EVOLUTIONARY AND MECHANISTIC STUDIES OF THE MULTIFUNCTIONAL INNATE IMMUNE PROTEIN S100A9

by

JOSEPH LESLIE HARMAN JR.

A DISSERTATION

Presented to the Department of Chemistry and Biochemistry
and the Graduate School of the University of Oregon
in partial fulfillment of the requirements
for the degree of
Doctor of Philosophy

June 2020

DISSERTATION APPROVAL PAGE

Student: Joseph Leslie Harman Jr.

Title: Evolutionary and Mechanistic Studies of the Multifunctional Innate Immune Protein S100A9

This dissertation has been accepted and approved in partial fulfillment of the requirements for the Doctor of Philosophy degree in the Department of Chemistry and Biochemistry by:

James Prell	Chairperson
Michael Harms	Advisor
Bradley Nolen	Core Member
Ken Prehoda	Core Member
Alice Barkan	Institutional Representative

and

Kate Mondloch	Interim Vice Provost and Dean of the Graduate School

Original approval signatures are on file with the University of Oregon Graduate School.

Degree awarded June 2020

DISSERTATION ABSTRACT

Joseph Leslie Harman Jr.

Doctor of Philosophy

Department of Chemistry and Biochemistry

June 2020

Title: Evolutionary and Mechanistic Studies of the Multifunctional Innate Immune Protein S100A9

Many questions remain regarding the molecular mechanisms by which proteins evolve new properties and functions. How do proteins evolve new functions without perturbing existing ones? How do changes in protein biophysics affect function? How are protein functions maintained, altered, or improved over time?

We used the multifunctional innate immune protein S100A9 (A9) as a model to dissect mechanisms of protein evolution. A9 performs two primary innate immune functions. As a homodimer, it potently stimulates inflammation by interacting with Toll-like receptor 4 (TLR4). As part of a heterocomplex with S100A8 (A8/A9), it is potently antimicrobial. A9 and the A8/A9 complex are further regulated by proteolysis: the A9 homodimer is highly proteolytically susceptible, while the A8/A9 heterocomplex is resistant. The evolutionary origins and mechanisms by which these functions arose are poorly understood, and the mechanism by which A9 activates TLR4 to drive inflammation is unknown.

We took an evolutionary biochemical approach to determine how A9 evolved its innate immune functions. Chapter I comprises an introduction. Chapter II

examines the role of pleiotropy in the evolution of A9 multifunctionality. We find that A9s gained proinflammatory activity and lost proteolytic resistance from a weakly proinflammatory, proteolytically resistant ancestral protein. A single ancient substitution had pleiotropic effects on A9 without affecting the A8/A9 complex, revealing a beneficial role for pleiotropy in the evolution of multifunctionality. Chapter III examines the biophysical mechanism by which A9 activates TLR4 to drive inflammation. We show that reverting the ancient substitution identified in chapter II compromises A9 activation of TLR4 by restricting access to a functionally necessary conformation of the protein. These findings highlight how subtle changes to a protein's conformational energy landscape can have critical impacts on protein evolution. In chapter IV, I outline ongoing work examining how later-diverging A9s have evolved more potent and promiscuous activation of TLR4 and determining how, mechanistically, A9 activates TLR4. This work provides novel insight into how a key innate immune protein evolved multifunctionality and highlights how basic changes in protein biophysics can have profound impacts on biological systems.

This dissertation includes both previously published/unpublished and co-authored material.

CURRICULUM VITAE

NAME OF AUTHOR: Joseph Leslie Harman Jr.

GRADUATE AND UNDERGRADUATE SCHOOLS ATTENDED:

University of Oregon, Eugene, OR
Willamette University, Salem, OR

DEGREES AWARDED:

Doctor of Philosophy, Chemistry, 2020, University of Oregon
Bachelor of Arts, Biochemistry, 2014, Willamette University

AREAS OF SPECIAL INTEREST:

Evolutionary Biology
Protein Engineering
Biophysics
Molecular Evolution

PROFESSIONAL EXPERIENCE:

PhD Candidate & Graduate Research Fellow, Harms Lab, University of Oregon, Eugene, OR, 2016-2020

PhD Student and Graduate Teaching Fellow, Department of Chemistry and Biochemistry, University of Oregon, Eugene, OR, 2014-2015

Undergraduate Researcher, Hoobler Lab, Department of Chemistry, Willamette University, Salem, OR, 2013-2014

Chemistry/Chemical Engineering Intern, Lawrence Berkeley National Laboratory, Berkeley, CA, Summer 2013.

Biochemistry Intern, Life Technologies, Eugene, OR, Summer 2012.

Biochemistry and Enology Intern, E&J Gallo Winery, Modesto, CA, Summer 2011.

GRANTS, AWARDS, AND HONORS:

National Institutes of Health, Genetics Training Grant, 2014-2017

Chemistry Department Distinguished Senior Award, Willamette University, 2014.

PUBLICATIONS:

Harman JL, Anderson JA, Reardon P, Warren GD, Costello SM, Connor P, Marqusee S, Harms MJ. "Biophysical studies of a conserved historical mutation reveal that S100A9 activates Toll-Like Receptor 4 through a non-native conformation." (in preparation)

Harman JL, Loes AN, Warren GD, Heaphy MC, Lampi KJ, Harms MJ (2020). "Evolution of multifunctionality through a pleiotropic substitution in the innate immune protein S100A9." *eLife*. 2020;9:e54100. DOI: 10.7554/eLife.54100

Harman JL (2014). "Synthesis of a Novel Pyridine-Substituted 7-Nitrobenz-(2,1,3)-Oxadiazol-4-yl (NBD) Fluorescent Compound to Monitor the Anti-Cancer Functionality of a NAMI-A Derivative." *Willamette University.* Thesis.

ACKNOWLEDGMENTS

I am sincerely grateful for the veritable village of mentors, colleagues, friends, and family who have helped me complete my PhD. I first thank my advisor, Michael Harms, for his constant mentorship and dedication to student success. I cannot speak highly enough of his intelligence, effort, and thoughtfulness as a scientist and mentor. I thank members of the Harms lab for years of mentorship and collaboration, thoughtful discussion and feedback, and friendship. Thank you to Andrea Loes, who was my first mentor in the Harms lab, provided a constant source of new and thoughtful ideas, and laid an excellent foundation for the work described in this dissertation. Thank you to Jeremy Anderson, an excellent mentor and friend who taught me protein NMR, spent many days commiserating with me in the cell culture room, and reminded me of the value of a good laugh. Thank you to previous lab members Luke Wheeler and Zach Sailer for mentorship and feedback early in my PhD and for staying in touch from afar through the Beer and Theory Society – a group that meets regularly to discuss scientific topics of interest outside of our immediate research. Thank you to Harms lab members Anneliese Morrison, Nick Frantz, and Michael Shavlik for friendship, feedback, and collegiality – I look forward to seeing where your work and interests take you. Thank you to Lauren Lehmann, Sophia Phillips, and Jon Muyskens for working with me to develop ongoing projects in the Harms lab, some of which are highlighted here. Thank you to the many excellent students who have worked in our lab – a special thanks to Maureen Heaphy and Gus Warren, who both contributed heavily to this work and were a pleasure to mentor.

Thank you to my committee members – Jim Prell, Brad Nolen, Ken Prehoda, and Alice Barkan – for constructive feedback and encouragement during my dissertation research. Thank you to the many graduate students, researchers, staff, and other members of my department for your kindness, collaboration, guidance, and support over the years.

Thanks are in order for several wonderful collaborators. Thank you to Patrick Reardon for enthusiastically contributing to protein NMR studies. Thank

you to Shawn Costello and Susan Marqusee for collaborating on mass spectrometry studies and for years of excellent discussions at various meetings – the Protein Folding Consortium meeting, in particular, shaped who I am as a scientist today. Thank you to members of the Bowman lab for help with molecular dynamics simulations and friendship. Thank you to the many labs and researchers with whom I've had wonderful discussions and interacted with over the years.

Thank you to previous mentors and educators at Willamette University for getting me excited about scientific research and encouraging me to pursue graduate studies – a special thanks to Sarah Kirk, Alison Fisher, Todd Silverstein, Chuck Williamson, and Tony Hoobler.

Thank you to my friends in Eugene and across the globe for your continued support and friendship. I particularly thank Berry Street and our local Eugene friends for many fun adventures, trips, constant encouragement, and regular laughs.

Thank you especially to my family. I thank my parents, Becky and Joe Harman, for their support, love, and encouragement over the years. Thank you to my siblings, Philip and Anna, and siblings-in-law, Andy, Trey, and Stephanie, for continued support and laughs. Thank you to my parents-in-law, Mary and Peter, for welcoming me into your family as one of your own. Thank you to my extended family for support and encouragement.

Most importantly, thank you to my wife, Lizzy. Thank you for your daily encouragement, kindness, and understanding. I truly couldn't have done this without you. Thank you for being my rock.

For my grandparents – James and Carol Harman and William and Judy Smith.

TABLE OF CONTENTS

Chapter Page

LIST OF FIGURES

Figure Page

Figure Page

Figure Page

CHAPTER I

INTRODUCTION

Organism-level evolution can be directly mapped to changes in protein biochemistry

Organisms have evolved to perform an incredible array of functions, some of which include growth,[1–3] movement,[4] conversion of nutrients to energy,[5,6] and defense against disease.[7,8] Underlying these macroscale processes are multitudes of biomolecules – including proteins, nucleic acids, and small molecules – that act in exquisite concert to produce organism-level functions. Proteins, in particular, mediate functions by performing essential biochemical processes that include catalysis,[9,10] molecular transport,[11,12] binding interactions,[13] and signaling.[2,3,14] Determining the mechanisms by which protein functions change over time is central to our understanding of how life as we know it has evolved.

Remarkably, the properties and functions of a protein are almost entirely encoded in its amino acid sequence.[15–17] The composition and ordering of amino acids within a protein dictates its chemistry, 3-dimensional structure, and biophysical properties, impacting protein function and ultimately organism-level outcomes.[18–20] Indeed, there are many well-established cases in which single amino acid changes within a protein can severely impact an organism. In humans, a single mutation in beta-hemoglobin – G7V – promotes hemoglobin aggregation and causes sickle cell anemia,[21] while mutations in the enzyme beta-hexosaminidase A

decrease the ability of the enzyme to break down fat in the nervous system, leading to Tay-Sachs disease.[22]

This physical reality of proteins – that amino acid changes within a protein can be directly mapped to functional outcomes in organisms – has profound implications for evolutionary biology. Evolutionary biologists have long sought mechanistic descriptions of how and why organisms have evolved various functions.[23,24] While much progress has been achieved in studying evolutionary processes at the level of both organismal traits and genetic composition,[2,3,25–28] comparatively less is known about the biochemical mechanisms by which proteins evolve new functions.[20,29–31] This gap in knowledge resulted in the creation of a relatively new field termed evolutionary biochemistry.[29,32–35] The primary goal of evolutionary biochemical research is to mechanistically describe evolution at the molecular level by combining the physical, mechanistic logic of biochemistry with the tools and reasoning of evolutionary biology.[29,32–34]

Evolutionary biochemistry has enabled mechanistic studies of protein evolution

Evolutionary biochemical studies over the last two decades have proven fruitful in determining the mechanisms by which a variety of protein functions have evolved. Early pioneering work combined phylogenetics,[36,37] ancestral sequence reconstruction,[38,39] and biochemical/functional characterization to examine ancient protein functions. These approaches – which are now hallmarks of evolutionary biochemistry – consist of constructing phylogenetic trees from alignments of

modern proteins,[37] estimating ancestral protein sequences within a phylogenetic tree using statistical methods (including maximum likelihood and Bayesian inference),[38–40] synthesizing reconstructed ancient proteins alongside their modern counterparts, and comparing them in functional experiments. Early studies using these techniques provided insight into ancient protein functions such as chromophore binding by ancestrally reconstructed archosaur rhodopsins,[41] diversification of hormone binding specificity in an ancestral steroid receptor,[42] and temperature-dependent GDP binding by an ancient bacterial EF-Tu protein.[43]

More recent evolutionary biochemical analyses have delved deeper into the biochemical and biophysical mechanisms underlying ancient changes in protein function. These studies often expand upon the techniques mentioned above by isolating functional transitions between ancestrally reconstructed proteins, identifying the key mutations that drove a functional transition, and using biophysical and biochemical techniques to determine how, mechanistically, key mutations cause changes in protein function. For example, a detailed mechanistic study of ancient steroid receptors revealed that key changes in receptor hormone specificity were mediated by two large-effect substitutions – out of 171 possible amino acid changes - that radically rewired the hydrogen bonding network and energetic landscape of the ancient steroid receptor.[44] A large-scale study of hemoglobins in birds adapted to living at higher altitudes found that while many avian hemoglobins convergently evolved tighter oxygen binding affinity, the mutations and molecular mechanisms by which this occurred often varied from species to species.[45] This study and many others highlight how introducing

mutations into different genetic backgrounds can have unpredictable outcomes – a phenomenon known as epistasis.[46] Understanding epistasis is a key area of evolutionary biochemical research, with many studies showing that epistasis profoundly shapes the evolutionary trajectories taken, or not taken, by proteins as they evolve.[47–49]

Modern evolutionary biochemical studies have made great progress in addressing key mechanistic questions in evolutionary biology. Studies examining the reversibility of protein functions have identified restrictive mutations that discourage reversion of protein function by physically or functionally restricting "backwards compatibility."[50,51] Permissive mutations along evolutionary trajectories have been identified that had no effect on protein function when they were acquired but were necessary for changes in function induced by later-occurring mutations.[52] Contingency in protein evolution has been elegantly examined using techniques such as deep mutational scanning on ancestrally reconstructed proteins, revealing that hundreds of different evolutionary trajectories and biochemical solutions are possible for a given protein function.[53,54] These studies and others not only provide unprecedented insight into mechanisms of protein evolution, but also facilitate an evolutionarily informed – and thus more targeted – analysis of the biophysical mechanisms underlying various protein functions.

How do proteins evolve multifunctionality?

The majority of studies in evolutionary biochemistry have focused on proteins with a single function. In contrast, few studies delve into the mechanisms by which proteins evolve multiple functions. Given the large number of multifunctional proteins present in biology, understanding how proteins evolve multifunctionality is central to our understanding of protein evolution and is a primary focus of this dissertation.

A paradox of multifunctional proteins is that while many exist in biology, it is generally believed that evolving multifunctionality is more challenging and under more constraint than evolving a single function.[55–57] Many studies in the field of protein engineering suggest that evolving multifunctionality is a problem of optimization. Adaptive constraint – or the complete incompatibility of two different functions in a protein simultaneously – has been elegantly demonstrated in studies of coenzyme use by β-isopropylmalate dehydrogenase (IMDH).[58] Constraints in enzyme evolution studies, such as stability-function tradeoffs, show that biophysical requirements constrain the evolution of proteins having even a single function, let alone multiple functions.[59–66] Finally, it is well-established that most mutations to a protein tend to be neutral or deleterious.[67–69] This suggests that multifunctionality is evolutionarily disfavored; the conditional probability that multiple mutations accumulate to produce multifunctionality without having a detrimental effect on a protein is much less likely than mutations accumulating to produce a single function.

Why, then, do so many multifunctional proteins exist in biology? While many studies suggest that evolving multifunctionality is challenging and subject to many constraints, these conclusions largely come from artificial directed evolution experiments and thus do not account for how multifunctional proteins evolve in nature. This raises a host of unanswered mechanistic questions. How do pre-existing functions in natural proteins constrain the acquisition of additional functions? How do natural proteins optimize multiple functions at once? What is the role of pleiotropy – defined as a single change that affects more than one function – in the evolution of protein multifunctionality? Experimental studies into how multifunctional proteins evolve, such as those presented in this dissertation, are necessary to address these questions.

Using evolutionary biochemistry to determine unknown mechanisms of protein functions

A central motivator of biological research is to determine how modern proteins perform their functions. We know the functions that many proteins perform in biology, but we often lack a mechanistic description of how a protein actually performs its function(s). Gaining a mechanistic understanding of how proteins function is critical for targeted drug design in human health,[70–73] rationally engineering proteins for desired functions,[74–78] and expanding our fundamental knowledge of biological systems.

Biochemists often determine unknown protein mechanisms by identifying the minimal set of residues that are important for function and then by

biochemically characterizing the mechanistic contributions of these residues to the function. A key challenge to this approach is that it can be difficult to determine which of the many protein residues to test. Modern mechanistic studies of proteins often require *a priori* knowledge of function from related proteins and/or large-scale mutagenesis screens that can be expensive, time-consuming, and difficult to interpret due to challenges such as epistatic interactions between mutations.

Evolutionary biochemical approaches such as phylogenetics and ancestral sequence reconstruction can alleviate some of these challenges because they estimate realized amino acid changes in proteins over their evolutionary history.[29] This powerfully constrains the search space of amino acids that are important for protein function, thus permitting more targeted studies of functionally relevant amino acid changes. By measuring the functions of just a few proteins in a phylogenetic tree, researchers can efficiently isolate intervals of functional change in a protein's evolutionary history and then rapidly identify functionally important and evolutionarily informed substitutions. This approach not only facilitates rapid dissection of mechanisms underlying protein function, but also provides valuable insight into mechanisms by which protein functions evolve.

S100A9 is a multifunctional protein in mammalian innate immunity

This dissertation seeks to examine outstanding topics in evolutionary biochemistry, including how proteins evolve multifunctionality and the use of evolutionary biochemical approaches in determining unknown mechanisms of protein function. To address these questions, we used the multifunctional innate

immune protein S100A9 (A9) as a model system. The following section details aspects of A9 biology that are pertinent to this work.

A9 and closely related protein S100A8 (A8) are two of the most abundant proteins in neutrophils[79,80] and are primary biomarkers for a variety of inflammatory diseases.[81–85] A9 and A8 are released by neutrophils at sites of infection as part of the innate immune system's first line of defense against pathogens,[86–90] where they perform several critical innate immune functions.

A9 is released into the extracellular space as both a homodimer and as a heterodimeric complex with A8 (the A8/A9 complex – often called calprotectin).[86,87,89,90] These two states of A9 have distinct innate immune functions. The A9 homodimer potently drives inflammation by directly interacting with proinflammatory cell surface receptor Toll-like receptor 4 (TLR4).[91–95] The A8/A9 heterocomplex exerts antimicrobial activity by sequestering transition metals away from microbes with extremely high affinity.[96–104] Finally, an additional layer of regulation is that A9 and the A8/A9 complex are differentially degraded by proteases; A9 is highly susceptible to proteolytic degradation, while the A8/A9 complex is resistant.[105–108] The differential proteolytic susceptibility of A9 and the A8/A9 complex is thought to regulate the relative abundances of these two states of A9 at sites of infection and has been proposed to play a functional role in A9 proinflammatory activity.[109]

The mechanism by which the A8/A9 complex exerts antimicrobial activity is well-established. Six histidines – four from A9 and two from A8 – form a biologically rare hexahistidine metal coordination site at the interface between the

two proteins that binds a variety of transition metals with extremely high affinity.[98,100–102,104,110] This metal binding site is unique to the A8/A9 complex within the S100 protein family. Most S100 proteins – including A8 and A9 – bind transition metals with varying affinities via a His_3Asp site that is largely conserved across the S100s.[111–113] While a small portion of the antimicrobial activity of the A8/A9 complex is attributed to the His_3Asp site, the distinct, broad-spectrum antimicrobial potency of the A8/A9 complex is largely due to the broad metal specificity and extremely high binding affinity of the hexahistidine site.[99,102–104]

The mechanism by which A9 activates TLR4 to drive inflammation, in contrast, is poorly understood. TLR4 activation by exogenous molecules, such as the bacterial cell wall component lipopolysaccharide (LPS), is mechanistically well-studied.[114,115] LPS is bound by the TLR4 cofactor CD14, which delivers LPS to a second cofactor, MD2.[116–118] Binding of LPS to a hydrophobic pocket within MD2 promotes dimerization of the TLR4/MD2 complex to form $(TLR4/MD2)_2$.[114,115,119] LPS-induced formation of the $(TLR4/MD2)_2$ complex leads to activation of the highly conserved Myd88 and TRIF/TRAM pathways, ultimately resulting in activation of the transcription factor NF-κB.[120–122] NF-κB activation drives cytokine production that produces an inflammatory cascade and recruits additional neutrophils to the site of infection.[123–125] LPS-induced activation of TLR4 is a primary driver of septic shock, making it a key therapeutic target.[126–128]

Few studies, in contrast, have examined the mechanism by which A9 activates TLR4 to drive inflammation. It was previously suggested that A9 binding

to zinc is required for TLR4 activation based on *in vitro* binding studies.[129] This theory was later disproven in a study where ablating zinc binding in A9 did not alter A9 activation of TLR4.[111] The disordered C-terminal extension of A9 (residues ~95-114) has also been shown to be dispensable for A9 activation of TLR4.[111] It is established that TLR4 cofactors MD2 and CD14 are both required for A9 activation of TLR4 across mammals,[130] suggesting a possible shared mechanism of activation with LPS. However, A9 is much larger than LPS and is simply too big to fit within the MD2 hydrophobic binding pocket in its native homodimeric state. It has been suggested that proteolytic fragments of A9 might activate TLR4,[109] as A9 is highly susceptible to proteolytic degradation and the extracellular space is rich in proteases. Indeed, proteolytic fragments of A9 have been described to modestly activate TLR4, although it is unclear whether this is the primary mode by which A9 activates TLR4. Previous work has identified pairs of point mutations to charged residues within A9 that disrupt A9 binding to TLR4 *in vitro*, leading to a proposed *in silico* docking model of A9 to TLR4.[109] However, the effect of these mutations on A9 activation of TLR4 were not characterized. Similar to LPS, A9 is a desirable therapeutic target due to its potent proinflammatory activity. Selectively inhibiting A9 activation of TLR4 could be an effective therapeutic strategy for reducing host-induced inflammation while retaining TLR4 sensing of exogenous danger signals like LPS. Designing such a strategy, however, is hampered by not knowing the molecular mechanism by which A9 activates TLR4. A key focus of this dissertation is using evolutionary biochemical approaches to expand our mechanistic understanding of how A9 activates TLR4 to drive inflammation.

Lastly, the role of proteolytic regulation in A9 biology is poorly understood. Various roles, both functional and regulatory, have been proposed for the differential proteolytic susceptibility of A9 and the A8/A9 complex. Proteolysis could selectively remove proinflammatory A9 from sites of inflammation, thus enriching for the antimicrobial A8/A9 complex.[105,107] It has also been suggested that proteolytic degradation of A9 could be required for proinflammatory activity, as proteolytic fragments of A9 are capable of activating TLR4.[109] Determining the mechanistic connection between A9 proteolytic regulation and the other innate immune functions of A9 is necessary for understanding how A9 balances multiple functional roles in mammalian innate immunity.

Evolutionary studies of S100A9 proinflammatory activity

Previous studies of A9 evolution have laid the groundwork for determining how A9 evolved multifunctionality and for examining the mechanism by which A9 activates TLR4. Phylogenetic analyses of A9 and closely related proteins revealed that A9s share a common amniote ancestral protein with three other S100 proteins – A8s, S100A12s (A12s), and MRP126s.[112,130] These four protein clades – which together form a group of S100 proteins called the calgranulins – form a polytomy, meaning that it is unclear exactly when each of these four clades diverged. However, MRP126s are only present in sauropsids (birds and reptiles), while A8s, A9s, and A12s are only present in mammals, indicating that the sauropsid MRP126 proteins are likely co-orthologous to the mammalian A8, A9, and A12 proteins.[130]

Previous work determined that A9s from multiple mammals (human, mouse, and opossum), as well as the chicken MRP126 protein, are each able to activate their respective species-matched TLR4s.[130] CD14 and MD2 were shown to be required for A9 and chicken MRP126 activation of TLR4.[130] Complementation experiments – in which TLR4 components from different species were swapped out and tested for activity – revealed that while CD14s from most species could complement TLR4 activation by most A9s, MD2s tended to only complement more closely related TLR4s and thus exhibited stronger species-specific co-evolution. Further, while A9s and LPS exhibited largely similar TLR4/MD2/CD14 requirements in complementation experiments, some differences were observed that suggest potential mechanistic differences in TLR4 activation between A9 and LPS.[130] These findings formed the foundation for much of the work that is presented in this dissertation.

Chapter-by-chapter breakdown of dissertation

Chapter II describes how A9 evolved multifunctionality, examines the evolutionary and functional interplay between A9 functions, and highlights a pleiotropic substitution that played a key role in A9 evolution.[131] We used phylogenetics, ancestral sequence reconstruction, and functional characterization to show that A9s and A8s evolved to form antimicrobial, proteolytically resistant heterocomplexes early in mammals. We found that A9 homodimers evolved potent proinflammatory activity and lost proteolytic resistance from a proteolytically resistant, weakly proinflammatory amniote ancestor. We identified a single key

substitution that occurred in the mammalian A9 ancestor and played key roles in the evolution of A9 multifunctionality. Reverting this substitution to its amniote ancestral state in human A9 rendered the protein proteolytically resistant and strongly decreased its proinflammatory activity, while introducing the modern substitution into the A9 amniote ancestor increased its proinflammatory activity. The substitution had no effect on the antimicrobial activity of the A8/A9 complex. Lastly, we showed that proteolytic degradation is not required for A9 activation of TLR4. These findings together show that A9 evolved multifunctionality by partitioning innate immune functions between the A9 homodimer and the A8/A9 complex. A key substitution had pleiotropic effects on the A9 homodimer – increasing its proinflammatory activity and driving a loss of proteolytic resistance – without affecting the antimicrobial activity or proteolytic resistance of the A8/A9 complex. This reveals that pleiotropy can play a beneficial role in the evolution of multifunctionality by selectively altering one protein functional state without affecting another. We propose that given the large number of proteins that have interaction partners and/or occupy multiple functional states, this is likely a common route by which multifunctional proteins arise in biology.

Chapter III consists of an in-depth biophysical analysis of the mechanism by which the historical mutation found in Chapter II decreases A9 proinflammatory activity. The goal of this study was to gain insight into the mechanism by which A9 activates TLR4 to drive inflammation. Thermodynamic studies of A9 and the A9 mutant – A9 M63F – revealed that the M63F mutation specifically stabilizes the calcium-bound form of A9, but not the calcium-free state. Structural

characterization of A9 M63F suggests that it does not significantly alter A9 structure. Given the lack of obvious structural changes, we hypothesized that the M63F mutation might alter some functionally important dynamic process of the protein. We observed no changes in dynamics on short timescales by NMR, but in long timescale hydrogen-deuterium exchange studies we observed significant local stabilization of A9 M63F compared to wildtype A9. This analysis enabled identification of a second critical amino acid position – F37 – that is well-positioned to form a pi-stacking interaction with the M63F mutation. Reverting F37 to its mammalian ancestral state in the A9 M63F background (A9 M63F F37L) negates the stabilizing effect of M63F and fully restores A9 proinflammatory activity. These findings show that M63F decreases A9 activation of TLR4 by locally stabilizing the protein via a direct interaction with a nearby residue. Given M63F does not appear to alter the native A9 protein structure, we propose that M63F inhibits access to an excited state conformation of A9 that is necessary for proinflammatory activation of TLR4. These findings provide unprecedented insight into the mechanism by which a key innate immune protein drives inflammation in mammals and highlight the utility of evolutionary biochemical analyses in dissecting poorly understood protein functions.

Chapter IV highlights ongoing studies examining how later-diverging A9 proteins have evolved proinflammatory activity and determining the mechanism by which A9 activates TLR4 to drive inflammation. We previously found that A9s evolved enhanced proinflammatory activity early in mammals. However, these studies revealed distinct differences in both A9 proinflammatory potency and cross-

reactivity with TLR4s from different species. In particular, human A9 potently activates TLR4s from various species better than species-matched A9s activate their respective TLR4s, and human TLR4 is much more potently activated by human A9 than by any other A9 protein. Further, the magnitude of A9 proinflammatory activity is often higher in later-diverging A9s. These findings suggest that A9 activation of TLR4 has been enhanced in later-diverging species. We also hypothesize that certain A9-TLR4 pairs, such as human A9 and human TLR4, have evolved increased specificity. This chapter highlights ongoing evolutionary biochemical analyses to examine these possibilities, including testing the proinflammatory activities of uncharacterized A9s from later diverging mammals such as elephants and identifying key amino acid substitutions that mediated changes in proinflammatory potency and TLR4 specificity.

Finally, I conclude chapter IV by describing ongoing orthogonal approaches for determining the mechanism by which A9 activates TLR4 to drive inflammation. We have implemented a high-throughput cell sorting and sequencing method for screening mutations in TLR4, MD2, and CD14 that alter LPS and/or A9 activation of TLR4. This unbiased approach complements previous evolutionary and mechanistic studies of A9 proinflammatory activity. We highlight how this approach is currently being employed to determine the A9:TLR4/MD2/CD14 interaction interface. We conclude by describing future characterizations of modern and ancestrally reconstructed A9s, TLR4s, MD2s, and CD14s from various species to examine mechanistic similarities and differences in A9 activation of TLR4 across mammals.

CHAPTER II

EVOLUTION OF MULTIFUNCTIONALITY THROUGH A PLEIOTROPIC SUBSTITUTION IN THE INNATE IMMUNE PROTEIN S100A9

AUTHOR CONTRIBUTIONS

Joseph Harman and Michael Harms conceptualized the study and designed the experiments. Michael Harms acquired the funding for the study. Joseph Harman, Gus Warren, and Maureen Heaphy performed the experiments. Michael Harms administered the project. Andrea Loes conducted phylogenetic analyses. Kirsten Lampi assisted in light scattering measurements. Joseph Harman analyzed the data and constructed figures. Joseph Harman and Michael Harms wrote the manuscript. All authors read and approved the manuscript.

INTRODUCTION

The innate immune system uses a small number of multifunctional proteins to respond to diverse immune challenges. Multifunctional immune proteins are critical for pathogen defense,[132–134] shaping host-associated microbial communities,[135] and well-regulated tissue growth.[136–138] They also drive pathological inflammation in disease, including autoimmune disorders, cancer, and cardiovascular disease.[96,139–142] These multifunctional proteins raise both mechanistic and evolutionary questions. How can one protein sequence satisfy the multiple constraints imposed by having multiple functions? How can multiple

functions evolve in one protein when, as a result of multifunctionality, each mutation likely has pleiotropic effects?[143–146]

One such multifunctional protein is S100A9 (A9), a small, soluble protein found at high concentrations in the extracellular space during an inflammatory response.[147] It has at least two key immune functions. As a homodimer, A9 potently activates inflammation via Toll-like receptor 4 (TLR4).[86,91–95,129,148–153] As a heterocomplex with S100A8 (A8/A9, also known as calprotectin), it is antimicrobial (Figure 2.1a).[97–104,110,154–158] A9 exacerbates endotoxin-induced shock in mice.[159] Both A9 and A8/A9 are primary biomarkers for many human inflammatory diseases.[81–85] Further, dysregulation of A9 is associated with various cancers, pulmonary disorders, and Alzheimer's disease.[81–85,93,160,161] Understanding the mechanisms by which A9 performs its innate immune functions is critical for developing treatments for A9-mediated diseases.

The mechanism of A8/A9 antimicrobial activity is well established: it sequesters a variety of transition metals through both a hexahistidine site and a His_3Asp site formed at the A8/A9 heterodimer interface, thereby limiting the concentrations of essential microbial nutrients in the extracellular space.[97–104,110,154–158] Other S100 proteins exert weaker antimicrobial activity via the His_3Asp site, which has lower metal binding affinity and binds fewer types of transition metals than the A8/A9 hexahistidine site.[101–103,110,156,162,163] In contrast, the proinflammatory mechanism of A9 is not well understood. A9 acts as a Damage-Associated Molecular Pattern (DAMP), activating NF-κB and other cytokines through Toll-like receptor 4 (TLR4).[86,91–95,129,148–153] The interaction

interface(s), affinity, and stoichiometry for the A9/TLR4 interaction are not known. A small region of A9 has been suggested to form part of the A9/TLR4 binding surface,[109] but no mutant of A9 has been identified that substantially compromises its activation of TLR4.

An additional layer of A9 immune function is that A9 and A8/A9 are thought to be regulated in the extracellular milieu by proteases. Neutrophils release multiple proteases along with A9 at sites of inflammation that regulate the inflammatory response.[164–169] A9 is very susceptible to proteolytic degradation, while A8/A9 is highly resistant (Figure 2.1a).[105,107] Proteolysis may serve to purge proinflammatory A9 from sites of inflammation and thus selectively enrich for antimicrobial A8/A9. There may even be a direct, functional link between A9 proteolytic degradation and inflammation. Proteolytic fragments of A9 are sufficient to activate TLR4,[109] and proinflammatory stimuli are thought to stabilize A9 homodimers against proteolytic degradation.[105] Directly testing the relationship between A9 proteolytic susceptibility and proinflammatory activity, however, has been challenging. There is no obvious way to selectively increase the proteolytic resistance of A9 and test its effect on A9 activation of TLR4, making it difficult to understand the relationship, if any, between these two functions.

We took an evolutionary biochemical approach to mechanistically dissect the evolution of A9 innate immune functions. Using phylogenetics, ancestral sequence reconstruction (ASR), and biochemical studies, we show that A9s evolved to form proteolytically resistant, antimicrobial A8/A9 complexes in early mammals. We find that A9 homodimers gained proinflammatory activity and lost

proteolytic resistance in the ancestor of therian mammals from a weakly proinflammatory, proteolytically resistant amniote ancestor. We identify a pleiotropic substitution that is necessary for A9 activation of TLR4, sufficient to increase TLR4 activation by the A9 amniote ancestor and played a role in loss of A9 proteolytic resistance. Mutating this site has minimal effect on A8/A9 antimicrobial activity or proteolytic resistance. Lastly, we show that proteolysis is not required for A9 activation of TLR4. Taken together, this work reveals that mammals concomitantly evolved A8/A9 antimicrobial activity, A9 proinflammatory activity, and a way to selectively regulate A9 inflammation via loss of A9 proteolytic resistance. These findings provide unprecedented mechanistic and evolutionary insight into A9 function and show how a single mutation can have pleiotropic effects in one functional state of a protein while not impacting another, thus facilitating the evolution of multifunctionality.

RESULTS

We first set out to establish when A9 evolved three innate immune properties: antimicrobial activity via formation of the A8/A9 complex, proinflammatory activation of TLR4 by A9 alone, and the differential proteolytic susceptibility of A9 and A8/A9.

A9s evolved to form antimicrobial A8/A9 complexes early in mammals

We sought to determine when A9 evolved to form the antimicrobial A8/A9 complex. We hypothesized that A8/A9 antimicrobial activity evolved in the

ancestor of therian mammals (the shared ancestor of marsupials and placental mammals) for several reasons. First, the broad-spectrum antimicrobial activity of human and mouse A8/A9 is well established.[97–104,110,155–158] Second, A9 and A8 genes are only found together in therian mammals (Figure 2.1b);[130] therefore the A8/A9 complex could not have arisen earlier than in the ancestor of therian mammals. Lastly, the residues composing the antimicrobial hexahistidine metal binding site are fully conserved across therian mammals (Figure AA1).

To determine whether the antimicrobial A8/A9 complex arose in the ancestor of therian mammals, we compared human A8/A9 to two previously uncharacterized A8/A9 complexes. We first tested the antimicrobial activity of A8/A9 from opossum, which is one of the earliest-diverging mammals relative to humans that possesses both of the S100A8 and S100A9 genes. Opossum and human A8 and opossum and human A9 have sequence identities of approximately 50%, respectively (Figure AA1). Following previous work,[104] we produced a cysteine-free variant of the complex to avoid the use of reducing agents in the antimicrobial assay. We confirmed that cysteine-free opossum A8/A9 formed a heterotetramer (46.8 ± 0.7 kDa) in the presence of calcium – like the human and mouse proteins[170] – using size exclusion chromatography coupled with multi-angle laser light scattering (SEC MALS, Figure AA2).

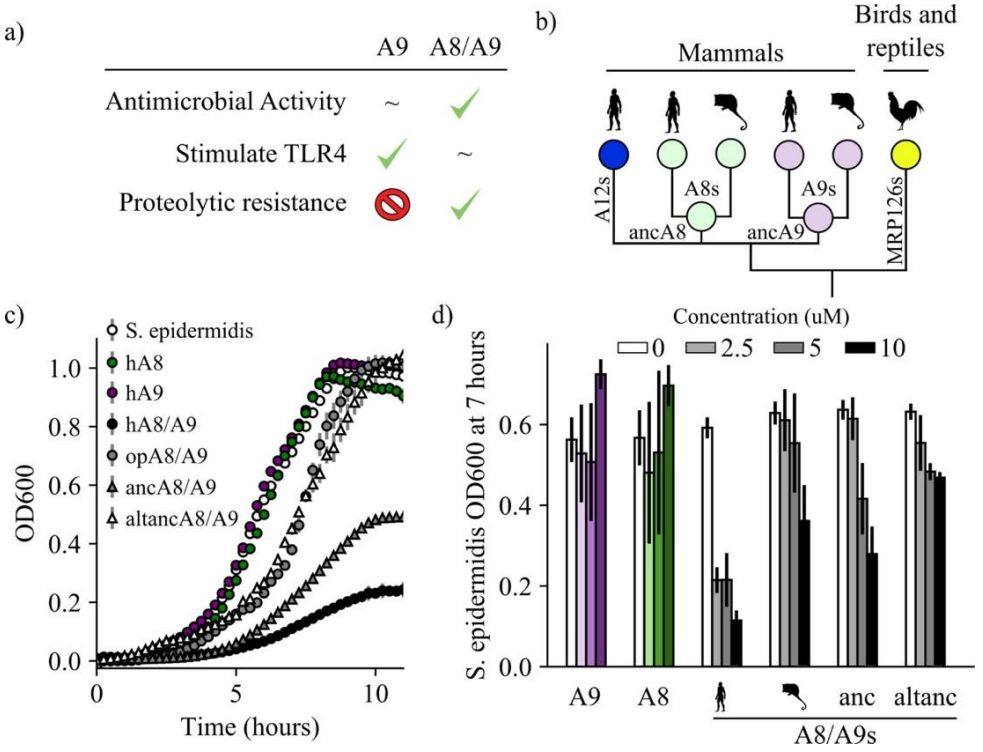


Figure 2.1. A9s evolved to form the antimicrobial A8/A9 complex early in mammals. (a) Table of A9 and A8/A9 properties. "~" represents weak or ambiguously characterized function, check marks and red "X" represent confirmed property (check) or lack thereof ("x"). (b) Schematic of previously published S100 protein tree. Colored nodes represent single protein sequences. Species cartoons shown are human, opossum, and chicken. Ancestrally reconstructed protein nodes are labeled. Branch lengths not to scale. (c) Representative growth curves for *Staphylococcus epidermidis* in the presence or absence of 10 μM S100 proteins. Each point represents optical density at 600 nm. *S. epidermidis* growth alone and in the presence of modern proteins are shown as circles, growth in the presence of ancestrally reconstructed proteins shown as triangles. Error bars are standard deviation of three technical replicates. (d) Percent of untreated *S. epidermidis* growth at 12 hours with S100 protein treatments. Data are average of three biological replicates. Error bars are standard error of the mean. Species cartoon labels are the same as in (b).

We measured cysteine-free opossum A8/A9 antimicrobial activity against a representative gram-negative bacterium, *Staphylococcus epidermidis*. In studies of A8/A9 from other species, activity against *S. epidermidis* tracked with the broad-

spectrum antimicrobial activity of the complex.[97] We assayed activity using a previously established *in vitro* antimicrobial assay that monitors bacterial growth in the absence or presence of S100 proteins (Figures 2.1c, AA3).[104] compare the activity of different proteins, we quantified inhibition at seven hours (Figure 2.1d). We observed a dose-dependent decrease in *S. epidermidis* growth in the presence of low micromolar concentrations of cysteine-free opossum A8/A9 (Figures 2.1c-d, AA3). The antimicrobial activity of opossum A8/A9 was weaker than that of human A8/A9: opossum A8/A9 delayed bacterial growth, while human A8/A9 both delayed growth and decreased bacterial carrying capacity (Figure 2.1c). It was previously found that cysteine-free human A8/A9 was potently antimicrobial,[104] while cysteine-free mouse A8/A9 exhibits weaker antimicrobial activity than wildtype mouse A8/A9[97] To determine whether the weaker activity of opossum A8/A9 was due to the removal of cysteines, we also measured the activity of wildtype opossum A8/A9 against *S. epidermidis*. We found that wildtype opossum A8/A9 had higher activity than cysteine-free opossum A8/A9 over the concentration range tested (Figure AA3). This suggests that the cysteines present in mouse and opossum A8/A9 play a role in their antimicrobial activity, unlike in human A8/A9. The antimicrobial activity of the opossum A8/A9 complex thus appears to be more similar to that of mouse A8/A9 than human A8/A9.

The shared antimicrobial activity of human, mouse, and opossum A8/A9 strongly suggests that the antimicrobial A8/A9 complex evolved in the ancestor of therian mammals. To test this further, we measured the ntimicrobial activity of ancestrally reconstructed therian mammalian A8/A9 (ancA8/A9 – Figure 2.1b). We

used our previously published phylogenetic tree[130] consisting of 172 S100 sequences to reconstruct therian mammalian ancestral A8 and A9 (ancA9 and ancA8 – Figure 2.1), which were used to form the ancA8/A9 complex. AncA8 and ancA9 had average posterior probabilities of 0.88 and 0.83, with sequence similarities to human A8 and A9 of 66% and 64%, respectively (Figure AA4). Average posterior probabilities in this range have been previously described as medium confidence reconstructions, with reconstructions characterized by others having average posterior probabilities as low as 0.7.[171] We confirmed that each protein was folded and had secondary structure content similar to that of human A8/A9 using far-UV circular dichroism (CD) spectroscopy (Figure AA5).

We then measured the antimicrobial activity of ancA8/A9 against *S. epidermidis*. We observed a potent reduction in *S. epidermidis* growth comparable to that of human A8/A9 (Figures 2.1c-d, AA3). To test for the robustness of this finding to phylogenetic uncertainty, we also tested the antimicrobial activity of an AltAll[171] reconstruction of ancA8/A9 against *S. epidermidis* (altancA8/A9, Figures 2.1c-d, AA4). In this reconstruction, we swapped all ambiguously reconstructed amino acid positions for their second-most likely state (see methods). AncA8/A9 and altancA8/A9 differ by 27 amino acids total (10 between ancA8 and altancA8 and 17 between ancA9 and altancA9 - Figure AA4). AltancA8/A9 exhibited antimicrobial activity against *S. epidermidis* similar to opossum A8/A9: it delayed growth but did not ultimately limit bacterial carrying capacity. While the hexahistidine site residues are conserved in ancA8/A9 and altancA8/A9 (Figure AA1), it appears that a subset of the ambiguously reconstructed 27 residues are

important for A8/A9 antimicrobial activity, perhaps affecting the orientation and/or affinity of the hexahistidine metal binding site.

Taken together, the antimicrobial activity of modern mammalian A8/A9 complexes (human, mouse, and opossum) and the antimicrobial activity of the reconstructed ancA8/A9 complex suggest that A9s evolved to form the antimicrobial A8/A9 complex in the ancestor of mammals.

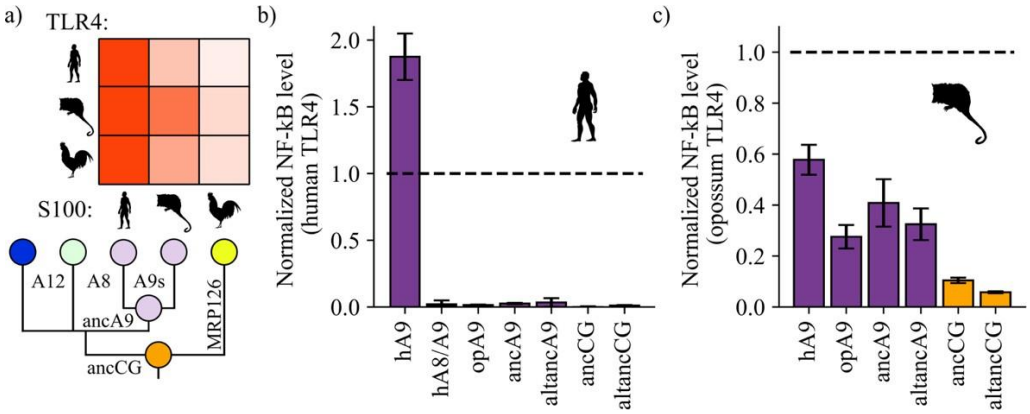


Figure 2.2. A9s gained proinflammatory activity from a weakly proinflammatory ancestor. (a) Schematic of previously measured proinflammatory activity of S100s against various TLR4s. Species labels on x and y-axes of heatmap are the same as Figure 2.1. Heatmap coloring is scaled to match 2 μM S100 activity levels measured in supplementary figure S2 of Loes et al. 2018.[130] (b) and (c) NF-κB production of human and opossum TLR4 in response to treatment with modern and ancestral S100 proteins. Bars represent average of >3 biological replicates, error bars are standard error of the mean. All values are background-subtracted and normalized to LPS positive control (see methods).

A9s evolved potent proinflammatory activity from a weakly active amniote ancestor

We next sought to determine when A9s evolved potent proinflammatory activity via activation of TLR4. Our previous work revealed that human A9 potently activates not only human TLR4 in functional assays, but also opossum and chicken

TLR4 (Figure 2.2a).[130] In contrast, chicken MRP126, the sauropsid ortholog of A9s, was found to be a weak activator of all TLR4s, including chicken TLR4. Both human and opossum A9 activate chicken TLR4 better than chicken MRP126 does. Two possibilities are consistent with these observations. Either mammalian A9s evolved enhanced proinflammatory activity from a less active amniote ancestral state, or A9s maintained a potent ancestral activity that was lost by chicken MRP126.

To differentiate between these two possibilities, we determined the ancestral proinflammatory activity of these proteins. We used ASR to reconstruct the shared amniote ancestor of A9s, A8s, A12s, and MRP126s. This group of proteins is known collectively as the "calgranulins", so we will refer to this ancestral protein as ancCG (ancestor of calgranulins). We also constructed an alternate, "alt All" version of this ancestor (altancCG, S4), which differed from ancCG by 8 amino acids. The average posterior probability of ancCG was 0.86 (table S4). We also expressed and purified ancA9 and altancA9 – the A9 subunits from the ancestral A8/A9 complexes described above. We confirmed that each protein was folded and had secondary structure content similar to that of modern S100s using far-UV CD spectroscopy (Figure AA5).

We then tested modern and ancestral S100s for activity against human TLR4. Following previous work,[130,159,172,173] we transiently transfected HEK293T cells with plasmids encoding TLR4 and its species-matched cofactors MD-2 and CD14, added purified S100 proteins to the growth media, and then measured output of luciferase under control of an NF-κB promoter. Consistent with previous

results,[130] we found that human A9 potently activated human TLR4, resulting in high levels of NF-κB production (Figures 2.2b, AA6). Human A8/A9 and opossum A9 exhibited much weaker activity against human TLR4. Lastly, we tested ancA9, altancA9, ancCG, and altancCG for activity against human TLR4 and observed weak or no activation for each ancestral protein. This result is unsurprising, as we previously found that human TLR4 is more specific than other amniote TLR4s: human TLR4 is activated much more potently by human A9 than by any other S100 protein (Figure 2.2a-b).[130] In contrast, TLR4s from other species (mouse, opossum, and chicken) appear to be more promiscuous and can be activated similarly by S100s from various species (Figure 2.2a).[130] This is consistent with lineage-specific coevolution between human TLR4 and human A9 – a confounding variable that makes assessment of ancestral S100 protein proinflammatory activity difficult using human TLR4.

We predicted that opossum TLR4 would be a better protein to probe ancestral S100 proinflammatory function because opossum TLR4 is broadly activated by A9s across mammals and gives little indication of lineage-specific coevolution.[130] We therefore tested the proinflammatory activity of ancA9, ancCG, and their corresponding alternate reconstructions against opossum TLR4. Corroborating previous results, human A9 strongly activated opossum TLR4, while opossum A9 activity was approximately half that of human A9 (Figures 2.2c, AA6-8). AncA9 and altancA9 activated opossum TLR4 to the same extent as opossum A9. AncCG and altancCG were the weakest activators of opossum TLR4, with activity approximately 25% or less than that of human A9. These findings suggest

that A9s evolved enhanced proinflammatory activity early in mammals from a weakly proinflammatory amniote ancestor, while A8/A9s and chicken MRP126 maintained weak, ancestral proinflammatory activity.

A9s evolved proteolytic susceptibility from a proteolytically resistant amniote ancestor

We next sought to determine when the differential proteolytic susceptibility of A9 and A8/A9 evolved. We used a simple *in vitro* assay to monitor S100 protein degradation over time in the presence of proteinase K, a potent non-specific serine protease (Figure 2.3a). Proteinase K was chosen both because of its low specificity and to mimic other serine proteases that A9 and A8/A9 encounter when released from neutrophils during an inflammatory response.[164–167,174] Proteolytic decay rates were estimated by fitting a single exponential decay function to the data (Figures 2.3b, AA9-12).

Human A8/A9 has been described as extremely resistant to proteases;[107] however, it has not been compared to S100 proteins besides human A8 and A9. To establish a baseline expectation for S100 protein proteolytic resistance, we characterized the proteolytic resistance of a broad set of human S100s against proteinase K. As previously shown,[107] human A9 and A8 alone were rapidly proteolytically degraded, while the human A8/A9 complex exhibited strong resistance (Figures 2.3c, AA9-10). Under our conditions, the degradation rates for human A8 and A9 were approximately three orders of magnitude faster than that of the human A8/A9 heterocomplex. We then characterized closely related protein

human S100A12 (A12), the chicken ortholog MRP126, and six distantly related human S100s.[112] Human A12, chicken MRP126, and five out of six more distantly related human S100s exhibited intermediate to strong proteolytic resistance, each degrading 1-2 orders of magnitude slower than human A8 or A9 but, on average, one order of magnitude faster than human A8/A9 (Figures 2.3c, AA9-10). Notably, human A12 and chicken MRP126 formed predominantly homodimers by SEC MALS under these conditions (Figure AA2), indicating that higher-order oligomerization (> 2 subunits) isn't required for S100 proteolytic resistance. Lastly, human A14 degraded faster than A9 or A8. This protein is evolutionarily distant[112] and therefore likely reflects independent evolution of this property. Taken together, these data show that the A8/A9 complex, A9, and A8 indeed fall at the extremes of human S100 proteolytic resistance; human A9 and A8 are among the fastest-degrading S100s tested, while human A8/A9 is one of the slowest.

To test whether A9 and A8 proteolytic susceptibility and A8/A9 resistance are conserved across mammals, we characterized mouse and opossum A9, A8, and A8/A9 for proteolytic resistance. Mouse A9 and A8 were found to be highly proteolytically susceptible and mouse A8/A9 strongly proteolytically resistant, matching the pattern observed for their human counterparts (Figures 2.3c, AA10). Opossum A9 and A8 were also highly proteolytically susceptible, while opossum A8/A9 was resistant (Figures 2.3c, AA10). This indicates that the susceptibility of A9s and A8s and the resistance of A8/A9 complexes is conserved across mammals.

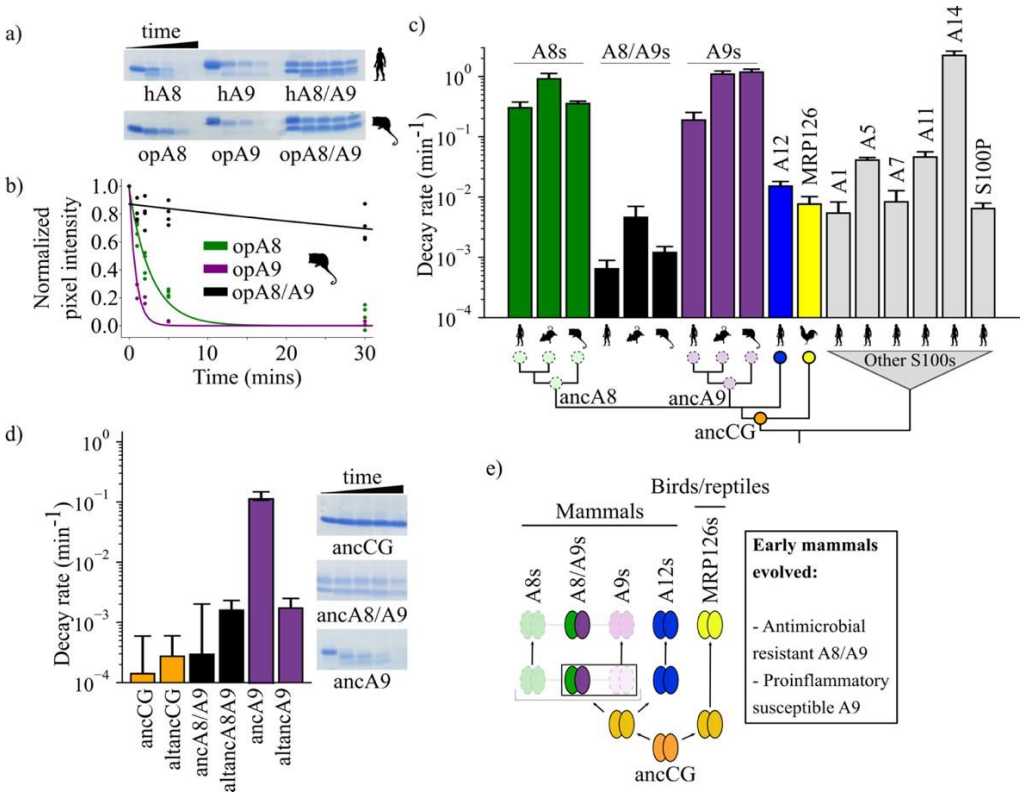


Figure 2.3. A9s lost proteolytic resistance from a proteolytically resistant amniote ancestor. (a) *In vitro* proteolytic resistance assay showing SDS-PAGE gel of S100 protein degradation via proteinase K over time. Gels were quantified using densitometry and normalized to the undigested protein band intensity. (b) A single exponential decay model was globally fit to the data to quantify decay rates. Points are biological replicates, lines are model fit to data. (c) S100 protein proteolysis rates mapped onto schematized S100 phylogeny. X-axis cartoon labels same as in Figure 2.1. Circles indicate proteolytic susceptibility (faded/dashed) and resistance (solid), with predicted resistance shown for ancA8, ancA9, and ancCG nodes. (d) Decay rates for ancestrally reconstructed proteins, with gels shown on the right. For panels (c) and (d), error bars are the square root of the diagonalized covariance matrix from the fit and the y-axis is in log scale. (e) Summary model for proposed evolution of A9 and A8/A9 innate immune properties. Box around A8/A9 and A9 indicate location in tree (ancestor of therian mammals) where immune functions evolved.

