Fall 2016

The fate of ICD-1 during misfolded protein induced apoptosis in caenorhabditis elegans

Kyle H. Perez
James Madison University

Follow this and additional works at: https://commons.lib.jmu.edu/honors201019

Part of the Cell Biology Commons, and the Molecular Genetics Commons

Recommended Citation
https://commons.lib.jmu.edu/honors201019/281

This Thesis is brought to you for free and open access by the Honors College at JMU Scholarly Commons. It has been accepted for inclusion in Senior Honors Projects, 2010-current by an authorized administrator of JMU Scholarly Commons. For more information, please contact dc_admin@jmu.edu.
The Fate of ICD-1 during Misfolded Protein Induced Apoptosis in Caenorhabditis elegans

An Honors College Project Presented to
the Faculty of the Undergraduate
College of Science and Mathematics
James Madison University

by Kyle H. Perez
May 2017

Accepted by the faculty of the Department of Biology, James Madison University, in partial fulfillment of the requirements for the Honors College.

FACULTY COMMITTEE:

Project Advisor: Tim Bloss, Ph.D.
Associate Professor, Department of Biology

Reader: Kim Slekar, Ph.D.
Associate Professor, Department of Biology

HONORS COLLEGE APPROVAL:

Bradley R. Newcomer, Ph.D.,
Dean, Honors College

PUBLIC PRESENTATION

This work is accepted for presentation, in part or in full, at the Honors Symposium on 21 April 2017.
Table of Contents

List of Figures and Tables 3

Acknowledgements 4

Abstract 5

Introduction 6

Methods and Materials 11

Results 15

Discussion 29

Future Research Investigations 34

References 39
List of Tables and Figures

Different Western blotting protocols yield different signal-to-noise ratios 17

Multiple sequence alignment of animal βNAC 18

Multiple sequence alignment of plant βNAC 19

Multiple sequence alignment of fungal βNAC 21

Animal βNAC multiple sequence alignment restricted to species with caspase cleavage sites 22

Multiple sequence alignment of C. elegans and human βNAC for docking site determination 23

Phyre 2 predicted model of C. elegans ICD-1 structure 24

Phyre 2 predicted model of C. elegans ICD-1 structure, colored for conservation by ConSurf 25

Structure of C. elegans NAC heterodimer predicted by Phyre 2 and HADDOCK 27
Acknowledgements

I would like to thank several people for their guidance and support over the course of the two studies in this investigation. I would like to thank Drs. Tim Bloss, Kim Slekar, and Raymond Enke for serving on my Honors committee and generously reviewing my work and suggesting areas of investigation. I would like to thank again the leaders of the joint laboratory, Drs. Bloss and Slekar for helping me get started in the field of apoptosis and molecular genetics research. Providing advice and suggestions, and for providing and maintaining an excellent workspace. It is an elaborate field, that 2 years ago was entirely new to me; I am grateful for the support and guidance you gave me in beginning to map it. I am also glad to have worked alongside the other undergraduate researchers in the lab, Kristen Hoffman, and Claire Gormley. The friendships made at the bench brought to the workspace the bond of a club, of which I am glad to be a part.

I would like to thank Dr. Berndsen from the university’s chemistry department for his advice on using the array of modelling and predictive software that I learned to use during the modelling based portion of this study.

I would also like to thank the Honors College for approving my project and the departments of biology and biotechnology for allowing me to use their space and resources to research a project of my own choosing, and for allowing it to be a part of my undergraduate degree. Additionally, this work would not have been possible without the access to research databases James Madison University kindly provides its students, nor the free software made available by the European Molecular Biology Laboratory, YASARA Biosciences, the Structural Bioinformatics Group at Imperial College, London, and Tel Aviv University.
Abstract

Severe misfolded protein stress initiates cellular responses that often result in the death of the affected cell, typically by apoptosis. An essential aspect of apoptosis is caspase-mediated cleavage of proteins that, once cleaved, further propagate death. One heterodimeric structure putatively targeted in this process in the nascent polypeptide-associated complex (NAC), a translational chaperone thought to help prevent misfolded protein stress in the ER. The purpose of this investigation was to determine whether the beta subunit of the NAC in C. elegans (ICD-1) is cleaved during the induction of apoptosis, with the hypothesis that ICD-1 is cleaved during stressed-induced apoptosis to propagate cell death. To test this hypothesis, I performed Western analyses of ICD-1 obtained from stressed and un-stressed populations of wild-type C. elegans to determine any stress-correlated differences in the protein length of ICD-1. To determine the possible structure of ICD-1 during this cleavage, I used protein modelling to evaluate the position of the putative caspase-cleavage site of monomeric ICD-1, the form the protein is thought to take during misfolded protein stress. Optimization of western blotting conditions for detecting ICD-1 were developed, but the final western blots on stressed and unstressed worms were inconclusive due to bacterial contamination of protein lysates. The amino acid sequence analyses and protein modelling algorithm analyses revealed the putative caspase cleavage site on monomeric ICD-1 is exposed and accessible to cleavage by a caspase. Sequence alignments and bioinformatics analyses for conservation revealed a high degree of variability localized to the region surrounding the putative caspase cleavage site, indicating evolutionary specialization that optimizes the cleavage of this site by the specific caspases found in the worm. Further investigation should develop uncontaminated western analysis results of the state of ICD-1 in stressed and unstressed worms. Further structural analyses could assess the nature of ICD-1 when bound by ICD-2 as well as x-ray crystallography to reliably resolve the structures of the ICD-1/ICD-2 heterodimer.
Introduction

Controlling the time and manner of a cell’s death is an important mechanism that adjudicates between the chaos of unmanaged damage on one hand and unwarranted self-destruction on the other. Controlled cell death plays an important role in the development of multicellular organisms (such as the shaping of organs and appendages and axonal pruning in the nervous system) and their protection against overwhelming cellular stress (Alberts et al., 2002; Wyllie, 1987). With regards to stress, cells must be able to assess stresses and ‘decide’ when it is unable to save itself and thus enter a controlled self-destruction to prevent damage to its neighbours. But, a low threshold for stress would mean inappropriate and overly frequent self-destruction, possibly resulting in disease. An accurate and precise balance exists between self-saving and self-destructive mechanisms with the goal of removing dangerous cells while maintaining stressed but salvageable cells. It is the purpose of this study to investigate that balance and the mechanisms by which it is tuned.

Apoptosis is a term that describes the controlled termination of a cell’s life sustaining processes and the subsequent packaging of organelles and other cellular materials for their redistribution to neighboring cells (Alberts et al., 2002). Apoptosis is also the method by which an organism culls certain pre-determined somatic cells to shape tissues and prune neural connections, which are necessary for proper development. Distinct from the controlled termination of apoptosis is un-controlled, toxic cell death known as necrosis. Necrosis occurs when cells are subject to such a degree of damage or abrupt fluctuation of homeostasis that they cannot recycle their constituent materials. In that case, vesicles such as peroxisomes burst, exposing their caustic chemicals and enzymes to the dying cells and inter-cellular spaces – thus causing damage to proximal cells and tissues (Golstein & Kroemer, 2007). Therefore, to avoid the hazards presented by necrosis most eukaryotic organisms have carefully coordinated the induction of apoptosis to disarm and recycle those cells that are on the verge of death.

