

ANALYZING LIGAND SPECIFICITY TO ASSESS THE  
EVOLUTION OF TLR4

by

CORINTHIA BROWN

A THESIS

Presented to the Department of Biology  
and the Robert D. Clark Honors College  
in partial fulfillment of the requirements for the degree of  
Bachelor of Science

February 2024

THESIS DEFENCE COMMITTEE APPROVAL

Approved:  Date: 2/23/2024


Dr. Mike Harms, Ph.D., Primary Advisor

Approved:  Date: 2/23/24

Kona Orlandi, Primary Graduate Student Advisor

Approved:  Date: 2/23/24

Dr. Daphne Gallagher, Ph.D., Robert D. Clark Honors College Faculty Representative

Approved:  Date: 02/23/24

Dr. Daniel Grimes, Ph.D., Biology Honors Faculty Representative

## **An Abstract of the Thesis of**

Corinthia Brown for the degree of Bachelor of Science  
in the Department of Biology to be taken June 2024

Title: Analyzing Ligand Specificity to Assess the Evolution of TLR4

Approved: Dr. Michael J. Harms  
Primary Thesis Advisor

Inflammation is a vital process our bodies use to remove foreign entities and help restore function to damaged tissue. However, when inflammation excessively activates it can lead to arthritis, neurodegeneration, and sepsis, which contributes to 11 million deaths a year.

Inflammation results from inflammatory cytokines produced by the NF- $\kappa$ B pathway, activated by Toll-Like Receptor 4 (TLR4). We know that certain lipopolysaccharides (LPS) present on gram-negative bacteria drive the dimerization of TLR4 and activate inflammatory cytokine production. However, we do not completely understand the rules which govern TLR4 activation, making it difficult to control this regulator of inflammation, particularly in clinical applications.

To understand how TLR4 activates, the Harms lab studies the alterations in function throughout this protein's evolutionary history. However, when the TLR4 in question is far from humans on the phylogenetic tree we experience difficulties observing the protein's activation. To observe distant species' TLR4 activities we developed a method which allows the outer portion (ectodomain) of the protein to bind to its specific ligand yet have an inner portion (transmembrane and TIR domain) which allows for interactions with the human proteins which bridge the connection between TLR4 and the inflammatory pathway.

## **Acknowledgements**

First and foremost, I would like to thank Mike Harms and Kona Orlandi for guiding and supporting me over the last few years. I am grateful for all the time and effort you have invested into my growth, and your encouragement to become the best I can be. In addition, I am thankful for all the support and feedback I have received from the entirety of the lab. It is vital to have a strong support system when entering or working in a lab and you all have given me everything I could ask for in spades.

Along with my time in the lab, my time in the CHC has provided many opportunities. However, I would not have been able to navigate and utilize the resources provided had it not been for my CHC representative Daphne Gallagher. I thank you for all the time you have spent encouraging me and suggesting alternatives when I believed I was stuck. Your guidance has allowed me to be confident in the path I have chosen and put my weight into it.

I am grateful for the support provided by the Center for Undergraduate Research and Engagement for awarding the Summer Undergraduate Research Fellowship which has funded this project.

Finally, I would not be here without my friends and family. Mom, Dad, I love you both so much, and I appreciate you beyond words for the support you have given me throughout my life. Shoutout to my roommates for their overflowing energy and love throughout these last four years. Knowing that I can count on you all for advice and foresight has kept me going through tough times and helped me write this thesis. Thank you.

## Table of Contents

Introduction	8
Toll-like Receptors	9
Lipopolysaccharide (LPS)	12
Toll-Like Receptor 4 Mechanism	14
TLR4 structural components and dimerization	16
Known TLR4 Ligand Specificity	18
Methods	24
Building The Plasmid	24
KLD Mutagenesis:	26
Best Conditions for KLD Mutagenesis	29
Activity Assay	30
Dual-Glo Luciferase Assay	31
Data Analysis	33
Results	35
Xenopus (Frog) TLR4 Activity in HEK293T cells	35
Testing the Effects of Helix Shifting on Human TLR4	40
TLR4 Homolog; CD180 Activity in HEK293T cells	42
TLR4 and MD2 Mutants to Change Ligand Specificity in Zebrafish	45
Discussion	47
Overview	47
Future directions	49
Glossary	51
Further Notes:	55
Further Data	58
Bibliography	60
Appendix	

## List of Figures

<b>Figure 1:</b> Structure of LPS (Lipopolysaccharide).....	12
<b>Figure 2:</b> Human TLR4 Dimerization Mechanism, and Ligand Binding Locations .....	15
<b>Figure 3:</b> The Three Main Domains Shown on a Human TLR4. ....	17
<b>Figure 4:</b> A phylogenetic tree heat map showing relative response to LPS variants LPS-R and lipid IV <sub>A</sub> in TLR4.....	19
<b>Figure 5:</b> Positions of Zebrafish Mutations in TLR4 and MD-2 .....	22
<b>Figure 6:</b> Diagram of how SLIC Cloning Works. ....	25
<b>Figure 7:</b> Primers used in hTLR4_hTLR4 +NK Mutant .....	26
<b>Figure 8:</b> KLD Mutagenesis Process. ....	28
<b>Figure 9:</b> Results of varying PCR conditions after being transformed with XL10 Gold bacteria. ....	29
<b>Figure 10:</b> Diagram of a General NF-kB Response Luciferase Reporter Assay .....	33
<b>Figure 11:</b> Diagram of Theorized Human vs Frog TLR4 NF-kB Pathway Activation .....	36
<b>Figure 12:</b> Ligand specificity of <i>Xenopus</i> TLR4 When Placed into a Human Cassette .....	37
<b>Figure 13:</b> <i>Xenopus</i> TLR4 Shifted Raw Renilla Signal.....	38
<b>Figure 14:</b> 3D models of what helix shifting may do to the orientation of TLR4.....	39
<b>Figure 15:</b> Luciferase Readings of Frog TLR4 Ligand specificity when Helix Shifted.....	40
<b>Figure 16:</b> Luciferase Readings of Helix Shifted hTLR4.....	41
<b>Figure 17:</b> Luciferase Readings on the Ligand Specificity of CD180 in a Human Cassette.....	43
<b>Figure 18:</b> (Figure 1 from Edwards et. al.) Shows the variation in structure between CD180 and TLR4 <sup>22</sup> . ....	44
<b>Figure 19:</b> Luciferase readings for zfTLR4 & zfMD-2 mutants normalized to human complex.....	46
<b>Figure 20:</b> CD180 & TLR4 Variant Raw Renilla Signal.....	58
<b>Figure 21:</b> Human TLR4 Shifted Raw Renilla Signal.....	59
<b>Figure 22:</b> zfMutant's Raw Renilla Signal .....	59

## **List of Tables**

Table of Primer Compositions for Mutagenesis/KLD	55
Table of SLIC/HIFI Reactions and their Primers	55

## Introduction

DNA is the genetic code for all life and encodes what molecules an organism can synthesize within their bodies to complete various tasks. DNA is used as a template for RNA, which is composed of four different nucleotides, adenine, cytosine, guanine, and uracil. Each triplet of nucleotides is read and made into an amino acid.

Proteins are polymers, chains of amino acids, encoded by DNA which perform most of the work in cells, and are vital for function, structure, and regulation of the body's organs. As proteins are so vital to the operation of a human body, a major field of study in biochemistry is devoted to understanding how proteins work. Often the proteins scientists study are large, with an average eukaryotic protein measuring around 400 amino acids long, with the largest being composed of 27,000 amino acids<sup>1,2</sup>! As there are 20 different amino acids commonly used in proteins, the possibilities for studying how the function of a protein can change if a portion of its amino acid sequence is altered are quite extensive. Therefore, scientists must create a pattern that helps them identify regions with which to study their protein of interest.

Nature has a brilliant study system which highlights changes in protein function over time due to changes in amino acid sequence. Evolution. Proteins, like the animals they are made in, change over time due to mutations in DNA. These mutations change the nucleotides which compose DNA which are made into the amino acids which proteins are composed of. When the nucleotide triplets are altered, say AAC  $\Rightarrow$  AAA, we get Lysine instead of Asparagine. This alteration would place a negatively charged residue where a polar residue previously resided, making that region of the protein more hydrophilic. A mutation as small as a single nucleotide can have dramatic effects on the function of the protein. A single nucleotide mutation can stop a

protein from folding correctly and taking the shape that it should. Protein misfolding is the cause of many devastating diseases such as Alzheimer's.

However, when a mutation does not break a protein, it may be incorporated, allowing different species to evolve similar proteins with varied functions over time. We, as scientists, can look at how protein function in related species has changed over time to connect the changes in phenotype to the changes in genotype. In other words, we can learn how the sequence connects to the protein's function.

To study the Toll-like receptor 4 (TLR4) protein's connection between sequence and function we need to obtain a series of protein sequences and their corresponding response to various ligands. When measuring what liposaccharides (LPS) each TLR4 responded to, and the degree to which it responded, we found that the further removed from humans the model species was, the lower the response signal. This became a major issue for characterizing species that lay below zebrafish on the phylogenetic tree as none of the measured response signals were statistically significant. Therefore, we needed to design a method of improving the TLR4 signal strength for these species.

## **Toll-like Receptors**

Mammalian immune systems are generally split into two groups: 1) Adaptive immunity which builds an immunological memory of pathogens which enter our body by synthesizing pathogen-specific receptors, and 2) innate immunity which provides a non-specific response to general structural patterns common to pathogens. These general structural components are known as pathogen-associated molecular patterns (PAMPs), as well as endogenously produced signals of tissue damage, known as damage-associated molecular patterns (DAMPs). Innate immunity is important as it is the body's first response to pathogens and wounds to prevent

illness. Studies in fruit flies have shown that Toll-like receptors are primary activators of innate immunity<sup>3</sup>. This ancient family of transmembrane pattern recognition receptors are present in species ranging from plants to mammals and are highly conserved. This makes studying TLR evolution an effective way to discover how to regulate their activation.

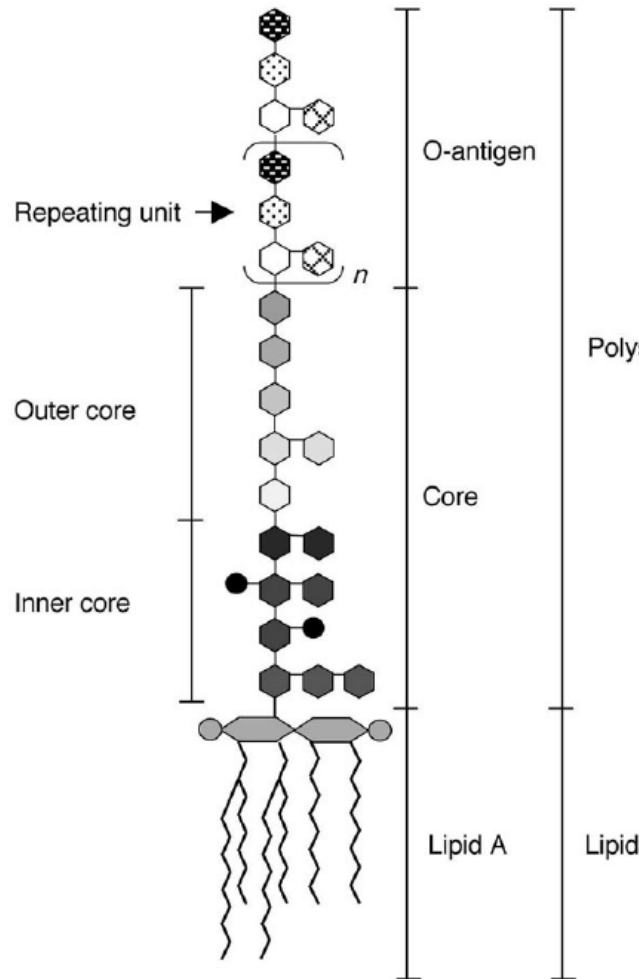
Several types of TLRs recognize PAMPs or DAMPs at the surface of immune cells and then initiate internal signaling cascades to activate NF- $\kappa$ B, a transcription factor that helps to orchestrate the production of inflammatory markers like cytokines. When cytokines are released from the cell, they signal the immune system to turn on and recruit other immune cells.

Inflammation is a process in which inflammatory mediators dilate blood vessels which allow for increased blood flow to injured tissues. The increased blood flow allows for more immune cells to enter and exit the area, allowing for faster response times from your body. In addition, fluid build-up in inflamed areas of your body help to flush out viruses and bacteria that may have entered<sup>4</sup>. However, inflammation cannot always be active as excessive inflammation damages tissue, eventually damaging organs. Therefore, it is important that the immune system can tightly regulate TLR activation and deactivation.

Within humans there are 10 TLRs which respond to various signals. For example, TLR1 activates the innate immune response due to triacyl lipopeptides present in bacteria, TLR5 responds to flagellin which are present on flagellin bacteria, and TLR4 responds to LPS from Gram-negative bacteria. LPS recognition by TLR4 is the primary cause of sepsis. Sepsis is a life-threatening complication of an infection that occurs when inflammatory markers in the bloodstream trigger inflammation throughout the body, known as a cytokine storm. As inflammation spreads it can cause severe damage to several organs, drastically lower blood pressure, and sometimes results in death.

Previous studies have demonstrated a direct effect of increased and decreased expression of TLR4 within mice and the contraction of sepsis. In Franck B. et. al. the authors demonstrated that mice with two or more copies of the TLR4 gene had an increased survivorship when initially exposed to a salmonella infection<sup>5</sup>. However, as the infection continued the mice with multiple copies of the TLR4 gene developed an excessive inflammatory immune response which led to the death of those specimens<sup>5</sup>. On the other hand, when a mutant strain of mice, which lacked the expression of TLR4 was exposed to lethal amounts of an inflammatory activator there was no response. The mice which had functioning TLR4 died due to an overactive inflammatory reaction<sup>6,7</sup>. Through this literature we can see that being able to control the activation of TLR4 would allow doctors to prevent damage caused by excessive inflammation while treating an infection.

## Lipopolysaccharide (LPS)



**Figure 1:** Structure of LPS (Lipopolysaccharide)

LPS is a component of Gram-negative bacteria's cell membranes, and is comprised of an O-antigen, a core, and Lipid A, an endotoxin<sup>8</sup>.

What is LPS, and why would the immune system evolve to respond in this manner if the results are so deadly? LPS is a component of Gram-negative bacteria's outer membranes and is composed of three sections: an O-antigen, a core oligosaccharide, and lipid A (Figure 1). The O-antigen portion of LPS is what is exposed to the outside world and participates in various interactions that we do not understand yet<sup>9</sup>. The O-antigen structure is highly variable even within a single population of bacteria and can be absent. LPS molecules are generally classified

into two groups, R = rough for when the O-antigen is absent, and S = smooth for when the O-antigen is present. The core oligosaccharide attaches the O-antigen to lipid A. The core oligosaccharide structure is relatively conserved within a species but can vary drastically between species. The highly conserved lipid A portion is responsible for the toxicity of LPS. Lipid A has two glucosamine units with phosphate groups and hydrophobic fatty acid acyl chains that are essential for bacteria to maintain the barrier property of their outer membrane<sup>8</sup>. The conserved lipid A molecular pattern is recognized by TLR4.

