

**UCLA**

**UCLA Electronic Theses and Dissertations**

**Title**

Modeling NFkappaB signaling to capture its dynamical features

**Permalink**

<https://escholarship.org/uc/item/0p60k6h2>

**Author**

Lin, Xiaofei

**Publication Date**

2021

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA  
Los Angeles

Modeling NFkappaB signaling to capture its dynamical features

A dissertation submitted in partial satisfaction of the  
requirements for the degree Doctor of Philosophy  
in Biochemistry, Molecular, and Structural Biology

by

Xiaofei Lin

2021

© Copyright by

Xiaofei Lin

2021

## ABSTRACT OF THE DISSERTATION

Modeling NF $\kappa$ B signaling to capture its dynamical features

by

Xiaofei Lin

Doctor of Philosophy in Biochemistry, Molecular, and Structural Biology

University of California, Los Angeles, 2021

Professor Alexander Hoffmann, Chair

Macrophages are immune sentinel cells that are distributed in every organ. Their physiological function is to detect pathogens, tissue damage, and immune cytokines to initiate and coordinate a multi-phased immune response that is appropriate for the immune threat. How macrophages specify the appropriate response remains unknown. However, recent experimental studies indicate that the dynamics of the signal-responsive transcription factor, nuclear factor kappa B (NF $\kappa$ B), constitute a temporal code that conveys to the nucleus information about the presence and type of immune threat in the extra-cellular environment. Here, I constructed a pipeline to fit a high resolution mathematical model of the NF $\kappa$ B signaling pathway to single cell experimental data. To address model fitting challenges due to high cell-to-cell variability, I developed a novel feature based objective function based on six so-called ‘signaling codons’ (i.e. duration, peak, total activity, oscillation content, etc.) identified as crucial for NF $\kappa$ B stimulus specificity using mutual information and classification analysis. In addition, I documented the performance of

varying optimization algorithms on our large parameter space of 95 biochemical reactions and identified sensitive parameters that specifically tune informative signaling codons. Applications of this high-resolution model include identifying key circuit design principles that encode the observed stimulus-specific use of signaling codons and pinpointing crucial sources of molecular noise that diminish NF $\kappa$ B information encoding.

The dissertation of Xiaofei Lin is approved.

Margot Quinlan

Jorge Torres

Roy Wollman

Alexander Hoffmann, Committee Chair

University of California, Los Angeles

2021

## DEDICATION

This dissertation is dedicated to anyone who was ever told 'you don't belong in science.'

Science is for everyone. You belong here.

## TABLE OF CONTENTS

<b>ABSTRACT OF THE DISSERTATION</b> .....	<b>ii</b>
<b>DEDICATION</b> .....	<b>v</b>
<b>LIST OF FIGURES</b> .....	<b>viii</b>
<b>LIST OF TABLES</b> .....	<b>x</b>
<b>ACKNOWLEDGEMENTS</b> .....	<b>xi</b>
<b>VITA</b> .....	<b>xiii</b>
<b>Chapter 1 : Introduction</b> .....	<b>1</b>
1.1 NF $\kappa$ B recognizes diverse stimuli to initiate appropriate physiological responses .....	2
1.2. Dynamic features of the NF $\kappa$ B temporal profile are suggested to carry information about a given stimulus and dose .....	2
1.3. Cell-to-cell variability in NF $\kappa$ B signaling may limit information encoding.....	3
1.4. Mathematical modeling of signaling systems provides insights to guide wet lab experimentation.....	4
1.5. Current NF $\kappa$ B model qualitatively fits experimental single cell data .....	4
1.6. Figures .....	6
<b>Chapter 2 : Quantitatively evaluating mathematical model fit to NF<math>\kappa</math>B dynamics</b> .....	<b>10</b>
2.1. Six ‘signaling codons’ of NF $\kappa$ B response possess information about the identity of the stimulus .....	11
2.2. Developing a feature based objective function for model fitting to single cell NF $\kappa$ B trajectories.....	11
2.3. Signaling codon discrepancies between initial model and ‘representative’ cells of heterogeneous NF $\kappa$ B response .....	13
2.4. Figures .....	15
<b>Chapter 3 : Optimization pipeline for fitting the NF<math>\kappa</math>B model to dynamic features</b> .....	<b>22</b>
3.1. Performance of varying optimization methods on our NF $\kappa$ B model.....	23
3.2. Quantifying NF $\kappa$ B model fit to multiple cells and stimulus conditions .....	24
3.3. Parameter sensitivity scan to constrain the NF $\kappa$ B model search space .....	25
3.4. Iterative approach to fit each module given shared pathway parameters .....	26
3.5. Optimized NF $\kappa$ B model to signaling codons .....	26
3.6. Figures .....	27
<b>Chapter 4 : Methods</b> .....	<b>35</b>
4.1. Experimental data generation .....	36
4.2. Baseline normalization for NF $\kappa$ B signaling trajectories .....	36
4.3. Scaling NF $\kappa$ B model trajectories from SI units to experimental AU .....	36
4.4. Signaling codon definitions .....	36

4.5. Defining non-responding cells.....	37
4.6. Selecting representative cells .....	37
4.7 Identifying sensitive parameters .....	38
4.8. Particle Swarm Optimization (PSO) to a single cell.....	38
4.9. Gradient Descent Optimization for the full NFκB model .....	38
<b>Chapter 5 : Conclusion.....</b>	<b>39</b>
5.1. Discussion .....	40
5.2. Figures .....	43
<b>References .....</b>	<b>44</b>

## LIST OF FIGURES

<b>Figure 1.1.</b> Functions of tissue-resident immune sentinel cells.....	6
<b>Figure 1.2.</b> Temporal coding to produce stimulus-specific gene expression.....	7
<b>Figure 1.3.</b> Reaction schema for the multi-stimulus model of NF $\kappa$ B activation.....	8
<b>Figure 1.4.</b> NF $\kappa$ B model simulations qualitatively fit experimental data at different stimuli and doses.....	9
<b>Figure 2.1.</b> Dynamic features of NF $\kappa$ B response contain information about the identity of the ligand.....	15
<b>Figure 2.2.</b> Model fit using time series mean squared error (MSE) is not guaranteed to preserve dynamic features.....	16
<b>Figure 2.3.</b> NF $\kappa$ B response shows single cell variability in the timing of oscillatory behavior....	18
<b>Figure 2.4.</b> Representative cells of NF $\kappa$ B response were identified based on signaling codon mode. ....	20
<b>Figure 2.5.</b> Model simulations of NF $\kappa$ B response possess signaling codons that deviate from representative cells.....	21
<b>Figure 3.1</b> Performance of varying optimization algorithms when fitting our NF $\kappa$ B model to a single cell.....	27
<b>Figure 3.2.</b> Parameter distributions for the lowest $\text{RMSD}_{\text{stimulus}}$ parameter sets for the five receptor proximal modules.....	29
<b>Figure 3.3.</b> Optimization pipeline iterates between core module and receptor-proximal modules using stimulus specific datasets.....	30
<b>Figure 3.4.</b> Gradient descent optimization finds the optimal parameter set to fit NF $\kappa$ B signaling codons. ....	31
<b>Figure 3.5.</b> Signaling codons of all NF $\kappa$ B datasets show improved fit to representative cells after feature based optimization.....	33

<b>Figure 3.6.</b> Optimized NF $\kappa$ B trajectories to signaling codons.....	34
<b>Figure 5.1.</b> Developing a mathematical model of NF $\kappa$ B response to simulate single cell heterogeneity.....	43

## LIST OF TABLES

<b>Table 2.1</b> Signaling codon weights ( $f_{nweight}$ ) for the feature based objective function.....	17
<b>Table 2.2.</b> Signaling codon metrics for the feature based objective function.....	19
<b>Table 3.1.</b> Stimulus weights for calculating $RMSD_{core}$ .....	28
<b>Table 3.2.</b> Sensitive parameters of the $NF\kappa B$ model.....	32

## ACKNOWLEDGEMENTS

Completing a PhD was an achievement both academically and in self-growth. I learned how to own my expertise, communicate effectively, advocate for myself, and champion myself when others would not. An incredible support network built in grad school made this journey possible.

First, I would like to thank my committee members, Dr. Alexander Hoffmann (advisor), Dr. Margot Quinlan, Dr. Jorge Torres, and Dr. Roy Wollman for their continued guidance and support in research and professional development throughout this dissertation work. Most notably, I would like to thank Alex for his mentorship, leadership, and enthusiasm for science. I first reached out to Alex during my second year of grad school, when many factors caused me to consider leaving my PhD. At that time, I told myself I would stay if I found a fantastic environment. During my first week as a rotation student who possessed a very different background in structural biology, I immediately felt the infectious enthusiasm from the lab that hooked me into immunology and math modeling. Over the years, Alex has continually proven to be a phenomenal mentor who supports both scientific training and the professional goals of all lab members.

Second, I would like to thank my lab mates for their valuable research insights and friendship. This dissertation would not have been possible without Dr. Ade Adelaja, who significantly contributed to the fluorescent microscopy workflow and data collection, and identification of informative dynamic features of NF $\kappa$ B signaling described in this thesis. Ade was my lab desk neighbor throughout my time in the Hoffmann lab. In addition to providing valuable research feedback as a colleague on a day-to-day basis, Ade has been a valued friend throughout my PhD years. I would also like to thank Dr. Simon Mitchell for being my first mentor in math modeling. Simon's approachability and willingness to mentor was foundational to my success as a growing computational biologist. Additional thanks to the remaining Hoffmann lab members who I had the pleasure of working with during my time at UCLA: Katherine Sheu, Ning Wang, Apeksha Singh, Diane Lefaudeux, Ying Tang, Stefanie Luecke, Catera Wilder, Koushik Roy, Marie Oliver

Metzig, Hari Vaidehi Narayanan, Yi Liu, Eason Lin, Kim Ngo, Quen Cheng, Brian Kim, Emilio Uriarte, Richard Ahn, Anup Mazumder, Carolina Chavez, Héctor Navarro, Kylie Farrell, Jennifer Chia, and Xiaolu Guo.

Third, my family and countless friends who I met during my time at UCLA have greatly contributed to my PhD journey. I would like to my family for providing me with the foundation that has made it possible for me to pursue graduate school. Next, I would like to thank my roommates, Christina Burghard and Neal Brodnik for their incredible support throughout the COVID-19 pandemic. A PhD is difficult its own, and even more challenging during a global pandemic. Christina and Neal kept me consistently entertained, fed, and supported during an isolating time of social distancing. I would also like to thank my fellow BMSB cohort members, Callie Glynn and Jennifer Ngo. I met Callie and Jen on Day 1 of grad school. Since that time, we've had countless adventures in Westwood while navigating our PhD milestones. Lastly, I would like to thank the PhD Balance team, which has shown me an incredible international community championing graduate student wellness. Most notably, I would like to thank Dr. Susanna Harris, founder of PhD Balance, for continually inspiring me to use social media to drive systemic change.

One of the key lessons I learned throughout this PhD is that while journeys require self-motivation, perseverance, and willingness to learn, surrounding yourself with a healthy environment is equally important because 1. healthy environments foster fantastic growth and 2. you deserve fantastic people and nothing less.

## VITA

### EDUCATION

- 2021                                      PhD Candidate, Biochemistry, Molecular, and Structural Biology  
University of California, Los Angeles | Los Angeles, CA
- 2016                                      Bachelor of Science, Biology  
New York University | New York, NY

### RESEARCH POSTER PRESENTATIONS

**Lin X**, Adelaja A, Taylor B, Hoffmann A. Developing a mathematical model of NF $\kappa$ B activity in single macrophages in response to pathogens and inflammatory cytokines. **UCLA QCBio Retreat, September 2019**. Los Angeles, CA

**Lin X**, Mitchell S, Tang Y, Adelaja A, Taylor B, Hoffmann A. Exploring the stimulus specificity of the dynamic NF $\kappa$ B signaling code. **International Conference on Systems Biology for Human Diseases, June 2018**. Los Angeles, CA.

**Lin X**, Leffler AE, Bonneau R. Analyzing the Molecular Interactions of the  $\alpha$ -GID/  $\alpha$ 4 $\beta$ 2 Receptor Complex: An Evaluation for Drug Design. **Rosetta Conference, July 2015**. Leavenworth, WA.

### FELLOWSHIPS

2018-2020                                      UCLA Cellular and Molecular Biology Training Grant

### AWARDS

March 2019                                      UCLA Grad Slam Audience Choice Award  
March 2018                                      NSF Graduate Research Fellowship Program (GRFP) Honorable Mention

# **Chapter 1: Introduction**

### *1.1 NFκB recognizes diverse stimuli to initiate appropriate physiological responses*

The transcription factor, nuclear factor kappa B (NFκB) is responsible for responding to diverse external stimuli and initiating the appropriate gene expression response for physiological functions such as lymphoid tissue development, immune, inflammatory, and environmental stress responses, and neuronal signaling<sup>1-3</sup> (Figure 1.1). Impairment of NFκB regulation causes human pathologies such as chronic inflammatory diseases and cancers<sup>4,5</sup>. How NFκB distinguishes each pathologic threat, injury, or inflammatory signal to produce the appropriate combination of gene expression is unknown. As a result, the NFκB regulatory network provides a wealth of valuable therapeutic targets that remain to be fully explored for clinical use.

Molecular characterization of the NFκB network indicates that NFκB is held in an inactive state by association with one of three IκB inhibitor proteins (IκBα, IκBε or IκBδ)<sup>6,7</sup>. In response to stimulation, the IκB kinase (IKK) phosphorylates NFκB-bound IκB proteins, targeting them for proteolysis through the ubiquitin-proteasome pathway, and allowing NFκB to translocate to the nucleus for transcription<sup>8</sup>. All external stimuli transduced by plasma membrane-bound receptors or subcellular organelles converge on IKK before promoting gene expression by NFκB for physiological response<sup>9</sup>. As a result, the functional pleiotropism of NFκB is suggested to rely on the ability of IKK to output a diverse set of NFκB temporal profiles in response to various upstream signals<sup>10</sup>. Understanding the molecular mechanisms of information encoding as signals are received at upstream receptor modules and converge on IKK to produce an NFκB response is valuable for elucidating the functional diversity of the NFκB system.