Several types of stress may cause a cell to initiate apoptosis; one such well-known stressor is misfolded-protein-stress. The accumulation of improperly folded proteins in a cell can interrupt important processes and cause physical damage to surrounding organelles. In response to misfolded protein stress, many cells initiate unfolded protein response (UPR) to manage
misfolded protein stress and repair the damage caused by it; and, if the stress becomes unmanageable they initiate apoptosis (Haynes, Titus, & Cooper, 2004). That coordinated process is realized by a specific set of enzymes are known to engage in cell-saving mechanisms during manageable stress. Principle among those is the nascent-polypeptide associated complex (NAC), which is composed of two orthologous subunits, the beta subunit (βNAC) and the alpha subunit (αNAC). The NAC normally acts as a translational chaperone to encourage the correct folding of peptides as they emerge from the ribosome and may also localize peptides to specific compartments of the cell including the mitochondria. During the UPR, the subunits of the NAC are known to dissociate and are believed to play individual roles in managing misfolded protein stress. The importance of the ratio of βNAC and αNAC to the UPR and induction of apoptosis has been shown by the evidence that the depletion of the Caenorhabditis elegans (C. elegans) homologue of βNAC, inhibitor of cell death 1 (ICD-1) leads to inappropriate apoptosis and morphological defects during embryogenesis (Arsenovic, Maldonado, Colleluori, & Bloss, 2012), and that the presence of ICD1 suppresses CED-3 independent apoptosis (Bloss, Witze, & Rothman, 2003).

Research with the model organism C. elegans supports the existence of a conserved pathway for the initiation of apoptosis. C. elegans is a ubiquitous soil nematode – found on every continent except Antarctica – whose apoptotic pathway is highly conserved with most other eukaryotes (Kaletta & Hengartner, 2006). The fate of every somatic cell of the organism is known; maps have been built that describe which cells live and give rise to descendants, and which are culled and when they are culled (Deppe et al., 1978). In virtue of that information, the precise effect of altering a component in the apoptotic pathway can determined by comparing the cell-counts and types to those under the wild-type, control conditions. Those features make C. elegans an ideal model organism for conducting studies in apoptosis that are relevant to humans.

In C. elegans, apoptosis is initiated by the joint action of four primary polypeptides: CED-3, CED-4, CED-9, and EGL-1. CED-3 is a caspase, a cysteine-based protease that targets sites containing aspartate residues, and is localized to the cytoplasm. CED-4 and CED-9 are localized to the mitochondria. CED-9 putatively binds to CED-4 to sequester it in its monomeric state, thus preventing it from obtaining an oligomeric conformation. During the induction of apoptosis, EGL-1 is expressed and localizes to the mitochondria. There it binds to CED-9 by means of a (BH3) C-terminal domain. By sequestering CED-9, EGL-1 is indirectly allowing
CED-4 monomers to bind with each other and attain an oligomeric conformation. Oligomeric CED-4 binds to and encourages the autocatalytic cleavage of pro-CED-3 to produce a catalytically active CED-3 (Huang & Strasser, 2000). CED-3, CED-4, and EGL-1 are pro-apoptotic, they contribute to the induction of apoptosis, whereas CED-9 functions to suppress apoptosis (Horvitz, 1999).

It is likely that CED-3 is directly tipping the balance away from cell-saving mechanisms and toward active apoptosis through caspase mediated cleavage of specific targets. Specifically, it is here suggested that CED-3 cleaves ICD-1 at a putative caspase cleavage site between residues 130 – 140. Since the depletion of ICD-1 relative to ICD-2, the C. elegans homologue of αNAC, is associated with the induction of apoptosis, it is proposed that ICD-1 is selectively cleaved by caspase-mediated cleavage and its self-saving functions terminated in order to initiate apoptosis. Whether this cleavage inactivates ICD-1 or instead converts it to a pro-apoptotic activity is not predicted, but the hypothesis remains that induction of stress-induced apoptosis in C. elegans involves the direct cleavage of ICD-1 by a caspase. To test this hypothesis I sought to investigate the plausibility that ICD-1 is cleaved under stress conditions. That was done both in vivo, by comparing protein lysates obtained from stressed and unstressed populations of C. elegans, and in silico, by examining the predicted structural conformations of ICD-1 both as a monomer and a heterodimer in complex with ICD-2.

Stressed conditions can be induced in C. elegans by treating the organism with chemicals that induce misfolded protein accumulation in the ER. Several chemicals contribute to ER stress when absorbed by C. elegans. Tunicamycin and Thapsigargin in particular are known to trigger ER stress and the subsequent responses in C. elegans, albeit through different pathways (Szegezdi, Logue, Gorman, & Samali, 2006). These chemicals stress C. elegans in such a way that would not be out of the ordinary range of stresses that the organism may be expected to encounter in its natural environment.

In this experiment mixed population protein lysates were extracted from wild-type worms under chemically induced stress and from wild-type unstressed populations. It is worth noting that to obtain samples of ICD-1 from colonies of C. elegans whole-protein extracts are obtained that include the proteins of C. elegans of all stages of development. The presence and size of ICD-1 were detected and measured through Western blot analysis.
To test the prediction that ICD-1 is cleaved in chemically stressed populations, I used western blotting analysis to look for a difference in protein size between ICD-1 from control populations of C. elegans and chemically stressed populations. Western blotting uses gel electrophoresis and immunostaining to separate by size the proteins from a lysate, transfer them to blotting filter, and then visually detect a protein of interest by its binding to a target-specific primary antibody. Here, Western blotting was used to compare the sizes of ICD-1 from populations of healthy wild-type C. elegans with the sizes of ICD-1 populations of wild-type C. elegans treated with either tunicamycin or thapsigargin. If ICD-1 is being cleaved in response to overwhelming misfolded-protein stress, then the size of ICD-1 from chemically stressed populations should be 3 kDa smaller, the predicted size of the portion of ICD-1 beyond the caspase cleavage site, than the size of ICD-1 from unstressed populations. Thus, I compared blots from unstressed and stressed populations to determine if there is a size difference indicative of caspase cleavage in the population most likely undergoing widespread apoptosis.

In addition to determining if ICD-1 is cleaved under stressed conditions, I have attempted to establish the predicted structure of ICD-1 in monomeric and heterodimeric form to determine the likelihood such that cleavage is possible at the putative cleavage site. To determine if ICD-1 is capable of being cleaved in vivo it is necessary to know what conformations ICD-1 adopts most often over the course of its functions in the cell, and whether or not its putative cleavage site is exposed on its surface during those conformations, or at very least, if that site is accessible to a caspase during those conformations. It is already known that monomeric ICD-1 can be cleaved in vitro by CED-3 at the putative caspase cleavage site, but the conformations which ICD-1 adopts during cellular functions may occlude that site, or otherwise make it inaccessible to CED-3 or another caspase. Since ICD-1 spends most of its time complexed with αNAC in a heterodimer to chaperone peptides emerging from the ribosome it is most relevant to investigate the state of ICD-1’s surface while it is in complex with αNAC. ICD-1 is also known to dissociate from αNAC during cell-saving functions. Since cells initiate cell-saving functions in response to stresses that may ultimately lead to apoptotic induction, and therefore the activation of caspases, it is also highly relevant to investigate the accessibility of monomeric ICD-1 to cleavage.

There are several protein modelling software packages that can produce a 3-D structure from a FASTA peptide sequence by using as templates the models of solved structures of similar
proteins. Protein modelling software systems report several parameters associated with the predicted structure. The level of confidence, or C-score, is the quality of the predicted structure, and is calculated based on the significance of the templates and their congruence with the sequence of the protein being modelled, as reported by RMSD scores. If the predicted structure is drawn with a sufficient level of confidence, then it is reasonable to use the predicted structure to make inferences about biochemical interactions of which a protein of interest is capable.

Protein modelling software is not always able to find templates to direct the predicted structure. In that case it performs *ab initio* modelling, which is based on the biochemical properties of the residues alone and is largely unreliable.

