HISTORICAL AND FUNCTIONAL INSIGHTS INTO TOLL-LIKE RECEPTOR 4 ACTIVATION BY LIPOPOLYSACCHARIDE AND CALGRANULINS

by

ANDREA N. LOES

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

December 2018

DISSERTATION APPROVAL PAGE

Student: Andrea N. Loes

Title: Historical and Functional Insights into Toll-like Receptor 4 Activation by Lipopolysaccharide and Calgranulins

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:

Victoria DeRose	Chairperson
Michael Harms	Advisor
Brad Nolen	Core Member
Karen Guillemin	Institutional Representative

and

Janet Woodruff-Borden	Vice Provost and Dean of the Graduate School

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

Degree awarded December 2018

© 2018 Andrea N. Loes

DISSERTATION ABSTRACT

Andrea N. Loes

Doctor of Philosophy

Department of Chemistry and Biochemistry

August 2018

Title: Historical and Functional Insights into Toll-like Receptor 4 Activation by Lipopolysaccharide and Calgranulins

Toll-like receptor 4 (TLR4) is an important vertebrate innate immune receptor. TLR4 recognizes both endogenous and exogenous danger signals to trigger an NF-kB dependent inflammatory response. While exogenous danger signal recognition is an essential part of pathogen response by the innate immune system, endogenous danger signal recognition by TLR4 can lead to chronic and pathological inflammation. Understanding the differences in recognition of these two types of danger signals would allow for independent modulation of pathogen and host triggered inflammatory response through TLR4. Here, we examine the evolution of activation of TLR4 by two agonists, pathogen-derived lipopolysaccharide and host-produced S100A9. We show that these two types of signals evolved earlier than previously thought. We identified TLR4 cofactors MD-2 and CD14 in amphibians and fish, and validated that zebrafish TLR4 can recognize LPS. By contrast, we find that S100A9 activation evolved in the ancestor of amniotes. We identified an ortholog of S100A9 in birds and reptiles capable of activating TLR4. Using comparative immunology, we found that the requirements for LPS and S100A9 activation are different. In addition to our evolutionary studies, we used molecular approaches to probe if zinc binding to S100A9 is necessary for TLR4 activation. We found that activation of TLR4 by S100A9 occurs even in the absence of

zinc. Finally, we describe how our evolutionary approach led to mechanistic hypotheses regarding TLR4 activation by both LPS and S100A9. This has led to ongoing projects in the Harms lab. This dissertation includes previously published and unpublished co-authored material.

CURRICULUM VITAE

NAME OF AUTHOR: Andrea N. Loes

GRADUATE AND UNDERGRADUATE SCHOOLS ATTENDED:

University of Oregon, Eugene, OR
Arizona State University, Tempe, AZ

DEGREES AWARDED:

Doctor of Philosophy, Chemistry, 2018, University of Oregon
Master of Science, Chemistry, 2015, University of Oregon
Bachelor of Science, Microbiology, 2013, Arizona State University
Bachelor of Science, Biochemistry, 2013, Arizona State University

AREAS OF SPECIAL INTEREST:

Biochemistry
Microbiology
Molecular Biology

PROFESSIONAL EXPERIENCE:

Graduate Research Assistant, Harms Lab, University of Oregon, Eugene, OR 2014-2018

Graduate Teaching Fellow, Department of Chemistry and Biochemistry, University of Oregon, Eugene, OR, 2013-2014

Undergraduate Researcher, Haydel Lab, Arizona State University, Tempe, AZ, 2011-2013

Undergraduate Researcher, Deutch Lab, Arizona State University West, Glendale, AZ, 2010-2011

Quality Control Intern, Marlyn Nutraceuticals, Phoenix, AZ, Summer 2009

GRANTS, AWARDS, AND HONORS:

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

Society for Molecular Biology and Evolution, Registration Award, 2017

National Science Foundation Graduate Research Fellowship, Honorable Mention, 2015

Adamson Family New Graduate Student Excellence Award, 2014

PUBLICATIONS:

Loes, A.N., R. Shi, M.J. Harms. "Zinc-binding and the C-terminal tail are not required for pro-inflammatory activity of S100A9". (*in prep*)

Loes, A.N., M.N. Hinman, K. Hamilton, L.E. Kaye, K.E. Guillemin, M.J. Harms. "Lipopolysaccharide activates zebrafish TLR4". (*in prep*)

Loes, A.N., J.T. Bridgham, M.J. Harms. 2018. "Coevolution of the Toll-like receptor 4 complex with calgranulins and lipopolysaccharide". *Front. Immunol.* 9:304. DOI: 10.3389/fimmu.2018.00304

Moghaddam, A.D., J.D. White, R.M. Cunningham, **A.N. Loes**, M.M. Haley and V.J. DeRose. 2014. "Convenient detection of metal–DNA, metal–RNA, and metal–protein adducts with a click-modified Pt (II) complex". *Dal. Trans.* 44(8), 3536-3539. DOI: 10.1039/c4dt02649g.

Loes, A.N., L. Ruyle, M. Arvizu, K. E. Gresko, A. L. Wilson and C. E. Deutch. 2013. "Inhibition of urease activity in the urinary tract pathogen *Staphylococcus saprophyticus*". *Lett. Appl. Microbiol.* 58(1):31-41. DOI: 10.1111/lam

ACKNOWLEDGMENTS

I wish to express sincere appreciation to Dr. Michael J. Harms for being an accessible advisor and endlessly excitable by new ideas for scientific projects. I would also like to thank him and members of the Harms lab for their thoughtful insight and creative ideas for the presentation of the work included in this thesis. I thank Joseph Harman, who worked closely with me to create “team calprotectin” and Dr. Jeremy Anderson who brought a new perspective and direction to our work in this system. Both of whom were excellent co-workers and teammates and I look forward to seeing where they bring this project in the future. I am also endlessly grateful to Ran Shi, who brought excitement back to this project at a time when I needed that encouragement.

I am thankful to have been a part of the Department of Chemistry and Biochemistry at UO, particularly for the supportive environment of friendship and community created by the graduate students. I also wish to thank my committee members in particular Dr. Victoria DeRose for being an invaluable supportive mentor throughout the process of completing my dissertation and Karen Guillemin for encouraging me to pursue challenging projects and for being a wonderful collaborator for completing the zebrafish work described here. I also thank Jamie Bridgham for being an excellent mentor and for training me in cell culture without which none of this work would have been possible.

Lastly, I thank my family, especially my sisters – Jaspar, Kaelyn, Gina, and Janelle for their support over the last five years. I am also incredibly grateful for my partner, Dr. Matthew Kast, without his support and encouragement I would not have completed this degree.

TABLE OF CONTENTS

Chapter Page

LIST OF FIGURES

Figure Page

Figure Page

Figure Page

LIST OF TABLES

CHAPTER I

HISTORICAL AND FUNCTIONAL INSIGHTS INTO TOLL-LIKE RECEPTOR 4 ACTIVATION BY LIPOPOLYSACCHARIDE AND CALGRANULINS

This chapter contains co-authored material. Portions of this chapter were published as Loes, A.N., J.T. Bridgham, M.J. Harms. 2018. "Coevolution of the Toll-like receptor 4 complex with calgranulins and lipopolysaccharide". Front. Immunol. 9:304. This chapter was written by me, with editorial assistance from M.J. Harms

Toll-like Receptors, an Important Line of Defense for the Innate Immune System

The immune system must recognize when a pathogen is present. Cell surface receptors, known as pattern recognition receptors serve as the first line of defense for the innate immune system. These receptors recognize conserved molecules from pathogens[1]. The ability to recognize these pathogenic signals and rapidly respond is an essential part of protecting the host from infection.

The Toll-like receptor (TLR) family is a well-studied class of pattern recognition receptors[2]. In 1991, the first Toll was discovered in *Drosophila*[3]. The TLR family expanded across vertebrates. To date, 13 TLRs have been identified in mammals[4,5].

The diversity of TLRs found across the mammals allows for recognition of diverse types of ligands, known as pathogen-associated molecular patterns (PAMPs)[6]. TLRs have been shown to recognize chemically and structurally different types of

biological molecules including nucleic acids, proteins, and small molecules via similarly structured but significantly diverged, leucine-rich repeats (LRRs)[4,6–9].

While the ectodomains of TLRs have diversified over time to specifically recognize certain types of molecules, the internal domains and downstream signaling cascades have been conserved across mammalian Tolls[5,10]. The internal domain of TLRs, known as the Toll-interleukin (IL)-1 receptor (TIR) domain is highly conserved across vertebrates[5,10]. TIR domains recognize adaptor molecules including Myd88 and TRIF which trigger highly conserved downstream signaling cascades leading to activation of the transcription factor NF-kB[11]. NF-kB activation triggers expression of inflammatory cytokines[3]. This functions as an alarm signal, recruiting neutrophils to the site of infection and stimulating an inflammatory response to control the spread of infection.

Lipopolysaccharide Recognition by Toll-like Receptor 4 and Septic Shock

One of the first identified mammalian Toll-like receptors and the most well-studied is Toll-like receptor 4 (TLR4). TLR4 recognizes lipopolysaccharide (LPS), a component of the cell wall of Gram-negative bacteria. Through LPS-recognition, TLR4 provides a highly sensitive and tuned alarm for presence of bacteria in the bloodstream. This early response from TLR4 amplifies the signal of bacterial presence in hosts and is essential for control of infection.

The inflammatory response triggered by LPS through TLR4 is a critical gateway for initiation of septic shock[12–14]. Sepsis is a common cause of death in hospitals in the United States. The high mortality rate associated with sepsis is due to the intense activation and amplification of an inflammatory response through TLR4 by LPS[12,14].

Development of antagonists of TLR4 activation by LPS, or inhibitors of downstream response elements has been a clinical goal for many years[15–18].

Binding and activation of TLR4 by LPS is relatively well understood at the molecular level. Like many other TLRs, activation of TLR4 requires homodimerization following ligand binding[19]. TLR4 activation by LPS requires two cofactors MD-2 and CD14[20–28]. MD-2 functions like a hydrophobic pocket for LPS molecules(Fig. 1.1)[17,29–32]. MD-2 interacts with TLR4 monomers, even in the absence of ligand[21]. LPS binding by MD-2 induces a conformational change and allows for formation dimer with an additional TLR4/MD-2 molecule[19,31]. MD-2 comprises a significant portion of the dimerization interface and is essential for LPS binding and activation of TLR4[17,23,32]. LPS is often found in the extracellular space as LPS-aggregates[33,34]. In order for independent LPS molecules to bind to TLR4/MD-2, these aggregates need to be broken up by LPS-binding protein (LBP)[35,36]. CD14 works in concert with LBP to deliver LPS molecules to the TLR4/MD-2 complex[35,37]. CD14 plays an important role in LPS concentration for a more sensitive response to LPS by TLR4[32,38,39].

Endogenous Molecule Recognition by TLR4 and Chronic Inflammation

In addition to its role in septic shock, Toll-like receptor 4 also plays a role in sterile inflammation[13,18,40]. Stressed or damaged tissues release a variety of endogenous molecules, known as damage-associated molecular patterns (DAMPS), which trigger the activation of inflammatory pathways[9,41]. Many of these molecules have been shown to be recognized by TLRs, indicating a dual role for these receptors in pathogen-mediated and host-mediated inflammatory responses[41,42].

TLR4 has been shown to recognize a large diversity of endogenous ligands – proteins, lipoproteins, lipids[43–47]. In contrast to the LPS response, endogenous activation of TLR4 is not well-characterized at the molecular level. However, endogenous activation of TLR4 plays a role in many inflammatory diseases including atherosclerosis[48,49], liver disease[50,51], kidney disease[52], pulmonary disease[53,54], and inflammatory bowel disease (IBD)[55–60].

S100A9 can act as a DAMP, activating a potent pro-inflammatory response through a direct interaction with TLR4[61–63]. S100A9 is present in hosts as both a highly pro-inflammatory homodimer and as a part of the antimicrobial heterocomplex calprotectin (CP)[61,64,65]. During IBD, CP is found at high levels in the gut and is directly correlated with the degree of inflammation[66]. Expression of S100A9 is driven by LPS, and functions as an amplification of pathogen signaling through TLR4[64].

Understanding sterile inflammation and endogenous activation of TLR4 is important, not only for tempering inflammation in chronic diseases but also in treating septic shock. As sterile inflammation and LPS-driven inflammation are not fully separate, disruption of the chronic signal can be therapeutically effective in disrupting the positive feedback loop of endogenous molecule driven inflammation. Deletion of S100A9, for example, has been shown to have a protective effect for septic shock[67]. Therefore, S100A9 has been identified as a potential therapeutic target to control chronic inflammation[68]. Further work examining how endogenous and exogenous activation of TLR4 differ mechanistically may lead to therapeutic strategies for controlling chronic inflammation in patients while leaving pathogen-sensing intact.

The molecular mechanism for endogenous activation of TLR4 by S100A9 is not known. Identifying how S100A9 activation differs from LPS activation of TLR4 is critical to the establishment of therapeutic strategies for disrupting this powerful pro-inflammatory signal. Further, establishing when these two types of signaling regimes evolved for TLR4 is critical to identifying appropriate model organisms for studying the importance of these types of signaling in human disease.

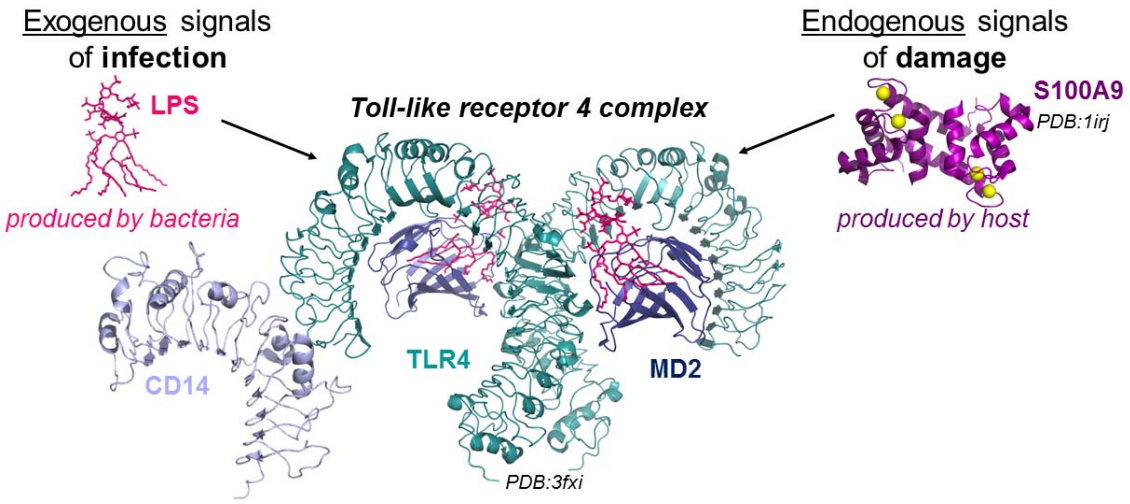

Figure 1.1. Toll-like receptor 4 activation by exogenous and endogenous danger signals Structure of TLR4 (turquoise) with co-factors MD-2 (navy) and CD14 (lilac) can be activated by LPS (pink) and S100A9 (violet).[69–71]

The primary aim of this work is to better understand the differences between endogenous activation of TLR4 by S100s and exogenous activation by LPS (Fig. 1.1). I use a combination of biochemical and evolutionary approaches to gain historical and functional insight into these two types of pro-inflammatory signals. With evolutionary biochemistry, we were able to discern differences across extant species and illuminate the core features for activation of this complex.

Molecular Evolution as a Powerful Lens for Biochemistry

A primary goal of biochemistry is understanding the physical basis for the function of biological molecules. Traditional biochemical methods primarily employ targeted analyses towards the identification of mutations which disrupt the function of interest[72,73]. A key drawback to this approach is that there are many types of perturbations that can disrupt a function but are not mechanistically informative. For complex molecular traits, it is often challenging to puzzle out the molecular and physical basis for a function from a non-functional mutant[74].

A complementary approach is to place the interaction in the context of its evolutionary history. Evolutionary biochemistry combines history and molecular mechanisms to reveal not only how proteins work, but also the process by which they came to work that way[75]. Two powerful tools for dissecting the evolution of protein function are phylogenetics and ancestral state reconstruction[76–80]. These approaches allow one to ask how changes in amino acid sequence, over the course of evolution, led to a change in protein function. This offers a distinct advantage over traditional mutagenesis approaches; rather than assessing how to break a function of interest, evolutionary approaches give one a blueprint to build an interesting function onto a historical protein framework[81]. These power and impact of these approaches will be discussed in more detail in the last chapter of this work.

Advantages of an Evolutionary Approach in Immunology

Evolutionary biochemistry has powerful advantages in studies of immune proteins. First, animal models are essential to pharmaceutical development. Yet, studies

in alternate organisms are not always directly applicable to human disease[82]. These challenges are most pronounced in infectious disease research, where the host-pathogen arms race has led to coevolution with host immune systems which result in differences even across relatively short evolutionary timescales.

Mice are a favored model organism for understanding human health. Mice and humans shared a common ancestor approximately 90 million years ago. This shared evolutionary history accounts for a great deal of the similarity in immune systems. However, many changes have occurred along the respective primate and rodent lineage since the common ancestor which results in a variety of differences, leading to difficulties in the ability to translate results in mice to human systems. It is essential, therefore that evolutionary history is considered when utilizing animal models. This allows one to separate the changes that have occurred along specific lineages from the core molecular basis for immune protein function.

An additional advantage of evolutionary biochemistry approaches for immunology stems from the relatively rapid evolution that has occurred within this system. Homology is often used to assign functions across related organisms. However, particularly within rapidly evolving systems, conservation in function at the macroscopic level is not always indicative of conservation in mechanism at the molecular level[83]. Organisms may contain the same protein performing the same function but given differences in the molecular basis of interactions, which arises from coevolution between complex components, identification of a disruptive mutation in one organism may have limited if any effect in another. This can be particularly challenging for drug development. Examination of compatibility of complex components across species can

alert one to lineage-specific coevolution and therefore potential challenges in translational studies.

Development of inhibitors of TLR4 has been hampered in part by differential response to potential inhibitory molecules between humans and preclinical animal models[84]. Several molecules function as antagonists in humans but agonists in mice[85–87]. Further, drugs have been developed to work in murine models that have been found not to function similarly in humans[88,89]. Given the significant cost of development of therapeutics and clinical testing in animal models and human subjects, a better understanding of the differences and similarities between human and animal models could provide a framework to begin development of potent and more broadly functional inhibitors of this pathway.

Evolutionary Biochemistry for Understanding TLR4 Activation by LPS and Calgranulins

Here, I describe the application of a combination of evolutionary and biochemical techniques to contribute to the understanding of LPS and calgranulin activation of TLR4. To take apart the evolution of TLR4 activation by both endogenous and exogenous ligands, my colleagues and I determined the species distribution of TLR4 activation by LPS and calgranulins. These studies contributed to the overall goal of isolating the evolutionary intervals in which these functions evolved. Next, we used a direct biochemical approach to understand the role for zinc in the interaction between S100A9 and TLR4. This work further contributes to our understanding of the molecular mechanism of S100A9 activation of TLR4. Finally, we used phylogenetics and ancestral

state reconstruction for calgranulins, TLR4, MD-2 and CD14 to generate hypothesis regarding the biochemical basis for functions exhibited by these proteins in extant species. These studies together demonstrate the power of combining an understanding of evolutionary history with biochemical analysis to break apart molecular mechanisms for interesting protein functions.

Chapter II describes our work investigating when LPS activation of TLR4 evolved. LPS recognition by TLR4 across the amniotes has been shown to be dependent on the presence of cofactors MD-2 and CD14[90–92]. The lack of identification of these cofactors in amphibians and fish has led some to conclude that LPS activation of TLR4 does not occur in fish[93,94]. However, we identified a putative zebrafish MD-2. Phylogenetic analysis and synteny support classification of this gene as an MD-2. Using a cell culture assay we show that expression of this protein with zebrafish TLR4a allows zebrafish TLR4 to activate an NF-kB cascade in response to LPS. By reporting the identification of MD-2 and CD14 in amphibians and fish, we show that this function of TLR4 evolved much earlier than previously thought. This also implies that fish may be a good model organism for understanding septic shock.

Chapter III examines when S100 activation of TLR4 evolved. TLR4 induces inflammation in response to both pathogen- and host-derived molecules. Lipopolysaccharide (LPS) recognition by TLR4 has been shown to occur across vertebrates, but endogenous signaling through TLR4 has not been validated outside of placental mammals. To determine whether endogenous danger signaling is shared across amniotes, we studied the evolution of TLR4-activation by the calgranulin proteins (S100A8, S100A9, and S100A12), a clade of host molecules that potently activate TLR4

in placental mammals. We performed phylogenetic and syntenic analysis and found MRP-126 — a gene in birds and reptiles — is likely orthologous to the mammalian calgranulins. We then used an *ex vivo* TLR4 activation assay to establish that calgranulin pro-inflammatory activity is not specific to placental mammals but is also exhibited by representative marsupial and sauropsid species. This activity is strongly dependent on the cofactors CD14 and MD2 for all species studied, suggesting a conserved mode of activation across the amniotes. Ortholog complementation experiments between the calgranulins, TLR4, CD14, and MD2 revealed extensive lineage specific-coevolution and multi-way interactions between components that are necessary for the activation of NF-κB signaling by calgranulins and LPS. Our work demonstrates that calgranulin activation of TLR4 evolved at least ~320 million years ago and has been conserved in the amniote innate immune system.

Chapter IV examines how S100A9 activates TLR4. S100A9 promotes a pro-inflammatory response through a direct interaction with Toll-like receptor 4 (TLR4). While the mechanism for LPS recognition by TLR4 is relatively well understood, the molecular requirements and mechanism for activation by S100 proteins is not known. The interaction between S100A9 and TLR4 has previously been shown to be regulated by zinc[68]. We find that the tail, which is predicted to undergo significant structural rearrangement upon zinc binding, is not directly involved in the interaction of S100A9 with TLR4. We further demonstrate that S100A9 is prone to reversible, zinc-dependent aggregation and show that this aggregation is dependent on the presence of metal binding residues in the canonical S100-transition metal binding site. However, we find that this

transition metal binding site is not necessary for the interaction with TLR4 and that activation of TLR4 by S100A9 occurs even in the absence of metal binding.

Finally, **Chapter V** demonstrates the power of utilizing phylogenetics and ancestral state reconstruction to gain insight into protein function. Phylogenetics on TLR4, MD-2, CD14, and the calgranulin proteins reveal some of the many different types of hypotheses evolutionary strategies generate. We show that by using ancestral state reconstruction we were able to polarize when a transition occurred in S100 proteins which led to proteolytic susceptibility. We also show how consideration for distribution of a function extant species can lead to identification of the historical substitutions which were important for the evolution of given functions. In addition, we show that co-evolutionary analysis can lead to identification of the core conserved molecular requirements for a function of interest and lead to hypotheses regarding how coevolution of complex components can lead to incompatibility between extant molecules. Finally, we show that phylogenetics can lead to identifications of the hierarchy of functions for a multi-functional molecule. These studies, while in preliminary stages, strongly demonstrate how an evolutionary perspective can lead to molecular level hypotheses that would not have been accessible with traditional molecular approaches.

CHAPTER II

LIPOPOLYSACCHARIDE ACTIVATES ZEBRAFISH TOLL-LIKE RECEPTOR 4

Portions of this chapter are co-authored. Experiments described here were conducted by me. This excerpt was written by me in collaboration with M.J. Harms.

INTRODUCTION

The zebrafish (*Danio rerio*) is a powerful model organism for studies of vertebrate innate immunity[95]. It has mature genetic resources, rapid generation time, a clear embryo, and facile germ-free derivation[96,97]. This makes it extremely useful as a complement to other model organisms for mechanistic studies of innate immune function.

A major difficulty for relating zebrafish and mammalian innate immunity is an apparent difference in how they respond to lipopolysaccharide (LPS). LPS is a component of the cell wall of Gram-negative bacteria that, in amniotes, plays a critical role in initiating inflammation in response to infection. LPS is the primary driver of septic shock and plays roles in Crohn's disease, inflammatory bowel disease, and other inflammatory disorders[49,51,54,56,57,98]. Across amniotes, LPS activates NF-kB signaling via Toll-like receptor 4 (TLR4)[90,91]. Peripheral proteins, including lipid binding protein (LBP) and cluster of differentiation 14 (CD14) bring LPS to TLR4[38,99,100]. They then load LPS into the essential cofactor MD-2 (also known as LY96), which forms a tight complex with TLR4[21]. Binding of LPS triggers dimerization of TLR4 and activation of NF-kB, largely through a Myd88-dependent pathway[17,101] (Fig. 2.1A).

Despite the central role of TLR4/MD-2 in LPS recognition in amniotes, it is currently believed that zebrafish do not respond to LPS via TLR4[93,94]. This makes it difficult to relate studies of inflammation in zebrafish back to mammals. It is also rather puzzling. Zebrafish possess three TLR4 genes (*tlr4ba*, *tlr4a1*, and *tlr4bb*), at least one of which is widely expressed (*tlr4ba*)[10]. Further, the downstream responses to LPS are similar between mammals and zebrafish. Like in amniotes, LPS activates NF-kB via a Myd88-dependent pathway in zebrafish[102,103]. Likewise, the expression patterns of genes induced by LPS stimulation are highly similar between mouse and zebrafish[104]. Why would the downstream LPS response in zebrafish be consonant with the LPS response in mammals if they use entirely different initial receptors?

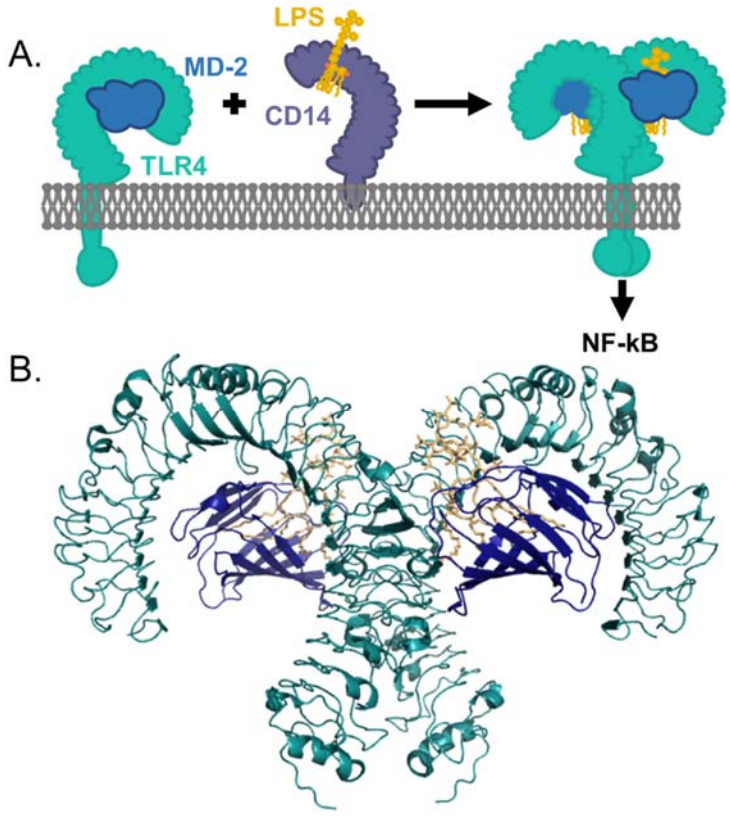

Figure 2.1. Mammalian TLR4 activation involves cofactors MD-2 and CD14. A. Schematic representation of LPS transfer from CD14 to TLR4-MD-2 complex. B. Human TLR4 (cyan) in complex with MD-2 (navy) with LPS (yellow) bound PDB: 3fxi.[71]

We set out to reevaluate the claim that zebrafish do not use TLR4 for LPS recognition. The current view is based on three lines of evidence. The most critical is the apparent lack of MD-2 in zebrafish and other bony fishes[93,94,105]. In amniotes, MD-2 is an essential TLR4 co-factor, as it forms the binding site at which LPS binds (Fig. 2.1B)[21]. Although zebrafish possess three different TLR4 genes, without MD-2 these cannot respond to LPS in the same manner as amniote TLR4[93,94]. The second line of evidence is that zebrafish TLR4s do not activate NF-kB in response to LPS in *ex vivo* assays, even when complemented with a mouse or human MD-2[93,94]. This has been interpreted as indicating that these receptors would not be able to respond to LPS, even if MD-2 was present.

Finally, it has been argued that the zebrafish "TLR4" genes are paralogs rather than orthologs to amniote TLR4[94]. In this evolutionary scenario, an ancestral Toll-like receptor gene duplicated, allowing subfunctionalization and/or neofunctionalization of each duplicate gene. After this event, tetrapods and bony fishes diverged. Finally, tetrapods lost one gene and retained the gene we now call TLR4; bony fishes, in contrast, lost TLR4 and retained the other gene. In this scenario, we would expect the "TLR4" proteins from bony fishes to behave differently than TLR4 from tetrapods, as the two genes had an opportunity to functionally diverge prior to the speciation of tetrapods and bony fishes.

We set out to carefully evaluate each of these lines of evidence using resources unavailable when the initial investigations of zebrafish TLR4 were performed. Using careful bioinformatics, we found an ortholog to mammalian MD-2 in multiple bony fish species. In an *ex vivo* functional assay, zebrafish MD-2 and *tlr4ba* activate NF-kB

signaling in response to LPS. Finally, we re-examine the evolutionary history of TLR4 in zebrafish. Both model-based phylogenetics and the genomic location of TLR4 in the gar (*Lepisosteus oculatus*) genome support assigning the TLR4 genes as co-orthologs of tetrapod TLR4.

Our results suggest that zebrafish TLR4 responds to LPS in the same manner as tetrapod TLR4s. This means that the LPS-induced activation of the TLR4/MD-2 complex evolved 110 million years earlier than previously appreciated. Further, this implies that studies of LPS-induced inflammation in zebrafish can be directly mapped to mammals.

RESULTS

Identification of an MD-2 like protein in amphibians and fishes

Our first question was whether we could find MD-2 in bony fishes. We used the human MD-2 protein sequence to BLAST against the zebrafish genome and transcriptomes. This returned no significant hits. We, therefore, took a more phylogenetically informed strategy.

Relative to humans, the earliest branching functionally characterized TLR4/MD-2 complex is in chickens. We, therefore "walked out" from amniotes towards fishes. We started with amphibians. We BLASTed human MD-2 against the *Xenopus laevis* genome. This revealed a hit to a hypothetical protein with 30% identity (OCT74818.1). When reverse-BLASTed against the human proteome, this hit returns MD-2. To validate the sequence, we compared it to a collection of functionally characterized MD-2 proteins from amniotes. We found that the *X. laevis* gene appeared to be N-terminally truncated. Using XenBase, we identified the full-length transcript in the transcriptome for *X. laevis*.

By BLASTing against available amphibian transcriptomes[106–109], we further identified putative MD-2 proteins in *Rhinella marina, and Odorrana margaretae* (Fig. AA1).

With these putative amphibian MD-2 sequences in hand, we returned to our search for a zebrafish MD-2. A BLAST against a zebrafish transcriptome using the *X. laevis* sequence revealed a likely transcript (23% identity). We then searched additional fish transcriptomes available from the Fish1TK project and identified a set of transcripts from three species that matched MD-2 (Fig. AA1). The putative MD-2 genes we identified in bony fishes were highly diverged. On average, they exhibited only 26% identity against human MD-2, and only ~40% identity relative to one another.

We next set out to assign the orthology of these putative MD-2 sequences. Our primary concern was that these newly identified sequences were paralogs of MD-2—likely the closely related protein, MD-1. We, therefore, built a phylogenetic tree to elucidate whether these newly identified sequences grouped with MD-1 or MD-2. We constructed an alignment of 294 MD-1 and MD-2 protein sequences sampled from amniotes, amphibians, and fishes and then used this to infer a maximum likelihood phylogeny (Fig. 2.2A).

The putative amphibian and bony fish MD-2 sequences grouped with the tetrapod MD-2 sequences with strong support (SH = 0.99). The MD-1/MD-2 protein tree largely reproduced the species tree, with the exception of amphibians. On the MD-1 lineage, amphibians form a polytomy with fishes at the base of the tree; on the MD-2 lineage, they are placed inside the amniote clade with a relatively short internal branch. This is likely an artifact of the small number of amphibian sequences available, as well as the rapid evolution of the family.

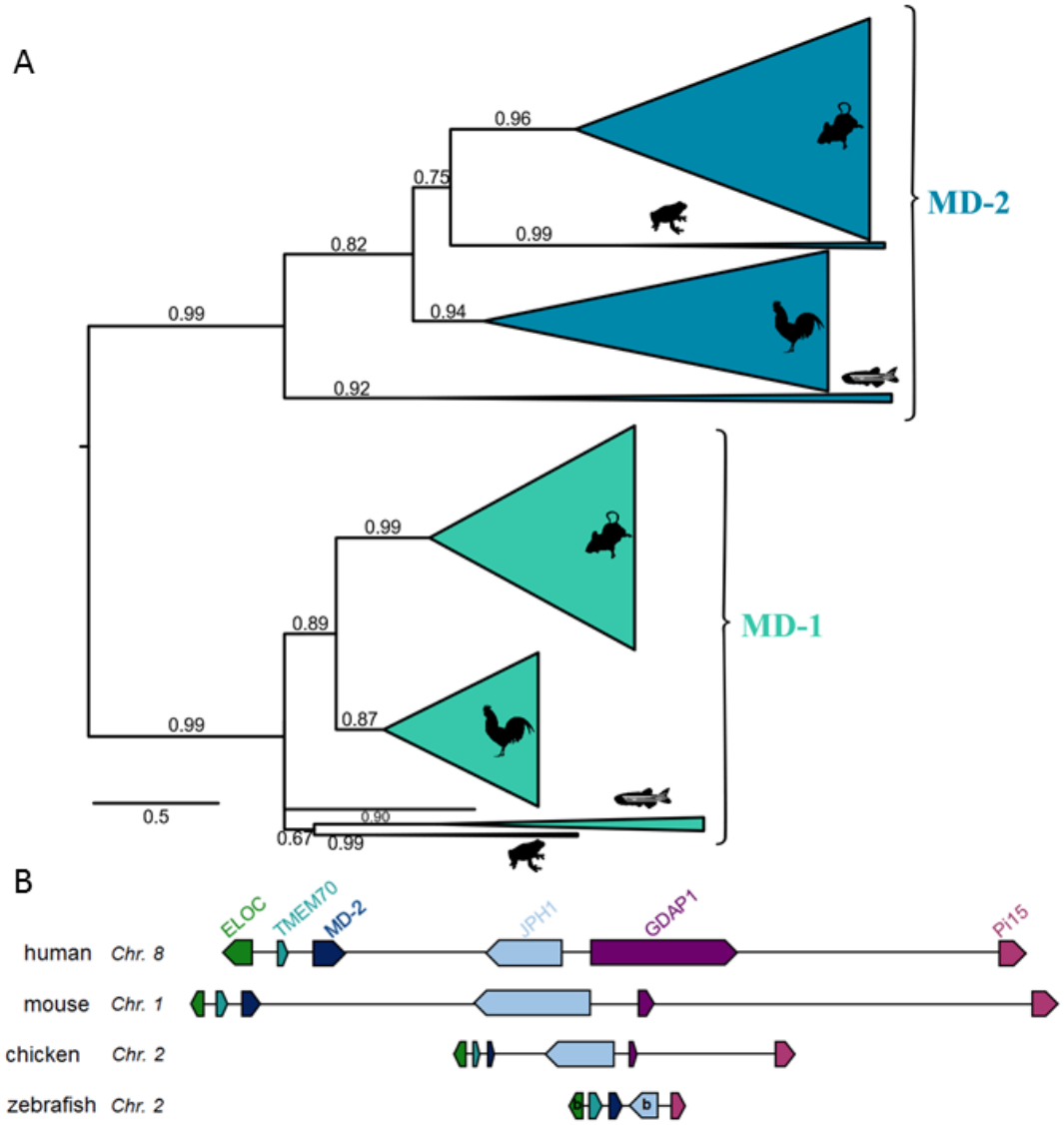

Figure 2.2. Phylogeny and synteny for identified protein in zebrafish support classification as an MD-2 A) Maximum likelihood phylogeny of MD-2 and MD-1 proteins. Wedges are collapsed clades of orthologs, with wedge height corresponding to the number of included taxa and wedge length indicating the longest branch length with the clade. Support values are SH-supports calculated using an approximate likelihood ratio test. Clades are colored to highlight MD-2 and MD-1 classification. The taxa included in each clade are noted on the tree by silhouettes of mammals (mouse), sauropsids (chicken), amphibians (frog), and fish (zebrafish). B) Genomic organization of genes surrounding MD-2 in vertebrates. Arrows for genes represent the coding strand. Approximate distances between genes are represented by the length of line for the selected chromosome.

Overall, the tree is consistent with a single gene duplication event sometime before the evolution of bony vertebrates. MD-1 and MD-2 were then preserved along most descendant lineages. This said, the protein sequences of MD-1 and, particularly, MD-2 are evolving rapidly. The total branch lengths between the last common ancestor of MD-2 to its human and zebrafish descendants are 2 and 2.44, respectively. Put another way, on each lineage, each position in the MD-2 sequence has changed its amino acid ~2-2.5 times over the last 430 million years. Only 7 of ~160 positions are in MD-2 are universally conserved across the clade.

To cross-validate the orthology of our newly identified *md2* genes, we next examined their location in the *X. laevis* and *D. rerio* genomes. We found that the synteny is consistent with *md2*s from other vertebrates (Fig. 2.2B). In five genomes sampled from across bony vertebrates—including *X. laevis* and *D rerio—md2* is located between *tmem70* and *jph1b* (Table AA1). This provides strong evidence that the amphibian and fish MD-2 genes are, in fact, orthologous to amniote MD-2.

Due to the genome duplication event that occurred along the zebrafish lineage, we also looked for a second copy of *md2*. We examined the genomic location of the *jph1a* paralog, but we were unable to identify an additional gene with any similarity to *md2*. It appears that an inversion may have occurred in this region, complicating identification by synteny alone. This said, no additional transcripts were identified within the zebrafish transcriptome with similarity to the identified zebrafish *md2* sequence. This is consistent with a loss of the duplicate copy of this gene.

Zebrafish TLR4a/MD-2 can activate NF-kB in response to lipopolysaccharide

Our phylogenetic and syntenic analysis are consistent with zebrafish MD-2 gene being homologous to amniote MD-2; however, this does not demonstrate that the protein is playing the same role played by amniote MD-2. Particularly given the low similarity between zebrafish MD-2 and its amniote orthologs, it was not clear that the zebrafish MD-2 would be capable of mediating the TLR4 response to MD-2. We, therefore, turned to an *ex vivo* cell culture assay to assess the ability of the zebrafish MD-2 to partner with zebrafish TLR4a and TLR4b for LPS activation. In this assay, we use luciferase to quantify NF-kB output in response to exogenously applied LPS.

Previous experiments found that neither TLR4a nor TLR4b activated in response to LPS when complemented with either human or mouse MD-2. One key difference between our experiments and those done previously is the sequence of TLR4a used. Previous investigators used a construct that was ~75 amino acids shorter than tetrapod TLR4s. This construct is missing both the signal peptide required to target TLR4 to the cell surface and a region of the protein that is likely critical for MD-2 binding (Fig. AA2). In contrast, we used a full-length ORF (XM_009307228.3). We started by co-transfecting cells with zebrafish MD-2 and zebrafish TLR4a or TLR4b and measuring NF-kB activation in response to exogenously applied LPS. We saw no activation (Fig. 2.3).

In these experiments, we are attempting to activate a TLR4/MD-2 complex without CD14—an important peripheral protein that brings LPS to TLR4/MD-2 complexes in amniotes, dramatically increasing the NF-kB response. We thus co-transfected TLR4a or TLR4b, zebrafish MD-2, and human CD14. In this context, we

observed potent activation of NF-kB by the TLR4a/MD-2 complex (Fig. 2.3). To verify that this depended on MD-2, rather than CD14, we tested the activation of TLR4 and CD14 without MD-2—this complex did not respond to LPS (Fig. 2.3).

