Towards an Understanding of S100A9 Activation of TLR4: Incorporating a Biochemical and Evolutionary Perspective

by

Lauren Olivia Chisholm

A dissertation accepted and approved in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

in Chemistry

Dissertation Committee:

Dr. Scott Hansen, Chair

Dr. Michael Harms, Advisor

Dr. Calin Plesa, Core Member

Dr. Karen Guillemin, Core Member

Dr. Matthew Barber, Institutional Representative

University of Oregon

Summer 2024

DISSERTATION ABSTRACT

Lauren Olivia Chisholm

Doctor of Philosophy in Chemistry

Title: Towards an Understanding of S100A9 Activation of TLR4: Incorporating a Biochemical and Evolutionary Perspective

The central puzzle of my dissertation work is understanding how two molecules with very different physiochemical properties activate the same receptor, Toll-like receptor 4 (TLR4). TLR4 is an innate immune receptor that responds to both the bacterial glycosylated phospholipid LPS and small soluble host proteins. Despite decades of work, we have little mechanistic understanding of how soluble proteins activate this receptor. S100A9 is one such soluble protein, or Damage Associated Molecular Pattern (DAMP), that activates inflammatory pathways via Toll-like receptor 4 (TLR4). This activity plays important homeostatic roles in tissue repair, but can also contribute to inflammatory diseases. The mechanism of activation is unknown. Learning more about the mechanism of S100A9-induced inflammation can improve our understanding of many disease pathologies, as well as providing a promising new therapeutic target. In this dissertation I describe my work addressing this gap in the literature, using biochemical, biophysical, computational, and evolutionary methods.

This dissertation includes previously published and unpublished co-authored material.

CURRICULUM VITAE

NAME OF AUTHOR: Lauren Olivia Chisholm

GRADUATE AND UNDERGRADUATE SCHOOLS ATTENDED:

University of Oregon, Eugene
University of Wisconsin Madison

DEGREES AWARDED:

Doctor of Philosophy, Chemistry, 2024, University of Oregon
Master of Science, Chemistry, 2021, University of Oregon
Bachelor of Science, Biochemistry, Mathematics, 2017, University of Wisconsin Madison

AREAS OF SPECIAL INTEREST:

Biochemistry
Biophysics
Molecular Biology
Protein Evolution

PROFESSIONAL EXPERIENCE:

Graduate Research and Teaching Assistant, University of Oregon, 2019-2024

Associate Scientist, Invenra, 2017-2019

Integrated Solutions and Engineering Intern, Promega, 2016

Biological Science Aid, USDA-ARS, 2015-2017

GRANTS, AWARDS, AND HONORS:

John Keana Graduate Student Fellowship, University of Oregon, 2023-2024

NIH T32 Molecular Biology and Biophysics Training Program, University of Oregon, 2020-2022

Graduate Student Award for Excellence in the Teaching of Chemistry, University of Oregon, 2020

Dean's First Year Merit Award, University of Oregon, 2019

PUBLICATIONS:

Chisholm LO, Jaeger NM, Murawsky HE, Harms MJ (2024) S100A9 interacts with a dynamic region on CD14 to activate Toll-like receptor 4. :2024.05.15.594416. Available from: https://www.biorxiv.org/content/10.1101/2024.05.15.594416v1

Chisholm LO, Orlandi KN, Phillips SR, Shavlik MJ, Harms MJ (2024) Ancestral Reconstruction and the Evolution of Protein Energy Landscapes. Annual Review of Biophysics 53:127–146. Available from: https://www.annualreviews.org/content/journals/10.1146/annurev-biophys-030722-125440

Chisholm LO, Jeon CK, Prell JS, Harms MJ (2024) Changing expression system alters oligomerization and proinflammatory activity of recombinant human S100A9. :2024.08.14.608001. Available from: https://www.biorxiv.org/content/10.1101/2024.08.14.608001v1

Shavlik MJ, Peterson M, Fitzgerald B, Kotamarti A, Chisholm LO, Marqusee S, Harms MJ. (2024). Single Substitutions Alter the Accessibility of Color in the Local Sequence Spaces of Evolving GFP-Like Proteins. Manuscript in preparation.

Dieterich Mabin ME, Brunet J, Riday H, Lehmann L (2021) Self-Fertilization, Inbreeding, and Yield in Alfalfa Seed Production. Front. Plant Sci. [Internet] 12. Available from: https://www.frontiersin.org/articles/10.3389/fpls.2021.700708/full

ACKNOWLEDGMENTS

First and foremost, I would like to thank my advisor Dr. Mike Harms for not only his scientific expertise and guidance, but his unwavering support and positive attitude. His commitment to mentorship is truly extraordinary, and something I hope to emulate. I'd also like to thank my dissertation advisory committee Dr. Scott Hansen, Dr. Karen Guillemin, Dr. Calin Plesa, and Dr. Matt Barber for their advice and support throughout graduate school.

This dissertation would not have been possible without my husband Clark and the love and support of both of our families. I'm especially grateful to my parents Chris and Greg Lehmann for their lifelong support and encouragement to pursue a career in science. Thank you to my best friends Christine Pelto (PE), Jen Dennin (M.S.), and (soon to be Dr.) Emily Nettesheim for their support (and commiseration), and the many hours they spent on the phone with me over the past 5 years.

I thank current and former members of the Harms lab for helpful discussion and input, with special thanks to Dr. Kona Orlandi, Natalie Jaeger, Dr. Michael Shavlik, Hannah Murawsky, Dr. Jon Muyskens, and Sophia Phillips. The work presented here would not have been possible without the facilities, resources and training shared by the Institute of Molecular Biology's GC3F, the Hansen Lab, and the Barber Lab. In particular I'd like to thank Dr. Jeff Bishop, Adam Fries, Benjamin Duewell, Dr. Scott Hansen, Dr. Kristin Kohler, Dr. Nicole Paterson, and Dr. Matt Barber. This dissertation benefited from access to the University of Oregon high performance computing cluster, Talapas. My work was funded by the Molecular Biology and Biophysics Training Grant (5T32GM007759), the John Keana Graduate Student Fellowship, and Michael Harms' R01 (NIGMS R01-GM146114).

DEDICATION

I dedicate this dissertation to my husband Clark Chisholm and our cat Jeff. Thanks for moving across the country with me, here's to round two!

I dedicate this dissertation to my mom Chris Lehmann, and my grandmother Stephanie Kolowith.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

Table Page

CHAPTER ONE

INTRODUCTION

This dissertation covers my contributions to four bodies of work, regarding protein-protein interactions and their evolution in the innate immune system.

Chapter II is a pre-print of a manuscript covering changes in proinflammatory activity and biophysical characteristics of recombinant S100A9 caused by changing the expression system used to produce protein. As lead author, I designed and conducted the majority of the experiments, as well as being the major contributor to writing and editing. Chae Kyung Jeon and James Prell designed and conducted key experiments. Michael Harms is corresponding author and provided experimental guidance and oversaw writing and editing. This manuscript is available on *bioRxiv*, and is currently submitted to *Protein Science*.

Chapter III is a pre-print of a manuscript detailing the interaction between the innate immune proteins S100A9 and CD14. As lead author, I designed and conducted the majority of the experiments, as well as being the major contributor to writing and editing. Natalie Jaeger and Hannah Murawsky contributed equally to the manuscript, designing and conducting key experiments. Michael Harms is corresponding author, and conducted molecular dynamics simulations and analysis, as well as providing experimental guidance and overseeing writing and editing. This manuscript is available on *bioRxiv*, and is currently in revision at the *Journal of Biological Chemistry*.

Chapter IV is unpublished work developing a method for screening the function of TLR4 variants in high throughput. This method was developed with the goal of studying the evolution of TLR4 function. As lead author, I designed and conducted the majority of the experiments, as

well as being the major contributor to writing and editing. Michael Harms contributed to data analysis, as well as providing experimental guidance and overseeing writing and editing.

Chapter V is a co-authored review of ancestral sequence reconstruction and its use in studying protein evolution and biophysics, published in *Annual Reviews in Biophysics*. As lead author, I oversaw writing and editing. Kona Orlandi, Sophia Phillips, and Michael Shavlik contributed equally to the writing of this review. Michael Harms is corresponding author and a major contributor to the writing and editing.

This introduction will provide background and context for the following chapters, with each chapter containing more specific information on the respective topic.

The Innate Immune System

Inflammation is mediated by the innate immune system, and is the body's first line of defense against infection or injury. The innate immune system is considered to be nonspecific and utilizes pattern recognition receptors (PRRs) to identify and respond to both internal and external danger signals. Different PRRs recognize different molecular patterns that share structural similarities, but are distinguishable between types of pathogen as well as host molecules. When a PRR is activated by its respective ligand, an inflammatory cascade is triggered, resulting in the release of various cytokines. Cytokine release stimulates further activation of the immune system and enables the body to mount the required immune response. This is an important part of a healthy immune response, however dysregulation of the innate immune system has grave consequences for human health. One in three deaths worldwide are linked to chronic inflammatory conditions[1].

Toll-Like Receptor 4

An important class of PRRs are the Toll-Like Receptors (TLRs). Humans have 10 different TLRs, each recognizes a particular external danger signal, or microbe associated molecular pattern (MAMPs). Some TLRs also recognize internal danger signals, or danger associated molecular patterns (DAMPs)[2]. TLRs consist of an extracellular pattern recognition leucine-rich-repeat domain, a transmembrane domain, and an intracellular signal mediating TIR (toll-IL-1 receptor) domain.

This dissertation focuses on the innate immune receptor TLR4. TLR4 is the canonical receptor for the MAMP lipopolysaccharide (LPS), a component of the outer membrane of gram-negative bacteria such as E.Coli[3]. Responding to and clearing gram-negative bacterial infection is a vital aspect of human immunity, but can have deadly consequences when dysregulated. Gram negative bacterial infections are the prevailing cause of sepsis, which accounts for 19.7% of deaths worldwide between 1990-2017[4]. Gram negative bacterial infections are so critical to human health that the discoverer of TLR4 won the Nobel Prize in 2011[5]. However, despite years of development and clinical trials, there are no TLR4 antagonist drugs currently approved[6].

The mechanism by which TLR4 recognizes LPS is relatively well understood. LPS, or endotoxin, consists of Lipid A, an oligosaccharide core, and the highly variable O-antigen[7]. Lipid A is the conserved molecular pattern recognized by TLR4. TLR4 utilizes two co-receptors to respond to LPS: MD-2 and CD14. CD14 is tethered to the cell membrane by a GPI-anchor, binds to extracellular LPS via an N-terminal binding pocket, and delivers LPS to TLR4/MD-2[8–10]. TLR4 and MD-2 form a heterodimer, MD-2 contains a large hydrophobic binding pocket, and together TLR4/MD-2 bind to LPS[11,12]. Upon LPS binding, TLR4/MD-2 forms a homodimer of heterodimers – 2(TLR4/MD-2). Formation of this homodimer results in conformational changes

to the TIR domains of TLR4, which in turn triggers adapter protein binding and downstream inflammation.

S100A9

TLR4 also recognizes and responds to the DAMP S100A9[13–15]. S100A9 is a small calcium binding protein that is highly abundant in neutrophils[16]. S100A9 exists as both a homodimer and heterodimer with the protein S100A8, known as calprotectin. As calprotectin, S100A9 is a vital anti-microbial protein, binding and sequestering transition metals that bacteria need to live[17]. This dissertation focuses largely on the homodimeric form of S100A9, which is secreted by neutrophils during various inflammatory states and activates the TLR4 complex. As is the case with LPS, activation by S100A9 has important consequences for human health: it is part of a healthy immune response[18,19], but has been associated with various inflammatory diseases such as cancer[20–22] and neurodegenerative diseases[23,24] .

The mechanism by which S100A9 activates TLR4 is poorly understood. S100A9 requires the same protein receptor components as LPS[25,26] (TLR4, MD2, and CD14). This is surprising because it S100A9 and LPS have radically different size, structure, and physiochemical properties. The lack of a biochemical mechanism has stunted discovery efforts for drugs targeting the S100A9/TLR4 interaction. For example, quinoline-3-carboxamide drugs were reported to target S100s—including S100A9—but failed in the clinic due to lack of specificity[27]. This leads to one of the key questions of my dissertation research: how does the TLR4 complex recognize S100A9? In this dissertation I describe the use of biochemical, biophysical, computational, and evolutionary techniques to further our understanding of the interaction between S100A9 and the TLR4 complex.

Protein Evolution

Studying protein evolution can provide important insight into a proteins function and biophysical features. TLR4 provides an excellent example for studying protein evolution. Proper TLR4 function is essential for human health, with even single amino acid mutations linked to increased susceptibility to gram-negative bacterial infection and sepsis[28–30]. TLR4 emerged in bony vertebrates, with the ability to recognize to LPS[31], and evolved the ability to recognize S100A9 in amniotes[25]. TLR4 is under intense evolutionary pressure to maintain the ability to respond to both the bacterial lipid LPS and the host protein S100A9.

LPS and S100A9 are expected to introduce competing evolutionary pressures on their shared receptor. Bacteria continuously evolve to avoid activating the immune system; this creates selection pressure for TLR4 to maintain its interaction with LPS, and thus should increase the rate of sequence evolution on TLR4. In contrast, two host proteins that interact are expected to avoid sequence changes to maintain their interaction. As a result, S100A9 is expected to create pressure to slow the evolution of TLR4. This leads to the second key question of my dissertation research: how does the TLR4 complex resolve these competing evolutionary pressures to maintain its response to both microbial and host activators? In this dissertation I address two methods for studying protein evolution: dN/dS ratios and ancestral sequence reconstruction.

dN/dS

Mutations occur on the DNA level, but selection acts at the amino acid level. A synonymous mutation is a change in the coding DNA sequence that does not result in an amino acid change – the codons are synonymous. A nonsynonymous mutation is a change in the coding

DNA sequence that does result in an amino acid change. Given a multiple sequence alignment of coding sequences for a protein within a clade, one can take the ratio of nonsynonymous : synonymous mutations (dN/dS, or omega) at each site to measure rates of selection[32,33].

A dN/dS ratio of 1 means that a mutation at a site is equally likely to be synonymous or nonsynonymous. This is interpreted as a site where changing the amino acid at this position has no effect on fitness. A ratio >1 indicates positive selection: this position experiences more amino acid changes than is expected based on the rate of synonymous substitutions, suggesting that selection is driving increased diversity at this site. A ratio <1 indicates negative selection: it is unfavorable for this position to undergo amino acid changes, diversifying this position decreases fitness. Given the short generation time of microbes, amino acids involved in microbe recognition tend to be under positive/diversifying selection. That is, the host protein is under selective pressure to diversify to keep up with the rapidly mutating bacteria.

In this case, TLR4 is under selective pressure to mutate rapidly to adapt to changing LPS structures and maintain recognition. Conversely, host protein-protein interactions evolve on the same time scale, and are facing selective pressure to be maintained. In the case of TLR4, there is selective pressure to maintain TLR4's interactions with MD-2 and CD14, as well as its ability to recognize S100A9. Thus, studying selective pressures at the amino acid level via dN/dS can provide predictive power to determine amino acid positions important to various functions. Increasing predictive power is especially useful in the case of very large proteins such as TLR4 (800+ aa). In this dissertation I describe the use of dN/dS ratios to inform site selection for mutational scanning.

Ancestral Sequence Reconstruction

Another powerful method for studying the evolution of protein function is ancestral sequence reconstruction (ASR). In ASR, a multiple sequence alignment of extant protein coding sequences is used to infer the protein sequences of ancestral nodes on the phylogenetic tree. These ancestral proteins can then be generated and characterized in the lab. Resurrecting ancestral states can help pinpoint when a function emerged, as well as aiding in determining causative mutations, or amino acid changes that are responsible for a change in function[34,35]. ASR can also be used to study protein folding, thermodynamics and kinetics. In the case of TLR4, ASR has been used to study how human TLR4 evolved the ability to distinguish between the agonist Lipid A, and the antagonist Lipid Iva[36]. In this dissertation I review the use of ASR to reveal the influence of protein energy landscapes on protein evolution.

Bridge to Chapter II:

In Chapter I I provided background information on the topic of the innate immune system and protein evolution. I described in brief what is known about the evolution and mechanism of the innate immune receptor TLR4, and its agonists LPS and S100A9. I also introduced two methods used to study protein evolution: dN/dS and ancestral sequence reconstruction. I introduced the two key questions of my thesis: 1) How does S100A9 activate TLR4? And 2) How did TLR4 evolve and maintain LPS and S100A9 recognition? In Chapter II, I begin to address question 1, and report my recent work on the impact of changing the expression system of recombinant S100A9. More specifically, I report changes to S100A9's biophysical characteristics and its ability to activate TLR4.

CHAPTER TWO

CHANGING EXPRESSION SYSTEM ALTERS OLIGOMERIZATION AND PROINFLAMMATORY ACTIVITY OF RECOMBINANT HUMAN S100A9.

*This chapter contains previously published coauthored material.

Chisholm LO, Jeon CK, Prell JS, Harms MJ (2024) Changing expression system alters oligomerization and proinflammatory activity of recombinant human S100A9. :2024.08.14.608001. Available from: https://www.biorxiv.org/content/10.1101/2024.08.14.608001v1

Author Contributions:

Lauren Chisholm and Michael Harms conceptualized the study and designed experiments. Lauren Chisholm and Chae Jeon conducted the experiments and data analysis. James Prell oversaw experiments conducted by Chae Jeon. Lauren Chisholm and Chae Jeon created the figures. Lauren Chisholm and Michael Harms wrote and edited the manuscript.

Introduction:

S100A9 is a small, dimeric calcium binding protein that is highly abundant in neutrophils [37]. S100A9 is a Damage Associated Molecular Pattern (DAMP) that activates inflammation via Toll-like receptor 4 (TLR4)[13,14,26,38]. S100A9 also forms a heterodimer with S100A8 known as calprotectin, an anti-microbial protein functioning in nutritional immunity [16]. As a DAMP, S100A9 activates the immune receptor TLR4 by an unknown mechanism. This activity has been demonstrated many times, in many different systems[2,13–15,25,26,38,39], and has been associated with negative outcomes in neurodegenerative diseases[23,24] and many cancers[20–22].

One challenge for studies of S100A9 activity is that its receptor, TLR4, is the canonical receptor for lipopolysaccharide (LPS), a component of the outer membrane of gram-negative bacteria[3,7,40]. LPS is a common contaminant in recombinant proteins purified from *E. coli*, potentially leading to spurious activation of TLR4. Most studies of DAMPs activating TLR4 have used proteins expressed in gram-negative bacteria[13,14,26,41], followed by purification steps to remove LPS. Because of this, some have suggested that DAMP activation of TLR4 is nothing more than an experimental artifact[42]. Further, another member of the S100 family, S100A3, has been shown to adopt completely a different structure when the expression system was changed[43].

We decided to remove LPS contamination at the source and recombinantly express S100A9 in eukaryotic cells. We show that, while S100A9 prepped out of *E. coli* ($S100A9^{ec}$) activates TLR4, S100A9 prepped out of insect cells ($S100A9^{in}$) does not. We rule out protein misfolding, post-translational modification, and simple LPS contamination as causes for the difference in activity. We show that $S100A9^{in}$ displays altered oligomeric states compared to $S100A9^{ec}$. We find disrupting oligomer formation of $S100A9^{in}$ using an *E. coli* disaggregase restores proinflammatory activity. This suggests that S100A9 can indeed activate TLR4; however, its oligomeric state is a key determinant of proinflammatory activity.

Results and Discussion:

Changing S100A9 expression system changes proinflammatory activity.

We set out to purify human S100A9 from several cell types to determine the extent to which S100A9 un-contaminated with LPS could activate TLR4. We recombinantly expressed human S100A9 from three cell types: Rosetta BL21(DE3) pLysS *E. coli* ($S100A9^{ec}$), HighFive insect cells ($S100A9^{in}$), and HEK293F human cells ($S100A9^{hek}$). We were unable to purify

S100A9^{hek} due to extensive proteolysis of S100A9 by human cells [44,45]. S100A9^{ec} and S100A9^{in} were readily expressed and purified to >99% purity by SDS-PAGE, so we focused on a comparison between these two recombinant proteins. For S100A9^{ec}, we purified the protein with 3 chromatography steps (Ni-NTA, followed by sequential anion exchange at pH 8, then pH 6); for S100A9^{in}, we achieved high purity in a single Ni-NTA step.

To measure TLR4 activity, we used a previously established activity assay[25,46,47]. In this assay, we transiently transfect HEK293T cells with plasmids encoding TLR4 and its co-receptors MD-2 and CD14, then measure TLR4 activity with a firefly luciferase behind an NF-κB promoter. In this way we can test the TLR4-specific proinflammatory activity of various agonists. To control for possible LPS contamination of recombinantly prepared proteins, we can include polymyxin B (PB), which binds to LPS and prevents LPS activation of TLR4 (Fig 2.1A)[25,26,47].

We first tested S100A9^{ec}, finding it activates TLR4 even with large amounts of polymyxin B (PB) (Fig 2.1A). This matches previous reports[13,25,26,48]. We next tested S100A9^{in}. To our surprise, it did not activate TLR4 (Fig 2.1A). To rule out a problem with our purification technique, we purchased and tested commercially available S100A9^{in} (Sino Biological). Like our prepared protein, commercial S100A9^{in} did not activate TLR4 (Fig 2.1B).

S100A9^{ec} does not deliver LPS.

We hypothesized that S100A9^{ec} was activating TLR4 by delivering LPS to TLR4, rather than S100A9 directly activating TLR4. To test this, we mimicked LPS contamination by pre-incubating S100A9^{in} with purified LPS. We then treated this sample with PB—in the same way

we treated S100A9^{ec}—and tested its activity. This treatment condition was also inactive (Fig 2.1B), ruling out LPS delivery by S100A9.

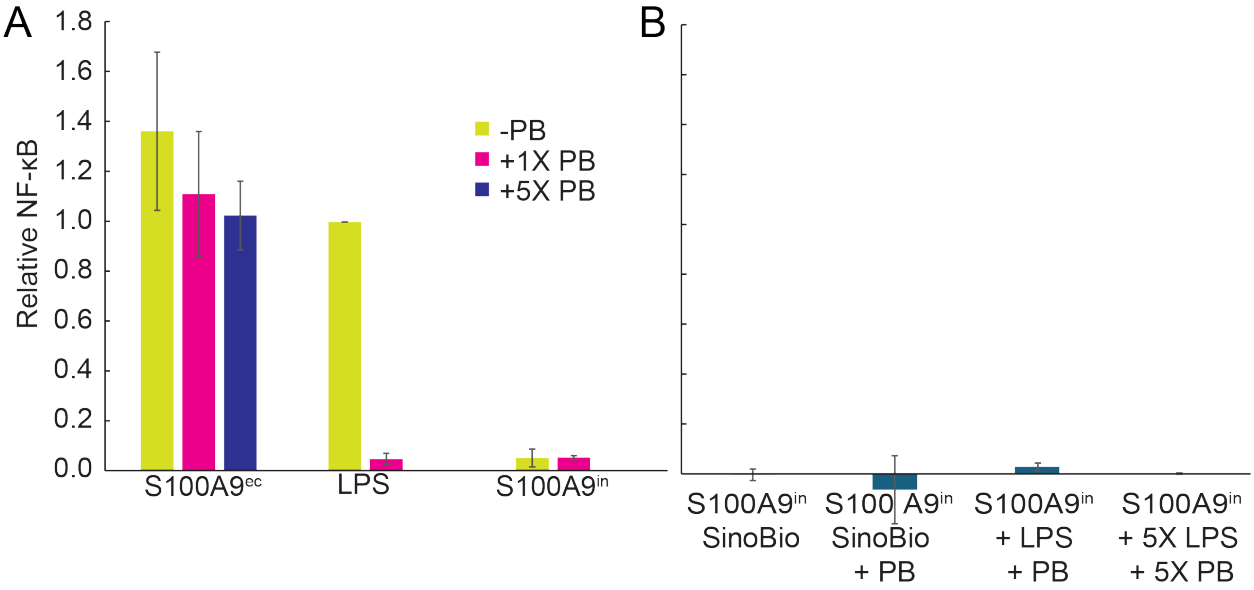


Figure 2.1: S100A9^{in} does not activate TLR4, and does not deliver LPS to TLR4. A) Relative NF-κB in response to S100A9^{ec} with multiple PB concentrations (0X, 1X and 5X), LPS with and without PB, and S100A9^{in} with and without PB. NF-κB output is normalized such that the reponse to LPS = 1.0. Bar color indicates amount of PB added. B) Relative NF-κB in response to S100A9^{in} that has been pre-incubated with LPS, (followed by the addition of PB), and to S100A9^{in} purchased from Sino Biological. For 1X: LPS concentration is 200ng/mL, 1X PB concentration is 200ug/mL, S100A9 concentration is 2uM. Data displayed is the average of 3+ biological replicates, error bars indicate standard error of the mean (SEM).

The difference is not due to a post-translational modification

We next hypothesized the difference was due to a post-translational modification of S100A9^{in} not present in S100A9^{ec}. S100A9 has one well characterized post-translational modification – a phosphorylation at position T113[49–51]. This has been shown to modulate certain activities of S100A9. For example phosphorylation inhibits S100A9-induced polymerization of tubulin[51]. It has also been reported that phosphorylation modulates the inflammatory state of calprotectin[52]. To assess whether S100A9^{in} might be phosphorylated at position T113, we performed western blots against the protein using either a generic anti-S100A9 antibody (1C22

Abnova) or antibody specific to S100A9 T113-p (#12782 Signalway Antibody). We found that the generic antibody recognized both $S100A9^{ec}$ and $S100A9^{in}$ (Fig 2.2A), but that the phosphorylation-specific antibody only recognized $S100A9^{in}$ (Fig 2.2A). We further validated the phosphorylation of T113 using top-down mass spectrometry (Oregon State University, Mass Spectrometry Core).

We hypothesized that phosphorylation was somehow modulating TLR4 activation. To test this hypothesis, we first attempted to dephosphorylate $S100A9^{in}$ by digesting with commercially available calf intestinal alkaline phosphatase (CIP – New England Biolabs). Our efforts at dephosphorylation were, however, unsuccessful as assessed by both western blot and MALDI-TOF mass spectrometry. We then turned to site-directed mutagenesis of $S100A9^{in}$ and $S100A9^{ec}$. To remove phosphorylation in the insect cell protein, we introduced $S100A9^{in}$ T113N; to mimic phosphorylation in the *E. coli* protein, we introduced $S100A9^{ec}$ T113D.

Neither removing the phosphorylation nor introducing a phosphomimetic had any effect: $S100A9^{in}$ T113N did not restore activity to $S100A9^{in}$ (Fig 2.2B), while $S100A9^{ec}$ T113D did not disrupt TLR4 activity from wildtype $S100A9^{ec}$ (Fig 2.2B). We also tested $S100A9^{in}$ T113N in a cysteine free background (C3S) (Fig 2.2B), to confirm that disulfide formation was not responsible for the difference in activity. These results suggest that some other feature of the protein, not the post-translational modification, differs between $S100A9^{in}$ and $S100A9^{ec}$.

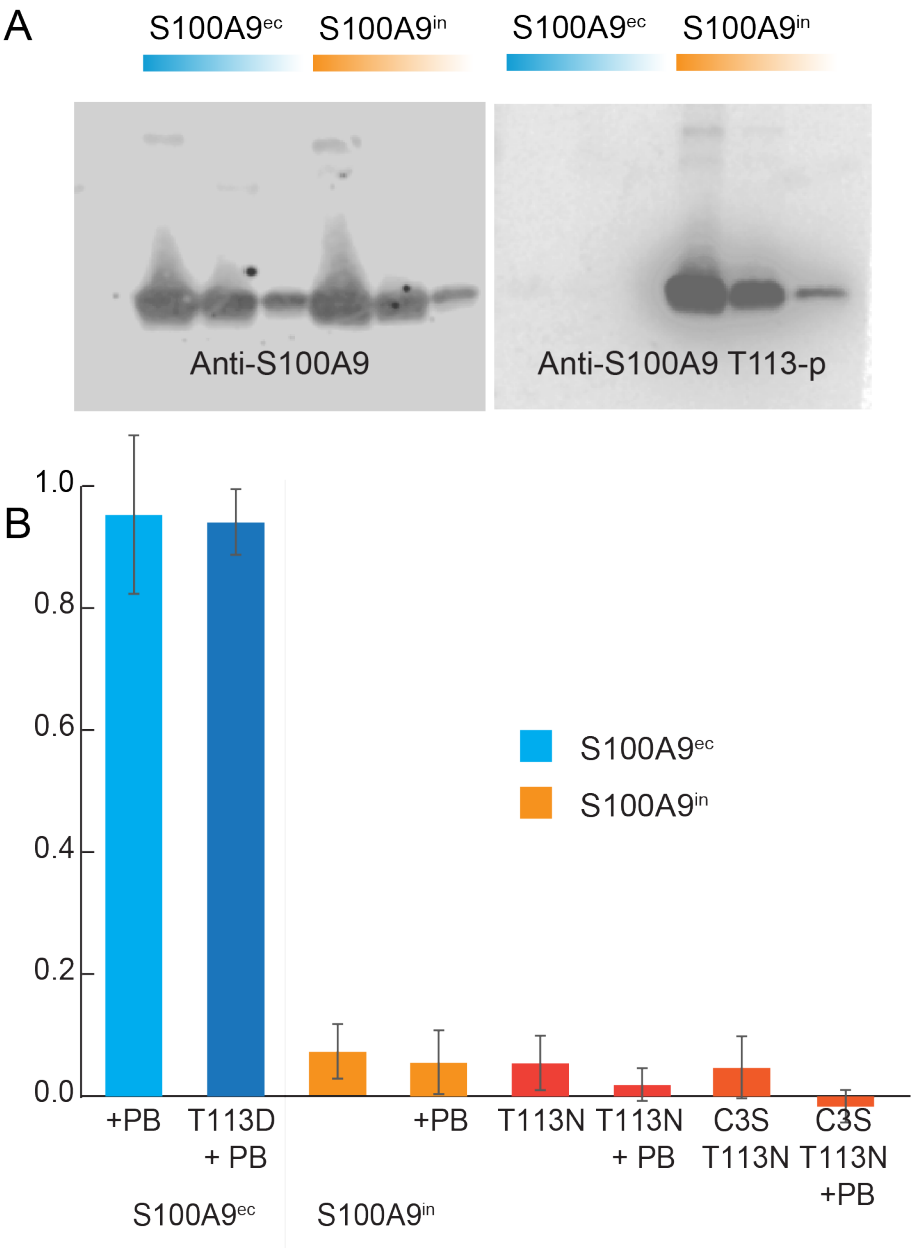


Figure 2.2: S100A9^{in} is phosphorylated, but this does not alter activity. A) western blot of S100A9^{ec} and S100A9^{in} with anti-S100A9 antibody (1C22 Abnova) and anti-S100A9 T113-p antibody (#12782 Signalway Antibody). The western blot images were cropped to exclude unnecessary white space, and converted to greyscale for improved visibility. B) Relative NF-κB in response to T113 mutations to S100A9^{in} and S100A9^{ec}. NF-κB is normalized such that wildtype S100A9^{ec} response = 1.0. PB concentration is 200ug/mL, S100A9 concentration is 2uM. Data displayed is the average of 3+ biological replicates, error bars indicate standard error of the mean (SEM).

S100A9^{in} and S100A9^{ec} likely have different high-order structures

We set out to determine other differences between S100A9^{ec} and S100A9^{in} that might explain the difference in activity. We measured three spectra: far-UV circular dichroism (CD), near-UV CD, and intrinsic fluorescence. We made these measurements both in the presence and absence of calcium, as S100A9 is known to undergo a conformational change in response to calcium binding[53].

Based on their far-UV CD spectra, both S100A9^{ec} and S100A9^{in} were primarily α-helical and responded similarly to the addition of calcium (Figure 2.3A). Likewise, we observed nearly identical intrinsic fluorescence spectra for both proteins (Figure 2.3B). The two purifications differed, however, in their near-UV CD spectra. S100A9^{in} exhibited a strong peak around 265 nm that was absent in S100A9^{ec}. This suggested a difference in the tertiary or quaternary structures of the two proteins.

Based on their far-UV CD spectra, both S100A9^{ec} and S100A9^{in} were primarily α-helical and responded similarly to the addition of calcium (Figure 2.3A). Likewise, we observed nearly identical intrinsic fluorescence spectra for both proteins (Figure 2.3B). The two purifications differed, however, in their near-UV CD spectra. S100A9^{in} exhibited a strong peak around 265 nm that was absent in S100A9^{ec}. This suggested a difference in the tertiary or quaternary structures of the two proteins.

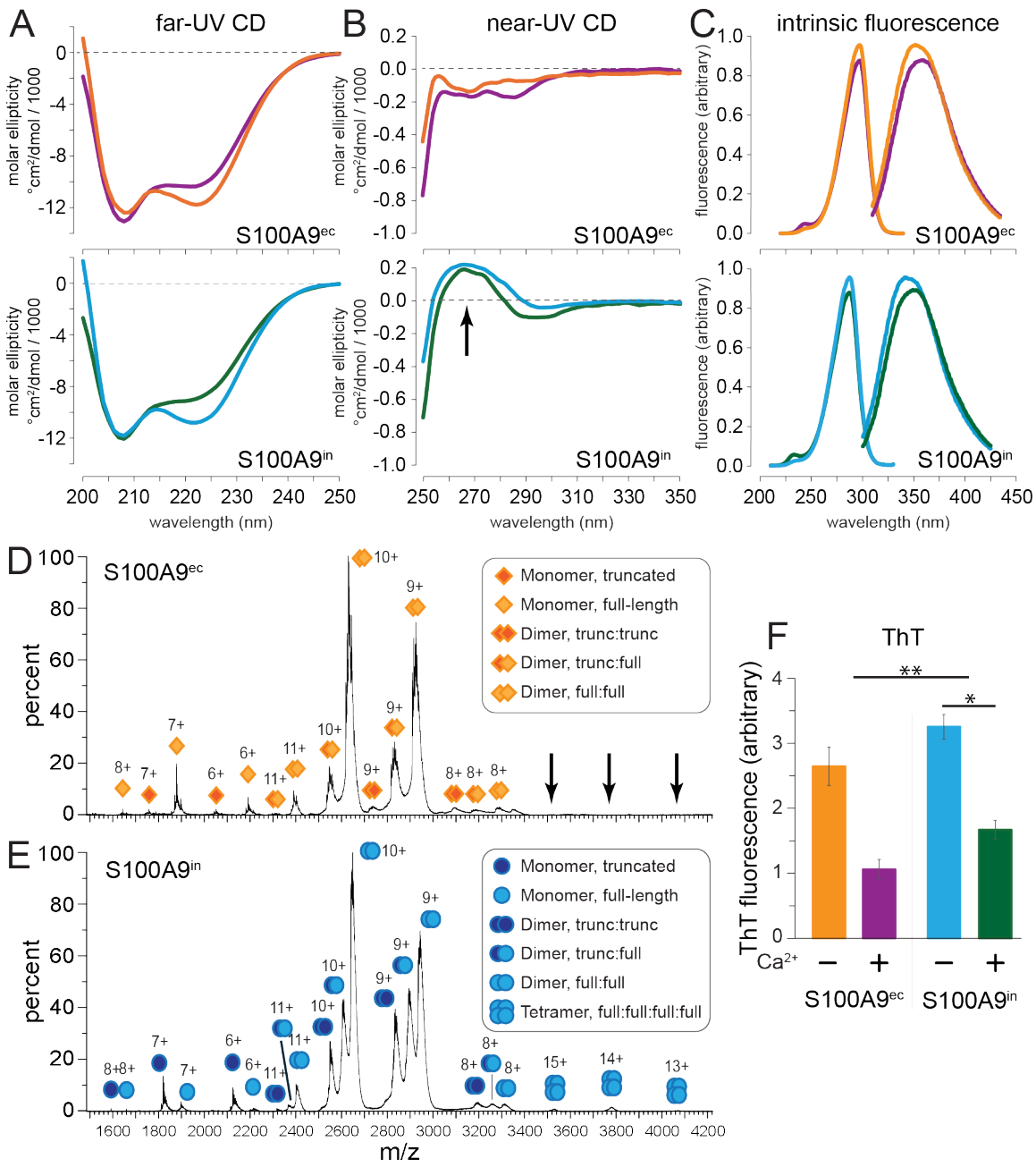
The proteins have different oligomeric and aggregation behaviors

We attempted to determine the tertiary structure of S100A9^{in} using x-ray crystallography but were unable to obtain usable crystals. To probe the quaternary structure, we used native mass spectrometry (nMS) to measure the oligomeric states of S100A9^{in} and S100A9^{ec} as previously

reported[54]. S100A9^{in} and S100A9^{ec} were measured at approximately the same bulk concentration. As expected, both proteins had peaks corresponding to dimers. S100A9^{in}, however, also populated a tetrameric form (Fig 2.3D & E). Of all species detected by native MS, the prominence of dimer peaks in both S100A9^{ec} and S100A9^{in} suggests that dimer form of S100A9 are the relevant and preferable species in solution.

S100A9 is also known to form amyloid like fibrils[55,56]. Therefore, we hypothesized that S100A9^{in} may also be more prone to amyloid fibril formation. Using the amyloid fibril specific fluorescent dye, ThioflavinT (ThT), we measured the capacity of S100A9^{in} and S100A9^{ec} to form these fibrils, in the presence and absence of calcium. S100A9^{in} exhibited a significantly higher ThT fluorescence compared to S100A9^{ec} regardless of buffer conditions ($p = 0.008$). This difference could be due to a difference in total amount of fibrils formed, or in the structure of the fibril. The addition of calcium significantly reduced ThT fluorescence for S100A9^{in} ($p = 0.035$), consistent with previous reports on S100A9^{ec} [55].