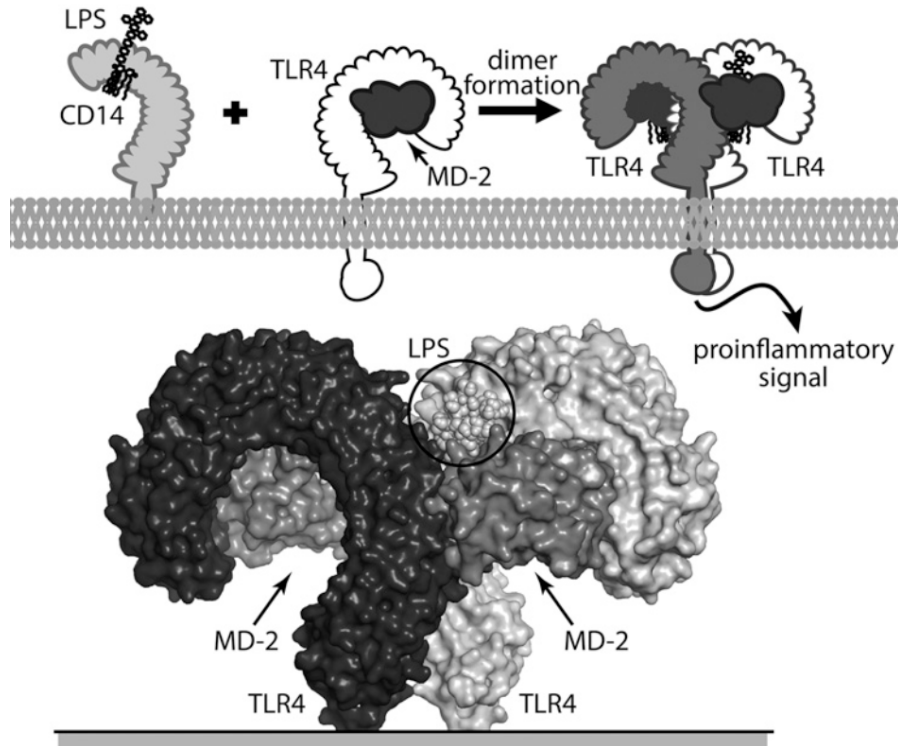
Although lipid A is highly conserved, some parts of its structure can vary between species or even within a single species. Most LPS molecules are hexa-acylated<sup>10</sup> but the number of acyl chains depends on the species and typically varies from four to seven. For instance, *Escherichia coli* LPS is primarily composed of LPS with six acyl chains, whereas LPS from *Rhodobacter Sphaeroides* often has five acyl chains. However, under certain environmental conditions, bacteria can change the distribution of their LPS acyl chain number. The acyl chain position and length, types of bonds, and presence of polar groups can also differ along with the number of phosphate groups on the sugar backbone. While the lipid A portion of LPS can vary to some degree, this highly conserved structural feature alone is a potent PAMP recognized by TLR4<sup>10</sup>.

Not all LPSs are built equal in the eyes of all TLR4s. Certain TLR4s from specific species, like humans, only react with specific forms of LPS. Other species' TLR4s may be less particular. For example, human TLR4 strongly reacts with LPS-R, a rough-form LPS purified from *E. coli* K12 strain with no O-antigen and six acyl-chains, but shows no activity when challenged with Lipid IVa, a biosynthetic precursor to LPS with no O-antigen and four acyl-chains. However, TLR4 from zebrafish shows the opposite response, with high immune pathway

activation when exposed to Lipid IVa, and very little activation with LPS-R<sup>11</sup>. Intriguingly, mouse TLR4 responds strongly to both LPS variants. We know from x-ray crystallography that human and mouse TLR4s can bind both LPS-R and Lipid IVa, but the orientation of Lipid IVa inhibits activation only in human TLR4<sup>12</sup>. We know the differences between human, mouse, and zebrafish TLR4 protein sequences, so how can we connect the differences in sequence to the differences in activity?

### **Toll-Like Receptor 4 Mechanism**

The mechanism of TLR4 activation by LPS in humans and mice is well established due to its implications in sepsis. TLR4 activation by LPS relies on co-factors which are not always conserved across species<sup>11</sup>. In humans, when Gram-negative bacteria are present, some LPS is released and bound by CD14, a protein present on the outside of immune cells. CD14 then transports the LPS molecule to the LPS-binding pocket within MD-2, a protein which forms a complex with TLR4. Once the TLR4/MD-2/LPS complex is formed, two TLR4/MD-2/LPS complexes dimerize<sup>13</sup>.



**Figure 2:** Human TLR4 Dimerization Mechanism, and Ligand Binding Locations

Above: The mechanism TLR4 uses to bind to LPS, dimerize, and release a proinflammatory<sup>14</sup>

Bottom: Rendering of the TLR4-MD2 complex showing areas of contact between the two proteins, alterations of which may influence activation of the NF-kB pathway.

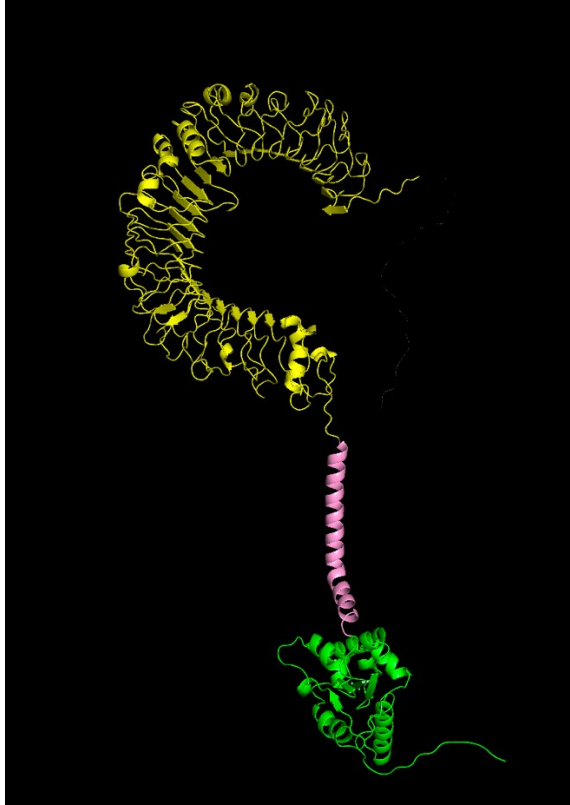
Once dimerized the signal is propagated within the cell to activate the NF-kB pathway.

TIRAP, a recruiter protein, constantly travels along the inside of the cell membrane to find then bind dimerized TLR4s to nucleate formation of a myddosome composed of MyD88 and IRAK family kinases. These myddosomes control further downstream signaling for the activation of the NF-kB transcription factor which upregulates the expression of inflammatory cytokines. The myddosome is analogous to a secondary messenger in that it allows for signal amplification and transduction. However, the structure lacks the graded nature of a secondary messenger, as it does not take into consideration concentration, signal half-lives, or deactivation kinetics. Therefore, an autophagosome must manually disassemble the myddosome to stop the inflammatory pathway

signal<sup>13</sup>. This manual disassembly process is partially why sepsis is so deadly. Once the signal has been transduced to this point, the body can only break so much of the myddosome down at a time.

### **TLR4 structural components and dimerization**

The structural features of TLR4 determine how two TLR4 complexes can dimerize. Within TLR4 there are three sections which must contain specific residues which allow it to exist and interact in that space favorably. These three sections are the ectodomain, transmembrane, and TIR domain in TLR4. The ectodomain is where the TLR4/MD2 complex lies, and interacts with the outside of the cell, including LPS from gram-negative bacteria. This portion of the protein is shown by the yellow region in Figure 3. The transmembrane domain is the portion of the protein which must span the cell's membrane therefore has many hydrophobic residues. To further increase like-with-like interactions proteins which travel through the cell membrane often take on an alpha-helix secondary structure. We see this pattern in TLR4 as well, shown in Figure 3 by the pink section between the ectodomain and TIR domain. Shown in Figure 3 by the green portion is the TIR domain, which transduces signals which originated outside of the cell to the inside. It is important that the TIR domain of a protein can interact with the internal proteins of a cell through binding.



**Figure 3:** The Three Main Domains Shown on a Human TLR4.

Yellow: Ectodomain, Pink: Transmembrane domain, Green: TIR domain. 3D model synthesized by Alphafold2, and color modifications and visualizations preformed via Pymol.

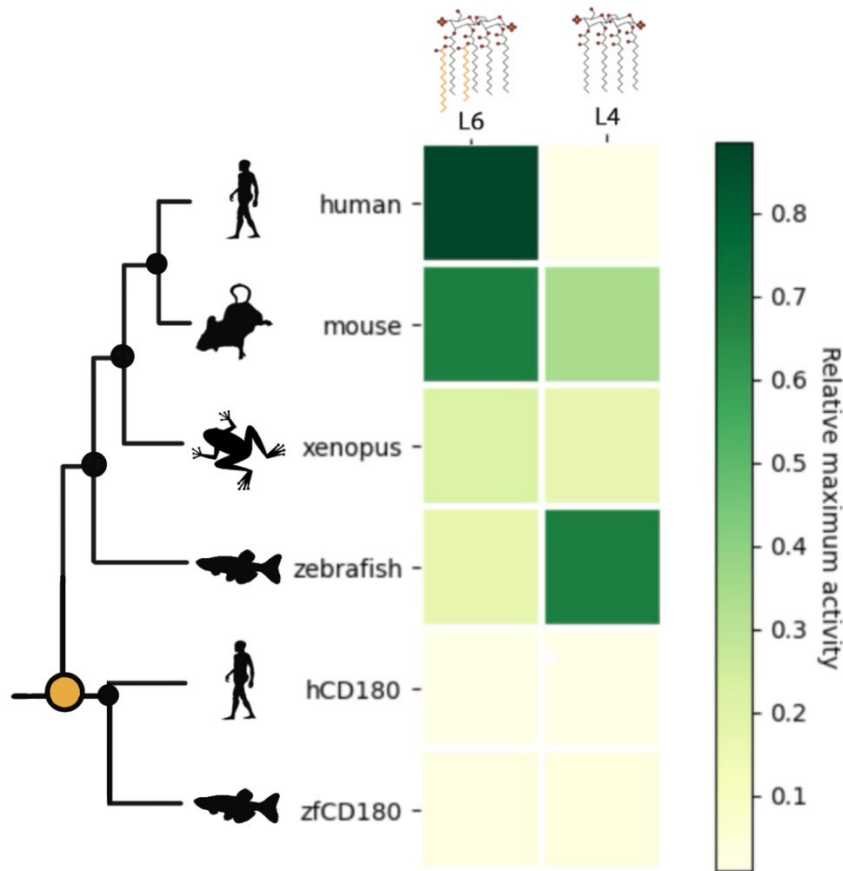
We understand how TLR4 and MD-2 form a complex and which residues are involved. These residues which bind ligands, and therefore confer ligand specificity lie in the ectodomain. However, we do not understand how ligand specificity is determined. A paper by Park, Beom and Lee, Jie-oh noted that the acyl chains of the Lipid A slip within MD-2's pockets and the two  $\text{PO}_4$  groups on Lipid A engage in charge and hydrogen bonding interactions with charged residues in the TLR4/MD-2 complex<sup>15</sup>. The LPS binding causes dimerization, as the LPS can create the additional binding interface between TLR4 and MD-2. This aspect of LPS creating an additional binding interface is what literature thought influenced TLR4's specificity. Previous studies looking at crystalline structures show that when LPS-R was placed into the human TLR4/MD-2 complex there was that binding interface protruding acyl-chain. On the other hand,

when crystalline structures of human TLR4/MD-2 with lipid IV<sub>A</sub> bound were examined no such acyl chain existed. This would seem to indicate that the acyl chain existence correlates to ligand specificity. However, when crystalline structures of mouse lipid IV<sub>A</sub> TLR4/MD-2 crystalline structures were examined this protruding acyl chain was absent, yet mice still reacted to lipid IV<sub>A</sub><sup>15</sup>. Therefore, this “binding interface” may be species specific, or be dependent on multiple factors.

Studies on other transmembrane proteins have shown that the transmembrane portion of the protein plays a part in the oligomerization of transmembrane proteins. For instance, a study by Popot found that although the  $\alpha$ -helices which constitute the transmembrane domain of a protein have only been considered as hydrophobic anchors, they play a role in oligomerization. In the case of glycoprotein A the transmembrane domain alone is sufficient for dimerization, even when removed from its normal protein context<sup>16</sup>. Therefore, the transmembrane domain is likely a key contributor to TLR4 dimerization success.

### **Known TLR4 Ligand Specificity**

To fully understand the evolutionary history of TLR4 specificity and the rules governing it, our lab wanted to characterize several early vertebrate TLR4 complexes to fill in the phylogenetic tree but kept running into problems of low activity or cell toxicity in their assay. The goal of this thesis project was to try to remedy this problem using molecular biology tools.



**Figure 4:** A phylogenetic tree heat map showing relative response to LPS variants LPS-R and lipid IV<sub>A</sub> in TLR4.

This table shows the relative activity in response to LPS as normalized to human L6 response. (Human L6 is 1) a stronger response is shown by a darker green, and a lower response is shown by a lighter green. The black dots show speciation, and the orange circle shows a duplication event. Heatmap base code provided by Sophia Phillips.

Fish are the earliest branching vertebrate that we know to have TLR4 and MD-2, however, very little characterization has been done for these proteins. Loes et. al showed zebrafish TLR4 can be lowly activated by LPS-R and we have recently found it can be more strongly activated by Lipid IV<sub>A</sub><sup>11</sup>. However, this signal is still much lower than the human TLR4 signal response to LPS-R in our assay and we do not understand why.

Amphibians diverged after fish but before mammals on the phylogenetic tree and although they have TLR4 complex components, these proteins have not been characterized. In

our lab's attempts to characterize frog TLR4 (from *Xenopus laevis*) we find that expressing these transgenic proteins kills the cells in our assay, resulting in low, variable signal.

In addition to characterizing fish and amphibian TLR4 proteins, our lab is interested in characterizing the TLR4 homolog CD180, which has a similar structure to TLR4 in that the ectodomain of the proteins are nearly identical. CD180 is thought to work with a cofactor that is homologous to MD-2, MD-1, to bind LPS. If we can analyze the ligand specificity and subsequent activity level for CD180, we can gain information on the ligand specificity and activity of the ancestor of TLR4 and CD180. This information would help us understand more clearly how TLR4 structure and function have evolved throughout evolutionary history. However, the TIR domain is missing from CD180. We therefore cannot test what historical LPS-binding specificity might be preserved in CD180 without the attachment of said region.

