EVOLUTIONARY AND BIOCHEMICAL CHARACTERIZATION OF EFFECTOR CASPASES

by

SUMAN SHRESTHA

Presented to the Faculty of the Graduate School of

The University of Texas at Arlington in Partial Fulfillment

of the Requirements

for the Degree of

DOCTOR OF PHILOSOPHY

THE UNIVERSITY OF TEXAS AT ARLINGTON

May 2021

ACKNOWEDGEMENTS

I would like to thank Dr. Clay Clark for his mentorship throughout the years. He has helped me to be not only a scientist but also a good human being. I would also like to thank my committee, Dr. Laura Mydlarz, Dr. Mark Pellegrino, Dr. Esther Betran, and Dr. Subhrangsu Mandal. I am grateful for scientific guidance and supportive critiques throughout my graduate student career. Special thanks to Dr. Mydlarz and Dr. Pellegino, and their lab members for the opportunity to collaborate and learn more. Thanks to Dr. Kimberly Bowles for all of the help.

I would like to thank my former lab mates Dr. Robert Grinshpon, Dr. Melvin Thomas, who taught me all the protein works. Thanks to my current labmates, Liqi Yao, Isha Joglekar, and Mithun Nag, for the good company and contribution to my research. I would like to thank undergraduates who worked with me, Jessica Tung, Godswill Nwaosu, and Tina Nguyen, for their assistance. Thanks to all my amazing friends and colleagues I have met in UTA, especially, Balan Ramesh (writing buddy), Adnan Qureshi, Sarbjeet Niraula, Santosh Dhamala, Swatantra Neupane, Raisa Amin, Farah Shamma, Fatema Ruma, Douja Chamseddine, and Bradford Dimos, for your company, support, and encouragement.

My family here, uncle and aunt, cousins- Rajkumar/Muna, Sunil/Sangita, and Samana, Niece- Rajani, nephews- Muraj, Sahas, and Suhas, thanks for their love and care, and making me feel home away from home.

March 31st, 2021

Dedication

This work is dedicated to my incredibly supportive family. My late dad would have been proud of his son's achievement. My mom had to overcome many obstacles and struggles to bring me up. You are a fighter and a real champion, mom. After losing our dad, my big brother has been my guardian and helped me financially throughout my college study. Thank you, brother, for your support. I wouldn't be here if you were not my brother. My loving and caring wife, I understand the past couple of years hasn't been easy, but you always supported and encouraged me. Thank you for your love and care. Last but not least, all my siblings, sister-in-law, brothers-in-law, nephews, nieces, thank you all for your unwavering support through this journey.

Abstract

Evolutionary and Biochemical Characterization of Effector Caspases

Suman Shrestha, PhD

The University of Texas at Arlington, 2021

Supervising professor: A. Clay Clark, Ph.D.

Caspases are an ancient class of cysteine-dependent aspartate-specific proteases that control apoptosis, responsible for cell differentiation, and maintain cellular homeostasis in multicellular organisms. Dysregulation of caspase functions leads to many human diseases, including cancer and neurological disorders. Due to the involvement of caspases in several diseases, it is crucial to understand how they are regulated. Extant human caspases have been studied extensively for over two decades. However, the success in the development of therapeutics against caspase dysfunctions has very little success to date. Hence, the evolutionary approaches for studying caspases could help understand their structure and function and the overall protein evolution. The caspase family is an excellent model to study protein evolution because all caspases are produced as zymogens that must be activated to gain full activity. The protein structures and substrate specificity are conserved through millions of years of evolution, and some regulatory features are ancient, and therefore common, while other features are modern and cluster-specific.

Here, we characterized coral caspases as a model to study caspase evolution and understand the role of caspases in coral bleaching. Furthermore, we examined substrate specificity and evolutionary folding trajectories of extant effector caspases and their resurrected common ancestor.

We identified and characterized two caspases, each from *Orbicella faveolata*, a disease-sensitive reef-building coral, and *Porites astreoides*, a disease-resistant reef-building coral. Our biochemical analyses suggest that corals have complex-apoptotic signaling cascades similar to those of vertebrates. We reported the first X-ray crystal structure of a coral caspase, which reveals conservation of the caspase-hemoglobinase fold in coral. An N-terminal peptide bound near the active site may serve as a regulatory exosite. Our Folding study shows that the effector caspases unfold by a minimum three-state equilibrium model at pH 7.5. The folding landscape was first established with the common ancestor and was then retained for >650 million years. Caspase-7 is closest to the common ancestor, whereas procaspases-6 and -3 show evolutionary changes, making them more stable. Moreover, modern human caspase-7 has lost stability as well as catalytic efficiency relative to their ancestors.

Firstly, this study provides evidence that apoptotic mechanisms are ancient and well conserved in metazoans. Secondly, the energy landscape of effector caspases is maintained for millions of years. Finally, this research sheds light on how caspase substrate specificity evolved.

Table of Contents

Chapter 1

Introduction

Caspases

Cysteine-dependent aspartate-directed proteases are an evolutionarily conserved group of proteases called “caspases.” Caspases are primarily associated with the regulation and execution of programmed cell death (PCD), or apoptosis, in all metazoans (1–3). Caspases are involved in the controlled dismantling of intracellular components while avoiding inflammation and damage to surrounding cells. Caspases are also involved in numerous non-apoptotic processes such as pyroptosis, cell differentiation, proliferation, and cellular remodeling (4, 5). Caspases were first reported in 1992 by two independent research groups as a protease responsible for activating the precursor of interleukin-1β hence named interleukin-1β-converting enzyme (ICE) (6, 7). After that, the search of caspase homologs began in other metazoan species and even in lower eukaryotes and prokaryotes in order to understand the evolution of caspases and apoptosis (8–11). Later, full details of caspases, including structure, activation pathways, and their roles in apoptosis, were studied in a model organism, *Caenorhabditis elegans* (12–14). In the last three decades, studies of apoptosis and caspases illustrate the roles of caspases as initiators and executioners in apoptosis; however, its role in non-apoptotic functions is still under study (15).

The human caspase family is broadly classified as apoptotic or inflammatory depending on their involvement in the life and death of a cell. Inflammatory caspases are cytokine activators and comprise caspase-1, -4, and -5, all of which have a caspase-recruitment domain (CARD) in the N-terminus (1, 16, 17) (Fig. 1). Within this

group, caspase-1 is the enzyme that has been well characterized and known for processing IL-1β involved in inflammation (18). The apoptotic caspases are further divided into initiators and effectors based on their entry into apoptotic cascade and domain organization (1, 11). Initiator caspases are produced as a monomer and activated by proximity-induced dimerization, which acts upstream in cascade (19). A monomer of initiator caspase consists of long N-terminal prodomain, Caspase-2, and -9 have CARD and caspase-8, and -10 have death effector domain (DED) (11) (Fig 1). Effector caspases are produced initially as a dimeric inactive form called zymogen, and a monomer consists of an N-terminal short prodomain connected with the C-terminal protease domain (Fig. 1). Caspases-3, -6, and -7 are effector caspases and are activated by cleavage of the intersubunit linker of zymogen by initiator caspases (11, 20).

Caspases-3, -6 and -7

A stable inactive zymogen of the effector caspase is thought to prevent inappropriate activation (11). The cleavage of the intersubunit linker between the large and small subunits of effector caspases allows conformational changes or rearrangement of the active site forming loop. The caspase then becomes active (21, 22). Among the effector caspases, caspase-3 is the primary effector caspase that cleaves the majority of substrates in cells undergoing apoptosis (23). Structurally, all three caspases are similar, whereas functionally, caspase-6 is somewhat different than caspase-3 and -7 (11). Caspase-3 and -7 are highly similar in terms of substrate specificity and with overlapping function in apoptosis, with few known exceptions. For

example, poly (ADP-ribose) polymerase (PARP) more efficiently by caspase-7 than caspase-3 (24). Caspase-6 has different specificity than the other two effectors, and few caspase-6 substrates are known, including lamin A (25). Together, there is no clear distinction among three effector caspases in apoptotic signaling, although orthologues of all three caspases are found in vertebrates as well as invertebrates. Caspase singly or in combination knocked out mice experiments suggest that all effector caspases have different roles in a different context however, the functions of caspase-6 and -7 are still not clear (16, 26).

Apoptotic pathways

The caspases are activated primarily by two core pathways: extrinsic pathway and intrinsic pathway. Schematics of apoptotic pathways is shown in figure 2.

Extrinsic pathway

Apoptotic signaling through this pathway triggers when extracellular ligands bind to the transmembrane receptors in the response of extracellular signal forming a death-inducing signaling complex (DISC). With the DISC formation, procaspase-8 gets activated and starts the cascade of caspase activation, which eventually activates effector caspases (17, 27, 28) (Fig. 2).

Intrinsic pathway

Various intracellular stresses and genotoxic damage trigger the intrinsic pathway or mitochondrial-mediated pathway. This pathway is mediated by Bax/Bak insertion into the mitochondrial membrane, inducing the release of cytochrome c by mitochondria. Cytochrome c then combines with Apaf-1 and procaspase-9 forming a heptameric

complex known as the apoptosome. Apoptosome formation results in caspase-9 activation, leading to the downstream activation of caspase-3, -6, and -7, which ultimately are responsible for the demolition of cells (17, 27, 28) (Fig. 2).

Caspases in human diseases

Dysregulation of caspases underlies human diseases, including cancer, autoimmune diseases, inflammatory disorders, and infectious pathologies (3, 4, 29–31). Defects in caspase activation cause inadequate cell death resulting in tumorigenesis, whereas overactivation of caspases causes excessive cell death resulting in neurogenerative disorders like Alzheimer's disease (3, 29). Mutations in *casp* genes are not frequent events, but malignant cells gain a capacity to breach the activation mechanism of caspases. Despite the low rate of mutation of the *casp* gene, the reduced expression of apoptotic caspases due to specific inactivating mutations has been reported in various cancers. Mutations have been found in all of the proapoptotic caspases causing caspase inactivation and leading to multiple forms of cancers such as neuroblastoma, carcinoma non-Hodgkin's lymphoma (NHL), lung cancer, and endometrial cancer (32–36). Numerous attempts have been made to use caspases as a target for cancer therapy, either activating or inhibiting, but the attempts have not been successful because of limitation in delivery or cross-reactivity (29). The long list of diseases caused due to dysregulation of caspases and inability to design better therapies has dire consequences in human health. Hence, the study of caspases to understand how these enzymes work and how they can be controlled will be beneficial

to understand how to modulate their function and develop therapeutics against caspase-related diseases.

Evolution of caspases

Following the discovery of caspases in *C. elegans,* researchers have been focused on their role in cell development and apoptosis. Since then, a large number of caspases have been identified in vertebrates and invertebrates as well, and caspase function is conserved throughout metazoans (37). After studying caspases in nematodes, the apoptotic network was thought to be relatively simple, with only one initiator and one effector caspase (10). Also, it was predicted that apoptotic regulatory networks evolved linearly from simple to complex manner (10). However, genome sequencing of many invertebrates like *Nematostella vectensis* and *Orbicella faveolata*, contradicts the simple ancestral apoptotic network by revealing the higher numbers of apoptotic components in these species (10, 38). For example, *Hydra magnipapillata* contains 17 homologs of caspases (39). Hence, the number of caspases is not the same in all species. The different number of caspases in individual species could be due to lineage-specific expansions and losses, although caspase has a complex evolutionary history(10). Also, a varying number of apoptotic machinery, including caspases, suggests the diversification of species-specific functions (26). These specific caspases may play novel roles in various species. Findings of caspase function in basal metazoans suggest that caspases possess a long evolutionary history and maybe as old as the advent of multicellular life forms (40). Therefore, it is beneficial to examine the

evolutionary history of caspases which could lead to new knowledge about the evolution of the caspase activity and its role in causing various diseases.

Cnidarians as a model to study the evolution of caspases

C. elegans and Drosophila were the model organisms to study caspases and apoptosis, which established the molecular bases of apoptosis. However, after the genome sequencing of evolutionarily older metazoans, cnidarians, it is clear that cell death of nematodes and insects does not reflect ancestral characteristics and can't be a model to study caspase evolution. The presence of multiple caspases as well as other apoptotic regulators, such as Bcl-2 and Apaf-1 in cnidarians like *Hydra magnipapillata, Hydractinia echinate, O. faveolata* and *Porites astreoides* suggests that cnidarians may have a complex apoptotic system more similar to that of vertebrates (38, 41, 42). Additionally, Moya et al. reported the interaction of caspase-8-like and FADD-like proteins in *Acropora millepora,* suggesting the presence of an extrinsic apoptotic pathway (43). Together, these characteristics of cnidaria indicate the level of complexity and the high degree of conservation of apoptotic networks in coral to mammals, although the functional interactions among the apoptotic components are unknown.

Moreover, corals are on the decline due to coral bleaching, which is defined as loss of the photosynthetic pigments and endosymbiotic *Symbiodinium* due to overheating or increasing ocean acidity, an effect of global warming (44, 45). Several studies have indicated the role of caspase and apoptosis in coral bleaching, showing caspase is one of the players in bleaching. Fuess et al. showed the different immune responses in disease susceptible species *Orbicella faveolata* and disease-resistant

species *Porites astreoides*. *O. faveolata* activated caspase-mediated apoptosis, and *P. astreoides* activated autophagic pathways (46). A few studies reported the differential expression of *casp* gene in the diseased corals (39, 47–49). Therefore, the characterization of caspases in corals would provide not only evolutionary perspectives of caspase and apoptotic networks but also a much better understanding of the roles of apoptosis in coral bleaching and stress responses.

Protein folding and caspase as a model

Translation produces an unfolded protein that must fold to the correct conformation to be functional. More than a half-century ago, the seminal studies of Anfinsen and colleagues established the fact that amino acid sequence in the polypeptide chain encodes all the information needed for a protein to fold into the correct three-dimensional structure (50). However, to date, how amino acid sequence dictates its three-dimensional atomic structure, the so-called protein folding problem has not been decoded despite many scientist's decades of work. The critical questions that remain unresolved are (i) what is the physical code that amino acid sequences determine three-dimensional structure? (ii) How do proteins fold rapidly and accurately? (iii) Is it possible to predict the structure of a protein from its amino acid sequence by developing a computer algorithm? To answer these questions a variety of experimental, theoretical, and computational tools have been created (51–53). Mainly, X-ray crystallography to determine the protein structure at an atomic level and equilibrium folding by chemical, thermal or pH denaturation to determine the folding mechanism of proteins have been two leading research fields in the quest of solving the folding

problem (54, 55). Solving the protein folding problem would be beneficial for understanding protein function in the cell as protein misfolding and aggregation are linked with the number of diseases (56, 57).

Most of the protein folding studies have focused on monomers and relatively simple dimers described by two-state pathways where only native state and unfolded state are populated. About two-thirds of proteins in eukaryotic organisms, however, are believed to be oligomeric (55, 58). For example, protein folding model systems include ribonuclease H (RNase H), lysozyme, cytochrome c, arc dimer, and dihydrofolate reductase. Studies of these proteins have provided knowledge of folding rates and pathways at the level of secondary and tertiary structure and some residue-specific information. However, monomeric systems lack the interchain interactions and coordination of proteins in a complex, which would be essential to understand their role in a biological settings (59).

Many studies focus on the evolution of protein function while the evolutionary constraints on folding are lacking. Over the course of time, the energy landscape – the ability of proteins to fold properly into the correct structure and reasonable time scale avoiding misfolding – must be maintained to function in the biological system. To date, however, there are not many studies to examine the evolution of folding landscape. Two recent studies on RNase H showed that *E. coli* RNase H and *Thermus thermophilus* RNase H are evolutionarily billions of years apart, but their folding pathways are conserved with a similar folding intermediate (53, 60). In addition, studies showed modern RNases H evolved to be kinetically more stable than their most recent common ancestor. Commonly we could expect the evolution of folding landscape to avoid kinetic

traps or misfolded states and improve kinetic stability over time; however, we do not have enough evidence to back that up. Understanding folding evolution is critical to learn the fitness and change in stability over time. A comparative study among the extant homologs enriched our understanding of the folding mechanism, and residue-specific function but did not give a clear picture of how folding mechanism changed from ancestor to modern proteins and its effect on biological functions. Most importantly, evolutionary folding studies reveal the root of the change in the energy landscape.

For the study of ancestors, ancestral state reconstruction (ASR) predicts the ancestral protein sequence by using evolutionary relationships of the modern homologs (61). Caspases are an ancient protease family that is well characterized functionally, but little is known about the evolution of the folding landscape and stability. Hence, we sought to examine how ancestral folding conformations modulate the function of extant caspase clusters. Characterizing ancestral caspase and the extant homologs will reveal the changes in the folding landscape which also can shed light on the sequence determinants of the caspase folding landscape.

Caspase substrate specificity

Substrate profiling of human caspases has been performed extensively by multiple approaches like peptide libraries (62), directed evolution screening (63), N-terminomics (64), Protomap (65), and PICS (66). All those studies agree at least on the requirement of aspartate at the P1 position of the substrate is conserved in all caspases due to conserved S1 binding pocket with less stringent specificity at P2 and P3 positions. Primarily, specificity of caspase is defined by the preference of amino acids at

the P4 position. Thus, based on amino acid preference at the P4 position, caspases can be classified into three groups. Group I (caspase-1, -4, -5, -14) prefer bulky amino acids such as tryptophan and leucine, group II (caspases-2, -3, -7) prefer charged amino acid mainly aspartate, and group III (caspases-6, -8, -9, -10) prefer hydrophobic amino acids such as valine, leucine, and isoleucine (2). Among effector caspases, the general consensus of preferred tetrapeptide substrate of caspase-3 and -7 is DEVD and VEID for caspase-6 (63). Substrate preference of caspase is determined mainly by the active site forming loops. However, it is reported that some of these differences could derive from exosites – substrate binding sites outside the active site. For example, caspase-7 uses the four lysine residues (38-KKKK-41) for the rapid proteolysis of the poly (ADP-ribose) polymerase 1 (PARP-1) (67), and caspase-6 uses tri-arginine (42-RRR-44) for the hydrolysis of protein substrates (68).

Evolutionary analysis of substrate specificity inferring sequences of ancestral proteins using ancestral state reconstruction (ASR) is more common. ASR allows vertical comparisons (comparing ancestral to modern enzymes) as well as horizontal comparisons (comparing modern enzymes from multiple species) of amino acid sequences to examine the evolutionary change in specificity. A previous study on the evolution of caspase specificity showed that the common ancestor of effector caspases was promiscuous (69). Later on, human caspase-6 developed specificity towards valine at P4 position, which evolved early after duplication of a common ancestor to the caspase-6 and caspase-3/7 lineages. Three vertical mutations were sufficient for the specificity change from the common ancestor to the caspase-6 ancestor (69). The development of therapeutics against caspases has been a difficult task due to the

conserved specificity and the resulting cross reactivity. Hence, a detailed understanding of the molecular basis of substrate specificity of caspases is critical for developing selective therapeutics. In this study, we took an evolutionary approach to study the substrate specificity of caspase-7 in more detail.

This dissertation is organized into 5 chapters. Chapter 2 describes the study of caspase evolution using coral as a model, where we have characterized caspases from two reef-building corals. Chapter 3 presents our data on the evolutionary folding landscape of effector caspases. Chapter 4 describes the evolution of substrate specificity of caspase-7 using ancestral protein reconstruction (APR). The last chapter will discuss all the results in brief then summarize the main findings of our study.

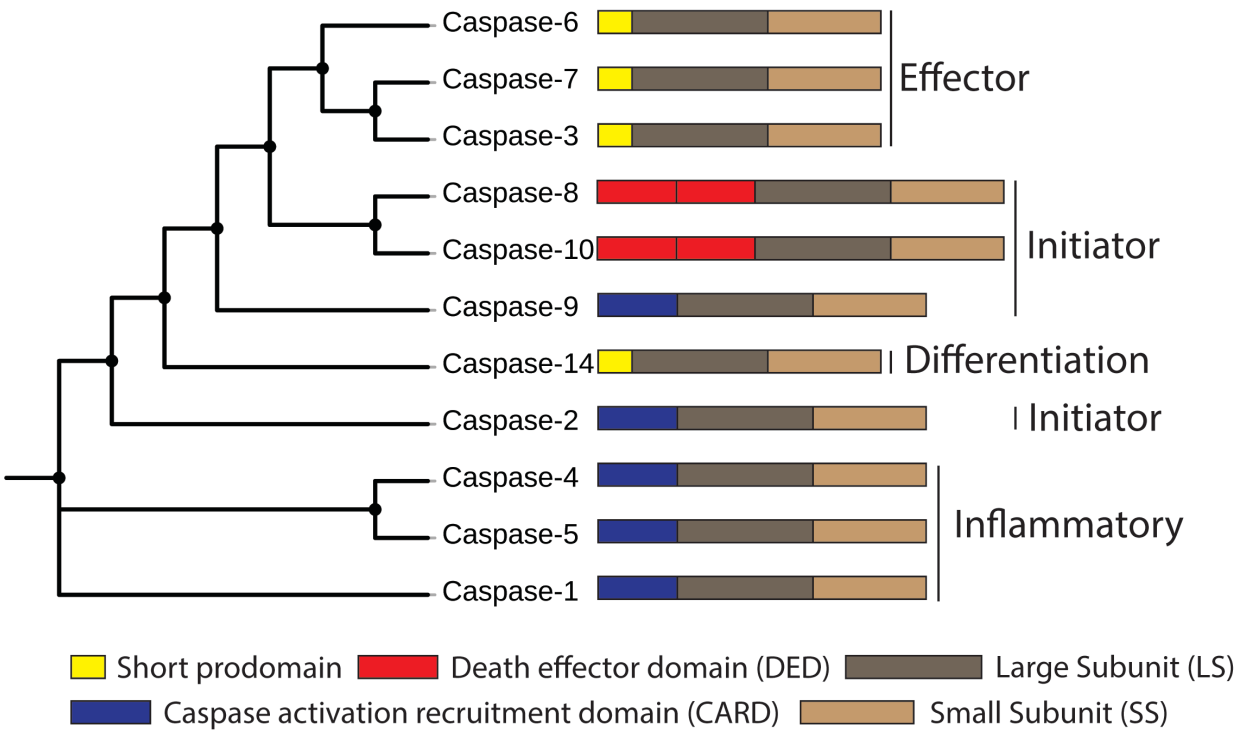


Figure 1: The Phylogenetic relationship of human caspases and schematics of the caspases showing the domain organization. A phylogenetic tree of human caspases was generated based on the alignment of the amino acid sequences by the maximum likelihood method. The physiological roles of caspases also indicated on right side of the schematics.

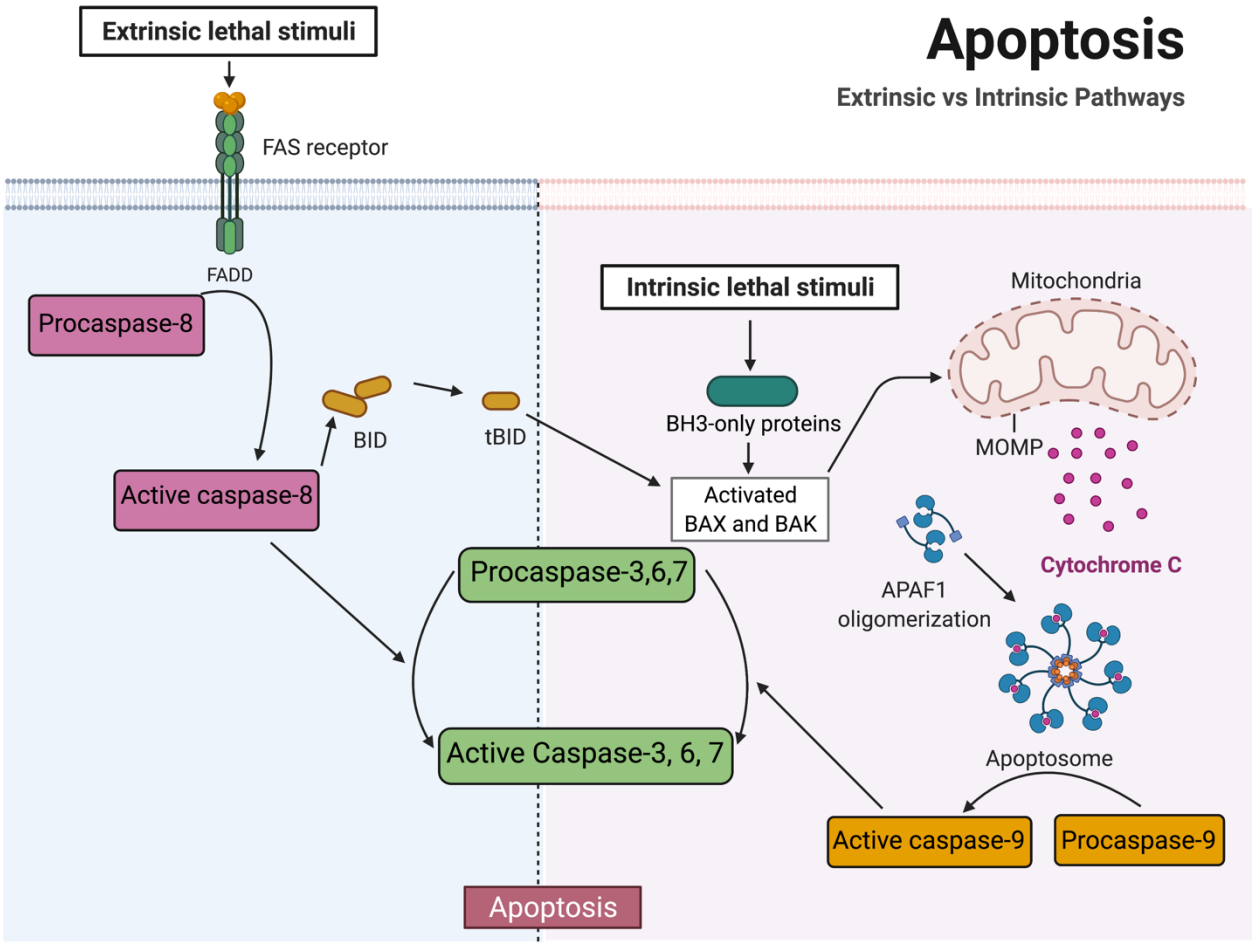


Figure 2: Extrinsic and Intrinsic pathways of apoptosis. Pathways are described in text. Abbreviations: FADD: Fas associated death domain, BID: BH3-interacting domain, tBID: truncated BID, MOMP: mitochondrial outer membrane permeabilization, APAF1: Apoptotic protease activating factor-1.

References

1. Stennicke, H. R., and Salvesen, G. S. (1999) Caspases: Preparation and characterization. *Methods A Companion to Methods Enzymol.* **17**, 313–319
2. Clark, A. C. (2016) Caspase Allostery and Conformational Selection. *Chem. Rev.* **116**, 6666–6706
3. Van Opdenbosch, N., and Lamkanfi, M. (2019) Caspases in Cell Death, Inflammation, and Disease. *Immunity*. **50**, 1352–1364
4. Shalini, S., Dorstyn, L., Dawar, S., and Kumar, S. (2015) Old, new and emerging functions of caspases. *Cell Death Differ.* **22**, 526–539
5. Bell, R. A. V., and Megeney, L. A. (2017) Evolution of caspase-mediated cell death and differentiation: Twins separated at birth. *Cell Death Differ.* **24**, 1359–1368
6. Cerretti, D. P., Kozlosky, C. J., Mosley, B., Nelson, N., Van Ness, K., Greenstreet, T. A., March, C. J., Kronheim, S. R., Druck, T., Cannizzaro, L. A., Huebner, K., and Black, R. A. (1992) Molecular cloning of the interleukin-1β converting enzyme. *Science (80-.).* 10.1126/science.1373520
7. Thornberry, N. A., Bull, H. G., Calaycay, J. R., Chapman, K. T., Howard, A. D., Kostura, M. J., Miller, D. K., Molineaux, S. M., Weidner, J. R., Aunins, J., Elliston, K. O., Ayala, J. M., Casano, F. J., Chin, J., Ding, G. J. F., Egger, L. A., Gaffney, E. P., Limjuco, G., Palyha, O. C., Raju, S. M., Rolando, A. M., Salley, J. P., Yamin, T. T., Lee, T. D., Shively, J. E., MacCross, M., Mumford, R. A., Schmidt, J. A., and Tocci, M. J. (1992) A novel heterodimeric cysteine protease is required for interleukin-1βprocessing in monocytes. *Nature*. 10.1038/356768a0
8. Wiens, M., Krasko, A., Perovic, S., and Müller, W. E. G. (2003) Caspase-mediated apoptosis in sponges: Cloning and function of the phylogenetic oldest apoptotic proteases from Metazoa. *Biochim. Biophys. Acta - Mol. Cell Res.* **1593**, 179–189
9. Song, Z., McCall, K., and Steller, H. (1997) DCP-1, a Drosophila cell death protease essential for development. *Science (80-.).* **275**, 536–540
10. Zmasek, C. M., Zhang, Q., Ye, Y., and Godzik, A. (2007) Surprising complexity of the ancestral apoptosis network. *Genome Biol.* **8**, R226
11. Fuentes-Prior, P., and Salvesen, G. S. (2004) The protein structures that shape caspase activity, specificity, activation and inhibition. *Biochem. J.* **384**, 201–232
12. Wu, Y.-C., and Xue, D. (2003) Programmed Cell Death in C. elegans. in *Essentials of Apoptosis*, pp. 135–144, Humana Press, Totowa, NJ, 10.1007/978-1-59259-361-3_9
13. Lettre, G., and Hengartner, M. O. (2006) Developmental apoptosis in C. elegans: a complex CEDnario. *Nat. Rev. Mol. Cell Biol.* **7**, 97–108
14. Yuan, J., Shaham, S., Ledoux, S., Ellis, H. M., and Horvitz, H. R. (1993) The C. elegans cell death gene ced-3 encodes a protein similar to mammalian interleukin-1β-converting enzyme. *Cell*. 10.1016/0092-8674(93)90485-9
15. Nakajima, Y., and Kuranaga, E. (2017) Caspase-dependent non-apoptotic processes in development. *Cell Death Differ.* **24**, 1422–1430
16. Kumar, S. (2007) Caspase function in programmed cell death. *Cell Death Differ.* **14**, 32–43

17. MacKenzie, S. H., and Clark, A. C. (2012) Death by caspase dimerization. *Adv. Exp. Med. Biol.* **747**, 55–73
18. Wei, Y., Fox, T., Chambers, S. P., Sintchak, J., Coll, J. T., Golec, J. M. C., Swenson, L., Wilson, K. P., and Charifson, P. S. (2000) The structures of caspases-1, -3, -7 and -8 reveal the basis for substrate and inhibitor selectivity. *Chem. Biol.* **7**, 423–432
19. Salvesen, G. S., and Dixit, V. M. (1999) Caspase activation: The induced-proximity model. *Proc. Natl. Acad. Sci. U. S. A.* 10.1073/pnas.96.20.10964
20. Julien, O., and Wells, J. A. (2017) Caspases and their substrates. *Cell Death Differ.* **24**, 1380–1389
21. Shi, Y. (2004) Caspase activation, inhibition, and reactivation: A mechanistic view. *Protein Sci.* **13**, 1979–1987
22. Shi, Y. (2004) Caspase activation: Revisiting the induced proximity model. *Cell*. **117**, 855–858
23. Slee, E. A., Adrain, C., and Martin, S. J. (2001) Executioner Caspase-3, -6, and -7 Perform Distinct, Non-redundant Roles during the Demolition Phase of Apoptosis. *J. Biol. Chem.* **276**, 7320–7326
24. Desroches, A., and Denault, J.-B. (2019) Caspase-7 uses RNA to enhance proteolysis of poly(ADP-ribose) polymerase 1 and other RNA-binding proteins. *Proc. Natl. Acad. Sci. U. S. A.* 10.1073/pnas.1909283116
25. Ruchaud, S., Korfali, N., Villa, P., Kottke, T. J., Dingwall, C., Kaufmann, S. H., and Earnshaw, W. C. (2002) Caspase-6 gene disruption reveals a requirement for lamin A cleavage in apoptotic chromatin condensation. *EMBO J.* **21**, 1967–1977
26. Sakamaki, K., and Satou, Y. (2009) Caspases: Evolutionary aspects of their functions in vertebrates. *J. Fish Biol.* **74**, 727–753
27. Boatright, K. M., and Salvesen, G. S. (2003) Mechanisms of caspase activation. *Curr. Opin. Cell Biol.* **15**, 725–731
28. Jan, R., and Chaudhry, G. e. S. (2019) Understanding apoptosis and apoptotic pathways targeted cancer therapeutics. *Adv. Pharm. Bull.* **9**, 205–218
29. McIlwain, D. R., Berger, T., and Mak, T. W. (2013) Caspase functions in cell death and disease. *Cold Spring Harb. Perspect. Biol.* **5**, 1–28
30. Wilson, C. H., and Kumar, S. (2018) Caspases in metabolic disease and their therapeutic potential. *Cell Death Differ.* **25**, 1010–1024
31. Yuan, J., Najafov, A., and Py, B. F. (2016) Roles of Caspases in Necrotic Cell Death. *Cell*. **167**, 1693–1704
32. Chen, K., Zhao, H., Hu, Z., Wang, L. E., Zhang, W., Sturgis, E. M., and Wei, Q. (2008) CASP3 polymorphisms and risk of squamous cell carcinoma of the head and neck. *Clin. Cancer Res.* 10.1158/1078-0432.CCR-08-1198
33. Kelly, J. L., Novak, A. J., Fredericksen, Z. S., Liebow, M., Ansell, S. M., Dogan, A., Wang, A. H., Witzig, T. E., Call, T. G., Kay, N. E., Habermann, T. M., Slager, S. L., and Cerhan, J. R. (2010) Germline variation in apoptosis pathway genes and risk of non-Hodgkin's lymphoma. *Cancer Epidemiol. Biomarkers Prev.* 10.1158/1055-9965.EPI-10-0581
34. Yang, X., Merchant, M. S., Romero, M. E., Tsokos, M., Wexler, L. H., Kontny, U., Mackall, C. L., and Thiele, C. J. (2003) Induction of caspase 8 by interferon γ renders some neuroblastoma (NB) cells sensitive to tumor necrosis factor-related

apoptosis-inducing ligand (TRAIL) but reveals that a lack of membrane TR1/TR2 also contributes to TRAIL resistance in NB. *Cancer Res.*
35. Soung, Y. H., Jeong, E. G., Ahn, C. H., Kim, S. S., Song, S. Y., Yoo, N. J., and Lee, S. H. (2008) Mutational analysis of caspase 1, 4, and 5 genes in common human cancers. *Hum. Pathol.* 10.1016/j.humpath.2007.10.015
36. Lee, W. K., Kim, J. S., Kang, H. G., Cha, S. I., Kim, D. S., Hyun, D. S., Kam, S., Kim, C. H., Jung, T. H., and Park, J. Y. (2009) Polymorphisms in the Caspase7 gene and the risk of lung cancer. *Lung Cancer*. **65**, 19–24
37. Sakamaki, K., and Satou, Y. (2009) Caspases: Evolutionary aspects of their functions in vertebrates. *J. Fish Biol.* **74**, 727–753
38. Shrestha, S., Tung, J., Grinshpon, R. D., Swartz, P., Hamilton, P. T., Dimos, B., Mydlarz, L., and Clark, A. C. (2020) Caspases from scleractinian coral show unique regulatory features. *J. Biol. Chem.* **295**, 14578–14591
39. Dunn, S. R., Phillips, W. S., Spatafora, J. W., Green, D. R., and Weis, V. M. (2006) Highly conserved caspase and Bcl-2 homologues from the sea anemone Aiptasia pallida: Lower metazoans as models for the study of apoptosis evolution. *J. Mol. Evol.* **63**, 95–107
40. Gorelick-Ashkenazi, A., Weiss, R., Sapozhnikov, L., Florentin, A., Tarayrah-Ibraheim, L., Dweik, D., Yacobi-Sharon, K., and Arama, E. (2018) Caspases maintain tissue integrity by an apoptosis-independent inhibition of cell migration and invasion. *Nat. Commun.* **9**, 2806
41. Lasi, M., Pauly, B., Schmidt, N., Cikala, M., Stiening, B., Käsbauer, T., Zenner, G., Popp, T., Wagner, A., Knapp, R. T., Huber, A. H., Grunert, M., Söding, J., David, C. N., and Böttger, A. (2010) The molecular cell death machinery in the simple cnidarian Hydra includes an expanded caspase family and pro- and anti-apoptotic Bcl-2 proteins. *Cell Res.* **20**, 812–825
42. Seipp, S., Wittig, K., Stiening, B., Böttger, A., and Leitz, T. (2006) Metamorphosis of Hydractinia echinata (Cnidaria) is caspase-dependent. *Int. J. Dev. Biol.* 10.1387/ijdb.052012ss
43. Moya, A., Sakamaki, K., Mason, B. M., Huisman, L., Forêt, S., Weiss, Y., Bull, T. E., Tomii, K., Imai, K., Hayward, D. C., Ball, E. E., and Miller, D. J. (2016) Functional conservation of the apoptotic machinery from coral to man: The diverse and complex Bcl-2 and caspase repertoires of Acropora millepora. *BMC Genomics*. **17**, 62
44. Palmer, C. V. (2018) Immunity and the coral crisis. *Commun. Biol.* **1**, 91
45. Mydlarz, L. D., Fuess, L. E., Mann, W., Pinzón C, J. H., and Gochfeld, D. J. (2016) Cnidarian immunity: from genomes to phenomes. in *The Cnidaria, Past, Present and Future*, pp. 1–855, Springer International Publishing, 10.1007/978-3-319-31305-4
46. Fuess, L. E., Pinzón C, J. H., Weil, E., Grinshpon, R. D., and Mydlarz, L. D. (2017) Life or death: disease-tolerant coral species activate autophagy following immune challenge. *Proceedings. Biol. Sci.* **284**, 20170771
47. Kvitt, H., Rosenfeld, H., and Tchernov, D. (2016) The regulation of thermal stress induced apoptosis in corals reveals high similarities in gene expression and function to higher animals. *Sci. Rep.* 10.1038/srep30359
48. Dunn, S. R., Thomason, J. C., Le Tissier, M. D. A., and Bythell, J. C. (2004) Heat

stress induces different forms of cell death in sea anemones and their endosymbiotic algae depending on temperature and duration. *Cell Death Differ.* **11**, 1213–1222
49. Dunn, S. R., and Weis, V. M. (2009) Apoptosis as a post-phagocytic winnowing mechanism in a coral-dinoflagellate mutualism. *Environ. Microbiol.* **11**, 268–276
50. ANFINSEN, C. B., HABER, E., SELA, M., and WHITE, F. H. (1961) The kinetics of formation of native ribonuclease during oxidation of the reduced polypeptide chain. *Proc. Natl. Acad. Sci. U. S. A.* 10.1073/pnas.47.9.1309
51. Walters, J., Milam, S. L., and Clark, A. C. (2009) *Chapter 1 Practical Approaches to Protein Folding and Assembly. Spectroscopic Strategies in Thermodynamics and Kinetics*, 1st Ed., Elsevier Inc., 10.1016/S0076-6879(08)04201-8
52. Roche, J., Royer, C. A., and Roumestand, C. (2017) Monitoring protein folding through high pressure NMR spectroscopy. *Prog. Nucl. Magn. Reson. Spectrosc.* **102**–**103**, 15–31
53. Lim, S. A., Bolin, E. R., and Marqusee, S. (2018) Tracing a protein's folding pathway over evolutionary time using ancestral sequence reconstruction and hydrogen exchange. *Elife*. 10.7554/eLife.38369
54. Onuchic, J. N., and Wolynes, P. G. (2004) Theory of protein folding. *Curr. Opin. Struct. Biol.* **14**, 70–75
55. Gloss, L. M. (2009) Equilibrium and kinetic approaches for studying oligomeric protein folding. *Methods Enzymol.* **466**, 325–357
56. Chiti, F., and Dobson, C. M. (2006) Protein misfolding, functional amyloid, and human disease. *Annu Rev Biochem*. 10.1146/annurev.biochem.75.101304.123901
57. Chiti, F., and Dobson, C. M. (2017) Protein misfolding, amyloid formation, and human disease: A summary of progress over the last decade. *Annu. Rev. Biochem.* **86**, 27–68
58. Dill, K. A., and MacCallum, J. L. (2012) The protein-folding problem, 50 years on. *Science (80-.).* **338**, 1042–1046
59. Rumfeldt, J. A. O., Galvagnion, C., Vassall, K. A., and Meiering, E. M. (2008) Conformational stability and folding mechanisms of dimeric proteins. *Prog. Biophys. Mol. Biol.* **98**, 61–84
60. Lim, S. A., Hart, K. M., Harms, M. J., and Marqusee, S. (2016) Evolutionary trend toward kinetic stability in the folding trajectory of RNases H. *Proc. Natl. Acad. Sci.* **113**, 13045–13050
61. Hochberg, G. K. A., and Thornton, J. W. (2017) Reconstructing Ancient Proteins to Understand the Causes of Structure and Function. *Annu. Rev. Biophys.* **46**, 247–269
62. Tucker, M. B., MacKenzie, S. H., Maciag, J. J., Dirscherl Ackerman, H., Swartz, P., Yoder, J. A., Hamilton, P. T., and Clay Clark, A. (2016) Phage display and structural studies reveal plasticity in substrate specificity of caspase-3a from zebrafish. *Protein Sci.* **25**, 2076–2088
63. Hill, M. E., Macpherson, D. J., Wu, P., Julien, O., Wells, J. A., and Hardy, J. A. (2016) Reprogramming Caspase-7 Specificity by Regio-Specific Mutations and Selection Provides Alternate Solutions for Substrate Recognition. *ACS Chem. Biol.* **11**, 1603–1612

64. Mahrus, S., Trinidad, J. C., Barkan, D. T., Sali, A., Burlingame, A. L., and Wells, J. A. (2008) Global sequencing of proteolytic cleavage sites in apoptosis by specific labeling of protein N termini. *Cell*. 10.1016/j.cell.2008.08.012
65. Dix, M. M., Simon, G. M., Wang, C., Okerberg, E., Patricelli, M. P., and Cravatt, B. F. (2012) Functional interplay between caspase cleavage and phosphorylation sculpts the apoptotic proteome. *Cell*. 10.1016/j.cell.2012.05.040
66. Schilling, O., and Overall, C. M. (2008) Proteome-derived, database-searchable peptide libraries for identifying protease cleavage sites. *Nat. Biotechnol.* 10.1038/nbt1408
67. Boucher, D., Blais, V., and Denault, J.-B. (2012) Caspase-7 uses an exosite to promote poly(ADP ribose) polymerase 1 proteolysis. *Proc. Natl. Acad. Sci. U. S. A.* **109**, 5669–5674
68. MacPherson, D. J., Mills, C. L., Ondrechen, M. J., and Hardy, J. A. (2019) Tri-arginine exosite patch of caspase-6 recruits substrates for hydrolysis. *J. Biol. Chem.* **294**, 71–88
69. Grinshpon, R. D., Shrestha, S., Titus-McQuillan, J., Hamilton, P. T., Swartz, P. D., and Clark, A. C. (2019) Resurrection of ancestral effector caspases identifies novel networks for evolution of substrate specificity. *Biochem. J.* **476**, 3475–3492

Chapter 2

Caspases from scleractinian coral show unique regulatory features

Suman Shrestha[a,1], Jessica Tung[a,1], Robert D. Grinshpon[b], Paul Swartz[b], Paul T. Hamilton[c], Bradford Dimos[a], Laura Mydlarz[a] and A. Clay Clark[a,*]

[a]Department of Biology, University of Texas at Arlington, Arlington, TX 76019, USA

[b]Department of Molecular and Structural Biochemistry, North Carolina State University, Raleigh, NC 27695, USA

[c]Department of Plant and Microbial Biology, North Carolina State University, Raleigh, NC 27695, USA

[1]Contributed equally to this work

[*]Corresponding author: A. Clay Clark

E-mail: clay.clark@uta.edu

Keywords: caspase; cysteine protease, allosteric regulation, substrate specificity, coral apoptosis, coral immunity, functional divergence substrate selection, card-caspase

This chapter is published in Journal of Biological Chemistry

How to cite: Shrestha, S., Tung, J., Grinshpon, R. D., Swartz, P., Hamilton, P. T., Dimos, B., ... & Clark, A. C. (2020). Caspases from scleractinian coral show unique regulatory features. *Journal of Biological Chemistry*, *295*(43), 14578-14591.

