Evolution of allostery and conformational dynamics of caspases involved in the extrinsic pathway of apoptosis

By

Mithun Nag Karadi Giridhar

Dissertation Submitted to the Graduate Faculty in Partial Completion of the Requirements for the Degree of Doctor of Philosophy at The University of Texas at Arlington

May 2023

Supervising Committee

Dr Allan Clay Clark, Principal investigator

Dr. Cara C Boutte,

Dr. Hyejin Moon,

Dr. Mark W. Pellegrino,

Dr. Matthew Fujita,

Dr. Young-Tae Kim.

ACKNOWLEDEMENTS

My mentor, Dr. Clay Clark, has provided me with invaluable guidance, mentorship, and support throughout my research journey. His insightful comments, constructive criticism, and unending encouragement were instrumental in shaping and enhancing my research and scientific abilities. As a music enthusiast, I was initially drawn to his home-made guitar building skills and wanted to acquire more than research skills from him, and life did not disappoint; I learned how to do science and build guitars, as well as attempted to learn his time management skills, which I'm still terrible at. I will also always treasure the opportunity I got to jam and play music with my supervisor. Overall, if given another opportunity, I would return to his laboratory. My committee members, Dr. Cara C Boutte, Dr. Hyejin Moon, Dr. Mark W. Pellegrino, Dr. Matthew Fujita, and Dr. Young-Tae Kim, from diverse fields of research have provided me with expert guidance, thoughtful feedback, valuable suggestions, and, most significantly driving me to comprehend and communicate more effectively. In addition, I would like to thank Dr. Kimberly Bowles at the life science core facility for providing expert knowledge on DNA work and assisting with a variety of other lab-related tasks to ensure the wellbeing of our lab.

My seniors Dr. Suman Shrestha and Dr. Liqi Yao, have been the most crucial mentors in the lab they've provided me with essential insights and technical expertise during my research work. They were absolutely the best seniors anyone could have asked for and are a very important part of my scientific progress and will carry their wisdom throughout my journey. Additionally, I am grateful to Jessica Tung and Godswill Nwaosu, who provided me with assistance and support throughout the duration of my project.I would like to express my deepest gratitude to my wiser colleague Isha Joglekar for her invaluable contribution and support throughout this journey. Her dedication, knowledge, and passion for science have been a constant source of inspiration and motivation for me. Her insights and perspectives have challenged me to think critically and have broadened my horizons. I would also like to recognize David Diaz, the devoted undergraduate researcher who chose our lab and is driving two projects that are currently in progress and evolving to be published in journals, a remarkable accomplishment. Last but not the least, I'd want to thank Melissa Fee, our lab manager, for managing the lab while also driving research work, which is so natural to her, and for which I've always pushed her to do more as she was more than a lab manager, and I'm sure her creativity will be very beneficial to the science community.

Finally, I am deeply grateful to my family for their unwavering support and encouragement, without which this accomplishment would not have been possible I have dedicated the next section totally toward them. My entire social circle was instrumental in my journey, and I am grateful to each and every one of them. Special thanks to Ankit Deo, my long-term roommate who has always been my well-wisher and the best musician I've known. Lastly, I'd like to thank Malhar for stepping in to helping me with one of the most critical problems of my life when all other options had been exhausted.

DEDICATION

My family has been the most important part of my path, and their never-say-die mentality in the face of adversity has driven me to be who I am today. This piece is entirely dedicated to them in recognition of their years of dedication. When times were tough, my mother, Jyothi Giridhar, wanted to send me to one of the best schools in Bangalore, and my father, Karadi Giridhar's selfless sacrifice enabled me to attend Bishop Cotton Boys School, from which I still carry the school's motto: 'Nex dextrorsum Nex sinistrorsum,' which means neither to the right nor to the left. I will always be driven to achieve my goals. On straight on! On cottonians on! neither to the left nor to the right till my grave. The most important things in my life were my father's sacrifice and my mother's vision. I would also like to thank the world for my wonderful sister Sushma Karadi Giridhar, without whom I would not have been motivated to investigate my deepest questions about what life is and why we are here. My mother always wanted me to be a doctor and serve people, but I'd like to say that I've chosen a path that is much more in line with her vision of me, and that my parents' numerous prayers have been answered. Because of the perfect composition and mixture of equal chromosomal copies that I inherited from them, this world has always been a great place for me, and my spirit has always danced through life, never compromising having fun and living life to the fullest. I'm constantly driven to be the finest son anybody could ask for In the eyes of my parents and the best brother I can be for my sister. These five years of research have satisfied one of my deepest desires in life, and I'm excited to continue learning. I'd want to thank my father and my country, India, for their support throughout my master's program, and I promise to give back to the people of my country what they've given me all these years.

I would want to thank my supervisor Dr. Clay Clark for allowing me to work under him despite my lack of experience in molecular biology, and I vow to do my best to pass on your knowledge to the next generation. I'd also like to thank the American people for their trust in me and for providing us with the grant to work, and I'd like to assure them that I've worked my hardest and tried my best to repay them. My heartfelt appreciation to the United States of America and its people, who are so committed to advancing research for the benefit of humanity and a constant source of inspiration.

ABSTRACT

Caspases are a class of cysteine proteases that play a crucial role in programmed cell death (apoptosis). We complete an evolutionary picture of folding of caspases in the extrinsic pathway of apoptosis. The folding models show that the conformational free energy of the monomeric fold is conserved in the extrinsic pathway of apoptosis. In addition, with molecular dynamics and mass spectrometry we also show that the small subunit is less stable than the large subunit and charges destabilize allosteric regions in caspases allowing for dynamics. In order to gain insight into the evolution of allostery in caspases across different species, we employ an evolutionary strategy that combines experimental and computational methods. The findings provide mechanistic details of allostery in caspases, highlighting a highly conserved process that has evolved diverse mechanisms of packing hydrophobic residues in the core, which controls the architecture of the active site to fine-tune the specificity and activity of caspases. Further we identify residues and interactions that mediate allosteric regulation and substrate recognition, which has implications for the development of novel drugs to treat a wide range of diseases, including cancer and autoimmune disorders. Finally, we identified the amino acid networks in the small subunit that have undergone distinct evolution in the initiator and effector caspase subfamilies, thus explaining their divergent oligomerization properties.

Table of Contents

INTRODUCTION:

Overview of Caspases and their Role in Apoptotic and Non-Apoptotic Processes.....1

References.....5

CHAPTER 1

Conserved folding landscape of monomeric initiator caspases

Abstract ……..11

Introduction.....12

Results……..…14

Discussion……26

Methods……29

References……35

Figures……...38

Supporting information……45

CHAPTER 2

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

Abstract ……..62

Introduction.....63

Results……..…65

Discussion……71

Material and Methods……74

References……76

Figures……...79

Supporting information……84

List of sequences used for ancestral reconstruction……..95

CHAPTER 3

Investigating the pH-Induced Destabilization Effect in Caspases

Abstract ……..128
Introduction.....129
Results……..…130
Material and Methods……132
Conclusion……133
Figures……..……....134
Supporting information……140
References……...143

CHAPTER 4

Caspase-8 dimerization constructs cloning with Gibson assembly

Introduction and Results……145
Conclusion……146
Figures……..……...147
Primers for Gibson cloning……148
References……...150

CHAPTER 5

Caspase-8 and cFLIP Proteins: probing the mechanism of homodimerization and heterodimerization in caspases using tandem systems.

Abstract……..151
Introduction……152
Results……..……157
Conclusion……160
Figures……..……...161
References……...166

CHAPTER 6

Resurrection of ancestral caspases to decipher the evolution of the initiator and executioner subfamily in the extrinsic pathway of apoptosis

Introduction……169
Results……..……171
Conclusion……..173
Figures……..……174
Supporting information……..……178
References……..……181

CHAPTER 7

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

Introduction……182
Results……..……183
Methods……184
Figures……..……186
Supporting information……..……187
References……..……191

Appendix

Igor pro fit commands……..……192
GROMACS commands for urea, water simulations and free energy landscapes……197
Gromacs in high performance computing: slurm commands……...205
MATLAB commands to generate energy landscapes……209
Jupyter notebook commands: python codes to analyze networks……...210
Violin plots, joint plots and pair plots using jupyter notebook……216

INTRODUCTION

Overview of Caspases and their Role in Apoptotic and Non-Apoptotic Processes

The study of animal development in the nineteenth century aroused questions about the occurrence, processes, and role of physiological cell death. But it wasn't until the mid-20th century that researchers realized that multicellular animals' cells could initiate a self-destruction mechanism called programmed cell death [1,2,3,4,5,6,7]. The model organism c elegans was instrumental in our understanding of the genetic basis of apoptosis[8]. Following the discovery of Ced3 and Ced 4 several homologous genes were identified in sponge, Hydra vulgaris, Drosophila melanogaster, zebrafish, humans and mice[7]. Eighteen caspases have been discovered in mammals with eleven described in humans[9]. Human caspases can be classified into apoptotic (Caspase 2,3,6,7,8,9,10) and inflammatory caspases (Caspase 1,4,5) based on their functions[10].

In the apoptotic family, activation of caspases is a biochemical signature of apoptosis, or programmed cell death. Caspases are classified into two types depending on the sequence in which they are activated: initiator caspases and effector caspases[11]. Caspases, like other proteases, are expressed in cells as inactive zymogens known as procaspase[12,13]. Activation of procaspase initiators differs from that of procaspase executioners. The monomeric initiator procaspases are activated after recruitment to a protein scaffold via their long prodomain called DED or CARD domains where they undergo dimerization followed by proteolytic cleavage[14]. Executioner caspases, such as procaspases-3, -6, and -7, have short prodomains and are expressed as constitutive dimers, and activation is induced by cleavage of the inter subunit linker by initiators or autoproteolysis[15,16,17]. The

apoptotic family can be classified into extrinsic or intrinsic pathway depending on whether the signal for apoptosis induction originates from outside or within the cell. In the extrinsic pathway of apoptosis Caspase 8 and 10 are initiators whereas caspase 3,6,and 7 are executioner caspases[18]. cFLIPL/cFLAR which is located in close proximity to the chromosomal location of Caspase 8 and 10 and is structurally similar to Caspase 8 and 10 but lacks the functional protease domain, regulates the function of caspase 8 and 10 by heterodimerization[19,20].

Caspase-8 is an apical caspase in the death receptor-induced cell death activation cascade[21]. Activation of death receptors like CD95/Fas/Trail recruits Fas-associated death domain, to which procaspase 8 is recruited via its DED domains and subsequent aggregation of additional procaspase 8 is believed to induce dimerization by proximity induced dimerization followed by cleavage [22,23]. Although the apoptotic activities of apoptotic caspases are well established, emergent non-apoptotic roles such as cell proliferation, differentiation, immunological response, stress, necroptosis, aging, synaptic plasticity, and tumor growth are less well understood[24]. Caspase 8 is reported to heterodimerize with cFLIP, a pseudoenzyme, in addition to homodimerization[25]. Caspase 8 heterodimerization with cFLIPL accomplishes its pro-survival roles. The caspase 8 conformation induced by heterodimerization has a restricted substrate repertoire, primarily inhibiting RIP1 and RIP3 from initiating necroptosis[26]. In High molecular weight platforms like DISC, activation of initiator caspases is followed by activation of effector caspases, with no point of return in the caspase cascade. It's not surprising, then, that molecular mechanisms exist to prevent initiator caspases from being activated during activities such as regulated or uncontrolled cell growth (tumor formation)[27]. Post translational modifications such as phosphorylation,

ubiquitylation, nitrosylation, dimerization, and limited proteolysis of initiator caspases, play a critical role in tightly regulating the molecule[28]. However, there are still numerous substantial gaps in our understanding of how these post translational modifications fine tune conformations and influence death vs non-death processes[29].

Computational tools applied to biomolecular systems, together with structural, thermodynamic, and kinetic data, aid in understanding the new viewpoint of proteins, where one sequence may code for several conformations, resulting in promiscuous activity or diverse functions[30,31]. Studies using antibody SPE7 revealed that structural isomers from the same sequence may increase antibody repertoire size while increasing autoimmunity and allergy[32]. Promiscuous activity was discovered in resurrected ancestral effector caspases which also exhibited a distinct folding landscape[33,34]. These studies highlight the epistatic nature of evolutionary changes on these statistical polypeptide chains to specialize regulation and function. NMR studies and Molecular Dynamics simulations have been indispensable for studying molecular motions [35, 36]. Protein folding research has uncovered folding processes, cooperative folding transitions, thermodynamic stability of mutants, and states that exist between unfolded and folded states, as well as their functional significance [37].Natural proteins as we observe them today are highly evolved complex systems[38]. Folding-on-binding, allostery, switching, and gating are examples of conformational changes tied to the energy landscape that biological polypeptides traverse [39]. The energy landscape provides the notion that diverse protein conformations generate low-lying energy minima. Aspects such as ligand binding, phosphorylation, native state symmetry, and misligation (metal ions) may influence folding landscapes[40]. Proteins with a wide range of functions

fold into similar shapes [41]. To understand how function and allostery in proteins restrict folding, understanding the folding energy landscape of families like caspases is essential [16,42]. In addition, the knowledge gained from such data will equip us with a new generation of caspase-targeting ideas, such as allosterically stabilizing certain conformations with small molecules to affect specific pathways, as opposed to the current drugs, which have largely failed due to issues such as lack of specificity, susceptibility, and limited efficacy as they all target the active site[45].

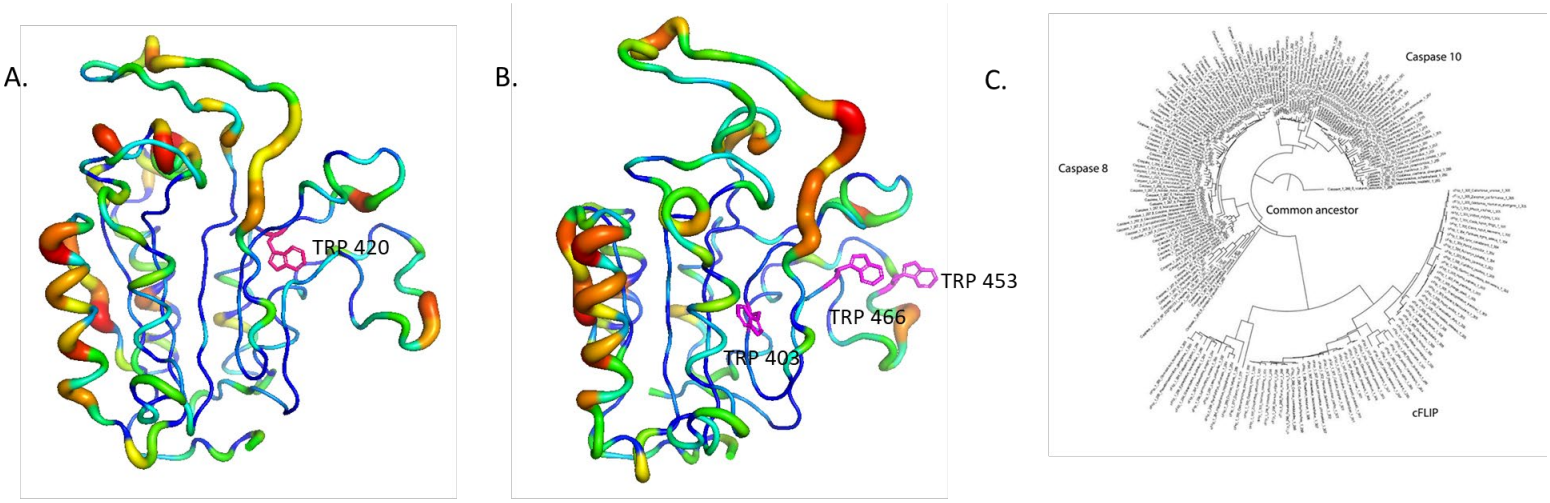


Figure 1: Conservation of amino acids and evolutionary relationship between Caspase 8 and cFLIP. **A.** Conservation of amino acids in the NMR structure of Caspase 8 compared to caspase 3,6,7,10 and cFLIP. Temperature B factor coloring depicting well-conserved regions as thin and blue (cold) and less conserved regions as thick and red (hot) (hot). Tryptophan residues are shown in magenta. **B.** Amino acid conservation in cFLIP models of caspase 8 2k7z compared to caspase 3,6,7,8 and 10. Temperature B factor coloring displaying well-conserved sections as thin and blue (cold) and poorly conserved regions as thick and red (hot). Tryptophan residues are colored magenta. **C.** Relationship phylogenetic of initiator caspases.

References

1. P. G. H. Clarke, S. Clarke, nineteenth century research on naturally occurring cell death and related phenomena - Brain Structure and Function. SpringerLink (1996) (May 10, 2022).
2. R. A. Lockshin, Z. Zakeri, Programmed cell death and apoptosis: origins of the theory - Nature Reviews Molecular Cell Biology. Nature (2001) (May 10, 2022).
3. , Programmed cell death—I. Cytology of degeneration in the intersegmental muscles of the Pernyi silkmoth. Programmed cell death—I. Cytology of degeneration in the intersegmental muscles of the Pernyi silkmoth - ScienceDirect (2003) (May 10, 2022).
4. , Requirement for RNA and protein synthesis for induced regression of the tadpole tail in organ culture - ScienceDirect. Requirement for RNA and protein synthesis for induced regression of the tadpole tail in organ culture - ScienceDirect (2004) (May 10, 2022).
5. J. F. R Kerr, A. H. Wyllie, A. R. Currie, Apoptosis: A Basic Biological Phenomenon with Wideranging Implications in Tissue Kinetics - British Journal of Cancer. Nature (1972) (May 10, 2022).
6. E. PMC, Europe PMC. Europe PMC (May 10, 2022).
7. J. C. Ameisen, On the origin, evolution, and nature of programmed cell death: a timeline of four billion years - Cell Death & Differentiation. Nature (2002) (May 10, 2022).
8. , Genetic control of programmed cell death in the nematode Caenorhabditis elegans - PubMed. PubMed (1999) (May 10, 2022).
9. , Caspases - ScienceDirect. Caspases - ScienceDirect (2015) (May 10, 2022).
10. , 1. , Caspases in Cell Death, Inflammation, and Disease. Caspases in Cell Death, Inflammation, and Disease - ScienceDirect (2019) (May 11, 2022).
11. S. Kumar, Caspase function in programmed cell death - Cell Death & Differentiation. Nature (2006) (May 11, 2022).
12. David R. McIlwain , Caspase Functions in Cell Death and Disease - PMC. PubMed Central (PMC) (2013) (May 11, 2022).

13. C. Pop, Y. Chen, B. Smith, K. Bose, B. Bobay, A. Tripathy, S. Franzen, and A. Clay Clark. Removal of the Pro-Domain Does Not Affect the Conformation of the Procaspase-3 Dimer Biochemistry 2001, 40, 47, 14224–14235 (October 30, 2001)
14. Anson, Francesca and Thayumanavan, S. and Hardy, Jeanne A Exogenous Introduction of Initiator and Executioner Caspases Results in Different Apoptotic Outcomes- JACS Au (July 8,2021)
15. Duclos, C., Champagne, A., Carrier, J. et al. Caspases play in traffic. Cell Death Dis 8, e2636 (2017). https://doi.org/10.1038/cddis.2017.55
16. Clay Calrk Caspase Allostery and Conformational Selection Chem. Rev. 2016, 116, 11, 6666–6706 (January 11, 2016)
17. Jad Walters, Cristina Pop, Fiona L. Scott, Marcin Drag, Paul Swartz, Carla Mattos, Guy S. Salvesen, A. Clay Clark; A constitutively active and uninhibitable caspase-3 zymogen efficiently induces apoptosis. Biochem J 15 December 2009; 424 (3): 335–345. doi: https://doi.org/10.1042/BJ20090825
18. Elmore, Susan. “Apoptosis: a review of programmed cell death.” Toxicologic pathology vol. 35,4 (2007): 495-516. doi:10.1080/01926230701320337
19. 11. A. R Safa, Roles of c-FLIP in Apoptosis, Necroptosis, and Autophagy - PMC. PubMed Central (PMC) (2013) (May 10, 2022).
20. Identification of novel mammalian caspases reveals an important role of gene loss in shaping the human caspase repertoire - PubMed. PubMed (2008) (May 10, 2022).
21. C. Ma,S. H. MacKenzie,A. Clay Clark Redesigning the procaspase-8 dimer interface for improved dimerization 17 January 2014 Protein Science https://doi.org/10.1002/pro.2426
22. T. Min Fu, Y. Li, A. Lu, Z. Li, P. R. Vajjhala, A. C. Cruz, D. B. Srivastava, F. DiMaio, P. A. Penczek, R. M. Siegel,K. J. Stacey, E. H. Egelman and H. Wu. Cryo-EM Structure of Caspase-8 Tandem DED Filament Reveals Assembly and Regulation Mechanisms of the Death Inducing Signaling Complex - PMC. PubMed Central (PMC) (2016) (May 11, 2022).

23. G. S. Salvesen and V. M. DixitCaspase activation: The induced-proximity model | PNAS. (n.d.). Retrieved May 11, 2022, from https://www.pnas.org/doi/10.1073/pnas.96.20.10964
24. Abraham MC, Shaham S . Death without caspases, caspases without death. Trends Cell Biol 2004; 14: 184–193.
25. Hillert, L.K., Ivanisenko, N.V., Busse, D. et al. Dissecting DISC regulation via pharmacological targeting of caspase-8/c-FLIPL heterodimer. Cell Death Differ 27, 2117–2130 (2020). https://doi.org/10.1038/s41418-020-0489-0
26. van Raam, Bram J, and Guy S Salvesen. “Proliferative versus apoptotic functions of caspase-8 Hetero or homo: the caspase-8 dimer controls cell fate.” Biochimica et biophysica acta vol. 1824,1 (2012): 113-22. doi:10.1016/j.bbapap.2011.06.005
27. Shi, Yigong. “Caspase activation, inhibition, and reactivation: a mechanistic view.” Protein science : a publication of the Protein Society vol. 13,8 (2004): 1979-87. doi:10.1110/ps.04789804
28. Parrish, Amanda B et al. “Cellular mechanisms controlling caspase activation and function.” Cold Spring Harbor perspectives in biology vol. 5,6 a008672. 1 Jun. 2013, doi:10.1101/cshperspect.a008672
29. A. V. Zamaraev, G. S. Kopeina, E.A. Prokhorova, B. Zhivotovsky, I. N. lavrik Posttranslational modification of caspases: The other side of apoptosis regulation. February 07, 2017 DOI:https://doi.org/10.1016/j.tcb.2017.01.003
30. Cui, Qiang, and Martin Karplus. “Allostery and cooperativity revisited.” Protein science : a publication of the Protein Society vol. 17,8 (2008): 1295-307. doi:10.1110/ps.03259908
31. 5. T. Hampton, The New View of Proteins | Tyler Hampton | Inference. Inference (2014) (May 11, 2022).
32. Leo James, Pietro Roversi, and Dan S. Tawfik, “Antibody Multispecificity Mediated by Conformational Diversity,” Science 299, no. 5,611 (2003): 1,362–67.
33. 6. R. D. Grinshpon, et al., Resurrection of ancestral effector caspases identifies novel networks for evolution of substrate specificity | Biochemical Journal | Portland Press. Portland Press (2019) (May 11, 2022).

34. L. Yao, P. Swartz, P. T. Hamilton, A. C. Clark, Remodeling hydrogen bond interactions results in relaxed specificity of Caspase-3 | Bioscience Reports | Portland Press. Portland Press (2021) (May 11, 2022).

35. Boehr DD, Dyson HJ, Wright PE. An NMR perspective on enzyme dynamics. Chem Rev. 2006 Aug;106(8):3055-79. doi: 10.1021/cr050312q. PMID: 16895318.

36. Karplus M, McCammon JA. Molecular dynamics simulations of biomolecules. Nat Struct Biol. 2002 Sep;9(9):646-52. doi: 10.1038/nsb0902-646. Erratum in: Nat Struct Biol 2002 Oct;9(10):788. PMID: 12198485.

37. Chen, Yiwen et al. “Protein folding: then and now.” Archives of biochemistry and biophysics vol. 469,1 (2008): 4-19. doi:10.1016/j.abb.2007.05.014

38. Sikosek, Tobias, and Hue Sun Chan. “Biophysics of protein evolution and evolutionary protein biophysics.” Journal of the Royal Society, Interface vol. 11,100 (2014): 20140419. doi:10.1098/rsif.2014.0419

39. Zhuravlev, P., & Papoian, G. (2010). Protein functional landscapes, dynamics, allostery: A tortuous path towards a universal theoretical framework. Quarterly Reviews of Biophysics, 43(3), 295-332. doi:10.1017/S0033583510000119

40. Ferreiro, Diego U et al. “Frustration in biomolecules.” Quarterly reviews of biophysics vol. 47,4 (2014): 285-363. doi:10.1017/S0033583514000092

41. Finkelstein AV, Gutun AM, Badretdinov AYa. Why are the same protein folds used to perform different functions? FEBS Lett. 1993 Jun 28;325(1-2):23-8. doi: 10.1016/0014-5793(93)81407-q. PMID: 8513888.

42. Morcos, Faruck et al. “Coevolutionary information, protein folding landscapes, and the thermodynamics of natural selection.” Proceedings of the National Academy of Sciences of the United States of America vol. 111,34 (2014): 12408-13. doi:10.1073/pnas.1413575111

43. Kakoli Bose and A. Clay Clark Dimeric Procaspase-3 Unfolds via a Four-State Equilibrium Process Biochemistry 2001 40 (47), 14236-14242DOI: 10.1021/bi0110387

44. Suman Shrestha A. Clay Clark Evolution of the folding landscape of effector caspases Journal of Biological Chemistry. (November 2021) DOI:https://doi.org/10.1016/j.jbc.2021.101249

45. Fischer, U., Schulze-Osthoff, K. Apoptosis-based therapies and drug targets. Cell Death Differ 12, 942–961 (2005). https://doi.org/10.1038/sj.cdd.4401556

Chapter 1:

Conserved folding landscape of monomeric initiator caspases

Mithun Nag,[1,] and A. Clay Clark[1,*]

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

Running title: Sequential folding of monomeric caspases

[*]Corresponding author: A. Clay Clark

Email: clay.clark@uta.edu

Key Words: caspase, apoptosis; folding landscape, protein evolution, evolutionary biology, apoptosis, protease, protein folding, fluorescence emission, circular dichroism, molecular dynamics simulations

Abstract

The apoptotic caspase subfamily evolved into two subfamilies - monomeric initiators and dimeric effectors. Sequence variations in the conserved caspase-hemoglobinase fold resulted in changes in oligomerization, enzyme specificity, and regulation, making caspases an excellent model for examining the mechanisms of molecular evolution in fine-tuning structure, function, and allosteric regulation. We examined the urea-induced equilibrium folding/unfolding of two initiator caspases, monomeric caspase-8 and $cFLIP_L$, over a broad pH range. Both proteins unfold by a three-state equilibrium mechanism that includes a partially folded intermediate. In addition, both proteins undergo a conserved pH-dependent conformational change that is controlled by an evolutionarily conserved mechanism. We show that the conformational free energy landscape of the caspase monomer is conserved in the monomeric and dimeric subfamilies. Molecular dynamics simulations in the presence or absence of urea, coupled with limited trypsin proteolysis and mass spectrometry, show that the small subunit is unstable in the protomer and unfolds prior to the large subunit. In addition, the unfolding of helix 2 in the large subunit results in disruption of a conserved allosteric site. Because the small subunit forms the interface for dimerization, our results highlight an important driving force for the evolution of the dimeric caspase subfamily through stabilizing the small subunit.

Introduction

Caspases are a family of enzymes that play critical roles in apoptosis and inflammation. In the apoptotic cascade, caspases function either in the intrinsic or extrinsic pathways, depending on the origin of the signal for apoptosis(1). In the extrinsic pathway of apoptosis, caspases evolved into two distinct subfamilies, namely initiator caspases and effector caspases, and their activation mechanisms differ yet are critical for the regulation of apoptosis(2). Caspases-8 and -10 are initiators of apoptosis, whereas caspases-3, -6, and -7 are effectors of apoptosis(3). Caspases are expressed in cells as zymogens and must be activated for full enzyme activity. The initiator procaspases exist as monomers and must dimerize to gain partial activity; dimerization is followed by cleavage of the zymogen, leading to full catalytic potential. In contrast, the effector procaspases-3, -6, and -7 are stable dimers and require only proteolytic processing to be activated (4, 5).

Caspases are an attractive system to study protein evolution due to the evolutionarily conserved fold that is utilized in both monomeric and dimeric subfamilies (6). The caspase-hemoglobinase fold that comprises the protease domain has a Rosmann-like organization (3-layer sandwich) and has been largely conserved for at least one billion years of evolution, although the sequence conservation is generally low (~15%) (7, 8). The caspase protomer is organized as a single polypeptide chain with an N-terminal pro-domain connected to the protease domain, and the protease domain is organized with a large subunit, intersubunit linker, and small subunit (Fig. 1A) (5). Although the zymogens of caspases-8 and -10 are produced as monomers in the cell, the proteins are enzymatically active only in the dimeric state (4). The death effector domains (DED) or caspase activation and recruitment domains (CARD) within the pro-domain of initiator caspases facilitate dimerization through interactions with similar motifs on oligomerization platforms, such as the death inducing signaling complex (DISC)(9). The intersubunit linker (IL) is cleaved following dimerization on the DISC or other platforms, which separates the large and small subunits and allows the active site to form. The two subunits of the protomer fold into a single unit with a six-stranded β-

sheet core and five α-helices on the surface (Fig. 1B). Cleavage of the IL (Loop 2 in Fig. 1B) leads to active site loop rearrangements and formation of the substrate binding pocket (10) .

In addition to its role in apoptosis, caspase-8 performs a range of non-apoptotic functions. For example, it forms a heterodimer with $cFLIP_L$, a pseudoenzyme, and the caspase-8:$cFLIP_L$ heterodimer functions in pro-survival pathways by blocking RIP1 and RIP3 from initiating necroptosis, a non-apoptotic type of cell death (11). The $cFLIP_L$/cFLAR gene is located in close proximity to those of caspases-8 and -10 on human chromosome 2q33-34, and the protein is structurally similar to caspase-8 and -10 but lacks a functional protease domain(12). $cFLIP_L$ evolved early in the caspase-8/-10 subfamily, and while it retains the caspase hemoglobinase fold, mutations in the active site prevent substrate binding (Fig. 1C and Supplemental Fig. S1) (13).

A comparison of the amino acid sequences of extrinsic caspases from all chordates shows that the β-sheet core is highly conserved, except for β-strand 2, but the helices on the protein surface are less conserved, particularly helix 2 (Fig. 1D). Overall, the conservation of the caspase structural scaffold makes it an excellent model for understanding the evolutionary events that led to species-specific changes in oligomerization, enzyme specificity, and regulation. Indeed, it is not clear how the conserved fold resulted in both monomeric and dimeric subfamilies or how oligomerization evolved as a key regulatory mechanism for caspase activity.

The assembly of the effector caspase dimer has been studied extensively, but little is known about the conformational landscape of the initiator caspases (6, 14). Studies of effector caspases-3,-6, and -7 from humans as well as caspase-3 from zebrafish show that the dimers fold and unfold via a four-state equilibrium pathway in which both dimeric and monomeric partially folded intermediates are well-populated (6, 14, 15). For the effector caspases, the folding was found to follow a four-state pathway ($N_2 \leftrightarrow I_2 \leftrightarrow 2I \leftrightarrow 2U$) in which the native dimer (N_2) unfolds to a partially folded dimeric intermediate (I_2), which dissociates into a partially folded monomer (I) prior to unfolding (U). While the folding landscape is conserved, differences in the relative population of

the folding intermediates provides flexibility for each caspase. Furthermore, studies of the common ancestor of effector caspases showed that the folding landscape was established more than 650 million years ago (6, 16). Overall, dimerization is important to the overall conformational free energy in a conserved folding landscape through contributing an additional 14-18 kcal mol^{-1} of free energy to the native dimer compared to that of the monomeric folding intermediate (5-7 kcal mol^{-1})(6, 14).

Aside from the conformational free energy obtained for the monomeric folding intermediate of dimeric caspases, there are no data on the folding of the monomeric caspases. That is, to date, the folding landscape of the protomer has been studied only in the context of a folding intermediate during dimer formation. Here, we examined the equilibrium folding and unfolding of human caspase-8 and $cFLIP_L$, and we show that the protomer folds through at least one well-populated partially folded intermediate prior to forming the native protein (N↔I↔U). The native protein is most stable near physiological pH and exhibits substantial loss of secondary structure at both lower and higher pH, such that the partially folded intermediate, I, predominates. In addition, data from molecular dynamics (MD) simulations at several pHs and in the presence of urea (5 M), followed by limited proteolysis and mass spectrometry, show that the small subunit is unstable within the protomer and unfolds prior to the large subunit. Finally, we showed previously that effector caspases undergo a pH-dependent conformational change, with a pKa of ~6 and that the conformational change resulted in an inactive enzyme, although the protein remained in the dimeric state(6, 15, 17). We observe a similar pH-dependent conformational change in the caspase-8 and $cFLIP_L$ protomers, suggesting that the effects of pH on the protein conformation may be due to an evolutionarily conserved mechanism. Altogether, the data show a conserved folding landscape for caspases and a role for dimerization in stabilizing the small subunit within the protomer, as well as an evolutionarily conserved pH-dependent conformational change present in all caspases.

Results

Caspase-8 and cFLIP$_L$ have intrinsic fluorescence probes that can be used to monitor conformational changes.

The construct of caspase-8 that we utilized comprises 264 amino acids (30 kDa), whereas that of cFLIP$_L$ is 294 amino acids (33.9 kDa) (Supplemental Fig S1)(20). Both constructs lack the pro-domain. Caspase-8 has one tryptophan residue (W420) in active site loop L3, and the tryptophan lines the S2 and S4 substrate binding pockets in the active structure (Supplemental Fig. S1, Fig. 1B). In contrast, cFLIP$_L$ contains three tryptophans (Supplemental Fig. S1, Fig 1C): W403 on β-sheet 5, W466 on β-sheet 6, and W453 on α-helix 5. Thus, for both proteins the tryptophans reside in the small subunit, and in cFLIP$_L$, one tryptophan (W453) forms part of the dimerization interface. In addition, caspase-8 and cFLIP$_L$ have 12 tyrosines (Supplemental Fig. S1) that are well-distributed throughout the structures. As described previously for effector caspases (6, 14, 15), we examined conformational changes in caspase-8 and in cFLIP$_L$ in the presence and absence of urea by observing changes in fluorescence emission following excitation at 280 nm or at 295 nm. While excitation at 280 nm monitors fluorescence emission of tryptophan and tyrosine residues, excitation at 295 nm is specific for tryptophan residues (21). We also monitored changes in secondary structure during unfolding via far-UV circular dichroism (CD).

The folding of caspase-8 and cFLIP$_L$ includes multiple intermediates

Native caspase-8 (that is, in the absence of urea) exhibits fluorescence emission maxima at 320 nm (Supplemental Fig. S2A) or 338 nm (Supplemental Fig. S2B) when excited at 280 nm or 295 nm, respectively, and far-UV CD spectra consistent with a well-packed secondary structure (Supplemental Fig. S2C). In comparison, native cFLIP$_L$ exhibits fluorescence emission maxima at 340 nm when excited at 280 nm (Supplemental Fig S2D) or 295 nm (Supplemental Fig S2E), suggesting that one or more tryptophan residues in cFLIP$_L$ is more exposed to solvent than is the single tryptophan in caspase-8. Like caspase-8, cFLIP$_L$ also exhibits far-UV spectra consistent with well-packed secondary structure (Supplemetal Fig. S2F). When caspase-8

(Supplemental Fig. S2A,B) and cFLIP$_L$ (Supplemental Fig. S2D,E) are incubated in 9 M urea-containing buffer at pH 7.5, both proteins exhibit a red-shift in fluorescence emission to 350 nm as well as a loss of secondary structure, demonstrating that the tryptophans are exposed to solvent and that the proteins are largely unfolded (Supplemental Fig. S2). In addition, at intermediate (4 M) to maximum (9 M) urea concentrations, caspase-8 exhibits two peaks in the fluorescence emission profile when excited at 280 nm, where emission maxima are observed at 305 nm and 355 nm (Supplemental Fig. S2A). As described previously for an ancestral caspase, the two peaks likely represent ionized and non-ionized tyrosinyl residues (6). In contrast, cFLIP$_L$ exhibits a blue shift to 330 nm when the protein is incubated in buffer containing intermediate concentrations of urea (Supplemental Fig. S2D,E).

To examine the unfolding of caspase-8 and cFLIP$_L$, we incubated proteins in urea-containing buffer, from 0 M to 9 M urea. Following equilibration, we monitored changes in fluorescence emission (following excitation at 280 nm or 295 nm) to examine changes in tertiary structure, and we monitored changes in far-UV CD to examine changes in secondary structure, as described previously (14). The data for caspase-8 (Fig. 2A) and cFLIP$_L$ (Fig. 2B), at pH 7.5, show a pre-transition between 0 M and ~1.5 M urea, where there is little to no change in the signal, followed by a cooperative change in signal between ~1.5 M and ~4 M urea. In the cooperative transition, the fluorescence emission decreases relative to the native conformation for caspase-8 but increases in the case of cFLIP$_L$. For caspase-8, the cooperative transition is similar for the three spectroscopic probes (Fig. 2A), except that the loss of secondary structure occurs at lower urea concentrations. In contrast, cFLIP$_L$ exhibits a higher fluorescence emission following the first transition (~4 M urea), and one observes a larger change when the protein is excited at 295 nm compared to excitation at 280 nm (Fig. 2B). A second cooperative transition occurs between ~4 M and 7 M urea. For both caspase-8 (Fig. 2A) and cFLIP$_L$ (Fig. 2B), the protein is largely unfolded at urea concentrations above 7 M. As expected for a monomer, the findings of our studies at two concentrations (2 and 6 uM) show that the folding parameters were concentration independent for both the proteins. Refolding data show that caspase-8 folds reversibly at pH 7.5 (Fig. 2A). In

contrast, cFLIP$_L$ refolds reversibly at urea concentrations greater than 3 M, but at lower concentrations the refolding signals did not recapitulate the unfolding signal (Fig. 2B). The lack of reversibility for cFLIP$_L$ at lower urea concentrations and at pH 7.5 is discussed in more detail below.

We have shown previously that changes in pH are an excellent perturbant for examining the caspase folding landscape (6, 15, 17). Both the protein conformation and the oligomeric state, in the case of human caspase-3, are sensitive to changes in pH, resulting in changes in protein stability. In order to determine whether the monomeric caspases undergo similar changes, we examined the equilibrium folding/unfolding over a broad pH range, from 3.5 to 9 for caspase-8 and for cFLIP$_L$. Although the full range of data are shown in supplemental figures for caspase-8 (Supplemental Fig. S3) and for cFLIP$_L$ (Supplemental Fig. S4), we show results for the lowest (Fig. 2C,D) and highest (Fig. 2E,F) pHs for caspase-8 and cFLIP$_L$, respectively, as a comparison to unfolding at pH 7.5 (Fig. 2A,B). We note that we were unable to obtain data at pH 5.5 for either protein due to protein aggregation. As described below, the far-UV CD signal of caspase-8 and of cFLIP$_L$ decreased below pH 7 and pH 6, respectively, suggesting a loss of secondary structure. Thus, we monitored only changes in fluorescence emission at pHs below 6.

Collectively, the data were fit to equilibrium folding models that best describe the folding/unfolding over the broad range of pH, and the results are shown as the solid lines in the figures (Fig. 2, Supplemental Fig. S3 and S4). The data for both caspase-8 and cFLIP$_L$ were best described by a three-state folding model in which the native protein unfolds through a partially folded intermediate prior to unfolding (N↔I↔U). For caspase-8, at all pHs, the fluorescence emission of the partially folded intermediate is red-shifted and has a higher fluorescence emission compared to the native protein (Supplemental Fig. S2 A,B), demonstrating that the single tryptophan residue is quenched in the native conformation relative to the partially folded intermediate or the unfolded conformation. The data for cFLIP$_L$, is best described by a three-state equilibrium folding model between pH 4.5 and 9 (N↔I↔U) and by a two-state folding model (I↔U) below pH 4.5. As described below for both proteins, one or more of the

conformational states is sensitive to changes in pH, which affects the relative population of the species during unfolding. In contrast to caspase-8, the fluorescence emission signal of the intermediate state for $cFLIP_L$ is blue-shifted and has a lower fluorescence emission signal compared to the native protein (Supplemental Fig. S2 D,E), demonstrating that one or more tryptophans transition to a more hydrophobic environment in the partially unfolded intermediate compared to the native state. Interestingly, the folding/unfolding of $cFLIP_L$ is reversible at pH below 6.5 (Fig. 2D) and above 8 (Fig. 2F), but at the pH range closer to neutral pH (pH 6.5 - pH 8) (Fig. 2B and Supplemental Fig. S4), folding is irreversible. In contrast, the folding/unfolding of caspase-8 is reversible at all pHs.

Global fitting of equilibrium unfolding data indicates that the stability of the protomer is conserved in all caspases

As described previously (21), the data at each pH for caspase-8 and $cFLIP_L$ were fit globally to the folding models described above in order to determine the conformational free energies and m-values associated with each transition. Results of the fits shown as the solid lines in Figure 2 and Supplemental Figures S3 and S4, are presented in Supplemental Tables S1 and S2. The free energy and cooperativity index (m-value) for each unfolding transition were estimated by fitting approximately twenty experimental replicates for fluorescence emission at protein concentrations of 2 μM and 6 μM and six replicates for far-UV CD. The data shown in the figures are averages of the replicates.

The results for caspase-8 show that, at pH 7.5, the total conformational free energy and m-value are 6.3 kcal mol^{-1} and 2.1 kcal mol^{-1} M^{-1}, respectively. Over the pH range of 6.5-9, the two transitions exhibit similar free energies in caspase-8, although the first transition (N↔I) has a somewhat higher conformational free energy (ΔG^0_1) compared to the second transition (I-U) (ΔG^0_2), ~3.7 kcal mol^{-1} *versus* 2.6 kcal mol^{-1} (Supplemental Table S1), as well as m-values (~1.5 kcal mol^{-1} M^{-1} *versus* 0.7 kcal mol^{-1} M^{-1}). The empirical relationship between m-values and surface area described by Scholtz and colleagues (22) suggests that more hydrophobic surface area is exposed in forming the intermediate conformation than during the unfolding of the intermediate. At lower and

higher pH, the conformational free energy of the first transition decreases while that of the second transition remains constant. The change in relative population of the native conformation is observed in the equilibrium folding data (Supplemental Fig. S3), in that, as the relative population of the native protein decreases, one observes a plateau between ~4 M and ~6 M urea, which reflects an increase in the relative population of the intermediate, I. In addition, the mid-point of the transition for N↔I decreases at lower pH. Indeed, at pH 3.5, the pre-transition region disappears, so it is difficult to obtain accurate fits to the first transition. Overall, the data obtained from the global fits are shown in Figures 3A, 3B, and Supplemental Table S1 for caspase-8 and demonstrate that the protein is maximally stable between pH 7-8. At lower and higher pH, the native protein is destabilized relative to the partially folded intermediate, I, such that the relative population of the intermediate increases. Thus, the change in the total conformational free energy *versus* pH is due to the destabilized native conformation. The changes in m-value *versus* pH also show the same trend. That is, the decrease in the m-value at lower pH reflects the increased relative population of the partially folded intermediate.

The global fits of the data for $cFLIP_L$ demonstrate that, unlike caspase-8, the first transition (N↔I) has a substantially lower conformational free energy (ΔG^0_1) compared to the second transition (I-U) (ΔG^0_2), ~2.2 kcal mol^{-1} *versus* 5.5 kcal mol^{-1} (Supplemental Table S2). Together, the data show a substantial increase in the stability of the partially folded intermediate of $cFLIP_L$ in comparison to that of caspase-8. Similar to caspase-8, however, the conformational free energy (ΔG^0_1) of first transition (N↔I) exhibits a pH dependence whereas that of the second transition (ΔG^0_2) is independent of pH (Supplemental Table S2 and Fig. 3C,D). Altogether, the data for folding/unfolding from pH 3.5-9 show that both caspase-8 and $cFLIP_L$ are sensitive to pH changes, similar to the effector caspase dimers (6, 14), due to destabilizing the native conformation relative to a partially folded intermediate. The higher m-values for the N↔I transition in caspase-8 indicates a larger exposure of hydrophobic surface area compared to the same transition for $cFLIP_L$, while the m-values for the second transition (I↔U) is higher for $cFLIP_L$ indicating a more compact conformation for the intermediate state in $cFLIP_L$. The

protomers of caspase-8 and cFLIP$_L$ exhibit a ΔG^0_{conf} of 6-8 kcal mol^{-1}, which is comparable to the monomeric intermediate observed in the equilibrium unfolding of executioner caspases-3 and -7 as well as that of the common ancestor of effector caspase dimers (6, 14). In those cases, the ΔG^0_{conf} of the monomeric folding intermediate was determined to be ~5-7 kcal mol^{-1} at pH 7.5 and 25^{0}C.

Using the values acquired from the global fits of the equilibrium unfolding data and the cooperativity indices determined for each transition (Supplemental Tables S1 and S2), we calculated the equilibrium distribution of species (N, I and U) for each protein at each pH and throughout the urea concentration range of 0 to 9 M, as described previously (17). The data are shown in Supplemental Figure S5 (caspase-8) and Supplemental Figure S6 (cFLIP$_L$). For caspase-8, at pH>6.5, the native species (N) is well-populated at low urea concentrations, from 0 M to 2 M, and the intermediate species (I) shows a maximum population at ~3 M urea. The unfolded fraction is well-populated above 5 M urea (Supplemental Fig S5). At pH<6.5, one observes an increase in the population of the folding intermediate, I, in the absence of urea, such that at pH 3.5, the “native” protein is an ensemble of native (N) and intermediate (I) conformations. Similar results are observed for cFLIP$_L$ in that the fraction of native species (N) decreases relative to the folding intermediate, I, in the absence of urea (Supplemental Fig S6). For cFLIP$_L$, however, the native conformation is not well-populated below pH 4.5.

We note that the thermodynamic parameters (ΔG^0_1 and m_1) reported for the first transition (N↔I), for cFLIP$_L$, from pH6 to pH8.5, are uncertain due to hysteresis. From pH 6 to pH 8.5, the fraction of the intermediate state is greatest at 4M urea (Fig S6), and the enzyme is shown to be reversible up to 4M, from pH6 to pH8.5 (fig S4). Furthermore, at pH 4 and 4.5, the intermediate is the predominant fraction at 0M urea, and since the enzyme is completely reversible, we can be certain of the thermodynamic parameters (ΔG^0_2 and m_2) reported for the second transition (I↔U). Given that data trends indicate that the monomeric fold has a conformational free energy of 5-7 kcal mol^{-1}, in the entire family of caspases in the extrinsic pathway of apoptosis, the conformational free energy of the first transition (ΔG^0_1), should fall within the reported

range, resulting in ~7-8 kcal mol^{-1} of total conformational free energy from pH6 to pH8.5 (Supplemental Table S2 and Fig. 3C,D).

Caspase-8 and cFLIP$_L$ undergo pH-dependent conformational changes

The effector caspase dimers have previously been shown to undergo a pH-induced conformational change that results in an enzymatically inactive dimeric conformation, with a pKa~6 for the transition (6, 17). Based on the data described above for caspase-8 and cFLIP$_L$, where we showed that the native conformation is sensitive to changes in pH, we examined changes in far-UV CD and fluorescence emission signals for the native protein *versus* pH, and we examined the midpoint of the folding transitions that are sensitive to pH changes, namely N↔I (ΔG^0_1) (Supplemental Fig. S7). Note that in supplemental figure S7A and B, the CD plots describing the secondary structure show a drop in secondary structure below pH6 along with a blue shift in the minima. To plot CD signal vs pH, we used CD signal at 230nm and 224.3nm respectively (for fitting), which did not capture the blue shifted minima at lower pH; therefore, the plots of CD signal vs pH in supplemental figure S7C and D are not indicative of the absence of secondary structure, but rather a drastic change in signal.First, the data show that caspase-8 exhibits a maximum far-UV CD signal between pH 7 and 8 (Supplemental Fig. S7C), whereas cFLIP$_L$ exhibits a maximum far-UV CD signal between pH 6 and 7 (Supplemental Fig. S7D), which is consistent with the pH range determined for maximum conformational stability (Fig. 3 A, C) for caspase-8 (pH 7-8) and cFLIP$_L$ (pH 6-7). As described above, we also determined the midpoints for the transitions from examining changes in the fraction of species, which again shows the effects of pH on the native conformation (N) but not the folding intermediate (I), for caspase-8 (Supplemental Fig. S7E) and cFLIP$_L$ (Supplemental Fig. S7F).

As described previously (23), we fit the pH-dependent transitions for the change in secondary structure (Supplemental Fig. S7C,D), transition midpoints (Supplemental Fig. S7E,F) and conformational free energies and m-values (Fig. 3A-D) to determine the pKa for the conformational change, and the fits are shown as the dashed lines in the figures. The results are reported in Supplemental Table S3 (caspase-8) and Supplemental Table S4 (cFLIP$_L$). Overall, the analysis shows two pH dependent

transitions, with pKa_1~5.6 and pKa_2~8.1, and little variation between the two proteins. We suggest that the variety of different probes, such as secondary structure, conformational free energy and m-values, and transition midpoints, likely report on the same conformational changes, regardless of the minor variations in the individual pKa values. In comparison to the dimeric effector caspases, the first transition is conserved, with pKa~6, while the second transition, with pKa~8.1, may be unique to the monomeric caspases. Thus, it appears that the caspase protomer undergoes a pH-dependent transition, regardless of oligomeric state. In the effector caspase dimer, the transition results in reversible formation of an enzymatically inactive intermediate (6, 14), whereas the monomeric caspases partially unfold. We note that in caspase-8, the pH-dependent formation of the folding intermediate, I, is reversible, whereas in $cFLIP_L$ the transition to I is not reversible at pHs close to physiological pH (see Supplemental Fig. S4, and described above). Because the first transition occurs in the monomer and dimer subfamilies, our data suggest that the property is inherent in the caspase protomer, and that the mechanism is conserved.

Molecular dynamics simulations in the presence of urea reveal the small subunit and helix 2 are unstable.

In order to further examine conformational changes in the caspase protomers, we performed molecular dynamics (MD) simulations for 200 ns using caspase-8 and $cFLIP_L$. The starting structures were modeled using the solution structure of the protease domain of procaspase-8 (PDB ID: 2K7Z) (19). As described in Methods, loop regions that connect β-strand 5 with active site loop 3, and α-helix 5 with β-strand 6 (see Supplemental Fig. S1A), were absent in the solution structure. The missing residues were modeled using Swiss modeler to produce a structure for the protomer of caspase-8 with contiguous sequence connectivity. The resulting model of the caspase-8 protomer was then used to generate the starting structure for the protomer of $cFLIP_L$. The solution structure of procaspase-8 provides a view of the protomer prior to oligomerization and substrate binding, because structures of the caspase-8 homodimer and of the caspase-8:$cFLIP_L$ heterodimer, solved by X-ray crystallography, contain

inhibitor bound to the caspase-8 active site. In the model used for MD simulations, the intact intersubunit linker prevents proper active site formation (see Fig. 4).

For each protein, the simulations were performed in the presence and absence of 5 M urea and at pH 3.5, 7.5, and 9. Representative time frames of zero, 100 ns, and 200 ns are shown in Figure 4 for both proteins. The data show that the small subunit unfolds due to helices 4 and 5 lifting away from the β-sheet. The unfolding of the helix 4/5 unit then pulls β-strand 6 away from the core such that the core structure of β2-1-3 (large subunit) remains intact, but the small subunit is largely unfolded. Within the large subunit, helices 2 and 3 separate from the β-sheet core and expose the core to solvent. Although similar processes occur at all pHs and for both proteins, one observes greater unfolding at pH 3.5 and pH 9 compared to pH 7.5.

We examined the unfolding of the proteins by monitoring the root mean square fluctuation (RMSF) of each amino and compared the results at each pH and in the presence and absence of urea for caspase-8 (Supplemental Figure S8A, B, C) and for $cFLIP_L$ (Supplemental Figure S9A, B, C). The RMSF varies with protonation states at pH 3.5, 7.5, and 9 for both caspase-8 and $cFLIP_L$, correlating with the extent of unfolding shown in Figure 4. In order to compare the changes in RMSF due to the presence of urea, we first subtracted the RMSF of simulations in the absence of urea from the RMSF in the presence of urea for caspase-8 (Supplemental Figure S8D, E, F) and $cFLIP_L$ (Supplemental Figure S9D, E, F). At pH 7.5, one observes an increase in RMSF of surface helices in the presence of urea. At pH 9, one observes an increase in the RMSF of helices 4 and 5 in the small subunit, while an increase in RMSF is observed throughout the protein at pH 3.5. We transformed the ΔRMSF values into b-factors and mapped them onto the structure to provide a visual representation of the unfolded regions on the structure. For both proteins, at pH 7.5 and in 5 M urea, helices 3 and 4 and the connecting loop are destabilized (Supplemental Fig S8G (caspase-8) and Supplemental Fig S9G ($cFLIP_L$)), and the fluctuations increase at lower and higher pH (Supplemental Fig S8 panels H, and I (caspase-8) and Supplemental Fig S9 panels H, and I ($cFLIP_L$)). In addition, at lower and higher pHs, the surface helices, particularly helices 2 and 3, show increased fluctuations leading to unfolding. Together, the data

from MD simulations show that the small subunit unfolds first and that a higher degree of unfolding occurs at pH 3.5 compared to higher pH.

Limited trypsin proteolysis of caspase-8 confirms that the small subunit is less stable than the large subunit.

As shown by equilibrium unfolding experiments (Figure 3), the native conformation of caspase-8 is most stable at pH 7-8, and it is destabilized at both higher and lower pH. In order to further examine changes in the protein conformation vs pH, we performed limited trypsin proteolysis of caspase-8 at pH 7.5 and pH 9. Trypsin activity curves from pH 7 to pH 9 display a comparable catalytic efficiency at pH 7 and pH 9, indicating that the activity is not substantially different, allowing us to examine pH-induced caspase-8 conformational changes (24). As described in methods, caspase-8 was treated with trypsin at pH 7.5 or at pH 9. Samples were collected at 15-minute time intervals until two hours, followed by two 30 minute intervals between two hours and three hours, and lastly a sample was collected after incubation overnight. Protein fragments were separated by SDS-PAGE (Fig. 5A,B). We note that the trypsin enzyme was not observed on the gel since the concentration of trypsin was low compared to that of caspase-8. We analyzed the changes in intensity of the native band at pH 7.5 and pH9, and then we fitted a single exponential equation to the data. The apparent rate constant that we obtained from this analysis is displayed in Supplemental Figure S11, and it reveals that the full-length protein (31 kDa) is cleaved with a half-time of 20 minutes at pH 9, in contrast to a half-time of 40 minutes at pH 7.5. This indicates that the enzyme is less stable at pH9, which is consistent with our conformational free energy estimates (Fig. 3A), which demonstrate a decrease in free energy values.