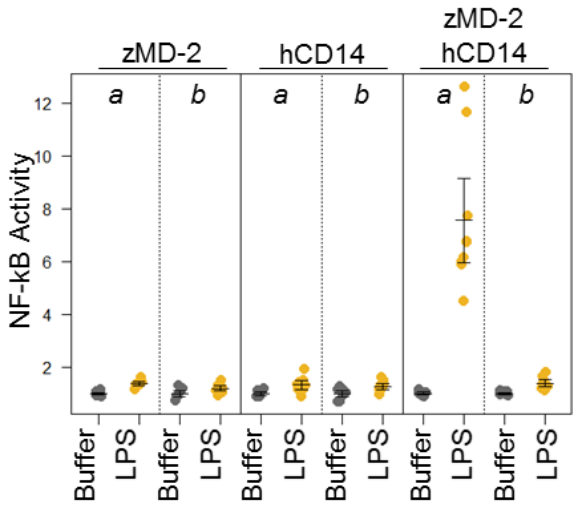

Figure 2.3. LPS activates zebrafish TLR4a with zebrafish MD-2 and human CD14. Activation of zebrafish TLR4a (*a*) and TLR4b (*b*) with 400 ng/mL LPS in the presence and absence of zebrafish MD-2 and human CD14. Points are the technical replicates from three biological replicates. Bold lines are the mean of the biological replicates. Error bars are a standard error on the mean of the biological replicates.

We also verified that the zebrafish TLR4a/MD-2 complex, complemented with human CD14, exhibited a dose-dependent response to LPS. This was the case (Fig. AA3). The concentrations of LPS needed for activation of the zebrafish TLR4/MD-2 complex was much higher than those needed for activation of the human proteins in these cells. However, we saw activation patterns comparable to that observed for the turtle LPS in the human background with matched MD2 and CD14[91].

Amniote CD14 and MD-2 allow LPS-induced activation of zebrafish TLR4a

Our results strongly support the hypothesis that zebrafish TLR4a/MD-2 can activate in response to LPS; however, this could only be done with the presence of a

supporting, mammalian protein (human CD14). To determine if this was an artifact of the human protein, we tested the LPS activation of TLR4a/MD-2 in the presence of human, mouse, opossum, and chicken CD14. We found that all but the chicken CD14 were able to support the activation of the complex (Fig. 2.4A). This suggests the activity induced by CD14 is a general property of CD14 proteins.

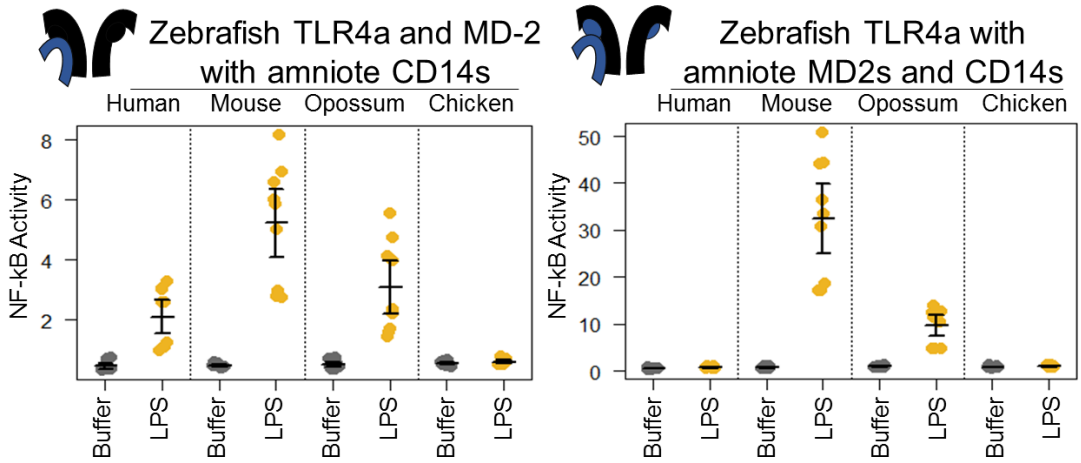

Figure 2.4. Shared molecular mechanism for cofactors CD14 and MD-2s across vertebrates A) Cross-reactivity of amniote CD14s with the zebrafish TLR4a/MD-2 complex, B) Cross-reactivity of amniote MD-2s with zebrafish TLR4a C) Conserved binding interface between MD-2 and N-terminus of TLR4 for vertebrates. Points are the technical replicates from three biological replicates. Bold lines are the mean of the biological replicates. Error bars are a standard error on the mean of the biological replicates.

We also looked for evidence of a zebrafish CD14 gene; however, we were unable to locate such a gene. The closest paralog to CD14 in zebrafish is TLR2. We therefore co-transfected zebrafish TLR4a/MD-2 with TLR2. This complex, however, did not activate in response to LPS (Fig. AA4). The simplest explanation for this failed activation is that TLR2 cannot perform the same role as CD14; however, this could also reflect inappropriate transfection conditions, a problem with the heterologous cell line (these

experiments were done in human cells), or a missing secondary cofactor (such as a fish LPS binding protein).

Finally, to see if zebrafish TLR4a was behaving similarly to amniote TLR4, we investigated whether MD-2 from other species to act in concert with zebrafish TLR4a. We co-transfected TLR4a in the presence of human, mouse, and opossum MD-2. We saw complementation by both mouse and opossum MD-2 for LPS activation of zebrafish TLR4a (Fig. 2.4B). This suggests that the requirements for activation by the TLR4/MD-2 complex have been conserved across vertebrates.

Zebrafish TLR4 genes are likely co-orthologous to tetrapod TLR4

The evidence supporting paralogy from previous phylogenetic analyses was mixed: a gene-tree supported orthology, while a syntenic analysis supported paralogy. We, therefore, revisited the question of orthology using genomic tools that were not available when the first analysis was performed.

As with our analysis of MD-2, we started with a phylogenetic tree and then turned to synteny. For the phylogenetic tree, we constructed a multiple sequence alignment containing 263 TLR4 sequences and 190 RP105 sequences (also known as CD180) as an outgroup. In the resulting maximum likelihood tree, TLR4 and RP105 form distinct, well-supported clades (Fig. 2.5A). Within the TLR4 clade, zebrafish TLR4a and TLR4b form a monophyletic group with other teleost fish TLR4s. The most straightforward interpretation of this result is TLR4a and TLR4b are co-orthologous to tetrapod TLR4s, and that the duplication of the *tlr4* gene in zebrafish represent lineage-specific duplication in bony fishes.

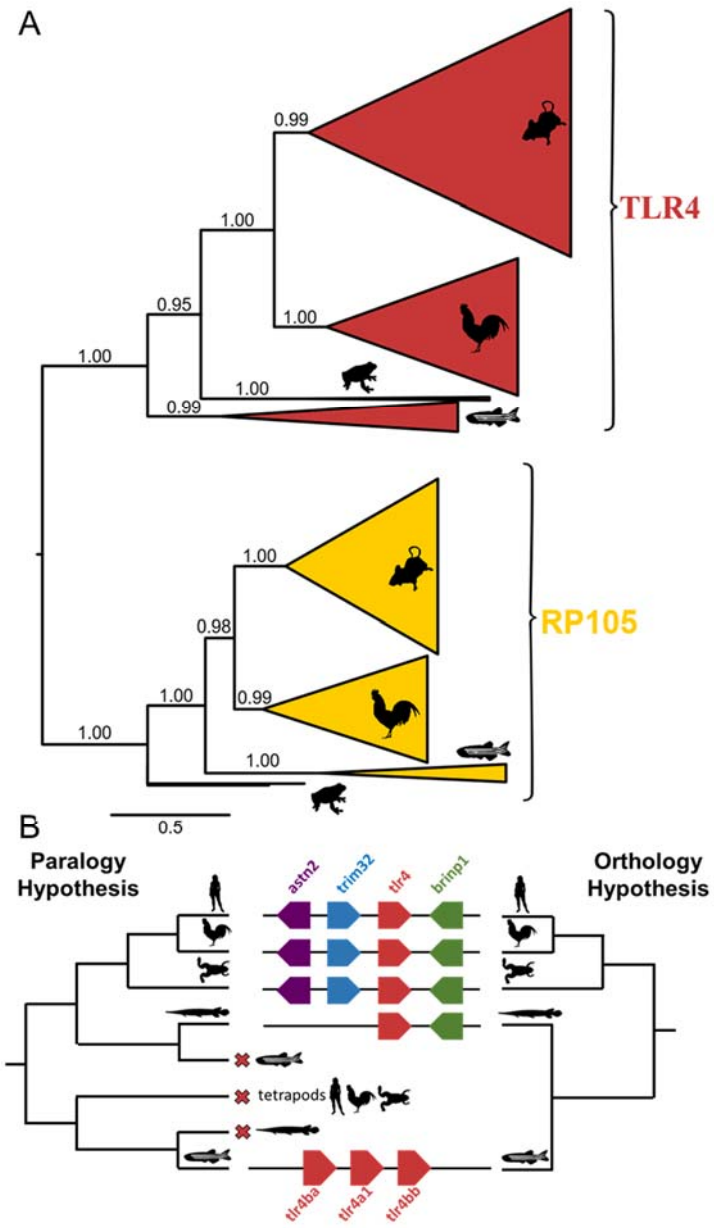

Figure 2.5. Phylogeny supports the classification of zebrafish TLR4 genes as orthologs despite genomic relocation. A) Maximum likelihood phylogeny of TLR4 and RP105. Wedges are collapsed clades of orthologs, with wedge height corresponding to the number of included taxa and wedge length indicating the longest branch length with the clade. Support values are SH-supports calculated using an approximate likelihood ratio test. Clades are colored to highlight TLR4 and RP105 classification. The taxa included in each clade are noted on the tree by silhouettes of mammals (mouse), sauropsids (chicken), amphibians (frog), and fish (zebrafish). B) Schematic representation of alternate interpretations of the synteny observations for fish TLR4. Synteny for TLR4 is conserved in amniotes, at least partially in amphibians, and minimally in gar. While the orthology hypothesis requires a single genomic relocation event, the paralogy hypothesis requires a number of independent losses along vertebrate lineages.

We next revisited the evidence from synteny. The zebrafish TLR4 genes are collinear on chromosome 13, suggesting that the three copies arose by tandem gene duplication (Fig. 2.5B). This is supported by our ML tree, in which TLR4a and TLR4b are sister to one another, but inside the genes from pike and gar (Fig. 2.5A).

The primary question is whether the gene that eventually became *tlr4ba, tlr4a1,* and *tlr4bb* duplicated before or after the speciation of bony fishes and tetrapods. The gar (*Lepisosteus oculatus*) genome provides an excellent tool with which to ask this question, as the gar is an early-branching bony fish. In tetrapods, TLR4 is consistently flanked by the genes *trim32* and *brinp1* (Table AA2). The gar *tlr4* gene, found on linkage group 21 is upstream of *brinp1*, like tetrapod *tlr4*. Further, many additional genes on linkage group 21 in gar are homologous to genes on chromosome 9 in humans, where human *tlr4* is located. The co-occurrence of these two genes is strong evidence that the gar TLR4 gene is orthologous to the tetrapod TLR4 (Fig. 2.5B).

Fig. 2.5B shows two trees, corresponding to the two evolutionary scenarios under consideration. In the *orthology* tree, the *tlr4* gene moved its location after the speciation of gar and zebrafish. This could have occurred, for example, via gene duplication, followed by loss of the gene in the original location. All of these gene copies are, however, co-orthologous to the tetrapod *tlr4*. In the *paralogy* tree, a gene duplication even preceded the speciation event separating tetrapods from bony fishes. This was then followed by three independent losses: a loss of the tetrapod *tlr4* on the bony fish lineage (after speciation of the gar), a loss of the tetrapod *tlr4* on the gar lineage, and a loss of the bony fish *tlr4* on the tetrapod lineage.

Based on this analysis, it is more parsimonious to place the zebrafish genes as co-orthologs of tetrapod *tlr4* than to posit an ancestral gene duplication event and differential loss along the tetrapod and teleost lineages. This is also consonant with the maximum likelihood phylogeny (Fig. 2.5A). Thus, although the genes surrounding the zebrafish *tlr4* genes are different than those surrounding the gar and tetrapod *tlr4* genes, the simplest interpretation of the available evidence is that the zebrafish *tlr4* are orthologs—not paralogs—to tetrapod *tlr4*.

DISCUSSION

This study provides the first evidence for LPS recognition by TLR4 outside of amniotes. We show that zebrafish TLR4a is capable of LPS sensing just as has been observed for mammalian TLR4. This is contrary to previous work which concluded a lack of LPS sensing by zebrafish TLR4 and therefore an alternate mechanism for LPS sensing in fish.

We report the identification of MD-2 in amphibians and fish. The lack of TLR4 in some fish species, the fact that MD-2, despite being only ~160 amino acids, contains 5 exons and the significant amount of divergence may have contributed to the difficulty others had previously in identifying a putative MD-2 transcript. Given the minimal conservation across the vertebrate MD-2s, examination of shared features will likely be useful in identifying the minimal molecular framework for MD-2 function.

Phylogenetic analysis and functional conservation support the classification of zebrafish TLR4 as orthologous to mammalian TLR4s. Previous studies have suggested that fish TLR4 may be paralogous rather than orthologous to mammalian TLR4. The lack

of tlr4 in some fish species[110] and the lack of conserved synteny were consistent with a paralogy hypothesis[94]. The paralogy hypothesis led many to consider the possibility that TLR4 may perform an alternate function in fish. However, the identification of MD-2, an essential adaptor protein for TLR4 function in mammals, led us to revisit this idea. Given the minimal syntenic conservation in gar and the functional conservation of zebrafish TLR4, we propose that zebrafish TLR4 is orthologous to mammalian TLR4 and that there was a translocation event in the teleost lineages and a subsequent loss of TLR4 in some fish species.

There are several known examples of ligands and inhibitors of the TLR4/MD-2 complex in mice that behave differently in humans[88,111,112]. Our previous work examining the cross-reactivity of TLR4/MD-2 complexes across the amniotes seemed to suggest significant lineage-specific coevolution of these proteins across amniotes; but, the ability of mouse and opossum MD-2 to complement zebrafish TLR4a indicates that coevolution between TLR4 and MD-2 is not a general trend but instead occurred along specific lineages. The differences in cross-specificity of the human proteins may parallel unique agonist/antagonist profiles that have previously been observed for human and mouse proteins. Further examination of the molecular basis for coevolution at the TLR4/MD-2 interface in humans may be informative in dissecting the mechanistic basis of species-specific activation.

The supporting players for LPS recognition in fish, outside of the TLR4/MD-2 complex, are not known. We were unable to identify a CD14-like molecule in fish. In mammals, CD14 works in tandem with lipid binding protein (LBP) to deliver LPS to the TLR4/MD-2 complex, thereby increasing sensitivity. LBP has also not yet been

identified in fish, the closest protein appears to be a homolog for mammalian bactericidal permeability-increasing protein (BPI)[99]. BPI shuttles LPS to high-density lipoproteins instead of to the TLR4/MD-2 complex[36,113,114]. Fish have previously been shown to be resistant to septic shock[110,115]. High concentrations of LPS are needed to activate teleost leukocytes[105,116–118]. The lack of CD14 and LBP, which increase responsiveness to LPS, could explain the lower sensitivity to LPS observed in teleost species.

These data provide a functional test for zebrafish MD-2 *ex vivo*, the next logical step is to examine if zebrafish MD-2 is important for LPS sensitivity *in vivo*. This work is ongoing, members of the Guillemin lab have generated a CRISPR knockout of zebrafish MD-2 and are currently comparing LPS sensitivity of this line to wild-type fish. This work will reveal if MD-2 plays the same role in septic shock in zebrafish as it does in mammals.

Zebrafish are popular as model organisms for immunological studies. The data presented here show that the TLR4/MD-2 complex is functionally conserved in early diverging vertebrates. We have shown that zebrafish can respond to LPS through a similar mechanism as observed in mammals. Further, the cross-reactivity of mammalian MD-2s and CD14s indicate that despite low sequence conservation in these proteins in vertebrates the molecular mechanism of activation also appears to be well conserved. Zebrafish may provide a powerful platform for studying TLR4-dependent inflammatory processes and septic shock in the future.

MATERIALS & METHODS

Bioinformatic analysis

We constructed curated databases of TLR4, RP105, MD-2 and MD-1 across the vertebrates. We obtained amino acids sequences of these proteins from NCBI, Ensembl, Fish1TK, amphibian transcriptomes[106–109], UniProt, and ZFIN. We constructed a multiple sequence alignment for TLR4 and RP105 and for MD-2 and MD-1 using MSAPROBS[119], followed by manual editing in MEGA[120]. We trimmed the alignment to remove highly variable (and therefore unalignable) regions. We used PHYML[121,122] with subtree pruning and re-grafting to construct the ML phylogeny. Pilot analyses revealed that the JTT substitution model with 8 rate categories and a floating gamma distribution parameter yielded the highest likelihood trees[123–125]. An AIC test was used to control for overfitting[126]. We rooted our trees at the duplication of these proteins in early vertebrates.

Cell culture and Transfection Conditions

Mammalian expression vectors containing human TLR4 and mouse TLR4 were obtained from Addgene (#13085 and #13086), originally deposited by Ruslan Medzhitov. Human CD14 and ELAM-Luc were also obtained from Addgene (#13645 and #13029) originally deposited by Doug Golenbock. Human MD-2 was obtained from the DNASU Repository (HsCD00439889) and contains a C-terminal V5-tag. Mouse MD-2 (UniProt #Q9JHF9) and CD14 (UniProt #P10810), as well as opossum MD-2 (UniProt #F6QBE6), CD14 (NCBI Accession #XP_007473804.1) and chicken MD-2 (UniProt #A0A1D5NZX9), and CD14 (UniProt #B0BL87) were designed to be free of restriction sites and codon optimized for human expression and purchased as mammalian expression vector constructs in pcDNA3.1(+) from Genscript (New Jersey, USA). Zebrafish TLR4a and MD-2 were also obtained from Genscript in pcDNA3.1(+) but were not codon

optimized. Zebrafish TLR4b was a gift from Carol Kim, we re-cloned this protein from its original vector into pcDNA3.1(+) to limit variability in expression due to differences in vector size and promoter.

Human embryonic kidney cells (HEK293T/17, ATCC CRL-11268) were maintained up to 30 passages in DMEM supplemented with 10% FBS at 37º C with 5% CO_2. For each transfection, a confluent 100 mm plate of HEK293T/17 cells was treated at room temperature with 0.25% Trypsin-EDTA in HBSS and resuspended with an addition of DMEM + 10% FBS. This was diluted 4-fold into fresh medium and 135 µL aliquots of resuspended cells were transferred to a 96-well cell culture treated plate. Transfection mixes were made with 10 ng of TLR4, 1 ng of CD14, 10 ng of MD-2, 1 ng of Renilla, 20 ng of ELAM-Luc, and 58 ng pcDNA3.1(+) per well for a total of 100 ng of DNA, diluted in OptiMEM to a volume of 10 µL/well. To the DNA mix, 0.5 µL per well of PLUS reagent was added followed by a brief vortex and RT incubation for 10 min. Lipofectamine was diluted 0.5 µL into 9.5 µL OptiMEM per well. This was added to the DNA + PLUS mix, vortexed briefly and incubated at RT for 15 min. The transfection mix was diluted to 65 µL/well in OptiMEM and aliquoted onto a plate. Cells were incubated with transfection mix overnight (20-22 hrs) and then treated with LPS. *E. coli* K-12 LPS (tlrl-eklps, Invivogen) was dissolved at 5 mg/mL in endotoxin-free water, aliquots were stored at -20º C. To avoid freeze-thaw cycles, working stocks of LPS were prepared at 10 ug/mL and stored at 4º C. To prepare treatments, LPS was diluted in 25% phosphate buffered saline, 75% DMEM. Cells were incubated with treatments for 4 hr. 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/Renilla luciferase activity, normalized to the buffer treated transfection control to compare fold-change in NF-kB activation for treatments.

BRIDGE TO CHAPTER III

In this chapter, we identified the first MD-2 outside of amniotes and showed that MD-2 is present in amphibians and fish. We validated that zebrafish MD-2 supports LPS recognition by zebrafish TLR4. This indicates that LPS signaling, and therefore recognition of exogenous danger signals, by TLR4 occurs across vertebrates. We next set out to determine when endogenous danger signal recognition by TLR4 evolved. S100A9 is an important pro-inflammatory molecule which activates TLR4 by a direct mechanism. S100A9 is specific to mammals. Chapter III describes how we identified MRP-126, an S100A9-ortholog in sauropsids, and validated that it activates TLR4. This work demonstrates that endogenous activation of TLR4 occurs at least across the amniotes.

CHAPTER III

COEVOLUTION OF THE TOLL-LIKE RECEPTOR 4 COMPLEX WITH CALGRANULINS AND LIPOPOLYSACCHARIDE

This chapter contains co-authored material. This chapter was published as Loes, A.N., J.T. Bridgham, M.J. Harms. 2018. "Coevolution of the Toll-like receptor 4 complex with calgranulins and lipopolysaccharide". Front. Immunol. 9:304. This chapter was written by me with editorial assistance M.J. Harms and experimental expertise from J.T. Bridgham.

INTRODUCTION

Calgranulin proteins—including S100A8, S100A9, and S100A12—potently activate inflammation via an interaction with Toll-like receptor 4 (TLR4)[47,61,63,68,127–130]. These Damage Associated Molecular Pattern (DAMP) proteins play important roles in wound healing and vascular development, but can also lead to upregulation and amplification of the inflammatory response in arthritis, atherosclerosis, and inflammatory bowel disease[62,66,131–135]. S100A9 has been identified as a potential drug target for inhibiting inflammation via inhibition of this pathway[68,136]. The molecular basis of calgranulin activation of TLR4 is not well understood but involves direct binding of the calgranulin to the TLR4/MD-2/CD14 complex in a calcium and zinc-dependent manner[68,137].

Calgranulin activation of TLR4 is often compared to lipopolysaccharide (LPS) activation of TLR4[47,68,127]. LPS is a Pathogen Associated Molecular Pattern (PAMP) found in the cell walls of Gram-negative bacteria. LPS activation of TLR4 is relatively

well understood[31]: CD14 binds to LPS and loads it on to an MD-2-TLR4 complex[21,25]. This induces a conformational change that leads to homodimerization of TLR4/MD-2 and triggers an NF-κB cascade[23,32]. In some ways, calgranulin activation is similar. S100A9 has been shown to directly interact with CD14 as well as the TLR4/MD-2 complex, possibly indicating a similar mechanism of activation[68,137]. In other ways, however, it is quite different. Most notably, LPS is a small molecule that is enveloped by MD-2—an unlikely activation pathway for a calgranulin protein[21,29,111,138–140]. Further, LPS and calgranulins induce distinct downstream inflammatory responses, suggesting different modes of action[63,141].

A better understanding of the differences between LPS and calgranulin activation of TLR4 may reveal the unique requirements for calgranulin activation of TLR4. This could allow elucidation of the molecular mechanism of calgranulin activation. Further, the ability to independently modulate PAMP and DAMP activity could allow suppression of pathological DAMP inflammation independently of the pathogen response.

An evolutionary lens provides a powerful means for dissecting these similarities and differences. Comparison of orthologs from different species is a natural way to reveal conserved—and presumably important—features of both DAMP and PAMP activation of TLR4. Additionally, studies of coevolution between the TLR4, MD-2, CD14, and calgranulins may reveal important species-specific interactions that, in turn, provide insight into the mode of activation for each agonist.

We set out to determine whether calgranulin activation of TLR4 was present outside of mammals. While LPS activation of TLR4 has been validated across amniotes[14,91,142,143], calgranulin activation has only been characterized in placental

mammals[61,68]. Using a combination of phylogenetics and functional characterization of amniote orthologs, we establish that calgranulin activation of TLR4 evolved at least in the last common ancestor of all amniotes. While the basic mode of action is conserved, complementation experiments reveal extensive coevolution between TLR4 complex members over time. Crucially, coevolution of components in this complex has different effects on LPS and calgranulins, revealing different molecular requirements for these two agonists.

RESULTS

Calgranulins are found across the amniotes

We first asked when the calgranulin proteins evolved. Calgranulins are members of the S100 protein family[144,145]. Previous phylogenies of the family allowed us to identify the close evolutionary relatives of mammalian calgranulins [144,146,147] but did not provide sufficient resolution to dissect the calgranulin clade itself. We constructed a curated multiple sequence alignment of 172 protein sequences from 30 species. All sequences were close homologs of the calgranulins, with the exception of S100B sequences included as an outgroup. We constructed a maximum-likelihood phylogeny for these proteins. This revealed eight well-supported clades corresponding to the paralog sequences we selected when we built the alignment (Fig. 3.1A). The mammalian calgranulins (S100A8, S100A9 and S100A12) formed a clade with MRP-126, a protein found in birds and reptiles (sauropsids), but not mammals.

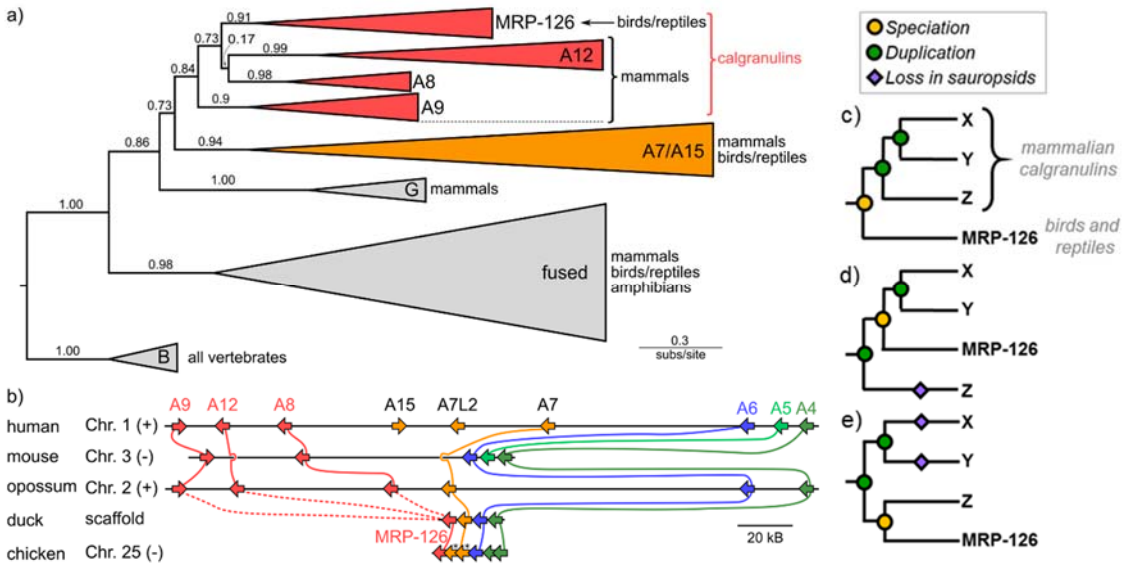

Figure 3.1. Phylogeny and synteny of the S100 family reveal a bird/reptile ortholog for mammalian calgranulin clade. A) Maximum-likelihood phylogeny of 172 sequences from 30 taxa for proteins within a subclade of the S100 family. C-terminal tails of sequences were truncated prior to resolving the phylogeny. Wedges are collapsed clades of orthologs, with wedge height corresponding to the number of included taxa and wedge length indicating the longest branch length with the clade. Support values are SH-supports calculated using an approximate likelihood ratio test. Clades are colored to highlight calgranulins (red) or S100A7/A15 (orange). The taxa included in each clade are noted on the tree. B) Genomic organization of S100 genes in amniote species. Arrows for genes represent the coding strand. Approximate distances between genes are represented by the length of line for the selected chromosome. Mouse and chicken chromosomes were flipped to place A6 in the same orientation between species. Colors indicate orthology: calgranulins (red), A7/A15 (orange), A6 (blue), A5 (light green), and A4 (dark green). Solid lines between genomes indicate well-supported orthology from the phylogenetic tree. Dashed lines indicate ambiguous orthology. Panels c), d) and e) show possible relationships between MRP-126 and the mammalian calgranulins. Mammalian calgranulins are denoted as X, Y, and Z. (Because their branching is ambiguous, we cannot resolve which letter represents which mammalian calgranulin in these trees). Icons indicate evolutionary events: speciation (yellow circle), duplication (green circle), or loss on sauropsid lineage (purple diamond). The panels are arranged from most parsimonious to least parsimonious. Note that, in the tree shown (a), two marsupial proteins are grouped within the sauropsid MRP-126 clade. These are likely marsupial S100A12, which would be consistent with their location in the genome (b) and the lack of other marsupial S100A12 genes. In our more extensive analyses, the placement of these proteins is labile within the calgranulin clade (Fig. AB1).

To validate the close relationship between these proteins, we examined the organization of these genes in the human, mouse, opossum, duck, and chicken genomes. We found that these genes exhibit synteny with one another, consistent with their placement together in the model-based phylogenetic analyses (Fig. 3.1B). In humans, the genes encoding S100A8, S100A9 and S100A12 are near the 5' end of a contiguous cluster of S100 genes, flanked on the 3' end by S100A6 and S100A4[148]. MRP-126 exists in an identical arrangement in its genome, albeit on the opposite strand (Table AB1). We attempted to verify this arrangement held for reptiles, but we were unable to find reptile genomes with contigs of sufficient length in this region to confidently assign synteny.

This analysis establishes MRP-126 as a sauropsid calgranulin; however, relationships within the calgranulin/MRP-126 clade were ambiguous. To try to better resolve the relationship between the calgranulins, we repeated our phylogenetic analysis using both a Bayesian approach and a second maximum-likelihood analysis with a larger set of 494 sequences. In both cases, we obtained similar results: S100A8, S100A9, S100A12, and MRP-126 form a clade, but these proteins could not be confidently placed relative to one another (Fig. AB1). One intriguing observation is that *Anolis carolinensis*, unlike other sauropsids, contains two calgranulin proteins. Both sequences are in our large ML tree (Fig. AB1). One protein is placed with other sauropsid MRP-126 proteins with high confidence. The other is placed within the calgranulins, but its orthology is ambiguous. This could be a lineage-specific duplication, a gene conversion, or a highly diverged ortholog to another S100 such as S100A7.

There are three basic scenarios consistent with our data (shown schematically in Fig. 3.1C-E). The most parsimonious would have MRP-126 co-orthologous to S100A8,

S100A9 and S100A12 (Fig. 3.1C). This requires two duplication events on the mammalian lineage after the divergence of mammals and sauropsids. Alternatively, MRP-126 could be orthologous to two of the mammalian proteins, requiring two duplication events and one loss (Fig. 3.1D). The least parsimonious scenario would have MRP-126 co-orthologous to one mammalian calgranulin. This would require two duplications and two losses (Fig. 3.1E). Alternate scenarios exist but require more duplications and losses. The first scenario seems most plausible, given its parsimony; however, our data are insufficient to fully resolve these relationships. Whatever the precise branching pattern within the calgranulin clade, our analysis supports classifying MRP-126 as a sauropsid calgranulin.

Finally, despite extensive BLAST queries, we were unable to locate any calgranulin orthologs in amphibians or bony fishes. Because of the large number of high-quality fish genomes, this likely indicates that calgranulins are specific to tetrapods. There are relatively few amphibian genomes, so it remains unclear whether the lack of calgranulins in amphibians represents poor sampling or a true origin in the amniote ancestor.

Calgranulin DAMP activity exists across amniotes

Our phylogenetic analyses reveal that calgranulins arose at least in the ancestor of amniotes. To determine when DAMP activity evolved, we "walked out" along the tree, recombinantly expressing and purifying calgranulin proteins from the human (*Homo sapiens*), mouse (*Mus musculus*), opossum (*Monodelphis domestica)* and chicken (*Gallus gallus*). For the mammals, we selected S100A9 proteins which, in the human and mouse,

are known to be potent DAMPs. We expressed chicken MRP-126 as a representative sauropsid calgranulin (Fig. 3.2A). After purification, we verified that all proteins were folded using far-UV circular dichroism spectroscopy. As expected for S100 proteins, all four proteins gave largely alpha-helical structures (Fig. 3.2B). We also built homology models of mouse S100A9, opossum S100A9 and chicken MRP-126 using the human S100A9 template (Fig. 3.2C). All these sequences could be readily accommodated in the S100 fold.

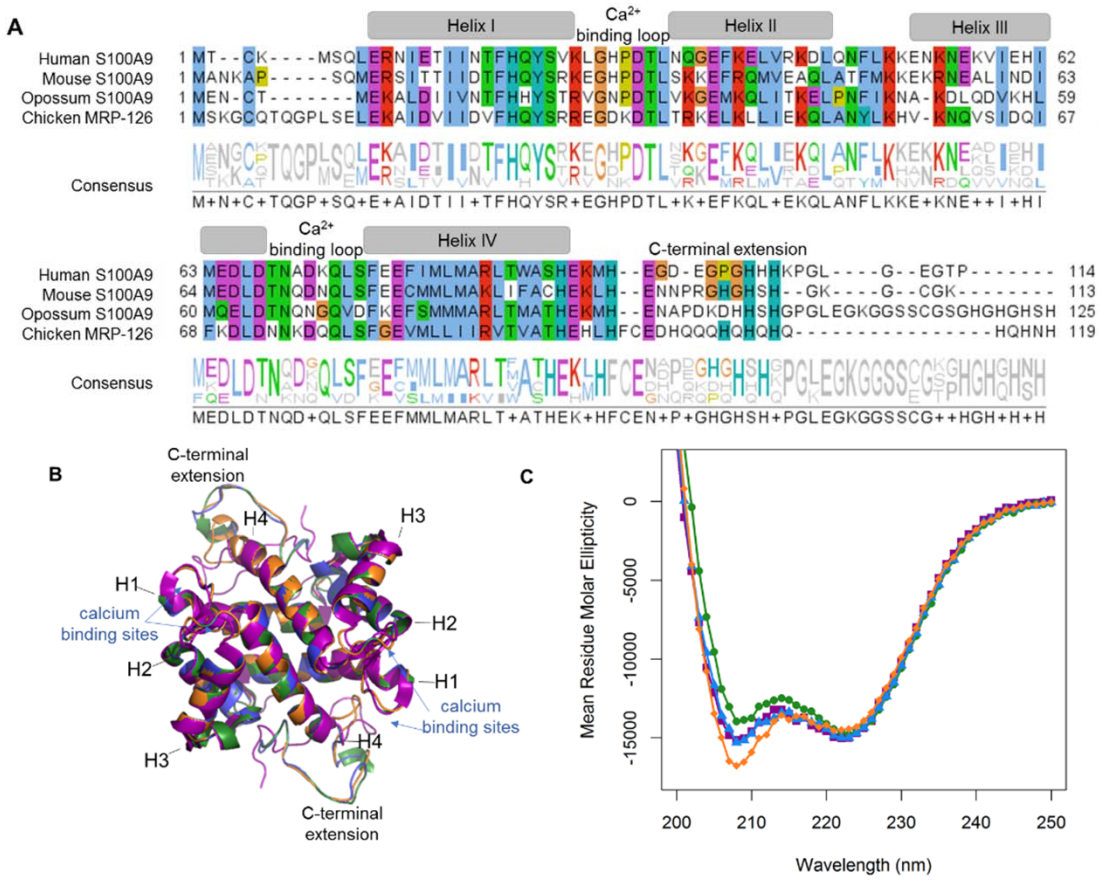

Figure 3.2. MRP-126, like mammalian calgranulins, is primarily alpha-helical with a disordered C-terminal extension. A) Multiple sequence alignment of mammalian calgranulins with chicken MRP-126. Amino acid residues are colored to show sites with similar chemical properties. B) Homology model of mouse S100A9 (green), opossum S100A9 (blue), and chicken MRP-126 (orange) aligned with human S100A9 (purple) calcium-bound NMR structure *PDB: 5i8n* C) Circular dichroism spectra of mammalian calgranulins and chicken MRP-126

We next asked how well each calgranulin could activate via TLR4. We assayed the TLR4-dependent activation of the inflammatory NF-κB pathway using an *ex vivo* cell culture assay[149,150]. We transiently transfected plasmid constructs encoding TLR4 and its cofactors MD-2 and CD14 into HEK293T/17 cells. These cells do not natively express TLR4, MD-2 or CD14 so TLR4-dependent activation of NF-κB depends on the heterologous expression of the transfected TLR4 complex components. We used a luciferase gene under the control of an NF-κB promoter as a reporter. The NF-κB pathway is conserved across bony vertebrates[5], making mammalian cells useful as a common host for amniote TLRs.

We first validated the assay by measuring the known PAMP and DAMP activation of TLR4 complexes by the human and mouse proteins[47,68,127]. We transfected genes encoding TLR4, MD-2, and CD14 into cells and measured activity under different treatment conditions (Fig. 3.3A, 3.3B). LPS induced significantly higher NF-κB response than a mock treatment. The addition of the LPS sequestering agent polymyxin B completely abolished this activity. We then treated transfected cells with recombinantly expressed S100A9—the most potent DAMP of the mammalian calgranulins. Both the human and mouse TLR4 complexes responded robustly to the addition of species-matched S100A9. All treatment buffers, with the exception of the LPS treatment, contained polymixin B to eliminate spurious activation via potential LPS contamination. Both LPS and S100A9 activation were strictly dependent on the presence of heterologous TLR4 (Fig. 3.3A, 3.3B).

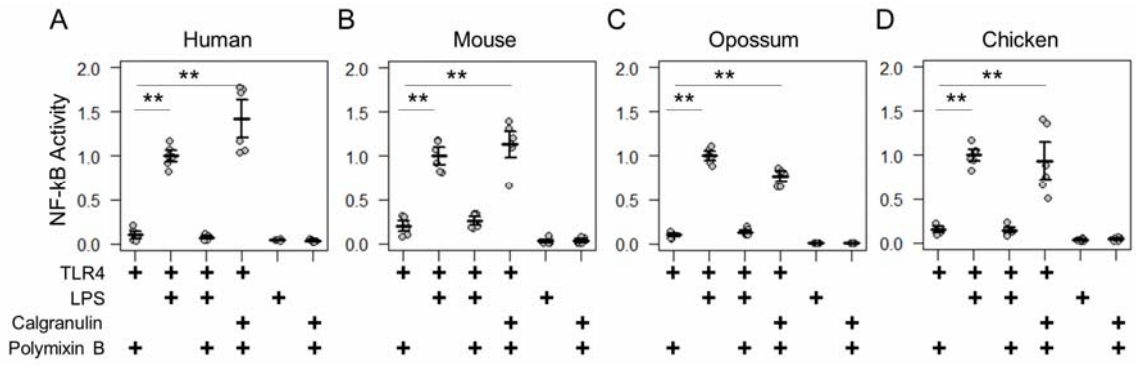

Figure 3.3. Calgranulin activation of TLR4 is shared across amniotes. NF-κB activity for species-matched TLR4/MD-2/CD14 complexes in the presence of calgranulins from A) human, B) mouse, C) opossum and D) chicken. A "+" in the table below each series indicates which components are included in that treatment. LPS is used as a positive control for expression and activation of the complex. Polymixin B is included to control for endotoxin-mediated activation of the complex. Activity is normalized to LPS activity of positive control within each biological replicate, (i.e. for human, each technical replicate is divided by the average LPS activation of hTLR4, hMD2, and hCD14 for that biological replicate). Points are the technical triplicates from three biological replicates. Bold lines are the mean of the biological replicates. Error bars are a standard error on the mean of the biological replicates. A two-tailed t-test was used to assess the significance of the difference in mean between the indicated series (**p-value < 0.01).

With the assay validated, we then turned our attention to the TLR4 complex from other amniotes. We started by assessing a marsupial, the opossum. TLR4, MD-2, CD14, and S100A9 are known to be expressed in marsupials[151,152], but their activity had not been characterized for LPS or DAMP activation. We transfected opossum TLR4 complex components and treated them with both LPS and recombinant opossum S100A9. We found an identical pattern of activation for the opossum proteins relative to those from human and mouse (Fig. 3.3C).

We next assessed the activation of the chicken TLR4 complex by LPS and MRP-126. Chicken MRP-126 is known to have a similar expression profile to mammalian calgranulins, with upregulated expression during bacterial infection[92,153], suggesting it could play a similar role in inflammation. Chicken CD14 had not previously been

functionally characterized for a role in LPS activation[92,154]. We transfected TLR4 complex components from the chicken. We observed an identical pattern of activation for the chicken as for the mammalian proteins for activation of the chicken TLR4 complex by LPS and recombinant MRP-126. (Fig. 3.3D).

