetics buffer prior to data acquisition. Each experiment was conducted using BLItz Pro 1.2 software and the instrument operated at an oscillation frequency of 2200Hz for the duration of each acquisition. The scheme for data acquisition was conducted as follows: Step 1 (baseline 1), the tip is equilibrated an additional 30s with kinetics buffer and instrument self-calibrates: Step 2 (analyte binding), p53-(1-73) variant or control sample was applied to the tip for 240s: Step 3 (baseline 2), tip is exposed to empty kinetics buffer for 120s: Step 4 (association), TAZ2 dilution at 300nM in 20mM Tris, 50mM NaCl, and 1mM DTT at pH=7.5 is applied for 300s(120s for TAZ1 acquisitions): And step 5 (dissociation), tip is exposed to kinetics buffer for 300s (120s for TAZ1 acquisitions). After analyzing the data and figure was generated using GraphPad Prism 8.1.0 software to approximate effective on-rate (k_{on}), off-rate(k_{off}), and K_D values.

6.8. References

- Bodenhausen, G., & Ruben, D. J. (1980). Natural abundance nitrogen-15 NMR by enhanced heteronuclear spectroscopy. *Chemical Physics Letters*, 69(1), 185-189.
- Abragam, A., & Abragam, A. (1961). *The principles of nuclear magnetism* (No. 32). Oxford university press.
- Brooks, C. L., & Gu, W. (2003). Ubiquitination, phosphorylation and acetylation: the molecular basis for p53 regulation. *Current opinion in cell biology*, 15(2), 164-171.
- Chao, C., Wu, Z., Mazur, S. J., Borges, H., Rossi, M., Lin, T., Wang, J.Y., Anderson, C.W., Appella, E., & Xu, Y. (2006). Acetylation of mouse p53 at lysine 317 negatively regulates p53 apoptotic activities after DNA damage. *Molecular and cellular biology*, 26(18), 6859-6869.
- Chen, L., & Chen, J. (2003). MDM2-ARF complex regulates p53 sumoylation. *Oncogene*, 22(34), 5348-5357.
- Chen, J., Marechal, V., & Levine, A. J. (1993). Mapping of the p53 and mdm-2 interaction domains. *Molecular and cellular biology*, 13(7), 4107-4114.
- Cho, Y., Gorina, S., Jeffrey, P. D., & Pavletich, N. P. (1994). Crystal structure of a p53 tumor suppressor-DNA complex: understanding tumorigenic mutations. *Science*, 265(5170), 346-355.
- Farrow, N. A., Zhang, O., Szabo, A., Torchia, D. A., & Kay, L. E. (1995). Spectral density function mapping using 15 N relaxation data exclusively. *Journal of biomolecular NMR*, 6(2), 153-162.

- Kay, L. E., Torchia, D. A., & Bax, A. (1989). Backbone dynamics of proteins as studied by nitrogen-15 inverse detected heteronuclear NMR spectroscopy: application to staphylococcal nuclease. *Biochemistry*, 28(23), 8972-8979.
- Lefèvre, J. F., Dayie, K. T., Peng, J. W., & Wagner, G. (1996). Internal mobility in the partially folded DNA binding and dimerization domains of GAL4: NMR analysis of the N–H spectral density functions. *Biochemistry*, 35(8), 2674-2686.
- Peng, J. W., & Wagner, G. (1992). Mapping of spectral density functions using heteronuclear NMR relaxation measurements. *Journal of Magnetic Resonance (1969)*, 98(2), 308-332.
- Peng, J. W., & Wagner, G. (1995). Frequency spectrum of NH bonds in eglin c from spectral density mapping at multiple fields. *Biochemistry*, 34(51), 16733-16752.
- Piotto, M., Saudek, V., & Sklenář, V. (1992). Gradient-tailored excitation for single-quantum NMR spectroscopy of aqueous solutions. *Journal of biomolecular NMR*, 2(6), 661-665.
- Vise, P. D., Baral, B., Latos, A. J., & Daughdrill, G. W. (2005). NMR chemical shift and relaxation measurements provide evidence for the coupled folding and binding of the p53 transactivation domain. *Nucleic acids research*, 33(7), 2061-2077.

Chapter 7- Biophysical and Functional Analysis of Cancer-linked Mutations within the Intrinsically Disordered Transactivation

Domain of p53¹

Abstract

p53 is one of the most studied human genes largely due to its critical roles in regulating cellular genotoxic stress response and DNA damage directed repair (DDR). However, little is known about the impact of cancer-linked mutations upon the intrinsically disordered N-terminal transactivation domain (TAD) of p53. In this work, we compared the structure and interactions of four common p53TAD variants, -K24N, -N29K/N30D, -D49Y, and -W53G with -WT through a combination of NMR and BLI techniques. NMR structural studies reveal only minor effects of the mutations on the structural ensemble of p53TAD, while RSDA reveals that mutants -N29K/N30D, -D49Y, and -W53G have subtle changes in internal motion. Subsequent interaction kinetics analysis from BLI with regulatory partners HDM2, CBP-TAZ1, and CBP-TAZ2 domains reveal p53TAD-WT shows k_{on} values approaching the diffusion limit with its tightest binding partners (HDM2 and CBP-TAZ2). These results indicate that changes in internal motion of the binding surface correlate with a loss of binding with these partners. In summary, the p53TAD conformational ensemble either samples natively or rapidly adopts a set of binding-compliant conformers in solution. Either local or distal mutations within this IDP can disproportionately compromise high-affinity interactions indirectly by subtly altering this conformational ‘landscape.’

¹This manuscript is under preparation for submission to a journal.

Abbreviations: CREB binding Protein (CBP); Human Double-minute-2 (HDM2); Intrinsically Disordered Proteins (IDPs); Post-translational modification (PTMs); Nuclear Magnetic Resonance (NMR); Reduced Spectral Density Approach (RSDA); Biolayer Interferometry (BLI)

7.1. Introduction

Since p53's discovery as a tumor suppressor it has been identified as one of the most frequently mutated genes in cancers (Baker, 1989; Hollstein, 1991; Kandath, 2013; Lawrence, 2014). This is due to p53's key role in regulating the DNA damage response (DDR), and it has become the most studied gene of the human genome (Dolgin, 2017). p53 is a 393 amino acid long protein, and it has been mapped to chromosome 17 in the p13.1 arm (Saha, 2015). This protein consists of 4 distinct domains: The intrinsically disordered (IDP) transactivation domain (TAD; M1-P60); the proline-rich domain (PRD; D61-S94); the DNA-binding domain (DBD; T102-K292); and the C-terminal domain or N-terminally regulated domain (NRD; P301-D393) (Cho, 1994). Upon stabilization of the p53 protein, it is capable of binding to DNA in monomeric, dimeric, and tetrameric oligomerization states which in turn activates unique gene sets that regulate downstream processes that dictate cell-fate (Allen, 2014; Fischer 2016). Genes induced by p53 activation span a multitude of biological processes relevant to a cell's response to genotoxic stresses. These processes include DNA damage and repair (DDR) proteins that include DNA damage-binding protein-2 (DDB2), cell-cycle arrest by cyclin-dependent kinase inhibitor 1A (CDKN1A), apoptosis by BCL2-binding component 3 (PUMA) and BCL2 associated X protein (BAX), metabolism through TP53-induced glycolysis and apoptosis regulator (TIGAR), and post-translational regulators of p53 by Human Double-minute-2 (HDM2) and p53-induced phosphatase 1 (Hafner, 2019).

Ubiquitously expressed in somatic cells, p53 is subject to regulation by post-translational modifications (PTMs) that include phosphorylation at S6/9/15/20/33/37/46/149/215/269/315/376/378/392 and at T18/55/81/150/155 (Cordenonsi, 2007; Tibbetts, 1999; Kalid and Ben-Tal, 2009; Sakaguchi, 1997; Sakaguchi, 1998; Oda, 2000; Bech-Otschir, 2001; Fraser, 2010; Craig, 1999; Schon, 2002; Katayama, 2004; Gatti, 2000; Waterman, 1998), acetylation at K120/164/305/317/320/372/373/381/382 (Sykes, 2009; Tang, 2008; Chao, 2006; Gu, 1997), sumoylation at K386 (Chen and Chen, 2003), ubiquitination at K24/370/372/373/381/382/386 (Peng, 2016; Lee, 2019), and methylation at K370/372/373/382 (Shi, 2007; Huang, 2006; Huang, 2010; Chuikov, 2004). Moreover, it has been estimated that p53 has greater than 300 unique PTM combinations in total (Bode, 2004; Brooks, 2003; Dehart, 2014). While most PTMs of p53 are site-specific and demonstrate a capacity to ‘fine-tune’ target gene selection and stability, there is only one bulk PTM regulatory process that suppresses or activates p53.

In vivo, the p53 transcription factor co-binds E3 ubiquitin ligase human double-minute-2 (HDM2) along with histone acetyltransferases (HATs) CREB binding protein (CBP) or, CBP paralog, p300. Under non-genotoxic stress conditions HDM2 binds the p53TAD (T18-L26) forming the first helix (AD1), while CBP/p300 domains bind to p53TAD (D42-T55) forming a second helix (AD2) (Dawson, 2003; Krois, 2016). When HDM2 is bound to the p53TAD it negatively regulates p53 stability through poly-ubiquitination of the NRD at residues K370/373/381/382/386 and p53 activity through partial occlusion of the TAD (Oliner, 1993; Chen, 1993). Upon integration of genotoxic stress signaling through kinases, residues S15/20 and T18 are phosphorylated and interrupts HDM2-p53TAD complex formation (Craig, 1999; Schon, 2002; Mavinahalli, 2010; Dumaz, 1999; Hafner, 2019; Tibbetts, 1999; Kalid and Ben-Tal, 2009; Matsumoto, 2006). This enables CBP/p300 to fully complex with the p53TAD (T18-L26, D42-

T55) promoting protective acetylation of p53 ubiquitination sites (Krois, 2016). It also facilitates CBP/p300 co-activator functionality through formation of the RSF1/p53/CBP complex resulting in histone remodeling at p53-dependent target genes to promote target mRNA expression (Min, 2018).