When mapped onto the S100 phylogeny, the most parsimonious explanation for these data is that the shared amniote ancestor—ancCG—was proteolytically resistant (Figure 2.3c). In this scenario, A12s, MRP126s, and A8/A9s conserved ancestral resistance, while A9s and A8s independently lost resistance early in

mammals. Alternatively, ancCG could have been proteolytically susceptible. This would mean that A9s and A8s maintained an ancestral susceptibility, while MRP126s, A12s, and A8/A9s each evolved novel proteolytic resistance.

To distinguish between these possibilities, we characterized ancestrally reconstructed S100s for proteolytic resistance. AncCG and altancCG exhibited extremely high proteolytic resistance (Figures 2.3d, AA11), with degradation rates 3-4 orders of magnitude slower than modern A8s or A9s and approximately one order of magnitude slower than modern A8/A9s. AncA8/A9 and altancA8/A9 also demonstrated high proteolytic resistance, with degradation rates approximately 2-3 orders of magnitude slower than A8s and A9s and comparable to modern A8/A9 complexes (Figures 2.3d, f). Together, these data paint a consistent picture: the amniote ancestor of A9s, ancCG, was strongly resistant to proteolytic degradation. Modern A9s and A8s lost proteolytic resistance from an ancestrally resistant state, while modern A12s, A8/A9 complexes, and MRP126s maintained the ancestral proteolytic resistance (Figure 2.3e).

Finally, we sought to better resolve when A9s acquired proteolytic susceptibility. We hypothesized that this occurred in the ancestor of mammalian A9s before the divergence of therian mammals and marsupials. To test this hypothesis, we measured the proteolytic susceptibility of therian mammalian ancA9 and found that it degraded rapidly. However, its alternative reconstruction (altancA9), was slow to degrade, with a rate two orders of magnitude slower than ancA9 and comparable to other highly resistant S100s. Because the descendants of ancA9 all exhibit proteolytic susceptibility, the simplest explanation is that

altancA9 is a low-quality reconstruction that does not capture the properties of the historical protein. Alternatively, proteolytic susceptibility could have been independently acquired along marsupial and placental mammal lineages.

A single substitution had pleiotropic effects on A9 proinflammatory activity and proteolytic resistance

We found above that A9 evolved to form the antimicrobial A8/A9 complex, gained potent proinflammatory activity, and lost proteolytic resistance over the narrow evolutionary interval after the divergence of mammals and sauropsids but before the divergence of placental mammals and marsupials. We next sought to determine how A9 evolved its antimicrobial and proinflammatory activities and lost proteolytic resistance.

The mechanism by which A9 evolved to form the antimicrobial A8/A9 complex is straightforward. After ancCG duplicated, additional histidines accumulated in the mammalian A8 and A9 ancestors that created the antimicrobil hexahistidine metal binding site in the A8/A9 complex (Figures 2.3e AA1). A8s acquired one additional histidine while retaining the three histidines present in ancCG, while A9s acquired two additional histidines via acquisition of a C-terminal extension (Figure AA1). While A9s evolved five of the six histidines composing the hexahistidine metal binding site, this was not sufficient to convey potent antimicrobial activity (Figure 2.1c). Instead, preservation of A8/A9 heterocomplex formation resulted in proper assembly of the complete antimicrobial hexahistidine site early in mammals. The quantitative difference between ancA8/A9 and

altancA8/A9 antimicrobial activity suggests that other amino acid changes tuned the antimicrobial activity of the molecule, but the core functionality is determined by whether the six histidine residues were present. This is independently supported by Brunjes Brophy et al., who showed that mutating the two C-terminal histidines in A9 is sufficient to strongly decrease the A8/A9 complex's antimicrobial activity.[110]

The mechanisms by which A9s gained proinflammatory activity and lost proteolytic resistance are less obvious, particularly because the mechanism by which A9 activates TLR4 is not well understood. We reasoned that we could identify functionally important amino acid substitutions by focusing on the evolutionary interval over which these properties evolved. We therefore compared the sequences of ancCG (weakly proinflammatory and resistant to proteolytic degradation) and ancA9 (potently proinflammatory and susceptible to proteolytic degradation). We further narrowed down sequence changes of interest by looking for residues conserved in modern A9s (Figure AA13). Finally, we focused on amino acid changes in helix III of A9, as this region is thought to be important for A9 activation of TLR4 based on *in vitro* binding studies and *in silico* docking studies.[109] Only one historical amino acid substitution met all three criteria: position 63 (human A9 numbering). This residue is a phenylalanine in both ancCG and altancCG, is conserved as a phenylalanine in 95% of modern A8s and A12s and has been substituted for a methionine or leucine (M/L) in 97% of A9s (Figure 2.4a).

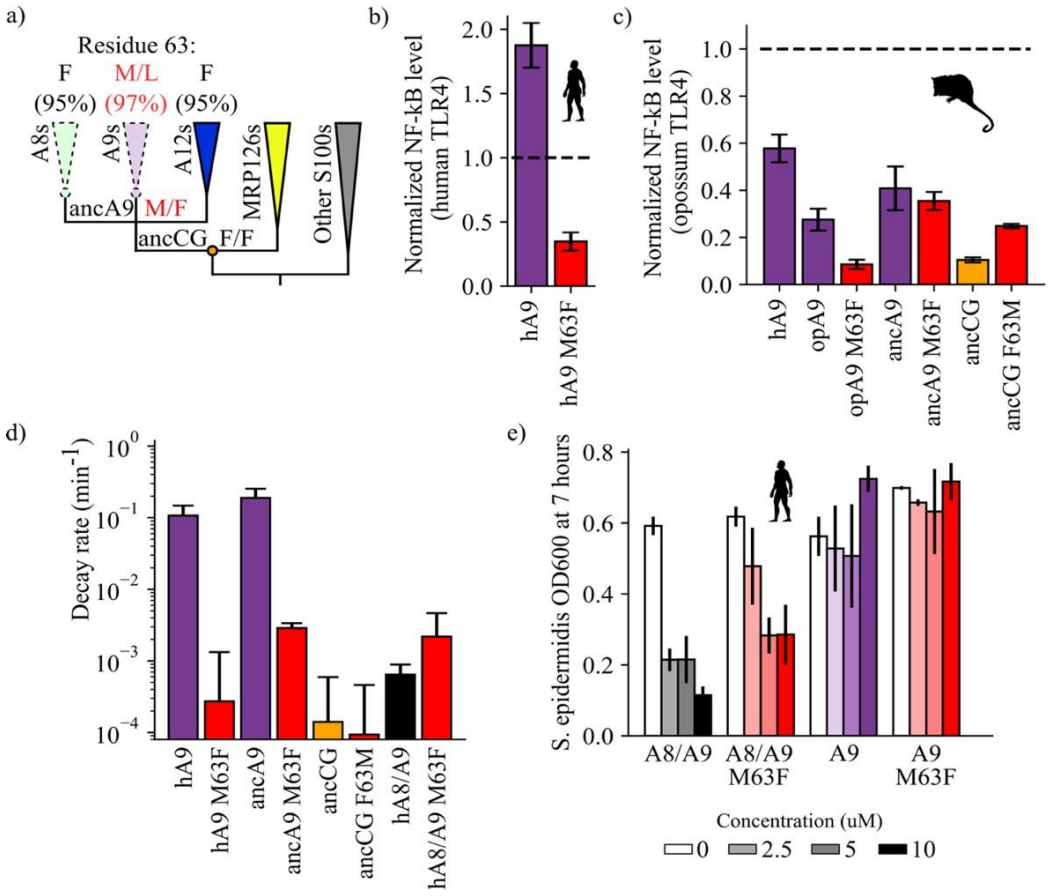


Figure 2.4. A single historical substitution alters A9 proinflammatory activity and proteolytic resistance without affecting properties of the A8/A9 complex. (a) Schematic S100 phylogenetic tree with the amino acid state of position 63 shown at key nodes. Wedges represent clades, colored as in Figure 2.1. Lines indicate proteolytic susceptibility (faded/dashed) and resistance (solid). Circles indicate characterized ancestors. Amino acid labels represent maximum likelihood state/alternate amino acid state for position 63 at ancestral nodes, while labels at clade tips represent percent conservation across modern S100 protein sequences. (b-c) NF-κB production of S100 point mutants at position 63 against human (b) and opossum (c) TLR4. (d) Proteolysis rates for S100 point mutants at position 63 (human A9 numbering). Error bars and y-axis are the same as in Figure 2.1. (e) Antimicrobial activity of hA9 and hA8/A9 with and without M63F mutation against *S. epidermidis.* Axes and error bars same as in Figure 2.1d.

We hypothesized that reverting this site to its amniote ancestral state—M63F—might affect A9 proinflammatory activity. We mutated this position to a phenylalanine in human A9 and opossum A9 and tested each protein for TLR4 activation. Strikingly, we found that introducing M63F into human A9 severely

compromised its ability to activate human TLR4 (Figures 2.4b, AA7). This was also true for opossum A9: introduction of M63F (human numbering) strongly decreased opossum A9 activation of opossum TLR4 (Figures 2.4c, AA8). We next introduced the forward substitution, F63M, into ancCG and tested its proinflammatory activity against opossum TLR4. We observed a modest increase in ancCG activity with the F63M substitution, with activity comparable to that of opossum A9 (Figures 2.4c, AA8).

For most proteins we studied, the amino acid at position 63 did indeed play an important role in determining the pro-inflammatory activity of A9. The effects of toggling position 63 between Met and Phe were not, however, universal. We introduced M63F into ancA9 and observed no change in proinflammatory activity (Figures 2.4c, AA8). Further, altancA9 has a Phe at position 63 but activates TLR4 in the assay (Figures 2.2c, AA8). Thus, while position 63 is an important contributor to activity in modern A9s, other substitutions were also important for the transition from a weakly pro-inflammatory ancestor to the modern set of potently pro-inflammatory A9s.

Because A9s lost proteolytic resistance and gained proinflammatory activity over the same evolutionary time interval, we reasoned that the F63M substitution might have also played a role in A9 loss of proteolytic resistance. To test this, we characterized the proteolytic resistance of human A9 M63F and ancA9 M63F. Strikingly, reversion of this single mutation rendered both ancA9 and human A9 strongly resistant to proteolytic degradation, decreasing their respective degradation rates by 1-2 orders of magnitude and approaching the degradation rates

of ancCG and various A8/A9 complexes (Figures 2.4d, AA12). To relate these findings to proteases that A9 might encounter at sites of inflammation, we also measured the proteolytic resistance of human A9 and human A9 M63F against two neutrophil-specific proteases – cathepsin G and neutrophil elastase (Figure AA14). Neutrophils release these proteases along with A9 at sites of inflammation, often through Neutrophil Extracellular Traps (NETs).[88,164–169] We found that M63F decreased the rate of human A9 degradation in the presence of cathepsin G and neutrophil elastase *in vitro* by approximately one order of magnitude, matching our results using proteinase K (Figure AA14). Lastly, we tested the effect of the forward mutation – F63M – on ancCG proteolytic resistance. We observed no change in resistance for ancCG F63M, indicating that additional substitutions were required to render ancA9 proteolytically susceptible. Together these data show that a single historical reversion is sufficient to render A9s proteolytically resistant, indicating that this position played a role in the loss of A9 proteolytic resistance early in therian mammals.

The pleiotropic substitution minimally affects the A8/A9 complex

A primary goal of this study was to understand the role of pleiotropy in the evolution of multifunctionality. M63F clearly has pleiotropic effects on A9, altering both its proinflammatory activity and proteolytic resistance (Figure 2.4b-d). We next asked whether introducing M63F would pleiotropically affect the antimicrobial A8/A9 complex. Position 63 is somewhat distant from the A8/A9 interface and the antimicrobial hexahistidine site (~10 Å in the manganese-bound

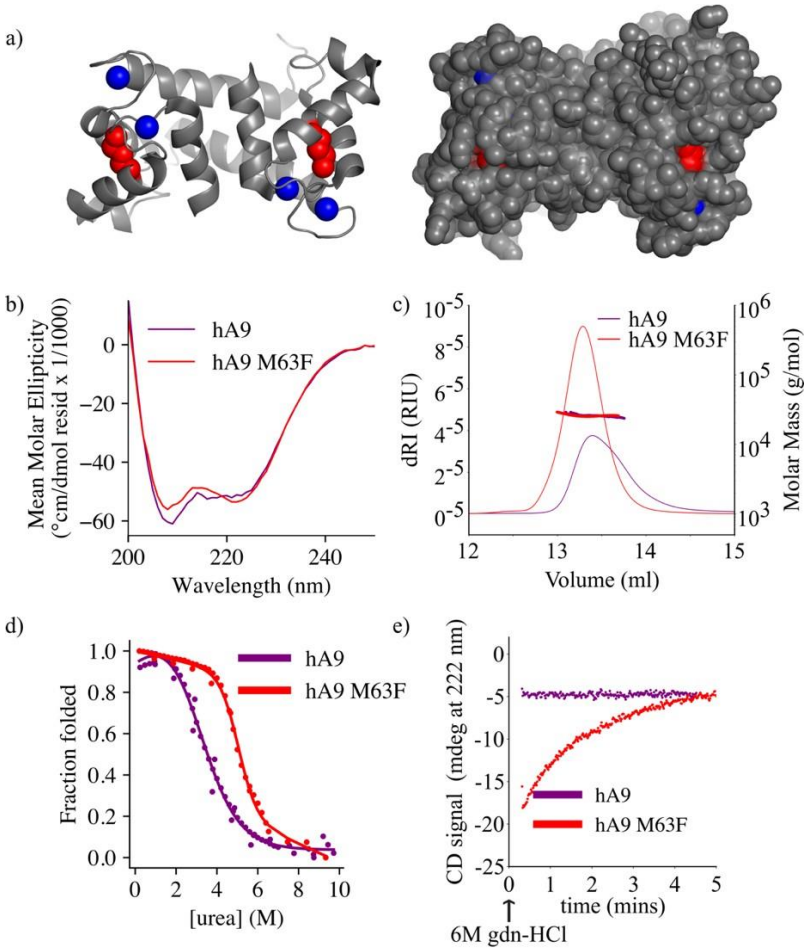
A8/A9 crystal structure);[33] we therefore hypothesized that M63F should not affect A8/A9 complex formation or function. To test this hypothesis, we introduced M63F into human A8/A9 and tested it for oligomeric state, proteolytic resistance, and antimicrobial activity against *S. epidermidis*. As predicted, human A8/A9 M63F predominantly formed a heterotetramer in the presence of calcium by SECMALS with a molecular weight similar to that of wildtype human A8/A9 (48.7 ± 4.2 kDa – Figure AA2). We found that human A8/A9 M63F was also strongly resistant to proteolytic degradation, similar to human A8/A9 (Figure 2.4d). Lastly, M63F had minimal impact on human A8/A9 antimicrobial activity against *S. epidermidis*, retaining potent antimicrobial activity (Figure 2.4e). In contrast, neither human A9 nor human A9 M63F were antimicrobial against *S. epidermidis* (Figure 2.4e). These findings suggest that this single amino acid position had important effects on the evolution of A9 activation of TLR4 and loss of proteolytic resistance without significantly impacting A8/A9 oligomeric state, proteolytic resistance, or antimicrobial activity.

M63F increases protein thermodynamic stability and decreases unfolding rate of human A9

We next asked what effect M63F has on the biophysical properties of human A9. Residue 63 sits in the middle of helix III of A9, pointing inward toward helix II, and is neither a core residue nor fully surface-exposed (Figure 2.5a).[175] Based on the published structure of human A9,[175] a Phe at position 63 could be plausibly tolerated without a steric clash. Using circular dichroism (CD) spectroscopy, we

found that the bulk secondary structure content of human A9 M63F was similar to that of hA9 (Figure 2.5b). We measured the oligomeric state of human A9 M63F by SEC MALS and found that it predominantly forms a homodimer in solution similarly to human A9, with no detectable monomers or larger oligomers (Figure 2.5c, 2.5f). These data together indicate that M63F does not significantly alter human A9's secondary structure or oligomeric state.

We then examined whether M63F alters the stability of human A9. We measured equilibrium unfolding curves for human A9 and human A9 M63F using CD spectroscopy and chemical denaturation via urea. We found that M63F appears to stabilize human A9, increasing the apparent free energy of unfolding by more than 4 kcal/mol and shifting the C_m by ~ 2M urea (Figures 2.5d, 2.5f, AA15). We also measured the unfolding kinetics of human A9 and human A9 M63F in the presence of calcium by spiking protein directly into 6M guanidinium hydrochloride (gdn-HCl) denaturant and monitoring its unfolding rate by CD spectroscopy. Strikingly, human A9 M63F takes several minutes to unfold under these conditions, while human A9 unfolds immediately within the dead time of the experiment (Figures 2.5e-f, AA16). We note that the folding pathway for A9 is complex and almost certainly not two-state—calcium binding, monomer folding, and dimerization all contribute—and thus we cannot reliably determine how M63F affects the stability of each of these potential folding intermediates. The large increase in apparent stability and unfolding rate suggests, however, that the mutation stabilizes some aspect of the folded structure.



f)

Protein	C_m (M urea)	ΔG (kcal/mol)	m-value (kcal/mol·M^{-1})	MW (g/mol)
hA9	2.96 ± 0.22	1.91 ± 0.14	0.65 ± 0.01	26.5 ± 0.5
hA9 M63F	5.03 ± 0.14	6.57 ± 0.18	1.31 ± 0.01	27.1 ± 2.2

Figure 2.5. M63F increases human A9 apparent stability by decreasing its unfolding rate. (a) Crystal structure of hA9 (PDB entry 1irj).[175] Cartoon depiction left, surface view right. Calcium ions are blue spheres. M63 is highlighted in red – two total for homodimeric A9. (b) Far-UV circular dichroism (CD) spectroscopy scans of hA9 and hA9 M63F. Data represent average of 3 scans. (c) SEC MALS analysis of hA9 and hA9 M63F oligomeric state. Solid lines are refractive index (left y-axis), points and molecular weights in table below represent molar mass calculated from light scattering detectors using ASTRA software (right y-axis - see methods). (d) Equilibrium chemical denaturation (urea) of 5 uM hA9 and hA9 M63F monitored by CD at 222 nm. Solid lines represent two-state unfolding model fit to data. (e) Kinetics of hA9 and hA9 M63F unfolding via chemical denaturation (guanidinium hydrochloride). Graph depicts one representative unfolding experiment (figure AA16). (f) Thermodynamic parameters estimated from (d) and molecular weights estimated from (c). Errors are standard deviations calculated from fit (see methods).

Proteolysis is not required for A9 activation of TLR4

The work above identified a mutation that, when introduced into human A9, increases the stability of the protein while also potently compromising its ability to activate TLR4. The mutation is not at a surface position and is therefore not likely a direct participant in the A9/TLR4 protein/protein interface. Further, the same mutation dramatically decreases the proteolytic susceptibility of the protein. One simple way to explain these observations would be if the proteolytic susceptibility itself was the feature that evolved to allow activation of TLR4. This would be consistent with a previous observation that proteolytic products of A9 activate TLR4.[109]

To test whether proteolysis itself was sufficient for activity, we engineered an alternate variant of A9 that was proteolytically susceptible. We introduced the M63A mutation into human A9, anticipating that the short alanine sidechain would not have the stabilizing effect of M63F. As expected, human A9 M63A was highly susceptible to proteolytic degradation, similar to wildtype human A9 (Figures 2.6a, AA12). We reasoned that if proteolysis is the primary determinant of A9 activation of TLR4, then proteolytically susceptible human A9 M63A should potently activate TLR4. Human A9 M63A, however, exhibited diminished proinflammatory activity, similar to human A9 M63F (Figures 2.6b, AA7). This indicates that the methionine at position 63 is important for A9 activation of TLR4. Further, we quantified the amount of human A9, human A9 M63F, and human A9 M63A before and after measuring TLR4 activity and observed no decrease in the amount of full-length protein remaining for wildtype human A9 or either mutant by western blot (Figure

2.6c). This indicates that A9 is not digested by extracellular proteases over the course of the *ex vivo* assay and that proteolysis is not necessary for A9 activation of TLR4.

Although proteolysis does not appear to be a requirement for TLR4 activation, this does not rule out that proteolysis could increase A9 proinflammatory activity by releasing proinflammatory fragments of A9. To test for this possibility, we treated human A9 with agarose-immobilized proteinase K for increasing amounts of time, removed the protease, and then measured the proinflammatory activity of A9 degradation products (Figure 2.6d). If proteolytic products of A9 are the most proinflammatory form of the protein, we might expect to observe a spike in TLR4 activation upon A9 digestion. Instead, we observed a steady decrease in human A9 activity with increasing digestion time. This suggests that full-length human A9 is the most potent activator of TLR4.

We did observe moderate activity for proteolytic products of human A9, as previously shown.[109] After 30 minutes of digestion, no detectable full-length A9 remains by western blot (< 30 ng, Figure 2.6d), but NF-κB production is still quite high, revealing that smaller fragments of A9 are sufficient to provide some degree of activation of TLR4. This raised the possibility that part of M63F's deleterious effect on proinflammatory activity could be to limit the release of active proteolytic fragments of A9. To test this, we also measured human A9 M63F activation of TLR4 after digestion for multiple hours (Figure 2.6d). Unlike wildtype, however, fragments of human A9 M63F did not activate TLR4—even after being liberated by the protease. This strongly suggests that the historical mutation induced a change

in the native structure or dynamics of the molecule to bring about increased activity, independent of its effect on proteolytic susceptibility.

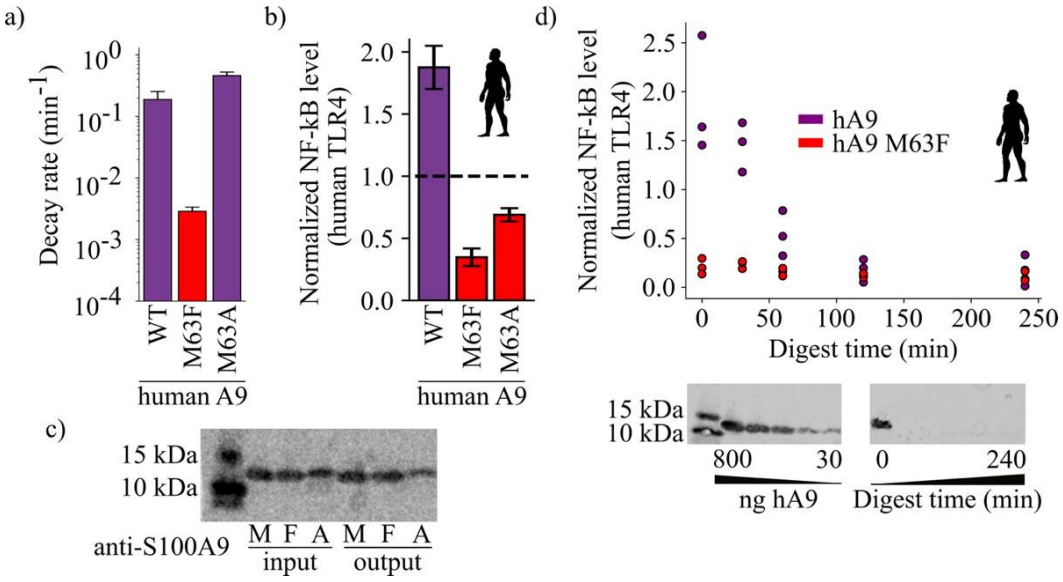


Figure 2.6. Proteolysis is not required for A9 activation of TLR4. (a) Proteolytic decay rates for human point mutants at position 63. Error bars and axes are the same as in Figure 2.3. (b) NF-κB production of human TLR4 in response to treatment with hA9, hA9 M63F, and hA9 M63A. Error bars the same as in Figure 2.2. (c) Western blot of hA9 and position 63 point mutants before and after proinflammatory activity assay. Left bands represent 10 and 15 kDa ladder. (d) NF-κB production of human TLR4 in response to hA9 and hA9 M63F pre-proteolyzed with proteinase K for increasing amounts of time. Points are biological replicates and are the average of three technical replicates. Western blots below depict the amount of full-length A9 remaining over time. Left blot shows antibody sensitivity to A9, right shows digestion time course samples. Ladder and antibody same as in (c).

DISCUSSION

The work presented here provides insight into how the multifunctional protein A9 evolved critical innate immune functions. We find that mammalian A9s gained enhanced proinflammatory activity and lost proteolytic resistance from a weakly proinflammatory, proteolytically resistant amniote ancestor. A single

substitution played a key role in the evolution of these properties without significantly affecting the antimicrobial activity of the A8/A9 heterocomplex. This work contributes to our mechanistic understanding of how A9 activates TLR4 to drive inflammation and clarifies the role of proteolysis in A9 innate immune function.

Innate immune functions of A9 continued to evolve within the mammals

Our data suggest that the proinflammatory and antimicrobial activities of A9 and the A8/A9 complex have undergone further optimization in placental mammals since these functions evolved. While the histidines composing the high-affinity metal binding site of A8/A9 complexes are conserved, we observed differences in antimicrobial potency for different A8/A9 complexes. In particular, human A8/A9 is one of the most potently antimicrobial A8/A9 complexes characterized. This suggests that further optimization of the metal binding site has occurred in along the human lineage within mammals. We also observed differences in activation of TLR4 by different A9s—human A9 is a potent, promiscuous activator of TLR4s from multiple species, while earlier-diverging A9s and other S100s exhibit weaker proinflammatory activity.[130] Future studies are necessary to understand how, mechanistically, later-diverging A9s and A8/A9 complexes have optimized these critical innate immune functions.

Why did A9s lose proteolytic resistance?

While proteolysis is not required for A9 activation of TLR4, it remains unclear why A9s lost proteolytic resistance. We suggest three possibilities. The first

is that loss of proteolytic resistance in A9s was simply a byproduct of evolving proinflammatory activity. No A9 characterized in this study, with the exception of the alternate reconstruction of ancA9, is both proteolytically resistant and potently proinflammatory. This indicates that the molecular requirements for A9 proteolytic resistance may be incompatible with those required for A9 activation of TLR4: A9s may have gained proinflammatory activity at the expense of proteolytic resistance. A second possibility is that A9 proteolytic susceptibility is being maintained to actively remove proinflammatory A9 from the cell and retain the antimicrobial A8/A9 complex. The last possibility for A9 loss of proteolytic resistance is adaptive constraint. There could be selection for some property of A9 or A8/A9 that we did not measure that is incompatible with A9 proteolytic resistance.

While we cannot explicitly distinguish between each of these possibilities, the end result is that A9s lost proteolytic resistance from a resistant ancestor. As A9s activate TLR4 in the protease-rich extracellular space, the functional result of A9 loss of proteolytic resistance is that A9s evolved a proteolytic "timer" concomitantly with evolving proinflammatory activity, all without affecting A8/A9 function.

Novel mechanistic insight into A9 activation of TLR4

Our findings suggest new directions for understanding how A9 potently activates TLR4. TLR4-driven inflammation has been the focus of intense study for over 20 years,[86,91,93–95,117,172,176–178] and the structural basis of TLR4 activation by exogenous agonists, such as the bacterial cell wall component lipopolysaccharide

(LPS), is well understood.[114] In contrast, little is known about how A9 activates TLR4. We have shown here that proteolytic degradation appears dispensable for activation; however, smaller fragments of the protein are sufficient activate TLR4 (Figure 2.6). Given the effect of mutating position 63 on A9 proinflammatory activity, we propose that the region surrounding it—helix III—is important for activity. This is independently supported by Vogl et al., who identified four pairs of double mutants within helix III (amino acids 64, 65, 73, and 77) that, when mutated to alanines in pairs, decrease A9 binding to TLR4 *in vitro*.[109] Biophysical characterization of hA9 M63F (Figure 2.5) indicates that it is more stable and unfolds more slowly, yet it maintains its bulk secondary structure and oligomeric state. The simplest explanation for these data is that M63F is affecting some functionally important dynamic process of the protein, possibly mediated by helix III, that is critical for A9 activation of TLR4. The proteolytic susceptibility of A9s also supports this hypothesis, as proteolysis is a dynamic process that often relies on substrate flexibility and local unfolding events to proceed.[179–183] Damage-Associated Molecular Patterns (DAMPs) often interact with their targets via hydrophobic surfaces;[184–186] one possibility is that A9 undergoes a local unfolding event that exposes a hydrophobic surface to interact with TLR4. This would mean that studies of the native structure of A9 might not be sufficient to gain mechanistic understanding of how it activates TLR4. Further work is required to understand the nature of the active functional state of A9.

Pleiotropic mutations can facilitate the evolution of multifunctionality

Finally, our results suggest a positive role for pleiotropy in the evolution of protein function. Pleiotropy is often viewed as a constraint on evolution: as functional complexity is added to a polypeptide sequence, it becomes increasingly challenging to introduce substitutions—and new functions—without perturbing existing ones.[57,143,145,146,187,188] Here, however, we find a single mutation that had beneficial pleiotropic effects on two important properties of A9: proinflammatory activity and proteolytic susceptibility. If A9 evolved potent proinflammatory activity without gaining susceptibility, it could potentially overstimulate inflammation simply by lingering in the extracellular milieu. Since both properties evolved at once, however, mammals evolved a proinflammatory molecule with a built-in "timer": they gained a new inflammatory signal while avoiding potentially deleterious effects. This shows how pleiotropy can positively contribute to the evolution of new functions.

This same mutation, in contrast, had little pleiotropic effect on another functional state of A9: the A8/A9 complex. The antimicrobial activity of the A8/A9 complex was insulated from any pleiotropic effects from the mutation because proinflammatory and antimicrobial activities were partitioned between A9 and the A8/A9 complex, respectively. A mutation arose in the A9 amino acid sequence and is thus present in both A9 and A8/A9 states, but we only observe effects on the A9 state. This shows that pleiotropic constraint can be reduced when protein functions are partitioned amongst different protein states.

These findings reveal the diversity of pleiotropic roles that a single mutation can play. It further shows how the deleterious pleiotropic effects of mutations can be reduced by partitioning protein functions and properties into different functional states, thus enabling the acquisition, optimization, and expansion of new protein functions. Given the vast diversity of protein functional domains and protein-protein interactions in biology, we suspect that this is a common occurrence in the evolution of protein multifunctionality.

MATERIALS AND METHODS

Phylogenetics and ancestral sequence reconstruction

We reconstructed ancestral sequences using a previously published a phylogenetic tree of S100 proteins containing 172 sequences from 30 amniote taxa (supplemental files 1-2).[130] We used PAML4 to generate maximum likelihood ancestors (marginal probability method)[36,38] using the previously-identified maximum likelihood (ML) substitution model ($LG+\Gamma_8$)[189] on the ML tree. To account for reconstruction uncertainty, we also generated "altAll" versions of each ancestor.[171] We took every site in which the alternate reconstruction had a posterior probability > 0.20 and substituted that amino acid into the maximum-likelihood ancestor. These alternate reconstructions had an average of 12 sequence differences relative to the maximum-likelihood ancestors (Figure AA4). They represent a "worst case" reconstruction relative to our best, maximum likelihood reconstruction.

We also investigated the effect of topological uncertainty on our reconstructed ancestors. In the published phylogenetic analysis, A8s, A9s, A12s, and MRP126s all formed distinct and well-supported clades; however, the branching pattern between these four clades could not be resolved with high confidence.[130] To explore how this uncertainty altered our reconstructed ancestral proteins, we constructed all 15 possible topologies for the A8, A9, A12, and MRP126 clades—i.e ((A8,A9),(A12,MRP126)), ((A8,A12),(MRP126,A9)), etc.—while maintaining species-corrected, within-clade topologies. We then optimized the tree branch lengths and substitution rates for each tree using PhyML.[190] Finally, we used PAML to reconstruct ancA9, ancCG, and ancA8 for all 15 possible arrangements of the MRP126, A12, A8, and A9 clades. The average number of sequence differences for ancestors reconstructed using different topologies was less than or equal to the number of sequence differences between the ML and altAll reconstructions (Table S1). Further, the sites that differed were a subset of those that differed between the ML and altAll reconstructions. Thus, the altAll reconstructions account for sequence changes due to both uncertainty given the ML tree and uncertainty due to topological uncertainty.

Cloning and mutagenesis

All S100 genes in this study were purchased as synthetic constructs in pUC57 vectors from Genscript. S100 genes (A8s, A9s, A12s, MRP126s, and ancestrally reconstructed genes) were sub-cloned into a pETDuet-1 (pD) vector (Millipore). A8s, A12s, MRP126s, and ancCGs were cloned into multiple cloning

site #1 (MCS1) of the pD vector, while A9s were cloned into MCS2. For expression and purification of A8/A9 heterocomplexes (A8/A9s), pD plasmids containing an A8 gene in MCS1 and an A9 gene in MCS2 were used as previously described.[191] Opossum A8 was sub-cloned into an MBP-LIC vector to yield a His-MBP-TEV-opA8 construct. For opossum A8/A9, the entire His-MBP-TEV-A8 construct was then sub-cloned into MCS1 of a pD vector containing a marsupial A9 in MCS2. Other S100s (A1, A5, A7, A11, A14, and P) were previously cloned into a pET28/30 vector to yield a TEV-cleavable N-terminal His tag.[112] Cysteine-free versions of all S100 genes, as well as point mutants, were prepared using site-directed mutagenesis (Agilent).

Protein expression and purification

Recombinant protein overexpression was conducted in *E. coli* BL21 (DE3) pLysS Rosetta cells. Cultures were innoculated in luria broth overnight at 37°C, shaking at 250 rpm, in the presence of ampicillin and chloramphenicol. The following day, 10 ml of saturated culture was diluted into 1.5 L of media with antibiotics, grown to OD_{600} = 0.6 – 1, and then induced overnight at 16°C using 1 mM IPTG. Cells were pelleted at 3,000 rpm for 20 min and stored at -20°C for no more than three months.

Lysates were prepared by vortexing pellets (3-5 g) in tris buffer (25 mM tris, 100 mM NaCl, pH 7.4) and incubating for 20 min at RT with DNAse I and lysozyme (ThermoFisher Scientific). Lysates were sonicated and cell debris was pelleted by centrifugation at 15,000 rpm at 4°C for > 20 min. All proteins were

purified on an Äkta PrimePlus FPLC using various 5 ml HiTrap columns (HisTrap FF (Ni-affinity), Q HP (anion exchange), SP FF (cation exchange), and MBPTrap HP (MBP) - GE Health Science). A1, A5, A7, A11, A14, and S100P were purified using a a TEV-cleavable His tag strategy used by our lab previously[1,3,4]. All other S100s, except for opossum A8 and opossum A8/A9, were purified in three steps using Ni-affinity chromatography in the presence of calcium followed by two rounds of anion exchange chromatography at different pHs. For Ni-affinity chromatography, proteins were eluted over a 50 ml gradient from 25-1000 mM imidazole in tris buffer. Peak elution fractions were pooled and placed in dialysis overnight at 4°C in 4 L of tris buffer (calcium-free) adjusted to pH 8. Anion exchange chromatography was then performed the following day over a 50 ml gradient from 100-1000 mM NaCl in pH 8 tris buffer. Fractions containing majority S100 were pooled and analyzed for purity on an SDS-PAGE gel. If trace contaminants remained, an additional anion exchange step was performed at pH 6 using the same elution strategy as for the previous anion exchange step.

Opossum A8 and A8/A9 lysates were prepared as above and then flowed over a nickel column, eluting over a 50 ml gradient from 25-1000 mM imidazole in tris buffer. Peak elution was pooled and the MBP tag was cleaved by incubation with ~1:5 TEV protease at 4°C overnight in 4 L of tris buffer. The MBP tag was then removed by flowing the sample over an MBPTrap column, step-eluting with 10 mM maltose. Additional MBP columns were run until all MBP was removed from the purified protein, assessed by SDS-PAGE. If necessary, an additional anion exchange step at pH 8 was performed to complete purification. All purified proteins

were dialyzed overnight at 4°C in tris buffer + 2 g/L Chelex-100 resin (Biorad), flash-frozen the following day in liquid nitrogen, and stored at -80°C.

Biophysical and biochemical characterization

For all experiments, protein aliquots were thawed fresh from freezer stocks and were either dialyzed in the appropriate experimental buffer overnight at 4°C or exchanged 3X into experimental buffer using 3K microsep spin concentrator columns (Pall Corporation). All samples were filter-sterilized using 0.1 um spin filters (EMD Millipore) prior to measuring concentration and using in experiments. Thawed aliquots were used for no more than one week before discarding. All concentrations were measured by Bradford assay and correspond to micromolar dimeric protein.

For *in vitro* proteolytic susceptibility experiments, proteins were dialyzed or exchanged into tris buffer + 1 mM $CaCl_2$. 12.5 uM S100 protein was treated with 5 uM monomeric Proteinase K from *Tritirachium album* (Sigma Aldrich), cathepsin G from human neutrophils (Athens Research), or neutrophil elastase from human neutrophils (Millipore Sigma) in thin-walled PCR tubes, which were held at a constant temperature of 25°C over the course of the experiment using a thermal cycler. Proteinase K activity was quenched at different time points by directly pipetting an aliquot of the reaction into an equal volume of 95% Laemmli SDS-PAGE loading buffer + 5% BME at 95°C in a separate thermal cycler. Time points were analyzed via SDS-PAGE, and gels were quantified by densitometry using in-house gel analysis software (https://github.com/harmslab/gelquant, v1.0). An

exponential decay function ($A_o e^{-kt}$) was fit to the data to extract the decay rate, floating A_o and k. Standard deviations were calculated from fits by taking the square root of the diagonalized covariance matrix and by error propagation.

Oligomeric states were measured using a superose 12 10/300 GL size exclusion column (Amersham Biosciences) with in-line concentration detection using refractive index (RI) and particle mass measured using a multiangle laser light scattering (MALS) instrument (Dawn Heleos, Wyatt Technology). Samples were concentrated to 0.5-2 mg/ml in tris buffer + 0.5 mM $CaCl_2$, 0.1 um sterile-filtered, and analyzed at a flow rate of 0.2 ml/min. Data were processed using manufacturer's software (Astra).

Circular dichroism (CD) and chemical denaturation experiments were performed using a Jasco J-815 CD spectrometer and spectroscopy-grade guanidine hydrochloride (gdn-HCl) or urea. Chemical denaturation was performed using 25 uM dimeric protein in tris buffer with $CaCl_2$, with tris substituted for spectroscopy-grade trizma. Reversible unfolding and refolding curves were constructed by making concentrated 100 uM protein stocks in either buffer or 6M gdm or 10M urea and then preparing protein dilutions in various concentrations of gdn-HCl or urea in buffer. Samples were left to equilibrate in denaturant between three hours and overnight to allow for equilibration and were then analyzed by CD. Unfolding/refolding equilibration was confirmed by comparing unfolded vs. refolded protein at the same concentration. CD signal was quantified at 222 nm in a 1 mm cuvette using a 1 nm bandwidth, standard sensitivity, and 2 second D.I.T. HT voltage was < 600 V. We fit a two-state unfolding model:

$$\frac{b_f + m_f x + (b_u + m_u x) e^{-\frac{\Delta G - mx}{RT}}}{1 + e^{-\frac{\Delta G - mx}{RT}}}$$

to the data to extract thermodynamic parameters, where $b_f, m_f, b_u,$ and m_u are the folded and unfolded baseline y-intercepts and slopes, ΔG is the unfolding free energy, m is the m-value, R = 0.001987 JxK^{-1}mol^{-1} and T = 298.15 K. Standard deviations were calculated from fits by taking the square root of the diagonalized covariance matrix and by error propagation. Apparent unfolding kinetics studies were performed using the above conditions by spiking concentrated protein stock directly into 6M gdm and immediately monitoring CD signal at 222 nm.

Cell lines

We purchased commercially distributed HEK293T cells from ATCC (CRL-11268). Because we are using this cell line as a host for heterologous transient transfections, the appropriate control for consistency between assays is the measurement of reporter output for a set of control plasmids and a panel of known treatments. Upon thawing each batch of cells, we run a positive control for ligand-induced response. We transfect the cells with plasmids encoding human CD14, human MD-2, human TLR4, renilla luciferase behind a constitutive promoter, and firefly luciferase behind an NF-KB promoter. We then characterize the raw luciferase output for five treatments: 1) mock, 2) LPS, 3) LPS + polymyxin B, 4) S100A9 + polymyxin B, and 5) S100A9 + 1.25x polymyxin B. This has a stereotypical pattern of responses in renilla luciferase (high for all) and firefly luciferase (low, high, low, high, high). To validate that this response is dependent

on the transfected TLR4 complex as opposed to the cells themselves, we repeat the experiment but exclude the TLR4 plasmid. This should give identical renilla luciferase values but no firefly luciferase output in response to any treatment. To ensure that the cells maintain their properties between passages, we repeat the mock, LPS, and LPS + polymyxin B control on every single experimental plate. This assay has a built-in control for mycoplasma contamination: high firefly luciferase signal in the absence of added agonist. This indicates that there is another source of TLR4-induced NF-kappa B output in the cells—most plausibly, contamination. This mycoplasma sensing approach is used in the commercially available HEK-BLUE mycoplasma detection kit (Invivogen). We discard any cells that exhibit high background values or reach 30 passages.

Functional Assays

The antimicrobial activity of S100s was measured against *S. epidermidis* using a well-established assay.[97,102,104,110,192] The day before, a 5 ml starter culture of *S. epidermidis* in tryptic soy broth (TSB) was grown overnight. The next day, the culture was diluted ~1:100 in TSB and grown for approximately 2 hours to an OD600 of ~0.8. Immediately prior to experiment, the *S. epidermidis* culture was again diluted 1:100 at a ratio of 62:38 experimental buffer (25 mM tris, 100 mM NaCl, 3 mM $CaCl_2$, pH 7.4):TSB. S100 proteins were exchanged into experimental buffer. Each well of a sterile 96-well plate was prepared with 40 ul of *S. epidermidis* diluted in experimental buffer + TSB, S100 protein at the desired concentration in experimental buffer, and then filled to 200 ul, maintaining a ratio of 62:38

experimental buffer:TB. *S. epidermidis* growth was monitored on a plate reader, measuring OD600 every 15 minutes for 13 hours. Each measurement was collected in technical triplicate and background-subtracted using a blank containing experimental buffer and TSB alone. Protein samples were confirmed to lack bacterial contamination by measuring S100 protein growth in experimental buffer and TSB lacking *S. epidermidis.*

All plasmids, cell culture conditions, and transfections for measuring the activity of S100s against TLR4s were identical to those previously described.[38,39130,172] Briefly, human embryonic kidney cells (HEK293T/17, ATCC CRL-11268) were maintained up to 30 passages in Dulbecco's Modified Eagle Media (DMEM) supplemented with 10% fetal bovine serum (FBS) at 37°C with 5% CO2. Lipopolysaccharide *E. coli* K-12 LPS (LPS - tlrl-eklps, Invivogen) aliquots were prepared at 5 mg/ml in endotoxin-free water and stored at -20^{o}C. Working solutions were prepared at 10 ug/ml and stored at 4^{o}C to avoid freeze-thaw cycles. S100 proteins were prepared by exchanging into endotoxin-free PBS and incubating with an endotoxin removal column (Thermo Fisher Scientific) for 2 hours. S100 LPS contamination was assessed by measuring activity with and without Polymyxin B, an LPS chelating agent (Figure AA6). LPS (200 ng per 100 ul well) or S100 (0.8, 0.4, 2, 4, or 5 uM dimer) treatments were prepared by diluting in 25:75 endotoxin-free PBS:serum-free Dulbecco's Modified Eagle Media (DMEM – Thermo Fisher Scientific). Polymyxin B (PB, 200 ug per 100 ul well) was added to all S100 experimental samples to limit background endotoxin contamination activity from recombinant protein preps. Cells were incubated with

treatments for 3 hours prior to assaying activity. The Dual-Glo Luciferase Assay System (Promega) was used to assay Firefly and Renilla luciferase activity of individual wells. Each NF-κB induction value shown represents the Firefly luciferase activity divided by the Renilla luciferase activity, background-subtracted using the LPS + PB activity for each TLR4 species and normalized to the activity of LPS alone for each TLR4 species to normalize between plates. All measurements were performed using three technical replicates per plate, a minimum of three biological replicates total, and a minimum of two separate protein preps.

For TLR4 activation measurements by A9 proteolytic products, 12.5 uM hA9 or hA9 M63F were incubated with 2.5 mg/ml Proteinase K immobilized to agarose at 37^{o}C for increasing amounts of time. The reaction was quenched by spin-filtering the sample to remove Proteinase K. 2 uM A9 proteolysis treatments were then added to cells as outlined above. Western blots were performed by running an SDS-PAGE gel and transferring to a nitrocellulose membrane. Membranes were blocked using Odyssey Blocking Buffer for 1 hour, incubated with 1:1000 mouse anti-S100A9 primary antibody (M13 clone 1CD22, Abnova) for 1 hour, and incubated with 1:10,000 IRDye Goat anti-mouse 800CW IgG (H+L, Licor) for 1 hour, with 3x5 min TBST washes in between each step. Blots were imaged using the Licor Odyssey Fc imaging system.

Species cartoons

All species cartoons were taken from the following websites: http://phylopic.org/image/c089caae-43ef-4e4e-bf26-973dd4cb65c5/,

http://phylopic.org/image/aff847b0-ecbd-4d41-98ce-665921a6d96e/, http://phylopic.org/image/0f6af3d8-49d2-4d75-8edf-08598387afde/http://phylopic.org/image/dde4f926-c04c-47ef-a337-927ceb36e7ef/.

We acknowledge Sarah Werning and David Liao as authors of the opossum and mouse cartoons respectively, which were made publicly available through the creative commons attributions 3.0 unported license (https://creativecommons.org/licenses/by/3.0/).

BRIDGE TO CHAPTER III

In this chapter, we determined that A9 gained potent proinflammatory activity and lost proteolytic resistance from a weakly proinflammatory, proteolytically resistant amniote ancestor.[131] We showed that A9 maintained the ability to form a complex with A8 post-gene duplication, resulting in the evolution of the antimicrobial, proteolytically resistant A8/A9 heterocomplex in mammals. We identified a single historical substitution – M63F - that occurred in the mammalian A9 ancestor and played a key role in A9 loss of proteolytic resistance and gain of proinflammatory activity. The historical substitution altered multiple A9 functions without significantly changing the A8/A9 complex, showing that pleiotropy can play a beneficial role in the evolution of protein multifunctionality. Lastly, we showed that proteolytic degradation of A9 is not necessary for A9 activation of TLR4. This chapter provides novel insight into how A9 evolved

multifunctionality and lays a foundation for determining the mechanism by which A9 activates TLR4 to drive inflammation. Chapter III builds upon these findings by dissecting the biophysical mechanism through which the M63F mutation alters A9 activation of TLR4.

CHAPTER III

BIOPHYSICAL STUDIES OF A CONSERVED HISTORICAL MUTATION SUGGEST THAT S100A9 ACTIVATES TLR4 THROUGH A NON-NATIVE CONFORMATION

AUTHOR CONTRIBUTIONS

Joseph Harman and Michael Harms conceptualized the study and designed the experiments. Michael Harms and Susan Marqusee acquired the funding for the study. Joseph Harman, Patrick Reardon, Gus Warren, Jeremy Anderson, Shawn Costello, and Patrick Connor performed the experiments. Michael Harms and Susan Marqusee administered the project. Joseph Harman, Patrick Reardon, Shawn Costello, and Jeremy Anderson analyzed the data. Joseph Harman constructed the figures. Joseph Harman wrote the manuscript with editorial assistance from Michael Harms. Hydrogen-deuterium exchange methods were contributed by Shawn Costello.

INTRODUCTION

S100A9 (A9) is one of the most abundant proteins in neutrophils[79,80] and potently activates inflammation via Toll-like Receptor 4 (TLR4).[86,91–95] It is a biomarker for various inflammatory diseases,[81–85] a damage-associated molecular pattern (DAMP) that signals tissue damage to the immune system,[92,151,153,186] and a drug target for chronic inflammatory disorders.[129,193] However, the mechanism by

which A9 activates TLR4 to drive inflammation is unknown. Determining how A9 induces inflammation via TLR4 is critical for understanding mechanisms by which DAMPs participate in mammalian innate immunity and for designing drugs to reduce aberrant A9-mediated inflammation.

Previous studies have yielded limited insight into how A9 interacts with TLR4. Lipopolysaccharide (LPS) – a small bacterial cell wall component – activates TLR4 by binding to a small hydrophobic pocket of MD2, a TLR4 cofactor required for activity.[117,121] It was previously proposed that smaller proteolytic fragments of A9 might activate TLR4 similarly to LPS, as A9 is readily proteolytically degraded and the extracellular space is rich in proteases.[109] We recently showed, however, that A9 proteolytic degradation is not required for TLR4 activation.[131] Additional studies have identified pairs of charged mutations that disrupt A9 binding to TLR4 *in vitro*, leading to a proposed *in silico* A9:TLR4 docking model.[109] However, the functional effects of these mutations on A9 activation of TLR4 were not characterized.

We recently identified a historical mutation in A9 that, when reverted to its amniote ancestral state in human A9 (A9 M63F), strongly compromised A9 activation of TLR4 in an *ex vivo* cell culture assay.[131] This was, to our knowledge, the first mutation that has been shown to alter A9 proinflammatory activity. Surprisingly, the M63F mutation is at an internal position of the protein, making it difficult to identify an A9:TLR4 interaction interface. We further found that M63F stabilizes A9 in the presence of calcium by decreasing its unfolding rate. This is puzzling because A9 activates TLR4 in the calcium-rich extracellular space, and

thus the calcium-bound form of A9 is thought to be the "active" state of the protein. How does M63F stabilize the "active" state of the protein, yet also reduce A9 activation of TLR4?

We set out to determine how, mechanistically, M63F alters A9 proinflammatory activity. We find that M63F stabilizes only the calcium-bound state – but not the calcium-free state – of A9. Biophysical characterization of the A9 M63F structure indicates that M63F does not appreciably alter the native structure of the protein. We find that A9 and A9 M63F exhibit similar dynamics over short timescales (ps - μs) in NMR NOE and CPMG experiments. However, we show that M63F strongly quenches long timescale dynamics in A9 using hydrogen-deuterium exchange (HDX) experiments. Mapping differences in exchange onto the A9 protein structure[175] reveals that M63F locally stabilizes A9 in regions previously proposed to interact with TLR4.[109] We identify a second phenylalanine (F37) within the stabilized region that is in close proximity to M63F. Mutating F37 to a leucine in the A9 M63F background (A9 M63F F37L) negates the stabilizing effect of M63F and fully restores A9 proinflammatory function. These findings show that M63F decreases A9 proinflammatory activity by locally stabilizing the protein via an interaction with a nearby residue – all without significantly altering the protein's native structure. We propose that A9 adopts a non-native conformation – distinct from the native calcium-bound structure – to activate TLR4. These findings propose a novel mechanism by which a key innate immune protein drives inflammation, enabling the development of treatment strategies to target the proinflammatory non-native state of S100A9.

