Investigation of the Evolution and Structural Basis of Caspases: Insights into Conformational Dynamics, Stability and Allostery

by

Isha Uday Joglekar

Dissertation Presented to the Graduate Faculty at The University of Texas at Arlington in Partial Fulfilment of the Requirements for the Degree of

Doctor of Philosophy

May 2023

Principal Investigator: Dr. Allan Clay Clark

Supervising Committee:

Dr. Joseph M. Boll

Dr. Matthew K. Fujita

Dr. Kayunta Johnson-Winters

Dr. Laura Mydlarz

Acknowledgements

To my advisor, Dr. Clay Clark, I am deeply grateful for your unwavering support and mentorship throughout my graduate studies. You have not only taught me the principles of good science but also the importance of effective communication. As Ben Feringa, Nobel Prize in Chemistry 2016, said, "When I draw a molecule in China or Argentina, it is the same molecule. People understand immediately without knowing Chinese or Spanish. That is beautiful." Thank you for your tireless efforts in helping me become a successful scientist and communicator. Your dedication to the scientific process and your passion for discovery have been inspirational, and I will always be grateful for the countless hours you spent helping me develop my research skills and scientific intuition. Above all, thank you for teaching me not only how to be a better scientist, but also a better human. Your kindness, compassion, and integrity have been a shining example to me, and I am honored to have had you as my advisor and mentor. I will always remember the lessons and guidance you have provided me, regardless of where my journey in science takes me next.

I would also like to thank my dissertation committee members Dr. Joseph Boll, Dr. Matthew Fujita, Dr. Kayunta Johnson-Winters and Dr. Laura Mydlarz for providing their time, effort, and expertise. Your invaluable feedback, guidance, and support have been instrumental in shaping my research and has helped me improve the quality of my work. I am especially grateful for your constructive criticism, which has challenged me to think deeply about my research and to consider different perspectives and approaches. Without your contributions, this dissertation would not have been possible. Thank you all for your time, patience, and dedication to my success. I would also like to thank Dr. Kimberly Bowles for all the help during the tough times with cloning, and her constant support for anything and everything throughout this journey.

I would also like to extend my appreciation to the former lab members Dr. Suman Shrestha and Dr. Liqi Yao, who taught me the fundamental techniques and experiments that formed the foundation of my research. I appreciate their patience and willingness to share their expertise, and I will always be grateful for their contributions to my research and training. I would also like to thank Mithun Nag for the insightful discussions, motivation, and support during my Ph.D. program. He has been an essential part of my academic journey, and I appreciate the countless hours we spent together in the lab, brainstorming ideas, and tackling challenges. I am grateful for his contributions to my academic and personal growth, and I cherish our friendship that we have developed along the way. Thanks to David, Jessica, Melissa and Will for the encouragement and contributions to my research. The camaraderie and support of the Clark lab have not only made my time in the lab enjoyable, but also highly productive.

And lastly I'd like to thank my wonderful roommates and friends, who have been my family here in the United States, providing me with support and encouragement as I pursued my academic goals. I want to express my appreciation for our engaging and thought-provoking physics and data science discussions. Your friendship and support have been invaluable to me throughout my academic journey. Thank you for being an amazing and supportive family, and for contributing to my academic and personal growth in immeasurable ways. I will always cherish the memories we have shared and the lessons I have learned from each one of you.

Dedication

This work is dedicated to my incredibly supportive family. Your love and support have been the pillars of my life, and for that, I am forever grateful. Your constant belief in me has been the driving force behind my success, so thank you for being my rock and support system. This is not just my achievement but ours, and I could not have done it without you.

Abstract

Investigation of the Evolution and Structural Basis of Caspases: Insights into Conformational Dynamics, Stability and Allostery

Isha Uday Joglekar

The University of Texas at Arlington, 2023

Supervising Professor: Dr. A. Clay Clark

Caspases, a class of cysteinyl proteases, are well known for their roles in apoptosis or programmed cell death. However, at lower levels of activation in the body, they also play other non-apoptotic roles in cell proliferation, cell differentiation, neuronal development, cell migration and motility, and embryonic development, all of which are poorly understood. Caspases provide an excellent model to study protein evolution since, they evolved from a common ancestral state into two subfamilies having different oligomeric properties: initiators, which exist as monomers, and effectors, which exist as dimers, under physiological conditions in a cell. However, the evolutionary basis for the divergence of the caspase-hemoglobinase fold into these subfamilies is not well understood.

Additionally, post-translational modifications (PTMs) are essential for tight regulation of caspases in cell death, and PTMs can significantly alter caspase function and promote non-apoptotic signaling pathways. Understanding these mechanisms is crucial for identifying new therapeutic targets for drug development in apoptotic cell death and associated disorders. While some PTMs such as phosphorylation and ubiquitination are well-established, the significance of others, including acetylation, are still being investigated. Understanding the mechanisms of PTMs in caspases could lead to the identification of new therapeutic targets for drug development, making this a crucial and emerging field for obtaining insights into the regulation of apoptotic cell death and associated disorders, as well as the development of effective therapies.

This study focuses on the folding landscape of monomeric caspases from two coral species, which diverged from each other about 300 million years ago and from

human caspases about 600 million years ago. The research shows that both coral caspases have high stability in the physiological pH range of 6 to 8, and they unfold via two partially folded intermediates, which are in equilibrium with the native and the unfolded state. Molecular dynamics simulations, limited proteolysis, and MALDI-TOF mass spectrometry indicate that the small subunit of the monomeric caspases is unstable and unfolds before the large subunit. Similar to the dimeric caspases, the monomeric coral caspases undergo a pH-dependent conformational change resulting from the titration of an evolutionarily conserved site. The findings suggest that all caspases share a conserved folding landscape, and a conserved allosteric site can be fine-tuned for species-specific regulation. Further, instability of the small subunit of the initiator subfamily highlights the significance of the evolution of the effector subfamily of caspases, wherein the small subunit is stabilized.

This study also investigates the relationship between modest evolutionary changes in the hydrophobic core and conformational dynamics of the caspase family. Ancestral reconstruction, molecular dynamics simulations, network analysis, and free energy landscapes identify a conserved scaffold that underlies the diverse conformational dynamics of caspases across the effector and the initiator subfamilies in chordates. This study demonstrates how modular modifications in the biophysical properties of amino acids that connect the conserved scaffold to other structural elements can give rise to dynamic conformations and highlight the energetic basis of conformational fluctuations that guide the evolution of monomeric and dimeric scaffolds in protein families. Further, a conserved network of residues crucial for allosteric communication is identified. This network of residues that has been conserved over a billion years suggests that although the residues that are post-translationally modified may necessarily not be conserved across caspases, their spatial positions are, and may utilize this conserved network of residues to communicate with the active site to alter caspase function. This knowledge will help us understand how caspases fold to contribute to the fitness of an organism while maintaining or altering the energy landscape that is guided by the trade-offs between protein stability and activity.

Key words: Apoptosis; protein evolution; caspases; protein folding.

Table of Contents

CHAPTER 1

Introduction: Overview of Caspases

It is estimated that a healthy adult body can produce 10^{10} cells per day and the same number of cells are required to be removed for maintaining homeostasis, which primarily occurs by apoptosis or autophagy. Apoptosis, a type of programmed cell death is carried out by caspases, which belong to a class of cysteinyl proteases that play an integral part in the cell development and apoptotic cell death as an evolutionarily conserved function.[1,2]

Apoptotic cell death is a defining characteristic of metazoans; however, the number of caspases identified is distinct for each species. The molecular mechanism of apoptosis was first discovered and studied in C. elegans, which then became one of the most popular models for studying apoptosis. Homologs of these CASP genes were eventually identified in D. melanogaster, gaining popularity as a model organism, and were also identified in all the vertebrates, wherein a much larger apoptotic network was identified in humans. [3,4]

Although the use of C. elegans and D. melanogaster as model organisms have provided key insights in establishing the molecular basis of apoptosis, the disparate cell death pathways have obscured its evolutionary origins.[5,6] However, genomic studies of cnidaria, the sister group to bilateria comprising of corals, jellyfish, sea anemone, revealed various genes which were earlier thought to be confined to the vertebrates, thereby demonstrating the extensive gene loss in C. elegans and D. melanogaster that accounts for the resulting apoptotic pathway which does not reflect the characteristics of

the ancestral metazoans. Since the cnidarians appear to have a full complement of apoptotic signaling cascade, the apoptotic pathways of cnidarians can explain the evolution of the vertebrate apoptotic network (Fig 1A).[6–10]Caspases are ubiquitous in metazoans and predate the cnidarian-bilaterian divergence which was about 600 million years ago. In view of the above, an emerging model to study caspases are reef-building corals of the order Scleractinia. Corals have a primitive immune system that includes only the innate immune pathways, but no adaptive immune pathways. The vertebrate apoptotic machinery appears to have retained several characteristics from its sister clade and research suggests that cnidarian apoptotic pathways may be as similar to and as complex as the vertebrate pathways.[11–13]

Over the course of evolution, each caspase has developed unique regulatory mechanisms like post-translational modification sites, yet, how these allosteric sites affect the activity of these enzymes is not well understood. Moreover, how caspases fold to the native conformation to contribute to the fitness of the organism while maintaining or altering the energy landscape that is guided by trade-offs between the stability and the activity of the protein remains elusive. [14–16]

Through hundreds of millions of years of evolution, the caspase family has developed common and unique allosteric sites, as well as conserved active sites due to overlapping substrate profiles. As a result of this, caspases provide as an excellent model system for studying protein evolution. Caspases evolved from an ancestral immune system into two different classes- inflammatory or apoptotic caspases. These apoptotic caspases further evolved into two subfamilies from an ancient, conserved caspase-hemoglobinase (CH) fold that has been retained for hundreds of millions of

2

years. The two subfamilies are characterized as either initiator caspases or effector caspases depending on their entry in the cell death cascade. Further, due to gene duplication events, an ancestor of initiators gave rise to four caspases (-8,-10,-18,cFLIP), whereas an ancestor of effectors gave rise to three caspases (-3,-6,-7) (Fig 1B).[1,17–19]

All caspases exist as inactive zymogens in the cell and oligomerization is vital for regulation of these enzymes. The caspase zymogens have an N-terminal pro-domain and a large and a small subunit which are connected by an intersubunit linker. Initiator caspases are present as monomers that require dimerization for complete activation, whereas the effector caspases are present as dimeric zymogens that require processing for complete activation (Fig 2). Dimerization is central for the activation of caspases since, the active site of these enzymes is comprised of loops from both monomers. The activation mechanisms are considerably well known with regards to the cleavage of the polypeptide chain and the subsequent rearrangements in the active site of the dimer. These principles governing the intramolecular interactions of monomers hold true for dimers as well along with the additional intramolecular interactions that are provided by the dimer interface. Although the evolutionary mechanisms in dimer formation are not very well understood, it is known that new protein-protein interactions involve amino acid substitutions or insertion/deletion of residues. Moreover, caspases have maintained specificity at the P1 site for aspartate residues in substrate proteins owing to evolutionary constraints, and thus caspases evolved allosteric mechanisms for fine tuning their activity.[1,17–21]

Post-translational modifications (PTMs) have a critical and essential role in the regulation of cell death. Caspases are modulated by PTMs in a highly dynamic manner. Following modification, the structure and accessibility of caspases may be altered, resulting in significant changes to their function, and even promoting non-apoptotic signaling pathways. Additionally, the interaction and competition between different PTMs, such as ubiquitination, SUMOylation, S-nitrosylation, and S-glutathionylation, as well as the degree of caspase modification, are crucial factors in determining the function of caspases. While the role of some PTMs, such as phosphorylation and ubiquitination, is well-established, the significance of others, including acetylation, is still being investigated. Understanding the mechanisms of PTMs in caspases could lead to the identification of new therapeutic targets for drug development, making this a crucial and emerging field for obtaining insights into the regulation of apoptotic cell death and associated disorders, as well as the development of effective therapies.[22–27]

Although caspases are well known for their role in apoptosis, they also have numerous other fate-determining roles in cell differentiation, neuronal remodeling, and inflammation.[28–31] Several autoimmune diseases, neurodegenerative disorders, and cancers frequently involve dysregulated apoptosis as a contributing factor. [32–38]The development of therapeutic strategies for these conditions may result from an understanding of how to selectively regulate the level of apoptosis by activating or inhibiting caspases. Due to the evolution of structural and functional differences, the monomeric and the dimeric caspases are regulated differently and play specific roles under different homeostatic conditions.[1,39,40]

Studying protein conformational landscapes is essential for understanding the relationship between protein evolution and function. The outdated notion of a single native structure has been replaced by the ensemble model, which recognizes that proteins sample a variety of conformations through local and global changes. Allostery is a crucial emergent property of this ensemble model, allowing proteins to shift their dynamics and access certain conformers for regulation. Evolution has also been linked to the ensemble view, wherein the same conserved three-dimensional native fold can adopt new functions by modulating the conformational sample space.[41–45] Exploring the relationship between protein evolution and conformational dynamics, can offer a more comprehensive understanding of the principles behind the emergence of new functions and how proteins adapt to new environments.

The study of conformational landscapes of protein families like caspases is particularly crucial as dysregulation of caspase activity has been linked to several disease states. However, the rational design of effective caspase inhibitors is complicated by the highly dynamic nature of these proteins, which can adopt different conformations with varying enzymatic activities and susceptibility to inhibition by small molecules.[46–49]

Recent advances in structural biology techniques have provided detailed information about the three-dimensional structure of caspases. However, understanding the dynamic behavior of caspases and their interactions with small molecule inhibitors requires a more comprehensive approach that combines experimental data with computational simulations.[50–53] By adopting this approach, we can acquire insights into the mechanisms behind protein function, evolution, and adaptation, and can better

understand the complex biological systems and develop new therapeutic interventions to address various diseases.

References

1. Clark AC (2016) Caspase allostery and conformational selection. Chem Rev 11:6666–6706.

2. Sakamaki K, Satou Y (2009) Caspases: evolutionary aspects of their functions in vertebrates. J Fish Biol 74:727.

3. Zmasek CM, Godzik A (2013) Evolution of the Animal Apoptosis Network. Cold Spring Harb Perspect Biol 5.

4. Ellis HM, Horvitz HR (1986) Genetic control of programmed cell death in the nematode C. elegans. Cell 44:817–829.

5. Irmler M, Hofmann K, Vaux D, Tschopp J (1997) Direct physical interaction between the Caenorhabditis elegans "death proteins" CED-3 and CED-4. FEBS Lett 406:189–190.

6. Mydlarz LD, Fuess L, Mann W, Pinzón JH, Gochfeld DJ (2016) Cnidarian immunity: From genomes to phenomes. The Cnidaria, past, present and Future: The World of Medusa and her Sisters 441–466.

7. Tchernov D, Kvitt H, Haramaty L, Bibbyd TS, Gorbunov MY, Rosenfeld H, Falkowski PG (2011) Apoptosis and the selective survival of host animals following thermal bleaching in zooxanthellate corals. Proc Natl Acad Sci 108:9905–9909.

8. Wiens M, Krasko A, Perovic S, Müller WEG (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. Richier S, Sabourault C, Courtiade J, Zucchini N, Allemand D, Furla P (2006) Oxidative stress and apoptotic events during thermal stress in the symbiotic sea anemone, Anemonia viridis. FEBS J 273:4186–4198.

10. Fuess LE, Mann WT, Jinks LR, Brinkhuis V, Mydlarz LD (2018) Transcriptional analyses provide new insight into the late-stage immune response of a diseased Caribbean coral. R Soc Open Sci 5.

11. Cleves PA, Shumaker A, Lee JM, Putnam HM, Bhattacharya D (2020) Unknown to Known: Advancing Knowledge of Coral Gene Function. Trends Genet 36:93–104.

12. Shrestha S, Tung J, Grinshpon RD, Swartz P, Hamilton PT, Dimos B, Mydlarz L, Clay Clark A (2020) Caspases from scleractinian coral show unique regulatory features. Journal of Biological Chemistry [Internet] 295:14578–14591.

13. Sakamaki K, Imai K, Tomii K, Miller DJ (2015) Evolutionary analyses of caspase-8 and its paralogs: Deep origins of the apoptotic signaling pathways. Bioessays 37:767–776.

14. Newton MS, Arcus VL, Gerth ML, Patrick WM (2018) Enzyme evolution: innovation is easy, optimization is complicated. Curr Opin Struct Biol 48:110–116.

15. Johansson KE, Lindorff-Larsen K (2018) Structural heterogeneity and dynamics in protein evolution and design. Curr Opin Struct Biol 48:157–163.

16. Modi T, Campitelli P, Kazan IC, Ozkan SB (2021) Protein folding stability and binding interactions through the lens of evolution: a dynamical perspective. Curr Opin Struct Biol 66:207–215.

17. Nag M, Clark AC (2023) Conserved folding landscape of monomeric initiator caspases. Journal of Biological Chemistry [Internet] 0:103075.

18. Shrestha S, Clark AC (2021) Evolution of the folding landscape of effector caspases. J Biol Chem 297,101249:1–12.

19. Yao L, Clark AC (2022) Comparing the folding landscapes of evolutionarily divergent procaspase-3. Biosci Rep 42.

20. Milam SL, Clark AC (2009) Folding and assembly kinetics of procaspase-3.

21. Bose K, Clark AC (2001) Dimeric procaspase-3 unfolds via a four-state equilibrium process. Biochemistry 40:14236–14242.

22. Parrish AB, Freel CD, Kornbluth S (2013) Cellular mechanisms controlling caspase activation and function. Cold Spring Harb Perspect Biol 5.

23. Zamaraev A V., Kopeina GS, Prokhorova EA, Zhivotovsky B, Lavrik IN (2017) Post-translational Modification of Caspases: The Other Side of Apoptosis Regulation. Trends Cell Biol 27:322–339.

24. Dagbay K, Eron SJ, Serrano BP, Velázquez-Delgado EM, Zhao Y, Lin D, Vaidya S, Hardy JA (2014) A multi-pronged approach for compiling a global map of allosteric regulation in the apoptotic caspases. Methods Enzymol 544:215.

25. Maciag JJ, Mackenzie SH, Tucker MB, Schipper JL, Swartz P, Clark AC Tunable allosteric library of caspase-3 identifies coupling between conserved water molecules and conformational selection. Proc Natl Acad Sci 113(41): E6080-E6088.

26. Cade C, Swartz P, MacKenzie SH, Clark AC (2014) Modifying caspase-3 activity by altering allosteric networks. Biochemistry 53:7582–7595.

27. Seyrek K, Ivanisenko N V., Richter M, Hillert LK, König C, Lavrik IN (2020) Controlling Cell Death through Post-translational Modifications of DED Proteins. Trends Cell Biol 30:354–369.

28. Schwerk C, Schulze-Osthoff K (2003) Non-apoptotic functions of caspases in cellular proliferation and differentiation. Biochem Pharmacol 66:1453–1458.

29. Hollville E, Deshmukh M (2018) Physiological functions of non-apoptotic caspase activity in the nervous system. Semin Cell Dev Biol 82:127.

30. Oberst A, Green DR (2011) It cuts both ways: reconciling the dual roles of caspase 8 in cell death and survival. Nature Reviews Molecular Cell Biology 2011 12:757–763.

31. Tsuchiya Y, Nakabayashi O, Nakano H, Lemarié A (2015) FLIP the Switch: Regulation of Apoptosis and Necroptosis by cFLIP. Int J Mol Sci 16(12):30321-30341.

32. Walters J, Pop C, Scott FL, Drag M, Swartz P, Mattos C, Salvesen GS, Clark AC (2009) A constitutively active and uninhibitable caspase-3 zymogen efficiently induces apoptosis. Biochemical Journal 424:335.

33. Olsson M, Zhivotovsky B (2011) Caspases and cancer. Cell Death Differ 18:1441.

34. Grunert M, Gottschalk K, Kapahnke J, Gü Ndisch S, Kieser A, Jeremias I (2012) The adaptor protein FADD and the initiator caspase-8 mediate activation of NF-κB by TRAIL. Cell Death Dis 3:414.

35. Zorn JA, Wille H, Wolan DW, Wells JA (2011) Self-Assembling Small Molecules Form Nanofibrils That Bind Procaspase-3 To Promote Activation. J. Am. Chem. Soc 133: 19630-19633.

36. Puck JM, Zhu S (2003) Immune disorders caused by defects in the caspase cascade. Curr Allergy Asthma Rep 3:378–384.

37. Mahajan A, Sharma G, Thakur K, Raza K, Singh G, Katare OP (2021) Autoimmune diseases and apoptosis: Targets, challenges, and innovations. Clinical Perspectives and Targeted Therapies in Apoptosis: Drug Discovery, Drug Delivery, and Disease Prevention:285–327.

38. You R, He X, Zeng Z, Zhan Y, Xiao Y, Xiao R (2022) Pyroptosis and Its Role in Autoimmune Disease: A Potential Therapeutic Target. Front Immunol 13:2531.

39. Shalini S, Dorstyn L, Dawar S, Kumar S (2015) Old, new and emerging functions of caspases. Cell Death Differ 22:526–539.

40. McIlwain DR, Berger T, Mak TW (2013) Caspase functions in cell death and disease. Cold Spring Harb Perspect Biol 5.

41. Murugan A, Husain K, B McLeish TC, Rodgers TL, Wilson MR (2013) Allostery without conformation change: modelling protein dynamics at multiple scales. Phys. Biol. 10 056004.

42. Tokuriki N, Tawfik DS (2009) Protein dynamism and evolvability. Science 324, pp 203–207.

43. Wodak SJ, Paci E, Dokholyan N V., Berezovsky IN, Horovitz A, Li J, Hilser VJ, Bahar I, Karanicolas J, Stock G, et al. (2019) Allostery in Its Many Disguises: From Theory to Applications. Structure 27:566–578.

44. Saavedra HG, Wrabl JO, Anderson JA, Li J, Hilser VJ (2018) Dynamic allostery can drive cold adaptation in enzymes. Nature 558, 324–328.

45. Kumar A, Butler BM, Kumar S, Ozkan SB (2015) Integration of structural dynamics and molecular evolution via protein interaction networks: a new era in genomic medicine. Curr Opin Struct Biol 35:135–142.

46. Fischer U, Schulze-Osthoff K Apoptosis-based therapies and drug targets. Cell Death Differ,12, pp 92-961.

47. Callus BA, Vaux DL (2006) Caspase inhibitors: viral, cellular and chemical. Cell Death & Differentiation 2007 14:1, pp 73–78.

48. Dhani S, Zhao Y, Zhivotovsky B (2021) A long way to go: caspase inhibitors in clinical use. Cell Death & Disease 2021 12:10, pp 1–13.

49. Ekert PG, Silke J, Vaux DL (1999) Caspase inhibitors. Cell Death Differ 6:1081–1086.

50. Bingöl EN, Serçinoğlu O, Ozbek P (2018) How do mutations and allosteric inhibitors modulate caspase-7 activity? A molecular dynamics study. Journal of Biomolecular Structure and Dynamics 37:3456–3466.

51. Sulpizi M, Rothlisberger U, Carloni P (2003) Molecular dynamics studies of caspase-3. Biophys J 84:2207–2215.

52. Campitelli P, Modi T, Kumar S, Ozkan SB (2020) The Role of Conformational Dynamics and Allostery in Modulating Protein Evolution.

53. Hu G, Doruker P, Li H, Demet Akten E (2021) Understanding Protein Dynamics, Binding and Allostery for Drug Design. Front Mol Biosci 8:250.

FIGURES

Figure 1
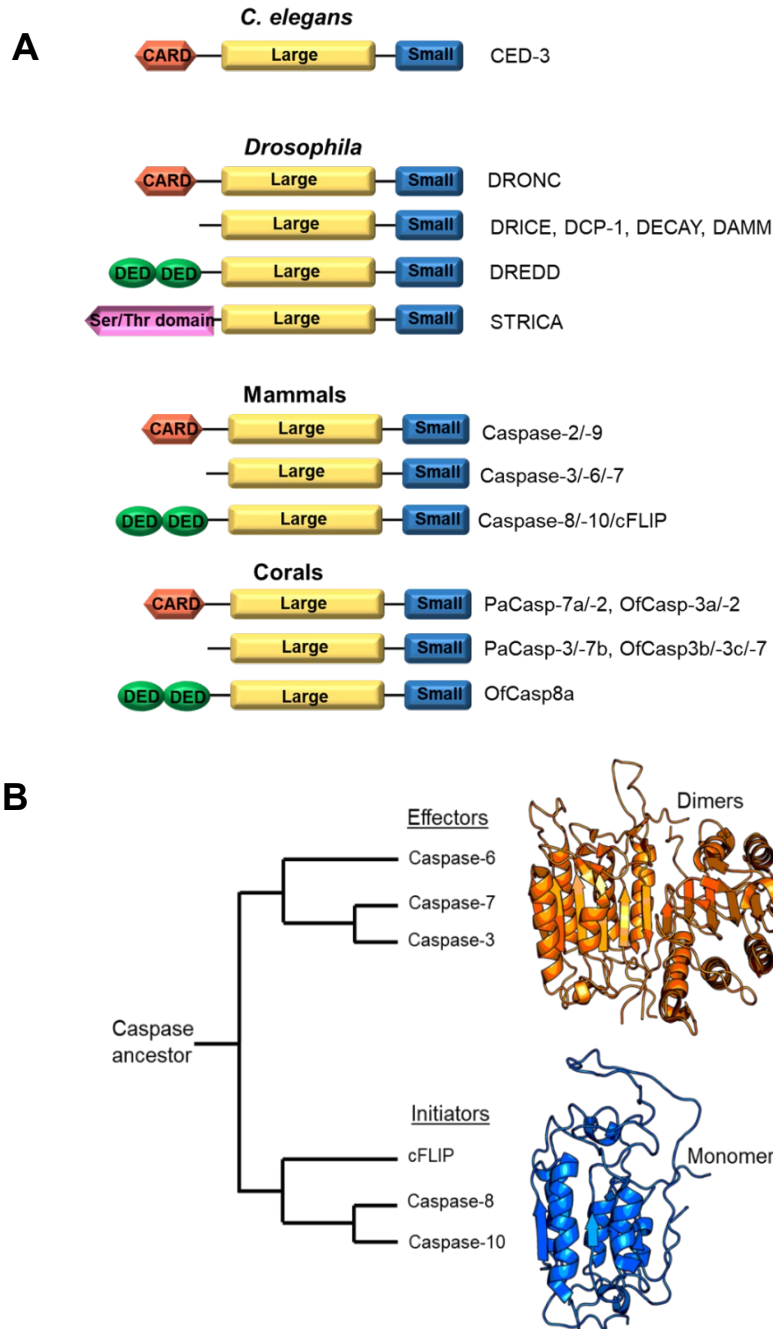
Figure 1. Domain organization and evolution of caspase sub-families. (A) Number of caspases and their domain organization in each of the species. The mammalian caspase repertoire is retained in caspases from corals. (B) All caspases evolved from a common ancestor into the initiator and the effector sub-families.

Figure 2
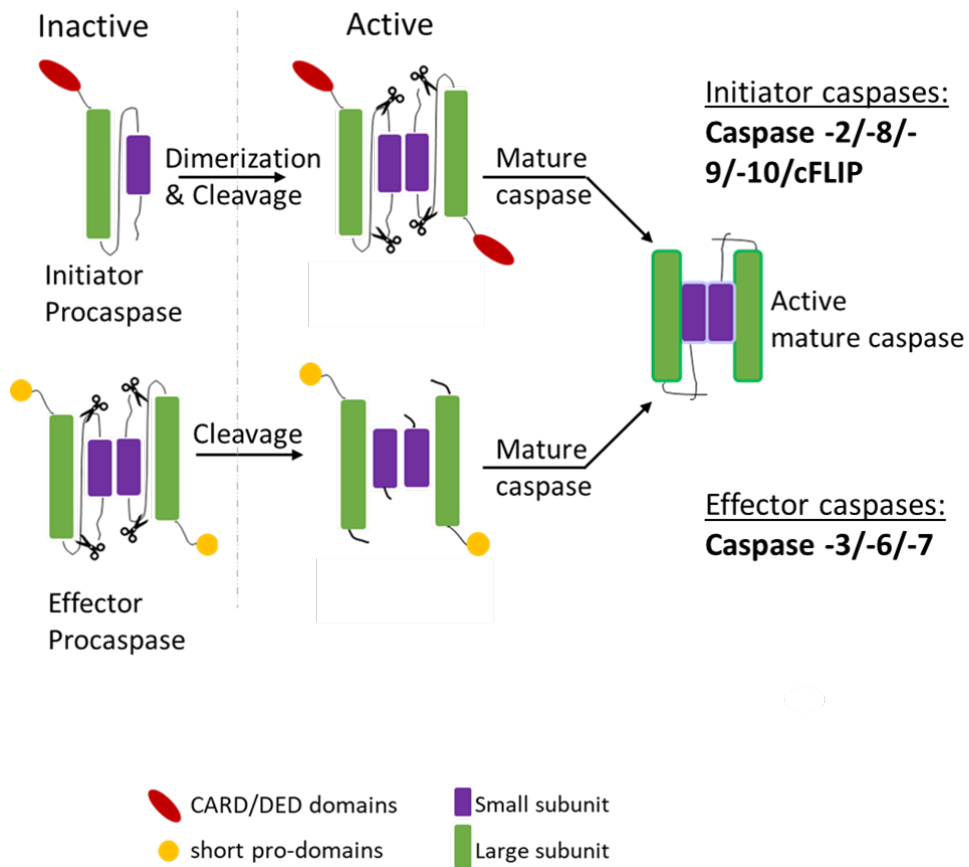
Figure 2. Activation mechanism of caspases involved in apoptosis. Initiator caspases are stable monomers and require dimerization for activation, whereas effector caspases are stable dimers that are activated by the cleavage of the inter-subunit linker.

CHAPTER 2

Sequential unfolding mechanisms of monomeric caspases

Isha Joglekar,[1,] and A. Clay Clark[1,*]

[1]Department of Biology, University of Texas at Arlington, Arlington, Texas, 76019

Running title: pH-effects on the stability of monomeric caspases

*Corresponding author: A. Clay Clark

Email: clay.clark@uta.edu

Abbreviations: DED, death-effector domain; CARD, caspase activation and recruitment domain; DISC, death inducing signaling complex, CH, caspase-hemoglobinase; PCP, procaspase; CP, common position; SASA, solvent accessible surface area; AEW, average emission wavelength; CD, circular dichroism; MD, molecular dynamics; NMR, nuclear magnetic resonance; RMSF, root mean square fluctuation; DTT, dithiothreitol, IPTG, isopropyl ß-D-1-thiogalactopyranoside.

Abstract

Caspases are evolutionarily conserved cysteinyl proteases that are integral in cell development and apoptosis. All apoptotic caspases evolved from a common ancestor into two distinct subfamilies with either monomeric (initiators) or dimeric (effectors) oligomeric states. The regulation of apoptosis is influenced by the activation mechanism of the two subfamilies, but the evolution of the well-conserved caspase-hemoglobinase fold into the two subfamilies is not well understood. We examined the folding landscape of monomeric caspases from two coral species over a broad pH range of 3 to 10.5. On an evolutionary timescale, the two coral caspases diverged from each other approximately 300 million years ago, and they diverged from human caspases about 600 million years ago. Our results indicate that both proteins have overall high stability, ~ 15 kcal mol^{-1} near the physiological pH range (pH 6 to pH 8), and unfold via two partially folded intermediates, I_1 and I_{2*}, that are in equilibrium with the native and the unfolded state. Like the dimeric caspases, the monomeric coral caspases undergo a pH-dependent conformational change resulting from the titration of an evolutionarily conserved site. Data from molecular dynamics simulations paired with limited proteolysis and MALDI-TOF mass spectrometry show that the small subunit of the monomeric caspases is unstable and unfolds prior to the large subunit. Overall, the data suggest that all caspases share a conserved folding landscape, that a conserved allosteric site can be fine-tuned for species-specific regulation, and that the subfamily of stable dimers may have evolved to stabilize the small subunit.

Key Words: caspase; dimerization; apoptosis; protein folding; protein evolution; fluorescence spectroscopy; circular dichroism; molecular dynamics simulations

Introduction

Caspases are a family of cysteine proteases and are well known for their role in apoptosis and inflammation[1]. At lower activation levels, however, caspases play essential roles in non-apoptotic functions like cell proliferation and differentiation, tissue regeneration, and neuronal development.[1,2] All caspases exist as inactive zymogens in the cell, and the caspase protomer comprises an N-terminal pro-domain and a large and a small subunit that are connected by an inter-subunit linker [3] (Fig. 1A). The apoptotic caspases evolved into two subfamilies and are classified as either initiator or effector caspases, depending on their entry into the cell death cascade. Initiator caspases exist as monomers and require dimerization for complete activation, whereas the effector caspases are dimers that require processing for complete activation. [4–6] Within the protomer of monomeric caspases, the pro-domain consists of an N-terminal CARD (caspase activation and recruitment domain) or two DEDs (death effector domain), which link caspase activation to death activation platforms in the cell, such as the death inducing signaling complex (DISC). In contrast, the dimeric effector caspases consist of short pro-domains and are activated via cleavage of the intersubunit linker by the initiator caspases. [1,7–9]

The caspase-hemoglobinase (CH) fold, which is a six-stranded ß-sheet core surrounded with atleast five α-helices on the protein surface, is a conserved feature among all caspases, specifically referring to their large and small subunits, excluding the prodomain (Fig.1B,C).[10]   [10]The protomer of effector caspases is cleaved at the intersubunit linker to separate the large and small subunits, resulting in the rearrangement of several active site loops and stabilization of the active conformation.

In contrast, initiator caspases are activated through dimerization, and subsequent chain cleavage is thought to stabilize the dimer. [11,12] Thus, the CH fold provides a basis for the evolution of regulation through oligomerization; enzyme specificity, resulting in overlapping but non-identical cellular substrates; and allosteric regulation, through both common and unique allosteric sites. As a result, caspases provide an excellent model for studying protein evolution. [4,13]

Previous studies in the model organisms *C. elegans* and *D. melanogaster* have provided critical insights into the molecular basis of the caspase cascade; however, both organisms utilize fewer caspases than in humans, resulting in apoptosis cascades that do not reflect the characteristics of the ancestral metazoans.[14–17] In contrast, genomic studies of cnidaria, the sister group to bilaterians comprising corals, jellyfish, and sea anemone, revealed multiple genes that were previously thought to be confined to vertebrates, thereby demonstrating the extensive gene loss in *C. elegans* and *D. melanogaster*. Indeed, the apoptotic genes in the cnidarians appear to complement the apoptotic signaling cascade observed in vertebrates, although much less is known about the cnidarian caspases compared to their counterparts in vertebrates.[18–22] We showed recently that a caspase-7 from *Porites astreoides* (PaCasp7a), a disease-resistant reef-building coral, and a caspase-3 from *Orbicella faveolata* (OfCasp3a), a disease-sensitive reef-building coral, utilize the conserved CH fold (Fig. 1C). The presence of an N-terminal CARD on a caspase that exhibits caspase-3/-6-like enzyme specificity suggests linkage of effector caspases to death platforms in coral. [23] The collective data from vertebrates and invertebrates show that caspases utilize a protein fold that has been conserved for at least one billion years and that evolutionary

modifications to the conformational landscape resulted in stable monomeric or dimeric subfamilies, changes in enzyme specificity, and changes in allosteric regulation. [4,13,24–26]

The folding landscape of human caspase-3 has been studied extensively.[27,28] Recent studies of other effector caspases from human (caspases-6 and -7) [4] and from zebrafish (caspases-3a and -3b) [13], as well as the common ancestor of effector caspases, [4,26] show a conserved folding landscape in which the native dimer (N_2) is in equilibrium with at least two partially folded intermediates - I_2 (partially folded dimer) and I (partially folded monomer) - prior to unfolding (U). The conserved landscape provides flexibility through mutations that either stabilize or destabilize the partially folded intermediates, allowing species-specific adjustments to the relative populations of partially folded states. In addition, all dimeric caspases examined to date undergo a pH-dependent conformational change, with pKa~6, resulting in an enzymatically inactive conformation. [4,13,29] The reversible conformational change may provide a mechanism to regulate caspase activity through localized changes in pH.

In contrast to the dimeric caspases, very little is known about the folding landscape of monomeric caspases, and the properties have been inferred from the monomeric folding intermediate (I) of dimeric caspases. [27] To further understand the evolution of the monomeric subfamily of caspases, we examined the folding landscape of coral caspases from *O. faveolata* and *P. astreoides*, which are about 300 million years apart on an evolutionary timescale and about 600 million years apart from human caspases. The two proteins (PaCasp7a and OfCasp3a) are monomeric (Fig. 1D) [23], are orthologs of the initiator family of caspases, and share a common ancestor with effector caspases (caspases-3, -6, -7) and with initiator caspases (caspases-2, -9, -8, -10) (Fig. 1E).

PaCasp7a and OfCasp3a have about 77% amino acid sequence identity with each other and about 35% sequence identity with human caspases -3/-7. [23,26]

Here, we examined the urea-induced equilibrium unfolding of the zymogens (procaspases) of PaCasp7a and of OfCasp3a over a broad pH range, from pH 3 to pH 10.5. The results are compared to the partially folded monomeric intermediate observed during procaspase-3 (PCP-3) folding and assembly. The data show that the pH-dependent conformational change observed in the dimers is also conserved in the monomers, suggesting an evolutionarily conserved regulatory mechanism. In addition, the monomeric caspases unfold via similar mechanisms where at least one partially folded intermediate is in equilibrium with the native conformation and the unfolded state. Overall, the conformational free energy is somewhat higher than that determined for the partially folded monomeric intermediate of dimeric caspases, and the native conformation is destabilized compared to the partially folded intermediates at low and high pH. Finally, the data show that the small subunit in the protomer unfolds prior to the large subunit. Together, the data indicate that evolution of the dimeric family of caspases was important for stabilizing the small subunit within the protomer.

Materials and Methods

Cloning, protein expression, and purification

Gibson cloning was used to clone PaCasp7a and OfCasp3a without their N-terminal CARD domain, and the active site cysteine was mutated to serine using site-directed mutagenesis, as described previously. [35] The inactive PaCasp7a and OfCasp3a zymogens were cloned into a pET-11a expression vector with a C-terminal HisTag, and

all proteins were expressed in *E. coli* BL21 (DE3) pLysS cells and purified as previously described. [27]

Phylogenetic analysis

Caspase sequences of representative species listed in Table S3 were retrieved from CaspBase,[30] coupled with top hits from BLAST and HMMER,[36] and multiple sequence alignments were generated using PROMALS3D[37] The optimal model of evolution for constructing a phylogenetic tree was determined using IQ-TREE, and the tree was constructed using the maximum likelihood approach using the Jones-Taylor Thornton model and distribution.[38] As a test for phylogeny, the tree was bootstrapped 1000 times.

Stock solutions and sample preparation for equilibrium unfolding

Equilibrium unfolding experiments were carried out as previously described. [32] Briefly, urea stock solutions (10 M) were prepared in citrate buffer (50 mM sodium citrate/citric acid, pH 3 to pH 5.5, 1 mM DTT), potassium phosphate buffer (50 mM potassium phosphate monobasic/potassium phosphate dibasic, pH 6.0-8.0, 1 mM DTT), and glycine-NaOH buffer (50 mM glycine/NaOH, pH 9 to pH 10.5, 1 mM DTT). For unfolding reactions, samples were prepared in the corresponding buffer with final urea concentrations between 0 M and 9 M. For refolding reactions, the protein was first incubated in a 9 M urea-containing buffer for ~6 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 M and 9 M. All solutions were prepared fresh for each experiment and were filtered (0.22 μm pore size) before use. Final protein

concentrations of 0.5 mM - 4 mM were used. The samples were incubated at 25°C for a minimum of 16 hours to allow for equilibration.

Fluorescence emission and CD measurements

Fluorescence emission was measured using a PTI C-61 spectrofluorometer (Photon Technology International) from 310 to 410 nm following excitation at 280 or 295 nm. Excitation at 280 nm follows tyrosinyl and tryptophanyl fluorescence emission, whereas excitation at 295 nm follows the tryptophanyl fluorescence emission. CD data were measured using a J-1500 CD spectropolarimeter (Jasco) between 210 and 260 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 to the equilibrium unfolding data

To visualize the different spectroscopic signals on a single scale, the raw data were corrected for buffer background and normalized between zero (unfolded) and one (native), as previously described.[32] The corresponding relative signal for intermediate species observed vary with pH .All data were fit globally as described previously.[27,28,32] Briefly, equilibrium unfolding data were collected for fluorescence emission (two excitations) and far-UV CD, and at several protein concentrations, between pH 3 and 10.5 for both PaCasp7a and OfCasp3a, providing six data sets at each pH. The data were fit globally to a two-state, three-state, or a four-state equilibrium folding model, as described. [32] Data collected between pH 6 and pH 8 were fit to a four-state equilibrium folding model shown in equation 1. In this model, the native protein unfolds via the presence of two monomeric intermediate species, I_1 and I_{2*}, prior to forming the the

unfolded state. The equilibrium constants K_1, K_2, and K_3 relate to equilibrium constants at respective unfolding steps.

$$N \overset{K_1}{\leftrightarrow} I_1 \overset{K_2}{\leftrightarrow} I_{2*} \overset{K_3}{\leftrightarrow} U \qquad \text{eqn (1)}$$
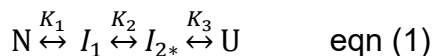
Data collected between pH 4.5 and 6, as well as between pH 9 and pH 10.5, were fit to a 3-state equilibrium folding model as described previously [27,28,32] and shown in Equation 2. In this model, the intermediate state (I_1) unfolds to the intermediate (I_{2*}) state before complete unfolding.

$$I_1 \overset{K_1}{\leftrightarrow} I_{2*} \overset{K_2}{\leftrightarrow} U \qquad \text{eqn (2)}$$

Data collected for PaCasp7a between pH 3 and pH 4 were fit to a two-state equilibrium folding model as described [32] and shown in equation 3,

$$I_{2*} \overset{K_1}{\leftrightarrow} U \qquad \text{eqn (3)}$$

Limited trypsin proteolysis

Proteins (6 mM) were digested with 0.5 ng/mL of trypsin in a buffer of 50 mM potassium phosphate, pH 7, 1 mM DTT at 25 °C. On addition of trypsin, aliquots were withdrawn at prescribed time intervals, and reactions were inhibited by adding SDS-PAGE buffer and heating to 100 °C for 10 minutes. The samples were frozen at -20°C until analyzed. Samples were analyzed with 4-20% SDS-PAGE gradient gels.

MALDI-TOF mass spectrometry

Protein bands generated by limited trypsin proteolysis and resolved on SDS-PAGE gels were excised and destained with a solution of acetonitrile and 50 mM ammonium

bicarbonate (1:1 v/v) for about 4 hours. The gel fragments were crushed in

microcentrifuge tubes, and the proteins were extracted with a solution of formic

acid/water/2-propanol (1:2:3 v/v/v) for 8 hours at room temperature. After extraction,

samples were centrifuged, and the supernatant was lyophilized and redissolved in

matrix solution (formic acid/water/2-propanol saturated with sinapinic acid). Protein was

retrieved for MS analysis using the dried-drop method of matrix crystallization [39] and

then analyzed using MALDI-MS (Axima Assurance Linear MALDI TOF). Laser power

was optimized and 5 shots were collected, with beam blanking turned on to blank high

intensity matrix peaks. Pulsed extraction was set to 2/3 maximum molecular weight.

Protein standards of known masses were used to calibrate time of flight. Data were

processed using data smoothing, baseline subtraction and threshold detection. Mass

analysis of protein digests and cleavage sites was performed using the protein

prospector software. [40]

MD simulations

PaCasp7a and OfCasp3a monomers were modelled using the Swiss-modeler program

that uses homology modeling algorithm with user-defined templates. [41] PaCasp7a and

OfCasp3a sequences were threaded onto the NMR structure of procaspase-8 (PDB

ID:2k7z), and the active site cysteine was mutated to serine using Pymol in both

proteins. The modeled proteins were protonated according to the calculated ionization

states of titratable groups at the specified pH using the H++ server [42]. The force field

parameters for urea were obtained as described previously, [43] and the urea molecule

was built using the Avogadro software. [44] A cubic box of 6 × 6 × 6 nm^3 containing 560

molecules of urea was generated to achieve a 5M concentration as described

previously.[45]The system was subjected to energy minimization with the steepest-descent algorithm down to a maximum gradient of 2000 kJ mol^{-1} nm^{-1} and was simulated for 1 ns with annealing from 300 to 0K under an isotropic pressure of 100 bar. The system was then relaxed for 1 ns at standard pressure, heated from 0 to 300K, then simulated for 1 ns at 300K. The size of the box at the completion of this equilibration technique was 3x3x3 nm^3. Using the Nosé-Hoover coupling algorithm, [46,47] we performed 100 ps MD simulations using the NVT (constant volume and temperature) and NPT (constant pressure and temperature) ensemble at 300K starting from the relaxed box. After heating the simulated system to 300K, a production run for each protein was conducted for 100 ps using the NPT ensemble. MD simulations were performed for 200 ns with GROMACS using the Amber99 force field and the TIP3P water model as described. [48] In addition, the MD trajectories were analyzed using GROMACS inbuilt tools such as gmx rmsf for computing the root mean square fluctuation of atomic positions in the trajectories and gmx sasa for computing the solvent accessible surface area.

Results

The monomeric procaspase PaCasp7a and OfCasp3a constructs are without an N-terminal CARD domain and consist of 260 and 258 amino acids, respectively, with a molecular weight of about ~31 kDa (Fig 1B). Since caspase activation can be autocatalytic under certain circumstances, such as when protein concentrations are high in heterologous expression systems, we replaced the active site cysteine with a serine residue (CP-C120S) for our equilibrium unfolding studies to prevent cleavage during expression in *E. coli*. We utilize the common position (CP) nomenclature,

described previously, that allows amino acid position descriptions and comparisons of all caspases from any organism. [30] We note that, the CP-C120S mutation prevents autocatalysis, but is not structurally perturbing. [27,31] As shown in the sequence alignment in Figure 1B, the proteins begin with a uniformly conserved tyrosine (CP-001) at the N-terminus of the protease domain (that is, large subunit-intersubunit linker-small subunit). PaCasp7a and OfCasp3a have two tryptophan residues each (CP-W160 and CP-W168), and both reside in the active site loop 3 (L3), which is in the small subunit of these caspases (Fig. 1A,D). In addition, PaCasp7a and OfCasp3a have 14 and 15 tyrosine residues, respectively, well distributed in the primary sequence (Fig. 1B, blue). Finally, residues in b-strand 6, which forms the dimer interface (Fig. 1B, CP-208 to CP-213, SSV/ISML), are more similar to those of the dimeric effector caspases (CIVSML) than those of the monomeric initiator caspases (QPTFTL), suggesting that other factors also contribute to stable dimer formation.

Urea-induced unfolding of PaCasp7a and of OfCasp3a

We examined changes in the tertiary structure at increasing urea concentrations by exciting the samples at 280 nm or at 295 nm to monitor changes in fluorescence emission of tyrosine and tryptophan residues (280 nm) or tryptophan residues only (295 nm) as a function of urea concentration. As shown in Supplemental Figure S1, native PaCasp7a (Supplemental Fig. S1, panels A and B) and OfCasp3a (Supplemental Fig. S1, panels D and E) have an emission maximum at 334 nm when excited at 280 nm, similar to that of human PCP-3 (procaspase-3) and 338 nm when excited at 295 nm, similar to that of PCP-7 (procaspase-7). The shift in the emission maxima when excited at 280 nm versus 295 nm demonstrates that tryptophan residues in the native state are

more solvent-exposed than the other aromatic residues. We note that, we have previously observed a similar shift in the emission maxima in PCP-3 and an ancestral caspase (PCP-CA) of the effector family.[4,27] The fluorescence emission maxima are red-shifted to ~347 nm in phosphate buffer containing 9 M urea, showing that the aromatic residues are exposed to solvent and that the proteins are fully unfolded under these solution conditions. Similarly, PaCasp7a (Supplemental Fig. S1, panel C) and OfCasp3a (Supplemental Fig. S1, panel F) have well-formed secondary structure that is disrupted in a 9 M urea-containing phosphate buffer, as shown by the significant change in the signal due to the loss of secondary structure, determined with far-UV circular dichroism (CD).

We examined the equilibrium unfolding of monomeric coral caspases as a function of urea concentration over the pH range of 3 to 10.5.. As expected for monomers, the findings of our studies at 0.5uM-4uM protein concentration suggest that the data were independent of concentration for both proteins and can be represented by the average as described earlier by Walters et al. (2009).[32] Further, refolding experiments for both proteins demonstrated reversible folding transitions. Representative data for the equilibrium unfolding of PaCasp7a and OfCasp3a at pH 3, pH 7, and pH 10 are shown in Figure 2. The data for pH 7 are described here, while those for lower and higher pHs are described below. In the case of PaCasp7a at pH 7 (Fig. 2A), one observes a pre-transition between 0 M and ~1.5 M urea in both the fluorescence emission and CD data, showing little to no change in the signal of the native protein. The pre-transition is followed by a cooperative change in the signal between ~1.5 M and ~2.5 M urea, where the signal for fluorescence emission increases. Interestingly, the transition is

accompanied by a near complete loss in secondary structure as shown by the decrease in CD signal. Following the cooperative change in signal, there is an apparent plateau between ~2.5 M and ~4.5 M urea when samples are excited at 295 nm, although one observes a modest decrease in the fluorescence signal in this region when samples are excited at 280 nm. At higher urea concentrations, one observes a second cooperative transition between ~4.5 M and ~6 M urea, demonstrating that the protein is unfolded at urea concentration >6 M. In comparison, OfCasp3a demonstrates a pre-transition between 0 M and ~0.5 M urea in both the fluorescence emission and CD data, followed by a cooperative change in the signal between ~0.5 M and ~1.5 M urea. Like PaCasp7a, the fluorescence emission signal increases, whereas the relative signal for CD decreases during this transition. Between ~2 M and ~4.5 M urea, one observes a decrease in the fluorescence signal followed by a second cooperative transition between ~5 M and ~6 M urea, beyond which the protein is unfolded. We note that the relative fluorescence emission signal is higher when samples are excited at 295 nm compared to those excited at 280 nm and that the relative fluorescence signal is higher overall than that observed by CD for both proteins.

Global fits to the equilibrium unfolding data

As described previously for the dimeric caspases[4,13,27], the data shown in Figure 2 were fit globally to an equilibrium folding model to determine the conformational free energies. The solid lines through the data points represent the global fits. The data at pH 7 (Fig. 2 A,B) suggest that the proteins unfold via a four-state mechanism in which the native monomer (N) isomerizes to a partially folded intermediate (I_1), followed by further unfolding through a second partially folded intermediate (I_{2*}) prior to unfolding (U)

(N $\leftrightarrow$ I$_1$ $\leftrightarrow$ I$_{2*}$ $\leftrightarrow$ U) (equation 1). We note that our equilibrium unfolding studies of

monomeric caspases revealed a second partially unfolded intermediate that differs from

the dimeric intermediate (I$_2$) observed in the unfolding of effector caspases. To

distinguish it from the latter, we are denoting it with an asterisk (I$_{2*}$). In general, the two

partially folded intermediate conformations, I$_1$ and I$_{2*}$, are observed in the data through

the plateau between ~2 M and ~4.5 M urea. While both intermediates exhibit a higher

relative fluorescence emission than the native conformation, the second intermediate,

I$_{2*}$, exhibits a lower fluorescence emission than does the first intermediate, I$_1$. At pH 7,

for PaCasp7a, the transition from native (N) to intermediate 1 (I$_1$) occurs with a free

energy of $\Delta G_1^{H_2O}$ = 2.2 ± 0.3 kcal mol^{-1}, and the transition from intermediate$_1$ (I$_1$) to

intermediate$_2$ (I$_{2*}$) has a free energy of $\Delta G_2^{H_2O}$ = 1.5 ± 0.4 kcal mol^{-1}. The free energy

change, $\Delta G_3^{H_2O}$, for the unfolding of the second intermediate to the unfolded state, I$_{2*}$ to

U, is 11.2 ± 0.1 kcal mol^{-1}, with a total conformational free energy of $\Delta G°_{conf}$ =14.9 kcal

mol^{-1}. The cooperativity indices m$_1$, m$_2$, and m$_3$ for the transitions are 1.05 ± 0.3 kcal

mol^{-1}M^{-1}, 1.10 ± 0.4 kcal mol^{-1}M^{-1}, and 2.30 ± 0.4 kcal mol^{-1}M^{-1}, respectively (Table 1).

For OfCasp3a at pH 7, the free energy change $\Delta G_1^{H_2O}$ for N to I$_1$ is 2.0 ± 0.4 kcal mol^{-1},

and that of I$_1$ to I$_{2*}$, $\Delta G_2^{H_2O}$ = 1.9 ± 0.5 kcal mol^{-1}. The free energy change $\Delta G_3^{H_2O}$, for I$_{2*}$ to

U, is 13.4 ± 0.6 kcal mol^{-1} with a total conformational free energy of $\Delta G°_{conf}$ = 17.4 kcal

mol^{-1}. The cooperativity indices m$_1$, m$_2$, and m$_3$ for the transitions are 1.80 ± 0.3 kcal

mol^{-1}M^{-1}, 1.20 ± 0.5 kcal mol^{-1}M^{-1}, and 2.60± 0.8 kcal mol^{-1}M^{-1}, respectively (Table 2).

Thus, the two proteins exhibit similar conformational free energies for unfolding.

Interestingly, the m-values indicate that approximately half of the buried surface area is

exposed in the first two unfolding transitions. Taken together, the data suggest that the

second intermediate, I_{2^*}, is characterized by a lack of secondary structure but with substantial buried hydrophobic surface. We previously showed that the partially folded monomer of human PCP-3 has a conformational free energy change of 7.2 ± 0.5 kcal mol^{-1} at pH 7, 25°C, so both monomeric coral caspases exhibit substantially higher conformational free energy.

pH effects on equilibrium unfolding of PaCasp7a and OfCasp3a

We showed previously that pH changes are an excellent perturbant for examining the conformational landscape of caspases as both the conformational stability and oligomeric state may be affected. [28] As shown in Figure 2 and Supplemental Figures S2-S5, we examined equilibrium folding of PaCasp7a and of OfCasp3a over the pH range of 3 to 10.5. While the data for the two extremes (pH 3 and pH 10.5) are summarized in Figure 2, the full profile is shown in Supplemental Figures S2-S5. We note that the equilibrium folding of OfCasp3a was not reversible between pH 4.5 and 6 due to protein aggregation, so data for those pHs are not considered in subsequent analyses. Similar to the folding model described at pH 7 above, the equilibrium unfolding of PaCasp7a and of OfCasp3a is well-described by a four-state model (equation 1) between pH 6 and 8. Fits of the data to the models described below are shown as the solid lines in the figures, and the $\Delta G°_{conf}$ and m-values are shown in Tables 1 and 2, and in Supplementary Tables S1 and S2. Finally, the pH dependence of the conformational stability and cooperativity indices (m-value) over the entire pH range (pH 3 to pH 10.5) for both PaCasp7a and OfCasp3a are summarized in Figure 3.