Fortunately, a three-dimensional structure of the interacting faces of βNAC and αNAC in a heterodimer has been solved by X-ray crystallography (Wang et al., 2010). However, those data alone reveal little about the conformation of the rest of the heterodimeric structure. Obtaining a predicted structure of the rest of the heterodimer will allow us to investigate whether ICD-1’s putative caspase cleavage site is exposed on its surface, or buried, without solving the entire structure with X-ray crystallography.

The conformation in which ICD-1 spends most of its time is that which it adopts when it is complexed in a heterodimer with ICD-2. ICD-1’s response to misfolded protein stress involves ICD-1’s dissociation from ICD-2. It is also proposed that ICD-1 may function as a chaperone while a monomer. Investigating the plausibility of the cleavage of ICD-1 as a monomer and as a heterodimer in complex with αNAC are of the greatest relevance, and comprises the aim of the present 3-D modelling-based portion of this study. That plausibility is ascertained by examining where the putative modelling-based portion of this study.

It would also be beneficial to this study to determine the degree to which the NAC peptide sequences are conserved within and between phylogenetic groups, what features are common or variable between groups, and if there are species specific features in the NAC, what those features mean in the context of a given species. For example, it would be informative to know if all eukaryotes share regions of high degrees of conservation, where those regions are, if they are likely to be involved in ICD-1 cleavage, and whether say, soil nematodes, share a distinct sequence unique to themselves alone (Crawford et al., 2012). The answers to those
questions would help reveal whether cleavage of ICD-1/βNAC is conserved in other organisms, what types of organisms share that function, whether some groups have developed differences in NAC structure and function and how their evolutionary history explains why they have developed those differences, and the different stresses and housekeeping functions pertinent to each phylogenetic group. If *C. elegans* belongs to a group of organisms distinguished by unique NAC structure or function, then it would provide an evolutionary context useful for concluding the extent to which the observations made in this study are applicable to other model organisms.

### Methods and Materials

The following Western blotting procedure consists of two primary parts: preparation of protein lysate samples and optimization of antibody incubation. Before either of those parts could be completed colonies of *C. elegans* had to be grown. Three sets of three treatments of agar plates of nematode growth medium were prepared. One set contained the chemical stressor thapsigargin, another contained the chemical stressor tunicamycin, and a set of control plates had no added stressors. Wild-type N2 *C. elegans* colonies were seeded on each of the plates and were given a week to grow.

Samples of protein lysates were prepared as follows. Sets of plates were pooled by treatment. Plates of a given treatment were placed on ice and washed with Tris-buffered saline (TBS), aspirated, and combined with 1 ml of RIPA lysis buffer (50mM Tris-HCl, 150 mM NaCl, 1% Nonidet P-40, 0.5% sodium deoxycholate, 0.1% sodium dodecyl sulphate (SDS), 1 mM sodium orthovanadate, 1 mM NaF, Roche protease inhibitors, pH 8.0). Each plate was then scraped lightly with a cell scraper. The suspended colonies of a given treatment were then transferred to a 50 ml tube. Each of the three tubes were labelled with the name of their respective treatments. Agitation was maintained until each tube was centrifuged at 4°C 16,000 xg for 20 minutes. The tubes were kept on ice and the supernatants were transferred to fresh
tubes as the pellets were discarded. Approximately 10 aliquots at 20 μl each were made from each treatment tube and were then stored at -80°C. Before each sample was used in loaded in gel electrophoresis it was thawed, combined with 20 μl of 2× laemmli sample buffer (4% SDS, 10% 2-mercaptoethanol, 20% glycerol, 0.004% bromophenol blue, 0.125 M Tris-HCl, pH 6.8), boiled at 95°C for 5 minutes, and the spun in a micro-centrifuge at 20°C at 16,000 xg for 1 minute.

Protein separation by gel electrophoresis was performed on 4-20% SDS-PAGE gels gradient gels with 20 μg of protein added to each lane. Each gel contained a three sets of four lanes: one lane of the molecular weight marker, one of the control unstressed population protein lysate, one of protein lysate from the population stressed with tunicamycin, and one of the protein lysate from the population stressed with thapsigargin. Each gel was run at 100 V for one hour.

The proteins, now separated by size, were then transferred to nitrocellulose membranes. The gel was submerged in 1x transfer buffer (15mM Tris, 190 mM glycine, 0.1% SDS at 8.3 pH) for 10-15 minutes. A transfer cassette was assembled, consisting of, outward to inside, plastic grids, plastic sponge pads, filter paper, the gel, and the transfer membrane, with the nitrocellulose transfer membrane on the cathode and the gel on the anode. The cassette was placed in a transfer tank filled with transfer buffer and run at -20°C for 18 hours. The filter was then cut such that each set of four different lanes were separated into 3 smaller filters. Each of the smaller filters of 4 lanes were treated with varying antibody incubation procedures described in general below.

To detect signal from the protein of interest the filters, now with protein embedded, were prepared for antibody incubation. The blotted transfer paper was blocked in 3% BSA in TBST at 20°C for 1 hour, and was then incubated in the primary monoclonal antibody specific to C. elegans ICD-1 overnight at 4°C. The blot was rinsed 3 times in TBST for 5 minutes each. The blot was then incubated with the HRP-conjugated secondary antibody solution for 1 hour. After the secondary incubation the blot was rinsed again 3 times with TBST for 5 minutes each. The filters, now with bands of the protein of interest indicated in indigo were dried, scanned, saved to computer .png files, and stored.

Initially, the conditions of the incubation procedure were set first at the lowest possible stringency to set a lower-bound for future attempts. The procedure included blocking the blot at 1% BSA for 30 minutes, rinsing the blot for 3 minutes once before and after the secondary
antibody incubation, with both incubations being performed at 20°C. The procedure was then performed at very high stringency, with all steps performed at -20°C, the blot blocked with 3% BSA for 1 hour, and washed 5 times in TBST for 5 minutes each before and after the secondary antibody incubation. Subsequent incubation steps were performed at intermediate levels of stringency, with some steps performed at -20°C and others at room temperature, varying BSA percent solutions for blocking and varying blocking durations, and with varying numbers of TBST washes with different durations. This was done until a procedure that optimizes signal from ICD-1 bands while reducing noise from other proteins was found.

The bands reported in each lane were to be compared to determine if a single band at the expected weight of 17.51 kDa was present in the unstressed population and a slightly smaller band of 14.41 kDa was present in the stressed population, indicating cleavage.

The methods describing the computer based creation of residue conservation data and the modelling of predicted structures of monomeric and heterodimeric ICD-1 are reported below. The steps built upon each other. Predictive structures of ICD-1 and αNAC were used in conjunction with residue conservation data from multiple sequence alignments to predict the structure of the heterodimer.

Sequence alignments to compare conservation within phylogenetic orders were created using both manually using Clustal Ω and computationally using ConSurf (-Celniker et al.; Ashkenazy, Erez, Martz, Pupko, & Ben-Tal, 2010; Ashkenazy et al., 2016; Berezin et al., 2004; Glaser et al., 2003; Landau et al., 2005). The first step in modelling ICD-1 was determining the degree of conservation between the ICD-1 sequences from several species of model organisms. The online software, Clustal Ω, provided by EMBL-EBI(Li et al., 2015), was used to create multiple sequence alignments of ICD-1 peptides from animals Lygus hesperus, Fopius arisanus, Clonorchis sinensis, Caenorhabditis elegans, Mus musculus, Homo sapiens and Salmo salar; plants, Gossypium arboreum, Anthurium amnicola, Arabidopsis thaliania, Nicotiana benthamiana, Dichanthelium oligosanthes, and Zea mays; fungi, Rhizoctonia solani, Cutaneotrichosporon oleaginosus, Cryptococcus amyloleutus, Cryptococcus depauperatus, Kwoniella dejecticola, Xylona heveae, Marissonia brunnea, Neurospora crassa, Rosellinia necatrix, Talaromyces islandicus, Spathaspora passalidarium (Y-27907), Candida albicans,
Candida tanzawawaensis (NRRL Y-17324), Saccharomyces cerevisiae (Beta-1 NAC), Saccharomyces cerevisiae (Beta-2 NAC), and Candida glabrata. The alignment of the sequences from those organisms was analyzed quantitatively for residue conservation by the program BoxShade. The results of those sequence alignments are reported below (Figure 2 – 5). Although reporting, with a high degree of confidence (high FOS score), regions of high levels of conservation common to all tested model organisms a region of specific interest with a high degree of conservation among C. elegans, human, mouse, and salmon (i.e. the putative caspase cleavage site) was not indicated as a region of conservation on the first sequence alignment because L. Hesperus, F. arisanus, and C. sinensis do not possess βNAC sequences with region homologous to that which contains the caspase cleavage site. A second sequence alignment was made of the ICD-1 sequences of animal model organisms which retain the domain containing the putative caspase cleavage site (Figure 5). The Phyre 2 model of ICD-1 in Figure 7 is color-coded using the multiple sequence alignment in Figure 5 to show to position of conserved residues on the predicted structure.