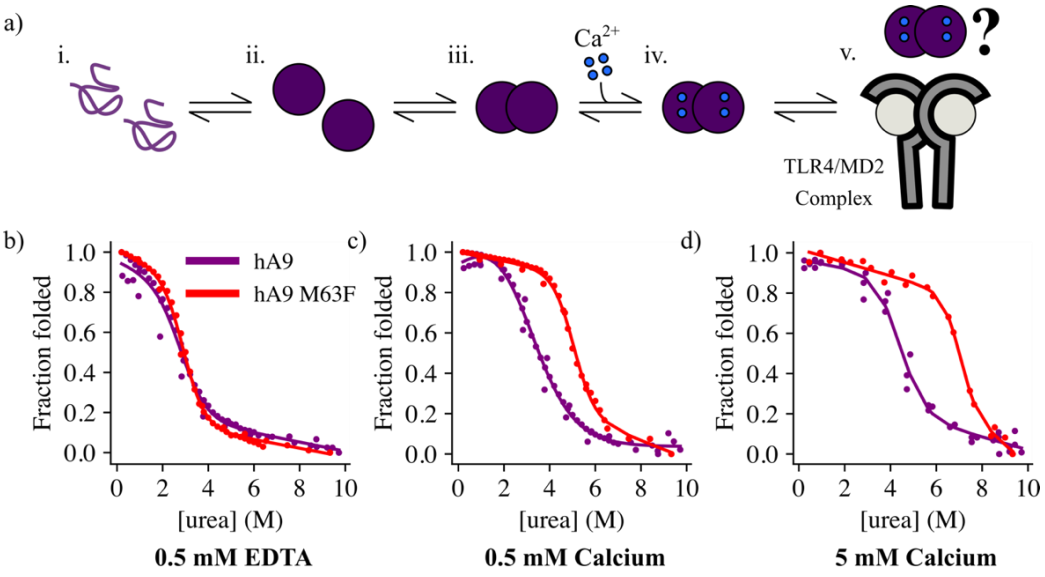
RESULTS

The M63F substitution stabilizes the calcium-bound state of A9

We first sought to determine which state(s) of A9 are stabilized by M63F along the A9 folding pathway. We previously showed that M63F stabilizes A9 in the presence of calcium.[131] However, A9 folding is a multi-step process that includes monomer folding, homodimerization, a conformational change, and then calcium binding (Figure 3.1A, i-iv). As the state of A9 that activates TLR4 is unknown, an additional state of A9 might also be required for A9 proinflammatory activity (figure 3.1A, v). The M63F mutation could hinder A9 activation of TLR4 by altering any of these folding or conformational intermediates or by stabilizing an off-pathway state.

To determine which state(s) of A9 are altered by M63F, we first measured the stability of A9 and A9 M63F in the absence of calcium. We performed equilibrium chemical denaturation experiments using urea and monitored protein folding using circular dichroism (CD) spectroscopy, as reported previously (see methods).[131] We fit an apparent two-state unfolding model to the data to extract the apparent folding free energy, m-value, and C_m (Figure 3.1b-e). In the absence of calcium, we observed little difference in apparent stability between A9 and A9 M63F at 5 uM protein concentration (2.91 ± 0.63 vs. 3.72 ± 0.12 kcal/mol – Figure 3.1b, 1e). We also measured the oligomeric states of calcium-free A9 and A9 M63F using size exclusion chromatography coupled with multi-angle laser light scattering (SECMALS). We found that both proteins form predominantly homodimers at low micromolar concentrations (Figure AB1), as previously observed in the presence

of calcium,[131] indicating that the M63F mutation does not significantly impact A9 homodimerization. This is unsurprising, as M63F is distant from the A9 dimerization interface[175] and the K_d of A9 homodimerization is sub-micromolar.[170] These findings indicate that the M63F mutation does not significantly alter the intrinsic stability of A9 monomers or the calcium-free A9 homodimer (Figure 1a, i-iii), but instead impacts a calcium-bound state of A9.



e)

Protein	Condition	ΔG (kcal/mol)	m-value (kcal/mol·M^{-1})	C_m (M urea)
hA9	EDTA	2.91 ± 0.63	1.03 ± 0.04	2.82 ± 0.62
hA9 M63F	EDTA	3.72 ± 0.12	1.27 ± 0.01	2.93 ± 0.10
hA9	0.5 mM Ca	1.91 ± 0.14	0.65 ± 0.01	2.96 ± 0.22
hA9 M63F	0.5 mM Ca	6.57 ± 0.18	1.31 ± 0.01	5.03 ± 0.14
hA9	5 mM Ca	4.83 ± 1.30	1.10 ± 0.06	4.41 ± 1.21
hA9 M63F	5 mM Ca	11.21 ± 2.81	1.61 ± 0.07	6.98 ± 1.77

Figure 3.1. M63F specifically stabilizes the calcium-bound form of A9. a) Folding pathway for A9. Unfolded monomers (i) fold (ii) and dimerize (iii), followed by calcium binding (iv). Calcium-bound A9 is then thought to activate the TLR4/MD2 complex (v). b-d) Equilibrium urea denaturation of A9 (purple) and A9 M63F (red) with b) 0.5 mM EDTA, c) 0.5 mM $CaCl_2$, or d) 5 mM $CaCl_2$ present. Points are biological replicates and solid lines are a two-state unfolding model fit to the data (see methods). e) Thermodynamic parameters calculated from fits in b-d). Errors are from fits (see methods).

To test whether M63F specifically alters the calcium-bound state of A9, we measured the stability of A9 and A9 M63F in the presence of calcium. We previously measured A9 and A9 M63F stability in the presence of 0.5 mM $CaCl_2$ (reproduced in figure 3.1c, 1e).[131] We expanded upon these findings by measuring A9 and A9 M63F stability with 5 mM $CaCl_2$ present (Figure 3.1d-e). We found that M63F stabilizes A9 in a calcium-dependent manner, affording 4.7 ± 0.2 kcal/mol and 6.4 ± 3.1 kcal/mol of additional stability at 0.5 mM and 5 mM $CaCl_2$, respectively. These findings show that M63F stabilizes the calcium-bound state(s) of A9 (Figure 3.1a, i-iii) without significantly affecting the stability of any calcium-free states (Figure 3.1a, iv-v).

M63F does not appear to significantly alter A9 structure

We next asked whether the structure of A9 is altered by the M63F mutation. We previously showed that the bulk secondary structure and oligomeric state of A9 are unchanged by M63F in the presence of calcium, indicating that M63F doesn't cause a large, global change in A9 structure.[131] However, this does not rule out the possibility that M63F causes a local structural change and/or stabilizes a misfolded, functionally inactive state of the protein. To test whether M63F alters the structure of A9, we characterized the A9 M63F mutant by NMR. We transferred and confirmed peak assignments for A9 M63F from the calcium-loaded wildtype A9 NMR structure[193] using TROSY-HSQC experiments. We successfully assigned 91 out of 114 A9 M63F peaks. The majority of unassigned peaks are within the A9 disordered C-terminal tail, many of which are also unassigned in the wildtype A9

structure.[193] We then compared changes in chemical shift for each assigned residue in A9 and A9 M63F (Figure 3.2a-b). We mapped the combined chemical shift differences between A9 and A9 M63F onto the A9 structure and found that the largest changes in chemical shift occur locally around the M63F mutation. Minimal changes in chemical shift are observed for regions of the protein that are distant from M63F. These findings suggest that M63F alters the local chemical environment of nearby residues but does not significantly affect the overall structure of A9.

M63F does not alter A9 dynamics on short timescales

We wondered whether M63F might inhibit A9 proinflammatory activity by altering a functionally important dynamic process of the protein. To test for A9 dynamics that are affected by M63F, we again turned to NMR. We probed protein dynamics on three timescales: ps-ns, ns-μs, and ms+. We examined A9 dynamics on the ps-ns timescale using a two-point Car-Purcell-Meiboom-Gill (CPMG)[194] experiment. We first measured wildtype A9 (Figure 3.3a). As expected from the A9 structure and previous work, the C-terminal tail of A9 exhibited rapid movement, as did loop regions. We then repeated the measurement for M63F (Figure 3.3a). There was no appreciable difference in CPMG ps-ns dynamics between WT and M63F (Figure 3.3a-b).

We then probed the ns-μs timescale using NOE relaxation-dispersion experiments.[195] We found that the bulk of A9 operated as a folded unit, with the tails moving independently on this timescale (Figure 3.3c). Similar to CPMG

experiments, we obtained comparable results for M63F (Figure 3.3c-d). These data together indicate that M63F does not alter A9 short timescale dynamics.

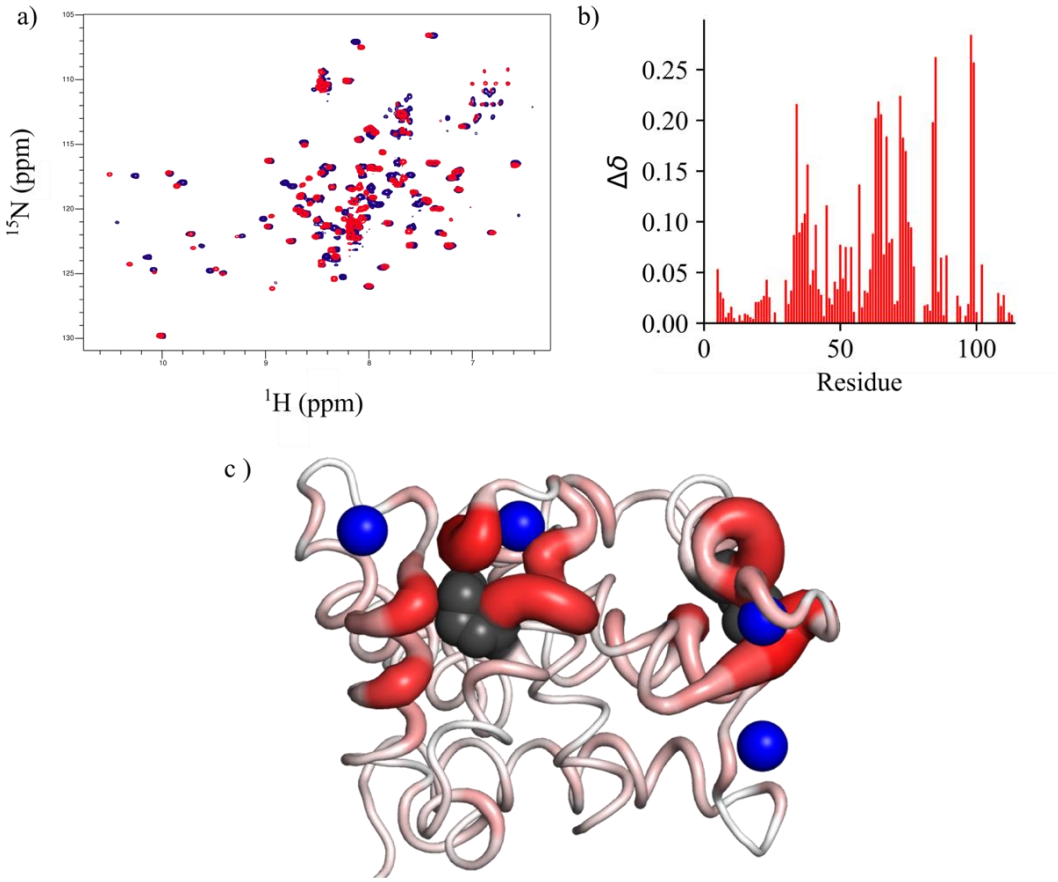


Figure 3.2. M63F locally perturbs nearby residues without affecting bulk A9 structure. a) Overlaid HSQC NMR spectra for A9 (purple) and A9 M63F (red). b) Combined chemical shift (CCS) differences between A9 and A9 M63F for 91 assigned residues (see methods). c) Data from b) mapped onto A9 crystal structure (PDB 1irj). Larger tubes and brighter red color indicate larger CCS difference. Calcium ions are blue spheres, residue 63 shown as grey spheres.

A9 long timescale dynamics are quenched by the M63F mutation

We next asked if M63F alters long timescale dynamics in A9. Supporting this possibility, we previously showed that M63F significantly decreases the unfolding rate of A9 such that A9 M63F unfolding can be observed over the course of several minutes.[131] To probe for changes in long timescale dynamics, we

measured hydrogen-deuterium exchange (HDX) by NMR for A9 and A9 M63F. We collected HSQC spectra for both proteins over the course of 24 hours (Figure 3.4a-d) and quantified peak intensities for each timepoint. As many peaks disappeared more rapidly than the first measured timepoint (~10 minutes), we used the Akaike Information Criterion (AIC) to determine whether a single exponential decay model or a linear model was more appropriate for extracting an exchange rate from the data for each residue (see methods). This is a conservative approach that places a lower limit on the exchange rate for each residue – if a linear model better describes the exchange process than a single exponential decay model (due to a lower AIC), then the exchange rate is simply assigned as faster than the fastest-exchanging residue for which a single exponential decay model was employed due to a lower AIC (see methods).

Using this approach, we calculated the log difference in exchange rate for each residue between A9 and A9 M63F (Figure 3.4e, AB3). Of the 93 residues assigned in A9 and A9 M63F, 49 residues exchanged too rapidly to quantify in both A9 and A9 M63F (Figure 3.4e, grey). Two residues exchanged too rapidly in A9 M63F to quantify but were measurable in wildtype A9 (Figure 3.4e – blue). We successfully quantified exchange rates for 26 residues in A9 M63F, 16 of which exchanged too rapidly in wildtype A9 to be measured (Figure 3.4e – pink and red, respectively). Calculated exchange rates for the remaining 18 assigned residues are not shown due to high error (> 3 log units). However, each of these 18 residues exchanged more slowly in A9 M63F than in A9 (Figure AB2-3).

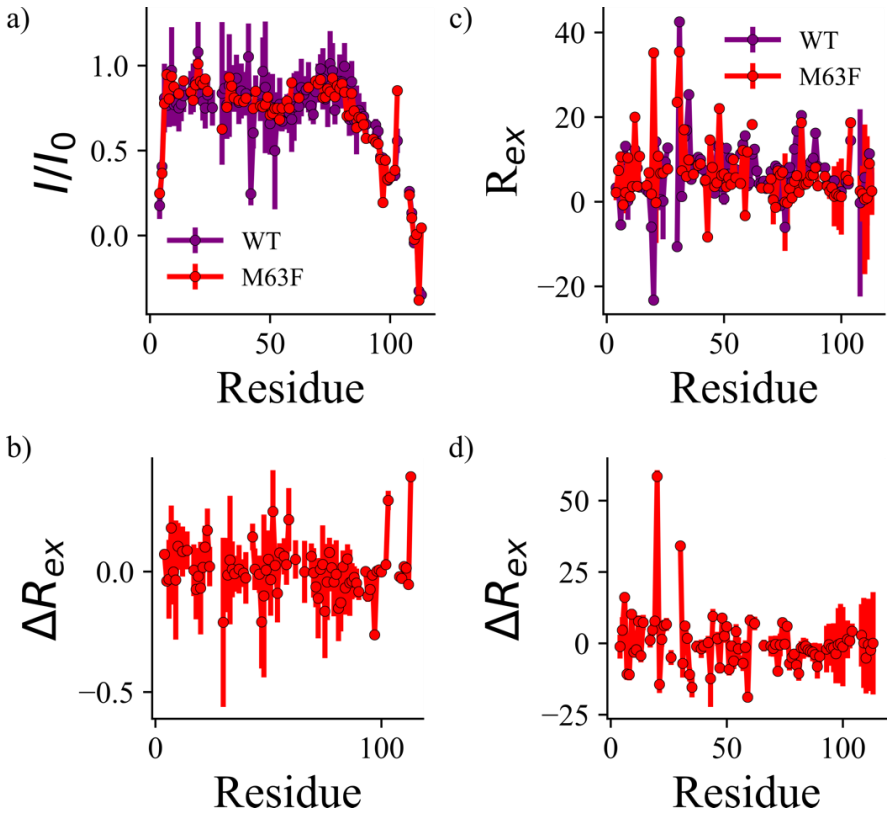


Figure 3.3. M63F does not alter A9 dynamics on short (ps-us) timescales. a) and c) 2-point CPMG and NOE NMR dynamics measurements, respectively. Differences in CPMG and NOE between A9 and M63F shown in b) and d). Points are average relaxation times, error bars are standard deviation.

Mapping differences in log exchange rate for each residue between A9 and A9 M63F onto the A9 crystal structure revealed that M63F strongly decreases exchange for multiple nearby residues, both in helix III where M63F resides and in helix II directly across from M63F. These data parallel the changes in combined chemical shift observed Figure 3.2. M63F thus hinders long timescale A9 dynamics by locally stabilizing nearby residues in the protein.

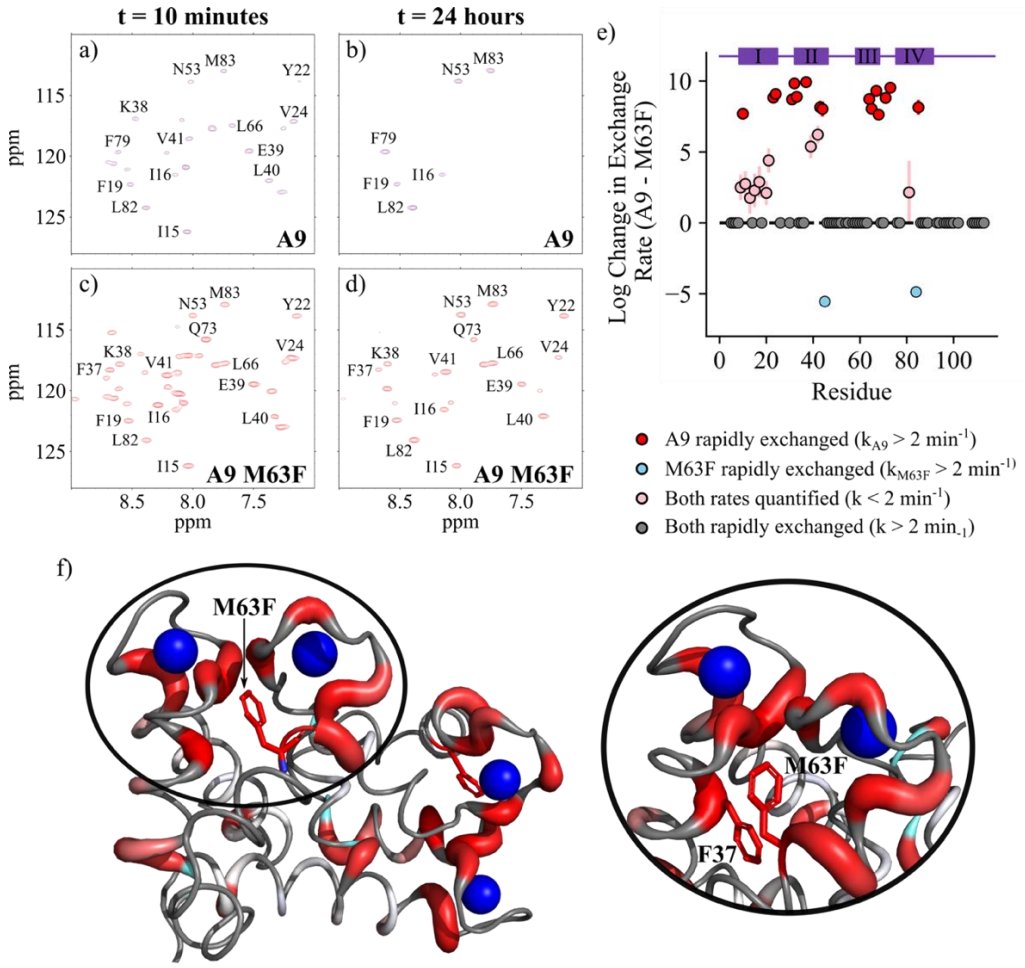


Figure 3.4. M63F strongly increases local A9 stability. a) NMR HSQC HDX spectra for wildtype A9 (a-b) and A9 M63F (c-d). a) and c) show peaks present after 10 minutes of exchange, while b) and d) show peaks present after 24 hours of exchange. X and y-axes are ^{1}H and ^{15}N chemical shifts (ppm). e) Log change in exchange rate for residues in A9 and A9 M63F. Points are exchange rate extracted from fit to HSQC data, error bars propagated from fits (see methods). red and light blue = residues that exchanged too fast to be measured only in A9 or M63F, respectively; pink = residues for which exponential decay was observed for both A9 and A9 M63F; grey = residues that exchanged too quickly to be observed for both A9 and M63F. A9 helices shown above (purple). Changes in exchange in e) mapped onto A9 crystal structure (PDB 1irj). Colors are the same as in e). Higher red intensity and larger tubes indicate residues that are more protected in A9 M63F. Positions 63 and 37 are indicated in black text. Calcium ions colored in dark blue.

M63F stabilizes A9 through a direct interaction with a nearby residue

We next asked how M63F locally stabilizes A9. We examined residues in the region of A9 that is stabilized by M63F and identified a phenylalanine at position 37 that is stabilized by M63F in NMR HDX experiments (Figure 3.4f). In both the A9 structure and in molecular dynamics simulations (unpublished data), F37 is well-positioned to interact with M63F. We reasoned that the increase in stability induced by M63F could be due to a pi-stacking interaction between M63F and F37. We therefore mutated F37 to a leucine, which is the ancestral state for this amino acid in the last common ancestor of amniotes and the most common amino acid at this position within calgranulin S100s.[130,131] We introduced F37L into both A9 and A9 M63F to produce A9 F37L and A9 M63F/F37L.

We tested the stability of the A9 F37L and A9 M63F/F37L variants using chemical denaturation and CD spectroscopy. In the absence of calcium, we observed no change in stability between wildtype A9 or any of the A9 variants (Figure 3.5a). We verified that the A9 variants were homodimers under these conditions using SECMALS (Figure AB1). This indicates that the F37L mutation does not appreciably affect the intrinsic stability of each monomer or the A9 dimerization constant, similarly to M63F.

In the presence of calcium, we found that both A9 F37L and A9 M63F/F37L exhibited wildtype A9 levels of stability (Figure 3.5b-c). This shows that mutating F37 in the A9 M63F background (A9 M63F/F37L) completely negates the stabilizing effect of M63F. We used the unfolding free energies for A9 and the three A9 variants to calculate the interaction energy between F63 and F37 (Figure 3.5d).

In the absence of calcium, $\Delta\Delta G_{F\text{-}F} = 0.43 \pm 0.66$ kcal/mol, affording no appreciable increase in stability. However, upon addition of 0.5 mM or 5 mM calcium, $\Delta\Delta G_{F\text{-}F} = 5.16 \pm 1.01$ and 6.13 ± 3.11 kcal/mol, respectively (Figure 3.5d, AB4). This shows that M63F stabilizes A9 by forming a strong interaction with F37. Mutating either residue in the F-F interaction is sufficient to restore wildtype stability.

We confirmed that mutating either phenylalanine in the F-F interaction negates the local stabilizing effect of M63F using HDX measurements by mass spectrometry. We monitored HDX for wildtype A9 and the three variants over the course of 24 hours, quenching the reaction at each timepoint using chemical denaturation at low pH. We then digested each exchanged protein sample using trypsin and quantified the number of deuterium ions present per peptide by mass spectrometry (Figure AB5). We calculated the percent deuterium uptake for each peptide and determined the change in percent uptake between A9 and the three variants (Figure 3.5e, AB5). Mimicking our findings by NMR, we found that M63F strongly decreased A9 hydrogen-deuterium exchange. In contrast, both A9 F37L and A9 M63F/F37L exhibited hydrogen-deuterium exchange that was comparable to or higher than A9 (Figure 3.4e), indicating that the F37L mutation locally destabilizes both wildtype A9 and A9 M63F. These findings further show that M63F primarily stabilizes A9 through a strong interaction with F37.

A9 variants bind calcium with high affinity

One explanation for why the M63F-F37 interaction stabilizes A9 is that it might alter A9 calcium binding, as M63F causes significant changes in chemical

environment (Figure 3.2) and stability (Figure 3.4, 3.5a-e) near A9 calcium binding sites. To test this possibility, we measured calcium binding by isothermal titration calorimetry (ITC) for A9 and the three A9 variants (Figure AB6-8). A9 monomers have two calcium binding sites – one with lower affinity between residues 23-36, and one with high affinity between residues 67-78.[170,196] We therefore fit a two-site binding model to the data (Figure AB6-8). We estimated $K_{d1} < 100$ nM for A9 and each A9 variant. However, due to the complexity of fitting a two-site binding model in which the sites have different binding affinities, we were unable to reliably quantify K_{d2} (Table AB7). Fitting a single-site model to each dataset yielded an identical K_d to the K_{d1} calculated in two-site models (Table AB7). We also found that the fraction competent (Ffx) parameter, which is used in ITC fitting to estimate the fraction of protein sample that is binding-competent, decreased for each A9 variant compared to wildtype A9. Ffx for wildtype A9 was estimated to be 0.98, while Ffx decreased to 0.21, 0.42, and 0.58 for A9 M63F, A9 M63F/F37L, and A9 F37L, respectively. Lastly, we were able to resolve a small, second binding transition only for A9 M63F. We separately fit this transition using a single-site binding model and estimated the $K_d = 741 \pm 134$ nM (Figure AB7-8). These findings suggest that calcium binding is modestly perturbed in each of the A9 variants. However, this perturbation does not correlate with changes in A9 stability, and each variant appears to retain sub-micromolar calcium binding affinity.

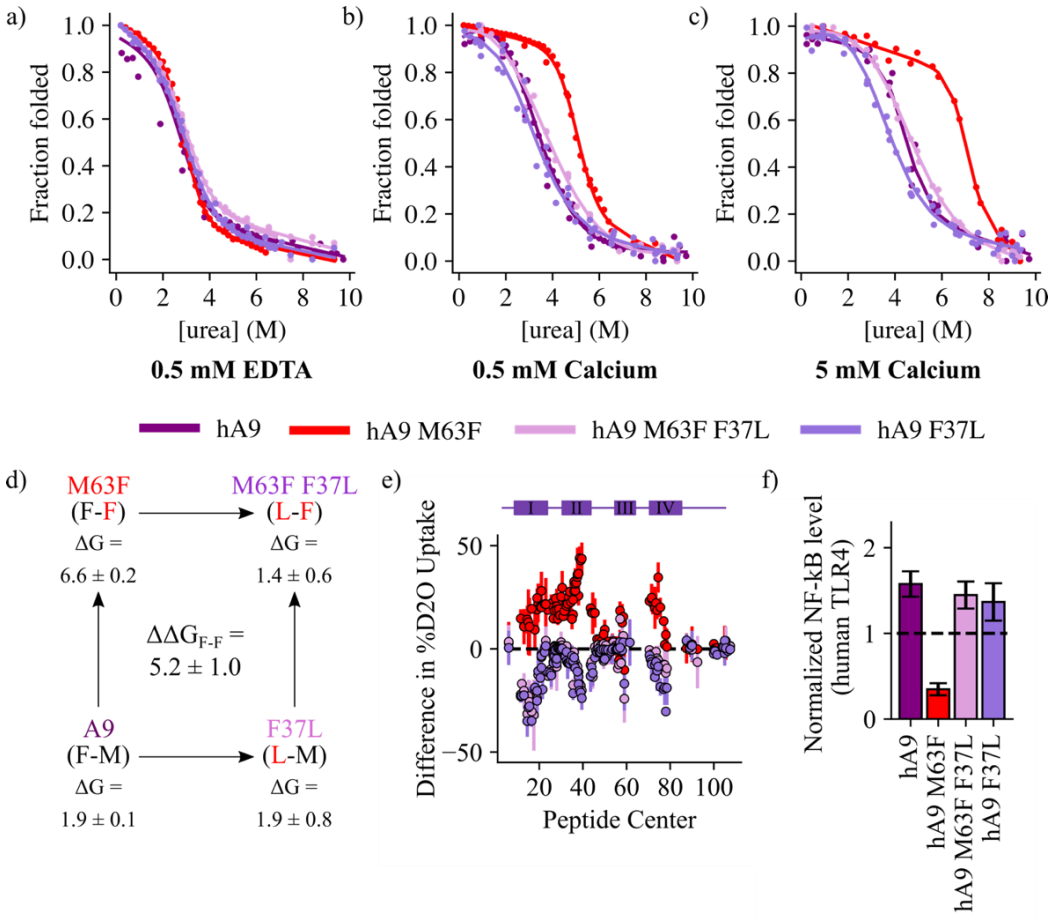


Figure 3.5. Disrupting the M63F-37F interaction restores A9 stability and TLR4 activation. a-c) Equilibrium urea denaturation of A9 (dark purple), A9 M63F (red), A9 M63F F37L (pink), and A9 F37L (light purple) with a) 0.5 mM EDTA, b) 0.5 mM $CaCl_2$, or c) 5 mM $CaCl_2$ present. Points are biological replicates and solid lines are a two-state unfolding model fit to the data (see methods). Free energy values are in kcal/mol. d) Mutant cycle thermodynamic values for A9 and three mutants at 500 uM $CaCl_2$. e) Differences in deuterium uptake from wildtype A9 after 1 hour for three A9 mutants measured by HDX-MS (see methods). Points represent one biological replicate, errors bars are standard error. f) Differences in TLR4 activation for A9 and three mutants. Colors same as in a-d). Data represent > 3 biological replicates, error bars are standard error of the mean.

Removing the F-F interaction fully restores A9 proinflammatory activity

Finally, we asked whether the M63F-F37 interaction is the mechanism by which M63F decreases A9 proinflammatory activity. If the F-F interaction is the primary inhibitor of A9 M63F proinflammatory activity, then removing this

interaction – via A9 M63F/F37L – should restore it. Alternatively, the M63F mutation could decrease A9 proinflammatory activity through another mechanism. To test this, we measured TLR4 activation by the A9 variants using a previously described a cell culture assay (Figure 3.5f, AB9). Strikingly, we found that both A9 F37L and A9 M63F/F37L activated TLR4 similarly to wildtype A9. This shows that M63F inhibits A9 proinflammatory activation of TLR4 by forming a strong pi-stacking interaction with nearby residue F37 to locally stabilize the protein.

DISCUSSION

M63F decreases A9 proinflammatory activity through a stabilizing molecular "staple"

This work reveals the mechanism by which a single mutation decreases A9 proinflammatory activation of TLR4. Lacking a mechanistic understanding of how A9 activates TLR4 has hampered the design of strategies for treating A9-induced chronic inflammation. The A9:TLR4 binding interface is unknown, and until recently, no mutation had been identified that alters A9 proinflammatory activity. Here we show that the previously identified M63F mutation decreases A9 activation of TLR4 by directly interacting with nearby F37, resulting in a strong local increase in stability. Disrupting this interaction by mutating either phenylalanine restores wildtype A9 stability and fully rescues A9 activation of TLR4, revealing that the F-F interaction is the primary determinant by which the M63F mutation inhibits A9 function. This finding provides valuable mechanistic

insight into how A9 activates TLR4 to drive inflammation, which we expand upon below.

A9 might activate TLR4 through a non-native conformation

We propose two possible explanations for how the F-F interaction inhibits A9 proinflammatory activity. The first is that the F-F interaction disrupts the native structure of A9. This could occur by stabilizing a misfolded intermediate, unfolding portions of the protein, or disrupting calcium binding. Alternatively, rather than disrupting the native structure of A9, the F-F interaction could restrict A9 from accessing a functionally important non-native conformation.

Supporting the possibility of a structural change, we do observe differences in chemical environment for several A9 residues near M63F by NMR. This suggests that M63F causes a local structural change. Multiple lines of evidence, however, suggest that the F-F interaction does not significantly alter A9 structure. We observe no differences in oligomeric state or bulk secondary structure between A9 and the three A9 variants, indicating no gross changes in protein folding or secondary structure content. By NMR, the majority of residues in A9 maintain their chemical environment upon introduction of M63F. The residues that do undergo a change in chemical environment are those near the M63F mutation, so it is unsurprising that these residues show larger chemical shift differences. This finding does not rule out a small, local structural change induced by M63F; however, it does further suggest that M63F does not induce a large structural change in the protein. Lastly, The A9 M63F/F37L and A9 F37L variants have the same stability

and activity as wildtype A9, showing that single substitutions at positions 63 and/or 37 alone do not functionally disrupt the protein. While we cannot explicitly rule out that the F-F interaction does not alter A9 structure, the majority of evidence suggests that it is not appreciably affected. To fully address this possibility, we are currently working to determine the structure of A9 M63F.

The F-F interaction could perturb calcium binding, as the largest chemical shift changes observed in A9 M63F by NMR occur near calcium binding sites. We do observe modest differences in calcium binding for each of the A9 variants by ITC. In particular, we resolved a second binding transition in A9 M63F with an estimated K_d of ~740 nM, whereas all other estimated K_ds for each variant are < 100 nM. This suggests that the the F-F interaction might decrease calcium binding affinity at one of the two calcium binding sites in A9. However, all experiments were performed at millimolar calcium concentrations to mimic the high concentrations of calcium present in the extracellular space. Under these conditions, even if the F-F interaction decreases calcium binding affinity from low to high nanomolar, A9 M63F will remain saturated with calcium. Further, the F-F interaction stabilizes A9 in a calcium-dependent manner, suggesting that calcium binding is not significantly altered. Previous work revealed that the A9 M63F mutant robustly forms a functional heterotetramer with S100A8, which is a known calcium-dependent process.[131] A9 proinflammatory function is also robust to other mutations, as A9 retained activity when mutating all zinc-binding residues simultaneously (residues 20, 30, 91, 95, 103-105) or when truncating the disordered C-terminal tail (residues 100-114).[111] These data together suggest that the F-F

interaction does not decrease A9 proinflammatory activity by disrupting calcium binding.

We propose, instead, that the F-F interaction prevents A9 from accessing a non-native conformation that is necessary to activate TLR4. Our data show that the F-F interaction strongly stabilizes nearby residues; the F-F interaction could prevent A9 binding to TLR4 by restricting access to these residues. In support of this, mutating charged amino acids between residues 64 -77 – near M63F – has been shown to decrease A9 binding to TLR4 *in vitro.*[109] This suggests that the F-F interaction decreases A9 activation of TLR4 by inhibiting A9 from accessing a binding-competent conformation or rearrangement.

A non-native mode of activation could explain why determining how A9 activates TLR4 has been challenging. A9 could adopt a non-native structure ranging from a simple conformational change – as is common amongst S100 proteins upon calcium binding[170,197] – to an exotic unfolding event or rearrangement, which would likely be necessary for A9 to mimic LPS and bind within the hydrophobic pocket of MD2. Our findings, which are consonant with other groups,[109] suggest that the local region of A9 surrounding the F-F interaction contains residues that are important for A9 activation of TLR4. We propose that this region of A9 should be a primary focus in ongoing studies to determine the exact nature of the state through which A9 activates TLR4 to drive inflammation.

Stability/function tradeoffs could be common in DAMPs

This work highlights how destabilizing protein structure can be important for conferring new protein functions. Stability/function tradeoffs are well-described in enzymes – for example, enzyme stability can decrease to increase flexibility, ease product release, and increase catalytic efficiency.[60,62,65,66,198] However, stability/function tradeoffs have not been broadly described for damage-associated molecular patterns (DAMPs) such as A9, which activate the innate immune system in response to tissue damage.[92,151,153,185,186] This is because the mechanisms by which many DAMPs elicit an immune response and drive inflammation are unknown. DAMPs often live double lives; many perform normal intracellular functions until tissue damage occurs, leading to DAMP release and induction of an immune response.[185,186] Hydrophobic surface exposure has been proposed as a mechanism by which DAMPs bind to immune receptors to trigger the immune system.[199] In the case of A9, we previously showed that A9s gained proinflammatory activity while simultaneously losing proteolytic resistance.[131] This suggests that DAMPs might have lost stability over time, allowing for exposure of core hydrophobic residues or other conformational rearrangements – such as those proposed here for A9 – to signal tissue damage to receptors like TLR4. Determining the role of stability and structural rearrangements in DAMP function will be critical for understanding how DAMPs mediate the innate immune response and for targeting pathological DAMPs to treat inflammatory disease.

MATERIALS AND METHODS

Cloning and mutagenesis

All S100 genes in this study were purchased as synthetic constructs in pUC57 vectors from Genscript. S100 genes were sub-cloned into a pETDuet-1 (pD) vector (Millipore) as described previously.[131] Cysteine-free versions of all S100 genes, as well as point mutants, were prepared using site-directed mutagenesis (Agilent).

Protein expression and purification

Recombinant protein overexpression was conducted in *E. coli* BL21 (DE3) pLysS Rosetta cells. Cultures were inoculated in luria broth overnight at 37°C, shaking at 250 rpm, in the presence of ampicillin and chloramphenicol. The following day, 10 ml of saturated culture was diluted into 1.5 L of media with antibiotics, grown to OD_{600} = 0.6 – 1, and then induced overnight at 16°C using 1 mM IPTG. Cells were pelleted at 3,000 rpm for 20 min and stored at -20°C for no more than three months.

Lysates were prepared by vortexing pellets (3-5 g) in tris buffer (25 mM tris, 100 mM NaCl, pH 7.4) and incubating for 20 min at RT with DNAse I and lysozyme (ThermoFisher Scientific). Lysates were sonicated and cell debris was pelleted by centrifugation at 15,000 rpm at 4°C for > 20 min. All proteins were purified on an Äkta PrimePlus FPLC using various 5 ml HiTrap columns (HisTrap FF (Ni-affinity), Q HP (anion exchange)). All S100s were purified in three steps using Ni-affinity chromatography in the presence of calcium followed by two

rounds of anion exchange chromatography at different pHs. For Ni-affinity chromatography, proteins were eluted over a 50 ml gradient from 25-1000 mM imidazole in tris buffer. Peak elution fractions were pooled and placed in dialysis overnight at 4°C in 4 L of tris buffer (calcium-free) adjusted to pH 8. Anion exchange chromatography was then performed the following day over a 50 ml gradient from 100-1000 mM NaCl in pH 8 tris buffer. Fractions containing majority S100 were pooled and analyzed for purity on an SDS-PAGE gel. If trace contaminants remained, an additional anion exchange step was performed at pH 6 using the same elution strategy as for the previous anion exchange step. All purified proteins were dialyzed overnight at 4°C in tris buffer + 2 g/L Chelex-100 resin (Biorad), flash-frozen the following day in liquid nitrogen, and stored at -80°C.

Biophysical and biochemical characterization

For all experiments, protein aliquots were thawed fresh from freezer stocks and were either dialyzed in the appropriate experimental buffer overnight at 4°C or exchanged 3X into experimental buffer using 3K microsep spin concentrator columns (Pall Corporation). All samples were filter-sterilized using 0.1 um spin filters (EMD Millipore) prior to measuring concentration and using in experiments. Thawed aliquots were used for no more than one week before discarding. All concentrations were measured by absorbance of protein in > 6M urea at 280 nm and correspond to micromolar dimeric protein.

Oligomeric state characterization by SECMALS

Oligomeric states were measured using a superose 12 10/300 GL size exclusion column (Amersham Biosciences) with in-line concentration detection using refractive index (RI) and particle mass measured using a multiangle laser light scattering (MALS) instrument (Dawn Heleos, Wyatt Technology). Samples were concentrated to 0.5-2 mg/ml in tris buffer + 0.5 mM $CaCl_2$, 0.1 um sterile-filtered, and analyzed at a flow rate of 0.2 ml/min. Data were processed using manufacturer's software (Astra).

CD Spectroscopy and chemical denaturation studies

Circular dichroism (CD) and chemical denaturation experiments were performed using a Jasco J-815 CD spectrometer and spectroscopy-grade guanidine hydrochloride (gdn-HCl) or urea. Chemical denaturation was performed in tris buffer with $CaCl_2$ or EDTA, with tris substituted for spectroscopy-grade trizma. Reversible unfolding and refolding curves were constructed by making concentrated 100 uM protein stocks in either buffer or 6M gdm or 10M urea and then preparing protein dilutions in various concentrations of gdn-HCl or urea in buffer. Samples were left to equilibrate in denaturant overnight and were then analyzed by CD. Unfolding/refolding equilibration was confirmed by comparing unfolded vs. refolded protein at the same concentration. CD signal was quantified at 222 nm in a 1 mm cuvette using a 1 nm bandwidth, standard sensitivity, and 2 second D.I.T. HT voltage was < 600 V. We fit a two-state unfolding model:

$$\frac{b_f + m_f x + (b_u + m_u x) e^{-\frac{\Delta G - mx}{RT}}}{1 + e^{-\frac{\Delta G - mx}{RT}}}$$

to the data to extract thermodynamic parameters, where $b_f, m_f, b_u,$ and m_u are the folded and unfolded baseline y-intercepts and slopes, ΔG is the unfolding free energy, m is the m-value, R = 0.001987 JxK^{-1}mol^{-1} and T = 298.15 K. Standard deviations were calculated from fits by taking the square root of the diagonalized covariance matrix and by error propagation.

Calcium binding measurements by ITC

ITC experiments were performed in 25 mM tris, 100 mM NaCl, pH 7.4 at 25°C. To ensure removal of calcium, protein samples were treated with 5 mM EDTA at room temperature for 1 hour and then exchanged 10X into experimental buffer to remove EDTA. Calcium concentration was confirmed to be < 10 uM per sample for 50 uM experimental protein samples by ICP-MS. Samples were equilibrated and degassed by centrifugation at 18, 000 × g at the experimental temperature for 30 minutes. All experiments were performed at on a MicroCal ITC-200 or a MicroCal VP-ITC (Malvern). Gain settings were determined on a case-by-case basis. A 750 rpm syringe stir speed was used for all ITC-200 experiments while 400 rpm speed was used for experiments on the VP-ITC. Spacing between injections ranged from 300s-900s depending on gain settings and relaxation time of the binding process and were optimized for each measured binding interaction. Titration data were globally fit to one- and two-site binding models using the Bayesian fitter in pytc (Figure AB7-8).

NMR experiments

Samples were prepared in 25 mM tris, 100 mM NaCl, 10 mM CaCl2, pH 7.4, 10% D2O. All protein concentrations ranged from 0.5 – 1 mM dimer. We collected 2D $^{1}H-{}^{15}N$ TROSY-HSQC NMR spectra and transferred published assignments to the hA9 spectrum (BMRB: 30017, PDB 5I8N), and then used 3D NOESY-TROSY spectra to verify the assignments. We unambiguously assigned 91 peaks of the 113 non-proline amino acids. All NMR experiments were performed at 25°C or 37°C on an 800 MHz (18.8T) Bruker spectrometer at Oregon State University. TROSY spectra were collected with 32 transients, 1024 direct points with a signal width of 12820, and 256 indirect points with a signal width of 2837 Hz in ^{15}N. NOESY-TROSYs were run with 8 transients, non-uniform sampling with 15% of data points used, and a 150 ms mixing time. All spectra were processed using NMRPipe;[200] data were visualized and assignments transferred using the CCPNMR analysis program.[201]

Relaxation Dispersion (RD) experiments were carried out using the pulse sequence developed by Kay and coworkers using a constant time (CT) of 50 ms at a field strength of 800 MHz with data collected at relaxation delays of 9.55, 0.25, and 0.0 (I_0) ms.[202] TROSY HSQC spectra were collected in an interleaved manner with 16 transients, a 90 ms T1 period, and a 2.5 second delay between transients for each relaxation delay at 19°C. Steady state hNOE values are a measurement of backbone flexibility on the ps-ns time scale. NOEs were calculated as the ratio of 1H-15N correlation peak volume in the spectra acquired with and without the proton saturation. {1H}-15N steady-state NOEs were obtained by recording

spectra with and without 1H presaturation, a duration of 3 s and a relaxation delay of 6 s.[203]

For hydrogen-deuterium exchange (HDX) experiments by NMR, protein samples were lyophilized and resuspended in 100% D2O immediately prior to beginning to measure 2D $^{1}H-^{15}N$ TROSY-HSQC spectra. Spectra were collected over 24 hours – every ~11 minutes for the first hour, then every hour up to ~8 hours, and then final spectra were measured the following day at 24 hours. Peak intensities (after NMRPipe processing) were quantified using Sparky.[204] We fit either a single exponential model or a linear model to each dataset, using the Akaike Information Criterion (AIC) to determine whether a single exponential decay model or a linear model was more appropriate (Figure AB2-3). All residues that exchanged too rapidly to be measured are reported as having an exchange rate faster than the fastest-exchanging residue that could be reliably quantified using an exponential model that passed the AIC filter.

Hydrogen-deuterium exchange mass spectrometry studies

Hydrogen exchange mass spectrometry (HX-MS) experiments were carried out using a LEAP PAL autosampler (LEAP Technologies, Carrboro, NC, USA).10 uM protein stocks were maintained at 4˚C. 3 uL of 10 uM protein sample was diluted into 27 uL of deuterated buffer at 20˚C. After a varying labeling time, 27 uL of labeled sample was quenched with 27 uL of 3.5M GdmCl , 1.5M Glycine pH 2.4 at 1˚C. 45 uL of quenched sample was then immediately injected into an Thermo UltiMate 3000 UHPLC system (Buffer A, Buffer B). For inline digestion,

samples were flowed over two immobilized protease columns (Upchurch C130B), one with fungal protease and one with pepsin (conjugated to POROS 20 AL Aldehyde Activated Resin, Thermo Scientific), at a flow rate of 100 uL/min of buffer A. Digested protein was then desalted on a trap column (Upchurch C-128 with POROS R2 beads) for 2 minutes at a flow rate of 300 uL/min. An acetonitrile gradient (10% buffer B to 60% buffer B over 9 minutes) eluted peptides from this trap column onto an analytical C-8 column (Thermo 72205–050565) for separation before injection into an ESI source for mass spectrometry analysis on a Thermo Q Exactive mass spectrometer operating in positive ion mode.

For every protein variant, a non-deuterated control was subjected to the same LC/MS protocol as the deuterated samples and the resulting peptides were identified using tandem MS (MS/MS). Peptide precursor spectra were acquired in data-dependent mode with the top 10 most abundant ions (charge state >= 2, <= 6) selected for fragmentation and product ion analysis. Following MS/MS acquisition precursor ions were excluded from further fragmentation for 4 seconds. The peptide identification software Byonic (Protein Metrics Inc.) was used to identify peptides from the tandem MS data. Following peptide identification deuterium incorporation for each peptide at each labeling time was determined by centroid analysis with the software HDExaminer (Sierra Analytics).

TLR4 activity assay and cell culture conditions

We purchased commercially distributed HEK293T cells from ATCC (CRL-11268). Because we are using this cell line as a host for heterologous transient

transfections, the appropriate control for consistency between assays is the measurement of reporter output for a set of control plasmids and a panel of known treatments. Upon thawing each batch of cells, we run a positive control for ligand-induced response. We transfect the cells with plasmids encoding human CD14, human MD-2, human TLR4, renilla luciferase behind a constitutive promoter, and firefly luciferase behind an NF-KB promoter. We then characterize the raw luciferase output for five treatments: 1) mock, 2) LPS, 3) LPS + polymyxin B, 4) S100A9 + polymyxin B, and 5) S100A9 + 1.25x polymyxin B. This has a stereotypical pattern of responses in renilla luciferase (high for all) and firefly luciferase (low, high, low, high, high). To validate that this response is dependent on the transfected TLR4 complex as opposed to the cells themselves, we repeat the experiment but exclude the TLR4 plasmid. This should give identical renilla luciferase values but no firefly luciferase output in response to any treatment. To ensure that the cells maintain their properties between passages, we repeat the mock, LPS, and LPS + polymyxin B control on every single experimental plate. This assay has a built-in control for mycoplasma contamination: high firefly luciferase signal in the absence of added agonist. This indicates that there is another source of TLR4-induced NF-kappa B output in the cells—most plausibly, contamination. This mycoplasma sensing approach is used in the commercially available HEK-BLUE mycoplasma detection kit (Invivogen). We discard any cells that exhibit high background values or reach 30 passages.

All plasmids, cell culture conditions, and transfections for measuring the activity of S100s against TLR4s were identical to those previously described.[111,130,131,172]

Briefly, human embryonic kidney cells (HEK293T/17, ATCC CRL-11268) were maintained up to 30 passages in Dulbecco's Modified Eagle Media (DMEM) supplemented with 10% fetal bovine serum (FBS) at 37°C with 5% CO2. Lipopolysaccharide *E. coli* K-12 LPS (LPS - tlrl-eklps, Invivogen) aliquots were prepared at 5 mg/ml in endotoxin-free water and stored at -20°C. Working solutions were prepared at 10 ug/ml and stored at 4°C to avoid freeze-thaw cycles. S100 proteins were prepared by exchanging into endotoxin-free PBS and incubating with an endotoxin removal column (Thermo Fisher Scientific) for 2 hours. S100 LPS contamination was assessed by measuring activity with and without Polymyxin B, an LPS chelating agent (Figure AA6). LPS (200 ng per 100 ul well) or S100 (0.8, 0.4, 2, 4, or 5 uM dimer) treatments were prepared by diluting in 25:75 endotoxin-free PBS:serum-free Dulbecco's Modified Eagle Media (DMEM – Thermo Fisher Scientific). Polymyxin B (PB, 200 ug per 100 ul well) was added to all S100 experimental samples to limit background endotoxin contamination activity from recombinant protein preps. Cells were incubated with treatments for 3 hours prior to assaying activity. The Dual-Glo Luciferase Assay System (Promega) was used to assay Firefly and Renilla luciferase activity of individual wells. Each NF-κB induction value shown represents the Firefly luciferase activity divided by the Renilla luciferase activity, background-subtracted using the LPS + PB activity for each TLR4 species and normalized to the activity of LPS alone for each TLR4 species to normalize between plates. All measurements were performed using three technical replicates per plate, a minimum of three biological replicates total, and a minimum of two separate protein preps.

BRIDGE TO CHAPTER IV

In Chapter III, we determine the mechanism by which a previously identified point mutation (M63F) decreases A9 activation of TLR4. We find that M63F interacts with a nearby residue to locally stabilize A9 using a variety of biophysical approaches. We show that disrupting either half of the interaction restores wildtype A9 stability and proinflammatory activation of TLR4. We suggest that this "molecular staple" does not disrupt A9 proinflammatory activity by altering its structure. Instead, we hypothesize that the interaction prevents A9 from accessing a non-native excited state conformation that is necessary for A9 activation of TLR4. We discuss the implications of this hypothesized novel mechanism for understanding A9 proinflammatory activity, treating chronic inflammation driven by A9, and expanding our understanding of the mechanisms by which damage-associated molecular patterns (DAMPs) trigger the innate immune system. In chapter IV, we highlight ongoing and future studies focused on determining how A9 proteins evolved to potently and promiscuously activate TLR4 and identifying where A9 interacts with TLR4, MD2, and CD14.

CHAPTER IV

MOLECULAR MAPPING OF CHANGES IN S100A9 PROINFLAMMATORY POTENCY, PROMISCUITY, AND THE S100A9:TLR4 INTERACTION

AUTHOR CONTRIBUTIONS

Joseph Harman, Lauren Lehmann, Sophia Phillips, and Michael Harms conceptualized the studies and designed the experiments. Michael Harms acquired the funding for the study. Joseph Harman, Lauren Lehmann, and Sophia Phillips performed the experiments. Michael Harms administered the project. Joseph Harman, Lauren Lehmann, Sophia Phillips, and Michael Harms analyzed the data. Joseph Harman and Michael Harms constructed the figures. Joseph Harman wrote this excerpt, with editorial assistance from Michael Harms. Lauren Lehmann and Sophia Phillips contributed to the methods section.

INTRODUCTION

A primary job of the innate immune system is to respond to damage and infection.[205,206] Pattern recognition receptors (PRRs) initiate a potent immune response by recognizing both exogenous pathogen molecules and endogenous signs of tissue damage and subsequently activating inflammation.[207–209] The mechanisms by which PRRs recognize exogenous pathogen-associated molecular patterns (PAMPs) and activate downstream inflammatory pathways are reasonably well understood. Comparatively less is known about how endogenous damage-associated molecular patterns (DAMPs)[185,186] activate PRRs to amplify

inflammation. Prolonged activation of inflammation via DAMPs contributes to many human inflammatory diseases.[184–186,210] Determining the mechanisms by which DAMPs drive inflammation through PRRs is critical to our understanding of innate immunity and for developing treatments for chronic inflammation.

TLR4 is a well-studied PRR that recognizes multiple PAMPs and DAMPs to activate inflammation.[117,121,122,176] The mechanism by which TLR4 is activated by PAMPs, such as the bacterial cell wall component lipopolysaccharide (LPS), is well understood.[115,117,121,122,125] LPS is delivered to a hydrophobic pocket of the TLR4 cofactor MD2 by a second cofactor, CD14.[115–117] LPS delivery promotes formation of the $(TLR4/MD2)_2$ heterotetramer and results in activation of downstream proinflammatory signaling pathways.[120,121] CD14 and MD2 are required for TLR4 activation by LPS across amniotes.[130]

In contrast, TLR4 activation by endogenous DAMPs is poorly understood. S100A9 (A9) is a highly abundant neutrophil DAMP that, upon release during an immune response, potently activates inflammation through TLR4.[86,91–95] Little is known about the mechanism by which A9 activates TLR4 to drive inflammation. Similarly to LPS, CD14 and MD2 are required for A9 activation of TLR4.[130] This suggests that LPS and A9 activate TLR4 through a similar mechanism. However, unlike LPS, the native state structure of A9 is too large to fit into the MD2 hydrophobic pocket. Smaller proteolytic fragments of A9 have thus been proposed to activate TLR4, as A9 is readily degraded by proteases and smaller A9 fragments could mimic LPS.[109] We previously showed, however, that proteolytic degradation is not a requirement for A9-induced TLR4 activation.[131] Ongoing work in our group

suggests that A9 might activate TLR4 through a non-native excited state conformation, but the exact nature of this proposed A9 excited state – and its binding interface to TLR4 – are unknown.

