STUDIES OF HUMAN ARMET AND OF PEA APHID TRANSCRIPTS OF SALIVA PROTEINS AND THE UNFOLDED PROTEIN RESPONSE

by

JAMES ROBERT BALTHAZOR

B.A., Fort Hays State University. 2011
B.S., Fort Hays State University. 2011

AN ABSTRACT OF A DISSERTATION

submitted in partial fulfillment of the requirements for the degree

DOCTOR OF PHILOSOPHY

Biochemistry and Molecular Biophysics Graduate Group
College of Arts and Sciences

KANSAS STATE UNIVERSITY
Manhattan, Kansas

2015

Abstract

Armet is a bifunctional protein that is apparently universally distributed among multicellular animal species, vertebrate and invertebrate alike. A member of the Unfolded Protein Response, (UPR) Armet promotes survival in cells that are under endoplasmic-reticulum (ER) stress. I have carried out biophysical studies on human Armet looking for compounds that bind to Armet and hence could reduce its anti-apoptotic function, thus potentially joining the growing class of pro-apoptotic drugs. Performed primarily with ^{1}H-^{15}N HSQC NMR, ligand studies showed that approximately 60 of the 158 residues are potentially involved with binding. The 60 residues are distributed throughout both domains and the linker suggesting multi-domain interaction with the ligand. Circular dichroism studies showed heat denaturation in a two-step unfolding process with independent unfolding of both domains of Armet with Tm values near 68˚C and 83°C with the C-terminal domain unfolding first, as verified by ^{1}H-^{15}N HSQC NMR measurements.

I also provide the first identification of UPR transcripts in pea aphids, *Acyrthosiphon pisum,* the genetic model among aphids. I measured transcript abundance with hope of finding future transcriptional targets for pest mitigation. I identified 74 putative pea aphid UPR components, and all but three of the components have higher transcript levels in aphids feeding on plants than those that fed on diets. This activated UPR state is attributed to the need for saliva proteins for plant feeding.

Because aphids are agriculturally significant pests, and saliva is pivotal to their feeding on host plants, genes that encode saliva proteins may be targets for pest mitigation. Here I have sought the aphid's saliva proteome by combining results obtained in several laboratories by proteomic and transcriptomic approaches on several aphid species. With these data I constructed a tentative saliva proteome for the pea aphid by compiling, collating, and annotating the data from several laboratories. I used RNA-seq to verify the transcripts in pea aphid salivary glands, thus expanding the proposed saliva proteome from approximately 50 components to around 130 components, I found that transcripts of saliva proteins are upregulated during plant feeding compared to diet feeding.

STUDIES OF HUMAN ARMET AND OF PEA APHID TRANSCRIPTS OF SALIVA PROTEINS AND THE UNFOLDED PROTEIN RESPONSE

by

JAMES ROBERT BALTHAZOR

B.A., Fort Hays State University. 2011
B.S., Fort Hays State University. 2011

A DISSERTATION

submitted in partial fulfillment of the requirements for the degree

DOCTOR OF PHILOSOPHY

Biochemistry and Molecular Biophysics Graduate Group
College of Arts and Sciences

KANSAS STATE UNIVERSITY
Manhattan, Kansas

2015

Approved by:

Major Professor
Gerald R. Reeck

Copyright

Abstract

Armet is a bifunctional protein that is apparently universally distributed among multicellular animal species, vertebrate and invertebrate alike. A member of the Unfolded Protein Response, (UPR) Armet promotes survival in cells that are under endoplasmic-reticulum (ER) stress. I have carried out biophysical studies on human Armet looking for compounds that bind to Armet and hence could reduce its anti-apoptotic function, thus potentially joining the growing class of pro-apoptotic drugs. Performed primarily with ^{1}H-^{15}N HSQC NMR, ligand studies showed that approximately 60 of the 158 residues are potentially involved with binding. The 60 residues are distributed throughout both domains and the linker suggesting multi-domain interaction with the ligand. Circular dichroism studies showed heat denaturation in a two-step unfolding process with independent unfolding of both domains of Armet with Tm values near 68˚C and 83°C with the C-terminal domain unfolding first, as verified by ^{1}H-^{15}N HSQC NMR measurements.

I also provide the first identification of UPR transcripts in pea aphids, Acyrthosiphon pisum, the genetic model among aphids. I measured transcript abundance with hope of finding future transcriptional targets for pest mitigation. I identified 74 putative pea aphid UPR components, and all but three of the components have higher transcript levels in aphids feeding on plants than those that fed on diets. This activated UPR state is attributed to the need for saliva proteins for plant feeding.

Because aphids are agriculturally significant pests, and saliva is pivotal to their feeding on host plants, genes that encode saliva proteins may be targets for pest mitigation. Here I have sought the aphid's saliva proteome by combining results obtained in several laboratories by proteomic and transcriptomic approaches on several aphid species. With these data I constructed a tentative saliva proteome for the pea aphid by compiling, collating, and annotating the data from several laboratories. I used RNA-seq to verify the transcripts in pea aphid salivary glands, thus expanding the proposed saliva proteome from approximately 50 components to around 130 components, I found that transcripts of saliva proteins are upregulated during plant feeding compared to diet feeding.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

ATF6 – Activating Transcription Factor 6

BiP – Binding Immunoglobulin Protein

CD – Circular Dichroism

CDNF – Cerebral Dopamine Neurotrophic Factor

CDNN – Secondary Structure Deconvolution Software

c-UPR – Constitutive Unfolded Protein response

DEPC – Diethylpyrocarbonate

DTT – Dithiothreitol

ER – Endoplasmic Reticulum

ERAD – Endoplasmic Reticulum Associated Degradation

ERSE I – Endoplasmic Reticulum Stress Response Element I

ERSE II – Endoplasmic Reticulum Stress Response Element II

GMO – Genetically Modified Organism

GRP78 – Glucose Regulating Protein 78

HSQC – Heteronuclear Single Quantum Coherence spectroscopy

IPTG - Isopropyl β-D-1-Thiogalactopyranoside

IRE1 – Inositol Requiring Enzyme 1

i-UPR – Inducible Unfolded Protein Response

KSU – Kansas State University

MANF – Mesencephalic Astrocyte Derived Neurotrophic Factor

MS – Mass Spectroscopy

NMR – Nuclear Magnetic Resonance

NOE – Nuclear Overhauser Effect

PDB – Protein Data Bank

PERK – Protein Endoplasmic Reticulum Kinase

RNA-seq – Ribonucleic Acid Sequencing

RPKM – Reads Per Kilobase Per Million

TCEP – Tris-2-Carboxyethyl Phosphine

UPR – Unfolded Protein Response

XBP1 – Xbox Binding Protein 1

Acknowledgements

I would like to express my gratitude to my advisor Jerry Reeck, for his support, patience, and encouragement throughout my graduate studies. It is not often that one finds an advisor and colleague that always finds the time for listening to the little problems and roadblocks that unavoidably appear in the course of performing research. His humor was needed and appreciated and his advice was essential to the completion of this dissertation. Through Jerry, and my academic committee I was taught innumerable lessons and insights on the workings of academic research. To my lab mates, Matt Aksamit, Jarrod Bechard, and Dr. Pinakin Sukthankar, I will remember and cherish the great times we had forever. I would also like to thank Dr. Raman Chandrasekar for the countless hours he spent dissecting insects for this project, this work was monumental in itself.

Last, but not certainly not least, I would like to thank my parents, my wife's parents, and especially my wife Heather for her understanding and love during the past few years. We are each other's greatest advocates and without her and my children being patient with daddy's work schedule I could not have completed this monumental achievement. Their support, encouragement, and love, was in the end what made this dissertation possible.

Dedication

This work is dedicated to my wife Heather and our children, without their love and support throughout the years I am sure this task would have been insurmountable.

For my Parker Joelle; Daddy loves you and will miss you all the days of my life.

Preface

This dissertation represents a culmination of work and learning that has taken place over a period of four years (2011 - 2015). The Reeck lab group consisted of a small group of people, driven in different directions, but with a collective goal of learning. The lab was a good place to develop ideas, but its members were paramount in forming friendships that will last throughout my life.

The first chapter describes the biophysical interaction of the human protein Armet and a multitude of ligands analyzed by nuclear magnetic resonance and circular dichroism spectroscopic techniques.

Chapter two presents a look into proteins of the unfolded protein response (UPR) within the pea aphid. This chapter aims to identify putative orthologs of human UPR members in the pea aphid and evaluate their expression levels by RNA-seq analysis whereas it is thought that the UPR is primarily upregulated in the salivary gland due to feeding.

As a continuation to chapter two but with a more direct focus, chapter three focuses on the proteins that are found within the salivary gland itself and more specifically the proteins of pea aphid saliva. RNA-seq validation of putative orthologs from many aphid species aid in the determination of a proposed saliva protein proteome in the pea aphid, the model insect for aphids and validate this method of bio-statistical research.

Chapter 1 - Studies of Human Armet

Literature Review

Nomenclature

The protein under investigation in this chapter was originally called ARMET, which stood for arginine rich mutated in early stage tumors (Shridhar et al., 1996). But I will use the name Armet, which is intended to be simply a tag. The term ARMET was coined due to polymorphisms found in the N-terminal arginine-rich region and a sequencing error that changed the ATG start codon to AGG. At the time of discovery, this polymorphism had been reported in a variety of solid tumors; however, these polymorphisms were later shown to exist in normal tissues and therefore being no longer tumor-related, rendering the term ARMET incorrect (Evron et al., 1997).

Armet is also known as MANF or mesencephalic astrocyte derived neurotrophic factor for its secretory neurotrophic effects and extracellular function (Lindholm et al., 2008). For its intracellular function, the term Armet has been more widely used, such as in the unfolded protein response (UPR) to endoplasmic reticulum (ER) stress.

Structure

The crystal and Nuclear Magnetic Resonance (NMR) solution structures of human Armet and the mouse Armet NMR solution structure have been solved and show a helix-rich protein composed of two domains as shown in Figure 1.1. I will refer to the domains as the N-terminal (residues 1-94) and C-terminal domains (residues 103-158) joined by a linker (residues 95-102) as defined by the NMR and crystallography structure determinations (Hellman et al., 2010, Hoseki et al., 2010). For the duration of this dissertation, structures of Armet will maintain an orientation of the N-terminal domain above the C-terminal domain, as shown in Figure 1.1.

The Nuclear Overhauser Effect (NOE) is a common phenomenon observed by NMR where the transfer of nuclear spin polarization from one nuclear spin population to another occurs via cross-relaxation (Anet et al., 1965). Thus, atoms that are in close proximity to each other (5 angstroms) can give a NOE signal, whereas spin coupling is observed only when the atoms are connected by 2–3 chemical bonds. This effect essentially shows atoms in respect to one another which makes the determination of the relative orientations of atoms in a molecule

possible, producing a three-dimensional structure. No Nuclear Overhauser Effect (NOE) signal was observed between residues in the two domains, showing relatively high fluctuation in the orientations of the two domains (Hellman et al., 2010). This lack of NOE-mediated inter-domain restraints, including missing long and medium range constraints, led the authors to conclude that "the domains are not tightly packed to each other, but instead tumble as independent structural modules separated by the flexible linker" (Hellman et al., 2010).

The N-terminal domain contains five α-helices and one 3-10 helix, and the C-terminal domain contains three α-helices. Within the C-terminal domain, the first helix (α 6) is loosely formed and the two consecutive helices (α 7 and α 8) run in parallel in a helix-loop-helix arrangement. Two of eight cysteine residues found in Armet are located in the C-terminal domain (Cys127 and Cys130) and form a CXXC motif residing in the loop which connects helices α 7 & α 8. The other cysteine residues are found in the N-terminal domain and form these disulfides: Cys6-Cys93, Cys9-Cys82, and Cys40-Cys51 (Hellman et al., 2010). Alternate disulfide arrangements have been reported however in pea aphid and mouse Armet. In a mass spectroscopy (MS) based approach on mouse Armet, two differences from the pairings listed above, namely the existence of Cys6-Cys9 and Cys82-Cys93 were reported (Mizobuchi et al., 2007). In the pea aphid, the same two pairs were also found using a MS approach (Wang et al., 2015). Wang et al. report that "Both approaches, the elucidation of Armet's 3-dimensional structure and MS of Armet peptides are valid; neither supplants or invalidates the other as regards the disulfide bonding pattern" (Wang et al., 2015). They present the following hypothesis; "that Armet, whether mammalian or insect in origin, has alternative disulfide arrangements in a portion of the N-terminal domain" (Wang et al., 2015). The authors further suggest that the possibility of alternative disulfide pairings in the N-terminal domain could be important functionally in understanding Armet's intracellular and extracellular roles.

Armet Tissue Expression

The Human Protein Atlas portal is a publicly available database which can be accessed online at http://www.proteinatlas.org/ where millions of images show the spatial distribution of human proteins and transcripts in tissues. As one of the proteins that have been studied, Armet has been identified in the following tissue types including 44 different normal human tissues, 20 different cancer types, as well as 46 different human cell lines. Western blots and antibody

validation show that Armet is produced in all tissues and show high expression levels in tissues such as the liver, pancreas, stomach, intestines, central nervous tissues, and endocrine glands (Uhlen et al., 2015).

Intracellular Functions of Armet

Human Armet contains a signal peptide (MRRMRRMWATQGLAVALALS) for secretion through the ER-Golgi pathway (Oh-hashi et al. 2012). Armet's gene has been identified as up-regulated by ER stress where it promotes survival in different cell lines (Airavaara et al., 2009). In other words, it has been shown to be a member of the unfolded protein response. ER stress can also cause upregulation of Armet in pancreatic and fibroblast cells (Lee et al., 2003, Mizobuchi et al., 2007, Apostolou et al., 2008, Airavaara et al., 2009). The accumulation of misfolded proteins in the ER causes ER stress that initiates the UPR, a cellular response to evaluate and respond to ER stress, and the UPR can function either adaptively or apoptotically (Oslowski et al., 2011).

Expression of Armet is analogous to that of the molecular chaperone BiP/GRP78 which is also a UPR member, but GRP78 was shown to be mediated by the endoplasmic reticulum stress response element 1 (ERSE-I) which is frequently found in the promoters of ER chaperone genes, whereas the upregulation of Armet was shown to be mediated by an endoplasmic reticulum stress response element 2 (ERSE-II) (Mizobuchi et al., 2007), the second UPR gene discovered to be regulated by an ERSE-II element after ATF6 (Kokame et al., 2000). ERSE-II likely contributes to quality control of proteins within the ER (Kokame et al., 2000, Mizobuchi et al., 2007). Armet, when over-expressed in HeLa cells, inhibited cell proliferation and ER stress-induced cell death (Apostolou et al., 2008). Armet also counteracts tunicamycin-induced ER stress and apoptosis in primary neurons (Yu et al., 2010) and serum starvation-induced apoptosis in cardiomyocytes (Tadimalla et al., 2008).

Extracellular Functions of Armet (MANF)

Armet also has an extracellular function, namely neurotrophic activity (Lindholm, 2010). In neural cells apoptosis is important to maintain the neuronal population and apoptosis is neutralized by the intervention of neurotrophic factors targeted to rescue apoptotic neurons from death (Hellman et al., 2010). "Parkinson's disease is a chronic, progressive neurodegenerative disease where dopaminergic cells die most prominently in the area of substantia nigra" (Hellman

et al., 2010). Armet has been found to be one of the most potent exogenous factors protecting and repairing the dopaminergic neurons in a rat 6-hydroxydopamine model of Parkinson's disease (Hellman et al., 2010). Armet also rescues cortical neurons in a rat stroke model, and aided in slowed neuronal apoptosis (Hellman et al., 2010).

Armet (MANF) has sequence similarity to one other neurotrophic factor, cerebral dopamine neurotrophic factor (CDNF), originally found in neural tissues, but also found in non-neural tissues similarly to Armet (Lindholm, 2010). While the details of both Armet & CDNF's function are still unclear, Armet has been shown to protect against cerebral ischemia in vivo interfering with apoptosis, improving the survival of dopaminergic neurons in vitro (Airavaara et al., 2009, Petrova et al, 2003)

Role of RTDL in Retention or Secretion

Armet has a C-terminal ER retention motif, RTDL. This motif targets Armet for retention in the ER lumen. The motif has been shown to bind to the KDEL receptor, but with weaker affinity than KDEL (Raykhel et al., 2007). Thus it is possible that under basal, unstressed conditions, low expression of MANF and other proteins with non-classical KDEL ER retention signals could allow for their complete retention (Glembotski et al., 2012). Then upon ER stress, levels of ER stress response gene products with ER retention motifs would increase, while KDEL receptor levels would not change (Llewellyn et al., 1997). Due to different affinities to the KDEL receptor between KDEL and RTDL, it might be that the RTDL ER retention motif allows for the partial secretion of Armet under ER stress. In other words, Armet and other non-KDEL ER retention signal containing proteins, may be secreted due to the inefficient retention in the ER during ER stress (Glembotski et al., 2012).

This idea was strengthened when a engineered mutant that lacked the an ER retention motif was found to be secreted while the wild type and an engineered mutant form carrying the KDEL sequence at the C-terminus was retained in the cell (Glembotski et al. 2012). Over expression of GRP78 resulted in retaining essentially all of these three variants of MANF showing that under some conditions the ER retention signal was not necessary to retain the mutant lacking an ER retention motif (Glembotski et al., 2012). An interaction between Armet and GRP78 was shown to be not directly dependent on the RTDL / KDEL sequence to retain

Armet within the ER, and that it also interacts in a non-calcium dependent fashion (Glembotski et al., 2012, Oh-hashi et al. 2012, Henderson et al., 2012).

High Throughput Screening

Surface plasmon resonance (SPR) sensing has been used to study bio-molecular binding events and their kinetics in a label-free way (Campbell et al., 2007). This method uses an optical phenomenon that enables the detection of unlabeled interactions in real time between proteins and potential ligands. The utilization of label-free SPR systems gives the advantage over labeled methods, with increased sensitivity and reduced costs due to less interference of the signal, and cost associated with coupling of a label to the target (Kooyman et al., 1988). The high throughput screen mentioned in this dissertation was completed at the University of Kansas High Throughput Screening Laboratory.

1H-15N HSQC NMR Ligand Binding

The use of 1H-15N HSQC NMR to evaluate binding is well established. For example in Rauthu et al. (2014), "Defining the Potential of Aglycone Modifications for Affinity/Selectivity Enhancement against Medically Relevant Lectins: Synthesis, Activity Screening, and HSQC-Based NMR Analysis," of the use of mapping chemical shift changes upon addition of a ligand is used. In this case, the ligand utilized was p-nitrophenyl lactopyranoside and was screened against human galectins 1, 3, & 7, identifying a proposed contact site and evaluating affinity and selectivity to each galectin.

Therapies Targeted at Other UPR Components: GRP78

Research to target UPR components for drug discovery is not new by any means. For instance, a peptidomimetic targeting strategy that used a GRP78 binding peptide, discovered by "epitope-mapping," coupled to the peptide apoptotic moiety $(KLAKLAK)_2$ selectively killed breast cancer cells that expressed surface-localized GRP78 (Miao et al., 2013). The apoptotic moiety, originally discovered as an antimicrobial peptide was shown to have a cytotoxic function when coupled with other peptides (Ma et al., 2012). The use of "epitope-mapping" was achieved by circulating a pool of antibodies elicited against tumors in cancer patients in a flow cytometer in the presence of GRP78. "Hits" against GRP78 identified the protein as a target in prostate and breast cancer (Miao et al., 2013). The highest efficacy binding peptide identified from the

epitope-mapping process was WIFPWIQL. When coupled with the apoptotic moiety the final construct was WIFPWIQL-GG-$_D$(KLAKLAK)$_2$ (Miao et al., 2013).

My Research Direction

The research outlined in this chapter is intended to be early stages in targeting Armet for drug development. The identification of a compound which would limit the anti-apoptotic nature of Armet, thus acting in a pro-apoptotic fashion, could have potential to combat diseases associated with Armet, for instance, cancer. This basic research direction has yielded well over 100 pro-apoptotic drugs in various stages of development, including activation of the UPR molecular target XBP1 with the drug Xanthohumol, which targets chronic lymphocytic leukemia (Reed et al., 2004, Lust et al. 2009). Three mammalian eIF2 kinases including protein ER kinase (PERK), has been shown to be activated with flavonoid compounds, which inhibit the growth of human leukemia cells (Ito et al., 1999). Armet, as a drug itself could supplement current treatment methods where anti-apoptotic therapies are desired, i.e. Parkinson's disease (Hellman et al., 2010).

Materials and Methods

Standard Recombinant Human Armet Expression and Purification

In conjunction with Dr. Raman Chandrasekar I expressed N-terminal tagged Armet. To express and purify N-terminal 6X His-tagged human Armet, the transcript was amplified by PCR from a plasmid containing a N-terminal 6×His-tagged human Armet gene using the following primers;

5'-GGCCCTCGAGCTACAAATCGGTCCG-3'

5'-GCCCATGGGCCACCACCACCACCACCACctgcggccgggcgac-3'.

After cloning, the product was inserted into pET-28a-c(+) vector using NcoI and XhoI sites and confirmation by sequencing was performed. The protein expression, using *E. coli* strain BL21 (DE3) transformed with the recombinant plasmid, was cultured at 37 ℃ using LB medium until OD 600 reached 0.6. The recombinant protein was induced with a 1 mM IPTG addition for 4 h at 30 ℃. Cells were harvested by centrifugation and resuspended in 50 mM Na_2HPO_4, 300 mM NaCl, 10 mM imidazole, pH 8.5, known as Buffer 1.

This suspension of *E. coli* in Buffer 1 was subjected to sonication on ice, 10 times with a 50% duty cycle, for 1 min at a time with a 1 min rest between cycles with a Model CV17 sonicator probe and Vibra-Cell TM375 controller module. If large clumps remained, the sample was subjected to additional rounds of sonication. Lysates were centrifuged at 14,000 g for 20 min, and the supernatant was loaded on a Ni-NTA agarose column (Qiagen #30230).

The column was washed with 40 column volumes of 50 mM Na_2HPO_4, 300 mM NaCl, and 20 mM imidazole, known as Buffer 2. Then, elution containing the N-terminal 6X His-tagged human Armet commenced with 5 column volumes of 50 mM Na_2HPO_4, 300 mM NaCl, 300 mM imidazole, known as Buffer 3.

Recombinant Human Armet Expression and Purification for NMR

Armet was harvested from cells grown with a ^{15}N-labled ammonium nitrate obtained from Cambridge Isotopes as its nitrogen source. The ^{15}N labeled protein was dialyzed overnight at 4 ℃ into 50 mM Tris-HCl, 100 mM NaCl, pH 7.0, which I will refer to as NMR Buffer.

To express ^{15}N labeled protein, the method of expression as indicated in the standard recombinant human Armet expression and purification described above was used with appropriate changes to the growth medium. A starter culture of *E. coli* strain BL21 (DE3) transformed with the recombinant plasmid was cultured overnight at 37 ℃, centrifuged and resuspended with M9 medium (12.8 g $Na_2HPO_4 \bullet 7H_2O$, 3 g KH_2PO_4, 0.25 g NaCl, 1 g $15NH_4Cl$ in 1 L distilled H_2O) in filter sterilized 2 mL of 1 M $MgSO_4$, 20 mL of 20 % glucose, 100 uL of 1 M $CaCl_2$ in a larger growth chamber until OD 600 reached 0.6.

The induction time of the labeled recombinant protein was 24 h due to the lack of LB nutrients with 1 mM IPTG at 30 ℃.

Identification and Use of Ligands

Ligand discovery came from collaboration between Reeck's lab and the High Throughput Screening Laboratory on Kansas University's Lawrence campus. The library, consisting of approximately 5,000 compounds was used by both a Fuji and Enspire Biacore system, where surface plasmon resonance high throughput screening yielded several hits as potential ligands for Armet, including tetracycline, several tetracycline derivatives and other compounds. The tetracycline derivatives were chosen for my work due to a μM dissociation constant from demeclocycline as determined by KU HTS and shown in Figure 1.2. At the time of this

dissertation, the Reeck group was in negotiations to perform the remaining dissociation studies in collaboration with the KU HTS laboratory.

The tetracyclines and derivatives and non-tetracycline compounds were chosen on multiple criterions, namely the dissociation constant for the tetracyclines and the fact that all compounds were readily available and relatively inexpensive. Mitoxanthrone, although it lacks one cyclic ring in comparison to tetracycline derivatives, was chosen for study due to its similar ring structure. Cefoperazone was chosen due to its crude similarity to that of an unfolded peptide.

Tetracycline and its derivatives, Figure 1.3 are a group of broad-spectrum antibiotics whose general usefulness has been reduced with the onset of antibiotic resistance, but which remain the treatment of choice for some specific indications. Two non-tetracycline ligands seen in Figure 1.4 were also utilized in NMR experiments, also identified as possible ligands by the KU HTS laboratory and may also be possible lead compounds.

Other possible ligands (luteolin and fisetin) were used in circular dichroism studies and are a subclass of flavonoids and are widely distributed in a variety of fruits and vegetables. Structurally, the flavonol contains three cyclic rings and is a ketone containing compound.

Nuclear Magnetic Resonance (NMR) Spectroscopy

^{1}H-^{15}N HSQC NMR experiments were performed at 25˚C on a 500 MHz Varian NMR Superconducting Spectrometer System equipped with pulsed field gradient accessory, four channel detection system, two waveform generators, and a 5 mm latest generation carbon enhanced Cold Probe. NMR tubes were purchased from Wilmad Lab Glass (535-PP-7). The tubes were 5 mm thin wall, 7” long, and intended for use in 500 MHz and higher field strength magnets.

For all ligand NMR experimentation, a 5 mM ligand concentration was used. The ligands were dissolved at a 1M stock concentration in DMSO and added to the protein solution prior to being placed in the NMR tube. ^{1}H-^{15}N HSQC spectra were acquired after addition of each ligand at 5 mM concentration at 3 mg/mL (approximately 150 μM) ^{15}N labeled N-terminal 6X His-tagged human Armet in 50 mM Tris-HCl, 100 mM NaCl, pH 7.0, supplemented with 5% (v/v) D_2O for 4 h. NMR data were analyzed by using MestReNova software.

Circular Dichroism Spectroscopy

Circular dichroism spectroscopy was performed with a Jasco 815 spectropolarimeter using either a jacketed 1.0 cm path length cell or an unjacketed 0.1 cm path length cell. Spectra from 190 to 260 nm were acquired at room temperature every 1 nm at 2 sec per data point and a 1 nm band pass. Thermal denaturation experiments were done in the same fashion over a temperature scale of 25°C to 95°C. N-terminal 6X His-tagged human Armet in 50 mM Tris-HCl, 100 mM NaCl, pH 7.0, was used for all circular dichroism experiments. For experiments including a potential ligand, the concentration of ligand was up to 10 mM.

Results

Ligand Binding Studies

The NMR tool used was two-dimensional ^{1}H-^{15}N HSQC spectroscopy on ^{15}N labeled Armet. This allowed mapping of signal changes when ligands were added to the protein. Residue assignments in my spectra were made using literature NMR assignments (Hellman et al., 2010). The ^{1}H-^{15}N HSQC reference spectrum (Armet without any ligand) from my work is shown in Figure 1.5. This showed somewhat better separation of signals than the literature NMR spectrum (Hellman et al., 2010). Residues unassigned in the literature, which remained unassigned by me, are shown in Table 1.

Tryptophan residue 123 (W123), seen with both backbone and sidechain signals, is important to note because W123 may be vital for binding where the possibility of π-π bonding between the W123 sidechain and cyclic structure of the ligands can occur. Other noteworthy signals include lysine 114 (K114), which shows altered signals in the presence of all ligands tested. These signals and numerous others are changed upon ligand binding.

I identified residues that are likely involved with ligand binding by their shifts in the ^{1}H-^{15}N HSQC spectrum in the presence of ligands. For example, see Figure 1.6 for a spectrum acquired in the presence of tetracycline. After spectra in the presence of ligands were acquired, a difference spectrum, Figure 1.7 (tetracycline ligand) was produced by subtraction of the reference spectrum from that of a spectrum acquired in the presence of a ligand. Changes are listed in Table 2 created by addition of each tetracycline ligand. These changes are reported as an increase, decrease, or shift in the signal. Changes in signal intensity are indicative of changes in the protein's mobility and may identify those residues that are altered in the presence of ligand

but do not directly interact with the ligand. Shifts in the signal, whether they be in the nitrogen or hydrogen environment, or most commonly both may indicate residues that bind and are in direct contact with the ligand.

Additional spectra used to create Table 2 are found in Appendix A. Residues with changed ^{1}H-^{15}N HSQC signals for tetracycline and derivatives are mapped in Figure 1.8 onto the three dimensional structure of Armet. Seen in Figure 1.8, the majority of residues involved with binding occur within the C-terminal domain and those residues with altered signals in the N-terminal domain and linker appear to form a “face” of binding on an interior portion where a “clamshell” binding mode of action is proposed.

Table 3 is an expansion of Table 2 and summarizes the changes in ^{1}H-^{15}N HSQC signals including two non-tetracycline ligands. Non-tetracycline ^{1}H-^{15}N HSQC ligand studies are mapped in Figure 1.9 onto the three dimensional structure of Armet. Figure 1.9 shows similarly to Figure 1.8 where the majority of residues involved with binding occur within the C-terminal domain and those residues with altered signals in the N-terminal domain and linker again appear to form a “face” of binding on an interior portion where a “clamshell” binding mode of action is proposed.

Residues that have altered ^{1}H-^{15}N HSQC signals in common between tetracycline and non-tetracyclines are mapped onto the three dimensional structure of Armet in Figure 1.10. Shown in Figure 1.10, the altered signals in residues in both the N-terminal and C-terminal domains form a "face" on the interior portions where a "clamshell" interaction would be likely to occur. It is important to note that with different ligands, the residues with altered signals are also different. This is best seen by the color coding in Figure 1.10 where red residues are associated with tetracycline and derivatives, green signals associated with non-tetracyclines, and blue indicates shared or common altered residues. All of these alterations in signal point to two recurrent observations, where the majority of altered signals are seen within the C-terminal domain and both domains form two “faces” where a proposed “clamshell” mode of binding would take place.

Only two residues which yield altered signals in every ligand tested, namely K114 and W123. Both of these residues are also identical in all vertebrates aligned in this study which may point to their importance for ligand binding due to their conservation. While W123 may directly interact with possible π-π binding to the cyclic portions of the tested ligands, K114 is not directly

identified in the proposed binding pocket. K114 may show altered signals due to accommodation of the ligand and not direct binding.

Circular Dichroism Spectroscopy Studies

From Armet's room temperature circular dichroism spectrum, seen in Figure 1.11, a helix content of 46.2% was estimated by the circular dichroism deconvolution software CDNN downloaded from http://www.photophysics.com/tutorials/cdnn-secondary-structure-analysis. From the NMR and crystallographic structural determinations I calculated 68% helix content. This indicates that the CDNN software reports lower than true observed values however it can still be utilized to gain insight into thermal denaturation process.

Figure 1.12 shows CD spectra of Armet at temperatures from 50-95°C. The change from α helix to random coil occurs where intermediates of both forms evidently occur. These intermediates could include uniform unfolding across both domains or, more likely, it is possible that the unfolding from helix to random coil could be occurring independently in each domain. An idealized representation of α helix and random coil CD signals are shown in Appendix A. Thermal denaturation monitored at 222 nm is shown in Figure 1.13. The two-step decline in the signal is centered at two Tm values of 65°C and 83˚C. Not shown, a β-turn content of approximately 20% for each temperature is predicted by CDNN. Figure 1.14 shows percent helix versus temperature as determined by CDNN analysis.

I hypothesized that ^{1}H-^{15}N HSQC NMR could evaluate if the either domain was unfolding first or in concert. But evaluating the two step unfolding by NMR at high temperatures was outside the instrument's normal operating limits therefore manipulation of transition temperatures of the two-step unfolding of Armet was achieved by modification of buffer conditions. Lowering the pH, as seen in Figure 1.15, shows that lowering pH affects both transition temperatures. Thermal denaturation of Armet in the presence of guanidinium chloride, shown in Figure 1.16, mirrors the results of the pH studies. A combination of guanidinium chloride and pH was also utilized and is shown in Figure 1.17 in the presence and absence of TCEP (tris-2-carboxyethyl-phosphine), a water soluble reducing agent utilized to break disulfide bonds.

In the presence of TCEP alone, shown in Figure 1.18, no significant change was noted in Tm values. None of these studies however could reduce the transition temperature to a level that was suitable for the NMR experiment I had in mind.

Discussion

1H-15N HSQC NMR Buffer Optimization

In Hellman et al. (2010), NMR spectra were recorded on a Varian Unity INOVA 800 NMR spectrometer, operating at 800 MHz. That instrument, while similar to Kansas State University's Varian spectrometer, possesses a higher field strength magnet, and like KSU's instrument it also utilized cryo-probe technology. Therefore, the increased signal resolution in spectra acquired at KSU may stem from the buffer conditions. Their buffer conditions were: 10 mM bis-Tris, 50 mM NaCl, pH 6.8, supplemented with 7% (v/v) D_2O, whereas my buffer condition was 50 mM Tris-HCl, 100 mM NaCl, pH 7.0, 5% (v/v) D_2O. This difference in buffers may account for our increased resolution due to added protein stability with increased salt concentration.

Ligand Binding Studies and Proposed Binding Site

Ligands that alter the ^{1}H-^{15}N HSQC signals may lay the ground work for future drug based ligand studies. Table 3 identifies commonalities between ligands I have studied. A heat map analysis to observe the commonalities between tetracycline derivatives and non-tetracyclines is shown. The heat map identified two residues, K118 and W123 that have altered ^{1}H-^{15}N HSQC signals with all ligands tested. These observations may indicate that both residues are vital to binding and site directed mutagenesis could better evaluate that observation.

Although more residues are affected in the C-terminal domain, my data indicates both domains interact with the ligands I have studied. Because both domains have altered ^{1}H-^{15}N HSQC signals upon addition of ligands, I propose that the domains could come together in a “clamshell” fashion around the ligand. This thought is further solidified Figure 1.10 which shows that residues in the N-terminal domain involved with binding are on an inner portion of where a “clamshell” interaction might take place. In Figure 1.19, I propose a "clamshell" binding model in Armet with the ligand tetracycline.

Tetracycline as a ligand has been evaluated in multiple crystallographic studies with the Tet Repressor protein (Aleksandrov et al., 2007). Using PyMOL, I looked at the Tet Repressor protein, PDB file 2TRT, and evaluated residues that appear to bind or interact with tetracycline. This evaluation looked at the possible interactions between the amino acid residues of the protein and tetracycline which include hydrophobic interaction with residues such as alanine, isoleucine, and leucine. Hydrogen bonding with tetracycline is evident with the residues tyrosine and histidine, and a possible stacking interaction between phenylalanine and one of the rings of tetracycline appears to be possible.

In contrast to claims that the linker was flexible based on NOE constraints (Hellman et al., 2010), I believe that the linker acts as a rather stiff section between the domains, with rotation about both ends. NMR studies aid this idea stems from the 15 lowest energy minima generated by Hellman et al. 2010, in which the linker maintains a fairly linear orientation with rotation about its ends. In my attempt to mirror this observation, seen in the "clamshell" representation in Figure 1.19, the linker maintains a linear orientation; however it allows rotation of the two domains to encompass the ligand.

Changes in Positions of the N- and C-terminal Domains in Armet

The X-ray crystallography of human Armet and solution NMR of human Armet with Protein Data Bank (PDB) identification numbers of 2W51 and 2KVD respectively are shown in Figure 1.20 in an overlay (Parkash et al., 2009, Hellman et al., 2010). The human Armet crystal structure is not complete in the C-terminus due flexible portions of protein which do not contribute to electron density. When the human Armet crystal structure is shown versus human NMR solution structure, they show good superimposition within the N-terminal domain. Variation in the flexible loops within the N-terminal domain is expected due to flexibility of the loop segments but the C-terminal domain did not superimpose well at all.

To achieve improved superimposition the linker region within the X-ray structure was clipped in PyMOL. This is shown with the human Armet X-ray crystallography and solution NMR Armet with PDB identification numbers of 2W51 and 2KVD, respectively, in Figure 1.21 (Parkash et al., 2009, Hellman et al., 2010). Now the crystal structure of Parkash et al. (2009), superimposes well with that of the NMR solution structure in both domains.

The online 3D structure prediction software I-TASSER was utilized to study commonalities between human, mouse, nematode, and pea aphid Armet. When these amino acid sequences were submitted, all of their sequence similarity and three dimensional structures linked back to the mouse and human X-ray and solution NMR determinations. In other words, proteins from *C. elegans* and *A. pisum* showed no similarity to proteins of known structure other than mouse and human Armet.

The Role of Armet's Linker

Shown in Figure 1.22 is the conserved nature of the linker residues across multiple vertebrate species. Low energy minima calculations in NMR structural determinations show a rather rigid linker which appears to have a motion similar to a rotor (Hellman et al., 2010). The rotor like movement gives traction to my hypothesis that movement at ends of the linker allows for the "clamshell" like action of the two domains coming together in concert gripping a potential ligand. Interestingly, the conservation of the linker region of Armet and its rotation may be paramount to the binding of ligands such as tetracyclines and natural binding partners of Armet. This idea somewhat contrasts the description by Hellman et al. (2010) where they state the linker is “flexible.”

Analysis to confirm the structure of the linker was further tested by evaluation of phi/psi angles for the linker residues determined by NMR. Plotted in a Ramachandran plot, two residues Q100 & I101 were identified to contain possible β strand structure. The residues L95 & Y97 show left handed α helix characteristic while K96, K99, & D102 indicate right handed α helix. The residue D98 was identified in a disallowed location in the Ramachandran plot indicating an error in structure at that position.

Lower Thermal Stability of Armet's C-terminal Domain & Roles of Disulfide Bonds

My thought that the C-terminal domain of Armet has a lower thermal stability stems from its less packed structure and possibly the disulfide arrangement. Armet contains 8 conserved cysteine residues through many species shown in Figure 1.22. In thermal denaturation studies in the presence of TCEP, a disulfide reducing agent, interesting results occur. Under conditions that the literature points toward complete reducing conditions, the Tm of Armet’s two step denaturation is unchanged. Of numerous possibilities, I outline three options. One is the complete reduction of disulfide bonds which have no effect on thermal stability. The second is

the partial reduction of disulfide bonds limited to the C-terminal domain with no effect on Tm for either domain. The third option could be the partial reduction of disulfides in both domains. I believe option two to be correct. As shown in Figure 1.23, the disulfide bond in the C-terminus is clearly solvent exposed and should be available to reduction. However, due to the size of TCEP, I feel that the compound would not be capable of accessing the three N-terminal disulfide bonds. These insights point toward the conclusion that the C-terminal domain's disulfide, does not aid in thermal stability.

Shown in Figure 1.14, a plotted analysis with the CDNN prediction tool developed by Dr. Gerald Bohm shows a shift from a α helix beginning around 65°C to a random coil in a two-step fashion with the last transition at approximately 78°C. I had hoped for a stabilization to occur with the addition of ligand, especially in the C-terminal domain; thermal unfolding did not show any change in the presence of the possible ligands fisetin and luteolin.

Relevance for Proposed Functions of N- and C-terminal Domains of Armet

Crystallographic studies of Armet led Parkash et al. (2009) to suggest the C-terminal domain has a disulfide isomerase activity, and the N-terminal structure was shown to be similar to saposins. The researchers make a marked jump to conclude that because the structure of the N-terminal domain resembles that of the human saposins the N-terminal domain may interact with lipids or membranes. This suggestion in 2009, while interesting, was based solely on structural similarity has still not been shown experimentally to date. The data presented in this chapter I think revokes the thought process as suggested by Parkash et al. (2009) that the two domains of Armet have marked separate functions. Here I show that the binding of ligands encompass the use of both domains, and that the domains work in concert. Although the statements of Parkash et al. (2009) were interesting, they lacked the evidence to show that the N-terminal domain of Armet was more than just saposin-like based in structure. As no binding studies to lipids or membranes have been published to corroborate their claims, I believe my model whereas the N-terminal & C-terminal domains both participate in binding of a ligand holds more substance.

Variation in Amino Acid Sequence in Mammalian Armet

In Figure 1.24, a phylogenetic tree shows the inferred evolutionary relationships among sequences aligned in Figure 1.22. As shown, a clear evolutionary tree beginning with *C. elegans*

as an “outgroup” to humans occurs with subtle divergences of the orthologs of Armet showing a branching pattern in agreement with views of vertebrate evolution (Hotton 1968).

Figure 1.25 shows the conserved residues from Figure 1.22's alignment in green and those residues that differ from human Armet's sequence in yellow. Figure 1.26 identifies residues that are involved with binding and conserved. There are 32 residues that are conserved and 29 that are not conserved within the set of aligned sequences. Of those binding to tetracycline and its derivatives versus non-tetracycline ligands, no discernible pattern can be found, indicating that neither type of ligand is more likely to be bound at a conserved or non-conserved residue with respect to the human sequence.

In any case, while not every identified binding residue interacts in the presence of each ligand, conservation of residues involved with binding is seen in Figure 1.22.

Possible Drug Development Strategies

Previously mentioned, Armet expression has been shown to be upregulated in many cancer cells (Miao et al., 2013). GRP78 has also been shown to be upregulated in many cancer cells and the mode of treatment outlined in Miao et al. (2013) could possibly be approached with Armet as the target for peptidomimetic studies for cancer therapy.

While it’s not understood yet which elements of tetracycline derivatives induce NMR spectral changes, the fact that all derivatives have the similar planar shape leads me to believe that the planar rings of the ligands are important for binding. As shown in Figure 1.3, the structures of the tetracycline derivatives differ slightly and with modification by synthetic chemists, they may be exploited to create a better ligand.

I also propose an interaction between residue W123 by π bonding to the ligands including Armet’s normal binding partners, namely unfolded proteins. Figure 1.1 indicates the location of the W123 sidechain and for the proposed interaction to take place, the C-terminal domain folding in the “clamshell” fashion would allow the W123 sidechain, which has been indicated in ^{1}H-^{15}N HSQC studies to have altered signals in every ligand tested to facilitate π bonding. This suggests an importance of the interaction I have proposed. As seen in the cartoon representation of a proposed “clamshell” binding in Figure 1.19, the possibility of a π-π interaction exists. The π-π interaction between the tryptophan and the cyclic rings of the derivatives tested might contribute to binding ligands containing cyclic rings.

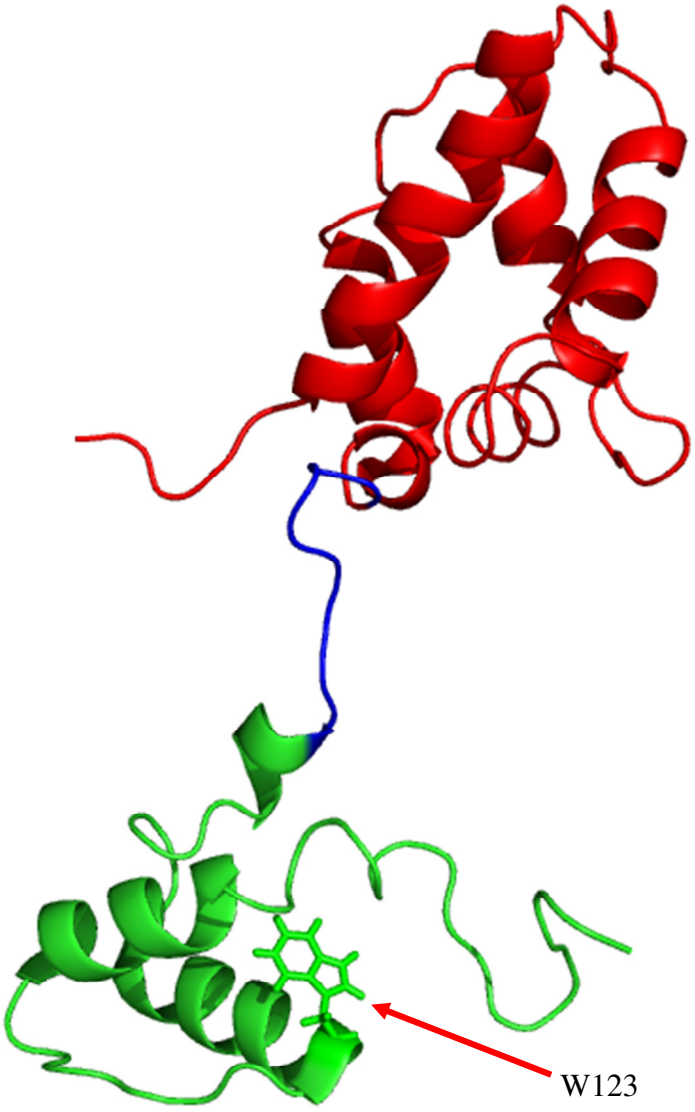


Figure 1.1 Lowest energy NMR structure of human Armet

Human Armet structure: N-terminal domain (red) linker region (blue) and C-terminal domain (green) PDB: 2KVD (Hellman et al., 2010) W123 sidechain is shown.

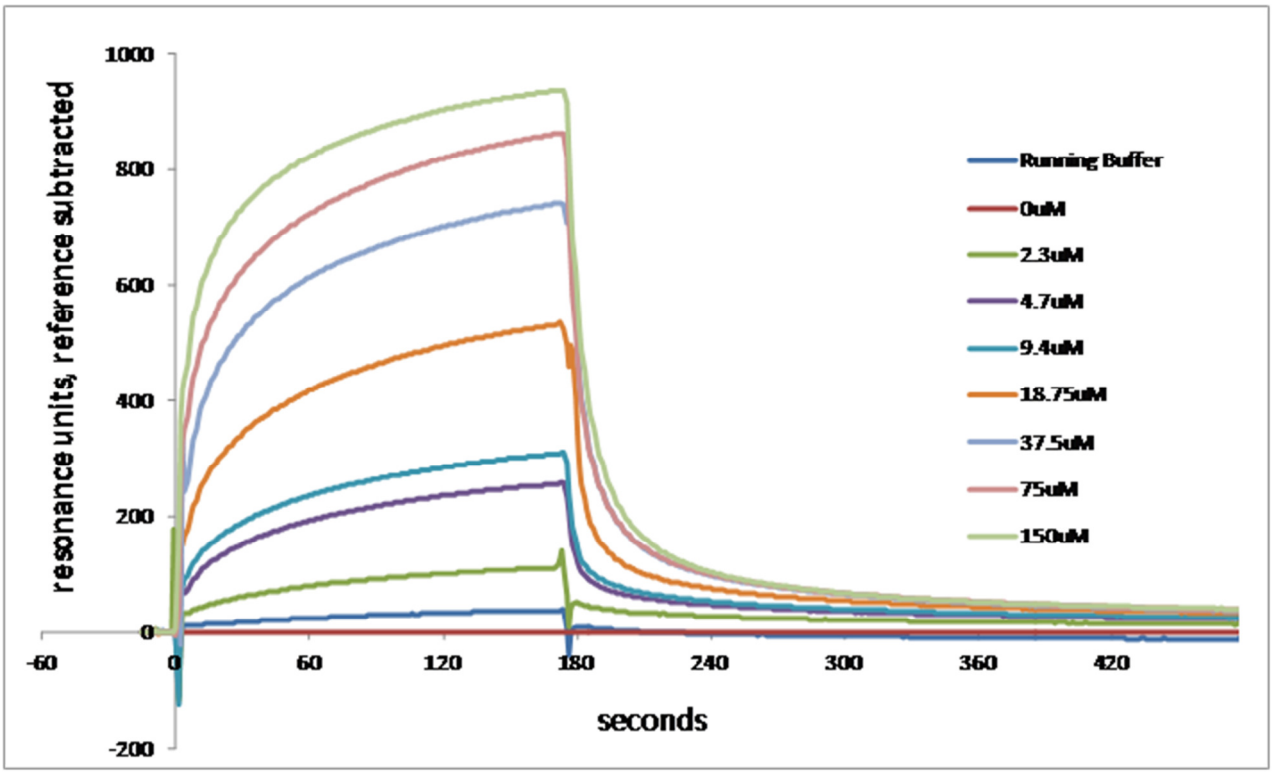


Figure 1.2 Determination of binding constant for demeclocycline and Armet

Demeclocycline binding analysis for ligand evaluation later used in binding experiments via NMR as determined at the KU HTS lab with surface plasmon resonance as a potential ligand. The data correspond to a dissociation constant of 18 μM measured at one half Vmax.

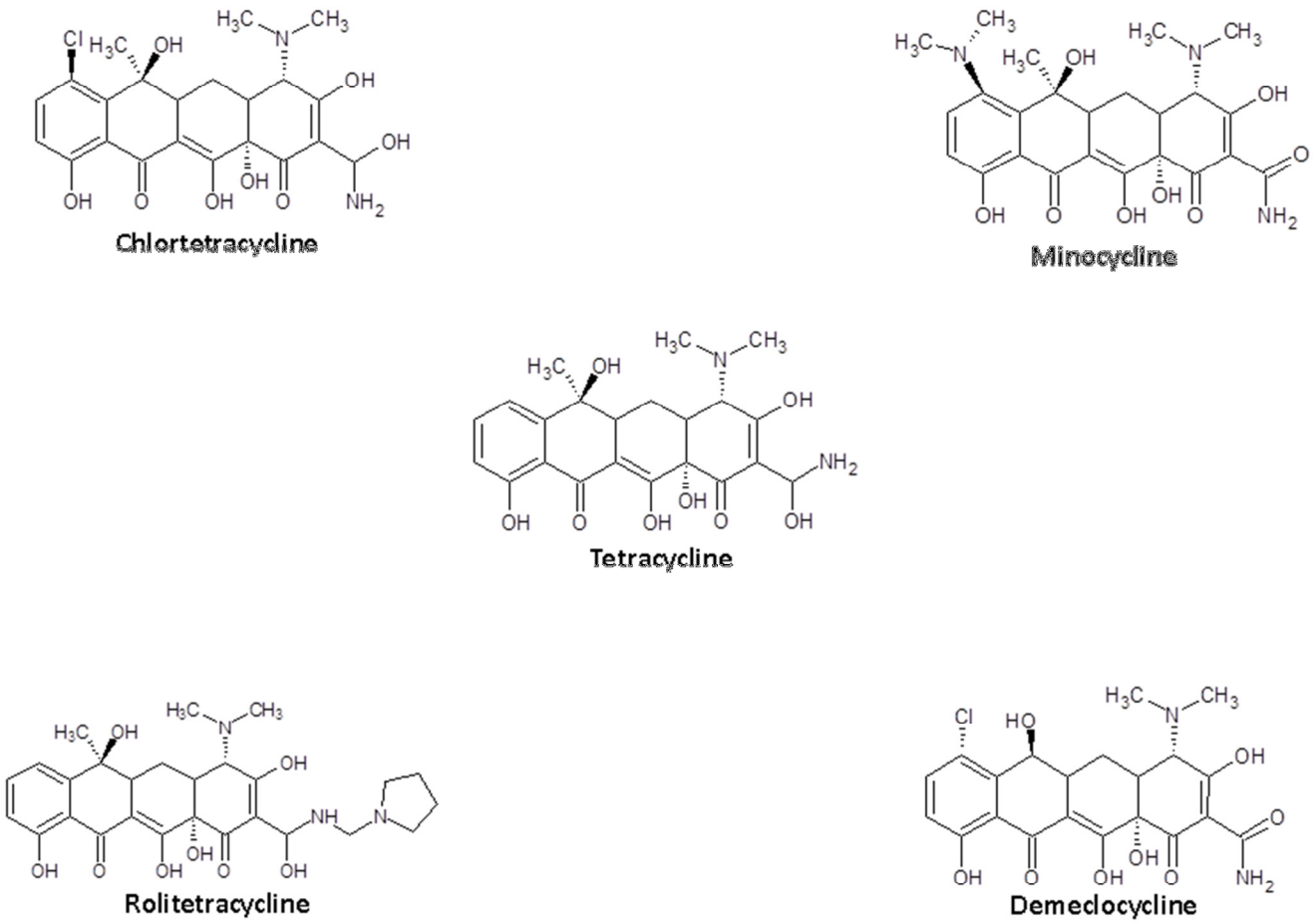


Figure 1.3 Structures of tetracycline and derivatives utilized in ^{1}H-^{15}N HSQC experiments

Images created using ChemDraw.

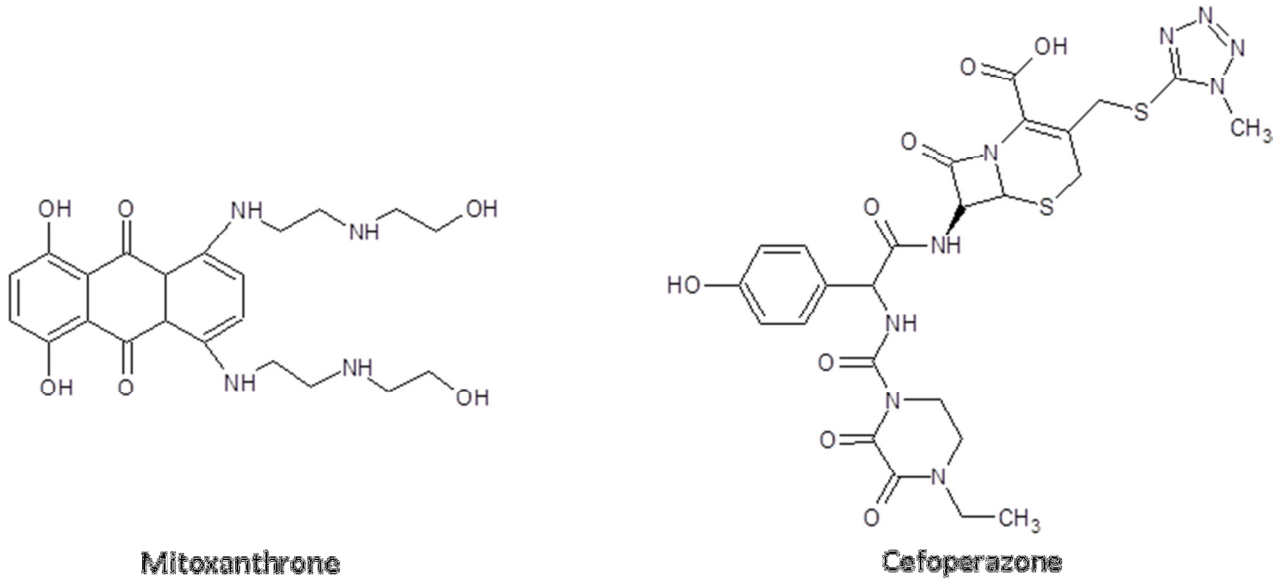


Figure 1.4 Structures of two non-tetracycline compounds utilized in ^{1}H-^{15}N HSQC experiments

Images created using ChemDraw.

