

Jannatul Ferdous

Candidate

Computer Science

Department

This dissertation is approved, and it is acceptable in quality
and form for publication:

Approved by the Dissertation Committee:

Melanie Moses, Chair

Bruna Jacobson, Member

G. Matthew Fricke, Member

Mousumi Roy, Member

Exploring Immune System through Computational Modeling: A Comprehensive Study of Lymph Nodes and Immune Response Scaling, Vaccine Efficacy, and Large-Scale Extreme First Passage Time

by

Jannatul Ferdous

B.S., Computer Science & Engineering, University of Dhaka, 2016

M.S., Computer Science, University of New Mexico, 2023

DISSERTATION

Submitted in Partial Fulfillment
of the Requirements for the Degree of

Doctor of Philosophy

Computer Science

The University of New Mexico

Albuquerque, New Mexico

July 2025

DEDICATION

This dissertation is dedicated to my loving husband, Parvez

my daughter Inaya, my parents

my siblings, Tauhid, Saifur, and Kaniz

ACKNOWLEDGMENTS

I would like to express my sincere gratitude to all those who have supported and guided me throughout the journey of completing this Ph.D. dissertation.

First and foremost, I am deeply thankful to my supervisor, Professor Melanie Moses, for her unwavering support, invaluable guidance, and continuous encouragement. She helped me grow as an independent researcher by teaching me how to approach a research problem, write a research paper, and effectively communicate the work to the audience. I highly valued our weekly meetings, which served as crucial checkpoints to keep me on track academically and provided plenty of encouragement. I will always admire her positive attitude, which I like to reflect in my professional and personal life. I am particularly grateful for her support during my family's medical emergency and during my pregnancy. I am profoundly thankful for the immeasurable contributions she made to my development.

I am deeply grateful to Dr. Matthew Fricke for his support and guidance throughout my research projects. His remarkable patience, insightful wisdom, and constant encouragement have had a profound impact on both my academic journey and personal growth. I have learned a great deal from his thoughtful approach to problem-solving and mentorship, and I aspire to embody the same patience and clarity in my professional path.

I would like to express my deepest gratitude to Dr. Judy L. Cannon for her continuous support throughout my Ph.D. journey. Her guidance, especially in validating the empirical data with the simulation data in my work, has been invaluable in the advancement of our research. As a member of my dissertation proposal committee, her thoughtful feedback shaped the direction of my work. Although she was unable to join my final defense committee, I remain deeply grateful for her continued generosity of time and wisdom.

I extend my appreciation to the members of my dissertation committee, Dr. Matthew Fricke, Dr. Bruna Jacobson, and Dr. Mousumi Roy, for their time and effort in reading and providing feedback on my dissertation. Their insightful comments and thoughtful suggestions have enriched the quality of this work.

I would like to acknowledge the National Science Foundation (NSF) for providing financial support in my doctoral studies. Their support has allowed me to focus on my research and dedicate the necessary time and resources to the successful completion of this dissertation.

I thank the UNM Center for Advanced Research Computing, supported in part by the NSF, for high performance computing resources.

I thank Chris Kempes, Sid Redner, Alan Friedman, and the Moses Biological Computation Lab for the helpful discussions.

I would like to express my gratitude to Dr. Ruian Ke, Dr. Ruy Ribeiro, and Dr. Jasmine Kreig from Los Alamos National Laboratory (LANL) for their time and contributions to my research. I have enjoyed and learned a lot from them during my internship at LANL.

I would like to thank all my fellow labmates at Moses Biological Computation Lab — Humayra, Julie, Carter, Jake, John, Akil, Antonio, Will, Venessa, Wayne, and Shannon for their continuous support and encouragement. Thank you for sharing your expertise, engaging in thoughtful discussions, and contributing to a stimulating and collaborative academic environment.

Heartfelt thanks go to my family members in Bangladesh for their unwavering support, understanding, and encouragement throughout the ups and downs of this challenging academic journey. Their love and belief in me have been a constant source of motivation.

I am forever grateful to my husband, Md Parvez Mollah, for his steadfast belief in me and for being my pillar of support in every step of this journey. Lastly and most importantly, I would like to thank my biggest cheerleader, my daughter Inaya Waziha. Her infectious laughter, and unconditional love have been my guiding light through every late night and early morning. In her hugs, I find the strength to carry on. She is my greatest motivation and my most profound joy.

ABSTRACT

The adaptive immune response is a complex defense mechanism that develops over time to recognize and eliminate pathogens with remarkable precision and durability. This dissertation investigates the dynamics, scaling, and efficiency of the adaptive immune response through a synthesis of computational modeling, mathematical analysis, and agent-based simulations. First, we analyze the topology of the lymphatic network and investigate the T cell search time to find the lymph node that is containing the matching dendritic cell. Second we show how the scaling of lymph node number and volume with body mass, leads to scale-invariant search times for T cells locating antigen-bearing dendritic cells across species. Third, we develop an analytical and numerical framework for extreme first passage time (EFPT) in confined three-dimensional volumes, revealing a transition from inverse-linear to inverse-logarithmic scaling of the fastest searcher’s discovery time as the number of searchers increases. This framework is validated against large-scale Monte Carlo simulations modeling T cells searching for a central target. Finally, we construct an agent-based model of B cell-mediated immunity to examine affinity maturation, antigenic drift, and vaccine efficacy against rapidly mutating viral variants. By representing receptors and epitopes in shape space and simulating somatic hypermutation and clonal selection, our model predicts population dynamics of B cells, antibodies, and antigens over repeated exposures. Together, these studies elucidate fundamental principles of immune surveillance and response timing, with implications for translational vaccine design and broader applications in search theory, chemical kinetics, and complex systems.

TABLE OF CONTENTS

List of Figures		xiii
List of Tables		xxviii
1 Introduction		1
1.1 Modeling the Lymphatic System		8
1.2 Scaling the Lymphatic System and Immune Response		9
1.3 Initial First Contact Time Dependency		10
1.4 Extreme First Passage Time in Large-Scale Models		12
1.5 Examining Affinity Maturation and Antigenic Drift with an Agent-Based Model		13
2 Modeling Immune Search Through the Lymphatic Network		15
2.1 Publication Notes		15
2.2 abstract		16
2.3 Introduction		17
2.4 Related Work		19
2.5 Methods		20

2.5.1	Lymphatic Network Simulation Algorithm	20
2.5.2	Comparing simulated graphs to observation	21
2.5.3	Search Algorithm	22
2.6	Results	24
2.6.1	Modeled Lymphatic Network	24
2.6.2	Predicted Time	25
2.7	Discussion	27
3	Bigger Is Faster: The Scalable Adaptive Immune Response	29
3.1	Publication Notes	29
3.2	Abstract	30
3.3	Introduction	31
3.4	Time to Initiate the Adaptive Immune Response Is the Same in Humans and Mice	34
3.5	Empirical Scaling of Spleen Size	36
3.6	Empirical Scaling of Lymph Node Size and Number	37
3.7	Initial First Contact Times with Respect to Animal Mass . . .	38
3.8	Discussion	43
3.9	Methods	47
3.9.1	Data Collection and Analysis for Animal Mass, Spleen Volume, Lymph Node Volume, and Lymph Node Number	47
3.9.2	Alternative Hypothesis of the Scaling of Spleen Volume	50
3.9.3	Statistical Analysis	51

3.9.4	Derivation of Initial and Mean First Contact Times . .	51
3.9.5	Agent Based Model	57
3.9.6	Fitting Exponents Using the ABM	58
3.10	Mean First Contact Time	62
3.10.1	Mean First Contact Times Within an Abstract Volume	63
3.11	Model Validation	65
3.11.1	Analysis of IFCT model effect on peak T cell population across different lymph node sizes	69
3.12	Caveats and Limitations	71

4 More is Faster: Why Population Size Matters in Biological

Search	78
4.1	Publication Notes 78
4.2	Abstract 80
4.3	Introduction 81
4.4	Predictions and Computational Methods 86
4.4.1	Summary of Scenarios 86
4.4.2	Analysis of t_{fin} 87
4.4.3	Analysis of t_{∞} 94
4.4.4	Computational Experiments 97
4.4.5	Fitting Exponents Using the ABM 98
4.5	Results: Comparing model predictions to simulated data . . . 100
4.6	Discussion 108

4.6.1	Summary	108
4.6.2	Interpretation of scenarios	109
5	An Analysis of Scaling of Extreme First Passage Time for Large-Scale Searcher Models	114
5.1	Abstract	114
5.2	Introduction	115
5.3	Results	120
5.4	Methods	125
5.5	Discussion	127
6	Investigating Vaccine Efficacy Against SARS-CoV-2 Variants: A Model-Based Analysis	130
6.1	Publication Notes	130
6.2	Abstract	130
6.3	Introduction	131
6.4	Results	135
6.4.1	Case 1: Small Antigenic Distance, High Degree of Overlap between Balls of Stimulation	137
6.4.2	Case 2: Moderate Antigenic Distance, Partial Overlap between Balls of Stimulation	140
6.4.3	Case 3: Larger Antigenic Distance, No Overlap between Balls of Stimulation	142
6.5	Methods	144

6.5.1	Computational Model	144
6.6	Discussion	153
	Bibliography	157

List of Figures

1.1	Simplified Schematic of T Cell Activation by Dendritic Cells .	3
1.2	Simplified Schematic of B Cell Activation and Function in Germinal Center	4
2.1	Comparison of simulated and observed lymphatic networks. (a) Mouse lymphatic network graph based on anatomical data with 36 nodes and 49 edges. edges. (b) Example simulated graph of the mouse lymphatic system. Algorithm parameters: $N_v = 36, N_{inp} = 13, N_l = 5, P_e = 0.851, P_o = 0.66.$	20

2.2 **Average Number of Hops to Find a LN with Cognate Antigen** after running the random walk on 500 observed and algorithm-based graph of mice and human. The **random bars** represent that $|V'|$ are randomly distributed over the graphs. There are 275 lymph nodes (LNs) containing the cognate T cell out of 996 lymph nodes (LNs) in human and 5 lymph nodes (LNs) out of 36 lymph nodes (LNs) in mice. The **cluster bars** represent $|V'|$ are distributed in clusters over the graphs. There are 275 lymph nodes (LNs) containing the cognate T cell out of 996 lymph nodes (LNs) in human and 5 lymph nodes (LNs) out of 36 lymph nodes (LNs) in mice. The **single bars** represent that there is only one lymph node (LN) ($|V'| = 1$) chosen at random carrying the cognate T cell out of 996 lymph nodes (LNs) in human and 36 lymph nodes (LNs) in mice. . . 25

3.1 **Lymphoid Organ Scaling with Mass.** Each data point represents a species shown on the log-log axis. The dashed lines show the regression fit. **(A) Spleen Volume** of 38 species is best fit by the regression cM^μ with $c = 1.2$ and exponent $\hat{\mu} = 1.1$ (95% CI [0.95, 1.2]). **(B) Lymph Node Volume** is best fit by $c_1M^\mu \ln(c_2M)$ with $c_1 = 1$, $c_2 = 1$, and $\hat{\mu} = 0.62$ (95% CI [0.55, 0.69]) for 16 species. **(C) Number of Lymph Nodes** for 10 species is best fit by cM^μ with $c = 4$ and $\hat{\mu} = 0.52$ (95% CI[0.40, 0.64]). The p-value of the exponents is significant at the 0.01 level. 36

3.2 **Theoretical Predictions for Search Time vs Animal Mass are Consistent with Simulations of T Cell Contact Times in Lymph Nodes. Panel A: For systemic infection** ($N_{DC} \propto M^{\frac{1}{2}}$), the predicted initial first-contact time (τ_{init}) scales as $\tau_{\text{init}} \propto M^{\mu} \ln(cM)$ with $\mu = -0.5$ (dashed red line). The fit (solid red line) to simulated data (red circles) gives $\hat{\mu} = -0.53$ (95% CI[-0.61, -0.45]). **Panel B. For small localized infections** ($N_{DC} \propto M^0$), the theory predicts the initial first contact time as $\tau_{\text{init}} \propto M^{\mu} \ln(cM)$ with $\mu = 0$ (dashed green line). The best fit (solid green line) to simulated data (green points) gives $\hat{\mu} = 0.0$ (95% CI [-0.01, 0.01]). 100 simulations were run for each LN volume that was estimated (from Eqn 1) for each value of M. The simulations for these two cases are consistent with theoretical predictions. 61

3.3	Theoretical Predictions for Mean Search Time vs Animal Mass are Consistent with Simulations of T Cell Contact Times in Lymph Nodes. Panel A: For systemic infection ($N_{DC} \propto M^{\frac{1}{2}}$), the predicted mean first-contact time is $\bar{\tau} \propto M^{\mu} \ln(cM)$ with, $\mu = 0$ (dashed blue line). The best fit (solid blue line) for the corresponding simulated data (blue points) gives the exponent $\hat{\mu} = 0$ (95% CI [-0.01, 0.01]). Panel B. For small localized infections ($N_{DC} \propto M^0$), the predicted mean first-contact time, is $\bar{\tau} \propto M^{\mu} \ln(cM)$ with $\mu = 0.5$ (dashed purple line). The best fit (solid purple line) to the simulated data (purple points) gives the value for the exponent, $\hat{\mu} = 0.50$ (95% CI [0.49, 0.51]). The simulations for all these two cases are consistent with theoretical predictions.	75
3.4	T Cell Population Growth From the IFCT Model Compared to Empirical Data and Median Model. Panel A depicts the T cell population over time in response to epitope NP118 for 5.8 days post-infection and Panel B depicts the same for epitope GP283 for 5.6 days post-infection. Blue data points are empirical data from [36] representing T cell populations at specified time points post-infection. The red data point is the Median model prediction of the peak T cell population. The green line shows the estimates from the IFCT model parameterized by data in [36].	76

3.5	T Cell Population Comparison Between Simplified Models for Smaller and Larger Lymph Nodes. The figure compares the log-transformed daywise T cell populations predicted by the Median model and IFCT model upto 8 days of exponential growth. The red dashed line represents the Median prediction for a smaller LN (60 initial T cells), and the red solid line corresponds to the Median prediction for a larger lymph node (1733 initial T cells). The green dashed line represents the IFCT model for a smaller LN (60 initial T cells), and the green solid line corresponds to the IFCT prediction for a larger lymph node (1733 initial T cells).	77
4.1	Graphical Representation of Simulation Cases.	88
4.2	The Three Simulations of Brownian Searchers in a Bounded Volume.	92

4.3	Distribution of Contact Times of All Searchers within One Volume for Case 0.	
	We show the distribution of the first contact times in two (out of 30) simulations of case 0. Panel A shows the distribution of 629 first contacts for 629 searchers with 2533 targets in a 921 mm ³ search volume. Panel B shows the distribution of 1989 first contact times for 1989 searchers with 8011 targets in a 3640 mm ³ search volume. Data are plotted as a histogram of probabilities of the initial first contact time of each searcher on the y-axis for the contact time given on the x-axis. The theoretical probability density functions (PDF) of a Gaussian distribution and an exponential distribution for the same mean and standard deviation as the data are shown for comparison. The data are visually consistent with an exponential distribution. The result of the goodness of fit tests (SSE and AIC) in Table 4.2 are consistent with our assumption of the exponential distribution of the first contact times.	101

4.4 t_{fin} Predictions Compared to Simulations Corresponding to Case 0: Constant Density, Multiple Targets.

Each red circle represents the time for the first searcher, in a population of N searchers, to contact its target, where N is varied along the x-axis. The initial first contact time for each replicated simulation is shown as red dots. t_{fin} is predicted to scale as N^{-1} (dashed green line, Equation (4.5)). The dotted purple line shows the regression through the simulated data: $E[t_{\text{fin}}] \propto N^{-1.05}$ with a 95% CI $[-1.1, -0.98]$ for the exponent, consistent with the exponent of -1 predicted by t_{fin} . Variance among the initial first contact times for a given N ranges from 0.25 to 0.36. In contrast, the mean search times (blue dot), averaged over all N searchers in each simulation, across the same values of N , are much larger and do not vary with N , and have extremely low variance, all less than 0.003. 100 replicates are used to calculate the mean first contact time and the initial first contact time for each value of N ; thus there are 100 blue and 100 red points for each value of N 102

4.5	t_{fin} and t_{∞} Predictions Compared to Simulations Corresponding to Case 1: Single Target, Constant Volume, Increasing N. Each red circle represents the time for the first searcher, in a population of N searchers, to contact its target, where N is varied along the x-axis. 30 simulations were run for each value of N , and the first contact in each simulation is shown. The pink dotted line connects the medians of the simulated data. The green line shows the t_{fin} prediction (Equation (4.7)) and the blue line shows the t_{∞} prediction (Equation (4.13)), both fitted to the simulated data. The regression through the simulated first contact points $N < 10^3$ is $N^{-1.1}$ with 95% CI $[-0.86, -1.3]$ consistent with the t_{fin} prediction for small N . The regression through data (pink line) where $N \geq 10^3$ gives $t_{\infty} \propto \log(N)^{-1.94}$ with a 95% CI $[-2.1, -1.8]$ which is slightly steeper than the $1/\log(N)$ prediction of t_{∞} . Prediction lines are shown with scaling constants set to fit the median of either the smallest (for t_{fin}) or largest (for t_{∞}) simulated data.	105
-----	--	-----

4.6	t_{fin} and t_{∞} Predictions Compared to Simulations Corresponding to Case 2: Single Target, Increasing Volume, Constant N. Symbols are replicated from Figure 5. The green line shows the $t_{\text{fin}} \propto L^3$ prediction and the blue line shows the $t_{\infty} \propto L^2$ prediction. The regression through the simulated data is $t \propto L^{2.17}$ with 95% CI [2.14, 2.2], close to the t_{∞} prediction.	106
4.7	t_{fin} and t_{∞} Predictions Compared to Simulations Corresponding to Case 3: Single Target, Increasing Volume, Increasing N. Symbols are replicated from Figure 5. The green line shows the $t_{\text{fin}} \propto 1$ prediction and the blue line shows the $t_{\infty} \propto L^2/\ln(N)$ prediction.	107

5.1	EFPT vs number of searchers within a small lymph node-sized search space.	
	The searchers mimic random T cell movement, recreated from chapter 4, Figure 4.5. Each red circle represents the time for the first searcher, in a population of N searchers, to contact its target, where N is varied along the x-axis. 30 simulations were run for each value of N, and the first contact in each simulation is shown. The pink dotted line connects the medians of the simulated data. The green line shows the t_{fin} prediction ($\propto \frac{1}{N}$) and the blue line shows the t_{∞} prediction ($\propto \frac{1}{\log(N)}$), both fitted to the simulated data. Prediction lines are shown with scaling constants set to fit the median of either the smallest (for t_{fin}) or largest (for t_{∞}) simulated data.	117

- 5.2 **EFPT over the number of searchers N for varying initial distance L between searchers and target.** Both axes are in logarithmic scale. (a) shows the analytical result of EFPT for varying numbers of searchers N and distance L . For any given N , EFPT increases with L . For each fixed L , EFPT is highest when N is very small. EFPT decreases sharply as N increases and almost saturates for large N . (b) shows the fitted curve of the two hypotheses (shadowed: $\frac{1}{N^\alpha}$ and dotted: $\frac{1}{\log N}$) and their transition points $(\Gamma_N, \Gamma_{EFPT})$. (c) shows the predicted minimum EFPT when the searchers follow the Brownian motion with 1 unit step at each time step. . . . 119
- 5.3 **Transition points of the number of searchers Γ_N and the Γ_{EFPT} over initial distance L .** Both axes are displayed on a logarithmic scale. The blue points represent the numerical data, and the red line is a best-fit curve. Panel (a) shows how Γ_N scales with L . The linear best-fit curve suggests that Γ_N **increases** with L . That is Γ_N increases about 3-fold when L doubles. Panel (b) shows how Γ_{EFPT} scales with L . The best-fit curve indicates that the transition in Γ_{EFPT} grows with L ; that is, when L doubles, Γ_{EFPT} also doubles. 121

5.4	Simulation and analytical EFPT for initial distance $L = 100$ unit over range of number of searchers N.	
	Both axes are displayed on a logarithmic scale. The maroon circles connected by the thin line show the analytic EFPT prediction. The red line shows the median of the 15 Monte Carlo simulations for hard speed cap $\Delta s = 1$ unit for each N and the green line shows 15 Monte Carlo simulations for diffusion coefficient $D = 1$ for each N . The red and green circles represent the simulation data illustrating the run-to-run variability. The dotted red line shows the theoretical expected minimum EFPT for hard cap step size Δs	123
6.1	Antigenic distance and balls of stimulation between different SARS-CoV-2 variants in a shape space diagram. (a) represents the smaller antigenic distance between 614G and Alpha, resulting in a high degree of overlap between their balls of stimulation. (b) represents a larger antigenic distance compared to panel (a) between 614G and Delta AY.4.2 variants with their slight overlapping ball of stimulation. (c) represents a larger antigenic distance between 614G and Omicron B.A.1 variants with no overlapping of their balls of stimulation. Although the figures show the antigens together, in the model the secondary antigen is introduced at day 40	136

6.2	Population dynamics for Case 1 where antigenic distance between primary antigen 614G, and secondary antigen Alpha is 0.81 Antigenic Unit (AU). This figure shows the simulation time courses following secondary challenge with the Alpha variant at day 40. (a) shows antigen concentration over time, with the secondary antigen (Alpha) introduced at day 40. (b) shows B-cell populations, showing naive, stimulated, plasma, and memory subsets. (c) shows viral load , antibody titer, and immune-complex (IC) formation following 614G and Alpha exposure.	138
6.3	Population dynamics for Case 2 where antigenic distance between primary antigen 614G, and secondary antigen Delta AY.4.2 is 1.73 Antigenic Unit (AU). This figure shows the simulation time courses following secondary challenge with the Delta AY.4.2 variant at day 40. (a) shows antigen concentration over time, with the secondary antigen (Delta AY.4.2) introduced at day 40. (b) shows B-cell populations, showing naive, stimulated, plasma, and memory subsets. (c) shows viral load , antibody titer, and immune-complex (IC) formation following 614G and Delta AY.4.2 exposure.	140
6.4	Population dynamics of 614G and Omicron B.A.2.	143

6.5	Population dynamics for Case 3 where antigenic distance between primary antigen 614G, and secondary antigen Omicron B.A.2. is 4.58 Antigenic Unit (AU).	
	This figure shows the simulation time courses following secondary challenge with the Omicron B.A.2. variant at day 40. (a) shows antigen concentration over time, with the secondary antigen (Omicron B.A.2.) introduced at day 40. (b) shows B-cell populations, showing naive, stimulated, plasma, and memory subsets. (c) shows viral load , antibody titer, and immune-complex (IC) formation following 614G and Omicron B.A.2. exposure.	143
6.6	Our simulation starts with initializing B cells and antigens in the shape space. Each point in the shape space represents a phenotype and each phenotype has multiple B cells. The phenotype groups containing multiple B cells are distributed uniformly in the shape space and antigens with a specific phenotype are plotted on the specific coordinate. The distance between the antigens and the B cells represents their antigenic distance. The model proceeds through submodels for stimulation, binding, mutation, differentiation, antibody secretion, antigen dynamics, and immune complex formation. The model keeps running until it reaches the specified time-step.	146

List of Tables

2.1	Summary statistics for observed and simulated graphs of mice and humans characterizing their topological properties. For the predicted graphs, we present the statistics obtained over 10,000 graphs for human and 500 for mice.	23
2.2	Predicted times for T cell to enter lymph node (LN) containing cognate dendritic cell (DC) based on hop count. <i>H. sapiens</i> and <i>M. musculus</i> lymph node (LN) residence times are taken to be 19 h. Time given in days (d).	27
3.1	Time to Initial Detection of Activated T Cells in Mice and Humans. Data are rounded to the nearest day (d). Means are calculated from the midpoints of each reported range of times observed in experiments. Remarkably, the mean, minimum, and maximum times are very similar in mice and humans.	35

3.2	DC-T Cell Initial First Contact Times considering 2 cases: i) assuming the number of dendritic cell (DC) in lymph node (LN) scales with $M^{0.5}$ (constant density of antigen-bearing DC from a systemic infection) and ii) a constant number of dendritic cell (DC) in lymph node (LN) ($N_{DC} = \text{Constant}$). We estimate times for each case, considering that T cells move in using either Brownian motion or a persistent random walk.	41
3.3	Published Spleen Volume, Lymph Node Volume and Lymph Node Numbers. Mean values and citations for body mass, spleen volume, lymph node (LN) volume, and LN number used in this paper. The number of lymph nodes (LNs) is the maximum reported value assuming that omissions are more likely than additions.	48
3.4	Notation Variables used in this work.	52
3.5	Experimental factor levels used in the ABM experiments. . . .	58
3.6	DC-T Cell Mean First Contact Times considering 2 cases: i) assuming the number of dendritic cell (DC) in lymph node (LN) scales with $M^{0.5}$ (constant density of antigen-bearing DC from a systemic infection) and ii) a constant number of dendritic cell (DC) in lymph node (LN) ($N_{DC} = \text{Constant}$). We estimate times for each case, considering that T cells move in using either Brownian motion or a persistent random walk.	63

3.7 Parameters and Peak Population estimates from the Median and IFCT models.	
The recruitment time and proliferation time are from [36]. The travel and activation time ($t_{travel+act}$) represents the corresponding time after subtracting the median first contact time (τ_{init}^{median}) from the recruitment time (t_{rec}). Peak T cell populations are after 5.8 days and 5.6 days post-infection for epitope NP118 and GP283 respectively. The median first contact time in the Median model is 22 hr for all T cells and the first contact time is different for every T cell in the IFCT model.	68
3.8 Comparison of T Cell Population Growth Predictions for Smaller and Larger Lymph Nodes Using Simplified Median and IFCT Models at Day 4 and Day 8.	
The table illustrates the median contact time and T cell population estimates after 4 and 8 days for a smaller LN (60 initial T cells) and a larger LN (1733 initial T cells) based on the two models. The Median model assumes simultaneous contact for all T cells, while the IFCT model allows for individual first contact times. X-fold growth represents the ratio between IFCT and Median model predictions for each scenario.	71

4.1	Comparison of t_∞ and t_{fin}. We assumed an exponential distribution of search times and a finite number of searchers in our model of t_{fin} . In contrast, t_∞ assumes an infinite number of searchers and that the fastest searcher will follow the shortest (and therefore, most direct) path to the target.	89
4.2	Statistical comparisons of the Gaussian and exponential distribution fit to the initial first contact times for case 0. The exponential distribution has lower mean SSE and AIC values than the Gaussian distribution, indicating a better fit to the data. 100 simulations were replicated for each number of searchers. Numbers are given to two significant digits.	103
6.1	Parameters and Notations	154
6.2	Mean age of B cells, and antibodies. The age is randomly assigned following an exponential distribution with mean age .	155

Chapter 1

Introduction

The immune system is a complex system of organs, tissues, cells, and molecules that defends the body against infections, cancer, and other threats. As the body's primary defense mechanism, it is designed to recognize and neutralize harmful agents, known as pathogens. Our first line of immune defense consists of the physical barrier (skin, mucous membrane), which prevents pathogens from entering the body [150]. If this barrier is breached, the innate immune response, the second line of defense, is activated. This non-specific defense mechanism involves immune cells such as macrophages that rapidly respond to a wide range of invaders through processes like phagocytosis. The adaptive immune response, the third line of immune defense, is a specific response that develops over time. Unlike the innate immune system, it's slow to respond initially. But after the first encounter it remembers the pathogen, ensuring future responses are much faster and effective. It is triggered by the

presence of antigens (fragments of pathogens) and involves the activation of T cells and B cells. The immune response is characterized by the interactions, relationships, associations, and interactions among these components, which define the behavior and functionality of the system.

A key event that initiates the adaptive immune response is the encounter between dendritic cells (DCs) and naive T cells. When a pathogen enters our body, DC collects and carries the fragments of the pathogen, known as antigens, to the closest lymph node. Inside the lymph node, naive T cells search for the antigens. Upon encountering a dendritic cell presenting a cognate antigen, naive T cells activate, proliferate, and travel through lymphatic vessels to the infected tissues to destroy the specific pathogen (Figure 1.1).

While dendritic cells carry processed (fragmented) antigens for T cells recognition, pathogens or their intact antigens in tissue fluid enter lymphatic capillaries and carried through lymphatic vessels to the nearest lymph node. B cells can directly bind to intact antigen via their B Cell Receptors (BCRs) and the help from activated helper T cells. B cells then enters into the germinal centers (GCs), [151]. where they proliferates, differentiate into plasma and memory B cells, and undergo somatic hypermutation Figure 1.2. Plasma cells secrete antibodies to fight against the pathogen; memory B cell remembers the antigen, protecting against reinfection; and somatic hypermutation mutates the B cell gene to generate diversity to bind the antigen with high affinity.

A fundamental aim of my research is to investigate the dynamics of the

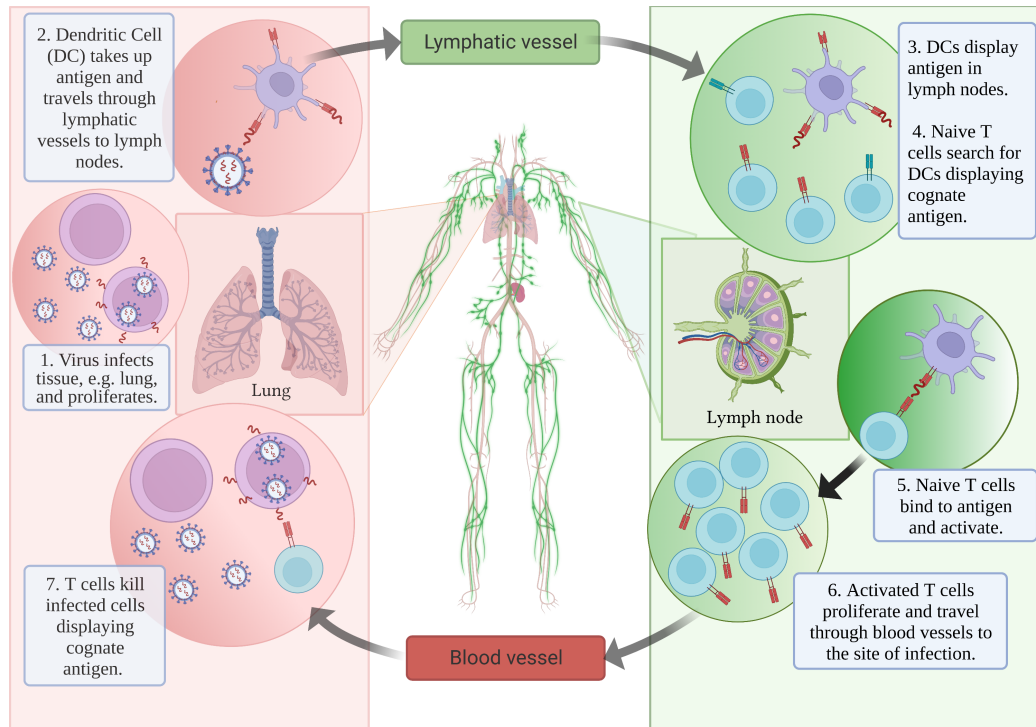
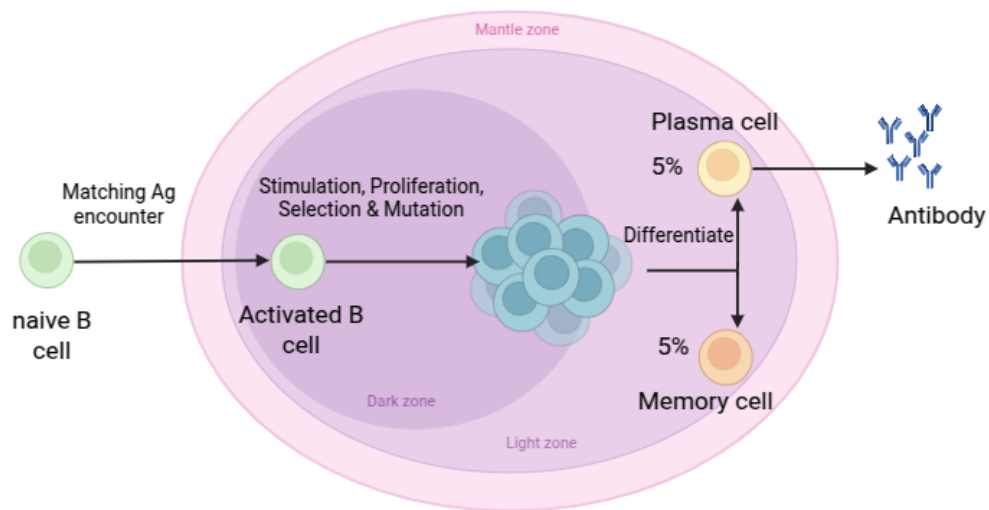


Figure 1.1: Simplified Schematic of T Cell Activation by Dendritic Cells. 1) A pathogen infects tissue, in this case for illustration, the lung. 2) DCs carry antigen from tissue through the lymphatic vessels to the draining lymph nodes (LNs). 3) DCs show antigen in the lymph node. 4) Naïve T cells search for cognate antigen presented on the surface of DCs. 5) T cells activate upon encountering DC presenting cognate antigen. 6) Activated T cells replicate exponentially, and some become killer T cells that travel through the bloodstream to the sites of inflamed, infected tissue. 7) T cells kill the infected cells that display cognate antigen. We model the timing of search and activation in steps 4 and 5, where the adaptive immune response is initiated; the timing of this process depends on LN size.



Created in BioRender.com bio

Figure 1.2: Simplified Schematic of B Cell Activation and Function in Germinal Center This figure shows the different stages of B cells after its encounter with an antigen. Once it binds with the antigens,

adaptive immune response during infection in order to gain a deeper understanding of its components' behavior and functionality within the immune response system. Understanding the complex interaction of the components in the adaptive immune response is important to effectively model the immune system and to make breakthroughs in drug discovery and disease treatment. A key aspect of this interaction is the initial first contact time, the time it takes for a cognate immune cell to encounter its target antigen, which fundamentally governs the timing and efficacy of the immune response. The concept of first contact time, popularly known as extreme first passage time (EFPT), is not only unique to immunology but also appears across diverse disciplines such as chemical kinetics, physics, and complex systems. For example, in the immune system, an immune response is initiated when the very first T cell encounters an antigen [30]; In chemical kinetics, for example, reaction rates are often limited by the time required for molecules to diffuse and interact [94]. Similarly, in ecological and search theory contexts, first-passage times dictate how efficiently searchers (e.g., predators, molecules, or information packets) locate their targets within large or disordered environments. [102]. These parallels suggest that biological systems may have evolved highly efficient mechanisms for locating targets in complex, noisy environments, offering inspiration for cross-disciplinary modeling approaches.

A major motivation to study the immune system is to develop effective vaccines, especially against rapidly mutating viruses. Viruses, especially those prone to rapid mutation, present significant challenges for comprehensive

modeling and long-term vaccine development. A prime example is SARS-CoV-2, which has garnered global attention since its emergence in late 2019. According to the CDC [55], approximately 1.2 million deaths have been reported in the United States due to this virus. Despite relentless and thorough efforts by researchers to understand and combat SARS-CoV-2, accurately modeling the behavior and impact of such viruses remains complex. This complexity is largely due to their high mutation rates, which often lead to the emergence of new variants that can evade existing immune responses. Consequently, predicting and modeling the effectiveness of vaccines across all variants of a rapidly mutating virus like SARS-CoV-2 has proven to be a difficult task. Therefore, understanding how the immune system initiates and scales its response is key to developing strategies that offer broader and more durable protection.

This dissertation addresses key questions regarding the dynamics, scaling, and efficiency of the adaptive immune response. The main contributions of this research are as follows:

- We investigate how the topology of the lymphatic network affects the time a T cell takes to reach the lymph node carrying antigen-bearing DCs. This investigation leads us to understand how the scaling of lymph node distribution can affect the expected time of finding the desired lymph node by T cell.
- We investigate how the lymph node number and volume scale with

body mass. This study is important to understand how the underlying population of T cells in a lymph node scales with body mass. Following the scaling of the volume with body mass, we study the time it takes for the adaptive immune response to occur in mammals while examining the interaction of cognate T cells and dendritic cells in the lymph node for systemic and local infection. This study of the first time interaction between T cells and antigen-bearing DC is extremely important as this one single event triggers the immune response through replication and mutation. By finding the scaling relationship between body mass and the contact times between T cell and DC, one can estimate the time to initiate the immune response in any-sized animal, and this information is particularly helpful in designing a vaccine covering a large section of animals.