Although a great deal is known concerning the effects of mutants within p53's DBD (compromise of oligomerization, DNA binding, overall stability) and its NRD (loss of PTM sites), however, there is a deficit in research concerning the effects of identified cancer-linked mutations within the TAD itself. The p53TAD, being an IDP, does not maintain a well-defined three-dimensional structure in isolation. It often adopts an ensemble of interconverting conformations. During binding, IDPs like p53TAD can adopt unique conformations from their ensembles that facilitate binding to each of their unique binding partners, thus enabling a large degree of binding promiscuity (Wright, 2015; Tompa, 2002). Therefore, it is important to define the conformational landscape of the free protein and its dynamics. Several recent studies conducted upon IDPs that fold upon binding investigated the dynamics of the isolated peptide to gain insights into these features (Salvi, 2016; Abyzov, 2016). Outside of *in silico* studies, the effects of cancer-linked mutations upon unbound p53TAD structure and relative dynamics is poorly understood. In addition, the functional impacts of p53TAD mutants with p53's strongest regulators, HDM2 and CBP/p300, are also poorly understood.

In this study, we selected four cancer-linked mutants within the p53TAD. These disorder promoting mutants included K24N, N29K/N30D, D49Y, and W53G. Using recombinantly

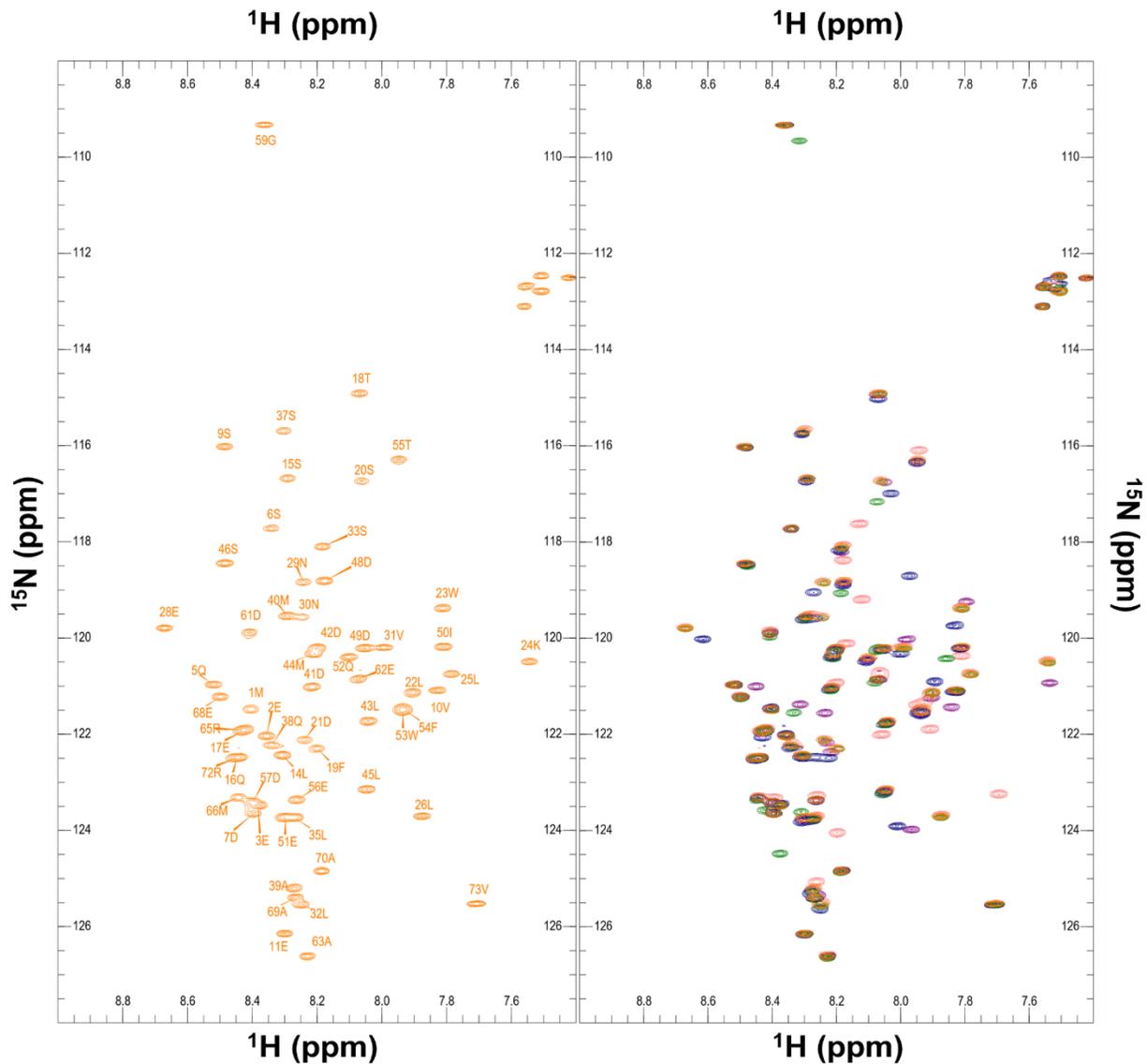


Figure 7.1. ^{15}N -HSQC of p53TAD. (Left) The complete assigned ^{15}N -HSQC spectrum of p53TAD-WT (orange) as derived from Vise et al., 2005. (Right) Overlay of p53TAD ^{15}N -HSQC spectra with K24N(blue), N29K/N30D (purple), D49Y (Pink), and W53G (Green).

expressed human p53TAD-(1-73) constructs in conjunction with solution NMR techniques, we quantitatively evaluated changes in structure at the ns-ps timescales in order to probe variation in conformer interconversion (Jensen, 2009). We used reduced spectral density approaches (RSDA) along with NMR to show that p53TAD-N29K/N30D created a small yet significant change in structural dynamics within the AD1 and AD2 helices, while -D49Y and -W53G mutants only affected the AD2 helix (Farrow, 1995). Then, interactions between the highest affinity regulatory

Chemical Shift Perturbation of p53TAD Variants

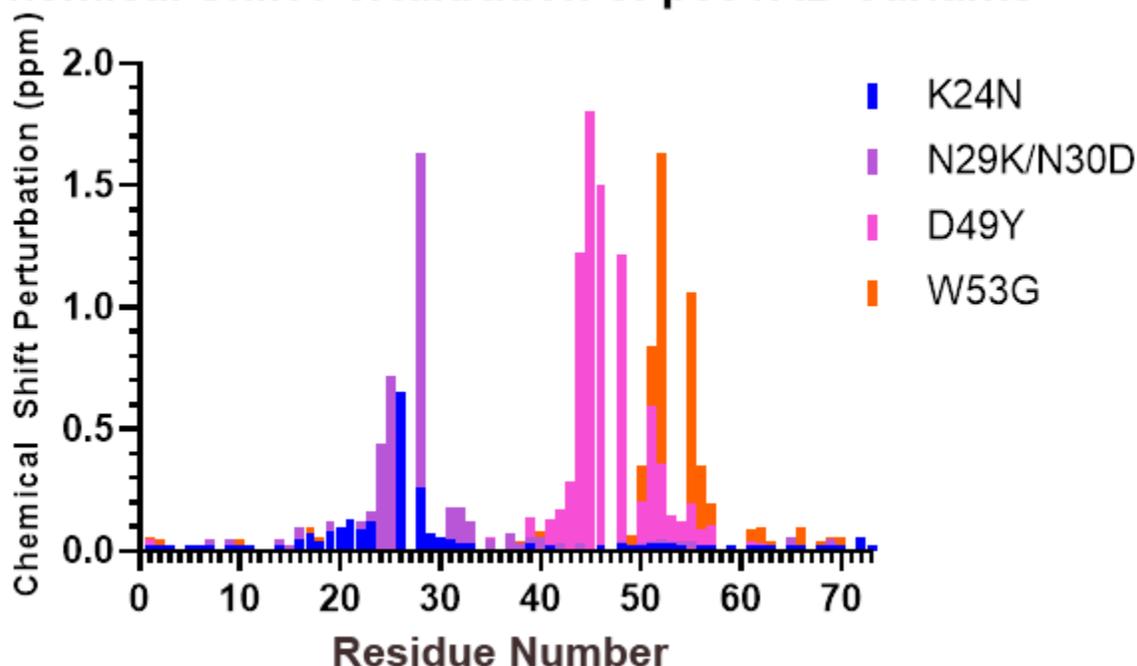


Figure 7.2. Chemical shift perturbation per residue of p53TAD variants relative to WT. The chemical shift perturbation of each assigned amino acid residue of p53TAD variants was calculated by measuring the total change (ppm) in ^{15}N and ^1H dimensions from the original assignment position in WT using the equation $\sqrt{(25(\Delta^1\text{H}_{\text{ppm}})^2 + (\Delta^{15}\text{N}_{\text{ppm}})^2)}$.

domains of *HsHDM2*(17-125), *MmCBP*-transcription adapter putative zinc finger (TAZ)-2 (340-439) and *MmCBP*-TAZ1 (1764-1850) were characterized through biolayer interferometric (BLI) techniques. BLI measurements of the association (k_{on}) and dissociation (k_{off}) kinetics of immobilized p53TAD variants with HDM2, CBP-TAZ1, and CBP-TAZ2 show an agreement with predicted fast exchange regime kinetics derived from previously reported ^{15}N -HSQC titration studies (Ferreon, 2009). The approximated $K_{\text{D}}(=k_{\text{off}}/k_{\text{on}})$ from kinetics also agrees with reported K_{D} measurements obtained from both isothermal calorimetric and ^{15}N -HSQC titrations (Ferreon, 2009). From these results, we conclude that changes in ensemble motion of p53TAD when overlapping with a partner relevant binding surface lowers binding efficiency. In addition, these

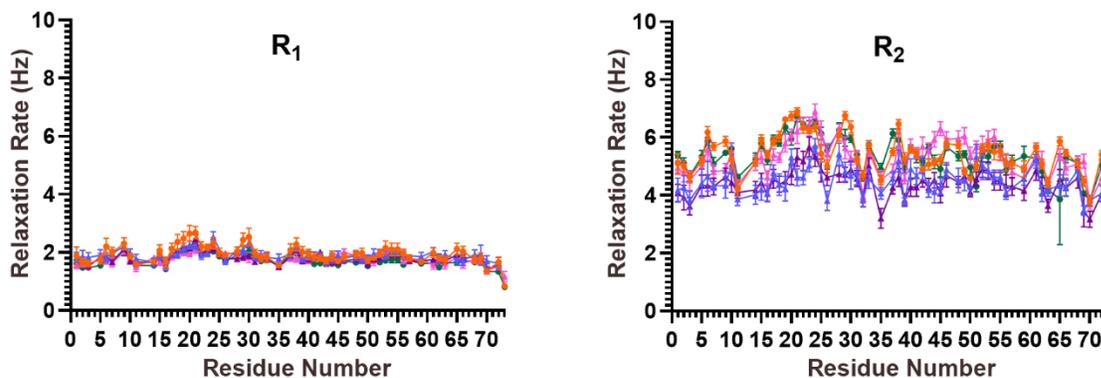


Figure 7.3. Spin-lattice (R_1) and spin-spin (R_2) relaxation rates of p53TAD variants. WT (orange), K24N (blue), N29K/N30D (purple), D49Y (pink), and W53G (green). Bars represent 95% confidence interval based upon exponential decay fitting.

changes in binding surface motions disproportionately impact binding partners with larger k_{on} values.