First, one observes that the midpoint of the first transition shifts to lower urea concentrations as the pH is lowered, such that the native protein is destabilized below

pH 6 (Supplemental Figures S2 and S4) resulting in the loss of the N-to-I_1 transition below pH~4.5. In other words, the "native" protein at pH 4.5 appears to comprise the I_1 partially folded state, so the data at pH 4.5 were best-fit to a three-state folding model ($I_1 \leftrightarrow I_{2^*} \leftrightarrow U$, equation 2). As the pH is lowered further, the mid-point of the transition of I_1 to I_{2^*} decreases such that the transition of I_1 to I_{2^*} is lost, and the data are best-fit to a two-state equilibrium folding model at pH 3 ($I_{2^*} \leftrightarrow U$, equation 3) (Fig. 2C,D and Supplemental Figures S2 and S4). A similar process is observed at higher pH. As the pH is increased above pH 9, the midpoint of the N to I_1 transition shifts to lower urea concentrations such that N is destabilized at pH>9, and the data are best fit first to a three-state equilibrium unfolding model ($I_1 \leftrightarrow I_{2^*} \leftrightarrow U$, equation 2, pH 9.5-10), then to a two-state equilibrium folding model at pH 10.5 ($I_{2^*} \leftrightarrow U$, equation 3) (Figure 2D,E and Supplemental Figures S2 and S4). For both PaCasp7a and OfCasp3a, the total conformational free energies demonstrate that the proteins are most stable near physiological pH and that the stability decreases at both lower and higher pH due to destabilizing first the native conformation then the first folding intermediate, I_1.

Overall, the data show that the higher stability near physiological pH, between pH 6 and ~8, with a $\Delta G°_{conf}$~16 kcal mol^{-1}, is largely due to the contribution of the partially folded intermediate, I_{2^*}, with $\Delta G_3^{H_2O}$~12 kcal mol^{-1}. While the stability of the native conformation and of the partially folded intermediate, I_1, decrease at lower and higher pH, the stability of I_{2^*} appears to be largely independent of pH and decreases only below pH 4 (Fig. 3A,C). The cooperativity index (m-value) is correlated to the change in accessible surface area (ΔASA) exposed to solvent during each unfolding transition. The total m-value (m_{total}) is similar for PaCasp7a and OfCasp3a, indicating a similar exposure of

hydrophobic area during each unfolding transition (Fig. 3B,D). Together, the data suggest that partially folded conformations are well-populated at lower and higher pHs and that the two monomeric caspases exist in their native conformation only between the pH range of 6 to 8.

The changes in the total conformational free energy shown in Figure 3 for both proteins are compared in Figure 4A and show three transitions over the pH range of 3 to 10.5, mostly due to the changes in the populations of N and I_1 at lower and higher pH as well as the lower stability of I_{2*} below pH 4. In order to further examine the conformational changes due to pH, we performed titrations of native protein versus pH, in the absence of urea, and measured changes in secondary structure at 232nm for PaCasp7a and 230nm for OfCasp3a(Fig. 4B). We also measured changes in average emission wavelength for both proteins following excitation at 280 nm or at 295 nm (Fig. 4C,D). As described above for urea-induced equilibrium unfolding, the proteins exhibit a substantial loss of secondary structure during the transition of N to I_1. Likewise, one observes a substantial loss of secondary structure when the native protein is titrated from pH 7 to pH 6 (Fig. 4B). While the secondary structure decreases at higher pH, the signal loss is lower than that observed at lower pH. In terms of the tertiary structure, the average emission wavelength is relatively constant at pH>6, and one observes a cooperative decrease below pH 6. We fit the data in Figure 4 to determine the pKa values for each transition, as described in Methods, and the fits are shown as the dashed lines in the figure. Although the data for OfCasp3a were not well-determined due to the lack of data at all pH values, we note that both proteins exhibit similar transitions and have identical pKa values. The fits to the total free energy, $\Delta G°_{conf}$ (Fig

4A) can be characterized by three pKa values between pH 3 and pH 10.5 for PaCasp7a and for OfCasp3a. We observe that, on decreasing the pH from 7 to 5, there is a decrease in the overall stability characterized by an estimated pKa of ~5.7. The transition correlates with the transition observed in the average emission wavelength of the native PaCasp7a and OfCasp3a (Fig 4C,D), which can be characterized by an estimated pKa ~5.3. Changes in the secondary structure for PaCasp7a and OfCasp3a show a sharp transition (Fig. 4B), with an estimated pKa of 6.1. We previously demonstrated that the dimers of PCP-3, PCP-6, and DrPCP-3b undergo a pH-dependent conformational change, with a pKa of 5.7. [4,13,29] Together, the data suggest that the decrease in the average emission wavelength and the conformational stability report the same pH-dependent conformational change in the monomeric coral caspases. That is, titration of one or more residues affects the transition of the native conformation to a partially folded intermediate. In the dimeric caspases, the transition is accompanied by a loss of enzymatic activity, although the protein remains dimeric. In the monomeric caspases, shown here, the transition is accompanied by a loss in secondary structure and a blue-shift in fluorescence emission. In addition, the proteins undergo a further transition with pKa~3.3. The data show that the transition is due to titration of the partially folded intermediate, I_{2*}. Interestingly, the transition that occurs at higher pH, with pKa~9.3, is not accompanied by a blue-shift in fluorescence emission even though the secondary structure is destabilized. The simplest model suggests that the native conformation is destabilized relative to the partially folded intermediates at both lower and higher pH. If this is true, then the fluorescence emission at higher pH

may not reflect the blue-shift until all secondary structure is lost, indicating a broader pH range would be required at higher pHs.

Molecular dynamics (MD) simulations and limited proteolysis show the small subunit is unstable

The equilibrium unfolding data described above demonstrate that the proteins have substantial conformational free energy between pH 5 and 10, with $\Delta G°_{conf}$ ~12-17 kcal mol^{-1} (Fig. 4A). Yet, based on data from urea-induced unfolding (Fig. 2) and from far-UV CD spectra (Fig. 4B), the proteins lose most of their secondary structure below pH 7 and above pH ~10. In addition, the proteins exhibit a cooperative unfolding transition following the apparent loss of secondary structure (Fig. 2), with a corresponding cooperativity index (m-value) that suggests a substantial hydrophobic core that remains solvent inaccessible. In contrast to the CD data, the average emission wavelength (AEW) does not change with pH>5, where the AEW remains constant at ~347 nm, which suggests that the tryptophan residues are in a hydrophilic environment (Fig. 4C,D). To further examine how the proteins may lose secondary structure at lower and higher pHs while retaining a partially folded conformation, we performed MD simulations at several pHs and in the presence and absence of 5 M urea.

PaCasp7a and OfCasp3a protomers were modeled using the procaspase-8 structure determined by NMR (PDB ID: 2k7z),[33] since no other structures are available for monomeric procaspases, and the active site cysteine was mutated to serine in both proteins (CP-C120S).The inter-subunit linker in PaCasp7a and OfCasp3a is intact as is observed in the model (Fig. 1B, Fig. S6). All ß-strands in the hydrophobic core of the

protein are well-formed except for ß6. Indeed, ß6 forms multiple contacts in the interface of the dimeric effector caspases but is less well-packed in the monomers.

MD simulations were carried out for 200 ns, which is sufficient time to observe changes in the protein conformation in the presence of urea. Data for PaCasp7a and for OfCasp3a in the absence of urea offer a baseline against which we compared data for both proteins in the presence of 5 M urea. Since the equilibrium unfolding studies suggest that both proteins are largely unfolded at 5M urea, we used 5M urea concentration in MD simulations to identify the unstable regions on the proteins. The unfolding of proteins during urea denaturation eventually exposes the hydrophobic core to the aqueous environment, resulting in the loss of hydrophobic contacts. The solvent-accessible surface area (SASA) can be used to determine changes in the solvent accessibility throughout the simulation [34]. Overall, the native PaCasp7a and OfCasp3a at pH 4, 7, and 9 in the absence of urea show a similar SASA of ~140 nm^2. However, in the presence of 5 M urea, the SASA of PaCasp7a and of OfCasp3a is much larger, ~195 nm^2, due to higher fluctuations as a result of unfolding (Fig. 5A,C). To examine residue-level changes, we computed root mean square fluctuations (RMSF) for both proteins in water vs 5 M urea. The RMSF data for proteins in water were subtracted from those in 5 M urea to determine regions of the protein with increased fluctuations in urea. In addition, the experiments were performed at pH 4, pH 7, and pH 9. The data show that, in the absence of urea, PaCasp7a and OfCasp3a show minimal fluctuations overall and that the active site loops as well as the N- and C-termini show the largest fluctuations (Supplemental Fig. S6). However, in the presence of 5 M urea, one observes elevated fluctuations in the region of helices 2 and 3, the inter-subunit linker,

and the small subunit, particularly helices 4 and 5 (Fig. 5C,D and Supplemental Fig. S6 B,D). We examined snapshots of the protein structures over the course of the simulations in 5 M urea, and the data show that the increased fluctuations in the small subunit as well as in helices 2 and 3 correlated with unfolding (Fig. 5B,D and Supplemental Fig. S6). At the end of the simulation, the small subunit is largely unfolded and helices 2 and 3 are pulled away from the protein, exposing the core ß-strands (Fig.S7). Similar results to those at pH 7 were observed at both low and high pH, except that the unfolding of the small subunit occurred earlier in the simulation, which may correlate with the lower $\Delta G°_{conf}$ observed in the equilibrium unfolding studies described above. Altogether, the data show that both proteins behave similarly, wherein the small subunit unfolds first while helices 2 and 3 separate from the body of the large subunit.

To further examine conformational changes in PaCasp7a and in OfCasp3a due to changes in pH, we performed limited proteolysis studies at pH 7 and 9 using trypsin. The results show that trypsin cleaves the proteins in discrete regions to generate several products (Supplemental Fig S8A-D). One observes that PaCasp7a and OfCasp3a are cleaved initially to produce slightly smaller, although mostly intact proteins, suggesting cleavages near the termini (Band 2 in Supplemental Fig. S8A,C). The proteins are then further cleaved to generate fragments of 26 (Band 2), 20 (Band 3), 18 (Band 4), 15 (Band 5), and 9 kDa (Band 6). We analyzed the digests by MALDI-TOF mass spectrometry, and the results show that the first two cleavages occur at CP-R161 and GP9-R02 (see Fig. 1A), respectively, generating the 26 kDa and 20 kDa fragments. CP-R161 is located in active site loop 3 (Fig. 6). In the protomer, loop 3 is disordered, so CP-R161 is solvent exposed (Fig. 6A). In contrast, loop 3 forms the

substrate binding pocket in the active dimer, so CP-R161 is more buried (Fig. 6B). In addition, GP9-R02 is located between a-helices 4 and 5, again suggesting that the two helices in the small subunit are unstable. Further cleavages of the C-terminus, inter-subunit linker, and CP-R018 on Loop 1 (R64) generate the remaining fragments. It has been reported that CP-R161 in the dimeric caspase-3 is cleaved by trypsin with a $t_{1/2}$~75 minutes at pH 7.2 and 25 °C. [29] However, CP-R161 is cleaved in both PaCasp7a and OfCasp3a, with a $t_{1/2}$~ 10 minutes. When considered with the result that most of the cleavages occur in the small subunit, then data suggest that the small subunit is unstable in the protomer. We note that the same cleavages occur at pH 9 as observed at pH 7, but with faster kinetics, again supporting the conclusion that the small subunit is less stable at higher pH (Supplemental Fig. S8B,D). Together, the results from MD simulations and limited proteolysis show that the small subunit of the monomeric coral caspases is unstable and unfolds prior to the large subunit.

Discussion

In this study, we characterized the equilibrium unfolding of monomeric caspases from two species of coral that are ~300 million years distant on an evolutionary timescale. Equilibrium unfolding data suggest that PaCasp7a and OfCasp3a unfold via two partially folded intermediates in equilibrium with the native and the unfolded state, showing maximum stability of ~17 kcal mol^{-1} near the physiological pH range (pH 6 to pH 8). The native state is destabilized outside of this pH range, and at the extreme pH the intermediate state (I_1) is also destabilized. The conformational free energy of monomeric caspases (~15 kcal mol^{-1}) is higher than that of the partially unfolded

monomeric intermediates observed during the folding and assembly of dimers, (~7 kcal mol^{-1}).

In the absence of urea, CD data indicate a substantial loss of secondary structure below pH 7 and above pH 8, yet the proteins have a conformational free energy of ~12 kcal mol^{-1}. The results suggest that intermediate I_{2*} is characterized as having a strong core with buried hydrophobic contacts. The MD simulations in the presence of urea also reveal that the small subunit is unstable and unfolds prior to the large subunit, while helices 2 and 3 and expose the core b-strands. Data from limited trypsin proteolysis also indicate that the small subunit is unstable in the protomer. The lower stability of the small subunit in the protomer suggests the importance of forming the dimer since b-strand 6 forms several inter-protomer contacts in the dimer.

In general, the conformational stability decreases when the pH is reduced from 7 to 5, and the transition occurs with a pKa of ~5.9. Similar pH-dependent conformational changes in the dimeric family of caspases have been reported[4,13,29]. Together, the data suggest that an evolutionarily conserved site is titrated with a pKa~6, indicating that either a histidine or an acidic residue controls an important conformational switch. For example, in the dimer, the switch results in the transition of the native dimer (N_2) to an enzymatically inactive conformation, I_2. [4,13,27,29] In the monomers, the switch also destabilizes the native conformation relative to an intermediate, which may affect the formation of the dimer. Nevertheless, the conformational switch observed in all caspases could be used as a regulatory mechanism to control the activity of caspases in cells, either through changes in the active site (dimeric caspases) or through destabilizing the protomer in monomeric caspases.

The folding landscape of the dimeric caspase subfamily is conserved and provides flexibility through changes in the population distributions of two folding intermediates, a monomeric (I) and a dimeric (I_2) intermediate. [4,13,27,28] Changes in the population of the two intermediates allow for species-specific modulations in overall protein stability. The caspase subfamilies evolved from a single ancestral scaffold with a conserved caspase-hemoglobinase fold that has been retained for over 650 million years, making them an excellent model for studying protein evolution. Previously, the folding mechanism of the resurrected common ancestor (CA) of the dimeric subfamily has been compared with the extant human caspases-3/-6/-7, and it was shown that the conformational free energy of the dimeric family ranges from 15 to 34 kcal/mol, [4] but there was little information regarding the monomer in the absence of dimerization.

PaCasp7a and OfCasp3a are evolutionarily distant from human caspases (600 million years), yet they provide a foundation for understanding the folding landscape of the monomeric family of caspases and the evolutionary events that led to the stable dimer. When considering the results described here with those of the dimeric caspases, we suggest a combined model where dimer formation results in the stabilization of the small subunit of the protomer. Evolutionary changes in the protomer that stabilize the small subunit may facilitate dimerization by decreasing fluctuations in regions of the small subunit that make up the bulk of contacts in the dimer.

Data availability

All data are contained in the article and supporting information.

Author contributions

I.J. and A.C.C conceptualization; I.J. and A.C.C methodology; I.J. investigation; I.J. and A.C.C writing-review and editing; A.C.C supervision.

References

1. Clark AC (2016) Caspase allostery and conformational selection. Chem Rev 11:6666–6706.

2. Shalini S, Dorstyn L, Dawar S, Kumar S (2015) Old, new and emerging functions of caspases. Cell Death Differ 22:526–539.

3. Chelur DS, Chalfie M (2007) Targeted cell killing by reconstituted caspases. Proc Natl Acad Sci 104:2283–2288.

4. Shrestha S, Clark AC (2021) Evolution of the folding landscape of effector caspases. J Biol Chem 297,101249:1–12.

5. Kesavardhana S, Malireddi RKS, Kanneganti TD (2020) Caspases in cell death, inflammation, and pyroptosis. Immunol 38:567–595.

6. McIlwain DR, Berger T, Mak TW (2013) Caspase functions in cell death and disease. Cold Spring Harb Perspect Biol 5.

7. Bao Q, Shi Y (2007) Apoptosome: a platform for the activation of initiator caspases. Cell Death Differ 14:56–65.

8. Riley JS, Malik A, Holohan C, Longley DB (2015) DED or alive: assembly and regulation of the death effector domain complexes. Cell Death Dis 6.

9. Mackenzie SH, Clark AC (2012) Death by caspase dimerization. Adv Exp Med Biol. 747:55–73.

10. Aravind L, Koonin EV (2002) Classification of the caspase-hemoglobinase fold: detection of new families and implications for the origin of the eukaryotic separins. Proteins 46:355–367.

11. Mackenzie SH, Clark AC (2008) Targeting cell death in tumors by activating caspases. Curr Cancer Drug Targets 8:98–109.

12. Parrish AB, Freel CD, Kornbluth S (2013) Cellular mechanisms controlling caspase activation and function. Cold Spring Harb Perspect Biol 5.

13. Yao L, Clark AC (2022) Comparing the folding landscapes of evolutionarily divergent procaspase-3. Biosci Rep 42.

14. Yuan J, Horvitz HR (1990) The Caenorhabditis elegans genes ced-3 and ced-4 act cell autonomously to cause programmed cell death. Dev Biol. 138:33–41.

15. Ellis HM, Horvitz HR (1986) Genetic control of programmed cell death in the nematode C. elegans. Cell 44:817–829.

16. Song Z, McCall K, Steller H (1997) DCP-1, a Drosophila cell death protease essential for development. Science (1979) 275:536–540.

17. Dorstyn L, Mills K, Lazebnik Y, 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.

18. Tchernov D, Kvitt H, Haramaty L, Bibby TS, Gorbunov MY, Rosenfeld H, Falkowski PG (2011) Apoptosis and the selective survival of host animals following thermal bleaching in zooxanthellate corals. Proc Natl Acad Sc 108:9905–9909.

19. Wiens M, Krasko A, Perovic S, Müller WE (2003) Caspase-mediated apoptosis in sponges: Cloning and function of the phylogenetic oldest apoptotic proteases from Metazoa. Biochim Biophys Acta 1593:179–189.

20. Salvesen G, Walsh C (2014) Functions of caspase 8: the identified and the mysterious. Semin Immunol. 26:246–252.

21. Kortschak RD, Samuel G, saint R, Miller DJ (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.

22. Mydlarz LD, Fuess LE, Mann W, Pinzón JH, Gochfeld DJ The cnidaria, past, present and future. In: Cnidarian immunity: From genomes to phenomes. ; 2016. pp. 441–466.

23. Shrestha S, Tung J, Grinshpon RD, Swartz P, Hamilton PT, Dimos B, Mydlarz L, Clark AC (2020) Caspases from scleractinian coral show unique regulatory features. J Biol Chem 295:14578–14591.

24. Xu L, Yuan S, Li J, Ruan J, Huang S, Yang M, Huang H, Chen S, Ren Z, Xu A (2011) The conservation and uniqueness of the caspase family in the basal chordate, amphioxus. BMC Biol 9.

25. Bell RA v, Megeney LA (2017) Evolution of caspase-mediated cell death and differentiation: twins separated at birth. Cell Death Differ 24:1359–1368.

26. Grinshpon RD, Shrestha S, Titus-Mcquillan J, Hamilton PT, Swartz PD, Clark AC (2019) Resurrection of ancestral effector caspases identifies novel networks for evolution of substrate specificity. Biochem J 476:3475–3492.

27. Bose K, Clark AC (2001) Dimeric procaspase-3 unfolds via a four-state equilibrium process. Biochemistry 40:14236–14242.

28. Bose K, Clark AC (2005) pH effects on the stability and dimerization of procaspase-3. Protein Sci 14:24–36.

29. Bose K, Pop C, Feeney B, Clark AC (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.

30. Grinshpon RD, Williford A, Titus-Mcquillan J, Clark AC (2018) The CaspBase: a curated database for evolutionary biochemical studies of caspase functional divergence and ancestral sequence inference. Protein Sci 27:1857–1870.

31. Gu Y, Wu J, Faucheul C, Lalanne J-L, Diu A, JLivingston D, S-SSu M (1995) Interleukin-13 converting enzyme requires oligomerization for activity of processed forms in vivo. EMBO J [Internet] 14:1923–1931. Available from: https://www.embopress.org/doi/10.1002/j.1460-2075.1995.tb07184.x

32. Walters J, Milam SL, Clark AC (2009) Practical approaches to protein folding and assembly: spectroscopic strategies in thermodynamics and kinetics. Methods Enzymol 455:1–39.

33. Keller N, Grütter MG, Zerbe O (2010) Studies of the molecular mechanism of caspase-8 activation by solution NMR. Cell Death Differ 17:710–718.

34. Zhang d, Lazim R (2017) Application of conventional molecular dynamics simulation in evaluating the stability of apomyoglobin in urea solution. Sci Rep 7.

35. Pop C, Chen YR, Smith B, Bose K, Bobay B, Tripathy A, Franzen S, Clark AC (2001) Removal of the pro-domain does not affect the conformation of the procaspase-3 dimer. Biochemistry 40:14224–14235.

36. Potter SC, Luciani A, Eddy SR, Park Y, Lopez R, Finn RD (2018) HMMER web server: 2018 update. Nucleic Acids Res 46:W200–W204.

37. Pei J, Kim BH, Grishin NV (2008) PROMALS3D: a tool for multiple protein sequence and structure alignments. Nucleic Acids Res 36:2295–2300.

38. Nguyen LT, Schmidt HA, von Haeseler A, Minh BQ (2015) IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol Biol Evol 32:268–274.

39. Cohen SL, Chait BT (1997) Mass spectrometry of whole proteins eluted from sodium dodecyl sulfate- polyacrylamide gel electrophoresis gels. Anal Biochem 247:257–267.

40. Chalkley RJ, Baker PR, Medzihradszky KF, Lynn AJ, Burlingame AL (2008) In-depth analysis of tandem mass spectrometry data from disparate instrument types. Mol Cell Proteomics 7:2386–2398.

41. Waterhouse A, Bertoni M, Bienert S, Studer G, Tauriello G, Gumienny R, Heer FT, de Beer TAP, Rempfer C, Bordoli L, et al. (2018) SWISS-MODEL: homology modelling of protein structures and complexes. Nucleic Acids Res 46:W296–W303.

42. Gordon JC, Myers JB, Folta T, Shoja V, Heath LS, Onufriev A (2005) H++: a server for estimating pK a s and adding missing hydrogens to macromolecules. Nucleic Acids Res:W368–W371.

43. Smith LJ, Berendsen HJC, van Gunsteren WF (2004) Computer simulation of urea-water mixtures: a test of force field parameters for use in biomolecular simulation. J Phys Chem B 108:1065–1071.

44. Hanwell MD, Curtis DE, Lonie DC, Vandermeersch T, Zurek E, Hutchison GR (2012) Avogadro: an advanced semantic chemical editor, visualization, and analysis platform. J Cheminform 4.

45. Stumpe MC, Grubmu H (2007) Aqueous Urea Solutions: Structure, Energetics, and Urea Aggregation. Available from: https://pubs.acs.org/sharingguidelines

46. Nosé S (1991) Constant Temperature Molecular Dynamics Methods. Progress of Theoretical Physics Supplement:1–46.

47. Hoover WG (1985) Canonical dynamics: Equilibrium phase-space distributions. Phys Rev A Gen Phys 31:1695–1697.

48. Abraham MJ, Murtola T, Schulz R, Páll S, Smith JC, Hess B, Lindahl E (2015) Gromacs: high performance molecular simulations through multi-level parallelism from laptops to supercomputers. SoftwareX 1–2:19–25ss

Figure 1.
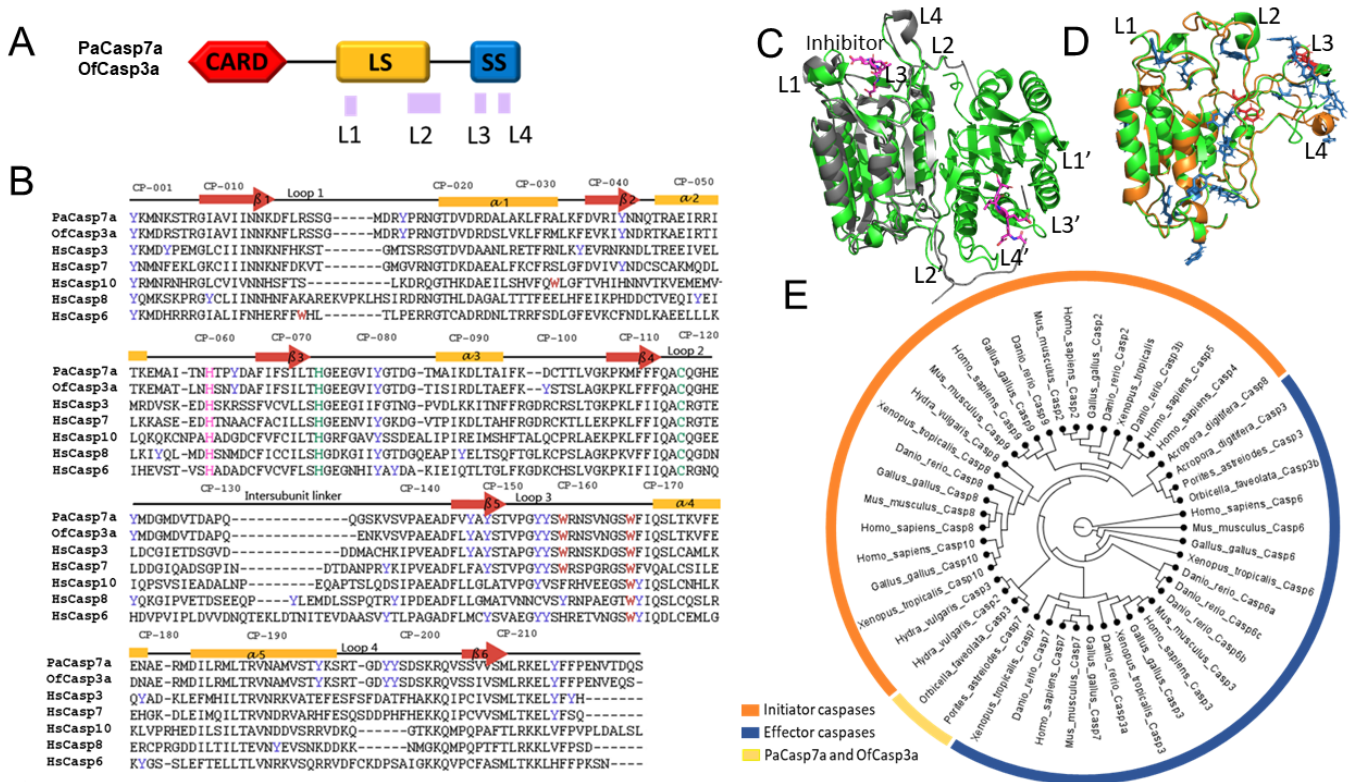
Figure 1. Phylogenetic relationship and caspase structure. (A) Domain organization of the protomer and location of active site loops, L1-L4, on coral caspases. The CARD domain precedes the large and the small subunit of the protease domain. (B) Multiple sequence alignment of coral caspases with human caspases showing secondary structural elements (loops, beta sheets, alpha helices) along with the common position (CP) numbers among caspases. Tyrosine residues (blue), tryptophan residues (red), active site catalytic residues His and Cys (green) and a conserved His residue (magenta) are highlighted. Sequences of the pro-domains are not shown. (C) Comparison of active dimeric PaCasp7a (green) (PDB ID: 6WI4) superimposed with HsCasp-3 protomer (gray) (PDB ID: 2J30) indicating active site loops L1-L4 on protomer 1 and L1'-L4' on protomer 2. (D) Predicted structures of the protomer of PaCasp7a (green) and of OfCasp3a (orange) built using NMR-model structure of HsproCasp8 (PDB ID: 2k7z) with an intact intersubunit linker, as described in the text. Tryptophan (red) and tyrosine (blue) residues are highlighted. (E) Phylogenetic tree of apoptotic caspases.

Figure 2.
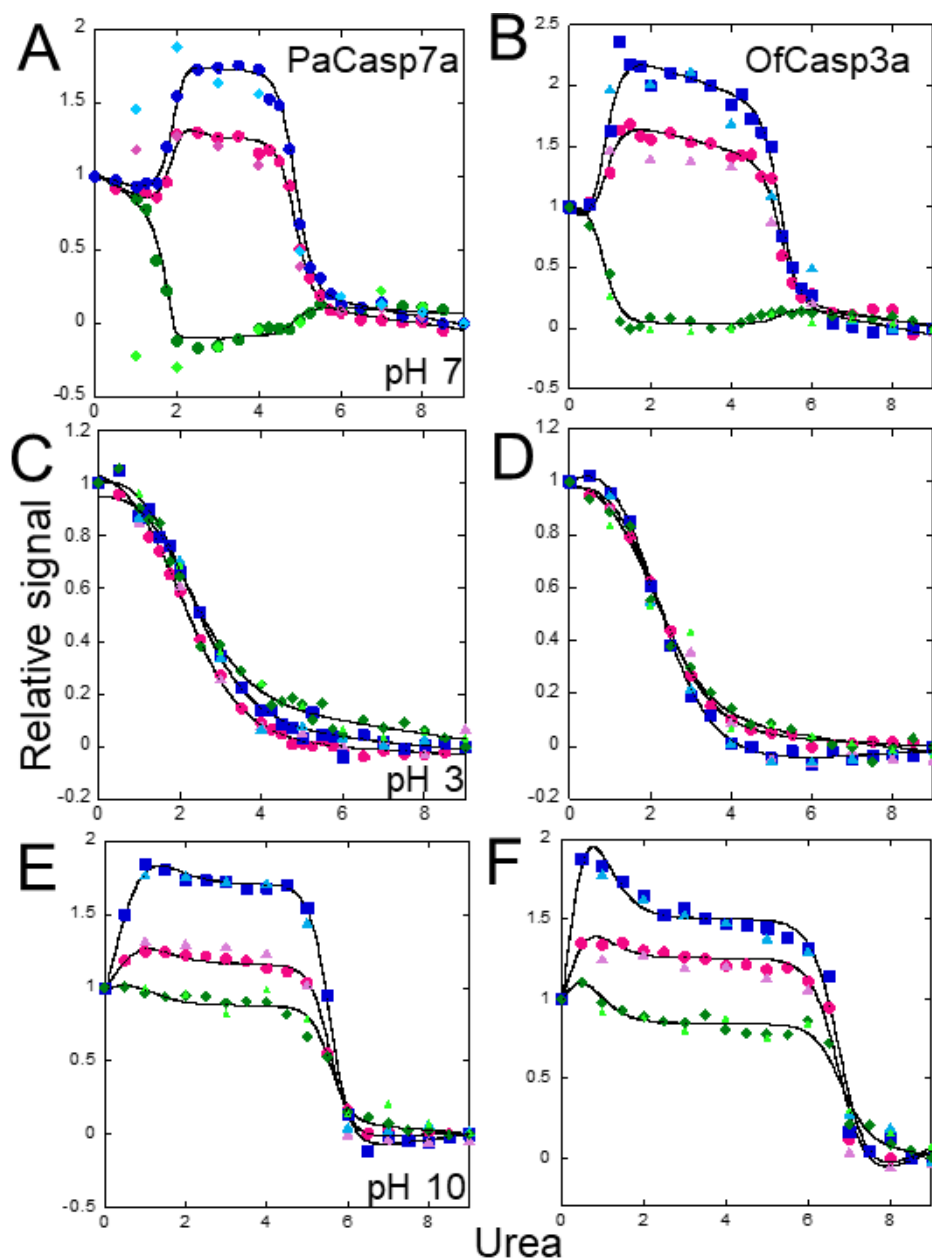
Figure 2. Normalized equilibrium unfolding data for PaCasp7a (left panels) at pH 7 (A), pH 3 (C) and pH 10 (E). Normalized equilibrium unfolding data for OfCasp3a (right panels) at pH 7 (B), pH 3 (D) and pH 10 (F). Colored solid symbols represent averaged raw data and solid lines through the data represent the global fits. Excitation at 280 nm- unfolding (●) and refolding (▲) , excitation at 295 nm - unfolding (■) and refolding (▲), CD - unfolding (◆) and refolding (▲)

Table 1.

pH	Equilibrium mechanism	Free energy change (ΔG kcal mol^{-1})	m-value (kcal mol^{-1} M^{-1})	ΔG^0_{total} (kcal mol^{-1})	m_{total} (kcal mol^{-1} M^{-1})
pH 3	I_2 to U	1.6 ± 0.2	0.80 ± 0.1	1.6 ± 0.2	0.80 ± 0.1
pH 7	N to I_1	2.2 ± 0.3	1.05 ± 0.3		
	I_1 to I_2	1.5 ± 0.4	1.10 ± 0.4		
	I_2 to U	11.5 ± 0.1	2.30 ± 0.3	14.9 ± 0.3	4.45 ± 0.4
pH 10	I_1 to I_2	0.9 ± 0.3	1.41 ± 0.3		
	I_2 to U	11.0 ± 0.1	1.97 ± 0.2	11.9 ± 0.8	3.38 ± 0.3

Table 1. Free energy changes and co-operativity index of PaCasp7a for each unfolding transition.

Table 2.

pH	Equilibrium mechanism	Free energy change (ΔG kcal mol $^{-1}$)	m-value (kcal mol $^{-1}$ M $^{-1}$)	ΔG^0_{total} (kcal mol $^{-1}$)	m_{total} (kcal mol $^{-1}$ M $^{-1}$)
pH 3	I_2 to U	1.7 ± 0.3	0.86 ± 0.1	1.7 ± 0.3	0.86 ± 0.1
pH 7	N to I_1	2.0 ± 0.4	1.80 ± 0.3		
	I_1 to I_2	1.9 ± 0.5	1.20 ± 0.5		
	I_2 to U	13.4 ± 0.6	2.60 ± 0.8	17.4 ± 0.5	5.6 ± 0.5
pH 10	I_1 to I_2	0.8 ± 0.2	1.50 ± 0.3		
	I_2 to U	11.5 ± 0.1	1.70 ± 0.2	12.3 ± 0.1	3.2 ± 0.2

Table 2. Free energy changes and co-operativity index of OfCasp3a for each unfolding transition.

Figure 3.
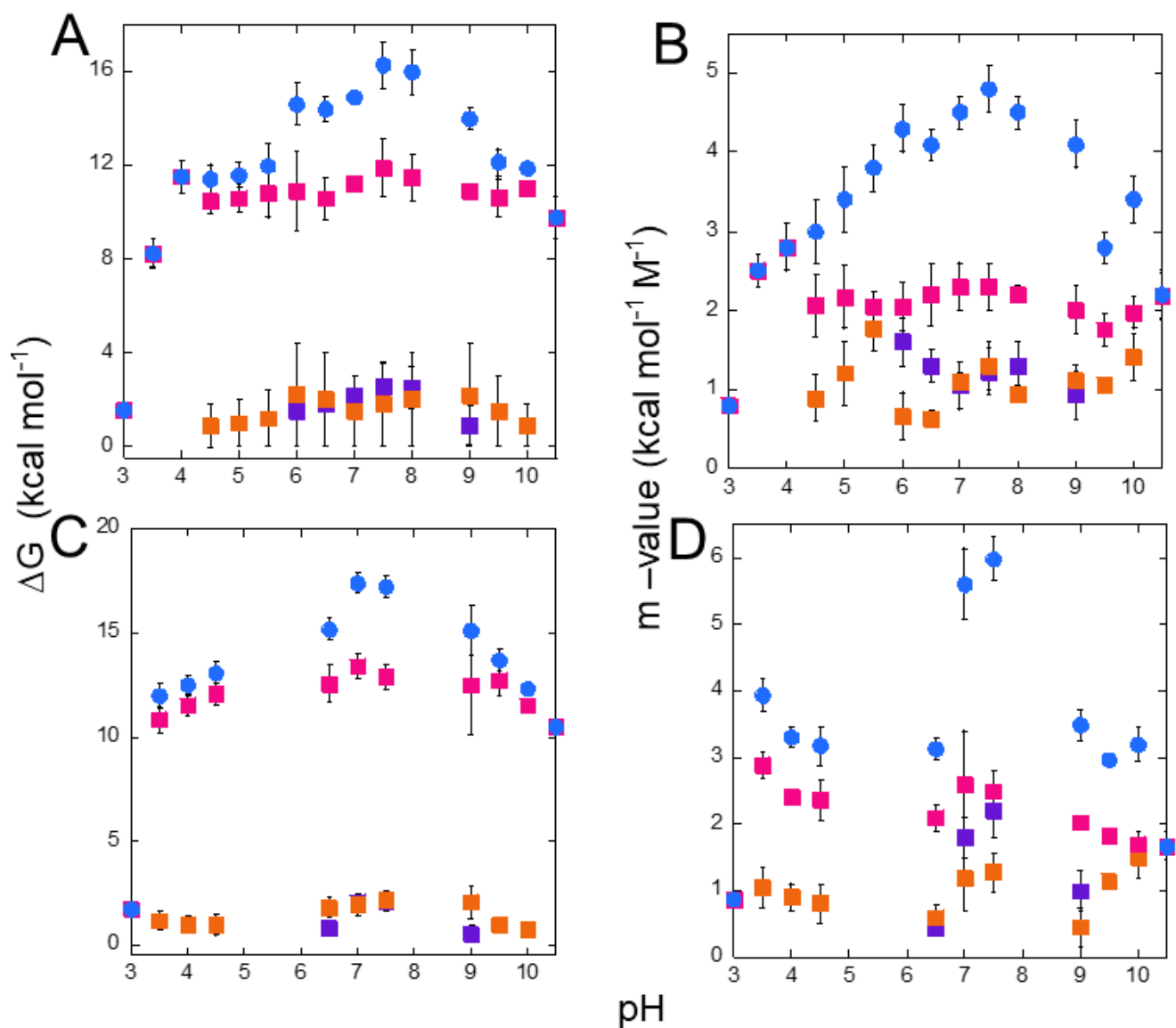
Figure 3. Conformational free energy as a function of pH. PaCasp7a (A,B), OfCasp3a (C,D). For panels A-D. the following symbols were used: ΔG_{total} (●), ΔG_1 (■), ΔG_2 (■), ΔG_3 (■). Cooperativity indices: m_{total} (●), m_1 (■), m_2 (■), m_3 (■). Error bars are from supplementary tables S1 and S2.

Figure 4.
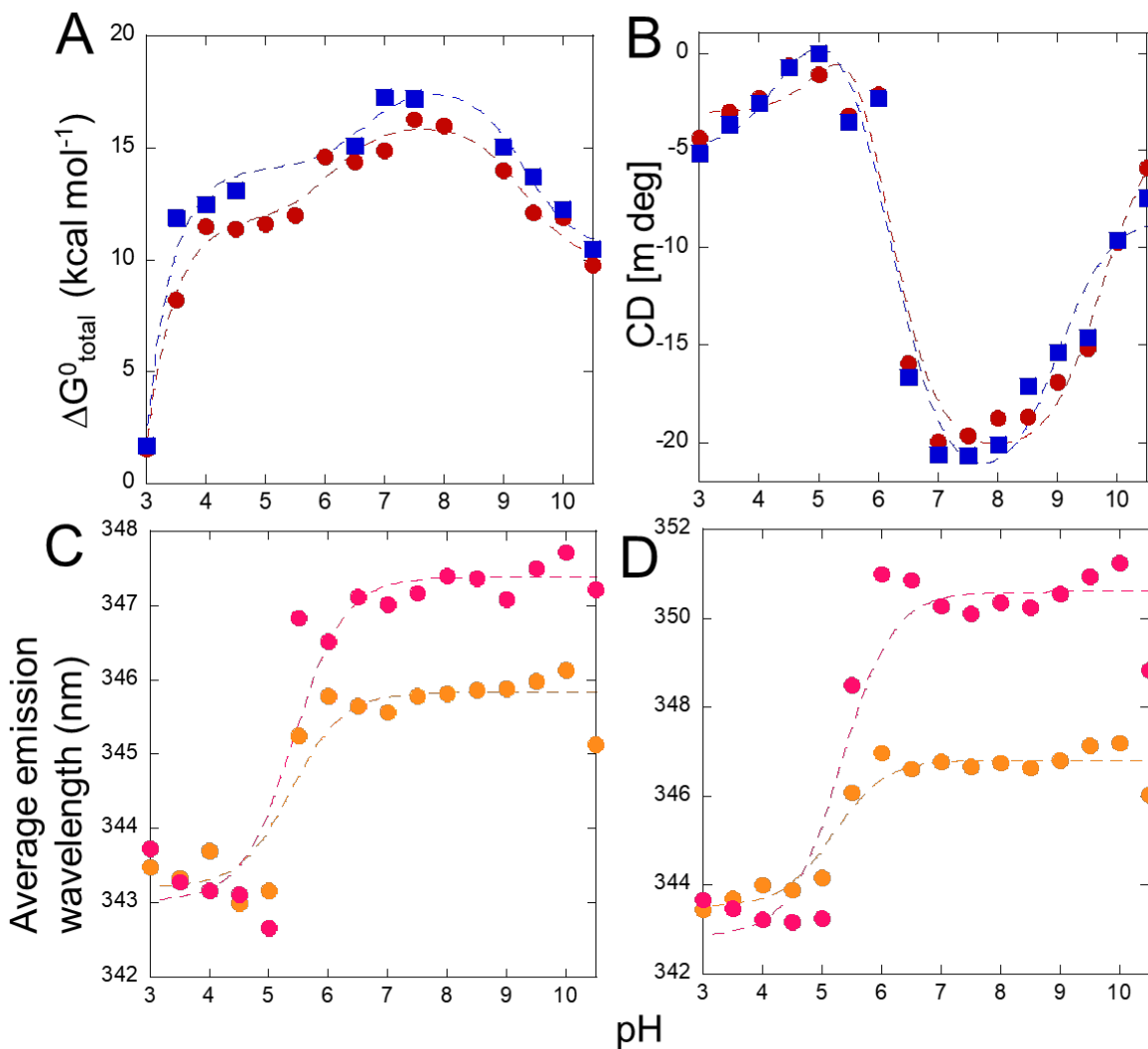
Figure 4. Effect of pH on monomeric coral caspases. (A) Comparison of the ΔG^0_{total} for PaCasp7a (●) and OfCasp3a (■). (B) Change in CD signal for native PaCasp7a at 232 nm(●) and native OfCasp3a at 230 nm (■). (C) Change in average emission wavelength versus pH for PaCasp7a (C) and for OfCasp3a (D). For panels C and D, samples were excited at 280nm (●) or 295nm (●). For A-D, dashed lines represent fits to the data to determine the pKa of the transitions, as described in the text.

Figure 5.
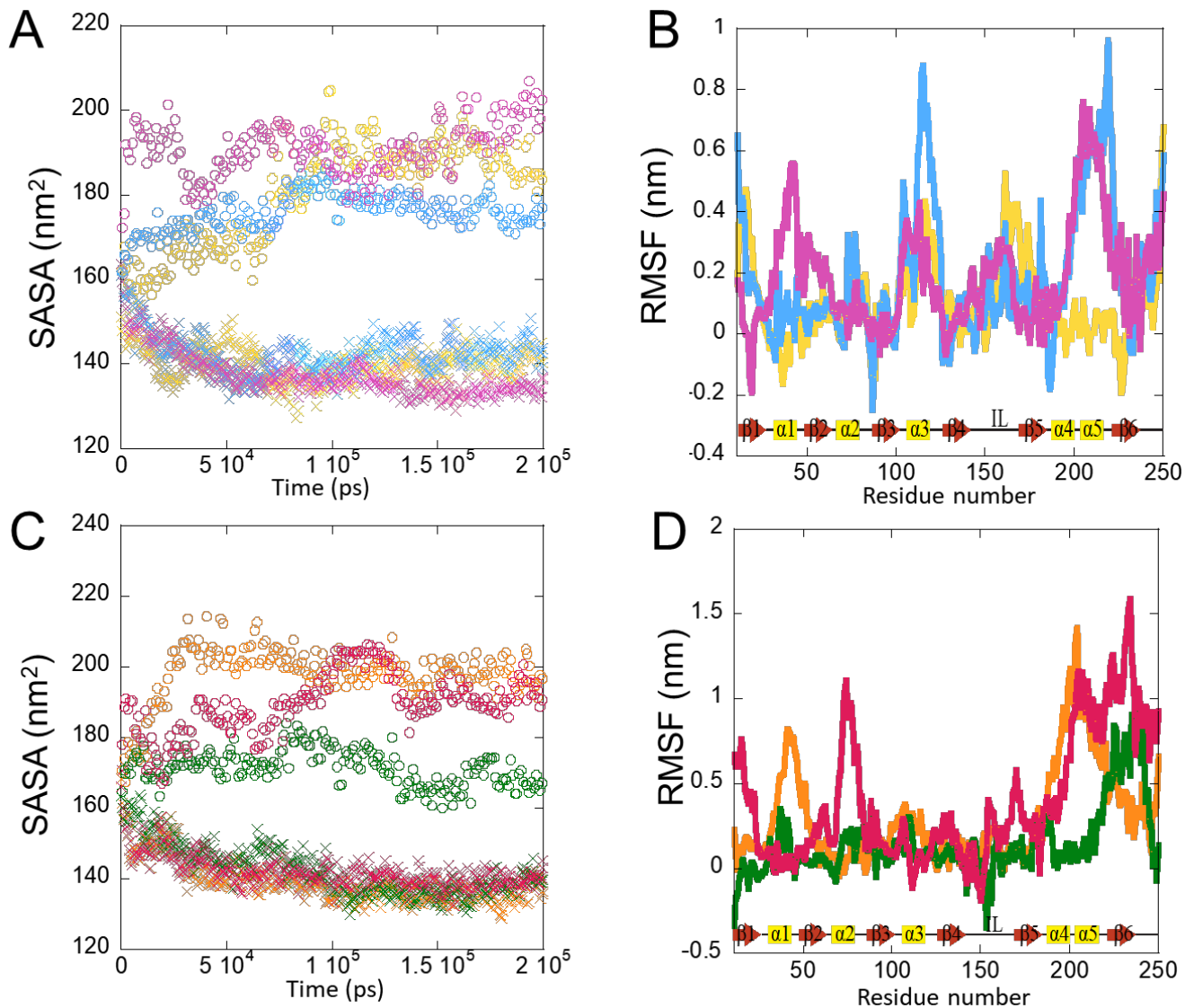
Figure 5. MD simulations of PaCasp7a and OfCasp3a. Panels A and C, solvent accessible surface area (SASA) of PaCasp7a (A) and of OfCasp3a (C) in water (crosses) and in 5 M urea (open circles). Panels B and D, ΔRMSF plots of 200 ns MD simulations with baseline correction as described in the text for PaCasp7a (B) and for OfCasp3a (D). For panels A and B, the following colors were used: pH 4 (yellow), pH 7 (blue), pH9 (magenta). For panels C and D, the following colors were used: pH 4 (orange), pH 7 (green), pH9 (pink).

Figure 6
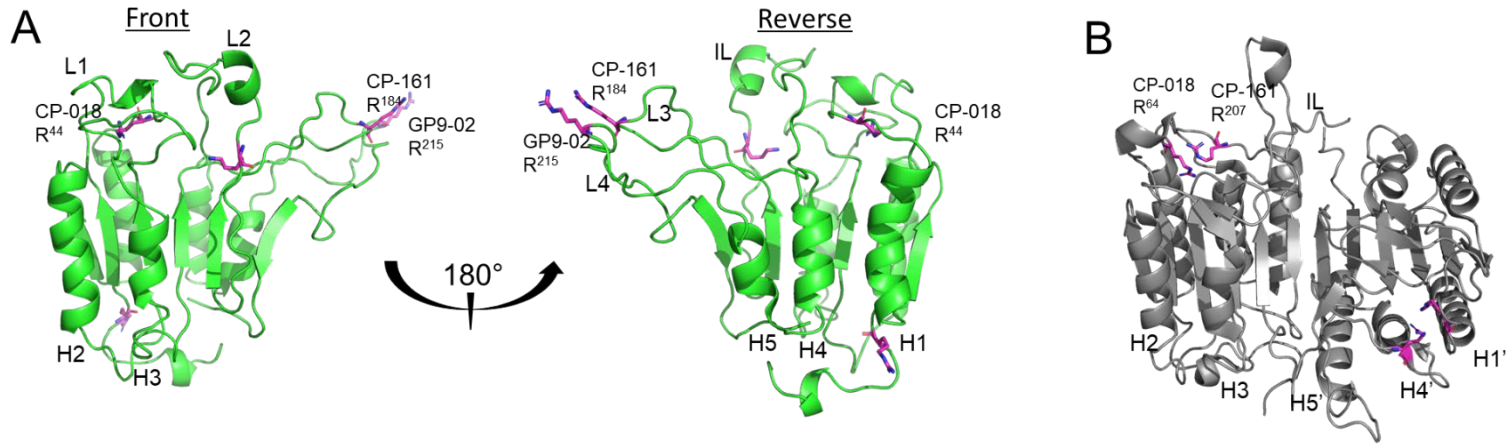
Figure 6. Limited trypsin proteolysis of PaCasp7a. (A) PaCasp7a protomer modeled using PDB ID:2k7z (B) Dimeric human caspase-3 (PDB ID:2J30). Common trypsin cleavage sites between the two structures are highlighted (magenta) and generate fragments as described in the text. α-helices (H1-H5) (on the first protomer) and H1'-H5' on the second protomer (B) and active site loops (L1-L4) are labeled.