The sequence alignment of human and C. elegans αNAC (Figure 6) was intended only as an index to identify residues of C. elegans ICD-1 that are homologous in sequence to human ICD-1. In that way, residues of C. elegans ICD-1 that are likely active in the heterodimerization of C. elegans ICD-1 and αNAC can be identified in reference to the map of residues of human ICD-1 known to be involved in heterodimerization (Wang et al., 2010).

The program ConSurf was used to determine the conservation of residues of ICD-1 from a large library of homologous proteins (figure 8). For consistency, the Phyre 2 predicted structure of ICD-1 was used as the known structure when mapping conservation in ConSurf. The maps of conservation were used to determine the most conserved residues of ICD-1 to use in conjunction with the map of residues involved in human NAC dimerization to select active residues for docking C. elegans NAC components. αNAC residues were selected if they paired with selected ICD-1 residues.

The structures of αNAC and ICD-1 were predicted first using I-TASSER (Ambrish, Alper, & Yang, 2010; Yang, 2008). However, of the five models produced by I-TASSER, the highest C-score associated with them was only -3.34 (out of a range of -5 to 2, with 2 being most confident). Since that score was too low to make meaningful inferences the structures were
predicted again with Phyre 2, another predictive protein modelling software (Kelley, Mezulis, Yates, Wass, & Sternberg, 2015).

The structure of αNAC was predicted from 3 models (c3lkxB, c1tr8A, and c4npbA), and the structure of ICD-1 was predicted using models c1tr8A, c3lkxA, and c4npbA.

The structure of the heterodimer of alpha-NAC and beta-NAC was predicted using the molecular docking software HADDOCK (De Vries, Van Dijk, & Bonvin, 2010) (Figure 9). PDB files of the structures of alpha-NAC and beta-NAC predicted by Phyre 2 were designated as the molecules to be docked. Active residues for the interaction were selected by using the multiple sequence alignments of animal beta-NAC and αNAC sequences to find the residues of C. elegans alpha NAC and βNAC that are homologous to residues which engaged in side-chain interactions in the interfacing of human alpha and beta NAC. Active residues were further refined by only selecting the top 4 most highly conserved residues, as reported by ConSurf. Active residues of ICD-1 were designated as 60, 64, 81, 87, 89, 94, 96, 101; and, active residues of αNAC were designated as 76, 80, 97, 103, 105, 110, 112, 117. Passive residues were determined automatically by the program.

To determine if cleavage is possible or likely for a given structure that structure is examined for the position and exposure of the predicted cleavage site and whether the protein that is predicted to cleave the site is capable of docking to the site long enough to achieve cleavage.

Results

Western analysis of ICD-1 produced variable results

The Western blotting protocol provided the clearest signal with minimal noise included blocking with 3% BSA for 1 hour, incubating the blot with the primary antibody overnight at -20°C, then
washing 3 times in TBST for 5 minutes each, incubating the blot with the secondary antibody at -20°C overnight, and then washing 5 times for five minutes each (Panel A in Figure 1). Three bands were visible in most lanes, one at 80 kDa, one at 27 kDa, and one at 17 kDa. The 80 kDa band had the strongest signal (Figure 1). The blot from the lowest stringency protocol displayed the strongest bands, the optimized protocol the second strongest, an intermediate protocol the third, and the highest stringency protocol the least. Although all blots displayed at least very faint bands, not all were visible once scanned to the computer, nor even when digitally enhanced.

**Multiple sequence alignments highlight cleavage site conservation and group specificity**

The multiple sequence alignments (MSA) created by Clustal Omega and annotated by Boxshade show multiple regions of high conservation within each of the three phylogenetic orders of ICD-1. All three MSAs were generated at 1.0 fractions of sequences, the highest setting while still identifying ample residues of greater than 80% conservation (Figures 2 – 4). The sequence of four residues constituting ICD-1’s putative caspase cleavage site (LGPD) were found, either in that sequence or in one functionally similar, several times in all three MSAs. *Lygus hesperus, Fopius aisanus,* and *Clonorchis sinensis* do not have the region of βNAC that contains the putative caspase cleavage site (Figure 2); mouse, human, and salmon ICD-1 have a putative caspase cleavage site that varies slightly in amino acids, with lysine being replaced by glutamine, glycine by serine or alanine, proline by valine or glycine, and aspartate remaining the same (Figure 2). All plants but two retained either the same sequence as C. elegans or one similar (Figure 3). Nine fungal ICD-1s had the LGPD sequence and four had a similar LGAE sequence of 16 compared (figure 4).

**ConSurf locates regions of conservation and variability on ICD-1 structure**

The results from ConSurf indicate that the sequences of ICD-1 that are most conserved among similar proteins are those known to be involved in NAC dimerization (residues 40 – 105). When the conservation color coding from ConSurf is overlain on the Phyre 2 predicted structure of ICD-1 it is observed that the region surrounding the putative caspase cleavage site is the protein’s most variable region (Figure 8).
**Structural predictions support the plausibility of ICD-1 cleavage**

The Phyre predicted structure of αNAC (visible as the magenta region in Figure 9) was generated with 50% of the structure being modelled at >90%. The sequences that were modelled with greater than 90% confidence are 70 to 122, and 148 to 197. In the Phyre 2 structural prediction of ICD-1 (Figure 7) 61% of residues were modelled at >90% confidence; those residues are 54 – 120, and 128 through 158. Residues 1 – 40 and 121 – 127 were modelled *ab initio*. Phyre 2 predicted structure of ICD-1 appears in figure 5. The predicted structure of the heterodimer, which was generated by docking αNAC and ICD-1 monomers in HADDOCK, had final clusters that covered 67% of possibilities with a z-score of 0.9. The residues that were selected as active in docking the subunits in HADDOCK were defined using a multiple sequence alignment of human and *C. elegans* αNAC and finding the *C. elegans* residues homologous to the human residues known to be involved in NAC dimerization (Figure 6).
Figures and Tables