### *1.2. Dynamic features of the NFκB temporal profile are suggested to carry information about a given stimulus and dose*

The Dynamical Code hypothesis suggests that information about an external stimulus is encoded in the temporal profile or dynamics of a signaling activity<sup>11-14</sup> (Figure 1.2a). Features of

network dynamics include exhibiting oscillatory, persistent, or transient activation, or quantitatively varying in amplitude, frequency, or duration of response. For the NF $\kappa$ B pathway, NF $\kappa$ B activity is suggested to be oscillatory in response to the proinflammatory cytokine tumor necrosis factor (TNF), but steady in fibroblasts responding to a component of the outer membrane of gram-negative bacteria, lipopolysaccharide (LPS)<sup>10,15,16</sup> (Figure 1.2b). Indeed, stimulus specific NF $\kappa$ B dynamics have been suggested to lead to differential gene expression and epigenomic reprogramming<sup>17,18</sup>. Such mediation of distinct cellular outcomes by disparate network dynamics has also been identified in the p53 DNA response pathway in human cell lines, and the Msn2 stress response in *S.cerevisiae*<sup>11,19</sup>.

### *1.3. Cell-to-cell variability in NF $\kappa$ B signaling may limit information encoding*

Most stimulus-specific time course measurements supporting the use of a dynamic code in the NF $\kappa$ B pathway were taken at the population-level. However, signaling networks need to transmit information from external stimuli in the presence of high cell-to-cell variability<sup>20</sup>. This variability can arise from the following: i) thermodynamic stochasticity in the reactions that directly control signaling ii) cell-to-cell differences in the abundance of network components or enzyme-catalyzed reactions between genetically identical cells<sup>21–23</sup>. Such variability has been shown to either limit or enhance the capacity for reliable biochemical information transduction<sup>24–27</sup>. More investigation is needed to determine how cell-to-cell variability impacts stimulus specificity and information encoding.

Thus far, single-cell studies of NF $\kappa$ B signaling have relied upon an exogenously-introduced fluorescent reporter in immortalized cell lines. These reporters have disadvantages such as demonstrable decreases in cellular ability to respond to inflammatory pathogen or cytokine signals, and photo toxicity caused by exposure to lasers or the production of reactive oxygen species by fluorescent protein moieties<sup>13</sup>. These drawbacks motivate the use of primary cells in which a fluorescent reporter is embedded into the endogenous NF $\kappa$ B gene locus. The

Hoffmann lab has recently generated a reporter mouse strain expressing a C-terminal mVenus fusion of the NF $\kappa$ B dimer, RelA/p65, from its endogenous locus to overcome these artifacts. mVenus offers the dual advantage of being less masked by autofluorescence in standard culture media, allowing long-term imaging using conventional epifluorescence microscopy.

#### *1.4. Mathematical modeling of signaling systems provides insights to guide wet lab experimentation*

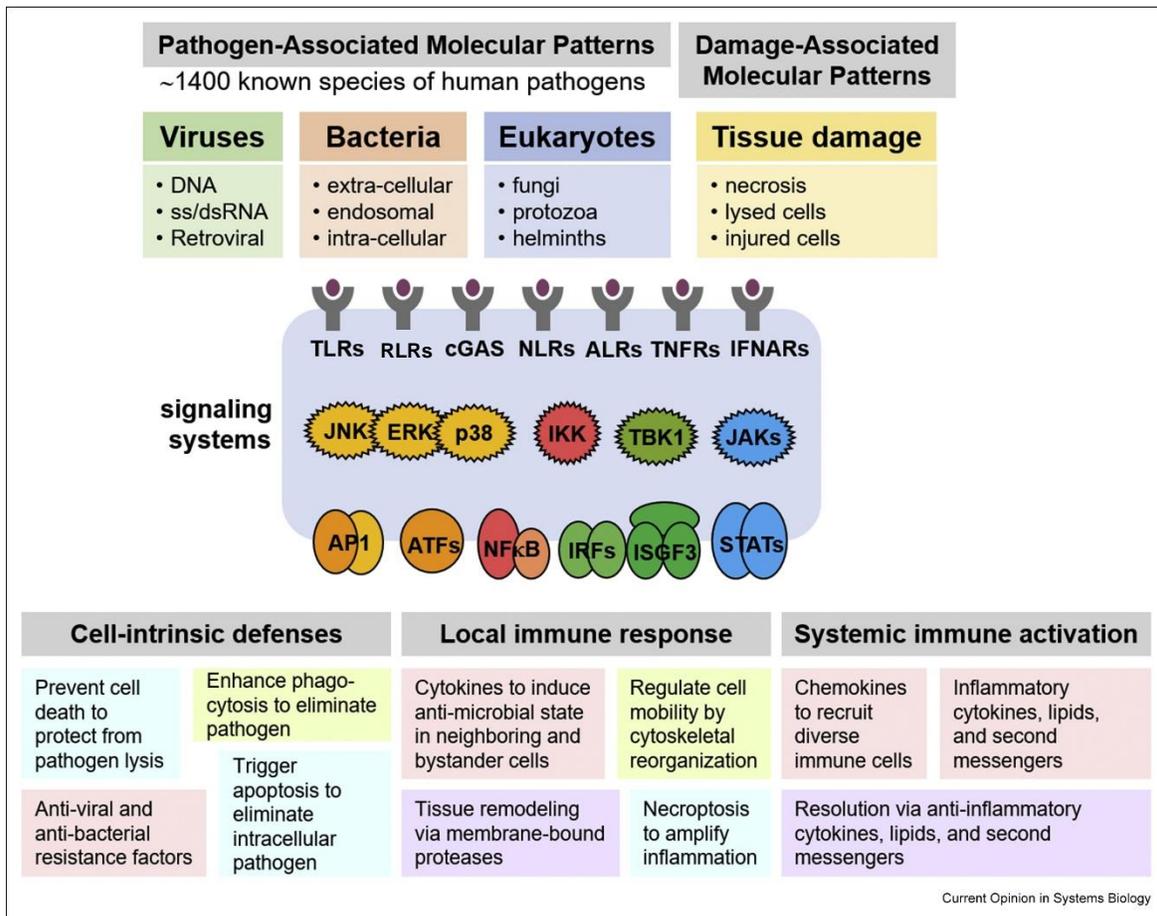
When studying signal transduction networks, such as the NF $\kappa$ B system, the role of individual components within the network is often difficult to discern based on experimental data alone. This idea is evident when a genetic mutation does not produce the predicted phenotype or when pharmacological treatments result in surprising side effects. Computational modeling, using mathematical equations to reconstruct networks, allows functional analysis of network behavior that can provide mechanistic insights unreachable by experimental studies<sup>13</sup>. Such insights include i) revealing network emergent properties (i.e. dynamic negative feedback regulation, redundant or overlapping mechanisms, and cross-talk between stimuli). ii) identifying compensation mechanisms after simulating knockouts, and iii) analyzing the contribution of each reaction to the network behavior using parameter sensitivity analysis<sup>28</sup>. These insights are valuable to the biomedical sciences for identifying which reactions are sensitive to modulation within ranges achievable by experimental tools and understanding how a drug with a known molecular target within a network will affect the network response.

#### *1.5. Current NF $\kappa$ B model qualitatively fits experimental single cell data*

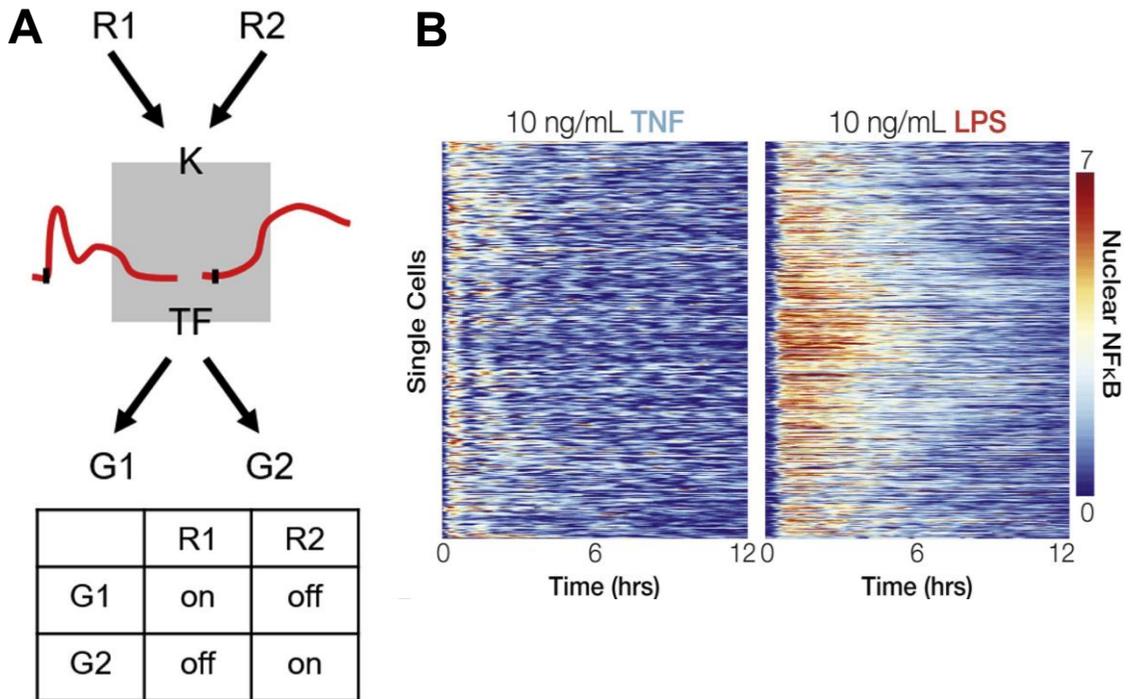
Many PAMPs are recognized by toll-like receptors (TLRs) which elicit NF $\kappa$ B activation to coordinate inflammatory and innate immune responses<sup>29,30</sup>. As a result, the Hoffmann lab has constructed an ordinary differential equation (ODE) based mathematical model of NF $\kappa$ B response using single cell signaling data in macrophages. This model contains a TNF receptor (TNFR) module, a variety of TLR modules, and the IKK/NF $\kappa$ B core module (Figure 1.3). This initial model

simulates deterministic single cell trajectories that qualitatively fit to nuclear NF $\kappa$ B temporal profiles at varying stimuli and doses (Figure 1.4). This qualitative model fitting is useful for predicting NF $\kappa$ B response at a coarse-grained level. However, fine grained model simulations and accurate analysis require thorough quantitative documentation of model fit. This thesis details a pipeline for quantitatively evaluating NF $\kappa$ B model fit to single cell data using informative dynamic features of NF $\kappa$ B response.