In independent studies, we treated caspase-8 with trypsin for 60 minutes at pH 7.5 and pH 9 (representing the 5th lane from the 0 time point lane in Figure 5A&B) and performed MALDI-TOF analysis to quantify the molecular weights of all the fragments (Supplemental Fig. S12). We picked the 60-minute time period because it yielded the greatest number of distinct products on the gel (Fig 5 A&B). To gain insight into the regions that are susceptible to the highest cleavages and hence represent the regions that are destabilized, we examined fragments with the highest intensities (top 10) at pH

7.5 and pH9 (Supplemental Fig. S11). All of the cleavage sites were determined (Supplemental Fig. S12) and were mapped onto the structure of caspase-8 (Fig. 5C) to illustrate that the largest contributions to the fragments in (fig 5A,B) are the result of cleavages occurring on the intersubunit linker (2 sites), small subunit (9 sites), and large subunit (11 sites) (2 sites). Cleavages in the small subunit account for 9 of the 15 identified sites, with cleavages on Loop 3 (1 site), Helix 4 (3 sites), Loop 4 (3 sites), and C-terminus (2 sites), whereas cleavages in the N-terminus are more limited to the N-terminus (2 sites) and loop1 (2sites). The majority of these cleavages (9 of 15) occur in the small subunit, indicating that the small subunit is generally less stable at both pH. At pH 9, additional cleavage sites are observed in loop 1 (2 sites), helix 1 (1 site), and the loop region between beta sheet 1 and the N terminus (Fig. 5C). These data suggest that the area surrounding helix 2 is destabilized, which reduces the stability of the large subunit and hence the molecule at pH9, resulting in a higher rate of cleavage at this pH. In addition, two high-intensity cleavage products of 4232 kDa and 4504 kDa (Supplemental Fig. S13) that map to helix 2 make it to the top 10 list of maximum cleavage products to support our conclusion along with the observation in MD simulations that suggests the large subunit is destabilized around helix 2. (fig 4A).

Altogether, the increased apparent rate constant at pH 9 vs pH 7.5 are consistent with the conformational changes monitored by changes in fluorescence emission and by CD, described above, which result in an increased exposure of cleavage sites. Several regions of the small subunit are accessible to cleavage by trypsin at both pH, while cleavage in the large subunit is more limited to the N-terminus and active site loop 1. Taken together, the results of limited trypsin proteolysis are consistent with the results of the MD simulations which show that the small subunit is less stable, and thus more accessible to cleavage by trypsin. In addition, the increased cleavage of helix 2, loop linking N-terminus to beta sheet 1, and loop 1 on the large subunit is consistent with the fluctuations found in MD simulations, as detailed previously, indicating that protonation destabilizes the large subunit around helix 2.

Discussion

Molecular evolution has broadened a limited set of ancient protein folds to perform varied biochemical tasks in a multitude of present-day species (25). The caspase family of enzymes has evolved into multiple members in several subfamilies, with new functions and allosteric regulation, while preserving the fold of a ~4 billion-year-old ancestral scaffold (26). We show here that caspase-8 and cFLIP$_L$ unfold at pH 7.5 by a three-state equilibrium folding pathway, where the native protein (N) unfolds to a partially folded intermediate (I) prior to unfolding (U). The overall conformational free energies of 6.2 kcal mol^{-1} (caspase-8) and of 7.7 kcal mol^{-1} (cFLIP$_L$) are similar to that determined previously for a monomeric folding intermediate of dimeric effector caspases, which is in the range of 5-7.0 kcal mol^{-1} (6, 14). In general, both caspase-8 and cFLIP$_L$ show similar properties, even though they are separated by ~300 million years of evolution from each other (27) and nearly 650 million years removed from the effector subfamily (6). Indeed, cFLIP$_L$ evolved to be a pseudo-enzyme that plays an important role in necroptosis through forming a heterodimer with caspase-8, while caspase-8 evolved as an activator of effector caspases (12). The similar conformational free energies of the monomers suggest that the folding landscape has been conserved throughout evolution, even as the two subfamilies evolved into multiple members with different substrate specificities and allosteric regulation.

The characterization of the native protomer of caspase-8 and of cFLIP$_L$ is consistent with the monomeric intermediate described previously for dimeric effector caspases. For example, the native conformation of the caspase-8 and the cFLIP$_L$ protomer has a partially disordered active site, as observed by limited trypsin proteolysis, yet the tryptophan residue near the active site is buried. In addition, the m-values for equilibrium folding demonstrate that the monomer is only partially folded. For example, Scholtz and colleagues described a correlation between unfolding m-values, the change in buried accessible surface area (ΔASA), and number of residues (22). For all caspases studied to date, the m-values for unfolding of the monomer range from 1.20-1.98 kcal mol^{-1} M^{-1} (Supplemental Tables S1 and S2 and (6, 14, 15). Using the equations described by Scholtz, the ΔASA for unfolding the caspase protomer is

~8,000-15,000 Å^2, corresponding to ~96-170 residues. Caspase protomers comprise ~260 amino acids, so the collective equilibrium folding data suggest that the “native” protomer contains substantial surface area exposed in loops and other unstructured regions. Our data, shown here, from MD simulations and limited trypsin proteolysis, suggest that the loss of buried surface area results from fluctuations in the small subunit and the surface helices.

We note that the assembly of the dimer of effector caspases does not occur through assembly of two pre-formed “native” protomers because an additional conformational change occurs after dimerization that involves active site loop rearrangements (14, 23, 28). The flexibility of the small subunit within the protomer likely results in a slow rate of dimerization. We showed previously that the rate of dimerization for procaspase-3 is ~70 M^{-1} sec^{-1}, which is very slow compared to many homodimers (18). Procaspase-8 has been shown to form homodimers in the presence of high concentrations of kosmotropes, such as sodium citrate at 1 M (29). Extrapolation of the measured dimerization rates to the absence of kosmotrope, however, suggested a second order rate of dimerization near zero. Decreasing the slow rate of dimerization by a small factor would essentially trap the protomer as a monomer, which shows the importance of the DED motifs and the activating platforms for increasing the local concentration of protomer to facilitate dimerization. Indeed, it was suggested that negative design elements in the dimer interface may decrease the rate of dimerization while simultaneously increase specificity (18). For example, in the case of caspase-8, F468, on β-strand 6 of one protomer, interacts with P466 in the interface of the second protomer, resulting in intersubunit stacking interactions. In the caspase-3 dimer, introduction of the V266H variant in the dimer interface resulted in a kinetically trapped monomer that slowly dimerized following rearrangement of the bulky histidine residues (30), demonstrating that reducing the very slow rate of dimerization effectively traps the protomer. We showed that optimizing the dimer interface by replacing P466 and F468 in caspase-8, increased the rate of dimerization but did not result in a stable dimer, suggesting that additional factors are required for dimerization aside from an optimized β-strand 6 (18). The data presented here suggest that a key factor resulting from

dimerization is the stabilization of the small subunit. Combined with our previous kinetic folding data for procaspase-3 (30), where we showed that the monomer forms a dimerization-competent species as well as other species that are not competent to form dimers, we suggest that the fluctuations in the small subunit result in an ensemble of protomer conformations which reduces the concentration of a dimerization-competent conformation and effectively decreases the second order rate of dimerization. Thus, the evolution of monomer and dimer caspase subfamilies appears to be a result of tailoring the conformational dynamics of the protomer. Evolutionarily, dimers may have originated as a consequence of stabilizing the small subunit, where a modest increase in the rate of dimerization would provide access to additional regions of the conformational landscape, that of the dimer, leading to substantial increases in conformational free energies of the dimer vs the monomer.

Different cellular compartments have been revealed to have different pH values. The pH of cytosol, endoplasmic reticulum, and nucleus is 7.2, mitochondria is 8, the golgi network is 6.5-6, and lysosomes is 4.7(31). Our findings reveal that protons in the environment can fine-tune the stability and hence the conformational landscape of these enzymes. Localizing caspase-8 and $cFLIP_L$ to different compartments could be a strategy used by cells to fine-tune conformational dynamics and affect certain pathways. In solid tumors with pH dysregulation, the initiator caspase-8 and $cFLIP_L$ can be substantially destabilized, affecting both apoptosis and necroptosis pathways(32).

It is worth noting that caspase conformations are finely regulated by dimerization, metal binding, post-translational modifications, and limited proteolytic cleavages (5). Most post-translational modifications occur in the large subunit, with just two sites known in the small subunit of caspase-8 (Supplemental Fig. S1). An evolutionarily conserved allosteric site is located at the base of helices 2 and 3 and near the residues in the N- and C-termini (33). Several post-translationally modified amino acids are known to be localized near the conserved allosteric site. For example, S150 and T152 in caspase-3, S347 in caspase-8, and T173 in caspase-7, are located at the base of helix 2, and their phosphorylation by p38 MAPK or by p21-activated kinase 2 (PAK2) decreases caspase activity (33–35). Our folding studies, presented here, demonstrate

that the native state of caspase-8 and of $cFLIP_L$ can be modulated by protonation/deprotonation through changes in pH. Results from our MD simulations show that helix 2 and 3 of the large subunit are unstable, suggesting that allosteric modulation, through ligand binding to the allosteric site and post-translational modifications, result in changes to the conformational ensemble to favor the partially folded conformation (I) by destabilizing the two surface helices near the allosteric site. Interestingly, all caspases have shown a pH-dependent conformational change, with pKa~6.

Methods

Cloning, protein expression, and protein purification

The plasmids for caspase-8 (36) and for $cFLIP_L$ (20) were obtained from the Addgene plasmid repository, and the catalytic cysteine was changed to alanine using site-directed mutagenesis. The purification steps were described previously (18). The concentration of caspase-8 was estimated using ε_{280} = 23,380 $M^{-1}cm^{-1}$. $cFLIP_L$ was purified as described previously (29), and the protein concentration was measured using ε_{280} = 34,380 $M^{-1}cm^{-1}$. We used constructs of caspase-8 and of $cFLIP_L$ where the pro-domain was removed, since the pro-domain has been shown to reduce solubility (18). In addition, to prevent auto-processing of caspase-8, we mutated the catalytic cysteine to alanine. Previous studies have shown that removal of the DED motifs in the pro-domain does not affect formation of the active homodimer, so the protein folds correctly in the absence of the pro-domain (19).

Sample preparation for equilibrium folding/unfolding

Folding/unfolding experiments were performed as described previously(21). Briefly, stock solutions of urea (10 M), citrate buffer (50 mM sodium citrate/citric acid, pH 3.5–5.5, 1 mM DTT), phosphate buffer (pH 6–8, 1 mM DTT), Tris buffer (50 mM Tris-HCl, pH 8.5-9, 1 mM DTT) were prepared as described (6, 14, 21). For unfolding studies, protein samples were prepared in buffer with urea concentrations ranging from 0 to 9 M.

The buffers used to prepare protein and urea solutions provided a range of pH, from 3.5-9, as shown in the figures. For renaturation experiments, stock protein was first incubated for three hours at 25 °C in an 8 M urea-containing buffer. The unfolded protein was then diluted into the appropriate buffer with urea concentrations ranging from 1 to 8 M were utilized. For all equilibrium folding/unfolding studies, the final protein concentration ranged from 2 to 6 μM. In both denaturation and renaturation studies, samples were incubated for a minimum of 16 hours at 25 °C.

Fluorescence emission and circular dichroism (CD) measurements

Fluorescence emission was acquired using a PTI C-61 spectrofluorometer (Photon Technology International, Birmingham, NJ) from 300 nm to 400 nm following excitation at 280 or 295 nm. Excitation at 280 nm follows tyrosinyl and tryptophanyl fluorescence emission, whereas excitation at 295 nm follows the tryptophanyl fluorescence emission. Experiments on folding/unfolding were carried out as previously described (6, 14, 21).In brief, stock solutions of urea (10 M), citrate buffer (50 mM sodium citrate/citric acid, pH 3.5-5.5, 1 mM DTT), phosphate buffer (pH 6-8, 1 mM DTT), and Tris buffer (50 mM Tris-HCl, pH 8.5-9, 1 mM DTT) were made as indicated (21). Protein samples were incubated in buffer containing urea concentrations ranging from 0 to 9 M for unfolding and at final pHs shown in the figures. For refolding, stock protein was incubated for three hours at 25 °C in buffer containing 8 M urea. The unfolded protein was then diluted into the appropriate buffer with urea concentrations ranging from 1 to 8 M. The final protein concentration ranged from 2 to 6 μM for all equilibrium folding/unfolding investigations. Samples were incubated at 25 °C for at least 16 hours in both denaturation and renaturation tests. CD measurements were recorded using a J-1500 spectropolarimeter (Jasco) between 220 and 240 nm. Fluorescence and CD spectra were measured using a 1 cm path length cuvette and constant temperature (25 °C). All data were corrected for buffer background.

Data analysis and global fits of equilibrium folding/unfolding data

The data were fit globally and interpreted as described previously (21). Briefly, fluorescence emission and CD data were collected between pH 3.5 and 9 for both proteins and at two protein concentrations (2 μM and 6 μM), which resulted in 15 different data sets at each pH. The data were fit to a 2-state or 3-state equilibrium folding model, as described below. The folding data for caspase-8, show a change in slope at around 3M urea from pH6.5 to pH 9, which is much more pronounced with the circular dichroism signal. The midpoint of this slope shifts to lower urea concentration from pH 5 to 3.5, requiring a 3-state model for fitting (Supplemental Fig S3). Therefore, for caspase-8, at all pHs, the data were best fit to a 3-state equilibrium folding model (eq. 1), where the native conformation unfolds to a partially folded intermediate (I) before unfolding (U).

$$N \overset{K_1}{\leftrightarrow} I \overset{K_2}{\leftrightarrow} U \qquad \text{eq. 1}$$

For cFLIP$_L$, the data were best-fit to a 3-state equilibrium folding model between pH 4.5 and pH 9 (eq. 1) since a simple 2-state cannot account for the change in curvature at about 3M-4M urea. Below pH 4.5, the data for cFLIP$_L$ were best fit to a 2-state model, as described by equation 2, in which the native conformation (N) is in equilibrium with the unfolded state (U).

$$N \overset{K_1}{\leftrightarrow} U \qquad \text{eq. 2}$$

Both folding models have been described in detail previously (21). Global fits of the equilibrium folding/unfolding data for both proteins and at each pH were performed in Igor Pro (WaveMetrics, Inc.) using the appropriate folding model from equations 1 or 2, as reported earlier (6, 14, 21).The results of the global fits are shown as the solid lines in the figures, and ΔG^0 and m-values obtained from the fits are provided in Supplemental Tables S1 and S2. Equilibrium constants acquired from the fits are related to the fraction of species (Supplemental Fig S5 and S6), therefore from the

global fits, one can determine the fraction of species, which has been described in detail for a three-state equilibrium model previously (21).

To determine pKa values for pH-dependent transitions, the data were fit as described previously (23, 37). The results of the fits are shown as the dashed lines in the figures, and pKa values obtained from the fits are provided in Supplemental Tables S3 and S4 and are described in the text.

MD simulations with and without urea

Loop regions missing in the NMR structure of caspase-8 (PDB ID: 2k7z) were modelled on Swiss-Model using the template of the NMR structure (38). The H++ server was used to protonate structures at pH 3.5 and 9, the salinity was 0.15 M, the internal dielectric was 10, and the external dielectric was 80. The orientation correction of asparagine, glutamine, and histidine groups based on van der Waals contacts and H-bonding was turned on (39). The urea molecule was created using Avogadro, as described (40). The PDB file of the urea molecule is attached to the supplementary material. Force field parameters for the simulations in urea were adopted from Smith and colleagues (41). The ratio of water and urea molecules added to the system to obtain a concentration of 5M $_{\text{urea was}}$ calculated as described (42). Gromacs molecular dynamics package was used to perform molecular dynamics simulations (43). Briefly, SPC water molecules were replaced with 560 molecules of urea (42). The structure of $cFLIP_L$ (PDB 3H11) was solved previously by X-ray crystallography (44) and shows the protein was cleaved in the inter subunit linker and is in a dimer competent conformation. Thus, we utilized the NMR structure of caspase-8 described above to model the $cFLIP_L$ protomer, also using Swiss-Model as described for caspase-8. The protein was placed in a cubic box of 6x6x6 nm^3 and dissolved either in SPC water or the previously described urea solution. Six sodium ions were used to neutralize the caspase-8 system, while one chloride ion was used to neutralize the $cFLIP_L$ system. Constraints on the positions of all heavy atoms were used for all equilibration runs. The system was then minimized using the steepest-descent method down to a F_{max} (maximum force) > 1000

kJ mol^{-1} nm^{-1}. NVT (constant temperature and volume) and NPT (constant temperature and pressure) equilibration was carried out for 100 ps using leap frog integrator every 2 fs for 50,000 steps. Bond angles and lengths were constrained using LINCS algorithm, and the vervet cut off scheme of 1 nm was used for both electrostatics and van der Waals forces. Temperature was maintained at 300K with a coupling constant of 0.1 ps using the V-rescale thermostats in NVT and NPT, equilibration was performed with a coupling constant of 0.1 ps for temperature. In NPT, equilibration pressure was kept constant at 1bar using Parrinello-Rahman pressure coupling with a 2 ps constant. The Coulomb cutoff distance was 1 nm, the Lennard-Jones cutoff distance was 1 nm, the Fast Fourier Transform grid maximum spacing was 0.16 nm, and the interpolation order was cubic(43, 45). Using the final step of the NPT ensemble, a 200ns production run was carried out., bond angles and lengths were constrained using LINCS algorithm, and the vervet cut off scheme of 0.9 nm was used for both electrostatics and van der Waals forces. Temperature was maintained at 300K with a coupling constant of 0.5 ps using the Nose-Hoover algorithm, pressure was maintained at 1 bar with a coupling constant of 1 ps using Parrinello-Rahman thermostat. The Coulomb cutoff distance was 0.9 nm, the Lennard-Jones cutoff distance was 0.9 nm, the Fast Fourier Transform grid maximum spacing was 0.12 nm, and the interpolation order was cubic(43, 45).

Limited proteolysis coupled with mass spectrometry

Caspase-8 (6 μM) was incubated overnight at 25 °C in a buffer of 50 mM phosphate, pH 7.5, or 50 mM Tris-HCl buffer, pH 9, with 1 mM DTT. Trypsin (0.15 ng) (New England Biolabs) was added following the removal of an aliquot at the zero-time point, and the reaction tube was incubated on a revolving mixer (46). Aliquots were removed every 15 minutes, and the reaction was stopped by adding SDS and incubating at 100 °C for five minutes. Samples were visualized by SDS-PAGE (4-20% gradient gel, Sure Page Gels, GenScript), and band intensity was determined using Image Lab (Bio-Rad). The data for band intensity *versus* time were fit to single exponential equations using Kaleidagraph (Synergy Software).

The samples from the limited proteolysis were prepared for mass spectrometry using Zip tips to remove salts and impurities. On the plate, 1 μL of sample and 1 μL of matrix were spotted and crystallized. The sinnapinic acid matrix was used to analyze molecular weight over 8 kDa and α-Cyano-4-hydroxycinnamic acid (CHCA) matrix was utilized below 6 kDa (47). Samples were analyzed by MALDI-MS (Axima Assurance) in linear mode. The cleavage locations were determined using MS-digest on Protein Prospector software version 6.4.2 (available at https://prospector.ucsf.edu/prospector) (48, 49). Results were mapped onto the modeled structure of caspase-8 using PyMOL molecular visualization system.

Conservation analysis

A total of 1000 sequences of caspases-3, -6, -7, -8, -10 and $cFLIP_L$ were obtained from caspbase (https://caspbase.uta.edu/) or NCBI and trimmed to have equal representations from each taxon in the chordate lineage resulting in 200 sequences (50). Site-specific conservation was determined using the Consurf server (https://consurf.tau.ac.il/consurf-old.php) and to map conservation as B-factors onto structures (51). Results were viewed in Jalview and subsequently in Pymol to generate figures (52).

References

1. van Opdenbosch, N., and Lamkanfi, M. (2019) Caspases in cell death, inflammation, and disease. *Immunity*. **50**, 1352–1364
2. Kumar, S. (2007) Caspase function in programmed cell death. *Cell Death Differ*. **14**, 32–43
3. Elmore, S. (2007) Apoptosis: A review of programmed cell death. *Toxicol Pathol*. **35**, 495–516
4. Anson, F., Thayumanavan, S., and Hardy, J. A. (2021) Exogenous introduction of initiator and executioner caspases results in different apoptotic outcomes. *JACS Au*. **1**, 1240–1256
5. Clark, A. C. (2016) Caspase allostery and conformational selection. *Chem Rev*. **116**, 6666–6706
6. Shrestha, S., and Clark, A. C. (2021) Evolution of the folding landscape of effector caspases. *Journal of Biological Chemistry*. **297**, 1–12
7. Medvedev, K. E., Kinch, L. N., Dustin Schaeffer, R., Pei, J., and Grishin, N. V. (2021) A fifth of the protein world: Rossmann-like proteins as an evolutionarily successful structural unit. *J Mol Biol*. **433**, 1–25
8. Ameisen, J. C. (2002) On the origin, evolution, and nature of programmed cell death: a timeline of four billion years. *Cell Death Differ*. **9**, 367–393
9. Pop, C., Fitzgerald, P., Green, D. R., and Salvesen, G. S. (2007) Role of proteolysis in caspase-8 activation and stabilization. *Biochemistry*. **46**, 4398–4407
10. Walters, J., Pop, C., Scott, F. L., Drag, M., Swartz, P., Mattos, C., Salvesen, G. S., and Clark, A. C. (2009) A constitutively active and uninhibitable caspase-3 zymogen efficiently induces apoptosis. *Biochem. J*. **424**, 335–345
11. Hillert, L. K., Ivanisenko, N. V., Busse, D., Espe, J., König, C., Sergey, P. E., Kolchanov, N. A., Ivanisenko, V. A., and Lavrik, I. N. (2020) Dissecting DISC regulation via pharmacological targeting of caspase-8/c-cFLIPL heterodimer. *Cell Death Differ*. **27**, 2117–2130
12. Tsuchiya, Y., Nakabayashi, O., Nakano, H., and Lemarié, A. (2015) FLIP the switch: Regulation of apoptosis and necroptosis by cFLIP. *Int J Mol Sci*. **16**, 30321–30341

13. Aravind, L., and Koonin, E. V. (2002) Classification of the caspase–hemoglobinase fold: Detection of new families and implications for the origin of the eukaryotic separins. *Proteins*. **46**, 355–367
14. Bose, K., and Clark, A. C. (2001) Dimeric procaspase-3 unfolds via a four-state equilibrium process. *Biochemistry*. **40**, 14236–14242
15. Yao, L., and Clark, A. C. (2022) Comparing the folding landscapes of evolutionarily divergent procaspase-3. *Biosci Rep*. **42**, 1–13
16. Grinshpon, R. D., Shrestha, S., Titus-Mcquillan, J., Hamilton, P. T., Swartz, P. D., and Clark, A. C. (2019) Resurrection of ancestral effector caspases identifies novel networks for evolution of substrate specificity. *Biochem J*. **476**, 3475–3492
17. Bose, K., and Clark, A. C. (2005) pH effects on the stability and dimerization of procaspase-3. *Protein Sci*. **14**, 24–36
18. Ma, C., MacKenzie, S. H., and Clark, A. C. (2014) Redesigning the procaspase-8 dimer interface for improved dimerization. *Protein Science*. **23**, 442–453
19. Keller, N., Mareš, J., Zerbe, O., and Grütter, M. G. (2009) Structural and biochemical studies on procaspase-8: New insights on initiator caspase activation. *Structure*. **17**, 438–448
20. Boatright, K. M., Deis, C., Denault, J.-B., Sutherlin, D. P., and Salvesen, G. S. (2004) Activation of caspases-8 and-10 by FLIPL. *Biochem. J*. **382**, 651–657
21. Walters, J., Milam, S. L., and Clark, A. C. (2009) Practical approaches to protein folding and assembly: Spectroscopic strategies in thermodynamics and kinetics. *Methods Enzymol*. **455**, 1–39
22. Scholtz, J. M., Grimsley, G. R., and Pace, C. N. (2009) Solvent denaturation of proteins and interpretations of the m value. *Methods Enzymol*. **466**, 549–565
23. Bose, K., Pop, C., Feeney, B., and Clark, A. C. (2003) An uncleavable procaspase-3 mutant has a lower catalytic efficiency but an active site similar to that of mature caspase-3. *Biochemistry*. **42**, 12298–12310
24. Crewther, W. C. (1953) The effect of pH and cations on the thermal denaturation of trypsin. *Aust J Biol Sci*. **6**, 597–616
25. Sikosek, T., and Chan, H. S. (2014) Biophysics of protein evolution and evolutionary protein biophysics. *J R Soc Interface*. **11**, 1–35

26. Ameisen, J. C. (2002) On the origin, evolution, and nature of programmed cell death: a timeline of four billion years. *Cell Death Differ*. **9**, 367–393
27. Kazuhiro Sakamaki (2014) The apoptotic initiator caspase-8: Its functional ubiquity and genetic diversity during animal evolution. *Mol Biol Evol*. **31**, 3282–3301
28. Pop, C., Chen, Y.-R., Smith, B., Bose, K., Bobay, B., Tripathy, A., Franzen, S., and Clark, A. C. (2001) Removal of the pro-domain does not affect the conformation of the procaspase-3 dimer. *Biochemistry*. **40**, 14224–14235
29. Pop, C., Oberst, A., Drag, M., van Raam, B. J., Riedl, S. J., Green, D. R., and Salvesen, G. S. (2011) FLIP(L) induces caspase 8 activity in the absence of interdomain caspase 8 cleavage and alters substrate specificity. *Biochem. J*. **433**, 447–457
30. Mackenzie, S. H., Schipper, J. L., England, E. J., Thomas, M. E., Blackburn, K., Swartz, P., and Clark, A. C. (2013) Lengthening the intersubunit linker of procaspase 3 leads to constitutive activation. *Biochemistry*. **52**, 6219–6231
31. Casey, J. R., Grinstein, S., and Orlowski, J. (2009) Sensors and regulators of intracellular pH. *Nature Reviews Molecular Cell Biology 2009 11:1*. **11**, 50–61
32. Swietach, P., Vaughan-Jones, R. D., Harris, A. L., and Hulikova, A. (2014) The chemistry, physiology and pathology of pH in cancer. *Philos Trans R Soc Lond B Biol Sci*. **369**, 1–9
33. Thomas, M. E., Grinshpon, R., Swartz, P., and Clark, A. C. (2018) Modifications to a common phosphorylation network provide individualized control in caspases. *Journal of Biological Chemistry*. **293**, 5447–5461
34. Alvarado-Kristensson, M., Melander, F., Leandersson, K., Rönnstrand, L., Wernstedt, C., and Andersson, T. (2004) p38-MAPK Signals Survival by phosphorylation of caspase-8 and caspase-3 in human neutrophils. *J Exp Med*. **199**, 449
35. Li, X., Wen, W., Liu, K., Zhu, F., Malakhova, M., Peng, C., Li, T., Kim, H. G., Ma, W., Yeon Cho, Y., Bode, A. M., Dong, Z., and Dong, Z. (2011) Phosphorylation of caspase-7 by p21-activated protein kinase (PAK) 2 inhibits chemotherapeutic drug-induced apoptosis of breast cancer cell lines. *Journal of Biological Chemistry*. **286**, 22291–22299

36. Zhou, Q., Snipas, S., Ortht, K., Muzio, M., Dixit, V. M., and Salvesen, G. S. (1997) Target protease specificity of the viral serpin CrmA: Analysis of five caspases. *Journal of Biological Chemistry*. **272**, 7797–7800
37. Bhuyan, A. K., and Udgaonkar, J. B. (2001) Folding of horse cytochrome c in the reduced state. *J Mol Biol*. **312**, 1135–1160
38. Waterhouse, A., Bertoni, M., Bienert, S., Studer, G., Tauriello, G., Gumienny, R., Heer, F. T., de Beer, T. A. P., Rempfer, C., Bordoli, L., Lepore, R., and Schwede, T. (2018) SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic Acids Res*. **46**, W296–W303
39. Gordon, J. C., Myers, J. B., Folta, T., Shoja, V., Heath, L. S., and Onufriev, A. (2005) H++: a server for estimating pKas and adding missing hydrogens to macromolecules. *Nucleic Acids Res*. **33**, W368–W371
40. Hanwell, M. D., Curtis, D. E., Lonie, D. C., Vandermeerschd, T., Zurek, E., and Hutchison, G. R. (2012) Avogadro: An advanced semantic chemical editor, visualization, and analysis platform. *J Cheminform*. **4**, 1–17
41. Smith, L. J., Berendsen, H. J. C., and van Gunsteren, W. F. (2004) Computer simulation of urea-water mixtures: A test of force field parameters for use in biomolecular simulation. *J Phys Chem*. **108**, 1065–1071
42. Stumpe, M. C., and Grubmu, H. (2007) Aqueous urea solutions: structure, energetics, and urea aggregation. *J Phys Chem B*. **111**, 6220–6228
43. Abraham, M. J., Murtola, T., Schulz, R., Páll, S., Smith, J. C., Hess, B., and Lindah, E. (2015) Gromacs: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX*. **1–2**, 19–25
44. Yu, J. W., Jeffrey, P. D., and Shi, Y. (2009) Mechanism of procaspase-8 activation by c-FLIPL. *Proc Natl Acad Sci U S A*. **106**, 8169–8174
45. Rocco, A. G., Mollica, L., Ricchiuto, P., Baptista, A. M., Gianazza, E., and Eberini, I. (2008) Characterization of the protein unfolding processes induced by urea and temperature. *Biophys J*. **94**, 2241–2251
46. Pop, C., Feeney, B., Tripathy, A., and Clark, A. C. (2003) Mutations in the procaspase-3 dimer interface affect the activity of the zymogen. *Biochemistry*. **42**, 12311–12320

47. Giebel, R., Worden, C., Rust, S. M., Kleinheinz, G. T., Robbins, M., and Sandrin, T. R. (2010) Microbial fingerprinting using matrix-assisted laser desorption ionization time-of-flight mass spectrometry (MALDI-TOF MS): Applications and challenges. *Adv Appl Microbiol*. **71**, 149–184
48. Chalkley, R. J., Hansen, K. C., and Baldwin, M. A. (2005) Bioinformatic methods to exploit mass spectrometric data for proteomic applications. *Methods Enzymol*. **402**, 289–312
49. Macur, K., Hagen, L., Ciesielski, T. M., Konieczna, L., Skokowski, J., Jenssen, B. M., Slupphaug, G., and Bączek, T. (2019) A targeted mass spectrometry immunoassay to quantify osteopontin in fresh-frozen breast tumors and adjacent normal breast tissues. *J Proteomics*. **208**, 103469
50. Grinshpon, R. D., Williford, A., Titus-Mcquillan, J., and Clark, A. C. (2018) The CaspBase: A curated database for evolutionary biochemical studies of caspase functional divergence and ancestral sequence inference. *Protein Sci*. **27**, 1857–1870
51. Landau, M., Mayrose, I., Rosenberg, Y., Glaser, F., Martz, E., Pupko, T., and Ben-Tal, N. (2005) ConSurf 2005: The projection of evolutionary conservation scores of residues on protein structures. *Nucleic Acids Res*. **33**, W299–W302
52. Waterhouse, A. M., Procter, J. B., Martin, D. M. A., Clamp, M., and Barton, G. J. (2009) Sequence analysis Jalview Version 2-a multiple sequence alignment editor and analysis workbench. *BIOINFORMATICS*. **25**, 1189–1191

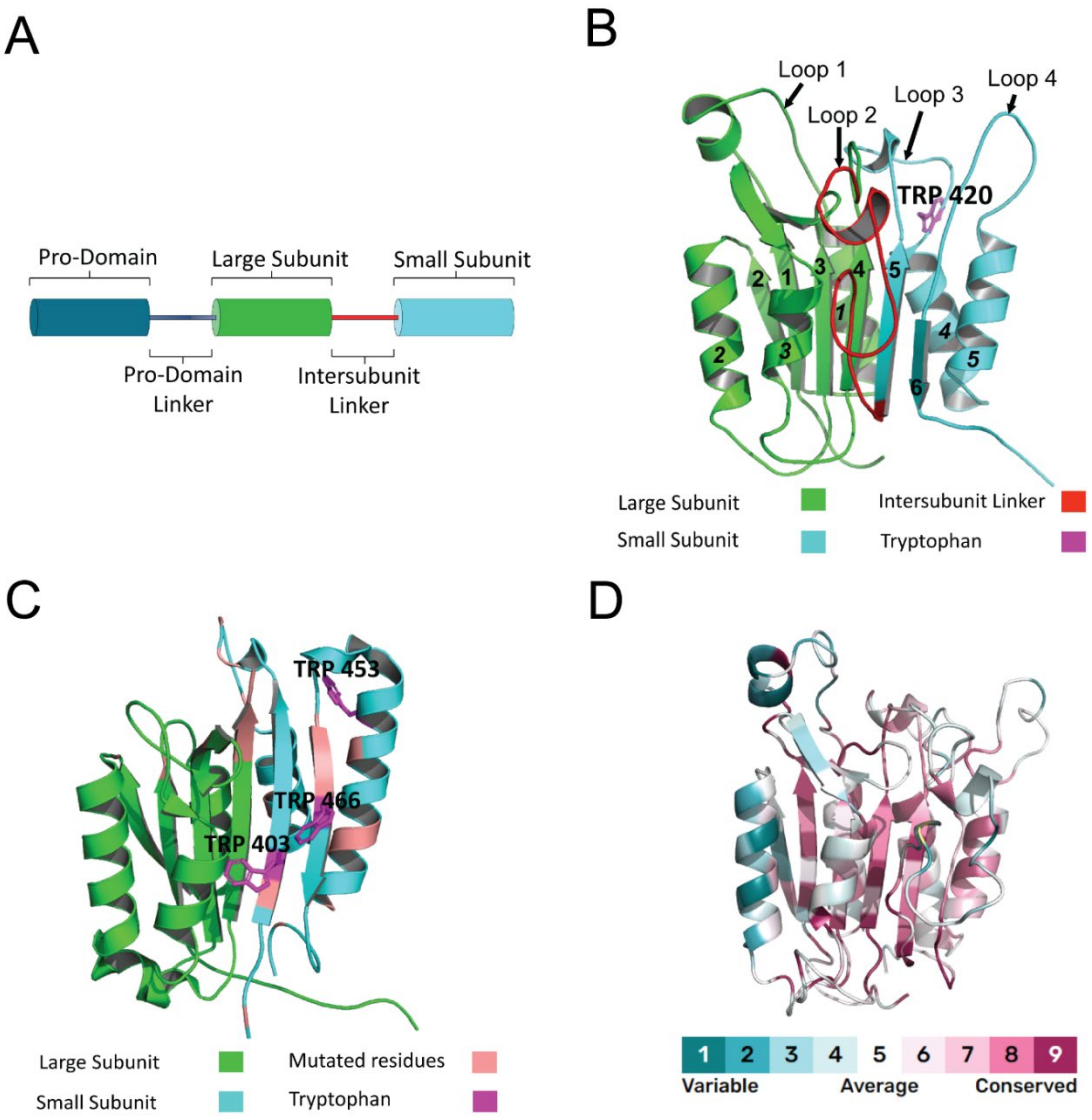


Figure 1.

Caspase structure and conservation. **(A)** Domain organization of initiator caspases**.** **(B)** Structure of caspase-8 (PDB ID: 6PX9) showing sequence of beta sheets (bold numbers) and helices (bold italics) in the protease domain. Large subunit, intersubunit linker, and small subunit are color-coded in accordance with the legend. Additionally, the only tryptophan in caspase 8 is labeled in magenta. **(C)** Structure of $cFLIP_L$ (PDB ID: 3H11) showing tryptophan residues in magenta. Residues that are conserved in all caspases but mutated in $cFLIP_L$ are shown in salmon. Large subunit and small subunit are color-coded as per the legend whereas the inter-subunit linker is missing as the crystal structure is that of processed mature enzyme. **(D)** Amino acid conservation of chordate caspases-3, -6, -7, -8, -10, and cFLIPL is illustrated on the structure of caspase-8 using the consurf conservation coloring scheme on the structure of procaspase-8 (6px9)

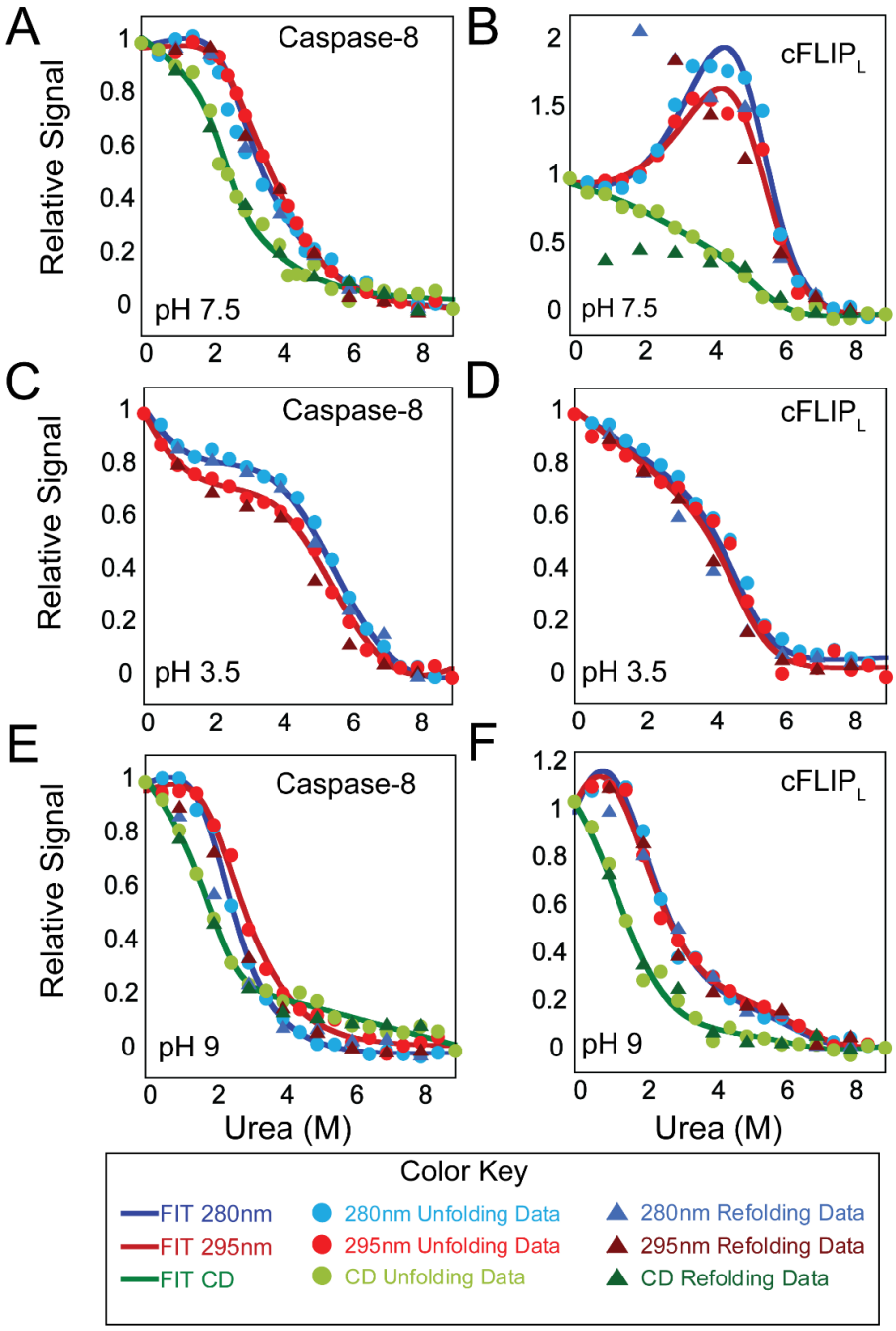


Figure 2. Equilibrium unfolding of caspase-8 and cFLIP$_L$ at pH 3.5, 7.5 and 9. Equilibrium unfolding of caspase-8 **(panels A, B, C)** and cFLIP$_L$ **(panels D, E, H)** monitored by fluorescence emission with excitation at 280 nm (●), 295 (●), and CD (●). Refolding of caspase 8 monitored by fluorescence at 280 nm (▲),295 (▲), and CD (▲). As described in the text, solid lines represent global fits to fluorescence emission data at 280nm (▬), 295nm (▬), and CD (▬). Thermodynamic parameters obtained from the fits are described in Supplemental Tables S1 and S2.

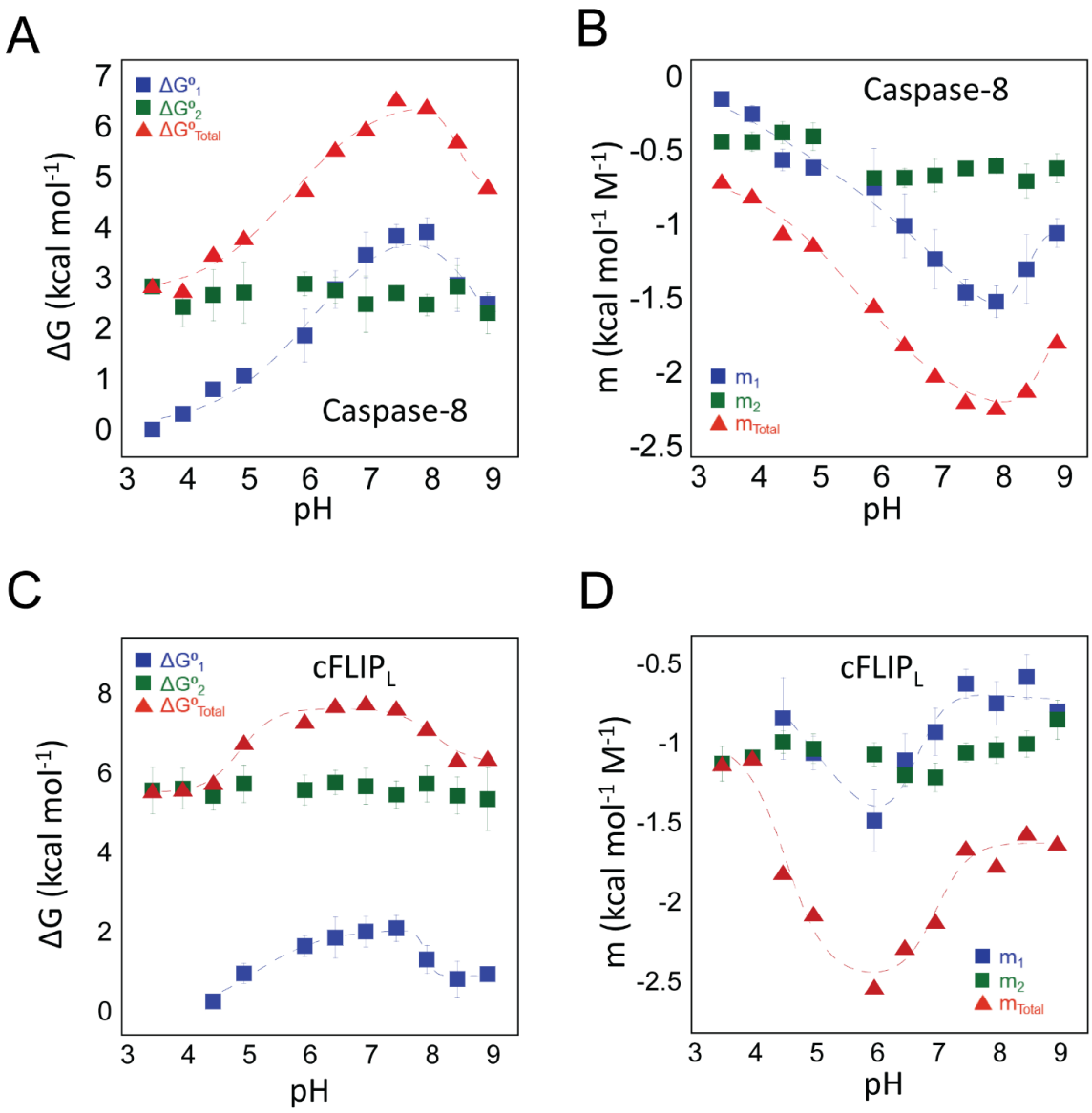


Figure 3: pH-dependent changes in conformational free energy of caspase-8 and cFLIP$_L$. Conformational free energies of **(A)** caspase-8 and **(C)** cFLIP$_L$ depicting the experimentally determined free energy values for native state - ΔG^o_1 (■), intermediate state - ΔG^o_2 (■), and total free energy - ΔG^o_{Total} (▲). Experimentally determined m-values for **(B)** caspase-8 and **(D)** cFLIP$_L$ depecting values for native state - m_1 (■), intermediate state - m_2 (■), and total m-value - m_{total} (▲). Dashed lines in panels A-D represent fits of the data to determine pKa values, as described in the text.

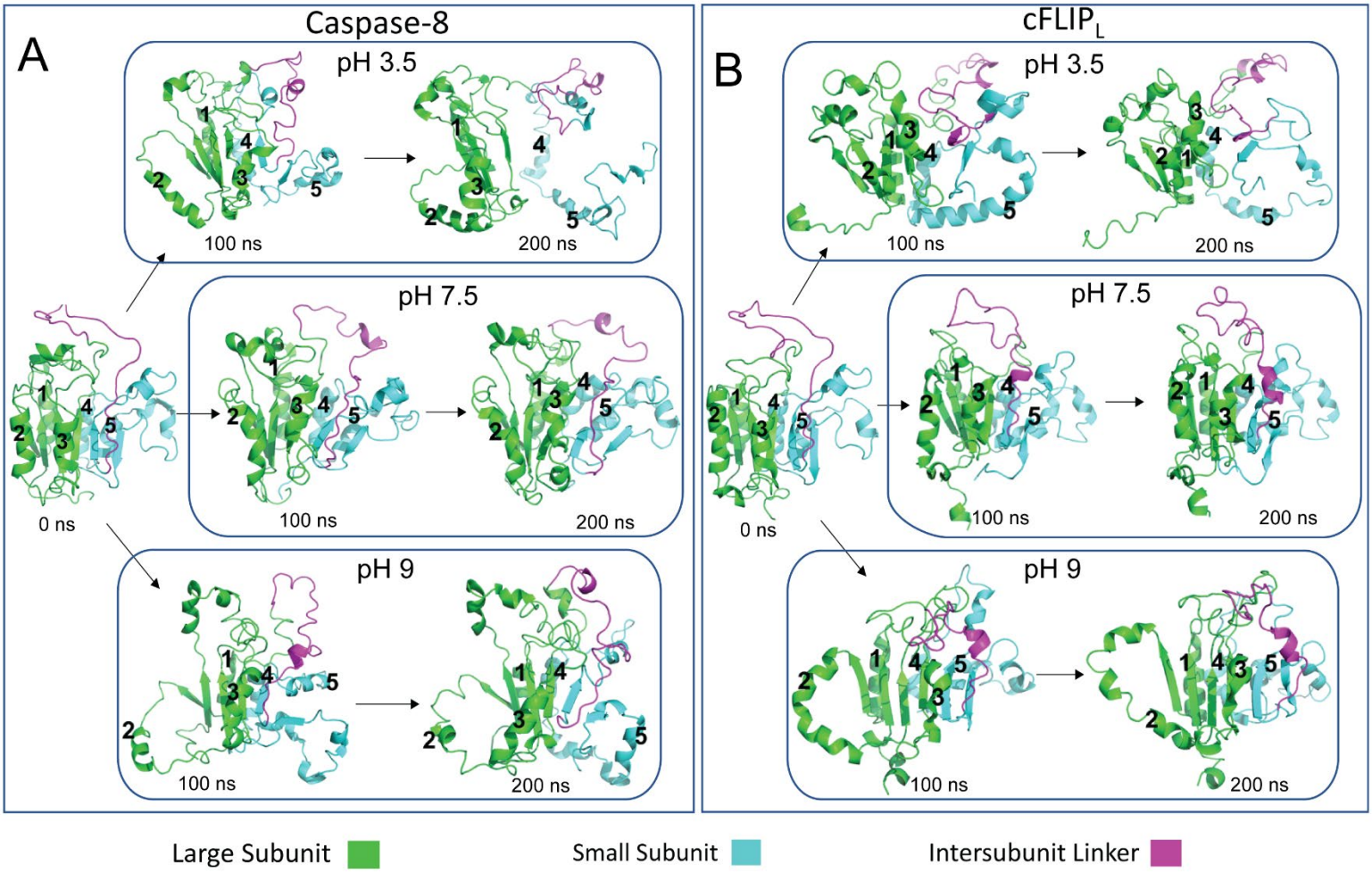


Figure 4: Molecular dynamics simulations of caspase-8 (A) and of cFLIP$_L$ **(B)** in 5 M urea at three different protonation states representing pH 3.5, pH 7.5 and pH 9 . For panels A and B, time points of zero, 100 ns, and 200 ns are shown. The large subunit and small subunit of caspase-8 and cFLIP$_L$, are shown in green and cyan, whereas the intersubunit linker is shown in magenta (refer color key inset). Helices 1 (back), 2(front) and 3(front) on the surface of the large subunit and helices 4(back) and 5(back) on the small subunit are labelled.

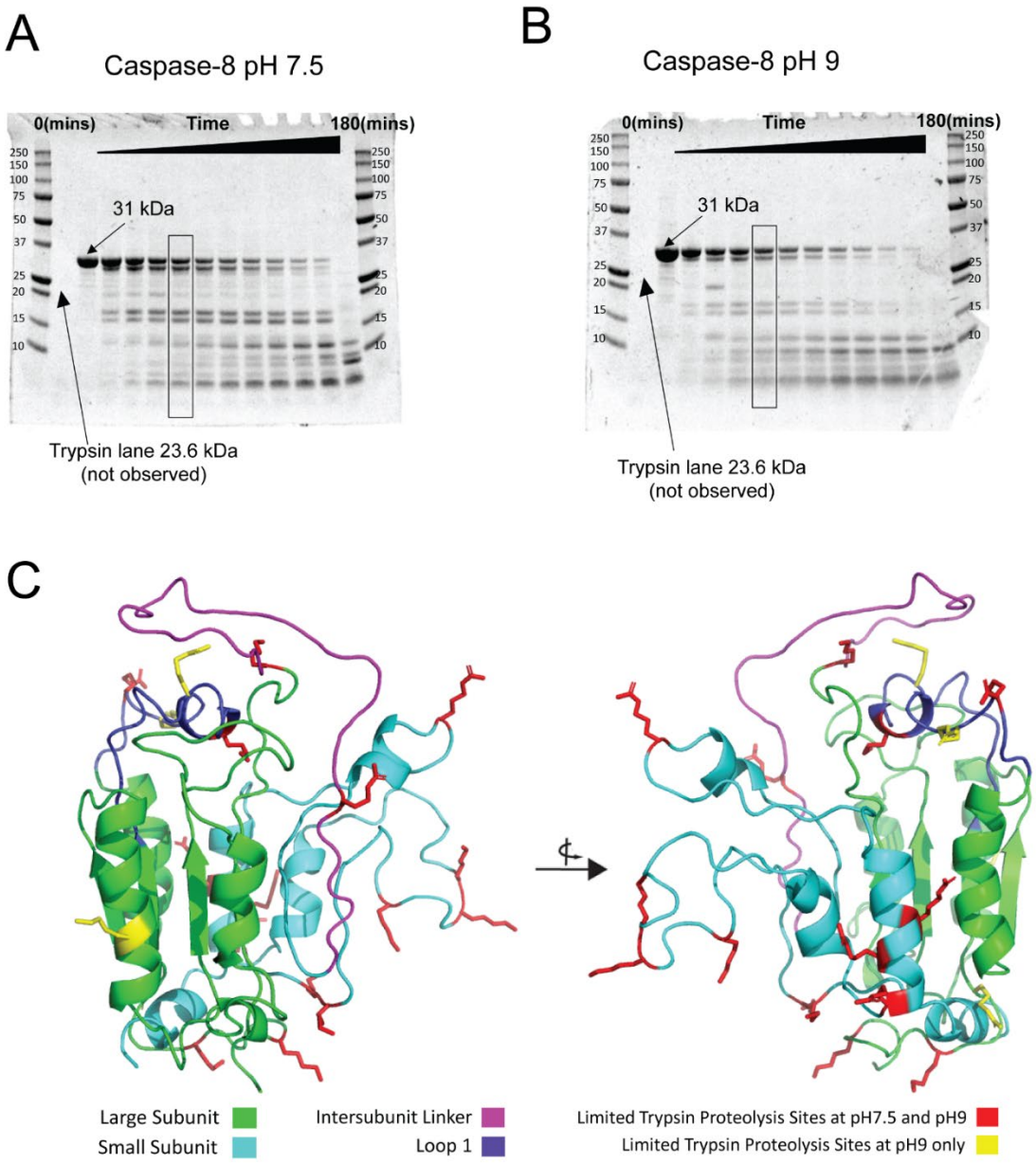


Figure 5. **Limited trypsin proteolysis of caspase-8 at pH 7.5 and pH 9. (A)** Results of cleavage of caspase-8 (31 kDa) at pH 7.5 **(B)** Results of cleavage of caspase-8 (31 kDa) at pH 9. Sizes of fragments are labeled. For panels A and B, molecular weight markers are shown in the first and last lanes, and O/N refers to samples incubated overnight. The fifth lane which refers to the products analyzed is boxed. **(C)** Cleavage products of trypsin proteolysis at pH7.5 and pH9, respectively, were determined by mass spectrometry as described in the text and represented on the structure of caspase-8 (PDB ID: 2K7Z). The small subunit is depicted in cyan, while the large subunit is depicted in green. Cleavage sites at both pH 7.5 and pH 9 are shown by red side chains, whereas cleavage products at pH 9 are represented by yellow side chains. The intersubunit linker is depicted in magenta, while loop 1 is indicated in purple.

Conserved folding landscape of monomeric initiator caspases

Mithun Nag,[1,] and A. Clay Clark[1,*]

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

Running title: Sequential folding of monomeric caspases

[*]Corresponding author: A. Clay Clark

Email: clay.clark@uta.edu

Key Words: caspase, apoptosis; folding landscape, protein evolution, evolutionary biology, apoptosis, protease, protein folding, fluorescence emission, circular dichroism, molecular dynamic

Supporting Information

Supplemental Table S1. Changes in free energy for each stage of caspase-8 folding/unfolding from pH 3.5 to pH 9. The errors represent the standard deviation from the mean (mean±SD).