Figure 2.3. S100A9^{in} has a different quaternary structure than S100A9^{ec}. Throughout the figure, orange indicates S100A9^{ec}, purple indicates S100A9^{ec} + calcium, blue indicates S100A9^{in}, and green indicates S100A9^{in} + calcium. A) Far-UV CD spectra of S100A9^{ec} (top) and S100A9^{in} (bottom). B) Near-UV CD spectra of S100A9^{ec} (top) and S100A9^{in} (bottom). The arrow indicates the peak present in S100A9^{in} but not S100A9^{ec}. C) Intrinsic fluorescence emission and excitation spectra of S100A9^{ec} (top) and S100A9^{in} (bottom). Excitation spectra (emission at 345 nm) are on the left; emission spectra (excitation at 288 nm) are on the right. D) Native mass spectrum of S100A9^{ec}. Inferred species are annotated as indicated in the key. (We observed full-length and truncated forms of S100A9, as indicated). The locations of the tetramer peaks observed for S100A9^{in} but not S100A9^{ec} are indicated with arrows. E) Native mass spectrum of S100A9^{in}, annotated similarly to panel D. F) Amyloid formation as measured by ThT fluorescence (ex/em: 450/480) after an 8-hour incubation at 37 °C. Bars show representative results for one biological replicate; error bars are standard error of three technical replicates. P-values were calculated using a paired 2-tailed Student's t-test on four biological replicates. ** $P < 0.01$; * $P < 0.05$


A
far-UV CD
molar ellipticity °cm²/dmol / 1000
S100A9ec
S100A9in
wavelength (nm)
B
near-UV CD
C
intrinsic fluorescence
fluorescence (arbitrary)
D
E
percent
m/z
Monomer, truncated
Monomer, full-length
Dimer, trunc:trunc
Dimer, trunc:full
Dimer, full:full
Tetramer, full:full:full:full
F
ThT
ThT fluorescence (arbitrary)
Ca2+

Disrupting oligomer – restores some activity

We next sought to understand why the two purifications gave different oligomeric states. We hypothesized the cause might be the bacterial protein SlyD, which binds weakly to Ni-NTA columns and can co-purify with recombinant proteins[57,58]. SlyD acts as a chaperone and disaggregase and could thus plausibly alter the oligomeric state and activity of recombinant S100A9. Further, because it is a chaperone, even a tiny amount of contaminant could lead to the result—consistent with the observation that S100A9^{ec} was >99% pure by SDS-PAGE.

We revisited our purification fractions and discovered that a protein with a molecular weight consistent with SlyD (~25 kDa) eluted from our Ni-NTA column just prior to S100A9. We first established that this fraction ("HisA"), on its own, was not sufficient to activate TLR4 above background (P = 0.097) (Figure 2.4A). To test whether it could, however, increase the ability of S100A9^{in} to activate TLR4, we added this fraction to S100A9^{in}. This led to a high NF-κB signal, even in the presence of large excess of PB to sequester LPS contamination (P = 0.0041) (Figure 2.4A). Having observed this change in activity, we next wanted to see if it corresponded to the predicted change in oligomeric state of S100A9. We added "HisA" to the protein and measured its native mass spectrum. All monomer and dimer species observed were comparable to the previous nMS of S100A9^{in}; however, addition of HisA eliminated the tetramer peaks in the native mass spectrum (Fig 2.4C & D).

To test whether SlyD was cause of the change in activity, we tested the ability of commercially available SlyD (AbCam) to restore S100A9^{in}'s ability to activate TLR4. Similar to HisA, SlyD alone does not activate TLR4 in the presence of high concentrations of PB (P = 0.38) (Figure 2.4B). When pre-incubated with S100A9^{in} followed by the addition of PB, S100A9^{in} exhibits activity above background (P = 0.015) (Figure 4B). This activity is less than that

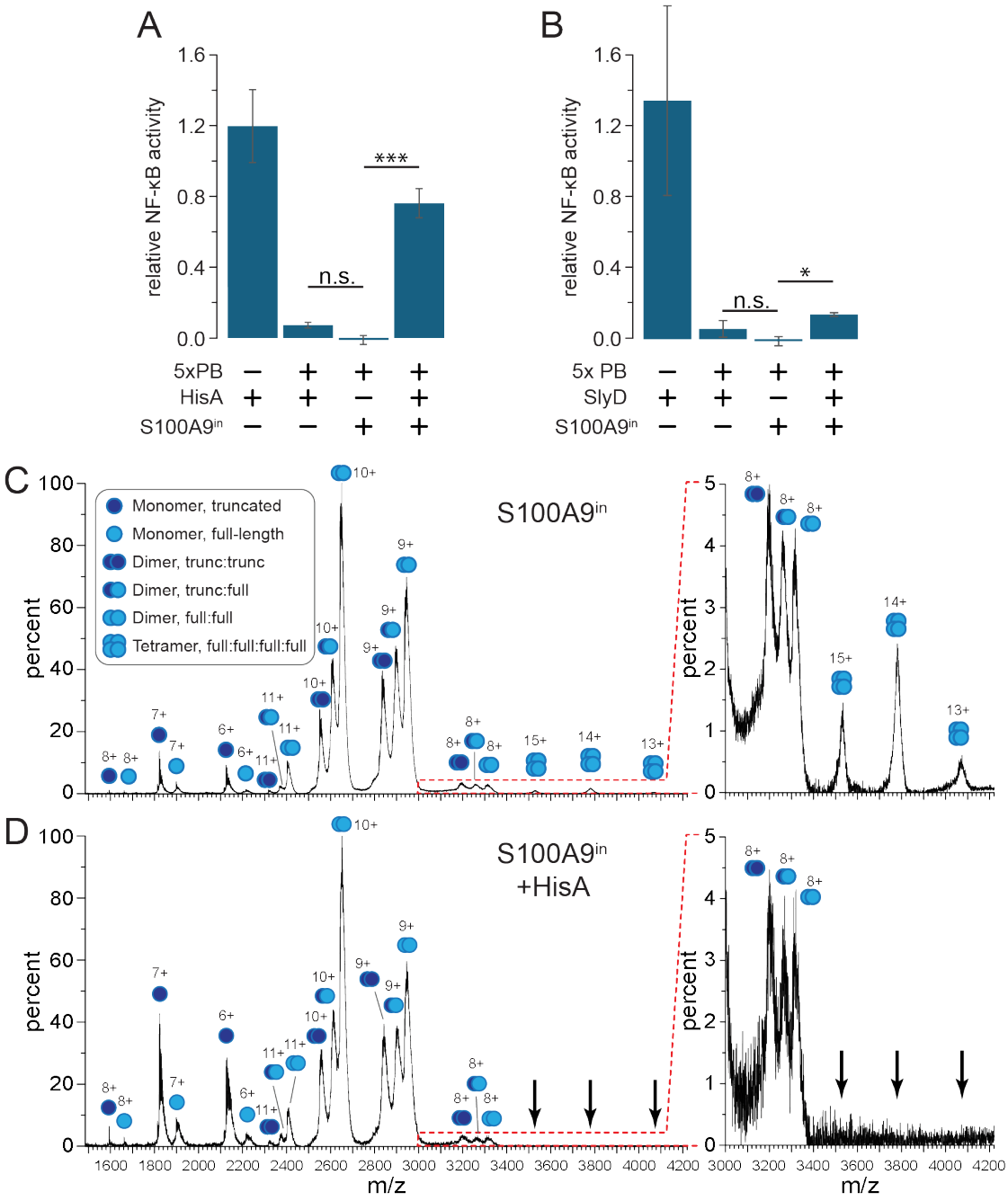


Figure 2.4: Bacterial protein contaminant SlyD alters S100A9 oligomeric state and activity. A) Purification fraction HisA restores some level of activity. B) Purified SlyD restores TLR4 activity, but not as potently as HisA. NF-κB is normalized such that Wildtype $S100A9^{ec}$ response = 1.0. PB concentration is 200 µg/mL, S100A9 concentration is 2 µM. SlyD and HisA concentration is 0.2 µM. Data displayed is the average of 3+ biological replicates, error bars indicate standard error of the mean (SEM). P-values were calculated using a paired 2-tailed Student's t-test. *** $P < 0.005$; * $P < 0.05$. C-D) Native mass spectra show disruption of tetramer with the addition of HisA fraction. Insets zoom in on tetramer region. Panel C shows the protein without the addition of HisA; panel D shows the protein after addition of HisA.

observed for HisA+S100A9^{in} (Figure 4A). This could be due to the concentration of SlyD in HisA differing from that in the purified sample, lower activity of the commercial preparation, or other factors in the HisA fraction contributing to the activity of the S100A9.

Conclusions:

We set out to purify LPS-free recombinant S100A9 using insect cells. We found that S100A9^{in} retained calcium binding and secondary structure but showed differences in tertiary structure and oligomeric state compared to S100A9^{ec}. Most importantly, S100A9^{in} lacks S100A9's reported proinflammatory activity. Our results also indicate that the oligomeric state of S100A9 is a major factor in its ability to activate the immune system. There are many open questions raised by our findings: What is the structure of S100A9^{in} and how does it differ from S100A9^{ec}? How does either structure compare to endogenously expressed human S100A9?

Our findings highlight the difficulty intrinsic in mapping results with recombinant proteins to their biological context. Indeed, this is not the first time changing the expression system of a recombinant S100 has turned the field on its head – S100A3 was found to adopt different structure when expressed in insect cells rather than *E. coli* [43].

Our findings also emphasize the delicate balance the immune system must maintain: even small changes to the tertiary structure and oligomeric state of S100A9 completely abolished its proinflammatory activity. Learning more about the native structure and oligomeric state of S100A9 will be critical to better understanding—and maybe someday intelligently modulating—its activity in the innate immune system.

Methods:

Native Mass Spectrometry (nMS):

Protein samples ($S100A9^{ec}$ and $S100A9^{in}$) were buffer exchanged into 200 mM ammonium acetate at pH 7 using Micro Bio-spin P6 columns (Bio-Rad, Hercules, CA). Buffer exchanged samples were then diluted to a working concentration of 10 μM. HisA was added in 1:1 volume with $S100A9^{in}$ immediately prior to the introduction to the instrument. These protein samples were analyzed with native MS on a Waters Synapt G2-S*i* quadrupole–ion mobility–time-of-flight instrument (Milford, MA). Samples were introduced using borosilicate capillary needles (prepared in-house with emitter i.d. ~450 nm) threaded with a platinum wire to allow nano-electrospray ionization (nESI). The electrospray was operated at capillary voltage of 0.3-0.5 kV with the sample cone and temperature at 25 V and 25 °C, respectively, in positive mode. The Trap cell was operated with an argon gas flow of 10 mL/min. The Collision Energy for both the Trap and Transfer cells was set to 5 V. For ion mobility separations, the IMS cell was pressurized at ~3.4 mbar nitrogen buffer gas. IM separation was performed with a traveling wave height of 30 V and wave velocity of 600 m/s. All native MS data were collected over the *m*/*z* range of 500-8000.

Recombinant S100A9 expression and purification from E. coli

We expressed and purified S100A9 as previously described [41,48]. Briefly, we expressed cysteine-free human S100A9 (C3S) in the pETDUET-1 vector in Rosetta BL21(DE3) pLysS *E. coli*. We purified S100A9 using three chromatography steps: immobilized metal ion affinity (HisTrap) at pH 7.4, anion exchange (HiTrap Q) at pH 8, followed by another anion exchange (HiTrap Q) at pH 6. We verified protein purity was >95% by SDS-PAGE. Proteins were stored

at -80 °C until needed. We determined protein concentration using A_{280} with an extinction coefficient of 6990 M^{-1} cm^{-1} (monomer). The protein concentrations reported in this manuscript are μM dimer.

Recombinant S100A9 expression in HighFive cells:

Recombinant expression of S100A9 in insect cells was performed using an existing protocol[59]. Bacmid DNA was generated by cloning human S100A9 into the pFastBac1 plasmid, which was then transformed into DH10Bac cells. The human S100A9 gene was purchased from Genscript (Clone ID: OHu25452C Accession No.: NM_002965.4). pFastBac1 and DH10Bac cells were a gift from Scott Hansen. Bacmid DNA was then transfected into Sf9 cells. Baculovirus generated from this transfection was further expanded to obtain a suitable volume and viral titer for protein expression. The resulting baculovirus was used to infect 1L HighFive cells for protein purification. Sf9 and HighFive cells were a gift from Scott Hansen. HighFive cells were harvested using centrifugation, and lysed using a dounce. The lysate was treated with EDTA-free protease inhibitor cocktail (Sigma Aldrich). We purified the S100A9 using one chromatographic step. The lysate was incubated with 1 mL of Ni-NTA agarose (Thermo Fisher) at 4 °C for 1 hour. We then washed the resin twice with 25 mL of 25 mM Tris, 100 mM NaCl, 5mM BME, 0.1% Tween pH 7.4 containing 25 mM imidazole. We eluted protein with 10 mL 25 mM Tris, 100 mM NaCl, 5mM BME, 0.1% Tween pH 7.4 containing 500 mM imidazole. We verified protein purity was >95% by SDS-PAGE. We concentrated and buffer exchanged proteins into 25 mM Tris, 100 mM NaCl, 5mM BME, 0.1% Tween pH 7.4, then flash-froze dropwise into liquid nitrogen. Proteins were stored at -80 °C until needed. We determined

protein concentration using A_{280} with an extinction coefficient of 6997 M^{-1} cm^{-1} (monomer). The protein concentrations reported in this manuscript are μM dimer.

NF-κB activity assay

We measured NF-κB activity, and normalized the data as previously described[25,41,48,60]. In brief, we co-transfected plasmids individually encoding TLR4, MD-2, CD14, and an Nf-kB luciferase reporter into HEK293T cells seeded into a 96 well plate. $S100A9^{in}$ and $S100A9^{ec}$ were buffer exchanged into endotoxin-free PBS using Pall microsep concentrator spin columns prior to treatment, and LPS was removed from $S100A9^{ec}$ using the Pierce High Capacity Endotoxin Removal Spin Columns. After stimulation with either LPS or S100A9 we then lyse cells and measure luminescence using the Promega Dual-Glo Luciferase kit. For data processing and normalization between experiments, each plate contained the following four treatments: mock (PBS), 200 ng/mL LPS, 200 ng/mL LPS with 200 μg/mL polymyxin B, and 2 μM S100A9 with 200 μg/mL polymyxin B.

Circular Dichroism:

Circular dichroism and intrinsic fluorescence measurements were collected on a J-815 CD spectrometer . All samples were dialyzed O/N at 4* prior to measurement in 25 mM Tris, 100mM NaCl, 2mM TCEP, pH7.4. Near UV (250-350nm) CD and intrinsic fluorescence were collected at 25uM S100A9 (monomer), in a 1cm cuvette. Far UV (200-250nm) CD was collected at 10uM S100A9 (monomer) in a 1mm cuvette. Samples with calcium contained 1mM CaCl2, followed by the addition of 5mM EDTA for no calcium measurements. Buffer measurements were collected both with and without calcium. Near and Far UV CD spectra were collected 3

times. Fluorescence excitation spectra was collected at emission 345nm, emission was collected at excitation 288nm. Using Jasco's software, the data was accumulated, buffer measurement was subtracted, and Savitsky-Golay filtered (level 9).

Western Blotting:

Blots were performed using either monoclonal mouse anti-S100A9 primary antibody, M13, clone 1C22 (Abnova) paired with IRDye 800CW Goat anti-Mouse IgG1 Secondary (Licor), or Polyclonal anti hS100A9 (Phospho-Thr113) Rabbit primary antibody #12782 (Signalway Antibody) paired with IRDye® 800CW Goat anti-Rabbit IgG Secondary Antibody (Licor).

Top Down Mass Spectrometry:

Top down mass spectrometry (TOF MS ES+ and FTMS) was performed by the mass spectrometry core at Oregon State University (data not shown).

Thioflavin T Fluorescence:

ThT fluorescence was measured as previously described[55]. Measurements were collected using 50uM ThT (Sigma Aldrich) and 50uM S100A9 (monomer), in a lidded Corning half area black plate after 8 hours of incubation, using an excitation/emission of 450/480. Buffer conditions: 50mM HEPES, pH7.4, 2mM TCEP, and EDTA-free protease inhibitor. Additionally, samples with calcium contained 1mM CaCl2, samples without calcium contained 1mM EDTA.

Bridge to Chapter III:

In Chapter II I characterized recombinant S100A9 purified out of insect cells ($S100A9^{in}$), and compared and contrasted $S100A9^{in}$ with the field standard $S100A9^{ec}$. $S100A9^{ec}$ and $S100A9^{in}$ share secondary structure and calcium binding, but differ strikingly in their tertiary/quaternary structure, oligomeric state, and proinflammatory TLR4 activity. This work raises many important questions about S100A9 features required for TLR4 recognition. Continuing my exploration of TLR4 complex recognition of S100A9, in Chapter III I follow up on a previous observation that the protein CD14 is an important co-receptor that enables S100A9 to activate TLR4. I utilize extensive mutagenesis, structural modeling, molecular dynamics simulations, in vitro biochemistry and ex vivo functional assays in this first attempt at a molecular characterization of the S100A9/CD14 interaction. This work brings us one step closer to unraveling the full mechanism by which S100A9 activates TLR4/MD-2.

CHAPTER THREE

S100A9 INTERACTS WITH A DYNAMIC REGION ON CD14 TO ACTIVATE TOLL-LIKE RECEPTOR 4.

*This chapter contains previously published co-authored material.

Chisholm LO, Jaeger NM, Murawsky HE, Harms MJ (2024) S100A9 interacts with a dynamic region on CD14 to activate Toll-like receptor 4. :2024.05.15.594416. Available from: https://www.biorxiv.org/content/10.1101/2024.05.15.594416v1

Author Contributions:

Lauren Chisholm and Michael Harms conceptualized the study and designed experiments. Lauren Chisholm, Natalie Jaeger, and Hannah Murawsky conducted the experiments and data analysis. Natalie Jaeger and Hannah Murawsky contributed equally to this manuscript. Michael Harms conducted the molecular dynamics simulations and analysis. Lauren Chisholm and Michael Harms created the figures, and wrote and edited the manuscript.

Introduction:

S100A9 is a small calcium-binding protein produced in copious quantities by neutrophils [16]. When released into the extracellular space, it acts as a DAMP (Damage Associated Molecular Pattern) that activates proinflammatory innate immune pathways [15] (Figure 3.1A). S100A9 plays important roles in the response to tissue damage and related processes such as angiogenesis [18,61]. When dysregulated, however, it contributes to a variety of poor health outcomes including cancer [20–22], neurodegenerative disorders [23,24], and chronic inflammatory diseases [62].

S100A9 activates inflammation through Toll-like receptor 4 (TLR4) [13,14,18,20]; however, its mechanism of action remains poorly understood (Figure 3.1B). There is some evidence that

S100A9 directly binds to TLR4 [14], and a handful of mutations to S100A9 and TLR4 are known to disrupt its activity in *in vitro* assays [48,60]. Developing a well-supported biochemical mechanism has, however, proved difficult. This lack of mechanistic understanding has had important consequences: a drug purportedly targeting S100A9 failed in phase III clinical trials due to lack of specificity [27].

Although it is unclear how S100A9 activates TLR4, the mechanism of action for TLR4's canonical ligand is well understood (Figure 3.1C). TLR4 is known for responding to lipopolysaccharide (LPS), a glycosylated phospholipid from the outer membrane of gram-negative bacteria [7,63]. It does so in concert with the adapter protein MD-2, which accommodates the acyl chains of LPS in a large hydrophobic pocket [12]. Binding of LPS promotes dimerization of TLR4, which then triggers the MyD88 inflammatory pathway via its intracellular TIR domain [7,9,10,12,64]. A third protein, CD14, plays an important supporting role by delivering LPS to the TLR4/MD-2 complex [65]. CD14 binds LPS in an N-terminal pocket, shielding the acyl chains of LPS from solvent [8,66–69]. It is anchored to the membrane by a C-terminal GPI anchor [65,70,71]. In addition to delivering LPS, CD14 promotes internalization of the LPS-bound TLR4/MD-2 complex [72,73], which triggers the TRIF-mediated inflammatory pathway [74].

TLR4/MD-2 and CD14 recognize LPS by its size and hydrophobicity, yet these features differ radically from S100A9 (Figure 3.1D). The portion of LPS recognized is ~2 kDa—about ten times smaller than the 26 kDa S100A9 dimer. Furthermore, LPS has multiple hydrophobic acyl chains and is prone to micelle formation, while S100A9 is highly soluble in water. How does a soluble protein activate a receptor tuned to recognize a small hydrophobic moiety?

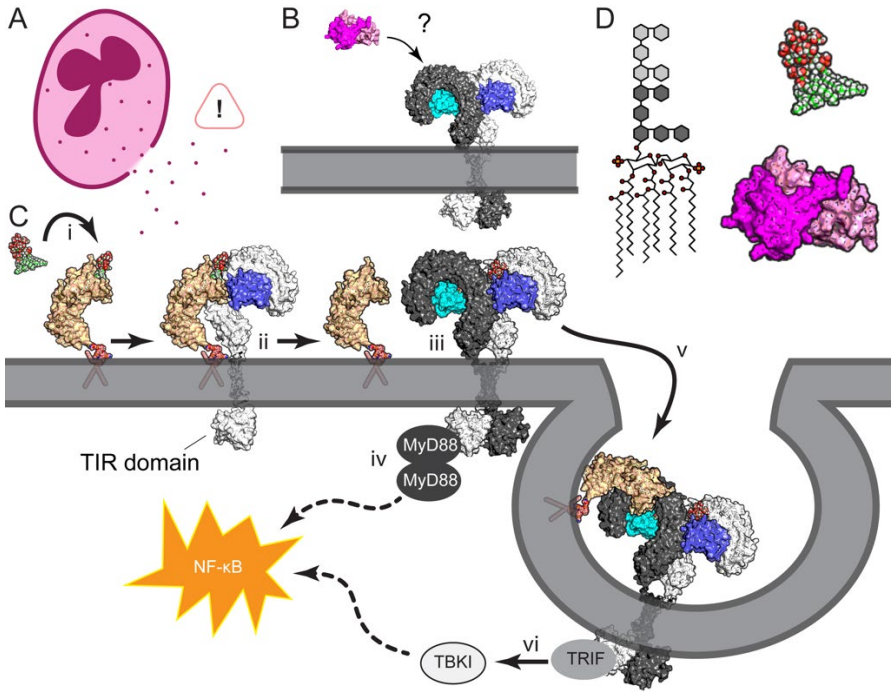


Figure 3.1: TLR4/MD2 responds to LPS and S100A9. A) S100A9 is released into the extracellular space by neutrophils, where it triggers inflammatory pathways. B) The mechanism by which the S100A9 dimer activates TLR4/MD-2 is unknown. The S100A9 structure is shown in pink/magenta: RCSB ID 1IRJ [75]. The TLR4/MD-2 structure is gray/blue: 3FXI [11]. C) In this scheme, lipopolysaccharide (LPS) is shown in spheres with green carbons; CD14 is shown in tan, with a modeled GPI linker in spheres. The CD14 structure is 4GLP [67]. The PubChem CID for the GPI linker is 145996624 [76]. TLR4 and MD-2 are shown as in B. In this mechanism, LPS binds to CD14 (i). CD14 then delivers LPS to a TLR4/MD-2 complex (ii), which homodimerizes with another TLR4/MD-2 heterodimer (iii). This brings the two TLR4 molecules' TIR domains together, allowing them to bind MyD88, which activates a multi-step pathway (dashed lines) leading to activation of NF-κB (iv). Membrane-anchored CD14 and TLR4/MD-2 can then later internalize (v), activating the TRIF-dependent pathway which further amplifies the NF-κB response (vi). D) Schematic structure of *E. coli* LPS. Hexagons are sugar moieties in the inner and outer core; lines are carbons; red circles are oxygens. Space-filling models of LPS and S100A9 are shown on the same scale to the right.

To gain a foothold to study this problem, we decided to focus on the interaction between CD14 and S100A9. In LPS signaling, CD14 is upstream of TLR4/MD-2, delivering LPS to the complex. We hypothesized that CD14 could be playing a similar role for S100A9. We reasoned that understanding the first step in the activation pathway would serve as a useful starting point for unraveling the total mechanism. In doing so, we are following up on previous observations

that S100A9 and CD14 interact directly *in vitro*, as well as co-localize and co-internalize in immune cells [26]. In line with this, we previously observed that TLR4, MD-2, and CD14 are all necessary for S100A9 to activate NF-κB signaling in a cell-based functional assay [25].

We set out to probe the role of CD14 in S100A9-mediated activation of TLR4/MD-2. We started with function: Does CD14 act as a delivery molecule for S100A9? If not, why does CD14 need to be present? We then turned to structure: What is the molecular basis for the S100A9/CD14 interaction? How does it compare to that of LPS? We tackled these questions using a combination of cell-based functional assays, site-directed mutagenesis, *in vitro* biochemical characterization, structural modeling, and molecular dynamics simulations. We found that the membrane anchor on CD14 was essential for activity with S100A9 and demonstrated that S100A9 signals more strongly through the TRIF pathway than LPS. This is consistent with CD14 promoting internalization of S100A9/TLR4 complexes, rather than behaving as a simple delivery molecule for S100A9. We also found that CD14 binds S100A9 and LPS at distinct, but overlapping, sites. The binding region of CD14 is highly dynamic, allowing it to take on different conformations and thus bind different molecules. This work is the first attempt at a molecular characterization of the S100A9/CD14 interaction, bringing us one step closer to unraveling the full mechanism by which S100A9 activates TLR4/MD-2.

Results:

CD14 improves the ability of S100A9 to activate the TLR4/MD-2 complex.

We first set out to replicate the previous finding that CD14 was necessary for S100A9 to activate TLR4 [25,26]. To do so, we used an established *in vitro* functional assay in which we transiently transfect HEK293T cells with plasmids encoding human TLR4, MD-2, CD14, and an

NF-κB luciferase reporter [25,36,48,60]. We then add potential agonists to the cell media and measure the expression of luciferase. HEK293T cells are excellent for such studies, as they natively express the required downstream effectors, but not TLR4, MD-2, or CD14 [77]. The TLR4-induced NF-κB response induced by either S100A9 or LPS is thus strictly dependent on the presence of these transfected components [25]. This allows us to study the relative contributions of each protein to the proinflammatory response with different agonists.

As a positive control for the assay, we first tested the ability of commercially available *E. coli* K-12 LPS to activate NF-κB signaling in the presence and absence of transfected CD14. As expected [65], we observed dose-dependent luciferase activity in the presence of CD14, but no activity its absence (Figure 3.2A).

We next repeated this experiment using recombinantly purified human S100A9 as an agonist. As we have done previously, we included polymyxin B (PB) in our S100A9 experiments [25,48,60]. PB sequesters LPS and prevents artifactual activation due to LPS contamination. As seen before [25,48,60], when we transfected plasmids encoding TLR4, MD-2, and CD14, we observed strong dose-dependent activation of NF-κB (Figure 3.2B). Transfecting only TLR4 and MD-2, but not CD14, dramatically lowered the response to S100A9. At 2 μM hS100A9, for example, NF-κB activity drops by 5-fold in the absence of CD14 (Figure 3.2B). (The presence of a small amount of S100A9 activity in the absence of CD14 is slightly different than one previous report [26], which found that CD14 was strictly required for S100A9 activity. This discrepancy could reflect a difference in the cell lines, as the previous experiments utilized THP1 and mouse BMDCs rather than HEK293T cells.)

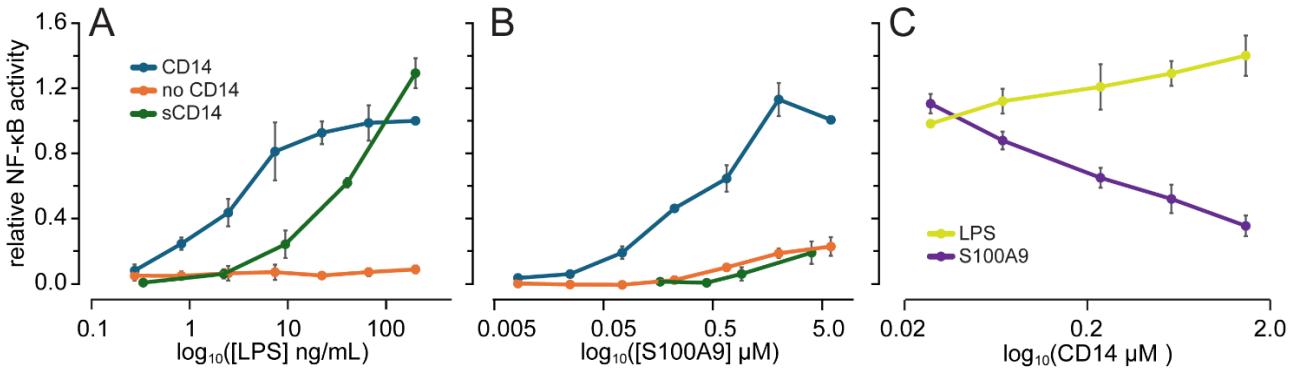


Figure 3.2: Membrane-anchored CD14 increases the response of TLR4/MD-2 to S100A9. A) NF-κB activity induced by increasing concentrations of purified LPS added to the growth media. The assay was done in HEK293T cells transiently transfected with human TLR4 and MD-2, +/- CD14. Series represent experiments done with CD14 transfected (blue), no CD14 transfected (orange), or purified soluble CD14 added to the growth media in the absence of transfected CD14 (green). Points are the means of at least three biological replicates. Error bars are standard error on the mean. B) NF-κB activity induced by increasing concentrations of purified human S100A9 to the growth media. Colors are the same as in panel A. C) Effect of adding increasing concentrations of purified soluble CD14 on the activity of cells transfected with CD14 in the presence of 20 ng/mL LPS (yellow) or 1 μM S100A9 (purple).

Soluble CD14 cannot replace membrane-bound CD14 for S100A9 signaling.

We hypothesized that CD14 is delivering S100A9 to TLR4/MD-2, analogous to its function for LPS. To test this hypothesis, we took advantage of a soluble form of CD14 (sCD14). Although CD14 is typically produced with a GPI-linker that anchors it to the cell membrane, cells also export a form without this post-translational modification [78]. sCD14 is present in the blood at ≥ 2 mg/L in healthy individuals, and is elevated during inflammation [79–81]. Unanchored sCD14 scavenges extracellular LPS from the environment and transfers it to any cells expressing the TLR4/MD-2 complex [68,78,82,83].

If CD14 is simply delivering S100A9 to TLR4/MD-2, we predicted that sCD14 would be able to replace anchored CD14 in our assay. To test this prediction, we expressed and purified sCD14 out of HEK293F cells. We then transfected HEK293T cells with plasmids encoding TLR4 and MD-2, but not CD14. In a separate tube, we pre-mixed purified sCD14 with either

LPS or S100A9. We then applied these sCD14/agonist mixtures like other treatments and measured the luciferase response.

As a control, we first tested the ability of sCD14 to deliver LPS. Consistent with previous experiments [83,84], sCD14 promoted LPS activation even in the absence of membrane-anchored (transfected) CD14 (green line, Figure 3.2A). We next tested the ability of sCD14 to deliver S100A9. Unlike LPS, sCD14 could not replace transfected CD14 for S100A9 activation (green line, Figure 3.2B). This result suggests that CD14's role in S100A9 activity is more complex than a simple delivery/drop-off mechanism.

As both forms of CD14 are present biologically, we next tested the impact of sCD14 in cells with CD14 transfected, to determine whether the addition of non-productive sCD14 blocks the productive S100A9-CD14 interaction. As one might expect, adding sCD14 to LPS treatments improves TLR4 activity, as both forms of CD14 present can deliver LPS to TLR4 (yellow line, Figure 3.2C). Conversely, sCD14 inhibits S100A9 activation of TLR4 in a dose-dependent manner (purple line, Figure 3.2C). This implies that sCD14 either sequesters S100A9 and prevents it from binding to membrane-associated CD14, or that it binds to the TLR4/MD-2 complex non-productively in the presence of S100A9.

Inhibition of TRIF dependent inflammation has a bigger effect on S100A9 than LPS-induced signaling.

Our results indicate that CD14 plays a role in the S100A9 activation of TLR4 beyond simple delivery (Figure 3.2). To help explain this, we turned to what is known about LPS-induced inflammation. In this pathway, one of the roles of membrane-associated CD14—but not soluble CD14—is to promote internalization of TLR4/MD-2 [72–74]. This activates the TRIF-

dependent proinflammatory pathway in addition to the primary MyD88-dependent pathway [74] (Figure 3.1C). Although the TRIF pathway is typically associated with Type-I Interferon productions, both pathways result in NF-κB production [85], and thus are both measured by our HEK293T functional assay.

We hypothesized that internalization is important for S100A9 activity, explaining the difference between soluble and membrane associated CD14 activity on S100A9 activity. One prediction from this hypothesis is that inhibition of TRIF-dependent signaling would inhibit S100A9 activity. To test our hypothesis, we utilized our functional assay to measure NF-κB activity in response to S100A9 or LPS in the presence of the TRIF inhibitor MRT67307. This small molecule binds to and blocks the kinases TBKI and IKKε [86,87], which operate downstream of TRIF (Figure 3.1C). We also tested the effect of the MyD88 inhibitor TJ-M2010-5 on activation by both ligands. This small molecule binds to the TIR domain of MyD88, blocking its homodimerization and thus ability to activate NF-κB [88,89].

This experiment revealed a significant difference between S100A9 and LPS activation of TLR4 in the presence of MRT67307, but not TJ-M2010-5 (Figure 3.3). Because TRIF signaling requires internalization, this implies that activation by S100A9 triggers internalization of TLR4/MD-2. Taken together with our experiments using soluble CD14 (Figure 3.2), this suggests that at least part of CD14's role in S100A9-induced activation of TLR4/MD-2 is promoting internalization of the complex. This aligns well with previous observations that S100A9 and CD14 co-internalize [26], and that a generic inhibitor of internalization, chloroquine, altered S100A9's ability to activate of TLR4 [13].

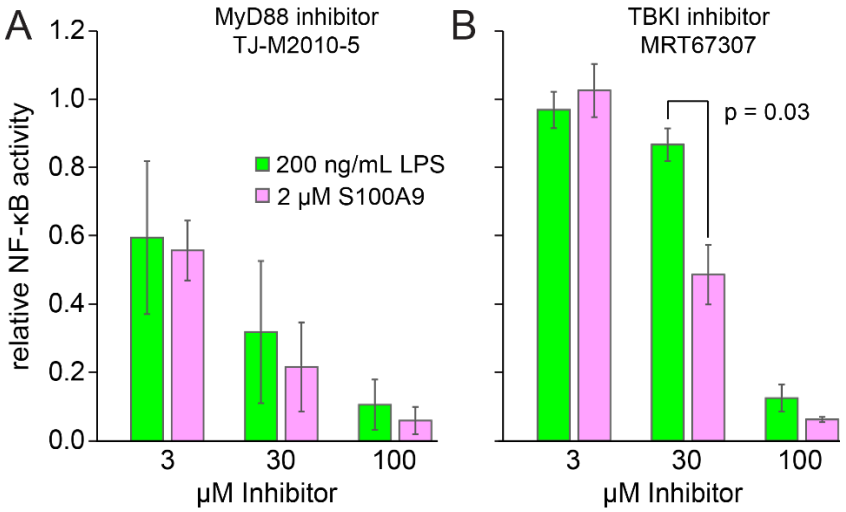


Figure 3.3. S100A9 activity may depend more on TRIF rather than MyD88 signaling. A) Effect of increasing concentrations of MyD88 inhibitor TJ-M2010-5 on NF-κB activity in response to 200 ng/mL LPS (green) or 2 μM S100A9 (pink). Bar heights are means of three biological replicates; error bars are standard errors on the mean. B) Effect of increasing concentrations of the TRIF inhibitor MRT67307 on activity induced by LPS or S100A9. Indicated p-value was calculated using a paired 2 sample students t-test. Colors are as in panel A.

LPS and S100A9 interact with different residues of CD14.

In addition to probing the function of CD14 in S100A9-dependent activation of TLR4/MD-2, we wanted to better understand the molecular basis for the interaction of S100A9 and CD14. Based on the different chemical structures of S100A9 and LPS (Figure 3.1) and the observed differences in S100A9 and LPS activation (Figures 3.2 and 3.3), we suspected that S100A9 and LPS interact differently with CD14, and that we could therefore separate their functions. To test this hypothesis, we studied the effects of an anti-CD14 antibody and mutagenesis to CD14 on LPS and S100A9 activity.

We first tested the ability of the anti-CD14 mAB MEM18 antibody to inhibit LPS and S100A9 activity in our assay. MEM18 is known to bind at amino acids 76-82 of human CD14 and block the CD14/LPS interaction (Figure 3.4A, yellow) [90]. We co-treated cells with MEM18 and either LPS or S100A9. As expected, MEM18 inhibited LPS activation (P = 0.018), but had

no effect on S100A9 activity (Figure 3.4B, 3C, yellow star). This result suggests that S100A9 interacts with a site on CD14 distinct from the LPS binding site.

We next employed site-directed mutagenesis, using the four following strategies to select mutations. 1) We introduced point mutants that had been previously reported to alter LPS activity [8]. 2) We selected amino acids near the LPS binding site based on the crystal structure of human CD14 (Figure 3.4A). This structure does not have LPS bound; however, the LPS binding pocket is known from previous biochemical experiments and docking studies [66,67]. 3) We mutagenized each of the sites in the MEM18 epitope individually and then all together. Finally, 4) We performed an alanine-scan of N-terminal residues 25-100 of CD14. This corresponds to the first 75 ordered amino acids in the protein (Figure 3.4A, purple). The first nineteen amino acids are a post-translationally removed signal peptide, while residues 20-25 are disordered in the structure. To maximize the effect size in the alanine scan, we introduced Ala mutations in blocks of three. For example, residues C25/E26/L27 were mutated to A25/A26/A27. We did not mutagenize existing alanine positions. For example, for the region A88/L89/R90, we introduced alanine at positions 89 and 90. For a list of all mutations we introduced, as well as their measured effects, see Table S3.1.

We used our functional assay to measure the NF-κB activity in response to 200 ng/mL LPS and 2 μM S100A9 for each of the CD14 mutants (Figure 3.4B). Many mutations had no effect, some moderately improved activity, and others moderately lowered activity. Only one set of mutations, converting every amino acid in the MEM18 epitope (residues 76-82) to alanine, completely disrupted activity. This did so for both LPS and S100A9, suggesting issues with CD14 expression or trafficking to the membrane.