7.2. Results

7.2.1. NMR resonance assignments of p53TAD variants

As established, the p53TAD has been described as a disordered protein in its unbound state. IDPs typically fail to display sufficient resonance dispersion in multi-dimensional heteronuclear NMR spectral approaches as the chemical environments for participating nuclei are often quite uniform. Interestingly, the NMR resonances for residues 1-73 of the p53 TAD were well dispersed and all 60 non-proline amide ^1H and ^{15}N resonance pairs were successfully assigned. Therefore, observational differences in structural features between p53TAD-WT and its variants, -K24N, -N29K/N30D, -D49Y, and -W53G, can be readily observed through acquisition of ^{15}N -HSQC data. Following procedures reported in Vise et al., 2005, p53TAD-WT was diluted in NMR buffer (50mM PBS with 1mM DTT at pH=6.5) to a concentration of 300uM. All 60 non-proline residues

Table 7.1. Statistics for Relaxation and Reduced Spectral Density Mapping for p53TAD Variants

P53TAD Variant	R1 (Hz)	R2 (Hz)	NOE	J(ω_H) (s *rad⁻¹ * 10⁻¹¹)	J(ω_N) (s *rad⁻¹ * 10⁻¹⁰)	J(0) (s *rad⁻¹ * 10⁻⁹)
WT	1.95±0.31	5.41±0.74	-0.47±0.27	4.93±0.91	4.56±0.80	1.85±0.25
K24N	1.91±0.19	4.46±0.56	-0.24±0.21	4.11±0.54	4.60±0.56	1.46±0.21
N29K/N30D	1.81±0.22	4.38±0.53	-0.21±0.23	3.76±0.61	4.39±0.60	1.45±0.19
D49Y	1.80±0.19	5.32±0.70	-0.46±0.34	4.54±0.92	4.22±0.58	1.85±0.27
W53G	1.72±0.22	5.34±0.70	-0.60±0.27	4.74±0.68	3.93±0.62	1.87±0.26

were assigned using ¹⁵N-HSQC correlation spectroscopy to an accuracy of less than 0.01ppm of previously reported values (Vise, 2005), as show in figure 7.1.

The chemical shifts for p53TAD-K24N, -N29K/N30D, -D49Y, and -W53G were assigned using p53TAD-WT chemical shift values as a reference. ¹⁵N-HSQC cross-peaks from p53TAD variant spectra that directly overlapped with -WT spectral positions were directly assigned. The NMR cross-peaks that didn't directly overlap within each mutant variant spectrum were assigned to the most proximal p53TAD-WT peak assignment. An overlay of all ¹⁵N-HSQC spectra of p53TAD variants makes it evident that changes in chemical environment due to mutations were constrained to less than 15% of observable peaks for any given p53TAD variant, as shown in figure 7.1. These results suggested that the global 'structure' or ensemble of free p53TAD did not significantly change with these mutations.

For quantitative analysis, the chemical shift perturbation of each assigned amino acid residue of p53TAD variants was calculated by measuring the total change (ppm) in ¹⁵N and ¹H dimensions from the original assignment position in WT using the equation [$\sqrt{(25(\Delta^1H_{ppm})^2 + (\Delta^{15}N_{ppm})^2)}$], as shown in figure 7.2. This quantitative analysis suggested that all significant perturbations in chemical shift for p53TAD variants were proximal to the site of each mutation which was expected as p53TAD is intrinsically dynamic. Interestingly, p53TAD-D49Y and

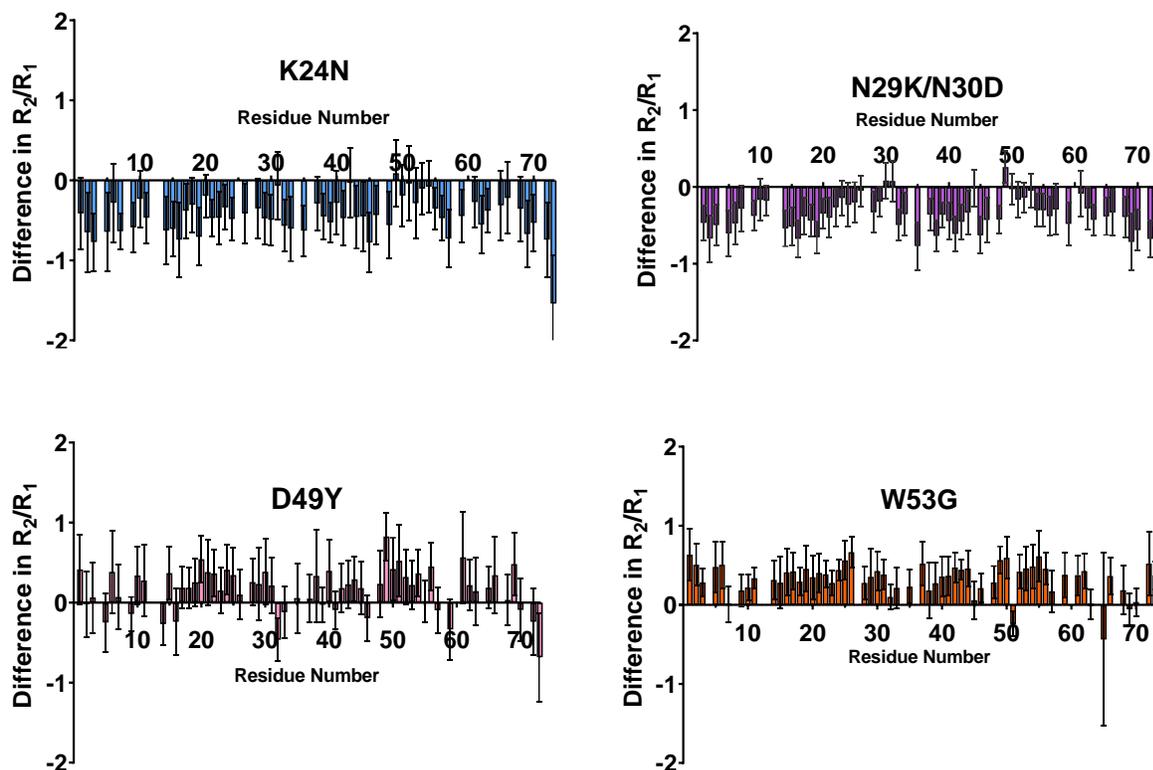


Figure 7.4. Difference in R_2/R_1 Between p53TAD Variants and -WT. Error bars reflect 95% confidence interval derived from R_1 and R_2 values. Negative R_2/R_1 correlates with shorter τ_c and suggests greater compactness. Positive R_2/R_1 correlates with longer τ_c and lowered compactness.

N29K/N30D manifested an asymmetric perturbation profile in both intensity and number of residues impacted suggesting a small local change in preferred conformations.

7.2.2. Relaxation rate measurements and reduced spectral density approaches for p53TAD variants

Because p53TAD variants displayed a relatively minor impact upon the native structure of p53TAD, a series of relaxation experiments were employed to assess the nuanced variances in conformational dynamics. Measurements of the spin-lattice relaxation rate (R_1), the