Figure 1: Different Western blotting protocols yield different signal-to-noise ratios.
Western blots obtained during the optimization of the western blotting procedure. This was intended to establish a protocol that yielded a high degree of signal specific to the ICD-1 protein, and minimal noise. All blots were digitally enhanced using the same parameters. A) the represents the most optimized comparing blotting protocol: from left to right tunicamycin stressed, unstressed, and the molecular weight ladder. B) represents the least stringent protocol: from left to right, tunicamycin-stressed, unstressed lysates, and a molecular weight ladder (not visible). C) represents a procedure of intermediate stringency: from left to right tunicamycin-stressed protein lysates (not visible), thapsigargin-stressed lysates (not visible), and a molecular weight ladder. D) represents the most stringent protocol: from left to right, a tunicamycin-stressed protein lysate (not visible), a thapsigargin-stressed lysate (not visible), and a molecular weight ladder (not visible).
Lygus. hesperus & 1 & --- ------------------------------------------------------------------ & MPITQEQLRKRAE  
F. arisanus & 1 & --- ------------------------------------------------------------------ & MPITQEQLRKRAE  
C. sinensis & 1 & --- ------------------------------------------------------------------ & MDSKAERIKLQQAQE  
C. elegans & 1 & --- ------------------------------------------------------------------ & MDSKAERIKLQQAQE  
Mus musculus & 1 & MRRTGAPTQADSRRGRRARGGWPQAEATPS--LPLGSGRGRSQMKEQKLAIKLAQ  
Homo sapiens & 1 & MRRTGAPQADSRRGRRARGGCPGEATLSQPPRGRQCFMPQEMTINQMEEKLAIKLAQ  
Salmo salar & 1 & --- ------------------------------------------------------------------ & MKEIINQMEEKLAIKLAQ  

**Figure 2:** Multiple sequence alignment of βNAC from animal model organisms. The multiple sequence alignment of βNAC peptide residue sequences from the model organisms *Lygus hesperus, Fopius arisanus, Clonorchis sinensis, Caenorhabditis elegans, Mus musculus, Homo sapiens, and Salmo salar*. Numbers to the right of the species names indicate on which residue the following sequence begins. The fraction of sequences that must agree for shading is 1.0 (FoS = 1.0), the most stringent setting for establishing conservation. Red indicates conservation of 80% or greater; blue indicates conservation between 60% and 80%, and gray indicates conservation less than 60%. The putative caspase cleavage site on *C. elegans* ICD-1 and homologous sequences of the other organisms site are highlighted in green. *L. hesperus, F. arisanus, and C. sinensis* sequences of βNAC end before reaching the point where the sequence homologous to the caspase cleavage site would be.
<table>
<thead>
<tr>
<th>Species</th>
<th>Alignment</th>
<th>FoS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gossypium_arbor</td>
<td>1</td>
<td>9.0</td>
</tr>
<tr>
<td>Anthurium_amnic</td>
<td>1</td>
<td>9.0</td>
</tr>
<tr>
<td>Arabidopsis_tha</td>
<td>1</td>
<td>9.0</td>
</tr>
<tr>
<td>Nicotiana_benthamiana</td>
<td>1</td>
<td>9.0</td>
</tr>
<tr>
<td>Dichanthelium</td>
<td>1</td>
<td>9.0</td>
</tr>
<tr>
<td>Zea_mays</td>
<td>1</td>
<td>9.0</td>
</tr>
<tr>
<td>consensus</td>
<td>1</td>
<td>9.0</td>
</tr>
</tbody>
</table>

**Figure 3: Multiple sequence alignment of βNAC sequences from plant model organisms.**

Species compared are listed on the far left; species included, from top to bottom, are Gossypium arboreum, Anthurium amnicola, Arabidopsis thaliana, Nicotiana benthamiana, Dichanthelium oligosanthes, and Zea mays. Numbers to the right of the species names indicate on which residue the following sequence begins. FoS = 9.0. Red indicates conservation of 80% or greater, blue indicates conservation between 60% and 80%, and gray indicates conservation lower than 60%. Sequences homologous to the *C. elegans* putative caspase cleavage site are highlighted in green.
<table>
<thead>
<tr>
<th>Species</th>
<th>Alignment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rhizoctonia sol</td>
<td>1 -------MOTSANSRVYIGKSVRRKVPVRSGP-VKAGGDKLAAALKLNLVQP</td>
</tr>
<tr>
<td>Cutaneotrichosp</td>
<td>1 -------MDKKDKLAKLQAVRIG-GKGTTPRQVKK-V-TSAQDDDRKLQALQLGVP</td>
</tr>
<tr>
<td>Cryptococcus am</td>
<td>1 MSLTPPTPIPLIVQSYNKT-GGKTPRQVKK-V-TATQDDDRKLQALQLGVP</td>
</tr>
<tr>
<td>Cryptococcus de</td>
<td>1 -------MPYI-GGTPRQVKK-V-TASQDDDRKLQALQLGVP</td>
</tr>
<tr>
<td>Kwonieilla dejec</td>
<td>1 -------MDKKDKLAKLQAVRIG-GCCTPPRQVKK-V-TASQDDDRKLQALQLGVP</td>
</tr>
<tr>
<td>Xylona heveae</td>
<td>1 -------MDHAKLARMQQSIVRIG-GKTPRKK-KKV-H-KNSSQDDDRKLQALQLGVP</td>
</tr>
<tr>
<td>Marissonia brun</td>
<td>1 -------MDMEKLMQSQSIVRIGDG-GGKTPRQVKK-V-KTGQDDDRKLQALQLGVP</td>
</tr>
<tr>
<td>Neurospora cras</td>
<td>1 MS--DVQERLKL--GARASIGIKGTKPRKV-KRA-P-ARSAGDDKQLQATLKLNNVQP</td>
</tr>
<tr>
<td>Rosellinia nect</td>
<td>1 MA--DVQELRLKFFGAAPARTGKGTPRHK--KK--V--ARPADDKQLQATLKLNNQTP</td>
</tr>
<tr>
<td>Talaromyces isl</td>
<td>1 -------MPVDEKALKLHKATA--KKGGSRIKAKKN----VEKEQDDDTLLEALGKATK</td>
</tr>
<tr>
<td>Spathaspora pas</td>
<td>1 -------MPVDEKALKLHKATA--KKGGSRIKAKKN----VEKEQDDDTLLEALGKATK</td>
</tr>
<tr>
<td>Candida albican</td>
<td>1 -------MPVDEKALKLHKATA--KKGGSRIKAKKN----VEKEQDDDTLLEALGKATK</td>
</tr>
<tr>
<td>Candida tanzawa</td>
<td>1 -------MPVDEKALKLHKLSAA--NKVGGRPRK1NKGGNLNNNNDKNTKLQAELKHLHMT</td>
</tr>
<tr>
<td>Saccharomyces c</td>
<td>1 -------MPIDQEXALKPKLSAN--NKVGGRPRK1NKGGNLNNNNDKNTKLQAELKHLHMT</td>
</tr>
<tr>
<td>Candida glabrata</td>
<td>1 -------MPIDQEXALKPKMSAG--NKVGGRPRK1KKTGSGSATGAKDDTDLHQLQALHLAVT</td>
</tr>
<tr>
<td>consensus</td>
<td>1 G R K D K L L K</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Species</th>
<th>Alignment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rhizoctonia sol</td>
<td>49 IAEPVEEMFQVDGSVHLFTAPKGMISILFVIIHIRAGSFKLKSIVHGLSLMTAVYG</td>
</tr>
<tr>
<td>Cutaneotrichosp</td>
<td>52 ITGEEVMFKEGNVLHFAQPV------------------------AVHAALPSMTAVYG</td>
</tr>
<tr>
<td>Cryptococcus am</td>
<td>58 IAQIEVEMFKEGNVLFAAAPRT-----------------------QVHAALPSMTAVYG</td>
</tr>
<tr>
<td>Cryptococcus de</td>
<td>41 ITGEEVMFKEGNVLHFAQPV------------------------QVHAALPSMTAVYG</td>
</tr>
<tr>
<td>Kwonieilla dejec</td>
<td>52 ITGEEVMFKEGNVLHFAQPV------------------------HALPSMTAVYG</td>
</tr>
<tr>
<td>Xylona heveae</td>
<td>51 IQAIEVEMFKEGNVLHFAQPV------------------------HASVPANTFAIYG</td>
</tr>
<tr>
<td>Marissonia brun</td>
<td>52 IQAIEVEMFKEGNVLHFAQPV------------------------HASVPANTFAIYG</td>
</tr>
<tr>
<td>Neurospora cras</td>
<td>54 IQAIEVEMFKEGNVLHFAQPV------------------------HASVPANTFAIYG</td>
</tr>
<tr>
<td>Rosellinia nect</td>
<td>56 IQAIEVEMFKEGNVLHFAQPV------------------------HASVPANTFAIYG</td>
</tr>
<tr>
<td>Talaromyces isl</td>
<td>51 IEIEEANFFRDKVLHFRVQV------------------------QGAAPANTFAIYG</td>
</tr>
<tr>
<td>Spathaspora pas</td>
<td>51 IEIEEANFFRDKVLHFRVQV------------------------QGAAPANTFAIYG</td>
</tr>
<tr>
<td>Candida albican</td>
<td>51 IEIEEANFFRDKVLHFRVQV------------------------QGAAPANTFAIYG</td>
</tr>
<tr>
<td>Candida tanzawa</td>
<td>51 VEGEEANFFRDKVLHFRVQV------------------------QGAAPANTFAIYG</td>
</tr>
<tr>
<td>Saccharomyces c</td>
<td>56 IENVAENFFRKVLHFRSVAV------------------------QIAPCNLTMTH</td>
</tr>
<tr>
<td>Candida glabrata</td>
<td>56 IDVAAENFFRKVLHFRSVAV------------------------QIAPCNLTMTH</td>
</tr>
<tr>
<td>consensus</td>
<td>61 i i E N F G V I H F N G</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Species</th>
<th>Alignment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rhizoctonia sol</td>
<td>109 AGQVKDLTELPNILQG-------------SLANLRLALASYSQMQRARRAQAAQGA-------GAPGADD</td>
</tr>
<tr>
<td>Cutaneotrichosp</td>
<td>91 PGQTELELPNIGLQG-------------SLANLRLALASYSQSTQQRQAEASKGK-------KGEGVGD</td>
</tr>
<tr>
<td>Cryptococcus am</td>
<td>97 PGQTELELPNIGLQG-------------SLANLRLALASYSQMTRARQAGAGG-------GAGEGG</td>
</tr>
<tr>
<td>Cryptococcus de</td>
<td>80 PGQTELELPNIGLQG-------------SLANLRLALASYSQMTRAAAAAAAGGAAGAG-EAAAG</td>
</tr>
<tr>
<td>Kwonieilla dejec</td>
<td>89 PGQTELELPNIGLQG-------------SLANLRLALASYSQMTRAAAAAAAGGAAGAKKLEE</td>
</tr>
<tr>
<td>Xylona heveae</td>
<td>88 NEKEELTELPNIGLQG-------------SLASLRKALASYSQSLQKEEGEE-----------------GDKEAD</td>
</tr>
<tr>
<td>Marissonia brun</td>
<td>88 NEGDKLELPNIGLQG-------------SLASLRKALASYSQSMKAEEDG-----------------EKKDDEDD</td>
</tr>
<tr>
<td>Neurospora cras</td>
<td>89 NEGDKLELPNIGLQG-------------SLASLRKALASYSQSMQKAEEDG-----------------EKKDDEDD</td>
</tr>
<tr>
<td>Rosellinia nect</td>
<td>93 AGEDKLELPNIGLQG-------------SLASLRKALASYSQSMKAEEDG-----------------GDKEAD</td>
</tr>
<tr>
<td>Talaromyces isl</td>
<td>88 YPQKTEILPLIQLPLQ------------------EADLIIQLAEQIQAGKTTPKFDNTG-AEAEAEAE</td>
</tr>
<tr>
<td>Spathaspora rae</td>
<td>88 YPQKTEILPLIQLPLQ------------------EADLIIQLAEQIQAGKTTPKFDNTG-AEAEAEAE</td>
</tr>
</tbody>
</table>