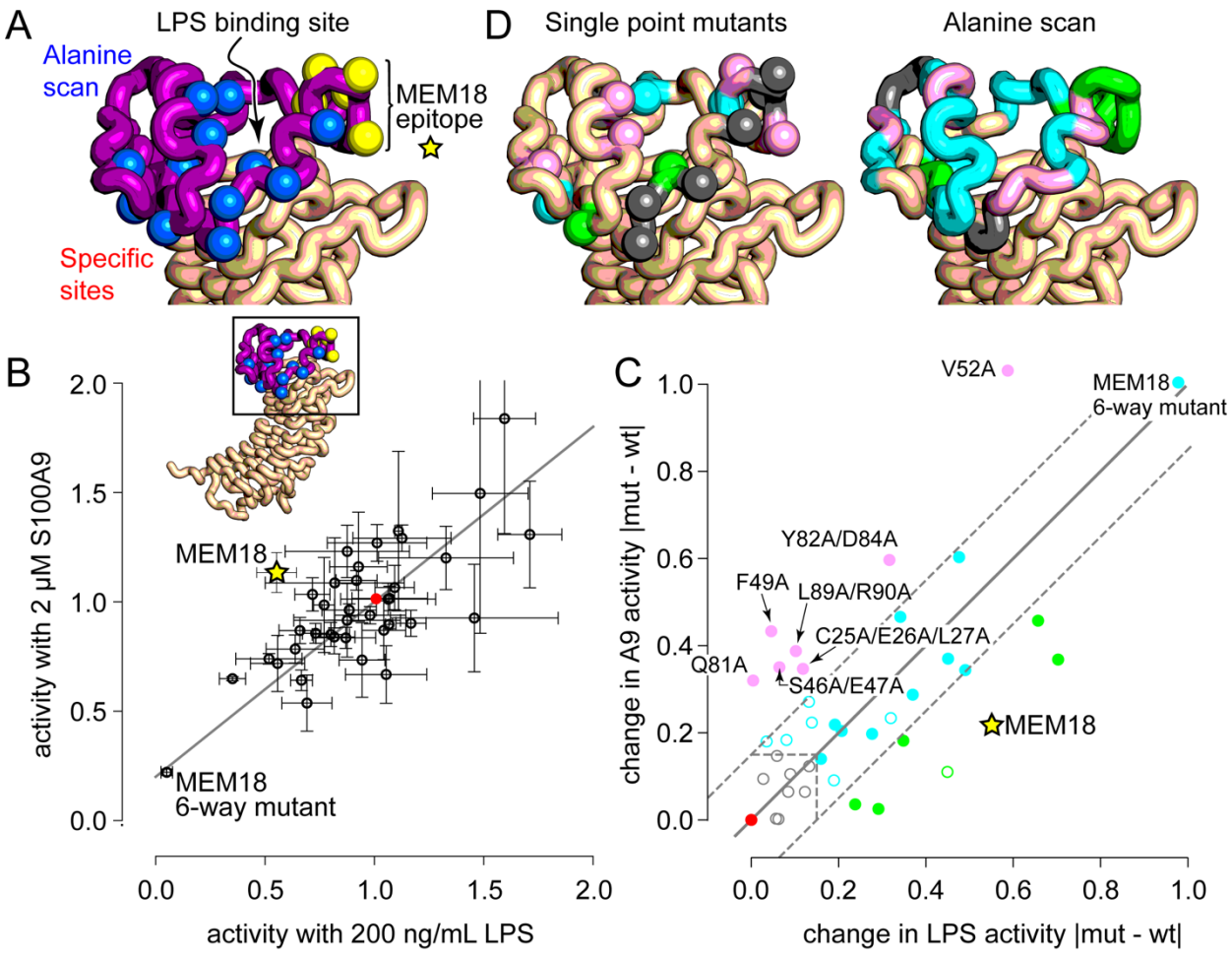


Figure 3.4: Mutations and antibody binding differentially affect LPS and S100A9 activity. A) Structure shows the region of CD14 under investigation. The purple region was probed with an alanine scan; the spheres are the C_β atoms of sites individually mutagenized; the yellow spheres are the known MEM18 epitope. B) Experimentally measured NF-κB activity in response to 200 ng/mL LPS or 2 μM S100A9 in cells with either wildtype CD14 and the MEM18 antibody (yellow star) or mutant versions of CD14 and no MEM18 (open circles). Points are the means of at least three biological replicates; error bars are standard errors on the mean. Activity is normalized to 1.0 for wildtype for both agonists (red circle). C) Magnitude of the change in LPS- and S100A9-induced NF-κB activity when antibody or CD14 mutations were introduced. This was calculated by the absolute value of the difference between the mutant and wildtype activity. Filled points have a mean difference at least two experimental standard deviations away from 0.0. Colors indicate effect type: larger effect on S100A9 (pink); larger effect on LPS (green); effect on both (blue); or no effect (gray). Wildtype CD14 is shown as a red point for reference. Mutations that had a larger effect on S100A9 are labeled. D-E) Results for single-mutants (D) and alanine scan (E) from panel C mapped onto the CD14 structure in the same orientation as in panel A. Colors match effect-type from panel C.

We were interested in mutations that had differential effects on LPS and S100A9, so we calculated the magnitude (i.e. absolute value) of the effect of each mutant on LPS- or S100A9-induced activity. We focused on magnitude rather than sign to maximize our ability to identify sites where mutations modulated activity, rather than attempting to dissect deleterious or favorable effects. The results of this analysis are shown in Figure 3.4C-E. Most mutations had either no effect (gray points) or had similar effects on both LPS- and S100A9-induced activity (blue points). Seven CD14 mutations had a larger effect on S100A9 than LPS (pink points), while five mutations had a larger effect on LPS than S100A9 (green points).

We then plotted the effects of these mutations onto the crystal structure of human CD14 (Figure 3.4D-E). Broadly, mutations that specifically altered S100A9-induced activity occurred in two regions. The first region was perturbed by the single mutations F49A and V52A, as well as the alanine scan mutations to C25A/E26A/L27A and S46A/E47A. The second region was perturbed by D81A, as well as alanine mutations to Y82A/D84A and L89A/R90A.

These S100A9-specific mutations were scattered among mutations that had different effects: altering both LPS and S100A9, changing LPS only, or having no effect at all (Figure 3.4D). This suggests that LPS and S100A9 share at least partially overlapping binding sites. We also observed some peculiar combined effects within the mutations we tested. The L89A/R90A mutation disrupted S100A9 more than LPS, while the R90A single mutant disrupted LPS more than S100A9. Likewise, D81A disrupted S100A9 and had little effect on LPS, despite being part of the MEM18 epitope that disrupts LPS but not S100A9.

A computational model predicts S100A9 interacts with CD14 via two interfaces.

To try and make sense of these results, we used AlphaFold2 [91–93] to generate a model of an S100A9 dimer docked to a CD14 monomer (Figure 3.5A). The top-ranked docking model has S100A9 bound to the N-terminus of CD14 via two interfaces, marked in cyan (Surface I) and green (Surface II) on the structure (Figure 3.5A). These surfaces correspond approximately to the two regions that our mutant screen highlighted as important. Based on this model, we hypothesized that we observed only moderate effects for our tested mutants because these mutants left one or the other interface intact. We reasoned that we would need to disrupt both interfaces to completely disrupt the ability of CD14 to interact with S100A9.

To test the two-interface docking model, we generated mutants of CD14 expected to disrupt either one or both surfaces. For Surface I, we noted that the wildtype sequence of CD14 contains bulky hydrophobic residues (W45, F49, etc.). Therefore, we chose to take an existing alanine mutant with a deleterious effect on S100A9 activation—S46A/E47A—and expand it to include the individually deleterious mutant F49A. For Surface II, we noted that the CD14 has many positively charged residues in this region. Specifically, three arginine residues (R90, R92, and R117) were in the vicinity of negatively charged amino acids on S100A9. We also had some evidence for the importance of this region, as the alanine mutant L89A/R90A improved S100A9 activity. To attempt to disrupt this interface, we introduced a negatively charged glutamate in place of arginine at all three of these positions.

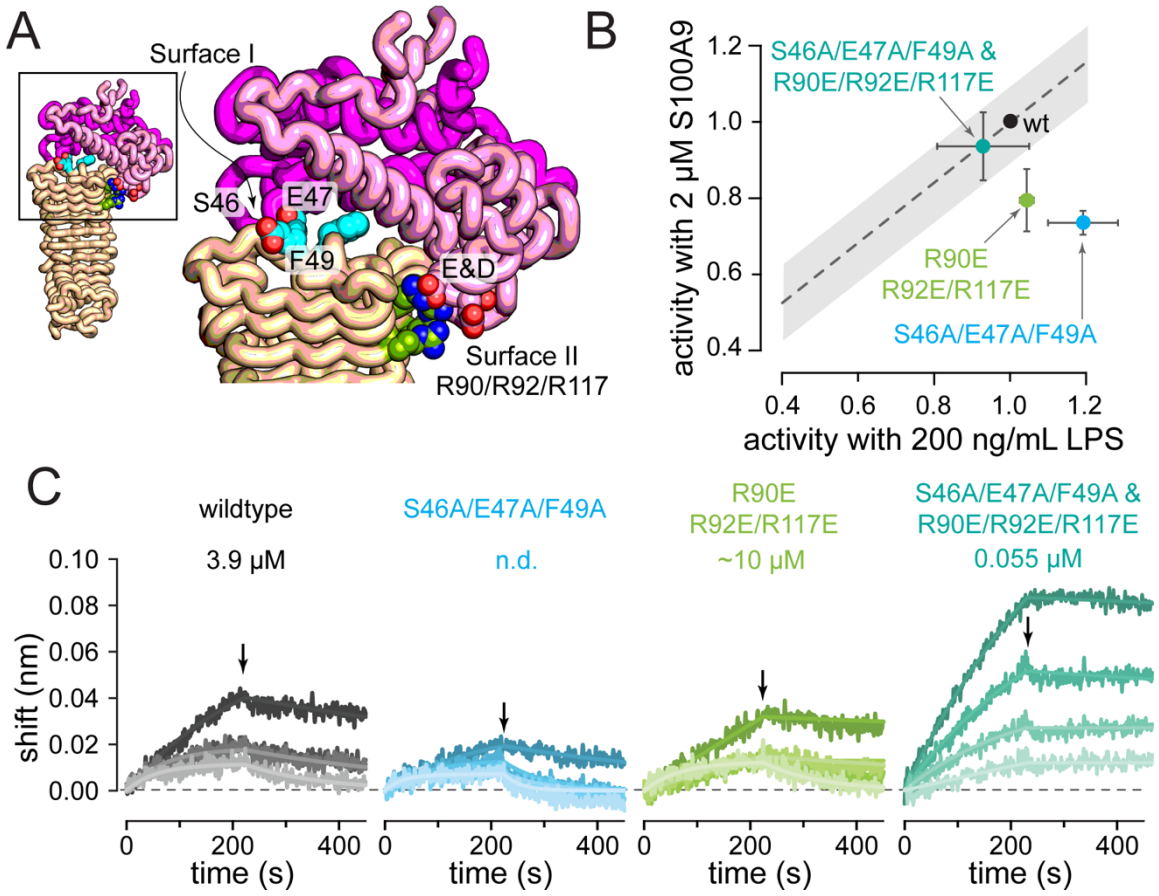


Figure 3.5. AlphaFold2 model predicts two binding interfaces on CD14. A) Structure showing the AlphaFold2 docking model. The S100A9 dimer is shown in pink (chain A) and magenta (chain B). CD14 is shown in tan. The residues we selected at the surface are shown as spheres: surface I is colored cyan, surface II is colored green. B) Experimentally measured NF-κB activity in response to 200 ng/mL LPS or 2 µM S100A9 in cells with wildtype CD14 and the indicated mutants of surface I, surface II, or surface I & II mutants. Points are the means of at least three biological replicates; error bars are standard errors on the mean. Activity is normalized to 1.0 for wildtype for both agonists (black circle). C) Bio-Layer Interferometry binding data for S100A9 interacting with sCD14. 50 nM biotinylated S100A9 was immobilized on a streptavidin sensor. Traces show response to four different concentrations of sCD14 (0.5µM, 1 µM, 2 µM, and 4 µM). Small arrow on each panel shows the point on that experiment where we switched from sCD14 to buffer. Model fits are shown as lines. Plots are representative from one biological replicate. Our best estimate of the S100A9/CD14 K_D for each variant are shown above each plot (see text).

Based on this reasoning, we generated three CD14 mutants: Surface I (S46A/E47A/F49A), Surface II (R90E/R92E/R117E), and Surface I & II (S46A/E47A/F49A & R90E/R92E/R117E). We tested the ability of these CD14 mutants to support both LPS and S100A9 activation of TLR4 (Figure 5B). As predicted, we found that the Surface I mutant lowered S100A9 activity by 30% ($P = 0.007$). It also slightly increased LPS activity ($p = 0.078$). Surface II mutant decreased S100A9 sensitivity by 20% ($P = 0.06$), and had no measurable effect on the LPS response. We next tested our combined Surface I & II mutant. To our surprise, the mutant was indistinguishable from wildtype CD14 when stimulated by either S100A9 or LPS (Figure 3.5B). The combined effect of two deleterious mutations was to have no effect at all.

One explanation of this observation is that these mutations are disrupting some feature of CD14 besides its ability to interact with S100A9. We know these proteins are being expressed and trafficked properly, as they promote activity with LPS. We wanted, however, to directly probe the hypothesis that CD14 and S100A9 interact at these sites, and that these mutations modulate binding.

To determine if these mutations altered binding, we measured the binding of wildtype CD14 and the three mutants to wildtype S100A9 using bio-layer interferometry (BLI). We biotinylated S100A9 and immobilized it on streptavidin sensors. We validated our experimental conditions with an anti-S100A9 mAb, measuring its K_D as 0.15 nM (Figure S3.1). We then measured soluble CD14 binding at concentrations between 0.5 μM and 4 μM. Despite extensive work to optimize the blocking and binding conditions, we struggled to collect high quality data for these proteins. Our traces were often barely above background. (The antibody, by contrast, gave excellent results; Figure S3.1). This is consistent with previously published surface plasmon resonance studies [26], which found that the CD14-S100A9 interaction gave low signal. Given

these difficulties, we applied a quality-control filter to all experiments, excluding any experiment for which the regressed binding model had $R^2 < 0.95$ when compared to the experimental data.

We started by measuring the interaction between S100A9 and wildtype sCD14. Of the four bio-replicates we attempted, only two yielded data that passed quality control. The mean K_D for these two replicates was 3.9 μM, with a 95% confidence interval of 0.2 to 80 μM (Figure 3.5C). This value is consistent with the previously measured K_D of 0.2 μM, accounting for different conditions and experimental methodologies [26].

We next introduced the Surface I and Surface II mutants and re-measured binding. The signal was even weaker for these proteins than for the wildtype protein: we could only reliably fit a model to one of our six biological replicates. The Surface I mutant yielded no fittable experiments (Figure 3.5D), while the Surface II mutant yielded one fittable experiment, with a K_D of 10 μM (Figure 3.5E). This apparent disruption of binding—manifesting as low BLI signal—is consistent with our functional assays, which found that both individual surface mutants lowered activity with S100A9 (Figure 3.5B). Intriguingly, when we combined the two surface mutants, we obtained much higher binding signal, with all three biological replicates yielding measurable signal (Figure 3.5E). The K_D for these three replicates was 55 nM, with a 95% confidence interval of 2 to 1000 nM.

Taken together, these mutant studies support the basic features of the docking model. Mutations at surface I and II perturb activity and binding. Importantly, the effects of the mutations track across both activity and binding: the mutants with decreased binding have lower activity, while the double mutant with increased binding recovers activity. At the same time, our simple reasoning based on charge and size complementarity was not able to predict the effect of mutations at both surfaces.

MD simulations reveal the N-terminus of CD14 is dynamic.

To better understand the molecular origins of these complicated patterns of mutational effects, we turned to atomistic molecular dynamics (MD) simulations. These simulations allowed us to relax the docking model, as well as compare the contacts made between CD14 and S100A9 with those seen for CD14 and LPS. We simulated CD14 alone, CD14 interacting with LPS, and CD14 interacting with S100A9. (See the methods for a detailed description of how we constructed our models and ran the calculations.) We ran three or four 500 ns simulations per condition, giving 1.5-2.0 μs of total simulation time for each.

We observed that the region of CD14 that binds to LPS and S100A9 was dynamic in all simulations. One of the consistent differences was the position of the "lid" formed by residues 25-55 (Figure 3.6A-C, yellow). In simulations of CD14 alone, this region collapsed into CD14's binding pocket, burying lid residues W45 and F49 (Figure 3.6A). This closed conformation shields the hydrophobic residues of the LPS binding pocket from water molecules. This conformation differs from the crystal structure of apo CD14, which has the hydrophobic LPS binding pocket open and exposed to water [67]. The conformation in the crystal structure is likely an artifact due to crystal-contacts stabilizing the open conformation (Figure S3.2).

In simulations with LPS, the aliphatic carbons of the LPS acyl chains interact with the hydrophobic binding pocket (Figure 3.6A). To accommodate the LPS molecule, the lid residues are displaced from the pocket. The lid residues noted above, W45 and F49, switch from interacting with hydrophobic residues in the binding pocket to interacting with the LPS acyl chains. This secures the LPS within the pocket and excludes water from interactions with hydrophobic regions of either LPS or the binding pocket. The CD14/LPS interaction is almost entirely hydrophobic in nature. We compared simulations of LPS and CD14 alone to the

simulations of the CD14/LPS complex, finding that LPS binding buries an average of 1360 Å^2 of nonpolar surface, but only 200 Å^2 of polar surface.

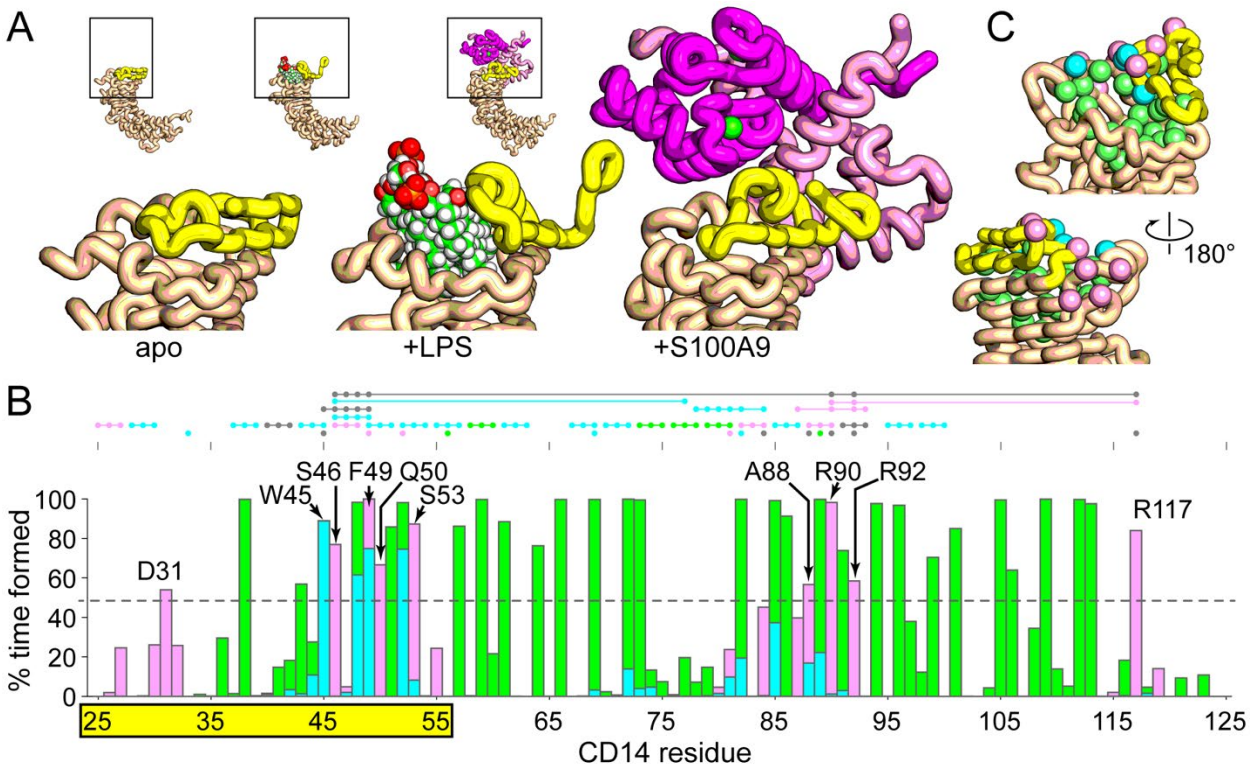


Figure 3.6. LPS and S100A9 interact with overlapping residues on CD14. A) Representative snapshots taken from simulations of CD14 with no ligand (left), LPS (middle), or S100A9 (right). The mobile lid (residues 25-55) is shown in yellow. S100A9 residues 1-3 and 91-114 are disordered and omitted for clarity. B) Bar plot shows the fraction of time each CD14 residue has at least one heavy atom within 4 Å of a heavy atom in S100A9 (pink) or LPS (green) across all simulations. Overlapping bars appear as blue. Residues participating in S100A9 interactions more than 50% of the time, but not LPS interactions, are annotated. Mobile lid residues W45 and F49 are also indicated, with the full span of the lid highlighted in yellow below the plot. Dots above the plot summarize the results from Figure 4C. Each point is a mutated residue; connected points indicate clones with multiple mutations. The color indicates the effect of that mutation from Figure 4C, where the mutation affects activation by: S100A9 alone (pink), LPS alone (green), both agonists (blue), or neither agonist (gray). C) Contacts populated more than 50% of the time in Figure 6B mapped onto the crystal structure of CD14. Color indicates the sorts of contacts made: S100A9 only (pink), LPS only (green), or both (blue).

In simulations with S100A9, CD14 interacts with similar, but distinct, residues when compared to LPS (Figure 3.6A). The CD14 lid residue F49 is in continuous contact with S100A9 over the course of the simulations, interacting with a hydrophobic patch formed by S100A9 residues M81, A84, W88, and the greasy aliphatic carbons of R85. In contrast, lid residue W45 bridges the LPS binding pocket and S100A9, interacting with both throughout the simulation. As one might expect for a protein-protein interaction, the CD14/S100A9 interface has more diverse interactions than the CD14/LPS interface. The interaction buries 880 $Å^2$ and 730 $Å^2$ of nonpolar and polar surface, respectively, reflecting both hydrophobic and hydrogen-bonding interactions across the interface.

A quantitative comparison of the CD14 residues in contact with LPS and S100A9 reveals the two agonists interact with distinct, but overlapping, sets of residues on CD14. Figure 6B shows the percent of time at least one atom from each residue of CD14 is in contact with LPS or S100A9. For simplicity, we defined any interaction formed more than 50% of the time as a contact. We found that LPS was in contact with 29 residues on CD14 (Figure 3.6B). Of these, 25 were specific to LPS, and not S100A9. As one might expect, these LPS-specific contacts are deep in the hydrophobic binding pocket of LPS (Figure 3.6C). S100A9, in contrast, only contacts 12 residues. Eight of these are specific to S100A9. These S100A9-specific contacts cluster outside the LPS binding pocket, both within the lid (D31, S46, Q50, and S53), as well as adjacent to the lid in the structure (A88, R90, R92, R117) (Figure 3.6C). These correspond closely to Surface I and Surface II from our binding experiments.

Finally, there are four residues, all in the lid region, that contact both S100A9 and LPS: W45, A48, F49, and V52. These residues take on different conformations and interactions depending on the ligand bound. When LPS is bound, they interact with acyl chains; when

S100A9 is bound, they interact with a hydrophobic patch on the protein; when nothing is bound, they interact with the hydrophobic pocket of CD14.

S100A9 binds to CD14 with competing binding modes.

Visual inspection of the simulations also revealed that the relative orientations and contacts between S100A9 and CD14 fluctuated over the course of the simulations (Figure S3.3). To explore this phenomenon further, we clustered simulation frames based on features describing the orientation and contacts of the two proteins (see methods for details). Briefly, we recorded which residues in CD14 and S100A9 had heavy atoms in contact, which residues participated in hydrogen bonds, and the rotation matrix necessary to align S100A9 from a given frame to the starting conformation. This yielded 228 features per frame. We then sampled frames every 0.5 ns and used k-means to identify clusters of conformations with shared features. This revealed two conformational clusters.

Figure 3.7A and B show representative snapshots taken from the two clusters we identified. In the one cluster—the "open" conformation—residues 25-55 and 71-89 of CD14 move apart, exposing the hydrophobic pocket used to bind LPS. S100A9 forms numerous contacts with CD14 in this conformation (black and blue spheres, Figure 3.7A). In the other cluster—the "closed" conformation—CD14 residues 25-55 and 71-89 interact with one another, sequestering the hydrophobic pocket from solvent. This conformation loses many of the S100A9 contacts seen in the open conformation (blue spheres) but gains a handful of new S100A9 contacts on the opposite side of the protein (orange arrow, Figure 3.7B).

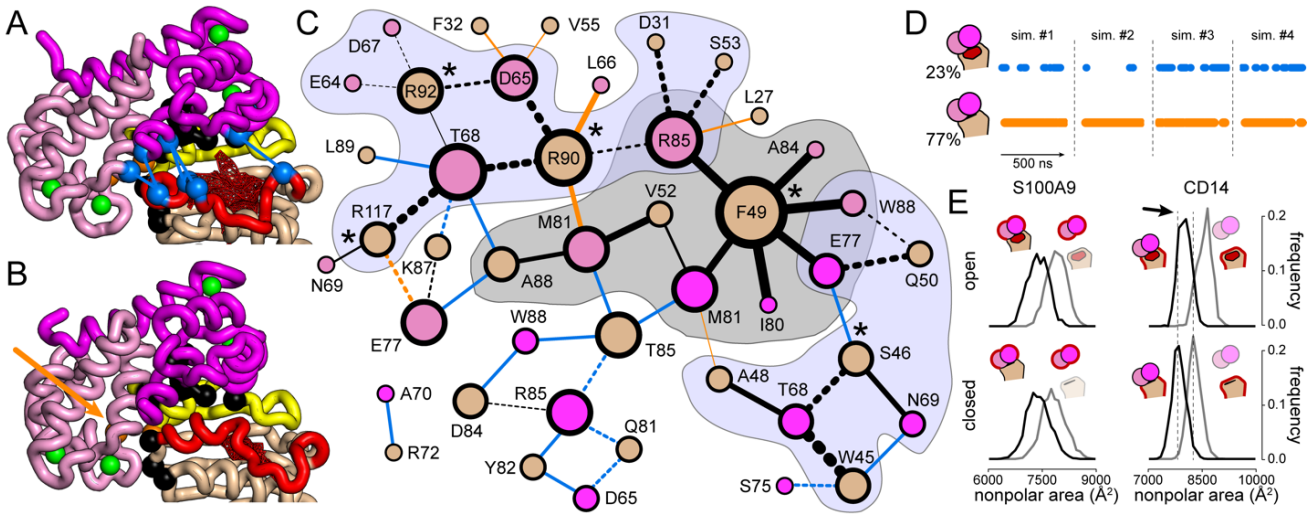


Figure 3.7. The S100A9/CD14 interface consists of competing interactions. A & B) Representative frames of the open (panel A) and closed (panel B) conformations of CD14 bound to S100A9. The S100A9 dimer is shown in pink (chain A) and magenta (chain B), with calcium ions shown as green spheres. Disordered S100A9 residues 1-3 and 91-114 are not shown for clarity. CD14 is shown in tan, with mobile regions shown in yellow (residues 25-55) and red (residues 71-89). Exposed hydrophobic surface in the binding pocket is shown as red mesh. The C_α atoms of residues participating in contacts between CD14 and S100A9 are shown in spheres whose colors indicate whether the contact is formed in both conformations (black), only open (blue), or only closed (orange). The closed-only contacts are on the opposite side of the structure, indicated with an orange arrow. C) Network of contacts formed between S100A9 and CD14 over the simulations. Nodes are amino acids in contact, with the color indicating the protein as in panels A & B. Node size indicates the number of contacts for that residue. Edges represent contacts. Edge width indicates the fraction of simulation time the interaction is formed from 10% (thinnest) to >90% (thickest). Edge color indicates whether the interaction is found in both conformations (black), specific to open (blue), or specific to closed (orange). A solid line is a non-polar contact; a dashed line is a polar contact. The gray region indicates the central hydrophobic surface between the molecules; the slate regions indicate adjacent networks of polar interactions. "*" icons indicate residues mutated in Figure 5. D) Plots show fluctuation between conformations over four replicate, 500 ns simulations. Points are frames where the proteins are in either the open (blue) or closed (orange) cluster. E) Nonpolar surface area for S100A9 (left) or CD14 (right) in simulation snapshots (black) or snapshots with the partner deleted (gray). The top row shows the results for the open conformation, the bottom row for the closed conformation. The shift in the black and gray distributions for CD14 (dashed lines) shows that the open conformation of CD14 has more exposed nonpolar surface than the open conformation.

The nature of the rearranged contacts can be seen by drawing the contacts as a network (Figure 3.7C). Notably, a central set of contacts is shared by both the open and closed conformations. CD14 residues F49, V52, and A88, form a central non-polar surface with carbons from S100A9 residues E77, I80, M81, A84, R85, and W88 (Figure 3.7C). This is surrounded by two networks of peripheral, largely polar, interactions. One network is built around CD14 residues W46, S46, and Q50, the other around CD14 residues R90, R92, and R117. These correspond to Surface I and II from Figure 3.5.

The open and closed conformations differ significantly in the number and nature of contacts on top of this shared core. The open conformation creates a whole new interface with residues 72-85 (blue spheres, Figure 3.7A; blue edges Figure 3.7C). This is adjacent to and complements the core interface formed between CD14 and S100A9. The closed conformation, by contrast forms only a few new contacts relative to the core network. These interactions tend to tie into existing core residues (for example, CD14 L27 interacting with S100A9 R85), or add new contacts between residues already in the network (for example, CD14 R90 with S100A9 M81).

Based on a naïve count of contacts, we would expect the open conformation to be favored over the closed conformation. It possesses 17 extra contacts, beyond the core network, between residues on S100A9 and CD14. When we studied the relative population of the two conformations over the course of the simulations; however, we found the opposite. Both conformations are populated in all four 500 ns simulations; however, the closed conformation is found about 3 times more often than the open conformation (77% versus 23%, Figure 3.7D). The bias towards the closed conformation is likely due to the exposure of hydrophobic surface area in the CD14 binding pocket in the open conformation (Figure 3.7A). This can be seen

quantitatively in the amount of surface area buried when S100A9 and CD14 interact in either the open or closed conformations (Figure 3.7E). S100A9 buries similar amounts of nonpolar surface when it interacts with CD14 in either the open or closed conformations. By contrast, the open form of CD14, by itself, more exposed hydrophobic surface than the closed form (Figure 3.7E).

The S100A9/CD14 interface is thus complex. Contacts between CD14 and S100A9 favor the open form of CD14, while the tendency to bury hydrophobic surface favors the closed form of CD14. As a result, the S100A9/CD14 complex fluctuates between the two conformations over the course of the simulation.

Discussion:

We set out to do an initial molecular characterization of the interaction between CD14 and S100A9. Using *in vitro* functional assays, we found that CD14 dramatically improves the activation of TLR4/MD-2 by S100A9. We also found significant differences between membrane-anchored and soluble CD14. Extensive mutagenesis and computational studies revealed that S100A9 interacts with the N-terminus of CD14, at a site distinct from the LPS interaction site. This region appears to be dynamic, allowing identical residues to participate in interactions with both lipid acyl chains and a soluble protein. This work is an important step towards understanding the mechanism by which S100A9 activates TLR4.

Proposed activation model

Our findings allow us to construct a model for how CD14 increases S100A9's ability to activate TLR4/MD-2. The key observations informing this model are: 1) Membrane-anchored CD14 is necessary for potent S100A9 activity (Figure 3.2B). 2) Soluble CD14 inhibits activation

of TLR4 by S100A9 in the presence of membrane-anchored CD14 (Figure 3.2C). 3) S100A9 signals through TRIF more than LPS does in these assays (Figure 3.3). 4) S100A9 interacts directly with CD14 (Figures 3.4 and 3.5).

These observations rule out a simple delivery mechanism in which CD14 drops off S100A9 to the TLR4/MD-2 complex (Figure 3.8A). One model that accounts for these observations has S100A9 binding directly to CD14, which then promotes internalization of TLR4/MD-2 (Figure 3.8B). This model is also consistent with existing work showing that blocking internalization differentially inhibits S100A9 activation of TLR4 [13], and that S100A9 and CD14 co-internalize [26]. In the context of this model, the simplest explanation for the inhibitory activity of sCD14 would be that sCD14 sequesters S100A9 and prevents the productive CD14-S100A9 interaction (Figure 3.8C).

We believe the model shown in Figure 3.8A-C is the simplest model that accounts for our functional observations. That said, even if this model is correct in broad outline, many questions remain. One of the most important questions is whether CD14/S100A9/TLR4/MD-2 assemble into a stable complex that is then internalized, or whether the interaction between TLR4/MD-2 and S100A9/CD14 is transient. Our data hint at the answer to this question. We found that weakening S100A9 binding is disruptive to CD14 function (Figure 3.5, mutating surfaces I and II alone), but strengthening the S100A9/CD14 interaction has no effect on activity (Figure 3.5, mutating surfaces I and II together). This result is consistent with quaternary complex formation; CD14 doesn't need to "let go" of S100A9 after the initial binding event. Formation of this complex may even promote internalization. This is, however, speculative: further work is required.

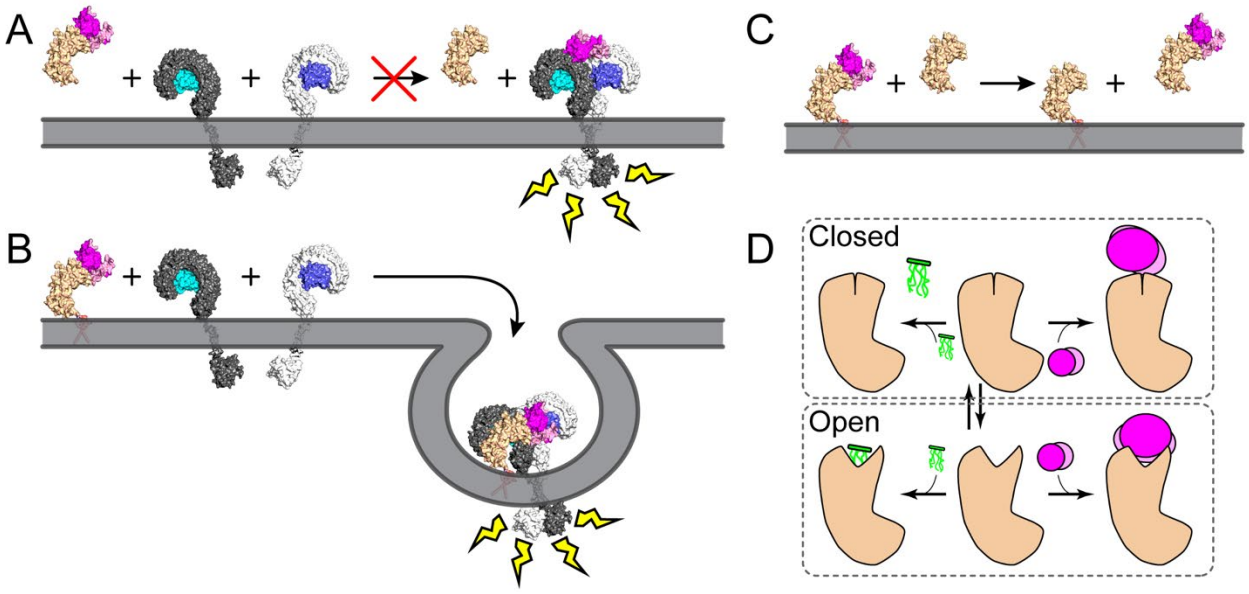


Figure 3.8. Models for the S100A9/CD14 interaction consistent with our data. A) Soluble CD14 (tan) cannot deliver S100A9 (pink) to TLR4/MD-2 (gray/blue) and trigger dimerization and activity. B) Membrane associated CD14 assembles a complex with TLR4/MD-2 and S100A9 that internalizes and triggers proinflammatory pathways. C) Soluble CD14 binds to and sequesters S100A9 away from membrane-associated CD14, thus preventing its activity. D) Schematic showing how CD14 binds to its ligands. CD14 (tan) is in equilibrium between a closed and open conformation. LPS (green molecule) binds exclusively to the open conformation; S100A9 (pink circles) can bind to either form in slightly different conformations.

Aspects of this model may also be wrong while the overall account remains correct. Our proposed mechanism by which sCD14 inhibits S100A9's ability to activate is plausible, but speculative. While sCD14 could compete for S100A9 as shown in Figure 3.8C, it could also compete with binding sites on TLR4/MD-2, inhibiting the interaction with membrane associated CD14.

Finally, it is possible that the model is mostly wrong. For example, it is possible that CD14 does not need to directly interact with TLR4/MD-2 to promote activation, but instead must be present to allow S100A9 to promote internalization. This could be achieved if S100A9 does not bind to TLR4/MD-2 at all, but instead interacts with and reorganizes the membrane. In this case, the role of CD14 could be to nucleate lipid rafts in the vicinity of TLR4/MD-2 via its GPI-

anchor. We favor a simple binding model because changes in S100A9/CD14 binding affinity correlate with changes in S100A9 activity, suggesting a role for binding (Figure 3.5). S100A9 is also known to interact directly with TLR4/MD-2, with a reported K_D of 3 nM [14], also favoring a binding model. Despite these hints, a non-binding mechanism has not been fully ruled out by our data or other reports in the literature.

Biological implications

Regardless of the precise mechanism of action, our results have three important biological implications. The first is that S100A9 can likely activate a slightly different set of downstream pathways than LPS. The MyD88 and TRIF pathways are temporally distinct and produce different inflammatory markers [85]. Indeed, other TLR4 agonists have been shown to induce differential activation of TRIF vs MyD88 dependent pathways [94,95] as a method of modulating the immune response. Because S100A9 appears to signal more strongly via TRIF than MyD88, it may activate different outputs than LPS.

The second implication is that only a subset of TLR4+ cells are likely sensitive to S100A9. This is because not all cells that express TLR4/MD-2 also express membrane-anchored CD14 [65]. Indeed, one of the roles of sCD14 is to bring LPS to cells that do not express the membrane anchored form. We predict that cells that express TLR4/MD-2, but not membrane-anchored CD14, will be sensitive to LPS but not S100A9.