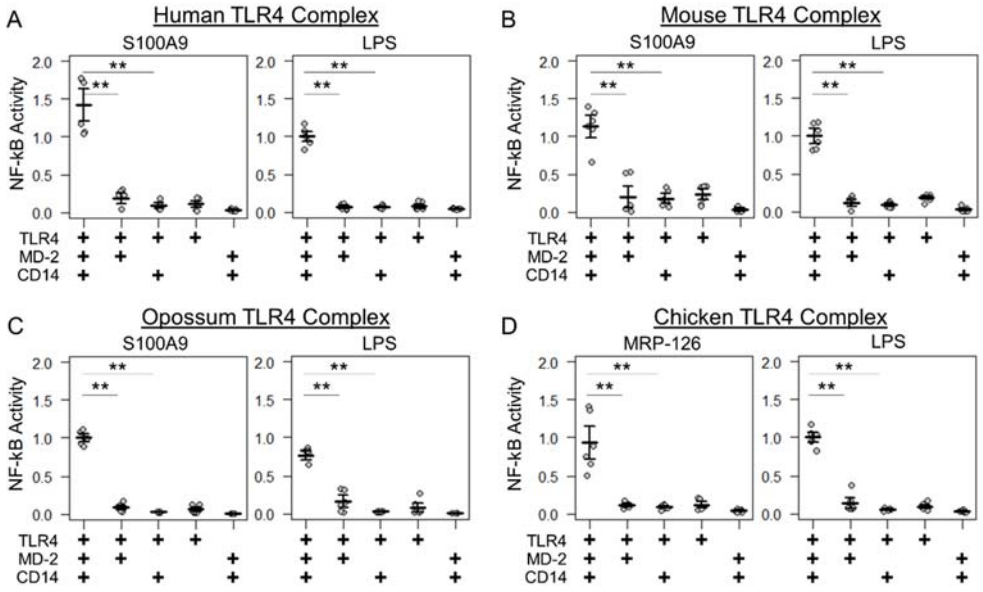

Figure 3.4. Amniote calgranulins have similar TLR4 complex requirements. NF-κB activation of TLR4 from A) human, B) mouse, C) opossum and D) chicken. NF-κB is normalized to LPS activation of the control complex for that species. Points are the technical replicates from three biological replicates. Bold lines are the mean of the biological replicates. Error bars are a standard error on the mean of the biological replicates. A "+" in the panel below indicates which components are included in the treatment. A two-tailed t-test was used to assess the significance of the difference in mean between the indicated series (**p-value < 0.01).

Calgranulins from across the amniotes activate TLR4 in a similar fashion

We next determined whether the same components were necessary and sufficient for calgranulin activation of TLR4 across the amniotes. Shared complex requirements are strong evidence for a common, ancestral mode of action. We first measured the ability of mammalian S100A9 and chicken MRP-126 to activate their species-matched TLR4 with

and without the cofactors MD-2 and CD14 present. We found that MD-2 was required for calgranulin activation of TLR4 in all species, as no MD-2-independent signaling was observed for any species tested (Fig. 3.4). CD14 also strongly contributed to signaling. This is consistent with previous observations that CD14 is involved in TLR4-dependent NF-κB signaling by human S100A9[137]. Some CD14-independent NF-κB signaling was observed in the human and mouse proteins; however, the addition of CD14 drastically improved the signal for activation. No significant signaling was observed for the opossum or chicken proteins in the absence of CD14 (Fig. 3.4).

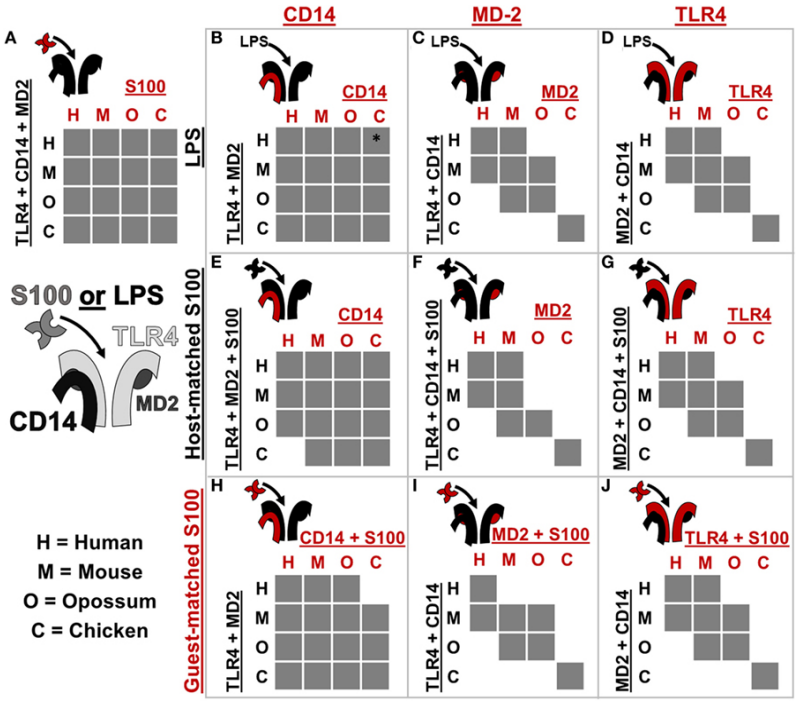

Figure 3.5. TLR4 complex components exhibit lineage-specific coevolution. Subpanels show activation of NF-κB signaling by receptor complexes assembled from host (black) and guest (red) proteins. Protein names on the left and top of each sub-panel indicate the components included. S100 is used to denote human S100A9, mouse S100A9, opossum S100A9 or chicken MRP-126. Letters indicate which species components come from human (H), mouse (M), opossum (O), or chicken (C). If the cross-species complex activates >2-fold above the buffer control, the corresponding box is filled gray. The icon to the top-left of each sub-panel indicates graphically which components are being combined. * indicates >2-fold activation above buffer control, but low total activation (see Fig. AB5)

We next tested whether calgranulins from across the amniotes could complement one another. We treated the TLR4 complex from each species with recombinant calgranulin from each species and then measured activation. We found that all calgranulins could activate all complexes (Fig. 3.5A), again supporting a shared ancestral mode of action. Dose-responsive activation occurs at concentrations in the micromolar range for all calgranulins against amniote TLR4/MD2/CD14 complexes (Fig. AB2).

Calgranulins and LPS signaling have overlapping, but different molecular requirements

Coevolution can lead to species-specific interactions that, in turn, reveal key determinants of activation[139,143]. The cross-reactivity of calgranulins against TLR4 complexes from different species reveals a shared mode of action; however, we were also interested in identifying any species-specific differences in activation. We, therefore, searched for coevolution between the calgranulins and members of the TLR4 complex. We took a complementation approach, adding a guest component from one species into a host complex of components from a different species (Fig. 5, AB3-AB8). We used this approach to test for lineage-specific evolution between TLR4, MD-2, CD14, and the calgranulins.

We first tested whether host-guest complexes were competent for LPS signaling. We used CD14 (Fig. 3.5B), MD-2 (Fig. 3.5C) and TLR4 (Fig. 3.5D) as guests and then measured activation upon the addition of LPS. We found that CD14 was non-specific and could complement its orthologs from other species (Fig. 3.5B). The lack of species-

specificity for CD14 is consistent with it acting as a peripheral protein that delivers LPS to the central TLR4/MD-2 complex[100,155].

In contrast, MD-2 and TLR4 exhibited strong species-specific variability in activation: many host-guest pairs were incompatible (Fig. 3.5C, 3.5D). The only proteins that could complement one another fully were the mouse/human and mouse/opossum pairs. Further, MD-2 and TLR4 gave essentially identical patterns of compatibility between species. This indicates strong coevolution between these two proteins. This likely arises as a result of the large, functionally critical interface formed between TLR4 and MD-2[20,21,23,29,139,156].

We repeated these experiments using calgranulins rather than LPS to activate the complex. We used calgranulin matched to the host complex and added guest CD14 (Fig. 3.5E), MD-2 (Fig. 3.5F), and TLR4 (Fig. 3.5G). This revealed a similar pattern to that of LPS signaling. CD14 was relatively nonspecific (Fig. 3.5E), while MD-2 and TLR4 were highly specific (Fig. 3.5F, 3.5G). Indeed, the MD-2/TLR4 pattern of activation for calgranulins was identical to that of LPS, with the exception of the human-mouse and mouse-opossum heterocomplex. A similar pattern was observed when examining activation by the calgranulin matched to the guest component (Fig. 3.5H-J).

The host-guest combinations that do not activate with either LPS or calgranulin likely fail to assemble into a productive complex, independent of the nature of the pro-inflammatory signal. The failure of the mouse-opossum complexes for calgranulin but not LPS, however, indicates that there are different requirements to activate the complex via LPS or calgranulins. This can be seen with a more detailed comparison of the opossum/mouse MD-2 complementation analysis (Fig. 3.6A-C). Opossum MD-2 can

stand in for mouse MD-2 for LPS signaling (Fig. 3.6A), but not for calgranulin signaling (Fig. 3.6B, 3.6C). This reveals that these two pro-inflammatory signals activate differently, despite identical complex requirements (Fig. 3.4).

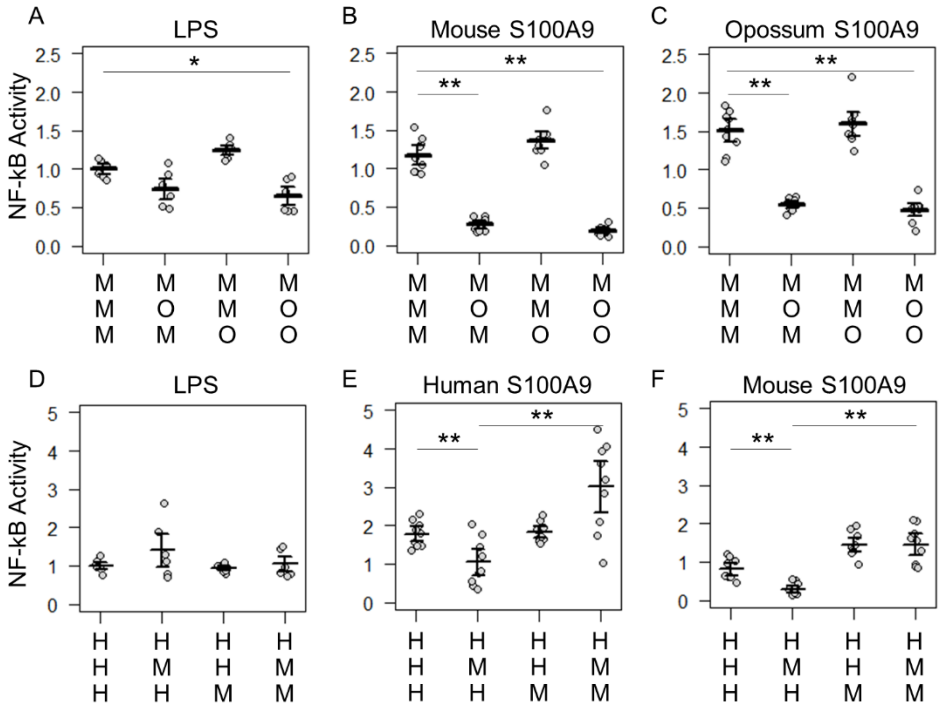

Figure 3.6. LPS and calgranulin exhibit different complex requirements. Panels a-c show NF-κB activation of mouse TLR4 with different combinations of mouse and opossum MD-2 and CD14. Activation by: A) LPS, B) mouse S100A9, and C) opossum S100A9. Panels D-F show NF-κB activation of human TLR4 with different combinations of human and mouse MD-2 and CD14. Activation by: D) LPS, E) human S100A9, and F) mouse S100A9. Letters indicate which species the component was taken from: human (H), mouse (M) or opossum (O). NF-κB is normalized to LPS activation of the control complex for that species. Points are the technical triplicates from three biological replicates. Bold lines are the mean of the biological replicates. Error bars are a standard error on the mean of the biological replicates. A "+" in the panel below indicates which components are included in the treatment. A two-tailed t-test was used to assess the significance of the difference in mean between the indicated series (*p-value < 0.05, **p-value < 0.01).

There were also quantitative differences between LPS and calgranulin activation of human/mouse heterocomplexes. Mouse MD-2 fully complements human MD-2 for

LPS activation of human TLR4 (Fig. 3.6D). In contrast, mouse MD-2 is less efficient than human MD-2 for calgranulin activation of the complex (Fig. 3.6E, 3.6F). Intriguingly, this difference can be offset by the addition of the mouse CD14 in addition to mouse MD-2 (Fig. 3.6E, 3.6F). This interaction between MD-2 and CD14 reveals that both components are important for calgranulin activation of TLR4 and that their roles are different for LPS and calgranulin.

Our analysis reveals lineage-specific coevolution between complex members. The dominant signal for coevolution is between TLR4 and MD-2. This impacts both LPS and calgranulin activation, likely because the interaction between MD-2 and TLR4 is important for complex assembly. There are, however, differences between which host-guest complexes are sensitive to LPS and calgranulins, revealing that LPS and calgranulins activate these complexes in subtly different ways. These differences may indicate that the interface calgranulins interact with has independent molecular requirements from LPS, possibly spanning components or inducing an alternate active conformation than LPS.

DISCUSSION

Our work reveals that the calgranulin (DAMP) and LPS (PAMP) activation of TLR4 evolved at least as early as the ancestor of amniotes. As all amniote calgranulins we tested activate their species-matched TLR4, this activity likely arose before the divergence of amniotes and mammals ~320 million years ago. Like LPS, activation of NF-κB by calgranulins is strictly dependent on the cofactor MD-2 and potently increased

by the cofactor CD14. These shared cofactor requirements suggest that both DAMP and PAMP use conserved, ancestral modes of activation.

We also find that DAMP and PAMP activation must occur via slightly different pathways. This can be seen in our complementation experiments. While opossum MD-2 can complement mouse MD-2 for LPS activation, it cannot complement mouse MD-2 for calgranulin activation. Likewise, mouse MD-2 requires mouse CD14 for calgranulin activation of human TLR4. These results could point to a direct interaction between MD-2 and S100A9 or a multi-way interaction between TLR4/MD-2/CD14 and calgranulins. Further analysis may help identify the residues involved in the TLR4/MD-2/calgranulin interface, as has been done for the MD-2/TLR4 interface by comparing human and other species activation by LPS[111,143]. Alternatively, this signal may be mediated indirectly, via the dimerization of TLR4/MD-2 – activation of TLR4 requires dimerization of TLR4/MD-2, in part, mediated by residues on MD-2[21,22,32,156].

We found that calgranulins are broadly cross-reactive to TLR4 across the amniotes. This is consistent with a fairly non-specific mechanism of action, such as binding via a non-specific hydrophobic patch on the calgranulin surface. Hydrophobicity has been proposed to be an important global danger signal – exposure of hydrophobic regions of folded proteins is an indicator of damage[157]. If a non-sequence dependent mechanism of activation, like hydrophobicity, is employed by the calgranulins for TLR4 activation, this may be a co-opted signaling mechanism which evolved prior to the calgranulins and is employed by other endogenous danger signals for TLRs.

It remains an open question when LPS activation evolved relative to calgranulin activation. One intriguing possibility is that the DAMP activation of TLR4 predates

PAMP activation. TLR4 is expressed in bony fishes, but the ligand for this receptor is unknown. If another class of hydrophobic DAMP activates TLR4 in fish, DAMP activation of TLR4 may even predate amniotes. We do not yet know when LPS activation of TLR4 evolved. Resolving if DAMP activation evolved prior to PAMP activation will require studying TLR4 activation in amphibians and bony fishes.

Previous studies have investigated the ability of zebrafish TLR4 to activate NF-kB signaling in response to LPS[93,94,110]. Because no orthologs to mammalian MD-2 and CD14 have been identified outside of amniotes, the researchers used mammalian CD14 and MD-2 in their experiments with zebrafish TLR4s. They observed no response to LPS. One interpretation of this result is that fish do not respond to LPS via TLR4. An observation that may support this is the loss of TLR4 in many fish lineages[93,94,110]. Our complementation studies suggest that MD-2 coevolution with TLR4 would likely prohibit cross-reactivity of mammalian MD-2 with fish TLR4, even if fish possessed a functional MD-2-like cofactor. LPS activation of TLR4 in fishes should thus be reexamined. This will, however, require identification of TLR4 cofactors from fish—if they exist—rather than cofactors from tetrapods.

Our experiments and evolutionary analysis have revealed that calgranulin activation of TLR4 evolved at least 180 million years earlier than previously appreciated. These results show that DAMP recognition occurred through TLR4 at least in early amniotes. For many years, the predominant theory of the innate immune system was that its primary role was to discriminate “self” from “non-self”. In the 1990’s, Matzinger proposed the danger hypothesis[158] – which describes the innate immune system as a general danger sensor, able to respond not just to pathogen molecules, but also to

endogenous signals indicative of damage. Our results are consistent with an ancient, general danger-sensing function for TLR4: at least across amniotes TLR4 plays a role in combating danger, not simply as a mechanism to distinguish "self" and "non-self".

MATERIALS & METHODS

Phylogenetics and genomic analysis

We constructed a curated database of S100 homologs from a consistent set of species. We obtained amino acids sequences of S100 proteins from a subset of the amniote species in Ensembl version_87[159]. We obtained additional bird and reptile S100A7 and MRP-126 sequences using the human calgranulin paralogs to BLAST against the NCBI database. This yielded a set of 172 sequences from 30 taxa (Data Sheet S1, S3). We constructed our multiple sequence alignment using MSAPROBS[119], followed by manual editing in MEGA[120]. We trimmed the alignment to remove highly variable (and therefore unalignable) C-terminal extensions, as well as the non-S100 domains of the fused S100 proteins. We used PHYML-SS[121,122] with subtree pruning and re-grafting to construct the ML phylogeny. Pilot analyses revealed that the LG substitution model with 8 rate categories and a floating gamma distribution parameter yielded the highest likelihood trees[123–125]. An AIC test was used to control for overfitting[126]. We rooted our trees using the divergence of S100B, an ancient S100 found across jawed vertebrates. These steps were repeated with a set of 494 sequences representing all amniote S100s within this clade of the S100 family (Data Sheet S2, S4, S5, S6).

We constructed our Bayesian tree using ExaBayes[160] using a single data partition, integrating over tree topologies, rates, and evolutionary models. We ran two replicate analyses, using three heated chains and a single cold chain per replicate. We ran for 5.6 million rounds, with a final average difference between replicates of 0.6 %. We discarded the first 15% of the analysis as burn-in and generated a 50% majority-rule consensus tree.

The syntenic analysis was done using the Ensembl synteny module[161] to map orthologs and homologs onto the chromosomes of species of interest (Table AB1). The genome assemblies were *Homo sapiens* (GRCh38.p10; GCA_000001405.25), *Mus musculus* (GRCm38.p5; GCA_000001635.7), *Monodelphis domestica* (monDom5; GCF_000002295.2), *Anas platyrhynchos* (BGI_duck_1.0; GCA_000355885.1), and *Gallus gallus* (Gallus_gallus-5.0; GCA_000002315.3). We attempted to include representative reptile genome; however, the S100 genes of interest were found on different contigs in the available reptile genomes.

Plasmids and recombinant protein preparation and characterization

Mammalian expression vectors containing human TLR4 and mouse TLR4 were a gift from Ruslan Medzhitov (Addgene plasmid #13085 and #13086) and human CD14 and ELAM-Luc were a gift from Doug Golenbock (Addgene plasmid #13645 and #13029). Human MD-2 was obtained from the DNASU Repository (HsCD00439889). Mouse MD-2 (UniProt #Q9JHF9) and CD14 (UniProt #P10810), as well as opossum TLR4 (UniProt #F6Y6W8), MD-2 (UniProt #F6QBE6), CD14 (NCB Accession #XP_007473804.1) and chicken TLR4 (UniProt #C4PCF3), MD-2 (UniProt #A0A1D5NZX9), and CD14 (UniProt #B0BL87) were designed to be free of restriction

sites and codon optimized for human expression and purchased as mammalian expression vector constructs in pcDNA3.1(+) from Genescript (New Jersey, USA).

Synthetic gene constructs for human (UniProt #P06702), mouse (UniProt #P31725), and opossum (UniProt #F7AJJ0) S100A9 and chicken MRP-126 (UniProt #P28318), were also designed to be free of common restriction sites and codon optimized and purchased as PUC57 constructs from Genewiz (New Jersey, USA). Genes were cloned into a modified 6xHis MBP LIC TEV vector with NcoI and HindIII to yield protein constructs with TEV-cleavable histidine tags. *E. coli* BL21(DE3) pLysS competent cells containing the expression vectors for S100 proteins were grown overnight at 37º C, diluted 1:150 into LB containing ampicillin and chloramphenicol, grown to mid-log phase (OD_{600} ~ 0.6-1), and induced with 1 mM isopropyl-β-d-1-thiogalactopyranoside overnight at 16º C with aeration. Bacterial pellets were harvested by centrifugation at 3000 rpm at 4º C for 20 min and stored at -20º C. Pellets (~5 g) were suspended by vortexing and lysed in 25 mL Buffer (25 mM Tris, 100 mM NaCl, 25 mM imidazole, pH 7.4) with 37.5 U DNase I (ThermoFisher Scientific) and 0.75 mg Lysozyme (ThermoFisher Scientific) by shaking at RT for approximately 1 hr. The lysate was clarified by centrifugation at 15,000 RPM for 50 min. at 4º C. Protein was purified using a 5 mL Ni^{2+}-NTA HisTrap column from (Healthcare GE) using an FPLC (Akta Biosciences) with gradient elution to HisB (25 mM Tris, 100 mM NaCl, pH 7.4, with 500 mM imidazole for human, mouse and chicken proteins, 1 M imidazole for opossum S100A9). Pooled elution peak of purified protein was cleaved with tobacco etch protease (TEV) overnight at RT. Cleaved protein was collected from a gradient elution of a 5 mL Ni-NTA column from HisA to HisB. Protein purity was assessed with SDS-PAGE and

pure fractions were pooled and dialyzed into phosphate buffered saline (PBS), 0.5 mM TCEP, pH 7.4. Protein was flash frozen in liquid nitrogen and fresh aliquots were thawed weekly for functional assays. TLR4 activation was tested with two independent preps of each protein to ensure that the results were not batch specific. Protein concentrations were measured using a Bradford assay.

Homology models of mouse S100A9, opossum S100A9 and chicken MRP-126 were prepared with Swiss-Model[162,163]. For secondary structure measurements, protein samples were prepared at 10 μM in endotoxin free PBS and far-UV circular dichroism data was collected between 200–250 nm using a J-815 CD spectrometer (Jasco) with a 1 mm quartz spectrophotometer cell (Starna Cells, Inc. Catalog No. 1-Q-1). Duplicate scans were collected for each protein and averaged. A blank buffer spectrum was subtracted from sample measurements and raw ellipticity was converted into mean molar ellipticity using concentration and the number of residues for each protein.

Cell culture and transfection conditions

Human embryonic kidney cells (HEK293T/17, ATCC CRL-11268) were maintained up to 30 passages in DMEM supplemented with 10% FBS at 37º C with 5% CO_2. For each transfection, a confluent 100 mm plate of HEK293T/17 cells was treated at room temperature with 0.25% Trypsin-EDTA in HBSS and resuspended with an addition of DMEM + 10% FBS. This was diluted 4-fold into fresh medium and 135 μL aliquots of resuspended cells were transferred to a 96-well cell culture treated plate. Transfection mixes were made with 10 ng of TLR4, 1 ng of CD14, 0.5 ng of MD-2, 0.1 ng of Renilla, 20 ng of ELAM-Luc, and 68.4 ng pcDNA3.1(+) per well for a total of 100 ng of DNA,

diluted in OptiMEM to a volume of 10 µL/well. To the DNA mix, 0.5 µL per well of PLUS reagent was added followed by a brief vortex and RT incubation for 10 min. Lipofectamine was diluted 0.5 µL into 9.5 µL OptiMEM per well. This was added to the DNA + PLUS mix, vortexed briefly and incubated at RT for 15 min. The transfection mix was diluted to 65 µL/well in OptiMEM and aliquoted onto a plate. Cells were incubated with transfection mix overnight (18-22 hrs) and then treated with protein (2 µM) or LPS (100 ng/well) mixtures (prepared in 25% phosphate buffered saline, 75% DMEM). *E. coli* K-12 LPS (tlrl-eklps, Invivogen) was dissolved at 5 mg/mL in endotoxin-free water, aliquots were stored at -20º C. To avoid freeze-thaw cycles, working stocks of LPS were prepared at 10 ug/mL and stored at 4º C. There has been some concern in testing recombinant DAMPs against TLR4 due to the potential presence of contaminating LPS in proteins which have been expressed in bacteria [7,9]. We tested our S100s in the presence of 50 µg/mL polymyxin B, an LPS binding agent to limit signaling from LPS contamination in recombinant protein preparations. This concentration of polymyxin B, while sufficient to eliminate signaling by 100 ng/mL of LPS had a minimal effect on the signaling by calgranulins (Fig. AB9). Cells were incubated with treatments for 4 hr. 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/Renilla luciferase activity, normalized to the LPS-treated transfection control for each species in order to normalize between plates. It should be noted that the level of constitutive activity was different between TLR4s from different species (Fig. AB10).

For cross-species comparisons, we used an internal standard to account for systematic differences in the magnitude of luciferase activation between biological replicates. We normalized measured activation values such that the LPS activation of the reference gene combination was 1.0. This is an appropriate normalization because we only compare the relative activation of gene/activator combinations on the same plate. For example, for each biological replicate plate used in the human/mouse complementation analysis, we divided the activation of all gene/activator combinations on that plate by the LPS activation hTLR4/hMD2/hCD14 observed for that plate. This means we are comparing the activation of other combinations (e.g. hA9 activation of hTLR4/mMD2/mCD14) relative to the internal standard (LPS activation hTLR4/hMD2/hCD14).

BRIDGE TO CHAPTER IV

In this chapter, we showed that activation of TLR4 by S100s occurs across the amniotes. We also show that S100 activation requires co-factors MD-2 and CD14, and that it appears to occur via a conserved, but a non-specific mechanism. Activation of human TLR4 by S100A9 has been shown to occur via a calcium and zinc-dependent mechanism. Zinc binding, however, has not been well characterized for S100A9 homodimers. In Chapter IV, we examine zinc binding by S100A9 homodimers and examine the role of zinc binding to S100A9 in activation of TLR4.

CHAPTER IV

ZINC-BINDING AND THE C-TERMINAL TAIL ARE NOT REQUIRED FOR PRO-INFLAMMATORY ACTIVITY OF S100A9

This chapter contains co-authored material. Experiments were performed by me and Ran Shi. ICP-MS data was collected by the Elemental Analysis Core at OHSU. This chapter was written by me with editorial assistance from Michael J. Harms

INTRODUCTION

Chronic inflammation is at the root of many human diseases. Targeted disruption of inflammatory signaling cascades is a core goal of therapeutic strategies for alleviating inflammation. S100A9 is a biomarker for inflammatory diseases[62,63,68,127,129] and has been identified as a potential therapeutic target[68]. Deletion of S100A9 in mice has a protective effect against septic shock[67], as this eliminates the positive feedback loop of chronic pathogenic inflammation propagated by S100A9. The potent pro-inflammatory activity of S100A9 is mediated through a direct interaction with TLR4[68]. The molecular basis for this interaction is not yet known. Understanding the molecular mechanism for S100A9 activation of TLR4 is important for being able to design targeted inhibitors of this interaction to regulate inflammation.

We can gain some insight into how S100A9 interacts with TLR4 by examining other members of the S100 family, which are predominately calcium-regulated signaling proteins. Like other members of the S100 family, S100A9 undergoes a characteristic conformational change upon Ca^{2+} binding which increases the solvent-accessible

hydrophobic surface area[164]. For many of the S100s, this hydrophobic surface has been shown to interact with short linear motifs of target proteins or peptides[165–168]. Recent data also suggests that it is this region of S100A9 that directly interacts with TLR4[169]. Further, the interactions between S100A9 and protein partners require calcium, suggesting S100A9 uses a similar binding mechanism for recognition of targets as other S100s[68,136,170].

While many S100 proteins bind to transition metals via a canonical site at the dimerization interface[146], S100A9 is unique in that binding to transition metals has been shown to modulate its interaction with protein partners, like Toll-like receptor 4 (TLR4)[68]. How zinc regulates interactions between S100A9 and TLR4 is not known. Here we set out to characterize how S100A9 interacts with zinc and how zinc binding by S100A9 is involved in activation of TLR4. We assessed whether deletion of the C-terminal extension or mutation of the transition metal binding site altered activation of TLR4. These experiments reveal that zinc is not required for activation of TLR4 in a cell culture assay either through direct binding or coordination of the C-terminal tail and that zinc-dependent aggregation is not required for activation of TLR4.

RESULTS

S100A9 homodimer binds to zinc

Zinc binding has not been characterized for the S100A9 homodimer. However, transition metal binding by the heterodimer of S100A8 and S100A9, known as calprotectin, has been extensively studied[171–173]. We set out to determine whether S100A9 homodimer binds to zinc via a similar mechanism observed for the heterodimer.

The heterodimer contains two distinct transition metal binding sites. Histidine residues in the C-terminal tail of S100A9 (H91, H95, H103-105) contribute to a high-affinity six-histidine manganese binding site in calprotectin[173] (Fig. 4.1). Additionally, N-terminal metal-coordinating residues in S100A9 (H20, D30) form part of a tetrahedral zinc binding site (Fig. 4.1). We sought to determine whether the zinc-binding sites in the S100A9 homodimer resemble either of the sites observed in the heterodimer. We designed two mutant constructs (Table AC1) of S100A9, a C-terminal truncation mutant (S100A9Δ99) and a metal binding mutant where all predicted zinc-coordinating histidines were substituted for asparagine, and the N-terminal aspartate was exchanged for serine. In our full metal site mutant, S100A9ΔZn, we also chose to mutate the N-terminal cysteine, as cysteine can coordinate zinc (C3S/H20N/D30S/H91N/H95N/H103-105N).

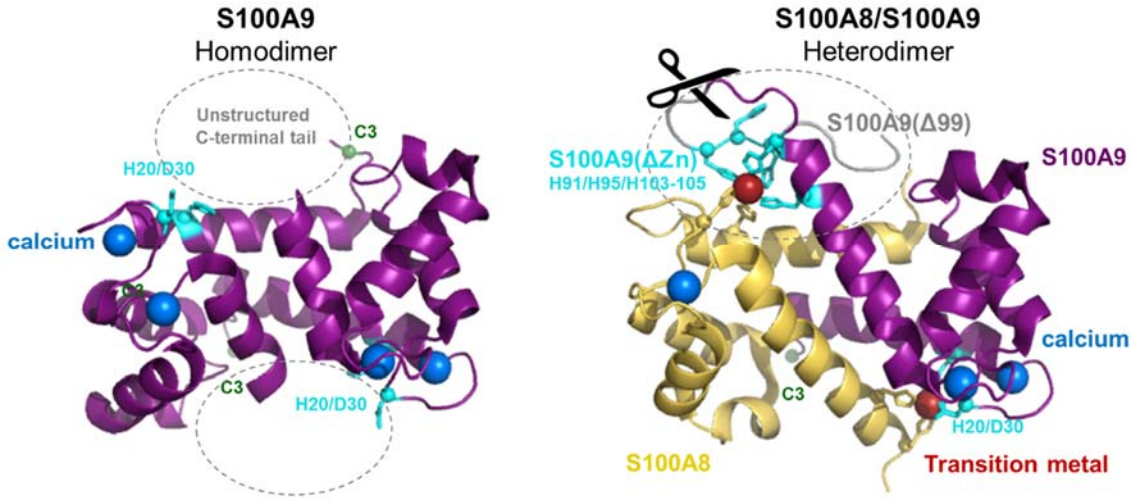

Figure 4.1. Heterodimer crystal structure leads to hypotheses for metal binding residues in S100A9 homodimer. Cartoon representations of crystal structures for S100A9 homodimer in the presence of calcium (PDB: 1irj)[70] and S100A8/S100A9 heterodimer (PDB:4ggf)[171] in the presence of calcium and manganese. The C-terminal tail (grey) of S100A9 is disordered in the absence of transition metals but becomes structured when histidines in the tail coordinate metal ions. Zinc binding residues of S100A9 have been identified for the heterodimer (cyan), mutation of these residues allows for identification of whether the homodimer binds to zinc at the same site as the heterodimer.

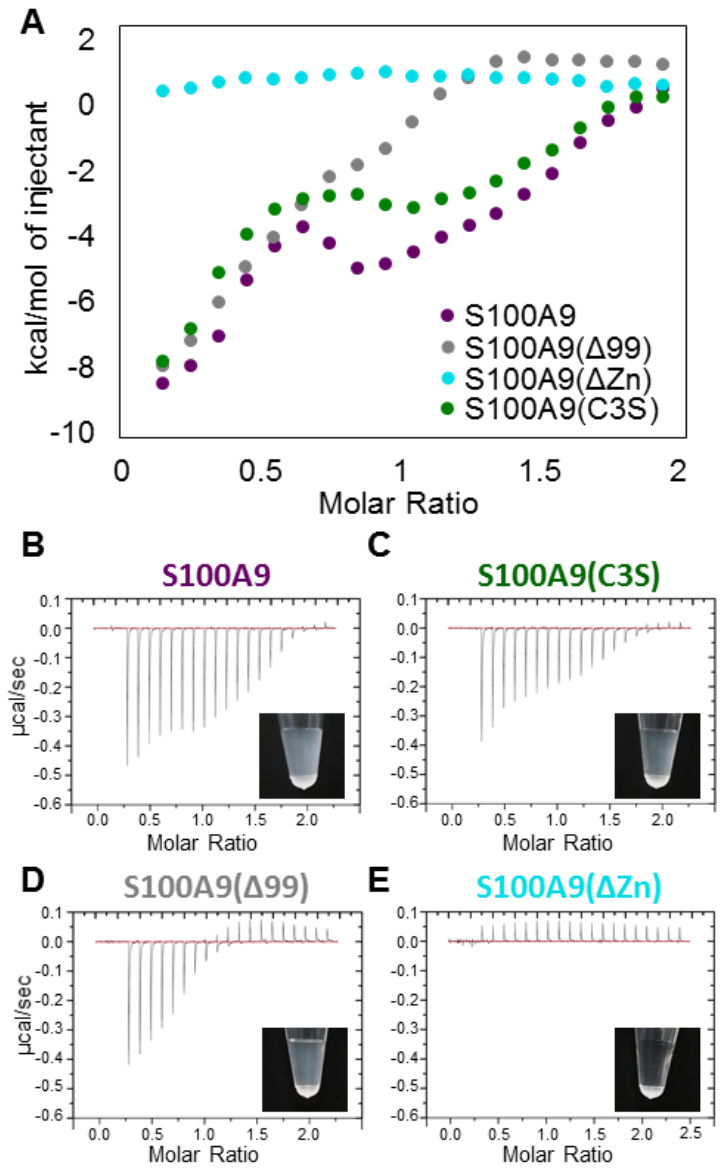

Figure 4.2. Metal binding by S100A9 homodimer. A) Metal binding mutants were constructed which disrupt zinc binding to S100A9 by ITC. Isotherms for metal binding mutants further show the distinct difference between B) wild-type, C) S100A9(C3S), D) S100A9(Δ99) and E) S100A9(ΔZn). Pictures are included to aggregation that occurred for S100A9, S100A9(C3S), S100A9(Δ99), but not for S100A9(ΔZn)

We examined zinc-binding to S100A9 in the presence of calcium, using isothermal calorimetry (ITC). The isotherm for wild-type S100A9 is consistent with binding (Fig. 4.2A, 4.2B). However, aggregation was observed within the sample cell upon addition of zinc. This aggregation led to difficulties in fitting these data to obtain a

reliable K_d. Despite the inability to calculate affinity, these experiments did give some insight into zinc binding by S100A9: zinc binding by S100A9 is exothermic, relatively low affinity, and appears to occur by two distinct processes.

To alleviate the aggregation issue, we substituted a serine for the N-terminal cysteine of S100A9. Previous studies have shown that an N-terminal cysteine in S100A2 comprises an additional zinc site which leads to aggregation of the protein. We predicted that mutagenesis of the N-terminal cysteine would disrupt aggregation but leave the metal binding site intact. However, zinc-binding for S100A9(C3S) was remarkably similar to the wildtype protein (Fig. 4.2A, 4.2C). Unfortunately, this mutant was still aggregation-prone.

We next tested the ability of the C-terminal truncation mutant, S100A9Δ99, to bind zinc. Again, we observed aggregation of S100A9(Δ99) within the sample cell. However, we do observe a difference in the binding isotherm for the C-terminal truncation mutant (Fig. 4.2A, 4.2D), suggesting that the tail participates in zinc binding by wild-type protein. Truncation of the C-terminal tail appears to disrupt one of the processes involved in zinc-binding by S100A9.

Finally, we tested the full-metal site mutant of S100A9 – S100A9ΔZn. The ITC curve for zinc binding by S100A9(ΔZn) is markedly different from that of both wild-type S100A9 and S100A9(Δ99) (Fig. 4.2A, 4.2E). Mutagenesis of the predicted metal binding residues in S100A9 appears to fully disrupt zinc binding. Additionally, S100A9ΔZn does not aggregate. This suggests that the aggregation observed for S100A9 in the presence of zinc is mediated by metal binding at the canonical S100 transition metal binding site.

Zinc binding to S100A9 results in reversible aggregation

S100 proteins are aggregation-prone[170,174,175]. We and others have observed zinc-dependent aggregation of S100A9 both *in vitro* and *in vivo*[176]. This aggregation behavior by S100A9 has been proposed to play a role regulating TLR4 activation[170], though it is not known whether the aggregate is the active pro-inflammatory complex or if aggregation is a mechanism for dampening inflammatory activity of S100A9. Aggregation occurs instantaneously at near equimolar concentrations of zinc for protein >10 μM. This aggregation was observed for wild-type S100A9 but also occurs with the C3S mutant (Fig. 4.2B, 4.2C), indicating that the aggregation behavior is not simply dependent on the presence of an N-terminal cysteine as has been proposed for S100A2.

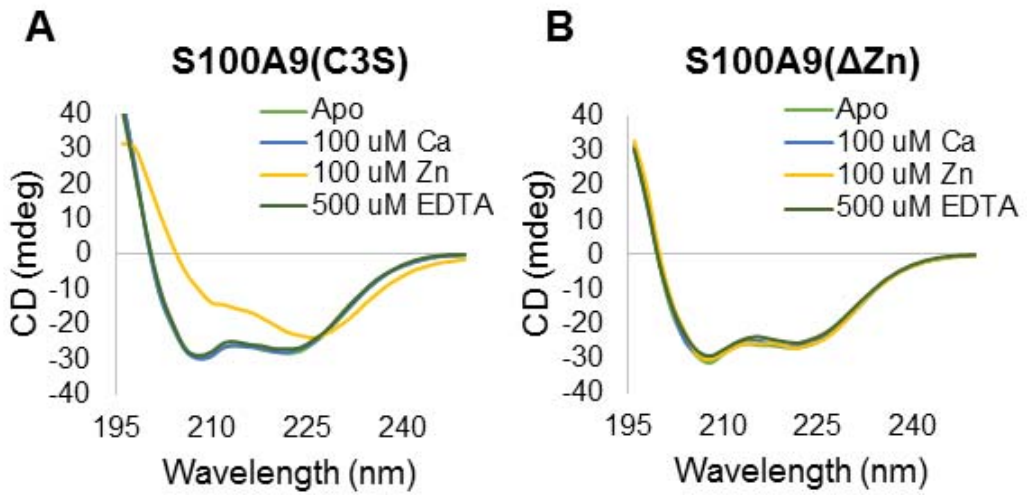

Figure 4.3. Reversible zinc-dependent aggregation of S100A9. A) Reversible aggregation of 20 μM S100A9(C3S) B) Aggregation is not observed for S100A9ΔZn. CD signal is shown for apo protein (light green) with the addition of 100 μM calcium (blue) and 100 μM zinc (yellow) by addition of 500 μM EDTA (dark green).

Our zinc-binding experiments had suggested an apparent loss of zinc-dependent aggregation of the S100A9(ΔZn). To examine this aggregation behavior more closely, we chose to use circular dichroism (CD). Protein aggregation is associated with a loss of

helical signal upon the addition of zinc. Interestingly, we found that aggregation of S100A9(C3S) is fully reversible with the addition of EDTA (Fig. 4.3A). Consistent with our ITC experiments, no aggregation was observed for S100A9(ΔZn) (Fig. 4.3B). This indicates that the aggregation behavior is a driven by zinc binding to the canonical transition metal binding site of S100A9. Given the lack of zinc-dependent aggregation observed for S100A9(ΔZn), this mutant allows us to test not only if zinc binding at this site is important for activation of TLR4 but also whether aggregation of S100A9 is the mechanism for zinc-dependence of TLR4 activation by S100A9.