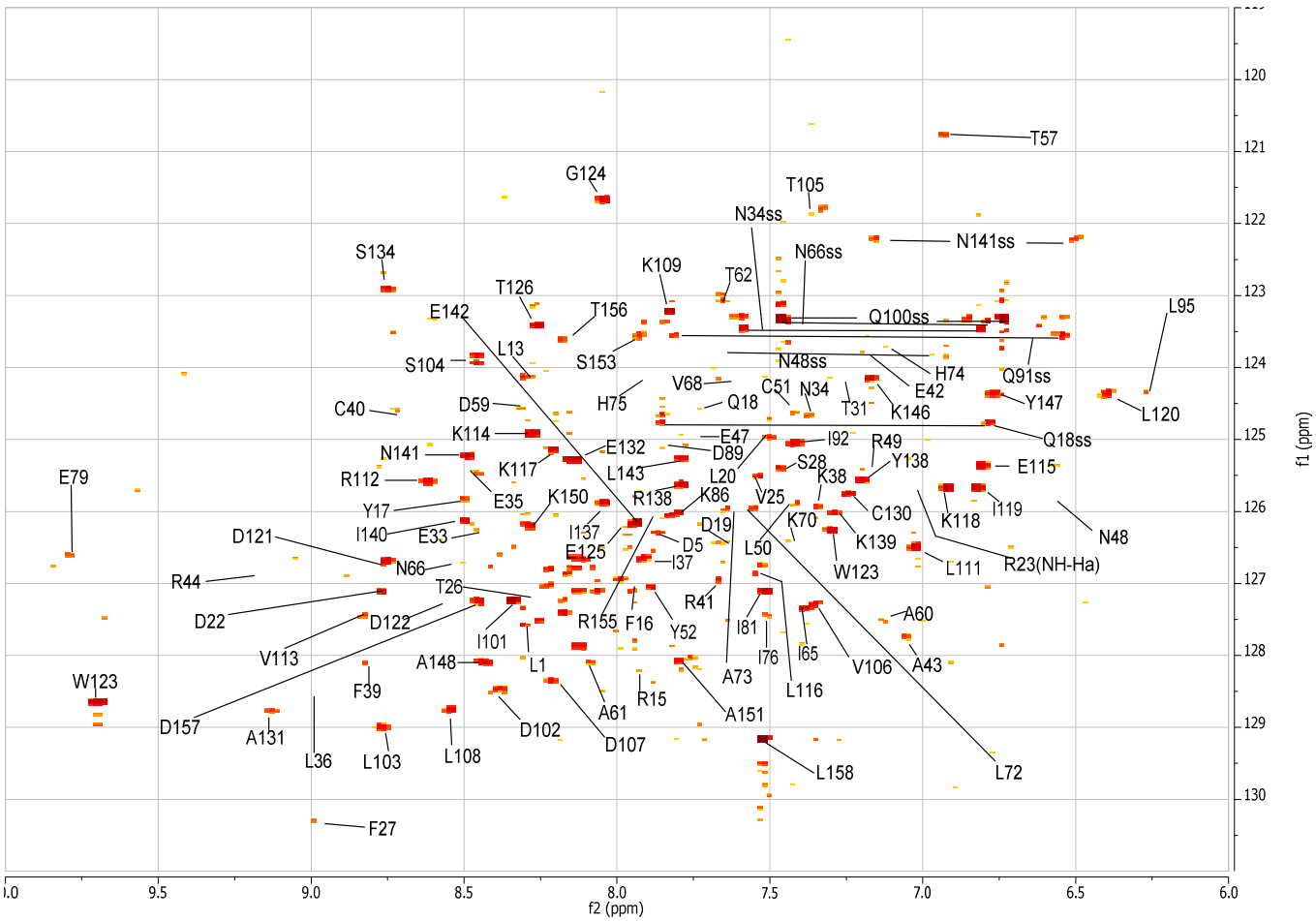


Figure 1.5 Reference ^{1}H-^{15}N HSQC spectrum of Armet

Acquired at 25 °C and annotated using the M-nova NMR suite (MestReNova Labs). Identification of signals achieved using assignments of Hellman et al. (2010). Unassigned residues are indicated in Table 1.

Unassigned Residues										
	Gly	**Cys**	**Glu**	**Val**	**Ile**	**Ser**	**Tyr**	**Phe**	**Lys**	**Asp**
Residue Number	4	6	7	8	10	11	12	16	21	59
	45	9	83				136		46	
	124	93	94						63	
	129	127							128	

Table 1 Residues not identified in the reference ^{1}H-^{15}N HSQC spectrum

Not all residues were assigned in the reference HSQC spectrum of Hellman et al. (2010). In addition to the unassigned residues, 6 proline residues did not generate signals.

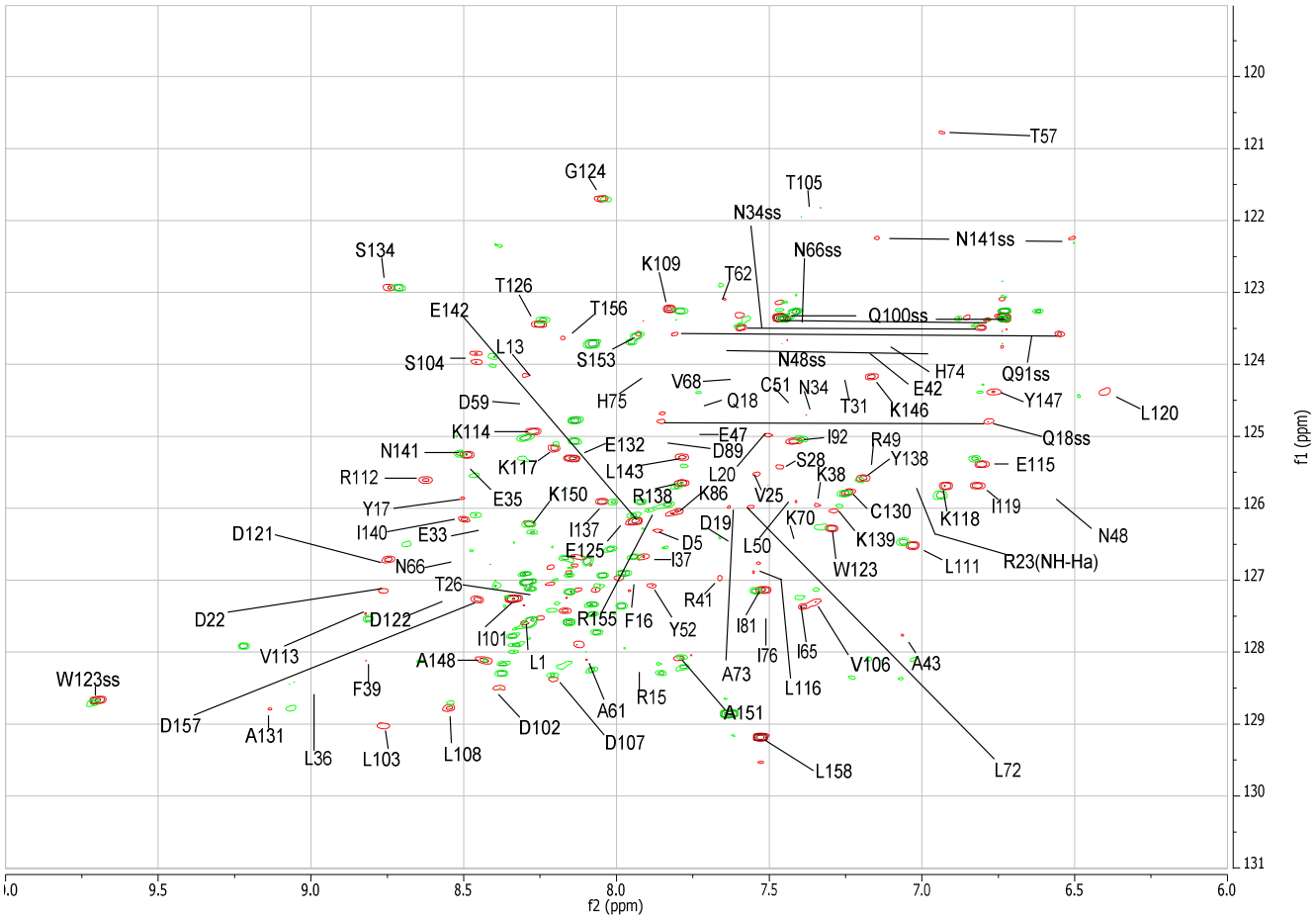


Figure 1.6 Overlay of ^{1}H-^{15}N HSQC spectra of Armet with and without 5 mM tetracycline

The protein without ligand is indicated in red while the spectrum containing the 5 mM tetracycline is shown in green. Changes in residues are determined by their movement in either the hydrogen or nitrogen ppm, F2 and F1 respectively or in intensity. Residues not identified (Table 1) and proline residues are not seen in this overlay. Residues that are altered are identified in Table 2.

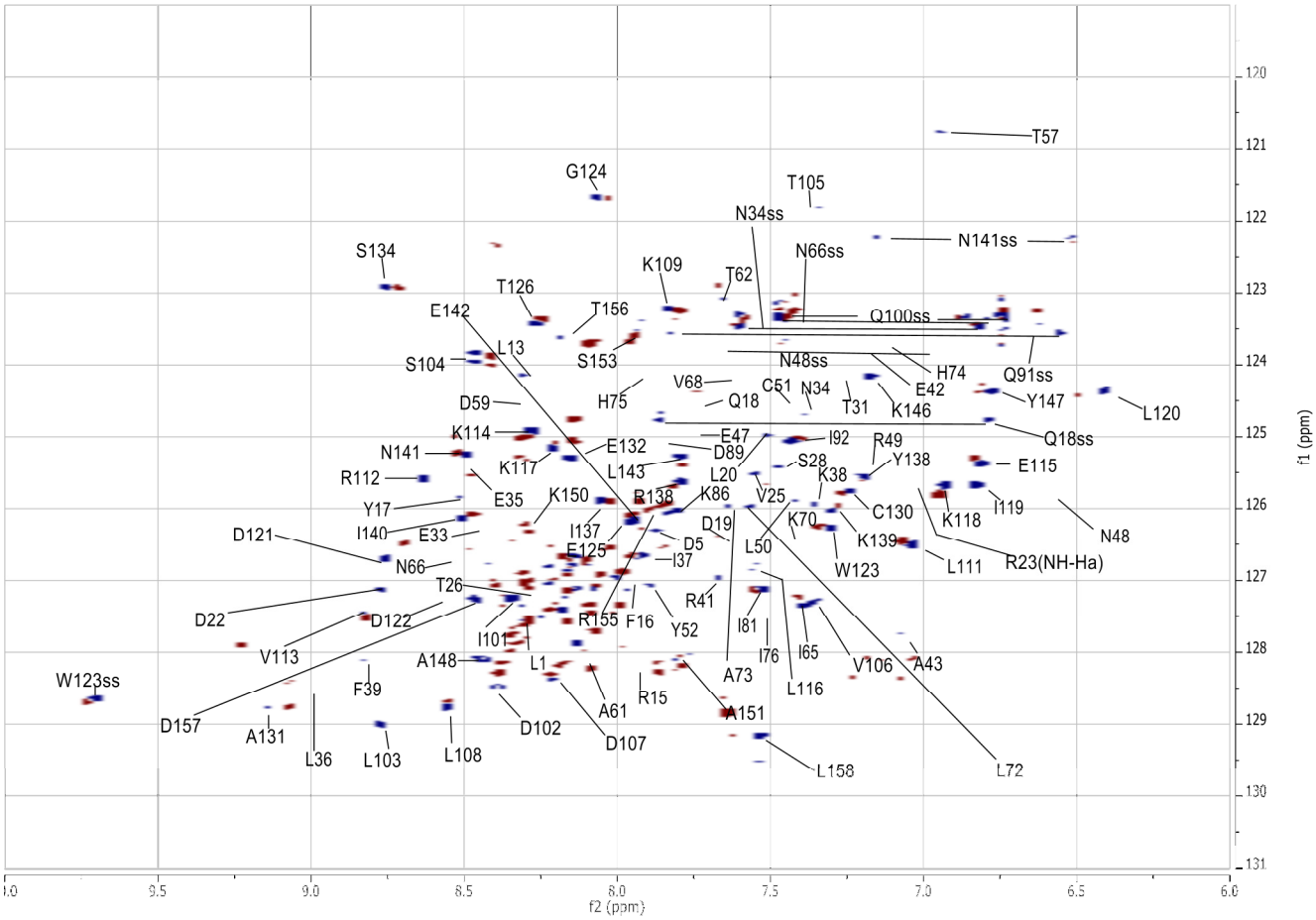


Figure 1.7 ^{1}H-^{15}N HSQC difference spectrum of Armet with and without 5 mM tetracycline

Difference spectrum is obtained from Figure 1.6. The signals indicated in red indicate a stronger signal without ligand present and a blue signal indicates an increased signal with the ligand present. These changes are identified in Tables 2 and 3. Changes in residues are determined by their movement in either the hydrogen or nitrogen ppm, F2 and F1 respectively or in their intensity.

Tetracycline			Chlortetracycline			Minocycline			Rolitetracycline			Demeclocycline		
Signal with Ligand			Signal with Ligand			Signal with Ligand			Signal with Ligand			Signal with Ligand		
Increased	Decreased	Shifted	Increased	Decreased	Shifted	Increased	Decreased	Shifted	Increased	Decreased	Shifted	Increased	Decreased	Shifted
L1	R44	T62	D122	I81	N34	Q100	I81	D5	N34	I81	K86	K86	T57	T62
E35	I81	I92	S153	I92	T62	D107	K139	I37	L158	K109	Q100	I101	S104	Q100
D107	I101	Q100		L111	S104	V113	I140	A43		L111	D107	E142	K109	D102
V113	D102	S104		K117	D107	S134	A151	K86		K114	K150		K114	D107
K118	L103	L108		Y136	K109	T156		D102		E115	T156		E115	I119
C130	K117	K109			K114			L103		K118			K118	W123
A151	I119	K114			E115			L111		I119			D157	T156
S153	A148	E115			K118			K114		W123				L158
		W123			W123			I119		G124				
		G124			G124			L120		T126				
		T126			T126			D121		C130				
		S134			C130			W123		S134				
		Y136			S134			E132		Y136				
		I140			N141			N141		N141				
		N141			A148			E142		D157				
		E142			L158			K150						
								L158						

Table 2 NMR Spectral changes in tetracycline derivatives

Signal changes in select in NMR studies. Changes in chemical shift as well as intensity with indicated ligand are indicated.

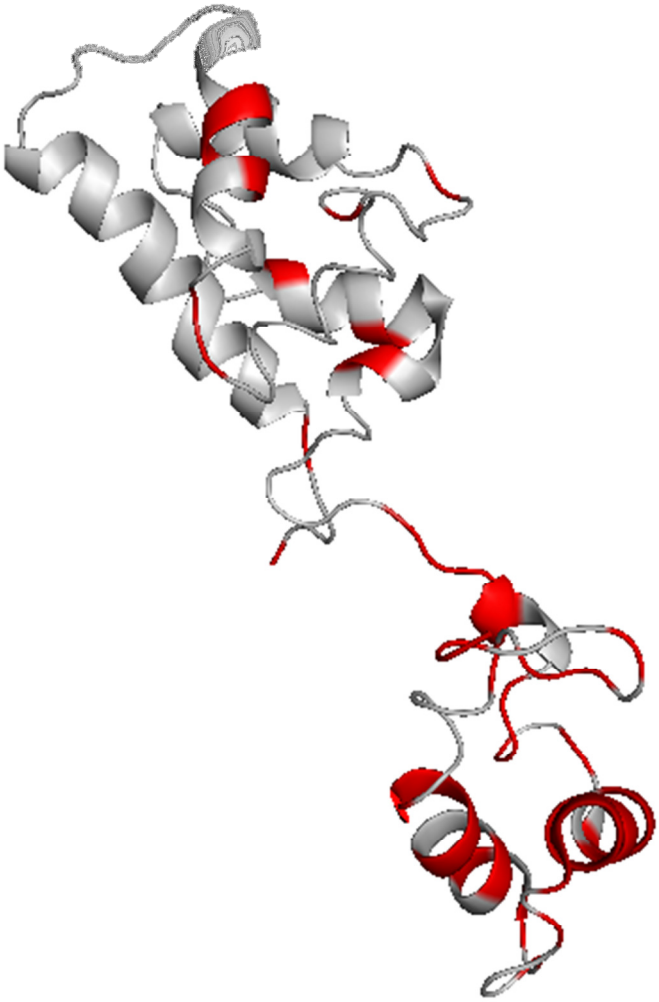


Figure 1.8 Residues with altered ^{1}H-^{15}N HSQC signals in the presence of tetracycline and tetracycline derivatives

Residues identified in Table 2 are indicated in red.

Domain	Residue	Tetracycline	Chlortetracycline	Minocycline	Rolitetracycline	Demeclocycline	Tetracycline Commonalities		Mitoxanthrone	Cefoparazone	All Commonalities
N-terminal	L1	Increase					1		Increase	Decrease	3
N-terminal	D5			Shift			1				1
N-terminal	L13						0		Shift		1
N-terminal	T26						0			Shift	1
N-terminal	N34		Shift		Increase		2				2
N-terminal	E35	Increase					1				1
N-terminal	I37			Shift			1				1
N-terminal	A43			Shift			1		Increase		2
N-terminal	R44	Decrease					1				1
N-terminal	T57					Decrease	1				1
N-terminal	T62	Shift	Shift			Shift	3		Increase		4
N-terminal	I65						0		Shift		1
N-terminal	N66						0		Increase	Increase	2
N-terminal	K70						0		Increase		1
N-terminal	I81	Decrease	Decrease	Decrease	Decrease		4			Decrease	5
N-terminal	K86			Shift	Shift	Increase	3				3
N-terminal	Q91						0			Decrease	1
N-terminal	I92	Shift	Decrease				2				2
Linker	Q100	Shift		Increase		Shift	3			Increase	4
Linker	I101	Decrease			Shift	Increase	3		Increase		4
Linker	D102	Decrease		Shift		Shift	3				3
C-terminal	L103	Decrease		Shift			2				2
C-terminal	S104	Shift	Shift			Decrease	3		Shift		4
C-terminal	T105						0		Shift		1
C-terminal	V106						0		Shift	Decrease	2
C-terminal	D107	Increase	Shift	Increase	Shift	Shift	5		Increase		6
C-terminal	L108	Shift					1		Increase		2
C-terminal	K109	Shift	Shift		Decrease	Decrease	4		Increase	Decrease	6
C-terminal	L111		Decrease	Shift	Decrease		3		Increase		4
C-terminal	V113	Increase		Increase			2				2
C-terminal	K114	Shift	Shift	Shift	Decrease	Decrease	5		Shift	Decrease	7
C-terminal	E115	Shift	Shift		Decrease	Decrease	4			Decrease	5
C-terminal	K117	Decrease	Decrease				2		Shift	Decrease	4
C-terminal	K118	Increase	Shift		Decrease	Decrease	4			Decrease	5
C-terminal	I119	Decrease		Shift	Decrease	Shift	4			Decrease	5
C-terminal	L120			Shift			1		Decrease		2
C-terminal	D121			Shift			1			Decrease	2
C-terminal	D122		Increase				1				1
C-terminal	W123	Shift	Shift	Shift	Decrease	Shift	5		Increase	Shift	7
C-terminal	G124	Shift	Shift		Decrease		3				3
C-terminal	E125						0			Increase	1
C-terminal	T126	Shift	Shift		Decrease		3		Increase	Decrease	5
C-terminal	C130	Increase	Shift		Decrease		3		Increase		4
C-terminal	E132			Shift			1		Increase	Increase	3
C-terminal	S134	Shift	Shift	Increase	Decrease		4				4
C-terminal	Y136	Shift	Decrease		Decrease		3				3
C-terminal	I137						0			Decrease	1
C-terminal	R138						0		Increase		1
C-terminal	K139			Decrease			1				1
C-terminal	I140	Shift		Decrease			2				2
C-terminal	N141	Shift	Shift	Shift	Decrease		4		Decrease	Decrease	6
C-terminal	E142	Shift		Shift		Increase	3			Decrease	4
C-terminal	Y147						0		Decrease		1
C-terminal	A148	Decrease	Shift				2		Shift	Decrease	4
C-terminal	K150			Shift	Shift		2				2
C-terminal	A151	Increase		Decrease			2		Shift		3
C-terminal	S153	Increase	Increase				2				2
C-terminal	T156			Increase	Shift	Shift	3				3
C-terminal	D157				Decrease	Decrease	2			Decrease	3
C-terminal	L158		Shift	Shift	Increase	Shift	4		Increase	Increase	6

Table 3 NMR Signal changes with all ligands by residue and domain

Signal changes identified by residue and ligand. The column titled tetracycline commonalities show a heat map style analysis of residues in common between tetracycline and derivatives. The column titled All commonalities is another heat map style analysis between all tested HSQC ligands.

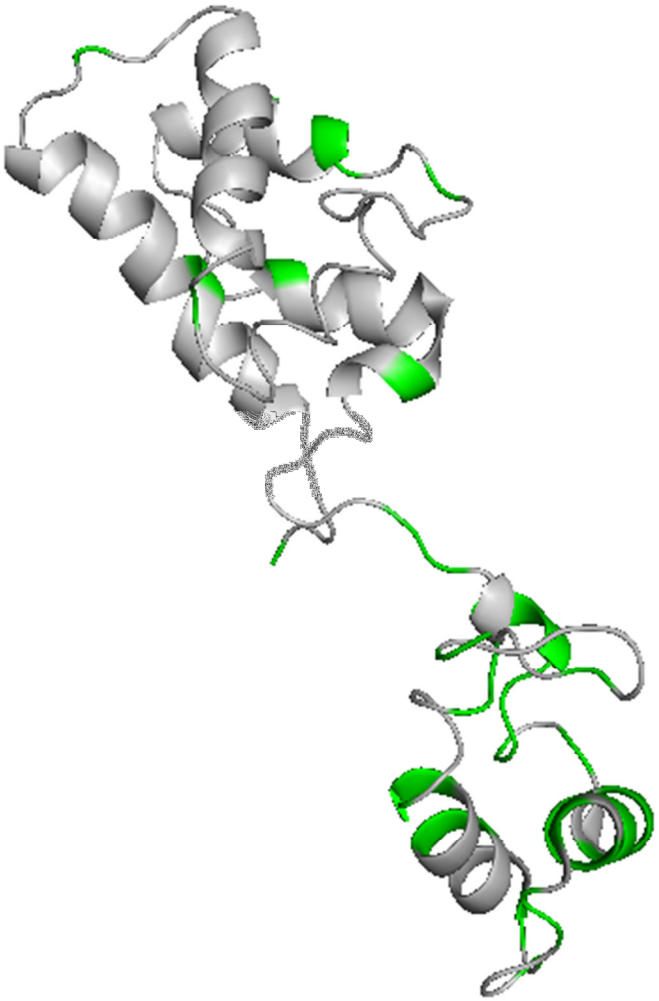


Figure 1.9 Residues with altered ^{1}H-^{15}N HSQC signals in the presence of mitoxanthrone and cefoperazone

Residues identified in Table 3 are indicated in green.

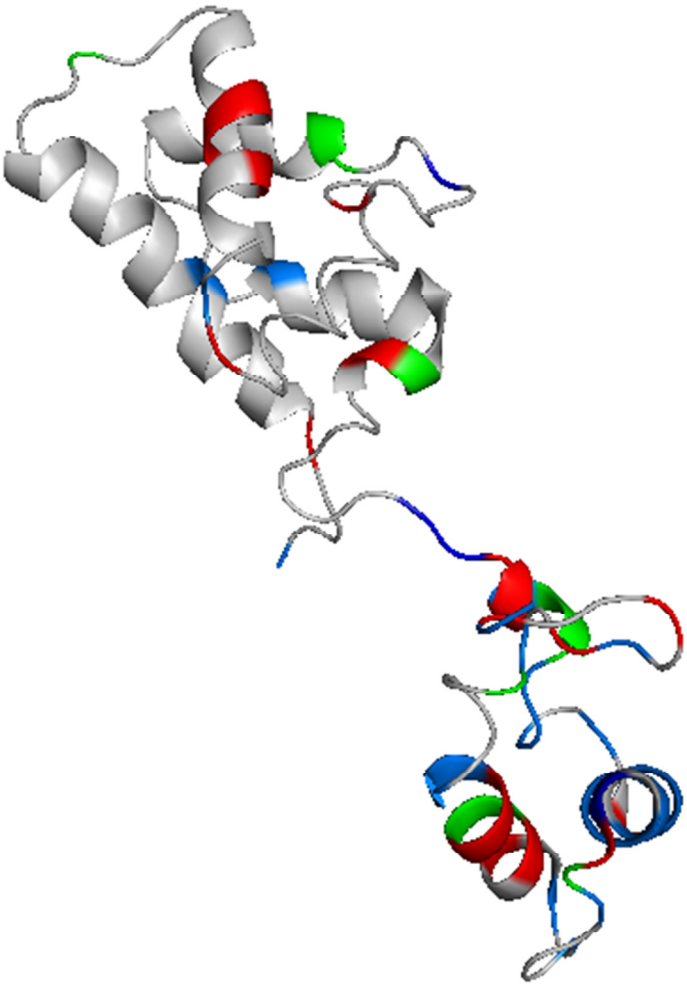


Figure 1.10 Residues with altered ^{1}H-^{15}N HSQC signals in the presence of tetracycline and non-tetracycline ligands

The coloring scheme is:

Tetracyclines altered residues (red)

Non-tetracyclines altered residues (green)

Shared or common altered residues (blue)

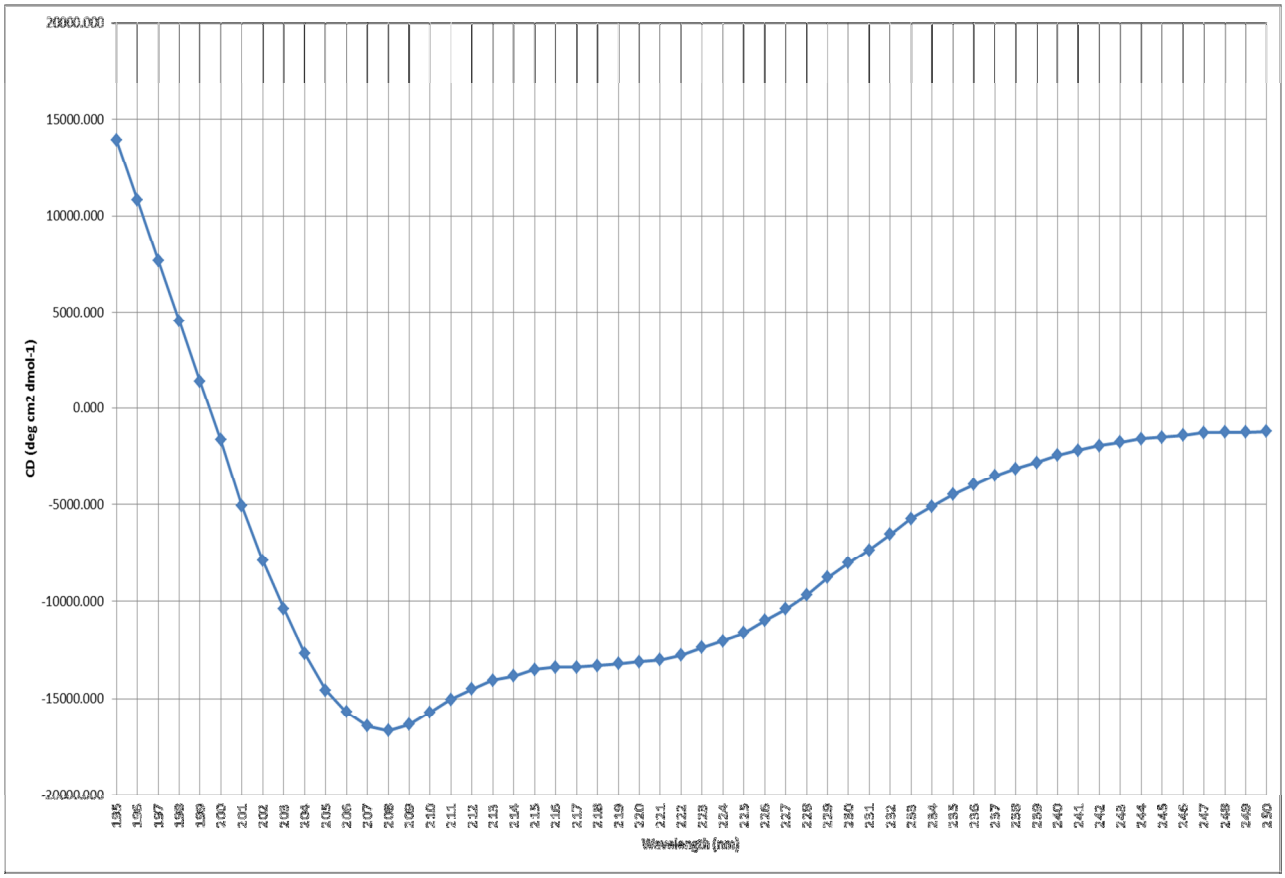


Figure 1.11 Circular dichroism spectrum of human Armet at 25°C

Spectrum showing characteristic α helix signal acquired at 25°C in 50 mM Tris-HCl 100 mM NaCl pH 7.0.

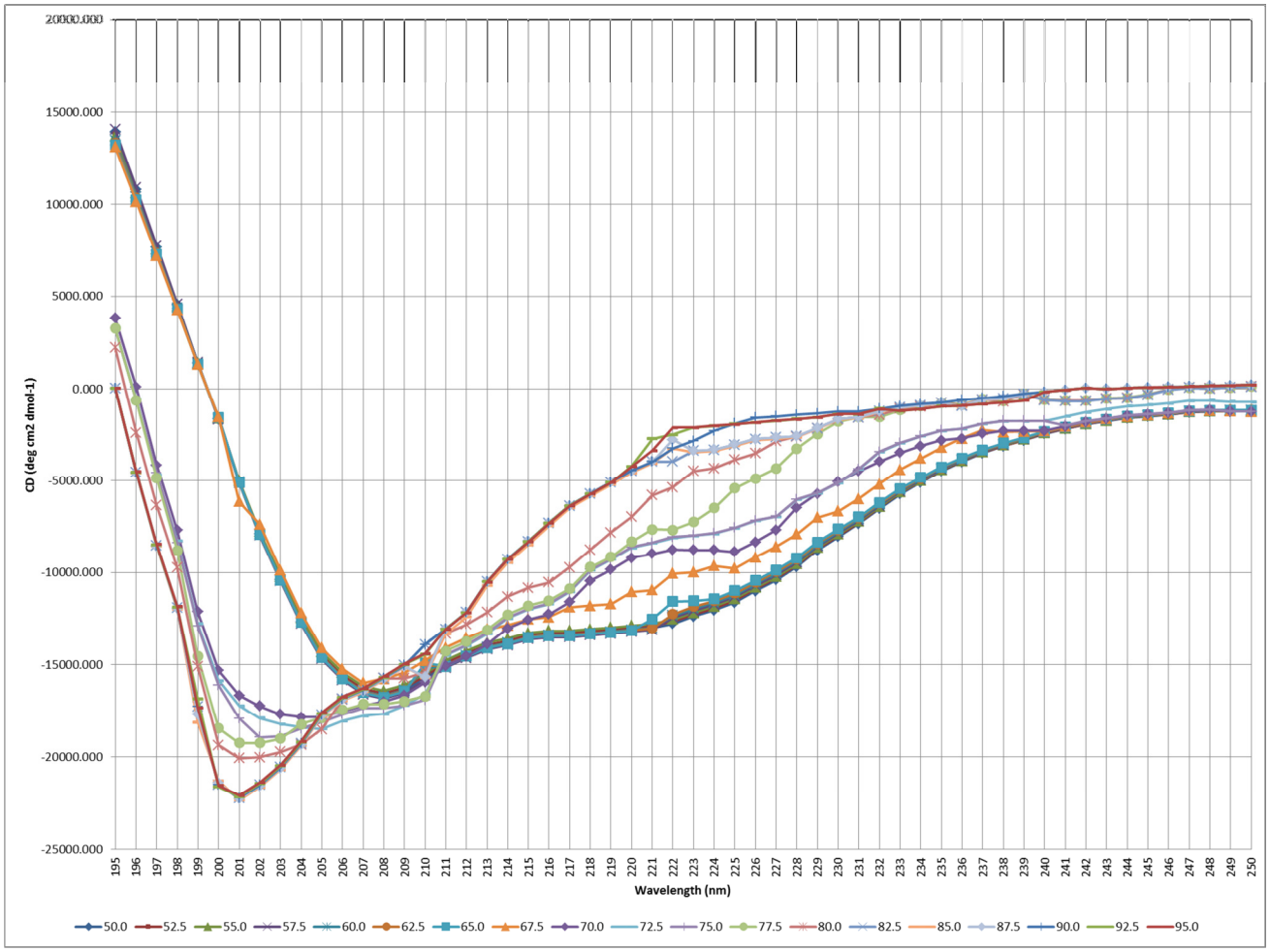


Figure 1.12 Thermal denaturation circular dichroism spectra of human Armet from 50 °C to 95°C

Spectra showing a change from the characteristic α helix content to that of a random coil.

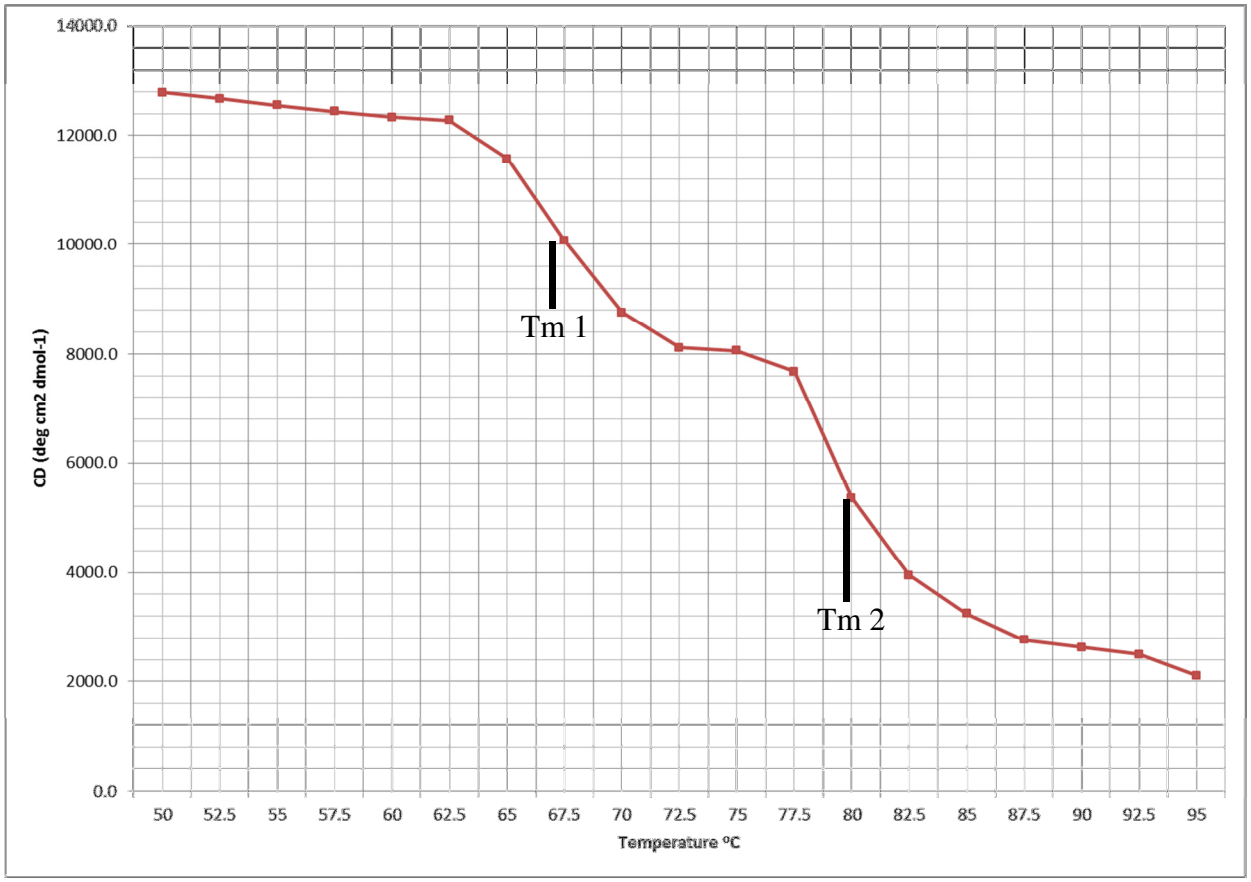


Figure 1.13 Thermal denaturation of human Armet monitored by circular dichroism

These spectra were acquired at 222 nm from 50 °C to 95°C

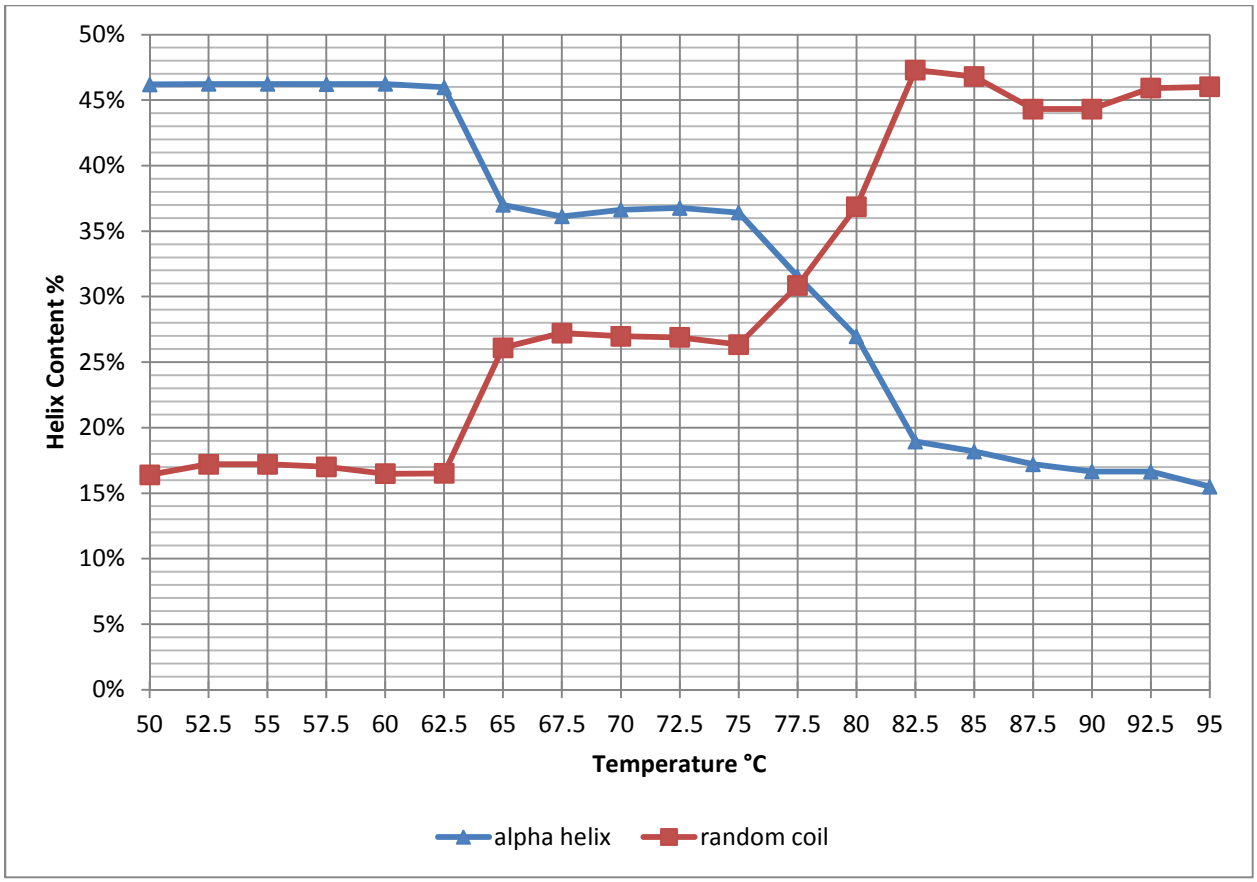


Figure 1.14 Percent helix versus temperature in thermal denaturation of Armet

Data from CDNN software analysis showing percent helix and random coil versus temperature. Not shown, a β turn characteristic that maintains a steady value at approximately 20%.

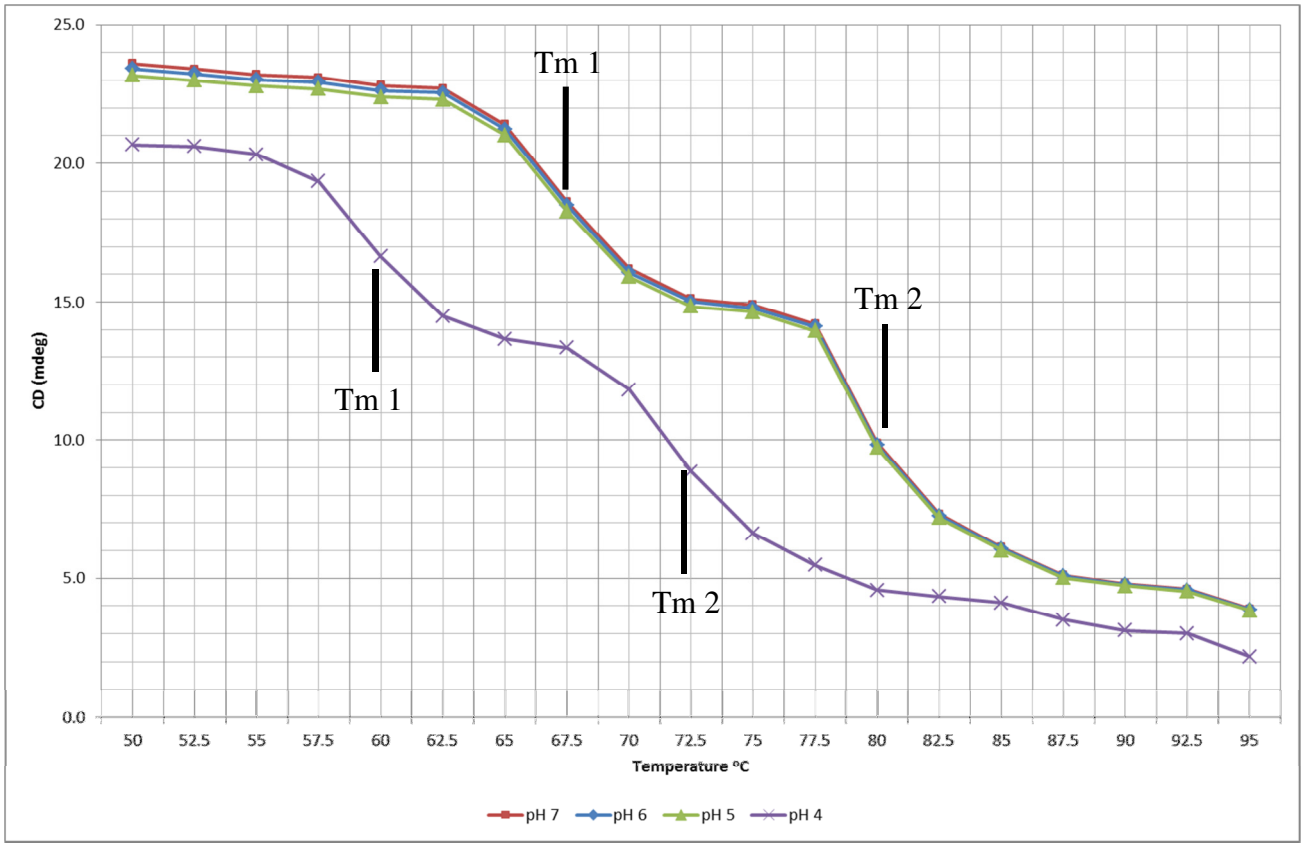


Figure 1.15 Thermal denaturation circular dichroism spectra of human Armet from 50°C to 95°C at varying pH

Thermal denaturation studies with altered pH in 50 mM Tris-HCl 100 mM NaCl.

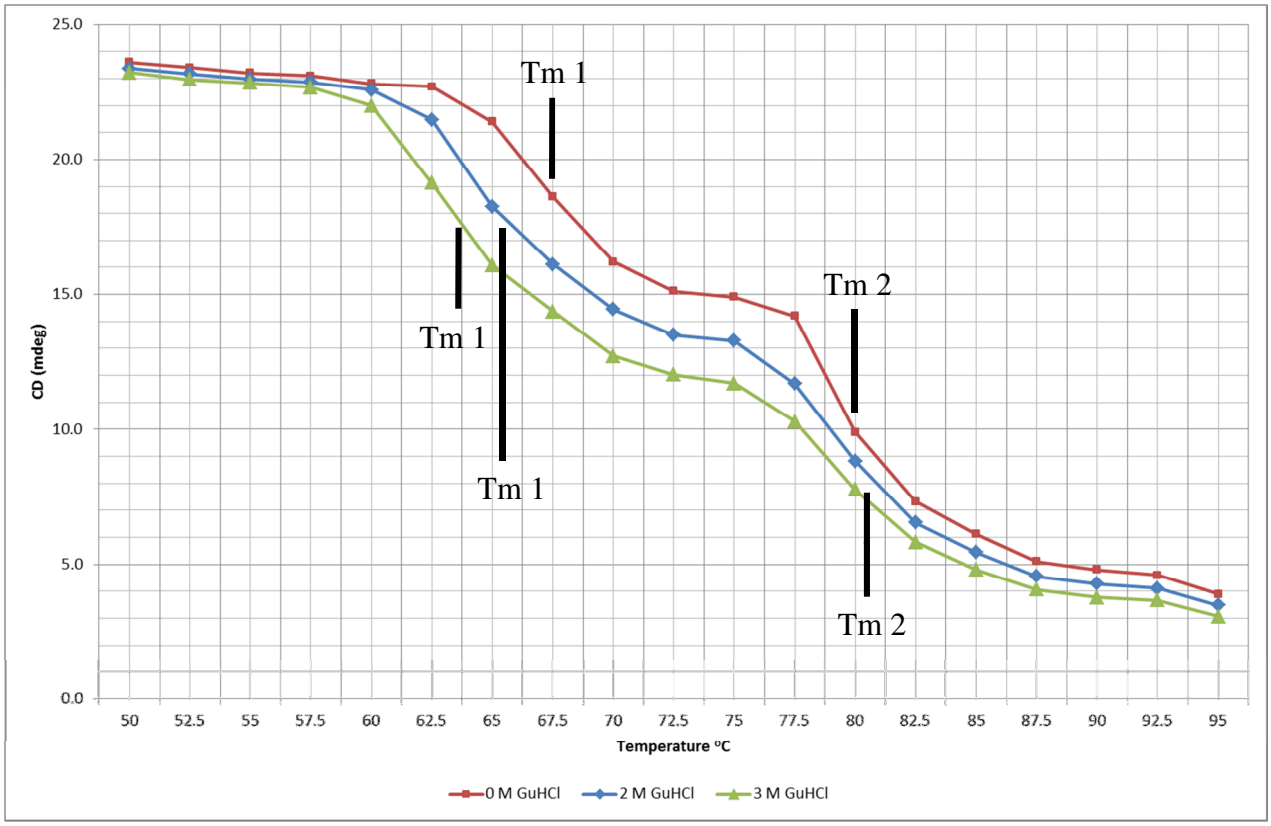


Figure 1.16 Thermal denaturation of Armet in guanidinium hydrochloride

Results obtained in the presence of guanidinium hydrochloride in 50 mM Tris-HCl 100 mM NaCl at pH 7.0

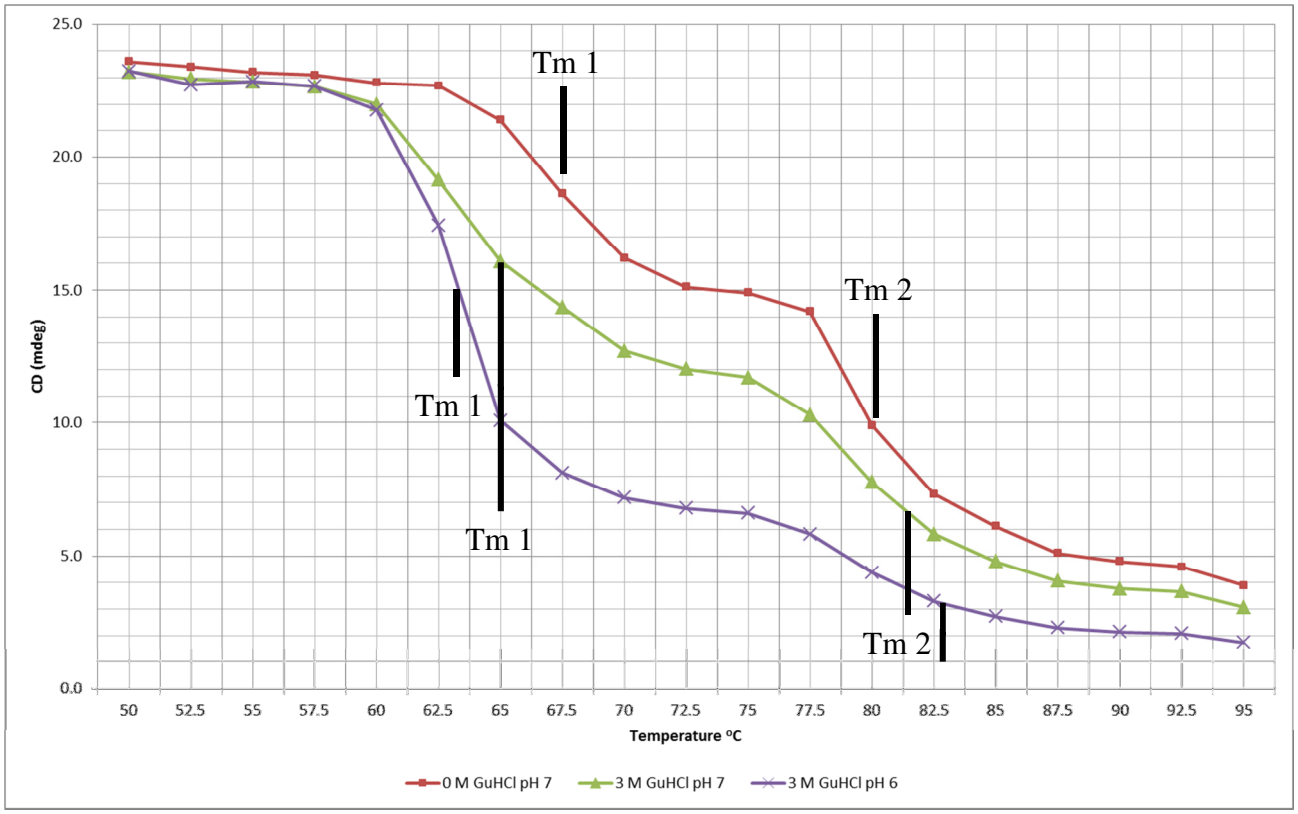


Figure 1.17 Thermal denaturation of Armet in guanidinium hydrochloride with altered pH

Thermal denaturation studies in the presence of guanidinium hydrochloride with altered pH conditions in 50 mM Tris-HCl 100 mM NaCl.

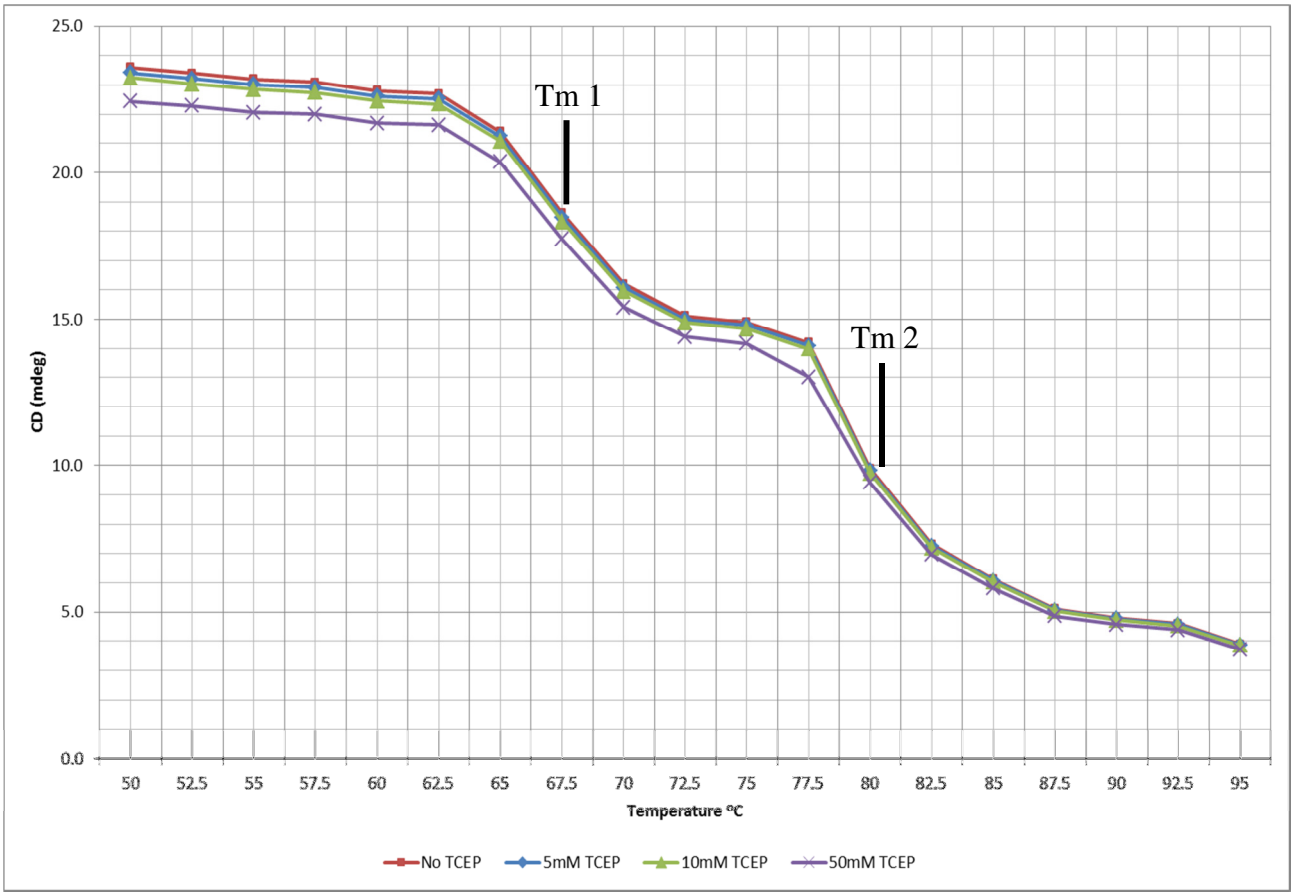


Figure 1.18 Thermal denaturation of Armet in the presence of TCEP

Thermal denaturation of Armet, measured at 222 nm from 50 °C to 95 °C in the presence of TCEP at several concentrations. No significant Tm shift is noted in the presence of the disulfide reducing agent.