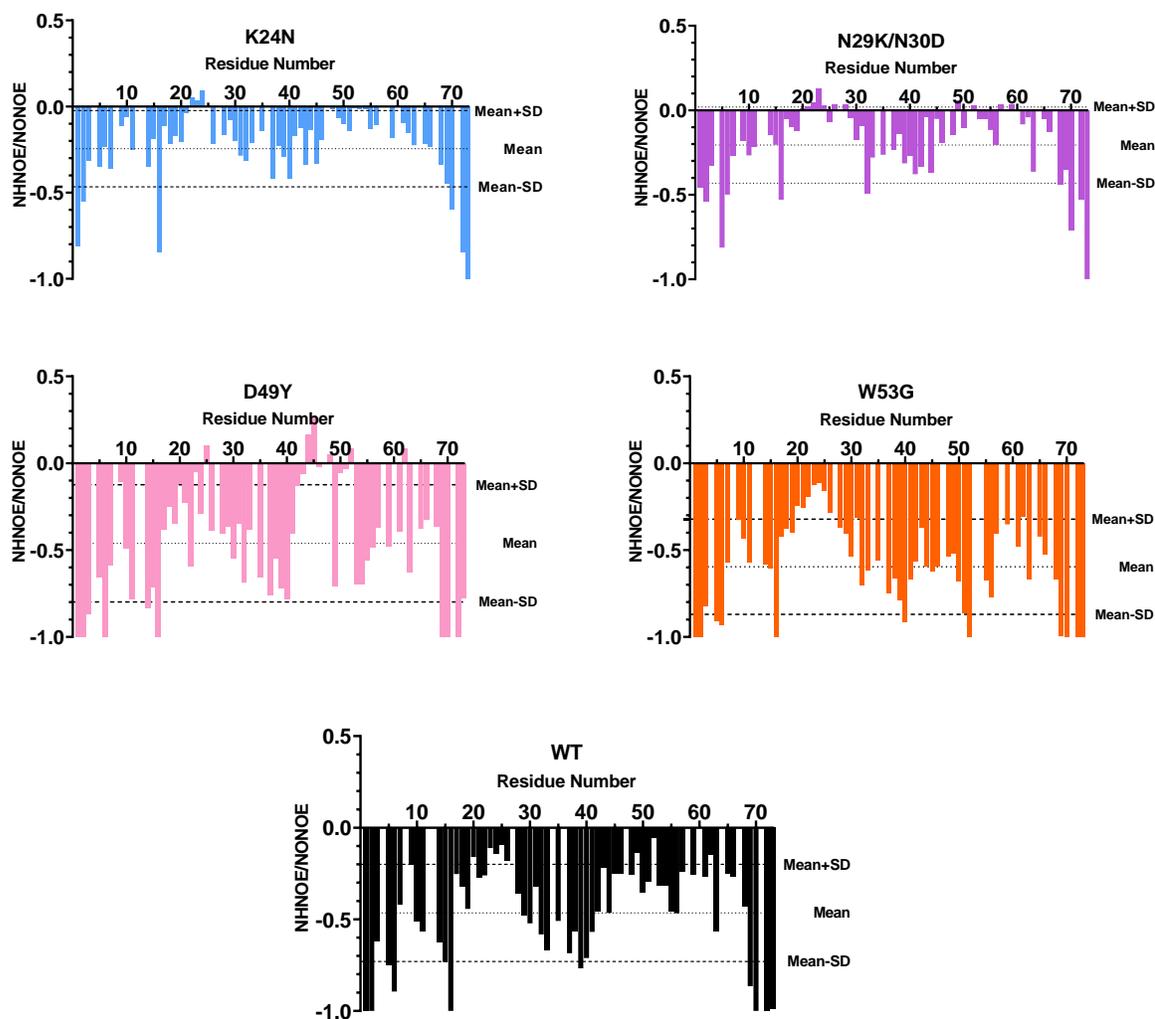


Figure 7.5. NHNOE/NONOE Ratio for Each p53TAD Variant. Mean reflects the average per residue NHNOE/NONOE value for that specific p53TAD variant and SD is the observed standard deviation per residue.

longitudinal relaxation rate, for p53TAD variants were carried out at a concentration of 500uM in NMR buffer with a total of 10 unique delays ranging from 10-1900ms for each variant. The measured peak height intensities of each assigned residue, with exception of p53TAD-K24N's L25 and -W53G's F54, were then fit to an exponential decay function to approximate the R_1 . The measurements of per residue longitudinal relaxation for each of the variants show only slight differences with p53TAD-WT, as shown in figure 7.3. The table 7.1. shows the highest measured R_1 value (1.95 ± 0.31 Hz) for p53TAD-WT while p53TAD variants display lower overall values.

Measurements of the spin-spin relaxation rate (R_2), the transverse relaxation rate, were acquired with similar conditions and number of delays ranging from 10-250ms. The calculated R_2 values for each p53TAD variant were also determined via peak height measurement and fitted to an exponential decay function, as shown in figure 7.3. Analysis of R_2 among p53TAD variants shows considerable variation as -K24N and -N29K/N30D were depreciated globally with respect to -WT, as shown in table 7.1. The relative ratio of R_2/R_1 , which is proportional to the rotational correlation time (τ_c), of each p53TAD variant was calculated and compared against WT values, as shown in figure 7.4. Analysis of the global R_2/R_1 ratio suggested that p53TAD-K24N and -N29K/N30D have markedly lower τ_c than -WT, while for -D49Y and -W53G τ_c were demonstrably higher. To better analyze the impacts of molecular motions, the ^1H - ^{15}N NHNOE ratio (NHNOE/NONOE) per residue for each p53TAD variant was acquired. Resultant NHNOE/NONOE ratios were then calculated based upon measured peak height for each of the assigned residues as presented in figure 7.5. The table 7.1. shows that NHNOE/NONOE values for p53TAD-K24N (-0.24 ± 0.21) and -N29K/N30D (-0.21 ± 0.23) are less negative in comparison to -D49Y (-0.46 ± 0.34), -W53G (-0.60 ± 0.27), and -WT (-0.47 ± 0.27). Typically, increases in NHNOE/NONOE values correlate with restriction in motion on the ps timescales and suggests that p53TAD-K24N and -N29K/N30D may undergo less internal motion than -WT. In contrast, p53TAD-W53G displays a more negative NHNOE/NONOE global ratio and showed a more pronounced negative ratio for residues 45 to 54

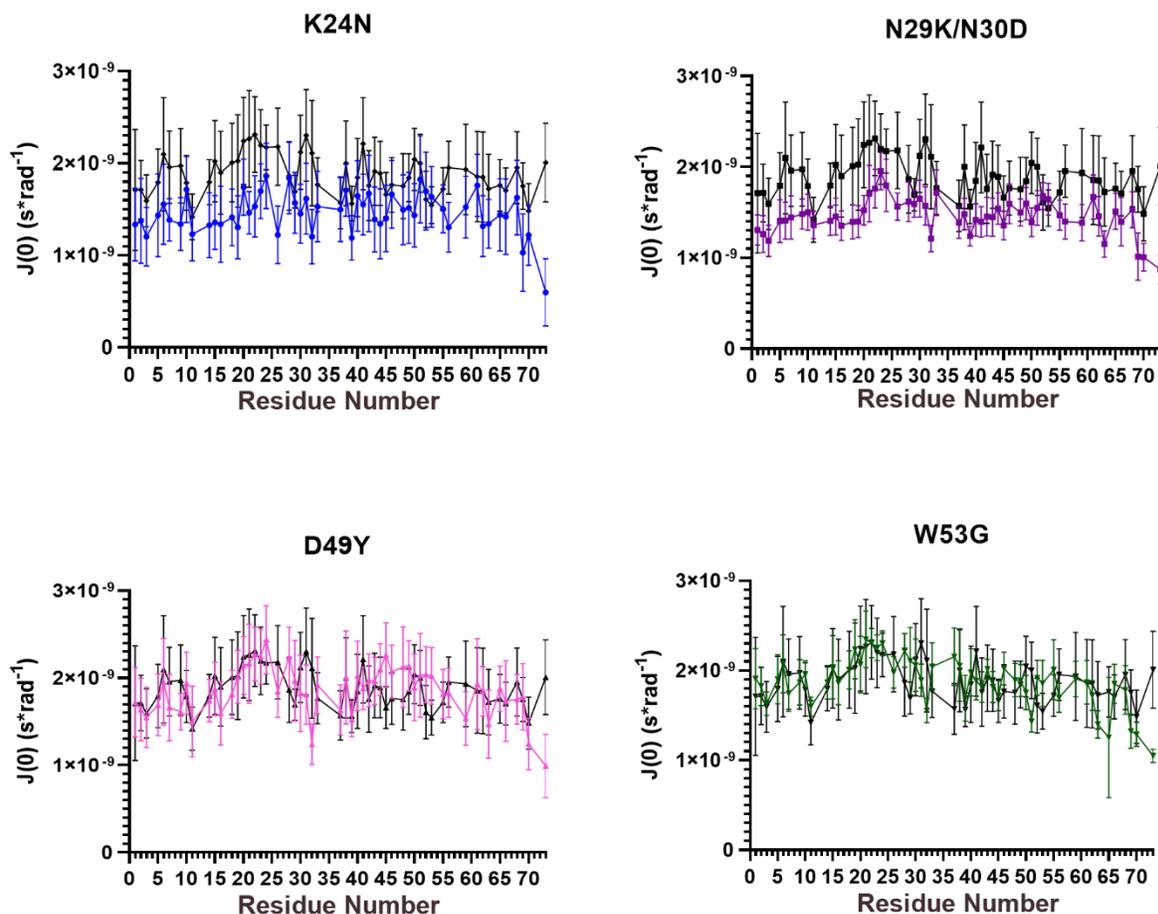


Figure 7.6. Reduced Spectral Density Approximation of $J(0)$ for p53TAD Variants. An approximation of relaxation power at a frequency of 0Hz. Color represents the corresponding variant while black represents WT values. Error bars represent the 95% confidence interval derived from R_1 and R_2 exponential decay fitting.

than -WT which inferred greater local internal motion. Moreover, p53TAD-D49Y maintained a very similar global NHNOE/NONOE ratio to -WT but displayed a significant increase in the ratio for residues 42 to 50, which may suggest that these residues experienced less internal motion.

To evaluate additional impacts upon protein dynamics of p53TAD through the introduction of mutation, a reduced spectral density approach (RSDA) was utilized. RSDA is an effective

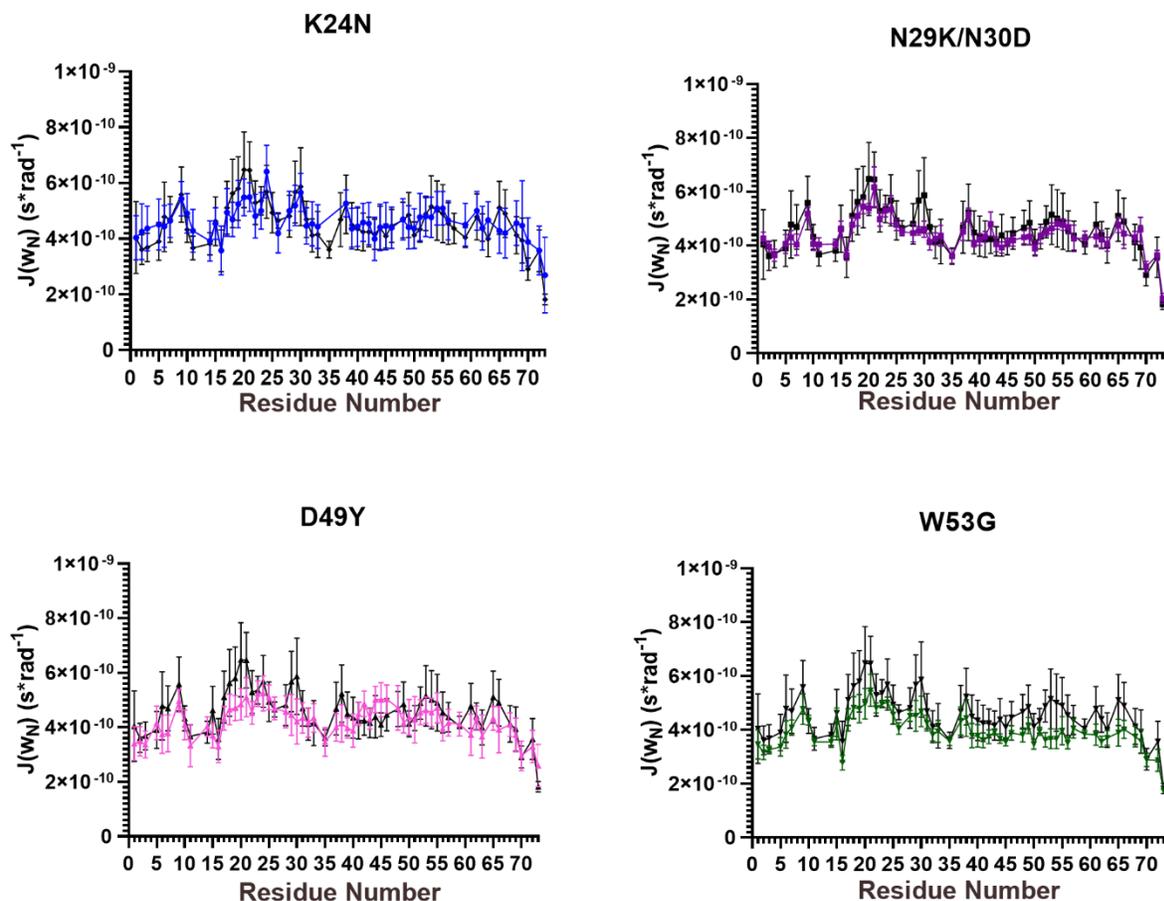


Figure 7.7. Reduced Spectral Density Approximation of $J(\omega_N)$ for p53TAD Variants. An approximation of relaxation power at a frequency of 50.7MHz, Larmor frequency of ^{15}N in 11.74T field. Color represents the corresponding mutant variant while black represents WT values. Error bars represent the 95% confidence interval derived from R_1 and R_2 exponential decay fitting.