Abstract

Coral reefs are experiencing precipitous declines around the globe with coral diseases and temperature-induced bleaching being primary drivers of these declines. Regulation of apoptotic cell death is an important component in the coral stress response. Although cnidaria contain complex apoptotic signaling pathways, similar to those in vertebrates, the mechanisms leading to cell death are largely unexplored. We identified and characterized two caspases each from *Orbicella faveolata*, a disease-sensitive reef-building coral, and *Porites astreoides*, a disease-resistant reef-building coral. The caspases are predicted homologs of human caspases-3 and -7, but OfCasp3a (*Orbicella faveolata* caspase-3a) and PaCasp7a (*Porites astreoides* caspase-7a) contain an amino-terminal caspase activation/recruitment domain (CARD) similar to human initiator/inflammatory caspases. OfCasp3b (*Orbicella faveolata* caspase-3b) and PaCasp3 (*Porites astreoides* caspase-3) have short pro-domains, like human effector caspases. We show that OfCasp3a and PaCasp7a are DxxDases, like human caspases-3 and -7, while OfCasp3b and PaCasp3 are more similar to human caspase-6, with VxxDase activity. Our biochemical analyses suggest a mechanism in coral in which the CARD-containing DxxDase is activated on death platforms, but the protease does not directly activate the VxxDase. We report the first X-ray crystal structure of a coral caspase, that of PaCasp7a determined at 1.57Å resolution. The structure reveals conservation of the caspase-hemoglobinase fold in coral and an N-terminal peptide bound near the active site that may serve as a regulatory exosite. The binding pocket has been observed in initiator caspases of other species, suggesting

mechanisms for the evolution of substrate selection while maintaining common activation mechanisms of CARD-mediated dimerization.

Introduction

Apoptotic cell death is thought to be a unique characteristic of metazoans, although its evolutionary origins are unclear. While caspases from human cells, and model organisms such as *C. elegans* and *Drosophila*, have been well studied both biochemically and structurally (1–6), little is known about caspase activity and regulation in other basal species (7). While *C. elegans* (3, 6) and *Drosophila* (8) were some of the first invertebrate caspases to be characterized, they have proven to be poor models for studying the evolution of the vertebrate apoptotic network as their networks utilize fewer caspases and regulatory proteins compared to higher eukaryotes. *C. elegans*, for example, utilizes only one effector caspase (CED-3), which also bears a CARD-motif necessary for its activation (9). Moreover, cytochrome C is not involved in the formation of the apoptosome in *Drosophila*, indicating that this organism lacks the intrinsic pathway found in humans (2). In contrast, it now appears that vertebrates have retained many characteristics of the apoptotic machinery found in sponges, sea anemone, and coral (10–12). Genomic studies of cnidarians, the sister group to the bilateria, revealed many genes that were previously thought to have been vertebrate innovations, demonstrating that the extensive gene loss in *C. elegans* and in *Drosophila* resulted in apoptotic pathways that do not reflect the characteristics of ancestral metazoans (13, 14). Basal metazoans, which appear to have a full complement of apoptotic signaling molecules, may therefore be more relevant to the evolutionary pathways of vertebrate apoptotic networks.

Cnidarians including reef-building corals from the genus Scleractinia are ecologically important organisms that are on the decline (15) and cell death has been

indicated to be important in these processes (16). The two primary drivers of coral declines are marine diseases affecting reef-building corals (17) as well as temperature induced loss of the coral's symbiont known as bleaching (15). Coral possess an innate immune system that both defends the animals against pathogenic organisms and also serves as general stress responses (18). Therefore, the coral immune system is critical in the response of these organisms to both coral diseases and bleaching. Activation of the innate immune system activates apoptotic pathways (19) however to date very few functional studies have been performed to characterize caspase structure and subsequent function in corals (20).

There are several examples pointing to the importance of apoptotic pathways and caspases in coral survival to both disease and temperature stress. An increase in expression of apoptosis-related genes was detected in a diseased Caribbean soft coral resulting in a visible inflammatory response (black-melanized appearance) (21). Also, executioner caspase genes were upregulated in the branching coral Acropora infected with white band disease (22). Several studies have gleaned important insights into coral apoptosis post-temperature stress by demonstrating that corals activate cell death responses following expulsion of their algal symbiont (19, 23–25). Specifically, the anti-apoptotic protein Bcl-2 in *Acropora millepora* is upregulated during temperature stress (26) indicating that this species likely has an intrinsic apoptosis mechanism as well as mechanisms to regulate this process. Interestingly it was shown that application of caspase inhibitors can prevent the death of bleached coral (10). Collectively, the data show the potential for complex apoptotic signaling pathways in coral, but data on

activation and control mechanisms, and how they compare to those in vertebrates, are lacking due to a dearth of biochemical characterization.

In order to gain insight into caspase activity and regulation in coral, we expressed and characterized two caspases each from two species of Caribbean reef-building corals, *Orbicella faveolata* and *Porites astreoides*. The two coral species are found on opposite ends of the stress-tolerance spectrum where the disease -susceptible *O. faveolata* activates caspase-mediated apoptotic pathways upon immune challenge, whereas the disease-tolerant *P. astreoides*, activates an adaptive autophagic response (19). These findings indicate that understanding the apoptotic machinery in corals likely has significant implication in understanding species stress tolerance. In this investigation we describe the structural composition of each species' caspase repertoire, and we use these data to functionally characterize both initiator and effector caspases from both species. Two proteins referred to as PaCasp7a and OfCasp3a based on sequences similarity to human caspases contain CARD motifs at the N-terminus, an unusual combination that has not been observed in caspases-3 or -7 enzymes from higher eukaryotes, and indeed these proteins function as initiator caspases. Additionally, two proteins PaCasp3 and OfCasp3b show canonical caspase-3/-7 structural organization, with short pro-domains and possess effector caspase function. We describe the first biochemical characterization of coral caspases and show that the PaCasp3 and OfCasp3b enzymes are not activated directly by the CARD-containing PaCasp7a and OfCasp3a, respectively. We also report the first X-ray crystal structure of a coral caspase, that of PaCasp7a determined at 1.57Å resolution, which reveals an N-terminal peptide bound near the active site that may serve as a regulatory

exosite. Overall, we find support for complex apoptotic mechanisms in these early metazoans, where the cellular machinery for both intrinsic and extrinsic apoptosis has ancient evolutionary origins.

Results

Caspases in two coral species: Phylogenetic analysis and domain organization

We examined seven caspase genes from *O. faveolata,* based on sequences obtained from previous transcriptomic and genomic data (Supplementary Fig. S1 and Supplementary Table S1) (19). The caspases were named based on the E-value from BLAST as well as the sequence similarity to the human orthologs. Results from examining the sequence homology and domain organization suggest that three of the caspases are apoptotic initiators and four are apoptotic effectors in *O. faveolata* (Fig. 1A). The sequence identities of the seven caspases compared to most human caspases are low, only ~35% (Table 1), so it is difficult to determine the nature of each coral caspase based solely on sequence comparisons with human orthologs. In addition, two caspases from *O. faveolata* contain an N-terminal CARD (caspase activation and recruitment domain) motif, similar to those in HsCasp2 and HsCasp9, and one caspase contains tandem DED (death effector domain) motifs, similar to that found in HsCasp8 (Fig. 1A). The remaining four proteins show domain organization similar to the human effector caspases, with short pro-domains (Fig. 1A).

In the case of *P. astreoides*, four caspase sequences consisted of two initiator-like caspases (called PaCasp7a and PaCasp2) and two effector-like caspases (called PaCasp7b and PaCasp3) (Fig. 1A and Supplementary Fig. S1). Similar to the results for

O. faveolata, the caspase sequences from *P. astreoides* also have only ~35% identity with human caspases, regardless of comparisons to initiator or effector caspases (Table 1). The sequences from the two coral species displayed much higher identity to putative homologs in the other coral species. For example, PaCasp7a has a 77% sequence identity with OfCasp3a, whereas PaCasp3 has 71 and 73% sequence identity, respectively, with OfCasp3b and OfCasp3c. Likewise, PaCasp2 demonstrates 76% sequence identity with OfCasp2, and PaCasp7b shares 60% identity with OfCasp7 (Fig. 1B).

A phylogenetic analysis of cnidarian and vertebrate caspases demonstrated that cnidarian caspases cluster in separate groups (Fig. 2A). All of the short pro-domain caspases, including PaCasp3 and OfCasp3b, cluster together between vertebrate effector (caspases-3/7) and initiator (caspases-8/10) caspases. Interestingly, the comparative genomics and phylogenetic analyses suggest that short cnidarian caspases, that is, those lacking a CARD or DED, share a common ancestor with vertebrate effector caspases-3 and -7 and with initiator caspases-8 and -10 (Fig. 2A). Homologs of caspase-8 in coral share the same clade with vertebrate caspases-8 and -10, and the CARD-containing OfCasp2 and PaCasp2 clustered with vertebrate caspase-2. With the exceptions of OfCasp2 and PaCasp2, the other CARD-containing coral caspases cluster with OfCasp3a and PaCasp7a and segregate into a different clade, although they share a common ancestor with vertebrate caspases-2 and -9.

We analyzed the CARD motifs of cnidarian caspases independently of the protease domains and compared them to the CARD motifs of vertebrate caspases-2 and -9 as well as that of CRADD (caspase-2 and RIPK1 domain containing adaptor with

death domain) motifs, which recruits caspase-2 to the PIDDosome (27) (Fig. 2B). The CARD motifs of coral caspases-3 and -7 cluster together but are more closely related to the CARD of caspase-2 than those of caspase-9 or CRADD. Based on this analysis, there appear to be many CARD-containing caspase-3-like proteins in cnidaria. At present, it is not clear why CARD-containing caspase-3-like proteins provide an advantage for coral development and/or symbiosis since the animals also contain initiator caspases that presumably activate the short pro-domain effector caspases. CARD-containing caspase-3-like proteins are rarely observed in vertebrate effector caspases. Fish-specific caspases have been found, such as the CARD-containing caspase-8 for example (28), but caspase-2 is, at present, the only characterized DxxDase with a CARD.

We chose two caspases from each species to characterize further, based on the sequence comparisons with human effector caspases-3, -6, or -7. In the case of *O. faveolata*, we chose two caspase-3-like proteins that showed 47% and 35% sequence identity, respectively, with HsCasp3, and we named the two proteins OfCasp3a and OfCasp3b, respectively (Fig. 1A and Table 1). Interestingly, despite predicted similarity to HsCasp3, OfCasp3a also has an N-terminal CARD motif. One caspase from *P. astreoides* demonstrated the highest sequence identity with HsCasp7 (44%) and was named PaCasp7a, even though it also contains a CARD motif (Fig. 1A and Table 1). The second protein from *P. astreoides* showed similar sequence identity to human caspases-3, -6, -7, and -8 (36-37%) (Fig. 1A and Table 1), but the protein does not have a DED motif like caspase-8 and the domain organization is more similar to that of caspase-3. Consequently, we named the protein PaCasp3. Overall, the low sequence identity

between the vertebrate and invertebrate caspases show that the classification is somewhat arbitrary without further biochemical characterizations of the proteins. Together, the phylogenetic analysis shows that the caspases from *P. astreoides* and *O. faveolata* have relatively low sequence identity (~40%) to mammalian caspases as well as other vertebrate families, but the proteins had much higher sequence identities to caspases from other cnidarian species, such as *Pocillopora damicornis, Stylophora pistillata,* and *Nematostella vectensis*.

An analysis of the coral caspase sequences shows that the proteins contain all of the conserved features that define a caspase. For example, each protein contains the catalytic dyad, histidine (CP-075) and cysteine (CP-117) (Fig. 3), where "CP" refers to the common position defined previously for caspases (29). The conserved sequence that contains the catalytic histidine (CP-115)-QACRG-(CP-119) is found in the four coral caspases, although PaCasp7a and OfCasp3a contain QACQG as in human caspase-8. One of the most highly variable regions, the intersubunit linker (IL) is the same length in OfCasp3b and PaCasp3 compared to that of HsCasp3, while those of PaCasp7a and OfCasp3a have 1 and 2 amino acids fewer than HsCasp3 respectively (Fig. 3).

Biochemical characterization of coral caspases

We examined the four coral caspases by size exclusion chromatography (SEC) since CARD-containing human caspases are monomers or mixtures of weak protomer-dimer (30). Because the IL of the procaspase monomer is cleaved during activation, the protomer is defined as a single unit that contains a large and small subunit and a single active site. Thus, the dimer consists of two protomers, or is more formally considered a

dimer of heterodimers. The data show that the CARD containing coral caspases, PaCasp7a and OfCasp3a, elute in a single peak with MW of 42.6 and 44 kDa, respectively. The sizes are larger than that of a protomer but smaller than a dimer (Supplementary Fig. S2 and Supplementary Table S5) suggesting that the proteins form weak dimers similar to the human initiator caspases. In contrast, the short pro-domain containing caspases, PaCasp3 and OfCasp3b, are dimers similar to the human effector caspases, with MW of 64.5 and 69.2 kDa, respectively (Supplementary Fig. S3 and Supplementary Table S5).

We also determined the mass of the large and small subunits by mass spectrometry. Caspase zymogens are cleaved in the IL, and the N-terminal CARD or pro-domain is removed during activation (30). The proteins also auto-process during overexpression in *E. coli*. The MW of the large and small subunits of each caspase, determined by MS, are shown in Supplementary Table S5. When compared to the sequences for each protein (Fig. 3), the data show that OfCasp3a and PaCasp7a are cleaved in the intersubunit linker after (CP-127)-DVTD-(CP-130), whereas OfCasp3b and PaCasp3 are cleaved after (CP-127)-VESD-(CP-130). The actual amino acid positions, in addition to the common position number, are shown in Fig. 1A, and the cleavage sites are indicated by the arrow in Fig. 3. In addition, the first twenty or thirty-one amino acids, respectively, in the pro-domains of OfCasp3b and PaCasp3 are removed following cleavage after VIGD (D^{20}) (OfCasp3b) or SSTD (D^{31}) (PaCasp3). The CARD motifs of OfCasp3a and of PaCasp7a are removed following cleavage after DEAD (D^{123}) and DQAD (D^{119}), respectively (Fig. 1A and Fig. 3). We note that there are potentially other

cleavage sites in the CARD motifs, but in our assays the CARD motif was completely removed.

We characterized the substrate specificity for each of the four coral caspases using substrate-phage display assays, as described previously (31). In these assays, we utilize two substrate-phage libraries that determine the P5-P1' substrate preferences, with either aspartate fixed at the P1 position (P5-xxxxDx-P1') or random (called 6x), and the results were the same for both libraries. The data show that PaCasp7a and OfCasp3a have Group II specificity, with a preference for aspartate in the P4 position (DxxDase) (Fig. 4A and Fig. 4B). In contrast, PaCasp3 and OfCasp3b prefer valine in the P4 position (VxxDase) (Figs 4C and 4D), which is defined as Group III specificity like HsCasp6.

The activities of PaCasp7a and of OfCasp3a were also examined using DEVD-AFC and VEID-AFC substrates. In all cases, however, the activity against the tetrapeptide substrates was very low due to K_M values >500 μM, so we could not reliably determine the steady-state catalytic parameters k_{cat} or K_M from the small peptide activity assays. In caspases, the K_M is thought to correlate with substrate binding (K_D), so the high K_M suggests poor binding of the small peptide.

Because of the low activity in small peptide assays, we tested the coral caspases for their ability to hydrolyze full-length (FL) human procaspases-3 and -6, which were made catalytically inactive due to mutation of the catalytic cysteine to serine (32). Thus, the proteins are incapable of undergoing self-proteolysis. As shown in Fig. 3, HsCasp3 is cleaved once in the intersubunit linker at CP-130 (IETD), while HsCasp6 contains two cleavage sites at CP-130 (DVVD) and at GP7-D17 (TEVD). Each procaspase substrate was incubated separately with an active coral caspase, and the reaction was monitored

over eight hours. Aliquots were removed and analyzed by SDS-PAGE (Fig. 5A). The results show that procaspase-3 was cleaved by PaCasp3 and by OfCasp3b, with little to no cleavage by PaCasp7a or by OfCasp3a. In contrast, procaspase-6 was cleaved by PaCasp7a and by OfCasp3a, but there was little to no cleavage by PaCasp3 or by OfCasp3b. Together, the data corroborate our results from substrate-phage display (Fig. 4) that identify PaCasp3 and OfCasp3b as VxxDases and PaCasp7a and OfCasp3a as DxxDases, respectively.

As described previously (33), we quantified the rate of hydrolysis of the two procaspase substrates by assessing the disappearance of the full-length procaspases-3 and -6, both ~32 kDa in size, and the appearance of the large (~20 kDa) and small (~10 kDa) subunits over the time course of the assay (Fig. 5B and Fig. 5C). The data were fit to a single exponential decay to approximate k_{cat}/K_M. The results show that PaCasp3 and OfCasp3b cleave procaspase-3 with hydrolysis rates of 31 $M^{-1}s^{-1}$ and 84 $M^{-1}s^{-1}$, respectively (Fig. 5B), while PaCasp7a and OfCasp3a cleaved procaspase-6 with hydrolysis rates of 159 $M^{-1}s^{-1}$ and 231 $M^{-1}s^{-1}$, respectively (Fig. 5C). We note that, although not quantified, both PaCasp3 and OfCasp3b cleave the procaspase-6 pro-peptide (TETD) at a much slower rate than that observed for cleavage of the intersubunit linker of procaspase-3 (IETD). Together, the biochemical data show that the coral caspases are weak enzymes, at least in the *in vitro* assays, with k_{cat}/K_M values ~10^2 $M^{-1}s^{-1}$.

Crystal structure of PaCasp7a

We attempted to crystalize all four of the coral caspases, and we were successful in obtaining diffraction quality crystals of PaCasp7a with an inhibitor (DEVD-CHO) bound in the active site. The crystals diffracted in the $P2_12_12_1$ space group, and we determined the structure to high resolution at 1.57 Å (Supplementary Table S4). The data show that the PaCasp7a is very similar to human caspases, with an RMSD of <1 Å compared to HsCasp3 (Fig. 6A). In the active site, the carboxylate group of the P4 aspartate hydrogen bonds to Asn^{315} (CP-162) on active site loop 3 (L3), the backbone amide of Arg^{356} (GP9-02) on L4, and through-water hydrogen bonds to Trp^{321} (CP-168) (on L3) as well as the backbone carbonyl of Arg^{356} (GP9-02) (on L4) (Fig. 6B). In general, the active site provides hydrophilic binding pockets for the P3 glutamate and P4 aspartate of the substrate, and a more hydrophobic binding pocket for the P2 valine side-chain (Fig. 6C), similarly to that of HsCasp3.

Both PaCasp7a and OfCasp3a contain a two-residue insertion in loop 1 (L1) of the active site (Fig. 3). The structure of PaCasp7a with inhibitor bound shows that the insertion extends the loop compared to HsCasp3 and results in an “RYP” motif in L1 (Fig. 3) near the catalytic histidine (Fig. 6D). Models of the PaCasp7a active site suggest that rotation of the loop results in intercalation of the Arg^{171} (GP2-17) between the catalytic His^{231} (CP-075) and Cys^{271} (CP-117) (Fig. 6E). In this orientation, the arginine side-chain hydrogen bonds with the carbonyl of Gly^{232} (CP-076) on active site loop 3 (L3) and clashes with the P1 and P2 positions of substrate. In addition, the tyrosine of the RYP motif forms a new hydrogen bond with the side-chain of Ser^{167} (GP2-05) in L1 (Fig. 6E). Altogether, the models suggest that in the absence of

substrate, rotation in L1 may stabilize an inactive conformation of the enzyme. We note, however, that MD simulations (50 ns) of the structural models show that the region of L1 that contains the RYP motif is very mobile, so if the RYP motif is indeed autoinhibitory, then the “RYP-In” conformation appears to be transient (Supplementary Fig. S3).

The structure of PaCasp7a also reveals a peptide bound on the protein surface near α-helices 1 and 4. The structure shows that amino acids in the N-terminus of PaCasp7a (N’-AKLFSFGG-C’) (N’-PD-A025 to PD-G018-C’ in the common position numbering) comprise the peptide, where the two phenylalanine side-chains bind in a hydrophobic pocket between the helices 1 and 4 (Fig. 6F). The binding pocket on the protein is formed by five hydrophobic residues on the two helices (L^{187} (CP-031), A^{190} (CP-034), L^{191} (CP-035), F^{330} (CP-177), A^{334} (CP-181)) as well as F^{381} and F^{382} (CP-217 and CP-218) at the C-terminus (Fig. 6F). The peptide also forms several hydrogen bonds with charged groups on the protein surface. We do not observe electron density for amino acids G^{128} (PD-017) - N^{141} (PD-004) (Fig. 3), but extensive interactions downstream of D^{142} (PD-003) result in an ordered structure that moves into the core of the protein. The fourteen disordered residues would provide ample distance to connect the peptide with the protease domain, and the data suggest that the intervening amino acids may hinder dimerization since they would be anchored near the dimer interface when the peptide is bound on the protein surface. The N-terminal end of the peptide is immediately downstream of the DQAD cleavage site that removes the CARD motif (Fig. 3), suggesting that the binding pocket on the protein surface may be used to position the N-terminal linker (between the CARD and protease domains) in the active site.

We searched the caspase structures in the protein data bank and found four examples of an N-terminal peptide bound between helices 1 and 4 – human caspases-1 (PDB ID 2H48) and -2 (PDB ID 3R7M), Dronc of *Drosophila melanogaster* (PDB ID 2FP3) and CED-3 of *Caenorhabditis elegans* (PDB ID 4M9R) (34–37) (Supplementary Figs. S4A – S4D). Interestingly, the region of the peptide that is disordered in PaCasp7a (G^{128} PD-017 - N^{141} PD-004) forms a short α-helix in caspases-1 and -2 and in CED-3 (Supplementary Figs. S4A – S4B and S4D). The short helix does not make contacts across the dimer interface but rather makes extensive intra-protomer contacts with the C-terminus of the protein. In the case of DRONC, the intervening peptide forms an extended structure that extends beyond the dimer interface and would clash with the second protomer of the dimer (Supplementary Fig. S4C). In all cases, the binding pocket between helices 1 and 4 is hydrophobic, and the peptide binds through insertion of one or more hydrophobic amino acids into the binding pocket as well as hydrogen bonds between side-chains on the protein surface and backbone atoms of the peptide. Therefore, the structures show a common theme in which the N-terminal peptide downstream of the pro-domain cleavage site binds to a hydrophobic pocket on the protein surface. The interactions likely stabilize the peptide in the binding pocket for cleavage.

Finally, we observed a similar hydrophobic pocket in human effector caspases (HsCasp3, PDB: 2J30; HsCasp6, PDB: 3S70; HsCasp7, PDB: 1F1J) (38–40) (Supplementary Figs. S4E – S4G). There is no evidence, however, from biochemical or structural data, that the N-terminal peptide of the short pro-domain caspases bind in the hydrophobic pocket. A comparison of the N-terminal sequences (Fig. 3) shows

significant divergence in the peptide of human effector caspases, so although the binding pocket is similar to that of PaCasp7a, the binding interactions with the peptide sequences are not similar. In HsCasp3 and HsCasp6, for example, the cleavage site is downstream of the putative binding sequence, so the entire peptide is removed from the N-terminus. Interestingly, in HsCasp7 the cleavage site is upstream of the binding region, but the sequence evolved into a tetra-lysine motif that has been shown to be an exosite for substrate selection in caspase-7 (41).

Discussion

Coral reefs are facing a significant decline due to increasing local and global stressors from disease, climate change, and pollution (42). While coral possess a robust innate immune system, the cellular mechanisms these organisms use to respond to disease have not been functionally characterized (43). In addition, elevated ocean temperatures have emerged as key threats to the long-term survival of coral reefs due to collapse of the coral-algal symbiosis (23, 42). Given the environmental consequences posed by coral disease and bleaching, understanding the molecular physiology behind coral immune responses will improve our understanding of coral declines (44). The two coral species used in this study *Orbicella faveolata* and *Porites astreoides* are both reef-building coral, but they lie on opposite ends of the disease response spectrum. Where *O. faveolata* is sensitive to disease and activates apoptotic responses to stress, *P. astreoides* is resistant to disease and activates autophagic responses to stress. Thus, these two species of coral represent an intriguing system to understand and characterize apoptotic mechanisms (21).

The phylogenetic analysis suggests that the caspase repertoire of reef-building corals cluster into two categories: coral effectors and coral initiators (Fig. 2). This indicates that in contrast to model systems such as *C. elegans* and *Drosophila,* corals likely possess caspases with discrete functions that mirror vertebrate caspases. Protein domain predictions for the caspase repertoire from our two focal species determined that the sequences in the initiator caspase family possess N-terminal CARD or DED domains similar to human caspases, while the predicted coral effector caspases contain short pro-domains, also a pattern reflected in human caspases. Overall, this comparative analysis demonstrates that it is likely that corals possess complex apoptosis networks similar to vertebrates.

Our functional characterization suggests that PaCasp7a and OfCasp3a may function similarly to caspase-2 since they exhibit DxxDase activity and contain an N-terminal CARD motif. In contrast, PaCasp3 and OfCasp3b share characteristics with effector caspase-6, with a short prodomain and VxxDase activity. Moreover, a phylogenetic analysis showed that OfCasp3a and PaCasp7a are close to vertebrate initiator caspases, whereas PaCasp3 and OfCasp3b are closer to effector caspases. Although the caspases exhibited low activity against peptide substrates, we were able to confirm the selection through cleavage assays of protein substrates. The results showed that the DxxDases (PaCasp7a and OfCasp3a) processed procaspase-6, which has a DVVD cleavage sequence in the intersubunit linker, but not procaspase-3, which contains a more hydrophobic recognition sequence recognized by caspase-8 (IETD). The opposite was true for PaCasp3 and OfCasp3b. In those cases, the enzymes processed procaspase-3 but not procaspase-6.

The biochemical data show that the two short prodomain caspases (OfCasp3b and PaCasp3) are most likely not the main executioner caspases in coral. Rather, the enzymes may function similarly to HsCasp6 which plays a role during cell development. In addition, the two effector caspases are not directly activated by the CARD-containing caspases, PaCasp7a or OfCasp3a. At present, it is not clear if the DxxDase activity of the two effector caspases functions as the primary executioner of apoptosis or if the proteins activate the as-yet-unidentified executioner caspase. Based on the caspases identified in coral, we suggest a model in which the PaCasp7a and OfCasp3a enzymes are activated on PIDDosome-like complexes (45, 46), similar to HsCasp2 (Fig. 7). Either the DxxDase activity is utilized to kill cells, like CED-3 in *C. elegans*, or the activated enzymes cleave the executioner. In the latter case, the PaCasp3 and OfCasp3b proteins would be indirectly activated by PaCasp7a and OfCasp3a, respectively, through the undefined executioner caspase. Alternatively, the OfCasp3a and PaCasp7a may be activated on apoptosome complexes, and the DxxDase activities could be used to activate downstream caspases or to execute apoptosis. The latter suggestion is consistent with the presence of coral caspase-2-like proteins that also contain CARD motifs (Fig. 1A). Caspase-8-like proteins containing DED motifs would also activate the executioner (and PaCasp3 or OfCasp3b) indirectly. The putative caspase-8 and caspase-3-executioner proteins have been identified, but not yet characterized, in coral (47–50). Further characterization of the apoptotic components will determine the signaling responses leading to activation of the caspases, in particular whether OfCasp3a and PaCasp7a are activated on PIDDosome or apoptosome complexes. Together, the data suggest that coral are responsive to death ligands as well as

metabolic changes in the cell, matching human extrinsic and intrinsic apoptotic pathways and establishing their importance in the context of coral health, disease and bleaching.

Finally, the data shown here for PaCasp7a, as well as previous structural data – human caspases-1 (PDB ID 2H48) and -2 (PDB ID 3R7M), Dronc of *Drosophila melanogaster* (PDB ID 2FP3) and CED-3 of *Caenorhabditis elegans* (PDB ID 4M9R), and human effector caspases (HsCasp3, PDB: 2J30; HsCasp6, PDB: 3S70; HsCasp7, PDB: 1F1J) (34–40) (Supplementary Figs. S4A – S4G) – identify a hydrophobic pocket on the protein surface between helices 1 and 4 in which a peptide sequence C-terminal to the processing sequence binds. The binding of the peptide may increase activity through improved binding of the recognition sequence in the active site and help position the linker near the CARD-motif for cleavage. Alternatively, the binding of the peptide to the pocket may affect substrate selection by demonstrating a preference for substrates with both a P1-P4 cleavage sequence and the downstream sequence that binds to the pocket. We also showed that the hydrophobic pocket is conserved in a wide range of species, with similar size and properties. The short pro-domain caspases appear to have retained the binding pocket on the protease domain, but the N-terminal peptide sequence diverged, suggesting that effector caspases may utilize the binding pocket as an exosite for substrate selection. In this case, for example, substrates with sequences that bind in the pocket, and are downstream of the cleavage site, may exhibit better binding compared to substrates that contain only the P1-P4 recognition sequences.

Conclusions

Coral have complex apoptotic signaling cascades, similar to those of vertebrates providing another line of evidence that apoptotic mechanisms are ancient and well conserved in metazoans. We have identified OfCasp3a and PaCasp7a as initiator caspases that appear to function similarly to HsCasp2, indicating that corals are responsive to metabolic changes in the cell. In addition, both *O. faveolata* and *P. astreoides* contain VxxDases similar to HsCasp6. Our data show that the enzymes from both species have similar biochemical properties and are activated by similar mechanisms. Given that coral caspases have relatively low sequence identity to human caspases, the designation of caspase function based on sequence similarity should undergo further biochemical characterization. Together, the data show that regulatory or physiological mechanisms rather than differences in the caspase cascade likely dictate the disease sensitivity of *O. faveolata* or disease resistance of *P. astreoides.* Potential differences could arise from response mechanisms in the death receptor or the PIDDosome activation platforms upstream of the caspase cascade. Our data suggest that differences in the receptor-mediated activation of caspases as well as cross-talk between the autophagic and apoptotic pathways in the two coral species lead to the different physiological responses.

Experimental procedures

Cloning, protein expression and protein purification

The codon optimized sequences of the four coral caspases, PaCasp3, PaCasp7a, OfCasp3a and OfCasp3b, were based on the sequences from previous

transcriptomic data (19) and were cloned into pET11a vector (Genescript, USA). All proteins contained a C-terminal His_6 tag and were expressed in *E. coli* BL21(DE3) pLysS cells and purified as previously described (32, 51).

Phylogenetic analysis

Caspase sequences of representative species were obtained from the CaspBase (caspbase.org) (29) along with BLAST top hits from HMMER (52), and multiple sequence alignments were obtained using MEGA 7 (53). The best model of evolution to construct a phylogenetic tree from our dataset was determined with ProtTest 3 (54) (https://github.com/ddarriba/prottest3), and the tree was computed with the maximum likelihood method in IQTREE, using the Jones-Taylor Thornton model (JTT) plus gamma distribution (55). The tree was bootstrapped 1000 times as a test of phylogeny. The accession numbers of all genes used for phylogenetic analysis are listed in Supplementary Tables S1 and S2.

Size exclusion chromatography

Proteins were examined using a Superdex75 Increase 10/300GL column on an AKTA-FPLC. The proteins were concentrated to 1-5 mg/mL and dialyzed in a buffer of 30 mM potassium phosphate, pH 7.5, containing 1 mM DTT for 4 hours. The column was equilibrated with two columns volume (50 mL) of the same buffer. Protein (200 μL) was loaded onto the column, and the column was resolved at a flow rate of 0.5 mL/min. The column was calibrated using the gel filtration LMW calibration kit (GE Health Sciences) following the manufacturer instructions.

Mass spectrometry

Matrix-assisted laser desorption/ionization (MALDI) analysis was done as described (56). In brief, proteins were resolved by SDS-PAGE on a 12.5% acrylamide gel, and then bands for the large and small subunits were excised. Each gel fragment was destained using a solution of acetonitrile and 50 mM ammonium bicarbonate (1:1 v/v) for 3 hrs. The gel fragments were then crushed in microcentrifuge tubes, and the proteins were extracted with 30 μL of a solution of formic acid/water/2-propanol (1:3:2 v/v/v) (FWI) for 8 hours at room temperature. After extraction, samples were centrifuged and supernatant was lyophilized then re-dissolved in 2 μL of MALDI matrix solution (FWI saturated with 4-hydroxy-α-cyano-cinnamic acid (4HCCA)). Dissolved protein was then retrieved for MS analysis using dried-drop method of matrix crystallization then analyzed by MALDI-MS (Axima Assurance Linear MALDI TOF).

Whole-protein cleavage assay

Enzyme specificity of the four coral caspases was first examined by cleavage of human procaspases-3 and -6 in time-course assays, as described previously (33). The procaspase substrate was diluted to a final concentration of 5 μM in a buffer of 150 mM Tris-HCl, pH 7.5, 50 mM NaCl, 1% sucrose, and 10 mM DTT at 37 °C. Reactions were started by the addition of respective coral caspase at a final concentration of 1 μM, and the total reaction volume was 2 mL. Aliquots of 100 μL were removed at times 30 sec, 1 min, 5 min, 15 min, 30 min, 45 min, 1 hour, 2 hour, 4 hour, 6 hour and 8 hour after the addition of active enzyme. Reactions were stopped by the addition of six-fold concentrated SDS-PAGE loading dye (20 μL) followed by incubation in boiling water for

5 minutes. Samples were loaded into a 16% resolving gel with a 4% stacking gel and electrophoresed for 1.5 hours at 80 volts followed by an increase in voltage to 190 V for an additional 4 hours. The change in density for the procaspase substrate over time as a result of cleavage was quantified using Image lab software (Bio-Rad), and the data were plotted with Kaleidagraph. As described previously (33), the data were fit to an exponential decay to determine the CF50 (cleavage of 50% of protein substrate), and the CF50 was used to calculate the rate of hydrolysis (M^{-1} sec^{-1}) using the equation $k = ((-\ln(P))/(E \cdot t))$. In this case, k is the rate of hydrolysis, P is the fraction cleaved (50%), E is the concentration at which CF50 is achieved (in molar), and t represents time (in seconds).

Enzyme assays and substrate-phage display

Enzyme activity was determined in a buffer of 150 mM Tris-HCl, pH 7.5, 50 mM NaCl, 10 mM DTT, 1% sucrose, 0.1% CHAPS (assay buffer) at 25 °C, as previously described (57, 58). The total reaction volume was 200 µL, and the final enzyme concentration was 10 nM. Following the addition of substrate (Ac-DEVD-AFC, Ac-VEID-AFC, Ac-LETD-AFC, Ac-LEHD-AMC, Ac-IETD-AMC), the samples were excited at 400 nm (AFC substrates) or 350 nm (AMC substrates), and fluorescence emission was monitored at 505 nm (for AFC substrates) or 450 nm (for AMC substrates) for 60 seconds using a PTI fluorometer (Photon Technology International, Edison, NJ, USA). The steady-state parameters, K_M and k_{cat}, were determined from plots of initial velocity *versus* substrate concentration.

Substrate phage display assays were performed as described (31, 58). Briefly, phage libraries consisting of caspase recognition sequences were bound to Ni-NTA resin. Enzyme (10-100 nM) was added to initiate the reaction, and samples were incubated between 3 and 20 hours. *E. coli* ER2738 cells were used to amplify the cleaved phage from previous rounds by infecting cells with the supernatant after enzyme incubation. The cells were grown for 4 hours, removed by centrifugation, and the supernatant was collected and used as the library for the following round of selection. Plaque counting was used to determine the endpoint of the experiment, when the number of phage bound to the resin was similar to the number of phage released during the treatment. The number of phage released during the reaction *versus* the control (without enzyme) was monitored to ensure progress in substrate selectivity.

X-ray crystallography

Protein structure predictions were performed using Swiss-Model (59) using human caspases-3, -6, and -7 as references (PDB ID 2J30, 3OD5, and 1F1J, respectively). For structure determination, the coral caspase proteins were dialyzed in a buffer of 10 mM Tris-HCl, pH 7.9, 100 mM NaCl, 1 mM DTT and concentrated to ~7 mg/mL. The molar extinction coefficients for the proteins were determined by ProtParam under reduced conditions (Supplementary Table S3). Inhibitor, Ac-DEVD-CHO (reconstituted in DMSO), was added at a 5:1 (w/w) inhibitor/protein ratio, and DTT and NaN_3 were added to final concentrations of 10 and 3 mM, respectively. Samples were incubated for 1 hour in the dark on ice. Hanging-drop vapor diffusion method was applied using 4 µL drops that contained equal volumes of protein and reservoir solutions

using the PEG/ion 2 screen (Hampton Research). PaCasp7a protein crystalized in a solution of 0.1 M sodium malonate pH 5.0, 12% w/v polyethylene glycol (PEG) 3350, and conditions were optimized such that the best diffracting crystals of PaCasp7a were obtained at 18 °C in a solution of 0.1 M sodium malonate, pH 4.9–5.1, 15–17% PEG 3350 (w/v), 10 mm DTT, and 3 mm NaN_3. Crystals for PaCasp7a appeared within 3 to 5 days and were briefly immersed in a cryogenic solution containing 20% PEG 4000, 80% reservoir solution. Crystals were stored in liquid nitrogen. We were unable to obtain diffraction quality crystals for the remaining coral caspases. Data sets were collected at 100 K at the SER-CAT synchrotron beamline (Advance Photon Source, Argonne National Laboratory, Argonne, IL). Each data set contained 180 frames at 1° rotation. The protein crystallized in the space group P 2_1 2_1 2_1 and was phased with a previously published HsCasp3 structure (PDB entry 2J30). Data reduction and model refinements were done using HKL2000, COOT, and Phenix, and a summary of the data collection and refinement statistics is shown in Supplementary Table S4. Molecular dynamics simulations were performed for 50 ns with GROMACS 4.5 (60) using the Amber99 force field (61) and the TIP3P water model (62), as previously described (63).

Data availability

The crystal structure for PaCasp7a has been deposited in the Protein Data Bank, www.wwpdb.org under PDB ID code: 6WI4.

Funding and additional information

This work was supported by a grant from the National Institutes of Health [grant number GM127654 (to A.C.C)] and by funds from UT Arlington [Office of the Vice President for Research (to A.C.C) and New Directions Award from the College of Science to A.C.C and L.D.M]. Use of the Advanced Photon Source was supported by the U.S. Department of Energy, Office of Science, Office of Basic Energy Sciences, under contract number W-31-109-ENG-38.

Conflict of interest

The authors declare that they have no conflicts of interest with the contents of this article.

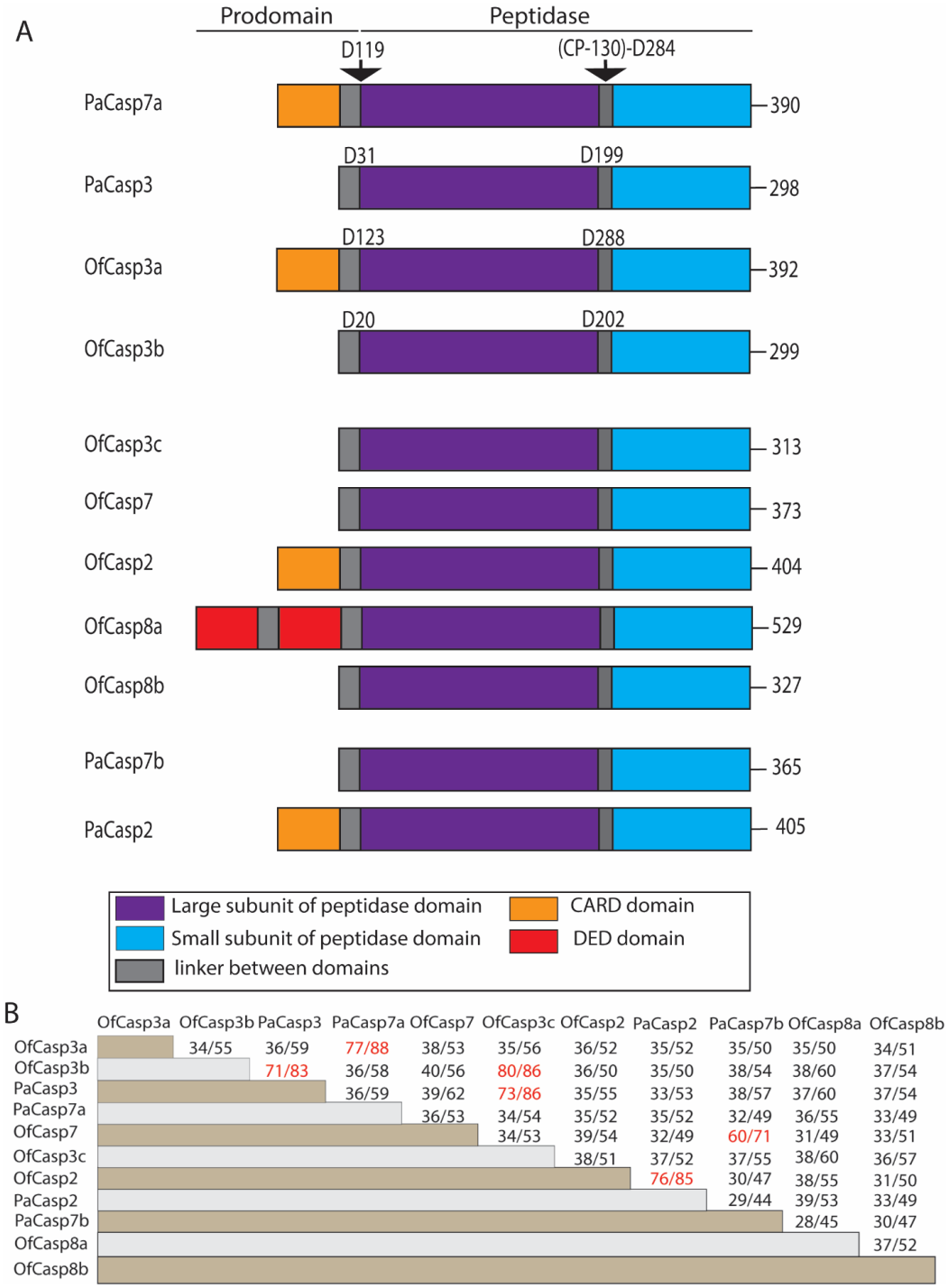


Figure 1: Domain organization and sequence comparison among caspases of *O. faveolata* and *P. astreoides*. (A) Domain organization of caspases in *O. faveolata* and *P. astreoides*. Processing site between large and small subunit, and after prodomain are noted in biochemically characterized caspase. (B) Protein sequence identity (%) and similarity (%) among coral caspases.