The final implication is that sCD14 might have an important function in attenuating the activity of S100A9. Soluble CD14 circulates in the blood in healthy individuals at 2 mg/L [79–81]. Because it inhibits the activation of TLR4 by S100A9, it could be part of mechanism to prevent over-activation of TLR4 by S100A9, thus preventing runaway inflammation. Taken together,

these results suggest that CD14 may play an important role in regulating and tuning S100A9's ability to activate TLR4.

Interaction via a dynamic, multi-functional site.

Putting together the complete mechanism by which S100A9 activates TLR4/MD-2 will require structural insight into how the protein promotes dimerization and, likely, internalization. We made a first step towards this goal by establishing how S100A9 interacts with CD14. S100A9 binds at a site that partially overlaps, but is distinct from, the LPS binding site on CD14. This region of CD14 is dynamic, fluctuating between an open and closed conformation (Figure 3.8D). The closed conformation is favored in the absence of ligand, shielding the hydrophobic pocket. The open conformation is favored when LPS binds because it sequesters the hydrophobic acyl chains of LPS away from water. Finally, both conformations are populated when S100A9 binds. The plastic nature of this region allows CD14 to bind to both LPS and S100A9, despite their low chemical similarity.

The complicated dynamics of this region may also explain our experimental results. Mutations of both Surface I and Surface II alter both activity and binding. This strongly implicates both surfaces in the S100A9/CD14 interaction. However, their effects were epistatic and thus difficult to rationalize in terms of simple charge (e.g., Arg to Glu) or size (e.g., Phe to Ala) reversals. If S100A9 does indeed interact with CD14 via two binding modes, mutations may perturb one mode and not the other. Or, more confusingly, a mutation that disrupts one binding mode might enhance another. This means a full dissection of the binding surface will likely require a combinatorial set of mutations to fully disrupt the network of interactions predicted by the MD simulations. We may also need to characterize multiple amino acid substitutions at each

site, allowing us to probe the importance of specific physiochemical features at each site. Although we have yet to completely dissect this interface, we have identified a plausible route for the complete separation of function for LPS and S100A9.

Future steps

The long-term goal of this project is to understand how S100A9 activates TLR4/MD-2. Our work provides a stepping-stone on the way to this outcome, as we can now ask how CD14 and S100A9, together, activate the complex. This provides a more nuanced starting point than just generically considering S100A9. For example, if the CD14/S100A9 complex physically interacts with TLR4/MD-2—as seems plausible—our work provides strong constraints on the orientation of the initial encounter, as we know where S100A9 interacts with CD14 and that both CD14 and TLR4/MD-2 are membrane anchored. This also sharpens our mechanistic questions. Rather than asking how S100A9 activates TLR4/MD-2, we can instead ask how CD14/S100A9 together promote internalization. Answering such questions will be critically important for understanding the basis for this biologically important, but poorly understood, proinflammatory mechanism.

Experimental Procedures:

Heterologous S100A9 expression and purification from E. coli

We expressed and purified S100A9 as previously described [48]. Briefly, we expressed cysteine-free human S100A9 (C3S) in the pETDUET-1 vector. We transformed *E. coli* Rosetta BL21(DE3) *pLysS*. We grew 1.5 L liquid cultures to an OD_{600} ~ 0.8 and then induced with 1 mM IPTG for 16 hours at 4°C. We harvested cells by centrifugation and lysed them via sonication.

We purified S100A9 using three chromatography steps: immobilized metal ion affinity (HisTrap) at pH 7.4, anion exchange (HiTrap Q) at pH 8, followed by another anion exchange (HiTrap Q) at pH 6. We verified protein purity was >95% by SDS-PAGE. We concentrated and buffer exchanged proteins into 25 mM Tris, 100 mM NaCl, pH 7.4, then flash-froze dropwise into liquid nitrogen. Proteins were stored at -80 °C until needed. We determined protein concentration using A_{280} with an extinction coefficient of 6990 M^{-1} cm^{-1} (monomer). The protein concentrations reported in this manuscript are μM dimer.

Heterologous sCD14 expression and purification from HEK293F cells

We expressed soluble CD14 and its mutants in Freestyle HEK293F cells. We started with the full-length human CD14 gene in the pcDNA3 backbone. pcDNA3-CD14 was a gift from Doug Golenbock (Addgene plasmid # 13645; http://n2t.net/addgene:13645; RRID:Addgene_13645). We designed primers to truncate CD14 at residue 367 (truncation at this residue to produce soluble CD14 has been previously reported [84]) and add a 6× His tag.

HEK293F cells were maintained in Freestyle293 Expression Medium, shaking at 135 rpm at 37 °C in 5% CO_2. We transfected the HEK293F cells and expressed protein following the manufacturer's instructions. Twenty-four hours prior to transfection, we seeded cells at 6-7 × 10^5 cells/mL in Freestyle 293F Expression Medium. The cell viability at the time of transfection was >90% as determined by trypan blue staining. We transfected 30 mL of cells at 1 × 10^6 cells/mL, using 293fectin and 1 μg of DNA per mL of cells. After 7 days of transfection, we harvested the supernatant and incubated it with 1 mL of Ni-NTA agarose at 4 °C for 1 hour. We then washed the resin twice with 25 mL of PBS containing 25 mM imidazole. We eluted protein with 5 mL PBS containing 500 mM imidazole. We buffer exchanged and concentrated the protein into PBS

using Pall Microsep spin columns, then flash froze dropwise into liquid nitrogen. We measured the protein concentration by A_{280}, using the calculated extinction coefficient of 31105 M^{-1} cm^{-1}.

Table 3.1. Table of key reagents used in Chapter III

Supplier	**Part No.**	**Item**
VWR	BNC471093-500	Anti-S100A9 Mouse Monoclonal Antibody (CF647) [clone: 47-8D3]
Millipore Sigma	70956-4	Rosetta™(DE3)pLysS Competent Cells
Promega	E2940	Dual-Glo® Luciferase Assay System
GatorBio	#160002	Streptavidin (SA) Probes
Invivogen	inh-mrt	mrt67307
Sigma	SML0694-5MG	TJ-M2010-5
Invivogen	tlrl-eklps	LPS from Escherichia coli K-12
ThermoFisherScientific	12347019	293fectin
ThermoFisherScientific	R90101	Ni-NTA Agarose
ThermoFisherScientific	18324-012	Lipofectamine
ThermoFisherScientific	11514-015	PLUS
Bio-Rad	MCA2185	CD14 antibody | MEM-18
NEB	M0554S	KLD enzyme mix
ThermoFisherScientific	21900	EZ-Link™ BMCC-Biotin
ThermoFisherScientific	R79007	FreeStyle 239F Cells

NF-κB activity assay

We assayed NF-κB activity as previously described [25,48,60]. We maintained HEK293T cells up to 30 passages in DMEM + 10% FBS + Antibiotic-Antimycotic, at 37 °C in 5% CO_2. When performing an assay, we transiently transfected these cells with appropriate plasmids in a 96-well tissue culture plate using Lipofectamine and Plus (ThermoFisher). For the receptor complex, we used pcDNA3 plasmids that individually encoded full-length human TLR4, MD-2, and CD14 genes under control of the CMV constitutive promoter. hTLR4 was a gift from Ruslan Medzhitov (Addgene plasmid # 13086; http://n2t.net/addgene:13086; RRID:Addgene_13086).

hMD-2_pcDNA3.1+ was purchased from Genscript (LY96_OHu26610C_pcDNA3.1(+)). pcDNA3-CD14 was a gift from Doug Golenbock (Addgene plasmid # 13645 ; http://n2t.net/addgene:13645 ; RRID:Addgene_13645). To measure NF-κB activity, we used the pGL3-elam-luc plasmid, which encodes the firefly luciferase behind one NF-κB promoter, and the pRL-TK plasmid, which encodes renilla luciferase behind the TK constitutive promoter. pGL3-ELAM-luc was a gift from Doug Golenbock (Addgene plasmid # 13029; http://n2t.net/addgene:13029; RRID:Addgene_13029). pRL-TK was purchased from Promega. We transfected a total of 100 ng DNA per well, diluted in Opti-Mem. The mix included 69 ng of empty pcDNA3 vector, as well as our experimental plasmids in the following amounts: 10 ng hTLR4, 0.5 ng hMD-2, 0.25 ng hCD14, 0.25 ng Renilla luciferase, and 20 ng firefly luciferase. We combined 65 µL of transfection mix with 135 µL of DMEM + FBS in each well. After 20 hours of transfection, we removed the transfection mix and replaced it with 100 µL/well of treatment.

Treatment mixes contained 75 µL DMEM and 25 µL treatment. We used *E. coli* K12 LPS (tlrl-klps, Invivogen). S100A9 was buffer exchanged into endotoxin-free PBS prior to treatment, using Pall Microsep concentrator spin columns. All S100A9 treatments included 200 µg/mL polymyxin B. We diluted treatment components to their desired concentration in endotoxin-free PBS (without Ca^{2+} or Mg^{2+}). After three hours of treatment, we measured luciferase activity using the Dual-Glo Luciferase Assay Kit (Promega). All measurements were performed in technical triplicate. For data processing and normalization between experiments, each plate contained the following four treatments applied to the cells transfected with the complete human TLR4/MD-2/CD14 complex: mock (PBS), 200 ng/mL LPS, 200 ng/mL LPS with 200 µg/mL polymyxin B, and 2 µM S100A9 with 200 µg/mL polymyxin B.

After reading each plate, we took the mean of the technical replicates for each condition. To control for transfection efficiency, we normalized our firefly luciferase signal (FF) to our Renilla luciferase signal. For a given sample and treatment *i*, the activity was:

$$activity_i = \frac{FF_i - FF_{mock}}{Renilla_i - Renilla_{mock}}.$$

To allow comparison of activity between plates collected on different days, we additionally normalized the activity of each sample and treatment to the wildtype TLR4/MD-2/CD14 response 200 ng/mL LPS or 2 μM S100A9:

$$normalized\ activity_i = \frac{activity_i}{activity_{wt}}.$$

All plots and analysis report the normalized activity.

Docking with AlphaFold2

We used the AlphaFold2 Google Colab notebook to predict the structure of the CD14/S100A9 complex [91–93]. We input a single CD14 sequence (UniProt ID: P08571) and two S100A9 sequences (Uniprot ID: P06702), thus setting the expected stoichiometry to one CD14 interacting with an S100A9 dimer. We used the top-ranked model for all further analyses.

BioLayer Interferometry (Gator)

We purified S100A9 C3S P114C expressed in *E. coli* using our standard method, then biotinylated the protein using the EZ-link BMCC kit, following the manufacturer's instructions. We confirmed biotinylation by MALDI-TOF mass spectrometry. The binding assays were performed using the GatorPrime BLI. Our running buffer consisted of 25 mM Tris (pH 7.4), 100 mM NaCl, 2 mM $CaCl_2$, 1% BSA, and 0.1% Tween-20. We included BSA and Tween-20 to

block non-specific binding. All proteins were buffer exchanged into running buffer prior to our experiments. We soaked streptavidin sensors in running buffer for a minimum of 15 minutes prior to our experiments to dissolve sucrose coating. For each experiment, we sequentially dipped sensors into wells of a 96-well plate containing the following conditions for the indicated times: buffer for 30 s, S100A9 at 50 nM for 120 s, buffer for 30 s, sCD14 at 0.5, 1.0, 2.0, or 4.0 μM for 240s, then buffer for 240 s. We ran our experiments at 30 °C, shaking at 1000 rpm during reads. Binding fitting was performed with the GatorBio Software.

Model construction for molecular dynamics simulations

For the CD14-alone simulations, we started with the crystal structure of the human protein (RCSB ID: 4GLP, [67]). In the crystal structure, there is a deep hydrophobic pocket exposed to solvent on the N-terminus of the protein. Within 5 ns of simulation time, the helix formed by residues 26-55 closed over the pocket, expelling water molecules and burying the hydrophobic surface of the pocket (Figure S2). This can be seen in Figure 6A, which shows CD14 in the closed conformation. Residues 26-55 (yellow) act as a "lid" that fills the pocket.

For the LPS simulations, we studied the interaction of CD14 with *E. coli* Lipid A (LA). LA has the six acyl chains, three glucosamines, and two phosphates of *E. coli* LPS, but does not possess the core sugars or O-antigen. Lipid A is the conserved molecular pattern recognized by TLR4/MD-2/CD14 [96]. We did not include the core or O-antigen because they consist of long, dynamic polymers of sugar moieties that would have been computationally expensive to model. To dock LA into CD14, we oriented LA with its acyl chains facing the CD14 binding pocket but with no atom closer than 5 Å. We then ran short (20 ns) docking simulations. We started ten simulations with the crystal structure of CD14 and another ten with a CD14 structure pre-

equilibrated by 100 ns of MD simulation. We found that LA engaged with and inserted into the CD14 pocket in 9 out of 10 of the crystal structure simulations and in 3 of the 10 pre-equilibrated simulations. This difference in success rate is because the lid residues started in the open conformation in the crystal structure, but started closed in the pre-equilibrated structure. For production runs, we arbitrarily selected two of the LA docking runs that started from the crystal structure and one that started from the pre-equilibrated structure.

For the S100A9 simulations, we started with the AlphaFold structure (Figure 3.5A). We also started simulations with S100A9 and CD14 in a variety of different orientations relative to one another. This included three different models generated using RosettaDock [97], three models with S100A9 in the same orientation as in the AlphaFold structure, but with CD14 in the closed state (achieved by pre-equilibration with 100 ns of simulation), and three models with S100A9 in an arbitrary orientation relative to the CD14 crystal structure. Only simulations that started with the AlphaFold CD14/S100A9 model were stable; in all other simulations, the S100A9 and CD14 drifted apart within ~50 ns. We therefore performed our CD14/S100A9 simulations using the AlphaFold structure as our starting conformation.

Molecular dynamics simulation parameters

For all simulations, we used GROMACS 2023 [98,99] with the CHARMM36 2021 forcefield (52) and TIP3P waters [100]. We generated lipid A coordinates and forcefield parameters using LPS Modeler [101] as implemented in CHARMM-GUI [102]. We placed Ca^{2+} ions in the AlphaFold structure of S100A9 by aligning the crystal structure of Ca^{2+}-bound S100A9 (RCSB ID: 1IRJ [75]) and then manually extracting the Ca^{2+} coordinates. Using the GROMACS pdb2gmx module, we constructed a cubic periodic solvent box 20 Å longer than the maximum model

dimension, neutralized the system by randomly placing 100 mM Na^+/Cl^- counter ions, added missing hydrogen atoms, and assigned protonation states at a pH of 7.0. We prepared the system for simulations with three final steps: 1) Steepest descent energy minimization; 2) 100 ps of position-restrained equilibration in the NVT ensemble (assigning initial velocities from a Maxwell distribution at 300 K); and 3) 100 ps of equilibration in the NPT ensemble. We restrained the positions of all non-solvent heavy atoms in our position-restrained simulations. We did our production runs using an NPT ensemble at 1 atmosphere and 300 K. We used isotropic Parrinello-Rahman pressure coupling [103,104] and velocity-rescaling temperature coupling [105]. We used LINCS for bond constraints [106,107], treated non-bonded interactions with a Verlet scheme [108], and captured long-range electrostatics using a 4th-order Particle Mesh Ewald approximation [109]. We did all calculations using the talapas high-performance computing cluster at the University of Oregon

Analysis of MD trajectories

We analyzed the results using a Visual Molecular Dynamics [110], python scripts using the MDAnalysis library [111,112], and PyMOL [113]. We calculated solvent-accessible surface areas using the freesasa library [114] with a solvent radius of 1.4 Å.

For our clustering analysis (Figure 3.7), we went through our simulations, calculating contacts, hydrogen bonds, and the orientation of S100A9 relative to CD14 for ~270,000 frames. We defined two residues as in contact if they had at least one pair of non-hydrogen atoms within 4 Å. We measured hydrogen bonds using the MDAnalysis HydrogenBondAnalysis package [115], defining hydrogen bonds as an oxygen or nitrogen within 3.0 Å of a polar proton where the donor/proton/acceptor angle was at least 150°. We calculated the orientation of S100A9 relative

to CD14 in two steps. First, we aligned each frame in the simulation to the starting frame using the C_α atoms of CD14 residues 100-200. This central region of CD14 is rigid and thus provides an approximately fixed reference against which to measure the orientation of S100A9. We then calculated the rotation matrix that would minimize the root-mean squared deviation of S100A9 from the CD14-aligned frame to S100A9 from the initial conformation [116,117]. We limited this analysis to the C_α atoms from residues 4-90 from both chains, as these residues were ordered throughout the simulations. The resulting rotation matrix has nine values; we treated each as its own feature in the downstream analysis.

We then encoded contacts, hydrogen bonds, and orientation of S100A9 as features. We scored each contact and hydrogen bond as present (1) or absent (0) in each frame. The raw orientation of S100A9 in each frame was given by nine floating point numbers between -1 and 1. We rescaled these values to be between 0-1 to match the scale of the hydrogen bond and contact features. We then applied a 100-frame sliding window. This transformed our Boolean contact and hydrogen bond scores to float values between 0 and 1. We then identified the set of hydrogen bonds and contacts that accounted for the top 99% of observed contact density. We dropped any hydrogen bonds or contacts outside this set. This yielded 228 final features: 86 hydrogen bonds, 133 contacts, and 9 rotation matrix elements. Finally, we sampled these features every 100th frame—matching the size of our smoothing window—yielding ~2,700 frames, each with 228 features.

We calculated the Euclidean distance between all frames and clustered using k-means as implemented in scikit-learn [118]. We generated between 2 and 20 clusters and selected the cluster with the highest Silhouette Score [119]. This proved to be two clusters (score of 0.35). We checked for robustness to feature choice by re-doing the analysis excluding all hydrogen bonds, contacts,

or orientation scores. We also used k-fold cross validation (k=10) sampling from all 228 features. In every case, we found two clusters with similar structural features in similar proportions across the simulations.

Bridge to Chapter IV:

In Chapter III this chapter I examined the role of TLR4's co-receptor CD14 in recognition of S100A9. I showed that CD14 dramatically improves the activation of TLR4/MD-2 by S100A9 using an in-vitro functional assay. I found that CD14 must be membrane anchored to aid in S100A9 recognition, and provided evidence for the importance of internalization in S100A9 activation of TLR4. I proposed an S100A9/CD14 binding model, supported by extensive mutagenesis and computational studies. This work is an important step towards understanding the mechanism by which S100A9 activates TLR4. Continuing my exploration of TLR4 complex recognition of S100A9, in Chapter IV I describe my work in developing a high throughput method for assessing mutational effects on TLR4 ligand recognition. By incorporating an evolutionary perspective into the design of my high throughput method, I address both of my thesis questions simultaneously: 1) How does S100A9 activate TLR4? And 2) How did TLR4 evolve and maintain LPS and S100A9 recognition?

CHAPTER FOUR

A HIGH-THROUGHPUT FUNCTIONAL ASSAY FOR SCREENING TLR4 VARIANTS *IN-VITRO*.

*This chapter contains previously unpublished coauthored material.

Author Contributions:

Lauren Chisholm and Michael Harms conceptualized the study and conducted data analysis. Lauren Chisholm designed and conducted the experiments, and created the figures. Lauren Chisholm and Michael Harms wrote and edited the manuscript.

Introduction:

Many proteins are multifunctional and thus must simultaneously satisfy very different constraints. Innate immune receptors are a classic case of this, as they must respond to signals from external pathogens (MAMPs) as well as endogenous danger signals (DAMPs)[2,120]. An excellent example of this is the TLR4/MD2/CD14 complex. TLR4 is best known for recognizing the MAMP LPS (lipopolysaccharide)[3,11], however it can also be activated by a variety of DAMPs such as the host protein S100A9[14,25,38]. Activation of the TLR4 complex initiates a signaling cascade that leads to the production of NF-κB and other inflammatory cytokines[121], which then further activates the innate immune system.

LPS and S100A9 likely induce different evolutionary pressures on TLR4/MD2/CD14. Receptors that respond to MAMPs are expected to be under positive selection due to evolutionary "arms races" with pathogens[122,123]. On the other hand, receptors that respond to

DAMPs are under the constraint to maintain these protein-protein interactions, which tends to slow evolution by purifying selection. Protein complexes that respond to both MAMPs and DAMPs, such as the TLR4 complex, must resolve these competing evolutionary pressures.

In addition to different evolutionary constraints, chemically LPS and S100A9 are radically different molecules. LPS is a small molecule with long hydrophobic acyl chains that bind to MD-2[7,11], whereas S100A9 is a small calcium binding protein with an unknown binding site[75]. LPS activation of TLR4 is extremely well characterized, however it remains to be shown how exactly S100A9 (or any other DAMP) activates TLR4. Both ligands require TLR4, MD-2, and CD14 for activity[25], but it is unknown where S100A9 binds to any component of the complex (Figure 4.1). TLR4 is a very large protein, with an extra-cellular domain over 600 amino acids long. This means that there are many sites to choose from when attempting to characterize the interaction.

We set out to develop an experimental method to determine how evolution resolves the competing selective pressures on TLR4 activity, and at the same time determine the sites required for S100A9 to interact with the TLR4/MD-2/CD14 complex. More specifically, we set out to: 1) determine how evolutionary "arms races" have shaped MAMP activation of TLR4, and 2) determine which residues are involved in TLR4 complex recognition of S100A9.

Defining the Problem:

Measuring selection by substitution rate

Positive and negative selection on individual sites within a protein can be inferred by calculating the dN/dS ratio – the rate of non-synonymous substitutions (dN) over the rate of synonymous substitutions (dS). A dN/dS ratio > 1 indicates positive (diversifying) selection and

a dN/dS ratio < 1 indicates negative (purifying) selection. Methods for calculating these dN/dS ratios in a maximum likelihood framework are well described and widely used. The software packages PAML and HYPHY allow one to calculate dN/dS ratios for each column in alignment of orthologous gene sequences under a variety of evolutionary scenarios[32,33,124–127]. They report dN/dS ratio, as well as various statistics to determine the statistical significance of the results.

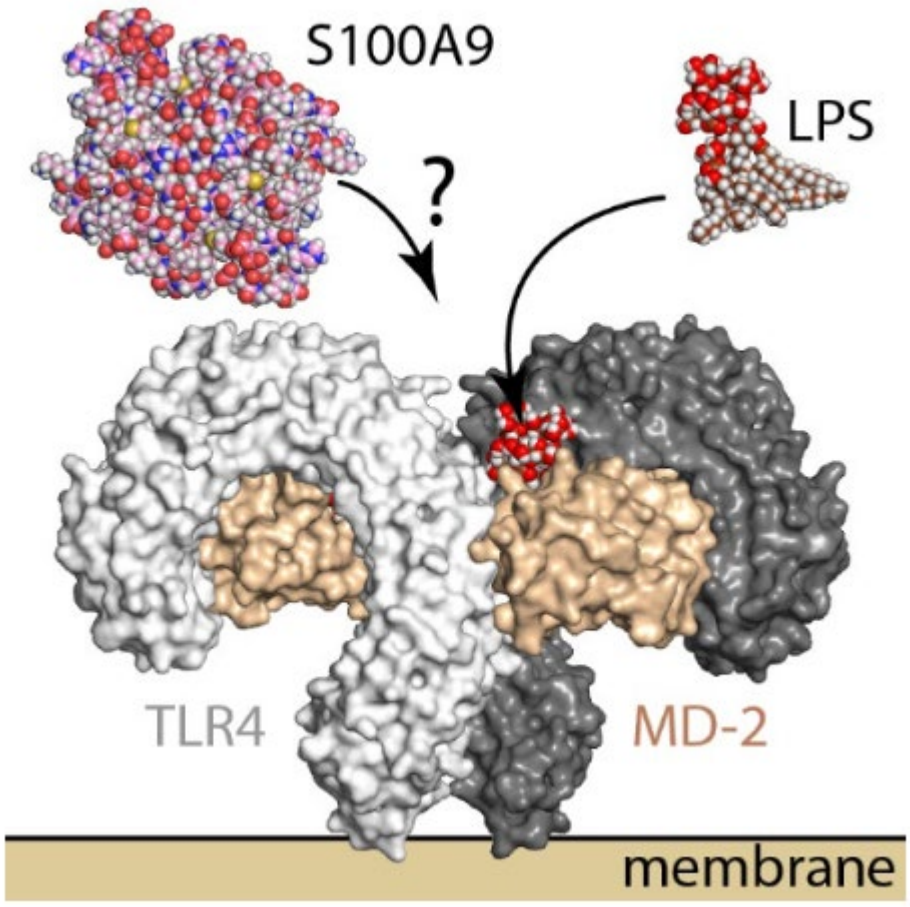


Figure 4.1: The question: how does TLR4 recognize both LPS and S100A9? Structures of S100A9 (PDB: 1IRJ[75]), LPS, and the homodimer of TLR4/MD-2 with LPS bound (PDB: 3FXI[11]) depicted to scale. Free LPS and S100A9 are shown as spheres, and colored by atom. TLR4 monomers are shown in white/grey, with MD-2 shown in tan.

Measuring Function of Variants

How can we correlate calculated selection with experimental results? There is an existing *in vitro* assay for testing TLR4 activation in response to LPS and other agonists[25,36,46,47,128]. In this assay, HEK293T cells are transiently transfected with plasmids encoding TLR4, MD-2, and CD14 under a constitutive promoter, and firefly luciferase under an NF-κB promoter. Transfected cells are then stimulated with agonist, resulting in NF-κB production. NF-κB activity is measured by lysing cells and measuring the resulting firefly luminescence. This assay enables quantitative measurement of the effect of mutations to TLR4 function. However, this assay is low-throughput and only allows for testing of a handful of mutations at once. Here we report the development of a high-throughput TLR4 functional assay, adapted from the existing low-throughput functional assay, enabling characterization of many TLR4 mutations simultaneously.

The Method:

The goal of this method is to identify sites under positive and negative selection on TLR4, and then correlate the calculated selection to mutational effects on activity using a high-throughput screen. We have already performed the described dN/dS calculations across all mammals, which revealed patches of positive (diversifying) and negative (purifying) selection. (Figure 4.2A, Table S4.1). There is a large patch of positive selection surrounding the LPS binding pocket, as has been previously reported[129]. There are also distinct separate patches of negative selection with unknown function (Figure 4.2A, Table S4.1). High-throughput screening of mutants in these regions will confirm whether these sites interact with LPS or DAMPs.

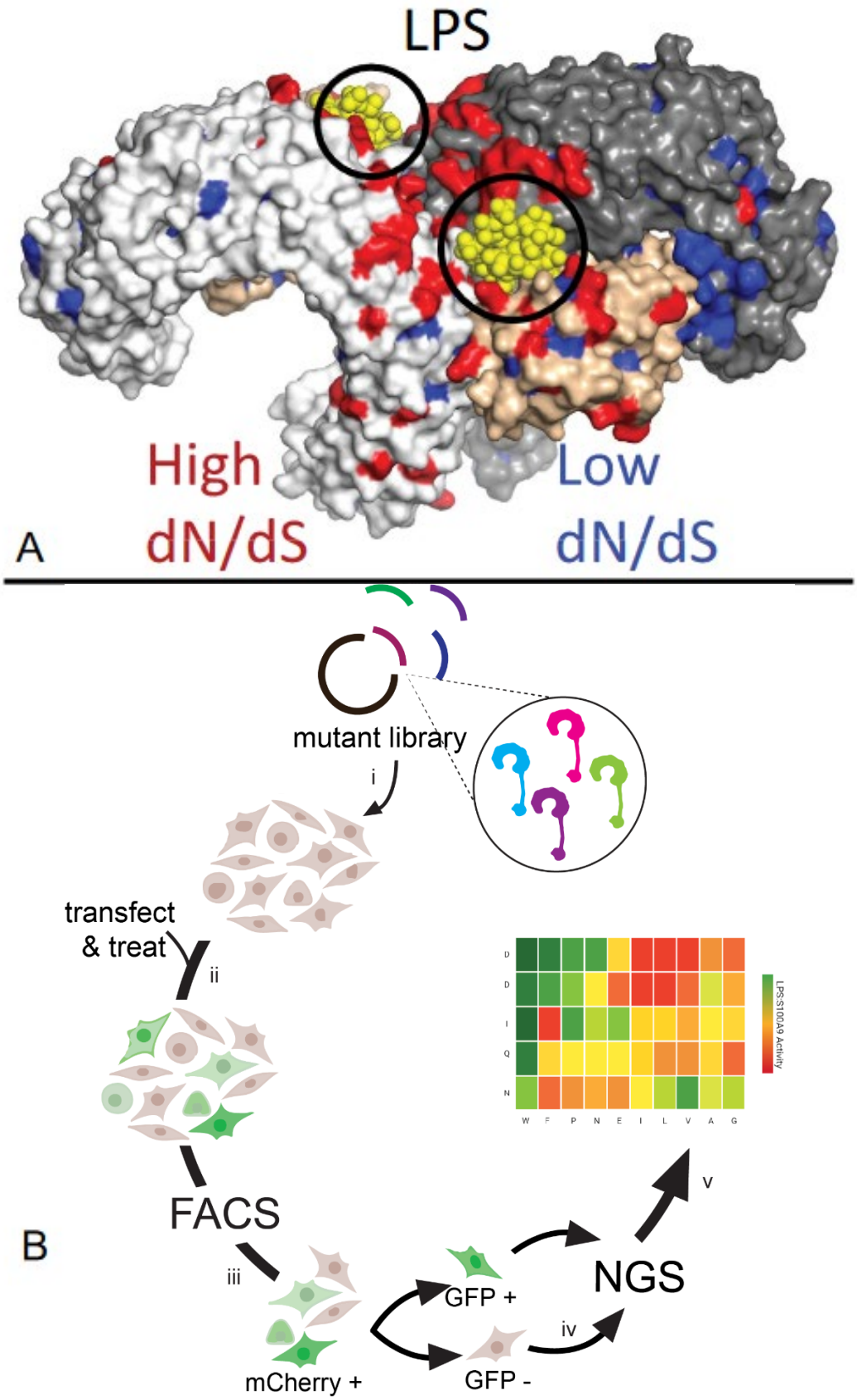
Two key changes to the existing low-throughput TLR4 functional assay enable the shift to high throughput. The first of these changes is changing from a luciferase NF-κB reporter to a

fluorescent GFP NF-κB reporter[130]. While a GFP reporter loses a great deal of sensitivity relative to a luciferase reporter[131], it enables the use of Fluorescence Activated Cell Sorting (FACS) coupled with Next-Generation Sequencing (NGS) to measure the TLR4 activity of single cells. The second change is the use of stable transfection to integrate a TLR4 library into HEK293 cells. Stable transfection results in one gene copy per cell, as opposed to hundreds or even thousands of gene copies per cell in transient transfection. One gene copy per cell ensures that each cell is expressing a single variant of TLR4, thus the genotype (TLR4 mutation) of the cell can be correlated with the phenotype (GFP expression).

In this method we generate a site saturation library at sites with elevated dN/dS ratios, where each selected site is substituted with all possible amino acids using a method adapted from the Bloom Lab[132]. We generate a plasmid library using a library of NNN primers across all single sites of desired mutagenesis, and overlap extension PCR. This method has the advantage of allowing for complete control of mutation rate, one round of overlap extension PCR = one mutation per gene, etc. As a quality check we use Sanger sequencing on a subset of the library to confirm successful mutagenesis. We stably transfect the site saturation library into FlpIn-293 cells (ThermoFisher). FlpIn-293 cells contain a FRT landing pad and a start codon, and the pcDNA5 library plasmid contains an FRT site and the Hygromycin resistance gene missing a start codon. Thus, when we co-transfect the library with a constitutively expressing Flp Recombinase, Hygromycin selection ensures successful integration. The landing pad method ensures that library variants are only inserted once per genome, thus each cell contains a unique TLR4 variant. After selection with Hygromycin and recovery, we freeze down library stocks for long term storage. To screen the TLR4 library, we seed cells into a 6 well plate and transiently transfect with co-factors MD-2 and CD14 under a constitutive promoter, alongside the GFP

reporter plasmid pSGN-Luc, which contains 8 NF-κB promoters upstream of the EGFP gene[130]. We include mCherry under a constitutive promoter is as a control for successful transient transfection. 22 hours post transfection cells we stimulate cells with either LPS, S100A9, or PBS (blank). Five hours after stimulation, we sort cells using the Sony SH800, and gate based on mCherry expression, keeping only cells expressing mCherry. Cells are then gated on GFP expression, GFP + or GFP –. Cells that express GFP contain active TLR4 variants, cells that do not express GFP contain broken TLR4 variants. We also collect an ungated control sample. We extract gDNA from cells using the NEB Monarch kit, and use this gDNA as template for PCR amplification of TLR4. We sequence the amplified TLR4 pools using NGS, and use the resulting sequence counts to calculate enrichment of each variant in each pool. We compare the enrichment of each variant in each pool to enrichment score in the unsorted pool. We then use the change in enrichment as a measure of a mutations effect on TLR4 function. A schematic of the method is depicted in Figure 4.2B.

Figure 4.2: The method. A) Calculated positive and negative selection on dimer of TLR4 (white and grey)/MD-2 (tan). Sites colored red indicate positive selection (high dN/dS), sites colored blue indicate negative selection (low dN/dS). Only sites with a dN/dS value with a confidence score of 95% or higher are colored. A full list of calculated selection rates is available in Table S5.1. B) Schematic of the method. i) A TLR4 variant library is stably integrated into FlpIn293 cells; ii) TLR4 library cells are transiently transfected with TLR4 cofactors, mCherry, and a GFP reporter, followed by treatment with TLR4 agonist; iii) cells are sorted based on mCherry expression (transient transfection control), then GFP expression (TLR4 activity); iv) sorted pools are used as input for next generation sequencing; v) NGS results are used to calculate variant enrichment in sorted pools.


LPS
High
dN/dS
Low
dN/dS
A
mutant library
i
transfect
& treat
ii
FACS
iii
mCherry +
GFP +
GFP -
iv
NGS
v
B
LPS:S100A9 Activity

Method Validation:

The method we describe here was validated through a pilot screen, sampling 16 amino acids of TLR4. Using the previously calculated dN/dS results (Figure 4.2A), as well as existing mutagenesis data, we designed the pilot screen to both assess the screening method and generate useful data. The 16 amino acid residues of TLR4 selected (Table S4.2), have either been tested experimentally—and thus have a known effect to TLR4 activity—or have inferred positive selection.

The 16-site saturation mutagenesis library containing only single mutants has a theoretical diversity of 320 variants, but due to the presence of double mutants likely has higher diversity. The entire input library was sequenced to assess true library diversity alongside sorted pools. Preliminary Sanger sequencing revealed primarily single mutants, as well as several double mutants. Stable transfection of FlpIn293 cells resulted in ~3,000 + integration events, or ~10X library coverage, as measured by colony count post Hygromycin selection. The library was screened against LPS and S100A9 in duplicate. Flow cytometry of the treated library showed that ~70% of cells were not successfully transiently transfected, and 1/3 of the successfully transfected library contained active variants (Figure 4.3B).

NGS of the input library after screening unveiled a problem with the method that must be addressed in future iterations. The diversity of the input library was heavily weighted towards wildtype (Figure 4.3A), with actual TLR4 mutants representing <50% of the sequence counts. This hindered our ability to draw strong conclusions from our experiments. This issue is also easily fixable in future iterations: changing the plasmid library generation method, and better quality control prior to library integration. Rather than using overlap extension PCR to generate

the library, we will utilize oligo assembly. Oligo assembly provides the same control over mutation type and frequency but enables easier removal of wildtype sequences. None of the oligos used in the assembly contain a full WT sequence, but the primers used in overlap extension can combine to form wild type. This change in library generation method, combined with NGS prior to integration instead of Sanger sequencing, will result in a library with the desired input diversity, in turn enabling a more effective high-throughput screen.

The results of the pilot screen revealed promising evidence for the function of the method. We assessed the success of the method in two ways. First, we plotted the change in enrichment for each genotype for the GFP and no GFP pools after treatment with LPS (Figure 4.3C). Changes in enrichment falling along a 1:1 line indicate TLR4 variants that changed in the same way regardless of phenotype. There is a clear population falling above the 1:1 line, indicating variants that are beneficial to TLR4 activity. Second, we compared the enrichment of the whole library compared to the enrichment of variants with early stop codons within the GFP pool. Early stop codons result in an incomplete, inactive protein and are expected to be de-enriched relative to the rest of the library. We saw the predicted effect; early stop codons are de-enriched relative to the whole library (Figure 4.3D). Taken together, these results provide promising evidence that our high-throughput method is functioning as expected for LPS. In contrast, library treated with S100A9 showed much weaker change in enrichment, with no clear population of variants off the 1:1 line (Figure 4.3E). Similarly, within the GFP pool the enrichment of early stops matched very closely with the rest of the library (Figure 4.3F). The distribution of all variants in the GFP pool for S100A9 treated cells was strikingly different from cells treated with LPS: the library treated with S100A9 was centered around zero, whereas the library treated with LPS had a much broader distribution of enrichment values (Figure 4.3D/F).

