Modeling the cytotoxic T cell response

by

Dennis Lai Chao

B.S.E., Princeton University, 1994

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

December 2004

Acknowledgments

I would like to thank my dissertation committee: Stephanie Forrest, Paul Helman, Rob Miller, Alan Perelson, and Lance Williams. Special thanks go to Miles Davenport, for being such a patient collaborator and for walking me through some of the murkier aspects of immunology, and to my officemates Gabriela Barrantes and Christy Warrender, who read very bad early drafts of my dissertation chapters.

Several funding agencies provided financial support during my time at UNM. I received a graduate fellowship from the National Science Foundation via the BIO Research Training Group in Ecological Complexity (NSF 9553623). This work was also partially supported by the Office of Naval Research (N00014-99-1-0417) and the Defense Advanced Research Projects Agency (AGR F30602-00-2-0584).

Modeling the cytotoxic T cell response

by

Dennis Lai Chao

ABSTRACT OF 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

December 2004

Modeling the cytotoxic T cell response

by

Dennis Lai Chao

B.S.E., Princeton University, 1994

Ph.D., Computer Science, University of New Mexico, 2004

Abstract

This work describes a computer model of the immune system's response to infection, specifically the cytotoxic T lymphocyte (CTL) response. CTLs play an important role in the control of infectious agents, and they are essential components of our defense against HIV, cancer, and other diseases of great public interest. Immunologists are interested in manipulating and enhancing the CTL response to these diseases, whether by vaccination or drug therapy, but the process can be difficult and ad hoc. A combination of animal experimentation, limited human testing, and simple mathematical models have been the primary sources of guidance in the efforts to address these diseases.

Computer models provide an alternative strategy for exploring immune system therapies. Recently developed laboratory techniques that have revealed and quantified many aspects of CTL behavior provide an unprecedented opportunity to develop detailed models. The model used in this work integrates many of these new findings into a coherent system that simulates an immune response to viral infection. This model reproduces many

of the phenomena seen in CTL responses but not captured by other mathematical or computer models and can be used to explore vaccination strategies.

The value of modeling goes beyond simply making predictions. It allows one to perform experiments difficult, or even impossible, to perform in the laboratory. For example, in a computer model one can replicate experiments exactly or choose to allow stochastic fluctuations to influence the outcome. In biological systems, achieving this level of control is impossible. Model-building can also be used as a vehicle for hypothesis testing by formulating one's assumptions about a system's behavior as a model. If the model's behavior does not match real-world experimental results, the initial assumptions can be changed and a new model built. The model presented here is the result of a series of such choices.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

My father always said if you translate a proverb from one language into another, you pass for a poet. The same for science. Work strictly within one area, and it's diminishing returns, hard to make progress. But translate a concept from its field for use where it is unknown, and it is always fresh and powerful. In buying outside, you are doing intellectual arbitrage. The rate limiting step in this is your willingness to continuously translate, to force strange languages to be yours, to live in between, to be everywhere and nowhere.

—Luca Turin, as quoted in *The Emperor of Scent* by Chandler Burr

It is easy for us to take our immune systems for granted. They usually rid our bodies of infectious agents quietly and reliably. The immune system only calls attention to itself when these pathogens are not effectively controlled and illness strikes. High-profile epidemics, such as AIDS and hepatitis C, have forced the public to learn more about immunity, and technical jargon like "viral load" and "T cell count" are entering common usage.

T cells play a major role in our body's defense against these viruses. In particular, one kind of T cell, the cytotoxic T lymphocyte (CTL), mounts attacks against cells infected with viruses and other intracellular pathogens. However, the CTL response is sometimes

deficient (as in the case of AIDS) or even harmful (which can result in autoimmune disorders such as diabetes and arthritis). Studying CTLs will advance our understanding of the pathology of these diseases and give us insight into potential treatment strategies.

Our knowledge of T cells is advancing rapidly. Less than 50 years ago, immunologists used coarse surgical methods to gather the first evidence of the existence and function of T cells (Miller, 1961; Claman et al., 1966). In the past decade, new technologies, such as CFSE labeling (Lyons and Parish, 1994), MHC tetramers (Altman et al., 1996), and two-photon imaging (Miller et al., 2002), have given scientists the unprecedented ability to observe T cells in vivo. I summarize current understanding of CTLs in the first half of Chapter 2. Despite the wealth of data now available, we are still struggling to understand how CTLs behave during an immune response. More sophisticated methods are needed to organize and integrate this information.

This work documents my attempt to understand and model the CTL response to infection. Modeling provides a framework in which to express the relationships among things in the world. By necessity, we simplify the real world to simulate the phenomena that interest us. Thus, the process of model formulation involves not only deciding which data are correct, but also selecting which are essential to replicate the phenomena of interest and which are not. Once a model is developed, it can be used to perform experiments that would be too difficult or even impossible to perform on the "real" system.

Immunological modeling is a relatively new field, and the second half of Chapter 2 describes the models most closely related to mine. The choice of modeling approach influences the kinds of knowledge one can incorporate and the results that can be produced. Most immunological models are mathematical—systems of equations that can be solved. Mathematical models are usually extremely simple in order to be tractable. Simplicity can make their results more robust and general, but it can also force the scientist to omit essential properties of the system. For example, much of the immunological data gained using the latest laboratory techniques is difficult to incorporate into these models. I have

constructed a computer model to study CTLs. Computer models can be a great deal more complex than mathematical models, and can therefore include more of these recent findings. A handful of other computer models have been used in immunology, but none have focused on CTL responses.

I define the behavior of CTLs and viruses in my model in Chapter 3. In this model, viruses infect healthy cells, and infected cells produce more virus. CTLs respond to the infection by reproducing rapidly and eliminating infected cells. The model integrates the findings from dozens of laboratory experiments into a single coherent description of CTL responses. Adding detail to CTL behavior in a model can be computationally expensive, but I use a stage-structured modeling approach that efficiently represents the actions of hundreds of billions of immune cells. With this model, one can begin to make predictions about how CTLs will behave in different circumstances. The immunological data used to construct the model are subject to interpretation, and several alternative assumptions are briefly listed in Appendix B.

Chapter 4 describes the rules the model uses to determine the strength of interactions between CTLs and antigens in the model. Most cells present a sample of their internal proteins on their surfaces, and CTLs have receptors to sense these proteins. This mechanism allows CTLs to detect if a cell contains abnormal proteins, such as those produced by viruses and intracellular bacteria. An individual CTL's receptors are specific to a small subset of proteins, and the body creates millions of different CTLs so it can detect a wide variety. The strength of the bond between a CTL's receptors and an infected cell's surface proteins determines how quickly the CTL can eliminate the infected cell. Accurate molecular simulations of the binding process are too complex to embed in a larger model of T cell behavior, so I use a simpler, abstract representation of this interaction. I am not as concerned with the mechanics of CTL–antigen interactions as with the fact that these interactions can have a range of strengths. Including a spectrum of CTLs with different binding characteristics allows the model to produce a greater variety of immunological

phenomena. A typical CTL response to antigen is composed of CTLs that have various affinities to the antigen, and the composition of the response can affect the ability of the immune system to eliminate the infection.

In Chapter 5, I use the model to provide possible explanations for phenomena observed in the laboratory and to make predictions that could be verified by experimentalists. I explored two facets of CTL behavior: detection and response. Before the immune system can resolve an infection, it must first be able to detect it. Detection is difficult because CTLs must distinguish between uninfected cells and those harboring viruses. The model includes a simplified representation of the process that the immune system uses to create a set of CTLs that is both accurate and efficient at making this distinction, and I used the model to quantify the efficiency of this process. After the pathogen is detected, CTLs can eliminate infected cells. I describe several experiments to demonstrate that the model's results agree with well-characterized CTL behavior, indicating that the model's representation of the CTL response is plausible.

The basic CTL model defined in Chapter 3 replicates responses in which the immune response clears an infection quickly. In situations in which it does not (e.g., chronic diseases such as AIDS), the model assumptions are not valid. The dynamics of prolonged immune responses are not well-characterized, so in Chapter 6 I extend the model based on one of many competing theories. A common feature of prolonged infections is the premature reduction of the CTL response, known as *immune exhaustion*. I test the effects of adding exhaustion, as well as the CTL response to a mutating pathogen.

Finally, I make a few concluding remarks in Chapter 7.

I have made efforts to make this work accessible to readers unfamiliar with immunology or computer science. Chapter 2 summarizes the immunology necessary to understand the model, and I have included a short glossary of immunological terms (which starts on page 127). For the readers who are familiar with T cell biology, I have compared my

results to those from laboratory experiments and have proposed additional experiments that could be performed to validate my results. At the ends of Chapters 3–6 are short summaries to make this work easier for all audiences to follow.

Enjoy.

Chapter 2

Background and related work

Scientists often have a naive faith that if only they could discover enough facts about a problem, these facts would somehow arrange themselves in a compelling and true solution.

—Theodosius Dobzhansky, *Mankind Evolving*

A good physicist is a man with original ideas. A good engineer is a person who makes a design that works with as few original ideas as possible.

—Freeman Dyson, *Adventures in Experimental Physics*

This chapter summarizes the biological and modeling background that informs my own work. Section 2.1 outlines the T cell biology necessary to understand my model. Because other immune cell types are not explicitly represented in my model, their interactions with T cells will be only briefly outlined. Section 2.2 reviews related models of the adaptive immune system. These models can be roughly grouped into mathematical and computer models. My model borrows a few techniques from prior computer models, but for efficiency, I use a stochastic stage-structured approach to modeling. This technique

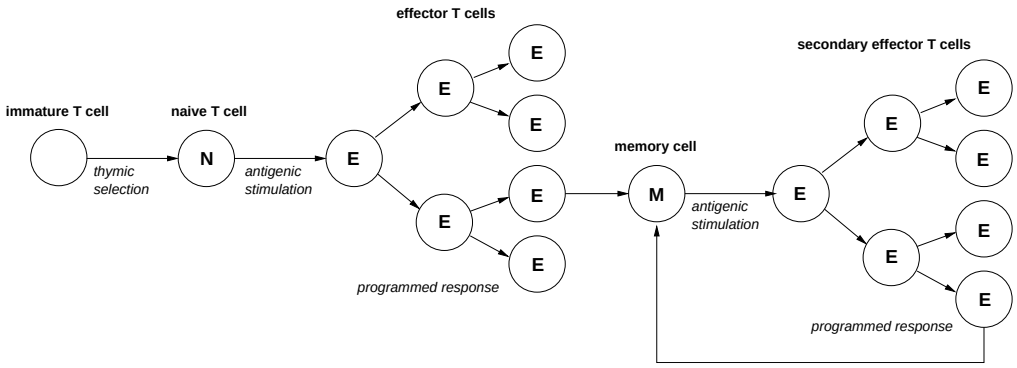


Figure 2.1: A simplified T cell life cycle. Immature T cells are subjected to thymic selection, and those that survive become naïve cells. Naïve cells, when exposed to antigen, become effector cells, which rapidly proliferate and eliminate infected cells in a primary response. At the end of the response, long-lived memory cells remain. When exposed to the same antigen, these memory cells participate in a secondary response in which they replicate and eliminate infected cells. Some of these secondary effectors then revert to memory cells.

allows the model to represent billions of discrete cells in a few thousand bytes of computer memory.

2.1 T cell biology

Cytotoxic T cells (CTLs) are essential for the control of viral infections. The life cycle of CTLs is summarized in Figure 2.1. Naïve T cells, or cells that have not yet been exposed to antigen, circulate through the body looking for antigen presenting cells (APCs) that express indications that the body is infected by virus. Once they receive stimulation from APCs and from another class of T cells known as helper T cells, CTLs begin their response to the infection. These stimulated CTLs, known as effector cells, circulate throughout the body to eliminate cells that are infected by the virus. It is their unique ability to distinguish between infected and uninfected cells that allows them to eliminate cells that harbor virus.

2.1.1 T cell receptors and repertoire

T cells have the unique ability to non-invasively view a cell's contents, allowing them to detect intra-cellular pathogens, because most cells present portions of their internal proteins on their cell surfaces. Presentation takes place when a cell processes a sample of its internal proteins into short peptide fragments that form complexes with cell surface proteins called major histocompatibility (MHC) class I molecules. There are hundreds of MHC class I alleles in humans (Marsh et al., 2002), and an individual can express as many as six of them. Each MHC type binds a particular set of peptides and is thus capable of presenting a different set of peptides than other MHCs. When a CTL binds to peptide–MHC complexes, it can initiate a series of actions that leads to the destruction of the infected cell.

One of the primary factors that determines whether a T cell binds to a cell is the *affinity* of its T cell receptor (TCR) for the peptide–MHC complexes. Each T cell expresses thousands of copies of identical receptors that bind to their cognate peptide–MHC complexes with high affinity. Thus, both the target cell peptides and the particular MHC type that presents the peptide play a role in determining affinity. The set of all TCR specificities in a body, on the order of 10^7 in humans (Arstila et al., 1999) and 10^6 in mice (Pannetier et al., 1993), comprise the T cell repertoire. *Avidity*, or the sum of the binding interactions between the receptors of a CTL and the surface of a target cell, determines whether a CTL recognizes the target. The number of copies of a particular peptide displayed by a target cell (its expression density) affects the avidity of the interaction. Due to thymic selection, described below, it is unlikely that a T cell will react to an uninfected cell—infected cells express foreign (e.g., virally encoded) peptides that make them subject to T cell responses. The antigenic peptides that stimulate T cells are known as *epitopes*.

T cell receptors are generated with seemingly random specificities, so many potentially harmful self-reactive ones are created. Most are screened out early in their maturation

process in the thymus, where they are exposed to a large array of the body's own peptides presented on MHC molecules. During *positive selection*, T cells that have an extremely weak avidity to self peptides bound to MHC are eliminated (Blackman et al., 1990). It is believed that this process eliminates T cells that have such poor avidity to MHC that they would not bind to any peptide–MHC pairs. *Negative selection* eliminates those that bind too tightly to MHC–self peptides (Kappler et al., 1987), ensuring that potentially self-reactive T cells are eliminated. This process might create "holes" in the repertoire that would allow antigenic epitopes that are sufficiently similar to self peptides to escape immune system detection. About 1–3% of pre-selection T cells pass both these "tests" and leave the thymus to join the peripheral repertoire as naïve T cells (Shortman et al., 1990).

2.1.2 T cell response

A naïve T cell remains quiescent until it receives antigenic stimulation from its cognate peptide–MHC complex. Larger antigen doses stimulate a greater fraction of naïve cells (perhaps by recruiting more low-affinity T cells) but probably do not affect the degree to which the individual cells are stimulated (Kaech and Ahmed, 2001). In other words, stimulation of individual cells might be "all or nothing." After stimulation, naïve cells appear to be committed to a *programmed response* that causes them to divide and acquire effector functions even in the absence of continuing antigenic stimulation (Kaech and Ahmed, 2001; van Stipdonk et al., 2001). For the first 24 hours, they do not replicate (Oehen and Brduscha-Riem, 1998; Gett and Hodgkin, 2000; Veiga-Fernandes et al., 2000; van Stipdonk et al., 2001), but after this initial phase, they can rapidly undergo a fixed number of divisions (up to 8 or more) (Kaech and Ahmed, 2001) once every 5 to 8 hours (Murali-Krishna et al., 1998; Gett and Hodgkin, 2000; van Stipdonk et al., 2001). After a few divisions, they acquire effector functions, such as cytotoxicity (Opferman et al., 1999; Auphan-Anezin et al., 2003). Effector CTLs kill target cells either by releasing perforins that create holes in the target cell's membrane

or by triggering apoptosis (i.e., cell suicide) in the target cell. Even during this period of rapid expansion, the cells have a high death rate, reducing net population growth. After initial expansion, the death rate dominates CTL kinetics and the population declines rapidly (Badovinac et al., 2002).

Effector cells can become impaired by over-stimulation by antigen. High doses or repetitive stimulation can cause *activation-induced cell death* in T cells (Shi et al., 1989; Hildeman et al., 2002). The death of over-stimulated cells causes the overall CTL response to diminish or disappear within a few days, a phenomenon known as *exhaustion* (Moskophidis et al., 1993). Chronic infection has been found to cause a progressive loss of function in effector CTL, starting with the inability to produce certain cytokines and ending in T cell death (Fuller and Zajac, 2003; Wherry et al., 2003). The memory cells created in the presence of antigen might also be impaired (Wherry et al., 2002), which indicates that the impairment could be an intrinsic property of the cell that does not change when antigen is removed.

2.1.3 T cell memory

After the activation and proliferation in response to an infection, most of the T cells activated in the response die, but a small subpopulation persists as *memory cells* (Murali-Krishna et al., 1999). Memory cells are able to mount a quicker and more aggressive response in future encounters with the same or closely related pathogens (Dutton et al., 1998). This *secondary response* can clear an infection before significant damage is inflicted upon the body. Immunological memory forms the basis of vaccination, in which an organism is exposed to viral antigens in order to build immune memory to the virus.

All effector T cells involved in a response to antigen appear to have the same probability of converting to memory cells cells (Sourdive et al., 1998; Busch et al., 1998a;

Blattman et al., 2000). The clonal composition of responding effectors in a primary infection is thus "mirrored" in the resulting memory population. However, this repertoire can be altered in the secondary response to antigen (Bousso et al., 2000). Some memory cell clones are preferentially recruited into the secondary response, resulting in their increased representation in memory.

It takes 2 or 3 weeks for a CTL to turn into a memory cell after the initial infection (Kaech et al., 2002). Therefore, memory cells are not likely to join the immune response that initially generated them. CTLs can die or form defective memory cells in the presence of persistent infection (Masopust et al., 2004). Therefore, if the immune system can not eliminate antigen quickly, the formation of immunological memory can be impaired.

Upon antigenic stimulation, memory cells begin to proliferate almost immediately and develop cytotoxicity within a few hours (Bachmann et al., 1999; Barber et al., 2003; Byers et al., 2003). They probably have the same sensitivity to antigen as naïve cells (Bachmann et al., 1999; Kersh et al., 2003), although some studies found their stimulation requirements to be lower (Pihlgren et al., 1996). Their replication rates are approximately the same as recently activated naïve cells. Memory cell-derived effectors die at a slower rate than effectors created in the primary response (Veiga-Fernandes et al., 2000; Grayson et al., 2002), giving them a faster accumulation rate and possibly allowing a larger portion of them to revert to memory. Presumably the shorter time to acquire effector functions, the larger starting populations, and their faster accumulation rates allow memory cells to clear infected cells much faster than naïve cells.

Homeostatic mechanisms appear to regulate the size of the memory pool, which remains approximately constant in size throughout an organism's lifetime (Rocha et al., 1989). New memory cells from heterologous infections appear to displace the memory cells from responses to prior infections (Selin et al., 1996; McNally et al., 2001). In the absence of immune system challenges, memory cells turn over slowly (Tough and Sprent,

1994; Dutton et al., 1998; Murali-Krishna et al., 1999).

2.1.4 Lymphocytic choriomeningitis virus: A model pathogen

Much of what we know about CTL responses in vivo comes from studies of lymphocytic choriomeningitis virus (LCMV) infections in mice (Armstrong and Lillie, 1934; Traub, 1935). LCMV stimulates a well-characterized CTL-mediated immune response, and infection is generally asymptomatic and eliminated quickly by the immune response (Lehmann-Grube, 1988). However, high doses of certain LCMV strains can cause chronic infection, resulting in immune exhaustion (Moskophidis et al., 1993). Because its major epitopes have been identified, the responses of cells specific to each epitope can be studied (Butz and Bevan, 1998). It is assumed that the CTL response in humans is analogous to that seen in this animal model. Using LCMV in inbred mouse lines as a model system allows researchers to observe CTL behavior in greater detail than would be possible in humans. The computer model described in Chapters 3 and 4 is calibrated using mouse data for this reason, but the model can be recalibrated using human data when it becomes available.