C-terminal tail and metal binding site of S100A9 are not required for TLR4 activation

We set out to test two hypotheses for how zinc might regulate the interaction of S100A9 with TLR4. First, the unique-to-the-family flexible C-terminal extension of S100A9 is proximal to the putative protein target binding region and disordered in the absence of transition metal (Fig. 4.1). Structuring of the C-terminal tail upon transition metal binding has been observed for the heterodimer of S100A8 and S100A9[171–173]. This structural rearrangement is a possible mechanism for zinc regulation of protein interactions for S100A9: The structured C-terminal tail may directly comprise part of the target binding surface, or alternatively, the disordered tail may inhibit target interactions in the absence of zinc by occluding the target binding surface. Zinc might also regulate the interaction between S100A9 and TLR4 via a non-tail-dependent mechanism, such as oligomer formation or aggregation. We set out to test these hypotheses for zinc-dependent activation of TLR4 by S100A9 using our zinc binding mutant and our C-terminal truncation mutant.

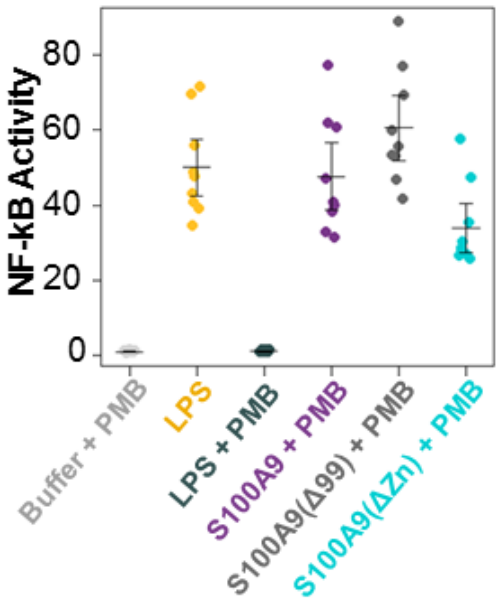

Figure 4.4. Zinc binding mutant is capable of activation of TLR4. We see no difference in activation of TLR4 following mutagenesis of the zinc binding site in S100A9. Activation by wildtype S100A9 (purple), S100A9(Δ99) (grey) and S100A9(ΔZn) (cyan) against human TLR4/MD-2/CD14 complex. Proteins were tested in the presence of polymixin B (PMB). Points show technical triplicates from biological replicates. Error bars show a standard error on the mean, mean is shown a horizontal line.

We observed that there was no difference in activation of TLR4 between wild-type S100A9 and S100A9(Δ99) (Fig. 4.4). This indicates that the C-terminal extension of S100A9 is not essential for activation of TLR4. However, we also considered that, in the absence of zinc, the C-terminal tail may be transiently obstructing the hydrophobic surface which has been shown in many S100s to be important for interacting with protein targets and has recently been shown to be important for the interaction between S100A9 and TLR4[169]. Consistent with this hypothesis, recent work examining the interaction of RAGE and S100A9 with HSQC has shown that binding of RAGE to S100A9 results in an increase in the peak intensities of the tail residues of S100A9[177]. For a disordered structure, this is consistent with a decrease in conformational dynamics. Binding to RAGE may push the tail out of the way of the target binding surface of S100A9.

If the tail is playing a transient role in obstruction of that surface, then we might expect truncation of the tail not to alter the interaction with TLR4. However, we may see a loss of activity if the tail was no longer able to bind to transition metals. Our metal binding mutant provided a means to test this hypothesis. To our surprise, we did not observe a significant difference in activation of TLR4 between S100A9 and S100A9(ΔZn) (Fig. 4.4). This indicates that metal binding at this site is not required for the interaction of S100A9 with TLR4. Further, this provides evidence that this zinc-dependent coordination of the C-terminal tail of S100A9 is not required for the interaction with TLR4 and that while the tail may occupy conformations that obstruct access to the hydrophobic surface in the absence of zinc, this is not restrictive for activation of TLR4. Finally, given the loss of zinc-dependent aggregation in this mutant, this indicates that aggregation does not appear to be the mechanism for zinc-regulation of the interaction with S100A9.

Activation of TLR4 by S100A9 occurs in the absence of zinc binding

Retention of activation of TLR4 by the full metal binding site deletion mutant was surprising given previous data suggests that zinc regulates the interaction with TLR4[68]. We wanted to validate that we had disrupted zinc binding under cell culture treatment conditions in this mutant and that activation of TLR4 was not dependent on the concentration of zinc in treatment conditions.

To validate that S100A9 binds to transition metals in TLR4 activation experiments and to confirm that we had disrupted zinc binding in the metal mutant, we tested the ability of S100A9 to strip metals from the protein treatment conditions. We

performed activation experiments in 25% buffer (20 mM HEPES, 100 mM NaCl, pH 7.4) and 75% media (Dulbecco's Modified Eagle Medium (DMEM) + 10% fetal bovine serum (FBS)). We prepared treatments as for the HEK cell culture assay and incubated these for 1 hour at 37°C with 5% CO_2. We used 3K filtration devices to separate protein from media. Using ICP-MS we determined the concentration of zinc in the filtrate (Table AC2). We compared the zinc concentration of untreated samples to samples treated with 2 μM protein. S100A9, S100A9Δ99, and S100A9(C3S) all show depletion of zinc from the media with respect to untreated sample, while S100A9ΔZn contains the same concentration of zinc as untreated (Fig. 4.5A). This indicates that wild-type protein is binding zinc under cell culture treatment conditions and that we have disrupted zinc binding in our metal site deletion mutant.

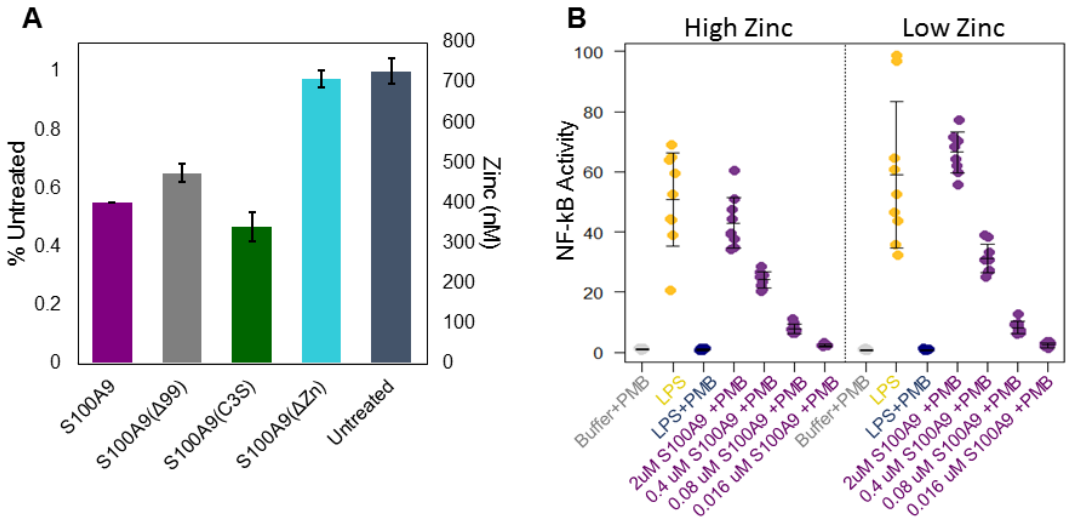

Figure 4.5. S100A9 activation of TLR4 does not require zinc binding. A) S100A9 binds to and depletes zinc from cell culture media. Zinc binding is disrupted for S100A9(ΔZn). Y-axis is calculated as % of zinc remaining as compared to untreated sample. Second y-axis shows average concentration of zinc detected in each sample. Error bars are standard deviation of two replicates. B) Activation of TLR4 by S100A9 in high and low zinc conditions. We tested activation by different concentrations of S100A9 in the presence and absence of 10% FBS which contributes zinc to the media. Points are technical triplicates from three biological replicates. Error bars show a standard error on the mean. Mean is shown as a horizontal line.

In earlier studies, we measured activation of TLR4 by S100A9 in media without FBS[90]. DMEM is a defined medium which does not contain any added zinc but has been shown to contain nanomolar levels of free zinc. We posited that there may not be sufficient zinc in the DMEM alone for S100A9 to be zinc-loaded. To ensure that there was sufficient zinc present within our experiments presented here, we chose to test activation with 75% DMEM + 10% FBS. The zinc concentration of DMEM + 10% FBS has been measured to be in the micromolar range. To determine if this difference in zinc concentration altered activation of TLR4 by S100A9, we compared activation of TLR4 by wild-type S100A9 in low zinc conditions (DMEM) and under conditions with high concentrations of zinc (DMEM with 10% FBS). We observed no change in the ability of S100A9 to activate TLR4 under these conditions (Fig. 4.5B). These results demonstrate that there is no difference in activation of TLR4 under high or low zinc conditions.

DISCUSSION

We find that zinc-binding involves many residues in S100A9. We show that there are two distinct processes for zinc-binding, one of which involves the C-terminal tail and the other does not. Additionally, we find that mutagenesis of the residues which participate in zinc binding by S100A9 in the heterodimer fully ablate metal binding by S100A9.

We also demonstrate that zinc is not required for activation of TLR4 by S100A9. Deletion of the transition metal binding site on S100A9 alters the ability of the protein to bind zinc under cell culture treatment conditions but does not alter the ability of S100A9 to activate TLR4. The lack of zinc-dependence for activation in this assay is surprising

given that it is contrary to previous studies which showed zinc-regulation of S100A9 binding to TLR4 *in vitro*[68]. However, our results here strongly suggest that the cell culture assay is measuring zinc-independent activation of TLR4 by S100A9.

Zinc-independent activation of TLR4 by S100A9 means that pro-inflammatory activity is separate from antibacterial activity of the protein. As part of the heterocomplex calprotectin, S100A9 is a potent antimicrobial protein through high-affinity metal chelation[65,171–173]. Our results here suggest that metal binding and therefore this nutrient sequestration-based antimicrobial activity is functionally separable from pro-inflammatory activity. The functional separation of these activities on S100A9 provides a means to piece apart the effect of these two functions of S100A9 during disease. One challenge in assessing the impact of S100A9 inhibition in hosts to limit inflammatory response, is predicting the overall effect of inhibition of S100A9's additional protective functions for the innate immune system. The antimicrobial activity of S100A9 may play a role in inhibiting proinflammatory response in hosts by controlling the spread of infection through nutrient sequestration.

We also show that the C-terminal tail of S100A9 is not required for the interaction with TLR4. Many biological functions both within cells and outside cells have been attributed to the flexible C-terminal tail of S100A9. The C-terminal peptide inhibits macrophage spreading and phagocytosis of adherent peritoneal lavage cells[178,179]. It is also directly involved in arachidonic acid binding and activation of oxidative burst[180]. The lack of involvement of the C-terminal in pro-inflammatory activity suggests that these functions, like antimicrobial activity, are functionally separable from TLR4-activation by S100A9. Further, given the exposure to proteases in the extracellular space,

one possibility is that in the C-terminal tail may be cleaved and functioning independently of S100A9. If the C-terminal tail is cleaved from S100A9 in the extracellular space, the independence of extracellular functions for the tail could be important for retention of pro-inflammatory activity post-cleavage.

Our results here suggest that the zinc-dependent aggregation is not required for activation by S100A9. Concentrations of zinc in the extracellular space are sufficient to induce aggregation of S100A9 *in vitro*, and these aggregates have been observed in hosts[170,176]. Understanding whether the aggregate was the active potently pro-inflammatory form of S100A9 or a built-in dampening mechanism for pro-inflammatory activity is an important step in understanding the pathology of S100A9. Further work is needed to address whether the zinc-induced aggregation results in a loss in pro-inflammatory activity. If so, Zn^{2+}-induced aggregation of S100A9 may not only be a mechanism for controlling the pro-inflammatory activity but also provide a pathway for sinking transition metals into abscesses. The heterodimer of S100A9/S100A8 has been found to be a large component of metal-protein abscesses found in infected tissue[181]. Aggregation could serve the dual-function by providing an initial pro-inflammatory response to control the spread of an infection and continuing to play a role in controlling the spread of infection through nutrient sequestration following aggregation.

Our experiments provide evidence of metal-independent activation of TLR4 by S100A9. Further work is needed to address the effect of additional perturbations to S100A9 on target binding interactions *in vitro*. Zinc-aggregation of S100A9 has previously been a technical challenge in these types of analyses, however, our observation of zinc-independence of TLR4 activation means direct binding studies can be

done in the absence of zinc. Finally, this work demonstrates that antimicrobial activity and arachidonic acid binding activity are functionally separate on S100A9 from pro-inflammatory activity, providing a future means to examine the independent roles of S100A9 in innate immune response.

MATERIALS & METHODS

Plasmids and recombinant protein preparation and characterization

Mammalian expression vectors were obtained from Addgene. The construct containing human TLR4 was made available by Ruslan Medzhitov (Addgene plasmid #13085). Constructs containing human CD14 and ELAM-Luc were obtained from Doug Golenbock (Addgene plasmid #13645 and #13029). Human MD-2 was obtained from the DNASU Repository (HsCD00439889).

A synthetic gene construct for human S100A9 (UniProt #P06702) was designed to be free of common restriction sites and codon optimized and purchased as PUC57 construct from Genscript (New Jersey, USA). S100A9 were cloned into a modified 6xHis MBP LIC TEV vector with NcoI and HindIII to yield a protein constructs with a TEV-cleavable histidine tag. Mutations to S100A9 were made using the QuikChange Lightning Mutagenesis Kit from Agilent Technologies (Santa Clara, CA).

To express S100A9 wild-type and mutants, *E. coli* BL21(DE3) pLysS competent cells containing the expression vectors for S100 proteins were grown overnight at 37ºC, diluted 1:150 into LB containing ampicillin and chloramphenicol, grown to mid-log phase (OD_{600} ~ 0.6-1), and induced with 1 mM isopropyl-β-d-1-thiogalactopyranoside overnight at 16º C with aeration. Bacterial pellets were harvested by centrifugation at

3000 rpm at 4° C for 20 min and stored at -20° C. Pellets (~5 g) were suspended by vortexing and lysed in 25 mL Buffer (25 mM Tris, 100 mM NaCl, 25 mM imidazole, pH 7.4) with 37.5 U DNase I (ThermoFisher Scientific) and 0.75 mg Lysozyme (ThermoFisher Scientific) by shaking at RT for approximately 1 hr. The lysate was clarified by centrifugation at 15,000 RPM for 50 min. at 4° C. Protein was purified using a 5 mL Ni^{2+}-NTA HisTrap column from (Healthcare GE) using an FPLC (Akta Biosciences) with gradient elution to HisB (25 mM Tris, 100 mM NaCl, pH 7.4, with 500 mM imidazole for human, mouse and chicken proteins, 1 M imidazole for opossum S100A9). Pooled elution peak of purified protein was cleaved with tobacco etch protease (TEV) overnight at RT. Cleaved protein was collected from a gradient elution of a 5 mL Ni-NTA column from HisA to HisB. Protein purity was assessed with SDS-PAGE and pure fractions were pooled and dialyzed into 20 mM HEPES, 100 mM NaCl, pH 7.4, treated with 10g/L chelex. For proteins containing cysteine (wild-type and S100A9Δ99, 0.5 mM TCEP was included in dialysis. Protein was flash frozen in liquid nitrogen and fresh aliquots were thawed and exchanged into fresh 20 mM HEPES, 100 mM NaCl, pH 7 (prepared with endotoxin-free water) with 3K Nanosep centrifugal devices prior to functional assays. While results shown here are replicates from a single protein prep, TLR4 activation was tested with two independent preps of each mutant to ensure that the results were not batch specific. Protein concentrations were measured using a Bradford assay.

For metal binding experiments, protein was exchanged into 20 mM HEPES, 100 mM NaCl, 500 μM TCEP, 200 μM $CaCl_2$, pH 7.4 and concentration was measured with a Bradford assay. Titrations were conducted at 25°C with stirring at 1000 rpm. The

titrations were performed all on the same day, with the same titrant (500 μM $ZnCl_2$, 20 mM HEPES, 100 mM NaCl, 500 μM TCEP, 200 μM $CaCl_2$, pH 7.4) in an ITC200 from MicroCal. Injections were 2 uL. Integrations were performed with Microcal Software.

For secondary structure measurements, protein samples were prepared at 20 μM in 10 mM Trizma, pH 7. Far-UV circular dichroism data was collected between 200–250 nm using a J-815 CD spectrometer (Jasco) with a 1 mm quartz spectrophotometer cell (Starna Cells, Inc. Catalog No. 1-Q-1). Assays for zinc-dependent aggregation were performed by adding 100 μM $CaCl_2$, then 100 μM $ZnCl_2$, and finally 500 μM EDTA. Aggregation was tested in duplicate for each protein. Raw ellipticity was converted into mean molar ellipticity using concentration and the number of residues for each protein.

Cell culture and transfection conditions

Human embryonic kidney cells (HEK293T/17, ATCC CRL-11268) were maintained up to 30 passages in DMEM supplemented with 10% FBS at 37º C with 5% CO_2. For each transfection, a confluent 100 mm plate of HEK293T/17 cells was treated at room temperature with 0.25% Trypsin-EDTA in HBSS and resuspended with an addition of DMEM + 10% FBS. This was diluted 4-fold into fresh medium and 135 μL aliquots of resuspended cells were transferred to a 96-well cell culture treated plate. Transfection mixes were made with 10 ng of TLR4, 1 ng of CD14, 0.5 ng of MD-2, 0.1 ng of Renilla, 20 ng of ELAM-Luc, and 68.4 ng pcDNA3.1(+) per well for a total of 100 ng of DNA, diluted in OptiMEM to a volume of 10 μL/well. To the DNA mix, 0.5 μL per well of PLUS reagent was added followed by a brief vortex and RT incubation for 10 min. Lipofectamine was diluted 0.5 μL into 9.5 μL OptiMEM per well. This was added to the

DNA + PLUS mix, vortexed briefly and incubated at RT for 15 min. The transfection mix was diluted to 65 µL/well in OptiMEM and aliquoted onto a plate. Cells were incubated with transfection mix overnight (18-22 hrs) and then treated with protein (0-10 µM) or LPS (100 ng/well) mixtures prepared in 25% buffer (20 mM HEPES, 100 mM NaCl, pH 7.4) and 75% cell culture media (DMEM + 10% FBS). *E. coli* K-12 LPS (tlrl-eklps, Invivogen) was dissolved at 5 mg/mL in endotoxin free water, aliquots were stored at -20° C. To avoid freeze-thaw cycles, working stocks of LPS were prepared at 10 ug/mL and stored at 4° C. There has been some concern in testing recombinant DAMPs against TLR4 due to the potential presence of contaminating LPS in proteins which have been expressed in bacteria [7,9]. We tested our S100s in the presence of 50 µg/mL polymyxin B, an LPS binding agent to limit signaling from LPS contamination in recombinant protein preparations. This concentration of polymyxin B eliminated signaling by 100 ng/mL of LPS but had a minimal effect on the signaling by S100A9. Cells were incubated with treatments for 4 hr. 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/Renilla luciferase activity.

Metal depletion experiments

Samples were prepared in the same way as HEK cell treatments. First, 150 uL of 8 uM protein was prepared in chelex treated 20 mM HEPES, 100 mM, pH 7.4, then protein was diluted 1:4 into DMEM + 10% FBS. Samples were incubated in cell culture incubator (37 C with 5% CO_2) for 1 hr and then filtered with 15 mL 3K microsep columns by centrifuging at 5000 xg for 20 min. Flow-through was collected and sent for

ICP-MS. ICPMS measurements were in the OHSU Elemental Analysis Core with partial support from NIH core grant S10RR025512.

For each sample, 100 µl of the filtered media solution was added to 1 ml of 1 % HNO3 (trace metal grade, Fisher) in a 15 ml acid-rinsed centrifuge tube (VWR). The final dilution factor was 11 (1100 µl / 100 µl). The concentrations listed in the supplement reflect the concentration in the original samples (Table AC2). Inductively coupled plasma mass spectroscopy (ICP MS) analysis was performed using an Agilent 7700x equipped with an ASX 500 autosampler. The system was operated at a radio frequency power of 1550 W, an argon plasma gas flow rate of 15 L/min, Ar carrier gas flow rate of 0.9 L/min. Elements were measured in kinetic energy discrimination (KED) mode using He gas (4.3 ml/min). Data were quantified using a 11-point (0, 0.5, 1, 2, 5, 10, 20, 50, 100, 500, 1000 ppb (ng/g) for Ca, Mn, Fe, Cu, and Zn using a multi-element standard. For each sample, data were acquired in triplicates and averaged. A coefficient of variance (CoV) was determined from frequent measurements of a sample containing ~10 ppb of Mn, Fe, Cu, and Zn. An internal standard (Sc, Ge, Bi) continuously introduced with the sample was used to correct for detector fluctuations and to monitor plasma stability. Elemental recovery was evaluated by measuring NIST reference material (water, SRM 1643f) and found to within 90 - 100% for all determined elements.

BRIDGE TO CHAPTER V

In this chapter, we described how we characterized zinc binding to S100A9 and determined that zinc binding is not required for activation of TLR4. We showed that zinc-dependent aggregation of S100A9 is dependent on the zinc binding to the canonical

S100 transition metal binding site and that disruption of this site does not alter TLR4 activation. We also demonstrate that the C-terminal tail of S100A9 is not required for TLR4 activation. This chapter employed traditional biochemical and functional assays to determine the role of zinc in S100A9 activation of TLR4. In Chapter V, we show how incorporating an evolutionary perspective in studies of S100A9, TLR4, MD-2, and CD14 have led to testable, mechanistic hypotheses for how these proteins function.

CHAPTER V

PHYLOGENETIC ANALYSIS LEADS TO MECHANISTIC HYPOTHESES REGARDING MECHANISM OF TLR4 ACTIVATION BY LIPOPOLYSACCHARIDE AND CALGRANULINS

This chapter contains co-authored material. Phylogenetic analysis and ancestral state reconstruction were performed by me. Experimental work was performed by me, J.L Harman, and G.L. Waddell. Figures were constructed by me and J.L. Harman. This excerpt was written by me, with portions of the methods section contributed by J.L. Harman and editorial assistance from Michael J. Harms.

INTRODUCTION

Evolutionary biology and biochemistry each offer unique insights into the diversity and complexity of biological systems. Employing these two perspectives simultaneously provides great advantages for understanding the how and the why in biology[182].

Early evolutionary biology was done exclusively at the organismal level. Initially, the potential value of molecular insights to evolutionary studies was largely ignored and even in some cases disputed[183,184]. However, by the early 1960s, the potential use of evolutionary theory to gain insight into molecular questions was gaining popularity. In 1961, Mayr argued that while "how" questions might be addressed with molecular methods, questions regarding "why" remain inaccessible[185]. To understand why a

biological phenomenon occurs, historical context is essential. Dobzhansky's famous quote, "Nothing in biology makes sense, except in light of evolution" also reflects this sentiment –mechanistic studies of biological molecules at the chemical and physical level are expanded by consideration of the evolutionary processes which drove their development [186].

Three primary paleomolecular methods are used to gain evolutionary insight into protein function: phylogenetics, ancestral state reconstruction (ASR), and co-evolutionary analysis[81]. Phylogenetics reveals the evolutionary relationship between extant proteins. Understanding the distribution of function across a phylogenetic tree leads to hypotheses for the timeline in which a new function arose. When coupled with sequence data, knowledge of the historical interval in which a function evolved can lead to mechanistic insight into how that function evolved[81]. ASR uses a phylogenetic tree and a multiple sequence alignment to infer the sequence of a protein at an ancestral node. Using ancestral state reconstruction, it is possible to elucidate the specific set of amino acid changes which occurred along a given interval and to reintroduce the minimal set of sequence changes into the appropriate historical background. These methods in combination permit one to retrace the historical steps that built a function onto a historical protein framework. This is different from traditional biochemical methods which identify important molecular components of a function by attempting to break that function in an extant protein. Finally, co-evolutionary analysis tests whether members of a protein complex are compatible with proteins from different organisms. This method can be used to determine if a function that is shared across a clade is mediated by a conserved molecular mechanism or if lineage-specific coevolution has occurred.

Over the past few decades, the use of these approaches has expanded, leading to mechanistic insights into protein evolution for a diversity of protein families[76–80,187]. Further, the generation of ancient proteins has allowed us to experimentally test hypotheses regarding ancient protein features and trends, such as stability and promiscuity[188–191]. Here we demonstrate the power of an evolutionary approach to biochemistry for the proteins addressed in this thesis: calgranulins, TLR4, MD-2, and CD14. The goal of this chapter is to convey the diversity of insight one gains from including an evolutionary perspective for biochemical studies, using different members of the complex as case studies.

POLARIZE TRANSITIONS WITHIN A PROTEIN FAMILY

From examination of modern proteins, it is not always clear when a transition of interest occurred. Defining the directionality of evolution of a function can reveal which function is "special" within a clade. For example, S100A8 and S100A9 form a heterocomplex calprotectin (CP) that is protease-resistant[192]. In contrast, homodimers of S100A8 and S100A9 are readily degraded by proteases[193,194]. Work has focused on determining the molecular basis for protease resistance of the heterodimer[192]; however, it could be that resistance is ancestral, and that rather than a gain of protease resistance in the heterodimer, the homodimers lost this feature. Distinguishing between these possibilities could allow us to identify the key mutations that are responsible for regulating protease susceptibility within this protein clade.

Ancestral state reconstruction of calgranulins reveals historical substitution important for proteolytic susceptibility of S100A9

In vivo, S100A9 is predominantly observed as a heterodimer. We hypothesized that homodimer degradation is the mechanism by which CP is selectively enriched *in vivo*. Looking at modern proteins, it is not clear whether a transition occurred from susceptible to resistant or resistant to susceptible? Phylogenetics and ASR provided a means to answer this question.

We set out to determine if proteolytic resistance was specific to the S100A8/S100A9 heterodimer or if S100s in general exhibit protease resistance. To distinguish between these possibilities, we tested the proteolytic susceptibility of S100 proteins across the phylogenetic tree. We chose to use proteinase K to test protease resistance as it is a robust protease which has previously been shown to degrade S100A8 and S100A9 homodimers but not CP. Additionally, proteinase K is a serine protease, mimicking neutrophil serine proteasses (NSPs) that get released at sites of inflammation. We observed that other S100s outside of the calgranulin clade of S100 proteins do exhibit resistance to degradation by proteinase K (Fig. 5.1). Further, S100A12 and MRP-126 exhibit protease resistance (Fig. 5.1). This indicates that proteolytic susceptibility appears to have evolved along the S100A8 and S100A9 lineages from a protease-resistant ancestor.

To determine when within the mammalian S100A8 and S100A9 lineages protease susceptibility evolved, we examined protease resistance of a marsupial S100A8 and S100A9. We chose to use opossum S100A8 and S100A9 proteins for this purpose. Examination of protease-susceptibility of opossum S100A8 and S100A9 revealed that

these extant proteins, like human S100A8 and S100A9, are susceptible to proteinase K (Fig. 5.1). This indicates that protease susceptibility of S100A8 and S100A9 appears to be shared across mammalian S100A8s and S100A9s. This suggests that the therian mammal ancestor of S100A8 and the therian mammal ancestor of S100A9 should also exhibit protease susceptibility. Direct measurement of protease susceptibility of these ancestral sequences would reveal if this activity evolved between amniote ancestor and the ancestor of therian mammals.

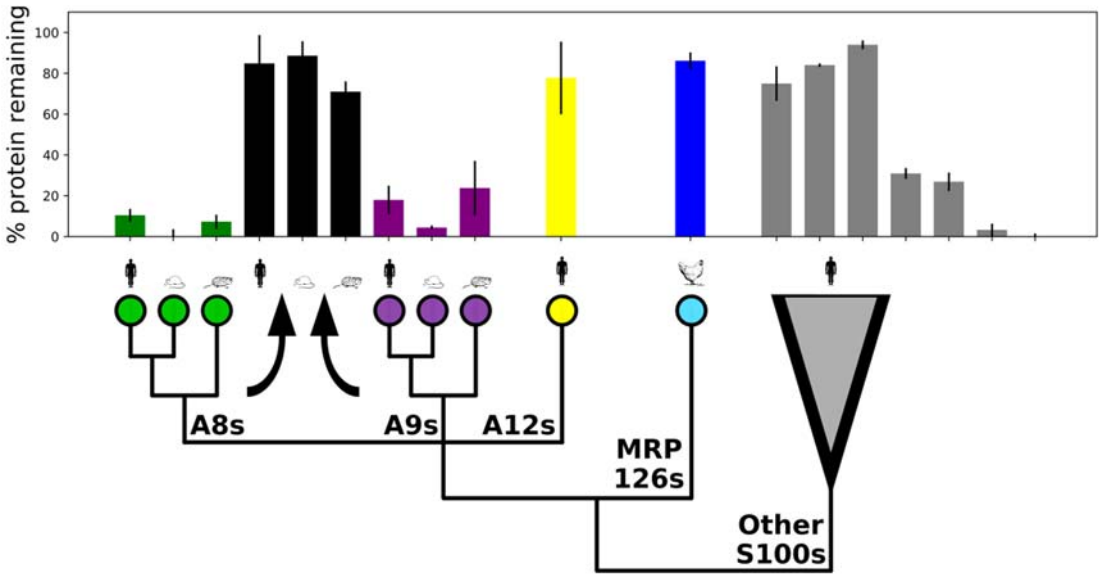

Figure 5.1. Human A8 and A9 homodimers uniquely lost proteolytic resistance in the S100s. A sampling of S100 proteolytic susceptibility across the phylogeny. Y-axis represents the percentage of the detectable protein remaining after 30 minutes of treatment with proteinase K. Bars represent average from multiple replicates. Error bars show standard error. Amniote A8s and A9s are readily degraded, while CPs are protease-resistant. Human A12 and chicken MRP126 are more resistant than A8 or A9 homodimers. 3 out of 7 non-calgranulin S100s tested were protease-resistant, while two exhibited intermediate resistance and two were rapidly degraded.

S100A8 and S100A9 evolved from a protease-resistant ancestor

To test the hypothesis that S100A8 and S100A9 evolved from a protease-resistant ancestor, we used ancestral state reconstruction to infer the sequence for an ancestral

calgranulin protein. One challenge in reconstructing the ancestor of S100A8 and S100A9 is that the exact topology within the calgranulin clade of S100 proteins is not able to be resolved[90]. We hypothesized that given that S100A8, S100A9, and S100A12 are all mammalian-specific proteins and that MRP-126 is specific to sauropsids that there may have been one ancestral calgranulin in early amniotes and that S100A8, S100A9 and S100A12 are the result of mammalian-specific gene duplication events.

There are two sources of uncertainty: topological and low signal at each site in the alignment. We investigated controls for both sources of uncertainty. Due to the difficulty in assigning a branching pattern within this clade of S100 proteins, we chose to reconstruct the sequences for ancestral calgranulins for species-corrected trees for all 15 potential topologies for the clade containing S100A8, S100A9, S100A12, and MRP-126. To address phylogenetic uncertainty, we chose to use an AltAll reconstruction[195], which contained substitutions to the next most likely residue at all sites where the second most likely residue had a posterior probability of >0.25. Examination of the sequences for the 15 possible reconstructions of ancestral calgranulin revealed that the phylogenetic uncertainty in the reconstruction was greater than the uncertainty in reconstructed sequences for the various topologies. Therefore, an AltAll reconstruction of the ancestral calgranulin tested not only the robustness of the functional measurement to phylogenetic uncertainty in the reconstruction but also tested the robustness of this measurement to the topological uncertainty in the tree.

We expressed and purified both the ancestral calgranulin and the AltAll reconstruction to determine if these proteins were susceptible to proteinase K degradation. Both the ancestral calgranulin (ancCG) and the AltAll ancestral calgranulin

(altancCG) were protease-resistant (Fig. 5.2). This is consistent with our hypothesis that S100A8 and S100A9 evolved from a protease-resistant ancestral state.

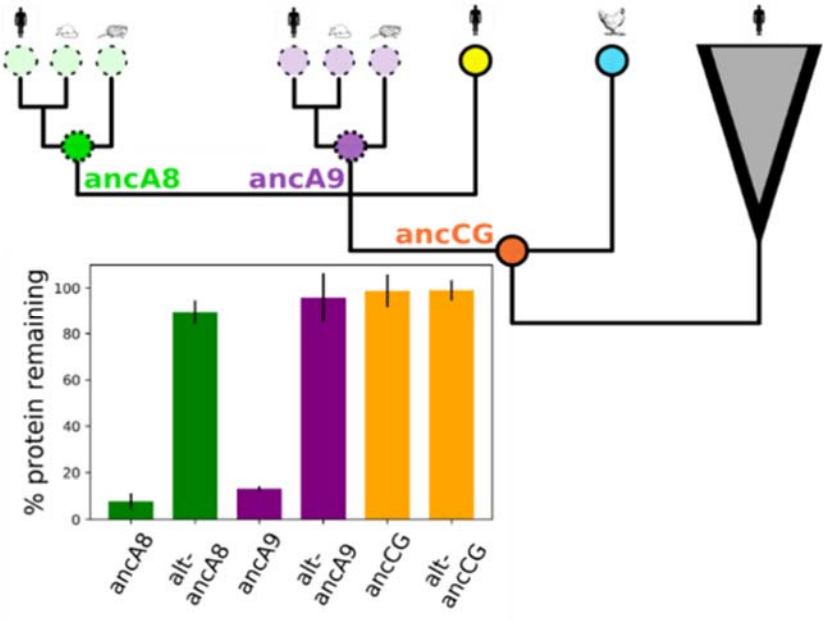

Figure 5.2. Proteolysis of ancestral calgranulins and ancestral S100A8 and S100A9 reveals proteolytic susceptibility evolved along the mammalian lineages for S100A8 and S100A9. Percentage of protein remaining following 30 min digestion with proteinase K is shown for the ML ancestral calgranulin (ancCG) and the AltAll (altancCG) in yellow. In addition, protease susceptibility for the ML (ancA8, ancA9) and the AltAll (altancA8, altancA9) for therian S100A8 (green) and S100A9 (purple). Cartoon tree is used to show which ancestors were used in this analysis. Y-axis represents the percentage of the detectable protein remaining after 30 minutes of treatment with proteinase K. Bars represent average from multiple replicates. Error bars are used to show standard error of measurements.

Proteolytic susceptibility for S100A8 and S100A9 evolved along mammalian lineages

Given that all extant mammalian S100A8 and S100A9s that we tested were protease susceptible, we hypothesized that the therian ancestor for S100A9 and S100A8 would also be susceptible to proteases. We therefore also reconstructed the ancestral sequences for therian S100A8 and S100A9 for the topology with the highest log-likelihood topology of the calgranulin clade. We expressed and purified these proteins, along with the AltAll reconstructions and tested for degradation by proteinase K. The

therian S100A8 and S100A9 were resistant to proteinase K, as expected (Fig. 5.2). However, the AltAll reconstructions were not resistant to proteinase K (Fig. 5.2). As the functional measurement for this ancestral node was not robust to phylogenetic uncertainty, we cannot assert that the protease-stability for this ancestral state with certainty. However, the difference in function between the ML reconstruction and the AltAll does provide a defined set of historical substitutions that occurred between the ancestral calgranulin and the extant S100 proteins that are useful for dissecting the evolution of protease susceptibility.

A single historical substitution restores protease resistance in human S100A9

For the therian ancestor of S100A9, there are 17 amino acid differences between the ML ancestor and the AltAll, 7 of these differences are in the C-terminal extension which is specific to S100A9. All or some subset of these substitutions must be responsible for the difference in the protease susceptibility between these two proteins. As the ancestral calgranulin and the AltAll calgranulin ancestor are both protease resistant, we chose to test only sites that are different between the ML and the AltAll ancestor of therian S100A9, but the same in the AltAll as in the calgranulin ancestor. This led to the identification of two potential residues that we hypothesized might be important for modulating this activity, N60D, and M63F. We mutated these sites and therian S100A9 (ancA9) and observed a gain of proteolytic resistance for this protein (Fig. 5.3A).

Next, we set out to test if either of these historical substitutions were able to revert human S100A9 back to the ancestral protease-resistant form (Fig. 5.3B). We started with

M63F as phenylalanine is present at this site in many extant mammalian protease-resistant S100s. We found that M63F was sufficient to induce protease resistance in human S100A9. Interestingly, S100A8 is protease susceptible but contains the ancestral phenylalanine at this position. This indicates that S100A8 independently evolved protease susceptibility by an alternate mechanism than S100A9.

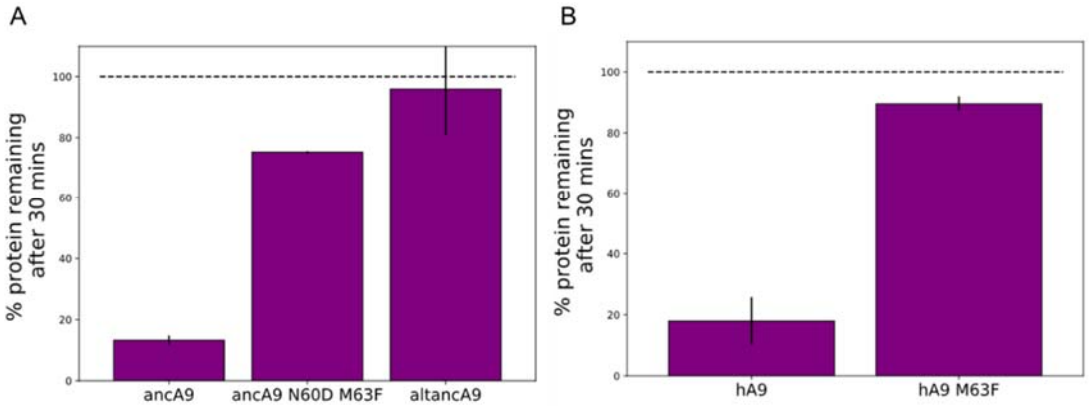

Figure 5.3. A single historical substitution altered protease susceptibility for S100A9. A. Introduction of N60D and M63F into ancestral S100A9 leads to protease-resistance, as observed in altancA9. B. Proteinase K susceptibility of human S100A9 is altered with the introduction of M63F. Y-axis represents the percentage of the detectable protein remaining after 30 minutes of treatment with proteinase K. Bars represent average from multiple replicates. Error bars are used to show standard error of measurements.

An evolutionary approach for the calgranulin proteins revealed that proteolytic susceptibility is a uniquely derived feature of the homodimers and not specific to the heterodimer. The result of this feature biologically is that the predominant form observed in hosts is the proteolytic-resistant heterodimer of S100A8 and S100A9, calprotectin. This work reveals that proteolytic resistance is not a unique feature of the heterodimer within the S100 family but rather that proteolytic resistance is an ancestral trait. We identified a single substitution that restores ancestral proteolytic resistance to human S100A9. Further, given that the ancestral residue is maintained at this site within modern

S100A8 it appears that proteolytic susceptibility evolved independently, by an alternate molecular mechanism along the S100A8 lineage.

IDENTIFY HISTORICAL SUBSTITUTIONS IMPORTANT FOR FUNCTION OF INTEREST

Sometimes, it is difficult to identify mutants of interest for a given function. If the molecular basis for a trait is not yet understood, it is often difficult to devise a directed experimental strategy towards an understanding of how a function works. Often, when faced with such a dilemma, one employs high-throughput methods to begin to break apart the function, which can often be insufficient, either because the modification has too small an effect (single point mutants) or to large an effect (massive deletions) which may alter more than just the given function of interest. ASR offers an alternative approach. One can focus on simply the set of historical substitutions which resulted in the evolution of a function of interest. This can lead to identification of the minimal set of substitutions necessary to alter a given function.

Identification of historical substitutions in the ectodomain of Toll-like receptor 4 that resulted in inhibition of ligand-independent activation

The highly conserved endodomain of TLR4 is a constitutive dimer[101,196,197]. Dimerization of the endodomain leads to activation of the NF-kB pro-inflammatory pathway[198]. However, in the absence of ligand, the ectodomain of TLR4 inhibits dimerization of the endodomain, thereby serving as a switch for ligand binding and ligand-dependent activation of the NF-kB pathway[196,199].