means of measuring variances in the spectral density function ($J(\omega)$) along the ^1H - ^{15}N bond vector at three distinct frequency regimes: 0.0Hz, ω_N (50.7MHz), and $0.87\omega_H$ (435MHz). By varying the frequencies assayed, this approach enables sampling of motions from ns-ps timescales. Although this approach is broadly applied to well-folded proteins, however, it does not rely on any assumptions concerning molecule motion (Farrow, 1995). This makes it especially effective at describing the dynamic behaviors observed for IDPs. Approximation of $J(\omega_H)$, $J(\omega_N)$, and $J(0)$ were carried out through integration of the observed R_1 , R_2 , and NHNOE/NONOE ratio values for each of the p53TAD variants in conjunction with equations as described in chapter 6. Excluding

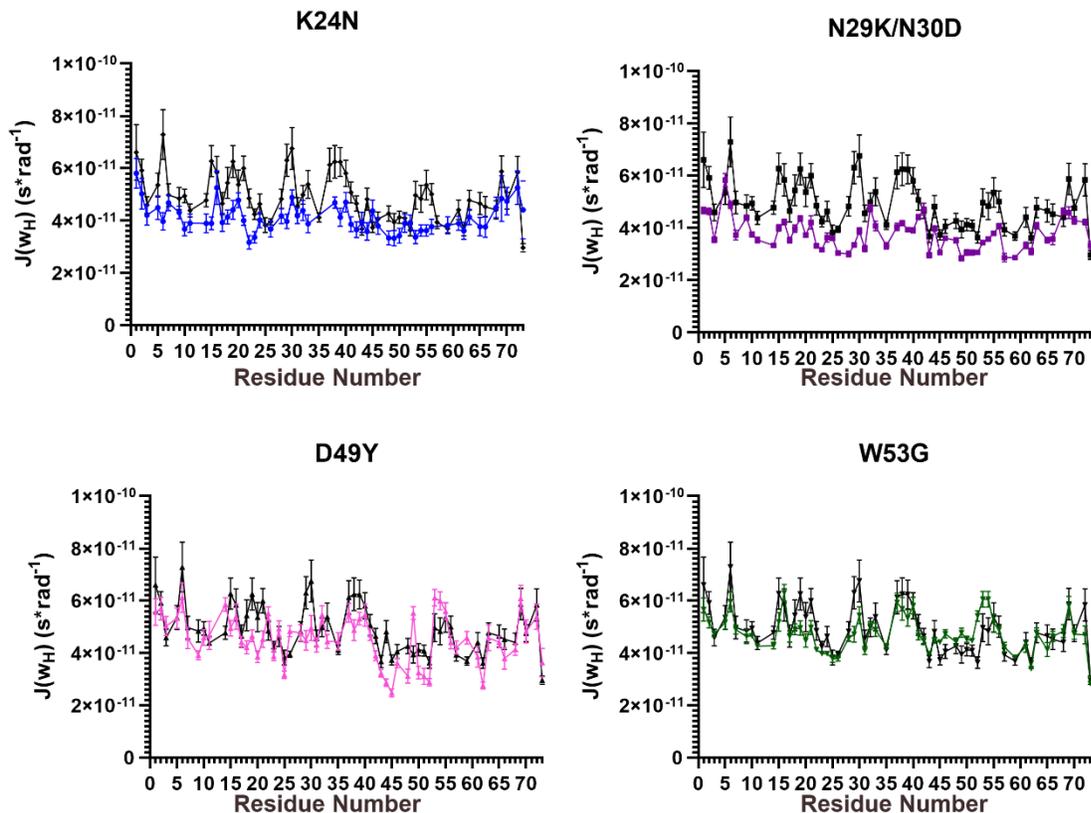


Figure 7.8. Reduced Spectral Density Approximation of $J(\omega_H)$ for p53TAD Variants. An approximation of relaxation power at a frequency of 435MHz, Larmor frequency of ^1H in 11.74T field. Color represents the corresponding mutant variant while black represents WT values. Error bars represent the 95% confidence interval derived from R_1 and R_2 exponential decay fitting.

residues L35 and D57 of p53TAD-K24N for poor curve fitting in R_1 and R_2 , approximation of $J(\omega_H)$, $J(\omega_N)$, and $J(0)$ for residues of all other p53TAD variants are summarized in figures 7.6-7.8.

$J(0)$ values inversely correlate with internal flexibility and directly correlate with tumbling of the ^1H - ^{15}N bond. Small values of $J(0)$ indicate greater internal flexibility of ^1H - ^{15}N bond. p53TAD-D49Y and -W53G displayed markedly similar $J(0)$ values ($\sim 1.85\text{s} \cdot \text{rad}^{-1} \cdot 10^{-9}$) per residue to -WT which is a strong indication of similar dynamics, as shown in table 7.1. Conversely,

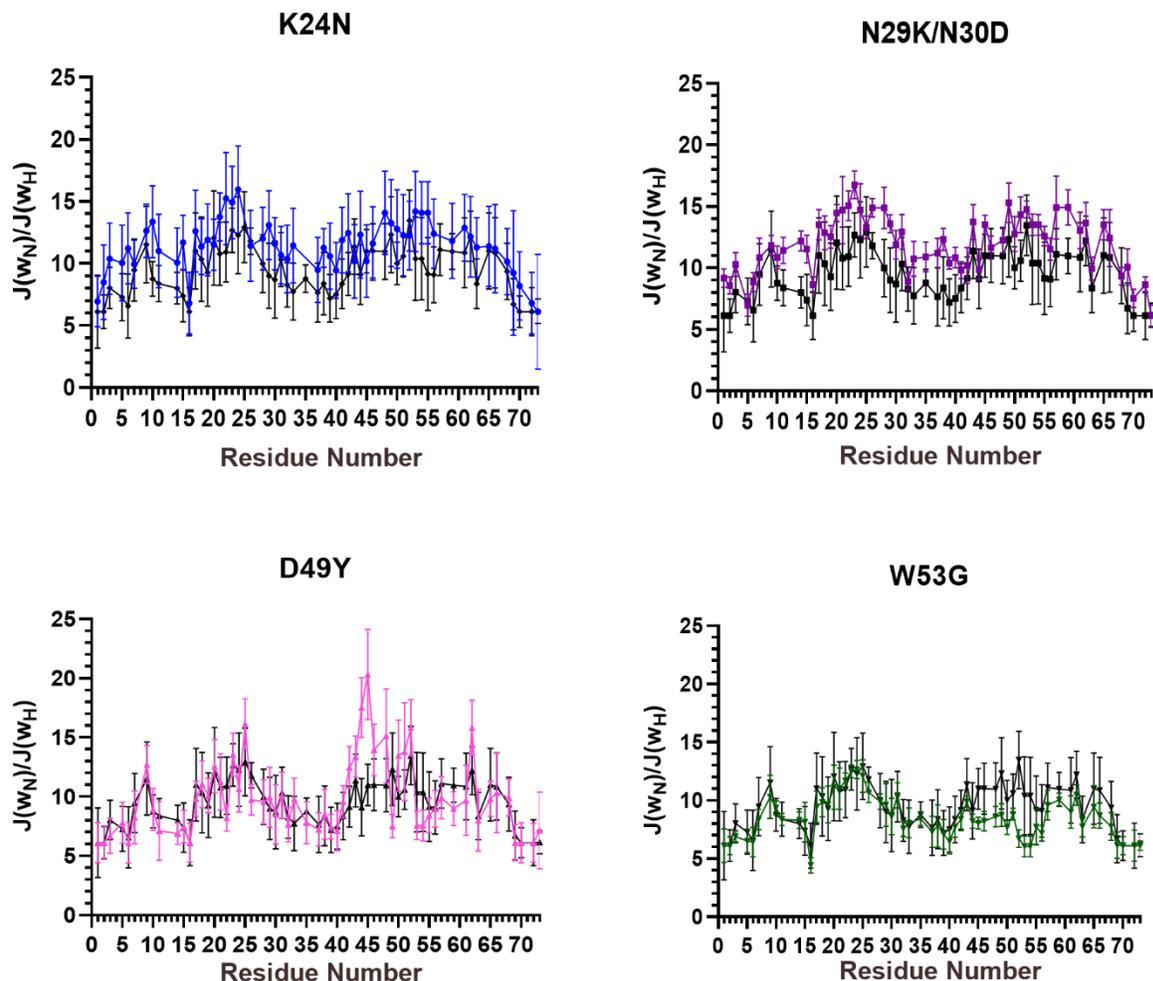


Figure 7.9. $J(\omega_N)/J(\omega_H)$ Spectral Density Ratio of p53TAD Variants. Larger ratio values correspond to shorter autocorrelation times of amide ^1H bond vectors with respect to amino acid residue side-chain autocorrelations. Error bars represent the 95% confidence interval calculated from R_1 and R_2 curve fitting.

both p53TAD-K24N and -N29K/N30D have generally lower $J(0)$ per residue with a more pronounced decrease observed for residues 14 to 21, as shown in figure 7.6. This marked decrease would be consistent with a diminished structure. Analysis of $J(\omega_N)$ displayed a low variance among p53TAD mutants with respect to -WT residue positions. These results suggested very little variance in sidechain to sidechain motion which is common among IDPs. Finally, higher frequency $J(\omega_H)$ values indicated faster timescale motions (ps) in a proteins' ^1H - ^{15}N bond vectors. Larger $J(\omega_H)$ values correlate to higher mobility while low values correspond to structured segments.