- Motivated by the study of the scaling of immune response time, we also explore the first interaction (EFPT) of searchers and targets in a confined space. This study provides a deeper understanding of target-searcher interactions in an immunological context. This research could potentially provide insight into the mechanisms of the adaptive immune response and how target-searcher interactions can influence the efficiency of this response in a confined space.
- We study EFPT in large-scale, confined 3D environments. We model how EFPT between searchers and targets changes as the number of

searchers increases. We explore the scaling behavior of EFPT and investigate two key hypotheses: EFPT scales inversely with the number of searchers N ($\text{EFPT} \propto \frac{1}{N}$) and EFPT scales inversely with the logarithm of N ($\text{EFPT} \propto \frac{1}{\log(N)}$). This investigation is critical not only in the immunological context but also in broader fields such as chemical kinetics, robotics, and information theory, where understanding efficient search in high-dimensional spaces is a foundational challenge.

- We study the population dynamics of B cells, antibodies, and antigens for repeated exposure of different antigen variants. Understanding this population dynamics is crucial for developing effective vaccination strategies and understanding the immune response to viral infections. We developed a model to investigate the varying efficacy of vaccines against SARS-CoV-2 variants. The model can be used to design vaccines through antigenic combinations that promote long-term immune protection by targeting a large number of SARS-CoV-2 variants.

1.1 Modeling the Lymphatic System

Our first study investigates how the topology of the lymphatic network affects the time required for T cells to search for the matching antigen presented by dendritic cells in lymph nodes. We have extended a previous method that mapped the human lymphatic network and inferred the topology of the lymphatic network in mice. By comparing the modeled and observed

topologies, we demonstrate that they are similar to each other and consistent with observed immune response times. This finding is crucial for translating immune response times in mice, where most experimental work is conducted, into anticipated immune response times in humans. Our analysis predicts that the lymphatic network’s topology facilitates fast immune response times for large systemic infections, despite the animal’s significant increase in mass. However, T cells may require more time to locate small, localized infections in larger animals.

Our analysis shows that the physical structure of the lymphatic network facilitates scale-invariant immune response. For large and systemic infections that require a large and fast response, T cells navigate the lymphatic network to find infected LN equally fast in large and small animals.

1.2 Scaling the Lymphatic System and Immune Response

Given the mounting threat of emerging zoonotic diseases to humanity, it is crucial to comprehend how the adaptive immune response varies across species. While laboratory species such as mice are commonly used as models to extrapolate immune function in humans, their physiology differs significantly from that of larger mammals. In [49], we demonstrated that the volume (V_{LN}) and number (N_{LN}) of lymph nodes exhibit a unique scaling relationship with

body mass (M) as follows,

$$V_{LN} \propto M^{\frac{1}{2}} \ln(cM)$$

$$N_{LN} \propto M^{\frac{1}{2}}$$

Critically, we showed that the first step of initiating the adaptive immune response occurs faster in the lymph nodes of larger animals, including humans than in smaller animals like mice. This finding challenges the conventional assumption that larger body size necessarily leads to slower immune response times [85, 160, 12]. Our work highlights the importance of the initial first-contact timing in the adaptive immune response, which is influenced by the number of immune system "searchers" available. Specifically, our results suggest that the timing of the adaptive immune response is proportionally faster in larger animals with more searchers, while the mean response time remains independent of the number of searchers. This study provides important insights into the fundamental mechanisms underlying the adaptive immune response across species and emphasizes the need for further research to fully understand the implications of these findings for human health.

1.3 Initial First Contact Time Dependency

The initial discovery time in biological search processes where multiple searchers attempt to find one or more targets is critical in initiating down-

stream events. This is evident in a wide range of contexts, from foraging behavior in ant colonies [102, 40] to immune surveillance, where the first T cell to encounter an antigen initiates the immune response [30]. While most research in biological search studies the mean contact time [120], physicists have studied the extreme first passage time (EFPT) [89, 88], which is the time it takes for the fastest searcher to find their target. However, most analytical frameworks for EFPT are limited to asymptotic conditions where the number of searchers approaches infinity. These models are less effective in realistic biological scenarios involving finite and often small populations of searchers. To address this limitation, we proposed a mathematical model that works for a small to moderate number of searchers in confined reflective volumes—representative of bounded biological environments such as lymph nodes. Our work suggests that when the density of searchers is low, the initial first contact time (IFCT) decreases approximately linearly with the number of searchers, a trend we termed IFCT hypothesis. As the density increases, the scaling shifts to a logarithmic decrease, consistent with the universal formula of EFPT. To arrive at this result, we considered searchers undergoing Brownian motion within a 3D reflective domain, initially placed at a fixed distance from a central target. We also examined the effect of scaling the search volume—specifically, the volume of a lymph node—while keeping the number of searchers constant. These findings are not only significant in the context of immunology but can also be generalized to other domains such as chemical kinetics, ecology, and molecular search processes.

1.4 Extreme First Passage Time in Large-Scale Models

Most prior studies focus either on single searcher dynamics [112] or asymptotic behavior as the number of searchers approaches infinity [91]. In contrast, EFPT—which characterizes the behavior of the first successful searcher among many—remains less well understood across the full range of searcher numbers.

In our previous work [51], we addressed this gap by demonstrating a scaling transition of EFPT in a bounded 3D search domain with reflective boundaries in an immunological context. We showed that EFPT decreases approximately linearly with a limited to moderate number of searchers and transitions to decreasing logarithmically as the number of searchers increases to infinity. This result, situated in an immunological context, modeled T cells searching for a single antigen within a confined volume.

In this work, we extend and generalize our previous work in several key ways: We test the robustness of the observed EFPT scaling transition for large-scale model, by adopting a more abstract model, removing immunological constraints. The model allows to study generic searcher-target interactions with unit-step Brownian motion. We develop and validate an analytical framework based on Lawley’s formulation [88] to predict the critical transition point between the two scaling regimes.

This investigation not only enhances our understanding of search efficiency

in immune responses but also contributes a generalizable model of search dynamics applicable to a wide range of disciplines, including applied physics, chemical kinetics, and robotics. It bridges a gap in existing literature by providing a full-spectrum view of EFPT scaling across different regimes of searcher populations.

1.5 Examining Affinity Maturation and Antigenic Drift with an Agent-Based Model

The immune system’s ability to recognize and respond to a vast array of antigens relies on the remarkable diversity of B cell receptors (BCRs), each with unique antigen specificity. Once the BCRs and antigen epitopes (a specific part of the antigen that B cell recognizes) matches at a certain level, B cells bind with the antigen and plasma B cells produce antibodies targeting the antigen to form immune complexes. Affinity is the strength of this binding between the B cell/antibody and the antigen. Antigenic distance refers to the degree of similarity between antigens and B cells/antibodies computed using the affinity.

In this work, we developed an agent-based model (ABM) to study how the adaptive immune system responds to evolving viral threats, particularly focusing on B cell-mediated immunity. Our model captures the interactions between B cells (naïve, plasma, and memory), antibodies, and antigens, representing their receptors and epitopes in a Euclidean shape space. Binding

between these agents is determined by Hamming distance, allowing us to simulate affinity-based interactions and the formation of immune complexes.

This work is important because the immune system’s ability to recognize and neutralize pathogens depends on the affinity. With repeated exposure to the same antigen, B cells undergo affinity maturation to produce antibodies with higher binding strength. However, as viruses mutate (antigenic drift), new variants can emerge that are increasingly distant from the original strain in antigenic space. This can reduce the effectiveness of pre-existing immunity. This issue has been well documented in cases of influenza, and there is a concern that it is occurring in SARS-CoV-2 given successive variants of concern (VOC).

Our model explores how such antigenic drift impacts immune recognition and cross-reactivity, especially in the context of immune imprinting (prior exposure influencing future responses).

Utilizing our immune response model, we predict vaccine efficacy for individuals receiving a wide-range of SARS-CoV-2 VOC. By simulating SARS-CoV-2 variants of concern (VOCs), we aim to better understand how existing immunity might respond—or fail to respond—to new variants. This insight can inform vaccine design and public health strategies.

In the following sections, I will be discussing related works, completed works, and my proposed work toward my complete dissertation work.

Chapter 2

Modeling Immune Search Through the Lymphatic Network

2.1 Publication Notes

Citation: Ferdous, J., Fricke, G.M., Moses, M.E. (2022). Modeling Immune Search Through the Lymphatic Network. In: Dorigo, M., et al. Swarm Intelligence. ANTS 2022. Lecture Notes in Computer Science, vol 13491. Springer, Cham. https://doi.org/10.1007/978-3-031-20176-9_30

Published: 29 October 2022

Formatting: The original published text has been preserved as much as possible while still adhering to the formatting requirements of this dissertation.

Funding: This work is funded by National Science Foundation (NSF), awards 2030037 & 2020247

Acknowledgements: We thank the UNM Center for Advanced Research Computing, supported in part by the NSF, for high performance computing resources. We also thank Moses Biological Computation Lab for constructive discussion.

2.2 abstract

The lymphatic system is a networked structure used by billions of immune cells, including T cells and Dendritic cells, to locate and identify invading pathogens. Dendritic cells carry pieces of pathogens to the nearest lymph node, and T cells travel through the lymphatic vessels and search within lymph nodes to find them. Here, we investigate how the topology of the lymphatic network affects the time it takes for this search to be completed. Building on prior work that maps out the human lymphatic network, we develop and extend a method to infer the lymphatic network topology of mice. We compare search times for the modeled and observed topologies and show that they are similar to each other and consistent with observed immune response times. This is relevant for translating immune response times in mice, where most experimental work occurs, into expected immune response times in humans. Our analysis predicts that for large systemic infections,

the topology of the lymphatic network allows immune response times to remain fast even as animal mass increases by orders of magnitude. This work advances our understanding of how the structure of the lymphatic network supports the swarm intelligence of the immune system. It also elucidates general principles relating swarm size and organization to search speed.

2.3 Introduction

Adaptive immunity evolved in vertebrates to recognize and remember novel pathogens, enabling a faster response time to subsequent infections. In contrast to most biological rates, which are systematically slower in larger animals (scaling as $M^{1/4}$, where M is body mass [85, 160, 12]), the adaptive immune response time is relatively invariant across several orders of magnitude of mammalian body mass [23, 14]. Immune response is a swarm intelligence problem with billions of interacting agents searching for pathogens without central control, and it is a model for scale-invariant search in swarms.

T cells are adaptive immune cells that can recognize novel pathogens in lymph nodes, and then replicate and disperse into tissues to find and kill cells infected by those pathogens. The movement of T cells through the lymphatic system increases contact with antigens and amplifies the immune response [154]. Similar to eusocial insects, information transmission in this liquid brain [149] is mediated through direct agent contact and chemical signals among agents that navigate complex and varied environments [109].

Each T cell can bind to a particular subset of cognate antigens. DCs gather antigen from tissues, travel to and enter nearby lymph nodes (LNs) through the lymphatic network, and display the antigen on their surfaces. T cells search LNs for DCs displaying cognate antigen, and if a match is made, the T cells activate, proliferate, and circulate to the site of infection where they kill infected cells. The time it takes to initiate an adaptive immune response depends on two factors: 1) the speed with which T cells travel through the lymphatic system to LNs containing DCs displaying antigen, and 2) how quickly T cells find those DCs once inside the LN.

In this work, we analyze T cell travel time through the lymphatic network to find DC's in mice and humans by extending the algorithm of Savinkov et al. [138], that models only the human lymphatic networks. While most lab studies that show how the immune system works are conducted on mice, most of the literature on modeling the lymphatic network is based on humans. The lack of data makes it challenging to build a general model of lymphatic networks for mice and other mammals. The model parameters are updated based on best-fit values by comparing empirically observed anatomical data with the graph resulting from the algorithm. We expand the network metrics used by the algorithm to better fit the model to empirical data. Using the inferred network model we compute the expected time for T cells to find LNs containing DCs presenting cognate antigen. We run a random walk search on the simulated and observed lymphatic networks to find the average time T cells need to reach the LNs containing cognate DCs. We find that the

generated and actual anatomical graphs have similar statistics. The resulting search time over the network is similar in mice and humans for systematic or mass-dependent infections, but it is longer in humans than in mice for small infections that only reach a single LN.

2.4 Related Work

Several studies have modeled the human lymphatic system [127, 155, 138]. In [155], the authors use computational geometry to build graph models of the human lymphatic network in order to explain the general features underlying the 3D structural organization of the lymphatic system. The model is based on available anatomical data (from the PlasticBoy project [2]), which estimates the lymphatic system’s structure and analyzes the topological properties of the resulting models. In [138], the authors developed and implemented a computational algorithm to generate the algorithm-based random graph of the human lymphatic system. Some fundamental characteristics of the observed data-based graph [155] and the algorithm-based graph of human lymphatic system graph models are analyzed.

In [162] Wiegel and Perelson hypothesize that LN number and size evolved to minimize two competing goals: the time to transport antigen from an infected area to the nearest LN and the time for immune cells to find the antigen inside the LN. Banerjee and Moses [13] use an ODE model to estimate that, empirically, immune response times are independent of host body size.

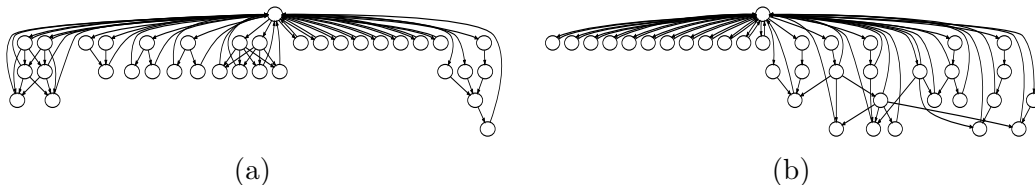


Figure 2.1: Comparison of simulated and observed lymphatic networks. **(a)** Mouse lymphatic network graph based on anatomical data with 36 nodes and 49 edges. **(b)** Example simulated graph of the mouse lymphatic system. Algorithm parameters: $N_v = 36$, $N_{inp} = 13$, $N_l = 5$, $P_e = 0.851$, $P_o = 0.66$.

2.5 Methods

2.5.1 Lymphatic Network Simulation Algorithm

Savinkov et al [138] developed an algorithm that generates a random directed human lymphatic network graph with no cycle from a reference human graph. We extend their work by adding another step to the algorithm to simulate T cells traveling through the circulatory system to enter LNs. The steps are given in Algorithm 1. We used data from [81] to create a reference graph of mice to compare with the simulated graph. Out of 5 input parameters in the algorithm, three parameters, number of nodes N_v , number of input nodes N_{inp} , and number of layers N_l are explicitly set to match the anatomy-based graph's properties. Based on the comparison metrics characterizing the topology of an anatomy-based graph (described in Section 2.5.2), the value of the other two parameters, probability of new edge creation P_e at each step and probability that the created edge connects nodes from different layers P_o ,

are set to produce graphs with similar topological structures.

2.5.2 Comparing simulated graphs to observation

We have used the following topological properties defined in [138] to compare the observed graph with the current state of the simulated graph for humans and mice: *The number of input nodes N_{inp} , Maximum degree of graph ΔG , Girth of the graph, g , The diameter of the graph, D , Radius of the graph, r , Average path length, I_G , The energy of the graph, E_n , The spectral radius of the graph, ρ , Edge density of the graph, ρ_d , The clustering coefficient, C (transitivity).* We also introduced the following graph properties to the list: *Number of separators, n_{sep} :* is the number of nodes removal of which disconnects the graph *n_{deg_i} :* is the number of nodes with degree i . *G_l :* is the average degree of nodes in each layer l and, *n_l :* is the number of nodes in each layer l .

Since number of node connections and layers are larger in larger animals, n_{deg_i} and n_l are also larger. Thus, the objective function has more parameters in larger animals. To produce a similar graph that matches these topological properties, we tune the parameters P_e and P_o . We collect these parameter values for the minimum value of the objective function, ω . For a number of properties, the objective function is defined as:

$$\omega = \sum_{i=1}^a \left(\frac{s_i(G) - s_i(G^*)}{s_i(G^*)} \right)^2 \quad (2.1)$$

where

$$s(G) = (n, m, n_{inp}, \Delta G, g, D, r, I_G, E_n, \rho, \rho_d, C, n_{sep}, n_{deg_1}, \dots, n_{deg_{max}}, G_1, \dots, G_l, n_1, \dots, n_l)^T$$

This objective function penalizes the topological discrepancies of graph G from the target graph G^* and weighs them with $(s_i(G^*))^{-2}$ to bring discrepancies of different components of vector s to a single scale.

2.5.3 Search Algorithm

To run the search algorithm, we randomly choose a source node n_s from which the T cell initiate a random walk through the graph. We consider that the LNs that contain matching DC, designated $V' \in V$, are distributed within the lymphatic network in three ways for different kinds of infections.

- *Random Systemic:* Systemic infections can spread to multiple lymph nodes throughout the body, i.e., in HIV. For this case, we assume that the V' are distributed randomly over the lymphatic network.
- *Clustered:* A cluster of LN can contain antigen if an animal gets a vaccine injection with inoculation dose adjusted to size, or if an animal breathes in a respiratory virus where the amount of inhaled virus is proportional to lung size. For such cases, we distribute the V' nodes in clusters. We randomly pick one node and run Breadth-First Search (BFS) to make the clusters. We exclude the circulation node 0 from being in the cluster.

Table 2.1: Summary statistics for observed and simulated graphs of mice and humans characterizing their topological properties. For the predicted graphs, we present the statistics obtained over 10,000 graphs for human and 500 for mice.

Parameter	Mice observed graph	Mice simulated graph	Human observed graph	Human simulated graph
$G(n, m)$	(36, 53)	(36, 49)	(996, 1117)	(996, 1029)
N_{inp}	13	13	357	357
Maximum degree, ΔG	24	26	8	16
Girth, g	3	3	3	4
Diameter, D	4	4	40	39.96
Radius, r	3	3	30	28
Average path length, l_G	1.34	1.42	12.79	15.3
Energy, E_n	37.17	36.40	1224.5	1190
Spectral radius, ρ	5.81	5.91	3.51	4.18
Edge density, ρ_d	0.04	0.04	0.001127	0.001038
Clustering coefficient, C	0.12	0.11	0.027	0.0004
Number of separators, n_{sep}	5	9	401	496

- *Single*: If an animal steps on a thorn and gets a local infection of a fixed size, or a mosquito bite transmits an illness into the blood, then the same small amount of infection is injected into the animal regardless of its size. For both of these cases, we randomly pick one node $|V'| = 1$ that contain cognate DC.

We compute the time it takes for each T cell using a random walk to reach the first LNs that contains DCs holding cognate antigen. We follow Perelson and Weigel's prediction that the number of LNs in mammals scale with $\propto M^{\frac{1}{2}}$ [120], for the random systemic and clustered scenarios, $|V'| \propto M^{\frac{1}{2}}$. For the uniform random and clustered V' , we assume the number of LNs that are bearing the cognate antigen-bearing DCs ($|V'|$ are 5 and 275 in mice and humans, respectively representing 7% and 3.6% of LN).

2.6 Results

2.6.1 Modeled Lymphatic Network

We run the extended algorithm to generate lymphatic networks for humans and mice. Figure 2.1a, and Figure 2.1b show the resulting observed and simulated graphs for mice. The first three parameters of the algorithm for mice are collected from [81]. For P_0 and P_e , we take their values that give the objective function's minimum value in Equation (2.1). They are compared numerically in Table 2.1 based on the topological properties, described in Section 2.5.2.

From Table 2.1 we can see that the properties are very similar for observed and simulated graphs for mice and humans. Some properties vary slightly, but the statistic from the objective function gives the overall best match of the simulated graph to the observed graph. We collect the time data the DC takes in humans and mice respectively to reach the LN containing cognate T cell from the infected area after running the random walk, shown in Figure 2.2. The time for T cells to encounter a target LN is shorter in humans than in mice for random and clustered target LNs. That is because there are more target LN in humans, and we consider only the time to find the first target LN. The search to find a single V' , takes much longer in humans because there are many more LN in humans (996) compared to mice (36).

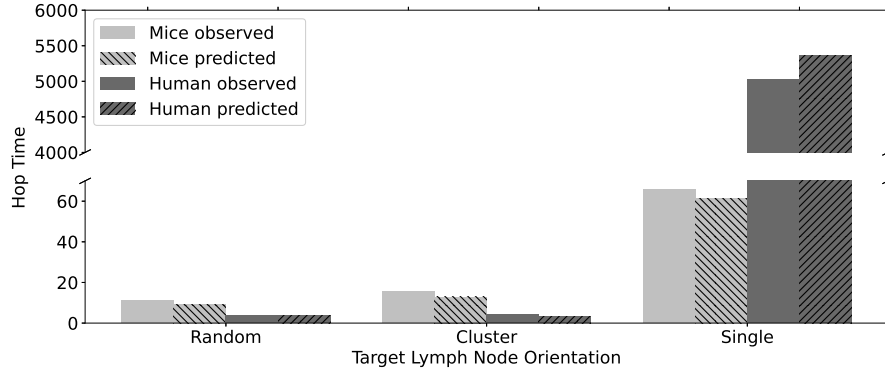


Figure 2.2: **Average Number of Hops to Find a LN with Cognate Antigen** after running the random walk on 500 observed and algorithm-based graph of mice and human. The **random bars** represent that $|V'|$ are randomly distributed over the graphs. There are 275 LNs containing the cognate T cell out of 996 LNs in human and 5 LNs out of 36 LNs in mice. The **cluster bars** represent $|V'|$ are distributed in clusters over the graphs. There are 275 LNs containing the cognate T cell out of 996 LNs in human and 5 LNs out of 36 LNs in mice. The **single bars** represent that there is only one LN ($|V'| = 1$) chosen at random carrying the cognate T cell out of 996 LNs in human and 36 LNs in mice.

2.6.2 Predicted Time

We compare the search time of a single T cell to find a target LN to actual immune response times to determine if our model predictions are reasonable. We calculate times from hop counts and estimates of the time between hops, shown in Table 2.2. Since we only model a portion of the overall adaptive immune response, that is, the time taken for a single T cell to conduct a random walk through the lymphatic network to find an infection, we cannot predict the speed of the overall immune response. For mice LN mean residence

time in LN per hop is approximately 13 h [154], and for sheep 19 h [101]. Since sheep and humans masses are similar (40 kg–160 kg for sheep [22] and 43 kg–140 kg for humans) [158], we approximate residence times in humans with those of sheep. Multiplying these residence times by the hop counts from Figure 2.2 results in Table 2.2. We find that the predicted time for a single T cell to find a LN with cognate antigen is on the same order as observed immune response times for systemic infections in mice and humans. According to [116, 106, 82, 153, 35] the mean adaptive immune response time in mice for influenza and LCMV infection is 5.3 days and in humans for SARS-CoV2 its 5.1 days [92, 74, 86]. This means that for systemic or whole-organ infections (where the number of LN increases with body mass), typical T cells can find the a LN with antigen during the time available to proliferate and amplify the growing immune response. In contrast, the time to find a single LN with antigen is orders of magnitude longer. This suggests that not many T cells would reach the single LN during the time of adaptive immune amplification. However, in small infections, a global response is likely not to be needed. We expect the T cells that reside in the local LN to be sufficient to respond to small local infection [13]. The actual timing depends on many factors, including the fraction of LN containing target DCs, V' and the number of cognate T cells searching for those DCs. We do not consider lymph vessel or blood residency times in these estimates, because those times are small relative to the time within LN [154].

Table 2.2: Predicted times for T cell to enter LN containing cognate DC based on hop count. *H. sapiens* and *M. musculus* LN residence times are taken to be 19 h. Time given in days (d).

	<i>M. musculus</i>				<i>H. sapiens</i>			
	Observed		Simulated		Observed		Simulated	
	Random	Clustered	Random	Clustered	Random	Clustered	Random	Clustered
Hops	11	15	9.3	13	3.8	4.1	3.9	3.4
Time (days)	5.9 d	8.3 d	5 d	7 d	3 d	3.3 d	3 d	2.7 d

2.7 Discussion

We simulated the lymphatic network for mice, ran a random walk process on the resulting graph, and predicted the time for a typical T cell, searching that graph for a LN with cognate antigen. We examined three scenarios corresponding to different infection patterns: random systemic infection, clustered infection, and infection in a single LN. Our results show that the time for each T cell to search for clustered and randomly distributed systemic infections in lymph nodes is on the same order as observed immune response times to systemic infections such as influenza and COVID-19 in humans and mice. In contrast, the time for a T cell to find a single LN is far longer, requiring thousands of network hops that would take years of search time in humans or a month in a mouse. However, we suggest that such long search times for small localized infections may be adaptive. For systemic infections that require a large response, T cells quickly discover LN with DCs presenting antigen, but T cells are not recruited to small local infections when they are not needed – local infections are responded to only by the small number of T

cells that already reside in the lymph node where the infection is presented on DC. This analysis shows that the physical structure of the lymphatic network facilitates scale-invariant immune response. For large and systemic infections that require a large and fast response, T cells navigate the lymphatic network to find infected LN equally fast in large and small animals. In one sense, the adaptive immune system exemplifies the kind of decentralized control typical in swarm intelligence: immune response is fast and adaptable based on the independent action of billions of immune cells that communicate locally and navigate complex tissue environments. However, the decentralized search is constrained by the network structure of the lymphatic system that provides a form of global guidance in physical space. That structure contributes to the extraordinary scalability of response.

Acknowledgements

We thank the UNM Center for Advanced Research Computing, supported in part by the NSF, for high performance computing resources, and NSF awards 2030037 & 2020247 for funding.

Chapter 3

Bigger Is Faster: The Scalable Adaptive Immune Response

3.1 Publication Notes

Authors: Jannatul Ferdous, Matthew G. Fricke, Judy L. Cannon, Melanie E. Moses **Submitted:** This paper is being submitted at Nature Scientific Reports for peer review.

Formatting: The original submitted text has been preserved as much as possible while still adhering to the formatting requirements of this dissertation.

Funding: This work is funded by National Science Foundation (NSF), awards 2030037 & 2020247

Acknowledgements: We thank the UNM Center for Advanced Research

Computing, supported in part by the NSF, for high performance computing resources. Thanks to Chris Kempes, Sid Redner, Alan Friedman, and the Moses Computational Biology Lab for the helpful discussions and reviews of earlier versions of this manuscript.

Data and materials availability: All data for spleen volume, LN volume, and LN number used in this paper are collected from the published literature and included in Table 3.3. Raw data files for initial and mean first contact time generated from our model are available in an online dataset (https://datadryad.org/stash/share/IkjXOE0jiqZ_rc4vdGIeqTYp9e08eCJkz654SyeFQ0A) All figures, except for Figure 1, are generated using Python 3 in a Jupyter Notebook and Adobe Illustrator. Figure 1 was generated using BioRENDER. The code for our agent-based model, mathematical analysis, and figure generation are available at <https://github.com/BCLab-UNM/BiggerIsFaster>.

3.2 Abstract

Understanding the speed of the adaptive immune response across mammals is important as humanity faces increasing threats from novel zoonotic pathogens. We show that the initial adaptive immune response time is equivalent in mice and humans. We explain this deviation from the expectation that larger animals have slower physiology due to an unusual relationship between the sizes and numbers of lymph nodes and animal size. We show that

pathogens are recognized more quickly in larger lymph nodes, allowing novel antigens to be discovered faster in larger lymph nodes, compensating for other physiological times slowing in larger animals. More generally, our analysis of initial first contact times introduced in this paper defines regimes under which any population of searchers gains speed proportional to population size, benefiting large population size for many search problems in biology.

3.3 Introduction

In mammals, most biological processes slow down as an animal’s size increases following a quarter-power scaling law [85, 160, 12]. While the cause of quarter-power scaling is debated [24, 95, 107], empirical observations consistently show that smaller mammals have faster physiology and life history, and larger mammals have slower rates over longer times [28, 161, 121, 137]. For example, humans who are 2500 times larger than mice, are predicted to have heart rates, breathing rates, and gestation times that are seven times slower than mice; observed values are within a factor of two of prediction: seven to 14 times slower [28, 31, 104]. However, the time to initiate a detectable adaptive immune response, which might theoretically be expected to slow by this same amount, remains surprisingly fast and consistent across species (Table 3.1), despite the challenges posed by longer distances for immune cells to travel in larger animals (Figure 1.1, step 2).

From an evolutionary perspective, it is logical that immune response times

scale faster than typical quarter-power scaling expectations since large animals could not survive a months-long delay in responding to exponentially growing pathogens. However, the mechanisms enabling larger mammals to respond as fast as smaller mammals remain unclear.

A second well-established scaling relationship is that organ size typically scales linearly across animals. For example, the heart, liver, and kidney are 1000 times larger in animals weighing 1000 times more [24]. In this paper, we show that this linear relationship holds for the spleen but not for LNs. LN size, number, and location vary with body mass due to their distribution throughout the body. We hypothesize that this non-linear allocation of LN size and number enhances immune response speed, compensating for other trade-offs in adaptive immunity (Figure 1.1, steps 4 and 5). We describe how more searchers in larger LN contribute to the unique scale invariance of the immune response.

Prior research has explored how LN size and number vary with body size and how that impacts the timing of the adaptive immune response [118, 7]. Perelson and Wiegel [120] theorized that if the benefits of larger LNs size and number were equally important and the total volume of LNs scales linearly with body mass, LN size and number should scale with the square root of body mass ($M^{\frac{1}{2}}$). Our study provides the first empirical data to test these theoretical predictions.

The benefit of more LNs is clear because a higher density of LNs reduces the mean distance between potential infection sites and the closest LN (Figure 1.1,

step 2) [13]. However, the benefit of larger LNs is unclear. [120] predicts that *typical* search times should be independent of the LN volume. That is, if the density of cells is approximately constant, then a typical T cell or B cell would find its target in approximately the same amount of time, regardless of LN volume (Figure 1.1 steps 3 and 4). Thus, the trade-off between LN size and number remains unclear. Why wouldn't a human have many thousands of LN, each as small as those of mice?

In previously published work [52], we present a mathematical model that predicts initial first contact times between searchers and targets distributed at random in a volume. Through simulations, we explore how the number of searchers, the distribution of searchers and targets, and the initial distances between searchers and targets affect initial first contact times. In this work, we develop a mathematical prediction for IFCT and test that prediction in simulations, to show that the first T cell contact with a DC is faster in larger LNs with larger T cell populations.

In most models, T cells contact cognate DC at the same time and start proliferating all at once. This is delayed in larger LNs due to the slower interaction between the last T cells and DCs. This delay implies that larger animals should take longer to initiate an immune response. However, we show that humans, which are thousands of times larger than mice, initiate an immune response just as quickly. Our IFCT model explains this observation. The IFCT model assumes that T cell proliferation begins as soon as the first T cell makes contact with a DC, which occurs earlier in larger LNs. As a

result, the IFCT model predicts significantly higher numbers of proliferating T cells in larger LNs compared to smaller ones, enabling larger animals to respond rapidly to infections with larger numbers of activated T cells.

3.4 Time to Initiate the Adaptive Immune Response Is the Same in Humans and Mice

We first establish that the timing of the first detectable adaptive immune response is similar in humans and mice. Table 3.1 shows that for a range of novel viral and bacterial pathogens, activated T cells are first detected in LN or tissues in both species within 4-10 days, with a typical detection time of 6 days. The time to detect activated T cells reflects the time for cells to move, activate, and proliferate, in addition to the search process we will focus on in the later part of this paper.

We note that the first detection of activated T cell populations is distinct from the peak T cell concentrations that are often measured in blood. It can take additional time to reach the peak after initial activation, particularly in larger animals. For example, peak T cell concentrations are observed in 5-10 days in mice [164, 106] and 14-28 days in macaques and humans [99, 144].

Table 3.1: **Time to Initial Detection of Activated T Cells in Mice and Humans.** Data are rounded to the nearest day (d). Means are calculated from the midpoints of each reported range of times observed in experiments. Remarkably, the mean, minimum, and maximum times are very similar in mice and humans.

<i>M. musculus</i> (24 g)	<i>H. sapiens</i> (62 kg)
Flu: 5 d [116, 106]	Dengue: 7 d [57]
Flu: 4 d–6 d [82]	Flu: 6 d [21]
Flu: 5 d–7 d [153]	LCMV: 4 d–5 d [35]
HSV: 5 d–7 d [19, 32]	RSV: 7 d–10 d [65]
LCMV: 5 d–7 d [72]	SARS-CoV2: 4 d [100, 86]
SARS-CoV2: 5 d–10 d [164]	SARS-CoV2: 6 d [77, 125]
SARS-CoV2: 7 d [141]	SARS-CoV2: 7 d [92, 74]
Staph: 6 d [139]	Staph: 7 d [20, 84]
Staph: 9 d [130]	
Mean: 6 d	Mean: 6 d
Min: 4 d	Min: 4 d
Max: 10 d	Max: 10 d

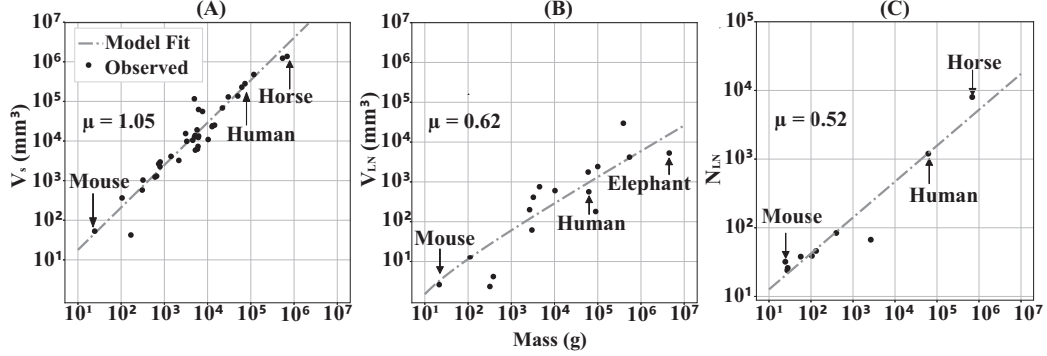


Figure 3.1: **Lymphoid Organ Scaling with Mass.** Each data point represents a species shown on the log-log axis. The dashed lines show the regression fit. **(A) Spleen Volume** of 38 species is best fit by the regression cM^μ with $c = 1.2$ and exponent $\hat{\mu} = 1.1$ (95% CI [0.95, 1.2]). **(B) Lymph Node Volume** is best fit by $c_1 M^\mu \ln(c_2 M)$ with $c_1 = 1$, $c_2 = 1$, and $\hat{\mu} = 0.62$ (95% CI [0.55, 0.69]) for 16 species. **(C) Number of Lymph Nodes** for 10 species is best fit by cM^μ with $c = 4$ and $\hat{\mu} = 0.52$ (95% CI [0.40, 0.64]). The p-value of the exponents is significant at the 0.01 level.

3.5 Empirical Scaling of Spleen Size

Figure 3.1(A) shows that spleen volume [103, 93] scales linearly with mass (also see Table 3.3). A linear regression on log-log transformed data was used to derive an exponent, $\hat{\mu}$ of 1.05 with 95% CI [0.95, 1.2] and with $R^2 = 0.91$. Although the data are consistent with the expectation of linear scaling of spleen size with M , we also found that the data are consistent with an additional logarithmic increase ($M \ln(cM)$) (See Section 3.9.2). Such a nonlinear scaling could accommodate the predicted logarithmic increase in lymphocyte diversity with M hypothesized in [120].

3.6 Empirical Scaling of Lymph Node Size and Number

Wiegel and Perelson [120] propose LN number and size scaling based on two key assumptions: first, maximizing LN volume and number are equally important, and second, the total LNs volume scales approximately linearly, proportional to body mass M (noting that scaling is also predicted to accommodate a small logarithmic increase in T cell diversity with mass). Based on these assumptions, they predict that the volume of a typical LN, V_{LN} , scales as follows, where c is a constant,

$$V_{\text{LN}} \propto M^{\frac{1}{2}} \ln(cM). \quad (3.1)$$

The number of LNs, N_{LN} , scale as,

$$N_{\text{LN}} \propto M^{\frac{1}{2}} \quad (3.2)$$

Figure 3.1(B) shows the scaling of LN volume with mammal mass for 16 species from mass 24 g (mice) to 4500 kg (Elephant). Regression of the form $V_{\text{LN}} = c_1 M^\mu \ln(c_2 M)$, produces, $c_1 = 1$, and $c_2 = 1$, $\hat{\mu} = 0.62$, 95% CI [0.55, 0.69]. This is close to, but slightly higher than, the predicted $\frac{1}{2}$ exponent. Excluding the logarithmic term would yield a higher exponent of $\hat{\mu} = 0.81$, 95% CI [0.56, 1.1]. Figure 3.1(C) also shows the scaling for LN number. The

best fit for 10 species, with mass ranging from 24g (mice) to 690 kg (Horse), for $N_{LN} = cM^\mu$ is $c = 3.8$, $\hat{\mu} = 0.52$, 95% CI [0.40, 0.64], consistent with the hypothesized exponent value of $\frac{1}{2}$. The data roughly align (though not perfectly) with the theoretical predictions given in Equations (3.1) and (3.2), as well as the linear scaling of spleen volume. Based on the scaling equation, we can roughly estimate human spleen volume, LN volume, and LN number relative to mice mass. Our theoretical expectation is that the spleen should be 2500 times larger, LN volume 350 times larger, and LN number 50 times greater in humans than in mice. Actual values from Table S1 are within a factor of two of these approximations. Given the more than three orders of magnitude difference in the sizes of humans and mice, predictions that are within a factor of two of empirical estimates are useful approximations, similar to the physiological scaling predictions of heart rates, breathing rates, and gestation times described in the introduction.