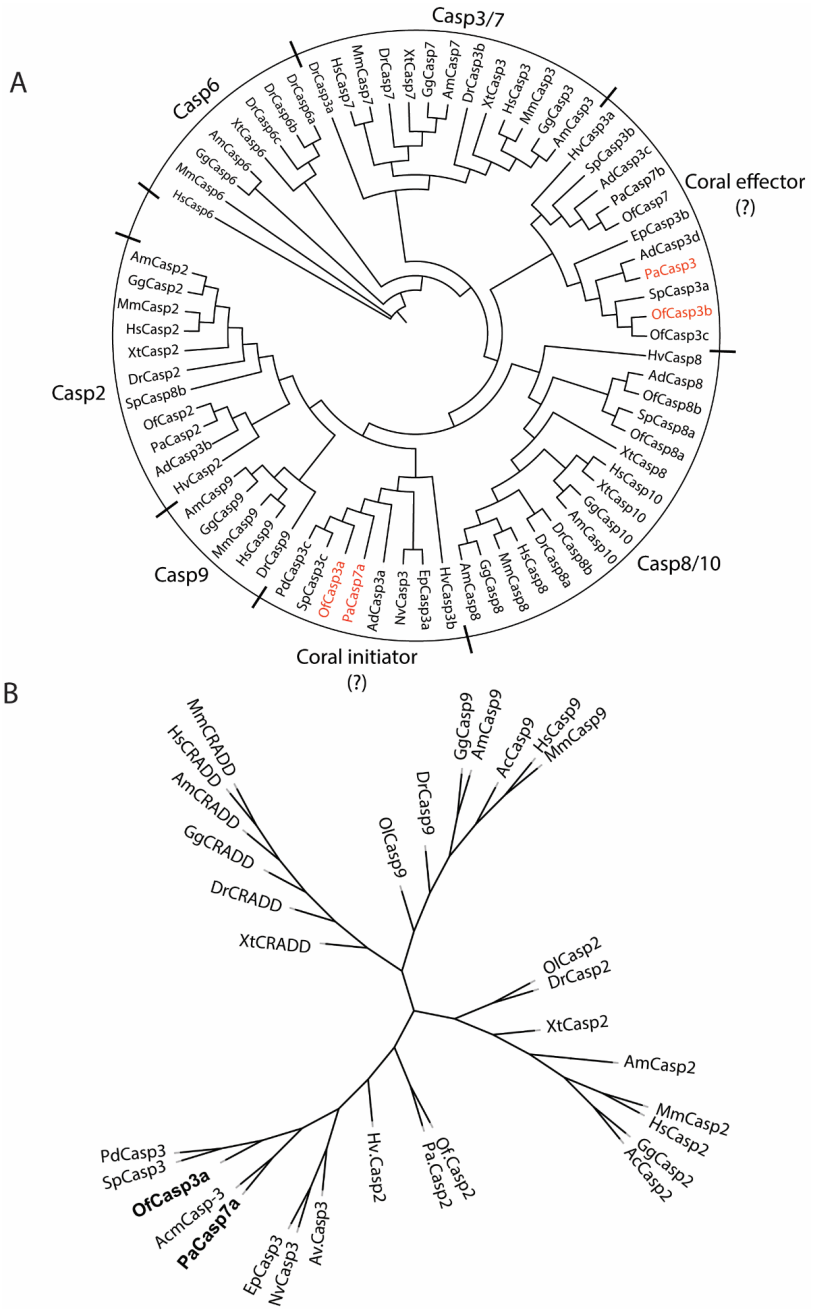


Figure 2: Coral caspase phylogenetic analysis. (A) Phylogenetic tree of cnidarian and vertebrate caspases. *Orbicella faveolata* (Of), *Porites astreoides* (Pa), *Pocillophora damicornis* (Pd), *Stylophora pistillata* (Sp), *Nematostella vectensis* (Nv), *Exaiptasia pallida* (Ep), *Hydra vulgaris* (Hv), *Acropora digitophora* (Ad), *Homo sapiens* (Hs), *Mus musculus* (Mm), *Gallus gallus* (Gg), *Alligator mississippiensis* (Am), *Xenopus laevis* (Xl), *Danio rerio* (Dr). Accession number of all used sequences are shown in Supplementary Tables S1 and S2. (B) Phylogenetic analysis of CARD domains of caspases and CRADDs (CASP2 and RIPK1 Domain containing Adaptor with Death Domain) between cnidarians and vertebrates.

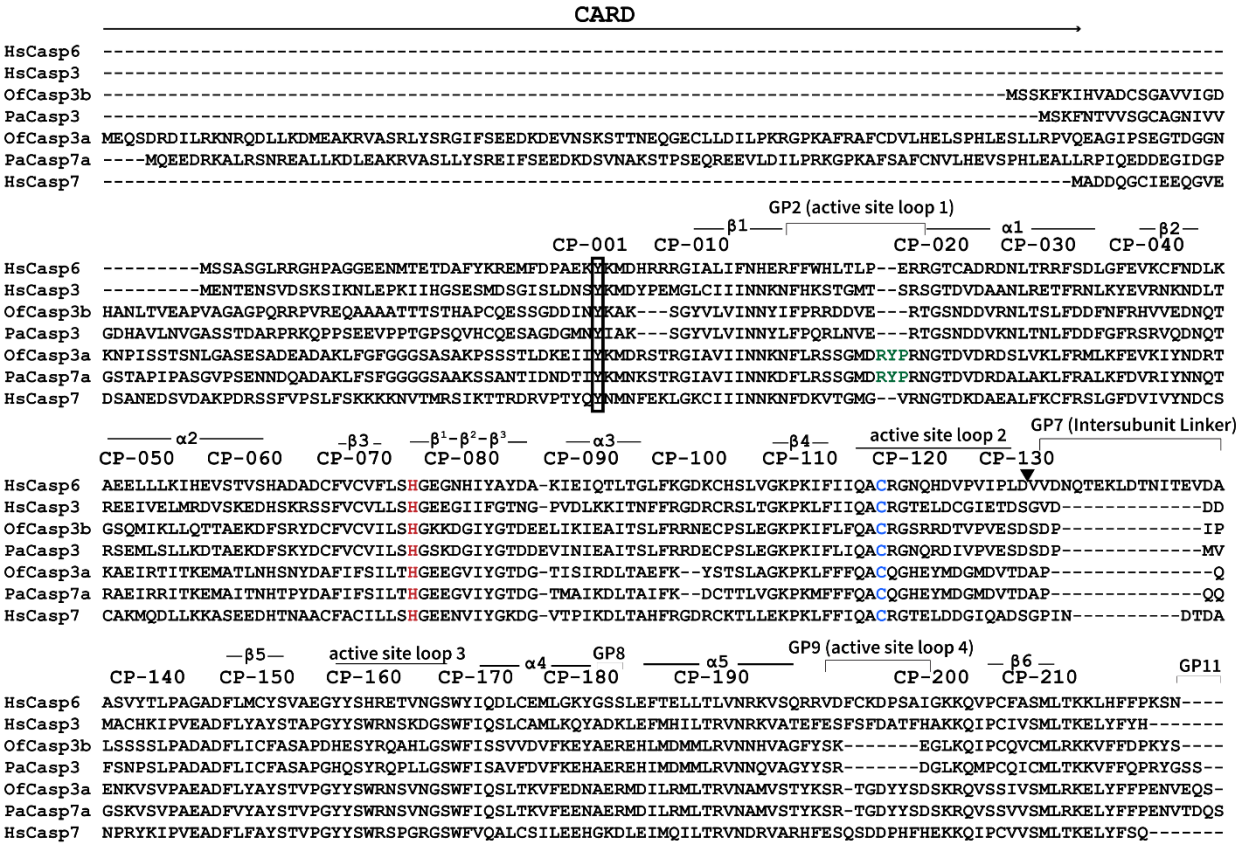


Figure 3: Biochemically characterized caspases of *O. faveolata* and *P. astreoides* aligned with human effector caspases. In the multiple sequence alignment (MSA), secondary structures (alpha helices, beta sheets, and loops) are indicated along with common position (CP) numbers among caspases. Gap positions, or sequences between common amino acid positions, are referred to as GP. Histidine (H) and Cysteine (C), which forms a catalytic dyad, are colored in red and blue respectively. “RYP” motif insertion in OfCasp3a and PaCasp7a are colored in green.

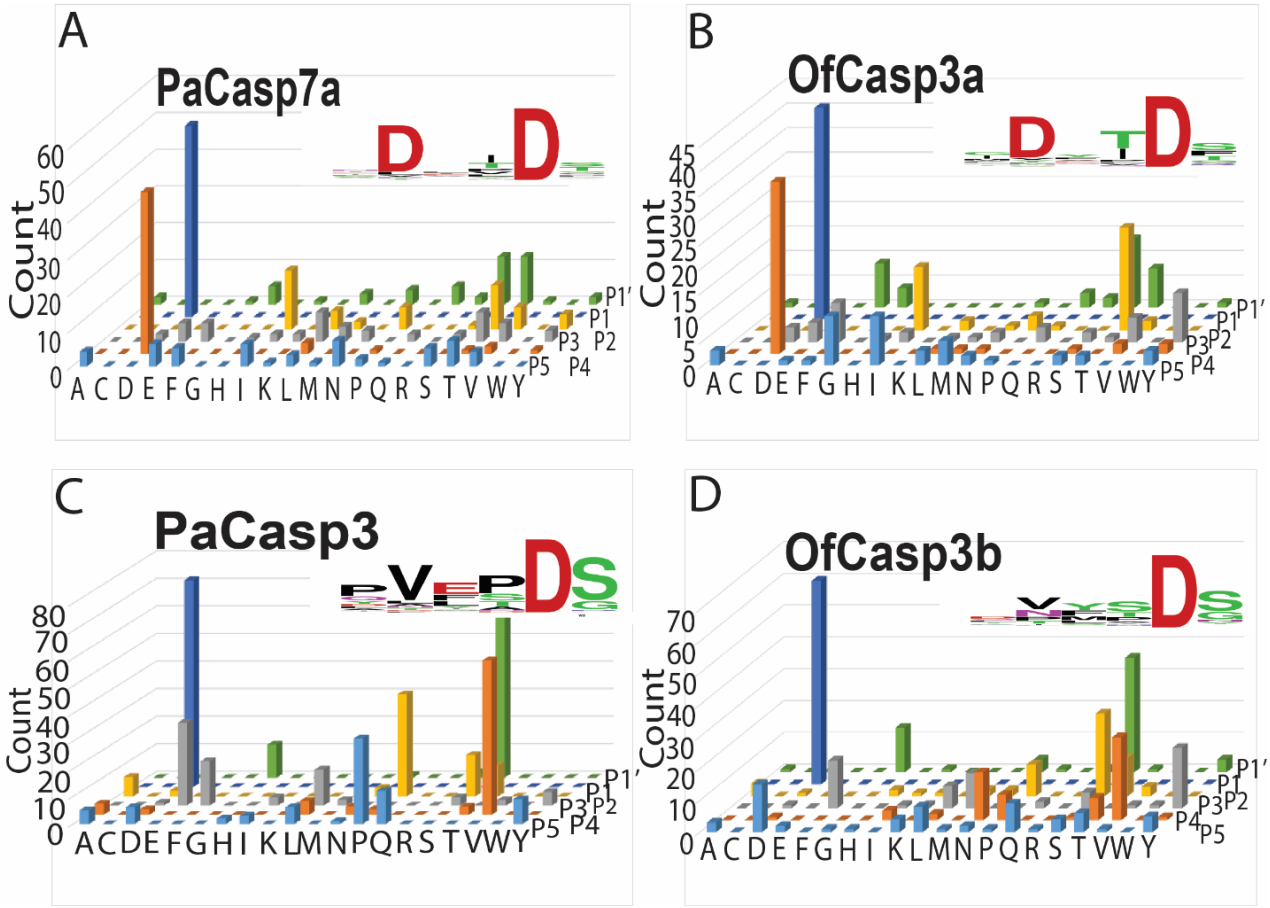


Figure 4: Substrate preference determined by substrate-phage display. Amino acid preferences shown for substrate positions P5-P4-P3-P2-P1-P1' for PaCasp7a (A), OfCasp3a (B), PaCasp3 (C) and OfCasp3b (D). Values of Y-axes indicate number of phage sequences containing the specified amino acid (Count). Amino acids are shown on the X-axes in single letter code. Web logos are also shown in inset of respective graph for same results.

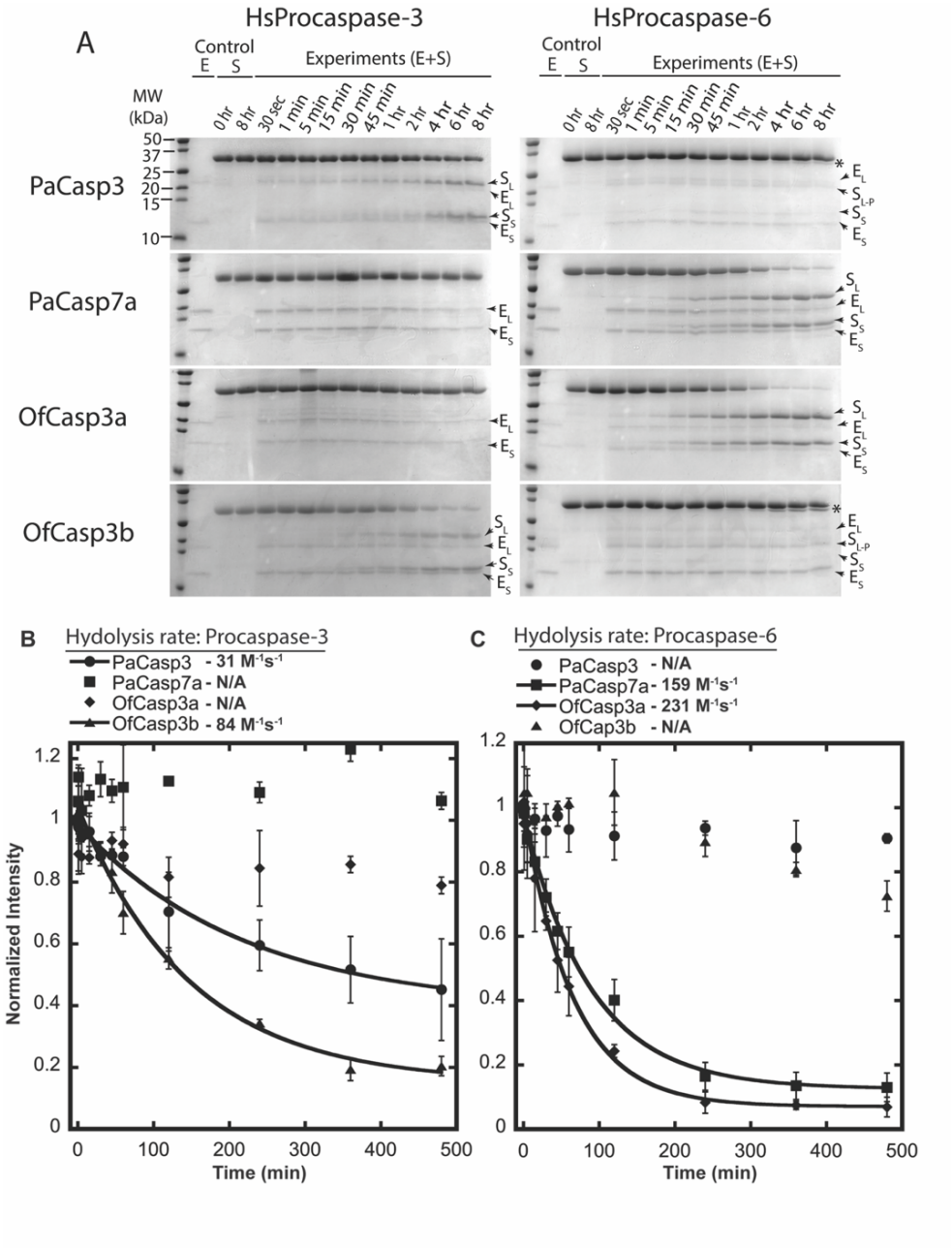


Figure 5: Cleavage kinetics of coral caspases using human procaspases-3 and -6 as a substrate. (A) Cleavage of full-length inactive HsCasp3 and HsCasp6 by coral caspases over time course. All cleaved products are labeled along with enzyme itself. (S_L- large subunit of substrate, S_S- small subunit of substrate, E_L- large subunit of enzyme, E_S- small subunit of enzyme, $S_{L\text{-}P}$- large subunit with prodomain cleaved, S- substrate and E+S- enzyme and substrate). Bands with "*" indicate only prodomains were removed from full-length substrate. (B, C) Quantification of procaspase bands relative to the control (substrate without enzyme after 8 hours incubation). Data were fit to a single exponential decay to calculate CF_{50} used to calculate hydrolysis rate of coral caspases (solid line). Procaspase-3 (B), Procaspase-6 (C). Error bars represent standard deviation from three different experiments.

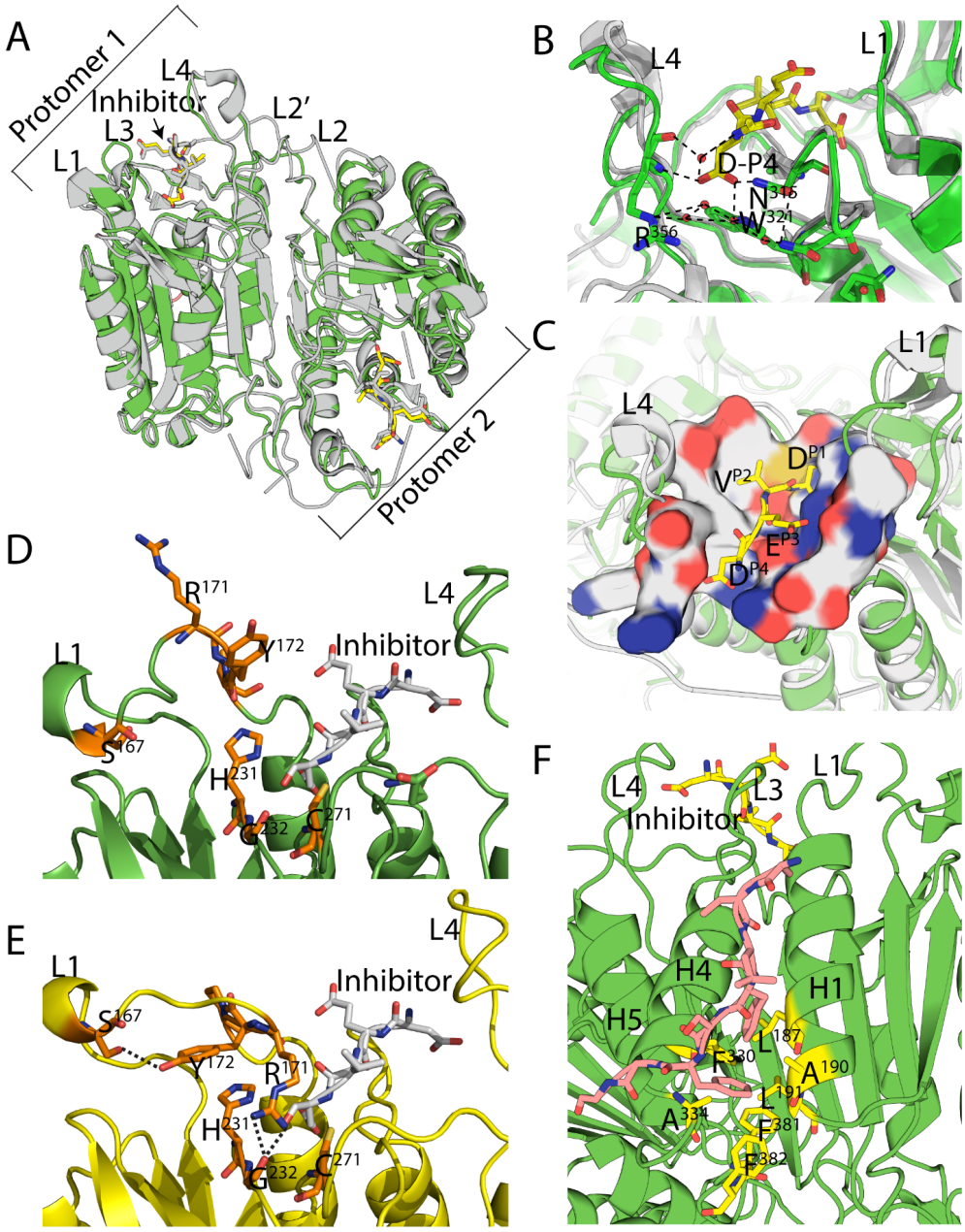


Figure 6. Structure of PaCasp7a. (A) Comparison of PaCasp7a (green) aligned with HsCasp3 (grey) (PDB ID: 2J30). (B) PaCasp7a active site bound with inhibitor DEVD-CHO. Dashed lines show hydrogen bonding network to the P4 aspartate. (C) Surface map of active site residues in PaCasp7 within 5 Å of the inhibitor (yellow sticks). Neutral charges are grey, negative charges are red, and positive charges are blue. (D) **"**Out" orientation, of "RYP" residues in loop 1 in crystal structure of PaCasp7a. (E) "In" orientation, of "RYP" residues in loop 1 in predicted model of PaCasp7a. Dashed lines show the hydrogen bonds formed by "R" and "Y" in "In" Orientation. (F) N-terminal peptide (orange) bound in hydrophobic pocket between helices 1 and 4. PaCasp7a residues that form the pocket are shown in yellow: L^{187} (CP-031), A^{190} (CP-034), L^{191} (CP-035), F^{330} (CP-177), A^{334} (CP-181) as well as F^{381} (CP-217) and F^{382} (CP-218) at the C-terminus.

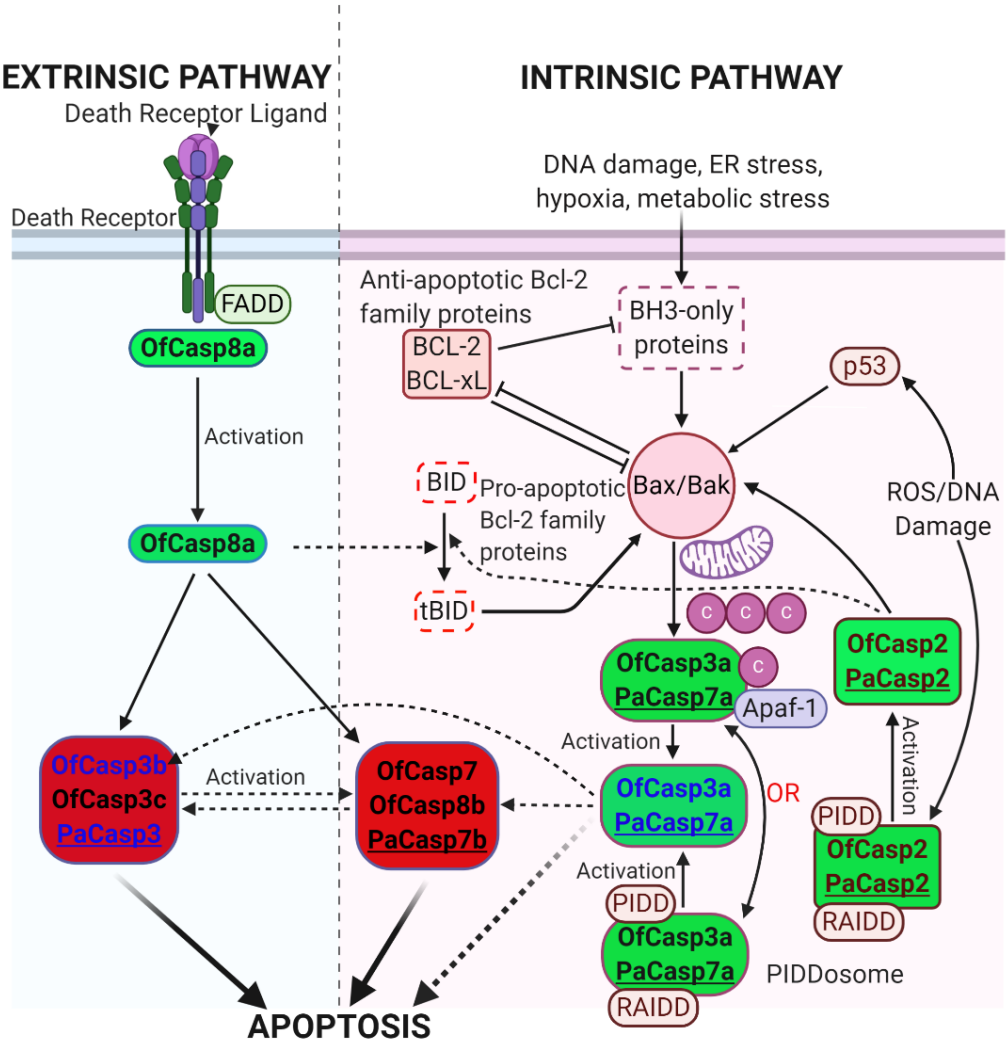


Figure 7. Proposed apoptotic pathways in coral compared to the apoptotic pathways in humans. All components in the pathways have homologs in *O. faveolata* and *P. astreoides* with the exception of BID (dotted box). A list of homologs is shown in Supplementary Table S6. Dotted lines indicate that links have not yet been shown experimentally. Caspases in green background are initiators and those in red background are effectors. Pa refers to *P. astreoides* and Of refers to *O. faveolata.* The four caspases characterized here are shown in blue.

Table 1: Protein sequence identity/similarity (%) with human caspases.

	HsCasp 3	HsCasp 7	HsCasp 6	HsCasp 2	HsCasp 8	HsCasp1 0	HsCasp 9
OfCasp7	37/54	38/52	32/49	28/43	34/50	33/53	34/50
OfCasp3c	36/58	35/56	35/52	32/48	37/53	37/54	33/49
OfCasp3b	35/60	32/57	33/52	32/49	37/55	33/53	34/50
OfCasp3a	47/69	45/65	38/54	29/48	39/54	39/55	28/46
OfCasp2	37/53	41/55	35/46	33/52	34/52	35/52	32/48
OfCasp8a	39/56	39/53	34/48	35/53	32/58	30/49	34/51
OfCasp8b	33/56	31/50	31/49	32/52	35/51	34/52	32/46
PaCasp3	36/58	37/59	36/56	33/49	37/54	34/54	35/50
PaCasp7a	43/65	44/60	36/53	28/46	37/53	37/53	28/44
PaCasp7b	38/54	37/52	34/49	29/46	31/48	30/50	33/48
PaCasp2	39/53	38/52	33/48	33/50	35/53	35/52	32/49

REFERENCES

1. Song, Z., McCall, K., and Steller, H. (1997) DCP-1, a Drosophila cell death protease essential for development. *Science (80-.).* **275**, 536–540
2. Dorstyn, L., Mills, K., Lazebnik, Y., and Kumar, S. (2004) The two cytochrome c species, DC3 and DC4, are not required for caspase activation and apoptosis in Drosophila cells. *J. Cell Biol.* **167**, 405–410
3. Ellis, H. M., and Horvitz, H. R. (1986) Genetic control of programmed cell death in the nematode C. elegans. *Cell*. **44**, 817–829
4. Schwartz, H. T., and Horvitz, H. R. (2007) The C. elegans protein CEH-30 protects male-specific neurons from apoptosis independently of the Bcl-2 homolog CED-9. *Genes Dev.* **21**, 3181–3194
5. Shaham, S., and Horvitz, H. R. (1996) Developing Caenorhabditis elegans neurons may contain both cell-death protective and killer activities. *Genes Dev.* **10**, 578–591
6. Yuan, J., and Horvitz, H. R. (1990) The Caenorhabditis elegans genes ced-3 and ced-4 act cell autonomously to cause programmed cell death. *Dev. Biol.* **138**, 33–41
7. Ramirez, M. L. G., and Salvesen, G. S. (2018) A primer on caspase mechanisms. *Semin. Cell Dev. Biol.* **82**, 79–85
8. Mydlarz, L. D., Fuess, L. E., Mann, W., Pinzón C, J. H., and Gochfeld, D. J. (2016) Cnidarian immunity: from genomes to phenomes. in *The Cnidaria, Past, Present and Future*, pp. 1–855, Springer International Publishing, 10.1007/978-3-319-31305-4
9. Irmler, M., Hofmann, K., Vaux, D., and Tschopp, J. (1997) Direct physical interaction between the Caenorhabditis elegans “death proteins” CED-3 and CED-4. *FEBS Lett.* **406**, 189–190
10. Tchernov, D., Kvitt, H., Haramaty, L., Bibbyd, T. S., Gorbunov, M. Y., Rosenfeld, H., and Falkowski, P. G. (2011) Apoptosis and the selective survival of host animals following thermal bleaching in zooxanthellate corals. *Proc. Natl. Acad. Sci. U. S. A.* **108**, 9905–9909
11. Wiens, M., Krasko, A., Perovic, S., and Müller, W. E. G. (2003) Caspase-mediated apoptosis in sponges: Cloning and function of the phylogenetic oldest apoptotic proteases from Metazoa. *Biochim. Biophys. Acta - Mol. Cell Res.* **1593**, 179–189
12. Furla, P., Sabourault, C., Zucchini, N., Allemand, D., Courtiade, J., and Richier, S. (2006) Oxidative stress and apoptotic events during thermal stress in the symbiotic sea anemone, Anemonia viridis. *FEBS J.* **273**, 4186–4198
13. Salvesen, G. S., and Walsh, C. M. (2014) Functions of caspase 8: The identified and the mysterious. *Semin. Immunol.* **26**, 246–252
14. Kortschak, R. D., Samuel, G., Saint, R., and Miller, D. J. (2003) EST analysis of the cnidarian Acropora millepora reveals extensive gene loss and rapid sequence divergence in the model invertebrates. *Curr. Biol.* **13**, 2190–2195
15. Hughes, T. P., Barnes, M. L., Bellwood, D. R., Cinner, J. E., Cumming, G. S., Jackson, J. B. C., Kleypas, J., Van De Leemput, I. A., Lough, J. M., Morrison, T. H., Palumbi, S. R., Van Nes, E. H., and Scheffer, M. (2017) Coral reefs in the Anthropocene. *Nature*. **546**, 82–90

16. Dunn, S. R., Schnitzler, C. E., and Weis, V. M. (2007) Apoptosis and autophagy as mechanisms of dinoflagellate symbiont release during cnidarian bleaching: Every which way you lose. *Proc. R. Soc. B Biol. Sci.* **274**, 3079–3085
17. Maynard, J., Van Hooidonk, R., Eakin, C. M., Puotinen, M., Garren, M., Williams, G., Heron, S. F., Lamb, J., Weil, E., Willis, B., and Harvell, C. D. (2015) Projections of climate conditions that increase coral disease susceptibility and pathogen abundance and virulence. *Nat. Clim. Chang.* **5**, 688–694
18. Mansfield, K. M., and Gilmore, T. D. (2019) Innate immunity and cnidarian-Symbiodiniaceae mutualism. *Dev. Comp. Immunol.* **90**, 199–209
19. Fuess, L. E., Pinzón C, J. H., Weil, E., Grinshpon, R. D., and Mydlarz, L. D. (2017) Life or death: disease-tolerant coral species activate autophagy following immune challenge. *Proceedings. Biol. Sci.* **284**, 20170771
20. Palmer, C. V., and Traylor-Knowles, N. (2012) Towards an integrated network of coral immune mechanisms. *Proc. R. Soc. B Biol. Sci.* **279**, 4106–4114
21. Fuess, L. E., Mann, W. T., Jinks, L. R., Brinkhuis, V., and Mydlarz, L. D. (2018) Transcriptional analyses provide new insight into the late-stage immune response of a diseased caribbean coral. *R. Soc. Open Sci.* **5**, 172062
22. Libro, S., Kaluziak, S. T., and Vollmer, S. V. (2013) RNA-seq profiles of immune related genes in the staghorn coral Acropora cervicornis Infected with white band disease. *PLoS One*. **8**, e81821
23. Kaniewska, P., Campbell, P. R., Kline, D. I., Rodriguez-Lanetty, M., Miller, D. J., Dove, S., and Hoegh-Guldberg, O. (2012) Major cellular and physiological impacts of ocean acidification on a reef building coral. *PLoS One*. **7**, e34659
24. Dunn, S. R., Thomason, J. C., Le Tissier, M. D. A., and Bythell, J. C. (2004) Heat stress induces different forms of cell death in sea anemones and their endosymbiotic algae depending on temperature and duration. *Cell Death Differ.* **11**, 1213–1222
25. Dunn, S. R., Bythell, J. C., Le Tissier, M. D. ., Burnett, W. J., and Thomason, J. C. (2002) Programmed cell death and cell necrosis activity during hyperthermic stress-induced bleaching of the symbiotic sea anemone Aiptasia sp. *J. Exp. Mar. Bio. Ecol.* **272**, 29–53
26. Pernice, M., Dunn, S. R., Miard, T., Dufour, S., Dove, S., and Hoegh-Guldberg, O. (2011) Regulation of apoptotic mediators reveals dynamic responses to thermal stress in the reef building coral acropora millepora. *PLoS One*. **6**, e16095
27. Park, H. H. (2012) Structural features of caspase-activating complexes. *Int. J. Mol. Sci.* **13**, 4807–4818
28. Sakamaki, K., and Satou, Y. (2009) Caspases: Evolutionary aspects of their functions in vertebrates. *J. Fish Biol.* **74**, 727–753
29. Grinshpon, R. D., Williford, A., Titus-McQuillan, J., and Clay Clark, A. (2018) The CaspBase: a curated database for evolutionary biochemical studies of caspase functional divergence and ancestral sequence inference. *Protein Sci.* **27**, 1857–1870
30. Clark, A. C. (2016) Caspase Allostery and Conformational Selection. *Chem. Rev.* **116**, 6666–6706
31. Tucker, M. B., MacKenzie, S. H., Maciag, J. J., Dirscherl Ackerman, H., Swartz, P., Yoder, J. A., Hamilton, P. T., and Clay Clark, A. (2016) Phage display and structural

studies reveal plasticity in substrate specificity of caspase-3a from zebrafish. *Protein Sci.* **25**, 2076–2088
32. Bose, K., Pop, C., Feeney, B., and Clark, A. C. (2003) An uncleavable procaspase-3 mutant has a lower catalytic efficiency but an active site similar to that of mature caspase-3. *Biochemistry*. **42**, 12298–12310
33. MacPherson, D. J., Mills, C. L., Ondrechen, M. J., and Hardy, J. A. (2019) Tri-arginine exosite patch of caspase-6 recruits substrates for hydrolysis. *J. Biol. Chem.* **294**, 71–88
34. Scheer, J. M., Romanowski, M. J., and Wells, J. A. (2006) A common allosteric site and mechanism in caspases. *Proc. Natl. Acad. Sci. U. S. A.* **103**, 7595–7600
35. Maillard, M. C., Brookfield, F. A., Courtney, S. M., Eustache, F. M., Gemkow, M. J., Handel, R. K., Johnson, L. C., Johnson, P. D., Kerry, M. A., Krieger, F., Meniconi, M., Muñoz-Sanjuán, I., Palfrey, J. J., Park, H., Schaertl, S., Taylor, M. G., Weddell, D., and Dominguez, C. (2011) Exploiting differences in caspase-2 and -3 S 2 subsites for selectivity: Structure-based design, solid-phase synthesis and in vitro activity of novel substrate-based caspase-2 inhibitors. *Bioorganic Med. Chem.* **19**, 5833–5851
36. Yan, N., Huh, J. R., Schirf, V., Demeler, B., Hay, B. A., and Shi, Y. (2006) Structure and activation mechanism of the Drosophila initiator caspase Dronc. *J. Biol. Chem.* **281**, 8667–8674
37. Huang, W., Jiang, T., Choi, W., Qi, S., Pang, Y., Hu, Q., Xu, Y., Gong, X., Jeffrey, P. D., Wang, J., and Shi, Y. (2013) Mechanistic insights into CED-4-mediated activation of CED-3. *Genes Dev.* **27**, 2039–2048
38. Feeney, B., Pop, C., Swartz, P., Mattos, C., and Clark, A. C. (2006) Role of loop bundle hydrogen bonds in the maturation and activity of (Pro)caspase-3. *Biochemistry*. **45**, 13249–13263
39. Liu, X., Zhang, H., Wang, X. J., Li, L. F., and Su, X. D. (2011) Get phases from arsenic anomalous scattering: De novo SAD phasing of two protein structures crystallized in cacodylate buffer. *PLoS One*. **6**, e24227
40. Wei, Y., Fox, T., Chambers, S. P., Sintchak, J., Coll, J. T., Golec, J. M. C., Swenson, L., Wilson, K. P., and Charifson, P. S. (2000) The structures of caspases-1, -3, -7 and -8 reveal the basis for substrate and inhibitor selectivity. *Chem. Biol.* **7**, 423–432
41. Boucher, D., Blais, V., and Denault, J.-B. (2012) Caspase-7 uses an exosite to promote poly(ADP ribose) polymerase 1 proteolysis. *Proc. Natl. Acad. Sci. U. S. A.* **109**, 5669–5674
42. Lesser, M. P. (2011) Coral Bleaching: Causes and Mechanisms. in *Coral Reefs: An Ecosystem in Transition*, pp. 405–419, Springer Netherlands, Dordrecht, 10.1007/978-94-007-0114-4_23
43. Cleves, P. A., Shumaker, A., Lee, J. M., Putnam, H. M., and Bhattacharya, D. (2020) Unknown to Known: Advancing Knowledge of Coral Gene Function. *Trends Genet.* **36**, 93–104
44. Weis, V. M., Davy, S. K., Hoegh-Guldberg, O., Rodriguez-Lanetty, M., and Pringle, J. R. (2008) Cell biology in model systems as the key to understanding corals. *Trends Ecol. Evol.* **23**, 369–376

45. Bhattacharya, D., Agrawal, S., Aranda, M., Baumgarten, S., Belcaid, M., Drake, J. L., Erwin, D., Foret, S., Gates, R. D., Gruber, D. F., Kamel, B., Lesser, M. P., Levy, O., Liew, Y. J., MacManes, M., Mass, T., Medina, M., Mehr, S., Meyer, E., Price, D. C., Putnam, H. M., Qiu, H., Shinzato, C., Shoguchi, E., Stokes, A. J., Tambutté, S., Tchernov, D., Voolstra, C. R., Wagner, N., Walker, C. W., Weber, A. P. M., Weis, V., Zelzion, E., Zoccola, D., and Falkowski, P. G. (2016) Comparative genomics explains the evolutionary success of reef-forming corals. *Elife*. **5**, e13288
46. Dimos, B. A., Butler, C. C., Ricci, C. A., Macknight, N. J., and Mydlarz, L. D. (2019) Responding to Threats Both Foreign and Domestic: NOD-Like Receptors in Corals. in *Integrative and Comparative Biology*, pp. 819–829, **59**, 819–829
47. Sakamaki, K., Imai, K., Tomii, K., and Miller, D. J. (2015) Evolutionary analyses of caspase-8 and its paralogs: Deep origins of the apoptotic signaling pathways. *BioEssays*. **37**, 767–776
48. Yu, X., Huang, B., Zhou, Z., Tang, J., and Yu, Y. (2017) Involvement of caspase3 in the acute stress response to high temperature and elevated ammonium in stony coral Pocillopora damicornis. *Gene*. **637**, 108–114
49. Lasi, M., Pauly, B., Schmidt, N., Cikala, M., Stiening, B., Käsbauer, T., Zenner, G., Popp, T., Wagner, A., Knapp, R. T., Huber, A. H., Grunert, M., Söding, J., David, C. N., and Böttger, A. (2010) The molecular cell death machinery in the simple cnidarian Hydra includes an expanded caspase family and pro- and anti-apoptotic Bcl-2 proteins. *Cell Res.* **20**, 812–825
50. Moya, A., Sakamaki, K., Mason, B. M., Huisman, L., Forêt, S., Weiss, Y., Bull, T. E., Tomii, K., Imai, K., Hayward, D. C., Ball, E. E., and Miller, D. J. (2016) Functional conservation of the apoptotic machinery from coral to man: The diverse and complex Bcl-2 and caspase repertoires of Acropora millepora. *BMC Genomics*. **17**, 62
51. MacKenzie, S. H., Schipper, J. L., England, E. J., Thomas, M. E., Blackburn, K., Swartz, P., and Clark, A. C. (2013) Lengthening the intersubunit linker of procaspase 3 leads to constitutive activation. *Biochemistry*. **52**, 6219–6231
52. Potter, S. C., Luciani, A., Eddy, S. R., Park, Y., Lopez, R., and Finn, R. D. (2018) HMMER web server: 2018 update. *Nucleic Acids Res.* **46**, W200–W204
53. Kumar, S., Stecher, G., and Tamura, K. (2016) MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol. Biol. Evol.* **33**, 1870–1874
54. Darriba, D., Taboada, G. L., Doallo, R., and Posada, D. (2011) ProtTest-HPC: Fast selection of best-fit models of protein evolution. in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, pp. 177–184, **6586 LNCS**, 177–184
55. Nguyen, L. T., Schmidt, H. A., Von Haeseler, A., and Minh, B. Q. (2015) IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* **32**, 268–274
56. Cohen, S. L., and Chait, B. T. (1997) Mass spectrometry of whole proteins eluted from sodium dodecyl sulfate- polyacrylamide gel electrophoresis gels. *Anal. Biochem.* **247**, 257–267
57. Thomas, M. E., Grinshpon, R., Swartz, P., and Clark, A. C. (2018) Modifications to a common phosphorylation network provide individualized control in caspases. *J. Biol. Chem.* **293**, 5447–5461

58. Grinshpon, R. D., Shrestha, S., Titus-McQuillan, J., Hamilton, P. T., Swartz, P. D., and Clark, A. C. (2019) Resurrection of ancestral effector caspases identifies novel networks for evolution of substrate specificity. *Biochem. J.* **476**, 3475–3492
59. Guex, N., Peitsch, M. C., and Schwede, T. (2009) Automated comparative protein structure modeling with SWISS-MODEL and Swiss-PdbViewer: A historical perspective. *Electrophoresis*. **30**, S162–S173
60. Pronk, S., Páll, S., Schulz, R., Larsson, P., Bjelkmar, P., Apostolov, R., Shirts, M. R., Smith, J. C., Kasson, P. M., van der Spoel, D., Hess, B., and Lindahl, E. (2013) GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit. *Bioinformatics*. **29**, 845–854
61. Wang, J., Cieplak, P., and Kollman, P. A. (2000) How well does a restrained electrostatic potential (RESP) model perform in calculating conformational energies of organic and biological molecules? *J. Comput. Chem.* **21**, 1049–1074
62. Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W., and Klein, M. L. (1983) Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **79**, 926–935
63. Maciag, J. J., Mackenzie, S. H., Tucker, M. B., Schipper, J. L., Swartz, P., and Clark, A. C. (2016) Tunable allosteric library of caspase-3 identifies coupling between conserved water molecules and conformational selection. *Proc. Natl. Acad. Sci. U. S. A.* **113**, E6080–E6088

Supplementary Information

Caspases from Scleractinian Coral Show Unique Regulatory Features

Suman Shrestha[a,1], Jessica Tung[a,1], Robert D. Grinshpon[b], Paul Swartz[b], Paul T. Hamilton[c], Bradford Dimos[a], Laura Mydlarz[a] and A. Clay Clark[a,2]

[a]Department of Biology, University of Texas at Arlington, Arlington, TX 76019, USA
[b]Department of Molecular and Structural Biochemistry, North Carolina State University, Raleigh, NC 27695, USA
[c]Department of Plant and Microbial Biology, North Carolina State University, Raleigh, NC 27695, USA

[1]Contributed equally to this work
[2]Address correspondence to A. Clay Clark at the Department of Biology, Box19498, 501 S. Nedderman Drive, 337 Life Science Building, University of Texas at Arlington Arlington, TX 76019. Phone: (817) 272-9226. Fax: (817) 272-2855.
E-mail: clay.clark@uta.edu

Key Words: caspase; coral apoptosis; functional divergence; substrate selection; card-caspase

Supplementary Table S1: Caspases from *Orbicella faveolata* and *Porites astreoides* with their assigned name based on sequence and domain similarity with human caspases, their respective accession number, and domains in their structure.