Supplementary information

Sequential unfolding mechanisms of monomeric caspases

Isha Joglekar,[1,] and A. Clay Clark[1,*]

[1]Department of Biology, University of Texas at Arlington, Arlington, Texas, 76019

Running title: pH-effects on the stability of monomeric caspases

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

This file includes:

Supplementary figures S1-S8

Supplementary tables S1-S3
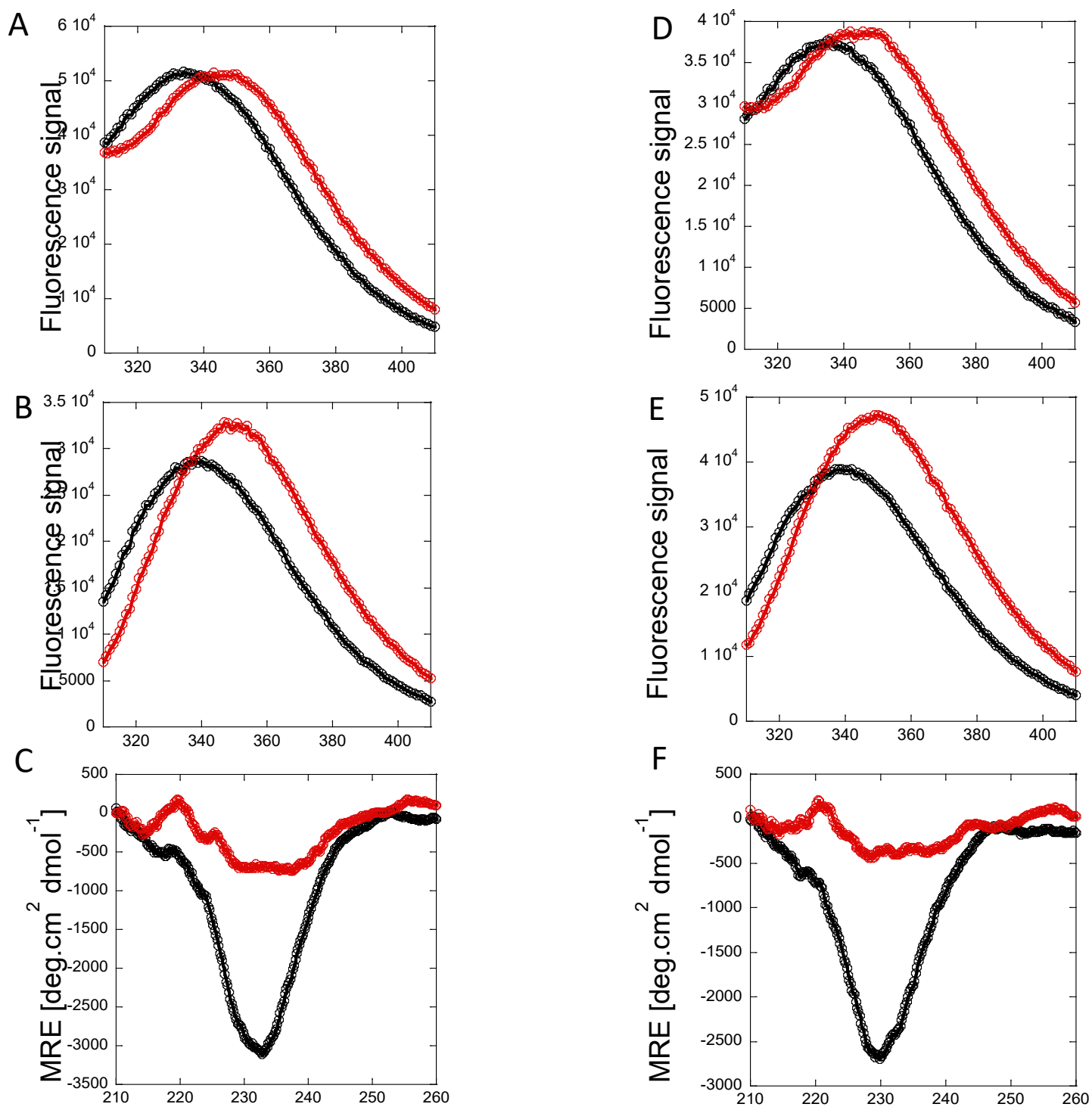
Figure S1. Fluorescence emission and circular dichroism spectra of PaCasp7a and of OfCasp3a at pH 7, 25°C. Fluorescence emission spectra of PaCasp7a following excitation at 280 nm (A) or 295 nm (B), and CD spectra (C). Fluorescence emission spectra of OfCasp3a following excitation at 280 nm (D) or 295 nm (E), and CD spectra (F). For A-F, the following symbols were used: proteins in buffer containing zero urea (●), or 9M urea (●)

Supplementary figure S2.
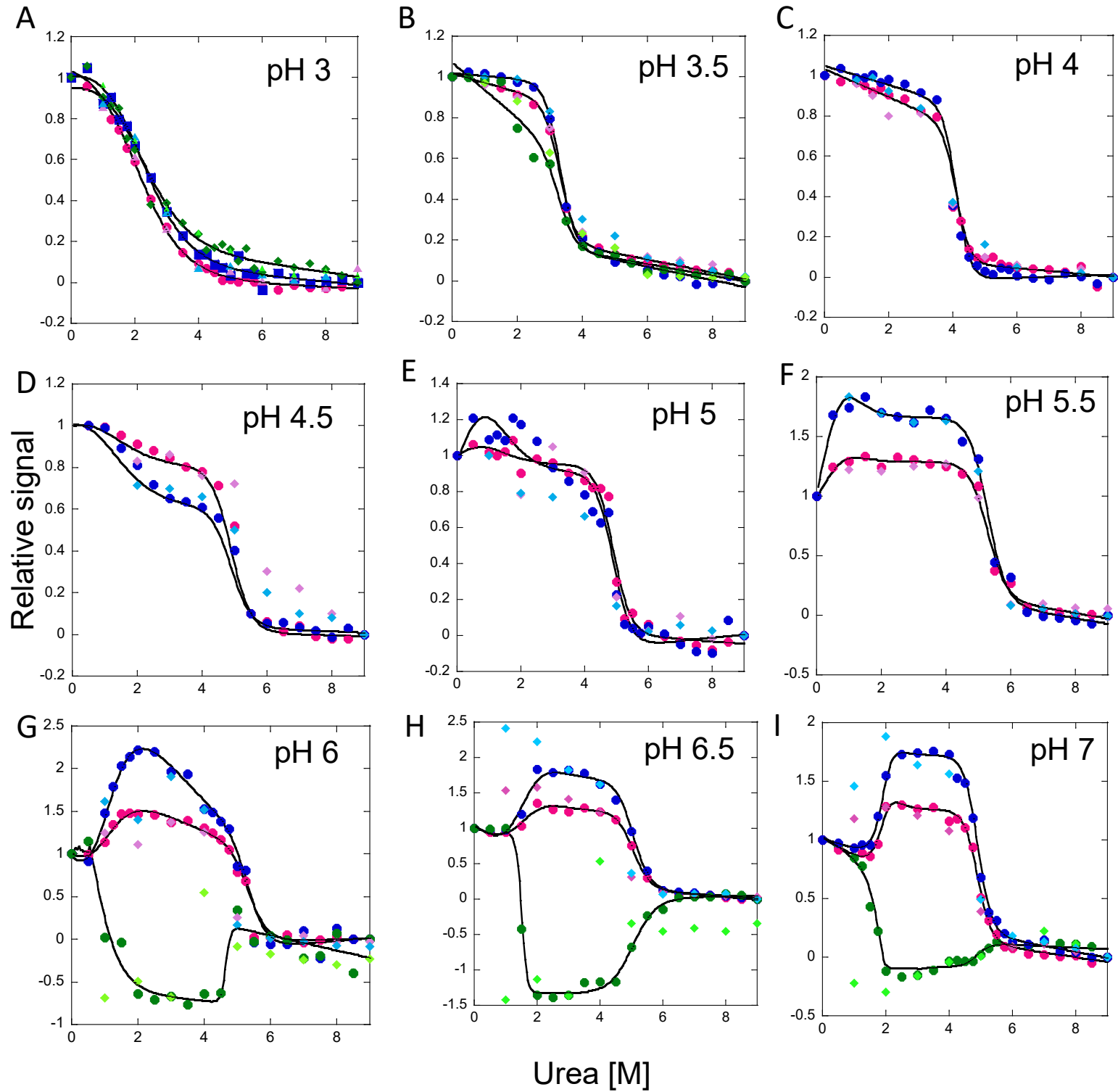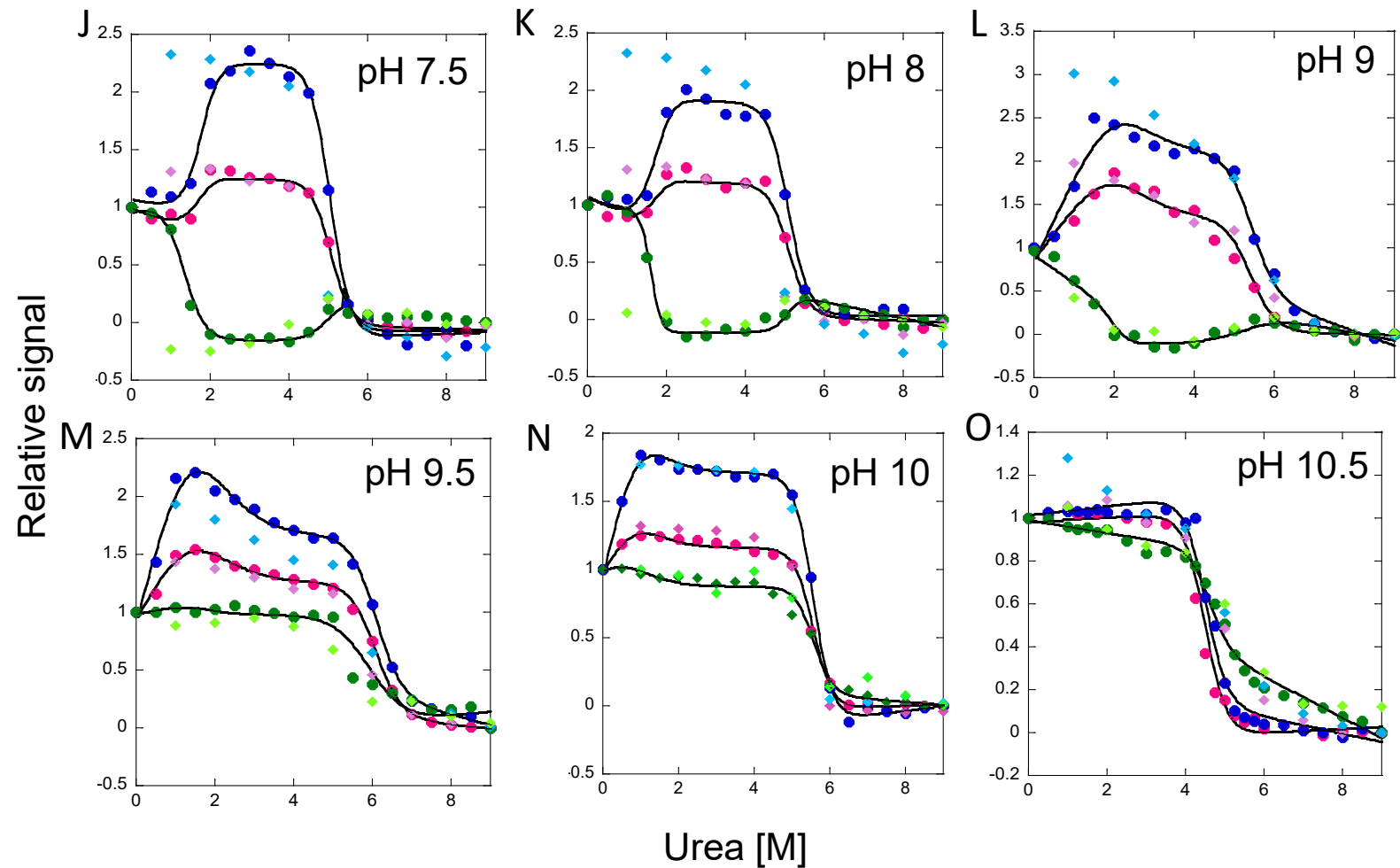
Figure S2. Normalized CD and equilibrium unfolding data for PaCasp7a from pH 3 to pH 10.5. Colored solid symbols represent averaged raw data and solid lines through the data represent the global fits, as described in the main text. For A-O, the following symbols were used: excitation at 280 nm- unfolding(⬤) and refolding(◆) , excitation at 295 nm - unfolding (⬤) and refolding (◆), CD - unfolding (⬤) and refolding (◆)

Supplementary figure S3.
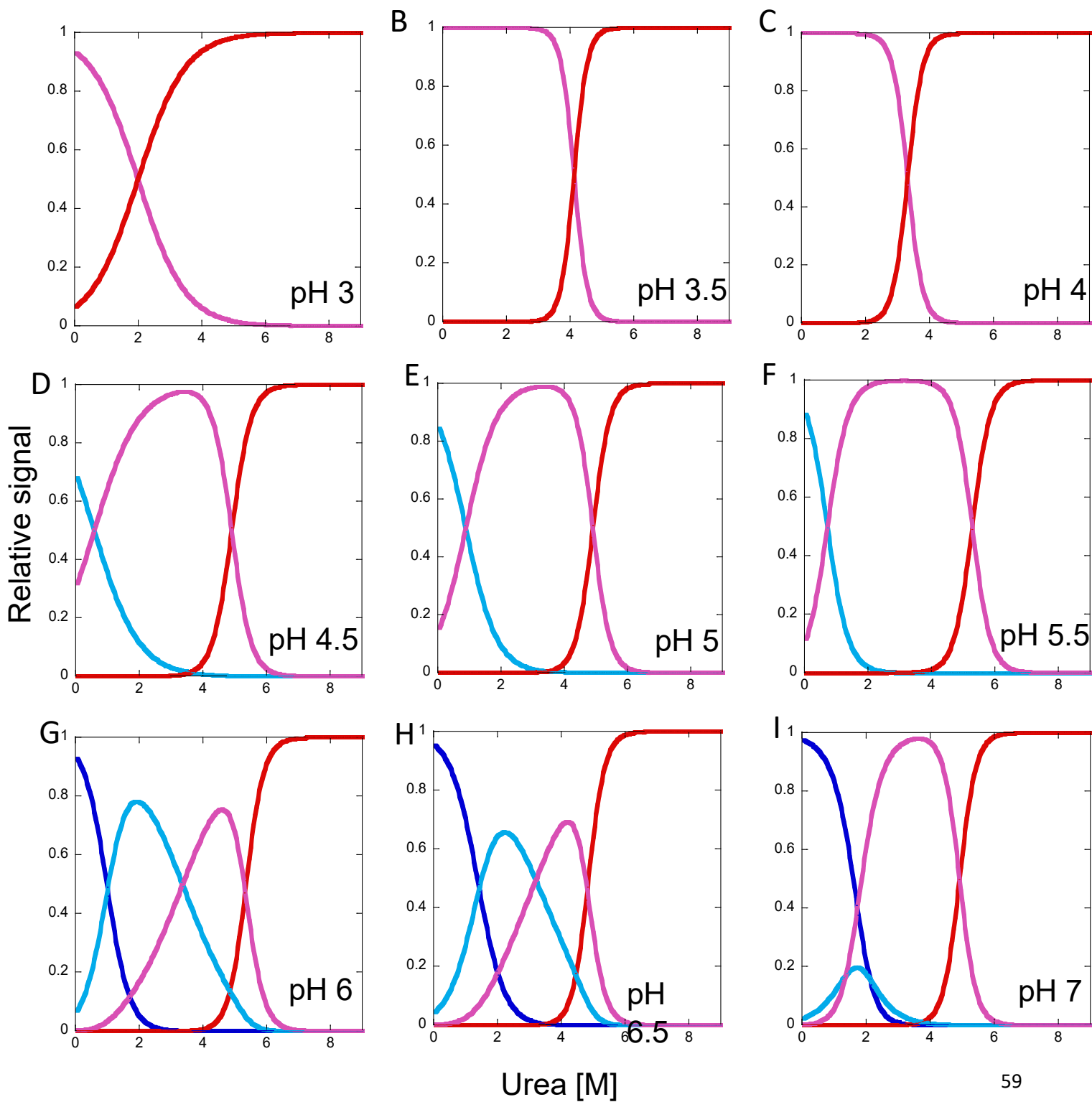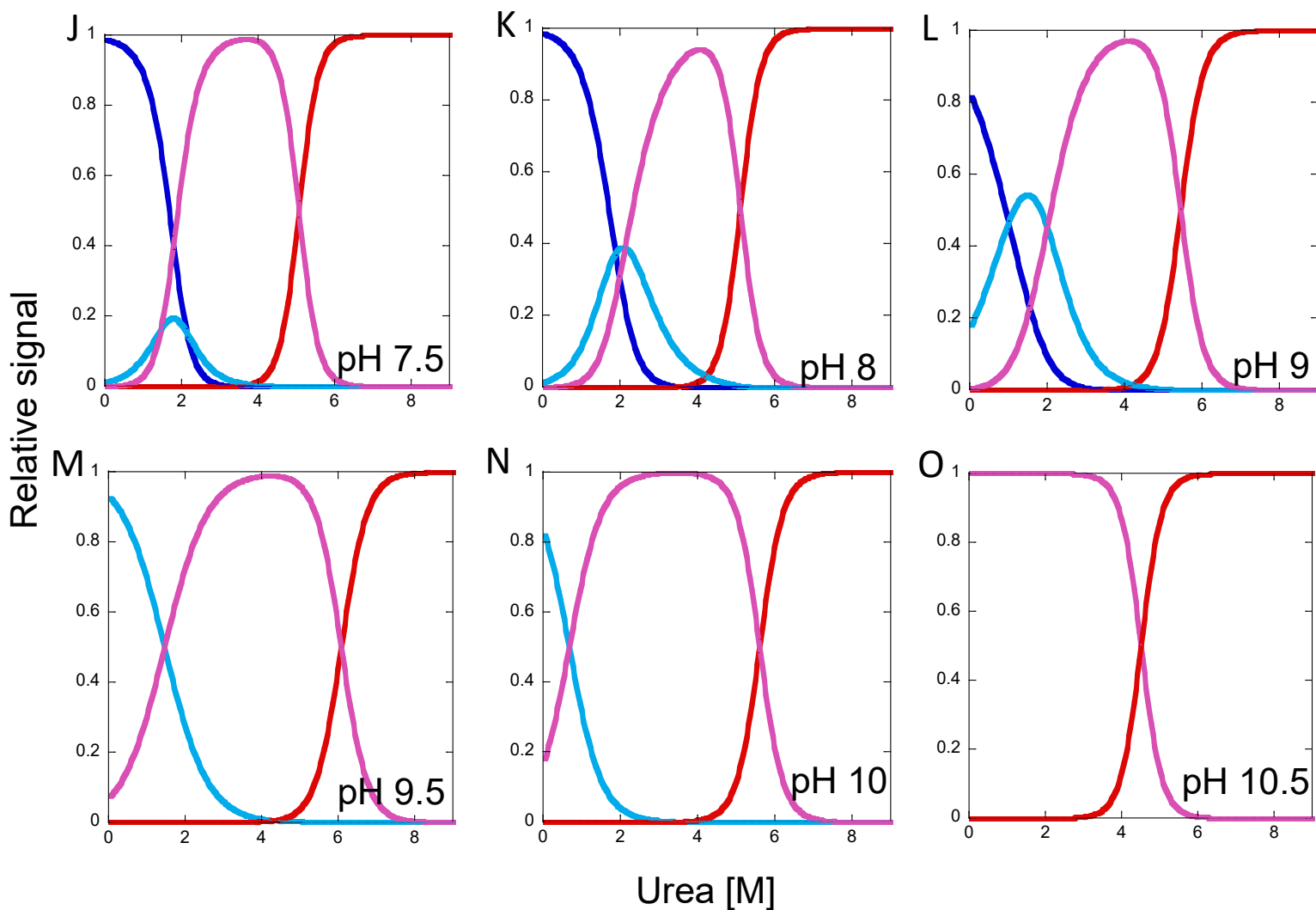
Figure S3.Fraction of species of PaCasp7a equilibrium unfolding as a function of urea concentration from pH 3 to pH 10.5. The fractions of native, intermediate$_1$, intermediate$_2$ and unfolded protein were calculated as a function of urea concentration from fits through the data at each pH. Fraction of species – native (–) , intermediate$_1$ (–) , intermediate$_2$ (–) and unfolded protein (–)

Supplementary figure S4.
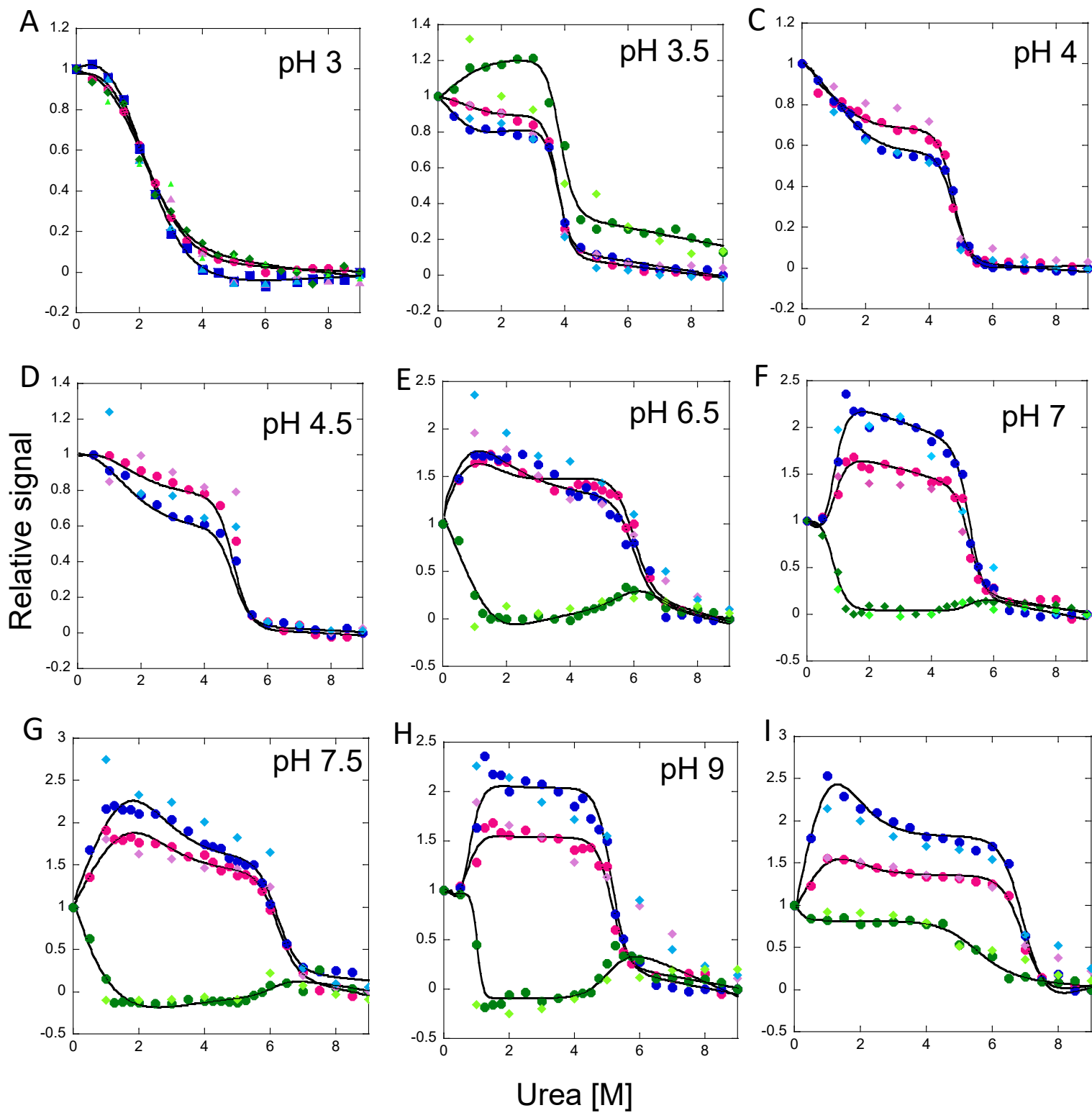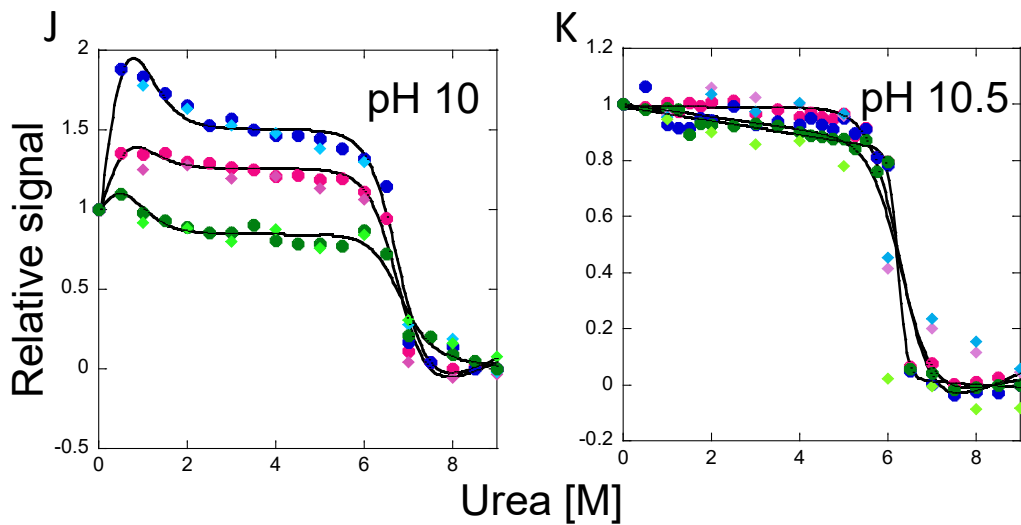
Figure S4. Normalized CD and equilibrium unfolding data for OfCasp3a from pH 3 to pH 10.5. Colored solid symbols represent averaged raw data and solid lines through the data represent the global fits, as described in the main text. For A-K, the following symbols were used: excitation at 280 nm- unfolding(⬤) and refolding(◆) , excitation at 295 nm - unfolding (⬤) and refolding (◆), CD data - unfolding (⬤) and refolding (◆)

Figure S5. Fraction of species of OfCasp3a equilibrium unfolding as a function of urea concentration from pH 3 to pH 10.5. The fractions of native, intermediate$_1$, intermediate$_2$ and unfolded protein were calculated as a function of urea concentration from fits through the data at each pH. Fraction of species - native(—) , intermediate$_1$ (—) , intermediate$_2$ (—) and unfolded protein (—)

Supplementary table S1. Summary of free energy changes and co-operativity index (m-values) for each transition in equilibrium unfolding of PaCasp7a

pH	ΔG_1 (kcal m^{-1})	m_1 (kcal m^{-1} M^{-1})	ΔG_2 (kcal m^{-1})	m_2 (kcal m^{-1} M^{-1})	ΔG_3 (kcal m^{-1})	m_3 (kcal m^{-1} M^{-1})	Total $\Delta G°_{conf}$ (kcal mol^{-1})
3					1.6 ± 0.2	0.80 ± 0.1	1.6 ± 0.2
3.5					8.3 ± 0.6	2.51 ± 0.2	8.3 ± 0.6
4					11.5 ± 0.7	2.80 ± 0.3	11.5 ± 0.7
4.5			0.9 ± 0.7	0.89 ± 0.3	10.5 ± 0.6	2.06 ± 0.4	11.4 ± 0.6
5			1.0 ± 0.5	1.21 ± 0.4	10.6 ± 0.6	2.17 ± 0.4	11.6 ± 0.5
5.5			1.2 ± 0.9	1.78 ± 0.3	10.8 ± 1.0	2.05 ± 0.2	12.0 ± 0.5
6	1.5 ± 0.2	1.61 ± 0.3	2.2 ± 0.8	0.66 ± 0.3	10.9 ± 1.7	2.05 ± 0.3	14.6 ± 0.9
6.5	1.8 ± 0.2	1.30 ± 0.2	2.0 ± 0.4	0.63 ± 0.1	10.6 ± 0.9	2.21 ± 0.4	14.4 ± 0.9
7	2.2 ± 0.3	1.05 ± 0.3	1.5 ± 0.4	1.10 ± 0.1	11.2 ± 0.1	2.30 ± 0.3	14.9 ± 0.5
7.5	2.6 ± 0.9	1.23 ± 0.3	1.8 ± 0.9	1.30 ± 0.3	11.9 ± 1.2	2.32 ± 0.3	16.3 ± 0.3
8	2.5 ± 0.9	1.30 ± 0.3	2.0 ± 1.0	0.95 ± 0.1	11.5 ± 1.0	2.21 ± 0.1	16.0 ± 1.0
9	0.9 ± 0.8	0.93 ± 0.3	2.2 ± 0.5	1.12 ± 0.2	10.9 ± 0.1	2.05 ± 0.3	13.9 ± 0.9
9.5			1.5 ± 0.4	1.05 ± 0.1	10.6 ± 0.8	1.75 ± 0.2	12.1 ± 0.3
10			0.9 ± 0.3	1.41 ± 0.3	11.0 ± 0.1	1.97 ± 0.2	11.9 ± 0.6
10.5					9.7 ± 0.9	2.18 ± 0.3	9.7 ± 0.2

Supplementary table S2. Summary of free energy changes and co-operativity index (m-values) for each transition in equilibrium unfolding of OfCasp3a

pH	ΔG_1 (kcal m $^{-1}$)	m_1 (kcal m $^{-1}$ M $^{-1}$)	ΔG_2 (kcal m $^{-1}$)	m_2 (kcal m $^{-1}$ M $^{-1}$)	ΔG_3 (kcal m $^{-1}$)	m_3 (kcal m $^{-1}$ M $^{-1}$)	Total $\Delta G°_{conf}$ (kcal mol $^{-1}$)
3					1.7 ± 0.3	0.86 ± 0.1	1.7 ± 0.3
3.5			1.1 ± 0.5	1.05 ± 0.3	10.8 ± 0.6	2.89 ± 0.2	12.0 ± 0.5
4			1.0 ± 0.4	0.90 ± 0.2	11.5 ± 0.5	2.41 ± 0.1	12.5 ± 0.4
4.5			1.0 ± 0.5	0.81 ± 0.3	12.1 ± 0.6	2.36 ± 0.3	13.1 ± 0.5
6.5	0.8 ± 0.2	0.44 ± 0.1	1.8 ± 0.5	0.60 ± 0.2	12.5 ± 0.9	2.09 ± 0.2	15.1 ± 0.5
7	2.0 ± 0.4	1.80 ± 0.3	1.9 ± 0.5	1.20 ± 0.5	13.4 ± 0.6	2.60 ± 0.8	17.3 ± 0.5
7.5	2.1 ± 0.5	2.20 ± 0.4	2.2 ± 0.4	1.28 ± 0.3	12.9 ± 0.6	2.50 ± 0.3	17.2 ± 0.5
9	0.5 ± 0.4	1.0 ± 0.3	2.0 ± 0.8	0.45 ± 0.3	12.5 ± 2.4	2.03 ± 0.1	15 ± 0.3
9.5			1.0 ± 0.3	1.13 ± 0.1	12.7 ± 0.7	1.83 ± 0.1	13.7 ± 0.5
10			0.8 ± 0.2	1.50 ± 0.3	11.5 ± 0.1	1.70 ± 0.2	12.3 ± 0.1
10.5					10.5 ± 0.2	1.67 ± 0.2	10.5 ± 0.2
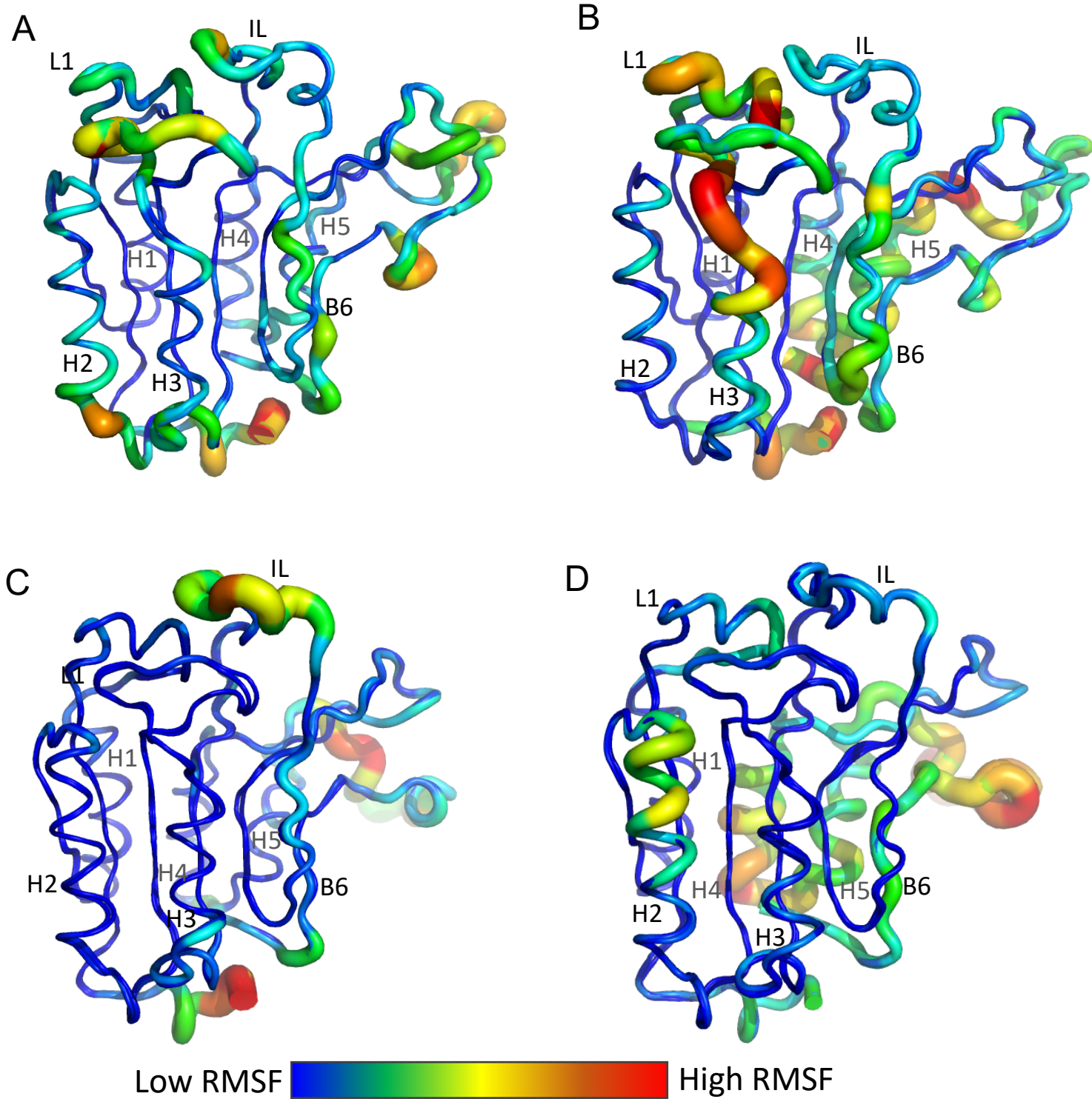
Figure S6. RMSF values from MD simulations converted to B-factors, overlaid and mapped onto the modeled structure of PaCasp7a (top panels) and OfCasp3a (bottom panels). PaCasp7a in water (A) or in 5M urea(B); OfCasp3a in water (C) or in 5M urea (D).

67

Supplementary figure S7.
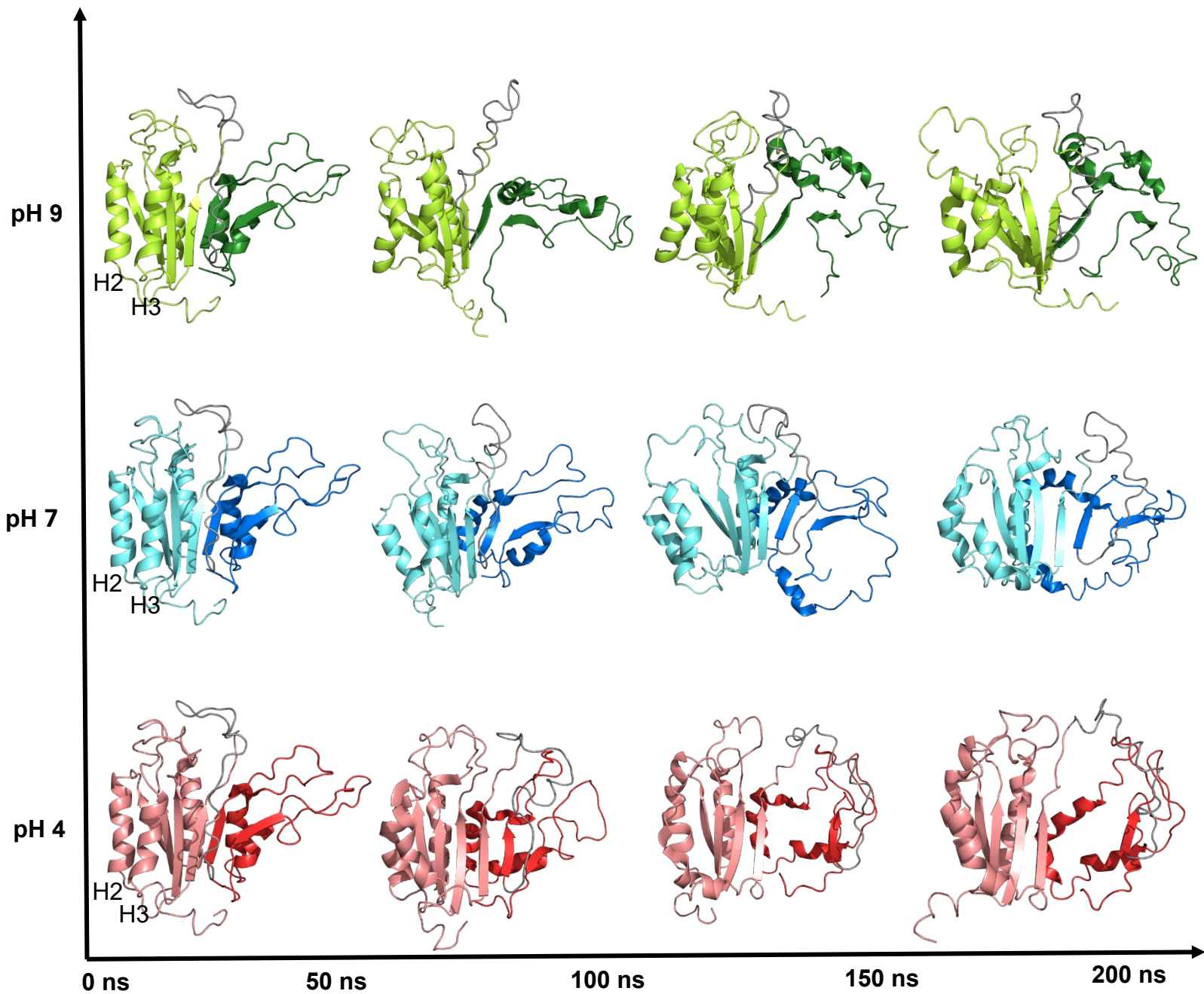
Figure S7. Urea MD snapshots of PaCasp7a at pH 4 (LS - salmon, SS - red), pH 7 (LS - cyan, SS - blue), and pH 9 (LS - limon, SS - green) with helices 2 and 3 (H2, H3) labeled on the structure at 0 ns.

Supplementary figure S8.
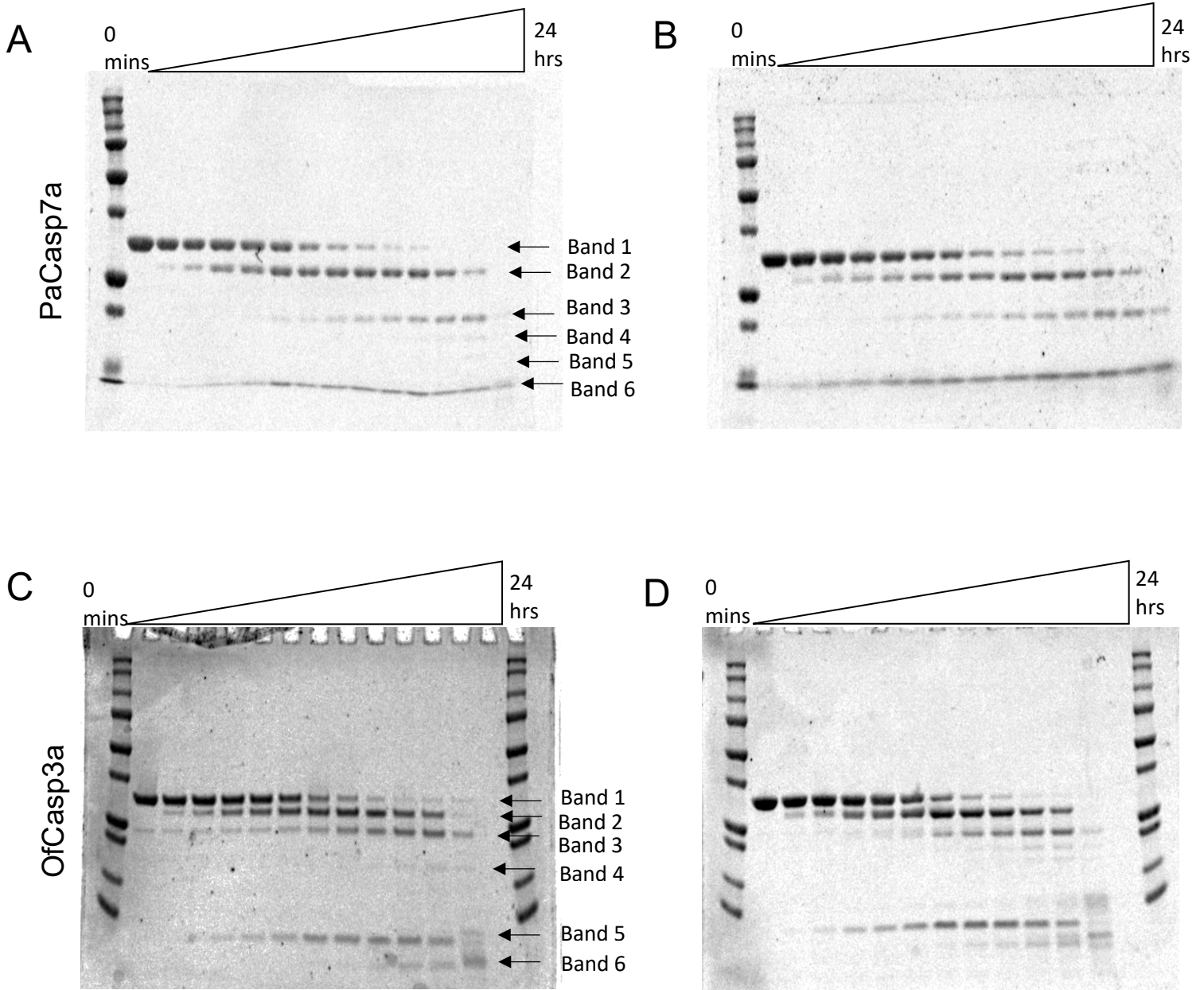
Figure S8. Limited trypsin digestion of PaCasp7a (A,B) and of OfCasp3a (C,D) at pH 7 (panels B and D) and pH 9 (panels C and D). The 31 kDa band observed at 0 min represents the native proteins, whereas the 26 kDa (band 2), 20 kDa (band 3), 18 kDa (band 4), 15 kDa (band 5), and 9 kDa (band 6) are the cleavage products as described in the text.

Supplementary table S3. Caspases used in the phylogenetic analysis
and their respective accession numbers.

Caspases	Accession number
Homo sapiens	
HsCasp-2	NP_116764.2
HsCasp-3	NP_004337.2
HsCasp-4	NP_001216.1
HsCasp-5	NP_004338.3
HsCasp-6	NP_001217.2
HsCasp-7	NP_001253985.1
HsCasp-8	NP_001219.2
HsCasp-9	NP_001220.2
HsCasp-10	NP_116759.2
Mus musculus	
MmCasp-2	NP_031636.1
MmCasp-3	NP_001271338.1
MmCasp-6	NP_033941.3
MmCasp-7	XP_006526679.1
MmCasp-8	NP_001264855.1
MmCasp-9	NP_056548.
Gallus gallus	
GgCasp-2	NP_001161173.1
GgCasp-3	NP_990056.1
GgCasp-6	NP_990057.1
GgCasp-7	XP_421764.3
GgCasp-8	NP_989923.1
GgCasp-9	XP_424580.5
GgCasp-10	XP_421936.4
Xenopus tropicalis	
XtCasp-2	XP_012809163.2
XtCasp-3	NP_001120900.1
XtCasp-6	NP_001011068.1
XtCasp-7	NP_001016299.1
XtCasp-8	XP_017953067.2
XtCasp-9	NP_001116935.1
XtCasp-10	NP_001015715.2

Caspases	Accession number
Danio rerio	
DrCasp-2	NP_001036160.1
DrCasp-3a	XP_001338890.2
DrCasp-3b	XP_005173133.1
DrCasp-6a	XP_005164109.1
DrCasp-6b	XP_017210076.1
DrCasp-6c	NP_001018333.1
DrCasp-7	XP_005156389.1
DrCasp-8	NP_001092089.1
DrCasp-9	NP_001007405.2
Coral caspases	
Acropora digitifera Casp-3	XP_015775441.1
Acropora digitifera Casp-8	XP_015761120.1
Hydra vulgaris Casp-2	NP_001274285.1
Hydra vulgaris Casp-3	XP_012557085.1
Hydra vulgaris Casp-8	XP_012562456.1
Orbicella faveolata Casp-3a	XP_020613409.1
Orbicella faveolata Casp-3b	XP_020630525.1

CHAPTER 3

Exploring the Conformational Landscape and Allosteric Networks of Caspases through Free Energy and Network Analysis

Isha Joglekar[1‡], Mithun Nag Karadi Giridhar [1‡], David A. Diaz[1], Ankit Deo[2] and A. Clay Clark[1*]

[1]Department of Biology, University of Texas at Arlington, Arlington, Texas, 76019

[2]Department of Information Systems, University of Texas at Arlington, Arlington, Texas, 76019

Running title: Evolutionary analysis of the conformational landscape of caspases

[*]Corresponding author: A. Clay Clark

Email: clay.clark@uta.edu

Key Words: caspase; dimerization; apoptosis; protein evolution; energy landscape, conformational dynamics, network analysis

Abstract

This work investigates the poorly known relationship between modest evolutionary changes in the hydrophobic core to maintain the fold and evolve conformational dynamics. Through ancestral reconstruction, molecular dynamics simulation, network analysis, and free energy landscapes, we identify a conserved, one-billion-year-old scaffold that underlies the diverse conformational dynamics of caspases across different families of effector and initiator caspases in chordates. Our findings demonstrate how modular modifications in the biophysical properties of amino acids that connect the conserved scaffold to other structural elements can give rise to dynamic conformations and highlight the energetic basis of conformational fluctuations that guide the evolution of monomeric and dimeric scaffolds in protein families. In addition, we identify a network of residues crucial for allosteric communication and the evolution of electron tunnel networks for effective signal transmission in caspases.

Introduction

Caspases are an ancient family of cysteinyl proteases that play a crucial role in apoptosis and inflammation (1). However, in addition to regulating apoptosis for maintaining cellular homeostasis, they are critical in cell proliferation and differentiation at lower levels of activation, for which the mechanism is poorly understood (1, 2). Within the apoptotic caspases, the two subfamilies evolved from a common ancestor over 700 million years ago (MYA), which provided a framework for the evolution of unique characteristics within the subfamilies – initiator caspases (-8,-10, cFLIP) that exist as monomers and require dimerization for complete activation, while effector caspases (-3,-6,-7) that exist as stable dimers and require processing for complete activation. Importantly, all caspases are synthesized as inactive zymogens within a cell, and the oligomeric form of the zymogen is a key to regulation (3–6). The caspase family offers an attractive model for understanding protein evolution since a common protein scaffold (the caspase-hemoglobinase fold) was utilized to develop distinct oligomeric states, modify enzyme specificity and acquire novel allosteric sites (1, 4, 7, 8).

Studies of protein families have identified characteristics that contribute to enzyme specificity. The protein scaffold provides the appropriate conformational dynamics that facilitate interaction and substrate binding, or the enzyme active sites provide the stereo-selective environment for ground or transition states. Biological macromolecules exist as ensembles of states that fluctuate around an average structure that is governed by Boltzmann's principle (1). Since proteins exhibit a broad range of molecular motions, their ability to form sub-states due to their dynamic nature is crucial to their function (9). Several static structures of proteins, such as those of caspases provide a plethora of

information on various ground states in the conformational space, but little regarding the transitions between states, activation barriers between the ground states, or high-energy states that may exist (1, 10). Conformational diversity of proteins has been associated with different aspects related to biological function such as enzyme catalysis, signal transduction, protein recognition specificities, and allostery among a few others (11). These observations suggest that proteins have evolved to employ rearrangements of dynamic residue interactions to perform specific functions, and that subtle changes to this environment can result in significant structural rearrangements while maintaining the overall fold (12). Ancestral protein reconstruction (APR) can help infer the evolutionary changes in protein structure and function over time, as well as identify residues and domains that may be responsible for changes in conformational dynamics (13, 14).

In proteins, the evolution of the hydrophobic core and its relevance in the evolution of conformational dynamics, subfamilies, and function are poorly known (15). We reconstructed several ancestral proteins along the tree from the common ancestor (ancestor of all) to extant caspases involved in the extrinsic pathway of apoptosis and performed molecular dynamics simulations on modeled and available structures of three distinct conformations: the active, which represents mature caspase enzymes with a cleaved inter-subunit linker and an ordered active site primed for catalysis; the dimeric, which represents the enzyme in its dimeric state but with a disordered active site and an intact inter-subunit linker; and the monomeric, which represents the enzyme in its monomeric form before dimerization. Using network analysis, we classified the hydrophobic core into optimally packed and flexible regions resembling the simple nuts-

and-bolts model (16), and demonstrated how the movement of specific residues with distinct biophysical properties in the initiator and executioner subfamilies, which stabilize the optimally packed residues in the monomeric and dimeric conformations, can remodel the entire network. With the help of the free energy landscape analyses, we demonstrate how this remodeling might reduce the entropy of flexible chains in the core, thereby moving the packing thermodynamically toward a stable active conformation. In addition, we identified a key aromatic phenylalanine residue that does not make contact in any conformation but is the central participant in electron transport and therefore allosteric communication.

Results

Conformational dynamics of monomeric and dimeric states are influenced by intermediately conserved residues

To examine the impact of urea on various conformations in the caspase family since the common ancestor, we employed the degree centrality metric in network analysis. The degree values for each amino acid position in the alignment (Fig S1) were separately averaged for each conformation, classified according to secondary structural elements, and illustrated using violin plots (Fig 1A-F) to provide an overview of the 51 simulations in water and 51 in urea in the active, dimeric, and monomeric conformations. For simulations in water (Fig 1A-C), the degree distributions reveal that residues in beta-sheets and alpha-helices are more likely to create high-degree interactions than residues in other secondary structural components, irrespective of the conformation. Since atomic contacts were utilized to calculate degree centrality, the data suggest that residues in helices and beta-sheets display the maximum non-covalent interactions.

In the dimeric and active conformations, the short beta-sheet (on helix 3) and the bottom loops make more high-degree contacts, indicating their significance in stabilizing the core and the active site, respectively. For simulations in urea, we observe a decrease in high-degree contacts among all secondary structural elements in the active (Fig 1A,D), dimeric (Fig 1B,E), and monomeric (Fig 1C,F) conformations; however, the active conformation maintains high-degree contacts relatively well in comparison to the dimeric and monomeric (Fig 1D-F), suggesting that the active conformation is the most stable followed by the dimeric and monomeric conformations. This has been illustrated on different conformations of caspase-8 (Fig1), which exhibit varying degrees of unfolding. Furthermore, subclassification of the degree scores into three categories based on Consurf conservation scores, high conservation (score of 9-8), intermediate conservation (score of 7-6) and variable (score below 5) in figure S2 A,B,C show that the variable residues make the least contacts among the categories.

In summary, the data indicate that the dimer-to-active conformation exhibits fewer modifications than the monomer-to-dimer transition (Fig1). Beta sheets and helices make more high-degree interactions than other secondary structural components, with conserved and intermediately conserved residues contributing the most (Fig S2). This shows that beta sheet-helix interactions affect evolution of caspase conformational stability, which is driven by small changes to intermediately conserved residues. Since the active conformation is very stable and undergoes minimal conformational dynamics, the dimeric and monomeric data were extensively explored to investigate the evolution of the caspase conformational landscape.

The conformational stability in the dimeric and monomeric conformations is regulated by the evolution of amino acid interactions in the small subunit.

Proteins can have thousands of atoms, and the mobility of each atom over time creates large datasets that are difficult to analyze and interpret. Principal component analysis (PCA) reduces massive data sets to their major principal components, revealing the most significant dynamics. Fig S3 shows the MD trajectories projected along the first eigenvector. In extant caspases, structures in the first principle component (PC1) space reveal higher structural displacements in the initiators when compared to the effectors.

In general, projections of the first two PCs were used to construct the free energy landscapes (FELs), since they explain most of the variance in the data and can be associated with the most significant changes occurring in the proteins. On comparing the FEL in water (Fig S4) to that of urea (Fig 2A), one observes an enhanced sampling of atoms in urea depicted by the larger conformational space. We note that, although 200 ns in 8M urea is inadequate for the molecule to unfold completely, these investigations can be useful in demonstrating relative stability across caspases, with a wider landscape indicating more unfolding. For instance, the FEL of caspase-6 depicts the least accessible conformational space (Fig. 2A), indicating the lowest degree of unfolding. Further, these results corroborate with experimental findings, that show caspase-6 has the largest conformational free energy of all effector caspases. There have been no experimental reports on the free energy of the initiators in a dimeric conformation. In Figure 2B&C we show metastable states extracted from the FEL minima, highlighted in Fig2A which correspond to the last observed metastable state in urea for effector caspase-3 and initiator caspase-8, in dimeric conformation. The small

subunit of caspase-3 (Fig 2B) unfolds to a lesser extent than the small subunit of caspase-8 (Fig 2C), and this trend is observed on average in the initiator and effector tree (Fig S5). FEL of the initiator subfamily (from AOI to extant caspases) is wider than that of the effector subfamily (from AOE to extant caspases) (Fig. 2A), indicating that initiators have evolved early on to have a wider conformational space in the dimeric conformation. Consequently, the broader landscapes observed in Fig. 2A for initiators correspond to a less stable small subunit in the dimeric conformation. Helices 2 and 3 detach from the beta sheets, a feature seen in caspase-3 (Fig 2B & Fig S5), cFLIP, and other ancestors (Fig S5) but not exclusive to either family.

For the monomeric conformation in urea, there is no discernible difference in the FEL between initiators and effectors (Fig S6) since they unfold extensively. For simulations in water, however, the metastable states recovered from the last minima (Fig S7) as the system evolves over time reveal that beta sheet 6 unfolds in caspase-3 (Fig. 2D), but the monomeric conformation of caspase-8 is relatively stable (Fig 2E) which is observed to be an average trend in the effector and initiator tree (Fig S8), respectively.

Further, we used network analysis to quantify the extent of unfolding with degree centrality measure in the monomeric and dimeric conformations (Fig S9). In the dimeric conformation, the initiators lose more high-degree contacts in the presence of urea, whereas the effectors maintain them relatively well, which is exemplified by the upper limits of the tails displaying higher than ~20 degrees/contacts (FigS9A) whereas the opposite is observed in the monomeric conformation(FigS9B). The ancestors of effectors and initiators (Fig S9A&B) demonstrate a similar pattern as the extant sub-families, whereas the ancestor of all exhibits a mixture of characteristics.

Overall, the data from FEL and network analysis, suggest that the ancestral state had mixed characteristics, which later evolved into the initiator and effector families, with the effectors stabilizing the dimeric conformation over time, whereas the initiators being relatively less stable as a function of interactions in the small subunit. On the other hand, the monomeric conformation is less stable in effectors in comparison to the initiators as a function of interactions in the anti-parallel beta-sheet.

Highly conserved DC and BC residues provide a scaffold for the evolution of subfamilies and conformational dynamics

Residues (nodes) with more connections are more likely to be influential, but a high degree does not always indicate importance (17). A residue with many links to low-degree residues may have less influence than one with fewer ties to high-degree residues. However, nodes with high betweenness centrality are important because they connect other nodes that can dramatically alter the flow of information across the system (18).

We averaged the degree and betweenness centrality data for each node position in the monomeric and dimeric conformations separately for the entire tree for the simulations in water, and pair plots (FigS10 A&B) show that the majority of the distribution falls below 1000 for betweenness centrality (BC) and below 20 for degree centrality (DC). Since residues with high DC are important for stability (Fig1) and high BC are critical for the flow of information, we filtered the node positions that, on average, have high DC, and BC (outside the range discussed above), and mapped these node positions onto the monomeric (Figure 3A) and dimeric structures (figure 3B) of caspase-8. In both the monomeric and dimeric conformations, a common network of residues representing

interactions (blue spheres in Fig 3A&B) between the beta sheets and helices that makes high-degree contacts, maps to the hydrophobic core, and because these residues are tightly interconnected, they also have a high betweenness centrality. In contrast, some residues only exhibit a high DC, BC in the monomeric (magenta spheres in Fig 3A&B) or dimeric state (yellow spheres in Fig 3A&B). High DC BC residues exclusive to the monomeric conformation map predominantly to the rear of the structure (magenta spheres in Fig 3A) and stabilize the packing of the common network of residues, whereas in the dimeric conformation, these residues move out, and high DC BC residues exclusive to the dimeric conformation (yellow spheres in Fig 3B) map predominantly to the active site and the bottom loops, which move in and connect with the common core, altering the packing of the common network of residues and stabilizing it.

The high DC and BC residues were further classified (into 3 categories as mentioned earlier) based on the conservation score computed using Consurf. These residues were colored by conservation scores and mapped onto structures (Fig 3C&D) and networks (Fig 3E&F). Intriguingly, the residues that are unique to either the monomeric or dimeric conformation, mostly belong to the intermediate and variable conservation group (Figure S1). Variable and intermediately conserved residues cluster between helices 2 and 3 and the beta-sheets (Fig 3C&D) that unfold in urea in some caspases in the dimeric configuration (Figure S5); in the back of the molecule, these residues cluster at helices 4 and 5, (Fig 3C&D) which lose connections in initiators but retain them in effectors; and these residues have evolved different biophysical characteristics in each family(Figure S1).

Intriguingly, when urea is added to these systems, the degrees (y-axis) of some communities drop while their betweenness (x-axis) remains the same (FigS10 C,D). In Figure 3 E,F we have mapped the high DC BC residues in water and urea onto the network maps of monomeric and dimeric caspase-8, respectively, even when urea is added, this community of residues stay connected. Nodes with high betweenness values in transition states have been shown to be crucial for nucleus folding (19). Since the networks from urea simulations are averaged networks that represent metastable states, these nodes represent stable interactions in transition states, resembling the characteristics of a folding nucleus.

In summary, a highly conserved network of residues in the front and rear that hasn't altered since the common ancestor (red spheres in Fig 3C&D) serves as a scaffold for the evolution of initiator and effector families, as well as conformational dynamics, in all caspases. The network of high DC BC residues at the back of the molecule is more conserved than the network at the front (Fig 3C&D). The stability of the monomeric conformation can be affected by subtle alterations to the intermediately conserved high BC DC residues localized to the rear network (Fig 3A). These residues (magenta spheres in Fig 3A) are unique to the monomeric conformation and exhibit varied biophysical properties between families. High DC BC residues unique to the dimeric conformation are localized to the bottom loops and the catalytic site in the front face(yellow spheres in Fig 3B). Several variable and intermediately conserved residues cluster around helices 2 and 3 at the front of the molecule (cyan and orange spheres in Fig 3D) which can affect the packing of the high DC BC residues exclusive to the dimeric conformation.

Residues at the core that do not participate in core stabilization form electron tunnels for allosteric communication

Electron transfer channels in proteins are pathways that allow for the movement of electrons between different parts of the protein or between the protein and its environment (17). These channels are typically formed by specific amino acid residues that are strategically positioned within the protein structure to facilitate electron transfer. The most common amino acids that form electron transfer channels are redox amino acids like cysteine, methionine, tryptophan, tyrosine, phenylalanine and histidine. These amino acids have electron-rich groups that can participate in electron transfer reactions (18, 19).

We utilized the emaps webserver to identify electron or hole hopping channels in the active, dimeric, and monomeric structures available on PDB, including the crystal structures for the ancestor of effectors and the ancestor of caspase-6. (13, 17). Since CP-F112, CP-F147, and CP-C070 in the front on beta sheets 4,5 and 2, respectively, and CP-F037 on helix 1 (figure 4A and Fig S1) at the rear are highly conserved residues with redox potentials located in the core of the protein, they are of particular interest because they do not form many connections and have a low betweenness centrality. Consequently, these residues do not significantly contribute to the monomeric or dimeric stability of the core. In addition, numerous other conserved and redox potential residues CP-H062, CP-C102 in the front and CP-Y001, CP-F037 and CP-M003 at the back, line the bottom of the molecule (Figure 4A).