3.7 Initial First Contact Times with Respect to Animal Mass

To derive a prediction for the time for the first T cell to find its cognate antigen-presenting DC within a LN, we first consider a generic search problem between a population of T cells, N_{TC} , and a population of DCs, N_{DC} that the searchers are looking for in a LN volume V . In Section 3.9.4, we prove

Derivation 1 for the initial first-contact time, τ_{init} :

$$\tau_{\text{init}} \propto \frac{\lambda}{N_{\text{TC}} N_{\text{DC}}}$$

Where λ is defined as the mean first-contact time in a volume between a single T cell and a single DC. Celli et al [27] showed that λ scales linearly with volume ($\lambda \propto V$). Since we have the product of N_{TC} and N_{DC} in the denominator, search times decrease linearly with the increase of both T cells and DCs. Using Derivation 1 we predict search times within LNs assuming Equation (3.1), i.e. that volume is proportional to $M^{\frac{1}{2}} \ln(cM)$. We also assume that the number of T cells of each clonal line is proportional to M , and therefore N_{TC} is proportional to $M^{\frac{1}{2}}$ in each of $M^{\frac{1}{2}}$ LNs. We model a simplified scenario in which DCs transports a single type of antigen to LNs, and we only consider the T cell population that is cognate to that antigen. We derived (Derivation 1) and simulated the model for the constant density of T cells. In this paper, we consider two bounding cases for the scaling of the number of antigen-carrying DCs.

In the first case, we assume that the number of DCs in an animal is proportional to M so that the number in any particular LNs is proportional to $M^{\frac{1}{2}}$. This mirrors the scaling of cognate T cells, which is relevant for systemic infections that increase linearly with M , like widespread respiratory infections such as SARS-CoV-2, affecting a substantial portion of lung tissue. In this case, the amount of antigen carried by DCs would be proportional to

lung volume and M . Since we assume that DCs carry a number of antigens to the draining LN that is proportional to $M^{\frac{1}{2}}$, then $M^{\frac{1}{2}}$ DCs would migrate to each LN. In Section 3.9.4, we derive Prediction 3.2, Prediction 3.1 by substituting $N_{TC} \propto M^{\frac{1}{2}}$ and $V_{LN} \propto M^{\frac{1}{2}} \ln(cM)$ into Derivation 1. Under this assumption that the number of DCs carrying a particular antigen in a LN is proportional to $M^{\frac{1}{2}}$, the expected τ_{init} will scale as $M^{-\frac{1}{2}} \ln(cM)$, predicting that τ_{init} is shorter in the larger LNs of larger animals. This prediction is validated through our IFCTs simulation shown in Figure 3.2(A), where the result (red circles) aligns closely with the prediction (red line). Additional analysis of this case with a constant density of searchers and targets is shown in [52] demonstrating that τ_{init} scale as inversely proportional to volume.

For the other extreme bounding case of the number of DCs, we consider a small localized infection, akin to a pinprick. Here, the initial infection is introduced at a single point, resulting in a small infection that does not scale with animal size. Thus, in this case, we model the number of DCs carrying antigen to each LN as a small constant, and the expected τ_{init} scales as $\ln(cM)$. We compare this prediction to our simulations, shown in Figure 3.2, panel B in green. The prediction and simulation align closely with each other, indicating that τ_{init} scales logarithmically with animal size. Table 3.2 shows that even in the extreme case with only a small constant number of DCs, T cells still contact those DCs fairly quickly even in a large animal.

Table 3.2 shows the predicted τ_{init} for both systemic and localized infections. The scaling relationships result in predicted τ_{init} of less than an hour in

Table 3.2: **DC-T Cell Initial First Contact Times** considering 2 cases: i) assuming the number of DC in LN scales with $M^{0.5}$ (constant density of antigen-bearing DC from a systemic infection) and ii) a constant number of DC in LN ($N_{DC} = \text{Constant}$). We estimate times for each case, considering that T cells move in using either Brownian motion or a persistent random walk.

	$N_{DC} \propto M^{\frac{1}{2}}$		$N_{DC} = \text{Constant}$	
Mass	Brownian Motion	Persistent Random Walk	Brownian Motion	Persistent Random Walk
24 g (Mouse)	31 min	28 min	17 min	14 min
1 kg	9.5 min	9.1 min	35 min	28 min
62 kg (Human)	1.7 min	1.6 min	55 min	60 min

humans and mice for both cases. While most of our simulations use Brownian motion to model T cell searchers, we also investigate the empirically observed persistent motion observed in LNs in our prior work by Fricke et al. [59]. Persistent motion decreases cell contact times by a relatively small factor. Figure 3.2 shows that both predicted and simulated τ_{init} for systemic and localized infections. This intuitive outcome arises because the time for the rare fortunate *first* contact is expedited when more T cells are present. However, the advantage of a large population doesn't benefit the typical T cell, as the last T cell-DC encounter takes longer in larger LNs.

In addition to modeling the first time for any T cell to contact its cognate DC, we also consider the mean first-contact times ($\bar{\tau}$) for a typical T cell to contact its cognate antigen-bearing DC using a constant density of T cells for

both systemic and local infection, following the approach outlined in [120] (See Supplementary Figure 3.3 and Supplementary Table 3.6). For the systemic infection, where $\tau_{init} \propto M^{-\frac{1}{2}} \ln(cM)$, $\bar{\tau}$ scales logarithmically with body mass as $M^0 \ln(cM)$ and for the local infection, where $\tau_{init} \propto M^0 \ln(cM)$, $\bar{\tau}$ scale as $M^{\frac{1}{2}} \ln(cM)$. τ_{init} emphasizes the critical role of timely first encounters in initiating an effective immune response, while $\bar{\tau}$ reflects the collective behavior of T cells in a population.

The initial first contact time has practical implications for immune system modeling. To demonstrate this we reanalyze the empirical data presented in [36] using the IFCT model (see Supplementary Section 3.11 for details). In [36], the model assumes that contact between all T cells and DCs happens simultaneously. We implement this assumption in a Median model (see Supplementary Section 3.11 for details) using the median first contact time from our simulations. We compare this to our IFCT model which accounts for the time for each individual T cell to first contact its DC target. After 5.8 days and 5.6 days post-infection for two different epitopes - NP118 and GP283, the Median model predicts peak T cell populations of 3.9×10^7 and 1.2×10^6 , respectively, whereas the IFCT model predicts more than double these values at 9.1×10^7 and 3.4×10^6 - nearly twice as large.

Additionally, we run simulations to estimate the peak T cell population in smaller and larger LN over 8 days in a simple model that focuses on the exponential growth of T cells upon contacting their target DC. The IFCT model predicts peak T cell populations that are 2.7 and 40 times larger than

the Median model estimations for smaller and larger LNs, respectively.

Thus, by accounting for the rare early first contact, we estimate larger T cell populations, particularly in larger LN.

3.8 Discussion

The time to initiate the adaptive immune response is similar in mice and humans despite the three-order of magnitude difference in their mass. This unusual mass invariance in adaptive immune response times is accompanied by an unusual scaling of the organs in which adaptive immunity is initiated. LN number and volume both scale sublinearly with mammal mass (M), with exponents close to the theoretically expected $\frac{1}{2}$. This contrasts with spleen and other organ volumes that scale approximately linearly with M . Theory predicts one-half exponents to arise if scaling up LN size and LN number have equal benefits. An obvious benefit of more LNs is the distance from a site of infection to the nearest LNs is reduced, reducing time to transport antigen to the LNs [120, 13].

Here we show the benefit of larger LNs: the search for antigen-bearing DCs happens faster in the larger LN. Given the observed scaling of LN volume ($\propto M^{\frac{1}{2}} \ln(cM)$), and assuming constant density of T cells and DCs in LN, we predict faster initial contact time in larger LN by a factor of $M^{-\frac{1}{2}} \ln(cM)$. We validated this prediction in our simulations (Figure 3.2, panel A red circles and lines).

This result explains the hypothesis proposed in [120] - there is a benefit to larger LN volume as well as larger LN number, and the approximately one-half exponents suggest that evolution has found a middle ground in which immune cells and antigens move quickly from sites of infection into nearby LN, and T cells find antigens quickly within larger LN.

Our IFCT model, parameterized to match empirical observations of T cell population growth, shows a twofold increase in T cell population compared to previous estimates that were based on the mean first contact time rather than the initial first contact time (Supplementary Section 3.11). The faster growth of T cell populations could fight pathogens more effectively than was previously modeled.

According to metabolic scaling theory, quarter-power metabolic scaling relationships [160, 12] arise from systematic increases in transport time through the cardiovascular network in larger animals. Here we argue that immune response times arise from systematic changes in the lymphatic network with animal size [50]. Unlike the cardiovascular network, which directs flow from a single, central heart, the decentralized lymphatic network allows immune cells to move through LNs with varying sizes, locations, and numbers. This presents an alternative, and more flexible, way to scale that allows the immune response to meet the evolutionary imperative for rapid response against exponentially growing pathogens.

Here, we have highlighted one advantage that the distributed lymphatic network provides: balancing the speed of transport to LNs with many small

LNs with the faster detection of antigen within a few large LNs. However, there are other constraints on LN size and number. For example, LN must be big enough to hold a sufficient diversity of B and T cells and a sufficient number of exponentially growing activated B and T cells during an infection; both of these may vary with animal size. It remains an open question to fully explain how the scaling of LN size and number, the complex dynamics of replicating T cells [37], and the movement of both antigen and T cells into LN [50] result in such similar times (6 days) to observe the first T cell replication in both mice and humans in Table 3.1.

More generally, our analysis shows a benefit of large size that has not been previously appreciated. While nearly all scaling relationships show that bigger animals are slower, here we show that first search times are faster in larger mammals. This makes sense intuitively - when there are more searchers, the first target is found faster. This phenomenon has been studied by physicists as extreme first passage times [90]. In contrast to our findings of a linear speedup with size, previous extreme first passage time (EFPT) analyses find a much slower speedup that is only logarithmic with the number of searchers. The differences arise because EFPT considers an infinite number of searchers, all starting their search at the same physical location, with search trajectories that overlap. In contrast, in the LN search problem, a finite number of dispersed searchers in the 3D volume of a LN can be considered independent of each other, leading to a much greater (linear) advantage of large search populations that we identify here.

The different scaling properties of initial first contact times and typical first contact times are particularly relevant when the first contact causes a cascade of downstream events. In the initiation of adaptive immunity, when cognate T cells contact DCs, the T cells replicate (Figure 1.1 step 5), and changes occur in the LN, including slowing the egress of other T cells. Thus, the first contact changes the dynamics of subsequent searches. Further, the exponential growth of T cells begins once the first contact is made. Subsequent T cell contacts can amplify the T cell response, but the initial first contact causes the first T cell replication (Figure 1.1, step 5) that produces activated T cells to locate infection in tissues (Figure 1.1, step 7). The first arrival time of T cells in tissue is important in controlling exponentially growing pathogens, as has been shown in response to SARS-CoV-2 infection [144] and in our simulations of the timing of T cell response [110].

Understanding the different scaling properties of initial and mean first contact times is also relevant for other immunological processes, for example, the B cell search for T cells in LN (modeled in [120]) and effector T cell search for infected cells in peripheral tissue (modeled in [110]). The analysis here suggests that initial contacts may happen faster in larger animals with more immune cells, but typical and last contacts (i.e., those that clear an infection) might take longer [?].

This variation in immune response can affect the timing and duration of infection and infectiousness in animals of different sizes; this, in turn, can affect how diseases spread across animal communities [44]. The distinct scaling

properties of first, typical, and last search times warrant further study in immunology and biology more broadly. The different times to achieve typical, first, and last search events affect any biological search that involves large numbers of searchers. For example, the first ant in a colony that finds food should similarly depend on colony size, and when that first event happens, communication of the food location changes the search times for the typical ant in the population [53, 109]. Similarly, the first individual with a rare genetic mutation that confers some fitness advantage occurs faster in larger populations and then changes the downstream dynamics. Thus, we suggest that understanding how the timing of the initial first successful search depends on the number of searchers is an essential and previously neglected question.

3.9 Methods

3.9.1 Data Collection and Analysis for Animal Mass, Spleen Volume, Lymph Node Volume, and Lymph Node Number

Using published data, we create a dataset of LN numbers for 10 species, LN volumes for 16 species, and spleen volumes for 36 species. The data and references used in our analysis are summarized in Table 3.3. When published animal **masses** are reported as a range, we use the midpoint in our analysis. We use the median when multiple values are published for the same LN.

Table 3.3: **Published Spleen Volume, Lymph Node Volume and Lymph Node Numbers.**

Mean values and citations for body mass, spleen volume, lymph node (LN) volume, and LN number used in this paper. The number of LNs is the maximum reported value assuming that omissions are more likely than additions.

Common Name	Binomial	Animal Mass	Spleen Vol.	LN Vol.	#LN (Max)
Lab Mouse	<i>Mus musculus</i>	24 g [148, 46, 1]	37 mm ³ [46]	2.9 mm ³ [46]	32 [81]
Egyptian tomb bat	<i>Taphozous perforatus</i>	27 g [148]			24 [4]
Greater Mouse Tailed Bat	<i>Rhinopoma microphyllum</i>	28 g [148]			26 [4]
Mongolian Gerbil	<i>Meriones unguiculatus</i>	57 g [148]			38 [80]
Hamster	<i>Mesocricetus auratus</i>	130 g [148]	260 mm ³ [60]		39 [80]
Egyptian Fruit Bat	<i>Rousettus aegyptiacus</i>	130 g [148]			46 [4]
Tree Shrew	<i>Tupaia belangeri chinensis</i>	140 g [145]		14 mm ³ [145]	
Tarsius spectrum	<i>Tarsius tarsier</i>	170 g [148]	30 mm ³ [113]		
Common marmoset	<i>Callithrix jacchu</i>	310 g [113]	400 mm ³ [113]		
Guinea Pig	<i>Cavia porcellus</i>	380 g [78]		4.2 mm ³ [66]	
Rat	<i>Rattus norvegicus</i>	320 g [148]	720 mm ³ [79]	2.4 mm ³ [136]	84 [80]
Central American squirrel	<i>Saimiri oerstedii</i>	610 g [113]	860 mm ³ [113]		
Common squirrel monkey	<i>Saimiri sciureus</i>	660 g [113]	900 mm ³ [113]		
Three-striped night monkey	<i>Aotus trivirgatus</i>	740 g [113]	1800 mm ³ [113]		
Cotton-top tamarin	<i>Saguinus oedipus</i>	790 g [113]	1600 mm ³ [113]		
Northern greater galago	<i>Galago garnettii</i>	800 g [113]	2100 mm ³ [113]		
Gorilla	<i>Cercopithecus aethiops</i>	1.4 kg [113]	2900 mm ³ [113]		
Tufted capuchin	<i>Cebus apella</i>	2.2 kg [113]	2300 mm ³ [113]		
Possum	<i>Trichosurus vulpecula</i>	2.7 kg [148]		200 mm ³ [76]	67 [76]
Rabbit	<i>Oryctolagus cuniculus</i>	3 kg [45]		62 mm ³ [45, 67, 70]	
White-faced capuchin	<i>Cebus capucinus</i>	3.1 kg [113]	11 000 mm ³ [113]		
Crab-eating macaque	<i>Macaca fascicularis</i>	3.3 kg [113, 143]	6800 mm ³ [113]	410 mm ³ [108]	
Cat	<i>Felis catus</i>	4.5 kg [10]		760 mm ³ [126, 140]	
Bonnet macaque	<i>Macaca radiata</i>	4.5 kg [113]	7300 mm ³ [113]		
Blue monkey	<i>Cercopithecus mitis</i>	4.9 kg [113]	81 000 mm ³ [113]		
Black howler	<i>Alouatta caraya</i>	5 kg [113]	8800 mm ³ [113]		
Southern pig-tailed macaque	<i>Macaca nemestrina</i>	5.2 kg [113]	9600 mm ³ [113]		
Rhesus macaque	<i>Macaca mulatta</i>	5.2 kg [113]	4100 mm ³ [113]		
Lar gibbon	<i>Hylobates lar</i>	5.7 kg [113]	13 000 mm ³ [113]		
Japanese macaque	<i>Macaca fuscata</i>	5.8 kg [113]	4400 mm ³ [113]		
Tana River mangabey	<i>Cercocebus galeritus</i>	6 kg [113]	5100 mm ³ [113]		
Collared mangabey	<i>Cercocebus torquatus</i>	5.1 kg [113]	8600 mm ³ [113]		
Mantled howler monkey	<i>Alouatta palliata</i>	6.2 kg [113]	44 000 mm ³ [113]		
Stump-tailed macaque	<i>Macaca arctoides</i>	6.2 kg [113]	9300 mm ³ [113]		
Black-handed spider monkey	<i>Ateles geoffroyi</i>	7.6 kg [113]	39 000 mm ³ [113]		
Koala	<i>Phascolarctos cinereus</i>	10 kg [39]		600 mm ³ [68]	
Formosan rock macaque	<i>Macaca cyclopis</i>	10 kg [113]	7600 mm ³ [113]		
Olive baboon	<i>Papio anubis</i>	13 kg [113]	16 000 mm ³ [113]		
Yellow baboon	<i>Papio cynocephalus</i>	13 kg [113]	17 000 mm ³ [113]		
Chacma baboon	<i>Papio ursinus</i>	15 kg [113]	18 000 mm ³ [113]		
Dog	<i>Canis lupus familiaris</i>	22 kg [54]	48 000 mm ³ [54]		
Chimpanzee	<i>Pan troglodytes</i>	30 kg [113]	90 000 mm ³ [113]		
Sheep	<i>Ovis aries</i>	50 kg [54]	95 000 mm ³ [54]		
Dolphin	<i>Sotalia fluviatilis</i>	60 kg [148]		1800 mm ³ [146]	
Human	<i>Homo sapiens</i>	62 kg [113, 158]	160 000 mm ³ [113]	560 mm ³ [114]	1200 [123]
Gorilla	<i>Gorilla gorilla</i>	65 kg [113]	180 000 mm ³ [113]		
Goat	<i>Capra hircus</i>	88 kg [48]		210 mm ³ [98]	
Pig	<i>Sus scrofa</i>	120 kg [148]	330 000 mm ³ [54]	2000 mm ³ [134, 135]	
Camel	<i>Camelus dromedarius</i>	390 kg [124]		30 000 mm ³ [3, 124]	
Cattle	<i>Bos taurus</i>	540 kg [54]	860 000 mm ³ [54]	4200 mm ³ [75]	
Horse	<i>Equus caballus</i>	690 kg [148]	950 000 mm ³ [54]		8000 [8]
Elephant	<i>Loxodonta africana</i>	4500 kg [71]		5300 mm ³ [71]	

When animal sex is reported, we use male values, because many studies that specifically report LN sizes for female animals are studying the effects of cancer, which impacts LN volume [4].

We compute the **spleen volume** from the mass using the spleen tissue density (1.1 g mL^{-1} [29]). When published **LN numbers** were given as a range, the upper limit is used in our analysis. We use the median values of the number of LNs reported for multiple strains of mouse. **LN volumes** are approximate estimates for each species, which are obtained using various methodologies in order to facilitate comparisons, as LNs volumes can vary significantly, even within a single individual, particularly in larger animals. For instance, Qatarneh et al. [123] give ranges from 2 mm to 38 mm for human LN diameter. Sometimes, the total volume of each LN is reported, but often only the width and length, or possibly just length, is reported. To estimate volume, we assume LNs are spherical in the case of a single measurement and ellipsoid in the case of two measurements. This may be a source of error in estimating the LN volume. Publications also vary in the number of samples used. When a publication reports values for multiple LNs or from multiple individuals, we take the median value. Different measurement methods, i.e., imaging vs. physical measurements using calipers, may introduce additional variation in the estimated volumes of LN. Although several sources measure a large subset of LNs in larger animals (Table 3.3), we are not aware of a complete size estimate for all LNs in any large animal. However, human LN size is particularly well studied, which provides some context to interpret variation

in LN size. Typically human LN is estimated to be approximately 1 cm in diameter, and diameters larger than 2 cm are considered abnormal or a sign of a disease [142, 38]. We use the mean value of the middle of the range in [115], which is exactly 10.25 mm, and consistent with the 1 cm rule of thumb.

3.9.2 Alternative Hypothesis of the Scaling of Spleen Volume

Although the spleen volume data are consistent with the expectation of linear scaling of spleen size, we also tested an alternative of nonlinear scaling based on the hypothesis that the total population of T cells and B cells should be proportional to mass M , multiplied by a logarithmic term ($\ln(cM)$, c = arbitrary constant, M = animal body mass)[120]. This assumes that the number of T cells in each clonal line (that recognize a specific antigen) is proportional to M , and the diversity of T cell clones increases logarithmically with M based on the diversity of receptors expected by shape-space theory [119, 162]. This prediction is consistent with humans' slightly larger T cell repertoire compared to mice [11, 26, 157, 73].

Including a logarithmic term in the regression results in a best-fit exponent of $\hat{\mu} = 0.99$. The Akaike's Information Criterion (AIC) [6] values for the models with and without logarithmic terms are 23 and 21, respectively. The lower AIC value slightly favors the model without logarithmic terms, but with such close AIC values, the theorized logarithmic term to accommodate

greater lymphocyte diversity in larger animals also is plausible.

3.9.3 Statistical Analysis

We assessed the goodness of fit of our regression models to empirical data by computing the R-squared value, using the *fit* function in MATLAB for the linear model (the LN number, spleen volume without the log term) and the Akaike’s Information Criterion (AIC) value, using the *fitnlm* function in MATLAB for the non-linear model (LN volume, spleen volume with the log term). We computed *p-values* using the *fit* function in MATLAB, which uses a two-tailed t-test.

3.9.4 Derivation of Initial and Mean First Contact Times

We derive mathematical equations to determine the expected first-contact times (τ) between T cells and different numbers of DCs carrying antigen within LNs. We follow the assumptions and approach in Perelson and Wiegel [120] and the Brownian motion model of contact time detailed in Celli et al. [27]. All the notations used in this paper are listed in Table 3.4.

Assumption 1. *We assume encounter times are memoryless, and therefore, exponentially distributed. Specifically, if the number of T cells that recognize a given antigen in a LN is N_{TC} , then the probability that one DC chosen at random encounters any of the cognate T cells in time t , $P(T_t)$, can be modeled*

Table 3.4: **Notation** Variables used in this work.

M	Body mass
V_{Spleen}	LN volume
V_{LN}	LN volume
N_{LN}	LN number
c	Arbitrary constant
N_{DC}	Number of DCs of antigen per type per LN
N_{TC}	Number of T cells of antigen per type per LN
λ	Mean first-contact time between intra-LN a single cognate DC and a T cell
t	Time interval for intra-LN cognate DC-T cell random encounter
τ_{init}	Expected value of intra-LN single cognate DC-T cell pair initial first-contact time
$\bar{\tau}$	Mean value of intra-LN cognate DC-T cells first-contact time
μ	scaling exponent

with an exponential distribution [27],

$$P(T_t) = \frac{N_{\text{TC}}}{\lambda} e^{-\frac{t N_{\text{TC}}}{\lambda}}$$

Note that we experimentally validated that the assumption that contact times are exponentially distributed is reasonable. We found that in all simulations, the exponential fit was better than a normal distribution fit according to the sum square error (SSE) and AIC (see Figure 3 and Table 2 in [52])

Definition 1. λ is the mean contact time between a single searcher and a single target within a volume V . Celli et al. [27] showed that $\lambda \propto V$.

Derivation 1 (General Scaling Formula for Expected Initial First Contact Time). The expected initial first contact time is equal to the mean first contact

time between a single searcher and a single target in a volume divided by the product of the number of searchers and the number of targets present in the volume.

Given there are N_{DC} targets and N_{TC} searchers, then from Assumption 1, the probability that at least one searcher encounters a target between time 0 and time t is,

$$P(T < t) = \frac{N_{\text{DC}}N_{\text{TC}}}{\lambda} e^{-\frac{tN_{\text{DC}}N_{\text{TC}}}{\lambda}}$$

To obtain the expected encounter time, we integrate over all time,

$$\tau_{\text{init}} = \int_0^\infty \frac{tN_{\text{DC}}N_{\text{TC}}}{\lambda} e^{-\frac{tN_{\text{DC}}N_{\text{TC}}}{\lambda}} dt$$

$$\tau_{\text{init}} = \frac{\lambda}{N_{\text{DC}}N_{\text{TC}}}$$

Derivation 1 states that the expected first contact time is inversely proportional to the product of the number of T cells and DCs. In other words, as the product of these two quantities increases, the expected first contact time decreases.

Assumption 2. *To model the initial and mean first contact time, we make the parsimonious assumption that T cells and DCs are initially uniformly distributed within the T cell zone of the LN.*

Derivation 2 (General Scaling Formula for the Mean First Contact Time). *The mean time a searcher takes to make first contact with a target scales as*

volume divided by the number of targets.

From Definition 1 λ scales with volume V . From Assumption 2 the cells are uniformly distributed. Following Perelson and Weigel [120], the time to find a target is inversely proportional to the number of targets (N_{DC}), yielding:

$$\bar{\tau} \propto \frac{V}{N_{\text{DC}}}$$

Assumption 3. *In order to model search times within a LN, we assume that the number of T cells of each clonal line in an animal is proportional to M , and therefore N_{TC} is proportional to $M^{\frac{1}{2}}$ in each of its $M^{\frac{1}{2}}$ LNs.*

Assumption 4. *Perelson and Wiegel [120] predict that the volume of a LN (V_{LN}) scales as $M^{\frac{1}{2}} \ln(cM)$ this is to accommodate the greater clonal diversity of larger mammals.*

Derivation 3 (Initial First Contact Time Between Cognate T Cell and DC Within the LN). *the expected initial first contact time between cognate populations of DCs and T cells scales logarithmically with mass divided by the number of DCs, i.e.,*

From Definition 1 we have $\lambda \propto V_{\text{LN}}$ and from Assumption 4 we have $V_{\text{LN}} \propto M^{\frac{1}{2}} \ln(cM)$, therefore

$$\lambda \propto M^{\frac{1}{2}} \ln(cM)$$

Substituting λ into Derivation 1 and $M^{\frac{1}{2}}$ for N_{TC} from Assumption 3, yields the expected initial first contact time in a LN:

$$\begin{aligned}\tau_{\text{init}} &= \frac{\lambda}{N_{\text{DC}}N_{\text{TC}}} \\ &\propto \frac{M^{\frac{1}{2}} \ln(cM)}{N_{\text{DC}}M^{\frac{1}{2}}} \\ \tau_{\text{init}} &\propto \frac{\ln(cM)}{N_{\text{DC}}}\end{aligned}$$

We derive τ_{init} for two relevant bounding cases: 1) the number of DCs responding to infection is proportional to M (i.e., a constant density of DCs respond to a systemic infection). Therefore, the number in a single LN is assumed proportional to $M^{\frac{1}{2}}$ in each of the $M^{\frac{1}{2}}$ LN. 2) when the number of DCs responding to an infection is constant, reflecting a localized infection. For each of these boundary cases, we derive the expected initial first contact time and the mean first contact time.

Prediction 3.1 (Initial First Contact with Constant Density of DCs). *The density of cognate DC is constant with respect to M , and the number of LNs scales with $M^{\frac{1}{2}}$. Hence the number of DCs (N_{DC}) per LN scales with $M^{\frac{1}{2}}$. Substituting the scaling of DCs in Derivation 3, we have $N_{\text{DC}} \propto M^{\frac{1}{2}} \Rightarrow \tau_{\text{init}} \propto M^{-\frac{1}{2}} \ln(cM)$.*

Prediction 3.2 (Initial First Contact with Constant Number of DCs). *Similarly, if the number of DCs presenting cognate antigen in LNs is constant as M increases then we substitute a constant for N_{DC} in Derivation 3. Then*

the expected initial first-contact time scales logarithmically with mass, i.e.,
 $N_{\text{DC}} = C \Rightarrow \tau_{\text{init}} \propto \ln(cM)$.

In addition to modeling the initial first time for a T cell to contact its cognate DC, we also consider the mean first-contact times ($\bar{\tau}$) for a typical T cell to contact its cognate antigen-bearing using a constant density of T cells and DCs, following the approach outlined by Perelson and Wiegel [120]. Consistent with their predictions, we find a logarithmic increase in $\bar{\tau}$ (Figure 3.2, panel A, top blue line). Considering the other extreme case of a constant number of DCs carrying the relevant antigen, we predict mean contact times to increase with $M^{\frac{1}{2}} \ln(cM)$.

Derivation 4 (Mean First Contact Time Between Cognate T Cell and DC Within the LN). *The mean time a T cell takes to make first contact with a particular DC displaying cognate antigen when the LN volume scales with mass, scales as $M^{\frac{1}{2}} \ln(cM)$ divided by the number of DCs, i.e.,*

$$\bar{\tau} \propto \frac{M^{\frac{1}{2}} \ln(cM)}{N_{\text{DC}}}$$

The result follows immediately from substituting the value of V_{LN} from Assumption 4 into Derivation 2.

Prediction 4.1 (Mean First Contact with Constant Density Scaling). *The density of cognate DCs is constant with respect to M , and the number of LNs scales with $M^{\frac{1}{2}}$. Therefore, N_{DC} per LN scales with $M^{\frac{1}{2}}$. Substituting the*

scaling of DCs in Derivation 4, the expected mean first-contact time scales logarithmically with mass, i.e., $N_{DC} \propto M^{\frac{1}{2}} \Rightarrow \bar{\tau} \propto \ln(cM)$.

Prediction 4.2 (Mean First Contact with Constant Number Scaling). *Similarly, if the number of DCs presenting cognate antigen in LNs is constant as M increases, then we substitute a constant for N_{DC} in Derivation 4. Then the expected mean first-contact time scales as mass to the one-half power, i.e., $N_{DC} = C \Rightarrow \bar{\tau} \propto M^{\frac{1}{2}} \ln(cM)$.*

3.9.5 Agent Based Model

To validate our mathematical model, we implement an agent based model, termed IFCT using the MASON libraries [97] in Java. The model simulates two types of agents - T cell (searchers) and DC (targets). Both are uniformly distributed within a cubic representation of a LN. LN space is modeled as a continuous Cartesian grid in three dimensions with fixed reflective boundaries.

IFCT model is run as a discrete-time simulation, with each time step representing one second. At the beginning of each simulation run, a predefined number of T cells and DCs are initialized (see Table 3.5). The positions of DCs are static throughout the simulation. T cells, on the other hand, move in each time step with either Brownian motion or a persistent random walk, which was modeled from empirical data in Fricke et al. [59] (Note that previous analysis [27, 120], assumed Brownian motion). Supplementary Table 3.6 contains the results from both types of walks, demonstrating their

Table 3.5: Experimental factor levels used in the ABM experiments.

Mass (M)	10 g	1×10^2 g	1×10^3 g	1×10^4 g	1×10^5 g
$V_{LN} \propto M^{\frac{1}{2} \ln(cM)}$	7.3 mm ³	46 mm ³	220 mm ³	920 mm ³	3600 mm ³
$N_{TC} \propto M^{\frac{1}{2}}$	20	63	200	630	2000
$N_{DC} \propto M^{\frac{1}{2}}$	80	250	800	2500	8000
$N_{DC} = constant$	200	200	200	200	200

similarity. Figure 3.2 and Supplementary Figure 3.3 shows the result of Brownian motion. Following the hypothesis given in [120], we assume T cells scale $\propto M^{\frac{1}{2}}$. We model T cell motion without considering collisions so that T cells pass through each other.

We run the IFCT model for the two bounding cases, one in which $N_{DC} \propto M^0$, i.e. a constant, and another where $N_{DC} \propto M^{\frac{1}{2}}$. Each case is simulated for five different animal mass (M) values. For each M, we run 100 experimental replicas for each combination of factors where the initial distribution of the T cells and DCs are stochastic. We assume that there is a contact between a T cell and DC when their centers are within 10 microns. The simulation result from IFCT model confirms that the distribution of contact times is exponential, an assumption that underpins the theoretical derivation.

3.9.6 Fitting Exponents Using the ABM

The first contact time between any cognate DC and T cell initiates the adaptive immune response. To compute exponent μ over M in the initial

first-contact time formulas from the experimental data, we first generate a generic formula by combining Prediction 3.1 and Prediction 3.2 as follows,

$$\tau \propto M^\mu \ln(cM)$$

For ease of presentation, we convert the generic formula to base 10 as follows,

$$\begin{aligned} \log_{10}(\tau) &= \log_{10}(aM^\mu \ln(cM)) \\ &= \log_{10}(a) + \mu \log_{10}(M) + \log_{10}(\ln(c) + \log_{10}(M) \ln(10)) \end{aligned}$$

If we set, $C_1 = \log_{10} a$ and $C_2 = \ln(c)$ then,

$$\log_{10}(\tau) = C_1 + \mu \log_{10}(M) + \log_{10}(C_2 + \log_{10}(M) \ln(10))$$

and exponentiating,

$$\tau = 10^{C_1 + \mu \log_{10}(M) + \log_{10}(C_2 + \log_{10}(M) \ln(10))}$$

Perelson and Wiegel in [120] ask how long it takes for a single B cell to encounter a T cell given assumptions Equations (3.1) and (3.2). This approach considers the mean time it takes any particular B cell to encounter a T cell. They proposed a formula for the mean first contact time, $\bar{\tau}$, which a B cell needs to make first contact with some T cell of the complementary clone

should scale like,

$$\bar{\tau} \propto \frac{V_{LN}}{N_{TC}} \quad (3.3)$$

We used the same method to compute the fitted exponent value for the mean first contact time in our code (see the Supplement), as described above for the initial first contact time.

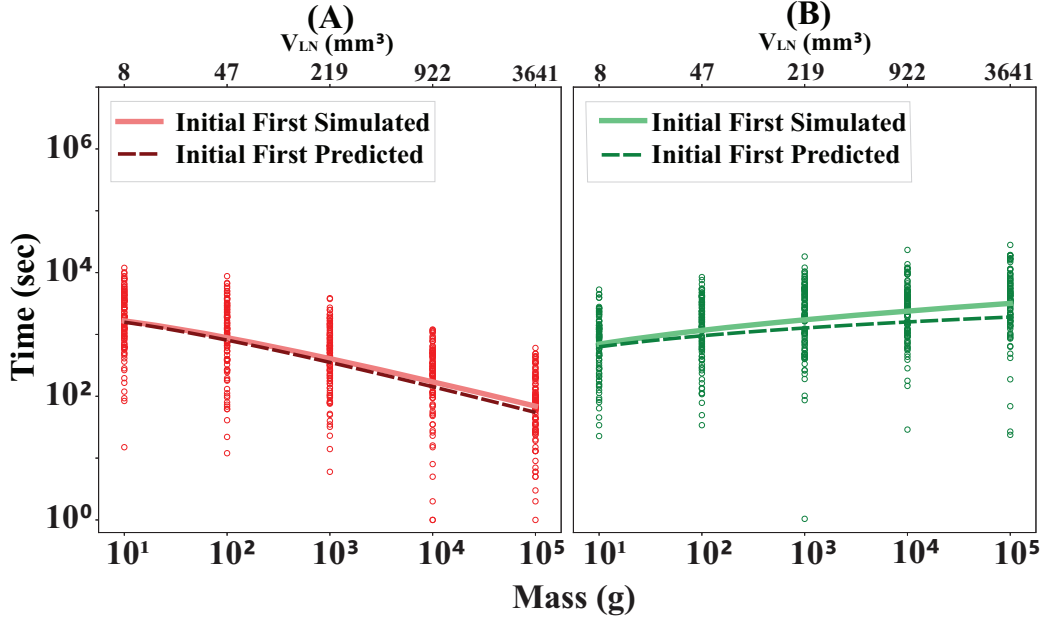


Figure 3.2: **Theoretical Predictions for Search Time vs Animal Mass are Consistent with Simulations of T Cell Contact Times in Lymph Nodes.** **Panel A:** For systemic infection ($N_{DC} \propto M^{\frac{1}{2}}$), the predicted initial first-contact time (τ_{init}) scales as $\tau_{\text{init}} \propto M^\mu \ln(cM)$ with $\mu = -0.5$ (dashed red line). The fit (solid red line) to simulated data (red circles) gives $\hat{\mu} = -0.53$ (95% CI[-0.61, -0.45]). **Panel B.** For small localized infections ($N_{DC} \propto M^0$), the theory predicts the initial first contact time as $\tau_{\text{init}} \propto M^\mu \ln(cM)$ with $\mu = 0$ (dashed green line). The best fit (solid green line) to simulated data (green points) gives $\hat{\mu} = 0.0$ (95% CI [-0.01, 0.01]). 100 simulations were run for each LN volume that was estimated (from Eqn 1) for each value of M. The simulations for these two cases are consistent with theoretical predictions.