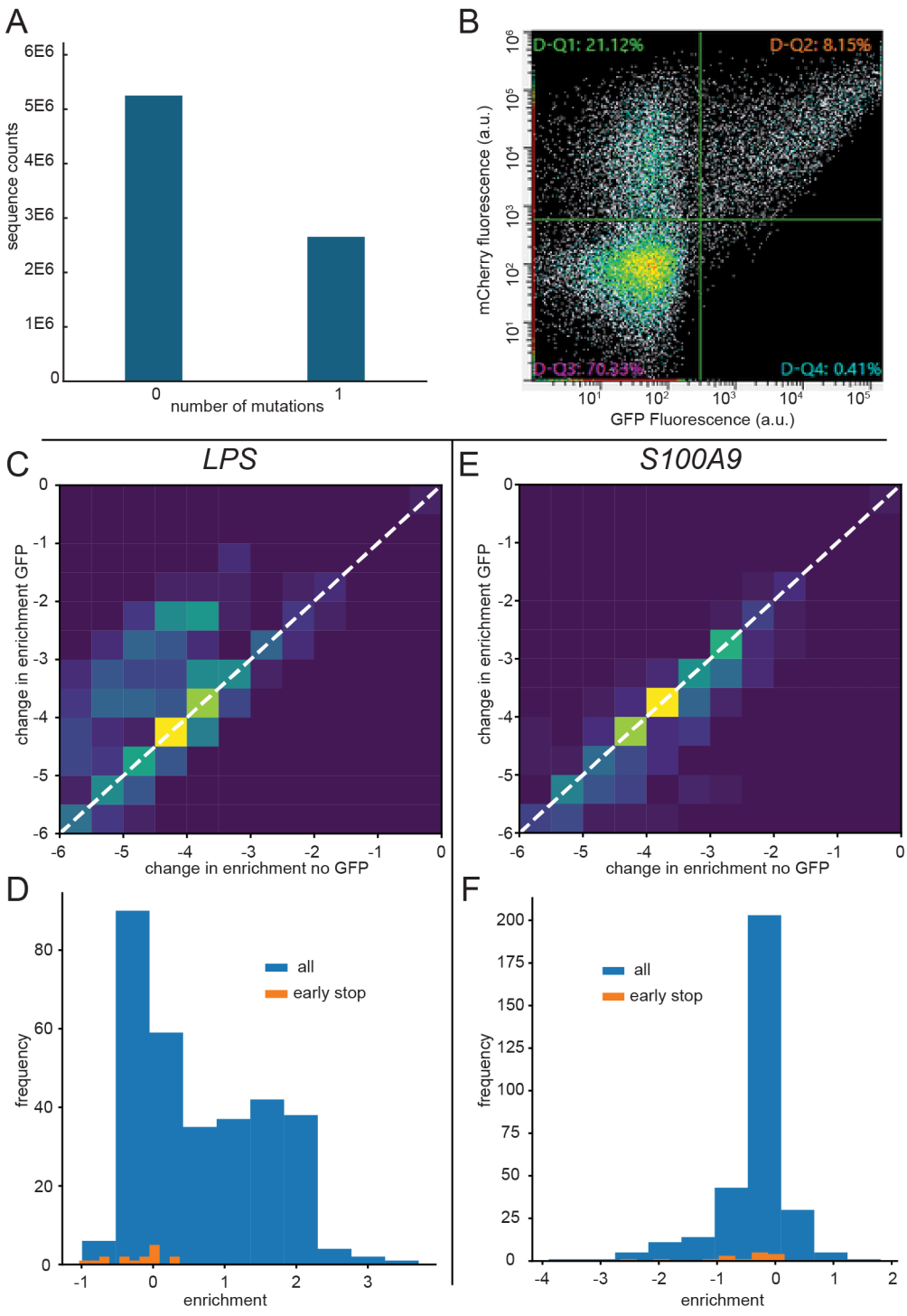


Figure 4.3: Pilot screen results. A) counts of mutations per gene of unsorted library. B) Cytometer dot plot with quadrant gates of mCherry vs GFP expression in library cells treated with LPS. Each dot is a single event. C) and E) Change in enrichment of GFP (active variants) vs no GFP (inactive variants). Sectors are colored by how many sequences fall within the quadrant. A hypothetical 1:1 line is plotted in a dashed white line. D) and F) enrichment frequency of the entire library (blue) vs early stops (orange).

Discussion:

The results of the pilot screen are mixed. When the library was treated with LPS, we observed strong changes in enrichment between the GFP and no GFP pools. However, S100A9 treatment conditions showed no such changes. The observed discrepancy between the screen results for S100A9 and LPS could be due to the lack of sensitivity of the GFP reporter. It could also be explained by the sites selected for the screen – all sites selected were chosen from within the region of TLR4 known to be important for LPS recognition, and were under positive selection. The lack of library diversity seen in Figure 4.3A may further explain the relatively weak enrichment of active clones in the screen results. Future iterations of the method with more diverse input libraries will probe the results of the pilot screen further, and allow stronger conclusions about mutational effects to be made.

We report here for the first time a method for measuring the function of TLR4 variants in high throughput. Our reported method will enable an experimental examination of the effects of competing evolutionary pressures within a single multifunctional protein for the first time. In addition to answering fundamental evolutionary questions, this method will also contribute to our mechanistic understanding of DAMP activity. DAMPs such as S100A9 are a very important class of molecule, with many effects on human health. However, very little, if anything, is known about the activation mechanisms of these proteins. This high-throughput method is a powerful tool that can be used to study mechanistic questions, by testing other TLR4 agonists or other TLRs, and evolutionary questions, for example testing ancestral TLRs.

Bridge to Chapter V:

In chapter IV I reported my progress on the development of a high throughput TLR4 activity screen. I described the method in detail, and reported the results of a small pilot screen of 16 amino acids of TLR4. Though there are still improvements to be made, I have created the first high throughput TLR4 activity assay, providing a powerful tool to study TLR4 ligand recognition from a mechanistic and evolutionary perspective. Continuing my exploration of protein evolution, in Chapter V I review another commonly used method to study protein evolution: ancestral sequence reconstruction (ASR). In ASR, one phylogenetically infers the sequences of ancient proteins, allowing characterization of their properties. Indeed, ASR has been used to study the evolution of TLR4. Chapter V reviews the use of ASR in studying the evolution of protein energy landscapes.

CHAPTER FIVE

ANCESTRAL RECONSTRUCTION AND THE EVOLUTION OF PROTEIN ENERGY LANDSCAPES.

*This chapter contains previously published coauthored material.

Chisholm LO, Orlandi KN, Phillips SR, Shavlik MJ, Harms MJ (2024) Ancestral Reconstruction and the Evolution of Protein Energy Landscapes. Annual Review of Biophysics 53:127–146. Available From: https://www.annualreviews.org/content/journals/10.1146/annurev-biophys-030722-125440

Author Contributions:

Michael Harms, Lauren Chisholm, Kona Orlandi, Sophia Phillips, and Michael Shavlik conceptualized the study. Each author contributed writing and editing to the manuscript and text was finalized by Michael Harms. Kona Orlandi, Sophia Phillips, and Michael Shavlik contributed equally to this manuscript. Figures were created and edited by each author and finalized by Michael Harms. Lauren Chisholm and Michael Harms administered the tasks for the project.

Introduction:

The sequence of a protein encodes a conformational energy landscape [133,134]. Some conformations are favored, while others are less so. The ensemble of conformations determines the function of the protein, with many functions depending on a protein fluctuating between multiple conformations [135–137]. This implies that understanding the evolution of protein function requires understanding how mutations alter the protein energy landscape [138,139].

The importance of an energy landscape view for understanding protein evolution can be seen in a simple engineered evolutionary trajectory. Researchers sequentially introduced mutations converting a primarily β-sheet protein into a primarily α-helical protein [140–142]. At the beginning of the trajectory, the α-helical state had a high energy and was thus unpopulated. The mutations then stabilized the α-helical conformation and destabilized the β-sheet conformation until, ultimately, the ground state of the protein switched from β-sheet to α-helical (Figure 5.1). A landscape view is needed to make sense of this change to the ground state, as the transition requires describing changes to both the β-sheet and α-helical conformations, as well as their relative energies at each evolutionary step. Like this engineered trajectory, natural protein evolution proceeds through mutations that tune the energy landscape [138,139,143–150].

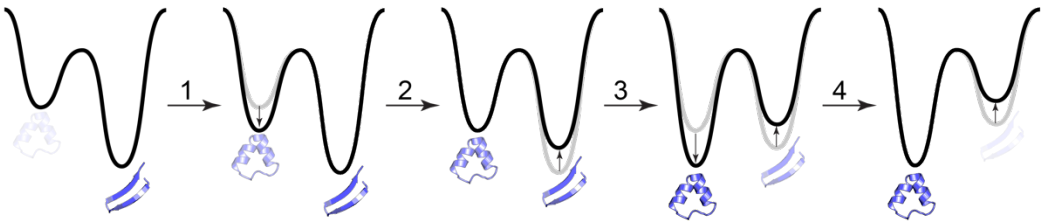


Figure 5.1. Protein evolution and energy landscapes. A schematic of a four-mutation evolutionary trajectory that switches a small protein from a β-sheet to α-helical conformation. The black lines show the energy landscape for each genotype; the opacity of each conformation shows its population at equilibrium; and the effects of mutations on the energy landscape are denoted with arrows. Mutations 1 and 3 stabilize the helical fold; mutations 2, 3, and 4 destabilize the sheet fold. Together, they switch the fold from β-sheet to α-helical.

Ancestral sequence reconstruction (ASR) is a powerful tool for revealing how energy landscapes evolve. In ASR, one uses the sequences of modern proteins and phylogenetic models to reconstruct the sequences of ancient proteins, which can then be characterized experimentally. ASR studies reveal when historical amino acid substitutions occurred and how they correlate with the acquisition of new protein features. This is an efficient means to identify residues

important for a given function [151,152]. Furthermore, ASR places protein features into a natural hierarchy: those that evolved first both set up and constrain those that evolved later [152].

ASR also provides information about how the energy landscapes of natural proteins evolve that is not readily accessible using other evolutionary analyses. For example, both ASR and co-evolutionary methods can be used to extract biophysical information from sequence alignments. ASR focuses on specific substitutions in historical proteins, allowing researchers to mechanistically dissect changes to the energy landscapes of specific family members. By contrast, co-evolutionary methods identify sites whose identities co-vary across massive sequence alignments [153,154]. This averages out sequence changes from individual evolutionary lineages, revealing architectural constraints shared by an entire protein family at the expense of detail about each family member. Likewise, experimental methods such as saturation mutagenesis and directed evolution are powerful for studies of how mutations alter energy landscapes [155,156]. However, these methods deal in artificial evolutionary trajectories occurring under laboratory conditions. ASR, by contrast, provides a window into the evolution of naturally occurring proteins, which evolve in an integrated biological context and are subject to diverse evolutionary processes. Thus, ASR provides specific evolutionary information that complements other bioinformatic and experimental methods.

In this review, we discuss how ASR methods have been used to uncover the evolution of energy landscapes in naturally occurring proteins. We briefly review the methods and logic of ASR, then discuss recent findings illustrating the evolution of energy landscapes and how this has shaped evolutionary trajectories. We describe how applying this lens to ASR studies helps us think about deep trends in protein evolution. Finally, we conclude with some current work that will help improve ASR as a tool for understanding the evolution of protein energy landscapes.

How Does Ancestral Sequence Reconstruction Work?

Before diving into how ASR has been used to dissect the evolution of energy landscapes, we provide a brief overview of the methods and statistics used in ASR. For more details, we recommend several recent reviews [157–161], as well as descriptions of software packages that explain how the calculations are done [162–164].

ASR was first proposed by Linus Pauling and Emile Zuckerkandl in 1963 [165]. They realized that analyzing the sequences of modern proteins in the context of an evolutionary tree could, in principle, allow them to reconstruct the sequences of ancient proteins. The advent of new statistical and computational approaches [166,167], massive databases of protein sequences, and cheap gene synthesis has led to an explosion in ASR studies to reconstruct the evolution of diverse protein features. These features include: enzymatic activity [146,168–174], thermodynamic stability [175–177], folding pathways [178,179], regulation [180], oligomeric state [181,182], binding specificity [183–186], fluorescent photoconversion [187,188], absorption wavelength [189–192], and many other functions [193–195].

ASR requires four steps. First, one defines a protein of interest and collects homologous sequences from diverse organisms (Figure 5.2A). Second, one constructs a multiple sequence alignment (MSA) from these sequences (Figure 5.2B). This alignment defines which sites are homologous—that is, arose by descent—in those sequences. Third, one uses the information from the MSA to construct a phylogenetic tree describing the evolutionary relationships between the sequences (Figure 5.2C). Fourth, and finally, one infers the sequences of ancestors by extrapolating backwards along the inferred tree (Figure 5.2C).

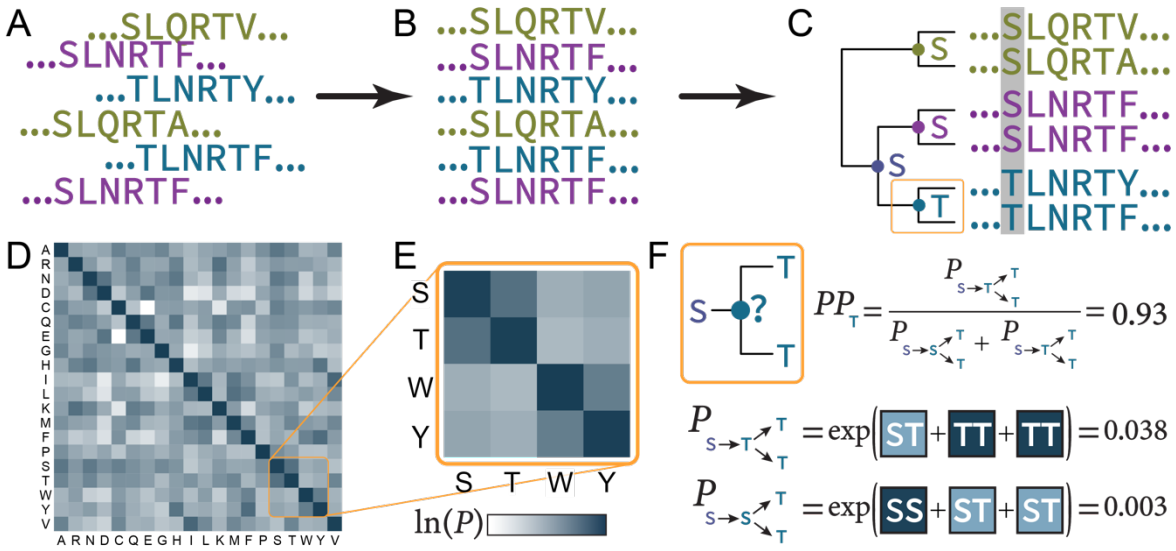


Figure 5.2: How Ancestral Sequence Reconstruction works. A) Download a diverse collection of modern sequences from online databases. B) Create a multiple sequence alignment (MSA). C) Infer a phylogenetic tree modeling the evolutionary history of the sequences in the MSA. The amino acids noted on the tree correspond to the site highlighted in gray in the MSA. D) The commonly used LG amino acid substitution matrix. Colors denote the log of the exchangeability of the amino acids denoted along the x- and y-axes. E) Subset of the LG matrix showing the relative substitution probabilities of a handful of chemically similar and dissimilar amino acids. F) Calculation of the posterior probability that the amino acid at ancestor "?" was most likely Thr (PP_T). We calculate the probability of two chains of evolutionary events: one with ancestral Thr (S→T, T→T, T→T), and the other with ancestral Ser (S→S, S→T, S→T). The chain with Thr is 13 times more likely than the chain with Ser (0.038 versus 0.003), giving a posterior probability that the ancestor was Thr of 0.93. (It is important to note that, in this toy example, we assumed that all branch lengths were identical).

To infer the phylogenetic tree and ancestors, most ASR studies use statistical models built on several assumptions: 1) sequences in the MSA arose by a strict bifurcation/branching process; 2) each site evolves independently; and 3) the relative probability of amino acid substitutions at each site has been the same over time. At the heart of such models is a substitution matrix encoding the probability of different evolutionary transitions (Figure 5.2D). These matrices are inferred from known protein sequences and thus tend to reproduce one's physiochemical intuitions. For example, in the commonly used LG matrix [196], the probability of a polar-to-polar Thr to Ser substitution is 33-times greater than a polar-to-aromatic Thr to Tyr

substitution (Figure 5.2E). Phylogenetics software finds the phylogenetic tree that maximizes the probability of observing the MSA given the substitution model.

Ancestors are then inferred using the phylogenetic tree. For most studies, ancestors are reconstructed using the marginal probability method [167]. For each site, at each ancestral node, ASR software determines the relative probability of all ancestral scenarios given the tree and MSA. This is shown schematically in Figure 5.2F for an ancestral site that could plausibly be Ser or Thr. In this case, Thr is favored over Ser because the evolutionary moves required to produce Thr (S→T, T→T, T→T) have a higher likelihood than those required for Ser (S→S, S→T, S→T). This difference is quantified with a posterior probability: the likelihood of the most likely set of events over the total likelihood of all events. A higher posterior probability indicates stronger support for the reconstructed amino acid at that site.

This core modeling approach is used in almost every ASR study. Additional terms may be added to better model specific evolutionary processes. These include modeling variable evolutionary rates across sites [197], using different substitution matrices for different sites in the alignment [198,199], and incorporating information from the species tree when inferring the gene tree [200]. Another key choice is whether to use a maximum likelihood approach, which infers the most plausible tree and ancestors, or a Bayesian approach, which infers a distribution of plausible trees and ancestors [201]. Given the experimental difficulties of characterizing a representative ensemble of ancestors and evolutionary trees, most ASR studies rely on a maximum likelihood approach.

The Logic of an Ancestral Sequence Reconstruction Study

The basic task in most ASR studies is to learn how a set of historical substitutions conferred a new function to an ancestral protein. A study to understand how protein feature *X* evolved would typically involve several steps [151,202]. First, find the most recent ancestor that did not have *X* (ancPreX) and the oldest ancestor that did have *X* (ancPostX). *X* evolved somewhere along the evolutionary branch between these two ancestors. Then, using experiments and/or computational analyses, identify the set of mutations that conferred *X*. This is usually a subset of the total sequence differences between ancPostX and ancPreX. Finally, dissect the mechanism by which the historical mutations conferred *X*, usually by studying their effect in the historical ancPreX genetic background.

We can see how this works in practice for the evolution of protein heterocomplexes [203]. ASR has been used to trace the evolution of several multi-component protein complexes including the V0 ring of V-ATPase [204], hemoglobin [205], and other proteins [203,206]. Figure 5.3 shows a schematic abstraction of these results for a hexameric complex assembled from four proteins (Figure 5.3, bottom right). The deepest reconstructed ancestor forms a homomeric hexamer (Figure 5.3, bottom left). By following the ways in which the assembly changes over subsequently more recent ancestors, one can identify the key substitutions and events that led to a highly specific modern complex. In this example, the first event was a duplication of the most ancient, homo-hexameric ancestor. Immediately after duplication, the subunits were interchangeable. This was followed by a mutation that created a hole in one subunit without altering assembly: The subunits remained interchangeable. A second mutation created a knob that can only be accommodated by being adjacent to a subunit with a hole, thus conferring a

specific assembly order. This process then repeated along further evolutionary branches, leading to the modern complex.

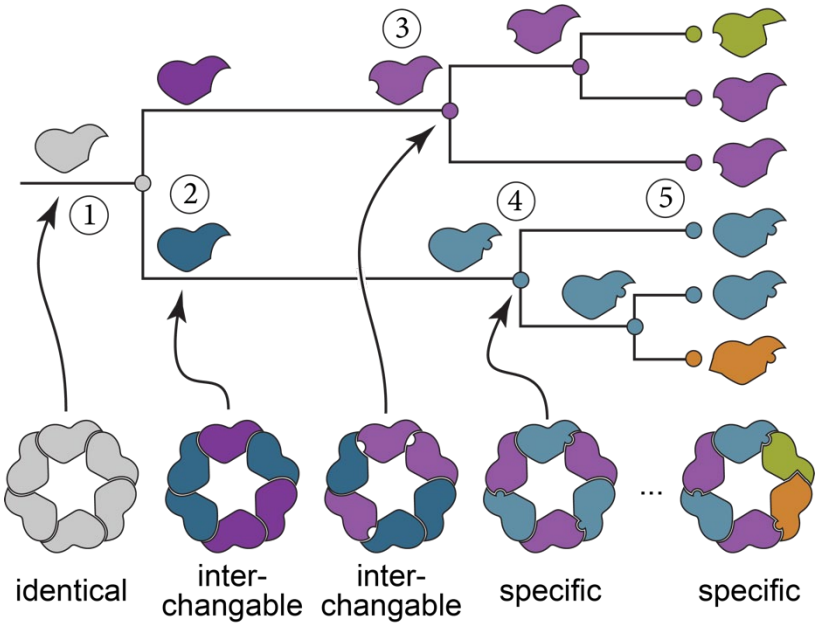


Figure 5.3: Ancestral Sequence Reconstruction can be used to trace the evolution of complicated protein features over time. This is a generalized view of the evolution of a heterohexamer, as found for multiple naturally occurring complexes [204–206]. (Step 1) The ancestral protein forms a homohexamer. (Step 2) The protein duplicates. The subunits have identical sequences and thus assemble in any stoichiometry and order. (Step 3) A mutation occurs to one subunit, creating a hole at the interface that is compatible with any assembly order. (Step 4) A mutation occurs on the other subunit, creating a knob at the interface that can only be accommodated by the hole on the other subunit. The proteins now form a specific hexamer with alternating subunits. (Step 5) Further duplications and mutations allow ever-more-specific complex assembly.

This example shows the basic logic of ASR studies, and how ASR can be used to pick apart the evolution of complicated, integrated protein features. With this strategy in mind, we can discuss how ASR has been used to learn about the evolution of protein function, specifically through the lens of energy landscapes.

Evolution of Folding Energy Landscapes

In this section, we examine how ASR has been used to study the evolution of energy landscapes. The study of folding energy landscapes is a natural place to start, as folding into the native state is a prerequisite for function in most proteins. What has ASR revealed about how the energy landscape controlling protein folding evolved?

One key question is how folding energy landscapes evolve in the first place. In one noteworthy example, researchers used ASR to unravel how a β-propeller fold evolved from much smaller subunits. A β-propeller is a closed repeat protein built from five tandem duplications of a short propeller-like motif [179,207]. Using lectin β-propeller evolution to model new fold emergence, Smock and colleagues resurrected an ancestral 47 aa protein encoding a single propellor-like motif [179]. This ancestral motif spontaneously forms a noncovalent pentamer in trans, mimicking the full β-propeller. Over evolutionary time, the gene encoding the monomer then duplicated, fused, and diversified to create new interfaces between subunits and, ultimately, a complete β-propeller. This evolutionary trajectory is remarkably similar to the evolutionary assembly of multi-protein complexes described in the previous section (Figure 5.3), suggesting that similar evolutionary mechanics can operate at the levels of both tertiary and quaternary structural assembly.

From the perspective of the energy landscape, one of the more intriguing aspects of this study was that it found that the folding constraints changed over evolutionary time. For the ancestor, function depended on stable folding of the monomer and efficient pentamer assembly. After duplication and fusion, new constraints emerged. One of the most notable was the requirement to avoid inappropriate β-sheet formation between subunits, which leads to misfolding. The transition from a motif assembling in trans to a full β-propeller thus required

smoothing the energy landscape by destabilizing—or, at least making kinetically inaccessible—misfolded forms of the protein.

The presence of a relatively complex folding energy landscape can also provide surprising opportunities for evolutionary optimization. One example comes from the folding of RNaseH [148,176,178]. This protein folds through a metastable, on-pathway intermediate. By characterizing ancestral proteins using hydrogen-deuterium exchange mass spectrometry, the Marqusee group found that the formation of the major folding intermediate has been conserved over billions of years.

To understand the evolutionary implications of the preserved intermediate, the researchers traced the evolution of RNaseH proteins starting from the ancestral protein and proceeding along the branches leading to mesophilic and thermophilic descendants. They discovered that preservation of the folding intermediate has allowed RNaseH proteins to decouple the evolution of thermodynamic stability (the proportion of molecules in the folded state at equilibrium) and kinetic stability (how long it takes before a protein unfolds once it reaches the native state) (Figure 5.4A). While kinetic stability increased on both lineages, thermodynamic stability only increased on the thermophilic lineage (Figure 5.4B). This was possible because kinetic stability increased by different mechanisms along the two lineages. On the thermophilic lineage, the mutations stabilized the native state, thus increasing both thermodynamic and kinetic stability (Figure 5.4C-D). On the mesophile lineage, in contrast, mutations destabilized the intermediate and folding transition state energy (Figure 5.4C,E). This increased kinetic stability by increasing the height of the unfolding barrier without altering the proportion of molecules folded at equilibrium.

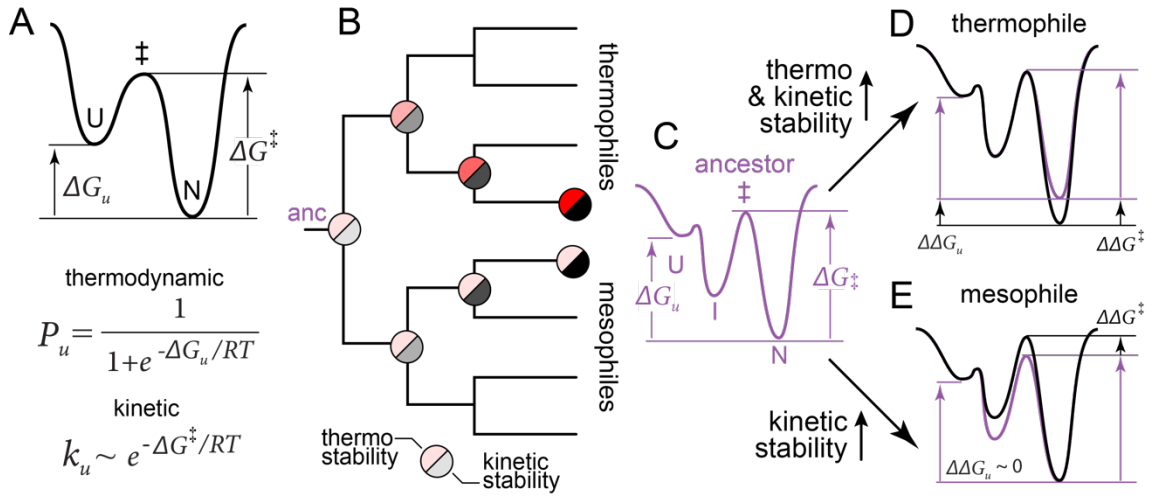


Figure 5.4. A folding intermediate decouples the evolution of thermodynamic and kinetic stability. A) Energy landscape illustrating thermodynamic stability (governed by ΔG_u) and kinetic stability (governed by $\Delta G^{\ddagger}$). B) Schematic tree showing evolution of RNaseH. The thermophile lineage showed increased kinetic stability (gray to black) and thermodynamic stability (pink to red). The mesophile lineage increased kinetic stability without altering thermodynamic stability. C) Energy landscape of the RNaseH ancestor (under conditions favoring folding). The protein folds through an on-pathway intermediate *I*. D) Energy landscape of a modern thermophile RNaseH. Relative to the ancestor (purple), the free energy of the native state decreased, increasing both thermodynamic ($\Delta\Delta G_u$) and kinetic ($\Delta\Delta G^{\ddagger}$) stability. E) Energy landscape of a modern mesophile RNaseH. The energy of the intermediate and folding barrier increased, increasing kinetic stability without altering thermodynamic stability.

Finally, ASR has recently been used to investigate how energy landscapes can encode multiple low energy conformations. In one recent example, researchers revealed how a few mutations stabilized an alternative conformation of the protein without destabilizing the existing native conformation [144]. They found a fascinating zigzagging evolutionary path, beginning with an ancestral protein that had a single native state. Mutations sequentially stabilized an alternate form of the protein, eventually causing the alternate conformation to become the most stable conformation. Further mutations rebalanced the energy landscape, giving both conformations nearly identical energies. As a result, the modern protein has two interconverting native states that allow the same gene to encode multiple functions.

Tuning the Energy Landscape to Confer New Functions

ASR has also revealed important changes to energy landscapes within the native state ensemble. Recently, for example, Yang and colleagues used ASR to study the evolution of new substrate specificity by the xenobiotic-degrading enzyme methyl-parathion hydrolase (MPH) [208]. MPH recently acquired the ability to act on four new organophosphate compounds. Yang and colleagues resurrected the last ancestral enzyme that was unable to recognize these substrates, then identified five mutations that were necessary and sufficient to confer the new activity. These mutations had two primary effects. First, they improved the ability of the enzyme to stabilize the appropriate transition state, thus increasing the rate of catalysis (Figure 5.5A). Second, these mutations reduced the prevalence of non-productive modes of binding (Figure 5.5B). In an energy landscape view, the mutations stabilized the transition state and destabilized several non-productive binding modes. In another study, researchers found that mutations away from the active site tuned the dynamics of the protein, and thus substrate specificity [168].

The importance of mutations disrupting non-productive conformations within the landscape has proven a common feature during the evolution of new activity and function. This has even held true for ASR studies that have dissected *de novo* evolution of enzyme active sites. For several different enzymes, residues within an existing binding site were repurposed when mutations altered the conformation of the site [146,171,209]. This reorganization of the binding site pre-positioned residues to bind the reaction transition state, lowering the reaction energy barrier and conferring a low level of enzymatic activity (Figure 5.5A). Subsequent mutations tuned the active site to increase activity. For both the evolution of a plant flavonoid biosynthetic enzyme and a bacterial cyclohexadienyl dehydratase [171,209], the final tuning stabilized the pocket, quenching non-productive dynamics and destabilizing non-productive interactions (Figure 5.5B).

Remarkably, many of the required changes to cyclohexadienyl dehydratase were distant from the active site [146,210], thus demonstrating an important role for the evolution of non-active site residues in tuning the energy landscape of evolving enzymes. We also note that this work powerfully demonstrates how ASR can identify residues important for function that may not be obvious from a simple structural analysis [151,152].

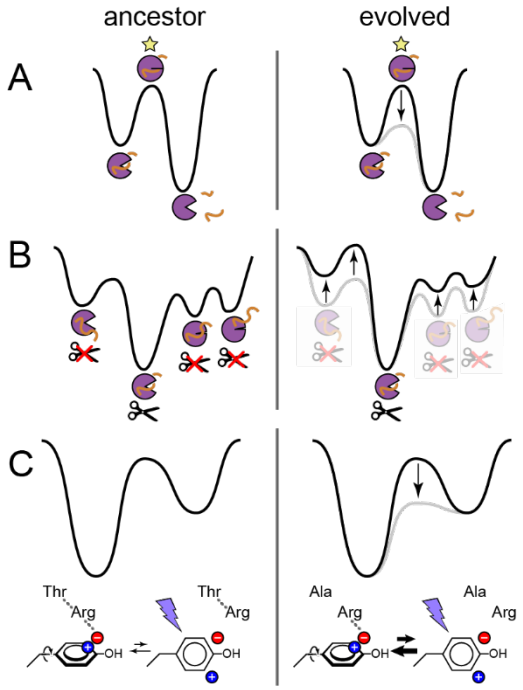


Figure 5.5. Evolution of function by alteration of energy landscapes. A) Icons show an enzyme (purple) operating on a substrate (orange). In the ancestor, the transition state energy (closed enzyme, yellow star) is too high to allow catalysis; mutations that lower the energy of the transition state convert the ancestral protein into an enzyme. B) An enzyme binds to a substrate in multiple conformations. One is productive (scissors); the other are non-productive (red "×"). Mutations that destabilize the non-productive conformations increase the bulk rate of catalysis. C) The slow step in the photoconversion of a Kaede green fluorescent protein (GFP)-like chromophore is breaking an ion pair near the chromophore (blue +, red -). A historical Thr to Ala mutation at a site away from the chromophore disrupted a polar network stabilizing the ion pair, thus increasing the rate of photoconversion.

The evolution of activity sometimes requires the addition of new dynamics rather than quenched dynamics. The Matz group dissected the evolution of photoconversion activity from an ancestral green fluorescent protein (GFP)-like protein from corals. The ancestral GFP-like

protein autocatalytically forms a green chromophore. Using ASR, the researchers identified a subset of historical substitutions that conferred the ability of the protein to photoconvert from green to red upon exposure to UV light [187]. Photoconversion requires that the chromophore rotate and pick up a nearby proton before interacting with an incoming photon. This rotation and proton pickup requires breaking an adjacent ion pair. One of the key mutations that occurred over this interval disrupted a hydrogen bond network stabilizing this ion pair, lowering the energetic barrier of rotating the chromophore and dramatically increasing the rate of photoconversion (Figure 5.5C) [188].

We highlighted only a few studies in this section, but there has been extensive work using ASR to reveal how small perturbations to the energy landscape have led to the evolution of new functions. Other examples include studies of the evolution of a binding protein from an enzyme by massively quenching the dynamics of the protein [150], conferring new enzyme activity by promoting oligomerization, and thus creating a new active site [181], creating a new binding interaction by mutations promoting an induced fit mechanism [211], mutations that tune allosteric networks [212–215], tuning loop dynamics to alter activity [180,216], and the possibility that altered heat capacity of the transition state was important for the evolution of increased enzyme activity [147].

Energy Landscapes Constrain Evolution

ASR has also revealed ways in which protein energy landscapes shape evolution. One example comes from our own work. We and our colleagues used ASR to study the evolution of a new proinflammatory function by the mammalian protein S100A9 [217,218]. Reverting a functionally important site in the human protein to its ancestral phenylalanine disrupted the function of the protein. The reversion creates a new Phe-Phe contact that stabilizes a non-

functional form of the protein. In wildtype S100A9, the non-functional form of the protein is quite close in energy to the native state—so close that a single new Phe-Phe interaction makes it the new native state (Figure 5.6A). A co-evolutionary analysis of S100A9 proteins from across mammals revealed that Phe can be found at either site in the Phe-Phe pair, but almost never at both sites together. This suggests that the protein has been evolving to avoid this deleterious contact, and that the energy landscape constrains how the protein can evolve.

Other studies have revealed similar constraints [219–221]. For example, the glucocorticoid receptor in bony vertebrates evolved ligand specificity through a conformational shift in a helix bordering the binding site. This was enabled by "permissive" mutations that stabilized the helix in its new conformation. Although many mutations can stabilize the protein, only a few can act as permissive mutations. This is because many of the stabilizing mutations shift the energy landscape to favor the active form even in the absence of ligand, thus making the receptor constitutively active [219].

Energy Landscapes Open and Close Evolutionary Trajectories

The Phe-Phe interaction discussed above is an example of the broad phenomenon of epistasis, where the effect of a mutation changes depending on the presence of other mutations (Figure 5.6B). Epistasis profoundly alters evolutionary outcomes by opening and closing evolutionary trajectories. Figure 5.6B shows a mutant cycle with two mutations a→A and b→B. The a→A mutation disrupts function on its own, but can be tolerated after the b→B mutation. This means the evolutionary trajectory ab→aB→AB is favored over the ab→Ab→AB trajectory. Epistasis introduces contingency into evolution [219,222,223]: a→A can only occur after the

permissive b→B mutation. Epistasis also causes entrenchment [189,223,224]: Once the restrictive a→A mutation occurs in the aB background, it prevents the reversion B→b.

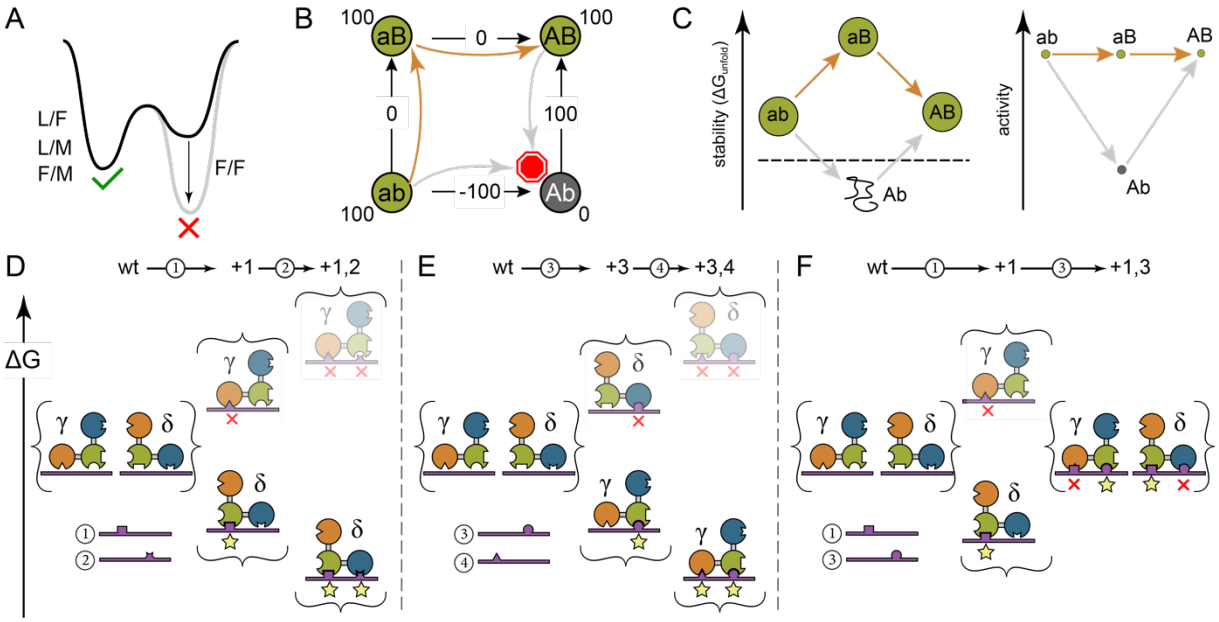


Figure 5.6. Epistasis arising from energy landscapes. A) The S100A9 energy landscape constrains evolution; Phe is tolerated at either site, but placing Phe at both sites stabilizes a non-functional form of the protein. B) Epistasis between mutations a→A and b→B shapes evolutionary trajectories. Genotypes ab, aB, and AB have the same activity (100; green); genotype Ab is nonfunctional (0; gray). Only the ab→aB→AB evolutionary trajectory is accessible without compromising function (orange arrows). Mutation a→A cannot be tolerated without the b→B mutation (bottom gray arrow, stop sign). Reversion of B→b cannot be tolerated after the a→A mutation (right gray arrow; stop sign). C) One biophysical explanation for the epistasis in panel B. Mutations a→A and b→B have opposite, additive effects on protein stability, leading to epistasis in activity. Genotype Ab is unfolded (left); therefore, there is less active protein in the cell and activity is low (right). D) Evolution of a new interaction between two macromolecules leads to epistasis. The purple molecule acquires mutations promoting interaction with the molecule drawn as colored circles. Arrows and letters above the panel indicate the evolutionary trajectory. The wildtype purple genotype allows two isoenergetic conformations (γ and δ). Mutations "1" and "2" to the purple molecule stabilize conformation δ (yellow stars) and destabilize conformation γ (red ×), this increases the affinity of the interaction. E) Identical to panel D, except mutations "3" and "4" to the purple molecule stabilize γ and destabilize δ. F) Epistasis mediated by the energy landscape. When introduced after "1", mutation "3" destabilizes the interaction because it "1" and "2" have opposite effects on γ and δ.