CP-F112 on beta-sheet 4, is a highly conserved central residue in the electron transport network, and this core hub has been retained for hundreds of millions of years (Fig 4A). CP-F112 undergoes a 180-degree rotation in the monomeric state which appears to correlate with large-scale active site loop movements. Laura Mario Pérez and colleagues have demonstrated how ring flipping at the core can cause large-scale conformational alterations without changing the fold (12). For instance, the ring flipping of CP-F112 in caspase-8 (Fig 4B) and caspase-3 (Fig 4C) can have an impact on the stacking of the aromatic residues in the upper half of the structure (Fig 4B). Since the highly conserved residues in the electron transport network are in the center and bottom loop regions(Fig 4A-C), each caspase has evolved a different signaling mechanism to communicate with the active site loops (Fig 4B-C). CP-F032 (rear) and CP-F068 (front) are highly conserved residues having high DC BC values, and they may participate in electron transport. When an electron is transferred to and received by an aromatic residue having a high DC BC, it can disrupt the pi-pi interactions, hydrophobic interactions, hydrogen bonding, and electrostatic interactions these residues make, thus affecting the packing of the molecule on a global scale, thus influencing the overall conformational dynamics and function in the family.

Discussion

Caspases are essential enzymes in the regulation of apoptosis, and inflammation has been extensively documented; however, the involvement of caspases in additional forms of cell death and non-cell death processes, such as differentiation and proliferation, remains poorly known(20). Understanding caspase conformational

landscapes, intermediate states, and mechanisms to fine-tune the conformational ensemble is crucial for comprehending regulation and, consequently, the fine-tuning of conformation to affect specific pathways (1). In this study, we used ancestral protein reconstruction and MD simulations to identify a conserved network of residues that provide a scaffold for the evolution of families and conformational dynamics, and with network analysis and FEL, we demonstrate how initiators and effectors stabilized the monomeric and dimeric folds and evolved conformational dynamics.

The folding funnel depicted in Figure 5A highlights our findings in conjunction with comprehensive experimental folding research conducted previously for initiator and effector caspase families (4, 7, 21). Network analysis for simulations in water and urea demonstrate that the active conformation is the most stable, followed by the dimeric and monomeric conformations (Fig1), which is illustrated as energy gradients for these conformations in the folding funnel (Fig 5A). The active conformation depicts the lowest minima in the folding funnel (Fig 5A) and indicates maximum stability, among all conformations. There is no experimental conformational free energy data for initiators in the dimeric conformation; however, our FEL analysis (Fig2 A,B) shows that the dimeric conformation is less stable in initiators than in effectors, resulting in a comparably lower minimum for initiators in the folding funnel (Fig 5A). Our results indicate that the relative instability is due to a less stable small subunit in initiators and these differential features evolved early in each lineage (Fig 2A,B). In the monomeric conformation, our data suggest that effectors are less stable than initiators due to a spontaneous loss of high-degree interactions between the anti-parallel beta-sheet and the core of the molecule, as observed in water simulations (Fig 2D,E), which has been represented by a higher

energy barrier for the initiators in the folding funnel model (Fig 5A). We can be sure that the high DC,BC residues (Fig 3E,F) that are highly conserved characterize the monomeric intermediate and represent the most stable interactions present ever since the evolution of the common ancestor.

The monomeric and dimeric conformations have the most conformational dynamics and are defined by a common network of residues (blue spheres in Fig 3A,B) that create above-average non-covalent interactions and are densely interconnected (high DC,BC). This common network of residues is highly conserved at the back of the molecule but less conserved at the front. The monomeric conformation is mainly stabilized by additional residues (high DC BC only in monomeric conformation) that stabilize the networks at the rear of the molecule (magenta spheres in Fig 3A, and orange in 3C). Interestingly, all the residues in high DC BC network are highly conserved in the rear network except for these additional residues that stabilize the monomeric fold. These residues have different biophysical properties in each subfamily, since high DC BC residues are the most critical to the network dynamics the initiators have evolved these residues to stabilize the rear network (stabilize the monomeric fold) whereas the effectors destabilized these residue interactions with the common network. This enables the effectors to remodel the network at the rear of the structure and access additional conformational space which may effectively lower the barrier to traverse to the dimeric state. Studies have revealed that 30 to 50 percent of the dimer stability originates from bottom interactions. We suggest that the packing of residues at the bottom influenced by the movement of the N-terminus can trigger the high DC,BC residues in the monomer to lose contact allowing for the common network to re-

organize and allow for the stabilization of active site loops. High DC,BC residues observed only in the dimer cluster around the front face near the active site loops and in the bottom which support our hypothesis. Overall, the network at the rear appear to evolve initiator and effector families whereas the less conserved front face appears to evolve conformational dynamics.

Further, we identified highly conserved redox potentials residues (Fig 5A) that line the common DC,BC residues in the monomeric and dimeric conformation that are implicated in electron transfer pathways at the front and back of the molecule. A network of charged residues line the bottom of the molecule and appear to act as charge sensors and relay the information to the electron transfer pathway. For example at the front face, electronic polarization state of several charged residues that are conserved at the bottom can be effected by stabilization of N-terminus and hence transmit this information by affecting the orientation of aromatic residues CP-F068, CP-F112 and CP-F147 of which Cp-F112 is the central player which can alter the aromatic stacking interactions, pi-pi interactions, hydrophobic interactions, and electrostatic interactions of residues in the high DC,BC cloud above it effecting the active site loops and hence activity.

Post-translational changes, such as the phosphorylation of CP-S104 in caspases (purple sphere in Fig. 5B), introduce a phosphate group (-PO3) to the hydroxyl group of serine, thereby forming a negatively charged residue. This disrupts the hydrogen bond that the serine was forming with the histidine CP-H062, which can convey the signal to CP-C070, thereby influencing the orientation of the central CP-F112. In addition, the electronic polarization state of multiple conserved charged residues at the bottom of the

structure may be altered, which can have a cascading impact on the electron clouds around CP-F068 and CP-F147 (Fig. 5B), forcing adjoining residues to reorient. Consequently, this phosphorylation signal can be amplified, and a coherent signal may deliver central CP-F112 with sufficient energy to rearrange the high DC, BC residues and affect the active site.

Materials and Methods

Homology modelling

For simulations in active configuration, we used mature caspases from the Protein Data Bank (PDB): caspases-3 (3dei), caspase-7 (1K86), caspase-6 (3NKF), caspase-8 (3kjq), AOE-1 (6PDQ), and pseudo enzyme cFLIP (3H11) (22–26). The rest of the ancestral enzymes were modelled based on the nearest available mature caspase structure in the evolutionary tree; for AOA-1,2,3, AOI-1. The ancestor of caspase-3/7 and caspase-6 were modeled based on the crystal structures of caspase-3 (3dei) and caspase-6 (3NKF), respectively. The ancestor of caspase 8/10 and caspase 10 were modeled based on the crystal structure of caspase 8(3KJQ), and the ancestor of cFLIP was modeled based on the crystal structure of cFLIP(3H11). Procaspase enzymes from the PDB, exhibiting inactive but dimeric structures, were utilized for the dimeric configuration: procaspase3 (4JQY), procaspase7(1K88), procaspase6(4N5D), and procaspase8 (6px9) (23, 26–29). Ancestor of caspase 3/7, AOE-1,2, and AOA-1,2,3 were modelled using procaspase3(4JQY) as template, while ancestor of caspase 6 was modelled using procaspase6(4N5D) as template. We used the procaspase8(6PX9) structure as a template for the entire initiator tree from AOI-1,2 to the extant caspases.

For the monomeric configuration, we modeled the complete tree using the procaspase 8(2k7z) structure, the only NMR solution structure available (30).

Ancestral protein reconstruction (APR)

To resurrect a highly probable sequence of the last common ancestral caspase of the chordates involved the extrinsic pathway of apoptosis, we utilized a database of curated caspase sequences from CaspBase (Grinshpon et al., 2018) that provided sequences from the initiator (caspase-8/-10/cFlip) and the effector (caspase-3/-6/-7) subfamilies in the chordate lineage. A total of 600 sequences were obtained to generate 3 databases comprising of 200 sequences each for APR. Representative taxa from various classes of Chordata (mammals, birds, fish, amphibians and reptiles) were chosen in each database to resurrect three probable ancestral sequences (AOA1,AOA2 and AOA3). Since the prodomain is subject to high sequence variation due to recombination, insertions, and deletions, we pruned the sequences on Jalview to remove the prodomains after PROMALS3D structure-based alignment. Ancestral protein reconstruction was carried out as previously described by Grinshpon et al.

Network analysis of amino acid interactions

To analyze the ancestral caspase networks, we utilized the open-source software Cytoscape(Shannon et al., 2003). Through SenseNet (Schneider & Antes, 2022) a Cytoscape plugin we can visualize and allocate measures of importance to amino acids by converting MD interaction timelines into protein structure networks. The networks were analyzed for degree and betweenness centrality for MD simulations performed in water and 8M urea. Simulations were imported onto SenseNet, and input parameters

were modified to examine Van der Waals, hydrophobic, electrostatic interactions with a distance cut-off of 4 angstroms. The interaction weights were set to sum and the average, in order to generate network interactions displaying the average degree centrality (DC) and betweenness centrality (BC) values for each residue derived from the entire simulation. Nodes displaying high DC and BC were further classified according to the conservation scores obtained from Consurf.

MD simulations and free energy landscape

MD simulations for 200ns were performed in water and 8M urea for all caspases (in Fig S1) in the active, monomeric and dimeric conformations as previously described. To study the concerted motions of caspases, and identify the most significant motions in the simulations, principal component analysis (PCA) was conducted for all protein atoms in the trajectory. The principal components (PCs) obtained from MD simulations in water and urea are essentially the eigenvector values from the covariance matrix, each corresponding to a change in the trajectory. The eigenvalues and eigenvectors were analyzed using the gmx anaeig tool and the principal components (PCs) with the largest motions were selected and plotted for comparison. These PCs provide the main information about the spread of datapoints in the conformational space, indicating the protein's global motion during simulations. To investigate the free energy landscape (FEL), the gmx sham tool was employed to combine the reaction coordinates of the PCs with the most significant movements. The FEL plots were generated using Matlab.

References

1.   A. C. Clark, Caspase allostery and conformational selection. *Chem Rev* 116, 6666–6706 (2016).

2.   S. Shalini, L. Dorstyn, S. Dawar, S. Kumar, Old, new and emerging functions of caspases (2015) https:/doi.org/10.1038/cdd.2014.216 (February 4, 2023).

3.   D. S. Chelur, M. Chalfie, Targeted cell killing by reconstituted caspases. *Proc Natl Acad Sci U S A* 104, 2283–2288 (2007).

4.   S. Shrestha, A. C. Clark, Evolution of the folding landscape of effector caspases. *Journal of Biological Chemistry* 297, 1–12 (2021).

5.   S. Kesavardhana, R. K. Subbarao Malireddi, T.-D. Kanneganti, Annual Review of Immunology Caspases in Cell Death, Inflammation, and Pyroptosis (2020) https:/doi.org/10.1146/annurev-immunol-073119 (February 4, 2023).

6.   D. R. McIlwain, T. Berger, T. W. Mak, Caspase Functions in Cell Death and Disease. *Cold Spring Harb Perspect Biol* 5, 1–28 (2013).

7.   L. Yao, A. C. Clark, Comparing the folding landscapes of evolutionarily divergent procaspase-3. *Biosci Rep* 42, 1–13 (2022).

8.   L. Aravind, E. v Koonin, Classification of the Caspase-Hemoglobinase Fold: Detection of New Families and Implications for the Origin of the Eukaryotic Separins (2002) https:/doi.org/10.1002/prot.10060 (November 29, 2022).

9.   L. Q. Yang, *et al.*, Protein dynamics and motions in relation to their functions: Several case studies and the underlying mechanisms. *J Biomol Struct Dyn* 32, 372–393 (2014).

10.   P. E. Wright, H. Jane Dyson, Intrinsically disordered proteins in cellular signalling and regulation (2015) https:/doi.org/10.1038/nrm3920 (February 4, 2023).

11.   T. E. Saldaño, A. M. Monzon, G. Parisi, S. Fernandez-Alberti, Evolutionary Conserved Positions Define Protein Conformational Diversity. *PLoS Comput Biol* 12, 1004775 (2016).

12.   L. Mariño Pérez, *et al.*, Visualizing protein breathing motions associated with aromatic ring flipping Protein dynamics induced by a tyrosine residue. *Nature* | 602, 695 (2022).

13.   R. D. Grinshpon, *et al.*, Resurrection of ancestral effector caspases identifies novel networks for evolution of substrate specificity. *Biochem J* 476, 3475–3492 (2019).

14.   G. K. A. Hochberg, J. W. Thornton, Reconstructing Ancient Proteins to Understand the Causes of Structure and Function. *https://doi.org/10.1146/annurev-biophys-070816-033631* 46, 247–269 (2017).

15.   A. Venkat, *et al.*, Modularity of the hydrophobic core and evolution of functional diversity in fold A glycosyltransferases. *Journal of Biological Chemistry* 298 (2022).

16.   S. Bromberg, K. A. Dill, Side-chain entropy and packing in proteins. *Protein Science* 3, 997–1009 (1994).

17.   R. N. Tazhigulov, J. R. Gayvert, M. Wei, K. B. Bravaya, eMap: A Web Application for Identifying and Visualizing Electron or Hole Hopping Pathways in Proteins (2019) https:/doi.org/10.1021/acs.jpcb.9b04816 (January 27, 2023).

18.   C. Wittekindt, M. Schwarz, T. Friedrich, T. Koslowski, Aromatic Amino Acids as Stepping Stones in Charge Transfer in Respiratory Complex I: An Unusual Mechanism Deduced from Atomistic Theory and Bioinformatics https:/doi.org/10.1021/ja900352t (February 5, 2023).

19.   M. Wang, J. Gao, P. Müller, B. Giese, Electron Transfer Electron Transfer in Peptides with Cysteine and Methionine as Relay Amino Acids** https:/doi.org/10.1002/anie.200900827 (February 5, 2023).

20.   S. Shalini, L. Dorstyn, S. Dawar, S. Kumar, Old, new and emerging functions of caspases (2015) https:/doi.org/10.1038/cdd.2014.216 (February 13, 2023).

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

22.   J. Q. Du, *et al.*, Isoquinoline-1,3,4-trione derivatives inactivate caspase-3 by generation of reactive oxygen species. *J Biol Chem* 283, 30205–30215 (2008).

23.   J. Chai, *et al.*, Crystal structure of a procaspase-7 zymogen: mechanisms of activation and substrate binding. *Cell* 107, 399–407 (2001).

24.   S. Vaidya, E. M. Velázquez-Delgado, G. Abbruzzese, J. A. Hardy, Substrate-induced conformational changes occur in all cleaved forms of caspase-6. *J Mol Biol* 406, 75–91 (2011).

25.   Z. Wang, *et al.*, Kinetic and structural characterization of caspase-3 and caspase-8 inhibition by a novel class of irreversible inhibitors. *Biochim Biophys Acta Proteins Proteom* 1804, 1817–1831 (2010).

26.   J. W. Yu, P. D. Jeffrey, Y. Shi, Mechanism of procaspase-8 activation by c-FLIPL. *Proc Natl Acad Sci U S A* 106, 8169–8174 (2009).

27.   N. D. Thomsen, J. T. Koerber, J. A. Wells, Structural snapshots reveal distinct mechanisms of procaspase-3 and-7 activation https:/doi.org/10.1073/pnas.1306759110 (February 5, 2023).

28.   J. Murray, *et al.*, Tailoring small molecules for an allosteric site on procaspase-6. *ChemMedChem* 9, 73–77 (2014).

29.   J. H. Xu, *et al.*, Integrative X-ray Structure and Molecular Modeling for the Rationalization of Procaspase-8 Inhibitor Potency and Selectivity. *ACS Chem Biol* 15, 575–586 (2020).

30.   N. Keller, J. Mareš, O. Zerbe, M. G. Grütter, Structural and Biochemical Studies on Procaspase-8: New Insights on Initiator Caspase Activation. *Structure* 17, 438–448 (2009).

FIGURES

Figure 1.
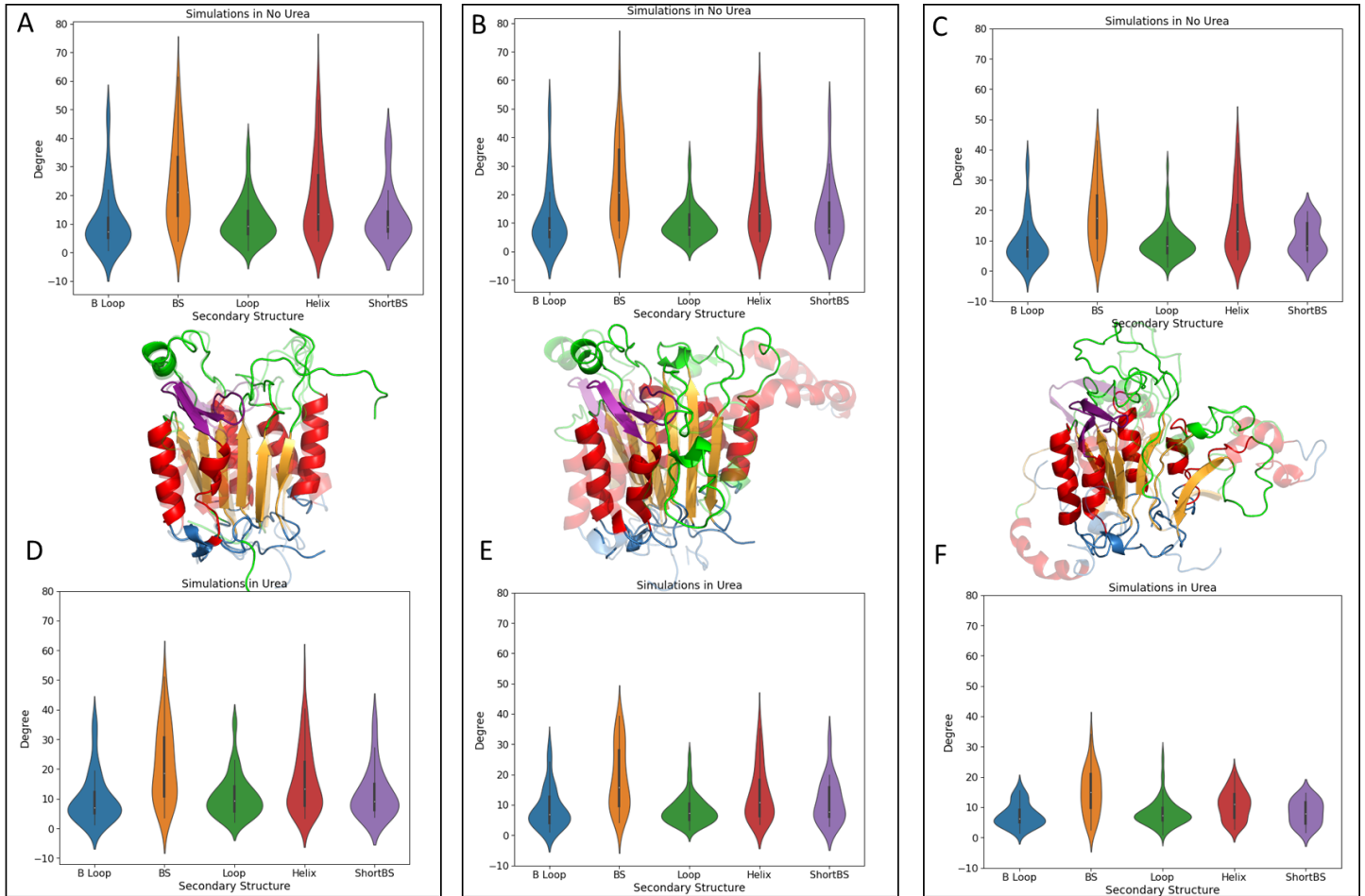
Figure 1. Violin plots representing the average degrees/contacts for the (A)active, (B)dimeric and (C)monomeric conformations in water and (D)active, (E)dimeric (F) monomeric conformations in 8M urea. Bottom loops (blue), beta-sheets (orange), catalytic site loops (green), helices (red) and the short beta-sheets (purple) are color-coded on the modeled caspase-8 structures. The solid structures represent the caspase-8 in water, and the translucent structures represent caspase-8 in 8M urea.

Figure 2.
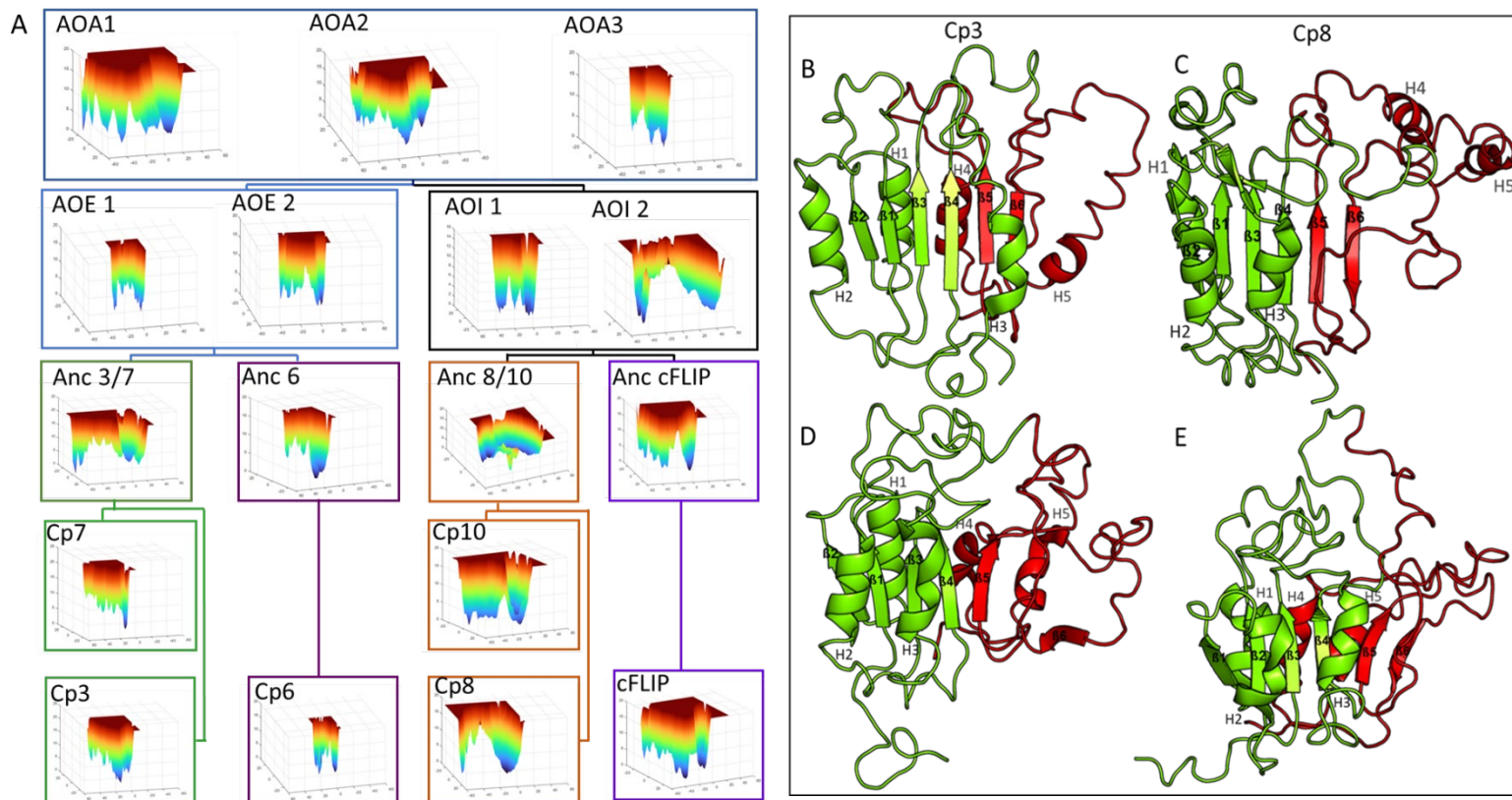
Figure 2. The free energy landscapes of the dimeric conformation of caspases in urea obtained from 200 ns MD simulations. (A) The FELs are generated as a function of projections of the MD trajectory onto the first (PC1) and the second (PC2) eigenvectors, respectively. Observed FEL minimas circled on caspase-3 and caspase-8 in (A) are represented as the metastable states visited during the simulations in (B) caspase-3 and (C) caspase-8. Observed minimas in the monomeric conformation for simulations in water are represented as the metastable states in (D) caspase-3 and (E) caspase-8

Figure 3.
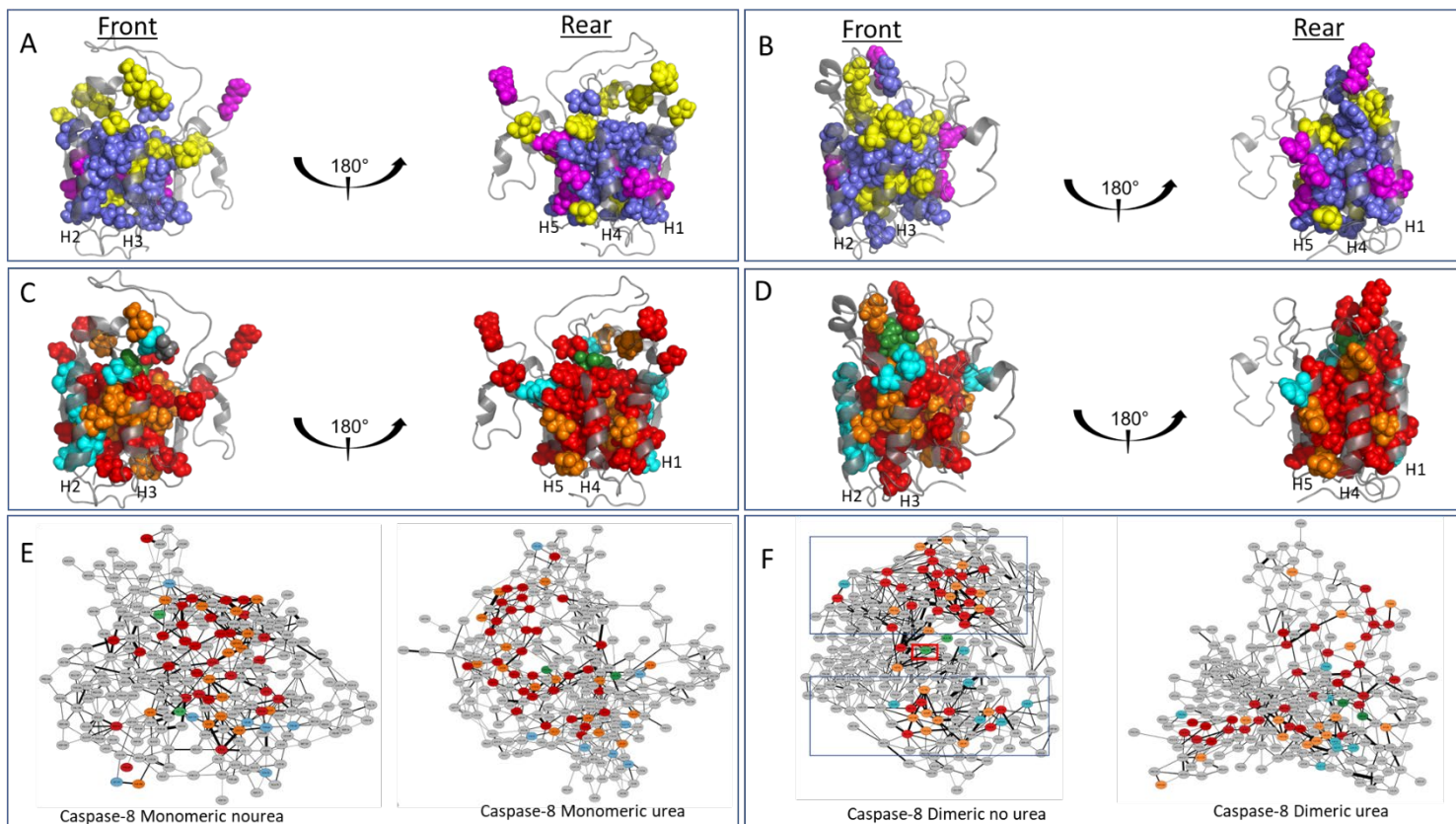
Figure 3. High DC and BC residues classified according to conservation in the monomeric and dimeric. Blue spheres represent the conserved network of high DC, BC residues; magenta spheres indicate high BC, DC residues unique to the monomeric, whereas yellow spheres indicate high BC, DC residues unique to the dimeric conformation that are mapped on the (A) monomeric conformation of caspase-8 and (B) dimeric conformation of caspase-8. High DC, BC residues classified as highly conserved (red spheres), intermediately conserved (orange spheres) and variable (cyan spheres) mapped on the (C) monomeric conformation of caspase-8 and (D) dimeric conformation of caspase-8. A 2D representation of the amino acid interaction networks in water and urea for the (E) monomeric and (F) dimeric conformations color-coded according to conservation scores.

Figure 4.
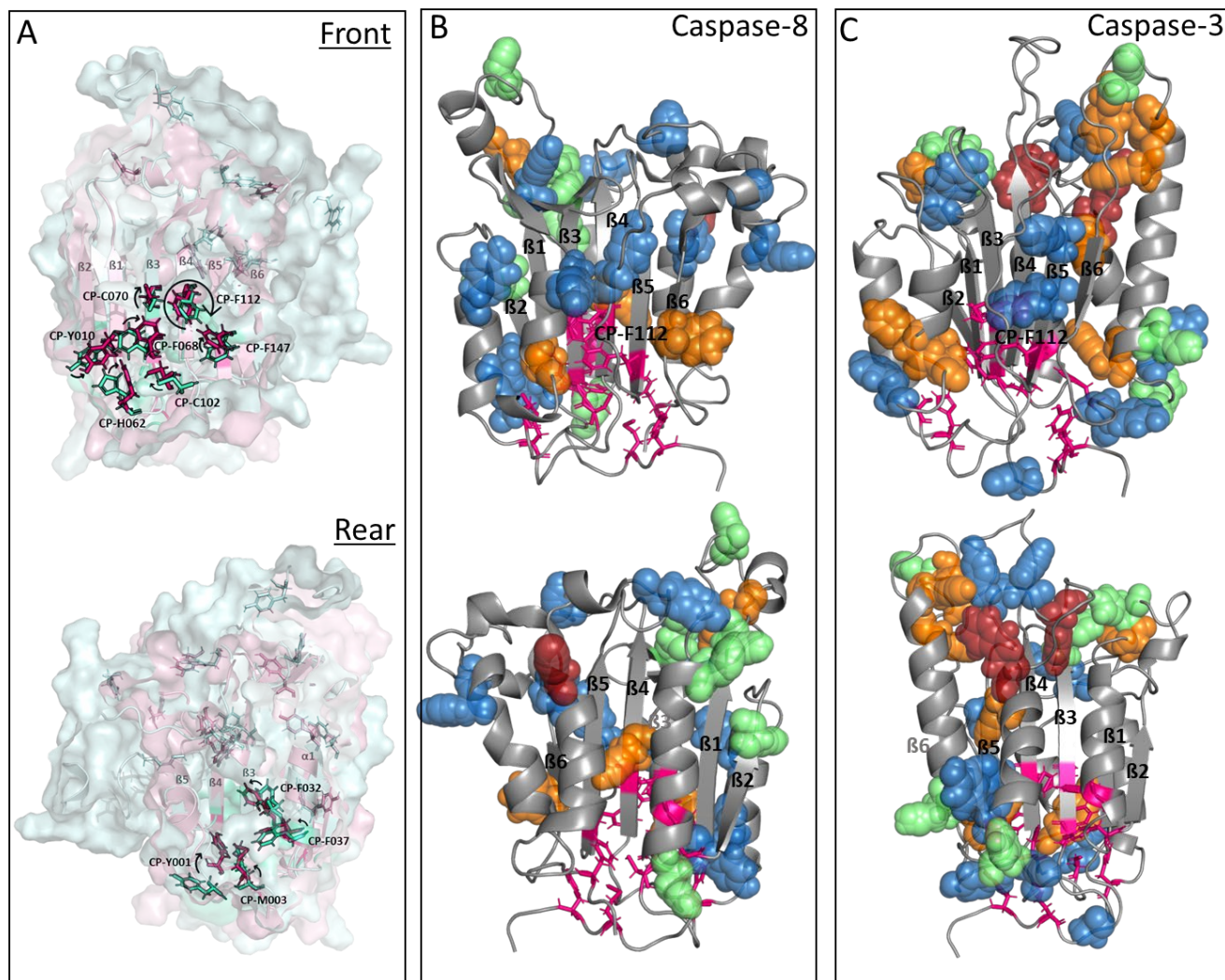
Figure 4. Conserved electron transport network in the caspase family. (A) A conserved network mapped onto the monomeric (pale cyan) and the dimeric (light pink) caspase-8 structures. All the residues in the conserved network are labeled according to the CP system and are depicted as sticks for the monomer (aquamarine sticks) and the dimer (pink sticks). The arrows indicate the relative motions of the residues as the protein remodels this network to transition from the monomer toward the dimer. Electron transport networks in the (B) dimeric conformation of caspase-8 and (C) the dimeric conformation of caspase-3 with the conserved network across families (pink sticks), and tyrosine (blue spheres), phenylalanine (orange spheres), histidine (green spheres) and tryptophan (red spheres) residues.

Figure 5.
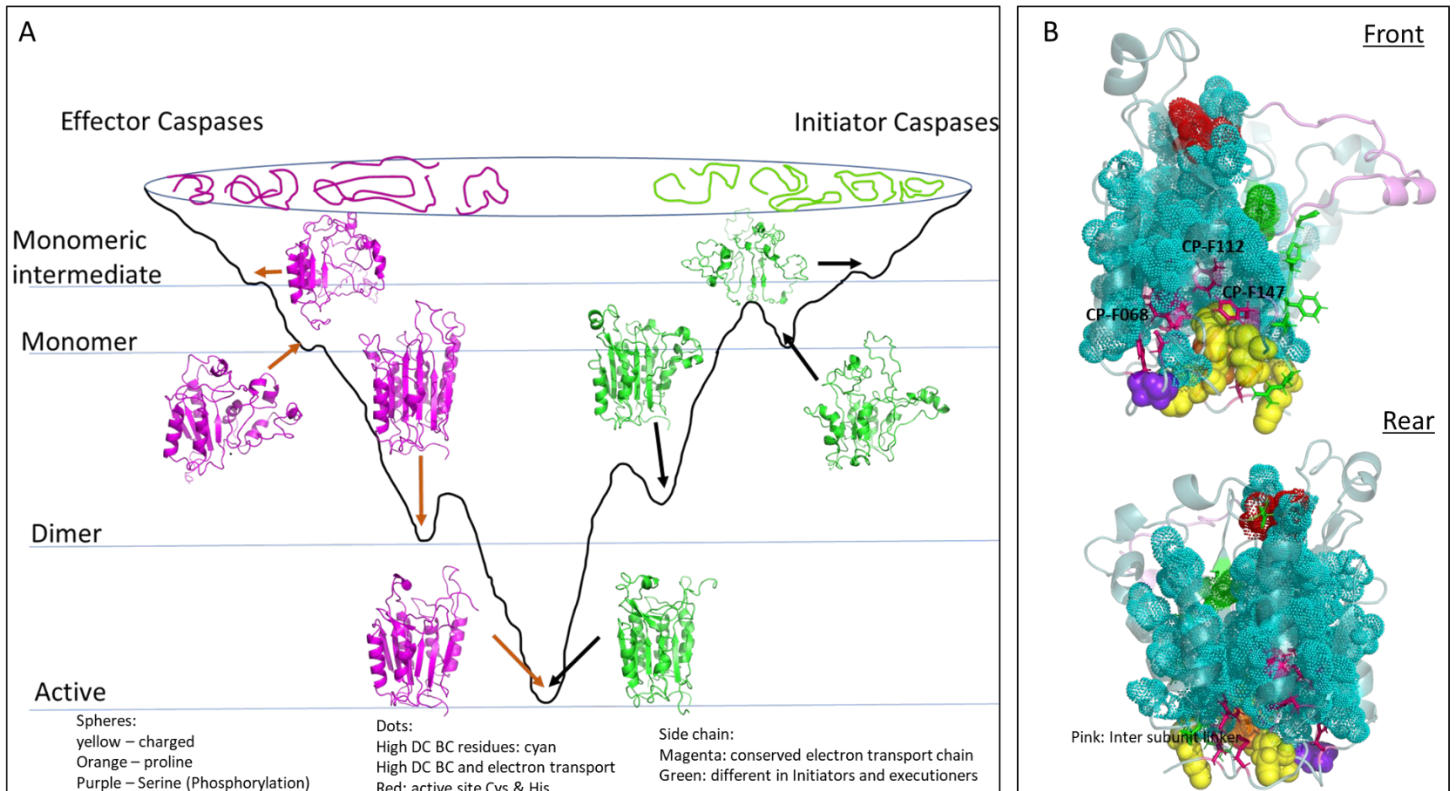
Figure 5. A folding funnel model for the conformational stability of caspases.

Supplementary information

Exploring the Conformational Landscape and Allosteric Networks of Caspases through Free Energy and Network Analysis

Isha Joglekar[1‡], Mithun Nag Karadi Giridhar [1‡], David A. Diaz[1], Ankit Deo[2] and A. Clay Clark[1*]

[1]Department of Biology, University of Texas at Arlington, Arlington, Texas, 76019

[2]Department of Information Systems, University of Texas at Arlington, Arlington, Texas, 76019

Running title: Evolutionary analysis of the conformational landscape of caspases

[*]Corresponding author: A. Clay Clark

Email: clay.clark@uta.edu

Figure S1.
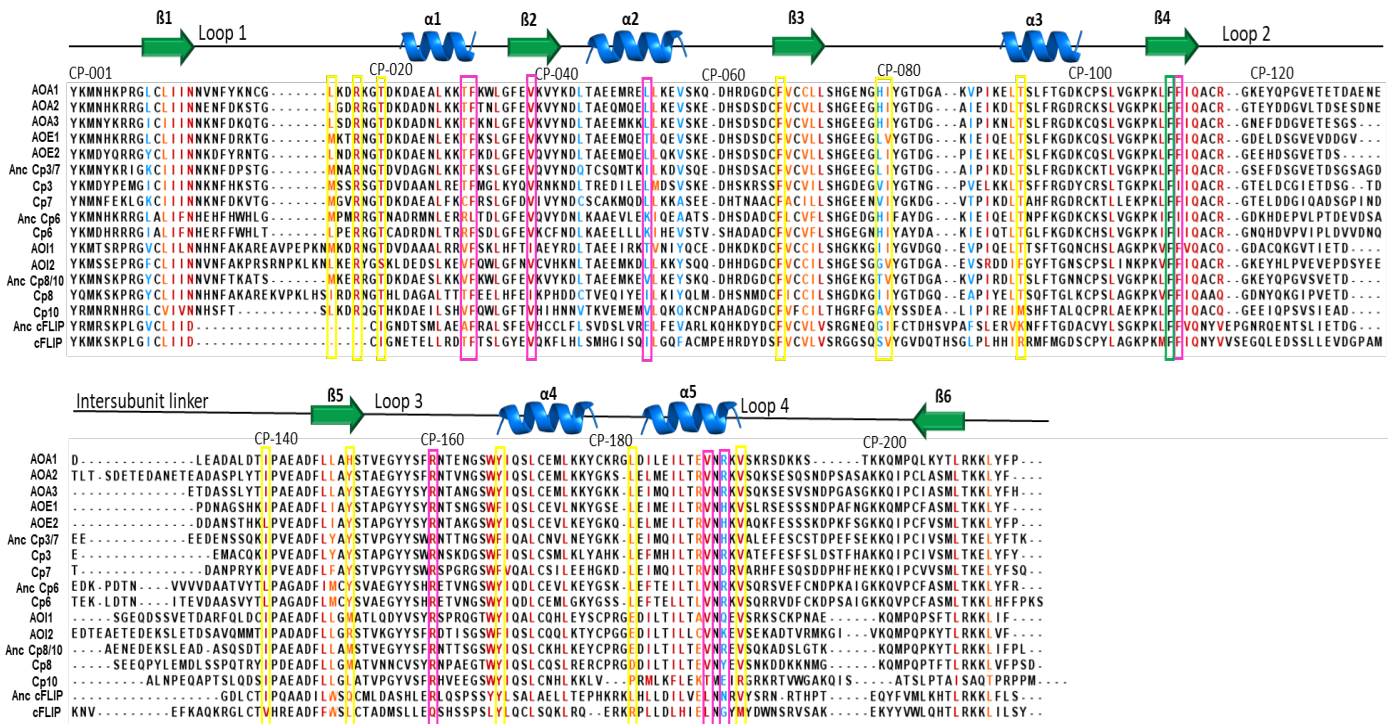
Figure S1. Multiple sequence alignment showing the secondary structural elements (loops, beta sheets, alpha helices) along with the common position numbers among the caspases. The colored residues represent the amino acids that have high DC, BC and are classified as conserved residues (red), intermediately conserved residues (orange) and variable residues (cyan) based on the conservation scores across the entire family. Residues exhibiting high DC, BC exclusively in the monomeric conformation are highlighted by yellow boxes, whereas those unique to the dimeric conformation are highlighted in magenta **boxes.**
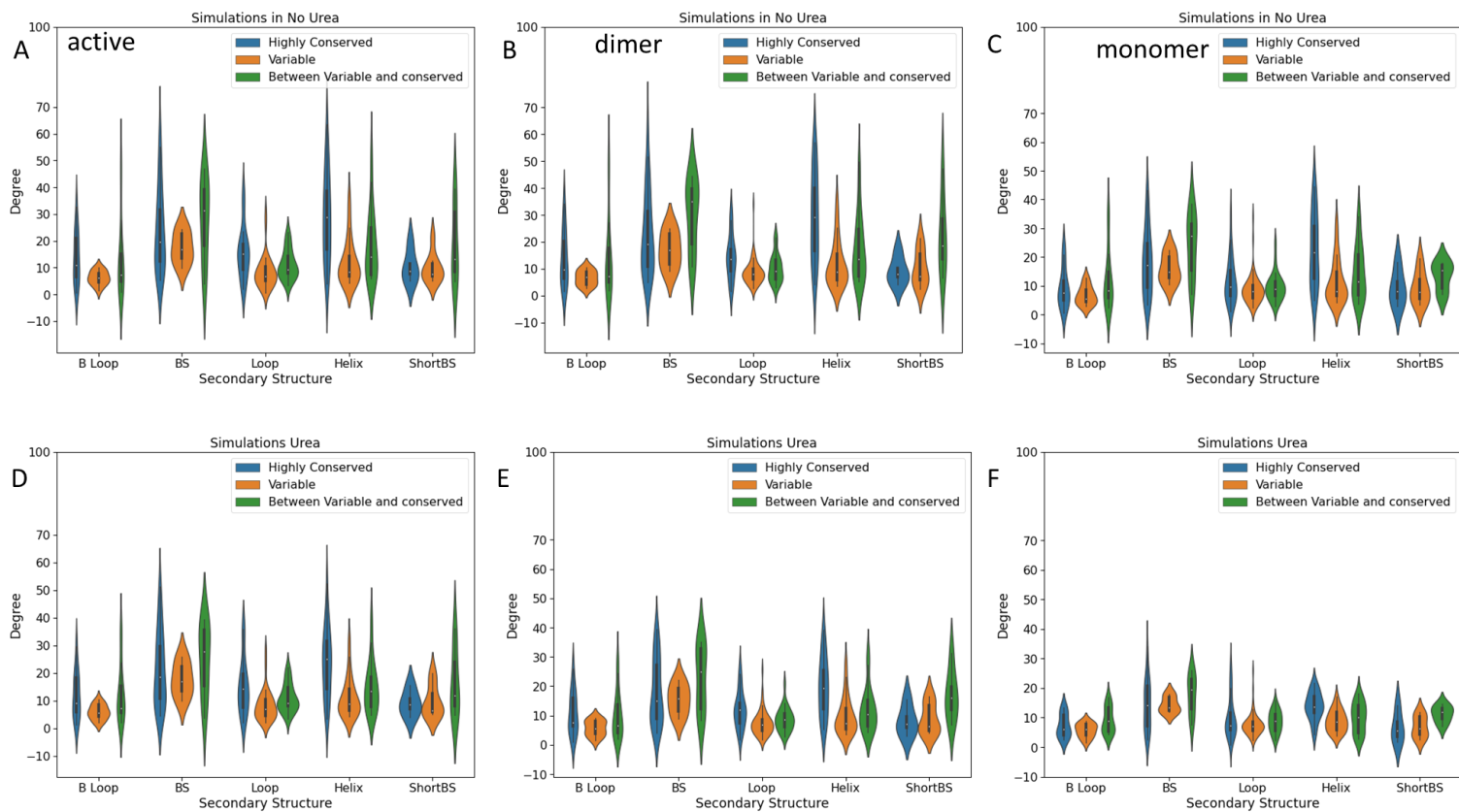
Figure S2. Violin plots representing the average degrees/contacts categorized as conserved (blue), intermediately conserved (green), and variable (orange) based on the conservation scores and grouped according to secondary structural elements for the (A)active, (B)dimeric and (C)monomeric conformations in water and (D)active, (E)dimeric (F) monomeric conformations in 8M urea.

Figure S3.
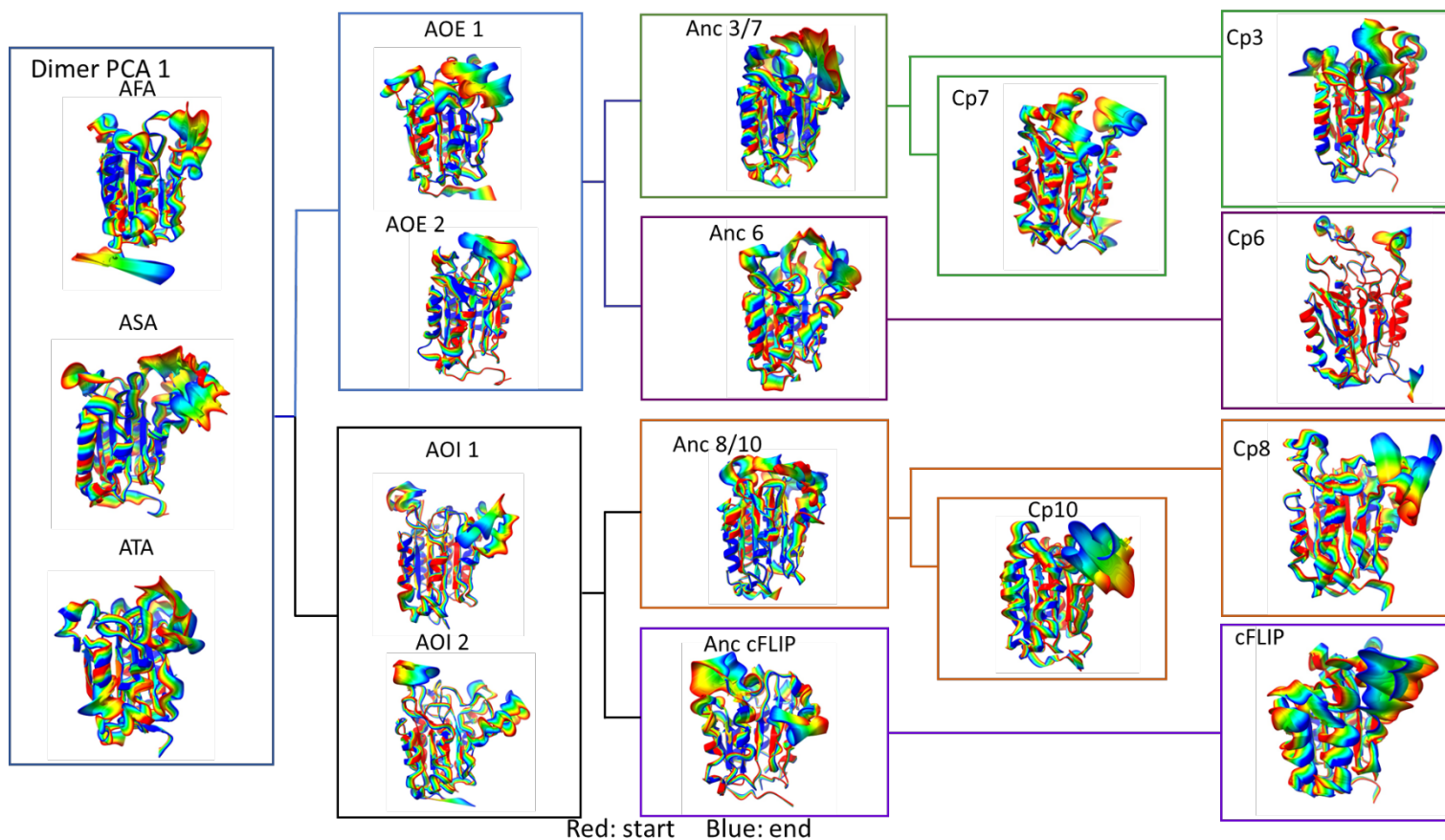
Figure S3. Large, concerted motions of the caspase family in the dimeric conformation in water described by eigenvector 1 (PC1). The structural displacements between the two extremes extracted from the projection of the trajectory onto the first PC are colored from blue to red to highlight the differences.

Figure S4.
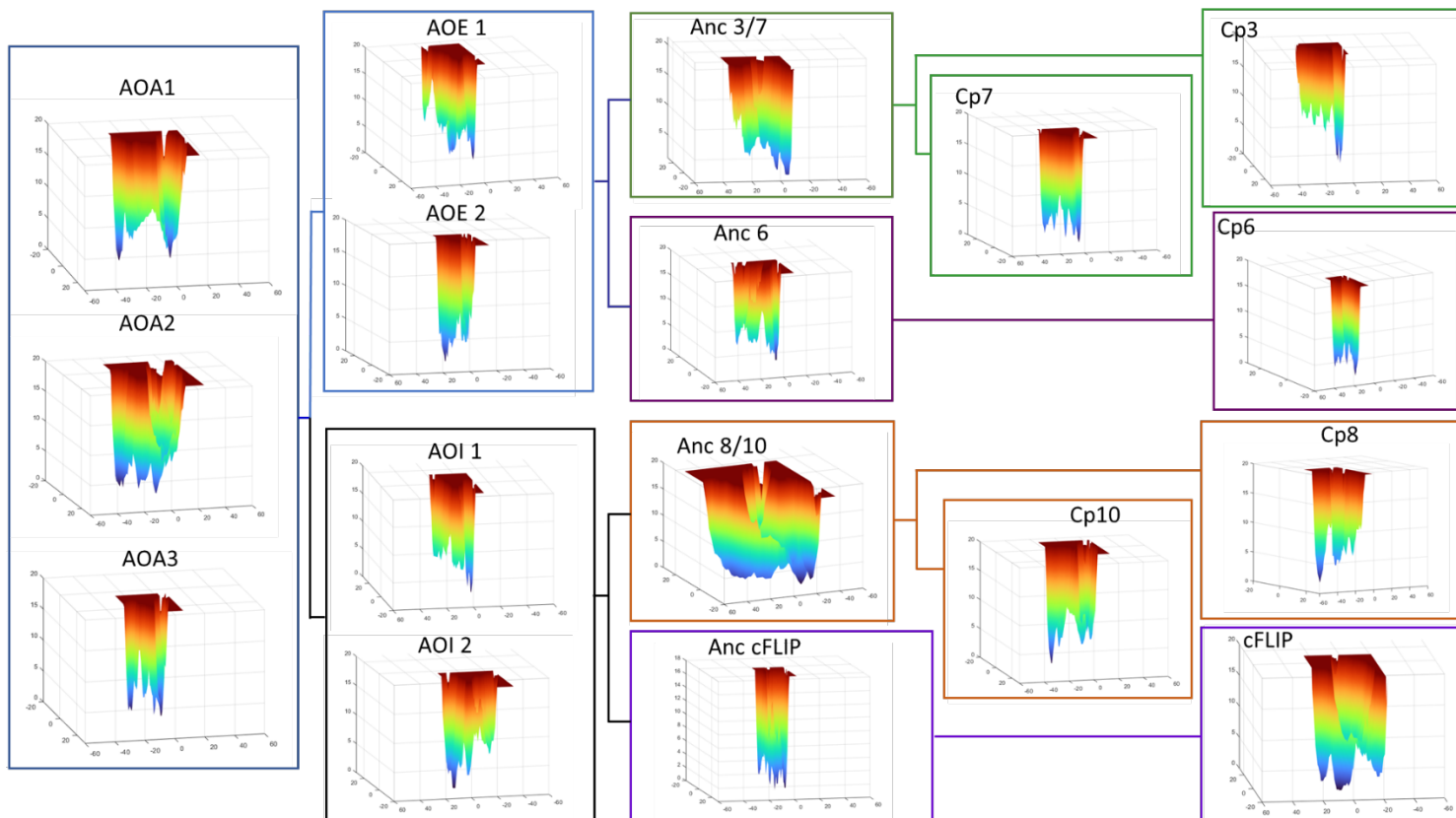
Figure S4. The free energy landscapes of the dimeric conformation of caspases in water obtained from 200 ns MD simulations. The FELs are generated as a function of projections of the MD trajectory onto the first (PC1) and the second (PC2) eigenvectors, respectively. Observed minimas represent the metastable states visited during the simulations.

Figure S5.
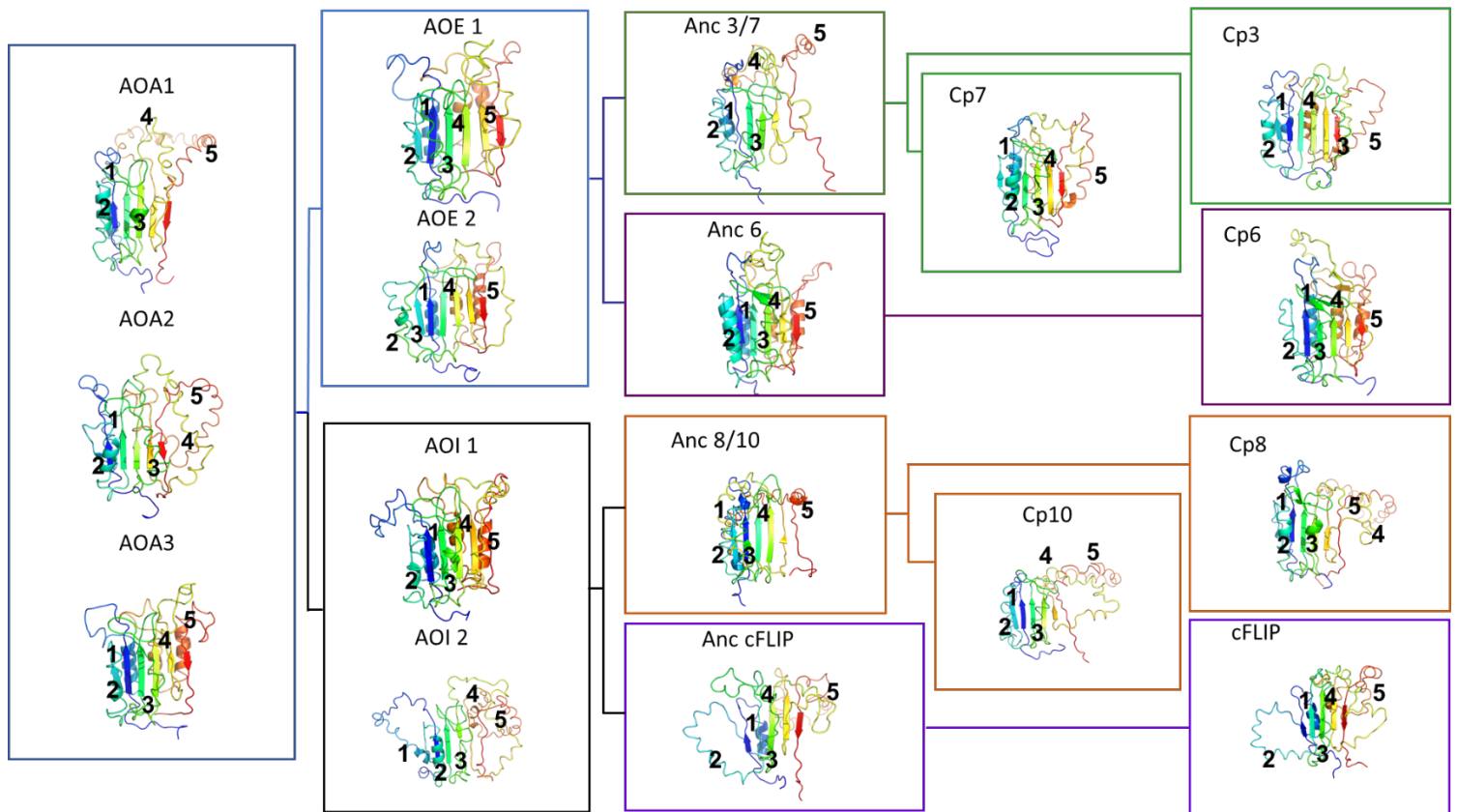
Figure S5. Metastable states of caspases in the dimeric conformation in urea indicating unstable regions.