Supplementary materials for Bigger is Faster: The Scalable Adaptive Immune Response

3.10 Mean First Contact Time

Table 3.6 shows the predicted values of mean first contact times $\bar{\tau}$ for both systemic and localized infections. For larger mammals, $\bar{\tau}$ is predicted to range from several days to weeks, which is far longer than the observed times for T cells to find their DC targets in the LN and replicate to detectable numbers. Thus, we suggest that the mean first contact time is not relevant for initiating the adaptive immune response, although it may be an important factor in determining the magnitude or timing of the peak T cell response.

Figure 3.3 shows that how both predicted and simulated $\bar{\tau}$ scale for systemic and localized infections. For systemic infection where N_{DC} scale as $M^{\frac{1}{2}}$, the prediction is $\bar{\tau} \propto M^0 \ln(cM)$ and the fit from the simulation gives

Table 3.6: **DC-T Cell Mean First Contact Times** considering 2 cases: i) assuming the number of DC in LN scales with $M^{0.5}$ (constant density of antigen-bearing DC from a systemic infection) and ii) a constant number of DC in LN ($N_{DC} = \text{Constant}$). We estimate times for each case, considering that T cells move in using either Brownian motion or a persistent random walk.

Mass	$N_{DC} \propto M^{\frac{1}{2}}$		$N_{DC} = \text{Constant}$	
	Brownian Motion	Persistent Random Walk	Brownian Motion	Persistent Random Walk
24 g (Mouse)	0.80 d	0.78 d	0.65 d	0.51 d
1 kg	1.7 d	1.8 d	9 d	7.1 d
62 kg (Human)	2.9 d	3 d	17 wk	13 wk

$\bar{\tau} \propto M^{0.01} \ln(cM)$. The slight increase in $\bar{\tau}$ for systemic infection in Figure 3.3 indicates that $\bar{\tau}$ increases logarithmically with mass, despite the near-zero exponent of M.

For a localized infection ($N_{DC} = \text{Constant}$), $\bar{\tau}$ scales as $M^{\frac{1}{2}}$. This intuitive outcome arises because the time for the rare fortunate *first* contact is expedited when more T cells are present. However, the advantage of a large population doesn't benefit the *typical* T cell, as they don't experience a faster encounter.

3.10.1 Mean First Contact Times Within an Abstract Volume

If the density of T cells and DCs are kept constant, $N_{TC} \propto N_{DC} \propto V$, then from Derivation 1,

$$\tau_{\text{init}} \propto V^{-1} \quad (3.4)$$

That is, given a constant density of T cells and DCs, the time for the initial first contact between a T cell and a DC is inversely proportional to the LN volume. When the density of the targets is constant, $N_{DC} \propto V$, then from Derivation 2, the mean first contact time is constant with respect to volume, in sharp contrast to t_{init} :

$$\bar{\tau} \propto V^0 \tag{3.5}$$

These simple equations predict that if a constant density of searchers look for a constant density of targets in a larger three-dimensional space, the first contact will happen faster, the speedup will be proportional to the volume in which the search takes place, and the mean first contact will be constant across all LN volumes. We simulate a uniform random distribution of searchers and targets in Figure 4 from [52] which shows that initial first contact times scale inversely proportional to volume. Figure 3.3 shows that for uniform random distribution of searchers and targets, the mean first contact is constant across volume as predicted by Equation (3.4) and Equation (3.5). The inverse linear dependence of τ_{init} on both N_{DC} and N_{TC} occurs because each searcher is assumed to search for each target independently. This assumption holds for searchers moving with a random walk in a three-dimensional space, leading to an exponential distribution of search times (see Figure 3 and Table 2 in [52])

3.11 Model Validation

We used experimental data from De Boer et. al. [36] to validate that our IFCT simulations provide biologically realistic estimates of the times for T cells to first contact their cognate DC. We compared contact times and resulting peak T cell populations from two models. We compare our IFCT model that explicitly represents the time that each T cell takes to contact its cognate DC to a Median model that uses the median first contact time as the estimated contact time for all T cells to contact their cognate DC. We show that the median contact time that we estimate from our IFCT model is consistent with data and assumptions in [36]. We also show that explicitly considering the variation in contact times leads to larger estimated peak T cell populations.

We set the parameters of our IFCT simulation to correspond to the empirical system modeled in [36]: 60 precursor T cells searching for DCs in a spleen of volume 46 mm^3 for two different epitopes (NP118 and GP283). We note that these empirical findings are close to our scaling estimate that a 37 mm LN should contain 63 T cells (Table 3.3).

Following [36], we consider four post-infection phases from the initial infection until the T cells reach their peak population:

1. travel time for a T cell and antigen carrying cognate DCs to travel to the spleen (t_{travel})

2. initial contact time between the T cells and their cognate antigen-bearing DC (τ_{init})
3. activation time it takes for a T cell to become fully activated upon encountering the cognate DC (t_{act})
4. proliferation time it takes for activated T cells to replicate (t_{rep}).

We define recruitment time (t_{rec}) as

$$t_{\text{rec}} = t_{\text{travel}} + \tau_{\text{init}} + t_{\text{act}}$$

The model of the mouse spleen [36], which we call the Median model, uses the same t_{rec} and the same τ_{init} for every T cell. In [36], t_{rec} was estimated to be 1.2 days for epitope NP118 and 1.7 days for epitope GP283. The model in [36] does not explicitly estimate τ_{init} . Here we estimate this time for all T cells as the median of all initial contact times ($\tau_{\text{init}}^{\text{median}}$) collected from our IFCT model.

In our IFCT simulation, we draw each contact (τ_{init}^i) from our simulation that ranges from 0.47 hrs to 150 hrs (0.019 days to 6.4 days) with a median ($\tau_{\text{init}}^{\text{median}}$) time of 22 h (0.9 days). The list of first contact times for each simulated T cell is available here: ¹.

To compare the IFCT model to the Median model, we compute the sum of the travel time and activation time ($t_{\text{travel}+\text{act}} = t_{\text{travel}} + t_{\text{act}}$) by subtracting

¹https://docs.google.com/spreadsheets/d/1ur3b4BpyfezHVAKFbyul0Y7N_X9-QGBA.BxZjYZuvsA/edit?usp=sharing

the median first contact time $\tau_{\text{init}}^{\text{median}}$ from the recruitment time t_{rec} .

$$t_{\text{travel+act}} = t_{\text{travel}} + t_{\text{act}} = t_{\text{rec}} - \tau_{\text{init}}^{\text{median}}$$

This gives us the summation of travel and activation time to be 6.8 hr and 18.8 hr for NP118 and GP283 epitopes respectively.

We calculate proliferation times (t_{rep}) as the inverse of proliferation rates r_{rep} in [36] given as 2.9 per day for epitope NP118 and 2.6 per day for epitope GP283. Thus t_{rep} is 0.39 days and 0.35 days for these epitopes. [36] estimated that proliferation stops, reaching the peak population t_{peak} , at day 5.8 for epitope NP118 and at day 5.6 for epitope GP283. The IFCT model assumes the exponential growth occurs between t_{rec}^i and t_{peak} , and we define that remaining time t_{rem}^i for each T cell i as follows,

$$t_{\text{rem}}^i = t_{\text{peak}} - t_{\text{travel+act}} - \tau_{\text{init}}^i$$

We continue this until $t_{\text{rem}}^i \geq t_{\text{rep}}$ so that we only count the T cells that complete proliferation before the t_{peak} . The total number of T cells (N_{peak}) at t_{peak} was calculated in the IFCT model as

$$N_{\text{peak}} = \sum_{i=1}^n e^{t_{\text{rem}}^i * r_{\text{rep}}}$$

In Figure 3.4, we show the correspondence of the mean T cell population dynamics of 10 simulations from the IFCT model with data from [36] over

Table 3.7: **Parameters and Peak Population estimates from the Median and IFCT models.** The recruitment time and proliferation time are from [36]. The travel and activation time ($t_{travel+act}$) represents the corresponding time after subtracting the median first contact time (τ_{init}^{median}) from the recruitment time (t_{rec}). Peak T cell populations are after 5.8 days and 5.6 days post-infection for epitope NP118 and GP283 respectively. The median first contact time in the Median model is 22 hr for all T cells and the first contact time is different for every T cell in the IFCT model.

Parameter	Symbol	Units	NP118	GP283
Recruitment Time	t_{rec}	days	1.2	1.7
Travel and Activation Time	$t_{travel+act}$	days	0.26	0.76
Proliferation Time	t_{rep}	days	0.35	0.39

Peak T cell Population Based on Fits		
Median Model Estimation	3.9×10^7	1.2×10^6
IFCT Model Estimation	9.1×10^7	3.4×10^6
Fold increase	2.3	2.1

time and with data from Median Model at t_{peak} . Table 3.7 shows that the IFCT estimated peak T cell populations at t_{peak} are twice as high as the Median model.

This analysis shows two things. First, the median first contact time τ_{init}^{median} calculated from our IFCT simulations is consistent with times in [36] when the recruitment time t_{rec} is broken down into travel, simultaneous contacts, and activation phases. Second, when we model the process in which each T cell finds its cognate antigen and then proliferates exponentially, the population of T cells begins growing earlier and reaches a higher peak compared to the

modeling of the average T cell in prior work.

3.11.1 Analysis of IFCT model effect on peak T cell population across different lymph node sizes

We further analyze T cell population growth for systemic infection ($N_{DC} \propto M^{0.5}$). For simplicity, we don't consider the travel time and activation time; we model the proliferation that begins after the first contact time τ_{init} between T cell and DC. We consider the T cell population reaches its peak at t_{peak} , giving us t_{rem} as

$$t_{rem} = t_{peak} - t_{init}$$

We compare the population in the Median and IFCT models for two cases: a smaller LN of 46 mm³ with 60 T cells and a larger LN of 3600 mm³ with 1733 T cells. The LN size (V_{LN}) and its corresponding T cell numbers (N_{TC}) are determined based on Assumption 4 ($V_{LN} \propto M^{0.5} \ln(cM)$) and Assumption 3 ($N_{TC} \propto M^{0.5}$). We assume the population reaches its peak at $t_{peak} = 8$ th day after the first T cell reaches the DC.

Figure 3.5 illustrates that the IFCT model begins replication immediately after the initial T cell contact (0.02 days for the smaller LN and 0.0015 days for the larger LN) which captures the critical early interactions between T cells and DC. In contrast, in the Median model, the initiation of the T cell replication is delayed until the median contact time (0.9 days for the smaller LN and 2.06 days for the larger LN). Table 3.8 shows the population at

day 4 and day 8 in both models. This early start in the IFCT model is consistent with biological evidence that T cells begin proliferating as soon as they encounter an antigen, a key event in initiating the immune response. As a result, the IFCT model predicts consistently higher T cell populations. By the time the Median model initiates proliferation, the IFCT model predicts 239 T cells in the smaller LN and 59700 T cells in the larger LN, which are 3.98 and 33.67 times higher, respectively, than the Median model's predictions at the median contact times of 0.9 and 2.06 days for smaller and larger LN. This substantial early proliferation in the IFCT model is particularly advantageous in larger LN with more T cells.

The ratio between the IFCT model predictions and the Median model predictions remains approximately proportional over time (2.68 for smaller LN and 41.44 for larger LN).

Additionally, Figure 3.5 shows that the T cell replication starts early in the smaller LN compared to the larger LN in the Median model. This is primarily because the median first contact time is longer in the larger LN than in the smaller LN. While the initial first contact occurs 13.33 times faster in the larger LN (0.0015 days) compared to the smaller LN (0.02 days), the final first contact is almost twice as slow in the larger LN (7.62 days) compared to the smaller LN (3.98 days). During the time difference between the median contact times of the smaller and larger LNs, the smaller LN undergoes approximately 11 more rounds of replication than the larger LN. However, in reality, replication in larger animals likely continues for a longer

Table 3.8: Comparison of T Cell Population Growth Predictions for Smaller and Larger Lymph Nodes Using Simplified Median and IFCT Models at Day 4 and Day 8. The table illustrates the median contact time and T cell population estimates after 4 and 8 days for a smaller LN (60 initial T cells) and a larger LN (1733 initial T cells) based on the two models. The Median model assumes simultaneous contact for all T cells, while the IFCT model allows for individual first contact times. X-fold growth represents the ratio between IFCT and Median model predictions for each scenario.

	Smaller LN w/ 60 T cells		Larger LN w/ 1733 T cells	
Median Contact Time (days)	0.90		2.1	
	After day 4	After day 8	After day 4	After day 8
Median Model Prediction	4.8×10^5	5.2×10^{10}	4.7×10^5	5.1×10^{10}
IFCT Model Prediction	1.3×10^6	1.4×10^{11}	1.9×10^7	2.1×10^{12}
Fold increase	2.7	2.7	40	41

duration, they may eventually catch up to the population in the smaller LN, despite the initial delay.

3.12 Caveats and Limitations

There are several caveats to consider in the empirical and modeling work we present here. First, all data we provide are species averages, and many are difficult to estimate from empirical data. Few studies have counted the total number of LN in different species, introducing potential variability. Additionally, estimating LN volumes accurately from one or two length measurements is difficult, particularly in large animals where LN sizes may vary dramatically within individuals. See Section 3.9 for further details on

potential biases from measurements. However, the approximate LN volumes reported in published literature show a clear trend with animal mass, allowing us to understand significant differences in LN size and number. This is validated by comparing data between the best-studied species – humans and mice. The distribution of LN volumes within animals remains a question for future studies.

Our mathematical models and simulations are, of course, simplifications of the complex search processes that occur in LNs. We model the search process as a Brownian random walk in a cube or sphere without considering the crowding or structures that may influence how T cells move in LNs, although we do consider alternate movements that mimic real T cell movement patterns [59] and find that realistic movement decreases search times by only a small percentage (Table 3.2 in the main file).

The assumption that T cells and DCs are uniformly distributed throughout the T cell zone at any given time is only an approximation. In reality, T cells and DCs enter LNs through lymphatic vessels over time. We simplify the complex dynamics of how T cells bind to different antigens presented by DC by considering only a single clonal line of T cells and a single antigen. It is unclear how many different clonal lines of T cells might match a particular antigen or the number of antigens DCs present from any given pathogen. There are empirical and theoretical suggestions that the clonal diversity of T cells increases slowly with M . Figure 3.1 in the main file shows spleen volume, and LN volume scaling could accommodate a linear increase in T

cells per clonal line and a logarithmic increase in the number of clonal lines. Still, more data are needed to test this hypothesis.

Because our IFCT model is focused on search times within LN, we do not address other potential causes or consequences of LN size and number. Physiological and anatomical constraints exist on the size, number, and placement of LN. LNs have multiple functions, for example, holding the population of exponentially growing B cells and T cells, which would logically also constrain the minimum size of LN. How these factors influence LN size remains an open question.

Finally, many factors determine the scaling of immune response times that we do not address here. The magnitude and scaling of the times to complete multiple steps in Figure 1.1 could be affected by the animal's size and metabolism. In particular, the transport of antigen from peripheral tissue to LN by DC would be longer in larger animals if DC speed is constant and LN are lower density and therefore further apart [13]. T cells might be expected to replicate more slowly in larger animals, following the typical quarter-power slowdown with body mass. This would mean it would take 7 times longer for T cells to replicate in humans vs. mice, but we are aware of no evidence of such systematic differences.

Perelson has hypothesized that LN might be metabolically privileged locations with sufficient metabolic resources for T cells to replicate as fast in humans as mice [120]. Finally, the timing of adaptive immune responses is intricately connected to innate immune responses, which have their own scaling

properties, e.g., those investigated in [43]. Others have identified differences in components of immune response across species [7, 34], suggesting a complex relationship between animal size and immune response that is important for translating observations between laboratory animals, animal sources of zoonotic disease and humans. While we leave a full accounting of factors that contribute to the scaling of adaptive immune response times to future work, the analysis we present here points to the decentralized architecture of the lymphatic network with sublinear increases in LN size and number with animal size as a key factor explaining how the adaptive immune response can be nearly as fast in large vs. small animals.

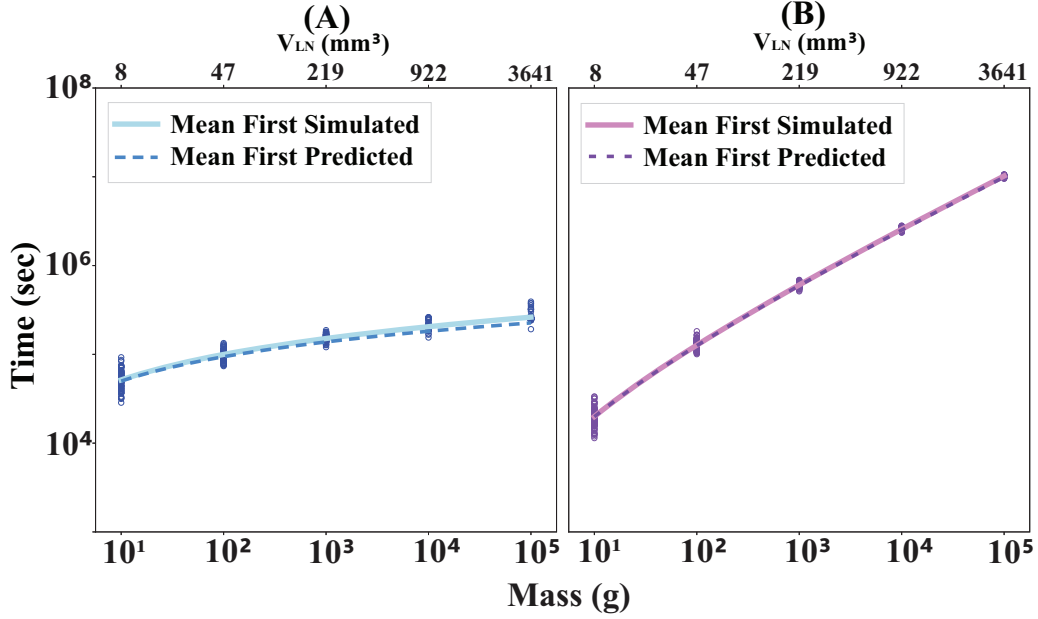


Figure 3.3: **Theoretical Predictions for Mean Search Time vs Animal Mass are Consistent with Simulations of T Cell Contact Times in Lymph Nodes.** **Panel A:** For systemic infection ($N_{DC} \propto M^{\frac{1}{2}}$), the predicted mean first-contact time is $\bar{\tau} \propto M^\mu \ln(cM)$ with, $\mu = 0$ (dashed blue line). The best fit (solid blue line) for the corresponding simulated data (blue points) gives the exponent $\hat{\mu} = 0$ (95% CI [-0.01, 0.01]). **Panel B.** For small localized infections ($N_{DC} \propto M^0$), the predicted mean first-contact time, is $\bar{\tau} \propto M^\mu \ln(cM)$ with $\mu = 0.5$ (dashed purple line). The best fit (solid purple line) to the simulated data (purple points) gives the value for the exponent, $\hat{\mu} = 0.50$ (95% CI [0.49, 0.51]). The simulations for all these two cases are consistent with theoretical predictions.

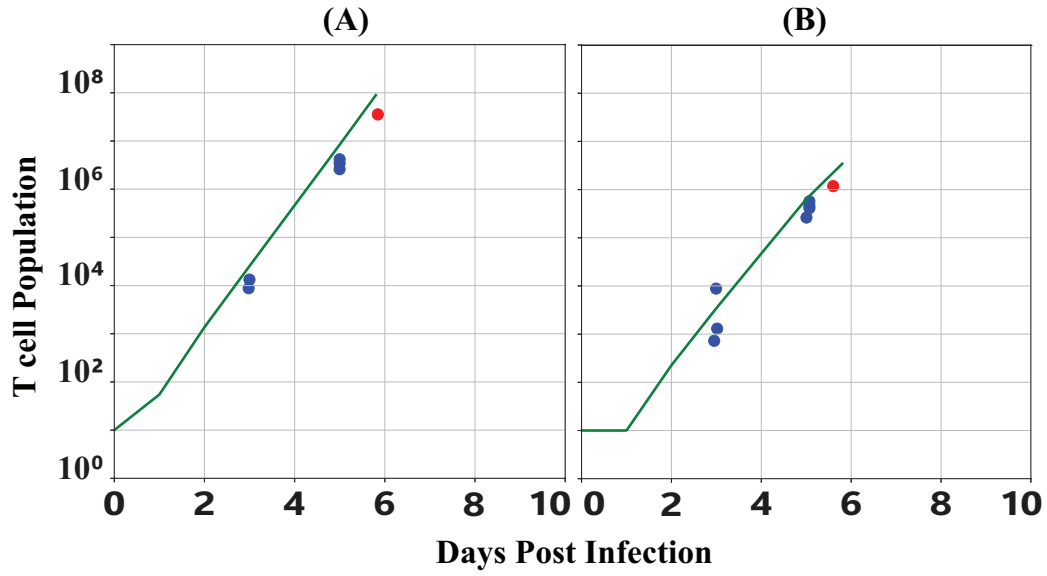


Figure 3.4: **T Cell Population Growth From the IFCT Model Compared to Empirical Data and Median Model.** Panel A depicts the T cell population over time in response to epitope NP118 for 5.8 days post-infection and Panel B depicts the same for epitope GP283 for 5.6 days post-infection. Blue data points are empirical data from [36] representing T cell populations at specified time points post-infection. The red data point is the Median model prediction of the peak T cell population. The green line shows the estimates from the IFCT model parameterized by data in [36].

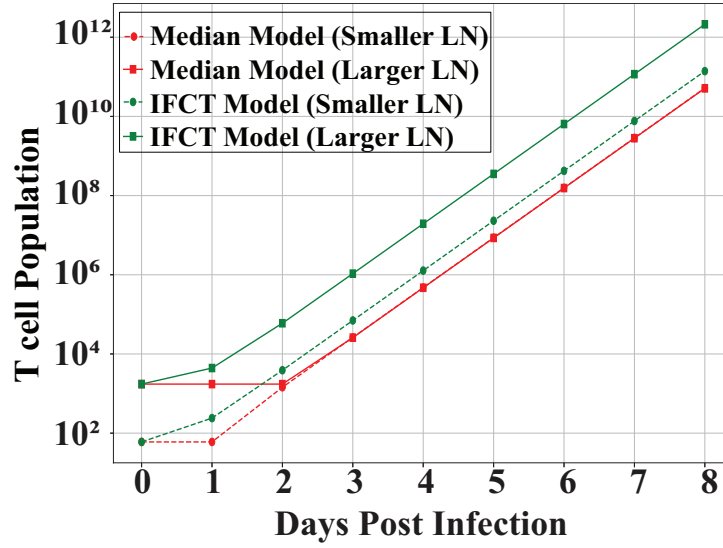


Figure 3.5: **T Cell Population Comparison Between Simplified Models for Smaller and Larger Lymph Nodes.** The figure compares the log-transformed daywise T cell populations predicted by the Median model and IFCT model upto 8 days of exponential growth. The red dashed line represents the Median prediction for a smaller LN (60 initial T cells), and the red solid line corresponds to the Median prediction for a larger lymph node (1733 initial T cells). The green dashed line represents the IFCT model for a smaller LN (60 initial T cells), and the green solid line corresponds to the IFCT prediction for a larger lymph node (1733 initial T cells).

Chapter 4

More is Faster: Why Population Size Matters in Biological Search

4.1 Publication Notes

Citation: Ferdous, J., Matthew Fricke, G. and Moses, M.E., 2024. More is Faster: Why Population Size Matters in Biological Search. *Journal of Computational Biology*, 31(5), pp.429-444.

Published online: 22 May 2024

Published ahead of print: 16 May 2024

Published in print: May 2024

Formatting: The original submitted text has been preserved as much as

possible while still adhering to the formatting requirements of this dissertation.

Funding: This work is funded by National Science Foundation (NSF), awards 2020247 & 2024520

Acknowledgements: We thank the UNM Center for Advanced Research Computing, supported in part by the NSF, for high performance computing resources. We thank Sid Redner, Judy Cannon, and the Moses Biological Computation Lab for helpful discussions that informed this manuscript. We thank anonymous reviewers for particularly helpful and insightful suggestions.

Author Contributions: JF wrote the software, performed computational experiments and created the figures under the supervision of GMF and MEM. JF, GMF, and MEM wrote the manuscript. MEM supervised JF on the selection of simulation parameters and place our results in context. JF derived the mathematical expressions under the supervision of GMF and MEM.

4.2 Abstract

Many biological scenarios have multiple cooperating searchers, and the timing of the initial first contact between any one of those searchers and its target is critically important. However, we are unaware of biological models that predict how long it takes for the first of many searchers to discover a target. We present a novel mathematical model that predicts initial first contact times between searchers and targets distributed at random in a volume. We compare this model to the extreme first passage time approach in physics that assumes an infinite number of searchers all initially positioned at the same location. We explore how the number of searchers, the distribution of searchers and targets, and the initial distances between searchers and targets affect initial first contact times. Given a constant density of uniformly distributed searchers and targets, the initial first contact time decreases linearly with both search volume and the number of searchers. However, given only a single target and searchers placed at the same starting location, the relationship between the initial first contact time and the number of searchers shifts from a linear decrease to a logarithmic decrease as the number of searchers grows very large. More generally, we show that initial first contact times can be dramatically faster than average first contact times and that initial first contact times decrease with the number of searchers while average search times are independent of the number of searchers. We suggest this is an underappreciated phenomenon in biology and other collective search

problems.

4.3 Introduction

There are many biological examples of a large number of searchers attempting to find one or more targets. Sometimes, the first target discovered by the first lucky searcher is particularly important because it results in a cascade of downstream events. For example, in ant colonies, many ants forage for food; when the first one is successful, it may use pheromones to recruit other searchers. Thus, the time for the first ant to discover food may have a disproportionate impact on resource collection rates. Similarly, naïve T cells search for cognate antigen in lymph nodes. When T cells bind with target antigen they activate and replicate in an exponentially growing population of cells. Thus, the timing of the first contact by the first T cell that finds its target is particularly important in initiating the adaptive immune response. In these examples, the time it takes for the initial discovery of a target (what we call the initial first contact time) may be more important than the average time of all of the searchers to find their first targets. Here we develop and analyze mathematical models and computer simulations to understand which factors affect initial first contact times.

In immune system models, it is common to consider the mean or median search time for a typical T cell to find its target, e.g., [120, 162, 87, 27, 50]. An analysis of initial first contact times in three-dimensional volumes has

only recently been investigated in a series of publications [90, 89, 91], but these models do not apply to T cell search that motivated our work. The derivations in these other studies apply to searchers that start in a single location and search for a single target, and they rely on population sizes that approach infinity. We call the predictions of initial first contact times from such models t_∞ .

In this work we derive initial first contact time predictions for finite populations of searchers, t_{fin} . This model initially assumes a number of searchers and targets distributed at uniform random within a volume, and we predict how t_{fin} scales as the number of searchers increases. We then test different search scenarios, assess when assumptions made by the t_{fin} and t_∞ models hold, and compare simulated initial first contact times predicted by t_{fin} and t_∞ .

While we were motivated to develop t_{fin} to predict how quickly the first T cell would find its target antigen in a lymph node, we also examine whether t_{fin} can explain other search scenarios relevant for social insects and other collective foragers. Many studies have investigated how colony size affects social insect foraging, finding that larger colonies are more complex [9], mass recruitment is more effective in larger ant colonies [16], but waggle dances are not necessarily more effective in larger bee colonies [42]. [122] found larger colonies searched an area faster and [5] found larger colonies found more events. However, none of these studies modeled whether the first event that triggered subsequent recruitment was found faster in larger colonies.

Some studies suggest this may be the case. For example, a field study by [40] found that larger honeybee colonies found resources faster, and as a result, overall foraging rates were better for larger colonies. Our previous field study[53] also found that larger seed harvester ant colonies found piles of seeds faster; however, neither of these studies had sufficiently detailed estimates of forager population size paired with and first discovery times to establish a quantitative relationship between the two. That is the focus of the models we build and analyze here.

Here we develop a model that predicts the initial first contact time for a finite number of searchers (t_{fin}). We simulate search in different scenarios in order to validate conditions under which our theoretical assumptions do, and do not, hold. Our goal in this paper is not to test model predictions with biological data (which we leave to future work), but we parameterize our simulations to reflect the numbers of T cell searchers, antigen-carrying dendritic cell targets, and lymph node volumes across mammals. We show that t_{fin} depends on the number of searchers (N). In scenarios where multiple searchers are involved, such as foraging ants, bees, or immune cells, a larger number of searchers can increase the likelihood of discovering a target quickly. We predict that more searchers exploring the environment increase the chances that one will stumble upon the target quickly, whether that target is a seed, a flower or an antigen. Our goal is to provide quantitative predictions and simulations that test this qualitative expectation given different numbers and placements of searchers and targets in different search volumes. In this paper

we develop a general model of initial first contact times, considering various placements of searchers and targets, and we compare our analysis for finite N to physics models of initial first contact times that assume an infinite number of searchers all starting in the same location. This work is a step toward a more general biological theory of first contact times. While we are unaware of quantitative models or data of initial first contact times, physicists have developed an extensive literature analyzing extreme first passage time (t_∞) [129, 88, 15, 159], an approach that predicts the time it takes for the fastest among an infinite number of searchers to find its target. However, t_∞ generally only considers search problems in which searchers all start at the same location and search for a single target, the number of searchers is assumed to approach infinity, and most work considers searchers on a one-dimensional line. To address these discrepancies from biological reality, we propose a mathematical model designed to predict first contacts in the immune system. Motivated by [49], t_{fin} is an idealized model of the initial first contact times between T cells and dendritic cells within the lymph node. This first contact is the critical step that initiates adaptive immunity and is a key factor in determining how long it takes T cells to control a viral infection. We consider a finite number of T cells as searchers and dendritic cells as targets, with both distributed uniformly in a bounded space representing the lymph node. We show in simulations that t_{fin} is proportional to the volume of the search space and inversely proportional to both the number of searchers and the targets. We investigate the influence of two factors that can reduce

search time when there are more searchers: 1) the initial distance between the searchers and the targets and 2) the chance that a searcher happens to take a direct path to the target. We expect that both the shortest distance to a target and the most direct path to the target will decrease when there are more searchers because both of these fortuitous events are more likely to occur at least once when there are more searchers. The major contributions of this paper are that we 1) develop a model (t_{fin}) that predicts initial first contact times for a finite number of randomly distributed searchers and targets, 2) compare t_{fin} to the t_{∞} model that assumes an infinite number of searchers all starting in one location, 3) analyze the models in four scenarios (cases 0-3) with varying searcher and target distributions in order to provide insights into the applicability of t_{fin} and t_{∞} in different biologically relevant scenarios, and 4) simulate idealized versions of those scenarios to examine how well the analytical predictions correspond to those scenarios. We consider the following cases: In case 0, we consider a random distribution of multiple searchers and multiple targets. We test whether t_{fin} assumptions are met and whether simulations are consistent with t_{fin} predictions. We then compare t_{fin} to t_{∞} , which assumes only a single target. Thus, for all other cases (1-3), we consider a single target located at the center of the search space ($T = 1$), and all searchers start at the same fixed point at some specified distance from the target (See Figure 4.1). The different predictions of the t_{∞} and t_{fin} approaches are summarized for each case in Table 4.1.

4.4 Predictions and Computational Methods

4.4.1 Summary of Scenarios

We show four scenarios in Figure 1. The t_{fin} and t_{∞} predictions of initial first contact times differ because each approach rests on different assumptions. The casewise predictions are summarized in Table 4.1.

- Case 0: For multiple searchers and multiple targets, both distributed at uniform random and with a constant density of searchers and targets, t_{fin} is expected to be inversely proportional to the volume of the search space and the number of searchers (N), Equation (4.5). One of the primary assumptions for t_{∞} is that all the searchers start from the same location; since that assumption does not hold in case 0, t_{∞} does not make a prediction for case 0.
- Case 1: For a fixed initial distance between all the searchers and a single target, t_{fin} is expected to decrease linearly with the number of searchers, according to Equation (4.7) (the same as case 0). In contrast, t_{∞} predicts the initial first contact times to decrease logarithmically with N (Equation (4.13)).
- Case 2: For a fixed number of searchers, t_{fin} is expected to increase proportionally to the cube of the distance between all the searchers starting from the same location and the target according to Equation (4.8). Our

analysis of t_∞ predicts that t_∞ scales quadratically with the distance in Equation (4.14).

- Case 3: If the density of searchers remains constant within a given volume, but all searchers initiate their search from a common starting point, t_{fin} is expected to be constant in Equation (4.9). t_∞ is expected to scale with length squared divided by N in Equation (4.15).

While t_{fin} and its underlying assumptions were originally formulated for Case 0, with multiple searchers and targets distributed randomly in a volume, we extend the examination of t_{fin} to cases 1-3 to assess whether it applies. Similarly, t_∞ and its assumptions were initially designed for Case 1 with the number of searchers approaching infinity. We explore how well the predictions of t_{fin} and t_∞ correspond to simulations given that the original assumptions may not hold in all cases. We note that our predictions for t_{fin} are based on an assumption that each searcher is independent, resulting in an exponential distribution of search times [27]. We will show that the assumption holds in case 0 (the original scenario that t_{fin} was developed for), but it may not hold for other cases.

4.4.2 Analysis of t_{fin}

We initially developed t_{fin} to predict the initial first contact time given randomly distributed searchers and targets in a volume. We assume that the contact times between the searchers and targets are memoryless random pro-

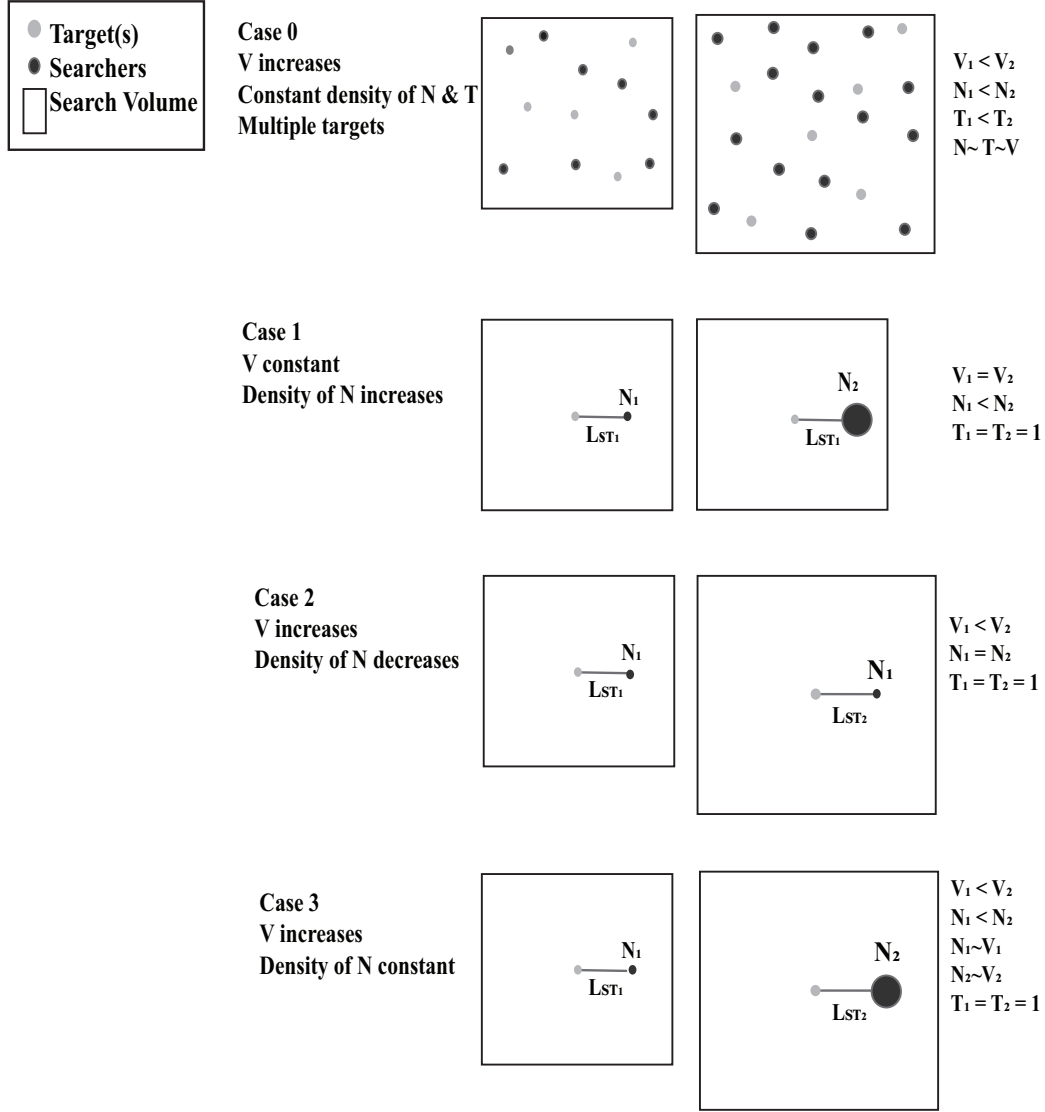


Figure 4.1: **Graphical Representation of the Simulated Cases.** In all cases there are multiple searchers (N). Here, V is the search space, T is the number of searchers, N is the number of searchers, L_{ST} is the length of the distance from the searchers to the target. In case 0, there are multiple targets and the targets and searchers are distributed at uniform random in the volume, and their density is constant as the volume increases. In cases 1 - 3, there is a single target and the searchers are all located at the same place at distance $L_{ST} = 0.25 L$ from the target so that searchers are equidistant from the target and the boundary. In case 1, the number of searchers N increases (indicated by the larger circle on the right) while L_{ST} and V are constant. In case 2, N is constant, and L_{ST} and V increase (indicated by the larger box on the right). In case 3, N , L_{ST} and V all increase (indicated by the larger box size and larger circle on the right). We assume L_{ST} is proportional to $V^{1/3}$.