2.2 Related work

Most immunological models can be classified into two categories: differential equation models and agent-based models. Differential equation models have a long history of success in immunology and other fields, but they have many shortcomings, listed in Section 2.2.1. Agent-based modeling is a relatively new approach, and its strengths complement mathematical models. Only a handful of agent-based models of the adaptive immune system exist, and representatives are described in Section 2.2.2. Agent-based modeling is computationally expensive, especially if one wants to simulate billions of immune cells,

so I chose to use an efficient stochastic stage-structured approach to modeling, outlined in Section 2.2.3.

2.2.1 Differential equation models

Differential equation models have long been used for immune system and viral infection modeling (Bell, 1970; Dibrov et al., 1977; Přikrylová et al., 1992; Perelson and Weisbuch, 1992; Ho et al., 1995; Nowak and Bangham, 1996; Bocharov, 1998; Perelson, 2002). In most of these models, populations of antigens and immune cells are represented as continuous variables, and systems of ordinary differential equations (ODEs) define their behaviors over time. Analytical techniques allow modelers to define regimes of system behavior and their associated parameters and initial conditions. For example, one can determine the model parameters for which an infection is effectively cleared by the immune system (Bocharov, 1998). The solutions capture the average behavior of large populations of perfectly mixed, identical individuals. Many techniques that could make these models more faithful to biological reality, such as adding time delays or age-structured partial differential equations (Antia et al., 2003), complicate solving the models analytically or numerically.

There are many simple differential equation models of the T cell response to antigen, several of which are reviewed in Nowak and May (2000) and Perelson (2002). These models are generally single-purpose models, by which I mean that they are purpose-built to match a small set of experimental data. Two differential equation T cell models are particularly closely related to my work. One, by Bocharov, fits a large set of T cell and virus data gathered from mice challenged with LCMV (Bocharov, 1998). The second, by Antia et al., is a model of the antigen-independent, or programmed, T cell response to antigen (Antia et al., 2003).

Bocharov (1998) describes an ODE model of the murine CTL response to LCMV.

In this model, there are 3 main variables: a virus population, a precursor CTL population (including both naïve and memory cells), and a non-replicating effector CTL population. The presence of virus induces precursor CTLs to proliferate and convert to effector CTLs, which clear virus. The effector population declines due to their lytic interactions with virus, activation-induced cell death (AICD) from exposure to high viral loads, and their own limited life span. The model was calibrated using experimental data from low-, moderate-, and high-dose infections of LCMV-D in C57BL/6 mice. A later version of the model included compartments representing different organs in a mouse (Bocharov et al., 2003).

As with most differential equation models, these models are stateless. In other words, the system has no memory and its behavior is determined solely by its current state. However, some immunological phenomena require the use of state, and these can be captured in these models by using delay differential equations. For example, it is assumed that prolonged high levels of antigen induce anergy in T cells. Therefore, the attrition due to anergy in the precursor T cell population is the product of the current population, the current viral load, and the viral load at time $t - \tau$, where t is the current time and τ is a constant. Thus, precursor T cell levels will decline when exposed to virus over time interval τ, but not when the interval is less than τ. Although the mathematical representation of this term is simple, the assumptions that it entails are not. Because the term depends only on the viral load at two time points, the viral load before or between these points has no effect. The "real-world" interpretation of delay differential terms is not obvious.

To my knowledge, the models described in Antia et al. (2003) are the first to include the programmed response of T cells. This inclusion is significant because it allows T cells to have state. Without state, T cell growth would be strictly antigen-dependent, only proliferating in the presence of antigen. The addition of state allows T cells to continue proliferating when the antigen load diminishes. In effect, the T cell response has "momentum," which makes it robust to fluctuations in antigen load. Antia et al. (2003) describes

two implementations of the same model, one as a partial differential equation and one as a set of ordinary differential equations. The results of the two are qualitatively similar. The models are at an early stage of development. In their discussion, the authors enumerate many extensions to their model that would make it more realistic. Many of these extensions are already implemented in my model, including a simple cell cycle model based on Smith and Martin (1973), explicit processes for cell division and death rather than a single net population growth process, and a one-day time lag before a cell's first division.

2.2.2 Agent-based models

Agent-based simulation is a promising technique made feasible with the advent of greater computer power. These simulations monitor the actions of a large number of simple entities, or agents, in order to observe their aggregate behavior. Each agent consists of state variables and a set of rules that governs its behavior, and agents can interact either directly with each other or indirectly through the environment. Because all individuals in a population are explicitly represented, they can have unique histories and behaviors. The combined behavior of these agents is observed in a simulation.

Agent-based modeling has many features suited to modeling the immune response. It is adept at incorporating stochastic events, which appear to be crucial in regulating immune function (Germain, 2001). A single chance event, such as the serendipitous recognition of a cancer antigen by a single cell in the immune system, can determine the fate of an organism (Ochsenbein et al., 2001). The addition of randomness to a model allows one to explore the distribution of possible outcomes, as in Detours and Perelson (2000), as opposed to only the single most likely one addressed by most mathematical models. This is especially valuable when studying immune responses, as even genetically identical individuals can exhibit different responses to the same antigen (Lin and Welsh, 1998; Bousso et al., 1998). Because small numbers of cells are involved in the beginning of an

immune response (Ehl et al., 1998; Bousso et al., 1999), using a discrete model might be more suitable in this context than a continuous one. The existing agent-based models of the adaptive immune system, such as IMMSIM (Celada and Seiden, 1992; Seiden and Celada, 1992; Kleinstein and Seiden, 2000), the B cell model of Smith et al. (Smith et al., 1999), and the self-nonself discrimination model of Langman and Cohn (Cohn et al., 2002; Langman et al., 2003), take advantage of these features. Another advantage of agent-based models is that by explicitly representing individual cells, they are in many ways closer to the modeled system. In contrast to population-level models, agent-based model parameters correspond to actual properties of the cells, and the output of these models can be processed so that they can be observed at any level, from the level of the individual cell to the whole organism.

An early immunological model described in Farmer et al. (1986) represents idiotypic network (Jerne, 1974) interactions among B cells. The authors outline the similarities between idiotypic networks and the classifier systems of Holland (Holland, 1986). The work introduces the use of binary strings to represent epitopes and receptors. A string match rule determines whether a receptor binds to an epitope based on the distance between their associated strings. If their strings are complementary, or nearly complementary, the receptor binds the epitope. Many other immunological models, including mine (Chapter 4), have adopted similar string representations of epitopes and receptors.

The most mature agent-based model of the immune system is probably IMMSIM (Celada and Seiden, 1992; Seiden and Celada, 1992). It is described as a "generalized" or "hyper" cellular automata model of the immune system, but within the individual "sites" it behaves like a typical spatially implicit agent-based system. Each site is populated with various kinds of entities, such as T cells, B cells, antigens, and antibodies. At each time step, each has a chance either to perform an action. These actions include interacting stochastically with other occupants of the same site and migrating to other sites. Thus, each site behaves like a well-mixed portion of an organism's immune system. Each

entity is associated with a bit-string, representing its receptors (if it is an immune cell) or epitopes (if it is an antigen). The likelihood of interactions is determined by a string match rule that quantifies the similarity between a receptor and an epitope. The IMMSIM group has published several papers that propose explanations for immunological observations (Morpurgo et al., 1995; Stewart et al., 1997; Kohler et al., 2000).

My model addresses only a subset of the immunological phenomena that IMMSIM does. While IMMSIM simulates a system with antigen presenting cells, B cells, helper T cells, and cytotoxic T cells, my model only attempts to capture the dynamics of cytotoxic T cells. I also choose not to include spatial effects because there is insufficient laboratory data to calibrate the distribution and movements of T cells throughout an organism. The smaller scope of my model allows the simulation to be run much more quickly, and thus more often, so the distribution of thousands of outcomes can be studied. Even more importantly, by limiting my model to CTL response, it can be more easily calibrated with empirical data. Many of the components of IMMSIM can not be accurately calibrated because the model includes such a large variety of cells and their interactions. There are many behaviors that are not yet quantified in biological systems, so IMMSIM must use arbitrary values. Although one can use models such as IMMSIM to make estimates of these unknown quantities (e.g., by running parameters sweeps and Monte Carlo simulations), the task becomes infeasible when there are too many unknown parameters.

While qualitative models might expose novel mechanisms that can explain certain phenomena, I believe that more useful predictions can be made by carefully calibrating the model with real-world data. For example, my model uses a realistic-sized T cell repertoire, while IMMSIM simulates an artificially small repertoire. Qualitative models give little indication of the frequency or magnitude of events. Quantitative modeling is essential in studying the immune system. Small differences in the quantity of a pathogen exposed to the immune system can mean the difference between immune system tolerance and a vigorous immune response. Slight changes in growth rates can greatly affect the outcome of

an immune response, as previous immune system simulations have postulated that there is a "race" between a rapidly reproducing pathogen and the immune response (Smith et al., 1999; Kohler et al., 2000). Seemingly minor and improbable events can trigger a cascade with significant consequences (Germain, 2001).

Derek Smith implemented a spatially implicit B cell model for studying influenza (Smith et al., 1999). In his model, all B cell receptors and antigenic epitopes are represented by strings. Binding affinity is determined by the Hamming distance (defined in Section A.1) between receptor and epitope strings. A high affinity match will cause a B cell to replicate with a high mutation rate (somatic hypermutation). These B cells release antibodies, which neutralize the antigen. An important contribution of Smith's work is the use of "lazy evaluation" to allow the model to accommodate a realistic-sized repertoire (Smith et al., 1998). The principle of lazy evaluation is to perform only the computations that are needed by the final result. Smith noted that the only B cells that are recruited into a response are those that bind sufficiently well to the antigen, and the remaining B cells are quiescent. In the model, these quiescent cells do not need to be instantiated. Therefore, rather than create a simulation with all $10^7 - 10^8$ B cell clones with distinct receptors, one only needs to include the $10^2 - 10^3$ that could actually respond to the antigen. The receptors of these responding cells are generated by creating random strings uniformly distributed close to the epitope. In the past, modelers would either need substantial computing resources to simulate a realistic number of cells (Detours and Perelson, 2000; Bernaschi and Castiglione, 2001) or use artificially small repertoires (Kleinstein and Seiden, 2000).

I adapted Smith's lazy evaluation technique to create only the CTLs that can respond to the antigens in the simulation. CTLs, unlike B cells, are subject to thymic selection and bind to MHC in addition to the antigen. These issues are solved by the models of Detours, described below. I take an additional step to reduce the computational cost of immunological simulation. Although the lazy evaluation technique reduces the number of clones, it

does not greatly reduce the number of cells involved in the response. Although the number of clones responding to an antigen is small, the number of cells is not. A handful of clones can generate millions of responding T cells. To make his simulation manageable, Smith makes each "agent" in his simulation represent 10 cells (Smith, 1997). Because the T cell response can be initiated by 25–50 cells (Ehl et al., 1998), a granularity of 10 cells might be too coarse—it could be important to allow cells to respond individually, not as groups. Section 2.2.3 describes the stage-structured modeling technique that allows me to achieve this fine granularity efficiently.

In contrast to Smith's dynamic model of B cell responses, Vincent Detours' model investigates the static properties of the naïve T cell repertoire without including response to antigen (Detours et al., 1999). This model uses strings to represent both the antigenic peptides and the portion of the MHC molecules that come into contact with the TCR. These two strings are concatenated then compared to the strings representing the TCRs to determine their affinity. Detours greatly increased the computational efficiency of his model by extending Smith's lazy generation technique to take thymic selection into account (Detours et al., 2001). In Smith's original algorithm, it was assumed that B cell receptor strings are distributed randomly across the universe of strings. Including thymic selection would violate this assumption in two ways. The first is that T cell receptors must bind to MHC as well as peptide, so TCRs have a non-random affinity to MHC. The second is that the T cell receptors must also have an intermediate affinity to self peptides because of positive and negative selection. Thus, the T cell receptors that respond to an epitope are not uniformly distributed—they are influenced by both the MHC and self peptide strings. Detours' complex algorithm takes these effects into account and can efficiently generate a TCR repertoire for a particular MHC–peptide string. His implementation is specific to his "xor" string matching rule (defined in Section A.2), in which the affinity between two strings is the sum of the bitwise xor of their digits. The parameters of the model are calibrated using real-world data (Detours et al., 2000) in order to allow for the quantitative exploration of certain T cell repertoire properties, such as alloreactiv-

ity (Detours and Perelson, 1999, 2000).

I have implemented his efficient T cell repertoire generation algorithm and generalized his algorithm to two other string matching rules: Hamming and Manhattan (described in Appendix A). For the alternative match rules, I use an algorithm that is less efficient but also less complex than Detours' original xor implementation. My implementation is described in Section 4.2.

The disadvantage of using other match rules is that the parameters used by the Detours et al. (2001) algorithm must be modified. The parameters are not easy to compute, and it is difficult to validate them except by comparing a variety of average statistics of the outputs from lazy and eager versions of the model. There is the danger that the lazy repertoire differs from the fully evaluated one in important but subtle ways that are not detected by the chosen statistical measures. Therefore, I have decided to choose a simpler but less computationally efficient approach. Because the pre-selection TCR repertoire is random and uniform over the universe of TCRs, it can be generated lazily in exactly the same manner as the B cell repertoire in Smith's work. The distances between each TCR from the pre-selection repertoire and all of the MHC–self peptide complexes can be computed to determine which cells survive to join the naïve repertoire. This approach can generate up to 100 times more TCRs than will actually join the naïve pool, but it is conceptually simpler than Detours' scheme and is thus less subject to error.

2.2.3 Stochastic stage-structured modeling

For computational efficiency, I use a stochastic stage-structured approach to modeling the cytotoxic T cell population (Chao et al., 2003). Stage-structured models have been used to model populations in ecology (Lefkovitch, 1965; Usher, 1966; Manly, 1990) but have rarely been applied to immune systems (e.g., Kleinstein and Singh (2001)). In stage-structured models, an individual's or cell's life cycle is divided into stages, such as devel-

opmental maturity or differentiation states. All individuals in a given stage are assumed to be identical. The transition probabilities between stages are specified, and at each time step, these probabilities are used to determine how many of the individuals in each stage transition to another stage. Stochasticity can be added to the model if needed, and the number of individuals that transition between two stages in a time step can be determined by drawing from a random distribution. Analytical techniques have been developed for studying these models, but when there are interacting populations (e.g., T cells and antigens), it is often easier to simply run the model on a computer multiple times and observe the distribution of outcomes. My modeling approach is described in more detail in Chapter 3.

By using discrete rather than continuous population variables and by explicitly specifying the actions and transitions of cells as probabilities per individual cell, my model enforces the realistic behavior of individual cells without the computational cost of representing each cell explicitly. The model attempts to strike a balance between the unrealistically small number of populations used by the analytical approaches described in Section 2.2.1 and and the unwieldy one-agent-per-cell implementations of the agent-based models described in Section 2.2.2. Because I do not intend to solve my system analytically, the model can accommodate multiple cell states. However, to make the model more efficient than an equivalent agent-based model, the number of possible cell states is reduced to a manageable number (described in section 3.2.4).

Chapter 3

The model

Science is what we understand well enough to explain to a computer. Art is everything else we do.

—Donald Knuth, from the Foreword to *A=B* by Marko Petkovsek, Herbert Wilf, and Doron Zeilberger

When I model I pretty much go blank. You can't think too much or it doesn't work.

—Paulina Porizkova

In this chapter, I describe my model of CTL response to infection. The model has two main subcomponents. One is a difference equation virus infection model (Section 3.1). In this model, virus infects healthy target cells, and the infected cells produce more virus. The other component is a stochastic stage-structured T cell life cycle model (Section 3.2). After T cells in the model are first stimulated by infected cells, they progress through a series of stages of differentiation, in which they proliferate, eliminate infected cells, then convert to memory cells. Including the T cell life cycle results in a more realistic portrayal of the dynamics of an immune response.

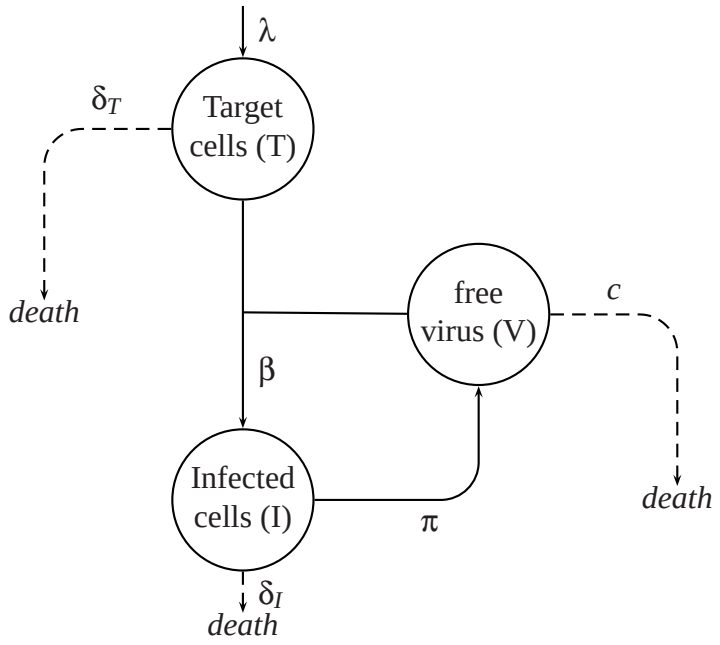


Figure 3.1: The virus infection model in the absence of an immune response. Virus (V) infects target cells (T), which become infected cells (I). Infected cells produce virus. A constant source replenishes the target cell population.

3.1 Virus dynamics

I adopt a standard model of viral infection previously used to describe human immunodeficiency virus (HIV) and hepatitis C virus (HCV) dynamics (Wei et al., 1995; Perelson et al., 1996; Neumann et al., 1998). In the absence of an immune response, the course of a viral infection is described by the following:

$$\dot{T} = \lambda - \delta_T T - \beta TV, \tag{3.1}$$

$$\dot{I} = \beta TV - \delta_I I, \tag{3.2}$$

$$\dot{V} = \pi I - cV \tag{3.3}$$

where T is the number of uninfected (or "target") cells, I is the number of infected cells, V is the number of virus particles, λ is the rate of uninfected cell production, π is the rate of virus production by infected cells, β is the infectivity parameter, δ_T is the death rate for target cells, δ_I is the death rate for infected cells, and c is the clearance rate for free virus. The system is represented schematically in Figure 3.1. Typically, after infection the viral load and the number of infected cells increase exponentially, peak, and then decline.

In my implementation, I use a difference equation version of the system of ODEs described by equations 3.1-3.3:

$$\Delta T = (\lambda - \delta_T T - \beta T V)\Delta t, \tag{3.4}$$

$$\Delta I = (\beta T V - \delta_I I)\Delta t, \tag{3.5}$$

$$\Delta V = (\pi I - cV)\Delta t \tag{3.6}$$

where $\Delta t = 10$ minutes. In order to include stochasticity, the terms in equations 3.4-3.6 are randomly drawn from the appropriate distributions at each time step, an approach similar to that taken in Kleinstein and Singh (2001). I assume that the variables are constant over the short interval Δt and are updated at the end of each time step. I also randomly permute the order in which the different infectious agents are updated using a Fisher-Yates shuffle (Fisher and Yates, 1938). The shuffling should eliminate any bias caused by the order in which these agents are updated. In Salfi (1974), the author notes that not all possible permutations can be created using such algorithms unless the random seed is extremely large. I believe that this potential problem does not noticeably affect the behavior of my model's implementation. For the production of uninfected cells and the virus production rate, I assume that they are governed by Poisson processes, and I draw from the Poisson distribution with their expected values as the mean (i.e., $\lambda\Delta t$ and $\pi I\Delta t$, respectively).