### **Could species mismatch between TLR4's TIR domain and internal signaling proteins cause low signal?**

The ectodomain of TLR4 interacts with LPS outside of the cell through MD-2, so the conformational change that begins the dimerization process results from the ectodomain. So, normally, we would think that if a signal is unusually low there may be an issue with the ectodomain dimerizing or binding to LPS properly. On the other hand, we know that the TIR domain must interact with TIRAP for the signal to be transmitted to the NF- $\kappa$ B pathway.

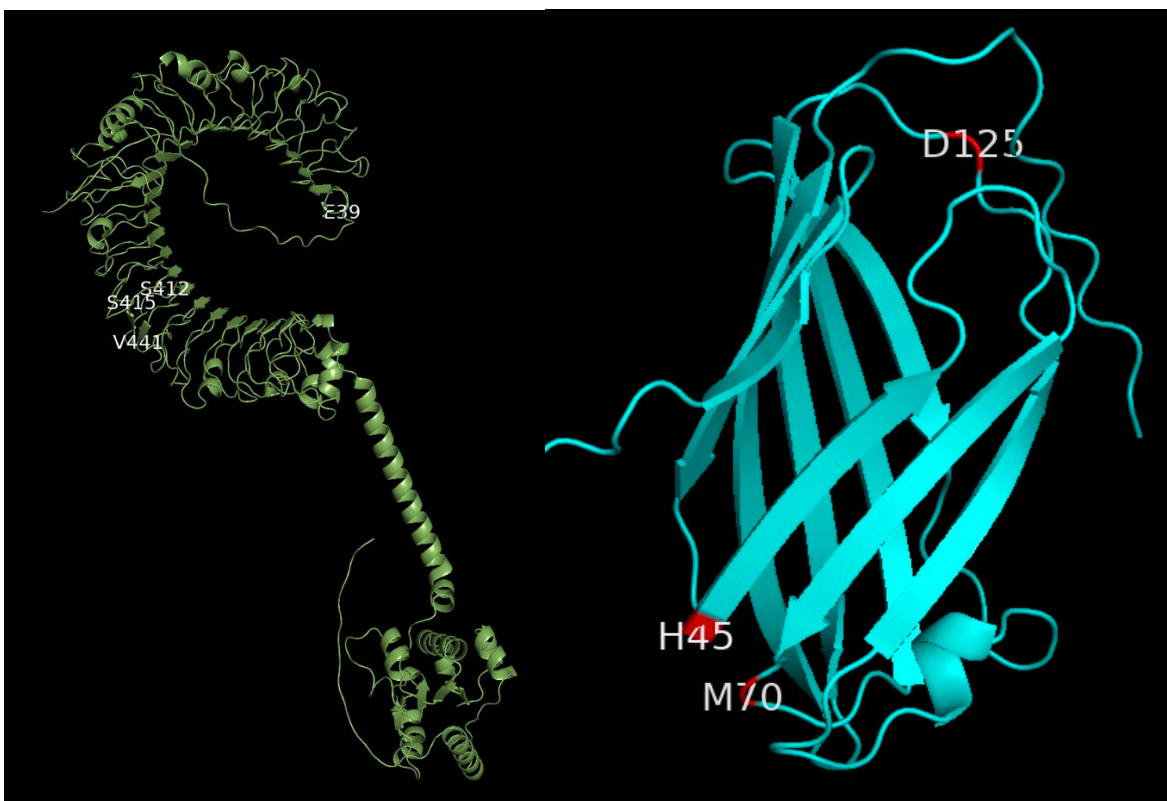
When we test the immune reaction to LPS of various species' TLR4 we do so using human cells. These cells have human signaling pathway proteins. So, we hypothesized that we often see very low signal or high levels of cytotoxicity for TLR4 from species distant from humans on the phylogenetic tree due to the TIR domain being unable to transmit the signal to the

NF- $\kappa$ B pathway. Therefore, we decided to investigate if an internal mismatch was reducing signal or causing toxicity.

To test for the possible mismatch between TLR4 and the signaling proteins, we built a genetic cassette containing human transmembrane and TIR domains downstream of a position which will allow us to place the ectodomains from any other TLR4s. We used this genetic cassette to test zebrafish and frog TLR4 specificity. We also used this cassette to see if adding a transmembrane and TIR domain to CD180 was sufficient to visualize the ectodomain's activity.

### **Changing zebrafish TLR4 ligand specificity with single point mutations**

Another portion of this project aimed to manipulate the ligand specificity of zebrafish TLR4 through single amino acid mutations in TLR4 and MD2. If a single amino acid variation altered the activity of TLR4 or MD-2, either through alterations in their ligand specificity or their degree of activation we could better understand the evolution of TLR4. To choose the specific amino acids in both zfTLR4 and zfMD-2 we turned to previous literature. Through reviewing the literature Sophia Phillips, a graduate student in the Harms lab decided that there were 7 amino acids of interest that could possibly alter the zebrafish TLR4's response to ligands, altering the amino acids at those positions to those seen in humans. These mutations of interest were: L441F, G415N, D412S, and G39E for TLR4, and D125K, M70R, and H45Y.



**Figure 5:** Positions of Zebrafish Mutations in TLR4 and MD-2

This diagram shows the positions of the discussed mutations on zebrafish TLR4 (Green structure on left) and Md-2 (Blue structure on right).

The position 441 was a location of interest due to its proximity to two residues that were previously studied, 440 and 444 which were shown to be involved in hydrophobic-based interactions between TLR4 and MD-2<sup>17</sup>. This amino acid would therefore be changed from (L) Leucine (NP) to Phenylalanine (NP) which has a larger side chain. This larger side chain should increase the non-polar contact and increase binding stability, and therefore ligand activation. The next mutations, G415N and D412S were chosen due to their proximity to S416 and N417 which have been shown to bind to MD-2 through hydrogen bonds and represent an area which is involved in the binding interface between TLR4 and MD-2. These residues would be changed from (G) glycine (NP) to (N) Asparagine (P), and (D) Aspartic acid (-) to (S) serine (P). We theorized that changing these residues in zebrafish to those seen in humans would increase the

number of binding contacts and improve signal response to ligands. Furthermore, G39E was chosen as a residue of interest due to its similarity to residue 42 in human TLR4 which is known to participate in hydrogen bonding with MD-2<sup>17</sup>. Changes at position 42 in human TLR4 were shown to alter ligand specificity. Therefore, an alteration from (G) glycine (NP) to (E) Glutamic Acid (-), the amino acid present in this position in humans, we would increase the number of polar contacts and should change ligand specificity, seeing a decreased signal response to Lipid IV<sub>A</sub>, and a larger response to LPS-R.

The first location for a mutation in MD-2 was the 125th amino acid which is involved with binding to TLR4 through a hydrogen bond and lies between the 126<sup>th</sup> and 124<sup>th</sup> amino acids which have been shown to bind to human TLR4 through hydrophobic interactions<sup>17</sup>. Furthermore, this amino acid had been connected to species specificity in human MD-2 by a previous paper, which noted this change in specificity may occur due to that residue being located at the dimerization interface, where differences in charge affect receptor dimerization<sup>18</sup>. Therefore, the goal was to change the original amino acid at that location (D) Aspartic acid (+) to a (K) Lysine (-), to observe the resulting effect on the MD-2's ability to then form a dimer with TLR4 in response to LPS-R and lipid IV<sub>A</sub>. The next mutations, M70R and H45K are adjacent to other residues, 69 and 42 which has been noted to affect species specificity when altered in human MD-2. Specifically, surface charge differences at these points affect the binding angle/the rigidity between TLR4/MD-2 indirectly effecting the TLR4/MD-2 interfaces<sup>18</sup>. Therefore, by altering the amino acid from (M) methionine (NP) to (R) Arginine (+) and (H) Histidine (+/P) to (K) Lysine (+) we should see an alteration in ligand specificity. Theoretically we should see a change in preference from reacting to lipid IV<sub>A</sub> to LPS-R as we make the zebrafish MD-2 more similar to human MD-2.

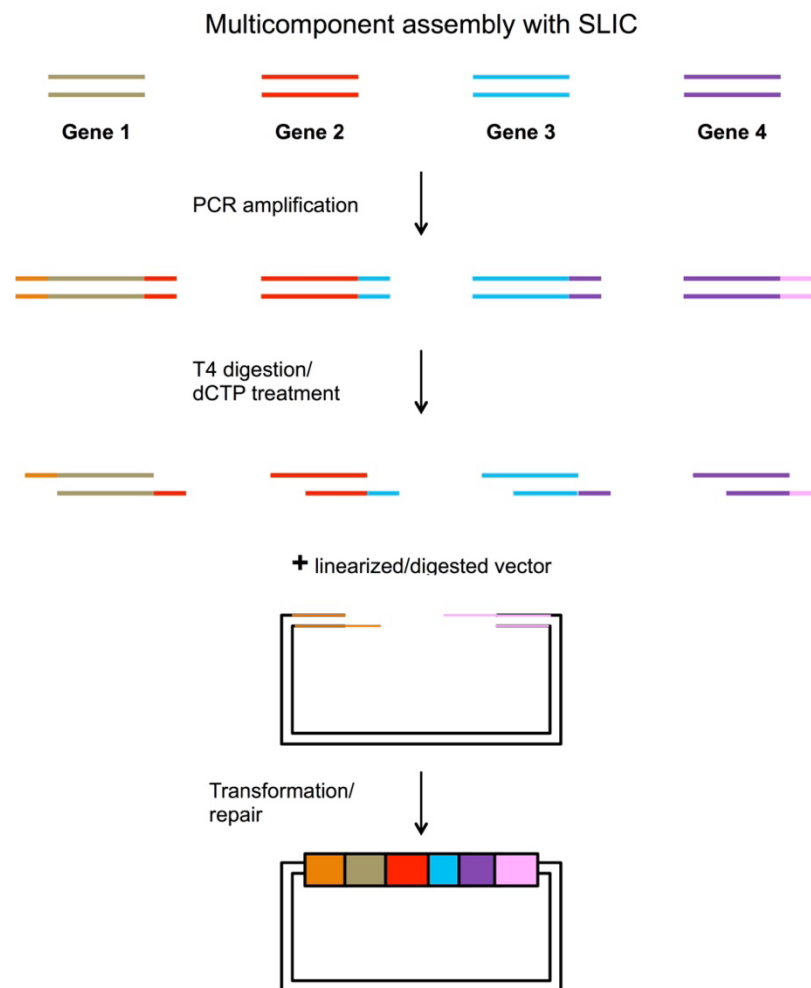
## Methods

Understanding how structural changes in TLR4 over evolutionary time have affected the ligand specificity and level of activation would be useful for developing pharmaceuticals to control inflammation. To understand which ligand causes each TLR4 protein to activate, we must place the desired TLR4 variant into an activity assay. The assay that has been developed for this purpose uses human embryonic kidney 293T (HEK 293T) cells. To eliminate species differences in protein interactions in the cytosol of human cells, we attached the ectodomain of each species' TLR4 to the transmembrane and TIR domain of human TLR4. We did this using SLIC, a molecular cloning technique. Next, to account for possible problems with the relative orientation of the extracellular to the intracellular domain for successful dimerization and activation, we used site-directed mutagenesis to shift the helix register at the interface of the ectodomain and transmembrane domain. In addition, we used site-directed mutagenesis to create the mutations of interest in zebrafish. This process allowed us to collect data on the ligand specificity of TLR4 from different species in the context of a human cell expression system.

### Building The Plasmid

To place the ectodomain of the species of interest onto the human cassette composed of a human transmembrane and TIR domain we needed to use SLIC. SLIC Cloning stands for sequence and ligation independent cloning, which allows for two or more sections of plasmids to be joined together. This process is useful as KLD mutagenesis can't join two or more plasmids together. If we wanted to test if the species of the TIR domain mattered in activating the NF- $\kappa$ B signal we needed to merge two plasmids together, one with its native ectodomain, and the other with the human TIR domain. This works by having single strand overlaps created by T4 DNA

polymerase, for each sequence you wish to add together, as you see in Figure 6. Then, when the plasmid fragments are transformed into bacteria, in this case XL-10 Gold super competent cells, the bacteria is able to ligate these fragments into a circularized plasmid and create copies of the plasmid during cell division<sup>19</sup>.



**Figure 6:** Diagram of how SLIC Cloning Works.

First the targeted sequences are multiplied through PCR. Next an overhang is made through T4 digestion and dCTP<sup>19</sup>

To modify these Franken-proteins to increase possible signaling we used point-mutagenesis methods. Point-mutagenesis is a molecular biology technique which allows you to alter the bases which make up DNA, which allows you to change the amino acid sequence in

proteins. This method uses a relatively small amount of the initial plasmid, which contains the gene of interest, and allows for insertions and deletions within the plasmid which are targeted via primers. These primers are small strings of bases which allow you to target your gene from within the plasmid's sequence by partially matching the bases contained within the template plasmid. Two methods of mutagenesis were used for this project, Quickchange Lightning, and KLD mutagenesis. The two are similar in mechanism, however, Quickchange Lightning uses overlapping primers, and KLD mutagenesis uses non-overlapping primers. Overall, a majority of the constructs created in this project used KLD mutagenesis, as this method had a higher tendency of producing colonies after being transformed into XL-10 Gold bacteria.