Given Name	Accession number	Domains
Orbicella faveolata		
OfCasp3a	XP_020613409.1	CARD, Peptidase_C14
OfCas3b	XP_020630525.1	Peptidase_C14
OfCasp7	XP_020613679.1	Peptidase_C14
OfCasp3c	XP_020630550.1	Peptidase_C14
OfCasp2	XP_020630531.1	CARD, Peptidase_C14
OfCasp8a	XP_020620405.1	DED, DED, Peptidase_C14
OfCasp8b	XP_020629413.1	Peptidase_C14
Porites astreoides		
PaCasp-3	comp-76580	Peptidase_C14
PaCasp-7a	comp-74978	CARD, Peptidase_C14
PaCasp2	comp-74936	CARD, Peptidase_C14
PaCas7b	comp-77018	Peptidase_C14

Supplementary Table S2: Caspases from invertebrates and vertebrates used in the phylogenetic analysis and their respective accession number.

Caspases name	Accession number
Acropora digitifera	
AdCasp3a	XP_015775441.1
AdCasp3b	XP_015766400.1
AdCasp3c	XP_015762449.1
AdCasp8	XP_015761120.1
AdCasp3d	XP_015753767.1
Alligator mississippiensis	
AmCasp6	XP_019355646.1
AmCasp8	XP_006272599.1
AmCasp10	XP_014449789.1
AmCasp7	XP_014450146.1
AmCasp2	XP_014464708.1
AmCasp3	XP_019336883.1
AmCasp9	XP_019355521.1
Danio rerio	
DrCasp2	NP_001036160.1
DrCasp3a	XP_001338890.2
DrCasp3b	XP_005173133.1
DrCasp6a	XP_005164109.1
DrCasp6b	XP_017210076.1
DrCasp6c	NP_001018333.1
DrCasp7	XP_005156389.1
DrCasp8a	NP_571585.2

DrCasp8b	NP_001092089.1
DrCasp9	NP_001007405.2
Exaiptasia pallida	
EpCasp3a	XP_020893866.1
EpCasp3b	XP_020905061.1
Gallus gallus	
GgCasp2	NP_001161173.1
GgCasp3	NP_990056.1
GgCasp6	NP_990057.1
GgCasp7	XP_421764.3
GgCasp8	NP_989923.1
GgCasp9	XP_424580.5
GgCasp10	XP_421936.4
Homo sapiens	
HsCasp2	NP_116764.2
HsCasp3	NP_004337.2
HsCasp6	NP_001217.2
HsCasp7	NP_001253985.1
HsCasp8	NP_001219.2
HsCasp9	NP_001220.2
HsCasp10	NP_116759.2
Hydra vulgaris	
HvCasp2	NP_001274285.1
HvCasp3a	XP_012557085.1
HvCasp3b	XP_002159783.3
HvCasp7	XP_012561656.1
HvCasp8	XP_012562456.1
Mus musculus	
MmCasp2	NP_031636.1
MmCasp3	NP_001271338.1
MmCasp6	NP_033941.3
MmCasp7	XP_006526679.1
MmCasp8	NP_001264855.1
MmCasp9	NP_056548.
Nematostella vectensis	
Nv.Casp3	XP_001633895.1
Pocillophora damicornis	
Pd.Casp3	XP_027037576.1
Stylophora pistillata	
SpCasp3a	XP_022784432.1
SpCasp3b	XP_022808070.1
Sp.Casp3c	PFX33553.1
SpCasp8a	XP_022796790.1
SpCasp8b	XP_022789601.1
Xenopus laevis	

XtCasp2	XP_012809163.1
XtCasp3	NP_001120900.1
XtCasp6	NP_001011068.1
XtCasp7	NP_001016299.1
XtCasp8	XP_017953067.1
XtCasp10	NP_001015715.2

Supplementary Table S3: Characteristics of coral caspases.

Composition [1]	Protein						
	OfCasp3a	OfCasp3b	PaCasp3	PaCasp7a	HsCasp3	HsCasp6	HsCasp7
Total Number Amino Acids	392	299	298	390	277	293	303
Ala (A)	24 (6.1%)	22 (7.4%)	16 (5.4%)	31 (7.9%)	12 (4.3%)	19 (6.5%)	18 (5.9%)
Arg (R)	26 (6.6%)	18 (6.0%)	16 (5.4%)	25 (6.4%)	14 (5.1%)	17 (5.8%)	15 (5.0%)
Asn (N)	19 (4.8%)	11 (3.7%)	16 (5.4%)	19 (4.9%)	15 (5.4%)	11 (3.8%)	14 (4.6%)
Asp (D)	29 (7.4%)	24 (8.0%)	23 (7.7%)	30 (7.7%)	20 (7.2%)	20 (6.8%)	27 (8.9%)
Cys (C)	3 (0.8%)	9 (3.0%)	9 (3.0%)	3 (0.8%)	8 (2.9%)	10 (3.4%)	11 (3.6%)
Gln (Q)	10 (2.6%)	10 (3.3%)	14 (4.7%)	12 (3.1%)	4 (1.4%)	7 (2.4%)	11 (3.6%)
Glu (E)	28 (7.1%)	17 (5.7%)	14 (4.7%)	26 (6.7%)	20 (7.2%)	20 (6.8%)	19 (6.3%)
Gly (G)	25 (6.4%)	16 (5.4%)	20 (6.7%)	23 (5.9%)	16 (5.8%)	19 (6.5%)	18 (5.9%)
His (H)	5 (1.3%)	9 (3.0%)	6 (2.0%)	5 (1.3%)	8 (2.9%)	12 (4.1%)	7 (2.3%)
Ile (I)	21 (5.4%)	16 (5.4%)	15 (5.0%)	21 (5.4%)	19 (6.9%)	13 (4.4%)	17 (5.6%)
Leu (L)	30 (7.7%)	19 (6.4%)	20 (6.7%)	25 (6.4%)	20 (7.2%)	26 (8.9%)	20 (6.6%)
Lys (K)	26 (6.6%)	17 (5.7%)	14 (4.7%)	24 (6.2%)	22 (7.9%)	20 (6.8%)	25 (8.3%)
Met (M)	12 (3.1%)	6 (2.0%)	9 (3.0%)	12 (3.1%)	10 (3.6%)	7 (2.4%)	7 (2.3%)
Phe (F)	19 (4.8%)	17 (5.7%)	17 (5.7%)	19 (4.9%)	15 (5.4%)	18 (6.1%)	17 (5.6%)
Pro (P)	13 (3.3%)	14 (4.7%)	17 (5.7%)	16 (4.1%)	7 (2.5%)	10 (3.4%)	12 (4.0%)
Ser (S)	42 (10.7%)	28 (9.4%)	28 (9.4%)	35 (9.0%)	26 (9.4%)	18 (6.1%)	21 (6.9%)
Thr (T)	22 (5.6%)	13 (4.3%)	10 (3.4%)	23 (5.9%)	16 (5.8%)	16 (5.5%)	15 (5.0%)
Trp (W)	2 (0.5%)	1 (0.3%)	1 (0.3%)	2 (0.5%)	2 (0.7%)	2 (0.7%)	2 (0.7%)
Tyr (Y)	16 (4.1%)	9 (3.0%)	9 (3.0%)	15 (3.8%)	10 (3.6%)	10 (3.4%)	9 (3.0%)
Val (V)	20 (5.1%)	23 (7.7%)	24 (8.1%)	24 (6.2%)	13 (4.7%)	18 (6.1%)	18 (5.9%)
Molecular Weight (Da)	44192.61	33437.75	33192.51	43727.11	31,608	33,310	34,277
pI	5.75	5.87	5.42	5.49	6.09	6.46	5.72
Extinction Coefficient (280 nm, $M^{-1}cm^{-1}$) [2]	34840	18910	18910	33350	26,500	25,900	24,410

[1] Parameters exclude the LEHHHHHH C-terminal tag.

[2] Assuming all cysteine residues are reduced.

Supplementary Table S4: PaCasp7a crystal statistics.

PDB Code		**Refinement**	
Data collection		**R_{work}/R_{free} (%)**	17.7/21
Wavelength	1.0	**Average B-factor ($Å^2$)**	17.76
Temperature(K)	100	**Macromolecules**	16.40
Space Group	P 2_1 2_1 2_1	**Solvent**	27.77
Cell Dimensions		**Wilson B-factor**	13.39
a, b, c (Å)	74.416 86.848 93.666	**R. m. s. deviations**	
α, β, γ (°)	90.00 90.00 90.00	**Bond length (Å)**	0.077
#Unique Reflection	83910	**Bond angle (°)**	5.004
Resolution (Å)	39.4 – 1.57	**MolProbity score**	4.84
R-meas	0.249	**Number of atoms**	
R-pim	0.117	**Protein**	3823
CC(1/2)	0.255	**Water**	586
Average I/σ	1.33	**Protein residues**	466
Completeness (%)	98.5	**MolProbity**	
Redundancy	4.9	**Ramachandran favored**	97.86%
Clash score	3.42	**Ramachandran outliers**	0.00%

		Rotamer outliers	0.24%

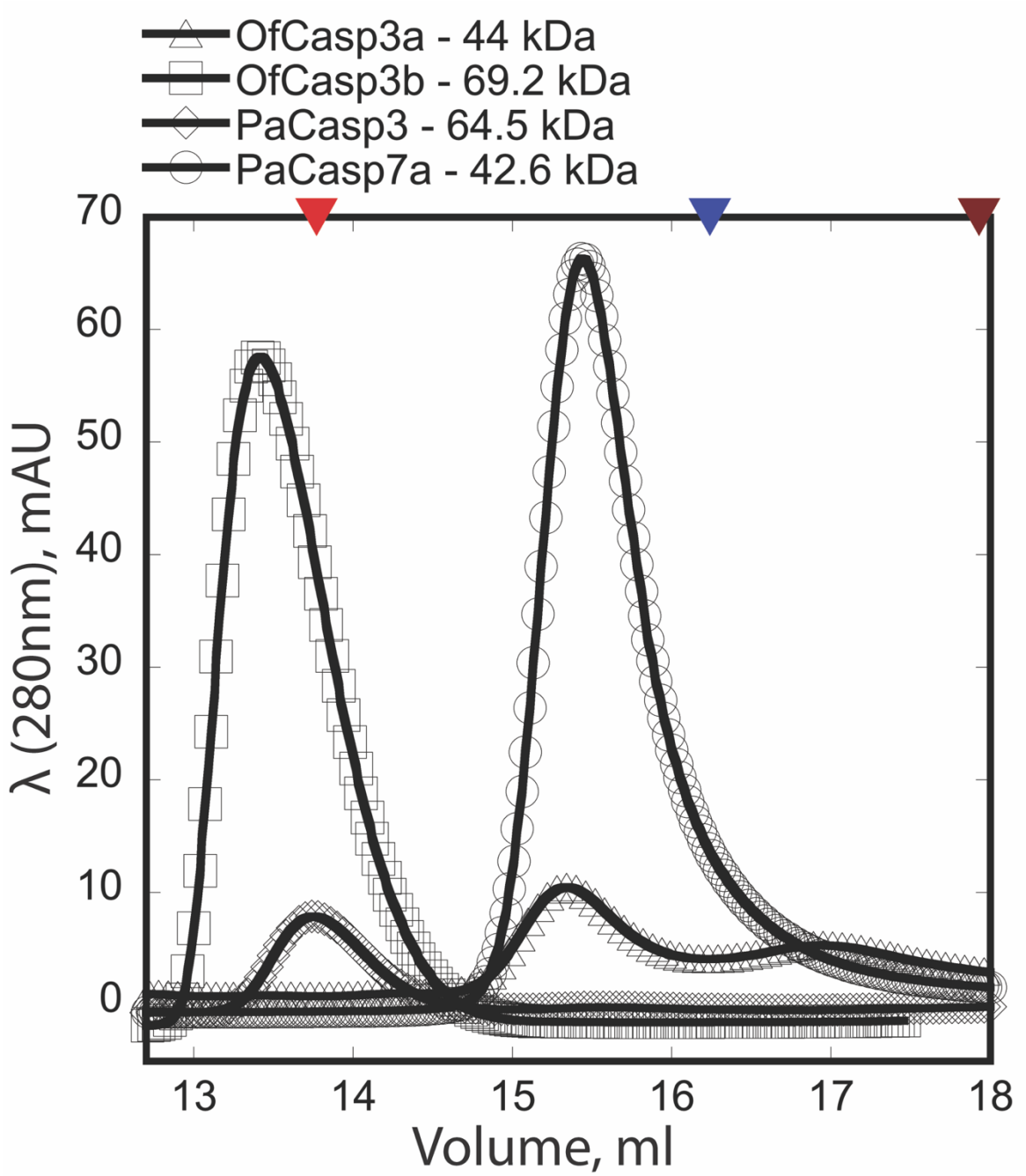
Supplementary Table S5: Molecular weight determined using ProtParam and mass spectrometry.

Protein	M.W. Full length (kDa) (Protparam)	M.W. Large Subunit (kDa) (Mass Spec.)	M.W. Small Subunit (kDa) (Mass Spec.)	M. W. A Protomer (cleaved IL and prodomain)
OfCasp3a	45.3	17.9	13.2	31.1
OfCasp3b	34.5	20.3	12.1	32.4
PaCasp3	34.3	18.9	12.9	31.8
PaCasp7a	44.8	17.8	13.3	31.1

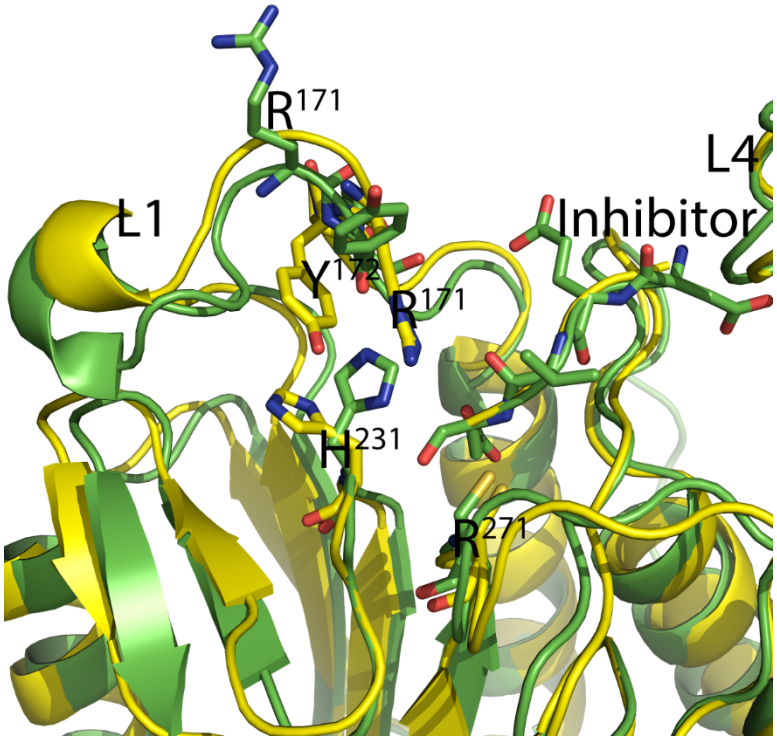
Supplementary Table S6: Homologs of components of human apoptotic pathways in *O. faveolata* and *P. astreoides*.

Protein_name	Uniprot ID (Human)	*Orbicella faveolata*	*Porites astreoides*
Bcl-2	P10415	XP_020601884.1	comp77460_c1_seq5.p1
Bcl-xL	Q07817	XP_020600881.1	comp71697_c0_seq4.p1
Bax	Q07812	XP_020620765.1	comp65965_c0_seq1.p1
Bak	Q16611	XP_020601942.1	comp78640_c3_seq2.p1
PIDD	Q9HB75	XP_020624054.1	comp63514_c0_seq4.p1
RAIDD	P78560	XP_020606624.1	comp74936_c0_seq2.p1
P53	P04637	XP_020628162.1	comp71445_c0_seq1.p1
Apaf-1	O14727	XP_020620792.1	comp76833_c0_seq1.p1
Cytochrome c	P99999	XP_02062022.1	Isotig14309.p1

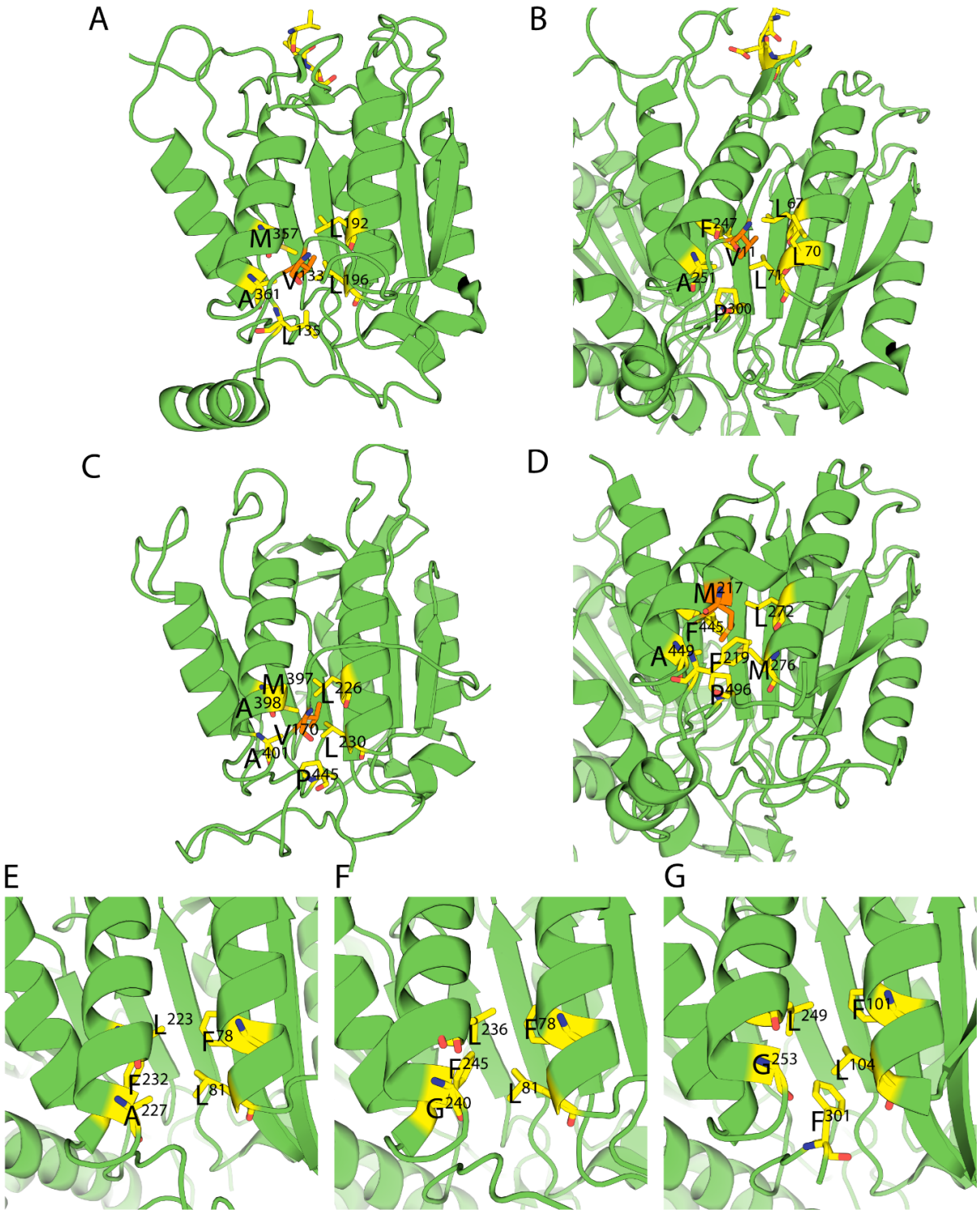
Supplementary Figure S1: Multiple sequence alignment of caspases from *O. faveolata* and *P. astreoides*.



Supplementary Figure S2: Chromatogram (A_{280}) of coral caspases. Peaks based on elution volume after FPLC analysis of the native oligomeric state through sizing column. Sigma-Aldrich gel filtration kit was used as a marker; Albumin (66 kDa, ▼), Carbonic anhydrase (29 kDa, ▼) and cytochrome c (12.4 kDa, ▼). OfCasp3a (*Orbicella faveolata* caspase-3a), OfCasp3b (*Orbicella faveolata* caspase-3b), PaCasp7a (*Porites astreoides* caspase-7a) and PaCasp3 (*Porites astreoides* caspase-3) refer to the four coral caspases characterized in the main text.



Supplementary Figure S3: Average structures of molecular dynamics (MD) simulations showing loop 1 (L1) containing “RYP” motif in the “in” *versus* “out” conformations. Green: Past7a crystal structure and “RYP out”, Yellow: PaCasp7a average structure from MD simulations of PaCasp7a model structure and “RYP in”. The data show that the two conformations do not interconvert on the timescale of the MD simulations (50 ns).



Supplementary Figure S4: Structures of caspases with N-terminal peptide bound between helices 1 and 4. A. HsCasp1 (PDB ID: 2H48), B. HsCasp2 (PDB ID: 3RJM), C. *D. melanogaster* initiator caspase Dronc (PDB ID: 2FP3), D. *C. elegans* caspase CED-3 (PDB ID: 4M9R), E-G. Hydrophobic pocket between helices 1 and 4 in human effector caspases. E. HsCasp3 (PDB ID: 2J30), F. HsCasp6 (PDB ID: 3S70), G. HsCasp7 (PDB ID: 1F1J).

Chapter 3

Evolution of the folding landscape of effector caspases

Suman Shrestha, and A. Clay Clark*

Department of Biology, University of Texas at Arlington, Arlington, TX 76019, USA

*Corresponding author: A. Clay Clark E-mail: clay.clark@uta.edu

Keywords: caspase, protein folding, folding landscape, protein evolution, oligomerization, evolutionary biology

The contents of this chapter are to be published

Abstract

Caspases are a family of cysteinyl proteases that control programmed cell death and maintain homeostasis in multicellular organisms. The caspase family is an excellent model to study protein evolution because all caspases are produced as zymogens (procaspases) that must be activated to gain full activity; the protein structures are conserved through millions of years of evolution; and some allosteric features arose with the early ancestor while others are more recent evolutionary events. The apoptotic caspases evolved from a common ancestor into two distinct subfamilies: monomers (initiator caspases) or dimers (effector caspases). Differences in activation mechanisms of the two subfamilies, and their oligomeric forms, play a central role in the regulation of apoptosis. Here, we examined changes in procaspase stability and changes in the folding landscape by characterizing human effector caspases and their common ancestor. The results show that the effector caspases unfold by a minimum three-state equilibrium model at pH 7.5, where the native dimer is in equilibrium with a partially folded monomeric (procaspase-7) or dimeric (procaspase-6) intermediate. In comparison, the unfolding pathway of procaspase-3 contains both oligomeric forms of the intermediate. Overall, the common ancestor forms a weak dimer which was stabilized early in the evolution of the subfamily. The folding landscape was first established with the common ancestor and was then retained for >650 million years. Procaspase-7 is closest to the common ancestor, whereas procaspases-6 and -3 show evolutionary changes that affect stability by stabilizing conformations of lower population in the ancestral ensemble.

Introduction

While folding landscapes of proteins have been studied for decades (1–3), many studies focused on small monomeric proteins or dimers with simple folding landscapes.. Studies of monomeric proteins have provided a wealth of information concerning the principles that govern intramolecular interactions during folding, but they do not consider intermolecular interactions provided by the interfaces of multimeric proteins (4). For some dimers, subunit interactions in dimeric interface lead to more complicated folding pathways compared to simple two-state behavior involving only native dimer (N_2) and unfolded monomers (U) (5, 6). Moreover, relatively little is known about the evolution of dimeric proteins compared to monomeric proteins (2, 7), although two-thirds of proteins form a multimeric assembly (8, 9). Thus, an understanding of the role of oligomerization in the folding landscape of a polypeptide sequence should include a consideration of the interface in assembly (10). In this regard, the caspase family of proteases is an attractive model system to study the evolution of protein folding landscapes. Caspases are a family of cysteinyl aspartate-specific proteases that initiate and execute programmed cell death and maintain cellular homeostasis in metazoans (11). There are two broad categories of caspases based on their function: inflammatory caspases (caspases -1, -4, and -5) and apoptotic caspases. The latter is further subdivided into two groups based on their entry into the apoptotic cascade. Initiator caspases (caspases-2, -8, -9 and -10) act upstream and activate downstream effector procaspases (caspases-3, -6 and -7), which execute the cell death function (12). In addition, the caspase-hemoglobinase fold has been conserved for >650 million

years(13), so the caspase family provides opportunities to examine the folding of monomers as well as the role of dimerization in the folding landscape.

More recently, evolutionary biochemists have used a new approach to study protein folding by examining the evolution of protein structure and function. Ancestral reconstruction methods enable one to use vertical comparisons (comparing ancestral to modern enzymes) as well as horizontal comparisons (comparing modern enzymes from multiple species) (14–16). The results of such studies potentially show the mechanisms by which changes in protein sequence have caused shifts in structure and function (17). The sequence determinants of protein structure-function, and substitutions revealed by the evolutionary analysis in common evolutionary nodes can then be introduced singly or in combination into ancestral backgrounds. Ultimately, by examining the ancestral reconstructions and changes that occur throughout evolution of the protein, one can determine the effects of historical mutations on protein structure, function, and physical properties (16).

The effector and initiator caspases evolved by gene duplication and divergence from a common ancestor (CA) more than 650 million years ago (Mya) (13). The CA provided a scaffold for the modern caspases to acquire distinct properties, such as formation of oligomers, changes in stability, enzyme specificity, and allosteric regulation(13). All caspases are produced in the cell as inactive zymogens that must be activated during the apoptotic cascade. In general, initiator procaspases are stable monomers, and dimerization is sufficient for activation, whereas effector procaspases are stable, yet inactive dimers, and are activated via cleavage by initiator caspases (Fig. 1). The oligomeric form of the zymogen and its activation mechanism is key to

regulation of apoptosis (11, 13). While the amino acid sequence identity is low between caspase subfamilies (~40%) (13), the caspase-hemoglobinase fold is well-conserved (18) even from distantly related species of vertebrates and invertebrates, such as human, *Danio rerio*, *Caenorhabditis elegans*, *Drosophila melanogaster,* and *Porites astreoides* (19–23).The structure of the procaspase monomer is characterized by a 6-stranded β-sheet core with several α-helices on the surface (4). Each monomer of the procaspase homodimer consists of approximately 300 amino acids organized into an N-terminal prodomain followed by a protease domain. The protease domain is further divided into a large and a small subunit that are connected by a short intersubunit linker. (Fig. 1B).

Previously, we showed that the human procaspase-3 dimer assembles by a four-state equilibrium mechanism in which the unfolded protein folds to a monomeric intermediate. Following dimerization of the intermediate, the protein undergoes a conformational change to form the native dimer ($2U \rightleftarrows 2I \rightleftarrows I_2 \rightleftarrows N_2$) (24). Dimerization results in a substantial increase in conformational free energy, $\Delta G°_{conf}$, for the dimer (~25 kcal mol^{-1}) *vs.* the monomer (~7 kcal mol^{-1}) at 25°C (24). Furthermore, by examining the changes in $\Delta G°_{conf}$ *vs.* pH, we showed that the procaspase-3 dimer dissociates at lower pH due to a pair of histidine residues that contribute to salt bridges across the dimer interface (25). Thus, our previous data showed that the per residue contribution to the total conformational free energy of the dimer ($\Delta G°_{conf}$) increases from 0.025 kcal/mol/aa to 0.044 kcal/mol/aa (24, 25). However, because human procaspase-3 is the only caspase in which the folding properties have been examined, the evolutionary trajectories that resulted in dimer formation remain unknown.

Previously, we used ancestral protein reconstruction (APR) techniques to resurrect the common ancestor (CA) of the caspase-3/6/7 subfamily (26). In those studies, we examined the robustness of reconstruction methods by resurrecting two sequences (called AncCP-Ef1 and AncCP-Ef2) from the pool of possible sequences of CA, and we characterized the proteins biochemically and structurally (13). Here, we examine the evolution of the caspase folding landscape using the ancestral reconstruction AncCP-Ef2 (referred to here as PCP-CA, procaspase-common ancestor). The results are compared to those for extant human procaspase-6 (called PCP6) and procaspase-7 (called PCP7) as well as our previous studies for procaspase-3 (called PCP3) (4, 24, 25). We examined urea-induced equilibrium unfolding over a broad pH range to compare the folding pathways of all three effector caspase subfamilies. The data show that the caspase folding landscape was first established and then retained for >650 million years. The common ancestor PCP-CA forms a weak dimer, and the dimer was stabilized early in the evolution of the subfamily. Of the three extant human effector caspases, caspase-6 is the most stable, while caspase-7 is the least stable. The folding landscape of procaspase-7 is more similar to that of the common ancestor, which is consistent with previous phylogenetic data that show caspase-7 is closest to the common ancestor (13). In procaspases-3 and -6, folding intermediates were stabilized later in evolution, and a pH-dependent dimer-tetramer equilibrium was established late in the evolution of caspase-6.

Experimental procedures

Cloning, protein expression, and protein purification

For all procaspase proteins, the catalytic cysteine (CP-117, colored blue in Figure 2) was mutated to serine using site-directed mutagenesis, as described previously (27). The inactive procaspases were cloned into pET11a expression vector with a C-terminal hexahistidine tag, and all proteins were expressed in *E. coli* BL21(DE3) pLysS cells and purified as previously described (28–30).

Size exclusion chromatography

Proteins were examined using a Superdex75 Increase 10/300GL column on an AKTA-FPLC. The proteins were concentrated to 1-5 mg/mL and dialyzed for four hours in a buffer of 30 mM potassium phosphate, pH 7.5, containing 1 mM DTT. The column was equilibrated with two-column volumes (50 mL) of the same buffer. Protein (200 μL) was loaded onto the column, and the column was resolved at a flow rate of 0.5 mL/min. The column was calibrated using the gel filtration LMW calibration kit (GE Health Sciences) following the manufacturer's instructions.

Sample preparation for equilibrium unfolding

Denaturation and renaturation experiments were carried out as described previously (31). Briefly, urea stock solutions (10 M) were prepared in citrate buffer (20 mM sodium citrate/citric acid, pH 4 to pH 5.5, 1 mM DTT), phosphate buffer (20 mM potassium phosphate monobasic and dibasic, pH 6 to 8, 1 mM DTT), or Tris buffer (20 mM Tris-HCl, pH 8.5, 1 mM DTT). For unfolding experiments, samples were prepared in

the corresponding buffer with final urea concentrations between 0 and 8 M. Stock protein in buffer was added such that the final concentrations are as shown in the figures. For the renaturation experiments, the protein was first incubated in an 8 M urea-containing buffer for 3 hours at 25 °C. The unfolded protein was then diluted with the corresponding buffer and urea such that the final urea concentrations were between 0.5 and 8 M. For all equilibrium unfolding experiments, protein concentrations from 0.5 to 4 μM were used. In both denaturation and renaturation experiments, the samples were incubated at 25 °C for a minimum of 16 h to allow for equilibration. This incubation time was found to be optimal to allow the protein to reach equilibrium at all urea concentrations.

Fluorescence and circular dichroism (CD) measurements

Fluorescence emission was acquired using a PTI C-61 spectrofluorometer (Photon Technology International, Birmingham, NJ) from 300 nm to 400 nm following excitation at 280 or 295nm. Excitation at 280 nm follows tyrosinyl and tryptophanyl fluorescence emission, whereas excitation at 295 nm follows the tryptophanyl fluorescence emission. CD measurements were recorded using a J-1500 CD spectropolarimeter (Jasco) between 220 and 240 nm. Fluorescence and CD spectra were measured using a 1 cm path length cuvette and constant temperature (25 °C). All data were corrected for buffer background.

Data analysis and global fits of the equilibrium unfolding data

The data were fit globally and interpreted as described previously (24, 25, 31). Briefly, fluorescence emission and CD data were collected between pH 8.5 and 4 for all three proteins and at three to four protein concentrations, which resulted in 9 - 12 data sets at each pH. The data were fit to a 2-state or 3-state equilibrium folding model, as described below. At pH 7.5, the data for PCP6 were best fit to a 3-state equilibrium folding model described with a dimeric intermediate in equilibrium with the native and unfolded protein, as shown in eq 1. In contrast, the data for PCP7 and PCP-CA were best fit to a 3-state equilibrium folding model described with a monomeric intermediate in equilibrium with the native and unfolded protein, as shown in eq 2.

$$N_2 \overset{K_1}{\Leftrightarrow} I_2 \overset{K_2}{\Leftrightarrow} 2U \qquad (1)$$

$$N_2 \overset{K_1}{\Leftrightarrow} 2I \overset{K_2}{\Leftrightarrow} 2U \qquad (2)$$

In both equations 1 and 2, K_1 and K_2 refer to equilibrium constants for the two steps, respectively. At pH 4, the data for PCP6 was best fit to a two-state equilibrium folding model, where the native dimer is in equilibrium with the unfolded protein, as shown in equation 3.

$$N_2 \overset{K_{eq}}{\Leftrightarrow} 2U \qquad (3)$$

In contrast to PCP6, at pH 4 the data for PCP7 was best fit to a 2-state model where the native monomer is in equilibrium with the unfolded protein, as shown in equation 4, and PCP-CA was best fit to a 3-state equilibrium folding model for a monomer, as shown in equation 5.

$$N \overset{K_{eq}}{\Leftrightarrow} U \qquad (4)$$

$$N \overset{K_1}{\Leftrightarrow} I \overset{K_2}{\Leftrightarrow} U \qquad (5)$$

For all proteins, the equilibrium folding/unfolding data at each pH were fit globally using the appropriate folding model from equations 1-5and the program Igor Pro (WaveMetrics, Inc.), as described previously (24, 25, 31). Results of the fits are shown as the solid lines in Figure 4 and Supplementary Figures 3-7, and in Supplementary Table 3 & 4.

Results

A protomer of PCP6, PCP7, and PCP-CA consists of 293, 303, and 275 amino acids, respectively (Table S1). Under some conditions, such as the high protein concentrations found in heterologous expression systems, effector caspase activation is autocatalytic. In order to prevent autoproteolysis during expression in *E. coli*, we substituted the active site cysteine with a serine residue for our equilibrium unfolding studies, as described previously with procaspase-3 (24).

PCP6 and PCP7 have two tryptophan residues, while PCP-CA has only one tryptophan (Fig. 2 & Table S1). In PCP6, one tryptophan resides in active site loop 1 (Fig. S2), whereas the second tryptophan resides in active site loop 3. In PCP7 and in PCP-CA, the tryptophans are found only in active site loop 3 (Fig. 2 &Fig. S2). PCP6, PCP7, and PCP-CA have 10, 9, and 13 tyrosine residues, respectively, and they are well distributed in the primary sequence (Fig. 2 & Fig. S2). As shown in Figs. 3A & 3B, native PCP6 has a fluorescence emission maximum at 319 nm while that of native PCP7 is at 340 nm, when excited either at 280 nm or 295 nm (Figs. 3C & 3D). In the case of an PCP-CA, the fluorescence emission maximum is 321 nm when excited at

280 nm and 330 nm when excited at 295 nm (Figs. 3E & 3F). Overall, the data show that the tryptophan residues in PCP7 are more solvent-exposed than those of PCP6 or of PCP-CA. In phosphate buffer containing 8 M urea, the fluorescence emission maximum is red shifted to ~350 nm following excitation at 280 nm or 295 nm in all three proteins, indicating that the proteins were largely unfolded under these solution conditions (Fig. 3). At intermediate concentrations of urea (3 to 5 M), the emission maxima were red-shifted in the case of PCP6 and PCP-CA, but blue-shifted in the case of PCP7, indicating that an equilibrium folding intermediate may be populated under these conditions. The changes in emission maxima are described more fully below.

Equilibrium unfolding of extant and ancestor caspases

Changes in the fluorescence emission and circular dichroism properties of PCP3 as a function of urea concentration have been described previously (24, 25). In this study, we examined the equilibrium unfolding of PCP6, PCP7, and PCP-CA at pH 7.5 as a function of urea concentration (0 - 8 M), and the results are shown in Fig. 4. Renaturation experiments of all three proteins demonstrated the folding transitions are reversible.

For PCP6 at pH 7.5, the data show little to no change in signal between 0 and ~2.5 M urea. One then observes a cooperative decrease in the signal between ~3.5 and 5 M urea, demonstrating a plateau between the native and unfolded signals. A second cooperative transition occurs between ~5 and 6.5 M urea (Figs. 4A & 4B). For PCP6, the relative signal of the plateau as well as the second cooperative transition are dependent on the protein concentration. In contrast, both PCP7 and PCP-CA show a

protein-concentration dependent increase in the relative signal between 0 and ~3 M urea, followed by a cooperative decrease in the signal between ~3 and 6 M urea to form the unfolded state (Figs. 4D-4E & 4G-4H). Thus, the protein concentration dependence to unfolding is similar for PCP7 and PCP-CA, where the dependence is observed in the first unfolding transition rather than the second transition as in the case of PCP6. The fluorescence emission data are similar regardless of excitation at 280 nm or 295 nm. The cooperative transitions of all three proteins through CD signal were similar to the fluorescence signal although there are few notable differences in terms of concentration dependence. Relative CD signal of varying concentration of PCP6 differs from ~0 - 2.5 M urea unlike fluorescence data however, there is no change in midpoint of transition. Again, the concentration dependence in CD signal of PCP7 and PCP-CA were not as robust as in fluorescence signal. The Concentration dependent signal varies from ~0 - 2 M of urea in PCP7 and ~2.5 - 5.5 M urea in PCP-CA. Overall the data suggest cooperative changes in both the tertiary and secondary structures during each unfolding transition (Fig. 4).

Global fitting of equilibrium unfolding data

The experimental design described above at pH 7.5 for monitoring fluorescence emission (four protein concentrations, each with two excitation wavelengths) and circular dichroism (three protein concentrations) provides 11 data sets for each protein that were fit globally to determine the free energy and the cooperativity indices (m-values) for each unfolding transition. In the case of PCP6, the data were best fit to the three-state equilibrium model described in equation 1. In this model, the dimeric native

conformation, N_2, isomerizes to a dimeric intermediate, I_2, and the dimeric intermediate dissociates and unfolds to monomers. The dissociation of I_2 to 2U leads to a protein-concentration dependent change in the mid-point of the second transition, as shown in Figures 4A-4C. Based on this model, we have determined the conformational free energy, $\Delta G°_{conf}$, and the m-values for each step in unfolding. The solid lines in Figs. 4A - 4C are the results of global fits of the model to the data. The free energy change, $\Delta G_1^{H_2O}$ and the cooperativity index, m_1, for the first step of unfolding, the isomerization of N_2 to I_2, are 8.4 ± 0.82 kcal/mol and 2.56 ± 0.25 kcal mol^{-1} M^{-1}, respectively (Table 1). The free energy change, $\Delta G_2^{H_2O}$, and cooperativity index, m_2, for the dissociation and complete unfolding of the dimeric intermediate to two unfolded monomeric proteins ($I_2 \rightleftarrows 2U$) are 24.4 ± 0.88 kcal/mol and 2.93 ± 0.15 kcal mol^{-1} M^{-1}, respectively. Overall, the data demonstrate that PCP6 is very stable, with the total conformational free energy of 32.8 kcal/mol at pH 7.5 (Table S2 & S3).

For PCP7 and PCP-CA, the global fits demonstrate that the data are well described by a three-state equilibrium model. In contrast to PCP6, described above, the partially folded intermediate is monomeric for PCP7 and for PCP-CA (equation 2). The solid lines in Figs. 4D – 4F (PCP7) and 4G – 4I (PCP-CA) are the results of fits of the model to the data. In the case of PCP7, the free energy change, $\Delta G_1^{H_2O}$, and the cooperativity index, m_1, for the first step of unfolding, the dissociation of N_2 to 2I, are 10.2 ± 0.16 kcal/mol and 1.27 ± 0.03 kcal mol^{-1} M^{-1}, respectively. The free energy change, $\Delta G_2^{H_2O}$, and cooperativity index, m_2, for the complete unfolding of the monomeric intermediate to unfolded monomeric proteins ($I \rightleftarrows U$) are 5.2 ± 0.03 kcal/mol and 1.22 ± 0.001 kcal mol^{-1} M^{-1}, respectively (Table 1). Similarly, for PCP-CA, $\Delta G_1^{H_2O}$,

and m_1 are 14.9 ± 0.14 kcal/mol and 2.02 ± 0.04 kcal mol^{-1} M^{-1}, respectively. For unfolding of the monomeric intermediate of PCP-CA (I ⇄ U), $\Delta G_2^{H_2O}$ and m_2, are 6.9 ± 0.31 kcal/mol and 1.21 ± 0.06 kcal mol^{-1} M^{-1}, respectively (Table 1). Overall, the data suggest a minimum three-state process in all effector caspases in which a well-populated intermediate is in equilibrium with the native and unfolded protein. Comparatively, of the three human effector caspases, PCP6 is the most stable with $\Delta G°_{conf}$ of 32.8 kcal/mol, PCP-CA is intermediate with $\Delta G°_{conf}$ of 21.8 kcal/mol, and PCP7 is the least stable with $\Delta G°_{conf}$ of 15.4 kcal/mol at pH 7.5 (Fig. 5).

pH effects on equilibrium unfolding of effector caspases

We showed previously that PCP3 undergoes pH dependent conformational changes, and the dimer dissociates below pH 5.5 (25). From our previous data, we suggested that dimer dissociation was due to a series of salt bridges across the dimer interface, which include two histidine residues. The other effector caspases also have charged amino acids that interact across the dimer interface, but only PCP3 contains the histidine residues. In order to determine the effects of pH on dimer stability, we performed equilibrium unfolding studies of PCP6, PCP7 and PCP-CA between pH 8.5 and pH 4 (Figure 4 & Figures S3 – S7). Similar to the data described above at pH 7.5, the folding/unfolding of PCP6 can be described by a three-state equilibrium model (eq. 1) over the range of pH, from 4.5 to 8.5, with $\Delta G°_{conf}$ ~30 kcal/mol. On the other hand, the unfolding data at pH 4 were best described by two-state equilibrium model where the native dimer is in equilibrium with the unfolded monomer (eq 3) (Figs. S3 & S4). Surprisingly, even at pH 4, PCP6 remains in a dimeric form, as demonstrated by the

protein-concentration dependence to unfolding (Figs. S3A & S3B). The $\Delta G°_{conf}$ is 27.6 kcal/mol at pH4, demonstrating the consistently high conformational free energy of the PCP6 dimer over a broad pH range.