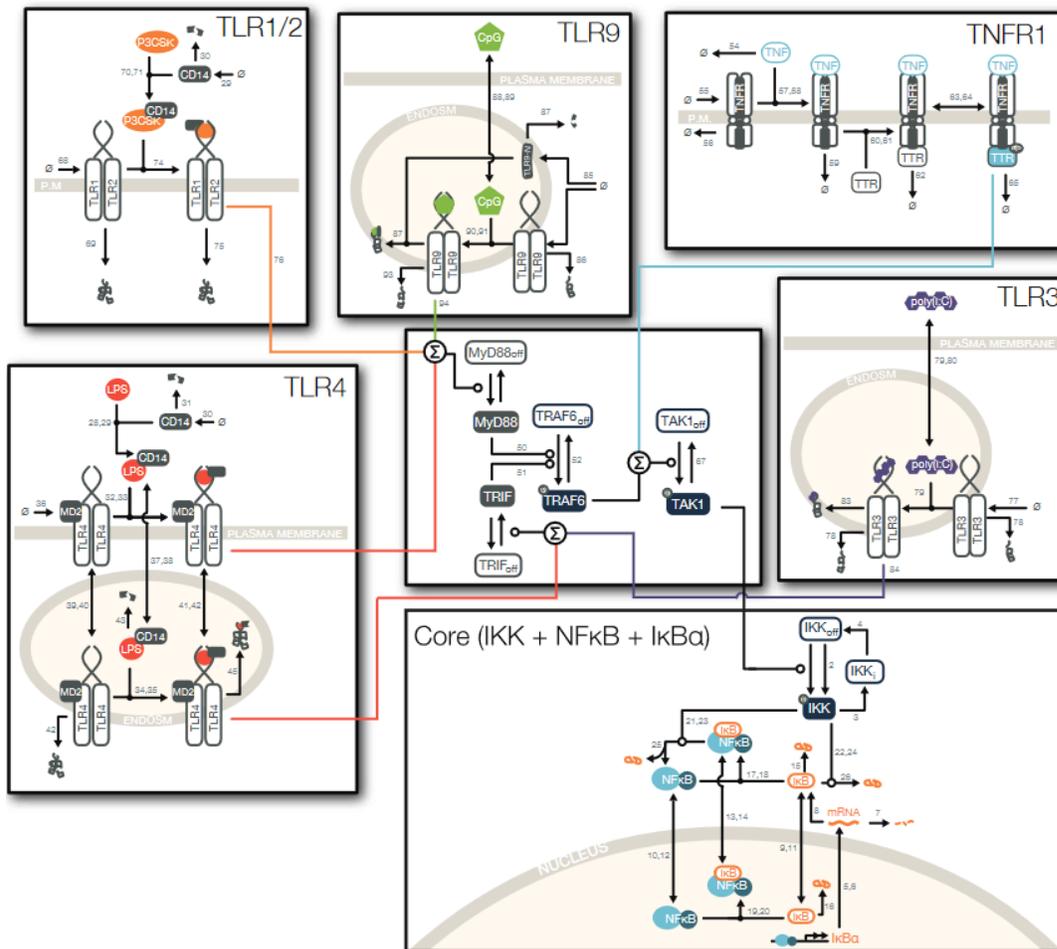
1.6. Figures



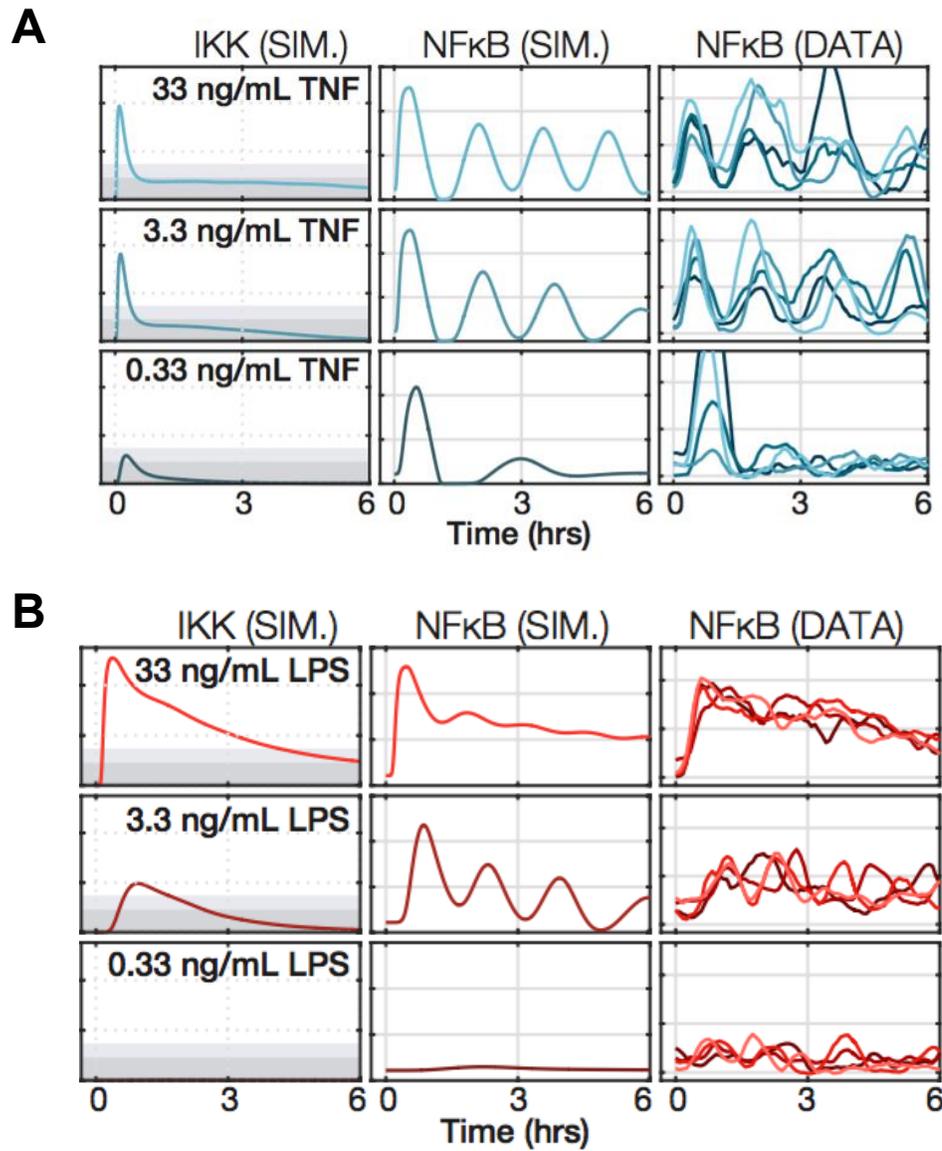
**Figure 1.1. Functions of tissue-resident immune sentinel cells.** Immune sentinel cells as such tissue resident macrophages or dendritic cells are capable of sensing diverse immune threats via dozens of different sensors and respond in numerous ways to regulate cell-intrinsic defenses, local immune responses, or systemic immune activation. Colors illustrate the diversity of functions but do not represent a color code. (Sheu, Luecke, Hoffmann, Current Opinions in Systems Biology 2019)



**Figure 1.2. Temporal coding to produce stimulus-specific gene expression.** A. Two receptors elicit distinct gene expression programs via a single shared pathway; this is achieved via stimulus-specific temporal patterns of signaling activity. (Sheu, Luecke, Hoffmann, *Current Opinions in Systems Biology* 2019). B. NF $\kappa$ B signaling dynamics vary between TNF and LPS. Each row is a single cell. (Adelaja, Taylor et al. *Immunity* 2021).



**Figure 1.3. Reaction schema for the multi-stimulus model of NFκB activation.** The five upstream receptor modules include TNFR, TLR4, TLR3, TLR1/2, and TLR9. These receptors recognize the following inflammatory cytokines and pathogen-associated molecular patterns (PAMPs): tumor necrosis factor (TNF), lipopolysaccharide (LPS), Polyinosinic:polycytidylic acid (poly I:C), synthetic triacylated lipoprotein (Pam3CSK), and CpG oligodeoxynucleotides (CpG) respectively. The output of each receptor-associated module is the IKK activity, which then serves as the sole input of the common IKK/NFκB signaling module. (Adelaja, Taylor et al. Immunity 2021).



**1.4. NFκB model simulations qualitatively fit experimental data at different stimuli and doses.** A. TNF model simulations show oscillatory character at varying doses to match the experimental data. B. LPS model simulations show sustained behavior at highest dose, oscillatory behavior at intermediate dose, and low response at lowest dose, consistent with the data (Adelaja, Taylor et al. Immunity 2021).

# **Chapter 2: Quantitatively evaluating mathematical model fit to NF $\kappa$ B dynamics**

## *2.1. Six 'signaling codons' of NF $\kappa$ B response possess information about the identity of the stimulus*

Information-theoretic algorithms based on estimates of channel capacity, a measure of the maximum distinction possible across a set of measured input conditions, can approximate the amount of information carried by dynamical encoding<sup>25</sup>. The Hoffmann lab has developed an algorithm to iteratively build combinations of dynamic features of NF $\kappa$ B response that optimize channel capacity<sup>31</sup>. Channel capacities were markedly enhanced when considering multiple dose-ligand conditions (Figure 2.1a), and six dynamical features (termed 'signaling codons') that optimized channel capacity were identified. These six signaling codons include duration, peak amplitude, speed, total activity, oscillatory content, and early vs late activity. In addition, eliminating a signaling codon led to loss of the certainty of  $_{\text{ligand}}$  prediction (Figure 2.1b). These results suggest that signaling codons of NF $\kappa$ B response encode information about a ligand identity and dose. A mathematical model that accurately recapitulates these six signaling codons is a powerful tool for investigating the molecular mechanisms that generate stimulus specific NF $\kappa$ B response.

## *2.2. Developing a feature based objective function for model fitting to single cell NF $\kappa$ B trajectories*

Quantitative model fitting to dynamic time series data has been historically difficult as standard distance metrics, such as mean squared error (MSE), highly penalize slight shifts in phase, amplitude, and frequency. As a result, these approaches are limited in their ability to preserve informative NF $\kappa$ B signaling codons. For example, NF $\kappa$ B response to inflammatory cytokine tumor necrosis factor (TNF) is characterized by high oscillatory content<sup>32</sup>. However, time series MSE optimization of NF $\kappa$ B response to TNF leads to loss of documented biological oscillations (Figure 2.2). In addition, time series data at single cell resolution is prone to technical noise that challenges model fitting to biological behavior. To address these limitations, this

chapter developed a feature based objective function (Equation 2.1) for model fitting to 1. preserve informative NFκB signaling codons and 2. reduce the effects of technical noise during model fitting by measuring the distance between biologically relevant features. RMSD refers to the distance between the model simulation and a single cell NFκB temporal profile.  $n$  represents the number of signaling codons,  $f_{nweight}$  represents the weighting factor for each signaling codon,  $f_{nexp}$  represents the experimental signaling codon value (z-scored), and  $f_{nmod}$  represents the model signaling codon value:

$$\text{RMSD} = \sqrt{\left(\frac{1}{n}\right) (f_{1weight}(f_{1exp} - f_{1mod})^2 + f_{2weight}(f_{2exp} - f_{2mod})^2 + \dots + f_{nweight}(f_{nexp} - f_{nmod})^2)} \quad (\text{Eq. 2.1})$$

$f_{nweight}$  was determined based on the known influence of the parameters on NFκB dynamic features. For example, the core module is documented to modulate NFκB oscillatory character<sup>33</sup>. As a result, the oscillatory codon was highly weighted when fitting the parameters of the core module. Alternatively, pathway-specific mechanisms upstream of IKK activity are documented to produce stimulus specific dynamic features that include peak amplitude, duration, early vs late activity, total activity, and speed<sup>16</sup>. When fitting the receptor proximal modules, the weighting of each signaling codon was stimulus specific and informed by classification analysis performed by Adelaja, Taylor et al. 2021 (Figure 2.1b). The  $f_{nweight}$  values for each module of our NFκB model is documented in Table 2.1.

Signaling codons were defined based the metrics identified by Adelaja, Taylor et al. 2021. As these metrics were originally applied to experimental trajectories only, metrics were adapted to be appropriate for our NFκB model simulations. Most notably, the oscillatory codon was originally defined based on a Fourier transform approach to handle experimental trajectories containing multiple frequencies due to technical noise. However, deterministic ODE models of biological networks lack technical noise, and Fourier transforms of simulations are not comparable with experimental trajectories. To address this issue, the Fourier transform approach was

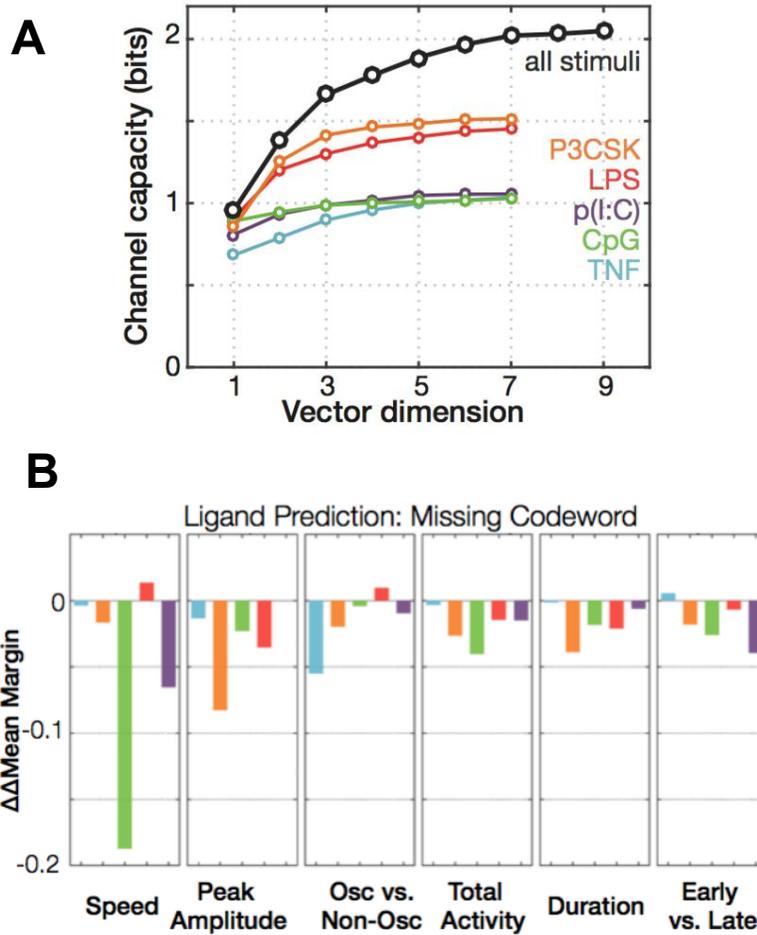
replaced with metrics that describe “sum to peak to trough ratio” and “number of oscillatory peaks” to calculate the oscillatory codon of our feature based objective function. In addition, the timing of oscillations during NF $\kappa$ B response is heterogeneous (Figure 2.3a) and using six signaling codons with oscillatory content over the full trajectory did not preserve oscillation timing (Figure 2.3b). As a result, the oscillatory codon is split into two separate signaling codons (‘early oscillatory’ and ‘late oscillatory’) to preserve information about the timing of oscillations during optimization. Ultimately, the feature based objective function is composed of seven signaling codons: duration, early vs late, early oscillatory, late oscillatory, peak amplitude, speed, total activity (Table 2.2). For each signaling codon, metrics are z-scored and averaged to determine the signaling codon value.

### *2.3. Signaling codon discrepancies between initial model and ‘representative’ cells of heterogeneous NF $\kappa$ B response*

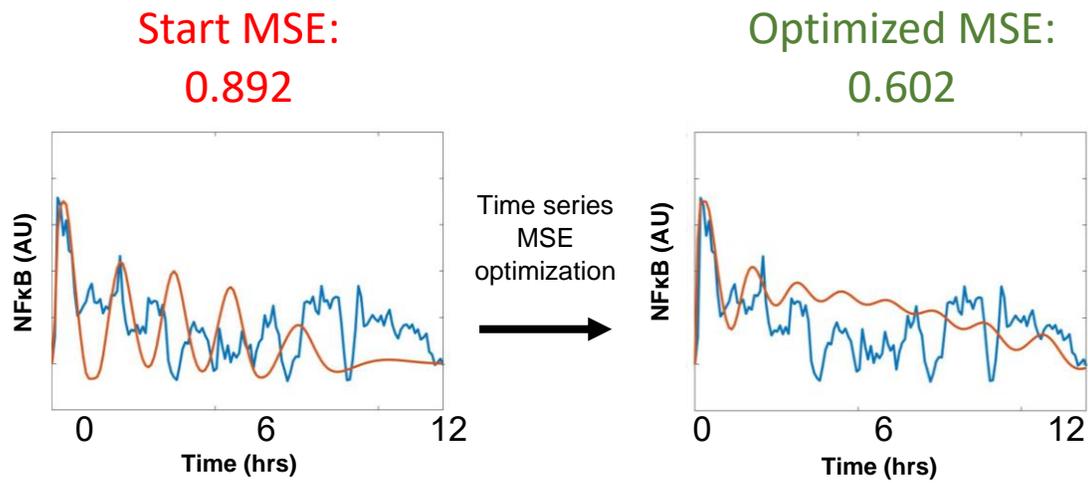
Using the endogenous mVenus reporter, the Hoffmann lab has generated a wealth of experimental data detailing NF $\kappa$ B dynamics in primary macrophages at single cell resolution. These datasets capture NF $\kappa$ B response to stimuli of the five upstream receptor modules of our NF $\kappa$ B model at varying doses. These single cell datasets show high cell-to-cell variability (Figure 1.2b), and averaging techniques to time series data, such as arithmetic mean of time points or dynamic time warping, may distort signaling codons captured by single cell NF $\kappa$ B response. To address this challenge, 10-20 ‘representative’ single cell trajectories of each stimulus and dose were selected for model fitting. Trajectories that possess signaling codon values that surround the experimental distribution peak were defined as ‘representative’ (Figure 2.4). The initial model possesses signaling codon values that deviate from representative cells in multiple stimuli and doses (Figure 2.5). Most notably, model simulations of TNF at 1ng/mL and 10ng/L show lower duration and total activity than the experimental data. In addition, model simulations of LPS at 10ng/mL and 100ng/mL possess higher speed than the experimental data. Chapter 3 describes

the pipeline for optimizing the NF $\kappa$ B model to signaling codons of representative cells of a stimulus and dose.