Primer Name	Primer Sequence (5' to 3')			
ins_6nt_after_45	5'-cacctgtcagatgaataagaataagaccatcattgggtgtgc-3'			
ins_6nt_after_45-antisense	5'-gacacaccaatgatggcttattcttattcatctgacaggtg-3'			

**Oligonucleotide information:**

Primer Name	Length (nt.)	Tm	Duplex Energy at 68 °C	Energy Cost of Mismatches
ins_6nt_after_45	42	78.19°C	-35.82 kcal/mole	23.95%
ins_6nt_after_45-antisense	42	78.19°C	-37.11 kcal/mole	21.81%

**Primer-template duplexes:**

Primer Name	Primer-Template Duplex
ins_6nt_after_45	5'-cacctgtcagatgaataagaataagaccatcattgggtgtgc-3'     atagtggacagtcacttattc-----tggtagtaaccacacagcca tatcacctgtcagatgaataag-----accatcattgggtgtgcggt
ins_6nt_after_45-antisense	3'-gtggacagtcacttattcttattctgtagtaaccacacag-5' 

**Figure 7:** Primers used in hTLR4\_hTLR4 +NK Mutant

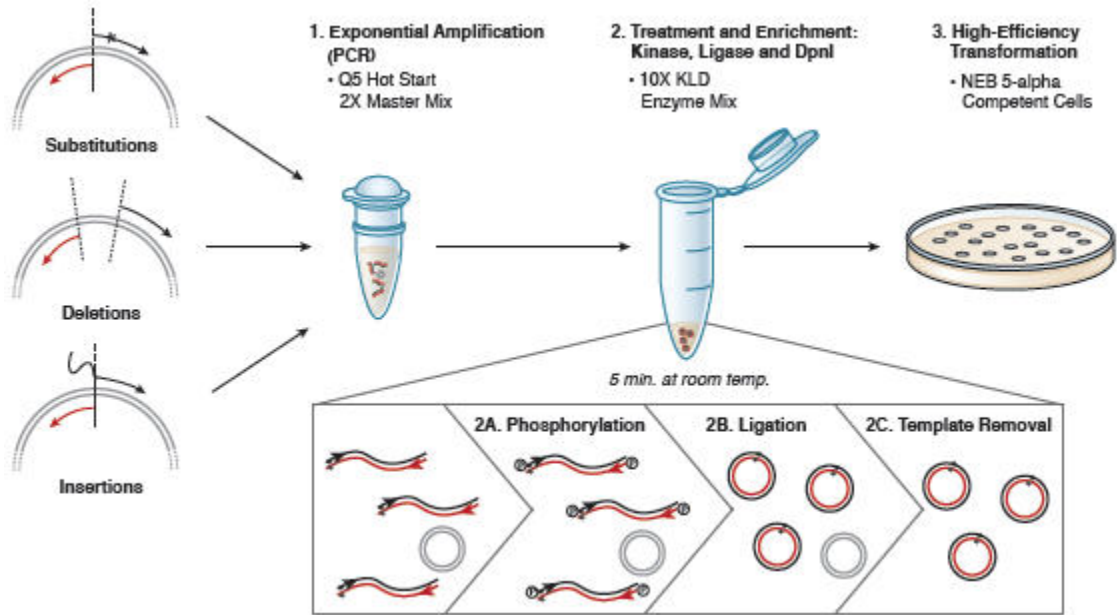
An example of the primer sequences used to create the hTLR4\_hTLR4 +NK mutants via KLD mutagenesis. Other primer compositions are listed in the further notes section.

*KLD Mutagenesis:*

KLD mutagenesis allows you to take template DNA and make point mutations to the bases to alter amino acid sequence. Different amino acids have different properties, such as glycine (small and non-polar) and asparagine (larger and polar). These two amino acids were switched in the G415N zebrafish TLR4 mutation to try and determine if altering those amino

acids would create more contacts between TLR4-MD2 sites, in turn altering the ligands which the protein interacted with. This was performed as the resulting NF-kB activity should be higher or lower if amino acids can/can't form the interactions which would hold the two proteins together. The more stable the TLR4-MD2 complex, the faster the dimerization of TLR4, and the more signal transmitted to the NF-kB pathway. In addition, the position at which the amino acids lie, particularly around the membrane matters a lot for protein function. KLD mutagenesis can remove certain bases from the sequence, which effects the placement, and therefore available interactions for the amino acids.

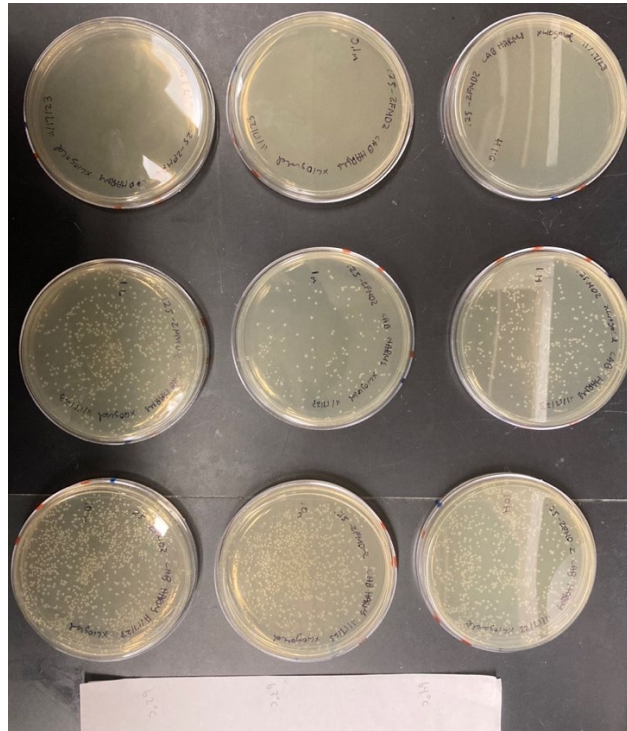
KLD mutagenesis starts with PCR amplification of your desired sequence, using your template DNA and primers. Next, the master mix containing kinase, ligase, and *dpn-1* is added to the PCR product. The kinase phosphorylates the end of the PCR product, which allows the ligase to bring the ends together to reform a circular plasmid. The *dpn-1* breaks down the template DNA still present, as both the product and template can be transformed into bacteria and contain the antibiotic resistance gene. By breaking down the remnants of the template you can ensure that more of the plated colonies will have the desired mutation<sup>20</sup>.



**Figure 8:** KLD Mutagenesis Process.

First, we design our primers with the sequence that we want and place it with a plasmid, which has a matching portion of a sequence. These ingredients are then mixed together and run through a PCR reaction. PCR, or polymerase-chain-reaction, is a process that allows for the sequence specified by the primers to increase in solution. These specified sequences are then phosphorylated, giving them sticky ends and ligated, allowing them to form full plasmids once more. Finally, the template, or plasmids without the modifications specified by the primers are removed from solution<sup>20</sup>.

## Best Conditions for KLD Mutagenesis



**Figure 9:** Results of varying PCR conditions after being transformed with XL10 Gold bacteria. The rows contain as follows: 0.1ul, 1ul, and 10ul of 1ng/ul template used. The columns correspond to the annealing temperatures used. The recommended annealing temperature was 63°C. The first column is the lower temperature at 62°C next to 63°C next to 64°C.

To obtain the proteins required for testing the zebrafish mutants and to make point mutations around the helix to perform a register shift we used KLD mutagenesis. As we needed a lot of point mutants, and colonies would often show odd mutations of repeating primers or random deletions when sequenced we needed to test what PCR process gave the most desirable colonies. Therefore, we took the D125K mutant and tried three conditions for both temperature and template concentration for the PCR. For temperature we wanted to try just above and just below the annealing temperature given by the NEB calculator to see if a more specific (lower temperature) or less specific (higher temperature) annealing temperature worked better. As the calculated temperature was 63°C for D125K we tried a 62°C, 63°C and 64°C annealing

temperature. For the template concentration we wanted to see how much template was needed to reliably produce colonies without having excess template that would not be broken down by Dpn-1, causing us to collect colonies without our desired mutation. Therefore, we used 0.1ul, 1ul, and 10ul of 1ng/ul of template, showing in Figure 9, going from top to bottom.

Looking at the transformation results in Figure 9 we can see that 0.1ng of template DNA was not sufficient for reliable colony growth. However, 1ng of template DNA gave many colonies without creating a bacterial lawn. This lawn-like growth is present in some parts of the 10ng plates. The range used for the annealing temperature did not seem to influence colony growth. However, when colonies from these plates were tested, we found that the plate with 1ng template and higher annealing temperature consistently had higher rates of containing the desired sequence. On the other hand, all of the 10ng template colonies showed the odd mutations that had haunted previous attempts of this procedure. The conclusion of this experiment was that template concentration had the largest effect on unintended mutations, with a high or normal annealing temperature being preferable.

### **Activity Assay**

Once we have the proteins which we wished to test, we need to place them into an activity assay to test their response to various LPS agonists in vitro. To collect data on the activity of TLR4, human embryonic kidney cells (HEK293T) were used. Plasmids encoding the mutated proteins under a constitutive, or always active, promoter were transfected into the HEK293T cells. The cells then express the mutant proteins and the immune response to certain LPS is measured using a luciferase-based assay. The Dual-Glo Luciferase assay system required us to transfect our cells with a plasmid containing the firefly luciferase gene under control of the

NF- $\kappa$ B transcription factor to report on NF- $\kappa$ B activation, as well as a plasmid with the *Renilla* luciferase gene with a constitutively active promoter to report on the viability of cells and their ability to express genes from transfected plasmids.

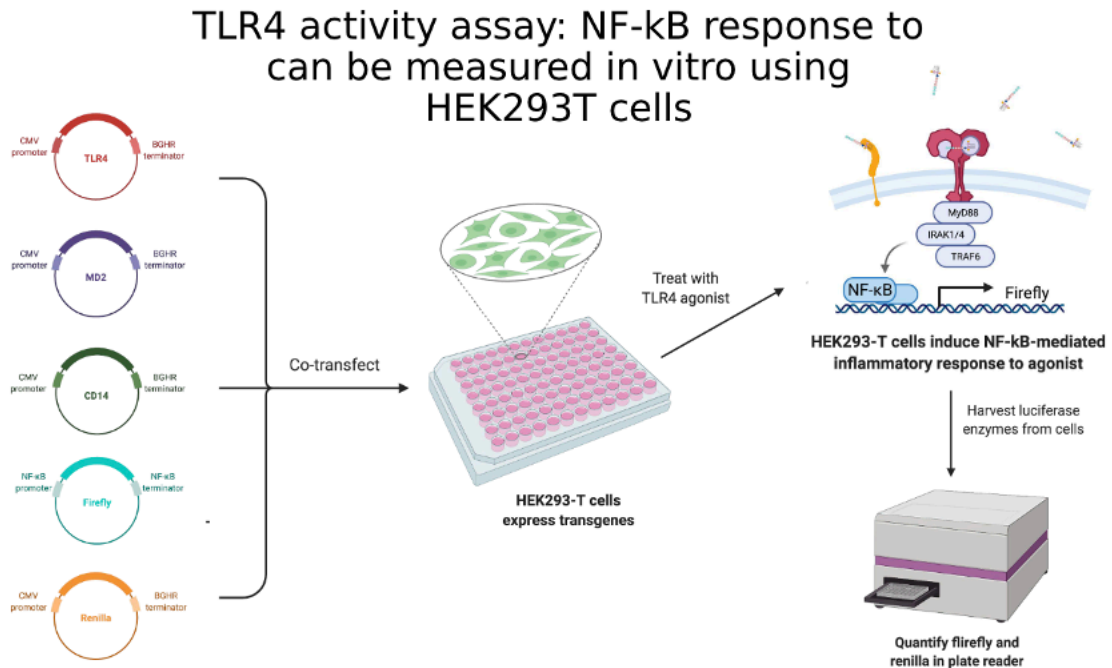
#### *Dual-Glo Luciferase Assay*

To transfect the cells for the Dual-Glo Luciferase Assay (Promega), we followed the protocol outlined by Loes, 2021<sup>11</sup>. HEK293T cells were grown in Dulbecco's Modified Eagle Media (DMEM) with 10% fetal bovine serum (FBS) in an incubator set for 5% CO<sub>2</sub> at 37°C. A confluent 100mm plate of cells was treated with 0.25% Trypsin-EDTA in Hank's Balanced Salt Solution to detach cells from the bottom of the plate and diluted to a 1:4 ratio with DMEM+FBS solution<sup>14</sup>. Then, 135ul of the cell solution was placed into each well in a 96 well plate.

All DNA solutions to be transfected were made in Opti-MEM with PLUS and Lipofectamine Reagents (ThermoFisher) to improve transfection efficiency in a total volume of 65ul for each well. Most transfection mixes contained 10ng TLR4, 1ng CD14, 20ng MD-2 along with 1ng Renilla, 20ng ELAM-Luc, and 48ng of empty vector pcDNA3.1(+) to bring the DNA mass to 100ng per well. However, the human control used 10ng TLR4, 0.5ng MD-2, 1ng CD14, along with 1ng Renilla, 20ng ELAM-Luc, and 67.5ng empty vector per well. Each TLR4 was paired with its matching co-factors. For instance, frog TLR4 (xenTLR4) was paired with xenMD-2 and xenCD14, except for zebrafish which do not natively have CD14, so mouse CD14 was used. In addition, the reporter assays included a row of vector-only transfections for Renilla background subtraction, which consisted of 100ng of pcDNA 3.1(+). Once the transfection mixes were placed onto the HEK293T cells they were allowed to recover for at least 20 hours before the treatment step was initiated.

All 200 uL of transfection media was removed from the cells in each well and replaced with 100 uL of treatment media made in 75% DMEM and 25% PBS. To test different TLR4 complex activity levels in response to LPS variants, each transfection type had three wells treated with 1ng LPS-R (except for human TLR4 which has a much higher response and therefore was treated with 0.1ng LPS-R for comparison's sake), three wells treated with 1ng Lipid IVA, and three wells treated only with buffer to establish a baseline. Once the treatments had been placed into their respective wells, the plate was placed into the incubator for three to four hours before luciferase activity was measured.

After incubation with TLR4 agonists, 60ul of media was discarded from each well in the plate and 30ul of Dual-Glo Luciferase Reagent, containing firefly luciferase substrate in lysis buffer, was added back. After a seven-minute wait step with the plate kept in the dark, the treated cells were scraped from the bottom of each well and 60 uL of cell solution from each well was transferred to an opaque plate for measurement. The luciferase activity levels were then quantified by measuring the luminescence in a microplate reader. After the firefly luciferase reading was taken, 30ul of the Dual-Glo Stop & Glo Reagent, containing firefly luciferase quenching reagents and *Renilla* luciferase substrate, was added to each of the wells and left to rest in the dark for seven minutes. Then, luminescence levels were measured and recorded.



**Figure 10:** Diagram of a General NF- $\kappa$ B Response Luciferase Reporter Assay

This diagram shows a visual representation of the Dual-glo Luciferase assay. First, the proteins of interest are transfected (inserted) into HEK293T cells which are left to grow for at least 20 hours. Then, the LPS of interest is placed in triplicate for each combination of TLR4 co-factors. The response to the LPS is then measured using a spectrometer and the relative signal as normalized to the human LPS-R levels is measured.

## Data Analysis

From the dual-glow assay there are two components which are measured, firefly and *Renilla* luciferase activity. The luciferase signal indicates the extent to which the pathway of interest is activated. A high value of luciferase indicates that TLR4 is consistently dimerizing, and inflammation would be observed. On the other hand, *Renilla* indicates the cell population of the measured well. The higher the signal the higher the rate of cell survival is and can indicate cytotoxicity. When there is a stronger luminescence reading for firefly luciferase with a stronger reading for *Renilla* luciferase we know the resulting signal is produced from activating the

pathway due to TLR4 dimerizing and promoting the NF-kB and luciferase reporter gene neighboring sequences. However, if the *Renilla* signal is low, it indicates a cell-wide impact, or cell death. Therefore, to understand our data we need to process the data. This way we only include luciferase signal which is due to activation of the reporter's pathway. To do this we simply divide the luminescence value of Luciferase by the value of *Renilla*, to get the degree of activation of the NF-kB pathway.