Enzyme	pH	Equilibrium Mechanism	ΔG1 (kcal mol^{-1})	m1 (kcal mol^{-1} M^{-1})	ΔG2 (kcal mol^{-1})	m2 (kcal mol^{-1} M^{-1})	ΔG Total (kcal mol^{-1})	m Total (kcal mol^{-1} M^{-1})
Caspase-8	3.5	N↔I↔U	0.2 ± 0.02	-0.27 ± 0.01	2.6 ± 0.12	-0.50 ± 0.01	2.8 ± 0.14	-0.77 ± 0.03
	4	N↔I↔U	0.3 ± 0.14	-0.34 ± 0.05	2.3 ± 0.37	-0.51 ± 0.06	2.6 ± 0.51	-0.85 ± 0.11
	4.5	N↔I↔U	0.8 ± 0.07	0.62 ± 0.07	2.5 ± 0.48	-0.45 ± 0.07	3.3 ± 0.44	-1.08 ± 0.13
	5	N↔I↔U	1.0 ± 0.09	-0.67 ± 0.04	2.6 ±0.57	-0.48 ± 0.09	3.6 ± 0.50	-1.15 ± 0.13
	6	N↔I↔U	1.8 ± 0.50	-0.80 ± 0.24	2.8 ±0.22	-0.74 ± 0.05	4.6 ± 0.72	-1.54 ± 0.29
	6.5	N↔I↔U	2.7 ± 0.35	-1.04 ± 0.20	2.6 ± 0.25	-0.74 ± 0.06	5.3 ± 0.65	-1.78 ± 0.26
	7	N↔I↔U	3.3 ± 0.43	-1.25 ± 0.18	2.4 ± 0.53	-0.72 ± 0.10	5.7 ± 0.41	-1.98 ± 0.29
	7.5	N↔I↔U	3.7 ± 0.21	-1.46 ± 0.08	2.6 ± 0.14	-0.68 ± 0.03	6.3 ± 0.48	-2.10 ± 0.06
	8	N↔I↔U	3.7 ± 0.27	-1.52 ± 0.09	2.4 ±0.20	-0.66 ± 0.05	6.1 ± 0.59	-2.18 ± 0.15
	8.5	N↔I↔U	2.8 ± 0.50	-1.31 ± 0.22	2.7 ±0.40	-0.76 ± 0.11	5.5 ± 0.90	-2.07 ± 0.32
	9	N↔I↔U	2.4 ± 0.09	-1.10 ± 0.09	2.2 ± 0.3	-0.60 ± 0.09	4.6 ± 0.63	-1.70 ± 0.09

Supplemental Table S2. Changes in free energy for each stage of $cFLIP_L$ folding/unfolding from pH 3.5 to pH 9. Note that the thermodynamic parameters

Enzyme	pH	Equilibrium Mechanism	ΔG1 (kcal mol^{-1})	m1 (kcal mol^{-1} M^{-1})	ΔG2 (kcal mol^{-1})	m2 (kcal mol^{-1} M^{-1})	ΔG Total (kcal mol^{-1})	m Total (kcal mol^{-1} M^{-1})
$cFLIP_L$	3.5	I↔U			5.7 ± 0.59	-1.10 ± 0.10	5.7 ± 0.59	-1.10 ± 0.10
	4	I↔U			5.7 ± 0.51	-1.06 ± 0.02	5.7 ± 0.51	-1.06 ± 0.02
	4.5	N↔I↔U	0.3 ± 0.17	-0.82 ± 0.26	5.6 ± 0.36	-0.97 ± 0.07	5.9 ± 0. 26	-1.78 ± 0.16
	5	N↔I↔U	1.0 ± 0.25	-1.03 ± 0.11	5.8 ±0.47	-1.01 ± 0.09	6.8 ± 0.36	-2.05 ± 0.10
	6	N↔I↔U	1.7 ± 0.26*	-1.46 ± 0.19*	5.7 ±0.38	-1.05 ± 0.07	7.4 ± 0.32*	-2.51 ± 0.13*
	6.5	N↔I↔U	1.9 ± 0.51*	-1.08 ± 0.16*	5.9 ± 0.31	-1.18 ± 0.06	7.8 ± 0.41*	-2.25 ± 0.11*
	7	N↔I↔U	2.1 ± 0.38*	-0.90 ± 0.15*	5.8 ± 0.45	-1.19 ± 0.09	7.9 ± 0.42*	-2.09 ± 0.12*
	7.5	N↔I↔U	2.2 ± 0.33*	-0.60 ± 0.09*	5.5 ± 0.34	-1.03 ± 0.06	7.7 ± 0.34*	-1.63 ± 0.07*
	8	N↔I↔U	1.4 ± 0.36*	-0.72 ± 0.14*	5.8 ±0.46	-1.02 ± 0.08	7.2 ± 0.41*	-1.74 ± 0.11*
	8.5	N↔I↔U	0.9 ± 0.45*	-0.56 ± 0.14*	5.5 ±0.47	-0.98 ± 0.08	6.4 ± 0.46*	-1.54 ± 0.11*
	9	N↔I↔U	1.0 ± 0.06	-0.77 ± 0.02	5.4 ± 0.79	-0.82 ± 0.12	6.4 ± 0.43	-1.60 ± 0.07

reported with asterix (*) are uncertain due to hysteresis. The errors represent the standard deviation from the mean (mean±SD).

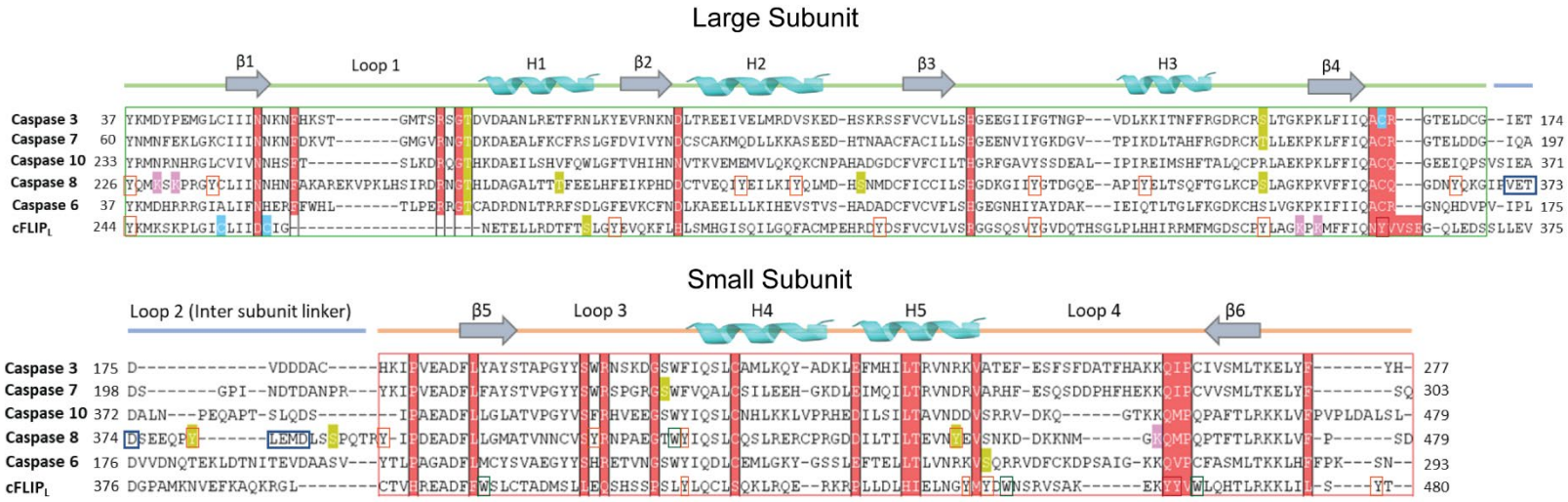
Supplemental Table S3. Determination of pKa for transitions of caspase-8.

Protein	Parameter	pka1	error pKa1	pka2	error pKa2
Caspase 8	ΔG_1	6.1	1.6	8.7	2.7
	m_1	6.1	1.9	8.6	0.8
	ΔG_2	-	-	-	-
	m_2	-	-	-	-
	Total ΔG	5.7	0.3	8.4	0.2
	Total m	5.7	0.3	8.7	2.2
	Native Fraction of Species	4.7	0.2	8.3	1.5
	Intermediate Fraction of Species	-	-	-	-
	Unfolded Fraction of Species	5.75	1.1	7.1	4
	Circular Dichroism	6.3	0.1	8.6	15
Average		5.7	0.8	8.3	3.7

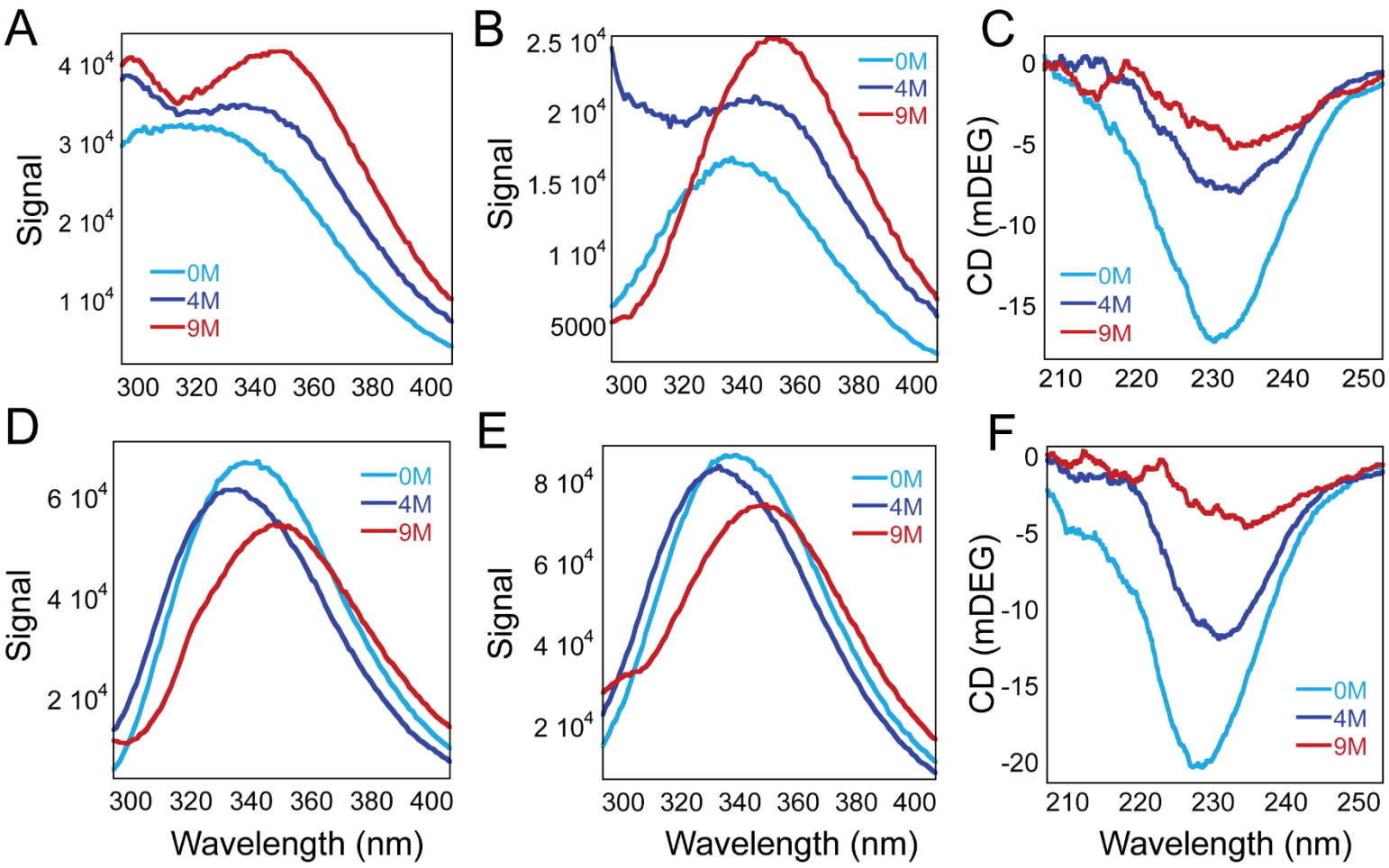
Supplemental Table S4. Determination of pKa for transitions of cFLIPL.

Protein	Parameter	pka1	error pKa1	pka2	error pKa2
$cFLIP_L$	ΔG_1	5.7	0.9	8.1	0.5
	m_z	4.6	0.3	7.6	1.1
	ΔG_2	-	-	-	-
	m_2	-	-	-	-
	Total ΔG	5.1	0.7	8.1	0.7
	Total m	4.5	0.2	7.0	0.3
	Native Fraction of Species	6.3	0.9	8.1	1.1
	Intermediate Fraction of Species	-	-	-	-
	Unfolded Fraction of Species	6.1	1.1	8.0	0.2
	Circular Dichroism	5.7	0.2	8.5	0.6
Average		5.4	0. 6	7.9	0.6

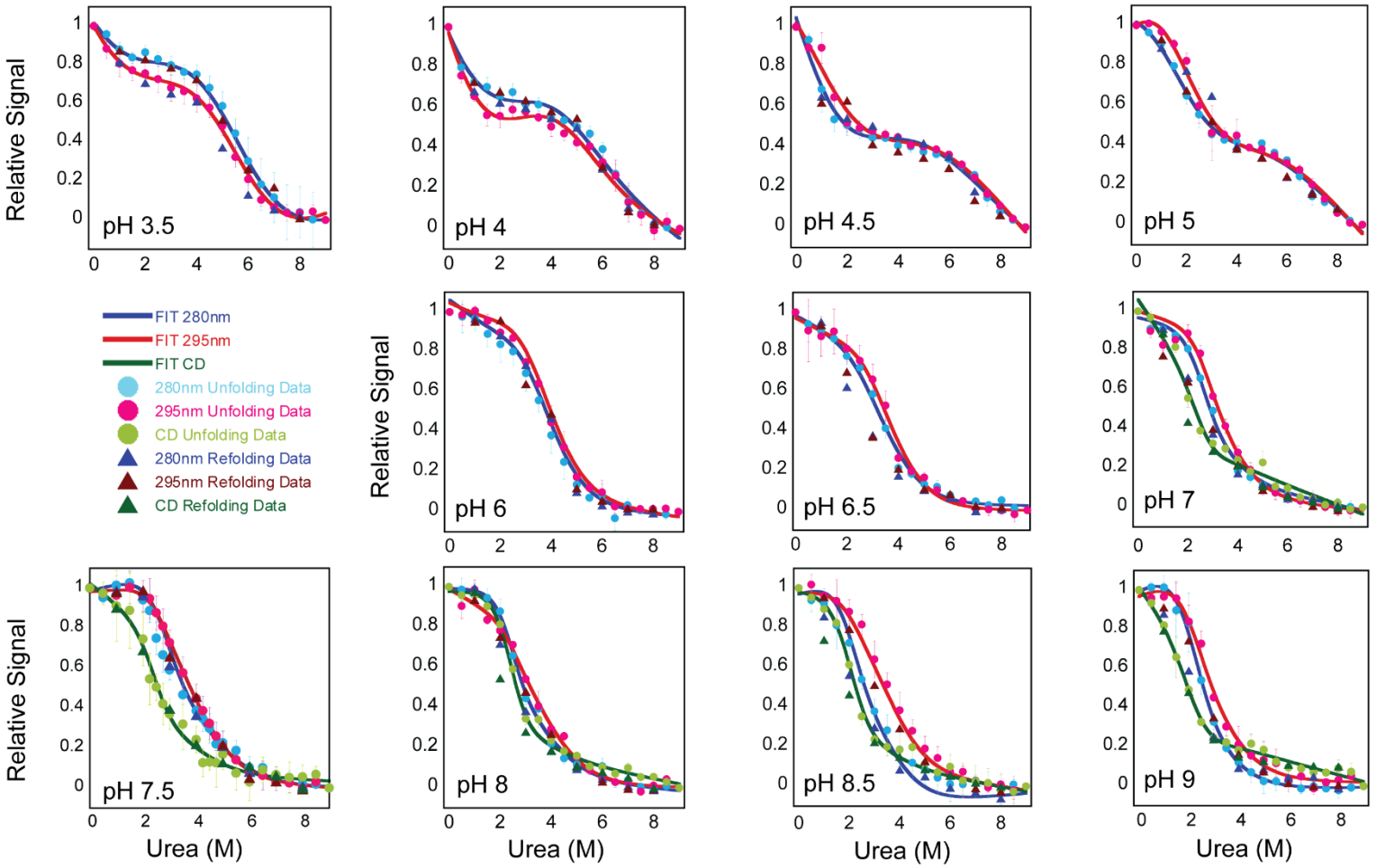
A



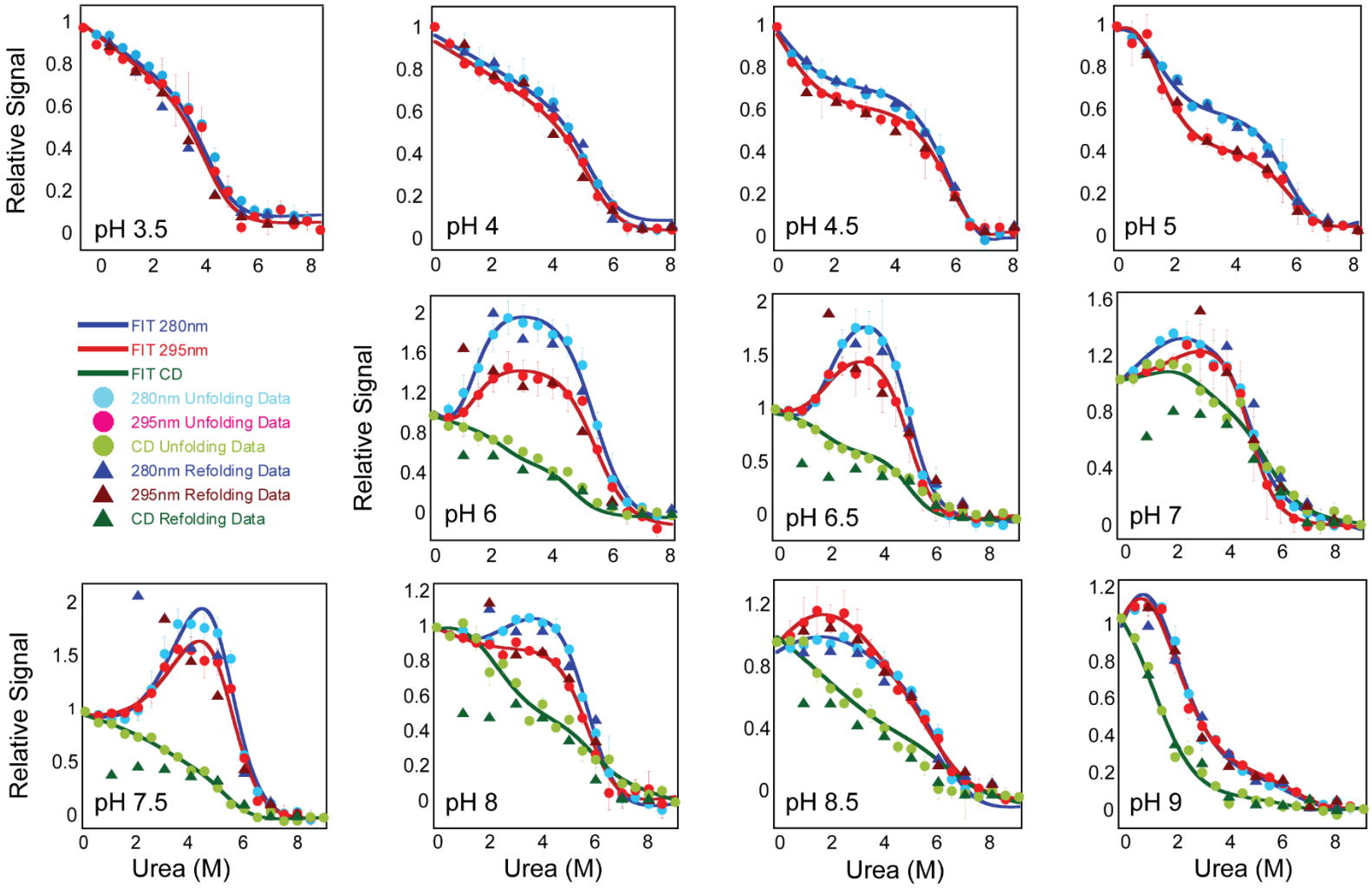
Supplemental Figure S1. Conservation, post- translational modifications and residues involved in active site stabilization. Multiple sequence alignment of caspases-3, -6, -7, -8, -10 and cFLIP$_L$. Boxed regions in red on alignment represent residues that are conserved in all caspases but mutated in cFLIP$_L$. The green box shows the large subunit, and the red box shows the small subunit. The cleavage sites in the intersubunit linker of caspase-8 are boxed in blue. Amino acids boxed in orange and green only for caspase-8 and cFLIP$_L$ represent tyrosine and tryptophan residues, respectively. Amino acids highlighted in green are phosphorylation sites, sites of ubiquitylation are shown in pink, and sites of nitrosylation are shown in blue.



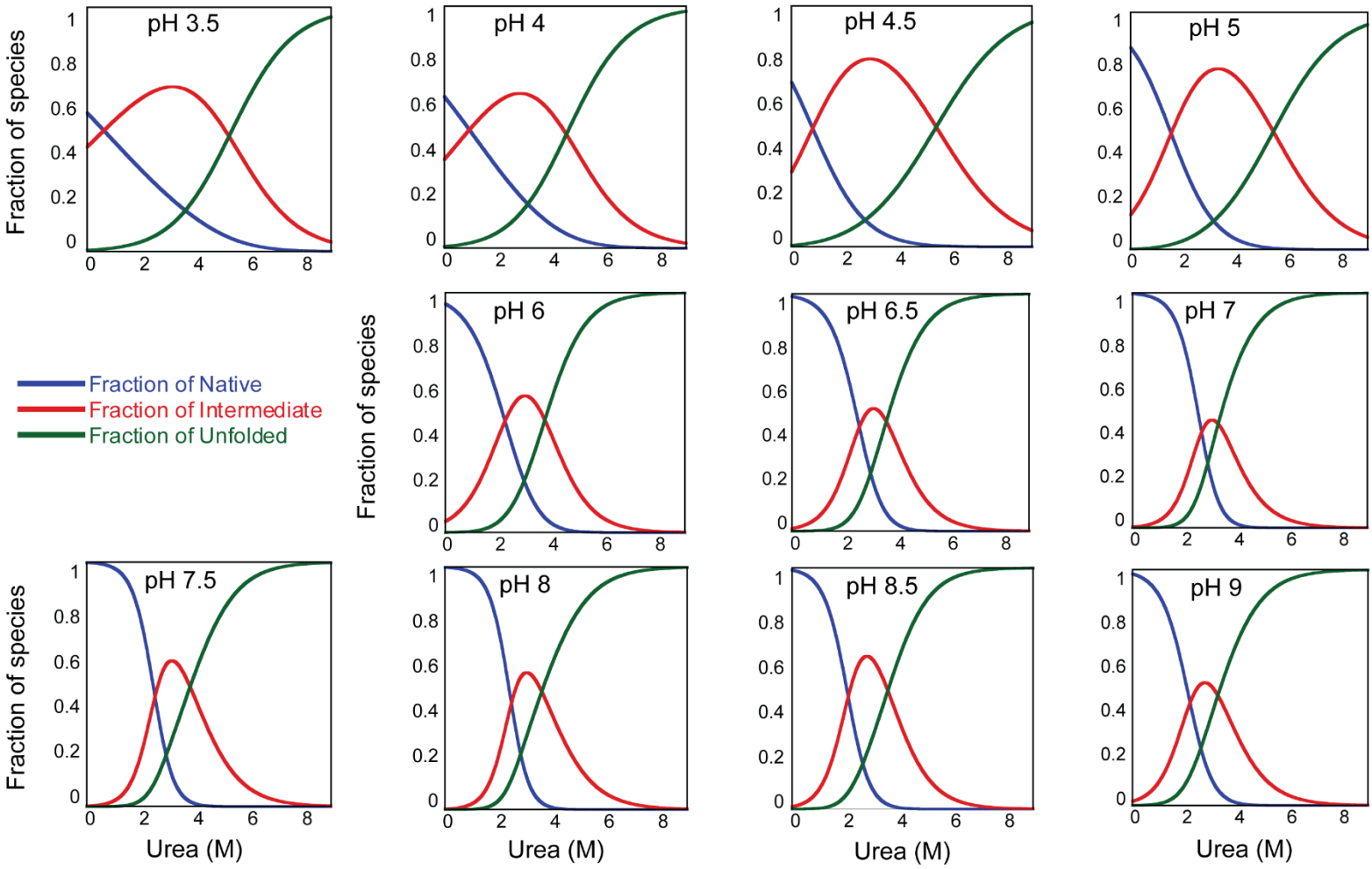
Supplemental Figure S2. Caspase-8 and cFLIP$_L$ fluorescence emission and circular dichroism spectra at pH7.5. Caspase-8 fluorescence emission spectra in 0 M, 4 M, and 8 M ure-containing buffer when excited at 280 nm **(A)**, 295 nm **(B)**, and circular dichroism spectra **(C)**. cFLIP$_L$ emission spectra in 0 M, 4 M, and 8 M urea-containing buffer when excited at 280 nm **(D)**, 295 nm **(E)**, and circular dichroism spectra **(F)**.



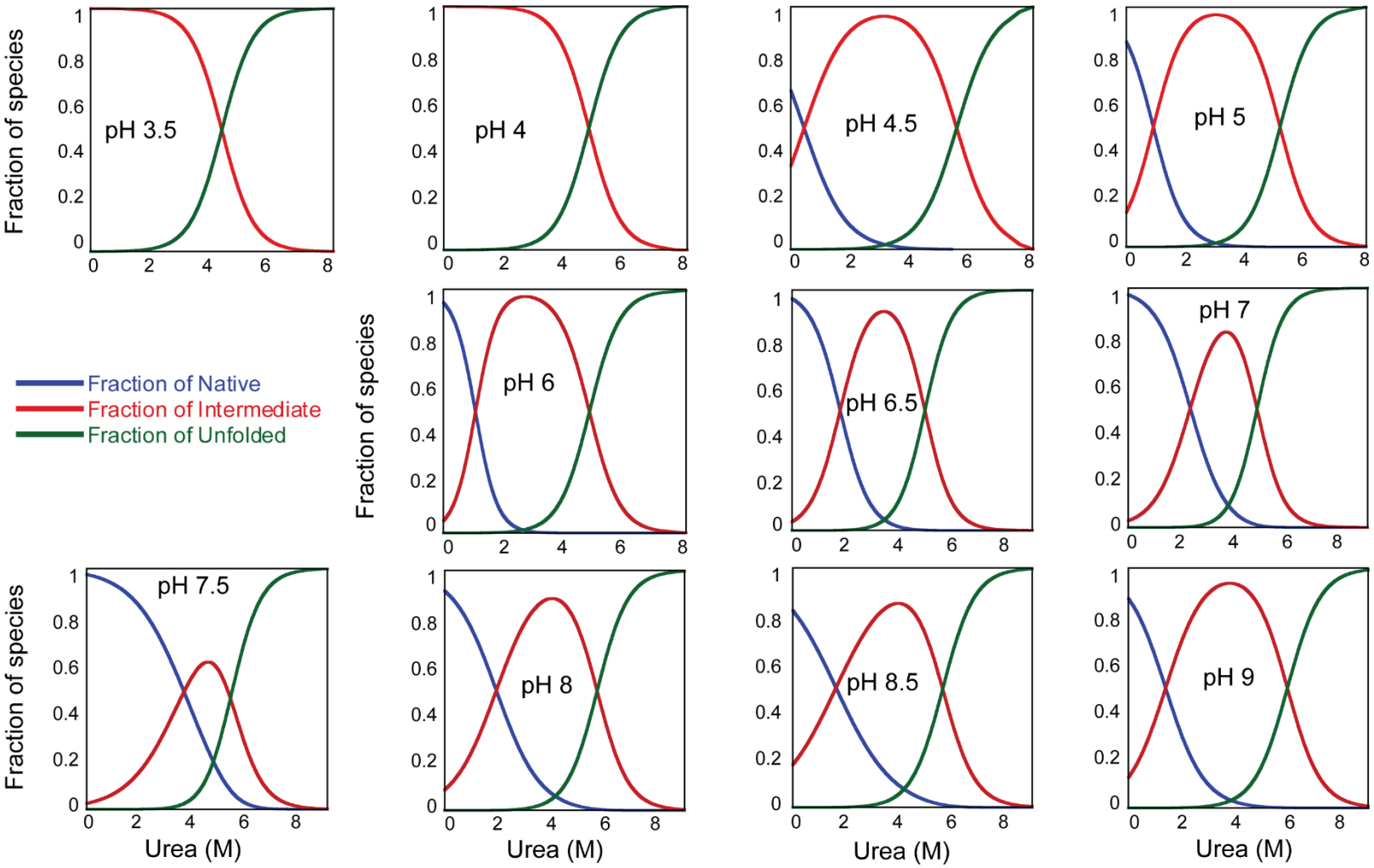
Supplemental Figure S3. Equilibrium unfolding monitored by fluorescence emission and circular dichroism for caspase-8 from pH 3.5 to 9. Solid lines represent global fits of the data as described in the text. Circles represent averaged unfolding data and triangles represent refolding data following fluorescence emission after excitation at 280 nm (blue) or 295 nm (red), and circular dichroism (green).



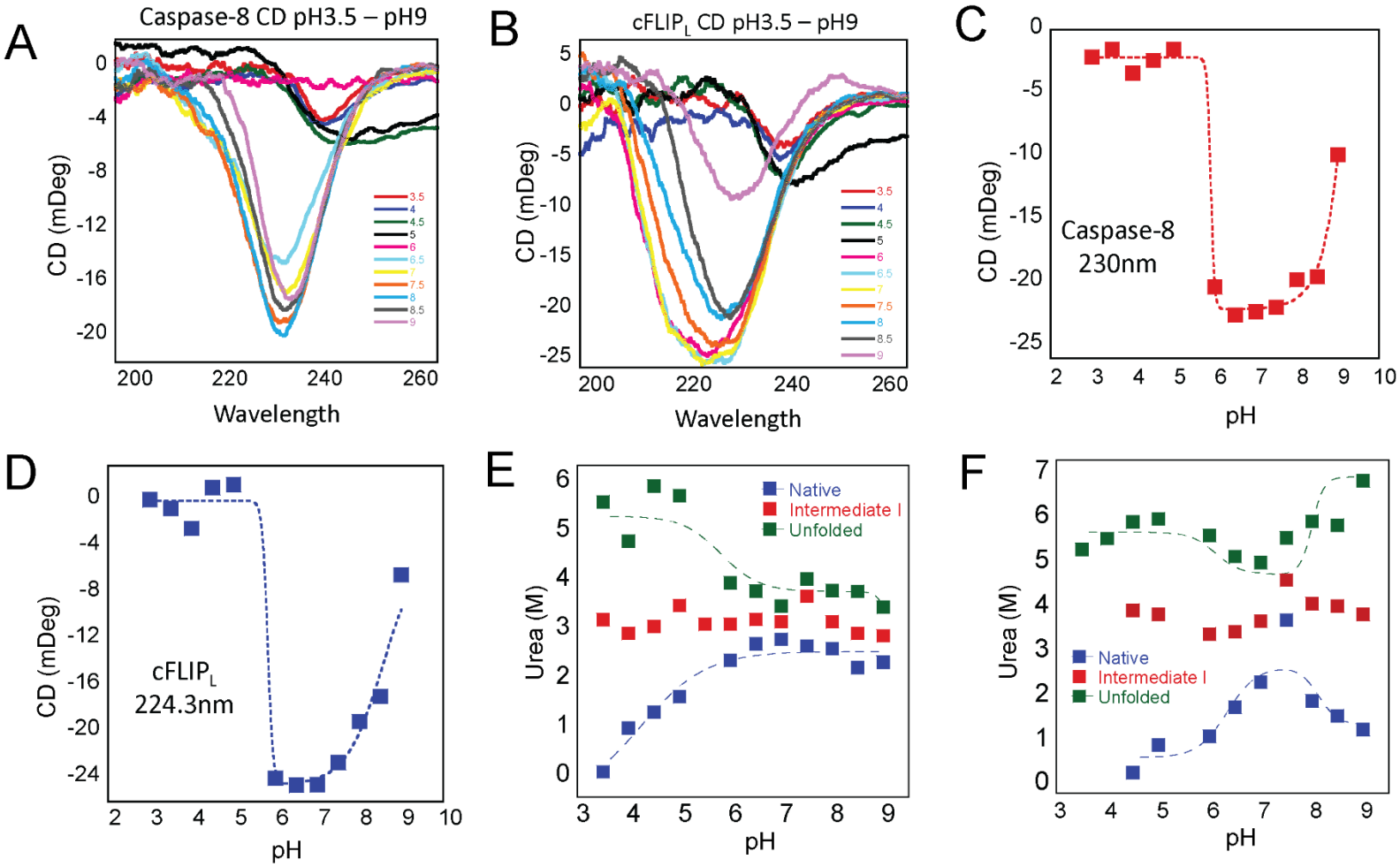
Supplemental Figure S4. Equilibrium unfolding monitored by fluorescence emission and circular dichroism for cFLIP$_L$ from pH 3.5 to 9. Solid lines represent global fits of the data as described in the text. Circles represent averaged unfolding data and triangles represent refolding data following fluorescence emission after excitation at 280 nm (blue) or 295 nm (red), and circular dichroism (green).



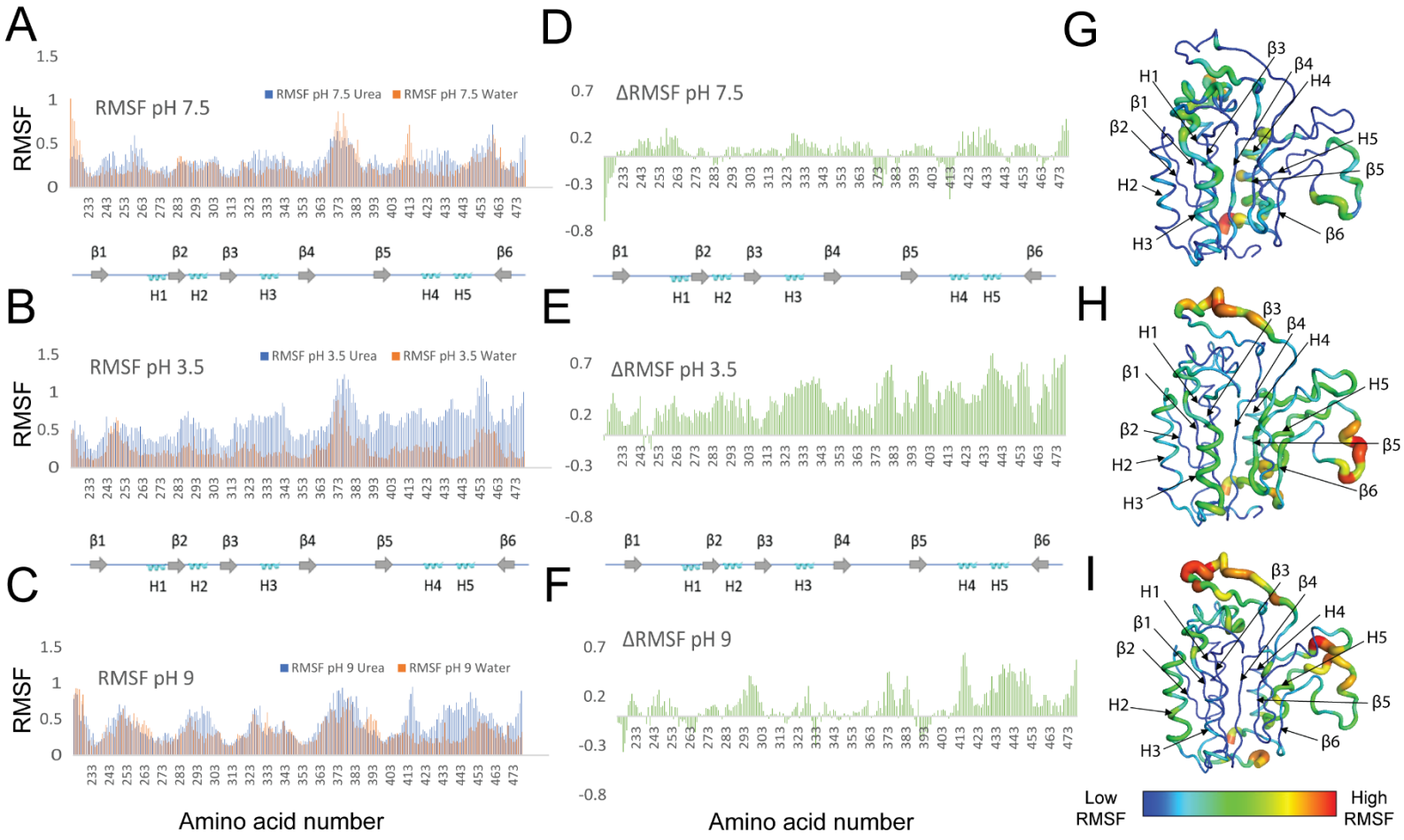
Supplemental Figure S5. Fraction of species of caspase-8 folding/unfolding as a function of urea concentration from pH 3.5 to 9. Fractions were determined from global fits of data in Supplemental Figure S3, as described in the text. For each panel, blue solid line represents fraction of native species (N), red solid line represents fraction of partially folded intermediate (I), and green solid line represents the fraction of unfolded species (U).



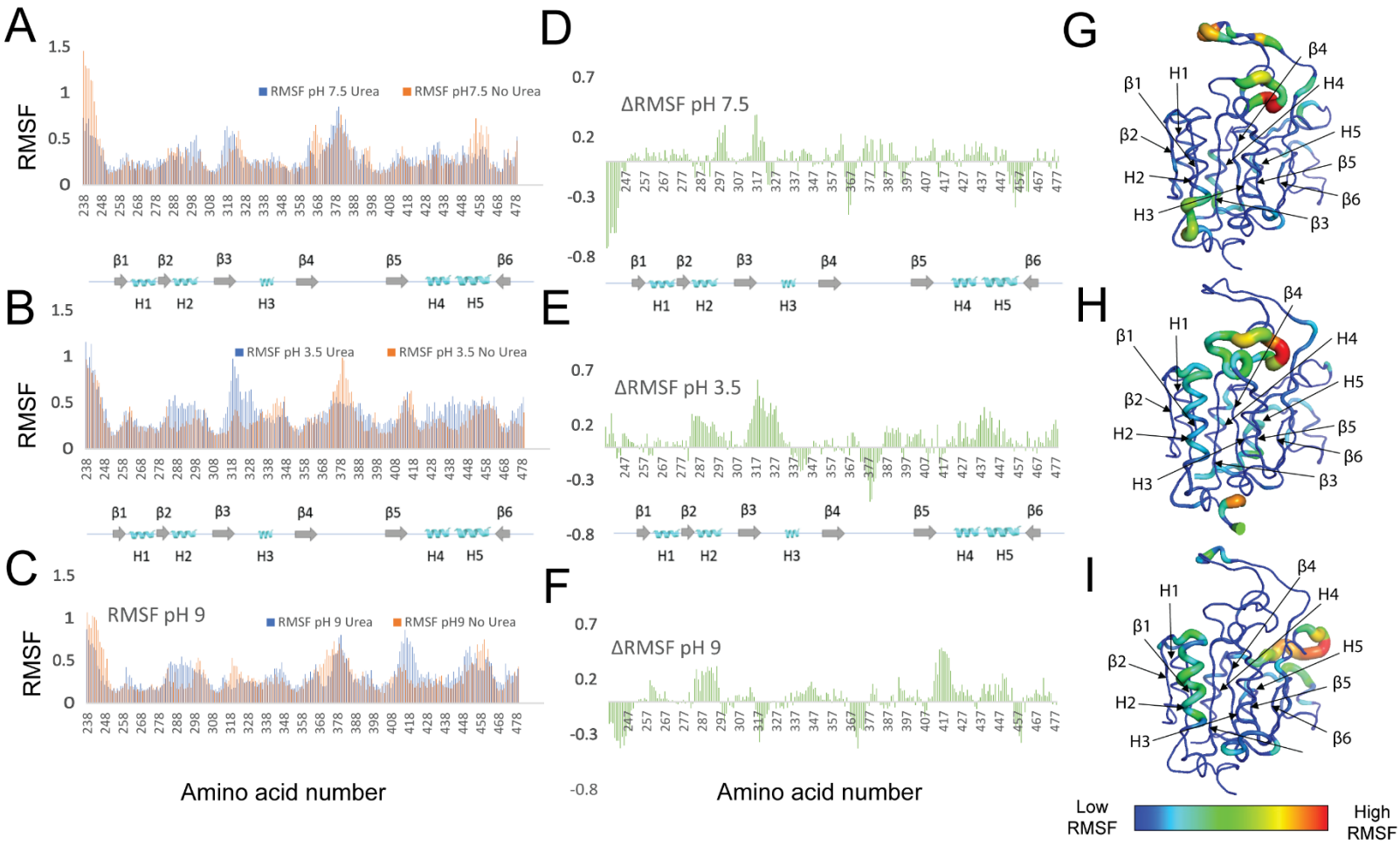
Supplemental Figure S6. Fraction of species of cFLIP$_L$ folding/unfolding as a function of urea concentration from pH 3.5 to 9. Fractions were determined from global fits of data in Supplemental Figure S4, as described in the text. For each panel, blue solid line represents fraction of native species (N), red solid line represents fraction of partially folded intermediate (I), and green solid line represents the fraction of unfolded species (U).



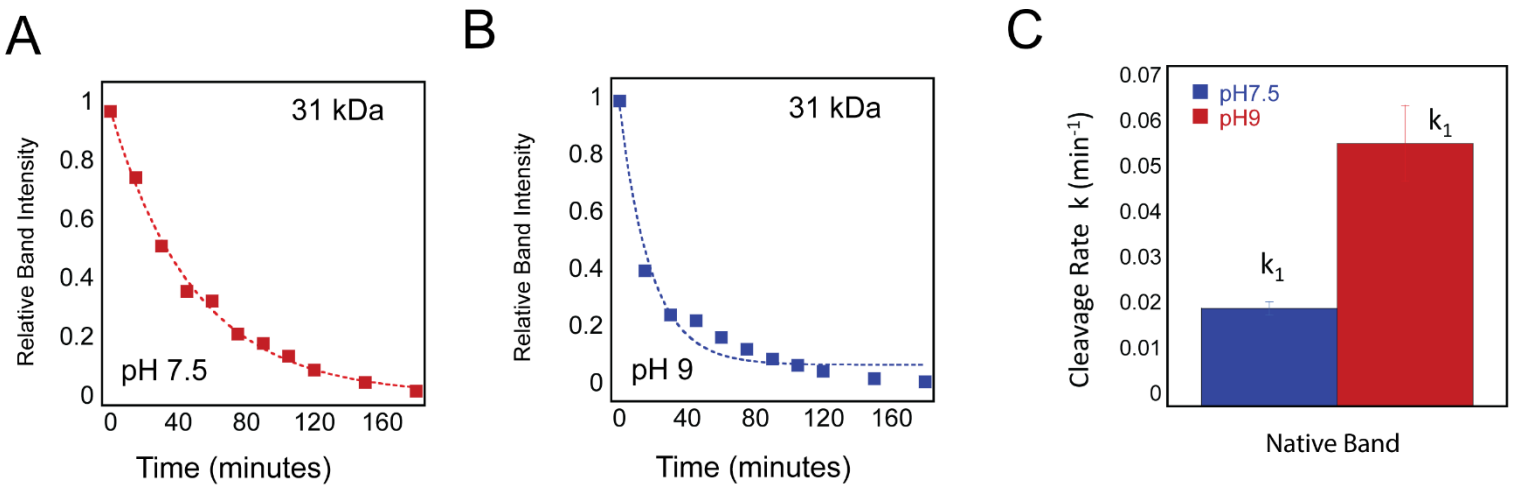
Supplemental Figure S7. Effects of pH on structure and fraction of species. **(A & B)** CD spectra scan showing a blue shift for the minima below pH6 for caspase-8 and cFLIP$_L$, respectively **(C & D)** Changes in far-UV CD signal as a function of pH for acquiring pKa fits in supplemental table S3&4 for caspase-8 (230nm) and cFLIP$_L$ (224.3nm), respectively. **(E)** The midpoint of the fraction of species determined from data in Supplemental Figure S5 for caspase-8. **(F)** The midpoint of the fraction of species determined from data in Supplemental Figure S6 for cFLIP$_L$. For panels A-C, the data were fit as described in the text to determine the pKa for each transition. Results of the fits are shown as the dashed lines, and the pKas are listed in Supplemental Tables S3 and S4.



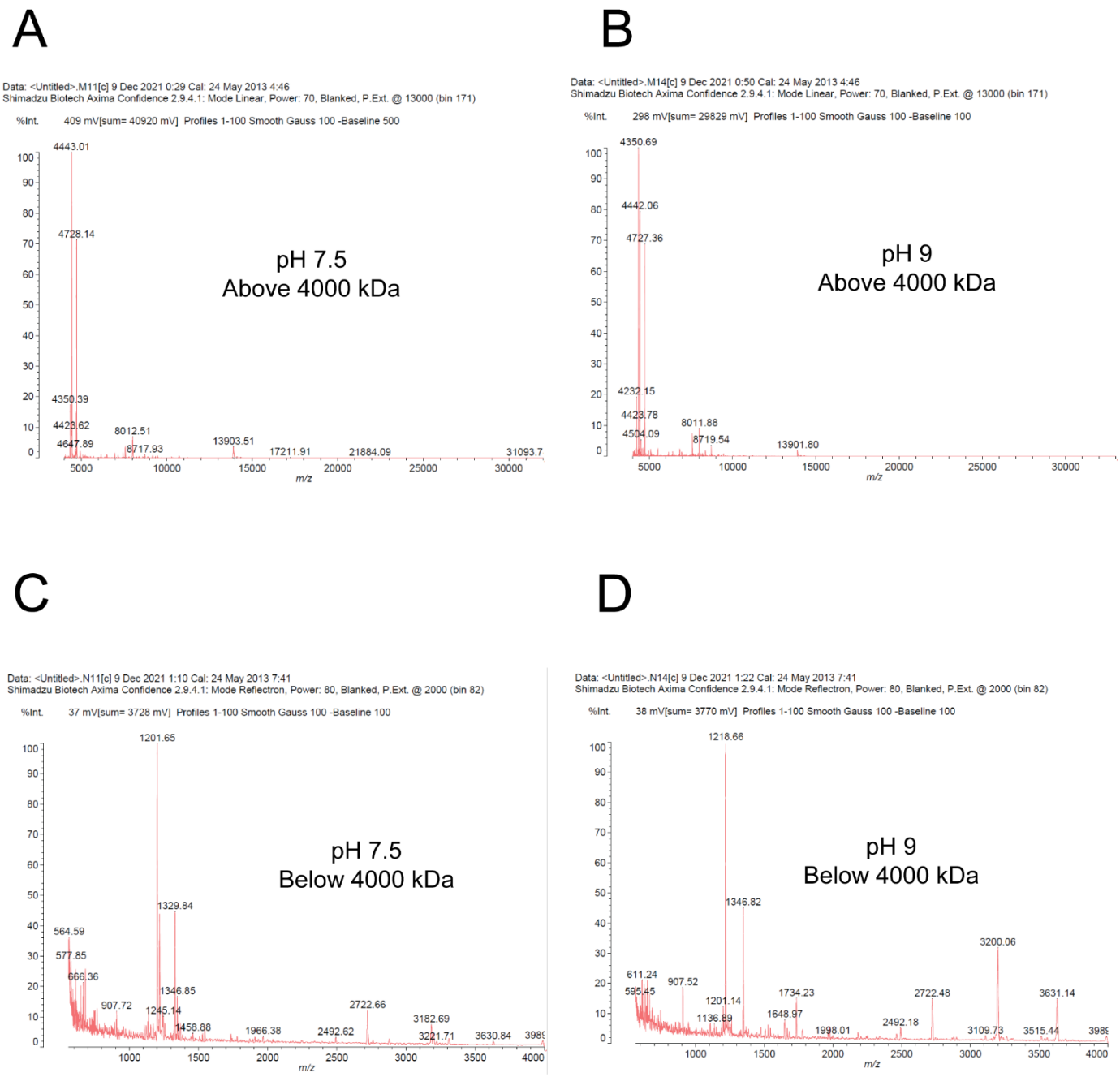
Supplemental Figure S8. MD simulations of caspase-8 in water and in urea. The root mean square fluctuations (RMSF) for caspase-8 in water and in urea are shown in orange and blue, respectively, at **(A)** pH 7.5 , **(B)** pH 3.5 and **(C)** pH 9. The difference in RMSF (ΔRMSF) between the urea and water simulations are shown at **(D)** pH 7.5, **(E)** pH 3.5, and **(F)** pH 9. The ΔRMSF were converted into B-factors and mapped onto the structure of caspase-8 (PDB ID: 2K7Z) at **(G)** pH 7.5, **(H)** pH 3.5, and **(I)** pH 9. The temperature coloring scheme shows lowest RMSF regions in blue, the highest in red, and a spectrum, which is shown by the color bar, is used to illustrate the scores in between.



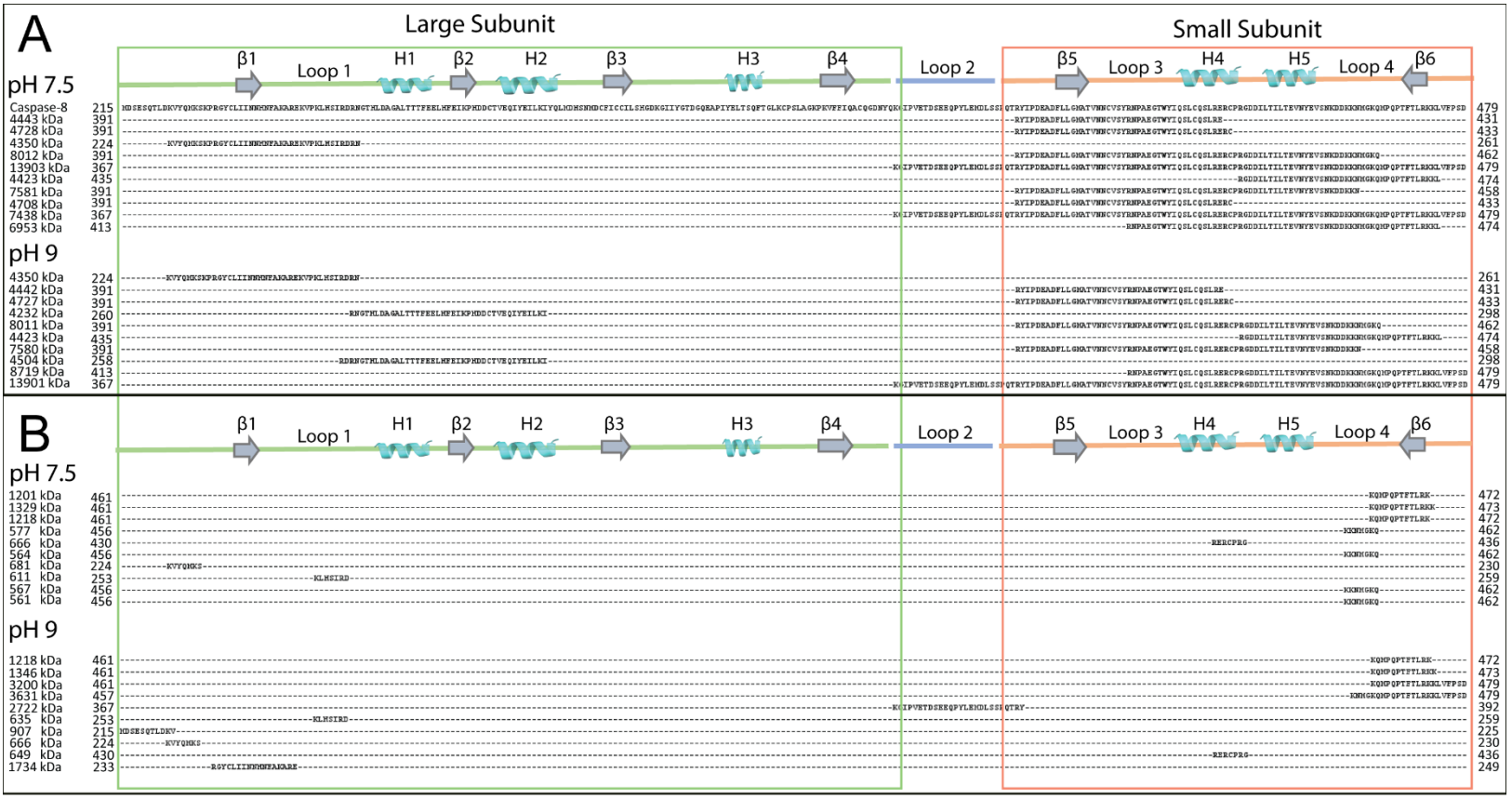
Supplemental Figure S9. MD simulations of cFLIP$_L$ in water and in urea. The root mean square fluctuations (RMSF) for cFLIP$_L$ in water and in urea are shown in orange and blue, respectively, at **(A)** pH 7.5 , **(B)** pH 3.5 and **(C)** pH 9. The difference in RMSF (ΔRMSF) between the urea and water simulations are shown at **(D)** pH 7.5, **(E)** pH 3.5, and **(F)** pH 9. The ΔRMSF were converted into B-factors and mapped onto the structure of caspase-8 (PDB ID: 2K7Z) at **(G)** pH 7.5, **(H)** pH 3.5, and **(I)** pH 9. The temperature coloring scheme shows lowest RMSF regions in blue, the highest in red, and a spectrum, which is shown by the color bar, is used to illustrate the scores in between.



Supplemental Figure S10. Kinetics of caspase 8 cleavage by trypsin at pH 7.5 vs pH 9. Intensities for native band cleavage were determined from data in Figure 5 for pH 7.5 **(A)** or pH 9 **(B)** and normalized to the zero-time point. The data were fit to a single exponential equation to determine the apparent rate constant for decay of native band. **(C)** Kinetics of cleavage of annotated bands in panels A and B were to fit to single exponential equation to determine the rate constants. Data for pH 7.5 are shown in blue, and data for pH 9 are shown in red. Apparent rate constant, k_1 was determined as described in the text.



Supplemental Figure S11. MALDI-TOF data showing cleavage fragments at the 60-minute time point of trypsin proteolysis as plots of m/z (mass/charge) vs intensity as percentage. Fragments above 4000 kDa were detected using linear mode for samples at **(A)** pH 7.5 and **(B)** pH 9 , while fragments below 4000 kDa were detected using reflectron mode for samples at **(C)** pH 7.5 and **(D)** pH 9.



Supplemental Figure S12. Cleavage fragments with the top 10 intensity at the 60-minute time point of limited proteolysis at pH 7.5 and pH 9 were determined by MALDI-TOF for cleavages with molecular weight (A) above 4000 kDa and (B) below 4000 kDa and aligned to the caspase-8 ΔDED sequence (topmost). The molecular weights are shown on the left with the alignment on the right while the numbers on the extremes of the alignment represent the residue number of the canonical caspase-8 sequence. The green box depicts the large subunit and the orange box depicts the small subunit to show that most of the fragments are from the small subunit and at pH 9 additional fragments from the large subunit appear.

Chapter 2

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 (Fig 1), 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 loose 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 addtion, 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 effect 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).