2.4. Figures



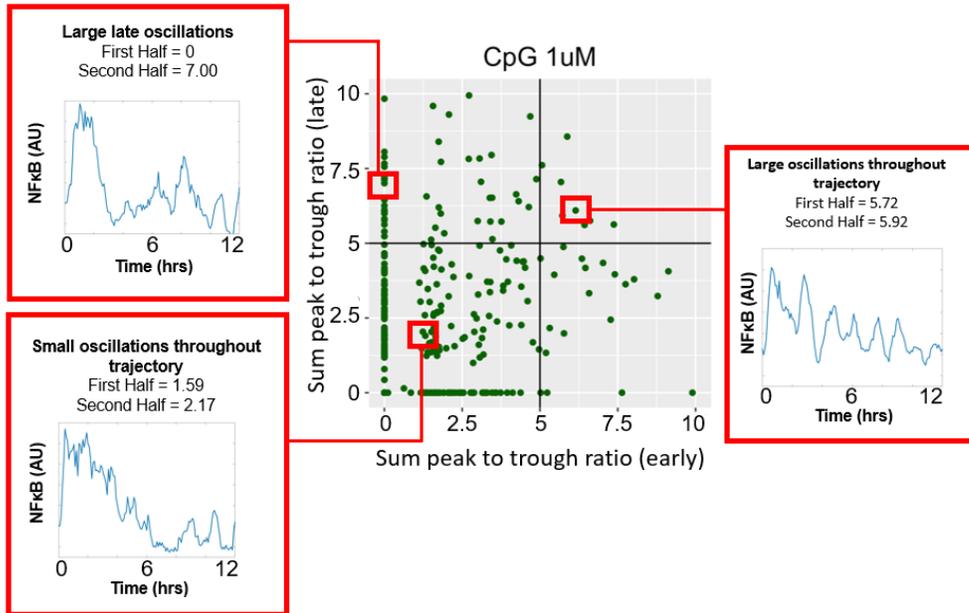
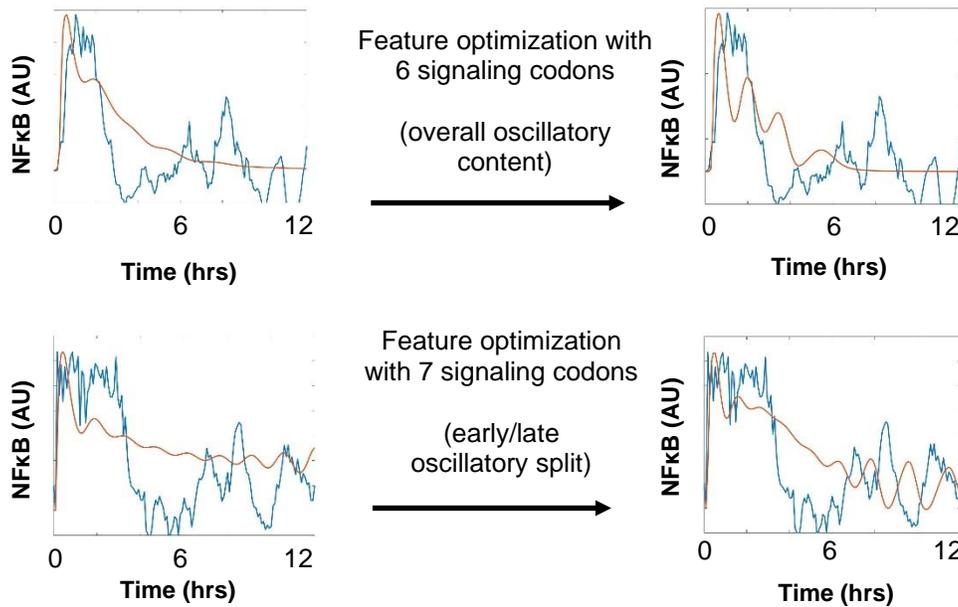
**Figure 2.1. Dynamic features of NFκB response contain information about the identity of the ligand.** A. Channel capacity as a function of the number of most informative metrics identified by Adelaja, Taylor et al., either using the entire dataset of all ligand types and doses (black line) or using the dose response data for each indicated ligand. B. The effect of each signaling codon on the certainty of ligand prediction: The loss in classification confidence when the indicated codon is missing from the set of six (versus all features). (Adelaja, Taylor et al. Immunity 2021)



**Figure 2.2. Model fit using time series mean squared error (MSE) is not guaranteed to preserve dynamic features.** Fitting the model (orange) to an experimental single cell NFκB response to TNF 10ng (blue) using time series MSE leads to dampened oscillations not consistent with the experiment.

	Signaling Codon Weight					
Signaling Codon	Core	LPS	TNF	Pam3CSK	Poly(I:C)	CpG
Duration	0	1	0.6	2	0.5	0.6
Early vs. Late	0	1	0.6	0.5	2	0.6
Early Oscillatory	3	0.5	1.5	0.25	0.25	0.3
Late Oscillatory	3	0.5	1.5	0.25	0.25	0.3
Peak Amplitude	0	1	0.6	2	0.5	0.6
Speed	0	1	0.6	0.5	2	3
Total Activity	0	1	0.6	0.5	0.5	0.6
Total	6	6	6	6	6	6

**Table 2.1 Signaling codon weights ( $f_{nweight}$ ) for the feature based objective function.**

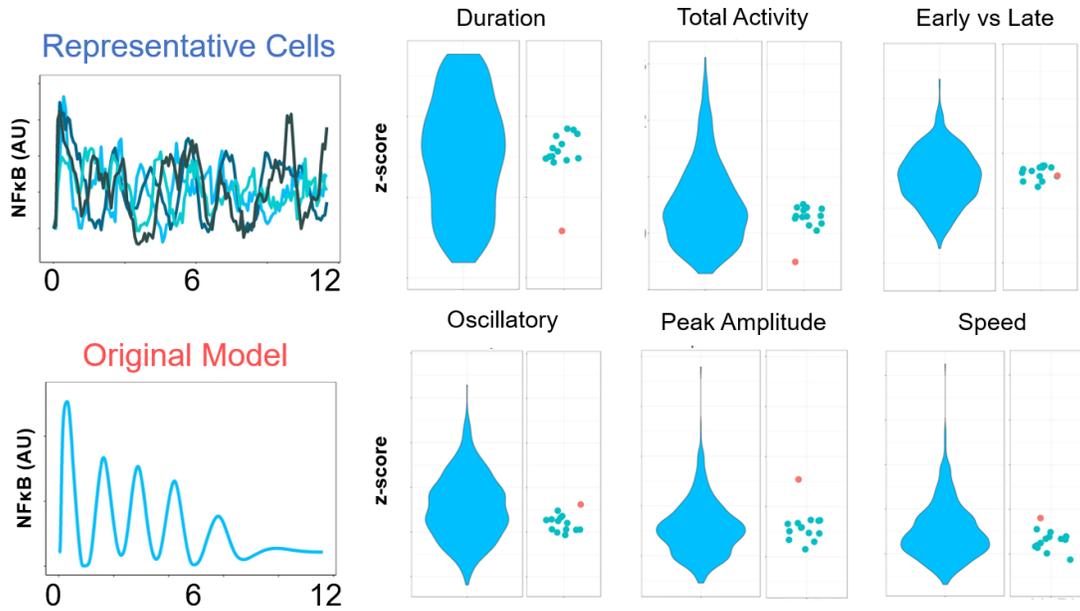
**A****B**

**Figure 2.3. NFKB response shows single cell variability in the timing of oscillatory behavior.** A. Heterogeneous oscillation timing is documented in NFKB response to CpG 1 $\mu$ M B. Feature based optimization using overall oscillatory content does not preserve timing of oscillatory behavior. (Condition - CpG 1 $\mu$ M) (Top) Splitting oscillatory codon into “early oscillatory” and “late oscillatory” preserves timing of oscillatory behavior (Condition - LPS 100ng). (Bottom)

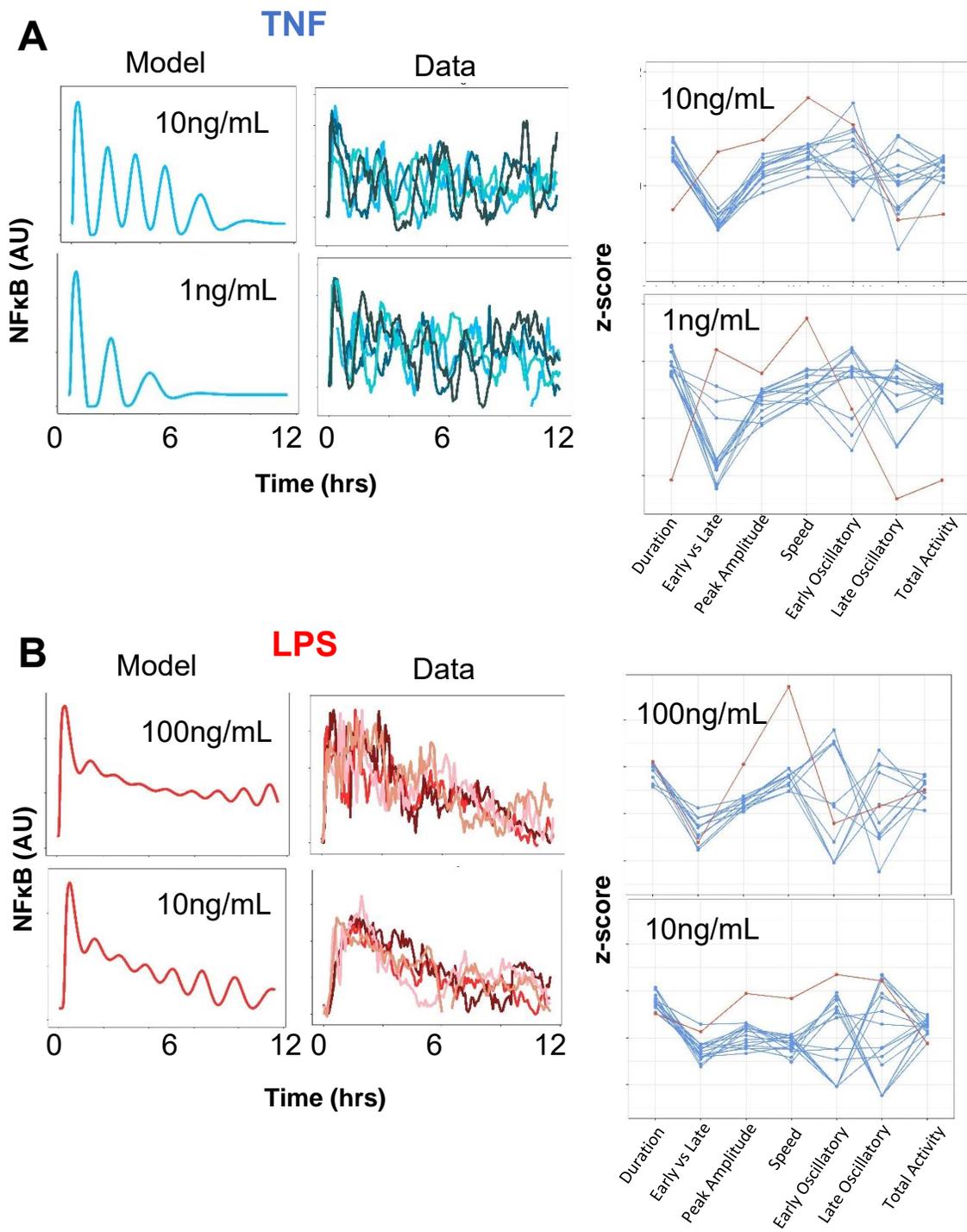
Signaling Codon Number	Metric	Signaling Codon Name	Unit
1	Number of Peaks	Duration	#
1	Total Duration	Duration	Hr.
2	Time to Half Maximal Activity	Early vs. Late	Hr.
3	Sum peak to trough ratios (early)	Early Oscillatory	A.U.
3	Number of oscillatory peaks (early)	Early Oscillatory	#
4	Late sum peak to trough ratios (late)	Late Oscillatory	A.U.
4	Number of oscillatory peaks (late)	Late Oscillatory	#
5	Fold Change	Peak Amplitude	A.U.
5	Maximum to Minimum Amplitude Difference	Peak Amplitude	A.U.
5	1 <sup>st</sup> Peak Amplitude	Peak Amplitude	A.U.
6	Maximum 1 <sup>st</sup> Peak Translocation Speed	Speed	A.U.
6	Time to 1 <sup>st</sup> Peak	Speed	Hr.
6	Derivative Between 5 & 10 minutes	Speed	A.U.
7	Maximum Total Activity	Total Activity	A.U.

**Table 2.2 Signaling codon metrics for the feature based objective function.**

# TNF 10ng



**Figure 2.4. Representative cells of NFκB response were identified based on signaling codon mode.** The duration, total activity, and peak amplitude of the original model (red dot, right panels) deviate from the signaling codons of the representative cells (blue dots, right panels)



**Figure 2.5. Model simulations of NFκB response possess signaling codons that deviate from representative cells.** Signaling codon comparisons between the model (red line) and representative cells (blue lines) for TNF doses (A) and LPS doses (B) identify model signaling codons that do not fit the data.