To stochastically determine the number of cells out of a population of identical cells that perform a certain action, such as dying, I draw randomly from the binomial distribution. In order to do this, I must convert continuous rates into probabilities that events

occur during a time step. If a process occurs at rate r, then the probability that it first occurs at time t is defined by the exponential distribution $\mathcal{E}(r) = re^{-rt}$. The probability that it occurs at or before time t is $1 - e^{-rt}$. Thus, rates r can be converted to probabilities that the processes occur in a time step Δt, $1 - e^{-r\Delta t}$. If there are n cells each with a probability p of performing an action, then drawing from the binomial $\mathcal{B}(n, p)$ is a computationally efficient way to determine the number of cells that perform the action. For example, I compute the number of uninfected cells T that are infected in each time step by converting their infection rate, βV, to the probability that they will become infected in a time step, $1 - e^{-\beta V \Delta t}$, and randomly drawing a value from $\mathcal{B}(T, 1 - e^{-\beta V \Delta t})$.

To validate my implementation of the infection dynamics difference equation model, I compared its results to an alternate version using Gillespie's Direct Method (Gillespie, 1977), which is an exact stochastic simulation technique that explicitly generates all discrete events rather than computing how many reactions occur in a given time step. Gillespie developed two algorithms for exact stochastic simulation of chemical reactions, the Direct Method and the First Reaction Method. Gibson and Bruck (2000) contains a good explanation of both. I use the Direct Method because it is more computationally efficient.

To convert the virus infection model to a Direct Method simulation, the difference equations are expressed as a set of parallel reactions:

$$\xrightarrow{\lambda} \quad T+1, I, V \tag{3.7}$$

$$T \xrightarrow{\delta_T} T-1, I, V \tag{3.8}$$

$$TV \xrightarrow{\beta} T-1, I+1, V \tag{3.9}$$

$$I \xrightarrow{\delta_I} T, I-1, V \tag{3.10}$$

$$I \xrightarrow{\pi} T, I, V+1 \tag{3.11}$$

$$V \xrightarrow{c} T, I, V-1 \tag{3.12}$$

Each of these reactions represents the conversion of reactants (terms on the left of the

arrow) to products (terms on the right). The products are written in terms of how a reaction affects each of the three state variables T, I, and V. For example, $T-1$ indicates that the number of target cells is decremented by 1, $V+1$ indicates that the number of virus are incremented by 1, and I indicates that the number of infected cells is not affected by the reaction. Reactions occur at rates proportional to the product of the quantities of the inputs times the reaction rate constant, which is written over the arrow of each reaction. For example, reaction 3.9 proceeds at rate βTV, and each "reaction event" eliminates one target cell T and produces one infected cell I while the virus level V remains constant. Reaction 3.7 does not depend on the presence of any inputs, so its rate is λ and it increases the number of target cells T by one.

The Direct Method simulation is initialized by setting the target cell, infected cell, and virus levels to the desired levels and setting the time elapsed to 0. One advances the state of the simulation by choosing one of the six reactions to occur, changing the state of the system according to the reaction chosen, and incrementing elapsed simulation time. For each iteration, a reaction is chosen randomly with probabilities proportional to their reaction rates. Thus, the fastest reaction is most likely to be selected, but the slowest reaction can be chosen. Note that these reaction rates are not constant—most of them depend on the current numbers of target cells, infected cells, and virus—so they need to be computed each time before a reaction is chosen. Once a reaction is chosen, the state of the system (i.e., T, I, and V) is updated to reflect the effects of the chosen reaction. The simulation time must then be advanced. One would expect that the time increment would depend on the speed of the reaction chosen, but it does not. The time increment is simply drawn randomly from an exponential distribution with the sum of all six reaction rates as its parameter: $\mathcal{E}(\lambda+\delta_T T+\beta TV+\delta_I I+\pi I+cV)$. This time is added to the total time elapsed. At this point, the effects on the system of performing one reaction have been computed. Subsequent reactions can be chosen in the same manner, typically until the simulation time reaches the desired value.

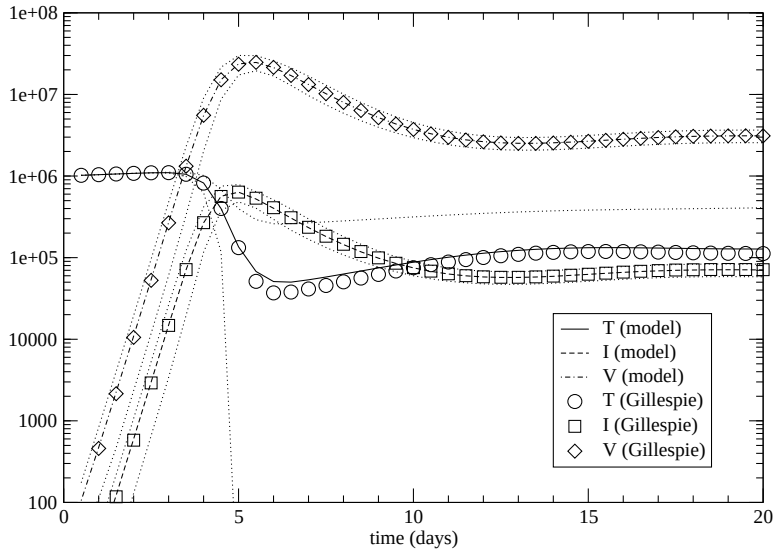


Figure 3.2: Comparison of the average outputs from the difference equation model and the Gillespie Direct method version. The averages (solid, dashed, and dot-dash lines) and standard deviations (indicated by the fine dotted lines) of 100 runs of the difference equation model and the averages from the Gillespie version (symbols with no lines) are plotted. The standard deviations from the Gillespie version are omitted for clarity. The variance for the number of target cells (T) is high because in some cases the virus would not infect any cells at all and the number of target cells remained constant (and high).

I compared the outputs of the Direct Method simulation with the difference equation version. The means and variances of the final outcomes appeared to be the same for various initial conditions (Figure 3.2), but I was also concerned about the distribution of outcomes, not just the low-order moments. To compare the distributions, I ran both versions 100 times then took the histograms of the final target cell, infected cell, and virus levels. I initialized both systems with a small number of viruses (50) so that the variance would be higher and the distribution of outcomes broader. The other parameters corresponded to a typical acute virus infection: $T_0 = 10^6$, $I_0 = 0$, $V_0 = 50$, $\lambda = 50000$, $\delta_T = 0.01$, $\beta = 2 \times 10^{-7}$, $\delta_I = 0.7$, $\pi = 100$, and $c = 2.3$. I recorded the system state at the beginnings of days 2 and 5. The results from my simulation and the Gillespie model seemed to have the same distributions at both time points (Figure 3.3).

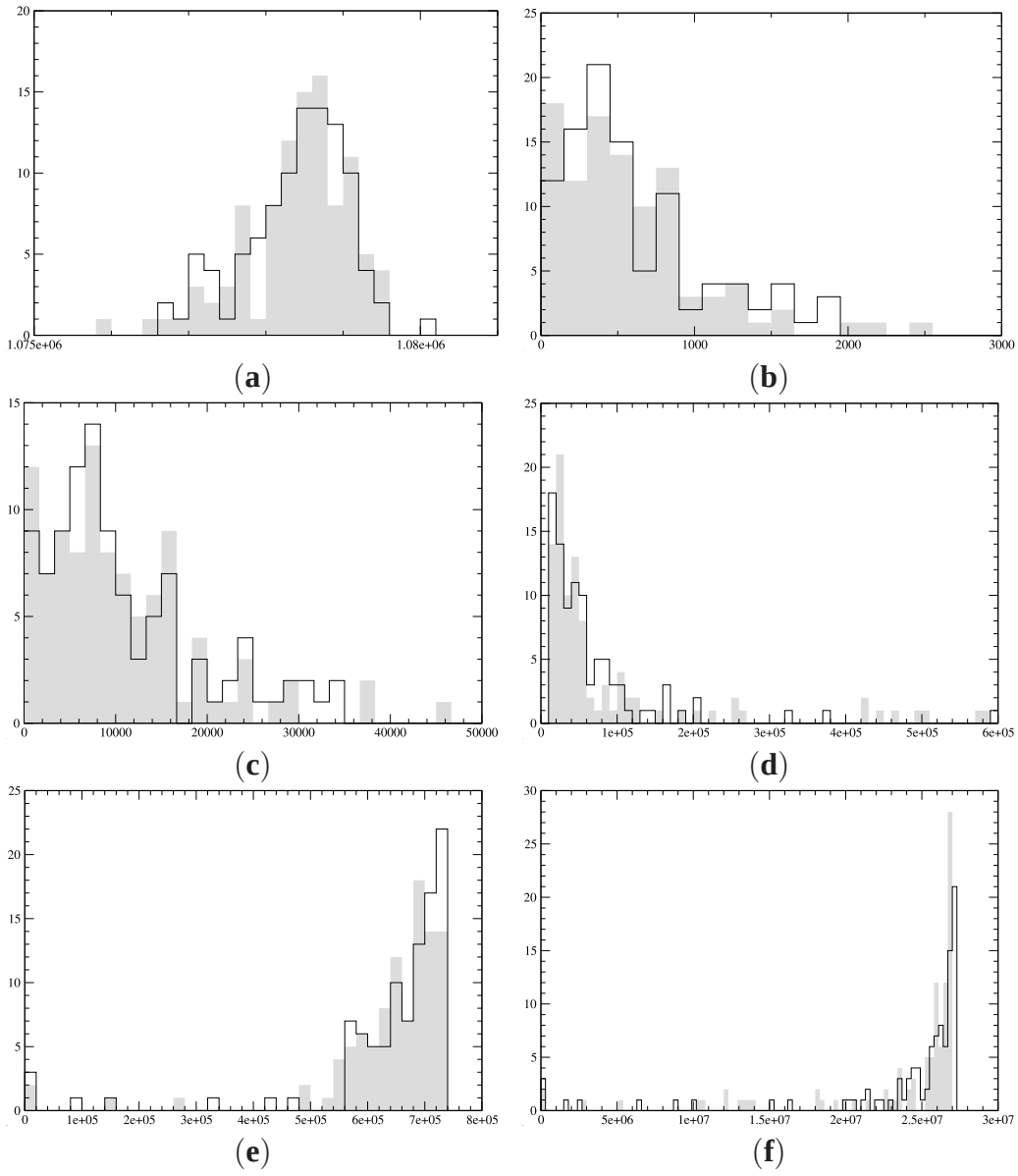


Figure 3.3: Comparison of the distribution of outputs of the difference equation model and the Gillespie Direct Method version. The histograms (a)–(c) show the distribution of target cell, infected cell, and virus populations at the beginning of day 2, while histograms (d)–(f) are for the beginning of day 5. The filled gray histograms represent the Direct Method outputs, while the open histogram bars are the difference equation outputs. The averages from these runs are shown in Figure 3.2.

The running time for a naïve implementation of the Gillespie Direct Method is at least three orders of magnitude longer than the difference equation model I use. In my model, it is assumed all probabilities (such as the probability that a cell becomes infected) can be treated as constant during a time step. Using this assumption, the effects of the actions of a population of identical cells during one time step can be computed in one operation. Using larger time steps (such as 30 minutes) decreased running time but produced a noticeably different distribution of outcomes than the Gillespie version, so I chose a 10-minute time step for the simulation runs.

3.2 The T cell life cycle

CTL dynamics are represented in a stochastic stage-structured model of T cell activation, proliferation, and differentiation. Infected cells from the infection dynamics model (described in Section 3.1) stimulate naïve T cells and are killed by effector T cells (depicted in Figure 3.4). The degree of T cell stimulation and infected cell clearance are determined by receptor binding rules.

3.2.1 Receptor binding

CTLs detect antigens when their TCRs bind sufficiently well to MHC–epitope complexes on the surfaces of the infected cells (Section 2.1.1). In the model, each antigen is associated with one or more epitopes, and each epitope is associated with an MHC type. A cell infected by this antigen expresses these epitopes, which reveal to the immune system that the cell contains pathogens. Each CTL in the model is associated with a single TCR specificity (implying that each CTL expresses one kind of TCR), which can detect a particular epitope. The strength of the binding interaction between the CTL's TCRs and the MHC–epitope complexes is defined to be the binding affinity. A TCR has high affinity for its

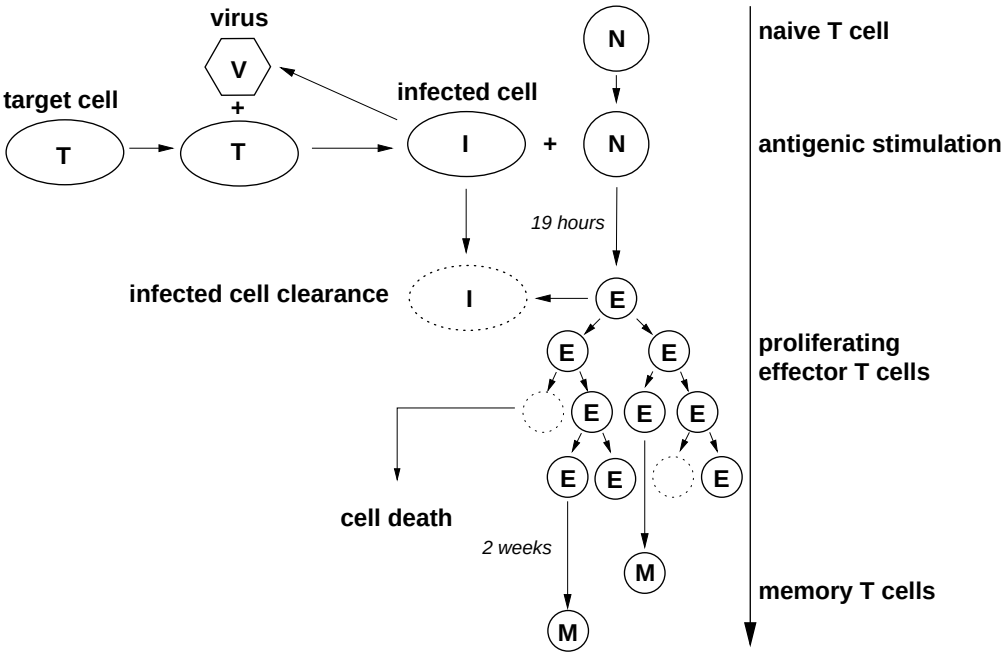


Figure 3.4: The process of infection and the life cycle of CTLs in the model. Target cells are infected by virus, and these infected cells generate more virus and interact with T cells. Naïve cells, when stimulated by antigen proliferate and become effector cells. The probability of a naïve cell being stimulated by antigen depends on the string distance between the TCR and the antigen-MHC complex. Most effectors die, but about 5% of these proliferating effector cells become memory cells. The memory cells can be stimulated to become effectors in a secondary response (not shown).

cognate epitope, lower affinity for related epitopes, and no affinity for unrelated epitopes. The model assigns affinity values for each combination of TCR and epitope. The model assigns each TCR a dissociation constant for each MHC–epitope complex, and affinity is inversely proportional to the dissociation constant. The model's representation of TCRs, MHC–peptide complexes, and the affinities between them are described in Chapter 4.

A CTL successfully detects an infected cell when it has a high avidity for the cell. The avidity that a CTL in the model has for an infected cell expressing a single epitope is the product of its TCR's affinity for the epitope multiplied by a scalar value, e, representing

the epitope's surface expression density. Thus, a high-affinity interaction can result in low avidity if the epitope's density is low. I assume that all cells infected with the same pathogen have the same epitope densities, making them all equally antigenic. If an infected cell expresses multiple epitopes, a CTL's avidity for it is the sum of its avidities for the various epitopes.

3.2.2 Effector recruitment from the naïve and memory cell pools

Infected cells stimulate naïve and memory cells, causing them to differentiate into effectors. Because a relatively small number of naïve cells are recruited into an immune response, I assume that they do not compete with each other for antigen, allowing the stimulation of each naïve clone to be computed independently. Antigenic stimulation takes the form of a saturating function (De Boer et al., 2001; Davenport et al., 2002):

$$\text{Stimulation} = \frac{\sum \frac{e_i I_i}{K_i}}{1 + \sum \frac{e_i I_i}{K_i}} \tag{3.13}$$

where K_i is the amount of antigen i required to generate half-maximal stimulation for the T cell, e_i is epitope density on cells infected by antigen i, and I_i is the number of infected cells expressing antigen i in the system. This expression is in agreement with the observation that CTL recruitment is proportional to epitope density (Wherry et al., 1999), but the response magnitude does not increase after a threshold density is reached (Vijh et al., 1998). I assume that naïve T cells are recruited into the immune response at a rate of γ multiplied by the stimulation, where $\gamma = 1$ day^{-1} is the maximum recruitment rate of T cells.

Naïve T cells specific to a particular antigen are in the same stage until they are stimulated. My model accommodates T cells of different antigen specificities by instantiating separate stage-based models for each, but for the purposes of discussion I will assume that there is only one T cell specificity. If there are multiple T cell clones, their execution order

is permuted using a Fisher-Yates shuffle (Fisher and Yates, 1938), as is done for multiple infectious agents. As naïve cells are stimulated, they must wait τ_n hours, representing the developmental time before a naïve cell begins its programmed response. To implement this delay, the cells are promoted through a series of $6\tau_n$ stages, with all cells in a stage moving to the next stage at each 10-minute time step. The cells in these stages do not interact with infected cells, but when they emerge after τ_n simulation hours, they become effectors and start responding to infected cells and dividing. In my model, I assume T cells take a minimum of 5 hours to divide, and that the first T cell divisions take place 24 hours after antigenic stimulation (Oehen and Brduscha-Riem, 1998; Gett and Hodgkin, 2000; Veiga-Fernandes et al., 2000; van Stipdonk et al., 2001), so I chose $\tau_n = 19$ hours.

Memory cells are recruited in the same manner as naïve cells except that I assume it takes only one hour ($\tau_m = 1$ hour) for a stimulated memory cell to begin its programmed response, reflecting the rapid response of memory cells to pathogens (Bachmann et al., 1999; Barber et al., 2003).

3.2.3 Clearance of infected cells

Because the CTL responses to different antigenic epitopes of the same pathogen do not appear to interfere with each other (Vijh et al., 1999), I model the immune response to multiple epitopes as the sum of independent responses to the individual epitopes. Therefore, I need only define the clearance of infected cells expressing a single epitope by many T cell clones. I assume that effector T cells of clone j, E_j, bind to infected cells I in reversible reactions (at rates k^b for binding and k^d for dissociation) to form complexes C_j, and that effectors bound in these complexes clear the infected cells at rate k^c:

$$E_j + I \underset{k_j^d}{\overset{k_j^b}{\rightleftarrows}} C_j \xrightarrow{k_j^c} E_j \tag{3.14}$$

Directly translating the above expression to a differential equation:

$$\dot{C}_j = k_j^b \hat{E}_j \hat{I} - (k_j^d + k_j^c) C_j \tag{3.15}$$

where $\hat{E}_j$ and $\hat{I}$ are unbound effectors and infected cells, respectively. Changing variables to total cells and conserving the number of infected cells, as suggested in Borghans et al. (1996), gives

$$\dot{C}_j = k_j^b (E_j - C_j)(I - \sum_k C_k) - (k_j^d + k_j^c) C_j \tag{3.16}$$

where $\sum_k C_k$ is the number of complexes of all effector cells of all specificities with I. Assuming quasi-steady state:

$$0 = k_j^b (E_j I - C_j I - E_j \sum_k C_k + C_j \sum_k C_k) - (k_j^d + k_j^c) C_j \tag{3.17}$$

Following De Boer and Perelson (1995), I approximate the solution to equation 3.17 by assuming the $C_j C_k$ terms are small enough to be omitted:

$$C_j \approx \frac{E_j I - E_j \sum_k C_k}{I + K_j} \tag{3.18}$$

where $K_j = \frac{k_j^d + k_j^c}{k_j^b}$.