Epistasis can arise directly from protein energy landscapes. One way in which epistasis can arise, well documented in a variety of ASR studies, is via the stability of the native state [220,221,225,226] (Figure 5.6C). A mutation that compromises protein stability may not be tolerated unless it is preceded by a different mutation that stabilizes the protein (Figure 5.6C). Such stability tradeoffs often occur during the evolution of new functions. Function-switching mutations often compromise protein stability by creating unsatisfied interactions that confer binding specificity or, in the case of enzymes, stabilize a transition state; these destabilizing effects are often offset by stabilizing mutations [220,221,225,226].

ASR has also revealed more subtle ways that mutations can perturb energy landscapes, thus changing the accessibility of evolutionary trajectories. Figure 5.6D-F shows a schematic landscape that starts with a weak interaction between two molecules. This weak interaction can occur via two, initially isoenergetic, conformations (γ and δ). If mutations 1 and 2 occur in the molecule, then they increase the overall favorability of the interaction by sequentially stabilizing δ and destabilizing γ (Figure 5.6D). Mutations 3 and 4 promote the same interaction, but instead stabilize γ over δ (Figure 5.6E). Starr and colleagues observed this phenomenon for the evolution of transcription factors interacting with specific DNA response elements [227]. Using a combination of ASR and deep mutational scanning, they found that there were several structurally different ways for an ancestral transcription factor to evolve to bind specific DNA sequences. Commitment to one binding model excluded the other binding model. This leads to profound epistasis. In the evolutionary trajectory shown in Figure 5.6E, mutation 3 stabilizes the interaction; however, the same mutation destabilizes the interaction in Figure 5.6F because mutation 1 has already committed the interaction to favor conformation δ rather than γ. The initial mutation selecting one conformation or the other entrenches that particular outcome.

These examples demonstrate pairwise epistasis between two mutations; however, ASR studies have also revealed examples of high-order epistasis among three or more mutations [187,208,221,224]. In three-way epistasis, for example, the effect of three mutations placed together cannot be predicted from their individual effects combined with any pairwise epistasis [228–230]. Such high-order interactions arise naturally on energy landscapes that populate more than two conformations [231]. In the ASR study described above, where the authors dissected the evolution of new enzymatic activity in MPH [208], they generated all 2^5 combinations of the five mutations that they identified as important to the new activity. They found extensive high-order epistasis among the mutations, mediated by subtle changes in structure and binding energy. This highly constrained the evolution of MPH activity. Out of the 120 (5!) possible paths through this functional landscape, only 19 were accessible between the ancestral and extant MPH, underscoring the importance of mutation order in evolution. This study joins a wealth of other ASR studies showing how subtle biophysical changes can profoundly alter evolutionary outcomes by inducing epistasis [219–221,227,232].

Increasingly Thermostable Ancestral States: An Artifact of the Energy Landscape?

Viewing protein evolution from the perspective of an energy landscape may also shed light on some puzzling observations from ASR studies. One such observation is that resurrected ancestral proteins often exhibit increasing thermostability as one moves deeper into the past (Figure 5.7A) [149,172,175,177,233–240]. This makes reconstructed proteins robust and opens the door for aggressive engineering approaches that would otherwise result in primarily dead proteins [161,241–246].

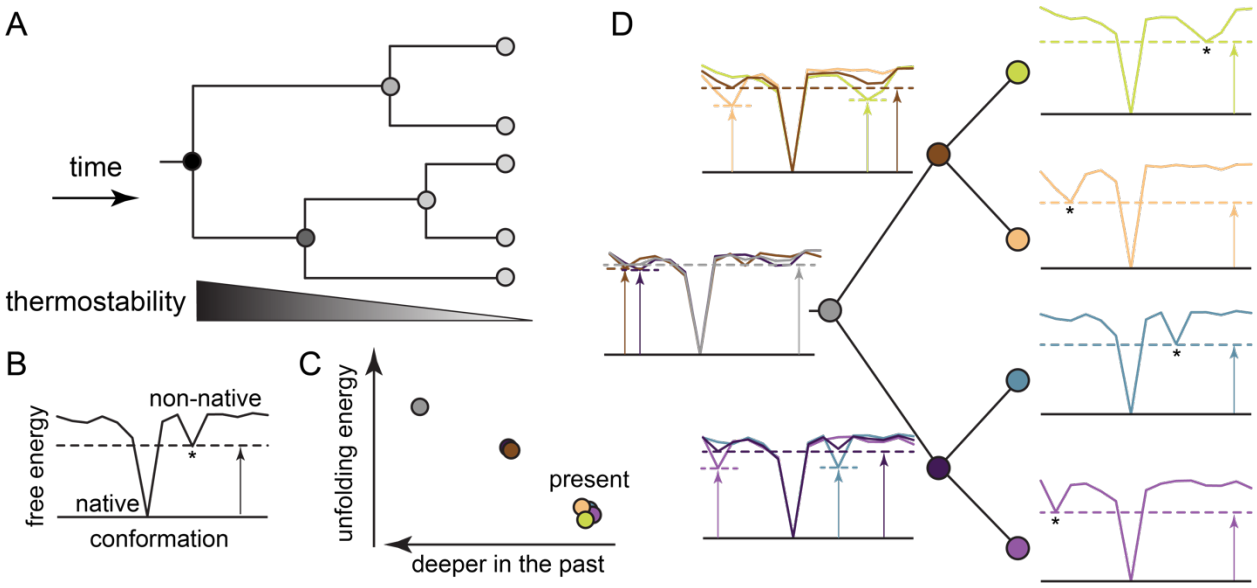


Figure 5.7. Evolution of consensus landscapes as a possible mechanism for ancestral stabilization. A) Ancestral sequence reconstruction studies have found that more ancient ancestors, on most lineages, have higher thermostability than later ancestors. Thermostability is represented as a gradient from high (dark) to low (light). B) An energy landscape with a protein that has a single high-energy non-native conformation. The arrow indicates the energy difference between the native state and the lowest energy excited state (star). C) A hypothetical trend in protein stability, with ancestral colors corresponding to the tree. D) Modern proteins on the right have the same native state but different excited states. Because of this, the evolutionary signal is stronger for the native state than for the excited states. When ancestors are reconstructed, the physical interactions encoding the native state are correctly inferred, leading to an accurate native state conformational energy. The evolutionary signal for the excited states is weak; therefore, the interactions stabilizing specific excited states are not accurately inferred. Ancestors have an "average" set of non-native interactions that is incompatible with specific non-native conformations. Thus, reconstructed ancestors maintain the native conformation energy, but sequentially average out the energies of non-native conformations, leading to a larger energy difference between the native and non-native conformations the deeper one goes back in time.

The origins of this trend, however, remain unclear. Some have argued that the trend toward higher stability in the past arises because ancestral environments were hotter, necessitating more stable ancient proteins [247,248]. This would make ASR a useful tool to study ancient environments [175,249]. One difficulty with this view is that temperature conditions varied widely over time and geographical location, making the uniformity of these trends across all lineages difficult to rationalize [250].

Others have argued that the trend towards hyperstability is an artifact of the methods used to reconstruct ancestors [251–254]. One possible cause of artifactually high stability of ancestral proteins would be an inappropriate introduction of consensus residues. Consensus proteins—where one substitutes the most common amino acid seen in a multiple sequence alignment at each site in the protein—are consistently hyperstable [251]. If the same stabilizing amino acid was acquired convergently on multiple lineages, ASR methods could incorrectly infer that the ancestor had that amino acid [254]. As a result, ancestral proteins would have, on average, an excess number of stabilizing amino acids and would thus prove artificially stable. This effect could also arise from phylogenetic model violation. Most common methods assume that the frequencies of amino acids at each site have not changed over time; if this is not true, the reconstructed ancestors could be biased towards amino acids with higher frequency in the alignment [252].

The evidence that ASR converges on consensus sequences remains mixed. Not all reconstructed ancestors appear consensus-like [251,253]. Furthermore, it is not clear that adding a handful of consensus mutations would be sufficient to confer hyperstability. Sternke and colleagues compared the measured effects of mutations across a database of proteins to their amino acid frequencies in the proteomes. There was low correlation between the relative stabilizing effect of the mutations and the frequency of amino acids, suggesting that the consensus explanation may not be sufficient to explain the stabilization effect [251]. Furthermore, given the degree of epistasis in proteins, it is not clear that a consensus mutation, introduced by itself, would be universally stabilizing across all backgrounds. This would be especially true for evolutionary changes in function that induce evolutionary Stokes shifts— reorganizations of

residue preference across sites after a large-effect substitution [255,256]. The origins of ancestral hyperstability thus remain unclear.

With the protein energy landscape in view, we propose another way in which ASR could artificially increase protein stability: evolution towards a consensus landscape rather than a consensus sequence. This idea is tentative and must be tested. The idea has three parts. A) As most proteins evolve, their native state is under purifying selection and thus remains relatively unchanged. The residues encoding interactions in the native state will evolve relatively slowly. B) In addition to the native conformation, protein energy landscapes possess relatively low-energy non-native conformations (Figure 5.7B). These conformations have no functional constraints—other than not becoming so populated that they disrupt function—and thus are randomly formed and destroyed as the protein sequence evolves [257]. C) When an ancestor is reconstructed, it will have higher quality signal for residues encoding the native conformation than for those encoding the non-native conformations. Because the sequences and identities of the non-native interactions change relatively rapidly over time, they are smeared out by the reconstruction, leading toward an "average" landscape dominated by the native structure (Figure 5.7C,D). The net result is a smoother energy landscape and a more dominant contribution to stability by the native state. This effect would become greater the further back one went in time, as one is averaging over a larger number of non-native conformations.

Limitations and Future Directions

While ASR has proven powerful, there are limitations to the approach. Some of these limitations are intrinsic. Evolutionary models will always be approximations of a complicated

historical process. Furthermore, some evolutionary events would require sequences from species that went extinct and thus have no modern sequences to include in an MSA.

Some limitations to ASR methods can, however, be addressed and improved. Substitution models (Figure 5.2D) are one important aspect of the method that can be improved upon. The most popular models assume that all sites in the protein have the same substitution probabilities and, on long time scales, converge to the same amino acid preferences [252]. It is not obvious, however, that one should treat the evolution of a site in the core of a protein with the same model one uses to treat a surface residue, nor that the preferences of amino acids at a site might change over time. Models have been developed that use different substitution models for different classes of sites. One model, for example, uses six matrices to treat sites classified based on their solvent accessibility and secondary structures [198]. Other work has been done to empirically define the amino acid preferences at sites using deep mutational scans [258]. Others have proposed using non-stationary models of amino acid frequencies, thus allowing amino acid preference to change over time [252,259]. Despite having clear utility for ASR studies [260]—particularly for biophysicists dissecting the detailed features of energy landscapes—many of these models have seen limited use in practice [260,261]. This is likely because such models have not been incorporated into mainstream phylogenetics software and require coding skill and/or detailed phylogenetics knowledge to use.

Other aspects of the models that can be improved include how they capture variations in evolutionary rate across sites and time, explicit models of insertion and deletion, methods to bring in outside information such as a species tree at the time of inference, attempts to treat co-variation between sites, and tree-search algorithms. These are active areas of development within the phylogenetics community [262,263], and continued improvements will almost certainly lead to

better reconstructed ancestors for biophysical study. That said, it is critical for researchers to keep in mind that ASR studies can only be done with confidence on protein families that adhere reasonably well to the assumptions of current phylogenetic models.

Conclusion

ASR has proven to be a powerful means to access information on how protein energy landscapes evolve. By separating the evolution of protein features in time, ASR provides a nuanced view of how mutations alter energy landscapes during the evolution of protein function. Furthermore, ASR has revealed how constraints due to the energy landscape have shaped—and continue to shape—the evolution of protein function. Continued methodological development promises to make the tool even more useful for biophysicists. As the approach is applied to ever more protein families, it will continue to reveal both how proteins acquired their amazingly diverse functions, and general principles that can be used to understand and maybe, someday, predict protein evolution.

Bridge to Chapter VI:

In Chapter V, we reviewed how ASR studies have been used to dissect the evolution of energy landscapes. When coupled to biophysical, biochemical, and functional characterization, ASR can reveal how historical mutations altered the energy landscape of ancient proteins, allowing the evolution of enzyme activity, altered conformations, binding specificity, oligomerization, and many other protein features. We also discussed ASR studies that reveal how energy landscapes have shaped protein evolution. Finally, we proposed that thinking about evolution from the perspective of an energy landscape can improve how we approach and

interpret ASR studies. Completing my exploration of protein evolution and innate immunity, in Chapter VI I summarize and tie together the findings of the previous chapters and provide concluding remarks.

CHAPTER SIX

CONCLUDING REMARKS

This dissertation detailed four studies centering around the interaction between S100A9 and the TLR4 complex, incorporating a biochemical and evolutionary perspective. I set out to answer two questions: 1) How does S100A9 activate TLR4? And 2) How did TLR4 evolve and maintain LPS and S100A9 recognition?

Chapter II explored the consequences of changing the expression system used to produce recombinant S100A9. I demonstrated several key differences between $S100A9^{in}$, and the field standard $S100A9^{ec}$. Both proteins share secondary structure and the ability to bind calcium, but $S100A9^{in}$ does not activate TLR4. $S100A9^{in}$ also differs in tertiary structure, and is more prone to higher order oligomer formation. I reported that disruption of higher order oligomer formation with an E.Coli disaggregase restores $S100A9^{in}$ activation of TLR4, which suggests that the oligomeric state of S100A9 determines proinflammatory activity.

In Chapter III I examined the role of TLR4's co-receptor CD14 in TLR4 recognition of S100A9, and proposed a structural model for the protein-protein interaction between S100A9 and CD14. I showed that CD14 markedly improves S100A9's ability to activate TLR4, and set forth evidence that this effect is due to CD14-dependent internalization. I also used computational modeling paired with extensive mutagenesis and functional testing to develop and provide support for a S100A9/CD14 docking model.

Chapter IV described the development of a method for studying the evolution and function of TLR4 in high throughput. I reported the use of calculated selection rates (dN/dS) across sites of TLR4 to inform site selection for a high throughput TLR4 activity assay. I

described my developed method, and the results of a pilot screen which assessed the validity of the method.

In Chapter V I reviewed the use of ancestral sequence reconstruction in studying protein energy landscapes. I covered the basics of an ASR study, and examined how energy landscapes both constrain and open evolutionary trajectories. I also provided an energy landscape perspective on observed trends in stability in reconstructed ancestors.

S100A9 was first identified as a possible drug target in 2009[14] based on its ability to directly activate TLR4, but 15 years later there remains no biochemical mechanism for this interaction. This dissertation described my contributions to this open problem in the field, incorporating biochemical and evolutionary techniques. I characterized for the first time recombinant S100A9 purified from eukaryotic cells, and unveiled key changes in structure and function. I reported the first attempt at a detailed characterization of the interaction between S100A9 and TLR4's co-receptor CD14. I pioneered the development of a first of its kind high throughput TLR4 activity assay. Though many open questions remain, my work provides valuable insights into the mechanism and evolution of S100A9 activation of TLR4 as well as providing new tools for use in future study of the interaction.

APPENDIX A

SUPPLEMENTAL INFORMATION FOR CHAPTER III

Appendix A is the supplementary information for Chapter III, it contains supplementary figures and a supplementary table referenced in Chapter III.

Table S3.1: Effect of CD14 mutations on S100A9 and LPS activity. Bio reps is the number of biological replicates, SEM is standard error of the mean, Fig is the figure where the data is displayed in the main text, where relevant.

CD14 genotype	agonist	agonist concentration (uM or ng/mL)	Bio reps	relative TLR4 activity	SEM	experiments performed by	Fig.
A61A G62A G63A	S100A9	0.5	2	0.61165523	0.041421	NMJ	
A61A G62A G63A	S100A9	2	2	0.84202834	0.045334	NMJ	4
A61A G62A G63A	LPS	10	2	0.18257309	0.016014	NMJ	
A61A G62A G63A	LPS	200	2	0.72328762	0.059926	NMJ	4
A88A L89A R90A	S100A9	0.5	2	0.950722	0.157754	NMJ	
A88A L89A R90A	S100A9	2	2	1.30996386	0.365212	NMJ	4
A88A L89A R90A	LPS	10	2	0.45274338	0.042531	NMJ	
A88A L89A R90A	LPS	200	2	1.10206686	0.016759	NMJ	4
A88E	S100A9	2	3	1.05164294	0.101495	LOC	4
A88E	LPS	200	3	1.08465476	0.089747	LOC	4
A88W	S100A9	0.5	3	0.85324386	0.021026	LOC	
A88W	S100A9	2	3	1.0707145	0.052642	LOC	4
A88W	LPS	10	3	0.18528188	0.043043	LOC	
A88W	LPS	200	3	0.89400792	0.098637	LOC	4
C25A E26A L27A	S100A9	0.5	3	0.96748169	0.084726	NMJ	
C25A E26A L27A	S100A9	2	3	1.27736081	0.060683	NMJ	4
C25A E26A L27A	LPS	10	3	0.51688914	0.254733	NMJ	
C25A E26A L27A	LPS	200	3	1.11832696	0.22481	NMJ	4
D28A D29A E30A	S100A9	0.5	2	0.75518468	0.289399	NMJ	
D28A D29A E30A	S100A9	2	2	1.14679691	0.249739	NMJ	4
D28A D29A E30A	LPS	10	2	0.05849923	0.089484	NMJ	
D28A D29A E30A	LPS	200	2	0.91934661	0.131754	NMJ	4
D76A A77A D78A	S100A9	0.5	2	0.66396248	0.009028	NMJ	
D76A A77A D78A	S100A9	2	2	0.97144229	0.217674	NMJ	4
D76A A77A D78A	LPS	10	2	0.18600752	0.082257	NMJ	

Table S3.1 Continued…

CD14 genotype	agonist	agonist concentration (uM or ng/mL)	Bio reps	relative TLR4 activity	SEM	experiments performed by	Fig.
D76A A77A D78A	LPS	200	2	0.76205636	0.047649	NMJ	4
D78A P79A R80A Q81A Y82A D84A	S100A9	0.008230453	3	-0.0021814	0.002635	LOC	
D78A P79A R80A Q81A Y82A D84A	S100A9	0.024691358	3	0.00110531	0.012343	LOC	
D78A P79A R80A Q81A Y82A D84A	S100A9	0.074074074	3	0.00651307	0.009933	LOC	
D78A P79A R80A Q81A Y82A D84A	S100A9	0.222222222	3	0.01730694	0.00723	LOC	
D78A P79A R80A Q81A Y82A D84A	LPS	0.274348422	3	0.01887388	0.019762	LOC	
D78A P79A R80A Q81A Y82A D84A	S100A9	0.666666667	3	0.0881389	0.032856	LOC	
D78A P79A R80A Q81A Y82A D84A	LPS	0.823045267	3	0.01514795	0.018107	LOC	
D78A P79A R80A Q81A Y82A D84A	S100A9	2	5	0.20593335	0.014928	LOC	4
D78A P79A R80A Q81A Y82A D84A	LPS	2.469135802	3	0.00973919	0.025464	LOC	
D78A P79A R80A Q81A Y82A D84A	S100A9	6	3	0.24239658	0.036924	LOC	
D78A P79A R80A Q81A Y82A D84A	LPS	7.407407407	3	0.00656383	0.024523	LOC	
D78A P79A R80A Q81A Y82A D84A	LPS	22.22222222	3	0.01833316	0.037496	LOC	
D78A P79A R80A Q81A Y82A D84A	LPS	66.66666667	3	0.00647868	0.030604	LOC	
D78A P79A R80A Q81A Y82A D84A	LPS	200	5	0.04254643	0.026391	LOC	4
D84A	S100A9	2	2	0.94847951	0.027928	LOC	4
D84A	LPS	200	2	0.87713006	0.124167	LOC	4
D84W	S100A9	0.5	3	0.45279155	0.010536	LOC	
D84W	S100A9	2	3	0.88283323	0.047094	LOC	4
D84W	LPS	10	3	0.0233146	0.006727	LOC	
D84W	LPS	200	3	0.44486087	0.119624	LOC	4
D84W A88W	S100A9	0.5	3	0.55969646	0.004528	LOC	
D84W A88W	S100A9	2	3	1.04291914	0.073533	LOC	4
D84W A88W	LPS	10	3	0.06122604	0.021255	LOC	
D84W A88W	LPS	200	3	0.53908132	0.062089	LOC	4
D84W T85W	S100A9	0.5	3	0.40960594	0.051809	LOC	
D84W T85W	S100A9	2	3	0.78647053	0.063121	LOC	4
D84W T85W	LPS	10	3	0.01246812	0.044783	LOC	
D84W T85W	LPS	200	3	0.21045553	0.01603	LOC	4
D84W T85W A88W	S100A9	0.5	3	0.74657922	0.088874	LOC	

Table S3.1 Continued…

CD14 genotype	agonist	agonist concentration (uM or ng/mL)	Bio reps	relative TLR4 activity	SEM	experiments performed by	Fig.
D84W T85W A88W	S100A9	2	3	1.12449985	0.011362	LOC	4
D84W T85W A88W	LPS	10	3	0.08830574	0.006112	LOC	
D84W T85W A88W	LPS	200	3	0.4425363	0.050926	LOC	4
E40A P41A Q42A	S100A9	0.5	2	0.42441017	0.095726	NMJ	
E40A P41A Q42A	S100A9	2	2	0.88249493	0.019671	NMJ	4
E40A P41A Q42A	LPS	10	2	0.53128394	0.028448	NMJ	
E40A P41A Q42A	LPS	200	2	1.05886637	0.015667	NMJ	4
E56K	S100A9	2	3	0.91185204	0.246027	HEM	4
E56K	LPS	200	3	1.44919567	0.383864	HEM	4
E58A I59A H60A	S100A9	0.5	2	0.78153179	0.044765	NMJ	
E58A I59A H60A	S100A9	2	2	1.0202035	0.076123	NMJ	4
E58A I59A H60A	LPS	10	2	0.14953222	0.125029	NMJ	
E58A I59A H60A	LPS	200	2	0.70878398	0.08024	NMJ	4
E67A P68A F69A	S100A9	0.5	2	0.78971129	0.240992	NMJ	
E67A P68A F69A	S100A9	2	2	1.07265728	0.206176	NMJ	4
E67A P68A F69A	LPS	10	2	0.30511457	0.301058	NMJ	
E67A P68A F69A	LPS	200	2	0.81063647	0.318153	NMJ	4
F49A	S100A9	0.008230453	3	0.00927959	0.008065	LOC	
F49A	S100A9	0.024691358	3	0.03039725	0.012654	LOC	
F49A	S100A9	0.074074074	3	0.11309288	0.013585	LOC	
F49A	S100A9	0.222222222	3	0.34268843	0.055591	LOC	
F49A	LPS	0.274348422	3	0.10567787	0.013999	LOC	
F49A	S100A9	0.666666667	3	0.56546777	0.101434	LOC	
F49A	LPS	0.823045267	3	0.23655288	0.017937	LOC	
F49A	S100A9	2	10	0.65358832	0.131738	LOC, HEM	4
F49A	LPS	2.469135802	3	0.41839013	0.045264	LOC, HEM	
F49A	S100A9	6	3	0.77711884	0.108486	LOC	
F49A	LPS	7.407407407	3	0.59747002	0.081854	LOC	
F49A	LPS	22.22222222	3	0.78648765	0.114436	LOC	
F49A	LPS	66.66666667	3	0.76178452	0.112745	LOC	
F49A	LPS	200	10	1.04569382	0.185748	LOC,HEM	4
F49A Q50A C51A	S100A9	0.5	3	0.40575198	0.07471	NMJ	
F49A Q50A C51A	S100A9	2	3	0.7040157	0.128323	NMJ	4
F49A Q50A C51A	LPS	10	3	0.16534609	0.08093	NMJ	
F49A Q50A C51A	LPS	200	3	0.54929575	0.124044	NMJ	4
F69A	S100A9	2	3	1.48237853	0.639999	HEM	4
F69A	LPS	200	3	1.47604902	0.218936	HEM	4

Table S3.1 Continued…

CD14 genotype	agonist	agonist concentration (uM or ng/mL)	Bio reps	relative TLR4 activity	SEM	experiments performed by	Fig.
G98A A99A A100A	S100A9	0.5	2	0.49892915	0.246548	NMJ	
G98A A99A A100A	S100A9	2	2	0.83676432	0.055032	NMJ	4
G98A A99A A100A	LPS	10	2	0.21612866	0.012982	NMJ	
G98A A99A A100A	LPS	200	2	0.79298461	0.08571	NMJ	4
L70A K71A R72A	S100A9	0.5	2	0.29215036	0.054211	NMJ	
L70A K71A R72A	S100A9	2	2	0.62726378	0.047155	NMJ	4
L70A K71A R72A	LPS	10	2	0.17640785	0.016058	NMJ	
L70A K71A R72A	LPS	200	2	0.6585598	0.049882	NMJ	4
L89A	S100A9	2	3	1.29437834	0.243854	HEM	4
L89A	LPS	200	3	1.70284608	0.146335	HEM	4
L95A T96A V97A	S100A9	0.5	2	0.39116598	0.115549	NMJ	
L95A T96A V97A	S100A9	2	2	0.82509621	0.077352	NMJ	4
L95A T96A V97A	LPS	10	2	0.37071744	0.278776	NMJ	
L95A T96A V97A	LPS	200	2	0.80845797	0.094767	NMJ	4
N37A F38A S39A	S100A9	0.5	2	0.65085811	0.049903	NMJ	
N37A F38A S39A	S100A9	2	2	0.72494641	0.02388	NMJ	4
N37A F38A S39A	LPS	10	2	0.12197674	0.110099	NMJ	
N37A F38A S39A	LPS	200	2	0.50984549	0.151249	NMJ	4
P43A D44A W45A	S100A9	0.5	4	0.60907178	0.178623	NMJ	
P43A D44A W45A	S100A9	2	4	0.85589658	0.134616	NMJ	4
P43A D44A W45A	LPS	10	4	0.46924407	0.123326	NMJ	
P43A D44A W45A	LPS	200	4	1.03516982	0.099926	NMJ	4
P79A R80A Q81A	S100A9	0.5	2	0.41795434	0.017029	NMJ	
P79A R80A Q81A	S100A9	2	2	0.63426218	0.009649	NMJ	4
P79A R80A Q81A	LPS	10	2	0.03583237	0.011373	NMJ	
P79A R80A Q81A	LPS	200	2	0.34305538	0.060244	NMJ	4
Q81A	S100A9	2	2	1.25582856	0.083989	LOC	4
Q81A	LPS	200	2	1.00422757	0.227389	LOC	4
R117E	S100A9	2	3	0.99831527	0.073242	LOC	
R117E	LPS	200	3	1.05512239	0.217573	LOC	
R33E	S100A9	2	3	0.88790007	0.059795	HEM	4
R33E	LPS	200	3	1.15954261	0.068734	HEM	4
R90E	S100A9	2	3	1.00158226	0.055776	LOC	
R90E	LPS	200	3	1.06224004	0.173218	LOC	
R90E R92E R117E	S100A9	2	3	0.79553906	0.099853	LOC	5
R90E R92E R117E	LPS	200	3	1.05696636	0.022151	LOC	5
R92E	S100A9	2	3	0.90137072	0.168111	LOC	

Table S3.1 Continued…

CD14 genotype	agonist	agonist concentration (uM or ng/mL)	Bio reps	relative TLR4 activity	SEM	experiments performed by	Fig.
R92E	LPS	200	3	0.86753081	0.24398	LOC	
S46A E47A A48A	S100A9	0.5	3	0.52948173	0.135566	NMJ	
S46A E47A A48A	S100A9	2	3	0.71970844	0.170587	NMJ	4
S46A E47A A48A	LPS	10	3	0.59584827	0.190521	NMJ	
S46A E47A A48A	LPS	200	3	0.93563885	0.10491	NMJ	4
S46A E47A A48A F49A	S100A9	2	3	0.73634718	0.038277	LOC	5
S46A E47A A48A F49A	LPS	200	3	1.20485643	0.112596	LOC	5
S46A E47A A48A F49A R90E R92E R117E	S100A9	2	3	0.93680358	0.102732	LOC	5
S46A E47A A48A F49A R90E R92E R117E	LPS	200	3	0.94210347	0.139949	LOC	5
S46C	S100A9	0.018518519	2	0.0624757	0.005232	LOC	
S46C	S100A9	0.111111111	2	0.29575691	0.035534	LOC	
S46C	LPS	0.390625	2	0.05625127	0.056925	LOC	
S46C	S100A9	0.666666667	2	0.81757911	0.211999	LOC	
S46C	LPS	3.125	2	0.08436959	0.015048	LOC	
S46C	S100A9	4	2	1.02505728	0.17503	LOC	
S46C	LPS	25	2	0.48787396	0.234586	LOC	
S46C	LPS	200	2	1.16415291	0.146688	LOC	4
T85A V86A K87A	S100A9	0.5	2	1.19321582	0.164902	NMJ	
T85A V86A K87A	S100A9	2	2	1.21681692	0.119664	NMJ	4
T85A V86A K87A	LPS	10	2	0.48833676	0.393585	NMJ	
T85A V86A K87A	LPS	200	2	0.86803924	0.284297	NMJ	4
T85W	S100A9	0.5	3	0.88432768	0.030311	LOC	
T85W	S100A9	2	3	1.29335454	0.102093	LOC	4
T85W	LPS	10	3	0.16284865	0.020433	LOC	
T85W	LPS	200	3	0.85891339	0.145531	LOC	4
T85W A88W	S100A9	0.5	3	0.85522043	0.007829	LOC	
T85W A88W	S100A9	2	3	1.37669509	0.060206	LOC	4
T85W A88W	LPS	10	3	0.04862281	0.012913	LOC	
T85W A88W	LPS	200	3	0.70332033	0.025253	LOC	4
V52A	S100A9	2	3	1.82482473	0.52622	HEM	4
V52A	LPS	200	3	1.58744225	0.141722	HEM	4
V52A S53A A54A	S100A9	0.5	2	0.50183271	0.010842	NMJ	
V52A S53A A54A	S100A9	2	2	0.77007818	0.063862	NMJ	4
V52A S53A A54A	LPS	10	2	0.51378994	0.058031	NMJ	
V52A S53A A54A	LPS	200	2	0.63014696	0.133579	NMJ	4

Table S3.1 Continued…

CD14 genotype	agonist	agonist concentration (uM or ng/mL)	Bio reps	relative TLR4 activity	SEM	experiments performed by	Fig.
V55A E56A V57A	S100A9	0.5	2	0.45573412	0.031117	NMJ	
V55A E56A V57A	S100A9	2	2	0.82130771	0.048192	NMJ	4
V55A E56A V57A	LPS	10	2	0.27504295	0.001586	NMJ	
V55A E56A V57A	LPS	200	2	0.86136246	0.136825	NMJ	4
V73A D74A A75A	S100A9	0.5	2	0.48401137	0.139807	NMJ	
V73A D74A A75A	S100A9	2	2	0.85453323	0.05978	NMJ	4
V73A D74A A75A	LPS	10	2	0.13380633	0.000892	NMJ	
V73A D74A A75A	LPS	200	2	0.65164653	0.096302	NMJ	4
V91A R92A R93A	S100A9	0.5	2	0.77719468	0.085335	NMJ	
V91A R92A R93A	S100A9	2	2	1.08407089	0.044607	NMJ	4
V91A R92A R93A	LPS	10	2	0.24994017	0.033298	NMJ	
V91A R92A R93A	LPS	200	2	0.91060921	0.093776	NMJ	4
W45A	S100A9	2	3	0.92465241	0.03976	LOC	4
W45A	LPS	200	3	0.97252061	0.163694	LOC	4
W45A S46A E47A A48- F49A	S100A9	2	3	0.8856116	0.091854	LOC	
W45A S46A E47A A48- F49A	LPS	200	3	1.08422483	0.072678	LOC	
W45A S46A E47A A48A F49A	S100A9	2	3	0.9319346	0.069636	LOC	
W45A S46A E47A A48A F49A	LPS	200	3	1.10969471	0.049529	LOC	
W45A S46A E47A A48A F49A R90E R92E R117E	S100A9	2	3	0.80618688	0.200947	LOC	
W45A S46A E47A A48A F49A R90E R92E R117E	LPS	200	3	0.82937242	0.268704	LOC	
W45C	S100A9	0.018518519	2	0.10066938	0.003603	LOC	
W45C	S100A9	0.111111111	2	0.34826358	0.058385	LOC	
W45C	LPS	0.390625	2	0.05364318	0.001721	LOC	
W45C	S100A9	0.666666667	2	0.97439097	0.174891	LOC	
W45C	LPS	3.125	2	0.14082586	0.032036	LOC	
W45C	S100A9	4	2	0.93141942	0.189867	LOC	
W45C	LPS	25	2	0.38954166	0.108016	LOC	
W45C	LPS	200	2	1.17535022	0.09088	LOC	4
W45C A77C	S100A9	0.018518519	2	0.09225071	0.003058	LOC	
W45C A77C	S100A9	0.111111111	2	0.35570102	0.029933	LOC	
W45C A77C	LPS	0.390625	2	-0.0042284	0.003978	LOC	
W45C A77C	S100A9	0.666666667	2	0.85583048	0.002058	LOC	
W45C A77C	LPS	3.125	2	0.05603889	0.008474	LOC	
W45C A77C	S100A9	4	2	0.80706832	0.020756	LOC	

Table S3.1 Continued…

CD14 genotype	agonist	agonist concentration (uM or ng/mL)	Bio reps	relative TLR4 activity	SEM	experiments performed by	Fig.
W45C A77C	LPS	25	2	0.31390285	0.134381	LOC	
W45C A77C	LPS	200	2	0.73274704	0.0428	LOC	4
Y82A	S100A9	2	5	1.1868813	0.144924	LOC, HEM	4
Y82A	LPS	200	5	1.31956285	0.308724	LOC, HEM	4
Y82A A83A D84A	S100A9	0.5	3	0.2445137	0.033047	NMJ	
Y82A A83A D84A	S100A9	2	3	0.52280401	0.128278	NMJ	4
Y82A A83A D84A	LPS	10	3	0.30139654	0.157843	NMJ	
Y82A A83A D84A	LPS	200	3	0.68361991	0.114632	NMJ	4

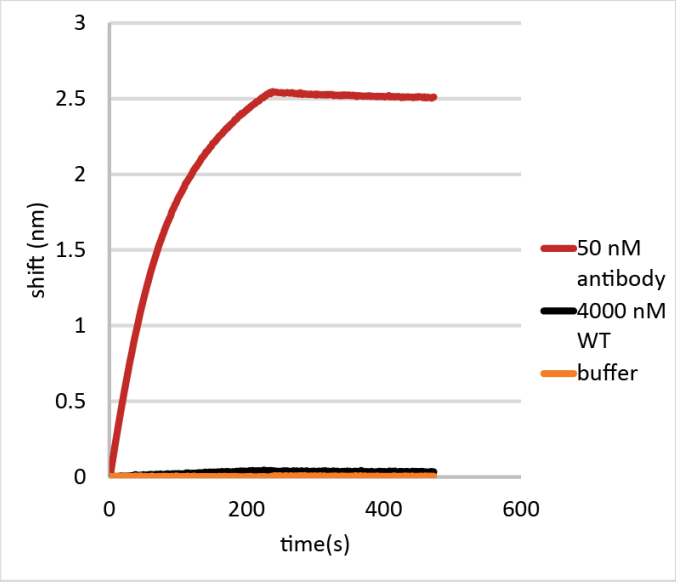


Figure S3.1: mAB antibody reveals S100A9 was successfully on the BLI sensor. Bio-Layer Interferometry binding data for immobilized S100A9 with different molecules flowed over the cell: 50 nM anti-S100A9 mAb (red), 4 μM sCD14 (black), and running buffer (orange). The S100A9 concentration was kept constant at 100 nM monomer. Binding plots are representative plots from one replicate. Binding was measured in triplicate at 50 nM and 100 nM mAb.

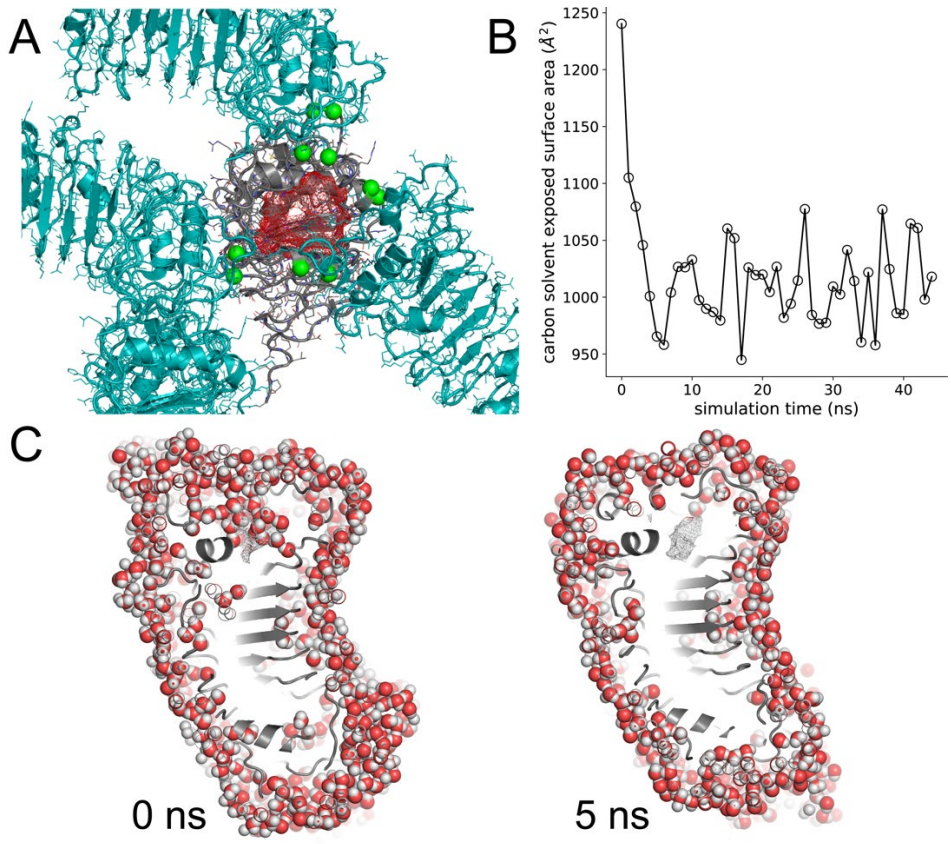


Figure S3.2. The CD14 crystal structure has a dynamic hydrophobic pocket propped open by crystal contacts. A) The crystal structure of human CD14 (RCSB ID: 4GLP). The CD14 monomer is the asymmetric unit, shown in gray. The four crystallographic symmetry mates adjacent to the binding pocket of the monomer are shown in teal. The green spheres are crystal contacts. In the crystal, CD14 has a large, concave hydrophobic surface (red mesh) surrounded by loops and helices supported by the crystal contacts. B) Plot shows solvent accessible surface area of CD14 carbons for snapshots taken ever nanosecond from an unrestrained MD simulation of the CD14 monomer. The exposed surface area drops by ~20% in ~5 ns. C) Sub-panels show slices lengthwise through the center of CD14, with the N-terminal hydrophobic pocket at the top and the C-terminus at the bottom. Waters near the protein are shown as spheres. The left sub-panel is the initial structure (0 ns); the right sub-panel is the structure after 5 ns of simulation time. Waters fill the hydrophobic pocket in the initial structure. By 5 ns, the lid helix (see Figure 6) closes and expels the water from the hydrophobic pocket.