Figure S6.
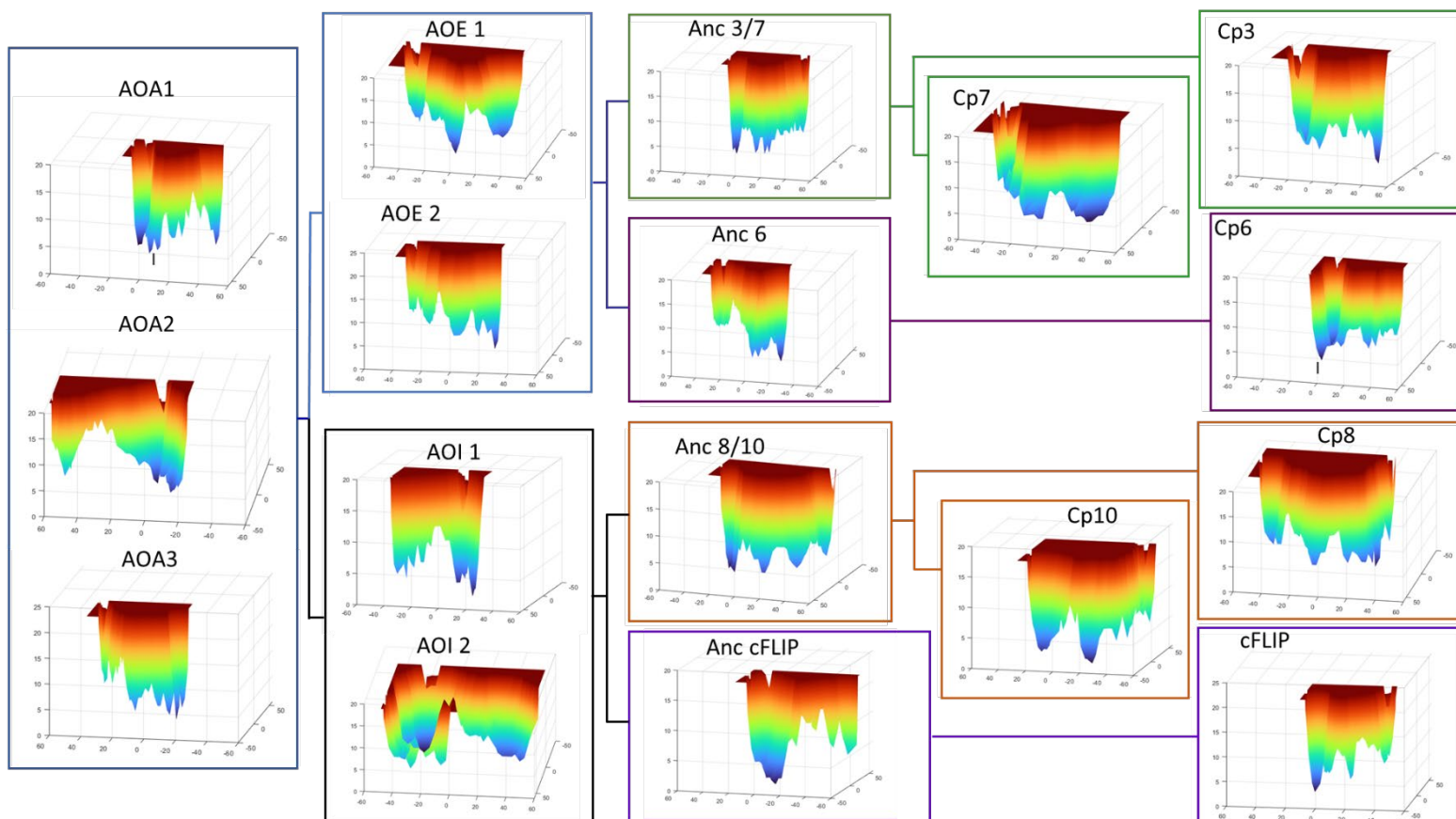
Figure S6. The free energy landscapes of the monomeric conformation of caspases in urea obtained from 200 ns MD simulations. The FELs are generated as a function of projections of the MD trajectory onto the first (PC1) and the second (PC2) eigenvectors, respectively. Observed minimas represent the metastable states visited during the simulations.

Figure S7.
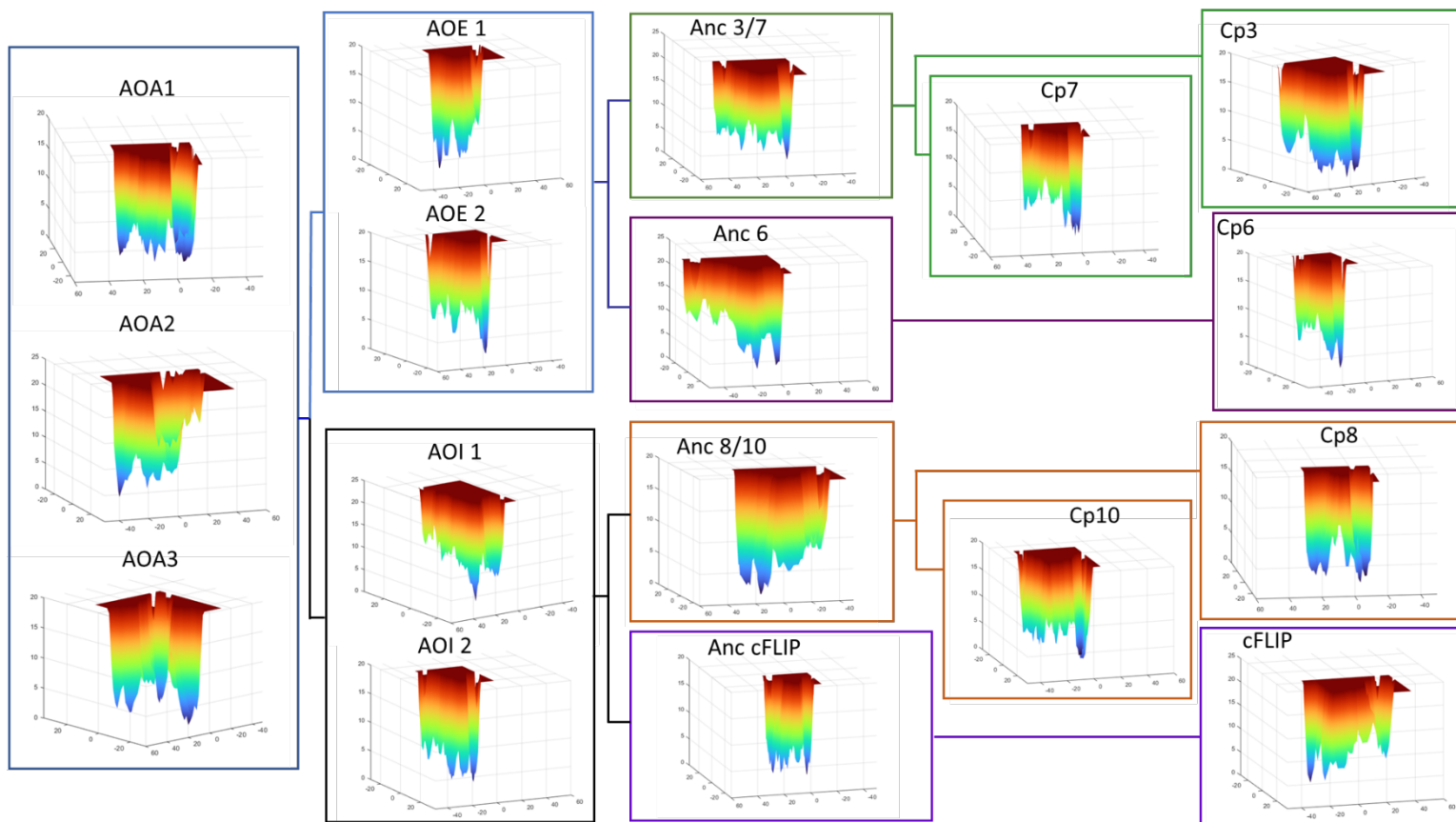
Figure S7. The free energy landscapes of the monomeric conformation of caspases in water obtained from 200 ns MD simulations. The FELs are generated as a function of projections of the MD trajectory onto the first (PC1) and the second (PC2) eigenvectors, respectively. Observed minimas represent the metastable states visited during the simulations.

Figure S8.
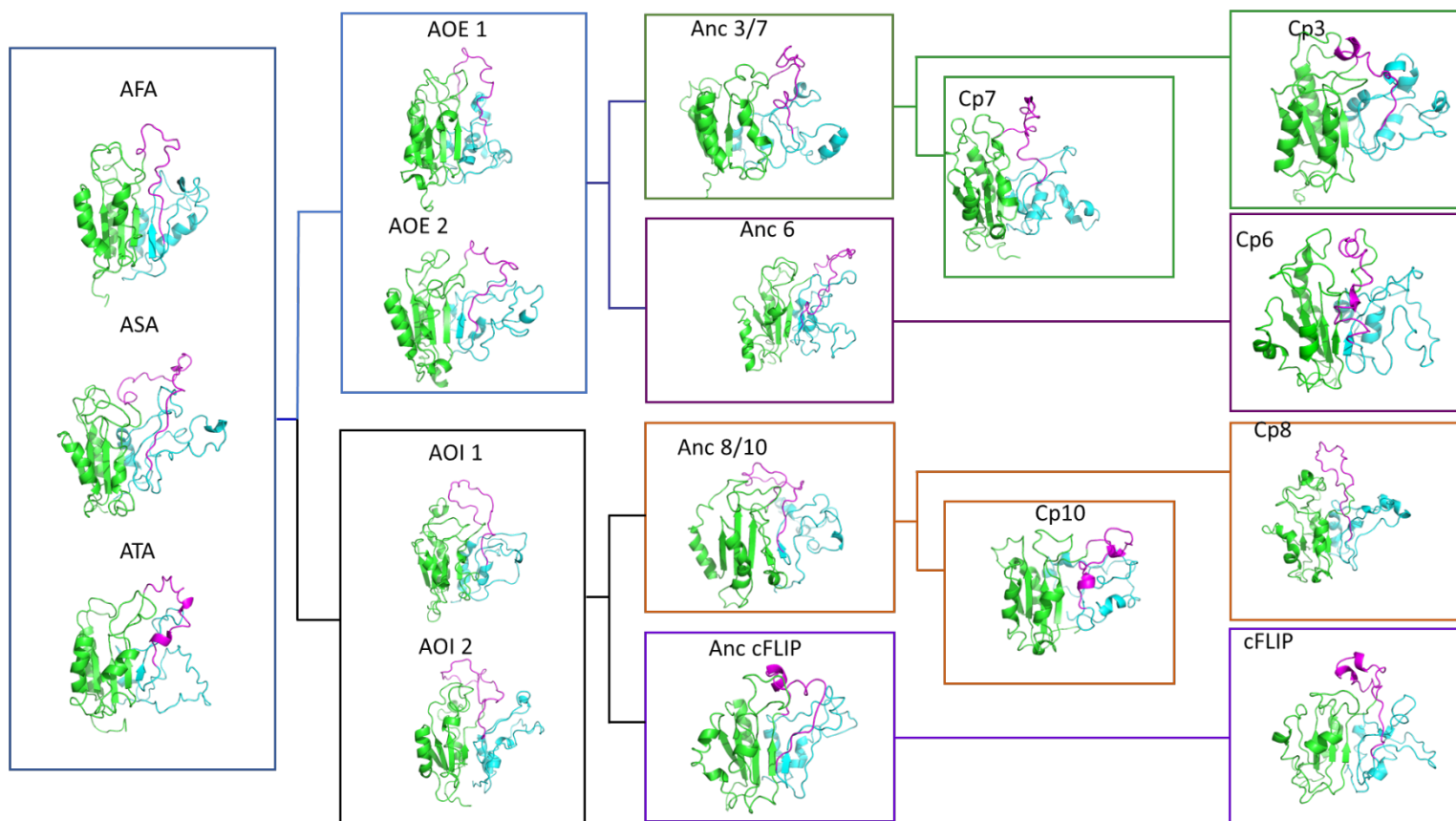
Figure S8. Metastable states extracted from the last observed minima in the FEL of the dimeric conformation in water

Figure S9.
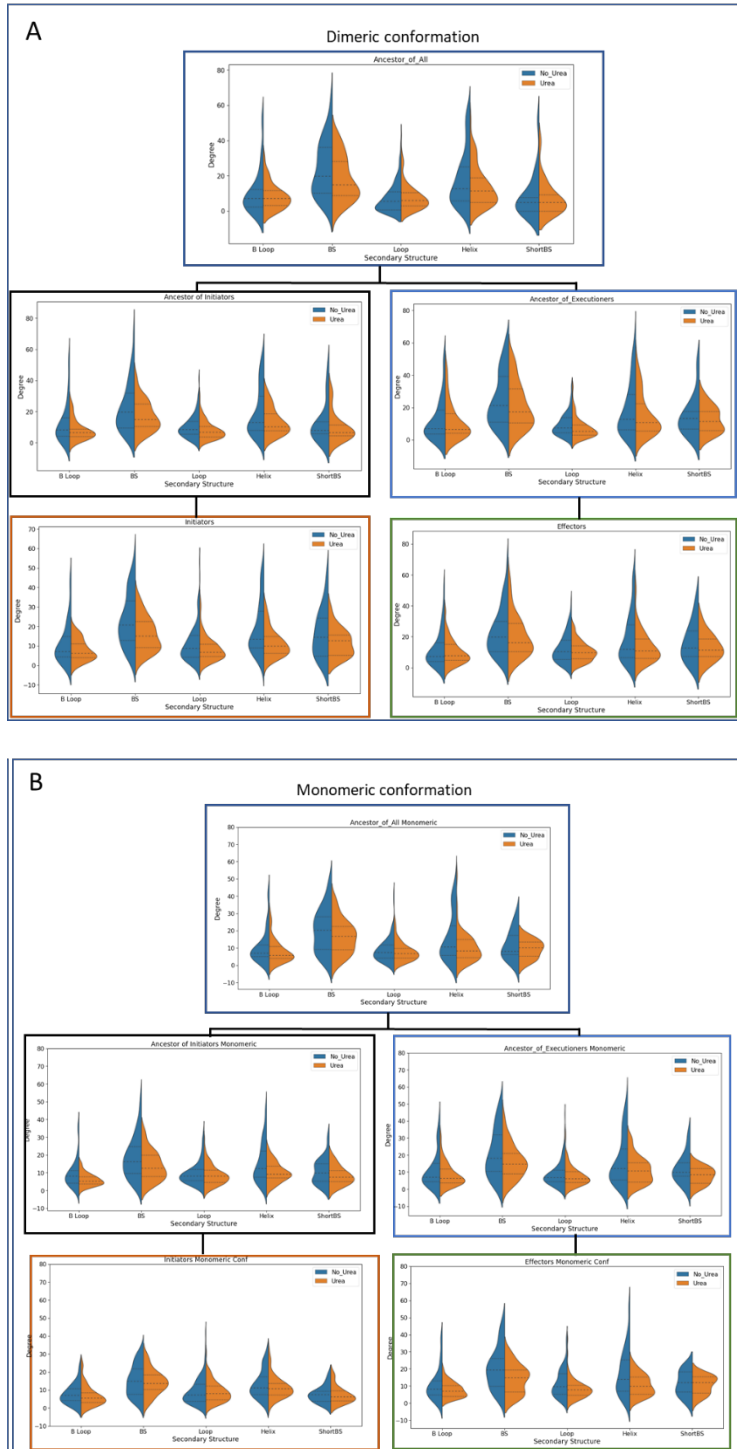
Figure S9. Violin plots illustrating the averaged high-degree contacts across the following groups: ancestor of all (AOA1/2/3), ancestor of initiators (AOI1/2), ancestor of effectors (AOE1/2), effector caspases (caspase -3/-6/-7) and initiator caspases (caspase -8/10/cFlip) in the (A) dimeric conformation in water (blue), urea (orange) and the (B) monomeric conformation in water (blue), urea (orange).

Figure S10.
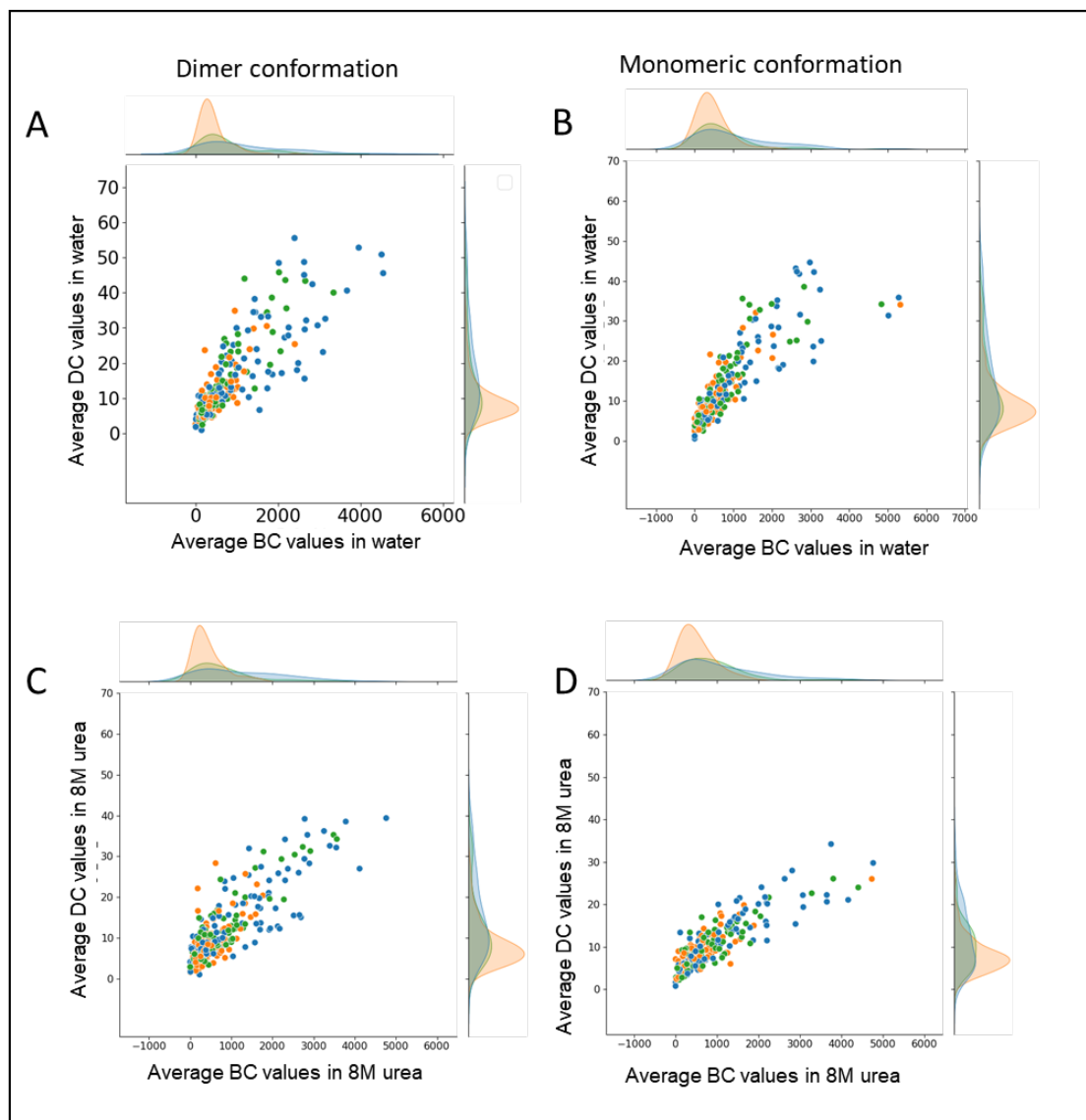
Figure S10. Pair plots displaying the average degree centrality on y-axis and average betweenness centrality on x-axis for dimeric structures in (A) water, (C) 8M urea, and the monomeric structures in (B) water, (D) 8M urea

Supplementary Fig. Legend

Supplementary Tables S1, **S2, and S3** Display the input sequence composition for ancestral reconstruction database of AOA1, AOA2, and AOA3.

Supplementary Figure S1. Displays the sequence composition of the CaspBase database, and from the sequences available in which phylum they fall under.

Supplementary Table 1

Table S1. displays the accession number of all caspases used in the database reconstruction for AOA1

AOA1				
Accession ID	Caspase	Species	Class	Order
NP_004337.2	Caspase-3 Isoform A Preprotein	Homo Sapiens	Mammalia	Primates
NP_0012713 38.1	Caspase-3	Mus Musculus	Mammalia	Rodentia
NP_571952.1	Caspase-3 apoptosis-related cysteine peptidase a	Danio Rerio	Fish	Cypriniformes
XP_0015171 22.2	Predicted: Caspase-3	Ornithorhynchus Anatinus	Mammalia	Monotremata

NP_0010030 42.1	Caspase-3	Canis Lupus Familiaris	Mammalia	Carnivora
NP_0011574 33.1	Caspase-3	Equus Caballus	Mammalia	Perissodactyla
NP_990056.1	Caspase-3	Gallus Gallus	Birds	Galliformes
NP_999296.1	Caspase-3	Sus Scrofa	Mammalia	Artiodactyla
NP_0010755 86.1	Caspase-3	Oryctolagus Cuniculus	Mammalia	Lagomorpha
NP_0011209 00.1	Caspase-3	Xenopus Tropicalis	Amphibia	Anura
XP_0113623 27.1	Predicted: Caspase-3	Pteropus Vampyrus	Mammalia	Chiroptera
XP_0037730 79.1	Predicted: Caspase-3	Sarcophilus Harrisii	Mammalia	Dasyuromorphia
XP_0044818 06.1	Predicted: Caspase-3	Dasypus Novemcinctus	Mammalia	Cingulata
XP_0061285 58.1	Predicted: Caspase-3 isoform X1	Pelodiscus Sinensis	Reptilia	Testudines
XP_0051492 03.1	Predicted: Caspase-3	Melopsittacus Undulatus	Birds	Psittaciformes
XP_0052820 31.1	Predicted: Caspase-	Chrysemys Picta Bellii	Reptilia	Testudines

	3 isoform X1			
XP_0043897 25.1	Predicted: Caspase-3	Trichechus Manatus Latirostris	Mammal ia	Sirenia
XP_0050451 77.1	Predicted: Caspase-3	Ficedula Albicollis	Birds	Passeriformes
XP_0193368 83.1	Predicted: Caspase-3	Alligator Mississippiensis	Reptilia	Crocodilia
XP_0075177 15.1	Predicted: Caspase-3	Erinaceus Europaeus	Mammal ia	Erinaceomorph a
NP_0010981 40.1	Caspase-3	Oryzias Latipes	Fish	Beloniformes
XP_0055002 35.1	Caspase-3	Columba Livia	Birds	Columbiformes
XP_0225392 61.1	Caspase-3	Astyanax Mexicanus	Fish	Characiformes
XP_0098102 05.1	Predicted: Caspase-3	Gavia Stellata	Birds	Gaviiformes
XP_0099586 77.1	Predicted: Caspase-3	Leptosomus Discolor	Birds	Leptosomiform es
XP_0085697 58.1	Predicted: Caspase-3	Galeopterus Variegatus	Mammal ia	Dermoptera

XP_0094730 93.1	Predicted: Caspase-3	Nipponia Nippon	Birds	Pelecaniformes
NP_0012905 81.1	Caspase-3	Esox Lucius	Fish	Esociformes
XP_0139102 23.1	Predicted: Caspase-3	Thamnophis Sirtalis	Reptilia	Squamata
NP_0011880 10.1	Caspase-3	Ictalurus Punctatus	Fish	Siluriformes
NP_0010812 25.1	Caspase-3	Xenopus Laevis	Amphibia	Anura
XP_0175037 48.1	Predicted: Caspase-3	Manis Javanica	Mammalia	Pholidota
NP_0012698 23.1	Caspase-3	Oreochromis Niloticus	Fish	Perciformes
XP_0208633 40.1	Caspase-3	Phascolarctos Cinereus	Mammalia	Diprotodontia
XP_0206583 85.1	Caspase-3	Pogona Vitticeps	Reptilia	Squamata
NP_001217.2	Caspase-6 Isoform Alpha Precursor	Homo Sapiens	Mammalia	Primates
XP_0168075 02.1	Predicted: Caspase-6 isoform X1	Pan Troglodytes	Mammalia	Primates

NP_033941.3	Caspase-6 Precursor	Mus Musculus	Mammalia	Rodentia
NP_113963.2	Caspase-6	Rattus Norvegicus	Mammalia	Rodentia
NP_001018333.1	Caspase-6	Danio Rerio	Fish	Cypriniformes
XP_545022.4	Caspase-6	Canis Lupus Familiaris	Mammalia	Carnivora
NP_990057.1	Caspase-6	Gallus Gallus	Birds	Galliformes
NP_001030496.1	Caspase-6	Bos Taurus	Mammalia	Artiodactyla
XP_008265713.1	Predicted: Caspase-6 isoform X1	Oryctolagus Cuniculus	Mammalia	Lagomorpha
NP_001011068.1	Caspase-6	Xenopus Tropicalis	Amphibia	Anura
XP_003221840.2	Predicted: Caspase-6	Anolis Carolinensis	Reptilia	Squamata
XP_006109127.2	Predicted: Caspase-6	Myotis Lucifugus	Mammalia	Chiroptera
XP_002197219.1	Predicted: Caspase-6	Taeniopygia Guttata	Birds	Passeriformes

XP_0079079 39.1	Predicted: Caspase-6	Callorhinchus Milii	Fish	Chimaeriformes
XP_0039724 93.2	Predicted: Caspase-6	Takifugu Rubripes	Fish	Tetraodontiformes
XP_0074374 51.1	Predicted: Caspase-6	Python Bivittatus	Reptilia	Squamata
XP_0044762 29.1	Predicted: Caspase-6 isoform X1	Dasypus Novemcinctus	Mammalia	Cingulata
XP_0059996 42.1	Predicted: Caspase-6	Latimeria Chalumnae	Fish	Coelacanthiformes
XP_0051486 83.1	Predicted: Caspase-6	Melopsittacus Undulatus	Birds	Psittaciformes
XP_0142688 41.1	Caspase-6 isoform X2	Maylandia Zebra	Fish	Perciformes
XP_0057976 75.2	Predicted: Caspase-6	Xiphophorus Maculatus	Fish	Cyprinodontiformes
XP_0052879 72.1	Predicted: Caspase-6	Chrysemys Picta Bellii	Reptilia	Testudines
XP_0043802 93.1	Predicted: Caspase-6	Trichechus Manatus Latirostris	Mammalia	Sirenia

XP_0193556 46.1	Predicted: Caspase-6 isoform X1	Alligator Mississippiensis	Reptilia	Crocodilia
XP_0055175 72.1	Predicted: Caspase-6	Pseudopodoces Humilis	Birds	Passeriformes
XP_0061623 04.2	Predicted: Caspase-6	Tupaia Chinensis	Mammal ia	Scandentia
XP_0055152 76.1	Predicted: Caspase-6 isoform X1	Columba Livia	Birds	Columbiformes
XP_0070545 43.1	Predicted: Caspase-6	Chelonia Mydas	Reptilia	Testudines
XP_0096636 99.1	Predicted: Caspase-6 isoform X1	Struthio Camelus Australis	Birds	Struthioniforme s
XP_0084978 31.1	Predicted: Caspase-6	Calypte Anna	Birds	Apodiformes
XP_0115727 91.1	Predicted: Caspase-6 isoform X1	Aquila Chrysaetos Canadensis	Birds	Accipitriformes
XP_0184255 27.1	Predicted: Caspase-6	Nanorana Parkeri	Amphibi a	Anura

XP_0152819 44.1	Predicted: Caspase-6 isoform X1	Gekko Japonicus	Reptilia	Squamata
NP_0010814 06.1	Caspase-6 L Homeolog	Xenopus Laevis	Amphibia	Anura
NP_0011177 43.1	Caspase-6 Precursor	Oncorhynchus Mykiss	Fish	Salmoniformes
NP_0012539 85.1	Caspase-7 Isoform Alpha Precursor	Homo Sapiens	Mammalia	Primates
NP_071596.1	Caspase-7	Rattus Norvegicus	Mammalia	Rodentia
NP_0010184 43.1	Caspase-7	Danio Rerio	Fish	Cypriniformes
XP_0015133 88.4	Predicted: Caspase-7	Ornithorhynchus Anatinus	Mammalia	Monotremata
XP_0056377 95.1	Caspase-7 isoform X1	Canis Lupus Familiaris	Mammalia	Carnivora
XP_421764.3	Predicted: Caspase-7	Gallus Gallus	Birds	Galliformes
XP_0209289 81.1	Caspase-7 isoform X3	Sus Scrofa	Mammalia	Artiodactyla

XP_017204061.1	Predicted: Caspase-7 isoform X2	Oryctolagus Cuniculus	Mammalia	Lagomorpha
NP_001016299.1	Caspase-7	Xenopus Laevis	Amphibia	Anura
XP_008112945.1	Predicted: Caspase-7 isoform X2	Anolis Carolinensis	Reptilia	Squamata
XP_011357571.1	Predicted: Caspase-7 isoform X1	Pteropus Vampyrus	Mammalia	Chiroptera
XP_003961640.1	Predicted: Caspase-7	Takifugu Rubripes	Fish	Tetraodontiformes
XP_012396447.1	Predicted: Caspase-7	Sarcophilus Harrisii	Mammalia	Dasyuromorphia
XP_004458121.1	Predicted: Caspase-7	Dasypus Novemcinctus	Mammalia	Cingulata
XP_006002865.1	Predicted: Caspase-7 isoform X1	Latimeria Chalumnae	Fish	Coelacanthiformes
XP_006135034.1	Predicted: Caspase-7	Pelodiscus Sinensis	Reptilia	Testudines

XP_0103311 89.1	Predicted: Caspase-7	Saimiri Boliviensis Boliviensis	Mammal ia	Primates
XP_0045675 46.1	Caspase-7 isoform X3	Maylandia Zebra	Fish	Perciformes
XP_0058044 67.1	Predicted: Caspase-7	Xiphophorus Maculatus	Fish	Cyprinodontifor mes
XP_0052939 40.1	Predicted: Caspase-7	Chrysemys Picta Bellii	Reptilia	Testudines
XP_0152031 44.1	Predicted: Caspase-7 isoform X1	Lepisosteus Oculatus	Fish	Lepisosteiform es
XP_0050488 27.1	Predicted: Caspase-7 isoform X2	Ficedula Albicollis	Birds	Passeriformes
XP_0144501 46.1	Predicted: Caspase-7 isoform X1	Alligator Mississippiensis	Reptilia	Crocodilia
XP_0040803 11.2	Caspase-7	Oryzias Latipes	Fish	Beloniformes
XP_0055208 11.1	Predicted: Caspase-7	Pseudopodoces Humilis	Birds	Passeriformes

XP_0211511 55.1	Caspase-7 isoform X1	Columba Livia	Birds	Columbiformes
NP_0012687 71.1	Caspase-7	Mesocricetus Auratus	Mammalia	Rodentia
XP_0050145 01.1	Caspase-7	Anas Platyrhynchos	Birds	Anseriformes
XP_0099658 79.1	Predicted: Caspase-7	Tyto Alba	Birds	Strigiformes
XP_0102887 94.1	Predicted: Caspase-7	Phaethon Lepturus	Birds	Phaethontiformes
XP_0094834 64.1	Predicted: Caspase-7	Pelecanus Crispus	Birds	Pelecaniformes
XP_0184199 81.1	Predicted: Caspase-7 isoform X1	Nanorana Parkeri	Amphibia	Anura
XP_0186181 89.1	Predicted: Caspase-7	Scleropages Formosus	Fish	Osteoglossiformes
NP_0010814 08.1	Caspase-7	Xenopus Laevis	Amphibia	Anura
NP_0010912 72.1	Caspase-7 S Homeolog	Xenopus Laevis	Amphibia	Anura
NP_001219.2	Caspase-8 Isoform	Homo Sapiens	Mammalia	Primates

	A Precursor			
NP_0011252 22.2	Caspase-8	Pongo Abelii	Mammalia	Primates
NP_0012648 55.1	Caspase-8 isoform 2	Mus Musculus	Mammalia	Rodentia
NP_071613.1	Caspase-8	Rattus Norvegicus	Mammalia	Rodentia
XP_0106007 86.1	Predicted: Caspase-8 isoform X1	Loxodonta Africana	Mammalia	Proboscidea
NP_571585.2	Caspase-8	Danio Rerio	Fish	Cypriniformes
NP_0010414 94.1	Caspase-8	Canis Lupus Familiaris	Mammalia	Carnivora
XP_0075016 84.1	Predicted: Caspase-8 isoform X1	Monodelphis Domestica	Mammalia	Didelphimorphia
XP_0145878 12.1	Predicted: Caspase-8	Equus Caballus	Mammalia	Perissodactyla
NP_989923.1	Caspase-8	Gallus Gallus	Birds	Galliformes
NP_0010269 49.2	Caspase-8	Sus Scrofa	Mammalia	Artiodactyla
NP_0010394 35.1	Caspase-8	Bos Taurus	Mammalia	Artiodactyla

XP_0179530 67.1	Predicted: Caspase-8	Xenopus Tropicalis	Amphibia	Anura
XP_0107117 55.1	Predicted: Caspase-8	Meleagris Gallopavo	Birds	Galliformes
XP_0143198 97.1	Predicted: Caspase-8	Myotis Lucifugus	Mammalia	Chiroptera
XP_0078828 91.1	Predicted: Caspase-8	Callorhinchus Milii	Fish	Chimaeriformes
XP_0116150 13.1	Predicted: Caspase-8	Takifugu Rubripes	Fish	Tetraodontiformes
NP_0012337 25.1	Caspase-8	Cricetulus Griseus	Mammalia	Rodentia
XP_0053063 09.1	Predicted: Caspase-8 isoform X1	Chrysemys Picta Bellii	Reptilia	Testudines
XP_0152149 52.1	Predicted: Caspase-8	Lepisosteus Oculatus	Fish	Lepisosteiformes
XP_0062725 99.1	Predicted: Caspase-8 isoform X1	Alligator Mississippiensis	Reptilia	Crocodilia
XP_0062726 09.1	Predicted: Caspase-8	Alligator Mississippiensis	Reptilia	Crocodilia

NP_0010982 58.1	Caspase-8	Oryzias Latipes	Fish	Beloniformes
XP_0129501 90.1	Predicted: Caspase-8 isoform X1	Anas Platyrhynchos	Birds	Anseriformes
XP_0099639 25.1	Predicted: Caspase-8	Tyto Alba	Birds	Strigiformes
XP_0097013 37.1	Predicted: Caspase-8	Cariama Cristata	Birds	Cariamiformes
XP_0098078 96.1	Predicted: Caspase-8	Gavia Stellata	Birds	Gaviiformes
XP_0175837 33.1	Predicted: Caspase-8 isoform X1	Corvus Brachyrhynchos	Birds	Passeriformes
XP_0099381 09.1	Predicted: Caspase-8	Opisthocomus Hoazin	Birds	Opisthocomiformes
XP_0100175 55.1	Predicted: Caspase-8	Nestor Notabilis	Birds	Psittaciformes
XP_0096825 48.1	Predicted: Caspase-8	Struthio Camelus Australis	Birds	Struthioniformes
XP_0207951 73.1	Caspase-8	Boleophthalmus Pectinirostris	Fish	Perciformes

XP_0185850 33.1	Predicted: Caspase-8 isoform X1	Scleropages Formosus	Fish	Osteoglossiformes
NP_0011871 27.1	Caspase-8	Ictalurus Punctatus	Fish	Siluriformes
NP_0010790 34.1	Caspase-8 L Homeolog	Xenopus Laevis	Amphibia	Anura
NP_116759.2	Caspase-10 Isoform 1 Preprotein	Homo Sapiens	Mammalia	Primates
NP_0011273 68.1	Caspase-10	Pongo Abelii	Mammalia	Primates
XP_0075016 70.1	Predicted: Capase-10	Monodelphis Domestica	Mammalia	Didelphimorphia
XP_0056017 38.1	Predicted: Capase-10	Equus Caballus	Mammalia	Perissodactyla
XP_421936.4	Predicted: Capase-10	Gallus Gallus	Birds	Galliformes
NP_0011551 12.1	Caspase-10	Sus Scrofa	Mammalia	Artiodactyla
NP_0010934 36.1	Caspase-10	Oryctolagus Cuniculus	Mammalia	Lagomorpha
NP_0010157 15.2	Caspase-10	Xenopus Tropicalis	Amphibia	Anura

XP_0112188 37.1	Predicted: Caspase-10 isoform X1	Ailuropoda Melanoleuca	Mammalia	Carnivora
XP_0081187 35.1	Predicted: Caspase-10 isoform X1	Anolis Carolinensis	Reptilia	Squamata
XP_0107117 45.1	Predicted: Capase-10	Meleagris Gallopavo	Birds	Galliformes
XP_0060824 64.1	Predicted: Capase-10	Myotis Lucifugus	Mammalia	Chiroptera
XP_0113674 68.1	Predicted: Caspase-10 isoform X1	Pteropus Vampyrus	Mammalia	Chiroptera
XP_0037753 43.2	Predicted: Capase-10	Sarcophilus Harrisii	Mammalia	Dasyuromorphia
XP_0129177 60.1	Predicted: Capase-10	Mustela Putorius Furo	Mammalia	Carnivora
XP_0081759 24.1	Predicted: Caspase-10 isoform X3	Chrysemys Picta Bellii	Reptilia	Testudines
XP_0043783 55.1	Predicted: Capase-10	Trichechus Manatus Latirostris	Mammalia	Sirenia

XP_0050492 34.1	Predicted: Capase-10	Ficedula Albicollis	Birds	Passeriformes
XP_0046748 84.1	Predicted: Capase-10	Condylura Cristata	Mammalia	Soricomorpha
XP_0178022 66.1	Caspase-10 Isoform X1	Papio Anubis	Mammalia	Primates
XP_0144497 89.1	Predicted: Caspase-10 isoform X1	Alligator Mississippiensis	Reptilia	Crocodilia
XP_0146402 22.1	Predicted: Capase-10	Ceratotherium simum simum	Mammalia	Perissodactyla
XP_0129501 89.1	Caspase-10	Anas Platyrhynchos	Birds	Anseriformes
XP_0143747 17.1	Predicted: Caspase-10 isoform X1	Alligator Sinensis	Reptilia	Crocodilia
XP_0102887 54.1	Predicted: Capase-10	Phaethon Lepturus	Birds	Phaethontiformes
XP_0094831 91.1	Predicted: Capase-10	Pelecanus Crispus	Birds	Pelecaniformes
XP_0097013 38.1	Predicted: Capase-10	Cariama Cristata	Birds	Cariamiformes

XP_0099242 72.1	Predicted: Capase-10	Haliaeetus Albicilla	Birds	Accipitriformes
XP_0099493 74.1	Predicted: Capase-10	Leptosomus Discolor	Birds	Leptosomiform es
XP_0089416 75.1	Predicted: Capase-10	Merops Nubicus	Birds	Coraciiformes
XP_0099381 08.1	Predicted: Capase-10	Opisthocomus Hoazin	Birds	Opisthocomifor mes
XP_0100175 44.1	Predicted: Capase-10	Nestor Notabilis	Birds	Psittaciformes
NP_0010814 10.1	Caspase-10 S Homeolog	Xenopus Laevis	Amphibi a	Anura
NP_0010831 30.1	Caspase-10 L Homeolog	Xenopus Laevis	Amphibi a	Anura
XP_0193787 51.1	Predicted: Caspase-10 isoform X1	Gavialis Gangeticus	Reptilia	Crocodilia
NP_0013007 01.1	cFLIP	Danio Rerio	Fish	Cypriniformes
XP_0081187 37.1	cFLIP	Anolis Carolinensis	Reptilia	Squamata
XP_0349930 49.1	cFLIP	Zootoca Vivipara	Reptilia	Squamata

XP_0152810 11.1	cFLIP	Gekko Japonicus	Reptilia	Squamata
XP_0267090 97.1	cFLIP	Athene Cunicularia	Birds	Strigiformes
XP_0098072 46.1	cFLIP	Gavia Stellata	Birds	Gaviiformes
XP_0094790 33.1	cFLIP	Pelecanus Crispus	Birds	Pelecaniformes
XP_0092829 34.1	cFLIP	Aptenodytes Forsteri	Birds	Sphenisciforme s
XP_0193787 55.1	cFLIP	Gavialis Gangeticus	Reptilia	Crocodilia
XP_0061180 11.1	cFLIP	Pelodiscus Sinensis	Reptilia	Testudines
XP_0075016 66.1	cFLIP	Monodelphis Domestica	Mammal ia	Didelphimorphi a
XP_0318142 08.1	cFLIP	Sarcophilus harrisii	Mammal ia	Dasyuromorphi a
XP_0192947 30.1	cFLIP	Panthera Pardus	Mammal ia	Carnivora
XP_0209303 60.1	cFLIP	Sus Scrofa	Mammal ia	Artiodactyla
XP_0324941 11.1	cFLIP	Phocoena Sinus	Mammal ia	Cetacea
XP_0274465 88.1	cFLIP	Zalophus Californianus	Mammal ia	Carnivora
XP_0194912 41.1	cFLIP	Hipposideros Armiger	Mammal ia	Chiroptera
NP_0011206 55.1	cFLIP	Homo Sapiens	Mammal ia	Primates

NP_0011251 40.1	cFLIP	Pongo Abelii	Mammal ia	Primates
XP_0283656 50.1	cFLIP	Phyllostomus Discolor	Mammal ia	Chiroptera
NP_0011884 45.1	cFLIP	Oryzias Latipes	Fish	Beloniformes
XP_0225959 48.1	cFLIP	Seriola Dumerili	Fish	Perciformes
XP_0261522 99.1	cFLIP	Mastacembelus Armatus	Fish	Synbranchiform es
XP_0199452 53.1	cFLIP	Paralichthys Olivaceus	Fish	Pleuronectiform es
NP_0012545 95.1	cFLIP	Gasterosteus Aculeatus	Fish	Gasterosteifor mes

Table S2. displays the accession number of all caspases used in the database reconstruction for AOA2

AOA 2				
Accenssio n ID	Caspase	Species	Class	Order
NP_00433 7.2	Caspase-3 isoform a preproprotein	Homo Sapiens	Mamma lia	Primates
NP_00101 2435.1	Caspase-3 isoform a preproprotein	Pan Troglodyte	Mamma lia	Primates
NP_57195 2.1	Caspase-3 apoptosis-related cysteine peptidase a	Danio Rerio	Fish	Cypriniform es
NP_00100 3042.1	Caspase-3	Canis Lupus Familiaris	Mamma lia	Carnivora

NP_001157433.1	Caspase-3	Equus Caballus	Mammalia	Perissodactyla
NP_990056.1	Caspase-3	Gallus Gallus	Birds	Galliformes
NP_001071308.1	Caspase-3	Bos Taurus	Mammalia	Artiodactyla
NP_001075586.1	Caspase-3	Oryctolagus Cuniculus	Mammalia	Lagomorpha
NP_001120900.1	Caspase-3	Xenopus Tropicalis	Amphibia	Anura
XP_014314596.1	Predicted: Caspase-3 isoform X1	Myotis Lucifugus	Mammalia	Chiroptera
XP_004610417.1	Predicted: Caspase-3	Sorex Araneus	Mammalia	Soricomorpha
XP_014351567.1	Predicted: Caspase-3	Latimeria Chalumnae	Coelacanthi	Coelacanthiformes
XP_006128558.1	Predicted: Caspase-3 isoform X1	Pelodiscus Sinensis	Reptilia	Testudines
XP_004682433.1	Predicted: Caspase-3	Condylura Cristata	Mammalia	Soricomorpha
XP_004579072.1	Predicted: Caspase-3	Ochotona Princeps	Mammalia	Lagomorpha
XP_007939063.1	Predicted: Caspase-3	Orycteropus afer afer	Mammalia	Tubulidentata
XP_006882296.1	Predicted: Caspase-3	Elephantulus Edwardii	Mammalia	Macroscelidea
NP_001098168.1	Caspase-3B	Oryzias Latipes	Fish	Beloniformes
XP_006916432.1	Predicted: Caspase-3	Pteropus Alecto	Mammalia	Chiroptera

XP_01411 4736.1	Predicted: Caspase-3	Pseudopodoces Humilis	Birds	Passeriform es
XP_00544 5415.1	Predicted: Caspase-3	Falco Cherrug	Birds	Falconiform es
XP_00705 4526.1	Predicted: Caspase-3	Chelonia Mydas	Reptilia	Testudines
XP_00503 0551.1	Caspase-3	Anas Platyrhynchos	Birds	Anseriform es
NP_00127 1764.1	Caspase-3	Macaca Fascicularis	Mamma lia	Primates
XP_00602 6683.1	Predicted: Caspase-3	Alligator Sinensis	Reptilia	Crocodilia
XP_00754 9682.1	Predicted: Caspase-3 isoform X1	Poecilia Formosa	Fish	Cyprinodon tiformes
XP_00997 0400.1	Predicted: Caspase-3	Tyto Alba	Birds	Strigiformes
XP_00947 9867.1	Predicted: Caspase-3	Pelecanus Crispus	Birds	Pelecanifor mes
XP_00992 2022.1	Predicted: Caspase-3	Haliaeetus Albicilla	Birds	Accipitrifor mes
NP_00129 0581.1	Caspase-3	Esox Lucius	Fish	Esociforme s
XP_01568 7286.1	Predicted: Caspase-3	Protobothrops Mucrosquamatus	Reptilia	Squamata
NP_00118 8010.1	Caspase-3	Ictalurus Punctatus	Fish	Siluriformes
NP_00108 1225.1	Caspase-3	Xenopus Laevis	Amphibi a	Anura
NP_00127 3018.1	Caspase-3	Capra Hircus	Mamma lia	Artiodactyla lia

NP_00126 9823.1	Caspase-3	Oreochromis Niloticus	Fish	Perciformes
NP_00121 7.2	Caspase-6 Isoform Alpha Precursor	Homo Sapiens	Mamma lia	Primates
NP_03394 1.3	Caspase-6 Precursor	Mus Musculus	Mamma lia	Rodentia
NP_11396 3.2	Caspase-6	Rattus Norvegicus	Mamma lia	Rodentia
XP_01058 8808.1	Predicted: Caspase-6	Loxodonta Africana	Mamma lia	Proboscide a
NP_00101 8333.1	Caspase-6	Danio Rerio	Fish	Cypriniform es
NP_99005 7.1	Caspase-6	Gallus Gallus	Birds	Galliformes
NP_00103 0496.1	Caspase-6	Bos Taurus	Mamma lia	Artiodactyla
NP_00101 1068.1	Caspase-6	Xenopus Tropicalis	Amphibi a	Anura
XP_01135 9747.1	Predicted: Caspase-6 Isoform X1	Pteropus Vampyrus	Mamma lia	Chiroptera
XP_01510 1393.1	Predicted: Caspase-6	Vicugna Pacos	Mamma lia	Artiodactyla
XP_00474 8146.1	Predicted: Caspase-6 Isoform X2	Mustela Putorius Furo	Mamma lia	Carnivora
XP_01442 7079.1	Predicted: Caspase-6	Pelodiscus Sinensis	Reptilia	Testudines
XP_00662 9925.2	Predicted: Caspase-6	Lepisosteus Oculatus	Holostei	Lepisosteifo rmes
XP_00438 0293.1	Predicted: Caspase-6	Trichechus Manatus Latirostris	Mamma lia	Sirenia

XP_00504 4995.1	Predicted: Caspase-6 Isoform X1	Ficedula Albicollis	Birds	Passeriform es
XP_01258 3463.1	Predicted: Caspase-6 Isoform X2	Condylura Cristata	Mamma lia	Soricomorp ha
XP_00688 1156.1	Predicted: Caspase-6	Elephantulus Edwardii	Mamma lia	Macroscelid ea
XP_00408 4686.1	Caspase-6	Oryzias Latipes	Fish	Beloniforme s
XP_01286 0317.1	Predicted: Caspase-6	Echinops Telfairi	Mamma lia	Afrosoricida
XP_01315 5138.1	Predicted: Caspase-6	Falco Peregrinus	Birds	Falconiform es
XP_00705 4543.1	Predicted: Caspase-6	Chelonia Mydas	Reptilia	Testudines
XP_02112 8354.1	Caspase-6	Anas Platyrhynchos	Birds	Anseriform es
XP_00602 5610.2	Predicted: Caspase-6	Alligator Sinensis	Reptilia	Crocodilia
XP_00756 0684.1	Predicted: Caspase-6 Isoform X1	Poecilia Formosa	Fish	Cyprinodon tiformes
XP_00996 6306.1	Predicted: Caspase-6	Tyto Alba	Birds	Strigiformes
XP_01029 0767.1	Predicted: Caspase-6	Phaethon Lepturus	Birds	Phaethontif ormes
XP_00948 3599.1	Predicted: Caspase-6	Pelecanus Crispus	Birds	Pelecanifor mes
XP_00969 9574.1	Predicted: Caspase-6	Cariama Cristata	Birds	Cariamifor mes
XP_00828 5572.1	Predicted: Caspase-6	Stegastes Partitus	Fish	Perciformes

XP_01911 0884.1	Predicted: Caspase-6 Isoform X1	Larimichthys Crocea	Fish	Perciformes
XP_01566 6700.1	Predicted: Caspase-6	Protobothrops Mucrosquamatus	Reptilia	Squamata
XP_01861 1669.1	Predicted: Caspase-6	Scleropages Formosus	Fish	Osteoglossi formes
NP_00108 1406.1	Caspase-6 L Homeolog	Xenopus Laevis	Amphibi a	Anura
XP_01937 7505.1	Predicted: Caspase-6 Isoform X1	Gavialis Gangeticus	Reptilia	Crocodilia
NP_00111 7743.1	Caspase-6 Precursor	Oncorhynchus Mykiss	Fish	Salmonifor mes
NP_00125 3985.1	Caspase-7 isoform alpha precursor	Homo Sapiens	Mamma lia	Primates
NP_07159 6.1	Caspase-7	Rattus Norvegicus	Mamma lia	Rodentia
XP_01058 7277.1	Predicted: Caspase-7 isoform X1	Loxodonta Africana	Mamma lia	Proboscide a
NP_00101 8443.1	Caspase-7	Danio Rerio	Fish	Cypriniform es
XP_42176 4.3	Predicted: Caspase-7	Gallus Gallus	Birds	Galliformes
XP_00269 8555.1	Predicted: Caspase-7	Bos Taurus	Mamma lia	Artiodactyla
NP_00101 6299.1	Caspase-7	Xenopus Tropicalis	Amphibi a	Anura
XP_01123 0112.1	Predicted: Caspase-7 isoform X2	Ailuropoda Melanoleuca	Mamma lia	Carnivora
XP_00811 2945.1	Predicted: Caspase-7 isoform X2	Anolis Carolinensis	Reptilia	Squamata

XP_01431 0783.1	Predicted: Caspase-7	Myotis Lucifugus	Mamma lia	Chiroptera
XP_00461 2023.2	Predicted: Caspase-7	Sorex Araneus	Mamma lia	Soricomorp ha
XP_00613 5034.1	Predicted: Caspase-7	Pelodiscus Sinensis	Reptilia	Testudines
XP_00515 4576.1	Predicted: Caspase-7 isoform X1	Melopsittacus Undulatus	Birds	Psittaciform es
XP_00574 7456.1	Predicted: Caspase-7 isoform X1	Pundamilia Nyererei	Actinopt erygii	Cichliforme s
XP_00468 1050.1	Predicted: Caspase-7	Condylura Cristata	Mamma lia	Soricomorp ha
XP_01445 0146.1	Predicted: Caspase-7 isoform X1	Alligator Mississippiensis	Reptilia	Crocodilia
XP_01464 1239.1	Predicted: Caspase-7	Ceratotherium simum simum	Mamma lia	Perissodact yla
XP_00683 1548.1	Predicted: Caspase-7	Chrysochloris Asiatica	Mamma lia	Afrosoricida
XP_00588 8715.1	Predicted: Caspase-7	Bos Taurus	Mamma lia	Artiodactyla
XP_00552 0811.1	Predicted: Caspase-7	Pseudopodoces Humilis	Birds	Passeriform es
XP_00523 9289.1	Predicted: Caspase-7 isoform X1	Falco Peregrinus	Birds	Falconiform es
NP_00126 8771.1	Caspase-7	Mesocricetus Auratus	Mamma lia	Rodentia
XP_00548 1147.1	Predicted: Caspase-7	Zonotrichia Albicollis	Birds	Passeriform es
XP_00757 4351.1	Predicted: Caspase-7 isoform X1	Poecilia Formosa	Fish	Cyprinodon tiformes

XP_00831 2002.1	Predicted: Caspase-7 isoform X1	Cynoglossus Semilaevis	Fish	Pleuronectif ormes
XP_00969 7487.1	Predicted: Caspase-7	Cariama Cristata	Birds	Cariamifor mes
XP_01019 8191.1	Predicted: Caspase-7	Colius Striatus	Birds	Coliiformes
XP_00827 5786.1	Predicted: Caspase-7	Stegastes Partitus	Fish	Perciformes
XP_00957 8529.1	Predicted: Caspase-7	Fulmarus Glacialis	Birds	Procellariifo rmes
XP_00992 3386.1	Predicted: Caspase-7	Haliaeetus Albicilla	Birds	Accipitrifor mes
XP_01000 6222.1	Predicted: Caspase-7 isoform X1	Chaetura Pelagica	Birds	Apodiforme s
XP_01841 9981.1	Predicted: Caspase-7 isoform X1	Nanorana Parkeri	Amphibi a	Anura
XP_01267 0088.1	Predicted: Caspase-7	Clupea Harengus	Fish	Clupeiform es
XP_01861 8189.1	Predicted: Caspase-7	Scleropages Formosus	Fish	Osteoglossi formes
NP_00109 1272.1	Caspase-7 S Homeolog	Xenopus Laevis	Amphibi a	Anura
NP_00121 9.2	Caspase-8 Isoform A Precursor	Homo Sapiens	Mamma lia	Primates
NP_00112 5222.2	Caspase-8	Pongo Abelii	Mamma lia	Primates
NP_00126 4855.1	Caspase-8 Isoform 2	Mus Musculus	Mamma lia	Rodentia
NP_07161 3.1	Caspase-8	Rattus Norvegicus	Mamma lia	Rodentia