Following the derivation from the Appendix of De Boer and Perelson (1995), the solution to equation 3.18 when there are multiple T cell clones is:

$$C_j \approx \frac{I E_j}{K_j + I + \sum_k E_k \frac{I + K_j}{I + K_k}} \tag{3.19}$$

Therefore, the clearance rate of I due to effectors of all specificities is:

$$\dot{I} = -\sum_j k_j^c C_j \approx \sum_j -k_j^c \frac{I E_j}{K_j + I + \sum_k E_k \frac{I + K_j}{I + K_k}} \tag{3.20}$$

For a system with only one T cell clone, E:

$$\dot{I} \approx \frac{-k^c I E}{K + I + E} \tag{3.21}$$

Expression 3.21 yields a dose-response relationship between effector cell numbers and the infected cell clearance rate that saturates at $k^c I$ as the number of effector cells increases, which agrees with experimental findings (Lehmann-Grube, 1988). It also includes a term for inter-clonal competition among the effector cells for infected cells expressing a single epitope. It appears that high- and low-avidity CTL lyse their targets at similar rates (Derby et al., 2001), so I set k^c to be the same for all T cell clones in my model. In LCMV responses, the value of k^c was found to be 12 day^{-1} (Barchet et al., 2000). Smaller populations of T cells might have higher per capita killing rates, but I assume that most of an infection is resolved while the effector cell population is large. In my model, increased avidity K affects the ability to detect and bind to infected cells at low concentrations of I. Multiple T cell clones clear infected cells at the rate described by equation 3.20, in which T cells compete for access to infected cells based on their avidities to them. High-avidity clones are more effective at clearing infected cells than low-avidity clones.

I assume that effector cell mediated clearance of infected cells is a Poisson process. From equation 3.21, one can determine the expected number of infected cells to be cleared in a time interval Δt to be $\dot{I}\Delta t$. I compute the number of infected cells that are cleared during Δt by randomly drawing from the Poisson distribution $\mathcal{P}(\dot{I}\Delta t)$ at each time step. This term is subtracted from the right side of equation 3.5 to include the effect of cytotoxic T cell clearance on the infected cell population.

3.2.4 T cell replication

I implement the programmed divisions of newly activated effector cells by keeping track of the number of times a cell divides. When a naïve cell is first stimulated, it joins the cohort of effector cells that have not yet divided. When it reproduces, it is moved with its daughter to the next division cohort. I adopt the transition probability cell cycle model described by Smith and Martin (1973), which has two phases: an A phase with a variable

residence time and a B phase that takes a fixed length of time to traverse. Cells start in phase *A*, in which the cells do not divide. At each time step, a cell has a constant probability of entering phase *B*, during which it divides in a fixed amount of time. At the end of the *B* phase, both the parent cell and the new daughter cell enter the *A* phase. This two-phase model enforces a minimum time to cell division. Without the fixed length *B* phase, some cells could divide an arbitrarily large number of times in a time interval, which is a characteristic of continuous models of cell replication.

To implement the Smith-Martin cell cycle model, each division cohort is subdivided into an *A* phase and a set of *B* phase sub-cohorts (Figure 3.5). To mimic the fixed length of time it takes a cell to traverse the B phase I allocate one *B* phase sub-cohort per time step that the cells remain in *B* phase, and move cells from one sub-cohort to the next at each simulation time step (Figure 3.5). I use 10-minute time steps, so to model cells remaining in the *B* phase for *n* hours, I use $6nB$ phase sub-cohorts per division cohort. At each time step, cells in the *A* phase of each division cohort transition to the *B* phase with a fixed probability.

I assume that the average cell cycle time of an effector T cell is 6 hours and that the minimum time to division is about 5 hours (van Stipdonk et al., 2001). Therefore the duration of the *B* phase is 5 hours and the average duration of the *A* phase is 1 hour. To simulate a 5 hour *B* phase using 10-minute simulation time steps, I use 30 sub-cohorts. To mimic the one hour average residence in the *A* phase, I assume the rate at which cells in *A* phase transition to *B* phase is 1 hour^{-1}. I convert this rate to the probability that *A* phase cells will transition to *B* phase in a time step in the manner described in Section 3.1 and draw from the binomial distribution to determine how many cells performed the transition. Because T cells with different specificities seem to expand at the same rate in vivo (Busch et al., 1998b), all cells in the model share the same cell cycle parameters. When a death rate of $\delta_E = 0.6$ day^{-1} is included (Veiga-Fernandes et al., 2000), the cell population grows at a rate of 0.092 hour^{-1}, or about 9-fold per day. T cells divide for

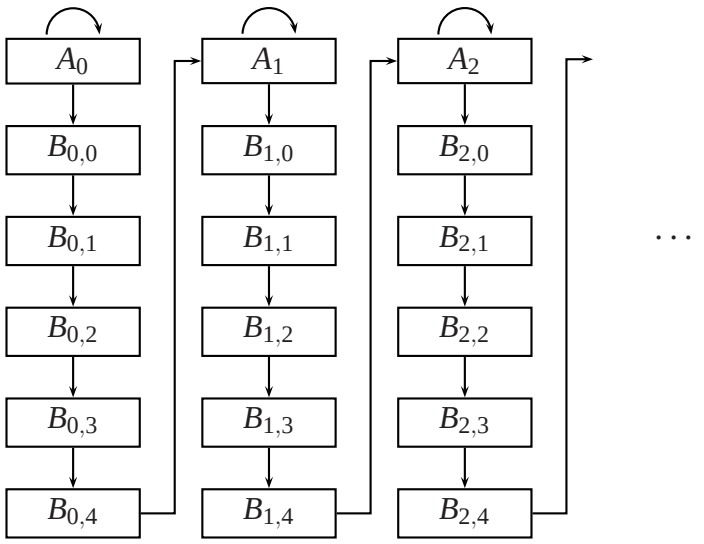


Figure 3.5: Implementation of the Smith and Martin two-phase cell cycle model (Smith and Martin, 1973). Each box represents the cells in a given stage, and the arrows represent possible transitions between stages. Note that cells in *A* phase can either remain in *A* phase or transition to *B* phase, while *B* phase cells progress at a fixed rate until they reach *A* phase. In this figure, each B sub-stage is one hour, and in the model implementation each sub-stage is 10 minutes.

about 5 days (Lehmann-Grube, 1988), which implies that a single naïve T cell can generate 60,000 effector cells, which agrees with experiment (Welsh and Selin, 2002). If one assumes that a T cell cannot divide more than 100 times, there could to be up to 3100 subpopulations of effector cells per T cell clone, or 100 A phase subpopulations and 3000 B phase subpopulations. These 3100 subpopulations efficiently represent the approximately 600,000 cells (i.e., 10 naïve cells per clone (Casrouge et al., 2000) and 60,000 effectors from each naïve cell) that can originate from a single clone in an immune response.

After their programmed divisions, the cells stop dividing (Badovinac et al., 2002). I assume that during the entire lifetime of the activated T cell, they are subject to the same high death rate δ_E. Thus, cell populations that have stopped dividing are subject to rapid population decline.

3.2.5 Memory

Some of the effector cells that proliferate during an immune response become long-lived memory cells. In the model, effector cells have a 2% per day chance of becoming memory cells after 5 cell divisions (Oehen and Brduscha-Riem, 1998; Opferman et al., 1999), which results in a final memory pool that is about 5% of the peak response (De Boer et al., 2001). The model assumes that all effector cells have an equal probability of converting to memory. Like naïve cells, memory cells become effector cells upon antigenic stimulation. I assume that memory cells have the same sensitivity to antigen as naïve cells, but they enter cell cycle only one hour after antigenic stimulation. Memory-derived effector cells have a lower death rate than naïve-derived effectors (Grayson et al., 2002), and I set this rate to be $\delta_{E_m} = 0.4$ day^{-1} (Veiga-Fernandes et al., 2000). Because memory-derived effectors have the same proliferation rate as naïve-derived effectors, this lower death rate allows them to experience higher net population growth.

3.3 Summary

A virus infection model and a CTL model interact to form a system that can simulate the CTL response to infection. The virus dynamics are adapted from a standard ODE model of infection. The CTL model captures the behavior of individual T cells, but it uses a computationally efficient stage-structured approach. Naïve CTLs are recruited into the immune response by infected cells at a rate proportional to their affinity to the antigen. Once recruited, they become effector cells, which rapidly proliferate and eliminate infected cells. After the response, some of these effector cells become long-lived memory cells while the rest die. Memory cells are dormant until they are stimulated by infected cells, after which they become effector cells. Many of the parameters used in the model are summarized in Table 3.1.

attribute	*value*
time step (Δt)	10 minutes
naïve cell clone size	10 cells[*]
maximum T cell recruitment rate (γ)	1 day^{-1}
delay before a stimulated naïve cell becomes an effector (τ_n)	19 hours[†]
delay before a stimulated memory cell becomes an effector (τ_m)	1 hour[‡]
naïve-derived active CTL death rate (δ_E)	0.6 day^{-1}[§]
memory-derived active CTL death rate (δ_{E_m})	0.4 day^{-1}[§]
time in B phase for CTL	5 hours[‖]
average CTL cell cycle time	6 hours[‖]
infected cell clearance rate (k^c)	12 day^{-1}[¶]

[*] Casrouge et al. (2000)
[†] Oehen and Brduscha-Riem (1998); Gett and Hodgkin (2000); Veiga-Fernandes et al. (2000); van Stipdonk et al. (2001)
[‡] Bachmann et al. (1999); Barber et al. (2003)
[§] Veiga-Fernandes et al. (2000)
[‖] van Stipdonk et al. (2001)
[¶] Barchet et al. (2000)

Table 3.1: A summary of model parameters.

Chapter 4

Representing the CTL repertoire

What is real is not the external form, but the essence of things... it is impossible for anyone to express anything essentially real by imitating its exterior surface.

—Constantin Brancusi

I'm afraid that if you look at a thing long enough, it loses all of its meaning.

—Andy Warhol

This chapter describes the model's abstract representation of TCR–peptide interactions that define the affinities of CTLs for infected cells. The CTL model described in Chapter 3 uses these affinity values to govern the behavior of cells. TCRs and peptides are represented as digit strings in the model, and the strength of interactions between them are determined by the similarity between their strings, as defined by a distance metric. Strings are defined in Section 4.1. The purpose of the model is not to mimic receptor–ligand binding, but to have a representation that supports a realistic number of CTL clones with different affinities to antigen. Section 4.2 is a high-level description of the procedure used

to generate the naïve CTL repertoire. This procedure is applied to create three different versions of the model, each with a different definition of string distance. Comparing the behavior of the three implementations in Chapters 5 and 6 will indicate how robust the results are to assumptions about antigenic distance. The three metrics are defined and calibrated in Appendix A. In the model, an antigen's epitopes are subject to random mutation, and this operation is defined in Section 4.3. The CTL model described in Chapter 3 requires affinity, not string distance, be defined, so the procedure for converting distance to affinity is in Section 4.4.

4.1 Strings and distances

Strings of digits represent the binding surfaces of receptors and ligands in the model, an abstraction used by several immunological models in the past (Farmer et al., 1986; Celada and Seiden, 1992; Detours et al., 1999; Smith et al., 1999; Bernaschi and Castiglione, 2001). The digits can take any value between 0 and $k-1$ inclusive, where k is the alphabet size. A random string, which one can think of as a sequence of amino acids, is generated for each self peptide in the simulation. It has been suggested that $10^3 - 10^5$ self peptides are involved in thymic selection (Bevan, 1997; Müller and Bonhoeffer, 2003; Bandeira and Faro, 2003), so the model creates 10,000 random "self peptide" strings for each of the three MHC alleles in the model. When a new antigen type is created in the model, random peptide strings are created to represent its epitopes. These strings represent the novel peptides that a cell infected with the antigen expresses. Thus, all cells infected by this antigen are associated with the same set of one or more epitope strings. Uninfected cells do not express any peptides in the model because it is assumed that CTLs do not interact with healthy cells.

The organism represented by the model has three MHC alleles. Because I assume that each distinct peptide in the "real" immune system is presented by a single MHC allele,

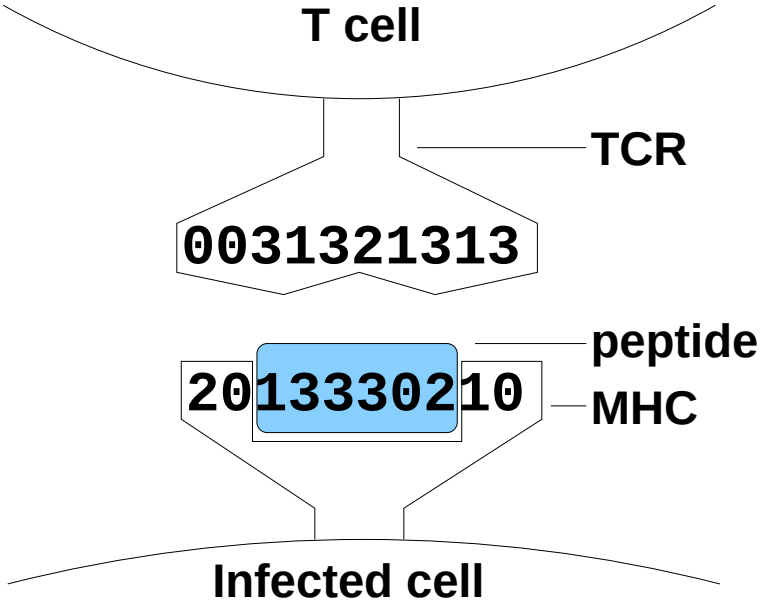


Figure 4.1: The digit string representation of TCR binding. Each T cell receptor, peptide, and MHC type is represented by a digit string. Peptide strings are concatenated with a string associated with one of the MHC types to form a single MHC–peptide complex string. Affinity is proportional to the similarity of the TCR string to the MHC–peptide complex string.

each peptide string in the model is associated with exactly one of the three MHC alleles. Because a TCR binds to both the peptide and parts of the MHC, each of the alleles is associated with a random digit string to represent the portion of MHC visible to the TCR. A peptide string is concatenated with its associated MHC's string to form a single MHC–peptide complex string that interacts with TCRs (Figure 4.1).

Each CTL in the simulation is assigned a randomly generated TCR string, which is the same length as the MHC–peptide complex strings (Figure 4.1). The similarity between a TCR string to an MHC–peptide complex string determines the affinity that the CTL has for an infected cell expressing that peptide. Each CTL is assumed to express many copies of the same TCR, so a single TCR string is sufficient to represent a CTL's specificity for antigen.

Antigenic distance, which has an inverse relationship with affinity, is a measure of how reactive an immune cell is to an antigen. If a CTL has a high affinity for an epitope, then it is antigenically close to it. In the model, a distance metric is used to formally define the distance between a CTL and an MHC–peptide complex string. Distance is inversely proportional to similarity. If the metric determines a TCR string and an MHC–peptide complex string to be close (similar), then they have a high-affinity interaction in the model. The distance metric can be defined in many ways, and the choice of metric might affect the CTL model's behavior. Therefore, I implement three different versions of the CTL model, each using a different definition of distance. These metrics are defined in Appendix A. For all three metrics, the distance between two strings is the sum of the distances between their corresponding digits. This constraint agrees with the observation that amino acid side chains of peptides seem to make independent contributions to the binding energy with the TCR (Hemmer et al., 1998).

4.2 Generating the naïve T cell repertoire

Thymic selection shapes the distribution of TCRs in the immune system, and the CTL model uses an analogous process to generate its naïve CTL repertoire (described in Section 4.2.1). A murine or human naïve repertoire consists of $10^6 - 10^7$ clones (Pannetier et al., 1993; Arstila et al., 1999; Casrouge et al., 2000), which would be computationally expensive to simulate. Because the purpose of the model is to observe the response to a set of antigens, only the tiny fraction CTLs that can respond to these antigens need to be instantiated. The procedure for generating only the responding CTLs is described in Section 4.2.2. I have implemented versions of the CTL model using three different distance metrics: Hamming, xor, and modified Manhattan distance. These metrics are defined and calibrated for the CTL model in Appendix A, and a summary of this calibration is presented in Table 4.2.

4.2.1 Thymic selection

In the immune system, the fate of CTLs during thymic selection depends on their affinity for MHC–self peptide complexes. The model subjects random pre-selection TCR strings to an analogous process. Random strings are generated to represent the TCRs of the pre-selection CTL repertoire. The distance between each of these pre-selection strings and all of the MHC–self peptide strings is computed. A positive selection process eliminates CTLs with TCR strings that are too far from (dissimilar to) all MHC–self peptide complexes, and a negative selection process eliminates those with TCRs that are too close (similar) to any MHC-self complex. Only CTLs with TCR strings that are an intermediate distance from MHC–self peptide complexes survive to form the naïve repertoire. The

	Mouse	Human	Hamming	xor	L_1'
# of self peptides	$10^4 - 10^5$*		30,000	30,000	30,000
# of MHC types	3	4	3	3	3
universe of TCRs (or # of possible TCR strings)		10^{15}†	1.47×10^{38}	1.18×10^{21}	1.13×10^{15}
# of pre-selection clones	$< 10^9$	10^{13}	8×10^7	2.5×10^8	2.5×10^8
# of naïve clones	$10^6 - 10^7$‡	10^7§	3.17×10^6	2.02×10^6	1.95×10^6
foreign peptide response frequency	$10^{-5} - 10^{-6}$		8.39×10^{-6}	1.27×10^{-5}	1.43×10^{-5}
thymic selection window size	1-3%		3.96%	0.807%	0.778%
% killed in negative selection	50-66%		46%	61%	70%
# of clones per epitope	10-20‖		26.6	25.7	27.9

* Bevan (1997); Müller and Bonhoeffer (2003); Bandeira and Faro (2003)
† Davis and Bjorkman (1988)
‡ Pannetier et al. (1993); Casrouge et al. (2000)
§ Arstila et al. (1999)
‖ Blattman et al. (2002)

Table 4.1: A summary of the values used to calibrate the different distance metrics. Biologically plausible values from studies of mice and humans are listed for comparison with the model's parameters.

model computes positive and negative selection distance thresholds to eliminate most of the random pre-selection CTLs, leaving only a small set of surviving naïve clones (Figure 4.2).

For a particular TCR string, I designate the nearest (most similar) MHC–self peptide complex string as its "selecting" peptide. The distance between a CTL's TCR string and its selecting peptide determines whether or not the CTL survives thymic selection. If the selecting peptide is too close, then the CTL is eliminated by negative selection; if it is too far, then it is eliminated by positive selection. Non-selecting self peptides, which are farther from the TCR than the selecting peptide, do not affect its chance of surviving selection.

The model's definition of the "intermediate distance" from self that ensures survival of pre-selection CTLs is derived from mouse data. In mice, 1–3% of pre-selection T cells survive thymic selection (Shortman et al., 1990), and about one-half to two-thirds of cells that survive positive selection are eliminated by negative selection. Therefore the model uses positive and negative selection thresholds such that 1–3% of pre-selection CTLs have selecting peptides at distances between these two thresholds, and about 1–2 times more pre-selection CTLs (i.e., 1–6%) have selecting peptides that are closer than the negative selection threshold.