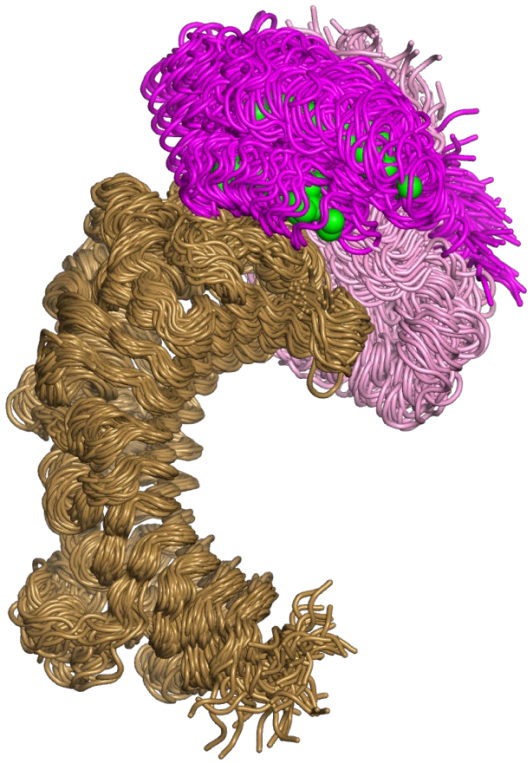


Figure S3.3. S100A9 fluctuates relative to CD14 over simulations. Figure shows overlay of 72 frames extracted from four replicate 500 ns unrestrained MD simulations of CD14 and S100A9, starting from the AlphaFold2 docking model. To reveal the movement of S100A9 relative to CD14, we aligned the core residues of CD14 (residues 100-200) from each frame to the starting structure. The S100A9 dimer is shown in pink (chain A) and magenta (chain B) with calcium ions shown as green spheres. CD14 is shown in brown. We omitted S100A9 residues 1-3 and 91-114 from the figure, as these residues are disordered and obscure the interaction.

APPENDIX B

SUPPLEMENTAL INFORMATION FOR CHAPTER IV

Appendix B is the supplementary information for Chapter IV, it contains supplementary tables referenced in Chapter IV.

Table S4.1: dN/dS calculations for TLR4 across mammals. Selection was calculated using PAML and HYPHY's MEME, FEL, and FUBAR[32,33,124–127]. PAML calculates dN/dS as ω and reports significance as P[ω >1], therefore ω values with a P > 0.95 are considered significant. MEME approximates dN/dS as LRT and reports significance as p-value, therefore LRT values with $p < 0.05$ are considered significant. FUBAR approximates dN/dS as β – α and reports P[α < β] therefore β – α values with P > 0.95 are considered significant (this is inverted for negative selection calculations). FEL calculates dN/dS as ω and reports significance as p-value, therefore $p < 0.05$ are considered significant. Note that only FEL and FUBAR calculate negative selection.

		PAML		*MEME*		*FUBAR*		*FEL*	
Amino Acid	***Selection***	**P[ω >1]**	**ω**	**LRT**	**p-value**	**β - α**	**P[α < β]**	**ω**	**p-value**
270	*positive*	0.988*	2.481 +- 0.173	6.29	0.02*	2.065	0.901*	Infinity	0.012*
276	*positive*	0.962*	2.441 +- 0.299	5.15	0.03*	1.621	0.86	3.677	0.069*
295	*positive*	0.971*	2.454 +- 0.264	0.58	0.41	1.167	0.464	2.023	0.445
300	*positive*	0.984*	2.475 +- 0.197	4.03	0.06*	1.803	0.82	5.667	0.045*
319	*positive*	0.991**	2.487 +- 0.145	7.15	0.01*	6.19	0.962*	17.692	0.009*
322	*positive*	1.000**	2.500 +- 0.004	0.59	0.41	3.078	0.777	2.321	0.444
347	*positive*	0.986*	2.478 +- 0.186	2.01	0.18	0.047	0.085	0.829	0.776
351	*positive*	1.000**	2.500 +- 0.023	5.05	0.04*	1.452	0.277	1.563	0.505
360	*positive*	0.972*	2.457 +- 0.258	3.09	0.1	1.622	0.837	5.22	0.087*
364	*positive*	0.979*	2.468 +- 0.223	2.97	0.11	0.821	0.396	1.762	0.538
365	*positive*	0.978*	2.466 +- 0.229	1.97	0.19	1.396	0.644	3.053	0.161

Table S4.1 Continued…

		PAML			*FUBAR*	FUBAR		FEL	
Amino Acid	***Selection***	**P[ω >1]**	**ω**	**LRT**	**p-value**	**β - α**	**P[α < β]**	**ω**	**p-value**
389	*positive*	0.988*	2.481 +- 0.173	3.37	0.09*	0.334	0.219	1.171	0.803
393	*positive*	0.972*	2.456 +- 0.260	1.73	0.21	0.242	0.177	0.865	0.853
394	*positive*	1.000**	2.500 +- 0.026	3.53	0.08*	6.238	0.918*	6.969	0.06*
400	*positive*	0.988*	2.481 +- 0.171	1	0.32	1.032	0.538	2.485	0.318
413	*positive*	0.975*	2.461 +- 0.246	4.64	0.05*	1.605	0.758	5.038	0.084*
437	*positive*	0.992**	2.487 +- 0.141	1.2	0.28	0.986	0.538	2.25	0.273
471	*positive*	0.997**	2.496 +- 0.083	8.65	0.01*	5.53	0.946*	7.397	0.008*
474	*positive*	0.956*	2.430 +- 0.324	2.28	0.16	1.449	0.726	4.521	0.131
500	*positive*	0.957*	2.433 +- 0.318	6.19	0.02*	2.591	0.784	5.848	0.053*
505	*positive*	0.998**	2.498 +- 0.062	0.99	0.32	1.922	0.451	1.983	0.347
517	*positive*	0.979*	2.467 +- 0.224	3.31	0.09*	0.959	0.507	1.989	0.413
520	*positive*	0.960*	2.437 +- 0.308	4.59	0.05*	1.165	0.641	2.482	0.225
537	*positive*	0.984*	2.475 +- 0.195	4.1	0.06*	1.43	0.732	3.858	0.127
648	*positive*	0.963*	2.442 +- 0.295	0.73	0.37	0.815	0.533	1.782	0.392
9	*negative*	.	.	.	.	-2.864	0	0	0
13	*negative*	.	.	.	.	-1.049	0.005	0	0.006
20	*negative*	.	.	.	.	-2.24	0.001	0	0
22	*negative*	.	.	.	.	-1.575	0.001	0	0.001
23	*negative*	.	.	.	.	-1.189	0.014	0	0.014
24	*negative*	.	.	.	.	-0.814	0.025	0	0.078
29	*negative*	.	.	.	.	-0.842	0.04	0	0.039
34	*negative*	.	.	.	.	-0.77	0.038	0	0.094
40	*negative*	.	.	.	.	-0.781	0.037	0	0.042

Table S4.1 Continued…

		PAML		*MEME*		*FUBAR*		*FEL*	
Amino Acid	***Selection***	**P[ω >1]**	**ω**	**LRT**	**p-value**	**β - α**	**P[α < β]**	**ω**	**p-value**
44	*negative*	.	.	.	.	-0.893	0.02	0	0.062
45	*negative*	.	.	.	.	-0.843	0.04	0	0.039
47	*negative*	.	.	.	.	-1.159	0.003	0	0.004
48	*negative*	.	.	.	.	-0.966	0.017	0	0.056
50	*negative*	.	.	.	.	-1.098	0.015	0	0.018
55	*negative*	.	.	.	.	-1.361	0.006	0	0.006
61	*negative*	.	.	.	.	-0.934	0.025	0	0.022
64	*negative*	.	.	.	.	-1.36	0.005	0	0.004
69	*negative*	.	.	.	.	-1.231	0.014	0	0.01
70	*negative*	.	.	.	.	-1.506	0.002	0	0.001
72	*negative*	.	.	.	.	-2.685	0.001	0	0
73	*negative*	.	.	.	.	-1.066	0.007	0	0.012
82	*negative*	.	.	.	.	-1.993	0.005	0	0.002
84	*negative*	.	.	.	.	-1.139	0.003	0	0.005
85	*negative*	.	.	.	.	-1.043	0.017	0	0.014
88	*negative*	.	.	.	.	-0.882	0.038	0	0.036
92	*negative*	.	.	.	.	-0.845	0.021	0	0.036
95	*negative*	.	.	.	.	-1.046	0.01	0	0.015
100	*negative*	.	.	.	.	-1.074	0.01	0	0.016
104	*negative*	.	.	.	.	-0.777	0.037	0	0.095
106	*negative*	.	.	.	.	-0.795	0.038	0	0.094
119	*negative*	.	.	.	.	-2.186	0.001	0	0
122	*negative*	.	.	.	.	-0.952	0.018	0	0.026

Table S4.1 Continued…

		PAML		*MEME*		*FUBAR*		*FEL*	
Amino Acid	***Selection***	**P[ω >1]**	**ω**	**LRT**	**p-value**	**β - α**	**P[α < β]**	**ω**	**p-value**
124	*negative*	.	.	.	.	-1.076	0.007	0	0.008
128	*negative*	.	.	.	.	-1.262	0.003	0	0.003
131	*negative*	.	.	.	.	-1.349	0.013	0	0.012
141	*negative*	.	.	.	.	-8.654	0	0	0
145	*negative*	.	.	.	.	-1.089	0.009	0	0.012
147	*negative*	.	.	.	.	-4.025	0	0	0
150	*negative*	.	.	.	.	-0.857	0.035	0	0.084
164	*negative*	.	.	.	.	-1.576	0.004	0	0.002
167	*negative*	.	.	.	.	-1.731	0.002	0	0.001
171	*negative*	.	.	.	.	-1.089	0.009	0	0.012
176	*negative*	.	.	.	.	-1.939	0.002	0	0.001
186	*negative*	.	.	.	.	-1.237	0.008	0	0.006
189	*negative*	.	.	.	.	-1.007	0.019	0	0.016
191	*negative*	.	.	.	.	-1.431	0.011	0	0.006
195	*negative*	.	.	.	.	-0.886	0.036	0	0.081
210	*negative*	.	.	.	.	-1.426	0.002	0	0.002
230	*negative*	.	.	.	.	-0.887	0.036	0	0.081
238	*negative*	.	.	.	.	-1.325	0.01	0	0.007
246	*negative*	.	.	.	.	-0.77	0.038	0	0.094
309	*negative*	.	.	.	.	-0.991	0.013	0	0.014
331	*negative*	.	.	.	.	-1.602	0.004	0	0.002
335	*negative*	.	.	.	.	-1.263	0.006	0	0.008
346	*negative*	.	.	.	.	-0.823	0.021	0	0.074

Table S4.1 Continued…

		PAML		*MEME*		*FUBAR*		*FEL*	
Amino Acid	***Selection***	**P[ω >1]**	**ω**	**LRT**	**p-value**	**β - α**	**P[α < β]**	**ω**	**p-value**
347	*negative*	.	.	.	.	-1.411	0.005	0	0.006
357	*negative*	.	.	.	.	-0.887	0.026	0	0.025
363	*negative*	.	.	.	.	-0.848	0.014	0	0.059
365	*negative*	.	.	.	.	-0.931	0.024	0	0.064
368	*negative*	.	.	.	.	-1.015	0.017	0	0.021
373	*negative*	.	.	.	.	-1.157	0.008	0	0.01
374	*negative*	.	.	.	.	-1.166	0.012	0	0.013
376	*negative*	.	.	.	.	-1.351	0.001	0	0.001
377	*negative*	.	.	.	.	-1.213	0.044	0	0.021
380	*negative*	.	.	.	.	-0.81	0.033	0	0.033
383	*negative*	.	.	.	.	-0.662	0.056	0	0.068
390	*negative*	.	.	.	.	-0.991	0.024	0	0.019
399	*negative*	.	.	.	.	-1.059	0.015	0	0.017
405	*negative*	.	.	.	.	-2.454	0	0	0
407	*negative*	.	.	.	.	-1.119	0.019	0	0.019
409	*negative*	.	.	.	.	-1.485	0.005	0	0.005
410	*negative*	.	.	.	.	-1.704	0.005	0	0.002
411	*negative*	.	.	.	.	-1.03	0.011	0	0.018
413	*negative*	.	.	.	.	-1.859	0.001	0	0
423	*negative*	.	.	.	.	-2.564	0.003	0	0.001
425	*negative*	.	.	.	.	-0.873	0.009	0	0.021
429	*negative*	.	.	.	.	-0.768	0.03	0	0.05
432	*negative*	.	.	.	.	-1.641	0.008	0	0.005

Table S4.1 Continued…

		PAML		*MEME*		*FUBAR*		*FEL*	
Amino Acid	***Selection***	**P[ω >1]**	**ω**	**LRT**	**p-value**	**β - α**	**P[α < β]**	**ω**	**p-value**
435	*negative*	.	.	.	.	-0.793	0.02	0	0.083
439	*negative*	.	.	.	.	-2.671	0.003	0	0.001
448	*negative*	.	.	.	.	-0.787	0.037	0	0.093
454	*negative*	.	.	.	.	-2.929	0.001	0	0
457	*negative*	.	.	.	.	-0.781	0.035	0	0.036
462	*negative*	.	.	.	.	-2.704	0.001	0	0
464	*negative*	.	.	.	.	-1.291	0.01	0	0.008
475	*negative*	.	.	.	.	-0.976	0.013	0	0.015
478	*negative*	.	.	.	.	-2.207	0.001	0	0
484	*negative*	.	.	.	.	-1.333	0.003	0	0.005
486	*negative*	.	.	.	.	-1.57	0.003	0	0.002
504	*negative*	.	.	.	.	-3.454	0	0	0
507	*negative*	.	.	.	.	-0.966	0.012	0	0.049
509	*negative*	.	.	.	.	-2.525	0.002	0	0.001
529	*negative*	.	.	.	.	-1.825	0.003	0	0.001
532	*negative*	.	.	.	.	-1.236	0.009	0	0.009
536	*negative*	.	.	.	.	-1.427	0.005	0	0.006
543	*negative*	.	.	.	.	-1.563	0.005	0	0.003
545	*negative*	.	.	.	.	-1.148	0.016	0	0.017
562	*negative*	.	.	.	.	-1.242	0.006	0	0.008
565	*negative*	.	.	.	.	-0.767	0.024	0	0.091
576	*negative*	.	.	.	.	-0.82	0.015	0	0.032
577	*negative*	.	.	.	.	-0.949	0.017	0	0.057

Table S4.1 Continued…

		PAML		*MEME*		*FUBAR*		*FEL*	
Amino Acid	***Selection***	**P[ω >1]**	**ω**	**LRT**	**p-value**	**β - α**	**P[α < β]**	**ω**	**p-value**
587	*negative*	.	.	.	.	-4.308	0	0	0
600	*negative*	.	.	.	.	-1.652	0.006	0	0.004

Table S4.2: List of amino acids selected for 16 site TLR4 library. Also included are the corresponding NNN overlap extension mutagenesis primers used to generate the library.

Amino Acid	Reverse Primer	Forward Primer
298	ataagtcaataatatcnnngaggtagtagtctaag	cttagactactacctcnnngatattattgacttat
299	taaataagtcaataatnnnatcgaggtagtagtct	agactactacctcgatnnnattattgacttattta
300	aattaaataagtcaatnnnatcatcgaggtagtag	ctactacctcgatgatnnnattgacttatttaatt
301	aacaattaaataagtcnnnaatatcatcgaggtag	ctacctcgatgatattnnngacttatttaattgtt
302	tcaaacaattaaataannnaataatatcatcgagg	cctcgatgatattattnnnttatttaattgtttga
305	aaacatttgtcaaacannnaaataagtcaataata	tattattgacttatttnnntgtttgacaaatgttt
309	gggaaaatgaagaaacnnntgtcaaacaattaaat	atttaattgtttgacannngtttcttcattttccc
319	cttttacccttttcaatnnncacactcaccagggaa	ttccctggtgagtgtgnnnattgaaagggtaaaag
321	aaaagtcttttaccctnnnaatagtcacactcacc	ggtgagtgtgactattnnnagggtaaaagactttt
322	aagaaaagtcttttacnnnttcaatagtcacactc	gagtgtgactattgaannngtaaaagacttttctt
324	aattataagaaaagtcnnntacccttttcaatagtc	gactattgaaagggtannngacttttcttataatt
325	cgaaattataagaaaannnttttacccttttcaata	tattgaaagggtaaaannnttttcttataatttcg
327	gccatccgaaattatannnaaagtcttttaccctt	aagggtaaaagactttnnntataatttcggatggc
344	gtttcaatgtgggaaannntccaaatttacagtta	taactgtaaatttggannnttttcccacattgaaac
347	gagatttgagtttcaannngggaaactgtccaaat	atttggacagtttcccnnnttgaaactcaaatctc
349	ttttgagagatttgagnnncaatgtgggaaactgt	acagtttcccacattgnnnctcaaatctctcaaaa

REFERENCES CITED

1. Pahwa R, Goyal A, Jialal I Chronic Inflammation. In: StatPearls. Treasure Island (FL): StatPearls Publishing; 2024. Available from: http://www.ncbi.nlm.nih.gov/books/NBK493173/

2. Gong T, Liu L, Jiang W, Zhou R (2020) DAMP-sensing receptors in sterile inflammation and inflammatory diseases. Nat Rev Immunol 20:95–112.

3. Poltorak A, He X, Smirnova I, Liu MY, Van Huffel C, Du X, Birdwell D, Alejos E, Silva M, Galanos C, et al. (1998) Defective LPS signaling in C3H/HeJ and C57BL/10ScCr mice: mutations in Tlr4 gene. Science 282:2085–2088.

4. Rudd KE, Johnson SC, Agesa KM, Shackelford KA, Tsoi D, Kievlan DR, Colombara DV, Ikuta KS, Kissoon N, Finfer S, et al. (2020) Global, regional, and national sepsis incidence and mortality, 1990–2017: analysis for the Global Burden of Disease Study. The Lancet 395:200–211.

5. Anon (2011) Nobel Prize to immunology. Nat Rev Immunol 11:714–714.

6. Romerio A, Peri F (2020) Increasing the Chemical Variety of Small-Molecule-Based TLR4 Modulators: An Overview. Front Immunol 11:1210.

7. Raetz CRH, Whitfield C (2002) Lipopolysaccharide Endotoxins. Annu Rev Biochem 71:635–700.

8. Cunningham MD, Shapiro RA, Seachord C, Ratcliffe K, Cassiano L, Darveau RP (2000) CD14 Employs Hydrophilic Regions to “Capture” Lipopolysaccharides. The Journal of Immunology 164:3255–3263.

9. da Silva Correia J, Soldau K, Christen U, Tobias PS, Ulevitch RJ (2001) 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 276:21129–21135.

10. Jagtap P, Prasad P, Pateria A, Deshmukh SD, Gupta S (2020) A Single Step in vitro Bioassay Mimicking TLR4-LPS Pathway and the Role of MD2 and CD14 Coreceptors. Front Immunol [Internet] 11. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6992608/

11. Park BS, Song DH, Kim HM, Choi B-S, Lee H, Lee J-O (2009) The structural basis of lipopolysaccharide recognition by the TLR4-MD-2 complex. Nature 458:1191–1195.

12. Nagai Y, Akashi S, Nagafuku M, Ogata M, Iwakura Y, Akira S, Kitamura T, Kosugi A, Kimoto M, Miyake K (2002) Essential role of MD-2 in LPS responsiveness and TLR4 distribution. Nat Immunol 3:667–672.

13. Riva M, Källberg E, Björk P, Hancz D, Vogl T, Roth J, Ivars F, Leanderson T (2012) Induction of nuclear factor-κB responses by the S100A9 protein is Toll-like receptor-4-dependent. Immunology 137:172–182.

14. Björk P, Björk A, Vogl T, Stenström M, Liberg D, Olsson A, Roth J, Ivars F, Leanderson T (2009) Identification of Human S100A9 as a Novel Target for Treatment of Autoimmune Disease via Binding to Quinoline-3-Carboxamides. PLoS Biol [Internet] 7. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2671563/

15. Chen B, Miller AL, Rebelatto M, Brewah Y, Rowe DC, Clarke L, Czapiga M, Rosenthal K, Imamichi T, Chen Y, et al. (2015) S100A9 Induced Inflammatory Responses Are Mediated by Distinct Damage Associated Molecular Patterns (DAMP) Receptors In Vitro and In Vivo. PLOS ONE 10:e0115828.

16. Berntzen HB, Fagerhol MK (1990) L1, a major granulocyte protein; isolation of high quantities of its subunits. Scandinavian Journal of Clinical and Laboratory Investigation 50:769–774.

17. Zygiel EM, Nolan EM (2018) Transition Metal Sequestration by the Host-Defense Protein Calprotectin. Annu Rev Biochem 87:621–643.

18. Ryckman C, Vandal K, Rouleau P, Talbot M, Tessier PA (2003) Proinflammatory activities of S100: proteins S100A8, S100A9, and S100A8/A9 induce neutrophil chemotaxis and adhesion. J Immunol 170:3233–3242.

19. Averill MM, Kerkhoff C, Bornfeldt KE (2012) S100A8 and S100A9 in Cardiovascular Biology and Disease. Arteriosclerosis, Thrombosis, and Vascular Biology 32:223–229.

20. Ehrchen JM, Sunderkötter C, Foell D, Vogl T, Roth J (2009) The endogenous Toll–like receptor 4 agonist S100A8/S100A9 (calprotectin) as innate amplifier of infection, autoimmunity, and cancer. Journal of Leukocyte Biology 86:557–566.

21. Markowitz J, Carson WE (2013) Review of S100A9 biology and its role in cancer. Biochimica et Biophysica Acta (BBA) - Reviews on Cancer 1835:100–109.

22. Hermani A, Hess J, Servi BD, Medunjanin S, Grobholz R, Trojan L, Angel P, Mayer D (2005) Calcium-Binding Proteins S100A8 and S100A9 as Novel Diagnostic Markers in Human Prostate Cancer. Clin Cancer Res 11:5146–5152.

23. Wang C, Iashchishyn IA, Pansieri J, Nyström S, Klementieva O, Kara J, Horvath I, Moskalenko R, Rofougaran R, Gouras G, et al. (2018) S100A9-Driven Amyloid-Neuroinflammatory Cascade in Traumatic Brain Injury as a Precursor State for Alzheimer's Disease. Scientific Reports 8:12836.

24. Zhang C, Liu Y, Gilthorpe J, van der Maarel JRC (2012) MRP14 (S100A9) Protein Interacts with Alzheimer Beta-Amyloid Peptide and Induces Its Fibrillization. PLoS One [Internet] 7. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3310843/

25. Loes AN, Bridgham JT, Harms MJ (2018) Coevolution of the Toll-Like Receptor 4 Complex with Calgranulins and Lipopolysaccharide. Front Immunol [Internet] 9. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5826337/

26. He Z, Riva M, Björk P, Swärd K, Mörgelin M, Leanderson T, Ivars F (2016) CD14 Is a Co-Receptor for TLR4 in the S100A9-Induced Pro-Inflammatory Response in Monocytes. PLoS One [Internet] 11. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4881898/

27. Pelletier M, Simard J-C, Girard D, Tessier PA (2018) Quinoline-3-carboxamides such as tasquinimod are not specific inhibitors of S100A9. Blood Adv 2:1170–1171.

28. Namath A, Patterson AJ (2011) Genetic Polymorphisms in Sepsis. Critical Care Nursing Clinics of North America 23:181–202.

29. Schröder NW, Schumann RR (2005) Single nucleotide polymorphisms of Toll-like receptors and susceptibility to infectious disease. The Lancet Infectious Diseases 5:156–164.

30. Netea MG, Wijmenga C, O'Neill LAJ (2012) Genetic variation in Toll-like receptors and disease susceptibility. Nat Immunol 13:535–542.

31. Loes AN, Hinman MN, Farnsworth DR, Miller AC, Guillemin K, Harms MJ (2021) Identification and Characterization of Zebrafish Tlr4 Coreceptor Md-2. The Journal of Immunology 206:1046–1057.

32. Yang Z (2007) PAML 4: Phylogenetic Analysis by Maximum Likelihood. Molecular Biology and Evolution 24:1586–1591.

33. Pond SLK, Frost SDW, Muse SV (2005) HyPhy: hypothesis testing using phylogenies. Bioinformatics 21:676–679.

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

35. Hochberg GKA, Thornton JW (2017) Reconstructing Ancient Proteins to Understand the Causes of Structure and Function. Annu Rev Biophys 46:247–269.

36. Anderson JA, Loes AN, Waddell GL, Harms MJ (2019) Tracing the evolution of novel features of human Toll-like receptor 4. Protein Science 28:1350–1358.

37. Edgeworth J, Gorman M, Bennett R, Freemont P, Hogg N (1991) Identification of p8,14 as a highly abundant heterodimeric calcium binding protein complex of myeloid cells. J Biol Chem 266:7706–7713.

38. Vogl T, Tenbrock K, Ludwig S, Leukert N, Ehrhardt C, van Zoelen MAD, Nacken W, Foell D, van der Poll T, Sorg C, et al. (2007) Mrp8 and Mrp14 are endogenous activators of Toll-like receptor 4, promoting lethal, endotoxin-induced shock. Nature Medicine 13:1042–1049.

39. Schiopu A, Cotoi OS (2013) S100A8 and S100A9: DAMPs at the crossroads between innate immunity, traditional risk factors, and cardiovascular disease. Mediators Inflamm 2013:828354.

40. Hoshino K, Takeuchi O, Kawai T, Sanjo H, Ogawa T, Takeda Y, Takeda K, Akira S (1999) Cutting Edge: Toll-Like Receptor 4 (TLR4)-Deficient Mice Are Hyporesponsive to

Lipopolysaccharide: Evidence for TLR4 as the Lps Gene Product. The Journal of Immunology 162:3749–3752.

41. Chisholm LO, Jaeger NM, Murawsky HE, Harms MJ (2024) S100A9 interacts with a dynamic region on CD14 to activate Toll-like receptor 4. :2024.05.15.594416. Available from: https://www.biorxiv.org/content/10.1101/2024.05.15.594416v1

42. Erridge C (2010) Endogenous ligands of TLR2 and TLR4: agonists or assistants? J Leukoc Biol 87:989–999.

43. Unno M, Kawasaki T, Takahara H, Heizmann CW, Kizawa K (2011) Refined Crystal Structures of Human Ca2+/Zn2+-Binding S100A3 Protein Characterized by Two Disulfide Bridges. Journal of Molecular Biology 408:477–490.

44. Riva M, He Z, Källberg E, Ivars F, Leanderson T (2013) Human S100A9 Protein Is Stabilized by Inflammatory Stimuli via the Formation of Proteolytically-Resistant Homodimers. PLOS ONE 8:e61832.

45. Nacken W, Kerkhoff C (2007) The hetero-oligomeric complex of the S100A8/S100A9 protein is extremely protease resistant. FEBS Lett 581:5127–5130.

46. Chow JC, Young DW, Golenbock DT, Christ WJ, Gusovsky F (1999) Toll-like Receptor-4 Mediates Lipopolysaccharide-induced Signal Transduction *. Journal of Biological Chemistry 274:10689–10692.

47. Rallabhandi P, Bell J, Boukhvalova MS, Medvedev A, Lorenz E, Arditi M, Hemming VG, Blanco JCG, Segal DM, Vogel SN (2006) Analysis of TLR4 Polymorphic Variants: New Insights into TLR4/MD-2/CD14 Stoichiometry, Structure, and Signaling1. The Journal of Immunology 177:322–332.

48. Harman JL, Loes AN, Warren GD, Heaphy MC, Lampi KJ, Harms MJ (2020) Evolution of multifunctionality through a pleiotropic substitution in the innate immune protein S100A9 Garrett WS, Laub MT, editors. eLife 9:e54100.

49. Lominadze G, Rane MJ, Merchant M, Cai J, Ward RA, McLeish KR (2005) Myeloid-Related Protein-14 Is a p38 MAPK Substrate in Human Neutrophils1. The Journal of Immunology 174:7257–7267.

50. Edgeworth J, Freemont P, Hogg N (1989) Ionomycin-regulated phosphorylation of the myeloid calcium-binding protein p14. Nature 342:189–192.

51. Vogl T, Ludwig S, Goebeler M, Strey A, Thorey IS, Reichelt R, Foell D, Gerke V, Manitz MP, Nacken W, et al. (2004) MRP8 and MRP14 control microtubule reorganization during transendothelial migration of phagocytes. Blood 104:4260–4268.

52. Schenten V, Plançon S, Jung N, Hann J, Bueb J-L, Bréchard S, Tschirhart EJ, Tolle F (2018) Secretion of the Phosphorylated Form of S100A9 from Neutrophils Is Essential for the Proinflammatory Functions of Extracellular S100A8/A9. Front Immunol 9:447.

53. Vogl T, Leukert N, Barczyk K, Strupat K, Roth J (2006) Biophysical characterization of S100A8 and S100A9 in the absence and presence of bivalent cations. Biochimica et Biophysica Acta (BBA) - Molecular Cell Research 1763:1298–1306.

54. Wheeler LC, Donor MT, Prell JS, Harms MJ (2016) Multiple Evolutionary Origins of Ubiquitous Cu2+ and Zn2+ Binding in the S100 Protein Family. PLOS ONE 11:e0164740.

55. Baronaitė I, Šulskis D, Kopūstas A, Tutkus M, Smirnovas V (2024) Formation of Calprotectin Inhibits Amyloid Aggregation of S100A8 and S100A9 Proteins. ACS Chem. Neurosci. 15:1915–1925.

56. Iashchishyn IA, Sulskis D, Nguyen Ngoc M, Smirnovas V, Morozova-Roche LA (2017) Finke–Watzky Two-Step Nucleation–Autocatalysis Model of S100A9 Amyloid Formation: Protein Misfolding as "Nucleation" Event. ACS Chem. Neurosci. 8:2152–2158.

57. Martino L, He Y, Hands-Taylor KLD, Valentine ER, Kelly G, Giancola C, Conte MR (2009) The interaction of the Escherichia coli protein SlyD with nickel ions illuminates the mechanism of regulation of its peptidyl-prolyl isomerase activity. The FEBS Journal 276:4529–4544.

58. Robichon C, Luo J, Causey TB, Benner JS, Samuelson JC (2011) Engineering Escherichia coli BL21(DE3) Derivative Strains To Minimize E. coli Protein Contamination after Purification by Immobilized Metal Affinity Chromatography. Applied and Environmental Microbiology 77:4634–4646.

59. Hansen SD, Huang WYC, Lee YK, Bieling P, Christensen SM, Groves JT (2019) Stochastic geometry sensing and polarization in a lipid kinase–phosphatase competitive reaction. Proceedings of the National Academy of Sciences 116:15013–15022.

60. Harman JL, Reardon PN, Costello SM, Warren GD, Phillips SR, Connor PJ, Marqusee S, Harms MJ (2022) Evolution avoids a pathological stabilizing interaction in the immune protein S100A9. Proceedings of the National Academy of Sciences 119:e2208029119.

61. Averill Michelle M., Kerkhoff Claus, Bornfeldt Karin E. (2012) S100A8 and S100A9 in Cardiovascular Biology and Disease. Arteriosclerosis, Thrombosis, and Vascular Biology 32:223–229.

62. Zhang X, Wei L, Wang J, Qin Z, Wang J, Lu Y, Zheng X, Peng Q, Ye Q, Ai F, et al. (2017) Suppression Colitis and Colitis-Associated Colon Cancer by Anti-S100a9 Antibody in Mice. Front Immunol 8:1774.

63. Galloway SM, Raetz CR (1990) A mutant of Escherichia coli defective in the first step of endotoxin biosynthesis. Journal of Biological Chemistry 265:6394–6402.

64. Huber RG, Berglund NA, Kargas V, Marzinek JK, Holdbrook DA, Khalid S, Piggot TJ, Schmidtchen A, Bond PJ (2018) A Thermodynamic Funnel Drives Bacterial Lipopolysaccharide Transfer in the TLR4 Pathway. Structure 26:1151-1161.e4.

65. Wright SD, Ramos RA, Tobias PS, Ulevitch RJ, Mathison JC (1990) CD14, a Receptor for Complexes of Lipopolysaccharide (LPS) and LPS Binding Protein. Science 249:1431–1433.

66. Kim J-I, Lee CJ, Jin MS, Lee C-H, Paik S-G, Lee H, Lee J-O (2005) Crystal Structure of CD14 and Its Implications for Lipopolysaccharide Signaling. J. Biol. Chem. 280:11347–11351.

67. Kelley SL, Lukk T, Nair SK, Tapping RI (2013) The Crystal Structure of Human Soluble CD14 Reveals a Bent Solenoid with a Hydrophobic Amino-Terminal Pocket. The Journal of Immunology 190:1304–1311.

68. Ryu J-K, Kim SJ, Rah S-H, Kang JI, Jung HE, Lee D, Lee HK, Lee J-O, Park BS, Yoon T-Y, et al. (2017) Reconstruction of LPS Transfer Cascade Reveals Structural Determinants within LBP, CD14, and TLR4-MD2 for Efficient LPS Recognition and Transfer. Immunity 46:38–50.

69. Hailman E, Lichenstein HS, Wurfel MM, Miller DS, Johnson DA, Kelley M, Busse LA, Zukowski MM, Wright SD (1994) Lipopolysaccharide (LPS)-binding protein accelerates the binding of LPS to CD14. Journal of Experimental Medicine 179:269–277.

70. Haziot A, Chen S, Ferrero E, Low MG, Silber R, Goyert SM (1988) The monocyte differentiation antigen, CD14, is anchored to the cell membrane by a phosphatidylinositol linkage. J Immunol 141:547–552.

71. Simmons D, Tan S, Tenen D, Nicholson-Weller A, Seed B (1989) Monocyte antigen CD14 is a phospholipid anchored membrane protein. Blood 73:284–289.

72. Zanoni I, Ostuni R, Marek LR, Barresi S, Barbalat R, Barton GM, Granucci F, Kagan JC (2011) CD14 Controls the LPS-Induced Endocytosis of Toll-like Receptor 4. Cell 147:868–880.

73. Ciesielska A, Matyjek M, Kwiatkowska K (2021) TLR4 and CD14 trafficking and its influence on LPS-induced pro-inflammatory signaling. Cell Mol Life Sci 78:1233–1261.

74. Jiang Z, Georgel P, Du X, Shamel L, Sovath S, Mudd S, Huber M, Kalis C, Keck S, Galanos C, et al. (2005) CD14 is required for MyD88-independent LPS signaling. Nat Immunol 6:565–570.

75. Itou H, Yao M, Fujita I, Watanabe N, Suzuki M, Nishihira J, Tanaka I (2002) The crystal structure of human MRP14 (S100A9), a Ca2+-dependent regulator protein in inflammatory process11Edited by R. Huber. Journal of Molecular Biology 316:265–276.

76. PubChem GPI-anchor amidated glycine. Available from: https://pubchem.ncbi.nlm.nih.gov/compound/145996624

77. Rouillard AD, Gundersen GW, Fernandez NF, Wang Z, Monteiro CD, McDermott MG, Ma'ayan A (2016) The harmonizome: a collection of processed datasets gathered to serve and mine knowledge about genes and proteins. Database 2016:baw100.

78. Yu B, Wright SD (1996) Catalytic properties of lipopolysaccharide (LPS) binding protein. Transfer of LPS to soluble CD14. J Biol Chem 271:4100–4105.

79. Lien E, Aukrust P, Sundan A, Müller F, Frøland SS, Espevik T (1998) Elevated levels of serum-soluble CD14 in human immunodeficiency virus type 1 (HIV-1) infection: correlation to disease progression and clinical events. Blood 92:2084–2092.

80. Sandler NG, Wand H, Roque A, Law M, Nason MC, Nixon DE, Pedersen C, Ruxrungtham K, Lewin SR, Emery S, et al. (2011) Plasma Levels of Soluble CD14 Independently Predict Mortality in HIV Infection. J Infect Dis 203:780–790.

81. Gonzalez-Quintela A, Alonso M, Campos J, Vizcaino L, Loidi L, Gude F (2013) Determinants of serum concentrations of lipopolysaccharide-binding protein (LBP) in the adult population: the role of obesity. PLoS One 8:e54600.

82. Mukherjee R, Kanti Barman P, Kumar Thatoi P, Tripathy R, Kumar Das B, Ravindran B (2015) Non-Classical monocytes display inflammatory features: Validation in Sepsis and Systemic Lupus Erythematous. Sci Rep 5:13886.

83. Frey EA, Miller DS, Jahr TG, Sundan A, Bazil V, Espevik T, Finlay BB, Wright SD (1992) Soluble CD14 participates in the response of cells to lipopolysaccharide. Journal of Experimental Medicine 176:1665–1671.