NP_57158 5.2	Caspase-8	Danio Rerio	Fish	Cypriniform es
XP_00766 3299.1	Predicted: Caspase-8	Ornithorhynchus Anatinus	Mamma lia	Monotremat a
NP_00104 1494.1	Caspase-8	Canis Lupus Familiaris	Mamma lia	Carnivora
NP_98992 3.1	Caspase-8	Gallus Gallus	Birds	Galliformes
NP_00102 6949.2	Caspase-8	Sus Scrofa	Mamma lia	Artiodactyla
NP_00103 9435.1	Caspase-8	Bos Taurus	Mamma lia	Artiodactyla
XP_01795 3067.1	Predicted: Caspase-8	Xenopus Tropicalis	Amphibi a	Anura
XP_01136 7467.1	Predicted: Caspase-8	Pteropus Vampyrus	Mamma lia	Chiroptera
XP_00460 1328.1	Predicted: Caspase-8	Sorex Araneus	Mamma lia	Soricomorp ha
XP_01238 0773.1	Predicted: Caspase-8	Dasypus Novemcinctus	Mamma lia	Cingulata
NP_00123 3725.1	Caspase-8	Cricetulus Griseus	Mamma lia	Rodentia
XP_00530 6309.1	Predicted: Caspase-8 isoform X1	Chrysemys Picta Bellii	Reptilia	Testudines
XP_00627 2599.1	Predicted: Caspase-8 isoform X1	Alligator Mississippiensis	Reptilia	Crocodilia
XP_00627 2609.1	Predicted: Caspase-8	Alligator Mississippiensis	Reptilia	Crocodilia
NP_00109 8258.1	Caspase-8	Oryzias Latipes	Fish	Beloniforme s

XP_00996 3925.1	Predicted: Caspase-8	Tyto Alba	Birds	Strigiformes
XP_00980 7896.1	Predicted: Caspase-8	Gavia Stellata	Birds	Gaviiformes
XP_00928 2932.1	Predicted: Caspase-8	Aptenodytes Forsteri	Birds	Sphenisciformes
XP_00987 6853.1	Predicted: Caspase-8	Apaloderma Vittatum	Birds	Trogoniformes
XP_00988 8484.1	Predicted: Caspase-8	Charadrius Vociferus	Birds	Charadriiformes
XP_00946 0551.1	Predicted: Caspase-8	Nipponia Nippon	Birds	Pelecaniformes
XP_01056 3026.1	Predicted: Caspase-8	Haliaeetus Leucocephalus	Birds	Accipitriformes
XP_01040 4904.1	Caspase-8	Corvus Cornix Cornix	Birds	Passeriformes
XP_02117 5080.1	Caspase-8	Fundulus Heteroclitus	Fish	Cyprinodontiformes
XP_01858 5033.1	Predicted: Caspase-8 isoform X1	Scleropages Formosus	Fish	Osteoglossiformes
NP_00118 7127.1	Caspase-8	Ictalurus Punctatus	Fish	Siluriformes
NP_00107 9034.1	Caspase-8 L Homeolog	Xenopus Laevis	Amphibia	Anura
XP_00892 9353.1	Predicted: Caspase-8	Manacus Vitellinus	Birds	Passeriformes
XP_00345 7507.2	Predicted: Caspase-8	Oreochromis Niloticus	Fish	Perciformes
XP_02144 4269.1	Predicted: Caspase-8 isoform X1	Oncorhynchus Mykiss	Fish	Salmoniformes

XP_02261 0907.1	Predicted: Caspase-8 isoform X1	Seriola Dumerili	Fish	Perciformes
NP_11675 9.2	Caspase-10 Isoform 1 Preprotein	Homo Sapiens	Mamma lia	Primates
NP_00112 7368.1	Caspase-10	Pongo Abelii	Mamma lia	Primates
XP_00564 0587.1	Caspase-10	Canis Lupus Familiaris	Mamma lia	Carnivora
XP_42193 6.4	Predicted: Caspase-10	Gallus Gallus	Birds	Galliformes
NP_00115 5112.1	Caspase-10	Sus Scrofa	Mamma lia	Artiodactyla
NP_00109 3436.1	Caspase-10	Oryctolagus Cuniculus	Mamma lia	Lagomorph a
NP_00101 5715.2	Caspase-10	Xenopus Tropicalis	Amphibi a	Anura
XP_00811 8735.1	Predicted: Caspase-10 isoform X1	Anolis Carolinensis	Reptilia	Squamata
XP_00445 8966.1	Predicted: Caspase-10	Dasypus Novemcinctus	Mamma lia	Cingulata
XP_02159 1334.1	Caspase-10	Ictidomys Tridecemlineatus	Mamma lia	Rodentia
XP_00817 5924.1	Predicted: Caspase-10 isoform X3	Chrysemys Picta Bellii	Reptilia	Testudines
XP_01444 9789.1	Predicted: Caspase-10 isoform X1	Alligator Mississippiensis	Reptilia	Crocodilia
XP_00457 7327.1	Predicted: Caspase-10	Ochotona Princeps	Mamma lia	Lagomorph a
XP_00686 2096.1	Predicted: Caspase-10	Chrysochloris Asiatica	Mamma lia	Afrosoricida

XP_008144768.1	Predicted: Caspase-10	Eptesicus Fuscus	Mammalia	Chiroptera
XP_012421961.1	Predicted: Caspase-10 isoform X1	Odobenus Rosmarus Divergens	Mammalia	Carnivora
XP_015452237.1	Predicted: Caspase-10 isoform X1	Pteropus Alecto	Mammalia	Chiroptera
XP_014104940.1	Predicted: Caspase-10	Pseudopodoces Humilis	Birds	Passeriformes
XP_004262912.2	Predicted: Caspase-10	Orcinus Orca	Mammalia	Cetacea
XP_005484241.1	Predicted: Caspase-10	Zonotrichia Albicollis	Birds	Passeriformes
XP_014374717.1	Predicted: Caspase-10 isoform X1	Alligator Sinensis	Reptilia	Crocodilia
XP_007190566.1	Predicted: Caspase-10 isoform X1	Balaenoptera Acutorostrata Scammoni	Mammalia	Cetacea
XP_009807895.1	Predicted: Caspase-10	Gavia Stellata	Birds	Gaviiformes
XP_010125639.1	Predicted: Caspase-10	Chlamydotis Macqueenii	Birds	Otidiformes
XP_008494202.1	Predicted: Caspase-10	Calypte Anna	Birds	Apodiformes
XP_009876852.1	Predicted: Caspase-10	Apaloderma Vittatum	Birds	Trogoniformes
XP_009888483.1	Predicted: Caspase-10	Charadrius Vociferus	Birds	Charadriiformes
XP_009460550.1	Predicted: Caspase-10	Nipponia Nippon	Birds	Pelecaniformes

XP_00950 0813.1	Predicted: Caspase-10	Phalacrocorax Carbo	Birds	Suliformes
XP_01056 3039.1	Predicted: Caspase-10	Haliaeetus Leucocephalus	Birds	Accipitriformes
XP_01157 4974.1	Predicted: Caspase-10	Aquila Chrysaetos Canadensis	Birds	Accipitriformes
NP_00108 1410.1	Caspase-10 S Homeolog	Xenopus Laevis	Amphibia	Anura
NP_00108 3130.1	Caspase-10 L Homeolog	Xenopus Laevis	Amphibia	Anura
XP_01752 7989.1	Predicted: Caspase-10	Manis Javanica	Mammalia	Pholidota
XP_01937 8751.1	Predicted: Caspase-10 isoform X1	Gavialis Gangeticus	Reptilia	Crocodilia
NP_00130 0701.1	cFLIP	Danio Rerio	Fish	Cypriniformes
XP_03302 5467.1	cFLIP	Lacerta Agilis	Reptilia	Squamata
XP_02066 7408.1	cFLIP	Pogona Vitticeps	Reptilia	Squamata
XP_00933 0029.1	cFLIP	Pygoscelis Adeliae	Birds	Sphenisciformes
XP_01940 9414.1	cFLIP	Crocodylus Porosus	Reptilia	Crocodilia
XP_00602 2153.1	cFLIP	Alligator Sinensis	Reptilia	Crocodilia
XP_00611 8011.1	cFLIP	Pelodiscus Sinensis	Reptilia	Testudines
XP_01000 5644.1	cFLIP	Chaetura Pelagica	Birds	Apodiformes

XP_01012 3198.1	cFLIP	Chlamydotis Macqueenii	Birds	Otidiformes
XP_00970 7893.1	cFLIP	Cariama Cristata	Birds	Cariamifor mes
XP_03018 0293.1	cFLIP	Lynx Pardinus	Mamma lia	Carnivora
XP_02093 0360.1	cFLIP	Sus Scrofa	Mamma lia	Artiodactyla
XP_02242 3099.1	cFLIP	Delphinapterus Leucas	Mamma lia	Cetacea
XP_02698 0099.1	cFLIP	Lagenorhynchus Obliquidens	Mamma lia	Cetacea
XP_02574 9261.1	cFLIP	Callorhinus Ursinus	Mamma lia	Carnivora
XP_00564 0588.2	cFLIP	Canis Lupus Familiaris	Mamma lia	Carnivora
XP_03516 1179.1	cFLIP	Callithrix Jacchus	Mamma lia	Primates
NP_00112 0655.1	cFLIP	Homo Sapiens	Mamma lia	Primates
NP_00112 5140.1	cFLIP	Pongo Abelii	Mamma lia	Primates
XP_02441 9473.1	cFLIP	Desmodus Rotundus	Mamma lia	Chiroptera
NP_00118 8445.1	cFLIP	Oryzias Latipes	Fish	Beloniforme s
XP_02328 4372.1	cFLIP	Seriola Lalandi	Fish	Perciformes
XP_02045 0436.1	cFLIP	Monopterus Albus	Fish	Synbranchif ormes

XP_03474 0679.1	cFLIP	Etheostoma Cragini	Fish	Perciformes
NP_00125 4595.1	cFLIP	Gasterosteus Aculeatus	Fish	Gasterostei formes

Table S3. displays the accession number of all caspases used in the database reconstruction for AOA3.

AOA 3				
Accenssio n ID	Caspase	Species	Class	Order
NP_0043 37.2	Caspase-3 isoform a preprotein	Homo Sapiens	Mammali a	Primates
NP_0010 12435.1	Caspase-3	Pan Troglodytes	Mammali a	Primates
NP_0012 71338.1	Caspase-3	Mus Musculus	Mammali a	Rodentia
NP_0370 54.1	Caspase-3	Rattus Norvegicus	Mammali a	Rodentia
NP_5719 52.1	Caspase-3 Apoptotsis- related cysteine peptidase a	Danio Rerio	Fish	Cypriniform es
NP_0010 03042.1	Caspase-3	Canis Lupus Familiaris	Mammali a	Carnivora
NP_0011 57433.1	Caspase-3	Equus Caballus	Mammali a	Perissodact yla
NP_9900 56.1	Caspase-3	Gallus Gallus	Birds	Galliformes
NP_9992 96.1	Caspase-3	Sus Scrofa	Mammali a	Artiodactyla

NP_0010 71308.1	Caspase-3	Bos Taurus	Mammalia	Artiodactyla
NP_0010 75586.1	Caspase-3	Oryctolagus Cuniculus	Mammalia	Lagomorpha
NP_0011 20900.1	Caspase-3	Xenopus Tropicalis	Amphibia	Anura
XP_00790 5080.1	Predicted: Caspase-3	Callorhinchus Milii	Chondrichthyes	Chimaeriformes
NP_0012 30975.1	Caspase-3	Cricetulus Griseus	Mammalia	Rodentia
XP_00612 8558.1	Predicted: Caspase-3 isoform X1	Pelodiscus Sinensis	Reptilia	Testudines
NP_0012 66895.1	Caspase-3	Saimiri Boliviensis	Mammalia	Primates
XP_00571 9459.1	Predicted: Caspase-3	Pundamilia Nyererei	Actinopterygii	Cichliformes
XP_01933 6883.1	Predicted: Caspase-3	Alligator Mississippiensis	Reptilia	Crocodilia
XP_00442 8794.1	Predicted: Caspase-3	Ceratotherium simum simum	Mammalia	Perissodactyla
XP_00683 4493.1	Predicted: Caspase-3	Chrysochloris Asiatica	Mammalia	Afrosoricida
NP_0010 98140.1	Caspase-3	Oryzias Latipes	Fish	Beloniformes
XP_00616 0500.1	Predicted: Caspase-3 isoform X2	Tupaia Chinensis	Mammalia	Scandentia
XP_01316 1017.1	Predicted: Caspase-3 isoform X1	Falco Peregrinus	Birds	Falconiformes
XP_00705 4526.1	Predicted: Caspase-3	Chelonia Mydas	Reptilia	Testudines

XP_01412 0317.1	Predicted: Caspase-3	Zonotrichia Albicollis	Birds	Passeriform es
XP_00597 2660.1	Predicted: Caspase-3	Pantholops Hodgsonii	Mammali a	Artiodactyla
XP_00602 6683.1	Predicted: Caspase-3	Alligator Sinensis	Reptilia	Crocodilia
XP_00909 9177.1	Predicted: Caspase-3 isoform X1	Serinus Canaria	Birds	Passeriform es
XP_01028 1454.1	Predicted: Caspase-3	Phaethon Lepturus	Birds	Phaethontif ormes
XP_00969 7353.1	Predicted: Caspase-3	Cariama Cristata	Birds	Cariamifor mes
XP_00957 0651.1	Predicted: Caspase-3	Fulmarus Glacialis	Birds	Procellariifo rmes
NP_0012 90581.1	Caspase-3	Esox Lucius	Fish	Esociforme s
NP_0011 88010.1	Caspase-3	Ictalurus Punctatus	Fish	Siluriformes
NP_0010 81225.1	Caspase-3	Xenopus Laevis	Amphibia	Anura
NP_0012 69823.1	Caspase-3	Oreochromis Niloticus	Fish	Perciformes
NP_0012 17.2	Caspase-6 Isoform Alpha Precursor	Homo Sapiens	Mammali a	Primates
NP_0339 41.3	Caspase-6 Precursor	Mus Musculus	Mammali a	Rodentia
NP_1139 63.2	Caspase-6	Rattus Norvegicus	Mammali a	Rodentia
NP_0010 18333.1	Caspase-6	Danio Rerio	Fish	Cypriniform es

NP_9900 57.1	Caspase-6	Gallus Gallus	Birds	Galliformes
XP_00565 6604.1	Predicted: Caspase-6 isoform X1	Sus Scrofa	Mammalia	Artiodactyla
NP_0010 30496.1	Caspase-6	Bos Taurus	Mammalia	Artiodactyla
NP_0010 11068.1	Caspase-6	Xenopus Tropicalis	Amphibia	Anura
XP_00322 1840.2	Predicted: Caspase-6	Anolis Carolinensis	Reptilia	Squamata
XP_01442 7079.1	Predicted: Caspase-6	Pelodiscus Sinensis	Reptilia	Testudines
XP_01935 5646.1	Predicted: Caspase-6 isoform X1	Alligator Mississippiensis	Reptilia	Crocodilia
XP_01441 8265.1	Predicted: Caspase-6	Camelus Ferus	Mammalia	Artiodactyla
XP_00441 1305.1	Predicted: Caspase-6	Odobenus Rosmarus Divergens	Mammalia	Carnivora
XP_00691 8975.1	Predicted: Caspase-6 isoform X1	Pteropus Alecto	Mammalia	Chiroptera
XP_00705 4543.1	Predicted: Caspase-6	Chelonia Mydas	Reptilia	Testudines
XP_01412 0908.1	Predicted: Caspase-6 isoform X1	Zonotrichia Albicollis	Birds	Passeriformes
XP_00746 3292.1	Predicted: Caspase-6	Lipotes Vexillifer	Mammalia	Cetacea
XP_00831 5389.1	Predicted: Caspase-6	Cynoglossus Semilaevis	Fish	Pleuronectiformes

XP_00884 1740.1	Predicted: Caspase-6	Nannospalax Galili	Mammalia	Rodentia
XP_00842 0898.1	Predicted: Caspase-6	Poecilia Reticulata	Fish	Cyprinodontiformes
XP_00957 8173.1	Predicted: Caspase-6	Fulmarus Glacialis	Birds	Procellariiformes
XP_00992 4247.1	Predicted: Caspase-6	Haliaeetus Albicilla	Birds	Accipitriformes
XP_01759 2325.1	Predicted: Caspase-6	Corvus Brachyrhynchos	Birds	Passeriformes
XP_00993 0434.1	Predicted: Caspase-6	Opisthocomus Hoazin	Birds	Opisthocomiformes
XP_01018 7950.1	Predicted: Caspase-6	Mesitornis Unicolor	Birds	Mesitornithiformes
XP_01932 6128.1	Predicted: Caspase-6	Aptenodytes Forsteri	Birds	Sphenisciformes
XP_01272 4337.1	Predicted: Caspase-6 isoform X1	Fundulus Heteroclitus	Fish	Cyprinodontiformes
XP_01842 5527.1	Predicted: Caspase-6	Nanorana Parkeri	Amphibia	Anura
XP_01911 0884.1	Predicted: Caspase-6 isoform X1	Larimichthys Crocea	Fish	Perciformes
XP_01386 3217.1	Predicted: Caspase-6	Austrofundulus Limnaeus	Fish	Cyprinodontiformes
XP_01528 1944.1	Predicted: Caspase-6 isoform X1	Gekko Japonicus	Reptilia	Squamata
XP_01566 6700.1	Predicted: Caspase-6	Protobothrops Mucrosquamatus	Reptilia	Squamata
XP_01853 9675.1	Predicted: Caspase-6	Lates Calcarifer	Fish	Perciformes

NP_0010 81406.1	Caspase-6 L Homeolog	Xenopus Laevis	Amphibia	Anura
NP_0011 17743.1	Caspase-6 Precursor	Oncorhynchus Mykiss	Fish	Salmonifor mes
NP_0012 53985.1	Caspase-7 Isoform Alpha Precursor	Homo Sapiens	Mammali a	Primates
NP_0715 96.1	Caspase-7	Rattus Norvegicus	Mammali a	Rodentia
NP_0010 18443.1	Caspase-7	Danio Rerio	Fish	Cypriniform es
XP_42176 4.3	Predicted: Caspase-7	Gallus Gallus	Birds	Galliformes
NP_0010 16299.1	Caspase-7	Xenopus Tropicalis	Amphibia	Anura
XP_00811 2945.1	Predicted: Caspase-7 isoform X2	Anolis Carolinensis	Reptilia	Squamata
XP_01509 1414.1	Predicted: Caspase-7	Vicugna Pacos	Mammali a	Artiodactyla
XP_00476 3887.1	Predicted: Caspase-7	Mustela Putorius Furo	Mammali a	Carnivora
XP_00592 8237.1	Predicted: Caspase-7 isoform X1	Haplochromis Burtoni	Actinopte rygii	Cichliforme s
XP_01241 0596.1	Predicted: Caspase-7	Trichechus Manatus Latirostris	Mammali a	Sirenia
XP_01337 1286.1	Predicted: Caspase-7 isoform X1	Chinchilla Lanigera	Mammali a	Rodentia
XP_01604 2302.1	Predicted: Caspase-7	Erinaceus Europaeus	Mammali a	Erinaceomo rpha
XP_00588 8715.1	Predicted: Caspase-7	Bos Mutus	Mammali a	Artiodactyla

XP_00814 2545.1	Predicted: Caspase-7 isoform X1	Eptesicus Fuscus	Mammalia	Chiroptera
XP_01444 8895.1	Predicted: Caspase-7	Tupaia Chinensis	Mammalia	Scandentia
XP_00544 3932.1	Predicted: Caspase-7 isoform X1	Falco Cherrug	Birds	Falconiformes
NP_0012 68771.1	Caspase-7	Mesocricetus Auratus	Mammalia	Rodentia
XP_00602 2892.1	Predicted: Caspase-7 isoform X1	Alligator Sinensis	Reptilia	Crocodilia
XP_00908 5424.1	Predicted: Caspase-7	Serinus Canaria	Birds	Passeriformes
XP_00843 4898.1	Predicted: Caspase-7 isoform X1	Poecilia Reticulata	Fish	Cyprinodontiformes
XP_00894 7331.1	Predicted: Caspase-7 isoform X1	Merops Nubicus	Birds	Coraciiformes
XP_00993 2919.1	Predicted: Caspase-7	Opisthocomus Hoazin	Birds	Opisthocomiformes
XP_01018 0183.1	Predicted: Caspase-7 isoform X1	Mesitornis Unicolor	Birds	Mesitornithiformes
XP_01001 0881.1	Predicted: Caspase-7	Nestor Notabilis	Birds	Psittaciformes
XP_00966 6939.1	Predicted: Caspase-7	Struthio Camelus Australis	Birds	Struthioniformes
XP_01089 5864.1	Predicted: Caspase-7	Esox Lucius	Fish	Esociformes
XP_02077 3763.1	Caspase-7	Boleophthalmus Pectinirostris	Fish	Perciformes
XP_01841 9981.1	Predicted: Caspase-7 isoform X1	Nanorana Parkeri	Amphibia	Anura

XP_01074 0374.1	Predicted: Caspase-7	Larimichthys Crocea	Fish	Perciformes
XP_01386 3024.1	Predicted: Caspase-7 isoform X1	Austrofundulus Limnaeus	Fish	Cyprinodon tiformes
XP_01861 8189.1	Predicted: Caspase-7	Scleropages Formosus	Fish	Osteoglossi formes
NP_0010 81408.1	Caspase-7	Xenopus Laevis	Amphibia	Anura
NP_0010 91272.1	Caspase-7 S Homeolog	Xenopus Laevis	Amphibia	Anura
XP_01941 1759.1	Predicted: Caspase-7 isoform X1	Crocodylus Porosus	Reptilia	Crocodilia
XP_02063 5566.1	Caspase-7	Pogona Vitticeps	Reptilia	Squamata
NP_0012 19.2	Caspase-8 Isoform A Precursor	Homo Sapiens	Mammalia	Primates
NP_0011 25222.2	Caspase-8	Pongo Abelii	Mammalia	Primates
NP_0012 64855.1	Caspase-8 Isoform 2	Mus Musculus	Mammalia	Rodentia
NP_0716 13.1	Caspase-8	Rattus Norvegicus	Mammalia	Rodentia
NP_5715 85.2	Caspase-8	Danio Rerio	Fish	Cypriniform es
NP_0010 41494.1	Caspase-8	Canis Lupus Familiaris	Mammalia	Carnivora
NP_9899 23.1	Caspase-8	Gallus Gallus	Birds	Galliformes
NP_0010 26949.2	Caspase-8	Sus Scrofa	Mammalia	Artiodactyla

NP_0010 39435.1	Caspase-8	Bos Taurus	Mammali a	Artiodactyla
XP_01795 3067.1	Predicted: Caspase-8	Xenopus Tropicalis	Amphibia	Anura
XP_01071 1755.1	Predicted: Caspase-8	Meleagris Gallopavo	Birds	Galliformes
NP_0012 33725.1	Caspase-8	Cricetulus Griseus	Mammali a	Rodentia
XP_01277 8477.1	Caspase-8	Maylandia Zebra	Fish	Perciformes
XP_00530 6309.1	Predicted: Caspase-8 isoform X1	Chrysemys Picta Bellii	Reptilia	Testudines
XP_01521 4952.1	Predicted: Caspase-8	Lepisosteus Oculatus	Holostei	Lepisosteifo rmes
XP_00437 8277.1	Predicted: Caspase-8 isoform X1	Trichechus Manatus Latirostris	Mammali a	Sirenia
XP_01983 0532.1	Predicted: Caspase-8	Bos Indicus	Mammali a	Artiodactyla
XP_01258 1231.1	Predicted: Caspase-8 isoform X1	Condylura Cristata	Mammali a	Soricomorp ha
XP_01280 3201.1	Predicted: Caspase-8	Jaculus Jaculus	Mammali a	Rodentia
XP_00627 2599.1	Predicted: Caspase-8 isoform X1	Alligator Mississippiensis	Reptilia	Crocodilia
XP_00627 2609.1	Predicted: Caspase-8	Alligator Mississippiensis	Reptilia	Crocodilia
NP_0010 98258.1	Caspase-8	Oryzias Latipes	Fish	Beloniforme s
XP_01295 0190.1	Caspase-8 isoform X1	Anas Platyrhynchos	Birds	Anseriform es

XP_01001 7555.1	Predicted: Caspase-8	Nestor Notabilis	Birds	Psittaciform es
XP_00968 2548.1	Predicted: Caspase-8	Struthio Camelus Australis	Birds	Struthionifo rmes
XP_00989 4783.1	Predicted: Caspase-8	Picoides Pubescens	Birds	Piciformes
XP_00987 6853.1	Predicted: Caspase-8	Apaloderma Vittatum	Birds	Trogonifor mes
XP_00988 8484.1	Predicted: Caspase-8	Charadrius Vociferus	Birds	Charadriifor mes
XP_00946 0551.1	Predicted: Caspase-8	Nipponia Nippon	Birds	Pelecanifor mes
XP_01056 3026.1	Predicted: Caspase-8	Haliaeetus Leucocephalus	Birds	Accipitrifor mes
XP_02117 5123.1	Caspase-8	Fundulus Heteroclitus	Fish	Cyprinodon tiformes
XP_01858 5033.1	Predicted: Caspase-8 isoform X1	Scleropages Formosus	Fish	Osteoglossi formes
NP_0011 87127.1	Caspase-8	Ictalurus Punctatus	Fish	Siluriformes
NP_0010 79034.1	Caspase-8 L Homeolog	Xenopus Laevis	Amphibia	Anura
XP_02144 4269.1	Caspase-8 isoform X1	Oncorhynchus Mykiss	Fish	Salmonifor mes
NP_1167 59.2	Caspase-10 Isoform 1 Preprotein	Homo Sapiens	Mammali a	Primates
NP_0011 27368.1	Caspase-10	Pongo Abelii	Mammali a	Primates
XP_42193 6.4	Predicted: Caspase-10	Gallus Gallus	Birds	Galliformes

NP_0011 55112.1	Caspase-10	Sus Scrofa	Mammalia	Artiodactyla
NP_0010 93436.1	Caspase-10	Oryctolagus Cuniculus	Mammalia	Lagomorpha
NP_0010 15715.2	Caspase-10	Xenopus Tropicalis	Amphibia	Anura
XP_00274 9670.1	Predicted: Caspase-10 isoform X1	Callithrix Jacchus	Mammalia	Primates
XP_00811 8735.1	Predicted: Caspase-10 isoform X1	Anolis Carolinensis	Reptilia	Squamata
XP_01235 4911.1	Predicted: Caspase-10 isoform X1	Nomascus Leucogenys	Mammalia	Primates
XP_00817 5924.1	Predicted: Caspase-10 isoform X3	Chrysemys Picta Bellii	Reptilia	Testudines
XP_00504 9234.1	Predicted: Caspase-10	Ficedula Albicollis	Reptilia	Crocodilia
XP_01444 9789.1	Predicted: Caspase-10 isoform X1	Alligator Mississippiensis	Reptilia	Crocodilia
XP_01604 2508.1	Predicted: Caspase-10	Erinaceus Europaeus	Mammalia	Erinaceomorpha
XP_01242 1961.1	Predicted: Caspase-10 isoform X1	Odobenus Rosmarus Divergens	Mammalia	Carnivora
XP_01541 4641.1	Predicted: Caspase-10	Myotis Davidii	Mammalia	Chiroptera
XP_01444 5038.1	Predicted: Caspase-10 isoform X1	Tupaia Chinensis	Mammalia	Scandentia
XP_00557 3964.1	Predicted: Caspase-10	Macaca Fascicularis	Mammalia	Primates

XP_00548 4241.1	Predicted: Caspase-10	Zonotrichia Albicollis	Birds	Passeriform es
XP_01437 4717.1	Predicted: Caspase-10 isoform X1	Alligator Sinensis	Reptilia	Crocodilia
XP_00709 0058.1	Predicted: Caspase-10	Panthera Tigris Altaica	Mammali a	Carnivora
XP_00908 6338.1	Predicted: Caspase-10	Serinus Canaria	Birds	Passeriform es
XP_00996 3924.1	Predicted: Caspase-10	Tyto Alba	Birds	Strigiformes
XP_01758 3676.1	Predicted: Caspase-10 isoform X1	Corvus Brachyrhynchos	Birds	Passeriform es
XP_00928 2933.1	Predicted: Caspase-10	Aptenodytes Forsteri	Birds	Spheniscifo rmes
XP_01016 5007.1	Predicted: Caspase-10	Antrostomus Carolinensis	Birds	Caprimulgif ormes
XP_01022 2872.1	Predicted: Caspase-10	Tinamus Guttatus	Birds	Tinamiform es
XP_01304 3023.1	Predicted: Caspase-10 isoform X1	Anser Cygnoides Domesticus	Birds	Anseriform es
XP_01494 1825.1	Predicted: Caspase-10	Acinonyx Jubatus	Mammali a	Carnivora
XP_01598 6581.1	Predicted: Caspase-10	Rousettus Aegyptiacus	Mammali a	Chiroptera
XP_01548 9705.1	Predicted: Caspase-10	Parus Major	Birds	Passeriform es
XP_01572 3627.1	Predicted: Caspase-10 isoform X1	Coturnix Japonica	Birds	Galliformes
NP_0010 81410.1	Caspase-10 S Homeolog	Xenopus Laevis	Amphibia	Anura

NP_0010 83130.1	Caspase-10 L Homeolog	Xenopus Laevis	Amphibia	Anura
XP_01752 7989.1	Predicted: Caspase-10	Manis Javanica	Mammalia	Pholidota
XP_01937 8751.1	Predicted: Caspase-10 isoform X1	Gavialis Gangeticus	Reptilia	Crocodilia
NP_0013 00701.1	cFLIP	Danio Rerio	Fish	Cypriniform es
XP_00811 8737.1	cFLIP	Anolis Carolinensis	Reptilia	Squamata
XP_03302 5467.1	cFLIP	Lacerta Agilis	Reptilia	Squamata
XP_02502 2289.1	cFLIP	Python Bivittatus	Reptilia	Squamata
XP_02670 9097.1	cFLIP	Athene Cunicularia	Birds	Strigiformes
XP_00947 9033.1	cFLIP	Pelecanus Crispus	Birds	Pelecanifor mes
XP_00602 2153.1	cFLIP	Alligator Sinensis	Reptilia	Crocodilia
XP_01934 7600.1	cFLIP	Alligator Mississippiensis	Reptilia	Crocodilia
XP_00997 6784.1	cFLIP	Tauraco Erythrolophus	Birds	Musophagif ormes
XP_00993 8107.1	cFLIP	Opisthocomus Hoazin	Birds	Opisthocom iformes
XP_02578 6690.1	cFLIP	Puma Concolor	Mammalia	Carnivora
XP_02093 0360.1	cFLIP	Sus Scrofa	Mammalia	Artiodactyla

XP_00719 0580.1	cFLIP	Balaenoptera Acutorostrata Scammoni	Mammali a	Cetacea
XP_02698 0099.1	cFLIP	Lagenorhynchus Obliquidens	Mammali a	Cetacea
XP_02744 6588.1	cFLIP	Zalophus Californianus	Mammali a	Carnivora
XP_00564 0588.2	cFLIP	Canis Lupus Familiaris	Mammali a	Carnivora
XP_01949 1241.1	cFLIP	Hipposideros Armiger	Mammali a	Chiroptera
XP_03516 1179.1	cFLIP	Callithrix Jacchus	Mammali a	Primates
NP_0011 20655.1	cFLIP	Homo Sapiens	Mammali a	Primates
NP_0011 25140.1	cFLIP	Pongo Abelii	Mammali a	Primates
NP_0011 88445.1	cFLIP	Oryzias Latipes	Fish	Beloniforme s
XP_02938 5985.1	cFLIP	Echeneis Naucrates	Fish	Perciformes
XP_03441 7045.1	cFLIP	Cyclopterus Lumpus	Fish	Scorpaenifo rmes
XP_02050 6830.2	cFLIP	Labrus Bergylta	Fish	Perciformes
NP_0012 54595.1	cFLIP	Gasterosteus Aculeatus	Fish	Gasterostei formes

Supplementary Figure 1. Figure S1 displays the groups of sequences in the ancestral databases used to reconstruct AOA1, AOA2 and AOA3.

	Caspase-3	Caspase-6	Caspase-7	Caspase-8	Caspase-10	cFLIP
■ Average Reptiles	5.291%	6.468%	4.569%	2.797%	10.092%	10.000%
■ Average Amphibians	1.058%	1.493%	2.030%	2.797%	2.752%	0.000%
■ Average Fish	14.286%	14.428%	19.797%	9.790%	0.000%	30.000%
■ Average Birds	26.455%	27.861%	25.381%	13.986%	30.275%	20.000%
■ Average Mammals	52.910%	49.751%	48.223%	70.629%	56.881%	40.000%

■ Average Mammals ■ Average Birds ■ Average Fish ■ Average Amphibians ■ Average Reptiles

CHAPTER 4

Characterizing the evolutionary pathways of caspase stability

Abstract

The caspase family provides an excellent model for studying protein evolution, since it evolved from a common ancestor and developed new properties such as enzyme specificity and regulation. Caspases are divided into apoptotic and inflammatory classes, with apoptotic caspases further divided into initiator and effector subfamilies in the apoptotic signaling pathway, which evolved from promiscuous ancestral proteins by selecting pre-existing suboptimal activity through amino acid substitutions. These subfamilies evolved into cell fate determinants with distinct substrate specificities, and the caspase-hemoglobinase (CH) fold is highly conserved across all chordates. The context of amino acid substitutions in protein evolution is important and directed evolutionary approaches and evolutionary biochemical methods are used to examine changes in sequence, structure, and function. Evolutionary analysis can help identify the impact of past mutations on protein properties by introducing mutations into ancestral backgrounds and examining changes during protein evolution. This study examines the folding landscape of the ancestor of all apoptotic chordate caspases involved in the extrinsic pathway (AOA) to understand how the structure and functions of the two subfamilies evolved over time. Our MD simulations data indicate that AOA1 is unstable, similar to initiator caspases, while AOA3 is stable, similar to effector caspases. Five specific mutations were introduced into AOA1 to make it more stable and mimic AOA3, and vice versa. The results showed that the mutations did make AOA1 more stable and AOA3 less stable.

Introduction

The caspase family is an attractive model for examining protein evolution as caspase subfamilies evolved from a common ancestor and developed new oligomeric states, enzyme specificity, and allosteric regulation.[1] Caspase genes and their functions are ancient and well-conserved in all metazoans, and they are believed to have evolved from an ancestral immune system. Caspases are divided into two classes, apoptotic and inflammatory caspases based on their functions, with apoptotic caspases further evolving into two subfamilies characterized as initiators (caspases-2, -8, -9, -10 and cFLIP) and effectors (caspases-3, -6, and -7) in the apoptotic signaling pathway.[2,3]

Caspases are produced as inactive zymogens, and dimerization is key to regulation. The initiator caspases exist as monomers, and the ability to form heterodimers versus homodimers in response to cellular conditions is a critical feature in cell fate decisions regarding the activation of necroptosis or apoptosis pathways. In contrast, effectors exist as homodimers under physiological conditions in cells, that are processed by initiators.[4] Caspases identify a tetrapeptide motif as their target for protein cleavage, although caspase-2 is an exception since it recognizes a pentapeptide sequence. The enzyme specificity in some cases is linked to exosites that aid in substrate selection. The peptide positions P1 to P4 are coordinated with their corresponding substrate pockets, S1 to S4, in the active site and the P1 residue is almost always an aspartate.[5,6] Caspases can be divided into three categories based on their recognition of the amino acid at the P4 position, as this determines their specificity. Group I caspases prefer bulky amino acids like W or H, group II for hydrophilic residues such as D or E, and group III for aliphatic residues such as I, L, or V. Although effector

caspases share similarities, caspases-3 and -7 exhibit group II specificity, while caspase-6 exhibits group III specificity. Based on degradome analyses, the selection made at P4 leads to overlapping but distinct substrate profiles.[7,8] During the evolution of chordates, the emergence of new substrate specificities for caspases played a crucial role in the development of the brain and nervous systems.[9]

Enzymes that exist today have evolved from ancestral proteins that exhibited promiscuity, through the selection of suboptimal activities via amino acid substitutions. While the caspase-8 subfamily evolved to become determinants of cell fate, with a mostly uniform selection of substrates ((I/L)EXD), modifications in effector caspases led to two different specificities: DxxD and VxxD. Comparative studies of modern enzymes can help identify critical active site residues. However, they are often insufficient in revealing the specific residues that are responsible for functional variability in large protein families.[10]

Despite the relatively low amino acid sequence identity, which is approximately 40%, between caspase subfamilies, the caspase-hemoglobinase (CH) fold remains highly conserved across all chordates.[4] The conserved CH fold consists of a six-stranded beta-sheet core with atleast five alpha-helices on the surface. The PCP homodimer consists of two monomers, with each monomer containing approximately 300 amino acids arranged into an N-terminal prodomain and a protease domain. The protease domain is then divided into large and small subunits, which are linked by a short intersubunit linker.[11]

Horizontal studies typically lack evolutionary context, and substitutions are not examined within the framework of protein epistasis, which influences the combination of

specific amino acids that evolve along distinct evolutionary paths. Directed evolutionary approaches, on the other hand, allow for a more comprehensive exploration of the sequence space.[12,13] One such study identified a particular combination of amino acids that can relax specificity in caspase-7, resulting in a change in substrate cleavage profiles in cells expressing the evolved-caspase-7 enzymes.[8] Evolutionary biochemical methods allow for the examination of the entire protein sequence and, at the same time, consider changes that took place between common ancestral proteins. Such methods can reveal the mechanisms that underlie how modifications in protein sequence led to changes in structure and function.[14] By utilizing evolutionary analysis, it becomes possible to identify the sequence determinants that influence protein structure and function, as well as the substitutions that were present at common evolutionary nodes. These substitutions can then be introduced individually or in combination into ancestral backgrounds. This approach can help elucidate the effects of previous mutations on protein structure, function, and physical properties by examining the ancestral reconstructions and the changes that occurred during protein evolution.[11,15]

Previous studies have demonstrated that the dimerization of human PCP-3 occurs through a four-state equilibrium mechanism, resulting in a considerable increase in the conformational free energy for the dimer in comparison to the monomer.[16] Moreover, ancestral protein reconstruction techniques were used to resurrect the common ancestor (CA) of the effector (caspase-3/-6/-7) subfamily. The resulting CA, referred to as PCP-CA, forms a weak dimer that was established and stabilized at an early stage in the evolution of the subfamily. The stability of the effector caspase subfamilies was also examined, with caspase-6 being the most stable and caspase-7

being the least stable.[11] Furthermore, previous studies examining the equilibrium

unfolding of the initiators (caspase-8 and cFLIP), have shown the presence of atleast

one well-populated partially folded intermediate prior to forming the native protein.

Similar to the effector caspases, both caspase-8 and cFLIP undergo a conformational

change that is dependent on pH, suggesting the existence of a conserved mechanism

across caspases. The study indicates the presence of a conserved folding landscape

for caspases, wherein dimerization helps in the stabilization of the small subunit within

the protomer.[17]

In this study, we describe the folding landscape of the ancestor of all apoptotic

chordate caspases involved in the extrinsic pathway (AOA) to provide insights into how

the structure and functions of the two sub-families have evolved over time (Fig 1A). To

examine the robustness of the reconstruction, we resurrected two sequences (AOA-1

and AOA-3) from a pool of probabilistic sequences for AOA and performed MD

simulations. Our MD simulations data in 8M urea reveal that AOA1 has an unstable

small subunit, similar to that observed in the initiator caspases, whereas AOA-3 is

stable, similar to the effector caspases. To examine this difference in stability between

AOA1 and AOA3 proteins, five specific mutations (L212M, L238I, Y240S, A46R, and

L246K) were made in AOA1 to make it more stable and mimic AOA3. To further

investigate the impact of these mutations, they were also introduced into AOA3 (M210L,

I242L, S244Y, R20A, K205L) to mimic the AOA1 protein. The results of our in-silico

studies showed that the five mutations did make AOA1 more stable and made AOA3

less stable. These proteins were further resurrected and analyzed using biophysical and

biochemical characterization.

Results

We note that, the data presented does not focus on the phylogenetic relationship between the caspases, but rather solely focuses on the folding landscape and the evolutionary changes. Fig 1A describes the evolutionary events that led to each caspase, with lines representing the different sub-families. However, the length of the line serves to illustrate the distinct subfamilies of caspases and not the evolutionary timelines. The reconstructed and resurrected ancestor of the chordate caspases involved in the extrinsic pathway of apoptosis has been represented in Fig 1B.

AOA1 has two tryptophan residues, one in the active site loop 1 and the other in the active site loop 3, whereas AOA3 has only one tryptophan residue in the active site loop 3. AOA1 and AOA3 have 11 and 12 tyrosine residues, respectively, and they are well distributed in the primary sequence (Fig 1C). A representation of the makeup of AOA1, AOA2 and AOA3 has been shown in Fig 2A. The phylogenetic composition of AOA 1, 2, and 3 can be represented through three distinct datasets, each consisting of 200 sequences that have been selectively chosen and designed to comprehensively represent chordates (Fig 2A). Our MD simulations data indicate that the initiator caspases have an unstable small subunit, having poor hydrophobic contacts similar to what is observed in AOA1, whereas the effectors have optimal hydrophobic contacts similar to what is observed in AOA3. It is worth noting that, the ancestral sequences AOA1 and AOA3 evolved from a pool of sequences that either have optimal or suboptimal hydrophobic contacts between helix 1 and 4, helix 4 and 5, and between beta-sheet 6 and helix 5, leading to the evolution of the two subfamilies of initiator and

effector caspases (Fig. 2B). This highlights the significance of dimerization for both protein stability and evolution as demonstrated using MD simulations.

We note that, AOA1 is inactive and cannot autoprocess itself during overexpression in E.coli, whereas AOA3 is an active enzyme able to undergo self-auto activation. To perform equilibrium unfolding assays, we have mutated the active site cysteine to serine in AOA3, and the equilibrium unfolding will be performed on this inactive mutant. In view of the above, equilibrium unfolding of AOA1 and AOA1-5 is discussed in this chapter.

Native AOA1 has a fluorescence emission maximum at 334 nm when excited either at 280 nm or 295 nm, whereas AOA1-5 has a fluorescence emission maximum at 334 nm when excited at 280 nm and 340 nm when excited at 295 nm. Overall, the data show that the tryptophan residues in AOA1-5 are more solvent exposed than those of AOA-1. In pH 7.5 phosphate buffer containing 9M urea, the fluorescence maximum emission is red shifted to ~350 nm, following excitation at 280 nm or 295 nm in both proteins, indicating that the proteins remain largely unfolded under these urea conditions (Fig. 3). At intermediate urea concentrations ~4M, the emission maxima were red-shifted in the case of AOA1-5, but were largely unaffected in the case of AOA1.

Changes in the fluorescence emission and CD spectra of the initiator caspases (caspase -8,-10) and effector caspases (caspase -3,-6,-7) have been described previously. In this study, we examined the equilibrium unfolding the ancestral caspases-AOA1 and AOA1-5 mutant caspase at pH 7.5 as a function of urea concentration (0-9M), and the results are shown in Fig 4. Renaturation experiments for both proteins demonstrated that the folding transitions are reversible.

For AOA1 and AOA1-5 at pH 7.5, both the fluorescence emission data and the CD data show little to no change in the signal between 0 to 2M urea. One then observes a cooperative decrease in the signal between 3M to 8M urea to form the unfolded state. The equilibrium unfolding experiments for both proteins were performed at varied protein concentrations (2uM- 4uM), and we observe no concentration dependence across the protein concentrations, as expected for a monomer.

We will further globally fit these data to determine the free energy and the cooperativity index (m-value) of each unfolding transition. The data for AOA3 (C140S) and AOA3-5 will be obtained and globally fit to determine these parameters and provide a comparison between the ancestors to gain insights into the evolution of caspase stability.

Materials and Methods

Protein expression and purification

The ancestral proteins AOA1, AOA1-5, AOA3, AOA3-5 in the expression vector pET21b with a C-terminal hexahistidine tag were expressed in E.coli BL21 (DE3) pLysS cells and purified as previously described. [18]

Phylogenetic analysis and ancestral protein reconstruction

To reconstruct and resurrect the highly probable sequence of the last common ancestral caspase involved in the extrinsic pathway of apoptosis for chordates, we employed a curated database of caspase sequences from CaspBase (Grinshpon et al., 2018). This database contained sequences from the initiator (caspase-8/-10/cFlip) and effector

(caspase-3/-6/-7) subfamilies within the chordate lineage. A total of 600 sequences were collected from this database and were used to generate three databases, each comprising 200 sequences for ancestral protein reconstruction (APR). Representative taxa from various classes of Chordata (mammals, birds, fish, amphibians, and reptiles) were included in each database to resurrect three potential ancestral sequences (AOA1, AOA2, and AOA3). As the prodomain is highly variable due to recombination, insertions, and deletions, we removed it from the sequences using Jalview after conducting a PROMALS3D structure-based alignment. The ancestral protein reconstruction was carried out using methods previously described by Grinshpon et al.[19]

Sample preparation for equilibrium unfolding

The methodology for equilibrium unfolding experiments was implemented as previously described.[20] In brief, stock solutions of urea (10 M) were prepared in potassium phosphate buffer (50 mM potassium phosphate monobasic/potassium phosphate dibasic, pH 7.5, 1 mM DTT. The protein samples were prepared in the corresponding buffer with urea concentrations ranging from 0 M to 9 M for unfolding reactions. For refolding reactions, the protein was initially incubated in a buffer containing 10 M urea for approximately 6 hours at 25°C. The unfolded protein was then diluted with the corresponding buffer and urea to achieve final urea concentrations ranging from 0.5 M to 8 M. All solutions were freshly prepared for each experiment and were filtered using a 0.22 μm pore size filter prior to use. Final protein concentrations ranging from 2 μM to 4 μM were employed. The samples were allowed to equilibrate at 25°C for at least 16 hours.

Fluorescence emission and CD measurements

Fluorescence emission was assessed using a PTI C-61 spectrofluorometer (Photon Technology International) within the wavelength range of 300-400 nm upon excitation at 280 or 295 nm. Excitation at 280 nm enables the measurement of tyrosine and tryptophan residues fluorescence emission, while excitation at 295 nm allows for the measurement of tryptophan fluorescence emission. CD data were recorded using a J-1500 CD spectropolarimeter (Jasco) within the wavelength range of 215-250 nm. Spectral measurements were acquired using a 1 cm path length cuvette and at a constant temperature of 25°C. All data were adjusted for buffer background.

Data analysis and global fits to the equilibrium unfolding data

To achieve a uniform representation of the diverse spectroscopic signals, the original data were adjusted for buffer background and normalized between zero (representing the unfolded state) and one (representing the native state), in accordance with prior reports.[18,20,21] The relative signals of the intermediate species observed varied with the pH. The data at pH 7.5 were globally fitted as previously described, encompassing equilibrium unfolding data obtained for fluorescence emission (upon two excitations) and far-UV CD, and across different protein concentrations. The fitting model involved a two-state and a three-state equilibrium folding model, based on the description provided. The data was fitted either to a two-state model (eqn 1) or a three state-model (eqn 2) as described below. The equilibrium constants K1 and K2 relate to the equilibrium constants at respective unfolding steps.

In a three-state model, the native state (N) unfolds to the intermediate state (I) before completely unfolding.

$$N \overset{K_1}{\leftrightarrow} I \overset{K_2}{\leftrightarrow} U \qquad\qquad \text{eqn (1)}$$

For a two-state model, the native state (N) completely unfolds (U) without the presence of an intermediate state.

$$N \overset{K_1}{\leftrightarrow} U \qquad\qquad \text{eqn (2)}$$

MD simulations

The ancestral caspases were modelled using the Swiss-modeler program that uses homology modeling algorithm with user-defined templates.[22] Ancestral sequences were threaded onto the NMR structure of procaspase-8 (PDB ID:2k7z). The force field parameters for urea were obtained as described previously, and the urea molecule was built using the Avogadro software. [23] A cubic box of $6 \times 6 \times 6$ nm^3 was generated to achieve an 8M concentration as described previously. The system was subjected to energy minimization with the steepest-descent algorithm down to a maximum gradient of 2000 kJ mol^{-1} nm^{-1} and was simulated for 1 ns with annealing from 300 to 0K under an isotropic pressure of 100 bar. The system was then relaxed for 1 ns at standard pressure, heated from 0 to 300K, then simulated for 1 ns at 300K.Using the Nosé-Hoover coupling algorithm, we performed 100 ps MD simulations using the NVT (constant volume and temperature) and NPT (constant pressure and temperature) ensemble at 300K starting from the relaxed box. After heating the simulated system to 300K, a production run for each protein was conducted for 100 ps using the NPT

ensemble. MD simulations were performed for 200 ns with GROMACS using the

Amber99 force field and the TIP3P water model as described. [24,25]

References

1. Bayles KW (2013) Bacterial programmed cell death: making sense of a paradox. Available from: www.nature.com/reviews/micro

2. Allocati N, Masulli M, Di Ilio C, De Laurenzi V (2015) Die for the community: an overview of programmed cell death in bacteria. Available from: www.nature.com/cddis

3. Crawford ED, Seaman JE, Ii AB, David DC, Babbitt PC, Burlingame AL, Wells JA (2040) Conservation of caspase substrates across metazoans suggests hierarchical importance of signaling pathways over specific targets and cleavage site motifs in apoptosis. Cell Death Differ [Internet] 19:2040–2048. Available from: http://prospector.ucsf.edu/prospector/

4. Clark AC (2016) Caspase allostery and conformational selection. Chem Rev 11:6666–6706.

5. Pore M, Ba ̤, Stró A, Yk ˙, Salvesen GS, Dra̤g M Caspase Substrates and Inhibitors. Available from: http://cshperspectives.cshlp.org/

6. Talanian R V., Quinlan C, Trautz S, Hackett MC, Mankovich JA, Banach D, Ghayur T, Brady KD, Wong WW (1997) Substrate specificities of caspase family proteases. Journal of Biological Chemistry 272:9677–9682.

7. Young Lee A, Chul Park B, Jang M, Cho S, Hee Lee D, Chul Lee S, Keun Myung P, Goo Park S Identification of caspase-3 degradome by two-dimensional gel electrophoresis and matrix-assisted laser desorption/ionization-time of flight analysis.

8. Hill ME, Macpherson DJ, Wu P, Julien O, Wells JA, Hardy JA (2016) Reprogramming Caspase-7 Specificity by Regio-Specific Mutations and Selection Provides Alternate Solutions for Substrate Recognition. ACS Chem. Biol 45:32

9. Gulyaeva N.V. (2003) Nonapoptotic Functions of Caspase3 in Nervous Tissue. Biochemistry 68:1171–1180.

10. Grinshpon RD, Shrestha S, Titus-Mcquillan J, Hamilton PT, Swartz PD, Clark AC (2019) Resurrection of ancestral effector caspases identifies novel networks for evolution of substrate specificity.

11. Shrestha S, Clark AC (2021) Evolution of the folding landscape of effector caspases. J Biol Chem 297,101249:1–12.

12. Sikosek T, Chan HS Biophysics of protein evolution and evolutionary protein biophysics.

13. Figliuzzi M, Jacquier H, Schug A, Tenaillon O, Weigt M Coevolutionary Landscape Inference and the Context-Dependence of Mutations in Beta-Lactamase TEM-1.

14. Harms MJ, Thornton JW (2013) Evolutionary biochemistry: revealing the historical and physical causes of protein properties. Nature Publishing Group [Internet]. Available from: www.nature.com/reviews/genetics

15. Pillai AS, Chandler SA, Liu Y, Signore A V, Cortez-Romero CR, Benesch JLP, Laganowsky A, Storz JF, Hochberg GKA, Thornton JW (2020) Origin of complexity in haemoglobin evolution. Nature 581.

16. Bose K, Clark AC (2001) Dimeric procaspase-3 unfolds via a four-state equilibrium process. Biochemistry 40:14236–14242.

17. Nag M, Clark AC (2023) Conserved folding landscape of monomeric initiator caspases. Journal of Biological Chemistry, 103075

18. Bose K, Pop C, Feeney B, Clark AC (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.

19. Grinshpon RD, Shrestha S, Titus-Mcquillan J, Hamilton PT, Swartz PD, Clark AC (2019) Resurrection of ancestral effector caspases identifies novel networks for evolution of substrate specificity. Biochem J 476:3475–3492.

20. Walters J, Milam SL, Clark AC (2009) Practical approaches to protein folding and assembly: spectroscopic strategies in thermodynamics and kinetics. Methods Enzymol 455:1–39.

21. Bose K, Clark AC (2005) pH effects on the stability and dimerization of procaspase-3. Protein Sci 14:24–36.

22. Waterhouse A, Bertoni M, Bienert S, Studer G, Tauriello G, Gumienny R, Heer FT, de Beer TAP, Rempfer C, Bordoli L, et al. (2018) SWISS-MODEL: homology modelling of protein structures and complexes. Nucleic Acids Res 46:W296–W303.

23. Hanwell MD, Curtis DE, Lonie DC, Vandermeersch T, Zurek E, Hutchison GR (2012) Avogadro: an advanced semantic chemical editor, visualization, and analysis platform. J Cheminform 4.

24. Abraham MJ, Murtola T, Schulz R, Páll S, Smith JC, Hess B, Lindahl E (2015) Gromacs: high performance molecular simulations through multi-level parallelism from laptops to supercomputers. SoftwareX 1–2:19–25.