The positive and negative selection thresholds are found using the distribution of expected distances between a random TCR string and its selecting peptide. The distribution was computed for each distance metric using the algorithm described in Detours et al. (1999). The expected fraction of pre-selection CTLs eliminated by negative selection is calculated by summing the distribution for all distances from zero to the negative selection threshold (Figure 4.2). The expected fraction of CTLs eliminated by positive selection is the summation of the distribution for all distances from the positive selection threshold to infinity. The CTLs that are between these two thresholds are in the "window" of distances that survive thymic selection in the model. Various combinations of positive and negative

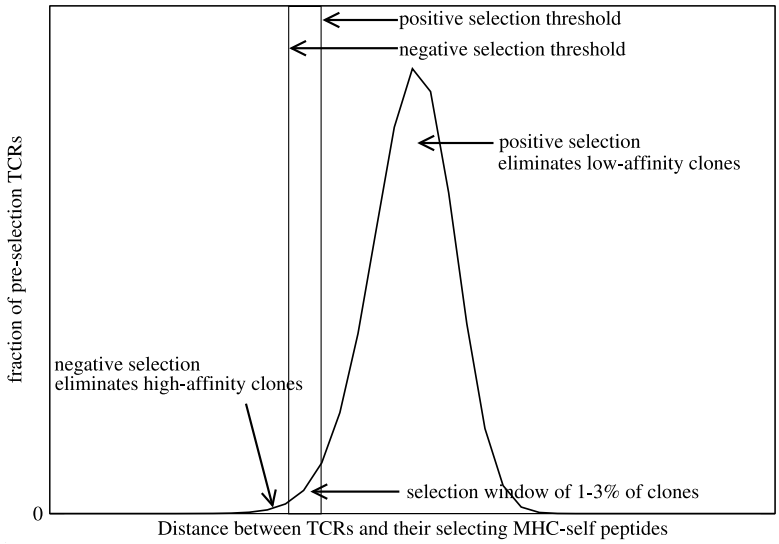


Figure 4.2: The thymic selection window computation. The expected distribution of distances between a random pre-selection TCR and the nearest MHC–self peptide complex is plotted. Positive selection eliminates the CTLs with TCRs that are to the "right" of the positive selection threshold, while negative selection eliminates those to the "left" of the negative selection threshold. Those that are between the two thresholds survive selection and become naïve cells.

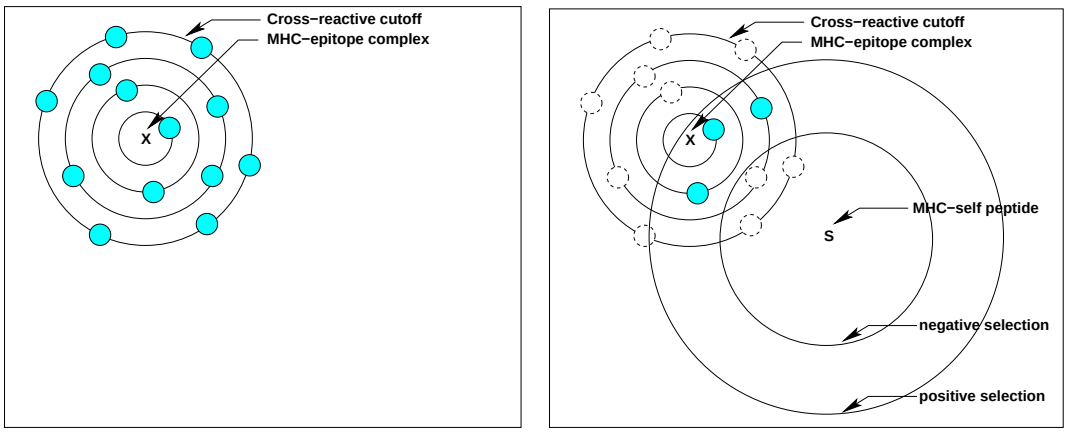
selection thresholds were tested to find a combination that satisfy the constraints derived from mouse data (Figure 4.2).

4.2.2 Lazy evaluation and the cross-reactive cutoff

Mice and humans have an estimated $10^6 - 10^7$ naïve CTL clones (Pannetier et al., 1993; Arstila et al., 1999; Casrouge et al., 2000), which exist to anticipate a seemingly infinite variety of pathogens. Most of these cells never have the opportunity to participate in a response to antigen during the organism's lifetime. A response to a single epitope usually involves only tens of CTL clones, and a single organism will be exposed to a limited number of antigens. Thus, only a tiny fraction of naïve CTLs will ever play a significant

role in clearing infections from an organism. In a simulation, we are interested only in this fraction. In fact, a "newborn" simulated organism would not need any naïve CTLs until it is exposed to antigen. Immediately before each exposure to antigen, the simulation can instantiate the cells that can respond to that particular antigen if they were not already created in a prior exposure to antigen. In this manner, only the CTLs that play an active role in the simulation are explicitly created. This procedure was formalized in Smith et al. (1998), in which the author adapts computer science's concept of *lazy evaluation* to determine which cells actually need to be instantiated. By creating only the cells that are necessary, the simulation is orders of magnitude more efficient.

Most CTLs are too antigenically distant from any given epitope to have any affinity for it, so only the tiny fraction of CTLs that have affinity for the antigen's epitopes are instantiated in a "lazy" simulation. The *cross-reactive cutoff* is the antigenic distance from an epitope beyond which immune cells, such as B or T cells, have negligible affinity for the epitope. Thus, when a simulation introduces a new antigen, only the naïve CTLs that are closer than the cross-reactive cutoff of the antigen's epitopes are created. In the model, the cross-reactive cutoff for MHC–foreign peptide complexes is set to be equal to thymic selection's negative selection threshold for MHC–self peptide complexes. This is based on the assumption that the purpose of negative selection is to rid the body of self-reactive CTLs, so the cells that could react to MHC–self peptide complexes in the body (i.e., those within the cross-reactive cutoff of these complexes) are exactly those that are removed by negative selection (i.e., those within the negative selection threshold). If only 10^{-5} of the repertoire responds to an epitope (Stockinger et al., 1980; Zinkernagel, 1996), then using lazy evaluation can reduce the number of cells created in a simulated response to an epitope by 5 orders of magnitude. In the past, modelers used artificially small repertoires (Kleinstein and Seiden, 2000) or required substantial computing resources to simulate a realistic-sized repertoire (Detours and Perelson, 2000). In my model, all of the active cells of a realistic-sized repertoire are represented.



(a) **(b)**

Figure 4.3: Lazy evaluation of CTLs. The pre-selection repertoire for a single epitope is depicted in Figure a). The $\times$ represents an MHC–epitope complex, and filled circles represent pre-selection CTLs. The distance between the epitope and a CTL in the figure is proportional to their antigenic distance. The pre-selection repertoire is generated by creating CTLs at each distance from 0 to the cross-reactive radius from the MHC–epitope complex. Figure b) depicts the repertoire after thymic selection against a single MHC–self peptide complex. Thymic selection eliminates those that are within the negative selection thresholds of any MHC–self peptide complex and those that are outside the positive selection thresholds of all MHC–self peptide complexes. The CTLs that do not survive thymic selection are drawn as empty dashed circles. The surviving cells (filled circles) mature to become naïve CTLs.

The first step in instantiating the CTLs that can respond to a particular epitope is to create a pre-selection repertoire for the epitope. Using lazy evaluation, only the pre-selection repertoire that is within the cross-reactive cutoff of the MHC–epitope complex is generated. These CTLs form a "sphere" of strings surrounding the MHC–epitope complex with a radius equal to the cross-reactive cutoff (Figure 4.3a). These CTLs can be created by generating random strings that are at distance 0 from the complex, then at distance 1, and so on until the cross-reactive cutoff distance is reached. See Smith et al. (1998) for a detailed description.

The number of CTLs that should be generated at each distance is based on the number

one would expect to find at each distance if all of the clones of the pre-selection repertoire were explicitly generated. For example, if the simulation uses MHC–peptide complex strings of length 10 and an alphabet of size 20, then the number of possible strings is 20^{10}, or about 10^{13}. Of these strings, only one is exactly the same as the complex's string. If the simulated organism has a pre-selection repertoire size of 10^8 CTLs, then the expected number of clones at distance 0 from the complex is $10^8/10^{13} = 10^{-5}$. The expected number of clones for distances greater than 0 depends on the particular distance metric chosen, and the calculations for each distance metric are described in Appendix A. The actual number of clones at each distance is chosen by drawing a random number from the binomial distribution with the expected value as the mean.

Once the number of pre-selection TCR strings at a given distance from an MHC–epitope complex is determined, random TCR strings are generated using the algorithms described in Appendix A. If the repertoire for one or more other MHC–epitope complexes was generated before, then care must be taken so that these new pre-selection CTLs do not "overlap" with them. When a new pre-selection TCR string falls within the cross-reactive cutoffs of a previously encountered epitope, that TCR is eliminated. These TCRs are in the region of TCR space that is stimulated by the current MHC–epitope complex and a previous one. Therefore, the TCRs in this region are created by lazy generation upon exposure to the first complex, and the second complex can stimulate those previously generated TCRs rather than create extra ones in this region covered by the other complex.

After the pre-selection cells are generated, they are subjected to a thymic selection process against all MHC–self peptide complexes. The distances between each clone and each MHC–self peptide complex is computed, and those clones that are too close to one of the self complexes or too far from all of them are eliminated (Figure 4.3b). The remaining clones enter the naïve repertoire.

4.3 Mutation

Genetic mutation is one of the many mechanisms that viruses have evolved to evade the immune response. When antigens replicate in the body, mutations can change or even eliminate the epitopes that cells infected by this antigen express. Over the course of an infection, random mutations can accumulate in an antigen's lineage, generating multiple competing variant strains in a single host. The CTL model implements antigenic mutation so these effects can be studied.

I assume that mutation makes random changes to the viral genome when it replicates, so the mutation rate in the CTL model is expressed as a probability of mutation per replication event. A mutation changes an antigen's epitopes by setting a single randomly chosen portion of an antigen's epitope string to a random value. For the xor and modified Manhattan distance versions, a mutation changes only a randomly chosen single digit. Because the Hamming distance version of the model uses much longer epitope strings (see Appendix A.1), strings are divided into groups of 8 digits, and a mutation sets all 8 digits of one randomly chosen group to random values.

In the CTL model, the number of mutations that occur depends on the virus dynamics. Recall the equations that govern virus dynamics from Section 3.1:

$$\Delta T = (\lambda - \delta_T T - \beta T V)\Delta t, \tag{4.1}$$

$$\Delta I = (\beta T V - \delta_I I)\Delta t, \tag{4.2}$$

$$\Delta V = (\pi I - cV)\Delta t \tag{4.3}$$

If one assumes that mutations occur when viruses replicate within a host cell, then the number of new mutant viruses that arise per time step is proportional to the virus production rate, $\pi I \Delta t$ from Equation 4.3. However, most of these new viruses will die (the cV term) without infecting cells, so there is no need for the simulation to generate all of the mutants. It is more efficient to create the mutants as cells are infected because only the

viruses that successfully infect a cell affect the host. Cells are infected at rate $\beta TV\Delta t$, so the number of new mutant cells generated in a time step is drawn from the binomial $\mathcal{B}(\beta TV\Delta t, p)$, where p is the mutation rate. The epitopes expressed by the original antigen are changed for each of these new mutants. Thus, each mutation event creates a single infected cell expressing what is likely to be a unique epitope.

Because a newly mutated antigen strain has an altered epitope, some of the CTLs that had responded to the original antigen will not respond to it. Figure 4.4 shows the distribution of the number of CTLs that can respond to one epitope and the number of those that can recognize the mutated epitope. For the results shown in this figure, mutation was performed by setting the first digit(s) of the epitope string to 0, so the peptide was unchanged if the digit was originally 0. For both the xor and L_1' metrics, about half of the mutations changed the epitope enough that none of the CTLs that had responded to the original epitope recognized the new epitope. In the Hamming metric version, mutation never allowed the epitope to evade all of the clones that had responded to the original.

4.4 Converting distance to affinity

The CTL model described in Chapter 3 requires that the affinity between a TCR and an MHC–peptide complex be defined. Affinity, which is the strength of the interaction between a TCR and an MHC–peptide complex, determines the rate at which quiescent CTLs are recruited into a response (Section 3.2.2) and how rapidly a particular clone clears infected cells (Section 3.2.3). The strength of interactions between TCRs and MHC–peptide complexes is proportional to their string distances in the model.

In the immune system, TCRs have sensitivities for antigen that can differ by orders of magnitude. Therefore, the affinity decreases exponentially with respect to string distance in the model. Affinities are defined for each of the three versions of the mode by computing

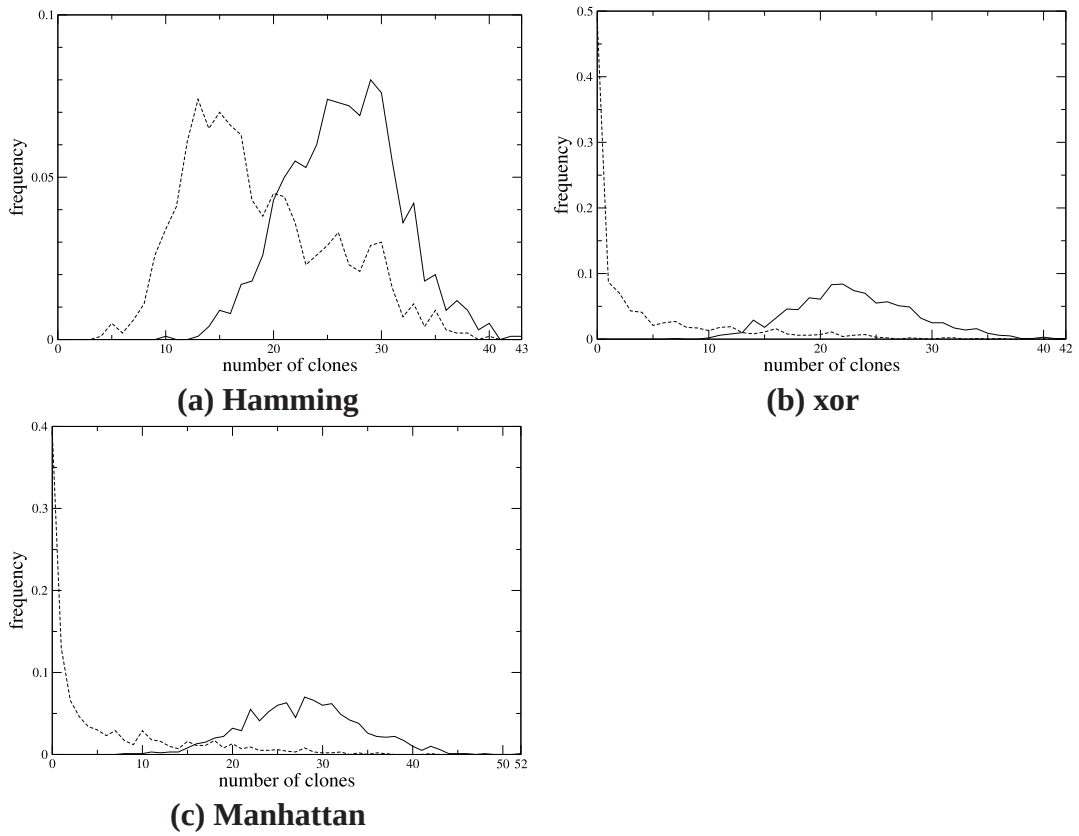


Figure 4.4: The repertoire overlap between mutated epitopes. TCRs were generated around a foreign peptide–self MHC complex. The distribution of the number of TCRs is indicated by the solid line. The epitope was mutated, and the distribution of the number of TCRs from the original epitope that could respond to the new one is indicated by the dashed line. The peptide was mutated by eight digits for the (a) Hamming results and by one digit for the (b) xor and (c) L_1' results. Results shown are the distribution of 1000 trials.

a dissociation constant, K, based on distance:

$$K_{xor} = 5,000 + 15,000 \times e^{(D_{xor}-115)/3} \quad (4.4)$$

$$K_H = 5,000 + 5,000 \times e^{2\times(D_H-31)} \quad (4.5)$$

$$K_{L1'} = 5,000 + 10,000 \times e^{2\times(D_{L1'}-15)} \quad (4.6)$$

where D_{xor} is the xor distance, D_H is the Hamming distance, and $D_{L1'}$ is the modified Manhattan distance. Affinity is inversely proportional to the dissociation constant (see Section 3.2.2). The constants in Equations 4.4–4.6 were chosen so that each epitope has a few high-affinity clones in the naïve repertoire (K in the range of $5000 - 10000$) and that the low-affinity clones have a dissociation constant that is between $10^5 - 10^7$. The distribution of naïve clone distances from and affinities for an MHC–epitope complex is shown in Figure 4.5. The affinity distributions for the different distance metrics could not be made equal, but they are qualitatively similar.

4.5 Summary

The CTL model uses digit strings to represent TCRs and MHC–peptide complexes. The affinity that a CTL has for an MHC–peptide complex is proportional to the similarity of the digit strings corresponding to the CTL's TCR and the complex, where similarity is defined by a string distance metric. Three versions of the model, each using a different distance metric (Hamming, xor, and modified Manhattan distance), are calibrated to match known thymic selection characteristics in mice. The model implements a process that represents thymic selection to produce a naïve CTL repertoire. The model represents an organism with $10^6 - 10^7$ CTL clones, but most of these clones do not need to be explicitly generated in a simulation. By creating only the TCRs that can respond to the epitopes used in a particular simulation, the simulation realizes an enormous savings in computation and memory, sometimes by as much as a factor of 10^5. An epitope mutation operation is defined to allow the model to simulate the evolution of pathogens in a host. Finally, formulas for converting string distances to the affinity values are given.

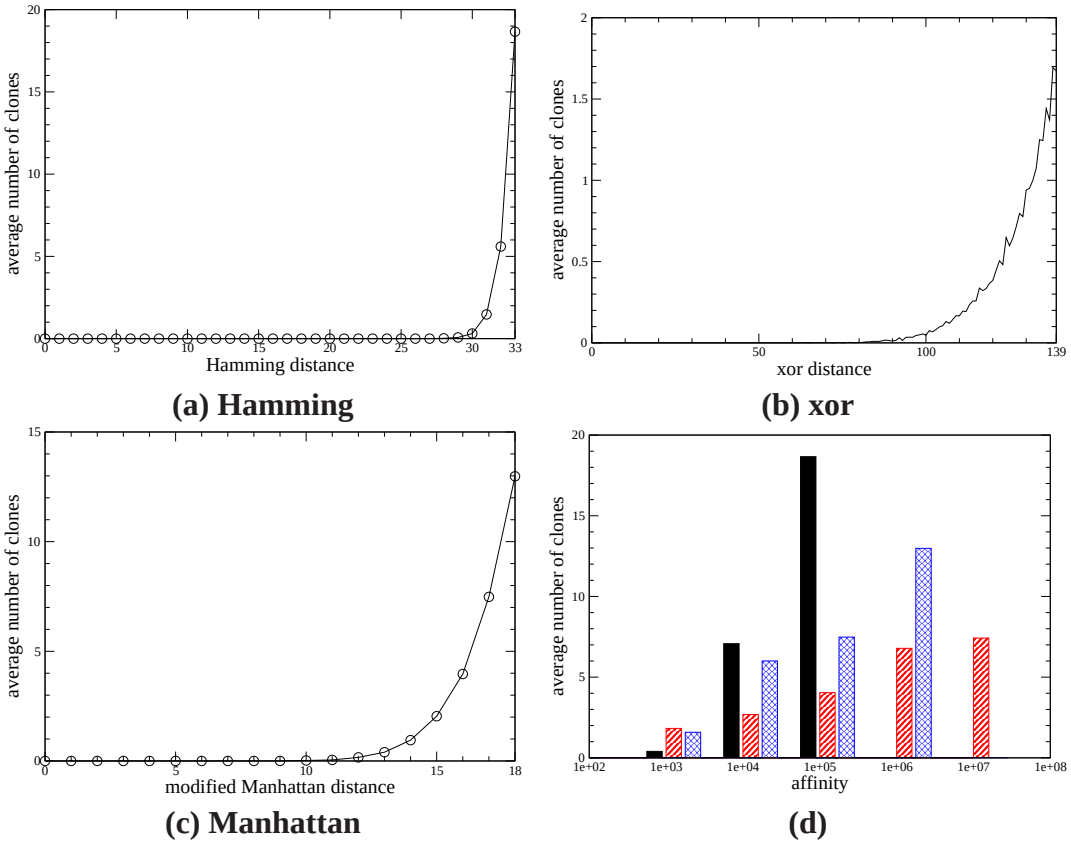


Figure 4.5: The distances of TCRs from an MHC–foreign peptide complex. CTL clones for a foreign peptide were generated using lazy evaluation. The distribution of distances between the clones and the foreign peptide-self MHC complex is plotted for (a) Hamming distance, (b) xor distance, and (c) modified Manhattan distance. The plots represent average results from 100 runs. In (d), the results in (a)–(c) are converted to affinities using Equations 4.4–4.6. The histogram plots the number of clones with a given affinity for the complex. Each bin of the histogram is larger than the preceding bin by a factor of 10. The solid bars represent Hamming distance, the diagonal bars represent xor distance, and the cross-hatched bars represent modified Manhattan distance.