In contrast, we were unable to examine the equilibrium folding/unfolding of PCP7 from pH 4.5 to pH7 because protein aggregation resulted in irreversible unfolding. Likewise, a similar result was obtained for PCP-CA from pH 4.5 to pH 6. At higher pH, the data for PCP7 were well described by the three-state equilibrium model described above for pH 7.5, in which the native dimer is in equilibrium with a monomeric intermediate (eq. 2). The conformational free energy($\Delta G°_{conf}$) was 18.4 kcal/mol (figs. S6E-6H), similar to the data described above for pH 7.5. At pH 4, however, the data for PCP7 were best fit by a two-state model as shown in equation 4, where the monomer unfolds to the unfolded monomer. Thus, the data show that the PCP7 dimer is destabilized relative to the monomer so that the ‘‘native’’ protein at pH 4 resembles the monomeric intermediate (I) present at higher pH. The decrease in dimer stability at lower pH, and the presence of the monomer at pH 4, is reflected in a lower conformational free energy, where $\Delta G°_{conf}$ is 5.8 kcal/mol (Figs. S6A-S6D).

Similar to PCP7, the data for PCP-CA at pH 6.5 to pH 8.5 were best described by a three-state equilibrium model in which a monomeric intermediate is in equilibrium with the native dimer and unfolded monomer (eq. 2) (Fig. S7E-7L). The total conformational free energy on unfolding of PCP-CA from pH 6.5 to pH 8.5 was ~20 kcal/mol. The data at pH 4 were the best fit to a three-state model in which the native monomer unfolds through a monomeric intermediate (eq 5), In this case, the $\Delta G°_{conf}$ of unfolding is 5.9 kcal/mol (Figs. S7A-7D). The free energies and m-values obtained from the fits of all

three proteins at all pH are shown in Figs. 5A & 5B. Due to aggregation upon refolding, however, we were unable to determine the precise pH range for the transition of dimer to monomer. Nevertheless, PCP7 and PCP-CA are similar to PCP3 (25) in that the dimer is destabilized at lower pH, such that the protein is a monomer at pH 4. The stability of the monomer is comparable between the three proteins (Table 1).

Together, the data show that the dimer is destabilized for all procaspases at lower pH. However, only PCP6 remains dimeric, while PCP7, PCP3, and PCP-CA are monomers at pH 4. Because the dimer contributes a substantial portion of the conformational free energy, PCP6 retains a high ΔG°_{conf} at all pHs, while the ΔG°_{conf} of PCP7, PCP3, and PCP-CA reflect the stability of the monomer at pH 4. Unlike PCP3, however, the pKas for dimer dissociation of PCP7 and PCP-CA are not known due to protein aggregation between pH 4.5 and 6.5.

The fraction of species versus urea concentration

For each pH, we calculated the equilibrium distribution of species over the urea concentration range of 0 to 8 M using the values of the free energies, the cooperativity indices determined for each transition, and four protein concentrations. (0.5, 1, 2, and 4 μM). The fraction of species corresponding to the global fits are shown in Fig. 6 and Figs. S5-S7. At pH 7.5, there is a cooperative decrease in native protein with a concomitant increase in the partially folded intermediate population between 0 and ~4 M urea. The dimeric (PCP6) or monomeric (PCP7 & PCP-CA) intermediate reaches a maximum at ~3 - 4 M urea and remains predominant up to ~4 – 5 M urea (Figs. 6 & S5 – S7). Furthermore, the unfolded state was fully populated by 6 M urea (PCP7) or 7 M

urea (PCP6 and PCP-CA). At pH 4, the major fraction of PCP6 “native” ensemble is a dimer (Fig. S4A) while the PCP7 (Fig. S6D) and PCP-CA (Fig. S7D) have the monomeric conformation.

Discussion

Effector caspases, caspase-3, -6, and -7, have a significant role in apoptosis and serve overlapping but nonredundant functions (32). Prior to activation, effector caspases exist in the form of zymogen (33). A stable zymogenic form of caspase is believed to be critical in the apoptosis regulation (32). Studying the folding landscape of the caspase homologs with the conserved structure in terms of evolutionary perspective can identify residues important for folding and function. Furthermore, effector caspases are obligate homodimers contrary to monomeric initiator caspases, and an evolutionary folding study could give insight into how dimeric caspases were evolved. A similar study in hemoglobin (Hb) showed that modern heterotetramer Hb evolved from a monomeric ancestor via a homodimer as an intermediate (16). Here, we examined the conservation of the folding landscape by determining the equilibrium folding/unfolding process of the common ancestor of effector caspases. By comparing the folding mechanism and stability (ΔG°_{conf}) between extant caspases and the common ancestor, we show that the landscape is conserved for over 650 million years. The data show that the folding and stability of PCP7, which is evolutionarily more similar to the ancestor, remains comparable to the common ancestor. Indeed, a major change that occurred during the evolution of the effector caspases was to stabilize the dimer. The weak dimer observed

in the common ancestor, with a dissociation constant in the high uM range, became a strong dimer, with a dissociation constant below 50 nM, in the three effector caspases.

We performed urea-induced unfolding monitored by intrinsic fluorescence emission and circular dichroism over a broad pH range (pH 8.5-4, 25 °C). All three proteins show reversible unfolding transitions, but only the folding of PCP6 was reversible over the entire pH range. For PCP7 and PCP-CA, the folding was reversible only at pH 4 and at higher pH, so we have focused on the conditions where folding is reversible. At higher pH, all proteins demonstrated a three-state equilibrium unfolding process. However, the folding intermediate present in PCP6 is a dimer compared to a monomer in PCP7 and PCP-CA. A previous study in the folding landscape of PCP3 showed the four-state unfolding pathway with a dimer and a monomeric intermediate. So, we suggest that the two intermediates are most likely present in the ancestor, but only the monomeric intermediate is sufficiently stable to appear in our spectroscopic assays. In addition, the intermediates may have been preferentially stabilized during evolution, such that PCP6 stabilized the dimeric intermediate while PCP7 retained the monomer as the most stable intermediate. In PCP3, both the partially folded monomer and dimer were stabilized sufficiently to populate both species. At this point, however, it is not clear when the changes occurred in the evolution of each subfamily and how these changes affect their functions. Further experiments with ancestral proteins in each subfamily and extant caspases from species other than human will be important to refine the evolutionary changes.

Change in conformational free energy ($\Delta G°_{conf}$) was generated by global fitting of the denaturation curves (31, 34). At pH 7.5, PCP6 is the most stable caspase ($\Delta G°_{conf}$ =

32.8 ± 1.7 kcal/mol), while PCP3 (24) and PCP-CA are notably less stable ($\Delta G°_{conf}$ = 26.1 ± 2.1 and 21.8 ± 0.5 kcal/mol, respectively) and PCP7 is the least stable ($\Delta G°_{conf}$ = 15.5 ± 0.2 kcal/mol). The denaturation graphs also show that the dimer of PCP6 is more stable than that of PCP7 or of PCP-CA, where the mid-point of dimer dissociation was approximately 5 M urea for PCP6 *versus* 2 M urea for PCP7 and PCP-CA. Although the folding landscape of caspases was evolutionarily conserved, stability differs among the extant caspases, which could be due to subsequent mutations along the evolution. Notably, change in salt bridges in the native structure of protein due to substitution have been reported as a key difference in the stability of proteins (35, 36). We compared the number of salt bridges among extant caspases using a protein tool (https://proteintools.uni-bayreuth.de/) developed by Höcker lab. It showed caspase-3, -6 and -7 consist 51 (PDB:1CP3), 40 (PDB:3S70) and 20 (PDB:1F1J) salt bridges, respectively, in their entire dimeric structure. We also counted salt bridges in available procaspase structures of all three proteins and didn't find significant differences from active caspases. It suggests that PCP7 being least stable among effector caspases could be due to the fewer salt bridges formed within the structure. The higher stability of PCP6 despite fewer salt bridges than PCP3 could be due to the reported fraction of tetramers in the native state (33). As some studies noted the trade-off between stability and activity (37, 38), we speculate that higher stability of PCP6 could play a regulatory function in apoptosis, allowing caspase-3 to be the main executioner caspase.

Total free energy change ($\Delta G°_{conf}$) of PCP6 seems to be least affected by pH as the lowest is 27.6 kcal/mol at pH 4, and the highest is 34.5 kcal/mol at pH 8. In contrast, $\Delta G°_{conf}$ of PCP7 and PCP-CA decreased significantly from high pH to low pH, similar to

the PCP3 (25). In our data, PCP7 is most stable at pH 8 with 18.4 kcal/mol and least stable at pH 4 with 5.8 kcal/mol free energy changes whereas, PCP-CA is most stable at pH 7.5 with 21.9 kcal/mol and least stable at pH 4 with 6 kcal/mol free energy changes. At pH 4, significant fraction of PCP6 remains dimeric while PCP7 and PCP-CA dissociated almost exclusively to the monomeric form as previously reported in PCP3, which is reflected well in their substantial decrease in stability with free energy ~5 kcal/mol (25). Since PCP7 and PCP-CA were unfolded irreversibly around pH 5 and 6, it is not clear to us at what pH proteins dissociate to monomers. PCP6 dimer is reasonably stable with 27 kcal/mol free energy change. This difference suggests that caspase-6 could be the main executioner caspase in a low pH environment instead of caspase-3. Caspase-6 having more specialized roles in specific physiological contexts than caspase-3 and -7 would support the idea of caspase-6 being a major effector at low pH. In addition, it is reported that execution of cell death at pH 5 in neurons – where caspase-6 plays a major role – during extracellular acidosis (39). Perhaps, close analysis of the difference in the dimeric interface among effector caspases could answer the reason behind the higher stability of PCP6 dimer at low pH in amino acid level.

Global fitting also calculated cooperative index (m-value), which relates to the accessible surface area (ΔASA) of a protein exposed to solvent when it unfolds (40). At pH 7.5 total m-value (m_{total}) of PCP6, PCP7 and PCP-CA are 5.5, 2.5, and 3.2 kcal mol^{-1} M^{-1} respectively. According to the equation developed by Scholtz and co-workers (40) to show the empirical relationships that correlate ΔASA with experimental m- values, as shown in eq 6:

$$m = 243 + 0.13(\Delta ASA) \quad [6]$$

The higher m_{total} of PCP6 illustrated that it exposed a larger surface area after complete unfolding, followed by PCP-CA and PCP7. Since all these proteins have similar residue numbers, m_{total} suggests that the PCP6 structure is more compact than the other two proteins in its native state. The m_{total} of PCP6 is similar between pH 8.5 and pH 6 (~5.5 kcal mol^{-1} M^{-1}) and then slightly decreases between pH 5.5 and pH 4 (~4 kcal mol^{-1} M^{-1}), even though m_2 remains constant throughout the pH range. In contrast, m_{total} of PCP7 and PCP-CA reduced significantly from ~3 kcal mol^{-1} M^{-1} to ~1 kcal mol^{-1} M^{-1}. The decrease in m-value at lower pH in PCP6 is due to loss of burial of hydrophobic surface area. Besides, the dimer's dissociation caused a significant loss of m-value in the case of PCP7 and PCP-CA (25).

In summary, the folding landscape of the effector caspases was established in the common ancestor (>650 Mya), and extant caspases fold differently and have different stability maintaining the caspase hemoglobinase-fold. The caspase family highlights how sequence changes and evolutionary processes can drive the folding landscape. Hence, the stability of the native enzyme and their response to changes in the environment can be fine-tuned in a species-specific manner. Moreover, knowledge of folding mechanisms among modern proteins and their ancestral state implies how the energy landscape is maintained and how oligomerization played a role in the evolution of modern-day homologs to serve a potential regulatory purpose. Also, elucidation of protein conformational changes and stability is a key to understanding protein function. Thermodynamic study in different environmental conditions can also provide useful information regarding the effect of solvent and various solutes on proteins' stability.

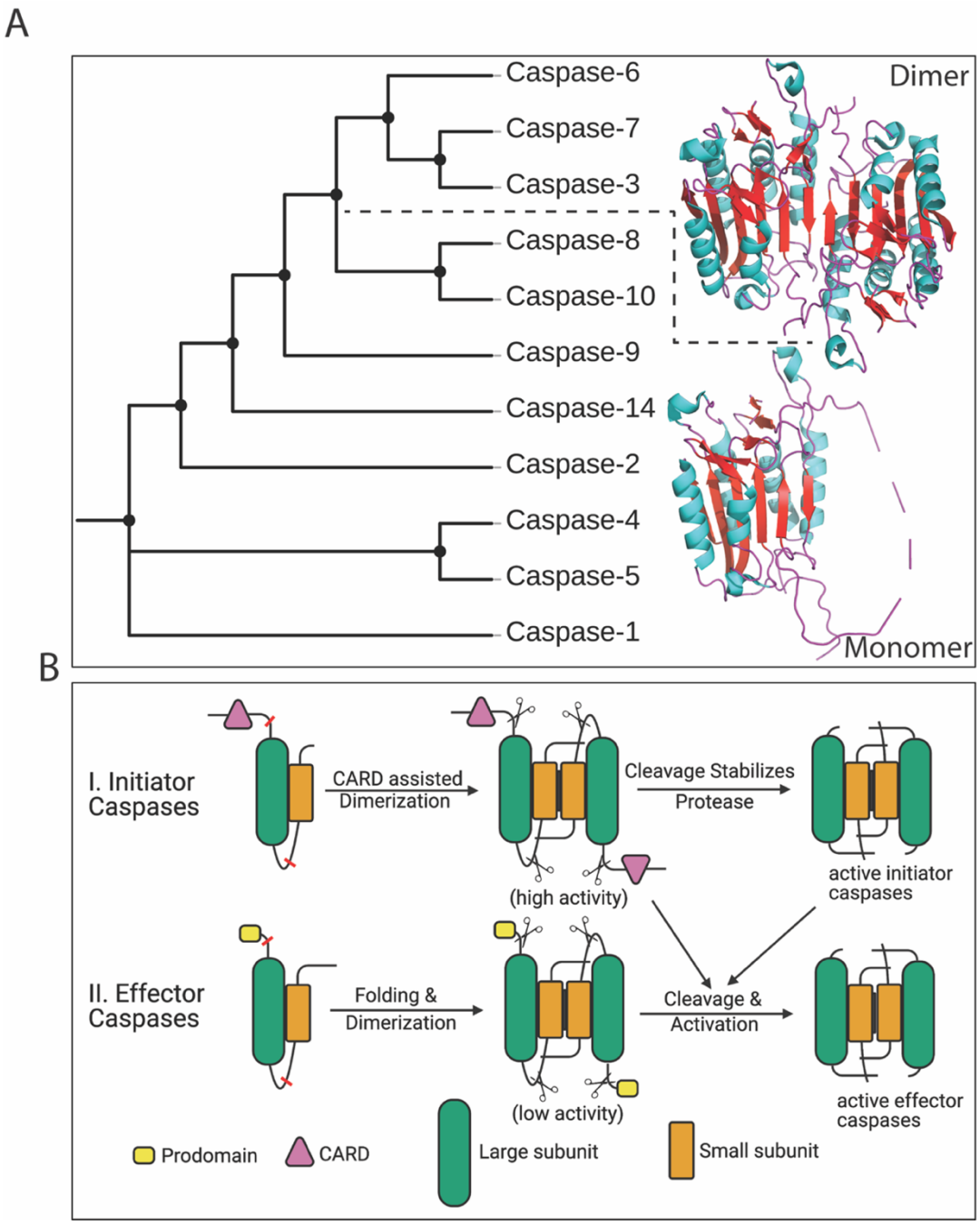


Figure 1: Phylogenetic relationship and activation mechanism of apoptotic caspases. (**A**) All caspases evolved from a common ancestor then inflammatory and initiator caspases evolved as a monomer and effectors as a dimer. (**B**) Initiator caspases are stable monomer and dimerization is enough for activation while effector caspases are stable dimer and activated by cleavage of an intersubunit linker by initiator caspases. Red lines represent cleavage sites and scissors represents cleavage.

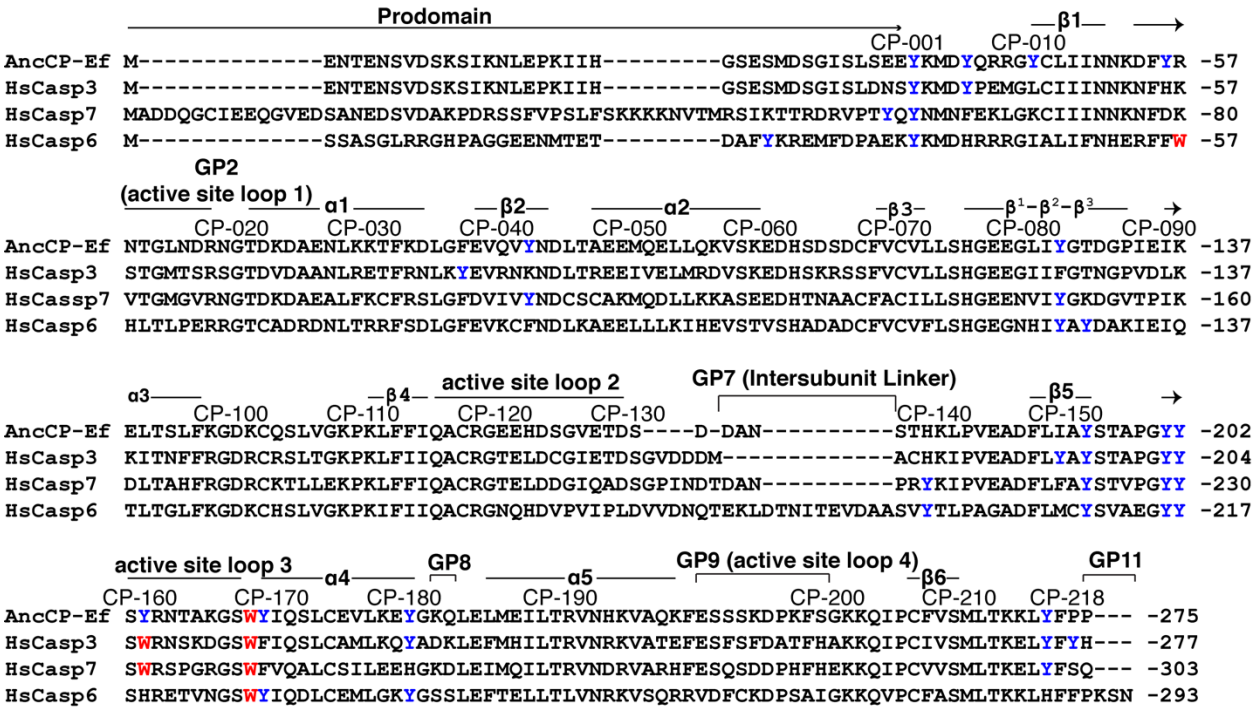


Figure 2: Multiple sequence alignment of human effector caspases with a common ancestor. CP refers to common position number of caspases, individual sequence number is indicated at the right side of sequences. Also, secondary structural elements are indicated, and tyrosine residues (blue) and tryptophan residues (red) are highlighted.

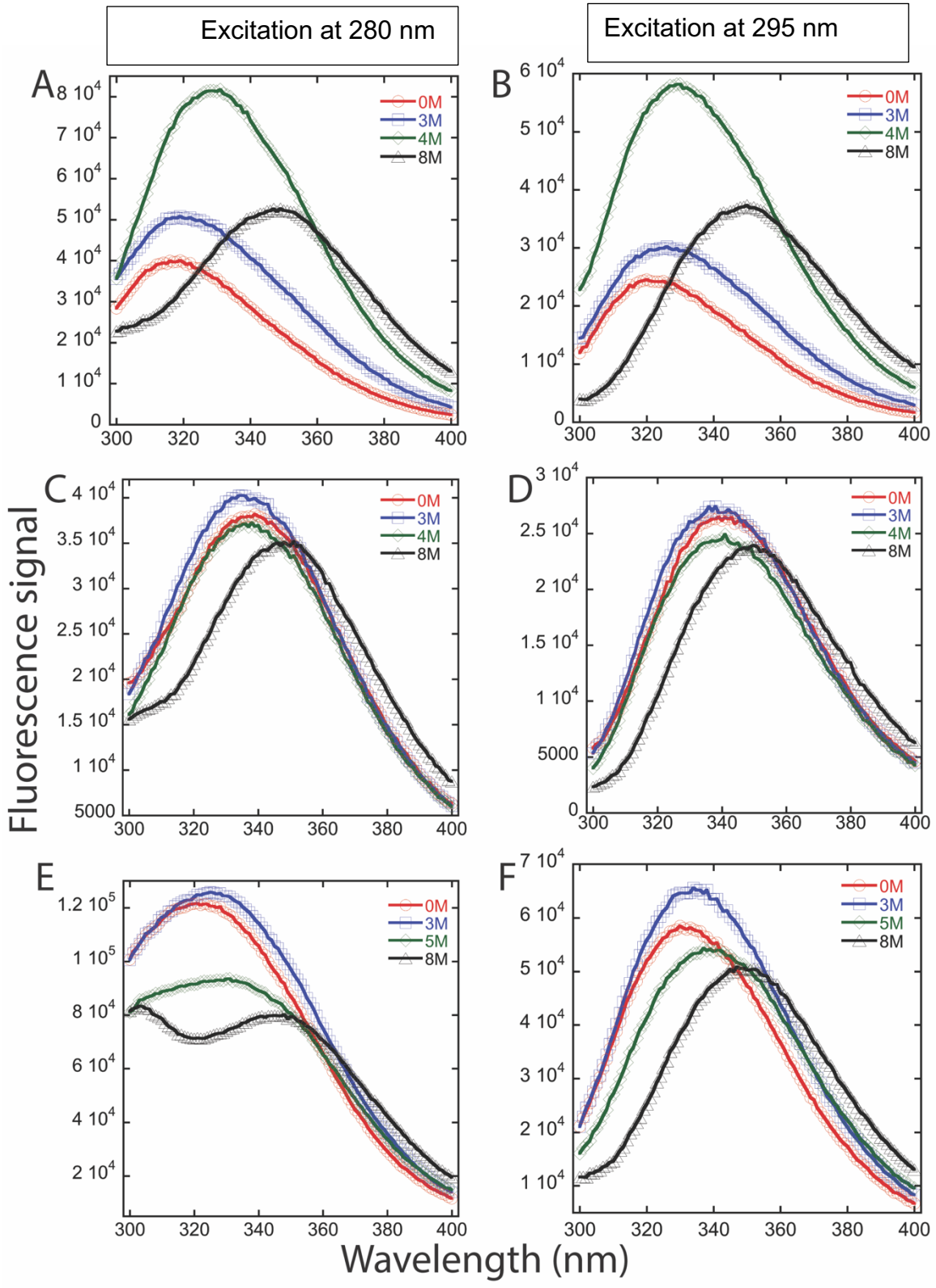


Figure 3: Fluorescence emission spectra of effector caspases following excitation at 280 nm (left) and 295 nm (right). Emission spectra of 2 μM, PCP6 (**A, B**), PCP7 (**C, D**), and AncCP-Ef (**E, F**) in 20 mm phosphate buffer, pH 7.5 containing urea (displayed in inset of each graph).

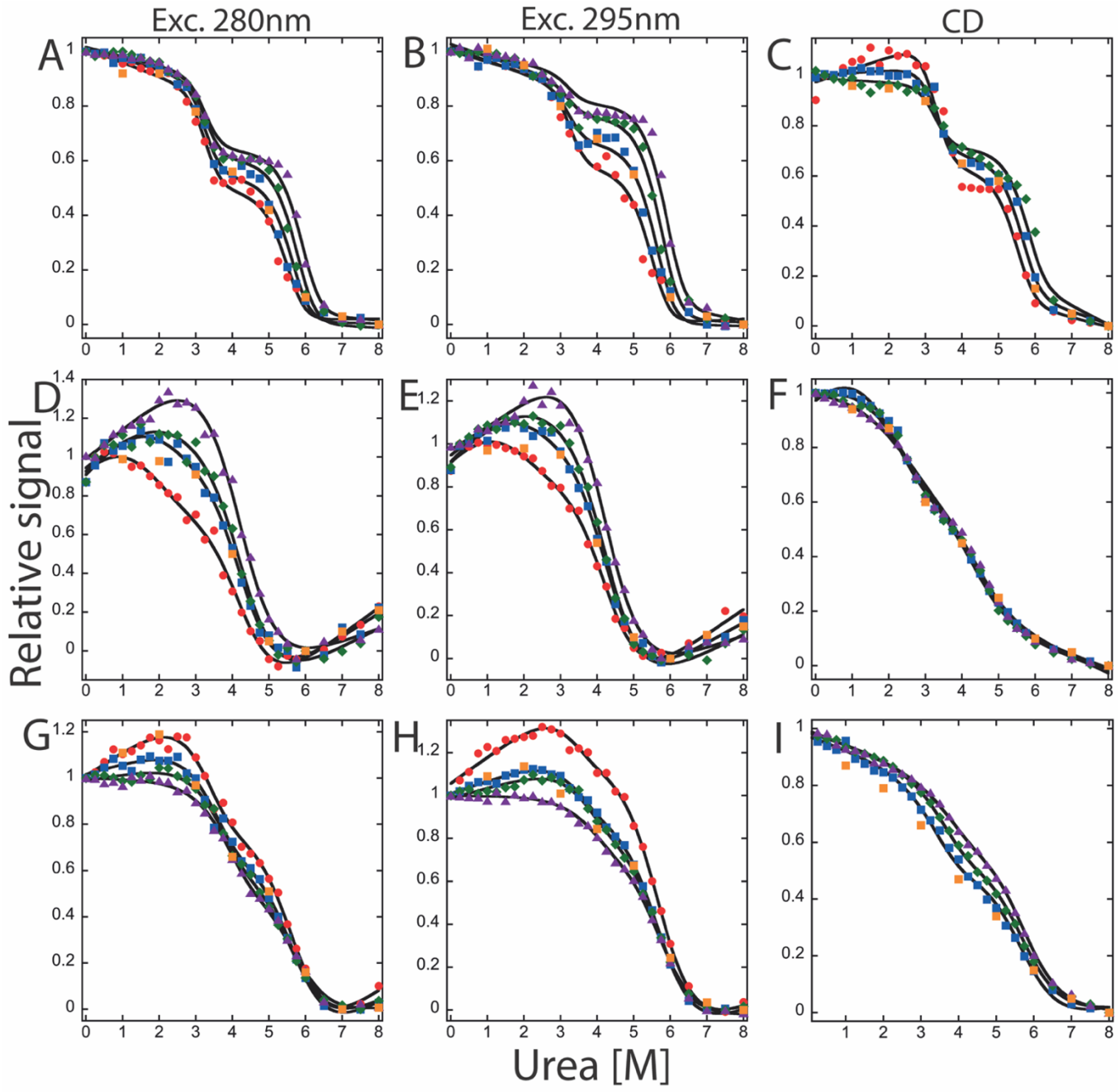


Figure 3: Equilibrium unfolding of effector caspases at pH 7.5. Equilibrium unfolding of PCP6 (**A, B, C**), PCP7 (**D, E, F**), and PCP-CA (**G, H, I**) monitored by fluorescence emission with excitation at 280 nm (left column), 295 nm (middle column) and circular dichroism (right column). Four different concentration of proteins were used is fluorescence emission and three different concentration of proteins were used in circular dichroism. Colored solid symbols represent raw data and corresponding solid lines represent the global fits of the data in an appropriate model described in text. Symbols are represented as follows. 0.5 μM (●), 1 μM (■), 2 μM (◆), and 4 μM (▲). Orange square (■) represents refolding data of 1uM protein.

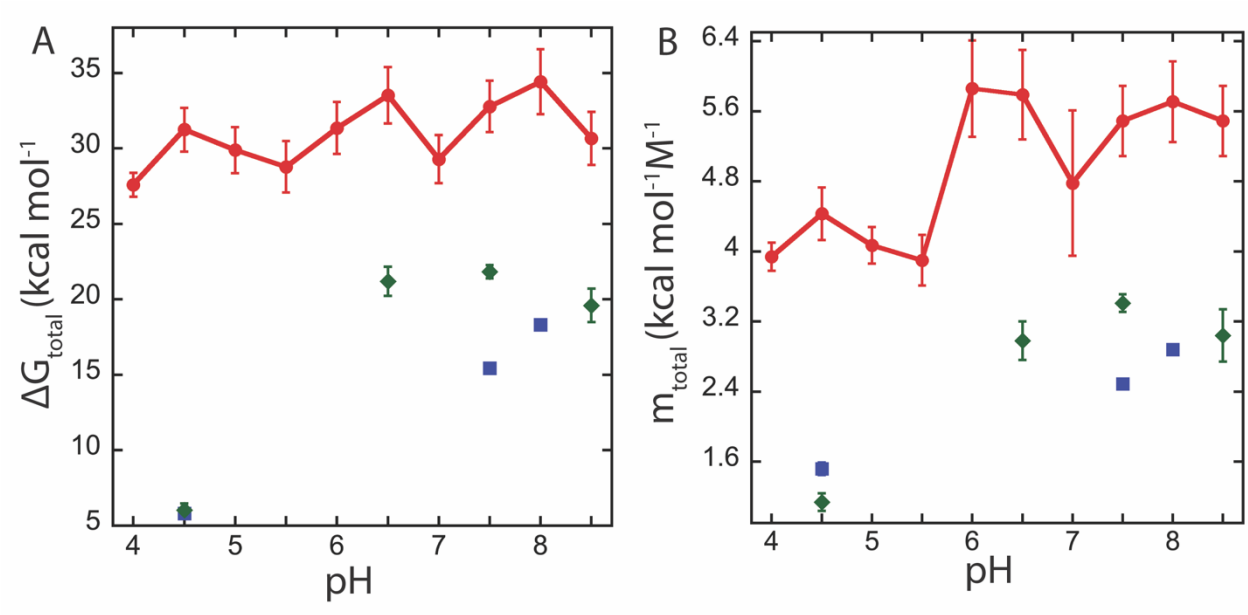


Figure 5: Comparison of total free energy (ΔG_{total}) (**A**), and total m-value (m_{total}) (**B**) changes as a function of pH between PCP6 (●), PCP7 (■) and PCP-CA (◆). Error bars represent standard deviation of respective parameters given by the global fitting.

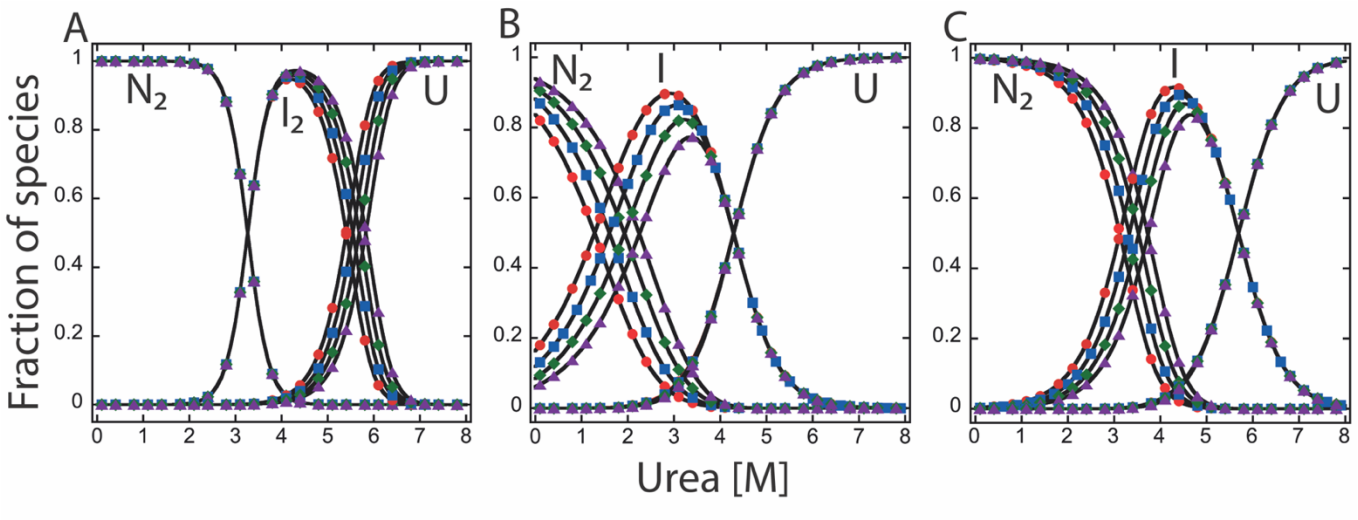


Figure 6: Fraction of species as a function of urea concentration at pH 7.5. Fraction of species of PCP6 (**A**), PCP7 (**B**), and AncCP-Ef (**C**). The fractions of native, intermediate, and unfolded protein were calculated as a function of urea concentration from fits of the data shown in figure 4. Protein concentrations symbols and colors are consistent with figure 4. N_2 refers to dimeric native protein, I_2 and I are dimeric and monomeric intermediates, and U refers to unfolded species.

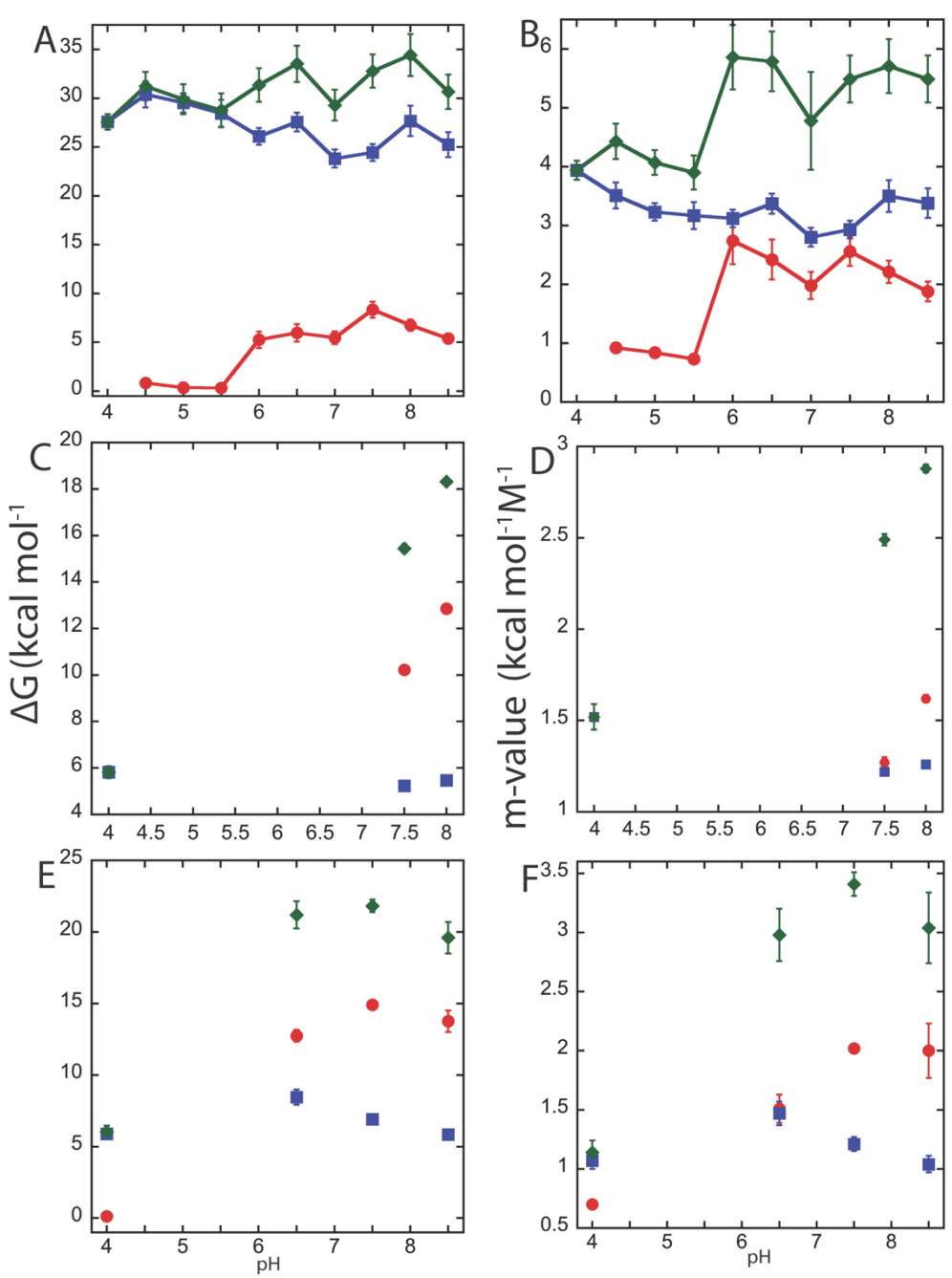


Figure 7: Free energy and m-value changes in each step of unfolding. The left column of plots shows free energy changes at each unfolding step of PCP 6 (**A**), PCP7 (**C**), and AncCP-Ef (**E**) as a function of pH. The right side of plots of m-value changes at each step of unfolding of PCP6 (**B**), PCP7(**D**), and AncCP-Ef (**F**) as a function of pH. Solid red circles (●) represent ΔG_1 and m_1 in respective graphs in corresponding pH. Similarly, solid blue squares (■) represent ΔG_2 and m_2, and solid green diamonds (◆) represent total change in free energy (ΔG_{total}) and total change in m-value (m_{total}). Error bars represent standard deviation of respective parameters given by the global fitting.

Table 1: Thermodynamic parameters of each step of folding/unfolding of extant and ancestral effector caspases. PCP3 data is taken from Bose and Clark, 2001 for the comparison.