(Figure continued below)
Figure 4: Multiple sequence alignment of βNAC sequences from fungal species. Organisms included in this multiple sequence alignment include: *Rhizoctonia solani*, *Cutaneotrichosporon oleaginosus*, *Cryptococcus amylolentus*, *Cryptococcus depauperatus*, *Kwoniella dejecticola*, *Xylona heveae*, *Marissonia brunnea*, *Neurospora crassa*, *Rosellinia necatrix*, *Talaromyces islandicus*, *Spathaspora passalidarium* (Y-27907), *Candida albicans*, *Candida tanzawawaensis* (NRRL Y-17324), *Saccharomyces cerevisiae* (Beta-1 NAC), *Saccharomyces cerevisiae* (Beta-2 NAC), and *Candida glabrata*. Numbers to the right of the species names indicates on which residue the following sequence begins. Red is ≥ 80% conservation, blue ≥ 60% and < 80% conservation. FoS = 1.0, the highest stringency for establishing conservation. The homologues of the *C. elegans* putative caspase cleavage site are highlighted in green.
Figure 5: Multiple sequence alignment of βNAC sequences from animal model organisms, excluding those with natively shortened sequences. This multiple sequence alignment compares βNAC polypeptide sequences from model organisms *C. elegans*, *H. sapiens*, *M. musculus*, and *S. salar*. Numbers preceding sequences identify the residue number upon which the sequence to the right begins. Letters in red indicate residues with 80% conservation or greater, and letters in blue indicate residues of conservation greater than or equal to 60% and less than 80%. FoS = 1.0, which is the highest stringency setting for determining conservation. The putative caspase cleavage site on ICD-1, and its homologues in the other animal species, is highlighted in green.
The alignment of *C. elegans* and human sequences of αNAC. Based on Figure 1 from Wang et al. the residues of human αNAC that are known to be active in NAC dimerization are identified in yellow. In that way, based on homology, the residues of *C. elegans* ICD-2 that are active in NAC dimerization can be identified. A similar procedure where by residues on human βNAC that are actively involved in dimerization were identified from Figure 1 of Wang et al and the homologous residues on ICD-1 were found by with reference to multiple sequences alignments containing both *C. elegans* and human βNAC sequences.
**Figure 7:** The Phyre 2 predicted structural model of C. elegans ICD-1. Phyre 2 is a protein modeling program that predicts the structure of a protein from its sequence of amino acids. Red colored residues are 80% conserved or greater; blue colored residues are greater than 60% conserved and less than 80% conserved; gray colored residues are less than 60% conserved; all with respect to multiple sequence alignment of βNAC sequences from *Caenorhabditis elegans, Homo sapiens, Mus musculus*, and *Salmo salar*. Aspartate 133 of the putative caspase cleavage site is indicated in yellow. The caspase cleavage site occupies a relatively conserved position (red behind yellow text) and is exposed on the inner surface of a cleft flanked on the left by histidine 19, (top of gray alpha-helix) and on the right by lysine 135 (right, gray), and out from the plane of the image by glutamate 137 (below, blue).
Figure 8: Two images of the Phyre 2 predicted structure of ICD-1 colored for conservation with data obtained from the multiple sequence alignment produced by ConSurf comparing conservation of residues between ICD-1 and 150 similar protein sequences taken from online proteomic libraries, selected in virtue of sequence homology with ICD-1. Magenta indicates regions of conservation and cyan indicates regions of variability. The darker the coloring the greater conservation or variability. Grey indicates areas that are on average between conservation and variability. Panel A shows the outside of the pocket containing the putative caspase cleavage site (aspartate 133). The caspase cleavage site is the visible as the section of peptide backbone in dark cyan between a section of light cyan to the left and a section of gray to the right. Panel B shows the inside of the pocket containing the caspase cleavage site. The site’s aspartate 133 is visible as the dark cyan residue directly behind the yellow text. To the upper right in dark cyan is the residue lysine 135. The inside of the pocket is mostly in gray. The alpha-helices forming the outer surfaces of the pocket are colored dark maroon on their outer surfaces and gray on their inner surfaces.
**Figure 9:** The predicted structure of the ICD-1 ICD-2 heterodimer generated by docking the Phyre 2 predicted structures in the docking simulator HADDOCK ICD-1 at the NAC dimerization domains. The resulting structure was predicted with a z-score of 0.9 and 67% of possible binding sites used; that is, a moderately low amount of error but a structure that is somewhat more compressed than expected, centered at the known NAC binding domain. In both panels the ICD-1 component of the heterodimer is colored in gray, red, and blue; red indicating conservation of 80% or greater, blue indicating conservation between 60% and 80%, and gray indicating conservation less than 60%, relative to the multiple sequence alignment of *Caenorhabditis elegans, Mus musculus, Salmo Salar, Homo sapiens, Lygus Hesperus, Fopius Arisanus*, and *Clonorchis sinensis*. In both panels the ICD-2 component of the heterodimer is in magenta. In panel A) the interior of the pocket containing the putative caspase cleavage site is visible in gray in the lower left-hand corner. The caspase cleavage site (aspartate 133) is exposed on the right-hand lip of that pocket. In Panel B) the outer side of the caspase cleavage site is visible on the reverse-S-shaped segment of protein backbone beneath the yellow text; it is also exposed from the back of the pocket as well as the interior. The addition of ICD-2 to form the heterodimer does not appear to occlude the caspase cleavage site on ICD-1, even in the present highly-compressed model.
Discussion