We hypothesized that inhibition of constitutive activity by the ectodomain increased along the human lineage. Mouse and opossum TLR4 complexes exhibit significant constitutive activity in a cell culture assay, while human TLR4 does not[200]. This distribution of function across the phylogenetic tree suggests that changes occurred along the human lineage which resulted in an increase in the ability of the ectodomain of TLR4 to inhibit ligand-independent dimerization of the endodomain.

Understanding the underlying molecular basis for inhibition of constitutive activity by the ectodomain of TLR4 is important for rational design of inhibitors of activation of this protein both in the presence and absence of ligand. Here, we used ancestral state reconstruction to identify substitutions which occurred on the human lineage, since the ancestor of humans and mice, which are involved in the inhibition of ligand-independent activation.

Horizontal mutagenesis for human and mouse TLR4 reveals epistasis

By testing N-terminal truncation mutants of human TLR4, a previous study found that a small segment of the extracellular domain (90 residues) is sufficient to inhibit ligand-independent activation of NF-kB cascade[196]. Examination of this region reveals two historical substitutions have occurred on the human lineage at sites where the ancestral state has been maintained in opossum and mouse. We hypothesized that these residues may be important for inhibiting ligand-independent activation in human TLR4. To test the role for these residues in regulating ligand-independent activation, we introduced the residues in human TLR4 at these sites into the mouse protein. Introduction of these substitutions into mouse TLR4 (C549Y and G559K) decreased constitutive

activation. This indicates that these sites are potentially important for ectodomain inhibition of constitutive activity in the human TLR4 (Fig. 5.4). Next, we introduced the mouse residues at these sites into human TLR4. These substitutions did not lead to an increase in constitutive activity (Fig. 5.4). This indicates that there are likely additional substitutions that occurred along this lineage that also regulate this activity.

Figure 5.4. Two sites which changed along the human lineage disrupt constitutive activity in the mouse TLR4 A. Alignment of 20-residue region previously shown to be important for ectodomain inhibition of constitutive activity for human, mouse, and opossum. Two sites are highlighted which changed specifically along the human lineage B. Introduction of the ancestral residue at these sites in the full-length and truncated flag-tagged human TLR4. Points represent technical triplicates from single biological replicate. Error bars are a standard error on the mean.

Inhibition of endodomain constitutive activity evolved on the human lineage

To test our hypothesis that this inhibition of constitutive activity evolved along the human lineage since the ancestor of humans and mice, we performed phylogenetic analysis and ancestral state reconstruction for TLR4. We compiled a database of 199 vertebrate TLR4 sequences using BLAST and created a multiple sequence alignment. Next, we constructed a species-corrected maximum-likelihood tree for vertebrate TLR4. Ancestral sequences were reconstructed using PAML[201], gaps were inferred based on parsimony and manually assigned.

This early reconstruction for the ancestor of human and mouse TLR4 contained the cysteine found in mouse and opossum in the first site of interest, but a glutamic acid in the second position, differing from both the mouse and opossum. The inferred sequence for TLR4 from the ancestor of human and mice was synthesized and cloned into a mammalian expression vector. We tested this protein for constitutive activation and observed that, like the mouse and opossum proteins, the TLR4 present in the ancestor of humans and mice was constitutive as well.

Historical substitutions are sufficient to disrupt inhibition of endodomain constitutive activity for truncated human TLR4

We next chose to introduce these changes into an N-terminally truncated human TLR4 to test if these substitutions altered inhibition of constitutive activity in a truncation mutant containing only the 90-residue region of the ectodomain which is sufficient for inhibiting constitutive dimerization of the endodomain. In this truncated protein, these reversions increase the constitutive activity of human TLR4 (Fig. 5.5). This suggests a historical importance for these substitutions in regulating the constitutive activity of the endodomain of TLR4.

This work is ongoing, I recently repeated the ancestral state reconstruction for TLR4 with additional sequence data to confirm the sequence within this region. This reconstructed ancestor contains the same residues at both sites of interest as observed in the mouse and opossum proteins, with high posterior probability. We plan to assess whether this ancestor also displays a level of constitutive activation consistent with that observed for mouse and opossum proteins. Further work is then needed to determine if a

glycine at position 561 is sufficient to disrupt the constitutive activity observed in the truncated human TLR4. We also plan to test the effect of the human substitutions in ancestral TLR4. This will reveal whether these substitutions are sufficient in the ancestral state to induce a reduction of constitutive activity.

Figure 5.5. Two historical substitutions which modulate constitutive activity. A. Structure of TLR4 with minimal 90-residue region sufficient for inhibition of constitutive activity highlighted in green and two residues shown in orange which were predicted to modulate this function B. Introduction of the ancestral residue at these sites in the full-length and truncated flag-tagged human TLR4. Points represent technical triplicates from single biological replicate. Error bars are a standard error on the mean.

Inhibition of ligand-independent activation by the ectodomain of TLR4 has been observed in the previous studies[196], but a hypothesis for the molecular mechanism for this trait was missing. Here, using variation in extant species to direct our study, we identified two sites which modulate this function. By identifying a minimal mutation set that alters this trait in extant TLRs we can now work to identify how inhibition of constitutive activity by the ectodomain of TLR4 is mediated. Is there protein partner which binds and inhibits activation of TLR4 in the absence of ligand? Do these sites stabilize a

conformation which prohibits dimerization of TLR4 in the absence of ligand? Future work is needed to investigate the mechanism by which these mutations alter this function.

REVEAL MOLECULAR BASIS FOR CONSERVED FUNCTION

Lineage-specific coevolution can lead to incompatibilities between extant proteins from different species. However, even when lineage specific coevolution has occurred, the core molecular basis for a function may be conserved. Co-evolution between protein partners can obscure the ability to identify what the molecular requirements are for the function of the complex. By employing methods which incorporate measurements of co-evolution within a system, we can disentangle these issues and isolate the core molecular mechanism of function.

Two residues are predictive of cross-compatibility between vertebrate TLR4 and MD-2s

TLR4 activation by LPS has previously been shown, by us and others, to be dependent on the presence species-specific co-factors[21,91,92,155,200]. This is most frequently observed for human TLR4 which has limited cross-species compatibility with other vertebrate MD-2s[23,85,86,111]. This lack of cross-specificity across vertebrate TLR4/MD-2 pairs indicates that lineage-specific co-evolution has occurred between these two proteins across the vertebrates. This co-evolution may have arisen, in part, due to the evolution of specificity toward different LPS chemotypes from different bacterial species.

Examination of TLR4/MD-2 pairs across the vertebrates reveals that the co-evolution between MD-2 and TLR4 is not a general trend across vertebrates but has occurred along specific lineages (Fig. 5.6). Opossum and mouse MD-2 are not only

compatible with each other but also with zebrafish TLR4. This indicates that like CD14, the molecular mechanism for MD-2 appears to be conserved across the vertebrates.

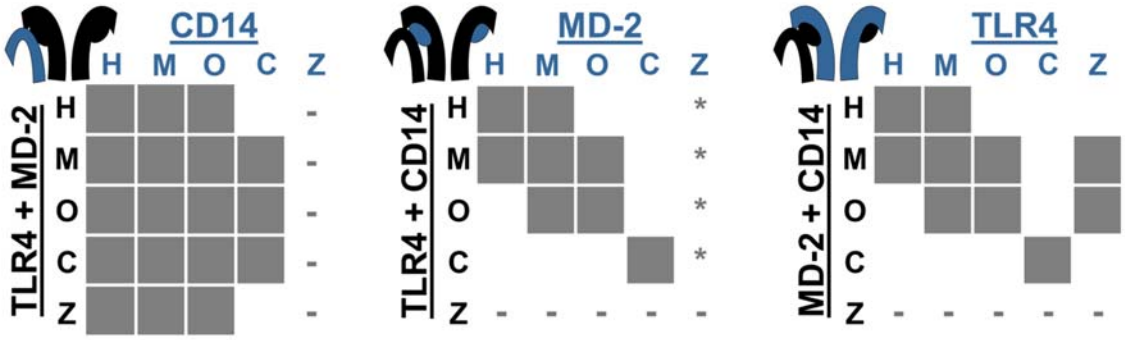

Figure 5.6. TLR4 co-factors MD-2 and CD14 show significant levels of cross-reactivity across vertebrates. Subpanels show activation of NF-κB signaling by receptor complexes assembled from host (black) and guest (red) proteins. Protein names on the left and top of each sub-panel indicate the components included. Letters indicate which species components come from human (H), mouse (M), opossum (O), or chicken (C). If the cross-species complex activates >2-fold above the buffer control, the corresponding box is filled gray. The icon to the top-left of each sub-panel indicates graphically which components are being combined. * indicates >2-fold activation above buffer control, but low total activation.

We looked for regions of MD-2 and TLR4 which are conserved across vertebrates to identify the molecular basis for compatibility of vertebrate MD-2s. We found that the interface between the N-terminal region of TLR4 and MD-2 is fully conserved across vertebrates (Fig. 5.7). Minimal sequence conservation is observed across vertebrate TLR4 and MD-2 sequences outside of this region.

Notably, despite being essential for LPS signaling, the dimerization interface between MD-2 and the C-terminal region of TLR4 is not conserved across species (Fig. 5.8). Differences within this region may account for the differences in LPS sensitivity observed across vertebrates. The composition of this loop is quite similar between mouse and human MD-2. This is surprising given the lack of compatibility of human MD-2 with opossum and zebrafish TLR4. Close examination of this loop of MD-2 reveals two

residues which are predictive of interaction cross-reactivity across the vertebrates (Fig. 5.8). These two prolines bracket and potentially structure the loop of MD-2 which interacts with the C-terminal region of TLR4.

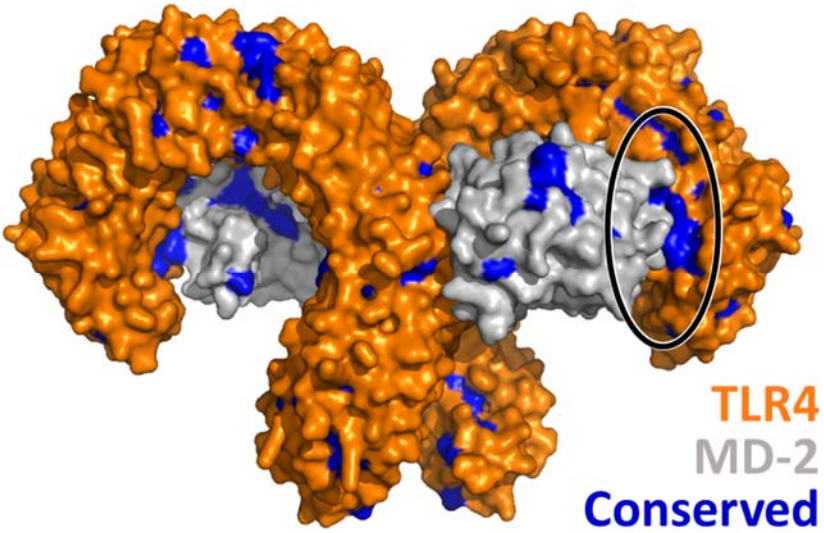

Figure 5.7 Interface between MD-2 and N-terminus of TLR4 appears to be conserved across the vertebrates. Structure of TLR4 (orange)/MD-2 (grey) complex (PDB: 3fxi) with residues which are conserved across vertebrates highlighted (blue).

Cross-compatibility studies with vertebrate MD-2 and TLR4 reveal that the molecular mechanism by which MD-2 functions is conserved across vertebrates. By examining substitutions which have occurred along specific lineage which lead to incompatibilities we were able to identify two sites which are predictive of cross-compatibility of MD-2 with vertebrate TLR4s. Testing if introduction of these residues into the human protein restores cross-compatibility of human MD-2 with opossum TLR4 will reveal if these two substitutions lead to species-specificity for TLR4/MD-2 pairs. Coevolution between TLR4/MD-2 pairs in vertebrates may have been driven by evolution for LPS specificity for different species. Further examination of how substitution of these sites in the human protein might lead to changes in species-specific recognition of LPS is needed.

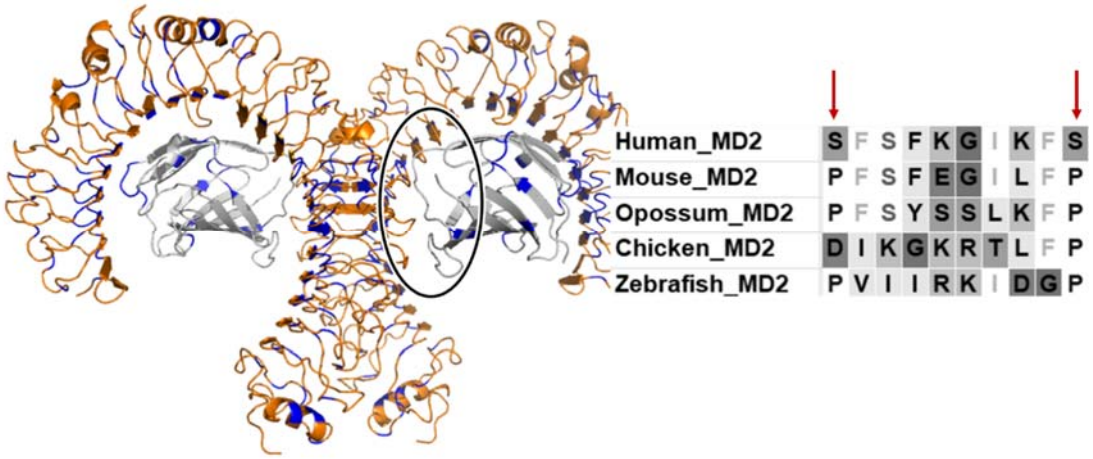

Figure 5.8 Dimerization interface between TLR4 and MD-2 important for activation by LPS is minimally conserved. A. Structure of TLR4 (orange)/MD-2 (grey) complex (PDB: 3fxi) with conserved residues highlighted (blue). Loop of MD-2 essential for dimerization of TLR4/MD-2 complexes and downstream activation is circled. B. Sequence of loop of MD-2 highlighted in A shows minimal conservation, two residues which are predictive of cross-compatibility of MD-2s across vertebrates are highlighted with red arrows.

DETERMINE THE HIERARCHY OF FUNCTIONS FOR A MULTI-FUNCTIONAL PROTEIN

Proteins often perform more than one function within a cell[202]. A given protein may have multiple protein partners or multiple protein functions. Understanding which of these functions is ancestral and when new functions evolved can reveal the hierarchy of functions. Phylogenetics can reveal when a new protein arose. Characterization of early branching extant proteins and examination of the functional profile of ancestral states can expose the order of evolution of functions for a protein of interest.

CD14 appears to have arose in early tetrapods

Sensitive recognition of LPS is important for early detection of pathogenic bacteria by the innate immune system. CD14 directly binds LPS and delivers it to the TLR4/MD-2 complex thereby increasing the sensitivity of immune cells to LPS[28,39,155].

CD14 has previously only been identified in amniotes[91,154,200]. However, here we show that it likely arose due to a duplication of TLR2 in early tetrapods.

CD14 has only been identified in amniotes and functional validation of sauropsid CD14 functioning similarly to mammalian CD14 is recent[91,200]. However, given our recent identification of a MD-2 in amphibians and fish, we revisited the question of when CD14 evolved. We identified a putative sequence for an amphibian CD14 annotated in the *Xenopus* genome. Using this sequence, we were able to identify additional amphibian CD14 molecules, but were not able unable to find a CD14 in fish, blasting with this sequence only revealed TLR2.

Duplication event resulting in CD14 appears to have occurred in early tetrapods

TLR2 and CD14 are sister clades (Fig. 5.9A). We reasoned that our inability to identify a CD14 in fish could be due to three alternative possibilities – 1. Fish lost CD14, 2. CD14 exists in fish but has significantly diverged and we are unable to identify it with a simple BLAST approach, 3. The duplication event that resulted in CD14 happened after the divergence of fish. To distinguish between these possibilities, we set out to determine when the duplication event occurred that resulted in CD14.

First, we constructed a database of TLR2 and CD14 proteins from chordates and constructed a maximum likelihood phylogeny, using the JTT model (Fig. 5.9B). The phylogeny for TLR2 and CD14 supports the hypothesis of a duplication event occurring after the divergence of fish. This is consistent with a model of a single TLR2 being present in sharks and fish with a duplication event occurring prior to the divergence of the tetrapod lineage.

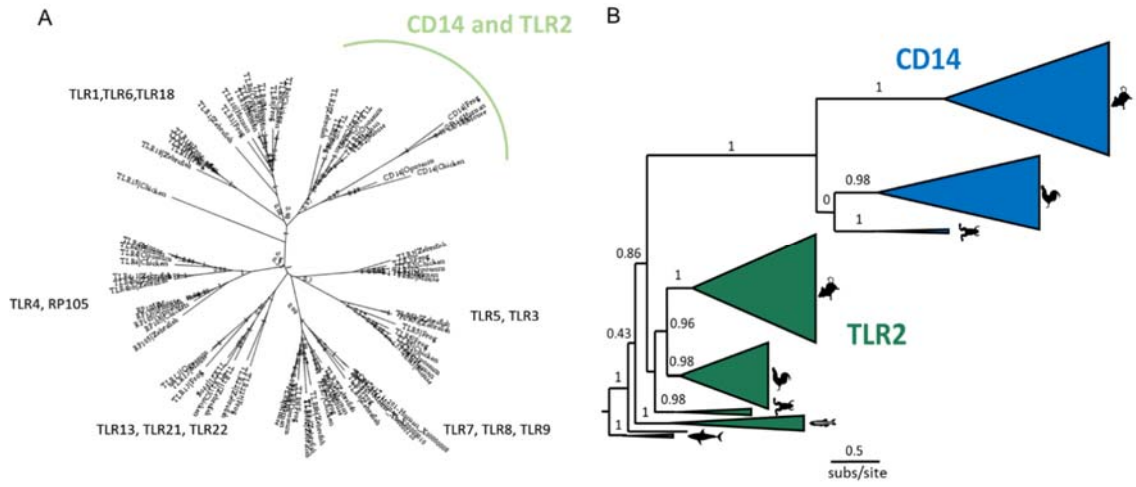

Figure 5.9. Phylogenetic analysis for CD14. A. Phylogeny for the TLR family. Phylogeny was constructed with all known TLRs from human, mouse, opossum, chicken, frog, and zebrafish. B. Phylogeny for vertebrate TLR2 and CD14. Phylogeny was constructed with all identified CD14s and TLR2s across vertebrates.

Difference in membrane linkage between mammalian and sauropsid CD14 is due to serial truncation of the endodomain

The phylogeny for TLR2 and CD14 reveals significant divergence along the amphibian CD14 lineage (Fig. 5.9B). Previously, it has been shown that the membrane linkage mechanism for mammalian and sauropsid CD14 is different. Chicken CD14 is a transmembrane protein while mammalian CD14s are GPI-coupled to the membrane[92]. Examination of the multiple sequence alignment reveals that this difference in membrane linkage is likely due to serial truncations of the internal TIR domain of TLR2 in the evolution of tetrapod CD14 (Fig. 5.10). Amphibian and sauropsid CD14 contain a C-terminal domain which aligns to a significant portion of the TIR domain of TLR2 while mammalian CD14s are approximately 80 residues shorter, containing a minimal C-terminal domain.

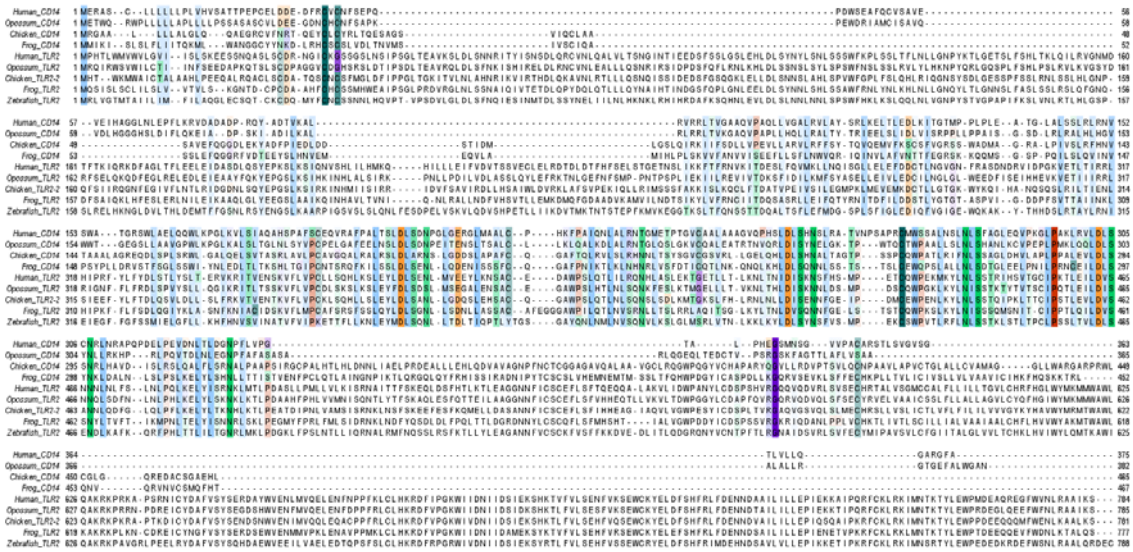

Figure 5.10. Alignment of TLR2 and CD14 from representative vertebrate species. Alignment was constructed for sequences from human, opossum, chicken, frog and zebrafish CD14 and TLR2. T-Coffee was used to present a structural alignment, gap in N-terminal region of CD14 was manually edited to reflect structural conservation of the N-terminal domain of TLR2 and CD14.

Pattern for sequence conservation in CD14 suggests ancient role in LPS-sensing

We have previously observed high cross-specificity of CD14s across the amniotes[200]. This is surprising given how much divergence has occurred along the mammalian lineage, as evidenced by the long-branch length for mammalian CD14s since the amniote ancestor. We tested whether the high level of divergence observed in the phylogeny was due to differences in the membrane linkage by constructing a phylogeny with just the ectodomain of CD14. We again observed long-branch lengths indicative of significant divergence since the amniote ancestor in the ectodomain (data not shown).

Examination of conservation of surface residues across vertebrate CD14s reveals minimal conservation across the structure. We highlighted residues with greater than 80% conservation across the tetrapod CD14s and found that the largest region of conservation on the protein surface was in the negative patch in the concave surface of the C-terminal region of the CD14 solenoid structure (Fig. 5.11). This region has previously been shown

to be important for formation of a transient complex with the basic patch of LBP and stimulation of NF-kB signaling[203]. Conservation of this region of CD14 across the vertebrates seems to suggest and early role for CD14 in LPS sensing.

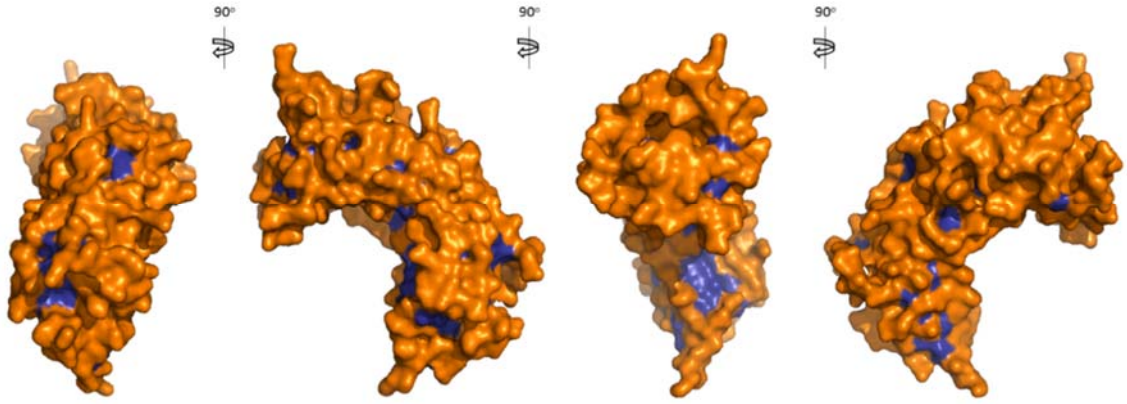

Figure 5.11. Minimal sequence conservation is observed across tetrapod CD14s. CD14 (PDB: 4glp)[28] is shown in orange, residues that are conserved across at least 80% of vertebrate CD14s are shown in blue.

TLR4 has previously been shown to be expressed in frogs[204], suggesting that LPS recognition in frogs may be occurring via this complex. We tested LPS recognition by the full complex of TLR4, MD-2, and CD14 from *Xenopus laevis* but did not observe activation of the full complex in an *ex vivo* cell culture assay. Given that we have previously observed some cross-specificity for individual components of the complex across vertebrates, we also examined whether any of the individual components, in combination with chicken TLR4, MD-2, and or CD14 could activate with LPS, we did not observe activation with *Xenopus* TLR4 or MD-2 in the presence of chicken complex components. We hypothesized that this could be due to a large amount of divergence within the signal peptide of frog TLR4. Further work is needed to assess whether use of a mammalian signal peptide on this receptor could lead to observations of activation within the cell culture system. Preliminarily experiments did show a slight increase in activation

of the chicken TLR4/MD-2 complex with the addition of *Xenopus* CD14 (Fig. 5.12). Additional work is needed to validate the ability of *Xenopus* CD14 to complement additional TLR4/MD-2 complexes to validate functionality of this complex.

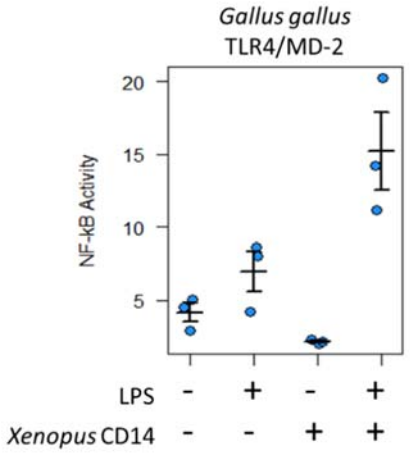

Figure 5.12. Frog CD14 increases LPS recognition by chicken TLR4/MD-2 complex. Preliminary analysis of LPS delivery by *Xenopus* CD14. LPS activation of chicken TLR4/MD-2 complex was assessed in the presence and absence of frog CD14. Points represent technical triplicate. Error bars are standard error with dash to represent mean for replicate.

In mammals, CD14 has been shown to have multiple functions in addition to its participation in LPS response[205]. One of these includes supporting peptidoglycan recognition by TLR2[206], which we hypothesis, given the shared binding motif for peptidoglycan between TLR2 and CD14 is likely the ancestral function of the molecule. Cross-reactivity has been observed across the amniotes for lipoprotein delivery by CD14, human CD14 can supports lipoprotein recognition by chicken TLR2[207]. This suggests that this role may not only be ancestral, but function via a conserved molecular mechanism. Further work may address which functions are ancestral to tetrapod CD14 and when LPS transfer and other functions evolved in CD14.

Finally, phylogenetic analysis revealed that CD14 may be tetrapod-specific. This is an important finding given the interest in presence of TLR4 co-factors outside of amniotes. Here we identify an amphibian CD14. We also show that the difference in membrane linkage between chicken and human CD14 is likely due to endodomain shortening along the mammalian lineage. This difference in membrane linkage may be important for generation of soluble CD14 which has been observed in mammals. Further work is needed to examine how this difference in membrane linkage affects function of CD14 between mammals and other tetrapods. Patterns of conservation in CD14 across tetrapods suggests that the role for LPS-sensing may be an ancient function of CD14. More work is needed to validate the role of CD14 as a TLR4-cofactor in LPS sensing in frogs.

DISCUSSION

Phylogenetics, ancestral state reconstruction, and co-evolutionary analysis are powerful approaches for studying biochemistry. Several different types of hypotheses were generated by applying these approaches to proteins within an innate immune pathway. We identified when important transitions occurred within protein families and the minimal historical substitutions needed to reproduce it in an ancestral background. In addition, we generated a testable hypothesis for the molecular basis of coevolution for a conserved protein complex. Finally, we showed how phylogenetic analysis can lead to identification of the hierarchy of functions for multi-functional protein.

This work highlights several insights for how proteins function and evolve. We show that single substitutions can have a large impact on protein function and that even

complex protein features can evolve rapidly with a small number of substitutions. We demonstrate that lack of cross-compatibility is not a good indicator of lack of function. Further, we show that identifying where lineage-specific coevolution has occurred can lead to insights into the molecular determinants of lack of cross-specificity and facilitate the parsing of co-evolutionary signal to reveal the core molecular determinants of function. Finally, we show how phylogenetic analysis can aid in identification of the early features of a multi-functional protein.

Homology is often used as an indicator of conserved functionality between species. Within the rapidly evolving immune system, conservation of function at the organismal level may not be indicative of conservation of molecular mechanism. Understanding how the molecular details of functions differ between species is essential when considering use of animal models for human disease. Coevolution between components of the TLR4/MD-2/CD14 complex have already led to challenges in identification and validation of antagonists of TLR4[27,85,86,89,111]. The methods described here provide a path towards identifying the conserved molecular details and mechanisms of activation, but also illuminating the molecular basis for species-specificity. These methods may also be applied towards other innate immune receptors, and may prove a powerful strategy towards disentangling differences between species in recognition of innate immune molecules.

These examples demonstrate the extraordinary power of considering evolutionary history in generating molecular hypotheses for protein function. Phylogenetics and ancestral state reconstruction led us to testable hypotheses for all four proteins addressed in this thesis: S100A9, TLR4, MD-2, and CD14. Merely by characterizing of extant

protein functions and examining the distribution of these functions across a phylogenetic tree we were able to identify historical substitutions which modulate these functions in extant proteins and delineate when new functions arose across within vertebrates.

MATERIALS AND METHODS

Phylogenetics and ancestral state reconstruction

For the S100 proteins, we constructed a curated database of homologs from a consistent set of species. We obtained amino acids sequences of S100 proteins from a subset of the amniote species in Ensembl version_87 [159]. We obtained additional bird and reptile S100A7 and MRP-126 sequences using the human calgranulin paralogs to BLAST against the NCBI database. This yielded a set of 172 sequences from 30 taxa (Data Sheet S1, S3). We constructed our multiple sequence alignment using MSAPROBS [119], followed by manual editing in MEGA [120]. We trimmed the alignment to remove highly variable (and therefore unalignable) C-terminal extensions, as well as the non-S100 domains of the fused S100 proteins. We used PHYML [121,122] with subtree pruning and re-grafting to construct the ML phylogeny. Pilot analyses revealed that the LG substitution model with 8 rate categories and a floating gamma distribution parameter yielded the highest likelihood trees [123–125]. An AIC test was used to control for overfitting [126]. We rooted our trees using the divergence of S100B, an ancient S100 found across jawed vertebrates. This allowed us to determine topology for the calgranulin clade with respect to S100A7, S100A15 and fused S100s. We species-corrected these clades using Archaeopteryx[208] according to NCBI taxonomic data[209,210], and published phylogenies of mammals[211], and sauropsids[212]. Given the difficulty in assigning topology within the

calgranulin clade using merely phylogenetic analysis. We constructed 15 possible trees for all possible topologies of A8, A9, A12, and MRP-126 and used these species corrected trees for reconstruction. The species corrected tree with the highest log-likelihood was used for ancestral state reconstruction of ancestral S100A8, S100A9 and ancestral calgranulins (Fig. AD1).

For TLR4, MD-2 and CD14 we constructed a curated database of protein sequences for each protein from an NCBI BLAST. Additional amphibian sequences were identified in amphibian transcriptomes available on dryad. We constructed multiple sequence alignments for each protein with MSAPROBS[119] followed by manual editing in MEGA[120]. Pilot analyses revealed that the JTT substitution model with 8 rate categories and a floating gamma distribution parameter yielded the highest likelihood trees[121,123–125]. We species-corrected these trees using Archaeopteryx[208] to reflect published vertebrate phylogenies[209–212]. Species-corrected trees are shown in the supplemental information for TLR4 (Fig. AD2), MD-2 (Fig. AD3), and CD14 (Fig. AD4).

To reconstruct ancestors, we used PAML[201,213]. We assigned gaps manually using parsimony. We generated the AltAll sequences for the therian ancestral S100A8 and S100A9 as well as the ancestral amniote calgranulin as described in Eick et al[195]. This incorporates uncertainty in the reconstruction by taking the next-best reconstruction at all ambiguous sites. We took each site at which the posterior probability of the next-best reconstruction was greater than 0.25 and the introduced that alternate reconstruction at the site of interest. These constructs were used in functional assays to assess if the functional observation at the nodes of interest was robust to phylogenetic uncertainty.

Protein expression and purification

All S100 genes in this study were purchased as synthetic constructs in pUC57 vectors from Genscript. Calgranulin 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 a pD vector, while A9s were cloned into MCS2. For expression and purification of A8/A9 heterocomplexes (CPs), pD plasmids containing an A8 gene in MCS1 and an A9 gene in MCS2 were used2. Marsupial A8s (opossum and the Tasmanian devil) were sub-cloned into an MBP-LIC vector to yield a His-MBP-TEV-A8 construct. For marsupial CPs, the entire His-MBP-TEV-A8 gene was then sub-cloned into MCS1 of a pD vector containing a marsupial A9 in MCS2. Non-calgranulin S100s (A1, A5, A7, A11, A14, P, and tunB) were previously cloned into a pET28/30 vector to yield a TEV-cleavable N-terminal His tag3. For cell culture assays, human codon-optimized S100 genes were sub-cloned into a pCDNA3 vector (Invitrogen). Finally, cysteine-free versions of all calgranulin genes, as well as point mutants, were prepared using site-directed mutagenesis (Agilent).

Recombinant proteins were expressed 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 + 7.5 ml of 40% glucose. 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). All non-calgranulin S100s were purified using a TEV-cleavable His tag strategy used by our lab previously[200]. All calgranulin S100s, with the exception of marsupial A8s and CPs, were purified in two steps using Ni-affinity chromatography followed by anion exchange chromatography. 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 cation exchange step was performed at pH 6 using the same elution strategy as for anion exchange.

Marsupial A8 and CP lysates were prepared as above and then flowed over a His column, eluting over a 25 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 + 1 mM EDTA. The MBP tag was then removed using amylose affinity chromatography in Tris buffer + 1mM EDTA, eluting stepwise with 10 mM maltose. 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.

Protease susceptibility assays

For all experiments, protein aliquots were thawed fresh from freezer stocks and dialyzed in the appropriate experimental buffer overnight at 4°C. Thawed aliquots were used for no more than several days (< 1 week) before discarding. All concentrations were measured by Bradford assay and correspond to uM dimeric protein. For in vitro proteolytic susceptibility experiments, proteins were exchanged into Tris buffer + 1 mM CaCl2. 12.5 uM dimeric S100 protein was treated with 5 uM Proteinase K from *Tritirachium album* (Sigma Aldrich) 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 using in-house gel analysis software (see GitHub).

Cell culture and transfection conditions

Human embryonic kidney cells (HEK293T/17, ATCC CRL-11268) were maintained up to 30 passages in DMEM supplemented with 10% FBS at 37º C with 5% CO_2. For each transfection, a confluent 100 mm plate of HEK293T/17 cells was treated at room temperature with 0.25% Trypsin-EDTA in HBSS and re-suspended with an addition of DMEM + 10% FBS. This was diluted 4-fold into fresh medium and 135 µL

aliquots of re-suspended cells were transferred to a 96-well cell culture treated plate. Transfection mixes were made with 10 ng of TLR4, 1 ng of CD14, 0.5 ng of MD-2, 0.1 ng of Renilla, 20 ng of ELAM-Luc, and 68.4 ng pcDNA3.1(+) per well for a total of 100 ng of DNA, diluted in OptiMEM to a volume of 10 µL/well. Additional TLR4 and MD-2 was needed for fish and amphibian constructs: 10 ng zebrafish MD-2 was used in experiments shown and 10 ng of human, mouse, opossum, and chicken MD-2 was therefore used for cross-complementation studies, in addition, to see activation with chicken TLR4/MD-2 we needed 10 ng of frog CD14 in transfections. To the DNA mix, 0.5 µL per well of PLUS reagent was added followed by a brief vortex and RT incubation for 10 min. Lipofectamine was diluted 0.5 µL into 9.5 µL OptiMEM per well. This was added to the DNA + PLUS mix, vortexed briefly and incubated at RT for 15 min. The transfection mix was diluted to 65 µL/well in OptiMEM and aliquoted onto a plate. Cells were incubated with transfection mix overnight (18-22 hrs) and then treated with protein (2 µM) or LPS (100 ng/well) mixtures (prepared in 25% phosphate buffered saline, 75% DMEM). *E. coli* K-12 LPS (tlrl-eklps, Invivogen) was dissolved at 5 mg/mL in endotoxin-free water, aliquots were stored at -20° C. To avoid freeze-thaw cycles, working stocks of LPS were prepared at 10 ug/mL and stored at 4° C. Cells were incubated with treatments for 4 hrs (or 6 hrs for studies done with zebrafish proteins). 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/Renilla luciferase activity, normalized to the LPS-treated transfection control for each species in order to normalize between plates. It should be noted that the level of constitutive activity was different between TLR4s from different

species. For constitutive TLR4 experiments, these studies were performed in the absence of TLR4 co-factors MD-2 and CD14.

BRIDGE TO CHAPTER VI

In this chapter, we showed how an evolutionary perspective has led to testable, mechanistic hypotheses for the four protein families described in this thesis. We described how the use of phylogenetics, ancestral state reconstruction, and co-evolutionary analysis can polarize functional transitions within a protein family, lead to the identification of residues important for a function of interest, determine the molecular mechanism for a conserved function, and determine the hierarchy of functions. In Chapter VI, the results from this entire work are summarized and the implications are further discussed.

CHAPTER VI

CONCLUDING REMARKS

The work presented in this thesis demonstrates the power of using a combination of evolutionary and biochemical approaches to learn how complex protein features and functions evolve. We determined when two types of agonists, pathogen-derived LPS and host-produced S100A9 evolved to activate TLR4. We found that both LPS and S100A9 activation of TLR4 evolved earlier than previously thought. We describe how we identified MD-2 and CD14, essential TLR4 cofactors for both S100A9 and LPS signaling, in amphibians and fish. We also show that an ortholog of S100A9 in birds and reptiles, MRP-126, is capable of activating TLR4. This work validates new model systems for studying human disease. By determining that endogenous activation of TLR4 occurs out to chickens, we demonstrate that amniote species contain the same endogenous positive feedback loop for TLR4 signaling that is observed in humans and mice. Further, by validating that zebrafish TLR4 recognizes LPS, we show that *Danio rerio*, a popular model organism for innate immune studies, could be used as a model for studying septic shock.

We also show that S100A9 activation of TLR4 does not require zinc. This means that antimicrobial activity, which is accomplished through zinc and manganese sequestration by S100s, is functionally separate from pro-inflammatory activity of this protein. We also show that the aggregation behavior of S100A9 in the presence of transition metals is mediated by metal binding to the canonical S100 transition metal binding site. Disruption of metal binding at this site allowed us to show that this

reversible, zinc-dependent aggregation behavior of S100A9 is not required for TLR4 activation. This suggests that zinc-dependent aggregation is not the mechanism by which S100A9 signals through TLR4. However, these results do lead to an alternate testable hypothesis, that aggregation of S100A9 may be a mechanism for dampening S100A9-driven pro-inflammatory activity.

Finally, we show how using co-evolutionary analysis, phylogenetics, and ancestral state reconstruction, led to mechanistic hypotheses regarding TLR4 activation by both LPS and S100A9. We demonstrate that these techniques can lead to identification of mutants that may be difficult to otherwise isolate. We also show how these methods can lead to identification of key features that are different between species and therefore aid in resolving complications in identification of suitable inhibitors due to difference between humans and animal models.

The hypotheses described here are merely a sampling of those that can be generated from phylogenetic data on these proteins. One promising area where these data could also be used to build mechanistic hypotheses is in identifying the molecular mechanisms of species-specific recognition of various chemotypes of LPS. Others have identified that MD-2 plays a role in species specific recognition of LPS molecules and have isolated some larger regions of TLR4 which are important for the variance observed between extant species[85–87,111,214,215]. By applying an evolutionary perspective, it may be possible to isolate the evolutionary interval within which these specificities arose. Further, examination of the profile of LPS recognition by ancestral TLR4s could potentially reveal what types of chemotypes of LPS were recognized by early vertebrate TLR4 complexes.