Proteins	Equilibrium mechanism	Free energy changes ($\Delta G°_{conf}$) (kcal/mol)	Cooperativity index (m-value) (kcal/mol/M)	Total $\Delta G°_{conf}$ (kcal/mol)	m_{total} (kcal/mol/M)
PCP3	$N_2 \rightleftharpoons I_2$	8.3 ± 1.3	2.8 ± 0.5	25.8 ± 2.8	4.5 ± 0.7
	$I_2 \rightleftharpoons 2I$	10.5 ± 1.0	0.5 ± 0.1		
	$2I \rightleftharpoons 2U$	7.0 ± 0.5	1.2 ± 0.1		
PCP6	$N_2 \rightleftharpoons I_2$	8.4 ± 0.8	2.6 ± 0.3	32.8 ± 1.5	5.5 ± 0.5
	$I_2 \rightleftharpoons 2U$	24.4 ± 0.9	2.9 ± 0.2		
PCP7	$N_2 \rightleftharpoons 2I$	10.2 ± 0.2	1.3 ± 0.1	15.4 ± 0.3	2.5 ± 0.2
	$2I \rightleftharpoons 2U$	5.2 ± 0.1	1.2 ± 0.1		
PCP-CA	$N_2 \rightleftharpoons 2I$	15 ± 0.1	2 ± 0.1	21.9 ± 1.3	3.2 ± 0.2
	$2I \rightleftharpoons 2U$	6.9 ± 0.3	1.2 ± 0.1		

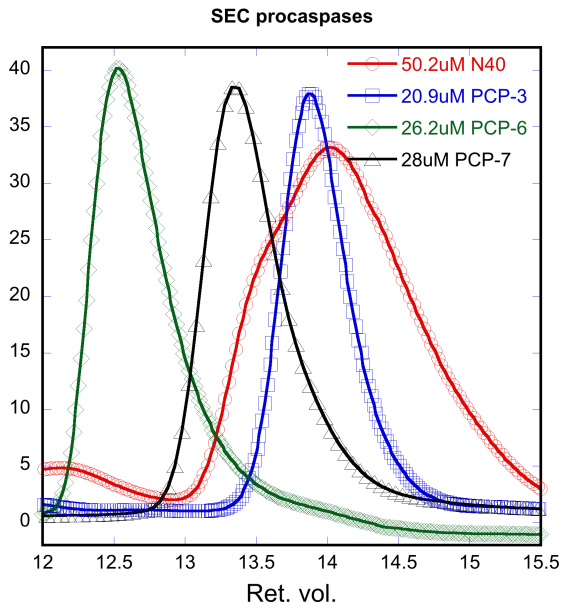
References

1. Doyle, C. M., Rumfeldt, J. A., Broom, H. R., Broom, A., Stathopulos, P. B., Vassall, K. A., Almey, J. J., and Meiering, E. M. (2013) Energetics of oligomeric protein folding and association. *Arch. Biochem. Biophys.* **531**, 44–64
2. Rumfeldt, J. A. O., Galvagnion, C., Vassall, K. A., and Meiering, E. M. (2008) Conformational stability and folding mechanisms of dimeric proteins. *Prog. Biophys. Mol. Biol.* **98**, 61–84
3. Röder, K., Joseph, J. A., Husic, B. E., and Wales, D. J. (2019) Energy Landscapes for Proteins: From Single Funnels to Multifunctional Systems. *Adv. Theory Simulations*. **2**, 1800175
4. Milam, S. L., and Clark, A. C. (2009) Folding and assembly kinetics of procaspase-3. *Protein Sci.* **18**, 2500–2517
5. Yao, J., and Wang, J. (2015) Neither two-state nor three-state: Dimerization of lambda cro repressor. *J. Phys. Chem. Lett.* **6**, 2022–2026
6. Zheng, W., Schafer, N. P., Davtyan, A., Papoian, G. A., and Wolynes, P. G. (2012) Predictive energy landscapes for protein-protein association. *Proc. Natl. Acad. Sci. U. S. A.* **109**, 19244–19249
7. Han, J. H., Batey, S., Nickson, A. A., Teichmann, S. A., and Clarke, J. (2007) The folding and evolution of multidomain proteins. *Nat. Rev. Mol. Cell Biol.* 10.1038/nrm2144
8. Lynch, M. (2012) The evolution of multimeric protein assemblages. *Mol. Biol. Evol.* **29**, 1353–1366
9. Lynch, M. (2013) Evolutionary diversification of the multimeric states of proteins. *Proc. Natl. Acad. Sci. U. S. A.* **110**, E2821–E2828
10. Hashimoto, K., and Panchenko, A. R. (2010) Mechanisms of protein oligomerization, the critical role of insertions and deletions in maintaining different oligomeric states. *Proc. Natl. Acad. Sci. U. S. A.* **107**, 20352–20357
11. Clark, A. C. (2016) Caspase Allostery and Conformational Selection. *Chem. Rev.* **116**, 6666–6706
12. MacKenzie, S. H., and Clark, A. C. (2012) Death by caspase dimerization. *Adv. Exp. Med. Biol.* **747**, 55–73
13. Grinshpon, R. D., Shrestha, S., Titus-McQuillan, J., Hamilton, P. T., Swartz, P. D., and Clark, A. C. (2019) Resurrection of ancestral effector caspases identifies novel networks for evolution of substrate specificity. *Biochem. J.* **476**, 3475–3492
14. Lim, S. A., and Marqusee, S. (2018) The burst-phase folding intermediate of ribonuclease H changes conformation over evolutionary history. *Biopolymers*. 10.1002/bip.23086
15. Lim, S. A., Hart, K. M., Harms, M. J., and Marqusee, S. (2016) Evolutionary trend toward kinetic stability in the folding trajectory of RNases H. *Proc. Natl. Acad. Sci.* **113**, 13045–13050
16. Pillai, A. S., Chandler, S. A., Liu, Y., Signore, A. V., Cortez-Romero, C. R., Benesch, J. L. P., Laganowsky, A., Storz, J. F., Hochberg, G. K. A., and Thornton, J. W. (2020) Origin of complexity in haemoglobin evolution. *Nature*. **581**, 480–485
17. Harms, M. J., and Thornton, J. W. (2013) Evolutionary biochemistry: Revealing the historical and physical causes of protein properties. *Nat. Rev. Genet.* **14**, 559–

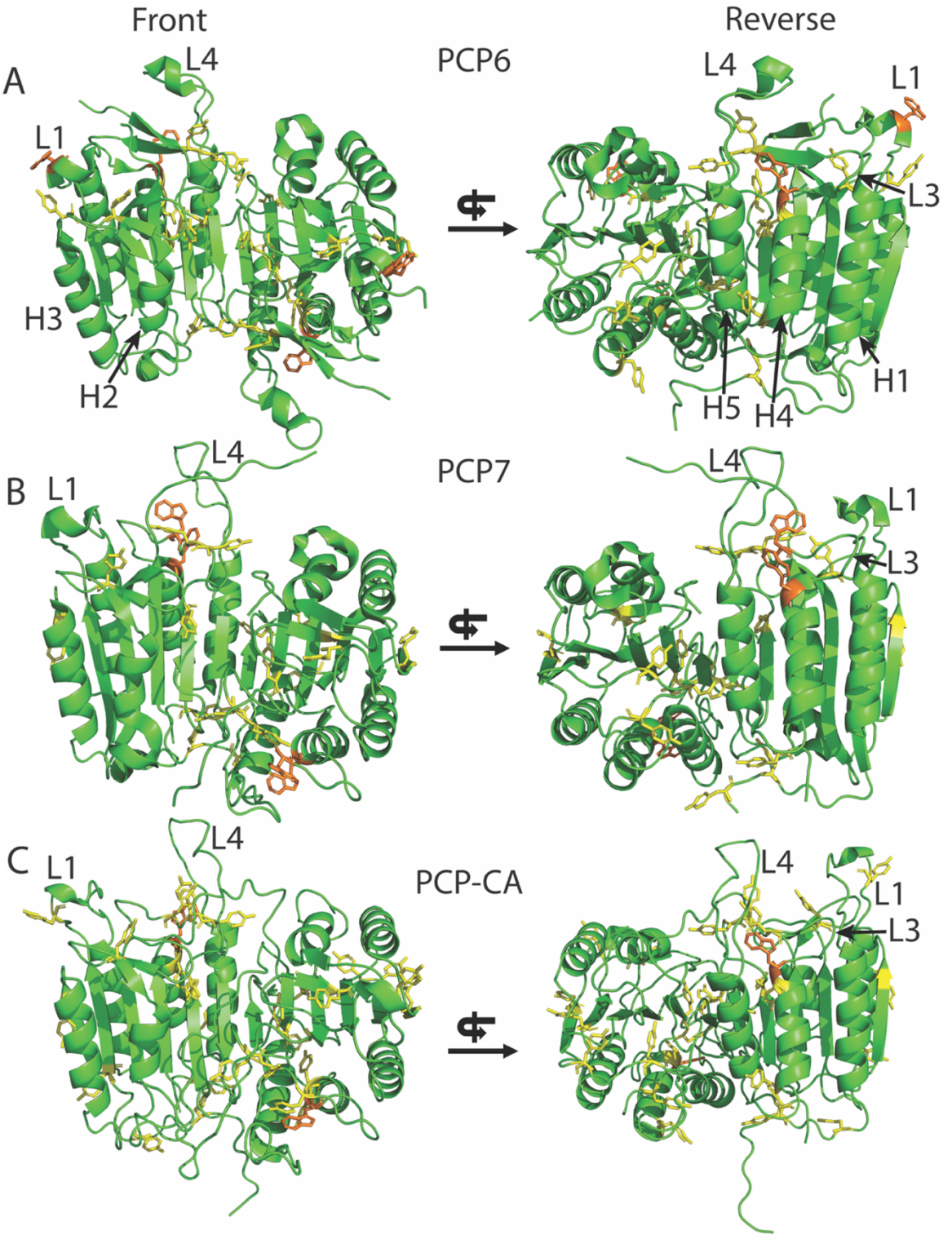
571
18. Yan, N., and Shi, Y. (2005) Mechanisms of apoptosis through structural biology. *Annu. Rev. Cell Dev. Biol.* 10.1146/annurev.cellbio.21.012704.131040
19. Tucker, M. B., MacKenzie, S. H., Maciag, J. J., Dirscherl Ackerman, H., Swartz, P., Yoder, J. A., Hamilton, P. T., and Clay Clark, A. (2016) Phage display and structural studies reveal plasticity in substrate specificity of caspase-3a from zebrafish. *Protein Sci.* **25**, 2076–2088
20. Feeney, B., Pop, C., Swartz, P., Mattos, C., and Clark, A. C. (2006) Role of loop bundle hydrogen bonds in the maturation and activity of (Pro)caspase-3. *Biochemistry*. **45**, 13249–13263
21. Huang, W., Jiang, T., Choi, W., Qi, S., Pang, Y., Hu, Q., Xu, Y., Gong, X., Jeffrey, P. D., Wang, J., and Shi, Y. (2013) Mechanistic insights into CED-4-mediated activation of CED-3. *Genes Dev.* **27**, 2039–2048
22. Yan, N., Huh, J. R., Schirf, V., Demeler, B., Hay, B. A., and Shi, Y. (2006) Structure and activation mechanism of the Drosophila initiator caspase Dronc. *J. Biol. Chem.* **281**, 8667–8674
23. Shrestha, S., Tung, J., Grinshpon, R. D., Swartz, P., Hamilton, P. T., Dimos, B., Mydlarz, L., and Clark, A. C. (2020) Caspases from scleractinian coral show unique regulatory features. *J. Biol. Chem.* **295**, 14578–14591
24. Bose, K., and Clark, A. C. (2001) Dimeric procaspase-3 unfolds via a four-state equilibrium process. *Biochemistry*. **40**, 14236–14242
25. Bose, K., and Clark, a C. (2005) pH effects on the stability and dimerization of procaspase-3. *Protein Sci.* **14**, 24–36
26. Merkl, R., and Sterner, R. (2016) Reconstruction of ancestral enzymes. *Perspect. Sci.* 10.1016/j.pisc.2016.08.002
27. Pop, C., Chen, Y. R., Smith, B., Bose, K., Bobay, B., Tripathy, A., Franzen, S., and Clark, A. C. (2001) Removal of the pro-domain does not affect the conformation of the procaspase-3 dimer. *Biochemistry*. **40**, 14224–14235
28. Bose, K., Pop, C., Feeney, B., and Clark, A. C. (2003) An uncleavable procaspase-3 mutant has a lower catalytic efficiency but an active site similar to that of mature caspase-3. *Biochemistry*. **42**, 12298–12310
29. MacKenzie, S. H., Schipper, J. L., England, E. J., Thomas, M. E., Blackburn, K., Swartz, P., and Clark, A. C. (2013) Lengthening the intersubunit linker of procaspase 3 leads to constitutive activation. *Biochemistry*. **52**, 6219–6231
30. Roschitzki-Voser, H., Schroeder, T., Lenherr, E. D., Frölich, F., Schweizer, A., Donepudi, M., Ganesan, R., Mittl, P. R. E., Baici, A., and Grütter, M. G. (2012) Human caspases in vitro: Expression, purification and kinetic characterization. *Protein Expr. Purif.* 10.1016/j.pep.2012.05.009
31. Walters, J., Milam, S. L., and Clark, A. C. (2009) *Chapter 1 Practical Approaches to Protein Folding and Assembly. Spectroscopic Strategies in Thermodynamics and Kinetics*, 1st Ed., Elsevier Inc., 10.1016/S0076-6879(08)04201-8
32. Thomsen, N. D., Koerber, J. T., and Wells, J. A. (2013) Structural snapshots reveal distinct mechanisms of procaspase-3 and -7 activation. *Proc. Natl. Acad. Sci. U. S. A.* **110**, 8477–8482
33. Stanger, K., Steffek, M., Zhou, L., Pozniak, C. D., Quan, C., Franke, Y., Tom, J., Tam, C., Elliott, J. M., Lewcock, J. W., Zhang, Y., Murray, J., and Hannoush, R.

N. (2012) Allosteric peptides bind a caspase zymogen and mediate caspase tetramerization. *Nat. Chem. Biol.* **8**, 655–660
34. Myers, J. K., Nick Pace, C., and Martin Scholtz, J. (1995) Denaturant m values and heat capacity changes: Relation to changes in accessible surface areas of protein unfolding. *Protein Sci.* **4**, 2138–2148
35. Basak, S., Paul Nobrega, R., Tavella, D., Deveau, L. M., Koga, N., Tatsumi-Koga, R., Baker, D., Massi, F., and Robert Matthews, C. (2019) Networks of electrostatic and hydrophobic interactions modulate the complex folding free energy surface of a designed βα protein. *Proc. Natl. Acad. Sci. U. S. A.* **116**, 6806–6811
36. Kleiner, D., Shmulevich, F., Zarivach, R., Shahar, A., Sharon, M., Ben-Nissan, G., and Bershtein, S. (2019) The interdimeric interface controls function and stability of Ureaplasma urealiticum methionine S-adenosyltransferase. *J. Mol. Biol.* **431**, 4796–4816
37. Cipolla, A., Delbrassine, F., Da Lage, J. L., and Feller, G. (2012) Temperature adaptations in psychrophilic, mesophilic and thermophilic chloride-dependent alpha-amylases. *Biochimie*. **94**, 1943–1950
38. Miller, S. R. (2017) An appraisal of the enzyme stability-activity trade-off. *Evolution (N. Y).* **71**, 1876–1887
39. Yermolaieva, O., Leonard, A. S., Schnizler, M. K., Abboud, F. M., and Welsh, M. J. (2004) Extracellular acidosis increases neuronal cell calcium by activating acid-sensing ion channel 1a. *Proc. Natl. Acad. Sci. U. S. A.* **101**, 6752–6757
40. Scholtz, J. M., Grimsley, G. R., and Pace, C. N. (2009) Chapter 23 Solvent Denaturation of Proteins and Interpretations of the m Value. in *Methods in enzymology*, pp. 549–565, **466**, 549–565

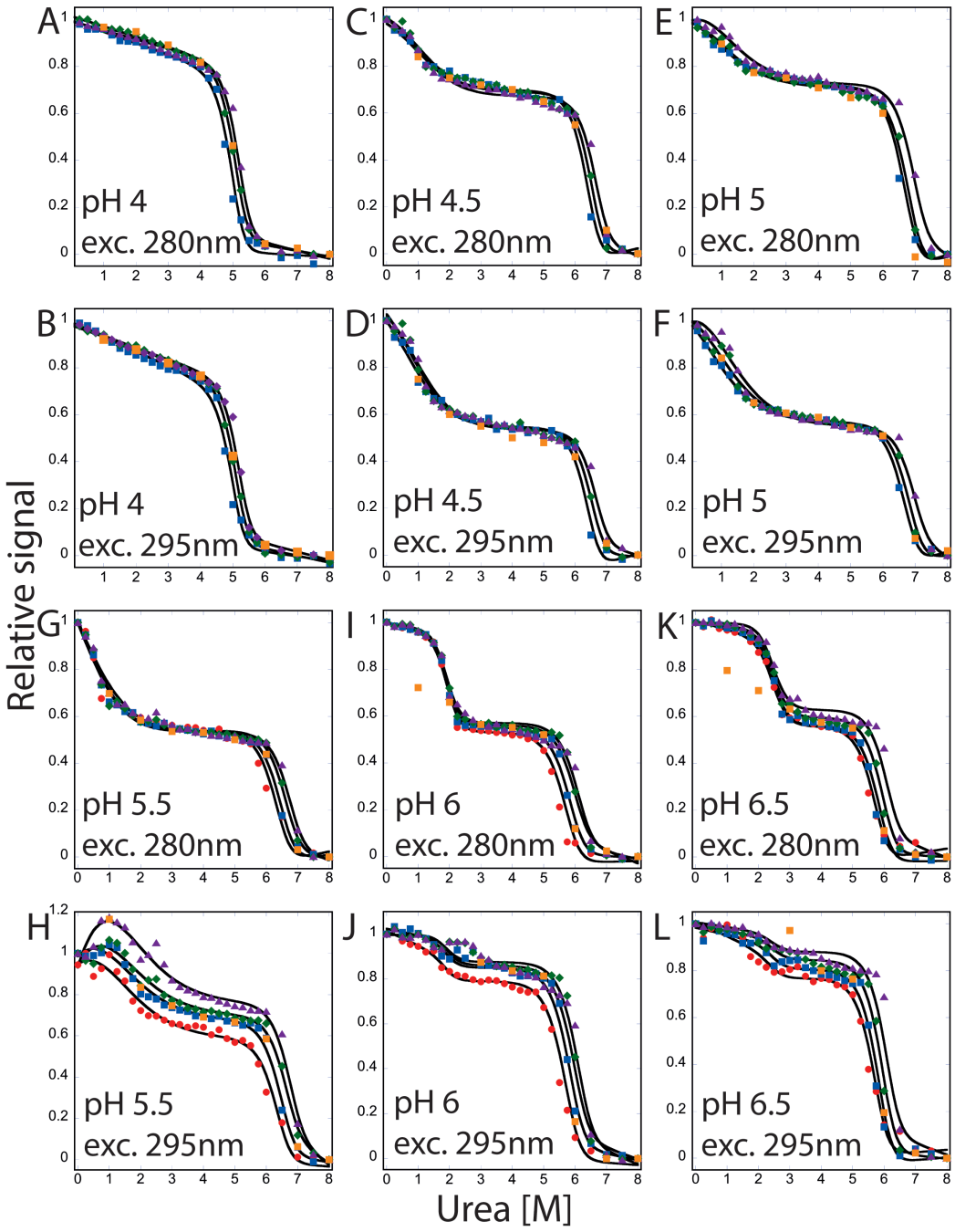
Proteins	Theoretical MW (kDa)	Exp. MW (kDa)
Caspase-3	63.2	63.1
Caspase-6	66.6	80.8
Caspase-7	68.6	70.1
AncCP-Ef	62.4	61.3



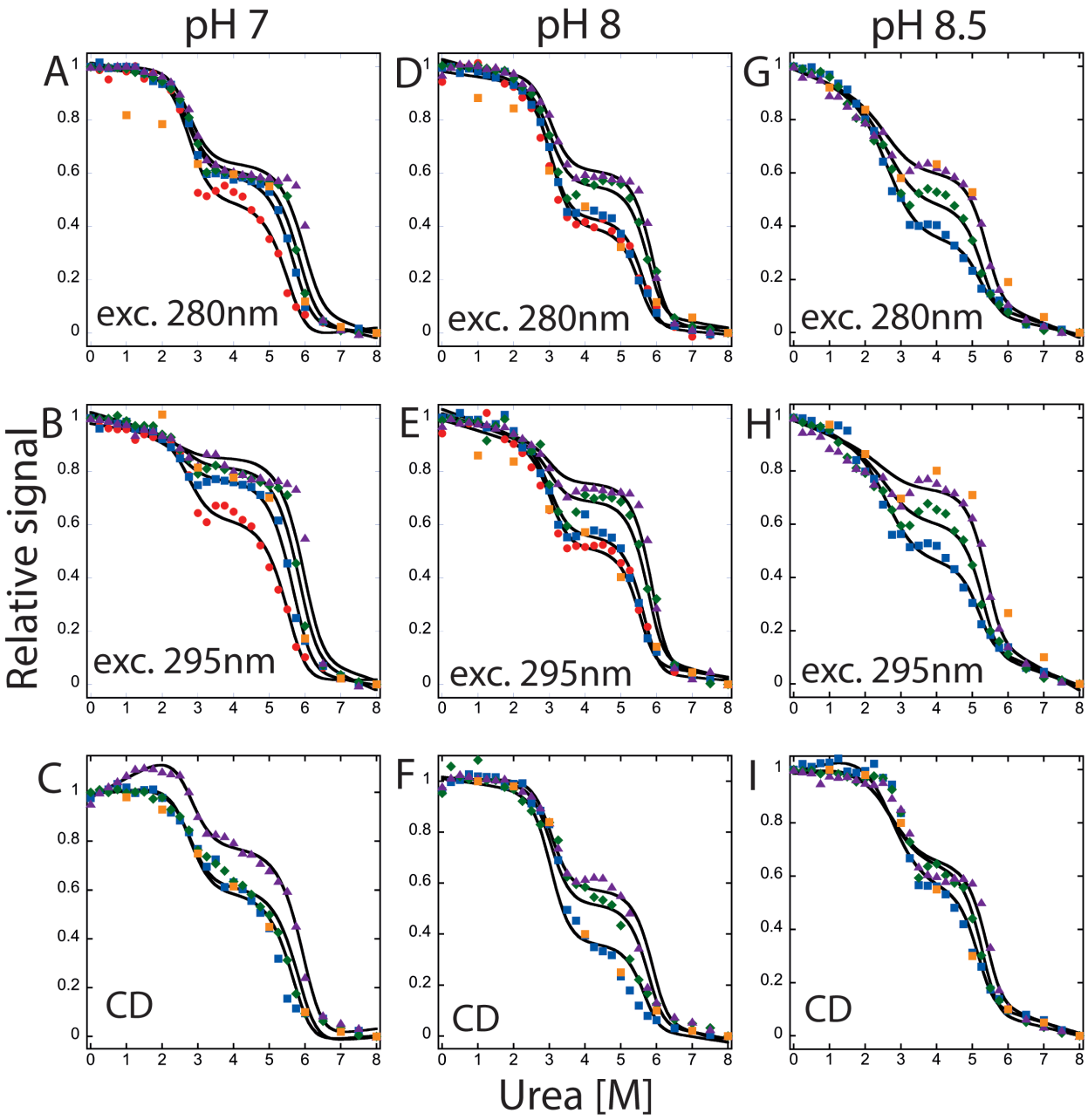
Supplementary Figure 1: Theoretical and experimental molecular weight of procaspase-3, -6, -7 and inactive AncCP-Ef2



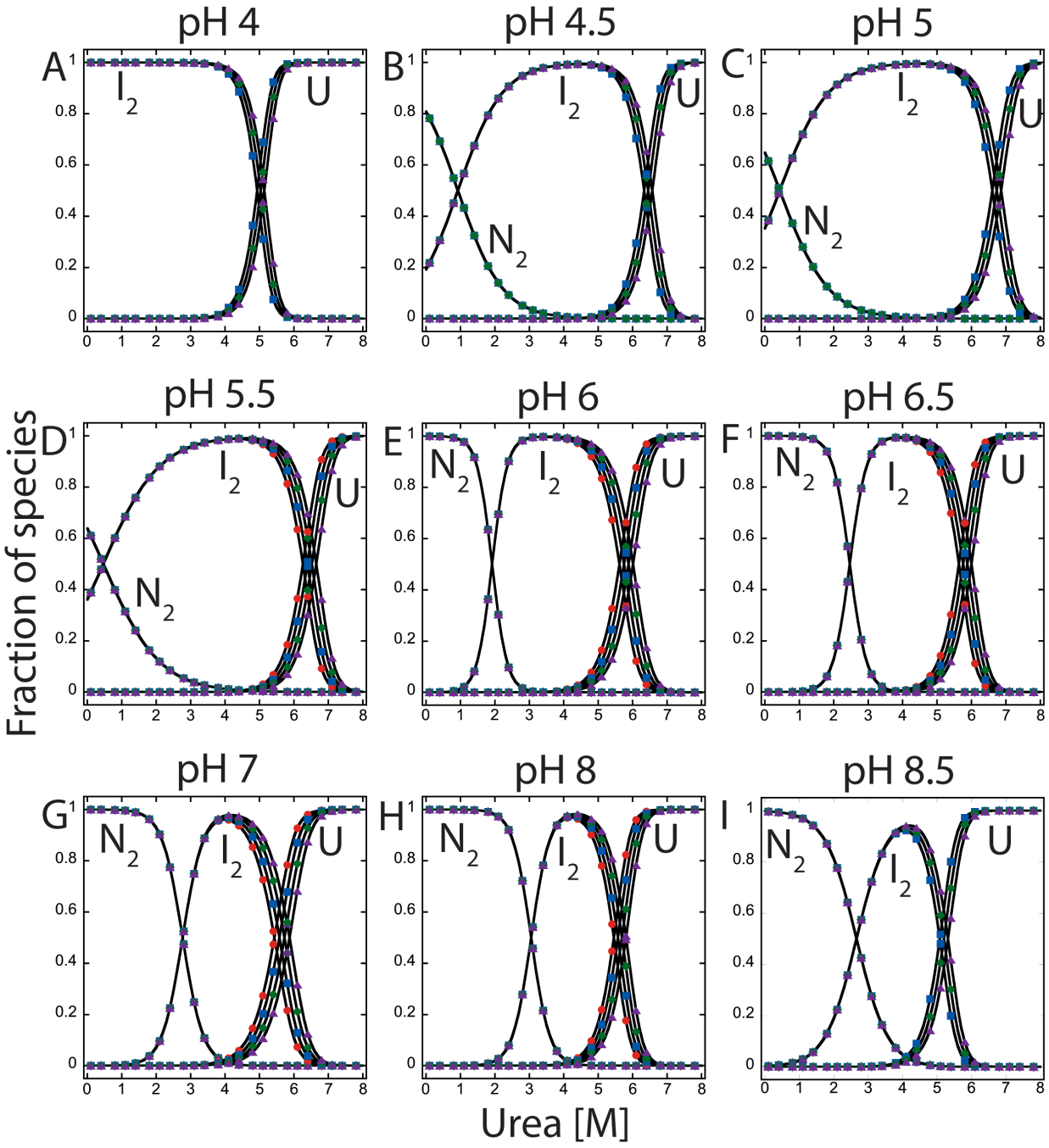
Supplemental Figure 2: Tryptophan residues (yellow sticks) and tryptophan residues (orange sticks) are shown in PCP6 (PDB:4nbk) (**A**), PCP7 (PDB:1K86) (**B**) and PCP-CA (model structure) (**C**).



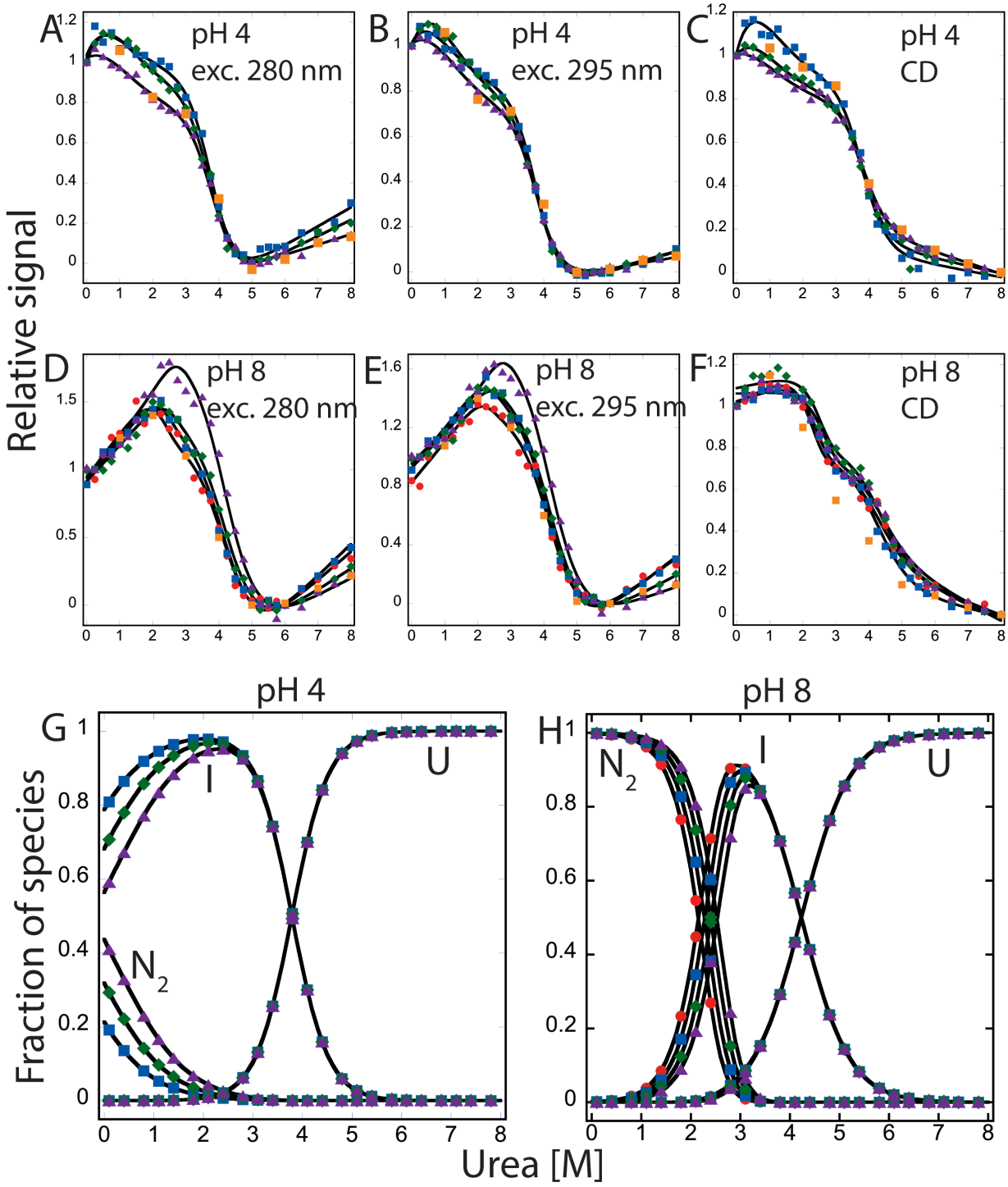
Supplementary Figure 3: Equilibrium unfolding by fluorescence emission of PCP6 from pH 4 to 6.5. Relative signal of fluorescence emission with excitation at 280 nm and 295 nm at pH 4 (**A, B**), pH 4.5 (**C, D**), pH 5 (**E, F**), pH 5.5 (**G, H**), pH 6 (**I, J**), and pH 6.5 (**K, L**). Three different concentrations of proteins were used from pH 4, 4.5 and 5, and four different concentration of proteins were used from pH 5.5, 6 and 6.5. Colored solid symbols represent raw data and corresponding solid lines represent the global fits of the data in an appropriate model described in text. Symbols are represented as follows. 0.5 μM (●), 1 μM (■), 2 μM (◆), and 4 μM (▲). Orange square (■) represents refolding data of 2 μM protein.



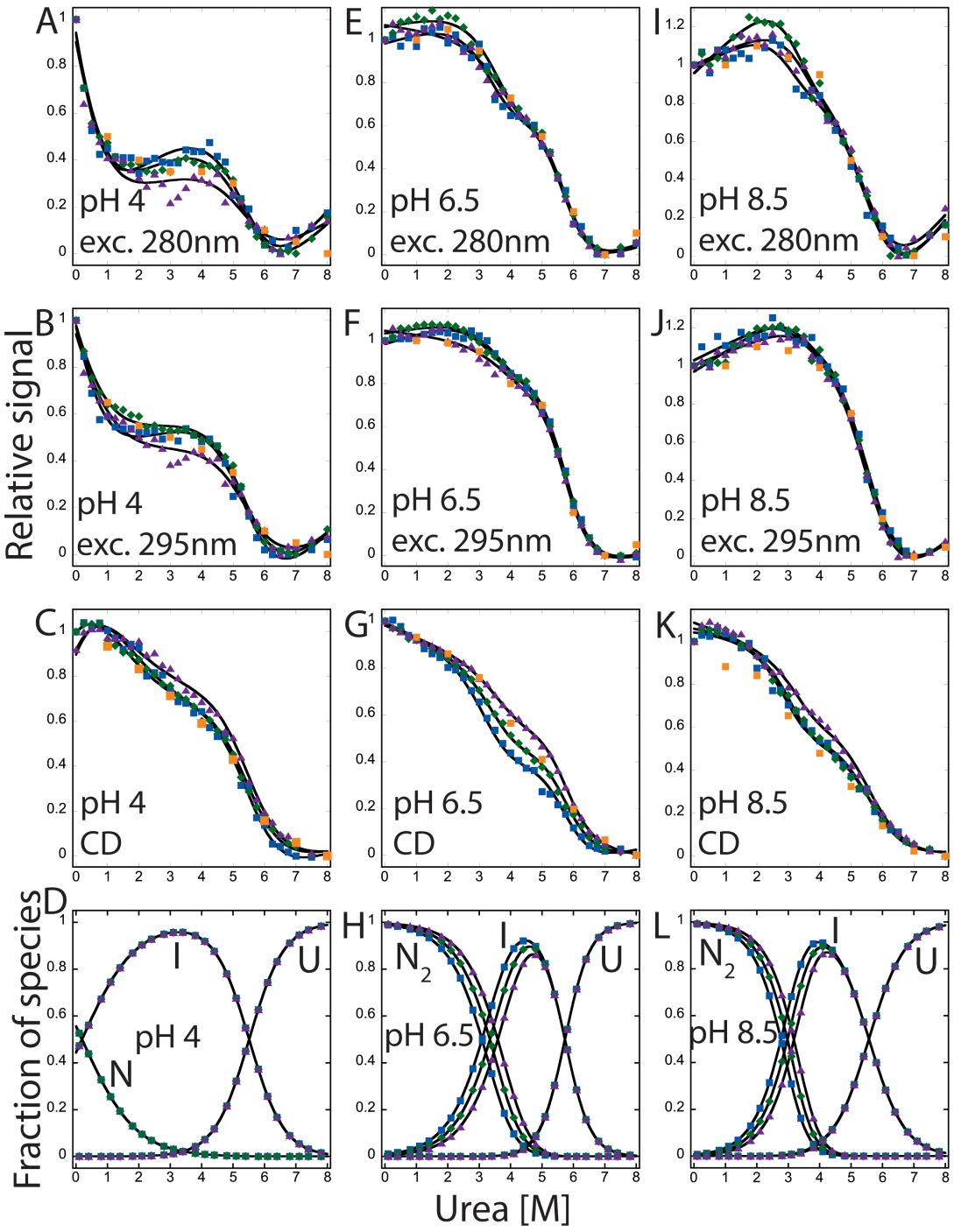
Supplementary Figure 4: Equilibrium unfolding of PCP6 from pH 7 to 8.5. Relative signal of fluorescence emission with excitation at 280 nm and 295 nm, and CD at pH 7 (**A, B, C**), pH 8 (**D, E, F**), and pH 8.5 (**G, H, I**). Four different concentrations of proteins were used at pH 7, and 8 fluorescence emission data, and three different concentration of proteins were used at pH 8 fluorescence data and all CD data. Colored solid symbols represent raw data and corresponding solid lines represent the global fits of the data in an appropriate model described in text. Symbols are consistent with previous figure.



Supplementary Figure 5: Fraction of species of PCP6 folding/unfolding as a function of urea concentration from pH 4 to 8.5. The fractions of native, intermediate, and unfolded protein were calculated as a function of urea concentration from fits of the data at respective pH. Protein concentrations symbols and colors are consistent with previous figures. N_2 refers to dimeric native protein, I_2 is dimeric intermediate, and U refers to unfolded species.



Supplementary Figure 6: Equilibrium unfolding fits and fraction of species of PCP7. Relative signal of fluorescence emission with excitation at 280 nm and 295 nm, and CD at pH 4 (**A, B, C**), and pH 8 (**D, E, F**), and calculated fraction of species of pH 4 (**G**), and pH 8 (**H**). Symbols are consistent with previous figures. N_2 refers to dimeric native protein, I is monomeric intermediate, and U refers to unfolded species.



Supplementary Figure 7: Equilibrium unfolding fits and fraction of species of PCP-CA. Relative signal of fluorescence emission with excitation at 280 nm and 295 nm, and CD at pH 4 (**A, B, C**), pH 6.5 (**E, F, G**), and pH 8.5 (**I, J, K**) and calculated fraction of species at pH 4 (**D**), pH 6.5 (**H**), and pH 8.5 (**L**). Symbols are consistent with previous figures. N_2 refers to dimeric native protein, N is monomeric "native" protein, I is monomeric intermediate, and U refers to unfolded species.

Supplementary Table 1: Amino acids composition and theoretical parameters of procaspase-3, -6, -7, and their common ancestor (AncCP-Ef).

Composition [1]	**Proteins**			
	HsCasp3	HsCasp6	HsCasp7	AncCP-Ef
Total Number Amino Acids	277	293	303	275
Ala (A)	12 (4.3%)	19 (6.5%)	18 (5.9%)	9 (3.3%)
Arg (R)	14 (5.1%)	17 (5.8%)	15 (5.0%)	7 (2.5%)
Asn (N)	15 (5.4%)	11 (3.8%)	14 (4.6%)	13 (4.7%)
Asp (D)	20 (7.2%)	20 (6.8%)	27 (8.9%)	20 (7.3%)
Cys (C)	8 (2.9%)	10 (3.4%)	11 (3.6%)	7 (2.5%)
Gln (Q)	4 (1.4%)	7 (2.4%)	11 (3.6%)	10 (3.6%)
Glu (E)	20 (7.2%)	20 (6.8%)	19 (6.3%)	25 (9.1%)
Gly (G)	16 (5.8%)	19 (6.5%)	18 (5.9%)	18 (6.5%)
His (H)	8 (2.9%)	12 (4.1%)	7 (2.3%)	6 (2.2%)
Ile (I)	19 (6.9%)	13 (4.4%)	17 (5.6%)	14 (5.1%)
Leu (L)	20 (7.2%)	26 (8.9%)	20 (6.6%)	25 (9.1%)
Lys (K)	22 (7.9%)	20 (6.8%)	25 (8.3%)	28 (10.2%)
Met (M)	10 (3.6%)	7 (2.4%)	7 (2.3%)	6 (2.2%)
Phe (F)	15 (5.4%)	18 (6.1%)	17 (5.6%)	12 (4.4%)
Pro (P)	7 (2.5%)	10 (3.4%)	12 (4.0%)	9 (3.3%)
Ser (S)	26 (9.4%)	18 (6.1%)	21 (6.9%)	26 (9.5%)
Thr (T)	16 (5.8%)	16 (5.5%)	15 (5.0%)	13 (4.7%)
Trp (W)	2 (0.7%)	2 (0.7%)	2 (0.7%)	1 (0.4%)
Tyr (Y)	10 (3.6%)	10 (3.4%)	9 (3.0%)	13 (4.7%)
Val (V)	13 (4.7%)	18 (6.1%)	18 (5.9%)	13 (4.7%)
Molecular Weight (Da)	31,608	33,310	34,277	31221
pI	6.09	6.46	5.72	5.2
Extinction Coefficient (280 nm, M-1cm-1) [2]	26,500	25,900	24,410	24870

[1] Parameters exclude the LEHHHHHH C-terminal tag.
[2] Assuming all cysteine residues are reduced.

Supplementary Table 2: Free energy changes (ΔG)of each step of folding/unfolding of extant and ancestral effector caspases. PCP3 data is taken from Bose and Clark, 2001 for the comparison.

pH	PCP6			PCP7			PCP-CA			PCP3			
	$\Delta G_1^{H_2O}$	$\Delta G_2^{H_2O}$	$\Delta G_{total}^{H_2O}$	$\Delta G_1^{H_2O}$	$\Delta G_2^{H_2O}$	$\Delta G_{total}^{H_2O}$	$\Delta G_1^{H_2O}$	$\Delta G_2^{H_2O}$	$\Delta G_{total}^{H_2O}$	$\Delta G_1^{H_2O}$	$\Delta G_2^{H_2O}$	$\Delta G_3^{H_2O}$	$\Delta G_{total}^{H_2O}$
4	-	27.6±0.8	27.6	-	5.8±0.3	5.8	0.1±0.1	5.9±0.3	6	-	-	3.8±0.2	3.8±0.2
4.2										-	1.3±0.3	3.7±0.3	5±0.6
4.5	0.9±0.1	30.4±1.4	31.3	-	-	-	-	-	-	-	1.7±0.6	5.3±0.5	7±1.1
4.75										0.8±0.2	9.2±0.1	5±0.2	15±0.5
5	0.4±0.5	29.5±1	29.9	-	-	-	-	-	-	3.1±0.6	9.6±0.2	5.3±0.3	18.1±1
5.5	0.3±0.3	28.5±1.4	28.8	-	-	-	-	-	-	3.3±0.1	11±1.9	6.8±1.8	21.1±3.8
6	5.3±0.9	26.1±0.9	31.4	-	-	-	-	-	-	6.1±1.0	10.3±0.3	5.6±0.2	22±1.5
6.5	6±0.9	27.6±1	33.6	-	-	-	12.8±0.4	8.5±0.5	21.3	5.9±0.8	10.7±0.6	5.9±0.6	22.6±2
7	5.5±0.7	23.8±0.2	29.3	-	-	-	-	-	-	8.3±1.3	10.5±1	7±0.5	25.8±2.8
7.5	8.4±0.8	24.4±0.9	32.8	10.2±0.2	5.2±0.1	15.4	15±0.1	6.9±0.3	21.9	7.9±0.1	9.7±0.3	7.2±0.5	24.7±0.9
8	6.8±0.6	27.7±1.6	34.5	12.9±0.1	5.5±1.3	18.4	-	-	-	5.8±0.8	9.6±0.2	7.8±0.4	23.2±1.4
8.5	5.4±0.5	25.3±1.3	30.8	-	-	-	13.8±0.8	5.8±0.3	19.6	4.9±0.6	9.6±0.3	6.2±0.6	20.7±1.5

Supplementary Table 3: Cooperative index (m-values) of each step of folding/unfolding of extant and ancestral effector caspases. PCP3 data is taken from Bose and Clark, 2001 for the comparison.

pH	PCP6			PCP7			PCP-CA			PCP3			
	m_1	m_2	m_{total}	m_1	m_2	m_{total}	m_1	m_2	m_{total}	m_1	m_2	m_2	m_{total}
4	-	3.9±0.2	3.9	-	1.5±0.1	1.5	0.7±0.1	1.1±0.1	1.8	-	-	1.1±0.1	1.1±0.1
4.2										-	1.5±0.2	0.9±0.1	2.4±0.3
4.5	0.9±0.1	3.5±0.2	4.4	-	-	-	-	-	-	-	2.2±0.4	1.2±0.1	3.4±0.5
4.75										2.8±0.1	0.9±0.2	1.2±0.1	4.9±0.4
5	0.8±0.1	3.2±0.2	4	-	-	-	-	-	-	2.3±0.3	0.2±0.1	1.2±0.1	3.7±0.5
5.5	0.7±0.1	3.2±0.2	3.9	-	-	-	-	-	-	3±0.1	0.4±0.2	1.5±0.2	4.9±0.5
6	2.7±0.4	3.1±0.2	5.8	-	-	-	-	-	-	3.4±0.5	0.2±0.1	1.3±0.1	4.9±0.7
6.5	2.4±0.3	3.4±0.2	5.8	-	-	-	1.5±0.1	1.5±0.1	3	2.6±0.4	0.4±0.1	1.3±0.1	4.3±0.6
7	2±0.2	2.8±0.2	4.8	-	-	-	-	-	-	2.8±0.5	0.5±0.1	1.2±0.1	4.5±0.7
7.5	2.6±0.3	2.9±0.2	5.5	1.3±0.1	1.2±0.1	2.5	2±0.1	1.2±0.1	3.2	2.8±0.1	0.4±0.1	1.2±0.1	4.4±0.3
8	2.2±0.2	3.5±0.3	5.7	1.6±0.1	1.3±0.1	2.9	-	-	-	2.8±0.4	0.4±0.1	1.5±0.1	4.7±0.6
8.5	1.9±0.2	3.4±0.3	5.3	-	-	-	2±0.2	1±0.1	3	1.9±0.2	0.4±0.1	1±0.1	3.3±0.4

Chapter 4

Evolutionary characterization of caspase-7 substrate specificity

Abstract

Apoptotic caspases are conserved in metazoans for more than 950 million years, and a series of gene duplications resulted in initiator and effector subfamilies. The effector caspase genes (caspase-3, -6, and -7) were evolved from a common ancestor more than 650 million years ago (Mya) when the gene for a common ancestor (CA) was duplicated to caspase-6 and caspase-3/7 lineages. The caspase-3/7 lineages again duplicated to form caspase-3, and -7. Since then, not only mammals but also some invertebrates consist of multiple effector caspases. All caspases cleave their substrates after aspartate residue (P1 position) but have different preference at the P4 position which defines their discrete cellular roles. We examined the evolution of substrate specificity in caspase-7, which prefers aspartate at the P4 residue similar to caspases-3 and different specificity than caspase-6, which prefers valine, by reconstructing the CA of effector caspases (AncCP-Ef) and the ancestors of caspase-7. We show that DxxDase function of caspase-7 defined early in the evolution, where AncCP-3/7 demonstrates a preference for Asp over Val at P4, however, the catalytic activity of caspase-7 decreases towards modern human caspases.

Introduction

Caspase subfamilies evolved from a common ancestor and developed new oligomeric states, enzyme specificity, and allosteric regulation; hence, the caspase family is an attractive model for examining protein evolution (1). Caspase genes and their functions are ancient and well conserved in all metazoans (2) and are thought to be evolved from an ancestral immune system (3). Caspases are divided into two classes, apoptotic and inflammatory caspases (Caspase-1, -4, and -5) based on their functions. Apoptotic caspases further evolved into two subfamilies initiator and effector caspases. Initiator caspases act upstream in the apoptotic cascade, activating effector caspases while effector caspases are the executioner of cell death. After gene duplication from the ancestral initiator and effector caspases, four initiator caspases (caspase-8, -10, -18, c-FLIP) and three effector caspases (caspase-3, -6, -7) were evolved, respectively (Fig. 1A) (1, 4).