However, to analyze our data, we not only have to consider what *Renilla* tells us about background signal, we need to account for random activation that occurs without any TLR4 around. So, to begin processing the data we take the firefly signal for each well and subtract the average firefly signal from buffer treated controls with the same transfection mix. This gives the buffer-subtracted firefly signal. To get the background subtracted *Renilla* signal, we take our *Renilla* signal for each well and subtract the average *Renilla* signal from wells transfected only with pcDNA. This gives us our *Renilla* signal above background luminescence in the plate. Then to see how much firefly is being produced from cells expressing protein, we take the firefly signal and divide it by the *Renilla* signal for each well. To compare between datasets, we normalize each measurement to human TLR4 treated with 0.1ng LPS-R by dividing the firefly/*Renilla* value for each well by the average of the firefly/*Renilla* value for human + 0.1ng LPS-R. Finally, we plot our data to assess relative TLR4 complex activity levels and their ligand specificity.

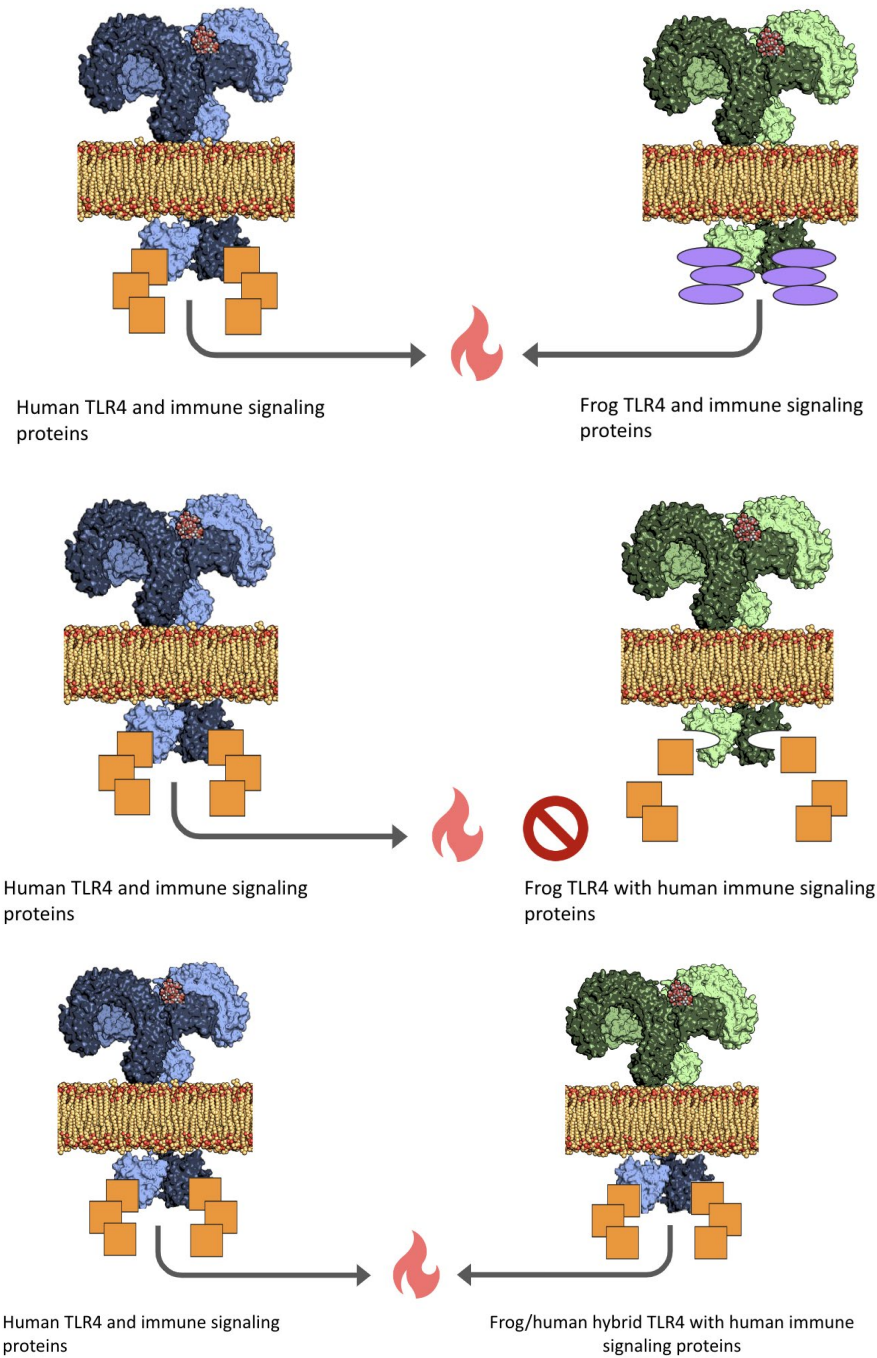
## Results

In previous experiments performed in the Harms lab, activity levels for other species' TLR4 when expressed in HEK293T cells were consistently lower than that of human TLR4. We theorized that this lack of signal may have been due to an internal signaling pathway mismatch of the TIR domain of the TLR4 protein and the downstream internal cellular proteins which activate the NF- $\kappa$ B pathway. Furthermore, we decided to test areas of interest within zebrafish TLR4 and MD-2 to see if a single point mutation was sufficient to increase signal or switch ligand specificity.

### **Xenopus (Frog) TLR4 Activity in HEK293T cells**

It would be very useful to characterize the TLR4 of certain species such as frogs, which seem to show signal, but not enough to be significant when error is accounted for. Therefore, this project aimed to resolve this issue, and increase signal strength and characterize these ambiguous proteins.

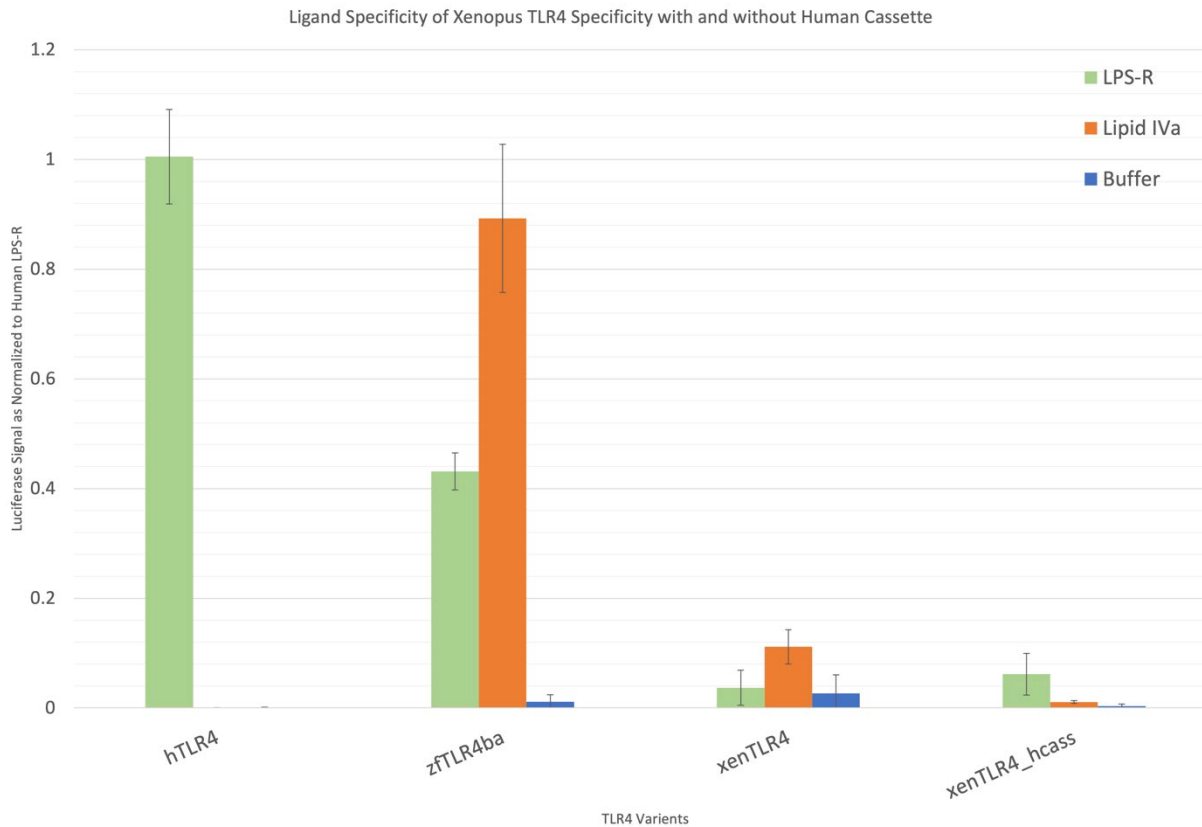
As frog TLR4 is from a distant species and was expressed in HEK293T cells a theory was that the internal proteins of the structure weren't able to effectively transduce the signal to activate the NF- $\kappa$ B pathway. To fix this issue the ectodomain of the low-signal TLR4 from frog was placed onto the transmembrane and TIR domains of the human TLR4 protein then transfected in with all the normal frog TLR4 co-factors.



**Figure 11:** Diagram of Theorized Human vs Frog TLR4 NF-kB Pathway Activation

A sequence of figures explaining the theory that the internal proteins of humans and frogs weren't similar enough to transduce the signal of LPS entering the system to activate the NF-kB pathway.

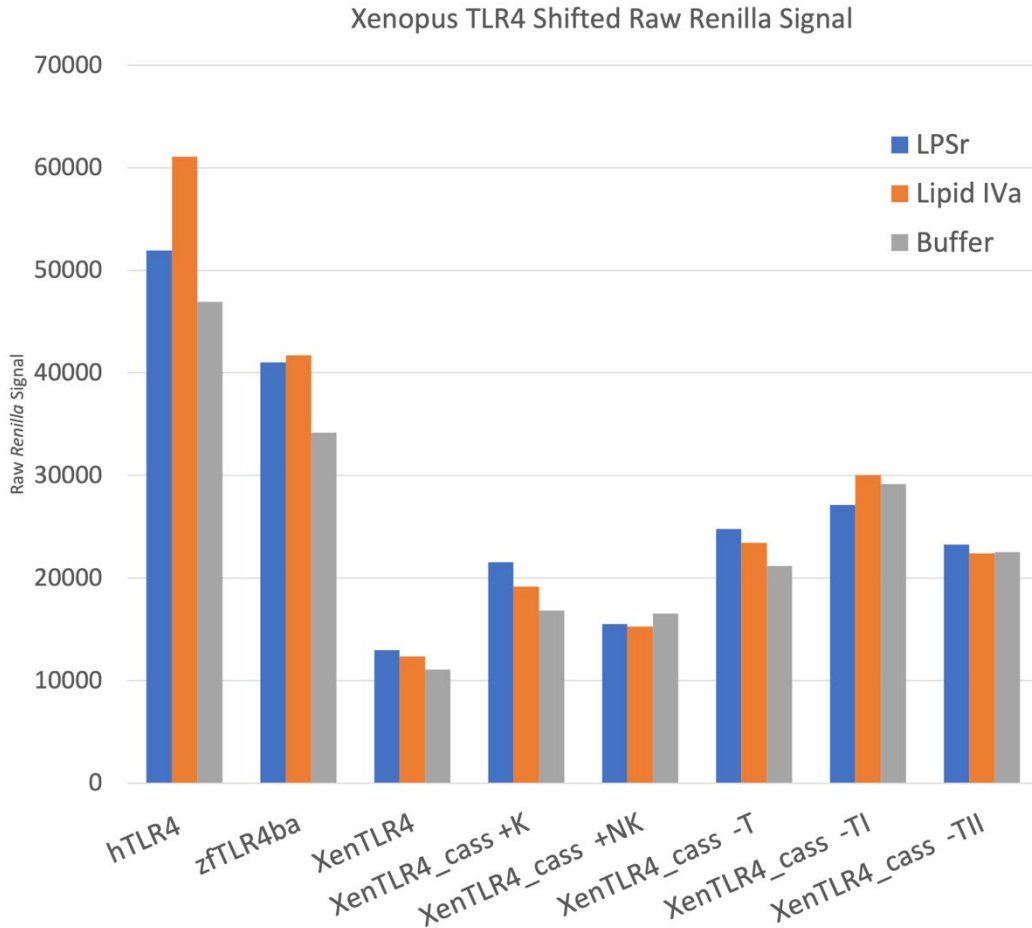
Therefore, we created a frog-human TLR4 hybrid through SLIC cloning and the activity was tested in a dual-glow luciferase activity assay using HEK293T cells.



**Figure 12:** Ligand specificity of *Xenopus* TLR4 When Placed into a Human Cassette

Luciferase activity of *Xenopus* TLR4 and co-factors (xenMD-2 & xenCD14) with and without a human cassette replacing the transmembrane domain and intracellular domain. Data based off of three bio-replications.

However, when we placed the human internal domains onto the *Xenopus* TLR4 we saw that there was a decrease in signal. It is important to note that the average Renilla signal over three biological replications was lower for *Xenopus* TLR4 and higher for *Xenopus* TLR4 in a human cassette. The higher *Renilla* signal indicates a lowered cytotoxic effect when the *Xenopus* TLR4 is placed into HEK293T cells.

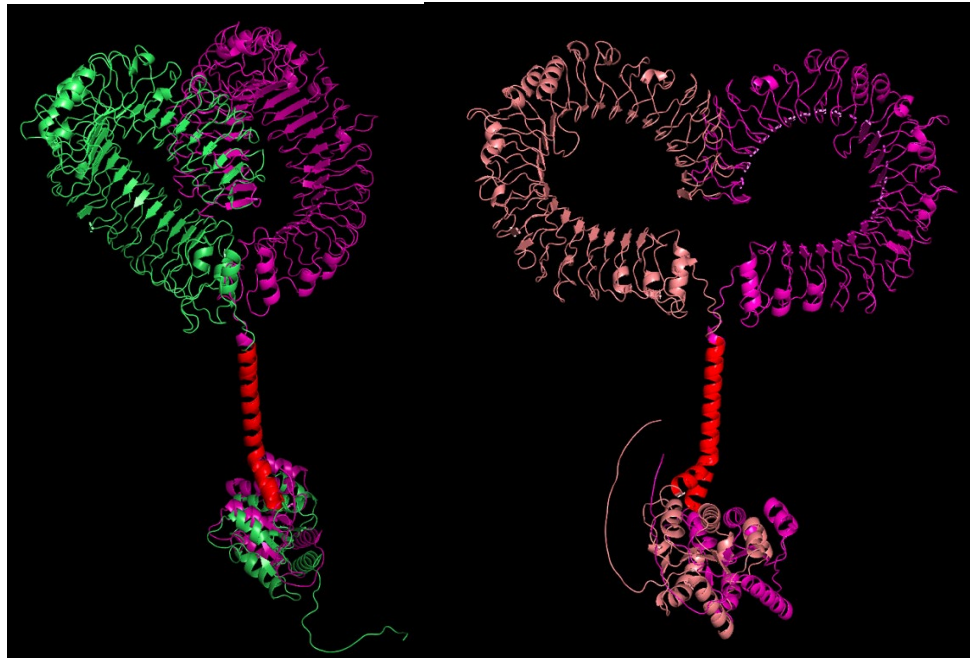


**Figure 13:** *Xenopus* TLR4 Shifted Raw Renilla Signal

This bar graph shows the relative Renilla values for the helix shifted *Xenopus* in and out of the cassette. The graph shows that the Renilla values increase when placed into a human cassette, but, once more decline when the hydrophobic region starts to be removed.