This work forms the basis for a number of ongoing projects in the Harms lab. Further characterization of the molecular mechanism by which S100A9 binds to metals and to protein targets is ongoing. Additional work is being conducted to identify where S100A9 binds to TLR4 and whether the same or different regions of TLR4 are important for LPS and S100A9 activation. Finally, further work is being done to understand evolution of new functions within these protein families. Overall, the work presented here not only gave us insight into the evolution and molecular mechanisms of TLR4 activation by LPS and S100A9 but will provide the basis for evolutionary biochemical studies of this important signaling pathway in years to come.

APPENDIX A

SUPPLEMENTAL MATERIAL FOR CHAPTER II

This section includes the figures and tables referenced in Chapter II. The experimental work was completed by me with editorial assistance by M.J. Harms.

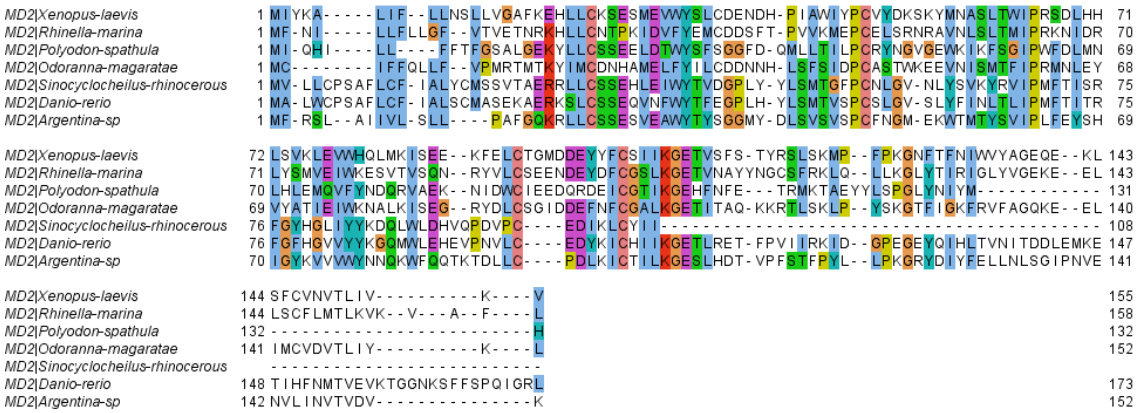

Figure AA1. Alignment of Newly Identified Fish and Amphibian MD-2 Proteins. MD-2 proteins were identified using transcriptomic data from amphibians (dryad) and fish (Fish1TK). Sequences were aligned with MSAPROBS. Colors are used to denote chemically similar residues and conservation.

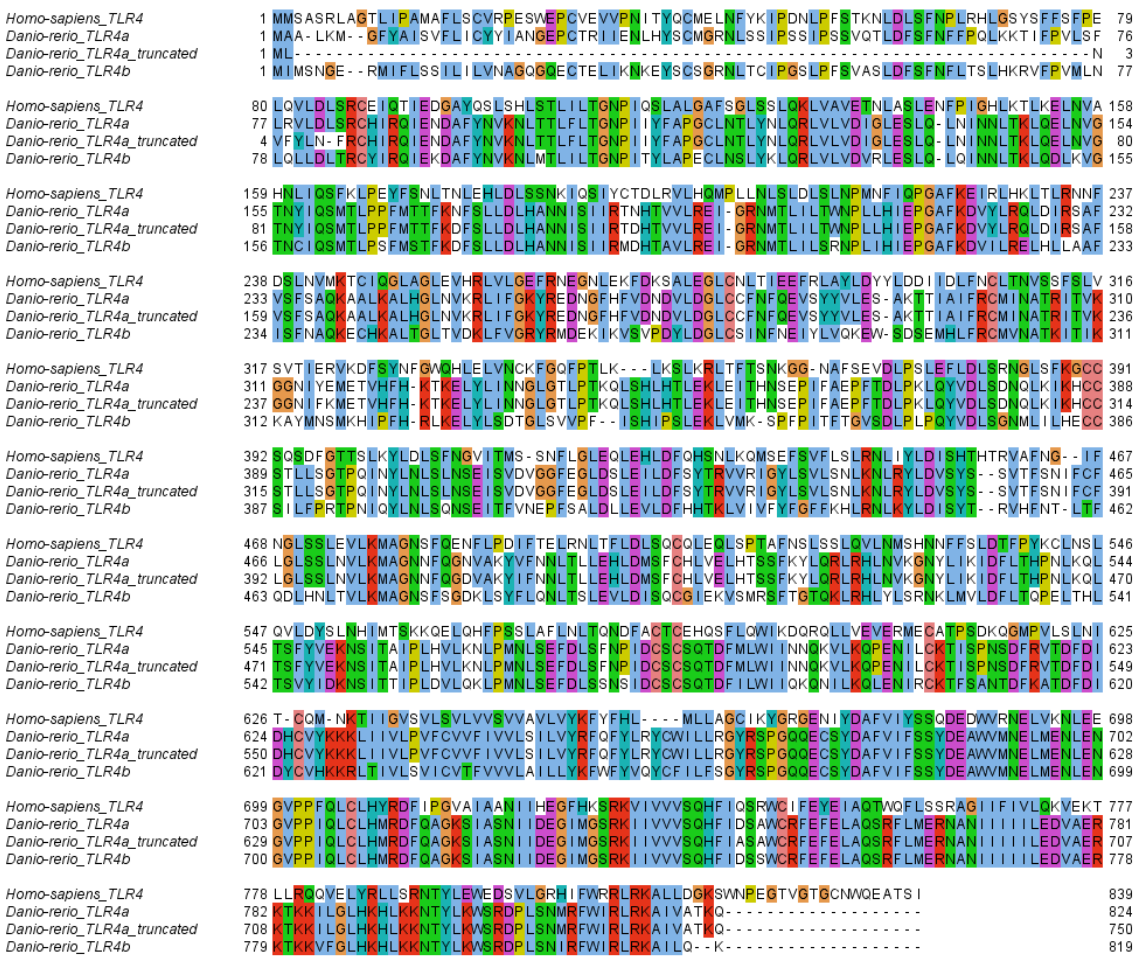

Figure AA2. Alignment of Zebrafish TLR4s with Human TLR4. Sequences were aligned with T-coffee. Colors are used to denote chemically similar residues and conservation.

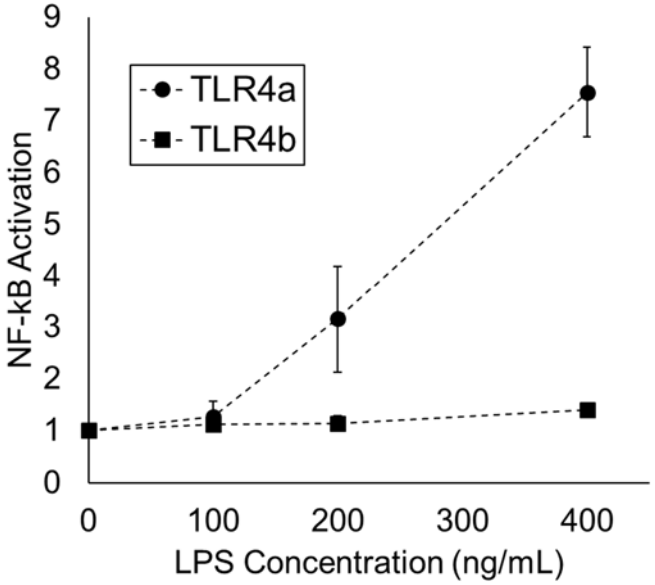

Figure AA3. Increasing LPS concentration leads to increasing response of zebrafish TLR4a in the presence of zebrafish MD-2 and human CD14, but not TLR4b. Points represent average activity detected at each concentration of LPS. Error bars are used to show the standard error for three replicates. Points are normalized to buffer control.

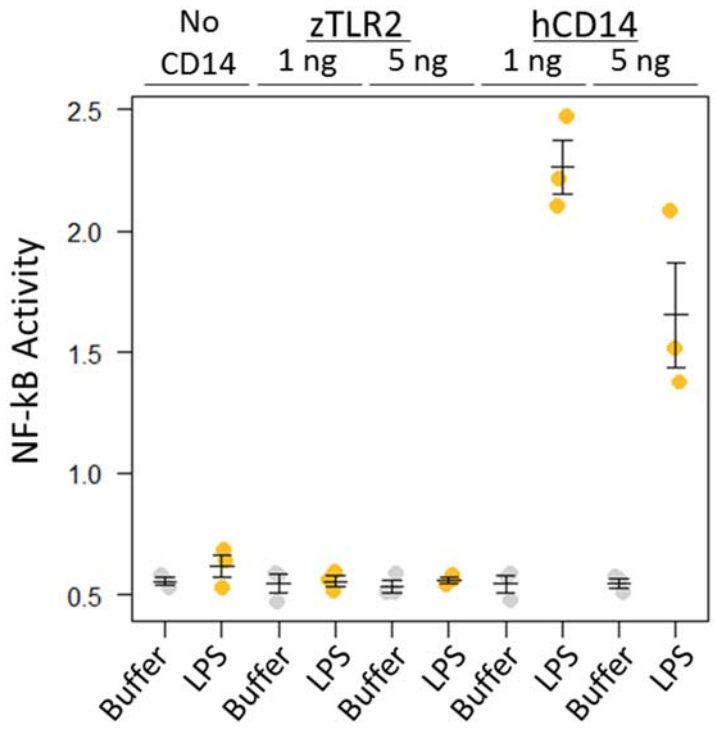

Figure AA4: Zebrafish TLR4a with MD-2 and TLR2. Activation of zebrafish TLR4a with MD-2 in the presence and absence of TLR2 or hCD14. Points are used to represent technical triplicates from three biological replicates. Error bars shown are a standard error on the mean.

Table AA1: **Genome locations for MD-2 proteins shown for synteny analysis.** Genomic locations were obtained from Ensembl database.

Species	Protein	Accession	Chromosome Location
Homo sapiens	ELOC	ENSG00000154582	8:73939169-73972287
Homo sapiens	TMEM70	ENSG00000175606	8:73972437-73982783
Homo sapiens	LY96	ENSG00000154589	8:73991352-74029087
Homo sapiens	JPH1	ENSG00000104369	8:74234700-74321328
Homo sapiens	GDAP1	ENSG00000104381	8:74321130-74488872
Homo sapiens	PI15	ENSG00000137558	8:74824534-74855029
Mus musculus	ELOC	ENSMUSG00000079658	1:16641725-16657042
Mus musculus	TMEM70	ENSMUSG00000025940	1:16665207-16678275
Mus musculus	LY96	ENSMUSG00000025779	1:16688051-16709611
Mus musculus	JPH1	ENSMUSG00000042686	1:16964560-17097889
Mus musculus	GDAP1	ENSMUSG00000025777	1:17145362-17164271
Mus musculus	PI15	ENSMUSG00000067780	1:17601901-17630939
Gallus gallus	ELOC	ENSGALG00000035417	2:118445078-118458839
Gallus gallus	TMEM70	ENSGALG00000033568	2:118458980-118463535
Gallus gallus	LY96	ENSGALG00000032701	2:118468953-118475495
Gallus gallus	JPH1	ENSGALG00000035845	2:118543324-118621765
Gallus gallus	GDAP1	ENSGALG00000042798	2:118637735-118644214
Gallus gallus	PI15	ENSGALG00000029264	2:118812721-118835079
Danio rerio	ELOCb	ENSDARG00000045359	2:30244376-30250311
Danio rerio	TMEM70	ENSDARG00000078773	2:30249977-30252347
Danio rerio	LY96	ENSDARG00000105462	2:30281444-30286249
Danio rerio	JPH1b	ENSDARG00000038826	2:30290719-30324610
Danio rerio	PI15b	ENSDARG00000061292	2:30368800-30376396

Table AA2: **Genome locations for TLR4 proteins shown for synteny analysis.** Genomic locations were obtained from Ensembl database.

Species	Protein	Accession	Chromosome Location
Homo sapiens	ASTN2	ENSG00000148219	9:116425225-117415070
Homo sapiens	TRIM32	ENSG00000119401	9:116687302-116701300
Homo sapiens	TLR4	ENSG00000136869	9:117704175-117724730
Homo sapiens	BRINP1	ENSG00000078725	9:119153458-119369467
Homo sapiens	CDK5RAP	ENSG00000136861	9:120388869-120580170
Mus musculus	ASTN2	ENSMUSG00000028373	4:65380803-66404611
Mus musculus	TRIM32	ENSMUSG00000051675	4:65604986-65616238
Mus musculus	TLR4	ENSMUSG00000039005	4:66827584-66930284
Mus musculus	BRINP1	ENSMUSG00000028351	4:68761514-68954397
Mus musculus	CDK5RAP	ENSMUSG00000039298	4:70216856-70410443
Gallus gallus	ASTN2	ENSGALG00000007038	17:3724742-3998741
Gallus gallus	TRIM32	ENSGALG00000034426	17:3818049-3819992
Gallus gallus	TLR4	ENSGALG00000007001	17:4083114-4088567
Gallus gallus	BRINP1	ENSGALG00000006997	17:4502517-4580960
Gallus gallus	CDK5RAP	ENSGALG00000041102	17:5006022-5105974
Xenopus tropocalis	ASTN2	ENSXETG00000021737	GL172772.1: 1,356,611-1,715,043
Xenopus tropocalis	TRIM32	ENSXETG00000021741	GL172772.1: 1,472,539-1,484,435
Xenopus tropocalis	TLR4	NA	GL172772:2026416- 2028407
Xenopus tropocalis	BRINP1	ENSXETG00000000244	GL172772.1: 2,180,027-2,198,495
Danio rerio	TLR4ba	ENSDARG00000019742	13:18514788-18517453
Danio rerio	TLR4a1	ENSDARG00000075671	13:18520738-18523378
Danio rerio	TLR4bb	ENSDARG00000022048	13:18523661-18528862
Lepisosteus oculatus	TLR4	ENSLOCG00000003751	LG21:5778346-5781017
Lepisosteus oculatus	BRINP1	ENSLOCG00000003763	LG21:6721875-6840675

APPENDIX B

SUPPLEMENTAL MATERIAL FOR CHAPTER III

This section includes the figures and tables referenced in Chapter II. The experimental work was completed by me. J.T. Bridgham contributed technical expertise. This excerpt was written by me with editorial assistance by M.J. Harms. This work was published as part of Loes, A.N., J.T. Bridgham, M.J. Harms. 2018. “Coevolution of the Toll-like receptor 4 complex with calgranulins and lipopolysaccharide”. Front. Immunol. 9:304.

Table AB1. Genome locations for proteins shown for synteny analysis in human, mouse, opossum, duck and chicken. Genomic locations obtained from Ensembl database.

Organism	ID	Protein	Genomic Location
Homo sapiens	ENSG00000163220	S100A9	1:153357854-153361027
Homo sapiens	ENSG00000163221	S100A12	1:153373706-153375649
Homo sapiens	ENSG00000143546	S100A8	1:153390032-153391188
Homo sapiens	ENSG00000184330	S100A7A	1:153416524-153423225
Homo sapiens	ENSG00000197364	S100A7L2	1:153437058-153439949
Homo sapiens	ENSG00000143556	S100A7	1:153457744-153460701
Homo sapiens	ENSG00000197956	S100A6	1:153534599-153536244
Homo sapiens	ENSG00000196420	S100A5	1:153537147-153541765
Homo sapiens	ENSG00000196154	S100A4	1:153543613-153550136
Mus musculus	ENSMUSG00000056071	S100A9	3:90692632-90695721
Mus musculus	ENSMUSG00000056054	S100A8	3:90668978-90670035
Mus musculus	ENSMUSG00000001025	S100A6	3:90612882-90624181
Mus musculus	ENSMUSG00000001023	S100A5	3:90608523-90611780
Mus musculus	ENSMUSG00000001020	S100A4	3:90603771-90606045
Monodelphis domestica	ENSMODG00000017406	S100A9	2:187687578-187692267
Monodelphis domestica	ENSMODG00000017410	S100A12	2:187668840-187671749
Monodelphis domestica	ENSMODG00000017403	S100A8	2:187727543-187728740
Monodelphis domestica	ENSMODG00000017402	S100A7	2:187744792-187750533
Monodelphis domestica	ENSMODG00000017400	S100A6	2:187846539-187848831
Monodelphis domestica	ENSMODG00000017397	S100A4	2:187864578-187865727
Gallus gallus	ENSGALG00000024272	S100A8	25:1885186-1886619
Gallus gallus	ENSGALG00000041826	S100A6	25:1874283-1875569
Gallus gallus	ENSGALG00000037599	S100A4	25:1868937-1871234
Anas platyrhynchos	ENSAPLG00000010189	MRP-126	KB742558.1: 42667-47205
Anas platyrhynchos	Unannotated	S100A7	KB742558: 39193-39333
Anas platyrhynchos	ENSAPLG00000010172	S100A6	KB742558.1: 32550-33337
Anas platyrhynchos	ENSAPLG00000010031	S100A4	KB742558.1: 30950-31363

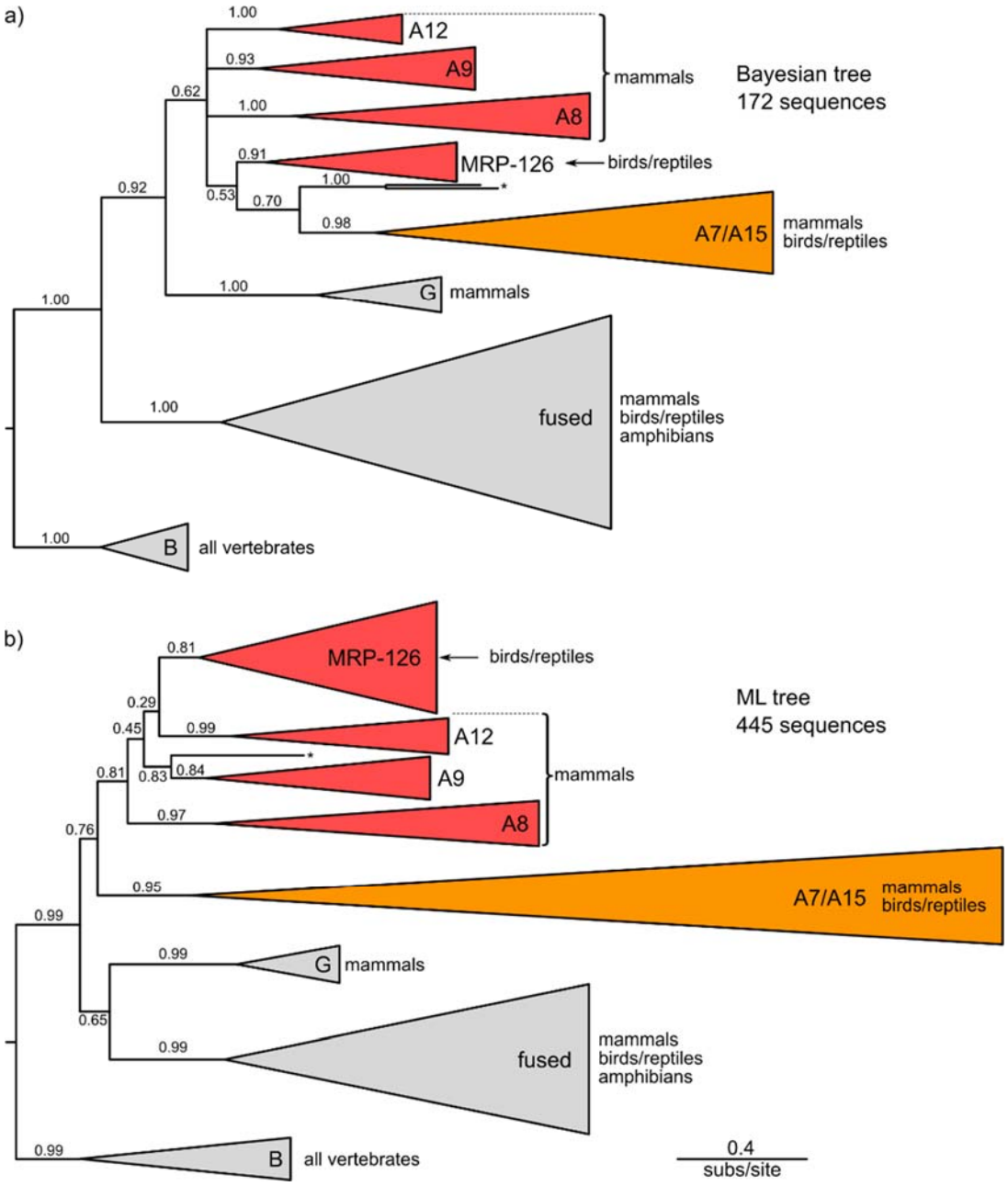

Figure AB1. Alternative phylogenies for a subset of the S100 family are consistent with MRP-126 designation as a calgranulin a) Bayesian phylogeny of 172 sequences and b) Maximum-likelihood phylogeny of 445 sequences for proteins within a subclade of the S100 family. C-terminal tails of sequences were truncated prior to resolving the phylogeny. Wedges are collapsed clades of orthologs, with wedge height corresponding to the number of included taxa and wedge length indicating the longest branch length with the clade. Support values are SH-supports calculated using an approximate likelihood ratio test. Clades are colored to highlight calgranulins (red) or S100A7/A15 (orange). The taxa included in each clade are noted on the tree. * indicates marsupial S100 proteins which according to genome location are likely S100A12.

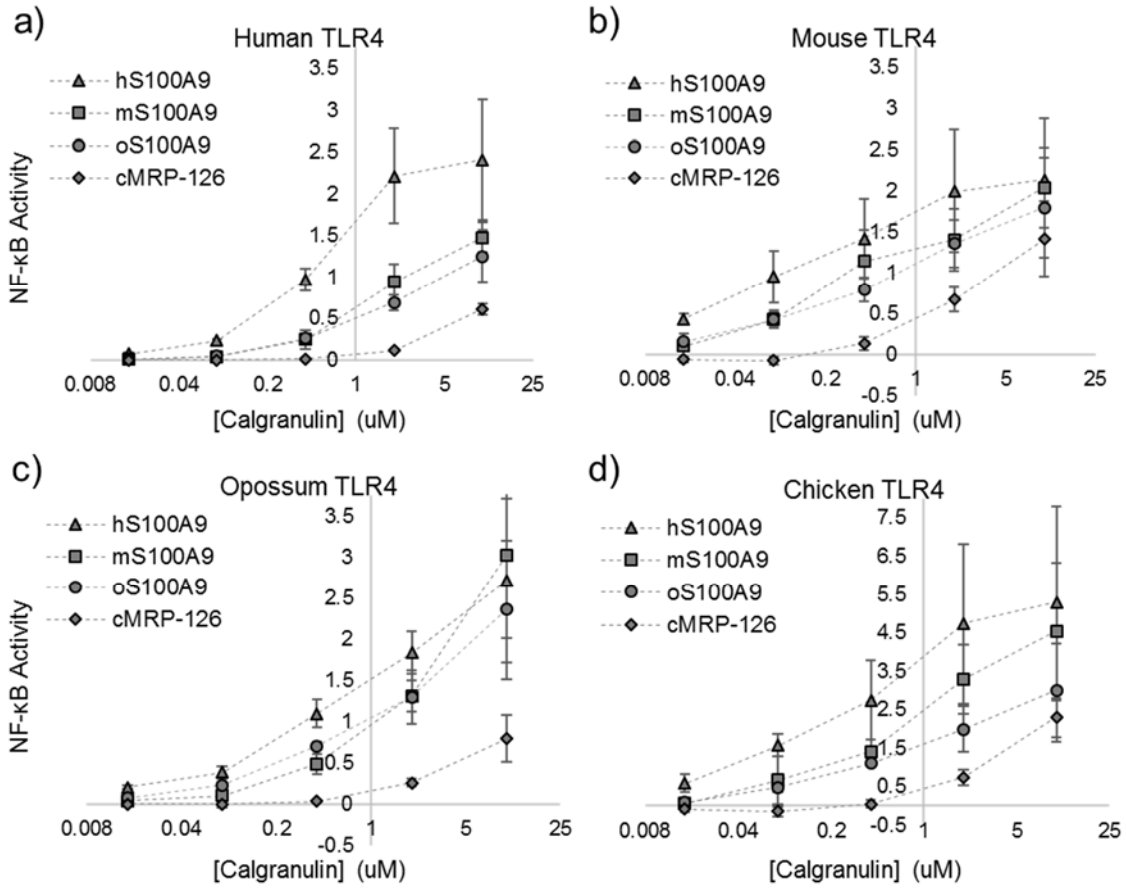

Figure AB2. Dose-response of calgranulins against alternate species TLR4/MD2/CD14 complexes NF-κB activity for calgranulins from human (▲), mouse (■), opossum (●), and chicken (♦) against TLR4/MD2/CD14 complexes from a) human, b) mouse, c) opossum and d) chicken. LPS was used as a positive control for expression and activation of the complex. Polymixin B is included to control for endotoxin-mediated activation of the complex. Activity is normalized to LPS activity of positive control within each biological replicate. Points are biological triplicates. Error bars for line plots are standard error of biological triplicates.

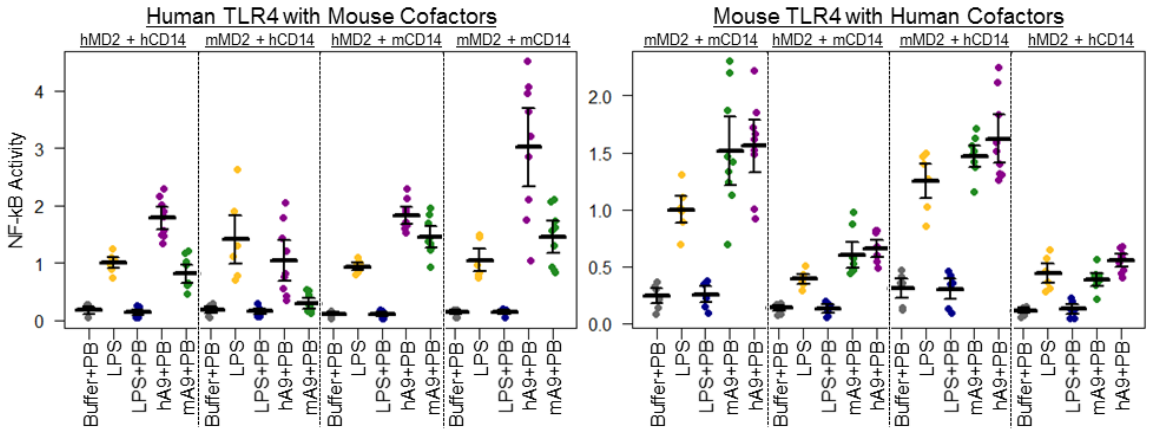

Figure AB3. Complementation of TLR4 cofactors: human vs. mouse. Activation of human and mouse TLR4 in the presence of human and mouse MD-2 and CD14 by both LPS and calgranulins. Transfection conditions are noted above each panel (i.e. mouse CD14 is shown as mCD14, and human CD14 is hCD14, respectively). NF-κB is normalized to LPS activation of the control complex for that species. Points are the technical triplicates from three biological replicates; error bars show standard error with mean shown as a bold line. Colors show treatment conditions: Buffer + PB (grey), LPS (yellow), LPS+PB (navy), human S100A9 (purple), mouse S100A9 (green).

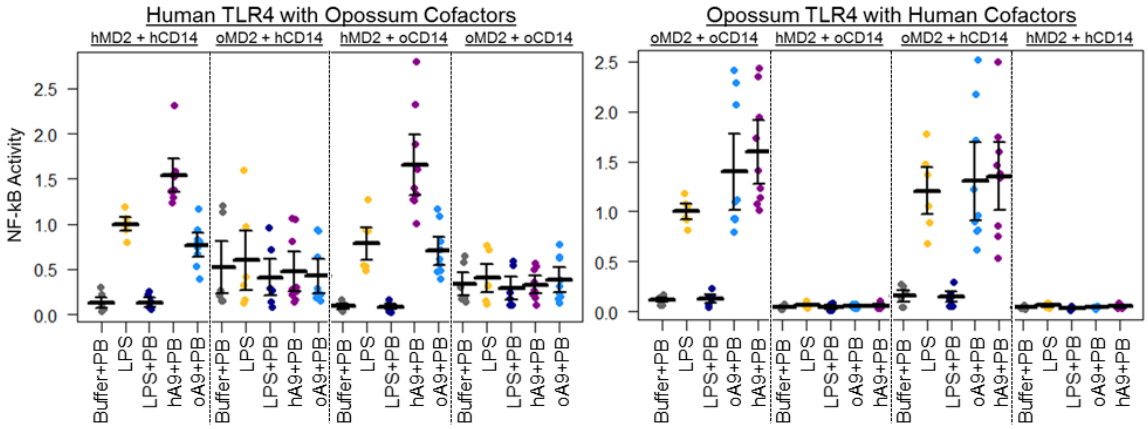

Figure AB4. Complementation of TLR4 cofactors: human vs. opossum. Activation of human and opossum TLR4 in the presence of human and opossum MD-2 and CD14 by both LPS and calgranulins. Transfection conditions are noted above each panel. NF-κB is normalized to LPS activation of the control complex for that species. Points are the technical triplicates from three biological replicates; error bars show standard error with mean shown as a bold line. Colors show treatment conditions: Buffer + PB (grey), LPS (yellow), LPS+PB (navy), human S100A9 (purple), opossum S100A9 (blue).

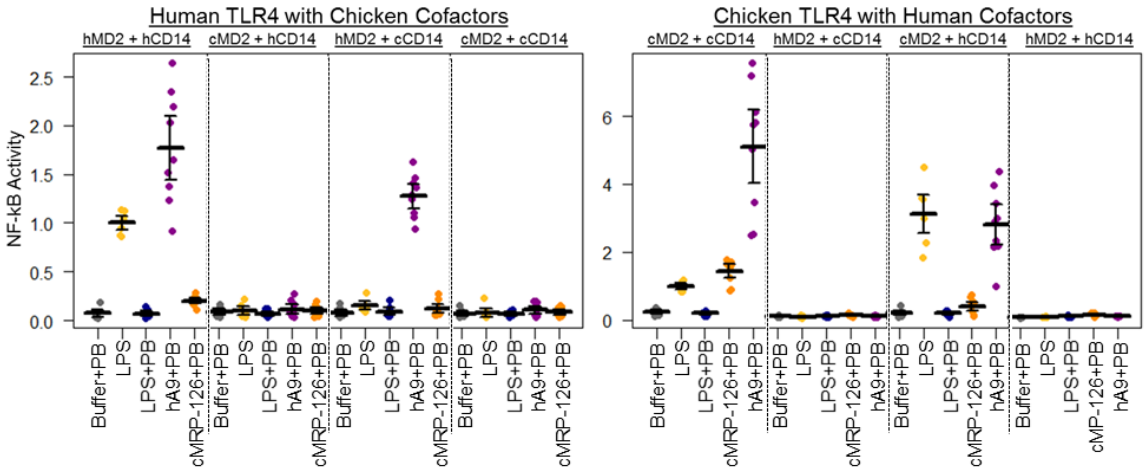

Figure AB5. Complementation of TLR4 cofactors: human vs. chicken. Activation of human and chicken TLR4 in the presence of human and chicken MD-2 and CD14 by both LPS and calgranulins. Transfection conditions are noted above each panel. NF-κB is normalized to LPS activation of the control complex for that species. Points are the technical triplicates from three biological replicates; error bars show standard error with mean shown as a bold line. Colors show treatment conditions: Buffer + PB (grey), LPS (yellow), LPS+PB (navy), human S100A9 (purple), chicken MRP-126 (orange).

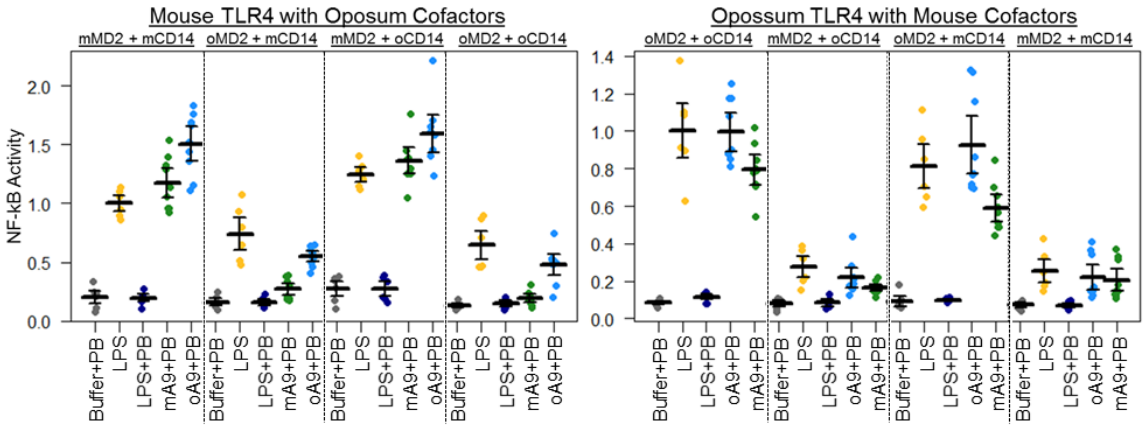

Figure AB6. Complementation of TLR4 cofactors: mouse vs. opossum. Activation of mouse and opossum TLR4 in the presence of mouse and opossum MD-2 and CD14 by both LPS and calgranulins. Transfection conditions are noted above each panel. NF-κB is normalized to LPS activation of the control complex for that species. Points are the technical triplicates from three biological replicates; error bars show standard error with mean shown as a bold line. Colors show treatment conditions: Buffer + PB (grey), LPS (yellow), LPS+PB (navy), mouse S100A9 (green), opossum S100A9 (blue).

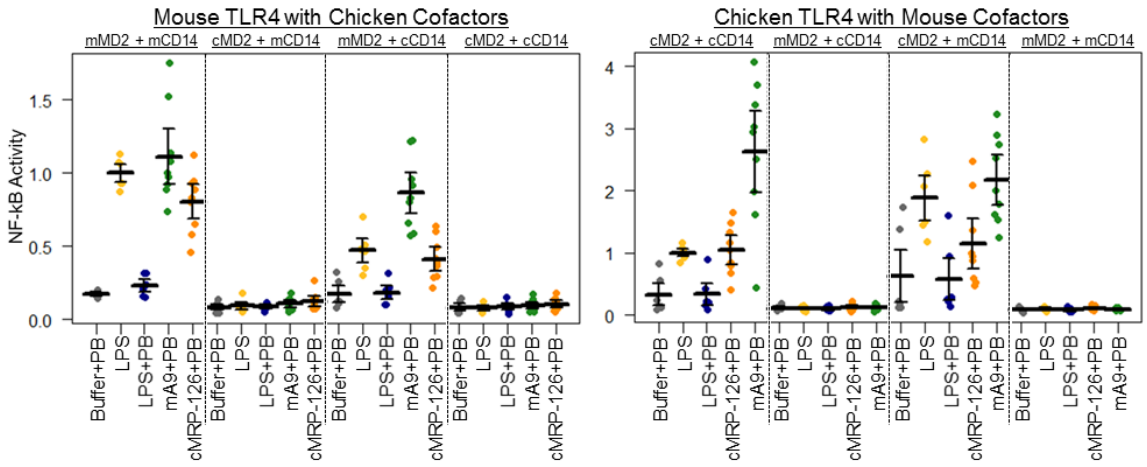

Figure AB7. Complementation of TLR4 cofactors: mouse vs. chicken. Activation of mouse and chicken TLR4 in the presence of mouse and chicken MD-2 and CD14 by both LPS and calgranulins. Transfection conditions are noted above each panel. NF-κB is normalized to LPS activation of the control complex for that species. Points are the technical triplicates from three biological replicates; error bars show standard error with mean shown as a bold line. Colors show treatment conditions: Buffer + PB (grey), LPS (yellow), LPS+PB (navy), mouse S100A9 (green), chicken MRP-126 (orange).

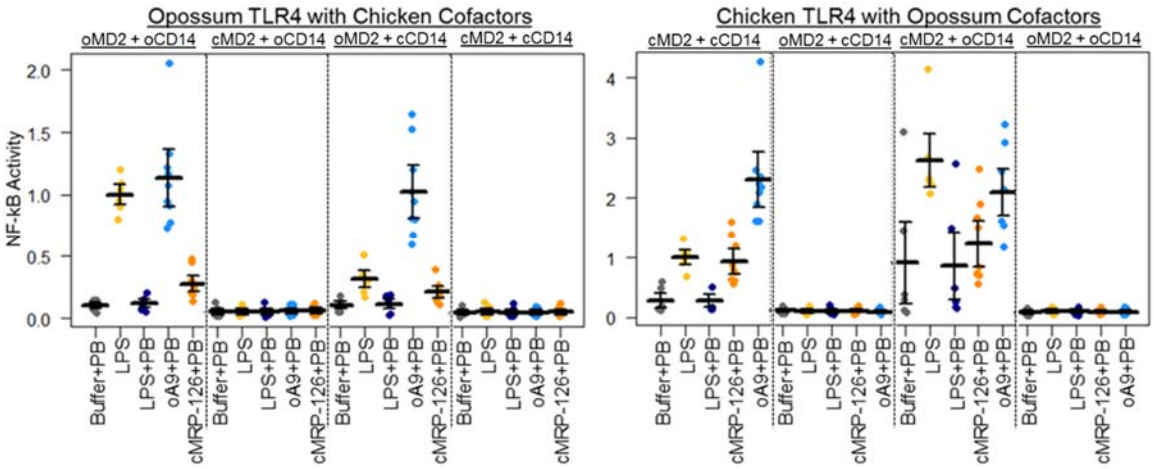

Figure AB8. Complementation of TLR4 cofactors: opossum vs. chicken. Activation of opossum and chicken TLR4 in the presence of opossum and chicken MD-2 and CD14 by both LPS and calgranulins. Transfection conditions are noted above each panel. NF-κB is normalized to LPS activation of the control complex for that species. Points are the technical triplicates from three biological replicates; error bars show standard error with mean shown as a bold line. Colors show treatment conditions: Buffer + PB (grey), LPS (yellow), LPS+PB (navy), opossum S100A9 (blue), chicken MRP-126 (orange).

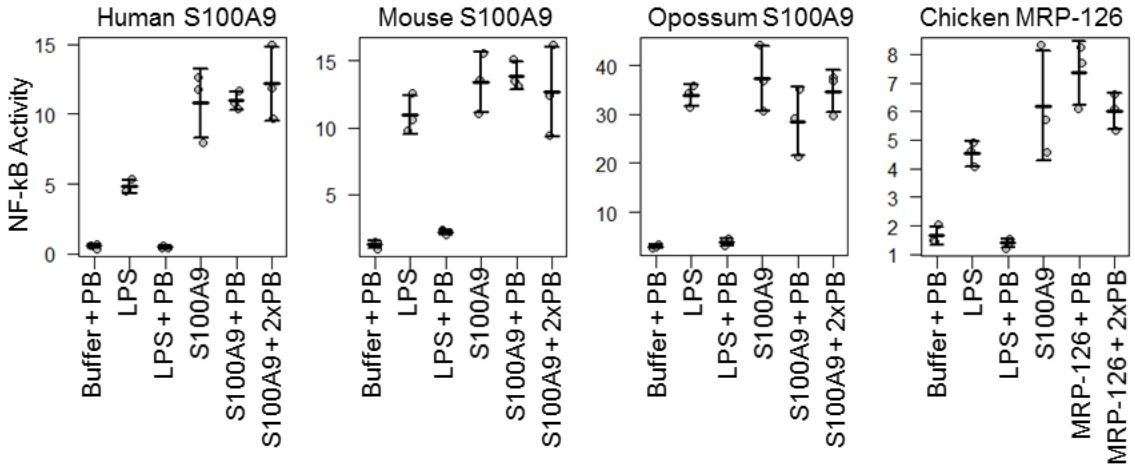

Figure AB9. Activation of TLR4 by recombinant proteins in the presence and absence of polymixin B. NF-κB activity for TLR4/MD2/CD14 complexes from amniotes treated with LPS (100 ng/mL), or 2 uM recombinant protein in phosphate buffered saline with and without 50 ug/mL or 100 ug/mL Polymixin B (PB). The ratio of Firefly luciferase to Renilla luciferase is shown. Points are technical triplicates from a single biological replicate. Bold line shows mean of technical replicates. Error bars are the standard deviation.