Figure 1

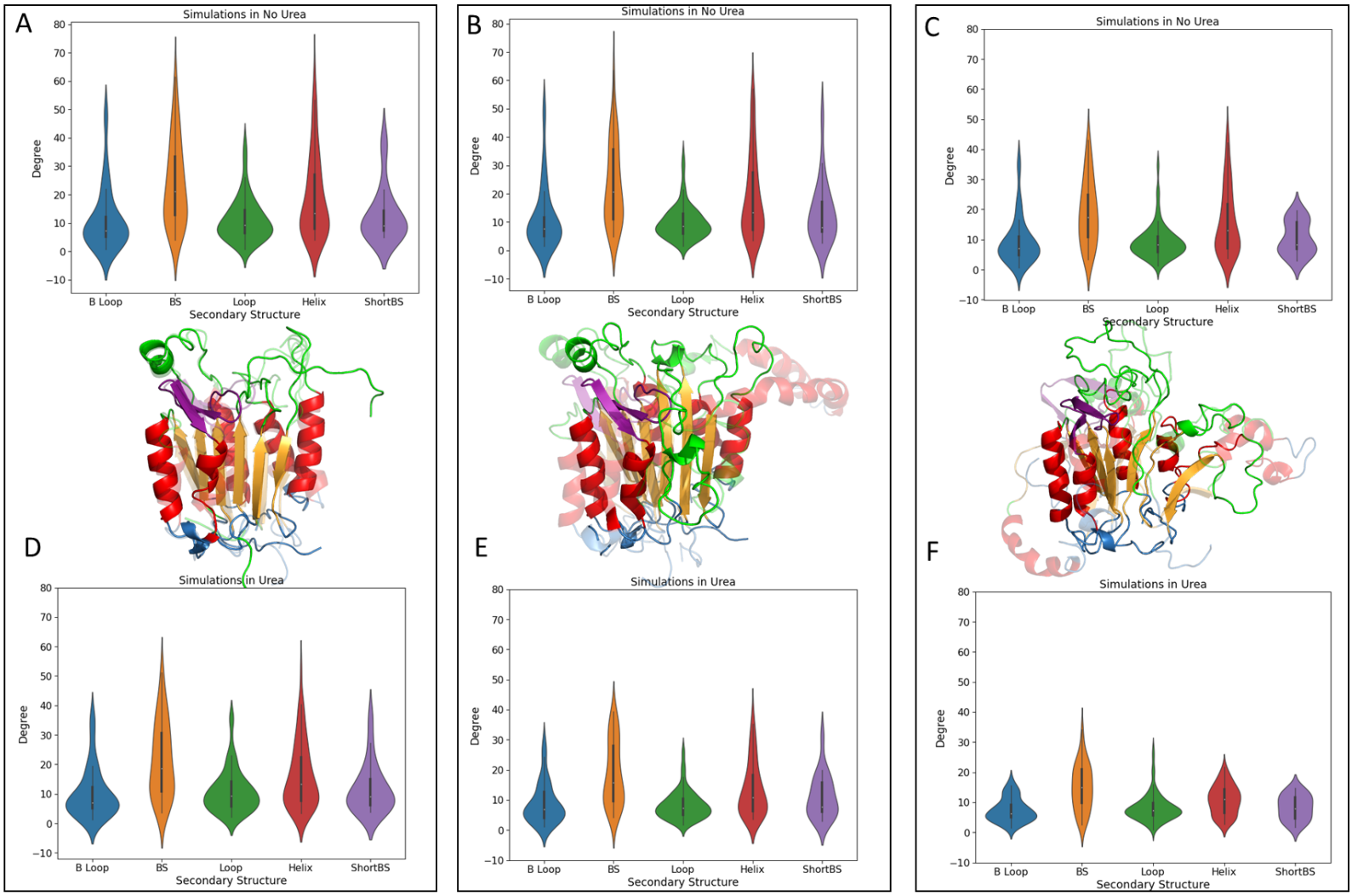


Figure 1. Violin plots representing the average degrees/contacts for the (A)active, (B)dimeric and (C)monomeric conformations in water and (D)active, (E)dimeric (F) monomeric conformations in 8M urea. Bottom loops (blue), beta-sheets (orange), catalytic site loops (green), helices (red) and the short 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

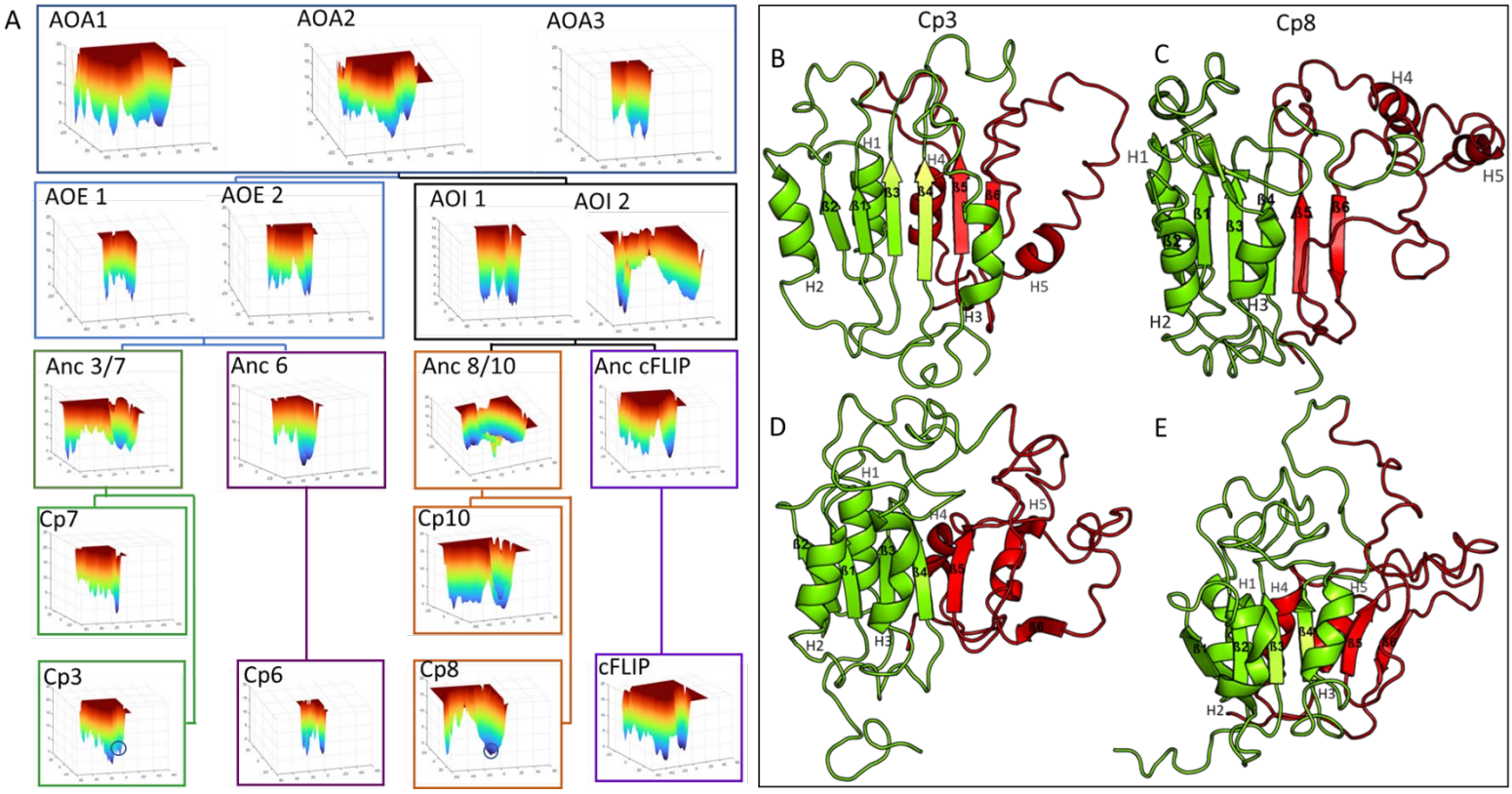


Figure 2. The free energy landscapes of the dimeric conformation of caspases in urea obtained from 200 ns MD simulations. (A) The FELs are generated as a function of projections of the MD trajectory onto the first (PC1) and the second (PC2) eigenvectors, respectively. Observed FEL minimas circled on 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

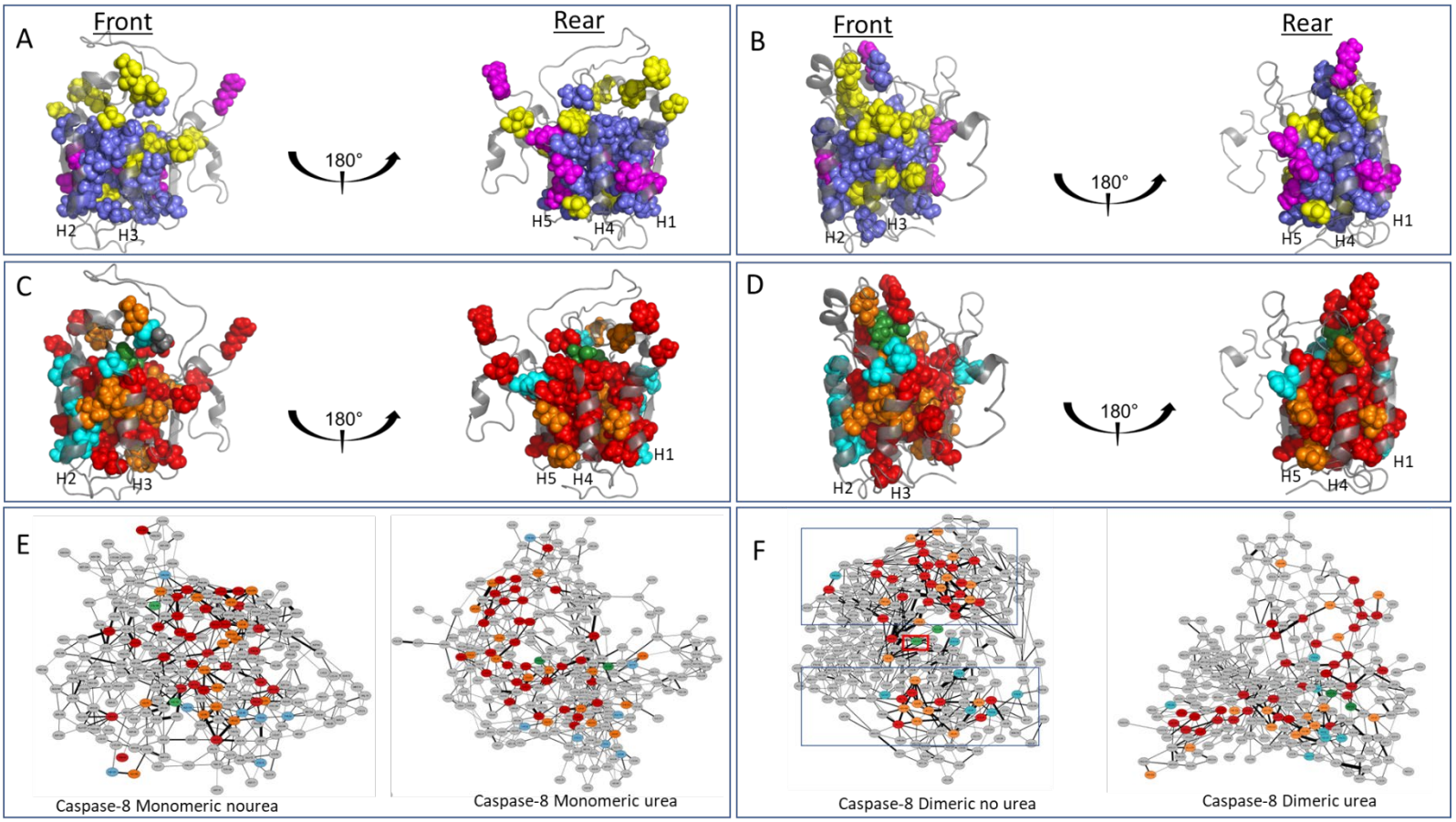


Figure 3. High DC and BC residues classified according to conservation in the monomeric and dimeric. Blue spheres represent the conserved network of high DC, BC residues; magenta spheres indicate high BC, DC residues unique to the monomeric, whereas yellow spheres indicate high BC, DC residues unique 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

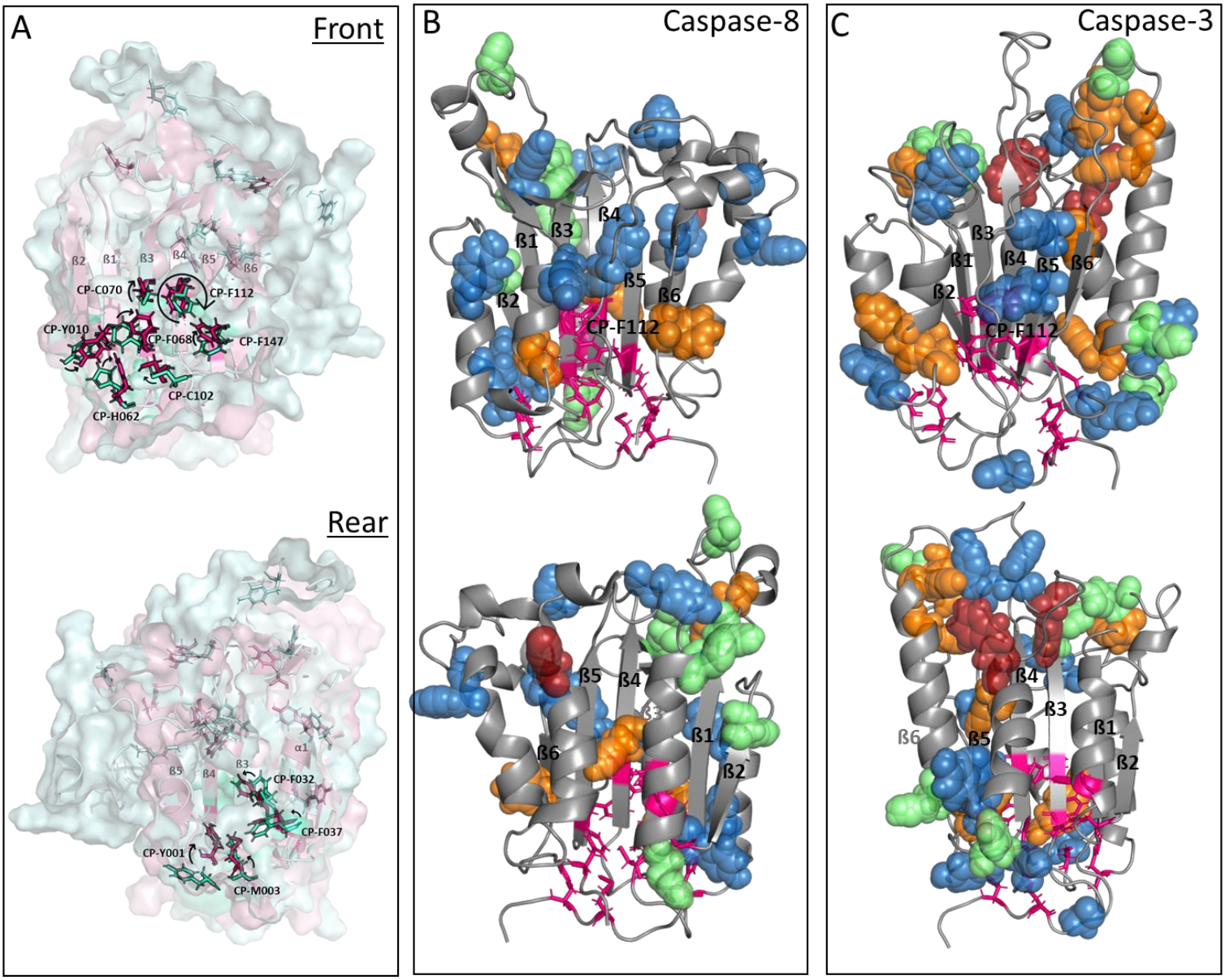


Figure 4. Conserved electron transport network in the caspase family. (A) A conserved network mapped onto the monomeric (pale cyan) and the dimeric (light pink) caspase-8 structures. All the residues in the conserved network are labeled according to the CP system and are depicted as sticks for the monomer (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

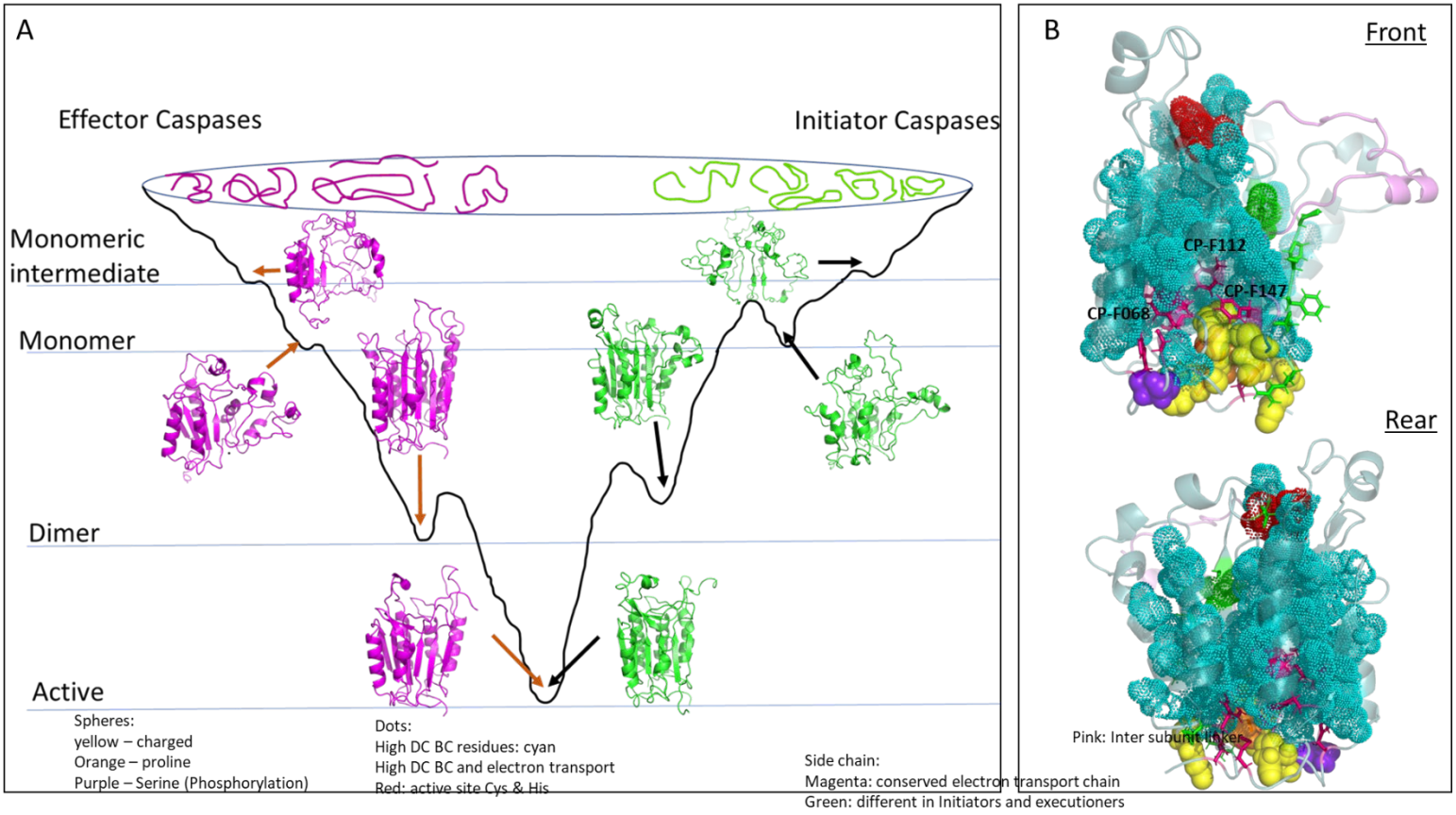


Figure 5. A. folding funnel model for the conformational stability of caspases. B. Model of caspases according to the classification.

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

Supporting Information

Figure S1

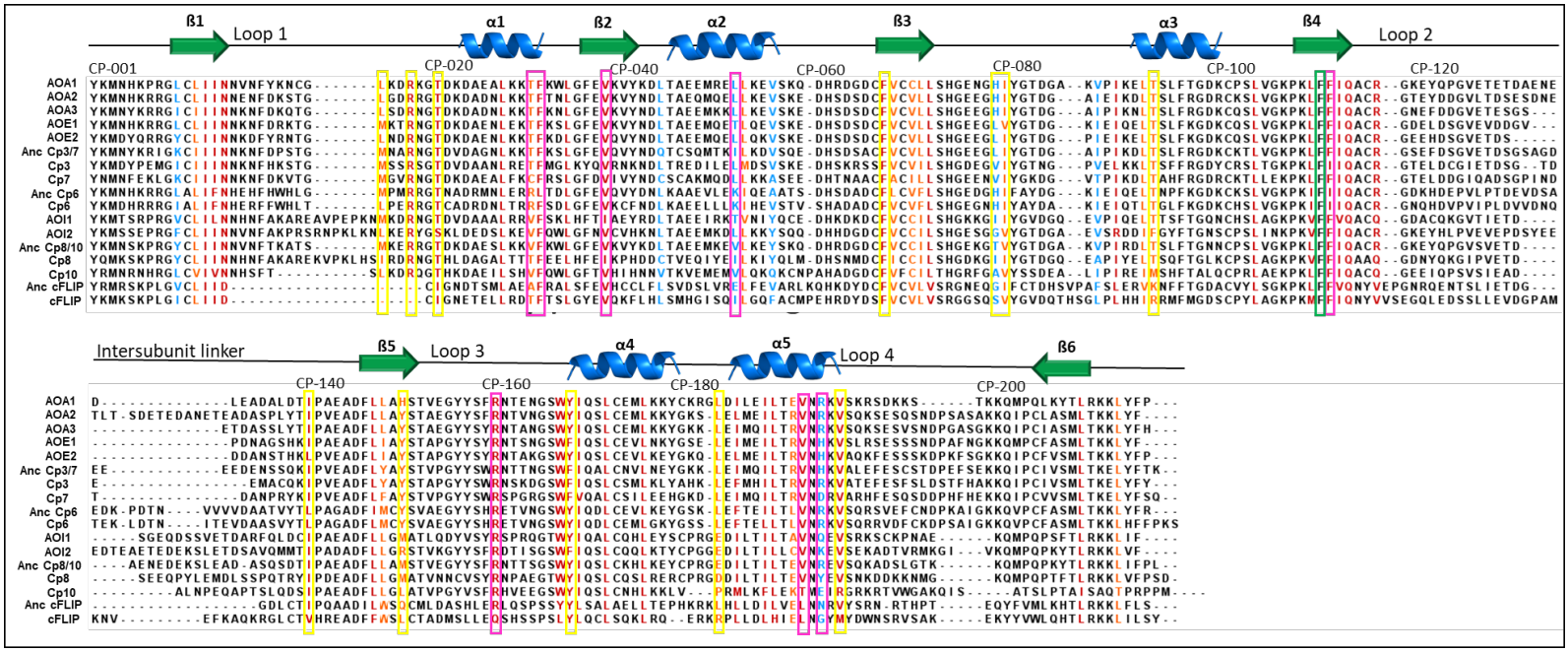


Figure S1. Multiple sequence alignment showing the secondary structural elements (loops, beta sheets, alpha helices) along with the common position numbers among the caspases. The colored residues represent the amino acids that have high DC, BC and are classified as conserved residues (red), intermediately 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

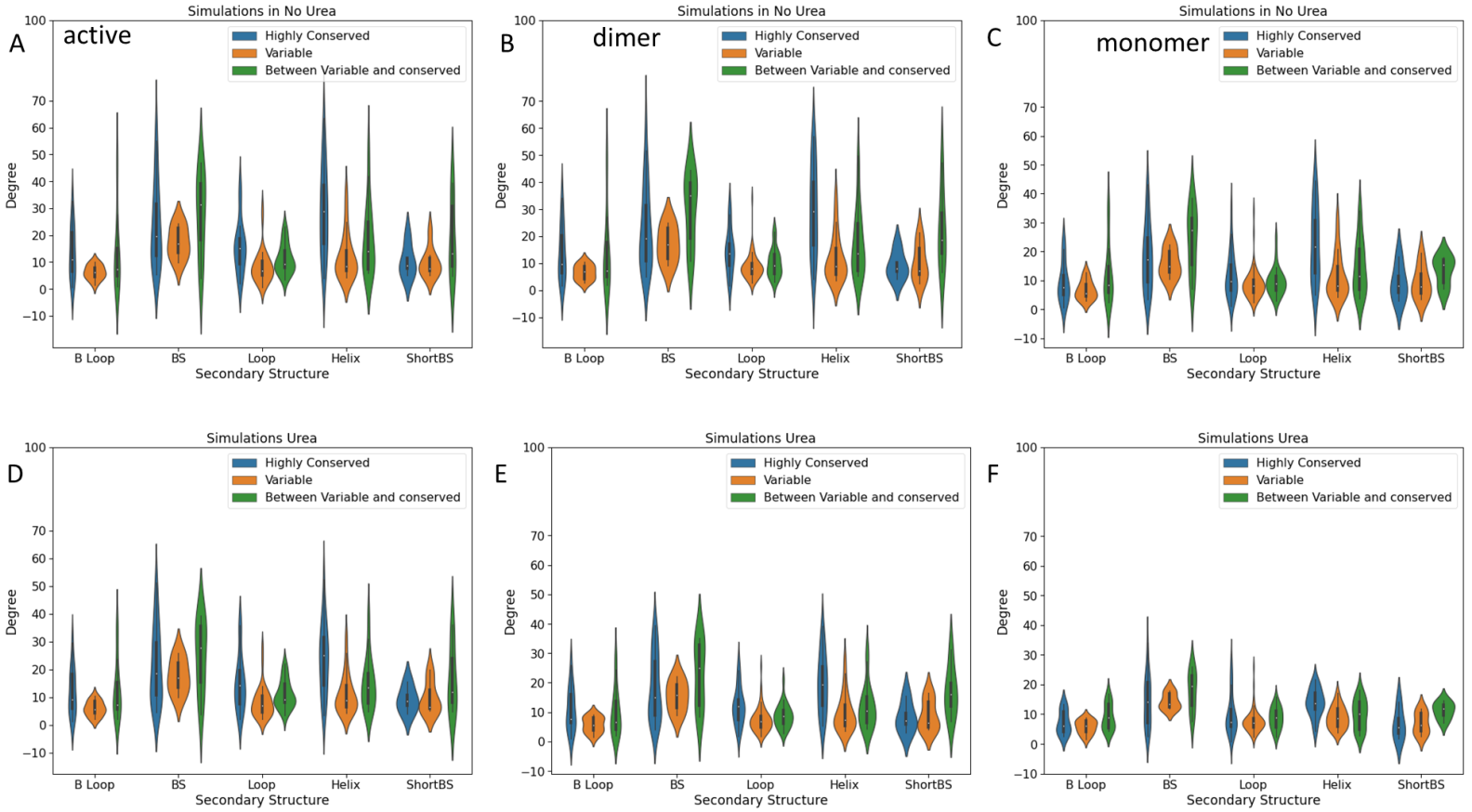


Figure S2. Violin plots representing the average degrees/contacts categorized as conserved (blue), intermediately conserved (green), and variable (orange) based on the conservation scores and grouped according to secondary structural elements for the (A)active, (B)dimeric and (C)monomeric conformations in water and (D)active, (E)dimeric (F) monomeric conformations in 8M urea.

Figure S3

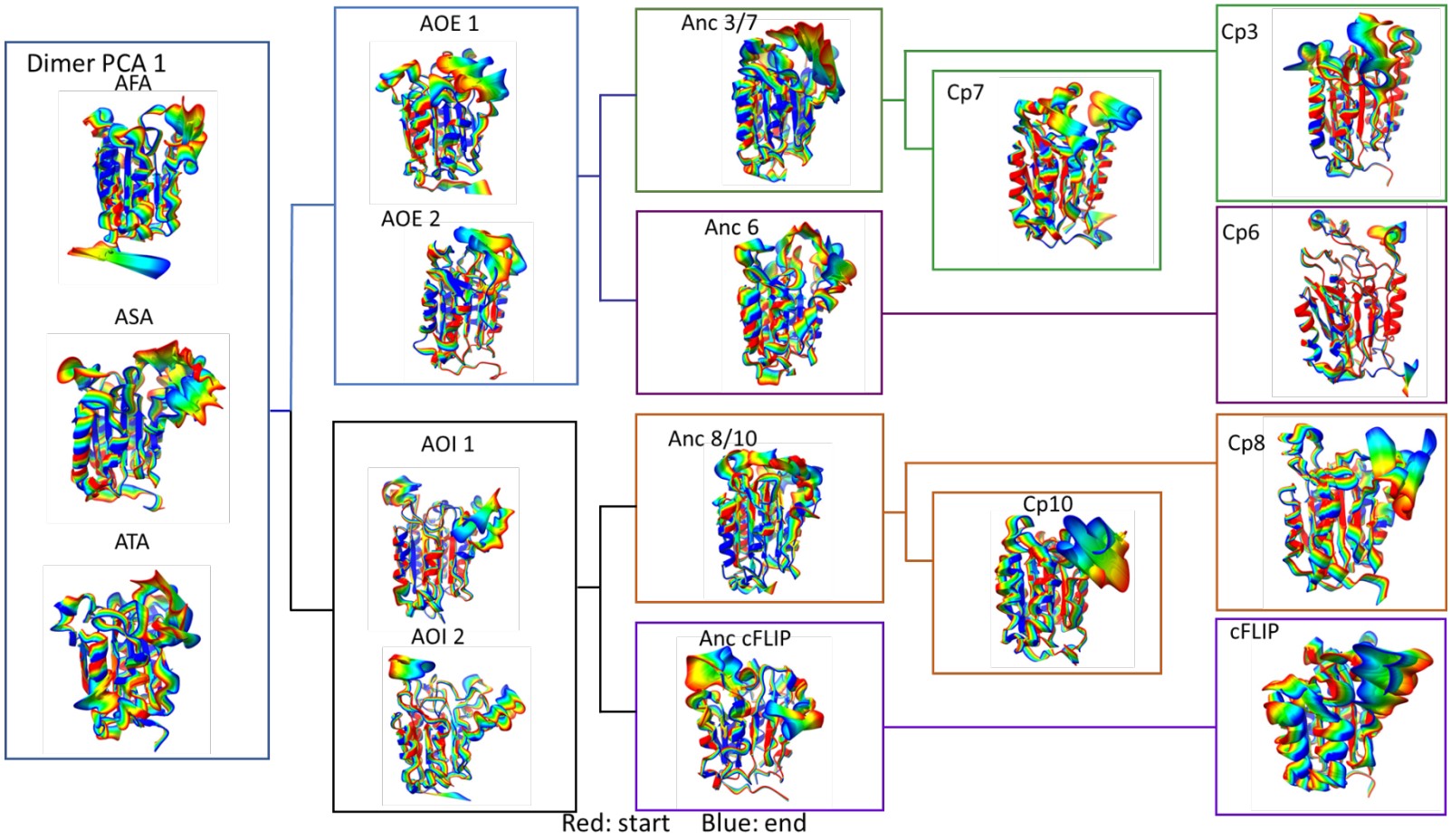


Figure S3. Large concerted motions of the caspase family in the dimeric conformation in water described by eigenvector 1 (PC1). The structural displacements between the two extremes extracted from the projection of the trajectory onto the first PC are colored from blue to red to highlight the differences.

Figure S4

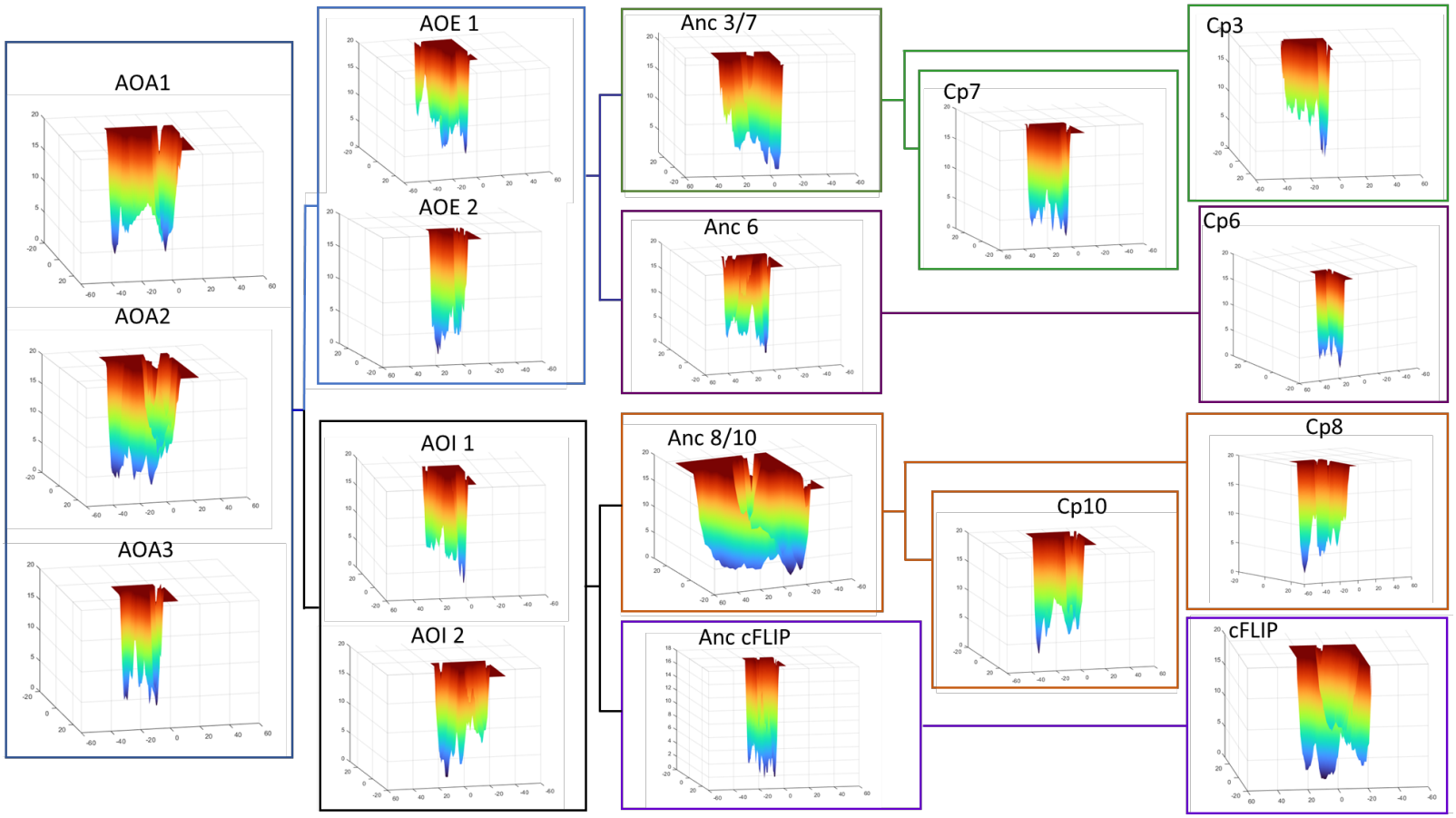


Figure S4. The free energy landscapes of the dimeric conformation of caspases in water obtained from 200 ns MD simulations. The FELs are generated as a function of projections of the MD trajectory onto the first (PC1) and the second (PC2) eigenvectors, respectively. Observed minimas represent the metastable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

Figure S5

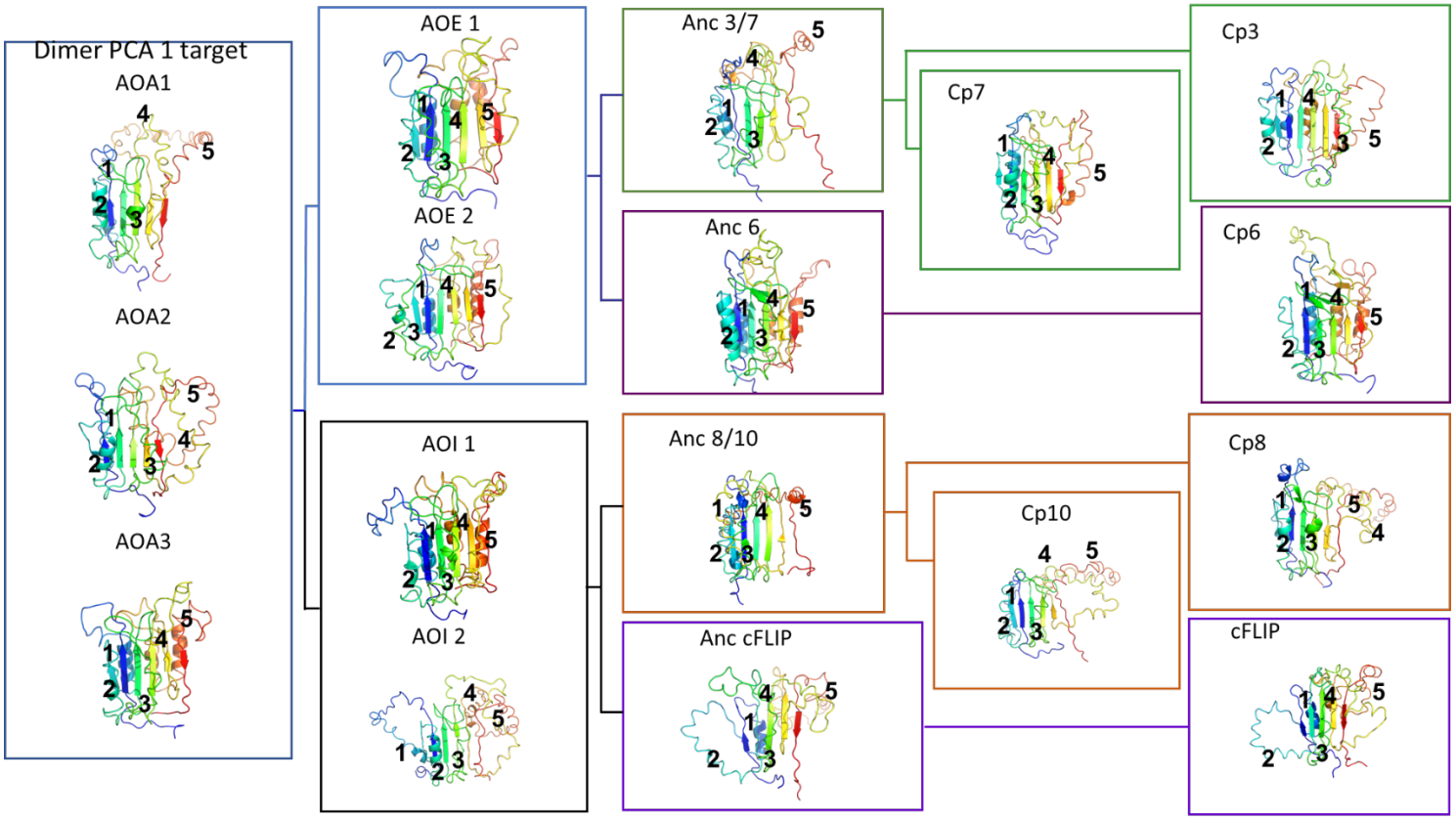


Figure S5. Metastable states of caspases in the dimeric conformation in urea indicating unstable regions.

Figure S6

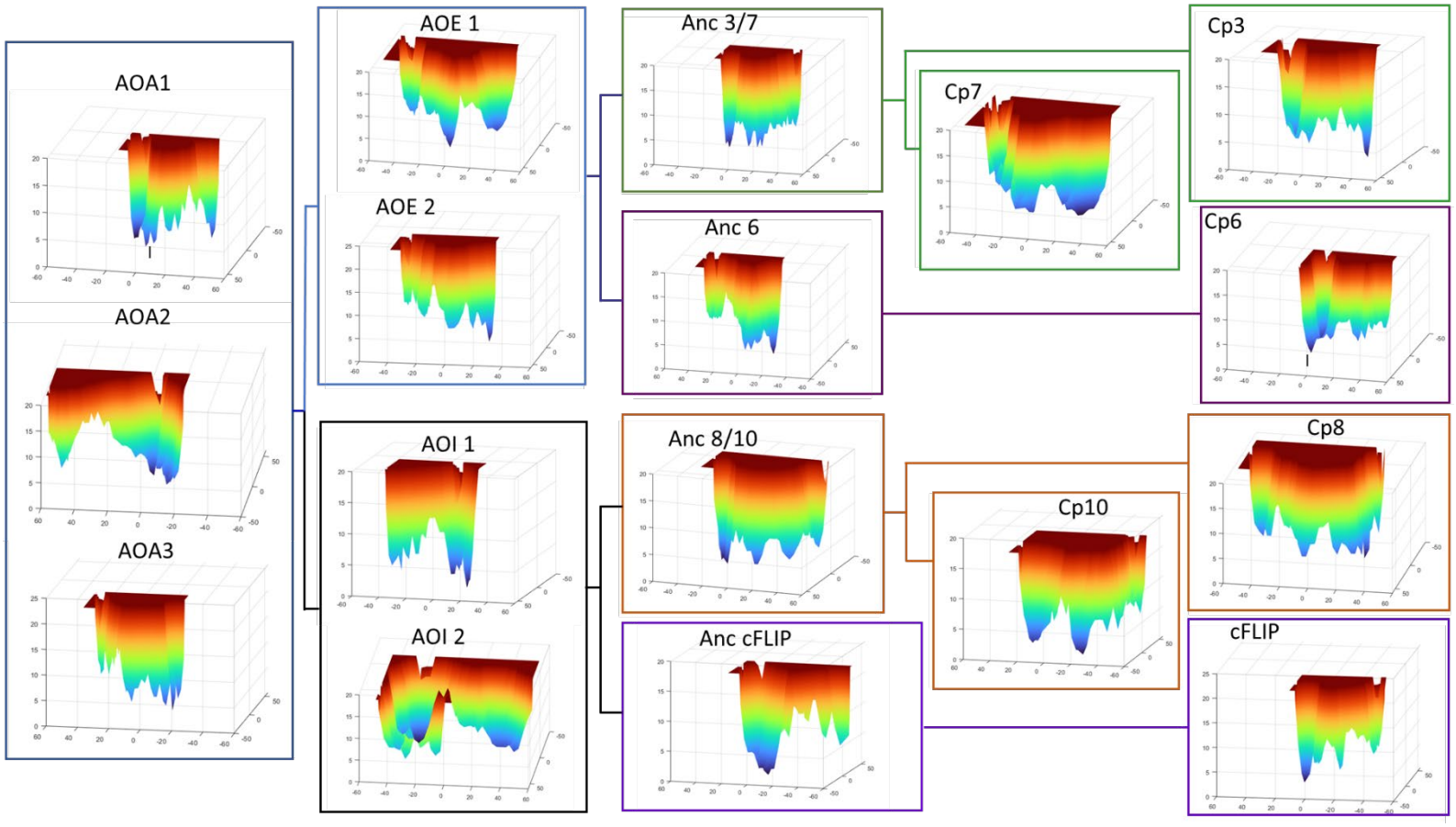


Figure S6. The free energy landscapes of the monomeric conformation of caspases in urea obtained from 200 ns MD simulations. The FELs are generated as a function of projections of the MD trajectory onto the first (PC1) and the second (PC2) eigenvectors, respectively. Observed minimas represent the metastable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

Figure S7

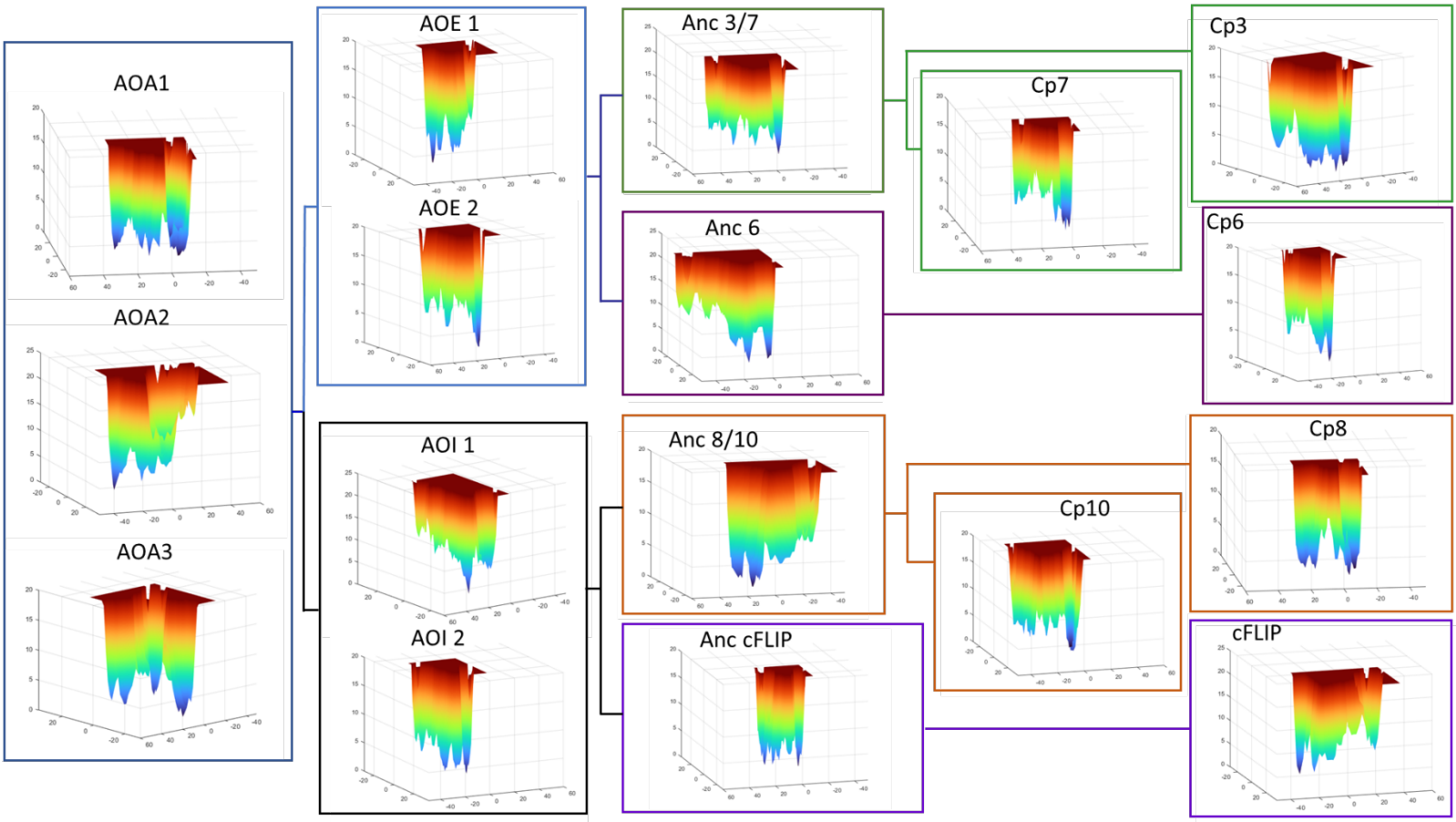


Figure S7. The free energy landscapes of the monomeric conformation of caspases in water obtained from 200 ns MD simulations. The FELs are generated as a function of projections of the MD trajectory onto the first (PC1) and the second (PC2) eigenvectors, respectively. Observed minimas represent the metastable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

Figure S8

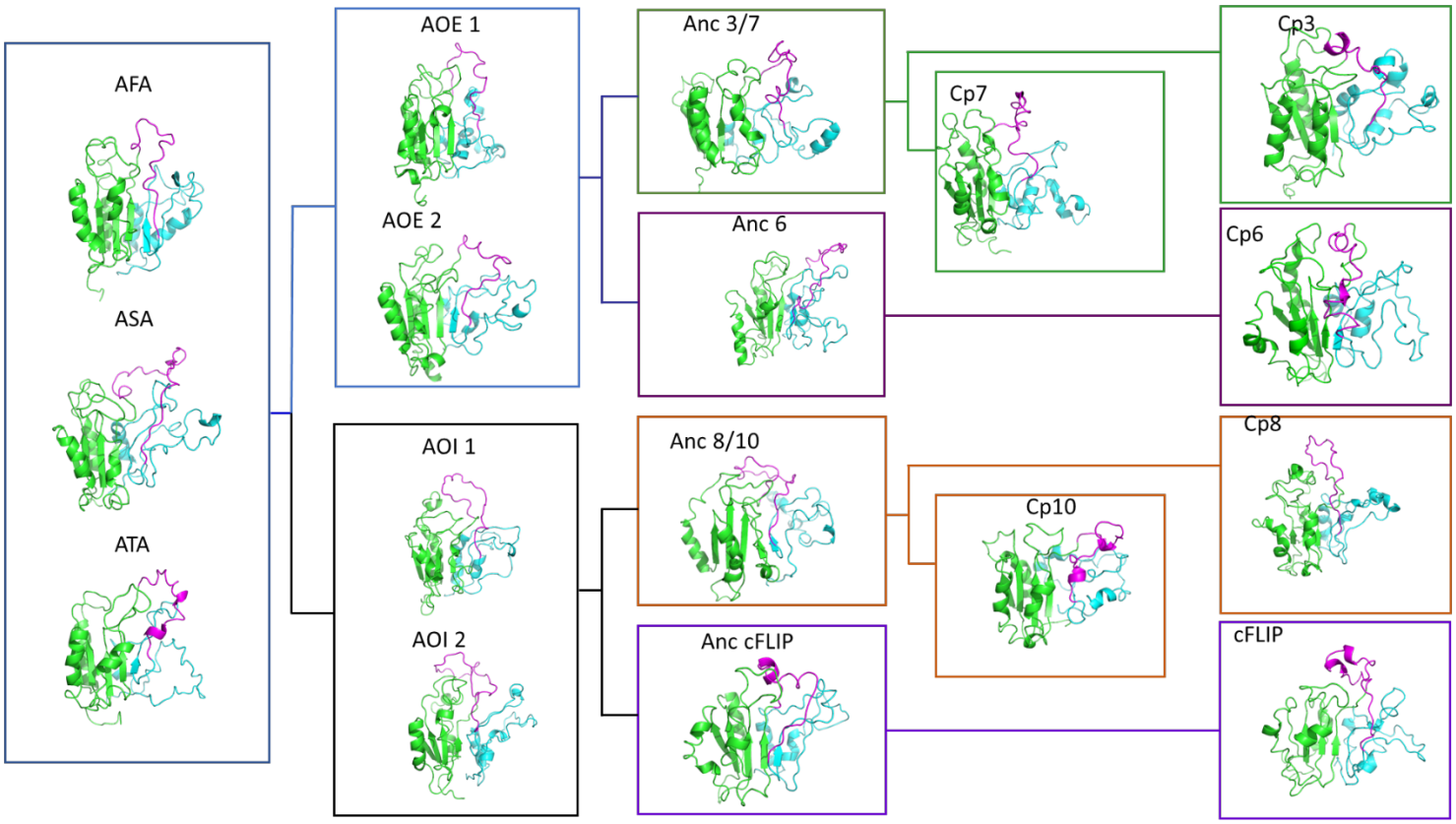


Figure S8. Metastable states extracted from the last observed minima in the FEL of the dimeric conformation in water

Figure S9

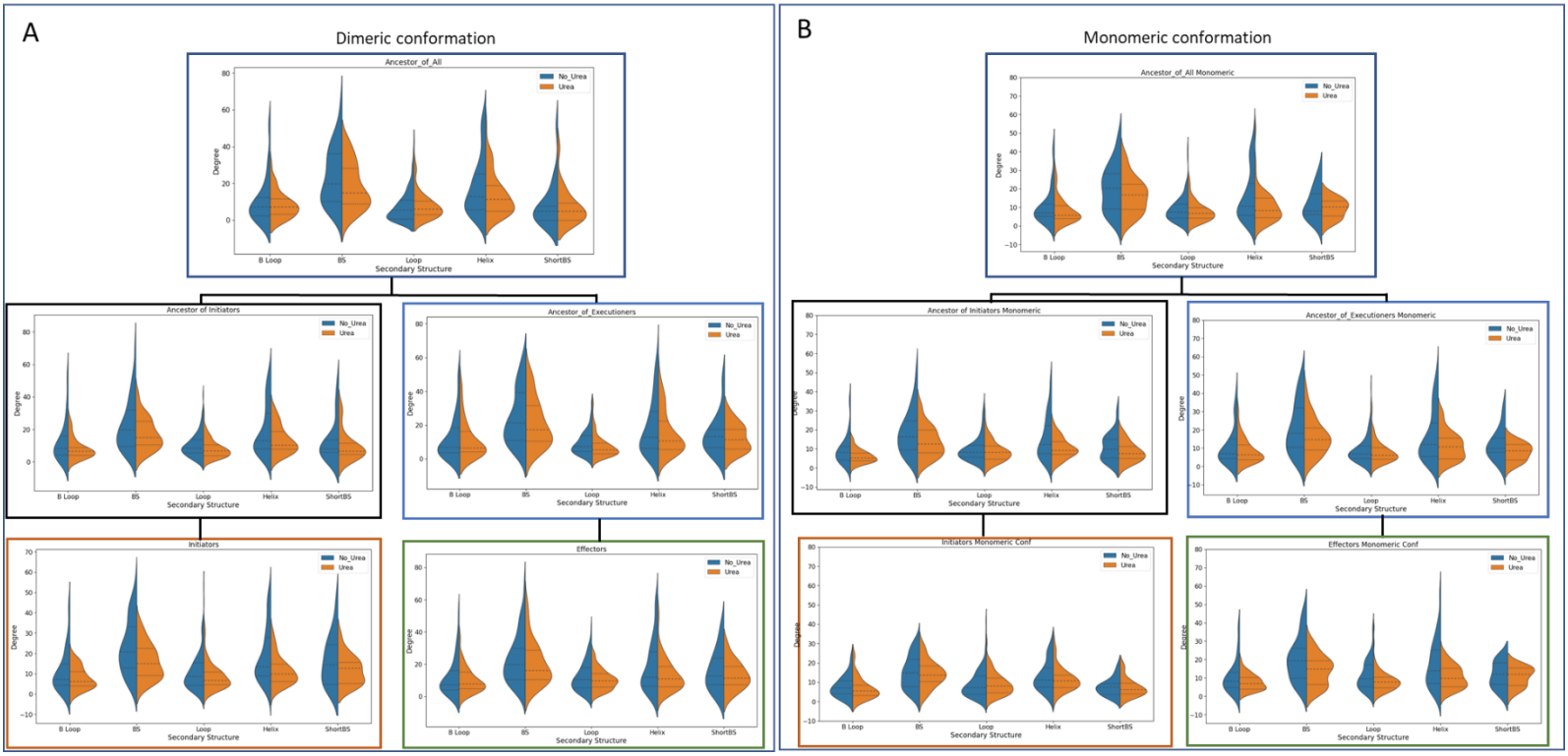


Figure S9. Violin plots illustrating the averaged high-degree contacts across the following groups: ancestor of all (AOA1/2/3), ancestor of initiators (AOI1/2), ancestor of effectors (AOE1/2), effector caspases (caspase -3/-6/-7) and initiator caspases (caspase -8/10/cFlip) in the (A) dimeric conformation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

Figure S10

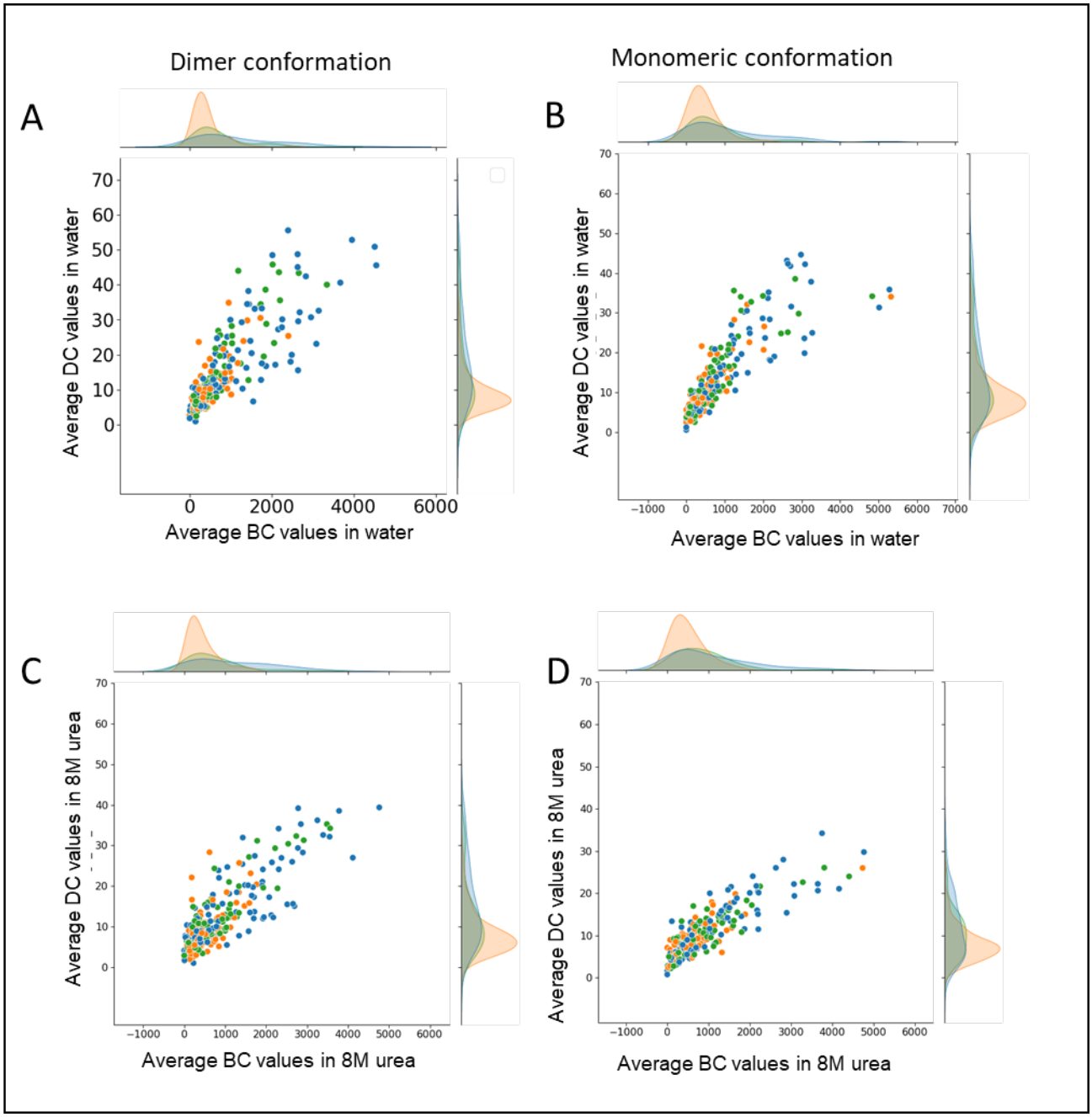


Figure S10. Pair plots displaying the average degree centrality on y-axis and average betweenness centrality on x-axis for dimeric structures in (A) water, (C) 8M urea, and the monomeric structures in (B) water, (D) 8M urea

Supplemental Fig. Legend

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

Supplemental Figure. S2. Graph S1 Displays the sequence composition of the CaspBase database, and from the sequences available in which phylum they fall under.