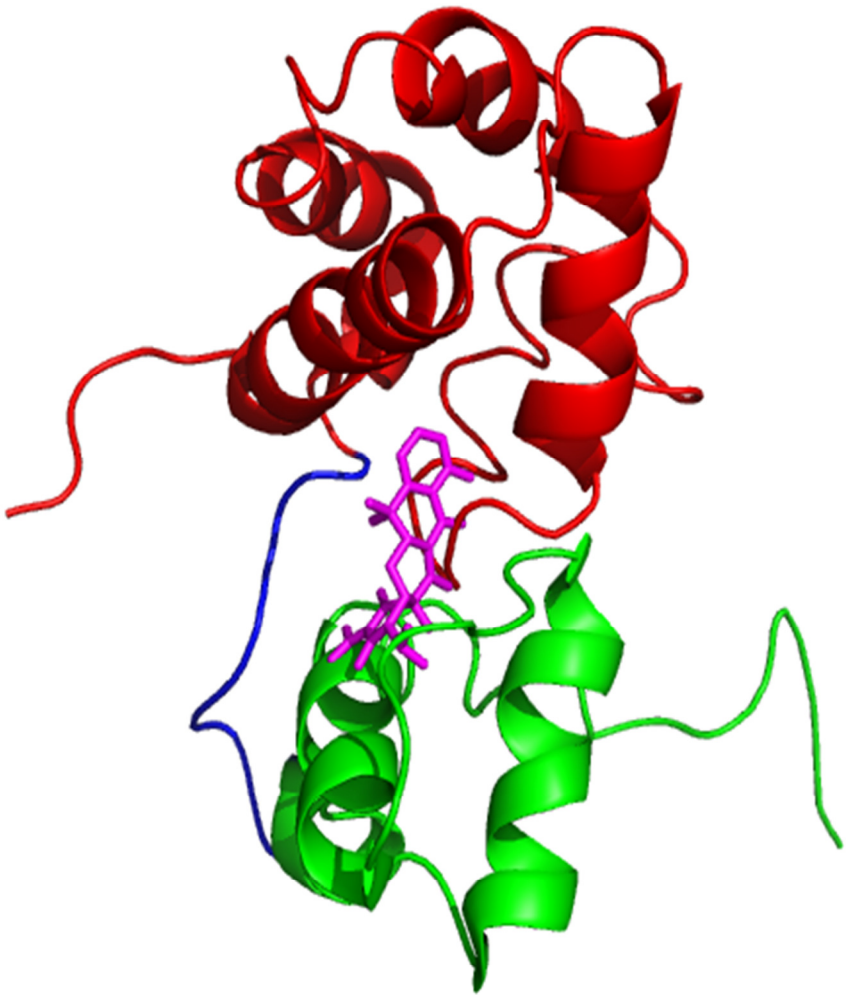


Figure 1.19 Possible mode of binding of tetracycline to Armet

Graphical concept of the N-terminal domain (red), linker (blue), and C-terminal domain (green) coming together in a "clamshell" type fashion around the ligand tetracycline (magenta). Note: based loosely on experimental data, just a theoretical representation

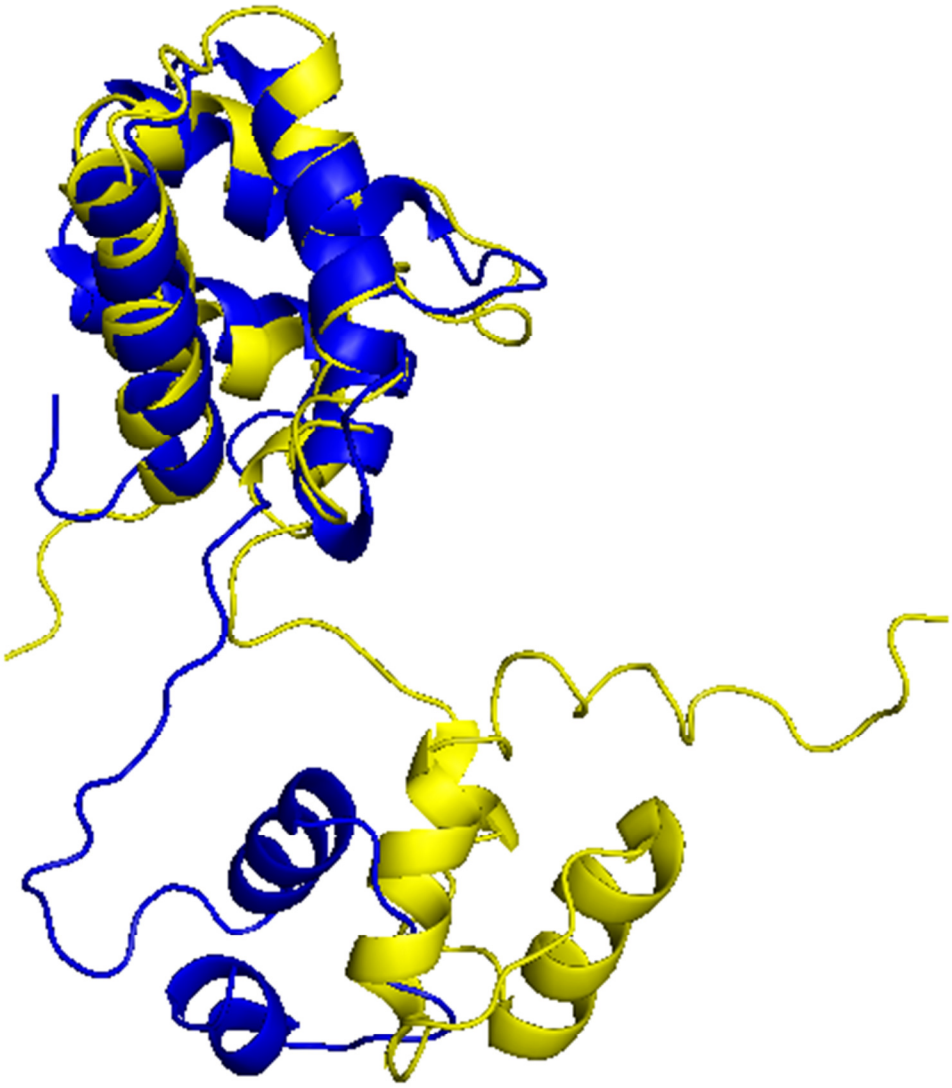


Figure 1.20 Superimposition of crystal and solution NMR structures of human Armet with clipped linker

Human Armet X-ray structure shown in blue, PDB: 2W51 (Parkash et al., 2009)

Human Armet NMR structure shown in yellow, PDB: 2KVD (Hellman et al., 2010)

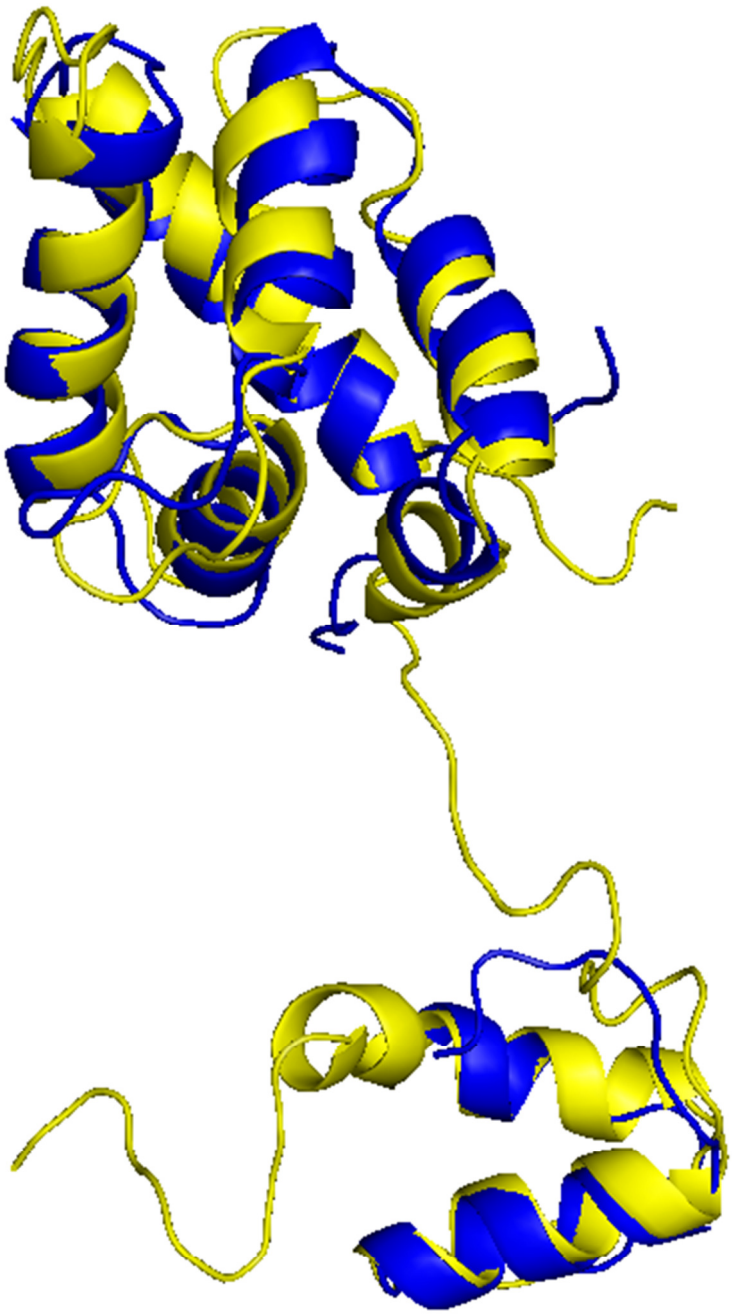
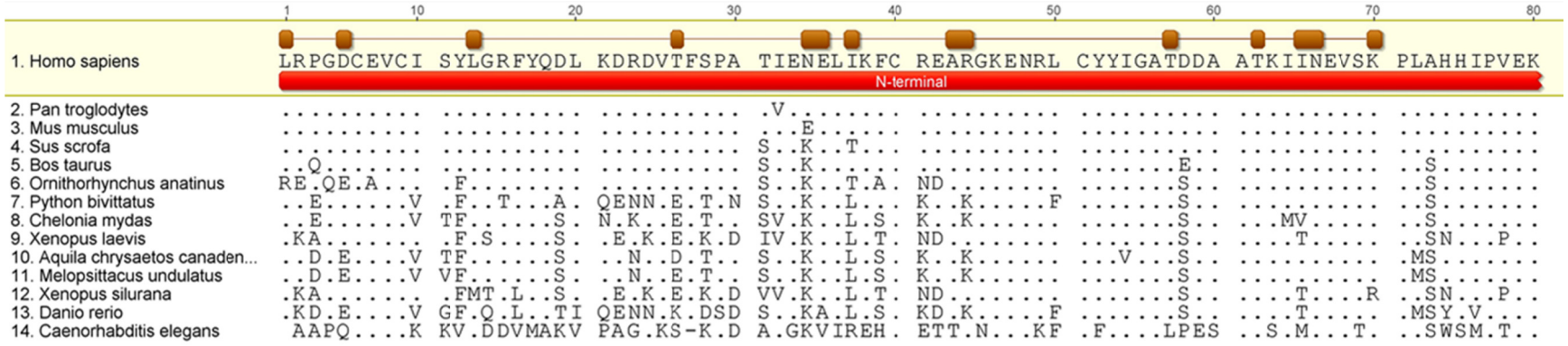
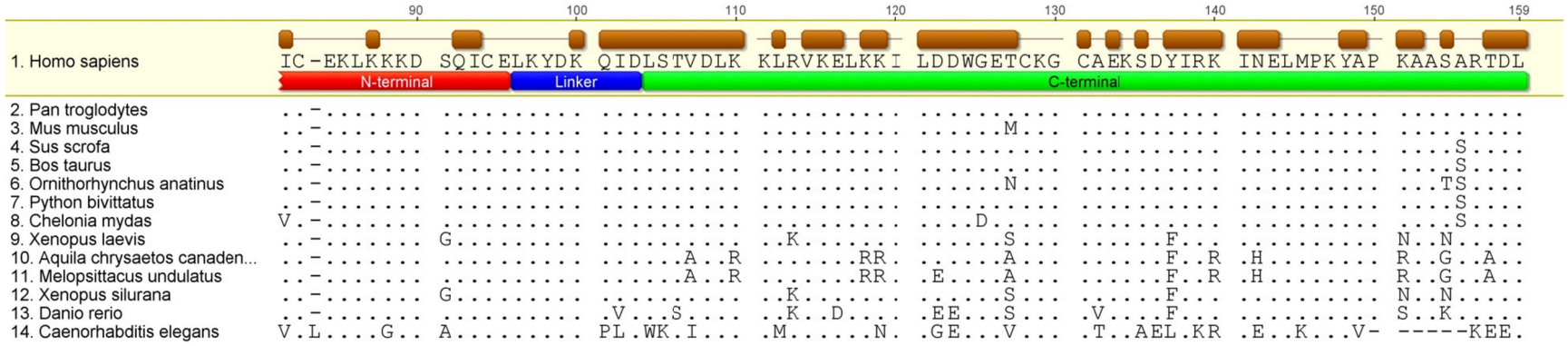


Figure 1.21 Superimposition of crystal and solution NMR structure of human Armet

Human Armet X-ray structure shown in blue, PDB: 2W51 (Parkash et al., 2009)

Human Armet NMR structure shown in yellow, PDB: 2KVD (Hellman et al., 2010)





(green) as
lues are
rs of
binding are

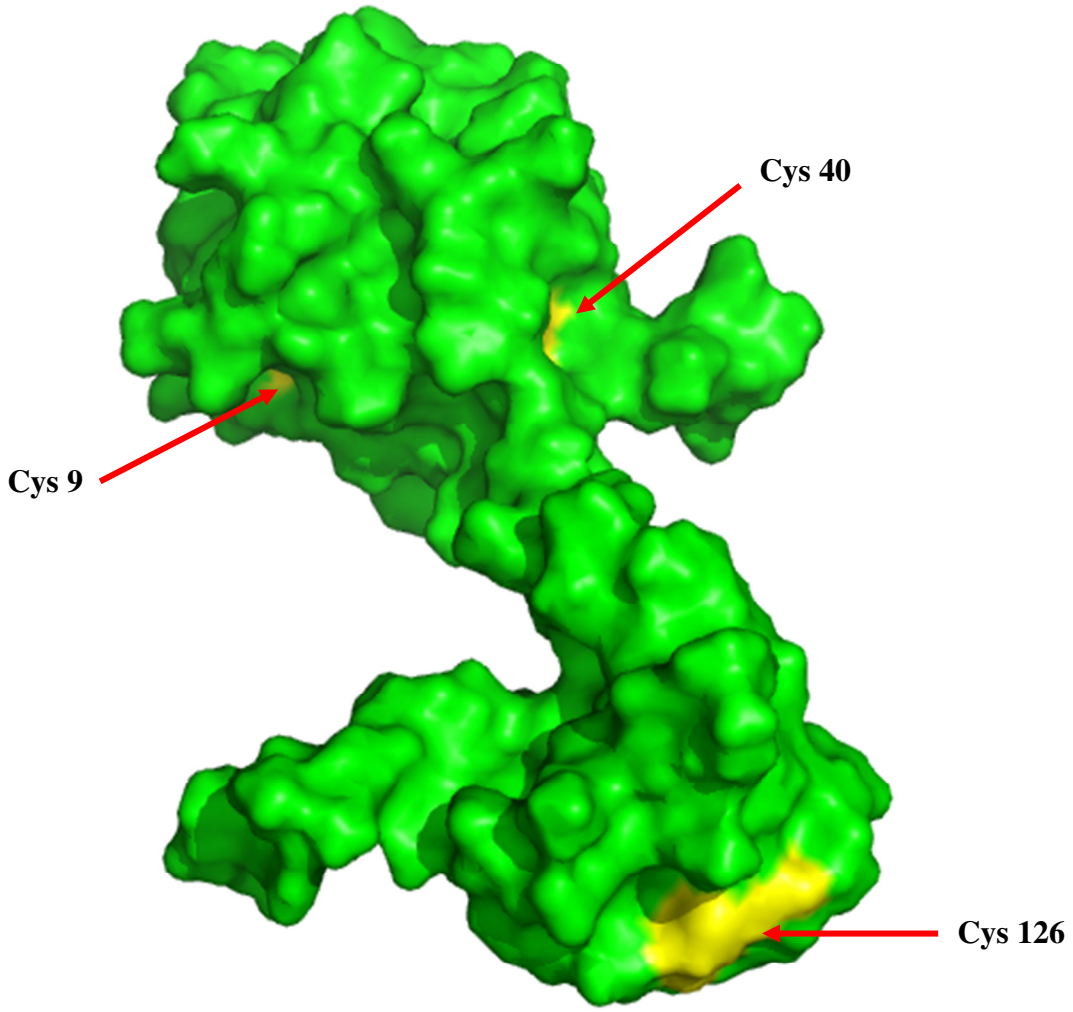


Figure 1.23 Armet C-terminal solvent accessible disulfide

Cysteine residues shown in yellow and identified with red arrows, indicate that the three disulfides within the N-terminal are not solvent accessible whereas the C-terminal disulfide is accessible. Note: one disulfide in the N-terminal domain is completely buried in the structure.

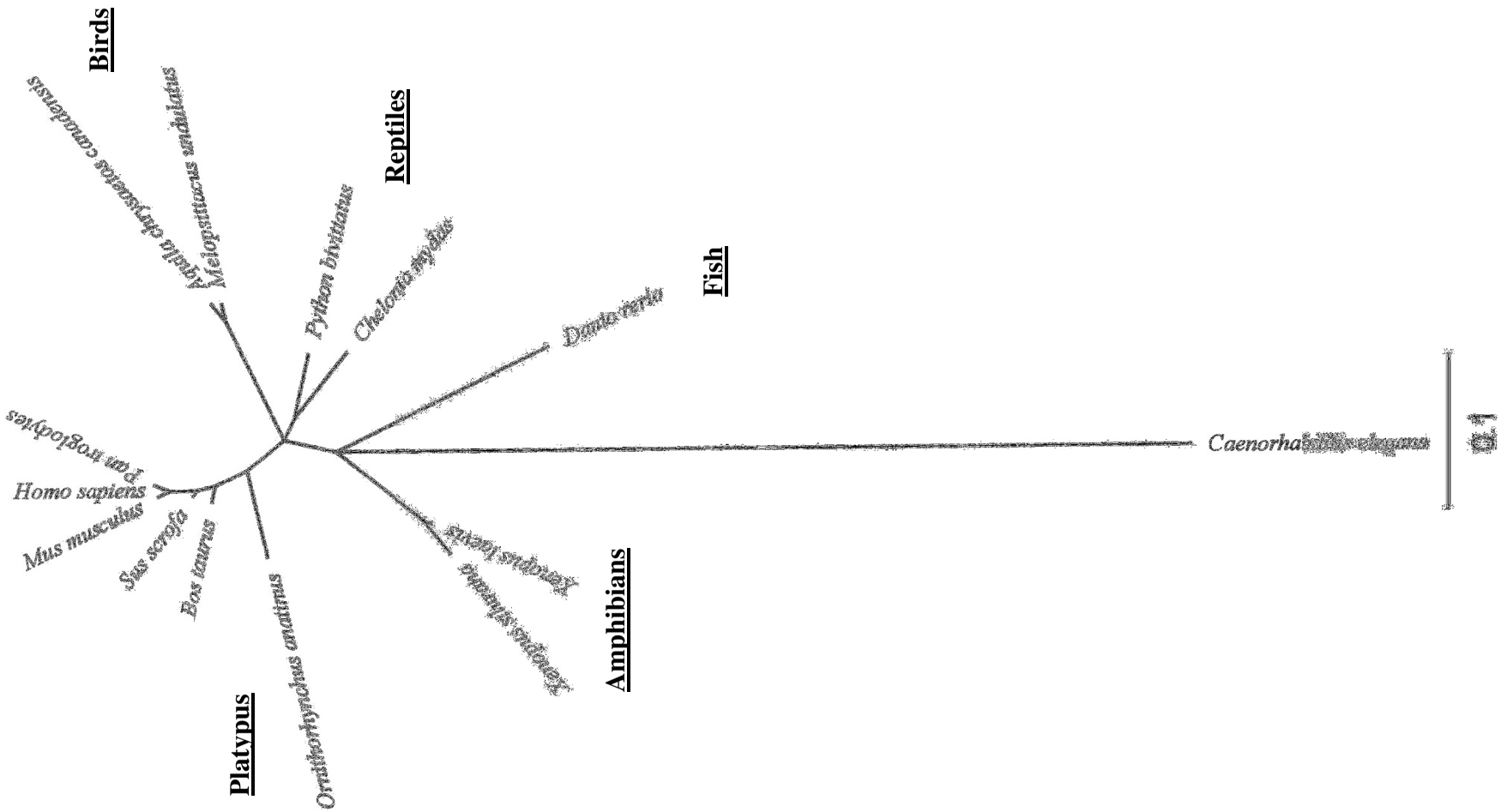


Figure 1.24 Consensus tree

Orthologous Armet proteins from Figure 1.22, shown in a consensus tree generated by the Geneious software suite.

Human (*Homo sapiens)*

Chimpanzee (*Pan troglodytes*)

Mouse (*Mus musculus*)

Pig (*Sus scrofa*)

Cattle (*Bos taurus*)

Platypus (*Ornithorhynchus anatinus*)

Burmese python (*Python bivattatus*)

Green sea turtle (*Chelonia mydas*)

African clawed frog (*Xenopus laevis*)

Golden eagle (*Aquila chrysaetos canadensis*)

Parakeet (*Melopsittacus undulatus*)

Western clawed frog (*Xenopus silurana*)

Zebrafish (*Danio rerio*)

Nematode (*Caenorhabditis elegans*)

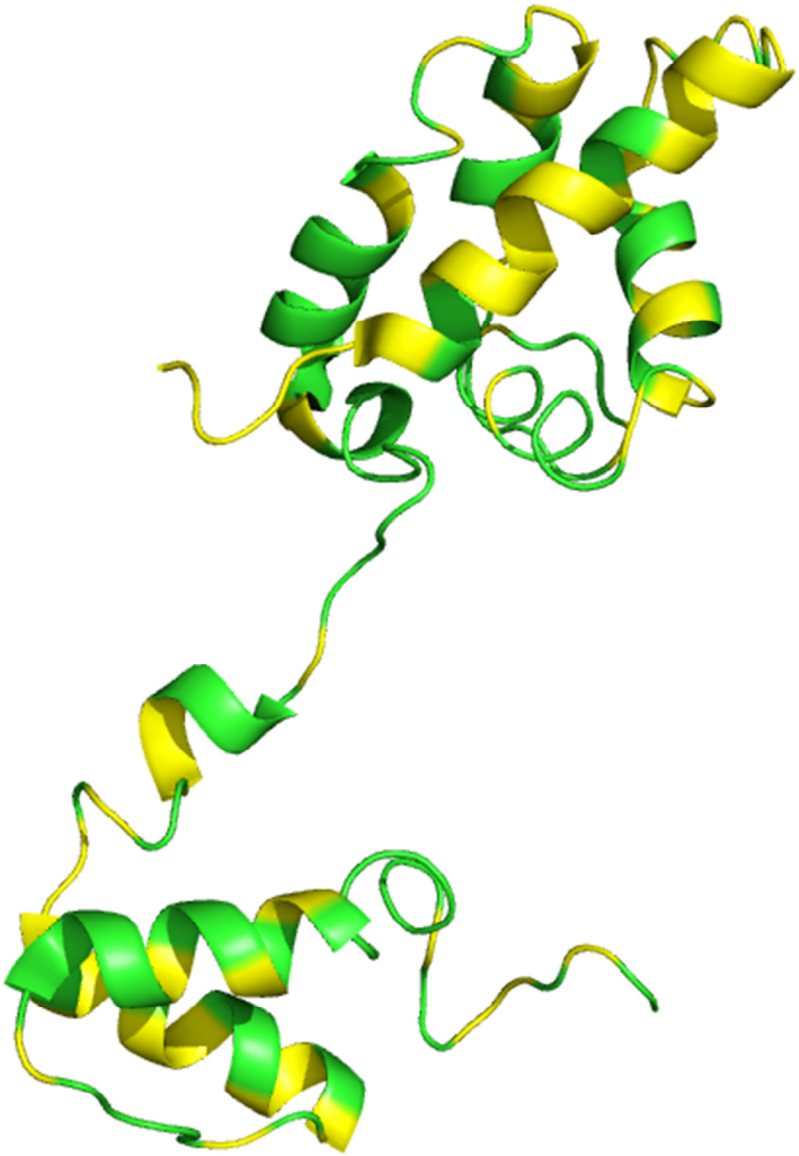


Figure 1.25 Conserved and non-conserved residues mapped onto the structure of human Armet

Graphical representation of conserved residues (green) and non-conserved residues (yellow) in vertebrate Armets.

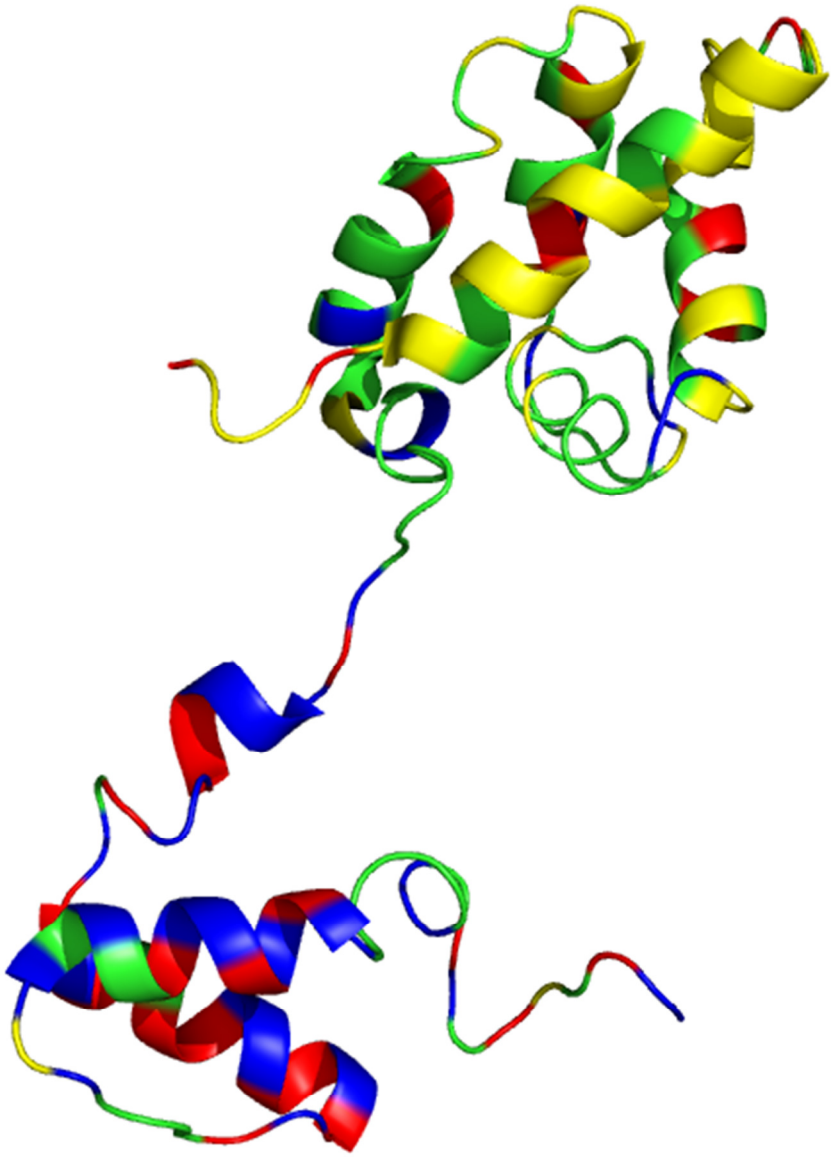


Figure 1.26 Conserved/non-conserved residues with potential ligand binding residues

- 59 conserved residues that do not appear to bind: green
- 32 conserved residues that do potentially bind: blue
- 38 non-conserved residues that do not appear to bind: yellow
- 29 non-conserved residues that do potentially bind: red

Chapter 2 - Transcripts of the Unfolded Protein Response in the Pea Aphid

Literature Review:

Aphids

Aphids are insects of the superfamily Aphidoidea within the order Hemiptera. Within the order Hemiptera, commonly known as true bugs, there are estimates of 50,000-80,000 species, all with a common arrangement of sucking mouthparts. These include, for example, insects such as cicadas, planthoppers, and shield bugs besides aphids (Polhemus et al., 2008). Aphids rely on their saliva to feed from a host plant's phloem sap utilizing a piercing and sucking action. There are over 4,000 species of aphids, and many of them feed on multiple host plants (Jaouannet et al., 2014). In terrestrial plants, there is essentially no part of the plant that is not attacked by an aphid, either above or below ground. Feeding on leaves, roots, and even bark, aphids, such as *Rhopalosiphum maidis* (corn aphid) and *Aphis glycines* (soybean aphid), have evolved into significant agricultural pests (Minks et al., 1989). Aphids commonly studied in laboratory settings include *Myzus persicae* (green peach aphid), *Diuraphis noxia* (Russian wheat aphid) and *Acyrthosiphon pisum* (pea aphid) due to their sequenced genomes.

Aphid species have host specificity ranging from strict monophagy, i.e., the grape phylloxera (*Daktulosphaira vitifoliae*), to polyphagy, i.e., the green peach aphid (*Myzus persicae)*, whose summer generations can develop on an exceedingly wide range of host plant species (Dixon 1987). Host plants for pea aphids include some legumes such as alfalfa, pea, clover, and broad bean plants (Blackman et al., 2000). There are more than 20 legume genera known to host pea aphids, however the entire host plant range is undetermined. The fava bean, *Vicia faba*, is commonly used to maintain laboratory insects.

Phloem sap is fed on by means of high pressure from within the sieve elements using an elongated stylet that pierces the plant tissues, probing to find the phloem sap while injecting saliva (Dinant et al., 2010). When the stylets have reached the phloem flow, the antennae fold back as an indication of feeding (Darcy et al., 2000). A cartoon diagram of a pea aphid feeding on a sieve element with its stylet is seen in Figure 2.1.

Pea aphid: Acyrthosiphon pisum

The pea aphid is significant scientifically because the genome is known (International Aphid Genomics Consortium, 2010) and due to its size. As the largest aphid commonly maintained in laboratories, it is able to be dissected more easily than other aphid species (Polhemus et al., 2008). It can be maintained easily in a laboratory setting. Although proteins that interact with plant defenses are well-characterized for pathogens such as bacteria, oomycetes, and nematodes, the equivalent molecules in aphids and other phloem-feeders are not well characterized but are being studied. For example in "Immunity and other defenses in pea aphids, Acyrthosiphon pisum," Gerardo et al. (2010) researchers have initiated studies on heat shock proteins and proteins of immunity; however no lab has identified members of the UPR.

Female pea aphids lay fertilized eggs in autumn that hatch the following spring. Of the hatched nymphs, they are all females, which undergo four molts before they reach sexual maturity (Simon et al., 2010). At sexual maturity, they begin to reproduce by viviparous parthenogenesis, like most aphids (Simon et al., 2010). Each adult female will give birth to 4-12 female nymphs per day, totaling around a hundred in her life cycle (Simon et al., 2010). These develop into mature females in approximately 7-10 days. The life span of an adult is about 30 days.

Through predation and parasitism, the highest population density of aphids during early summer begins to decrease where the lengthening of night triggers the production of one generation of sexual males and oviparous females. Once inseminated, the females lay eggs in the autumn to restart the aphid life cycle (Simon et al., 2010).

In pea aphids two morphs exist, winged and wingless. The winged morphs may be triggered by overcrowding and poor food quality and it is then that the winged aphids can colonize other host plants (Braendle et al., 2006). When pea aphids feed on crops such as alfalfa, they can act as vectors for plant viruses that may retard growth, reduce yield, and cause death in the host plant. Aphid feeding on crops such as alfalfa can significantly reduce feed value, where black fungus and sooty mold grows on the honeydew excreted by the aphid, which reduces palatability to livestock (Mulder et al., 2013). The pea aphid is considered the most agronomical detrimental aphid on pea and alfalfa crops (van Emden et al., 2007).

Agricultural Threat and Current Pest Management Strategy

According to the Division of Agricultural Sciences and Natural Resources at Oklahoma State University, pea aphids are a significant threat to alfalfa crops, causing toxic effects in the plant and monetary losses estimated at 100 million dollars a year worldwide (Blackman et al., 2000, Mulder et al., 2013). A loss estimate for aphid species throughout the world is approximately 1 billion dollars a year (Blackman et al., 2000). For pea aphids, infestations begin in early March and last through late May in areas that grow alfalfa. States such as Oklahoma, Kansas, Texas, Nebraska, and Iowa are affected by the pea aphid and damage to alfalfa crops range from stunted growth to death when insects are present for several weeks (Mulder et al., 2013).

A genetically modified organism (GMO), namely a wheat crop, to repel aphids has been attempted with failed result (Bruce et al., 2015). The GMO crop failed to repel aphids any more effectively than ordinary crops in a 3 million dollar trial. A wheat crop engineered to emit an odor that deters aphids in the hope of reducing the amount of pesticides required by plants, nicknamed "whiffy wheat," were successful in lab tests, but succumbed to aphids when trialed in the field (Bruce et al., 2015). At a significant cost, the experiment was conducted in the United Kingdom at the Rothamsted Agricultural Institution from 2012 to 2013. Although researchers had hoped to create a strain of wheat capable of deterring aphids from eating the crops and spreading plant viruses and infections, it failed (Bruce et al., 2015).

Current strategies for mitigating pea aphid infestations, according to the 2015 Kansas State University Online Research and Extension Guide, include 12 insecticides and identifies the two most widely used, Chlorpyrifos and Dimate. Harmful pesticides may persist on harvested crops and in the environment, to the detriment of human health and environment. Because current aphid pest strategy is limited to early cutting of crops and population control by pesticides, any new potential pest mitigation solution would be desired.

The Unfolded Protein Response (UPR)

Discovered by Peter Walter in 1994 in yeast, the UPR is an ER response to aggregation of proteins in the secretory pathway (Walter et al., 2011). In eukaryotic cells, secretory and transmembrane proteins are folded and modified in the lumen of the ER (Alberts et al., 2002). Although it is unknown if the UPR is activated during normal protein synthesis, estimates

suggest approximately 30% of newly synthesized proteins are rapidly degraded when cells are not under ER stress, possibly as a result of improper protein folding (Schubert et al., 2000). Thus, an increase in the translation of secretory proteins would inflict a major problem for the cell in the absence of the UPR due to an increased aggregation of misfolded proteins.

Contingent on proper folding, proteins are transported to the Golgi apparatus for secretion. Impairment of proper folding can be caused by various factors, such as chemical compounds, mutations in genes involved in ER quality control and increased secretion of proteins, resulting in the accumulation of unfolded or misfolded proteins in the ER, collectively called ER stress (Balch et al., 2008, Schroder et al. 2005). An estimate of the concentration threshold for apoptosis from ER stress due to unfolded proteins is approximately 100 mg/mL (Naidoo et al., 1999, Stevens et al., 1999). Prior to UPR mediated apoptosis at the estimated 100 mg/mL threshold, an unknown concentration of protein aggregation “triggers” the UPR and activates a complex signal transduction pathway that conveys information about the protein folding status in the ER lumen and then deals with the situation.

The UPR’s major function is to increase protein folding capacity, therefore decreasing unfolded protein load. However, if this major function cannot be achieved, and the cell is unable to re-establish ER homeostasis, the cells undergo death by UPR mediated apoptosis (Kimura et al., 2010). The activation threshold is unknown but if the claims of Schubert et al. (2000) are correct, and 30% of newly synthesized proteins in the ER are degraded due to improper folding, the value must be higher than 30%. The UPR is likely to be functioning constantly, adjusting in response to physiological conditions as suggested in the paper from Matus et al. (2008), "The Stress Rheostat: An interplay Between the Unfolded Protein Response (UPR) and Autophagy in Neurodegeneration." The authors suggest that the unfolded protein response is not activated in an on/off function but continually changes in regards to stress.

I show a schematic of activation of the UPR in Figure 2.2 when the accumulation of unfolded or misfolded proteins in ER have stimulated the UPR pathway. Glucose-regulated protein 78 (GRP78) binds to the exposed hydrophobic surface area of unfolded or misfolded proteins and because it is no longer binding to its ER membrane-binding partners inositol requiring enzyme 1 (IRE1), activating transcription factor 6 (ATF6), and protein kinase RNA-like ER kinase (PERK), it triggers the downstream events (Schroder et al., 2005). The PERK pathway mediates cell cycle arrest and protein translation attenuation slowing the expression of

new non-UPR proteins (Schroder et al., 2005). When IRE1 is activated, it alternative splices xbox binding protein 1 (XBP1) and the spliced form translocates into the nucleus where it activates a set of target genes which increase production of protein disulfide isomerases (PDIs) and chaperones such as heat shock proteins (HSP) (Yoshida et al., 2001, Tigges et al., 2006). An active, Golgi translocated ATF6 is cleaved by proteases into a smaller fragment, called ATF6 p50 (50kDa) (Haze et al., 1999, Ye et al., 2000, Bommiasamy et al., 2009). ATF6 p50, a transcription factor, is translocated to the nucleus, binds to an ERSE element and activates another set of genes such as itself (ATF6) that may or may not overlap with XBP1 target genes (Haze et al., 1999, Ye et al., 2000, Bommiasamy et al., 2009). Acting in a positive feedback loop, the UPR activated gene products include the UPR key regulators themselves such as glucose regulating protein 78 (GRP78), IRE1, ATF6, XBP1 and PERK in addition to proteins that are involved in protein folding, glycosylation, degradation and lipid synthesis, where examples of protein folding components are identified in Table 4 (Schroder et al., 2005). As previously stated, the activated UPR-specific target genes increase protein-folding capacity, however if the UPR cannot rescue the cell from the protein folding mediated ER stress, the cell will undergo apoptosis (Schroder et al., 2005). Of the three pathways, IRE1/XBP1 and ATF6 pathways are UPR-specific, while the PERK pathway is shared with other cellular stress pathways including those involved in amino acid deprivation, infection with double-stranded RNA viruses, and mechanical stress. According to Schroder et al. (2005), the UPR pathway is centralized by the UPR-specific transcriptional events, where detection of the overall UPR-specific transcriptional activation should provide a means to monitor or even quantify levels of UPR activation in cells under ER stress.

Several subsystem processes such as transducers, chaperones, endoplasmic reticulum associated degradation (ERAD), and pro/anti apoptotic functions (Hetz et al., 2013) have been identified in the UPR, where the primary function of ER stress is to induce the expression of ER chaperones, decrease new protein synthesis, and enhance the degradation of proteins accumulated in the ER by way of ERAD within the cytosol.

Unfolded Protein Response Subsystems

Transducers of the UPR

The UPR in mammals and invertebrates is initiated by three ER transmembrane sensors, IRE1, ATF6, and PERK (Sidrauski et al., 1997, Bertolotti et al., 2000, Ron et al., 2007). As seen in Figure 2.2, the three transducers of the UPR modulate downstream responses in an attempt to adapt to and avoid chronic ER stress and ultimately apoptosis. In yeast, IRE1 is the only transducer (Sidrauski et al., 1997, Ron et al., 2007).

Protein Folding & Chaperones of the UPR

The most widely studied components of the UPR are chaperones including glucose regulating proteins, protein disulfide isomerases, calnexin, and calreticulin.

Glucose regulating proteins (GRPs), namely GRP78 (also known as BiP and a HSP70), GRP94 (also known as HSP90B1), GRP170, and GRP75 are stress-inducible molecular chaperones that belong to heat shock protein (HSP) families (Lee 2014). GRPs are found in the ER and regulate protein quality control and metabolic balance (Lee 2014). As chaperones, these GRPs facilitate protein folding and assembly, as well as the export of misfolded proteins for degradation. A well-studied component with chaperone activity, known by multiple names including GRP78 (glucose regulated protein 78), BiP (binding immunoglobulin protein), and HSPA5 (heat shock 70 kDa protein 5) is located in the lumen of the ER (Ting et al., 1988, Hendershot et al., 1994). GRP78 has been shown to bind to newly synthesized proteins as they are translocated into the ER and assists them in subsequent folding. GRP78 is an essential component in translocation across the ER membrane of proteins destined for degradation by the proteasome (Delom et al., 2001). GRP78 is an abundant protein under all growth conditions but it is upregulated under ER stress (Delom et al., 2001). GRP78 can interact with another protein of the UPR, Armet, and can retain Armet in the ER lumen. It can also interact with the transducer IRE1 when the UPR is not yet initiated (Ryoo et al., 2007). Other chaperones that can bind unfolded proteins include GRP94 and GRP170, with the numbers associated corresponding to their molecular masses (Lee 2001).

Both calnexin and calreticulin are ER associated modulators of calcium transport (Camacho et al., 2003), and they aid in maintaining the integrity and homeostasis of the ER under ER stress. Although the molecular mechanisms underlying ER stress-induced apoptosis are not completely understood, evidence suggests that ER and mitochondria cooperate to signal cell death. Calnexin and calreticulin aid in regulation of calcium transfer from the ER to the

mitochondria. This regulation is key in maintaining control of pro-survival or pro-death pathways (Malhotra et al., 2013).

Calnexin, a molecular chaperone, aids in the translocation of nascent polypeptides and in the folding and quality control of newly synthesized proteins (Bukau et al., 2000, Fewell et al., 2001, Williams, 2006). Structurally, calnexin is an ER transmembrane protein, with a large luminal domain, and a short cytosolic tail.

Calreticulin is also a molecular chaperone located in the lumen of the ER and plays an important role in the folding of newly synthesized proteins in the ER lumen (Saito et al., 1999). Shown to be an important component in development, calreticulin deficient mice have embryonic lethality, and if the mice survive, possess heart defects (Mesaeli et al., 1999). Calreticulin is also associated with several cancer disease states; its expression is either up or down regulated in particular cancers such as metastatic melanoma, squamous cell carcinoma, and colon cancer (Dissemond et al., 2004, Ogino et al., 2003, Brunagel et al., 2003).

The recognition of misfolded or mutated proteins depends on the detection of substructures within proteins such as exposed hydrophobic regions or free cysteines (unpaired disulfide bonds) in the form of cysteine residues and immature glycans (Williams et al., 2006). In glycan processing, for example, the lectin-type chaperones calnexin/calreticulin provide immature glycoproteins the ability to reach their native conformation (Williams et al., 2006). The UPR has 21 mammalian protein disulfide isomerases, encoded by different genes, including components such as protein disulfide isomerase family member A3 (PDIA3) and CCAAT-enhancer-binding protein homologous protein (CHOP also known as CEBP) which attempt to allow recovery of disulfide bonds to achieve the correct tertiary structure. The family of protein disulfide isomerases are important in maintaining function and structure in secreted proteins (Wilkinson et al., 2004). The UPR strives to return unfolded proteins back to their native state, and to eliminate the aggregation of misfolded proteins. The family of protein disulfide isomerases, as ER enzymes, catalyzes the formation and breakage of disulfide bonds between cysteine residues within proteins as they fold (Gruber et al., 2006). This allows proteins to refold until the correct arrangement of disulfide bonds occurs, forming the lowest energy minima in their fully folded state. Thus by catalyzing disulfides, protein disulfide isomerases (PDIs) can be considered to be chaperones. PDI functions together with ER oxidoreductase (Ero1) by using the oxidizing power of molecular oxygen to create de novo

disulfide bonds in a folding protein (Hatahet et al., 2007). An exchange of disulfide bonds takes place from Ero1 to PDI to the target folding protein, in conjunction with an electron flow in a reverse direction, from target protein to PDI to Ero1.

This process of protein folding is vital to mitigating ER stress returning to homeostasis within the ER. The return of a target protein's functional shape or conformation alleviates stress. Failure to fold back into native structure generally produces inactive proteins, and an over accumulation of these proteins trigger UPR mediated apoptosis (Schroder et al., 2005).

Pro-apoptotic and Anti-apoptotic Signal Induction from the UPR

In contrast to necrosis, which is a form of traumatic cell death that results from acute cellular injury, apoptosis generally confers advantages during an organism's life cycle except in UPR mediated cellular death (Alberts et al., 2008). For example, the differentiation of fingers and toes in a developing human embryo occurs because cells between the fingers apoptose; resulting in digit separation. Unlike necrosis, apoptosis produces cell fragments called apoptotic bodies that phagocytic cells are able to engulf, and quickly remove before the contents of the cell can spill out onto surrounding cells and cause damage (Alberts et al., 2008). While many components are anti-apoptotic, some of the UPR components such as ER nucleus signaling 2 (Ern2), mitogen activated protein kinase 8 (MAPK8), and mitogen activated protein kinase 9 (MAPK9) are involved in inducing apoptosis when the UPR is overwhelmed (Szegezdi et al., 2006, Oslowski et al., 2015).

The UPR is a double edged sword when it comes to apoptosis. When the estimated 100 mg/mL protein aggregation threshold is reached, the UPR can no longer maintain homeostasis, and the fate of the cell shifts towards apoptosis (Tsang et al., 2010). In other words, the UPR is a measurement and response tool for the homeostasis of protein aggregation.

Between 50 and 70 billion cells die each day due to apoptosis in the average human adult although it is unknown how many die due to overwhelming the UPR (Karam et al., 2010). It has also been shown that defective apoptotic processes have been implicated in an extensive variety of diseases. Hyperactive apoptosis can cause atrophy, whereas decreased rates result in uncontrolled cell proliferation, such as cancer (Karam et al., 2010).

ERAD Mediated by Ubiquitination

By biochemically studying fractionated yeast cells, Brodsky and McCracken coined the now widely used term ERAD, while establishing the first *in vitro* system to study ERAD, (McCracken et al., 1996, Werner et al., 1996, Brodsky et al., 1999).

Ubiquitin is a small regulatory protein that has been found in almost all tissues of eukaryotic organisms, not a UPR member itself; however it is utilized by components of the UPR. Ubiquitin directs proteins to cellular compartments, including the proteasome used in ERAD where proteins are recycled and destroyed within the cytosol. Seen in Figure 2.2, the downstream activation of ERAD with IRE1/XBP1 pathway is in concert with the activation of chaperones. Ubiquitin consists of 76 amino acid residues with a C-terminal tail containing seven lysine residues. It is highly conserved among eukaryotic species, from human to yeast, with 96% sequence identity (Kimura et al., 2010). Ubiquination is an enzymatic, protein post-translational modification process in which the carboxylic acid of the terminal glycine from the di-glycine motif in the activated ubiquitin forms an amide bond to the epsilon amine of the lysine in the modified protein (Amerik et al., 2000). The ubiquination of terminally misfolded proteins, caused by a cascade of enzymatic reactions, marks the protein for ERAD. Following successive addition of ubiquitin molecules to lysine residues of the previously attached ubiquitin, a polyubiquitin chain is formed. After the polyubiquinated protein is produced it is recognized by specific subunits in the 19S capping complexes of the 26S proteasome (Aravind et al., 1998). The protein and attached chain are fed into the central chamber of the 20S core region at the proteolytically active site. Ubiquitin is released for reuse by deubiquinating enzymes before degradation of the protein. However, the proteasomal degradation takes place in the cytoplasm. The ER membrane anchored RING finger containing ubiquitin ligases Hrd1, a UPR component, and non-component Doa10 are the major mediators of substrate ubiquination during ERAD (Vembar et al., 2008). The tail anchored membrane protein Ubc6 as well as Ubc1 and the Cue1 dependent membrane bound Ubc7 are the ubiquitin conjugating enzymes involved in ERAD (Vembar et al., 2008).

Transcription & Translation Factors

To increase the protein folding capacity of the ER, UPR transcription factors such as XBP1, cAMP responsive element binding protein 3 (CREB3), and others, enhance the expression of genes encoding ER-resident chaperones and foldases and promote ER expansion (Bommiasamy et al. 2009, Sriburi et al., 2004).

Translation factors such as eukaryotic translation initiation factor (EIF2A) also play a role in reduction of protein synthesis under ER stress. Protein synthesis is inhibited through PERK-induced phosphorylation of eIF2, a translation factor that when modified, leads to a loss of translation initiation complexes (DuRose et al., 2009, Harding et al., 2000).

Insect UPR Components, Including Armet

Aphids deliver proteins in their saliva to host plants during feeding, and Armet is one of the components in Aphid salivary gland containing a signal secretion peptide (Carolan et al., 2009). Aphid Armet is reported as having intracellular and extracellular roles as in its mammalian counterpart although not specifically in saliva. Wang et al. (2015) characterized the aphid protein, demonstrated that its promoter is responsive to ER stress, and that its extracellular role is as a secreted effector protein that facilitates successful aphid feeding on host plants. By interfering with the expression of Armet, Wang et al. (2015) undermined the compatible interaction between aphids and their host plants.

Although Wang et al. (2015) identified the secretive function of Armet during feeding, they stress that the neurotrophic role of Armet is presumed present as well. In gene knockouts in another insect, namely *D. melanogaster,* it was shown that Armet was essential for development and was lethal to larvae, which may be attributed to the neurotrophic role of Armet (Palgi et al., 2009).

It has been shown that dsRNA injection into the hemolymph of the pea aphid is effective in transcript knockdown of protein C002 (Mutti et al., 2006). The knockdown studies in protein C002 and Armet indicate that the RNAi effect is present in aphids and give traction to possible studies involving dsRNA injections targeted at a UPR component's mRNA.

Working on *D. melanogaster,* UPR has centered on the three transducers. Investigations have shown that one transducer and its cofactor, IRE1 and XBP1, respectively, are essential genes during development of fly cell lines and *in vivo* (Ryoo et al., 2007). These studies while relevant used the ER stress inducer tunicamycin. Other studies have measured survival time also in the presence of the inducer tunicamycin, identifying natural variation among 114 drosophila lines that lend insights into the polymorphisms attributed to have putative roles in ER stress (Chow et al., 2013). One point to be made is that these studies use a compound to measure ER stress.

UPR in C. elegans

Transcriptional profiling in *C. elegans* revealed two aspects of the UPR (Shen et al., 2005). The inducible UPR pathway (i-UPR), directs cells to respond to acute environmental stress, whereas the constitutive UPR pathway (c-UPR) is an essential component for normal development (Shen et al., 2005). Components such as PDI and PDI-2 are members of the i-UPR, where IRE-1 and PERK, are members of the c-UPR (Shen et al., 2005). Researchers concluded that in the i-UPR pathway, IRE-1 and its cofactor XBP1 "act in a linear process that dominates transcriptional regulation to reshape the secretory pathway and adjust cellular functions involved in calcium and phospholipid homeostasis, cell proliferation and death, anti-oxidative stress, metabolism, energy generation, cytoskeletal structure, and mitochondrial function" (Shen et al., 2005). The researchers suggest that work in *C. elegans* might provide a missing link between the yeast and the mammalian UPRs.

A strain of mutant *C. elegans*, SJ17, has been identified as having a flaw in the UPR, characterized by expression of hsp-4 gene. Researchers show that the mutants are incapable of inducing hsp-4 when stressed with tunicamycin, a common UPR inducer (Glover-Cutter et al., 2013). The mutants, when treated with DTT and tunicamycin, exhibit slow growth and do not progress beyond the L2 larval stage. The observed deficiency was attributed to the xbp-1, a transcription factor and UPR member (Glover-Cutter et al., 2013).

Cholesterol Regulation

Cholesterol regulation is important in the UPR in mammals. It has been shown that the change of cholesterol and lipid perturbation in biological membranes can influence and indeed activate the UPR (Xie et al., 2006, Volmer et al., 2013). Insects do not synthesize cholesterol and instead obtain it from their diet, where they may modify it, possibly with the components MBTPS1, INSIG1, and SREBF1. In humans, MBTPS1 is a serine protease that activates SREBF1 function. INSIG1 functions in humans by mediating feedback control of cholesterol synthesis and has been shown to block SREBF1 function. SREBF1 functions in humans as a transcriptional activator required for lipid homeostasis where it binds to the sterol regulatory element and effects cholesterol synthesis. Descriptions of these components were found at www.uniprot.org and are based solely on human function. In aphids, these components are may be involved with the modification of cholesterol obtained from diet.

Materials and Methods:

Search for Pea Aphid Putative Orthologs of Human UPR Proteins

The human UPR components studied in this chapter were found by literature review of the following papers: Bertolotti et al. (2000), Ron et al. (2007), Chakrabarti et al. (2011), Hetz et al. (2011), Oslowski et al. (2011), Kuny et al. (2012).

Utilizing BLASTn, human UPR transcripts were used as query sequences to find the putative orthologous transcripts in the pea aphid. This search encompassed the 91 human UPR genes in Table 4 which are color coordinated by function. "Hits" from this search made up my pea aphid UPR list. Several gene duplications in human UPR components corresponded to single genes in the pea aphid. With duplications removed, the final component list for my studies totaled 74 components.

Dissections

120 diet-fed insects and 120 plant-fed insects were dissected as follows by Dr. Chandrasekar Raman in the Reeck lab. Heads dissected were from wingless, asexual pea aphids from the clone LSR1 line. The location of the salivary glands in an aphid head is represented in the cartoon in Figure 2.1. Prior to dissection, the lab bench, dissection slides, and gloved hands of the researcher were cleaned with a solution of 0.1% DEPC treated water, followed by application of RNaseZap (Sigma-Aldrich #R2020). After all surfaces were allowed to dry, 5-10 uL of RNALater was placed on top of a dissection slide. The dissections then followed the method of removing an aphid from its feeding state, placing it into the small amount of RNAlater on the surface of a microscope slide and immediately starting the dissection. Each subsequent aphid was dissected each time in a fresh droplet of RNAlater. The dissection was achieved by using a bent 22 ga needle. After grasping the aphid in forceps, antennae and eyes were removed followed by the decapitation from the exoskeleton and placement of the head into an RNase/DNase-free collection micro centrifuge tube containing 50 µL of RNAlater. Dissection with the head removal method averaged 1 h for 120 insects.

Diet Feeding

Pea aphids were collected in sterile Petri dishes directly from faba bean plants, *Vicia faba*. These aphids were reared on Akey-Beck diet Table 5 (Akey and Beck, 1971, 1972) for 48

h. The feeding apparatus consisted of a 1 oz container (Dart#100PC) with a thinly stretched piece of Parafilm over the opening. Diet was spread on top of the Parafilm and another Parafilm piece was stretched over the top of the diet enclosing it. After approximately 70 aphids were inside the container was inverted over a yellow piece of paper to attract the aphids to the diet to feed. The feeding ensued by piercing the first layer of Parafilm with their stylets and sucking the diet from between the two layers of Parafilm. A 48 h feeding was conducted, aphids were removed and dissections ensued as detailed above.

RNA Isolation for RNA-seq

Following dissection, immediate RNA isolation was conducted. Prior to any RNA isolation, surface sterilization of all instruments and lab bench tops was performed with RNaseZap.

To the dissected heads, 100 µL of QIAzol reagent (Qiagen #79306) was added to 120 aphid salivary glands in 50 µL of RNAlater in a DNase/RNase free 1.5 mL Eppendorf microcentrifuge tube. A rotating pestle (USA Scientific, 1.5 mL microcentrifuge tube pestle) was used to homogenize the glands with a battery powered rotating tissue homogenizer (Argos Technology, Pestle Motor Mixer) for 2-3 min until no tissue remained intact. Following an addition of 900 µL of QIAzol, and the samples were allowed to stand for 3 min at room temperature. After 3 min, 1 µL of gDNA Eliminator from an RNeasy Kit (Qiagen #74104) was added to reduce genomic DNA contamination from the aqueous phase during phase separation. The sample was then subjected to vortexing several times with 200 µL chloroform to ensure even distribution of reagents in the sample. The sample was stored at room temperature for 10 min, followed by centrifugation at 12000 x g for 15 min at 4°C.

Two distinct layers within the sample formed after centrifugation; the top layer was a clear, aqueous layer, with a pink, organic layer on the bottom. The clear layer was transferred into a new, RNase/DNase free 1.5 mL Eppendorf micro-centrifuge tube, and the organic layer was discarded.

Following the sample transfer into the new centrifuge tube, 500 µL of chilled isopropanol was added and was allowed to stand for 10 min at room temperature facilitating the precipitation of RNA. Centrifugation was then performed at 12000 x g for 15 min at 4°C, which formed a pellet of RNA. After removal of the liquid, the RNA pellet was washed twice with 500 µL of

chilled ethanol. After washing, a 15 min air drying at room temperature was utilized to evaporate excess ethanol in the sample.

The dried RNA pellet was subsequently dissolved in 30 µL of RNase-Free water. (Qiagen #129112) Of that sample, 3 µL was removed for Bioanalyzer analysis to determine if the RNA quality was suitable for RNA-seq by the Illumina Mi-Seq platform. The reported RNA integrity number for the head RNA isolation for plant-fed and diet-fed states respectively were 5.5 and 5.9. Although a good typical RNA integrity number is higher than 7 on a scale of 1 to 10 with other eukaryotes, insect preparations do not follow that standard and are more closely evaluated by 18s rRNA and 28s rRNA peaks. For insects, the electropherogram as shown in Figures 2.3 and 2.4, the area under the 28s rRNA peak is about two times smaller than 18srRNA indicating good quality. The entire remaining sample was utilized in the generation of the cDNA library.

RNA Isolation from Heads for RNA-seq

As with the salivary gland isolation, RNA was isolated in a similar fashion. After aphids were dissected by removing the antennae and eyes prior to decapitation and placement into RNAlater an additional homogenizing step was used increasing the time to 5 min prior to adding the final volume of QIAzol.

Thermal Cycler Programs for cDNA Library Preparation

The synthesis of the cDNA library conducted at The Integrated Genomics Facility at Kansas State University followed the programs listed below in a thermal cycler:

mRNA denaturation: 65°C for 5 min; hold at 4°C

mRNA elution 1: 80°C for 2 min; hold at 25°C

Elution 2-Frag-Prime: 94°C for 8 min; hold at 4°C

1st strand: 25°C for 10 min; 42°C for 50 min; 70°C for 15 min; hold at 4°C

2nd strand: 16°C for 1h; hold at 16°C

End repair: 30°C for 30 min; hold at 4°C

ATAIL70: 37°C for 30 min; 70°C for 5 min; hold at 4°C

Ligation: 30°C for 10 min

PCR: 98°C for 30 s; (15 cycles of) 98°C for 10 s, 60°C for 30 s, 72°C for 30 s, 72°C for 5 min; hold at 10°C

Purification and Fragmentation of mRNA

The Kansas State University Integrated Genomics Facility generated cDNA libraries for RNA-seq analysis. To synthesize the cDNA library, a TruSeq RNA Sample Preparation Kit (Illumina #RS-122-2001) was used. For each feeding type, plant-fed and diet-fed, the total RNA isolated respectively was added to 50 µL of magnetic RNA Purification Beads that were intended to bind poly-A tails of the mRNA followed by mixing. Following incubation in the thermal cycler under the program *mRNA denaturation,* the sample was allowed to reach 4°C and then incubated at room temperature for 5 min.