This investigation has attempted to determine if in *C. elegans* the chaperone protein, inhibitor of cell death 1 (ICD-1) is being cleaved during the induction of apoptosis. The pathway of apoptosis is highly conserved at many levels of metazoan life, especially between *C. elegans* and humans. Determining if there is a mechanism by which apoptosis is decisively engaged and how would be highly informative for any human process that involves apoptosis; such being neuronal development, neurodegenerative pathologies, aging, and developmental abnormalities. If it can be shown that ICD-1 is selectively cleaved as the step by which apoptosis is committed, then that information would provide a target for treatments of diseases where apoptosis occurs in an unregulated or overly-sensitive fashion, such as Parkinson’s disease. At first I intended to use Western analysis to provide evidence of ICD-1’s cleavage in populations of *C. elegans* under apoptotic conditions. Protein lysates from worms treated with the chemical stressors tunicamycin and thapsigargin to induce unfolded protein response and thus apoptotic conditions and unstressed wild-type worms were compared to determine changes in ICD-1’s size from one population to the other. I succeeded in adjusting the stringency of the Western blotting protocol to give results with a high degree of signal for ICD-1 and a low amount of noise. However, the bacterial contamination of the *C. elegans* colonies rendered the results of the comparative Western analysis inconclusive. In the absence of physical evidence of ICD-1’s cleavage I continued using computer modeling and comparative proteomics to explore the possibility of ICD-1’s cleavage. Multiple sequence alignments using βNAC sequences from several model organisms were made to map residues of ICD-1 and its βNAC homologues for conservation across and within phylogenetic groups. The program Phyre 2 was used to predict the unsolved structure of monomeric ICD-1, the conformation in which ICD-1 is predicted to be during the unfolded protein response, and from which the plausibility of caspase mediated cleavage could be determined. The program ConSurf was used to find 150 other polypeptides from online proteomic libraries with sequences similar to ICD-1 and then construct a multiple sequence alignment to map the conservation of ICD-1’s sequence among a large resource of proteins. The program HADDOC was used to predict the structure of the heterodimer of ICD-1 and ICD-2, a conformation ICD-1 may be in to some degree during the induction of the UPR. The predictive
models show that the putative caspase cleavage site on ICD-1 (residues 130 – 140) are not buried in the molecule but exposed on its surface on the edge of a pocket, for the both the monomer and the heterodimer. There is also enough surrounding space for the active residue of a caspase to dock to the site. Therefore, it is likely that ICD-1 is accessible to cleavage during the UPR. When the conservation mapping is applied to the predictive 3D protein models it is revealed that the region surrounding the putative caspase cleavage site on ICD-1 is the most variable region among the 150 proteins similar to ICD-1. That points toward a species specific mechanism for caspase cleavage, which is consistent with the fact that many caspases do have species specific active residues.

The series of Western blotting experiments were uninformative of the states of ICD-1 from chemically stressed versus unstressed wild-type populations, and therefore we were unable to conclude that chemically induced unfolded protein stress is associated with ICD-1 cleavage. However, we were also unable to conclude that ICD-1 is not cleaved during said conditions. I did identify the large protein that was detected in the Western blots as likely the bacterial homolog of ICD-1. Its presence in the lysates can be explained by the bacterial food source used to feed C. elegans populations. Further work characterizing the nature of ICD-1 during misfolded protein stress will be necessary.

The structural predictions from this study support the possibility that monomeric ICD-1, and to a lesser degree, heterodimeric ICD-1, are cleavable in vivo at the putative caspase cleavage site. The caspase cleavage site is exposed on the surface of the peptide on the edge of a pocked formed by two alpha-helices. The side chain of the site’s active aspartate residue points into the space of the pocket and its backbone is exposed on the outer surface. There are also histidine and glutamate residues flanking the cleavage site, possibly to facilitate an ionic interaction in caspase binding. Monomeric ICD-1 is thought to become available during the induction of the unfolded protein response, which is then followed by apoptosis when unfolded protein stress exceeds a threshold. Concurrent with the release of ICD-1 from ICD-2 during misfolded protein stress is the activation of caspase. Thus, the presence of both monomeric ICD-1 with a highly exposed putative caspase cleavage site and apoptotic caspases supports the possibility that vulnerable ICD-1 is exposed to and cleaved by CED3 or other caspases under
stress-inducing conditions. Both human βNAC and Drosophila Bicaudal/βNAC have been shown to be cleaved by a caspase in vitro, and it is known that there is a high degree of conservation between human and C. elegans apoptotic machinery. Therefore, it is possible that the predicted structure of monomeric C. elegans ICD-1 would fold in such a way that the putative caspase cleavage site would be exposed on its surface and accessible to binding and cleavage by a caspase endogenous to C. elegans.

Although putative caspase cleavage site of heterodimeric ICD-1 is somewhat exposed, it is not likely that it would come into contact with caspases and undergo cleavage. The caspase cleavage site and the surrounding structure are largely in the same conformation as in the monomer. The only difference is that a lobe of ICD-2 juts out near the region of ICD-1 that has the cleavage site. The lobe of ICD-2 does not occlude the caspase cleavage site, but seems to form the boundaries of a bowl-like negative space, like a continuation of the pocket housing the cleavage site. The model of the heterodimer is more compressed than expected, so there may be more space between ICD-1 and ICD-2. But, even with the compaction in the model as it appears in figure 9, it still seems plausible that a caspase could dock to the open space where the cleavage site is exposed on the protein’s surface. The abundance of the NAC heterodimer decreases as the abundance of monomeric NAC subunits increases and caspases are released from the mitochondria. Once the caspases are released they would more likely encounter monomeric ICD-1 than the heterodimeric form. Therefore, the release of caspases will likely lead to the cleavage of monomeric ICD-1 regardless of the vulnerability of the heterodimeric ICD-1. Since the NAC heterodimer is involved in localizing peptides to the mitochondria it would be placed in close proximity to the origin of the caspases. How ICD-1 interacts with other proteins and how those interactions may have been shaped by evolution is revealed by the conservation of ICD-1 residues.