Supplemental Figure 1

Table S1 : showing the accession number of all caspases used in the database reconstructions

AOA1				
Accenssion ID	Caspase	Species	Class	Order
NP_004337.2	Caspase-3 Isoform A Preprotein	Homo Sapiens	Mammalia	Primates
NP_001271338.1	Caspase-3	Mus Musculus	Mammalia	Rodentia
NP_571952.1	Caspase-3 apoptosis-related cysteine peptidase a	Danio Rerio	Fish	Cypriniformes
XP_001517122.2	Predicted: Caspase-3	Ornithorhynchus Anatinus	Mammalia	Monotremata
NP_001003042.1	Caspase-3	Canis Lupus Familiaris	Mammalia	Carnivora
NP_001157433.1	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_001075586.1	Caspase-3	Oryctolagus Cuniculus	Mammalia	Lagomorpha

NP_001120900.1	Caspase-3	Xenopus Tropicalis	Amphibia	Anura
XP_011362327.1	Predicted: Caspase-3	Pteropus Vampyrus	Mammalia	Chiroptera
XP_003773079.1	Predicted: Caspase-3	Sarcophilus Harrisii	Mammalia	Dasyuromorphia
XP_004481806.1	Predicted: Caspase-3	Dasypus Novemcinctus	Mammalia	Cingulata
XP_006128558.1	Predicted: Caspase-3 isoform X1	Pelodiscus Sinensis	Reptilia	Testudines
XP_005149203.1	Predicted: Caspase-3	Melopsittacus Undulatus	Birds	Psittaciformes
XP_005282031.1	Predicted: Caspase-3 isoform X1	Chrysemys Picta Bellii	Reptilia	Testudines
XP_004389725.1	Predicted: Caspase-3	Trichechus Manatus Latirostris	Mammalia	Sirenia
XP_005045177.1	Predicted: Caspase-3	Ficedula Albicollis	Birds	Passeriformes
XP_019336883.1	Predicted: Caspase-3	Alligator Mississippiensis	Reptilia	Crocodilia
XP_007517715.1	Predicted: Caspase-3	Erinaceus Europaeus	Mammalia	Erinaceomorpha
NP_001098140.1	Caspase-3	Oryzias Latipes	Fish	Beloniformes
XP_005500235.1	Caspase-3	Columba Livia	Birds	Columbiformes
XP_022539261.1	Caspase-3	Astyanax Mexicanus	Fish	Characiformes
XP_009810205.1	Predicted: Caspase-3	Gavia Stellata	Birds	Gaviiformes
XP_009958677.1	Predicted: Caspase-3	Leptosomus Discolor	Birds	Leptosomiformes
XP_008569758.1	Predicted: Caspase-3	Galeopterus Variegatus	Mammalia	Dermoptera
XP_009473093.1	Predicted: Caspase-3	Nipponia Nippon	Birds	Pelecaniformes
NP_001290581.1	Caspase-3	Esox Lucius	Fish	Esociformes

XP_013910223.1	Predicted: Caspase-3	Thamnophis Sirtalis	Reptilia	Squamata
NP_001188010.1	Caspase-3	Ictalurus Punctatus	Fish	Siluriformes
NP_001081225.1	Caspase-3	Xenopus Laevis	Amphibia	Anura
XP_017503748.1	Predicted: Caspase-3	Manis Javanica	Mammalia	Pholidota
NP_001269823.1	Caspase-3	Oreochromis Niloticus	Fish	Perciformes
XP_020863340.1	Caspase-3	Phascolarctos Cinereus	Mammalia	Diprotodontia
XP_020658385.1	Caspase-3	Pogona Vitticeps	Reptilia	Squamata
NP_001217.2	Caspase-6 Isoform Alpha Precursor	Homo Sapiens	Mammalia	Primates
XP_016807502.1	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_007907939.1	Predicted: Caspase-6	Callorhinchus Milii	Fish	Chimaeriformes
XP_003972493.2	Predicted: Caspase-6	Takifugu Rubripes	Fish	Tetraodontiformes
XP_007437451.1	Predicted: Caspase-6	Python Bivittatus	Reptilia	Squamata
XP_004476229.1	Predicted: Caspase-6 isoform X1	Dasypus Novemcinctus	Mammalia	Cingulata
XP_005999642.1	Predicted: Caspase-6	Latimeria Chalumnae	Fish	Coelacanthiformes
XP_005148683.1	Predicted: Caspase-6	Melopsittacus Undulatus	Birds	Psittaciformes
XP_014268841.1	Caspase-6 isoform X2	Maylandia Zebra	Fish	Perciformes
XP_005797675.2	Predicted: Caspase-6	Xiphophorus Maculatus	Fish	Cyprinodontiformes
XP_005287972.1	Predicted: Caspase-6	Chrysemys Picta Bellii	Reptilia	Testudines
XP_004380293.1	Predicted: Caspase-6	Trichechus Manatus Latirostris	Mammalia	Sirenia
XP_019355646.1	Predicted: Caspase-6 isoform X1	Alligator Mississippiensis	Reptilia	Crocodilia
XP_005517572.1	Predicted: Caspase-6	Pseudopodoces Humilis	Birds	Passeriformes
XP_006162304.2	Predicted: Caspase-6	Tupaia Chinensis	Mammalia	Scandentia
XP_005515276.1	Predicted: Caspase-6 isoform X1	Columba Livia	Birds	Columbiformes
XP_007054543.1	Predicted: Caspase-6	Chelonia Mydas	Reptilia	Testudines
XP_009663699.1	Predicted: Caspase-6 isoform X1	Struthio Camelus Australis	Birds	Struthioniformes
XP_008497831.1	Predicted: Caspase-6	Calypte Anna	Birds	Apodiformes

XP_011572791.1	Predicted: Caspase-6 isoform X1	Aquila Chrysaetos Canadensis	Birds	Accipitriformes
XP_018425527.1	Predicted: Caspase-6	Nanorana Parkeri	Amphibia	Anura
XP_015281944.1	Predicted: Caspase-6 isoform X1	Gekko Japonicus	Reptilia	Squamata
NP_001081406.1	Caspase-6 L Homeolog	Xenopus Laevis	Amphibia	Anura
NP_001117743.1	Caspase-6 Precursor	Oncorhynchus Mykiss	Fish	Salmoniformes
NP_001253985.1	Caspase-7 Isoform Alpha Precursor	Homo Sapiens	Mammalia	Primates
NP_071596.1	Caspase-7	Rattus Norvegicus	Mammalia	Rodentia
NP_001018443.1	Caspase-7	Danio Rerio	Fish	Cypriniformes
XP_001513388.4	Predicted: Caspase-7	Ornithorhynchus Anatinus	Mammalia	Monotremata
XP_005637795.1	Caspase-7 isoform X1	Canis Lupus Familiaris	Mammalia	Carnivora
XP_421764.3	Predicted: Caspase-7	Gallus Gallus	Birds	Galliformes
XP_020928981.1	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_010331189.1	Predicted: Caspase-7	Saimiri Boliviensis Boliviensis	Mammalia	Primates
XP_004567546.1	Caspase-7 isoform X3	Maylandia Zebra	Fish	Perciformes
XP_005804467.1	Predicted: Caspase-7	Xiphophorus Maculatus	Fish	Cyprinodontiformes
XP_005293940.1	Predicted: Caspase-7	Chrysemys Picta Bellii	Reptilia	Testudines
XP_015203144.1	Predicted: Caspase-7 isoform X1	Lepisosteus Oculatus	Fish	Lepisosteiformes
XP_005048827.1	Predicted: Caspase-7 isoform X2	Ficedula Albicollis	Birds	Passeriformes
XP_014450146.1	Predicted: Caspase-7 isoform X1	Alligator Mississippiensis	Reptilia	Crocodilia
XP_004080311.2	Caspase-7	Oryzias Latipes	Fish	Beloniformes
XP_005520811.1	Predicted: Caspase-7	Pseudopodoces Humilis	Birds	Passeriformes
XP_021151155.1	Caspase-7 isoform X1	Columba Livia	Birds	Columbiformes
NP_001268771.1	Caspase-7	Mesocricetus Auratus	Mammalia	Rodentia
XP_005014501.1	Caspase-7	Anas Platyrhynchos	Birds	Anseriformes
XP_009965879.1	Predicted: Caspase-7	Tyto Alba	Birds	Strigiformes
XP_010288794.1	Predicted: Caspase-7	Phaethon Lepturus	Birds	Phaethontiformes

XP_009483464.1	Predicted: Caspase-7	Pelecanus Crispus	Birds	Pelecaniformes
XP_018419981.1	Predicted: Caspase-7 isoform X1	Nanorana Parkeri	Amphibia	Anura
XP_018618189.1	Predicted: Caspase-7	Scleropages Formosus	Fish	Osteoglossiformes
NP_001081408.1	Caspase-7	Xenopus Laevis	Amphibia	Anura
NP_001091272.1	Caspase-7 S Homeolog	Xenopus Laevis	Amphibia	Anura
NP_001219.2	Caspase-8 Isoform A Precursor	Homo Sapiens	Mammalia	Primates
NP_001125222.2	Caspase-8	Pongo Abelii	Mammalia	Primates
NP_001264855.1	Caspase-8 isoform 2	Mus Musculus	Mammalia	Rodentia
NP_071613.1	Caspase-8	Rattus Norvegicus	Mammalia	Rodentia
XP_010600786.1	Predicted: Caspase-8 isoform X1	Loxodonta Africana	Mammalia	Proboscidea
NP_571585.2	Caspase-8	Danio Rerio	Fish	Cypriniformes
NP_001041494.1	Caspase-8	Canis Lupus Familiaris	Mammalia	Carnivora
XP_007501684.1	Predicted: Caspase-8 isoform X1	Monodelphis Domestica	Mammalia	Didelphimorphia
XP_014587812.1	Predicted: Caspase-8	Equus Caballus	Mammalia	Perissodactyla
NP_989923.1	Caspase-8	Gallus Gallus	Birds	Galliformes
NP_001026949.2	Caspase-8	Sus Scrofa	Mammalia	Artiodactyla
NP_001039435.1	Caspase-8	Bos Taurus	Mammalia	Artiodactyla
XP_017953067.1	Predicted: Caspase-8	Xenopus Tropicalis	Amphibia	Anura
XP_010711755.1	Predicted: Caspase-8	Meleagris Gallopavo	Birds	Galliformes

XP_014319897.1	Predicted: Caspase-8	Myotis Lucifugus	Mammalia	Chiroptera
XP_007882891.1	Predicted: Caspase-8	Callorhinchus Milii	Fish	Chimaeriformes
XP_011615013.1	Predicted: Caspase-8	Takifugu Rubripes	Fish	Tetraodontiformes
NP_001233725.1	Caspase-8	Cricetulus Griseus	Mammalia	Rodentia
XP_005306309.1	Predicted: Caspase-8 isoform X1	Chrysemys Picta Bellii	Reptilia	Testudines
XP_015214952.1	Predicted: Caspase-8	Lepisosteus Oculatus	Fish	Lepisosteiformes
XP_006272599.1	Predicted: Caspase-8 isoform X1	Alligator Mississippiensis	Reptilia	Crocodilia
XP_006272609.1	Predicted: Caspase-8	Alligator Mississippiensis	Reptilia	Crocodilia
NP_001098258.1	Caspase-8	Oryzias Latipes	Fish	Beloniformes
XP_012950190.1	Predicted: Caspase-8 isoform X1	Anas Platyrhynchos	Birds	Anseriformes
XP_009963925.1	Predicted: Caspase-8	Tyto Alba	Birds	Strigiformes
XP_009701337.1	Predicted: Caspase-8	Cariama Cristata	Birds	Cariamiformes
XP_009807896.1	Predicted: Caspase-8	Gavia Stellata	Birds	Gaviiformes
XP_017583733.1	Predicted: Caspase-8 isoform X1	Corvus Brachyrhynchos	Birds	Passeriformes
XP_009938109.1	Predicted: Caspase-8	Opisthocomus Hoazin	Birds	Opisthocomiformes
XP_010017555.1	Predicted: Caspase-8	Nestor Notabilis	Birds	Psittaciformes
XP_009682548.1	Predicted: Caspase-8	Struthio Camelus Australis	Birds	Struthioniformes
XP_020795173.1	Caspase-8	Boleophthalmus Pectinirostris	Fish	Perciformes

XP_018585033.1	Predicted: Caspase-8 isoform X1	Scleropages Formosus	Fish	Osteoglossiformes
NP_001187127.1	Caspase-8	Ictalurus Punctatus	Fish	Siluriformes
NP_001079034.1	Caspase-8 L Homeolog	Xenopus Laevis	Amphibia	Anura
NP_116759.2	Caspase-10 Isoform 1 Preprotein	Homo Sapiens	Mammalia	Primates
NP_001127368.1	Caspase-10	Pongo Abelii	Mammalia	Primates
XP_007501670.1	Predicted: Capase-10	Monodelphis Domestica	Mammalia	Didelphimorphia
XP_005601738.1	Predicted: Capase-10	Equus Caballus	Mammalia	Perissodactyla
XP_421936.4	Predicted: Capase-10	Gallus Gallus	Birds	Galliformes
NP_001155112.1	Caspase-10	Sus Scrofa	Mammalia	Artiodactyla
NP_001093436.1	Caspase-10	Oryctolagus Cuniculus	Mammalia	Lagomorpha
NP_001015715.2	Caspase-10	Xenopus Tropicalis	Amphibia	Anura
XP_011218837.1	Predicted: Caspase-10 isoform X1	Ailuropoda Melanoleuca	Mammalia	Carnivora
XP_008118735.1	Predicted: Caspase-10 isoform X1	Anolis Carolinensis	Reptilia	Squamata
XP_010711745.1	Predicted: Capase-10	Meleagris Gallopavo	Birds	Galliformes
XP_006082464.1	Predicted: Capase-10	Myotis Lucifugus	Mammalia	Chiroptera
XP_011367468.1	Predicted: Caspase-10 isoform X1	Pteropus Vampyrus	Mammalia	Chiroptera
XP_003775343.2	Predicted: Capase-10	Sarcophilus Harrisii	Mammalia	Dasyuromorphia

XP_012917760.1	Predicted: Capase-10	Mustela Putorius Furo	Mammalia	Carnivora
XP_008175924.1	Predicted: Caspase-10 isoform X3	Chrysemys Picta Bellii	Reptilia	Testudines
XP_004378355.1	Predicted: Capase-10	Trichechus Manatus Latirostris	Mammalia	Sirenia
XP_005049234.1	Predicted: Capase-10	Ficedula Albicollis	Birds	Passeriformes
XP_004674884.1	Predicted: Capase-10	Condylura Cristata	Mammalia	Soricomorpha
XP_017802266.1	Caspase-10 Isoform X1	Papio Anubis	Mammalia	Primates
XP_014449789.1	Predicted: Caspase-10 isoform X1	Alligator Mississippiensis	Reptilia	Crocodilia
XP_014640222.1	Predicted: Capase-10	Ceratotherium simum simum	Mammalia	Perissodactyla
XP_012950189.1	Caspase-10	Anas Platyrhynchos	Birds	Anseriformes
XP_014374717.1	Predicted: Caspase-10 isoform X1	Alligator Sinensis	Reptilia	Crocodilia
XP_010288754.1	Predicted: Capase-10	Phaethon Lepturus	Birds	Phaethontiformes
XP_009483191.1	Predicted: Capase-10	Pelecanus Crispus	Birds	Pelecaniformes
XP_009701338.1	Predicted: Capase-10	Cariama Cristata	Birds	Cariamiformes
XP_009924272.1	Predicted: Capase-10	Haliaeetus Albicilla	Birds	Accipitriformes
XP_009949374.1	Predicted: Capase-10	Leptosomus Discolor	Birds	Leptosomiformes
XP_008941675.1	Predicted: Capase-10	Merops Nubicus	Birds	Coraciiformes
XP_009938108.1	Predicted: Capase-10	Opisthocomus Hoazin	Birds	Opisthocomiformes
XP_010017544.1	Predicted: Capase-10	Nestor Notabilis	Birds	Psittaciformes

NP_001081410.1	Caspase-10 S Homeolog	Xenopus Laevis	Amphibia	Anura
NP_001083130.1	Caspase-10 L Homeolog	Xenopus Laevis	Amphibia	Anura
XP_019378751.1	Predicted: Caspase-10 isoform X1	Gavialis Gangeticus	Reptilia	Crocodilia
NP_001300701.1	cFLIP	Danio Rerio	Fish	Cypriniformes
XP_008118737.1	cFLIP	Anolis Carolinensis	Reptilia	Squamata
XP_034993049.1	cFLIP	Zootoca Vivipara	Reptilia	Squamata
XP_015281011.1	cFLIP	Gekko Japonicus	Reptilia	Squamata
XP_026709097.1	cFLIP	Athene Cunicularia	Birds	Strigiformes
XP_009807246.1	cFLIP	Gavia Stellata	Birds	Gaviiformes
XP_009479033.1	cFLIP	Pelecanus Crispus	Birds	Pelecaniformes
XP_009282934.1	cFLIP	Aptenodytes Forsteri	Birds	Sphenisciformes
XP_019378755.1	cFLIP	Gavialis Gangeticus	Reptilia	Crocodilia
XP_006118011.1	cFLIP	Pelodiscus Sinensis	Reptilia	Testudines
XP_007501666.1	cFLIP	Monodelphis Domestica	Mammalia	Didelphimorphia
XP_031814208.1	cFLIP	Sarcophilus harrisii	Mammalia	Dasyuromorphia
XP_019294730.1	cFLIP	Panthera Pardus	Mammalia	Carnivora
XP_020930360.1	cFLIP	Sus Scrofa	Mammalia	Artiodactyla
XP_032494111.1	cFLIP	Phocoena Sinus	Mammalia	Cetacea
XP_027446588.1	cFLIP	Zalophus Californianus	Mammalia	Carnivora

XP_019491241.1	cFLIP	Hipposideros Armiger	Mammalia	Chiroptera
NP_001120655.1	cFLIP	Homo Sapiens	Mammalia	Primates
NP_001125140.1	cFLIP	Pongo Abelii	Mammalia	Primates
XP_028365650.1	cFLIP	Phyllostomus Discolor	Mammalia	Chiroptera
NP_001188445.1	cFLIP	Oryzias Latipes	Fish	Beloniformes
XP_022595948.1	cFLIP	Seriola Dumerili	Fish	Perciformes
XP_026152299.1	cFLIP	Mastacembelus Armatus	Fish	Synbranchiformes
XP_019945253.1	cFLIP	Paralichthys Olivaceus	Fish	Pleuronectiformes
NP_001254595.1	cFLIP	Gasterosteus Aculeatus	Fish	Gasterosteiformes

Table S2

AOA 2				
Accenssion ID	Caspase	Species	Class	Order
NP_004337.2	Caspase-3 isoform a preproprotein	Homo Sapiens	Mammalia	Primates
NP_001012435.1	Caspase-3 isoform a preproprotein	Pan Troglodyte	Mammalia	Primates
NP_571952.1	Caspase-3 apoptosis-related cysteine peptidase a	Danio Rerio	Fish	Cypriniformes
NP_001003042.1	Caspase-3	Canis Lupus Familiaris	Mammalia	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_014114736.1	Predicted: Caspase-3	Pseudopodoces Humilis	Birds	Passeriformes
XP_005445415.1	Predicted: Caspase-3	Falco Cherrug	Birds	Falconiformes
XP_007054526.1	Predicted: Caspase-3	Chelonia Mydas	Reptilia	Testudines
XP_005030551.1	Caspase-3	Anas Platyrhynchos	Birds	Anseriformes
NP_001271764.1	Caspase-3	Macaca Fascicularis	Mammalia	Primates
XP_006026683.1	Predicted: Caspase-3	Alligator Sinensis	Reptilia	Crocodilia
XP_007549682.1	Predicted: Caspase-3 isoform X1	Poecilia Formosa	Fish	Cyprinodontiformes
XP_009970400.1	Predicted: Caspase-3	Tyto Alba	Birds	Strigiformes
XP_009479867.1	Predicted: Caspase-3	Pelecanus Crispus	Birds	Pelecaniformes

XP_009922022.1	Predicted: Caspase-3	Haliaeetus Albicilla	Birds	Accipitriformes
NP_001290581.1	Caspase-3	Esox Lucius	Fish	Esociformes
XP_015687286.1	Predicted: Caspase-3	Protobothrops Mucrosquamatus	Reptilia	Squamata
NP_001188010.1	Caspase-3	Ictalurus Punctatus	Fish	Siluriformes
NP_001081225.1	Caspase-3	Xenopus Laevis	Amphibia	Anura
NP_001273018.1	Caspase-3	Capra Hircus	Mammalia	Artiodactyla
NP_001269823.1	Caspase-3	Oreochromis Niloticus	Fish	Perciformes
NP_001217.2	Caspase-6 Isoform Alpha Precursor	Homo Sapiens	Mammalia	Primates
NP_033941.3	Caspase-6 Precursor	Mus Musculus	Mammalia	Rodentia
NP_113963.2	Caspase-6	Rattus Norvegicus	Mammalia	Rodentia
XP_010588808.1	Predicted: Caspase-6	Loxodonta Africana	Mammalia	Proboscidea
NP_001018333.1	Caspase-6	Danio Rerio	Fish	Cypriniformes
NP_990057.1	Caspase-6	Gallus Gallus	Birds	Galliformes
NP_001030496.1	Caspase-6	Bos Taurus	Mammalia	Artiodactyla
NP_001011068.1	Caspase-6	Xenopus Tropicalis	Amphibia	Anura
XP_011359747.1	Predicted: Caspase-6 Isoform X1	Pteropus Vampyrus	Mammalia	Chiroptera
XP_015101393.1	Predicted: Caspase-6	Vicugna Pacos	Mammalia	Artiodactyla
XP_004748146.1	Predicted: Caspase-6 Isoform X2	Mustela Putorius Furo	Mammalia	Carnivora
XP_014427079.1	Predicted: Caspase-6	Pelodiscus Sinensis	Reptilia	Testudines
XP_006629925.2	Predicted: Caspase-6	Lepisosteus Oculatus	Holostei	Lepisosteiformes

XP_004380293.1	Predicted: Caspase-6	Trichechus Manatus Latirostris	Mammalia	Sirenia
XP_005044995.1	Predicted: Caspase-6 Isoform X1	Ficedula Albicollis	Birds	Passeriformes
XP_012583463.1	Predicted: Caspase-6 Isoform X2	Condylura Cristata	Mammalia	Soricomorpha
XP_006881156.1	Predicted: Caspase-6	Elephantulus Edwardii	Mammalia	Macroscelidea
XP_004084686.1	Caspase-6	Oryzias Latipes	Fish	Beloniformes
XP_012860317.1	Predicted: Caspase-6	Echinops Telfairi	Mammalia	Afrosoricida
XP_013155138.1	Predicted: Caspase-6	Falco Peregrinus	Birds	Falconiformes
XP_007054543.1	Predicted: Caspase-6	Chelonia Mydas	Reptilia	Testudines
XP_021128354.1	Caspase-6	Anas Platyrhynchos	Birds	Anseriformes
XP_006025610.2	Predicted: Caspase-6	Alligator Sinensis	Reptilia	Crocodilia
XP_007560684.1	Predicted: Caspase-6 Isoform X1	Poecilia Formosa	Fish	Cyprinodontiformes
XP_009966306.1	Predicted: Caspase-6	Tyto Alba	Birds	Strigiformes
XP_010290767.1	Predicted: Caspase-6	Phaethon Lepturus	Birds	Phaethontiformes
XP_009483599.1	Predicted: Caspase-6	Pelecanus Crispus	Birds	Pelecaniformes
XP_009699574.1	Predicted: Caspase-6	Cariama Cristata	Birds	Cariamiformes
XP_008285572.1	Predicted: Caspase-6	Stegastes Partitus	Fish	Perciformes
XP_019110884.1	Predicted: Caspase-6 Isoform X1	Larimichthys Crocea	Fish	Perciformes
XP_015666700.1	Predicted: Caspase-6	Protobothrops Mucrosquamatus	Reptilia	Squamata
XP_018611669.1	Predicted: Caspase-6	Scleropages Formosus	Fish	Osteoglossiformes
NP_001081406.1	Caspase-6 L Homeolog	Xenopus Laevis	Amphibia	Anura

XP_019377505.1	Predicted: Caspase-6 Isoform X1	Gavialis Gangeticus	Reptilia	Crocodilia
NP_001117743.1	Caspase-6 Precursor	Oncorhynchus Mykiss	Fish	Salmoniformes
NP_001253985.1	Caspase-7 isoform alpha precursor	Homo Sapiens	Mammalia	Primates
NP_071596.1	Caspase-7	Rattus Norvegicus	Mammalia	Rodentia
XP_010587277.1	Predicted: Caspase-7 isoform X1	Loxodonta Africana	Mammalia	Proboscidea
NP_001018443.1	Caspase-7	Danio Rerio	Fish	Cypriniformes
XP_421764.3	Predicted: Caspase-7	Gallus Gallus	Birds	Galliformes
XP_002698555.1	Predicted: Caspase-7	Bos Taurus	Mammalia	Artiodactyla
NP_001016299.1	Caspase-7	Xenopus Tropicalis	Amphibia	Anura
XP_011230112.1	Predicted: Caspase-7 isoform X2	Ailuropoda Melanoleuca	Mammalia	Carnivora
XP_008112945.1	Predicted: Caspase-7 isoform X2	Anolis Carolinensis	Reptilia	Squamata
XP_014310783.1	Predicted: Caspase-7	Myotis Lucifugus	Mammalia	Chiroptera
XP_004612023.2	Predicted: Caspase-7	Sorex Araneus	Mammalia	Soricomorpha
XP_006135034.1	Predicted: Caspase-7	Pelodiscus Sinensis	Reptilia	Testudines
XP_005154576.1	Predicted: Caspase-7 isoform X1	Melopsittacus Undulatus	Birds	Psittaciformes
XP_005747456.1	Predicted: Caspase-7 isoform X1	Pundamilia Nyererei	Actinopterygii	Cichliformes
XP_004681050.1	Predicted: Caspase-7	Condylura Cristata	Mammalia	Soricomorpha
XP_014450146.1	Predicted: Caspase-7 isoform X1	Alligator Mississippiensis	Reptilia	Crocodilia
XP_014641239.1	Predicted: Caspase-7	Ceratotherium simum simum	Mammalia	Perissodactyla
XP_006831548.1	Predicted: Caspase-7	Chrysochloris Asiatica	Mammalia	Afrosoricida

XP_005888715.1	Predicted: Caspase-7	Bos Taurus	Mammalia	Artiodactyla
XP_005520811.1	Predicted: Caspase-7	Pseudopodoces Humilis	Birds	Passeriformes
XP_005239289.1	Predicted: Caspase-7 isoform X1	Falco Peregrinus	Birds	Falconiformes
NP_001268771.1	Caspase-7	Mesocricetus Auratus	Mammalia	Rodentia
XP_005481147.1	Predicted: Caspase-7	Zonotrichia Albicollis	Birds	Passeriformes
XP_007574351.1	Predicted: Caspase-7 isoform X1	Poecilia Formosa	Fish	Cyprinodontiformes
XP_008312002.1	Predicted: Caspase-7 isoform X1	Cynoglossus Semilaevis	Fish	Pleuronectiformes
XP_009697487.1	Predicted: Caspase-7	Cariama Cristata	Birds	Cariamiformes
XP_010198191.1	Predicted: Caspase-7	Colius Striatus	Birds	Coliiformes
XP_008275786.1	Predicted: Caspase-7	Stegastes Partitus	Fish	Perciformes
XP_009578529.1	Predicted: Caspase-7	Fulmarus Glacialis	Birds	Procellariiformes
XP_009923386.1	Predicted: Caspase-7	Haliaeetus Albicilla	Birds	Accipitriformes
XP_010006222.1	Predicted: Caspase-7 isoform X1	Chaetura Pelagica	Birds	Apodiformes
XP_018419981.1	Predicted: Caspase-7 isoform X1	Nanorana Parkeri	Amphibia	Anura
XP_012670088.1	Predicted: Caspase-7	Clupea Harengus	Fish	Clupeiformes
XP_018618189.1	Predicted: Caspase-7	Scleropages Formosus	Fish	Osteoglossiformes
NP_001091272.1	Caspase-7 S Homeolog	Xenopus Laevis	Amphibia	Anura
NP_001219.2	Caspase-8 Isoform A Precursor	Homo Sapiens	Mammalia	Primates
NP_001125222.2	Caspase-8	Pongo Abelii	Mammalia	Primates
NP_001264855.1	Caspase-8 Isoform 2	Mus Musculus	Mammalia	Rodentia

NP_071613.1	Caspase-8	Rattus Norvegicus	Mammalia	Rodentia
NP_571585.2	Caspase-8	Danio Rerio	Fish	Cypriniformes
XP_007663299.1	Predicted: Caspase-8	Ornithorhynchus Anatinus	Mammalia	Monotremata
NP_001041494.1	Caspase-8	Canis Lupus Familiaris	Mammalia	Carnivora
NP_989923.1	Caspase-8	Gallus Gallus	Birds	Galliformes
NP_001026949.2	Caspase-8	Sus Scrofa	Mammalia	Artiodactyla
NP_001039435.1	Caspase-8	Bos Taurus	Mammalia	Artiodactyla
XP_017953067.1	Predicted: Caspase-8	Xenopus Tropicalis	Amphibia	Anura
XP_011367467.1	Predicted: Caspase-8	Pteropus Vampyrus	Mammalia	Chiroptera
XP_004601328.1	Predicted: Caspase-8	Sorex Araneus	Mammalia	Soricomorpha
XP_012380773.1	Predicted: Caspase-8	Dasypus Novemcinctus	Mammalia	Cingulata
NP_001233725.1	Caspase-8	Cricetulus Griseus	Mammalia	Rodentia
XP_005306309.1	Predicted: Caspase-8 isoform X1	Chrysemys Picta Bellii	Reptilia	Testudines
XP_006272599.1	Predicted: Caspase-8 isoform X1	Alligator Mississippiensis	Reptilia	Crocodilia
XP_006272609.1	Predicted: Caspase-8	Alligator Mississippiensis	Reptilia	Crocodilia
NP_001098258.1	Caspase-8	Oryzias Latipes	Fish	Beloniformes
XP_009963925.1	Predicted: Caspase-8	Tyto Alba	Birds	Strigiformes
XP_009807896.1	Predicted: Caspase-8	Gavia Stellata	Birds	Gaviiformes
XP_009282932.1	Predicted: Caspase-8	Aptenodytes Forsteri	Birds	Sphenisciformes
XP_009876853.1	Predicted: Caspase-8	Apaloderma Vittatum	Birds	Trogoniformes

XP_009888484.1	Predicted: Caspase-8	Charadrius Vociferus	Birds	Charadriiformes
XP_009460551.1	Predicted: Caspase-8	Nipponia Nippon	Birds	Pelecaniformes
XP_010563026.1	Predicted: Caspase-8	Haliaeetus Leucocephalus	Birds	Accipitriformes
XP_010404904.1	Caspase-8	Corvus Cornix Cornix	Birds	Passeriformes
XP_021175080.1	Caspase-8	Fundulus Heteroclitus	Fish	Cyprinodontiformes
XP_018585033.1	Predicted: Caspase-8 isoform X1	Scleropages Formosus	Fish	Osteoglossiformes
NP_001187127.1	Caspase-8	Ictalurus Punctatus	Fish	Siluriformes
NP_001079034.1	Caspase-8 L Homeolog	Xenopus Laevis	Amphibia	Anura
XP_008929353.1	Predicted: Caspase-8	Manacus Vitellinus	Birds	Passeriformes
XP_003457507.2	Predicted: Caspase-8	Oreochromis Niloticus	Fish	Perciformes
XP_021444269.1	Predicted: Caspase-8 isoform X1	Oncorhynchus Mykiss	Fish	Salmoniformes
XP_022610907.1	Predicted: Caspase-8 isoform X1	Seriola Dumerili	Fish	Perciformes
NP_116759.2	Caspase-10 Isoform 1 Preprotein	Homo Sapiens	Mammalia	Primates
NP_001127368.1	Caspase-10	Pongo Abelii	Mammalia	Primates
XP_005640587.1	Caspase-10	Canis Lupus Familiaris	Mammalia	Carnivora
XP_421936.4	Predicted: Caspase-10	Gallus Gallus	Birds	Galliformes
NP_001155112.1	Caspase-10	Sus Scrofa	Mammalia	Artiodactyla
NP_001093436.1	Caspase-10	Oryctolagus Cuniculus	Mammalia	Lagomorpha
NP_001015715.2	Caspase-10	Xenopus Tropicalis	Amphibia	Anura
XP_008118735.1	Predicted: Caspase-10 isoform X1	Anolis Carolinensis	Reptilia	Squamata

XP_004458966.1	Predicted: Caspase-10	Dasypus Novemcinctus	Mammalia	Cingulata
XP_021591334.1	Caspase-10	Ictidomys Tridecemlineatus	Mammalia	Rodentia
XP_008175924.1	Predicted: Caspase-10 isoform X3	Chrysemys Picta Bellii	Reptilia	Testudines
XP_014449789.1	Predicted: Caspase-10 isoform X1	Alligator Mississippiensis	Reptilia	Crocodilia
XP_004577327.1	Predicted: Caspase-10	Ochotona Princeps	Mammalia	Lagomorpha
XP_006862096.1	Predicted: Caspase-10	Chrysochloris Asiatica	Mammalia	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_009500813.1	Predicted: Caspase-10	Phalacrocorax Carbo	Birds	Suliformes
XP_010563039.1	Predicted: Caspase-10	Haliaeetus Leucocephalus	Birds	Accipitriformes
XP_011574974.1	Predicted: Caspase-10	Aquila Chrysaetos Canadensis	Birds	Accipitriformes
NP_001081410.1	Caspase-10 S Homeolog	Xenopus Laevis	Amphibia	Anura
NP_001083130.1	Caspase-10 L Homeolog	Xenopus Laevis	Amphibia	Anura
XP_017527989.1	Predicted: Caspase-10	Manis Javanica	Mammalia	Pholidota
XP_019378751.1	Predicted: Caspase-10 isoform X1	Gavialis Gangeticus	Reptilia	Crocodilia
NP_001300701.1	cFLIP	Danio Rerio	Fish	Cypriniformes
XP_033025467.1	cFLIP	Lacerta Agilis	Reptilia	Squamata
XP_020667408.1	cFLIP	Pogona Vitticeps	Reptilia	Squamata
XP_009330029.1	cFLIP	Pygoscelis Adeliae	Birds	Sphenisciformes
XP_019409414.1	cFLIP	Crocodylus Porosus	Reptilia	Crocodilia
XP_006022153.1	cFLIP	Alligator Sinensis	Reptilia	Crocodilia
XP_006118011.1	cFLIP	Pelodiscus Sinensis	Reptilia	Testudines
XP_010005644.1	cFLIP	Chaetura Pelagica	Birds	Apodiformes
XP_010123198.1	cFLIP	Chlamydotis Macqueenii	Birds	Otidiformes
XP_009707893.1	cFLIP	Cariama Cristata	Birds	Cariamiformes
XP_030180293.1	cFLIP	Lynx Pardinus	Mammalia	Carnivora
XP_020930360.1	cFLIP	Sus Scrofa	Mammalia	Artiodactyla
XP_022423099.1	cFLIP	Delphinapterus Leucas	Mammalia	Cetacea

XP_026980099.1	cFLIP	Lagenorhynchus Obliquidens	Mammalia	Cetacea
XP_025749261.1	cFLIP	Callorhinus Ursinus	Mammalia	Carnivora
XP_005640588.2	cFLIP	Canis Lupus Familiaris	Mammalia	Carnivora
XP_035161179.1	cFLIP	Callithrix Jacchus	Mammalia	Primates
NP_001120655.1	cFLIP	Homo Sapiens	Mammalia	Primates
NP_001125140.1	cFLIP	Pongo Abelii	Mammalia	Primates
XP_024419473.1	cFLIP	Desmodus Rotundus	Mammalia	Chiroptera
NP_001188445.1	cFLIP	Oryzias Latipes	Fish	Beloniformes
XP_023284372.1	cFLIP	Seriola Lalandi	Fish	Perciformes
XP_020450436.1	cFLIP	Monopterus Albus	Fish	Synbranchiformes
XP_034740679.1	cFLIP	Etheostoma Cragini	Fish	Perciformes
NP_001254595.1	cFLIP	Gasterosteus Aculeatus	Fish	Gasterosteiformes

Table S3

AOA 3				
Accenssion ID	Caspase	Species	Class	Order
NP_004337.2	Caspase-3 isoform a preprotein	Homo Sapiens	Mammalia	Primates
NP_001012435.1	Caspase-3	Pan Troglodytes	Mammalia	Primates
NP_001271338.1	Caspase-3	Mus Musculus	Mammalia	Rodentia
NP_037054.1	Caspase-3	Rattus Norvegicus	Mammalia	Rodentia
NP_571952.1	Caspase-3 Apoptotsis-related cysteine peptidase a	Danio Rerio	Fish	Cypriniformes

NP_001003042.1	Caspase-3	Canis Lupus Familiaris	Mammalia	Carnivora
NP_001157433.1	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_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_007905080.1	Predicted: Caspase-3	Callorhinchus Milii	Chondrichthyes	Chimaeriformes
NP_001230975.1	Caspase-3	Cricetulus Griseus	Mammalia	Rodentia
XP_006128558.1	Predicted: Caspase-3 isoform X1	Pelodiscus Sinensis	Reptilia	Testudines
NP_001266895.1	Caspase-3	Saimiri Boliviensis	Mammalia	Primates
XP_005719459.1	Predicted: Caspase-3	Pundamilia Nyererei	Actinopterygii	Cichliformes
XP_019336883.1	Predicted: Caspase-3	Alligator Mississippiensis	Reptilia	Crocodilia
XP_004428794.1	Predicted: Caspase-3	Ceratotherium simum simum	Mammalia	Perissodactyla
XP_006834493.1	Predicted: Caspase-3	Chrysochloris Asiatica	Mammalia	Afrosoricida
NP_001098140.1	Caspase-3	Oryzias Latipes	Fish	Beloniformes
XP_006160500.1	Predicted: Caspase-3 isoform X2	Tupaia Chinensis	Mammalia	Scandentia
XP_013161017.1	Predicted: Caspase-3 isoform X1	Falco Peregrinus	Birds	Falconiformes
XP_007054526.1	Predicted: Caspase-3	Chelonia Mydas	Reptilia	Testudines
XP_014120317.1	Predicted: Caspase-3	Zonotrichia Albicollis	Birds	Passeriformes

XP_005972660.1	Predicted: Caspase-3	Pantholops Hodgsonii	Mammalia	Artiodactyla
XP_006026683.1	Predicted: Caspase-3	Alligator Sinensis	Reptilia	Crocodilia
XP_009099177.1	Predicted: Caspase-3 isoform X1	Serinus Canaria	Birds	Passeriformes
XP_010281454.1	Predicted: Caspase-3	Phaethon Lepturus	Birds	Phaethontiformes
XP_009697353.1	Predicted: Caspase-3	Cariama Cristata	Birds	Cariamiformes
XP_009570651.1	Predicted: Caspase-3	Fulmarus Glacialis	Birds	Procellariiformes
NP_001290581.1	Caspase-3	Esox Lucius	Fish	Esociformes
NP_001188010.1	Caspase-3	Ictalurus Punctatus	Fish	Siluriformes
NP_001081225.1	Caspase-3	Xenopus Laevis	Amphibia	Anura
NP_001269823.1	Caspase-3	Oreochromis Niloticus	Fish	Perciformes
NP_001217.2	Caspase-6 Isoform Alpha Precursor	Homo Sapiens	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
NP_990057.1	Caspase-6	Gallus Gallus	Birds	Galliformes
XP_005656604.1	Predicted: Caspase-6 isoform X1	Sus Scrofa	Mammalia	Artiodactyla
NP_001030496.1	Caspase-6	Bos Taurus	Mammalia	Artiodactyla
NP_001011068.1	Caspase-6	Xenopus Tropicalis	Amphibia	Anura
XP_003221840.2	Predicted: Caspase-6	Anolis Carolinensis	Reptilia	Squamata
XP_014427079.1	Predicted: Caspase-6	Pelodiscus Sinensis	Reptilia	Testudines

XP_019355646.1	Predicted: Caspase-6 isoform X1	Alligator Mississippiensis	Reptilia	Crocodilia
XP_014418265.1	Predicted: Caspase-6	Camelus Ferus	Mammalia	Artiodactyla
XP_004411305.1	Predicted: Caspase-6	Odobenus Rosmarus Divergens	Mammalia	Carnivora
XP_006918975.1	Predicted: Caspase-6 isoform X1	Pteropus Alecto	Mammalia	Chiroptera
XP_007054543.1	Predicted: Caspase-6	Chelonia Mydas	Reptilia	Testudines
XP_014120908.1	Predicted: Caspase-6 isoform X1	Zonotrichia Albicollis	Birds	Passeriformes
XP_007463292.1	Predicted: Caspase-6	Lipotes Vexillifer	Mammalia	Cetacea
XP_008315389.1	Predicted: Caspase-6	Cynoglossus Semilaevis	Fish	Pleuronectiformes
XP_008841740.1	Predicted: Caspase-6	Nannospalax Galili	Mammalia	Rodentia
XP_008420898.1	Predicted: Caspase-6	Poecilia Reticulata	Fish	Cyprinodontiformes
XP_009578173.1	Predicted: Caspase-6	Fulmarus Glacialis	Birds	Procellariiformes
XP_009924247.1	Predicted: Caspase-6	Haliaeetus Albicilla	Birds	Accipitriformes
XP_017592325.1	Predicted: Caspase-6	Corvus Brachyrhynchos	Birds	Passeriformes
XP_009930434.1	Predicted: Caspase-6	Opisthocomus Hoazin	Birds	Opisthocomiformes
XP_010187950.1	Predicted: Caspase-6	Mesitornis Unicolor	Birds	Mesitornithiformes
XP_019326128.1	Predicted: Caspase-6	Aptenodytes Forsteri	Birds	Sphenisciformes
XP_012724337.1	Predicted: Caspase-6 isoform X1	Fundulus Heteroclitus	Fish	Cyprinodontiformes
XP_018425527.1	Predicted: Caspase-6	Nanorana Parkeri	Amphibia	Anura
XP_019110884.1	Predicted: Caspase-6 isoform X1	Larimichthys Crocea	Fish	Perciformes
XP_013863217.1	Predicted: Caspase-6	Austrofundulus Limnaeus	Fish	Cyprinodontiformes

XP_015281944.1	Predicted: Caspase-6 isoform X1	Gekko Japonicus	Reptilia	Squamata
XP_015666700.1	Predicted: Caspase-6	Protobothrops Mucrosquamatus	Reptilia	Squamata
XP_018539675.1	Predicted: Caspase-6	Lates Calcarifer	Fish	Perciformes
NP_001081406.1	Caspase-6 L Homeolog	Xenopus Laevis	Amphibia	Anura
NP_001117743.1	Caspase-6 Precursor	Oncorhynchus Mykiss	Fish	Salmoniformes
NP_001253985.1	Caspase-7 Isoform Alpha Precursor	Homo Sapiens	Mammalia	Primates
NP_071596.1	Caspase-7	Rattus Norvegicus	Mammalia	Rodentia
NP_001018443.1	Caspase-7	Danio Rerio	Fish	Cypriniformes
XP_421764.3	Predicted: Caspase-7	Gallus Gallus	Birds	Galliformes
NP_001016299.1	Caspase-7	Xenopus Tropicalis	Amphibia	Anura
XP_008112945.1	Predicted: Caspase-7 isoform X2	Anolis Carolinensis	Reptilia	Squamata
XP_015091414.1	Predicted: Caspase-7	Vicugna Pacos	Mammalia	Artiodactyla
XP_004763887.1	Predicted: Caspase-7	Mustela Putorius Furo	Mammalia	Carnivora
XP_005928237.1	Predicted: Caspase-7 isoform X1	Haplochromis Burtoni	Actinopterygii	Cichliformes
XP_012410596.1	Predicted: Caspase-7	Trichechus Manatus Latirostris	Mammalia	Sirenia
XP_013371286.1	Predicted: Caspase-7 isoform X1	Chinchilla Lanigera	Mammalia	Rodentia
XP_016042302.1	Predicted: Caspase-7	Erinaceus Europaeus	Mammalia	Erinaceomorpha
XP_005888715.1	Predicted: Caspase-7	Bos Mutus	Mammalia	Artiodactyla
XP_008142545.1	Predicted: Caspase-7 isoform X1	Eptesicus Fuscus	Mammalia	Chiroptera
XP_014448895.1	Predicted: Caspase-7	Tupaia Chinensis	Mammalia	Scandentia

XP_005443932.1	Predicted: Caspase-7 isoform X1	Falco Cherrug	Birds	Falconiformes
NP_001268771.1	Caspase-7	Mesocricetus Auratus	Mammalia	Rodentia
XP_006022892.1	Predicted: Caspase-7 isoform X1	Alligator Sinensis	Reptilia	Crocodilia
XP_009085424.1	Predicted: Caspase-7	Serinus Canaria	Birds	Passeriformes
XP_008434898.1	Predicted: Caspase-7 isoform X1	Poecilia Reticulata	Fish	Cyprinodontiformes
XP_008947331.1	Predicted: Caspase-7 isoform X1	Merops Nubicus	Birds	Coraciiformes
XP_009932919.1	Predicted: Caspase-7	Opisthocomus Hoazin	Birds	Opisthocomiformes
XP_010180183.1	Predicted: Caspase-7 isoform X1	Mesitornis Unicolor	Birds	Mesitornithiformes
XP_010010881.1	Predicted: Caspase-7	Nestor Notabilis	Birds	Psittaciformes
XP_009666939.1	Predicted: Caspase-7	Struthio Camelus Australis	Birds	Struthioniformes
XP_010895864.1	Predicted: Caspase-7	Esox Lucius	Fish	Esociformes
XP_020773763.1	Caspase-7	Boleophthalmus Pectinirostris	Fish	Perciformes
XP_018419981.1	Predicted: Caspase-7 isoform X1	Nanorana Parkeri	Amphibia	Anura
XP_010740374.1	Predicted: Caspase-7	Larimichthys Crocea	Fish	Perciformes
XP_013863024.1	Predicted: Caspase-7 isoform X1	Austrofundulus Limnaeus	Fish	Cyprinodontiformes
XP_018618189.1	Predicted: Caspase-7	Scleropages Formosus	Fish	Osteoglossiformes
NP_001081408.1	Caspase-7	Xenopus Laevis	Amphibia	Anura
NP_001091272.1	Caspase-7 S Homeolog	Xenopus Laevis	Amphibia	Anura
XP_019411759.1	Predicted: Caspase-7 isoform X1	Crocodylus Porosus	Reptilia	Crocodilia
XP_020635566.1	Caspase-7	Pogona Vitticeps	Reptilia	Squamata

NP_001219.2	Caspase-8 Isoform A Precursor	Homo Sapiens	Mammalia	Primates
NP_001125222.2	Caspase-8	Pongo Abelii	Mammalia	Primates
NP_001264855.1	Caspase-8 Isoform 2	Mus Musculus	Mammalia	Rodentia
NP_071613.1	Caspase-8	Rattus Norvegicus	Mammalia	Rodentia
NP_571585.2	Caspase-8	Danio Rerio	Fish	Cypriniformes
NP_001041494.1	Caspase-8	Canis Lupus Familiaris	Mammalia	Carnivora
NP_989923.1	Caspase-8	Gallus Gallus	Birds	Galliformes
NP_001026949.2	Caspase-8	Sus Scrofa	Mammalia	Artiodactyla
NP_001039435.1	Caspase-8	Bos Taurus	Mammalia	Artiodactyla
XP_017953067.1	Predicted: Caspase-8	Xenopus Tropicalis	Amphibia	Anura
XP_010711755.1	Predicted: Caspase-8	Meleagris Gallopavo	Birds	Galliformes
NP_001233725.1	Caspase-8	Cricetulus Griseus	Mammalia	Rodentia
XP_012778477.1	Caspase-8	Maylandia Zebra	Fish	Perciformes
XP_005306309.1	Predicted: Caspase-8 isoform X1	Chrysemys Picta Bellii	Reptilia	Testudines
XP_015214952.1	Predicted: Caspase-8	Lepisosteus Oculatus	Holostei	Lepisosteiformes
XP_004378277.1	Predicted: Caspase-8 isoform X1	Trichechus Manatus Latirostris	Mammalia	Sirenia
XP_019830532.1	Predicted: Caspase-8	Bos Indicus	Mammalia	Artiodactyla
XP_012581231.1	Predicted: Caspase-8 isoform X1	Condylura Cristata	Mammalia	Soricomorpha
XP_012803201.1	Predicted: Caspase-8	Jaculus Jaculus	Mammalia	Rodentia
XP_006272599.1	Predicted: Caspase-8 isoform X1	Alligator Mississippiensis	Reptilia	Crocodilia

XP_006272609.1	Predicted: Caspase-8	Alligator Mississippiensis	Reptilia	Crocodilia
NP_001098258.1	Caspase-8	Oryzias Latipes	Fish	Beloniformes
XP_012950190.1	Caspase-8 isoform X1	Anas Platyrhynchos	Birds	Anseriformes
XP_010017555.1	Predicted: Caspase-8	Nestor Notabilis	Birds	Psittaciformes
XP_009682548.1	Predicted: Caspase-8	Struthio Camelus Australis	Birds	Struthioniformes
XP_009894783.1	Predicted: Caspase-8	Picoides Pubescens	Birds	Piciformes
XP_009876853.1	Predicted: Caspase-8	Apaloderma Vittatum	Birds	Trogoniformes
XP_009888484.1	Predicted: Caspase-8	Charadrius Vociferus	Birds	Charadriiformes
XP_009460551.1	Predicted: Caspase-8	Nipponia Nippon	Birds	Pelecaniformes
XP_010563026.1	Predicted: Caspase-8	Haliaeetus Leucocephalus	Birds	Accipitriformes
XP_021175123.1	Caspase-8	Fundulus Heteroclitus	Fish	Cyprinodontiformes
XP_018585033.1	Predicted: Caspase-8 isoform X1	Scleropages Formosus	Fish	Osteoglossiformes
NP_001187127.1	Caspase-8	Ictalurus Punctatus	Fish	Siluriformes
NP_001079034.1	Caspase-8 L Homeolog	Xenopus Laevis	Amphibia	Anura
XP_021444269.1	Caspase-8 isoform X1	Oncorhynchus Mykiss	Fish	Salmoniformes
NP_116759.2	Caspase-10 Isoform 1 Preprotein	Homo Sapiens	Mammalia	Primates
NP_001127368.1	Caspase-10	Pongo Abelii	Mammalia	Primates
XP_421936.4	Predicted: Caspase-10	Gallus Gallus	Birds	Galliformes
NP_001155112.1	Caspase-10	Sus Scrofa	Mammalia	Artiodactyla
NP_001093436.1	Caspase-10	Oryctolagus Cuniculus	Mammalia	Lagomorpha

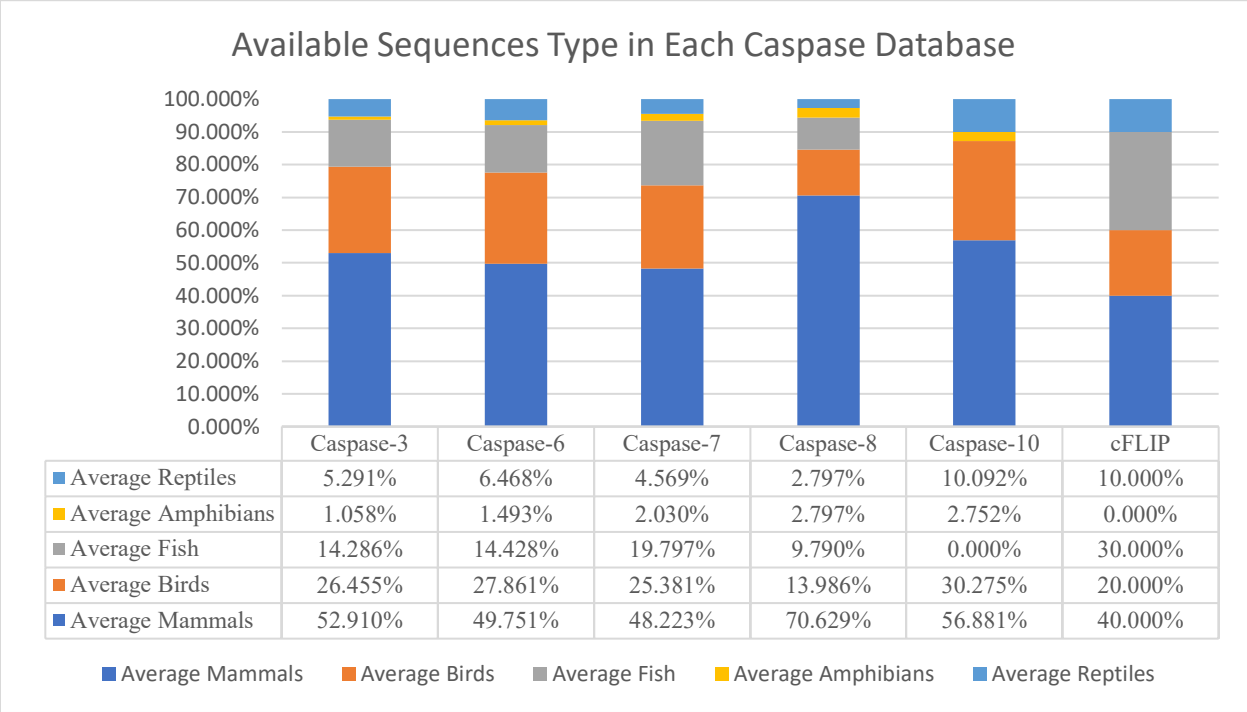
NP_001015715.2	Caspase-10	Xenopus Tropicalis	Amphibia	Anura
XP_002749670.1	Predicted: Caspase-10 isoform X1	Callithrix Jacchus	Mammalia	Primates
XP_008118735.1	Predicted: Caspase-10 isoform X1	Anolis Carolinensis	Reptilia	Squamata
XP_012354911.1	Predicted: Caspase-10 isoform X1	Nomascus Leucogenys	Mammalia	Primates
XP_008175924.1	Predicted: Caspase-10 isoform X3	Chrysemys Picta Bellii	Reptilia	Testudines
XP_005049234.1	Predicted: Caspase-10	Ficedula Albicollis	Reptilia	Crocodilia
XP_014449789.1	Predicted: Caspase-10 isoform X1	Alligator Mississippiensis	Reptilia	Crocodilia
XP_016042508.1	Predicted: Caspase-10	Erinaceus Europaeus	Mammalia	Erinaceomorpha
XP_012421961.1	Predicted: Caspase-10 isoform X1	Odobenus Rosmarus Divergens	Mammalia	Carnivora
XP_015414641.1	Predicted: Caspase-10	Myotis Davidii	Mammalia	Chiroptera
XP_014445038.1	Predicted: Caspase-10 isoform X1	Tupaia Chinensis	Mammalia	Scandentia
XP_005573964.1	Predicted: Caspase-10	Macaca Fascicularis	Mammalia	Primates
XP_005484241.1	Predicted: Caspase-10	Zonotrichia Albicollis	Birds	Passeriformes
XP_014374717.1	Predicted: Caspase-10 isoform X1	Alligator Sinensis	Reptilia	Crocodilia
XP_007090058.1	Predicted: Caspase-10	Panthera Tigris Altaica	Mammalia	Carnivora
XP_009086338.1	Predicted: Caspase-10	Serinus Canaria	Birds	Passeriformes
XP_009963924.1	Predicted: Caspase-10	Tyto Alba	Birds	Strigiformes
XP_017583676.1	Predicted: Caspase-10 isoform X1	Corvus Brachyrhynchos	Birds	Passeriformes
XP_009282933.1	Predicted: Caspase-10	Aptenodytes Forsteri	Birds	Sphenisciformes
XP_010165007.1	Predicted: Caspase-10	Antrostomus Carolinensis	Birds	Caprimulgiformes

XP_010222872.1	Predicted: Caspase-10	Tinamus Guttatus	Birds	Tinamiformes
XP_013043023.1	Predicted: Caspase-10 isoform X1	Anser Cygnoides Domesticus	Birds	Anseriformes
XP_014941825.1	Predicted: Caspase-10	Acinonyx Jubatus	Mammalia	Carnivora
XP_015986581.1	Predicted: Caspase-10	Rousettus Aegyptiacus	Mammalia	Chiroptera
XP_015489705.1	Predicted: Caspase-10	Parus Major	Birds	Passeriformes
XP_015723627.1	Predicted: Caspase-10 isoform X1	Coturnix Japonica	Birds	Galliformes
NP_001081410.1	Caspase-10 S Homeolog	Xenopus Laevis	Amphibia	Anura
NP_001083130.1	Caspase-10 L Homeolog	Xenopus Laevis	Amphibia	Anura
XP_017527989.1	Predicted: Caspase-10	Manis Javanica	Mammalia	Pholidota
XP_019378751.1	Predicted: Caspase-10 isoform X1	Gavialis Gangeticus	Reptilia	Crocodilia
NP_001300701.1	cFLIP	Danio Rerio	Fish	Cypriniformes
XP_008118737.1	cFLIP	Anolis Carolinensis	Reptilia	Squamata
XP_033025467.1	cFLIP	Lacerta Agilis	Reptilia	Squamata
XP_025022289.1	cFLIP	Python Bivittatus	Reptilia	Squamata
XP_026709097.1	cFLIP	Athene Cunicularia	Birds	Strigiformes
XP_009479033.1	cFLIP	Pelecanus Crispus	Birds	Pelecaniformes
XP_006022153.1	cFLIP	Alligator Sinensis	Reptilia	Crocodilia
XP_019347600.1	cFLIP	Alligator Mississippiensis	Reptilia	Crocodilia
XP_009976784.1	cFLIP	Tauraco Erythrolophus	Birds	Musophagiformes
XP_009938107.1	cFLIP	Opisthocomus Hoazin	Birds	Opisthocomiformes

XP_025786690.1	cFLIP	Puma Concolor	Mammalia	Carnivora
XP_020930360.1	cFLIP	Sus Scrofa	Mammalia	Artiodactyla
XP_007190580.1	cFLIP	Balaenoptera Acutorostrata Scammoni	Mammalia	Cetacea
XP_026980099.1	cFLIP	Lagenorhynchus Obliquidens	Mammalia	Cetacea
XP_027446588.1	cFLIP	Zalophus Californianus	Mammalia	Carnivora
XP_005640588.2	cFLIP	Canis Lupus Familiaris	Mammalia	Carnivora
XP_019491241.1	cFLIP	Hipposideros Armiger	Mammalia	Chiroptera
XP_035161179.1	cFLIP	Callithrix Jacchus	Mammalia	Primates
NP_001120655.1	cFLIP	Homo Sapiens	Mammalia	Primates
NP_001125140.1	cFLIP	Pongo Abelii	Mammalia	Primates
NP_001188445.1	cFLIP	Oryzias Latipes	Fish	Beloniformes
XP_029385985.1	cFLIP	Echeneis Naucrates	Fish	Perciformes
XP_034417045.1	cFLIP	Cyclopterus Lumpus	Fish	Scorpaeniformes
XP_020506830.2	cFLIP	Labrus Bergylta	Fish	Perciformes
NP_001254595.1	cFLIP	Gasterosteus Aculeatus	Fish	Gasterosteiformes

Supplemental Figure 2

Graph S1



	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%

CHAPTER 3

Investigating the pH-Induced Destabilization Effect in Caspases

ABSTRACT

This study evaluated the impact of various mutations on the pH-dependent destabilization of caspases. Histidine and lysine residues in caspase-8 and aspartic acid residues in caspase-3 were mutated, and the stability of the resulting mutant proteins was determined by titration tests and other biophysical methods. None of the mutations significantly contributed to the pH-dependent instability, showing that it is likely due to a network impact of amino acids rather than a single essential amino acid. Trypsin cleavage experiments revealed that the mutant proteins were highly destabilized, although circular dichroism analysis revealed that their secondary structure was unaffected. The findings implies that these mutations may affect the conformational stability of this family of enzymes, moving them across ensembles, and may have an allosteric influence. Further research is needed to explore the role of these residues in electron transport in caspases which can be useful to design drugs that can interfere with these residues to treat specific pathways that rely on residue modifications that are essential to cellular function.