Table 4.1: **Comparison of t_∞ and t_{fin} .** We assumed an exponential distribution of search times and a finite number of searchers in our model of t_{fin} . In contrast, t_∞ assumes an infinite number of searchers and that the fastest searcher will follow the shortest (and therefore, most direct) path to the target.

Experimental variation in N , T and V	t_∞ prediction	t_{fin} prediction
Case 0 (N , T and V increase; multiple searchers and targets are distributed at uniform random)	no prediction	$E[t_{\text{fin}}] \propto N^{-1}$
Case 1 (N increases, V is constant; searchers start at one location; single target)	$E[t_\infty] \propto \frac{1}{\ln(N)}$	$E[t_{\text{fin}}] \propto \frac{1}{N}$
Case 2 (N is constant, V increases; searchers start at one location; single target)	$E[t_\infty] \propto L^2$	$E[t_{\text{fin}}] \propto L^3$
Case 3 (N and V increase; searchers start at one location; single target)	$E[t_\infty] \propto \frac{L^2}{\ln N}$	$E[t_{\text{fin}}] \propto 1$

cesses, such that the first contact times of each searcher follow an exponential distribution. This is shown in case 0 as a bounded cube with N searchers looking for a number of targets (T). If the probability that a searcher finds the target in time t is $P(T_t)$, then the probability $P(T_t)$ that at least one searcher encounters a target between time 0 and time t can be modeled with an exponential distribution,

$$P(T < t) = \frac{NT}{\lambda} e^{\frac{-tNT}{\lambda}} \quad (4.1)$$

Here, λ is the mean contact time between a single searcher and a single target within a volume V . [27] showed that

$$\lambda \propto V \quad (4.2)$$

To obtain the expected encounter time, we integrate over all time,

$$E[t_{\text{fin}}] = \int_0^\infty \frac{tNT}{\lambda} e^{\frac{-tNT}{\lambda}} dt = \frac{\lambda}{NT} \quad (4.3)$$

By replacing λ with V from Equation (4.2) in Equation (4.3) we get,

$$E[t_{\text{fin}}] \propto \frac{V}{NT} \quad (4.4)$$

- **Case 0 (N, T, and L all increase, Multiple Targets):**

Here we assume there are multiple searchers and targets and the density

of searchers and targets are constant, $N \propto T \propto V$, then from Equation (4.4) we get,

$$E[t_{\text{fin}}] \propto V^{-1} \propto N^{-1} \quad (4.5)$$

That is, given a constant density of searchers and targets, the time for the initial first contact between a searcher and a target is inversely proportional to the search volume and the number of searchers. In this scenario, the initial first contact predicted t_{fin} speeds up linearly with N .

In Figure 4.3 we show that simulations of case 0 generate exponentially distributed contact times among all of the searchers in a simulation, consistent with the primary assumption in t_{fin} . In (Figure 4.4) we show that the initial first contact times produced by the simulations are consistent with the (t_{fin}) prediction. Next, in order to compare our t_{fin} prediction to the t_{∞} prediction, we consider 3 other cases with searchers starting at the same location and searching for a single target. Thus we set the number of targets $T = 1$, simplifying Equation (4.4) to

$$E[t_{\text{fin}}] \propto \frac{V}{N} \quad (4.6)$$

In addition to allowing us to compare the predictions of the t_{fin} and t_{∞} approaches, these scenarios also give us insights into the contribution of two factors that determine t_{fin} : (1) the initial distance between the lucky first searcher and the nearest target versus (2) the directness of the path of the

lucky first searcher toward the target because in these scenarios we fix (1) the distance between all searchers and the target.

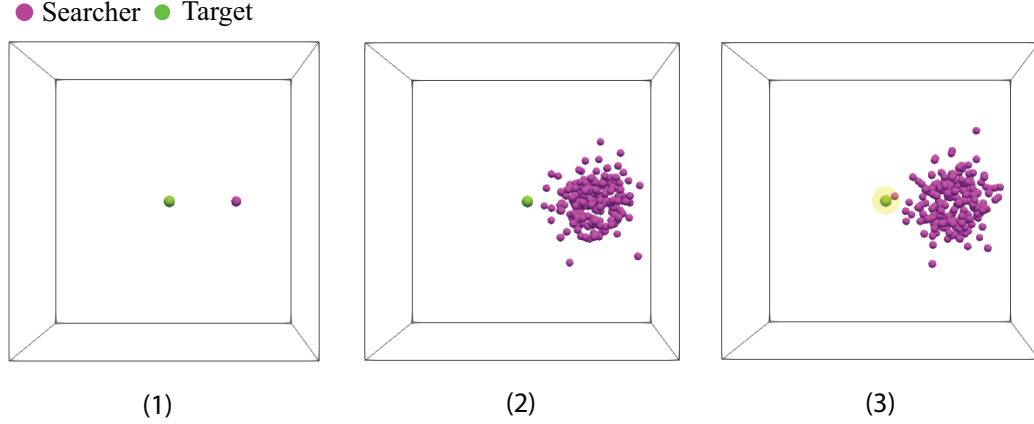


Figure 4.2: **Schematic Showing Simulations of Sases 1 - 3.** Multiple searchers (pink) looking for a target (green) in a confined cubic region, with the target placed at the center. The figure (A) shows the initial setup where all the searchers are in the same location with a fixed distance $L_{ST} = 0.25L$ from the target at the center of a fixed volume (Case 1). (B) the search progress at the 250th step, with the searchers moving via Brownian motion and (C) the first contact between a searcher and the target (Highlighted in yellow). **In Case 1**, we fixed volume at 8 mm^3 and systematically increased the number of searchers (N) from 5 to 100 000. **In Case 2**, we increased V from 8 to 33 000 mm^3 (L from 2 to 32) with fixed N . **In Case 3**, we increased both V and N with the N ranging from 240 to 1 000 000 by scaling with the volume ($N \propto V$). **In Case 0**, with randomly distributed searchers and multiple targets (not shown) we increased V , T and N with the N ranging from 20 to 2000 ($N \propto V$), T ranging from 80 to 8000, and V ranging from 8 to 790. Each simulation was replicated 30 times for each combination of parameters.

- **Case 1 (N increases, V is constant, Single Target):**

Since V is constant and N increases, Equation (4.6) is simplified to:

$$E[t_{\text{fin}}] \propto \frac{V}{N} \propto \frac{1}{N} \quad (4.7)$$

This case with a single target in a fixed volume gives the same prediction as case 0 for multiple randomly distributed searchers and targets in an increasing volume: as N increases, the expected initial first contact time decreases linearly.

- **Case 2 (N is constant, V increases, Single Target):**

Since $V = L^3$ and N is constant, Equation (4.6) becomes,

$$E[t_{\text{fin}}] \propto \frac{V}{N} \propto L^3 \quad (4.8)$$

In this scenario, the expected initial first contact time increases linearly with the volume. As the search volume (and therefore L) increase, it takes more time for the searchers to cover the increased space required to find the target.

- **Case 3 (N and V both increase, Single Target)** In this scenario we set $N \propto V$, so that Equation (4.6) predicts,

$$E[t_{\text{fin}}] \propto \frac{V}{N} \propto 1 \quad (4.9)$$

Thus if density is constant, t_{fin} is predicted to be constant across volumes.

4.4.3 Analysis of t_∞

The t_∞ prediction was developed by physicists to describe the time for the first of an infinite number of searchers to find a target in an idealized search, originally on a one-dimensional line [159, 129]. The theory has recently been expanded to more dimensions and more complex searches [88, 15, 90, 131], but most formulations consider only very large N that approaches infinity. t_∞ predicts that,

$$E[t_\infty] \propto \frac{L_{ST}^2}{4D \ln(N)}, \quad N \rightarrow \infty \quad (4.10)$$

where N is the number of searchers, D is a diffusivity, and L_{ST} is the distance between the identical starting location of all searchers and the target. The model assumes searchers move via a random walk. The time for the first searcher to reach the target is, in essence, the rate of spread of the furthest of the diffusing searchers. That diffusion creates a Gaussian distribution of distances from the start location. The logarithmic term in the denominator comes from determining the time it takes for the tail of the distribution to reach the target.

The full derivation of Equation (4.10) is in [90] (Equation 5-7, Section 2.1). t_∞ predicts that the distance traveled by the fastest searcher increases slowly. As the number of searchers increases, the time required for the first searcher to reach the target decreases as a slow logarithmic function of N . Note that this formulation of t_∞ doesn't specifically address how the volume of the

search domain affects search time; it considers only the distance between searcher and target. This is because the approach assumes that the fastest searcher will be the one that has taken the most direct path from the start location to the target; any searcher that meandered through the full volume of the search space would not be the first searcher to arrive, and therefore would not be relevant to the initial first contact time. This argument only holds for large N that approaches infinity. However, many biological search processes have a finite number of searchers within a bounded volume, for example, immune cells search in the volume of a lymph node or an infected organ, and foraging animals search within a 2D or 3D territory. We consider only the 3D case here and set L_{ST} to scale with the side L of the volume V of the search domain such that ,

$$L_{ST} \propto L \propto V^{1/3} \quad (4.11)$$

For simplicity, we consider a constant rate of diffusivity (D). By replacing L_{ST} with L and D with a constant, Equation (4.10) gives us a simplified expression for t_∞ in a bounded volume of length L :

$$E[t_\infty] \propto \frac{L^2}{\ln(N)}, \quad N \rightarrow \infty \quad (4.12)$$

We consider three cases for Equation (4.12).

- **Case 1 (N increases, L is constant, Single Target):**

In this case, we assume that L_{ST} remains fixed. As the number of searchers N increases, we reformulate the equation Equation (4.12) by ignoring scaling constants to obtain a prediction from t_∞ ,

$$E[t_\infty] \propto \frac{L^2}{\ln(N)} \propto \frac{1}{\ln(N)} \quad N \rightarrow \infty \quad (4.13)$$

This equation predicts that as N grows, the expected time for finding the target decreases logarithmically with N . The diminishing returns of adding more searchers are evident, indicating that large groups of searchers only slightly reduce the search time compared to smaller groups when the distance (L) is fixed.

- **Case 2 (N is constant, L increases, Single Target):**

In this case, we keep N constant while V (and therefore L_{ST}) increase. Then, Equation (4.12) becomes,

$$E[t_\infty] \propto L^2 \quad (4.14)$$

In this scenario, t_∞ is predicted to increase with L^2 while volume increases with L^3 .

- **Case 3 (N and L increase, Single Target):** In this case, N increases proportional to V , leading to:

$$E[t_\infty] \propto \frac{L^2}{\ln(N)}, \quad \text{where } N \propto V \propto L^3 \quad (4.15)$$

The different predictions from t_∞ and t_{fin} are summarized in Table 4.1.

4.4.4 Computational Experiments

In order to test the t_{fin} prediction for Case 0 and the competing predictions of t_{fin} and t_∞ for cases 1-3, we conduct a series of experiments using an agent-based model implemented in Biodynamico [18]. We consider that a contact happens when the centers of a searcher and a target are within $10\ \mu\text{m}$ distance. Hence the physical sizes of searchers and targets do not affect search times. To simulate T cells searching for dendritic cells in the lymph node, we utilized a model of Brownian motion parameterized from empirical data of T cell movement in lymph nodes from [59]. Figure 4.2 illustrates the experimental setup of our simulations designed to test the predictions of the t_{fin} and t_∞ approaches to estimating the initial first contact times in the 4 cases shown in Figure 4.1.

The scarcity of empirical data presents a challenge in directly matching the numerical simulations to real biological systems. We acknowledge this limitation and took steps to align the simulation parameters in case 0 with available data from the immune system. We chose search volumes to match the range of sizes of mammalian lymph nodes and estimated numbers of searchers to match the number of T cells in each clonal line that search for antigen-bearing dendritic cells in each lymph node and also estimated the number of targets for case 0 to match the number of dendritic cells in lymph node carrying the same type of antigens. The purpose of the

computational simulations is to establish a proof of principle showing how the initial first contact times vary with N and various other specifications of the search problem. Our intent here is not to test our predictions with biological data, but rather to demonstrate through simulations how various mathematical assumptions and predictions hold given simple simulations using finite numbers of searchers, targets, and volumes that approximate different biological scenarios. Our simulations validate cases where different approaches (t_{fin} and t_{∞}) are (and are not) able to predict biologically relevant phenomena. All the data used in our analysis are summarized in an online dataset.¹

4.4.5 Fitting Exponents Using the ABM

In order to fit model predictions to simulated data, we estimate the best-fit exponents that relate initial first contact times to the dependent variable in each scenario. We first consider the generic formula in Equation (4.4) as follows,

$$E[t_{\text{fin}}] \propto \frac{V^i}{N^j T^k} \quad (4.16)$$

¹<https://docs.google.com/spreadsheets/d/1AiWJ2onZYN5F-42LNLv-14kUluWyQhGSD3JFk-mNT4Q/edit?usp=sharing>

For ease of presentation, we convert Equation (4.16) to base 10:

$$\begin{aligned}\log_{10}(E[t_{\text{fin}}]) &= \log_{10}\left(\frac{aV^i}{N^jT^k}\right) \\ &= \log_{10} a + i \log_{10}(V) - j \log_{10}(N) - k \log_{10}(T)\end{aligned}\quad (4.17)$$

If we set, $C_1 = \log_{10} a$ then,

$$\log_{10}(E[t_{\text{fin}}]) = C_1 + i \log_{10}(V) - j \log_{10}(N) - k \log_{10}(T) \quad (4.18)$$

and exponentiating,

$$E[t_{\text{fin}}] = 10^{C_1 + i \log_{10}(V) - j \log_{10}(N) - k \log_{10}(T)} \quad (4.19)$$

For Case 0, where numbers of searchers and targets maintain a constant density as volume increase (i.e., $V \propto N$ and $T \propto N$), we further simplify eq. (4.18) because N , V , and T scale isometrically, thus we can substitute a single scaling exponent to reflect the scaling in terms of N :

$$\log_{10}(E[t_{\text{fin}}]) = C_1 + \mu \log_{10}(N)$$

We use linear regression on log-transformed simulation data to determine the value of the scaling exponent μ . This provides the final scaling relationship

between t_{fin} and N for case 0.

$$E[t_{\text{fin}}] = 10^{C_1 + \mu \log_{10}(N)}$$

We fit the scaling constant C_1 to the simulated data.

Similarly, for Case 1 where there is a single target ($\log_{10}(1) = 0$), N increases and V is constant ($i \log_{10}(V) = C_2$), we simplify eq. (4.18) as,

$$\begin{aligned} \log_{10}(E[t_{\text{fin}}]) &= C_1 + C_2 - j \log_{10}(N) - 0 \\ &= C_{1'} - j \log_{10}(N) \end{aligned}$$

where $C_{1'} = C_1 + C_2$. We again use linear regression on log-transformed simulation data to determine the value of the scaling exponent j . We use the same procedure to estimate scaling exponents for cases 2 and 3 for t_{fin} ; and for t_{∞} for Cases 1-3.

4.5 Results: Comparing model predictions to simulated data

The t_{fin} approach to estimating the initial first contact times was developed for case 0 in which a uniform random distribution of searchers and targets are placed in a volume with constant density so that N and T increase linearly with V . The key assumption in t_{fin} is that initial first contact times (the

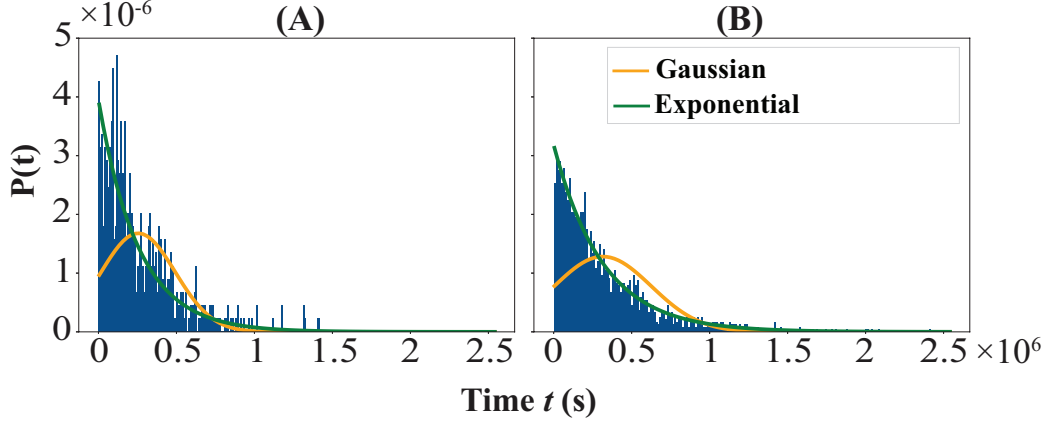


Figure 4.3: **Distribution of Contact Times of All Searchers within One Volume for Case 0.** We show the distribution of the first contact times in two (out of 30) simulations of case 0. Panel A shows the distribution of 629 first contacts for 629 searchers with 2533 targets in a 921 mm³ search volume. Panel B shows the distribution of 1989 first contact times for 1989 searchers with 8011 targets in a 3640 mm³ search volume. Data are plotted as a histogram of probabilities of the initial first contact time of each searcher on the y-axis for the contact time given on the x-axis. The theoretical probability density functions (PDF) of a Gaussian distribution and an exponential distribution for the same mean and standard deviation as the data are shown for comparison. The data are visually consistent with an exponential distribution. The result of the goodness of fit tests (SSE and AIC) in Table 4.2 are consistent with our assumption of the exponential distribution of the first contact times.

first time any one of the searchers comes into contact with a target) are exponentially distributed within each simulation. This allows us to predict that the first time the *first* searcher in a given simulation will contact a target is $\frac{1}{N}$. We found that in all case 0 simulations, an exponential distribution was a good fit. We compare to a Gaussian to demonstrate the difference in sum of squares error (SSE) and Akaike information criteria (AIC) (Table 4.2).

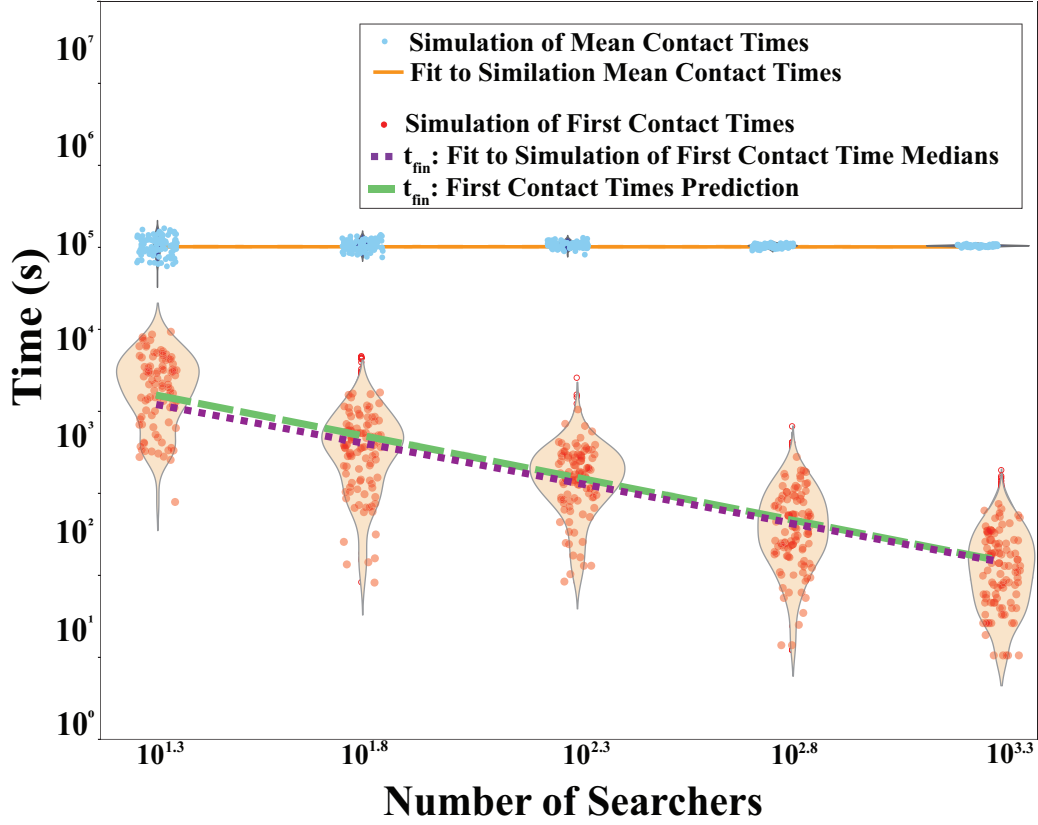


Figure 4.4: t_{fin} Predictions Compared to Simulations Corresponding to Case 0: Constant Density, Multiple Targets. Each red circle represents the time for the first searcher, in a population of N searchers, to contact its target, where N is varied along the x-axis. The initial first contact time for each replicated simulation is shown as red dots. t_{fin} is predicted to scale as N^{-1} (dashed green line, Equation (4.5)). The dotted purple line shows the regression through the simulated data: $E[t_{\text{fin}}] \propto N^{-1.05}$ with a 95% CI $[-1.1, -0.98]$ for the exponent, consistent with the exponent of -1 predicted by t_{fin} . Variance among the initial first contact times for a given N ranges from 0.25 to 0.36. In contrast, the mean search times (blue dot), averaged over all N searchers in each simulation, across the same values of N , are much larger and do not vary with N , and have extremely low variance, all less than 0.003. 100 replicates are used to calculate the mean first contact time and the initial first contact time for each value of N ; thus there are 100 blue and 100 red points for each value of N .

Table 4.2: Statistical comparisons of the Gaussian and exponential distribution fit to the initial first contact times for case 0. The exponential distribution has lower mean SSE and AIC values than the Gaussian distribution, indicating a better fit to the data. 100 simulations were replicated for each number of searchers. Numbers are given to two significant digits.

		Number of searchers				
		20	63	200	630	2000
Exponential	SSE	$8.1.2 \times 10^{-8}$	4.6×10^{-10}	4.6×10^{-11}	4.6×10^{-12}	7×10^{-13}
	AIC	2600	2800	3000	3200	3400
Gaussian	SSE	1.3×10^{-8}	6.7×10^{-10}	1.2×10^{-10}	3.7×10^{-11}	1.9×10^{-11}
	AIC	2600	3000	3300	3900	4300

In Figure 4.3, we show example probability distributions from two sets of simulations. These and the other simulations show a close fit to an exponential distribution.

Given that the simulation validated the assumption that first contact times are exponentially distributed for case 0, we then test the prediction that $t_{\text{fin}} \propto \frac{1}{N}$ from (Equation (4.5)). In figs. 4.4 to 4.7 we compare model predictions with regression fits for the simulation data. The predicted scaling relationship with respect to N or V are depicted by adjusting the vertical intercept through a constant fit. For t_{fin} , the scaling constant was chosen to align the predicted line with the data corresponding to the smallest N , and for t_{∞} , the constant was selected to match the largest simulated N . Consequently, these figures are intended to demonstrate how the simulation results conform to the model predictions regarding the systematic variation in initial first contact times as a function of N or V . The fitting procedure is

detailed in Section 4.4.5.

Figure 4.4 shows a very close correspondence between the t_{fin} prediction and the simulated data. We also show that initial first contact times are orders of magnitude faster than the mean first contact times. The mean search time is (unsurprisingly) unaffected by N , and it is far less variable. This demonstrates that our model for t_{fin} accurately predicts $\frac{1}{N}$ scaling for its intended use case, and that initial first contact times are substantially faster than mean contact times. We now test whether t_{fin} can be extended to 3 search scenarios with a single target and all searchers placed at the same initial location. We compare t_{fin} predictions to t_{∞} predictions that assume an infinite number of searchers. For case 1 we compare t_{fin} and t_{∞} predictions given a single target in a constant search volume while increasing N across simulations. This was the scenario t_{∞} was designed for, assuming N approaches infinity. Figure 4.5 shows that for small $N < 10^3$ initial first contact times decrease approximately linearly with N , consistent with our t_{fin} prediction that initial first contact times are proportional to $\frac{1}{N}$; for large $N \geq 10^3$, simulations are approximately consistent with the t_{∞} prediction, proportional to $\frac{1}{\ln(N)}$ (Equation (4.13)). Thus, even in this search scenario that differs from the original scenario envisioned by the t_{fin} approach, t_{fin} approximates initial first contact times for biologically relevant numbers of searchers, while t_{∞} better approximates initial first contact times for larger N .

For case 2 we compare t_{fin} and t_{∞} predictions to simulations where N is

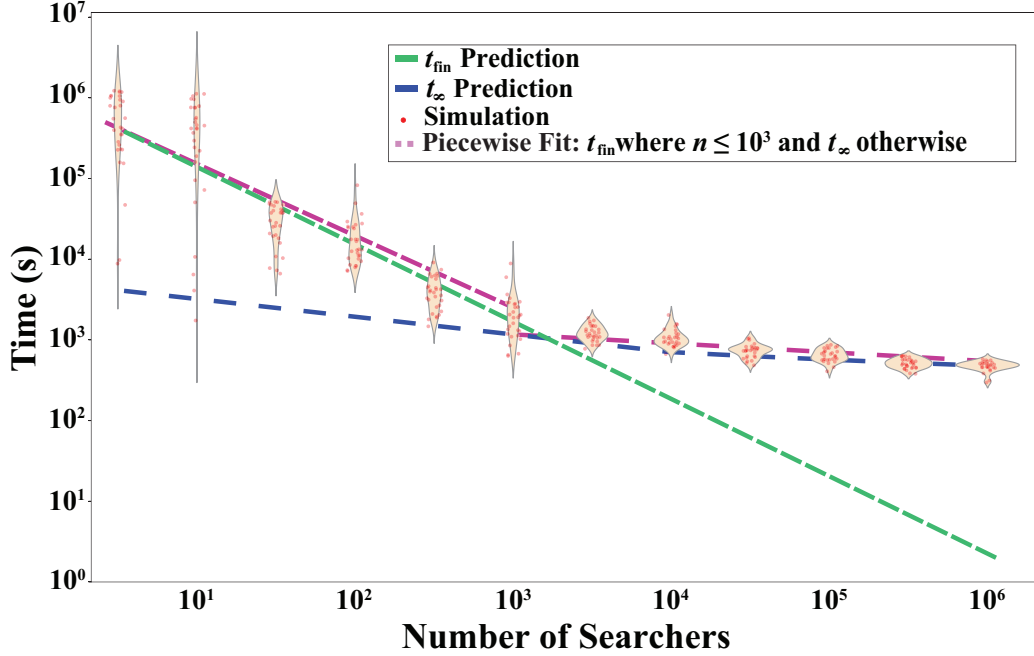


Figure 4.5: t_{fin} and t_{∞} Predictions Compared to Simulations Corresponding to Case 1: Single Target, Constant Volume, Increasing N . Each red circle represents the time for the first searcher, in a population of N searchers, to contact its target, where N is varied along the x-axis. 30 simulations were run for each value of N , and the first contact in each simulation is shown. The pink dotted line connects the medians of the simulated data. The green line shows the t_{fin} prediction (Equation (4.7)) and the blue line shows the t_{∞} prediction (Equation (4.13)), both fitted to the simulated data. The regression through the simulated first contact points $N < 10^3$ is $N^{-1.1}$ with 95% CI $[-0.86, -1.3]$ consistent with the t_{fin} prediction for small N . The regression through data (pink line) where $N \geq 10^3$ gives $t_{\infty} \propto \log(N)^{-1.94}$ with a 95% CI $[-2.1, -1.8]$ which is slightly steeper than the $1/\log(N)$ prediction of t_{∞} . Prediction lines are shown with scaling constants set to fit the median of either the smallest (for t_{fin}) or largest (for t_{∞}) simulated data.

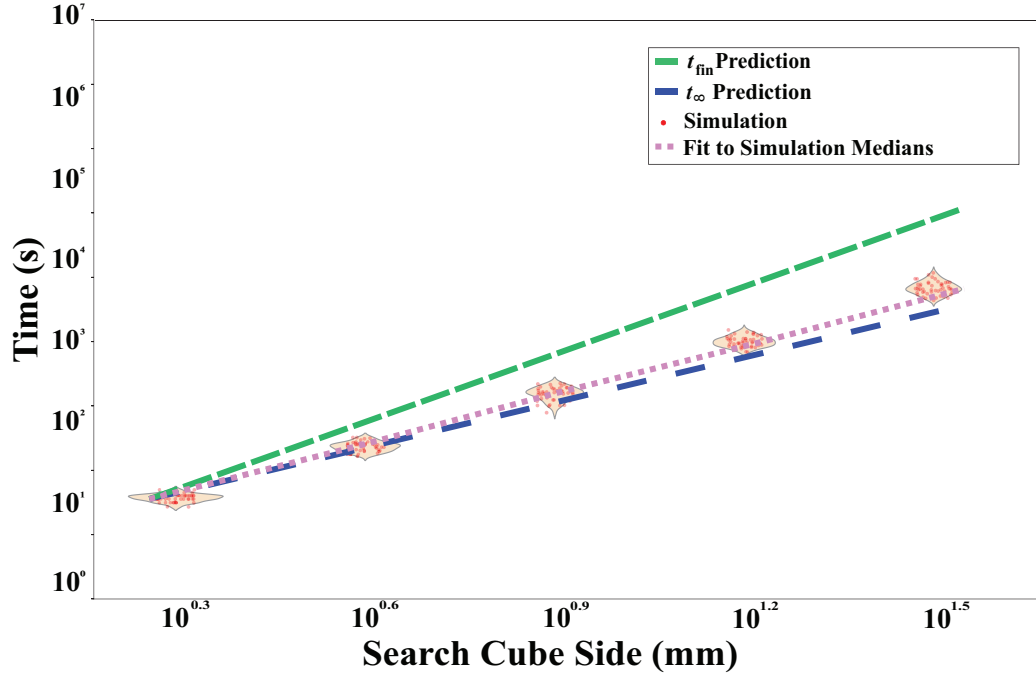


Figure 4.6: t_{fin} and t_{∞} Predictions Compared to Simulations Corresponding to Case 2: Single Target, Increasing Volume, Constant N . Symbols are replicated from Figure 5. The green line shows the $t_{\text{fin}} \propto L^3$ prediction and the blue line shows the $t_{\infty} \propto L^2$ prediction. The regression through the simulated data is $t \propto L^{2.17}$ with 95% CI [2.14, 2.2], close to the t_{∞} prediction.

fixed and V increases. Figure 4.6 shows that the simulations are very close to the t_∞ prediction that time is proportional to L^2 , and considerably less steep than the t_{fin} prediction which is proportional to L^3 . Similarly, the simulation tests of case 3 in which both N and V increase are considerably closer to the t_∞ prediction ($L^2/\ln(N)$) than the constant t_{fin} prediction.

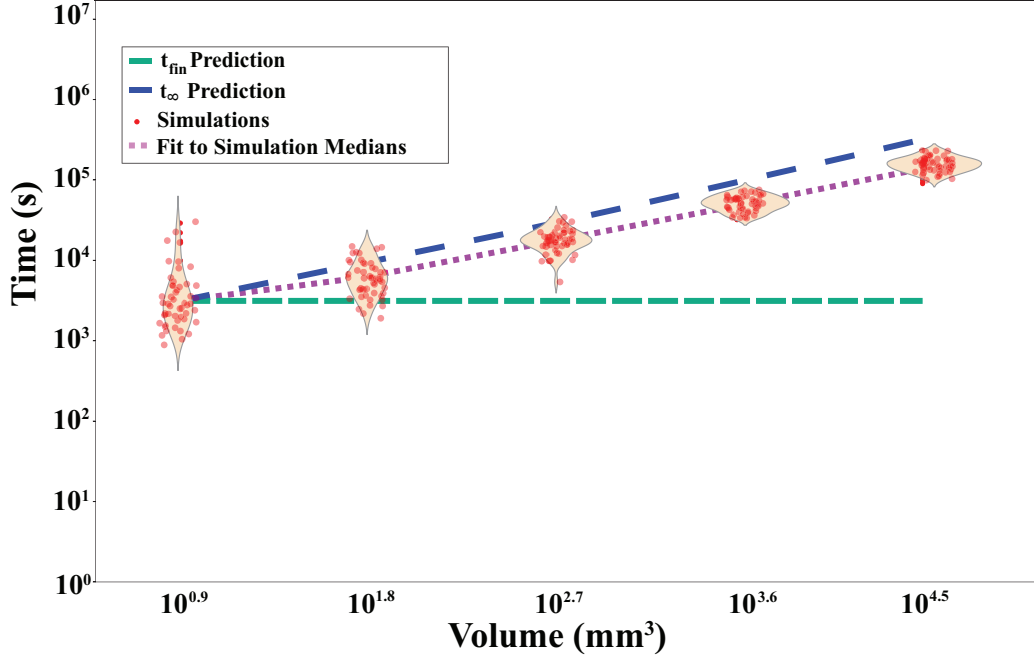


Figure 4.7: t_{fin} and t_∞ Predictions Compared to Simulations Corresponding to Case 3: Single Target, Increasing Volume, Increasing N . Symbols are replicated from Figure 5. The green line shows the $t_{\text{fin}} \propto 1$ prediction and the blue line shows the $t_\infty \propto L^2/\ln(N)$ prediction.

4.6 Discussion

4.6.1 Summary

We developed an analytical model, t_{fin} , designed to predict how quickly the first of many searchers finds its first target, given that searchers and targets are distributed randomly in a volume. Simulations validate the key t_{fin} assumption that search times are exponentially distributed (Figure 4.3) and the key prediction that initial first contacts scale as $\frac{1}{N}$ (Figure 4.4) for case 0 with a uniform random distribution of searchers and targets that are kept at constant density as volume increases. In order to compare t_{fin} to another model of initial first contact times (t_{∞}), we conduct a set of simulations where all searchers start at the same location and search for a single target. We find, for case 1 (where the distance between searchers and target is fixed), the t_{fin} $\frac{1}{N}$ prediction is consistent with simulations for small N , and the t_{∞} $\frac{1}{\ln(N)}$ prediction is close to simulations for large N (Figure 4.5). In cases 2 and 3, we systematically increase the distance between the initial placement of searchers and the target. The initial first contact times are close to t_{∞} predictions and quite far from t_{fin} predictions (Figure 4.6 and Figure 4.7). Our analysis of t_{fin} in case 0 suggests that the fastest searcher is one that is lucky in two respects: it happens to be placed near a target, and it takes a relatively direct path to the target. Because t_{fin} predictions do not hold in cases 2 and 3 when distance increases, this suggests that $1/N$ scaling only

holds when distances between searchers and targets are fixed. We suggest that the existence of lucky first searchers is an important benefit of collective search, particularly when the success of the first searcher causes important downstream events.

4.6.2 Interpretation of scenarios

In this paper we simulated idealized searches that correspond to the volume of lymph nodes across mammal sizes with estimates of the numbers of T cell searchers and dendritic cell targets within those lymph nodes, where the first T cell that find a dendritic cell presenting cognate antigen initiates the adaptive immune response motivated by ongoing work [49]. However, we suggest that the initial first contact times are relevant to a broader range of biological phenomena.