To discard the supernant, the tube was then placed on a magnetic stand for 5 min to isolate the RNA-bound magnetic beads. Following a wash with 200 µL of bead washing buffer, the sample tube containing the beads was placed back on the magnetic tube rack for 5 min. Again with another wash, the supernatant was removed from the sample and discarded in the same fashion.

To the sample tube containing the beads, 50 µL of elution buffer was added and mixed, followed by an incubation in the thermal cycler with the program *mRNA elution 1*.
Once the sample reached 25°C, 50 µL of bead binding buffer was added and mixed, incubation followed at room temperature for 5 min. Following the same method as before, the tube containing the sample and beads were placed in the magnetic tube stand for 5 min to allow the supernatant to be discarded. Another subsequent was with 200 µL of bead washing buffer, followed by another 5 min on the magnetic stand again allowed for the removal and discarding of the supernatant. To the beads containing the sample, 19.5 µL of Elute, Prime, Fragment mix was added and mixed. Incubation with the program *Elution-2-Frag-Prime* in the thermal cycler was utilized to elute RNA from the beads, and after the sample had reached 4°C, it was briefly centrifuged and placed back on the magnetic stand for 5 min. In similar fashion as previously discussed the supernant was removed but then placed into a fresh PCR tube.

Synthesis of First Strand cDNA

To synthesize the first strand cDNA, 17 µL of the supernatant that contained the fragmented and primed mRNA was removed and placed in a new PCR tube. To this tube 1st Strand Master Mix (+SuperScriptII) was added and mixed and the thermal cycler incubation was

performed under the program *1st Strand*. After the program had completed, the sample tube then contained single stranded cDNA.

Synthesis of Second Strand cDNA

To achieve double stranded cDNA, the above sample was allowed to reach 4°C. To the sample, 25 µL of 2nd Strand Master Mix was added and mixed. Thermal cycler incubation under the program *2nd Strand* was utilized and following incubation, the tube then contained double stranded cDNA.

Purification of Double Stranded cDNA

To purify the double stranded cDNA, it was allowed to reach room temperature, and it was transferred to a 1.7 mL tube containing 90 µL of AmpureXP beads with mixing. Thermal incubation for 15 min at room temperature was allowed prior to being placed on a magnetic stand for 5 min. The double strand cDNA was attached to the beads which allowed the supernatant to be removed and discarded. To the beads, 200 µL of 80% ethanol was added and care was taken to not disturb the beads in the tube during a 30 s incubation. The supernant was discarded in the same fashion previously described. A subsequent washing step was completed as previously described and following the wash the tube was dried for 15 min. After drying an addition of 62.5 µL of Resuspension buffer was added and the sample was mixed when removed from the magnet. A 2 min incubation at room temperature was performed followed by replacement of the tube on the magnet for 5 min. The tube containing the sample was removed from the magnet and a 60 µL fraction of purified ds cDNA supernatant was transferred to a new tube.

End Repair and Reaction Clean-up

Once purified double stranded cDNA was isolated, it was necessary to perform an end repair step as well as an overall reaction clean-up. To the purified double stranded cDNA, 40 µL of end repair mix was added and mixed. The thermal cycler program *End Repair* was utilized and the sample was transferred to a new tube. To this new tube, 160 µL of AmpureXP beads were added and mixed, followed by 15 min incubation at room temperature. Again the sample was placed on the magnetic stand for 5 min to allow the supernant to be removed and discarded. To the sample tube, 200 µL of 80% ethanol was again added without disturbing the beads with a

30 s incubation. Once the supernant was removed an ethanol was repeated again followed by air drying for 15 min. Again 20 µL of Resuspension buffer was added and mixed with room temperature incubation for 2 min. Once incubation was complete, the sample was placed in the magnetic stand for 5 min allowing for a 17.5 µL fraction of the supernatant to be removed and placed in a fresh PCR tube.

Adenylation of 3' Ends, Adapter Ligation, and Reaction Clean-up

Following the end repair steps and reaction clean up, adenylation of 3' ends and ligation and more clean-up steps were needed post ligation. To begin the adenylation of the 3' end of the cDNA library, 12.5 µL of A-Tailing Mix was added to the sample and mixed. Thermal cycling program *ATAIL70* was utilized and when the sample had reached 4°C, it was removed.
For the Illumina Mi-Seq platform, adapters must be ligated onto the cDNA library so that sequencing can be performed. This ligation was achieved by adding 2.5 µL of Resuspension buffer and 2.5 µL of ligation mix to the sample tube with mixing. Following the thermal cycler program Ligation, it was removed, and a 5 µL aliquot of Stop ligation buffer was added and mixed to the sample tube.

For clean-up of the reaction mixture, another use of 42 µL of AmpureXP beads was added to the sample followed by mixing and a15 min incubation at room temperature. Following the established procedure the tube was placed on a magnetic stand and the supernatant was removed and discarded. Again without disturbing the beads, 200 µL of 80% ethanol was added, incubated for 30 s prior to discarding of the supernatant. In the same previously describe fashion, the ethanol wash was repeated followed by a 15 min sample drying time. Post drying, 62.5 µL of Resuspension buffer was added, mixed, incubated at room temperature for 2 min prior to being placed on a magnetic stand for 5 min. Removal of the 50 µL supernatant into a new 1.7 mL centrifuge tube followed with the addition of 50 µL of AmpureXP beads. A second clean-up was performed in the same fashion and the sample was then incubated for 15 min at room temperature prior to being placed back on the magnetic stand for 5 min. Samples were again washed twice with 200 µL of ethanol and incubated for 30 s prior to drying at room temperature for 15 min. To the sample tube, 22.5 µL of Resuspension Buffer was added and incubated for 2 min. Separation of the supernatant was performed by placing the tube on the magnetic stand for 5 min and removing a 20 µL sample of the supernatant for transfer to a new fresh PCR tube.

DNA Fragment Enrichment

Now that the purified cDNA was 3' polyadenylated and ligand adapted, the cDNA library was enriched by PCR. To the newly formed library, 5 µL of PCR primer cocktail and 25 µL of PCR master mix were added with mixing. The subsequent thermal cycling program *PCR* allowed for amplification of the cDNA library.

PCR Product Clean-Up

Post enrichment, the tube was removed from the thermal cycler and clean-up steps were again performed to purify the PCR product. To the tube, 50 µL of AmpureXP beads were added and mixed, followed by 15 min incubation at room temperature prior to being placed on a magnetic stand. Removal of the supernant allowed the beads to be washed with 200 µL of 80% ethanol, and incubated for 30 s. Again the supernatant was removed, and the ethanol wash was repeated. A 15 min air drying at room temperature preceded the addition of 32.5 µL of Resuspension Buffer with mixing and 2 min incubation at room temperature. Following incubation the sample was placed on a magnetic stand for 5 min and 30 µL of the supernatant was transferred to a new 1.7 mL centrifuge tube. This final volume of the supernatant would be the final cDNA library that was to be sequenced by RNA-seq on the Illumina Mi-Seq platform.

RNA-seq Library Validation and Sequencing

Prior to sequencing, the cDNA libraries were verified by an Agilent Bioanalyzer 2100. Equilibrated at room temperature for 30 min, 25 µL of DNA dye was added to DNA gel matrix. This solution was centrifuged at 1500 x g for 10 min post mixing of the two components. The 9 µL sample of Gel-Dye Mix was loaded into a specific well denoted as "G" on a DNA 7500 chip on the priming station.

Subsequent sequencing of the cDNA libraries at IGF-KSU on the Illumina Mi-Seq platform generated the following results:

RNA isolation: dissected head tissues

19,998,120 paired-end reads for the plant-fed cDNA library

10,516,022 paired-end reads for the diet-fed cDNA library.

All reads were 250 bases in length.

RNA-seq Read Mapping

To map the RNA-Seq reads the Assemble program in the software package Geneious was utilized. A full list of the mRNA transcripts used as a "reference genome" has been listed in Table 6. The assemblies of reads were mapped to the reference genome under the Medium-Low Sensitivity setting, which allows 10 gaps per read, and requires 18 consecutive bases to be identical to match a read to the genome. The allowed mismatch percentages for single bases are up to 20%.

RPKM Calculations

For the following calculations of reads per kilobase of exon per million reads mapped (RPKM) are listed below. R_M is the number of reads mapped to a reference sequence, L_T is the length of the reference transcript, and R_T is the total number of RNA-seq reads. Multiplying by 10^9 is a normalization factor. This factor is the mean length of a transcript in the transcriptome (1,000 base pairs) times one million.

$$RPKM = \left(\frac{R_M}{\left(\frac{L_T}{1{,}000}\right)\left(\frac{R_T}{1{,}000{,}000}\right)} \right)$$

$$RPKM = 10^9 * \left(\frac{R_M}{L_T * R_T} \right)$$

An RPKM calculation from a diet fed aphid by head dissection is outlined below for the UPR component Armet.

$$Armet\ diet\ fed\ RPKM = 14.53\ = 10^9 * \left(\frac{233}{1{,}525 * 10{,}516{,}022} \right)$$

Results:

In Figure 2.5 alignments of the amino acid sequence of GRP78 and Armet with their respective human orthologs show the level of amino acid residue conservation in two pairs of orthologs. The protein alignment of GRP78/BiP has a pairwise identity of 80.3% and the Armet

alignment has an identity of 46.5%. The e-value from the ortholog BLASTp pea aphid putative ortholog searches for GRP78 and Armet are 0.0 and 5e-42 respectively.

RNA isolated from heads of plant-fed and diet-fed pea aphids was isolated and submitted to the Integrated Genomics Facility at Kansas State University (IGF-KSU) for quality analysis by Agilent 2100 Bioanalyzer.

Bioanalyzer electropherograms of plant and diet-fed aphid RNA, Figure 2.3 & Figure 2.4, respectively, showed RNA isolations with good quality, suitable for RNA-seq cDNA library synthesis. The genomics facility then generated cDNA libraries, analysis of which is shown in Figures 2.6 and 2.7.

Table 4 indicates gene names and descriptions of the 91 human UPR components included in this research. In later tables, the pea aphid putative orthologs maintain the human gene names and descriptions. Some components have been identified by multiple gene names and for this dissertation, only one gene name is provided. An exception to this naming scheme is the identification of 4 canonical component names, which include GRP78, GRP94, PERK, and IRE1 where these identifications are located within the description in later tables.

RNA-Seq validation of transcripts from salivary gland dissections are shown in Table 6 from the combined plant and diet fed libraries. The validation of transcripts showed identification of UPR putative orthologs in pea aphids were present.

Head dissections validated with RNA-Seq reads from plant-fed and diet-fed salivary gland libraries were assembled for each of the 74 pea aphid UPR mRNA transcripts, using each transcript as a "reference genome." Table 7 shows the 74 mRNA transcripts that were utilized. Table 7 also reports the number of reads for each feeding state, transcript length, RPKM, and RPKM fold change as calculated by division of the plant-fed RPKM by the diet-fed RPKM. Transcripts reported by mRNA reference number in Table 7 are organized by the RPKM fold change ratio. The fold changes range from 4.92 to 0.52, with an average of 1.98 and standard deviation of 0.85, with only three ratios being less than 1. The C-terminal 4 residues are shown in Table 7 for each encoded pea aphid protein and those that are known to be ER retention motifs are indicated in yellow.

In Figure 2.8, the RPKM fold change for plant versus diet feeding for the entire pea aphid gene set is shown, as well as the RPKM fold change for UPR components. The mean RPKM fold change value for the entire gene set was 1.61 while the UPR components mean was 1.98. A

student's unpaired T test was utilized to compare the two means and at a 95% confidence interval and the difference was significant using the following data.

	Gene Set	UPR Components
Mean	1.6100	1.9800
SD	0.9085	0.8520
SEM	0.0062	0.1026
N	21501	74

Unpaired t test results

P value and statistical significance:

The two-tailed P value equals 0.0007

By conventional criteria, this difference is considered to be extremely statistically significant indicating that there is a difference between the means.

Confidence interval:

The mean of Gene Set minus UPR Components equals -0.37

95% confidence interval of this difference: From -0.6 to -0.2

Figure 2.9 aligns the nucleotide sequences of GRP78 and Armet and the nucleotide e-values after BLASTn pea aphid putative ortholog searches for GRP78 and Armet are 1e-156 and 4e-91 respectively and are used to show examples of putative orthology and locations for possible dsRNA generation.

Discussion:

UPR Activation and "Triggering"

Many literature sources indicate that the UPR is "triggered" when a certain threshold is reached, but my view of the UPR mirrors that of Matus et al. (2008) that the UPR is always on and is not "triggered," although not always functioning at 100% capacity. In other words, there is never a time when the UPR is simply on or off like a light switch but it functions like a rheostat always changing in regards to need.

Statistical Comparison of UPR Components versus the Entire Gene Set

Utilizing a t-test, the comparison of RPKM fold change means from the gene set (1.61) and UPR components (1.98) show that the difference is statistically significant at a 95%

confidence. At 1.61, the gene set's RPKM mean indicates that the entire gene set is upregulated as previously defined in all transcripts in plant feeding versus diet feeding. Until further replication is pursued, this phenomenon may be attributed to the mediation of plant defenses. The importance of this comparison of means is that the UPR components' expression is different from the overall population of the gene set. The increased mean value of RPKM fold change in UPR components confirms the hypothesis that the UPR is upregulated during plant feeding in aphids.

The Presence and Lack of ER Retention Signals

As seen in Table 7, a number of UPR components contain ER retention motifs targeting the encoded protein for the ER lumen. Because the UPR has a wide reach within the cell, not all UPR components must contain an ER retention motif. Components of the UPR occur not only in the ER, some are translocated to the nucleus for signaling and the cytosol for ERAD. It makes sense that many of the components do not possess an ER retention motif when addressing the entirety of the UPR, in comparison to just protein folding which does occur within the ER. This is seen in Table 7 where the chaperone proteins that are found within the lumen of the ER indeed possess a ER retention signal in the pea aphid. It is also apparent that transducers IRE1, PERK, and ATF6 do not contain an ER retention signal, which is logical because they are anchored to the ER membrane.

Analysis of RPKM Ratios

The head is comprised of roughly 50% neural tissue and 50% salivary gland. Because the UPR is present in all tissues, the data derived from head dissections presumably lowers the RPKM ratios from the values that would occur in salivary glands alone. While it is assumed that the salivary gland UPR components are upregulated, the neural tissues should not be affected since there are relatively few signaling components in comparison to saliva components. In other words, I assume that all changes in RPKM values in head RNA stem from salivary gland RNA in plant feeding versus diet feeding.

For an example, if a component of the UPR had a RPKM ratio of 2 in salivary glands, in the presence of 50% neural tissue from a head dissection, the value would be 1.5 for head RNA assuming that neural tissue UPR transcript levels remain unchanged. Therefore, the ratio of

plant-fed to diet-fed RPKM in salivary gland RNA is presumably higher than reported in this work, that is, for RPKM ratios different from 1.0.

Upregulation of the UPR in Plant Feeding

I hypothesized transcripts of the UPR are upregulated in the salivary gland. The use of heads versus salivary glands for this work was performed to lessen RNA degradation. During plant feeding, the increase of UPR RPKM values is attributed to the increased secretion of saliva proteins that activate the UPR. With insights from chapter 3 of this dissertation, upregulation of the UPR and saliva proteins go hand in hand.

All but 3 of the pea aphid transcripts studied had higher expression in plant-fed than in diet-fed heads. The range of the fold change was 4.92 to 0.52 in head RNA isolations. The five highest fold changes were for TOR1A, PPIA, BAX, CALR, and PFDN5 proteins. The components TOR1A, PPIA, CALR, & PFDN5 are chaperones that solidify the hypothesis that the increased secretion of saliva proteins requires further activation of UPR components, namely chaperones. The component BAX is associated with apoptosis and may be increased due to cells activating UPR mediated apoptosis. Because feeding on plants is much more complex than feeding on artificial diets and hence may require more saliva proteins requiring more UPR induction. Plants have defense mechanisms that protect them from invading pathogens or insect pests (Fürstenberg-Hägg et al., 2013), and some pea aphid saliva proteins may help circumvent these systems. In addition to mitigating plant defenses, the proteins of saliva themselves may aid in digestion of host plant proteins, facilitating absorption of nutrients (Fürstenberg-Hägg et al., 2013).

Of the 4 PDI transcripts found, each had a higher RPKM in RNA from plant feeding versus diet feeding giving support to the idea that the UPR is upregulated during plant feeding. It makes sense that the upregulation of PDI transcripts occurs to promote proper disulfide formation in proteins within the ER. As seen in all but three transcripts, the data presented here also coincides with the idea of the UPR being upregulated during plant feeding.

A New Method for Measuring UPR Upregulation

In contrast to studies outlined previously in *C. elegans* (Shen et al., 2005) and *D. melanogaster* (Palgi et al., 2009) using the compound tunicamycin as a method for inducing ER stress, here I have used a natural method of measuring upregulation in UPR components, namely

plant feeding versus diet feeding. This upregulation indicates the presence of ER stress and therefore upregulation of the UPR.

While many studies have been completed with mutant nematodes and variant lines of flies with altered UPR components, I have not found a similar natural induction and measurement of the UPR by feeding. This coupled without the use of tunicamycin induced ER stress may lead to new avenues UPR research in other Hemiptera insects which use saliva to feed on plants with this experimental method.

Comparison of UPR Components versus C. elegans

A comparison to *C. elegans* indicates that feeding of aphids on plants versus diets can mimic the induction of the UPR in nematodes induced with tunicamycin. Fold changes seen for my natural method of UPR induction *via* plant feeding range from 4.92 to 0.52 with an average fold change of 1.98. The non-natural tunicamycin UPR induction measured in Shen et al. (2005) possesses a range of fold changes from 4.94 to 0.98 with an average of 1.67. When comparing these means by T-test, the difference of means is statistically significant, however the ranges of upregulation are extremely close indicating that this method of natural induction of the UPR by feeding is valid and comparable to established ER stress induction methods.

Although the list from Shen et al. (2005) contained different components from my list, their measurement of induction was similar to my studies. There was a small overlap of components in both of our lists including 12 inducible UPR components and 2 constitutive components identified as putative orthologs in the pea aphid.

Silencing UPR Components and Pesticide Free Pest Mitigation

The long term goal of this research is to gain insight into reducing pea aphid fecundity or causing pea aphid death without the use of pesticides. Documentaring UPR components in the pea aphid may be an important aspect for future dsRNA silencing studies, reducing the viability of the pea aphid.

Obviously the most promising goal is using genetically modified crops to combat the pest without the use of pesticides. Transforming a plant to produce dsRNA against an insect component for transcript knockdown is something that has been achieved (Todd et al., 2008, Liu et al., 2015).

Using genetically modified crops targeting insects is not a new idea. For example, Bt corn does not use dsRNA, but combat pests without the use of pesticides. In Bt corn, a *Bacillus thuringiensis* toxin was inserted which encodes a protein targeting coleopteran or lepidopteran insect pests (Gordon et al., 2007).

Allowing a plant to combat insects with dsRNA is promising, but there might be challenges to be addressed such as the identification of the sequence similarity of targeted UPR components in pea aphids versus humans. Reported on Monsanto's research and development products page (http://www.monsanto.com/products/pages/corn-pipeline.aspx) an example of a dsRNA producing corn crop, knockdowns a target in a rootworm. This corn has been engineered specifically to produce a double-stranded RNA, in this case to inactivate a gene called Snf7 that is essential for moving proteins around in the rootworm (Bolognesi et al., 2012).

A nucleotide alignment shown in Figure 2.9A of the human and pea aphid putative orthologs of the canonical component GRP78/BiP showed a 65.8% pairwise identity between the two sequences. But choosing a region of sequence targeted toward pea aphids without targeting humans can be achieved as indicated in the figure in red boxes. The typical length of dsRNA is 19-22 nucleotides long. In off target mRNA transcripts, if 2 or more nucleotides are mismatched with the dsRNA, no off target effects will be seen. This alignment and evaluation of nucleotide sequences is crucial in ensuring that researchers do not target human or non-target insect transcripts with transgenic crops. While a transgenic crop could theoretically naturally evolve to target a human or another insect over thousands or hundreds of thousands of years, it is not likely.

A non-canonical component, Armet in a similar nucleotide alignment shown in Figure 2.9B shows a pairwise identity of 48.8% It too could be utilized as a good target for pest mitigation with a sequences identified in the figure that could be used for pea aphid silencing.

Both examples above for knockdown studies do not account for any other off target silencing effects, and further nucleotide (BLASTn) searches of the final dsRNA would need to be completed. This search ensures that the nucleotides that are chosen for the dsRNA do not share sequence similarity to other human mRNA sequences.

Another challenge to address is the uptake of the dsRNA itself. It is unclear if a dsRNA in a crop would have uptake into the hemolymph by way of the gut. A major hurdle to first address is the delivery of dsRNA by feeding measuring its efficacy. Without uptake into the

hemolymph dsRNA will not have any effect on tissues other than the gut which may not kill or reduce aphid fecundity.

To choose a component for knockdown studies, multiple approaches may be used. The least economical method is to test all 91 components with dsRNA feeding studies, where each dsRNA is contained within a diet evaluating fecundity and longevity. A better method may be to knockdown chaperones, evaluating aphid fecundity and lifespan for the components of the UPR that have been the most widely researched in other organisms. Lastly, I would propose a method that encompasses some of the earlier two methods. Knockdown of chaperones and at least one component from each subsystem individually or in concert with other targets, may give the best insights into future targets for silencing experiments and transgenic plant production. In any case, my work presented here in conjunction with any of the above methods could give good insight into the importance of the UPR on plant feeding insects.

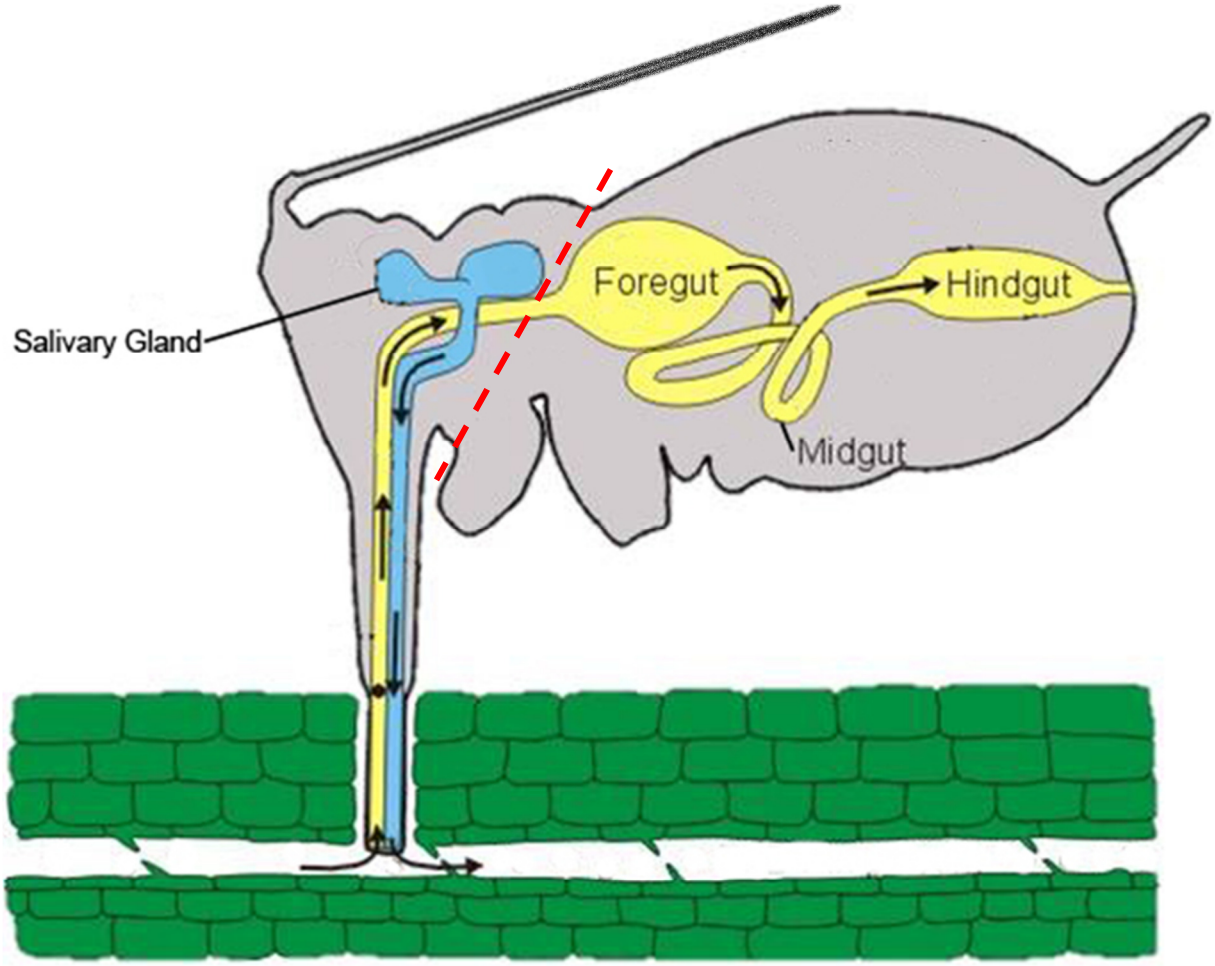


Figure 2.1 Cross section of pea aphid feeding on phloem element

Image obtained from D'Arcy et al. (2000) from a Google image search, originally originating from The American Phytopathological Society retrieved at: www.apsnet.org, modified for this dissertation. The location of dissection for head removal in RNA isolations is indicated by the red dotted line.

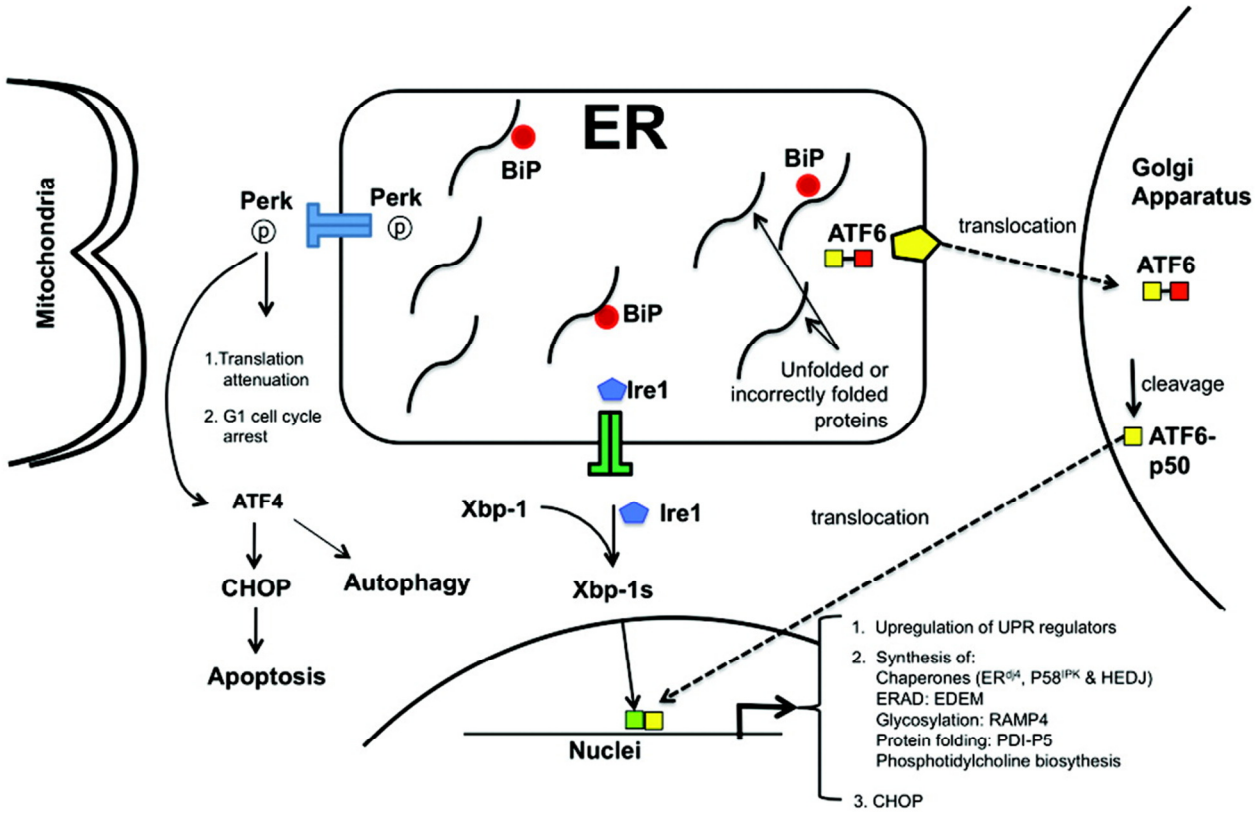


Figure 2.2 Schematic of the UPR pathway activated by the accumulation of unfolded or incorrectly folded proteins, by sequestering BiP

Reprinted with permission from Du, Z., Treiber, D., McCoy, R.E. Unfolded Protein Response (UPR) During CHO Cell Production Culture. Developments in Biotechnology and Bioprocessing. American Chemical Society. Copyright (2013) American Chemical Society.

Table 4 List of 91 human UPR components with descriptions

List of 91 Human UPR gene names & descriptions. Color coordinated by function.

ERAD	Brown
Ubiquination	Purple
Cholesterol	Gray
Transducers	Green
Apoptosis	Red
Transcription	Blue
Translation	Orange
Protein Folding	Yellow

Gene Name	Description
AMFR	Autocrine motility factor receptor, E3 ubiquitin protein ligase
ARMET	Mesencephalic astrocyte derived neurotrophic factor
ATF4	Activating transcription factor 4
ATF6A	Activating transcription factor 6 alpha
ATF6B	Activating transcription factor 6 beta
BAX	BCL2-associated X protein
CALR	Calreticulin
CANX	Calnexin
CCT4	Chaperonin containing TCP1 subunint 4
CCT7	Chaperonin containing TCP1 subunint 7
CEBPD	CCAAT/enhancer binding protein
CREB3	cAMP responsive element binding protein 3
CREB3L3	cAMP responsive element binding protein 3-like 3
DDIT3	DNA damage inducible transcript 3
DERL1	Derlin 1
DERL2	Derlin 2
DNAJB2	DnaJ (Hsp40) homolog subfamily B member 2
DNAJB9	DnaJ (Hsp40) homolog subfamily B member 9
DNAJC10	DnaJ (Hsp40) homolog subfamily C member 10
DNAJC3	DnaJ (Hsp40) homolog subfamily C member 3
DNAJC4	DnaJ (Hsp40) homolog subfamily C member 4
EDEM1	ER degradation enhancer, mannosidase alpha-like 1
EDEM3	ER degradation enhancer, mannosidase alpha-like 3
EIF2A	Eukaryotic intiation factor 2 alpha
EIF2AK3	Eukaryotic intiation factor 2 alpha kinase 3 (PERK)
EIF2B	Eukaryotic intiation factor 2 beta
ERN1	Inositol-requiring enzyme 1 (IRE1)
ERN2	Endoplasmic reticulum to nucleus signaling 2 (IRE1B)
ERO1	Endoplasmic oxidoreductin 1
ERO1L	Endoplasmic oxidoreductin 1 like
ERO1LB	Endoplasmic oxidoreductin 1 like beta

ERAD	Brown
Ubiquination	Purple
Cholesterol	Gray
Transducers	Green
Apoptosis	Red
Transcription	Blue
Translation	Orange
Protein Folding	Yellow

Gene Name	Description
ERP44	Endoplasmic reticulum protein 44
FBXO6	F-box protein 6
GANAB	Glucosidase alpha neutral AB
GANC	Gulcosidase alpha neutral C
GRP78	Glucose regulating protein 78kDa
GRP75	Glucose regulating protein 75kDa
GRP170	Glucose regulating protein 170kDa
HERPUD1	Homocysteine-inducible ER stress-inducible, ubiquitin-like domain member 1
HSP90B1	Heat shock protein 90kDa beta (GRP94) member 1
HSPA1B	Heat shock protein 1B
HSPA1L	Heat shock 70kDa protein 1-like
HSPA2	Heat shock 70kDa protein 2
HSPA4	Heat shock 70kDa protein 4
HSPA4L	Heat shock protein 4-like
HSPH1	Heat shock 105kDa/110kDa protein 1
HTRA2	HtrA serine peptidase 2
HTRA4	HtrA serine peptidase 4
INSIG1	Insulin induced gene 1
INSIG2	Insulin induced gene 2
MAPK10	Mitogen activated protein kinase 10
MAPK8	Mitogen activated protein kinase 8
MAPK9	Mitogen activated protein kinase 9
MBTPS1	Membrane-bound transcription factor peptidase, site 1
MBTPS2	Membrane-bound transcription factor peptidase, site 2
NPLOC4	Nuclear protein localization 4 homolog
NUCB1	Nucleobindin 1
OS9	Osteosarcoma amplified 9
PDIA	Protein disulfide isomerase
PDIA3	Protein disulfide isomerase A, member 3
PDIA5	Protein disulfide isomerase A, member 5
PDIA6	Protein disulfide isomerase A, member 6

ERAD	Brown
Ubiquination	Purple
Cholesterol	Gray
Transducers	Green
Apoptosis	Red
Transcription	Blue
Translation	Orange
Protein Folding	Yellow

Gene Name	Description
PFDN2	Prefoldin subunit 2
PFDN5	Prefoldin subunit 5
PPIA	Peptidylprolyl isomerase A
PPP1R15A	Protein phosphatase 1, regulatory subunit 15A
PRKCSH	Protein kinase C substrate 80K-H
RNF139	Ring finger protein 139
RNF5	Ring finger protein 5, E3 ubiquitin protein ligase
RPN1	Ribophorin 1
SCAP	SREBF chaperone
SEC62	SEC62 homolog (S. cerevisiae)
SEC63	SEC63 homolog (S. cerevisiae)
SEL1L	Sel-1 suppressor of lin-12-like
SELS	VIMP VCP-interacting membrane protein
SERP1	Stress-associated endoplasmic reticulum protein 1
SIL1	SIL1 homolog, ER chaperone (S. cerevisiae)
SREBF1	Sterol regulatory element binding transcription factor 1
SREBF2	Sterol regulatory element binding transcription factor 2
SYVN1	Synovial apoptosis inhibitor 1
TCP1	T-complex 1
TOR1A	Torsin family 1, member A
TRAF2	TNF receptor-associated factor 2
UBE2G2	Ubiquitin-conjugating enzyme E2G 2
UBXN4	UBX domain protein 4
UFD1L	Ubiquitin fusion degradation 1 like
UGGT1	UDP-glucose glycoprotein glucosyltransferase 1
UGGT2	UDP-glucose glycoprotein glucosyltransferase 2
USP14	Ubiquitin specific peptidase 14
VCP	Valosin containing protein
XBP1	X-box binding protein 1

ERAD	Brown
Ubiquination	Purple
Cholesterol	Gray
Transducers	Green
Apoptosis	Red
Transcription	Blue
Translation	Orange
Protein Folding	Yellow

Essential Amino acids:	
L-Arginine HCl	12.5 mM
L-Histidine	7.5 mM
L-Isolucine	7.5 mM
L-Leucine	7.5 mM
L-Lysine HCl	7.5 mM
L-Methionine	2.5 mM
L-Phenylalanine	2.5 mM
L-Threonine	7.5 mM
L-Tryptophan	2.5 mM
L-Valine	7.5 mM
Nonessential amino acids:	
L-Alanine	5 mM
L-Asparagine	12.5 mM
L-Aspartic acid	12.5 mM
L-Cysteine HCl	2.5 mM
L-Cysteine	0.2 mM
L-Glutamic acid	7.5 mM
L-Glutamine	15 mM
Glycine	1 mM
L-Proline	5 mM
L-Serine	5 mM
L-Tyrosine	0.5 mM
Gamma amino butyric acid	2 mM

Trace Metals:	
Cupric chloride	14 µM
Ferric chloride	49 µM
Magnesium (II) chloride	40 µM
Zinc sulfate	30 µM
Salts, Buffers, and Sterol:	
Calcium citrate	0.175 mM
Cholesterol benzoate	50 µM
Potasium phosphate	18.37 mM
Sodium chloride	0.127 mM
Magnesium chloride	9.837 mM
Choline chloride	3.579 mM
Vitamins:	
p-Aminobenzoic acid	0.73 mM
Ascorbic acid	5.68 mM
D-Calcium pantothenate	0.21 mM
Folic acid	22 mM
Inositol (meso) dihydrate	1.39 µM
Nicotinc acid	0.812 mM
Pyridoxine HCl	0.21 mM
Thiamine HCl	74 µM
Sugars:	
Sucrose	0.5 mM

Table 5 Artificial diet (Akey and Beck 1971, 1972)

Pea aphids were fed for 48 h on this artificial diet, which is referred as the Akey/Beck diet throughout this dissertation.

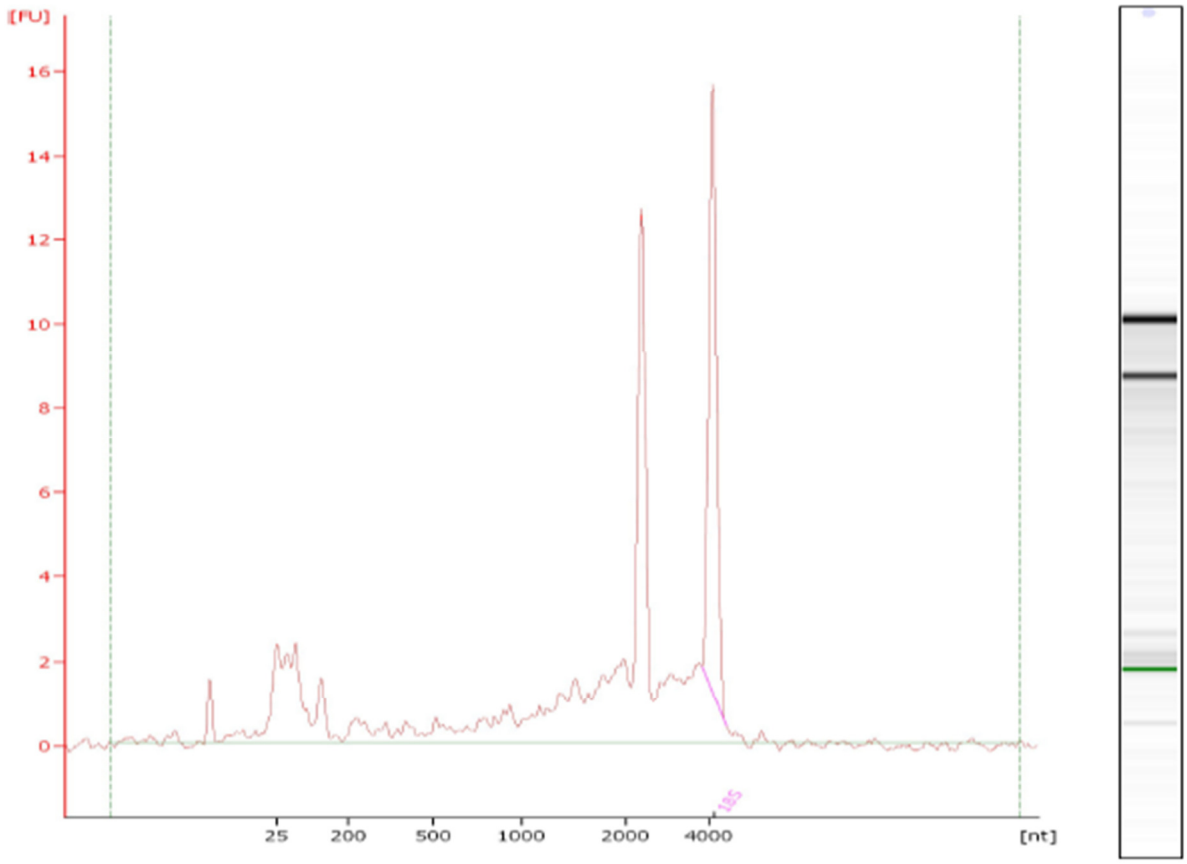


Figure 2.3 Bioanalyzer electropherogram of plant-fed aphid RNA

RNA isolated from plant-fed reared pea aphids were analyzed on an Agilent 2100 Bioanalyzer for determination of RNA quality.

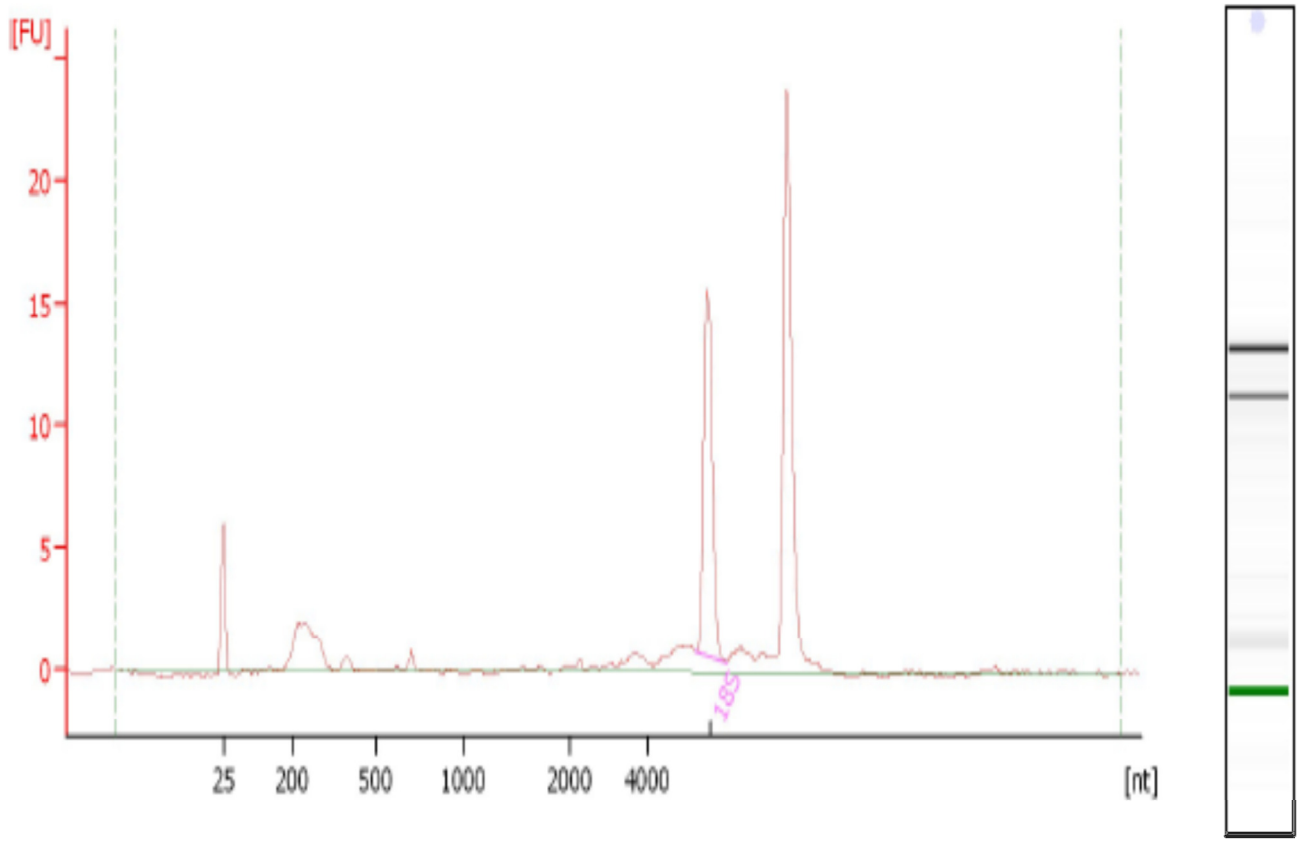


Figure 2.4 Bioanalyzer electropherogram of diet-fed aphid RNA

RNA isolated from diet-fed reared pea aphids were analyzed on an Agilent 2100 Bioanalyzer for determination of RNA quality.

Table 6 Verification of reads with salivary gland dissections

Reads generated by RNA-seq were mapped to each individual transcript open reading frame as a "reference genome." The number of salivary gland reads and RPKM values, calculated as described in the text are given for each transcript. Color coding of each transcript is continued.

NCBI mRNA Identification (aphid)	Gene Name (human)	Transcript Length	Salivary Gland Isolation	
			Reads Mapped	RPKM
XM_003242736.1	AMFR	4407	3893	5.93
XM_003247466.1	ATF4	2337	13714	33.75
XM_003245077.1	ATF6A	1935	1238	3.66
XM_001948762.2	BAX	1626	1105	4.18
XM_003240040.1	BiP	2911	104431	203.28
XM_001945770.2	CALR	2299	104161	231.58
XM_001948045.2	CANX	5205	21792	27.53
XM_001948927.2	CCT4	2594	9451	22.81
XM_003246689.1	CCT7	2500	11122	27.01
XM_001949174.2	Cebpd	4415	2822	4.50
XM_001951715.2	CREB3	2970	18938	42.35
XM_003246689.2	DDIT3	1954	21	0.10
XM_001951549.2	DERL1	1625	10438	35.45
NM_001162746.2	DERL2	3033	8328	17.57
XM_003247556.1	DNAJB2	1226	6176	28.04
XM_001949024.1	DNAJB9	1954	5960	17.26
NM_001162097.2	DNAJC3	3137	19279	38.80
XM_001946233.1	DNAJC4	1010	1199	6.56
XM_001945860.2	EDEM1	3269	7187	14.61
XM_003245436.1	EIF2A	1947	2831	8.38
XM_001942883.2	EIF2B	1267	8658	38.06
XM_001951459.2	ERO1	1425	2224	9.99
XM_001950428.2	ERP44	3318	26043	50.78
XM_001943249.2	FBXO6	2466	2178	5.29
XM_003244000.1	GANAB	4148	8493	13.93
XM_001946431	GRP170	4478	16031	25.70
XM_001948031	GRP75	3605	68637	1810.51
XM_001950766.2	HERPUD1	2335	8970	25.16
XM_001948902.2	HSP90B1	3358	65238	126.68
XM_001951172.2	HSPA1L	3316	206274	333.89
XM_001951757.2	HSPA4	3376	14161	26.41
XM_001945735.2	HTRA2	1706	4923	17.77
XM_003242717.1	HTRA4	1438	157	0.67
XM_001944194.2	INSIG1	1797	459	1.54
XM_001943638.2	IRE1	3782	21515	35.27
XM_001949506.2	MANF	1525	9752	44.09
XM_003242827.1	MAPK8	2740	6248	14.26

NCBI mRNA Identification (aphid)	Gene Name (human)	Transcript Length	Salivary Gland Isolation	
			Reads Mapped	RPKM
XM_001952362.2	MBTPS1	4002	3988	6.61
XM_001950402.2	MBTPS2	2577	1553	4.15
XM_001951793.2	NPLOC4	2487	3671	9.46
XM_001946280.2	NUCB1	2921	6620	14.23
XM_001944320.2	OS9	1374	13251	59.09
XM_008184943.1	PDIA	2931	27463	67.27
XM_001950371.2	PDIA3	2212	28283	76.73
XM_008188836.1	PDIA5	824	21005	2424.06
XM_001948267.2	PDIA6	2357	21648	873.39
XM_003245614.1	PERK	3862	2503	4.29
XM_003240262.1	PFDN2	1478	21	0.10
XM_001162260.2	PFDN5	1236	3137	14.01
XM_001945068.2	PPIA	1029	22085	104.79
XM_001945556.3	PPP1R15A	1291	21	0.10
XM_001948968.2	PRKCSH	2174	9543	29.80
XM_001943758.2	RNF139	4454	2757	4.30
XM_001950468	RNF5	1352	21	0.10
XM_003243279.1	RPN1	3911	10474	17.25
XM_003242865.1	SCAP	4722	4680	6.75
XM_001949921.2	SEC62	2131	10387	30.93
XM_003242649.1	SEC63	3199	13884	27.25
XM_003240171.1	SEL1L	3841	15650	28.40
XM_003248234.1	SELS	3849	1017	1.79
XM_001946233.1	SERP1	3269	7775	15.58
XM_001943931.2	SIL1	2255	10602	32.68
XM_001947517.2	SREBF1	4166	2920	4.75
XM_001943033.2	TCP1	2077	7795	24.61
XM_001946078.2	TOR1A	3134	927	2.03
XM_001948320.2	TRAF2	2299	2308	6.41
NM_001162605.1	UBE2G2	985	12436	46.50
XM_003241322.1	UBXN4	3255	4478	8.55
XM_001945406.2	UFD1L	1616	1981	8.23
XM_001948648.2	UGGT1	21726	20929	6.74
XM_001944664.2	UGGT2	5251	11437	14.69
XM_003244618.1	USP14	2362	5532	13.97
XM_001948341.2	VCP	2787	1844	4.32
XM_003248521.1	XBP1	4027	16392	28.47

ERAD	Brown
Ubiquination	Purple
Cholesterol	Gray
Transducers	Green
Apoptosis	Red
Transcription	Blue
Translation	Orange
Protein Folding	Yellow

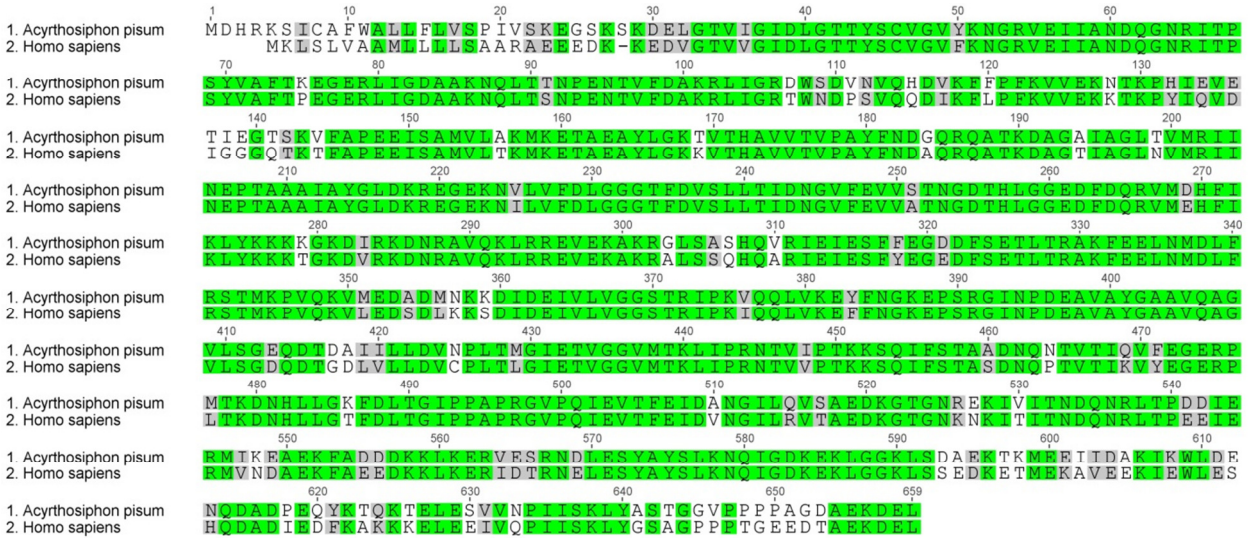
Table 7 Comparative analysis of 74 UPR in diet and plant-fed libraries by RNA-seq

Reads generated by RNA-seq were mapped to each individual transcript open reading frame as a "reference genome". Salivary gland RPKM values as well as head RPKM and fold changes values, calculated as described in the text. ER retention motifs are indicated in yellow if possessed by the pea aphid proteins encoded by the represented mRNAs.