Evolutionary analyses of A9 proinflammatory activity have provided new insight into the mechanism by which A9s evolved to activate TLR4. We recently showed that mammalian A9s evolved potent proinflammatory activity from a weakly active amniote ancestor.[131] We identified a single substitution that played a key role in the evolution of A9 proinflammatory activity. Earlier work revealed that later-diverging (eutherian) A9s activate TLR4s from multiple species more potently and promiscuously than earlier-diverging (therian) A9s.[130] Complementation experiments – in which TLR4 components from different species were systematically swapped and tested for activity – have revealed subtle differences in TLR4/MD2/CD14 requirements across mammalian A9s compared to LPS, suggesting overlapping but different molecular mechanisms for TLR4 activation by A9 and LPS. These findings together present new avenues for determining how the abundant neutrophil DAMP A9 evolved to drive inflammation.

This study outlines ongoing and prospective work to dissect the mechanism by which A9s evolved to activate TLR4s in mammals. We highlight two complementary approaches – one using evolutionary biochemical analyses and the other using a high-throughput mutational screen – to determine how A9s evolved to activate TLR4 and identify the A9:TLR4/MD2/CD14 interaction interface. Preliminary data suggests that important changes in A9 proinflammatory activity

occurred over the narrow evolutionary interval between the therian and eutherian A9 ancestors. Of the amino acid differences between these two ancestral states of A9, we find that nearly half cluster within a small region of A9 that was previously proposed to bind to TLR4 *in vitro.*[109] We show preliminary efforts to determine which of these substitutions are important for the evolution of A9 proinflammatory activity. Finally, we reveal the results of a pilot cell sorting and sequencing experiment to identify mutants of MD2 that alter LPS and/or A9 proinflammatory activity. This complementary high-throughput approach shows promise for determining the A9:TLR4 binding interface and can be readily employed on A9s, TLR4s, MD2s, and CD14s from different species. Combining these orthogonal strategies will enable determination of the mechanisms by which A9s activate TLR4s to drive inflammation in mammals. These studies will increase our understanding of how DAMPs evolved key roles in mammalian innate immunity and provide a mechanistic basis for developing strategies to alleviate chronic inflammation.

RESULTS

A9s evolved enhanced proinflammatory activity in eutherian mammals

We began by asking why later-diverging A9 proteins activate TLR4s from various amniotes more strongly and promiscuously than earlier-diverging A9 proteins. Previous work showed that therian mammalian A9s evolved enhanced proinflammatory activity from a weakly active amniote ancestor (Figure 4.1, AC1).[130,131] While these findings revealed that the proinflammatory activity of A9s

improved in early mammals, they do not explain how later-diverging A9 proteins, such as human and mouse A9, evolved such potent activity (Figure 4.1, AC1). Further, there are many amino acid changes between human and mouse A9 and the therian mammalian A9 ancestor (33/99 and 42/99, respectively, excluding the disordered C-terminal tail – Figure AC2), making it difficult to determine which substitutions mattered for increasing A9 proinflammatory activity.

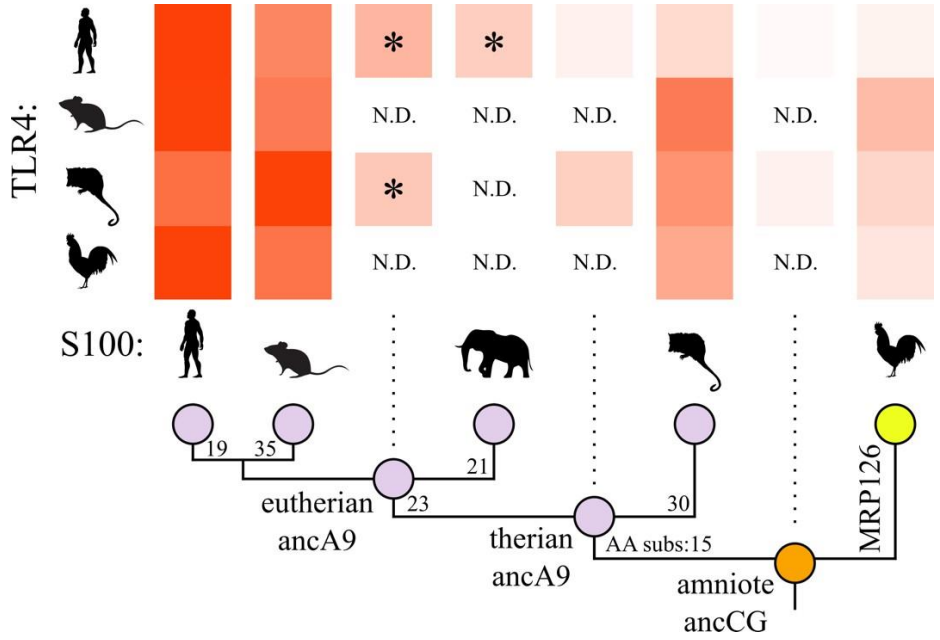


Figure 4.1. Later-diverging A9s activate amniote TLR4s more potently and promiscuously than earlier-diverging A9s. Data are synthesized from two previous studies;[130,131] * = preliminary data in this study. Species cartoons are human, mouse, elephant, opossum, and chicken. Chicken MRP126 is indicated in yellow and the amniote calgranulin ancestor is indicated in orange. A9 nodes are purple. Red intensities are scaled from 0-1 for each TLR4 species (table AC1). In cases where multiple values are reported, red intensity is the average value of reported measurements. N.D. = no data. Numbers in tree are the number of amino acid substitutions between nodes, excluding the poorly conserved C-terminal tail (residues 100-114).

To narrow the interval over which human and mouse A9 evolved potent proinflammatory activity, we measured the activities of two additional eutherian A9 proteins to compare to human and mouse A9. We selected elephant A9 because elephants are among the earliest-diverging eutherian species. We also used ancestral sequence reconstruction to infer the eutherian mammalian A9 ancestor (eutherian ancA9). Eutherian ancA9 was reconstructed with very high confidence (average posterior probability = 0.96), so an AltAll[171] control reconstruction was not necessary.

We first compared elephant A9 and eutherian ancA9 against human A9 and previously characterized therian ancA9[131] for human TLR4 activation. We found that elephant A9 and eutherian A9 had approximately 40% and 60% of the activity of human A9 against human TLR4, respectively (Figure 4.2a, AC1). Both proteins activated human TLR4 more strongly than the earlier therian A9 ancestor, reflecting the higher activity previously observed for other eutherian A9 proteins. We also tested eutherian ancA9 for activity against opossum TLR4 and found that it had activity comparable to therian ancA9 and lower than human A9 (Figure 4.2b). These findings indicate that some, but not all, of the increase in later-diverging A9 proinflammatory potency is accounted for in the amino acid changes in eutherian ancA9. This also suggests that the potent activity of human and mouse A9 might have evolved even later than the eutherian ancA9 ancestor.

Preliminary studies testing how eutherian A9s evolved increased proinflammatory activity

We next sought to identify the amino acid changes between therian ancA9 and eutherian ancA9 that led to increased proinflammatory activity in eutherian A9s. There are 23 amino acid differences between therian ancA9 and eutherian ancA9, 10 of which occur within a contiguous stretch of 14 amino acids in helix III of the protein (Figure 4.2c, AC2). We previously found that a substitution in helix III at position 63 played a key role in A9s evolving increased proinflammatory activity.[131] Ongoing work by our group, outlined in chapter III of this dissertation, suggests that instability in helix III of A9 might be important for A9 activation of TLR4. Lastly, work by others has shown that mutating pairs of charged amino acids within residues 64-77 altered human A9 binding to TLR4 *in vitro*.[109]

We hypothesized that charge-altering substitutions that occurred between the therian and eutherian A9 ancestors in helix III could be important for increasing eutherian A9 proinflammatory activity. We selected 5 charge-altering substitutions in this region and introduced them together into therian A9 to produce 5X therian A9 (N51K, D55N, P56E, A57K, Q64E; human A9 numbering – Figure 2c, AC2). We confirmed that 5X therian A9 was folded and had similar helical content to therian ancA9 using circular dichroism spectroscopy (Figure AC3). However, we observed no difference in activation of human TLR4 between 5X therian A9 and therian A9 (Figure 2d). This indicates one of two possibilities. The 5 selected substitutions could simply not be important for increasing eutherian A9 proinflammatory activity. Alternatively, we could have introduced a negative

epistatic interaction by introducing some, but not all, of the substitutions required to increase therian A9 proinflammatory activity. Determining which substitutions drove an increase in A9 proinflammatory activity between the therian and eutherian A9 ancestors is an area of ongoing study in our group.

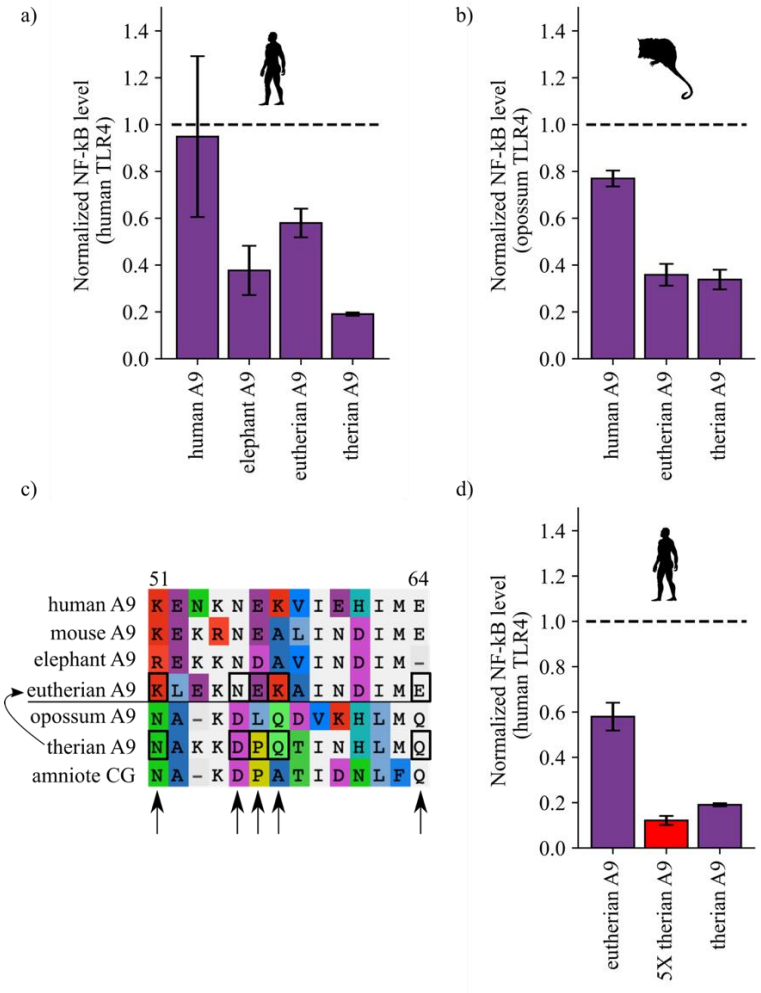


Figure 4.2. Elephant A9 and the reconstructed eutherian A9 ancestor activate human TLR4 better than earlier-diverging A9s. (a), (b), and (d) NF-κB production of human and opossum TLR4 in response to treatment with modern and ancestral S100 proteins. Bars represent average of >3 biological replicates, error bars are standard error of the mean. All values are background-subtracted and normalized to LPS positive control (see methods). (c) Sequence alignment of residues 51-64 of various A9 proteins. Colored residues are non-consensus sites. Boxes show the five mutations introduced into therian ("5X therian A9" – N51K, D55N, P56E, A57K, Q64E; human A9 numbering).

Mapping the A9:TLR4 interaction interface using a high-throughput mutant screen

The ongoing evolutionary studies of A9 described above focus on the A9 side of the A9:TLR4 interaction. Here we describe an orthogonal approach to examine the other half of the A9:TLR4 interaction by focusing on TLR4, MD2, and CD14. Understanding where A9 interacts with TLR4, MD2, and CD14 is necessary for determining how A9 activates TLR4 to drive inflammation.

To address this, we are developing a high-throughput mutant screen to determine where A9 interacts with TLR4, MD2, and CD14. The primary goal of the screen is to identify mutations in these TLR4 components that decrease A9 activation of TLR4 without altering LPS activity. Finding mutations that only decrease A9 activation, but not LPS activation, will allow us to filter out mutations that simply "break" the TLR4 complex.

The mutant screen is modified from a previously used cell-culture based TLR4 activation assay.[130,131,172] First, a mutant library of TLR4, MD2, or CD14 is produced using random mutagenesis, aiming for an average mutation rate of ~1. The mutant library is then transfected into HEK293T cells along with other necessary TLR4 components. This includes a GFP plasmid that reports on downstream NF-kB production, similarly to the luciferase reporter plasmid used in this study and in previous work (Figure 4.2). Transfected cells are treated with buffer, LPS, or A9 and sorted based on GFP fluorescence using fluorescence-activated cell sorting (FACS). Lastly, Illumina sequencing is used to determine which mutants are enriched in each treatment group. This approach is amenable to

testing mutant libraries of TLR4, MD2, and CD14, and can be applied to TLR4 components and A9s from a variety of species.

As a pilot study, we performed a screen of LPS vs. human A9 activating human TLR4 using a mutant human MD2 library. We chose to mutate human MD2 because it is small (~160 amino acids), amenable to short sequencing reads with high coverage, and has a structurally well-defined interaction with LPS.[114] We confirmed that an increase in GFP signal could be observed over background for cells transfected with the MD2 library and treated with LPS or A9 (Figure 4.3a-c). We then used FACS to sort cells treated with either buffer, LPS, or human A9 as GFP^+ or GFP^- (Figure 4.3d-e). Six uniquely barcoded Illumina libraries were prepared from each sorted condition using PCR and subjected to high-throughput sequencing using a MiSeq nano run. We measured a combined ~380,000 reads for the six libraries and the input MD2 mutant library. The MD2 library had an average of ~2.5 mutations per sequence, which matched the average number of mutations observed for each treatment group (Figure AC4). We filtered results to exclude sequences with an early stop codon, more than 10 mutations, or less than 6 total reads. After filtering, a total of 510 unique mutant sequences were found to be shared between all treatment conditions.

We then quantified enrichment for the 510 unique mutant sequences in the LPS vs. human A9 treatment conditions. We calculated the frequency of each clone in each library, and then compared enrichment in GFP^+ compared to GFP^- pools for LPS versus human A9 (Figure 4.3f). We focused on the top single mutants that were most significantly de-enriched for human A9 activity but remained the same

or were enriched for LPS activity (Figure 4.3, red). These are the top candidate mutations for decreasing A9 activation of TLR4 without altering LPS activation. For comparison, we also identified the top single mutants that were most significantly de-enriched for LPS activity but remained the same or were enriched for A9 activity (Figure 4.3, purple). We identified top single mutants with a de-enrichment score > $\log_2(1)$, an enrichment score < $\log_2(0.6)$, and a log2(de-enriched/enriched) score < -1 (Figure AC5). Mutant sequences with more than one mutation were discarded from this analysis to avoid having to account for epistatic interactions. Using these filters, we identified a total of 10 and 12 candidate mutations that showed A9 or LPS-specific de-enrichment, respectively. Three of these sites were the same position in both categories, although they were different amino acid mutations. These residues are therefore not shown in Figure 4.3 but are listed in supplemental figure AC5.

Mapping the remaining top A9-specific and LPS-specific mutations onto the TLR4/MD2 structure[114] reveals that they are scattered throughout MD2 (Figure 4.3g). As expected, LPS-specific mutations largely occurred near LPS and tended to cluster near the bottom of the MD2 hydrophobic pocket. No A9-specific mutations were inside the MD2 hydrophobic pocket, suggesting that A9 interacts with the exterior of MD2. Several A9-specific mutations were distant from LPS near the bottom of MD2, suggesting that A9 might interact with the bottom of MD2. Follow-up studies are required to validate the role of these residues in A9 activation of TLR4. This pilot study highlights the potential of an unbiased high-throughput

mutant screen for identifying where A9 interacts with TLR4 en route to determining the mechanism by which A9 activates TLR4 to drive inflammation.

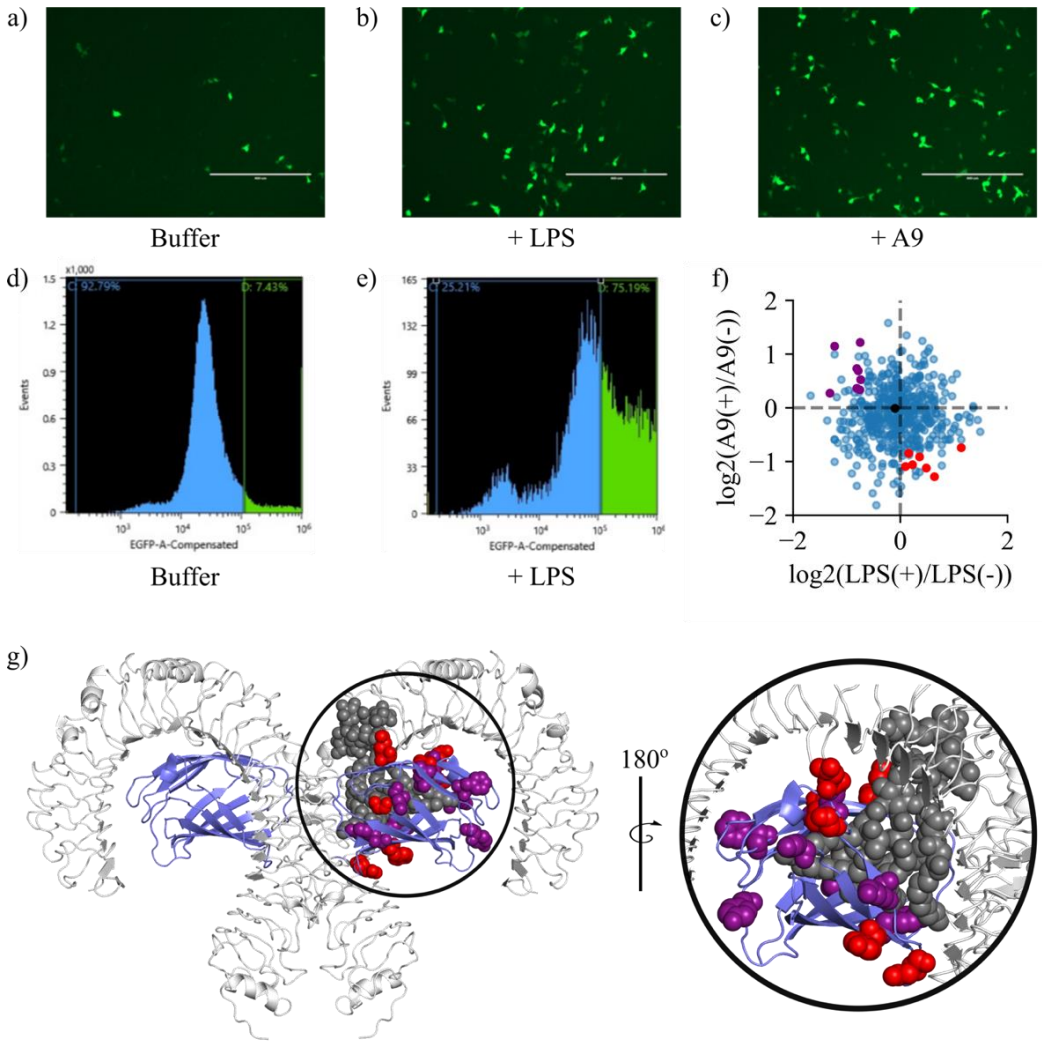


Figure 4.3. An MD2 mutant screen identifies candidate mutations that selectively decrease A9 activation of TLR4 without altering LPS activation. (a-c) Microscopy images of GFP signal in cells expressing mutant MD2 library and treated with (a) buffer, (b) LPS, or (c) human A9. (d-e) FACS of cells from (a) and (b), respectively. GFP^+ = green, GFP^- = blue. (f) Enrichment of MD2 mutants for A9 and LPS activity. Each point is a unique MD2 mutant containing at least one mutation. The most enriched mutants that decrease A9 GFP signal without decreasing LPS GFP signal are shown in red, opposite highlighted in purple (decrease LPS GFP signal without decreasing A9 GFP signal). Black point is wildtype MD2. (g) Mutants in (f) mapped onto TLR4 structure (PDB 3FXI) as red and purple spheres. TLR4 chains shown in white, MD2 chains shown in blue. LPS shown as spheres in grey. Circled zoom-in to the right shows back of MD2.

DISCUSSION

Later-diverging A9s have evolved to more potently and promiscuously activate TLR4s

The data presented here suggests that A9s have continued to evolve improved proinflammatory activity over time. Previous work showed that human and mouse A9 evolved to be highly potent activators of TLR4 from a modestly active therian mammalian ancestor.[130,131] Preliminary findings in this study suggest that some of these improvements, but not all, occurred between the therian and eutherian A9 ancestors. The eutherian A9 ancestor and elephant A9 each had higher activities than the earlier therian A9 ancestor against human TLR4 (Figure 4.1 & 4.2) but did not exhibit the same potency as human A9. This suggests that later-diverging A9s may have evolved increased proinflammatory activity in a stepwise fashion.

A9s do, however, exhibit complex patterns of both proinflammatory potency and promiscuity. For example, previous work showed that opossum A9 activates both human and opossum TLR4 better than the therian A9 ancestor.[131] In this study, eutherian A9 activates human TLR4 better than therian A9, but both proteins have equivalent activity against opossum TLR4 (Figure 4.1 & 4.2). This suggests that modern A9s in different species could have independently evolved enhanced proinflammatory activity from less active ancestors. This could be a product of A9s co-evolving with their species-matched TLR4s, resulting in increased proinflammatory activity evolving independently in human, mouse, and opossum A9. Further studies, including testing elephant A9 and eutherian ancA9 against additional TLR4s, are necessary for dissecting when and how A9s evolved

potent and promiscuous proinflammatory activity. Testing the supraprimate A9 ancestor (the shared ancestor of humans and mice) for TLR4 activation will also help explain the potency and promiscuity of human and mouse A9 and help determine whether patterns of potency and promiscuity are lineage-specific or have a common evolutionary origin.

A multi-pronged approach for determining the mechanism by which A9 activates TLR4

This study highlights the utility of combining evolutionary and biochemical approaches to determine mechanisms of protein functions. Determining the amino acid changes underlying how proteins change functions provides mechanistic insight into how protein functions work.[29] This strategy is unveiling a comprehensive picture of how A9s have evolved to drive inflammation in mammals. By mapping the proinflammatory activities of different A9s against different TLR4s onto their evolutionary history and comparing them at the amino acid level, we have isolated a region of A9 – helix III – that has acquired many substitutions over time and appears to be important for A9 proinflammatory function. Previous work by our group and others suggests that stability and charge distribution within helix III may be important for A9 TLR4 activation.[109,131] Future evolutionary and biochemical studies are required to fully understand the mechanism by which A9 activates TLR4, but this work suggests that we are focusing our efforts in the right direction.

Our pilot screen of MD2 mutations that alter A9 proinflammatory activity also highlights promising avenues for determining where A9 interacts with TLR4, MD2, and CD14. Previous work showed that A9s and LPS have overlapping, but subtly different, requirements for TLR4 activation.[130] The preliminary results presented here hint that A9 interacts with the exterior of MD2, as opposed to LPS, which inserts into the MD2 hydrophobic pocket. This makes sense if A9 maintains its native size and shape upon binding to TLR4; A9 is simply too big to mimic LPS binding within the MD2 hydrophobic pocket. However, the pilot screen also shows overlap between candidate residues that selectively alter A9 and LPS activity. This suggests that the A9 and LPS binding interfaces might overlap to some degree. Directly measuring the quantitative effects of the candidate MD2 mutations on A9 proinflammatory activity will be necessary to test these hypotheses. This approach can be readily applied to A9s, TLR4s, MD2s, and CD14s from multiple species, enabling future determination of differences in binding interface and ultimately activation mechanism by A9s across mammals. This will provide valuable context in studies employing, for example, mouse models to inform A9-induced inflammation in humans.

How did endogenous DAMPs evolve to drive inflammation?

Open questions remain regarding how endogenous damage-associated molecular patterns (DAMPs) such as A9 evolved to signal tissue damage to the innate immune system. Many DAMPs perform normal functions inside of healthy cells – and are thus "invisible" to the immune system – until cell damage occurs,

leading to DAMP release and induction of an innate immune response.[184–186] It remains unclear whether DAMP function in proteins is accidental, under some form of selection, dependent on the protein being studied, or some combination of each.

Previous work on the evolution of A9 multifunctionality suggests that A9 DAMP function is not accidental. We found that a single substitution increased A9 proinflammatory activity without affecting other functions of A9, suggesting that A9 DAMP function has been specifically maintained in mammals.[131] Alternatively, it has been proposed that common biochemical features of proteins, such as hydrophobicity,[199] could themselves serve as DAMPs. Many pattern recognition receptors (PRRs) like TLR4 bind a variety of diverse molecules that share no obvious consensus motif.[207–209] This suggests that DAMPs could function through a non-specific mechanism such as exposure of hydrophobic residues by proteins in the harsh extracellular space. Further studies examining how DAMPs have evolved to activate the innate immune system are required to address these questions.

The orthogonal evolutionary and biochemical approaches described in this work show promise for enabling future studies of DAMP evolution. These strategies can be used to isolate when and how DAMP functions have evolved, dissect mechanisms by which DAMPs activate PRRs, and determine common features of DAMP evolution. Determining how DAMPs have evolved and how they work will enhance our understanding of the innate immune system and enable more targeted intervention in inflammatory disease.

MATERIALS AND METHODS

Phylogenetics and ancestral sequence reconstruction

We reconstructed ancestral sequences using a previously published a phylogenetic tree of S100 proteins containing 172 sequences from 30 amniote taxa.[130] We used PAML4 to generate maximum likelihood ancestors (marginal probability method)[36,38] using the previously-identified maximum likelihood (ML) substitution model (LG+Γ_8)[189] on the ML tree.

Cloning and mutagenesis

All S100 genes in this study were purchased as synthetic constructs in pUC57 vectors from Genscript. S100 genes were sub-cloned into a pETDuet-1 (pD) vector (Millipore) as described previously. Cysteine-free versions of all S100 genes were prepared using site-directed mutagenesis (Agilent).

Protein expression and purification

Recombinant protein overexpression was conducted in *E. coli* BL21 (DE3) pLysS Rosetta cells. Cultures were innoculated in luria broth overnight at 37°C, shaking at 250 rpm, in the presence of ampicillin and chloramphenicol. The following day, 10 ml of saturated culture was diluted into 1.5 L of media with antibiotics, grown to OD_{600} = 0.6 – 1, and then induced overnight at 16°C using 1 mM IPTG. Cells were pelleted at 3,000 rpm for 20 min and stored at -20°C for no more than three months.

Lysates were prepared by vortexing pellets (3-5 g) in tris buffer (25 mM tris, 100 mM NaCl, pH 7.4) and incubating for 20 min at RT with DNAse I and lysozyme (ThermoFisher Scientific). Lysates were sonicated and cell debris was pelleted by centrifugation at 15,000 rpm at 4°C for > 20 min. All proteins were purified on an Äkta PrimePlus FPLC using various 5 ml HiTrap columns (HisTrap FF (Ni-affinity), Q HP (anion exchange)). All S100s were purified in three steps using Ni-affinity chromatography in the presence of calcium followed by two rounds of anion exchange chromatography at different pHs. For Ni-affinity chromatography, proteins were eluted over a 50 ml gradient from 25-1000 mM imidazole in tris buffer. Peak elution fractions were pooled and placed in dialysis overnight at 4°C in 4 L of tris buffer (calcium-free) adjusted to pH 8. Anion exchange chromatography was then performed the following day over a 50 ml gradient from 100-1000 mM NaCl in pH 8 tris buffer. Fractions containing majority S100 were pooled and analyzed for purity on an SDS-PAGE gel. If trace contaminants remained, an additional anion exchange step was performed at pH 6 using the same elution strategy as for the previous anion exchange step. All purified proteins were dialyzed overnight at 4°C in tris buffer + 2 g/L Chelex-100 resin (Biorad), flash-frozen the following day in liquid nitrogen, and stored at -80°C.

Biophysical and biochemical characterization

For all experiments, protein aliquots were thawed fresh from freezer stocks and were either dialyzed in the appropriate experimental buffer overnight at 4°C or exchanged 3X into experimental buffer using 3K microsep spin concentrator

columns (Pall Corporation). All samples were filter-sterilized using 0.1 um spin filters (EMD Millipore) prior to measuring concentration and using in experiments. Thawed aliquots were used for no more than one week before discarding. All concentrations were measured by absorbance of protein in > 6M urea at 280 nm and correspond to micromolar dimeric protein. Circular dichroism (CD) experiments were performed using a Jasco J-815 CD spectrometer.

TLR4 activity assay and cell culture conditions

We purchased commercially distributed HEK293T cells from ATCC (CRL-11268). Because we are using this cell line as a host for heterologous transient transfections, the appropriate control for consistency between assays is the measurement of reporter output for a set of control plasmids and a panel of known treatments. Upon thawing each batch of cells, we run a positive control for ligand-induced response. We transfect the cells with plasmids encoding human CD14, human MD-2, human TLR4, renilla luciferase behind a constitutive promoter, and firefly luciferase behind an NF-KB promoter. We then characterize the raw luciferase output for five treatments: 1) mock, 2) LPS, 3) LPS + polymyxin B, 4) S100A9 + polymyxin B, and 5) S100A9 + 1.25x polymyxin B. This has a stereotypical pattern of responses in renilla luciferase (high for all) and firefly luciferase (low, high, low, high, high). To validate that this response is dependent on the transfected TLR4 complex as opposed to the cells themselves, we repeat the experiment but exclude the TLR4 plasmid. This should give identical renilla luciferase values but no firefly luciferase output in response to any treatment. To

ensure that the cells maintain their properties between passages, we repeat the mock, LPS, and LPS + polymyxin B control on every single experimental plate. This assay has a built-in control for mycoplasma contamination: high firefly luciferase signal in the absence of added agonist. This indicates that there is another source of TLR4-induced NF-kappa B output in the cells—most plausibly, contamination. This mycoplasma sensing approach is used in the commercially available HEK-BLUE mycoplasma detection kit (Invivogen). We discard any cells that exhibit high background values or reach 30 passages.

All plasmids, cell culture conditions, and transfections for measuring the activity of S100s against TLR4s were identical to those previously described.[38,39130,172] Briefly, human embryonic kidney cells (HEK293T/17, ATCC CRL-11268) were maintained up to 30 passages in Dulbecco's Modified Eagle Media (DMEM) supplemented with 10% fetal bovine serum (FBS) at 37°C with 5% CO2. Lipopolysaccharide *E. coli* K-12 LPS (LPS - tlrl-eklps, Invivogen) aliquots were prepared at 5 mg/ml in endotoxin-free water and stored at -20^{o}C. Working solutions were prepared at 10 ug/ml and stored at 4^{o}C to avoid freeze-thaw cycles. S100 proteins were prepared by exchanging into endotoxin-free PBS and incubating with an endotoxin removal column (Thermo Fisher Scientific) for 2 hours. S100 LPS contamination was assessed by measuring activity with and without Polymyxin B, an LPS chelating agent (Figure AA6). LPS (200 ng per 100 ul well) or S100 (0.8, 0.4, 2, 4, or 5 uM dimer) treatments were prepared by diluting in 25:75 endotoxin-free PBS:serum-free Dulbecco's Modified Eagle Media (DMEM – Thermo Fisher Scientific). Polymyxin B (PB, 200 ug per 100 ul well) was added to all S100

experimental samples to limit background endotoxin contamination activity from recombinant protein preps. Cells were incubated with treatments for 3 hours prior to assaying activity. The Dual-Glo Luciferase Assay System (Promega) was used to assay Firefly and Renilla luciferase activity of individual wells. Each NF-κB induction value shown represents the Firefly luciferase activity divided by the Renilla luciferase activity, background-subtracted using the LPS + PB activity for each TLR4 species and normalized to the activity of LPS alone for each TLR4 species to normalize between plates. All measurements were performed using three technical replicates per plate, a minimum of three biological replicates total, and a minimum of two separate protein preps.

FACS and high-throughput MD2 mutant screen

A mutant human MD2 library in a pDuo plasmid backbone was prepared using the Genemorph II EZ clone kit (Agilent). 10 ng of MD2 plasmid were used to obtain approximately 1 mutation per plasmid. Cells were transfected as described above, substituting the firefly luciferase plasmid with a dual firefly luciferase/GFP plasmid under an NF-kB promoter and wildtype MD2 with the mutant MD2 library. Cells were transfected for 24 hours, and transfections were scaled up to 100 mm dishes to increase cell yield for sorting. Transfection mix was removed and cells were treated with PBS buffer, 1.5 ug/ul LPS, or 5 uM purified human S100A9 protein for another 6 hours. Treated cells were sorted on a Sony SH800 cell sorter into GFP- (intensity < 10^4) or GFP+ (intensity > 10^4) bins. Plasmid DNA was extracted from sorted cells and barcoded Illumina libraries were prepared by PCR.

High-throughput sequencing was performed on an Illumina MiSeq instrument set on nano mode.

Species cartoons

All species cartoons were taken from the following websites: http://phylopic.org/image/c089caae-43ef-4e4e-bf26-973dd4cb65c5/, http://phylopic.org/image/aff847b0-ecbd-4d41-98ce-665921a6d96e/, http://phylopic.org/image/0f6af3d8-49d2-4d75-8edf-08598387afde/http://phylopic.org/image/dde4f926-c04c-47ef-a337-927ceb36e7ef/.

We acknowledge Sarah Werning and David Liao as authors of the opossum and mouse cartoons respectively, which were made publicly available through the creative commons attributions 3.0 unported license (https://creativecommons.org/licenses/by/3.0/).

BRIDGE TO CHAPTER V

In chapter IV, we highlight ongoing work examining how A9s have evolved potent proinflammatory activity in mammals and determining the mechanism by which A9 activates TLR4 to drive inflammation. We show preliminary results that suggest eutherian A9s, such as those from humans, mice, and elephants, have evolved increased proinflammatory activity from the therian mammalian ancestor. This increase, however, is not fully accounted for in the reconstructed eutherian A9 ancestor; further work is needed to determine why supraprimate A9s, such as those

from humans and mice, are highly potent and promiscuous activators of amniote TLR4s. We also show the results of a pilot screen that focuses on determining the TLR4/MD2/CD14 interaction interface with A9. We identify candidate mutations in human MD2 that appear to decrease A9 activation of TLR4 without affecting LPS activity and prescribe follow-up studies to determine the role of these residues in A9 activation of TLR4. We outline future studies to determine how A9s have evolved potent proinflammatory activity and the mechanism by which A9 activates TLR4 to drive inflammation. We conclude chapter IV by prescribing future studies to examine how DAMPs have evolved. Chapter V concludes this dissertation by summarizing the work presented in chapters II-IV, discussing the contributions of this work to the fields of evolutionary biochemistry and immunology, and outlining future directions for studying S100A9 and the evolution of DAMPs in innate immunity.

CHAPTER V

SUMMARY AND CONCLUDING REMARKS

Evolutionary biochemistry is powerful for determining how proteins evolve new functions

The work presented here highlights the value of studying biological functions in an evolutionary context.[29] Two of the most commonly asked questions in biology are "how does it work?" and "why does it work this way?" This dissertation demonstrates how one can address both questions simultaneously by focusing mechanistic, functional studies through the lens of evolution. This approach is broadly applicable to proteins across biology, enabling a richer understanding of how biological systems evolve and how they work.

We provide new insight into how multifunctional proteins evolve using the innate immune protein S100A9 (A9) as a case study.[131] We showed that mammalian A9 proteins underwent radical functional change – gain of potent proinflammatory and antimicrobial activities and loss of proteolytic resistance – from a less functional amniote ancestor. By isolating this interval of evolutionary change in A9s between the ancestors of amniotes and mammals, we identified several amino acid changes that impacted each of these functions. A single substitution, in particular, played a key role in A9s gaining proinflammatory activity and losing proteolytic resistance without affecting its antimicrobial activity. Mammals thus evolved a potent proinflammatory molecule that can be quickly turned off by endogenous proteases, all while retaining antimicrobial function to

combat pathogens. This demonstrates how pleiotropy can be advantageous in protein evolution; rather than serving as a constraint,[57,145,146,187] pleiotropy facilitated the evolution of a multifunctional innate immune protein.

This work lays the foundation for future evolutionary studies of A9 proinflammatory activity. We find that A9 proinflammatory potency and promiscuity vary across mammalian A9 proteins, with potentially significant changes occurring later in eutherian A9s. This suggests a complex pattern of co-evolution between A9s and TLR4s in different species; A9s from some species activate many different TLR4s potently, while others exhibit more specific activity against their species-matched TLR4s or are generally poor activators. This could mean that A9s in different species activate TLR4s through overlapping but slightly different mechanisms, or that TLR4 complexes themselves have evolved different properties in different species, perhaps in response to species-specific pathogens or other damage-associated molecular patterns (DAMPs) besides A9. Future studies of how A9s have evolved these complex patterns of activity, as outlined in chapter IV of this dissertation, are necessary to answer these questions.

Mechanistic insight into A9 proinflammatory activity

Using an evolutionary biochemical approach, we identified, to our knowledge, the only known point mutation (M63F) of A9 that decreases its proinflammatory activity. It is doubtful that this mutation would have been identified outside of an evolutionary context, as it is on the inside of the protein and

does not stand out in any obvious way. This finding enabled a mechanistic dissection of A9 activation of TLR4 that would not have otherwise been possible.

Determining how the M63F mutation decreases A9 proinflammatory activity led us to propose a new mechanism by which A9 activates TLR4 to drive inflammation. Through a detailed biophysical analysis, we found that the M63F mutation decreases A9 proinflammatory activity by locally stabilizing the protein through a direct interaction with a nearby residue. The residues that are stabilized by this interaction overlap with residues that have been previously proposed to be important for A9 binding to TLR4.[109] Breaking this interaction restores wildtype A9 stability and proinflammatory activity, indicating that this region of A9 is important for its function and might require a rearrangement or conformational change to activate TLR4. Recent studies in drug design and protein engineering focus on specifically targeting transiently exposed protein surfaces and conformations to inhibit pathological proteins that have been previously deemed "undruggable."[211,212] These findings will direct future studies toward determining the nature of the hypothesized proinflammatory non-native conformation of A9, thus informing the design of treatments to specifically target A9-mediated chronic inflammation.

Future mechanistic and evolutionary studies to understand DAMP functions

We prescribe the use of complementary approaches to evolutionary biochemistry for determining mechanisms by which proteins evolve and perform functions. We highlight the challenges associated with determining where A9

interacts with TLR4, MD2, and CD14 and mechanistically understanding how A9 proinflammatory function works. The findings presented here provide new insight into how A9 activates TLR4 and will help focus future studies toward important aspects of A9 proinflammatory function. Orthogonal approaches, however – such as the high-throughput mutant screening strategy of TLR4 components described in chapter IV – will be necessary to fully determine where and how A9 interacts with TLR4 to drive inflammation. Creatively combining evolutionary and biochemical approaches, particularly in a high-throughput fashion, will be a cornerstone of future mechanistic and evolutionary studies of protein functions.

This work raises important questions regarding the mechanisms by which DAMPs have evolved to activate the immune system. We found that the DAMP S100A9 lost stability en route to evolving proinflammatory activity. Intriguingly, the easily degraded A9 forms a proteolytically resistant A8/A9 complex that has different functions, thus allowing A9 to sample two different states of function and stability. The relative abundance of A9 in each of these states is likely mediated by multiple factors, including relative levels of A8 and A9 expression under different conditions and whether the protein is within neutrophils or released into the extracellular space. Many DAMPs perform other functions inside cells until they are released due to tissue damage.[184–186] Perhaps other DAMPs, similar to A9, also have binding partners that protect them from degradation until DAMP function is required in the harsh environment of the extracellular space.

Loss of stability could also be a common feature of DAMPs. As demonstrated for A9, loss of stability could serve to simultaneously enable DAMP

function and create a simple mechanism for regulating it. Studies examining evolutionary trends in protein stability indicate that many proteins are marginally stable because selection only requires a protein to be "stable enough," suggesting that protein stability is not necessarily under selection.[213–215] Different rules may apply, however, to DAMPs, as lower stability may be important for both DAMP function and for maintaining a mechanism to rapidly remove pathological inflammation from the body. Future studies examining the role of stability in DAMP evolution and function will aid in addressing these questions.

This dissertation contributes to the field of evolutionary biochemistry by providing mechanistic insight into how the innate immune protein S100A9 evolved multiple functions. We highlight evolutionary and biochemical approaches for studying mechanisms of protein evolution that can be broadly applied to many biological systems. This work lays a foundation for determining how A9 and other DAMPs have evolved critical innate immune functions and will help inform strategies to alleviate pathological DAMP function.

APPENDIX A

SUPPLEMENTAL MATERIAL FOR CHAPTER II

Supplemental Figures

This section includes supplemental figures referenced in Chapter II.

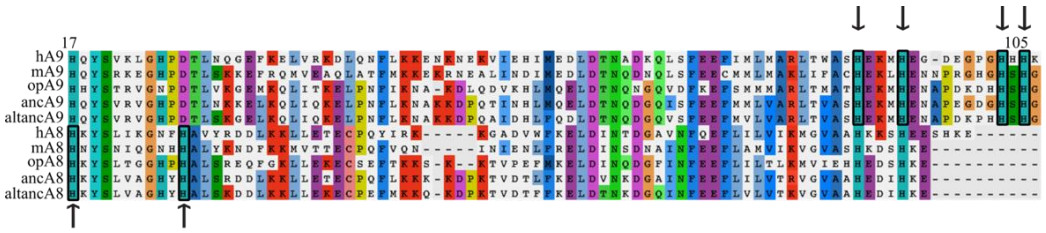
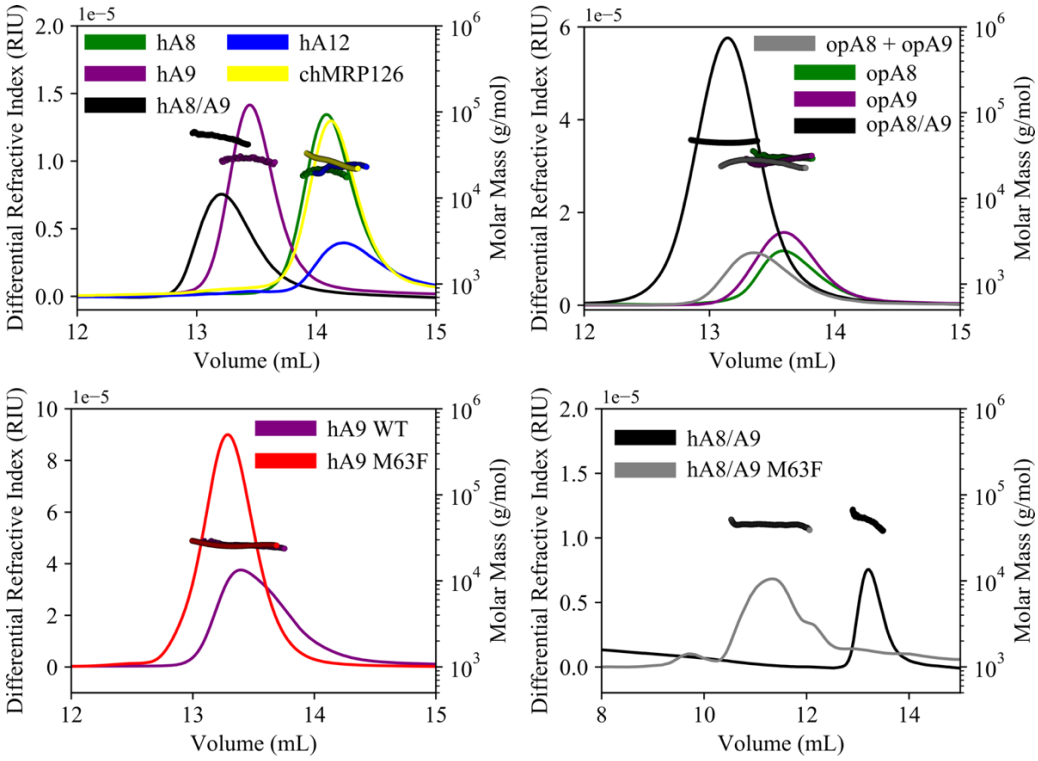


Figure AA1. Hexahistidine site conservation of modern and ancestrally reconstructed A8s and A9s. Human (h), mouse (m), opossum (op), maximum likelihood therian mammalian ancestors (anc), and AltAll ancestors (altanc) shown. Alignment truncated to show conservation of key hexahistidine site metal binding residues (boxed + arrows). A8s conserve two (positions 17 and 27 in human A8), while A9s conserve four (positions 91, 95, 103, and 105 of human A9). Consensus residues for alignment are highlighted.



Protein	Mw (kDa)	Error	Polydispersity (Mw/Mn)	Error	Mass fraction (%)
hA9	26.5	2.00%	1.004	2.74%	96.5
hA9 M63F	27.1	8.20%	1.058	9.24%	90
hA8/A9	47.8	4.80%	1.046	7.28%	94.6
hA8/A9 M63F	48.7	8.70%	1.049	9.77%	88.7
hA8	18.5	7.40%	1.073	10.60%	89.6
hA9	27.4	6.10%	1.061	7.42%	91.6
chMRP126	28.1	15.60%	1.056	18.97%	84.9
hA12	23.5	14.20%	1.031	19.19%	80.2
opA8	31.7	9.3%	1.017	12.25%	87.5
opA9	29.7	7.6%	1.019	10.41%	91.9
opA8/A9	46.8	1.4%	1.003	1.93%	92.6
opA8 + opA9	25.7	9.6%	1.012	14.01%	93.2

Figure AA2. Oligomeric state analysis of S100 proteins by SECMALS. Differential refractive index (left y-axis, lines) and calculated molecular weights from light scattering detectors (right y-axis, points) for modern S100 proteins used in this study. h = human, op = opossum, ch = chicken species. Opossum A8 + A9 sample is an equimolar mixture of opossum A8 and A9 homodimers. Table below shows summary data calculated using Wyatt Astra software.

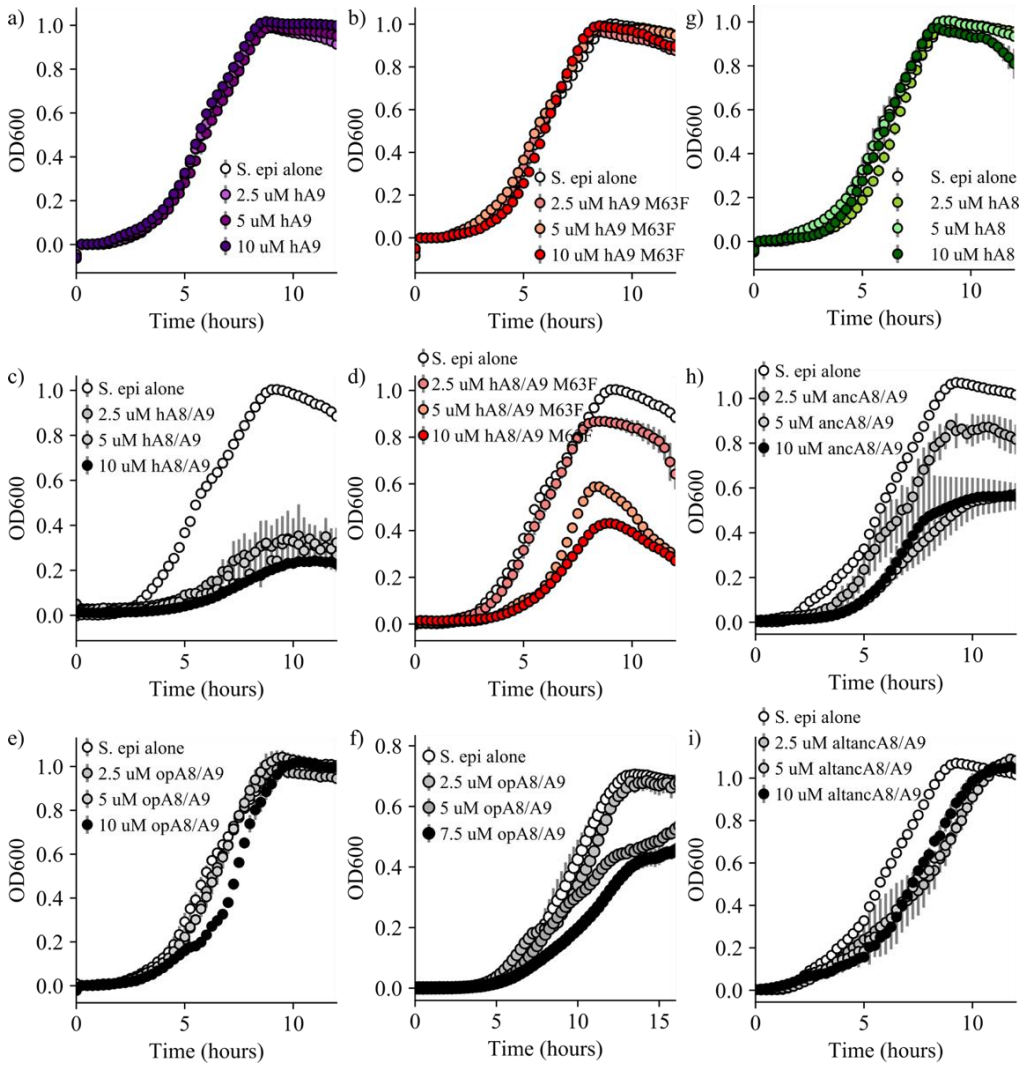


Figure AA3. *S. epidermidis* growth curves in the presence of S100 proteins. Representative *S. epidermidis* growth curves in the presence of a) human A9, b) human A9 M63F, c) human A8/A9, d) human A8/A9 M63F, e) opossum A8/A9 (cysteine-free), f) opossum A8/A9 (containing cysteines), g) ancA8/A9, and h) altancA8/A9. Error bars are the standard deviation for 3 technical replicates, points show one representative biological replicate for each protein at four different concentrations.

Table AA4. Average posterior probabilities for ancestrally reconstructed proteins.

Ancestrally reconstructed protein	Average posterior probability	# of sites that differ from the ML sequence	Average # sites that differ from ML sequence for reconstructions on alternate topologies
ancA9	0.83	--	5
altancA9	0.81	10	--
ancA8	0.88	--	1
altancA8	0.86	17	--
ancCG	0.86	--	8
altancCG	0.84	8	--

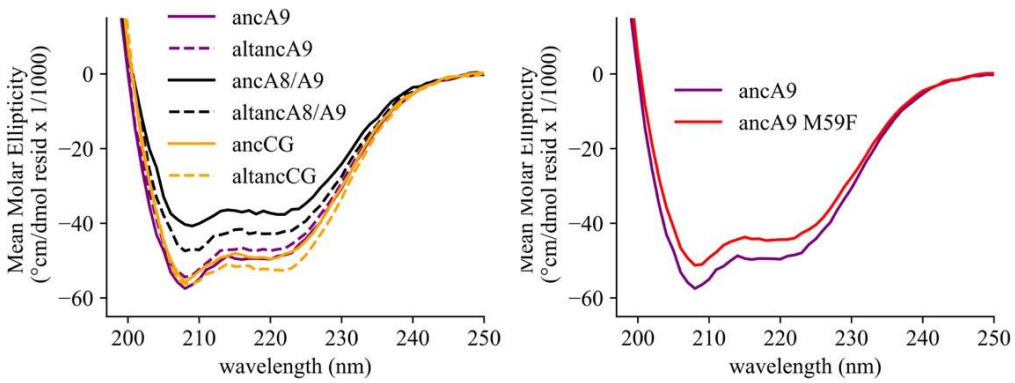


Figure AA5. Circular dichroism spectroscopy measurements of ancestral proteins. Data shown are the average of 3 scans. Solid lines are maximum likelihood ancestral proteins, dotted lines are alt-all ancestors (colored the same as matching maximum likelihood ancestor for comparison).