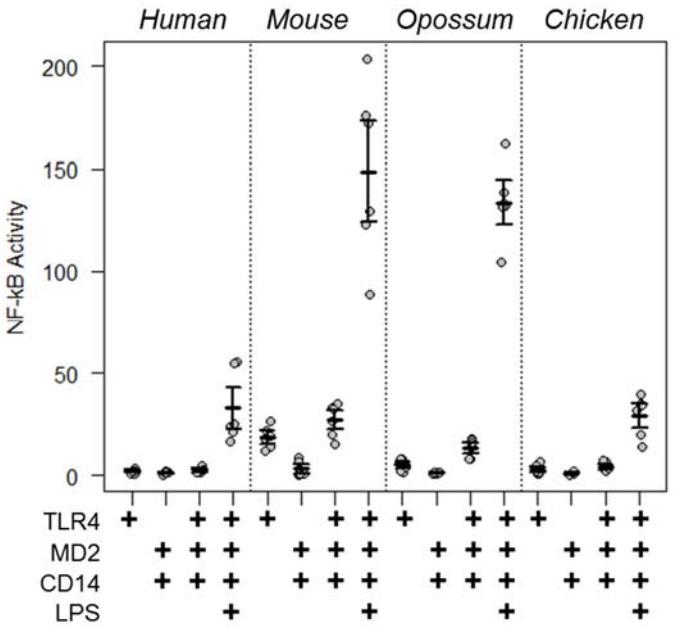

Figure AB10. Constitutive activity of TLR4 complexes from different species NF-κB activity for TLR4/MD2/CD14 complexes from amniotes treated with phosphate buffered saline (PBS) with and without LPS (100 ng/mL). Polymixin B (50 ug/mL) was included in buffer samples to control for endotoxin-mediated activation of the complex. A "+" in the table below each series indicates which components are included in that treatment. The ratio of Firefly luciferase to Renilla luciferase is shown. Points are the technical replicates from three biological replicates; error bars show standard error with mean shown as a bold line, "+" in the panel below indicates which components are included in the treatment.

APPENDIX C

SUPPLEMENTAL MATERIAL FOR CHAPTER IV

This section includes the figures and tables referenced in Chapter IV. This section contains co-authored material. The experimental work was completed by me and Ran Shi with supervised by me and M.J. Harms. This chapter was written by me with editorial assistance M.J. Harms.

Table AC1. Amino acid sequences for proteins used in this study. Wildtype protein sequence for human S100A9 was obtained from UniProt. Mutagenesis was performed with QuikChange lightning kit. Mutants were expressed in BL21(DE3) Rosetta pLyss cells.

Protein	Sequence
S100A9(WT)	MTCKMSQLERNIETIINTFHQYSVKLGHPDTLNQGEFKELVRKDLQNFLK KENKNEKVIEHIMEDLDTNADKQLSFEEFIMLMARLTWASHEKMHEGDE GPGHHHKPGLGEGTP
S100A9(C3S)	MT**S**KMSQLERNIETIINTFHQYSVKLGHPDTLNQGEFKELVRKDLQNFLK KENKNEKVIEHIMEDLDTNADKQLSFEEFIMLMARLTWASHEKMHEGDE GPGHHHKPGLGEGTP
S100A9(∆99)	MTCKMSQLERNIETIINTFHQYSVKLGHPDTLNQGEFKELVRKDLQNFLK KENKNEKVIEHIMEDLDTNADKQLSFEEFIMLMARLTWASHEKMHEGDE
S100A9(∆Zn)	MT**S**KMSQLERNIETIINTF**N**QYSVKLGHP**S**TLNQGEFKELVRKDLQNFLK KENKNEKVIEHIMEDLDTNADKQLSFEEFIMLMARLTWAS**N**EKM**N**EGDE GPG**NNN**KPGLGEGTP

Table AC2. S100A9 depletes metals from cell culture media. Metal concentrations identified with ICP-MS in DMEM and in DMEM + 10% FBS treated with 2 uM protein.

Sample	Rep	Ca [mM]	Mn [nM]	Fe [nM]	Cu [nM]	Zn [nM]
S100A9(WT)	1	1.122547	<20	<56	39.90628	405.7573
S100(Δ99)	1	1.128009	<20	<56	41.25464	465.7686
S100A9(C3S)	1	1.038788	<20	<56	39.8894	320.2949
S100A9(ΔZn)	1	1.103102	<20	<56	44.2999	705.4366
DMEM+FBS	1	1.12145	<20	<56	39.64886	716.2602
DMEM	1	1.053144	<20	<56	16.99926	<104
S100A9(WT)	2	1.076186	<20	<56	33.46436	407.3709
S100(Δ99)	2	1.103707	<20	<56	40.11286	497.2861
S100A9(C3S)	2	1.104335	<20	<56	39.28092	370.9397
S100A9(ΔZn)	2	1.08796	<20	<56	33.82844	736.1877
DMEM+FBS	2	1.063412	<20	<56	53.21472	761.2194
DMEM	2	1.043663	<20	62	43.80738	176.6499

APPENDIX D

SUPPLEMENTAL MATERIAL FOR CHAPTER V

This section includes the supplemental figures referenced in Chapter V. This section contains co-authored material. The experimental work was completed by me, Joseph Harman, and Grace Waddell, supervised by me and M.J. Harms. This section was written by me with editorial assistance M.J. Harms.

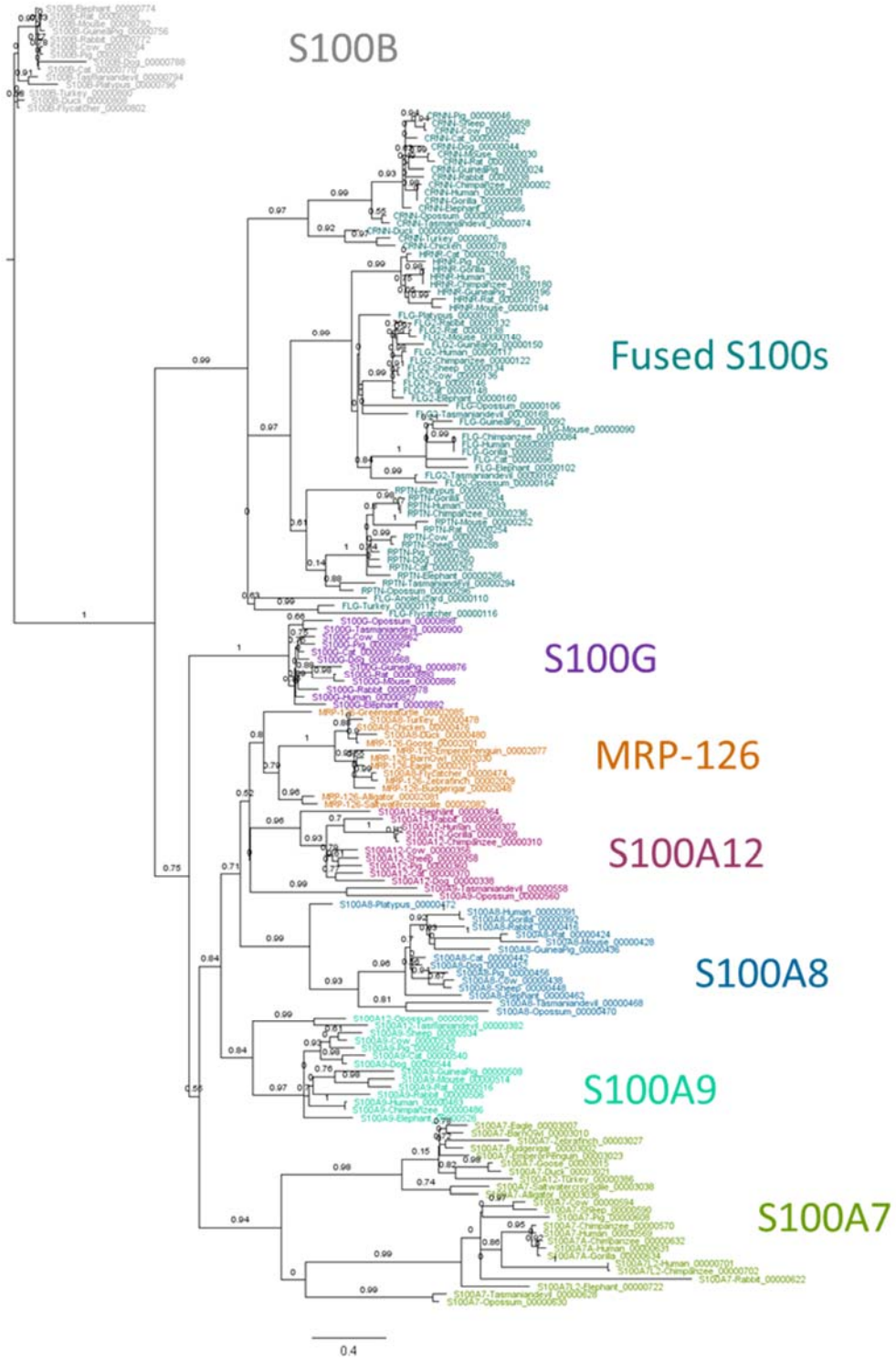

Figure AD1. Phylogeny for calgranulin tree with highest log-likelihood. Fifteen possible topologies for the calgranulin clade were used in reconstructions. This topology which had the highest log-likelihood was used for ancestral state reconstruction. Support values are SH-supports calculated using an approximate likelihood ratio test.

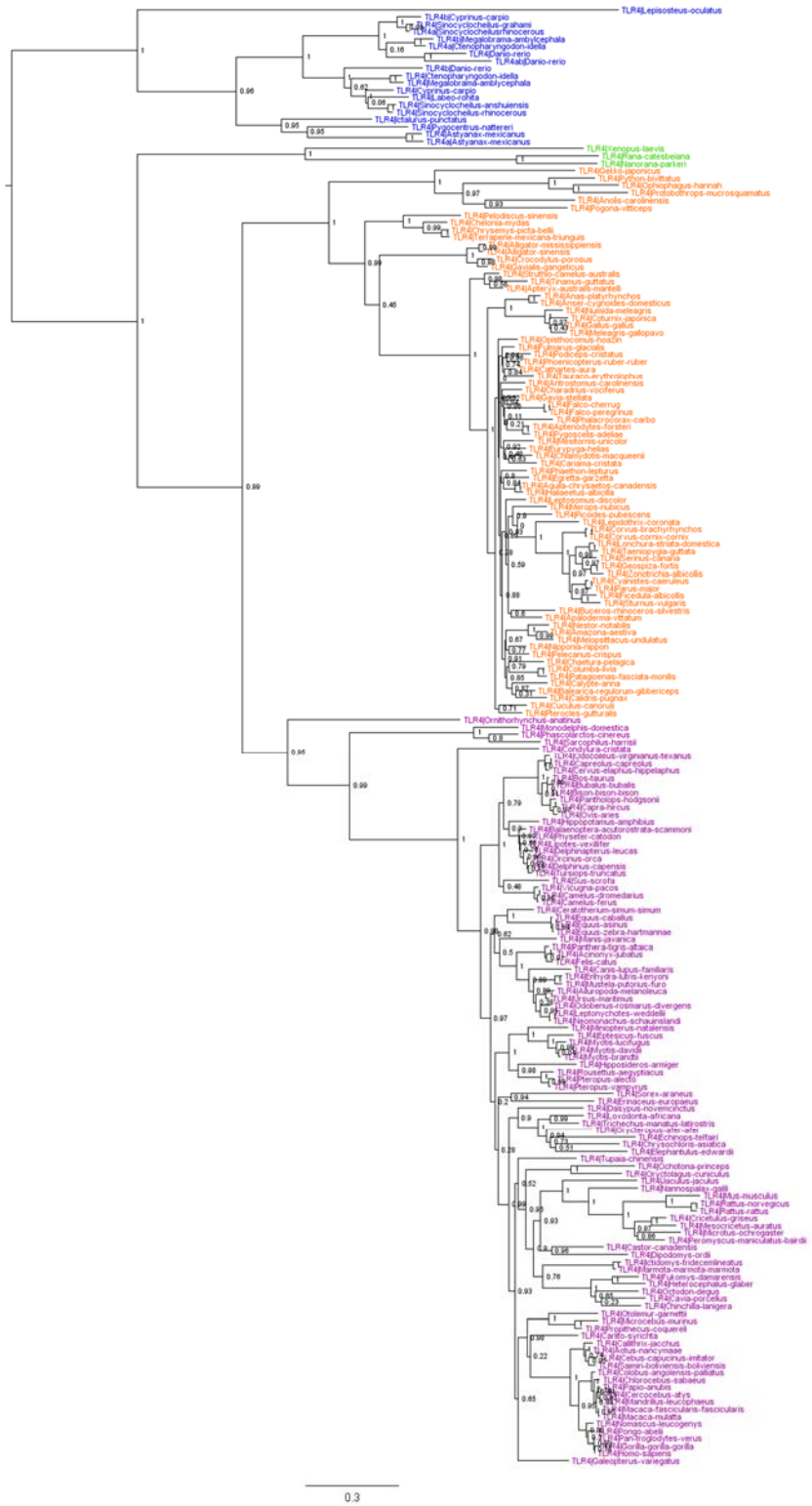

Figure AD2. Species corrected phylogeny of TLR4. A maximum likelihood phylogeny of TLR4 was solved using the JTT model and manually species-corrected. Genus and species are shown for tip labels. Colors are used to show clades of fish (blue), amphibians (green), birds and reptiles (orange), and mammals (purple). Support values are SH-supports calculated using an approximate likelihood ratio test.

Figure AD3. Species corrected phylogeny for MD-2. A maximum likelihood phylogeny of MD-2 was solved using the JTT model and manually species-corrected. Genus and species are shown for tip labels. Colors are used to show clades of fish (blue), amphibians (green), birds and reptiles (orange), and mammals (purple). Support values are SH-supports calculated using an approximate likelihood ratio test.

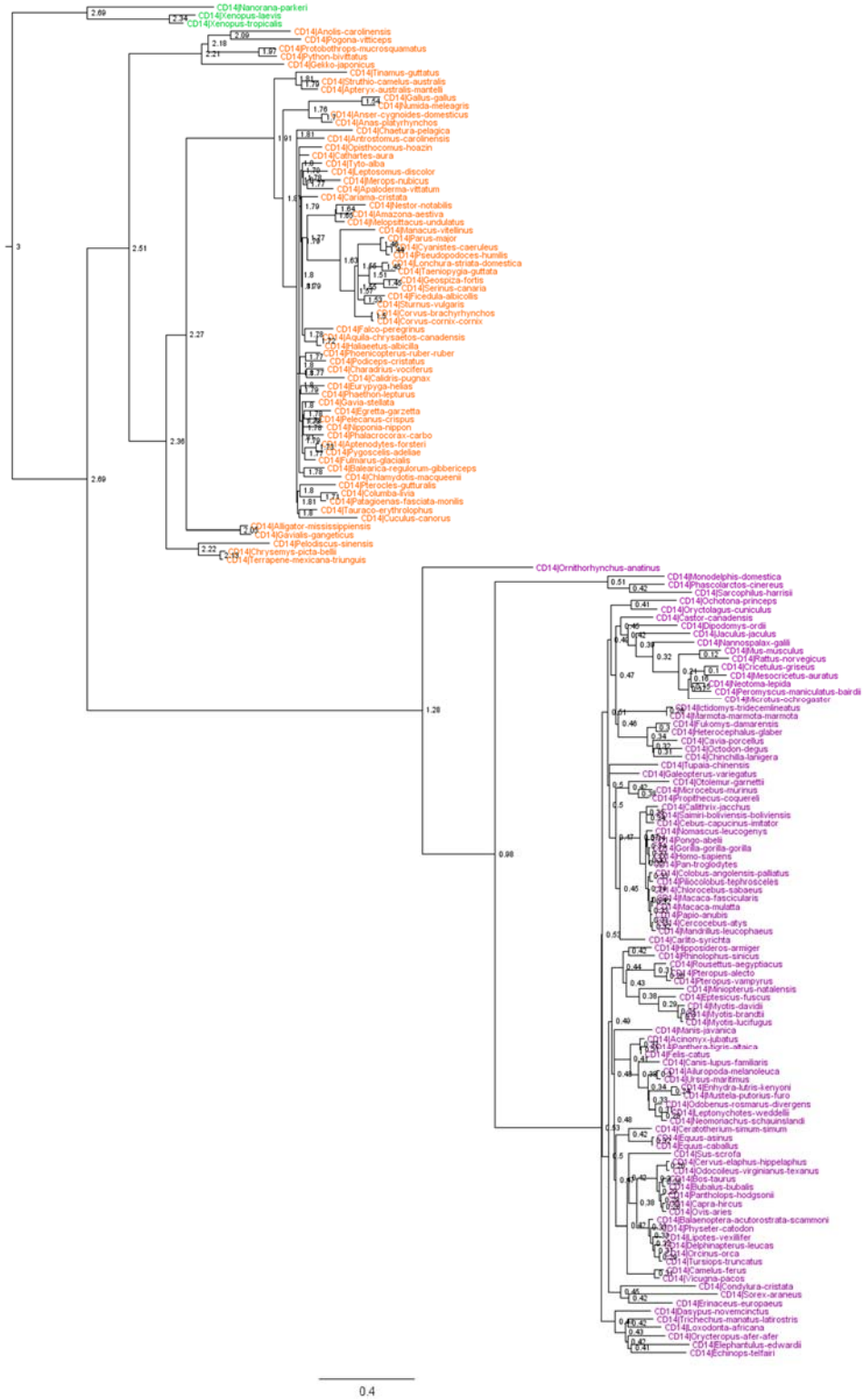

Figure AD4. Species-corrected phylogeny for CD14. A maximum likelihood phylogeny of TLR4 was solved using the JTT model and manually species-corrected. Genus and species are shown for tip labels. Colors are used to show clades of fish (blue), amphibians (green), birds and reptiles (orange), and mammals (purple). Support values are SH-supports calculated using an approximate likelihood ratio test.

REFERENCES CITED

(1) Hayashi, T.; Nakamura, T.; Takaoka, A. Pattern Recognition Receptors. *Nihon Rinsho Meneki. Gakkai Kaishi* **2011**, *34* (5), 329–345. https://doi.org/10.2177/JSCI.34.329.

(2) 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.

(3) Belvin, M. P.; Anderson, K. V. A Conserved Signalling Pathway: The Drosophila Toll-Dorsal Pathway. *Annu. Rev. Cell Dev. Biol.* **1996**, *12* (1), 393–416. https://doi.org/10.1146/annurev.cellbio.12.1.393.

(4) Hopkins, P. A.; Sriskandan, S. Mammalian Toll-like Receptors: To Immunity and Beyond. *Clin. Exp. Immunol.* **2005**, *140* (3), 395–407. https://doi.org/10.1111/j.1365-2249.2005.02801.x.

(5) Beutler, B.; Rehli, M. Evolution of the TIR, Tolls and TLRs: Functional Inferences from Computational Biology. In *Toll-Like Receptor Family Members and Their Ligands*; Springer Berlin Heidelberg, 2002; pp 1–21. https://doi.org/10.1007/978-3-642-59430-4_1.

(6) Roach, J. C.; Glusman, G.; Rowen, L.; Kaur, A.; Purcell, M. K.; Smith, K. D.; Hood, L. E.; Aderem, A. The Evolution of Vertebrate Toll-like Receptors. *Proc. Natl. Acad. Sci. U. S. A.* **2005**, *102* (27), 9577–9582. https://doi.org/10.1073/pnas.0502272102.

(7) Tsan, M.-F.; Gao, B. Endogenous Ligands of Toll-like Receptors. *J. Leukoc. Biol.* **2004**, *76* (3), 514–519. https://doi.org/10.1189/jlb.0304127.

(8) Leulier, F.; Lemaitre, B. Toll-like Receptors — Taking an Evolutionary Approach. *Nat. Rev. Genet.* **2008**, *9* (3), 165–178. https://doi.org/10.1038/nrg2303.

(9) Schaefer, L. Complexity of Danger: The Diverse Nature of Damage-Associated Molecular Patterns. *J. Biol. Chem.* **2014**, *289* (51), 35237–35245. https://doi.org/10.1074/jbc.R114.619304.

(10) Meijer, A. H.; Gabby Krens, S. .; Medina Rodriguez, I. A.; He, S.; Bitter, W.; Ewa Snaar-Jagalska, B.; Spaink, H. P. Expression Analysis of the Toll-like Receptor and TIR Domain Adaptor Families of Zebrafish. *Mol. Immunol.* **2004**, *40* (11), 773–783. https://doi.org/10.1016/J.MOLIMM.2003.10.003.

(11) Ghosh, S.; May, M. J.; Kopp, E. B. NF-κB AND REL PROTEINS: Evolutionarily Conserved Mediators of Immune Responses. *Annu. Rev. Immunol.* **1998**, *16* (1), 225–260. https://doi.org/10.1146/annurev.immunol.16.1.225.

(12) Gallay, P.; Heumann, D.; Le Roy, D.; Barras, C.; Glauser, M. P.; Mauri, D.; Burns, K.; Riederer, B. M.; Akira, S.; Calandra, T. Lipopolysaccharide-Binding Protein as a Major Plasma Protein Responsible for Endotoxemic Shock. *Proc. Natl. Acad. Sci. U. S. A.* **1993**, *90* (21), 9935–9938. https://doi.org/10.1073/pnas.90.21.9935.

(13) Billiar Gregory Gibson, T. R.; Sodhi, C.; Watkins, S.; Meihong Deng, D.; Scott, M. J.; Loughran, P. Specific Functions of TLR4 during Sepsis − Responses Are Regulated by Cell Type Clearance, and Systemic Inflammatory Lipopolysaccharide Clearance, Bacterial. **2018**. https://doi.org/10.4049/jimmunol.1300496.

(14) Poltorak, A. Defective LPS Signaling in C3H/HeJ and C57BL/10ScCr Mice: Mutations in Tlr4 Gene. *Science (80-.).* **1998**, *282* (5396), 2085–2088. https://doi.org/10.1126/science.282.5396.2085.

(15) Leon, C. G.; Tory, R.; Jia, J.; Sivak, O.; Wasan, K. M. Discovery and Development of Toll-like Receptor 4 (TLR4) Antagonists: A New Paradigm for Treating Sepsis and Other Diseases. *Pharm. Res.* **2008**, *25* (8), 1751–1761. https://doi.org/10.1007/s11095-008-9571-x.

(16) White, A. F. B.; Demchenko, A. V. Modulating LPS Signal Transduction at the LPS Receptor Complex with Synthetic Lipid A Analogues; 2014; pp 339–389. https://doi.org/10.1016/B978-0-12-800128-8.00005-4.

(17) Kim, H. M.; Park, B. S.; Kim, J.-I.; Kim, S. E.; Lee, J.; Oh, S. C.; Enkhbayar, P.; Matsushima, N.; Lee, H.; Yoo, O. J.; et al. Crystal Structure of the TLR4-MD-2 Complex with Bound Endotoxin Antagonist Eritoran. *Cell* **2007**, *130* (5), 906–917. https://doi.org/10.1016/j.cell.2007.08.002.

(18) Kuzmich, N. N.; Sivak, K. V; Chubarev, V. N.; Porozov, Y. B.; Savateeva-Lyubimova, T. N.; Peri, F. TLR4 Signaling Pathway Modulators as Potential Therapeutics in Inflammation and Sepsis. *Vaccines* **2017**, *5* (4). https://doi.org/10.3390/vaccines5040034.

(19) Saitoh, S.; Akashi, S.; Yamada, T.; Tanimura, N.; Matsumoto, F.; Fukase, K.; Kusumoto, S.; Kosugi, A.; Miyake, K. Ligand-Dependent Toll-like Receptor 4 (TLR4)-Oligomerization Is Directly Linked with TLR4-Signaling. *J. Endotoxin Res.* **2004**, *10* (4), 257–260. https://doi.org/10.1177/09680519040100041001.

(20) Kobayashi, M.; Saitoh, S.; Tanimura, N.; Takahashi, K.; Kawasaki, K.; Nishijima, M.; Fujimoto, Y.; Fukase, K.; Akashi-Takamura, S.; Miyake, K. Regulatory Roles for MD-2 and TLR4 in Ligand-Induced Receptor Clustering. *J. Immunol.* **2006**, *176* (10), 6211–6218. https://doi.org/10.4049/jimmunol.176.10.6211.

(21) 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. https://doi.org/10.1038/ni809.

(22) Fujimoto, T.; Yamazaki, S.; Eto-Kimura, A.; Takeshige, K.; Muta, T. The Amino-Terminal Region of Toll-like Receptor 4 Is Essential for Binding to MD-2 and Receptor Translocation to the Cell Surface. *J. Biol. Chem.* **2004**, *279* (46), 47431–47437. https://doi.org/10.1074/jbc.M408724200.

(23) Walsh, C.; Gangloff, M.; Monie, T.; Smyth, T.; Wei, B.; McKinley, T. J.; Maskell, D.; Gay, N.; Bryant, C. Elucidation of the MD-2/TLR4 Interface Required for Signaling by Lipid IVa. *J. Immunol.* **2008**, *181* (2), 1245–1254. https://doi.org/10.4049/jimmunol.181.2.1245.

(24) Silva, A.; Wagner, B.; McKenzie, H. C.; Desrochers, A. M.; Furr, M. O. An Investigation of the Role of Soluble CD14 in Hospitalized, Sick Horses. *Vet. Immunol. Immunopathol.* **2013**, *155* (4), 264–269. https://doi.org/10.1016/j.vetimm.2013.08.007.

(25) da Silva Correia, J.; Soldau, K.; Christen, U.; Tobias, P. S.; Ulevitch, R. J. Lipopolysaccharide Is in Close Proximity to Each of the Proteins in Its Membrane Receptor Complex. Transfer from CD14 to TLR4 and MD-2. *J. Biol. Chem.* **2001**, *276* (24), 21129–21135. https://doi.org/10.1074/jbc.M009164200.

(26) Ohnishi, T.; Muroi, M.; Tanamoto, K.-I. The Lipopolysaccharide-Recognition Mechanism in Cells Expressing TLR4 and CD14 but Lacking MD-2. **2007**. https://doi.org/10.1111/j.1574-695X.2007.00281.x.

(27) Muroi, M.; Ohnishi, T.; Tanamoto, K. Regions of the Mouse CD14 Molecule Required for Toll-like Receptor 2- and 4-Mediated Activation of NF-κB. *J. Biol. Chem.* **2002**, *277* (44), 42372–42379. https://doi.org/10.1074/jbc.M205966200.

(28) Kim, J.-I.; Lee, C. J.; Jin, M. S.; Lee, C.-H.; Paik, S.-G.; Lee, H.; Lee, J.-O. Crystal Structure of CD14 and Its Implications for Lipopolysaccharide Signaling. *J. Biol. Chem.* **2005**, *280* (12), 11347–11351. https://doi.org/10.1074/jbc.M414607200.

(29) 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.

(30) Vasl, J.; Prohinar, P.; Gioannini, T. L.; Weiss, J. P.; Jerala, R. Functional Activity of MD-2 Polymorphic Variant Is Significantly Different in Soluble and TLR4-Bound Forms: Decreased Endotoxin Binding by G56R MD-2 and Its Rescue by TLR4 Ectodomain. *J. Immunol.* **2008**, *180* (9), 6107–6115. https://doi.org/10.4049/jimmunol.180.9.6107.

(31) Jerala, R. Structural Biology of the LPS Recognition. *Int. J. Med. Microbiol.* **2007**, *297* (5), 353–363. https://doi.org/10.1016/j.ijmm.2007.04.001.

(32) Park, B. S.; Lee, J.-O. Recognition of Lipopolysaccharide Pattern by TLR4 Complexes. *Exp. Mol. Med.* **2013**, *45* (12), e66. https://doi.org/10.1038/emm.2013.97.

(33) Mueller, M.; Lindner, B.; Dedrick, R.; Schromm, A. B.; Seydel, U. Endotoxin: Physical Requirements for Cell Activation. *J. Endotoxin Res.* **2005**, *11* (5), 299–303. https://doi.org/10.1177/09680519050110050701.

(34) Mueller, M.; Lindner, B.; Kusumoto, S.; Fukase, K.; Schromm, A. B.; Seydel, U. Aggregates Are the Biologically Active Units of Endotoxin. *J. Biol. Chem.* **2004**, *279* (25), 26307–26313. https://doi.org/10.1074/jbc.M401231200.

(35) Yu, B.; Wright, S. D. Catalytic Properties of Lipopolysaccharide (LPS) Binding Protein. Transfer of LPS to Soluble CD14. *J. Biol. Chem.* **1996**, *271* (8), 4100–4105. https://doi.org/10.1074/JBC.271.8.4100.

(36) Tobias, P. S.; Soldau, K.; Iovine, N. M.; Elsbach, P.; Weiss, J. Lipopolysaccharide (LPS)-Binding Proteins BPI and LBP Form Different Types of Complexes with LPS. *J. Biol. Chem.* **1997**, *272* (30), 18682–18685. https://doi.org/10.1074/JBC.272.30.18682.

(37) Kim, S. J.; Kim, H. M. Dynamic Lipopolysaccharide Transfer Cascade to TLR4/MD2 Complex via LBP and CD14. *BMB Rep.* **2017**, *50* (2), 55–57. https://doi.org/10.5483/BMBREP.2017.50.2.011.

(38) Schütt, C.; Bernheiden, M.; Grunwald, U.; Stelter, F.; Menzel, R.; Müller, H. P.; Fan, X.; Jack, R. S. Implications for a General Role of LPS-Binding Proteins (CD14, LBP) in Combating Bacterial Infections. *J. Endotoxin Res.* **1999**, *5* (1–2), 75–80. https://doi.org/10.1177/09680519990050010601.

(39) Shapiro, R. A.; Cunningham, M. D.; Ratcliffe, K.; Seachord, C.; Blake, J.; Bajorath, J.; Aruffo, A.; Darveau, R. P. Identification of CD14 Residues Involved in Specific Lipopolysaccharide Recognition. *Infect. Immun.* **1997**, *65* (1), 293–297.

(40) Zettel, K.; Korff, S.; Zamora, R.; Morelli, A. E.; Darwiche, S.; Loughran, P. A.; Elson, G.; Shang, L.; Salgado-Pires, S.; Scott, M. J.; et al. Toll-Like Receptor 4 on Both Myeloid Cells and Dendritic Cells Is Required for Systemic Inflammation and Organ Damage after Hemorrhagic Shock with Tissue Trauma in Mice. *Front. Immunol.* **2017**, *8*, 1672. https://doi.org/10.3389/fimmu.2017.01672.

(41) 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.

(42) Gordon, S. Pattern Recognition Receptors: Doubling Up for the Innate Immune Response. *Cell* **2002**, *111* (7), 927–930. https://doi.org/10.1016/S0092-8674(02)01201-1.

(43) Taylor, K. R.; Trowbridge, J. M.; Rudisill, J. A.; Termeer, C. C.; Simon, J. C.; Gallo, R. L. Hyaluronan Fragments Stimulate Endothelial Recognition of Injury through TLR4. *J. Biol. Chem.* **2004**, *279* (17), 17079–17084. https://doi.org/10.1074/jbc.M310859200.

(44) Termeer, C.; Benedix, F.; Sleeman, J.; Fieber, C.; Voith, U.; Ahrens, T.; Miyake, K.; Freudenberg, M.; Galanos, C.; Simon, J. C. Oligosaccharides of Hyaluronan Activate Dendritic Cells via Toll-like Receptor 4. *J. Exp. Med.* **2002**, *195* (1), 99–111. https://doi.org/10.1084/JEM.20001858.

(45) Johnson, G. B.; Brunn, G. J.; Kodaira, Y.; Platt, J. L. Receptor-Mediated Monitoring of Tissue Well-Being via Detection of Soluble Heparan Sulfate by Toll-like Receptor 4. *J. Immunol.* **2002**, *168* (10), 5233–5239. https://doi.org/10.4049/JIMMUNOL.168.10.5233.

(46) Ohashi, K.; Burkart, V.; Flohé, S.; Kolb, H. Receptor-4 Complex Putative Endogenous Ligand of the Toll-Like Cutting Edge: Heat Shock Protein 60 Is a. *J Immunol Ref.* **2018**, *164*, 558–561. https://doi.org/10.4049/jimmunol.164.2.558.

(47) Ehrchen, J. M.; Sunderkotter, 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.

(48) Roshan, M. H. K.; Tambo, A.; Pace, N. P. The Role of TLR2, TLR4, and TLR9 in the Pathogenesis of Atherosclerosis. *Int. J. Inflam.* **2016**, *2016*, 1–11. https://doi.org/10.1155/2016/1532832.

(49) Curtiss, L. K.; Tobias, P. S. Emerging Role of Toll-like Receptors in Atherosclerosis. *J. Lipid Res.* **2009**, *50 Suppl* (Suppl), S340-5. https://doi.org/10.1194/jlr.R800056-JLR200.

(50) Shimamoto, A.; Chong, A. J.; Yada, M.; Shomura, S.; Takayama, H.; Fleisig, A. J.; Agnew, M. L.; Hampton, C. R.; Rothnie, C. L.; Spring, D. J.; et al. Inhibition of Toll-like Receptor 4 With Eritoran Attenuates Myocardial Ischemia-Reperfusion Injury. *Circulation* **2006**, *114* (1_suppl), I-270-I-274. https://doi.org/10.1161/CIRCULATIONAHA.105.000901.

(51) Tsung, A.; Hoffman, R. A.; Izuishi, K.; Critchlow, N. D.; Nakao, A.; Chan, M. H.; Lotze, M. T.; Geller, D. A.; Billiar, T. R. Hepatic Ischemia/reperfusion Injury Involves Functional TLR4 Signaling in Nonparenchymal Cells. *J. Immunol.* **2005**, *175* (11), 7661–7668.

(52) Cunningham, P. N.; Wang, Y.; Guo, R.; He, G.; Quigg, R. J. Role of Toll-like Receptor 4 in Endotoxin-Induced Acute Renal Failure. *J. Immunol.* **2004**, *172* (4), 2629–2635.

(53) Simpson, J. L.; Grissell, T. V; Douwes, J.; Scott, R. J.; Boyle, M. J.; Gibson, P. G. Innate Immune Activation in Neutrophilic Asthma and Bronchiectasis. *Thorax* **2007**, *62* (3), 211–218. https://doi.org/10.1136/thx.2006.061358.

(54) Baumgarten, G.; Knuefermann, P.; Wrigge, H.; Putensen, C.; Stapel, H.; Fink, K.; Meyer, R.; Hoeft, A.; Grohé, C. Role of Toll-like Receptor 4 for the Pathogenesis of Acute Lung Injury in Gram-Negative Sepsis. *Eur. J. Anaesthesiol.* **2006**, *23* (12), 1041–1048. https://doi.org/10.1017/S0265021506001098.

(55) Fort, M. M.; Mozaffarian, A.; Stöver, A. G.; Correia, J. da S.; Johnson, D. A.; Crane, R. T.; Ulevitch, R. J.; Persing, D. H.; Bielefeldt-Ohmann, H.; Probst, P.; et al. A Synthetic TLR4 Antagonist Has Anti-Inflammatory Effects in Two Murine Models of Inflammatory Bowel Disease. *J. Immunol.* **2005**, *174* (10), 6416–6423.

(56) Oostenbrug, L. E.; Drenth, J. P. H.; de Jong, D. J.; Nolte, I. M.; Oosterom, E.; van Dullemen, H. M.; van der Linde, K.; te Meerman, G. J.; van der Steege, G.; Kleibeuker, J. H.; et al. Association between Toll-like Receptor 4 and Inflammatory Bowel Disease. *Inflamm. Bowel Dis.* **2005**, *11* (6), 567–575.

(57) De Jager, P. L.; Franchimont, D.; Waliszewska, A.; Bitton, A.; Cohen, A.; Langelier, D.; Belaiche, J.; Vermeire, S.; Farwell, L.; Goris, A.; et al. The Role of the Toll Receptor Pathway in Susceptibility to Inflammatory Bowel Diseases. *Genes Immun.* **2007**, *8* (5), 387–397. https://doi.org/10.1038/sj.gene.6364398.

(58) Dheer, R.; Santaolalla, R.; Davies, J. M.; Lang, J. K.; Phillips, M. C.; Pastorini, C.; Vazquez-Pertejo, M. T.; Abreu, M. T. Intestinal Epithelial Toll-Like Receptor 4 Signaling Affects Epithelial Function and Colonic Microbiota and Promotes a Risk for Transmissible Colitis. *Infect. Immun.* **2016**, *84* (3), 798–810. https://doi.org/10.1128/IAI.01374-15.

(59) Cario, E. Toll-like Receptors in Inflammatory Bowel Diseases: A Decade Later. *Inflamm. Bowel Dis.* **2010**, *16* (9), 1583–1597. https://doi.org/10.1002/ibd.21282.

(60) Lu, Y.; Li, X.; Liu, S.; Zhang, Y.; Zhang, D. Toll-like Receptors and Inflammatory Bowel Disease. *Front. Immunol.* **2018**, *9*, 72. https://doi.org/10.3389/fimmu.2018.00072.

(61) Ryckman, C.; Vandal, K.; Rouleau, P.; Talbot, M.; Tessier, P. A. Proinflammatory Activities of S100: Proteins S100A8, S100A9, and S100A8/A9 Induce Neutrophil Chemotaxis and Adhesion. *J. Immunol.* **2003**, *170* (6), 3233–3242. https://doi.org/10.4049/jimmunol.170.6.3233.

(62) 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.

(63) Chen, B.; Miller, A. L.; Rebelatto, M.; Brewah, Y.; Rowe, D. C.; Clarke, L.; Czapiga, M.; Rosenthal, K.; Imamichi, T.; Chen, Y.; et al. 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.

(64) 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.

(65) Kehl-Fie, T. E.; Chitayat, S.; Hood, M. I.; Damo, S.; Restrepo, N.; Garcia, C.; Munro, K. A.; Chazin, W. J.; Skaar, E. P. *Nutrient Metal Sequestration by Calprotectin Inhibits Bacterial Superoxide Defense, Enhancing Neutrophil Killing of Staphylococcus Aureus*; 2011; Vol. 10. https://doi.org/10.1016/j.chom.2011.07.004.

(66) Leach, S. T.; Yang, Z.; Messina, I.; Song, C.; Geczy, C. L.; Cunningham, A. M.; Day, A. S. Serum and Mucosal S100 Proteins, Calprotectin (S100A8/S100A9) and S100A12, Are Elevated at Diagnosis in Children with Inflammatory Bowel Disease. *Scand. J. Gastroenterol.* **2007**, *42* (11), 1321–1331. https://doi.org/10.1080/00365520701416709.

(67) Dai, J.; Kumbhare, A.; Youssef, D.; McCall, C. E.; Gazzar, M. El. Intracellular S100A9 Promotes Myeloid-Derived Suppressor Cells during Late Sepsis. *Front. Immunol.* **2017**, *8*, 1565. https://doi.org/10.3389/fimmu.2017.01565.

(68) 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.

(69) Kim, J.-I.; Lee, C. J.; Jin, M. S.; Lee, C.-H.; Paik, S.-G.; Lee, H.; Lee, J.-O. Crystal Structure of CD14 and Its Implications for Lipopolysaccharide Signaling. *J. Biol. Chem.* **2005**, *280* (12), 11347–11351. https://doi.org/10.1074/jbc.M414607200.

(70) 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.

(71) 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.

(72) Weiss, G. A.; Watanabe, C. K.; Zhong, A.; Goddard, A.; Sidhu, S. S. *Rapid Mapping of Protein Functional Epitopes by Combinatorial Alanine Scanning*; Sunesis Pharmaceuticals, Inc, 2000.

(73) Morrison, K. L.; Weiss, G. A. Combinatorial Alanine-Scanning. *Curr. Opin. Chem. Biol.* **2001**, *5* (3), 302–307. https://doi.org/10.1016/S1367-5931(00)00206-4.

(74) Carlson, G. M.; Fenton, A. W. What Mutagenesis Can and Cannot Reveal About Allostery. *Biophys. J.* **2016**, *110* (9), 1912–1923. https://doi.org/10.1016/J.BPJ.2016.03.021.

(75) Pauling, L.; Zuckerkandl, E. Chemical Paleogenetics. *Acta Chem. Scand.* **1963**, *17*, S9–S16.

(76) Kuang, D.; Yao, Y.; Maclean, D.; Wang, M.; Hampson, D. R.; Chang, B. S. W. Ancestral Reconstruction of the Ligand-Binding Pocket of Family C G Protein-Coupled Receptors. *Proc. Natl. Acad. Sci. U. S. A.* **2006**, *103* (38), 14050–14055. https://doi.org/10.1073/pnas.0604717103.