$J(\omega_H)$ analysis shows an interesting effect as p53TAD-K24N and N29K/N30D both lack variation and are lower than observed values for -WT, as shown in figure 7.8. These studies suggested a restriction in the degrees of freedom available for ^1H - ^{15}N bond vector motion while simultaneously inferring diminished native structure. Conversely, p53TAD-D49Y and -W53G present the opposite effect as -WT values for $J(\omega_H)$ are still partially retained. To better approximate the relative steric restriction per residue of the p53TAD variants, the $J(\omega_N)/J(\omega_H)$ ratio was calculated, as shown in figure 7.9. This compensates for motions that influence relaxation similarly in ps-ns timescales. The similarity in global ratios for most residues in p53TAD-D49Y and -W53G suggested near identical per residue flexibility to -WT while larger global ratios of -K24N and -N29K/N30D suggested less flexibility. Examining further, p53TAD variants -N29K/N30D, -D49Y, and -W53G show subtle but distinct local effects in local motion. p53TAD-N29K/N30D displays a greater $J(\omega_N)/J(\omega_H)$ ratio for residues 10-40 and 55-59, suggesting a slight increase in steric restriction mostly concerted around the AD1 helix. Similarly, P53TAD-D49Y $J(\omega_N)/J(\omega_H)$ ratio suggests an increase in steric restriction for residues 42-50. These results were consistent with chemical shift perturbation as both variants showed the most pronounced changes in local chemical shifts. In contrast, p53TAD-W53G showed a reduction in $J(\omega_N)/J(\omega_H)$ ratio for residues 45 to 54 suggesting that this region had become less sterically restricted.

7.2.3. Impact upon binding kinetics between p53TAD variants and CBP-TAZ2, CBP-TAZ1, and HDM2

Assessment of the impact of p53TAD variants upon binding interactions with regulatory partners was accomplished using a BLI method. This approach assesses the thermodynamic and

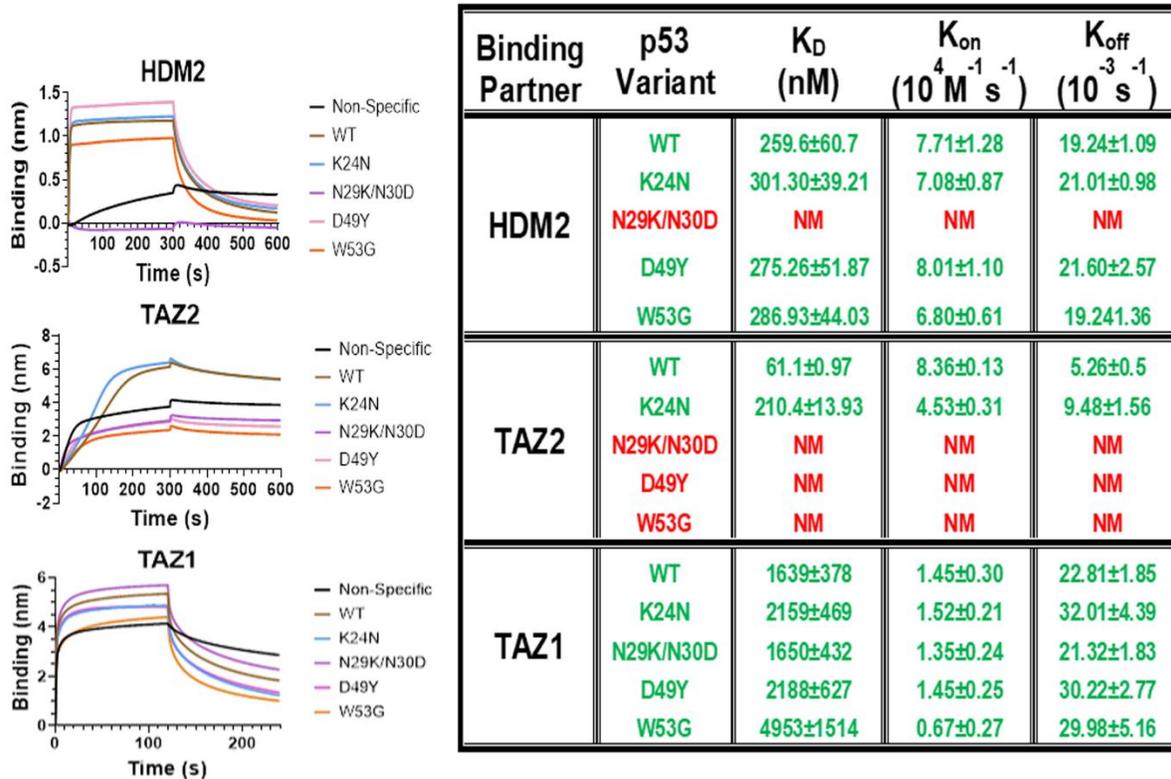


Figure 7.10. Binding of p53TAD Variants to Binding Partners CBP-TAZ2, CBP-TAZ1, and N-HDM2. (Left) 100ng/uL of P53TAD variants were applied to Ni-NTA tip prior to the introduction of 300nM TAZ2, 3uM HDM2, or 30uM TAZ1. Binding response was calibrated to 0 at the start of association phase. All binding curves represent average response of two complete association phase, first 120s or 300s, and dissociation phase, the second 120s or 300s. The Non-Specific had no p53 variants applied to the tip prior to binding partner introduction and is used as a control. (Right) Summary of binding kinetics between p53TAD variants and binding partners. Each interaction was measured with two replicates at 3 unique concentrations using BLI with an Ni-NTA tip. Concentrations used were 2uM, 3uM, and 4uM for HDM2, 300nM, 600nM, and 900nM for TAZ2, and 20uM, 30uM, and 40uM for TAZ1. K_{on} and K_{of} rates were then used to determine the effective k_D . NM indicates that the value could not be accurately measured at this concentration range. (This work was done in collaboration with Indhujah Thevarajan)

kinetic impacts of each p53TAD variant (-WT, -K24N, -N29K/N30D, -D49Y, and -W53) interaction with each binding partner (HDM2, CBP-TAZ1, and CBP-TAZ2) by measuring small changes in the effective length of the probe tip (nm) using interferometry. Using reported K_D values for p53TAD-WT with binding partners, a series of binding kinetics assays were conducted with two replicates at concentrations of 300nM, 600nM, and 900nM for CBP-TAZ2, 2uM, 3uM, and

4 μ M for HDM2, and 20 μ M, 30 μ M, and 40 μ M CBP-TAZ1 to all p53 variants. All acquisitions were baseline subtracted using the dilution buffer as a control, see figure 7.10. Using Blitz Pro1.2 software, the association (k_{on}) and dissociation (k_{off}) rates were then calculated based upon non-linear regression fitting by least squares with the mean and standard deviation for all 6 measurements reported. Results suggest that p53TAD-WT has k_{on} values of $7.71 \cdot 10^4 \text{M}^{-1} \text{s}^{-1}$, $1.45 \cdot 10^4 \text{M}^{-1} \text{s}^{-1}$, and $8.36 \cdot 10^4 \text{M}^{-1} \text{s}^{-1}$ with k_{off} values of $19.24 \cdot 10^{-3} \text{s}^{-1}$, $22.81 \cdot 10^{-3} \text{s}^{-1}$, and $5.26 \cdot 10^{-3} \text{s}^{-1}$ for interactions with HDM2, CBP-TAZ1, and CBP-TAZ2 respectively. The calculated k_{on} values are consistent with quickly equilibrating interactions on the NMR timescales as previously predicted (Ferreon, 2009). The effective K_D s for each interaction were then approximated using $k_{off}/k_{on}=K_D$, and also agreed with previously reported values (Ferreon, 2009).

All p53TAD variants that displayed measurable binding with HDM2, CBP-TAZ1, and CBP-TAZ2 retained k_{on} and k_{off} values within 2-4 fold of -WT values. P53TAD variants with no detectable interaction (NM, not measured, as shown in figure 7.10.) are present for CBP-TAZ2:p53TAD (variants -N29K/N30D, D49Y, and -W53G) and HDM2:p53TAD (variant -N29K/N30D) binding and absent in CBP-TAZ1:p53TAD binding measurements. These data suggest a likely diminishment in k_{on} as both CBP-TAZ2:p53TAD and HDM2:p53TAD complexes have faster k_{on} interactions approaching $10^5 \text{M}^{-1} \text{s}^{-1}$ while CBP-TAZ2:p53TAD complex is slightly slower at the $\sim 10^4 \text{M}^{-1} \text{s}^{-1}$. These data have been summarized in figure 7.10.

7.3. Discussion

p53 has long been the focus of clinical research as its functional loss is frequently identified in cancer patients. While extensive research has been conducted showing mutations within the DBD (T102-K292) and the NRD (P301-D393) can compromise either structure and/or sites of post-translation modification, little focus has been applied to evaluating mutational outcomes within the TAD (M1-P60) as it is an intrinsically dynamic region (IDR or IDP). Recent studies have attempted to address changes in dynamics observed in IDR/IDPs by varying field strength or presence/absence of binding partners. However, it is not known how the dynamics of IDPs change from introduction of point mutations at select residue positions as the complex conformational landscape renders most predictions concerning changes in structure, dynamics, and biophysical activities moot.