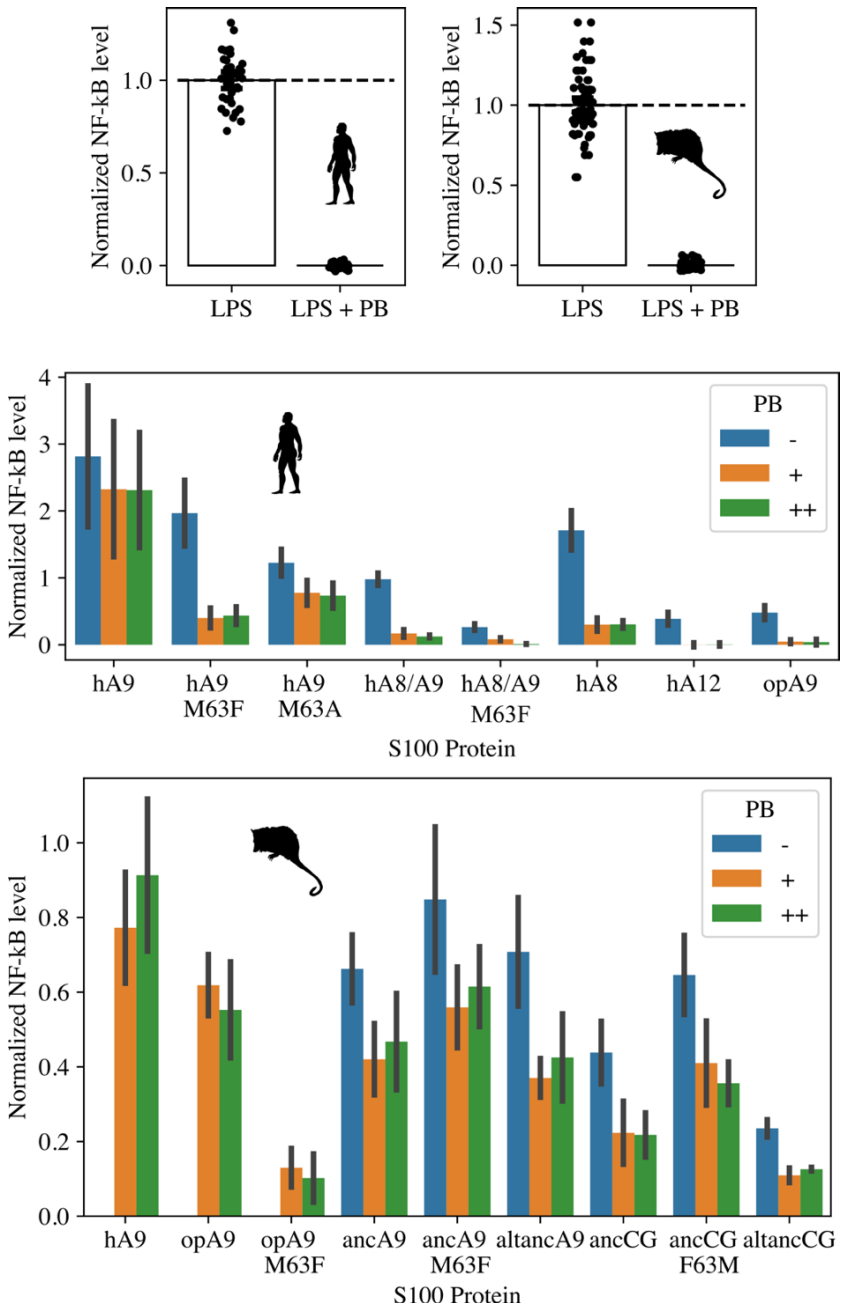


Figure AA6. Validation of polymyxin B treatment for endotoxin contamination. Top panel: Activity of LPS (0.2 ng/ul) against human (left) and opossum TLR4 (right) is inhibited by the addition of Polymyxin B (PB, 0.2 ug/ul). Middle and bottom panels: S100 activation of human (middle) and opossum (bottom) TLR4 with no PB, 0.2 ug/ul PB (+) and 0.25 ug/ul PB (++). No-PB data were not collected for hA9, opA9, and opA9 M60F against opossum TLR4 (bottom panel). Data were background-subtracted using the LPS + PB control and normalized to LPS activity against either human or opossum TLR4. Error bars are standard deviation.

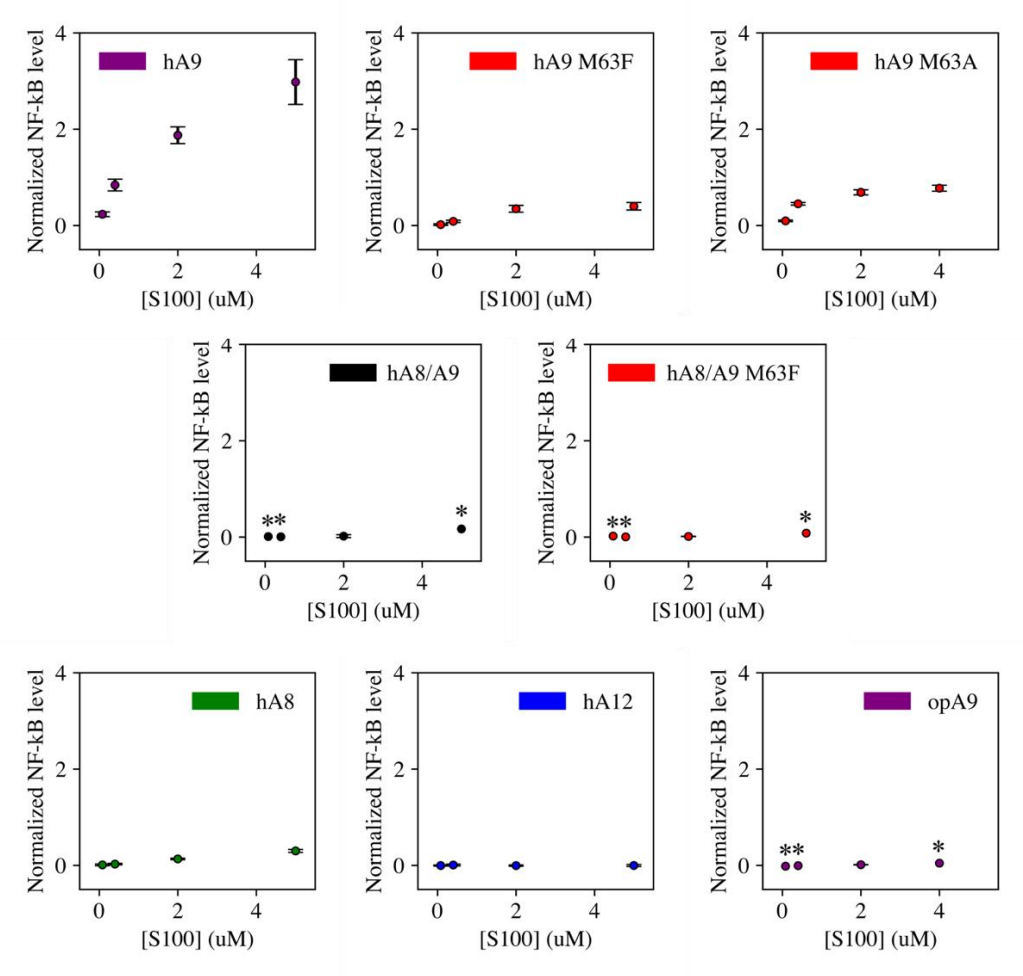


Figure AA7. Human TLR4 dose curves. Points are the average of >3 biological replicates each consisting of 3 technical triplicates, error bars are standard error of the mean. An asterisk (*) indicates a concentration at which a single biological replicate was measured. Data were background-subtracted using the LPS + PB control and normalized to LPS activity against human TLR4.

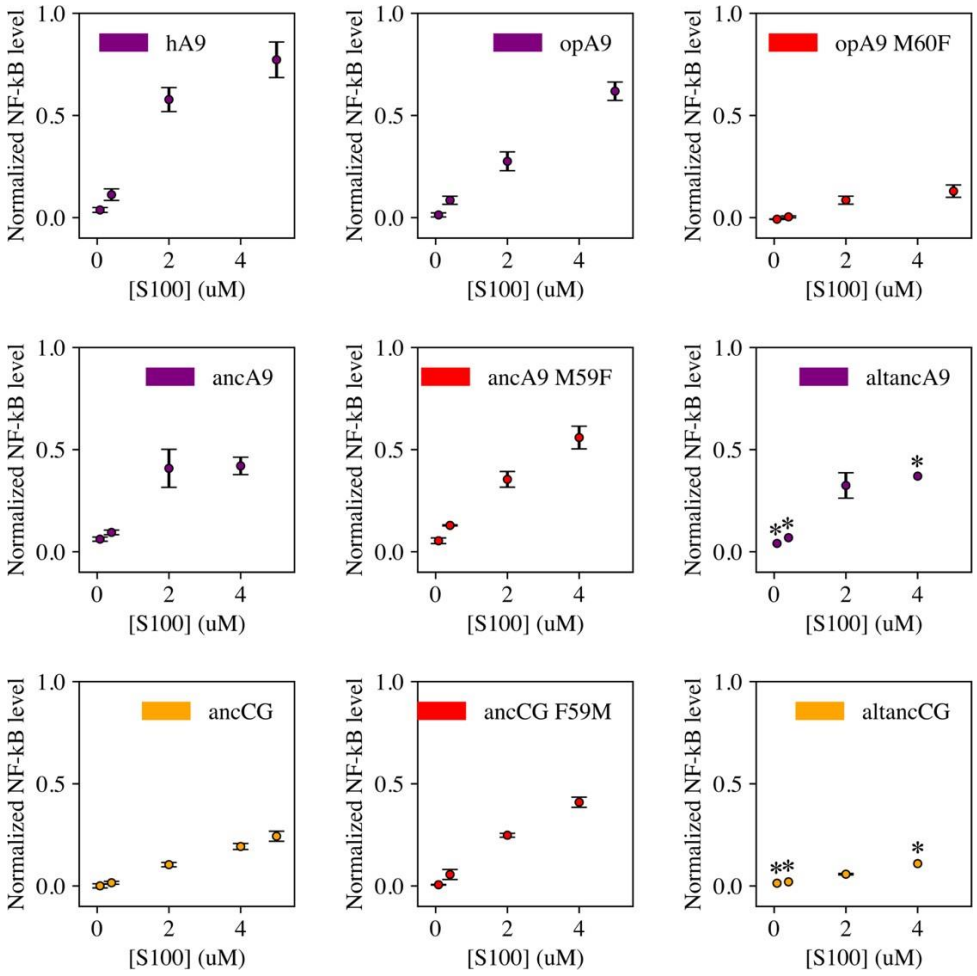


Figure AA8. Opossum TLR4 dose curves. Points are the average of >3 biological replicates each consisting of 3 technical triplicates, error bars are standard error of the mean. An asterisk (*) indicates a concentration at which a single biological replicate was measured. Data were background-subtracted using the LPS + PB control and normalized to LPS activity against opossum TLR4.

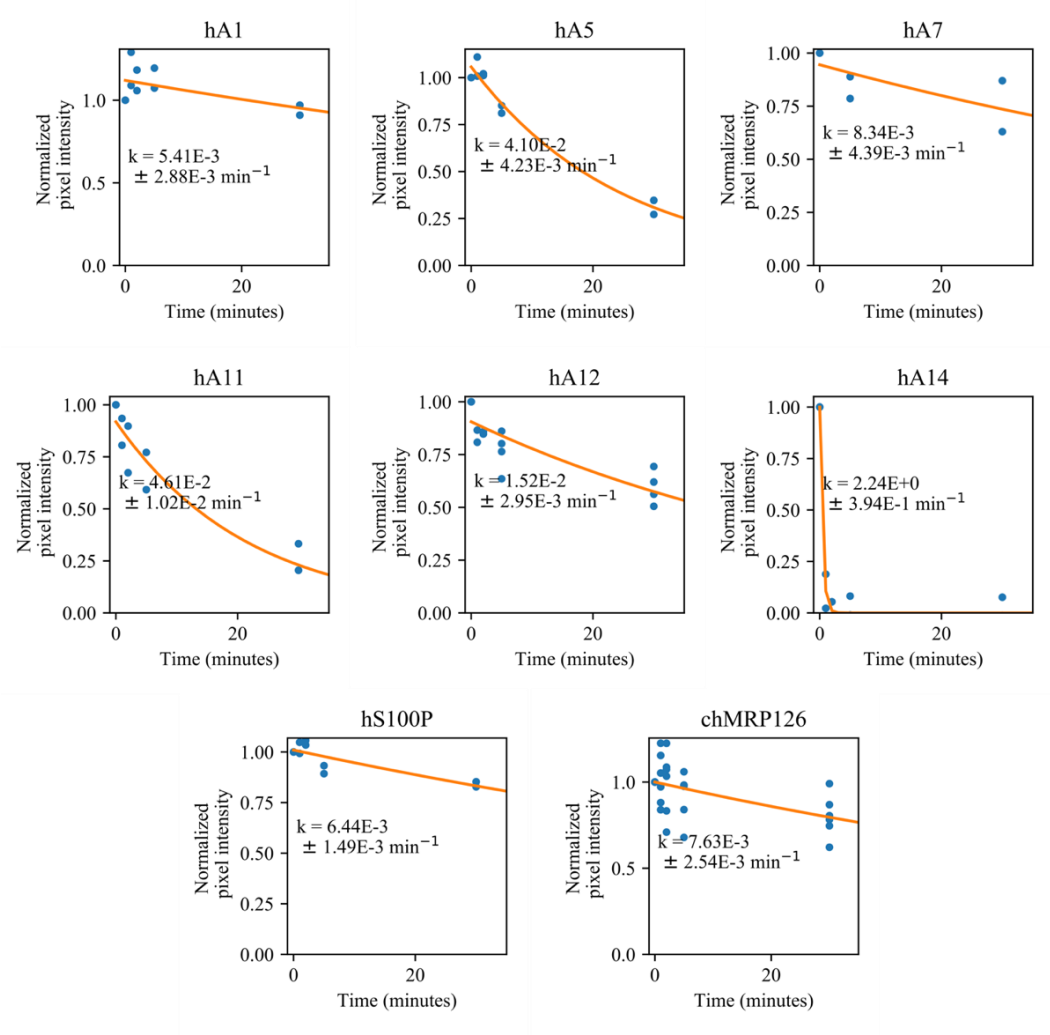


Figure AA9. Extant S100 protein proteolysis fits. Blue dots are biological replicates, orange line is a single exponential decay fit (see methods). Protein is listed at the top. Pixel intensity was quantified by densitometry from SDS-PAGE gels.

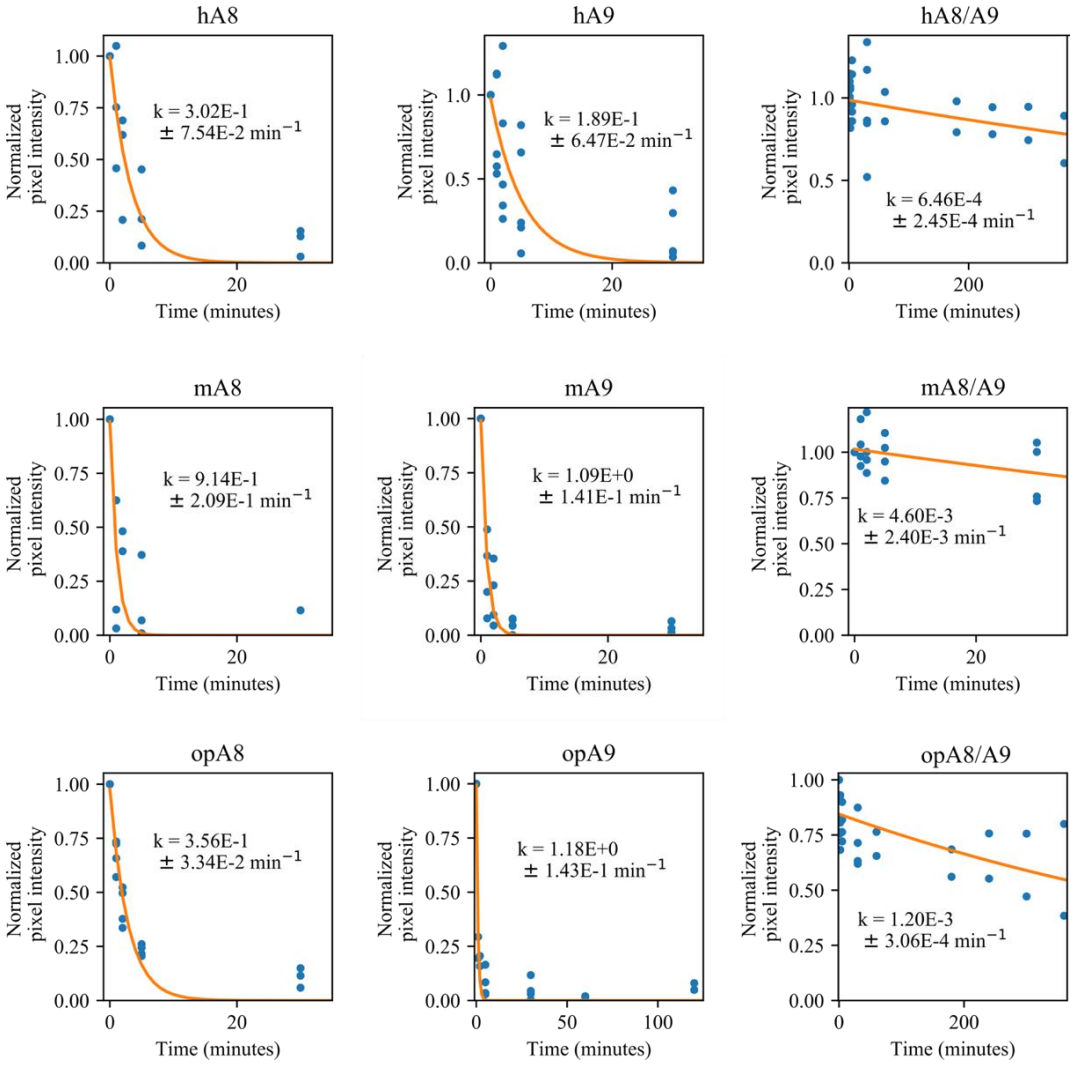


Figure AA10. Mammalian A8, A9, and A8/A9 complex proteolysis fits. Blue dots are biological replicates, orange line is a single exponential decay fit (see methods). Protein is listed at the top. Pixel intensity was quantified by densitometry from SDS-PAGE gels. Longer time points were collected for proteins with slower degradation rates (see x-axis).

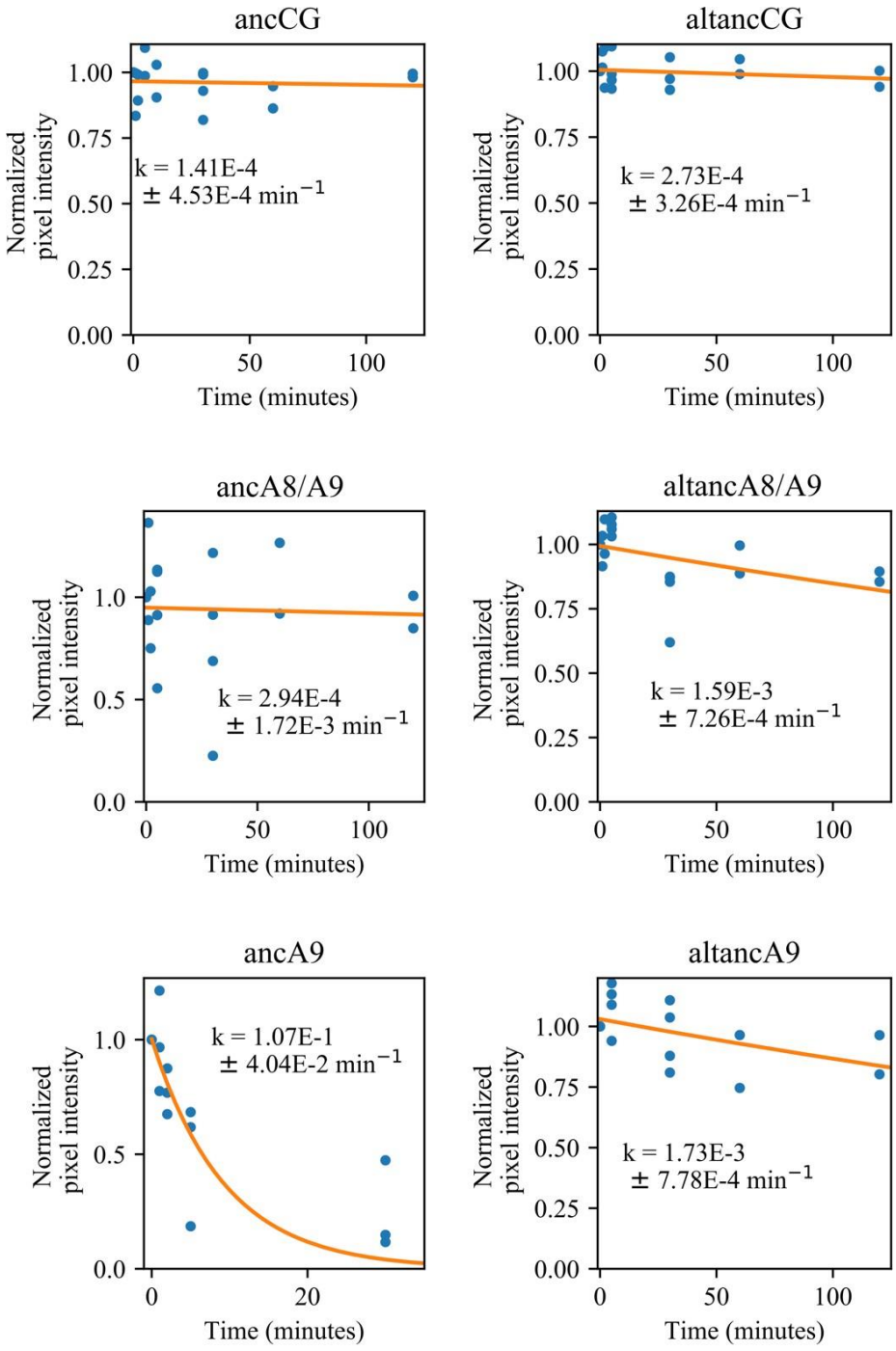


Figure AA11. Ancestral S100 proteolysis fits. Blue dots are biological replicates, orange line is a single exponential decay fit (see methods). Protein is listed at the top. Pixel intensity was quantified by densitometry from SDS-PAGE gels. Longer time points were collected for proteins with slower degradation rates (see x-axis).

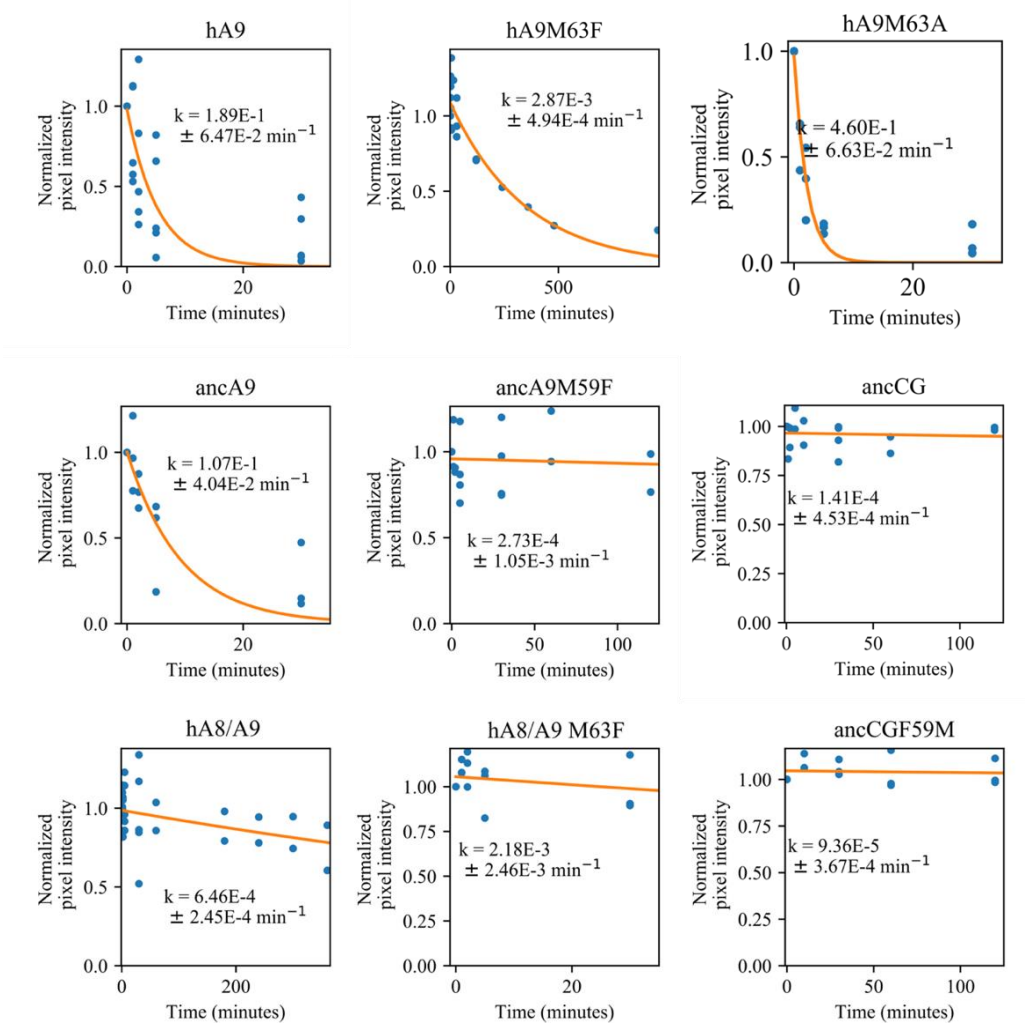


Figure AA12. Mutant S100 proteolysis fits. Blue dots are biological replicates, orange line is a single exponential decay fit (see methods). Protein is listed at the top. Pixel intensity was quantified by densitometry from SDS-PAGE gels. Longer time points were collected for proteins with slower degradation rates (see x-axis).

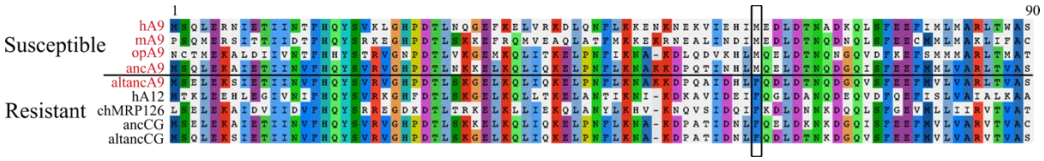


Figure AA13. Sequence alignment highlighting identification of position 63. S100 protein sequences are grouped into proteolytically susceptible (top) or resistant and potently proinflammatory (red text) or not (black text). Only the first 90 residues out of 114 total were examined as the disordered A9 tail (residues ~93-114) are highly variable and the tail is dispensable for A9 proinflammatory activity. Residues are colored when found to be the consensus residue for a column.

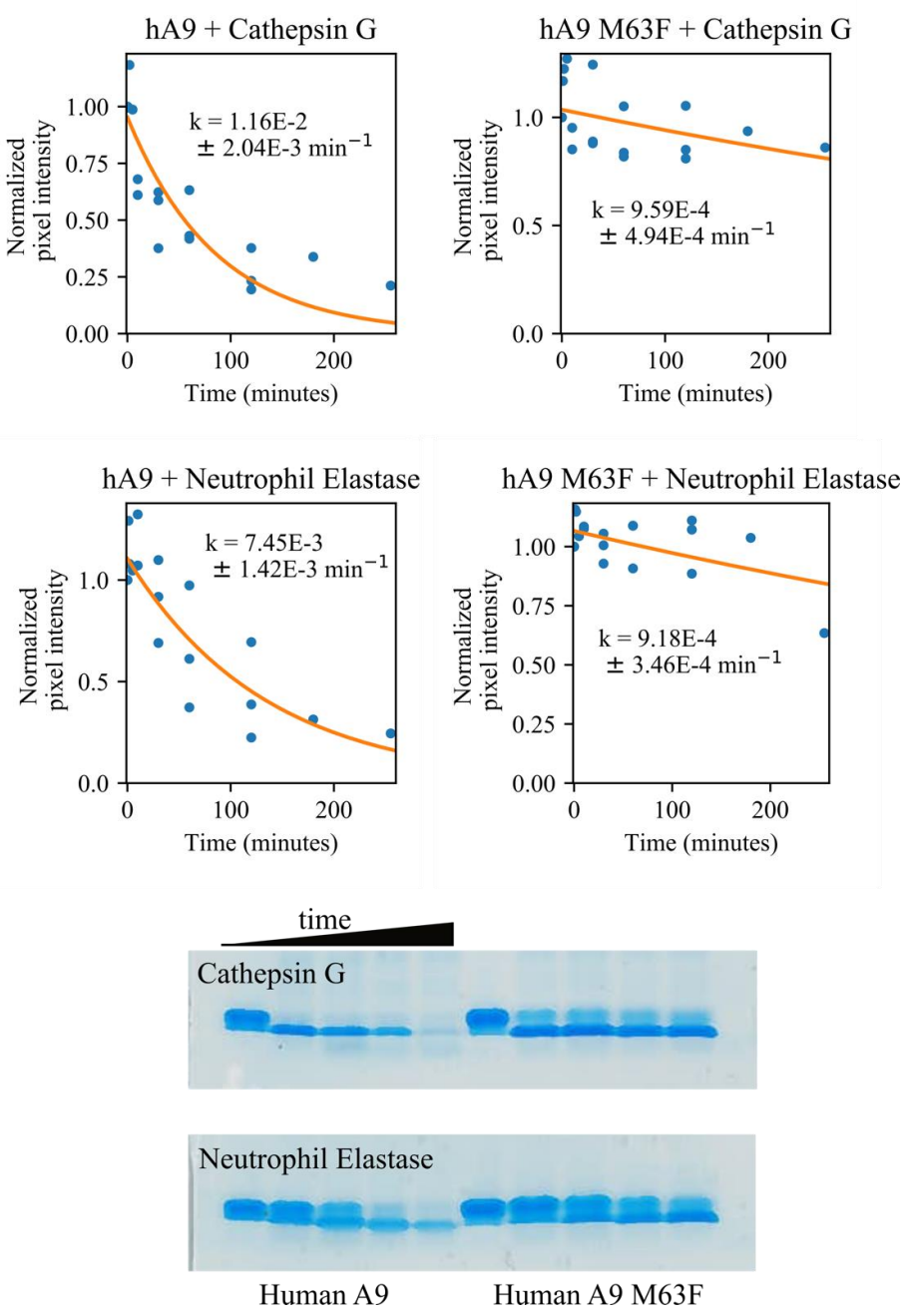
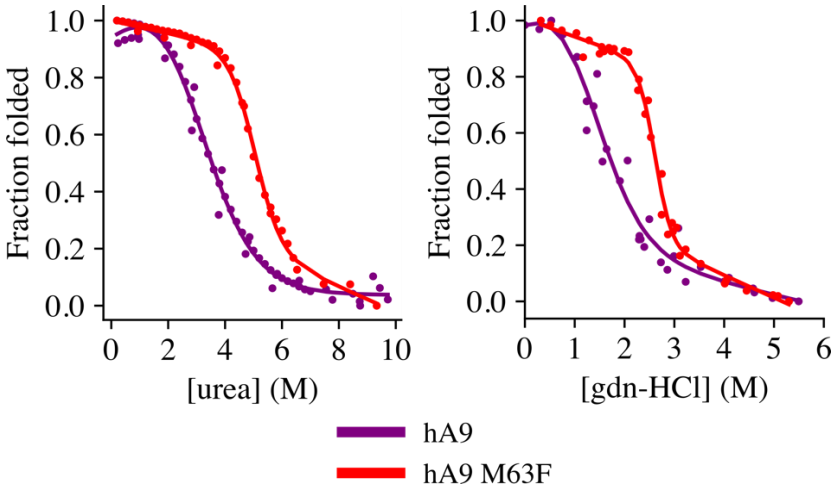


Figure AA14. Human A9 and A9 M63F proteolytic degradation using neutrophil proteases. Blue dots are biological replicates, orange line is a single exponential decay fit (see methods). Protein is listed at the top of each graph. Pixel intensity was quantified by densitometry from SDS-PAGE gels. Longer time points were collected for proteins with slower degradation rates (see x-axis). Representative SDS-PAGE gels are shown for 0-120 minute digestion with each protease.



Protein	Denaturant	ΔG (kcal/mol)	m-value (kcal/mol x M^{-1})	C_m (M)
Human A9	Urea	1.91 ± 0.14	0.65 ± 0.01	2.96 ± 0.22
Human A9 M63F	Urea	6.57 ± 0.18	1.31 ± 0.01	5.03 ± 0.14
Human A9	Gdn-HCl	1.62 ± 1.30	1.25 ± 0.22	1.3 ± 1.07
Human A9 M63F	Gdn-HCl	9.36 ± 1.70	3.59 ± 0.26	2.61 ± 0.51

Figure AA15. A9 equilibrium chemical denaturation experiments. Chemical denaturation experiments using urea (left) and guanidinium hydrochloride (right). Graphs represent > 3 replicates. Human A9 is shown in purple, A9 M63F in red. An apparent two-state unfolding model was fit to the data to estimate thermodynamic parameters, shown as a solid line.

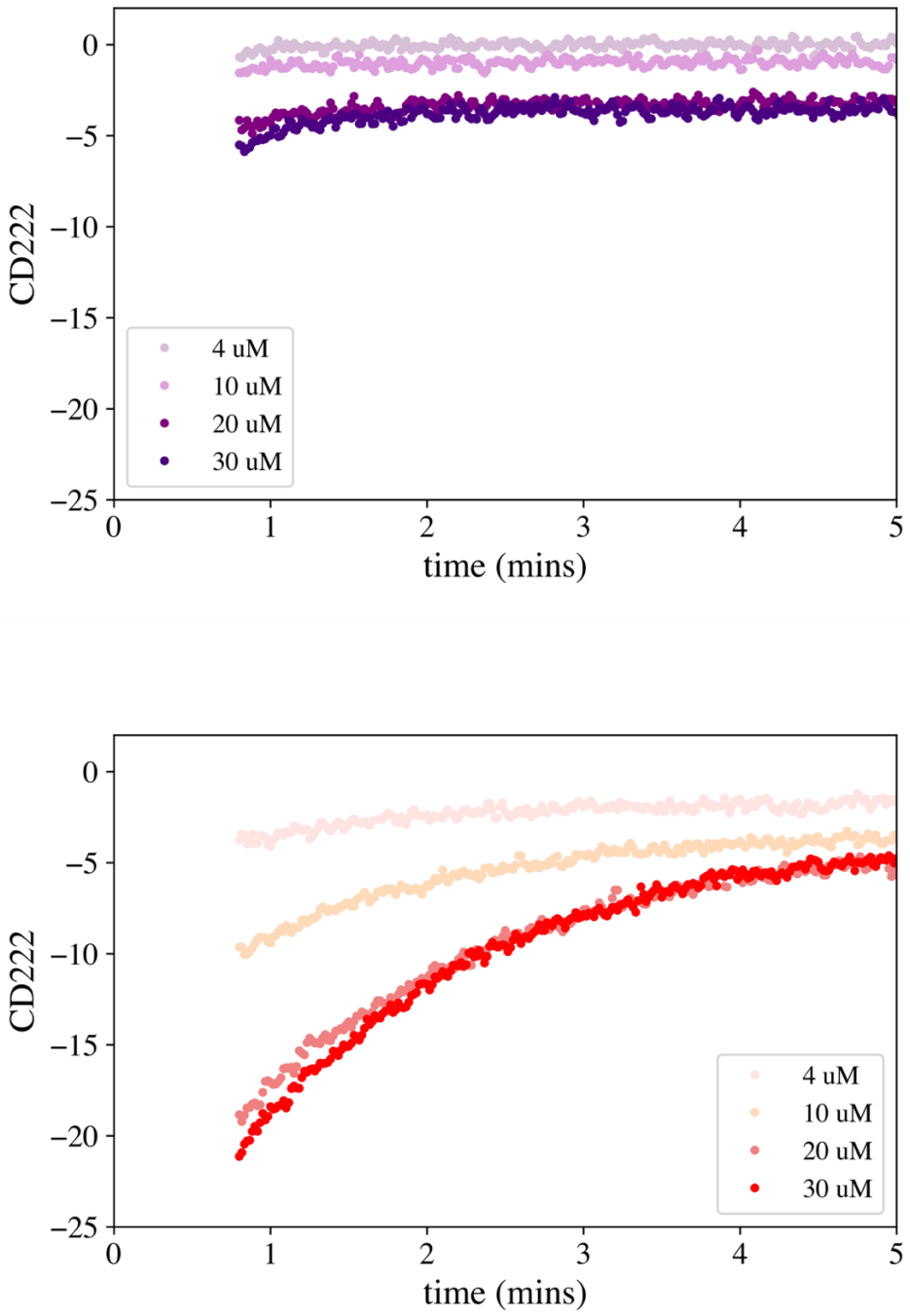


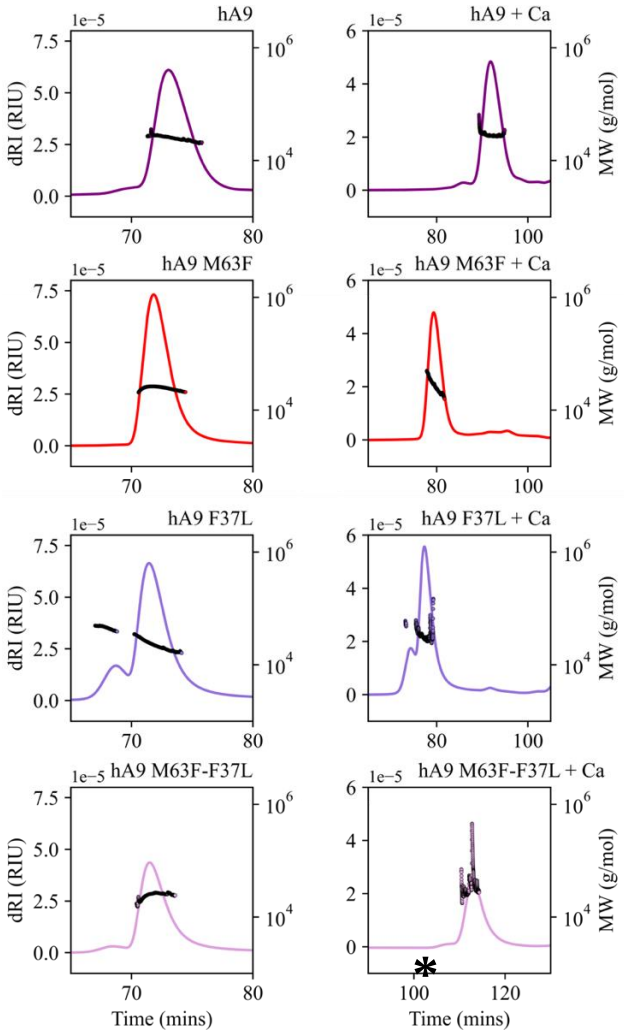
Figure AA16. A9 kinetic chemical denaturation experiments. Time course measurement of hA9 (top) and hA9 M63F (bottom) unfolding upon addition of 6M gdn-HCl. Each curve is a single replicate at one concentration, monitoring CD signal at 222 nm (y-axis).

APPENDIX B

SUPPLEMENTAL MATERIAL FOR CHAPTER III

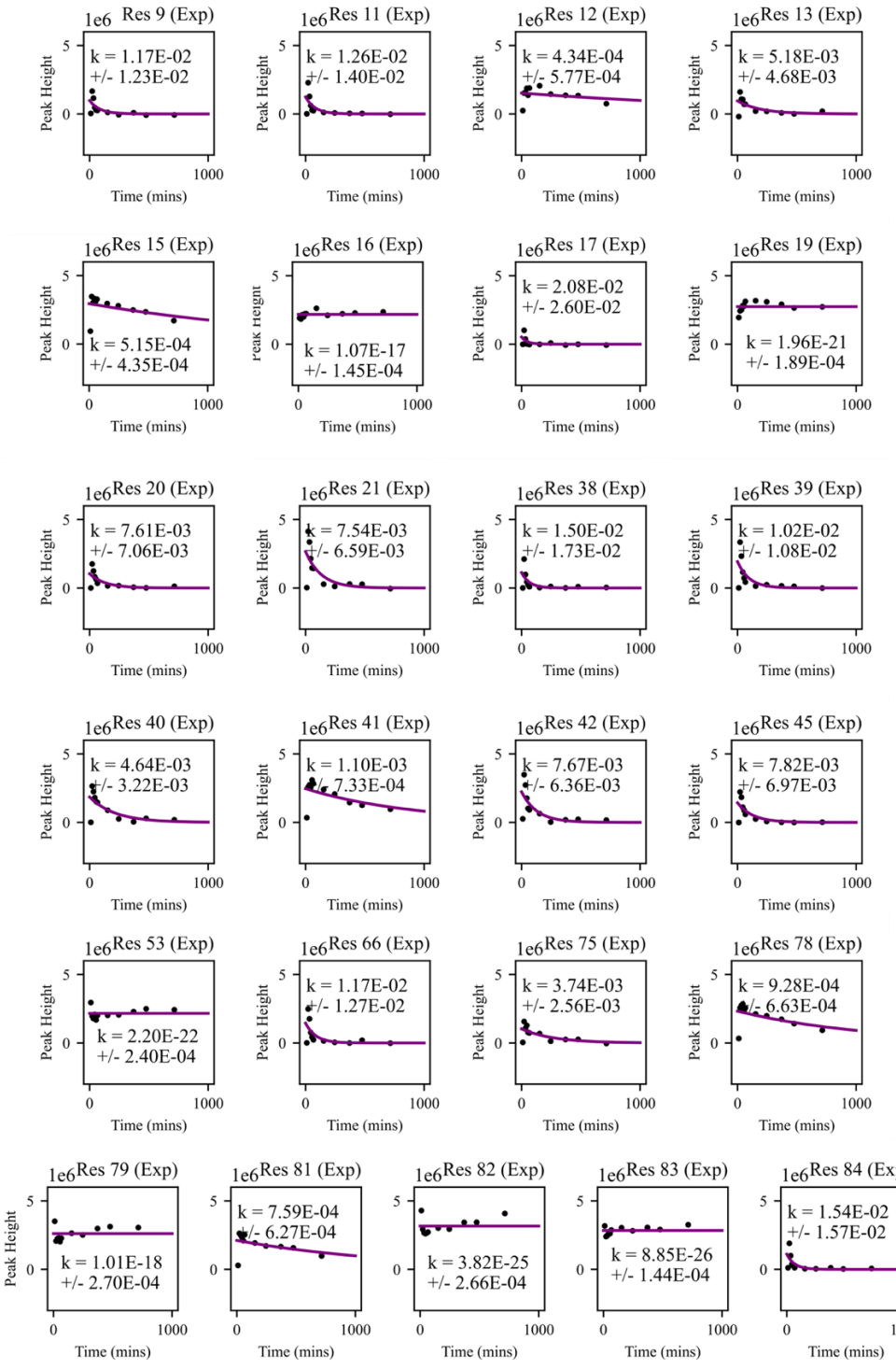
Supplemental Figures

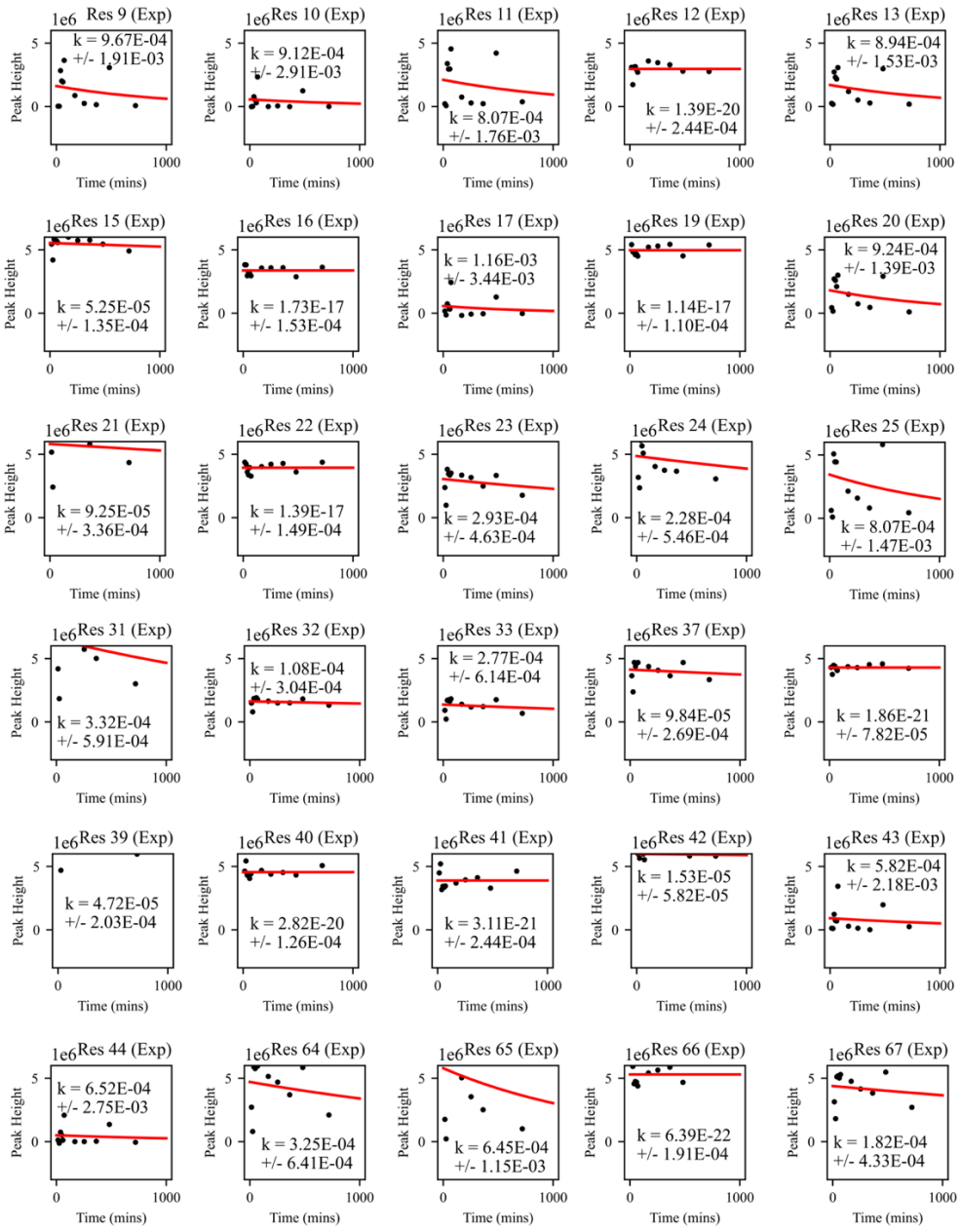
This section includes supplemental figures referenced in Chapter III.



Protein	Mw (kDa)		Polydispersity (Mw/Mn)		Mass Fraction (%)	
	- Ca	*+ Ca*	*- Ca*	*+ Ca*	*- Ca*	*+ Ca*
hA9	25.1 ± 0.9	31.0 ± 2.6	1.01 ± 0.05	1.08 ± 0.10	96.1	96.6
hA9 M63F	25.0 ± 0.2	29.8 ± 3.4	1.00 ± 0.01	1.12 ± 0.16	100	97.2
hA9 F37L	23.5 ± 0.4	33.9 ± 4.9	1.04 ± 0.02	1.07 ± 0.19	85.1	82.9
hA9 M63F-F37L	25.3 ± 0.9	*46.6 ± 28.2	1.02 ± 0.05	*1.74 ± 0.86	94.2	*100

Figure AB1. Oligomeric state analysis of A9 and A9 variants by SECMALS. Differential refractive index (left y-axis, lines) and calculated molecular weights from light scattering detectors (right y-axis, points) for A9 proteins used in this study. * indicates preliminary low-quality measurement that could not be completed due to the coronavirus pandemic. Ca = calcium, Mw = molecular weight. Table shows summary data calculated using Wyatt Astra software.


1e6 Res 9 (Exp)
k = 1.17E-02
+/- 1.23E-02
1e6 Res 11 (Exp)
k = 1.26E-02
+/- 1.40E-02
1e6 Res 12 (Exp)
k = 4.34E-04
+/- 5.77E-04
1e6 Res 13 (Exp)
k = 5.18E-03
+/- 4.68E-03
1e6 Res 15 (Exp)
k = 5.15E-04
+/- 4.35E-04
1e6 Res 16 (Exp)
k = 1.07E-17
+/- 1.45E-04
1e6 Res 17 (Exp)
k = 2.08E-02
+/- 2.60E-02
1e6 Res 19 (Exp)
k = 1.96E-21
+/- 1.89E-04
1e6 Res 20 (Exp)
k = 7.61E-03
+/- 7.06E-03
1e6 Res 21 (Exp)
k = 7.54E-03
+/- 6.59E-03
1e6 Res 38 (Exp)
k = 1.50E-02
+/- 1.73E-02
1e6 Res 39 (Exp)
k = 1.02E-02
+/- 1.08E-02
1e6 Res 40 (Exp)
k = 4.64E-03
+/- 3.22E-03
1e6 Res 41 (Exp)
k = 1.10E-03
+/- 7.33E-04
1e6 Res 42 (Exp)
k = 7.67E-03
+/- 6.36E-03
1e6 Res 45 (Exp)
k = 7.82E-03
+/- 6.97E-03
1e6 Res 53 (Exp)
k = 2.20E-22
+/- 2.40E-04
1e6 Res 66 (Exp)
k = 1.17E-02
+/- 1.27E-02
1e6 Res 75 (Exp)
k = 3.74E-03
+/- 2.56E-03
1e6 Res 78 (Exp)
k = 9.28E-04
+/- 6.63E-04
1e6 Res 79 (Exp)
k = 1.01E-18
+/- 2.70E-04
1e6 Res 81 (Exp)
k = 7.59E-04
+/- 6.27E-04
1e6 Res 82 (Exp)
k = 3.82E-25
+/- 2.66E-04
1e6 Res 83 (Exp)
k = 8.85E-26
+/- 1.44E-04
1e6 Res 84 (Exp)
k = 1.54E-02
+/- 1.57E-02
Peak Height
Time (mins)
0
1000
5


1e6 Res 9 (Exp)
k = 9.67E-04
+/- 1.91E-03
1e6 Res 10 (Exp)
k = 9.12E-04
+/- 2.91E-03
1e6 Res 11 (Exp)
k = 8.07E-04
+/- 1.76E-03
1e6 Res 12 (Exp)
k = 1.39E-20
+/- 2.44E-04
1e6 Res 13 (Exp)
k = 8.94E-04
+/- 1.53E-03
1e6 Res 15 (Exp)
k = 5.25E-05
+/- 1.35E-04
1e6 Res 16 (Exp)
k = 1.73E-17
+/- 1.53E-04
1e6 Res 17 (Exp)
k = 1.16E-03
+/- 3.44E-03
1e6 Res 19 (Exp)
k = 1.14E-17
+/- 1.10E-04
1e6 Res 20 (Exp)
k = 9.24E-04
+/- 1.39E-03
1e6 Res 21 (Exp)
k = 9.25E-05
+/- 3.36E-04
1e6 Res 22 (Exp)
k = 1.39E-17
+/- 1.49E-04
1e6 Res 23 (Exp)
k = 2.93E-04
+/- 4.63E-04
1e6 Res 24 (Exp)
k = 2.28E-04
+/- 5.46E-04
1e6 Res 25 (Exp)
k = 8.07E-04
+/- 1.47E-03
1e6 Res 31 (Exp)
k = 3.32E-04
+/- 5.91E-04
1e6 Res 32 (Exp)
k = 1.08E-04
+/- 3.04E-04
1e6 Res 33 (Exp)
k = 2.77E-04
+/- 6.14E-04
1e6 Res 37 (Exp)
k = 9.84E-05
+/- 2.69E-04
k = 1.86E-21
+/- 7.82E-05
1e6 Res 39 (Exp)
k = 4.72E-05
+/- 2.03E-04
1e6 Res 40 (Exp)
k = 2.82E-20
+/- 1.26E-04
1e6 Res 41 (Exp)
k = 3.11E-21
+/- 2.44E-04
1e6 Res 42 (Exp)
k = 1.53E-05
+/- 5.82E-05
1e6 Res 43 (Exp)
k = 5.82E-04
+/- 2.18E-03
1e6 Res 44 (Exp)
k = 6.52E-04
+/- 2.75E-03
1e6 Res 64 (Exp)
k = 3.25E-04
+/- 6.41E-04
1e6 Res 65 (Exp)
k = 6.45E-04
+/- 1.15E-03
1e6 Res 66 (Exp)
k = 6.39E-22
+/- 1.91E-04
1e6 Res 67 (Exp)
k = 1.82E-04
+/- 4.33E-04
Peak Height
Time (mins)
0
5
1000

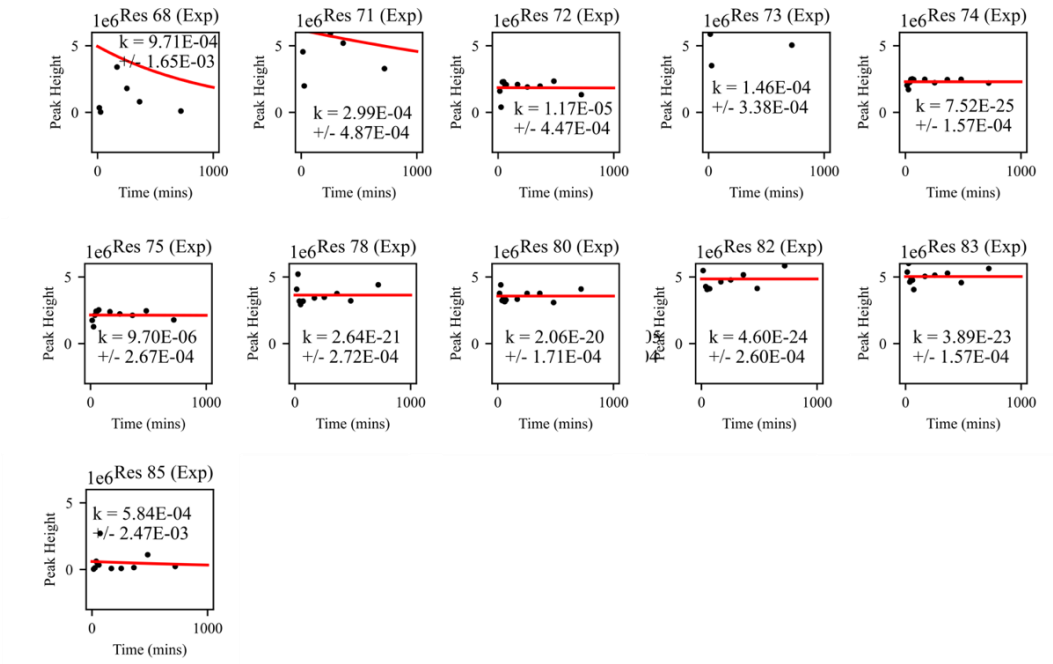


Figure AB2. Exponential fits for A9 and A9 M63F NMR HDX experiments. Points are HSQC peak intensity at a given timepoint, lines are single exponential fits to data (purple = A9, red = A9 M63F). Fitting details are described in methods.