UPR Components

NCBI mRNA Identification (aphid)	Gene Name (human)	Description (human)	Transcript Length	Salivary Gland Isolation		Head Isolation				RPKM Fold Change	Aphid C-terminal Tetra Peptide
						Diet Fed		Plant Fed			
				Reads Mapped	RPKM	Reads Mapped	RPKM	Reads Mapped	RPKM		
XM_003248234.1	SELS	sel-1 suppressor of lin-12-like	3849	1017	1.79	0	0.00	483	6.27	∞	MRER
XM_001946078.2	TOR1A	torsin family 1, member A (torsin A)	3134	927	2.03	14	0.42	131	2.09	4.920	SNLI
XM_001945068.2	PPIA	peptidylprolyl isomerase A (cyclophilin A)	1029	22085	104.79	1091	100.82	9901	481.14	4.772	GQLS
XM_001948762.2	BAX	BCL2-associated X protein	1626	1105	4.18	25	1.46	172	5.29	3.618	SVFR
XM_001945770.2	CALR	Calreticulin	2299	104161	231.58	2067	85.50	13497	293.57	3.434	HDEL
XM_001162260.2	PFDN5	prefoldin subunit 5	1236	3137	14.01	118	9.08	755	30.54	3.365	TENK
XM_001951793.2	NPLOC4	nuclear protein localization 4 homolog	2487	3671	9.46	98	3.75	627	12.61	3.364	RDIN
XM_001949024.1	DNAJB9	DnaJ (Hsp40) homolog, subfamily B, member 9	1954	5960	17.26	251	12.22	1536	39.31	3.218	DTLP
XM_003246689.1	CCT7	chaperonin containing TCP1 subunit 7	2500	11122	27.01	398	15.14	2341	46.82	3.093	GRPM
XM_001948267.2	PDIA6	protein disulfide isomerase family A, member 6	2357	21648	65.93	517	20.86	3010	63.86	3.062	KEEL
XM_001950371.2	PDIA3	protein disulfide isomerase family A, member 3	2212	28283	76.73	920	39.55	4876	110.23	2.787	KHEL
XM_008188836.1	PDIA5	protein disulfide isomerase family A, member 5	824	21005	183.00	599	69.13	3105	188.43	2.726	KHEL
XM_001943033.2	TCP1	t-complex 1	2077	7795	24.61	384	17.58	1981	47.69	2.713	AGEL
XM_001949506.2	MANF	Armet	1525	9752	44.09	233	14.53	1190	39.02	2.686	KEEL
XM_003245077.1	ATF6A	Activating transcription factor 6A	1935	1238	3.66	34	1.67	171	4.42	2.645	LPSY
XM_001948341.2	VCP	valosin containing protein	2787	1844	4.32	61	2.08	304	5.45	2.621	APRS
XM_001948927.2	CCT4	Chaperonin containing TCP1 subunit 4	2594	9451	22.81	402	14.74	1989	38.34	2.602	TRGY
XM_001951757.2	HSPA4	heat shock 70kDa protein 4	3376	14161	26.41	553	15.58	2728	40.41	2.594	GNDA
XM_001948031	GRP75	glucose regulating protein 75	3605	68637	136.68	2924	77.13	12801	177.56	2.302	KDEL
NM_001162746.2	DERL2	degradation in endoplasmic reticulum protein 2	3033	8328	17.57	252	7.90	1074	17.71	2.241	RQND
XM_001948902.2	HSP90B1	heat shock protein 90kDa beta (Grp94), member 1	3358	65238	126.68	1326	37.55	5646	84.08	2.239	HDEL
XM_001942883.2	EIF2B	eukaryotic translation initiation factor 2 beta	1267	8658	38.06	464	34.82	1919	75.74	2.175	QLQL
XM_001948045.2	CANX	Calnexin	5205	21792	27.53	962	17.58	3939	37.84	2.153	TRKD
XM_001946280.2	NUCB1	nucleobindin 1	2921	6620	14.23	312	10.16	1236	21.16	2.083	NKNQ
NM_001162097.2	DNAJC3	DnaJ (Hsp40) homolog, subfamily C, member 3	3137	19279	38.80	413	12.52	1611	25.68	2.051	FNFN
XM_003241322.1	UBXN4	UBX domain protein 4	3255	4478	8.55	149	4.35	581	8.93	2.050	TQQL
XM_003244000.1	GANAB	Glucosidase, Alpha; Neutral AB	4148	8493	13.93	331	7.59	1289	15.54	2.048	ITLL
XM_001952362.2	MBTPS1	membrane-bound transcription factor peptidase, site 1	4002	3988	6.61	160	3.80	619	7.73	2.034	GYNL
XM_003247466.1	ATF4	Activating transcription factor 4	2337	13714	33.75	804	32.71	3028	64.79	1.980	GLLN
XM_003240040.1	BiP	glucose regulating protein 78	2911	104431	203.28	2058	67.23	7726	132.72	1.974	KDEL
XM_001950766.2	HERPUD1	homocysteine-inducible, endoplasmic reticulum stress-inducible	2335	8970	25.16	193	7.86	714	15.29	1.945	PDII
XM_003243279.1	RPN1	ribophorin I	3911	10474	17.25	352	8.56	1287	16.46	1.923	TQKN
XM_001950468	RNF5	ring finger protein 5, E3 ubiquitin protein ligase	1352	21	0.10	190	13.36	681	25.19	1.885	
XM_001945406.2	UFD1L	ubiquitin fusion degradation 1 like	1616	1981	8.23	117	6.88	409	12.66	1.838	TKKN
XM_003242827.1	MAPK8	mitogen-activated protein kinase 8	2740	6248	14.26	296	10.27	999	18.23	1.775	QPIR
XM_001946233.1	DNAJC4	DnaJ (Hsp40) homolog, subfamily C, member 4	1010	1199	6.56	96	9.04	322	15.94	1.764	IVKK
XM_001946431	GRP170	glucose regulating protein 170	4478	16031	#REF!	330	7.01	1104	12.33	1.759	HTEL
XM_003247556.1	DNAJB2	DnaJ (Hsp40) homolog, subfamily B, member 2	1226	6176	28.04	337	26.14	1116	45.52	1.741	AYGH
XM_003240171.1	SEL1L	Sel-1 suppressor	3841	15650	28.40	417	10.32	1359	17.69	1.714	PQNV
XM_003242649.1	SEC63	Translocation protein SEC63	3199	13884	27.25	573	17.03	1866	29.17	1.712	DVED
XM_001944320.2	OS9	osteosarcoma amplified 9, endoplasmic reticulum lectin	1374	13251	59.09	202	13.98	657	23.91	1.710	NKYY
XM_001950428.2	ERP44	Thioredoxin domain containing protein 4	3318	26043	50.78	592	16.97	1921	28.95	1.706	KEEL
XM_001945556.3	PPP1R15A	protein phosphatase 1, regulatory subunit 15A	1291	21	0.10	9163	674.93	29578	1145.65	1.697	
XM_001949174.2	Cebpd	CCAAT/enhancer binding protein delta	4415	2822	4.50	317	6.83	1016	11.51	1.685	PHLQ
XM_001950402.2	MBTPS2	membrane-bound transcription factor peptidase, site 2	2577	1553	4.15	50	1.85	160	3.10	1.683	KIIN
XM_001947517.2	SREBF1	sterol regulatory element binding transcription factor 1	4166	2920	4.75	91	2.08	291	3.49	1.682	SVTD
XM_001951172.2	HSPA1L	heat shock 70kDa protein 1-like	3316	206274	333.89	11929	342.09	37005	558.03	1.631	EEVD
XM_003242865.1	SCAP	sterol regulatory element binding transcription factor chaperone	4722	4680	6.75	152	3.06	466	4.93	1.612	TKED
XM_001943249.2	FBXO6	F-box only protein 6	2466	2178	5.29	117	4.51	358	7.26	1.609	AAEA
XM_001948968.2	PRKCSH	protein kinase C substrate 80K-H	2174	9543	29.80	297	12.99	884	20.33	1.565	HDEL
XM_001949921.2	SEC62	Translocation protein SEC62	2131	10387	30.93	669	29.85	1957	45.92	1.538	AQDT
XM_003245436.1	EIF2A	Eukaryotic translation initiation factor 2 alpha	1947	2831	8.38	145	7.08	419	10.76	1.520	NEEE
XM_003248521.1	XBP1	X-box binding protein 1	4027	16392	28.47	584	13.79	1677	20.82	1.510	PMQT
XM_001943758.2	RNF139	ring finger protein 139	4454	2757	4.30	144	3.07	408	4.58	1.490	ADNS
XM_001951715.2	CREB3	cAMP responsive element binding protein 3	2970	18938	42.35	883	28.27	2484	41.82	1.479	SESY
XM_003244618.1	USP14	ubiquitin specific peptidase 14 (tRNA-guanine transglycosylase)	2362	5532	13.97	374	15.06	1041	22.04	1.464	SVSS
XM_001943931.2	SIL1	SIL1 nucleotide exchange factor	2255	10602	32.68	111	4.68	302	6.70	1.431	PVLE
XM_003242736.1	AMFR	autocrine motility factor receptor	4407	3893	5.93	289	6.24	764	8.67	1.390	SKTD
XM_003246689.2	DDIT3	DNA-damage-inducible transcript 3 (CHOP)	1954	21	0.10	343	16.69	903	23.11	1.384	
XM_001945735.2	HTRA2	HtrA serine peptidase 2	1706	4923	17.77	285	15.89	750	21.98	1.384	HSTI
XM_001951459.2	ERO1	ER oxidoreductin	1425	2224	9.99	130	8.68	332	11.65	1.343	QLFA
XM_008184943.1	PDIA	protein disulfide isomerase family A, member	2931	27463	#REF!	2778	90.13	7050	120.28	1.335	KEEL
XM_001946233.1	SERP1	Stress-associated endoplasmic reticulum protein 1	3269	7775	15.58	185	5.38	444	6.79	1.262	IRSA
NM_001162605.1	UBE2G2	ubiquitin-conjugating enzyme E2G 2	985	12436	46.50	1400	135.16	3228	163.87	1.212	PTSK
XM_001948648.2	UGGT1	UDP-glucose glycoprotein glucosyltransferase 1	21726	20929	6.74	3580	15.67	8233	18.95	1.209	FWKQ
XM_001945860.2	EDEM1	ER degradation enhancer, mannosidase alpha-like 1	3269	7187	14.61	193	5.61	443	6.78	1.207	LGAI
XM_001943638.2	IRE1	Inositol requiring enzyme 1	3782	21515	35.27	1611	40.51	3379	44.68	1.103	TSEQ
XM_001944664.2	UGGT2	UdP-glucose glycoprotein glucosyltransferase 2	5251	11437	14.69	450	8.15	938	8.93	1.096	HTEL
XM_003242717.1	HTRA4	HtrA serine peptidase 4	1438	157	0.67	17	1.12	35	1.22	1.083	RKMV
XM_001948320.2	TRAF2	TNF receptor-associated factor 2	2299	2308	6.41	117	4.84	232	5.05	1.043	IVAV
XM_001951549.2	DERL1	degradation in endoplasmic reticulum protein 1	1625	10438	35.45	642	37.57	1245	38.31	1.020	GQQQ
XM_003245614.1	PERK	Eukaryotic translation initiation factor 2 alpha kinase 3	3862	2503	4.29	196	4.83	297	3.85	0.797	KLQK
XM_001944194.2	INSIG1	insulin induced gene 1	1797	459	1.54	18	0.95	22	0.61	0.643	GRKS
XM_003240262.1	PFDN2	prefoldin subunit 2	1478	21	0.10	3	0.19	3	0.10	0.526	VNRD

A.



B.

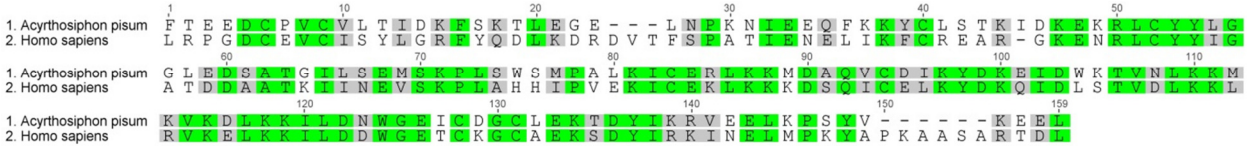


Figure 2.5 Human and pea aphid protein alignments of the proteins Armet and GRP78.

A. Protein alignment of human and pea aphid GRP78/BiP

B. Protein alignment of human and pea aphid Armet

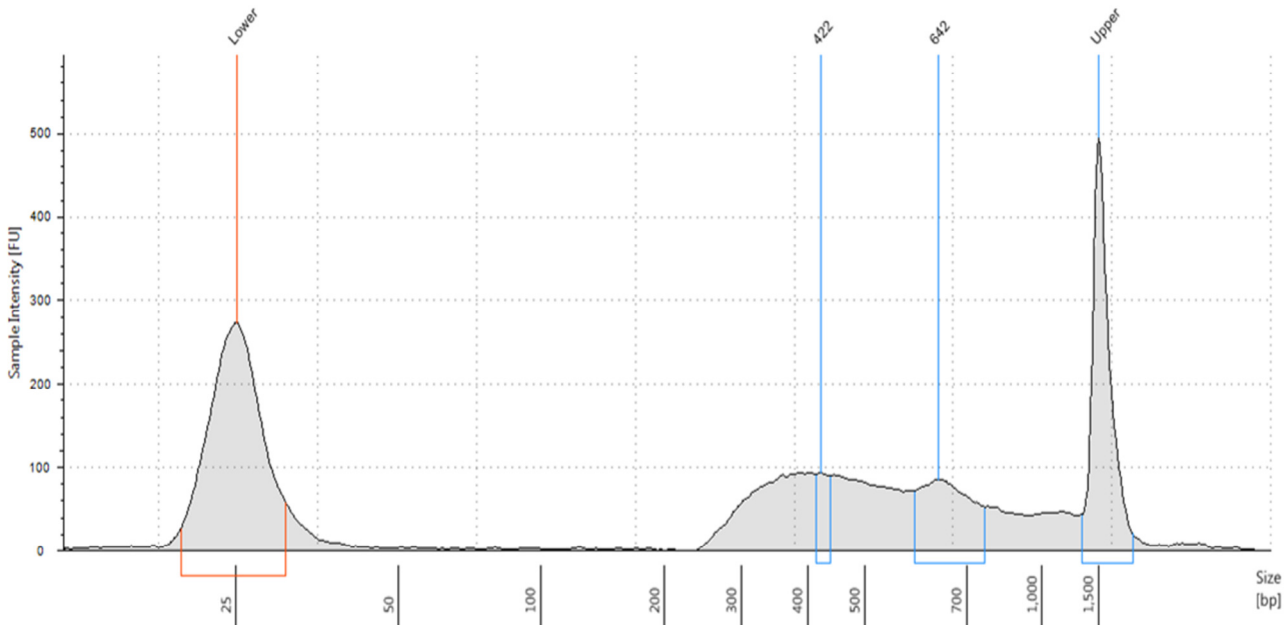


Figure 2.6 Bioanalyzer electropherogram of plant-fed aphid cDNA library

Generated cDNA library from isolated RNA from faba bean reared pea aphids analyzed on an Agilent 2100 Bioanalyzer.

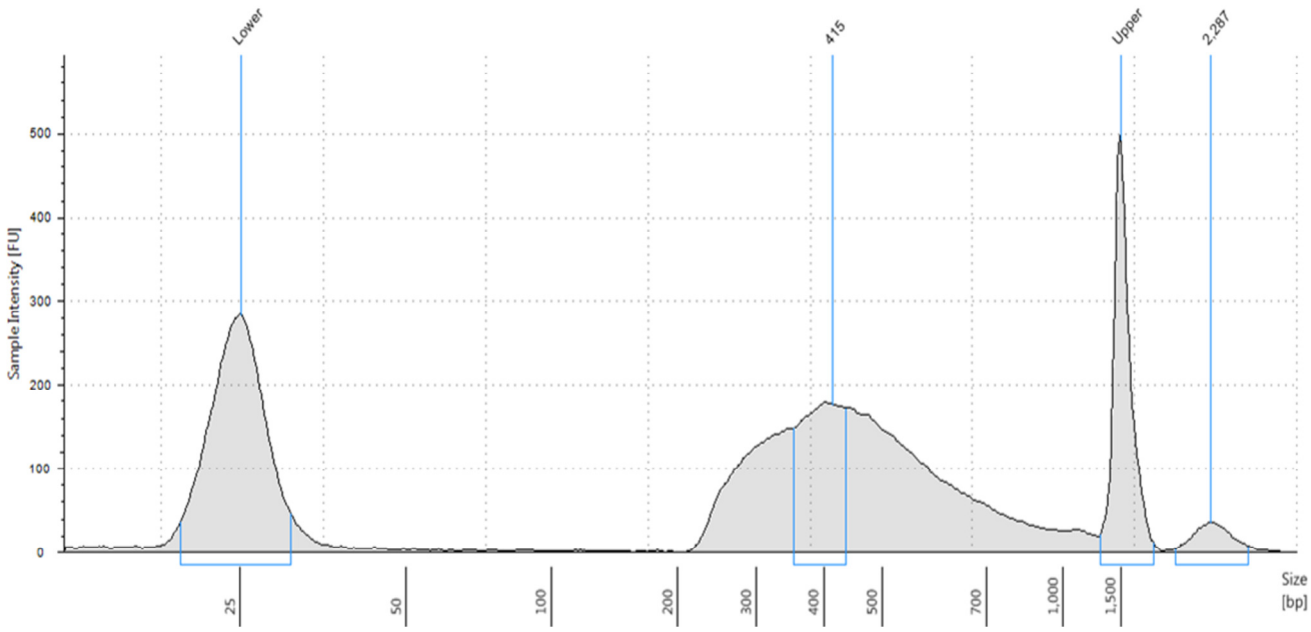


Figure 2.7 Bioanalyzer electropherogram of diet-fed aphid cDNA library

Generated cDNA library from isolated RNA from diet reared pea aphids analyzed on an Agilent 2100 Bioanalyzer.

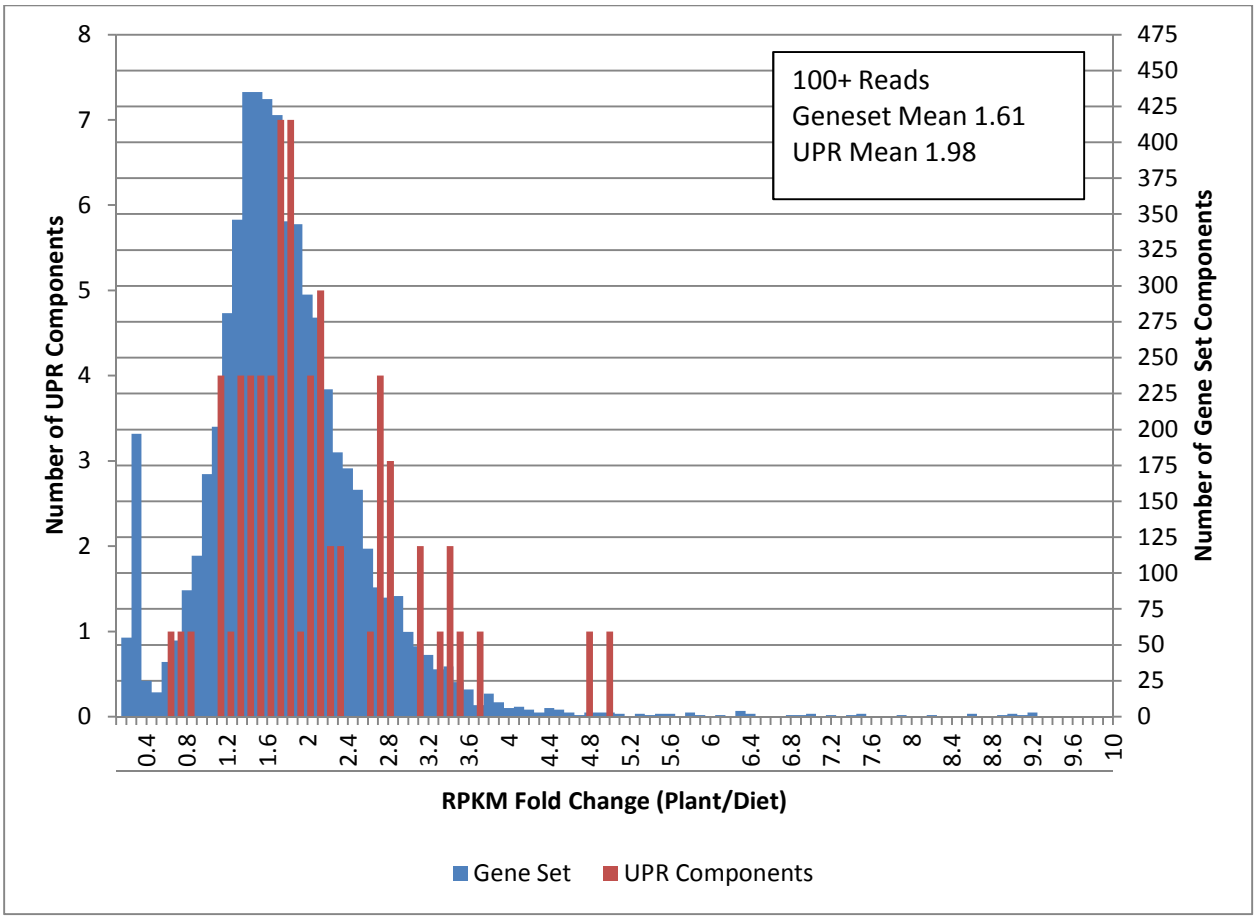


Figure 2.8 RPKM fold change ratio, plant versus diet feeding in the entire gene set and UPR components

RPKM fold change in both UPR components and the entire gene set with a minimum 100 read threshold for UPR components measured by RNA-seq.

Figure 2.9 Human and pea aphid nucleotide alignments of the proteins Armet and GRP78.

A. Nucleotide alignment of human and pea aphid GRP78/BiP with red boxes indicating location for RNAi effect by generation of dsRNA.
    - The first box identifies a segment of RNA with 5 identities and 14 non-identities out of 19
    - The second box identifies a segment of RNA with 8 identities and 11 non-identities out of 19
B. Nucleotide alignment of human and pea aphid Armet with red boxes indicating location for RNAi effect by generation of dsRNA
    - The first box identifies a segment of RNA with 3 similarities out of 19
    - The second box identifies a segment of RNA with 1 similarities out of 19

A.

B.

```
                    1         10        20        30        40        50        60        70
1. Acyrthosiphon pisum  CAAAATTAGTGATGTGTCTTTAAAATTTGTTAACGCTTGTCCGCTTCATTAAAATTAAAAAGTATTAATTGTCGTGA
2. Homo sapiens

                    80        90       100       110       120       130       140       150
1. Acyrthosiphon pisum  ACGACTTTTTAAAAAATACTAAAACAATCTTTTTT----------------TGTACTCAATCCCTATATTTTTCTA
2. Homo sapiens                                          AGCTACGGCGCGCGGCCGGGACTTGGAGGCGGTGCGGCGC-G
                   160       170       180       190       200       210       220       230
1. Acyrthosiphon pisum  TAATATACATATAAATAAATTTAGTTTATTTATAAACAAATAATACTTGA-CATAATTCTGAATATTGAGTTGTAA
2. Homo sapiens         GCGGGTGCGGTTCAGTCGGTCGGCGGCGGCAGCGGAGGAGGAGGAGGAGGAGGAGGATGAGGAGGA-TGAGGAG---
                   240       250       260       270       280       290       300
1. Acyrthosiphon pisum  AATTCTATTTGACAATGGATAAACACATATTGTTGGTATGTGTGTTTTTCATTGTATTTCATGTATTTCAAGCACAA
2. Homo sapiens         -----GATGTG-----GGCCACGCAGGGGCTGGCGGTGGCGCTGGCTCTGAGCGTGCTGCCGGGCAGCCGGGCGCT-
                   310       320       330       340       350       360       370       380
1. Acyrthosiphon pisum  TCGCGAACATTTACCGAAGAAGATTGTCCTGTTTGCGTACTTACAATTGAT-AAGTTTTCAAAAACACTTCAACGCG
2. Homo sapiens         --GCGGCCGGGCGACTGCGAAGTTTGTATTTCTTATCTGGGAAGATTTTACCAGGACCTCAAAGACA-GAGATGTCA
                   390       400       410       420       430       440       450       460
1. Acyrthosiphon pisum  AATTAAATCCAAAAAATATTGAAGAACAATTCAAAAAATATTGTTTGTCTACAAAGATTGATAAGGAAAAAAGATTG
2. Homo sapiens         CATTCTCACCAGCCACTATTGAAAACGAACTTATAAAGTTCTGCC-GGGAAGCAAGA--GGCAAAGAGAATCGGTTG
                   470       480       490       500       510       520       530
1. Acyrthosiphon pisum  TGTTACTATTTGGGCGGATTAGAAGATTCTGCAACTGGTATATTAAGTGAAATGTCAAAACCTTTATCGTGGTCCAT
2. Homo sapiens         TGCTACTATATCGGGCCACAGATGATGCAGCCACCAAAATCATCAATGAGGTATCAAAGCCTCTGGCCCACCACAT
                   540       550       560       570       580       590       600       610
1. Acyrthosiphon pisum  GCCGGCACTTAAAATCTGCGA-ACGCTTAAAGAAAATGGATGCTCAAGTCTGCGATATTAAATATGACAAAGAATC
2. Homo sapiens         CCCTGTGGAGAAGATCTGTGAGAAGCTTA-AGAAGAAGGACAGCCAGATATGTGAGCTTAAGTATGACAAGCAGATC
                   620       630       640       650       660       670       680       690
1. Acyrthosiphon pisum  GATTGGAAAACAGTCAACTTAAAGAAGATGAAAGTTAAAGATTTGAAGAAAATATTAGATAACTGGGGAGAGATTTG
2. Homo sapiens         GACCTGAGCACAGTGGACCTGAAGAAGCTCCGAGTTAAAGAGCTGAAGAAGATTCTGGATGACTGGGGGGAGACATG
                   700       710       720       730       740       750       760       770
1. Acyrthosiphon pisum  TGATGGTTGTTTAGAAAAGACAGATTATAT-CAAAAGAGTTGAAGAATTAAAACC---ATCTTACGTCAAGGAAG--
2. Homo sapiens         CAAAGGCTGTGCAGAAAAGTCTGACTACATCCGGAAGA-TAAATGAACTGATGCCTAAATATGCCCCAAGGCAGCC
                   780       790       800       810       820       830       840
1. Acyrthosiphon pisum  AGTTATAAATAATACAAATTATTATCCTACGCATTGTATAAATTTGCTAATTTTTTTAATACAATATTGTTTTTTGT
2. Homo sapiens         AGT----GCACGGACCGATTTGTAGTCTGCTCAATCTCT---GTTGCACCTGAGGGGGAAAAAACA---GTTCAACT
                   850       860       870       880       890       900       910       920
1. Acyrthosiphon pisum  ATTGTGTTCCATAA-----TTTATTTAAGATATAAAATAAA-GGGCTGAAGTTTAACAAAACC-AAACAAAATTGAA
2. Homo sapiens         GCTTACTCCCAAAACAGCCTTTTTGTAATTTATTTTTTAAGTGGGCTCC----TGACAATACTGTATCAGATGTGAA
                   930       940       950       960       970       980       990      1,000
1. Acyrthosiphon pisum  TAAAAAA
2. Homo sapiens         GCCTGGAGCTTTCCTGATGATGCTGGCCCTACAGTACCCCCATGAGGGGATTCCCTTCCTTCTGTTGCTGGTGTACT
                          1,010     1,020     1,030     1,040     1,050     1,060     1,070
1. Acyrthosiphon pisum
2. Homo sapiens         CTAGGACTTCAAAGTGTGTCTGGGATTTTTTTATTAAAGAAAAAAAATTTCTAGCTGTCCTTGCAGAATTATAGTGA
                          1,080     1,090     1,100     1,110     1,120     1,130     1,140 1,143
1. Acyrthosiphon pisum
2. Homo sapiens         ATACCAAAATGGGGTTTTGCCCCAGGAGGCTCCTAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
```

Chapter 3 - Saliva Protein Transcripts in the Pea Aphid

Literature Review:

Aphid saliva is pivotal to the feeding of aphids on host plants (Miles 1999; Tjallingii 2006). The aphid's salivary gland secretome and saliva proteome has been sought after by several laboratories (references given below under Proteomics and Transcriptomics sections). The salivary gland secretome is all proteins that are secreted from the salivary gland, whereas the saliva proteome is proteins that are found within saliva. The saliva proteome is a subset of the salivary gland secretome. Researchers have used two primary modes of research, proteomics and transcriptomics, to identify the salivary gland secretome and saliva proteome. Therefore this literature review is divided into these two approaches. Within each, papers are organized by date, but researchers who worked in the pea aphid are listed first.

Proteomics

Carolan et al. 2009 - Acyrthosiphon pisum (Pea Aphid)

In "The Secreted Salivary Proteome of the Pea Aphid *Acyrthosiphon pisum* Characterized by Mass Spectrometry," Carolan et al. (2009) identified a total of 9 proteins in pea aphid saliva using a proteomic GE-LC-MS/MS and LC-MS/MS approach. 40,000 aphids were allowed to feed on diet contained in Parafilm sachets. The 200 mL of diet was diluted to 250 mL with PBS. The diluted diet was concentrated and treated with a clean-up kit to prepare for gel electrophoresis. Using two analytical replicates of 6 pooled collections for SDS-PAGE, the proteins were visualized *via* silver nitrate based stains in semi-reducing conditions. Bands were excised from the gel and subjected to digestion with trypsin. The separated tryptic peptides were subjected to LC-MS/MS and utilizing a TurboSEQUEST algorithm in BioWorks v3.2 identifying transcripts in NCBI's non-redundant database of the pea aphid genome.

Carolan et al. 2011 - Acyrthosiphon pisum (Pea Aphid)

In Carolan et al. (2011), (Predicted Effector Molecules in the Salivary Secretome of the Pea Aphid (*Acyrthosiphon pisum*): a Dual Transcriptomic/Proteomic Approach), the authors utilized a parallel analysis of proteins and transcripts to identify a large list of salivary secretome members. For the proteomic analysis identifying 20 proteins, Carolan utilized aphid salivary

gland dissection and mirrored his previous publication in 2009 with one dimensional analysis, and added the use of two dimensional SDS-PAGE gels with MALDI-TOF mass spectroscopy coupled with MASCOT searches on salivary gland homogenates. Although the proteins identified were from salivary gland homogenates, the proteins reported were previously observed in saliva or to an effector that was secreted by other phytopathogenic organisms such as nematodes and fungi.

The transcriptomic approach was by Reeck's group and will be discussed in the transcriptomic section under Reeck 2011.

Harmel et al. 2008 - Myzus persicae (Green Peach Aphid)

In Harmel et al. (2008), "Identification of Aphid Salivary Proteins: a Proteomic Investigation of *Myzus persicae*," the authors identified 9 proteins and reported the putative orthologs in the pea aphid by accession number, two of which, maltase 2-like (ACYPI009042) and 3-hydroxyacyl-CoA dehydrogenase type-2 (ACYPI56654) were newly identified saliva components. Researchers' primary use of proteomics to identify these components came by either directly in-solution digesting or utilizing a two dimensional SDS-PAGE before trypsin digestion coupled with mass spectroscopy. The use of silver staining allowed visualization of the bands to be digested and analyzed by first excision, destaining, and digestion. After digestion and subsequent LC MS/MS analysis, the Mascot search engine was used to obtain peptide sequences. BLAST searches through pea aphid ESTs allowed the identification of the aforementioned saliva proteins.

Cooper et al. 2010 - Diuraphis noxia (Russian Wheat Aphid)

In Cooper et al. (2010), "Salivary Protein of Russian Wheat Aphid (Hemiptera: Aphididae,)" the authors identified four saliva proteins which were identified as the putative orthologs of pea aphid RNA Helicase (ACYPI007670), pathogen-associated molecular pattern (PAMP) (ACYPI005439), Zinc binding dehydrogenase (ACYPI009182), and Unknown Protein 23 (ACYPI005882). They used three different diets consisting of a pure water diet, amino acid diet, and a sucrose diet. Approximately 450 aphids were placed in each plate to collect saliva. Stylet sheaths remained in the parafilm after rinsing. 25 Plates were pooled which total 11000-11500 aphids and the pooled collections were concentrated. For analysis, both one dimensional

SDS-mass spectrometry analysis and two dimensional SDS-mass spectrometry analysis were used in conjunction with Bradford Assays.

Using three diet compositions, including a pure water diet, a 15% sucrose diet, and an amino acid diet (100 mM serine, 100 mM methionine, 100 mM aspartic acid) in 15% sucrose. There was a significant difference in protein amount produced on each diet as indicated below.

Pure water diet: 0.052 +/- 0.02 ng/aphid

Sucrose diet: 0.66 +/- 0.09 ng/aphid

Amino acid diet: 0.14 +/- 0.01 ng/aphid

Alkaline phosphatase activity was only detectable in aphid probed sucrose diets but not water, amino acid, or control diets. In gel digests followed by mass spectroscopy and subsequent analysis using MASCOT software led to the determination from fragments of proteins of the four protein's amino acid sequences in the Russian wheat aphid which led to BLAST analysis against the pea aphid EST database, in turn identifying the pea aphid putative orthologs.

Rao et al. 2013 - Sitobion avenae & Metopolophium dirhodum (Grain Aphids)

In Rao et al. (2013), "Proteomic Profiling of Cereal Aphid Saliva Reveals Both Ubiquitous and Adaptive Secreted Proteins," the saliva identified proteins β-galactosidase precursor (ACYPI007650), actin-related protein 3-like (ACYPI000064), unannotated protein (ACYPI000113), and glucose dehydrogenase (ACYPI005582) were identified in saliva from *S. avenae* and *M. dirhodum.* 40,000 aphids' saliva was collected by pooling protein concentrates from 50 diet preparations. Non protein contaminants were removed from the final concentrate using a two dimensional clean-up kit followed by one dimensional SDS-PAGE and visualized with silver staining. Visible protein bands were excised and digested overnight with trypsin, and processed with LC MS/MS. Using the TurboSEQUEST algorithm in BioWorks v3.2 protein sequences were derived which allowed BLAST searches to correlate with pea aphid proteins and subsequent identification of the accession numbers. Of the proteins identified, the four listed above were the only proteins that had not been previously identified in other studies.

Transcriptomics

Reeck et al. in Carolan et al. 2011 - Acyrthosiphon pisum (Pea Aphid)

Utilizing Sanger sequencing, BLAST2GO, and an R-statistic for salivary gland enrichment, transcripts were identified that were enriched in salivary glands and encoded proteins with a signal secretion sequence. The R-statistic is a method developed by Stekel et al. (2000) to identify EST contigs that are abundant in individual tissues, in this case the salivary gland. Transcripts which contain an R-statistic value greater than 7 identify the transcript as enriched in salivary glands. Following BLASTx searches, ACYPI accession numbers were identified for each transcript. In total, 42 components were reported in Carolan et al. (2011) and were considered to be saliva proteome.

Ramsey et al. 2007 - Myzus persicae (Green Peach Aphid)

In Ramsey et al. (2007), "Genomic Resources for *Myzus persicae:* EST sequencing, SNP Identification, and Microarray Design," the authors sequenced from 16 *M. persicae* cDNA libraries to generate 26,669 expressed sequence tags (ESTs). Of those ESTs 3233 were from the salivary gland library, encoding 2242 unigenes of the green peach aphid salivary gland.

To identify proteins of saliva, sequences expressed in the salivary glands that were predicted to have signal peptides were more closely examined. Of the 45 such proteins derived from salivary glands, 15 proteins were predicted to contain an anchor sequence, and therefore excluded from the list of possible saliva proteins. The authors thus proposed 30 proteins of the saliva proteome.

Bos et al. 2010 - Myzus persicae (Green Peach Aphid)

In Bos et al. (2010), "A Functional Genomics Approach Identifies Candidate Effectors from the Aphid Species *Myzus persicae* (Green Peach Aphid)," a pipeline to identify candidate effectors is approached. One such protein, newly identified, was Mp42 (ACYPI010222) using the following methodology. A 5919 EST library from *M. persicae* salivary glands was processed through gene annotation software yielding 3233 protein coding sequences. These sequences were subjected to signal peptide prediction (SignalP3.0) reducing the proteins to 304 which contained a signal peptide. Of those containing signal peptides, a blastp analysis was performed and it removed redundant sequences bringing the total proteins containing a signal peptide to 134. Those peptides were checked for the presence of transmembrane domains (TMHMM v2.0) reducing the results down to 115 predicted secreted proteins. After full length sequences were identified and the presence of polymorphisms within *M. persicae* and other species were

analyzed, the similarities of predicted secreted *A. pisum* salivary gland proteins were compared identifying homologs yielding a pool of 46 candidate effectors. Of those 46 candidates, two were not found in the *A. pisum* indicating that they are unique to *M. persicae.* Multiple candidates were removed because they were exclusive to head dissections, leaving the remaining candidates *Mp1, Mp2, Mp10, Mp30, Mp42, Mp47, Mp50,* and *MpC002*. Of those candidates, *Mp42* was the only orthologous component that had not already been discovered in other laboratories.

Atamian et al. 2012 - Macrosiphum euphorbiae (Potato Aphid)

In Atamian et al. (2012), "In Planta Expression or Delivery of Potato Aphid *Macrosiphum euphorbiae* Effectors Me10 and Me23 Enhances Aphid Fecundity," interactions between two candidate effectors were found to affect aphid fecundity when overexpressed in the host plant. The authors identify 5 pea aphid putative orthologs from saliva. This line of research establishes another important aspect that could be useful when targeting UPR components in the pea aphid by dsRNA. Both effectors Me10 and Me23 were shown to increase fecundity suggesting that they possess the ability to suppress the host plant defenses on the feeding aphid and both were identified as glutathione peroxidases. While the research focused on Me10 and Me23 due to their ability to elucidate changes in aphid fecundity, Me13, Me17, Me14, Me20, and Me25 had not been previously identified as salivary gland proteins. To achieve these putative orthologs, 200 *M. euphorbiae* salivary glands were dissected and evaluated with RNA-seq. With reciprocal TBLASTX analysis, 551 *M. euphorbiae* contigs were identified with sequences orthologous to 460 *A. pisum* transcripts. Of those, signal peptide prediction (SignalP4.0) reduced the number to 125 and of which they were further reduced due to being previously identified by Carolan et al. (2011). There were components that were unable to be identified and Atamian et al. attribute that to gaps in the sequencing of the salivary gland.

Materials and Methods:

Dissections

Aphids head dissections were from wingless, asexual pea aphids from the clone LSR1 line. Previously indicated in Chapter 2, aphids were allowed to feed in the diet-fed and plant-fed states.

To test whether transcripts were found in salivary glands of the pea aphid, dissections of salivary glands were also performed. Salivary gland dissections were completed in the following manner by Dr. Raman Chandrasekar. Prior to dissection, the bench, dissection slides, and gloved hands of the researcher were cleaned with a solution of 0.1% DEPC treated water, followed by application of RNaseZap (Sigma-Aldrich #R2020). After all surfaces were allowed to dry, 100-200 uL of RNALater was placed on top of dissection slide. The insect was again placed in the RNAlater on the surface of the slide; grasping the aphid with forceps at the abdomen, the antennae were removed with a small 22ga needle. To remove the exoskeleton from the head of the aphid, a bent 22ga needle was used. With the exoskeleton removed, both pairs of primary and accessory salivary glands were exposed. The isolated pairs of primary and accessory salivary glands from 120 aphids were removed and placed in RNase/DNase-free centrifuge tubes containing 50 µL of RNAlater. Approximately 30 aphids were dissected in the solution of RNAlater before switching to another new slide. Collection of salivary glands took place over 3 or more hours and aphids were placed into sterile petri dishes beside the microscope and not removed directly from their respective feeding state.

RNA Isolation and Sequencing of Heads on Plant & Diet-fed States:

After feeding and dissection, RNA isolations described in chapter two were used to purify RNA. RNA-seq using the same platform as previously indicated through the Integrated Genomics Facility was then utilized to measure and quantify the transcriptional levels of UPR components in both fed states.

Results:

Overall the goal of this study was to identify components predicted to be in pea aphid saliva utilizing all available proteomic and transcriptomic data from other aphid species. It is important to note that annotations from previous researchers were carried over in this study. The term “unannotated protein1” is ambiguous and would be better stated as “unknown protein.” However, at this point any manipulation of the annotations would remove the ties to the originating research for each component. This study identifies those putative orthologs by AphidBase identification number, and a compiled list from the proteomic approaches is outlined in Table 8 and transcriptomic approaches in Table 9. With duplications removed, Table 10 shows the entire list from both approaches.

RNA isolated from isolated heads or salivary glands of plant-fed and diet-fed insects was isolated and submitted to the Integrated Genomics Facility at Kansas State University (IGF-KSU) for quality analysis by Agilent 2100 Bioanalyzer. The bioanalyzer profiles for plant-fed and Akey-Beck diet-fed states are shown in Figure 2.3 and Figure 2.4, and showed good quality RNA, suitable for RNA-seq cDNA library synthesis. The cDNA library synthesis bioanalyzer profiles are shown in Figure 2.5 and Figure 2.6.

RPKM values on RNA from salivary gland dissections establishing that the transcripts were indeed found within the salivary gland are reported in Table 11. Reported in Table 12 are the RPKM values and RPKM plant-fed to diet-fed ratios from head dissections.

Signal secretion peptides are short N-terminal peptides present in the majority of newly synthesized proteins that are destined towards the secretory pathway. According to statistics available at the SignalP 3.0 website, the average eukaryotic signal peptide is 22 amino acids in length. Anchor peptides function as one would assume, anchoring the newly produced protein to a membrane. The encoded proteins for all studied transcripts were analyzed to predict the presence of a signal secretion peptide, anchor, and ER retention signal. All sequences were processed through the SignalP 3.0 server (http://www.cbs.dtu.dk/services/SignalP-3.0/), and the probabilities that a signal peptide or anchor existed were calculated through the Hidden Markov model within the program. These results indicated in Table 13, show the accession number, description, SignalP3.0 result with cleavage site, and predicted anchor probability, as well as ER retention signals for the encoded proteins. All transcripts are shown to encode a signal peptide which would indicate probable secretion into saliva or other extracellular fluids such as hemolymph.

Analysis of means of RPKM fold change for plant versus diet feeding for the entire pea aphid gene set and salivary components was completed with a T-test. The mean RPKM fold change value for the entire gene set was 1.61 while the salivary components mean was 1.95. A student's unpaired T test was utilized to compare the two means and at a 95% confidence interval and the difference was significant using the following data. A plot representing the gene set and salivary component data is shown in Figure 3.1.

	Gene Set	Salivary Components
Mean	1.6100	1.9500
SD	0.9085	0.8824

SEM	0.0062	0.0802
N	21501	121

Unpaired t test results

P value and statistical significance:

The two-tailed P value equals 0.0001

By conventional criteria, this difference is considered to be extremely statistically significant indicating that there is a difference between the means.

Confidence interval:

The mean of Gene Set minus UPR Components equals -0.34

95% confidence interval of this difference: From -0.5 to -0.2

Discussion:

Multifaceted Approach to Identify Salivary gland Secretome Proteins

Work from enzyme assays to the more recent proteomic and transcriptomic approaches have been used to identify secretome components, although no one lab has used all methods and available data coupled with RNA-seq data to attempt to define the full saliva proteome of an aphid. Here I attempt to complete the saliva proteome for the pea aphid. It is important to note that although the proteins represented here are assumed to be in saliva that may not be the case. The only components that can be definitively shown as saliva proteins are those that were studied by analysis of saliva or as in the case of protein C002 and Armet, have been found in plants after aphid feeding. In other words, this work does identify some proteins of saliva, but may be better portrayed as a secretome of the salivary gland, where most of the studied transcripts are likely to be in saliva.

Throughout this chapter one important realization is that this work is a compilation of many researchers that work in different species. By utilizing the sequence data that each researcher has procured whether by proteomic or transcriptomic means, I have been able to build a comprehensive saliva proteome in the pea aphid.

Analysis of Saliva Proteins

After reads were mapped, with the Geneious software suite, RPKM values for each transcript were calculated. These RPKM values were used to generate ratios of plant-fed to diet-

fed RPKM which indicate trends in up and down regulated transcripts. The results of these analyses are found in Table 12. As in Chapter 2, the majority of the transcripts studied had higher expression in plant-fed salivary gland libraries. The range of the fold change was 5.970 to 0.209 in head isolations. The five highest fold changes were for ACYPI54712 (unknown protein), ACYPI009182 (Zinc binding dehydrogenase), ACYPI009625 (EMP24), ACYPI56654 (3-hydroxyacyl-CoA dehydrogenase type-2), and ACYPI007677 (Calreticulin) proteins in head isolations.

In Feng et al. (2014), "Characterization of an Aphid-specific, Cysteine-rich Protein Enriched in Salivary Glands," a pea aphid transcript was studied which was first identified by Carolan et al. (2011). Reported as enriched in salivary glands, the protein was not included in the Carolan et al. (2011) studies as a saliva protein but as a possible member of the salivary gland secretome. This component, identified as the aphid specific cysteine rich protein (ACYPI39568) was analyzed using both a proteomic and transcriptomic approach. The transcript's message was expressed, the protein was purified, and antibodies were also synthesized. The protein sequence was analyzed *via* SignalP3.0 and was found to contain a signal peptide. Immunohistochemistry *via* the antibodies produced were able to locate the high expression of the transcript/protein in the salivary glands. Double stranded RNA for ACYPI39568 was produced for feeding and injection studies in live aphids. It was discovered that aphids had increased transcript levels when feeding on plants than when feeding on an artificial diet. Interestingly the interference of ACYPI39568 expression did not affect the survival rate of aphids on plants.

In Feng et al. (2012), "Polymorphisms in Salivary-gland Transcripts of Russian Wheat Aphid Biotypes 1 and 2," saliva secretome components were analyzed. Although those components were first identified by Carolan et al. (2011), further analyses including non-synonymous and synonymous mutations were analyzed. Of the 17 sequences, 2 were not able to be annotated. For four transcripts (those encoding a coated-vesicle membrane protein, a peroxidase and the two non-annotatable proteins) there was no polymorphism detected. The other 13 transcripts all had observable polymorphisms between the biotypes at the nucleotide and predicted protein level.

Armet as a Saliva Protein

One of the transcripts that has been studied as a saliva component is Armet, an aphid ortholog addressed in Chapter 1, represented at 2.54 fold higher in plant-fed head libraries in comparison to diet-fed libraries. Armet is secreted into plants during pea aphid feeding (Wang et al., 2015).

Statistical Comparison of Salivary Secretome Components versus the Entire Gene Set

Utilizing a t-test, the comparison of RPKM fold change means from the gene set (1.61) and UPR components (1.95) show that the difference is statistically significant at a 95% confidence. The importance of this comparison of means indicates that the salivary secretome component's expression is different from the overall population of the gene set. The increased mean value of RPKM fold change in salivary components confirms the hypothesis that proteins of saliva are upregulated during plant feeding in aphids.

After close inspection of the top 20 RPKM fold change values in the gene set, an interesting component which appears to be an alternative gene to a component in the saliva component list was discovered. This component, β-galactosidase, putatively identified to a human ortholog of β-galactosidase, was identified by AphidBase ID ACYPI001373. In comparison to the component identified as β-galactosidase (ACYPI007650) in this dissertation as a saliva protein, ACYPI007650 and ACYPI001373 showed a 54.49% identity at the protein level.

Diet-fed Upregulation of Some Saliva Proteins

A total of 15 transcripts were found to have a RPKM fold change ratio under one indicating higher expression in diet-fed aphid heads than in plant-fed heads. The values of RPKM ratio for those 15 components range from 0.209 to 0.985. Those transcripts which expressed RPKM ratios less than one are Unannotated Protein 2, Unannotated Protein 3, M1 zinc metalloprotease, Dipeptidyl carboxypeptidase, Peptidase M1, MCO1, Unknown protein 11, Glucose Dehydrogenase, CLIP-domain serine protease, Zinc-dependent Phospholipase C, Unknown protein 34, Me25, Cadherin, Maltase 2-like, and AHNAK nucleoprotein (desmoyokin).

Protein Disulfide Isomerases

Another interesting point is the presence of protein disulfide isomerases. As reported in Chapter 2, each PDI transcript found, had a higher RPKM in RNA from plant feeding versus diet feeding giving support to the idea that the UPR is upregulated during plant feeding due to increased secretion of salivary proteins. While it doesn't give insight as to the importance or presence of PDIs in saliva, it does show that they are found in salivary gland tissues.

Expectations of Results and Generation of the Largest Saliva Proteome

Data corroborates the expectation that nearly all transcripts would upregulate in plant-fed states due to the fact that aphid feeding on plants is a much more complex method than their feeding on artificial diets. Whereas plants have numerous defense mechanisms that protect them from invading pathogens or insect pests, many pea aphid saliva proteins may help circumvent these systems. As for feeding, it is logical that the fold change is higher in plant-fed states versus diet feeding on many levels. Initially one can assume that the complexity of feeding on another living organism with defense mechanisms versus a petri dish is paramount. Although some components have been found to be inducible under the plant-fed state, it isn't difficult to attribute that increase to mitigation of plant defenses or the digestion of complex nutrients.

One question that may not be answered is why 15 components are not upregulated during plant feeding. While one may assume that all transcripts should be upregulated in plant feeding due to mitigation of plant defenses or the digestion of complex nutrients, there may be instances where some components are not needed for plant feeding versus diet feeding. An example of this may be seen in the component glucose dehydrogenase (ACYPI000986). It is possible that the sugar concentration, 0.5 mM, in the Akey & Beck diet is higher than the physiological sugar levels in the host plant accounting for the increased transcript level in diet fed aphids. This one example is an idea that does not entirely answer the above question, but it may give way for other lines of thought on the diet fed upregulation of the 15 components.

Identification of this "master list" of pea aphid saliva proteins will no doubt continue. My list is by no means a final list of the proteins that encompass saliva. It is also a good starting point for other aphid species. For instance the Russian wheat aphid would be another aphid which is easy to maintain in laboratory settings and although the size of the insect is much

smaller than the pea aphid, it would be a good target species for pest mitigation due to its agricultural impact worldwide.

Obviously the goal outlined in Chapter 2 to produce genetically modified crops is also viable with this set of saliva transcripts. As previously indicated in Chapter 2, the most informative yet the least economical approach would be to test fecundity and lifespan knockdown studies with each identified component of this saliva proteome.

The production of a crop targeted at either UPR or saliva components that reduces aphid fecundity or increases mortality would be an agricultural benefit. It is unclear which system, the UPR or saliva proteins will provide the best set of targets for knockdown studies. It may be a combination of both because the UPR is upregulated during the production of the salivary proteins in plant feeding. Ultimately this chapter has laid the groundwork for a comprehensive saliva proteome in a model aphid species which may allow identification of these components and other species.

Carolan 2009 (pea aphid)
ACYPI000733 – Dipeptidylcarboxy peptidase
ACYPI008911 – Dipeptidylcarboxy peptidase
ACYPI009427 – M1 Zinc metalloprotease
ACYPI010198 – Unknown protein
ACYPI007868 – Unknown protein
ACYPI000113 – Glucose dehydrogenase [FAD quinone]-like
ACYPI003308 – Regucalcin-like
ACYPI009881 – Unknown protein
ACYPI005582 – Centrosomal protein of 104kDa

Harmel 2008 (green peach aphid)
ACYPI000986 – Glucose dehydrogenase
ACYPI56654 – 3-Hydroxyacyl-CoA dehydrogenase type-2
ACYPI009042 – Alpha-amylase

Cooper 2010 (Russian wheat aphid)
ACYPI005766 – unknown protein
ACYPI009182 – Zinc binding dehydrogenase
ACYPI005439 – Phosphatase activator protein phosphatase 2A activator
ACYPI007670 – RNA helicase

Rao 2013 (grain aphid)
ACYPI000113 – Glucose dehydrogenase
ACYPI000288 – Glucose dehydrogenase
ACYPI000817 – Peroxidase
ACYPI002298 – Trehalase
ACYPI23752 – Carbonic anhydrase
ACYPI007650 – Beta-galactosidase precursor
ACYPI001857 – Yellow e-3 like protein
ACYPI000064 – Actin

Table 8 List of salivary proteins identified by proteomics

Organized by source and identified by AphidBase accession number, transcripts found by ortholog search in the pea aphid with confirmed in the salivary gland tissue. Duplications removed between multiple researchers with the component that was duplicated placed with the researcher who first identified the component.

Reeck et al. 2011 (pea aphid)
ACYPI000472 – unknown protein
ACYPI000490 – unknown protein
ACYPI000852 – unknown protein
ACYPI001152 – unknown protein
ACYPI001099 – unknown protein
ACYPI001541 – unknown protein
ACYPI001843 – unknown protein
ACYPI002891 – Cadherin
ACYPI003695 – unknown protein
ACYPI007553 – unknown protein
ACYPI008567 – unknown protein
ACYPI009427 – M1 zinc metalloprotease
ACYPI38795 – unknown protein
ACYPI43360 – unknown protein
ACYPI45001 – unknown protein
ACYPI56502 – unknown protein
ACYPI001271 – unknown protein
ACYPI001606 – unknown protein
ACYPI001887 – unknown protein
ACYPI002172 – unknown protein
ACYPI002476 – Inositol monophosphatase
ACYPI004198 – Lipophorin precursor
ACYPI007406 – unknown protein
ACYPI009881 – unknown protein
ACYPI39568 – Aphid specific cysteine rich protein
ACYPI008617 – C002
ACYPI002439 – Glutathione peroxidase(ApGPx1)
ACYPI008001 – Armet
ACYPI000733 – Dipeptidylcarboxy peptidase
ACYPI008224 – unknown protein
ACYPI003917 – SCP_GAPR-1 like
ACYPI000228 – glucose dehydrogenase
ACYPI000986 – glucose dehydrogenase
ACYPI49603 – unknown protein
ACYPI000223 – unknown protein

Feng et al. 2012
ACYPI52702 – Cathepsin B
ACYPI006974 – Cathepsin L
ACYPI008182 – Juvenile Hormone Binding Protein Homolog
ACYPI082770 – MCO1 (Laccase)
ACYPI007300 – Endoribonuclease
ACYPI000002 – Sucrase
ACYPI007065 – Contig_37
ACYPI071951 – Peptidase M1

ACYPI001523 – Chorin Peroxidase H6
ACYPI002497 – Transmembrane 876-like
ACYPI008911 – Dipeptidyl carboypeptidase
ACYPI009625 – EMP24-like
ACYPI008617 – C002
ACYPI002298 – Trehalase

Reeck et al. 2014 (pea aphid: 8 unique)
ACYPI38240 – ApGPx2
ACYPI000768 – Maltase-A1
ACYPI088277 – unknown protein
ACYPI005041 – unknown protein
ACYPI28317 – unknown protein
ACYPI063417 – unknown protein
ACYPI081664 – unknown protein

Bos et al. 2010 (green peach aphid)
ACYPI006346 – Mp1
ACYPI001774 – Mp2
ACYPI000097 – Mp10
ACYPI001610 – Mp30
ACYPI010222 – Mp42
ACYPI005818 – Mp47
ACYPI23752 – Mp50 (Carbonic anhydrase II)

Atamian et al. 2012
ACYPI56566 – Me13
ACYPI21663 – Me14
ACYPI53825 – Me17
ACYPI21412 – Me20
ACYPI006300 – Me25

Ramsey et al. 2007
ACYPI55147 – unknown protein
ACYPI001719 – unknown protein
ACYPI003601 – unknown protein
ACYPI009919 – chromatin STP2
ACYPI000558 – Rab GTPase domain containing protein
ACYPI56620 – cuticular protein
ACYPI54712 – unknown protein
ACYPI51013 – unknown protein
ACYPI48849 – unknown protein
ACYPI48356 – unknown protein
ACYPI46095 – unknown protein
ACYPI45769 – yellow-g2
ACYPI45597 – unknown protein

ACYPI42782 – similar to Cg9849
ACYPI073649 – AHNAK
ACYPI089376 – CG2839
ACYPI004591 – chromatin STP2
ACYPI080546 – Glutathione S trans.
ACYPI26959 – peroxidase
ACYPI004866 – similar to CG11699
ACYPI006775 – similar to CG2471
ACYPI010168 – similar to CG5861
ACYPI003247 – similar to CG6583
ACYPI001706 – Der1-like domain
ACYPI007387 – ring finger-like 185
ACYPI002976 – tetraspanin 29FA
ACYPI24281 – unknown protein
ACYPI22506 – unknown protein
ACYPI080156 – unknown protein
ACYPI010151 – unknown protein
ACYPI007022 – unknown protein
ACYPI006124 – unknown protein
ACYPI005838 – unknown protein
ACYPI004394 – unknown protein
ACYPI003780 – unknown protein
ACYPI003602 – unknown protein
ACYPI003327 – unknown protein
ACYPI001445 – unknown protein
ACYPI000797 – unknown protein
ACYPI000707 – unknown protein
ACYPI000422 – unknown protein
ACYPI071317 – Zinc-dependant phospholipase C
ACYPI008945 – unknown protein
ACYPI009585 – unknown protein

Carolan et al. 2011
ACYPI000119 – Disulfide isomerase
ACYPI000288 – Glucose dehydrogenase
ACYPI002258 – M1 zinc metalloprotease
ACYPI002298 – Trehalase
ACYPI002622 – Calreticulin
ACYPI003917 – SCP GAPR-1
ACYPI007677 - Calreticulin
ACYPI008001 – Armet
ACYPI008370 – CLIP-domain serine protease
ACYPI39568 – Aphid specific cysteine rich protein

Table 9 List of salivary proteins identified by transcriptomics

Organized by source and identified by AphidBase accession number, transcripts found by putative ortholog search in the pea aphid and confirmed in the salivary gland tissue. Duplications removed between multiple researchers with the component that was duplicated placed with the researcher who first identified the component.

Carolan 2009
LOC100575164 – Dipeptidylcarboxy peptidase (Reeck 2014)
ACYPI008911 – Dipeptidylcarboxy peptidase
ACYPI009427 – M1 Zinc metalloprotease
ACYPI010198 – Aminopeptidase N-like
ACYPI007868 – Aminopeptidase N-like
ACYPI000113 – Glucose dehydrogenase[FAD quinone]-like
ACYPI003308 – Regucalcin-like
ACYPI009881 – Putative Sheath Protein(Carolan 2011)

Harmel 2008
ACYPI000986 – Glucose dehydrogenase
ACYPI56654 – 3-Hydroxyacyl-CoA dehydrogenase type-2
ACYPI009042 – Maltase 2-like

Cooper 2010
ACYPI005766 – Centrosomal protein of 104kDa
ACYPI005882 – Glucose dehydrogenase
ACYPI009182 – Zinc binding dehydrogenase
ACYPI005439 – Serine/threonine-protein phosphatase 2A activator-like
ACYPI007670 – RNA helicase

Rao 2013
ACYPI007650 – Beta-galactosidase precursor
ACYPI000064 – Actin

Reeck et al. 2011
ACYPI000472 – unknown protein
ACYPI000490 – unknown protein
ACYPI000558 – unknown protein
ACYPI000852 – unknown protein
ACYPI001152 – unknown protein
ACYPI001099 – unknown protein
ACYPI001541 – unknown protein
ACYPI001843 – unknown protein
ACYPI002891 – Cadherin
ACYPI003695 – unknown protein
ACYPI005818 – unknown protein
ACYPI007553 – unknown protein
ACYPI008667 – unknown protein
ACYPI009427 – M1 zinc metalloprotease
ACYPI009919 – unknown protein
ACYPI38795 – unknown protein
ACYPI43360 – unknown protein
ACYPI45001 – unknown protein
ACYPI56502 – unknown protein
ACYPI001271 – unknown protein
ACYPI001606 – unknown protein
ACYPI001887 – unknown protein
ACYPI002172 – unknown protein
ACYPI002476 – Inositol monophosphatase
ACYPI004198 – Lipophorin precursor
ACYPI006345 – unknown protein
ACYPI007406 – unknown protein
ACYPI009881 – Putative sheath protein
ACYPI55147 – unknown protein
ACYPI39568 – Aphid specific cysteine rich protein
ACYPI003601 – unknown protein
ACYPI008617 – C002
ACYPI002489 – Glutathione peroxidase(ApGPx1)

Feng et al. 2012
ACYPI52702 – Cathepsin B
ACYPI006874 – Cathepsin L
ACYPI008182 – Juvenile Hormone Binding Protein Homolog
ACYPI082770 – MC01 (Laccase)
ACYPI007300 – Endoribonuclease
ACYPI000002 – Sucrase
ACYPI007065 – Contig_37
ACYPI071951 – Peptidase M1
ACYPI001523 – Chorin Peroxidase H6
ACYPI002497 – Transmembrane 87B-like
ACYPI008911 – Dipeptidyl carboypeptidase
ACYPI009625 – EMP24-like

Reeck et al. 2014
ACYPI38240 – ApGPx2
ACYPI000768 – Maltase-A1
ACYPI089277 – unknown protein
ACYPI005041 – unknown protein
ACYPI28317 – unknown protein
ACYPI009585 – unknown protein
ACYPI063417 – unknown protein
ACYPI031664 – unknown protein

Ramsey et al. 2007
ACYPI001719 – unknown protein
ACYPI56520 – cuticular protein
ACYPI54712 – unknown protein
ACYPI51013 – unknown protein
ACYPI48849 – unknown protein
ACYPI48356 – unknown protein
ACYPI46095 – unknown protein
ACYPI45769 – yellow-g2
ACYPI45597 – unknown protein
ACYPI42782 – similar to Cg9849
ACYPI073648 – AHNAK
ACYPI089376 – CG2839
ACYPI004591 – chromatin STP2
ACYPI000546 – Glutathione S trans.
ACYPI28959 – peroxidase
ACYPI004866 – similar to CG11699
ACYPI006775 – similar to CG2471
ACYPI010168 – similar to CG5881
ACYPI008247 – similar to CG6593
ACYPI001706 – Der1-like domain
ACYPI007387 – ring finger-like 185
ACYPI002976 – tetraspanin 29FA
ACYPI24281 – unknown protein
ACYPI22506 – unknown protein
ACYPI080156 – unknown protein
ACYPI010151 – unknown protein
ACYPI007022 – unknown protein
ACYPI006124 – unknown protein
ACYPI00583 – unknown protein
ACYPI004394 – unknown protein
ACYPI003780 – unknown protein
ACYPI009602 – unknown protein
ACYPI003327 – unknown protein
ACYPI001445 – unknown protein
ACYPI000797 – unknown protein
ACYPI000707 – unknown protein
ACYPI000422 – unknown protein
ACYPI071317 – Zinc-dependant phospholipase C

Bos et al. 2010
ACYPI23752 – Carbonic anhydrase II
ACYPI010222 – Mp42

Atamian et al. 2012
ACYPI56566 – Me13
ACYPI21663 – Me14
ACYPI53825 – Me17
ACYPI21412 – Me20
ACYPI006300 – Me25

Carolan et al. 2011
ACYPI000119 – Disulfide isomerase
ACYPI000288 – Glucose dehydrogenase
ACYPI002258 – M1 zinc metalloprotease
ACYPI002298 – Trehalase
ACYPI002622 – Calreticulin
ACYPI008917 – SCP GAPR-1
ACYPI007677 - Calreticulin
ACYPI008001 – Armet
ACYPI008370 – CLIP-domain serine protease
ACYPI39568 – Aphid specific cysteine rich protein

Table 10 List of salivary proteins identified by proteomics and transcriptomics

Organized by source and identified by AphidBase accession number, transcripts found by putative ortholog search in the pea aphid with confirmed in the salivary gland tissue. Duplications removed between proteomic and transcriptomic approaches with the component that was duplicated placed with the researcher who first identified the component.