25. Rocco AG, Mollica L, Ricchiuto P, Baptista AM, Gianazza E, Eberini I (2008) Characterization of the protein unfolding processes induced by urea and temperature. Biophys J. 94; pp 2241-2251
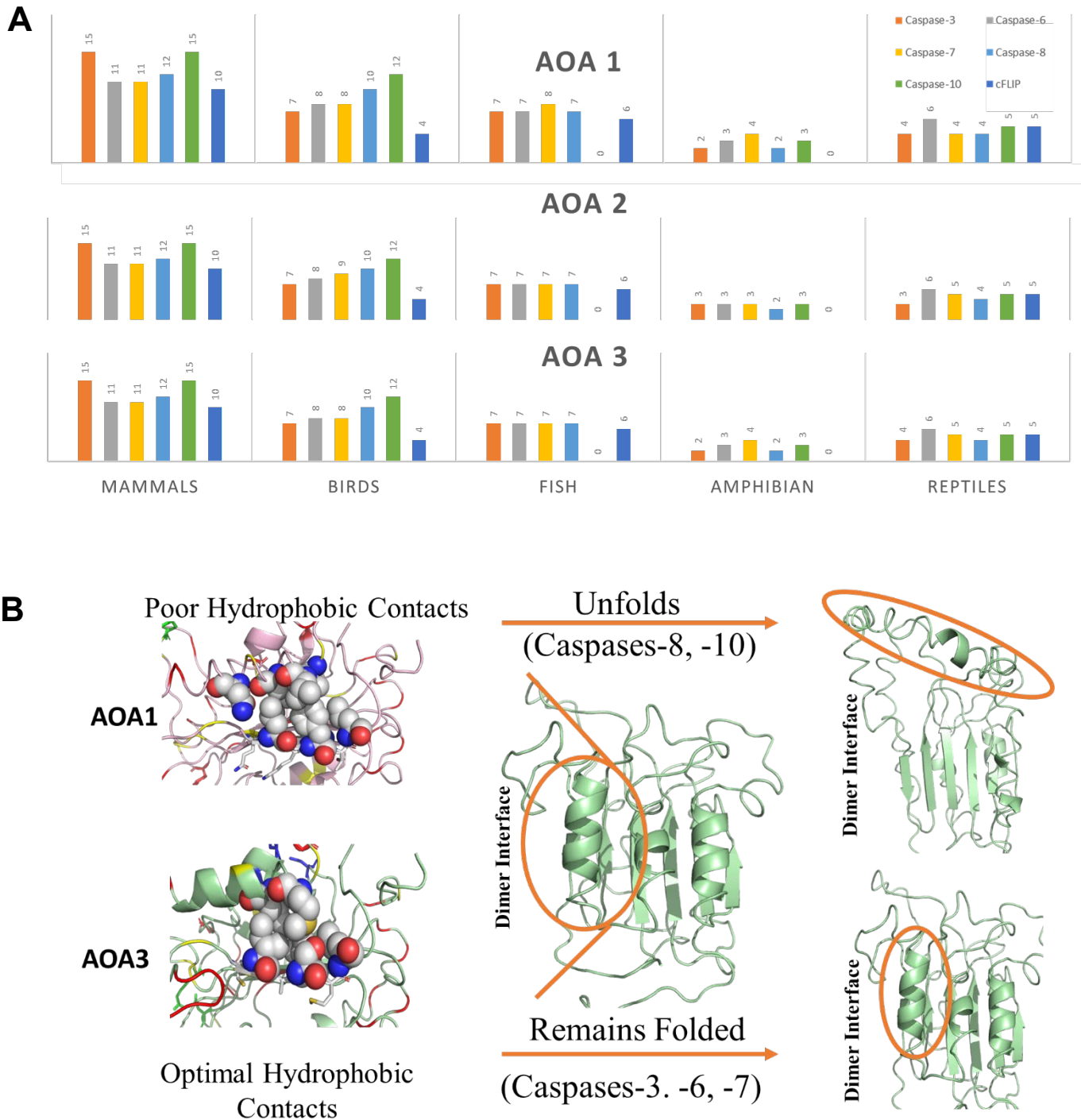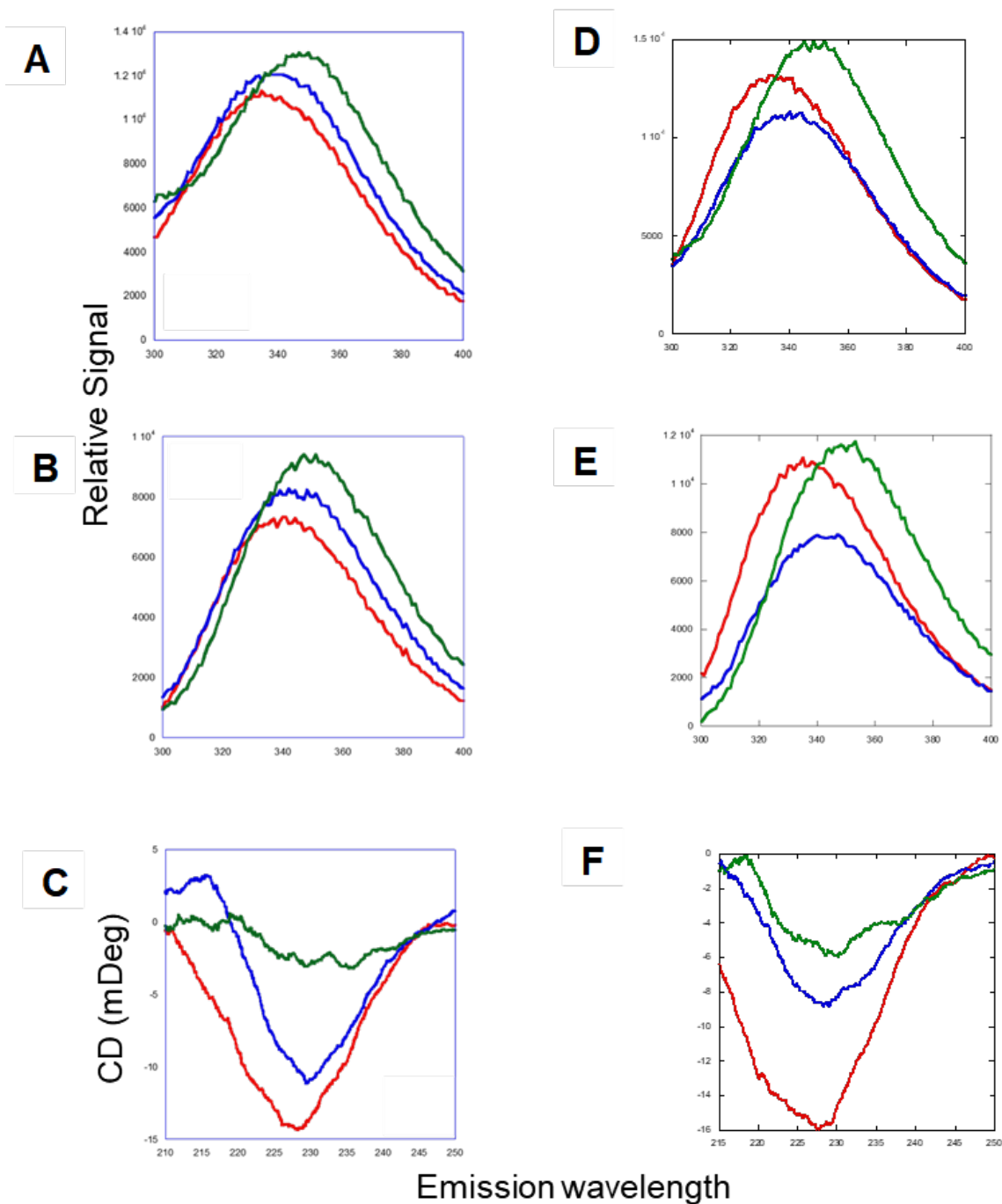
Figures

Figure 1
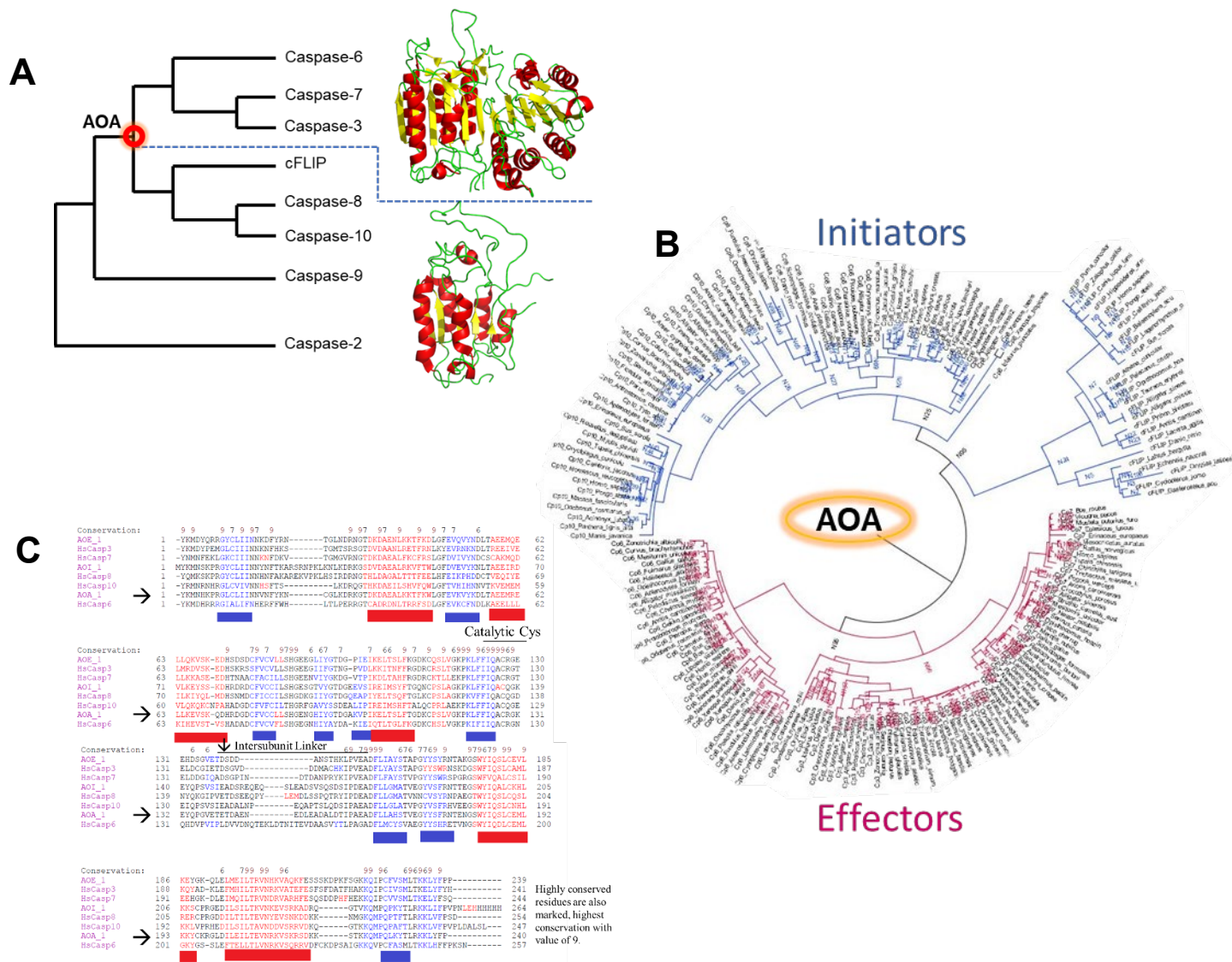
Figure 1. **Phylogenetic relationship and multiple sequence alignment** (A) All caspases evolved from a common ancestor, initiator caspases (caspase -2,-8,-9,-10, cFLIP) evolved as monomers, whereas effector caspases (caspase-3, -6, -7) evolved as dimers. (B)Phylogenetic Tree displays the species sequence originating from a common ancestor (AOA). (C)Multiple sequence alignment of initiator and effector caspases with the ancestral caspase enzymes. Beta-sheets are in blue and helices are in red.

Figure 2. Phylogenetic composition of the ancestral database and differences in the hydrophobic contacts between AOA1 and AOA3, thus affecting stability. (A) Representation of the phylogenetic make up of AOA 1, 2, and 3. Each dataset compiled totals to 200 sequences, and display differences in sequence selection. (B) The ancestor of all apoptotic caspases (AOA) evolved from a pool of sequences that either have optimal or suboptimal hydrophobic contacts in two units of 20 structure, leading to the two subfamilies of initiator and effector caspases

Figure 3

Figure 3. Fluorescence emission and circular dichroism spectra of AOA1 and AOA1-5 at pH 7.5, 25°C. Fluorescence emission spectra of AOA1 following excitation at 280 nm (A) or 295 nm (B), and CD spectra (C). Fluorescence emission spectra of AOA1-5 following excitation at 280 nm (D) or 295 nm (E), and CD spectra (F). For A-F, proteins in buffer containing 0M urea (—), 4M urea (—) and 9M urea (—).

176
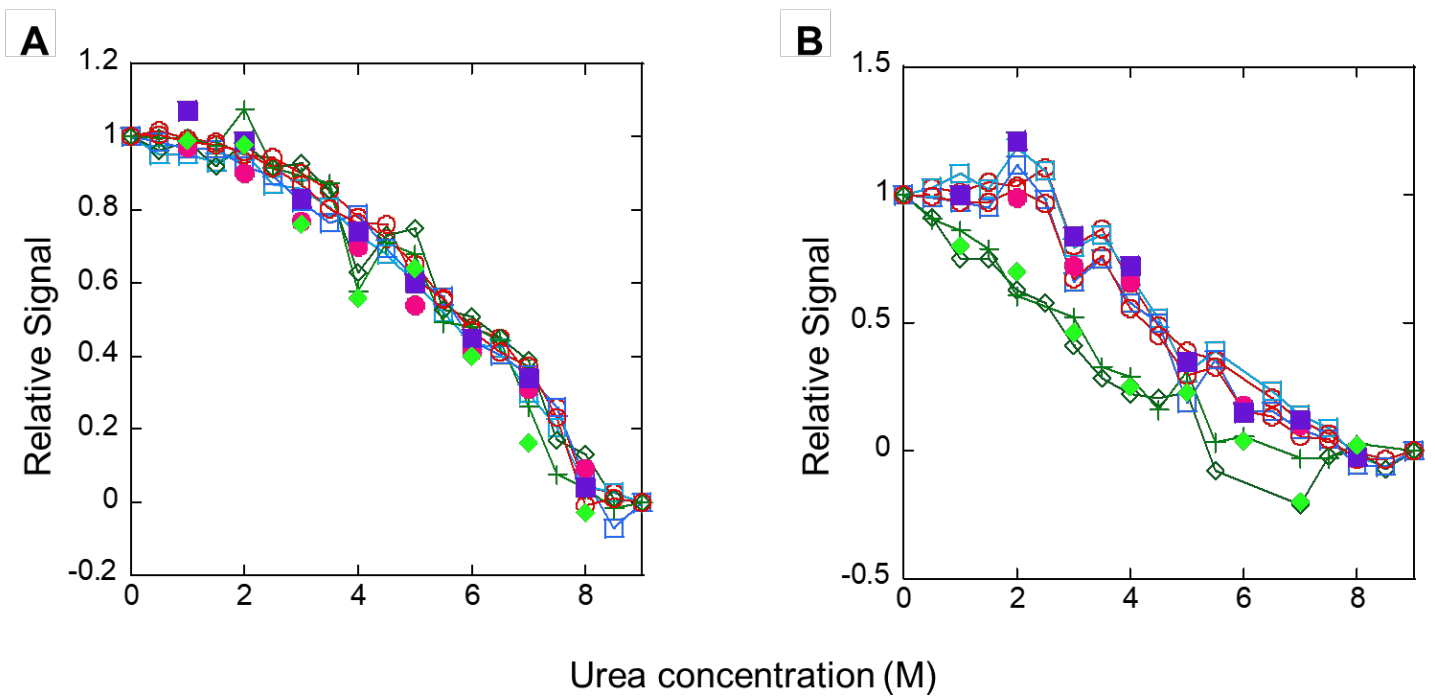
Figure 4

Figure 4. Equilibrium unfolding of AOA1 and AOA1-5 at pH 7.5. Equilibrium unfolding of AOA1 (A) and AOA1-5 (B) monitored by fluorescence emission with excitation at 280nm (○), 295 nm (□), and CD (◇) and refolding at 280nm (●), 295 nm (■), and CD (◆)

CHAPTER 5

Resurrection of an ancient inflammatory locus reveals switch to caspase-1 specificity on a caspase-4 scaffold

Betsaida Bibo-Verdugo [1], Isha Joglekar [2], Mithun N. Karadi Giridhar [2], Monica L. Ramirez [3], Scott J. Snipas [1], A. Clay Clark [2], Marcin Poreba [4], Guy S. Salvesen [1]

[1] Sanford Burnham Prebys Medical Discovery Institute, La Jolla, California, USA

[2] Department of Biology, University of Texas at Arlington, Arlington, Texas, USA

[3] Department of Pharmacology, University of California San Diego, La Jolla, California, USA

[4] Department of Bioorganic Chemistry, Wroclaw University of Science and Technology, Wroclaw, Poland

Abbreviations: ACC, 7-amino-4-carbamoylmethylcoumarin; AFC, 7-amino-4-trifluoromethylcoumarin; ASR, ancestral sequence reconstruction; CARD, caspase activation and recruitment domain; DAMP, damage-associated molecular pattern; IL-1β, interleukin-1β; IL-18, interleukin-18; pro-IL-1β, inactive precursor of IL-1β; pro-IL-18, inactive precursor of IL-18

**Please note that the abstract and the introduction has been modified to highlight the results; however, the rest of the content in this chapter is our original contribution and remains unchanged from the original article.

Abstract

This study discusses the distinct phenotypes resulting from programmed cell death mechanisms of apoptosis and pyroptosis, which are executed by the caspase family of proteases. Pyroptosis is highly inflammatory and triggers the innate immune system to clear pathogens. Here, we investigated the evolution of the inflammatory caspase locus in Carnivora and to determine the origin of the caspase-1 and caspase-4-like gene using ancestral protein reconstruction. The study aimed to discern the unique characteristics of the reconstructed ancestor, as well as the evolutionary processes that facilitated the evolution of caspase-1 function on a caspase-4 scaffold. Previous research indicates that the dog genome has a single gene encoding a "hybrid inflammatory caspase," which evolved from the loss of a chromosomal region encoding the caspase-1 catalytic domain and exon 1 of CASP4.

Introduction

The programmed cell death mechanisms of apoptosis and pyroptosis are executed by the caspase family proteases, which have distinctive cleavage specificity resulting in distinct phenotypes. Apoptosis is noninflammatory, while pyroptosis is lytic and highly inflammatory, triggering the response of the innate immune system and leading to the clearance of pathogens.[1–5] Pyroptosis execution depends on the cleavage of gasdermin D by inflammatory caspases, resulting in the release of the lytic N-terminal domain.[6–8] The pyroptotic pathways are driven by inflammatory caspases, with caspase-1 activating the canonical pathway and caspase-4 in mice and caspase-4 and caspase-5 in humans initiating the noncanonical pathway.[8–10] The release of damage-associated molecular patterns (DAMPs) and inflammatory cytokines such as IL-1β and IL-18 occurs during pyroptosis, and the proteolytic efficiency differences among caspases highlight a key difference in their contribution to innate immunity.[11–14]

The inflammatory caspases are composed of a caspase activation and recruitment domain (CARD) and a catalytic domain with proteolytic function.[15] CASP1 and CASP4 genes are usually contiguous, except in primates where CASP5 is located between them.[16] In the dog genome, there is a single gene encoding a "hybrid inflammatory caspase," which contains a caspase-1 CARD followed by a caspase-4 CARD and a catalytic domain. Early research indicates that this gene evolved from the loss of a chromosomal region encoding the caspase-1 catalytic domain and exon 1 of CASP4.[17]

Previous studies investigating the inflammatory cell death in Carnivora mammals demonstrated that the catalytic domain of the dog inflammatory caspase exhibited

catalytic behavior similar to that of caspase-1, despite having a sequence identity with caspase-4.[18] Furthermore, it appears that the gene for caspase-1 and caspase-4-like enzymes if prevalent in mammals belonging to this order.[18,19] In view of the above, Bibo-Verdugo *et al*. (2022) aimed to investigate the evolution of the inflammatory caspase locus in Carnivora and determine the origin of the caspase-1 and caspase-4-like gene. To accomplish this objective, we utilized ancestral protein reconstruction to examine the characteristics of the most probabilistic caspase ancestor that served as a precursor to the extant inflammatory caspases in Carnivora.

The process of phylogenetic tree construction via ancestral protein reconstruction entails statistical analysis of sequence conservation and substitutions of extant proteins within a given family.[20] Through this analysis, the resurrected protein can be identified as the ancestral node from which the Carnivora inflammatory caspases diverged. This study aimed to discern the unique characteristics of the reconstructed ancestor, as well as the evolutionary processes that facilitated the evolution of caspase-1 function on a caspase-4 scaffold.

Results

Resurrection of a Carnivora inflammatory caspase ancestor

Interested in the evolutionary process that generated the caspase-1 function on the caspase-4 catalytic domain scaffold, we utilized ancestral sequence reconstruction (ASR) to examine the characteristics of the protein from which the Carnivora inflammatory caspases descended. ASR calculates a phylogenetic tree based on statistical analysis of sequence conservation and substitutions of existing proteins within

a family.[21] These relationships allow for calculation of sequences that represent the diverging nodes within the phylogenetic tree and thus the ancestor of each branch. The resurrected protein represents the node from which the Carnivora inflammatory caspases diverged. We termed this caspase "node 22" based on the position on the phylogenetic tree. The results of ASR analysis are site-specific probabilities for each position in the protein sequence.[21,22] Only 5% of the node 22 caspase sequence was identified as ambiguous, defined as sites with <70% probability (Fig 5A).All ambiguous residues in the node 22 sequence are indicated in Fig. S5. The node 22 sequence shares more than 80% identity with the dog inflammatory caspase and human caspase-4 but only 60% identity with human caspase-1. Because node 22 is the predicted ancestor of the catalytic domain of the Carnivora inflammatory caspases, we expected that these proteins would have the same specificity. Accordingly, we hypothesized that the Carnivora inflammatory caspase evolved from a protein that should have been able to convert pro-IL-1β.

Methods

Phylogenetic trees and computation of ancestral sequences

Seeking proteins with the caspase CARD–CARD–catalytic domain arrangement, we used BLAST on UniProt[23], Ensembl[24], and the National Center for Biotechnology Information to retrieve related proteins employing the dog inflammatory caspase (UniProt: A9YEF4) as a query. The origin of caspase-4 is found in early mammals; hence, we focused on Mammalia in our homology search. To resurrect a highly probable sequence of the last common ancestor of the Carnivora clade (node 22), we utilized a database of curated caspase sequences (CaspBase.org) that provided

inflammatory caspase protein sequences from the chordate lineage[25] (Table S2).

PROMALS3D (prodata.swmed.edu/promals3d) generated structure-based alignments[26], and sequences were pruned on Jalview (jalview.org)[27] to remove the CARDs so that we could focus our analysis on the catalytic domain. Finally, ancestral protein reconstruction proceeded as previously described by Grinshpon et al.[20] Structural model of the ancestral reconstructed caspase was obtained by the PHYRE2 protein fold recognition server (http://www.sbg.bio.ic.ac.uk/~phyre2/html/page.cgi?id=index).[28]

References

1. Salvesen GS, Hempel A, Coll NS (2016) Protease signaling in animal and plant-regulated cell death. FEBS Journal:2577–2598.

2. Martin SJ, Henry CM, Cullen SP (2012) A Perspective on Mammalian Caspases as Positive and Negative Regulators of Inflammation. Mol Cell 46:387–397.

3. Cookson BT, Brennan MA (2001) Pro-inflammatory programmed cell death [2]. Trends Microbiol 9:113–114.

4. Man SM, Karki R, Kanneganti TD (2017) Molecular mechanisms and functions of pyroptosis, inflammatory caspases and inflammasomes in infectious diseases. Immunol Rev 277:61–75.

5. Van Opdenbosch N, Lamkanfi M (2019) Caspases in Cell Death, Inflammation, and Disease. Immunity 50:1352–1364.

6. Kayagaki N, Stowe IB, Lee BL, O'Rourke K, Anderson K, Warming S, Cuellar T, Haley B, Roose-Girma M, Phung QT, et al. (2015) Caspase-11 cleaves gasdermin D for non-canonical inflammasome signalling. Nature 2015 526:7575 526:666–671.

7. Ding J, Wang K, Liu W, She Y, Sun Q, Shi J, Sun H, Wang DC, Shao F (2016) Pore-forming activity and structural autoinhibition of the gasdermin family. Nature 2016 535:7610, 111–116.

8. Martinon F, Burns K, Tschopp J (2002) The Inflammasome: A molecular platform triggering activation of inflammatory caspases and processing of proIL-β. Mol Cell 10:417–426.

9. Kayagaki N, Warming S, Lamkanfi M, Walle L Vande, Louie S, Dong J, Newton K, Qu Y, Liu J, Heldens S, et al. (2011) Non-canonical inflammasome activation targets caspase-11. Nature 2011 479:7371,117–121

10. Viganò E, Diamond CE, Spreafico R, Balachander A, Sobota RM, Mortellaro A (2015) Human caspase-4 and caspase-5 regulate the one-step non-canonical inflammasome activation in monocytes. Nat Commun

11. Heilig R, Dick MS, Sborgi L, Meunier E, Hiller S, Broz P (2018) The Gasdermin-D pore acts as a conduit for IL-1β secretion in mice. Eur J Immunol 48:584–592.

12. Heilig R, Dick MS, Sborgi L, Meunier E, Hiller S, Broz P (2018) The Gasdermin-D pore acts as a conduit for IL-1β secretion in mice. Eur J Immunol 48:584–592.

13. Monteleone M, Stanley AC, Chen KW, Brown DL, Bezbradica JS, von Pein JB, Holley CL, Boucher D, Shakespear MR, Kapetanovic R, et al. (2018) Interleukin-1β

Maturation Triggers Its Relocation to the Plasma Membrane for Gasdermin-D-Dependent and -Independent Secretion. Cell Rep 24:1425–1433.

14. Xia S, Zhang Z, Giri Magupalli V, Lorenzo Pablo J, Dong Y, Vora SM, Wang L, Fu T-M, Jacobson MP, Greka A, et al. (2021) Gasdermin D pore structure reveals preferential release of mature interleukin-1. Nature 593:607.

15. Ramirez MLG, Salvesen GS (2018) A primer on caspase mechanisms. Semin Cell Dev Biol 82:79–85.

16. Martinon F, Tschopp J (2006) Inflammatory caspases and inflammasomes: master switches of inflammation. Cell Death & Differentiation 2007 14:1, 10–22.

17. Eckhart L, Ballaun C, Hermann M, Vandeberg JL, Sipos W, Uthman A, Fischer H, Tschachler E Identification of Novel Mammalian Caspases Reveals an Important Role of Gene Loss in Shaping the Human Caspase Repertoire.

18. Digby Z, Tourlomousis P, Rooney J, Boyle JP, Bibo-Verdugo B, Pickering RJ, Webster SJ, Monie TP, Hopkins LJ, Kayagaki N, et al. (2021) Evolutionary loss of inflammasomes in the Carnivora and implications for the carriage of zoonotic infections. Cell Rep 36.

19. Devant P, Cao A, Kagan JC (2021) Evolution-inspired redesign of the LPS receptor caspase-4 into an interleukin-1β-converting enzyme. Sci Immunol 6.

20. Grinshpon RD, Shrestha S, Titus-Mcquillan J, Hamilton PT, Swartz PD, Clark AC (2019) Resurrection of ancestral effector caspases identifies novel networks for evolution of substrate specificity.

21. Harms MJ, Thornton JW (2010) Analyzing protein structure and function using ancestral gene reconstruction. Curr Opin Struct Biol 20:360–366.

22. Eick GN, Bridgham JT, Anderson DP, Harms MJ, Thornton JW Robustness of Reconstructed Ancestral Protein Functions to Statistical Uncertainty

23. Apweiler R, Bairoch A, Wu CH, Barker WC, Boeckmann B, Ferro S, Gasteiger E, Huang H, Lopez R, Magrane M, et al. UniProt: the Universal Protein knowledgebase.

24. Howe KL, Achuthan P, Allen J, Allen J, Alvarez-Jarreta J, Ridwan Amode M, Armean IM, Azov AG, Bennett R, Bhai J, et al. (2021) Ensembl 2021. Nucleic Acids Res 49:D884–D891.

25. Grinshpon RD, Williford A, Titus-McQuillan J, Clay Clark A (2018) The CaspBase: a curated database for evolutionary biochemical studies of caspase functional divergence and ancestral sequence inference. Protein Science 27:1857–1870.

26. Pei J, Grishin N V. (2014) PROMALS3D: Multiple protein sequence alignment enhanced with evolutionary and three-dimensional structural information. Methods in Molecular Biology 1079:263–271.

27. Waterhouse AM, Procter JB, Martin DMA, Clamp M, Barton GJ (2009) Jalview Version 2-a multiple sequence alignment editor and analysis workbench. BIOINFORMATICS APPLICATIONS NOTE 25:1189–1191.

28. Kelley LA, Mezulis S, Yates CM, Wass MN, Sternberg MJE (2015) The Phyre2 web portal for protein modeling, prediction and analysis. Nature Protocols 2015 10:6 10:845–858.

Figures

Figure 5A.

We note that the following figure is referred as Fig 5A. in the published article. We shared the pymol file that contained conservation probabilities, and Bibo-Verdugo represented the structure using a different perspective in the published work.
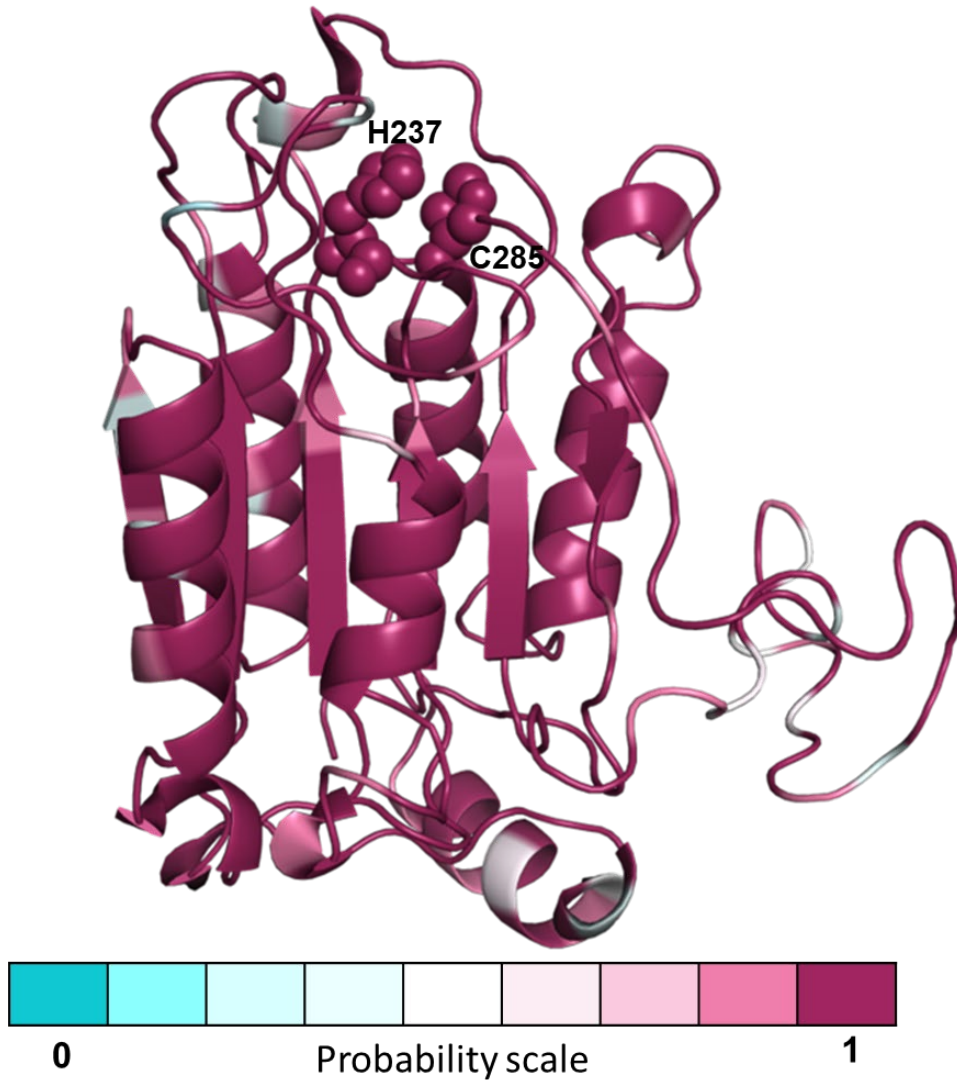
Figure 5. Site-specific probabilities of caspase node 22 and comparison with extant inflammatory caspases. A, caspase node 22 structural model was obtained by using the Phyre2 portal[126], and each protein site was colored according to the site-specific posterior probability. The catalytic residues His-237 and Cys-285 are shown as spheres.

Supplementary information

Resurrection of an ancient inflammatory locus reveals switch to caspase-1 specificity on a caspase-4 scaffold

Betsaida Bibo-Verdugo [1], Isha Joglekar [2], Mithun N. Karadi Giridhar [2], Monica L. Ramirez [3], Scott J. Snipas [1], A. Clay Clark [2], Marcin Poreba [4], Guy S. Salvesen [1]

[1] Sanford Burnham Prebys Medical Discovery Institute, La Jolla, California, USA

[2] Department of Biology, University of Texas at Arlington, Arlington, Texas, USA

[3] Department of Pharmacology, University of California San Diego, La Jolla, California, USA

[4] Department of Bioorganic Chemistry, Wroclaw University of Science and Technology, Wroclaw, Poland

Keywords: caspase; cell death; inflammation; interleukin 1; protein evolution

Abbreviations: ACC, 7-amino-4-carbamoylmethylcoumarin; AFC, 7-amino-4-trifluoromethylcoumarin; ASR, ancestral sequence reconstruction; CARD, caspase activation and recruitment domain; DAMP, damage-associated molecular pattern; IL-1β, interleukin-1β; IL-18, interleukin-18; pro-IL-1β, inactive precursor of IL-1β; pro-IL-18, inactive precursor of IL-18

**Please note that the supplementary information only relevant to our contribution has been included in this section.

Supplementary table 2. List of taxa and caspases used in the reconstruction of a Carnivora inflammatory caspase ancestor.

Caspase	Species
Caspase-1	Alligator mississippiensis
Caspase-1	Aptenodytes forster
Caspase-1	Bos taurus
Caspase-1	Camelus bactrianus
Caspase-1	Camelus dromedarius
Caspase-1	Camelus ferus
Caspase-1	Callithrix jacchus
Caspase-1	Castor canadensis
Caspase-1	Cebus capucinus imitator
Caspase-1	Chlorocebus sabaeus
Caspase-1	Colobus angolensis
Caspase-1	Columba livia
Caspase-1	Cricetulus griseus
Caspase-1	Danio rerio
Caspase-1	Dasypus novemcinctus
Caspase-1	Delphinapterus leucas
Caspase-1	Echinops telfairi
Caspase-1	Equus asinus
Caspase-1	Equus caballus
Caspase-1	Equus przewalskii
Caspase-1	Fukomys damarensis
Caspase-1	Galeopterus variegatus
Caspase-1	Gallus gallus
Caspase-1	Gorilla gorilla gorilla
Caspase-1	Heterocephalus glaber
Caspase-1	Hipposideros armiger
Caspase-1	Homo sapiens
Caspase-1	Jaculus jaculus
Caspase-1	Lepisosteus oculatus
Caspase-1	Macaca fascicularis
Caspase-1	Macaca mulatta
Caspase-1	Macaca nemestrina
Caspase-1	Mandrillus leucophaeus
Caspase-1a	Maylandia zebra
Caspase-1	Melopsittacus undulatus
Caspase-1	Microcebus murinus
Caspase-1	Microtus ochrogaster
Caspase-1	Mus caroli
Caspase-1	Mus musculus

Caspase-1	Mus Pahari
Caspase-1	Myotis brandtii
Caspase-1	Myotis davidii
Caspase-1	Nannospalax galili
Caspase-1	Nomascus leucogenys
Caspase-1	Octodon degus
Caspase-1	Otolemur garnettii
Caspase-1	Orcinus orca
Caspase-1	Oreochromicus niloticus
Caspase-1a	Oreochromus niloticus
Caspase-1	Orycteropus afer afer
Caspase-1	Oryctogalus cuniculus
Caspase-1	Pan paniscus
Caspase-1	Pan troglodytes
Caspase-1	Papio Anubis
Caspase-1	Peromyscus manicula
Caspase-1	Pongo abellii
Caspase-1	Propithecus coquereli
Caspase-1	Pteroptus vampyrus
Caspase-1	Pygocentrus nattereri
Caspase-1	Rattus norvegicus
Caspase-1	Rhinopthecus bieti
Caspase-1	Rhinopthecus roxellana
Caspase-1	Rousettus aegyptiacus
Caspase-1	Saimiri boliviensis
Caspase-1	Sus scrofa
Caspase-1	Trichechus manatus
Caspase-1	Tupaia chinensis
Caspase-1	Tursiops truncatus
Caspase-1	Vicugna pacos
Caspase-1a	Xenopus laevis
Caspase-1b	Xenopus laevis
Caspase-4	Aotus nancymaae
Caspase-4	Bos taurus
Caspase-4	Callithrix jacchus
Caspase-4	Cebus capucinus imitator
Caspase-4	Cerocebus atys
Caspase-4	Chinchilla lanigera
Caspase-4	Chlorocebus sabaeus
Caspase-4	Colobus angolensis
Caspase-4	Equus caballus
Caspase-4	Gorilla gorilla gorilla
Caspase-4	Heterocephalus glaber
Caspase-4	Homo sapiens
Caspase-4	Jaculus jaculus

Caspase-4	Macaca fascicularis
Caspase-4	Macaca mulatta
Caspase-4	Macaca nemestrina
Caspase-4	Mandrillus leucophaeus
Caspase-4	Mus caroli
Caspase-4	Mus musculus
Caspase-4	Mus pahari
Caspase-4	Mustela putorious furo
Caspase-4	Myotis lucifugus
Caspase-4	Nannospalax galili
Caspase-4	Nomascus leucogenys
Caspase-4	Ochotona princeps
Caspase-4	Oryctolagus cuniculus
Caspase-4	Otolemur garnettii
Caspase-4	Ovis aries
Caspase-4	Pan troglodytes
Caspase-4	Papio anubis
Caspase-4	Pongo abelii
Caspase-4	Rattus norvegicus
Caspase-4	Rhinopthecus roxellana
Caspase-4	Saimiri boliviensis
Caspase-4	Sarcophilus harrisi
Caspase-4	Tupaia chinensis
Caspase-5	Dipodomys ordii
Caspase-5	Gorilla gorilla gorilla
Caspase-5	Homo sapiens
Caspase-5	Macaca fascicularis
Caspase-5	Macaca mulatta
Caspase-5	Macaca nemestrina
Caspase-5	Mandrillus leucophaeus
Caspase-5	Nomascus leucogenys
Caspase-5	Pan troglodytes
Caspase-5	Papio anubis
Caspase-5	Pongo abelii
Caspase-5	Rhinopthecus roxellana
Caspase-5	Saimiri boliviensis
Hybrid Caspase	Acinonyx jubatus
Hybrid Caspase	Ailuropoda melanoleuca
Hybrid Caspase	Canis lupus familiaris
Hybrid Caspase	Enhydra lutris kenyoni
Hybrid Caspase	Eumetopias jubatus
Hybrid Caspase	Felis catus
Hybrid Caspase	Halichoerus grypus
Hybrid Caspase	Hyaena hyaena
Hybrid Caspase	Leptonychotes weddellii

Hybrid Caspase	Lontra canadensis
Hybrid Caspase	Lynx canadensis
Hybrid Caspase	Lynx pardinus
Hybrid Caspase	Mirounga leonina
Hybrid Caspase	Neomonachus schauinslandi
Hybrid Caspase	Odobenus rosmarus
Hybrid Caspase	Panthera pardus
Hybrid Caspase	Panthera tigris
Hybrid Caspase	Phoca vitulina
Hybrid Caspase	Puma concolor
Hybrid Caspase	Ursus arctos horribilis
Hybrid Caspase	Ursus maritimus
Hybrid Caspase	Vulpes vulpes
Hybrid Caspase	Zalophus california

Supplementary figure 4. Phylogenetic tree of the inflammatory caspases used for the ancestral protein reconstruction. The last Carnivora inflammatory caspase ancestor, node 22, is highlighted on the red branch comprising the Carnivora inflammatory caspases, which are often annotated as hybrid caspases.
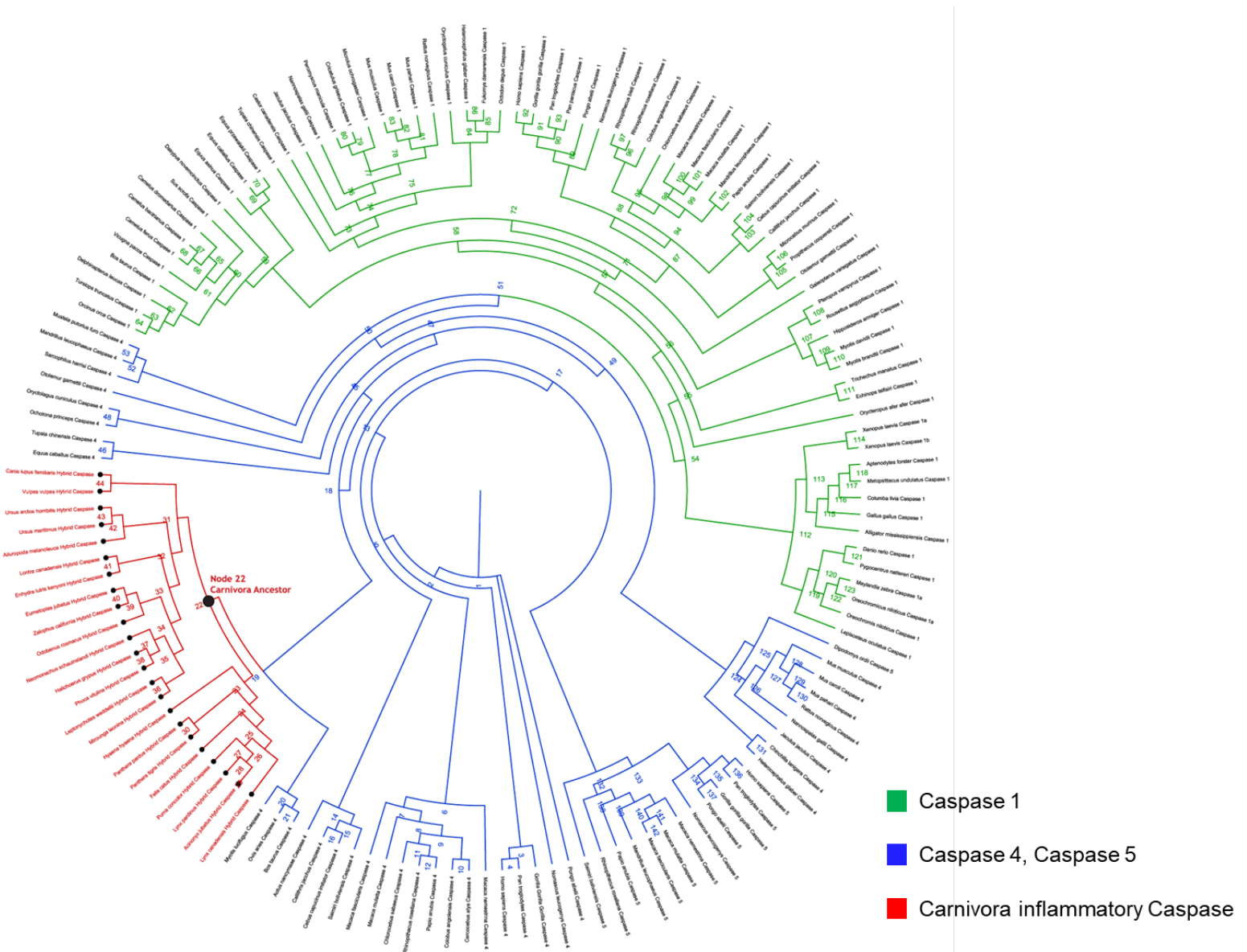
Supplementary figure 5. Sequence alignment of node 22, human caspases-1 and -4, and the dog inflammatory caspase catalytic domains. Ambiguous residues are represented with asterisks and percent identity values of node 22 to other inflammatory caspases are shown in parenthesis. Catalytic residues His-237 and Cys-285 are indicated in bold letters. Amino acids within loop-341 involved in substrate interaction are underlined. Residue 342 is a major difference within inflammatory caspases and is highlighted in gray. The caspase-1 numbering system is used.

```
hCasp-1        VKLCSLEEAQRIWKQKSAEIYPIMDKSSRTRLALIICNEEFDSIPRRTGAEVDITGMTML
Dog inf casp   LKLCPPETFVKMYKEKAEEIYPIKERKDRTRLALIICNIEFDHLSTRDGAELDIAGMESL
hCasp-4        LKLCPHEEFLRLCKERAEEIYPIKERNNRTRLALIICNTEFDHLPPRNGADFDITGMKEL
Node 22        LKLCPHEEFVKLCKERAEEIYPIKERKDRTRLALIICNTEFDHLPPRNGADLDIAGMKRL
                         *  *                                        *        *

                                           237
                                            |
hCasp-1        LQNLGYSVDVKKNLTASDMTTELEAFAHRPEHKTSDSTFLVFMSHGIREGICGKKHSEQV
Dog inf casp   LEGLGYSVVVKRKLTAKGMESVLREFAARPEHKSSDSTFLVLMSHGILNGICGTAHSVEN
hCasp-4        LEGLDYSVDVEENLTARDMESALRAFATRPEHKSSDSTFLVLMSHGILEGICGTVHDEKK
Node 22        LEGLGYSVDVKEKLTAKDMESVLRAFAARPEHKSSDSTFLVFMSHGILSGICGTTHSPEN
                   *                 *                              *    *  *

                                      285
                                       |
hCasp-1        PDILQLNAIFNMLNTKNCPSLKDKPKVIIIQACRGDSPGVVWFKDSVGVSGNLSLPTTEE
Dog inf casp   PDVLAYDTIFQIFNNRHCLNLKDKPKVIIIQACRGENPGELWVSDSPKASTDSWTHQPLM
hCasp-4        PDVLLYDTIFQIFNNRNCLSLKDKPKVIIVQACRGANRGELWVRDSPASLEVASSQSSEN
Node 22        PDVLPYDTIFQIFNNRNCLSLKDKPKVIIVQACRGENLGELWVSDSPAAPTDSSQSPEN
                                                           *      *  *

                                  342
                                   |
hCasp-1        FEDDAIKKAHIEKDFIAFCSSTPDNVSWRHPTMGSVFIGRLIEHMQEYACSCDVEEIFRK
Dog inf casp   LQSDAIHKVHVEKDFIAFCSSTPHNVSWRHITKGSLFIAQLITCFQKYSWCCHLEGVFRK
hCasp-4        LEEDAVYKTHVEKDFIAFCSSTPHNVSWRDSTMGSIFITQLITCFQKYSWCCHLEEVFRK
Node 22        LEDDAIYKVHVEKDFIAFCSSTPHNVSWRDVTKGSLFITQLITCFQKYSWCCHLEEVFRK
                   *     *                   |_____|
                                                *
                                              341-loop
hCasp-1        VRFSFEQPDGRAQMPTTERVTLTRCFYLFPGH (60%)
Dog inf casp   VQQSFEKPDVKAQMPTIERVSMTRYFYLFPGH (85.7%)
hCasp-4        VQQSFETPRAKAQMPTIERLSMTRYFYLFPGN (84.2%)
Node 22        VQQSFEKPNVKAQMPTIERLSMTRYFYLFPGN
```

194

CHAPTER 6

Examining the basis of pH-dependent conformational change in caspases

Abstract

This study investigated the role of highly conserved charged amino acid residues responsible for the pH-dependent conformational change observed in the equilibrium unfolding studies of effector and initiator caspases from H. *sapiens*, O. *faveolata*, P. *astreiodes* and D. *rerio*. We hypothesized that pH-sensitive amino acid residues such as histidine, aspartate or lysine, may be crucial to the conformational changes observed. To investigate this, we mutated certain highly conserved residues in human caspase-8 and human caspase-3, which are involved in the extrinsic pathway of apoptosis. Our findings suggest that neither of the mutants contribute significantly to the conformational change, indicating that this pH-dependent destabilization is likely due to modifications to a network of amino-acid interactions or a combination of the mutated sites but not a single amino acid.

Introduction

Caspases exist as inactive zymogens in normal cells and are activated through proteolytic processing and dimerization. Executioner caspases, such as caspase-3/-6/-7, are inactive because their active sites are not properly oriented, and upon maturation, the active-site loops undergo reorientations that result in the formation of the catalytic groove, rotation of the cysteine, formation of the oxyanion hole, and stabilization of the active site through the formation of new hydrogen bonds.[1,2]

Previous studies have shown that procaspase-3 is catalytically competent but with lower activity than the mature enzyme. The activity is due to the opening of the active site and rearrangement of three of the four active-site loops that ultimately allow the cysteine to rotate toward the substrate binding site. The intracellular pH also affects the activation of caspase-3, with a decrease in pH increasing the rate of autoactivation of procaspase-3.[3–5] Moreover, procaspase-3 undergoes a pH-dependent conformational change, and the dimer dissociates below pH 5.5.[6] Recent studies on procaspase-6/-7 have also suggested a similar pH-dependent conformational change with a pKa ~5.9. In contrast to procaspase-3, the data for procaspase-6/-7 show that the proteins remain dimeric, although the dimer is destabilized at a lower pH.[7] Furthermore, studies on initiator caspases-8/cFLIP have suggested a similar pH-dependent conformational change that is also observed in effector caspases from other organisms like D. *rerio* and initiator caspases from O. *faveolata* and P. *astreoides*.[8–10]

Ionizable amino acid residues play crucial roles in protein binding and enzyme action mechanisms, as well as influencing protein structure, stability, and solubility.[11,12] The pKa values of these residues are important because they determine their protonation

196

state and interactions with their environment. The local environment strongly influences pKa values, and pH-dependent changes in protein stability can regulate essential cellular processes.[13,14] These residues can form electrostatic interactions or salt bridges with other charged or polar amino acids in the protein, either stabilizing or destabilizing protein conformations.[15] In some instances, charged residues located in allosteric sites can act as switches triggering conformational changes, subsequently leading to the activation of downstream signaling pathways. Overall, charged residues are crucial for allosteric regulation as they play a pivotal role in the conformational changes that occur within a protein. These residues help to modify interactions between distinct regions of the protein, thereby functioning as essential components of the intricate network of interactions that regulate protein activity and function.[16,17]

Since equilibrium folding studies conducted on all caspases so far have shown a pH-dependent conformational change, we investigated the role of conserved charged residues using pH titration studies.[6–10] This study aimed to determine the crucial amino-acids responsible for this pH-dependent destabilization, based on the hypothesis that this conformational change observed is due to pH-sensitive amino acids like histidine, aspartate or lysine residues which have been conserved over hundreds of millions of years of evolution. Consequently, the study involved the systematic mutation of highly conserved residues in human caspase-8 and human caspase-3 which are involved in the extrinsic pathway of apoptosis.

Results

As shown in Fig 1A, two highly conserved aspartate residues D70 and D192, mapped onto the crystal structure of caspase-3 (PDB ID:2j30) and in Fig 1B, three highly

conserved residues K136, K138 and H89, mapped onto the NMR structure of caspase-8 (PDB ID:2k7z) were potential amino-acids that were suspected to be responsible for the pH-dependent conformational change observed in all caspases. These five sites were mutated to alanine, and we performed pH titration experiments as well as equilibrium unfolding experiments to monitor the changes in the stability in these mutants and compare them to the wild type proteins.

The pH-titration was measured using fluorescence emission and circular dichroism to monitor changes in the secondary as well as tertiary structure of the proteins. On mutating the potential residues, we expected a significant change in the transition with a pKa ~ 6 which is observed in all other wild-type caspases. Our pH titration experiments suggested that neither of the mutants significantly contributed to the conformational change as we decreased the pH. The data indicate that the pH-dependent conformational change observed is likely due to modifications to a network of amino-acid interactions, and not just a single amino-acid (Fig 2). The fitting parameters of this transition observed indicate that the pKa of caspase3-D70A is slightly greater than that observed with the other mutants, however, this difference is not significant (Fig 3). From the fitting parameters, one observes that procaspase-3 has twice the number of protons compared to the other proteins, suggesting that all the others are indeed monomers. As is observed from our expression and purification studies, the caspase3-D70A mutation and caspase3-D192A mutation prevent caspase-3 from becoming a dimer, since it does not auto process itself and remains inactive.

Further, we performed limited trypsin proteolysis assay to demonstrate that the mutant forms were less stable than the wild type (Fig 4). One observes that the mutant

forms of caspase-8 and caspase-3 get rapidly cleaved within the first fifteen minutes itself when compared to the wild type procaspase-3 and procaspase-8. However, circular dichroism data suggest that the secondary structure is still maintained in the mutants, regardless of the loss in stability. This suggests that the mutant forms have regions that are fluctuating due to destabilization of the protein. We also examined the equilibrium unfolding of these mutant forms and the data suggest that the mutants lack the cooperativity observed in the equilibrium unfolding of the wild type procaspases.

Together, the data suggest that these mutations affect the overall stability of caspases, however, these point mutants do not show a significant change in the pH-dependence transitions. This indicates that a network of interacting ionizable residues or a combination of the studied mutants may be responsible for the pH-dependent conformational change observed across all caspases, but not a single point mutation. Further mutational studies and biophysical characterization may provide insights on the mechanisms of pH-dependent instability in the caspase family of proteases.

Materials and Methods

Protein expression and purification in E.coli

All mutant proteins were expressed in pET21b using 0.5mM IPTG and overnight induction at 25 °C. The proteins were purified under denaturing conditions since, they appeared as inclusion bodies. The pellets were resuspended in 6M urea overnight and were spun down the next day at 15000 rpm for 30 minutes. The supernatant was collected and diluted to a final urea concentration of 3M and was then permitted to bind to the Ni-NTA resin for affinity chromatography. The proteins were refolded on the

affinity chromatography column by adding buffer (100mM NaCl-50mM Tris, pH 7.9) without urea to the column. Proteins were eluted using a gradient imidazole concentration from 0 mM to 500 mM. The protein containing fraction was dialyzed and further concentrate to the desired concentration.

pH titrations

The experimental setup involved the addition of a 100 mM H2SO4 solution in 15 µL increments, 26 times, to a cuvette containing 2 mL of each mutant and the wild type solution. The protein samples were excited at 295 nm, and the emitted light was captured between 300-400 nm using a PTI C-61 spectrofluorometer connected to a Hamilton titrator. However, as the signal was not strong enough at 2 µM, a separate circular dichroism titration was performed at 4 µM using a JASCO titrator. The pH at each step was calculated by referring to a standard curve (shown below) which was established by adding the same amount of acid to the buffer 26 times. Finally, a Mettler Toledo 751-4MM conductivity cell, 4 mm was used to measure the pH at each step.
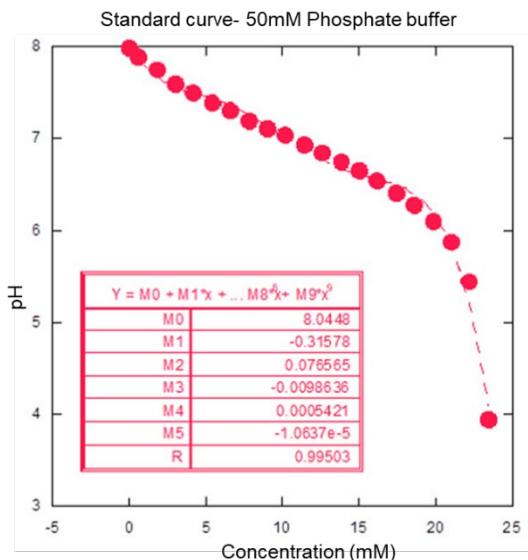
Standard curve- 50mM Phosphate buffer

$Y = M0 + M1^*x + \dots M8^*x^8 + M9^*x^9$	
M0	8.0448
M1	-0.31578
M2	0.076565
M3	-0.0098636
M4	0.0005421
M5	-1.0637e-5
R	0.99503

Limited trypsin proteolysis

Proteins (6 µM) were digested with 0.5 ng/µL of trypsin in a buffer of 50 mM potassium phosphate, pH 7, 1 mM DTT at 25 °C. On addition of trypsin, aliquots were withdrawn at prescribed time intervals, and reactions were inhibited by adding SDS-PAGE buffer and heating to 100 °C for 10 minutes. The samples were frozen at -20°C until analyzed. Samples were analyzed with 4-20% SDS-PAGE gradient gels.

References

1. Bose K, Pop C, Feeney B, Clark AC (2003) An Uncleavable Procaspase-3 Mutant Has a Lower Catalytic Efficiency but an Active Site Similar to That of Mature Caspase-3. Biochemistry. 42(42):12298-12310.