To better characterize the circumstances under which t_{fin} makes valid predictions, we compare our analytical predictions, to the similar Extreme First Passage Time models developed by physicists to describe the time for a population of searchers to find a single target. That theory (that we call t_{∞}) employs the useful simplifying assumptions of an infinite number of searchers diffusing in one dimension to find a single target. Specifically, we compare the t_{fin} predictions to t_{∞} predictions extended by Lawley [91] to predict the initial first contact times in three dimensions. While t_{fin} predicts that larger populations of searchers find their targets proportionally faster, t_{∞} predicts a much slower logarithmic speedup in search times as N increases.

We compared t_∞ predictions to those of t_{fin} under three experimental cases. In case 1, when we hold the size of the search volume constant (and consequently the distance between the initial location of the searchers and the target are held constant), we find that simulated initial first contact times are consistent with the linear decrease with N predicted by t_{fin} when N is less than 1000; for $N > 1000$ simulations are consistent with the t_∞ logarithmic prediction (Figure 4.5). This result suggests that when the only variation across experiments is N , the t_{fin} predictions are reasonable approximations for biologically relevant numbers of searchers. We note that in this case the key assumption holds that search times are exponentially distributed among the searchers within each simulation. However, there is a crossover, such that t_∞ applies when N approaches infinity. Once N is sufficiently large there is only a logarithmic benefit to increasing N . We expect that the values of N for which the $\frac{1}{N}$ scaling regime holds depends on the details of how far apart the searchers are placed from the target, the behavior of searchers at the boundaries, and the dimension of the system (i.e. whether searchers are in 2D or 3D). Some of these issues are explored in [131] which also found a transition from $\frac{1}{N}$ to $\frac{1}{\ln(N)}$ scaling although they presented the transition in more general dimensionless terms rather than for a particular number of searchers. We leave additional analysis to future work and suggest that interdisciplinary approaches among physicists, biologists, and computer scientists are needed to further understand which factors most influence initial first contact times in different scenarios with biologically relevant numbers of searchers.

In cases 2 and 3 when we simulate increasing distances between the target and initial searcher placements, t_{fin} predictions are not close to simulated results (Figure 4.6 and Figure 4.7). We predict that t_{fin} doesn't hold in these cases because the most important factor in these cases is that the distance between searchers and targets increases across simulations. This sheds light on an important feature of case 0, where searchers and targets are distributed at random and with constant density. In case 0, the distances between the initial placement of searchers and targets are determined by a random process, and therefore, when there are more searchers and targets in a larger volume, the *closest* distance between a searcher and the nearest target will decrease. Thus, the lucky first searcher will be one that happens to be very close to its target as well as one that happens to take a relatively direct path to that nearby target. The t_{∞} predictions developed by physicists only consider the directness of the path because that approach assumes a fixed distance.

The $\frac{1}{N}$ scaling of initial first contact times suggests a substantial and underappreciated advantage of collective search. Case 0 suggests that in a bigger lymph node with 100 times more T cell searchers and dendritic cell targets, T cell activation would occur 100 times faster. Case 1 suggests that a bee colony (or in 2D, an ant colony) with 100 more searchers starting in a single nest and foraging independently in a fixed-size territory, would find a single resource (e.g. a rare patch of flowers or food) 100 times faster. These advantages would not directly accrue to the average searcher, which would not

find a target any faster if it continued to search independently. It is only the lucky first searcher that finds a target faster in a larger population. However, if there is any signal (e.g., waggle dance or pheromone communication) or structural change (T cell replication or changes in influx and efflux to and from lymph nodes) following the first target discovery, then subsequent searchers could modify their search to take advantage of that information. This is consistent with findings from [40] that larger bee colonies found resources faster and overall foraging rates were greater in larger colonies. Similarly, the first lucky mutation in a population or the lucky first receptor to bind in a cellular interaction, means that larger populations of individual agents have a lucky first encounter faster and the whole system can benefit from that lucky first encounter. Thus, the $\frac{1}{N}$ initial first contact time confers a significant advantage to a large population size. However, we note a caveat that real biological systems may deviate in important ways from the simple simulations we used here. For example, the density of cognate T cells may not be constant in lymph nodes, and forager territory may scale non-linearly with the forager number. The transition from search times that scale as $\frac{1}{N}$, to search times that scale with $\frac{1}{\ln(N)}$ is also important for analyzing the costs and benefits of increased population size. In many realistic scenarios, more searchers can lead to a proportionally faster search, but as size increases that benefit may diminish. This suggests that, combined with other tradeoffs, initial first contact times may contribute to an optimal colony size under different conditions. Our findings are relevant not only for

understanding collective search in biology but also for engineering collective searchers in swarm robotics, a field that takes inspiration from collective search in biology [41, 96, 58, 69, 152]. Understanding the transition from linear to logarithmic dependence on the number of searchers and the shift from cubic to squared initial distance between the searchers and the target dependence with increasing search volume points to important constraints on scalability, a key concern in swarm robotics. We suggest that swarm robotic systems can both benefit from an understanding of initial first contact times, and serve as real-world testbeds that can help to develop and refine biologically relevant theories about how initial first contact times depend on numbers, search areas, or volumes and the distribution of searchers and targets in space. Direct biological data to test the initial first contact time predictions are scarce. In other work [49] we show that t_{fin} predictions are consistent with the time to initiate adaptive immunity, but the details of how quickly each individual T cell finds its target are poorly understood. We hope that this paper will encourage biologists to report initial first contact times in empirical studies.

Chapter 5

An Analysis of Scaling of Extreme First Passage Time for Large-Scale Searcher Models

5.1 Abstract

In various fields, such as immunology, chemical kinetics, collective behavior, and complex systems, estimating the first time a searcher among a group of searchers locates a target, popularly known as Extreme First Passage Time (EFPT), is important as it often triggers the series of actions. Motivated by our previous work on T cell-antigen interaction, here we investigate the EFPT in a more generic context for various numbers of searchers ranging from 1 to infinity, starting from the same location in a bounded, three-dimensional

domain. We analyze the scaling behavior of EFPT with varying initial distances between the searchers and the target for a large-scale model. We implement an analytical framework based on extreme value theory and the short-time asymptotics of the single-searcher survival probability and show how the analytical results match our simulation data. We demonstrate that for a small number of searchers, EFPT decreases linearly with an exponent whereas for a large number of searchers, it transitions to a logarithmic decrease. We identify the critical transition point between linear and logarithmic decrease, for varying initial distance and show the scaling of the transition point with the initial distance between searchers and target. This generalizable EFPT model offers estimators for the optimal deployment of searchers for various distance ensembles in biological, chemical, and engineered systems.

5.2 Introduction

In Biology, chemistry, and other complex systems, a key question is how to estimate the contact time between searchers and targets, known as the first passage time (FPT). Most studies focus on either a single searcher [112] or an idealized scenario with an infinite number of diffusive searchers [91], mostly highlighting the mean of the contact times [120]. However, it often is the very first contact time by the first searcher, primarily known as EFPT, that triggers subsequent courses of action. For example, in the immune system, an immune response is initiated when the very first T cell encounters an antigen

[30]; in chemical reactions, the initial collision between reactant molecules can determine the reaction pathway [94]; and in complex systems such as ant colonies, other foragers can be recruited once the first ant finds a food source [102].

Weiss, in his 1983 paper [159], introduced the concept of EFPT, which describes the fastest search time among a pool of searchers [129, 128, 133]. However, analytically finding the EFPT is quite challenging because of its dependency on very rare events. While typical searchers tend to wander in search of a target, the fastest searcher will happen to follow a more direct path, typically the shortest geodesic path to the target [61, 64, 88].

Grebenkov et al. [62, 63] talked about the FPT (the paper referred to it as First Reaction Time (FRT)) in the context of the diffusion process, focusing on the distribution of FPT. The authors argued that the probability distribution of FPT can be divided into multiple regions, with each region being influenced by factors such as the diffusion coefficient, target size, search space radius, and system geometry. However, the authors suggested that their framework may not be applicable in an unbounded or extremely large search space, where the dynamics of FPTs could behave differently from those observed in confined systems.

However, none of these studies tend to find the EFPT covering the full range of possible numbers of searchers, from 1 to infinity. In our previous work [51], we provided a solution for EFPT in the context of immunology where we demonstrated that for a fixed, bounded, reflective search volume,

EFPT scales with $t_{fin} \propto 1/N$ for a limited value of $N (< 10^3)$, and the scaling transitions to $t_{\infty} \propto \frac{1}{\log(N)}$ as N becomes extremely large Figure 5.1. Here N = the number of searchers.

In [51], we studied the search process for a single absorbing target located at the center of a three-dimensional search space with a complete reflecting boundary. Multiple searchers, initially placed at a distance L from the target, move within the search space following Brownian motion, which is used to model the movement of T cells in the immune system. In this work, we first

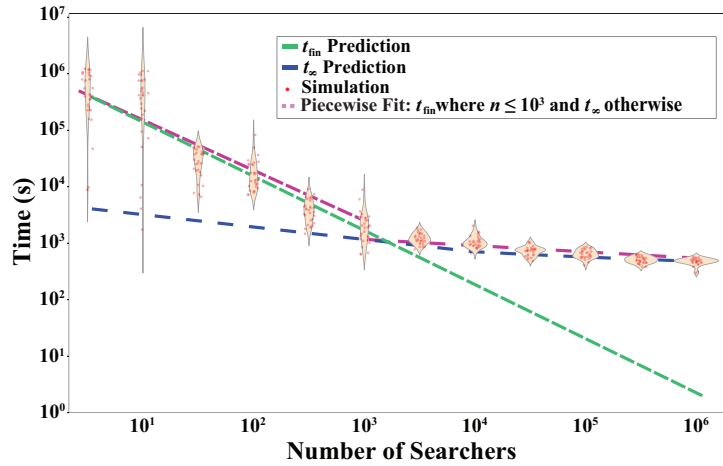


Figure 5.1: **EFPT vs number of searchers within a small lymph node-sized search space.** The searchers mimic random T cell movement, recreated from chapter 4, Figure 4.5. Each red circle represents the time for the first searcher, in a population of N searchers, to contact its target, where N is varied along the x-axis. 30 simulations were run for each value of N , and the first contact in each simulation is shown. The pink dotted line connects the medians of the simulated data. The green line shows the t_{fin} prediction ($\propto \frac{1}{N}$) and the blue line shows the t_{∞} prediction ($\propto \frac{1}{\log(N)}$), both fitted to the simulated data. Prediction lines are shown with scaling constants set to fit the median of either the smallest (for t_{fin}) or largest (for t_{∞}) simulated data.

investigate whether this transition observed in our previous study remains valid for larger initial distances L between the searchers and the target. We extend the findings from an immunological context to a more general model where searchers take one unit step at each time. We confirm the existence of the transition and then aim to identify the critical point at which EFPT shifts between the two scaling behaviors. Given the computational challenges of determining EFPT for very large N within a large search space in simulation, we develop an analytical approach to identify the transition point. While several previous studies have proposed analytical methods to compute EFPT, many have focused on small initial distances or small target-to-search space radius ratios [88, 62]. We adopt the approach outlined by Lawley in [88] to develop the analytical model for large-scale models.

In this paper, we apply extreme value theory to find an EFPT approximation $E(\tau)$ for any number of searchers in a specific initial distance between searchers and target. This approximation can be easily applied in many diverse scenarios, as it depends on only a few properties of the survival probability of a single FPT. Survival probability is the probability that the searcher, starting at $L \geq 0$ distance from the target, hasn't reached the target up to time τ_1 .

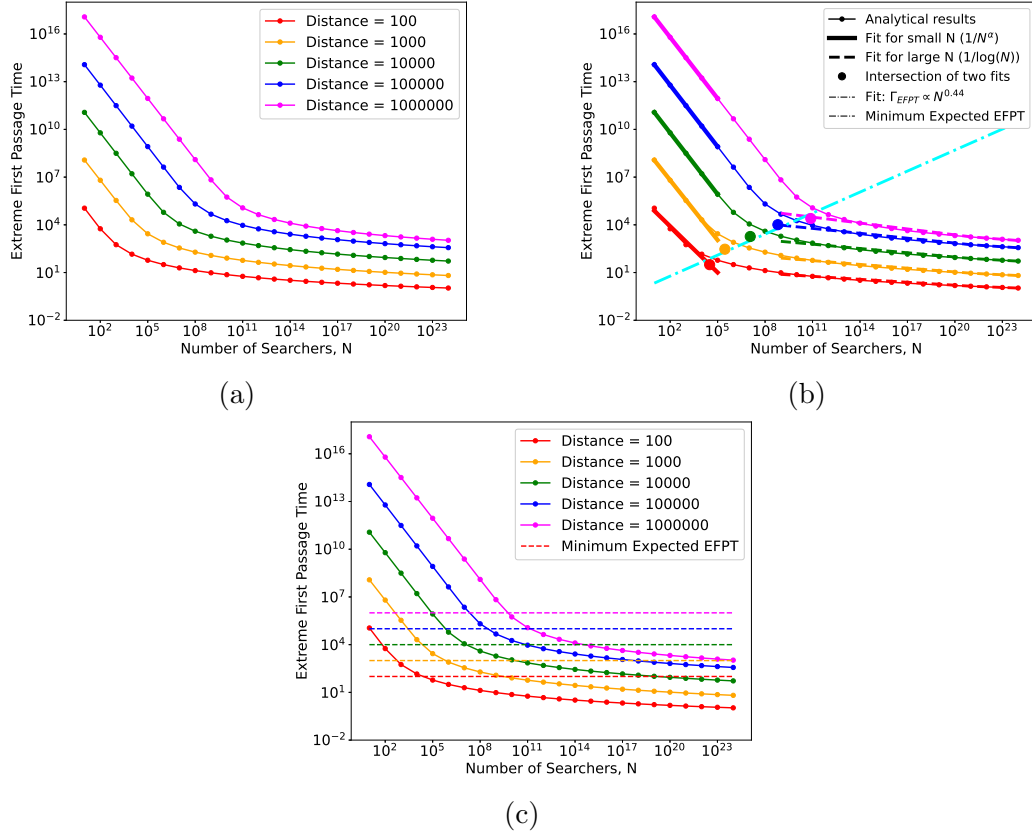


Figure 5.2: **EFPT over the number of searchers N for varying initial distance L between searchers and target.** Both axes are in logarithmic scale. (a) shows the analytical result of EFPT for varying numbers of searchers N and distance L . For any given N , EFPT increases with L . For each fixed L , EFPT is highest when N is very small. EFPT decreases sharply as N increases and almost saturates for large N . (b) shows the fitted curve of the two hypotheses (shadowed: $\frac{1}{N^\alpha}$ and dotted: $\frac{1}{\log N}$) and their transition points (Γ_N, Γ_{EFPT}). (c) shows the predicted minimum EFPT when the searchers follow the Brownian motion with 1 unit step at each time step.

5.3 Results

Figure 5.2 panel (a) shows how EFPT scales over N for various initial distances L . As L increases, the EFPT for any fixed N also increases. Within each curve (fixed L), EFPT is largest when N is very small but decreases substantially as N grows. Eventually, beyond a threshold in N , the EFPT levels off, indicating that the system has reached a saturation point once N is sufficiently large (Figure 5.2 panel (b)). This saturation indicates that adding additional searchers does not significantly lower the EFPT.

However, upon analyzing the data, we observed that for small values of N , the scaling is better described by a modified power-law relationship.

$$E[\tau] = \frac{1}{N^\alpha} \quad (5.1)$$

where the exponent α is not exactly equal to 1. Our initial hypothesis was that EFPT scales inversely with N for small values of N ,

$$E[\tau] = \frac{1}{N}$$

A linear regression shows that α increases with L , implying a steeper decrease in EFPT as N grows when L is large (Figure 5.2 panel (b), shaded solid line). This suggests the benefit of adding additional searchers when the number of searchers is limited is more pronounced at greater distances. When N becomes very large, the EFPT scales asymptotically as $\frac{1}{\log(N)}$ (Figure 5.2,

panel (b) dotted line). This analytical result is consistent with our previously published simulation results. Figure 5.2 panel (b) also shows the transition points Γ_N, Γ_{EFPT} from $\frac{1}{N^\alpha}$ hypothesis to $\frac{1}{\log N}$ one.

Figure 5.2 panel (c) shows the minimum cutoff EFPT for searchers that are following the Brownian motion with 1 unit step at each time step.

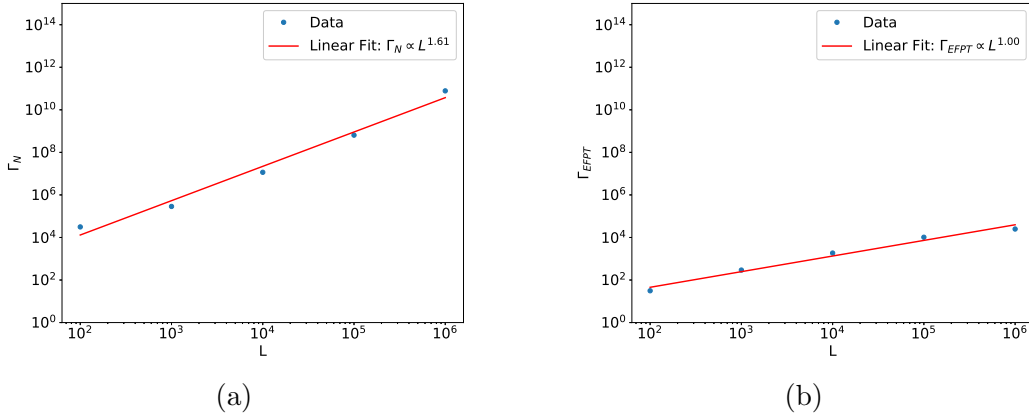


Figure 5.3: **Transition points of the number of searchers Γ_N and the Γ_{EFPT} over initial distance L .** Both axes are displayed on a logarithmic scale. The blue points represent the numerical data, and the red line is a best-fit curve. Panel (a) shows how Γ_N scales with L . The linear best-fit curve suggests that Γ_N **increases** with L . That is Γ_N increases about 3-fold when L doubles. Panel (b) shows how Γ_{EFPT} scales with L . The best-fit curve indicates that the transition in Γ_{EFPT} grows with L ; that is, when L doubles, Γ_{EFPT} also doubles.

We find the transition point $(\Gamma_N, \Gamma_{EFPT})$ between these two hypotheses $(\frac{1}{N^\alpha}$ and $\frac{1}{\log(N)})$ (See Figure 5.2, panel (b), large dots). Our analysis suggests that the transition point Γ_N scales as linearly positive with L , meaning that with the increase of L , the transition point Γ_N will occur for larger number of searchers (Figure 5.3, panel (a)). In other words, the saturation of EFPT

will occur for a large number of searchers in larger L . The transition point for EFPT Γ_{EFPT} scales linearly positive with L , meaning as L increases, the transition point for EFPT of these two hypotheses increases as well (Figure 5.3, panel(b)). These two scaling figures suggest that in larger initial distance, the saturation occurs late and for a relatively larger value of N . That is, as we increase the initial distance, the saturation of EFPT (that is, it can not be any significantly faster) will occur slowly, and to reach that point N will be relatively larger than for a smaller initial distance. Therefore for a smaller initial distance $L1$, if transition occurs at $EFPT_{L1}$ at the number of searchers $N1$, then in larger initial distance $L2(L1 < L2)$, the transition at $EFPT_{L2}$ will occur slower ($EFPT_{L1} < EFPT_{L2}$) at the number of searchers $N2$ where $N2 > N1$.

For a constant one unit step length motion Δs , one would expect that the minimum EFPT would be the same as L ,

$$T_{min} = \frac{L}{\Delta s}$$

shown in Figure 5.2 panel (c) (dotted line). But our analytical result keeps dropping with the increasing number of searchers; it can even predict times below T_{min} .

In classical first passage time theory, each searcher follows a Brownian motion with no hard speed limit, so that its motion is characterized solely by a diffusion coefficient D .

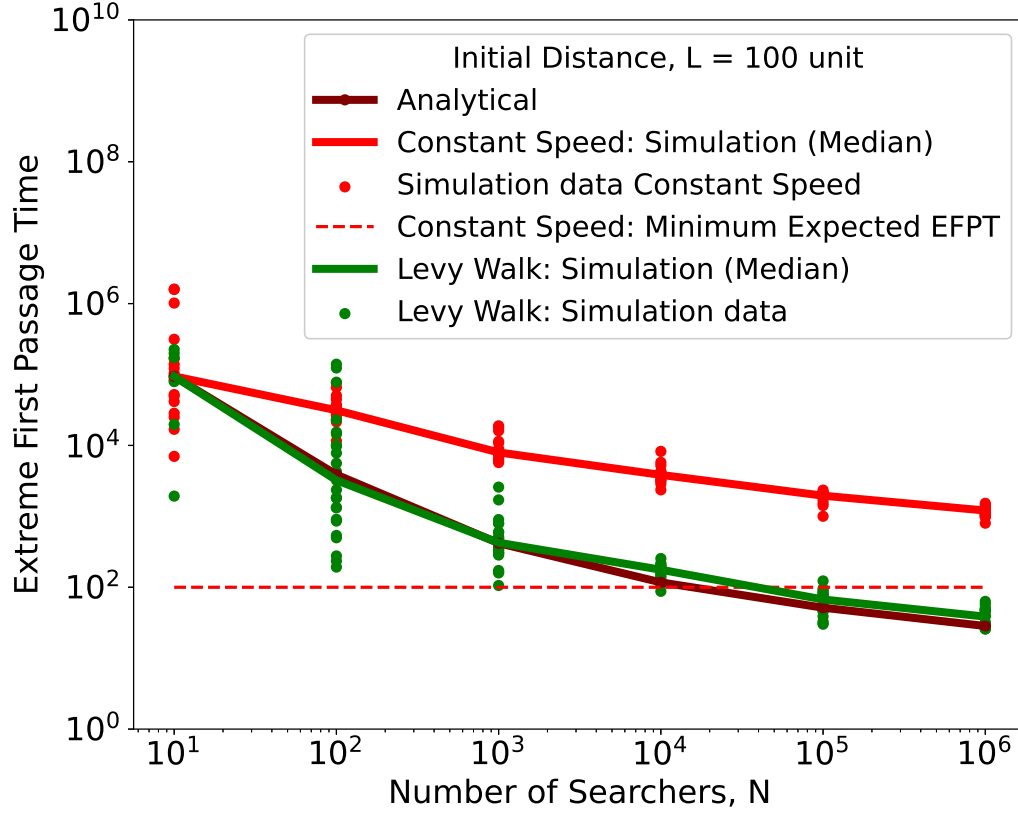


Figure 5.4: **Simulation and analytical EFPT for initial distance $L = 100$ unit over range of number of searchers N .** Both axes are displayed on a logarithmic scale. The maroon circles connected by the thin line show the analytic EFPT prediction. The red line shows the median of the 15 Monte Carlo simulations for hard speed cap $\Delta s = 1$ unit for each N and the green line shows 15 Monte Carlo simulations for diffusion coefficient $D = 1$ for each N . The red and green circles represent the simulation data illustrating the run-to-run variability. The dotted red line shows the theoretical expected minimum EFPT for hard cap step size Δs .

Because true Brownian paths have infinite propagation speed, the diffusion model EFPT prediction continues to decrease for larger N and eventually falls below the ballistic lower limit T_{min} , the time required for a searcher to

traverse distance L in a straight line at maximum speed Δs .

To establish this speed-limit effect, we have simulated two cases for our agent-based model in EFPT. In case one, we enforce a fixed step length $\Delta s = 1$ per time step, so no searcher can arrive in fewer than L steps.

In the second case, we remove the hard speed cap $\Delta s = 1$ by endowing each searcher with pure diffusive motion with diffusion coefficient $D = 1$.

As shown in Figure 5.4 for $L = 100$, when speed is capped by the ballistic limit ($\Delta s = 1$ unit), the simulated EFPT (red line) goes slower than the analytic result (maroon line). For large N , the analytic EFPT prediction curve continues its $\propto \frac{1}{\log N}$ decline and eventually undercuts the ballistic limit L , whereas our simulated EFPT with speed cap $\Delta s = 1$ per time step asymptotes to L . On the other hand, the Lévy walk diffusion model shows that the simulation result can find the target quicker compared to single step simulation, aligning closely to the predicted analytical result. The rare, arbitrarily long displacements of the Lévy walk model allow at least one searcher to traverse the full distance in far fewer than L steps, thereby undercutting the ballistic limit. As N increases, these extreme sampling events become almost certain, driving the EFPT down by orders of magnitude below the deterministic speed-cap case.

In sum, the divergence between speed cap and Lévy walk curves reflects the contrast between (a) a constant-speed model, in which the maximum propagation speed enforces a hard lower bound on first-passage times, and (b) an unconstrained diffusive model, in which variability in step size grows

with N . Future work will extend this comparison to larger values of L , in order to map out more precisely the transition from the ballistic regime to the diffusive asymptote.

5.4 Methods

To understand how the number of searchers N affects the EFPT, consider $N \geq 1$ independent and identical diffusive searchers. Let $\tau_1, \tau_2, \dots, \tau_N$ be their independent and identically distributed (iid) first passage times (FPTs) to reach some target. The first time one of these searchers finds the target is defined as

$$T_N := \min\{\tau_1, \tau_2, \dots, \tau_N\}. \quad (5.2)$$

The survival probability for a single searcher is computed as,

$$S(t) = \mathbb{P}(\tau > t) = 1 - 2\sqrt{\frac{L^2}{\pi D t}} \sum_{j=1}^{\infty} \exp\left(-\frac{(j+0.5)^2 L^2}{D t}\right) \quad (5.3)$$

Assume that $S(t)$ has the short-time behavior,

$$1 - S(t) \sim A t^p \exp\left(-\frac{C}{t}\right), \quad \text{as } t \rightarrow 0^+. \quad (5.4)$$

for some constants $A > 0, C > 0$, and $p \in \mathbb{R}$.

The distribution converges to a Gumbel distribution X as follows,

$$\frac{T_N - b_N}{a_N} \xrightarrow{d} X, \quad \text{as } N \rightarrow \infty, \quad (5.5)$$

where,

$$\begin{cases} a_N = \frac{b_N}{\ln(A N)}, & b_N = \frac{C}{\ln(A N)}, & \text{if } p = 0, \\ a_N = \frac{b_N}{p(1+W)}, & b_N = \frac{C}{pW}, & \text{if } p \neq 0. \end{cases}$$

where $W_0(z)$ denotes the principal branch of the Lambert W function and $W_{-1}(z)$ denotes the lower branch of the Lambert W function [33].

Following [88] EFPT is given by

$$E[T_N] = b_N - \gamma a_N + o(a_N) \quad (5.6)$$

where $\gamma = 0.5772$ is the Euler–Mascheroni constant. Here $f(N) = o(a_N)$ means $\lim_{N \rightarrow \infty} a_N f(N) = 0$.

We account for discretization effects in our agent-based simulation by approximating the higher-order term $o(a_N)$ as

$$o(a_N) = \frac{L}{\Delta s} \times a_N \quad (5.7)$$

where Δs is fixed step size per unit time. The resulting expression maintains consistent time units.

In the constant speed brownian walk simulation, we consider $\Delta s = 1$ unit per step. In the Lévy walk simulation, each cell draws the step length δ at every time step from a power-law distribution [163, 156] as follows:

$$p(\delta) = (\mu - 1)l_{min}^{\mu-1}\delta^{-\mu} \quad \delta \geq l_{min} \quad (5.8)$$

Here l_{min} sets the minimum step length and μ controls the heaviness of the tail. In our implementation, we choose $l_{min} = 0.1$ unit and $\mu = 3$ so that $\langle \delta^2 \rangle = 6$ which yield the diffusion coefficient $D = \frac{\langle \delta^2 \rangle}{6} = 1$ in three dimensions.

5.5 Discussion

In this work, we have developed a unified analytical framework for the Extreme First Passage Time (EFPT) of N independent diffusive searchers in a bounded three-dimensional domain. By combining extreme-value theory with the short-time asymptotics of the single-searcher survival probability, we derived closed-form approximations that capture three distinct regimes of EFPT scaling:

- a modified power-law: $E[\tau] \propto \frac{1}{N^\alpha}$ for small to moderate N
- a logarithmic decay $E[\tau] \propto \frac{1}{\log(N)}$ for $N \rightarrow \infty$ in the continuum limit, and
- an absolute saturation of EFPT for larger values of N based on two different motions:

- Under fixed-step Brownian motion ($\Delta s = 1$ unit), EFPT cannot fall below the ballistic traversal time and hence $E[\tau] \xrightarrow{N \rightarrow \infty} L$
- In contrast, in the Lévy walk simulation, each searcher’s step length is drawn from a heavy-tailed distribution. This allows at least one searcher to have a rare, extremely long jump in a single step. Therefore, for large values of N , the EFPT continues to decrease and can even fall below the ballistic limit ($\propto L$).

We identified the transition point Γ_N at which EFPT shifts from the $N^{-\alpha}$ regime to the saturation (or logarithmic) regime and showed that Γ_N decreases with increasing initial distance L , whereas the EFPT at that transition, Γ_{EFPT} , grows linearly with L . These counter-intuitive scalings arise because a finite step length imposes a hard lower bound L on the fastest-searcher time. To confirm and calibrate these analytical predictions, it is essential to carry out targeted simulations:

- Validation for range of N and L : Run Monte Carlo agent-based simulations for a range of N and L to verify the exponent α and the onset of saturation.
- Robustness to Simulation Parameters: Test sensitivity to time-step size Δs dimensionality, and domain geometry, along with Lévy walk variability.

Looking ahead, several directions can extend this work:

- Heterogeneous Environments: Incorporate spatial obstacles or variable diffusivity to model more realistic search landscapes.
- Multiple Targets & Cooperative Search: Generalize to scenarios with multiple absorbing targets or interacting searchers.
- Immune-System Applications: Integrate EFPT theory into detailed agent-based immune models to predict how lymph node architecture and cell motility shape early antigen detection times.

By coupling our analytical insights with comprehensive simulation validation and by broadening the theoretical scope, this research lays the groundwork for optimizing search strategies across physics, chemistry, biology, and complex systems.

Chapter 6

Investigating Vaccine Efficacy Against SARS-CoV-2 Variants: A Model-Based Analysis

6.1 Publication Notes

Authors: Jannatul Ferdous, Jasmine Kreig, Ruian Ke, Ruy Ribeiro

6.2 Abstract

Affinity maturation—the process by which produced antibodies increase in affinity for antigen—occurs during the course of an immune response. Re-

peated exposures to the same antigen will produce antibodies of successively greater affinities. However, as antigen move away in antigenic distance from the initial strain (antigenic drift), the ability of the body to cross-reactively neutralize the antigen decreases. This issue has been well documented in cases of influenza and there is a concern it is occurring in SARS-CoV-2 given successive variants of concern (VOC). Such VOCs would be less susceptible to any immune protection gained from vaccination and prior infection. We modeled adaptive immunity using an agent-based model (ABM) that considers B cells (naïve, plasma, memory), antibodies, and antigens to investigate the varying efficacy of vaccines against SARS-CoV-2 variants. We represent receptor (B cells, antibodies) and epitope (antigens) proteins in Euclidean shape space, simulating binding between these agents based on Hamming distance. We also consider the formation of immune complexes—free antibodies bound to antigen which limits the antigen’s ability to infect more cells. Our findings aim to elucidate observed differences in vaccine efficacy between first-time and repeated SARS-CoV-2 variants, providing insights into vaccination strategies for antigenically variable pathogens necessitating repeated vaccination efforts.

6.3 Introduction

SARS-CoV-2 has gained significant attention since its occurrence in late 2019. According to CDC [55] approximately 1.2 million deaths have been reported in the United States. Although researchers continuous quest for information on

SARS-CoV-2 has been tireless and consummate, the comprehensive modeling of SARS-CoV-2 remains challenging due in part to its high mutation nature. The virus’s rapid mutation often leads to the emergence of new variants, rendering the existing vaccine less effective against the new variant.

On top of that, the complex behavior of the humoral immune response makes it even more complicated to model, especially covering all the variants of SARS-CoV-2. More specifically, the ability of somatic hypermutation to increase antibody affinity over time [47, 48, 105] in response to repeated exposure to SARS-CoV-2 is nonetheless insufficient to neutralize all emerging variants, given the virus’s rapid mutation rate.

[17] has discussed a basic and generic computational model based on clonal selection theory in adult animals. The study shows how a population of lymphocytes can expand in response to antigenic stimulation, predicting dose–response curves under various generic conditions—but without specifying a particular antigen or detailing how parameter values were chosen. In contrast, our model focuses explicitly on SARS-CoV-2 epitopes, grounding parameter selection in experimentally measured binding affinities and mutation rates to capture realistic antigen-antibody dynamics.

Hybrid immunity refers to the combined immunity from a prior infection and vaccination and has been shown to cause highly effective B cell and T cell responses capable of neutralizing diverse SARS-CoV-2 variants [132]. However, the study doesn’t define the extent of the diverse SARS-CoV-2 variants that the hybrid immunity can neutralize. The experiments were

also relatively short-term (3–18 months), making us wonder how the immune imprinting evolves over a longer period of time. The also doesn’t talk about how the immunity adapts to the SARS-CoV-2 mutations that escaped the previously imprinted responses. By embedding successive antigen exposures into our agent-based simulation, we quantify the durability of hybrid-imprinted responses over long time and map the evolving neutralization landscape across emerging variants.

Kepler et al.’s optimal-control analysis of somatic hypermutation rates revealed that, in a simplified single-compartment setting, B cells may employ phasic mutation schedules—periodically re-entering germinal-center-like compartments to maximize affinity maturation [83]. Park and colleagues found that although hybrid immunity confers the highest neutralization titers against a broad panel of SARS-CoV-2 variants, highly divergent strains (e.g., Omicron BA.1) still partially escape, and cross-reactivity wanes over months [117]. Our work bridges this gap by explicitly modeling antigenic distance and variant mutational trajectories: we simulate how successive exposure to evolving spike proteins reshapes the EFPT of high-affinity B-cell clones, thereby predicting which future variants might evade existing imprinting and when booster strategies will be most effective.

In this paper, we investigate the population dynamics governing B cells, antibodies, and antigens within an immunological context for different variants of SARS-CoV-2 using a computational model. This model is motivated by Derek Smith et. al. [147] on antigenic distance on the influenza virus where

the authors considered a string of 20 symbols representing the receptors of B cells, antibodies, and antigens. The antigenic distance between two antigens is measured by the Hamming distance of their strings. The same concept calculates the antigenic distance between an antibody (or B cell) and an antigen. 0 antigenic distance means the best match or the highest affinity, and the higher the distance the less the affinity.

Our model addresses the concept of immune imprinting of SARS-CoV-2, where the immune system prioritizes responding with antibodies specific to the first antigen. It also investigates how the pre-existing antibodies effectively react to neutralize the repeated exposures to different variants of the antigens of high and low affinity.

The main application of our model is to understand the immune imprinting in SARS-CoV-2 variants of different antigenic distances. The model shows the effect of immune imprinting on the immune response to the changing variant of SARS-CoV-2. It studies why different variants may have different degrees of protection due to pre-existing immunity. The model can predict the cross-reactivity of antibodies from primary SARS-CoV-2 infection to different SARS-CoV-2 variants by integrating antigenic distance. Our model studies the immune response trends as antigenic variation scales up, providing insights into vaccine design and immune resilience against evolving pathogens. The model can be used to design vaccines through antigenic combinations that promote long-term immune protection by targeting a large number of SARS-CoV-2 variants.

6.4 Results

In this paper, we evaluate three cases based on the antigenic distance or affinity between a primary antigen (SARS-CoV-2 614G) and a secondary variant. We consider a ball of stimulation around the antigen in an antigenic map to show the extent of affinity for B cells and antibodies that can bind to the antigen. In each case, we depict the extent of overlap of their balls of stimulation in an antigenic space Figure 6.1 and discuss the resulting cross-reactivity of B cells, antibodies and antigens.