In this work, we have attempted to evaluate changes in p53TAD by using clinically derived cancer mutants with NMR spectroscopic measurements at single field strength (11.74T). The observation of all chemical shift perturbations for the p53TAD variants with respect to -WT infers that the structural impact is minor. Our results suggest that the chemical environment experienced for most of the amine ^{15}N - ^1H pairs within the backbone are similar for all p53TAD variants outside the sites of mutation, which indicates no detectable change in ensemble conformation. However, analysis of the R_2/R_1 ratios, which is a surrogate assay for assessing rotational correlation time τ_c , suggest that p53TAD-K24N and -N29K/N30D may be either less structured or more compact than -WT, while -D49Y and -W53G may be more structured or less compact; as observed transverse relaxation rate ($R_{2,\text{obs}}$) is biased by contributions of the chemical exchange rate (R_{ex}), an unbiased $R_{2,\text{actual}}$ cannot be deconvoluted using this experimental approach alone. Analysis using the ^{15}N heteronuclear NOE ratio (NHNOE/NONOE), which enables sensitive measurements of motions

less than 0.5ns timescales, was used to assay structural rigidity. NHNOE/NONOE ratios reveal that p53TAD-K24N and -N29K/N30D display less relaxation and by extension are more rigid than -WT, while the converse is true of -D49Y and -W53G. To integrate the data obtained from all three experimental assays, the RSDA was used to calculate $J(\omega)$ for all p53TAD variants at 0MHz, 50.7MHz (Larmor frequency of ^{15}N , ω_N), and 435MHz ($0.87 * \text{Larmor frequency of } ^1\text{H}$, ω_H) using equations as described in chapter 6. Global values of $J(0)$ per residue largely reflect observations from R_2/R_1 analysis as both calculations are contingent upon $R_{2,\text{obs}}$ values. Calculations of $J(\omega_N)$ showed very little change in values among p53TAD-WT and its mutants. However, changes in $J(\omega_H)$ values between p53TAD variants proved informative. $J(\omega_H)$ values for p53TAD-K24N and -N29K/N30D were markedly lower than -D49Y, -W53G, and -WT. This would suggest a loss of the native ensembles structure while indicating a restriction in ^1H - ^{15}N bond vector. To confirm this observation, the degree of steric hindrance was approximated via the ratio of $J(\omega_N)/J(\omega_H)$ calculations. The $J(\omega_N)/J(\omega_H)$ ratio analysis suggests p53TAD-K24N and -N29K/N30D are more sterically confined overall. The residues V10 to M40 and I50 to G59 of p53TAD-N29K/N30D are specifically less mobile which compromises the motions of residues from both the AD1 and AD2 helices. Conversely, $J(\omega_N)/J(\omega_H)$ ratio analysis also suggests P53TAD-D49Y and -W53G are more localized to the AD2 helix in their effects upon ensemble motion. p53TAD-W53G promotes greater flexibility for residues 45 to 54 and -D49Y displays pronounced rigidity for residues 42 to 50. Intriguingly, $J(\omega_N)/J(\omega_H)$ ratio analysis agrees quite well with outcomes from BLI kinetics assays as p53TAD-K24N displays little effect with HDM2 (binding surface T18-L26) and CBP-TAZ1/2 (binding surface D42-T55) domains. P53TAD-K24N's middling impact in both binding kinetics and dynamics studies is rather peculiar as these assays alone would suggest that the mutation is only coincidental in identified cancers. However, this is unlikely as recent studies have

identified residue 24 to be another site of poly-ubiquitination for degradation by mitochondrial ubiquitin ligase 1 (Peng, 2016; Lee, 2019). p53TAD-D49Y and -W53G both compromise interactions solely with CBP-TAZ domains and disproportionately effect the CBP-TAZ2 domain. p53TAD-N29K/N30D's outcome is also unique in that it compromises interactions with both the repressor and activator (CBP-TAZ2 and HDM2) but this remains consistent with the observed changes in motion along binding surfaces belonging to both the activator and the repressor. Fascinatingly, there is a diminished impact upon CBP-TAZ1:p53TAD binding for mutants that showed compromise in CBP-TAZ2:p53TAD interactions. Although both domains have 40-fold difference in p53TAD-WT binding sensitivity, they do share a nearly identical binding surface along p53TAD. We suggest that the difference in CBP-TAZ2 sensitivity to mutation is largely due to its observed high k_{on} value near $8.36 \cdot 10^4 \text{M}^{-1}\text{s}^{-1}$. Such a high k_{on} value so close to the diffusion limit for binding indicates that the CBP-TAZ2:p53TAD complex formation occurs almost immediately. It could then be inferred that in the free state p53TAD-WT adopts conformations which facilitate binding with its partners. And, that local or distal mutations within the p53TAD can affect high-affinity interactions through changing its rate of conformational interconversion.

7.4. Conclusions

While it has been well established that loss-of-function point mutations throughout p53's DBD leads to structural compromise and NRD mutations largely compromise regulation through ablation of PTM sites, little is known about the structural and biological consequences of point mutations within the TAD as it is an IDP. Through this work we analyzed four common p53TAD mutations, -K24N, -N29K/N30D, -D49Y, and -W53G against -WT using a combination of NMR spectroscopic techniques and BLI studies with domains from its suppressor, HDM2, and its

activator, CBP-TAZ1/2. NMR structural studies revealed only minor effects to the structural ensemble for p53TAD-K24N and -N29K/N30D mutants displaying compaction or diminishing structure. Assessment of p53TAD variants through RSDA revealed that in the case of mutant -N29K/N30D the motion of residues involved in binding to both CBP-TAZ and HDM2 domains are affected. While p53TAD-D49Y and -W53G have small localized differences in dynamics that affects the motions of residues solely involved in interactions with CBP-TAZ domains. These subtle detected differences in ps timescale motions correlate with observed loss of binding to regulatory partners with sub-micromolar affinities. This implies that p53TAD-WT in solution readily adopts binding compliant conformations. In addition, local or distal mutations within the p53TAD can compromise high-affinity interactions mostly through changing its rate of conformational interconversion. Future analysis of the p53TAD through molecular dynamics approaches and further RSDA by NMR studies with PTM versions of select mutations will elucidate biological consequences, which may improve prognostic outcomes.

7.5. References

- Abragam, A., & Abragam, A. (1961). *The principles of nuclear magnetism* (No. 32). Oxford university press, London.
- Abyzov, A., Salvi, N., Schneider, R., Maurin, D., Ruigrok, R. W., Jensen, M. R., & Blackledge, M. (2016). Identification of dynamic modes in an intrinsically disordered protein using temperature-dependent NMR relaxation. *Journal of the American Chemical Society*, *138*(19), 6240-6251.
- Allen, M. A., Andrysiak, Z., Dengler, V. L., Mellert, H. S., Guarnieri, A., Freeman, J. A., Sullivan, K.D., Galbraith, M.D., Luo, X., Kraus, W.L. & Dowell, R. D. (2014). Global analysis of p53-regulated transcription identifies its direct targets and unexpected regulatory mechanisms. *Elife*, *3*, e02200.
- Baker, S. J., Fearon, E. R., Nigro, J. M., Preisinger, A. C., Jessup, J. M., Ledbetter, D. H., Barker, D.F., Nakamura, Y., White, R., & Vogelstein, B. (1989). Chromosome 17 deletions and p53 gene mutations in colorectal carcinomas. *Science*, *244*(4901), 217-221.
- Bech-Otschir, D., Kraft, R., Huang, X., Henklein, P., Kapelari, B., Pollmann, C., & Dubiel, W. (2001). COP9 signalosome-specific phosphorylation targets p53 to degradation by the ubiquitin system. *The EMBO journal*, *20*(7), 1630-1639.
- Bode, A. M., & Dong, Z. (2004). Post-translational modification of p53 in tumorigenesis. *Nature Reviews Cancer*, *4*(10), 793-805.
- Brooks, C. L., & Gu, W. (2003). Ubiquitination, phosphorylation and acetylation: the molecular basis for p53 regulation. *Current opinion in cell biology*, *15*(2), 164-171.

- Chao, C., Wu, Z., Mazur, S. J., Borges, H., Rossi, M., Lin, T., Wang, J.Y., Anderson, C.W., Appella, E., & Xu, Y. (2006). Acetylation of mouse p53 at lysine 317 negatively regulates p53 apoptotic activities after DNA damage. *Molecular and cellular biology*, 26(18), 6859-6869.
- Chen, L., & Chen, J. (2003). MDM2-ARF complex regulates p53 sumoylation. *Oncogene*, 22(34), 5348-5357.
- Chen, J., Marechal, V., & Levine, A. J. (1993). Mapping of the p53 and mdm-2 interaction domains. *Molecular and cellular biology*, 13(7), 4107-4114.
- Cho, Y., Gorina, S., Jeffrey, P. D., & Pavletich, N. P. (1994). Crystal structure of a p53 tumor suppressor-DNA complex: understanding tumorigenic mutations. *Science*, 265(5170), 346-355.
- Chuikov, S., Kurash, J. K., Wilson, J. R., Xiao, B., Justin, N., Ivanov, G. S., McKinney, K., Tempst, P., Prives, C., Gamblin, S.J., & Barlev, N. A. (2004). Regulation of p53 activity through lysine methylation. *Nature*, 432(7015), 353-360.
- Cordenonsi, M., Montagner, M., Adorno, M., Zacchigna, L., Martello, G., Mamidi, A., Soligo, S., Dupont, S., & Piccolo, S. (2007). Integration of TGF- β and Ras/MAPK signaling through p53 phosphorylation. *Science*, 315(5813), 840-843.
- Craig, A. L., Burch, L., Vojtesek, B., Mikutowska, J., Thompson, A., & Hupp, T. R. (1999). Novel phosphorylation sites of human tumour suppressor protein p53 at Ser20 and Thr18 that disrupt the binding of mdm2 (mouse double minute 2) protein are modified in human cancers. *Biochemical Journal*, 342(1), 133-141.