A graduate student at the University of Oregon, Andrew Holston had presented research on helix shifting, the idea that by adding or taking away amino acids which make up alpha-helix structures you can move the positions of binding domains in proteins. So, we hypothesized that when we spliced the two proteins together there was a change in position of the exodomain which was decreasing TLR4/MD-2 dimerization. To ensure that there was no mismatch in the orientation between the domains we added and removed 2~3 amino acids from around the splice

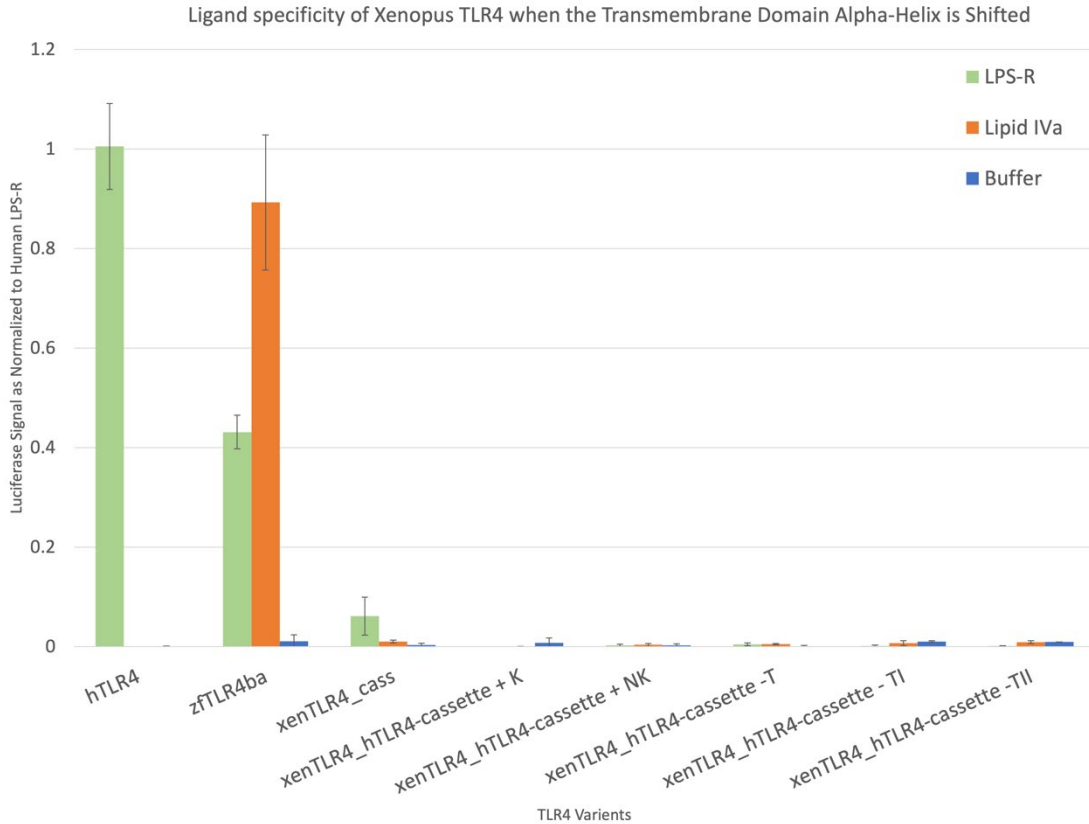
point. As one rotation of an alpha helix takes 3.6 amino acids this would allow us to see if the rotation of the alpha helix would affect the dimerization of the complex, and subsequent signal<sup>21</sup>.



**Figure 14:** 3D models of what helix shifting may do to the orientation of TLR4.

3D models were obtained through use of alpha fold co-lab and rendered using pymol. (-TII xenTLR4 mutant as compared to normal xenTLR4 shown on left, and +NK xenTLR4 mutant compared to normal xenTLR4 shown on right.) The magenta structure shown in comparison to each helix-shifted mutant is the normal xenTLR4.

This figure highlights the effect of altering the alpha-helix on the ectodomain's location. When a helix-shifted mutant like the -TII xenTLR4 (shown by the green structure) is aligned with a normal xenTLR4 (shown as the magenta structure) we notice that the ectodomains are in a similar, but not quite identical formation. However, when we look at how a +NK xenTLR4 (shown by the salmon structure) looks when the transmembrane domain is aligned with normal xenTLR4 we see that the ectodomains are in completely altered positions. We hypothesized that when the hTLR4 and xenTLR4 were placed together this changed the ectodomain location and is therefore had a negative effect on dimerization and resulted in a decreased signal.



**Figure 15:** Luciferase Readings of Frog TLR4 Ligand specificity when Helix Shifted

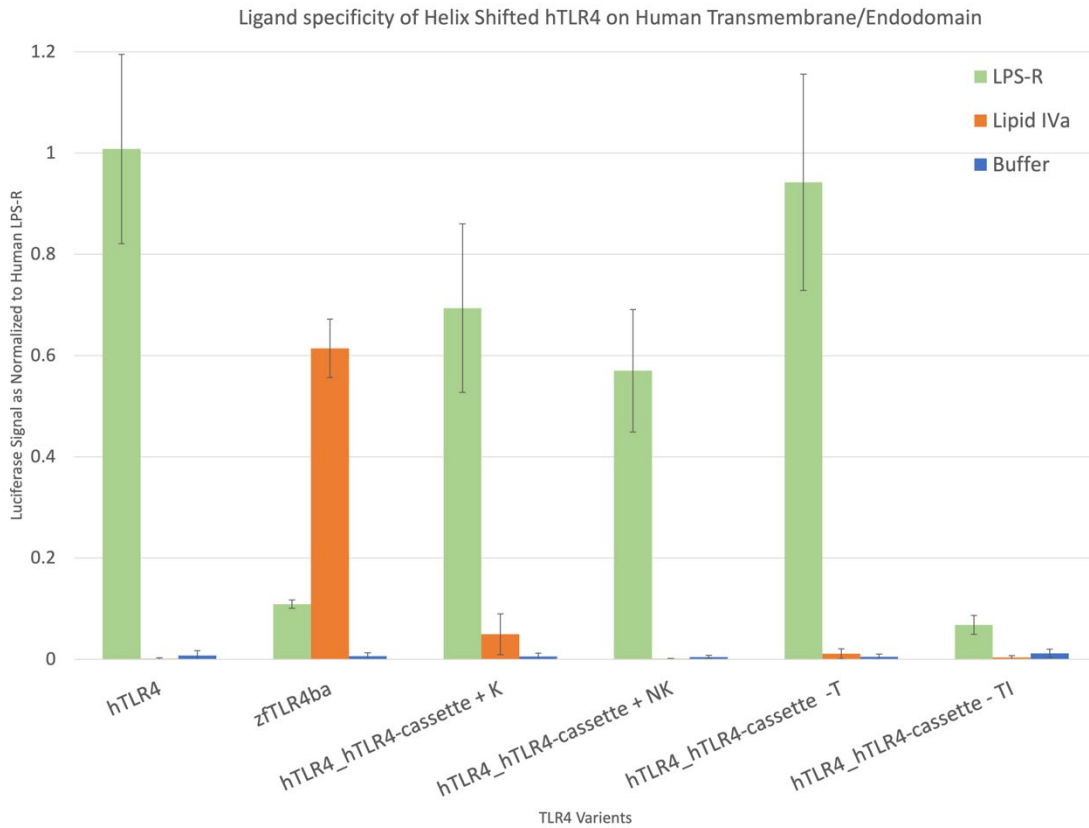
Luciferase Readings of Frog TLR4 Ligand specificity when Helix Shifted *Xenopus* (Frog) TLR4 Activity in HEK293T Cells when Placed into a Human TM/TIR Cassette. This graph shows the average signal for hybrid *Xenopus*/human TLR4 (xenTLR4\_hTLR4-cassette) when exposed to LPS-R Lipid IVa and buffer. The data is normalized to human TLR4 (hTLR4) and uses hTLR4 and zebrafish TLR4 (zfTLR4) as a basis to determine that the ligand can activate the NF-kB system. Data based off three bio-replications.

Unfortunately, even with the addition and deletion of amino acids around the alpha helix we did not see an increase in luciferase signal. This would indicate that simply shifting the alpha-helix around does not significantly increase signal in xenTLR4\_hcass.

### Testing the Effects of Helix Shifting on Human TLR4

After finding relatively no change in *Xenopus* TLR4 despite the addition of a human TIR domain we wanted to test if helix shifting around the TIR domain had any effect at all. As we

had little activation no matter what we did on *Xenopus* TLR4 we wanted to test the theory on human TLR4 which has a larger baseline signal to observe.



**Figure 16:** Luciferase Readings of Helix Shifted hTLR4

Ligand Specificity of Helix Shifted hTLR4 around its TIR domain. In this activity assay up to two amino acids were added or deleted to observe if the shift in helical structure had an effect on the protein's response to LPS.

When the ectodomain was moved, it did not seem to have any significant effect on the resultant signal. Only when a hydrophobic amino acid was removed from the alpha-helix did we see that the signal decreased. Interestingly we also see a small increase in buffer signal which may act as an indicator of an altered transmembrane domain in other constructs.

### **TLR4 Homolog; CD180 Activity in HEK293T cells**

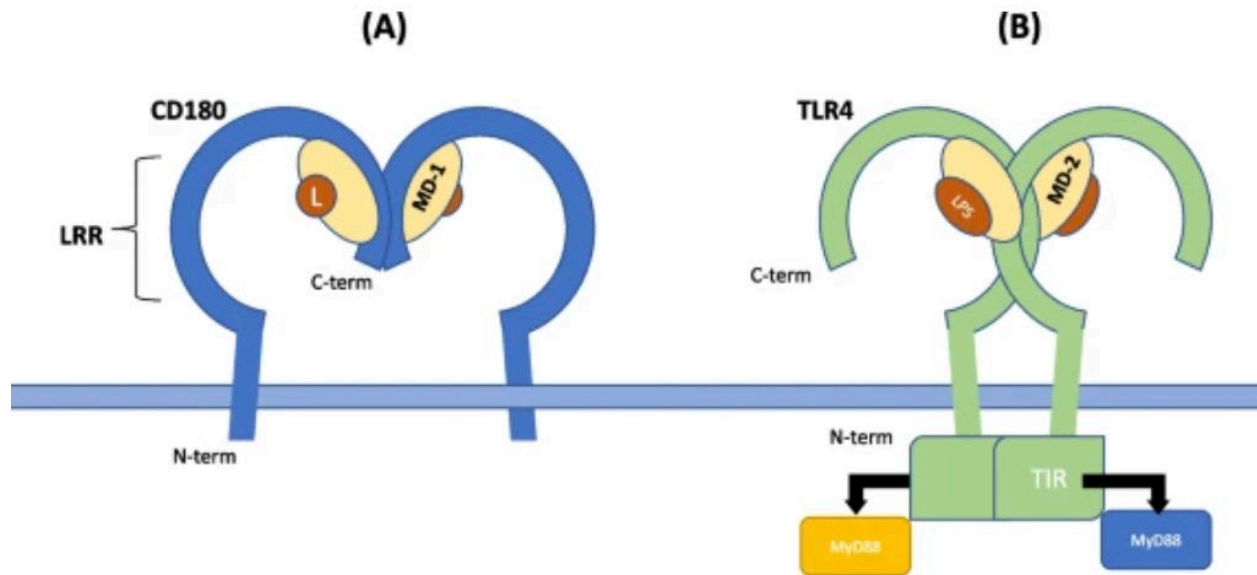
Early in vertebrate evolution, there was a point of divergence, likely a whole genome duplication, which created a duplicate of TLR4 without the TIR domain, known as CD180 or RP105, and a duplicate of MD-2, called MD-1<sup>22</sup>. These proteins form a complex and are thought to possibly bind LPS, similar to TLR4 and MD-2, and are part of the immune system<sup>22</sup>. As CD180 is a homolog of TLR4 we were interested in testing if there was a response that would occur in the presence of LPS if it had a TIR domain to transduce a signal. So, through SLIC cloning using the same cassette as described above, a human transmembrane and TIR domain were attached to the ectodomain of CD180 and tested in a Dual-Glo activity assay. We tested two variations of CD180, from human and zebrafish, with both their respective MD-1 and MD-2 co-factors.



**Figure 17:** Luciferase Readings on the Ligand Specificity of CD180 in a Human Cassette

Ligand Specificity of CD180 with Human TIR domain. This graph shows combinations of CD180 in cassette tested with both MD-1 and MD-2.

Human CD180 generally showed no response at all, even with the TIR domain attached. Although zebrafish CD180 had a larger signal than human CD180 the error bars essentially cover the whole signal, suggesting that the difference is not statistically significant. This activity assay placed both the literature agreed co-factor to CD180, MD-1 and TLR4's co-factor MD-2 due to a consideration into how CD180 dimerizes when compared to TLR4.



**Figure 18:** (Figure 1 from Edwards et. al.) Shows the variation in structure between CD180 and TLR4<sup>22</sup>.

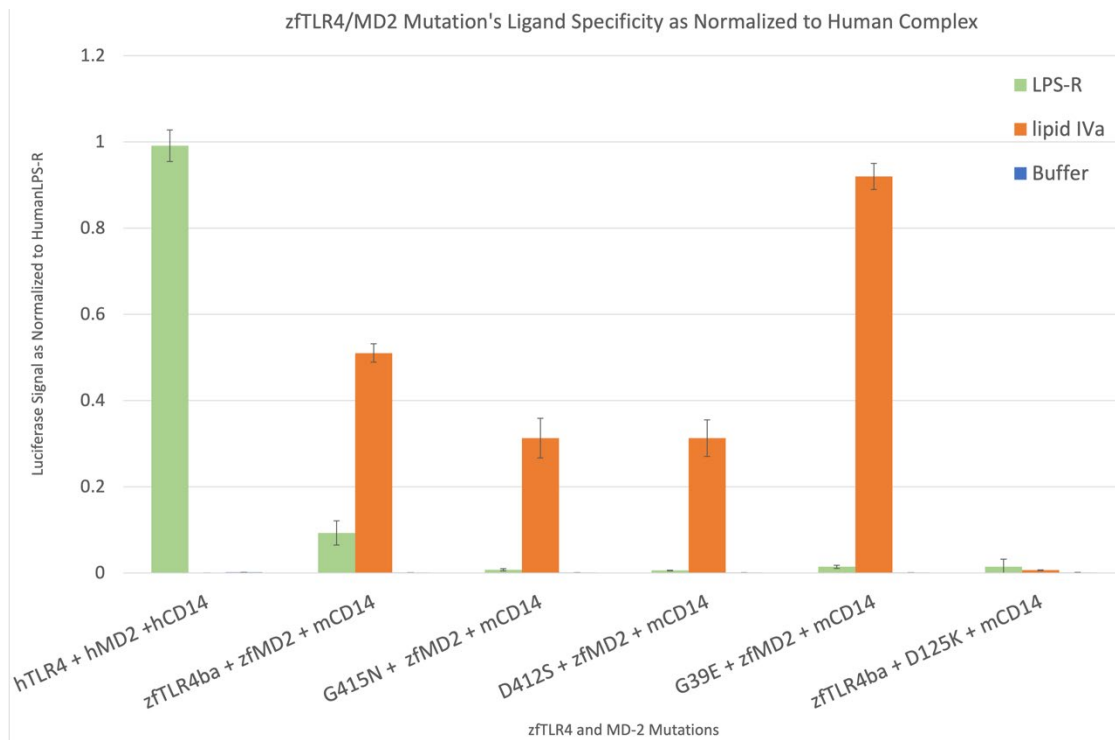
This highlights that the TLR4/MD-2 complex binds in a different region and conformation than CD180/MD-1 complex.