Chapter 5

Results

If you know exactly what you're going to do, what's the good in doing it?

—Pablo Picasso

The definition of insanity is doing the same thing over and over and expecting different results.

—Benjamin Franklin

The CTL model reproduces phenomena seen in cell culture and in laboratory mice. The effects that thymic selection has on the model's naïve CTL repertoire are described in Section 5.1. Section 5.2 illustrates the basic dynamics of the acute CTL response using only one or two clones per epitope. A realistic number of clones is introduced in Section 5.3, which describes the clonal composition of responses. Most of the results presented in Sections 5.2 and 5.3 were first published in Chao et al. (2003), Chao et al. (2004a), and Chao et al. (2004b).

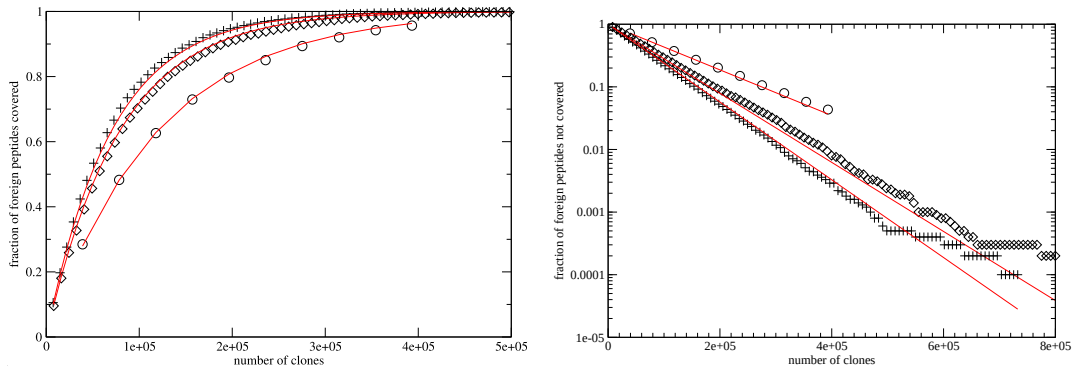
5.1 Consequences of thymic selection

Thymic selection transforms a random CTL repertoire into one that can detect foreign peptides while ignoring self peptides. The model's implementation of thymic selection performs an analogous function on a set of random TCR strings, and its effects can be observed by comparing the repertoire before and after selection. Even though the model creates three MHC alleles and mice and humans generally express more than one, I discuss many of the results as if there were only one MHC type. I assume that MHC restriction (Zinkernagel and Doherty, 1974) is strong enough that a CTL can only interact with peptides presented on the MHC type that presented its selecting peptide in the thymus. This assumption simplifies the following discussion without loss of generality.

5.1.1 CTL repertoire coverage of foreign peptides

The number of foreign peptides that a CTL repertoire can recognize in the model is a function of the number of clones. I define *coverage* as the percentage of foreign peptides that are detected by at least one CTL clone. Figure 5.1 plots the relationship between the number of naïve clones and coverage, which initially increases with the number of clones then quickly saturates. To determine the foreign peptide coverage of a CTL repertoire, I generated 10,000 random "foreign" peptide strings and counted the fraction that was detected by the model's CTL repertoire, which is randomly generated. Because I was measuring the coverage of a whole repertoire, the CTLs were explicitly generated and subjected to thymic selection against 30,000 self peptides, and lazy evaluation was not used.

The probability that a foreign peptide is "covered" by at least one clone can be estimated using the foreign peptide response frequency f, which is calculated for each of the distance metrics used by the CTL model in Appendix A. If a single CTL covers a fraction



(a) **(b)**

Figure 5.1: Foreign peptide coverage by CTLs. 10,000 random foreign peptide strings were generated and each was associated with one of three MHC types. In (a), the fraction of these MHC–foreign peptide complexes detected by at least one CTL of a set of n clones is plotted against the size of the repertoire for Hamming (○), xor (◇), and L_1' (+) distance versions of the model. In (b), the fraction *not* covered is plotted. The lines indicate the expected values using Equation 5.1.

f of all possible foreign peptides, then the fraction of space *not* covered is $1-f$. The probability that a foreign peptide is not covered by a set of n distinct CTLs is $(1-f)^n$. Thus, the foreign peptide coverage A_n of n clones is:

$$A_n = 1-(1-f)^n \tag{5.1}$$

This prediction fits results from the model for all distance metrics (Figure 5.1). This indicates that coverage can be estimated accurately using only the foreign peptide response frequency, without considering the other properties of the distance metrics. This result also implies that the CTLs that survive selection cover the space of foreign peptides with the same efficiency as one would expect of the same number of randomly generated pre-selection CTLs.

An effective CTL response requires multiple clones per epitope, so coverage of an antigen's epitopes by a single clone does not guarantee its clearance by the immune system. To estimate the probability that a foreign peptide is covered by multiple clones, I first determine the probability that a foreign peptide is covered by exactly m clones out of a

repertoire of size n:

$$A_{n,m} = f^m(1-f)^{n-m}C(n,m) \tag{5.2}$$

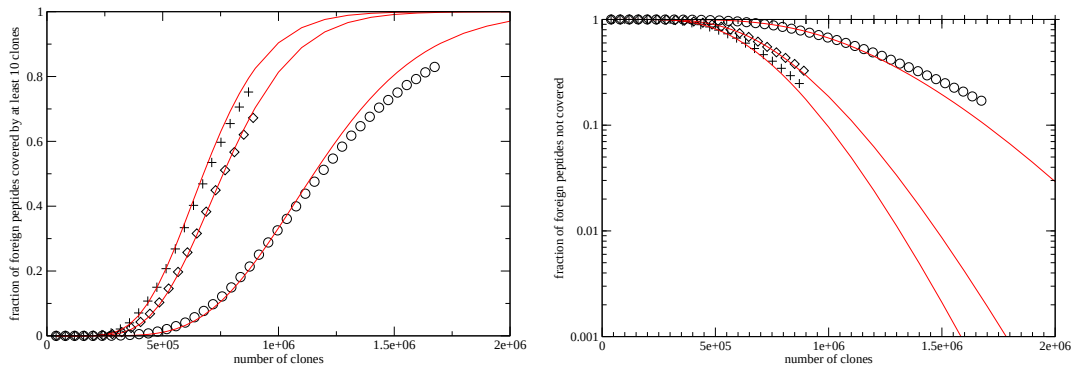
where $C(n,m)$ is the number of distinct combinations of m objects that can be drawn from a set of n and is equal to $\frac{n!}{(n-m)!m!}$. It has been observed that 10–20 clones respond to an immunodominant LCMV epitope in mice (Blattman et al., 2002), so I will assume that an effective response requires at least 10 clones per epitope. The probability that an epitope is covered by 10 or more clones is equal to 1 minus the probability that the epitope is covered by fewer than 10 clones:

$$\sum_{i=10}^{\infty} A_{n,i} = 1 - \sum_{i=0}^{9} A_{n,i} \tag{5.3}$$

This function is plotted and compared to results from the model in Figure 5.2. The probability that a foreign peptide is detected by enough naïve CTL clones to mount an effective response reaches 99% when the number of naïve clones is between 10^6 and 10^7, which happens to be the number of CTL clones in a mouse (Pannetier et al., 1993; Casrouge et al., 2000).

5.1.2 CTL affinity for MHC is correlated with affinity for self peptides

In the model, thymic selection eliminates CTLs solely on the basis of their affinities for MHC–self peptide complexes. Before selection, TCR strings are random, so a TCR's affinity for MHC is independent of its affinity for self peptides. Thymic selection introduces a dependence between a surviving CTL's affinity for MHC and its affinity for peptide. After selection, affinity for MHC has an inverse relationship with affinity for self peptides (Figure 5.3). Because thymic selection allows only CTLs with a very narrow range of affinities for their closest MHC–self peptide complexes to survive, if a surviving CTL has a certain



(a) **(b)**

Figure 5.2: Foreign peptide coverage by at least 10 CTLs. 10,000 random foreign peptide strings were generated and each was associated with one of three MHC types. In (a), the fraction of these MHC–foreign peptide complexes detected by at least ten CTLs of the model's naïve repertoire is plotted against the size of the repertoire for xor ($\diamond$) and L_1' (+) distance versions of the model. In (b), the fraction *not* covered is plotted. The size of the repertoire that could be created by the model was limited by the computer's memory capacity, so there are no empirical results for large numbers of clones. The lines indicate the expected values using Equation 5.3.

affinity for its selecting peptide, then its affinity for MHC must fall within a very narrow range for it to have survived selection.

The MHC-binding portion of a TCR determines its peptide binding degeneracy, which is a measure of the number of different peptides with which a single TCR can bind. Because TCRs bind to peptide presented by MHC, those having high affinity for MHC can bind to a much larger set of peptides, and thus have a higher degeneracy, than those that bind poorly to MHC. Therefore, peptide binding degeneracy and affinity for MHC are correlated. The affinity of a thymically selected CTL for self peptide and its peptide binding degeneracy are so closely linked that the standard explanations for the roles of negative and positive selection are reconsidered in the following sections.

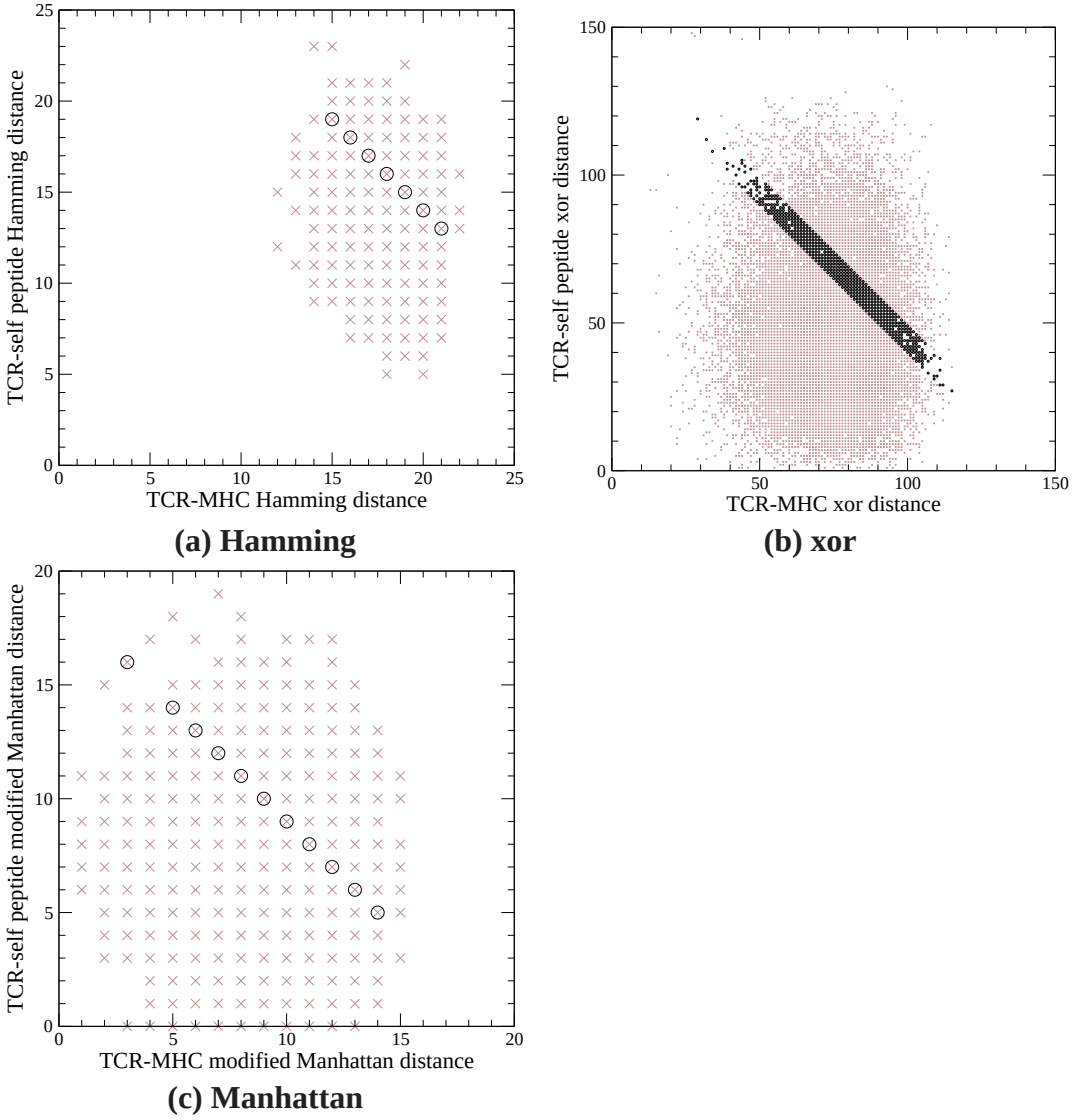


Figure 5.3: Thymic selection introduces a dependency between MHC and self-peptide affinity in CTLs. For 100 random foreign peptides presented by MHC, realistic-sized pre- and post-thymic selection repertoires were generated using lazy evaluation. The antigenic distance between each TCR's peptide-binding region and its selecting self peptide is on the x-axis, and the distance between each TCR's MHC-binding region and the presenting MHC is on the y-axis. Three distance rules were tested: (a) Hamming, (b) xor, and (c) modified Manhattan distances. The pre-selection TCRs are represented by light $\times$s, and the post-selection TCRs by black $\circ$s.

5.1.3 Negative selection increases peptide binding specificity

Immunologists assume that negative selection eliminates CTLs with high affinity for self peptides. In the model, negative selection eliminates CTLs with high affinity for MHC–self peptide complexes, so both CTLs with high affinity for self peptides *and* those with high affinity for MHC are removed. Figure 5.4 compares the pre-selection CTL repertoire with the repertoire that survives negative selection. Comparing the distribution of distances between TCRs and MHC before and after negative selection (without subjecting them to positive selection), one can see that negative selection decreases the average affinity for MHC. By eliminating TCRs with high affinity for MHC, negative selection only allows those with higher specificity for their cognate peptides to survive. This agrees with the suggestion that negative selection increases the specificity that TCRs have to foreign peptides (Huseby et al., 2003; Slifka et al., 2003).

TCRs can be generated in vitro so that they are not subject to thymic selection. In Holler et al. (2003), TCRs were selected in vitro to have high affinity for a particular set of MHC–foreign peptide complexes. In the experiment, it was found that cells expressing these TCRs tended to react to self peptides (Holler et al., 2003). One would expect these CTLs to have high affinity for both the foreign peptide and its presenting MHC. The consequence of having high affinity for MHC would be highly degenerate peptide binding, allowing them to react to self peptides also. Presumably, these cells would have been eliminated by negative selection because of their high affinity for MHC, not because they have high affinity for a self peptide.

5.1.4 Positive selection maximizes peptide binding degeneracy

It is widely believed that the purpose of positive selection is to eliminate CTLs with such low affinity for MHC that they would not be likely to bind to foreign peptides presented by MHC. Some have even suggested that self peptides are just “stand-ins” for foreign

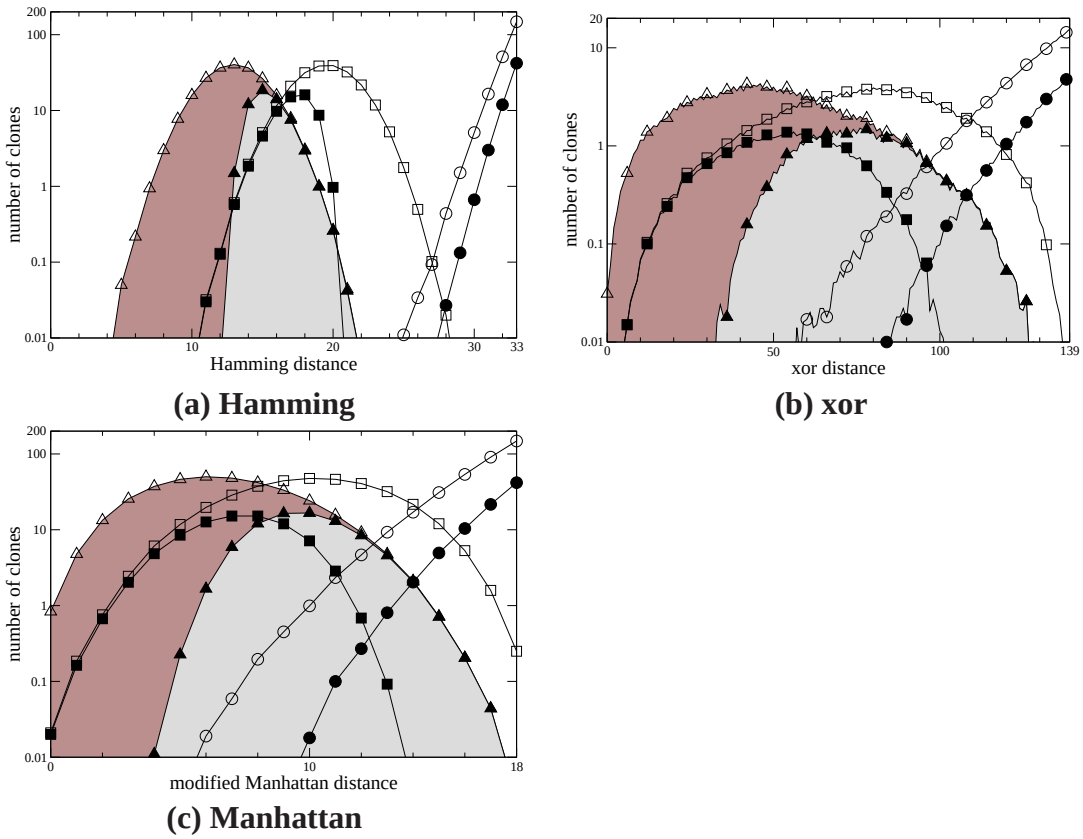


Figure 5.4: Effects of negative selection on TCR distance to foreign peptides. A pre-selection TCR repertoire was lazily generated for a single foreign peptide presented by MHC. This repertoire was then subjected to negative but not positive selection. The distributions of antigenic distances from the MHC–foreign peptide complex to TCRs from the pre- and post-selection repertoires are plotted. The number of TCRs at each distance from the MHC–foreign peptide are indicated by ○ for the pre-selection repertoire and ● for the post-selection repertoire. The number of TCRs whose peptide-binding region are at each distance from the foreign peptide are indicated by □ for the pre-selection repertoire and ■ for post-selection. The number of TCRs whose MHC-binding region are at each distance from the MHC that presented the foreign peptide are indicated by △ and the dark shaded region for the pre-selection repertoire, and ▲ and the light shaded region for the post-selection distribution. Three distance measures were used: (a) Hamming, (b) xor, and (c) modified Manhattan. The results shown are the averages from 1000 different trials for each distance metric.

peptides during positive selection (Goldrath and Bevan, 1999). I believe that self peptides play an essential but overlooked role in positive selection. Although positive selection in the model tends to eliminate CTLs with low affinity for MHC, some of these CTLs can be "rescued" by having high affinity for a self peptide. Conversely, CTLs with moderate affinity for MHC can be "damned" by having low affinity for all self peptides. Therefore, positive selection does not simply purge the repertoire of CTLs with low affinity for MHC—it removes CTLs that have "sub-optimal" affinity for MHC given the CTL's affinity for its selecting peptide.