- Dawson, R., Müller, L., Dehner, A., Klein, C., Kessler, H., & Buchner, J. (2003). The N-terminal domain of p53 is natively unfolded. *Journal of molecular biology*, 332(5), 1131-1141.
- DeHart, C. J., Chahal, J. S., Flint, S. J., & Perlman, D. H. (2014). Extensive post-translational modification of active and inactivated forms of endogenous p53. *Molecular & Cellular Proteomics*, 13(1), 1-17.
- Dolgin, E. (2017). The most popular genes in the human genome. *Nature*, 551(7681), 427.
- Dumaz, N., Milne, D. M., & Meek, D. W. (1999). Protein kinase CK1 is a p53-threonine 18 kinase which requires prior phosphorylation of serine 15. *FEBS letters*, 463(3), 312-316.
- Farrow, N. A., Zhang, O., Szabo, A., Torchia, D. A., & Kay, L. E. (1995). Spectral density function mapping using ^{15}N relaxation data exclusively. *Journal of biomolecular NMR*, 6(2), 153-162.
- Ferreon, J. C., Lee, C. W., Arai, M., Martinez-Yamout, M. A., Dyson, H. J., & Wright, P. E. (2009). Cooperative regulation of p53 by modulation of ternary complex formation with CBP/p300 and HDM2. *Proceedings of the National Academy of Sciences*, 106(16), 6591-6596.
- Fraser, J. A., Madhumalar, A., Blackburn, E., Bramham, J., Walkinshaw, M. D., Verma, C., & Hupp, T. R. (2010). A Novel p53 Phosphorylation Site within the MDM2 Ubiquitination Signal II. A model in which phosphorylation at ser269 induces a mutant conformation to p53. *Journal of Biological Chemistry*, 285(48), 37773-37786.
- Fischer, N. W., Prodeus, A., Malkin, D., & Gariépy, J. (2016). p53 oligomerization status modulates cell fate decisions between growth, arrest and apoptosis. *Cell Cycle*, 15(23), 3210-3219.

- Gatti, A., Li, H. H., Traugh, J. A., & Liu, X. (2000). Phosphorylation of human p53 on Thr-55. *Biochemistry*, 39(32), 9837-9842.
- Gu, W., & Roeder, R. G. (1997). Activation of p53 sequence-specific DNA binding by acetylation of the p53 C-terminal domain. *Cell*, 90(4), 595-606.
- Hafner, A., Bulyk, M. L., Jambhekar, A., & Lahav, G. (2019). The multiple mechanisms that regulate p53 activity and cell fate. *Nature Reviews Molecular Cell Biology*, 20(4), 199-210.
- Hollstein, M., Sidransky, D., Vogelstein, B., & Harris, C. C. (1991). p53 mutations in human cancers. *Science*, 253(5015), 49-53.
- Huang, J., Perez-Burgos, L., Placek, B. J., Sengupta, R., Richter, M., Dorsey, J. A., Kubicek, S., Opravil, S., Jenuwein, T., & Berger, S. L. (2006). Repression of p53 activity by Smyd2-mediated methylation. *Nature*, 444(7119), 629-632.
- Huang, J., Dorsey, J., Chuikov, S., Zhang, X., Jenuwein, T., Reinberg, D., & Berger, S. L. (2010). G9a and Glp methylate lysine 373 in the tumor suppressor p53. *Journal of Biological Chemistry*, 285(13), 9636-9641.
- Jensen, M. R., Markwick, P. R., Meier, S., Griesinger, C., Zweckstetter, M., Grzesiek, S., Bernado, P., & Blackledge, M. (2009). Quantitative determination of the conformational properties of partially folded and intrinsically disordered proteins using NMR dipolar couplings. *Structure*, 17(9), 1169-1185.
- Kalid, O., & Ben-Tal, N. (2009). Study of MDM2 binding to p53-analogues: affinity, helicity, and applicability to drug design. *Journal of chemical information and modeling*, 49(4), 865-876.

- Kandoth, C., McLellan, M. D., Vandin, F., Ye, K., Niu, B., Lu, C., Xie, M., Zhang, Q., McMichael, J.F., Wyczalkowski, M.A., & Leiserson, M. D. (2013). Mutational landscape and significance across 12 major cancer types. *Nature*, *502*(7471), 333-339.
- Katayama, H., Sasai, K., Kawai, H., Yuan, Z. M., Bondaruk, J., Suzuki, F., Fujii, S., Arlinghaus, R.B., Czerniak, B.A., & Sen, S. (2004). Phosphorylation by aurora kinase A induces Mdm2-mediated destabilization and inhibition of p53. *Nature genetics*, *36*(1), 55-62.
- Kay, L. E., Torchia, D. A., & Bax, A. (1989). Backbone dynamics of proteins as studied by nitrogen-15 inverse detected heteronuclear NMR spectroscopy: application to staphylococcal nuclease. *Biochemistry*, *28*(23), 8972-8979.
- Khan, S. N., Charlier, C., Augustyniak, R., Salvi, N., Déjean, V., Bodenhausen, G., Lequin, O., Pelupessy, P., & Ferrage, F. (2015). Distribution of pico-and nanosecond motions in disordered proteins from nuclear spin relaxation. *Biophysical journal*, *109*(5), 988-999.
- Krois, A. S., Ferreon, J. C., Martinez-Yamout, M. A., Dyson, H. J., & Wright, P. E. (2016). Recognition of the disordered p53 transactivation domain by the transcriptional adapter zinc finger domains of CREB-binding protein. *Proceedings of the National Academy of Sciences*, *113*(13), E1853-E1862.
- Lawrence, M. S., Stojanov, P., Mermel, C. H., Robinson, J. T., Garraway, L. A., Golub, T. R., Meyerson, M., Gabriel, S.B., Lander, E.S., & Getz, G. (2014). Discovery and saturation analysis of cancer genes across 21 tumour types. *Nature*, *505*(7484), 495-501.
- Lee, M. S., Lee, S. O., Lee, M. K., Yi, G. S., Lee, C. K., Ryu, K. S., & Chi, S. W. (2019). Solution structure of MUL1-RING domain and its interaction with p53 transactivation domain. *Biochemical and biophysical research communications*, *516*(2), 533-539.

- Lefèvre, J. F., Dayie, K. T., Peng, J. W., & Wagner, G. (1996). Internal mobility in the partially folded DNA binding and dimerization domains of GAL4: NMR analysis of the N–H spectral density functions. *Biochemistry*, *35*(8), 2674-2686.
- Matsumoto, M., Furihata, M., & Ohtsuki, Y. (2006). Posttranslational phosphorylation of mutant p53 protein in tumor development. *Medical molecular morphology*, *39*(2), 79-87.
- Mavinahalli, J. N., Madhumalar, A., Beuerman, R. W., Lane, D. P., & Verma, C. (2010). Differences in the transactivation domains of p53 family members: a computational study. *BMC genomics*, *11*(1), S5.
- Oda, K., Arakawa, H., Tanaka, T., Matsuda, K., Tanikawa, C., Mori, T., Nishimori, H., Tamai, K., Tokino, T., Nakamura, Y., & Taya, Y. (2000). p53AIP1, a potential mediator of p53-dependent apoptosis, and its regulation by Ser-46-phosphorylated p53. *Cell*, *102*(6), 849-862.
- Oliner, J. D., Pietenpol, J. A., Thiagalingam, S., Gyuris, J., Kinzler, K. W., & Vogelstein, B. (1993). Oncoprotein MDM2 conceals the activation domain of tumour suppressor p53. *Nature*, *362*(6423), 857-860.
- Peng, J. W., & Wagner, G. (1992). Mapping of spectral density functions using heteronuclear NMR relaxation measurements. *Journal of Magnetic Resonance (1969)*, *98*(2), 308-332.
- Peng, J. W., & Wagner, G. (1995). Frequency spectrum of NH bonds in eglin c from spectral density mapping at multiple fields. *Biochemistry*, *34*(51), 16733-16752.
- Peng, J., Ren, K. D., Yang, J., & Luo, X. J. (2016). Mitochondrial E3 ubiquitin ligase 1: a key enzyme in regulation of mitochondrial dynamics and functions. *Mitochondrion*, *28*, 49-53.

- Saha, T., Kar, R. K., & Sa, G. (2015). Structural and sequential context of p53: A review of experimental and theoretical evidence. *Progress in biophysics and molecular biology*, *117*(2-3), 250-263.
- Sakaguchi, K., Sakamoto, H., Lewis, M. S., Anderson, C. W., Erickson, J. W., Appella, E., & Xie, D. (1997). Phosphorylation of serine 392 stabilizes the tetramer formation of tumor suppressor protein p53. *Biochemistry*, *36*(33), 10117-10124.
- Sakaguchi, K., Herrera, J. E., Saito, S. I., Miki, T., Bustin, M., Vassilev, A., Anderson, C.W., & Appella, E. (1998). DNA damage activates p53 through a phosphorylation–acetylation cascade. *Genes & development*, *12*(18), 2831-2841.
- Salvi, N., Abyzov, A., & Blackledge, M. (2016). Multi-timescale dynamics in intrinsically disordered proteins from NMR relaxation and molecular simulation. *The journal of physical chemistry letters*, *7*(13), 2483-2489.
- Schon, O., Friedler, A., Bycroft, M., Freund, S. M., & Fersht, A. R. (2002). Molecular mechanism of the interaction between MDM2 and p53. *Journal of molecular biology*, *323*(3), 491-501.
- Shi, X., Kachirskaia, I., Yamaguchi, H., West, L. E., Wen, H., Wang, E. W., Dutta, S., Appella, E., & Gozani, O. (2007). Modulation of p53 function by SET8-mediated methylation at lysine 382. *Molecular cell*, *27*(4), 636-646.
- Sykes, S. M., Stanek, T. J., Frank, A., Murphy, M. E., & McMahon, S. B. (2009). Acetylation of the DNA binding domain regulates transcription-independent apoptosis by p53. *Journal of Biological Chemistry*, *284*(30), 20197-20205.
- Tang, Y., Zhao, W., Chen, Y., Zhao, Y., & Gu, W. (2008). Acetylation is indispensable for p53 activation. *Cell*, *133*(4), 612-626.

- Tompa, P. (2002). Intrinsically unstructured proteins. *Trends in biochemical sciences*, 27(10), 527-533.
- Tibbetts, R. S., Brumbaugh, K. M., Williams, J. M., Sarkaria, J. N., Cliby, W. A., Shieh, S. Y., Taya, Y., Prives, C., & Abraham, R. T. (1999). A role for ATR in the DNA damage-induced phosphorylation of p53. *Genes & development*, 13(2), 152-157.
- Vise, P. D., Baral, B., Latos, A. J., & Daughdrill, G. W. (2005). NMR chemical shift and relaxation measurements provide evidence for the coupled folding and binding of the p53 transactivation domain. *Nucleic acids research*, 33(7), 2061-2077.
- Waterman, M. J., Stavridi, E. S., Waterman, J. L., & Halazonetis, T. D. (1998). ATM-dependent activation of p53 involves dephosphorylation and association with 14-3-3 proteins. *Nature genetics*, 19(2), 175-178.
- Wright, P. E., & Dyson, H. J. (2015). Intrinsically disordered proteins in cellular signalling and regulation. *Nature reviews Molecular cell biology*, 16(1), 18-29.