Table 11 Verification of transcripts by salivary gland dissection

Reads generated by RNA-seq were mapped to each individual transcript open reading frame as a "reference genome". The number of salivary gland reads and RPKM values, calculated as described in the text are given for each transcript.

A color key indicated below identify transcript names as follows: purple (with white text): pea aphid transcripts corresponding to transcripts studied in Russian wheat aphids (Cui et al., 2012), blue (with white text): transcripts of proteins identified in (Bos et al., 2010), red: pea aphid transcripts of proteins identified in green peach aphid (Harmel et al., 2008), dark green (with white text): pea aphid transcripts of proteins identified in Russian wheat aphids (Cooper et al., 2010), blue (with black text): pea aphid transcripts of proteins identified in English grain aphid, rose grain aphid, and pea aphid (Rao et al., 2013), light green: pea aphid salivary gland enriched transcripts (Carolan et al., 2011), yellow: pea aphid salivary gland enriched transcripts (Balthazor et al., 2015), tan: green peach aphid transcripts corresponding to transcripts in the pea aphid (Ramsey et al., 2007), and peach: potato aphid transcripts corresponding to transcripts in the pea aphid (Atamian et al., 2012).

Color Key:

Atamian et al.
Balthazor et al.
Bos et al.
Carolan et al.
Cooper et al.
Feng et al.
Harmel et al.
Ramsey et al.
Rao et al.

Saliva Proteins				
Accession	Identification	Transcript Length	Salivary Gland Isolation	
			Reads Mapped	RPKM
ACYPI54712	unknown protein 13	728	7947	78.37
ACYPI009182	Zinc binding dehydrogenase	2328	2369	7.31
ACYPI009625	EMP24 like	1782	9189	37.02
ACYPI56654	3-hydroxyacyl-CoA dehydrogenase type-2	1244	864	4.99
ACYPI007677	Calreticulin	2299	50573	157.92
ACYPI002622	Calreticulin	2014	105539	376.19
ACYPI21412	Me20	1966	24413	89.14
ACYPI52702	Cathepsin B	1020	765	5.38
ACYPI009585	unknown protein 10	387	8322	154.37
ACYPI000490	Unannotated Protein 5	1162	708258	4375.65
ACYPI008926	Disulfide isomerase	2357	25591	77.94
ACYPI005818	Unannotated Protein 24	832	62152	536.28
ACYPI008667	Unannotated Protein 17	933	122762	944.58
ACYPI010222	Mp42	1130	36220	230.11
ACYPI005594	Disulfide isomerase	2013	28542	101.79
ACYPI53825	Me17	1017	19676	138.89
ACYPI002172	Unannotated Protein 18	1306	237379	1304.84
ACYPI45769	major royal jelly protein (yellow-g2)	2257	4140	13.17
ACYPI56566	Me13	856	13487	113.11
ACYPI45001	Unannotated Protein 10	1330	629606	3398.40
ACYPI001887	Unannotated Protein 26	976	34090	250.75
ACYPI007406	Unannotated Protein 9	1046	433606	2975.92
ACYPI24281	unknown protein 19	1430	8831	44.33
ACYPI010168	similar to CG5861-PA	884	3478	28.24
ACYPI008001	Armet	1525	9530	44.86
ACYPI003695	Unannotated Protein 25	612	16635	195.13
ACYPI000002	Sucrase	1293	25596	142.11
ACYPI001541	Unannotated Protein 13	887	99919	808.69
ACYPI004866	similar to CG11699-PA	4277	3451	5.79
ACYPI089376	CG2839	687	254246	2656.77
ACYPI38240	ApGPx2	1921	16006	59.82
ACYPI45597	unknown protein 18	1754	39522	161.76
ACYPI001523	Chorin Peroxidase H6	2322	2537	7.84
ACYPI001719	Unannotated Protein 15	1202	387130	2312.12
ACYPI42782	similar to CG9849-PA	1520	8049	38.02
ACYPI56502	Unannotated Protein 20	1198	233257	1397.77
ACYPI39568	Aphid specific cysteine rich protein	1387	1018208	5270.08
ACYPI55147	Unannotated Protein 12	1617	901	4.00
ACYPI55148	Unannotated Protein 27	3484	256238	527.99
ACYPI001271	Unannotated Protein 7	1018	169536	1195.56
ACYPI081664	unknown protein 12	1941	8997	33.28
ACYPI003247	similar to CG6583-PA	1027	1030	7.20
ACYPI005439	PAMP	1605	2070	9.26
ACYPI43360	Unannotated Protein 28	867	42955	355.67
ACYPI007065	Contig_37	1787	12008	48.24
ACYPI004394	unknown protein 28	1181	104363	634.39
ACYPI001706	similar to Der1-like domain family	1602	9964	44.65
ACYPI008617	C002	1020	528171	3717.34
ACYPI56620	cuticular protein	1207	40683	241.97
ACYPI21663	Me14	1489	4724	22.78
ACYPI007387	similar to ring finger protein 185	1654	3572	15.50
ACYPI003327	unknown protein 31	5941	10623	12.84
ACYPI001606	Unannotated Protein 14	1030	143828	1002.45
ACYPI006124	unknown protein 26	3228	17917	39.85
ACYPI007650	Beta-galactosidase precursor	2140	248	0.83
ACYPI007022	unknown protein 25	1161	2950	18.24
ACYPI005041	unknown protein 8	1782	1186	4.78
ACYPI002439	ApGPx1	2467	76627	222.98
ACYPI48356	unknown protein 16	3058	11370	26.69
ACYPI006346	Unannotated Protein 6	925	707062	5487.48
ACYPI23752	Carbonic anhydrase II	1125	32652	208.36
ACYPI007670	RNA Helicase	3944	4719	8.59

Saliva Proteins				
Accession	Identification	Transcript Length	Salivary Gland Isolation	
			Reads Mapped	RPKM
ACYPI005838	unknown protein 27	1901	21730	82.06
ACYPI002976	Tetraspanin 29Fa	2624	8390	22.95
ACYPI002497	Transmembrane 87B-like	2568	3146	8.79
ACYPI000768	Maltase-A1	2164	3603	11.95
ACYPI005882	Unknown protein 23	2064	4454	15.49
ACYPI008224	Me10	1047	644020	4415.81
ACYPI009919	Unannotated Protein 22	1056	31803	216.20
ACYPI000119	Disulfide isomerase	2637	25745	70.09
ACYPI088277	unknown protein 7	2070	142	0.49
ACYPI48849	unknown protein 15	3085	55	0.13
ACYPI26959	Peroxidase	3947	19175	34.88
ACYPI004198	Lipophorin precursor	8748	72165	59.22
ACYPI002476	Inositol Monophosphatase	1667	1755	7.56
ACYPI46095	unknown protein 17	1742	4023	16.58
ACYPI003780	unknown protein 29	5947	20553	24.81
LOC100575164	DPC	2410	307969	917.38
ACYPI001152	Unannotated Protein 30	2941	167450	408.74
ACYPI003602	unknown protein 30	3688	33264	64.75
ACYPI010151	unknown protein 22	7345	9179	8.97
ACYPI000852	Unannotated Protein 11	694	25929	268.22
ACYPI004591	chromatin STP2	1979	242781	880.70
ACYPI009881	Putative Sheath Protein	1348	1942185	10343.28
ACYPI002258	M1 zinc metalloprotease	2966	30394	73.57
ACYPI006974	Cathepsin L	2102	42488	145.11
ACYPI003917	SCP GAPR-1	2172	154603	510.99
ACYPI001445	unknown protein 32	7423	2103	2.03
ACYPI51013	unknown protein 14	2382	476	1.43
ACYPI007300	Endoribonuclease	5838	3438	4.23
ACYPI008182	Juvenile Hormone Binding Protein Homolog	1349	20561	109.42
ACYPI001099	Unannotated Protein 19	1350	124758	663.43
ACYPI009755	Disulfide isomerase	2931	100086	245.14
ACYPI006775	similar to CG2471-PA	4478	6186	9.92
ACYPI22506	unknown protein 20	1057	445	3.02
ACYPI002298	Trehalase	2637	25095	68.32
ACYPI080156	unknown protein 21	2086	10366	35.67
ACYPI000797	unknown protein 33	3080	2025	4.72
ACYPI38795	Unannotated Protein 31	1080	472	3.14
ACYPI001843	Unannotated Protein 23	2647	43443	117.82
ACYPI080546	Glutathione S transferase D10	1055	9630	65.53
ACYPI000422	unknown protein 35	9393	636939	486.80
ACYPI000288	Glucose Dehydrogenase	2335	269427	828.35
ACYPI007553	Unannotated Protein 29	1076	260383	1737.23
ACYPI28317	unknown protein 9	2760	5600	14.57
ACYPI003601	Unannotated Protein 16	3067	24380	57.07
ACYPI000558	Unannotated Protein 21	2936	75216	183.91
ACYPI000472	Unannotated Protein 3	1439	627952	3132.73
ACYPI009427	M1 zinc metalloprotease	1548	48018	222.68
ACYPI008911	Dipeptidyl carboxypeptidase	2722	201696	531.94
ACYPI071951	Peptidase M1	3458	144814	300.64
ACYPI082770	MCO1 (Laccase)	2154	1290	4.30
ACYPI063417	unknown protein 11	285	96368	2427.42
ACYPI000986	Glucose Dehydrogenase	3596	235771	470.68
ACYPI008370	CLIP-domain serine protease	2771	13495	34.96
ACYPI071317	Zinc-dependent Phospholipase C	1151	8595	53.61
ACYPI000707	unknown protein 34	2961	4892	11.86
ACYPI006300	Me25	2744	6706	17.54
ACYPI002891	Cadherin	2956	11450	27.81
ACYPI009042	Maltase 2-like	2574	56911	158.72
ACYPI073648	AHNAK nucleoprotein (desmoyokin)	12789	3503	1.97

Table 12 Comparative analysis of 121 saliva proteins in diet and plant-fed libraries by RNA-seq

Reads generated by RNA-seq were mapped to each individual transcript open reading frame as a "reference genome". The number of head reads, RPKM, and head RPKM fold changes were calculated as described in the text. Each transcript encodes a protein that contains a signal peptide and determination was achieved using SignalP 3.0.

A color key indicated below identify transcript names as follows: purple (with white text): pea aphid transcripts corresponding to transcripts studied in Russian wheat aphids (Cui et al., 2012), blue (with white text): transcripts of proteins identified in (Bos et al., 2010), red: pea aphid transcripts of proteins identified in green peach aphid (Harmel et al., 2008), dark green (with white text): pea aphid transcripts of proteins identified in Russian wheat aphids (Cooper et al., 2010), blue (with black text): pea aphid transcripts of proteins identified in English grain aphid, rose grain aphid, and pea aphid (Rao et al., 2013), light green: pea aphid salivary gland enriched transcripts (Carolan et al., 2011), yellow: pea aphid salivary gland enriched transcripts (Balthazor et al., 2015), tan: green peach aphid transcripts corresponding to transcripts in the pea aphid (Ramsey et al., 2007), and peach: potato aphid transcripts corresponding to transcripts in the pea aphid (Atamian et al., 2012).

Color Key:

Atamian et al.
Balthazor et al.
Bos et al.
Carolan et al.
Cooper et al.
Feng et al.
Harmel et al.
Ramsey et al.
Rao et al.

Saliva Proteins							
Accession	Identification	Transcript Length	Head Isolation				RPKM Fold Change
			Diet Fed		Plant Fed		
			Reads Mapped	RPKM	Reads Mapped	RPKM	
ACYPI54712	unknown protein 13	728	17	2.22	193	13.26	5.970
ACYPI009182	Zinc binding dehydrogenase	2328	49	2.00	528	11.34	5.666
ACYPI009625	EMP24 like	1782	184	9.82	1309	36.73	3.741
ACYPI56654	3-hydroxyacyl-CoA dehydrogenase type-2	1244	28	2.14	196	7.88	3.681
ACYPI007677	Calreticulin	2299	2067	85.50	13497	293.57	3.434
ACYPI002622	Calreticulin	2014	2084	98.40	13582	337.22	3.427
ACYPI21412	Me20	1966	257	12.43	1558	39.63	3.188
ACYPI52702	Cathepsin B	1020	1	0.09	6	0.29	3.155
ACYPI009585	unknown protein 10	387	76	18.67	455	58.79	3.148
ACYPI000490	Unannotated Protein 5	1162	5731	469.00	33954	1461.15	3.115
ACYPI008926	Disulfide isomerase	2357	518	20.90	3011	63.88	3.057
ACYPI005818	Unannotated Protein 24	832	949	108.47	5283	317.52	2.927
ACYPI008667	Unannotated Protein 17	933	1320	134.54	7152	383.32	2.849
ACYPI010222	Mp42	1130	172	14.47	920	40.71	2.813
ACYPI005594	Disulfide isomerase	2013	931	43.98	4899	121.70	2.767
ACYPI53825	Me17	1017	197	18.42	1007	49.51	2.688
ACYPI002172	Unannotated Protein 18	1306	1704	124.07	8636	330.66	2.665
ACYPI45769	major royal jelly protein (yellow-g2)	2257	55	2.32	272	6.03	2.601
ACYPI56566	Me13	856	155	17.22	763	44.57	2.589
ACYPI45001	Unannotated Protein 10	1330	4392	314.02	21587	811.62	2.585
ACYPI001887	Unannotated Protein 26	976	518	50.47	2543	130.29	2.582
ACYPI007406	Unannotated Protein 9	1046	3650	331.83	17724	847.31	2.553
ACYPI24281	unknown protein 19	1430	61	4.06	296	10.35	2.552
ACYPI010168	similar to CG5861-PA	884	129	13.88	623	35.24	2.540
ACYPI008001	Armet	1525	233	14.53	1124	36.86	2.537
ACYPI003695	Unannotated Protein 25	612	282	43.82	1343	109.73	2.504
ACYPI000002	Sucrase	1293	453	33.32	2139	82.72	2.483
ACYPI001541	Unannotated Protein 13	887	589	63.15	2773	156.33	2.476
ACYPI004866	similar to CG11699-PA	4277	105	2.33	489	5.72	2.449
ACYPI089376	CG2839	687	1258	174.13	5821	423.69	2.433
ACYPI38240	ApGPx2	1921	355	17.57	1634	42.53	2.420
ACYPI45597	unknown protein 18	1754	505	27.38	2297	65.49	2.392
ACYPI001523	Chorin Peroxidase H6	2322	170	6.96	773	16.65	2.391
ACYPI001719	Unannotated Protein 15	1202	5358	423.88	24358	1013.32	2.391
ACYPI42782	similar to CG9849-PA	1520	157	9.82	701	23.06	2.348
ACYPI56502	Unannotated Protein 20	1198	1076	85.41	4803	200.48	2.347
ACYPI39568	Aphid specific cysteine rich protein	1387	12071	827.59	53699	1935.98	2.339
ACYPI55147	Unannotated Protein 12	1617	2149	126.38	9533	294.80	2.333
ACYPI55148	Unannotated Protein 27	3484	2149	58.66	9533	136.82	2.333
ACYPI001271	Unannotated Protein 7	1018	2052	191.68	9015	442.82	2.310
ACYPI081664	unknown protein 12	1941	29	1.42	126	3.25	2.285
ACYPI003247	similar to CG6583-PA	1027	66	6.11	285	13.88	2.271
ACYPI005439	PAMP	1605	86	5.10	371	11.56	2.268
ACYPI43360	Unannotated Protein 28	867	346	37.95	1478	85.24	2.246
ACYPI007065	Contig_37	1787	237	12.61	1010	28.26	2.241
ACYPI004394	unknown protein 28	1181	917	73.84	3838	162.50	2.201
ACYPI001706	similar to Der1-like domain family	1602	281	16.68	1171	36.55	2.191
ACYPI008617	C002	1020	5129	478.17	21300	1044.22	2.184
ACYPI56620	cuticular protein	1207	3774	297.33	15639	647.91	2.179
ACYPI21663	Me14	1489	64	4.09	265	8.90	2.177
ACYPI007387	similar to ring finger protein 185	1654	180	10.35	734	22.19	2.144
ACYPI003327	unknown protein 31	5941	521	8.34	2122	17.86	2.142
ACYPI001606	Unannotated Protein 14	1030	1557	143.75	6308	306.24	2.130
ACYPI006124	unknown protein 26	3228	382	11.25	1547	23.96	2.130
ACYPI007650	Beta-galactosidase precursor	2140	28	1.24	113	2.64	2.122
ACYPI007022	unknown protein 25	1161	280	22.93	1111	47.85	2.086
ACYPI005041	unknown protein 8	1782	46	2.45	176	4.94	2.012
ACYPI002439	ApGPx1	2467	1198	46.18	4583	92.89	2.012
ACYPI48356	unknown protein 16	3058	299	9.30	1128	18.45	1.984
ACYPI006346	Unannotated Protein 6	925	6959	715.41	25754	1392.24	1.946
ACYPI23752	Carbonic anhydrase II	1125	407	34.40	1493	66.36	1.929
ACYPI007670	RNA Helicase	3944	212	5.11	761	9.65	1.888

Saliva Proteins Cont.

Accession	Identification	Transcript Length	Head Isolation				
			Diet Fed		Plant Fed		RPKM Fold Change
			Reads	RPKM	Reads	RPKM	
ACYPI005838	unknown protein 27	1901	225	11.26	804	21.15	1.879
ACYPI002976	Tetraspanin 29Fa	2624	131	4.75	458	8.73	1.838
ACYPI002497	Transmembrane 87B-like	2568	109	4.04	381	7.42	1.838
ACYPI000768	Maltase-A1	2164	240	10.55	809	18.69	1.773
ACYPI005882	Unknown protein 23	2064	479	22.07	1552	37.60	1.704
ACYPI008224	Me10	1047	8707	790.81	28095	1341.82	1.697
ACYPI009919	Unannotated Protein 22	1056	383	34.49	1234	58.43	1.694
ACYPI000119	Disulfide isomerase	2637	590	21.28	1894	35.92	1.688
ACYPI088277	unknown protein 7	2070	477	21.91	1485	35.87	1.637
ACYPI48849	unknown protein 15	3085	1	0.03	3	0.05	1.578
ACYPI26959	Peroxidase	3947	188	4.53	561	7.11	1.569
ACYPI004198	Lipophorin precursor	8748	804	8.74	2394	13.68	1.566
ACYPI002476	Inositol Monophosphatase	1667	124	7.07	369	11.07	1.565
ACYPI46095	unknown protein 17	1742	124	6.77	368	10.56	1.561
ACYPI003780	unknown protein 29	5947	171	2.73	505	4.25	1.553
LOC100575164	DPC	2410	4165	164.34	12281	254.82	1.551
ACYPI001152	Unannotated Protein 30	2941	6662	215.41	19617	333.54	1.548
ACYPI003602	unknown protein 30	3688	847	21.84	2493	33.80	1.548
ACYPI010151	unknown protein 22	7345	350	4.53	1026	6.98	1.541
ACYPI000852	Unannotated Protein 11	694	486	66.59	1412	101.74	1.528
ACYPI004591	chromatin STP2	1979	2802	134.64	7968	201.33	1.495
ACYPI009881	Putative Sheath Protein	1348	27662	1951.38	78099	2897.12	1.485
ACYPI002258	M1 zinc metalloprotease	2966	1317	42.22	3683	62.09	1.471
ACYPI006974	Cathepsin L	2102	2906	131.47	8052	191.55	1.457
ACYPI003917	SCP GAPR-1	2172	3921	171.67	10738	247.21	1.440
ACYPI001445	unknown protein 32	7423	118	1.51	319	2.15	1.422
ACYPI51013	unknown protein 14	2382	19	0.76	50	1.05	1.384
ACYPI007300	Endoribonuclease	5838	195	3.18	511	4.38	1.378
ACYPI008182	Juvenile Hormone Binding Protein Homolog	1349	2240	157.90	5802	215.07	1.362
ACYPI001099	Unannotated Protein 19	1350	1072	75.51	2766	102.45	1.357
ACYPI009755	Disulfide isomerase	2931	2779	90.16	7051	120.29	1.334
ACYPI006775	similar to CG2471-PA	4478	1166	24.76	2937	32.80	1.325
ACYPI22506	unknown protein 20	1057	71	6.39	176	8.33	1.304
ACYPI002298	Trehalase	2637	632	22.79	1522	28.86	1.266
ACYPI080156	unknown protein 21	2086	147	6.70	352	8.44	1.259
ACYPI000797	unknown protein 33	3080	141	4.35	329	5.34	1.227
ACYPI38795	Unannotated Protein 31	1080	58	5.11	135	6.25	1.224
ACYPI001843	Unannotated Protein 23	2647	398	14.30	921	17.40	1.217
ACYPI080546	Glutathione S transferase D10	1055	76	6.85	174	8.25	1.204
ACYPI000422	unknown protein 35	9393	9779	99.00	22339	118.92	1.201
ACYPI000288	Glucose Dehydrogenase	2335	4424	180.17	9454	202.46	1.124
ACYPI007553	Unannotated Protein 29	1076	1961	173.31	4127	191.79	1.107
ACYPI28317	unknown protein 9	2760	180	6.20	370	6.70	1.081
ACYPI003601	Unannotated Protein 16	3067	833	25.83	1710	27.88	1.079
ACYPI000558	Unannotated Protein 21	2936	932	30.19	1746	29.74	0.985
ACYPI000472	Unannotated Protein 3	1439	9779	646.22	17826	619.45	0.959
ACYPI009427	M1 zinc metalloprotease	1548	256	15.73	449	14.50	0.922
ACYPI008911	Dipeptidyl carboxypeptidase	2722	199	6.95	347	6.37	0.917
ACYPI071951	Peptidase M1	3458	2312	63.58	3998	57.81	0.909
ACYPI082770	MCO1 (Laccase)	2154	370	16.33	623	14.46	0.885
ACYPI063417	unknown protein 11	285	1368	456.45	2239	392.84	0.861
ACYPI000986	Glucose Dehydrogenase	3596	4629	122.41	7543	104.89	0.857
ACYPI008370	CLIP-domain serine protease	2771	3653	125.36	5545	100.06	0.798
ACYPI071317	Zinc-dependent Phospholipase C	1151	162	13.38	229	9.95	0.743
ACYPI000707	unknown protein 34	2961	1009	32.40	1418	23.95	0.739
ACYPI006300	Me25	2744	332	11.51	461	8.40	0.730
ACYPI002891	Cadherin	2956	424	13.64	523	8.85	0.649
ACYPI009042	Maltase 2-like	2574	1496	55.27	756	14.69	0.266
ACYPI073648	AHNAK nucleoprotein (desmoyokin)	12789	813	6.05	323	1.26	0.209

Table 13 Saliva proteome components secretion and anchor probability with ER retention signals

Listed in ascending order by AphidBase accession number, signal peptide prediction with cleavage site is identified with membrane anchor probability. ER retention signals are indicated with a yes followed by the four C-terminal amino acid residues in each encoded protein.

Accession	Identification	SignalP 3.0 HMM Result	Cleavage site	Anchor Probability	ER Retention Signal
ACYPI000002	Sucrase	0.948	21-22	0.000	no
ACYPI000119	Disulfide isomerase	0.753	32-33	0.080	Yes (KEEL)
ACYPI000288	Glucose Dehydrogenase	0.938	24-25	0.043	no
ACYPI000422	unknown protein 35	0.999	19-20	0.000	no
ACYPI000472	Unannotated Protein 3	0.999	26-27	0.000	no
ACYPI000490	Unannotated Protein 5	0.970	22-23	0.003	no
ACYPI000558	Unannotated Protein 21	0.975	25-26	0.010	no
ACYPI000707	unknown protein 34	0.981	31-32	0.000	no
ACYPI000768	Maltase-A1	0.896	20-21	0.000	no
ACYPI000797	unknown protein 33	0.888	18-19	0.000	no
ACYPI000852	Unannotated Protein 11	1.000	25-26	0.000	no
ACYPI000986	Glucose Dehydrogenase	0.983	22-23	0.004	no
ACYPI001099	Unannotated Protein 19	0.981	22-23	0.010	no
ACYPI001152	Unannotated Protein 30	0.999	23-24	0.001	no
ACYPI001271	Unannotated Protein 7	0.999	23-24	0.001	Yes (KEDK)
ACYPI001445	unknown protein 32	0.890	18-19	0.000	no
ACYPI001523	Chorin Peroxidase H6	1.000	34-35	0.000	no
ACYPI001541	Unannotated Protein 13	0.798	27-28	0.161	no
ACYPI001606	Unannotated Protein 14	0.977	24-25	0.000	no
ACYPI001706	similar to Der1-like domain family	0.718	32-33	0.040	no
ACYPI001719	Unannotated Protein 15	0.982	18-19	0.017	no
ACYPI001843	Unannotated Protein 23	0.924	25-26	0.012	no
ACYPI001887	Unannotated Protein 26	0.763	20-21	0.201	no
ACYPI002172	Unannotated Protein 18	0.536	25-26	0.245	no
ACYPI002258	M1 zinc metalloprotease	1.000	19-20	0.000	no
ACYPI002298	Trehalase	0.973	20-21	0.002	no
ACYPI002439	ApGPx1	0.905	28-29	0.016	no
ACYPI002476	Inositol Monophosphatase	0.944	19-20	0.000	no
ACYPI002497	Transmembrane 87B-like	0.966	21-22	0.008	no
ACYPI002622	Calreticulin	1.000	23-24	0.000	Yes (HDEL)
ACYPI002891	Cadherin	0.957	18-19	0.018	no
ACYPI002976	Tetraspanin 29Fa	0.899	61-62	0.880	no
ACYPI003247	similar to CG6583-PA	0.999	23-24	0.000	no
ACYPI003327	unknown protein 31	0.894	43-44	0.239	no
ACYPI003601	Unannotated Protein 16	0.994	19-20	0.005	no
ACYPI003602	unknown protein 30	0.919	18-19	0.000	no
ACYPI003695	Unannotated Protein 25	0.830	19-20	0.000	no
ACYPI003780	unknown protein 29	0.817	31-32	0.000	no
ACYPI003917	SCP GAPR-1	0.996	23-23	0.004	no
ACYPI004198	Lipophorin precursor	0.999	19-20	0.000	no
ACYPI004394	unknown protein 28	0.909	29-30	0.085	no
ACYPI004591	chromatin STP2	0.985	28-29	0.001	no
ACYPI004866	similar to CG11699-PA	0.918	50-51	0.483	no
ACYPI005041	unknown protein 8	0.902	18-19	0.009	no
ACYPI005439	PAMP	0.996	24-25	0.000	no
ACYPI005594	Disulfide isomerase	0.999	20-21	0.001	YES KHEL
ACYPI005818	Unannotated Protein 24	0.998	21-22	0.000	no
ACYPI005882	unknown protein 27	0.989	21-22	0.000	no
ACYPI006300	Unknown protein 23	0.998	23-24	0.000	no
ACYPI006346	unknown protein 26	0.999	19-20	0.000	no
ACYPI006300	Me25	0.998	23-24	0.000	no
ACYPI006346	Unannotated Protein 6	0.999	19-20	0.000	no
ACYPI006775	similar to CG2471-PA	0.997	18-19	0.000	no
ACYPI006974	Cathepsin L	0.999	19-20	0.000	no
ACYPI007022	unknown protein 25	0.998	21-22	0.001	no
ACYPI007065	Contig_37	0.917	25-26	0.058	Yes (HTEL)
ACYPI007300	Endoribonuclease	1.000	21-22	0.000	no
ACYPI007387	similar to ring finger protein 185	0.799	18-19	0.000	no
ACYPI007406	Unannotated Protein 9	0.998	22-23	0.001	no
ACYPI007553	Unannotated Protein 29	0.957	22-23	0.012	no
ACYPI007650	Beta-galactosidase precursor	0.996	23-24	0.003	no

Accession	Identification	SignalP 3.0 HMM Result	Cleavage site	Anchor Probability	ER Retention Signal
ACYPI007670	RNA Helicase	0.999	19-20	0.000	no
ACYPI007677	Calreticulin	0.995	23-24	0.005	no
ACYPI008001	Armet	0.993	20-21	0.005	no
ACYPI008182	Juvenile Hormone Binding Protein Homolog	0.989	20-21	0.001	no
ACYPI008224	Me10	0.608	27-28	0.142	no
ACYPI008370	CLIP-domain serine protease	0.942	19-20	0.025	no
ACYPI008617	C002	0.909	23-24	0.046	no
ACYPI008667	Unannotated Protein 17	0.938	28-29	0.017	no
ACYPI008911	Dipeptidyl carboxypeptidase	0.994	25-26	0.006	no
ACYPI008926	Disulfide isomerase	0.997	18-19	0.000	Yes (KEEL)
ACYPI009042	Maltase 2-like	0.991	21-22	0.002	no
ACYPI009182	Zinc binding dehydrogenase	0.999	20-21	0.000	no
ACYPI009427	M1 zinc metalloprotease	0.992	19-20	0.000	no
ACYPI009585	unknown protein 10	0.963	20-21	0.001	no
ACYPI009625	EMP24 like	0.939	39-40	0.059	no
ACYPI009755	Disulfide isomerase	0.998	18-19	0.000	YES (KDEL)
ACYPI009881	Putative Sheath Protein	0.992	25-26	0.000	no
ACYPI009919	Unannotated Protein 22	0.853	22-23	0.050	no
ACYPI010151	unknown protein 22	0.879	25-26	0.000	no
ACYPI010168	similar to CG5861-PA	0.981	23-24	0.015	no
ACYPI010222	Mp42	0.508	22-23	0.448	no
ACYPI063417	unknown protein 11	0.895	24-25	0.054	no
ACYPI071317	Zinc-dependent Phospholipase C	0.901	15-16	0.000	no
ACYPI071951	Peptidase M1	0.826	28-29	0.160	no
ACYPI073648	AHNAK nucleoprotein (desmoyokin)	0.798	17-18	0.000	no
ACYPI080156	unknown protein 21	0.697	24-25	0.947	no
ACYPI080546	Glutathione S transferase D10	0.778	28-29	0.769	no
ACYPI081664	unknown protein 12	0.883	42-43	0.934	no
ACYPI082770	MCO1 (Laccase)	0.988	27-28	0.700	no
ACYPI088277	unknown protein 7	0.908	18-19	0.001	no
ACYPI089376	CG2839	0.912	25-26	0.069	no
ACYPI21412	Me20	0.996	25-26	0.001	no
ACYPI21663	Me14	0.747	19-20	0.132	no
ACYPI22506	unknown protein 20	0.801	36-37	0.000	no
ACYPI23752	Carbonic anhydrase II	0.995	21-22	0.000	no
ACYPI24281	unknown protein 19	0.959	25-26	0.021	no
ACYPI26959	Peroxidase	0.999	19-20	0.000	no
ACYPI28317	unknown protein 9	0.959	26-27	0.018	no
ACYPI38240	ApGPx2	0.533	18-19	0.175	no
ACYPI38795	Unannotated Protein 31	0.992	16-17	0.000	no
ACYPI39568	Aphid specific cysteine rich protein	0.828	28-29	0.039	no
ACYPI42782	similar to CG9849-PA	0.547	29-30	0.432	no
ACYPI43360	Unannotated Protein 28	0.997	22-23	0.000	no
ACYPI45001	Unannotated Protein 10	0.585	28-29	0.291	no
ACYPI45597	unknown protein 18	0.999	20-21	0.000	no
ACYPI45769	major royal jelly protein (yellow-g2)	0.522	16-17	0.000	no
ACYPI46095	unknown protein 17	0.501	24-25	0.262	no
ACYPI48356	unknown protein 16	0.601	18-19	0.000	no
ACYPI48849	unknown protein 15	0.645	18-19	0.000	no
ACYPI51013	unknown protein 14	0.598	24-25	0.000	no
ACYPI52702	Cathepsin B	0.997	20-21	0.000	no
ACYPI53825	Me17	0.999	24-25	0.001	no
ACYPI54712	unknown protein 13	0.983	19-20	0.000	no
ACYPI55147	Unannotated Protein 12	1.000	18-19	0.000	no
ACYPI55148	Unannotated Protein 27	1.000	18-19	0.000	no
ACYPI56502	Unannotated Protein 20	0.730	28-29	0.147	no
ACYPI56566	Me13	0.996	22-23	0.004	no
ACYPI56620	cuticular protein	0.997	18-19	0.002	no
ACYPI56654	B-hydroxyacyl-CoA dehydrogenase type-2	0.997	32-33	0.008	no
LOC100575164	DPC	0.993	23-24	0.003	no

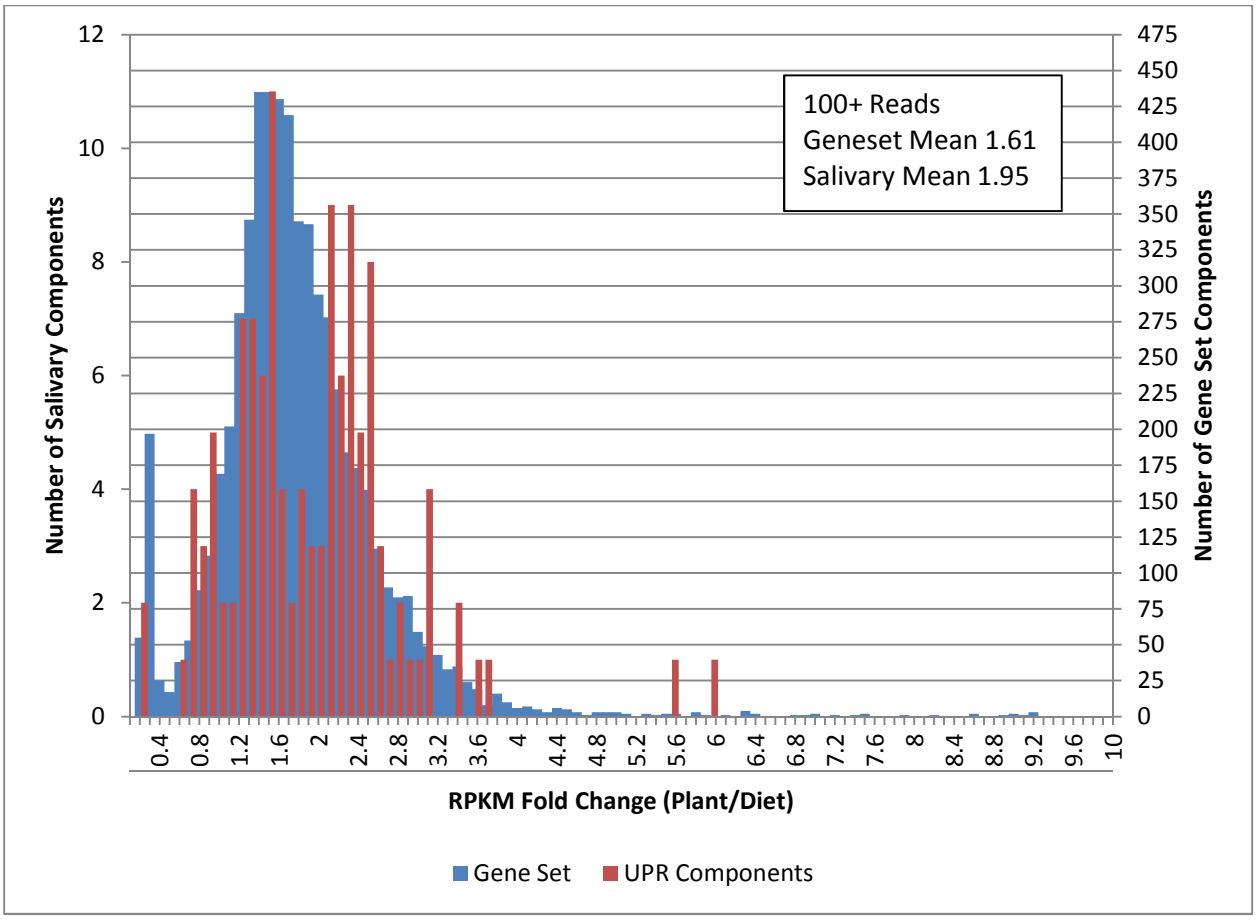


Figure 3.1 RPKM fold change ratio, plant versus diet feeding in the entire gene set and predicted transcripts of saliva

RPKM fold change in both predicted saliva transcripts and the entire gene set with a minimum 100 read threshold for salivary components measured by RNA-seq.

References:

Chapter 1:

Airavaara, M., Shen, H., Kuo, C.C., Peränen, J., Saarma, M., Hoffer, B., Wang, Y. (2009) Widespread cortical expression of MANF by AAV serotype 7: localization and protection against ischemic brain injury. J. Comp. Neurol. (225) 104-113.

Anet, F.A.L., Bourn, A.J.R. (1965) Nuclear Magnetic Resonance Spectral Assignments from Nuclear Overhauser Effects. Journal of the American Chemical Society. (87) 5250-5251.

Apostolou, A., Shen, Y., Liang, Y., Luo, J., Fang, S., (2008) Armet, a UPR-upregulated protein, inhibits cell proliferation and ER stress-induced death. Exp. Cell Res. (314) 13.

Campbell, C.T., Kim, G. (2007) SPR microscopy and its applications to high-throughput analyses of biomolecular binding events and their kinetics. Biomaterials. (28) 2380–2392.

Evron, E., Cairns, P., Halachmi, N., Ahrendt, S.A., Reed, A.L. (1997) Normal polymorphism in the incomplete trinucleotide repeat of the arginine-rich protein gene. Cancer Res. (57) 2888–2889.

Gerardo, N.M., Altincicek, B., Anselme, C., Atamian, H., Barribeau, S.M., de Vos, M., Duncan, E.J., Evans, J.D., Gabaldón, T., Ghanim, M., Heddi, A., Kaloshian, I., Latorre, A., Moya, A., Nakabachi, A., Parker, B.J., Pérez-Brocal, V., Pignatelli, M., Rahbé, Y., Ramsey, J.S., Spragg, C.J., Tamames, J., Tamarit, D., Tamborindeguy, C., Vincent-Monegat, C., Vilcinskas, A. (2010) Immunity and other defenses in pea aphids, Acyrthosiphon pisum. Genome Biol. (11) r21.

Glembotski, C., Thuerauf, D., Huang, C., Vekich, J., Gottlieb, R., Doroudgar, S. (2012) Mesencephalic astrocyte-derived neurotrophic factor protects the heart from ischemic damage and is selectively secreted upon sarco/endoplasmic reticulum calcium depletion. J Biol Chem. (287) 25893–25904.

Hellman, M., Peränen, J., Saarma, M., Permi, P. (2010) 1H, 13C and 1N resonance assignments of the human mesencephalic astrocyte-derived neurotrophic factor. Biomolecular NMR assignments. (2) 215-217.

Hellman, M., Arumae, U., Yu, L.Y., Lindholm, P., Peranen, J., Saarma, M., Permi, P. (2011) Mesencephalic astrocyte-derived neurotrophic factor (MANF) has a unique mechanism to rescue apoptotic neurons. J Biol Chem. (286) 2675–2680.

Henderson, M.J., Richie, C.T., Airavaara, M., Wang. Y., Harvey, B.K. (2013) Mesencephalic astrocyte-derived neurotrophic factor (MANF) secretion and cell surface binding are modulated by KDEL receptors. J Biol Chem. (288) 4209-4225.

Hoseki, J., Saskawa, H., Yamaguchi, Y., Maeda, M., Kubota, H., Kato, K., Nagata, K. (2010) Solution structure and dynamics of mouse ARMET. Febs Lett. (584) 1536-1542.

Hotton III, N. (1968) The evidence of evolution. New York. American Heritage Publishing Company, INC.

Ibañez, C.F. (1998) Emerging themes in structural biology of neurotrophic factors. Trends Neurosci. (21) 438–444.

Ito, T., Warnken, S.P., May, W.S. (1999) Protein synthesis inhibition by flavonoids: roles of eukaryotic initiation factor 2 alpha kinases. Biochem Biophys Res Commun (265) 589-594.

Kim, I., Xu, W., Reed, J.C. (2008) Cell death and endoplasmic reticulum stress: disease relevance and therapeutic opportunities. Nature reviews Drug discovery. (7) 1013–1030.

Kokame, K., Kato, H., Miyata, T. (2001) Identification of ERSE-II, a new cis-acting element responsible for the ATF6-dependent mammalian unfolded protein response. J Biol Chem (276) 9199–9205.

Kooyman, R.P., Kolkman, H., Van Gent, J., Greve, J., (1998) Surface plasmon resonance immunosensors: sensitivity considerations, Anal. Chim. Acta. (213) 35-45.

Lee, A.H., Iwakoshi, N.N., Glimcher, L.H. (2003) XBP-1 regulates a subset of endoplasmic reticulum resident chaperone genes in the unfolded protein response. Mol. Cell. Biol. (23) 7448–7459.

Llewellyn, D.H., Roderick, H.L., Rose, S. (1997) KDEL receptor expression is not coordinatedly up-regulated with ER stress-induced reticuloplasmin expression in HeLa cells. Biochem. Biophys. Res. Commun. (240) 36–40.

Lindholm, P., Pera¨nen, J., Andressoo, J.O., Kalkkinen, N., Kokaia, Z., Lindvall, O., Timmusk, T. Saarma, M. (2008) MANF is widely expressed in mammalian tissues and differently regulated after ischemic and epileptic insults in rodent brain. Mol. Cell. Neurosci., (39) 356–371.

Lust, S., Vanhoecke, B., Vang, M., Boelens, J., Vanm, H., Kaileh, M. (2009) Xanthohumol activates the proapoptotic arm of the unfolded protein response in chronic lymphocytic leukemia. Anticancer research. (29) 3797–3805.

Ma, X., Xi, L., Luo, D., Liu, R., Li, S., Liu, Y., (2012) Anti-Tumor Effects of the Peptide TMTP1-GG-D(KLAKLAK)2 on Highly Metastatic Cancers. PLoS ONE (7) e42685.

Malhotra, J.D., Kaufman, R.J. (2011) ER stress and its functional link to mitochondria: role in cell survival and death. Cold Spring Harb Perspect Biol. 3:a004424.

Miao, Y.R., Eckhardt, B.L., Cao, Y., (2013) Inhibition of established micrometastases by targeted drug delivery via cell surface-associated GRP78. Clin Cancer Res. (19) 2107–2116.

Mizobuchi, N., Hoseki, J., Kubota, H., Toyokuni, S., Nozaki, J., Naitoh, M., Koizumi, A. Nagata, K. (2007) ARMET is a soluble ER protein induced by the unfolded protein response via ERSE-II element. Cell Struct. Funct. (32) 41–50.

Oh-Hashi, K., Tanaka, K., Koga, H., Hirata, Y., Kiuchi, K. (2012) Intracellular trafficking and secretion of mouse mesencephalic astrocyte-derived neurotrophic factor. Mol Cell Biochem (363) 35–41.

Parkash, V., Lindholm, P., Peränen, J., Kalkkinen, N., Oksanen, E., Saarma, M., Leppänen, V.M., Goldman, A. (2009) The structure of the conserved neurotrophic factors MANF and CDNF explains why they are bifunctional. Protein Engineering. (4) 233-241.

Perez-Ramirez, B., Steckert, J.J. (2005). Therapeutic proteins: methods and protocols. Smales, C.M., James, D.C. Eds. Humana Press Inc, Totowa, NJ. Volume (308) 301-318.

Petrova, P., Raibekas, A., Pevsner, J., Vigo, N., Anafi, M., Moore, M.K., Peaire, A.E., Shridhar, V., Smith, D.I., Kelly, J., Durocher, Y., Commissiong, J.W. (2003) MANF: a new mesencephalic, astrocyte-derived neurotrophic factor with selectivity for dopaminergic neurons. J. Mol. Neurosci. (20) 173–188.

Rauthu, S. R., Shiao, T. C., André, S., Miller, M. C., Madej, É., Mayo, K. H., Gabius, H.-J. and Roy, R. (2015) Defining the Potential of Aglycone Modifications for Affinity/Selectivity Enhancement against Medically Relevant Lectins: Synthesis, Activity Screening, and HSQC-Based NMR Analysis. ChemBioChem, (16) 126–139.

Raykhel, I., Alanen, H., Salo, K., Jurvansuu, J., Nguyen, V.D., Latva-Ranta, M., Ruddock, L. (2007) A molecular specificity code for the three mammalian KDEL receptors J. Cell Biol. (179) 1193–1204.

Shridhar, V., Rivard, S., Shridhar, R., Mullins, C., Bostick, L., Sakr, W., Grignon, D., Miller, O.J., Smith, D.I. (1996) A gene from human chromosomal band 3p21.1 encodes a highly conserved arginine-rich protein and is mutated in renal cell carcinomas. Oncogene. (12) 1931–1939.

Tadimalla, A., Belmont, P.J., Thuerauf, D.J., Glassy, M.S., Martindale, J.J., (2008) Mesencephalic astrocyte-derived neurotrophic factor is an ischemia-inducible secreted endoplasmic reticulum stress response protein in the heart. Circ Res. (103) 1249–1258.

Uhlén, M., Fagerberg, L., Hallström, B.M., Lindskog, C., Oksvold, P., Mardinoglu, A., Sivertsson, A., Kampf, C., Sjöstedt, E., Asplund, A., Olsson, I.M., Edlund, K., Lundberg, E., Navani, S., Szigyarto, C.A., Odeberg, J., Djureinovic, D., Takanen, J.O., Hober, S., Alm, T., Edqvist, P.H., Berling, H., Tegel, H., Mulder, J., Rockberg, J., Nilsson, P., Schwenk, J.M., Hamsten, M., Feilitzen, K., Forsberg, M., Persson, L., Johansson, F.,

Zwahlen, M., Heijne, G., Nielsen, J., Pontén, F. (2015) Tissue-based map of the human proteome. Science (6220) 1260419

Wang, W., Dai, H., Zhang, Y., Chandrasekar, R., Luo, L., Hiromasa, Y., Sheng, C., Peng, G., Chen, S., Tomich, J.M., Reese, J., Edwards, O., Kang, L., Reeck, G., Cui, F. (2015) Armet is an effector protein mediating aphid-plant interactions. FASEB J. (5) 2032-2045.

Yu, Y.Q., Liu, L.C., Wang, F.C., Liang, Y., Cha, D.Q., Zhang, J.J., Shen, Y.J., Wang, HP, Fang, S, Shen, YX (2010) Induction profile of MANF/ARMET by cerebral ischemia and its implication for neuron protection. J Cereb Blood Flow Metab. (30) 79-91.

Chapter 2:

Akey, D.H., Beck, S.D. (1972) Nutrition of the pea aphid, *Acyrthosiphon pisum*: Requirements for trace metals, sulphur, and cholesterol. Journal of insect physiology, (18) 1901-1914.

Aksamit, M.S. (2014) Bioinformatic Analysis of Pea Aphid Salivary Gland Transcripts. MSc Thesis. Kansas State University, USA.

Alberts, B., Johnson, A., Lewis, J., Raff, M., Roberts, K., Walter, P., (2002) Molecular Biology of the Cell. 4th edition. New York: Garland Science;. The Endoplasmic Reticulum.

Atamian, H.S., Chaudhary, R., Dal Cin, V., Bao, E., Girke, T., Kaloshian, I. (2013). In planta expression or delivery of potato Aphid Macrosiphum euphorbiae effectors Me10 and Me23 enhances aphid fecundity. Mol. Plant Microbe Interact. (26) 67–74.

Blackman, R.L. Eastop, V.F. (2000) Aphids on the World's Crops, Second Edition. John Wiley & Sons with the Natural History Museum, London.

Bolognesi, R., Ramaseshadri, P., Anderson, J., Bachman, P., Clinton, W., Flannagan, R., Ilagan, O., Lawrence, C., Levine, S., Moar, W., (2012) Characterizing the mechanism of action of double-stranded RNA activity against western corn rootworm (Diabrotica virgifera virgifera LeConte). PLoS One 7: e47534.

Bommiasamy, H., Back, S. H., Fagone, P., Lee, K., Meshinchi, S., Vink, E., Sriburi, R., Frank, M., Jackowski, S., Kaufman, R. J., Brewer, J. W. (2009) ATF6alpha induces XBP1-independent expansion of the endoplasmic reticulum. J. Cell Sci. (122) 1626–1636.

Bos, J.I., Prince, D., Pitino, M., Maffei, M.E., Win, J., Hogenhout, S.A. (2010) A functional genomics approach identifies candidate effectors from the aphid species *Myzus persicae* (green peach aphid). PLoS genetics. (6) e1001216.

Braendle, C., Davis, G. K., Brisson, J. A., Stern, D. L. (2006) Wing dimorphism in aphids. Heredity (97) 192–199.

Bruce, T., Aradottir, G.I., Smart, L.E., Martin, J.L., Caulfield, J.C., Doherty, A., Sparks, C.A., Woodcock, C.M., Birkett, M.A., Napier, J.A., Jones, H.D., Pickett, J.A. (2015) The first

crop plant genetically engineered to release an insect pheromone for defense. Scientific Reports 5, Article number 11183.

Brunagel, G., Shah, U., Schoen, R.E., Getzenberg, R.H. (2003) Identification of calreticulin as a nuclear matrix protein associated with human colon cancer. J. Cell Biochem. (89) 238–243.

Camacho, P., John, L., Li, Y., Paredes, R.M., Roderick, H.L. (2003). Calnexin and calreticulin, ER associated modulators of calcium transport in the ER. Molecular Biology Intelligence Unit. 126–132.

Carolan, J.C., Fitzroy, C.I., Ashton, P.D., Douglas, A.E., Wilkinson, T.L. (2009) The secreted salivary proteome of the pea aphid Acyrthosiphon pisum characterized by mass spectrometry. Proteomics. (9) 2457-2467.

Carolan, J.C., Caragea, D., Reardon, K.T., Mutti, N.S., Dittmer, N., Pappan, K., Cui, F., Castaneto, M., Poulain, J., Dossat, C., Tagu, D., Reese, J.C., Reeck, G.R., Wilkinson, T.L., Edwards, O.R. (2011) Predicted effector molecules in the salivary secretome of the pea aphid (*Acyrthosiphon pisum*): a dual transcriptomic/proteomic approach. Journal of Proteome Research. (10) 1505-1518.

Chakrabarti, A., Chen, A.W., Varner, J.D. (2011). A review of the mammalian unfolded protein response. Biotechnology and Bioengineering, (108) 2777–2793.

Chow, C.Y., Wolfner, M.F., Clark, A.G. (2013) Using natural variation in Drosophila to discover previously unknown endoplasmic reticulum stress genes. Proc Natl Acad Sci USA. (110) 9013–9018.

Cooper, W.R., Dillwith, J.W., Puterka, G.J. (2010). Salivary proteins of Russian wheat aphid (Hemiptera: Aphididae). Environmental entomology. (39) 223-231.

Cui, F., Smith, M., Reese, J., Edwards, O., Reeck G.R. (2012) Polymorphisms in salivary-gland transcripts of Russian wheat aphid biotypes 1 and 2. Insect Science, (19) 429-440.

D'Arcy, C.J., Domier, L.L. (2000) Barley yellow dwarf. The plant health instructor. DOI: 10.1094/PHI-I-2000-1103-01 Updated 2005.

Delom, F., Mallet, B., Carayon, P., Lejeune, P.J. (2001) Role of extracellular molecular chaperones in the folding of oxidized proteins. Refolding of colloidal thyroglobulin by protein disulfide isomerase and immunoglobulin heavy chain-binding protein. J. Biol. Chem. (276) 21337–21342.

Dinant S, Bonnemain J-L, Girousse C, Kehr J. (2010) Phloem sap intricacy and interplay with aphid feeding. C R Biol. (333) 504–515

Dissemond, J., Busch, M., Kothen, T., Mors, J., Weimann, T.K., Lindeke, A. (2004) Differential downregulation of endoplasmic reticulum-residing chaperones calnexin and calreticulin in human metastatic melanoma. Cancer Letters. (203) 225–231.

Dixon, A.F.G. (1987) The way of life of aphids: host specificity, speciation and distribution. In: Minks AK, Harrewijn P, editors. Aphids. New York: Elsevier. 197-207.

Du, Z., Treiber, D., McCoy, R.E. (2013) Unfolded Protein Response (UPR) During CHO Cell Production Culture. Developments in Biotechnology and Bioprocessing. American Chemical Society. (2) 19-32.

Fürstenberg-Hägg, J., Zagrobelny, M., Bak, S. (2013) Plant defense against insect herbivores. Int. J. Mol. Sci. (14) 10242-10297.

Gordon, K., Waterhouse, P. M. (2007). RNAi for insect-proof plants. Nat. Biotechnol. (25) 1231–1232.

Guo, K, Wang, W., Luo, L., Chen, J., Guo, Y., Cui, F. (2014) Characterization of an aphid-specific, cysteine-rich protein enriched in salivary glands. Biophysical Chemistry, (189) 25-32.

Harmel, N., Létocart, E., Cherqui, A., Giordanengo, P., Mazzucchelli, G., Guillonneau, F., De Pauw, E., Haubruge, E., Francis, F. (2008). Identification of aphid salivary proteins: a proteomic investigation of *Myzus persicae*. Insect molecular biology. (17) 165-174.

Haze, K., Yoshida, H., Yanagi, H., Yura, T., Mori, K. (1999) Mammalian transcription factor ATF6 is synthesized as a transmembrane protein and activated by proteolysis in response to endoplasmic reticulum stress. Mol. Biol. Cell (10) 3787–3799.

Hendershot, L.M., Valentine, V.A., Lee, A.S., Morris, S.W., Shapiro, D.N. (1994) Localization of the gene encoding human BiP/GRP78, the endoplasmic reticulum cognate of the HSP70 family, to chromosome 9q34. Genomics (20) 281–284.

Hetz, C., Thielen, P., Matus, S., Nassif, M., Court, F., Kiffin, R., Martinez, G., Cuervo, A.M., Brown, R.H., Glimcher, L.H. (2009) XBP-1 deficiency in the nervous system protects against amyotrophic lateral sclerosis by increasing autophagy. Genes Dev. (23) 2294-2306.