# **Chapter 3: Optimization pipeline for fitting the $NF_{\kappa}B$ model to dynamic features**

### 3.1. Performance of varying optimization methods on our NFκB model

A common task when constructing mathematical models is parameter estimation or identifying areas of the parameter space that are consistent with experimental data and prior knowledge. For biological models, many parameters cannot be determined directly by available experimental techniques. Optimization methods are needed to search the parameter space for agreement between model and data. In addition, there is a wide range of optimization methods that possess varying computational times and ability to achieve minima<sup>34</sup>. As a result, optimization research has been a long documented component on parameter estimation for biological models<sup>35-37</sup>.

To determine the optimization algorithm most appropriate for parameterizing our NFκB model, the performance of three optimization approaches when fitting a single cell were documented. Parameter spaces were restricted between fold changes within x0.1-10 of the original parameter set to constrain computational time and stay within biologically relevant values documented by Adelaja, Taylor et al. 2021:

1. *Random Sampling* – A sample size of 100,000 randomly sampled parameter sets was documented to effectively cover the x0.1-10 search space. The parameter sets that possess the lowest RMSDs out of 100,000 samples were evaluated as the optimal fit.
2. *Gradient Descent (Local Optimization)* - Given a random starting parameter set, parameters were iteratively varied and the RMSD is calculated at each iteration. RMSDs lower than the previous iteration were accepted. The process continues until the RMSD converges at a local minima. Each start seed produces an independent local minima, and the lowest RMSDs at convergence were evaluated as the best fit.
3. *Particle Swarm Optimization (PSO) (Global Optimization)* - A number of candidate parameter sets ('particles') were defined and moved around the search space. Information about the parameter space is shared among particles such that all particles move toward a global minimum. The swarm defines a single global minimum as the optimal fit.

The performance of these three optimization algorithms is documented in Figure 3.1. Random sampling requires the least computational time but does not reach the lowest minimum. PSO has the highest computational time but achieves the lowest RMSD minimum by magnitude. Gradient descent optimization, when given 50 starting positions, has an intermediate computational time in comparison to the two other methods, and achieves minima slightly higher than PSO. However, increasing the number of starting positions to 200 led to minima with RMSDs comparable to PSO. All three optimization methods produce lowest RMSD trajectories with comparable dynamic features. Gradient descent optimization was selected as the preferred method for fitting our NFκB model given lower computational time in comparison to PSO. In addition, gradient descent's multiple solution output allowed for characterization of the variation of dynamic features at low RMSD parameter sets.

### 3.2. Quantifying NFκB model fit to multiple cells and stimulus conditions

The above optimization method performances were documented when calculating RMSDs for a single cell. However, quantitatively fitting the NFκB model requires optimization to multiple representative cells and stimulus conditions. Specifically, fitting the core module required all datasets, as the core parameters are involved in response to all stimuli. Alternatively, fitting the receptor proximal modules upstream of IKK required using stimulus specific datasets. RMSD calculations for each dose, stimulus, and the core module are detailed in Equations 3.1, 3.2, and 3.3.

$$\text{(Eq. 3.1)} \quad RMSD_{dose} = \frac{RMSD_{1\_cell} + RMSD_{2\_cell} + RMSD_{3\_cell} + \dots + RMSD_{n\_cell}}{n_{cell}}$$

$$\text{(Eq. 3.2)} \quad RMSD_{stimulus} = \frac{RMSD_{1\_dose} + RMSD_{2\_dose} + \dots + RMSD_{n\_dose}}{n_{dose}}$$

$$\text{(Eq. 3.3)} \quad RMSD_{core} = \frac{RMSD_{1\_stimulus}^{s_{1\_weight}} + \dots + RMSD_{n\_stimulus}^{s_{n\_weight}}}{n_{stimulus}}$$

$RMSD_{n\_cell}$  is described by Equation 2.1.  $n_{cell}$  represents the number of representative cells.  $n_{dose}$  represents the number of doses for a stimulus.  $n_{stimulus}$  represents the number of stimuli datasets.  $s_{n\_weight}$  is the weighting factor of each stimulus when calculating the RMSD of the core ( $RMSD_{core}$ ). Table 3.1 documents  $s_{n\_weight}$  values. TNF datasets are highly weighted given that TNF is the only host inflammatory cytokine represented in these data. Signaling codon weights for  $RMSD_{core}$  and  $RMSD_{stimulus}$  are documented in Table 2.1.

### 3.3. Parameter sensitivity scan to constrain the $NF\kappa B$ model search space

Our  $NF\kappa B$  model has a large parameter space with 95 biochemical reactions. Searching the entire parameter space was not required, as not all parameters influence signaling codons and there is no experimental or computational justification for changing their value. As a result, the search space was limited to sensitive parameters, or parameters in which small changes in parameter value led to large changes in RMSD. Constraining the optimization to sensitive parameters restricted the search space to parameters that control signaling codons and reduced computational time.

A parameter sensitivity scan was implemented to identify sensitive parameters. For each module of the  $NF\kappa B$  model, 100,000 random parameter sets were sampled and  $RMSD_{stimulus}$  and  $RMSD_{core}$  were calculated for the modules of the  $NF\kappa B$  model. The distribution of parameter sets with RMSDs in the lowest 1% out of 100,000 random samples were identified (Figure 3.2). These lowest RMSD parameter distributions were then compared with randomly sampled uniform distributions via the Kolmogorov–Smirnov (KS) test. Parameters were identified as sensitive if the KS test gave a p-value  $<1 \times 10^{-16}$ . 34 out of 95 parameters were identified as sensitive. Table 3.2 lists the identified sensitive parameters and their fold change bounds during model optimization. The fold change constraints for core module reaction 6, parameter 4 ( $I\kappa B\alpha$  mRNA transcription/processing/maturation delay) were set within x1-1.5 to maintain documented  $NF\kappa B$  oscillation periods of 70-110 min<sup>38</sup>.

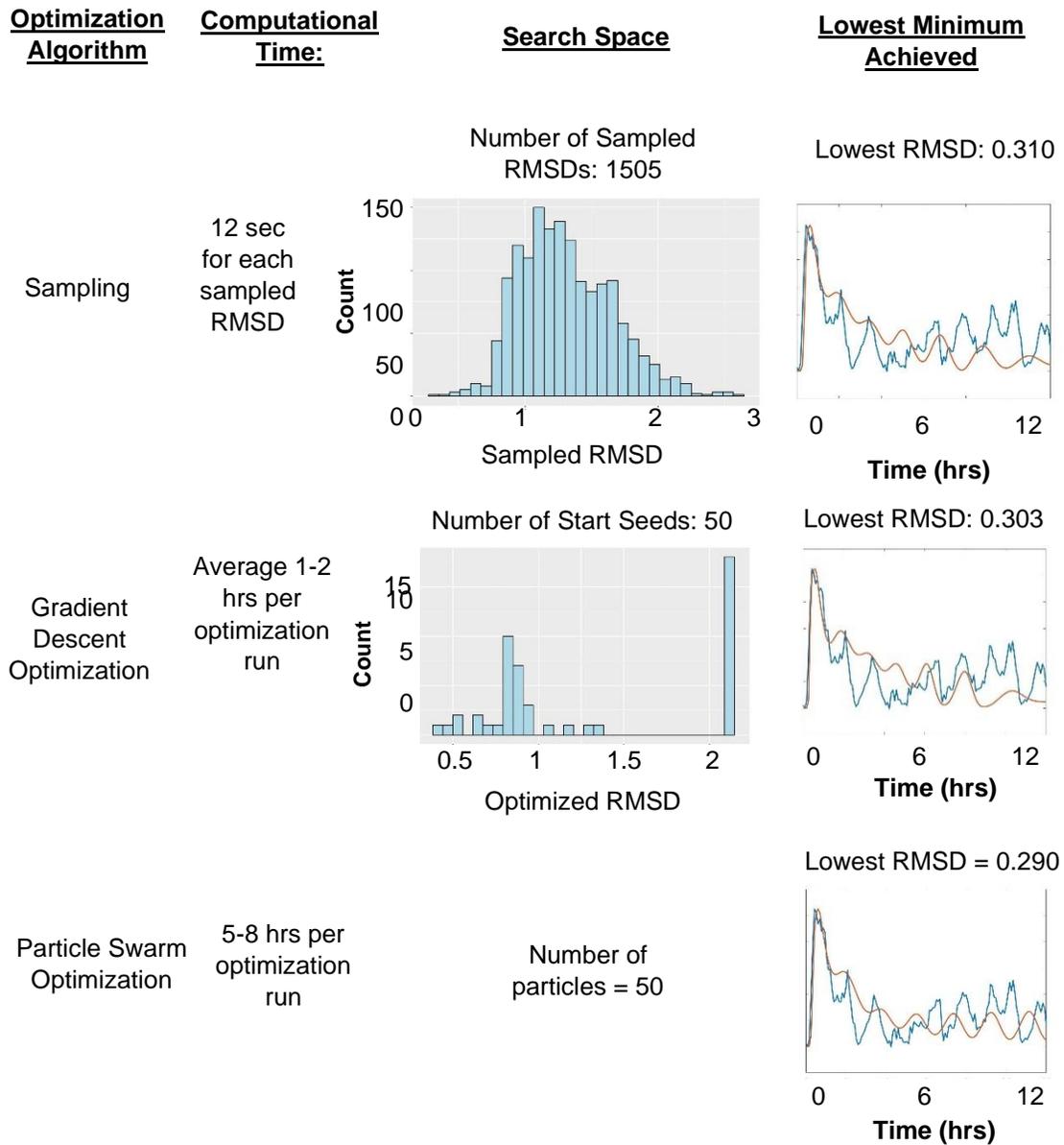
### *3.4. Iterative approach to fit each module given shared pathway parameters*

As each module of the NF $\kappa$ B model required fitting to stimulus specific datasets, an iterative approach that fits each module separately was implemented. First, gradient descent optimization of the sensitive parameters determined a parameter set that gave the optimal  $\text{RMSD}_{\text{core}}$ . Next, these optimized core module parameters were used as the starting point to fit the receptor proximal modules. At the end of the iteration, the optimization returns to the core module after optimizing the receptor proximal modules. Figure 3.3. details the iterative module fitting process. Figure 3.4 shows a schematic of gradient descent optimization of the TLR9 module.

### *3.5. Optimized NF $\kappa$ B model to signaling codons*

All modules of the NF $\kappa$ B model were fit to the 17 experimental conditions using the feature based objective function and gradient descent optimization. The optimized parameter fold changes are documented in Table 3.2. Signaling codons were markedly improved for all conditions (Figure 3.5). Notable signaling codon improvements to the data include: persistent duration throughout the 12 hour time course for all doses of TNF, and lessened oscillatory character at CpG 100nM and LPS 3ng/mL, and longer oscillatory periods for all Poly(I:C) doses (Figure 3.6).