Table AB3. Exponential fit data table for A9 and A9 M63F NMR HDX experiments. Data corresponding to Chapter III, Figure 3.4b (residues that exchanged too quickly to measure for both A9 and A9 M63F – and therefore do not differ in rate – are not included).

residue	a9_ln_rate	a9_ln_err	m63f_ln_rate	m63f_ln_err	diff	diff_err
9	-4.4441	1.0429	-6.941	1.974	2.497	0.894
10	0.6931	0.0000	-7.000	3.190	7.693	0.415
11	-4.3779	1.1189	-7.122	2.180	2.744	0.893
13	-5.2635	0.9038	-7.020	1.712	1.757	1.102
15	-7.5706	0.8437	-9.855	2.579	2.285	1.188
17	-3.8715	1.2480	-6.755	2.955	2.884	1.112
20	-4.8776	0.9266	-6.987	1.509	2.109	0.840
21	-4.8879	0.8743	-9.288	3.629	4.400	0.848
23	0.6931	0.0000	-8.136	1.581	8.830	0.179
24	0.6931	0.0000	-8.386	2.392	9.079	0.263
31	0.6931	0.0000	-8.009	1.778	8.702	0.204
32	0.6931	0.0000	-9.135	2.821	9.828	0.287
33	0.6931	0.0000	-8.191	2.214	8.884	0.249
37	0.6931	0.0000	-9.226	2.730	9.919	0.275
39	-4.5836	1.0537	-9.962	4.296	5.378	0.822
42	-4.8710	0.8302	-11.089	3.810	6.218	0.627
43	0.6931	0.0000	-7.448	3.751	8.142	0.461
44	0.6931	0.0000	-7.335	4.216	8.028	0.525
45	-4.8508	0.8912	0.693	0.000	-5.544	-0.161
64	0.6931	0.0000	-8.032	1.973	8.725	0.226
65	0.6931	0.0000	-7.346	1.779	8.039	0.221
67	0.6931	0.0000	-8.614	2.382	9.307	0.256
68	0.6931	0.0000	-6.937	1.698	7.630	0.223
71	0.6931	0.0000	-8.116	1.631	8.809	0.185
73	0.6931	0.0000	-8.833	2.316	9.526	0.243
81	-7.1834	0.8264	-9.329	4.713	2.145	2.230
84	-4.1741	1.0207	0.693	0.000	-4.867	-0.210
85	0.6931	0.0000	-7.446	4.239	8.140	0.521

Table AB4. A9 mutant cycle chemical denaturation data. Parameters for a two-state unfolding model fit to data in Figure 3.4a-c (see methods). F-F interaction energies at different calcium concentrations shown below, calculated from dG values listed in upper table (see Figure 3.4d). Err = error from fit (diagonalized covariance matrix). Errors in bottom F-F interaction table are propagated from errors in upper table.

Protein	[Ca] (mM)	dG	err	m	err	Cm	err
hA9	0	2.9	0.6	1.0	0.0	2.8	0.6
hA9	500	1.9	0.1	0.6	0.0	3.0	0.2
hA9	5000	4.8	1.3	1.1	0.1	4.4	1.2
f37l	0	2.8	0.1	0.9	0.0	3.1	0.1
f37l	500	1.9	0.8	0.7	0.0	2.9	1.3
f37l	5000	2.2	0.3	0.7	0.0	3.3	0.5
m63f	0	3.7	0.1	1.3	0.0	2.9	0.1
m63f	500	6.6	0.2	1.3	0.0	5.0	0.1
m63f	5000	11.2	2.8	1.6	0.1	7.0	1.8
double	0	3.2	0.1	1.0	0.0	3.2	0.1
double	500	1.4	0.6	0.4	0.0	3.2	1.3
double	5000	2.5	0.2	0.6	0.0	4.4	0.3

F-F interaction	[Ca] (mM)	ddG	err	ddm	err	ddCm	err
	0	0.4	0.7	0.1	0.0	0.0	0.6
	0.5	5.2	1.0	0.9	0.0	1.7	1.8
	5	6.1	3.1	0.6	0.1	1.5	2.2

Table AB5. A9 mutant cycle HDXMS data. Summary exchange data after 3600 seconds for peptides from HDXMS experiments (corresponding to Figure 3.5e). See methods for collection details.

Protein State	Start	End	Sequence	Deut Time (sec)	max D	#D	%D	CI (#D)	Stddev
WT	2	8	TSKMSQL	3600	5	3.005	60.109	2.149	0.239
WT	6	13	SQLERNIE	3600	6	2.925	48.749		0
WT	8	13	LERNIE	3600	4	1.665	41.628	0.166	0.067
WT	8	14	LERNIET	3600	5	2.327	46.537	0.157	0.063
WT	9	15	ERNIETI	3600	5	2.325	46.504	0.157	0.063
WT	9	25	ERNIETIINTFHQYSVK	3600	15	7.444	49.627	1.063	0.668
WT	9	32	ERNIETIINTFHQYSVKLGHPDTL	3600	21	9.89	47.097	1.016	0.409
WT	9	37	ERNIETIINTFHQYSVKLGHPDTLNQGEF	3600	26	13.041	50.159	0.37	0.353
WT	9	39	ERNIETIINTFHQYSVKLGHPDTLNQGEFKE	3600	28	13.623	48.654	2.48	0.998
WT	10	17	RNIETIIN	3600	6	2.268	37.796	1.291	0.144
WT	11	17	NIETIIN	3600	5	1.789	35.789	0.687	0.277
WT	11	22	NIETIINTFHQY	3600	10	4.007	40.067	6.114	0.68
WT	11	25	NIETIINTFHQYSVK	3600	13	6.048	46.52	2.399	0.966
WT	12	17	IETIIN	3600	4	1.299	32.483	0.52	0.209
WT	12	25	IETIINTFHQYSVK	3600	12	5.716	47.63	3.259	0.363
WT	12	37	IETIINTFHQYSVKLGHPDTLNQGEF	3600	23	11.18	48.609	3.812	0.424
WT	12	39	IETIINTFHQYSVKLGHPDTLNQGEFKE	3600	25	11.683	46.731		0
WT	14	17	TIIN	3600	2	0.78	38.98	0.046	0.019
WT	14	22	TIINTFHQY	3600	7	3.161	45.161	0.144	0.058
WT	14	25	TIINTFHQYSVK	3600	10	4.646	46.457	0.303	0.122
WT	14	37	TIINTFHQYSVKLGHPDTLNQGEF	3600	21	10.927	52.033	0.3	0.285
WT	14	39	TIINTFHQYSVKLGHPDTLNQGEFKE	3600	23	12.162	52.88	0.524	0.5
WT	14	40	TIINTFHQYSVKLGHPDTLNQGEFKEL	3600	24	12.037	50.155	0.394	0.317
WT	15	25	IINTFHQYSVK	3600	9	3.953	43.92	0.115	0.109
WT	15	36	IINTFHQYSVKLGHPDTLNQGE	3600	19	8.886	46.767	0.575	0.463
WT	15	37	IINTFHQYSVKLGHPDTLNQGEF	3600	20	9.585	47.923	0.259	0.247
WT	15	39	IINTFHQYSVKLGHPDTLNQGEFKE	3600	22	10.243	46.561	0.356	0.385
WT	15	48	IINTFHQYSVKLGHPDTLNQGEFKELVRKDLQNF	3600	31	15.056	48.569	0.546	0.22
WT	18	22	TFHQY	3600	3	1.024	34.131	0.172	0.069
WT	18	25	TFHQYSVK	3600	6	2.895	48.245	0.101	0.096
WT	19	25	FHQYSVK	3600	5	2.205	44.096	0.367	0.148
WT	19	39	FHQYSVKLGHPDTLNQGEFKE	3600	18	9.72	54	2.417	0.973
WT	23	29	SVKLGHP	3600	4	2.512	62.795	0.126	0.12
WT	23	30	SVKLGHPD	3600	5	2.468	49.368	0.146	0.139
WT	23	32	SVKLGHPDTL	3600	7	3.608	51.536	0.337	0.321
WT	23	33	SVKLGHPDTLN	3600	8	4.364	54.547	0.203	0.193
WT	23	34	SVKLGHPDTLNQ	3600	9	5.049	56.1	0.227	0.216
WT	23	37	SVKLGHPDTLNQGEF	3600	12	6.603	55.029	0.197	0.188

WT	23	38	SVKLGHPDTLNQGEFK	3600	13	6.888	52.983	0.631	0.254
WT	23	39	SVKLGHPDTLNQGEFKE	3600	14	7.789	55.636	0.955	0.91
WT	23	40	SVKLGHPDTLNQGEFKEL	3600	15	7.697	51.316	0.143	0.186
WT	26	32	LGHPDTL	3600	4	1.475	36.87	0.127	0.051
WT	26	33	LGHPDTLN	3600	5	2.209	44.18	0.378	0.152
WT	26	37	LGHPDTLNQGEF	3600	9	4.511	50.121	0.312	0.126
WT	26	38	LGHPDTLNQGEFK	3600	10	4.683	46.83	3.105	0.346
WT	26	39	LGHPDTLNQGEFKE	3600	11	5.261	47.825	0.453	0.182
WT	26	40	LGHPDTLNQGEFKEL	3600	12	5.883	49.022	0.104	0.042
WT	26	48	LGHPDTLNQGEFKELVRKDLQNF	3600	20	10.659	53.297	0.69	0.278
WT	30	37	DTLNQGEF	3600	6	3.533	58.875	0.635	0.256
WT	30	39	DTLNQGEFKE	3600	8	3.84	47.994	0.408	0.164
WT	30	40	DTLNQGEFKEL	3600	9	5.449	60.546	0.061	0.024
WT	31	37	TLNQGEF	3600	5	2.916	58.321	0.508	0.205
WT	31	38	TLNQGEFK	3600	6	3.553	59.222	0.137	0.055
WT	31	39	TLNQGEFKE	3600	7	4.044	57.775	0.37	0.149
WT	31	40	TLNQGEFKEL	3600	8	4.404	55.046	0.307	0.124
WT	33	38	NQGEFK	3600	4	2.376	59.391	0.242	0.097
WT	33	39	NQGEFKE	3600	5	2.709	54.176	0.137	0.055
WT	33	40	NQGEFKEL	3600	6	3.301	55.022	0.515	0.207
WT	34	39	QGEFKE	3600	4	1.921	48.024	0.033	0.013
WT	34	40	QGEFKEL	3600	5	2.439	48.778	0.371	0.149
WT	35	40	GEFKEL	3600	4	2.087	52.175	0.291	0.117
WT	36	40	EFKEL	3600	3	1.457	48.58	0.327	0.132
WT	36	62	EFKELVRKDLQNFLKKENKNEKVIEHI	3600	25	12.844	51.376	1.533	0.617
WT	37	40	FKEL	3600	2	0.967	48.356	0.248	0.1
WT	38	47	KELVRKDLQN	3600	8	3.883	48.533	0.215	0.258
WT	38	48	KELVRKDLQNF	3600	9	4.462	49.574	0.142	0.185
WT	40	47	LVRKDLQN	3600	6	3.15	52.499	0.848	0.342
WT	40	49	LVRKDLQNFL	3600	8	5.269	65.864	0.531	0.427
WT	40	60	LVRKDLQNFLKKENKNEKVIE	3600	19	9.849	51.839	1.151	0.927
WT	40	64	LVRKDLQNFLKKENKNEKVIEHIME	3600	23	12.084	52.54	0.273	0.295
WT	41	49	VRKDLQNFL	3600	7	4.057	57.953	0.094	0.038
WT	41	50	VRKDLQNFLK	3600	8	4.763	59.538	0.122	0.146
WT	41	52	VRKDLQNFLKKE	3600	10	5.622	56.219	0.126	0.198
WT	41	54	VRKDLQNFLKKENK	3600	12	6.085	50.712	0.237	0.309
WT	41	57	VRKDLQNFLKKENKNEK	3600	15	6.621	44.142	0.285	0.448
WT	41	60	VRKDLQNFLKKENKNEKVIE	3600	18	8.945	49.695	0.479	0.457
WT	41	61	VRKDLQNFLKKENKNEKVIEH	3600	19	8.638	45.465	0.604	0.653
WT	44	48	DLQNF	3600	3	1.877	62.573	0.19	0.181
WT	44	49	DLQNFL	3600	4	2.686	67.142	0.163	0.155
WT	44	50	DLQNFLK	3600	5	3.666	73.317	0.157	0.063
WT	44	54	DLQNFLKKENK	3600	9	5.612	62.351	0.297	0.283
WT	44	61	DLQNFLKKENKNEKVIEH	3600	16	8.733	54.58	0.598	0.57
WT	45	50	LQNFLK	3600	4	2.688	67.203	0.196	0.079
WT	45	52	LQNFLKKE	3600	6	3.078	51.303	0.272	0.171
WT	45	54	LQNFLKKENK	3600	8	4.06	50.751	0.267	0.254
WT	45	57	LQNFLKKENKNEK	3600	11	4.671	42.462	0.252	0.302
WT	45	61	LQNFLKKENKNEKVIEH	3600	15	7.213	48.09	0.88	1.145
WT	47	50	NFLK	3600	2	1.54	77.013	0.102	0.041

WT	47	51	NFLKK	3600	3	2.325	77.503	0.181	0.073
WT	47	52	NFLKKE	3600	4	2.701	67.523	0.085	0.081
WT	47	54	NFLKKENK	3600	6	3.089	51.481	0.742	0.299
WT	47	57	NFLKKENKNEK	3600	9	3.847	42.74	0.88	0.354
WT	47	61	NFLKKENKNEKVIEH	3600	13	5.716	43.973	0.4	0.382
WT	47	63	NFLKKENKNEKVIEHIM	3600	15	5.204	34.694	0.563	0.063
WT	48	57	FLKKENKNEK	3600	8	3.436	42.948	0.839	0.338
WT	48	60	FLKKENKNEKVIE	3600	11	4.823	43.843	0.296	0.282
WT	48	61	FLKKENKNEKVIEH	3600	12	4.756	39.631	0.244	0.318
WT	48	63	FLKKENKNEKVIEHIM	3600	14	4.336	30.972	0.243	0.316
WT	48	64	FLKKENKNEKVIEHIME	3600	15	8.124	54.16	1.502	1.21
WT	48	65	FLKKENKNEKVIEHIMED	3600	16	5.621	35.131	0.751	0.716
WT	48	66	FLKKENKNEKVIEHIMEDL	3600	17	7.79	45.821	6.339	0.706
WT	49	60	LKKENKNEKVIE	3600	10	4.236	42.356	0.267	0.32
WT	49	61	LKKENKNEKVIEH	3600	11	4.36	39.632	0.284	0.37
WT	49	66	LKKENKNEKVIEHIMEDL	3600	16	7.619	47.619	0.689	0.824
WT	50	62	KKENKNEKVIEHI	3600	11	4.397	39.974	0.979	0.394
WT	50	66	KKENKNEKVIEHIMEDL	3600	15	3.216	21.441	1.827	0.736
WT	58	63	VIEHIM	3600	4	1.837	45.937	0.27	0.03
WT	59	63	IEHIM	3600	3	1.349	44.962	1.68	0.187
WT	64	78	EDLDTNADKQLSFEE	3600	13	6.619	50.914	0.414	0.167
WT	64	79	EDLDTNADKQLSFEEF	3600	14	6.388	45.63		0
WT	65	74	DLDTNADKQL	3600	8	4.301	53.76	0.223	0.09
WT	65	75	DLDTNADKQLS	3600	9	5.105	56.726	0.405	0.163
WT	65	76	DLDTNADKQLSF	3600	10	5.115	51.148	0.287	0.274
WT	65	79	DLDTNADKQLSFEEF	3600	13	5.651	43.473	0.377	0.359
WT	66	75	LDTNADKQLS	3600	8	4.235	52.943	1.306	0.526
WT	66	78	LDTNADKQLSFEE	3600	11	5.542	50.377	0.363	0.228
WT	67	78	DTNADKQLSFEE	3600	10	5.362	53.616	0.689	0.278
WT	67	79	DTNADKQLSFEEF	3600	11	4.896	44.506	0.49	0.197
WT	70	76	ADKQLSF	3600	5	2.975	59.506		0
WT	70	79	ADKQLSFEEF	3600	8	3.38	42.244	0.524	0.211
WT	71	76	DKQLSF	3600	4	2.401	60.03	0.567	0.228
WT	73	79	QLSFEEF	3600	5	2.117	42.348	0.109	0.103
WT	74	79	LSFEEF	3600	4	1.346	33.643	0.268	0.108
WT	75	79	SFEEF	3600	3	0.827	27.576	0.044	0.018
WT	75	80	SFEEFI	3600	4	1.35	33.748	0.249	0.1
WT	84	114	ARLTWASHEKMHEGDEGPGHHHKPGLGEGTP	3600	26	22.267	85.644	0.487	0.464
WT	85	88	RLTW	3600	2	1.427	71.365	0.176	0.071
WT	85	91	RLTWASH	3600	5	2.406	48.129	0.223	0.212
WT	85	114	RLTWASHEKMHEGDEGPGHHHKPGLGEGTP	3600	25	21.738	86.951	1.302	0.524
WT	86	91	LTWASH	3600	4	2.225	55.617	0.274	0.11
WT	86	97	LTWASHEKMHEG	3600	10	2.219	22.187	0.321	0.259
WT	87	91	TWASH	3600	3	1.397	46.564	0.369	0.149
WT	87	114	TWASHEKMHEGDEGPGHHHKPGLGEGTP	3600	23	20.945	91.067	0.706	0.763
WT	94	114	MHEGDEGPGHHHKPGLGEGTP	3600	16	1.562	9.764	0.367	0.477

WT	95	114	HEGDEGPGHHHKPGLGEGTP	3600	15	3.188	21.253	0.51	0.486
WT	96	110	EGDEGPGHHHKPGLG	3600	11	1.849	16.81	0.148	0.141
WT	96	114	EGDEGPGHHHKPGLGEGTP	3600	14	3.075	21.966	0.138	0.216
WT	97	110	GDEGPGHHHKPGLG	3600	10	1.818	18.181	0.1	0.095
WT	97	114	GDEGPGHHHKPGLGEGTP	3600	13	2.984	22.952	0.169	0.22
WT	98	110	DEGPGHHHKPGLG	3600	9	1.889	20.992	0.115	0.149
WT	98	114	DEGPGHHHKPGLGEGTP	3600	12	3.075	25.621	0.179	0.233
WT	99	110	EGPGHHHKPGLG	3600	8	1.8	22.497	0.083	0.079
WT	99	114	EGPGHHHKPGLGEGTP	3600	11	3.028	27.523	0.22	0.263
F37L	2	8	TSKMSQL	3600	5	2.981	59.612	0.681	0.076
F37L	6	13	SQLERNIE	3600	6	4.351	72.512		0
F37L	8	13	LERNIE	3600	4	2.587	64.687	0.284	0.114
F37L	8	14	LERNIET	3600	5	3.466	69.312	0.392	0.158
F37L	9	15	ERNIETI	3600	5	3.449	68.986	0.431	0.173
F37L	9	25	ERNIETIINTFHQYSVK	3600	15	10.257	68.379	0.62	0.591
F37L	9	32	ERNIETIINTFHQYSVKLGHPDTL	3600	21	14.024	66.781	2.934	1.181
F37L	9	37	ERNIETIINTFHQYSVKLGHPDTLNQGEL	3600	26	15.972	61.432	1.473	0.926
F37L	9	39	ERNIETIINTFHQYSVKLGHPDTLNQGELKE	3600	28	16.938	60.493	0.477	0.621
F37L	10	17	RNIETIIN	3600	6	4.369	72.809	0.917	0.369
F37L	11	17	NIETIIN	3600	5	3.212	64.233	0.706	0.284
F37L	11	22	NIETIINTFHQY	3600	10	8.056	80.559	8.172	0.91
F37L	11	25	NIETIINTFHQYSVK	3600	13	9.316	71.664	1.02	0.411
F37L	12	17	IETIIN	3600	4	2.412	60.304	0.57	0.229
F37L	12	25	IETIINTFHQYSVK	3600	12	7.01	58.418	3.039	1.223
F37L	12	37	IETIINTFHQYSVKLGHPDTLNQGEL	3600	23	14.14	61.479	5.213	0.58
F37L	12	39	IETIINTFHQYSVKLGHPDTLNQGELKE	3600	25	14.877	59.508	1.859	1.168
F37L	14	17	TIIN	3600	2	1.477	73.827	0.196	0.079
F37L	14	22	TIINTFHQY	3600	7	4.587	65.522	0.413	0.166
F37L	14	25	TIINTFHQYSVK	3600	10	6.048	60.481	0.21	0.2
F37L	14	37	TIINTFHQYSVKLGHPDTLNQGEL	3600	21	12.348	58.802	0.626	0.597
F37L	14	39	TIINTFHQYSVKLGHPDTLNQGELKE	3600	23	13.432	58.401	0.423	0.551
F37L	14	40	TIINTFHQYSVKLGHPDTLNQGELKEL	3600	24	13.722	57.174	1.336	1.274
F37L	15	25	IINTFHQYSVK	3600	9	4.771	53.008	0.229	0.184
F37L	15	36	IINTFHQYSVKLGHPDTLNQGE	3600	19	9.145	48.132	0.777	0.488
F37L	15	37	IINTFHQYSVKLGHPDTLNQGEL	3600	20	10.224	51.121	0.398	0.379
F37L	15	39	IINTFHQYSVKLGHPDTLNQGELKE	3600	22	11.303	51.375	0.339	0.441
F37L	15	48	IINTFHQYSVKLGHPDTLNQGELKELVRKDLQNF	3600	31	17.267	55.701		0
F37L	18	22	TFHQY	3600	3	1.125	37.513	0.261	0.105
F37L	18	25	TFHQYSVK	3600	6	2.994	49.904	0.131	0.124
F37L	19	25	FHQYSVK	3600	5	2.132	42.643	0.332	0.209
F37L	19	39	FHQYSVKLGHPDTLNQGELKE	3600	18	9.642	53.569	0.684	0.651
F37L	23	29	SVKLGHP	3600	4	2.543	63.569	0.086	0.082
F37L	23	30	SVKLGHPD	3600	5	2.487	49.731	0.11	0.105

F37L	23	32	SVKLGHPDTL	3600	7	3.592	51.314	0.35	0.333
F37L	23	33	SVKLGHPDTLN	3600	8	4.367	54.587	0.215	0.204
F37L	23	34	SVKLGHPDTLNQ	3600	9	5.08	56.441	0.189	0.18
F37L	23	37	SVKLGHPDTLNQGEL	3600	12	6.71	55.913	0.231	0.22
F37L	23	38	SVKLGHPDTLNQGELK	3600	13	7.314	56.26	0.637	0.256
F37L	23	39	SVKLGHPDTLNQGELKE	3600	14	8.308	59.343	0.507	0.66
F37L	23	40	SVKLGHPDTLNQGELKEL	3600	15	8.569	57.129	0.259	0.337
F37L	26	32	LGHPDTL	3600	4	1.486	37.148	0.145	0.058
F37L	26	33	LGHPDTLN	3600	5	2.2	44.008	0.299	0.12
F37L	26	37	LGHPDTLNQGEL	3600	9	4.64	51.559	0.287	0.115
F37L	26	38	LGHPDTLNQGELK	3600	10	5.363	53.625	0.668	0.269
F37L	26	39	LGHPDTLNQGELKE	3600	11	6.001	54.551	0.171	0.163
F37L	26	40	LGHPDTLNQGELKEL	3600	12	6.741	56.176	0.222	0.212
F37L	26	48	LGHPDTLNQGELKELVRKDLQNF	3600	20	13.353	66.764	0.242	0.195
F37L	30	37	DTLNQGEL	3600	6	3.867	64.449	0.713	0.287
F37L	30	39	DTLNQGELKE	3600	8	5.563	69.532	0.688	0.277
F37L	30	40	DTLNQGELKEL	3600	9	6.142	68.244	0.445	0.179
F37L	31	37	TLNQGEL	3600	5	3.085	61.709	0.543	0.219
F37L	31	38	TLNQGELK	3600	6	3.902	65.03	0.342	0.137
F37L	31	39	TLNQGELKE	3600	7	4.336	61.946	0.382	0.154
F37L	31	40	TLNQGELKEL	3600	8	5.143	64.287	0.316	0.127
F37L	33	38	NQGELK	3600	4				
F37L	33	39	NQGELKE	3600	5				
F37L	33	40	NQGELKEL	3600	6	3.963	66.048	0.361	0.145
F37L	34	39	QGELKE	3600	4				
F37L	34	40	QGELKEL	3600	5	3.293	65.867	0.287	0.116
F37L	35	40	GELKEL	3600	4	2.898	72.458	0.278	0.112
F37L	36	40	ELKEL	3600	3	2.093	69.765	0.241	0.097
F37L	36	62	ELKELVRKDLQNFLKKENKNEKVIEHI	3600	25				
F37L	37	40	LKEL	3600	2	1.445	72.228	0.167	0.067
F37L	38	47	KELVRKDLQN	3600	8	4.937	61.714	1.789	1.124
F37L	38	48	KELVRKDLQNF	3600	9	5.262	58.472	0.507	0.483
F37L	40	47	LVRKDLQN	3600	6	3.852	64.2	0.643	0.259
F37L	40	49	LVRKDLQNFL	3600	8	5.607	70.093	0.303	0.289
F37L	40	60	LVRKDLQNFLKKENKNEKVIE	3600	19	10.57	55.633	0.515	0.557
F37L	40	64	LVRKDLQNFLKKENKNEKVIEHIME	3600	23	12.1	52.61	0.537	0.216
F37L	41	49	VRKDLQNFL	3600	7	4.129	58.988	1.348	0.15
F37L	41	50	VRKDLQNFLK	3600	8	4.902	61.28	0.234	0.28
F37L	41	52	VRKDLQNFLKKE	3600	10	5.802	58.021	0.141	0.222
F37L	41	54	VRKDLQNFLKKENK	3600	12	6.325	52.71	0.252	0.328
F37L	41	57	VRKDLQNFLKKENKNEK	3600	15	6.885	45.903	0.322	0.48
F37L	41	60	VRKDLQNFLKKENKNEKVIE	3600	18	9.157	50.873	0.459	0.437
F37L	41	61	VRKDLQNFLKKENKNEKVIEH	3600	19	8.813	46.387	0.905	0.729
F37L	44	48	DLQNF	3600	3	1.88	62.652	0.188	0.179
F37L	44	49	DLQNFL	3600	4	2.728	68.211	0.164	0.157
F37L	44	50	DLQNFLK	3600	5	3.742	74.839	0.318	0.128
F37L	44	54	DLQNFLKKENK	3600	9	5.652	62.799	0.394	0.159
F37L	44	61	DLQNFLKKENKNEKVIEH	3600	16	8.763	54.767	0.52	0.495
F37L	45	50	LQNFLK	3600	4	2.704	67.607	0.183	0.074
F37L	45	52	LQNFLKKE	3600	6	3.37	56.174	0.42	0.169

F37L	45	54	LQNFLKKENK	3600	8	4.118	51.472	0.158	0.151
F37L	45	57	LQNFLKKENKNEK	3600	11	4.887	44.427	0.223	0.267
F37L	45	61	LQNFLKKENKNEKV IEH	3600	15	7.141	47.603	0.829	1.078
F37L	47	50	NFLK	3600	2	1.544	77.217	0.106	0.043
F37L	47	51	NFLKK	3600	3	2.475	82.492	0.116	0.047
F37L	47	52	NFLKKE	3600	4	2.716	67.901	0.066	0.063
F37L	47	54	NFLKKENK	3600	6	3.118	51.966	0.318	0.128
F37L	47	57	NFLKKENKNEK	3600	9	3.948	43.87	0.781	0.314
F37L	47	61	NFLKKENKNEKVIE H	3600	13	5.74	44.154	0.292	0.278
F37L	47	63	NFLKKENKNEKVIE HIM	3600	15	5.094	33.961	0.821	0.33
F37L	48	57	FLKKENKNEK	3600	8	3.452	43.154	0.887	0.357
F37L	48	60	FLKKENKNEKVIE	3600	11	4.886	44.418	0.285	0.271
F37L	48	61	FLKKENKNEKVIEH	3600	12	4.786	39.887	0.212	0.275
F37L	48	63	FLKKENKNEKVIEHI M	3600	14	4.438	31.699	0.417	0.397
F37L	48	64	FLKKENKNEKVIEHI ME	3600	15	7.621	50.809	1.298	1.403
F37L	48	65	FLKKENKNEKVIEHI MED	3600	16	5.651	35.319	0.954	0.769
F37L	48	66	FLKKENKNEKVIEHI MEDL	3600	17	7.914	46.552	0.549	0.345
F37L	49	60	LKKENKNEKVIE	3600	10	4.247	42.475	0.15	0.195
F37L	49	61	LKKENKNEKVIEH	3600	11	4.357	39.605	0.21	0.274
F37L	49	66	LKKENKNEKVIEHIM EDL	3600	16	7.11	44.435	0.555	0.529
F37L	50	62	KKENKNEKVIEHI	3600	11	4.331	39.373	1.013	0.408
F37L	50	66	KKENKNEKVIEHIME DL	3600	15	5.744	38.294	1.806	0.727
F37L	58	63	VIEHIM	3600	4	1.714	42.854	1.713	0.191
F37L	59	63	IEHIM	3600	3	0.999	33.31	3.229	0.359
F37L	64	78	EDLDTNADKQLSFE E	3600	13	7.354	56.572	1.036	0.417
F37L	64	79	EDLDTNADKQLSFE EF	3600	14	8.24	58.858	4.187	0.466
F37L	65	74	DLDTNADKQL	3600	8	4.549	56.856	0.409	0.165
F37L	65	75	DLDTNADKQLS	3600	9	5.562	61.799	0.359	0.144
F37L	65	76	DLDTNADKQLSF	3600	10	5.752	57.519	0.329	0.313
F37L	65	79	DLDTNADKQLSFEEF	3600	13	7.432	57.171	0.202	0.193
F37L	66	75	LDTNADKQLS	3600	8	4.834	60.426	1.113	0.448
F37L	66	78	LDTNADKQLSFEE	3600	11	6.388	58.074	0.511	0.321
F37L	67	78	DTNADKQLSFEE	3600	10	5.999	59.991	0.673	0.271
F37L	67	79	DTNADKQLSFEEF	3600	11	6.058	55.075	2.904	1.169
F37L	70	76	ADKQLSF	3600	5	3.288	65.751	0.242	0.097
F37L	70	79	ADKQLSFEEF	3600	8	4.853	60.666	0.697	0.281
F37L	71	76	DKQLSF	3600	4	2.743	68.585	0.154	0.062
F37L	73	79	QLSFEEF	3600	5	3.093	61.862	0.377	0.359
F37L	74	79	LSFEEF	3600	4	2.093	52.337	0.779	0.314
F37L	75	79	SFEEF	3600	3	1.738	57.926	0.174	0.166
F37L	75	80	SFEEFI	3600	4	2.105	52.636	0.778	0.313
F37L	84	114	ARLTWASHEKMHE GDEGPGHHHKPGLG EGTP	3600	26	22.887	88.027	0.393	0.247
F37L	85	88	RLTW	3600	2	1.345	67.256	0.126	0.051
F37L	85	91	RLTWASH	3600	5	2.381	47.618	0.167	0.159
F37L	85	114	RLTWASHEKMHEG DEGPGHHHKPGLGE GTP	3600	25	22.021	88.082	0.281	0.176
F37L	86	91	LTWASH	3600	4	2.152	53.789	0.277	0.112
F37L	86	97	LTWASHEKMHEG	3600	10	2.189	21.888		0
F37L	87	91	TWASH	3600	3	1.35	45.016	0.329	0.133

F37L	87	114	TWASHEKMHEGDEGPGHHHKPGLGEGTP	3600	23	21.604	93.929	0.333	0.268
F37L	94	114	MHEGDEGPGHHHKPGLGEGTP	3600	16	1.913	11.957	0.538	0.7
F37L	95	114	HEGDEGPGHHHKPGLGEGTP	3600	15	2.918	19.45	0.465	0.187
F37L	96	110	EGDEGPGHHHKPGLG	3600	11	1.729	15.718	0.222	0.179
F37L	96	114	EGDEGPGHHHKPGLGEGTP	3600	14	3.163	22.592	0.123	0.194
F37L	97	110	GDEGPGHHHKPGLG	3600	10	1.744	17.444	0.114	0.108
F37L	97	114	GDEGPGHHHKPGLGEGTP	3600	13	3.115	23.961	0.169	0.219
F37L	98	110	DEGPGHHHKPGLG	3600	9	1.53	17.003	0.111	0.089
F37L	98	114	DEGPGHHHKPGLGEGTP	3600	12	3.096	25.801	0.127	0.166
F37L	99	110	EGPGHHHKPGLG	3600	8	1.768	22.1	0.201	0.191
F37L	99	114	EGPGHHHKPGLGEGTP	3600	11	3.054	27.76	0.202	0.242
M63F	2	8	TSKMSQL	3600	5	2.864	57.273	1.675	0.186
M63F	6	13	SQLERNIE	3600	6	1.469	24.483		0
M63F	8	13	LERNIE	3600	4	1.077	26.925	0.215	0.087
M63F	8	14	LERNIET	3600	5	1.736	34.721	3.752	0.418
M63F	9	15	ERNIETI	3600	5	1.787	35.744	0.75	0.302
M63F	9	25	ERNIETIINTFHQYSVK	3600	15	6.055	40.367	2.725	1.713
M63F	9	32	ERNIETIINTFHQYSVKLGHPDTL	3600	21	5.141	24.482		0
M63F	9	37	ERNIETIINTFHQYSVKLGHPDTLNQGEF	3600	26	8.142	31.316	0.481	0.302
M63F	9	39	ERNIETIINTFHQYSVKLGHPDTLNQGEFKE	3600	28				
M63F	10	17	RNIETIIN	3600	6	1.415	23.588	0.162	0.065
M63F	11	17	NIETIIN	3600	5	1.142	22.842	1.422	0.158
M63F	11	22	NIETIINTFHQY	3600	10	2.039	20.388	1.279	0.515
M63F	11	25	NIETIINTFHQYSVK	3600	13	3.575	27.498	1.152	0.464
M63F	12	17	lETIIN	3600	4	0.87	21.762	0.497	0.2
M63F	12	25	lETIINTFHQYSVK	3600	12	3.045	25.372	1.892	0.211
M63F	12	37	lETIINTFHQYSVKLGHPDTLNQGEF	3600	23	6.113	26.577	0.621	0.25
M63F	12	39	lETIINTFHQYSVKLGHPDTLNQGEFKE	3600	25				
M63F	14	17	TIIN	3600	2	0.391	19.563	0.038	0.015
M63F	14	22	TIINTFHQY	3600	7	1.439	20.561	0.025	0.01
M63F	14	25	TIINTFHQYSVK	3600	10	2.623	26.225	0.191	0.12
M63F	14	37	TIINTFHQYSVKLGHPDTLNQGEF	3600	21	6.691	31.86	0.199	0.19
M63F	14	39	TIINTFHQYSVKLGHPDTLNQGEFKE	3600	23	6.542	28.445	0.744	0.083
M63F	14	40	TIINTFHQYSVKLGHPDTLNQGEFKEL	3600	24	7.081	29.503	2.986	1.202
M63F	15	25	IINTFHQYSVK	3600	9	2.144	23.826	0.219	0.209
M63F	15	36	IINTFHQYSVKLGHPDTLNQGE	3600	19	6.091	32.057	0.67	0.638
M63F	15	37	IINTFHQYSVKLGHPDTLNQGEF	3600	20	5.883	29.416	0.106	0.101
M63F	15	39	IINTFHQYSVKLGHPDTLNQGEFKE	3600	22	6.07	27.59	0.457	0.495
M63F	15	48	IINTFHQYSVKLGHPDTLNQGEFKELVRKDLQNF	3600	31	8.734	28.175	1.068	0.671
M63F	18	22	TFHQY	3600	3	0.18	6.008	0.077	0.031
M63F	18	25	TFHQYSVK	3600	6	1.593	26.549	0.04	0.038
M63F	19	25	FHQYSVK	3600	5	1.123	22.468	0.658	0.413

M63F	19	39	FHQYSVKLGHPDTLNQGEFKE	3600	18	5.611	31.174	13.135	1.462
M63F	23	29	SVKLGHP	3600	4	2.454	61.348	0.127	0.121
M63F	23	30	SVKLGHPD	3600	5	2.41	48.198	0.161	0.153
M63F	23	32	SVKLGHPDTL	3600	7	2.595	37.076	0.298	0.24
M63F	23	33	SVKLGHPDTLN	3600	8	2.823	35.282	0.157	0.15
M63F	23	34	SVKLGHPDTLNQ	3600	9	3.587	39.857	0.22	0.21
M63F	23	37	SVKLGHPDTLNQGEF	3600	12	4.708	39.234	0.178	0.17
M63F	23	38	SVKLGHPDTLNQGEFK	3600	13	4.509	34.682	0.907	0.365
M63F	23	39	SVKLGHPDTLNQGEFKE	3600	14	4.43	31.641	0.304	0.245
M63F	23	40	SVKLGHPDTLNQGEFKEL	3600	15	4.384	29.227	0.152	0.198
M63F	26	32	LGHPDTL	3600	4	0.464	11.596	0.128	0.051
M63F	26	33	LGHPDTLN	3600	5	0.727	14.534	0.266	0.107
M63F	26	37	LGHPDTLNQGEF	3600	9	2.639	29.32	0.241	0.097
M63F	26	38	LGHPDTLNQGEFK	3600	10	2.503	25.031	0.429	0.173
M63F	26	39	LGHPDTLNQGEFKE	3600	11	2.51	22.817	0.152	0.145
M63F	26	40	LGHPDTLNQGEFKEL	3600	12	2.561	21.341	0.253	0.102
M63F	26	48	LGHPDTLNQGEFKELVRKDLQNF	3600	20	5.508	27.541	1.247	0.502
M63F	30	37	DTLNQGEF	3600	6	2.173	36.221	0.963	0.387
M63F	30	39	DTLNQGEFKE	3600	8	2.511	31.391	0.09	0.01
M63F	30	40	DTLNQGEFKEL	3600	9	2.515	27.946	0.133	0.054
M63F	31	37	TLNQGEF	3600	5	2.194	43.885	0.381	0.154
M63F	31	38	TLNQGEFK	3600	6	2.128	35.461	0.294	0.118
M63F	31	39	TLNQGEFKE	3600	7	2.094	29.915	0.233	0.094
M63F	31	40	TLNQGEFKEL	3600	8	2.15	26.871	0.164	0.066
M63F	33	38	NQGEFK	3600	4	1.115	27.88	0.402	0.162
M63F	33	39	NQGEFKE	3600	5	1.104	22.072	0.349	0.141
M63F	33	40	NQGEFKEL	3600	6	1.213	20.213	0.198	0.08
M63F	34	39	QGEFKE	3600	4	0.477	11.92	0.305	0.123
M63F	34	40	QGEFKEL	3600	5	0.619	12.372	0.073	0.029
M63F	35	40	GEFKEL	3600	4	0.331	8.278	0.083	0.033
M63F	36	40	EFKEL	3600	3	0.174	5.786	0.025	0.01
M63F	36	62	EFKELVRKDLQNFLKKENKNEKVIEHI	3600	25	10.512	42.047	2.182	0.878
M63F	37	40	FKEL	3600	2	0.093	4.668	0.028	0.011
M63F	38	47	KELVRKDLQN	3600	8	2.307	28.843	0.176	0.21
M63F	38	48	KELVRKDLQNF	3600	9	2.882	32.017	0.089	0.116
M63F	40	47	LVRKDLQN	3600	6	2.113	35.217	0.86	0.541
M63F	40	49	LVRKDLQNFL	3600	8	3.876	48.447	1.302	0.524
M63F	40	60	LVRKDLQNFLKKENKNEKVIE	3600	19	5.949	31.311		0
M63F	40	64	LVRKDLQNFLKKENKNEKVIEHIFE	3600	23	11.51	50.045	2.214	0.891
M63F	41	49	VRKDLQNFL	3600	7				
M63F	41	50	VRKDLQNFLK	3600	8	4.382	54.769	0.169	0.22
M63F	41	52	VRKDLQNFLKKE	3600	10	5.165	51.649	0.179	0.281
M63F	41	54	VRKDLQNFLKKENK	3600	12	5.632	46.932	0.359	0.43
M63F	41	57	VRKDLQNFLKKENKNEK	3600	15	6.238	41.588	0.347	0.516
M63F	41	60	VRKDLQNFLKKENKNEKVIE	3600	18	8.303	46.127	0.465	0.374
M63F	41	61	VRKDLQNFLKKENKNEKVIEH	3600	19	8.343	43.909	0.673	0.728
M63F	44	48	DLQNF	3600	3	1.843	61.434	0.24	0.229
M63F	44	49	DLQNFL	3600	4	2.762	69.042	0.271	0.258
M63F	44	50	DLQNFLK	3600	5	3.572	71.443	0.52	0.209
M63F	44	54	DLQNFLKKENK	3600	9	5.627	62.527	0.519	0.326

M63F	44	61	DLQNFLKKENKNEK VIEH	3600	16	8.268	51.676	1.518	0.611
M63F	45	50	LQNFLK	3600	4	2.652	66.312	0.261	0.105
M63F	45	52	LQNFLKKE	3600	6	2.654	44.24	0.158	0.099
M63F	45	54	LQNFLKKENK	3600	8	4.032	50.399	0.323	0.308
M63F	45	57	LQNFLKKENKNEK	3600	11	4.625	42.044	0.461	0.552
M63F	45	61	LQNFLKKENKNEKV IEH	3600	15	6.641	44.272	0.317	0.412
M63F	47	50	NFLK	3600	2	1.516	75.806	0.105	0.042
M63F	47	51	NFLKK	3600	3	2.33	77.652	0.238	0.096
M63F	47	52	NFLKKE	3600	4	2.667	66.663	0.096	0.091
M63F	47	54	NFLKKENK	3600	6	3.19	53.161	0.534	0.215
M63F	47	57	NFLKKENKNEK	3600	9	3.856	42.843	0.832	0.335
M63F	47	61	NFLKKENKNEKVIE H	3600	13	5.699	43.841	0.476	0.453
M63F	47	63	NFLKKENKNEKVIE HIF	3600	15				
M63F	48	57	FLKKENKNEK	3600	8	3.441	43.007	0.968	0.39
M63F	48	60	FLKKENKNEKVIE	3600	11	4.656	42.324	0.454	0.433
M63F	48	61	FLKKENKNEKVIEH	3600	12	4.672	38.931	0.233	0.303
M63F	48	63	FLKKENKNEKVIEHI F	3600	14	5.114	36.532	0.253	0.329
M63F	48	64	FLKKENKNEKVIEHI FE	3600	15	5.348	35.652	0.633	0.51
M63F	48	65	FLKKENKNEKVIEHI FED	3600	16	5.834	36.461	1.253	0.504
M63F	48	66	FLKKENKNEKVIEHI FEDL	3600	17	5.638	33.164	0.375	0.488
M63F	49	60	LKKENKNEKVIE	3600	10	4.18	41.802	2.045	0.228
M63F	49	61	LKKENKNEKVIEH	3600	11	4.39	39.909	0.614	0.664
M63F	49	66	LKKENKNEKVIEHIF EDL	3600	16	5.132	32.074	0.272	0.38
M63F	50	62	KKENKNEKVIEHI	3600	11	4.35	39.542	1.644	0.662
M63F	50	66	KKENKNEKVIEHIFE DL	3600	15	4.75	31.67	0.378	0.361
M63F	58	63	VIEHIF	3600	4				
M63F	59	63	IEHIF	3600	3				
M63F	64	78	EDLDTNADKQLSFE E	3600	13	3.764	28.954	0.952	0.383
M63F	64	79	EDLDTNADKQLSFE EF	3600	14	3.637	25.975	1.343	0.541
M63F	65	74	DLDTNADKQL	3600	8	2.447	30.586	0.225	0.091
M63F	65	75	DLDTNADKQLS	3600	9	2.278	25.314	4.91	0.546
M63F	65	76	DLDTNADKQLSF	3600	10	2.702	27.018	0.3	0.241
M63F	65	79	DLDTNADKQLSFEEF	3600	13	3.059	23.534	0.438	0.353
M63F	66	75	LDTNADKQLS	3600	8	2.251	28.141	0.291	0.117
M63F	66	78	LDTNADKQLSFEE	3600	11	3.355	30.499	3.05	0.339
M63F	67	78	DTNADKQLSFEE	3600	10	3.142	31.417	0.665	0.268
M63F	67	79	DTNADKQLSFEEF	3600	11	2.831	25.736	0.559	0.225
M63F	70	76	ADKQLSF	3600	5	1.461	29.228	0.072	0.008
M63F	70	79	ADKQLSFEEF	3600	8	1.77	22.124	0.309	0.124
M63F	71	76	DKQLSF	3600	4	1.017	25.426	0.282	0.113
M63F	73	79	QLSFEEF	3600	5	1.444	28.877	0.075	0.071
M63F	74	79	LSFEEF	3600	4	1.265	31.626	0.535	0.215
M63F	75	79	SFEEF	3600	3	0.623	20.781	0.102	0.041
M63F	75	80	SFEEFI	3600	4	1.31	32.751	0.516	0.208
M63F	84	114	ARLTWASHEKMHE GDEGPGHHHKPGLG EGTP	3600	26	21.644	83.248	0.521	0.497
M63F	85	88	RLTW	3600	2	1.423	71.134	0.484	0.195
M63F	85	91	RLTWASH	3600	5	2.359	47.187	0.291	0.277

M63F	85	114	RLTWASHEKMHEGDEGPGHHHKPGLGEGTP	3600	25	21.69	86.761	1.363	0.549
M63F	86	91	LTWASH	3600	4	2.222	55.538	0.347	0.14
M63F	86	97	LTWASHEKMHEG	3600	10	2.235	22.347	0.24	0.228
M63F	87	91	TWASH	3600	3	1.329	44.315	0.453	0.182
M63F	87	114	TWASHEKMHEGDEGPGHHHKPGLGEGTP	3600	23	20.849	90.648	0.501	0.478
M63F	94	114	MHEGDEGPGHHHKPGLGEGTP	3600	16	1.134	7.087	0.431	0.173
M63F	95	114	HEGDEGPGHHHKPGLGEGTP	3600	15	2.932	19.547	0.896	0.563
M63F	96	110	EGDEGPGHHHKPGLG	3600	11	1.858	16.893	0.136	0.129
M63F	96	114	EGDEGPGHHHKPGLGEGTP	3600	14	3.005	21.463	0.151	0.238
M63F	97	110	GDEGPGHHHKPGLG	3600	10	1.824	18.236	0.102	0.097
M63F	97	114	GDEGPGHHHKPGLGEGTP	3600	13	2.944	22.647	0.225	0.292
M63F	98	110	DEGPGHHHKPGLG	3600	9	1.924	21.383	0.121	0.131
M63F	98	114	DEGPGHHHKPGLGEGTP	3600	12	3.147	26.221	0.17	0.221
M63F	99	110	EGPGHHHKPGLG	3600	8	1.747	21.838	0.104	0.099
M63F	99	114	EGPGHHHKPGLGEGTP	3600	11	2.871	26.102	0.215	0.279
F37L/M63F	2	8	TSKMSQL	3600	5	2.833	56.668	0.406	0.045
F37L/M63F	6	13	SQLERNIE	3600	6	3.701	61.689		0
F37L/M63F	8	13	LERNIE	3600	4	2.549	63.728	0.305	0.123
F37L/M63F	8	14	LERNIET	3600	5	3.173	63.453	0.207	0.023
F37L/M63F	9	15	ERNIETI	3600	5	3.341	66.817	2.529	0.281
F37L/M63F	9	25	ERNIETIINTFHQYSVK	3600	15	10.433	69.556	1.238	0.997
F37L/M63F	9	32	ERNIETIINTFHQYSVKLGHPDTL	3600	21	12.067	57.46		0
F37L/M63F	9	37	ERNIETIINTFHQYSVKLGHPDTLNQGEL	3600	26	16.449	63.264	2.24	0.249
F37L/M63F	9	39	ERNIETIINTFHQYSVKLGHPDTLNQGELKE	3600	28	16.459	58.781	0.484	0.579
F37L/M63F	10	17	RNIETIIN	3600	6	4.159	69.308	0.633	0.255
F37L/M63F	11	17	NIETIIN	3600	5	2.934	58.681	0.479	0.053
F37L/M63F	11	22	NIETIINTFHQY	3600	10	7.215	72.149	1.041	0.419
F37L/M63F	11	25	NIETIINTFHQYSVK	3600	13	8.745	67.268	1.531	0.616
F37L/M63F	12	17	IETIIN	3600	4	2.326	58.162	0.76	0.306
F37L/M63F	12	25	IETIINTFHQYSVK	3600	12	7.234	60.287	3.035	0.338
F37L/M63F	12	37	IETIINTFHQYSVKLGHPDTLNQGEL	3600	23	13.728	59.685	9.583	1.067
F37L/M63F	12	39	IETIINTFHQYSVKLGHPDTLNQGELKE	3600	25	14.269	57.077	3.356	0.373
F37L/M63F	14	17	TIIN	3600	2	1.387	69.357	0.207	0.083
F37L/M63F	14	22	TIINTFHQY	3600	7	4.858	69.407	0.806	0.324
F37L/M63F	14	25	TIINTFHQYSVK	3600	10	5.818	58.184	0.194	0.185
F37L/M63F	14	37	TIINTFHQYSVKLGHPDTLNQGEL	3600	21	12.23	58.237	0.492	0.469