INTRODUCTION

Caspases are primarily known for their role in programmed cell death or apoptosis, which is an important physiological process in the development and maintenance of multicellular organisms(1). However, recent research has suggested that caspases also play important roles in non-apoptotic functions, such as inflammation, immunity, and cell differentiation(2).

Caspases also play a role in immune system regulation. Caspase-8, for example, is involved in T-cell receptor signaling and activation of the immune response(3). Caspase-8 deficiency can lead to impaired T-cell activation and increased susceptibility to infection. In addition, caspases have been shown to be involved in the maturation and differentiation of immune cells, such as B-cells and dendritic cells(3). Another non-apoptotic function of caspases is their role in cell differentiation(2). Caspase-3, for example, has been shown to be involved in the differentiation of neurons and skeletal muscle cells(4). Caspase-3 activation leads to the cleavage of specific substrates involved in cell differentiation, such as laminin and the transcription factor MyoD(4). Furthermore, caspases have also been implicated in other cellular processes such as autophagy, DNA damage repair, and metabolism(5). In summary, while caspases are primarily known for their role in apoptosis, they play a crucial role in various non-apoptotic functions, highlighting their importance in the overall functioning of living organisms(2). Examining the conformational changes are critical for understanding these less understood non apoptotic functions of caspases.

Conformational changes in allosteric hotspots can be brought about by charged residues(6). Allosteric hotspots are specific regions of a protein that are critical for transmitting signals from one part of the protein to another, often leading to a change in the protein's activity or function(7). These hotspots can be located far away from the protein's active site, and changes in their conformation can modulate the protein's activity in response to various stimuli. Charged residues, such as aspartate (D),

glutamate (E), lysine (K), and arginine (R), are known to be involved in many protein-protein interactions and can play a crucial role in the conformational changes associated with allosteric regulation(8). For example, charged residues can form salt bridges or electrostatic interactions with other charged residues or polar amino acids in the protein, stabilizing or destabilizing certain conformations(9). In some cases, charged residues located in allosteric hotspots can act as switches, triggering a conformational change in the protein(7). For instance, the binding of calcium ions to calmodulin induces a conformational change that is mediated by a network of charged residues, including D28, E31, and R39(10). These charged residues help to stabilize the conformational change induced by the calcium ions, leading to the activation of downstream signaling pathways(10).

Overall, charged residues can play a critical role in the conformational changes associated with allosteric regulation. By modulating the interactions between different regions of a protein, charged residues can act as key players in the complex network of interactions that govern protein function and activity.

In the first chapter and several other previous studies done by others in the lab, it was found that all caspases in the folding landscape show a drop in stability at pH below 6.5(11–13). This study aimed to identify the critical residue responsible for this transition, based on the hypothesis that it was likely due to a pH-sensitive amino acid and that the critical residue responsible for pH-dependent destabilization could be a pH-sensitive amino acid such as histidine, aspartic acid, or lysine. Based on this observation, in this study we proceeded to systematically mutate specific residues in caspase-8 and caspase-3, which are highly conserved in caspases involved in the extrinsic apoptotic pathway.

RESULTS

With reference to caspase-8 **(Figure 1A)**, histidine residue at position 89 (which is 304 in full length canonical sequence) and K136/138 (351/353 in full length) were individually suspected to be responsible for the pH-dependent destabilization in all caspases. In this study we mutated H89 to alanine and K136/138 to Alanine, in separate experiments and measured the effect of titration and other biophysical studies

to measure the stability of the mutant protein. With reference to Caspase-3 **(Figure 1B)**, two aspartic acid residues at positions 70 and 192 (D70 and D192) that were suspected to be involved in pH-dependent destabilization were also mutated and studies to answer our hypothesis.

In the titration experiment, at least six repetitions were done, as shown in the **supplemental figure S1**. And all the data were fitted using Igor Pro using the equation depicted in **Figure 2**. Also, the code is displayed in the appendix. The results indicated that none of the mutations significantly contributed to the decrease in pH, indicating that the network, and not a single amino acid, is responsible, and not a single mutant. **Figure 3**, fitting parameters indicate that the pKa of only the D70A mutant is marginally greater than that of the other mutants, but not significantly so. As this mutation is found on helix 1, which is the most stable element according to the MD simulations in chapter 3, this might have a significant impact on the system's pKa. The fact that only procaspase-3 has twice as many protons as the other enzymes does indicates that the others are all monomers. This also shows that the D70A and D192A mutants of caspase-3 convert the enzyme to a monomer, which may be the result of an inhibition of the electron transfer routes detailed in chapter 3.

In addition, we performed trypsin cleavage assays to demonstrate that these systems are significantly destabilized, as all mutants, as shown in **Figure 4 B,C,E,F**, are rapidly cleaved in the first 15 minutes compared to the wild type, as shown in **Figure 4 A,D**, for caspase-3 and caspase-8, respectively. Circular dichroism analysis revealed that the secondary structure of all mutants is still intact, indicating that the enzyme has not lost its structure but has lost its stability. This could indicate that the enzyme is fluctuating wildly in multiple ensembles as a result of these mutations, which may have an allosteric effect. Then, we examined the folding of these enzymes to demonstrate that the D70A mutant lacks the cooperation observed in the folding of other caspases. Our findings suggest that these mutations may change the enzyme's conformational stability, shifting it across ensembles. Since the pH studies yielded a negative result, these studies were halted due to a lack of time. We hope that others will be able to demonstrate how these mutations can affect electron transport in caspases and design

drugs that can interfere with these residues to treat specific pathways that rely on residue modifications that are essential to function.

Materials and Methods

Caspase-8 H89A, K136/138A, Caspase-3 D70A, and D192A were all created in pet 21b plasmid and donated by genescript. All mutations occurred as inclusion bodies, necessitating the use of denaturing purification. The supernatant was discarded following the French press, and the pellet was resuspended overnight in 6M urea. Tip: Add 6M urea (powder) directly to the pellet and then the buffer (100mM NaCl, 50mM Tris, pH8) to make 6M urea for uniform mixing and protein release from inclusion bodies. The solution was then spun down for 30 minutes at 15,000 rpm. The supernatant was diluted with 3M urea and permitted to bind to the resin in an affinity chromatography column. On the column itself, the protein was refolded by adding a wash buffer that lacked urea and included only the above-mentioned buffer composition. Subsequently, a gradient elution from 0 mM Imidazole to 500 mM Imidazole was carried out using the buffer described previously. The protein-containing fraction was pooled, dialyzed (using the same buffer as above to remove imidazole), and concentrated for further research. In instances where the fraction was not pure, an anion exchange column with Sepharose beads and a salt gradient from 50 to 500 mM NaCl in the same buffer was used. Since refolding on column resulted in an impure fraction of caspase-3 D70A, the entire purification procedure was carried out with urea. Thus, 6M urea was present throughout, and the protein was refolded through dialysis following affinity chromatography, yielding a pure fraction.

Titration experiment:

Using a Hamilton titrator connected to a PTI C-61 spectrofluorometer, a 15ul solution of 100mM H_2SO_4 was added 26 times to a cuvette containing a 2ml solution of the each mutant and the wild type. Every time the proteins were excited at 295nm and the emission was captured between 300nm and 400nm, the proteins were excited at 295nm. As 2uM was insufficient to capture the signal, circular dichroism titration was done separately at 4uM. In the experiment, the same processes were performed

utilizing circular dichroism and a titrator from JASCO. Using the standard curve (supplementary figure S1) established by adding the same quantity of acid to the buffer 26 times, the pH at each step was calculated and measured using a Mettler Toledo 751-4MM Conductivity Cell, small sample, 4 mm.

Conclusion

Our data indicate that the pH-induced instability observed in all caspases at pH 6.5 is not due to a single amino acid, but rather a network effect of amino acids or possibly a combination of the studied mutations. Further research may shed insight on the mechanism of this enzyme family-specific instability.

Figure 1

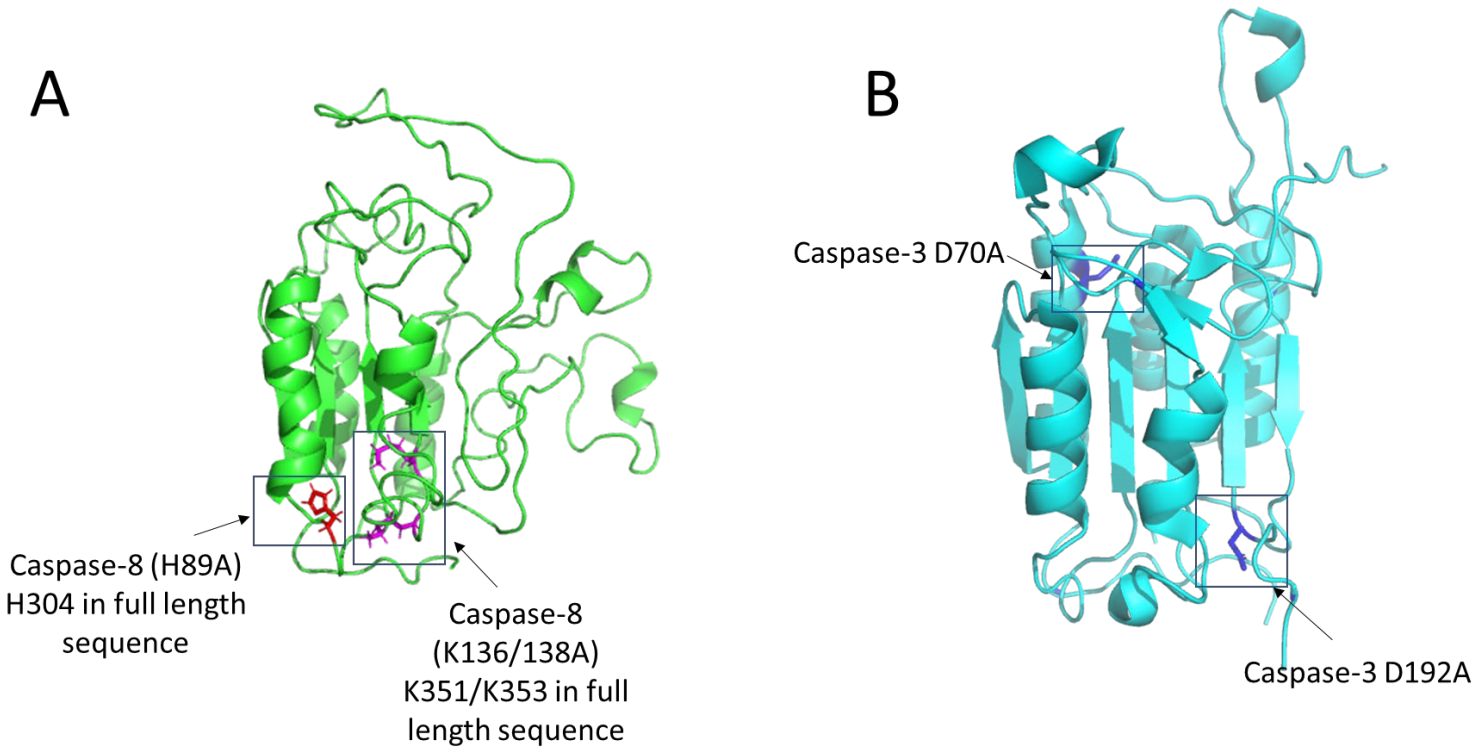
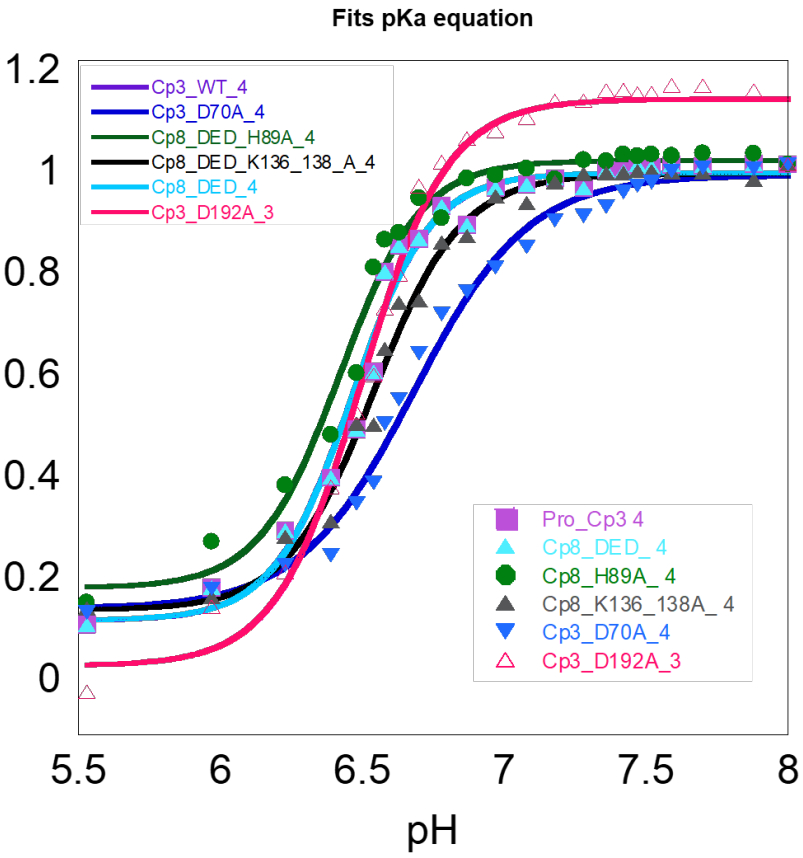


Figure 1. A. Mutants that were purified and worked with for caspase-8 H304 (referred to as H86A) and K351/K353 (referred to as K136/K138A). (B) Caspase-3 mutants D70A and D192A

Figure 2



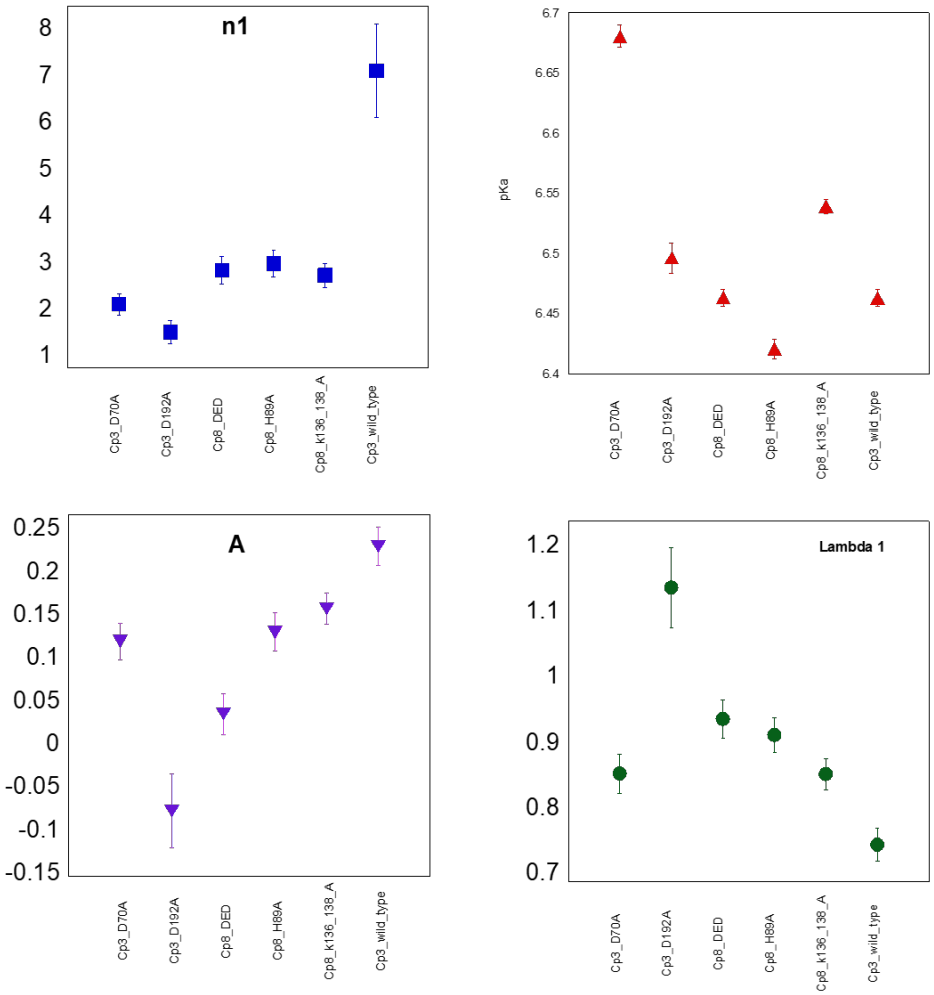
Fitting equation:

$$\langle\lambda\rangle = \frac{(A + \Delta\langle\lambda_{(1)}\rangle) \times 10^{n1(pH-pKa)}}{1+10^{n1(pH-pKa)}}$$

Figure 2. Titration curves of caspase-8, caspase-8 H89A, caspase-8 K136/138A and procaspase-3, procaspase-3 D70A, procaspase-3 D192A. Raw data for all the proteins is shown in supplemental figure S3.

Figure 3

A



B

$$\langle\lambda\rangle = \frac{(A + \Delta\langle\lambda_{(1)}\rangle) \times 10^{n1(pH-pKa)}}{1+10^{n1(pH-pKa)}}$$

Figure 3.A. Fitting parameters and their values obtained for all the caspase mutants in figure 1. **B.** Fitting equation for reference

Figure 4

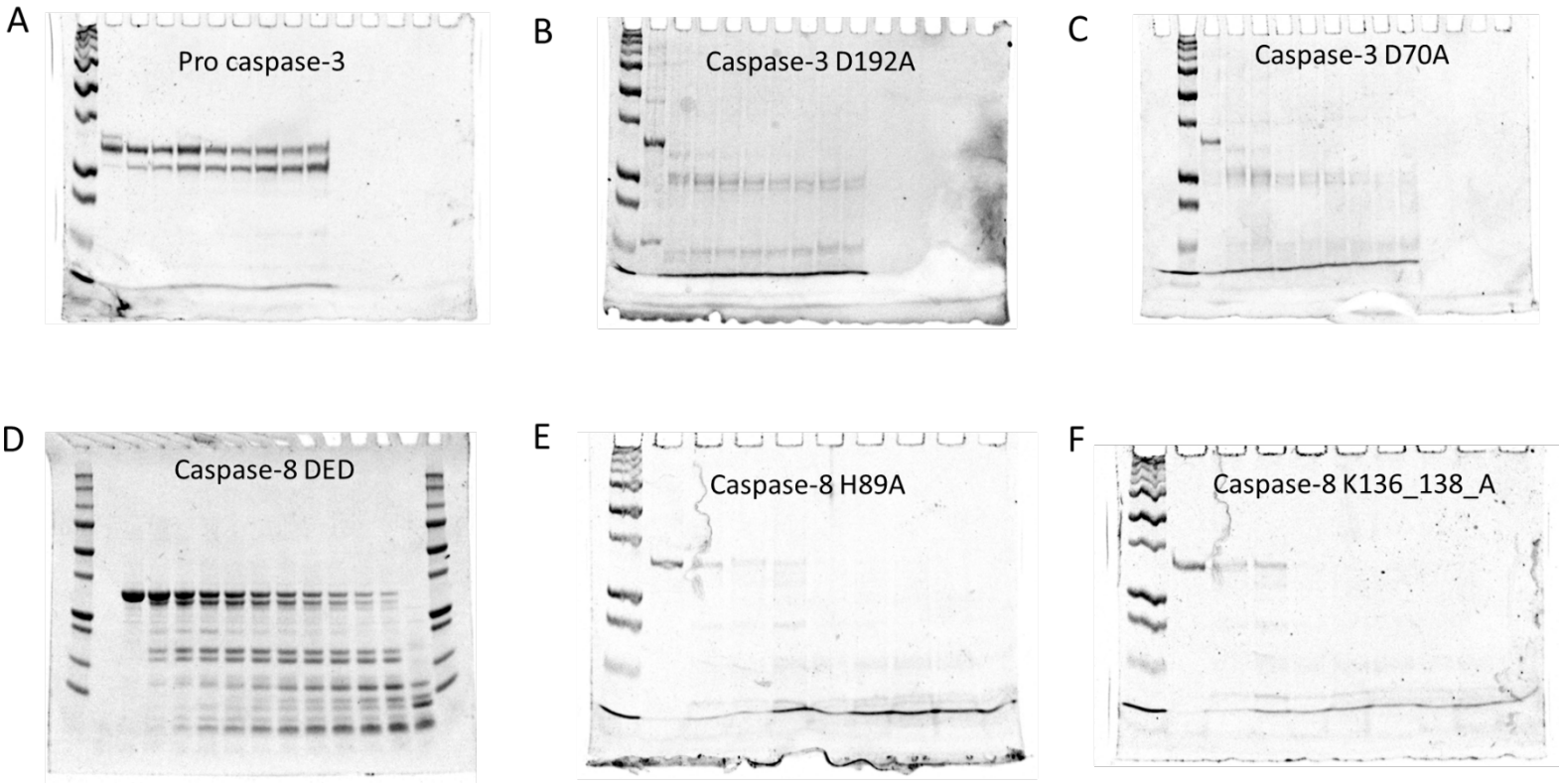


Figure 4. A. Limited trypsin proteolysis for (A) Procaspase-3, (B) Caspase-3 D192A (C) Caspase-3 D70A (D) Caspase-8 DED, (E) Caspase-8 H89A and Caspase-8 K136/138A shows that the mutants are less stable in structure.

Figure 5

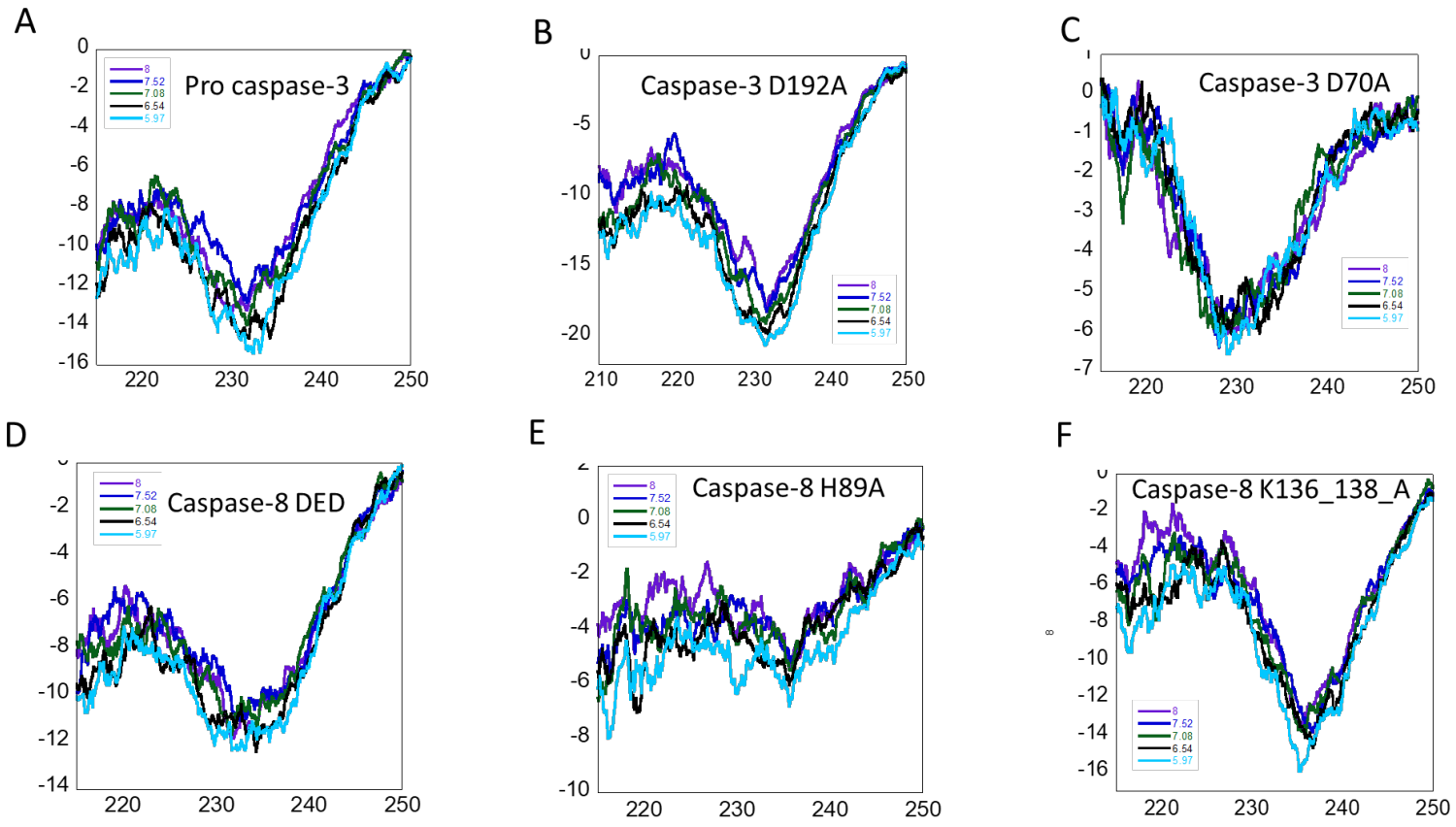


Figure 5. (A) Circular dichrosim for (A) Procaspase-3, (B) Caspase-3 D192A (C) Caspase-3 D70A (D) Caspase-8 DED, (E) Caspase-8 H89A and (F) Caspase-8 K136/138A shows that the mutants have a well defined secondary structure in spite of being destabilized

Figure 6

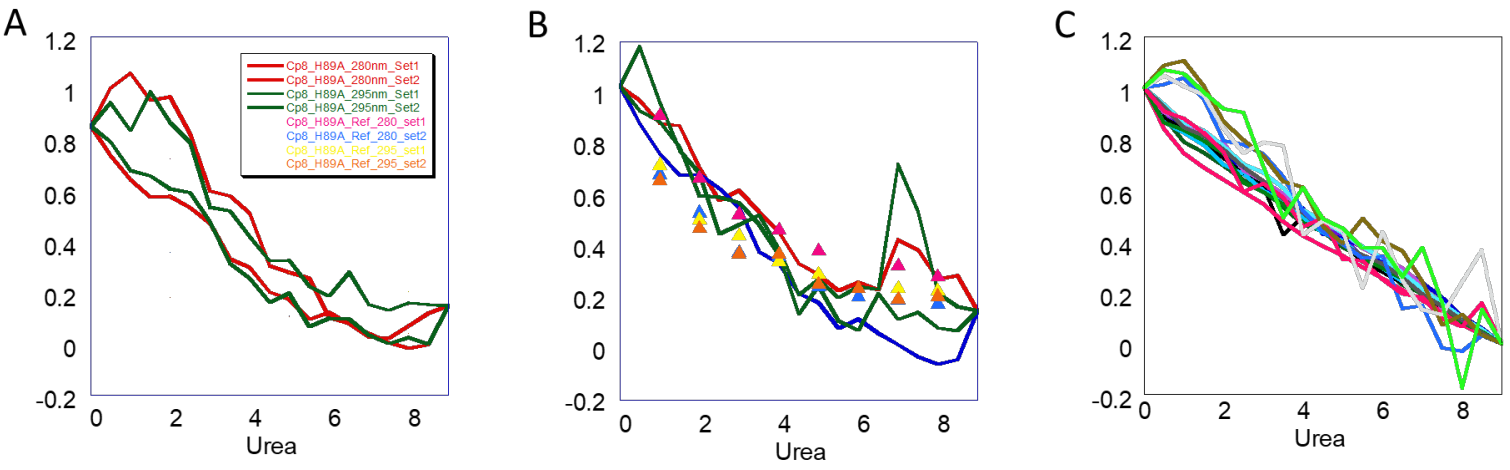
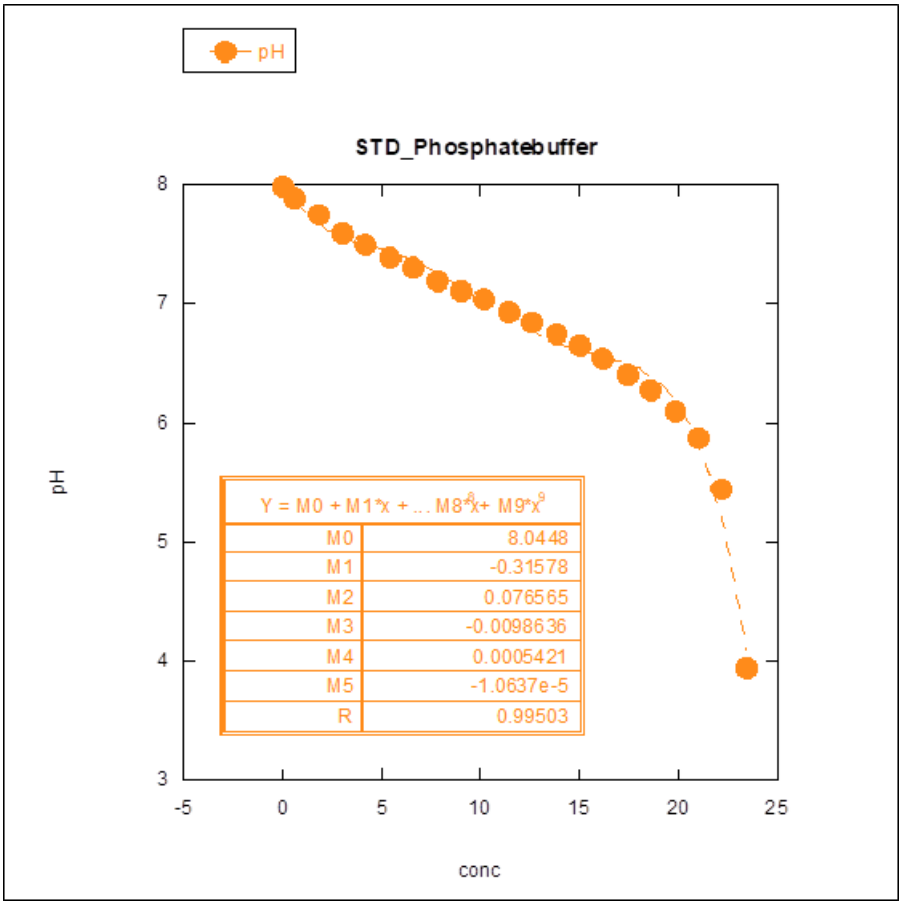


Figure 6. Folding studies with the mutants for (A) Caspase-8 H89A (B) Caspase-8 K136/138A (C) Caspase-3 D70A. Interestingly the caspase-3 D70A mutants shows that co-operativity is lost in the folding process for this process. Raw data is shown in supplemental figure 2.

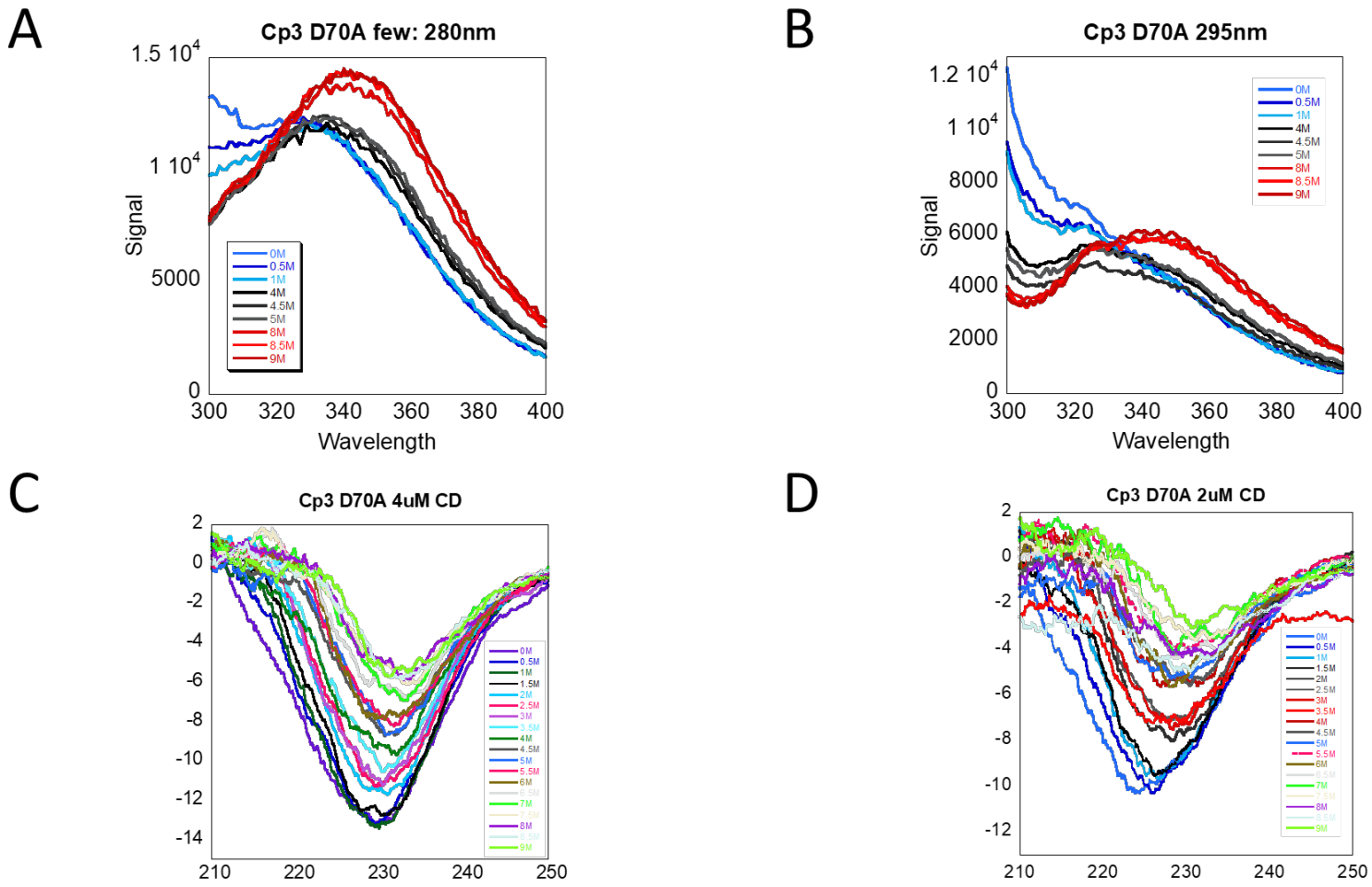
Supporting Information

Supplemental Figure 1



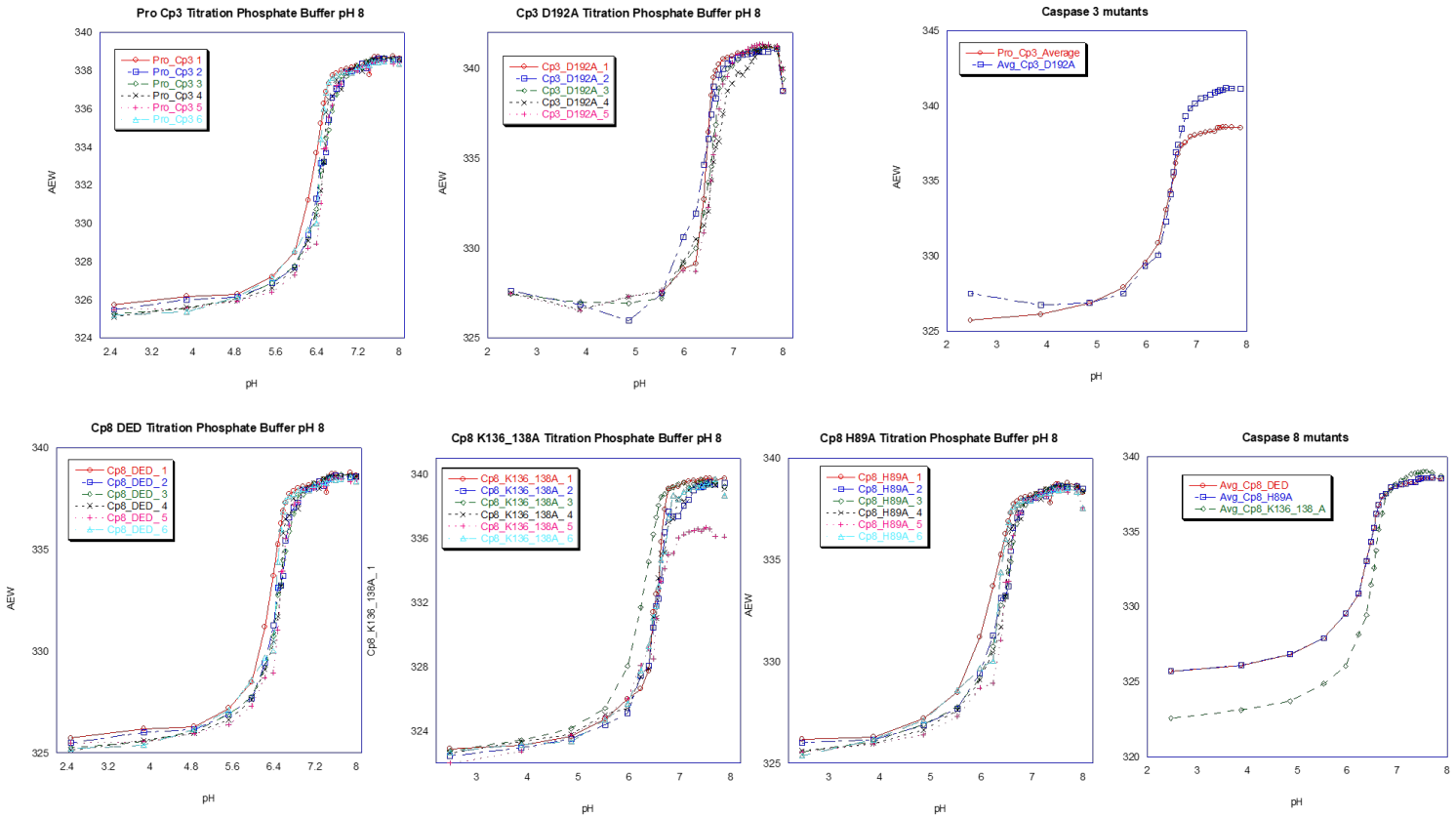
Supplemental Figure 1. Phosphate buffer standard curve generated from adding sulfuric acid

Supplemental Figure 2



Supplemental Figure 2. A,B fluorescence curves obtained from unfolding of caspase-3 D70A. C&D Circular dichroism curves obtained from the unfolding of caspase-3 D70A

Supplemental Figure 3



Supplemental Figure 3. All the titration curves obtained from the fluorescence, the data was transformed and normalized to look at the titration.

References

1. Adams, J. M., and Cory, S. (2002) Apoptosomes: Engines for caspase activation. Curr Opin Cell Biol. 14, 715–720

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

3. Salmena, L., Lemmers, B., Hakem, A., Matysiak-Zablocki, E., Murakami, K., Billie Au, P. Y., Berry, D. M., Tamblyn, L., Shehabeldin, A., Migon, E., Wakeham, A., Bouchard, D., Yeh, W. C., McGlade, J. C., Ohashi, P. S., and Hakem, R. (2003) Essential role for caspase 8 in T-cell homeostasis and T-cell-mediated immunity. Genes Dev. 17, 883

4. Fernando, P., Kelly, J. F., Balazsi, K., Slack, R. S., and Megeney, L. A. Caspase 3 activity is required for skeletal muscle differentiation. [online] www.pnas.orgcgidoi10.1073pnas.162172899 (Accessed March 8, 2023)

5. Mcilwain, D. R., Berger, T., Mak, T. W., Baehrecke, E. H., Green, D. R., Kornbluth, S., and Salvesen, G. S. Additional Perspectives on Cell Survival and Cell Death available at Harbor Laboratory Press at University of Texas at Arlington on March 8. 10.1101/cshperspect.a008656

6. Weinkam, P., Chen, Y. C., Pons, J., and Sali, A. (2013) Impact of mutations on the allosteric conformational equilibrium. J Mol Biol. 425, 647–661

7. Reynolds, K. A., McLaughlin, R. N., and Ranganathan, R. (2011) Hot spots for allosteric regulation on protein surfaces. Cell. 147, 1564–1575

8. Debler, E. W., Jain, K., Warmack, R. A., Feng, Y., Clarke, S. G., Blobel, G., and Stavropoulos, P. (2016) A glutamate/aspartate switch controls product specificity in a protein arginine methyltransferase. Proc Natl Acad Sci U S A. 113, 2068–2073

9. Zhou, H.-X., and Pang, X. (2018) Electrostatic Interactions in Protein Structure, Folding, Binding, and Condensation. 10.1021/acs.chemrev.7b00305

10. Hoffman, L., Stein, R. A., Colbran, R. J., and McHaourab, H. S. (2011) Conformational changes underlying calcium/calmodulin-dependent protein kinase II activation. EMBO Journal. 30, 1251–1262
11. Shrestha, S., and Clark, A. C. (2021) Evolution of the folding landscape of effector caspases. Journal of Biological Chemistry. 297, 1–12
12. Bose, K., and Clark, A. C. (2001) Dimeric procaspase-3 unfolds via a four-state equilibrium process. Biochemistry. 40, 14236–14242
13. Nag, M., and Clark, A. C. (2023) Conserved folding landscape of monomeric initiator caspases. Journal of Biological Chemistry. 0, 103075

CHAPTER 4:

Caspase-8 dimerization constructs cloning with Gibson assembly

Introduction and Results

High conc of kosmotropic salt Sodium citrate (1M) has been used to Dimerize initiator caspases which are physiologically monomers, however fusion proteins Like FKBP are predicted to mimic the dimerization by death domains more faithfully (1, 2). Uncleavable caspase 8 with prohibitive mutations at the interdomain cleavage site is less stable and shows lesser enzymatic activity than cleavable version (1). For homodimerization DmrB domain or Fv domain is fused to the N terminal region of Caspase 8 ΔDED (starting at Amino acid M206) by a 4 Glycine linker and a TEV protease sequence SRGSH (Fig 1.a). For Heterodimerization DmrA domain or FKBP is fused to the N terminal of caspase 8 ΔDED (starting at amino M206) by a 4 Glycine linker and TEV protease sequence, DmrC or FRB domain is fused to cFLIPL sequence with the same linker architecture as the others (Fig 1.b).

B/B homodimerizer is an organic molecule like AP20187 when added in stoichiometric concentrations to fusion proteins with DmrB it induces self-association of two DmrB which brings the molecules in proximity mimicking homodimerization (Fig 1.c). A/C heterodimerizer is like the organic molecule AP21967 which induces association of fusion proteins containing DmrA and DmrC (Fig 1.d). The organic molecules were acquired from Takara bio.

The DmrA and B constructs were cloned in house, using Fv-hCaspase 8-2A-GFP and Fv-hCaspase 8 (C/A)-2A-GFP, obtained from addgene plasmid repository and we cloned the regions of inserts into pet 21b using Gibson cloning representing DmrA and DmrB constructs with caspase-8 and linker as shown in Figure 1A (3). All the fragments were designed on SnapGene software (www.snapgene.com) using the inbuilt program. The Primer sequences are given in the supplementary information. Gibson assembly was acquired from NEB for further in vitro cloning. More information on the Gibson cloning is given below.

Gibson cloning is a molecular biology technique utilized to combine several DNA fragments into a longer, functional DNA sequence. The procedure was created in 2009 by Dr. Daniel Gibson, after whom it is called(4).

The Gibson cloning procedure consists of three steps:

Overlap extension PCR: In this stage, the DNA fragments to be assembled are amplified with overlapping ends using polymerase chain reaction (PCR). Afterwards, the PCR products are purified. The purified PCR products are combined with a mixture of enzymes including T5 exonuclease, Phusion polymerase, Taq DNA ligase, and a buffer for assembly. Together, these enzymes digest the ends of the PCR products, revealing their complimentary sequences, and then ligate them to form a single, bigger DNA molecule.The DNA is inserted into a host organism, such as bacteria or yeast, where it is duplicated and expressed.Traditional cloning techniques, such as restriction enzyme digestion and ligation, are surpassed by the Gibson cloning technique, which offers various advantages. It can be used to simultaneously assemble several DNA segments without the need for specific restriction enzyme sites, and it can also be used to incorporate site-directed mutations or alterations into the formed DNA. The cloned constructs were transformed into BL21 lemo DE3, competent cells which were acquired form NEB and we tried to purify the hybrid protein in-vitro. However, the constructs failed to express and we speculate that this could be due to codon optimization problems as the original constructs were mammalian plasmids.

Conclusion

Since we were unable to express the constructs, we hypothesize that codon optimization was the primary issue, but frame shift effect or other faults may have been introduced as a result of the polymerases utilized. However, as no one has attempted to purify these systems in-vitro to our knowledge, it is possible that they are also hazardous to the Ecoli system. Thus, we advise future studies to improve the codons and experiment with other expression methods to Ecoli BL21 lemo DE3 systems.

Figure 1.

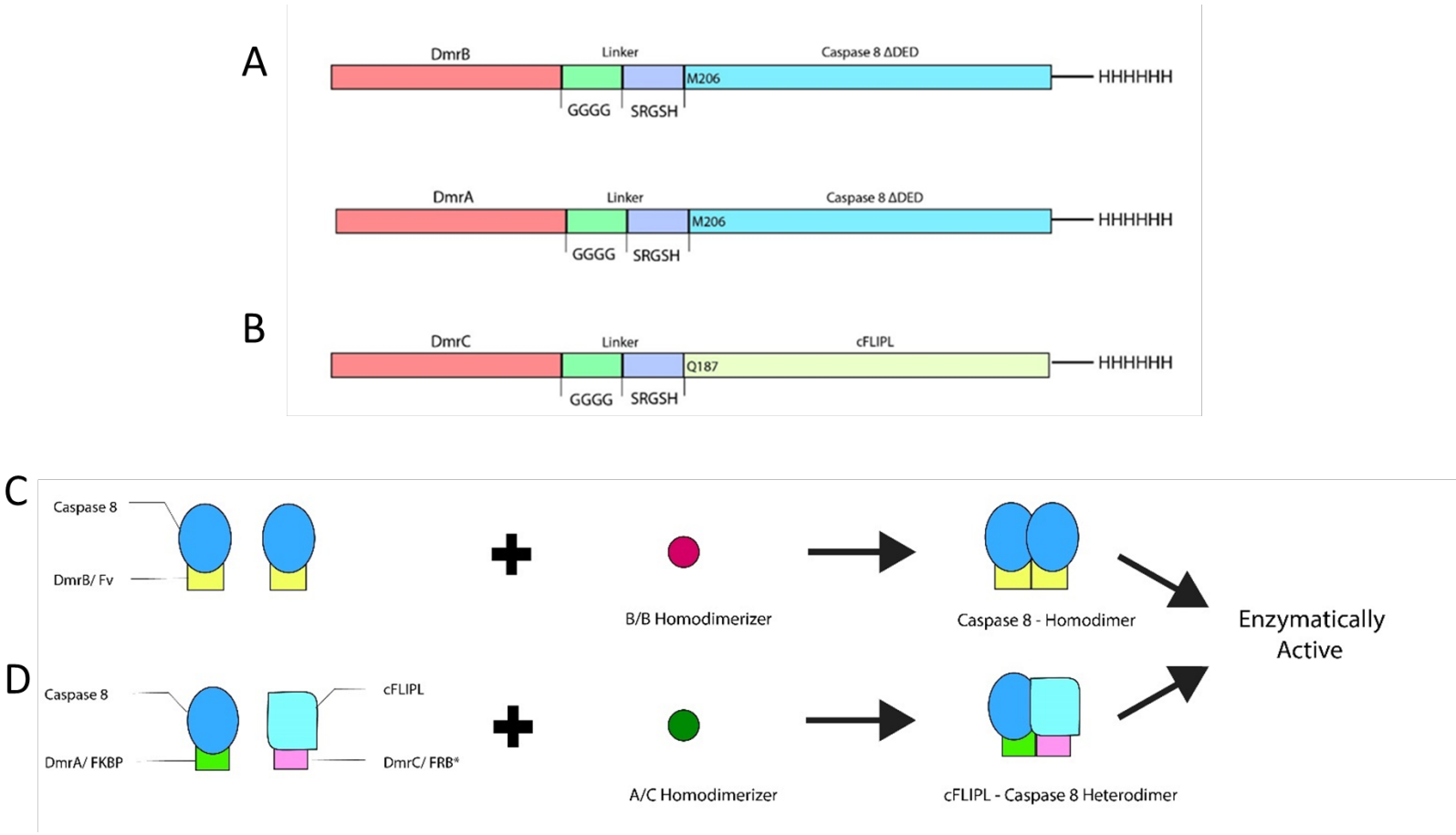


Figure 1. 2D architecture of Fusion Proteins (**A**) Homodimerization system: DmrB or Fv domain is fused to caspase 8 ΔDED with a 4 glycine linker and a tobacco etch virus protease sequence SRGSH. (**B**) Heterodimerization system: DmrA or FKBP domain is fused to caspase 8 ΔDED with a 4 glycine linker and a tobacco etch virus protease sequence SRGSH. DmrC or FRB domain is fused to cFLIPL with a 4 glycine linker and a tobacco etch virus protease sequence SRGSH Fusion Proteins dimerization mechanism. (**C**) Caspase 8 ΔDED fusion protein with DmrB forming a homodimer upon addition of a tightly binding organic molecule B/B Homodimer (**D**) Caspase 8 ΔDED -DmrA and cFLIPL-DmrC fusion proteins forming a Heterodimer upon addition of tightly binding organic molecule A/C Heterodimer.