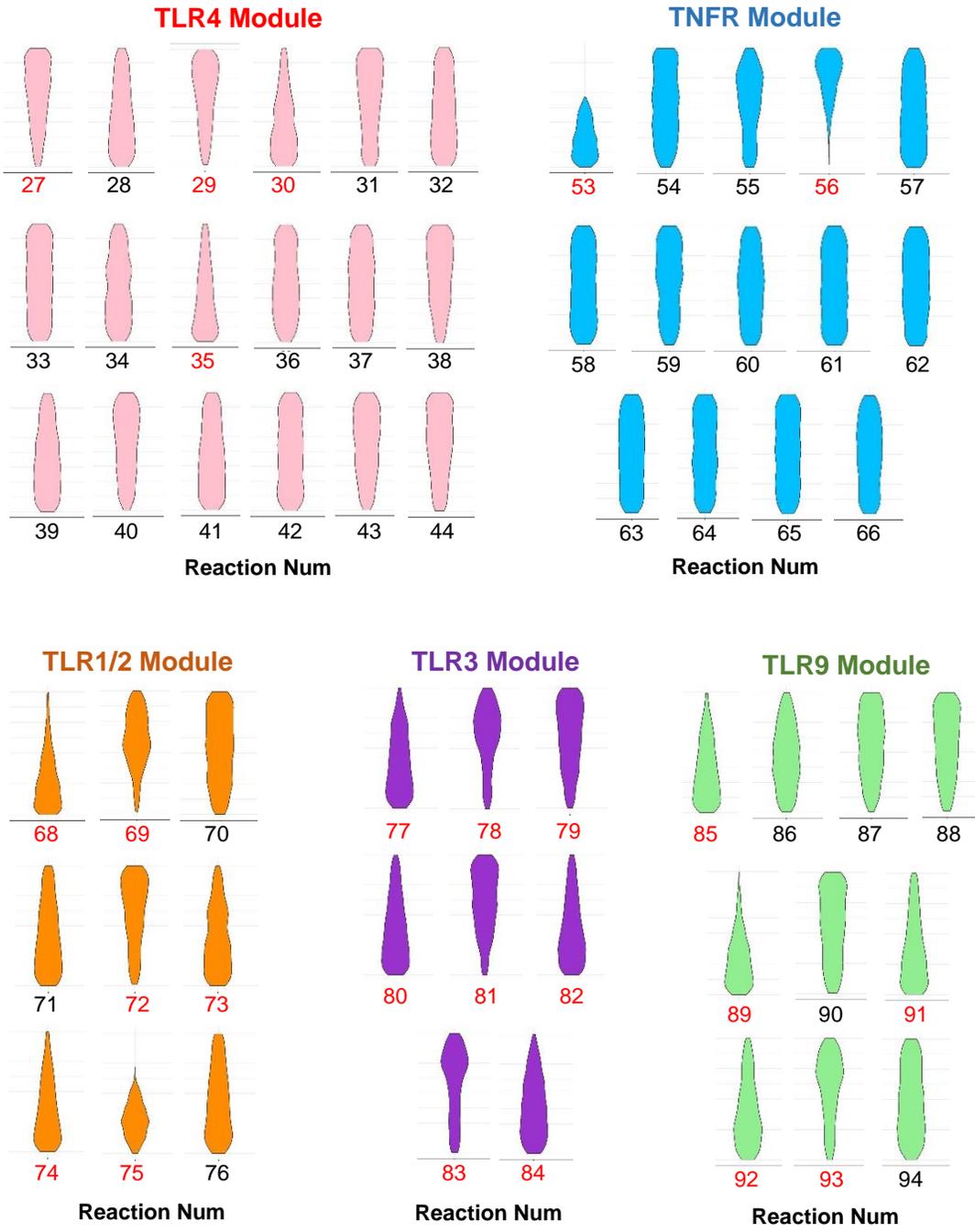
### 3.6. Figures



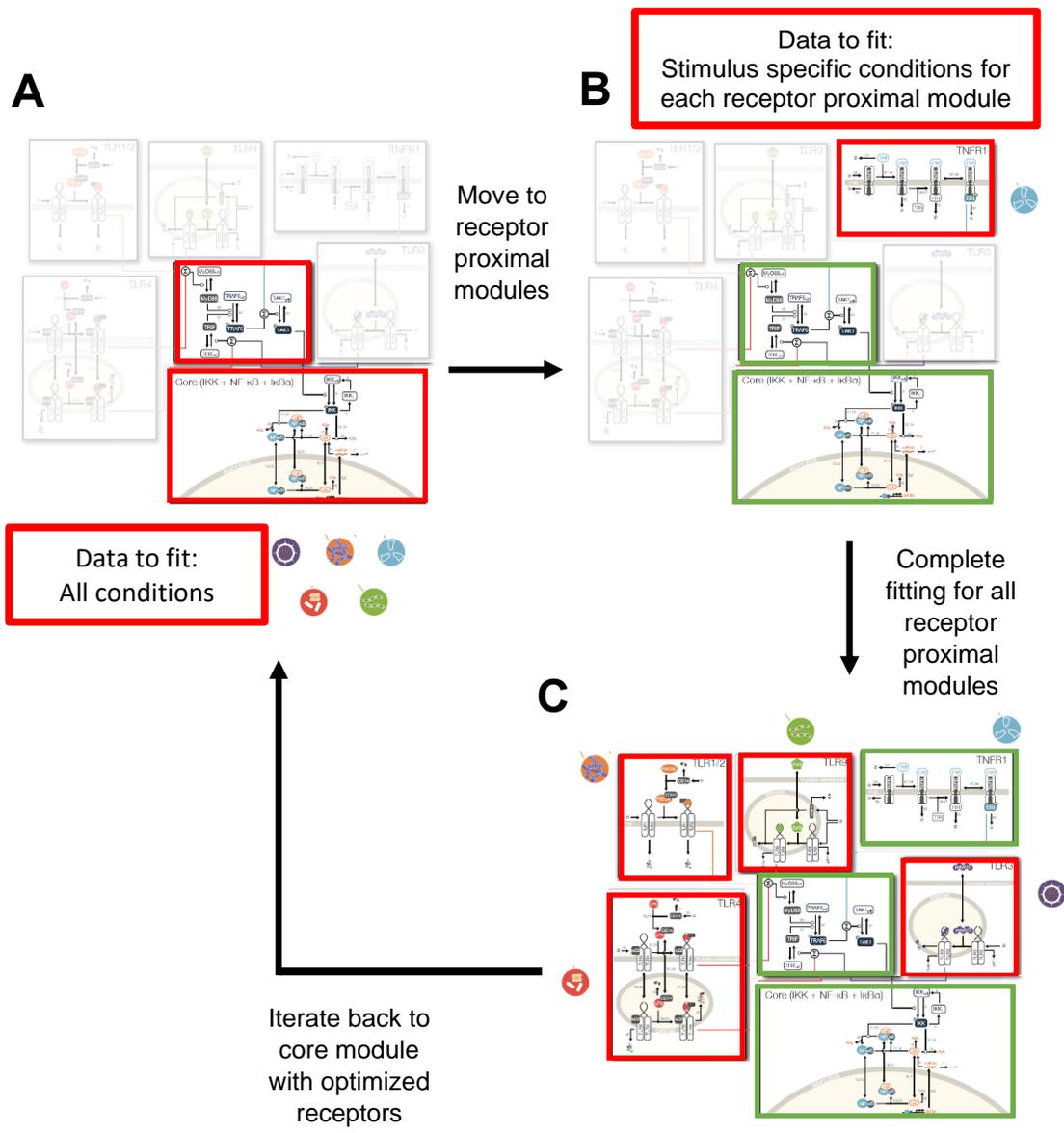
**Figure 3.1 Performance of varying optimization algorithms when fitting our NF $\kappa$ B model to a single cell.**

Stimulus	Dose	Source	Stimulus Weight $S_{n\_weight}$
CpG	1 $\mu$ M	PAMP	1.25
CpG	333nM		
CpG	100nM		
CpG	33nM		
LPS	100ng	PAMP	1.25
LPS	33ng		
LPS	10ng		
LPS	3ng		
LPS	1ng		
Pam3CSK	1 $\mu$ g	PAMP	1.25
Pam3CSK	100ng		
Pam3CSK	10ng		
TNF	10ng	Host Inflammatory Cytokine	5
TNF	1ng		
Poly(I:C)	100 $\mu$ g	PAMP	1.25
Poly(I:C)	33 $\mu$ g		
Poly(I:C)	10 $\mu$ g		

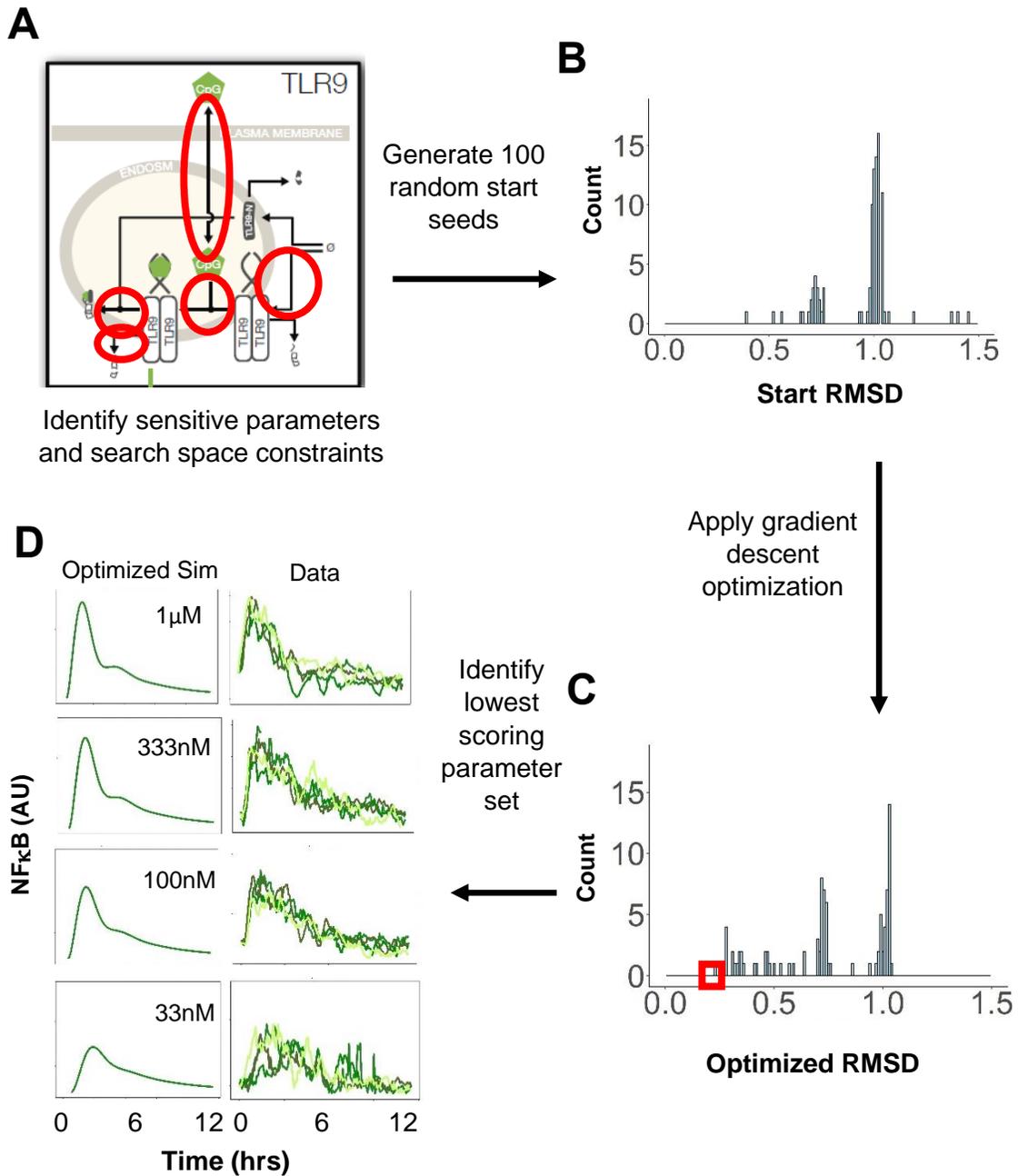
**Table 3.1. Stimulus weights for calculating  $RMSD_{core}$ .**



**Figure 3.2. Parameter distributions for the lowest  $\text{RMSD}_{\text{stimulus}}$  parameter sets for the five receptor proximal modules.** Sensitive parameters are labeled in red. The distribution for each plot ranges from  $\log(-1)$  to  $\log(1)$ .



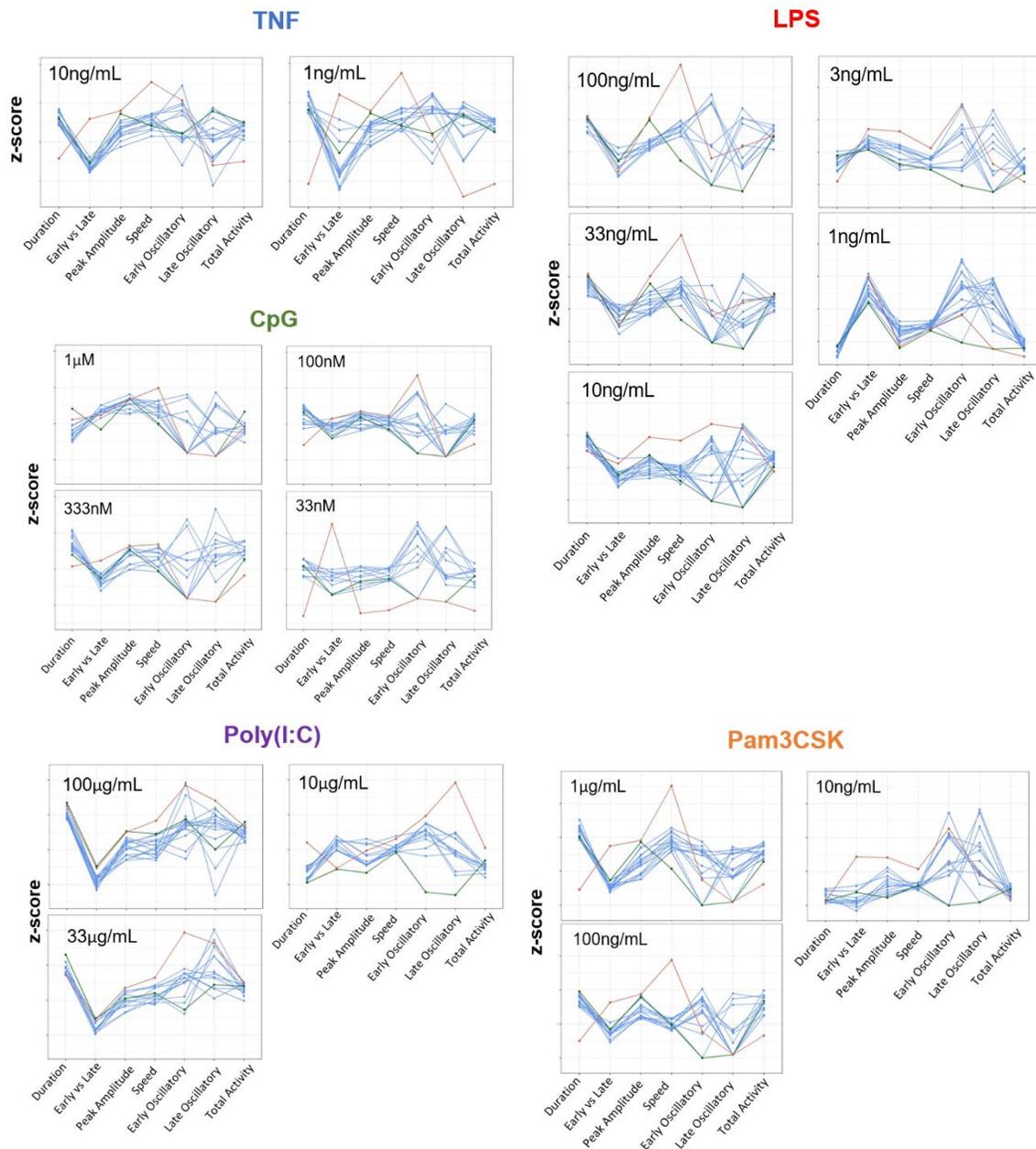
**Figure 3.3. Optimization pipeline iterates between core module and receptor-proximal modules using stimulus specific datasets.** A. The parameters of the core module are first allowed to vary (red box) to fit all datasets while the receptor proximal module parameters remain fixed (faded boxes). B. The optimized core module then remains fixed (green boxes) while the parameters of each of the receptor proximal modules are fit to stimulus-specific datasets. (TNFR module shown) C. One round of optimization is completed once the remaining receptor proximal modules are fit (red boxes). The next iteration begins with the optimized parameter set from (C) as the starting point.



**Figure 3.4. Gradient descent optimization finds the optimal parameter set to fit NF $\kappa$ B signaling codons.** A. For each module (TLR9 shown), sensitive parameters and search space constraints are identified. B. 100 random start parameter sets are generated using the constraints from (A) and start RMSDs are calculated. C. Gradient descent is applied to each start seed so that the distribution shifts to lower RMSD values. Lowest RMSD candidate is shown in red box. D. Optimized model trajectories of the lowest RMSD solution next to representative experimental cells.