84. Juan TS-C, Kelley MJ, Johnson DA, Busse LA, Hailman E, Wright SD, Lichenstein HS (1995) Soluble CD14 Truncated at Amino Acid 152 Binds Lipopolysaccharide (LPS) and Enables Cellular Response to LPS (∗). Journal of Biological Chemistry 270:1382–1387.

85. Kawai T, Akira S (2010) The role of pattern-recognition receptors in innate immunity: update on Toll-like receptors. Nat Immunol 11:373–384.

86. Petherick KJ, Conway OJL, Mpamhanga C, Osborne SA, Kamal A, Saxty B, Ganley IG (2015) Pharmacological Inhibition of ULK1 Kinase Blocks Mammalian Target of Rapamycin (mTOR)-dependent Autophagy. J Biol Chem 290:11376–11383.

87. Clark K, Peggie M, Plater L, Sorcek RJ, Young ERR, Madwed JB, Hough J, McIver EG, Cohen P (2011) Novel cross-talk within the IKK family controls innate immunity. Biochem J 434:93–104.

88. Xie L, Jiang F-C, Zhang L-M, He W-T, Liu J-H, Li M-Q, Zhang X, Xing S, Guo H, Zhou P (2016) Targeting of MyD88 Homodimerization by Novel Synthetic Inhibitor TJ-M2010-5 in Preventing Colitis-Associated Colorectal Cancer. JNCI: Journal of the National Cancer Institute 108:djv364.

89. Miao Y, Ding Z, Zou Z, Yang Y, Yang M, Zhang X, Li Z, Zhou L, Zhang L, Zhang X, et al. (2020) Inhibition of MyD88 by a novel inhibitor reverses two-thirds of the infarct area in myocardial ischemia and reperfusion injury. Am J Transl Res 12:5151–5169.

90. Juan TS-C, Hailman E, Kelley MJ, Busse LA, Davy E, Empig CJ, Narhi LO, Wright SD, Lichenstein HS (1995) Identification of a Lipopolysaccharide Binding Domain in CD14 between Amino Acids 57 and 64 (∗). Journal of Biological Chemistry 270:5219–5224.

91. Mirdita M, Schütze K, Moriwaki Y, Heo L, Ovchinnikov S, Steinegger M (2022) ColabFold: making protein folding accessible to all. Nat Methods 19:679–682.

92. Evans R, O'Neill M, Pritzel A, Antropova N, Senior A, Green T, Žídek A, Bates R, Blackwell S, Yim J, et al. (2022) Protein complex prediction with AlphaFold-Multimer. :2021.10.04.463034. Available from: https://www.biorxiv.org/content/10.1101/2021.10.04.463034v2

93. Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, Tunyasuvunakool K, Bates R, Žídek A, Potapenko A, et al. (2021) Highly accurate protein structure prediction with AlphaFold. Nature 596:583–589.

94. Bonhomme D, Santecchia I, Vernel-Pauillac F, Caroff M, Germon P, Murray G, Adler B, Boneca IG, Werts C (2020) Leptospiral LPS escapes mouse TLR4 internalization and TRIF-associated antimicrobial responses through O antigen and associated lipoproteins. PLoS Pathog [Internet] 16. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7447051/

95. Tan Y, Zanoni I, Cullen TW, Goodman AL, Kagan JC (2015) Mechanisms of Toll-like receptor 4 endocytosis reveal a common immune-evasion strategy used by pathogenic and commensal bacteria. Immunity 43:909–922.

96. Bryant CE, Spring DR, Gangloff M, Gay NJ (2010) The molecular basis of the host response to lipopolysaccharide. Nat Rev Microbiol 8:8–14.

97. Wang C, Bradley P, Baker D (2007) Protein-protein docking with backbone flexibility. J Mol Biol 373:503–519.

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

99. Van Der Spoel D, Lindahl E, Hess B, Groenhof G, Mark AE, Berendsen HJC (2005) GROMACS: Fast, flexible, and free. Journal of Computational Chemistry 26:1701–1718.

100. MacKerell AD, Bashford D, Bellott M, Dunbrack RL, Evanseck JD, Field MJ, Fischer S, Gao J, Guo H, Ha S, et al. (1998) All-atom empirical potential for molecular modeling and dynamics studies of proteins. J Phys Chem B 102:3586–3616.

101. Lee J, Patel DS, Ståhle J, Park S-J, Kern NR, Kim S, Lee J, Cheng X, Valvano MA, Holst O, et al. (2019) CHARMM-GUI Membrane Builder for Complex Biological Membrane Simulations with Glycolipids and Lipoglycans. J. Chem. Theory Comput. 15:775–786.

102. Jo S, Kim T, Iyer VG, Im W (2008) CHARMM-GUI: A web-based graphical user interface for CHARMM. Journal of Computational Chemistry 29:1859–1865.

103. Parrinello M, Rahman A (1982) Strain fluctuations and elastic constants. The Journal of Chemical Physics 76:2662–2666.

104. Nosé S, Klein ML (1983) Constant pressure molecular dynamics for molecular systems. Molecular Physics 50:1055–1076.

105. Bussi G, Donadio D, Parrinello M (2007) Canonical sampling through velocity rescaling. The Journal of Chemical Physics 126:014101.

106. Hess B, Bekker H, Berendsen HJC, Fraaije JGEM (1997) LINCS: A linear constraint solver for molecular simulations. Journal of Computational Chemistry 18:1463–1472.

107. Hess B (2008) P-LINCS: A Parallel Linear Constraint Solver for Molecular Simulation. J. Chem. Theory Comput. 4:116–122.

108. Swope WC, Andersen HC, Berens PH, Wilson KR (1982) A computer simulation method for the calculation of equilibrium constants for the formation of physical clusters of molecules: Application to small water clusters. The Journal of Chemical Physics 76:637–649.

109. Darden T, York D, Pedersen L (1993) Particle mesh Ewald: An N·log(N) method for Ewald sums in large systems. The Journal of Chemical Physics 98:10089–10092.

110. Humphrey W, Dalke A, Schulten K (1996) VMD: Visual molecular dynamics. Journal of Molecular Graphics 14:33–38.

111. Michaud-Agrawal N, Denning EJ, Woolf TB, Beckstein O (2011) MDAnalysis: A Toolkit for the Analysis of Molecular Dynamics Simulations. J Comput Chem 32:2319–2327.

112. Gowers R, Linke M, Barnoud J, Reddy T, Melo M, Seyler SL, Domański J, Dotson D, Buchoux S, Kenney I, et al. MDAnalysis: A Python Package for the Rapid Analysis of Molecular Dynamics Simulations. In: ; 2016. pp. 98–105.

113. Schrödinger, LLC (2015) The PyMOL Molecular Graphics System, Version 1.8.

114. Mitternacht S (2016) FreeSASA: An open source C library for solvent accessible surface area calculations. Available from: https://f1000research.com/articles/5-189

115. Smith P, Ziolek RM, Gazzarrini E, Owen DM, Lorenz CD (2019) On the interaction of hyaluronic acid with synovial fluid lipid membranes. Phys Chem Chem Phys 21:9845–9857.

116. Theobald DL (2005) Rapid calculation of RMSDs using a quaternion-based characteristic polynomial. Acta Crystallogr A 61:478–480.

117. Liu P, Agrafiotis DK, Theobald DL (2010) Fast determination of the optimal rotational matrix for macromolecular superpositions. J Comput Chem 31:1561–1563.

118. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, Blondel M, Prettenhofer P, Weiss R, Dubourg V, et al. (2011) Scikit-learn: Machine Learning in Python. Journal of Machine Learning Research 12:2825–2830.

119. Rousseeuw PJ (1987) Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. Journal of Computational and Applied Mathematics 20:53–65.

120. Zindel J, Kubes P (2020) DAMPs, PAMPs, and LAMPs in Immunity and Sterile Inflammation. Annual Review of Pathology: Mechanisms of Disease 15:493–518.

121. Lu Y-C, Yeh W-C, Ohashi PS (2008) LPS/TLR4 signal transduction pathway. Cytokine 42:145–151.

122. Sironi M, Cagliani R, Forni D, Clerici M (2015) Evolutionary insights into host–pathogen interactions from mammalian sequence data. Nat Rev Genet 16:224–236.

123. Daugherty MD, Malik HS (2012) Rules of Engagement: Molecular Insights from Host-Virus Arms Races. Annu. Rev. Genet. 46:677–700.

124. Weaver S, Shank SD, Spielman SJ, Li M, Muse SV, Kosakovsky Pond SL (2018) Datamonkey 2.0: A Modern Web Application for Characterizing Selective and Other Evolutionary Processes. Molecular Biology and Evolution 35:773–777.

125. Delport W, Poon AFY, Frost SDW, Kosakovsky Pond SL (2010) Datamonkey 2010: a suite of phylogenetic analysis tools for evolutionary biology. Bioinformatics 26:2455–2457.

126. Pond SLK, Frost SDW (2005) Datamonkey: rapid detection of selective pressure on individual sites of codon alignments. Bioinformatics 21:2531–2533.

127. Kosakovsky Pond SL, Poon AFY, Velazquez R, Weaver S, Hepler NL, Murrell B, Shank SD, Magalis BR, Bouvier D, Nekrutenko A, et al. (2020) HyPhy 2.5—A Customizable Platform for Evolutionary Hypothesis Testing Using Phylogenies. Molecular Biology and Evolution 37:295–299.

128. Harman JL, Loes AN, Warren GD, Heaphy MC, Lampi KJ, Harms MJ (2020) Evolution of multifunctionality through a pleiotropic substitution in the innate immune protein S100A9 Garrett WS, Laub MT, editors. eLife 9:e54100.

129. Tschirren B, Råberg L, Westerdahl H (2011) Signatures of selection acting on the innate immunity gene Toll-like receptor 2 (TLR2) during the evolutionary history of rodents. Journal of Evolutionary Biology 24:1232–1240.

130. Kuri P, Ellwanger K, Kufer TA, Leptin M, Bajoghli B (2017) A high-sensitivity bi-directional reporter to monitor NF-κB activity in cell culture and zebrafish in real time. J. Cell. Sci. 130:648–657.

131. Fan F, Wood KV (2007) Bioluminescent Assays for High-Throughput Screening. ASSAY and Drug Development Technologies 5:127–136.

132. Bloom JD (2014) An Experimentally Determined Evolutionary Model Dramatically Improves Phylogenetic Fit. Mol Biol Evol 31:1956–1978.

133. Onuchic JN, Luthey-Schulten Z, Wolynes PG (1997) Theory of protein folding: the energy landscape perspective. Annual review of physical chemistry 48:545–600.

134. Weber G Energetics of ligand binding to proteins. In: Advances in protein chemistry. Vol. 29. Elsevier; 1975. pp. 1–83.

135. Bahar I, Lezon TR, Yang L-W, Eyal E (2010) Global dynamics of proteins: bridging between structure and function. Annual review of biophysics 39:23–42.

136. Motlagh HN, Wrabl JO, Li J, Hilser VJ (2014) The ensemble nature of allostery. Nature 508:331–339.

137. Wei G, Xi W, Nussinov R, Ma B (2016) Protein Ensembles: How Does Nature Harness Thermodynamic Fluctuations for Life? The Diverse Functional Roles of Conformational Ensembles in the Cell. Chem. Rev. 116:6516–6551.

138. Corbella M, Pinto GP, Kamerlin SCL (2023) Loop dynamics and the evolution of enzyme activity. Nat Rev Chem:1–12.

139. Tokuriki N, Tawfik DS (2009) Protein Dynamism and Evolvability. Science 324:203–207.

140. Alexander PA, He Y, Chen Y, Orban J, Bryan PN (2009) A minimal sequence code for switching protein structure and function. Proceedings of the National Academy of Sciences 106:21149–21154.

141. He Y, Chen Y, Alexander PA, Bryan PN, Orban J (2012) Mutational Tipping Points for Switching Protein Folds and Functions. Structure 20:283–291.

142. Sikosek T, Krobath H, Chan HS (2016) Theoretical Insights into the Biophysics of Protein Bi-stability and Evolutionary Switches. PLOS Computational Biology 12:e1004960.

143. Damry AM, Jackson CJ (2021) The evolution and engineering of enzyme activity through tuning conformational landscapes. Protein Engineering, Design and Selection 34:gzab009.

144. Dishman AF, Tyler RC, Fox JC, Kleist AB, Prehoda KE, Babu MM, Peterson FC, Volkman BF (2021) Evolution of fold switching in a metamorphic protein. Science 371:86–90.

145. Gardner JM, Biler M, Risso VA, Sanchez-Ruiz JM, Kamerlin SCL (2020) Manipulating Conformational Dynamics To Repurpose Ancient Proteins for Modern Catalytic Functions. ACS Catal. 10:4863–4870.

146. Kaczmarski JA, Mahawaththa MC, Feintuch A, Clifton BE, Adams LA, Goldfarb D, Otting G, Jackson CJ (2020) Altered conformational sampling along an evolutionary trajectory changes the catalytic activity of an enzyme. Nat Commun 11:5945.

147. Nguyen V, Wilson C, Hoemberger M, Stiller JB, Agafonov RV, Kutter S, English J, Theobald DL, Kern D (2017) Evolutionary drivers of thermoadaptation in enzyme catalysis. Science 355:289–294.

148. Nixon CF, Lim SA, Sailer ZR, Zheludev IN, Gee CL, Kelch BA, Harms MJ, Marqusee S (2021) Exploring the Evolutionary History of Kinetic Stability in the α-Lytic Protease Family. Biochemistry 60:170–181.

149. Okafor CD, Pathak MC, Fagan CE, Bauer NC, Cole MF, Gaucher EA, Ortlund EA (2018) Structural and Dynamics Comparison of Thermostability in Ancient, Modern, and Consensus Elongation Factor Tus. Structure 26:118-129.e3.

150. Whitney DS, Volkman BF, Prehoda KE (2016) Evolution of a Protein Interaction Domain Family by Tuning Conformational Flexibility. J. Am. Chem. Soc. 138:15150–15156.

151. Harms MJ, Thornton JW (2010) Analyzing protein structure and function using ancestral gene reconstruction. Current Opinion in Structural Biology 20:360–366.

152. Hochberg GKA, Thornton JW (2017) Reconstructing Ancient Proteins to Understand the Causes of Structure and Function. Annu Rev Biophys 46:247–269.

153. Marks DS, Hopf TA, Sander C (2012) Protein structure prediction from sequence variation. Nat Biotechnol 30:1072–1080.

154. Rivoire O, Reynolds KA, Ranganathan R (2016) Evolution-Based Functional Decomposition of Proteins. PLOS Computational Biology 12:e1004817.

155. Fowler DM, Fields S (2014) Deep mutational scanning: a new style of protein science. Nature methods 11:801–807.

156. Otten R, Pádua RAP, Bunzel HA, Nguyen V, Pitsawong W, Patterson M, Sui S, Perry SL, Cohen AE, Hilvert D, et al. (2020) How directed evolution reshapes energy landscapes in enzymes to boost catalysis. Science 370:1442–1446.

157. Arenas M Methodologies for Microbial Ancestral Sequence Reconstruction. In: Luo H, editor. Environmental Microbial Evolution: Methods and Protocols. Methods in Molecular Biology. New York, NY: Springer US; 2022. pp. 283–303. Available from: https://doi.org/10.1007/978-1-0716-2691-7_14

158. Liberles DA Ancestral Sequence Reconstruction. OUP Oxford; 2007.

159. Merkl R, Sterner R (2016) Reconstruction of ancestral enzymes. Perspectives in Science 9:17–23.

160. Selberg AGA, Gaucher EA, Liberles DA (2021) Ancestral Sequence Reconstruction: From Chemical Paleogenetics to Maximum Likelihood Algorithms and Beyond. J Mol Evol 89:157–164.

161. Spence MA, Kaczmarski JA, Saunders JW, Jackson CJ (2021) Ancestral sequence reconstruction for protein engineers. Curr Opin Struct Biol 69:131–141.

162. Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, von Haeseler A, Lanfear R (2020) IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. Molecular Biology and Evolution 37:1530–1534.

163. Musil M, Khan RT, Beier A, Stourac J, Konegger H, Damborsky J, Bednar D (2021) FireProtASR: A Web Server for Fully Automated Ancestral Sequence Reconstruction. Briefings in Bioinformatics 22:bbaa337.

164. Orlandi KN, Phillips SR, Sailer ZR, Harman JL, Harms MJ (2023) Topiary: Pruning the manual labor from ancestral sequence reconstruction. Protein Science 32:e4551.

165. Linus Pauling, Emile Zuckerlandl (1963) Chemical paleogenetics. Acta Chemica Scandinavica 17:S9–S16.

166. Yang Z (2007) PAML 4: Phylogenetic Analysis by Maximum Likelihood. Molecular Biology and Evolution 24:1586–1591.

167. Yang Z, Kumar S, Nei M (1995) A new method of inference of ancestral nucleotide and amino acid sequences. Genetics 141:1641–1650.

168. Bailleul G, Yang G, Nicoll CR, Mattevi A, Fraaije MW, Mascotti ML (2023) Evolution of enzyme functionality in the flavin-containing monooxygenases. Nat Commun 14:1042.

169. Busch F, Rajendran C, Heyn K, Schlee S, Merkl R, Sterner R (2016) Ancestral Tryptophan Synthase Reveals Functional Sophistication of Primordial Enzyme Complexes. Cell Chemical Biology 23:709–715.

170. Chiang C-H, Wymore T, Rodríguez Benítez A, Hussain A, Smith JL, Brooks CL, Narayan ARH (2023) Deciphering the evolution of flavin-dependent monooxygenase stereoselectivity using ancestral sequence reconstruction. Proceedings of the National Academy of Sciences 120:e2218248120.

171. Clifton BE, Kaczmarski JA, Carr PD, Gerth ML, Tokuriki N, Jackson CJ (2018) Evolution of cyclohexadienyl dehydratase from an ancestral solute-binding protein. Nat Chem Biol 14:542–547.

172. Furukawa R, Toma W, Yamazaki K, Akanuma S (2020) Ancestral sequence reconstruction produces thermally stable enzymes with mesophilic enzyme-like catalytic properties. Sci Rep 10:15493.

173. Garcia AK, McShea H, Kolaczkowski B, Kaçar B (2020) Reconstructing the evolutionary history of nitrogenases: Evidence for ancestral molybdenum-cofactor utilization. Geobiology 18:394–411.

174. Rauwerdink A, Lunzer M, Devamani T, Jones B, Mooney J, Zhang Z-J, Xu J-H, Kazlauskas RJ, Dean AM (2016) Evolution of a Catalytic Mechanism. Molecular Biology and Evolution 33:971–979.

175. Gaucher EA, Thomson JM, Burgan MF, Benner SA (2003) Inferring the palaeoenvironment of ancient bacteria on the basis of resurrected proteins. Nature 425:285–288.

176. Hart KM, Harms MJ, Schmidt BH, Elya C, Thornton JW, Marqusee S (2014) Thermodynamic System Drift in Protein Evolution. PLOS Biology 12:e1001994.

177. Malcolm BA, Wilson KP, Matthews BW, Kirsch JF, Wilson AC (1990) Ancestral lysozymes reconstructed, neutrality tested, and thermostability linked to hydrocarbon packing. Nature 345:86–89.

178. Lim SA, Bolin ER, Marqusee S (2018) Tracing a protein's folding pathway over evolutionary time using ancestral sequence reconstruction and hydrogen exchange. eLife 7:e38369.

179. Smock RG, Yadid I, Dym O, Clarke J, Tawfik DS (2016) De Novo Evolutionary Emergence of a Symmetrical Protein Is Shaped by Folding Constraints. Cell 164:476–486.

180. Sang D, Pinglay S, Wiewiora RP, Selvan ME, Lou HJ, Chodera JD, Turk BE, Gümüş ZH, Holt LJ (2019) Ancestral reconstruction reveals mechanisms of ERK regulatory evolution Kuriyan J, Kern D, editors. eLife 8:e38805.

181. East NJ, Clifton BE, Jackson CJ, Kaczmarski JA (2022) The role of oligomerization in the optimization of cyclohexadienyl dehydratase conformational dynamics and catalytic activity. Protein Science 31:e4510.

182. Holinski A, Heyn K, Merkl R, Sterner R (2017) Combining ancestral sequence reconstruction with protein design to identify an interface hotspot in a key metabolic enzyme complex. Proteins: Structure, Function, and Bioinformatics 85:312–321.

183. Baldwin MW, Toda Y, Nakagita T, O'Connell MJ, Klasing KC, Misaka T, Edwards SV, Liberles SD (2014) Evolution of sweet taste perception in hummingbirds by transformation of the ancestral umami receptor. Science 345:929–933.

184. Howard CJ, Hanson-Smith V, Kennedy KJ, Miller CJ, Lou HJ, Johnson AD, Turk BE, Holt LJ (2014) Ancestral resurrection reveals evolutionary mechanisms of kinase plasticity Ferrell J, editor. eLife 3:e04126.

185. McKeown AN, Bridgham JT, Anderson DW, Murphy MN, Ortlund EA, Thornton JW (2014) Evolution of DNA specificity in a transcription factor family produced a new gene regulatory module. Cell 159:58–68.

186. Siddiq MA, Hochberg GK, Thornton JW (2017) Evolution of protein specificity: insights from ancestral protein reconstruction. Current Opinion in Structural Biology 47:113–122.

187. Field SF, Matz MV (2010) Retracing Evolution of Red Fluorescence in GFP-Like Proteins from Faviina Corals. Molecular Biology and Evolution 27:225–233.

188. Kim H, Zou T, Modi C, Dörner K, Grunkemeyer TJ, Chen L, Fromme R, Matz MV, Ozkan SB, Wachter RM (2015) A Hinge Migration Mechanism Unlocks the Evolution of Green-to-Red Photoconversion in GFP-like Proteins. Structure 23:34–43.

189. Castiglione GM, Chang BS (2018) Functional trade-offs and environmental variation shaped ancient trajectories in the evolution of dim-light vision Lupas AN, Wittkopp PJ, Lupas AN, Carleton K, editors. eLife 7:e35957.

190. Van Nynatten A, Castiglione GM, de A. Gutierrez E, Lovejoy NR, Chang BSW (2021) Recreated Ancestral Opsin Associated with Marine to Freshwater Croaker Invasion Reveals Kinetic and Spectral Adaptation. Molecular Biology and Evolution 38:2076–2087.

191. Yokoyama S, Radlwimmer FB (2001) The molecular genetics and evolution of red and green color vision in vertebrates. Genetics 158:1697–1710.

192. Yokoyama S, Tada T, Zhang H, Britt L (2008) Elucidation of phenotypic adaptations: Molecular analyses of dim-light vision proteins in vertebrates. Proceedings of the National Academy of Sciences 105:13480–13485.

193. Chi PB, Liberles DA (2016) Selection on protein structure, interaction, and sequence. Protein Science 25:1168–1178.

194. Cortez LM, Morrison AJ, Garen CR, Patterson S, Uyesugi T, Petrosyan R, Sekar RV, Harms MJ, Woodside MT, Sim VL (2022) Probing the origin of prion protein misfolding via reconstruction of ancestral proteins. Protein Science 31:e4477.

195. Thornton JW (2004) Resurrecting ancient genes: experimental analysis of extinct molecules. Nat Rev Genet 5:366–375.

196. Le SQ, Gascuel O (2008) An Improved General Amino Acid Replacement Matrix. Molecular Biology and Evolution 25:1307–1320.

197. Yang Z (1996) Among-site rate variation and its impact on phylogenetic analyses. Trends in Ecology & Evolution 11:367–372.

198. Le SQ, Gascuel O (2010) Accounting for Solvent Accessibility and Secondary Structure in Protein Phylogenetics Is Clearly Beneficial. Systematic Biology 59:277–287.

199. Moshe A, Pupko T (2019) Ancestral sequence reconstruction: accounting for structural information by averaging over replacement matrices. Bioinformatics 35:2562–2568.

200. Groussin M, Hobbs JK, Szöllősi GJ, Gribaldo S, Arcus VL, Gouy M (2015) Toward More Accurate Ancestral Protein Genotype–Phenotype Reconstructions with the Use of Species Tree-Aware Gene Trees. Molecular Biology and Evolution 32:13–22.

201. Pagel M, Meade A, Barker D (2004) Bayesian Estimation of Ancestral Character States on Phylogenies. Systematic Biology 53:673–684.

202. Straub K, Merkl R Ancestral Sequence Reconstruction as a Tool for the Elucidation of a Stepwise Evolutionary Adaptation. In: Sikosek T, editor. Computational Methods in Protein Evolution. Methods in Molecular Biology. New York, NY: Springer; 2019. pp. 171–182. Available from: https://doi.org/10.1007/978-1-4939-8736-8_9

203. Mallik S, Tawfik DS, Levy ED (2022) How gene duplication diversifies the landscape of protein oligomeric state and function. Current Opinion in Genetics & Development 76:101966.

204. Finnigan GC, Hanson-Smith V, Stevens TH, Thornton JW (2012) Evolution of increased complexity in a molecular machine. Nature 481:360–364.

205. Pillai AS, Chandler SA, Liu Y, Signore AV, Cortez-Romero CR, Benesch JLP, Laganowsky A, Storz JF, Hochberg GKA, Thornton JW (2020) Origin of complexity in haemoglobin evolution. Nature 581:480–485.

206. Schulz L, Guo Z, Zarzycki J, Steinchen W, Schuller JM, Heimerl T, Prinz S, Mueller-Cajar O, Erb TJ, Hochberg GKA (2022) Evolution of increased complexity and specificity at the dawn of form I Rubiscos. Science 378:155–160.

207. Lee J, Blaber M (2011) Experimental support for the evolution of symmetric protein architecture from a simple peptide motif. Proceedings of the National Academy of Sciences 108:126–130.

208. Yang G, Anderson DW, Baier F, Dohmen E, Hong N, Carr PD, Kamerlin SCL, Jackson CJ, Bornberg-Bauer E, Tokuriki N (2019) Higher-order epistasis shapes the fitness landscape of a xenobiotic-degrading enzyme. Nat Chem Biol 15:1120–1128.

209. Kaltenbach M, Burke JR, Dindo M, Pabis A, Munsberg FS, Rabin A, Kamerlin SCL, Noel JP, Tawfik DS (2018) Evolution of chalcone isomerase from a noncatalytic ancestor. Nat Chem Biol 14:548–555.

210. Joho Y, Vongsouthi V, Spence MA, Ton J, Gomez C, Tan LL, Kaczmarski JA, Caputo AT, Royan S, Jackson CJ, et al. (2023) Ancestral Sequence Reconstruction Identifies Structural Changes Underlying the Evolution of Ideonella sakaiensis PETase and Variants with Improved Stability and Activity. Biochemistry 62:437–450.

211. Wilson C, Agafonov RV, Hoemberger M, Kutter S, Zorba A, Halpin J, Buosi V, Otten R, Waterman D, Theobald DL, et al. (2015) Using ancient protein kinases to unravel a modern cancer drug's mechanism. Science 347:882–886.

212. Hadzipasic A, Wilson C, Nguyen V, Kern N, Kim C, Pitsawong W, Villali J, Zheng Y, Kern D (2020) Ancient origins of allosteric activation in a Ser-Thr kinase. Science 367:912–917.

213. Natarajan C, Signore AV, Bautista NM, Hoffmann FG, Tame JRH, Fago A, Storz JF (2023) Evolution and molecular basis of a novel allosteric property of crocodilian hemoglobin. Current Biology 33:98-108.e4.

214. Park Y, Patton JEJ, Hochberg GKA, Thornton JW (2020) Comment on "Ancient origins of allosteric activation in a Ser-Thr kinase." Science 370:eabc8301.

215. Schupfner M, Straub K, Busch F, Merkl R, Sterner R (2020) Analysis of allosteric communication in a multienzyme complex by ancestral sequence reconstruction. Proceedings of the National Academy of Sciences 117:346–354.

216. Boucher JI, Jacobowitz JR, Beckett BC, Classen S, Theobald DL (2014) An atomic-resolution view of neofunctionalization in the evolution of apicomplexan lactate dehydrogenases Levitt M, editor. eLife 3:e02304.

217. Harman JL, Loes AN, Warren GD, Heaphy MC, Lampi KJ, Harms MJ (2020) Evolution of multifunctionality through a pleiotropic substitution in the innate immune protein S100A9 Garrett WS, Laub MT, editors. eLife 9:e54100.

218. Harman JL, Reardon PN, Costello SM, Warren GD, Phillips SR, Connor PJ, Marqusee S, Harms MJ (2022) Evolution avoids a pathological stabilizing interaction in the immune protein S100A9. Proceedings of the National Academy of Sciences 119:e2208029119.

219. Harms MJ, Thornton JW (2014) Historical contingency and its biophysical basis in glucocorticoid receptor evolution. Nature 512:203–207.

220. Kumar A, Natarajan C, Moriyama H, Witt CC, Weber RE, Fago A, Storz JF (2017) Stability-Mediated Epistasis Restricts Accessible Mutational Pathways in the Functional Evolution of Avian Hemoglobin. Molecular Biology and Evolution 34:1240–1251.

221. Tufts DM, Natarajan C, Revsbech IG, Projecto-Garcia J, Hoffmann FG, Weber RE, Fago A, Moriyama H, Storz JF (2015) Epistasis Constrains Mutational Pathways of Hemoglobin Adaptation in High-Altitude Pikas. Molecular Biology and Evolution 32:287–298.

222. Duan S, Govorkova EA, Bahl J, Zaraket H, Baranovich T, Seiler P, Prevost K, Webster RG, Webby RJ (2014) Epistatic interactions between neuraminidase mutations facilitated the emergence of the oseltamivir-resistant H1N1 influenza viruses. Nat Commun 5:5029.

223. Starr TN, Flynn JM, Mishra P, Bolon DNA, Thornton JW (2018) Pervasive contingency and entrenchment in a billion years of Hsp90 evolution. Proceedings of the National Academy of Sciences 115:4453–4458.

224. Bridgham JT, Ortlund EA, Thornton JW (2009) An epistatic ratchet constrains the direction of glucocorticoid receptor evolution. Nature 461:515–519.

225. Gong LI, Suchard MA, Bloom JD (2013) Stability-mediated epistasis constrains the evolution of an influenza protein Pascual M, editor. eLife 2:e00631.

226. Studer RA, Christin P-A, Williams MA, Orengo CA (2014) Stability-activity tradeoffs constrain the adaptive evolution of RubisCO. Proceedings of the National Academy of Sciences 111:2223–2228.

227. Starr TN, Picton LK, Thornton JW (2017) Alternative evolutionary histories in the sequence space of an ancient protein. Nature 549:409–413.

228. Horovitz A, Fersht AR (1990) Strategy for analysing the co-operativity of intramolecular interactions in peptides and proteins. Journal of Molecular Biology 214:613–617.

229. Horovitz A, Serrano L, Avron B, Bycroft M, Fersht AR (1990) Strength and co-operativity of contributions of surface salt bridges to protein stability. Journal of Molecular Biology 216:1031–1044.

230. Weinreich DM, Lan Y, Wylie CS, Heckendorn RB (2013) Should evolutionary geneticists worry about higher-order epistasis? Current Opinion in Genetics & Development 23:700–707.

231. Morrison AJ, Wonderlick DR, Harms MJ (2021) Ensemble epistasis: thermodynamic origins of nonadditivity between mutations. Genetics 219:iyab105.

232. Jalal ASB, Tran NT, Stevenson CE, Chan EW, Lo R, Tan X, Noy A, Lawson DM, Le TBK (2020) Diversification of DNA-Binding Specificity by Permissive and Specificity-Switching Mutations in the ParB/Noc Protein Family. Cell Reports 32:107928.

233. Akanuma S, Nakajima Y, Yokobori S, Kimura M, Nemoto N, Mase T, Miyazono K, Tanokura M, Yamagishi A (2013) Experimental evidence for the thermophilicity of ancestral life. Proceedings of the National Academy of Sciences 110:11067–11072.

234. Butzin NC, Lapierre P, Green AG, Swithers KS, Gogarten JP, Noll KM (2013) Reconstructed Ancestral Myo-Inositol-3-Phosphate Synthases Indicate That Ancestors of the Thermococcales and Thermotoga Species Were More Thermophilic than Their Descendants. PLOS ONE 8:e84300.

235. Garcia AK, Schopf JW, Yokobori S, Akanuma S, Yamagishi A (2017) Reconstructed ancestral enzymes suggest long-term cooling of Earth's photic zone since the Archean. Proceedings of the National Academy of Sciences 114:4619–4624.

236. Gaucher EA, Govindarajan S, Ganesh OK (2008) Palaeotemperature trend for Precambrian life inferred from resurrected proteins. Nature 451:704–707.

237. Iwabata H, Watanabe K, Ohkuri T, Yokobori S, Yamagishi A (2005) Thermostability of ancestral mutants of Caldococcus noboribetus isocitrate dehydrogenase. FEMS Microbiology Letters 243:393–398.

238. Miyazaki J, Nakaya S, Suzuki T, Tamakoshi M, Oshima T, Yamagishi A (2001) Ancestral Residues Stabilizing 3-Isopropylmalate Dehydrogenase of an Extreme Thermophile: Experimental Evidence Supporting the Thermophilic Common Ancestor Hypothesis. The Journal of Biochemistry 129:777–782.

239. Risso VA, Gavira JA, Mejia-Carmona DF, Gaucher EA, Sanchez-Ruiz JM (2013) Hyperstability and Substrate Promiscuity in Laboratory Resurrections of Precambrian β-Lactamases. J. Am. Chem. Soc. 135:2899–2902.

240. Thomas A, Cutlan R, Finnigan W, van der Giezen M, Harmer N (2019) Highly thermostable carboxylic acid reductases generated by ancestral sequence reconstruction. Commun Biol 2:1–12.

241. Emond S, Petek M, Kay EJ, Heames B, Devenish SRA, Tokuriki N, Hollfelder F (2020) Accessing unexplored regions of sequence space in directed enzyme evolution via insertion/deletion mutagenesis. Nat Commun 11:3469.

242. Gomez-Fernandez BJ, Risso VA, Rueda A, Sanchez-Ruiz JM, Alcalde M (2020) Ancestral Resurrection and Directed Evolution of Fungal Mesozoic Laccases. Applied and Environmental Microbiology 86:e00778-20.

243. Hobbs HT, Shah NH, Shoemaker SR, Amacher JF, Marqusee S, Kuriyan J (2022) Saturation mutagenesis of a predicted ancestral Syk-family kinase. Protein Science 31:e4411.

244. Risso VA, Sanchez-Ruiz JM, Ozkan SB (2018) Biotechnological and protein-engineering implications of ancestral protein resurrection. Current Opinion in Structural Biology 51:106–115.

245. Schenkmayerova A, Pinto GP, Toul M, Marek M, Hernychova L, Planas-Iglesias J, Daniel Liskova V, Pluskal D, Vasina M, Emond S, et al. (2021) Engineering the protein dynamics of an ancestral luciferase. Nat Commun 12:3616.

246. Trudeau DL, Tawfik DS (2019) Protein engineers turned evolutionists—the quest for the optimal starting point. Current Opinion in Biotechnology 60:46–52.

247. Scotese CR, Song H, Mills BJW, van der Meer DG (2021) Phanerozoic paleotemperatures: The earth's changing climate during the last 540 million years. Earth-Science Reviews 215:103503.

248. Woese CR (1987) Bacterial evolution. Microbiological Reviews 51:221–271.

249. Garcia AK, Kaçar B (2019) How to resurrect ancestral proteins as proxies for ancient biogeochemistry. Free Radical Biology and Medicine 140:260–269.

250. Wheeler LC, Lim SA, Marqusee S, Harms MJ (2016) The thermostability and specificity of ancient proteins. Current Opinion in Structural Biology 38:37–43.

251. Sternke M, Tripp KW, Barrick D Chapter Seven - The use of consensus sequence information to engineer stability and activity in proteins. In: Tawfik DS, editor. Methods in Enzymology. Vol. 643. Enzyme Engineering and Evolution: General Methods. Academic Press; 2020. pp. 149–179. Available from: https://www.sciencedirect.com/science/article/pii/S0076687920302500

252. Susko E, Roger AJ (2013) Problems With Estimation of Ancestral Frequencies Under Stationary Models. Systematic Biology 62:330–338.

253. Trudeau DL, Kaltenbach M, Tawfik DS (2016) On the Potential Origins of the High Stability of Reconstructed Ancestral Proteins. Molecular Biology and Evolution 33:2633–2641.

254. Williams PD, Pollock DD, Blackburne BP, Goldstein RA (2006) Assessing the Accuracy of Ancestral Protein Reconstruction Methods. PLOS Computational Biology 2:e69.

255. Pollock DD, Thiltgen G, Goldstein RA (2012) Amino acid coevolution induces an evolutionary Stokes shift. Proceedings of the National Academy of Sciences 109:E1352–E1359.

256. Risso VA, Manssour-Triedo F, Delgado-Delgado A, Arco R, Barroso-delJesus A, Ingles-Prieto A, Godoy-Ruiz R, Gavira JA, Gaucher EA, Ibarra-Molero B, et al. (2015) Mutational Studies on Resurrected Ancestral Proteins Reveal Conservation of Site-Specific Amino Acid Preferences throughout Evolutionary History. Molecular Biology and Evolution 32:440–455.

257. Taverna DM, Goldstein RA (2002) Why are proteins marginally stable? Proteins: Structure, Function, and Bioinformatics 46:105–109.

258. Hilton SK, Bloom JD (2018) Modeling site-specific amino-acid preferences deepens phylogenetic estimates of viral sequence divergence. Virus Evolution 4:vey033.

259. Ho LST, Susko E (2022) Ancestral state reconstruction with large numbers of sequences and edge-length estimation. J. Math. Biol. 84:21.

260. Arenas M, Bastolla U (2020) ProtASR2: Ancestral reconstruction of protein sequences accounting for folding stability. Methods in Ecology and Evolution 11:248–257.

261. Wheeler LC, Anderson JA, Morrison AJ, Wong CE, Harms MJ (2018) Conservation of Specificity in Two Low-Specificity Proteins. Biochemistry 57:684–695.

262. Del Amparo R, Arenas M (2022) Consequences of Substitution Model Selection on Protein Ancestral Sequence Reconstruction. Molecular Biology and Evolution 39:msac144.

263. Holland BR, Ketelaar-Jones S, O'Mara AR, Woodhams MD, Jordan GJ (2020) Accuracy of ancestral state reconstruction for non-neutral traits. Sci Rep 10:7644.