Primers for Gibson Cloning:

A_Assembled_Caspase8_DED_2A_DmrB_Pet21b

A_Fragment.FOR
TGGTGGTGGTGGTGCTCGAGATCAGAAGGGAAGACAAGTTTTTTTCTTAGTG

A_Vector_OfavCasp7.REV
AACTTGTCTTCCCTTCTGATCTCGAGCACCACCACC

A_Fragment.REV AAGAAGGAGATATACATATGGGAGTGCAGGTGGAGACTATCT

A_Vector_OfavCasp7.FOR
ATAGTCTCCACCTGCACTCCCATATGTATATCTCCTTCTTAAAGTTAAACAAAATTATTTCTAGAGG

B_Assembled_Caspase8_DED_CA_DmrB_Pet21b

B_Fragment.FOR
TGGTGGTGGTGGTGCTCGAGATCAGAAGGGAAGACAAGTTTTTTTCTTAGTG

B_Vector_OfavCasp7.REV
AACTTGTCTTCCCTTCTGATCTCGAGCACCACCACC

B_Fragment.REV AAGAAGGAGATATACATATGGGAGTGCAGGTGGAGACTATCT

B_Vector_OfavCasp7.FOR
ATAGTCTCCACCTGCACTCCCATATGTATATCTCCTTCTTAAAGTTAAACAAAATTATTTCTAGAGG

C_Assembled_cFLIP_DmrC

C_Fragment 1.REV
TGGTGGTGGTGGTGCTCGAGTGTGTAGGAGAGGATAAGTTTCTTTCTCAGAGT

C_Vector_OfavCasp7.REV
AACTTATCCTCTCCTACACACTCGAGCACCACCACC

C_Fragment 1.FOR
GAATCTCAAAGGGCGGCGGACAAGCAGCAATCCAAAAGAGTCTCA

C_Fragment 2.REV CTCTTTTGGATTGCTGCTTGTCCGCCGCCCTTTGAGATTCG

C_Fragment 2.FOR AAGAAGGAGATATACATATGATCCTCTGGCATGAGATGTGG

C_Vector_OfavCasp7.FOR
CACATCTCATGCCAGAGGATCATATGTATATCTCCTTCTTAAAGTTAAACAAA
ATTATTTCTAGAGGG

D_Assembled_cp8_DED_CA_DmrA_with_linker

D_Fragment 1.FOR
TGGTGGTGGTGGTGCTCGAGATCAGAAGGGAAGACAAGTTTTTTTCTTAGT
G

D_Vector_Ofav3c.REV AACTTGTCTTCCCTTCTGATCTCGAGCACCACCACCAC

D_Fragment 1.REV AGCTTCTAAAACTGGAAGGAAGTGGAGGTGGAGGTTCTAGAG

D_Fragment 2.REV
CTAGAACCTCCACCTCCACTTCCTTCCAGTTTTAGAAGCTCCACA

References

1. Mocarski, E. S., Upton, J. W., and Kaiser, W. J. (2012) Viral infection and the evolution of caspase 8-regulated apoptotic and necrotic death pathways. 10.1038/nri3131

2. Reddy Chichili, V. P., Kumar, V., and Sivaraman, J. (2013) Linkers in the structural biology of protein-protein interactions. Protein Science. 22, 153–167

3. Oberst, A., Pop, C., Tremblay, A. G., Blais, V., Denault, J. B., Salvesen, G. S., and Green, D. R. (2010) Inducible dimerization and inducible cleavage reveal a requirement for both processes in caspase-8 activation. J Biol Chem. 285, 16632–16642

4. Gibson, D. G., Young, L., Chuang, R. Y., Venter, J. C., Hutchison, C. A., and Smith, H. O. (2009) Enzymatic assembly of DNA molecules up to several hundred kilobases. Nature Methods 2009 6:5. 6, 343–345

CHAPTER 5

Caspase-8 and cFLIP Proteins: probing the mechanism of homodimerization and heterodimerization in caspases using tandem systems.

ABSRACT

This study examines the dimerization of caspase-8 and cFLIP, as well as the factors that govern it. The tryptophan residues in caspase-8 and cFLIP are employed as probes to study conformational changes and assess if the molecules can adopt a dimeric structure. The results demonstrate that attaching the proteins with a linker and bringing them close together is insufficient to facilitate dimerization via folding of the otherwise less stable and disordered small subunit. Experiments with potassium iodide and acrylamide reveal that the two quenchers interact with the protein in distinct ways, with acrylamide quenching being dictated by protein accessibility and potassium iodide quenching being based on protein structure and environment. The result implies that additional components play a role in the dimerization process, and additional studies will be undertaken to determine stability and test additional possibilities.

INTRODUCTION

When the death receptor is activated, FADD (Fas-associated protein with the death domain) is recruited to form the DISC complex (death-inducing signaling complex)(1). The DISC then attracts and activates caspase 8, signifying apoptosis in the subsequent steps (1). Nevertheless, there is a competing mechanism in which cFLIP (cellular FLICE-inhibitory protein) attaches to the DISC rather than caspase 8, preventing apoptotic signaling(2). This is because cFLIP is a pseudoenzyme that shares the caspase hemoglobinase fold with caspase-8 and can bind to the same site on FADD (2). The cooperative and hierarchical binding of Caspase 8 and cFLIP at the DISC has been demonstrated(3). Caspase 8 utilizes its DED1 pocket to attach to the FL motif of the FADD molecule. Whereas cFLIP links to caspase 8 DED2 via its DED2 domain's pocket. cFLIP binds to heterodimeric caspase 8 more preferentially than homodimeric caspase 8 Multiple caspase 8 and cFLIP produce death effector filaments, as demonstrated by cellular research(3). Caspase 8 is recruited to the DISC via its interaction with FADD, and cFLIP can subsequently bind to the same site on FADD to form a heterodimeric complex with caspase 8. Autoprocessing of this complex leads to the production of active caspase 8 and cFLIP-CASP8 heterodimers, both of which are capable of activating downstream apoptotic signaling pathways (3). In both cases the binding of DED domains appears to change the conformation of the protease domain which can then mature to the active state by induced conformational changes.

Until now, proximity-induced dimerization has been proposed as the mechanism for initiator caspase dimerization. While the Induced Proximity model summarizes many facts concisely, it does not explain the precise mechanism by which initiator caspases dimerize(4). In chapter 2, we demonstrate that the stability of the small subunit is a key factor in dimerization, and in chapter 3, we demonstrate that the enzyme must pass many checkpoints before assuming its dimeric conformation. Furthermore, we hypothesize that amino acids in the electron transport channel are critical for communication and so may be involved in the dimerization process. Dimerization could be the result of conformational changes induced by bringing two protomers together, as demonstrated by the proximity-induced dimerization for which we demonstrate in chapter 3 that the rear networks must be reorganized, a process that does not involve a barrier in effector caspases (due to unfolding of beta sheet 6).

Several studies, as described in chapter 5, have exploited inducible dimerization systems to dimerize and activate caspase-8. Changes in an enzyme's active site can alter its conformation, which in turn can impact the conformation of the allosteric site and the binding of allosteric effectors, according to studies(5). Consequently, the inducible dimerization systems may be forcing this two-way communication by modifying the active site loops and other regions near the dimer interface to interact, thereby causing the allosteric pocket to change conformation and influencing the overall conformation to be driven to the active state(6).Hence with the view of the above dimerization strategies employed various researchers in caspases could be due to bringing them into close proximity and hence allowing the otherwise unstable small

subunit to fold and hence proceed to the dimeric state by reorganizing the network of amino acid interaction networks.

Caspase-8 has also been investigated in the presence of high concentrations of sodium citrate (kosmotrope), which has been shown to promote the protein's tendency to dimerize. Kosmotropes are ions or molecules that can stabilize protein structures and promote protein-protein interactions by reducing the solubility of nonpolar groups in water (7). Sodium citrate is a kosmotrope and has been shown to increase the stability of some proteins and protein complexes. This may occur due to the effect of the kosmotrope on the stability of the protein structure or through direct interactions with the protein surface. It is thought that the kosmotrope-induced stabilization of the protein structure may reduce the activation energy required for dimerization, allowing the protein to more easily form the dimer interface necessary for activation(8).

Certain cysteine residues in several proteins are susceptible to reversible oxidation and reduction, resulting in conformational and functional alterations. These cysteine residues are capable of forming disulfide bonds, which can stabilize protein structure and facilitate protein-protein interactions(9). In certain instances, the redox status of these cysteine residues can be altered by electron transport routes. In the mitochondrial electron transport chain, for example, electrons are moved from NADH and FADH2 to oxygen by a series of redox processes, thereby generating a proton gradient across the inner mitochondrial membrane(10). This proton gradient can be utilized to drive ATP synthesis, but it can also be used to control the redox status of cysteine residues in proteins such as transcription factors (11). This is demonstrated by the transcription factor NF-κB, which can be controlled by redox signals(12). The

inhibitor protein IκB typically sequesters NF-κB in the cytoplasm (12). When IκB is phosphorylated and degraded, however, NF-κB is liberated and translocated to the nucleus, where it promotes the transcription of genes implicated in inflammation and immunity (13). The redox status of NF-κB cysteine residues can influence its capacity to bind with IκB and undergo nuclear translocation (13). Specifically, formation of disulfide bonds between cysteine residues in the DNA-binding domain of NF-κB can facilitate dimerization and nuclear translocation (11).

Throughout the HIV-1 replication cycle, the HIV-1 integrase protein is an important enzyme that enables the integration of viral DNA into the host cell genome(14). Three domains make up the protein: the N-terminal domain (NTD), the catalytic core domain (CCD), and the C-terminal domain (CTD) (CTD). It has been demonstrated that the NTD of HIV-1 integrase is involved in protein dimerization and the development of the intasome complex, which is responsible for integrating viral DNA into the host cell genome(14). Specifically, a conserved tryptophan residue at position 235 in the NTD has been linked to protein dimerization and intasome assembly(15). The tryptophan residue at position 235 is implicated in ligand binding and conformational changes that enhance protein dimerization, according to studies(16). The tryptophan residue specifically produces a hydrophobic pocket that can bind to small molecule inhibitors and stabilize the dimer interface. Furthermore, mutation of the tryptophan residue at position 235 to alanine reduces protein dimerization and intasome assembly, highlighting the significance of this residue in facilitating protein-protein interactions(16). Hence, other redox amino acids may also be implicated in caspase dimerization. In addition, we explain in chapter 3 how M403 in caspase-8 is trapped in a hydrophobic

pocket and how the inversion of F356 can assist in modifying this interaction and therefore altering the conformation. Thus, differing charges could influence the conformation, as found in initiator caspase -8 and cFLIP, in which the enzyme is less stable at lower and higher pH values (pKa 5.6 and pKa 8.5). The first transition is observed in all caspases, whereas the second transition is exclusive to initiator caspases. Hence, the cysteine residue may have a role in these research. While the common transition at pKa 5.6 may also be involved, as dimeric caspase-3 has been demonstrated to shift to a monomeric shape at a lower pH, pKa 5.6 may also play a role(17).

To test the above hypothesis that dimerization could be caused by thermodynamics leading to the folding of the small subunit when they are brought close together or by critical redox residues, we employed a tandem system with two caspase 8 DED in tandem or caspase 8 -cFLIPL in tandem (figure 1) with linkers between the proteins. By creating a covalent bond between proteins, linkers have been utilized to trap weak or temporary interactions (18). Linker length and composition are crucial for obtaining oligopeptides with conformational stability(18). Glycine repeats are typically utilized as linkers, but an excess of glycine might confer flexibility and so interfere with conformational stability energetics. The GSGSG linker is superior since it is structurally expanded without kinks (x-ray crystallography data). Serine in the linker facilitates interactions with polar solvents and nitrogen in the main chain, hence promoting conformational and energy stability(19). The linker GGSGSGSGSGG was chosen, and the two terminal glycines provide flexibility. Based on the crystal structure of the caspase 8 DED – cFLIPL heterodimer (3H11), the distance between the C terminal end

of caspase 8 and the N terminal end of cFLIPL48 is 22 angstroms (20). According to the crystal structure of the caspase 8 Homodimer, the distance between the N terminus of one caspase 8 protein and the C terminal of the other protein is 19 Angstrom (21). Taking into account 3 angstrom for each residue, the length of the linker is 27 angstrom. Figure 1A depicts how the 2D architecture of the 8 DED tandem was built. Caspase 8 DED C terminal was coupled to the N terminal of cFLIPL, as depicted as 2D structure in Figure 1B.

RESULTS

The caspase hemoglobinase domain of caspase-8 and cFLIP are depicted in Figure 2A and B, respectively. The figure illustrates the tryptophans in caspase-8 and cFLIP. The tryptophans in cFLIP are optimally positioned in the small subunit and can be used as probes to determine whether this region undergoes dimerization in the tandem system, whereas the tryptophan in caspase-8 is buried in a hydrophobic pocket and must move to a hydrophilic environment in the dimeric state. Following the fluorescence signal of the tryptophans is therefore an excellent approach for tracking the conformation and determining if the molecule can transition to the dimeric conformation.

The enzyme was purified under denaturing conditions and refolded on the affinity chromatography column as described in chapter 5. Two copies of the fluorescence signal were examined. And the results (figure 3 A,B) demonstrate that the signal for the tandem (Cp8 cFLIP in figure3) is additive (Cp8 cFLIP add_ in figure3), i.e., when the

signal of caspase-8 and cFLIP are acquired separately and then added, the result is identical to that obtained for the tandem. This indicates that, despite being allowed to fold in a crowded test tube with both proteins attached, the proteins fold to their monomeric state without aggregating or proceeding to the dimeric state. Hence, these results suggest that other factors play a role in the dimerization process in these systems. When casapse-8 and cFLIP are refolded together in a test tube (figure 4), they follow the same pattern as the dimer, demonstrating that the linker has no influence on the fold and that positioning the enzymes in tandem has no effect on the fold either. Our results suggest that tethering two proteins together with a linker and bringing them close together is insufficient to promote dimerization via folding of the otherwise less stable and disordered small subunit.

At pKa 6.5 and pKa 8.5, when the enzyme is less stable, additional signals will be collected. At the above-mentioned pH, the network interaction may be weakened because, as demonstrated in Chapter 4, this is not due to the influence of a titratable conserved amino acid, but rather a network-wide effect. Nevertheless, mutations of cysteines at pKa 8.5 have not been conducted to determine the pKa at which the conformational change occurs. Consequently, the current results suggest that the disordered small subunit does not allow homo or hetero dimerization to occur; therefore, the other hypothesis, that dimierization is a result of affecting the redox amino acids, will be tested with similar studies to those described above at pH 6.5 and pH 8.5, and further experiments will be conducted to determine stability if a positive result is obtained.

Potassium iodide and acrylamide quenching is a typical approach for studying the exposure of tryptophan residues in proteins to determine conformations. The procedure requires adding a quencher, such as potassium iodide or acrylamide, to the protein solution in order to quench the fluorescence of the tryptophan residues. The degree of quenching is proportional to the accessibility of tryptophan residues to the quencher, which is dependent on the protein's conformation and environment.

Caspase-8 DED quenching experiments in potassium iodide (figure 5A) and acrylamide (figure 5B) demonstrate that at pH 8.5, the titration curves for potassium iodide exhibit a unique transition that is distinct from the unfolded signal quenching. Yet, with acrylamide, the results are the same regardless of pH. Noting that the quenching experiments were not conducted at pH 6.5, where the lower transition occurs, the experiment will be performed if necessary if the second hypothesis is supported by a positive result.

If the quenching signal exhibits the same impact at all pH values when acrylamide is used, but a distinct signal when potassium iodide is used, this demonstrates that the two quenchers interact with the protein in different ways. Acrylamide quenching largely quenches the fluorescence of tryptophan residues via a collisional mechanism in which the quencher collides with the excited state of the tryptophan residue and loses the energy as heat (22). Hence, the degree of quenching is essentially determined by the accessibility of the tryptophan residues to the quencher, as opposed to the protein's shape or environment. In contrast, potassium iodide quenches tryptophan fluorescence via a more complex mechanism involving the production of a tryptophan-potassium iodide complex, which can subsequently interact

with additional tryptophan residues and result in a non-linear quenching behavior. The development of this complex is likely reliant on the protein's shape and environment, as well as the potassium iodide concentration. If the quenching signal has a similar effect at all pH values with acrylamide, this demonstrates that the tryptophan residues are accessible to the quencher at all pH values. However, the various signals observed when potassium iodide was used to quench the signal at different pH values imply that the conformation and environment of the protein, as well as the potassium iodide concentration, have a role in the quenching behavior. The increased unfolded signal quenching (Figure 5) compared to folded signal is owing to the fact that the tryptophan residues in the unfolded form are more exposed and accessible to the quencher. Further research, such as additional tests or structural analysis, may be required to completely comprehend the mechanism underlying the reported results.

CONCLUSION

The aim of this study was to examine the dimerization process of caspase-8 and cFLIP, two important apoptosis-related proteins, by analyzing conformational changes using fluorescence and quenching experiments. The results indicate that linking the two proteins with a linker and bringing them close together is insufficient to promote dimerization through folding of the small subunit. In addition, the study discovered that the quenching behavior of tryptophan residues in the protein was dependent on the conformation and environment of the protein, as well as the quencher employed. The findings suggest that other factors play a role in the dimerization process of these proteins, and additional research is necessary to determine the dimerization mechanism. This study provides important insights into the conformational changes and

dimerization of caspase-8 and cFLIP, which could have implications for understanding their role in apoptosis and possibly developing new therapeutics for diseases related to apoptosis.

Figure 1

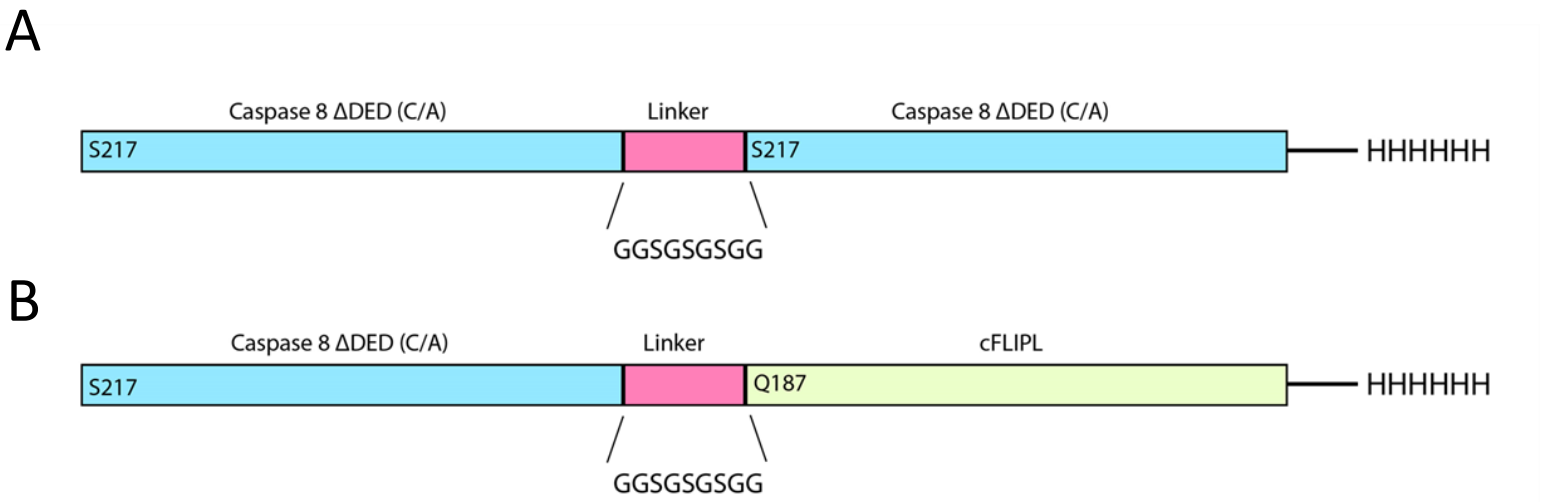


Figure 1: a schematic of the tandem constructions utilized in this work. (A) Caspase-8 (in cyan) – caspase-8 (cyan) tandem construct depicted with a linker (magenta), the starting amino acid, and his tag. (B) Caspase-8 (cyan) and cFLIP (green) in tandem with a linker (magenta), the starting amino acid, and his tag.

Figure 2

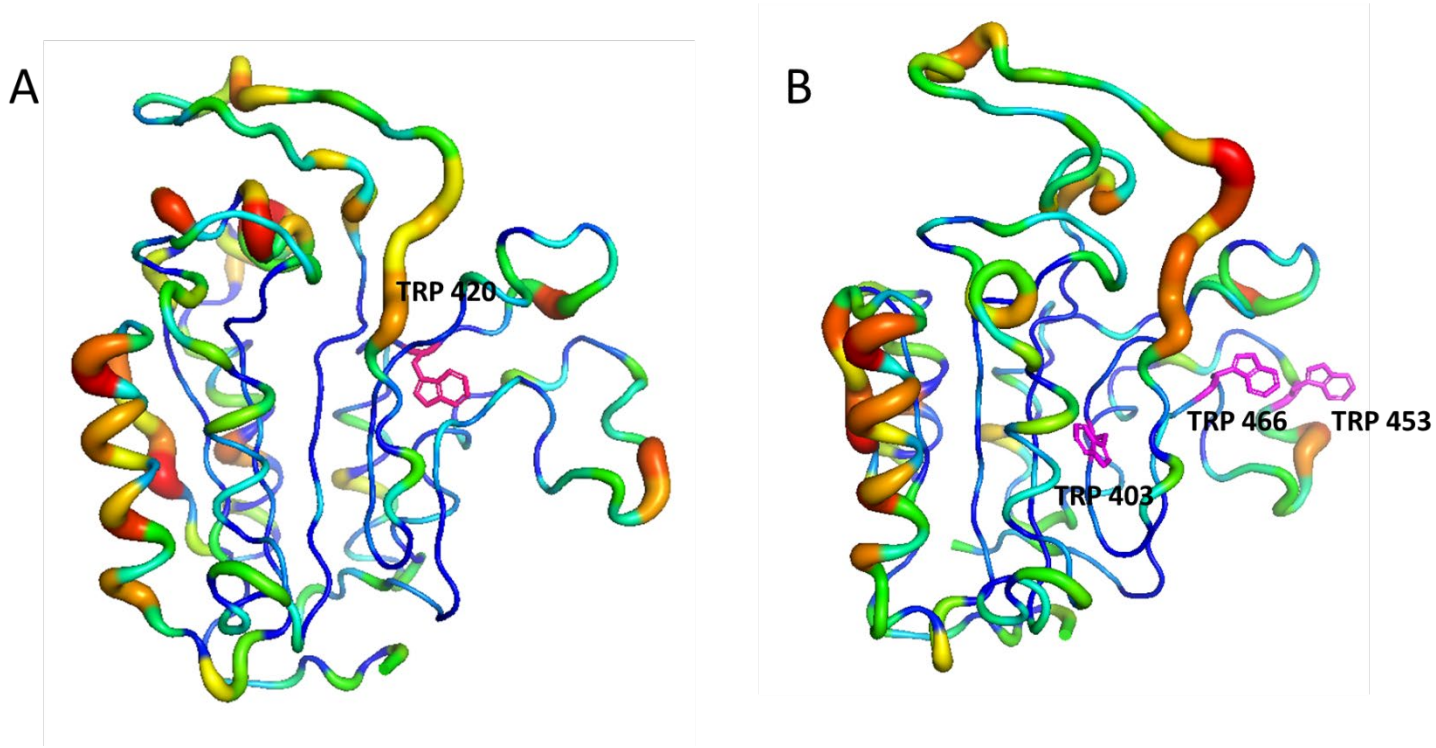


Figure 2: B-factor conservation overlaid onto the structures of (A) caspase-8 (2k7z) and (B) cFLIP, which was modelled based on the 2k7z structure, with tryptophan residues shown in magenta depicting that the intrinsic probes are excellent for probing transition to the dimeric conformation.

Figure 3

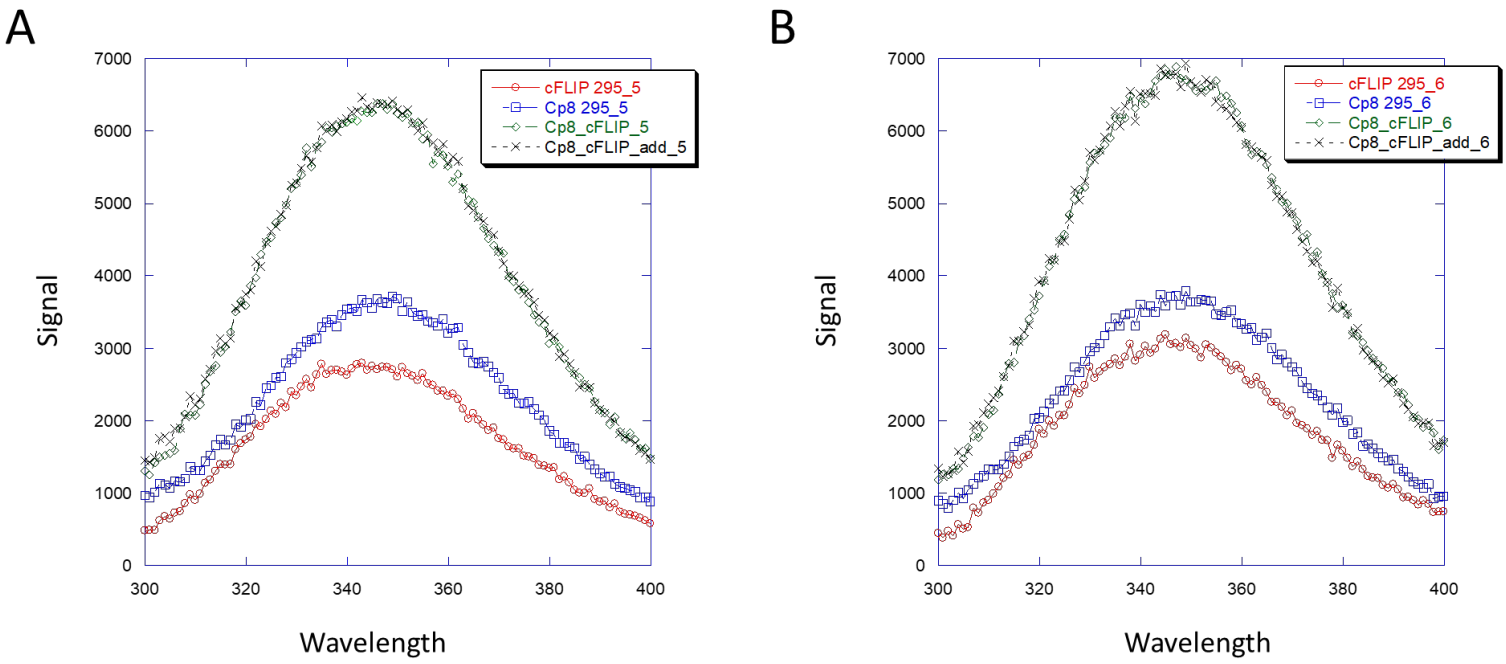


Figure 3: The native fluorescence emission signal of caspase-8 (blue) and cFLIP (red), the additive signal of the two (black), and the tandem caspase-8 – cFLIP signal (green). A and B are experimental duplicates.

Figure 4

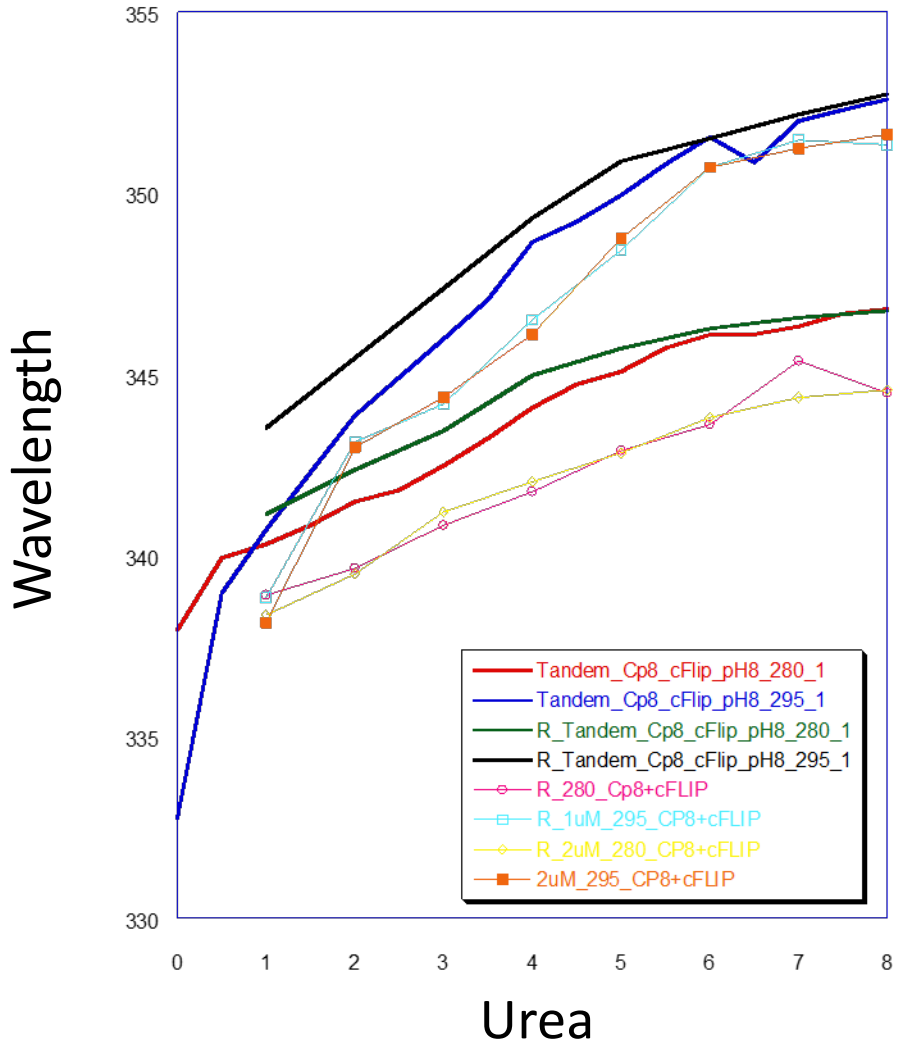


Figure 4: The illustration depicts tandems unfolding and refolding as solid lines. While caspase-8 and cFLIP were refolded together in-vitro (squares and circles in thinner lines), the signal indicates that the linker had no influence on the folding.

Figure 5

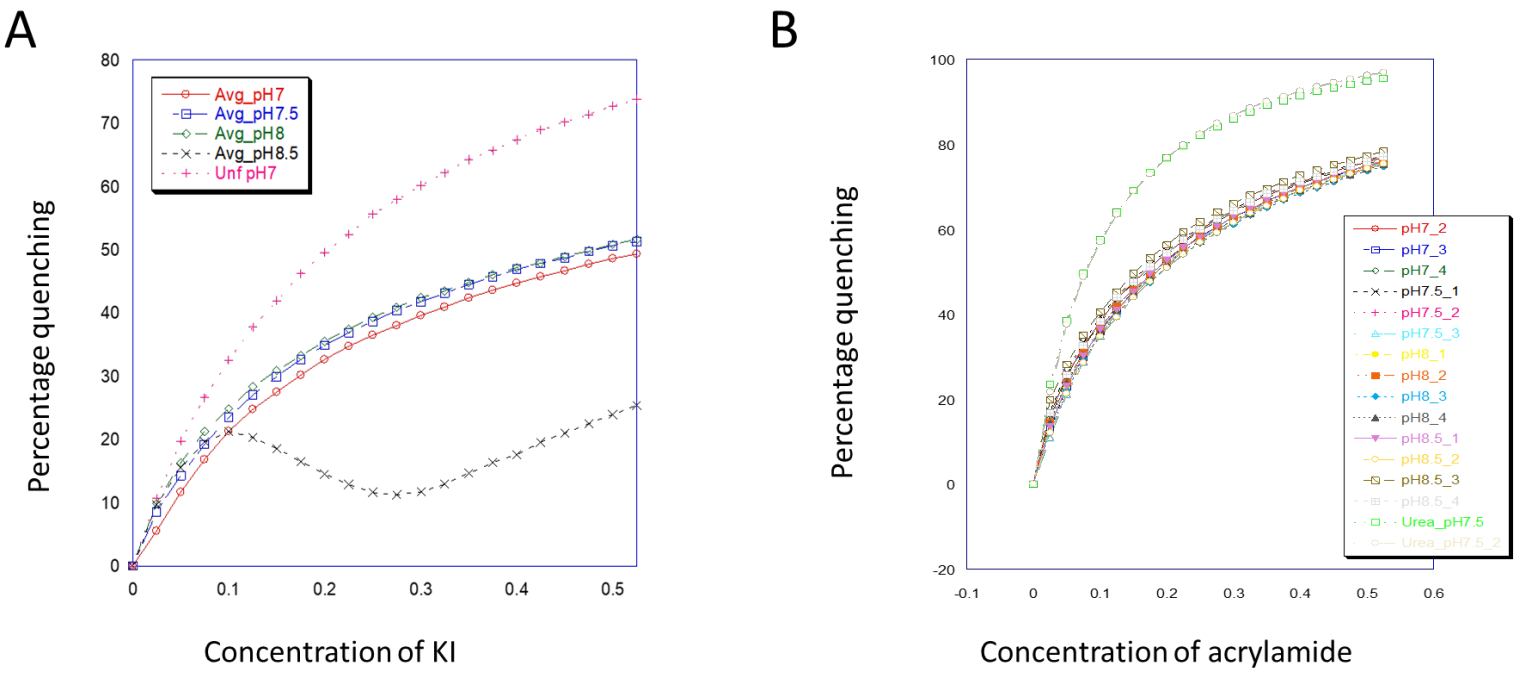


Figure 5: Caspase-8 quenching by potassium iodide and acrylamide. (A) Potassium iodide quenching of caspase-8 suggests that the unfolded signal is quenched more, indicating that the tryptophans are more exposed to solvent, and quenching at pH 8.5 indicates that the signal is distinct from those observed at other pH values. (B) Acrylamide quenching exhibits a similar trend to that of potassium iodide quenching, with increased quenching of unfolded caspase-8; however, unlike potassium iodide, which exhibits a difference at pH 8.5, acrylamide quenching is similar at all pH.

References

1. Davis, A. R., Lotocki, G., Marcillo, A. E., Dietrich, W. D., and Keane, R. W. (2007) FasL, Fas, and Death-Inducing Signaling Complex (DISC) Proteins are Recruited to Membrane Rafts after Spinal Cord Injury. *https://home.liebertpub.com/neu*. **24**, 823–834

2. Tsuchiya, Y., Nakabayashi, O., Nakano, H., and Lemarié, A. (2015) FLIP the Switch: Regulation of Apoptosis and Necroptosis by cFLIP. 10.3390/ijms161226232

3. Hughes, M. A., Powley, I. R., Jukes-Jones, R., Horn, S., Feoktistova, M., Fairall, L., Schwabe, J. W. R., Leverkus, M., Cain, K., and MacFarlane, M. (2016) Co-operative and Hierarchical Binding of c-FLIP and Caspase-8: A Unified Model Defines How c-FLIP Isoforms Differentially Control Cell Fate. *Mol Cell*. **61**, 834

4. Shi, Y. (2004) Caspase Activation: Revisiting the Induced Proximity Model. *Cell*. **117**, 855–858

5. Guo, J., and Zhou, H.-X. (2016) Protein Allostery and Conformational Dynamics HHS Public Access. *Chem Rev*. **116**, 6503–6515

6. Oberst, A., Pop, C., Tremblay, A. G., Blais, V., Denault, J. B., Salvesen, G. S., and Green, D. R. (2010) Inducible Dimerization and Inducible Cleavage Reveal a Requirement for Both Processes in Caspase-8 Activation. *J Biol Chem*. **285**, 16632

7. Boatright, K. M., Deis, C., Denault, J.-B., Sutherlin, D. P., and Salvesen, G. S. (2004) Activation of caspases-8 and-10 by FLIP L. *Biochem. J*. **382**, 651–657

8. Boatright, K. M., Renatus, M., Scott, F. L., Sperandio, S., Shin, H., Pedersen, I. M., Ricci, J. E., Edris, W. A., Sutherlin, D. P., Green, D. R., and Salvesen, G. S. (2003) A unified model for apical caspase activation. *Mol Cell*. **11**, 529–541

9. Fass, D., and Thorpe, C. (2018) Chemistry and Enzymology of Disulfide Cross-linking in Proteins Graphical abstract HHS Public Access. *Chem Rev*. **118**, 1169–1198

10. Cooper, G. M. (2000) The Mechanism of Oxidative Phosphorylation. [online] https://www.ncbi.nlm.nih.gov/books/NBK9885/ (Accessed March 10, 2023)

11. Barford, D. (2004) The role of cysteine residues as redox-sensitive regulatory switches. *Curr Opin Struct Biol*. **14**, 679–686

12. Lingappan, K. (2018) NF-κB in Oxidative Stress. *Curr Opin Toxicol*. **7**, 81

13. Christian, F., Smith, E. L., and Carmody, R. J. (2016) cells The Regulation of NF-κB Subunits by Phosphorylation. 10.3390/cells5010012

14. Delelis, O., Carayon, K., Saïb, A., Deprez, E., and Mouscadet, J. F. (2008) Integrase and integration: Biochemical activities of HIV-1 integrase. *Retrovirology*. **5**, 1–13

15. Engelman, A. N., and Kvaratskhelia, M. (2022) Multimodal Functionalities of HIV-1 Integrase. *Viruses*. 10.3390/V14050926

16. G, E., K, G., A, A., S, M., H, M., R, S., Y, H., FD, B., and G, V. D. (2022) Structure of a HIV-1 IN-Allosteric Inhibitor Complex at 2.93 Å Resolution: Routes to Inhibitor Optimization. 10.1101/2022.06.09.495524

17. Nag, M., and Clark, A. C. (2023) Conserved folding landscape of monomeric initiator caspases. *Journal of Biological Chemistry*. **0**, 103075

18. Reddy Chichili, V. P., Kumar, V., and Sivaraman, J. (2013) Linkers in the structural biology of protein-protein interactions. *Protein Science*. **22**, 153–167

19. Argos, P. (1990) An investigation of oligopeptides linking domains in protein tertiary structures and possible candidates for general gene fusion. *J Mol Biol*. **211**, 943–958

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

21. Blanchard, H., Donepudi, M., Tschopp, M., Kodandapani, L., Wu, J. C., and Grütter, M. G. (2000) Caspase-8 specificity probed at subsite S(4): crystal structure of the caspase-8-Z-DEVD-cho complex. *J Mol Biol*. **302**, 9–16

22. Mátyus, L., Szöllosi, J., and Jenei, A. (2006) Steady-state fluorescence quenching applications for studying protein structure and dynamics. *J Photochem Photobiol B*. **83**, 223–236

CHAPTER 6

Resurrection of ancestral caspases to decipher the evolution of the initiator and executioner subfamily in the extrinsic pathway of apoptosis

Isha Joglekar[1‡], Mithun Nag Karadi Giridhar [1‡], David A. Diaz[1], Melissa Fee[1] and A. Clay Clark[1*]

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

Please note that this work is in progress: Only the preliminary data is displayed

Running title: Exploring the Evolution of Caspase Subfamilies in Apoptosis by Resurrecting Ancestral Proteins

[*]Corresponding author: A. Clay Clark

Email: clay.clark@uta.edu

Key Words: caspase; ancestral reconstruction; apoptosis; protein evolution; conformational dynamics, allostery.

INTRODUCTION

Enzymes evolve function and conformational dynamics through genetic mutations, which result in changes in the amino acid sequence of the protein. Over time, these mutations can accumulate and lead to changes in enzyme function and structure(1). One way that enzymes can evolve new functions while preserving their fold is through gene duplication and divergence(1). When a gene is duplicated, one copy can continue to perform the original function while the other copy can evolve new functions through mutations (1). This allows the enzyme to explore new functional space without compromising its original function. In terms of conformational dynamics, enzymes can evolve changes in their active site or allosteric sites that allow for new interactions with substrates or regulatory molecules(2). These changes can be subtle, such as the addition or removal of a single amino acid, or more dramatic, such as the repositioning of entire domains within the enzyme structure (2). Despite these changes, enzymes generally maintain their overall fold because it is critical for their stability and function(3). The fold of an enzyme is determined by its tertiary and quaternary structure, which is stabilized by a network of non-covalent interactions between amino acids. These interactions help to maintain the structural integrity of the enzyme and ensure that it can perform its catalytic function(3).

The evolutionary origins of caspases are thought to be very ancient, with homologs of caspase-like proteases found in both prokaryotes and eukaryotes(4). However, the specific caspases involved in the extrinsic pathway of apoptosis (caspase-3, 6, 7, 8, 10, and cFLIP) likely evolved later in metazoan evolution, as these caspases are generally only found in multicellular animals. The caspases in the extrinsic pathway

of apoptosis, including caspase-3, 6, 7, 8, 10, and cFLIP, evolved in early vertebrates around 500-600 million years ago during the emergence of the immune system(5). These caspases are found in all vertebrates and are involved in various cellular processes, including apoptosis, inflammation, and immune response(5).

In addition to their apoptotic roles, these caspases also play non-apoptotic roles in various physiological processes. For example, caspase-8 and cFLIP have been implicated in the regulation of immune response, necroptosis and inflammation, while caspase-3 has been shown to be involved in the differentiation of muscle cells and the maintenance of neuronal homeostasis. Caspase-7 has been shown to play a role in autophagy, a cellular process involved in the recycling of cellular components. Furthermore, caspase-6 has been implicated in the regulation of axonal degeneration and neuronal development. Overall, caspases in the extrinsic pathway play a diverse range of roles in addition to their well-known apoptotic function (6, 7).

However, it is currently not well know how these enzymes have evolved these diverse functions during the course of evolution from the common ancestor while preserving their fold. Ancestral protein reconstruction is a powerful tool that allows researchers to study the evolution of protein function and dynamics. In this study, we reconstructed the ancestor of all caspases in the extrinsic pathway of apoptosis and, using three databases, we reconstructed three ancestors from which all present caspases evolved. Due to the probabilistic nature of the reconstructions, we chose to characterize a pool of sequences to examine probabilistic ancestral sequences. These proteins are discussed within the framework of molecular dynamics simulations and network analysis in chapter 2. We possess three ancestors of all enzymes (AOA), of

which AOA1 and AOA3 were successfully expressed. In chapter 2, we demonstrate that AOA1 displays characters resembling initiators and AOA3 displays characters resembling executioners. Similar to caspase-6, the free energy landscape of AOA3 indicates that it is highly stable. In this study, we therefore characterize the two progenitors and alter hydrophobic networks that, in theory, can stabilize the structure and cause the enzyme to have either an initiator-like or an executioner-like character. These hydrophobic networks, which are located at the back of the molecule, may contain the key to whether an enzyme is a monomer or a dimer; hence, this may have contributed to the divergence of initiators and executioners into two subfamilies.

RESULTS

AOA3 and AOA1 are depicted in figure 1A, respectively. AOA3 is an active enzyme that auto-cleaves and generates two cleavages; a small quantity of uncleaved protomer is also detected at the same location as AOA1. AOA1 is not processed and largely behaves as a monomer. The molecular mass was measured by examining the cleavage, and the MALDI-TOF data (figure 1B) indicates a molecular mass of 12.8 kDa and 16.87 kDa. In supplementary figure S1, the cleavage position estimated from molecular weights indicates that the cleavage occurs at the VETE site and as a result generates these fragments in the large subunit and small subunit, and as additional peaks are observed near the base of the 12.6 kDa peak in figure 1A, the SETD site may also be cleaved. Further mutational tests are required to determine if the cleavages occur in both locations or only in one. Nonetheless, the MALDI results indicate that the

cleavage at VETE occurs with greater intensity than in any other site. The molecular weight of AOA1 is determined to be 29.6 kDa.

The mutants that were expressed for AOA1 to mimic AOA3 and AOA3 to imitate AOA1 are depicted in supplemental figure S2 A, with the mutational locations highlighted. The molecular weights of the mutants were measured by MALDI-TOF and are 29.6 kDa for AOA1-5 (supplemental figure S2A) and 29.78 kDa for AOA3-5 (supplemental figure S2B) mutants. Since the AOA3-5 mutation renders the enzyme inactive, the molecular weight of AOA3 is also 29.8 kDa prior to cleavage. Note that the sequence and the mutational information for all the wild type and mutant ancestors is provided in supplemental figure S3

Then, we evaluated the activity of Ac-DEVD-AFC, Ac-IETD-AFC, and Ac-VEID-AFC. The results of the activity indicate that the enzyme has a low level of activity compared to caspase-3 (figure 2A), but is promiscuous. Figure 2B demonstrates that when an inhibitor is added, the enzyme's activity decreases, providing us an indication of which inhibitors to utilize for crystallization. Figure 2B depicts a scale that is tenfold lower than in figure 2A, indicating that the addition of an inhibitor tenfold reduces the enzyme's activity.

In figures 3A and 3B, the fluorescence emission at various urea concentrations between 300nm and 400nm for AOA1 at 280nm and 295nm, respectively, is depicted. Figure 3C depicts how circular dichroism is used to monitor the loss of secondary structure as urea concentrations rise. Overall, the converted and normalized signal from all probes

(CD and fluorescence) indicates that the folding pathway transition follows a three-state model with an intermediate.

Preliminary data from trypsin proteolysis (figure 4) utilizing AOA1, AOA1-5, and AOA3-5 indicate that not enough enzyme is used for AOA1-5 and AOA3-5, and so the data cannot be examined. This component of the to-be-obtained results need additional optimization. With AOA1, it appears that the cleavage rate is comparable to that of caspase-8, indicating that the enzyme is as stable as caspase-8, given that the same amount of trypsin (0.2 uM) was employed for cleavage. These investigations will be conducted to gain a comprehension of the less stable regions and to confirm results from urea simulations (in progress, data not displayed) comparable to the chapter1 study.

CONCLUSION

Overall, studying ancestral enzymes can provide important insights into the evolution of enzyme function and help us understand how little changes in an enzyme's amino acid sequence can alter its activity and specificity. These discoveries can be used to influence the development of new enzymes for biotechnological applications and to better understand the molecular basis of disease.

Urea is a chemical denaturant that breaks protein natural shape by breaking hydrogen bonds and other non-covalent interactions, causing the protein to unfold. Experimenting on protein stability may be time-consuming and expensive, but simulations can give quick insights into protein thermodynamics and conformational dynamics. Simulations

can also reveal structural properties of proteins that are critical for stability and function, such as the existence of disordered or flexible areas, which are difficult to investigate experimentally. Note that simulations of the ancestral proteins are not shown, but these experimental results will be compiled and linked to the simulations for the article that will be published once all of the experimental data is collected, so this write up is more of a guide for future studies, but we believe the conclusions will still hold as far as the preliminary data suggest.

Figure 1

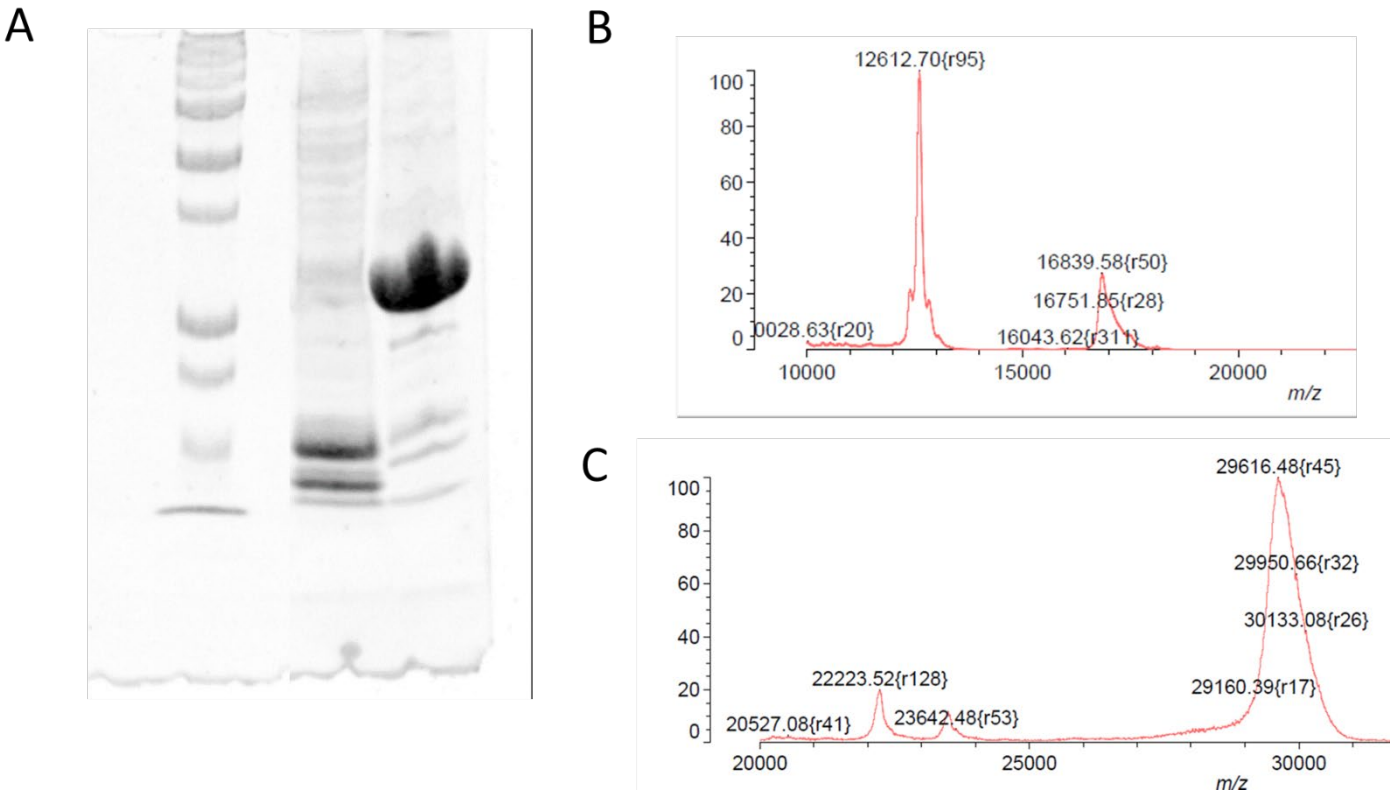


Figure 1. Experimental characterization of resurrected Ancestral proteins. (A) AOA1 and AOA3 were purified in the lab and their masses were determined by MALDI-TOF for (B) AOA3 and (C) AOA1.

Figure 2

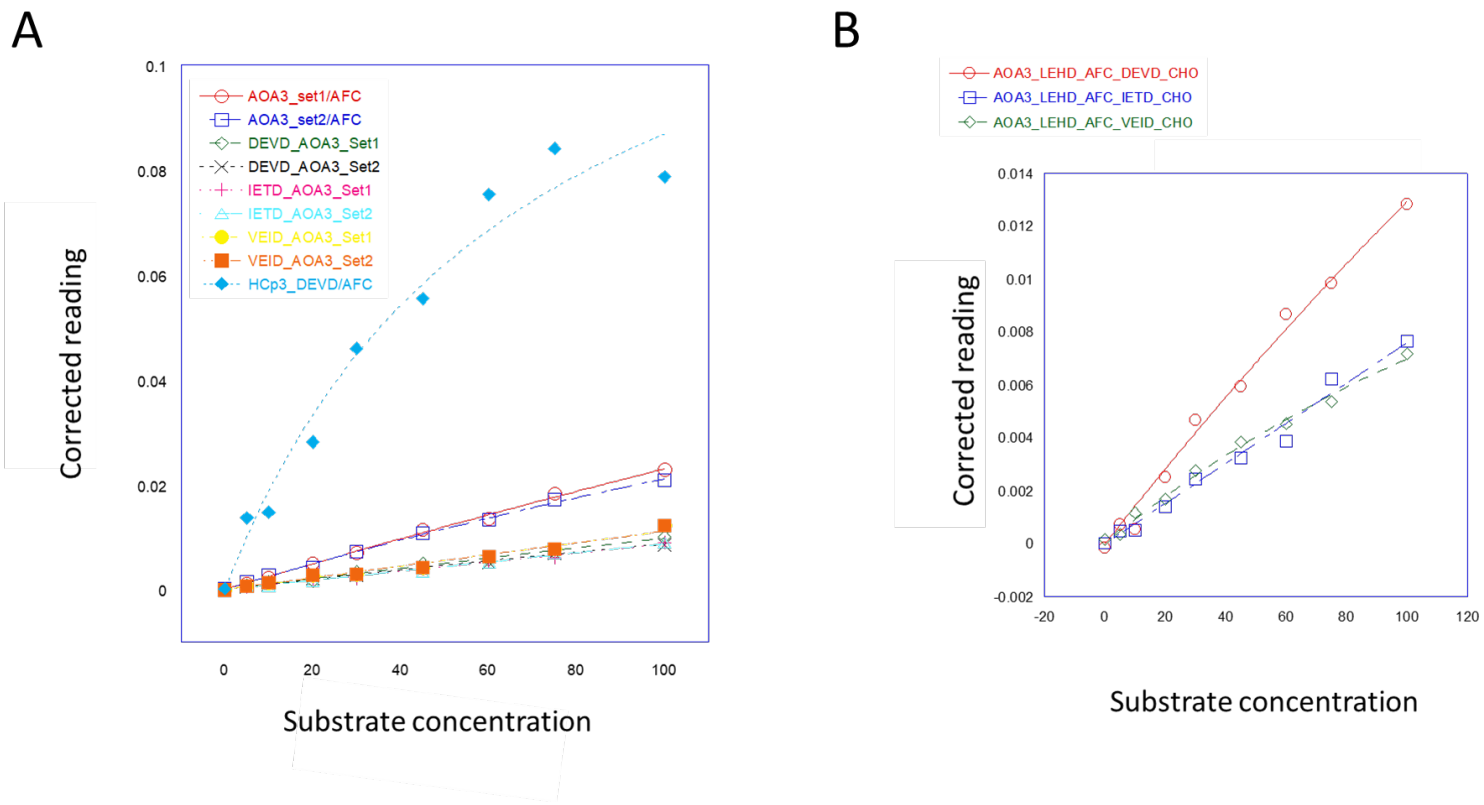


Figure 2. Activity assay to determine the activity of active AOA3. (A) Activity of AOA3 determined for Ac-DEVD-AFC, Ac-VEID-AFC and Ac-IETD-AFC compared to that of caspase-3 in Ac-DEVD-AFC. (B) Activity assay done with Ac-DEVD-CHO, Ac-IETD-CHO and Ac-VEID-CHO along with their AFC counterparts.