F37L/M63F	14	39	TIINTFHQYSVKLGHPDTLNQGELKE	3600	23	13.39	58.218	0.608	0.79
F37L/M63F	14	40	TIINTFHQYSVKLGHPDTLNQGELKEL	3600	24	14.109	58.786	14.807	1.648
F37L/M63F	15	25	IINTFHQYSVK	3600	9	4.313	47.921	0.777	0.489
F37L/M63F	15	36	IINTFHQYSVKLGHPDTLNQGE	3600	19	9.571	50.374	1.622	0.653
F37L/M63F	15	37	IINTFHQYSVKLGHPDTLNQGEL	3600	20	10.154	50.771	1.153	0.128
F37L/M63F	15	39	IINTFHQYSVKLGHPDTLNQGELKE	3600	22	11.102	50.463	0.256	0.334
F37L/M63F	15	48	IINTFHQYSVKLGHPDTLNQGELKELVRKDLQNF	3600	31	16.449	53.063		0
F37L/M63F	18	22	TFHQY	3600	3	1.104	36.796	0.035	0.014
F37L/M63F	18	25	TFHQYSVK	3600	6	2.888	48.131	0.058	0.055
F37L/M63F	19	25	FHQYSVK	3600	5	2.008	40.167	0.607	0.244
F37L/M63F	19	39	FHQYSVKLGHPDTLNQGELKE	3600	18	9.137	50.759	0.467	0.445
F37L/M63F	23	29	SVKLGHP	3600	4	2.481	62.025	0.056	0.053
F37L/M63F	23	30	SVKLGHPD	3600	5	2.455	49.108	0.072	0.069
F37L/M63F	23	32	SVKLGHPDTL	3600	7	3.555	50.782	0.094	0.09
F37L/M63F	23	33	SVKLGHPDTLN	3600	8	4.258	53.22	0.134	0.128
F37L/M63F	23	34	SVKLGHPDTLNQ	3600	9	4.942	54.912	0.133	0.126
F37L/M63F	23	37	SVKLGHPDTLNQGEL	3600	12	6.638	55.313	0.152	0.122
F37L/M63F	23	38	SVKLGHPDTLNQGELK	3600	13	7.191	55.314	0.538	0.217
F37L/M63F	23	39	SVKLGHPDTLNQGELKE	3600	14	7.652	54.654	0.267	0.348
F37L/M63F	23	40	SVKLGHPDTLNQGELKEL	3600	15	8.287	55.248	0.185	0.24
F37L/M63F	26	32	LGHPDTL	3600	4	1.473	36.823	0.148	0.06
F37L/M63F	26	33	LGHPDTLN	3600	5	2.123	42.464	0.11	0.044
F37L/M63F	26	37	LGHPDTLNQGEL	3600	9	4.39	48.781	0.339	0.136
F37L/M63F	26	38	LGHPDTLNQGELK	3600	10	5.275	52.752	0.583	0.235
F37L/M63F	26	39	LGHPDTLNQGELKE	3600	11	5.675	51.593	0.159	0.151
F37L/M63F	26	40	LGHPDTLNQGELKEL	3600	12	6.485	54.039	0.151	0.144
F37L/M63F	26	48	LGHPDTLNQGELKELVRKDLQNF	3600	20	12.895	64.475	1.208	0.973
F37L/M63F	30	37	DTLNQGEL	3600	6	3.731	62.189	0.774	0.312
F37L/M63F	30	39	DTLNQGELKE	3600	8	5.241	65.512	0.608	0.245
F37L/M63F	30	40	DTLNQGELKEL	3600	9	5.877	65.302	0.369	0.149
F37L/M63F	31	37	TLNQGEL	3600	5	3.065	61.307	0.333	0.134
F37L/M63F	31	38	TLNQGELK	3600	6	3.771	62.842	0.23	0.093
F37L/M63F	31	39	TLNQGELKE	3600	7	4.156	59.378	0.323	0.13
F37L/M63F	31	40	TLNQGELKEL	3600	8	4.929	61.616	0.206	0.083

F37L/M63F	33	38	NQGELK	3600	4				
F37L/M63F	33	39	NQGELKE	3600	5				
F37L/M63F	33	40	NQGELKEL	3600	6	3.812	63.54	0.031	0.013
F37L/M63F	34	39	QGELKE	3600	4				
F37L/M63F	34	40	QGELKEL	3600	5	3.122	62.432		0
F37L/M63F	35	40	GELKEL	3600	4	2.765	69.127	0.181	0.073
F37L/M63F	36	40	ELKEL	3600	3	1.994	66.464	0.169	0.068
F37L/M63F	36	62	ELKELVRKDLQNFLKKENKNEKVIEHI	3600	25				
F37L/M63F	37	40	LKEL	3600	2	1.372	68.58	0.129	0.052
F37L/M63F	38	47	KELVRKDLQN	3600	8	4.5	56.25	1.934	0.215
F37L/M63F	38	48	KELVRKDLQNF	3600	9	5.021	55.786	0.577	0.55
F37L/M63F	40	47	LVRKDLQN	3600	6	3.968	66.134	0.325	0.31
F37L/M63F	40	49	LVRKDLQNFL	3600	8	5.198	64.974	0.278	0.112
F37L/M63F	40	60	LVRKDLQNFLKKENKNEKVIE	3600	19	10.313	54.281	1.244	0.782
F37L/M63F	40	64	LVRKDLQNFLKKENKNEKVIEHIFE	3600	23	12.311	53.525	0.283	0.338
F37L/M63F	41	49	VRKDLQNFL	3600	7				
F37L/M63F	41	50	VRKDLQNFLK	3600	8	4.771	59.636	0.214	0.231
F37L/M63F	41	52	VRKDLQNFLKKE	3600	10	5.852	58.521	0.167	0.249
F37L/M63F	41	54	VRKDLQNFLKKENK	3600	12	6.15	51.249	0.133	0.144
F37L/M63F	41	57	VRKDLQNFLKKENKNEK	3600	15	6.801	45.341	0.191	0.285
F37L/M63F	41	60	VRKDLQNFLKKENKNEKVIE	3600	18	9.015	50.086	0.449	0.361
F37L/M63F	41	61	VRKDLQNFLKKENKNEKVIEH	3600	19	8.738	45.989	0.574	0.547
F37L/M63F	44	48	DLQNF	3600	3	1.82	60.671	0.165	0.157
F37L/M63F	44	49	DLQNFL	3600	4	2.722	68.051	0.128	0.122
F37L/M63F	44	50	DLQNFLK	3600	5	3.6	72.001	0.235	0.095
F37L/M63F	44	54	DLQNFLKKENK	3600	9	5.561	61.79	1.189	0.132
F37L/M63F	44	61	DLQNFLKKENKNEKVIEH	3600	16	8.572	53.574	0.363	0.346
F37L/M63F	45	50	LQNFLK	3600	4	2.618	65.452	0.136	0.055
F37L/M63F	45	52	LQNFLKKE	3600	6	3.428	57.131	0.59	0.238
F37L/M63F	45	54	LQNFLKKENK	3600	8	4.06	50.749	0.108	0.103
F37L/M63F	45	57	LQNFLKKENKNEK	3600	11	4.704	42.766	0.099	0.119
F37L/M63F	45	61	LQNFLKKENKNEKVIEH	3600	15	6.742	44.945	0.454	0.543
F37L/M63F	47	50	NFLK	3600	2	1.492	74.602	0.029	0.012
F37L/M63F	47	51	NFLKK	3600	3	2.316	77.212	0.442	0.178
F37L/M63F	47	52	NFLKKE	3600	4	2.634	65.838	0.037	0.035

F37L/M63F	47	54	NFLKKENK	3600	6	3.153	52.542	0.667	0.268
F37L/M63F	47	57	NFLKKENKNEK	3600	9	3.962	44.027	0.39	0.157
F37L/M63F	47	61	NFLKKENKNEKVIEH	3600	13	5.768	44.366	0.253	0.241
F37L/M63F	47	63	NFLKKENKNEKVIEHIF	3600	15	6.622	44.148	0.214	0.204
F37L/M63F	48	57	FLKKENKNEK	3600	8	3.492	43.656	0.8	0.322
F37L/M63F	48	60	FLKKENKNEKVIE	3600	11	4.65	42.273	0.226	0.182
F37L/M63F	48	61	FLKKENKNEKVIEH	3600	12	4.82	40.169	0.082	0.107
F37L/M63F	48	63	FLKKENKNEKVIEHIF	3600	14	5.401	38.58	0.107	0.14
F37L/M63F	48	64	FLKKENKNEKVIEHIFE	3600	15	5.951	39.677	0.408	0.389
F37L/M63F	48	65	FLKKENKNEKVIEHIFED	3600	16	6.721	42.005	0.68	0.548
F37L/M63F	48	66	FLKKENKNEKVIEHIFEDL	3600	17	7.199	42.346	0.426	0.51
F37L/M63F	49	60	LKKENKNEKVIE	3600	10	4.13	41.297	0.485	0.391
F37L/M63F	49	61	LKKENKNEKVIEH	3600	11	4.366	39.69	0.115	0.15
F37L/M63F	49	66	LKKENKNEKVIEHIFEDL	3600	16	6.616	41.353	0.518	0.771
F37L/M63F	50	62	KKENKNEKVIEHI	3600	11	4.247	38.613	0.331	0.133
F37L/M63F	50	66	KKENKNEKVIEHIFEDL	3600	15	6.842	45.614	0.628	0.679
F37L/M63F	58	63	VIEHIF	3600	4				
F37L/M63F	59	63	IEHIF	3600	3				
F37L/M63F	64	78	EDLDTNADKQLSFEE	3600	13	6.874	52.876	0.474	0.191
F37L/M63F	64	79	EDLDTNADKQLSFEEF	3600	14	7.096	50.684	1.108	0.446
F37L/M63F	65	74	DLDTNADKQL	3600	8	4.346	54.326	0.141	0.057
F37L/M63F	65	75	DLDTNADKQLS	3600	9	5.232	58.132		0
F37L/M63F	65	76	DLDTNADKQLSF	3600	10	5.367	53.668	0.534	0.43
F37L/M63F	65	79	DLDTNADKQLSFEEF	3600	13	6.082	46.781	0.386	0.311
F37L/M63F	66	75	LDTNADKQLS	3600	8	4.672	58.403	0.344	0.138
F37L/M63F	66	78	LDTNADKQLSFEE	3600	11	5.892	53.565		0
F37L/M63F	67	78	DTNADKQLSFEE	3600	10	5.503	55.034	0.387	0.156
F37L/M63F	67	79	DTNADKQLSFEEF	3600	11	5.394	49.037	0.072	0.029
F37L/M63F	70	76	ADKQLSF	3600	5	3.338	66.759	0.16	0.065
F37L/M63F	70	79	ADKQLSFEEF	3600	8	4.081	51.016	0.481	0.194
F37L/M63F	71	76	DKQLSF	3600	4	2.622	65.541	0.685	0.276
F37L/M63F	73	79	QLSFEEF	3600	5	2.495	49.905	0.254	0.242
F37L/M63F	74	79	LSFEEF	3600	4	1.657	41.426	0.764	0.308
F37L/M63F	75	79	SFEEF	3600	3	1.176	39.202	0.115	0.11
F37L/M63F	75	80	SFEEFI	3600	4	1.728	43.204	1.056	0.425

F37L/M63F	84	114	ARLTWASHEKMHEGDEGPGHHHKPGLGEGTP	3600	26	22.494	86.515	0.195	0.186
F37L/M63F	85	88	RLTW	3600	2	1.352	67.579	0.167	0.067
F37L/M63F	85	91	RLTWASH	3600	5	2.428	48.568	0.159	0.151
F37L/M63F	85	114	RLTWASHEKMHEGDEGPGHHHKPGLGEGTP	3600	25	21.974	87.898	0.22	0.21
F37L/M63F	86	91	LTWASH	3600	4	2.222	55.558	0.184	0.074
F37L/M63F	86	97	LTWASHEKMHEG	3600	10	2.861	28.606	1.752	0.705
F37L/M63F	87	91	TWASH	3600	3	1.409	46.963	0.108	0.044
F37L/M63F	87	114	TWASHEKMHEGDEGPGHHHKPGLGEGTP	3600	23	21.395	93.02	0.227	0.245
F37L/M63F	94	114	MHEGDEGPGHHHKPGLGEGTP	3600	16	1.673	10.456	0.584	0.76
F37L/M63F	95	114	HEGDEGPGHHHKPGLGEGTP	3600	15	2.608	17.385	0.786	0.494
F37L/M63F	96	110	EGDEGPGHHHKPGLG	3600	11	1.9	17.275	0.16	0.153
F37L/M63F	96	114	EGDEGPGHHHKPGLGEGTP	3600	14	3.119	22.277	0.069	0.109
F37L/M63F	97	110	GDEGPGHHHKPGLG	3600	10	1.684	16.842	0.114	0.108
F37L/M63F	97	114	GDEGPGHHHKPGLGEGTP	3600	13	3.082	23.708	0.078	0.101
F37L/M63F	98	110	DEGPGHHHKPGLG	3600	9	1.552	17.239	0.112	0.07
F37L/M63F	98	114	DEGPGHHHKPGLGEGTP	3600	12	3.078	25.654	0.107	0.14
F37L/M63F	99	110	EGPGHHHKPGLG	3600	8	1.749	21.858	0.097	0.092
F37L/M63F	99	114	EGPGHHHKPGLGEGTP	3600	11	2.888	26.251	0.256	0.277

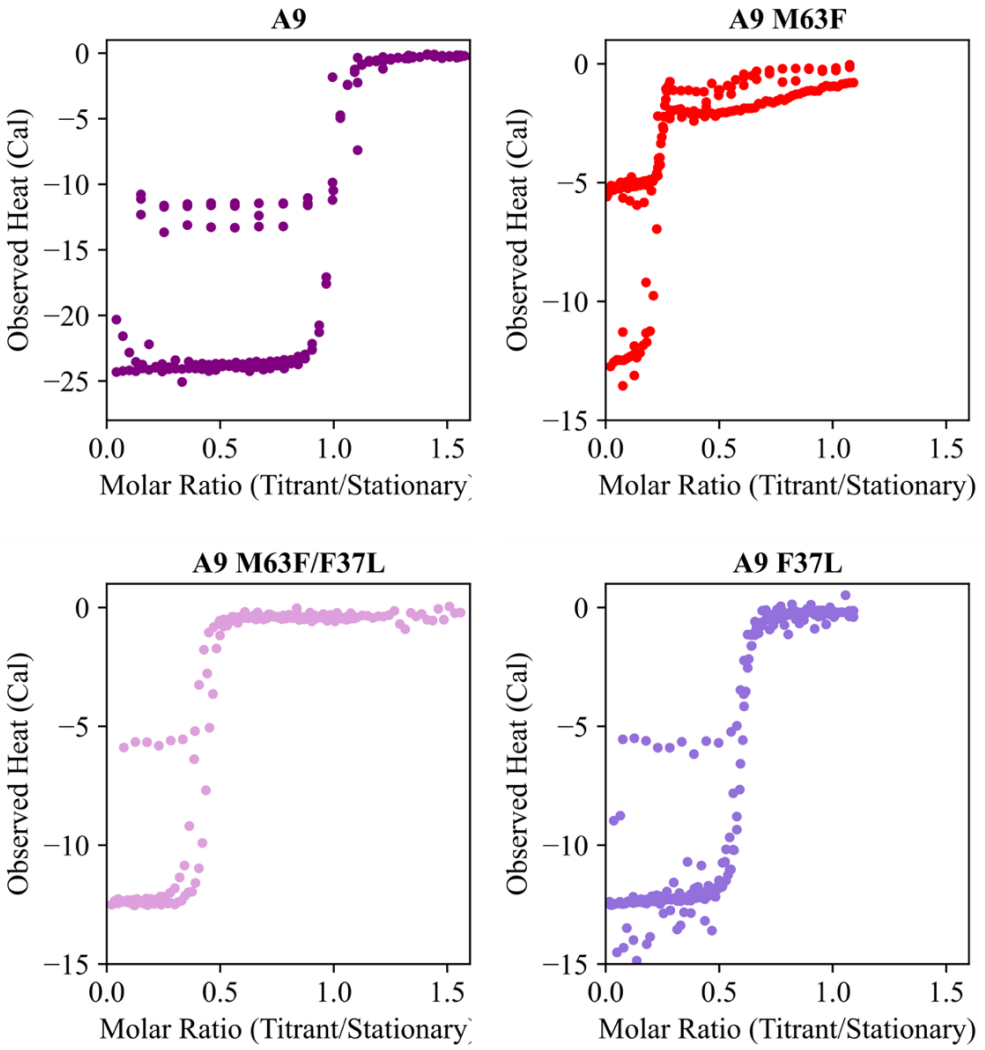
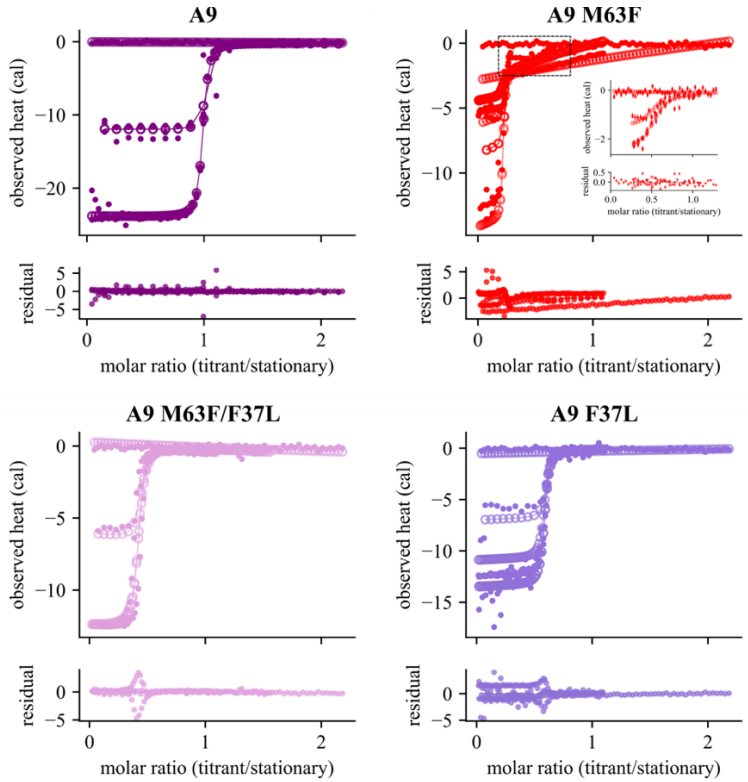


Figure AB6. Calcium binding by A9 and A9 variants by ITC. A global single-site model was fit to each dataset using the software package pytc. Filled points are experimental data, lines and unfilled points are model fit to data. Model parameters are shown in table. Experimental and fitting details are described in methods.



Protein	Kd (nM)	Fraction Competent	# sites fit
A9	31.0 ± 0.5	0.98	1
A9 M63F	26.8 ± 1.2	0.21	1
A9 M63F/F37L	79.8 ± 1.8	0.42	1
A9 F37L	35.9 ± 0.7	0.58	1
A9 M63F transition 2	741 ± 134	0.51	1
A9	31.6 ± 0.5	0.98	2
	-121.1 ± 177.3		2
A9 M63F	15.3 ± 0.8	0.21	2
	31.4 ± 12.4		2
A9 M63F/F37L	73.3 ± 1.8	0.42	2
	138.7 ± 96.8		2
A9 F37L	34.7 ± 0.8	0.58	2
	119.7 ± 145.2		2

Figure AB7. Models fit to calcium binding data for A9 and A9 variants by ITC. A global 1-site model was fit to each dataset using the software package pytc. Filled points are experimental data, lines and unfilled points are model fit to data. Inset in top right shows 1-site model fit to A9 M63F transition 2.

Table AB8. Fit parameters for A9 and A9 variant calcium binding in Figure AB5. Fit values are shown for both 1-site and 2-site fits. See methods for experimental and modeling details.

parameter	value	stdev	bot_95	top_95	protein	# sites fit
global_beta1	32228707.4	516464.707	31306594.3	33305691.6	a9	1
global_dH1	-11897.044	9.21924946	-11915.066	-11878.719	a9	1
fx_competent	0.97668994	0.00022442	0.97625225	0.97713034	a9	1
global_heat	-2856.4478	335.673135	-3515.5008	-2197.0496	a9	1
global_intercept	-0.0201415	0.01712789	-0.0537238	0.01338561	a9	1
Kd1 (nM)	31.0282379	0.4972272			a9	1
global_beta1	37368358.9	1700405.99	34160692	40844014.5	m63f	1
global_dH1	-11302.535	34.0337762	-11369.215	-11235.807	m63f	1
fx_competent	0.21321373	0.00031129	0.2126012	0.21382461	m63f	1
global_heat	66587.2227	482.976545	65636.8108	67540.0848	m63f	1
global_intercept	-2.8414565	0.01215331	-2.8652757	-2.8174116	m63f	1
Kd1 (nM)	26.7606079	1.21771198			m63f	1
global_beta1	1348888.54	244553.037	938451.453	1895774.43	m63f transition 2	1
global_dH1	-4741.0219	171.225406	-5095.9744	-4420.9606	m63f transition 2	1
fx_competent	0.51241495	0.00833148	0.49589132	0.528807	m63f transition 2	1
global_heat	-1817.3107	869.780106	-3516.2575	-87.250582	m63f transition 2	1
global_intercept	-0.0519736	0.0244351	-0.1001031	-0.0046571	m63f transition 2	1
Kd1 (nM)	741.351097	134.406703			m63f transition 2	1
global_beta1	12531395.6	280374.66	12000820.4	13099428	double	1
global_dH1	-12669.36	27.0161299	-12722.165	-12616.776	double	1
fx_competent	0.41663275	0.00035293	0.41594566	0.41733575	double	1
global_heat	-15149.272	701.067449	-16532.036	-13777.601	double	1
global_intercept	0.26408788	0.02211314	0.2207673	0.30765762	double	1
Kd1 (nM)	79.7995714	1.78541787			double	1
global_beta1	27874569.1	533262.676	26868598.1	29003287.3	f37l	1
global_dH1	-12974.424	19.5578747	-13012.294	-12935.951	f37l	1
fx_competent	0.5813502	0.00023856	0.58088217	0.58181775	f37l	1
global_heat	10114.5387	643.693665	8860.82272	11393.0581	f37l	1
global_intercept	-0.5177881	0.02108731	-0.5597927	-0.4765572	f37l	1
Kd1 (nM)	35.874994	0.68631717			f37l	1
global_beta1	31602549.7	523013.761	30562551	32693725.7	a9	2
global_dH1	-11895.491	9.47241505	-11913.764	-11877.096	a9	2

global_beta2	-8254604.9	12082635.6	-42090471	-101419.82	a9	2
global_dH2	-19412348	30659573.4	-113781321	-229820.55	a9	2
fx_competent	0.97680545	0.00023428	0.97634434	0.97726643	a9	2
global_heat	-3793.4025	486.23839	-4749.7951	-2850.7882	a9	2
global_intercept	-0.0067445	0.01759122	-0.0409632	0.02802263	a9	2
Kd1 (nM)	31.6430165	0.52368348			a9	2
Kd2 (nM)	-121.1445	177.324645			a9	2
global_beta1	65520718	3620225.94	58846801.9	72978388.1	m63f	2
global_dH1	-11370.8	31.2793439	-11433.701	-11310.969	m63f	2
global_beta2	31869685	12604834.1	13899746.2	59008843.7	m63f	2
global_dH2	-286504293	118091251	-558099533	-130652747	m63f	2
fx_competent	0.2111714	0.00032247	0.21054289	0.21180318	m63f	2
global_heat	79653.0862	531.156243	78629.5717	80722.5173	m63f	2
global_intercept	-2.7800567	0.01200469	-2.804679	-2.7566439	m63f	2
Kd1 (nM)	15.262348	0.84329278			m63f	2
Kd2 (nM)	31.3777811	12.4102804			m63f	2
global_beta1	13645115.8	335003.031	13006693.9	14326242.1	double	2
global_dH1	-12661.685	26.7346176	-12713.987	-12608.944	double	2
global_beta2	7209599.17	5031803.73	1338493.06	19518143.7	double	2
global_dH2	-40111078	32288816.9	-126691382	-8613791.3	double	2
fx_competent	0.41563233	0.00035602	0.41491589	0.41632745	double	2
global_heat	-11873.374	800.026603	-13405.073	-10322.201	double	2
global_intercept	0.25660546	0.02245748	0.21223876	0.2994366	double	2
Kd1 (nM)	73.2862961	1.79926148			double	2
Kd2 (nM)	138.703966	96.8058164			double	2
global_beta1	28786224.4	640662.075	27538058.1	30041288.3	f37l	2
global_dH1	-12973.954	20.4551602	-13013.688	-12934.026	f37l	2
global_beta2	8351379.7	10128121.7	335357.312	36483552.4	f37l	2
global_dH2	-37938328	48225191.8	-171767783	-1730218.3	f37l	2
competent	0.58122387	0.00024219	0.58074397	0.5816949	f37l	2
global_heat	11443.4824	800.852515	9886.11337	13025.9061	f37l	2
global_intercept	-0.529113	0.02211435	-0.5733811	-0.4860587	f37l	2
Kd1 (nM)	34.7388385	0.7731426			f37l	2
Kd2 (nM)	119.740694	145.215326			f37l	2

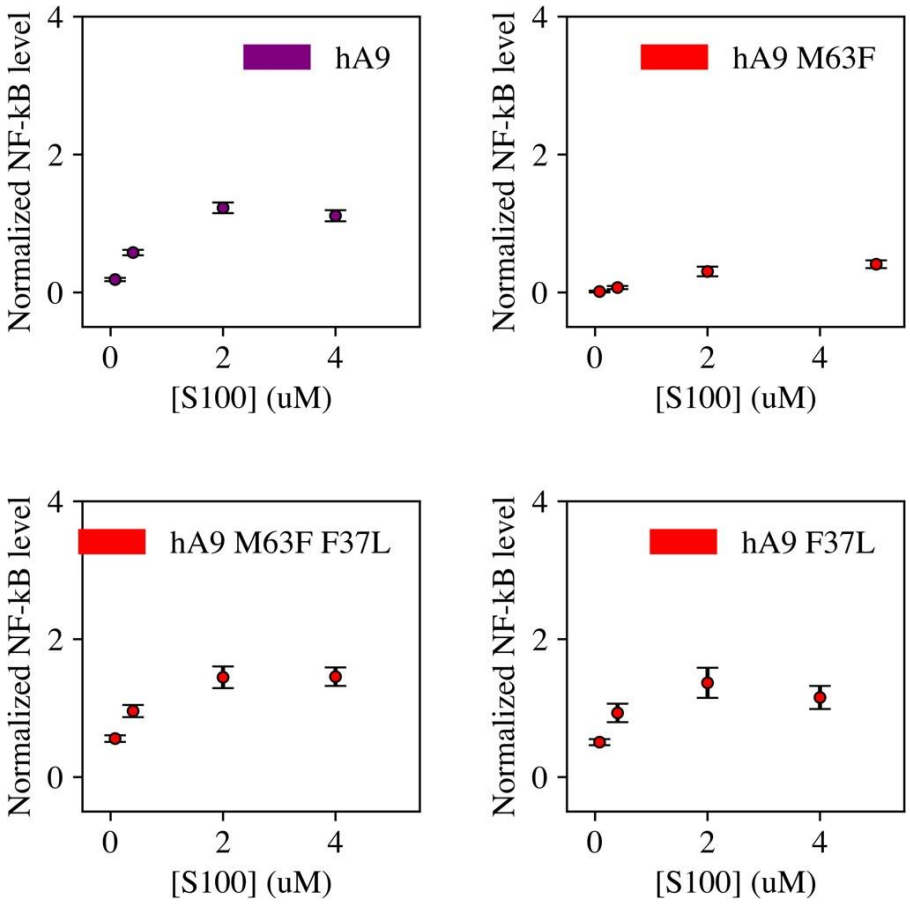


Figure AB9. A9 mutant cycle TLR4 activation dose curves. Dose curves corresponding to data in Figure 3.5f.

APPENDIX C

SUPPLEMENTAL MATERIAL FOR CHAPTER IV

Supplemental Figures

This section includes supplemental figures referenced in Chapter IV.

Table AC1. Activation of amniote TLR4s by modern and ancestrally reconstructed S100 proteins. Compiled data from multiple studies. Protein species indicated; Anc indicates maximum likelihood reconstructed ancestral protein (see referenced studies and/or methods section for details). All TLR4/MD2/CD14 components are from same species. NF-kB values are mean ± standard error for > 3 biological replicates. Note that differences in values between studies are partially due to changes in protein preparation and treatment. See methods for details of cell culture assay. Values in scaled average activity column put each set of TLR4 species measurements on a scale of 0-1 and correspond to heatmap color values in figure 4.1.

S100 Protein	TLR4/MD2/ CD14 species	NF-kB activity (background-subtracted, normalized to LPS)	Scaled Average NF-kB activity	Reference
Human A9	Human	2.1 ± 0.5 1.6 ± 0.2 **1.0 ± 0.3*	1 ± 0.4	Loes et al. 2018 Harman et al 2020 *this study
Mouse A9		1.0 ± 0.1	0.6 ± 0.1	Loes et al. 2018
Elephant A9		**0.4 ± 0.1*	*0.4* ± 0.1	**this study*
Eutherian AncA9		**0.6 ± 0.1*	*0.6* ± 0.1	**this study*
Opossum A9		0.6 ± 0.1 0.01 ± 0.003	0.2 ± 0.1	Loes et al. 2018 Harman et al. 2020
Therian AncA9		0.03 ± 0.01 **0.2 ± 0.01*	0.1 ± 0.01	Harman et al. 2020 **this study*
Chicken MRP126		0.1 ± 0.01	0.1 ± 0.01	Loes et al. 2018
Amniote AncCG		0.0 ± 0.01	0 ± 0.01	Harman et al. 2020
Human A9	Mouse	2 ± 0.6	1 ± 0.3	Loes et al. 2018
Mouse A9		1.4 ± 0.5	0.7 ± 0.3	Loes et al. 2018
Opossum A9		1.4 ± 0.4	0.7 ± 0.2	Loes et al. 2018

Chicken MRP126		0.7 ± 0.1	0.4 ± 0.1	Loes et al. 2018
Human A9	Opossum	1.8 ± 0.3 0.6 ± 0.1 0.8 ± 0.03	0.8 ± 0.2	Loes et al. 2018 Harman et al. 2020 **this study*
Mouse A9		1.4 ± 0.3	1 ± 0.2	Loes et al. 2018
Eutherian AncA9		0.4 ± 0.04	*0.3 ± 0.03*	**this study*
Opossum A9		1.3 ± 0.3 0.3 ± 0.1	0.6 ± 0.2	Loes et al. 2018 Harman et al. 2020
Therian AncA9		0.4 ± 0.1 0.3 ± 0.04	0.3 ± 0.1	Harman et al. 2020 **this study*
Chicken MRP126		0.3 ± 0.1	0.2 ± 0.1	Loes et al. 2018
Amniote AncCG		0.1 ± 0.01	0.1 ± 0.01	Harman et al. 2020
Human A9	Chicken	4.5 ± 2.0	1 ± 0.4	Loes et al. 2018
Mouse A9		3.3 ± 1.0	0.7 ± 0.2	Loes et al. 2018
Opossum A9		2.0 ± 0.7	0.4 ± 0.2	Loes et al. 2018
Chicken MRP126		0.6 ± 0.1	0.1 ± 0.02	Loes et al. 2018

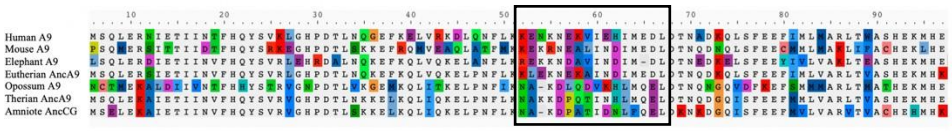


Figure AC2. Alignment of S100A9s, highlighting highly mutated region.

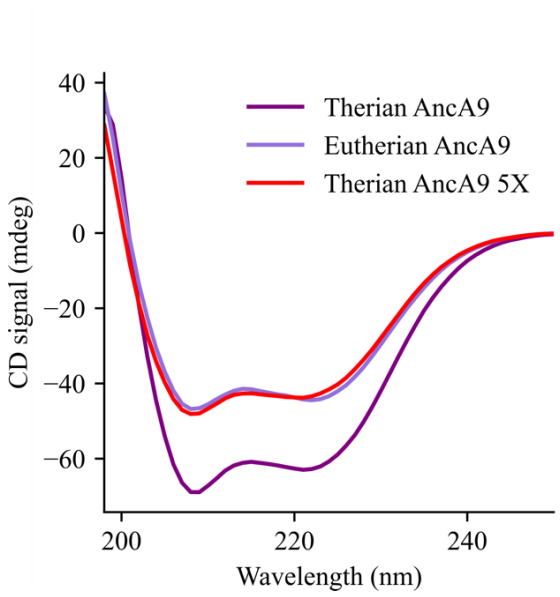


Figure AC3. CD spectroscopy of therian ancA9 and 5X mutant therian ancA9. Dark purple - therian ancA9, light purple - eutherian ancA9, red – "5X therian ancA9"; therian A9 with mutations N51K, D55N, P56E, A57K, Q64E (human A9 numbering). Average of 3 scans.

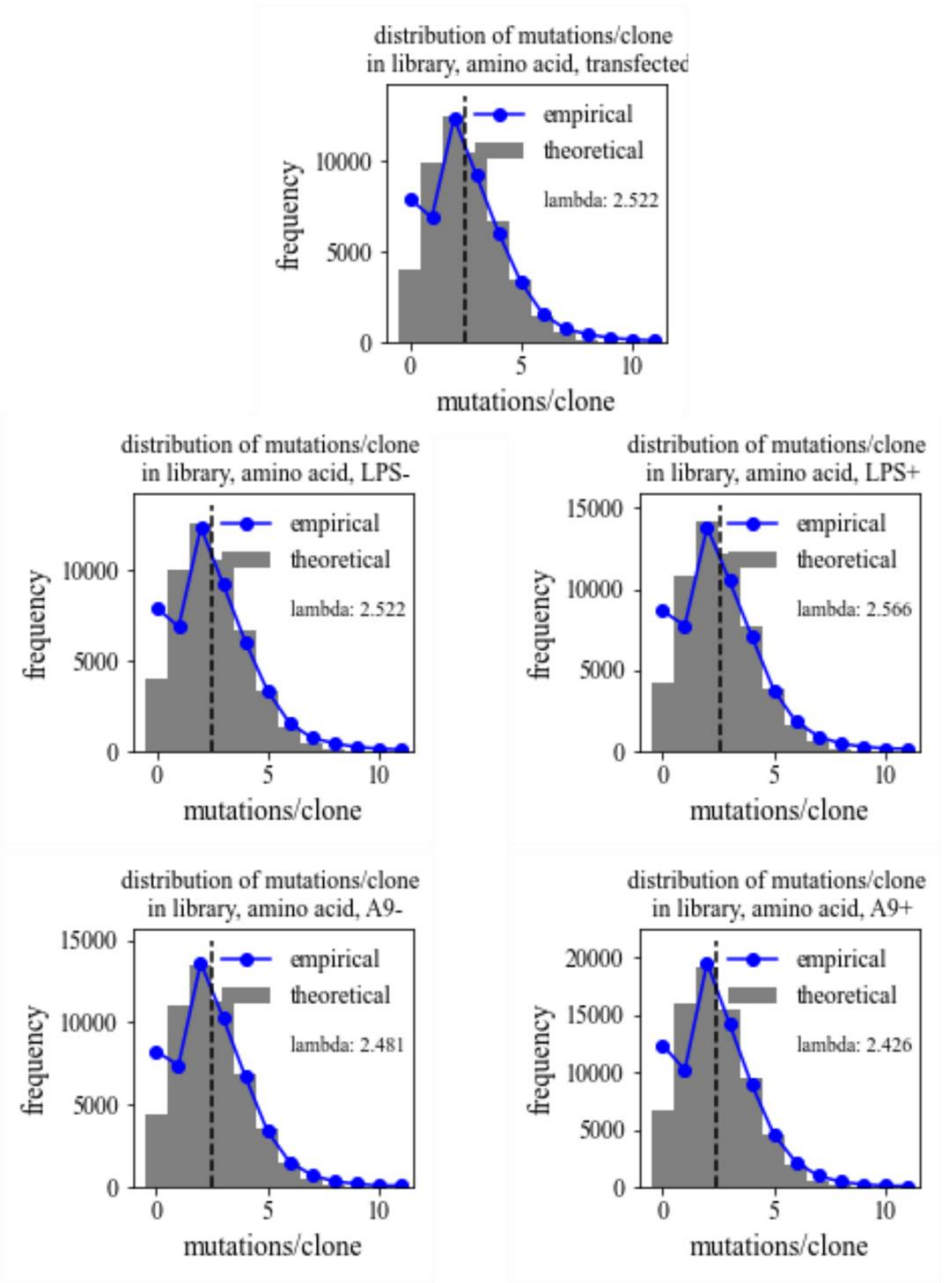


Figure AC4. Mutation rates for MD2 mutant library treatment conditions. Average number of mutations and distribution of number of mutations from high-throughput sequencing data for library (top), LPS GFP^- (top left), LPS GFP^+ (top right), A9 GFP^- (top left), A9 GFP^+ (top right).

Table AC5. Top single MD2 mutants that alter LPS or A9 activation of TLR4. Enrichment cutoff for LPS and A9 defined as log_2enrichment > 0, log_2de-enrichment < -0.7 (for example. de-enrichment in A9 when calculating enrichment for LPS), and a log2(de-enriched/enriched) score < -1. See Chapter III results and methods for further details.

log2(LPS enrichment)	log2(A9 enrichment)	Mutant	Enrichment
0.637	-1.282	I117L	LPS over A9
1.137	-0.743	Y79N	LPS over A9
0.485	-1.121	K58E	LPS over A9
0.359	-1.198	V134I	LPS over A9
0.227	-1.060	G97R	LPS over A9
0.359	-0.913	D99V	LPS over A9
0.096	-1.093	N86K	LPS over A9
0.374	-0.751	K132E	LPS over A9
0.167	-0.880	I66L	LPS over A9
0.158	-0.847	A30T	LPS over A9
-0.756	0.335	Y42F	A9 over LPS
-0.854	0.278	I66N	A9 over LPS
-0.811	0.364	F64I	A9 over LPS
-0.740	0.525	N26K	A9 over LPS
-0.963	0.327	K132I	A9 over LPS
-1.033	0.302	V134D	A9 over LPS
-0.785	0.690	D100G	A9 over LPS
-0.811	0.731	K72M	A9 over LPS
-1.313	0.272	L61F	A9 over LPS
-0.778	0.954	I63T	A9 over LPS
-0.749	1.217	K91E	A9 over LPS
-1.226	1.146	Y131C	A9 over LPS

REFERENCES CITED

(1) Srivastava, M.; Simakov, O.; Chapman, J.; Fahey, B.; Gauthier, M. E. A.; Mitros, T.; Richards, G. S.; Conaco, C.; Dacre, M.; Hellsten, U.; Larroux, C.; Putnam, N. H.; Stanke, M.; Adamska, M.; Darling, A.; Degnan, S. M.; Oakley, T. H.; Plachetzki, D. C.; Zhai, Y.; Adamski, M.; Calcino, A.; Cummins, S. F.; Goodstein, D. M.; Harris, C.; Jackson, D. J.; Leys, S. P.; Shu, S.; Woodcroft, B. J.; Vervoort, M.; Kosik, K. S.; Manning, G.; Degnan, B. M.; Rokhsar, D. S. The Amphimedon Queenslandica Genome and the Evolution of Animal Complexity. *Nature* **2010**, *466* (7307), 720–726. https://doi.org/10.1038/nature09201.

(2) Martin, B. L.; Kimelman, D. Wnt Signaling and the Evolution of Embryonic Posterior Development. *Curr. Biol.* **2009**, *19* (5), R215–R219. https://doi.org/10.1016/j.cub.2009.01.052.

(3) Nichols, S. A.; Dirks, W.; Pearse, J. S.; King, N. Early Evolution of Animal Cell Signaling and Adhesion Genes. *Proc. Natl. Acad. Sci.* **2006**, *103* (33), 12451–12456. https://doi.org/10.1073/pnas.0604065103.

(4) Patterson, T.; Thomas, L.; Wilcox, C.; Ovaskainen, O.; Matthiopoulos, J. State–Space Models of Individual Animal Movement. *Trends Ecol. Evol.* **2008**, *23* (2), 87–94. https://doi.org/10.1016/j.tree.2007.10.009.

(5) Livingstone, D. R. Origins and Evolution of Pathways of Anaerobic Metabolism in the Animal Kingdom. *Am. Zool.* **1991**, *31* (3), 522–534. https://doi.org/10.1093/icb/31.3.522.

(6) Muller, M.; Mentel, M.; van Hellemond, J. J.; Henze, K.; Woehle, C.; Gould, S. B.; Yu, R.-Y.; van der Giezen, M.; Tielens, A. G. M.; Martin, W. F. Biochemistry and Evolution of Anaerobic Energy Metabolism in Eukaryotes. *Microbiol. Mol. Biol. Rev.* **2012**, *76* (2), 444–495. https://doi.org/10.1128/MMBR.05024-11.

(7) Buchmann, K. Evolution of Innate Immunity: Clues from Invertebrates via Fish to Mammals. *Front. Immunol.* **2014**, *5*. https://doi.org/10.3389/fimmu.2014.00459.

(8) Flajnik, M.; Dupasquier, L. Evolution of Innate and Adaptive Immunity: Can We Draw a Line? *Trends Immunol.* **2004**, *25* (12), 640–644. https://doi.org/10.1016/j.it.2004.10.001.

(9) El-Sayed, M. A.; Logunov, S. On the Molecular Origin of the Protein Catalysis of the Primary Process in Bacteriorhodopsin Photosynthesis: Retinal Photoisomerization. *Pure Appl. Chem.* **1997**, *69* (4), 749–754. https://doi.org/10.1351/pac199769040749.

(10) Holliday, G. L.; Almonacid, D. E.; Mitchell, J. B. O.; Thornton, J. M. The Chemistry of Protein Catalysis. *J. Mol. Biol.* **2007**, *372* (5), 1261–1277. https://doi.org/10.1016/j.jmb.2007.07.034.

(11) Lee, M. C. S.; Miller, E. A.; Goldberg, J.; Orci, L.; Schekman, R. BI-DIRECTIONAL PROTEIN TRANSPORT BETWEEN THE ER AND GOLGI. *Annu. Rev. Cell Dev. Biol.* **2004**, *20* (1), 87–123. https://doi.org/10.1146/annurev.cellbio.20.010403.105307.

(12) Rothman, J. E. Mechanisms of Intracellular Protein Transport. *Nature* **1994**, *372* (6501), 55–63. https://doi.org/10.1038/372055a0.

(13) Jones, S.; Thornton, J. M. Review Principles of Protein-Protein Interactions. *Proc Natl Acad Sci USA* **1996**, 8.

(14) Sachs, K. Causal Protein-Signaling Networks Derived from Multiparameter Single-Cell Data. *Science* **2005**, *308* (5721), 523–529. https://doi.org/10.1126/science.1105809.

(15) Baier, F.; Copp, J. N.; Tokuriki, N. Evolution of Enzyme Superfamilies: Comprehensive Exploration of Sequence–Function Relationships. *Biochemistry* **2016**, *55* (46), 6375–6388. https://doi.org/10.1021/acs.biochem.6b00723.

(16) Fowler, D. M.; Araya, C. L.; Fleishman, S. J.; Kellogg, E. H.; Stephany, J. J.; Baker, D.; Fields, S. High-Resolution Mapping of Protein Sequence-Function Relationships. *Nat. Methods* **2010**, *7* (9), 741–746. https://doi.org/10.1038/nmeth.1492.

(17) Sangar, V.; Blankenberg, D. J.; Altman, N.; Lesk, A. M. Quantitative Sequence-Function Relationships in Proteins Based on Gene Ontology. *BMC Bioinformatics* **2007**, *8* (1), 294. https://doi.org/10.1186/1471-2105-8-294.

(18) Allan Drummond, D.; Wilke, C. O. The Evolutionary Consequences of Erroneous Protein Synthesis. *Nat. Rev. Genet.* **2009**, *10* (10), 715–724. https://doi.org/10.1038/nrg2662.

(19) Drummond, D. A.; Wilke, C. O. Mistranslation-Induced Protein Misfolding as a Dominant Constraint on Coding-Sequence Evolution. *Cell* **2008**, *134* (2), 341–352. https://doi.org/10.1016/j.cell.2008.05.042.

(20) Bershtein, S.; Serohijos, A. W.; Shakhnovich, E. I. Bridging the Physical Scales in Evolutionary Biology: From Protein Sequence Space to Fitness of Organisms and Populations. *Curr. Opin. Struct. Biol.* **2017**, *42*, 31–40. https://doi.org/10.1016/j.sbi.2016.10.013.

(21) Kan, Y. Wa. Polymorphism of DNA Sequence Adjacent to Human F-Globin Structural Gene: Relationship to Sickle Mutation. *Proc Natl Acad Sci USA* **1978**, 5.

(22) Arpaia, E.; Dumbrille-Ross, A.; Maler, T.; Neote, K.; Tropak, M.; Troxel, C.; Stirling, J. L.; Pitts, J. S.; Bapat, B.; Lamhonwah, A. M.; Mahuran, D. J.; Schuster, S. M.; Clarke, J. T. R.; Lowden, J. A.; Gravel, R. A. Identification of an Altered Splice Site in Ashkenazi Tay-Sachs Disease. *Nature* **1988**, *333* (6168), 85–86. https://doi.org/10.1038/333085a0.

(23) Darwin, C.; Kebler, L. *On the Origin of Species by Means of Natural Selection, or, the Preservation of Favoured Races in the Struggle for Life*; J. Murray: London, 1859.

(24) Simpson, G. G. Organisms and Molecules in Evolution. *Science* **1964**, *146* (3651), 1535–1538.

(25) Queiroz, A. D. Contingent Predictability in Evolution: Key Traits and Diversification. *Syst. Biol.* **2002**, *51*, 13.

(26) Stearns, S. C. Life History Evolution: Successes, Limitations, and Prospects. *Naturwissenschaften* **2000**, *87* (11), 476–486. https://doi.org/10.1007/s001140050763.

(27) Asplen, M. K.; Bruns, E.; David, A. S.; Denison, R. F.; Epstein, B.; Kaiser, M. C.; Kaser, J. M.; Lacroix, C.; Mohl, E. K.; Quiram, G.; Prescott, K.; Stanton-Geddes, J.; Vincent, J. B.; Wragg, P. D.; May, G. DO TRADE-OFFS HAVE EXPLANATORY POWER FOR THE EVOLUTION OF ORGANISMAL INTERACTIONS? *Evolution* **2012**, *66* (5), 1297–1307. https://doi.org/10.1111/j.1558-5646.2011.01573.x.

(28) Piersma, T.; Drent, J. Phenotypic Flexibility and the Evolution of Organismal Design. *Trends Ecol. Evol.* **2003**, *18* (5), 228–233. https://doi.org/10.1016/S0169-5347(03)00036-3.

(29) Harms, M. J.; Thornton, J. W. Evolutionary Biochemistry: Revealing the Historical and Physical Causes of Protein Properties. *Nat. Rev. Genet.* **2013**, *14* (8), 559–571. https://doi.org/10.1038/nrg3540.

(30) .Newton, M. S.; Arcus, V. L.; Gerth, M. L.; Patrick, W. M. Enzyme Evolution: Innovation Is Easy, Optimization Is Complicated. *Curr. Opin. Struct. Biol.* **2018**, *48*, 110–116. https://doi.org/10.1016/j.sbi.2017.11.007.

(31) Alhindi, T.; Zhang, Z.; Ruelens, P.; Coenen, H.; Degroote, H.; Iraci, N.; Geuten, K. Protein Interaction Evolution from Promiscuity to Specificity with Reduced Flexibility in an Increasingly Complex Network. *Sci. Rep.* **2017**, *7*, 44948. https://doi.org/10.1038/srep44948.

(32) Dean, A. M.; Thornton, J. W. Mechanistic Approaches to the Study of Evolution: The Functional Synthesis. *Nat. Rev. Genet.* **2007**, *8* (9), 675–688. https://doi.org/10.1038/nrg2160.

(33) Harms, M. J.; Thornton, J. W. Analyzing Protein Structure and Function Using Ancestral Gene Reconstruction. *Curr. Opin. Struct. Biol.* **2010**, *20* (3), 360–366. https://doi.org/10.1016/j.sbi.2010.03.005.

(34) Thornton, J. W. Resurrecting Ancient Genes: Experimental Analysis of Extinct Molecules. *Nat. Rev. Genet.* **2004**, *5* (5), 366–375. https://doi.org/10.1038/nrg1324.

(35) Zuckerkandl, E.; Pauling, L. Molecules as Documents of Evolutionary History. *J. Theor. Biol.* **1965**, *8* (2), 357–366. https://doi.org/10.1016/0022-5193(65)90083-4.

(36) Yang, Z. PAML 4: Phylogenetic Analysis by Maximum Likelihood. *Mol. Biol. Evol.* **2007**, *24* (8), 1586–1591. https://doi.org/10.1093/molbev/msm088.

(37) Yang, Z.; Rannala, B. Molecular Phylogenetics: Principles and Practice. *Nat. Rev. Genet.* **2012**, *13* (5), 303–314. https://doi.org/10.1038/nrg3186.

(38) Ziheng Yang; Sudhir Kumar; Masatoshi Nei. A New Method of Inference of Ancestral Nucleotide and Amino Acid Sequences. *Genetics* **1995**, No. 141, 1641–1650.

(39) Gumulya, Y.; Gillam, E. M. J. Exploring the Past and the Future of Protein Evolution with Ancestral Sequence Reconstruction: The 'retro Approach to Protein Engineering. *Biochem. J.* **2017**, *474* (1), 1–19. https://doi.org/10.1042/BCJ20160507.

(40) Merkl, R.; Sterner, R. Ancestral Protein Reconstruction: Techniques and Applications. *Biol. Chem.* **2016**, *397* (1), 1–21. https://doi.org/10.1515/hsz-2015-0158.

(41) Chang, B. S. W.; Kazmi, M. A.; Sakmar, T. P. Synthetic Gene Technology: Applications to Ancestral Gene Reconstruction and Structure-Function Studies of Receptors. In *Methods in Enzymology*; Elsevier, 2002; Vol. 343, pp 274–294. https://doi.org/10.1016/S0076-6879(02)43142-4.

(42) Thornton, J. W. Resurrecting the Ancestral Steroid Receptor: Ancient Origin of Estrogen Signaling. *Science* **2003**, *301* (5640), 1714–1717. https://doi.org/10.1126/science.1086185.

(43) Gaucher, E. A.; Thomson, J. M.; Burgan, M. F.; Benner, S. A. Inferring the Palaeoenvironment of Ancient Bacteria on the Basis of Resurrected Proteins. *Nature* **2003**, *425* (6955), 285–288. https://doi.org/10.1038/nature01977.

(44) Harms, M. J.; Eick, G. N.; Goswami, D.; Colucci, J. K.; Griffin, P. R.; Ortlund, E. A.; Thornton, J. W. Biophysical Mechanisms for Large-Effect Mutations in the Evolution of Steroid Hormone Receptors. *Proc. Natl. Acad. Sci.* **2013**, *110* (28), 11475–11480. https://doi.org/10.1073/pnas.1303930110.

(45) Natarajan, C.; Hoffmann, F. G.; Weber, R. E.; Fago, A.; Witt, C. C.; Storz, J. F. Predictable Convergence in Hemoglobin Function Has Unpredictable Molecular Underpinnings. *Science* **2016**, *354* (6310), 336–339. https://doi.org/10.1126/science.aaf9070.

(46) Phillips, P. C. Epistasis — the Essential Role of Gene Interactions in the Structure and Evolution of Genetic Systems. *Nat. Rev. Genet.* **2008**, *9* (11), 855–867. https://doi.org/10.1038/nrg2452.

(47) Sailer, Z. R.; Harms, M. J. High-Order Epistasis Shapes Evolutionary Trajectories. *PLOS Comput. Biol.* **2017**, *13* (5), e1005541. https://doi.org/10.1371/journal.pcbi.1005541.

(48) Sailer, Z. R.; Harms, M. J. Molecular Ensembles Make Evolution Unpredictable. *Proc. Natl. Acad. Sci.* **2017**, 201711927. https://doi.org/10.1073/pnas.1711927114.

(49) Sailer, Z. R.; Harms, M. J. Uninterpretable Interactions: Epistasis as Uncertainty. **2018**. https://doi.org/10.1101/378489.

(50) Bridgham, J. T.; Ortlund, E. A.; Thornton, J. W. An Epistatic Ratchet Constrains the Direction of Glucocorticoid Receptor Evolution. *Nature* **2009**, *461* (7263), 515–519. https://doi.org/10.1038/nature08249.

(51) Storz, J. F. Compensatory Mutations and Epistasis for Protein Function. *Curr. Opin. Struct. Biol.* **2018**, *50*, 18–25. https://doi.org/10.1016/j.sbi.2017.10.009.

(52) Bloom, J. D.; Gong, L. I.; Baltimore, D. Permissive Secondary Mutations Enable the Evolution of Influenza Oseltamivir Resistance. *Science* **2010**, *328* (5983), 1272–1275. https://doi.org/10.1126/science.1187816.

(53) Harms, M. J.; Thornton, J. W. Historical Contingency and Its Biophysical Basis in Glucocorticoid Receptor Evolution. *Nature* **2014**, *512* (7513), 203–207. https://doi.org/10.1038/nature13410.

(54) Starr, T. N.; Flynn, J. M.; Mishra, P.; Bolon, D. N. A.; Thornton, J. W. Pervasive Contingency and Entrenchment in a Billion Years of Hsp90 Evolution. **2018**. https://doi.org/10.1101/189803.

(55) Zaretsky, J. Z.; Wreschner, D. H. Protein Multifunctionality: Principles and Mechanisms. *Transl. Oncogenomics* **2008**, *2008* (1), 99–136.

(56) Kirschner, K.; Bisswanger, H. Multifunctional Proteins. *Annu. Rev. Biochem.* **1976**, *45* (1), 143–166.

(57) Pavličev, M.; Cheverud, J. M. Constraints Evolve: Context Dependency of Gene Effects Allows Evolution of Pleiotropy. *Annu. Rev. Ecol. Evol. Syst.* **2015**, *46* (1), 413–434. https://doi.org/10.1146/annurev-ecolsys-120213-091721.

(58) Miller, S. P.; Lunzer, M.; Dean, A. M. Direct Demonstration of an Adaptive Constraint. *Science* **2006**, *314* (5798), 458–461. https://doi.org/10.1126/science.1133479.

(59) Tokuriki, N.; Stricher, F.; Serrano, L.; Tawfik, D. S. How Protein Stability and New Functions Trade Off. *PLoS Comput. Biol.* **2008**, *4* (2), e1000002.