This hypothesis can be tested using engineered thymic selection environments with only one positively selecting peptide. In Kraj et al. (2001), the specificities of two of CTLs positively selected on a single MHC–peptide complex were characterized. One CTL was very specific to a peptide similar to the selecting peptide. The other CTL was specific to a peptide that was unrelated to the selecting peptide, and it had a high peptide binding degeneracy. I postulate that the first CTL had a high affinity for the selecting peptide and a low affinity for MHC, and the second had low affinity for the peptide and high affinity for MHC. More studies will be needed to determine the relationship between the affinities that CTLs have for self peptide and for MHC.

5.1.5 Epitopes and self peptides

Although the CTLs in the model have "maximal" peptide binding degeneracies, the CTLs that are close to self peptides have lower peptide binding degeneracies. Epitopes that are close to self peptides would be covered by these highly specific CTLs. In the model, the average binding degeneracy of CTLs that have affinity for an epitope is correlated with the epitope's distance from the nearest self peptide for two of the three distance metrics tested (Figure 5.5). For the Hamming distance version (Figure 5.5a), the binding degeneracy is not affected by the distance between the epitope and self peptides. This divergence from

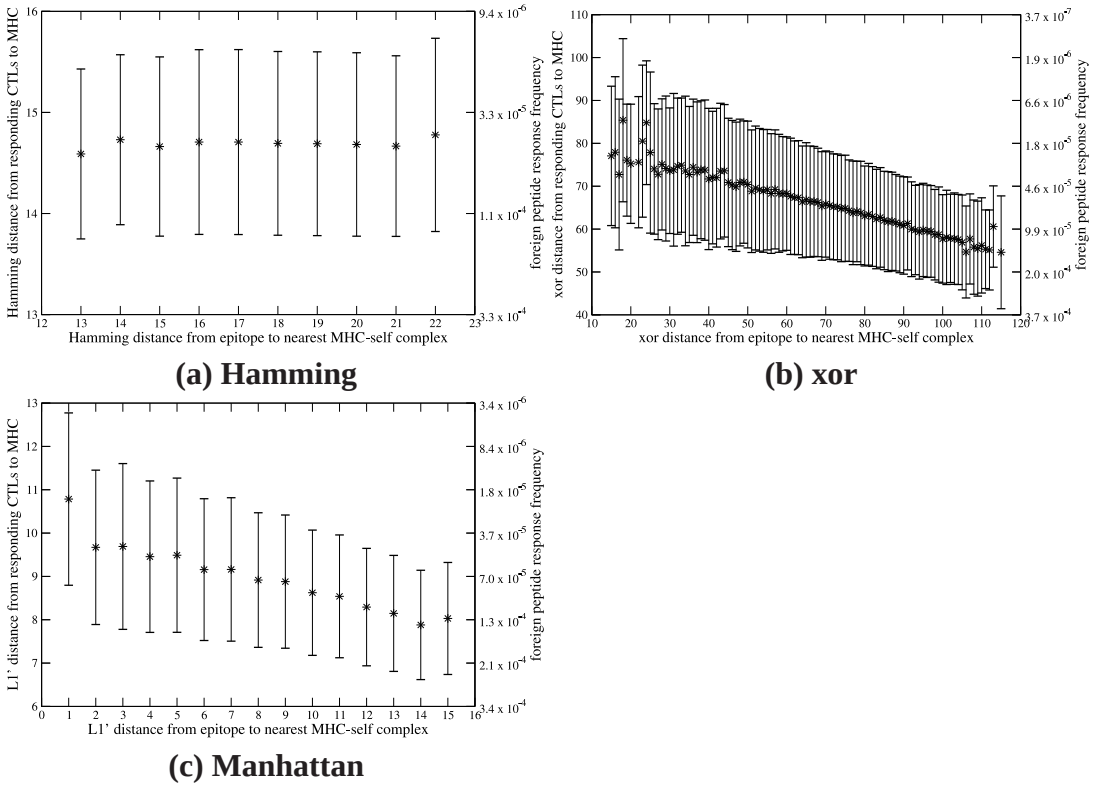


Figure 5.5: CTL to MHC distance vs. distance between an epitope and its closest self peptide. 10,000 random epitopes were generated, and the distances between these epitopes and their nearest self peptides were measured. A new CTL repertoire was created for each of these epitopes, and the average and standard deviation of the distances from their TCRs to the MHC presenting the epitopes is plotted against the epitope–selecting peptide distance. The TCR–MHC distance determines a CTL's foreign peptide response frequency, and the corresponding frequency for each distance is displayed on the y-axis on the right of each plot. The results shown are from simulations using three distance metrics: (a) Hamming, (b) xor, and (c) modified Manhattan distances.

the other distance metrics (xor and modified Manhattan) could be a property of Hamming distance or it might simply be the choice of Hamming distance parameters used by the model.

The peptide binding degeneracy of CTLs could affect the ability of the immune system to eliminate a mutating pathogen. Pathogens can escape the immune system's response

when their epitopes mutate. These modified epitopes might not be recognized by the CTLs that respond to infected cells expressing the original epitope, so the immune system would need to mount a new response against the mutant. A CTL with high peptide binding degeneracy might be able to recognize both an epitope and its variants, which would make it difficult for a new mutant to survive. Thus, even a quickly mutating pathogen would generate few surviving variants so its "effective" mutation rate would be low. A CTL that is too specific could be easier to escape through mutation. If most CTLs responding to an epitope that is similar to a self peptide are highly specific, then the antigen's variants would have a greater chance of surviving the immune response. Because their mutated progeny would be the most viable, such antigens would have the highest "effective" mutation rates. If the effective mutation rate increases as the epitopes become more similar to self peptides, then the immune system drives the pathogen to express epitopes that mimic self. Once an epitope is sufficiently similar to a self peptide, the immune system would be unable to detect it because negative selection eliminates CTLs too close to self.

5.1.6 Naïve repertoire generation efficiency

The generation of the naïve CTL repertoire is an expensive process, both in the body and in the model. If over 95% of randomly generated CTLs are purged during thymic selection, then for each CTL that joins the naïve repertoire, over 19 are eliminated in the thymus. I measured the efficiency of CTL generation in the model. On average 30 CTL clones respond to each epitope, so one would expect the model to generate $20 \times 30 = 600$ pre-selection CTL clones per epitope. In practice, only about 10 pre-selection CTL clones are generated to produce each naïve clone (in Table 5.1, divide the number of CTL clones generated by the number surviving selection), or about 300 pre-selection clones per epitope.

This efficiency in the model appears to be a consequence of using lazy evaluation (Sec-

	Hamming	xor	L_1'
number of pre-selection clones	8×10^7	2.5×10^8	2.5×10^8
thymic selection window	34	140–149	19
number of CTL clones generated	222.8 ± 14.9	230.9 ± 15.0	359.9 ± 18.4
clones killed in positive selection	31.3 ± 5.9	40.0 ± 13.6	53.9 ± 17.4
clones killed in negative selection	164.9 ± 13.1	168.1 ± 19.4	278.0 ± 24.2
clones surviving selection	26.6 ± 5.1	23.1 ± 5.5	27.9 ± 6.4

Table 5.1: The efficiency of T cell repertoire generation in the model. A pre-selection repertoire was generated around a single MHC–foreign peptide complex using lazy evaluation. This repertoire was subjected to positive and negative selection to produce naïve CTL clones. The average and standard deviation for 1000 trials for each of the three distance metrics are shown.

tion 4.2.2). The model uses lazy evaluation so that it does not generate the pre-selection CTLs that are outside the cross-reactive cutoff of an MHC–epitope complex, reducing the computational and memory requirements of the simulation by several orders of magnitude. However, lazy evaluation introduces an unexpected additional efficiency gain.

In the "real" immune system, it is believed that positive selection eliminates more pre-selection clones than negative selection: about 95% by positive selection and $<5\%$ by negative selection (Section 4.2.1). In the model, these proportions are reversed—about 75% of pre-selection clones specific to a single MHC–epitope complex are eliminated by negative selection, and only about 15% are eliminated by positive selection (Table 5.1). If the role of positive selection is to rid the body of pre-selection CTLs that are unlikely to respond to any MHC–epitope complexes, then lazy evaluation makes this task less important because it already ensures that all pre-selection CTLs generated are within the cross-reactive radius of an MHC–epitope complex. If the model generated all pre-selection clones, and not just the ones that could respond to an epitope, then positive selection would eliminate a higher proportion of cells. This hypothesis could be tested in the lab by measuring the number of pre-selection clones that respond to a particular MHC–epitope complex, then comparing this to the number that are eliminated by positive and negative selection. I believe that

negative selection would eliminate more pre-selection clones than positive selection.

Although the cells purged by positive selection would have been capable of responding to antigen, there is a tradeoff between the cost of maintaining these cells and the possible benefit of having them during an immune response. For many pre-selection cells, the probability of their responding to an antigen during an organism's lifetime is disproportionately small, and positive selection preferentially eliminates them. Although positive selection eliminates about 95% of all pre-selection CTLs in the body, the model indicates that positive selection reduces the number that could respond to an epitope only by about 60% (in Table 5.1, divide the number of clones killed in positive selection by the sum of the number killed in positive selection and the number of clones surviving selection).

5.1.7 Is the TCR repertoire optimized to detect foreign peptides?

There is a striking similarity between the model's CTL repertoire and Reduced Coulomb Energy (RCE) networks (Reilly et al., 1982). RCE networks are used to classify inputs into various categories. During a training phase, they are exposed to examples from each of the desired categories. Every training example is assigned a radial basis function detector. A detector is a hypersphere that covers a set of input values similar to each example, and any input that falls within a detector is considered to be of the same category as the example associated with the detector. The radius (size) of each detector is adjusted so that it covers as much of input space as possible without covering a training example from a different category. Thus, RCE networks attempt to cover as much of input space as possible without misclassifying inputs.

The model's CTL repertoire behaves like an RCE network. CTLs are radial basis function detectors that cover portions of peptide space classified as "foreign." The regions that are not covered by CTLs are implicitly considered "self." Like radial basis function detectors, CTLs can cover different-sized portions of space based on their peptide binding de-

generacies. Unlike standard RCE networks, the CTL repertoire is trained using examples from only one category—pre-selection CTLs are exposed to self peptides in the thymus. Therefore, rather than tuning detectors like an RCE network, the CTL model creates an excess of random detectors and eliminates those that are not optimal (Figure 5.6). These pre-selection CTLs must be screened to eliminate those that detect self peptides, a task accomplished by negative selection. Positive selection eliminates CTLs that do not cover enough peptides. Like the RCE network detectors, CTL coverage should be as broad as possible without covering a self peptide. The farther the distance between the CTL and its selecting self peptide, the more degenerate its peptide binding should be. Although the generation of CTLs is quite different than the RCE network training approach, the set of detectors that is generated by these processes have similar properties.

I believe that foreign peptide coverage is improved by using smaller detectors close to self and larger detectors that are far from self. Using different-sized detectors gives the immune system the ability to increase the fineness of foreign peptide coverage near self peptides, regions in which it must be highly discriminating, and have coarser coverage farther from self peptides. Thus, the wide range of peptide binding degeneracies observed in the lab could be a key component of antigen detection.

5.2 Basic dynamics of the CTL response

The following experiments illustrate the basic properties of the model using only one or two CTL clones. Although CTL responses normally involve many clones (and this case is covered in Section 5.3), it is easier to analyze the behavior of a single large population of T cells sharing the same specificity. The virus infection parameters used in these experiments are in Table 5.2.

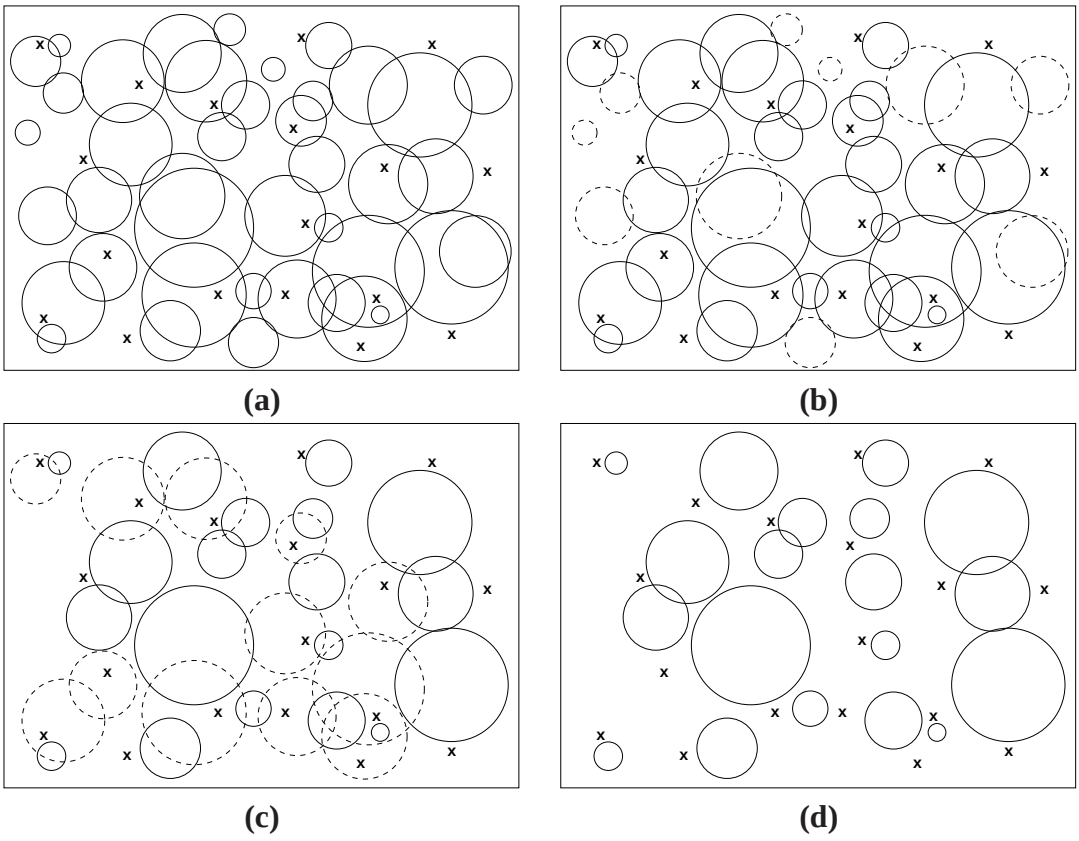


Figure 5.6: Thymic selection optimizes foreign peptide coverage. Figure (a) is a representation of the pre-selection CTL repertoire. Self peptides are denoted as $\times$s, and the CTLs are represented as circles, with the areas covered by each circle representing the peptides that the CTLs can bind. The distances between the $\times$s and the centers of the circles are proportional to the antigenic distances between the self peptides and the peptide-binding portions of the TCRs. In (b), CTLs that are eliminated during positive selection are indicated with dashed lines. CTLs that are eliminated by negative selection are indicated with dashed lines in (c). The CTLs that survive thymic selection are shown in (d).

5.2.1 Primary and secondary immune responses

I simulated the primary and secondary responses to an acute infection (Figure 5.7). For this trial, I was interested in testing the overall dynamics of the T cell response in the model rather than attempting to match the results to a particular laboratory experiment. I

attribute	*value*
susceptible cell population (T)	10^6 cells
susceptible cell production rate (λ)	10^5 cells/day
susceptible cell death rate (δ_T)	0.1 day^{-1}
virus infection rate (β)	2×10^{-7}
virus production rate (π)	100 day^{-1}
virus clearance rate (c)	2.3 day^{-1}
infected cell death rate (δ_I)	0.8 day^{-1}

Table 5.2: A summary of infection parameters used in Section 5.2.

simulated the injection of 500 viral units into a mouse with a single high-affinity T cell clone of 50 cells. The primary response began after approximately one day. It peaked at day 9 then declined and formed a stable memory pool. At day 28, an identical injection was administered, and the secondary response was faster and larger than the primary (Figure 5.7). The secondary response began almost immediately after secondary exposure to the virus, and the lower death rate of memory-derived effectors caused the T cell population to increase more rapidly. The secondary response also created a larger pool of stable memory cells. Therefore, the simulated mouse's immune memory could be "boosted" by multiple exposures to the same antigen, making future responses to it even more effective.

5.2.2 The programmed response

One of the implications of the programmed T cell response (described in Section 3.2.4) is that the immune response is initiated by antigen but its outcome is antigen-independent. If this is true, then removing antigen after the start of a response should not affect its dynamics. This was tested in mice infected by *L. monocytogenes* (Mercado et al., 2000; Badovinac et al., 2002). Antibiotics were administered to eliminate the infection 24 hours after inoculation, which quickly removed all antigen. The peak of the T cell response occurred at the same time in the antibiotic-treated mice and in non-treated control mice.

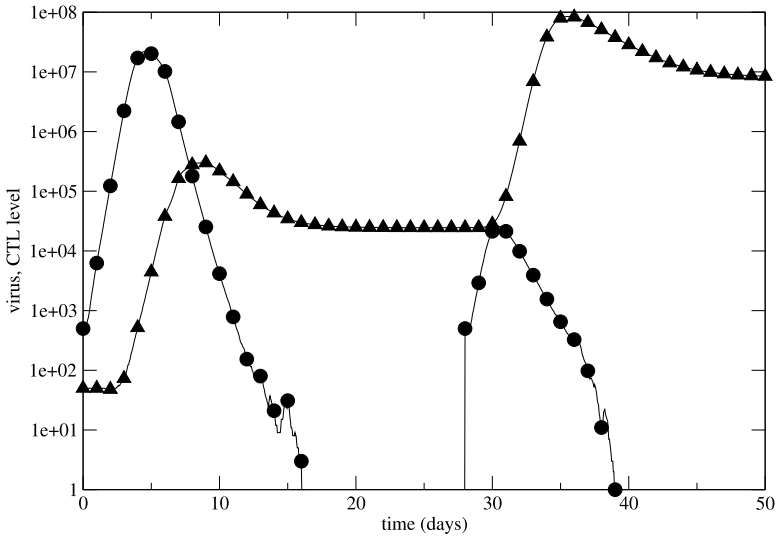


Figure 5.7: Primary and secondary CTL responses to viral infection. The primary exposure to the virus (●) is on day 0 and the secondary exposure at day 28. The number of T cells (▲) specific to this virus includes naïve, effector, and memory cells.

The elimination of the infectious agent caused only a small reduction in the magnitude of the response. Therefore, the elimination of antigen did not greatly affect the timing or magnitude of the T cell response.

The model gives qualitatively similar results in a system using LCMV parameters (Figure 5.8). Since antibiotic effects are not immediate and do not directly remove bacteria in mice infected with *L. monocytogenes*, I chose to eliminate all LCMV at 36 hours post-infection instead of 24. Eliminating antigen caused the peak viral load of the response to occur one day earlier and decrease only slightly in magnitude. The reduced response in the model was due to the shortened recruitment time of naïve cells.