(77) 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.

(78) Thomson, J. M.; Gaucher, E. A.; Burgan, M. F.; De Kee, D. W.; Li, T.; Aris, J. P.; Benner, S. A. Resurrecting Ancestral Alcohol Dehydrogenases from Yeast. *Nat. Genet.* **2005**, *37* (6), 630–635. https://doi.org/10.1038/ng1553.

(79) Yokoyama, S.; Yang, H.; Starmer, W. T. Molecular Basis of Spectral Tuning in the Red- and Green-Sensitive (M/LWS) Pigments in Vertebrates. *Genetics* **2008**, *179* (4), 2037–2043. https://doi.org/10.1534/genetics.108.090449.

(80) Field, S. F.; Matz, M. V. Retracing Evolution of Red Fluorescence in GFP-Like Proteins from Faviina Corals. *Mol. Biol. Evol.* **2010**, *27* (2), 225–233. https://doi.org/10.1093/molbev/msp230.

(81) 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.

(82) Ernst, P. B.; Carvunis, A.-R. Of Mice, Men and Immunity: A Case for Evolutionary Systems Biology. *Nat. Immunol.* **2018**, *19* (5), 421–425. https://doi.org/10.1038/s41590-018-0084-4.

(83) Baker, C. R.; Booth, L. N.; Sorrells, T. R.; Johnson, A. D. Protein Modularity, Cooperative Binding, and Hybrid Regulatory States Underlie Transcriptional Network Diversification. *Cell* **2012**, *151* (1), 80–95. https://doi.org/10.1016/J.CELL.2012.08.018.

(84) Miller, S. I.; Ernst, R. K.; Bader, M. W. LPS, TLR4 and Infectious Disease Diversity. *Nat. Rev. Microbiol.* **2005**, *3* (1), 36–46. https://doi.org/10.1038/nrmicro1068.

(85) Akashi, S.; Nagai, Y.; Ogata, H.; Oikawa, M.; Fukase, K.; Kusumoto, S.; Kawasaki, K.; Nishijima, M.; Hayashi, S.; Kimoto, M.; et al. Human MD-2 Confers on Mouse Toll-like Receptor 4 Species-Specific Lipopolysaccharide Recognition. *Int. Immunol.* **2001**, *13* (12), 1595–1599. https://doi.org/10.1093/intimm/13.12.1595.

(86) Anwar, M. A.; Panneerselvam, S.; Shah, M.; Choi, S. Insights into the Species-Specific TLR4 Signaling Mechanism in Response to Rhodobacter Sphaeroides Lipid A Detection. *Sci. Rep.* **2015**, *5* (1), 7657. https://doi.org/10.1038/srep07657.

(87) Muroi, M.; Ohnishi, T.; Tanamoto, K.-I. MD-2, a Novel Accessory Molecule, Is Involved in Species-Specific Actions of Salmonella Lipid A. *Infect. Immun.* **2002**, *70* (7), 3546–3550. https://doi.org/10.1128/IAI.70.7.3546-3550.2002.

(88) Zimmer, S. M.; Liu, J.; Clayton, J. L.; Stephens, D. S.; Snyder, J. P. Paclitaxel Binding to Human and Murine MD-2. *J. Biol. Chem.* **2008**, *283* (41), 27916–27926. https://doi.org/10.1074/jbc.M802826200.

(89) Kawasaki, K.; Akashi, S.; Shimazu, R.; Yoshida, T.; Miyake, K.; Nishijima, M. Mouse Toll-like Receptor 4/MD-2 Complex Mediates Lipopolysaccharide-Mimetic Signal Transduction by Taxol. *J. Biol. Chem.* **2000**, *275* (4), 2251–2254. https://doi.org/10.1074/jbc.275.4.2251.

(90) 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*, 304. https://doi.org/10.3389/fimmu.2018.00304.

(91) Zhou, Y.; Liang, Q.; Li, W.; Gu, Y.; Liao, X.; Fang, W.; Li, X. Characterization and Functional Analysis of Toll-like Receptor 4 in Chinese Soft-Shelled Turtle Pelodiscus Sinensis. *Dev. Comp. Immunol.* **2016**, *63*, 128–135. https://doi.org/10.1016/j.dci.2016.05.023.

(92) Keestra, A. M.; van Putten, J. P. M. Unique Properties of the Chicken TLR4/MD-2 Complex: Selective Lipopolysaccharide Activation of the MyD88-Dependent Pathway. *J. Immunol.* **2008**, *181* (6), 4354–4362. https://doi.org/10.4049/jimmunol.181.6.4354.

(93) Sepulcre, M. P.; Alcaraz-Pérez, F.; López-Muñoz, A.; Roca, F. J.; Meseguer, J.; Cayuela, M. L.; Mulero, V. Evolution of Lipopolysaccharide (LPS) Recognition and Signaling: Fish TLR4 Does Not Recognize LPS and Negatively Regulates NF-kappaB Activation. *J. Immunol.* **2009**, *182* (4), 1836–1845. https://doi.org/10.4049/jimmunol.0801755.

(94) Sullivan, C.; Charette, J.; Catchen, J.; Lage, C. R.; Giasson, G.; Postlethwait, J. H.; Millard, P. J.; Kim, C. H. The Gene History of Zebrafish tlr4a and tlr4b Is Predictive of Their Divergent Functions. *J. Immunol.* **2009**, *183* (9), 5896–5908. https://doi.org/10.4049/jimmunol.0803285.

(95) Yoder, J. A.; Nielsen, M. E.; Amemiya, C. T.; Litman, G. W. Zebrafish as an Immunological Model System. *Microbes Infect.* **2002**, *4* (14), 1469–1478. https://doi.org/10.1016/S1286-4579(02)00029-1.

(96) Renshaw, S. A.; Trede, N. S. A Model 450 Million Years in the Making: Zebrafish and Vertebrate Immunity. *Dis. Model. Mech.* **2012**, *5* (1), 38–47. https://doi.org/10.1242/dmm.007138.

(97) Hall, C.; Flores, M. V.; Chien, A.; Davidson, A.; Crosier, K.; Crosier, P. Transgenic Zebrafish Reporter Lines Reveal Conserved Toll-like Receptor Signaling Potential in Embryonic Myeloid Leukocytes and Adult Immune Cell Lineages. *J. Leukoc. Biol.* **2009**, *85* (5), 751–765. https://doi.org/10.1189/jlb.0708405.

(98) Fischbach, W.; Becker, W.; Smith, C.; Becker, K.; Friedrich, A. W.; Rueffer, A.; Dobos, G. J.; Roth, J.; Foell, D. Clinical Relevance of Activity Parameters in Crohn's Disease Estimated by the Faecal Excretion of 111In-Labeled Granulocytes. *Digestion* **1991**, *50* (3–4), 149–152. https://doi.org/10.1136/GUT.27.1.92.

(99) Stenvik, J.; Solstad, T.; Strand, C.; Leiros, I.; Jørgensen T, T. Ø. Cloning and Analyses of a BPI/LBP cDNA of the Atlantic Cod (Gadus Morhua L.). *Dev. Comp. Immunol.* **2004**, *28* (4), 307–323.

(100) 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.

(101) Zhang, H.; Tay, P. N.; Cao, W.; Li, W.; Lu, J. Integrin-Nucleated Toll-like Receptor (TLR) Dimerization Reveals Subcellular Targeting of TLRs and Distinct Mechanisms of TLR4 Activation and Signaling. *FEBS Lett.* **2002**, *532* (1–2), 171–176.

(102) Bates, J. M.; Akerlund, J.; Mittge, E.; Guillemin, K. Intestinal Alkaline Phosphatase Detoxifies Lipopolysaccharide and Prevents Inflammation in Zebrafish in Response to the Gut Microbiota. *Cell Host Microbe* **2007**, *2* (6), 371–382. https://doi.org/10.1016/J.CHOM.2007.10.010.

(103) van der Sar, A. M.; Stockhammer, O. W.; van der Laan, C.; Spaink, H. P.; Bitter, W.; Meijer, A. H. MyD88 Innate Immune Function in a Zebrafish Embryo Infection Model. *Infect. Immun.* **2006**, *74* (4), 2436–2441. https://doi.org/10.1128/IAI.74.4.2436-2441.2006.

(104) Forn-Cuní, G.; Varela, M.; Pereiro, P.; Novoa, B.; Figueras, A. Conserved Gene Regulation during Acute Inflammation between Zebrafish and Mammals. *Sci. Rep.* **2017**, *7*, 41905. https://doi.org/10.1038/srep41905.

(105) Sepulcre, M. P.; López-Castejón, G.; Meseguer, J.; Mulero, V. The Activation of Gilthead Seabream Professional Phagocytes by Different PAMPs Underlines the Behavioural Diversity of the Main Innate Immune Cells of Bony Fish. *Mol. Immunol.* **2007**, *44* (8), 2009–2016. https://doi.org/10.1016/J.MOLIMM.2006.09.022.

(106) Qiao, L.; Yang, W.; Fu, J.; Song, Z. Transcriptome Profile of the Green Odorous Frog (*Odorrana Margaretae*). *PLoS One* **2013**, *8* (9), e75211. https://doi.org/10.1371/journal.pone.0075211.

(107) Nourisson, C.; Carneiro, M.; Vallinoto, M.; Sequeira, F. De Novo Transcriptome Assembly and Polymorphism Detection in Ecologically Important Widely Distributed Neotropical Toads from the *Rhinella Marina* Species Complex (Anura: Bufonidade. *Genomic Resour. Notes Accept. 1 August 2014-30 Sept. 2014* **2014**. https://doi.org/10.5061/dryad.3jm3n.

(108) Genomic Resources Development Consortium; Arthofer, W.; Banbury, B. L.; Carneiro, M.; Cicconardi, F.; Duda, T. F.; Harris, R. B.; Kang, D. S.; Leaché, A. D.; Nolte, V.; et al. Genomic Resources Notes Accepted 1 August 2014-30 September 2014. *Mol. Ecol. Resour.* **2015**, *15* (1), 228–229. https://doi.org/10.1111/1755-0998.12340.

(109) Matsunami, M.; Kitano, J.; Kishida, O.; Michimae, H.; Miura, T.; Nishimura, K. Transcriptome Analysis of Predator- and Prey-Induced Phenotypic Plasticity in the Hokkaido Salamander (*Hynobius Retardatus*). *Mol. Ecol.* **2015**, *24* (12), 3064–3076. https://doi.org/10.1111/mec.13228.

(110) Iliev, D. B.; Roach, J. C.; Mackenzie, S.; Planas, J. V; Goetz, F. W. Endotoxin Recognition: In Fish or Not in Fish? *FEBS Lett.* **2005**, *579* (29), 6519–6528. https://doi.org/10.1016/j.febslet.2005.10.061.

(111) Oblak, A.; Jerala, R. The Molecular Mechanism of Species-Specific Recognition of Lipopolysaccharides by the MD-2/TLR4 Receptor Complex. *Mol. Immunol.* **2015**, *63* (2), 134–142. https://doi.org/10.1016/j.molimm.2014.06.034.

(112) Walsh, C.; Gangloff, M.; Monie, T.; Smyth, T.; Wei, B.; McKinley, T. J.; Maskell, D.; Gay, N.; Bryant, C. Elucidation of the MD-2/TLR4 Interface Required for Signaling by Lipid IVa. *J. Immunol.* **2008**, *181* (2).

(113) Wiese, A.; Brandenburg, K.; Lindner, B.; Schromm, A. B.; Carroll, S. F.; Rietschel, E. T.; Seydel, U. Mechanisms of Action of the Bactericidal/Permeability-Increasing Protein BPI on Endotoxin and Phospholipid Monolayers and Aggregates †. *Biochemistry* **1997**, *36* (33), 10301–10310. https://doi.org/10.1021/bi970176m.

(114) Dentener, M. A.; Von Asmuth, E. J.; Francot, G. J.; Marra, M. N.; Buurman, W. A. Antagonistic Effects of Lipopolysaccharide Binding Protein and Bactericidal/permeability-Increasing Protein on Lipopolysaccharide-Induced Cytokine Release by Mononuclear Phagocytes. Competition for Binding to Lipopolysaccharide. *J. Immunol.* **1993**, *151* (8), 4258–4265.

(115) Berczi, I.; Bertók, L.; Bereznai, T. Comparative Studies on the Toxicity of Escherichia Coli Lipopolysaccaride Endotoxin in Various Animal Species. *Can. J. Microbiol.* **1966**, *12* (5), 1070–1071. https://doi.org/10.1139/m66-143.

(116) Stafford, J. L.; Ellestad, K. K.; Magor, K. E.; Belosevic, M.; Magor, B. G. A Toll-like Receptor (TLR) Gene That is Up-Regulated in Activated Goldfish Macrophages. *Dev. Comp. Immunol.* **2003**, *27* (8), 685–698. https://doi.org/10.1016/S0145-305X(03)00041-7.

(117) MacKenzie, S.; Planas, J. V.; Goetz, F. W. LPS-Stimulated Expression of a Tumor Necrosis Factor-Alpha mRNA in Primary Trout Monocytes and in Vitro Differentiated Macrophages. *Dev. Comp. Immunol.* **2003**, *27* (5), 393–400. https://doi.org/10.1016/S0145-305X(02)00135-0.

(118) Pelegrín, P.; García-Castillo, J.; Mulero, V.; Meseguer, J. Interleukin-1β Isolated from a Marine Fish up-Regulated Expression in Macrophages Following Activation with Lipopolysaccharide and Lymphokines. *Cytokine* **2001**, *16* (2), 67–72. https://doi.org/10.1006/CYTO.2001.0949.

(119) Liu, Y.; Schmidt, B.; Maskell, D. L. MSAProbs: Multiple Sequence Alignment Based on Pair Hidden Markov Models and Partition Function Posterior Probabilities. *Bioinformatics* **2010**, *26* (16), 1958–1964. https://doi.org/10.1093/bioinformatics/btq338.

(120) Tamura, K.; Stecher, G.; Peterson, D.; Filipski, A.; Kumar, S. MEGA6: Molecular Evolutionary Genetics Analysis Version 6.0. *Mol. Biol. Evol.* **2013**, *30* (12), 2725–2729. https://doi.org/10.1093/molbev/mst197.

(121) 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.

(122) Le, S. Q.; Gascuel, O. Accounting for Solvent Accessibility and Secondary Structure in Protein Phylogenetics Is Clearly Beneficial. *Syst. Biol.* **2010**, *59* (3), 277–287. https://doi.org/10.1093/sysbio/syq002.

(123) Le, S. Q.; Gascuel, O. An Improved General Amino Acid Replacement Matrix. *Mol. Biol. Evol.* **2008**, *25* (7), 1307–1320. https://doi.org/10.1093/molbev/msn067.

(124) Hordijk, W.; Gascuel, O. Improving the Efficiency of SPR Moves in Phylogenetic Tree Search Methods Based on Maximum Likelihood. *Bioinformatics* **2005**, *21* (24), 4338–4347. https://doi.org/10.1093/bioinformatics/bti713.

(125) Anisimova, M.; Gascuel, O. Approximate Likelihood-Ratio Test for Branches: A Fast, Accurate, and Powerful Alternative. *Syst. Biol.* **2006**, *55* (4), 539–552. https://doi.org/10.1080/10635150600755453.

(126) Akaike, H. Information Theory and an Extension of the Maximum Likelihood Principle; Springer New York, 1998; pp 199–213. https://doi.org/10.1007/978-1-4612-1694-0_15.

(127) Riva, M.; Källberg, E.; Björk, P.; Hancz, D.; Vogl, T.; Roth, J.; Ivars, F.; Leanderson, T. Induction of Nuclear Factor-κB Responses by the S100A9 Protein Is Toll-like Receptor-4-Dependent. *Immunology* **2012**, *137* (2), 172–182. https://doi.org/10.1111/j.1365-2567.2012.03619.x.

(128) 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.

(129) Austermann, J.; Friesenhagen, J.; Fassl, S. K.; Ortkras, T.; Burgmann, J.; Barczyk-Kahlert, K.; Faist, E.; Zedler, S.; Pirr, S.; Rohde, C.; et al. Alarmins MRP8 and MRP14 Induce Stress Tolerance in Phagocytes under Sterile Inflammatory Conditions. *Cell Rep.* **2014**, *9* (6), 2112–2124. https://doi.org/10.1016/j.celrep.2014.11.020.

(130) 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-κB Pathways. *Immunology* **2015**, *144* (1), 79–90. https://doi.org/10.1111/imm.12352.

(131) 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.

(132) Chen, G. Y.; Nuñez, G. Sterile Inflammation: Sensing and Reacting to Damage. *Nat. Rev. Immunol.* **2010**, *10* (12), 826–837. https://doi.org/10.1038/nri2873.

(133) 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.

(134) Cesaro, A.; Anceriz, N.; Plante, A.; Pagé, N.; Tardif, M. R.; Tessier, P. A. An Inflammation Loop Orchestrated by S100A9 and Calprotectin Is Critical for Development of Arthritis. *PLoS One* **2012**, *7* (9), e45478. https://doi.org/10.1371/journal.pone.0045478.

(135) Foell, D.; Wittkowski, H.; Ren, Z.; Turton, J.; Pang, G.; Daebritz, J.; Ehrchen, J.; Heidemann, J.; Borody, T.; Roth, J.; et al. Phagocyte-Specific S100 Proteins Are Released from Affected Mucosa and Promote Immune Responses during Inflammatory Bowel Disease. *J. Pathol.* **2008**, *216* (2), 183–192. https://doi.org/10.1002/path.2394.

(136) Björk, P.; Källberg, E.; Wellmar, U.; Riva, M.; Olsson, A.; He, Z.; Törngren, M.; Liberg, D.; Ivars, F.; Leanderson, T. Common Interactions between S100A4 and S100A9 Defined by a Novel Chemical Probe. *PLoS One* **2013**, *8* (5), e63012. https://doi.org/10.1371/journal.pone.0063012.

(137) 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.

(138) Mattis, D. M.; Chervin, A. S.; Ranoa, D. R.; Kelley, S. L.; Tapping, R. I.; Kranz, D. M. Studies of the TLR4-Associated Protein MD-2 Using Yeast-Display and Mutational Analyses. *Mol. Immunol.* **2015**, *68* (2), 203–212. https://doi.org/10.1016/j.molimm.2015.08.008.

(139) Vasl, J.; Oblak, A.; Gioannini, T. L.; Weiss, J. P.; Jerala, R. Novel Roles of Lysines 122, 125, and 58 in Functional Differences between Human and Murine MD-2. *J. Immunol.* **2009**, *183* (8), 5138–5145. https://doi.org/10.4049/jimmunol.0901544.

(140) Gioannini, T. L.; Teghanemt, A.; Zhang, D.; Esparza, G.; Yu, L.; Weiss, J. Purified Monomeric ligand.MD-2 Complexes Reveal Molecular and Structural Requirements for Activation and Antagonism of TLR4 by Gram-Negative Bacterial Endotoxins. *Immunol. Res.* **2014**, *59* (1–3), 3–11. https://doi.org/10.1007/s12026-014-8543-y.

(141) Simard, J.-C.; Cesaro, A.; Chapeton-Montes, J.; Tardif, M.; Antoine, F.; Girard, D.; Tessier, P. A. S100A8 and S100A9 Induce Cytokine Expression and Regulate the NLRP3 Inflammasome via ROS-Dependent Activation of NF-κB1. *PLoS One* **2013**, *8* (8), e72138. https://doi.org/10.1371/journal.pone.0072138.

(142) Sauter, K. S.; Brcic, M.; Franchini, M.; Jungi, T. W. Stable Transduction of Bovine TLR4 and Bovine MD-2 into LPS-Nonresponsive Cells and Soluble CD14 Promote the Ability to Respond to LPS. *Vet. Immunol. Immunopathol.* **2007**, *118* (1–2), 92–104. https://doi.org/10.1016/j.vetimm.2007.04.017.

(143) Ohto, U.; Fukase, K.; Miyake, K.; Shimizu, T. Structural Basis of Species-Specific Endotoxin Sensing by Innate Immune Receptor TLR4/MD-2. *Proc. Natl. Acad. Sci.* **2012**, *109* (19), 7421–7426. https://doi.org/10.1073/pnas.1201193109.

(144) Zimmer, D. B.; Eubanks, J. O.; Ramakrishnan, D.; Criscitiello, M. F. Evolution of the S100 Family of Calcium Sensor Proteins. *Cell Calcium* **2013**, *53* (3), 170–179. https://doi.org/10.1016/j.ceca.2012.11.006.

(145) Marenholz, I.; Heizmann, C. W.; Fritz, G. S100 Proteins in Mouse and Man: From Evolution to Function and Pathology (Including an Update of the Nomenclature). *Biochem. Biophys. Res. Commun.* **2004**, *322* (4), 1111–1122. https://doi.org/10.1016/j.bbrc.2004.07.096.

(146) 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.

(147) Kwek, J. H. L.; Wynne, A.; Lefèvre, C.; Familari, M.; Nicholas, K. R.; Sharp, J. A. Molecular Evolution of a Novel Marsupial S100 Protein (S100A19) Which Is Expressed at Specific Stages of Mammary Gland and Gut Development. *Mol. Phylogenet. Evol.* **2013**, *69* (1), 4–16. https://doi.org/10.1016/j.ympev.2013.05.005.

(148) Ravasi, T.; Hsu, K.; Goyette, J.; Schroder, K.; Yang, Z.; Rahimi, F.; Miranda, L. P.; Alewood, P. F.; Hume, D. A.; Geczy, C. Probing the S100 Protein Family through Genomic and Functional Analysis. *Genomics* **2004**, *84* (1), 10–22. https://doi.org/10.1016/j.ygeno.2004.02.002.

(149) Chow, J. C.; Young, D. W.; Golenbock, D. T.; Christ, W. J.; Gusovsky, F. Toll-like Receptor-4 Mediates Lipopolysaccharide-Induced Signal Transduction. *J. Biol. Chem.* **1999**, *274* (16), 10689–10692. https://doi.org/10.1074/JBC.274.16.10689.

(150) Rallabhandi, P.; Bell, J.; Boukhvalova, M. S.; Medvedev, A.; Lorenz, E.; Arditi, M.; Hemming, V. G.; Blanco, J. C. G.; Segal, D. M.; Vogel, S. N. Analysis of TLR4 Polymorphic Variants: New Insights into TLR4/MD-2/CD14 Stoichiometry, Structure, and Signaling. *J. Immunol.* **2006**, *177* (1), 322–332. https://doi.org/10.4049/jimmunol.177.1.322.

(151) Cui, J.; Cheng, Y.; Belov, K. Diversity in the Toll-like Receptor Genes of the Tasmanian Devil (*Sarcophilus Harrisii*). *Immunogenetics* **2015**, *67* (3), 195–201. https://doi.org/10.1007/s00251-014-0823-0.

(152) Daly, K. A.; Lefévre, C.; Nicholas, K.; Deane, E.; Williamson, P. CD14 and TLR4 Are Expressed Early in Tammar (Macropus Eugenii) Neonate Development. *J. Exp. Biol.* **2008**, *211* (8).

(153) Rychlik, I.; Elsheimer-Matulova, M.; Kyrova, K. Gene Expression in the Chicken Caecum in Response to Infections with Non-Typhoid Salmonella. *Vet. Res.* **2014**, *45* (1), 119. https://doi.org/10.1186/s13567-014-0119-2.

(154) Wu, Z.; Rothwell, L.; Hu, T.; Kaiser, P. Chicken CD14, Unlike Mammalian CD14, Is Trans-Membrane rather than GPI-Anchored. *Dev. Comp. Immunol.* **2009**, *33* (1), 97–104. https://doi.org/10.1016/j.dci.2008.07.008.

(155) Juan, T. S.-C.; Hailman, E.; Kelley, M. J.; Wright, S. D.; Lichenstein, H. S. Identification of a Domain in Soluble CD14 Essential for Lipopolysaccharide (LPS) Signaling but Not LPS Binding. *J. Biol. Chem.* **1995**, *270* (29), 17237–17242. https://doi.org/10.1074/jbc.270.29.17237.

(156) Teghanemt, A.; Re, F.; Prohinar, P.; Widstrom, R.; Gioannini, T. L.; Weiss, J. P. Novel Roles in Human MD-2 of Phenylalanines 121 and 126 and Tyrosine 131 in Activation of Toll-like Receptor 4 by Endotoxin. *J. Biol. Chem.* **2008**, *283* (3), 1257–1266. https://doi.org/10.1074/jbc.M705994200.

(157) Seong, S.-Y.; Matzinger, P. Opinion: 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.

(158) Matzinger, P. The Danger Model: A Renewed Sense of Self. *Science* **2002**, *296* (5566), 301–305. https://doi.org/10.1126/science.1071059.

(159) Yates, A.; Akanni, W.; Amode, M. R.; Barrell, D.; Billis, K.; Carvalho-Silva, D.; Cummins, C.; Clapham, P.; Fitzgerald, S.; Gil, L.; et al. Ensembl 2016. *Nucleic Acids Res.* **2016**, *44* (D1), D710–D716. https://doi.org/10.1093/nar/gkv1157.

(160) Aberer, A. J.; Kobert, K.; Stamatakis, A. ExaBayes: Massively Parallel Bayesian Tree Inference for the Whole-Genome Era. *Mol. Biol. Evol.* **2014**, *31* (10), 2553–2556. https://doi.org/10.1093/molbev/msu236.

(161) Herrero, J.; Muffato, M.; Beal, K.; Fitzgerald, S.; Gordon, L.; Pignatelli, M.; Vilella, A. J.; Searle, S. M. J.; Amode, R.; Brent, S.; et al. Ensembl Comparative Genomics Resources. *Database* **2016**, *2016*, bav096. https://doi.org/10.1093/database/bav096.

(162) Arnold, K.; Bordoli, L.; Kopp, J.; Schwede, T. The SWISS-MODEL Workspace: A Web-Based Environment for Protein Structure Homology Modelling. *Bioinformatics* **2006**, *22* (2), 195–201. https://doi.org/10.1093/bioinformatics/bti770.

(163) Biasini, M.; Bienert, S.; Waterhouse, A.; Arnold, K.; Studer, G.; Schmidt, T.; Kiefer, F.; Cassarino, T. G.; Bertoni, M.; Bordoli, L.; et al. SWISS-MODEL: Modelling Protein Tertiary and Quaternary Structure Using Evolutionary Information. *Nucleic Acids Res.* **2014**, *42* (W1), W252–W258. https://doi.org/10.1093/nar/gku340.

(164) 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 - Mol. Cell Res.* **2006**, *1763* (11), 1298–1306. https://doi.org/10.1016/J.BBAMCR.2006.08.028.

(165) Weber, D. J.; Rust, R. R.; Baldisseri, D. M. Structure of the Negative Regulatory Domain of p53 Bound to S100B(ββ). *Nat. Struct. Biol.* **2000**, *7* (7), 570–574. https://doi.org/10.1038/76797.

(166) Lewit-Bentley, A.; Réty, S.; Sopkova, J.; Renouard, M.; Osterloh, D.; Gerke, V.; Tabaries, S.; Russo-Marie, F. The Crystal Structure of a Complex of p11 with the Annexin II N-Terminal Peptide. *Nat. Struct. Biol.* **1999**, *6* (1), 89–95. https://doi.org/10.1038/4965.

(167) Kiss, B.; Duelli, A.; Radnai, L.; Kékesi, K. A.; Katona, G.; Nyitray, L. Crystal Structure of the S100A4-Nonmuscle Myosin IIA Tail Fragment Complex Reveals an Asymmetric Target Binding Mechanism. *Proc. Natl. Acad. Sci. U. S. A.* **2012**, *109* (16), 6048–6053. https://doi.org/10.1073/pnas.1114732109.

(168) Lee, Y.-T.; Dimitrova, Y. N.; Schneider, G.; Ridenour, W. B.; Bhattacharya, S.; Soss, S. E.; Caprioli, R. M.; Filipek, A.; Chazin, W. J. Structure of the S100A6 Complex with a Fragment from the C-Terminal Domain of Siah-1 Interacting Protein: A Novel Mode for S100 Protein Target Recognition † ‡. *Biochemistry* **2008**, *47* (41), 10921–10932. https://doi.org/10.1021/bi801233z.

(169) Vogl, T.; Stratis, A.; Wixler, V.; Völler, T.; Thurainayagam, S.; Jorch, S. K.; Zenker, S.; Dreiling, A.; Chakraborty, D.; Fröhling, M.; et al. Autoinhibitory Regulation of S100A8/S100A9 Alarmin Activity Locally Restricts Sterile Inflammation. *J. Clin. Invest.* **2018**, *128* (5), 1852–1866. https://doi.org/10.1172/JCI89867.

(170) 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.

(171) 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.; et al. Molecular Basis for Manganese Sequestration by Calprotectin and Roles in the Innate Immune Response to Invading Bacterial Pathogens. *Proc. Natl. Acad. Sci. U. S. A.* **2013**, *110* (10), 3841–3846. https://doi.org/10.1073/pnas.1220341110.

(172) 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.

(173) 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.

(174) Carvalho, S. B.; Botelho, H. M.; Leal, S. S.; Cardoso, I.; Fritz, G.; Gomes, C. M. Intrinsically Disordered and Aggregation Prone Regions Underlie β-Aggregation in S100 Proteins. *PLoS One* **2013**, *8* (10), e76629. https://doi.org/10.1371/journal.pone.0076629.

(175) Fritz, G.; Botelho, H. M.; Morozova-Roche, L. A.; Gomes, C. M. Natural and Amyloid Self-Assembly of S100 Proteins: Structural Basis of Functional Diversity. *FEBS J.* **2010**, *277* (22), 4578–4590. https://doi.org/10.1111/j.1742-4658.2010.07887.x.

(176) Yanamandra, K.; Alexeyev, O.; Zamotin, V.; Srivastava, V.; Shchukarev, A.; Brorsson, A.-C.; Tartaglia, G. G.; Vogl, T.; Kayed, R.; Wingsle, G.; et al. Amyloid Formation by the Pro-Inflammatory S100A8/A9 Proteins in the Ageing Prostate. *PLoS One* **2009**, *4* (5), e5562. https://doi.org/10.1371/journal.pone.0005562.

(177) 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. *BBA - Proteins Proteomics* **2016**. https://doi.org/10.1016/j.bbapap.2016.08.008.

(178) Pagano, R. L.; Sampaio, S. C.; Juliano, L.; Juliano, M. A.; Giorgi, R. The C-Terminus of Murine S100A9 Inhibits Spreading and Phagocytic Activity of Adherent Peritoneal Cells. *Inflamm. Res.* **2005**, *54* (5), 204–210. https://doi.org/10.1007/s00011-005-1344-y.

(179) Pagano, R. L.; Moraes, N. F.; De Lorenzo, B. H.; Coccuzzo Sampaio, S.; Mariano, M.; Giorgi, R. Inhibition of Macrophage Functions by the C-Terminus of Murine S100A9 Is Dependent on B-1 Cells. *Mediators Inflamm.* **2014**, *2014*, 836491. https://doi.org/10.1155/2014/836491.

(180) Sopalla, C.; Leukert, N.; Sorg, C.; Kerkhoff, C. Evidence for the Involvement of the Unique C-Tail of S100A9 in the Binding of Arachidonic Acid to the Heterocomplex S100A8/A9. *Biol. Chem.* **2002**, *383* (12), 1895–1905. https://doi.org/10.1515/BC.2002.213.

(181) Corbin, B. D.; Seeley, E. H.; Raab, A.; Feldmann, J.; Miller, M. R.; Torres, V. J.; Anderson, K. L.; Dattilo, B. M.; Dunman, P. M.; Gerads, R.; et al. Metal Chelation and Inhibition of Bacterial Growth in Tissue Abscesses. *Science* **2008**, *319* (5865), 962–965. https://doi.org/10.1126/science.1152449.

(182) 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.

(183) Simpson, G. G. Organisms and Molecules in Evolution. *Sci. New Ser.* **1964**, *146* (3651), 1535–1538.

(184) Aronson, J. D. "Molecules and Monkeys": George Gaylord Simpson and the Challenge of Molecular. *Hist. Philos. Life Sci.* **2002**, *24* (4), 441–465.

(185) Mayr, E. Cause and Effect in Biology. *Science.* **1961**, *134* (3489), 1501–1506.

(186) Dobzhansky, T. Biology, Molecular and Organismic. *Am. Zool.* **1964**, *4*, 443–452.

(187) Akanuma, S.; Iwami, S.; Yokoi, T.; Nakamura, N.; Watanabe, H.; Yokobori, S.; Yamagishi, A. Phylogeny-Based Design of a B-Subunit of DNA Gyrase and Its ATPase Domain Using a Small Set of Homologous Amino Acid Sequences. *J. Mol. Biol.* **2011**, *412* (2), 212–225. https://doi.org/10.1016/J.JMB.2011.07.042.

(188) Voordeckers, K.; Brown, C. A.; Vanneste, K.; van der Zande, E.; Voet, A.; Maere, S.; Verstrepen, K. J. Reconstruction of Ancestral Metabolic Enzymes Reveals Molecular Mechanisms Underlying Evolutionary Innovation through Gene Duplication. *PLoS Biol.* **2012**, *10* (12), e1001446. https://doi.org/10.1371/journal.pbio.1001446.

(189) Clifton, B. E.; Jackson, C. J. Ancestral Protein Reconstruction Yields Insights into Adaptive Evolution of Binding Specificity in Solute-Binding Proteins. *Cell Chem. Biol.* **2016**, *23* (2), 236–245. https://doi.org/10.1016/J.CHEMBIOL.2015.12.010.

(190) Hobbs, J. K.; Shepherd, C.; Saul, D. J.; Demetras, N. J.; Haaning, S.; Monk, C. R.; Daniel, R. M.; Arcus, V. L. On the Origin and Evolution of Thermophily: Reconstruction of Functional Precambrian Enzymes from Ancestors of Bacillus. *Mol. Biol. Evol.* **2012**, *29* (2), 825–835. https://doi.org/10.1093/molbev/msr253.

(191) Akanuma, S.; Nakajima, Y.; Yokobori, S.-I.; Kimura, M.; Nemoto, N.; Mase, T.; Miyazono, K.-I.; Tanokura, M.; Yamagishi, A. Experimental Evidence for the Thermophilicity of Ancestral Life. https://doi.org/10.1073/pnas.1308215110.

(192) 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.

(193) Riva, M.; He, Z.; Källberg, E.; Ivars, F.; Leanderson, T.; Guignard, F.; Mauel, J.; Markert, M.; Schafer, B.; Heizmann, C.; et al. 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.

(194) 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.

(195) 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**, *34* (2), msw223. https://doi.org/10.1093/molbev/msw223.

(196) Panter, G.; Jerala, R. The Ectodomain of the Toll-like Receptor 4 Prevents Constitutive Receptor Activation. *J. Biol. Chem.* **2011**, *286* (26), 23334–23344. https://doi.org/10.1074/jbc.M110.205419.

(197) Godfroy, J. I.; Roostan, M.; Moroz, Y. S.; Korendovych, I. V.; Yin, H. Isolated Toll-like Receptor Transmembrane Domains Are Capable of Oligomerization. *PLoS One* **2012**, *7* (11), e48875. https://doi.org/10.1371/journal.pone.0048875.

(198) Gay, N. J.; Symmons, M. F.; Gangloff, M.; Bryant, C. E. Assembly and Localization of Toll-like Receptor Signalling Complexes. *Nat. Rev. Immunol.* **2014**, *14* (8), 546–558. https://doi.org/10.1038/nri3713.

(199) Daringer, N. M.; Schwarz, K. A.; Leonard, J. N. Contributions of Unique Intracellular Domains to Switchlike Biosensing by Toll-like Receptor 4. *J. Biol. Chem.* **2015**, *290* (14), 8764–8777. https://doi.org/10.1074/jbc.M114.610063.

(200) 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* (FEB). https://doi.org/10.3389/fimmu.2018.00304.

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

(202) Albert, R. Scale-Free Networks in Cell Biology. *J. Cell Sci.* **2005**, *118* (Pt 21), 4947–4957. https://doi.org/10.1242/jcs.02714.

(203) Ryu, J.-K.; Kim, S. J.; Rah, S.-H.; Kang, J. I.; Jung, H. E.; Lee, D.; Lee, H. K.; Lee, J.-O.; Park, B. S.; Yoon, T.-Y.; et al. Reconstruction of LPS Transfer Cascade Reveals Structural Determinants within LBP, CD14, and TLR4-MD2 for Efficient LPS Recognition and Transfer. *Immunity* **2017**, *46* (1), 38–50. https://doi.org/10.1016/J.IMMUNI.2016.11.007.

(204) Ishii, A.; Kawasaki, M.; Matsumoto, M.; Tochinai, S.; Seya, T. Phylogenetic and Expression Analysis of Amphibian *Xenopus* Toll-like Receptors. *Immunogenetics* **2007**, *59* (4), 281–293. https://doi.org/10.1007/s00251-007-0193-y.

(205) Zanoni, I.; Granucci, F. Role of CD14 in Host Protection against Infections and in Metabolism Regulation. *Front. Cell. Infect. Microbiol.* **2013**, *3*, 32. https://doi.org/10.3389/fcimb.2013.00032.

(206) Manukyan, M.; Triantafilou, K.; Triantafilou, M.; Mackie, A.; Nilsen, N.; Espevik, T.; Wiesmüller, K.-H.; Ulmer, A. J.; Heine, H. Binding of Lipopeptide to CD14 Induces Physical Proximity of CD14, TLR2 and TLR1. *Eur. J. Immunol.* **2005**, *35* (3), 911–921. https://doi.org/10.1002/eji.200425336.

(207) Keestra, A. M.; de Zoete, M. R.; van Aubel, R. A. M. H.; van Putten, J. P. M. The Central Leucine-Rich Repeat Region of Chicken TLR16 Dictates Unique Ligand Specificity and Species-Specific Interaction with TLR2. *J. Immunol.* **2007**, *178* (11), 7110–7119. https://doi.org/10.4049/JIMMUNOL.178.11.7110.

(208) Han, M. V; Zmasek, C. M. phyloXML: XML for Evolutionary Biology and Comparative Genomics. *BMC Bioinformatics* **2009**, *10* (1), 356. https://doi.org/10.1186/1471-2105-10-356.

(209) Benson, D. A.; Karsch-Mizrachi, I.; Lipman, D. J.; Ostell, J.; Sayers, E. W. GenBank. *Nucleic Acids Res.* **2009**, *37* (Database), D26–D31. https://doi.org/10.1093/nar/gkn723.

(210) Sayers, E. W.; Barrett, T.; Benson, D. A.; Bryant, S. H.; Canese, K.; Chetvernin, V.; Church, D. M.; DiCuccio, M.; Edgar, R.; Federhen, S.; et al. Database Resources of the National Center for Biotechnology Information. *Nucleic Acids Res.* **2009**, *37* (Database), D5–D15. https://doi.org/10.1093/nar/gkn741.

(211) Murphy, W. J.; Eizirik, E.; O'Brien, S. J.; Madsen, O.; Scally, M.; Douady, C. J.; Teeling, E.; Ryder, O. A.; Stanhope, M. J.; de Jong, W. W.; et al. Resolution of the Early Placental Mammal Radiation Using Bayesian Phylogenetics. *Science (80-.).* **2001**, *294* (5550), 2348–2351. https://doi.org/10.1126/science.1067179.

(212) Prum, R. O.; Berv, J. S.; Dornburg, A.; Field, D. J.; Townsend, J. P.; Lemmon, E. M.; Lemmon, A. R. A Comprehensive Phylogeny of Birds (Aves) Using Targeted next-Generation DNA Sequencing. *Nature* **2015**, *526* (7574), 569–573. https://doi.org/10.1038/nature15697.

(213) Yang, Z.; Kumar, S.; Nei, M. A New Method of Inference of Ancestral Nucleotide and Amino Acid Sequences. *Genetics* **1995**, *141* (4), 1641–1650.

(214) Rehli, M. Of Mice and Men: Species Variations of Toll-like Receptor Expression. *Trends Immunol.* **2002**, *23* (8), 375–378. https://doi.org/10.1016/S1471-4906(02)02259-7.

(215) Lizundia, R.; Sauter, K.-S.; Taylor, G.; Werling, D. Host Species-Specific Usage of the TLR4-LPS Receptor Complex. *Innate Immun.* **2008**, *14* (4), 223–231. https://doi.org/10.1177/1753425908095957.