(60) Beadle, B. M.; Shoichet, B. K. Structural Bases of Stability–Function Tradeoffs in Enzymes. *J. Mol. Biol.* **2002**, *321* (2), 285–296. https://doi.org/10.1016/S0022-2836(02)00599-5.

(61) Ellison, P. T. Evolutionary Tradeoffs. *Evol. Med. Public Health* **2014**, *2014* (1), 93–93. https://doi.org/10.1093/emph/eou015.

(62) Klesmith, J. R.; Bacik, J.-P.; Wrenbeck, E. E.; Michalczyk, R.; Whitehead, T. A. Trade-Offs between Enzyme Fitness and Solubility Illuminated by Deep Mutational Scanning. *Proc. Natl. Acad. Sci.* **2017**, *114* (9), 2265–2270. https://doi.org/10.1073/pnas.1614437114.

(63) Nagatani, R. A.; Gonzalez, A.; Shoichet, B. K.; Brinen, L. S.; Babbitt, P. C. Stability for Function Trade-Offs in the Enolase Superfamily "Catalytic Module" †,‡. *Biochemistry* **2007**, *46* (23), 6688–6695. https://doi.org/10.1021/bi700507d.

(64) Studer, R. A.; Christin, P.-A.; Williams, M. A.; Orengo, C. A. Stability-Activity Tradeoffs Constrain the Adaptive Evolution of RubisCO. *Proc. Natl. Acad. Sci.* **2014**, *111* (6), 2223–2228.

(65) Wang, X.; Minasov, G.; Shoichet, B. K. Evolution of an Antibiotic Resistance Enzyme Constrained by Stability and Activity Trade-Offs. *J. Mol. Biol.* **2002**, *320* (1), 85–95. https://doi.org/10.1016/S0022-2836(02)00400-X.

(66) Tokuriki, N.; Jackson, C. J.; Afriat-Jurnou, L.; Wyganowski, K. T.; Tang, R.; Tawfik, D. S. Diminishing Returns and Tradeoffs Constrain the Laboratory Optimization of an Enzyme. *Nat. Commun.* **2012**, *3* (1), 1257. https://doi.org/10.1038/ncomms2246.

(67) Charlesworth, B.; Morgan, M. T.; Charlesworth, D. The Effect of Deleterious Mutations on Neutral Molecular Variation. 15.

(68) Sniegowski, P. D.; Gerrish, P. J.; Johnson, T.; Shaver, A. The Evolution of Mutation Rates: Separating Causes from Consequences. 10.

(69) Charlesworth, B.; Charlesworth, D. Some Evolutionary Consequences of Deleterious Mutations. In *Mutation and Evolution*; Woodruff, R. C., Thompson, J. N., Eds.; Contemporary Issues in Genetics and Evolution; Springer Netherlands: Dordrecht, 1998; Vol. 7, pp 3–19. https://doi.org/10.1007/978-94-011-5210-5_1.

(70) Cohen, S. Comparative Biochemistry and Drug Design for Infectious Disease. *Science* **1979**, *205* (4410), 964–971. https://doi.org/10.1126/science.382357.

(71) Goginashvili, A.; Zhang, Z.; Erbs, E.; Spiegelhalter, C.; Kessler, P.; Mihlan, M.; Pasquier, A.; Krupina, K.; Schieber, N.; Cinque, L.; Morvan, J.; Sumara, I.; Schwab, Y.; Settembre, C.; Ricci, R. Using Ancient Protein Kinases to Unravel a Modern Cancer Drug's Mechanism. *Science* **2015**, *347* (6224), 878–882. https://doi.org/10.1126/science.aaa2628.

(72) Schwartz, P. A.; Murray, B. W. Protein Kinase Biochemistry and Drug Discovery. *Bioorganic Chem.* **2011**, *39* (5–6), 192–210. https://doi.org/10.1016/j.bioorg.2011.07.004.

(73) Chevalier, A.; Silva, D.-A.; Rocklin, G. J.; Hicks, D. R.; Vergara, R.; Murapa, P.; Bernard, S. M.; Zhang, L.; Lam, K.-H.; Yao, G.; Bahl, C. D.; Miyashita, S.-I.; Goreshnik, I.; Fuller, J. T.; Koday, M. T.; Jenkins, C. M.; Colvin, T.; Carter, L.; Bohn, A.; Bryan, C. M.; Fernández-Velasco, D. A.; Stewart, L.; Dong, M.; Huang, X.; Jin, R.; Wilson, I. A.; Fuller, D. H.; Baker, D. Massively Parallel de Novo Protein Design for Targeted Therapeutics. *Nature* **2017**, *550* (7674), 74–79. https://doi.org/10.1038/nature23912.

(74) Baker, D. What Has de Novo Protein Design Taught Us about Protein Folding and Biophysics? *Protein Sci.* **2019**, *28* (4), 678–683. https://doi.org/10.1002/pro.3588.

(75) Butterfield, G. L.; Lajoie, M. J.; Gustafson, H. H.; Sellers, D. L.; Nattermann, U.; Ellis, D.; Bale, J. B.; Ke, S.; Lenz, G. H.; Yehdego, A.; Ravichandran, R.; Pun, S. H.; King, N. P.; Baker, D. Evolution of a Designed Protein Assembly Encapsulating Its Own RNA Genome. *Nature* **2017**, *552* (7685), 415–420. https://doi.org/10.1038/nature25157.

(76) Huang, P.-S.; Boyken, S. E.; Baker, D. The Coming of Age of de Novo Protein Design. *Nature* **2016**, *537* (7620), 320–327. https://doi.org/10.1038/nature19946.

(77) Boyken, S. E.; Chen, Z.; Groves, B.; Langan, R. A.; Oberdorfer, G.; Ford, A.; Gilmore, J. M.; Xu, C.; DiMaio, F.; Pereira, J. H.; Sankaran, B.; Seelig, G.; Zwart, P. H.; Baker, D. De Novo Design of Protein Homo-Oligomers with Modular Hydrogen-Bond Network–Mediated Specificity. 9.

(78) Jäckel, C.; Kast, P.; Hilvert, D. Protein Design by Directed Evolution. *Annu. Rev. Biophys.* **2008**, *37* (1), 153–173. https://doi.org/10.1146/annurev.biophys.37.032807.125832.

(79) Hessian, P.; Edgeworth, J.; Hogg, N. MRP-8 and MRP-14, Two Abundant Ca2+-Binding Proteins of Neutrophils and Monocytes. *J. Leukoc. Biol.* **1993**, *53*, 197–204.

(80) Edgeworth, J.; Gormanl, M.; Bennett, R.; Freemontl, P.; Hogg, N. Identification of P8,14 as a Highly Abundant Heterodimeric Calcium Binding Protein Complex of Myeloid Cells. 8.

(81) Vogl, T.; Eisenblätter, M.; Völler, T.; Zenker, S.; Hermann, S.; van Lent, P.; Faust, A.; Geyer, C.; Petersen, B.; Roebrock, K.; Schäfers, M.; Bremer, C.; Roth, J. Alarmin S100A8/S100A9 as a Biomarker for Molecular Imaging of Local Inflammatory Activity. *Nat. Commun.* **2014**, *5* (1), 4593. https://doi.org/10.1038/ncomms5593.

(82) Hara, A.; Sakamoto, N.; Ishimatsu, Y.; Kakugawa, T.; Nakashima, S.; Hara, S.; Adachi, M.; Fujita, H.; Mukae, H.; Kohno, S. S100A9 in BALF Is a Candidate Biomarker of Idiopathic Pulmonary Fibrosis. *Respir. Med.* **2012**, *106* (4), 571–580. https://doi.org/10.1016/j.rmed.2011.12.010.

(83) Obry, A.; Lequerré, T.; Hardouin, J.; Boyer, O.; Fardellone, P.; Philippe, P.; Le Loët, X.; Cosette, P.; Vittecoq, O. Identification of S100A9 as Biomarker of Responsiveness to the Methotrexate/Etanercept Combination in Rheumatoid Arthritis Using a Proteomic Approach. *PLoS ONE* **2014**, *9* (12), e115800. https://doi.org/10.1371/journal.pone.0115800.

(84) Horvath, I.; Jia, X.; Johansson, P.; Wang, C.; Moskalenko, R.; Steinau, A.; Forsgren, L.; Wågberg, T.; Svensson, J.; Zetterberg, H.; Morozova-Roche, L. A. Pro-Inflammatory S100A9 Protein as a Robust Biomarker Differentiating Early Stages of Cognitive Impairment in Alzheimer's Disease. *ACS Chem. Neurosci.* **2016**, *7* (1), 34–39. https://doi.org/10.1021/acschemneuro.5b00265.

(85) Huang, C.-H.; Kuo, C.-J.; Liang, S.-S.; Chi, S.-W.; Hsi, E.; Chen, C.-C.; Lee, K.-T.; Chiou, S.-H. Onco-Proteogenomics Identifies Urinary S100A9 and GRN as Potential Combinatorial Biomarkers for Early Diagnosis of Hepatocellular Carcinoma. *BBA Clin.* **2015**, *3*, 205–213. https://doi.org/10.1016/j.bbacli.2015.02.004.

(86) Lee, N. R.; Park, B. S.; Kim, S. Y.; Gu, A.; Kim, D. H.; Lee, J.-S.; Kim, I. S. Cytokine Secreted by S100A9 via TLR4 in Monocytes Delays Neutrophil Apoptosis by Inhibition of Caspase 9/3 Pathway. *Cytokine* **2016**, *86*, 53–63. https://doi.org/10.1016/j.cyto.2016.07.005.

(87) Rammes, A.; Roth, J.; Goebeler, M.; Klempt, M.; Hartmann, M.; Sorg, C. Myeloid-Related Protein (MRP) 8 and MRP14, Calcium-Binding Proteins of the S100 Family, Are Secreted by Activated Monocytes via a Novel, Tubulin-Dependent Pathway. *J. Biol. Chem.* **1997**, *272* (14), 9496–9502.

(88) O'Donoghue, A. J.; Jin, Y.; Knudsen, G. M.; Perera, N. C.; Jenne, D. E.; Murphy, J. E.; Craik, C. S.; Hermiston, T. W. Global Substrate Profiling of Proteases in Human Neutrophil Extracellular Traps Reveals Consensus Motif Predominantly Contributed by Elastase. *PLoS ONE* **2013**, *8* (9), e75141. https://doi.org/10.1371/journal.pone.0075141.

(89) Schenten, V.; Plançon, S.; Jung, N.; Hann, J.; Bueb, J.-L.; Bréchard, S.; Tschirhart, E. J.; Tolle, F. Secretion of the Phosphorylated Form of S100A9 from Neutrophils Is Essential for the Proinflammatory Functions of Extracellular S100A8/A9. *Front. Immunol.* **2018**, *9*. https://doi.org/10.3389/fimmu.2018.00447.

(90) Urban, C. F.; Ermert, D.; Schmid, M.; Abu-Abed, U.; Goosmann, C.; Nacken, W.; Brinkmann, V.; Jungblut, P. R.; Zychlinsky, A. Neutrophil Extracellular Traps Contain Calprotectin, a Cytosolic Protein Complex Involved in Host Defense against Candida Albicans. *PLoS Pathog.* **2009**, *5* (10), e1000639. https://doi.org/10.1371/journal.ppat.1000639.

(91) He, Z.; Riva, M.; Björk, P.; Swärd, K.; Mörgelin, M.; Leanderson, T.; Ivars, F. CD14 Is a Co-Receptor for TLR4 in the S100A9-Induced Pro-Inflammatory Response in Monocytes. *PLOS ONE* **2016**, *11* (5), e0156377. https://doi.org/10.1371/journal.pone.0156377.

(92) Tsai, S.-Y.; Segovia, J. A.; Chang, T.-H.; Morris, I. R.; Berton, M. T.; Tessier, P. A.; Tardif, M. R.; Cesaro, A.; Bose, S. DAMP Molecule S100A9 Acts as a Molecular Pattern to Enhance Inflammation during Influenza A Virus Infection: Role of DDX21-TRIF-TLR4-MyD88 Pathway. *PLoS Pathog.* **2014**, *10* (1), e1003848. https://doi.org/10.1371/journal.ppat.1003848.

(93) Källberg, E.; Vogl, T.; Liberg, D.; Olsson, A.; Björk, P.; Wikström, P.; Bergh, A.; Roth, J.; Ivars, F.; Leanderson, T. S100A9 Interaction with TLR4 Promotes Tumor Growth. *PLoS ONE* **2012**, *7* (3), e34207. https://doi.org/10.1371/journal.pone.0034207.

(94) Laouedj, M.; Tardif, M. R.; Gil, L.; Raquil, M.-A.; Lachhab, A.; Pelletier, M.; Tessier, P. A.; Barabé, F. S100A9 Induces Differentiation of Acute Myeloid Leukemia Cells through TLR4. *Blood* **2017**, *129* (14), 1980–1990. https://doi.org/10.1182/blood-2016-09-738005.

(95) Gao, H.; Zhang, X.; Zheng, Y.; Peng, L.; Hou, J.; Meng, H. S100A9-Induced Release of Interleukin (IL)-6 and IL-8 through Toll-like Receptor 4 (TLR4) in Human Periodontal Ligament Cells. *Mol. Immunol.* **2015**, *67* (2), 223–232. https://doi.org/10.1016/j.molimm.2015.05.014.

(96) Shabani, F.; Farasat, A.; Mahdavi, M.; Gheibi, N. Calprotectin (S100A8/S100A9): A Key Protein between Inflammation and Cancer. *Inflamm. Res.* **2018**. https://doi.org/10.1007/s00011-018-1173-4.

(97) Hadley, R. C.; Gu, Y.; Nolan, E. M. Initial Biochemical and Functional Evaluation of Murine Calprotectin Reveals Ca(II)-Dependence and Its Ability to Chelate Multiple Nutrient Transition Metal Ions. *Biochemistry* **2018**. https://doi.org/10.1021/acs.biochem.8b00309.

(98) Gagnon, D. M.; Brophy, M. B.; Bowman, S. E. J.; Stich, T. A.; Drennan, C. L.; Britt, R. D.; Nolan, E. M. Manganese Binding Properties of Human Calprotectin under Conditions of High and Low Calcium: X-Ray Crystallographic and Advanced Electron Paramagnetic Resonance Spectroscopic Analysis. *J. Am. Chem. Soc.* **2015**, *137* (8), 3004–3016. https://doi.org/10.1021/ja512204s.

(99) Damo, S. M.; Kehl-Fie, T. E.; Sugitani, N.; Holt, M. E.; Rathi, S.; Murphy, W. J.; Zhang, Y.; Betz, C.; Hench, L.; Fritz, G. Molecular Basis for Manganese Sequestration by Calprotectin and Roles in the Innate Immune Response to Invading Bacterial Pathogens. *Proc. Natl. Acad. Sci.* **2013**, *110* (10), 3841–3846.

(100) Besold, A. N.; Gilston, B. A.; Radin, J. N.; Ramsoomair, C.; Culbertson, E. M.; Li, C. X.; Cormack, B. P.; Chazin, W. J.; Kehl-Fie, T. E.; Culotta, V. C. The Role of Calprotectin in Withholding Zinc and Copper from Candida Albicans. *Infect. Immun.* **2017**, IAI.00779-17. https://doi.org/10.1128/IAI.00779-17.

(101) Nakashige, T. G.; Zhang, B.; Krebs, C.; Nolan, E. M. Human Calprotectin Is an Iron-Sequestering Host-Defense Protein. *Nat. Chem. Biol.* **2015**, *11* (10), 765–771. https://doi.org/10.1038/nchembio.1891.

(102) Nakashige, T. G.; Stephan, J. R.; Cunden, L. S.; Brophy, M. B.; Wommack, A. J.; Keegan, B. C.; Shearer, J. M.; Nolan, E. M. The Hexahistidine Motif of Host-Defense Protein Human Calprotectin Contributes to Zinc Withholding and Its Functional Versatility. *J. Am. Chem. Soc.* **2016**, *138* (37), 12243–12251. https://doi.org/10.1021/jacs.6b06845.

(103) Hayden, J. A.; Brophy, M. B.; Cunden, L. S.; Nolan, E. M. High-Affinity Manganese Coordination by Human Calprotectin Is Calcium-Dependent and Requires the Histidine-Rich Site Formed at the Dimer Interface. *J. Am. Chem. Soc.* **2013**, *135* (2), 775–787. https://doi.org/10.1021/ja3096416.

(104) Brophy, M. B.; Hayden, J. A.; Nolan, E. M. Calcium Ion Gradients Modulate the Zinc Affinity and Antibacterial Activity of Human Calprotectin. *J. Am. Chem. Soc.* **2012**, *134* (43), 18089–18100. https://doi.org/10.1021/ja307974e.

(105) Riva, M.; He, Z.; Källberg, E.; Ivars, F.; Leanderson, T. Human S100A9 Protein Is Stabilized by Inflammatory Stimuli via the Formation of Proteolytically-Resistant Homodimers. *PLoS ONE* **2013**, *8* (4), e61832. https://doi.org/10.1371/journal.pone.0061832.

(106) Stephan, J. R.; Yu, F.; Costello, R. M.; Bleier, B. S.; Nolan, E. M. Oxidative Post-Translational Modifications Accelerate Proteolytic Degradation of Calprotectin. *J. Am. Chem. Soc.* **2018**. https://doi.org/10.1021/jacs.8b06354.

(107) Nacken, W.; Kerkhoff, C. The Hetero-Oligomeric Complex of the S100A8/S100A9 Protein Is Extremely Protease Resistant. *FEBS Lett.* **2007**, *581* (26), 5127–5130. https://doi.org/10.1016/j.febslet.2007.09.060.

(108) Stephan, J. R.; Nolan, E. M. Calcium-Induced Tetramerization and Zinc Chelation Shield Human Calprotectin from Degradation by Host and Bacterial Extracellular Proteases. *Chem Sci* **2016**, *7* (3), 1962–1975. https://doi.org/10.1039/C5SC03287C.

(109) Vogl, T.; Stratis, A.; Wixler, V.; Völler, T.; Thurainayagam, S.; Jorch, S. K.; Zenker, S.; Dreiling, A.; Chakraborty, D.; Fröhling, M.; Paruzel, P.; Wehmeyer, C.; Hermann, S.; Papantonopoulou, O.; Geyer, C.; Loser, K.; Schäfers, M.; Ludwig, S.; Stoll, M.; Leanderson, T.; Schultze, J. L.; König, S.; Pap, T.; Roth, J. Autoinhibitory Regulation of S100A8/S100A9 Alarmin Activity Locally Restricts Sterile Inflammation. *J. Clin. Invest.* **2018**. https://doi.org/10.1172/JCI89867.

(110) Brunjes Brophy, M.; Nakashige, T. G.; Gaillard, A.; Nolan, E. M. Contributions of the S100A9 C-Terminal Tail to High-Affinity Mn(II) Chelation by the Host-Defense Protein Human Calprotectin. *J. Am. Chem. Soc.* **2013**, *135* (47), 17804–17817. https://doi.org/10.1021/ja407147d.

(111) Loes, A. N.; Shi, R.; Harms, M. J. *Zinc-Independent Activation of Toll-like Receptor 4 by S100A9*; preprint; Biochemistry, 2019. https://doi.org/10.1101/796219.

(112) Wheeler, L. C.; Donor, M. T.; Prell, J. S.; Harms, M. J. Multiple Evolutionary Origins of Ubiquitous Cu2+ and Zn2+ Binding in the S100 Protein Family. *PLOS ONE* **2016**, *11* (10), e0164740. https://doi.org/10.1371/journal.pone.0164740.

(113) Wheeler, L. C.; Harms, M. J. Human S100A5 Binds Ca2+ and Cu2+ Independently. *BMC Biophys.* **2017**, *10* (1). https://doi.org/10.1186/s13628-017-0040-y.

(114) Park, B. S.; Song, D. H.; Kim, H. M.; Choi, B.-S.; Lee, H.; Lee, J.-O. The Structural Basis of Lipopolysaccharide Recognition by the TLR4–MD-2 Complex. *Nature* **2009**, *458* (7242), 1191–1195. https://doi.org/10.1038/nature07830.

(115) Triantafilou, M.; Triantafilou, K. Lipopolysaccharide Recognition: CD14, TLRs and the LPS-Activation Cluster. *Trends Immunol.* **2002**, *23* (6), 301–304. https://doi.org/10.1016/S1471-4906(02)02233-0.

(116) FinleySO, F.; Leturcqll, D. Analysis of Lipopolysaccharide Binding ByCD14. 6.

(117) Nagai, Y.; Akashi, S.; Nagafuku, M.; Ogata, M.; Iwakura, Y.; Akira, S.; Kitamura, T.; Kosugi, A.; Kimoto, M.; Miyake, K. Essential Role of MD-2 in LPS Responsiveness and TLR4 Distribution. *Nat. Immunol.* **2002**, *3* (7), 667–672. https://doi.org/10.1038/ni809.

(118) Viriyakosol, S.; Tobias, P. S.; Kitchens, R. L.; Kirkland, T. N. MD-2 Binds to Bacterial Lipopolysaccharide. 9.

(119) Visintin, A.; Latz, E.; Monks, B. G.; Espevik, T.; Golenbock, D. T. Lysines 128 and 132 Enable Lipopolysaccharide Binding to MD-2, Leading to Toll-like Receptor-4 Aggregation and Signal Transduction. *J. Biol. Chem.* **2003**, *278* (48), 48313–48320. https://doi.org/10.1074/jbc.M306802200.

(120) Biswas, S. K.; Bist, P.; Dhillon, M. K.; Kajiji, T.; del Fresno, C.; Yamamoto, M.; Lopez-Collazo, E.; Akira, S.; Tergaonkar, V. Role for MyD88-Independent, TRIF Pathway in Lipid A/TLR4-Induced Endotoxin Tolerance. *J. Immunol.* **2007**, *179* (6), 4083–4092. https://doi.org/10.4049/jimmunol.179.6.4083.

(121) Tanimura, N.; Saitoh, S.; Matsumoto, F.; Akashi-Takamura, S.; Miyake, K. Roles for LPS-Dependent Interaction and Relocation of TLR4 and TRAM in TRIF-Signaling. *Biochem. Biophys. Res. Commun.* **2008**, *368* (1), 94–99. https://doi.org/10.1016/j.bbrc.2008.01.061.

(122) Lu, Y.-C.; Yeh, W.-C.; Ohashi, P. S. LPS/TLR4 Signal Transduction Pathway. *Cytokine* **2008**, *42* (2), 145–151. https://doi.org/10.1016/j.cyto.2008.01.006.

(123) Li, W.; Feng, G.; Gauthier, J. M.; Lokshina, I.; Higashikubo, R.; Evans, S.; Liu, X.; Hassan, A.; Tanaka, S.; Cicka, M.; Hsiao, H.-M.; Ruiz-Perez, D.; Bredemeyer, A.; Gross, R. W.; Mann, D. L.; Tyurina, Y. Y.; Gelman, A. E.; Kagan, V. E.; Linkermann, A.; Lavine, K. J.; Kreisel, D. Ferroptotic Cell Death and TLR4/Trif Signaling Initiate Neutrophil Recruitment after Heart Transplantation. *J. Clin. Invest.* **2019**, *129* (6), 2293–2304. https://doi.org/10.1172/JCI126428.

(124) Zhou, H.; Andonegui, G.; Wong, C. H. Y.; Kubes, P. Role of Endothelial TLR4 for Neutrophil Recruitment into Central Nervous System Microvessels in Systemic Inflammation. *J. Immunol.* **2009**, *183* (8), 5244–5250. https://doi.org/10.4049/jimmunol.0901309.

(125) Mizgerd, J. P.; Lupa, M. M.; Spieker, M. S. NF-KB P50 Facilitates Neutrophil Accumulation during LPS-Induced Pulmonary Inflammation. *BMC Immunol.* **2004**, 11.

(126) Lamping, N.; Dettmer, R.; Schröder, N. W.; Pfeil, D.; Hallatschek, W.; Burger, R.; Schumann, R. R. LPS-Binding Protein Protects Mice from Septic Shock Caused by LPS or Gram-Negative Bacteria. *J. Clin. Invest.* **1998**, *101* (10), 2065–2071. https://doi.org/10.1172/JCI2338.

(127) Opal, S. M.; Scannon, P. J.; Vincent, J.; White, M.; Carroll, S. F.; Palardy, J. E.; Parejo, N. A.; Pribble, J. P.; Lemke, J. H. Relationship between Plasma Levels of Lipopolysaccharide (LPS) and LPS-Binding Protein in Patients with Severe Sepsis and Septic Shock. *J. Infect. Dis.* **1999**, *180* (5), 1584–1589. https://doi.org/10.1086/315093.

(128) Downey, J. S.; Han, J. CELLULAR ACTIVATION MECHANISMS IN SEPTIC SHOCK. 9.

(129) Björk, P.; Björk, A.; Vogl, T.; Stenström, M.; Liberg, D.; Olsson, A.; Roth, J.; Ivars, F.; Leanderson, T. Identification of Human S100A9 as a Novel Target for Treatment of Autoimmune Disease via Binding to Quinoline-3-Carboxamides. *PLoS Biol.* **2009**, *7* (4), e1000097. https://doi.org/10.1371/journal.pbio.1000097.

(130) Loes, A. N.; Bridgham, J. T.; Harms, M. J. Coevolution of the Toll-Like Receptor 4 Complex with Calgranulins and Lipopolysaccharide. *Front. Immunol.* **2018**, *9*. https://doi.org/10.3389/fimmu.2018.00304.

(131) Harman, J.; Loes, A.; Warren, G.; Heaphy, M.; Lampi, K.; Harms, M. Evolution of Multifunctionality Through a Pleiotropic Substitution in the Innate Immune Protein S100A9. *eLife* **2020**. https://doi.org/10.7554/eLife.54100.

(132) Auvynet, C.; Rosenstein, Y. Multifunctional Host Defense Peptides: Antimicrobial Peptides, the Small yet Big Players in Innate and Adaptive Immunity: AMPs, the Small yet Big Players of Immunity. *FEBS J.* **2009**, *276* (22), 6497–6508. https://doi.org/10.1111/j.1742-4658.2009.07360.x.

(133) Postel, S.; Küfner, I.; Beuter, C.; Mazzotta, S.; Schwedt, A.; Borlotti, A.; Halter, T.; Kemmerling, B.; Nürnberger, T. The Multifunctional Leucine-Rich Repeat Receptor Kinase BAK1 Is Implicated in Arabidopsis Development and Immunity. *Eur. J. Cell Biol.* **2010**, *89* (2–3), 169–174. https://doi.org/10.1016/j.ejcb.2009.11.001.

(134) Gudmundsson, G. H.; Agerberth, B. Neutrophil Antibacterial Peptides, Multifunctional Effector Molecules in the Mammalian Immune System. *J. Immunol. Methods* **1999**, *232* (1–2), 45–54. https://doi.org/10.1016/S0022-1759(99)00152-0.

(135) Singh, V.; Yeoh, B. S.; Chassaing, B.; Zhang, B.; Saha, P.; Xiao, X.; Awasthi, D.; Shashidharamurthy, R.; Dikshit, M.; Gewirtz, A.; Vijay-Kumar, M. Microbiota-Inducible Innate Immune Siderophore Binding Protein Lipocalin 2 Is Critical for Intestinal Homeostasis. *Cell. Mol. Gastroenterol. Hepatol.* **2016**, *2* (4), 482-498.e6. https://doi.org/10.1016/j.jcmgh.2016.03.007.

(136) Säemann, M. D.; Haidinger, M.; Hecking, M.; Hörl, W. H.; Weichhart, T. The Multifunctional Role of MTOR in Innate Immunity: Implications for Transplant Immunity. *Am. J. Transplant.* **2009**, *9* (12), 2655–2661. https://doi.org/10.1111/j.1600-6143.2009.02832.x.

(137) Bals, R.; Wilson, J. M. Cathelicidins - a Family of Multifunctional Antimicrobial Peptides. *Cell. Mol. Life Sci. CMLS* **2003**, *60* (4), 711–720. https://doi.org/10.1007/s00018-003-2186-9.

(138) Leask, A.; Abraham, D. J. The Role of Connective Tissue Growth Factor, a Multifunctional Matricellular Protein, in Fibroblast Biology. *Biochem. Cell Biol.* **2003**, *81* (6), 355–363. https://doi.org/10.1139/o03-069.

(139) Ehrchen, J. M.; Sunderkötter, C.; Foell, D.; Vogl, T.; Roth, J. The Endogenous Toll-like Receptor 4 Agonist S100A8/S100A9 (Calprotectin) as Innate Amplifier of Infection, Autoimmunity, and Cancer. *J. Leukoc. Biol.* **2009**, *86* (3), 557–566. https://doi.org/10.1189/jlb.1008647.

(140) Salama, I.; Malone, P. S.; Mihaimeed, F.; Jones, J. L. A Review of the S100 Proteins in Cancer. *Eur. J. Surg. Oncol. EJSO* **2008**, *34* (4), 357–364. https://doi.org/10.1016/j.ejso.2007.04.009.

(141) Gruden, M. A.; Davydova, T. V.; Wang, C.; Narkevich, V. B.; Fomina, V. G.; Kudrin, V. S.; Morozova-Roche, L. A.; Sewell, R. D. E. The Misfolded Pro-Inflammatory Protein S100A9 Disrupts Memory via Neurochemical Remodelling Instigating an Alzheimer's Disease-like Cognitive Deficit. *Behav. Brain Res.* **2016**, *306*, 106–116. https://doi.org/10.1016/j.bbr.2016.03.016.

(142) Peng, S.; Sun, X.; Wang, X.; Wang, H.; Shan, Z.; Teng, W.; Li, C. Myeloid Related Proteins Are Up-Regulated in Autoimmune Thyroid Diseases and Activate Toll-like Receptor 4 and pro-Inflammatory Cytokines in Vitro. *Int. Immunopharmacol.* **2018**, *59*, 217–226. https://doi.org/10.1016/j.intimp.2018.04.009.

(143) Lee, M. W.; Lee, E. Y.; Wong, G. C. L. What Can Pleiotropic Proteins in Innate Immunity Teach Us about Bioconjugation and Molecular Design? *Bioconjug. Chem.* **2018**, *29* (7), 2127–2139. https://doi.org/10.1021/acs.bioconjchem.8b00176.

(144) Bomblies, K. Pleiotropic Effects of the Duplicate Maize FLORICAULA/LEAFY Genes Zfl1 and Zfl2 on Traits Under Selection During Maize Domestication. *Genetics* **2005**, *172* (1), 519–531. https://doi.org/10.1534/genetics.105.048595.

(145) Stearns, F. W. One Hundred Years of Pleiotropy: A Retrospective. *Genetics* **2010**, *186* (3), 767–773. https://doi.org/10.1534/genetics.110.122549.

(146) Paaby, A. B.; Rockman, M. V. The Many Faces of Pleiotropy. *Trends Genet.* **2013**, *29* (2), 66–73. https://doi.org/10.1016/j.tig.2012.10.010.

(147) Berntzen, H. B.; Fagerhol, M. K. L1, a Major Granulocyte Protein; Isolation of High Quantities of Its Subunits. *Scand. J. Clin. Lab. Invest.* **1990**, *50* (7), 769–774. https://doi.org/10.1080/00365519009091071.

(148) Vogl, T.; Gharibyan, A. L.; Morozova-Roche, L. A. Pro-Inflammatory S100A8 and S100A9 Proteins: Self-Assembly into Multifunctional Native and Amyloid Complexes. *Int. J. Mol. Sci.* **2012**, *13* (12), 2893–2917. https://doi.org/10.3390/ijms13032893.

(149) Duan, L.; Wu, R.; Zhang, X.; Wang, D.; You, Y.; Zhang, Y.; Zhou, L.; Chen, W. HBx-Induced S100A9 in NF-KB Dependent Manner Promotes Growth and Metastasis of Hepatocellular Carcinoma Cells. *Cell Death Dis.* **2018**, *9* (6). https://doi.org/10.1038/s41419-018-0512-2.

(150) Shepherd, C. E.; Goyette, J.; Utter, V.; Rahimi, F.; Yang, Z.; Geczy, C. L.; Halliday, G. M. Inflammatory S100A9 and S100A12 Proteins in Alzheimer's Disease. *Neurobiol. Aging* **2006**, *27* (11), 1554–1563. https://doi.org/10.1016/j.neurobiolaging.2005.09.033.

(151) Schiopu, A.; Cotoi, O. S. S100A8 and S100A9: DAMPs at the Crossroads between Innate Immunity, Traditional Risk Factors, and Cardiovascular Disease. *Mediators Inflamm.* **2013**, *2013*, 1–10. https://doi.org/10.1155/2013/828354.

(152) Kang, J. H.; Hwang, S. M.; Chung, I. Y. S100A8, S100A9 and S100A12 Activate Airway Epithelial Cells to Produce MUC5AC via Extracellular Signal-Regulated Kinase and Nuclear Factor- κ B Pathways. *Immunology* **2015**, *144* (1), 79–90. https://doi.org/10.1111/imm.12352.

(153) Chen, B.; Miller, A. L.; Rebelatto, M.; Brewah, Y.; Rowe, D. C.; Clarke, L.; Czapiga, M.; Rosenthal, K.; Imamichi, T.; Chen, Y.; Chang, C.-S.; Chowdhury, P. S.; Naiman, B.; Wang, Y.; Yang, D.; Humbles, A. A.; Herbst, R.; Sims, G. P. S100A9 Induced Inflammatory Responses Are Mediated by Distinct Damage Associated Molecular Patterns (DAMP) Receptors In Vitro and In Vivo. *PLOS ONE* **2015**, *10* (2), e0115828. https://doi.org/10.1371/journal.pone.0115828.

(154) Besold, A. N.; Culbertson, E. M.; Nam, L.; Hobbs, R. P.; Boyko, A.; Maxwell, C. N.; Chazin, W. J.; Marques, A. R.; Culotta, V. C. Antimicrobial Action of Calprotectin That Does Not Involve Metal Withholding. *Metallomics* **2018**. https://doi.org/10.1039/C8MT00133B.

(155) Clark, H. L.; Jhingran, A.; Sun, Y.; Vareechon, C.; de Jesus Carrion, S.; Skaar, E. P.; Chazin, W. J.; Calera, J. A.; Hohl, T. M.; Pearlman, E. Zinc and Manganese Chelation by Neutrophil S100A8/A9 (Calprotectin) Limits Extracellular *Aspergillus Fumigatus* Hyphal Growth and Corneal Infection. *J. Immunol.* **2016**, *196* (1), 336–344. https://doi.org/10.4049/jimmunol.1502037.

(156) Liu, J. Z.; Jellbauer, S.; Poe, A. J.; Ton, V.; Pesciaroli, M.; Kehl-Fie, T. E.; Restrepo, N. A.; Hosking, M. P.; Edwards, R. A.; Battistoni, A.; Pasquali, P.; Lane, T. E.; Chazin, W. J.; Vogl, T.; Roth, J.; Skaar, E. P.; Raffatellu, M. Zinc Sequestration by the Neutrophil Protein Calprotectin Enhances Salmonella Growth in the Inflamed Gut. *Cell Host Microbe* **2012**, *11* (3), 227–239. https://doi.org/10.1016/j.chom.2012.01.017.

(157) Baker, T. M.; Nakashige, T. G.; Nolan, E. M.; Neidig, M. L. Magnetic Circular Dichroism Studies of Iron(II) Binding to Human Calprotectin. *Chem Sci* **2017**, *8* (2), 1369–1377. https://doi.org/10.1039/C6SC03487J.

(158) Nisapakultorn, K.; Ross, K. F.; Herzberg, M. C. Calprotectin Expression Inhibits Bacterial Binding to Mucosal Epithelial Cells. *Infect. Immun.* **2001**, *69* (6), 3692–3696. https://doi.org/10.1128/IAI.69.6.3692-3696.2001.

(159) Vogl, T.; Tenbrock, K.; Ludwig, S.; Leukert, N.; Ehrhardt, C.; van Zoelen, M. A. D.; Nacken, W.; Foell, D.; van der Poll, T.; Sorg, C.; Roth, J. Mrp8 and Mrp14 Are Endogenous Activators of Toll-like Receptor 4, Promoting Lethal, Endotoxin-Induced Shock. *Nat. Med.* **2007**, *13* (9), 1042–1049. https://doi.org/10.1038/nm1638.

(160) Kim, H.-J.; Kang, H. J.; Lee, H.; Lee, S.-T.; Yu, M.-H.; Kim, H.; Lee, C. Identification of S100A8 and S100A9 as Serological Markers for Colorectal Cancer. *J. Proteome Res.* **2009**, *8* (3), 1368–1379. https://doi.org/10.1021/pr8007573.

(161) Averill, M. M.; Kerkhoff, C.; Bornfeldt, K. E. S100A8 and S100A9 in Cardiovascular Biology and Disease. *Arterioscler. Thromb. Vasc. Biol.* **2012**, *32* (2), 223–229. https://doi.org/10.1161/ATVBAHA.111.236927.

(162) Bozzi, A.; Nolan, E. M. Avian MRP126 Restricts Microbial Growth through Ca(II)-Dependent Zn(II) Sequestration. *Biochemistry* **2019**, acs.biochem.9b01012. https://doi.org/10.1021/acs.biochem.9b01012.

(163) Cunden, L. S.; Nolan, E. M. Bioinorganic Explorations of Zn(II) Sequestration by Human S100 Host-Defense Proteins. *Biochemistry* **2018**. https://doi.org/10.1021/acs.biochem.7b01305.

(164) Henry, C. M.; Sullivan, G. P.; Clancy, D. M.; Afonina, I. S.; Kulms, D.; Martin, S. J. Neutrophil-Derived Proteases Escalate Inflammation through Activation of IL-36 Family Cytokines. *Cell Rep.* **2016**, *14* (4), 708–722. https://doi.org/10.1016/j.celrep.2015.12.072.

(165) Stapels, D. A.; Geisbrecht, B. V.; Rooijakkers, S. H. Neutrophil Serine Proteases in Antibacterial Defense. *Curr. Opin. Microbiol.* **2015**, *23*, 42–48. https://doi.org/10.1016/j.mib.2014.11.002.

(166) Kessenbrock, K.; Dau, T.; Jenne, D. E. Tailor-Made Inflammation: How Neutrophil Serine Proteases Modulate the Inflammatory Response. *J. Mol. Med.* **2011**, *89* (1), 23–28. https://doi.org/10.1007/s00109-010-0677-3.

(167) Heutinck, K. M.; ten Berge, I. J. M.; Hack, C. E.; Hamann, J.; Rowshani, A. T. Serine Proteases of the Human Immune System in Health and Disease. *Mol. Immunol.* **2010**, *47* (11–12), 1943–1955. https://doi.org/10.1016/j.molimm.2010.04.020.

(168) Janoff, A. Neutrophil Proteases in Inflammation. *Annu. Rev. Med.* **1972**, *23* (1), 177–190. https://doi.org/10.1146/annurev.me.23.020172.001141.

(169) Jerke, U.; Hernandez, D. P.; Beaudette, P.; Korkmaz, B.; Dittmar, G.; Kettritz, R. Neutrophil Serine Proteases Exert Proteolytic Activity on Endothelial Cells. *Kidney Int.* **2015**, *88* (4), 764–775. https://doi.org/10.1038/ki.2015.159.

(170) Streicher, W. W.; Lopez, M. M.; Makhatadze, G. I. Modulation of Quaternary Structure of S100 Proteins by Calcium Ions. *Biophys. Chem.* **2010**, *151* (3), 181–186. https://doi.org/10.1016/j.bpc.2010.06.003.

(171) Eick, G. N.; Bridgham, J. T.; Anderson, D. P.; Harms, M. J.; Thornton, J. W. Robustness of Reconstructed Ancestral Protein Functions to Statistical Uncertainty. *Mol. Biol. Evol.* **2016**, msw223. https://doi.org/10.1093/molbev/msw223.

(172) Anderson, J. A.; Loes, A. N.; Waddell, G. L.; Harms, M. J. Tracing the Evolution of Novel Features of Human Toll-like Receptor 4. *Protein Sci.* **2019**, pro.3644. https://doi.org/10.1002/pro.3644.

(173) Ehrhardt, C.; Schmolke, M.; Matzke, A.; Knoblauch, A.; Will, C.; Wixler, V.; Ludwig, S. Polyethylenimine, a Cost-Effective Transfection Reagent. *Signal Transduct.* **2006**, *6* (3), 179–184. https://doi.org/10.1002/sita.200500073.

(174) Fu, Z.; Thorpe, M.; Akula, S.; Chahal, G.; Hellman, L. T. Extended Cleavage Specificity of Human Neutrophil Elastase, Human Proteinase 3, and Their Distant Ortholog Clawed Frog PR3—Three Elastases With Similar Primary but Different Extended Specificities and Stability. *Front. Immunol.* **2018**, *9*, 2387. https://doi.org/10.3389/fimmu.2018.02387.

(175) Itou, H.; Yao, M.; Fujita, I.; Watanabe, N.; Suzuki, M.; Nishihira, J.; Tanaka, I. The Crystal Structure of Human MRP14 (S100A9), a Ca2+-Dependent Regulator Protein in Inflammatory Process. *J. Mol. Biol.* **2002**, *316* (2), 265–276. https://doi.org/10.1006/jmbi.2001.5340.

(176) Ibrahim, Z. A.; Armour, C. L.; Phipps, S.; Sukkar, M. B. RAGE and TLRs: Relatives, Friends or Neighbours? *Mol. Immunol.* **2013**, *56* (4), 739–744. https://doi.org/10.1016/j.molimm.2013.07.008.

(177) Poltorak, A. Defective LPS Signaling in C3H/HeJ and C57BL/10ScCr Mice: Mutations in Tlr4 Gene. *Science* **1998**, *282* (5396), 2085–2088. https://doi.org/10.1126/science.282.5396.2085.

(178) Prince, L. R.; Whyte, M. K.; Sabroe, I.; Parker, L. C. The Role of TLRs in Neutrophil Activation. *Curr. Opin. Pharmacol.* **2011**, *11* (4), 397–403. https://doi.org/10.1016/j.coph.2011.06.007.

(179) Guharoy, M.; Bhowmick, P.; Tompa, P. Design Principles Involving Protein Disorder Facilitate Specific Substrate Selection and Degradation by the Ubiquitin-Proteasome System. *J. Biol. Chem.* **2016**, *291* (13), 6723–6731. https://doi.org/10.1074/jbc.R115.692665.

(180) Fontana, A.; Polverino de Laureto, P.; De Filippis, V.; Scaramella, E.; Zambonin, M. Probing the Partly Folded States of Proteins by Limited Proteolysis. *Fold. Des.* **1997**, *2* (2), R17–R26. https://doi.org/10.1016/S1359-0278(97)00010-2.

(181) Hubbard, S. J. The Structural Aspects of Limited Proteolysis of Native Proteins. *Biochim. Biophys. Acta BBA - Protein Struct. Mol. Enzymol.* **1998**, *1382* (2), 191–206. https://doi.org/10.1016/S0167-4838(97)00175-1.

(182) Ottesen, M. Induction of Biological Activity by Limited Proteolysis. *Annu. Rev. Biochem.* **1967**, *36* (1), 55–76. https://doi.org/10.1146/annurev.bi.36.070167.000415.

(183) Imoto, T.; Yamada, H.; Ueda, T. Unfolding Rates of Globular Proteins Determined by Kinetics of Proteolysis. *J. Mol. Biol.* **1986**, *190* (4), 647–649. https://doi.org/10.1016/0022-2836(86)90250-0.

(184) Garg, A. D.; Nowis, D.; Golab, J.; Vandenabeele, P.; Krysko, D. V.; Agostinis, P. Immunogenic Cell Death, DAMPs and Anticancer Therapeutics: An Emerging Amalgamation. *Biochim. Biophys. Acta BBA - Rev. Cancer* **2010**, *1805* (1), 53–71. https://doi.org/10.1016/j.bbcan.2009.08.003.

(185) Rubartelli, A.; Lotze, M. T. Inside, Outside, Upside down: Damage-Associated Molecular-Pattern Molecules (DAMPs) and Redox. *Trends Immunol.* **2007**, *28* (10), 429–436. https://doi.org/10.1016/j.it.2007.08.004.

(186) Bianchi, M. E. DAMPs, PAMPs and Alarmins: All We Need to Know about Danger. *J. Leukoc. Biol.* **2007**, *81* (1), 1–5. https://doi.org/10.1189/jlb.0306164.

(187) He, X. Toward a Molecular Understanding of Pleiotropy. *Genetics* **2006**, *173* (4), 1885–1891. https://doi.org/10.1534/genetics.106.060269.

(188) Hirano, T. Molecular Basis Underlying Functional Pleiotropy of Cytokines and Growth Factors. *Biochem. Biophys. Res. Commun.* **1999**, *260* (2), 303–308. https://doi.org/10.1006/bbrc.1999.0609.

(189) Jones, D. T.; Taylor, W. R.; Thornton, J. M. The Rapid Generation of Mutation Data Matrices from Protein Sequences. *Bioinformatics* **1992**, *8* (3), 275–282. https://doi.org/10.1093/bioinformatics/8.3.275.

(190) Guindon, S.; Dufayard, J.-F.; Lefort, V.; Anisimova, M.; Hordijk, W.; Gascuel, O. New Algorithms and Methods to Estimate Maximum-Likelihood Phylogenies: Assessing the Performance of PhyML 3.0. *Syst. Biol.* **2010**, *59* (3), 307–321. https://doi.org/10.1093/sysbio/syq010.

(191) Futami, J.; Atago, Y.; Azuma, A.; Putranto, E. W.; Kinoshita, R.; Murata, H.; Sakaguchi, M. An Efficient Method for the Preparation of Preferentially Heterodimerized Recombinant S100A8/A9 Coexpressed in Escherichia Coli. *Biochem. Biophys. Rep.* **2016**, *6*, 94–100. https://doi.org/10.1016/j.bbrep.2016.03.009.

(192) Cunden, L. S.; Gaillard, A.; Nolan, E. M. Calcium Ions Tune the Zinc-Sequestering Properties and Antimicrobial Activity of Human S100A12. *Chem Sci* **2016**, *7* (2), 1338–1348. https://doi.org/10.1039/C5SC03655K.

(193) Chang, C.-C.; Khan, I.; Tsai, K.-L.; Li, H.; Yang, L.-W.; Chou, R.-H.; Yu, C. Blocking the Interaction between S100A9 and RAGE V Domain Using CHAPS Molecule: A Novel Route to Drug Development against Cell Proliferation. *Biochim. Biophys. Acta BBA - Proteins Proteomics* **2016**, *1864* (11), 1558–1569. https://doi.org/10.1016/j.bbapap.2016.08.008.

(194) Carr, H. Y.; Purcell, E. M. Effects of Diffusion on Free Precession in Nuclear Magnetic Resonance Experiments. *Phys. Rev.* **1954**, *94* (3), 630–638. https://doi.org/10.1103/PhysRev.94.630.

(195) Ferrage, F.; Piserchio, A.; Cowburn, D.; Ghose, R. On the Measurement of 15N–{1H} Nuclear Overhauser Effects. *J. Magn. Reson.* **2008**, *192* (2), 302–313. https://doi.org/10.1016/j.jmr.2008.03.011.

(196) Vogl, T.; Leukert, N.; Barczyk, K.; Strupat, K.; Roth, J. Biophysical Characterization of S100A8 and S100A9 in the Absence and Presence of Bivalent

Cations. *Biochim. Biophys. Acta BBA - Mol. Cell Res.* **2006**, *1763* (11), 1298–1306. https://doi.org/10.1016/j.bbamcr.2006.08.028.

(197) Otterbein, L. R.; Kordowska, J.; Witte-Hoffmann, C.; Dominguez, R. Crystal Structures of S100A6 in the Ca2⁺-Free and Ca2⁺-Bound States: The Calcium Sensor Mechanism of S100 Proteins Revealed at Atomic Resolution. 11.

(198) Knies, J.; Cai, F.; Weinreich, D. M. Enzyme Efficiency but Not Thermostability Drives Cefotaxime Resistance Evolution in TEM-1 β-Lactamase. *Mol. Biol. Evol.* **2017**, msx053. https://doi.org/10.1093/molbev/msx053.

(199) Seong, S.-Y.; Matzinger, P. Hydrophobicity: An Ancient Damage-Associated Molecular Pattern That Initiates Innate Immune Responses. *Nat. Rev. Immunol.* **2004**, *4* (6), 469–478. https://doi.org/10.1038/nri1372.

(200) Delaglio, F.; Grzesiek, S.; Vuister, G.; Zhu, G.; Bax, A. A Multidimensional Spectral Processing System Based on UNIX Pipes. *J Biomol NMR* **1995**, No. 6, 277–293. https://doi.org/10.1007/BF00197809.

(201) Skinner, S. P.; Goult, B. T.; Fogh, R. H.; Boucher, W.; Stevens, T. J.; Laue, E. D.; Vuister, G. W. Structure Calculation, Refinement and Validation Using *CcpNmr Analysis*. *Acta Crystallogr. D Biol. Crystallogr.* **2015**, *71* (1), 154–161. https://doi.org/10.1107/S1399004714026662.

(202) Hansen, D. F.; Vallurupalli, P.; Kay, L. E. An Improved ^{15}N Relaxation Dispersion Experiment for the Measurement of Millisecond Time-Scale Dynamics in Proteins †. *J. Phys. Chem. B* **2008**, *112* (19), 5898–5904. https://doi.org/10.1021/jp074793o.

(203) Farrow, N. A.; Muhandiram, R.; Singer, A. U.; Pascal, S. M.; Kay, C. M.; Gish, G.; Shoelson, S. E.; Pawson, T.; Forman-Kay, J. D.; Kay, L. E. Backbone Dynamics of a Free and a Phosphopeptide-Complexed Src Homology 2 Domain Studied by 15N NMR Relaxation. *Biochemistry* **1994**, *33* (19), 5984–6003. https://doi.org/10.1021/bi00185a040.

(204) Lee, W.; Tonelli, M.; Markley, J. L. NMRFAM-SPARKY: Enhanced Software for Biomolecular NMR Spectroscopy. *Bioinformatics* **2015**, *31* (8), 1325–1327. https://doi.org/10.1093/bioinformatics/btu830.

(205) Portou, M. J.; Baker, D.; Abraham, D.; Tsui, J. The Innate Immune System, Toll-like Receptors and Dermal Wound Healing: A Review. *Vascul. Pharmacol.* **2015**, *71*, 31–36. https://doi.org/10.1016/j.vph.2015.02.007.

(206) Akira, S.; Uematsu, S.; Takeuchi, O. Pathogen Recognition and Innate Immunity. *Cell* **2006**, *124* (4), 783–801. https://doi.org/10.1016/j.cell.2006.02.015.

(207) Zhang, Q.; Zmasek, C. M.; Godzik, A. Domain Architecture Evolution of Pattern-Recognition Receptors. *Immunogenetics* **2010**, *62* (5), 263–272. https://doi.org/10.1007/s00251-010-0428-1.
(208) Takeuchi, O.; Akira, S. Pattern Recognition Receptors and Inflammation. *Cell* **2010**, *140* (6), 805–820. https://doi.org/10.1016/j.cell.2010.01.022.

(209) Kawai, T.; Akira, S. The Role of Pattern-Recognition Receptors in Innate Immunity: Update on Toll-like Receptors. *Nat. Immunol.* **2010**, *11* (5), 373–384. https://doi.org/10.1038/ni.1863.

(210) Versluys, M.; Tarkowski, Ł. P.; Van den Ende, W. Fructans As DAMPs or MAMPs: Evolutionary Prospects, Cross-Tolerance, and Multistress Resistance Potential. *Front. Plant Sci.* **2017**, *7*. https://doi.org/10.3389/fpls.2016.02061.

(211) Porter, J. R.; Moeder, K. E.; Sibbald, C. A.; Zimmerman, M. I.; Hart, K. M.; Greenberg, M. J.; Bowman, G. R. Cooperative Changes in Solvent Exposure Identify Cryptic Pockets, Switches, and Allosteric Coupling. *Biophys. J.* **2019**, *116* (5), 818–830. https://doi.org/10.1016/j.bpj.2018.11.3144.

(212) Singh, S.; Bowman, G. R. Quantifying Allosteric Communication via Both Concerted Structural Changes and Conformational Disorder with CARDS. *J. Chem. Theory Comput.* **2017**, *13* (4), 1509–1517. https://doi.org/10.1021/acs.jctc.6b01181.

(213) Wheeler, L. C.; Lim, S. A.; Marqusee, S.; Harms, M. J. The Thermostability and Specificity of Ancient Proteins. *Curr. Opin. Struct. Biol.* **2016**, *38*, 37–43. https://doi.org/10.1016/j.sbi.2016.05.015.

(214) Bloom, J. D.; Raval, A.; Wilke, C. O. Thermodynamics of Neutral Protein Evolution. *Genetics* **2007**, *175* (1), 255–266. https://doi.org/10.1534/genetics.106.061754.

(215) Taverna, D. M.; Goldstein, R. A. Why Are Proteins Marginally Stable? *Proteins Struct. Funct. Genet.* **2002**, *46* (1), 105–109. https://doi.org/10.1002/prot.10016.