All caspases are produced in the cell as inactive zymogens that must be activated following cell signaling changes. In general, initiator caspases are stable monomer, and proximity-induced dimerization is enough for activation whereas, effector caspases are stable dimer and are activated via cleavage of previously existing inactive dimer by initiator caspases (5). The oligomeric form of the zymogen and its activation mechanism is key to the regulation of apoptosis (6). Monomeric initiator caspases form either homodimer or heterodimer based on cell signaling to activate apoptotic or necroptosis pathways respectively (7). In contrast, effector caspases being an obligate homodimer, their cellular functions developed through their ability to cleave substrates (4, 8). Multiple studies in the enzyme families have identified the features that contribute

to the enzyme specificity (9, 10). Usually, the substrate-binding site of an enzyme provides the proper conformational scaffold, which facilitates substrate-binding, so mutation of any amino acids in that scaffold could change the enzyme specificity required by the cellular environment (11–13).

All caspases recognize tetrapeptide sequences except caspase-2 (recognize pentapeptide) in their target substrate (14). In some caspases, exosite also plays a role in substrate binding and specificity (15–17). Tetrapeptide motifs (P1-P4) on the peptide sit well in the substrate-binding pocket (S1-S4) in the active site, where P1 residue is highly conserved and almost always an aspartate (4). Hence, specificity is mainly determined by the amino acid preference at the P4 position. On this basis, caspase are sub-categorized into three groups: group I prefers a bulky residue (W, H, Y); group II prefers hydrophilic residues (D, E); and group III prefers aliphatic residues (I, L, V) (Fig. 1B). Although, substrate-binding site of effector caspases is relatively closely related, few mutations cause the change in specificity. Caspase-6 prefers hydrophobic residues (group III specificity), while caspase-3 and -7 prefer charged amino acids (group II specificity) at the P4 position (18). In the evolution of the chordate, caspase substrate specificities were important in the brain and nervous systems' developmental stages, so what may appear to be subtle changes in enzyme selection have enormous consequences in cellular development.

Effector caspases have two distinct specificities-DxxD (caspase-3, and -7) versus VxxD (caspase-6) (4, 19). A previous study in caspase-6 lineage showed that modern enzymes evolved from promiscuous ancestral proteins through amino acid substitutions (1). VxxDase specificity of caspase-6 evolved earlier in the caspase-6 evolution from a

promiscuous common ancestor of caspase-3/6/7 due to the evolution of the charged network in the caspase-6 subfamily (1). However, the evolutionary trajectories that resulted in the distinct DxxDase substrate specificities in caspase-3 and -7 lineages from the promiscuous ancestor are unknown, which laid us to compare among modern caspases which have DxxDase specificity and their ancestors. Vertical and horizontal comparison among the effector caspase could help find the key active site residue or set of residues responsible for diversity in an evolutionary context.

Vertical comparison can be made by inferring ancestral protein sequences using the ancestral state reconstruction (ASR) technique. ASR is a powerful computational tool used to study protein evolution which predicts the ancestral protein sequences inferring the sequences of ancestral proteins using multiple sequence alignment and phylogenetic relationship of modern proteins (20–23). Once the ancestral protein sequence is established, the corresponding coding DNA is synthesized. The ancestral protein is then expressed and characterized experimentally. ASR has been done with various proteins, but *in vivo* expression has yet to be studied. Hence, in this study, we have expressed resurrected proteins in *C. elegans,* intending to develop a model system to study ancestral proteins.

C. elegans possesses a single caspase compared to multiple caspases and complex apoptotic machinery in mammals (24). Despite the discrepancy between the numbers of regulator proteins, the necessary apoptosis process is evolutionarily conserved (25, 26). The primary cell death regulators in *C. elegans* are EGL-1, CED-9, CED-4, and CED- 3 (24, 27). All of these regulators have their mammalian counterparts. To briefly summarize the apoptotic cascade in *C. elegans*: EGL-1 is the key activator of

apoptotic cell death and determines which cells die and survive (28) (Fig. 2). Generally, 131/1090 of cells in *C. elegans* undergo programmed cell death (29, 30). Among these regulators, EGL-1 first binds with CED-9, causing a significant conformational change in CED-9 which releases CED-4 from the CED-9/CED-4 complex (31)(32). CED-4 translocates from the mitochondria to the perinuclear membrane (33). The released CED-4 dimers then oligomerize to form a funnel-shaped CED-4 octamer, which may recruit two CED-3 zymogens and facilitate its auto-catalytic activation through zymogen dimerization (34) (Fig. 2).

To determine the evolution of DxxD specificity in caspase-7, we reconstructed ancestral proteins for the common ancestor (CA) of all three effector caspases (called AncCP-Ef1 and AncCP-Ef2) (1), CA of the caspase-3/7 branch (AncCP-3/7), CA of caspase-7 (AncCP-7), CA of caspase-7 from amniotes (reptiles/birds/mammals) (AncCP-7Am, where 'Am' refers to amniotes), CA of caspase-7 from mammals (AncCP-7M, where 'M' refers to mammals) (Fig. 3A). We then expressed AncCP-Ef2 and AncCP-7 in *C. elegans* germline and showed that they rescued cell death function. Moreover, enzyme activity assay and substrate phage display indicate that the selection of Asp over Val occurred early in the evolution before duplication into caspase-3 and -7 lineages. After duplication, catalytic efficiency increased in AncCP-7. Later on, decrease up to the human caspase-7 due to an increase in K_M.

Results

Ancestral state reconstruction of effector caspases and caspase-7 lineage

Reconstruction of ancestral effector caspases was done as described in our previous publication (1). To examine the evolution of changes in substrate specificity from a common ancestor of effector caspase to the extant caspases, we reconstructed the ancestral proteins for the effector caspases and caspase-7, which include common ancestor (CA) of mammalian caspases (AncCP-M), CA of caspase-7 from mammals, reptiles, and birds (AncCP-Am), CA of caspase-7 including fishes (AncCP-7) and then to the common ancestor of caspase-3, and -7 (AncCP-3/7) (Fig. 3A). In the CA of effector caspases, we carried out two separate APR experiments by using two representative datasets and comparing their posterior probabilities to examine the robustness of AncCP-Ef reconstruction. The two CA of effector caspases, named AncCP-Ef1 and AncCP-Ef2, represent the same pool of possible ancestors from the different reconstructions with different sets of data, and both proteins were resurrected and characterized to corroborate the robustness of the APR. AncCP-Ef1 and AncCP-Ef2 have 80% sequence identity and differ in 39 sites mainly in less conserved sites of the protein. The difference between AncCP-Ef1 and AncCP-Ef2 has little effect on enzyme activity (1).

APR techniques can reveal change in protein sequences from ancestors to the modern proteins leading to the evolution of new functions, neofunctionalization. The caspase-3, -6 and, -7 diverged from a CA ~650 Mya into the caspase-6 and caspase-3/-7 lineages, which again diverged into caspase-3 and caspase-7 ~450 Mya (1) (Fig. 3A). In the horizontal comparison, caspase-7 has 56% amino acid sequence identity

with caspase-3 and 37% with caspase-6. In vertical comparison, caspase-7 has 83% sequence identity with AncCP-7M excluding prodomain then as evolutionary time increases. Sequence identity with caspase-7 decreases to 79%, 75%, 71% to AncCP-7Am, AncCP-7, and AncCP-3/7, respectively (Fig. 3B). The common ancestors AncCP-Ef1 and AncCP-Ef2 have 58% and 60% sequence identity with caspase-7, respectively. Prodomain sequences does not affect enzymatic function in vitro and are removed during zymogen maturation (8, 35). Since prodomain sequences are highly variable, we used the sequences from extant proteins for consistency. One hundred and twenty-five amino acids out of 260 were conserved in the protease domain. Active site and protein cores are conserved among all caspase-7 ancestors. Variable sequences are mostly prevalent in intersubunit linker and surface helices, particularly helices 2 and 3 (Fig. 4).

Reconstructed ancestral caspase support cell death in vivo in *C. elegans*

To examine whether resurrected ancestral caspases support programmed cell death in a cellular environment, we expressed two of the reconstructed proteins (AncCP-Ef2 and AncCP-7) in egl-1 (gf);ced-3 (lf) strains of *C. elegans*. In hermaphrodite worms, TRA-1 binds to the TRA-1 binding site downstream of the EGL-1 transcription unit, thereby blocking an HSN activity (hermaphrodite specific neuron) transcriptional activator of EGL-1 (36) (Fig. 2). This results in the repression of *egl-1* transcription and ensures HSN survival. HSNs are essential for hermaphrodites to lay eggs. In egl-1 gain-of-function allele hermaphrodites, TRA-1 is active but unable to bind to the mutant TRA-1 binding site. This allows the HSN activator to mediate egl-1 transcriptional activation as in males, resulting in the inappropriate death of HSNs. Therefore, the egg-laying

muscles innervated by these neurons fail to contract, and the worm is unable to lay its eggs (36). This inappropriate cell death occurs in a CED-3- dependent manner. Our experiment took egl-1 (gf);ced-3 (lf) strain of worms to express our protein of interest and rescue the cell death function, thereby giving egg-laying defective phenotype. As expected, both ancestral caspases, AncCP-Ef2 and AncCP-7, rescue cell death function by giving egg-laying defective phenotype and proved that they execute the cell death function in vivo (Fig. 5C & 5D). As a control, we used egl-1 (gf), egl-1 (gf); ced-3 (lf) as a positive and negative control of egg-laying defective phenotype. In addition, we injected ced-3 plasmid in egl-1 (gf);ced-3 (lf) to make sure ced-3 itself works in these experimental settings (Fig. 5A & 5B).

The specificity of caspase-7 towards DxxDase evolved early before duplication

We examined the enzyme activity of extant human caspase-7 leading up to a common ancestor of caspase-3/6/7 against two tetrapeptide sequences (DEVD and VEID), representing specificities of caspases-3/7, and -6, respectively (Fig. 6). The activity of all proteins tested here for VEID was below the detection limit ($\sim 5\times10^{2}$ M^{-1} s^{-1}) for the assay due to high K_M values. Catalytic efficiency (k_{cat}/K_M) values are as low as $\sim 2.8\times10^{3}$ M^{-1} s^{-1} for AncCP-Ef2 to as high as $\sim 1.1\times10^{6}$ M^{-1} s^{-1} for AncCP-7 (Fig. 6A). The data shows that the common ancestor of all effector caspases has very low activity against tetrapeptide inhibitor, which increased by ~200 fold in AncCP-3/7 and increased by ~400 fold to the AncCP-7. Then, catalytic efficiency decreased from AncCP-7 to extant human caspase-7 linearly, with caspase-7 having k_{cat}/K_M value $\sim 4.8\times10^{4}$ M^{-1} s^{-1}, ~20 fold lower than AncCP-7. kcat value of caspase-7 ancestral proteins

is ~5 s^{-1} while Km value of ~10 mM (Figs. 6B & 6C). In extant human caspase-7, kcat decreased to 2.3 $s^{-1,}$ and Km value increased to 50.5 mM, suggesting low catalytic efficiency than ancestral protein is due to weak binding of the enzyme to the substrate in the ancestral proteins. Together, the data for AncCP-Ef2 show that the CA of effector caspases has a low activity for the tested tetrapeptides. Activity significantly improved to the caspase-7 ancestors and decreased to the extant caspase-7, particularly regarding aspartate at P4.

We also compared the substrate specificity using substrate-phage selection. As shown in Figure 7A, HsCaspase-7 shows preference (P5–P1') for GDYTDF. Similarly, all other caspase-7 ancestral proteins also select "Asp" in the P4 position (Figs. 7B-7D). AncCP-Ef2 showed a clear preference of Asp at P4, with an overall P5-P1' selection of ED(V/E)PDS (Fig. 7F), although it has very low activity against DEVD tetrapeptide substrate. However, AncCP-Ef1 showed promiscuity with the selection of both valine and aspartate along with leucine, isoleucine, threonine, and alanine. AncCP-3/7 also prefers the Asp at the P4 position. An overall P5–P1' selection of x(D/A/V)x(P/V)D(S/T) (Fig. 7E) (x, refers to any amino acids), but it also selects Alanine and Valine. However, it didn't cleave VEID tetrapeptide substrate and had the highest catalytic efficiency against DEVD substrate among modern and ancestral caspase-7. Along with AncCP-Ef2 and AncCP-3/7, all other caspase-7 lineage proteins follow similar specificity at the P4 position. There was no preference for a particular amino acid in the P3 position among all the proteins. However, Glu and Val are the most prevalent ones. Similarly, at the P2 position not a particular amino acid preference for any amino acids, while surprisingly, proline is the most frequent one following threonine, valine, and methionine

(Fig. 7). Overall, all the caspase-7 lineage ancestor prefers aspartate at P4 position with no specific preference in P3 and P2 position and selecting small amino acids in P1' and P5 position, consistent with the caspase-3 and -6 (1, 37, 38). Together, the ancestor of effector caspases was less active and not specific, which changed after duplication to caspase-6 and caspase-3/7 lineages. Caspase-3 and -7 specificity towards "Asp" at P4 position evolved before their own duplication.

Discussion

Over 50 years ago, Zuckerkandl and Pauling pioneered the research in evolutionary biochemistry of protein to understand structure-function (39, 40). The evolutionary study helps determine the cause of change in function along the phylogeny, which could be due to random chance, changes in energy landscapes, and selection pressure (41). Comparative study of the structure and function of the extant proteins is somewhat limited in finding the evolutionary map of genetic changes. This is where APR comes into play as it adds the dimension of evolutionary time and fitness to the structure-function relationship. In addition, it provides powerful tools to characterize evolutionary changes in proteins experimentally. Thus, after inferring ancestral protein using APR, biochemical and biophysical methods can be applied to determine the evolutionary trajectory of proteins (42, 43).

In this study, along with the biochemical characterization, we also expressed two ancestral proteins in *C. elegans*. It is a unique system for studying the components of apoptosis since the apoptotic mechanism is conserved, and it has fewer apoptotic cellular machinery than humans and other vertebrates (25, 44). We expressed ancestral

caspase in egl-1(gf); ced-3(lf) strain of the with an expectation of getting egg-laying defective worms in the case of ancestral protein executing the apoptosis. As expected, the cell death function of ced-3 in worms was rescued by AncCP-Ef2 and AncCP-7, which proved that they are functional in vivo. In comparing the bagging size or the number of hatched eggs in the belly of worms, a positive control (ced-3) and AncCP-7 had relatively more hatched eggs than AncCP-Ef2, reflection of the activity and specificity of an enzyme. Further study can be done in worms to understand the effect of specificity and tissue-specific expression in detail. The main finding of this particular experiment of ced-3 and ancestral proteins is despite being evolutionarily very far apart, the basic mechanism of the cell death is conserved. Across species, studies like this could give us insight into the evolutionary changes in protein at different times. More specifically, *C. elegans* study helps to correlate the characteristics of proteins in mammals and invertebrates.

Caspase-7 prefers “Asp” at the P4 position, and k_{cat}/K_M against DEVD tetrapeptide was 4.8×10^4 M^{-1} s^{-1} which is similar to the previous studies (37). Relative to the extant human caspase-7, AncCP-Ef2 was ~20-fold less active, which then increased by ~200 and ~400-fold to the AncCP-3/7 and AncCP-7, respectively, or AncCP-7 is ~20-fold more active than extant human caspase-7. Several studies suggest that proteins have changed from lower to higher specificity over a long evolutionary time (45, 46), which looks to be true for caspase-7 ancestral proteins until its duplication to the caspase-3 and -7. Evolutionary study suggests, promiscuity of an enzyme might have provided the ability to perform diverse functions in primordial organisms (47). Another study suggested, substrate specificity increases following gene

duplication (48), which is in agreement with our data. The duplication of caspase-6 and 3/7 lineage followed the change in specificity of VxxDase and DxxDase, respectively, which then become more specific and more active after duplication to the caspase-3 and -7. However, few nodes after duplication rate of catalysis of caspase-7 decreased. It is not clear yet what causes the decrease of activity in extant human caspase-7 and other functional changes related to catalysis. We also speculate that caspase-7 in other species than a human could be as active as AncCP-7 and be a major executioner instead of caspase-3.

Overall, the data suggest that the conformational landscape of the ancestral effector caspase remains in modern caspases. Few substitutions in the ancestral protein sequence cause the neofunctionalization in the extant enzymes. For example, out of forty-two evolutionary changes among AncCP-Ef1 and AncCP-6An (the common ancestor of caspase-6), only three vertical substitutions were responsible for the change in specificity of promiscuous AncCP-Ef1 to VxxDase in caspase-6 lineage (1). Directed evolution study by Hill et al. altered the caspase-7 specificity to mirror the caspase-6 with just four mutations [CP-158 (Y230V), CP-160 (W232Y), CP-162 (S234V), and CP-GP-03 (Q276D)] (37). Here we found three of those four amino acids are not conserved in the ancestral effector, supporting those amino acids' substitutions to be more substrate-specific. Substrate phage display results show that AncCP-Ef2 selected Asp at the P4 site, which could be due to its low reactivity rather than being highly specific. However, into the next node, AncCP-3/7, it mostly selected Asp along with Ala and Val. The proportion of valine at the P4 site decreased in the AncCP-7 and AncCP-7An. In contrast, it again decreased to the AncCP-7M and extant caspase-7, which supports the

enzyme activity data and suggests that catalytic efficiency decreased, and specificity is compromised. Besides, other substrate positions P2, P3, and P5, are not that specific, as shown in our phage display data.

In summary, the APR study is effective for characterizing evolutionary changes to infer functional changes in the extant proteases. Determination of substitution of amino acids from the ancestor to the extant proteins and their role in substrate selection may provide the strategies for engineering caspases with desired substrate selection as well as manipulation of their activity. In terms of caspase specificity, the data show that enzyme specificity was established early in the evolution of caspase-7 and that allosteric regulation likely followed through subsequent evolution, which may have caused the changes in catalytic efficiency of modern protein. A significant decrease in the catalytic efficiency of modern caspase-7 could be due to the overlapping function of the caspase-7 with caspase-3. In other words, it could be an indication of redundancy. Furthermore, ancestral caspases rescued the cell death function in C. elegans, and C. elegans can be developed as a model to further the detailed study of resurrected proteins.

Materials and methods

Ancestral protein reconstruction, enzyme activity assay and substrate-phage display methodology were similar as stated in chapter 2.

Bacterial strain for feeding worms and seeding plates

The worms were grown and maintained on standard Nematode Growth Medium (NGM) (49) plates seeded with *E.coli* OP50 strain at 20°C. A single colony of *E.coli* OP50 on LB plate was inoculated in LB broth and grown overnight at 37°C. Liquid culture of *E.coli* OP50 was disseminated onto NGM plates and left to dry at room temperature for 4-5 days. Seeded plates were stored at 4°C.

C. elegans strains

C. elegans strains were obtained from the Caenorhabditis Genetics Center (CGC) and include: N2 Bristol, MT1082: *egl-1(n487)*, and MT2563: *ced-3(n1165)*; *egl-1(n487).*

Cloning

Plasmids were constructed using standard genetic techniques. The PCR fragments of *a ced-3* promotor (*ced-3p*) and *ced-3* cDNA were generated from the N2 strain genome and inserted in pPD49.26 (L754, Addgene number) between PstI and XmaI, and XmaI and SacI respectively, to make *ced-3p::ced-3* construct. Similarly, we designed *ced-3p::*AncCP-7 and *ced-3p::*AncCP-Ef2 insert in the L754 vector. Primer sequences used to PCR amplify each rescue fragment were as follows*: ced-3p:* tttctgcagacgtctctttctatatata and tatcccgggcttcggctgatggttgacct, *ced-3*: aaacccgggatgatgcgtcaagatagaa and tttgagctcttagacggcagagtttcgt, *AncCP-7*: tatacccgggatggcggacgatcagg and atagagctcttaaaaatacagctctttggtca, and AncCP-Ef2: tatacccgggatggagaacaccgaaaacag and tatagagctcttacggcgggaagtacagtt.

Germline transformation

We used germline transformation to rescue the egg-laying defective phenotype associated with apoptosis executed by caspase using standard techniques. All three clones mentioned above were microinjected at 10 ng/µl concentration along with *myo-2p*::mCherry plasmid at 5 ng/µl as a transgenic marker in the *egl-1 (gf)*; *ced-3 (lf)* strain. At least two independent transgenic lines with the *myo-2p::*mCherry marker in the pharynx were used to test for rescue.

Microscopy

All transgenic animals were imaged using a Zeiss AxioCam MRm mounted on a Zeiss Imager Z2 microscope based on the expression of mCherry fluorescence in the pharynx. To visualize the phenotype and imaging, worms were anesthetized using 2.5 mM sodium azide in S-Basal and arranged on agarose pad-lined glass microscopy slides in Zeiss Observer Z1 upright microscope with differential interface contrast (DIC).

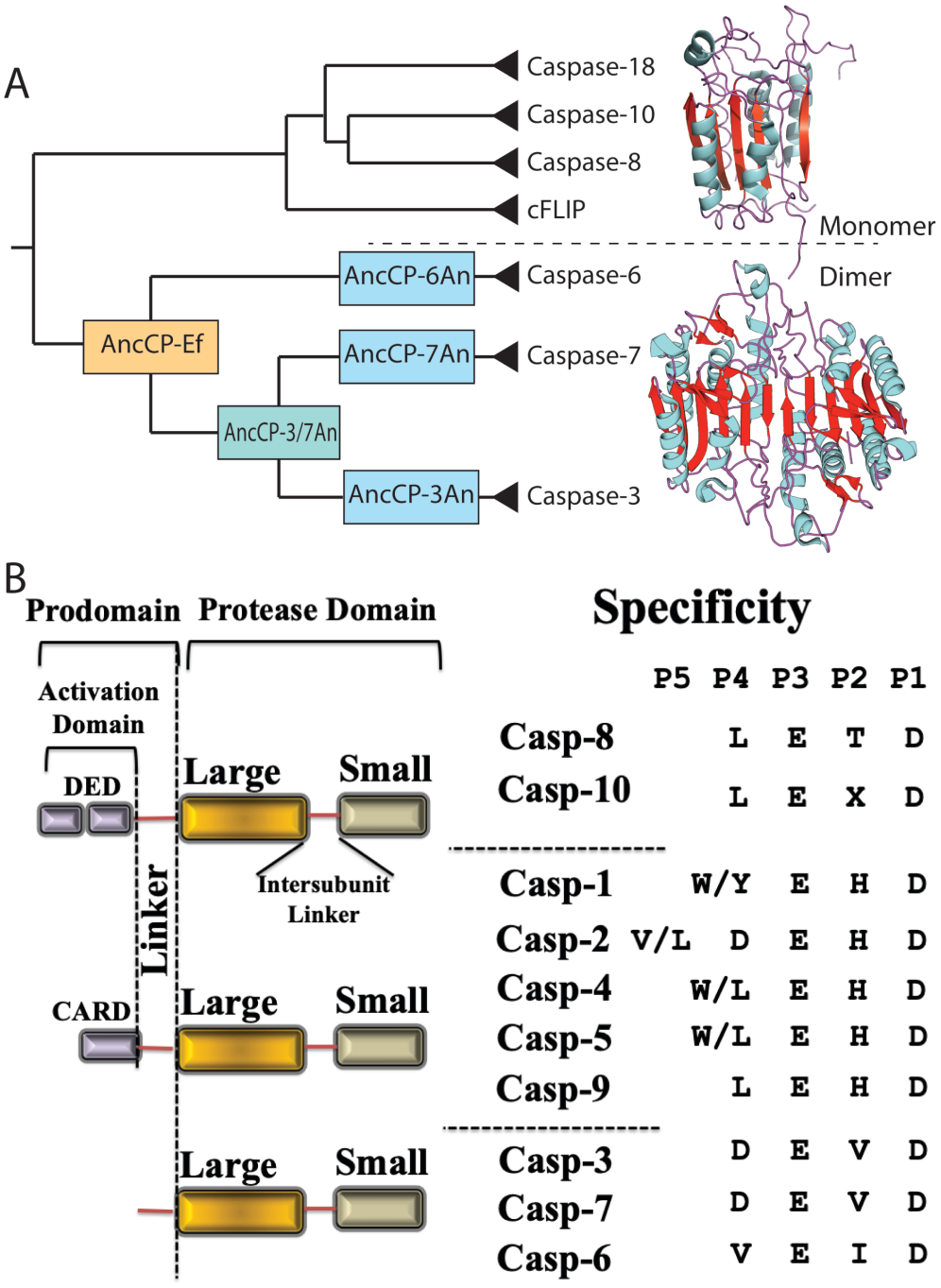


Figure 1: Caspase phylogeny, domain organization and specificity. (A) Evolutionary relationship of initiator (monomer) and effector (dimer) caspases. AncCP-Ef is the CA to effector caspases-3, -6, and -7. AncCP-6An is the CA to the caspase-6 lineage. (B) Domain organization and enzyme specificity for 10 human caspases. DED refers to death effector domain, and CARD refers to caspase activation and recruitment domain. The caspase protomer consists of a large and small subunit separated by an intersubunit linker (IL), which is cleaved upon maturation, and the active caspase is a dimer of protomers. All caspases recognize aspartate at the P1 position of the substrate, and specficity is primarily determined by the amino acid at the P4 position.

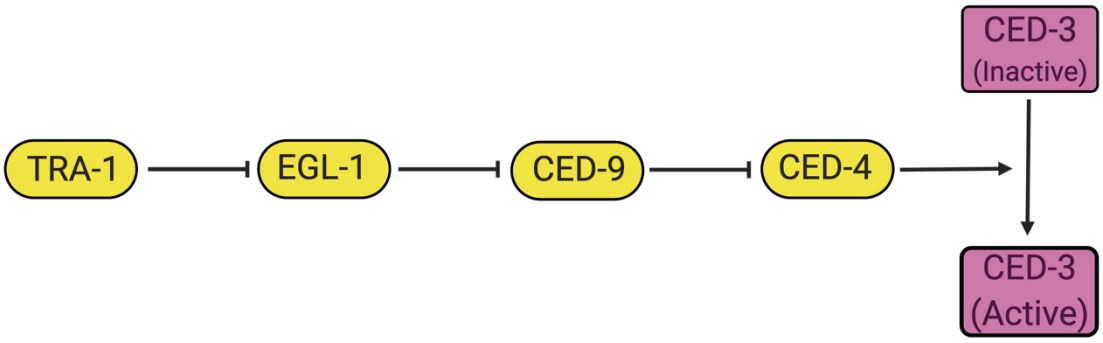
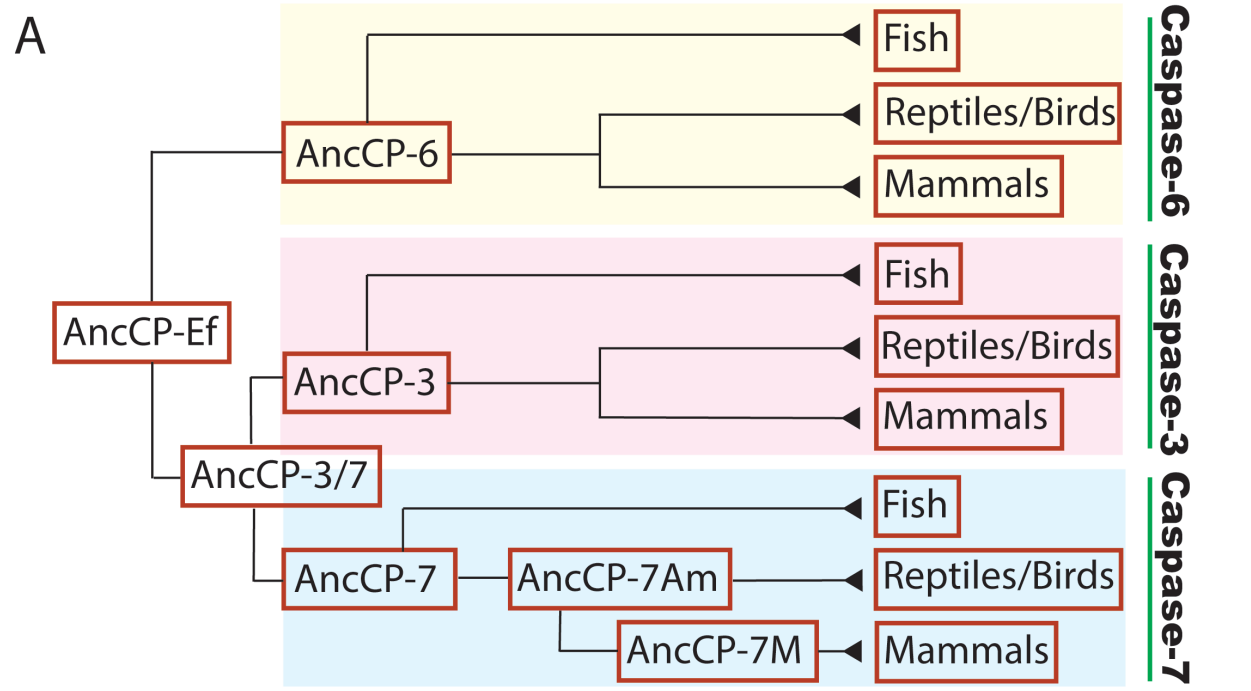


Figure 2: Apoptotic pathway in *C. elegans*. (perpendicular line indicates inhibition and arrow indicates activation.



B

	AncCP-Ef1	AncCP-Ef2	AncCP-3/7	AncCP-7	AncCP-7Am	AncCP-7M	HsCaspase-7
AncCP-Ef1		80/88	78/87	64/77	64/79	65/80	58/77
AncCP-Ef2			76/88	67/80	65/80	66/82	60/78
AncCP-3/7				76/87	78/89	76/87	71/83
AncCP-7					86/96	82/93	75/86
AncCP-7Am						94/97	79/87
AncCP-7M							83/91
HsCaspase-7							

Figure 3: Ancestral caspase reconstruction and their amino acid identity. (A) Reconstructed protein nodes and their phylogenetic relationship with modern caspases. (AncCP-Ef – common ancestor (CA) of caspase-3/6/7, AncCP-3/7 – CA of caspase 3/7, AncCP-6 – CA of caspase-6, AncCP-3 - CA of caspase-3, AncCP-7 – CA of caspase-7, AncCP-7Am – CA of caspase-7 of amniotes, AncCP-7M – CA of caspase-7 of mammals. (B) Comparison of ancestral and human caspase-7 percent amino acid identity.

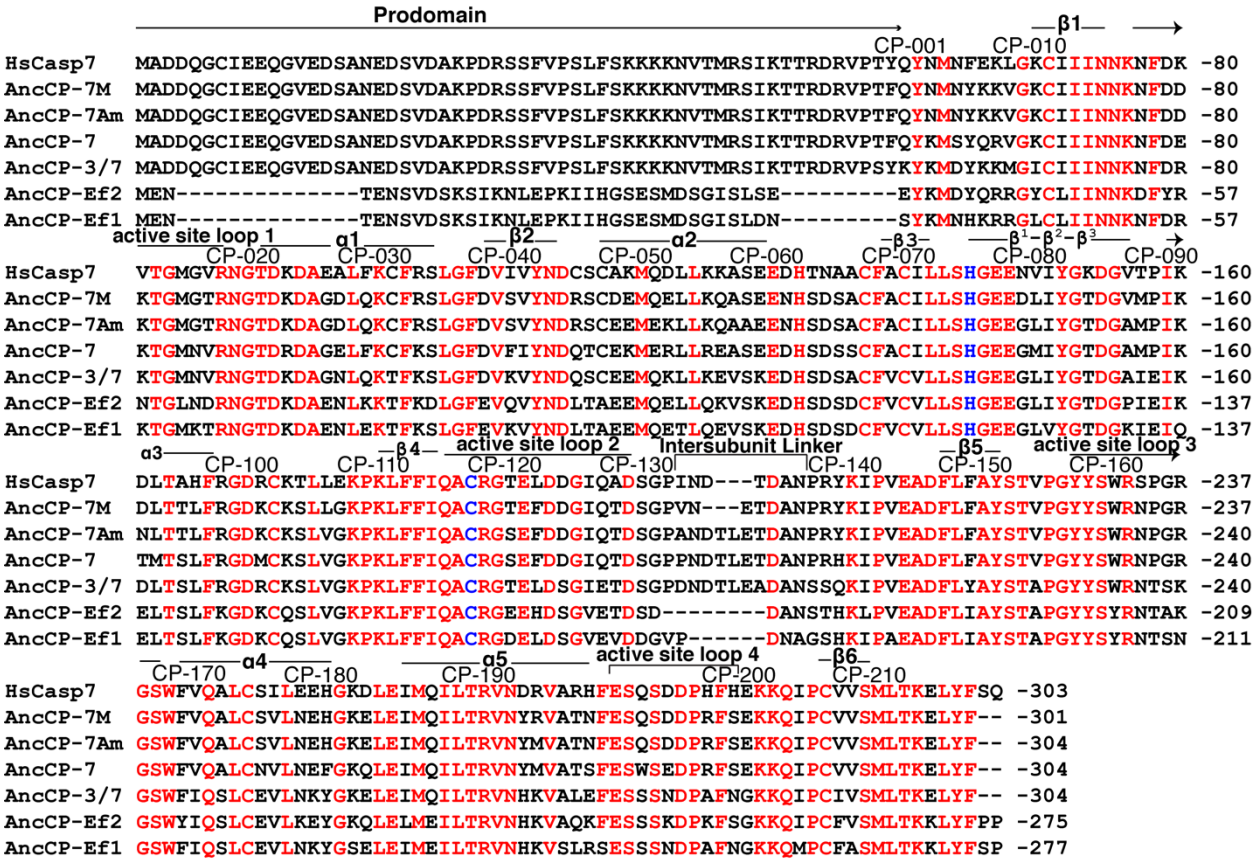


Figure 4: Sequence of human caspase-7 (HsCasp7) compared with caspase-7 ancestral proteins. Numbering refers to CP, secondary structural elements and active site loops are indicated.

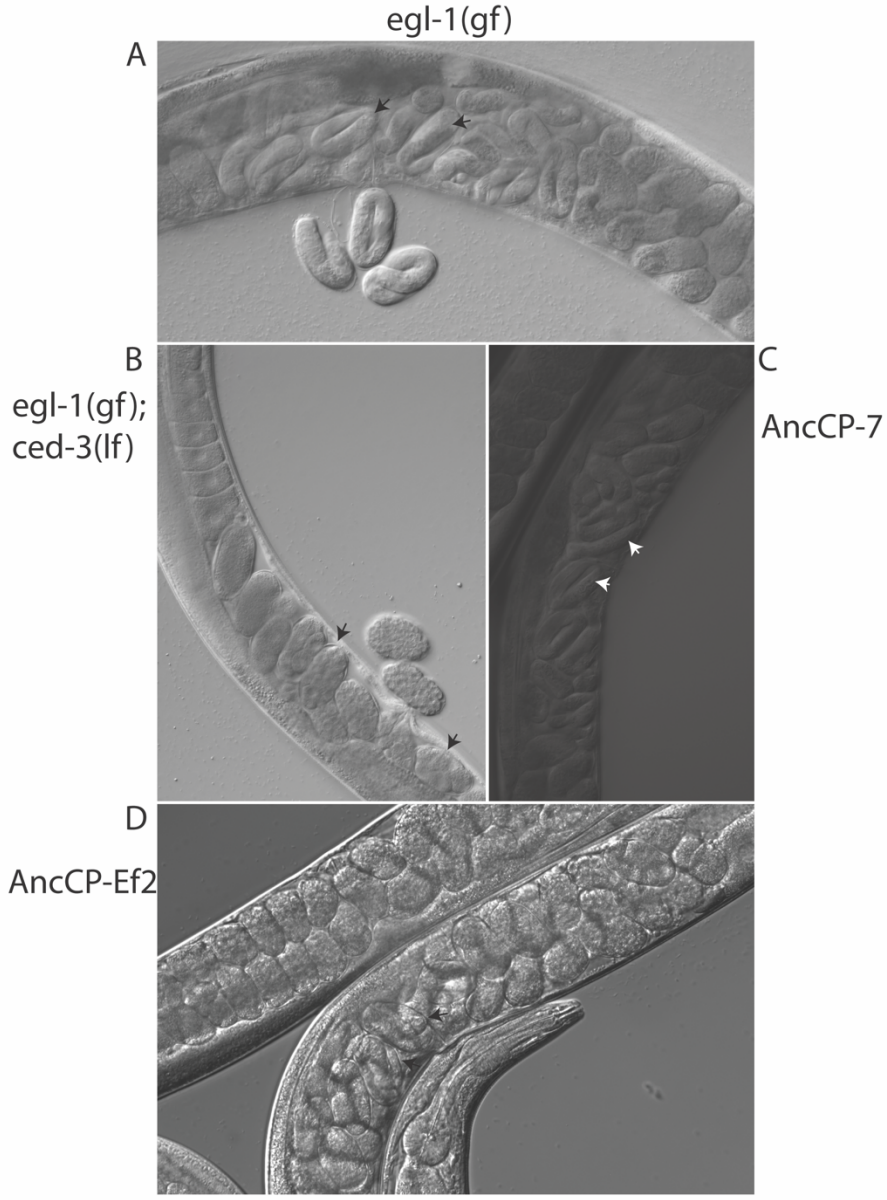


Figure 5: Differential Interface Contrast (DIC) image of gravid adult *C. elegans* strains egl-1 (gf) (positive control) (**A**), egl-1 (gf); ced-3 (lf) (negative control) (**B**) and worms expressing AncCP-7 (**C**), and AncCP-Ef2 (**D**) in egl-1(gf); ced-3 (lf) background. AncCP-7 and AncCP-Ef2 both rescued the cell death function giving the egg-laying defective phenotype. The arrows indicate "bagging" morphological changes that occur when embryos are hatched in utero.

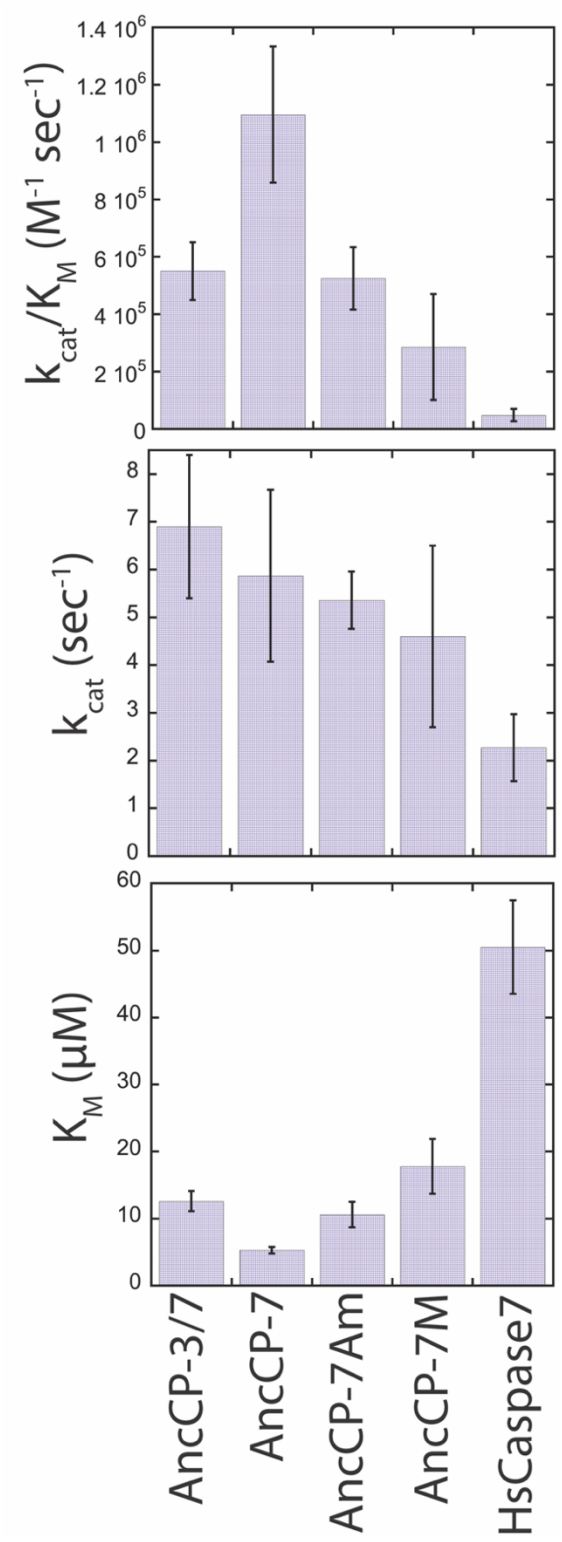
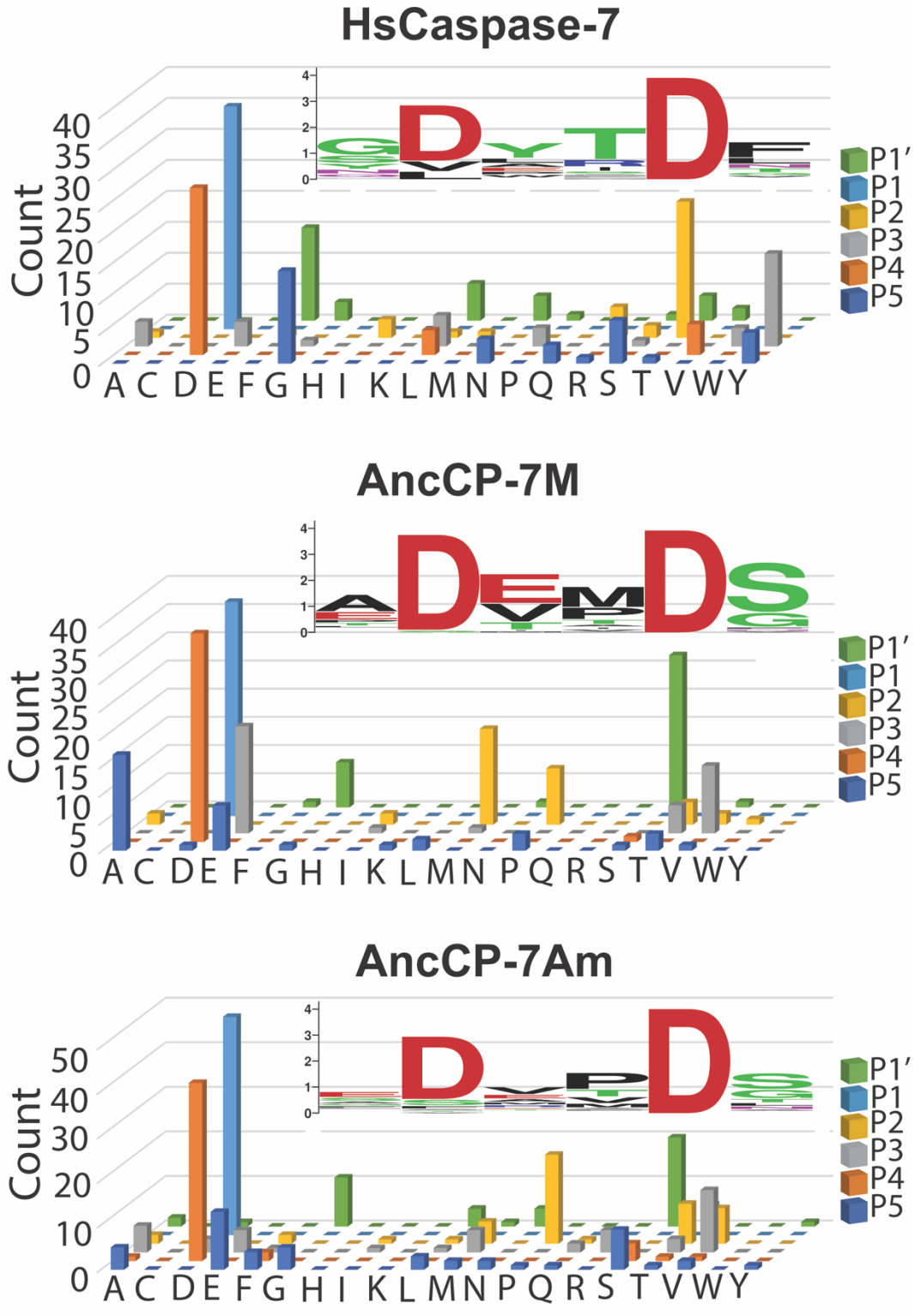


Figure 6: Enzyme activity of ancestral caspase-7. Catalytic parameters kcat/Km (A), kcat (B) and Km (C) of DEVD substrate cleavage.