According to ConSurf, a bioinformatics tool well-equipped to provide conservation data, the regions most conserved in C. elegans ICD-1 and other similar structures are the regions surrounding the NAC dimerization domain (mostly between residues 24 and 96) and the distal face of the alpha-helix on the C-terminal side of the putative caspase cleavage site (included in 141 – 150). Interestingly, the region of greatest variability is the surrounding the putative caspase cleavage site (130 – 134). The pocket containing the cleavage site has the greatest variability; the
The greater degree of conservation in the NAC dimerization domain indicates that there is a selective force for retaining the same pattern of residues in the domain across many phylogenetic lineages. Also, since the structures of αNAC and ICD-1 are very similar across this domain, it is likely that models of αNAC were included in analyzing the conservation of ICD-1 sequence, which supports the accepted notion that αNAC and ICD-1 share evolutionary history in a single enzyme, as is observed in archeabacteria, and that a NAC domain interacts best with a domain very similar to itself.

Models with less than 70% of their structure modelled at 90% confidence or greater are ineligible for submission to 3DLigandSite. Inferences made from such models should be considered cautiously; putative functions must be considered in reference to the confidence by which its respective section of structure was modelled. The models of ICD-1 and αNAC were modelled by Phyre 2 with less than 70% at 90% confidence or greater. The regions surrounding ICD-1’s putative caspase cleavage site were fortunately predicted at greater than 90% so the inferences made about the exposure and accessibility of that site on the monomer can be made with sufficient confidence. However, since structure of the heterodimer was predicted using the whole structures of the monomers the prediction of the heterodimer’s structure has associated with it the compound uncertainty of both monomers and the simulated docking interaction. Therefore, the following analysis of monomeric ICD-1’s structure has stronger grounds for acceptance than analysis of the predicted structure of the heterodimer.

Unlike the regions involved in NAC dimerization, the region surrounding the putative caspase cleavage site has a very high degree of variability. It is the only region of ICD-1 that is variable. One explanation for the reported variability is that the models compared for conservation included αNAC from a variety of species; since αNAC does not contain a caspase cleavage site, they would differ greatly from ICD-1 at the region containing its caspase cleavage site. Another explanation for the variability is that the region interacts with species specific proteins or ligands. Not all phylogenetic groups may have developed or preserved a caspase cleavage site. And, if several distant groups have preserved a caspase cleavage site at that region it may be the case that the caspas of those groups have developed differences and require a different region to bind with stability. C. elegans has at most two caspases, whereas humans have
as many as 12. There are slight species specific differences between caspases of different origins, and the caspases of a single organism likely differ from each other due to selection to perform functions unique to each organism. The aspartate of the site is highly variable, even when compared with the rest of the region surrounding it.

The multiple sequence alignments of ICD-1 from several important multicellular model organisms shows that many organisms retain the site identified as the putative caspase cleavage site. That supports the idea that the putative caspase cleavage site is functional and that the function it performs is important to many, but not all, multicellular organisms. The fact that the LGPD sequence is very highly conserved in fungi and is not as highly conserved in animals points to species specific selective pressures and species specific solutions for adapting to those pressures. That *Lygus hesperus*, *Fopius arisanus*, *Clonorchis sinensis* and *Anthurium amnicola* have ICD-1 sequences that terminate before reaching the point where other species have a putative caspase cleavage site may mean that if cleaving ICD-1 performs a function in the unfolded protein response or other scenarios, those species have been preserved while maintaining ICD-1 continuously in a state that is equivalent to its cleavage in other species. If the cleavage of ICD-1 leads to the apoptotic death of a cell in which that cleavage occurs, then some species have other processes at play that prevent a cleaved length ICD-1 from initiating apoptosis; or, the cleavage of ICD-1 is not sufficient for the induction of apoptosis. On the other hand, there are evidently some species which express ‘full-length’ ICD-1, but with without a putative caspase cleavage site where one would be expected. The map of residue conservation drawn by ConSurf from online protein libraries of proteins similar to *C. elegans* beta-NAC shows that the putative caspase cleavage site is the most variable sequence of ICD-1-like proteins. And, although that result may be skewed by alignment with αNAC sequences which do not contain putative caspase cleavage sites, it nonetheless means that there are many organisms that are able to function with a ICD-1 without a caspase cleavage site. Since the site seems to be regularly conserved in most multicellular organisms, the bulk of ICD-1 sequences without putative caspase cleavage sites are likely single-celled organisms. As NAC is known to be involved in cell-fate determination, transcription, and cell saving processes a function that alters its structure and function would more likely be useful in organisms of a variety of cell types and thus different conditions that require different responses to stress and transcription; that is, a multicellular organism.
Based on the models obtained in this study, it seems that ICD-1’s putative caspase cleavage site is exposed and accessible during its physiologically relevant conformations (monomeric and heterodimeric). Furthermore, ICD-1’s chaperone and cell-saving mechanisms generally bring it most areas of the cell, especially to compartments where active caspases that could cleave it. Therefore, if ICD-1 can be cleaved by a caspase in vivo, it is likely the activation of the caspase during apoptosis that is sufficient for triggering cleavage of ICD-1. The conservation data from multiple sequence alignments support the cleavage of ICD-1 being an important feature of multicellular life as it is found in identical states across the three multicellular kingdoms, but also that its role has been compensated for by other processes in the organisms that lack it. This study was intended to provide information to direct future experiments to confirm the predictions made here. This report concludes with suggestions for further studies.

**Future Research Investigations**

The best option for following up the inconclusive Western blotting experiment of this study is to repeat the experiment using the optimized protocol with the fresh samples of wild-type C. elegans. Since it is possible that the procedure was optimized for the incorrect protein (the fungal homolog rather than C. elegans ICD-1) the Western blotting procedure may have to be re-optimized.

The lack of evidence for ICD-1’s cleavage during the chemically directed ER stress does not exclude the possibility that ICD-1 is cleaved under other conditions associated with apoptosis. To provide a condition during which apoptosis is certain, C. elegans populations may be depleted for αNAC. Although αNAC depletion is an artificial condition, it is known to initiate apoptosis through ER stress (Arsenovic et al., 2012).

Since the putative caspase cleavage sites on monomeric ICD-1, and to a lesser degree the heterodimer, are exposed their cleavage may be possible and further studies could be invested in
to confirm that finding. To provide structural models with greater certainty X-ray crystallography could be used to formally solve the structures of C. elegans Alpha NAC and ICD-1. Kinetic studies could also be performed with the C. elegans heterodimer and the apoptotic caspases in *vitro* at physiologically relevant relative concentrations. An experiment could also be done by which CED-3 is tagged with a green fluorescent protein and αNAC is tagged with a red fluorescent protein, and the active cysteine residue on CED-3 is changed to an inactive residue, and the population of worms is chemically stressed to induce the unfolded protein response and their cells are viewed with microscopy to determine if CED-3 is docking near αNAC, which would mean CED-3 is binding to the ICD-1 component of a heterodimer.

Since it is likely that ICD-1 is accessible for cleavage during at least its monomeric conformation, it would be beneficial to learn which, if any, caspases and other proteases endogenous to C. elegans may realize that cleavage. One possible study would be to perform western blots with stressed and unstressed populations of worms with single caspase genes either depleted or knocked out. If there is no evidence of cleavage when one or more caspase gene is knocked out, then that caspase is likely the one that cleaves ICD-1. Another experiment could be to perform kinetic tests with C. elegans monomeric and ICD-1 and the NAC heterodimer and a library of caspases and proteases from C. elegans’ proteome. Those enzymes that can cleave the NAC and its component in *vitro* would then be examined by cell sorting to determine if they may come in contact with ICD-1 during ICD-1’s physiological functions.
References