Module	Reaction Num	Param Num	Parameter Name	Fold Change Lower Bound	Fold Change Upper Bound	Optimized Fold Change
Core	6	4	I $\kappa$ B $\alpha$ mRNA transcription/ processing/ maturation delay	1	1.5	1.423
Core	14	1	Nuclear export of I $\kappa$ B $\alpha$ -NF $\kappa$ B	0.1	10	2.493
Core	21	1	IKK-I $\kappa$ B $\alpha$ -NF $\kappa$ B association	0.1	10	0.368
Core	25	1	Phosphorylation/ degradation of complexed I $\kappa$ B $\alpha$	0.1	10	0.172
Core	52	1	Activation of TAK1 by TRAF6	0.1	10	0.288
Core	67	2	Hill coefficient for IKK activation	0.1	10	0.632
TLR4	27	1	CD14-LPS association	0.1	10	1.111
TLR4	29	1	CD14 synthesis	0.1	10	0.353
TLR4	30	1	CD14 degradation	0.1	10	0.316
TLR4	35	1	Synthesis rate of TLR4	0.1	10	0.154
TNFR	53	1	TNF degradation	0.1	10	0.531
TNFR	56	1	TNF/TNFR Association	0.1	10	6.006
TLR1/2	68	1	TLR1/2 synthesis	0.1	10	0.427
TLR1/2	69	1	TLR1/2 degradation	0.1	10	1.477
TLR1/2	72	1	Association of CD14/P3CSK and TLR2	0.1	10	6.134
TLR1/2	73	1	Degradation of CD14-P3CSK	0.1	10	2.184
TLR1/2	74	1	Dissociation of CD14/P3CSK and TLR2	0.1	10	4.300
TLR1/2	75	1	Degradation of TLR2/P3CSK	0.1	10	0.254
TLR3	77	1	TLR3 synthesis	0.1	10	0.811
TLR3	78	1	TLR3 degradation	0.1	10	0.730
TLR3	79	1	Poly(I:C) internalization	0.1	10	6.770
TLR3	80	1	Poly(I:C) exit from endosome	0.1	10	9.967
TLR3	81	1	Poly(I:C)-TLR3 association	0.1	10	0.348
TLR3	82	1	Poly(I:C)-TLR3 dissociation	0.1	10	3.900
TLR3	83	1	Bound poly(I:C)-TLR3 degradation	0.1	10	2.593
TLR3	84	1	Activation of TRIF by bound TLR3	0.1	10	3.994
TLR9	85	1	TLR9 synthesis	0.1	10	0.155
TLR9	89	1	CpG exchange from endosome	0.1	10	0.252
TLR9	91	1	CpG-TLR9 dissociation	0.1	10	0.100
TLR9	92	1	Mediated degradation of bound TLR9 by TLR9 N-terminus fragment	0.1	10	0.879
TLR9	93	1	Bound CpG-TLR9 degradation	0.1	10	0.332

**Table 3.2. Sensitive parameters of the NF $\kappa$ B model.**



**Figure 3.5. Signaling codons of all NFκB datasets show improved fit to representative cells after feature based optimization.** Each line connects signaling codons of the same cell. Original model signaling codons – red line. Optimized model – green line. Representative cells – blue lines.

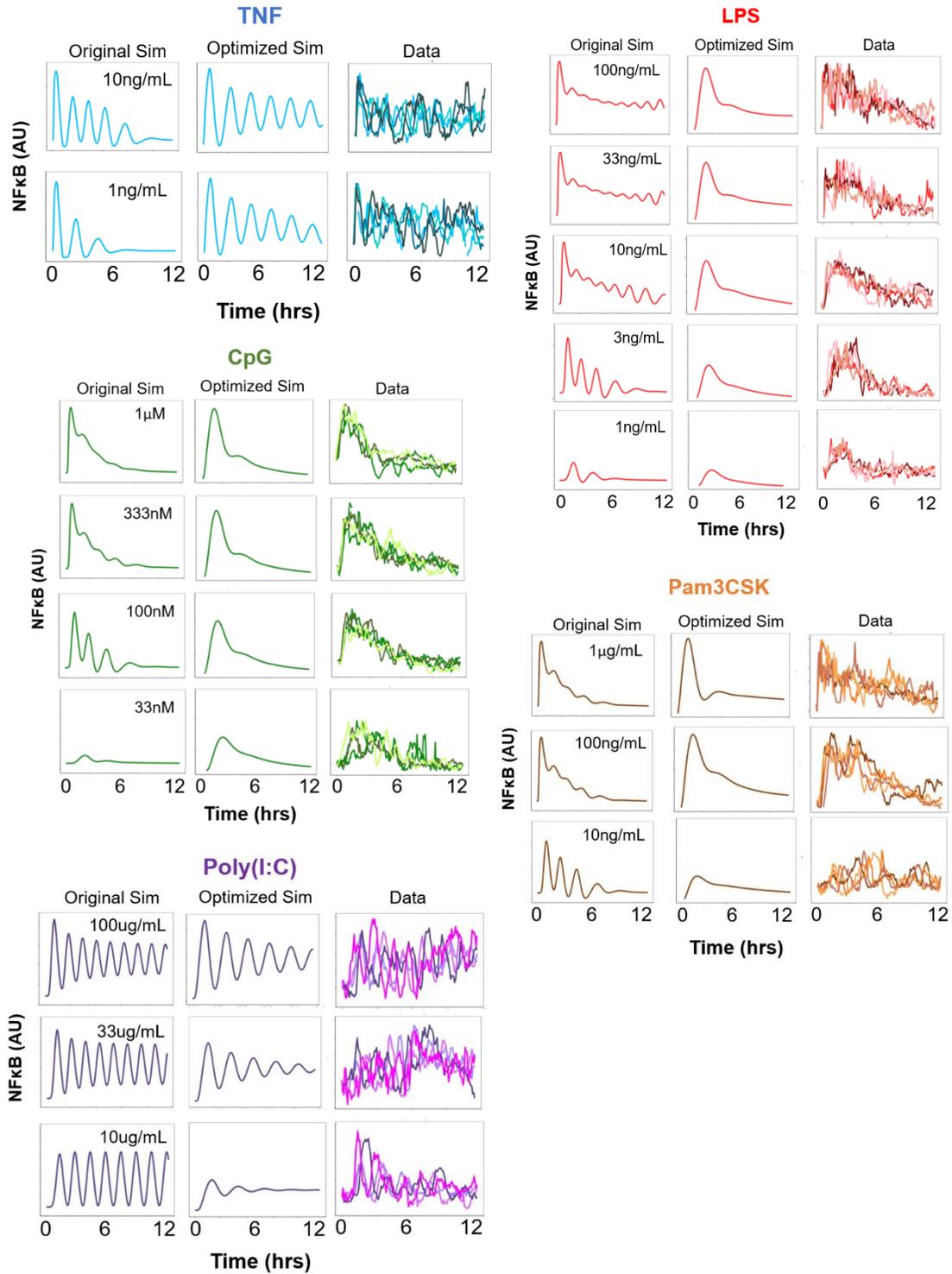


Figure 3.6. Optimized NFκB trajectories to signaling codons.

# **Chapter 4: Methods**

#### 4.1. Experimental data generation

Bone marrow-derived macrophages (BMDMs) generated from an mVenus-RelA knock-in reporter mouse strain were stimulated within a live cell microscopy workflow with indicated ligands in paracrine-TNF-blocked conditions using soluble TNFRI. Nuclear NF $\kappa$ B levels at single cell resolution were measured at a 5 minute frame rate. The measured fluorescence intensity was further normalized to image background levels by using an automated image analysis workflow (<https://github.com/Adewunmi91/MACKtrack>). Experimental data was generated by collaborator Dr. Adewunmi Adelaja and is available upon request.

#### 4.2. Baseline normalization for NF $\kappa$ B signaling trajectories

Signaling codons were calculated from 12 hour single cell NF $\kappa$ B signaling trajectories. For each trajectory, the stimulus was added at the second frame (after 5 minutes of tracking). The trajectory baselines were normalized such that either the first or second frame, whichever frame was lower, (pre-stimulus period) was shifted to 0 AU.

#### 4.3. Scaling NF $\kappa$ B model trajectories from SI units to experimental AU

Model trajectories were scaled to experimental AU ( $model_{AU}$ ) using Equation 4.1.  $max_{AU}$  is the average maximum NF $\kappa$ B in AU achieved by the representative cells at the highest dose of each stimulus.  $min_{AU}$  is 0 as all experimental trajectories were normalized to a baseline of 0 AU.  $max_{SI}$  is the maximum NF $\kappa$ B in SI units achieved by the published model (Adelaja, Taylor et al. Immunity 2021) at the top dose of each stimulus.  $min_{SI}$  is the model steady state value of NF $\kappa$ B in SI units before stimulation.  $model_{SI}$  is the simulated NF $\kappa$ B time series in SI units.

$$\text{(Eq. 4.1)} \quad model_{AU} = \frac{(max_{AU} - min_{AU})(model_{SI} - min_{SI})}{max_{SI} - min_{SI}}$$

#### 4.4. Signaling codon definitions

Experimental trajectories were first smoothed using the ‘moving’ method for the smooth function in the Curve Fitting Toolbox from MathWorks. Smoothing span was set to 0.03. Adelaja, Taylor

et al. 2021 signaling codon metrics were used for duration, peak amplitude, total activity, speed, and early vs late. The oscillatory codon metrics were adapted to 'sum peak to trough ratios' and 'number of oscillatory peaks.' Peaks were identified using the findpeaks function from the Signal Processing Toolbox of MathWorks. Peaks due to technical noise were identified and removed if the prominence value was less than 0.15 AU and neighboring peaks were less than 35 min apart. Peaks were identified as indicative of NF $\kappa$ B oscillations if the time between peaks fell within 50-180 min. For non-responding cells, all signaling codon metrics were assigned values of 0 by default, except the fold change metric which was assigned a value of 1.

#### *4.5. Defining non-responding cells*

Each cell was identified as a 'non-responder' if over 85% of the time series frames fell below the non-responding baseline of 1 AU. These parameters were benchmarked such that unstimulated conditions were identified to contain more than 95% non-responding cells. The representative response of single cell datasets was characterized as 'non-responding' if over 50% of cells in that dataset were labeled as 'non-responders.'

#### *4.6. Selecting representative cells*

Representative cells were selected for all conditions in which over 50% of cells were identified as responders. For each condition, non-responding cells were removed and signaling codon distributions for duration, peak amplitude, early vs late, oscillatory, speed, and total activity were generated using responding cells only. The mode was calculated for each signaling codon distribution. Cells were identified as 'representative' if their signaling codon values were located within thresholds of the distribution mode. The size of upper and lower thresholds from the mode were initially defined as 5% of the interquartile range (IQR). Both thresholds were iteratively increased by 5% until 10-20 representative cells were identified.

#### 4.7 Identifying sensitive parameters

For each module of the NF $\kappa$ B model, 100,000 random parameter sets were sampled in log space with a range of -1 to 1 fold change, and  $RMSD_{stimulus}$  and  $RMSD_{core}$  were calculated for the modules of the NF $\kappa$ B model. The distribution of parameter sets with RMSDs in the lowest 1% out of 100,000 random samples were identified (Figure 3.2). These lowest RMSD parameter distributions were then compared with randomly sampled uniform distributions via the Kolmogorov–Smirnov (KS) test using the `k.test` function in R. Parameters were identified as sensitive if the KS test gave a p-value  $<1 \times 10^{-16}$ .

#### 4.8. Particle Swarm Optimization (PSO) to a single cell

PSO was implemented using the `particleswarm` function from the Optimization Toolbox of MathWorks using 50 particles.  $RMSD_{cell}$  was applied as the distance metric for optimization performance documentation in Chapter 3. `HybridFcn` was set to 'fmincon.' Parameters were searched at log scale within the fold changes described by Table 3.2.

#### 4.9. Gradient Descent Optimization for the full NF $\kappa$ B model

The `fminsearch` function from the Optimization Toolbox of MathWorks was used to apply gradient descent optimization to our NF $\kappa$ B model.  $RMSD_{cell}$  was used as the distance metric for optimization performance documentation in Chapter 3. The distance metrics for fitting the receptor proximal modules and the core module were  $RMSD_{stimulus}$  (Equation 3.2) and  $RMSD_{core}$  (Equation 3.3) respectively. The termination tolerance on the parameter values (`TolX`) was  $1e-03$  while the termination tolerance on the function value (`TolFun`) was kept at the default of  $1e-04$ . Parameters were searched at log scale within the fold changes described by Table 3.2. For each module, 200 random start seeds were optimized and the lowest 5-10 solutions based on RMSD were evaluated as candidates of best fit.

# **Chapter 5: Conclusion**

## 5.1. Discussion

This dissertation work provides a framework for mathematical modeling biological processes given single cell heterogeneity, large parameter search spaces, and sizable experimental datasets for model fitting. Most notably, I developed a feature-based approach for optimizing a mathematical model of NF $\kappa$ B response to fit signaling codons informative of the identity of the stimulus and dose. My model fitting approach addressed key challenges of recapitulating single cell data, including: 1. handling high technical noise of single cell datasets, which limits measurement of biological information, and 2. defining a ‘representative’ NF $\kappa$ B response given biological heterogeneity. In addition, I developed an optimization pipeline for fitting our large NF $\kappa$ B model of 95 biochemical reactions to 17 single cell NF $\kappa$ B signaling datasets of five stimuli at varying doses.

In **Chapter 2**, I constructed a feature-based objective function based on the informative signaling codons of NF $\kappa$ B response identified by Adelaja, Taylor et al. 2021. Standard model fitting approaches for time series data calculate the distance between model simulations and time series data points directly, which subjects model simulations to the effects of technical noise and does not preserve biologically documented dynamic features (Figure 2.2). My feature-based approach addresses this challenge by fitting to informative signaling codons directly to preserve biological behavior and information. In addition, I developed visualization approaches to document signaling codon discrepancies between model and experiment in a more rigorous and quantitative manner (Figure 2.5). For each condition, my feature-based approach addressed biological heterogeneity by fitting to ‘representative’ NF $\kappa$ B response defined by signaling codon values at the mode of each signaling codon distribution. This definition was fitting for our NF $\kappa$ B datasets with single mode signaling codon distributions.