- Case 1: The antigenic distance between the two variants is small, resulting in an overlap between the balls of stimulation (Figure 6.1a). We choose SARS-CoV-2 variant 614G as the primary antigen and Alpha as the secondary antigen for this case. The antigenic distance between these two variants is 0.81 Antigenic Unit (AU). The overlap allows the high cross-reactivity of the B cells and antibody from the primary antigen towards the secondary antigen Alpha.
- Case 2: We consider 614G (primary) and Delta AY.4.2 for SARS-CoV-2 variants with antigenic distance 1.56 AU between them for Case 2. The antigenic distance between the two variants is larger compared to case 1, but the ball of stimulation still overlaps (Figure 6.1b). The reduced overlap leads to lower cross-reactivity. As a result, secondary exposure to Delta AY.4.2 elicits a weaker recall response from memory clones

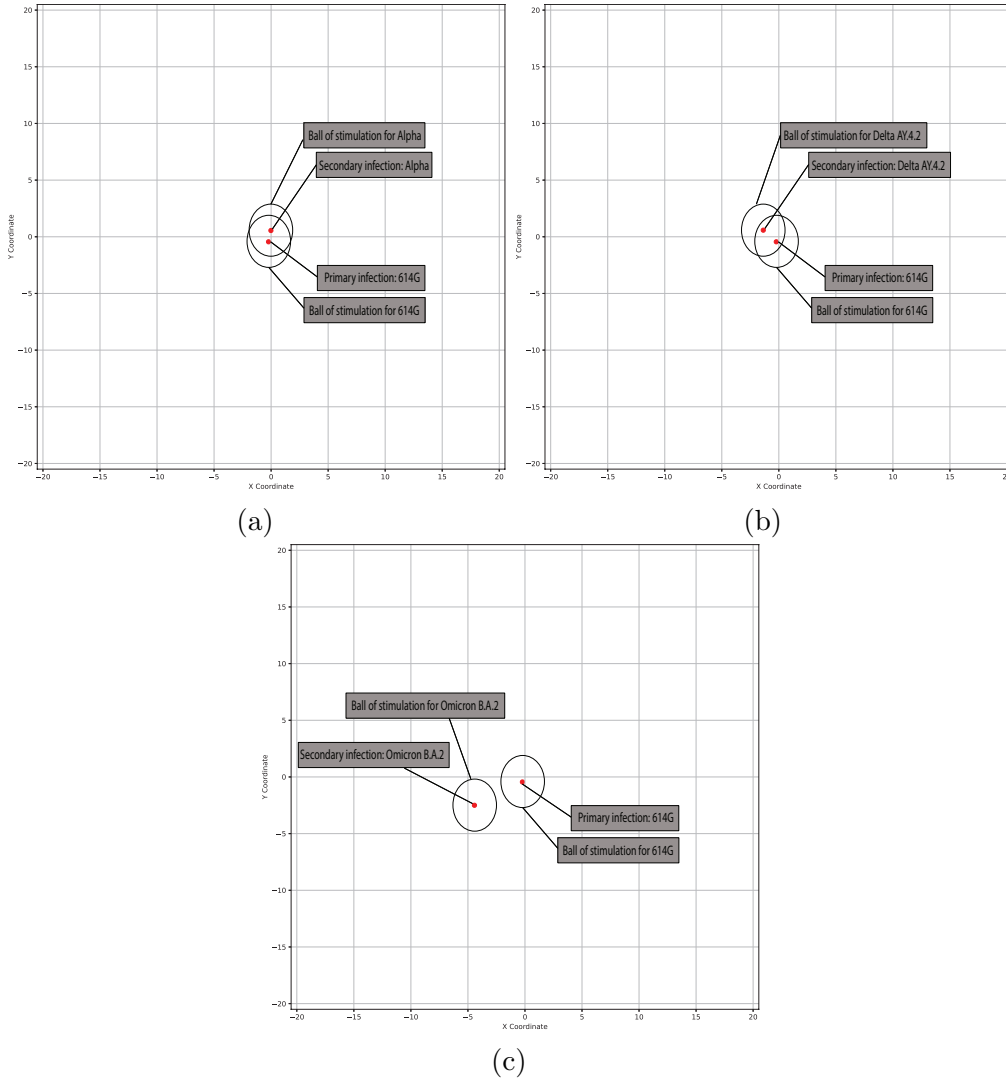


Figure 6.1: Antigenic distance and balls of stimulation between different SARS-CoV-2 variants in a shape space diagram. (a) represents the smaller antigenic distance between 614G and Alpha, resulting in a high degree of overlap between their balls of stimulation. (b) represents a larger antigenic distance compared to panel (a) between 614G and Delta AY.4.2 variants with their slight overlapping ball of stimulation. (c) represents a larger antigenic distance between 614G and Omicron B.A.1 variants with no overlapping of their balls of stimulation. Although the figures show the antigens together, in the model the secondary antigen is introduced at day 40

originally generated against 614G.

- Case 3: The antigenic distance between the two variants is significantly larger resulting in no overlapping between their balls of stimulation (Figure 6.1c). We consider 614G and Omicron B.A.2 for SARS-CoV-2 variants with an antigenic distance of 4.58 AU. In this scenario, pre-existing B cells and antibodies specific to 614G exhibit no binding to Omicron B.A.2 because of the large antigenic distance.

6.4.1 Case 1: Small Antigenic Distance, High Degree of Overlap between Balls of Stimulation

When the antigenic distance between the two variants is small, resulting in an overlap between the balls of stimulation (Figure 6.1a). We choose SARS-CoV-2 variant 614G as the primary antigen and Alpha as the secondary antigen for this case. The antigenic distance between these two variants is 0.81 AU. This overlap enables high cross-reactivity: antibodies and memory B cells elicited by 614G bind effectively to Alpha.

After the primary antigen, 614G, is removed from the system through the formation of immune complexes and clearance of immune complexes and infected cells, the secondary antigen, Alpha, is introduced at day 40. Figure 6.2a shows that alpha is cleared almost immediately. Because Alpha is antigenically not very distant from 614G, the existing antibodies from 614G have a higher affinity with Alpha and effectively neutralize the Alpha variant

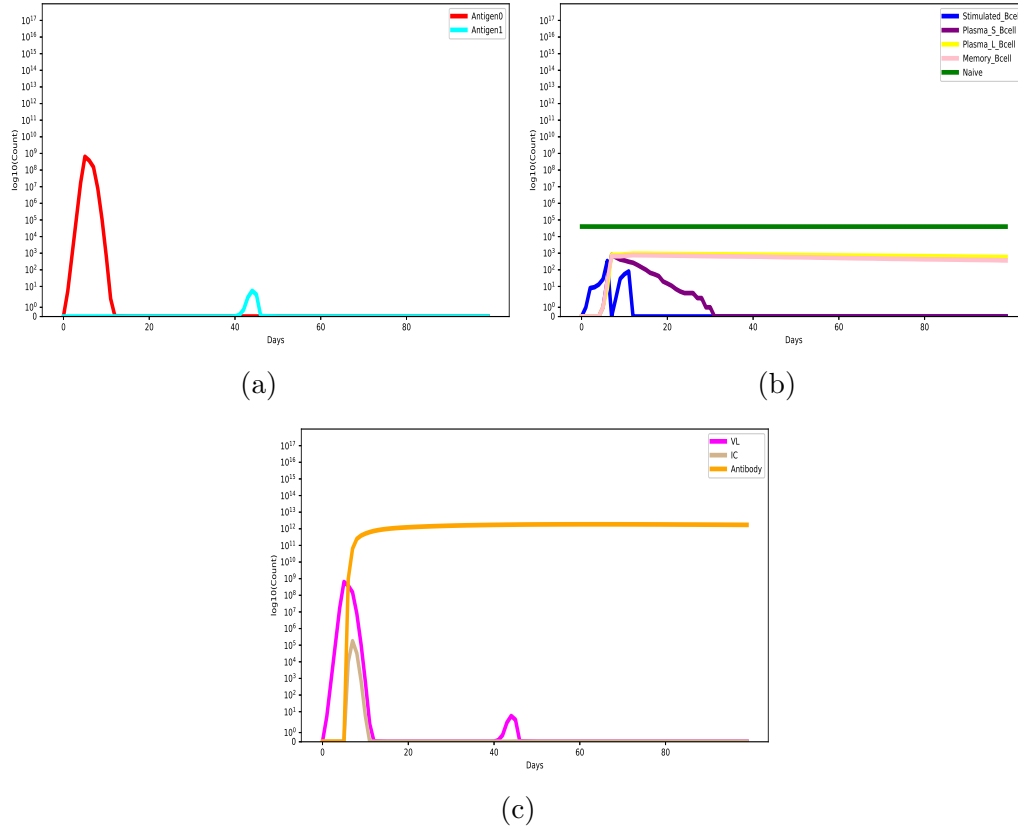


Figure 6.2: **Population dynamics for Case 1** where antigenic distance between primary antigen 614G, and secondary antigen Alpha is **0.81 AU**. This figure shows the simulation time courses following secondary challenge with the Alpha variant at day 40. (a) shows antigen concentration over time, with the secondary antigen (Alpha) introduced at day 40. (b) shows B-cell populations, showing naive, stimulated, plasma, and memory subsets. (c) shows viral load , antibody titer, and immune-complex (IC) formation following 614G and Alpha exposure.

immediately.

In Figure 6.2b, the population of naive B cell doesn't seem to change. That is because out of 40000 naive B cells, only 314 can be activated by the antigens as they are within the ball of stimulation. The remaining naive B cells

stays as they are for being outside of the ball of stimulation, meaning lower affinity towards the antigen. Even among these, stimulation probability scales with individual affinity, so only the highest-affinity clones expand modestly. The stimulated B cells initially increase as the primary antigen grows. Once the antigen is cleared, the stimulated B cells decay based on their age limit. Because Alpha is removed so rapidly, stimulated B cells peak briefly for Alpha and then decay according to their programmed lifespan, yielding negligible net expansion.

Stimulated B cells undergo differentiation after a four day maturation delay, generating long-lived plasma cells, short-lived plasma cells, and memory B cells. Plasma long-lived (Plasma L) cells and memory B cells have a very high longevity; therefore, they persist at steady levels for a long time and again increase after the introduction of the secondary antigen. Plasma short-lived (Plasma S) cells, on the other hand, have comparably lower longevity, resulting in a transient population. Consequently, antibody titers rise sharply following Plasma S and Plasma L production and remain elevated, ensuring a rapid secondary response to an antigenically similar variant Figure 6.2c.

The Immune Complex (IC) start forming once the plasma B cells start producing the antibodies at day 4 and decline as the antigens start to decay. Although the IC for secondary antigen Alpha is showing almost 0, it is because the duration is captured after each day and the IC are rapidly generated and diminish as the pre-existing antibodies react rapidly to the antigens.

6.4.2 Case 2: Moderate Antigenic Distance, Partial Overlap between Balls of Stimulation

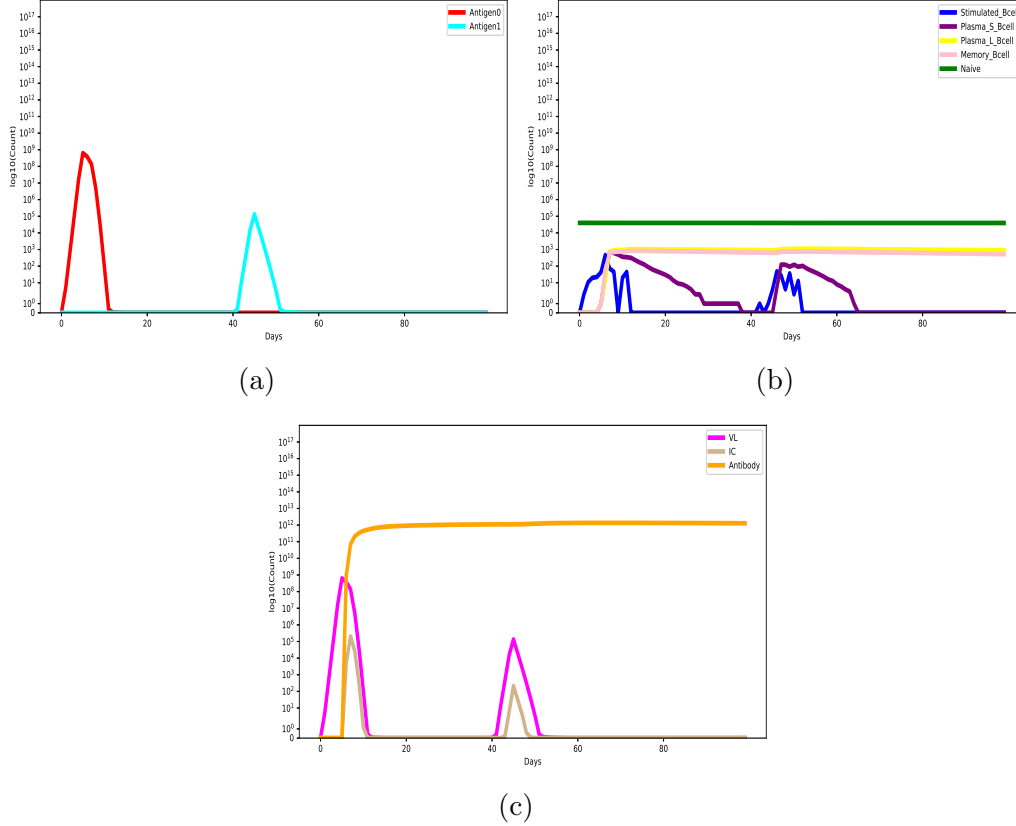


Figure 6.3: **Population dynamics for Case 2 where antigenic distance between primary antigen 614G, and secondary antigen Delta AY.4.2 is 1.73 AU.** This figure shows the simulation time courses following secondary challenge with the Delta AY.4.2 variant at day 40. (a) shows antigen concentration over time, with the secondary antigen (Delta AY.4.2) introduced at day 40. (b) shows B-cell populations, showing naive, stimulated, plasma, and memory subsets. (c) shows viral load , antibody titer, and immune-complex (IC) formation following 614G and Delta AY.4.2 exposure.

The antigenic distance between the two variants is larger compared to case

1, but the ball of stimulation still overlaps (Figure 6.1b). This overlap reduces the cross-reactivity of the existing B cells and antibodies with the secondary antigen. We consider 614G (primary) and Delta AY.4.2 for SARS-CoV-2 variants with an antigenic distance of 1.73 AU between them. Figure 6.3a shows that the secondary antigen Delta follows similar trajectory as the primary antigen 614G, but starts to decline shortly because of the affinity of the pre-existing antibodies from the primary antigen with Delta. Additionally, because of having a moderate affinity, the memory B cells recognizes the secondary antigen and start producing plasma cells in higher rate, resulting in a rapid antibody production.

The stimulated B cell population shows a small, transient peak immediately after the introduction of the primary. However, this peak is significantly smaller than during the primary response to the secondary antigen, but larger than what we see in case 1. This indicates minimal activation of naive B cells or memory B cells that briefly proliferate before differentiating into plasma cells, because of the clearance of antigens by the pre-existing antibodies and rapid production of antibodies from plasma cells because of the memory B cells.

6.4.3 Case 3: Larger Antigenic Distance, No Overlap between Balls of Stimulation

In this case, the antigenic distance between two variants (614G and Omicron B.A.2) are too large that there is no overlap between the ball of stimulations, resulting in no cross-reactivity of B cells and antibodies from primary antigen 614G towards the secondary antigen (Omicron B.A.2).

In Case 3 ($\Delta d = 4.58AU$), the antigenic distance between the primary 614G and secondary Omicron BA.2 exceeds the 2 AU radius of ball of stimulation, resulting in no cross-reactivity by pre-existing 614G specific B cells and antibodies.

Upon Omicron BA.2 challenge at day 40, the antigen concentration rises rapidly to a pronounced peak—reflecting unaffected viral replication by the pre-existing B cells and antibodies from 614G shown in Figure 6.4a. Figure 6.4b shows that the memory B cells specific to 614G remain flat, confirming zero reactivation for Omicron B.A.2. Stimulated B cells begin to appear after a brief lag as Omicron B.A.2.-specific B cells get activated; they peak and then decay according to their programmed lifespan. Plasma cells (short- and long-lived) and new memory B cells emerge only after the four-day maturation delay, leading to a gradual buildup rather than the abrupt spike seen when cross-reactivity is present.

Figure 6.4c shows the viral load which mirrors the antigen curve, peaking well before any antibody-mediated control. Antibody titers remain at the low

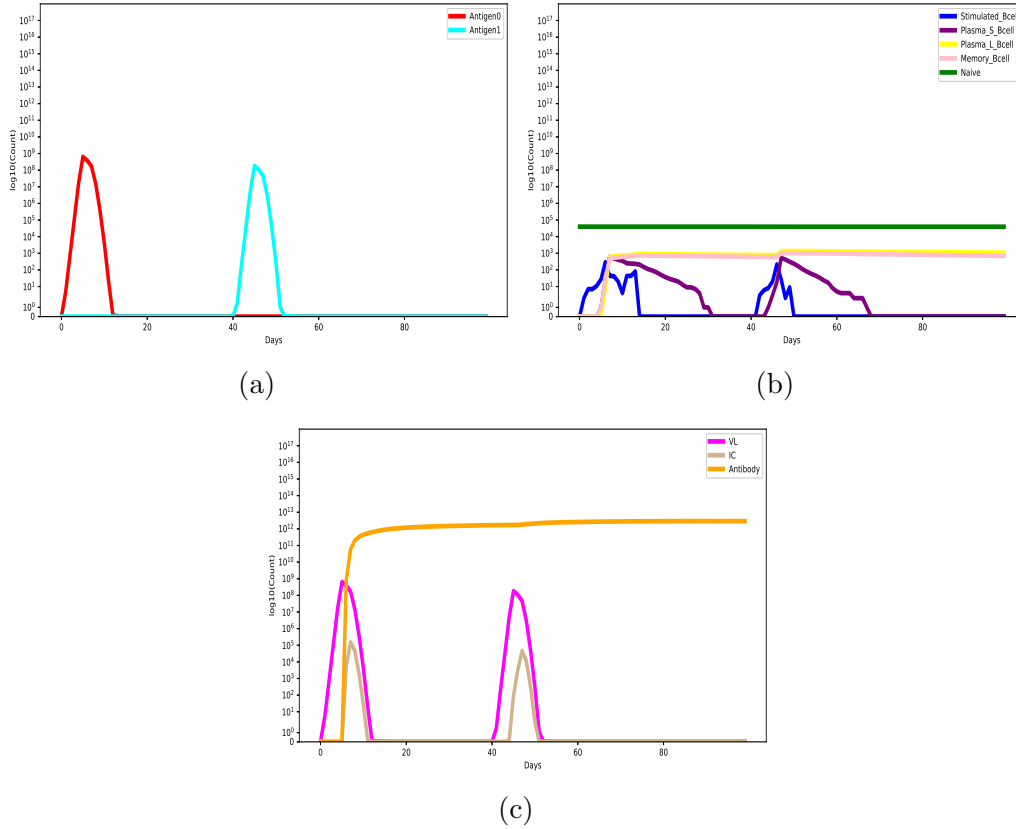


Figure 6.4: Population dynamics of 614G and Omicron B.A.2.

Figure 6.5: **Population dynamics for Case 3 where antigenic distance between primary antigen 614G, and secondary antigen Omicron B.A.2. is 4.58 AU.** This figure shows the simulation time courses following secondary challenge with the Omicron B.A.2. variant at day 40. (a) shows antigen concentration over time, with the secondary antigen (Omicron B.A.2.) introduced at day 40. (b) shows B-cell populations, showing naive, stimulated, plasma, and memory subsets. (c) shows viral load, antibody titer, and immune-complex (IC) formation following 614G and Omicron B.A.2. exposure.

baseline of residual anti-614G IgG until Omicron-specific plasma cells begin secreting new antibodies, at which point titers climb. Immune complexes (IC) are essentially zero during the initial replication phase (no binding by pre-existing antibodies) and rise only once the Omicron B.A.2.-specific antibodies reach sufficient concentration, driving antigen clearance. Together, these panels demonstrate that when antigenic distance places the secondary variant outside the original stimulation radius, hybrid immunity confers no immediate protection: the system must launch a wholly new B-cell response, resulting in a delayed antibody surge, belated immune-complex formation, and a correspondingly later decline in viral load.

6.5 Methods

6.5.1 Computational Model

We initialize a 40 by 40 2D grid, called shape space, to represent antigenic distances between different variants of antigens. Shape space diagram provides a visual representation of the affinities between multiple B cells/antibodies and antigens through the antigenic distance, where each point represents a unique B cell and antigen phenotype. Antigenic distance, indicating the affinity between antigens and B cells, was calculated as the hamming distance between points in this space. Stochastic agent-based model simulates the B cells in the shape space.

We created an initial concentration of a specific SARS-CoV-2 variant antigen (N_{Ag}) in the shape space. The position of the SARS-CoV-2 variant in the shape space diagram is calculated from [111]. The model then generates an initial number of B cells N_{BC} and uniformly distribute them in groups within the same 2D grid. Each group contains up to 100 B cells of similar phenotype meaning each group corresponds to specific coordinates within the grid and all cells of a group share identical coordinates in the shape space. We considered 6 hours as one time-step. Each B cell is initially considered a naive B cell. During each time step, they can shift to activated B cells, stimulated B cells, plasma L cells, plasma S cells, memory cells, and mutated cells in the simulation. Figure 6.6 shows the framework of our model.

- **Activation:** The activation mechanism involves identifying B cell neighbors within a specified distance from the antigens. We assume B cells within the specified distance have enough affinity with the antigen to get activated. In our model, we consider this activation distance $r = 2$ AU and we call this region ball of stimulation. We compute the affinity α_i between B cell/antibody of phenotype i and the antigen following this equation:

$$\alpha_i = 10^{-d_i} \quad (6.1)$$

Here d_i is the antigenic distance from the B cell/antibody of phenotype i to the antigens. This equation indicates that B cell/antibody with lower d_i has higher affinity. Maximum binding affinity is 1 when the

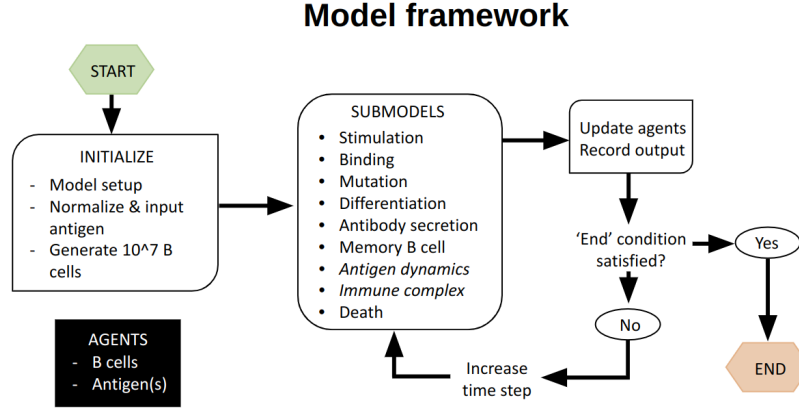


Figure 6.6: Our simulation starts with initializing B cells and antigens in the shape space. Each point in the shape space represents a phenotype and each phenotype has multiple B cells. The phenotype groups containing multiple B cells are distributed uniformly in the shape space and antigens with a specific phenotype are plotted on the specific coordinate. The distance between the antigens and the B cells represents their antigenic distance. The model proceeds through submodels for stimulation, binding, mutation, differentiation, antibody secretion, antigen dynamics, and immune complex formation. The model keeps running until it reaches the specified time-step.

distance is zero.

- Stimulation: Not all the activated B cells can bind to the antigen and get stimulated. Activated B cells can get stimulated based on their affinity. The model employs a stimulation mechanism at each time step to govern the interactions between B cells of phenotype i and antigens, where it computes the probability of binding ($P(binding)_i$) to the antigen based on affinity. If the B cell is within the ball of stimulation, it is assigned

a binding probability as follows:

$$P(binding)_i = C_b \times \alpha_i \quad (6.2)$$

Here C_b is the binding co-efficient. We assume $C_b = 0.1$.

This formula ensures that not every stimulated B cell binds with the antigen. Stimulation depends on the distance between B cells and antigens. B cells with lower affinity or whom are in close proximity to the antigen, have a higher probability of getting bound to the antigen.

- **Mutation:** Once stimulated, B cells can mutate leading to a diverse repertoire of B cells with a varying affinity for the antigen. During the mutation phase, the B cell changes its phenotype by choosing a random angle from a uniform distribution and a distance from an exponential distribution. Suppose the newly changed phenotype coordinate is within the specified radius r from the antigen. In that case, its affinity is updated based on its distance from the antigens and it can get stimulated in the next time step. Otherwise, it stays in the shape space but doesn't participate in any further simulation. We assume the probability of mutation $P(mutation)$ is 1% indicating a lower occurrence of mutation.
- **Proliferation and Differentiation:** Each stimulated B cell undergoes rapid clonal expansion, generating r_{clone_s} 2 more daughter cells at each time step expressing the same B cell phenotype. We assume each B

cell has a mean lifetime of 108 hr. Therefore, each stimulated B cell can produce a maximum of 36 daughter B cells. Each daughter cell has a 5% probability $P(memory)$ of being a memory B cell and 5% of being a plasma cell $P(plasma)$. Among those plasma cells, 75% are short-lived plasma cells $P(S)$ and 25% are long-lived plasma cells $P(L)$. All other cells remain stimulated cells and get the same probability in the next time step until they die. The mother cell dies (removed from the system) after each proliferation. After the 8th generation of proliferation, the new daughter cell becomes a plasma cell with a 50% chance of a plasma cell and 50% of a memory cell.

- Plasma Cell: At each time step, the plasma cells can secrete antibodies at a constant rate s_{Ab} with a certain probability. Short-lived plasma cells secrete $s_{Ab}^S = 2000$ antibodies per sec [150] and long-lived plasma cells secrete $s_{Ab}^L = 2000$ antibodies per hour. The high secretion rate of short-lived plasma cells ensures an immediate large production of antibodies to neutralize the pathogen quickly. The long-lived plasma cell sustains low levels of antibody production over extended periods. Their slower secretion rate reflects their role in maintaining baseline immunity rather than responding to active infection.
- Memory B Cell: Memory B cell undergoes rapid proliferation to produce antibodies for a repeated antigen if the memory B cell

falls within the ball of activation to the new variant of antigen. It proliferated $r_{clone_m} = 4$ daughter cells with a 75% probability of differentiating into plasma cells and a 25% probability to memory cells. Memory cells act as first responders in subsequent infections and they live for a long time in the system.

- **Antibody:** The antibody grows at a constant rate from plasma cells and decays following an exponential decay. The decay method iterates through each antibody phenotype, considering the current count of antibodies associated with that phenotype as follows:

$$\frac{dAb_i(t)}{dt} = -r_{Ab} \cdot Ab_i(t) \quad (6.3)$$

Here r_{Ab} is the antibody decay rate (per time step) and $Ab_i(t)$ the antibody count of phenotype i at time t . We compute r_{Ab} from the half-life (h_{Ab}) of antibodies as follow:

$$r_{Ab} = \frac{\ln(2)}{h_{Ab}} \quad (6.4)$$

The half-life of the antibody is 21 days ****Cite Ruy's paper****.

- **Antigen Dynamics:** Antigen level changes over time through clearance and replication dynamics. Antigen follows an exponential growth until it reaches the peak and follows an exponential decay afterward:

$$Ag(t + 1) = Ag(t) \times e^{-r_{Ag}t} \quad (6.5)$$

Here r_{Ag} is the growth and decay rate, and $Ag(t)$ is the antigen count at time t . r_{Ag} is 5 per day for the growth and 0.5 per day as the decay rate. The formula of exponential growth or decay arises from the solution to a first-order ordinary differential equation (ODE).

$$\frac{Ag(t)}{dt} = r_{Ag} \cdot Ag(t) \quad (6.6)$$

[56] mentioned that SARS-CoV-2 reaches its peak around day 4. Hence the decay rate is activated after day 4. The binding with antibodies also affects the decay of antigens.

- **Lifespan:** Each B cell, except naive B cells, has a certain lifespan and undergoes apoptosis after reaching that lifespan. Because of rapid turnover of naive b cells from bone marrow, they are considered to have an infinite lifespan for the purpose of this study. At different stages, the B cell is assigned a random lifetime from an exponential distribution, where their half-life mentioned in Table 6.2 corresponds to the mean value of the distribution. At each time step, if the B cell exceeds its assigned lifespan, a the cell death is simulated by removing it from the system.
- **Immune Complex (IC):** At each time step, if antibodies are present,

antigens bind to these antibodies and create an IC structure (IC_i) for each antibody phenotype i . In our model, we assume only one antigen type is present at a specific time. IC_i formation follows a quasi-equilibrium equation:

$$Ag + Ab_i \rightleftharpoons [K^+][K^-]IC_i \quad (6.7)$$

Here Ag represents the free antigens and Ab_i represents the free antibodies of phenotype i .

The quasi-equilibrium equation, representing the equilibrium state of ICs, is solved following a first order differential equation:

$$\frac{dIC_i}{dt} = K^+[Ab_i][Ag] - K^-[IC_i] \quad (6.8)$$

The free antigen Ag and the free antibodies Ab_i can be computed from the total antigens T_{Ag} and total antibodies T_{Ab_i} of certain phenotypes i as follows:

$$T_{Ag} = Ag + \sum_{i=1}^N IC_i \quad (6.9)$$

$$T_{Ab_i} = Ab_i + IC_i \quad (6.10)$$

We assume the formation of ICs is a two step process involving T_{Ag} and T_{Ab_i} of certain phenotypes i :

- We assume no initial ICs when the system first reaches equilibrium.

Therefore, the total antigen T_{Ag} is equal to the free antigen, Ag_i and the total antibody T_{Ab_i} is equal to the free antibody Ab_i of phenotype i . ICs formation is initiated by antibodies with the highest affinity for antigens.

- Once the system reaches equilibrium and there is ICs, we assume $\frac{dIC_i}{dt} = 0$. From Equations (6.8) to (6.10), the equilibrium condition can be written as,

$$\begin{aligned}
& K^+[Ab_i][Ag] - K^-[IC_i] = 0 \tag{6.11} \\
& \Leftrightarrow K^+[T_{Ab_i} - IC_i][T_{Ag} - IC_i] - K^-[IC_i] = 0 \\
& \Leftrightarrow K^+[T_{Ab_i} - IC_i][T_{Ag} - IC_i] = K^-[IC_i] \\
& \Leftrightarrow [T_{Ab_i} - IC_i][T_{Ag} - IC_i] = \frac{K^-[IC_i]}{K^+} \\
& \Leftrightarrow T_{Ab_i}T_{Ag} - IC_iT_{Ag} - T_{Ab_i}IC_i + IC_i^2 = K_{d_i}[IC_i] \\
& \Leftrightarrow IC_i^2 - IC_iT_{Ag} - T_{Ab_i}IC_i - K_{d_i}[IC_i] + T_{Ab_i}T_{Ag} = 0 \\
& \Leftrightarrow IC_i^2 - IC_i(T_{Ag} + T_{Ab_i} + K_{d_i}) + T_{Ab_i}T_{Ag} = 0 \\
& \Leftrightarrow IC_i = \frac{(T_{Ag} + T_{Ab_i} + K_{d_i}) - \sqrt{(T_{Ag} + T_{Ab_i} + K_{d_i})^2 - 4T_{Ag}T_{Ab_i}}}{2}
\end{aligned}$$

Here $K_{d_i} = \frac{K^-}{K^+}$ is the dissociation constant and computed as:

$$K_{d_i} = \frac{1}{10^{-8} \times \alpha_i} \tag{6.12}$$

The equilibrium concentrations of ICs are adjusted to ensure consistency with the total antigens. If necessary, normalization is performed to maintain the balance.

The clearance of IC is modeled using an exponential decay, where the IC_i at time $t + 1$ is given by,

$$IC_i(t + 1) = IC_i(t) \times e^{-r_{IC}} \quad (6.13)$$

Here, r_{IC} = IC decay rate per time step. We consider that ICs decay rate is 23 per day [25].

6.6 Discussion

In this study, we developed an agent-based model of SARS-CoV-2 specific adaptive immunity that explicitly represents B-cell, antibody, and antigen populations in a two-dimensional shape space. By embedding successive exposures to variant antigens (614G $\rightarrow$ Alpha, Delta AY.4.2, Omicron BA.2), we showed how antigenic distance governs the degree of cross-reactivity, shaping the speed and magnitude of immune responses. When antigenic distance is small (Case 1, 0.81 AU), pre-existing antibodies and memory B cells rapidly neutralize the challenge, yielding minimal viral replication and negligible de novo clonal expansion. At moderate distances (Case 2, 1.73 AU), a partial overlap of stimulation radius produces an intermediate

Table 6.1: Parameters and Notations

Parameter	Description	Value	Unit
r	Radius of the ball of stimulation	2	AU
d_i	Antigenic distance between a B cell/antibody of phenotype i		AU
α_i	Affinity between a B cell/antibody of phenotype i with the antigen		
$P(binding)_i$	Probability of binding of phenotype i with antigens		
C_b	B cell binding co-efficient towards an antigen	0.1	
$P(mutation)$	Probability of mutation	0.01	
$P(memory)$ 2	Probability of daughter B cells to be the memory B cell	0.05	
$P(plasma)$ 2	Probability of daughter B cells to be the plasma B cell	0.05	
$P(S)$ 2	Probability of plasma B cells to be the short-lived	0.75	
$P(L)$ 2	Probability of plasma B cells to be the long-lived	0.25	
s_{Ab}^S	Short-lived plasma cell antibody secretion rate	2000	sec^{-1}
s_{Ab}^L	Long-lived plasma cell antibody secretion rate	2000	hr^{-1}
r_{clone_s}	Production rate of each stimulated B cell	2	$timestep^{-1}$
r_{clone_m}	Production rate of each memory B cell after recognizing the new variant	4	$timestep^{-1}$
r_{Ab}	antibody decay rate		$timestep^{-1}$
$Ab_i(t)$	antibody count of phenotype i at time t		$timestep^{-1}$
h_{Ab}	half-life of antibodies	21	day [?]
r_{Ag}	growth and decay rate	5 & 0.5	day^{-1}
$Ag(t)$	antigen count at time t		
r_{IC}	IC decay rate	23	day^{-1}

Table 6.2: Mean age of B cells, and antibodies. The age is randomly assigned following an exponential distribution with mean age

	Value (days)	Unit
Activated & Stimulated B cell	4.5	days
Plasma L cell	200	days
Plasma S cell	3	days
Memory Cell	126	days
Antibody	21	days

phenotype: attenuated but still accelerated antibody kinetics relative to the primary response. When distance exceeds the stimulation radius (Case 3, 4.58 AU), hybrid immunity confers no immediate protection, and the system must re-launch a response, delaying peak antibody titers and viral clearance. These results quantitatively recapitulate empirical observations of variant-dependent vaccine efficacy and underscore the importance of antigenic mapping in anticipating immune escape.

Future Work

- **Model Calibration through Data Fitting:** We plan to fit our model outputs to longitudinal viral-load and antibody-titer datasets from clinical cohorts for SARS-CoV-2. By employing nonlinear least-squares and the Approximate Bayesian Computation method for parameter estimation, we are adjusting key parameters—such as antigen growth/decay rates, B cell death rates, antibody secretion/decay rates, and binding coefficients

to minimize discrepancies between simulated and observed trajectories.

- Computation of half-maximal effective concentration (EC_{50}): We will leverage our agent-based model to generate in silico dose-response curves. We will take antibodies raised against a primary SARS-CoV-2 variant and perform serial dilutions. For each dilution level, we will measure the fraction of viral inhibition (or antigen neutralization) in vitro against a panel of variant antigens. By plotting inhibition probability versus dilution factor, we will identify the dilution at which 50 % inhibition occurs, defined as the EC_{50} , for each variant.

Bibliography

- [1] Body weight information for j:nu (007850).
- [2] Human lymphatic system 3d model. https://www.plasticboy.co.uk/store/Human_Lymphatic_System_no_textures.html, Last accessed 30 April, 2022.
- [3] EM Abdel-Magied, AAM Taha, AA Al-Qarawi, and MG Elfaki. The parotid, mandibular and lateral retropharyngeal lymph nodes of the camel (*camelus dromedarius*). *Anatomia, histologia, embryologia*, 30(4):199–203, 2001.
- [4] Dalia Y Abel-Galil, Abdel-Rahman Tawfik, and Abdel-Hakim Saad. Anatomical and comparative distribution of lymph nodes in some egyptian chiroptera.
- [5] Frederick R. Adler and Deborah M. Gordon. Information collection and spread by networks of patrolling ants. *The American Naturalist*, 140(3):373–400, 1992. PMID: 19426049.