Hetz, C., Chevet, E., Harding, H. (2013). Targeting the unfolded protein response in disease. Nature Rev Drug Discovery (12) 703-719.

Hetz, C., Martinon, F., Rodriguez, D., Glimcher, L.H. (2011) The unfolded protein response: integrating stress signals through the stress sensor IRE1α. Physiological Reviews (91) 1219-1243.

The International Aphid Genomics Consortium (2010) Genome Sequence of the Pea Aphid Acyrthosiphon pisum. PLoS Biol (8) e1000313.

Jaouannet, M., Rodriguez, P. A., Thorpe, P., Lenoir, C. A., Macleod, R., Escudero-Martinez, C., (2014) Plant immunity in plant-aphid interactions. Front Plant Sci. (5) 663.

Kansas State University Online Research and Extension Guide (2015) Retrieved: 6/5/2015. http://www.ksre.ksu.edu/bookstore/pubs/mf809.pdf.

Kuny, S., Gaillard, F., Sauve, Y. (2012) Differential gene expression in eyecup and retina of a mouse model of Stargardt-like macular dystrophy (STGD3). Invest Ophthalmol Vis Sci. (53) 664-675.

Lee, A.S. (2001) The glucose-regulated proteins: stress induction and clinical applications. Trends in Biochemical Sciences (26) 504–510.

Liu, F., Wang, X.D., Zhao, Y.Y., Li, Y.J., Liu, Y.C., Sun, J. (2015) Silencing the HaAK gene by transgenic plant-mediated RNAi impairs larval growth of Helicoverpa armigera. Int J Biol Sci. (11) 67-74.

Mesaeli, N., Nakamura, K., Zvaritch, E., Dickie, P., Dziak, E., Krause, K.H. (1999) Calreticulin is essential for cardiac development. J. Cell Biol. (144) 857–868.

Miles, P.W. (1968) Insect secretions in plants. Ann. Rev. Phytopathol. (6) 137–64.

Miles, P.W. (1999) Aphid saliva. Biol Rev (74) 41–85.

Minks, A.K., Harrewijn, P., editors. (1989) Aphids: their biology, natural enemies and control. New York: Elsevier.

Ogino, T., Bandoh, N., Hayashi, T., Miyokawa, N., Harabuchi, Y., Ferrone S. (2003) Association of tapasin and HLA class I antigen down-regulation in primary maxillary sinus squamous cell carcinoma lesions with reduced survival of patients. Clin. Cancer Res. (9) 4043–4051.

Oslowski, C.M., Urano, F. (2011) Measuring ER stress and the unfolded protein response using mammalian tissue culture system. Methods Enzymol. (490) 71–92.

Palgi, M., Lindstrom, R., Peranen, J., Piepponen, T.P., Saarma, M., Heino, T.I. (2009) Evidence that DmMANF is an invertebrate neurotrophic factor supporting dopaminergic neurons. Proc. Natl. Acad. Sci. USA, (106) 2429–2434

Pilcher, C., Rice, M. (2005) Economic impact of soybean aphid. Iowa State University Integrated Crop Management. Department of Entomology. Retrieved: 6/10/2015. http://www.ipm.iastate.edu/ipm/icm/node/53.

Polhemus, J.T., Polhemus, D.A. (2008) Global diversity of true bugs (Heteroptera; Insecta) in freshwater animal diversity assessment. Developments in Hydrobiology (198) 379-391.

Rai K. M., Singh, S.K., Bhardwaj, A., Kumar, V., Lakhwani, D., Srivastava, A., Jena, S.N., (2013) Large-scale resource development in Gossypium hirsutum L. by 454 sequencing of genic-enriched libraries from six diverse genotypes. Plant Biotechnology Journal (11) 953–963.

Ramsey, J.S., Wilson, A.C.C., de Vos, M., Sun, Q., Tamborindeguy, C., Winfield, A., Malloch, G., Smith, D.M., Fenton, B., Gray, S.M., (2007) Genomic resources for *Myzus persicae*: EST sequencing, SNP identification, and microarray design. BMC Genomics. (8) 423.

Rao, S.A., Carolan, J.C., Wilkinson, T.L. (2013) Proteomic profiling of cereal aphid saliva reveals both ubiquitous and adaptive secreted proteins. PloS one. (8) e57413.

Ryoo, H.D., Steller, H. (2007) Unfolded protein response in Drosophila: why another model can make it fly. Cell Cycle. (6:7) 830-835.

Saito, Y., Ihara, Y., Leach, M.R., Cohen-Doyle, M.F., Williams D.B. (1999) Calreticulin functions in vitro as a molecular chaperone for both glycosylated and non-glycosylated proteins. EMBO J. (18) 6718–6729.

Schroder, M., Kaufman, R. J. (2005) The mammalian unfolded protein response. Annu. Rev. Biochem. (74) 739–789.

Simon, J. C., Stoeckel, S., Tagu, D. (2010) Evolutionary and functional insights into reproductive strategies of aphids. Comptes Rendus Biologies (333) 488–496.

Stekel, D.J., Git, Y., Falciani, F. (2000) The comparison of gene expression from multiple cDNA libraries. Genome Research. (10) 2055-2061.

Tigges, M., Fussenegger, M. (2006) XBP1-based engineering of secretory capacity enhances the productivity of Chinese hamster ovary cells. Metab. Eng. (8) 264–272.

Ting, J., Lee, A.S. (1988). Human gene encoding the 78,000-dalton glucose-regulated protein and its pseudogene: structure, conservation, and regulation. DNA (7) 275–286.

Tjallingii, W.F. (2006) Salivary secretions by aphids interacting with proteins of phloem wound responses. J Exp Bot (57) 739–745.

Tsang, K.Y., Chan, D., Bateman, J.F., Cheah, K.S. (2010) In vivo cellular adaptation to ER stress: survival strategies with double-edged consequences. J. Cell Sci., (123) 2145–2154.

van Emden, H., Harrington, R., (2007) Life cycles and polymorphism, aphids as crop pests. CAB International, London, 69-87.

Walter, P., Ron, D. (2011) The unfolded protein response: from stress pathway to homeostatic regulation. Science. (334) 1081-1086.

Wilkinson, B., Gilbert, H.F. (2004). Protein disulfide isomerase. Biochim Biophys Acta (1699) 35–44.

Ye, J., Rawson, R. B., Komuro, R., Chen, X., Dave, U. P., Prywes, R., Brown, M. S., Goldstein, J. L. (2000) ER stress induces cleavage of membranebound ATF6 by the same proteases that process SREBPs. Mol. Cell. (6) 1355–1364

Yoshida, H., Matsui, T., Yamamoto, A., Okada, T., Mori, K. (2001) XBP1 mRNA is induced by ATF6 and spliced by IRE1 in response to ER stress to produce a highly active transcription factor. Cell (107) 881–891.

Chapter 3:

Alberts, B., Johnson, A., Lewis, J., Raff, M., Roberts, K., Walter, P. (2008) Chapter 18 Apoptosis: Programmed cell death eliminates unwanted cells. Mol. Biol. Cell. 1115.

Amerik, A.Y., Hochstrasser, M. (2000) Analysis of the deubiquitinating enzymes of the yeast S. cerevisiae. J. Biol. Chem. (10) 981-992.

Apostolou, A., Shen, Y., Liang, Y., Luo, J., Fang, S. (2008) Armet, a UPR-upregulated protein, inhibits cell proliferation and ER stress-induced death. Exp. Cell Res. (314) 13.

Aravind, L., Ponting, C.P. (1998) Homologues of 26S proteasome subunits are regulators of transcription and translation. Prot. Sci. (7) 1250-1254.

Arkane, Y., Muthukrishnan, S., Kramer, K.J., Specht, C..A., Tomoyasu, Y., Lorenzen, M.D., Kanost, M., Beeman, R.W., (2005) The Tribolium chitin synthase genes TcCHS1 and TcCHS2 are specialized for synthesis of epidermal cuticle and mid gut peritrophic matrix. Insect Mol. Biol. (14) 453-463.

Attardo, G.M., Higgs, S., Klingler, K.A., Vanlandingham, D.L., Raikhel, A.S., (2003) RNA interference-mediated knockdown of a GATA factor reveals a link to anautogeny in the mosquito Aedesae gypti. Proc. Natl. Acad. Sci. USA. (100) 13374–13379.

Bettencourt, R., Terenius, O., Paye, I. (2002) Hemolin gene silencing by ds-RNA injected into Cecropia pupae is lethal to next generation embryos. Insect Mol. Biol. (11) 267-271.

Balch, W.E., Morimoto, R.I., Dillin, A., Kelly, J.W. (2008) Adapting proteostasis for disease intervention. Science. (319) 916-919.

Bertolotti, A., Zhang, Y., Hendershot, L.M., Harding, H.P., Ron, D. (2000) Dynamic interaction of BiP and ER stress transducers in the unfolded-protein response. Nature Cell Biol. (2) 326-332.

Bommiasamy, H., Back, S., Fagone, P., Lee, K., Meshinchi, S., Vink, E., Sriburi, R., Frank, M., Jackowski, S., Kaufman, R.J., Brewer, J.W. (2009) ATF6α induces XBP1-independent expansion of the endoplasmic reticulum. J. Cell Sci. (122) 1626-1636.

DuRose, J.B., Scheuner, D., Kaufman, R.J., Rothblum, L.I., Niwa, M. (2009) Phosphorylation of eukaryotic translation initiation factor 2α coordinates rRNA transcription and translation inhibition during endoplasmic reticulum stress. Mol. Cell Biol. (29) 4295-4307.

Fire, A., Xu, S., Montgomery, M.K., Kostas, S.A., Driver, S.E., Mello, C.C. (1998) Potent and specific genetic interference by double-stranded RNA in Caenorhabditis elegans. Nature. (391) 806-811.

Goto, A,. Blandin, S., Royet, J., Reichhart, J.M., Levanshina, E.A. (2003) Silencing of Toll pathway components by direct injection of double-strated RNA into Drosophila adult flies. Nucleic Acids Res. (22) 6619-6623.

Gruber, C.W., Cemazar, M., Heras, B., Martin, J.L., Craik, D.J. (2006) Protein disulfide isomerase: the structure of oxidative folding. Trends Biochem. Sci. (31) 455-64.

Harding, H.P., Novoa, I., Zhang, Y., Zeng, H., Wek, R., Schapira, M., Ron, D. (2000) Regulated translation initiation controls stress-induced gene expression in mammalian cells. Mol. Cell (6) 1099-1108.

Hatahet, F., Ruddock, L.W. (2007) Substrate recognition by the protein disulfide isomerases. FEBS J. (274) 5223-5234.

Jaubert-Possamai, S. (2007) Gene knockdown by RNAi in the pea aphid Acyrthosiphon pisum. BMC Biotechnol. (7) 63.

Karam, J.A., Hsieh, JT. (2009) Anti-Cancer strategy of transitional cell carcinoma of bladder based on induction of different types of programmed cell deaths: Apoptosis in carcinogenesis and chemotherapy. Springer. 25-50.

Kimura, Y., Tanaka, K. (2010) Regulatory mechanisms involved in the control of ubiquitin homeostasis. J. Biol. Chem. (147) 793-8.

Koumenis, C. (2006) ER stress, hypoxia tolerance and tumor progression. Curr. Mol. Med. (6) 55-69.

Levin, D.M., Breuer, L.N., Zhuang, S., Anderson, S.A., Nardi, J.B., Kanost, M.R., (2005) A hemocyte-specific integrin required for hemocytic encapsulation in the tobacco hornworm, Manduca sexta. Insect Biochem. Mol. Biol. (35) 369-80.

Lipson, K.L., Fonseca, S.G., Urano, F. (2006) Endoplasmic reticulum stress-induced apoptosis and auto-immunity in diabetes. Curr. Mol. Med. (6) 71-77.

Moenner, M., Pluquet, O., Bouchecareilh, M., Chevet, E. (2007) Integrated endoplasmic reticulum stress responses in cancer. Cancer Res. (67) 10631-10634.

Mulder, P., Seuhs, K. Oklahoma State University. Department of Entomology. Current report on alfalfa forage insect control. retrieved 7/13/2014. http://pods.dasnr.okstate.edu/docushare/dsweb/Get/Document-2364/EPP-7150web.pdf

Mutti, M.S., Park, Y., Reese, J.C., Reeck, G.R. (2006) RNAi knockdown of a salivary transcript leading to lethality in the pea aphid, Acyrthosiphon pisum. J. Insect Sci. (6) 37-38.

Naidoo, N. (2009) ER and aging: protein folding and the ER stress response. Ageing Res. Rev. (8) 150-159.

Nassif, M. Matus, S. Castillo, K. Hetz, C. (2010) Amyotrophic lateral sclerosis pathogenesis: a journey through the secretory pathway. Antioxid. Redox Signal. (13) 1955-1989.

Pitino, M., Coleman, A.D., Maffei, M.E., Ridout, C.J., Hogenhout, S.A. (2011) Silencing of aphid genes by dsRNA feeding from plants. PLoS Biol. (6) 25709.

Rajagopal, R., Sivakumar, S., Agrawal, N., Malhotra, P., Bhatnagar, R.K. (2002) Silencing of midgutaminopeptidase N of Spodoptera litura by double-stranded RNA establishes its role as Bacillus thuringiensis toxin receptor. J. Biol. Chem. (277) 46849-46851.

Ron, D., Walter, P., (2007) Signal integration in the endoplasmic reticulum unfolded protein response. Nat. Rev. Mol. Cell Biol. (8) 519-529.

Schell, J., Van Montagu, M. (1977) The Ti-plasmid of Agrobacterium tumefaciens, a natural vector for the introduction of nif genes in plants. Basic Life Sci. (9) 159-79.

Schroder, M., Kaufman, R.J., (2005) The mammalian unfolded protein response. Ann. Rev. Biochem. (74) 739-789.

Schubert, U., Anton, L.C., Gibbs, J., Norbury, C.C., Yewdell, J.W., Bennink, J.R. (2000) Rapid degradation of a large fraction of newly synthesized proteins by proteasomes. Nature. (404) 770-774.

Shen, X., Ellis, R.E., Sakaki, K., Kaufman, R.J. (2005) Genetic interactions due to constitutive and inducible gene regulation mediated by the unfolded protein response in C. elegans. PLoS Genet. (3) e37

Sidrauski, C., Walter, P. (1997) The transmembrane kinase IRE1p is a site-specific endonuclease that initiates mRNA splicing in the unfolded protein response. Cell. (90) 1031-1039.

Sriburi, R., Jackowski, S., Mori, K., Brewer, J.W. (2004) XBP1: a link between the unfolded protein response, lipid biosynthesis, and biogenesis of the endoplasmic reticulum. J. Cell Biol. (167) 35-41.

Stevens, F.J., Argon, Y., (1999) Protein folding in the ER. Semin. Cell. Dev. Biol. (10) 443-454.

Todd, D.J., Lee, A.H., Glimcher, L.H. (2008) The endoplasmic reticulum stress response in immunity and autoimmunity. Nature Rev. Immunol. (8) 663-674.

Uhlirova, M., Foy, B.D., Beaty, B.J., Olson, K.E., Riddiford, L.M., Jindra, M. (2003) Use of Sindbis virus-mediated RNA interference to demonstrate a conserved role of Broad-Complex in insect metamorphosis. Proc. Natl. Acad. Sci. USA. (100) 15607-15612.

Vembar, S.S., Brodsky, J.L. (2008)One step at a time: endoplasmic reticulum-associated degradation. Nature Rev. Mol. Cell Biol. (12) 944-57.

Volmer, R., van der Ploeg, K., Ron, D. (2013) Membrane lipid saturation activates endoplasmic reticulum unfolded protein response transducers through their transmembrane domains. Proc. Natl. Acad. Sci. USA. (110) 4628-33.

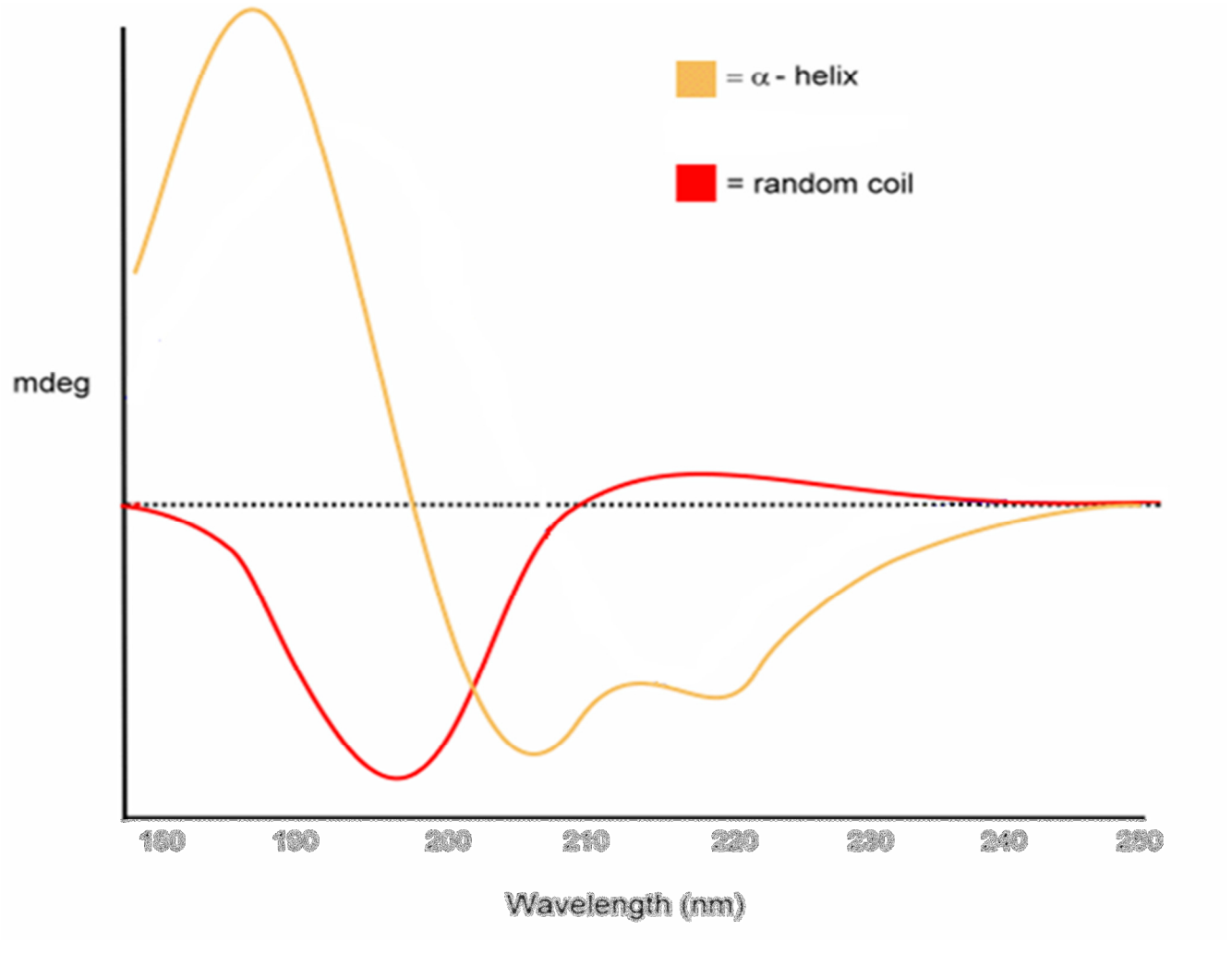
Williams, D.B. (2006) Beyond lectins; the calnexin/calreticulin chaperone system of the endoplasmic reticulum. J. Cell Sci. (119) 615-23.

Xie, Y., Tulenko, T.N. (2006) Activation of the unfolded protein response by cholesterol enrichment in human endothelium. FASEB J. (20) Meeting Abstracts.

Zamore, P.D., Tuschl, T., Sharp, P.A., Bartel, D.P. (2000) RNAi: Double-stranded RNA directs the ATP-dependent cleavage of mRNA at 21 to 23 nucleotide intervals. Cell (101) 25-33.

Zhou, X., Oi, F.M., Scharf, M.E. (2006) Social exploitation of hexamerin: RNAi reveals a major caste-regulatory factor in termites. Proc. Natl. Acad. Sci. USA. (103) 4499-4504.

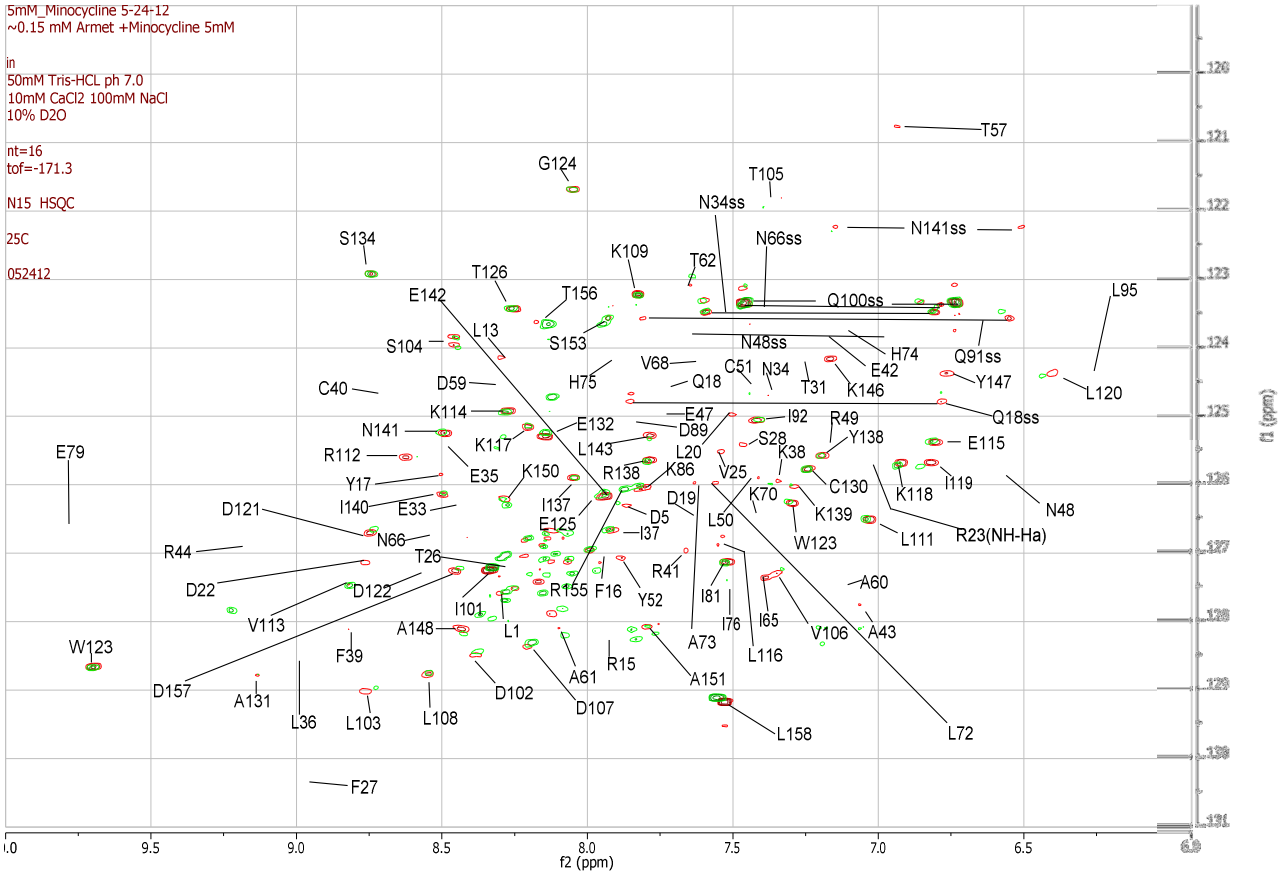
Appendix A - CD Reference Spectrum plus Ligand & Difference NMR Spectra of Armet with Select Ligands



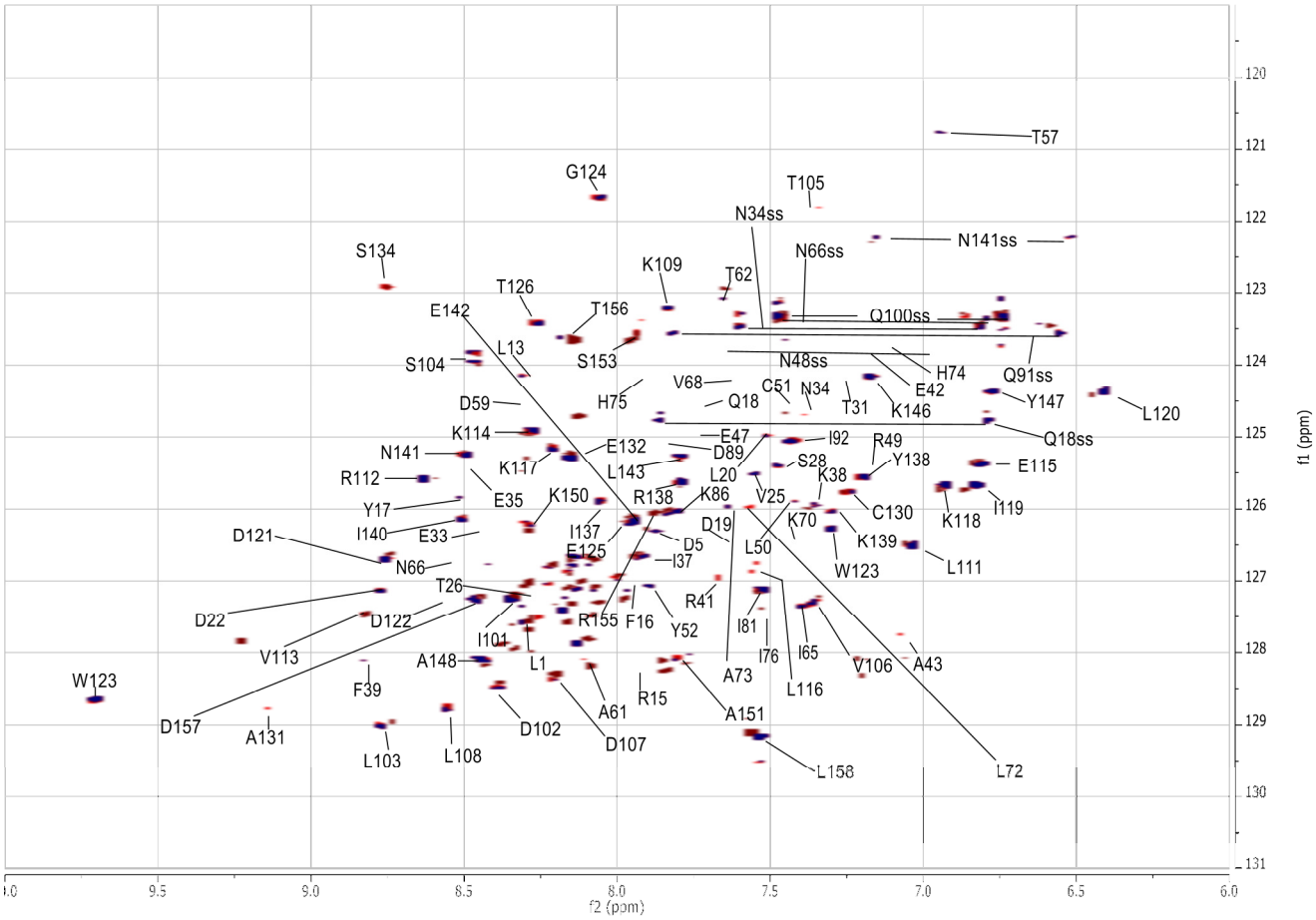
Idealized circular dichroism secondary structure

Image modified and obtained by Google image search from

http://www.proteinchemist.com/cd/cdspec.html

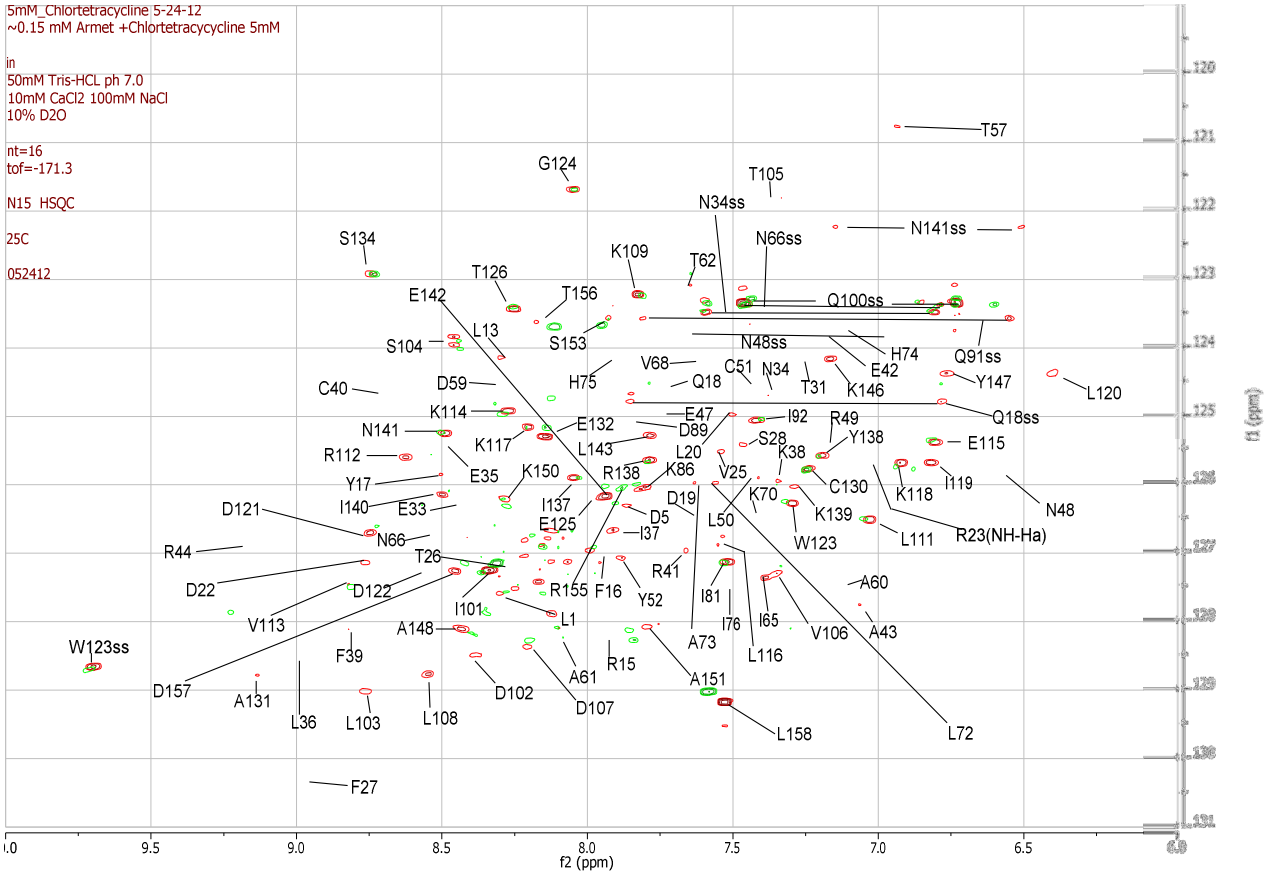


NMR overlay spectrum of human Armet protein and Armet plus 5 mM minocycline. The protein without ligand is indicated in red while the spectrum containing the 5 mM minocycline is shown in green. Changes in residues are determined by their movement in either the hydrogen or nitrogen ppm, F2 and F1 respectively.



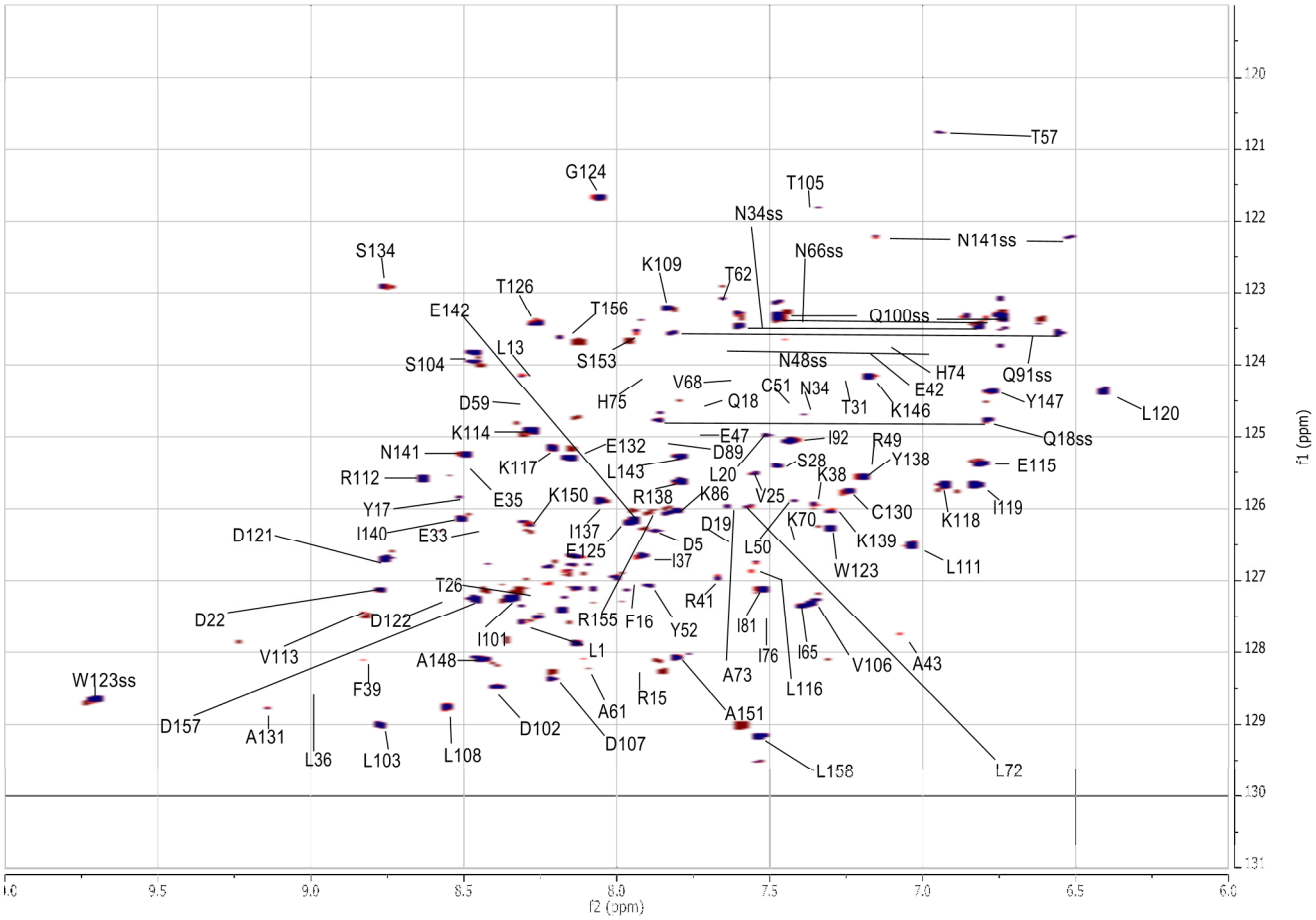
NMR difference spectrum of human Armet protein and Armet plus 5 mM minocycline

The signal indicated in red is indicative of an increased signal without ligand present and a blue signal indicates an increased signal with the ligand present. These changes are identified and qualified in table 1. Changes in residues are determined by their movement in either the hydrogen or nitrogen ppm, F2 and F1 respectively.



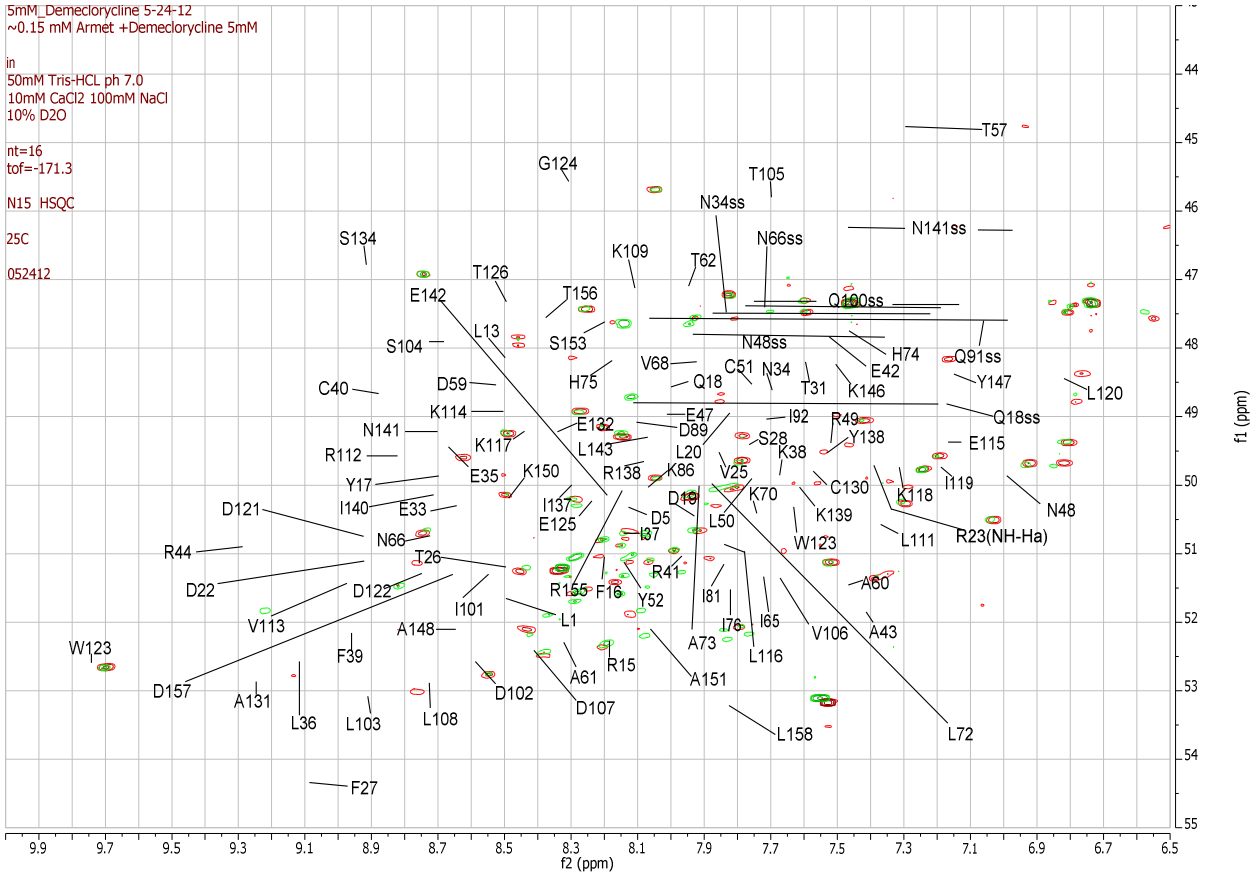
NMR overlay spectrum of human Armet protein and Armet plus 5 mM chlortetracycline

The protein without ligand is indicated in red while the spectrum containing the 5 mM chlortetracycline is shown in green. Changes in residues are determined by their movement in either the hydrogen or nitrogen ppm, F2 and F1 respectively.



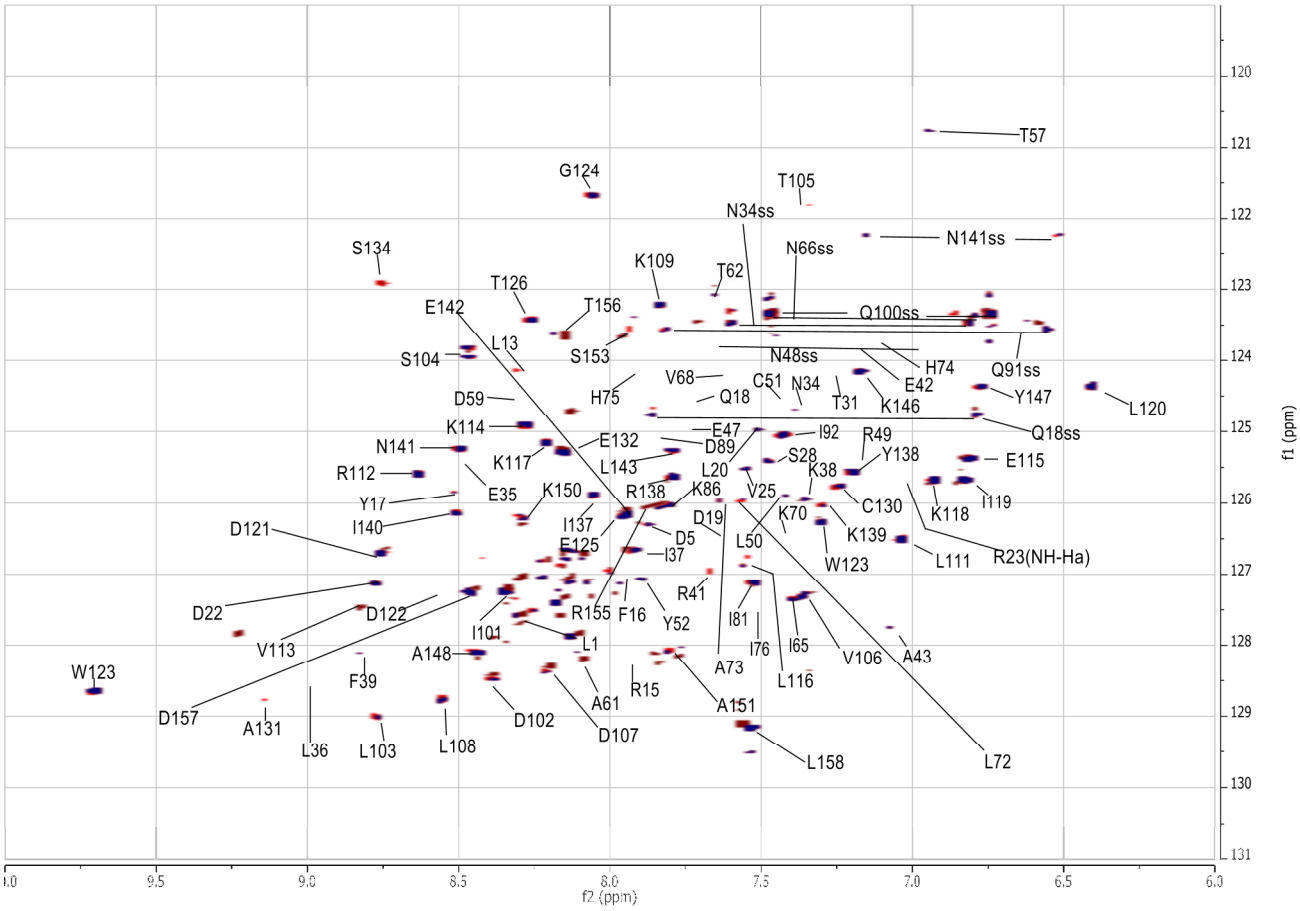
NMR difference spectrum of human Armet protein and Armet plus 5 mM chlortetracycline

The signal indicated in red is indicative of an increased signal without ligand present and a blue signal indicates an increased signal with the ligand present. These changes are identified and qualified in table 1. Changes in residues are determined by their movement in either the hydrogen or nitrogen ppm, F2 and F1 respectively.



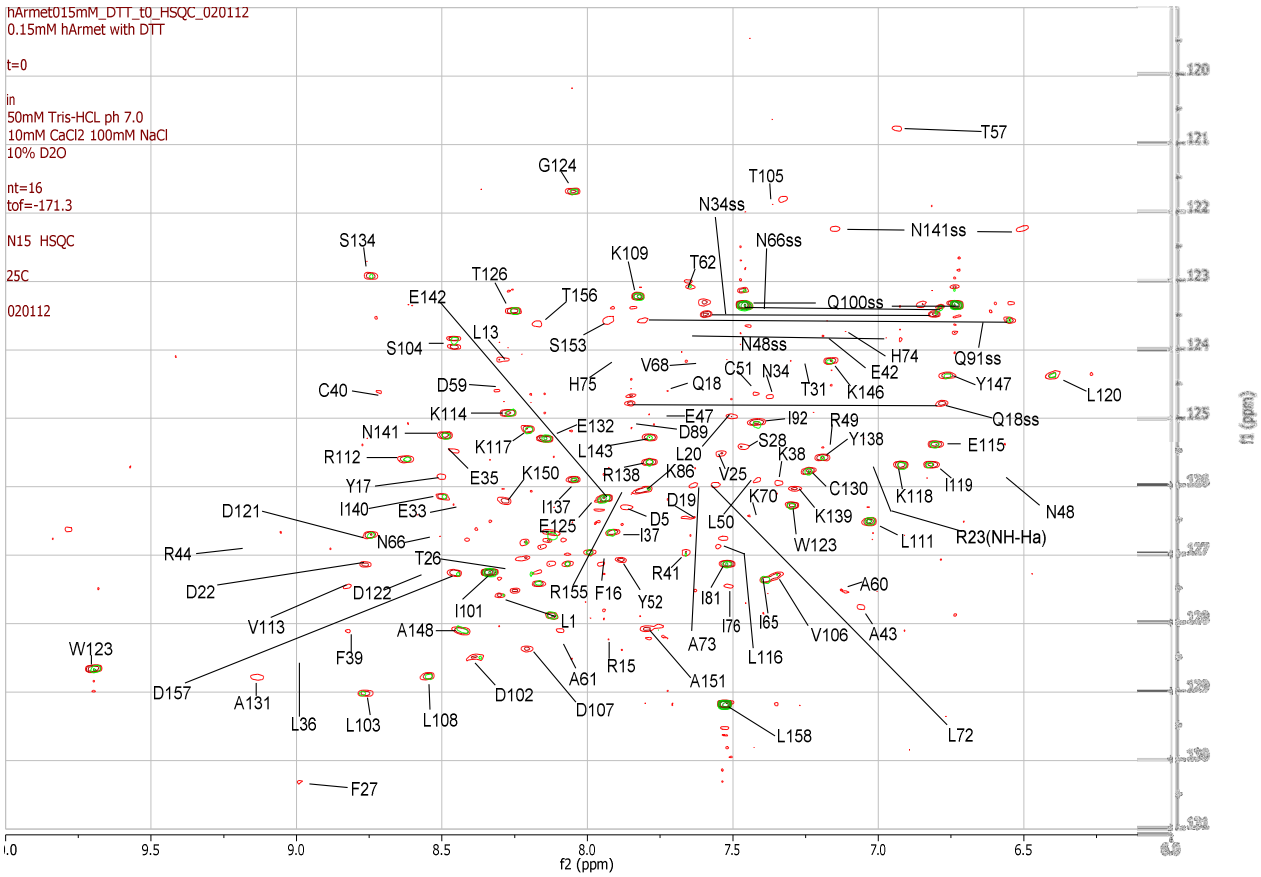
NMR overlay spectrum of human Armet protein and Armet plus 5 mM demeclocycline

The protein without ligand is indicated in red while the spectrum containing the 5 mM demeclocycline is shown in green. Changes in residues are determined by their movement in either the hydrogen or nitrogen ppm, F2 and F1 respectively.



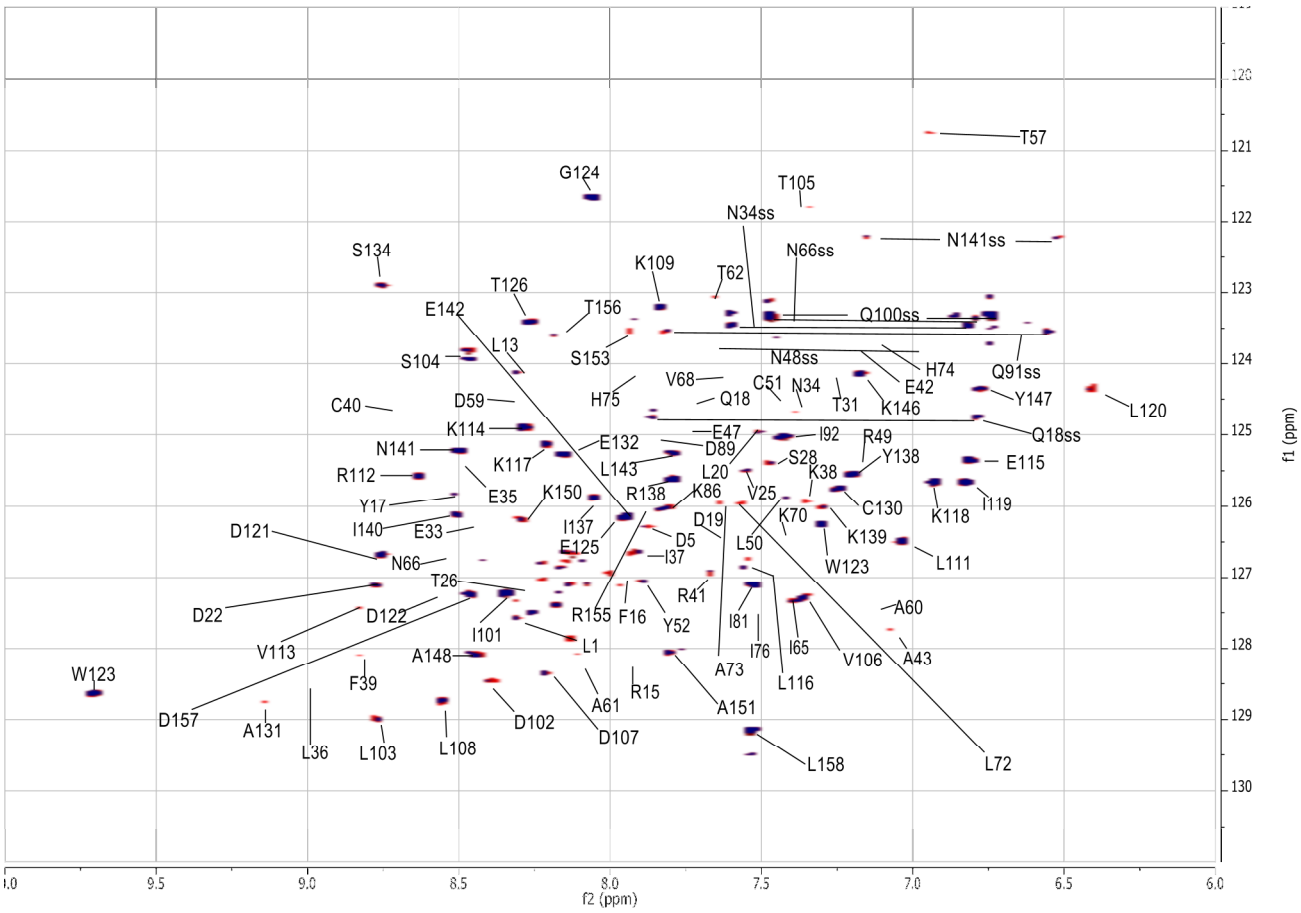
NMR difference spectrum of human Armet protein and Armet plus 5 mM demeclocycline

The signal indicated in red is indicative of an increased signal without ligand present and a blue signal indicates an increased signal with the ligand present. These changes are identified and qualified in table 1. Changes in residues are determined by their movement in either the hydrogen or nitrogen ppm, F2 and F1 respectively.



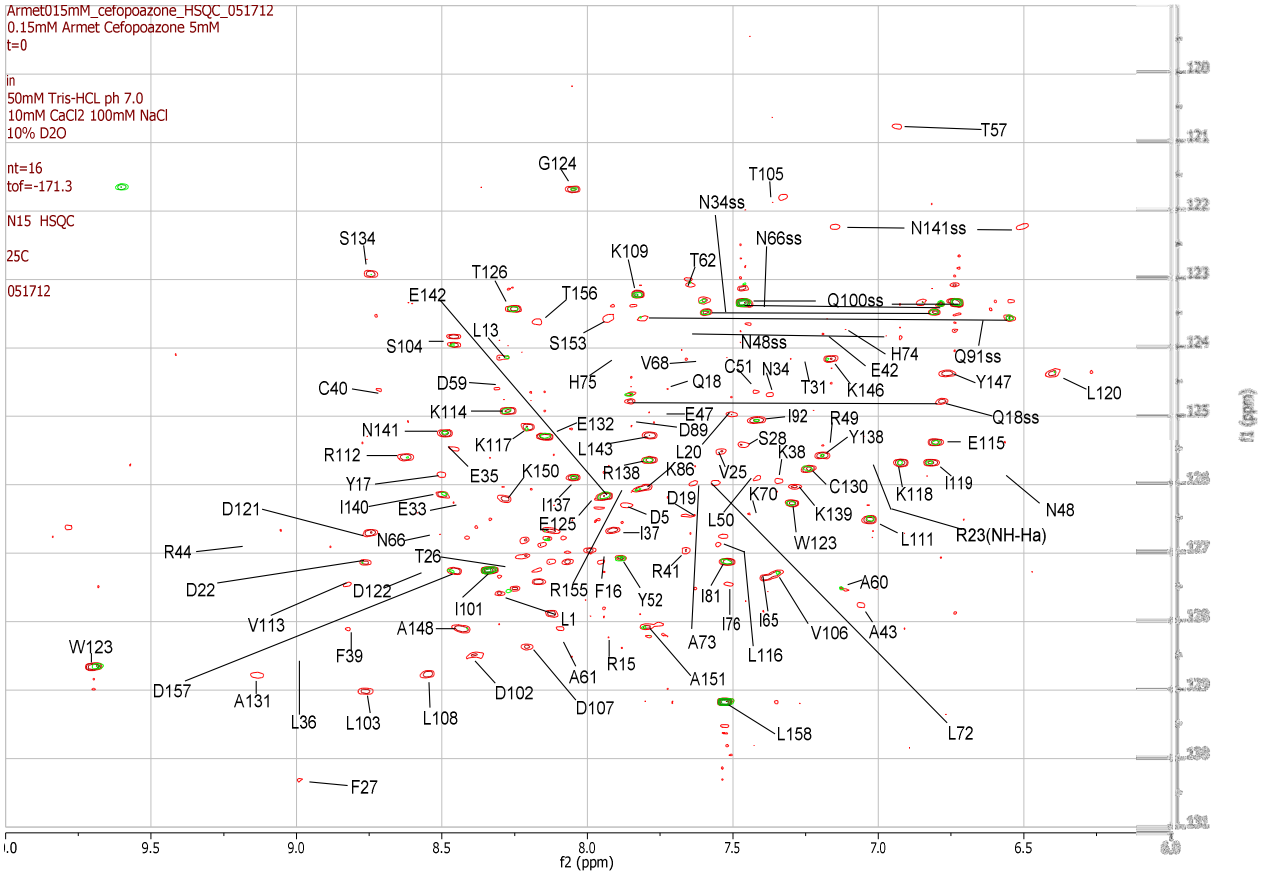
NMR overlay spectrum of human Armet protein and Armet plus 5 mM DTT

The protein without ligand is indicated in red while the spectrum containing the 5 mM DTT is shown in green. Changes in residues are determined by their movement in either the hydrogen or nitrogen ppm, F2 and F1 respectively.