The weighting of each signaling codon for my feature-based objective function (Table 2.1) was informed by mutual information and classification analysis by Adelaja, Taylor et al. 2021. For

example, given that eliminating the speed codon greatly decreased the certainty of ligand prediction for CpG (Figure 2.1b), the speed codon was highly weighted when fitting the TLR9 module. These weights were determined qualitatively and led to successful optimization of our NF $\kappa$ B model (Figure 3.5-3.6). However, the weighting of signaling codons can be determined in a more quantitative approach to further improve the optimization workflow.

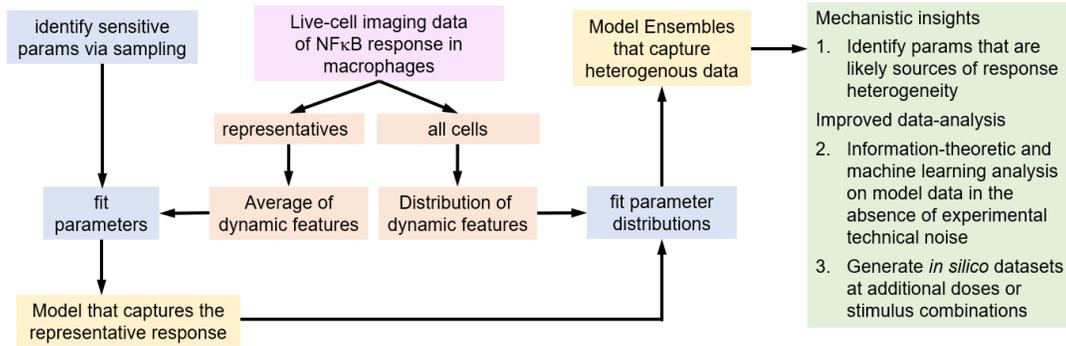
In **Chapter 3**, I developed an optimization pipeline for fitting our large NF $\kappa$ B model of 95 biochemical reactions to single cell NF $\kappa$ B signaling datasets at varying stimuli and doses. After documenting the performance of multiple optimization algorithms, I selected gradient descent optimization as the preferred method for fitting our NF $\kappa$ B model given reasonable computational time and ability to achieve minima. Given the large parameter search space of 95 biochemical reactions, I identified sensitive parameters in our NF $\kappa$ B model that specifically tuned signaling codons. Lastly, I developed an approach that iteratively optimizes the core and receptor proximal modules using stimulus specific datasets appropriate for each module. This optimization pipeline successfully improved the fit of our NF $\kappa$ B model to signaling codons in all 17 experimental conditions.

One observation for the optimized NF $\kappa$ B model trajectories with sustained behavior was that only 12% of the total NF $\kappa$ B in the cell was translocated into the nucleus at top doses of stimulus. This magnitude is not consistent with documented NF $\kappa$ B dynamics in which 50% or more of the total NF $\kappa$ B in the cell translocates into the nucleus during stimulation. In addition, nuclear NF $\kappa$ B fold change was reported to determine appropriate TNF-induced gene expression<sup>39</sup>. The current optimization workflow loses information about the magnitude of NF $\kappa$ B activation when the model simulations are scaled to experimental AU. The ability of a low magnitude NF $\kappa$ B activation to fit sustained dynamics suggests that these trajectories operate at the low end of the IKK regime, which the Hopf bifurcation analysis allows for non-oscillatory

behavior. These low NF $\kappa$ B activation trajectories are inconsistent with the finding that MyD88 ligands produce high and sustained IKK levels<sup>40</sup>. To address this issue, the feature-based objective function (Equation 2.1) can be updated to include information about the percentage of total NF $\kappa$ B found in the nucleus during activation.

Future directions for the NF $\kappa$ B model are to simulate biological heterogeneity. Current studies have explored various approaches, including linear mixed effects modeling, to estimate parameter distributions for biological systems<sup>41–43</sup>. However, these methods distort dynamic features using averaging techniques or are limited to objective functions that optimize to time series data points directly. The feature-based objective function developed in this dissertation can be applied to estimate parameter distributions for other biological systems with evidence of a dynamic code. In addition, the pipeline for optimizing the model to representative cell response is valuable for identifying the appropriate parameter ranges for parameter distributions (Figure 5.1). Applications of these single cell resolution models include determining key circuit design principles for stimulus specific encoding, identifying biological sources of single cell heterogeneity, analyzing information flow through intermediates in the signal transduction network that are difficult to characterize experimentally, and generate *in silico* datasets composed of thousands of doses unreachable by experiments.

## 5.2. Figures



**Figure 5.1. Developing a mathematical model of NFκB response to simulate single cell heterogeneity.**

# References

1. Gerondakis, S. *et al.* Unravelling the complexities of the NF- $\kappa$ B signalling pathway using mouse knockout and transgenic models. *Oncogene* **25**, 6781–6799 (2006).
2. Sheu, K. M., Luecke, S. & Hoffmann, A. Stimulus-specificity in the responses of immune sentinel cells. *Curr. Opin. Syst. Biol.* **18**, 53–61 (2019).
3. Chawla, M., Roy, P. & Basak, S. Role of the NF- $\kappa$ B system in context-specific tuning of the inflammatory gene response. *Curr. Opin. Immunol.* **68**, 21–27 (2021).
4. Bacher, Susanne; Schmitz, M. L. The NF $\kappa$ B Pathway as a Potential Target for Autoimmune Disease Therapy. *Bentham Sci. Publ.* **10**, 2827–2837 (2004).
5. Lawrence, T. The Nuclear Factor NF- $\kappa$ B Pathway in Inflammation. *Cold Spring Harb. Lab. Press* **1**, 1–10 (2009).
6. Kearns, J. D., Basak, S., Werner, S. L., Huang, C. S. & Hoffmann, A. I $\kappa$ B $\epsilon$  provides negative feedback to control NF- $\kappa$ B oscillations, signaling dynamics, and inflammatory gene expression. *J. Cell Biol.* **173**, 659–664 (2006).
7. Sun, S. C., Ganchi, P. A., Ballard, D. W. & Greene, W. C. NF-kappa B controls expression of inhibitor I kappa B alpha: evidence for an inducible autoregulatory pathway. *Science (80- )*. **259**, 1912 LP – 1915 (1993).
8. Ghosh, S., May, M. J. & Kopp, E. B. NF- $\kappa$ B AND REL PROTEINS: Evolutionarily Conserved Mediators of Immune Responses. *Annu. Rev. Immunol.* **16**, 225–260 (1998).
9. Pahl, H. L. Activators and target genes of Rel/NF- $\kappa$ B transcription factors. *Oncogene* **18**, 6853–6866 (1999).
10. Hoffmann, A., Levchenko, A., Scott, M. L. & Baltimore, D. The I $\kappa$ B-NF $\kappa$ B signaling module: temporal control and selective gene activation. *Science* **298**, 1241–5 (2002).
11. Purvis, J. E. & Lahav, G. Encoding and Decoding Cellular Information through Signaling Dynamics. *Cell* **152**, 945–956 (2013).
12. Behar, M. & Hoffmann, A. Understanding the temporal codes of intra-cellular signals. *Curr. Opin. Genet. Dev.* **20**, 684–693 (2010).
13. Basak, S., Behar, M. & Hoffmann, A. Lessons from mathematically modeling the NF- $\kappa$ B pathway. *Immunol. Rev.* **246**, 221–238 (2012).
14. Hoffmann, A. & Baltimore, D. Circuitry of nuclear factor kappaB signaling. *Immunol. Rev.* **210**, 171–186 (2006).
15. Covert, M. W., Leung, T. H., Gaston, J. E. & Baltimore, D. Achieving Stability of Lipopolysaccharide-Induced NF- $\kappa$ B Activation. *Science (80- )*. **309**, 1854–1857 (2005).
16. Werner, S. L., Barken, D. & Hoffmann, A. Stimulus Specificity of Gene Expression Programs Determined by Temporal Control of IKK Activity. *Science (80- )*. **309**, 1857–1861 (2005).
17. Cheng, Q. J. *et al.* NF $\kappa$ B dynamics determine the stimulus-specificity of epigenomic reprogramming in macrophages. *bioRxiv* 2020.02.18.954602 (2020) doi:10.1101/2020.02.18.954602.
18. Sen, S., Cheng, Z., Sheu, K. M., Chen, Y. H. & Hoffmann, A. Gene Regulatory Strategies that Decode the Duration of NF $\kappa$ B Dynamics Contribute to LPS- versus TNF-Specific

- Gene Expression. *Cell Syst.* **10**, 169-182.e5 (2020).
19. Hao, N. & O'Shea, E. K. Signal-dependent dynamics of transcription factor translocation controls gene expression. *Nat. Struct. Mol. Biol.* **19**, 31–39 (2012).
  20. Pelkmans, L. Using Cell-to-Cell Variability—A New Era in Molecular Biology. *Science (80-. )*. **336**, 425 LP – 426 (2012).
  21. Elowitz, M. B., Levine, A. J., Siggia, E. D. & Swain, P. S. Stochastic gene expression in a single cell. *Science* **297**, 1183–6 (2002).
  22. Snijder, B. & Pelkmans, L. Origins of regulated cell-to-cell variability. *Nat. Rev. Mol. Cell Biol.* **12**, 119–125 (2011).
  23. Raj, A. & van Oudenaarden, A. Nature, nurture, or chance: stochastic gene expression and its consequences. *Cell* **135**, 216–226 (2008).
  24. Tang, Y. *et al.* Quantifying information accumulation encoded in the dynamics of biochemical signaling. *Nat. Commun.* **12**, 1272 (2021).
  25. Cheong, R., Rhee, A., Wang, C. J., Nemenman, I. & Levchenko, A. Information transduction capacity of noisy biochemical signaling networks. *Science* **334**, 354–8 (2011).
  26. Selimkhanov, J. *et al.* Accurate information transmission through dynamic biochemical signaling networks. *Science (80-. )*. **346**, 1370–1373 (2014).
  27. Kellogg, R. A. & Tay, S. Noise Facilitates Transcriptional Control under Dynamic Inputs. *Cell* **160**, 381–392 (2015).
  28. Kearns, J. D. & Hoffmann, A. Integrating computational and biochemical studies to explore mechanisms in NF- $\kappa$ B signaling. *J. Biol. Chem.* **284**, 5439–43 (2009).
  29. Takeuchi, O. & Akira, S. Pattern Recognition Receptors and Inflammation. *Cell* **140**, 805–820 (2010).
  30. O'Neill, L. A. J., Golenbock, D. & Bowie, A. G. The history of Toll-like receptors — redefining innate immunity. *Nat. Rev. Immunol.* **13**, 453–460 (2013).
  31. Adelaja, A. *et al.* Six distinct NF $\kappa$ B signaling codons convey discrete information to distinguish stimuli and enable appropriate macrophage responses. *Immunity* **54**, 916-930.e7 (2021).
  32. Wang, Yunjiao; Paszek, Pawel; Horton, Caroline A.; Yue, Hong; White, Michael R.H.; Kell, Douglas B.; Muldoon, Mark R.; Broomhead, D. S. A systematic survey of the response of a model NF- $\kappa$ B signalling pathway to TNF $\alpha$  stimulation. *J. Theor. Biol.* **297**, 137–147 (2012).
  33. Ohshima, Daisuke; Ichikawa, K. Regulation of NF- $\kappa$ B Oscillation by Nuclear Transport: Mechanisms Determining the Persistency and Frequency of Oscillation. *PLoS One* **10**, e0127633 (2015).
  34. Cedersund, G., Samuelsson, O., Ball, G., Tegnér, J. & Gomez-Cabrero, D. Optimization in Biology Parameter Estimation and the Associated Optimization Problem. in *Uncertainty in Biology: A Computational Modeling Approach* (eds. Geris, L. & Gomez-Cabrero, D.) 177–197 (Springer International Publishing, 2016). doi:10.1007/978-3-319-21296-8\_7.

35. Banga, J. R. Optimization in computational systems biology. *BMC Syst. Biol.* **2**, 47 (2008).
36. Gomez-Cabrero, D., Compte, A. & Tegner, J. Workflow for generating competing hypothesis from models with parameter uncertainty. *Interface Focus* **1**, 438–449 (2011).
37. Greenberg, H. J., Hart, W. E. & Lancia, G. Opportunities for Combinatorial Optimization in Computational Biology. *INFORMS J. Comput.* **16**, 211–231 (2004).
38. Mengel, B., Krishna, S., Jensen, M. H. & Trusina, A. Theoretical analyses predict A20 regulates period of NF- $\kappa$ B oscillation. (2009).
39. Lee, R. E. C., Walker, S. R., Savery, K., Frank, D. A. & Gaudet, S. Fold Change of Nuclear NF- $\kappa$ B Determines TNF-Induced Transcription in Single Cells. *Mol. Cell* **53**, 867–879 (2014).
40. Cheng, Z., Taylor, B., Ourthiague, D. R. & Hoffmann, A. Distinct single-cell signaling characteristics are conferred by the MyD88 and TRIF pathways during TLR4 activation. *Sci. Signal.* **8**, ra69–ra69 (2015).
41. Llamosi, Art emis; Gonzalez-Vargas, Andres M.; Versari, Cristian; Cinquemani, Eugenio; Ferrari-Trecate, Giancarlo; Hersen, Pascal; Batt, G. What Population Reveals about Individual Cell Identity: Single-Cell Parameter Estimation of Models of Gene Expression in YeastNo Title. *PLoS Comput. Biol.* **12**, <https://doi.org/10.1371/journal.pcbi.1004706> (2016).
42. Yao, J., Pilko, A. & Wollman, R. Distinct cellular states determine calcium signaling response. *Mol. Syst. Biol.* **12**, 894 (2016).
43. Dharmarajan, L., Kaltenbach, H.-M., Rudolf, F. & Stelling, J. A Simple and Flexible Computational Framework for Inferring Sources of Heterogeneity from Single-Cell Dynamics. *Cell Syst.* **8**, 15-26.e11 (2019).