- [6] Hirotugu Akaike. A new look at the statistical model identification. *IEEE transactions on automatic control*, 19(6):716–723, 1974.
- [7] Christian L Althaus. Of mice, macaques and men: scaling of virus dynamics and immune responses, 2015.
- [8] Philip L Altman, Dorothy S Dittmer, et al. Biology data book, volume iii., 1974.
- [9] Carl Anderson and Daniel W McShea. Individual versus social complexity, with particular reference to ant colonies. *Biological reviews*, 76(2):211–237, 2001.
- [10] N Anna Toenjes. *Felis catus (On-line)*, *Animal Diversity Web*, 2014 (accessed Nov 24, 2020).
- [11] T Petteri Arstila, Armanda Casrouge, Véronique Baron, Jos Even, Jean Kanellopoulos, and Philippe Kourilsky. A direct estimate of the human $\alpha\beta$ t cell receptor diversity. *Science*, 286(5441):958–961, 1999.
- [12] Jayanth R Banavar, Melanie E Moses, James H Brown, John Damuth, Andrea Rinaldo, Richard M Sibly, and Amos Maritan. A general basis for quarter-power scaling in animals. *Proceedings of the National Academy of Sciences*, 107(36):15816–15820, 2010.
- [13] Soumya Banerjee and Melanie Moses. Scale invariance of immune system response rates and times: perspectives on immune system architecture

- and implications for artificial immune systems. *Swarm Intelligence*, 4(4):301–318, 2010.
- [14] Soumya Banerjee, Alan S Perelson, and Melanie Moses. Modelling the effects of phylogeny and body size on within-host pathogen replication and immune response. *Journal of The Royal Society Interface*, 14(136):20170479, 2017.
- [15] K Basnayake, Zeev Schuss, and David Holcman. Asymptotic formulas for extreme statistics of escape times in 1, 2 and 3-dimensions. *Journal of Nonlinear Science*, 29:461–499, 2019.
- [16] Madeleine Beekman, David J. T. Sumpter, and Francis L. W. Ratnieks. Phase transition between disordered and ordered foraging in pharaoh’s ants. *Proceedings of the National Academy of Sciences*, 98(17):9703–9706, 2001.
- [17] George I Bell. Mathematical model of clonal selection and antibody production. *Journal of theoretical biology*, 29(2):191–232, 1970.
- [18] Lukas Breitwieser, Ahmad Hesam, Jean De Montigny, Vasileios Vavourakis, Alexandros Iosif, Jack Jennings, Marcus Kaiser, Marco Manca, Alberto Di Meglio, Zaid Al-Ars, et al. Biodynamo: a modular platform for high-performance agent-based simulation. *Bioinformatics*, 38(2):453–460, 2022.

- [19] Gary J Brenner, Nicholas Cohen, and Jan A Moynihan. Similar immune response to nonlethal infection with herpes simplex virus-1 in sensitive (balb/c) and resistant (c57bl/6) strains of mice. *Cellular immunology*, 157(2):510–524, 1994.
- [20] Aisling F Brown, Alison G Murphy, Stephen J Lalor, John M Leech, Kate M O’Keeffe, Micheál Mac Aogáin, Dara P O’Halloran, Keenan A Lacey, Mehri Tavakol, Claire H Hearnden, et al. Memory th1 cells are protective in invasive staphylococcus aureus infection. *PLoS pathogens*, 11(11):e1005226, 2015.
- [21] TA Brown, BR Murphy, J Radl, JJ Haaijman, and J Mestecky. Subclass distribution and molecular form of immunoglobulin a hemagglutinin antibodies in sera and nasal secretions after experimental secondary infection with influenza a virus in humans. *Journal of clinical microbiology*, 22(2):259–264, 1985.
- [22] Melinda J. Burrill. Sheep. In *World Book*. World Book Inc, 2004.
- [23] Jessica M Cable, Brian J Enquist, and Melanie E Moses. The allometry of host-pathogen interactions. *PLoS One*, 2(11):e1130, 2007.
- [24] William A Calder. *Size, function, and life history*. Harvard University Press, Cambridge, UK, 1984.

- [25] E Fabian Cardozo-Ojeda and Alan S Perelson. Modeling hiv-1 within-host dynamics after passive infusion of the broadly neutralizing antibody vrc01. *Frontiers in Immunology*, 12:710012, 2021.
- [26] Armanda Casrouge, Emmanuel Beaudoin, Sophie Dalle, Christophe Pannetier, Jean Kanellopoulos, and Philippe Kourilsky. Size estimate of the $\alpha\beta$ tcr repertoire of naive mouse splenocytes. *The Journal of Immunology*, 164(11):5782–5787, 2000.
- [27] Susanna Celli, Mark Day, Andreas J Müller, Carmen Molina-Paris, Grant Lythe, and Philippe Bousso. How many dendritic cells are required to initiate a t-cell response? *Blood, The Journal of the American Society of Hematology*, 120(19):3945–3948, 2012.
- [28] Eric L Charnov, Robin Warne, and Melanie Moses. Lifetime reproductive effort. *The American Naturalist*, 170(6):E129–E142, 2007.
- [29] Vasiliki Chatzaraki, Janette Verster, Carlo Tappero, Michael J Thali, Wolf Schweitzer, and Garyfalia Ampanozi. Spleen measurements with reference to cause of death and spleen weight estimation: A study on postmortem computed tomography. *Journal of Forensic Radiology and Imaging*, 18:24–31, 2019.
- [30] Tom Chou and Maria R D’Orsogna. First passage problems in biology. In *First-passage phenomena and their applications*, pages 306–345. World Scientific, 2014.

- [31] JoanM Clarke, JR Shelton, GR Venning, John Hamer, and Sue Taylor. The rhythm of the normal human heart. *The Lancet*, 308(7984):508–512, 1976.
- [32] Richard M Coles, Scott N Mueller, William R Heath, Francis R Carbone, and Andrew G Brooks. Progression of armed ctl from draining lymph node to spleen shortly after localized infection with herpes simplex virus 1. *The Journal of Immunology*, 168(2):834–838, 2002.
- [33] Robert M Corless, Gaston H Gonnet, David EG Hare, David J Jeffrey, and Donald E Knuth. On the lambert w function. *Advances in Computational mathematics*, 5:329–359, 1996.
- [34] Emily Cornelius Ruhs, Daniel J Becker, Samantha J Oakey, Ololade Ogunsina, M Brock Fenton, Nancy B Simmons, Lynn B Martin, and Cynthia J Downs. Body size affects immune cell proportions in birds and non-volant mammals, but not bats. *Journal of Experimental Biology*, 224(13):jeb241109, 2021.
- [35] Rob J De Boer, Dirk Homann, and Alan S Perelson. Different dynamics of cd4+ and cd8+ t cell responses during and after acute lymphocytic choriomeningitis virus infection. *The Journal of Immunology*, 171(8):3928–3935, 2003.
- [36] Rob J De Boer, Mihaela Oprea, Rustom Antia, Kaja Murali-Krishna, Rafi Ahmed, and Alan S Perelson. Recruitment times, proliferation,

- and apoptosis rates during the cd8+ t-cell response to lymphocytic choriomeningitis virus. *Journal of virology*, 75(22):10663–10669, 2001.
- [37] Rob J De Boer and Alan S Perelson. Quantifying t lymphocyte turnover. *Journal of theoretical biology*, 327:45–87, 2013.
- [38] Adrianus J De Langen, Pieter Raijmakers, Ingrid Riphagen, Marinus A Paul, and Otto S Hoekstra. The size of mediastinal lymph nodes and its relation with metastatic involvement: a meta-analysis. *European journal of cardio-thoracic surgery*, 29(1):26–29, 2006.
- [39] Robert Degabriele and TJ Dawson. Metabolism and heat balance in an arboreal marsupial, the koala (*phascolarctos cinereus*). *Journal of comparative physiology*, 134(4):293–301, 1979.
- [40] Matina C. Donaldson-Matasci, Gloria DeGrandi-Hoffman, and Anna Dornhaus. Bigger is better: honeybee colonies as distributed information-gathering systems. *Animal Behaviour*, 85(3):585–592, 2013.
- [41] Marco Dorigo, Guy Theraulaz, and Vito Trianni. Reflections on the future of swarm robotics. *Science Robotics*, 5(49):eabe4385, 2020.
- [42] Anna Dornhaus, Franziska Klügl, Christoph Oechslein, Frank Puppe, and Lars Chittka. Benefits of recruitment in honey bees: effects of ecology and colony size in an individual-based model. *Behavioral Ecology*, 17(3):336–344, 01 2006.

- [43] Cynthia J Downs, Ned A Dochtermann, Ray Ball, Kirk C Klasing, and Lynn B Martin. The effects of body mass on immune cell concentrations of mammals. *The American Naturalist*, 195(1):107–114, 2020.
- [44] Cynthia J Downs, Laura A Schoenle, Barbara A Han, Jon F Harrison, and Lynn B Martin. Scaling of host competence. *Trends in parasitology*, 35(3):182–192, 2019.
- [45] Anja-A Dunne, S Plehn, S Schulz, A Levermann, A Ramaswamy, BM Lippert, and JA Werner. Lymph node topography of the head and neck in new zealand white rabbits. *Laboratory animals*, 37(1):37–43, 2003.
- [46] Vasiliki Economopoulos, Jennifer C Noad, Shruti Krishnamoorthy, Brian K Rutt, and Paula J Foster. Comparing the mri appearance of the lymph nodes and spleen in wild-type and immuno-deficient mouse strains. *PLoS One*, 6(11):e27508, 2011.
- [47] Herman N Eisen. Affinity enhancement of antibodies: how low-affinity antibodies produced early in immune responses are followed by high-affinity antibodies later and in memory b-cell responses. *Cancer immunology research*, 2(5):381–392, 2014.
- [48] Herman N Eisen and Gregory W Siskind. Variations in affinities of antibodies during the immune response. *Biochemistry*, 3(7):996–1008, 1964.

- [49] Jannatul Ferdous, G. Matthew Fricke, Judy L. Cannon, and Melanie E. Moses. Bigger is faster: The scalable adaptive immune response. *in submission*.
- [50] Jannatul Ferdous, G. Matthew Fricke, and Melanie E. Moses. Modeling immune search through the lymphatic network. In *Swarm Intelligence*, pages 332–340, Cham, 2022. Springer International Publishing.
- [51] Jannatul Ferdous, George Matthew Fricke, and Melanie E Moses. More is faster: Why population size matters in biological search. *Journal of Computational Biology*, 31(5):429–444, 2024.
- [52] Jannatul Ferdous, George Matthew Fricke, and Melanie E. Moses. More is faster: Why population size matters in biological search. *Journal of Computational Biology*, 31(5):429–444, 2024. PMID: 38754139.
- [53] Tatiana P. Flanagan, Kenneth Letendre, William R. Burnside, et al. Quantifying the effect of colony size and food distribution on harvester ant foraging. *PLOS ONE*, 7(7):1–9, 07 2012.
- [54] Peter Flood. The anatomy of the domestic animals. *Australian Veterinary Journal*, 60, 07 1983.
- [55] Centers for Disease Control and Prevention. Covid data tracker. <https://covid.cdc.gov/covid-data-tracker/#datatracker-home>. Accessed: 2024-12-17.

- [56] Jennifer K Frediani, Richard Parsons, Kaleb B McLendon, Adrianna L Westbrook, Wilbur Lam, Greg Martin, and Nira R Pollock. The new normal: delayed peak sars-cov-2 viral loads relative to symptom onset and implications for covid-19 testing programs. *Clinical Infectious Diseases*, 78(2):301–307, 2024.
- [57] Heather Friberg, Hema Bashyam, Tomoko Toyosaki-Maeda, James A Potts, Thomas Greenough, Siripen Kalayanarooj, Robert V Gibbons, Ananda Nisalak, Anon Srikiatkachorn, Sharone Green, et al. Cross-reactivity and expansion of dengue-specific t cells during acute primary and secondary infections in humans. *Scientific reports*, 1(1):1–9, 2011.
- [58] G Matthew Fricke, François Asperti-Boursin, Joshua Hecker, et al. From microbiology to microcontrollers: Robot search patterns inspired by t cell movement. In *Artificial Life Conference Proceedings*, volume ECAL 2013: The Twelfth European Conference on Artificial Life of *ALIFE 2023: Ghost in the Machine: Proceedings of the 2023 Artificial Life Conference*, pages 1009–1016. MIT Press One Rogers Street, Cambridge, MA 02142-1209, USA, 09 2013.
- [59] G Matthew Fricke, Kenneth A Letendre, Melanie E Moses, and Judy L Cannon. Persistence and adaptation in immunity: T cells balance the extent and thoroughness of search. *PLoS computational biology*, 12(3):e1004818, 2016.

- [60] Gilbert H Friedell, Joseph D Sherman, and Perry G Rigby. The effect of foreign protein on the blood and reticuloendothelial system of the hamster. *Blood*, 21(1):80–88, 1963.
- [61] Aljaž Godec and Ralf Metzler. Universal proximity effect in target search kinetics in the few-encounter limit. *Physical Review X*, 6(4):041037, 2016.
- [62] Denis Grebenkov, Ralf Metzler, and Gleb Oshanin. Defocusing of first-reaction times. In *Target Search Problems*, pages 55–76. Springer, 2024.
- [63] Denis S Grebenkov, Ralf Metzler, and Gleb Oshanin. Strong defocusing of molecular reaction times results from an interplay of geometry and reaction control. *Communications Chemistry*, 1(1):96, 2018.
- [64] Denis S Grebenkov and J-F Rupprecht. The escape problem for mortal walkers. *The Journal of chemical physics*, 146(8), 2017.
- [65] Aleks Guvenel, Agnieszka Jozwik, Stephanie Ascough, Seng Kuong Ung, Suzanna Paterson, Mohini Kalyan, Zoe Gardener, Emma Bergstrom, Satwik Kar, Maximillian S Habibi, et al. Epitope-specific airway-resident cd4+ t cell dynamics during experimental human rsv infection. *The Journal of clinical investigation*, 130(1):523–538, 2020.
- [66] R Hadek. The lymph nodes of the guinea-pig. *British Veterinary Journal*, 107(12):487–493, 1951.

- [67] Shigeaki Hamaguchi, Iwai Tohnai, Akira Ito, Kenji Mitsudo, Toshio Shigetomi, Masafumi Ito, Hiroyuki Honda, Takeshi Kobayashi, and Minoru Ueda. Selective hyperthermia using magnetoliposomes to target cervical lymph node metastasis in a rabbit tongue tumor model. *Cancer science*, 94(9):834–839, 2003.
- [68] JONATHAN J Hanger and TREVOR J Heath. Topography of the major superficial lymph nodes and their efferent lymph pathways in the koala (*phascolarctos cinereus*). *Journal of anatomy*, 177:67, 1991.
- [69] Joshua P Hecker and Melanie E Moses. Beyond pheromones: evolving error-tolerant, flexible, and scalable ant-inspired robot swarms. *Swarm Intelligence*, 9:43–70, 2015.
- [70] Christoph U Herborn, Thomas C Lauenstein, Florian M Vogt, Randall B Lauffer, Jorg F Debatin, and Stephan G Ruehm. Interstitial mr lymphography with ms-325: characterization of normal and tumor-invaded lymph nodes in a rabbit model. *American Journal of Roentgenology*, 179(6):1567–1572, 2002.
- [71] Aijima Hiroshi, Hara Hiroyuki, Harada Tomonori, and Hoshi Hajime. Histological findings of the lymph nodes in a case of an african elephant. 日大医学雑誌, 77(1):19–22, 2018.

- [72] Dirk Homann, Luc Teyton, and Michael Oldstone. Differential regulation of antiviral t-cell immunity results in stable cd8+ but declining cd4+ t-cell memory. *Nature medicine*, 7(8):913–919, 2001.
- [73] Iuliana Ionita-Laza, Christoph Lange, and Nan M. Laird. Estimating the number of unseen variants in the human genome. *Proceedings of the National Academy of Sciences*, 106(13):5008–5013, 2009.
- [74] Anita S Iyer, Forrest K Jones, Ariana Nodoushani, Meagan Kelly, Margaret Becker, Damien Slater, Rachel Mills, Erica Teng, Mohammad Kamruzzaman, Wilfredo F Garcia-Beltran, et al. Dynamics and significance of the antibody response to sars-cov-2 infection. *MedRxiv*, 2020.
- [75] Peter GG Jackson, Peter D Cockcroft, Samantha Elmhurst, et al. *Clinical examination of farm animals*. Number i9780632057061. Wiley Online Library, Online, 2002.
- [76] RONALD Jackson and ROGER S Morris. A study of the topography of the lymphatic system of the australian brushtail possum (*trichosurus vulpecula*). *Journal of anatomy*, 188(Pt 3):603, 1996.
- [77] Terry C Jones, Guido Biele, Barbara Mühlemann, Talitha Veith, Julia Schneider, Jörn Beheim-Schwarzbach, Tobias Bleicker, Julia Tesch, Marie Luisa Schmidt, Leif Erik Sander, et al. Estimating infectiousness throughout sars-cov-2 infection course. *Science*, 373(6551):eabi5273, 2021.

- [78] Don R Joseph. The ratio between the heart-weight and body-weight in various animals. *The Journal of experimental medicine*, 10(4):521, 1908.
- [79] J Kabat-Koperska, A Kolasa-Wołoskiuk, B Wojciuk, I Wojciechowska-Kosko, P Roszkowska, B Krasnodebska-Szponder, E Paczkowska, K Safranow, E Golembiewska, B Machaliński, et al. Changes in the immune system of female wistar rats after exposure to immunosuppressive treatment during pregnancy. *Scandinavian journal of immunology*, 83(6):418–426, 2016.
- [80] Yoshitsugu Kawashima. The lymph system in rodents. *Japanese Journal of Veterinary Research*, 20(1-2):35–36, 1972.
- [81] Yoshitsugu Kawashima, Makoto Sugimura, Yann-Ching Hwang, and Norio Kudo. The lymph system in mice. *Japanese Journal of Veterinary Research*, 12(4):69–78, 1964.
- [82] Rachael Keating, Melissa Y Morris, Wen Yue, Cory E Reynolds, Tarsha L Harris, Scott A Brown, Peter C Doherty, Paul G Thomas, and Maureen A McGargill. Potential killers exposed: tracking endogenous influenza-specific cd8+ t cells. *Immunology and cell biology*, 96(10):1104–1119, 2018.

- [83] Thomas B Kepler and Alan S Perelson. Somatic hypermutation in b cells: an optimal control treatment. *Journal of theoretical biology*, 164(1):37–64, 1993.
- [84] Riad Khatib, Kathleen Riederer, Sajjad Saeed, Leonard B Johnson, Mohamad G Fakih, Mamta Sharma, M Shamse Tabriz, and Amir Khosrovaneh. Time to positivity in staphylococcus aureus bacteremia: possible correlation with the source and outcome of infection. *Clinical infectious diseases*, 41(5):594–598, 2005.
- [85] Max Kleiber. Body size and metabolic rate. *Physiological reviews*, 27(4):511–541, 1947.
- [86] Maximilian Koblishke, Marianna T Traugott, Iris Medits, Felicia S Spitzer, Alexander Zoufaly, Lukas Weseslindtner, Cara Simonitsch, Tamara Seitz, Wolfgang Hoepler, Elisabeth Puchhammer-Stöckl, et al. Dynamics of cd4 t cell and antibody responses in covid-19 patients with different disease severity. *Frontiers in medicine*, 7, 2020.
- [87] Matthew F Krummel, Frederic Bartumeus, and Audrey Gérard. T cell migration, search strategies and mechanisms. *Nature Reviews Immunology*, 16(3):193–201, 2016.
- [88] Sean D Lawley. Distribution of extreme first passage times of diffusion. *Journal of Mathematical Biology*, 80(7):2301–2325, 2020.

- [89] Sean D Lawley. Extreme first-passage times for random walks on networks. *Physical Review E*, 102(6):062118, 2020.
- [90] Sean D Lawley. Universal formula for extreme first passage statistics of diffusion. *Physical Review E*, 101(1):012413, 2020.
- [91] Sean D Lawley and Jacob B Madrid. A probabilistic approach to extreme statistics of brownian escape times in dimensions 1, 2, and 3. *Journal of Nonlinear Science*, 30(3):1207–1227, 2020.
- [92] Qing Lei, Yang Li, Hong-yan Hou, Feng Wang, Zhu-qing Ouyang, Yandi Zhang, Dan-yun Lai, Jo-Lewis Banga Ndzouboukou, Zhao-wei Xu, Bo Zhang, et al. Antibody dynamics to sars-cov-2 in asymptomatic covid-19 infections. *Allergy*, 2020.
- [93] Steven M Lewis, Adam Williams, and Stephanie C Eisenbarth. Structure and function of the immune system in the spleen. *Science immunology*, 4(33):eaau6085, 2019.
- [94] Katja Lindenberg, Ralf Metzler, and Gleb Oshanin. *Chemical Kinetics: beyond the textbook*. World scientific, 2019.
- [95] SL Lindstedt and WA Calder III. Body size, physiological time, and longevity of homeothermic animals. *The Quarterly Review of Biology*, 56(1):1–16, 1981.

- [96] Qi Lu, G Matthew Fricke, John C Ericksen, et al. Swarm foraging review: Closing the gap between proof and practice. *Current Robotics Reports*, 1:215–225, 2020.
- [97] Sean Luke, Gabriel Catalin Balan, Liviu Panait, Claudio Cioffi-Revilla, and Sean Paus. Mason: A java multi-agent simulation library. In *Proceedings of Agent 2003 Conference on Challenges in Social Simulation*, volume 9, 2003.
- [98] JH Magilton, K Tanudimadja, and NG Ghoshal. The lymph nodes and lymph vessels of the thoracic viscera of the goat (*capra hircus*) 1. *Anatomia, Histologia, Embryologia*, 2(4):316–326, 1973.
- [99] Sameer-ul-Salam Mattoo and Jinjong Myoung. T cell responses to sars-cov-2 in humans and animals. *Journal of Microbiology*, pages 1–14, 2022.
- [100] Conor McAloon, Áine Collins, Kevin Hunt, Ann Barber, Andrew W Byrne, Francis Butler, Miriam Casey, John Griffin, Elizabeth Lane, David McEvoy, et al. Incubation period of covid-19: a rapid systematic review and meta-analysis of observational research. *BMJ open*, 10(8):e039652, 2020.
- [101] Margaret M McDaniel and Vitaly V Ganusov. Estimating residence times of lymphocytes in ovine lymph nodes. *Frontiers in immunology*, page 1492, 2019.

- [102] Hannah W McKenzie, Mark A Lewis, and Evelyn H Merrill. First passage time analysis of animal movement and insights into the functional response. *Bulletin of mathematical biology*, 71:107–129, 2009.
- [103] Reina E Mebius and Georg Kraal. Structure and function of the spleen. *Nature reviews immunology*, 5(8):606–616, 2005.
- [104] MK Meijer, BM Spruijt, LFM Van Zutphen, and V Baumans. Effect of restraint and injection methods on heart rate and body temperature in mice. *Laboratory animals*, 40(4):382–391, 2006.
- [105] Luka Mesin, Jonatan Ersching, and Gabriel D Victora. Germinal center b cell dynamics. *Immunity*, 45(3):471–482, 2016.
- [106] Hongyu Miao, Joseph A Hollenbaugh, Martin S Zand, Jeanne Holden-Wiltse, Tim R Mosmann, Alan S Perelson, Hulin Wu, and David J Topham. Quantifying the early immune response and adaptive immune response kinetics in mice infected with influenza a virus. *Journal of virology*, 84(13):6687–6698, 2010.
- [107] Joyce Mordenti. Man versus beast: pharmacokinetic scaling in mammals. *Journal of pharmaceutical sciences*, 75(11):1028–1040, 1986.
- [108] Akiko Moriyama, Junko Fujishima, Tomohiro Furukawa, Tsuyoshi Yoshikawa, Rinya Kodama, Yuji Sasaki, Takaharu Nagaoka, Yasuhiro Kamimura, Hiroshi Maeda, Takuya Hirai, et al. Quantitative analyses of lymphoid tissue in the spleen, lymph nodes, and peyer’s patches in

- cynomolgus monkeys. *Journal of Veterinary Medical Science*, pages 1107110573–1107110573, 2011.
- [109] Melanie E. Moses, Judy L. Cannon, Deborah M. Gordon, et al. Distributed adaptive search in t cells: Lessons from ants. *Frontiers in Immunology*, 10, 2019.
 - [110] Melanie E Moses, Steven Hofmeyr, Judy L Cannon, Akil Andrews, Rebekah Gridley, Monica Hinga, Kirtus Leyba, Abigail Pribisova, Vanessa Surjadidjaja, Humayra Tasnim, et al. Spatially distributed infection increases viral load in a computational model of sars-cov-2 lung infection. *PLoS computational biology*, 17(12):e1009735, 2021.
 - [111] Anna Z Mykytyn, Melanie Rissmann, Adinda Kok, Miruna E Rosu, Debby Schipper, Tim I Breugem, Petra B van den Doel, Felicity Chandler, Theo Bestebroer, Maurice de Wit, et al. Antigenic cartography of sars-cov-2 reveals that omicron ba. 1 and ba. 2 are antigenically distinct. *Science Immunology*, 7(75):eabq4450, 2022.
 - [112] Jae Dong Noh and Heiko Rieger. Random walks on complex networks. *Physical review letters*, 92(11):118701, 2004.
 - [113] Charles L Nunn. Spleen size, disease risk and sexual selection: a comparative study in primates. *Evolutionary Ecology Research*, 4(1):109–131, 2002.

- [114] Beatriz Ogassavara, Raul Renato Tucunduva Neto, Romeu Rodrigues de Souza, and Maria José Tucunduva. Ultrasound evaluation of the morphometric patterns of lymph nodes of the head and neck in young and middle-aged individuals. *Radiologia brasileira*, 49(4):225–228, 2016.
- [115] Beatriz Ogassavara, Raul Renato Tucunduva Neto, Romeu Rodrigues de Souza, and Maria José Tucunduva. Ultrasound evaluation of the morphometric patterns of lymph nodes of the head and neck in young and middle-aged individuals. *Radiologia brasileira*, 49(4):225–228, 2016.
- [116] SALLY L Owens, JW Osebold, and YC Zee. Dynamics of b-lymphocytes in the lungs of mice exposed to aerosolized influenza virus. *Infection and immunity*, 33(1):231–238, 1981.
- [117] Young-Jun Park, Dora Pinto, Alexandra C Walls, Zhuoming Liu, Anna De Marco, Fabio Benigni, Fabrizia Zatta, Chiara Silacci-Fregni, Jessica Bassi, Kaitlin R Sprouse, et al. Imprinted antibody responses against sars-cov-2 omicron sublineages. *Science*, 378(6620):619–627, 2022.
- [118] Alan S Perelson, Jason G Bragg, and Frederik W Wiegel. The complexity of the immune system: scaling laws. *Complex systems science in biomedicine*, pages 451–459, 2006.
- [119] Alan S Perelson and George F Oster. Theoretical studies of clonal selection: minimal antibody repertoire size and reliability of self-non-self discrimination. *Journal of theoretical biology*, 81(4):645–670, 1979.

- [120] Alan S Perelson and Frederik W Wiegel. Scaling aspects of lymphocyte trafficking. *Journal of theoretical biology*, 257(1):9–16, 2009.
- [121] Robert H Peters and Robert Henry Peters. *The ecological implications of body size*, volume 2. Cambridge university press, Cambridge, UK, 1986.
- [122] Stefan Popp and Anna Dornhaus. Ants combine systematic meandering and correlated random walks when searching for unknown resources. *iScience*, 26(2):105916, 2023.
- [123] Sharif M Qatarneh, Ion-Christian Kiricuta, Anders Brahme, Ulf Tiede, and Bengt K Lind. Three-dimensional atlas of lymph node topography based on the visible human data set. *The Anatomical Record Part B: The New Anatomist: An Official Publication of the American Association of Anatomists*, 289(3):98–111, 2006.
- [124] DE Rahmoun, MA Lieshchova, and MA Fares. Morphological and radiological study of lymph nodes in dromedaries in algeria. *Regulatory Mechanisms in Biosystems*, 11(2):330–337, 2020.
- [125] Balram Rai, Anandi Shukla, and Laxmi Kant Dwivedi. Incubation period for covid-19: a systematic review and meta-analysis. *Journal of Public Health*, pages 1–8, 2021.
- [126] Marc Henry Ratzlaff. The superficial lymphatic system of the cat. *Lymphology*, 3(4):151–159, 1970.

- [127] Narender P Reddy, Thomas A Krouskop, and Paul H Newell Jr. A computer model of the lymphatic system. *Computers in biology and medicine*, 7(3):181–197, 1977.
- [128] S Redner and B Meerson. Redundancy, extreme statistics and geometrical optics of brownian motion. comment on” redundancy principle and the role of extreme statistics in molecular and cellular biology” by z. schuss et al. *Physics of life reviews*, 28:80–82, 2019.
- [129] Sidney Redner. *A guide to first-passage processes*. Cambridge university press, 2001.
- [130] Miranda J Ridder, Aubrey KG McReynolds, Hongyan Dai, Michele T Pritchard, Mary A Markiewicz, and Jeffrey L Bose. Kinetic characterization of the immune response to methicillin-resistant staphylococcus aureus subcutaneous skin infection. *Infection and Immunity*, 90(7):e00065–22, 2022.
- [131] Sunghan Ro and Yong Woon Kim. Parallel random target searches in a confined space. *Physical Review E*, 96(1):012143, 2017.
- [132] Lauren B Rodda, Peter A Morawski, Kurt B Pruner, Mitchell L Fahning, Christian A Howard, Nicholas Franko, Jennifer Logue, Julie Eggenberger, Caleb Stokes, Inah Golez, et al. Imprinted sars-cov-2-specific memory lymphocytes define hybrid immunity. *Cell*, 185(9):1588–1601, 2022.

- [133] Dmitri A Rusakov and Leonid P Savtchenko. Extreme statistics may govern avalanche-type biological reactions: Comment on “redundancy principle and the role of extreme statistics in molecular and cellular biology” by z. schuss, k. basnayake, d. holcman. *Physics of life reviews*, 28:85–87, 2019.
- [134] LI Saar. Lymph nodes of the head, neck and shoulder region of swine. *Iowa State University Veterinarian*, 25(3):3, 1962.
- [135] LI Saar and R Getty. The lymph nodes and the lymph vessels of the abdominal wall, pelvic wall and the pelvic limb of swine. *Iowa State University Veterinarian*, 26(2):9, 1963.
- [136] Mustafa Sarsilmaz, Y Gumusalan, H Hamdi Celik, et al. Classification of the thoracic and abdominal lymph nodes of the rat. *Turk J Med Res*, 5:185–91, 1994.
- [137] Van M Savage, James F Gillooly, William H Woodruff, Geoffrey B West, Andrew P Allen, Brian J Enquist, and James H Brown. The predominance of quarter-power scaling in biology. *Functional Ecology*, 18(2):257–282, 2004.
- [138] Rostislav Savinkov, Dmitry Grebennikov, Darya Puchkova, Valery Chereshevnev, Igor Sazonov, and Gennady Bocharov. Graph theory for modeling and analysis of the human lymphatic system. *Mathematics*, 8(12):2236, 2020.

- [139] Mathias Schmalzer, Naja J Jann, Fabrizia Ferracin, and Regine Landmann. T and b cells are not required for clearing staphylococcus aureus in systemic infection despite a strong tlr2–myd88-dependent t cell activation. *The Journal of immunology*, 186(1):443–452, 2011.
- [140] Elke Schreurs, Kathelijn Vermote, Virginie Barberet, Sylvie Daminet, Heike Rudorf, and Jimmy H Saunders. Ultrasonographic anatomy of abdominal lymph nodes in the normal cat. *Veterinary Radiology & Ultrasound*, 49(1):68–72, 2008.
- [141] Isabel Schulien, Janine Kemming, Valerie Oberhardt, Katharina Wild, Lea M Seidel, Saskia Killmer, Franziska Daul, Marilyn Salvat Lago, Annegrit Decker, Hendrik Luxenburger, et al. Characterization of pre-existing and induced sars-cov-2-specific cd8+ t cells. *Nature medicine*, 27(1):78–85, 2021.
- [142] LH Schwartz, J Bogaerts, R Ford, L Shankar, P Therasse, S Gwyther, and EA Eisenhauer. Evaluation of lymph nodes with recist 1.1. *European journal of cancer*, 45(2):261–267, 2009.
- [143] Andrew M Seacat, Peter J Thomford, Kris J Hansen, Geary W Olsen, Marvin T Case, and John L Butenhoff. Subchronic toxicity studies on perfluorooctanesulfonate potassium salt in cynomolgus monkeys. *Toxicological Sciences*, 68(1):249–264, 2002.

- [144] Alessandro Sette and Shane Crotty. Adaptive immunity to sars-cov-2 and covid-19. *Cell*, 184(4):861–880, 2021.
- [145] Nan Shi, Wei Xia, Ketong Ji, Yiwei Feng, Hua Li, Guangyao He, and Anzhou Tang. Anatomy and nomenclature of tree shrew lymphoid tissues. *Experimental Animals*, 71(2):173–183, 2022.
- [146] Fernanda Menezes De Oliveira E Silva, Juliana Plácido Guimarães, Jociery Einhardt Vergara-parente, Vitor Luz Carvalho, Ana Carolina Oliveira De Meirelles, Miriam Marmontel, Juliana Shimara Pires Ferrão, and Maria Angelica Miglino. Morphological analysis of lymph nodes in odontocetes from north and northeast coast of brazil. *The Anatomical Record*, 297(5):939–948, 2014.
- [147] Derek J Smith, Stephanie Forrest, David H Ackley, and Alan S Perelson. Variable efficacy of repeated annual influenza vaccination. *Proceedings of the National Academy of Sciences*, 96(24):14001–14006, 1999.
- [148] Felisa A Smith, S Kathleen Lyons, SK Morgan Ernest, Kate E Jones, Dawn M Kaufman, Tamar Dayan, Pablo A Marquet, James H Brown, and John P Haskell. Body mass of late quaternary mammals: ecological archives e084-094. *Ecology*, 84(12):3403–3403, 2003.
- [149] Ricard Solé, Melanie Moses, and Stephanie Forrest. Liquid brains, solid brains, 2019.

- [150] Lauren M Sompayrac. *How the immune system works*. John Wiley & Sons, 2022.
- [151] Marisa Stebegg, Saumya D Kumar, Alyssa Silva-Cayetano, Valter R Fonseca, Michelle A Linterman, and Luis Graca. Regulation of the germinal center response. *Frontiers in immunology*, 9:2469, 2018.
- [152] Mohamed S. Talamali, Arindam Saha, James A. R. Marshall, et al. When less is more: Robot swarms adapt better to changes with constrained communication. *Science Robotics*, 6(56):eabf1416, 2021.
- [153] Shin-ichi Tamura and Takeshi Kurata. Defense mechanisms against influenza virus infection in the respiratory tract mucosa. *Jpn J Infect Dis*, 57(6):236–47, 2004.
- [154] Johannes Textor, Sarah E Henrickson, Judith N Mandl, Ulrich H Von Andrian, Jürgen Westermann, Rob J De Boer, and Joost B Beltman. Random migration and signal integration promote rapid and robust t cell recruitment. *PLoS computational biology*, 10(8):e1003752, 2014.
- [155] Rufina Tretyakova, Rostislav Savinkov, Gennady Lobov, and Gennady Bocharov. Developing computational geometry and network graph models of human lymphatic system. *Computation*, 6(1):1, 2017.
- [156] Gandimohan M Viswanathan, Sergey V Buldyrev, Shlomo Havlin, Marcos GE da Luz, Ernesto P Raposo, and H Eugene Stanley. Optimizing the success of random searches. *nature*, 401(6756):911–914, 1999.

- [157] Ulf G Wagner, Kerstin Koetz, Cornelia M Weyand, and Jörg J Goronzy. Perturbation of the t cell repertoire in rheumatoid arthritis. *Proceedings of the National Academy of Sciences*, 95(24):14447–14452, 1998.
- [158] Sarah Catherine Walpole, David Prieto-Merino, Phil Edwards, John Cleland, Gretchen Stevens, and Ian Roberts. The weight of nations: an estimation of adult human biomass. *BMC public health*, 12(1):439, 2012.
- [159] George H Weiss, Kurt E Shuler, and Katja Lindenberg. Order statistics for first passage times in diffusion processes. *Journal of Statistical Physics*, 31:255–278, 1983.
- [160] Geoffrey B West, James H Brown, and Brian J Enquist. A general model for the origin of allometric scaling laws in biology. *Science*, 276(5309):122–126, 1997.
- [161] Senior Scientist Geoffrey B West et al. *Scaling in biology*. Oxford University Press on Demand, Oxford, UK, 2000.
- [162] Frederik W Wiegel and Alan S Perelson. Some scaling principles for the immune system. *Immunology and cell biology*, 82(2):127–131, 2004.
- [163] Vasily Zaburdaev, Sergey Denisov, and Joseph Klafter. Lévy walks. *Reviews of Modern Physics*, 87(2):483–530, 2015.
- [164] Zhen Zhuang, Xiaomin Lai, Jing Sun, Zhao Chen, Zhaoyong Zhang, Jun Dai, Donglan Liu, Yuming Li, Fang Li, Yanqun Wang, et al. Mapping

and role of t cell response in sars-cov-2–infected mice. *Journal of Experimental Medicine*, 218(4), 2021.