2. Shi Y (2002) Mechanisms of caspase activation and inhibition during apoptosis. Mol Cell 9:459–470.

3. Roy S, Bayly CI, Gareau Y, Houtzager VM, Kargman S, Keen SLC, Rowland K, Seiden IM, Thornberry NA, Nicholson DW (2001) Maintenance of caspase-3 proenzyme dormancy by an intrinsic "safety catch" regulatory tripeptide. Proc Natl Acad Sci USA 98:6132–6137.

4. Chai J, Wu Q, Shiozaki E, Srinivasula SM, Alnemri ES, Shi Y (2001) Crystal structure of a procaspase-7 zymogen: Mechanisms of activation and substrate binding. Cell 107:399–407.

5. Riedl SJ, Fuentes-Prior P, Renatus M, Kairies N, Krapp S, Huber R, Salvesen GS, Bode W (2001) Structural basis for the activation of human procaspase-7. Proc Natl Acad Sci USA 98:14790

6. Bose K, Clark ACC (2005) pH effects on the stability and dimerization of procaspase-3. Protein Science 14:24–36.

7. Shrestha S, Clark AC (2021) Evolution of the folding landscape of effector caspases. Journal of Biological Chemistry 297:101249.

8. Yao L, Clark AC (2022) Comparing the folding landscapes of evolutionarily divergent procaspase-3. Biosci Rep 42.

9. Joglekar I, Clark AC (2023) Sequential unfolding mechanisms of monomeric caspases. bioRxiv 2023.01.04.522771.

10. Nag M, Clark AC (2023) Conserved folding landscape of monomeric initiator caspases. Journal of Biological Chemistry ,103075.

11. Grimsley GR, Scholtz JM, Pace CN (2009) A summary of the measured pK values of the ionizable groups in folded proteins. Protein Sci 18:247–251.

12. Kukić, Farrell D, Søndergaard CR, Bjarnadottir U, Bradley J, Pollastri G, Nielsen JE (2010) Improving the analysis of NMR spectra tracking pH-induced conformational changes: removing artefacts of the electric field on the NMR chemical shift. Proteins 78:971–984.

13. Gunner MR, Mao J, Song Y, Kim J (2006) Factors influencing the energetics of electron and proton transfers in proteins. What can be learned from calculations. Biochimica et Biophysica Acta (BBA) - Bioenergetics 1757:942–968.

14. Srivastava J, Barber DL, Jacobson MP (2007) Intracellular pH sensors: Design principles and functional significance. Physiology 22:30–39.

15. Zhou H-X, Pang X (2018) Electrostatic Interactions in Protein Structure, Folding, Binding, and Condensation.

16. Hoffman L, Stein RA, Colbran RJ, McHaourab HS (2011) Conformational changes underlying calcium/calmodulin-dependent protein kinase II activation. EMBO J 30:1251–1262.

17. Di Russo N V., Estrin DA, Martí MA, Roitberg AE (2012) pH-Dependent Conformational Changes in Proteins and Their Effect on Experimental pKas: The Case of Nitrophorin 4. PLoS Comput Biol 8:1002761.

Figure 1
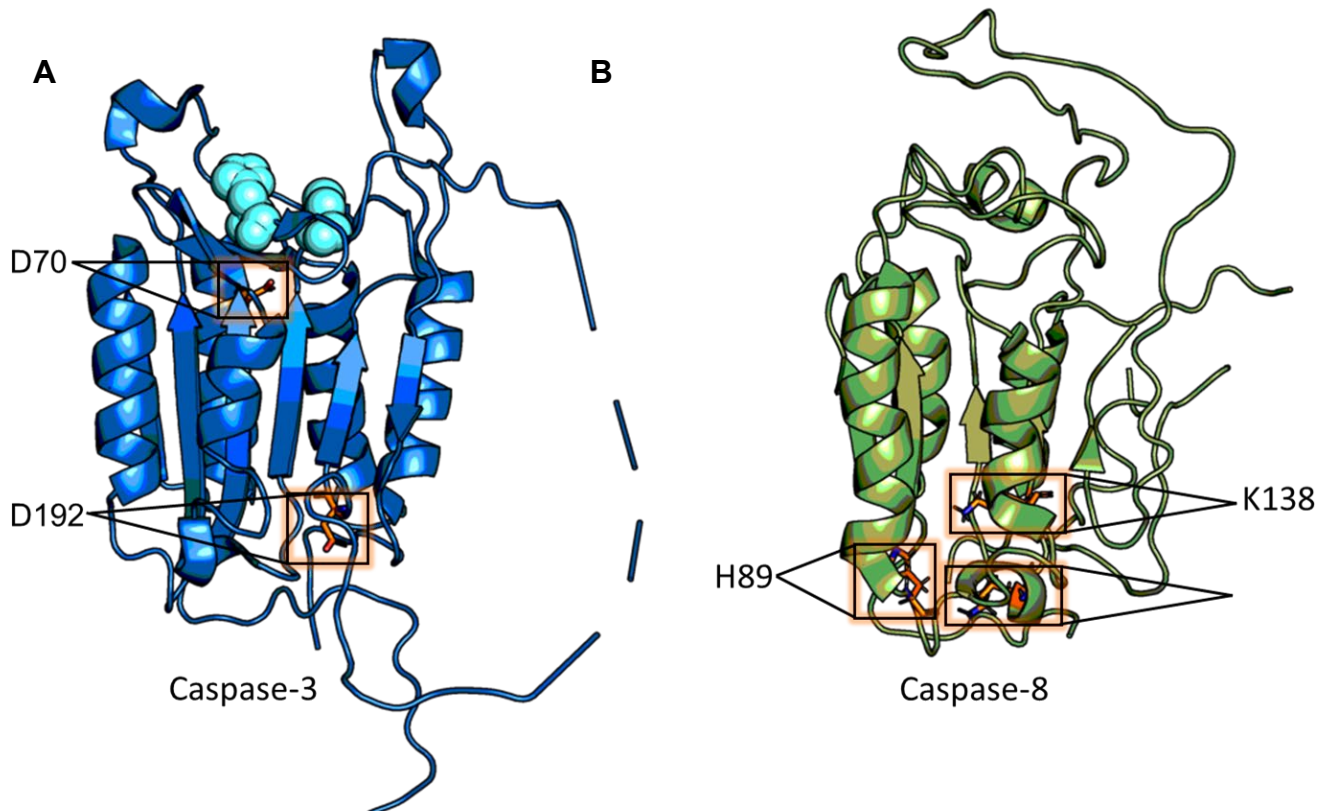
Figure 1. Identification of highly conserved residues potentially contributing to the pH-dependent conformational change observed in all caspases. (A) Two highly conserved sites mapped onto PDB ID: 2j30 and mutated to alanine- D70A and D192A in caspase-3. The active site residues are shown as cyan spheres. (B) Three highly conserved sites mapped onto PDB ID: 2k7z and mutated to alanine- K136/138A and H89A in casp-8.

Figure 2
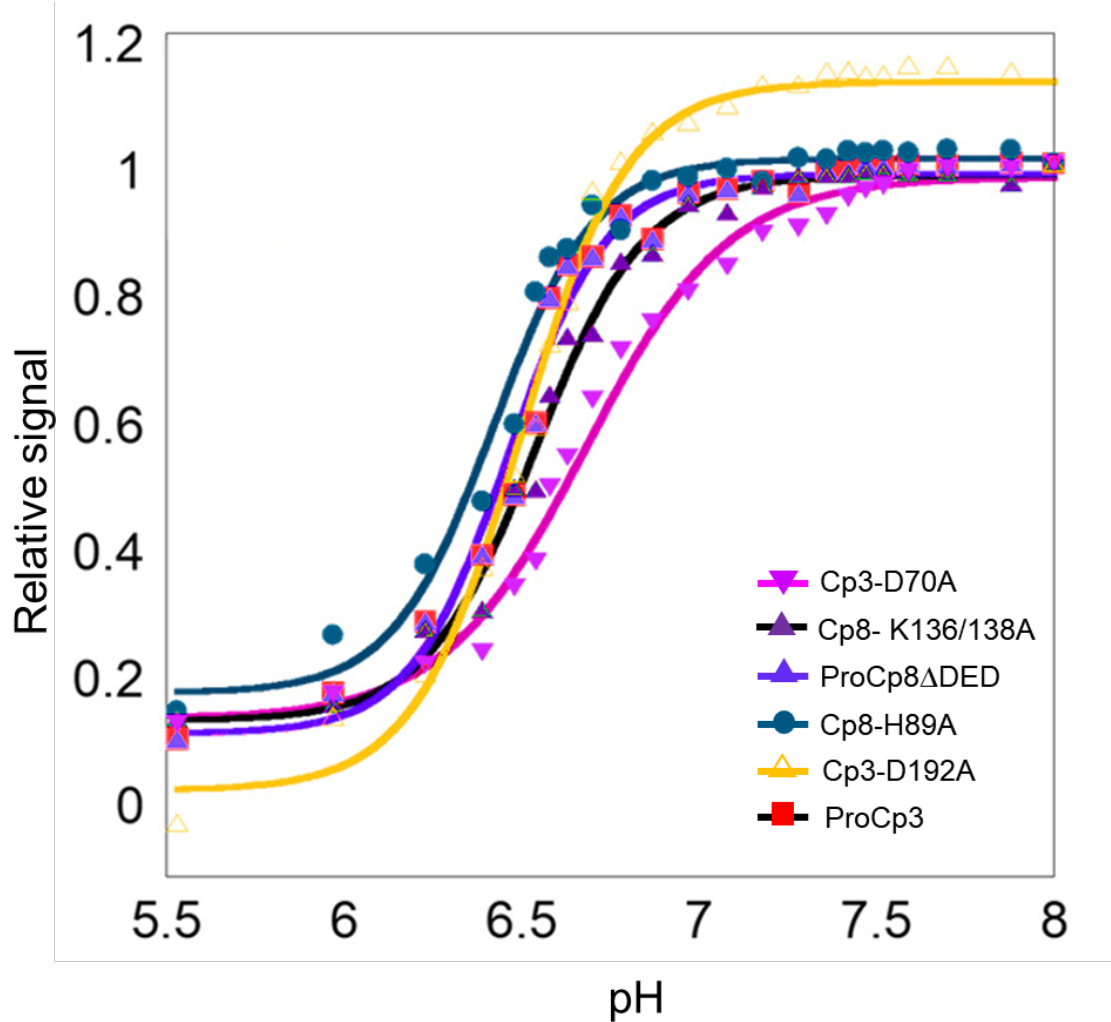
Figure 2. pH titration curves of caspase3-D70A, caspase8-D136/138A, procaspase8ΔDED, caspase8-H89A, caspase3-D192A, procaspase-3.

Figure 3

$$\langle\lambda\rangle = \frac{(A + \Delta\langle\lambda_{(1)}\rangle) \times 10^{n1(pH-pKa)}}{1+10^{n1(pH-pKa)}}$$
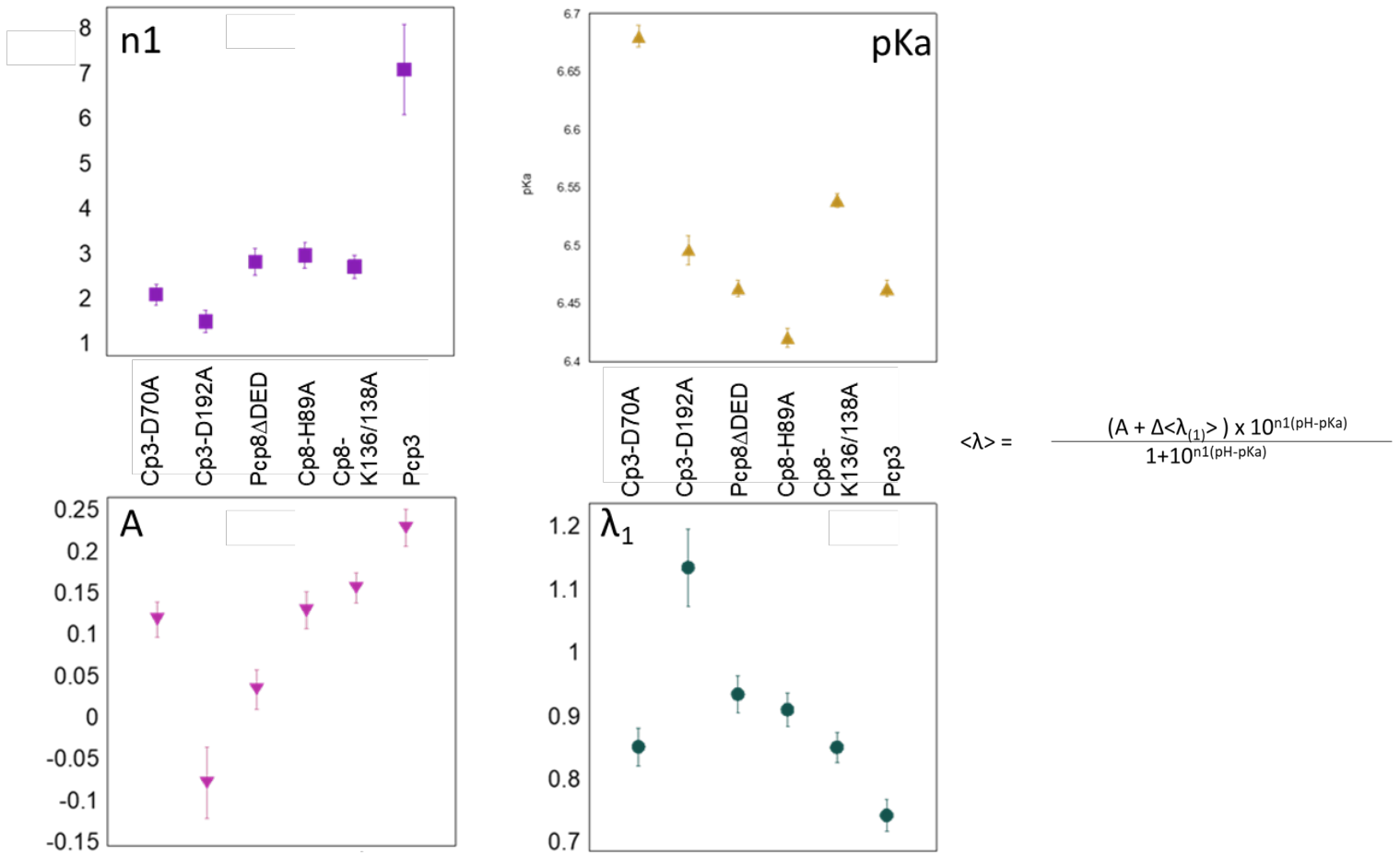
Figure 3. Fitting parameters from the equation obtained for all caspase mutants and wild type.

Figure 4
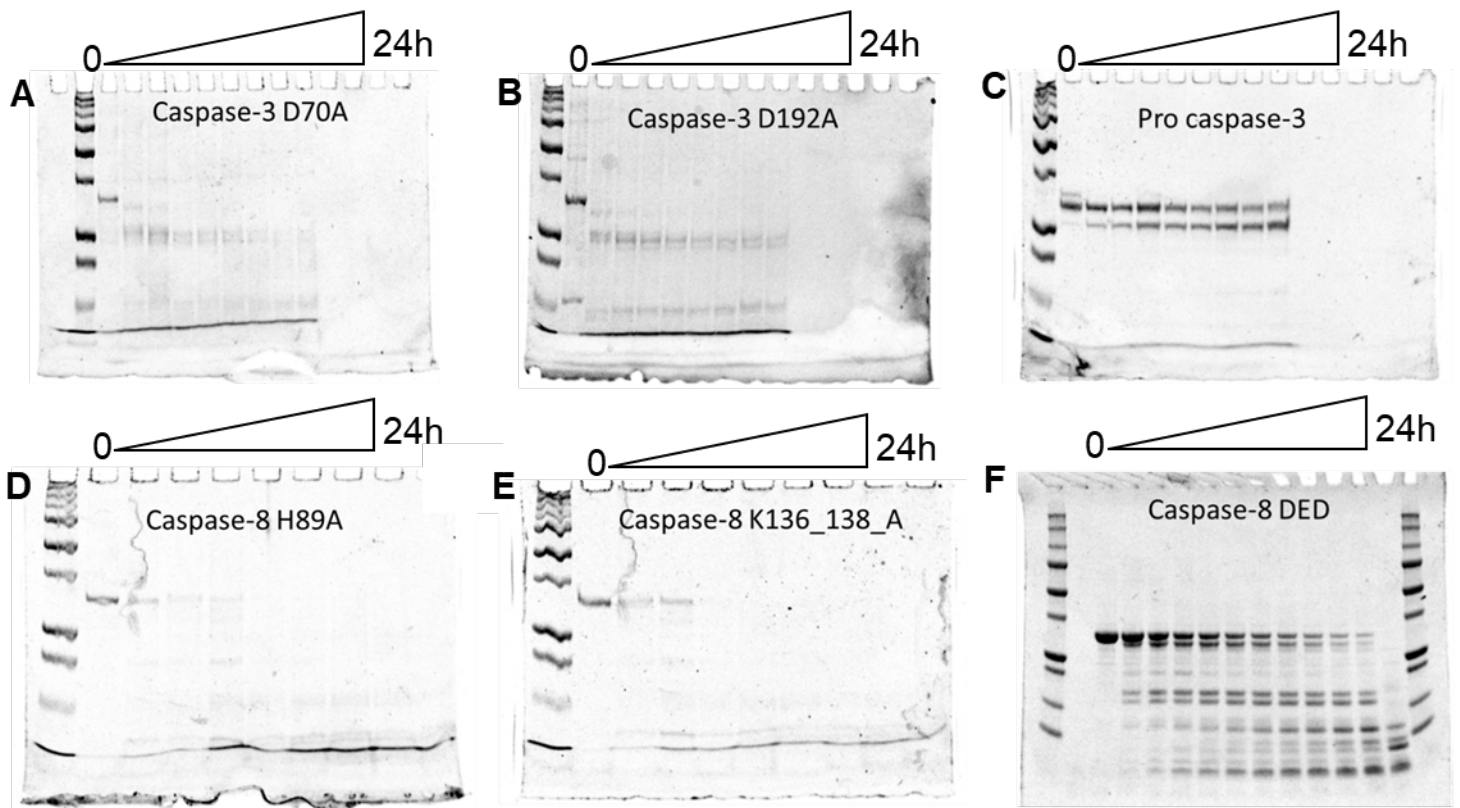
Figure 4. Limited trypsin proteolysis of WT ProCp-3/-8 and mutants. (A)Caspase3-D70A

(B) Caspase3-D192A (C) Procaspase-3 (D) Caspase8-H89A (E) Caspase8-K136/138A

(F)Procaspase8ΔDED
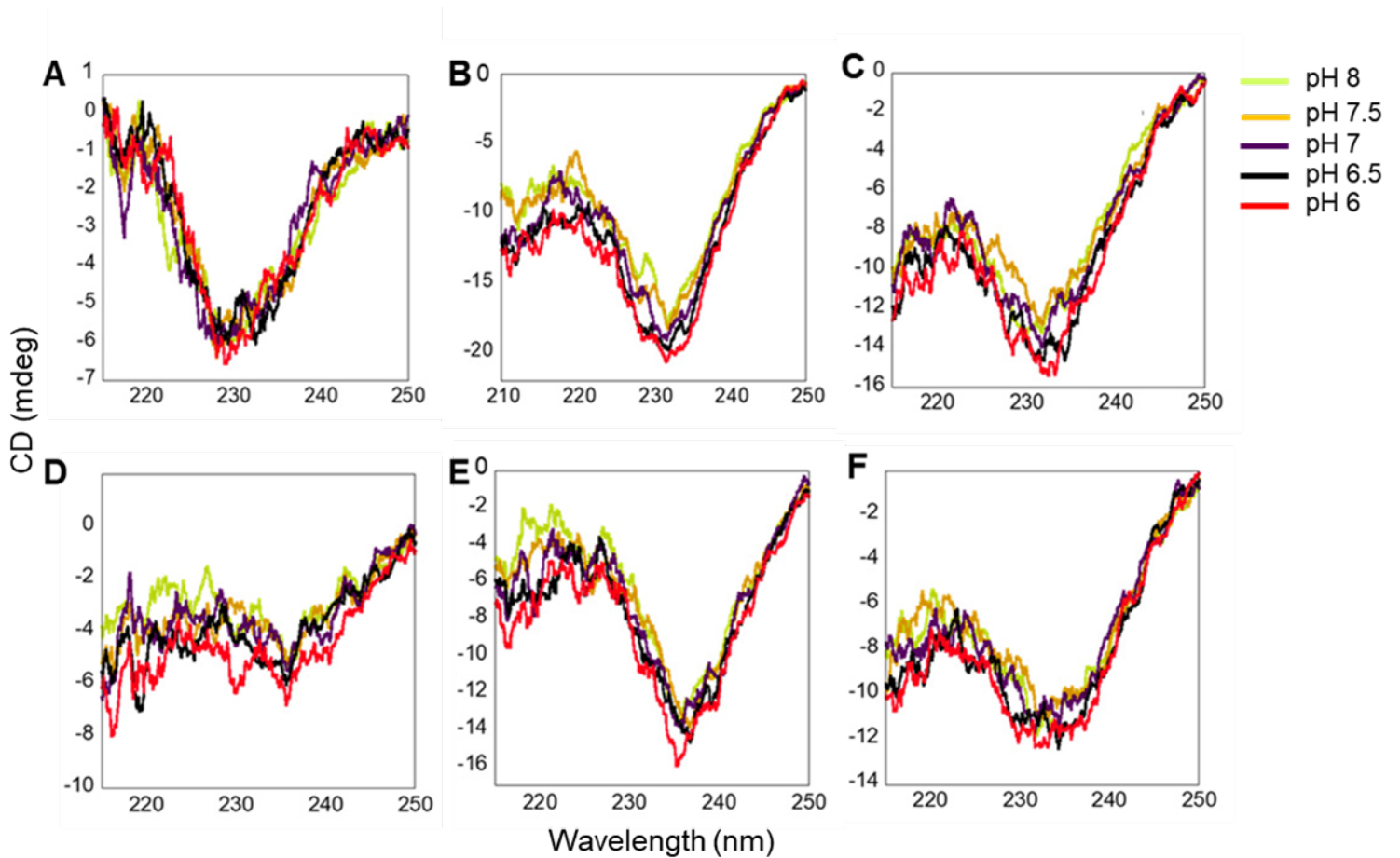
Figure 5

Figure 5. CD spectra for pH titration using 50mM phosphate buffer. (A) Caspase3-D70A

(B) Caspase3-D192A (C) Procaspase-3 (D) Caspase8-H89A (E) Caspase8-K136/138A

(F)Procaspase8ΔDED

Appendix

Igor Pro 2-state fit procedure:

```
Function IIstatemonomer(w,x) : FitFunc
	Wave w
	Variable x

	//CurveFitDialog/ These comments were created by the Curve Fitting dialog. Altering them will
	//CurveFitDialog/ make the function less convenient to work with in the Curve Fitting dialog.
	//CurveFitDialog/ Equation:
	//CurveFitDialog/ Variable  K=EXP(-((w_0+w_1*x)/(0.001987*298)))
	//CurveFitDialog/ Variable  Fu=(K/(1+K))
	//CurveFitDialog/ Variable  Fd=(1/(1+K))
	//CurveFitDialog/ Variable  Yd=w_2+w_3*x
	//CurveFitDialog/ Variable  Yu=w_4+w_5*x
	//CurveFitDialog/ Variable  y=Yd*Fd+Yu*Fu
	//CurveFitDialog/
	//CurveFitDialog/ f(x) =	y
	//CurveFitDialog/
	//CurveFitDialog/
	//CurveFitDialog/ End of Equation
	//CurveFitDialog/ Independent Variables 1
	//CurveFitDialog/ x
	//CurveFitDialog/ Coefficients 6
	//CurveFitDialog/ w[0] = w_0
	//CurveFitDialog/ w[1] = w_1
	//CurveFitDialog/ w[2] = w_2
	//CurveFitDialog/ w[3] = w_3
```

```
        //CurveFitDialog/ w[4] = w_4
    //CurveFitDialog/ w[5] = w_5



        Variable      K=EXP(-((w[0]+w[1]*x)/(0.001987*298)))
        Variable      Fu=(K/(1+K))
        Variable      Fd=(1/(1+K))
        Variable      Yd=w[2]+w[3]*x
        Variable      Yu=w[4]+w[5]*x
        Variable      y=Yd*Fd+Yu*Fu
        return y
End
```

Igor Pro 3-state fit procedure:

```
Function IIIstatemonomer(w,x) : FitFunc
        Wave w
        Variable x
        //CurveFitDialog/ These comments were created by the Curve Fitting dialog.
Altering them will
        //CurveFitDialog/ make the function less convenient to work with in the Curve
Fitting dialog.
        //CurveFitDialog/ Equation:
        //CurveFitDialog/ Variable  K1=EXP(-((w_0+w_1*x)/(0.001987*298)))
        //CurveFitDialog/ Variable  K2=EXP(-((w_2+w_3*x)/(0.001987*298)))
        //CurveFitDialog/ Variable  Fu=((K1*K2)/(1+K1+K1*K2))
        //CurveFitDialog/ Variable  Fi=(K1/(1+K1+K1*K2))
        //CurveFitDialog/ Variable  Fd=(1/(1+K1+K1*K2))
        //CurveFitDialog/ Variable  Yd=w_4+w_5*x
        //CurveFitDialog/ Variable  Yu=w_6+w_7*x
        //CurveFitDialog/ Variable  y=Yd*Fd+w_8*Fi+Yu*Fu
```

//CurveFitDialog/

//CurveFitDialog/ f(x) =   y

//CurveFitDialog/

//CurveFitDialog/

//CurveFitDialog/ End of Equation

//CurveFitDialog/ Independent Variables 1

//CurveFitDialog/ x

//CurveFitDialog/ Coefficients 9

//CurveFitDialog/ w[0] = w_0

//CurveFitDialog/ w[1] = w_1

//CurveFitDialog/ w[2] = w_2

//CurveFitDialog/ w[3] = w_3

//CurveFitDialog/ w[4] = w_4

//CurveFitDialog/ w[5] = w_5

//CurveFitDialog/ w[6] = w_6

//CurveFitDialog/ w[7] = w_7

//CurveFitDialog/ w[8] = w_8

Variable   K1=EXP(-((w[0]+w[1]*x)/(0.001987*298)))

Variable   K2=EXP(-((w[2]+w[3]*x)/(0.001987*298)))

Variable   Fu=((K1*K2)/(1+K1+K1*K2))

Variable   Fi=(K1/(1+K1+K1*K2))

Variable   Fd=(1/(1+K1+K1*K2))

Variable   Yd=w[4]+w[5]*x

Variable   Yu=w[6]+w[7]*x

Variable   y=Yd*Fd+w[8]*Fi+Yu*Fu

return y

End

Igor Pro 4-state fit procedure

Function IVstatemonomer(w,x) : FitFunc

       Wave w

       Variable x

       //CurveFitDialog/ These comments were created by the Curve Fitting dialog. Altering them will

       //CurveFitDialog/ make the function less convenient to work with in the Curve Fitting dialog.

       //CurveFitDialog/ Equation:

       //CurveFitDialog/ Variable K1=EXP(-((w_0+w_1*x)/(0.001987*298)))

       //CurveFitDialog/ Variable K2=EXP(-((w_2+w_3*x)/(0.001987*298)))

       //CurveFitDialog/ Variable   K3=EXP(-((w[4]+w[5]*x)/(0.001987*298)))

       //CurveFitDialog/ Variable Fn=((1)/(1+K1+K1*K2+K1*K2*K3))

       //CurveFitDialog/ Variable Fa=((K1)/(1+K1+K1*K2+K1*K2*K3))

       //CurveFitDialog/ Variable   Fb=((K1*K2)/(1+K1+K1*K2+K1*K2*K3))

       //CurveFitDialog/ Variable Fu=((K1*K2*K3)/(1+K1+K1*K2+K1*K2*K3))

       //CurveFitDialog/ Variable Yn=w[6]+w[7]*x

       //CurveFitDialog/ Variable Yu=w[8]+w[9]*x

       //CurveFitDialog/ Variable y=Yu*Fn+w[10]*Fa+w[11]*Fb+Yu*Fu

       //CurveFitDialog/

       //CurveFitDialog/ f(x) =   y

       //CurveFitDialog/

       //CurveFitDialog/

       //CurveFitDialog/ End of Equation

       //CurveFitDialog/ Independent Variables 1

       //CurveFitDialog/ x

       //CurveFitDialog/ Coefficients 12

       //CurveFitDialog/ w[0] = w_0

       //CurveFitDialog/ w[1] = w_1

       //CurveFitDialog/ w[2] = w_2

```
//CurveFitDialog/ w[3] = w_3
//CurveFitDialog/ w[4] = w_4
//CurveFitDialog/ w[5] = w_5
//CurveFitDialog/ w[6] = w_6
//CurveFitDialog/ w[7] = w_7
//CurveFitDialog/ w[8] = w_8
//CurveFitDialog/ w[9] = w_9
//CurveFitDialog/ w[10] = w_10
//CurveFitDialog/ w[11] = w_11


Variable      K1=EXP(-((w[0]+w[1]*x)/(0.001987*298)))
Variable      K2=EXP(-((w[2]+w[3]*x)/(0.001987*298)))
Variable  K3=EXP(-((w[4]+w[5]*x)/(0.001987*298)))
Variable      Fn=((1)/(1+K1+K1*K2+K1*K2*K3))
Variable      Fa=((K1)/(1+K1+K1*K2+K1*K2*K3))
Variable      Fb=((K1*K2)/(1+K1+K1*K2+K1*K2*K3))
Variable      Fu=((K1*K2*K3)/(1+K1+K1*K2+K1*K2*K3))
Variable      Yn=w[6]+w[7]*x
Variable      Yu=w[8]+w[9]*x
Variable      y=Yn*Fn+w[10]*Fa+w[11]*Fb+Yu*Fu


return y


End
```

Molecular dynamics (MD) simulations using GROMACS

Commands for simulations in water and urea

gmx pdb2gmx -f pdbname.pdb -o protein.gro -p protein.top

gmx editconf -f protein.gro -d 1.0 -o protein_box.gro

gmx insert-molecules -f protein_box.gro -ci urea_original1.pdb -o box_prot_urea.gro -nmol 560

gmx solvate -cp box_prot_urea.gro -cs spc216.gro -o b4em.gro -p protein.top

(add the following section after position restrain and before water)

; Include urea topology

#include "urea_comp_simulation_of_urea_paper_G96BOND.itp"

#ifdef POSRES_URE

; Position restraint for each urea molecule

[position_restraints]

; i	funct	fcx	fcy	fcz
1	1	1000	1000	1000
2	1	1000	1000	1000
3	1	1000	1000	1000
6	1	1000	1000	1000

#endif

gmx grompp -f grompp1.mdp -c b4em.gro -p protein.top -o em.tpr

gmx genion -s em.tpr -o b4em.gro -pname NA -np 12(change) -pq 1

  -- Choose Group 13 (solvent)

  -- Change protein.top using VI

    -- remove equal number of water

    -- add appropriate number of NA (or CL) to match genion

  -- Re-run grompp

gmx mdrun -s em.tpr -o em.trr -c b4NVT.gro -g em.log -e em.edr

gmx energy -f em.edr -o potential.xvg

xmgrace potential.xvg

gmx grompp -f New_NVT.mdp -c b4NVT.gro -r b4NVT.gro -p protein.top -o NVT.tpr

gmx mdrun -s NVT.tpr -o NVT.trr -c b4NPT.gro -g NVT.log -e NVT.edr

gmx energy -f NVT.edr -o temperature.xvg

xmgrace temperature.xvg

gmx grompp -f New_NPT.mdp -c b4NPT.gro -r b4NPT.gro -p protein.top -o NPT.tpr

gmx mdrun -s NPT.tpr -o NPT.trr -c b4md.gro -g NPT.log -e NPT.edr

gmx energy -f NPT.edr -o Pressure.xvg

xmgrace Pressure.xvg

gmx energy -f NPT.edr -o Density.xvg

xmgrace Density.xvg

gmx grompp -f Run.mdp -c b4md.gro -r b4md.gro -p protein.top -o md.tpr

gmx mdrun -s md.tpr -o md.trr -c md.gro -g md.log -e md.edr &

tail -f md.log

gmx trjconv -f md.trr -s md.tpr -skip 50 -center -pbc nojump -o protein_out.pdb

;calculate reduced trajectory file for following calculations

gmx trjconv -f md.trr -s md.tpr -o md_reduced.trr -center -pbc nojump

;calculate radius of gyration

gmx gyrate -f md_reduced.trr -s md.tpr -o protein_radgyration.xvg

;calculate rmsf, average structure, and B-factors on avg structure

gmx rmsf -f md_reduced.trr -s md.tpr -o md_rmsf.xvg -ox md_avg.pdb -oq md_bfactor.pdb -res

;calculate internal H-bonds

gmx hbond -f md_reduced.trr -s md.tpr -num md_hbond_internal.xvg

;calculate H-bonds with water

gmx hbond -f md.trr -s md.tpr -num md_hbond_solvent.xvg

Add the following details to a file named:

'urea_comp_simulation_of_urea_paper_G96BOND.itp' while performing simulations in urea so that gromacs can use these coordinates in the simulation.

```
[ moleculetype ]
; molname    nrexcl
URE          3


[ atoms ]
    1 C        1 URE      C     1    0  12.01000     ; amber C  type
    2 O        1 URE      O     2    0  16.00000     ; amber O  type
    3 N        1 URE     N1     3    0  14.01000    ; amber N  type
    4 H        1 URE    H11     4    0   1.00800    ; amber H  type
    5 H        1 URE    H12     5    0   1.00800    ; amber H  type
    6 N        1 URE     N2     6    0  14.01000    ; amber N  type
    7 H        1 URE    H21     7    0   1.00800    ; amber H  type
    8 H        1 URE    H22     8    0   1.00800    ; amber H  type


[ bonds ]
; ai aj fu b0 kb, b0 kb
 4   3 2 0.1000  18700000  0.1000  18700000
 8   6 2 0.1000  18700000  0.1000  18700000
 3   5 2 0.1000  18700000  0.1000  18700000
 3   1 2 0.1350  10300000  0.1350  10300000
 6   1 2 0.1350  10300000  0.1350  10300000
 6   7 2 0.1000  18700000  0.1000  18700000
 1   2 2 0.1265  13100000  0.1265  13100000


[ pairs ]
; ai aj fu
 2   4 1
 2   5 1
 2   7 1
 2   8 1
```

```
 3  7 1
 3  8 1
 4  6 1
 5  6 1


[ angles ]
; ai aj ak fu th0 kth ub0 kub th0 kth ub0 kub
 2  1  3 1 121.4  690    121.4  690
 2  1  6 1 121.4  690    121.4  690
 3  1  6 1 117.2  636    117.2  636
 1  3  4 1 120.0  390    120.0  390
 1  3  5 1 120.0  390    120.0  390
 4  3  5 1 120.0  445    120.0  445
 1  6  7 1 120.0  390    120.0  390
 1  6  8 1 120.0  390    120.0  390
 7  6  8 1 115.6300  261.96    115.6300  261.96


[ dihedrals ]
; ai aj ak al fu cos(sigma) m kphi
 2  1  3  4 11  -1.0  2  41.8  -1.0  2  41.8
 2  1  6  7 11  -1.0  2  41.8  -1.0  2  41.8


[ dihedrals ]
; ai aj ak al fu xi0 kxi xi0 kxi
 1  6  3  2 2  0.00  510    0.00  510
 3  4  1  5 2  0.00  510    0.00  510
 6  8  1  7 2  0.00  510    0.00  510
```

Additional files for energy minimization step in MD simulations

```
; minim.mdp - used as input into grompp to generate em.tpr
; Parameters describing what to do, when to stop and what to save
integrator  = steep         ; Algorithm (steep = steepest descent minimization)
emtol       = 1000.0        ; Stop minimization when the maximum force < 1000.0
kJ/mol/nm
emstep      = 0.01          ; Minimization step size
nsteps      = 50000         ; Maximum number of (minimization) steps to perform

; Parameters describing how to find the neighbors of each atom and how to calculate
the interactions
nstlist         = 1         ; Frequency to update the neighbor list and long range forces
cutoff-scheme   = Verlet    ; Buffered neighbor searching
ns_type         = grid      ; Method to determine neighbor list (simple, grid)
coulombtype     = PME       ; Treatment of long range electrostatic interactions
rcoulomb        = 1.0       ; Short-range electrostatic cut-off
rvdw            = 1.0       ; Short-range Van der Waals cut-off
pbc             = xyz       ; Periodic Boundary Conditions in all 3 dimensions
```

NVT file:
```
title               = OPLS Lysozyme NVT equilibration
define              = -DPOSRES  ; position restrain the protein
; Run parameters
integrator          = md        ; leap-frog integrator
nsteps              = 50000     ; 2 * 50000 = 100 ps
dt                  = 0.002     ; 2 fs
; Output control
nstxout             = 500       ; save coordinates every 1.0 ps
nstvout             = 500       ; save velocities every 1.0 ps
nstenergy           = 500       ; save energies every 1.0 ps
```

```
nstlog              = 500       ; update log file every 1.0 ps
; Bond parameters
continuation        = no        ; first dynamics run
constraint_algorithm = lincs    ; holonomic constraints
constraints         = h-bonds   ; bonds involving H are constrained
lincs_iter          = 1         ; accuracy of LINCS
lincs_order         = 4         ; also related to accuracy
; Nonbonded settings
cutoff-scheme       = Verlet    ; Buffered neighbor searching
ns_type             = grid      ; search neighboring grid cells
nstlist             = 10        ; 20 fs, largely irrelevant with Verlet
rcoulomb            = 1.0       ; short-range electrostatic cutoff (in nm)
rvdw                = 1.0       ; short-range van der Waals cutoff (in nm)
DispCorr            = EnerPres  ; account for cut-off vdW scheme
; Electrostatics
coulombtype         = PME       ; Particle Mesh Ewald for long-range electrostatics
pme_order           = 4         ; cubic interpolation
fourierspacing      = 0.16      ; grid spacing for FFT
; Temperature coupling is on
tcoupl              = V-rescale           ; modified Berendsen thermostat
tc-grps             = Protein Non-Protein ; two coupling groups - more accurate
tau_t               = 0.1   0.1           ; time constant, in ps
ref_t               = 300   300           ; reference temperature, one for each group, in K
; Pressure coupling is off
pcoupl              = no        ; no pressure coupling in NVT
; Periodic boundary conditions
pbc                 = xyz       ; 3-D PBC
; Velocity generation
gen_vel             = yes       ; assign velocities from Maxwell distribution
gen_temp            = 300       ; temperature for Maxwell distribution
gen_seed            = -1        ; generate a random seed
```

NPT file:

```
title                = OPLS Lysozyme NPT equilibration
define               = -DPOSRES  ; position restrain the protein
; Run parameters
integrator           = md        ; leap-frog integrator
nsteps               = 50000     ; 2 * 50000 = 100 ps
dt                   = 0.002     ; 2 fs
; Output control
nstxout              = 500       ; save coordinates every 1.0 ps
nstvout              = 500       ; save velocities every 1.0 ps
nstenergy            = 500       ; save energies every 1.0 ps
nstlog               = 500       ; update log file every 1.0 ps
; Bond parameters
continuation         = yes       ; Restarting after NVT
constraint_algorithm = lincs     ; holonomic constraints
constraints          = h-bonds   ; bonds involving H are constrained
lincs_iter           = 1         ; accuracy of LINCS
lincs_order          = 4         ; also related to accuracy
; Nonbonded settings
cutoff-scheme        = Verlet    ; Buffered neighbor searching
ns_type              = grid      ; search neighboring grid cells
nstlist              = 10        ; 20 fs, largely irrelevant with Verlet scheme
rcoulomb             = 1.0       ; short-range electrostatic cutoff (in nm)
rvdw                 = 1.0       ; short-range van der Waals cutoff (in nm)
DispCorr             = EnerPres  ; account for cut-off vdW scheme
; Electrostatics
coulombtype          = PME       ; Particle Mesh Ewald for long-range electrostatics
pme_order            = 4         ; cubic interpolation
fourierspacing       = 0.16      ; grid spacing for FFT
; Temperature coupling is on
```

tcoupl   = V-rescale   ; modified Berendsen thermostat

tc-grps   = Protein Non-Protein   ; two coupling groups - more accurate

tau_t   = 0.1   0.1   ; time constant, in ps

ref_t   = 300   300   ; reference temperature, one for each group, in K

; Pressure coupling is on

pcoupl   = Parrinello-Rahman   ; Pressure coupling on in NPT

pcoupltype   = isotropic   ; uniform scaling of box vectors

tau_p   = 2.0   ; time constant, in ps

ref_p   = 1.0   ; reference pressure, in bar

compressibility   = 4.5e-5   ; isothermal compressibility of water, bar^-1

refcoord_scaling   = com

; Periodic boundary conditions

pbc   = xyz   ; 3-D PBC

; Velocity generation

gen_vel   = no   ; Velocity generation is off

Run file:

title   = UREA RUN

; Run parameters

integrator   = md   ; leap-frog integrator

nsteps   = 100000000 ; 20 * 500000 = 1000 ps (200 ns)

dt   = 0.002   ; 2 fs

; Output control

nstxout   = 2500   ; output coordinates every 5 ps

nstvout   = 0   ; 0 for output frequency of nstxout,

nstfout   = 0   ; nstvout, and nstfout

nstenergy   = 5000   ; save energies every 10.0 ps

nstlog   = 5000   ; update log file every 10.0 ps

; Bond parameters

continuation   = yes   ; Restarting after NPT

constraint_algorithm   = lincs   ; holonomic constraints

```
constraints          = h-bonds   ; bonds involving H are constrained
lincs_iter           = 1         ; accuracy of LINCS
lincs_order          = 4         ; also related to accuracy
; Neighborsearching
cutoff-scheme        = Verlet    ; Buffered neighbor searching
ns_type              = grid      ; search neighboring grid cells
nstlist              = 10        ; 20 fs, largely irrelevant with Verlet scheme
rcoulomb             = 0.9       ; short-range electrostatic cutoff (in nm)
rvdw                 = 0.9       ; short-range van der Waals cutoff (in nm)
; Electrostatics
coulombtype          = PME       ; Particle Mesh Ewald for long-range electrostatics
pme_order            = 4         ; cubic interpolation
fourierspacing       = 0.12      ; grid spacing for FFT
; Temperature coupling is on
tcoupl               = Nose-Hoover        ; more accurate thermostat
tc-grps              = Protein Non-Protein   ; two coupling groups - more accurate
tau_t                = 0.5    0.5          ; time constant, in ps
ref_t                = 300    300          ; reference temperature, one for each group, in K
; Pressure coupling is on
pcoupl               = Parrinello-Rahman     ; Pressure coupling on in NPT
pcoupltype           = isotropic            ; uniform scaling of box vectors
tau_p                = 1.0                  ; time constant, in ps
ref_p                = 1.0                  ; reference pressure, in bar
compressibility      = 4.5e-5              ; isothermal compressibility of water, bar^-1
; Periodic boundary conditions
pbc                  = xyz       ; 3-D PBC
; Dispersion correction
DispCorr             = EnerPres  ; account for cut-off vdW scheme
; Velocity generation
gen_vel              = no        ; Velocity generation is off
```

REMD mdp file:

```
title               = PaCasp7a in urea run
; Run parameters
integrator          = md        ; leap-frog integrator
nsteps              = 5000000  ; 2 * 500000 = 1000 ps (1 ns) (from lysozyme in water)
Heres its 500000 * 10 for 10ns
dt                  = 0.002     ; 2 fs
; Output control
nstxout             = 2500      ; output coordinates every 5 ps
nstvout             = 0         ; 0 for output frequency of nstxout,
nstfout             = 0         ; nstvout, and nstfout
nstenergy           = 5000      ; save energies every 10.0 ps
nstlog              = 5000      ; update log file every 10.0 ps
; Bond parameters
continuation        = yes       ; Restarting after NPT
constraint_algorithm = lincs    ; holonomic constraints
constraints         = h-bonds   ; bonds involving H are constrained
lincs_iter          = 1         ; accuracy of LINCS
lincs_order         = 4         ; also related to accuracy
; Neighborsearching
cutoff-scheme       = Verlet    ; Buffered neighbor searching
ns_type             = grid      ; search neighboring grid cells
nstlist             = 10        ; 20 fs, largely irrelevant with Verlet scheme
rcoulomb            = 0.9       ; short-range electrostatic cutoff (in nm)
rvdw                = 0.9       ; short-range van der Waals cutoff (in nm)
; Electrostatics
coulombtype         = PME       ; Particle Mesh Ewald for long-range electrostatics
pme_order           = 4         ; cubic interpolation
fourierspacing      = 0.12      ; grid spacing for FFT
; Temperature coupling is on
tcoupl              = Nose-Hoover           ; modified Berendsen thermostat
```

tc-grps   = Protein Non-Protein   ; two coupling groups - more accurate

tau_t   = 0.5   0.5   ; time constant, in ps

ref_t   = TEMP   TEMP   ; reference temperature, one for each group, in K

; Pressure coupling is on

pcoupl   = Parrinello-Rahman   ; Pressure coupling on in NPT

pcoupltype   = isotropic   ; uniform scaling of box vectors

tau_p   = 1.0   ; time constant, in ps

ref_p   = 1.0   ; reference pressure, in bar

compressibility   = 4.5e-5   ; isothermal compressibility of water, bar^-1

; Periodic boundary conditions

pbc   = xyz   ; 3-D PBC

; Dispersion correction

DispCorr   = EnerPres ; account for cut-off vdW scheme

; Velocity generation

gen_vel   = no   ; Velocity generation is off

Commands for generating free energy landscapes on Gromacs using in-built Gromacs tools

gmx trjconv -f md.trr -o md.xtc

gmx covar -f md.xtc -s md.tpr -o eigenval.xvg -v eigenvect.trr -xpm covara.xpm

choose (protein for every atom)

gmx xpm2ps -f covara.xpm -o covara.eps -do covar.m2p

gmx anaeig -v eigenvect.trr -f md.xtc -s md.tpr -first 1 -last 2 -proj proj_eig.xvg -2d 2d_proj.xvg

gmx sham -f 2d_proj.xvg -ls gibbs.xpm -notime

gmx xpm2ps -f gibbs.xpm -o gibbs.eps -rainbow red

python xpm2txt.py -f gibbs.xpm -o gibbs.txt

python fel3d.py -f gibbs.tx

Gromacs commands for running MD simulations on UTA HPC

Note: These are slurm commands and the partition can vary depending on the run time, either long or normal partition.

```
#!/bin/bash

#SBATCH --job-name=Cp3_1us

#SBATCH --partition=long

#SBATCH --time=12-00:00:00

#SBATCH -e slurm-%j.err

#SBATCH --mail-user=ishauday.joglekar@mavs.uta.edu

#SBATCH --nodes=1

export LD_LIBRARY_PATH=/home/joglekari/Paper3/1us_Cp3/D192A


module load mpi/2021.3.0

module load libfabric/1.13.0

module load  mkl/2021.3.0

module load gromacs/2020.6


gmx_mpi grompp -f em.mdp -c b4em.gro -p protein.top -o em.tpr

gmx_mpi mdrun -s em.tpr -o em.trr -c b4nvt.gro -g em.log -e em.edr

gmx_mpi grompp -f New_NVT.mdp -c b4nvt.gro -r b4nvt.gro -p protein.top -o nvt.tpr

gmx_mpi mdrun -s nvt.tpr -o nvt.trr -c b4npt.gro -g nvt.log -e nvt.edr

gmx_mpi grompp -f New_NPT.mdp -c b4npt.gro -r b4npt.gro -p protein.top -o npt.tpr

gmx_mpi mdrun -s npt.tpr -o npt.trr -c b4md.gro -g npt.log -e npt.edr

gmx_mpi grompp -f Run.mdp -c b4md.gro -r b4md.gro -p protein.top -o
D_WTCasp3_md.tpr

gmx_mpi mdrun -s D_WTCasp3_md.tpr -o D_WTCasp3_md.trr -c D_WTCasp3_md.gro
-g D_WTCasp3_md.log -e D_WTCasp3_md.edr

echo 1 1|gmx_mpi trjconv -f D_WTCasp3_md.trr -s D_WTCasp3_md.tpr -skip 50 -
center -pbc nojump -o D_WTCasp3_out.pdb
```

echo 1 1|gmx_mpi trjconv -f D_WTCasp3_md.trr -s D_WTCasp3_md.tpr -o md_red_D_WTCasp3.trr -center -pbc nojump

echo 1|gmx_mpi gyrate -f md_red_D_WTCasp3.trr -s D_WTCasp3_md.tpr -o D_WTCasp3_radgyration.xvg

echo 1|gmx_mpi rmsf -f md_red_D_WTCasp3.trr -s D_WTCasp3_md.tpr -o D_WTCasp3_rmsf.xvg -ox D_WTCasp3_md_avg.pdb -oq D_WTCasp3_bfactor.pdb -res

gmx_mpi trjconv -f D_WTCasp3_md.trr -o D_WTCasp3_md.xtc

echo 1 1|gmx_mpi covar -f D_WTCasp3_md.xtc -s D_WTCasp3_md.tpr -o D_WTCasp3_eigenval.xvg -v D_WTCasp3_eigenvect.trr -xpm D_WTCasp3_covar.xpm

gmx_mpi xpm2ps -f D_WTCasp3_covar.xpm -o D_WTCasp3_covar.eps -do D_WTCasp3_covar.m2p

echo 1 1|gmx_mpi anaeig -v D_WTCasp3_eigenvect.trr -f D_WTCasp3_md.xtc -s D_WTCasp3_md.tpr -first 1 -last 2 -proj D_WTCasp3_proj_eig.xvg -2d D_WTCasp3_2d_proj.xvg

gmx_mpi sham -f D_WTCasp3_2d_proj.xvg -ls D_WTCasp3_gibbs.xpm -notime

gmx_mpi xpm2ps -f D_WTCasp3_gibbs.xpm -o D_WTCasp3_gibbs.eps -rainbow red

echo 1 1|gmx_mpi anaeig -v D_WTCasp3_eigenvect.trr -s D_WTCasp3_md.tpr -f D_WTCasp3_md.xtc -extr D_WTCasp3_ext1.pdb -first 1 -last 1 -nframes 400

echo 1 1|gmx_mpi anaeig -v D_WTCasp3_eigenvect.trr -s D_WTCasp3_md.tpr -f D_WTCasp3_md.xtc -extr D_WTCasp3_ext2.pdb -first 2 -last 2 -nframes 400

Jupyter notebook commands

Note-This code is written for finding the interactions between the conserved amino acids and segregating them according to specific amino acid properties such as aromatic, polar, non-polar, positive and negative. Firstly, upload the MD simulation file on Cytoscape and extract the interactions and open the inetractions file using MS-Excel and rename the columns to SRC,DEST and CONTACTS. Make another file which has all the highly conserved residues.

import pandas as pd

import numpy as np

Change #1. Change the file name accordingly

```python
fname = 'D_Cp3_No_urea.csv'
df = pd.read_csv(fname)
df.head()
import os

# folder path
dir_path = os.getcwd()

# list to store files
res = []
fs_files = []
# Iterate directory
for file in os.listdir(dir_path):
    # check only text files
    if file.endswith('.csv'):
        res.append(file)
    if file.endswith('.xlsx'):
        fs_files.append(file)
print(res)
print(fs_files)

#Group the interacting partners
for f in res:
    name = f
    if f.startswith('FS_') or f.startswith('Node'):
        #print('FS file', f)
        continue
    else:
        print(f)
        df = pd.read_csv(f)
```

```python
    try:

        groups = df.groupby('SRC')['DEST'].apply(list)
    except:
        continue
    df1 = groups.reset_index(name = 'CONTACTS')
    df1['DEGREE'] = ''
    for i in range(len(df1)):
        c = len(df1.iloc[i]['CONTACTS'])
        df1.loc[i]['DEGREE'] = c
#print(df1)
    newname = name+'_partners.csv'
    df1.to_csv(newname,index=False)
```

Change #2. Conserved contacts and their degrees.

```python
df_fs = pd.read_csv('FS.csv')
df_file = pd.read_csv('D_Cp3_No_urea.csv_partners.csv')
df_fs['Node'] = ''

df_fs = df_fs.fillna(-1)

df_fs['Partners']=''
df_fs['Degree']=''

# Clean the node
for i in range(len(df_fs)):

    # Change it to integers
    df_fs['A'].iloc[i] = int(df_fs['A'].iloc[i])
```

```python
    # Avoid -1's
    if df_fs['A'].iloc[i] == -1.0:
        continue
    else:
        df_fs['Node'].iloc[i] = df_fs['Cp3'].iloc[i] + '-' + str(df_fs['A'].iloc[i])


    for j in range(len(df_file)):
        if df_file['SRC'].iloc[j] == df_fs['Node'].iloc[i]:
            contacts = df_file['CONTACTS'].iloc[j]
            degree = df_file['DEGREE'].iloc[j]
            break


    df_fs['Partners'].iloc[i]  = contacts
    df_fs['Degree'].iloc[i] = degree
```

Change #3. Change file name accordingly.

```python
df_fs.to_csv('D_Cp3_FScontacts.csv',index=False)
```

We have now generated files which have all the contacts that the conserved residues make, and each conserved residue has a degree (no. of contacts). Now, we will make another file with 20 rows (for the 20 amino acids) and 2 columns (RES and PROPERTY). Further, we can run the codes mentioned below to extract information about the interactions between the conserved residues.

#Here node attributes is the file containing 20 rows and 2 columns

```python
df_node_att = pd.read_csv('Node_attributes.csv')

df_fs['PROPERTY'] = ''

#df_node_att = pd.get_dummies(df_node_att, columns=['PROPERTY'])

df_fs

df_file = df_file.drop('Unnamed: 0',axis=1)

attr={}

for i in range(len(df_node_att)):
```

```python
attr[df_node_att.NODE[i]] = df_node_att.PROPERTY[i]
df_fs.Partners.iloc[0]
import re
for rowno in range(len(df_fs)):

    prop = [ ]

    row = df_fs.Partners.iloc[rowno]
    elements = row.split(',')
    print(elements)
    for i in range(len(elements)):
        res1 = (" ".join(re.split("[^a-zA-Z]*", elements[i]))).strip()
        res1 = res1.replace(" ","")

        print(res1)

        prop.append(attr[res1])

    df_fs['PROPERTY'].iloc[rowno] = prop
    df_fs
```

#df_fs should generate a table on jupyter notebook showing the interacting partners, degree and the property of the amino acids.

Change#4. Change the file name accordingly to save it.

```python
df_fs.to_csv('D_Cp3_FScontacts_updated.csv',index=False)
```

Jupyter notebook commands to generate seaborn violin plots

```
import numpy as np

import pandas as pd

import matplotlib.pyplot as plt

import seaborn as sns


%matplotlib inline


import os


# folder path

dir_path = os.getcwd()


# list to store files

files = []
# Iterate directory

for file in os.listdir(dir_path):

    # check only text files

    if file.endswith('.csv'):

        files.append(file)
print(files)
```

#Change file name accordingly

```
D_ARO = pd.read_csv('top_IL_all_violin.csv')

plt.figure(figsize=(40,15))

sns.violinplot(data=D_ARO,y='DC',x='Caspase', hue='Family')

plt.title('DC_all_topIL')

#plt.yticks([-10,0,10, 20, 30,40,50,60, 70, 100])
```