HsCaspase-7
Count
A C D E F G H I K L M N P Q R S T V W Y
P1′
P1
P2
P3
P4
P5
AncCP-7M
Count
A C D E F G H I K L M N P Q R S T V W Y
P1′
P1
P2
P3
P4
P5
AncCP-7Am
Count
A C D E F G H I K L M N P Q R S T V W Y
P1′
P1
P2
P3
P4
P5

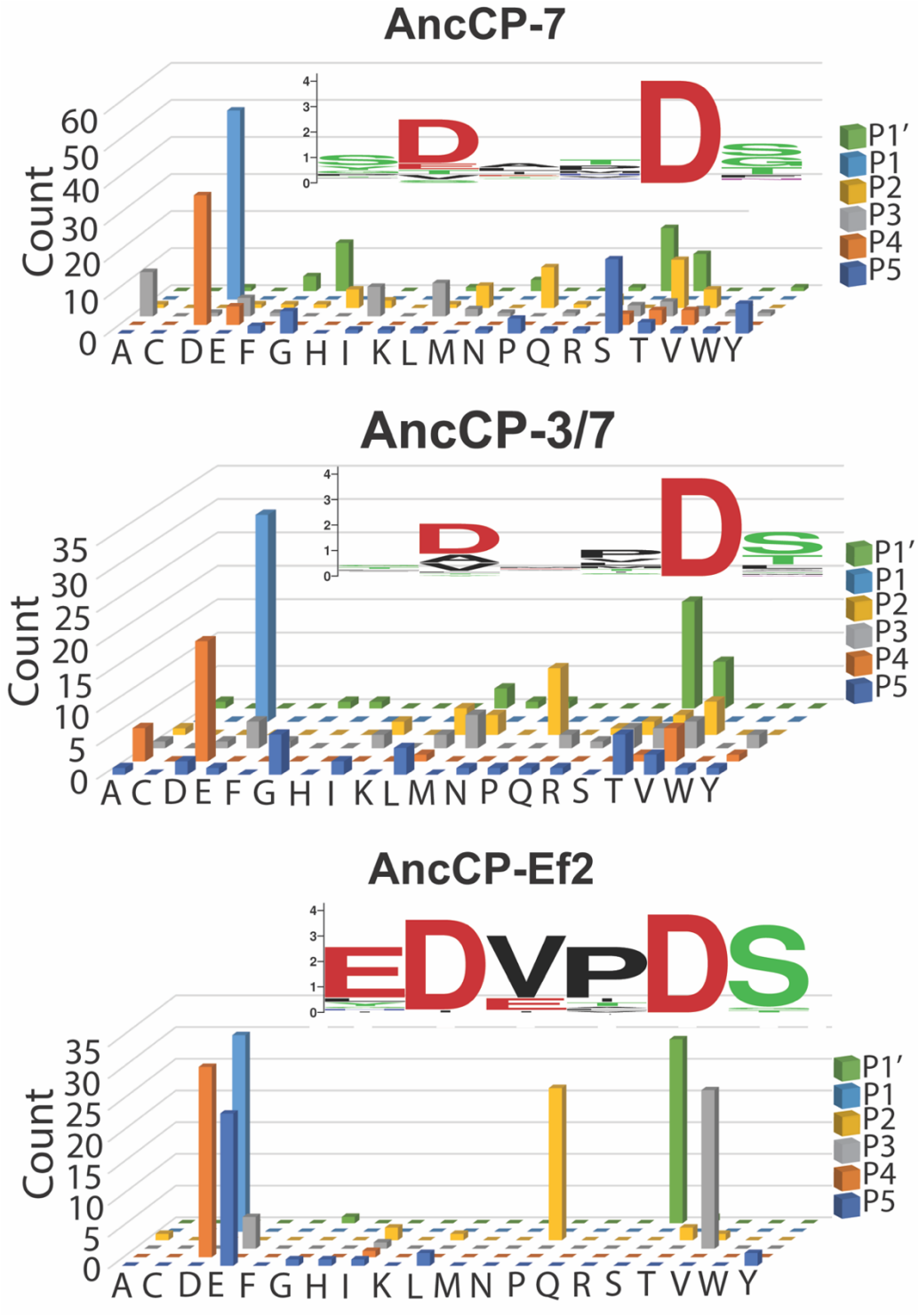


Figure 7: Substrate preferences determined by substrate-phage display. Amino acid preferences shown for substrate positions P5-P1′ for HsCaspase-7 (**A**), AncCP-7M (**B**), AncCP-7Am (**C**), AncCP-7 (**D**), AncCP-3/7 (**E**), and AncCP-Ef2 (**F**). Values on the Y-axes indicate number of phage sequences containing the specified amino acid.

References

1. Grinshpon, R. D., Shrestha, S., Titus-McQuillan, J., Hamilton, P. T., Swartz, P. D., and Clark, A. C. (2019) Resurrection of ancestral effector caspases identifies novel networks for evolution of substrate specificity. *Biochem. J.* **476**, 3475–3492
2. Zmasek, C. M., Zhang, Q., Ye, Y., and Godzik, A. (2007) Surprising complexity of the ancestral apoptosis network. *Genome Biol.* **8**, R226
3. Crawford, E. D., Seaman, J. E., Barber, A. E., David, D. C., Babbitt, P. C., Burlingame, A. L., and Wells, J. A. (2012) Conservation of caspase substrates across metazoans suggests hierarchical importance of signaling pathways over specific targets and cleavage site motifs in apoptosis. *Cell Death Differ.* **19**, 2040–2048
4. Clark, A. C. (2016) Caspase Allostery and Conformational Selection. *Chem. Rev.* **116**, 6666–6706
5. MacKenzie, S. H., and Clark, A. C. (2012) Death by caspase dimerization. *Adv. Exp. Med. Biol.* **747**, 55–73
6. Understanding enzyme function evolution from .pdf
7. Salvesen, G. S., and Walsh, C. M. (2014) Functions of caspase 8: The identified and the mysterious. *Semin. Immunol.* **26**, 246–252
8. Dagbay, K., Eron, S. J., Serrano, B. P., Velázquez-Delgado, E. M., Zhao, Y., Lin, D., Vaidya, S., and Hardy, J. A. (2014) *A multipronged approach for compiling a global map of allosteric regulation in the apoptotic caspases*, 10.1016/B978-0-12-417158-9.00009-1
9. MacGregor, E. A., Janeček, Š., and Svensson, B. (2001) Relationship of sequence and structure to specificity in the α-amylase family of enzymes. *Biochim. Biophys. Acta - Protein Struct. Mol. Enzymol.* **1546**, 1–20
10. Penning, T. M., and Jez, J. M. (2001) Enzyme redesign. *Chem. Rev.* **101**, 3027–3046
11. Pabis, A., Risso, V. A., Sanchez-Ruiz, J. M., and Kamerlin, S. C. (2018) Cooperativity and flexibility in enzyme evolution. *Curr. Opin. Struct. Biol.* **48**, 83–92
12. Mildvan, A. S. (1974) Mechanism of Enzyme Action. *Annu. Rev. Biochem.* **43**, 357–399
13. Mitchell, J. B. (2017) Enzyme function and its evolution. *Curr. Opin. Struct. Biol.* **47**, 151–156
14. Poręba, M., Strózyk, A., Salvesen, G. S., and Drąg, M. (2013) Caspase substrates and inhibitors. *Cold Spring Harb. Perspect. Biol.* **5**, a008680
15. Thomas, M. E., Grinshpon, R., Swartz, P., and Clark, A. C. (2018) Modifications to a common phosphorylation network provide individualized control in caspases. *J. Biol. Chem.* **293**, 5447–5461
16. MacPherson, D. J., Mills, C. L., Ondrechen, M. J., and Hardy, J. A. (2019) Tri-arginine exosite patch of caspase-6 recruits substrates for hydrolysis. *J. Biol. Chem.* **294**, 71–88
17. Scheer, J. M., Romanowski, M. J., and Wells, J. A. (2006) A common allosteric site and mechanism in caspases. *Proc. Natl. Acad. Sci. U. S. A.* **103**, 7595–7600
18. Julien, O., and Wells, J. A. (2017) Caspases and their substrates. *Cell Death Differ.* **24**, 1380–1389

19. Wei, Y., Fox, T., Chambers, S. P., Sintchak, J., Coll, J. T., Golec, J. M. C., Swenson, L., Wilson, K. P., and Charifson, P. S. (2000) The structures of caspases-1, -3, -7 and -8 reveal the basis for substrate and inhibitor selectivity. *Chem. Biol.* **7**, 423–432
20. Joy, J. B., Liang, R. H., McCloskey, R. M., Nguyen, T., and Poon, A. F. Y. (2016) Ancestral Reconstruction. *PLoS Comput. Biol.* 10.1371/journal.pcbi.1004763
21. Eick, G. N., Bridgham, J. T., Anderson, D. P., Harms, M. J., and Thornton, J. W. (2017) Robustness of Reconstructed Ancestral Protein Functions to Statistical Uncertainty. *Mol. Biol. Evol.* **34**, 247–261
22. Saito, S., Saito, C. T., Nozawa, M., and Tominaga, M. (2019) Elucidating the functional evolution of heat sensors among Xenopus species adapted to different thermal niches by ancestral sequence reconstruction. *Mol. Ecol.* **28**, 3561–3571
23. Gumulya, Y., and Gillam, E. M. J. (2017) Exploring the past and the future of protein evolution with ancestral sequence reconstruction: The “retro” approach to protein engineering. *Biochem. J.* **474**, 1–19
24. Conradt, B., Wu, Y.-C., and Xue, D. (2016) Programmed Cell Death During Caenorhabditis elegans Development. *Genetics*. **203**, 1533–62
25. Potts, M. B., and Cameron, S. (2011) Cell lineage and cell death: Caenorhabditis elegans and cancer research. *Nat. Rev. Cancer*. **11**, 50–58
26. Wu, Y.-C., and Xue, D. (2003) Programmed Cell Death in C. elegans. in *Essentials of Apoptosis*, pp. 135–144, Humana Press, Totowa, NJ, 10.1007/978-1-59259-361-3_9
27. Malin, J. Z., and Shaham, S. (2015) Cell Death in C. elegans Development. in *Current Topics in Developmental Biology*, pp. 1–42, **114**, 1–42
28. Nehme, R., and Conradt, B. (2008) egl-1: A key activator of apoptotic cell death in C. elegans. *Oncogene*. **27**, S30–S40
29. Peden, E., Killian, D. J., and Xue, D. (2008) Cell death specification in C. elegans. *Cell Cycle*. **7**, 2479–2484
30. Mangahas, P. M., and Zhou, Z. (2005) Clearance of apoptotic cells in Caenorhabditis elegans. *Semin. Cell Dev. Biol.* **16**, 295–306
31. Irmler, M., Hofmann, K., Vaux, D., and Tschopp, J. (1997) Direct physical interaction between the Caenorhabditis elegans “death proteins” CED-3 and CED-4. *FEBS Lett.* **406**, 189–190
32. Shaham, S., and Horvitz, H. R. (1996) Developing Caenorhabditis elegans neurons may contain both cell-death protective and killer activities. *Genes Dev.* **10**, 578–591
33. Chen, F., Hersh, B. M., Conradt, B., Zhou, Z., Riemer, D., Gruenbaum, Y., and Horvitz, H. R. (2000) Translocation of C. elegans CED-4 to nuclear membranes during programmed cell death. *Science (80-.).* **287**, 1485–1489
34. Wang, X., and Yang, C. (2016) Programmed cell death and clearance of cell corpses in Caenorhabditis elegans. *Cell. Mol. Life Sci.* **73**, 2221–2236
35. Roschitzki-Voser, H., Schroeder, T., Lenherr, E. D., Frölich, F., Schweizer, A., Donepudi, M., Ganesan, R., Mittl, P. R. E., Baici, A., and Grütter, M. G. (2012) Human caspases in vitro: Expression, purification and kinetic characterization. *Protein Expr. Purif.* 10.1016/j.pep.2012.05.009

36. Conradt, B., and Horvitz, H. R. (1999) The TRA-1A Sex Determination Protein of C. elegans Regulates Sexually Dimorphic Cell Deaths by Repressing the egl-1 Cell Death Activator Gene. *Cell*. **98**, 317–327
37. Hill, M. E., Macpherson, D. J., Wu, P., Julien, O., Wells, J. A., and Hardy, J. A. (2016) Reprogramming Caspase-7 Specificity by Regio-Specific Mutations and Selection Provides Alternate Solutions for Substrate Recognition. *ACS Chem. Biol.* **11**, 1603–1612
38. Julien, O., Zhuang, M., Wiita, A. P., O'Donoghue, A. J., Knudsen, G. M., Craik, C. S., and Wells, J. A. (2016) Quantitative MS-based enzymology of caspases reveals distinct protein substrate specificities, hierarchies, and cellular roles. *Proc. Natl. Acad. Sci. U. S. A.* **113**, E2001–E2010
39. ZUCKERKANDL, E., and PAULING, L. (1965) Evolutionary Divergence and Convergence in Proteins. in *Evolving Genes and Proteins*, 10.1016/B978-1-4832-2734-4.50017-6
40. Zuckerkandl, E., and Pauling, L. (1965) Molecules as documents of evolutionary history. *J. Theor. Biol.* **8**, 357–366
41. Harms, M. J., and Thornton, J. W. (2013) Evolutionary biochemistry: Revealing the historical and physical causes of protein properties. *Nat. Rev. Genet.* **14**, 559–571
42. Merkl, R., and Sterner, R. (2016) Ancestral protein reconstruction: Techniques and applications. *Biol. Chem.* **397**, 1–21
43. Furukawa, R., Toma, W., Yamazaki, K., and Akanuma, S. (2020) Ancestral sequence reconstruction produces thermally stable enzymes with mesophilic enzyme-like catalytic properties. *Sci. Rep.* **10**, 15493
44. Lettre, G., and Hengartner, M. O. (2006) Developmental apoptosis in C. elegans: a complex CEDnario. *Nat. Rev. Mol. Cell Biol.* **7**, 97–108
45. Risso, V. A., Gavira, J. A., Mejia-Carmona, D. F., Gaucher, E. A., and Sanchez-Ruiz, J. M. (2013) Hyperstability and substrate promiscuity in laboratory resurrections of precambrian β-lactamases. *J. Am. Chem. Soc.* **135**, 2899–2902
46. Risso, V. A., Gavira, J. A., and Sanchez-Ruiz, J. M. (2014) Thermostable and promiscuous Precambrian proteins. *Environ. Microbiol.* **16**, 1485–1489
47. Pandya, C., Farelli, J. D., Dunaway-Mariano, D., and Allen, K. N. (2014) Enzyme promiscuity: Engine of evolutionary innovation. *J. Biol. Chem.* **289**, 30229–30236
48. Jensen, R. A. (1976) Enzyme Recruitment in Evolution of New Function. *Annu. Rev. Microbiol.* **30**, 409–425
49. Chaudhuri, J., Parihar, M., and Pires-daSilva, A. (2010) An introduction to worm lab: From culturing worms to mutagenesis. *J. Vis. Exp.* 10.3791/2293

Chapter 5

Discussion and conclusions

Our study of two reef-building corals, *Orbicella faveolata* and *Porites astreoides*, has confirmed the presence of multiple copies of caspases in corals. Biochemical characterization of caspases from these two species proved that corals have both initiator and effector apoptotic caspases. Although this study did not characterize, *O. faveolata* also contains caspase-8-like protein with death effector domains (DED), suggesting the existence of extrinsic apoptotic pathways in corals. Similarly, other studies in coral reported the presence of many caspases and other apoptotic components like in humans and other vertebrates (1–5). Studies also showed that evolution of caspase was strongly influenced by gene duplication and gene loss (6, 7). Gene loss could be why model organisms *C. elegans* and *Drosophila* have fewer caspases than some coral species and vertebrates. For example, a primitive species *Hydra magnipapillata* consists of 17 caspases in total (4), *Danio rerio* consists of three copies of caspase-3 (8), *Lethenteron reissneri* consists of two copies of caspase-3 and three copies of caspase-7 (9). Nevertheless, all these findings suggest (i) Cnidarian apoptotic network is complex like vertebrates and can be a model to study evolution caspase and apoptosis overall (ii) It emphasizes the view of complex apoptotic pathways evolved early in metazoan evolution and that they may have contributed significantly to stabilizing the multicellular state (iii) Simple network seen in present-day nematodes and insects are misleading and hints that complex apoptotic pathways on vertebrates are not due to a gradual increase in network complexity.

Furthermore, coral caspases characterized in our study have two types of substrate specificities: DxxDase and VxxDase, similar to the human apoptotic caspase that emphasize functional conservation. Based on similar domain organization and substrate specificity of initiator caspases, PaCasp7a and OfCasp3a with vertebrate initiator caspases, we can speculate that coral apoptotic cascade is initiated either by forming apoptosome or PIDDosome. It suggests that coral response towards both death ligand and metabolic changes and confirms the presence of extrinsic and intrinsic apoptotic pathways. We also reported the structure of PaCasp7a, which confirms that the caspase-hemoglobinase fold conservation for millions of years. Interestingly, we found a potential regulatory exosite in the structure of PaCasp7a which is conserved in caspases of other species such as *C. elegans*, *Drosophila*, human caspase-2, giving another line of evidence of the conservation of structure and function.

In another study, we investigated how the folding landscape of caspases evolved using ancestral state reconstruction (ASR) and determining the equilibrium folding/unfolding process. Folding properties of a common ancestor of effector caspases (PCP-CA) and extant human effector caspases show that the folding landscape is conserved for over 650 million years. The data show the folding and stability of PCP7, which is evolutionarily more similar to the ancestor, remains comparable to the common ancestor (10). At the same time, PCP3 and PCP6 dimers become more stable, overall reflecting higher stability. At optimal pH, all proteins demonstrated a three-state equilibrium unfolding process with either monomeric or dimeric intermediate. The folding intermediate present in PCP6 is a dimer while a monomer in PCP7 and PCP-CA. A previous study in the folding landscape of PCP3

showed the four-state unfolding pathway with a dimer and a monomeric intermediate (11). So, we suggest that the two intermediates are most likely present in the ancestor. However, only the monomeric intermediate is sufficiently stable to appear in our spectroscopic assays. In addition, the intermediates may have been preferentially stabilized during evolution, such that PCP6 stabilized the dimeric intermediate while PCP7 retained the monomer as the most stable intermediate. In PCP3, both the partially folded monomer and dimer were stabilized sufficiently to populate both species. However, it is not clear when the changes occurred in the evolution of each subfamily and how these changes affect their functions.

In the pH study, at pH 4 significant fraction of PCP6 remains dimeric with 27.6 kcal/mol total free energy change (ΔG°conf), while PCP3, PCP7, and PCP-CA mostly dissociated to the monomers with ~6 kcal/mol free energy change (12). This difference suggests that caspase-6 could be the main executioner caspase in a low pH environment instead of caspase-3. Caspase-6 having more specialized roles in specific physiological contexts than caspase-3 and -7 would support the idea of caspase-6 being a major effector at low pH. Furthermore, the study of the difference in the dimeric interface among effector caspases could answer why caspase-6 remains dimer at lower pH. Proper cellular function of proteins is related to the stability and structural conformations; therefore, selective pressures that shape the protein folds may have resulted in the adaptive advantage to the caspase-6 to be stable at low pH. Additionally, pH-dependent stability difference among the proteins illustrates energy landscape contributes to the physicochemical mechanisms of proteins.

In the study of evolution substrate specificity, we used ASR techniques to map the substrate specificity and activity of the common ancestor of effector caspases (AncCP-Ef) to the extant human caspases with four nodes in between. AncCP-Ef was promiscuous with low activity (10). After duplication to the caspase-6 lineage and caspase-3/7 lineage, proteins became more specific with increased catalytic efficiency while it again decreased in caspase-7 lineage after duplication to the caspase-3 and -7. At this time, we do not know which amino acids are responsible for activity and specificity change. Hill et al. altered the caspase-7 specificity to mirror the caspase-6 with just four mutations near the active site (13). Three of those mutated residues are not conserved in the promiscuous AncCP-EF, suggesting the importance of those residues on substrate specificity. Recent studies underscored that exosites and allosteric sites are critical to substrate recognition by several proteases (14, 15). Therefore, caspase-7 exosite or allosteric sites must be responsible for the change in specificity and increase or decrease in catalytic efficiency. It was clear from the study that caspase-7 is becoming less active and less specific than its ancestors. It could be due to the overlapping function of caspase-7 with the main executioner caspase-3. Together, the substitution of amino acids from the ancestor to the extant proteins over time has changed the functions of proteins to adapt to the cellular environment and execute their roles.

In summary, our study in coral showed that corals have complex apoptotic signaling cascades, similar to those of vertebrates suggesting that caspases are ancient and well conserved in metazoans. Given that corals are declining due to coral bleaching, further study of corals is vital to dissect the apoptotic pathways and study

apoptosis evolution. The study of the evolutionary folding landscape showed that the native fold of effector caspases is robust to sequence changes and has been maintained throughout >650 million years of evolution. A study of the evolution of caspase-7 specificity showed that enzyme specificity was established early in the evolution of caspase-7 and that allosteric regulation likely followed through subsequent evolution, which may have caused the changes in catalytic efficiency of modern protein. Overall, caspase is an attractive model to study protein evolution. Understanding caspase and its functions using multiple approaches would help us develop therapeutics against diseases caused by caspase disorders.

References

1. Dunn, S. R., Phillips, W. S., Spatafora, J. W., Green, D. R., and Weis, V. M. (2006) Highly conserved caspase and Bcl-2 homologues from the sea anemone Aiptasia pallida: Lower metazoans as models for the study of apoptosis evolution. *J. Mol. Evol.* **63**, 95–107

2. Pernice, M., Dunn, S. R., Miard, T., Dufour, S., Dove, S., and Hoegh-Guldberg, O. (2011) Regulation of apoptotic mediators reveals dynamic responses to thermal stress in the reef building coral acropora millepora. *PLoS One*. **6**, e16095

3. Moya, A., Sakamaki, K., Mason, B. M., Huisman, L., Forêt, S., Weiss, Y., Bull, T. E., Tomii, K., Imai, K., Hayward, D. C., Ball, E. E., and Miller, D. J. (2016) Functional conservation of the apoptotic machinery from coral to man: The diverse and complex Bcl-2 and caspase repertoires of Acropora millepora. *BMC Genomics*. **17**, 62

4. Lasi, M., Pauly, B., Schmidt, N., Cikala, M., Stiening, B., Käsbauer, T., Zenner, G., Popp, T., Wagner, A., Knapp, R. T., Huber, A. H., Grunert, M., Söding, J., David, C. N., and Böttger, A. (2010) The molecular cell death machinery in the simple cnidarian Hydra includes an expanded caspase family and pro- and anti-apoptotic Bcl-2 proteins. *Cell Res.* **20**, 812–825

5. Kvitt, H., Rosenfeld, H., Zandbank, K., and Tchernov, D. Regulation of Apoptotic Pathways by Stylophora pistillata (Anthozoa, Pocilloporidae) to Survive Thermal Stress and Bleaching. 10.1371/journal.pone.0028665

6. Eckhart, L., Ballaun, C., Hermann, M., VandeBerg, J. L., Sipos, W., Uthman, A., Fischer, H., and Tschachler, E. (2008) Identification of novel mammalian caspases reveals an important role of gene loss in shaping the human caspase repertoire. *Mol. Biol. Evol.* 10.1093/molbev/msn012

7. Zmasek, C. M., Zhang, Q., Ye, Y., and Godzik, A. (2007) Surprising complexity of the ancestral apoptosis network. *Genome Biol.* **8**, R226

8. Howe, K., Clark, M. D., Torroja, C. F., Torrance, J., Berthelot, C., Muffato, M., Collins, J. E., Humphray, S., McLaren, K., Matthews, L., McLaren, S., Sealy, I., Caccamo, M., Churcher, C., Scott, C., Barrett, J. C., Koch, R., Rauch, G.-J., White, S., Chow, W., Kilian, B., Quintais, L. T., Guerra-Assunção, J. A., Zhou, Y., Gu, Y., Yen, J., Vogel, J.-H., Eyre, T., Redmond, S., Banerjee, R., Chi, J., Fu, B., Langley, E., Maguire, S. F., Laird, G. K., Lloyd, D., Kenyon, E., Donaldson, S., Sehra, H., Almeida-King, J., Loveland, J., Trevanion, S., Jones, M., Quail, M., Willey, D., Hunt, A., Burton, J., Sims, S., McLay, K., Plumb, B., Davis, J., Clee, C., Oliver, K., Clark, R., Riddle, C., Eliott, D., Threadgold, G., Harden, G., Ware, D., Mortimer, B., Kerry, G., Heath, P., Phillimore, B., Tracey, A., Corby, N., Dunn, M., Johnson, C., Wood, J., Clark, S., Pelan, S., Griffiths, G., Smith, M., Glithero, R., Howden, P., Barker, N., Stevens, C., Harley, J., Holt, K., Panagiotidis, G., Lovell, J., Beasley, H., Henderson, C., Gordon, D., Auger, K., Wright, D., Collins, J., Raisen, C., Dyer, L., Leung, K., Robertson, L., Ambridge, K., Leongamornlert, D.,

McGuire, S., Gilderthorp, R., Griffiths, C., Manthravadi, D., Nichol, S., Barker, G., Whitehead, S., Kay, M., Brown, J., Murnane, C., Gray, E., Humphries, M., Sycamore, N., Barker, D., Saunders, D., Wallis, J., Babbage, A., Hammond, S., Mashreghi-Mohammadi, M., Barr, L., Martin, S., Wray, P., Ellington, A., Matthews, N., Ellwood, M., Woodmansey, R., Clark, G., Cooper, J., Tromans, A., Grafham, D., Skuce, C., Pandian, R., Andrews, R., Harrison, E., Kimberley, A., Garnett, J., Fosker, N., Hall, R., Garner, P., Kelly, D., Bird, C., Palmer, S., Gehring, I., Berger, A., Dooley, C. M., Ersan-Ürün, Z., Eser, C., Geiger, H., Geisler, M., Karotki, L., Kirn, A., Konantz, J., Konantz, M., Oberländer, M., Rudolph-Geiger, S., Teucke, M., Osoegawa, K., Zhu, B., Rapp, A., Widaa, S., Langford, C., Yang, F., Carter, N. P., Harrow, J., Ning, Z., Herrero, J., Searle, S. M. J., Enright, A., Geisler, R., Plasterk, R. H. A., Lee, C., Westerfield, M., de Jong, P. J., Zon, L. I., Postlethwait, J. H., Nüsslein-Volhard, C., Hubbard, T. J. P., Crollius, H. R., Rogers, J., and Stemple, D. L. (2013) The zebrafish reference genome sequence and its relationship to the human genome. *Nature*. **496**, 498–503

9. Liu, Y., Xu, X., Wang, X., Zhu, T., Li, J., Pang, Y., and Li, Q. (2021) Analysis of the lamprey genotype provides insights into caspase evolution and functional divergence. *Mol. Immunol.* **132**, 8–20

10. Grinshpon, R. D., Shrestha, S., Titus-McQuillan, J., Hamilton, P. T., Swartz, P. D., and Clark, A. C. (2019) Resurrection of ancestral effector caspases identifies novel networks for evolution of substrate specificity. *Biochem. J.* **476**, 3475–3492

11. Bose, K., and Clark, A. C. (2001) Dimeric procaspase-3 unfolds via a four-state equilibrium process. *Biochemistry*. **40**, 14236–14242

12. Bose, K., and Clark, a C. (2005) pH effects on the stability and dimerization of procaspase-3. *Protein Sci.* **14**, 24–36

13. Hill, M. E., Macpherson, D. J., Wu, P., Julien, O., Wells, J. A., and Hardy, J. A. (2016) Reprogramming Caspase-7 Specificity by Regio-Specific Mutations and Selection Provides Alternate Solutions for Substrate Recognition. *ACS Chem. Biol.* **11**, 1603–1612

14. Boucher, D., Blais, V., and Denault, J.-B. (2012) Caspase-7 uses an exosite to promote poly(ADP ribose) polymerase 1 proteolysis. *Proc. Natl. Acad. Sci. U. S. A.* **109**, 5669–5674

15. Clark, A. C. (2016) Caspase Allostery and Conformational Selection. *Chem. Rev.* **116**, 6666–6706

Appendix-1

Folding study of common ancestor of caspase-7 (AncCP-7)

A protomer of AncCP-7 consists of 304 amino acids. It has two tryptophan residues, both in loop 3 and 11 tyrosine residues distributed in the primary sequence.

Equilibrium unfolding of extant and ancestor caspases

In this study, we examined the equilibrium unfolding of AncCP-7 at pH 7.5 and 6 by fluorescence emission and circular dichroism (pH 7.5 only) as a function of urea concentration. We collected fluorescence emission data of 4 different concentration (0.5, 1, 2, and 4 μM) at pH 7.5 and all other experiments with three concentrations (1, 2, and 4 μM). Renaturation experiments of all three proteins demonstrated the folding transitions are reversible.

At both pH 7.5 and 6, the data show little to no change in signal between 0 and ~2.5 M urea. One then observes a cooperative increase in the signal and demonstrates a plateau between ~3.5 and 5.5 M urea. A second cooperative transition occurs between ~5.5 M and 7 M urea (Figs. 1A-1C & Figs. 2A-2B). The fluorescence emission data are similar regardless of excitation at 280 nm or 295 nm at both pH. The cooperative transitions through CD signal are similar to the fluorescence signal with a notable difference. At first transition, CD signal decreased while it increased significantly in fluorescence emission. At all cases protein concentration dependence was observed in second transition. Overall, the data suggest cooperative changes in both the tertiary and secondary structures during each unfolding transition.

Global fitting of equilibrium unfolding data

The 11 different datasets at pH 7.5 and 6 different datasets at pH 6 were fitted globally to determine the free energy and the cooperativity indices (m-values) for each unfolding transition. The data were best fit to the three-state equilibrium model. In this model, the dimeric native conformation, N_2, isomerizes to a dimeric intermediate, I_2, and the dimeric intermediate dissociates and unfolds to monomers. The dissociation of I_2 to 2U leads to a protein-concentration dependent change in the mid-point of the second transition, as shown in (Figs. 1A-1C & Figs. 2A-2B). Based on this model, we have determined the conformational free energy, $\Delta G°_{conf}$, and the m-values for each step of unfolding. The solid lines in (Figs. 1A-1C & Figs. 2A-2B) are the results of global fits of the model to the data. The free energy change at pH 7.5, $\Delta G_1^{H_2O}$ and the cooperativity index, m_1, for the first step of unfolding, the isomerization of N_2 to I_2, are 14.7 ± 1.8 kcal/mol and 5.0 ± 0.6 kcal mol^{-1} M^{-1}, respectively (Table 1). The free energy change, $\Delta G_2^{H_2O}$, and cooperativity index, m_2, for the dissociation and complete unfolding of the dimeric intermediate to two unfolded monomeric proteins ($I_2 \rightleftarrows 2U$) are 28.3 ± 0.6 kcal/mol and 3.3 ± 0.1 kcal mol^{-1} M^{-1}, respectively. At pH 6, $\Delta G_1^{H_2O}$ and m_1 decreased to half while $\Delta G_2^{H_2O}$ and m_2 remains similar (Table 1). Overall, the data demonstrate that AncCP-7 unfolds by three-state process with dimeric intermediate and is very stable, with the total conformational free energy of 43 kcal/mol at pH 7.5 which decreased slightly to 36.4 kcal/mol at pH 6 (Table 1).

For both pH, we calculated the equilibrium distribution of species over the urea concentration using the values of the free energies, the cooperativity indices determined for each transition. The fraction of species corresponding to the global fits are shown in

Figs. 1D & 2C. At pH 7.5, protein remains in native form up to 2 M urea while it started to dissociate into intermediate at pH 6. The mid-point of first transition is ~3 M at pH 7.5 which decreased slightly at pH6 while in second transition mid-point is similar at both pH (~6 M urea). With the addition of ~2 M urea native protein decreased cooperatively with a concomitant increase in the partially folded intermediate population between 2 and ~3.5 M urea. The dimeric intermediate reaches a maximum at ~3.5 M urea and remains predominant up to ~5 M urea (Figs. 1D & 2C). Furthermore, the protein was completely unfolded by 7 M urea.

Table 3: Thermodynamic parameters of each step of folding/unfolding of AncCP-7.

pH	Free energy changes ΔG^{o}_{conf}			m-values		
	$\Delta G_1^{H_2O}$ (kcal/mol)	$\Delta G_2^{H_2O}$ (kcal/mol)	$\Delta G_{total}^{H_2O}$ (kcal/mol)	m_1 (kcal/mol/M)	m_2 (kcal/mol/M)	m_{total} (kcal/mol/M)
7.5	14.7 ± 1.8	28.3 ± 0.6	43 ± 2.4	5.0 ± 0.6	3.3 ± 0.1	8.3 ± 0.7
6	7.2 ± 1.6	29.2 ± 1.7	36.4 ± 3.3	2.7 ± 0.6	3.5 ± 0.3	6.2 ± 0.9

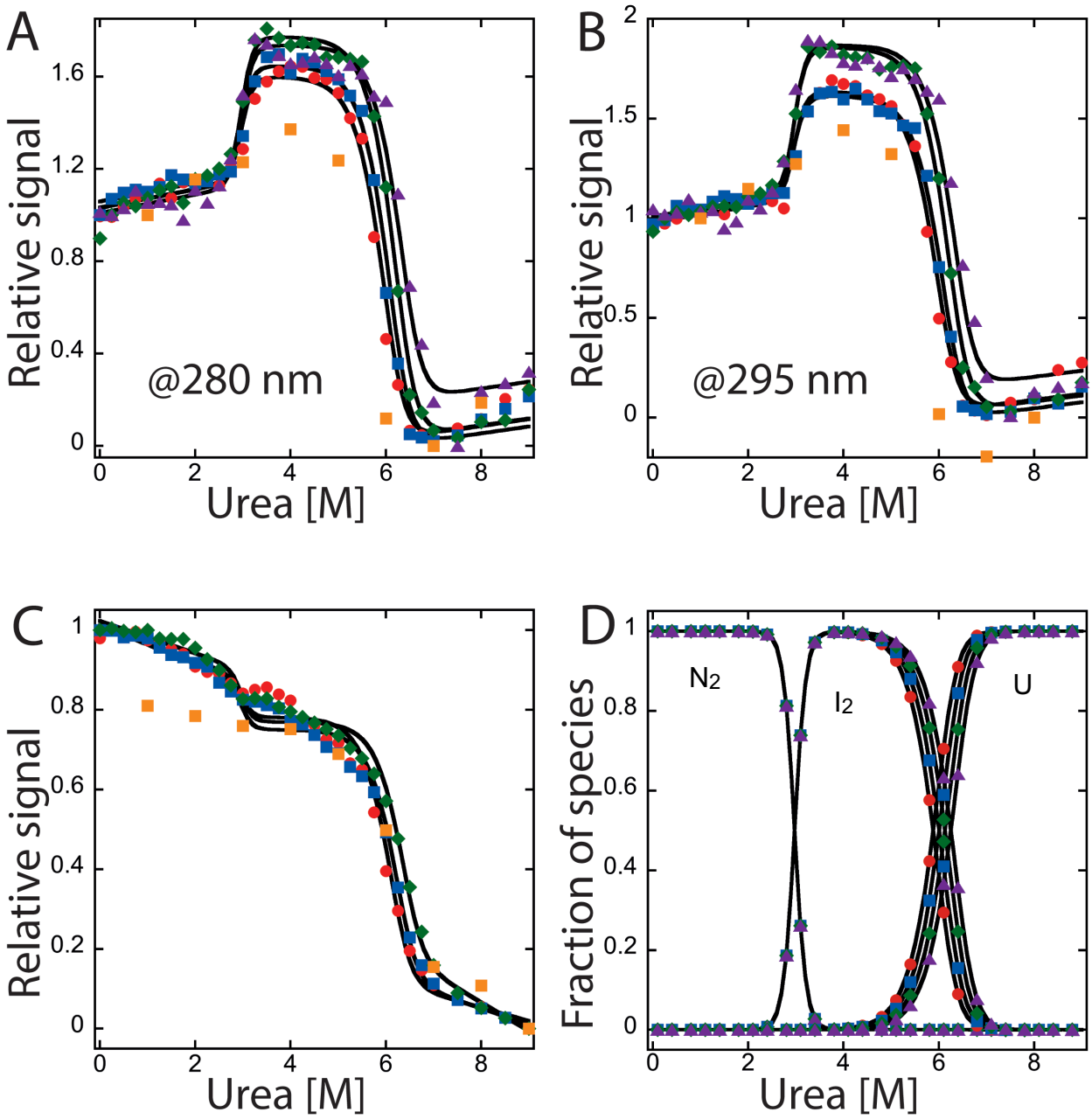


Figure 1: Equilibrium unfolding of AncCP-7 at pH 7.5. Relative signal of fluorescence emission with excitation at 280 nm (**A**), 295 nm (**B**) and CD (**C**). Fraction of species as a function of urea (**D**). Colored solid symbols represent raw data and corresponding solid lines represent the global fits of the data in an appropriate model described in text. Symbols are represented as follows. 0.5 μM (●), 1 μM (■), 2 μM (◆), and 4 μM (▲). Orange square (■) represents refolding data of 1 μM protein. N_2 refers to dimeric native protein, I_2 is a dimeric intermediate, and U refers to unfolded species.

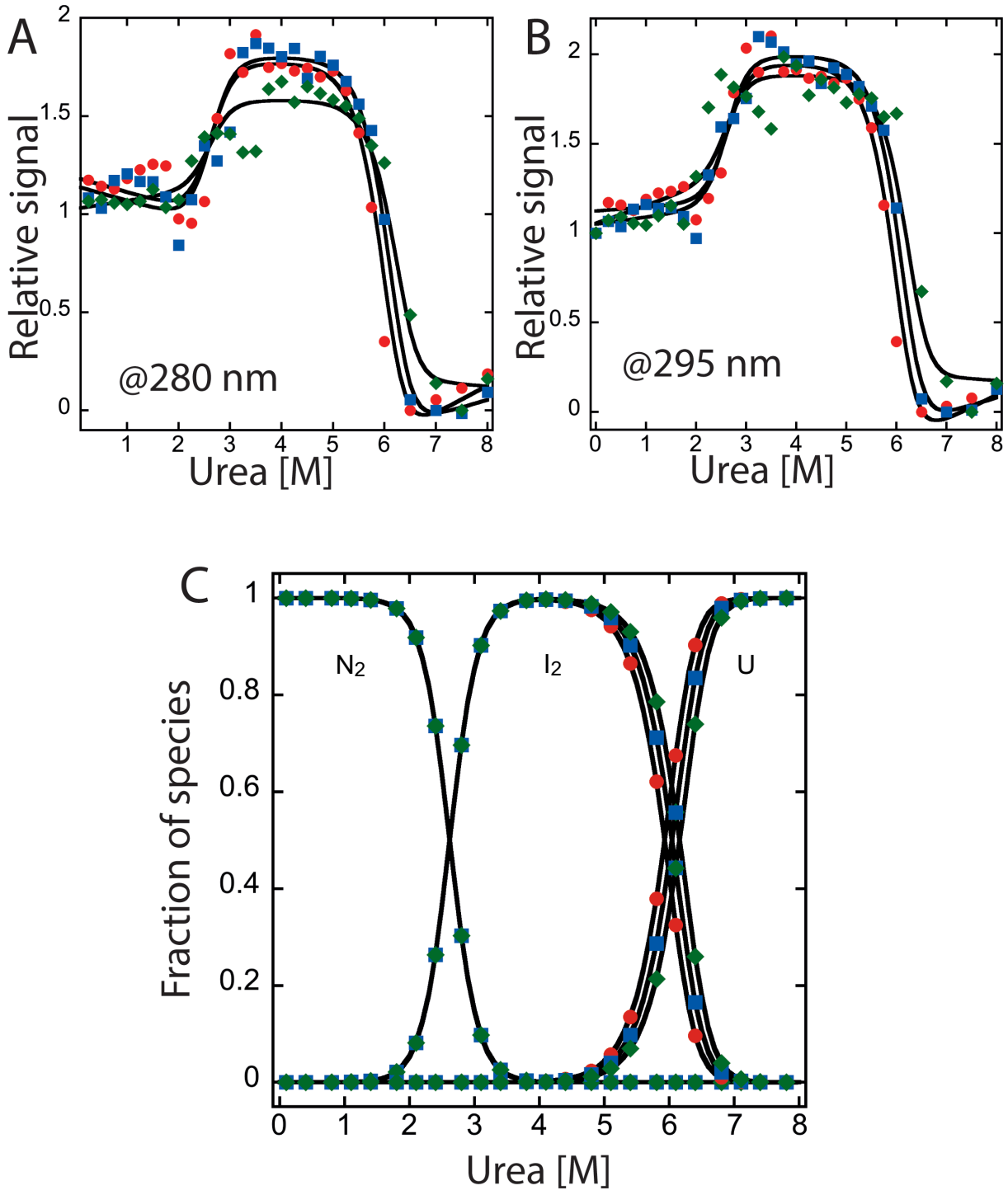


Figure 2: Equilibrium unfolding of AncCP-7 at pH 6.5. Relative signal of fluorescence emission with excitation at 280 nm (**A**), and 295 nm (**B**). Colored solid symbols represent raw data and corresponding solid lines represent the global fits of the data in an appropriate model described in text. Fraction of species as a function of urea concentrations (**C**). Symbols and colors are consistent with previous figure.