Therefore, it was hypothesized that even if CD180 dimerized in the assay using MD-1 the intracellular portions wouldn't be able to interact. So, if there remained sufficient binding interfaces for the MD-2 to drive the dimerization in CD180 it could dimerize in a manner which allowed the intracellular portions of the protein to interact with the internal proteins which transduce the signal to the NF-kB pathway.

From this data we can assume that the two homologs, TLR4 and CD180 have diverged too far, and CD180's ectodomain is no longer able to complete the steps necessary for NF-kB activation. Perhaps a future project could be to preform ASR (Ancestral sequence reconstruction) to see if an earlier point in the divergence of the two homologs had similar activity to their TLR4s or activated at all.

## **TLR4 and MD2 Mutants to Change Ligand Specificity in Zebrafish**

The final section of this project involved looking at the effect of single point mutations in zebrafish TLR4 and MD-2. These mutations were chosen due to previous literature noting the positions effect on ligand specificity and binding interfaces between TLR4 and MD-2. To see if a point mutation was sufficient for switching ligand specificity in zebrafish, we wanted to swap the amino acid in those positions to the amino acids that humans expressed in those positions. By switching these amino acids, we hypothesized that there would be an increase in the luciferase signal in response to LPS-R. Ultimately, not all of the mutants that were designed were collected by the end of this project. We were able to create and sequence G415N, D412S, G39E, (TLR4 mutations) and D125K (MD-2 mutation).



**Figure 19:** Luciferase readings for zfTLR4 & zfMD-2 mutants normalized to human complex. Data based on two biological replications. All x-axis labels are organized by TLR4+MD-2+CD14. As zebrafish are missing native CD14 we used mouse CD14, as the mouse complex is known to respond to all LPS.

Looking at this graph we can see that none of the point mutations were sufficient to alter the ligand specificity. However, we can see that signal strength was increased in the G39E mutation and decreased in both the G415N and D412S mutations. The singular MD-2 mutant we were able to create was quite interesting as the amino acid substitution seemed to break the operation of the complex.

## Discussion

### Overview

Understanding how proteins evolve helps us understand how a protein functions. Therefore, to better understand how TLR4 dimerizes and activates inflammation we need to understand how the protein has evolved to interact with different LPS over evolutionary time. To know which LPS each species' TLR4 responds to we placed them into HEK 293T cell assays. However, certain TLR4 variants like *Xenopus* do not have a strong enough response to determine what LPS they respond to. So, we theorized that the issue that was causing a low signal response was an internal signaling mismatch. To ensure that the TLR4 proteins could transduce the signal to the internal pathway we spliced together the ectodomain of the species of interest and placed it onto a human transmembrane and TIR domain.

The data collected shows that simply placing a human TIR domain onto a separate species' ectodomain does not increase signal. Interestingly, conjoining the two sets of domains may not directly increase the signal, however it does increase the *Renilla* signal produced by the protein containing cell. This indicates that the cytotoxicity of the protein is lower when in cassette verses out of it.

Therefore, we hypothesized that the cassette was a step in the right direction, and that there was a structural issue that was introduced when we spliced the two protein's domains together. The alpha helix that was adjacent to the region where we spliced the proteins together had the potential to shift the rotation of the entire ectodomain of the TLR4 protein. So, we tried adjusting the rotation by adding and subtracting amino acids around the area. Ultimately, shifting the helix didn't have a major effect on the activity of the protein. Rather, the main effect we

observed was that the size of the hydrophobic region going through the membrane was sensitive to alterations, and severely decreased the signal response to deleting the non-polar amino acid.

This human TLR4 transmembrane and TIR domain cassette tool was used to explore the ligand specificity of the TLR4 homolog, CD180, which is missing the signal transducing TIR domain. CD180 is interesting to look at in the context of TLR4 evolution as it has a shared ancestor, yet the function of the CD180 complex has completely diverged. The specific contacts which allow the CD180 to have a novel function can tell us how we can alter TLR4 activity in the body. So, to find the ligand specificity we placed the human transmembrane and TIR domain onto the CD180 ectodomain in a HEK293T cell assay. Similar to the *Xenopus* TLR4 the CD180\_hTLR4 constructs did not have a sufficient activity to determine their ligand specificity.

Finally, we wanted to test whether we could alter zebrafish TLR4 ligand specificity by replacing certain amino acids in TLR4 and MD-2 with residues determined to be important for specificity from studies in other organisms. When we placed the mutated proteins into an activity assay, we saw that none of the mutations acted as we expected, and only resulted in modifications in signal strength rather than alterations in ligand specificity. This is particularly evident in the mutants G39E (zfTLR4) and D125K (zfMD-2) where we saw both a major increase and decrease in signal strength.

In the end, the data collected throughout this project shows that point mutations occurring at residues of interest make clear differences in signal strength. However, our data did not show that a single point mutation is sufficient for changing ligand specificity in zebrafish. In addition, simply placing a human transmembrane and TIR domain onto a low-signal TLR4 ectodomain is not sufficient for increasing signal strength. It is important to note that placing the *Xenopus* TLR4 into cassette did lower cytotoxicity in those transfected cells, so the cassette may be a

viable component for increasing signal strength. However, we now know that shifting the alpha-helix going through the transmembrane domain does not strongly affect subsequent TLR4 activity. Information about the effects of a human cassette on the ectodomains of various TLR4s, and if the rotation of the alpha-helix effects TLR4 activity can influence the steps we take in the future to measure the ligand specificity of these difficult proteins.

### **Future directions**

This project may have covered a lot of ground but there's so much left to explore. First, this project was unable to explore some of the topics that were originally in discussion due to difficulties encountered performing mutagenesis. One of the casualties of the mutagenesis process were the L441F, M70R, and H45Y zebrafish mutations that were not grown. As the other mutations that were tested had effects different than what would be expected by literature it would be worthwhile to measure the effects of these mutations on TLR4 activity.

However, our main goal is to fill in our understanding of the evolution of TLR4. To do this, there are several TLR4 complexes that remain uncharacterized. Our lab has reconstructed the ancestors of tetrapod, teleost, and bony vertebrate TLR4 complexes using ASR, ancestral sequence reconstruction. Ancestral TLR4 complexes for the bony vertebrate, tetrapod and teleost ancestors were resurrected by other lab members using Topiary, a bioinformatic and phylogenetic tool described in Orlandi, 2023<sup>23</sup>. Characterizing these ancestral complexes will help us to understand what sequence changes in evolutionary history have defined the ligand specificities of present day TLR4 complexes in different species. A future direction for this project would be to place the teleost and bony vertebrate ancestral TLR4 ectodomains into the human transmembrane and TIR domain cassette to see if a species mismatch could be the reason why we see low signal for these complexes.

A future project could focus on species that are closely related to *Xenopus* which have a full native TLR4 complex, which includes TLR4, MD-2 and CD14 proteins. Two species have been identified by the harms lab to fit these conditions: Pike and Caecilians. By determining their ligand specificity, we would help to clarify if LPS-R or Lipid IVa recognition is the preserved ancestral trait, or if LPS variant recognition evolves very rapidly to respond to the environment.

## Glossary

**Acyl Chains:** A carbon chain attached to an acyl moiety which contains an R-group attached to a carbon double bonded to an oxygen and a functional group. The number of acyl chains in the lipid A portion of LPS is important for TLR4 dimerization.

**Adaptive immunity:** Activates second to the innate immune system. Uses antibodies which bind and destroy invading pathogens. A more targeted response than innate immune system but takes time to construct the antibodies.

**Amino acids:** Building blocks for proteins composed of a carbon backbone, an amino group, a carboxyl group, and an R-chain which differs between the twenty different amino acids.

**Antigen:** A molecule which binds to a specific antigen receptor which is produced by immune cells such as B and T cells. Antigens can be proteins, polysaccharides, lipids, or nucleic acids.

**CD14:** A co-factor in TLR4 activation, it can exist in a soluble or membrane bound form. CD14 binds to LPS and delivers it to the MD-2 TLR4 complex to induce dimerization.

**CD180:** A homolog of TLR4 which diverged and lost its endodomain. This protein remains functional in the immune system recognizing LPS and helping to activate B-cells.

**Conserved Sequence:** Portions of the genome are similar in amino acid sequence across species. Indicates that selection has preserved the sequence.

**Cytokine:** Small proteins which are important for secondary signaling, secreted as part of the immune system.

**DAMP (Damage Associated Molecular Pattern):** Molecules released from the organism due to damage. For example, S100A9 which activates the TLR4 based pathway.

**Dimerization:** Process by which two molecular subunits are joined together.

**DNA:** Deoxyribonucleic acid, a polymer that contains the genetic code, which determines the function, growth, and reproduction of organisms.

**Endodomain:** The portion of a transmembrane protein which lies on the inside of the cell membrane.

**Evolution:** A directional change in a species or population over time.

**Ectodomain:** The portion of a transmembrane protein which lies on outside of the cell membrane.

**Genotype:** The genetic information that makes up an organism. Also, may refer to the sequence which encodes for a specific gene.

**Gram-negative Bacteria:** Bacteria which have a special cell envelope, which contain the LPS which activates the TLR4 based NF-kB pathway. The cell envelope provides a high antibiotic resistance to these bacterial strains.

**Hydrophilic:** The molecular property which makes it favorable for the molecule to interact with water. Typically, the molecule is polar or charged.

**Hydrophobic:** The molecular property which makes it unfavorable for the molecule to interact with water. Typically, the molecule is hydrophobic.

**Innate immunity:** The body's non-specific response to pathogens. Connected to inflammation, and physical protection, such as skin.

**Ligand:** A small molecule which binds to another molecule, typically as a component of a pathway or reaction.

**Ligand Specificity:** Describes the type and strength of preference for a molecule to bind to a ligand. For instance, to describe hTLR4's ligand specificity we would note that it reacts strongly in the presence of LPS-R, and not so much Lipid IVa.

**LPS (Lipopolysaccharide):** A component of Gram-negative bacterial cell membranes, and the ligand which causes TLR4 to dimerize. Components of LPS (O-antigen, core oligosaccharide, and Lipid A) can vary depending on the bacteria's strain and species. This variation causes TLR4 to dimerize only to specific LPS.

**Firefly luciferase:** An enzyme derived from fireflies which catalyzes the oxidation of a luciferin, causing light to be exuded. Used as a reporter of desired pathway activation in a Dual-Glo assay.

**MD-1:** A co-factor of CD180 which brings LPS for cell signaling as part of the immune system.

**MD-2:** A co-factor of TLR4 which brings LPS for dimerization of TLR4, and subsequent activation of the NF- $\kappa$ B pathway to occur.

**Model Organisms:** Organisms which help scientists to study biological systems. Typically, the traits of these species include being able to mature rapidly, having lots of offspring, having a sequenced/well understood genome, and have some trait which makes them ideal for the system of study. For instance, many labs on the Knight campus use rats in their bioengineering work as the grafts and gels are easier to implant than in mice.

**Moiety:** A chemical group which possesses specific elements arranged in certain manners. For instance, a hydroxyl group is a moiety which looks like -OH.

**NF- $\kappa$ B Pathway:** A pathway activated in response to PAMPs or DAMPs which releases inflammatory cytokines. The main pathway for activating inflammation in mammals.

**PAMP (Pathogen Associated Molecular Pattern):** Molecules which are associated with pathogens which generally contain a moiety which is recognized PAMP-receptors such as TLR4. The PAMP for TLR4 is LPS.

**Phenotypes:** The physical traits that result from the genotype.

**Proteins:** Composed of amino acids, proteins are vital for the function, structure, and regulation of the body. TLR4 is a protein which is around 848 amino acids long.

**Renilla:** *Renilla* luciferase is an enzyme which comes from *Renilla reniformis*, a coral called the sea pansy. Used in conjunction with firefly luciferase in a Dual-Glo assay, *Renilla* helps to measure activation of the pathway when there is no signal that should be occurring. Using *Renilla* helps to normalize data from the assay.

**Sepsis:** Caused by a cytokine cascade, sepsis occurs when bacteria enter the bloodstream and results in overactivation of inflammation. This destroys the body from the inside, harming organs and raising the body's temperature to dangerous levels. Sepsis is associated with many medical outcomes and is the cause of 1 in 5 deaths according to the WHO.

**TLR4 (Toll-like Receptor 4):** A transmembrane protein which induces inflammation through the NF- $\kappa$ B pathway as part of the innate immune system. TLR4 is highly conserved across evolution yet has differing ligand specificities which depend on the species.

**Transmembrane domain:** The portion of a transmembrane protein which lies in the cell membrane. As such the amino acids which make up this portion of the protein are hydrophobic.