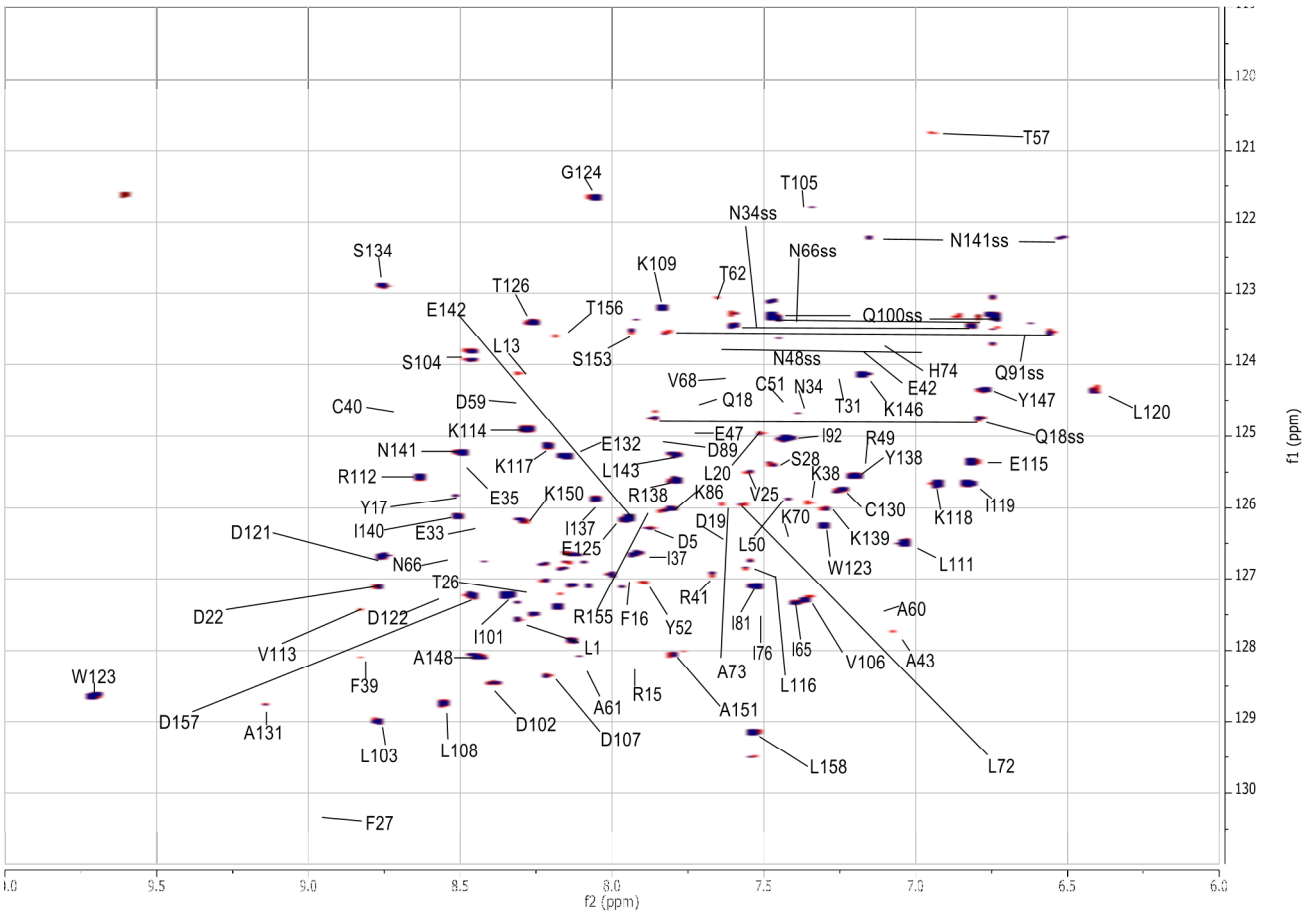
NMR difference spectrum of human Armet protein and Armet plus 5 mM DTT

The signal indicated in red is indicative of an increased signal without ligand present and a blue signal indicates an increased signal with the ligand present. These changes are identified and qualified in table 1. Changes in residues are determined by their movement in either the hydrogen or nitrogen ppm, F2 and F1 respectively.



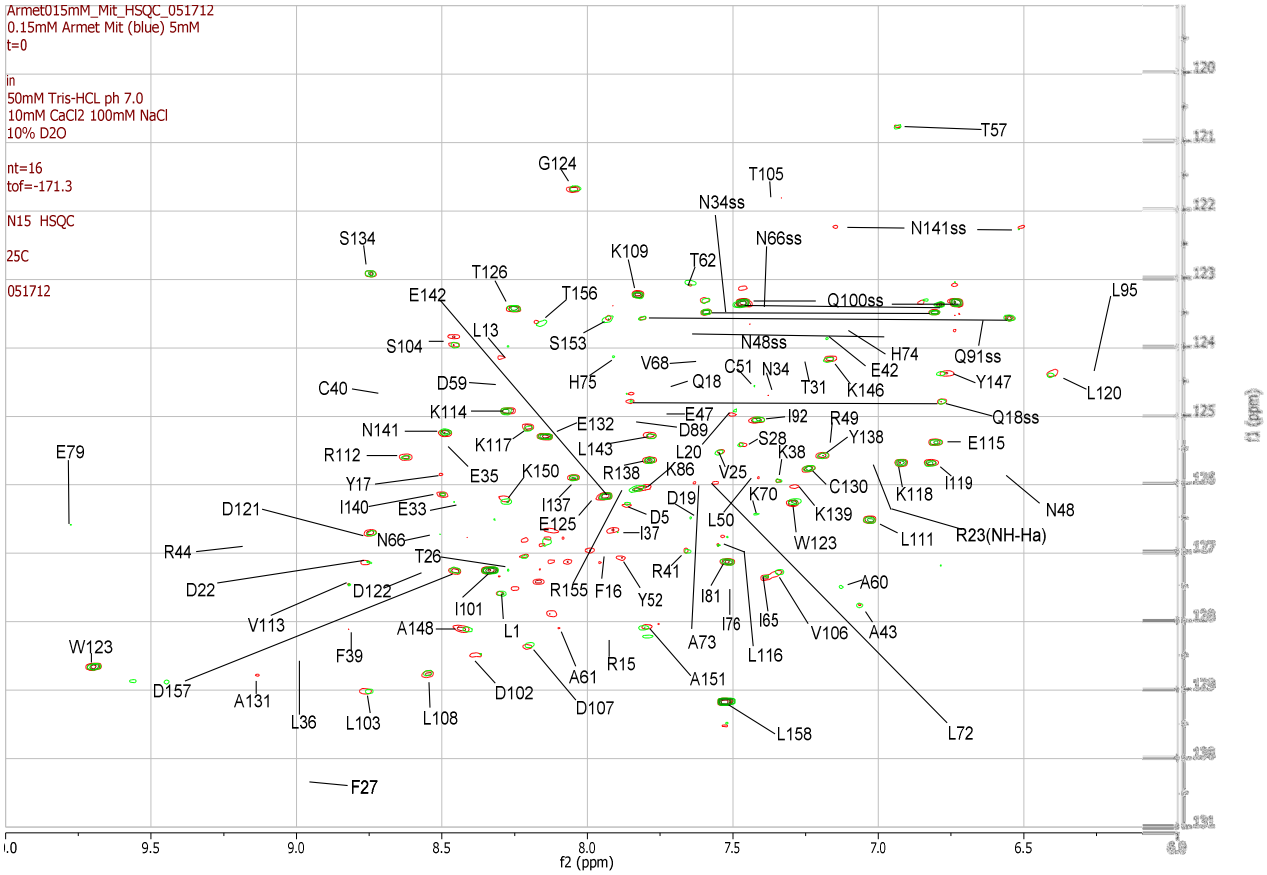
NMR overlay spectrum of human Armet protein and Armet plus 5 mM cefoperazone

The protein without ligand is indicated in red while the spectrum containing the 5 mM cefoperazone is shown in green. Changes in residues are determined by their movement in either the hydrogen or nitrogen ppm, F2 and F1 respectively.



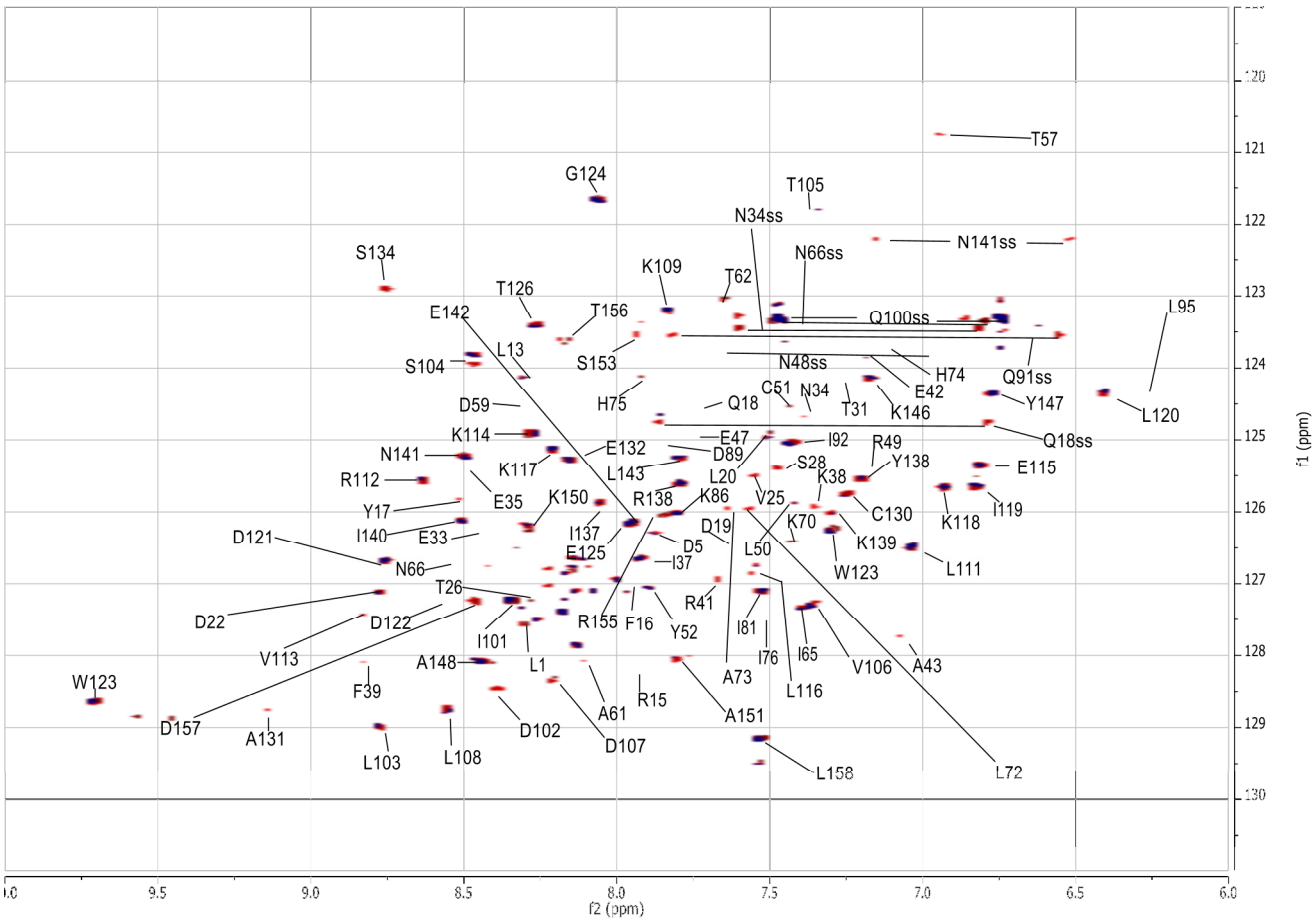
NMR difference spectrum of human Armet protein and Armet plus 5 mM cefoperazone

The signal indicated in red is indicative of an increased signal without ligand present and a blue signal indicates an increased signal with the ligand present. These changes are identified and qualified in table 1. Changes in residues are determined by their movement in either the hydrogen or nitrogen ppm, F2 and F1 respectively.



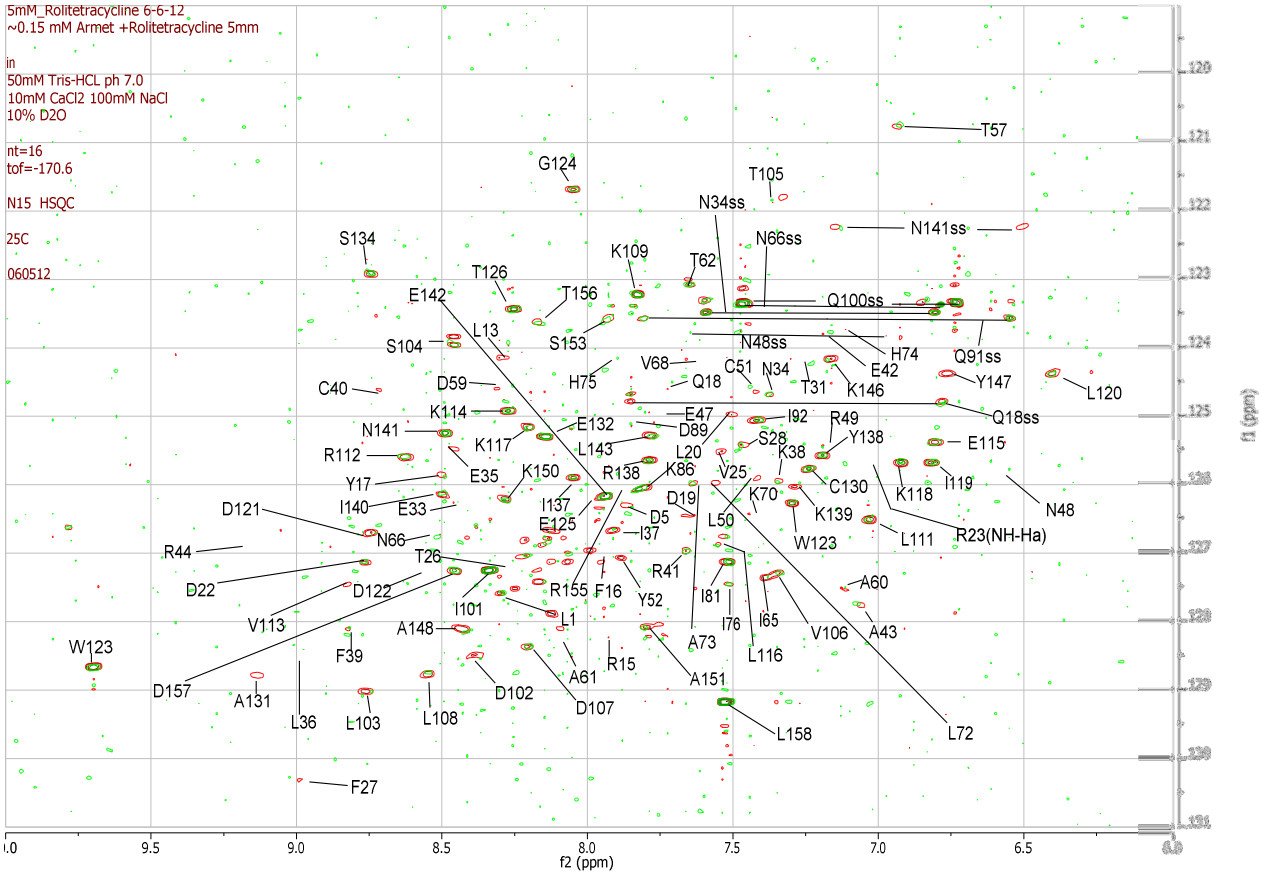
NMR overlay spectrum of human Armet protein and Armet plus 5 mM mitoxanthrone

The protein without ligand is indicated in red while the spectrum containing the 5 mM mitoxanthrone is shown in green. Changes in residues are determined by their movement in either the hydrogen or nitrogen ppm, F2 and F1 respectively.



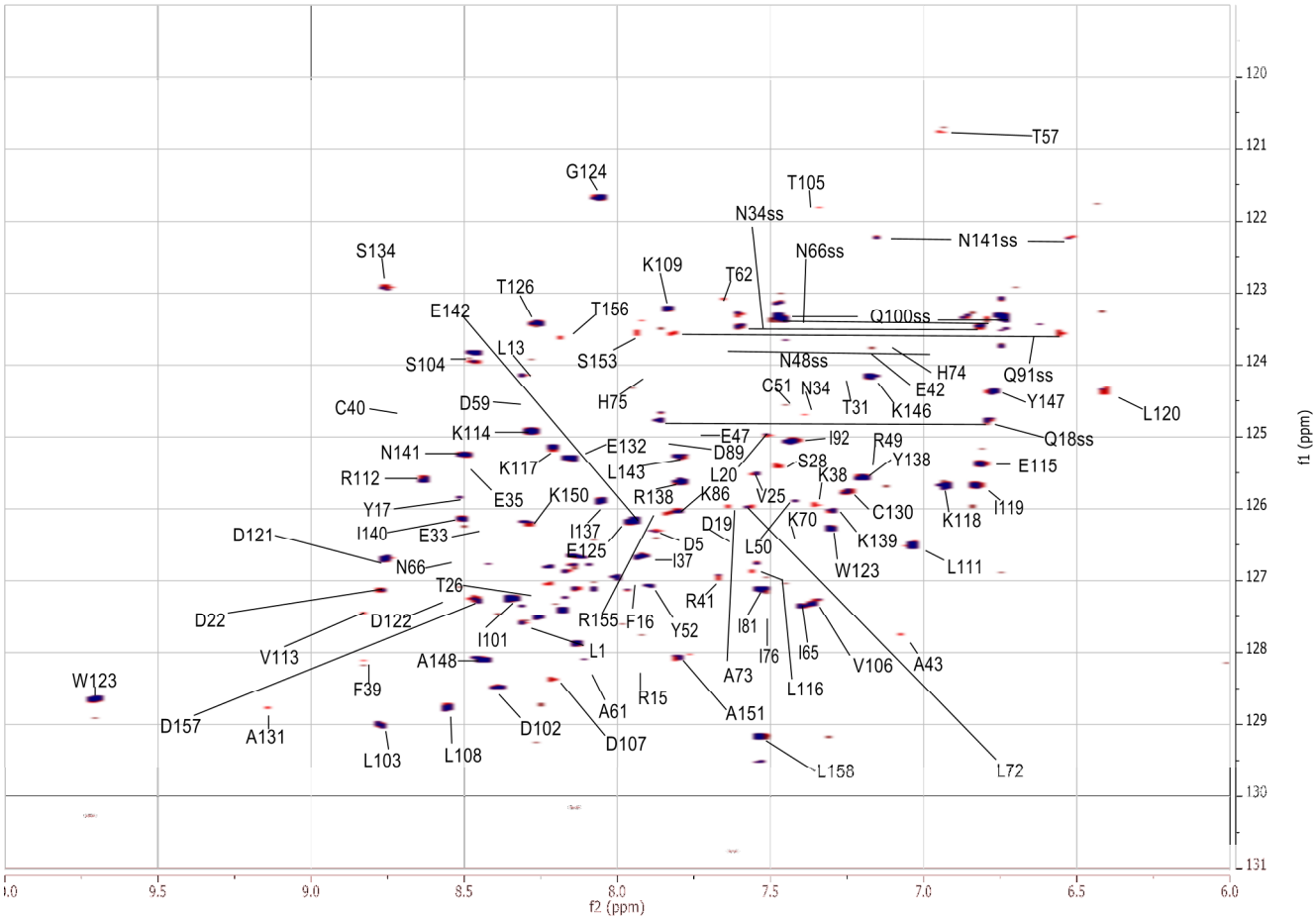
NMR difference spectrum of human Armet protein and Armet plus 5 mM mitoxanthrone

The signal indicated in red is indicative of an increased signal without ligand present and a blue signal indicates an increased signal with the ligand present. These changes are identified and qualified in table 1. Changes in residues are determined by their movement in either the hydrogen or nitrogen ppm, F2 and F1 respectively.



NMR overlay spectrum of human Armet protein and Armet plus 5 mM rolitetracycline

The protein without ligand is indicated in red while the spectrum containing the 5 mM rolitetracycline is shown in green. Changes in residues are determined by their movement in either the hydrogen or nitrogen ppm, F2 and F1 respectively.



NMR difference spectrum of human Armet protein and Armet plus 5 mM rolitetracycline

The signal indicated in red is indicative of an increased signal without ligand present and a blue signal indicates an increased signal with the ligand present. These changes are identified and qualified in table 1. Changes in residues are determined by their movement in either the hydrogen or nitrogen ppm, F2 and F1 respectively.

Appendix B - Human UPR List

List of human UPR genes which indicate description, human and aphid gene ID, alternative names, and official gene name of the 91 components utilized in Chapter 2.

Unfolded Protein Response: Genes, Proteins, & Functions

Official Name	AKA	Additional Roles	Comments	Human Gene ID	Aphid Gene ID	Gene Name
Primary UPR Role: ER Transducers						
Inositol-requiring enzyme-1	IRE1; IRE1P; IRE1a; hIRE1pCG4583; dire-1; dIRE1; DmaCG4583; Ire; ire-1; Ire-1; IRE-like; ire1; IRE1	Transcription factor	The protein encoded by this gene is the ER to nucleus signalling 1 protein, a human homologue of the yeast Ire1 gene product. This protein possesses intrinsic kinase activity and an endoribonuclease activity and it is important in altering gene expression as a response to endoplasmic reticulum-based stress signals.	2081	100164567	ERN1
Activating transcription factor 6	ATF6A		accumulation of misfolded pro causes proteolytic cleavage of ATF6 - cytosolic portion moves to nucleus to act as a transcription factor to increase transcription of ER Chaperones	22926	100571138	ATF6
Eukaryotic translation initiation factor 2 alpha kinase 3	Pek; Perk; A1427929	Translation Regulation	Patients with mutations in this gene develop Wolcott-Rallison syndrome	9451	100166491	EIF2AK3
Primary UPR Role: Transcription Factor						
X-box binding protein 1	XBP2; TREB5; XBP-1		transcription factor that regulates MHC class II genes by binding to a promoter element referred to as an X box. This gene product is a bZIP protein, which was also identified as a cellular transcription factor that binds to an enhancer in the promoter of the T cell leukemia virus type 1 promoter. It may increase expression of viral proteins by acting as the DNA binding partner of a viral transactivator. It has been found that upon accumulation of unfolded proteins in the endoplasmic reticulum (ER), the mRNA of this gene is processed to an active form by an unconventional splicing mechanism that is mediated by the endonuclease inositol-requiring enzyme 1 (IRE1). The resulting loss of 26 nt from the spliced mRNA causes a frame-shift and an isoform XBP1(S), which is the functionally active transcription factor. The isoform encoded by the unspliced mRNA, XBP1(U), is constitutively expressed, and thought to function as a negative feedback regulator of XBP1(S), which shuts off transcription of target genes during the recovery phase of ER stress. A pseudogene of XBP1 has been identified and localized to chromosome 5.	7494	100574802	XBP1
Activating transcription factor 4 (tax-responsive enhancer element B67)	CREB2; TXREB; CREB-2; TAXREB67		induces CHOP gene - JNK inactivation downregulates expression	468	100161513	ATF4
cAMP responsive element binding protein 3	LZIP; LUMAN		transcription factor that is a member of the leucine zipper family of DNA binding proteins. binds to the cAMP-response element and regulates cell proliferation. interacts with host cell factor C1, which also associates with the herpes simplex virus (HSV) protein VP16 that induces transcription of HSV immediate-early genes. This protein and VP16 both bind to the same site on host cell factor C1. proposed interaction between this protein and host cell factor C1 plays a role in the establishment of latency during HSV infection. This protein also plays a role in leukocyte migration, tumor suppression, and endoplasmic reticulum stress-associated protein degradation.	10488	100166610	CREB3
cAMP responsive element binding protein 3-like 3	CREBH; CREB-H; HYST1481		This gene encodes a member of the basic-leucine zipper family and the AMP-dependent transcription factor family. The encoded protein is localized to the endoplasmic reticulum and acts as a transcription factor activated by cyclic AMP stimulation. The encoded protein binds the cyclic AMP response element (CRE) and the box-B element and has been linked to acute inflammatory response, hepatocellular carcinoma, triglyceride metabolism, and hepcidin expression.	84699	100166610	CREB3L3
activating transcription factor 6 beta	G13; CREBL1; CREB-RP		transcription factor in the (UPR) pathway during ER stress. Either as a homodimer or as a heterodimer with ATF6-alpha, the protein binds to the ER stress response element, interacting with nuclear transcription factor Y to activate UPR target genes. normally found in the membrane of the ER; however, under ER stress, the N-terminal cytoplasmic domain is cleaved from the rest of the protein and translocates to the nucleus. Two transcript variants encoding different isoforms have been found for this gene.	1388	100571138	ATF6B
TNF receptor-associated factor 2	TRAP; TRAP3; MGC45012		The protein encoded by this gene is a member of the TNF receptor associated factor (TRAF) protein family. TRAF proteins associate with, and mediate the signal transduction from members of the TNF receptor superfamily. This protein directly interacts with TNF receptors, and forms a heterodimeric complex with TRAF1. This protein is required for TNF-alpha-mediated activation of MAPK8/JNK and NF-kappaB. The protein complex formed by this protein and TRAF1 interacts with the inhibitor-of-apoptosis proteins (IAPs), and functions as a mediator of the anti-apoptotic signals from TNF receptors. The interaction of this protein with TRADD, a TNF receptor associated apoptotic signal transducer, ensures the recruitment of IAPs for the direct inhibition of caspase activation. BIRC2/c-IAP1, an apoptosis inhibitor possessing ubiquitin ligase activity, can ubiquitinate and induce the degradation of this protein, and thus potentiate TNF-induced apoptosis. Multiple alternatively spliced transcript variants have been found for this gene, but the biological validity of only one transcript has been determined	7186	100159489	TRAF2
Primary UPR Role: Translation Regulation						
[illegible]	[illegible]		[illegible]	83939	100167878	EIF2A
initiation factor eIF-2B alpha subunit				1965	100168717	eif2
Primary UPR Role: Anti-Apoptosis						
Armet	ARP; MANF	Protein Binding	localized in the endoplasmic reticulum (ER) and golgi, and is also secreted. Reducing expression of this gene increases susceptibility to ER stress-induced death and promotes cell proliferation	7873	100167188	MANF
heat shock 70kDa protein 5 (glucose-regulated protein, 78kDa)	BiP; MIF2; GRP78		The protein encoded by this gene is a member of the heat shock protein 70 (HSP70) family. It is localized in the lumen of the endoplasmic reticulum (ER), and is involved in the folding and assembly of proteins in the ER. As this protein interacts with many ER proteins, it may play a key role in monitoring protein transport through the cell	3309	100163748	HSPA5
CCAAT/enhancer binding protein (CEBP), delta	CELF; CRP3; C/EBP-delta; NF-IL6-beta	Transcription factor	The protein encoded by this intronless gene is a bZIP transcription factor which can bind as a homodimer to certain DNA regulatory regions. It can also form heterodimers with the related protein CEBP-alpha. The encoded protein is important in the regulation of genes involved in immune and inflammatory responses, and may be involved in the regulation of genes associated with activation and/or differentiation of macrophages	1052	100162848	Cebpd
BCL2-associated X Protein	BCL2L4		form hetero- or homodimers and act as anti- or pro-apoptotic regulators. This protein forms a heterodimer with BCL2, and functions as an apoptotic activator. This protein is reported to interact with, increase the opening of the mitochondrial voltage-dependent anion channel (VDAC) - leads to the loss in membrane potential releases cytochrome c. The expression of this gene is regulated by the tumor suppressor P53 and has been shown to be involved in P53-mediated apoptosis.	581	100164650	BAX

Primary UPR Role: Apoptosis						
Endoplasmic reticulum to nucleus signalling 2	Ern2; Ire2; Ire1b	Transcription factor	The protein encoded by this gene is the ER to nucleus signalling 2 protein, a human homologue of the yeast Ire1 gene product. This protein possesses intrinsic kinase activity and an endoribonuclease activity and it is important in altering gene expression as a response to endoplasmic reticulum-based stress signals.	26918	100164567	ERN2
mitogen-activated protein kinase 8	JNK; JNK1; PRKM8; SAPK1; JNK-46; JNK1A2; SAPK1c; JNK21B/2		MAP kinase - serine/threonine specific member of the MAP kinase family. MAP kinases act as an integration point for multiple biochemical signals, and are involved in a wide variety of cellular processes such as proliferation, differentiation, transcription regulation and development	5599	100163276	MAPK8
mitogen-activated protein kinase 9	JNK2; Prkm9; AB51083; p54aSAPK		MAP kinase - related to cytochrom c apoptosis p/w	5601	100163276	MAPK9
Calreticulin	RO; CRT; SSA; cC1qR	Transcription factor/ Protein Folding / Unfolded Protein Binding	acts as a major Ca(2+)-binding (storage) protein in the lumen of the endoplasmic reticulum. It is also found in the nucleus, suggesting that it may have a role in transcription regulation. Calreticulin binds to the synthetic peptide KLGFFKR, which is almost identical to an amino acid sequence in the DNA-binding domain of the superfamily of nuclear receptors. Systemic lupus erythematosus is associated with increased autoantibody titers against calreticulin but calreticulin is not a Ro/SS-A antigen.	811	100161399	CALR
Protein phosphatase 1, regulatory subunit 15A	GADD34	Translation Regulation	This gene is a member of a group of genes whose transcript levels are increased following stressful growth arrest conditions and treatment with DNA-damaging agents. The induction of this gene by ionizing radiation occurs in certain cell lines regardless of p53 status, and its protein response is correlated with apoptosis following ionizing radiation.	23645	**N/A**	PPP1R15A
Heat shock protein 1B	Hsp70; hsp68; Hsp70-1; Hsp70.1	Unfolded Protein Binding	In conjunction with other heat shock proteins, this protein stabilizes existing proteins against aggregation and mediates the folding of newly translated proteins in the cytosol and in organelles. It is also involved in the ubiquitin-proteasome pathway through interaction with the AU-rich element RNA-binding protein 1. The gene is located in the major histocompatibility complex class III region, in a cluster with two closely related genes which encode similar proteins	15511	100163748	HSPA1B
HtrA serine peptidase 2	OMI; PARK13; PRSS25	Unfolded Protein Binding / ERAD	serine protease localized in the endoplasmic reticulum and interacts with an alternatively spliced form of mitogen-activated protein kinase 14. The protein has also been localized to the mitochondria with release to the cytosol following apoptotic stimulus, thought to induce apoptosis by binding the apoptosis inhibitory protein baculoviral IAP repeat-containing 4. Nuclear localization of this protein has also been observed. Alternate splicing of this gene results in two transcript variants encoding different isoforms.	27429	100169637	HTRA2
Valosin containing protein	p97; TERA; ALS14; IBMPFD	Unfolded Protein Binding / ERAD	member of a family that includes putative ATP-binding proteins involved in vesicle transport and fusion, 26S proteasome function, and assembly of peroxisomes. This protein, as a structural protein, is associated with clathrin, and heat-shock protein Hsc70, to form a complex. It has been implicated in a number of cellular events that are regulated during mitosis, including homotypic membrane fusion, spindle pole body function, and ubiquitin-dependent protein degradation.	7415	100164710	VCP
mitogen-activated protein kinase 10	JNK3; JNK3A; PRKM10; SAPK1b; p493F12; p54bSAPK		MAP kinase - interacts with Beta-arrestin2, stimulated by MAPK4. This protein is a neuronal-specific form of c-Jun N-terminal kinases (JNKs). Through its phosphorylation and nuclear localization, this kinase plays regulatory roles in the signaling pathways during neuronal apoptosis.	5602	100163276	MAPK10
Primary UPR Role: Cholesterol Regulation						
membrane-bound transcription factor peptidase, site 2	S2P; IFAP; KFSDX; BRESEK	ERAD	This gene encodes a intramembrane zinc metalloprotease, which is essential in development. This protease functions in the signal protein activation involved in sterol control of transcription and the ER stress response. Mutations in this gene have been associated with ichthyosis follicularis with atrichia and photophobia (IFAP syndrome); IFAP syndrome has been quantitatively linked to a reduction in cholesterol homeostasis and ER stress response	51360	100162197	MBTPS2
membrane-bound transcription factor peptidase, site 1	S1P; PCSK8; SKI-1	ERAD / Transcription Factor	central role in the regulation of lipid metabolism in cells. sterol-regulated subtilisin-like serine protease that cleaves ER membrane-bound sterol regulatory element-binding proteins, a run that initiates the 2-step proteolytic process. transcriptionally active fragments of SREBPs are released from the membrane for translocation to the nucleus. integral membrane in ER, bulk located in the ER lumen. synthesized inactive preproprotein that is self-activated by an intramolecular cleavage that generating mature proteins.	8720	100167456	MBTPS1
insulin induced gene 1	CL6; CL-6		Oxysterols regulate cholesterol homeostasis through the liver X receptor (LXR)- and sterol regulatory element-binding protein (SREBP)-mediated signaling pathways. insulin-induced gene. It encodes an (ER) membrane protein that plays a critical role in regulating cholesterol concentrations in cells. This protein binds to the sterol-sensing domains of SREBP cleavage-activating protein (SCAP) and HMG CoA reductase, and is essential for the sterol-mediated trafficking of the two proteins	3638	100169168	INSIG1
insulin induced gene 2	ubiquitin carboxyl-terminal hydrolase calypso-like		The protein encoded by this gene is highly similar to the protein product encoded by gene INSIG1. Both INSIG1 protein and this protein are endoplasmic reticulum proteins that block the processing of sterol regulatory element binding proteins (SREBPs) by binding to SREBP cleavage-activating protein (SCAP), and thus prevent SCAP from escorting SREBPs to the Golgi.	51141	100169168	INSIG2
sterol regulatory element binding transcription factor 2	SREBP2; bHLHd2		This gene encodes a ubiquitously expressed transcription factor that controls cholesterol homeostasis by stimulating transcription of sterol-regulated genes. The encoded protein contains a basic helix-loop-helix-leucine zipper (bHLH-Zip) domain.	6721	100164413	SREBF2
Primary UPR Role: ER Associated Degradation						
autocrine motility factor receptor, E3 ubiquitin protein ligase	GP78; RNF45	Ubiquitination	This locus encodes a glycosylated transmembrane receptor. Its ligand, autocrine motility factor, is a tumor motility-stimulating protein secreted by tumor cells. The encoded receptor is also a member of the E3 ubiquitin ligase family of proteins. It catalyzes ubiquitination and endoplasmic reticulum-associated degradation of specific proteins.	267	100573944	AMFR
homocysteine-inducible, endoplasmic reticulum stress-inducible, ubiquitin-like domain member 1	SUP; HERP; Mif1	Ubiquitination	The accumulation of unfolded proteins in the endoplasmic reticulum (ER) triggers the ER stress response. This response includes the inhibition of translation to prevent further accumulation of unfolded proteins, the increased expression of proteins involved in polypeptide folding, known as the unfolded protein response (UPR), and the destruction of misfolded proteins by the ER-associated protein degradation (ERAD) system. This gene may play a role in both UPR and ERAD. Its expression is induced by UPR and it has an ER stress response element in its promoter region while the encoded protein has an N-terminal ubiquitin-like domain which may interact with the ERAD system. This protein has been shown to interact with presenilin proteins and to increase the level of amyloid-beta protein following its overexpression.	9709	100165580	HERPUD1
sel-1 suppressor of lin-12-like (C. elegans)	SEL1L1; PRO1063; SEL1-LIKE	Ubiquitination	The protein encoded by this gene is part of a protein complex required for the retrotranslocation or dislocation of misfolded proteins from the endoplasmic reticulum lumen to the cytosol, where they are degraded by the proteasome in a ubiquitin-dependent manner. Alternatively spliced transcript variants encoding different isoforms have been found for this gene	6400	100162766	SEL1L

SEC62 homolog (S. cerevisiae)	HTP1; TP-1; Dtrp1; TLOC1	Ubiquitination	The Sec61 complex is the central component of the protein translocation apparatus of the endoplasmic reticulum (ER) membrane. The protein encoded by this gene and SEC63 protein are found to be associated with ribosome-free SEC61 complex. It is speculated that Sec61-Sec62-Sec63 may perform post-translational protein translocation into the ER. The Sec61-Sec62-Sec63 complex might also perform the backward transport of ER proteins that are subject to the ubiquitin-proteasome-dependent degradation pathway. The encoded protein is an integral membrane protein located in the rough ER.	7095	100168981	SEC62
UBX domain protein 4	UBXD2; UBXDC1; ersalin	Ubiquitination	UBXD2 is an integral membrane protein of the endoplasmic reticulum (ER) that binds valosin-containing protein (VCP; MIM 601023) and promotes ER-associated protein degradation (ERAD).	23190	100572415	UBXN4
derlin 1	DER1; DER-1		member of the derlin family. Members of this family participate in the ER-associated degradation response and retrotranslocate misfolded or unfolded proteins from the ER lumen to the cytosol for proteasomal degradation. This protein recognizes substrate in the ER and works in a complex to retrotranslocate it across the ER membrane into the cytosol. This protein may select cystic fibrosis transmembrane conductance regulator protein (CFTR) for degradation as well as unfolded proteins in Alzheimer's disease.	79139	100160408	DERL1
derlin 2	FLANa; F-LANa; CGI-101; F-LAN-1		Proteins that are unfolded or misfolded in the endoplasmic reticulum (ER) must be refolded or degraded to maintain the homeostasis of the ER. DERL2 is involved in the degradation of misfolded glycoproteins in the ER.	51009	100160760	DERL2
F-box protein 6	FBG2; FBS2; FBX6; Fbx6b		This gene encodes a member of the F-box protein family which is characterized by an approximately 40 amino acid motif, the F-box. The F-box proteins constitute one of the four subunits of the ubiquitin protein ligase complex called SCFs (SKP1-cullin-F-box), which function in phosphorylation-dependent ubiquitination. The F-box proteins are divided into 3 classes: Fbws containing WD-40 domains, Fbls containing leucine-rich repeats, and Fbxs containing either different protein-protein interaction modules or no recognizable motifs. The protein encoded by this gene belongs to the Fbxs class, and its C-terminal region is highly similar to that of rat NFB42 (neural F-box 42 kDa) which may be involved in the control of the cell cycle.	26270	100161254	FBXO6
HtrA serine peptidase 4			member of the HtrA family of proteases. The encoded protein contains a putative signal peptide, an insulin growth factor binding domain, a Kazal protease inhibitor domain, a conserved trypsin domain and a PDZ domain. Based on studies on other related family members, this enzyme may function as a secreted oligomeric chaperone protease to degrade misfolded secretory proteins. Other human HtrA proteins have been implicated in arthritis, tumor suppression, unfolded stress response, apoptosis, and aging.	203100	100570848	HTRA4
nuclear protein localization 4 homolog (S. cerevisiae)	NPL4		The protein encoded by this gene forms a complex with two other proteins, ubiquitin fusion degradation 1 like protein and valosin-containing protein, and this complex is necessary for the degradation of ubiquitinated proteins. In addition, this complex controls the disassembly of the mitotic spindle and the formation of a closed nuclear envelope after mitosis. Mutations in this gene have been associated with Catch 22 syndrome as well as cardiac and craniofacial defects. Alternative splicing results in multiple transcript variants encoding different isoforms. A related pseudogene has been identified on chromosome 18.	55666	100163186	NPLOC4
nucleobindin 1	NUC; CALNUC		This gene encodes a member of a small calcium-binding EF-hand protein family. The encoded protein is thought to have a key role in Golgi calcium homeostasis and Ca(2+)-regulated signal transduction events.	4924	100164693	NUCB1
osteosarcoma amplified 9, endoplasmic reticulum lectin	OS-9; ERLEC2		This gene encodes a protein that is highly expressed in osteosarcomas. This protein binds to the hypoxia-inducible factor 1 (HIF-1), a key regulator of the hypoxic response and angiogenesis, and promotes the degradation of one of its subunits. Alternate transcriptional splice variants, encoding different isoforms, have been characterized.	10956	100168261	OS9
synovial apoptosis inhibitor 1, synoviolin	DERS; HRD1		This gene encodes a protein involved in endoplasmic reticulum (ER)-associated degradation. The encoded protein removes unfolded proteins, accumulated during ER stress, by retrograde transport to the cytosol from the ER. This protein also uses the ubiquitin-proteasome system for additional degradation of unfolded proteins. Sequence analysis identified two transcript variants that encode different isoforms.	84447	100164346	SYVN1
Primary UPR Role: Folding Quality Control						
ER degradation enhancer, mannosidase alpha-like 1	EDEM	ERAD		9695	100161586	EDEM1
stress-associated endoplasmic reticulum protein 1		Translation Regulation		27230	100161196	SERP1
ER degradation enhancer, mannosidase alpha-like 3	C1orf22	Ubiquitination	Quality control in the endoplasmic reticulum (ER) ensures that only properly folded proteins are retained in the cell through recognition and degradation of misfolded or unassembled proteins. EDEM3 belongs to a group of proteins that accelerate degradation of misfolded glycoproteins in the ER.	80267	100161586	EDEM3
glucosidase, alpha; neutral AB	G2AN; GLUII			23193	100168783	GANAB
glucosidase, alpha; neutral C			Glycosyl hydrolase enzymes hydrolyze the glycosidic bond between two or more carbohydrates, or between a carbohydrate and a non-carbohydrate moiety. This gene encodes a member of glycosyl hydrolases family 31. This enzyme hydrolyses terminal, non-reducing 1,4-linked alpha-D-glucose residues and releases alpha-D-glucose. This is a key enzyme in glycogen metabolism and its gene localizes to a chromosomal region (15q15) that is associated with susceptibility to diabetes.	2595	100168783	GANC
protein kinase C substrate 80K-H	PCLD; PLD1; C19P1; PKCSH; ACE-82		This gene encodes the beta-subunit of glucosidase II, an N-linked glycan-processing enzyme in the endoplasmic reticulum (ER). This protein is an acidic phospho-protein known to be a substrate for protein kinase C. Mutations in this gene have been associated with the autosomal dominant polycystic liver disease (PCLD). Alternatively spliced transcript variants encoding distinct isoforms have been observed.	5589	100161041	PRKCSH
ribophorin 1	OST1; RBPH1		This gene encodes a type I integral membrane protein found only in the rough endoplasmic reticulum. The encoded protein is part of an N-oligosaccharyl transferase complex that links high mannose oligosaccharides to asparagine residues found in the Asn-X-Ser/Thr consensus motif of nascent polypeptide chains. This protein forms part of the regulatory subunit of the 26S proteasome and may mediate binding of ubiquitin-like domains to this proteasome.	6184	100167805	RPN1
UDP-glucose glycoprotein glucosyltransferase 2	UGT2; HUGT2; UGCGL2		UDP-glucose:glycoprotein glucosyltransferase (UGT) is a soluble protein of the endoplasmic reticulum (ER) that selectively reglucosylates unfolded glycoproteins, thus providing quality control for protein transport out of the ER.	55757	100162033	UGGT2
Primary UPR Role: Heat Shock Protein						

DnaJ (Hsp40) homolog, subfamily B, member 2	HSJ1; DSMA5; HSJ-1; HSPF3	Unfolded Protein Binding / Protein Folding	This gene is almost exclusively expressed in the brain, mainly in the neuronal layers. It encodes a protein that shows sequence similarity to bacterial DnaJ protein and the yeast homologs. In bacteria, this protein is implicated in protein folding and protein complex dissociation. Alternatively spliced transcript variants encoding different isoforms have been described for this gene.	3300	100165812	DNAJB2
DnaJ (Hsp40) homolog, subfamily B, member 9	MDG1; ERdj4; MDG-1; MSTP049; MSTP049	Unfolded Protein Binding / Protein Folding	member of the J protein family. J proteins function in many cellular processes by regulating the ATPase activity of 70 kDa heat shock proteins. This gene is a member of the type 2 subgroup of DnaJ proteins. The encoded protein is localized to the endoplasmic reticulum. This protein is induced by endoplasmic reticulum stress and plays a role in protecting stressed cells from apoptosis.	4189	100165140	DNAJB9
DnaJ (Hsp40) homolog, subfamily C, member 4	HSPF2; MCG18; DAN3C4	Unfolded Protein Binding / Protein Folding		3338	100163347	DNAJC4
SEC63 homolog (S. cerevisiae)	ERdj2; SEC63L; DNAJC23; PRO2507	Unfolded Protein Binding / Protein Folding	The Sec61 complex is the central component of the protein translocation apparatus of the (ER) membrane. encoded by this gene and SEC62 protein are found to be associated with ribosome-free SEC61 complex. speculated that Sec61-Sec62-Sec63 may perform post-translational protein translocation into the ER. The complex might also perform the backward transport of ER proteins that are subject to the ubiquitin-proteasome-dependent degradation pathway. integral membrane protein in the rough ER.	11231	100159107	SEC63
DnaJ (Hsp40) homolog, subfamily C, member 3	P58; HP58; ERdj6; PRKRI; P58IPK		This gene encodes a protein with multiple tetratricopeptide repeat (TPR) motifs as well as the highly conserved J domain found in DNAJ chaperone family members. It is a member of the tetratricopeptide repeat family of proteins and acts as an inhibitor of the interferon-induced, dsRNA-activated protein kinase (PKR).	5611	100165162	DNAJC3
heat shock 70kDa protein 1-like	HSP70T; hum70t; HSP70-1L; HSP70-HOM		This gene encodes a 70kDa heat shock protein. In conjunction with other heat shock proteins, this protein stabilizes existing proteins against aggregation and mediates the folding of newly translated proteins in the cytosol and in organelles. The gene is located in the major histocompatibility complex class III region, in a cluster with two closely related genes which also encode isoforms of the 70kDa heat shock protein	3305	100159065	HSPA1L
heat shock 70kDa protein 4	RY; APG-2; HSPH2; hsp70; hsp70RY; HS24/P52			3308	100163455	HSPA4
heat shock 105kDa/110kDa protein 1	105kDa; Hsp105; Hsp110; hsp-17L; AI790491			10808	100163455	HSPH1
Primary UPR Role: Protein Disulfide Isomerization						
VIMP VCP-interacting membrane protein	SELS; ADO15; SBBI8; SEPS1; AD-015	Apoptosis	encodes a selenoprotein, which contains a selenocysteine (Sec) residue at its active site. Studies suggest that this protein may regulate cytokine production, and thus play a key role in the control of the inflammatory response. Two alternatively spliced transcript variants encoding the same protein have been found for this gene	55829	100569049	SELS
Sterol regulatory element binding transcription factor 1	SREBP1; bHLHd1; SREBP-1c	Transcription Factor / Cholesterol Regulation	encodes a transcription factor that binds to the sterol regulatory element-1 (SRE1), which is a decamer flanking the low density lipoprotein receptor gene and some genes involved in sterol biosynthesis. The protein is synthesized as a precursor that is attached to the nuclear membrane and endoplasmic reticulum. Following cleavage, the mature protein translocates to the nucleus and activates transcription by binding to the SRE1	6720	100164413	SREBF1
Primary UPR Role: Protein Folding						
Endoplasmic oxidoreductin-1	ARO1; T9N14.18; T9N14_18	Disulfide Isomerization	endoplasmic oxidoreductin1 - helps ER proteins form disulfides	30001	100169648	ERO1
Protein disulfide isomerase family A, member 3	P58; ER60; ERp57; ERp60; ERp60; ERp61; GRP57; GRP58; IP-PLC; HuT17083	Disulfide Isomerization / Apoptosis	protein of the endoplasmic reticulum that interacts with lectin chaperones calreticulin and calnexin to modulate folding of newly synthesized glycoproteins	2923	100164598	PDIA3
DNA-damage-inducible transcript 3	CEBPZ; CHOP; CHOP10; CHOP-10; GADD153	Disulfide Isomerization / Apoptosis / Transcription factor	Protein Disulfide Isomerization, functions as a dominant-negative inhibitor by forming heterodimers with other C/EBP members, such as C/EBP and LAP (liver activator protein), and preventing their DNA binding activity. The protein is implicated in adipogenesis	1649	**N/A**	DDIT3
DnaJ (Hsp40) homolog, subfamily C, member 10	JPDI; MTHr; ERdj5; PDIA19	Disulfide Isomerization / ERAD / HSP / Unfolded Protein Binding	gene encodes an endoplasmic reticulum co-chaperone which is part of the endoplasmic reticulum-associated degradation complex. Alternatively spliced transcript variants encoding multiple isoforms have been observed for this gene	54431	100164598	DNAJC10
Endoplasmic reticulum protein 44	PDIA10; TXNDC4	Disulfide Isomerization / Folding Quality Control	also known as thioredoxin domain-containing protein 4	23071	100158679	ERP44
ERO1-like (S. cerevisiae)	ERO1A; ERO1-alpha	Protein Folding		30001	100169648	ERO1L
heat shock protein 4 like	94kDa; APG-1; Osp94; AI461691	Protein Folding		22824	100163455	HSPA4L
prefoldin subunit 2	PFD2	Unfolded Protein Binding	This gene encodes a member of the prefoldin beta subunit family. The encoded protein is one of six subunits of prefoldin, a molecular chaperone complex that binds and stabilizes newly synthesized polypeptides, thereby allowing them to fold correctly. The complex, consisting of two alpha and four beta subunits, forms a double beta barrel assembly with six protruding coiled-coils.	5202	100574163	PFDN2
peptidylprolyl isomerase A (cyclophilin A)	CYPA; CYPH	Unfolded Protein Binding	member of the peptidyl-prolyl cis-trans isomerase (PPIase) family. PPIases catalyze the cis-trans isomerization of proline imidic peptide bonds in oligopeptides and accelerate the folding of proteins. a cyclosporin binding protein and may play a role in cyclosporin A-mediated immunosuppression. can also interact with several HIV proteins, including p55 gag, Vpr, and capsid protein, and has been shown to be necessary for the formation of infectious HIV virions.	5478	100162086	PPIA
SIL1 homolog, endoplasmic reticulum chaperone (S. cerevisiae)	BAP; MSS; ULC5	Unfolded Protein Binding	resident endoplasmic reticulum (ER), N-linked glycoprotein with an N-terminal ER targeting sequence, 2 putative N-glycosylation sites, and a C-terminal ER retention signal. This protein functions as a nucleotide exchange factor for another unfolded protein response protein. Mutations in this gene have been associated with Marinesco-Sjogren syndrome	64374	100164154	SIL1
t-complex 1	CCT1; CCTa; D6S230E; CCT-alpha; TCP-1-alpha	Unfolded Protein Binding	molecular chaperone that is a member of the chaperonin containing TCP1 complex (CCT), also known as the TCP1 ring complex (TRiC). This complex consists of two identical stacked rings, each containing eight different proteins. Unfolded polypeptides enter the central cavity of the complex and are folded in an ATP-dependent manner. The complex folds various proteins, including actin and tubulin. three pseudogenes that appear to be derived from this gene have been found.	6950	100158833	TCP1
torsin family 1, member A (torsin A)	DQ2; DYT1	Unfolded Protein Binding	The protein encoded by this gene is a member of the AAA family of adenosine triphosphatases (ATPases), is related to the Clp proteases/heat shock family and is expressed prominently in the substantia nigra pars compacta. Mutations in this gene result in the autosomal dominant disorder, torsion dystonia 1	1861	100161851	TOR1A

calnexin	CNX; P90; IP90	Unfolded Protein Binding	This gene encodes a member of the calnexin family of molecular chaperones. The encoded protein is a calcium-binding, endoplasmic reticulum (ER)-associated protein that interacts transiently with newly synthesized N-linked glycoproteins, facilitating protein folding and assembly. It may also play a central role in the quality control of protein folding by retaining incorrectly folded protein subunits within the ER for degradation	821	100163271	CANX
chaperonin containing TCP1, subunit 4 (delta)	SRB; Cctd; CCT-DELTA	Unfolded Protein Binding	The chaperonin containing TCP1 (MIM 186980) complex (CCT), also called the TCP1 ring complex, consists of 2 back-to-back rings, each containing 8 unique but homologous subunits, such as CCT4. CCT assists the folding of newly translated polypeptide substrates through multiple rounds of ATP-driven release and rebinding of partially folded intermediate forms. Substrates of CCT include the cytoskeletal proteins actin (see MIM 102560) and tubulin (see MIM 191130), as well as alpha-transducin (MIM 139330)	10575	100162102	CCT4
chaperonin containing TCP1, subunit 7 (eta)	CCTH; CCTHETA; NIP7-1; TCP1ETA	Unfolded Protein Binding	This gene encodes a molecular chaperone that is a member of the chaperonin containing TCP1 complex (CCT), also known as the TCP1 ring complex (TRiC). This complex consists of two identical stacked rings, each containing eight different proteins. Unfolded polypeptides enter the central cavity of the complex and are folded in an ATP-dependent manner. The complex folds various proteins, including actin and tubulin.	10574	100168452	CCT7
UDP-glucose glycoprotein glucosyltransferase 1	UGT1; HUGT1; UGCCL1	Unfolded Protein Binding / Folding Quality Control	UDP-glucose:glycoprotein glucosyltransferase (UGT) is a soluble protein of the endoplasmic reticulum (ER) that selectively reglucosylates unfolded glycoproteins, thus providing quality control for protein transport out of the ER	56886	100159331	UGGT1
prefoldin subunit 5	MM1; MM-1; PFD5	Unfolded Protein Binding / Folding Quality Control	member of the prefoldin alpha subunit family. protein is one of six subunits of prefoldin, a molecular chaperone complex that binds and stabilizes newly synthesized polypeptides, thereby allowing them to fold correctly. The complex, consisting of two alpha and four beta subunits, forms a double beta barrel assembly with six protruding coiled-coils. protein may also repress the transcriptional activity of the proto-oncogene c-Myc.	5204	100166994	PFDN5
heat shock protein 90kDa beta (Grp94), member 1	Grp94; gp96; endoplasmin; HSPIP; TRA1		This gene encodes a member of a family of adenosine triphosphate(ATP)-metabolizing molecular chaperones with roles in stabilizing and folding other proteins. The encoded protein is localized to melanosomes and the endoplasmic reticulum. Expression of this protein is associated with a variety of pathogenic states, including tumor formation. There is a microRNA gene located within the 5' exon of this gene.	7184	100169283	HSP90B1
Primary UPR Role: Ubiquitination						
ring finger protein 139	RCA1; TRC8; HRCA1		The protein encoded by this gene is a multi-membrane spanning protein containing a RING-H2 finger. This protein is located in the endoplasmic reticulum, and has been shown to possess ubiquitin ligase activity. This gene was found to be interrupted by a t(3;8) translocation in a family with hereditary renal and non-medullary thyroid cancer. Studies of the Drosophila counterpart suggested that this protein may interact with tumor suppressor protein VHL, as well as with COPS5/JAB1, a protein responsible for the degradation of tumor suppressor CDKN1B/P27 KIP.	11236	100160538	RNF139
ring finger protein 5, E3 ubiquitin protein ligase	RMA 1; RING5		The protein encoded by this gene contains a RING finger, which is a motif known to be involved in protein-protein interactions. This protein is a membrane-bound ubiquitin ligase. It can regulate cell motility by targeting paxillin ubiquitination and altering the distribution and localization of paxillin in cytoplasm and cell focal adhesions.	6048	100166523	RNF5
ubiquitin-conjugating enzyme E2G 2	UBC7		The modification of proteins with ubiquitin is an important cellular mechanism for targeting abnormal or short-lived proteins for degradation. Ubiquitination involves at least three classes of enzymes: ubiquitin-activating enzymes, or E1s, ubiquitin-conjugating enzymes, or E2s, and ubiquitin-protein ligases, or E3s. This gene encodes a member of the E2 ubiquitin-conjugating enzyme family. The encoded protein shares 100% sequence identity with the mouse counterpart. This gene is ubiquitously expressed, with high expression seen in adult muscle. Three alternatively spliced transcript variants encoding distinct isoforms have been found for this gene.	7327	100159962	UBE2G2
ubiquitin fusion degradation 1 like (yeast)	UFD1		The protein encoded by this gene forms a complex with two other proteins, nuclear protein localization-4 and valosin-containing protein, and this complex is necessary for the degradation of ubiquitinated proteins. In addition, this complex controls the disassembly of the mitotic spindle and the formation of a closed nuclear envelope after mitosis. Mutations in this gene have been associated with Catch 22 syndrome as well as cardiac and craniofacial defects. Alternative splicing results in multiple transcript variants encoding different isoforms. A related pseudogene has been identified on chromosome 18.	7353	100166745	UFD1L
ubiquitin specific peptidase 14 (tRNA-guanine transglycosylase)	TGT		This gene encodes a member of the ubiquitin-specific processing (UBP) family of proteases that is a deubiquitinating enzyme (DUB) with His and Cys domains. This protein is located in the cytoplasm and cleaves the ubiquitin moiety from ubiquitin-fused precursors and ubiquitinylated proteins. Mice with a mutation that results in reduced expression of the ortholog of this protein are retarded for growth, develop severe tremors by 2 to 3 weeks of age followed by hindlimb paralysis and death by 6 to 10 weeks of age. Alternate transcriptional splice variants, encoding different isoforms, have been characterized.	9097	100162453	USP14
Primary UPR Role: Unfolded Protein Binding						
SREBF chaperone		Transcription Factor / Cholesterol Regulation	This gene encodes a protein with a sterol sensing domain (SSD) and seven WD domains. In the presence of cholesterol, this protein binds to sterol regulatory element binding proteins (SREBPs) and mediates their transport from the ER to the Golgi. The SREBPs are then proteolytically cleaved and regulate sterol biosynthesis.	22937	100168369	SCAP
ERO1-like beta (S. cerevisiae)				56605	100169648	ERO1LB
heat shock 70kDa protein 2	HSP70-2; HSP70-3			3306	100159065	HSPA2