Figure 3

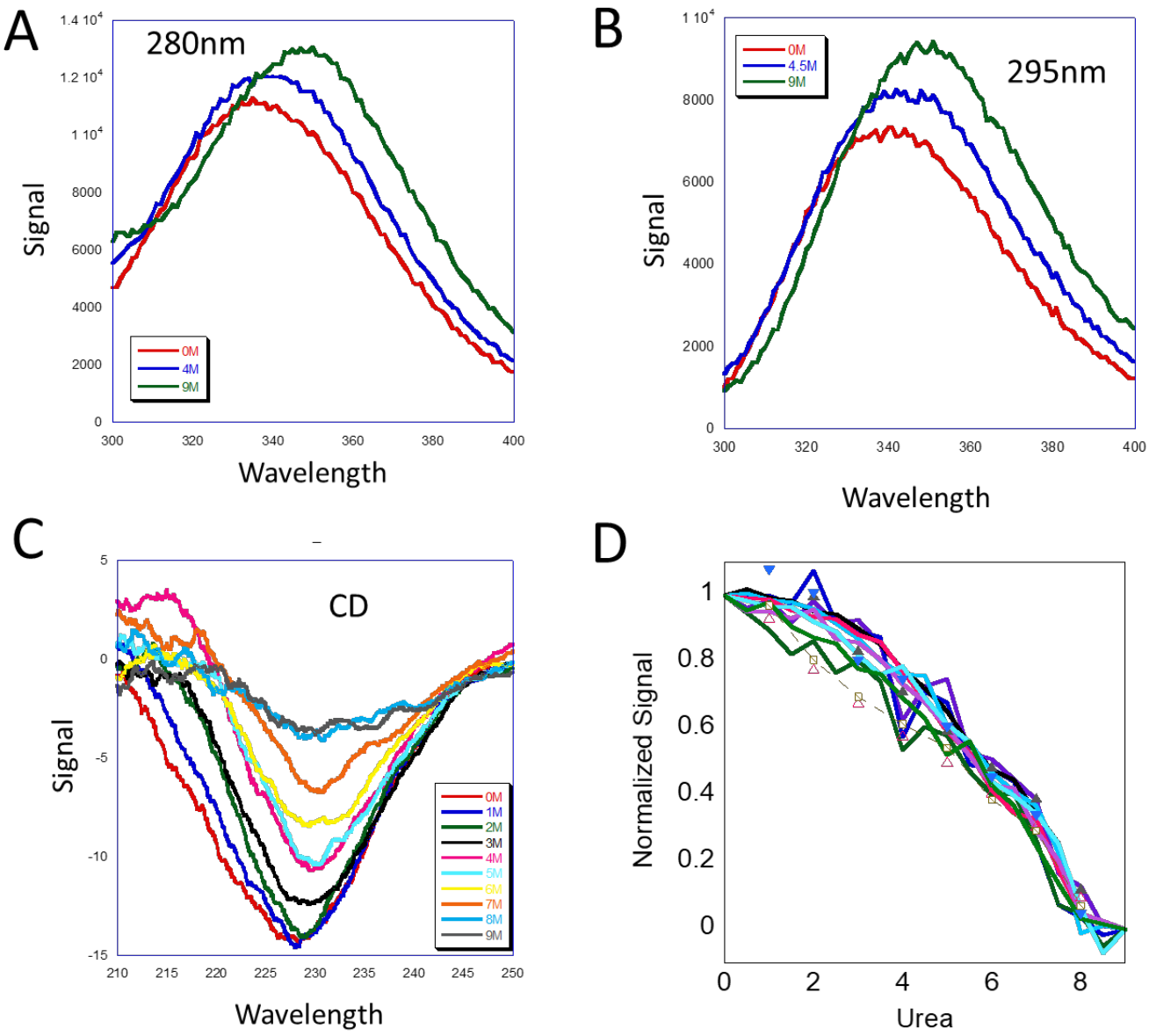


Figure 3. Urea induced denaturation studies of AOA1 with fluorescence spectroscopy and circular dichroism. AOA1 unfolding studied by excitation at (A) 280nm and (B) 295n, the emission was collected from 300- 400nm. The signal shows a blue shift as the protein unfolds. (C) Circular dichroism was another probe used to monitor the unfolding of AOA1. (D) Average emission wavelength was normalized and plotted for fitting, data shows that the protein can refold.

Figure 4

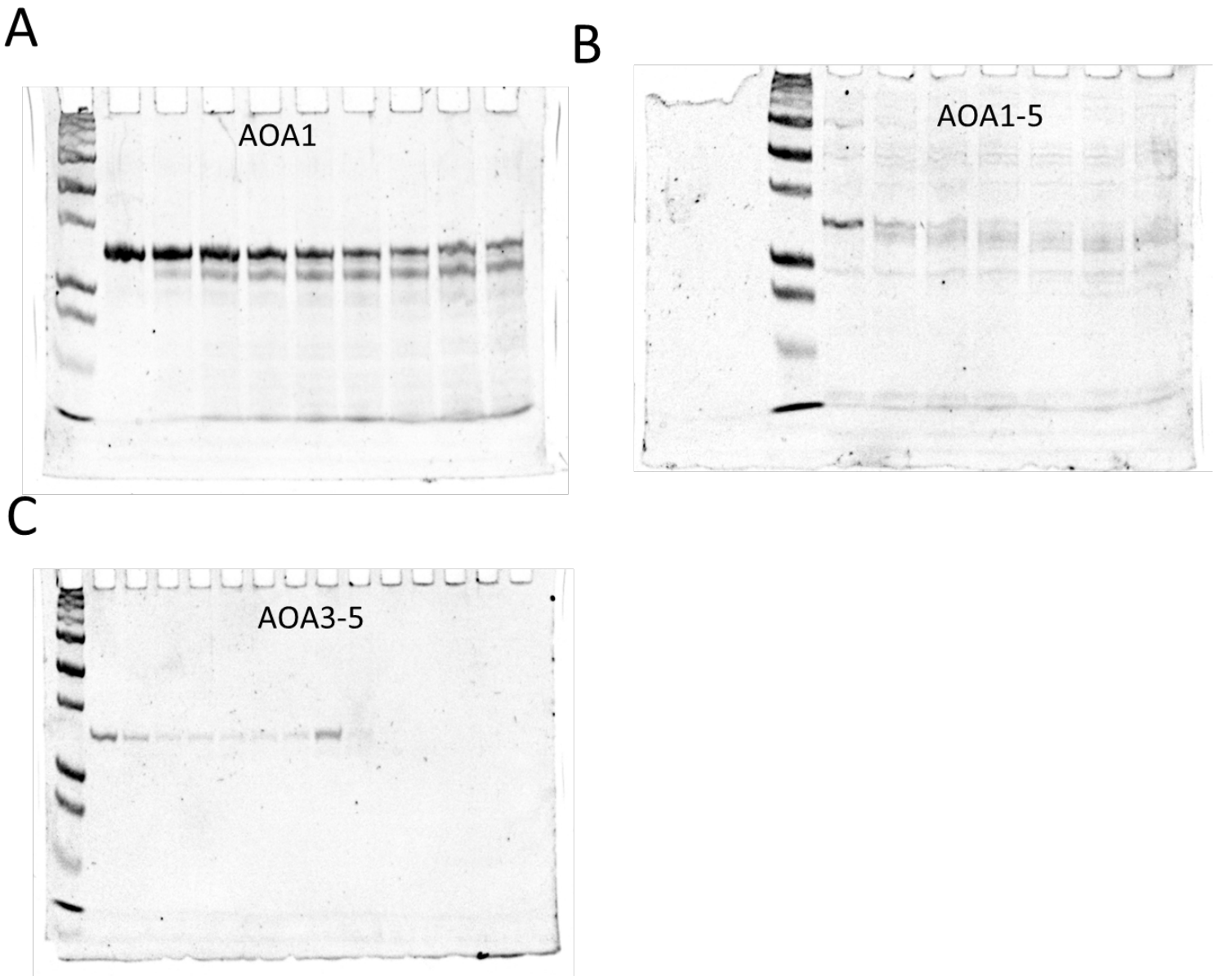


Figure 4. Limited trypsin proteolysis for obtaining structural information of (A) AOA1 its mutant (B) AOA1-5 and (C) AOA3-5. Note that these experiments have to be repeated to optimize the enzyme concentration.

Supporting Information

Supplemental figure S1

Large subunit

```
         10         20         30         40         50         60
MSQSKPSSPS EQYKMNYKRR GICIIINNKN FDKQTGLSDR NGTDKDADNL KKTFKNLGFE

         70         80         90        100        110        120
VKVYNDLTAE EMKKLLKEVS KEDHSDSDCF VCVLLSHGEE GHIYGTDGAI PIKNLTSLFR

        130        140        150
GDKCQSLVGK PKLFFIQACR GNEFDDGVET
```

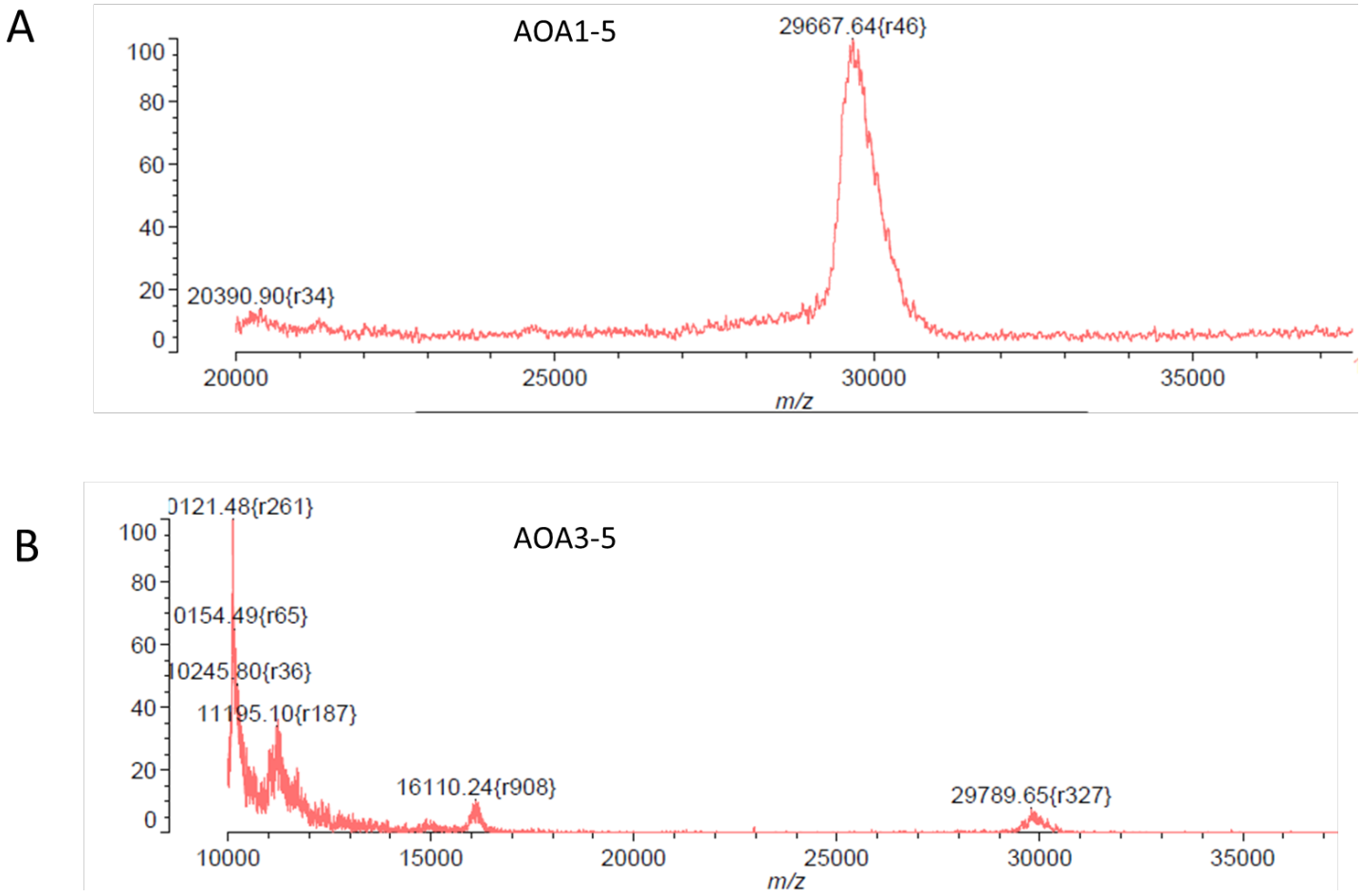
Small subunit

```
        10         20         30         40         50         60
SGSETDASSL YTIPAEADFL LAYSTAEGYY SYRNTANGSW YIQSLCEMLK KYGKKLEIMQ

        70         80         90        100        110
ILTRVNRKVS QKSESVSNDP GASGKKQIPC IASMLTKKLY FHPPSNLEHH HHHH
```

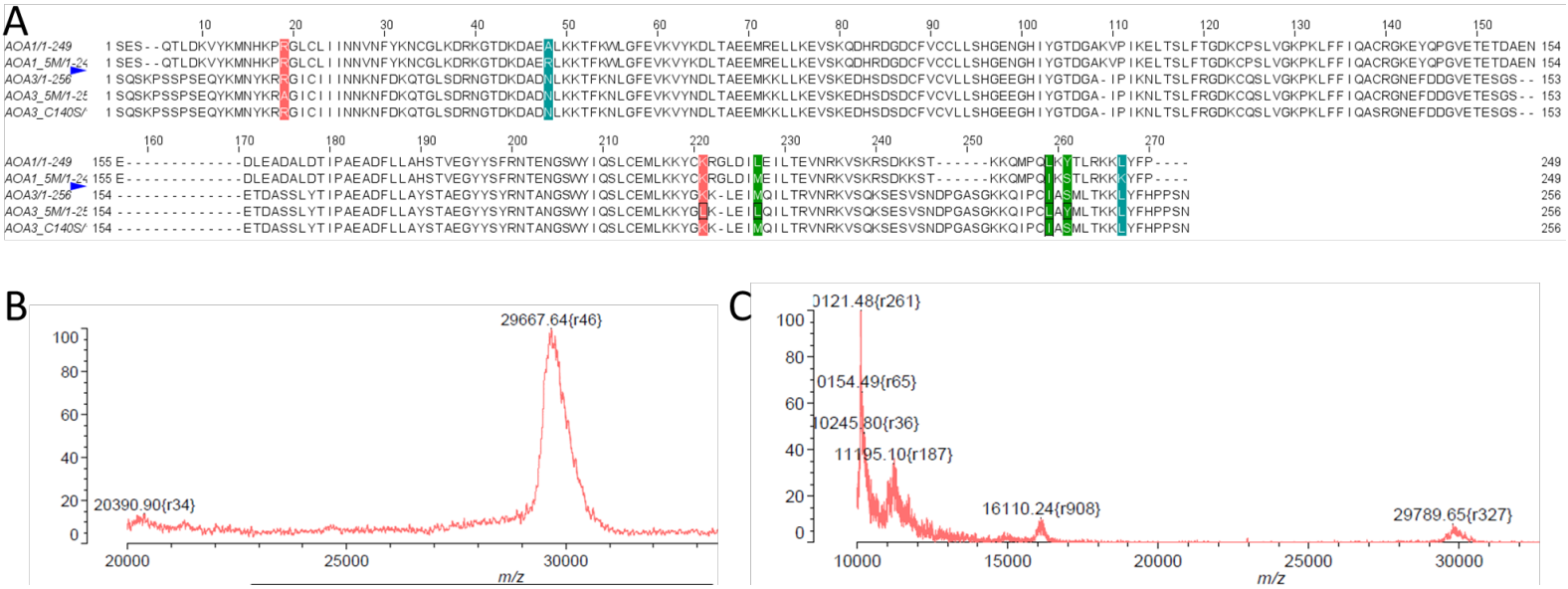
Supplemental figure S1. MALDI-TOF was used to determine the autocatalytic cleavage site VETE for the active AOA3 enzyme, and the figure depicts the fragments that arise from this autocatalytic activity.

Supplemental figure S2



Supplemental figure S2. MALDI-TOF was utilized to assess the molecular weight of AOA1 (AOA1-5) and AOA3 (AOA3-5) mutations. (A) The mass of AOA1-3 was determined to be 29.6 kDa, while the mass of AOA3-5 was determined to be 29.7 kDa (B).

Supplemental figure S3



Supplemental figure S3. Sequence of all the mutants that have been purified in the lab.

References

1. Copley, S. D. (2020) Evolution of new enzymes by gene duplication and divergence. *FEBS J*. **287**, 1262
2. Caroline, S., Kamerlin, L., Sanchez-Ruiz, J. M., Petrovic´1, D., Petrovic´1, P., and Risso, V. A. Headline review Conformational dynamics and enzyme evolution. 10.1098/rsif.2018.0330
3. Tyzack, J. D., Furnham, N., Sillitoe, I., Orengo, C. M., and Thornton, J. M. (2017) Understanding enzyme function evolution from a computational perspective. *Curr Opin Struct Biol*. **47**, 131–139
4. Asplund-Samuelsson, J., Bergman, B., and Larsson, J. (2012) Prokaryotic Caspase Homologs: Phylogenetic Patterns and Functional Characteristics Reveal Considerable Diversity. *PLoS One*. **7**, 49888
5. Mcilwain, D. R., Berger, T., Mak, T. W., Baehrecke, E. H., Green, D. R., Kornbluth, S., and Salvesen, G. S. Additional Perspectives on Cell Survival and Cell Death available at Harbor Laboratory Press at University of Texas at Arlington on March 12. 10.1101/cshperspect.a008656
6. Yu, J. W., Jeffrey, P. D., and Shi, Y. (2009) Mechanism of procaspase-8 activation by c-FLIPL. *Proc Natl Acad Sci U S A*. **106**, 8169–8174
7. Schwerk, C., and Schulze-Osthoff, K. (2003) Non-apoptotic functions of caspases in cellular proliferation and differentiation. *Biochem Pharmacol*. **66**, 1453–1458

CHAPTER 7

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

Received 24 January 2022, Revised 5 April 2022, Available online 12 April 2022, Version of Record 26 May 2022.

Please note that the introduction has been altered to emphasize the findings, but the remainder of the material is our contribution, with the same numbering, content, and citations as the article.

INTRODUCTION

Inflammatory caspases are a group of caspases that play important roles in immune response and inflammatory signaling pathways. They are responsible for cleaving pro-inflammatory cytokines such as IL-1β and IL-18, which are involved in a variety of inflammatory processes. The evolution of these caspases and their specific functions have been of interest to researchers. One particular aspect of interest is the evolutionary process that generated the caspase-1 function on the caspase-4 catalytic domain scaffold.

Earlier work by Robert D. Grinshpon and colleagues paved the way for the Clark lab to utilize ancestral reconstruction techniques to perform biochemical research which resulted in two publications, ‘Resurrection of ancestral effector caspases identifies novel networks for evolution of substrate specificity’ and ‘Evolution of the folding landscape of effector caspases’. Following their work we followed to utilize ancestral reconstruction studies in our work which we mastered and had the chance to collaborate with Betsaida Bibo-Verdugo and Guy S. Salvesen followed by other authors Monica L. Ramirez , Scott J. Snipas and Marcin Poreba. Ancestral sequence reconstruction (ASR) was utilized 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. The resurrected protein represents the node from which the Carnivora inflammatory caspases diverged. The study aimed to identify the characteristics of this protein and the evolutionary process that led to the emergence of the Caspase-1 function on the 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

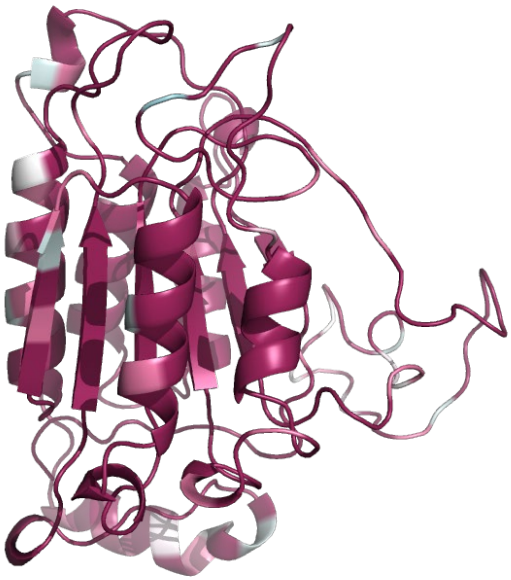
statistical analysis of sequence conservation and substitutions of existing proteins within a family. These relationships allow for calculation of sequences that represent the diverging nodes within the phylogenetic tree and thus the ancestor of each branch. The resurrected protein represents the node from which the Carnivora inflammatory caspases diverged. We termed this caspase “node 22” based on the position on the phylogenetic tree (Figs. 4A and S4). The results of ASR analysis are site-specific probabilities for each position in the protein sequence (33, 34). 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 (Fig. S5). 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 (50), Ensembl (51), and the National Center for Biotechnology Information to retrieve related proteins employing the dog inflammatory caspase (UniProt: A9YEF4) as a query. The Figure 7. A hypothetical process in the evolution of the Carnivora inflammatory caspase. Before the Carnivora, there were two inflammatory caspases, caspase-1 and caspase-4. Caspase-1 was lost in a Carnivora ancestor possibly upon a

first selective pressure for which production of biologically active IL-1β was detrimental. Later, an early Carnivora encountered other selective pressures that required the IL-1β response, and caspase-4 gained this activity. Extant Carnivora possess an inflammatory caspase with caspase-4 sequence homology but caspase-1 catalytic function. IL-1β, interleukin-1β. Evolution of an inflammatory caspase locus J. Biol. Chem. (2022) 298(6) 101931 7 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 (52) (Table S2). PROMALS3D (prodata.swmed.edu/promals3d) generated structure-based alignments (53), and sequences were pruned on Jalview (jalview.org) (54) 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. (55). 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) (56).

Figure 5A



Color Key

0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9

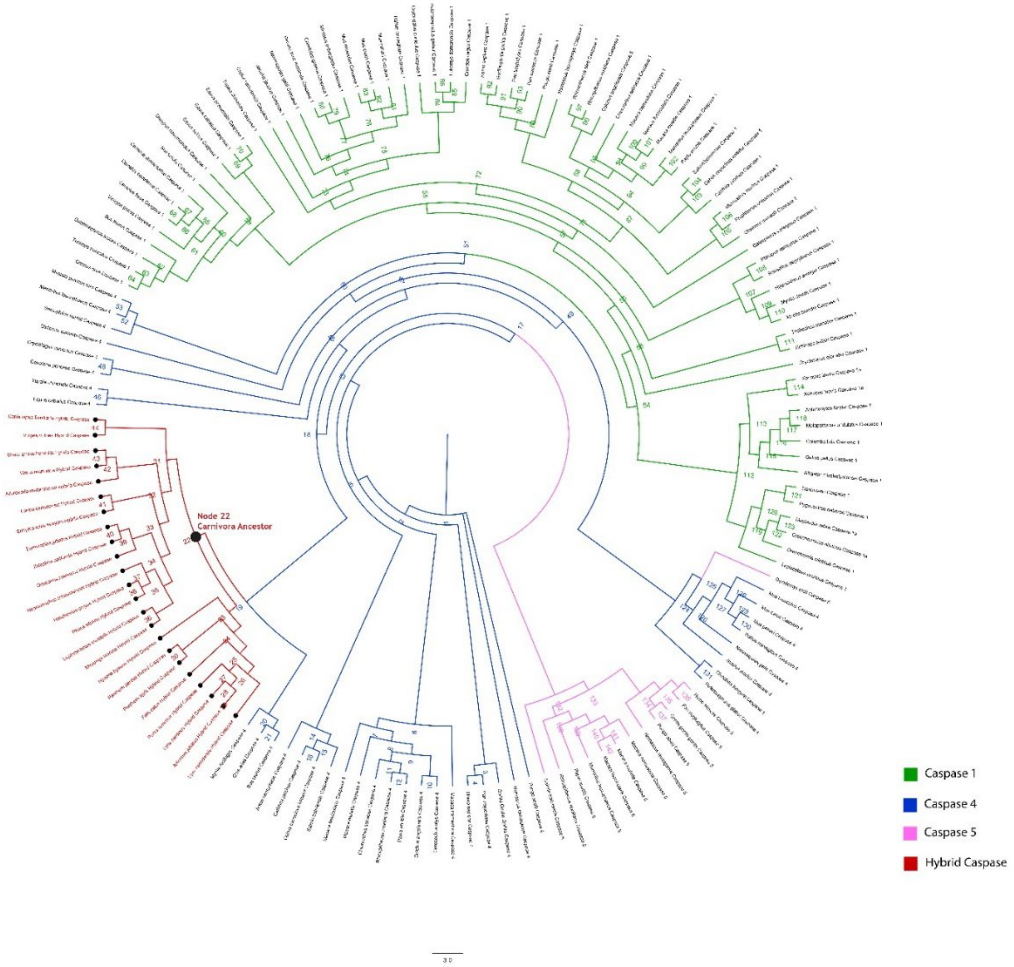
Supporting Information

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

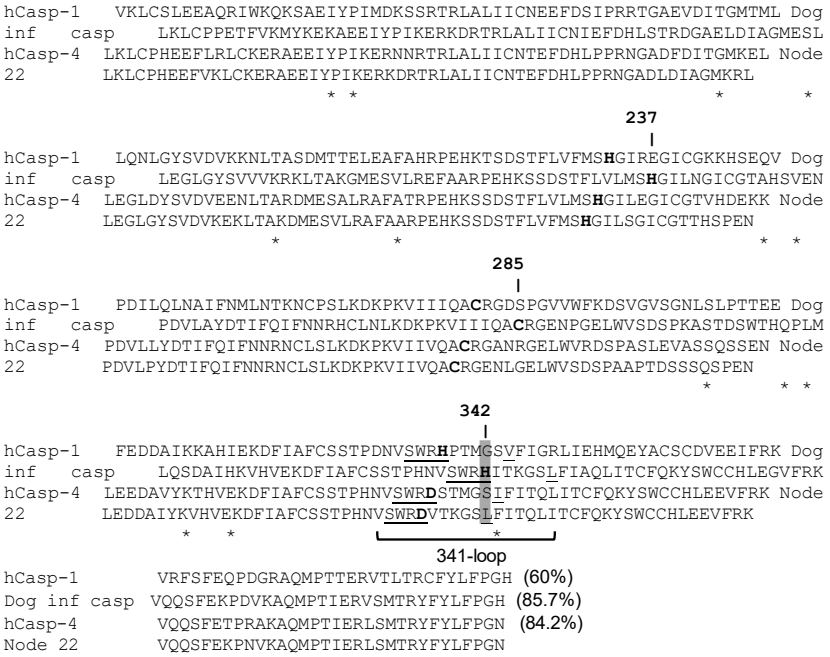
Caspase	Species	Caspase	Species
Caspase-1	Alligator mississippiensis	Caspase-1	Mandrillus leucophaeus
Caspase-1	Aptenodytes forster	Caspase-1a	Maylandia zebra
Caspase-1	Bos taurus	Caspase-1	Melopsittacus undulatus
Caspase-1	Camelus bactrianus	Caspase-1	Microcebus murinus
Caspase-1	Camelus dromedarius	Caspase-1	Microtus ochrogaster
Caspase-1	Camelus ferus	Caspase-1	Mus caroli
Caspase-1	Callithrix jacchus	Caspase-1	Mus musculus
Caspase-1	Castor canadensis	Caspase-1	Mus Pahari
Caspase-1	Cebus capucinus imitator	Caspase-1	Myotis brandtii
Caspase-1	Chlorocebus sabaeus	Caspase-1	Myotis davidii
Caspase-1	Colobus angolensis	Caspase-1	Nannospalax galili
Caspase-1	Columba livia	Caspase-1	Nomascus leucogenys
Caspase-1	Cricetulus griseus	Caspase-1	Octodon degus
Caspase-1	Danio rerio	Caspase-1	Otolemur garnettii
Caspase-1	Dasypus novemcinctus	Caspase-1	Orcinus orca
Caspase-1	Delphinapterus leucas	Caspase-1	Oreochromicus niloticus
Caspase-1	Echinops telfairi	Caspase-1a	Oreochromus niloticus
Caspase-1	Equus asinus	Caspase-1	Orycteropus afer afer
Caspase-1	Equus caballus	Caspase-1	Oryctogalus cuniculus
Caspase-1	Equus przewalskii	Caspase-1	Pan paniscus
Caspase-1	Fukomys damarensis	Caspase-1	Pan troglodytes
Caspase-1	Galeopterus variegatus	Caspase-1	Papio Anubis
Caspase-1	Gallus gallus	Caspase-1	Peromyscus manicula
Caspase-1	Gorilla gorilla gorilla	Caspase-1	Pongo abellii
Caspase-1	Heterocephalus glaber	Caspase-1	Propithecus coquereli
Caspase-1	Hipposideros armiger	Caspase-1	Pteroptus vampyrus
Caspase-1	Homo sapiens	Caspase-1	Pygocentrus nattereri
Caspase-1	Jaculus jaculus	Caspase-1	Rattus norvegicus
Caspase-1	Lepisosteus oculatus	Caspase-1	Rhinopthecus bieti
Caspase-1	Macaca fascicularis	Caspase-1	Rhinopthecus roxellana
Caspase-1	Macaca mulatta	Caspase-1	Rousettus aegyptiacus
Caspase-1	Macaca nemestrina	Caspase-1	Saimiri boliviensis

Caspase	Species	Caspase	Species
Caspase-1	Sus scrofa	Caspase-4	Papio anubis
Caspase-1	Trichechus manatus	Caspase-4	Otolemur garnettii
Caspase-1	Tupaia chinensis	Caspase-4	Ovis aries
Caspase-1	Tursiops truncatus	Caspase-4	Pan troglodytes
Caspase-1	Vicugna pacos	Caspase-4	Pongo abelii
Caspase-1a	Xenopus laevis	Caspase-4	Rattus norvegicus
Caspase-1b	Xenopus laevis	Caspase-4	Rhinopthecus roxellana
Caspase-4	Aotus nancymaae	Caspase-4	Saimiri boliviensis
Caspase-4	Bos taurus	Caspase-4	Sarcophilus harrisi
Caspase-4	Callithrix jacchus	Caspase-4	Tupaia chinensis
Caspase-4	Cebus capucinus imitator	Caspase-5	Dipodomys ordii
Caspase-4	Cerocebus atys	Caspase-5	Gorilla gorilla gorilla
Caspase-4	Chinchilla lanigera	Caspase-5	Homo sapiens
Caspase-4	Chlorocebus sabaeus	Caspase-5	Macaca fascicularis
Caspase-4	Colobus angolensis	Caspase-5	Macaca mulatta
Caspase-4	Equus caballus	Caspase-5	Macaca nemestrina
Caspase-4	Gorilla gorilla gorilla	Caspase-5	Mandrillus leucophaeus
Caspase-4	Heterocephalus glaber	Caspase-5	Nomascus leucogenys
Caspase-4	Homo sapiens	Caspase-5	Pan troglodytes
Caspase-4	Jaculus jaculus	Caspase-5	Papio anubis
Caspase-4	Macaca fascicularis	Caspase-5	Pongo abelii
Caspase-4	Macaca mulatta	Caspase-5	Rhinopthecus roxellana
Caspase-4	Macaca nemestrina	Caspase-5	Saimiri boliviensis
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		

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



References

3. Harms, M. J., and Thornton, J. W. (2010) Analyzing protein structure and function using ancestral gene reconstruction. Curr. Opin. Struct. Biol. 20,360–366

34. Eick, G. N., Bridgham, J. T., Anderson, D. P., Harms, M. J., and Thornton, J. W. (2016) Robustness of reconstructed ancestral protein functions to statistical uncertainty. Mol. Biol. Evol. 34, 247–261

50. Consortium, T. U. (2020) UniProt: The universal protein knowledgebase in 2021. Nucl. Acids Res. 49, D480–D489

51. Howe, K. L., Achuthan, P., Allen, J., Allen, J., Alvarez-Jarreta, J., Amode, M. R., Armean, I. M., Azov, A. G., Bennett, R., Bhai, J., Billis, K., Boddu, S., Charkhchi, M., Cummins, C., Da Rin Fioretto, L., et al. (2020) Ensembl 2021. Nucl. Acids Res. 49, D884–D891

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

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

54. Waterhouse, A. M., Procter, J. B., Martin, D. M. A., Clamp, M., and Barton, G. J. (2009) Jalview Version 2—a multiple sequence alignment editor and analysis workbench. Bioinformatics 25, 1189–1191

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

56. Kelley, L. A., Mezulis, S., Yates, C. M., Wass, M. N., and Sternberg, M. J. E. (2015) The Phyre2 web portal for protein modeling, prediction and analysis. Nat. Protoc. 10, 845–858

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
```

Igor Pro pKa fits for titration:

```
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
```

GROMACS commands for urea, water simualtions and free energy landscapes

MD commands in 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
```

Note add the following details (not in bold below) in a file named: 'urea_comp_simulation_of_urea_paper_G96BOND.itp' so that it can be used in the urea simualtions, see md commands in urea above for reference.

```
[ 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
```

MDP files

Energy minimization file:

```
; 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                = Urea and Cp8 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
```

Groamcs in high performance computing: slurm commands

The following section contains slurm files for UTA HPC: Note partitions can be changed to long and normal. The time limit is more for long and normal to run bigger simualtions.

```
#!/bin/bash
#SBATCH --job-name=D_Cp6_exhel_urea
#SBATCH --partition=short
#SBATCH --time=4-00:00:00
#SBATCH -e slurm-%j.err
```

```
#SBATCH --mail-user=mithunnag.karadigiridhar@mavs.uta.edu
#SBATCH --nodes=1
export LD_LIBRARY_PATH=/home/karadigiridhm/Urea/Anc_cFLIP

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_Cp6_exhel_md.tpr
gmx_mpi mdrun -s D_Cp6_exhel_md.tpr -o D_Cp6_exhel_md.trr -c D_Cp6_exhel_md.gro -g
D_Cp6_exhel_md.log -e D_Cp6_exhel_md.edr
echo 1 1|gmx_mpi trjconv -f D_Cp6_exhel_md.trr -s D_Cp6_exhel_md.tpr -skip 50 -center -pbc nojump -o
D_Cp6_exhel_out.pdb
echo 1 1|gmx_mpi trjconv -f D_Cp6_exhel_md.trr -s D_Cp6_exhel_md.tpr -o md_red_D_Cp6_exhel.trr -
center -pbc nojump
echo 1|gmx_mpi gyrate -f md_red_D_Cp6_exhel.trr -s D_Cp6_exhel_md.tpr -o
D_Cp6_exhel_radgyration.xvg
echo 1|gmx_mpi rmsf -f md_red_D_Cp6_exhel.trr -s D_Cp6_exhel_md.tpr -o D_Cp6_exhel_rmsf.xvg -ox
D_Cp6_exhel_md_avg.pdb -oq D_Cp6_exhel_bfactor.pdb -res
gmx_mpi trjconv -f D_Cp6_exhel_md.trr -o D_Cp6_exhel_md.xtc
echo 1 1|gmx_mpi covar -f D_Cp6_exhel_md.xtc -s D_Cp6_exhel_md.tpr -o D_Cp6_exhel_eigenval.xvg -
v D_Cp6_exhel_eigenvect.trr -xpm D_Cp6_exhel_covar.xpm
gmx_mpi xpm2ps -f D_Cp6_exhel_covar.xpm -o D_Cp6_exhel_covar.eps -do D_Cp6_exhel_covar.m2p
echo 1 1|gmx_mpi anaeig -v D_Cp6_exhel_eigenvect.trr -f D_Cp6_exhel_md.xtc -s D_Cp6_exhel_md.tpr
-first 1 -last 2 -proj D_Cp6_exhel_proj_eig.xvg -2d D_Cp6_exhel_2d_proj.xvg
gmx_mpi sham -f D_Cp6_exhel_2d_proj.xvg -ls D_Cp6_exhel_gibbs.xpm -notime
gmx_mpi xpm2ps -f D_Cp6_exhel_gibbs.xpm -o D_Cp6_exhel_gibbs.eps -rainbow red
echo 1 1|gmx_mpi anaeig -v D_Cp6_exhel_eigenvect.trr -s D_Cp6_exhel_md.tpr  -f D_Cp6_exhel_md.xtc
-extr D_Cp6_exhel_ext1.pdb -first 1 -last 1 -nframes 400
```

echo 1 1|gmx_mpi anaeig -v D_Cp6_exhel_eigenvect.trr -s D_Cp6_exhel_md.tpr -f D_Cp6_exhel_md.xtc -extr D_Cp6_exhel_ext2.pdb -first 2 -last 2 -nframes 400

Note to run from stop use: '-cpi state' in the run line of slurm commands if the simualtion stops half way due to time limit

Slurms Files for XSEDE supercomputer: Intel Knights landing nodes:

Note run on compute nodes:

```
#!/bin/bash
#SBATCH -J NPT_2            # job name
#SBATCH -e NPT_2.%j.err     # error file name
#SBATCH -o NPT_2.%j.out     # output file name
#SBATCH -N 20               # request 1 node
#SBATCH -n 48               # request all 48 cores
#SBATCH -p normal           # designate queue
#SBATCH -t 48:00:00         # designate max run time
#SBATCH -A TG-BIO220069     # charge job to myproject

module load gcc/9.1.0  mvapich2/2.3.7
module load  intel/19.1.1 impi/19.0.9
module load gromacs/2022.1
ibrun gmx_mpi_knl mdrun -s em.tpr -o em.trr -c b4NVT.gro -g em.log -e em.edr
ibrun gmx_mpi_knl grompp -maxwarn 2 -f New_NVT.mdp -c b4NVT.gro -r b4NVT.gro -p protein.top -o NVT.tpr
ibrun gmx_mpi_knl mdrun -s NVT.tpr -o NVT.trr -c b4NPT.gro -g NVT.log -e NVT.edr
ibrun gmx_mpi_knl grompp -maxwarn 2 -f New_NPT.mdp -c b4NPT.gro -r b4NPT.gro -p protein.top -o NPT.tpr
ibrun gmx_mpi_knl mdrun -s NPT.tpr -o NPT.trr -c b4md.gro -g NPT.log -e NPT.edr
ibrun gmx_mpi_knl grompp -f Run.mdp -c b4md.gro -r b4md.gro -p protein.top -o md.tpr
ibrun gmx_mpi_knl mdrun -s md.tpr -o md.trr -c md.gro -g md.log -e md.edr
```

To run simualtions from stop

```
ibrun gmx_mpi_knl mdrun -cpi state -s M_AO810_md.tpr -o M_AO810_md.trr -c M_AO810_md.gro -g M_AO810_md.log -e M_AO810_md.edr
```

The following generates the free energy landscapes.

```
#!/bin/bash
#SBATCH -J FEL            # job name
```

```
#SBATCH -e FEL.%j.err       # error file name
#SBATCH -o FEL.%j.out       # output file name
#SBATCH -N 1                # request 1 node
#SBATCH -n 48               # request all 48 cores
#SBATCH -p normal           # designate queue
#SBATCH -t 48:00:00         # designate max run time
#SBATCH -A TG-BIO220069     # charge job to myproject


module load gcc/9.1.0  mvapich2/2.3.7
module load  intel/19.1.1 impi/19.0.9
module load gromacs/2022.1

echo 1 1|ibrun gmx_mpi_knl trjconv -f md.trr -s md.tpr -skip 50 -center -pbc nojump -o D_Cp10_out.pdb
echo 1 1|ibrun gmx_mpi_knl trjconv -f md.trr -s md.tpr -o md_red_D_Cp10.trr -center -pbc nojump
echo 1|ibrun gmx_mpi_knl -f md_red_D_Cp10.trr -s D_Cp10_md.tpr -o D_Cp10_radgyration.xvg
echo 1|ibrun gmx_mpi_knl rmsf -f md_red_D_Cp10.trr -s D_Cp10_md.tpr -o D_Cp10_rmsf.xvg -ox
D_Cp10_md_avg.pdb -oq D_Cp10_bfactor.pdb -res
ibrun gmx_mpi_knl trjconv -f md.trr -o D_Cp10_md.xtc
echo 1 1|ibrun gmx_mpi_knl covar -f D_Cp10_md.xtc -s D_Cp10_md.tpr -o D_Cp10_eigenval.xvg -v
D_Cp10_eigenvect.trr -xpm D_Cp10_covar.xpm
ibrun gmx_mpi_knl xpm2ps -f D_Cp10_covar.xpm -o D_Cp10_covar.eps -do D_Cp10_covar.m2p
echo 1 1|ibrun gmx_mpi_knl anaeig -v D_Cp10_eigenvect.trr -f D_Cp10_md.xtc -s D_Cp10_md.tpr -first 1
-last 2 -proj D_Cp10_proj_eig.xvg -2d D_Cp10_2d_proj.xvg
ibrun gmx_mpi_knl sham -f D_Cp10_2d_proj.xvg -ls D_Cp10_gibbs.xpm -notime
ibrun gmx_mpi_knl xpm2ps -f D_Cp10_gibbs.xpm -o D_Cp10_gibbs.eps -rainbow red
```

Other useful commands on command line:

To run on interactive session:

Idev

To find the module use:

module spider gromacs/2022.1

Modules to load:

module load intel/19.1.1 impi/19.0.9
module load gromacs/2022.1

MATLAB commands to generate energy landscape:

```
>> M_AFA = table2array(Alltogether);
>> x_M_AFA = M_AFA(:,1);
>> y_M_AFA = M_AFA(:,2);
>> z_M_AFA = M_AFA(:,3);
>> xv_M_AFA = linspace(min(x_M_AFA), max(x_M_AFA), 1025);
>> yv_M_AFA = linspace(min(y_M_AFA), max(y_M_AFA), 1025);
>> [X_M_AFA, Y_M_AFA] = meshgrid(xv_M_AFA, yv_M_AFA);
>> Z_M_AFA = griddata(x_M_AFA, y_M_AFA, z_M_AFA, X_M_AFA, Y_M_AFA);

>> M_Anc_cFLIP = table2array(Alltogether1);
>> x_M_Anc_cFLIP = M_Anc_cFLIP(:,1);
>> y_M_Anc_cFLIP = M_Anc_cFLIP(:,2);
>> z_M_Anc_cFLIP = M_Anc_cFLIP(:,3);
>> xv_M_Anc_cFLIP = linspace(min(x_M_Anc_cFLIP), max(x_M_Anc_cFLIP), 1025);
>> yv_M_Anc_cFLIP = linspace(min(y_M_Anc_cFLIP), max(y_M_Anc_cFLIP), 1025);
>> [X_M_Anc_cFLIP, Y_M_Anc_cFLIP] = meshgrid(xv_M_Anc_cFLIP, yv_M_Anc_cFLIP);
>> Z_M_Anc_cFLIP = griddata(x_M_Anc_cFLIP, y_M_Anc_cFLIP, z_M_Anc_cFLIP, X_M_Anc_cFLIP,
Y_M_Anc_cFLIP);
```

Notes: Here the variable M_AFA (line 1 of command) was assigned to the left hand side which displays a function tablw2array which is used to convert the table format of excel into an array that is used by matlab. The excel files is named as Alltogethe and hence has to be mentioned in brackets so that the variable name is assigened to this array. In line 2,3and 4 we are assiging x,y and z values with different variable names. Note: (:,1) menas the first column. Next other variable are assigned as and they are important in creating the surface here is the explaning for each;

linspace is a function in MATLAB that generates a vector of linearly spaced, equally spaced values between two specified endpoints. The syntax for the function is linspace(start, stop, n), where start is the initial value, stop is the final value, and n is the number of points to be generated between start and stop. For instance, the vector [0 1 2 3 4 5 6 7 8 9 10] would be generated by linspace(0, 10, 11).

meshgrid is a function that produces a coordinate grid for charting surfaces and vector fields. The syntax of the function is [X,Y] = meshgrid(x,y), where x and y are corresponding vectors of x and y coordinates. The output matrices X and Y have the same dimension and represent all conceivable combinations of the input vectors. meshgrid is frequently employed in conjunction with other functions, such as surf and quiver, to plot surfaces and vector fields.

griddata is a function used to interpolate scattered data onto a grid. Z = griddata(x,y,z,Xi,Yi) is the function syntax, where x and y are vectors of x and y coordinates of the scattered data, z is a vector of the matching function values, and Xi and Yi are matrices describing the grid locations where the interpolated values will be computed. Z is the output matrix that contains interpolated values at the specified grid points. griddata utilizes different interpolation techniques, including linear, cubic, and nearest neighbor interpolation, to compute interpolated values.

Now the image has to be potted. Use the following commands below:

```
>> figure(1)
>> surf(X_M_AFA, Y_M_AFA, Z_M_AFA);
>> shading interp
hold on
>> surf(X_M_Anc_cFLIP, Y_M_Anc_cFLIP, Z_M_Anc_cFLIP);
>> shading interp
```

Here for the bracket next to figure a number can be assigned that opens up a new figure window with a differnt figure for a different number. Surf plots the surface from the variable assigned in the previous code. Shading interp: Shading can be changed to different colors, refer to the manual for generating different figures. Note that hold on is used to plot all the landscapes in the same image overlaid onto each other. If differnet figures are to be generate for AFA and Anc_cFLIP then remove hold on and add >>figure(2) in its place.

Jupyter notebook commands

Note that after each bold text add the command in different cells.

```
import pandas as pd
import numpy as np
```

Extract all files in folder use the following:

```
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('.xlsx'):
        files.append(file)
print(files)
```

Split dataframes

```
for f in files:

    name = f.rstrip('.xlsx')
    newname = name +'_aligned.csv'

    df = pd.read_excel(f)
    cols = df.columns
    print('\n',name)
    print(cols)

    df_cf = df[[cols[0]]]
    df_ref = df.drop(df.columns[0], axis=1)
    print(df_cf.columns)
    print(df_ref.columns)

    # NEW DATAFRAME

    df_final = pd.DataFrame(columns = ['CFNO','NAME','REFNO','DC','BC'])

    df = pd.read_excel(f)

    empty = []
    cols = df.columns
    orig_cols = cols
    c_cols = list(df[cols[0]])
    r_cols = list(df[cols[2]])

    for i in c_cols:
        if i not in r_cols:
            empty.append(i)

    final_list= [x for x in empty if ~np.isnan(x)]

    df = df.rename(columns={cols[0]: "CFNO", cols[1]: "NAME", cols[2]: "REFNO",
                            cols[3]: "DC", cols[4]: "BC"})

    df_cf = df[['CFNO']].copy()
    df_ref = df.drop(['CFNO'], axis=1)

    df_cf = df_cf.fillna(-1)
    df_ref = df_ref.fillna(-1)

    df_final = pd.DataFrame(columns = ['CFNO','NAME','REFNO','DC','BC'])
```

```
    cfno = list(df_cf.CFNO)

    for ind in cfno:

        if ind in final_list:
            print(ind,' in final list')
            row = {'CFNO':ind, 'NAME':np.nan, 'REFNO':np.nan, 'DC':np.nan, 'BC':np.nan}
            df_final = df_final.append(row, ignore_index=True)

        else:
            for j in range(len(df_ref)):
                val = df_ref['REFNO'].iloc[j]
                if ind == val:
                    row = {'CFNO':ind, 'NAME':df_ref['NAME'].iloc[j], 'REFNO':val,
                        'DC':df_ref['DC'].iloc[j], 'BC':df_ref['BC'].iloc[j]}
                    print(ind,val)
                    df_final = df_final.append(row, ignore_index=True)
                    break
                else:
                    j=j+1

    df_final = df_final.rename(columns = {'CFNO':name+'_CFNO', 'NAME':name+'_NAME',
'REFNO':name+'_REFNO',
                            'DC':name+'_DC','BC':name+'_BC'})
    df_final = df_final.replace(-1, '')
    df_final.to_csv(newname, index=False)
```

Note that a number of input files can be provided and the output will be the same name but with _aligned (refer code for idea) but the format of the input files should be as follows (else wont work):

CF_no	Residue	A_AFA	A_AFA_De	A_AFA_Betweeness
10	A/LEU	6	13.09	189
11	A/ASP	7	6.2	544
12	A/LYS	8	4.09	21
13	A/VAL	9	16.21	385
14	A/TYR	10	15.26	1622

Here the first column is the reference column. Now the third column will be aligned to the first, hence 10 which is in the 5th row will be move up and aligned to 10 on the first row, while this is done column 2,4and 5 will move with column 3 and align the values. Please note that the first column was acquired from consurf alignment and the other columns are acquired from cytoscape: where the amino acid number (column 3), amino acids name and chain (column 2), degree values for that row (column 4) and betweens

values for that row (column 5) are provided in this example. This code can be manipulated to align other different values according to user sepcification, the code only take into cosideration the column number and not the names.

Jupyter command to merge specific column of files in a folder:

```
import pandas as pd
import numpy as np
```

Extract all files

```
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)
```

Total files in folder:

```
len(files)
```

Merge all files in the folder

```
dfs = pd.read_csv(files[0])

for i in range(1,len(files)):
    dfn = pd.read_csv(files[i])
    dfs = pd.concat([dfs,dfn], axis=1)
dfs.to_csv('Merge_No_Urea.csv',index=False)
```

Check the document: display only the first 5 row

```
df1 = pd.read_csv('A_AFA_aligned.csv')
df2 = pd.read_csv('A_Anc_37_aligned.csv')
df1.head()
```

For moving only files ending with certain names:

```
import pandas as pd
import numpy as np
```

Extract all files

```
import os
import shutil
# folder path
dir_path = os.getcwd()
#outpath = 'C:\Users\mxk1003\OneDrive - University of Texas at Arlington\Desktop\Ankit\No_Urea\BC'
# list to store files
files = []
# Iterate directory
for file in os.listdir(dir_path):
    # check only text files
    if file.endswith('_DC.csv'):
        files.append(file)
print(files)
```

Merging only the ending with files

```
dfs = pd.read_csv(files[0])

for i in range(1,len(files)):
    dfn = pd.read_csv(files[i])
    dfs = pd.concat([dfs,dfn], axis=1)
dfs.to_csv('Merge_No_Urea_DC.csv',index=False)
```

Only extract some columns from files:

```
for f in files:

    name = f.rstrip('.xlsx')
    newname = name +'_aligned_BC.csv'

    df = pd.read_excel(f)
    cols = df.columns
    print('\n',name)
    print(cols)

    df_cf = df[[cols[0]]]
    df_ref = df.drop(df.columns[0], axis=1)
    print(df_cf.columns)
    print(df_ref.columns)

    # NEW DATAFRAME

    df_final = pd.DataFrame(columns = ['CFNO','NAME','REFNO','DC','BC'])

    df = pd.read_excel(f)

    empty = []
    cols = df.columns
    orig_cols = cols
```

```
    c_cols = list(df[cols[0]])
    r_cols = list(df[cols[2]])

    for i in c_cols:
        if i not in r_cols:
            empty.append(i)

    final_list= [x for x in empty if ~np.isnan(x)]

    df = df.rename(columns={cols[0]: "CFNO", cols[1]: "NAME", cols[2]: "REFNO",
                            cols[3]: "DC", cols[4]: "BC"})

    df_cf = df[['CFNO']].copy()
    df_ref = df.drop(['CFNO'], axis=1)

    df_cf = df_cf.fillna(-1)
    df_ref = df_ref.fillna(-1)

    df_final = pd.DataFrame(columns = ['CFNO','NAME','REFNO','DC','BC'])

    cfno = list(df_cf.CFNO)

    for ind in cfno:

        if ind in final_list:
            print(ind,' in final list')
            row = {'CFNO':ind, 'NAME':np.nan, 'REFNO':np.nan, 'DC':np.nan, 'BC':np.nan}
            df_final = df_final.append(row, ignore_index=True)

        else:
            for j in range(len(df_ref)):
                val = df_ref['REFNO'].iloc[j]
                if ind == val:
                    row = {'CFNO':ind, 'NAME':df_ref['NAME'].iloc[j], 'REFNO':val,
                        'DC':df_ref['DC'].iloc[j], 'BC':df_ref['BC'].iloc[j]}
                    print(ind,val)
                    df_final = df_final.append(row, ignore_index=True)
                    break
                else:
                    j=j+1

    df_final = df_final.rename(columns = {'CFNO':name+'_CFNO', 'NAME':name+'_NAME',
'REFNO':name+'_REFNO',
                            'DC':name+'_DC','BC':name+'_BC'})
    df_final = df_final.replace(-1, '')
    df_final = df_final.drop([name+'_CFNO',name+'_NAME',name+'_REFNO',name+'_DC'], axis =1)
    df_final.to_csv(newname, index=False)
```

Violin Plots, Joint plots and pair plots using jupyter notebook

Violin Plots:

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns

%matplotlib inline
```

Next cell

```
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)
```

assign variable name

```
Avg_loop_out = pd.read_csv('Avg_DC_BC_Loop_out.csv')
```

Violin plot side ways

```
sns.violinplot(x=Avg_all['Average_DC_No_urea'], y=Avg_all['short_SS'])
sns.set_context('paper', font_scale=3)
```

For figure:

```
plt.figure(figsize=(10,8))
sns.violinplot(x=Avg_all['short_SS'], y=Avg_all['Average_DC_No_urea'])
plt.yticks([-10,0,10, 20, 30,40,50,60, 70])
plt.ylabel("Degree")
plt.xlabel("Secondary Structure")
plt.title('Simulations in No Urea')
sns.set_context('paper', font_scale=1.8)
```

Adding a 3rd dimension (hue):

```
plt.figure(figsize=(10,8))
sns.violinplot(x=Avg_all['short_SS'], y=Avg_all['Average_DC_Loop_out' ], hue =Avg_all['Conservation'])
plt.yticks([-10,0,10, 20, 30,40,50,60, 70, 100])
plt.legend(loc=0)
plt.ylabel("Degree")
plt.xlabel("Secondary Structure")
plt.title('Simulations with Loop out')
```

```
sns.set_context('paper', font_scale=1.8)
```

For pair plots:

```
Avg_all = pd.read_csv('Avg_DC_BC_all.csv')
sns.pairplot(Avg_all, hue='short_SS')
```

Joint plots:

```
sns.jointplot(x='Average_BC_No_urea', y='Average_DC_No_urea',data=Avg_all,hue=
'Conservation',height=10,legend=False , s=100 )
plt.legend(loc=0)
plt.yticks([0,10, 20, 30,40,50,60, 70])

#sns.set_context('paper', font_scale=3)
```