**TIR:** Toll/interleukin receptor, the portion of a TLR4 protein responsible for the propagation of downstream signals.

## Further Notes:

Table of Primer Compositions for Mutagenesis/KLD

Resulting Plasmid	Starting Plasmid	Forward Primer	Reverse Primer
hTLR4_hTLR4 +K	hTLR4_pcDNA3	5'-cacctgtcagatgaataagaagaccatcattggtgtgc-3'	5'-gacacaccaatgatggctcttattcatctgacaggtg-3'
hTLR4_hTLR4 +NK	hTLR4_pcDNA3	5'-cacctgtcagatgaataagaataagaccatcattggtgtgc-3'	5'-gacacaccaatgatggctcttattcatctgacaggtg-3'
hTLR4_hTLR4 -T	hTLR4_pcDNA3	5'-aagatcattgggtgtcggg-3'	5'-gatcttattcatctgacaggtgatattc3'
hTLR4_hTLR4 -TI	hTLR4_pcDNA3	5'-aagattgggtgtcggc-3'	5'-aatcttattcatctgacaggtgatattcaaac-3'
Xen_TLR4 +K	XenTLR4_pcDNA3	5'-gaccctgaactgtccatgacaagaagaccatcattggtg-3'	5'-caccaatgatggtcttgcattggaacagttcagggtc-3'
Xen_TLR4 +NK	XenTLR4_pcDNA3	5'-gaccctgaactgtccatgacaagaagaccatcattggtg-3'	5'-caccaatgatggtcttgcattggaacagttcagggtc-3'
Xen_TLR4 -T	XenTLR4_pcDNA3	5'-gaccgacacaccaatgattgcatggaacagttcag-3'	5'-ctgaactgtccatgacaatcattggtgtcggc-3'
Xen_TLR4 -TI	XenTLR4_pcDNA3	5'-aggaccgacacaccaatgattgcatggaacagttcag-3'	5'-tgaactgtccatgacaatcattggtgtcggc-3'
Xen_TLR4 -TII	XenTLR4_pcDNA3	5'-accctgaactgtccatgacaggtgtcggg-3'	5'-accgacacacctgcatggaacagttcagggt-3'
ZfTLR4_ba-D412S	ZfTLR4_ba	5'-ageGTGGGAGGATTTGAAG-3'	5'-AACAGAAATTTCTGAGTTTAGG-3'
ZfTLR4_ba-G415N	ZfTLR4_ba	5'-aacTTTGAAGGACTTGATTTCG-3'	5'-TCCCACATCAACAGAAATTC-3'
ZfTLR4_ba-L441F	ZfTLR4_ba	5'-tttTCCAACCTAAAGAATCTGAG-3'	5'-AACAGAAAGGTATCCAATG-3'
ZfTLR4_ba-G39E	ZfTLR4_ba	5'-gaaAGAAACCTCAGCAGC-3'	5'-CATGCATGAGTAATGAAGATTC-3'
ZfMD2-D125K	ZfTLR4_ba	5'-aaaGGCCCTGAGGGAGAG-3'	5'-AATCTTCCTTATAATGACTGGGAAAG-3'
ZfMD2-H45Y	ZfTLR4_ba	5'-tatTACCTGTCAATGACTGTTTC-3'	5'-CAATGGTCCCTCAAAGG-3'
ZfMD2-M70R	ZfTLR4_ba	5'-cgcTTCACAATCACTCGTTTTG-3'	5'-TGGGATTAGCGTCAAATTTATAAAG-3'

Table of SLIC/HIFI Reactions and their Primers

Resulting Plasmid	Starting Plasmid	Forward Primer	Reverse Primer
*all TIR hTLR4 inserts use these primers	hTLR4_pcDNA3	5'-accatcattgggtgtcggc-3'	5'- CTTGTCGCATCGTCTTTGTAGTctggctcaccagga- 3'

Xen_hTLR4	hTLR4_pcDNA3 & XenTLR4_pcDNA3	5'- ccagactacaagacgatgacgacaagCTGAATGGCTG CATCGAGGAGAT-3'	5'- cacactgaggaccgacacaccaatgatggtTGTTCATGGAACAG TTCAGGGTCAG-3'
ZfCD180_hTLR4	hTLR4_pcDNA3 & ZfCD180	5'-GCAACACCATGGGCGAG-3'	GTCACATTGCAGTTTAGCATCAATAATCCT-3'
hCD180_hTLR4	hTLR4_pcDNA3 & hCD180	5'- ccagactacaagacgatgacgacaagTGGGATCAGAT GTGCATTGAG-3'	5'- cacactgaggaccgacacaccaatgatggtCCCACAGGAAAGC TTGACATC-3'
Esox-Lucius_hTLR4	hTLR4_pcDNA3 & Esox-LuciusTLR4	5'- gacgatgacgacaagCAGGTGCTGCCAAAGATG- 3'	5'- cacaccaatgatggttattGTAGGTACAGCTTCCAGGT- 3'
Esox-Lucius_MD2_hTLR4	Esox- LuciusMD2_pcDNA3 & hMD2_pcDNA3	For Pike Backbone: 5'-tactgaagctCAGCGGACCCTGC-3'  For hMD2 insert: 5'- gatcgccaccATGTTACCATTCTGTTTTTTC CAC-3'	For Pike Backbone: 5'-tggtaacatGGTGGCGGATCCG-3'  For hMD2 insert: 5'- GGGTCCGTGAGCTTCAGTAAATATGGAAGAA AAC-3'
ZfTLR4ba_hTLR4	hTLR4_pcDNA3 & ZfTLR4ba_pcDNA3	5'- ccagactacaagacgatgacgacaagGAGCCATGCAC TCGAATT-3'	5'- cacactgaggaccgacacaccaatgatggtTTTTTCTGTAGA CACAGTGGTC-3'
ZfTLR4bb_hTLR4	hTLR4_pcDNA3 & ZfTLR4bb_pcDNA3	5'-gactactgtgtgcaagaagaaagaccatcattggtg-3'	5'-caccaatgatggtctttttcttctgtcacacagtagtc-3'
ZfTLR4al_hTLR4	hTLR4_pcDNA3 & ZfTLR4al_pcDNA3	5'-ttgaccactgtgtctacaagaaaaaaagaccatcattggtg-3'	5'-cacaccaatgatggtctttttcttctgtacacagtggtcaa-3'

Table of Primers used for zfTLR4 & MD2 point mutants for testing effects on activity

Resulting Plasmid	Starting Plasmid	Forward Primer	Reverse Primer
ZfTLR4ba_hTLR4-L441F	ZfTLR4ba_hTLR4	5'-agcGTGGGAGGATTTGAAG-3'	5'-AACAGAAATTTCTGAGTTTAGG-3'
ZfTLR4ba_hTLR4-G415N	ZfTLR4ba_hTLR4	5'-aacTTTGAAGGACTTGATTCG-3'	5'-TCCCACATCAACAGAAATTC-3'
ZfTLR4ba_hTLR4-D412S	ZfTLR4ba_hTLR4	5'-tttTCCAACCTAAAGAATCTGAG-3'	5'-AACAGAAAGGTATCCAATG-3'
ZfTLR4ba_hTLR4-G39E	ZfTLR4ba_hTLR4	5'-gaaAGAAACCTCAGCAGC-3'	5'-CATGCATGAGTAATGAAGATTC-3'
ZfMD2-D125K	ZfMD2	5'-aaaGGCCCTGAGGGAGAG-3'	5'- AATCTCCTTATAATGACTGGGAAAG- 3'
ZfMD2-M70R	ZfMD2	5'-tatTACCTGTCAATGACTGTTTC-3'	5'-CAATGGTCCTCAAAGG-3'
ZfMD2-H45Y	ZfMD2	5'-cgcTTCACAATCACTCGTTTTG-3'	5'- TGGGATTAGCGTCAAATTTATAAG- 3'

## Raw Data

Raw data and figures collected can be requested from the Harms lab.

## Table of Resources used in Research

Purpose	Website URL
To help design KLD primers	<a href="https://nebasechanger.neb.com/">https://nebasechanger.neb.com/</a>
To check annealing temperatures and GC%	<a href="https://tmcalculator.neb.com/#!/main">https://tmcalculator.neb.com/#!/main</a>
To help design Quik-Change Lightning Primers	<a href="https://www.agilent.com/store/primerDesignProgram.jsp">https://www.agilent.com/store/primerDesignProgram.jsp</a>

To create the 3D models of the TLR4 and MD2 proteins	<a href="https://colab.research.google.com/github/sokrypton/ColabFold/blob/main/AlphaFold2.ipynb#scrollTo=kOblAo-xetgx">https://colab.research.google.com/github/sokrypton/ColabFold/blob/main/AlphaFold2.ipynb#scrollTo=kOblAo-xetgx</a>
--	---

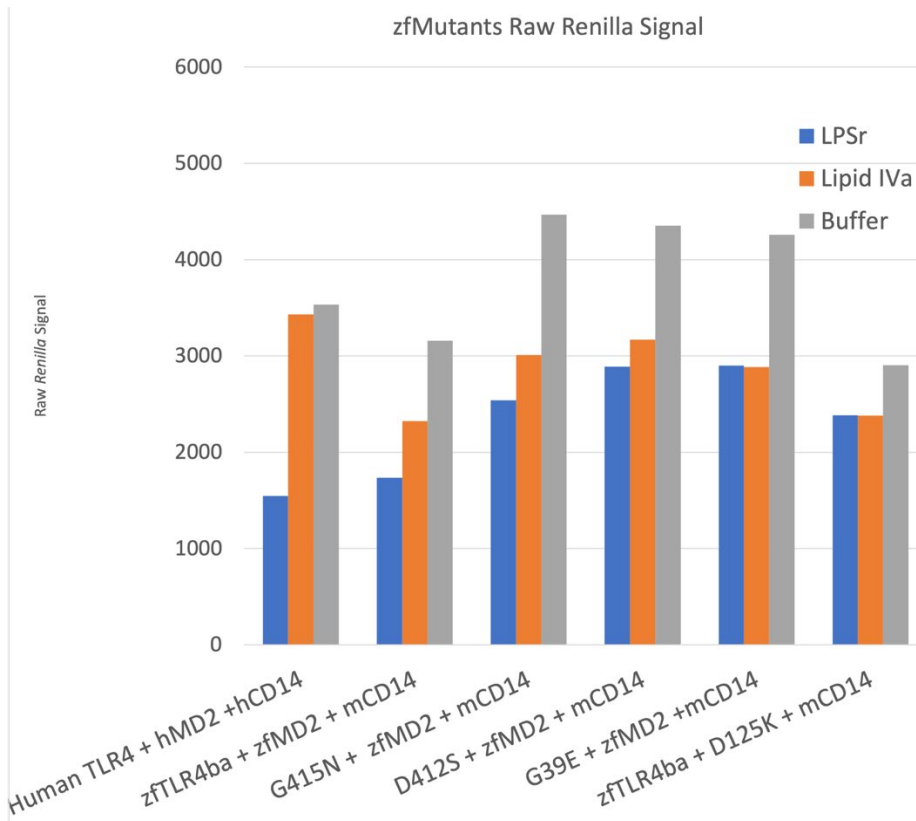
**Further Data**



**Figure 20:** CD180 & TLR4 Variant Raw Renilla Signal



**Figure 21:** Human TLR4 Shifted Raw Renilla Signal



**Figure 22:** zfMutant's Raw Renilla Signal

## Bibliography

1. Brocchieri, L. & Karlin, S. Protein length in eukaryotic and prokaryotic proteomes. *Nucleic Acids Res* **33**, (2005).
2. Sela, B. A. Titin: some aspects of the largest protein in the body. *Harefuah* vol. 141 Preprint at (2002).
3. Gay, N. J. & Keith, F. J. Drosophila Toll and IL-1 receptor [5]. *Nature* vol. 351 Preprint at <https://doi.org/10.1038/351355b0> (1991).
4. What is an Inflammation? *Institute for Quality and Efficiency in Health Care (IQWiG)* (2018).
5. Bihl, F. *et al.* Overexpression of Toll-Like Receptor 4 Amplifies the Host Response to Lipopolysaccharide and Provides a Survival Advantage in Transgenic Mice. *The Journal of Immunology* **170**, (2003).
6. Johnson, G. B., Brunn, G. J. & Platt, J. L. Cutting Edge: An Endogenous Pathway to Systemic Inflammatory Response Syndrome (SIRS)-Like Reactions through Toll-Like Receptor 4. *The Journal of Immunology* **172**, (2004).
7. Tsujimoto, H. *et al.* Role of toll-like receptors in the development of sepsis. *Shock* vol. 29 Preprint at <https://doi.org/10.1097/SHK.0b013e318157ee55> (2008).
8. Ogawa, T., Asai, Y., Makimura, Y. & Tamai, R. Chemical structure and immunobiological activity of Porphyromonas gingivalis lipid A. *Frontiers in Bioscience* **12**, (2007).
9. Whitfield, C., Williams, D. M. & Kelly, S. D. Lipopolysaccharide O-antigens-bacterial glycans made to measure. *Journal of Biological Chemistry* vol. 295 Preprint at <https://doi.org/10.1074/jbc.REV120.009402> (2020).
10. Steimle, A., Autenrieth, I. B. & Frick, J. S. Structure and function: Lipid A modifications in commensals and pathogens. *International Journal of Medical Microbiology* vol. 306 Preprint at <https://doi.org/10.1016/j.ijmm.2016.03.001> (2016).
11. Loes, A. N. *et al.* Identification and Characterization of Zebrafish Tlr4 Coreceptor Md-2. *The Journal of Immunology* **206**, (2021).
12. Kim, H. M. *et al.* Crystal Structure of the TLR4-MD-2 Complex with Bound Endotoxin Antagonist Eritoran. *Cell* **130**, (2007).
13. Kagan, J. C., Magupalli, V. G. & Wu, H. SMOCs: Supramolecular organizing centres that control innate immunity. *Nature Reviews Immunology* vol. 14 Preprint at <https://doi.org/10.1038/nri3757> (2014).
14. Anderson, J. A., Loes, A. N., Waddell, G. L. & Harms, M. J. Tracing the evolution of novel features of human Toll-like receptor 4. *Protein Science* **28**, (2019).
15. Park, B. S. & Lee, J. O. Recognition of lipopolysaccharide pattern by TLR4 complexes. *Experimental and Molecular Medicine* vol. 45 Preprint at <https://doi.org/10.1038/emm.2013.97> (2013).
16. Popot, J. L., Engelman, D. M., Zaccai, G. & de Vitry, C. The ‘microassembly’ of integral membrane proteins: applications & implications. *Progress in clinical and biological research* vol. 343 Preprint at (1990).
17. Park, B. S. *et al.* The structural basis of lipopolysaccharide recognition by the TLR4-MD-2 complex. *Nature* **458**, (2009).

18. Meng, J., Drolet, J. R., Monks, B. G. & Golenbock, D. T. MD-2 residues tyrosine 42, arginine 69, aspartic acid 122, and leucine 125 provide species specificity for lipid IVA. *Journal of Biological Chemistry* **285**, (2010).
19. Gearing, M. Plasmids 101: Sequence and Ligation Independent Cloning (SLIC). *Addgene* (2015).
20. KLD Enzyme Mix. *New England Biolabs*.
21. Nguyen, K. T., Le Clair, S. V., Ye, S. & Chen, Z. Orientation determination of protein helical secondary structures using linear and nonlinear vibrational spectroscopy. *Journal of Physical Chemistry B* **113**, (2009).
22. Edwards, K. *et al.* The role of CD180 in hematological malignancies and inflammatory disorders. *Molecular Medicine* vol. 29 Preprint at <https://doi.org/10.1186/s10020-023-00682-x> (2023).
23. Orlandi, K. N., Phillips, S. R., Sailer, Z. R., Harman, J. L. & Harms, M. J. Topiary: Pruning the manual labor from ancestral sequence reconstruction. *Protein Science* **32**, (2023).