Incorporating the programmed response might be essential to modeling the efficacy of vaccinations. Vaccines often use attenuated strains of pathogens that have diminished or no reproductive capacity and are rapidly cleared from the system. Since the purpose of vaccination is to induce a large response in order to build a large pool of specific memory

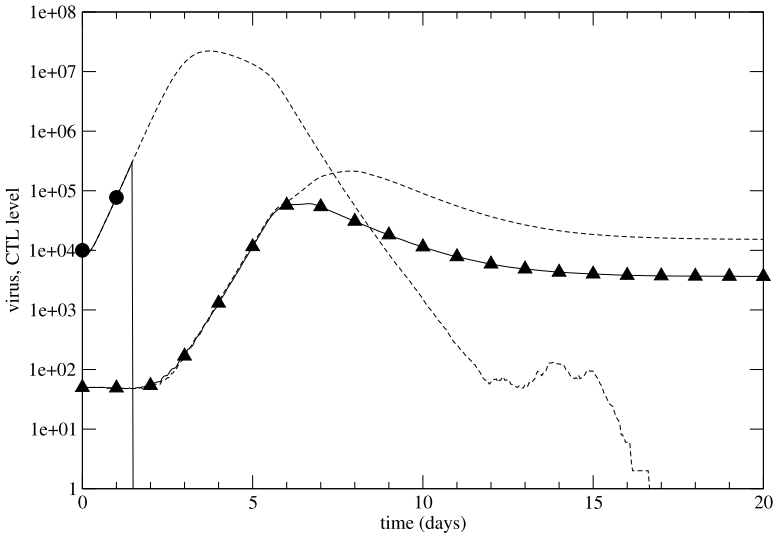


Figure 5.8: T cell response to an infection interrupted by treatment. The starting dose of the antigen (●) was 10,000 virus particles. The antigen was removed from the system after 36 hours. The T cell response (▲) is not significantly affected by the removal of antigen. For simplicity, only a single T cell specificity and a single antigenic epitope were used. The antigen and T cell levels of the control case, in which the antigen is not removed, are plotted for comparison (dashed lines).

cells, then a large dose of an attenuated virus might be effective even if the virus level drops rapidly. If the T cell response were totally antigen-dependent, short periods of antigenic stimulation would not stimulate an adequate response.

5.2.3 Naïve population size effects

The size of the initial naïve cell population can affect the outcome of an infection. Presumably, increasing the number of naïve cells can result in an earlier and larger response to infection. This hypothesis was tested experimentally in mice (Ehl et al., 1998). The number of naïve cells in mice was experimentally increased before infection in order to determine how the number of responding naïve cells affects the T cell response to an acute infection. It was estimated that about 50 naïve cells respond to LCMV in a normal

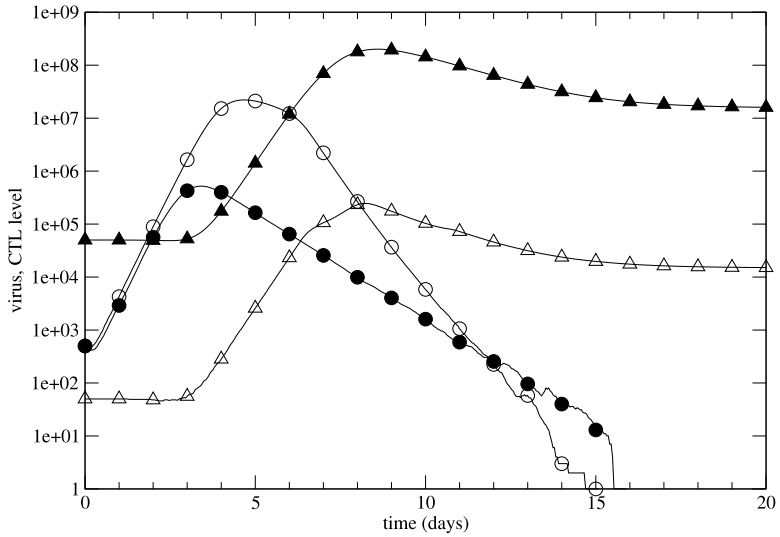


Figure 5.9: The effect of increasing the number of naïve cells. One model run was initialized with 50 naïve cells (△) and a viral load of 500 (○). The other model run started with 50,000 naïve cells (▲) and the same initial virus load (●).

mouse (Ehl et al., 1998), and the number was raised to 50,000 by adoptive transfer from donor mice. Increasing the number of naïve cells by 1000-fold moved the peak viral load of the infection between 1 and 2 days earlier and reduced the viral load by about 2 logs. In other words, the infection did not reach high levels. The model's results are in agreement with these experiments; after increasing the number of naïve cells from 50 to 50,000, the peak virus load was one day earlier and about 2 logs smaller than in the control case (Figure 5.9).

Surprisingly, the augmented immune response did not clear the infection more quickly in the model. The virus's reproductive rate is limited by the number of uninfected cells. A virus that is too prolific can exhaust the supply of new cells to infect. A weak immune response might allow the virus to infect most healthy cells. Once this happens, the virus is easier to eliminate because its spread is slowed. A stronger response might restrict viral spread early enough so that a large pool of uninfected cells is maintained, so the virus is

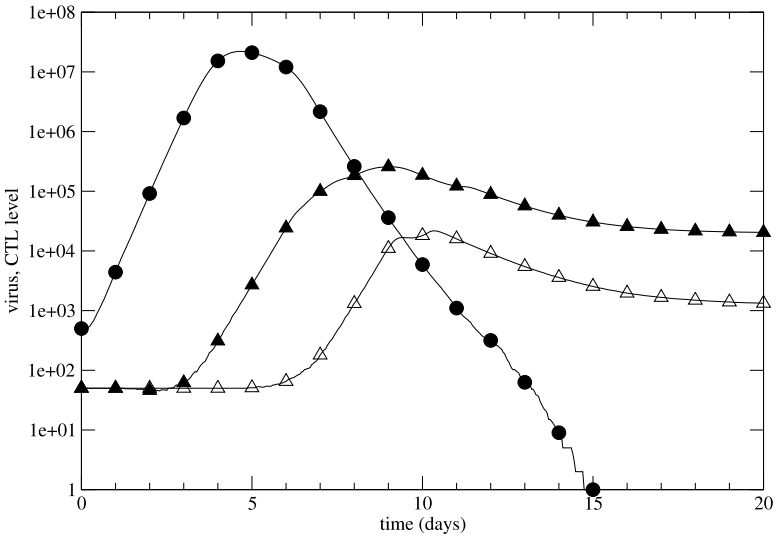


Figure 5.10: High- and low-avidity responses. The simulated viral load (●) is set to 500 on day 0. The high-avidity clone (▲) peaks about two days earlier than the low-avidity clone (△).

able to infect new cells as the response eliminates older infected cells. This effect can be observed in Figure 5.9. Note that the decline of viral load is slower when it did not peak at high levels.

5.2.4 High- and low-avidity responses

To study the clonal composition of the T cell response, I ran the model with a virus with a single epitope and two T cell clones with different avidities to this epitope, a high-avidity clone ($K = 7.8 \times 10^3$) and a low-avidity one ($K = 4.5 \times 10^7$). I assumed both clones initially contained 50 naïve cells each. The peak of the high-avidity clone's response is over one log greater than and over one day earlier than the low-avidity one (Figure 5.10). Scenarios involving larger numbers of clones are presented in the following section.

5.3 The clonal composition of T cell responses

One of the strengths of the model is that it can create a large repertoire of CTLs with different avidities to various antigens. Perhaps 20 T cell clones respond to a single epitope (Maryanski et al., 1996; Blattman et al., 2002). These clones have affinity not only for the epitope in question, but for a range of related epitopes. In a system subjected to heterologous infections, memory cells that cross-react to multiple antigens might be an essential part of our immune responses (Welsh and Selin, 2002). For example, it has been found that the CTL response to a particular hepatitis C epitope cross-reacts with an influenza A epitope. Thus, one may gain partial protection from one pathogen by exposure to an unrelated one. The digit string implementation, which implicitly defines an affinity between a TCR and any epitope, allows one to model the effect of heterologous infections over an organism's lifetime.

I simulated the response of a mouse with a realistic-sized repertoire to a viral infection. I used the xor distance rule (Appendix A.2) with an alphabet size of 128 and set the MHC string length to be 4 digits and the peptide string length to be 6 digits. The simulated mouse had 2.5×10^8 T cell specificities before thymic selection, but only about 200 of these were explicitly generated by the simulations, the remaining clones falling outside the cross-reactive cutoff of the antigen. Of these, approximately 20–30 survived the thymic selection process against 30,000 randomly generated self peptides to join the naïve repertoire (Table 5.1).

5.3.1 The primary response

I simulated the primary CTL response to a viral infection. Early in infection, antigenic levels were too low to stimulate T cell proliferation, so the naïve T cell population was stable. As the virus infected cells, the higher-affinity CTLs were stimulated and their

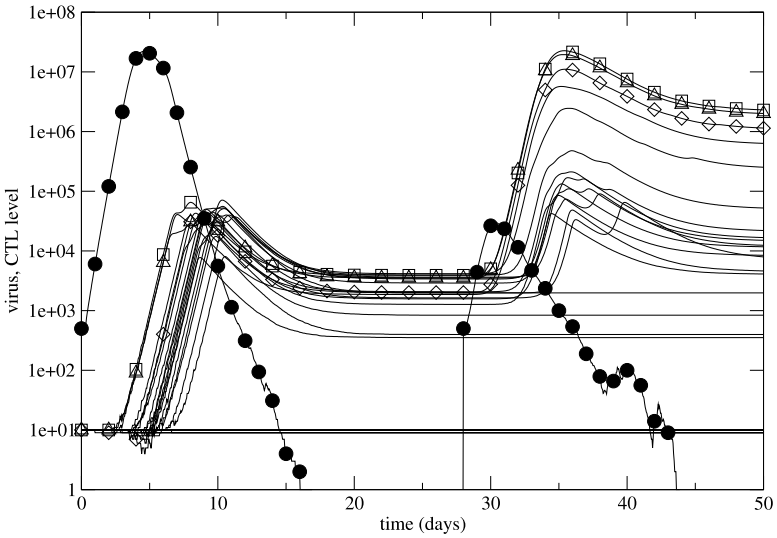


Figure 5.11: Primary and secondary CTL responses to a viral infection. 500 viral units were injected on days 0 and 28. The virus levels are indicated by • and the number of CTLs in the three highest-affinity clones as □, △, and ♢ (in decreasing order of affinity). Lower-affinity clones are represented by lines with no markers. Each CTL clone initially has 10 unstimulated naïve cells.

probability of entering the response increased. Low-affinity CTLs were later stimulated to join the immune response when antigen reached sufficiently high levels (Fig. 5.11). Thus, the entry of clones into the response was staggered, with progressively lower affinity clones tending to enter the response later. A similar observation has been made in murine systems: the contribution of a T cell clone to an immune response is largely determined by the time of its entry into the response (Bousso et al., 1999). Low-affinity clones sometimes responded more quickly than high-affinity ones because the simulation is stochastic. With a more slowly growing virus, this occurred less often because the more gradual rise in antigen levels led to a greater delay between the times of stimulation of high- and low-affinity T cells (data not shown).

Even among syngeneic mice, the CTLs involved in a primary response can have a variable mix of affinities for antigen (Bousso et al., 2000). Similarly, different runs of

the model with identical initial parameters had different responding clones. Because the initial number of cells in a single clone is small, stochastic effects play a large role in the composition of the primary response. In the model, a newly stimulated naïve T cell must survive a high death rate between the time of antigenic stimulation and the beginning of its programmed response, so that on average only 6 cells out of the 10 from a particular T cell clone survive to proliferate. Because the model is discrete and assumes that proliferation is antigen-independent, a response that begins with 1 to 6 proliferating cells will peak between 60,000 and 360,000 effector cells. This agrees with the estimate that only 1 to 6 cells per clone initiate CTL responses in mice and that individual clones produce between 4×10^4 and 3.7×10^5 cells at the peak of the response (Bousso et al., 1999). As a consequence of the antigen-independent proliferation of CTLs, memory levels formed by the primary response in the model are proportional to the initial number of cells that successfully enter proliferation because a constant fraction of effector cells formed convert to memory (about 5%).

The average affinity of T cells changed dramatically during the response to infection in the model. I define the average affinity of the response to be the inverse of the average K_d value (defined in Section 4.4) of all CTLs. Three days after infection, the average affinity rose (i.e., the average K_d fell) rapidly as high-affinity clones expanded (Fig. 5.12). The rising antigen levels progressively crossed the stimulation threshold of lower and lower affinity cells and recruited them into the response. As the T cell response peaked, the average affinity dropped (i.e., K_d rose) as the contribution of low-affinity clones to the overall response increased and the programmed expansion of high-affinity cells ended. The average affinity stabilized after day 10 as memory cells formed and dominated the population. These trends agree with observations made during experimental infection of mice with paramyxovirus simian virus 5: high-affinity $CD8^+$ T cell clones were exclusively detected early in the CTL response at day 3, but low-affinity clones comprised $\sim$50% of the response by day 5 post-infection (Gray et al., 2003). Similarly in the model, low-affinity clones comprised half of the response after day 7 post-infection. I also measured the

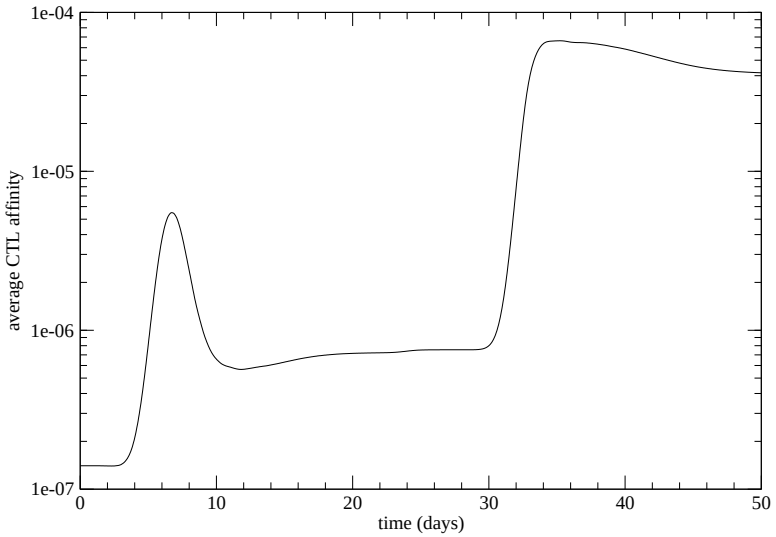


Figure 5.12: The average CTL affinity during primary and secondary responses to antigen. Affinity is 1 divided by the dissociation constant K_d, which is defined as the amount of antigen required to induce half-maximal stimulation in a CTL. 500 viral units were injected on days 0 and 28. The data plotted are the average values from 100 experiments.

affinity of the response as the ratio of low-affinity CTLs to high-affinity CTLs. This ratio rapidly dropped at the beginning of the CTL response then rose after day 7 (Fig. 5.13), which agrees qualitatively with observations in mice following infection with recombinant vaccinia expressing a well-characterized peptide antigen from ovalbumin: this ratio was initially high, dropped by day 6 post-infection, and returned to a high value in the memory population after the primary response (Alexander-Miller, 2000).

5.3.2 The secondary response

I simulated a secondary response to antigen by injecting additional virus into the system 28 days after a primary challenge. The T cell clonal hierarchy in the secondary response was more consistent across different simulation runs than that observed in the primary response. In the simulations of the secondary response to virus, I found that the same

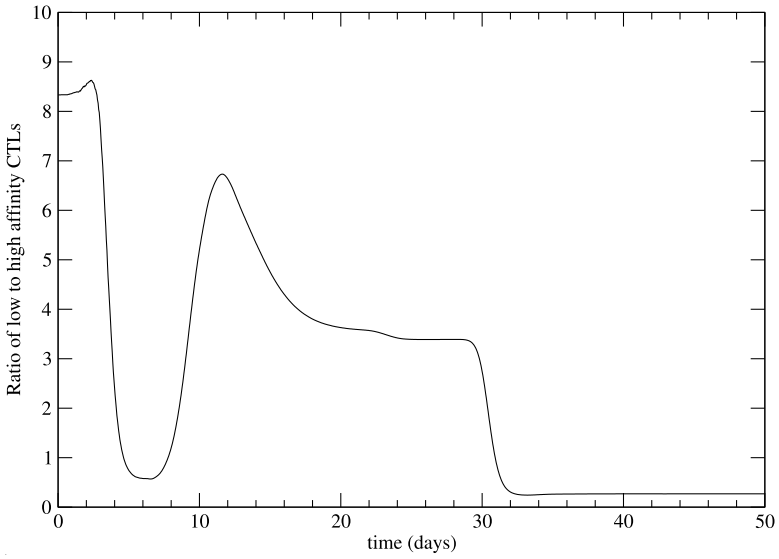


Figure 5.13: The ratio of low- to high-affinity T cells during a primary and secondary response to antigen. 500 viral units were injected on days 0 and 28. The data plotted are the ratios of the number of cells of the 26 lower-affinity clones to the 3 highest-affinity clones averaged over 100 experiments.

highest-affinity T cell clones were dominant, while a variable mix of lower-affinity clones comprised a small fraction of the response. The recruitment of high-affinity memory cells drove a second increase in average T cell affinity for antigen (Fig. 5.12).

The model results agree with observations that the clonal composition of the secondary response in mice varies less than the primary among syngeneic animals (Bousso et al., 2000; Bachmann et al., 1997; Busch et al., 1998a; Blattman et al., 2000; Kedzierska et al., 2004), that the secondary response is composed of a smaller set of responding clones (Savage et al., 1999), and that while the primary response recruits a mix of high- and moderate-affinity clones, the secondary preferentially recruits high-affinity clones (Estcourt et al., 2002). In the simulations, this consistency of the secondary response compared to the primary occurs because of the larger number of cells involved. As discussed above, precursor frequencies are low in the primary response, allowing stochas-

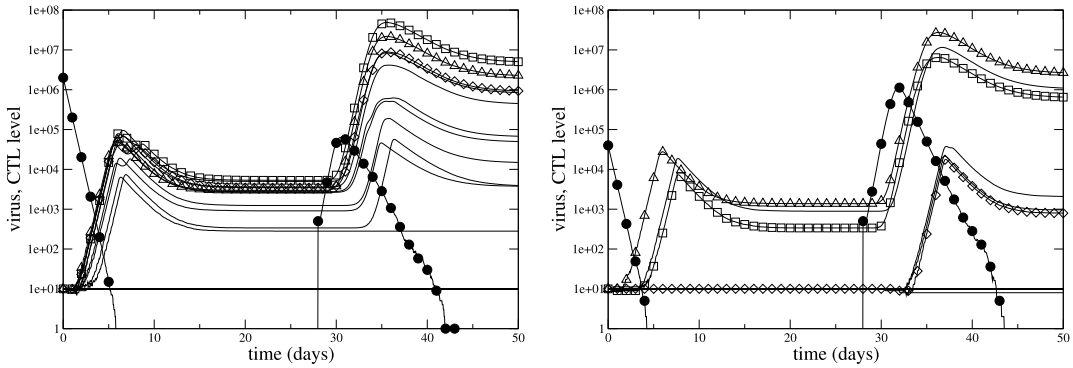


Figure 5.14: CTL responses to non-replicating virus followed by replicating virus challenge. The non-replicating virus dose on day 0 was (a) 2×10^6 units and (b) 4×10^4 units. The replicating virus challenge of 5,000 viral units was administered on day 28 in both experiments. Virus levels are indicated by • and the number of CTLs in the three highest-affinity clones as □, △, and ◊ (in decreasing order of affinity). Lower affinity clones are represented by lines with no markers.

tic effects to determine whether the first cell to proliferate will come from a high- or low-affinity clone. By contrast, there is a large number of cells per clone in the secondary response, and the hierarchy of responding cells is therefore much more stable among simulation runs.

5.3.3 Non-replicating antigen

I simulated immunization with 2×10^6 viral units of non-replicating antigen. This immunization created a sharp spike in the antigen level that rapidly decayed. The high initial antigen load maximally stimulated all T cells with an affinity above a certain threshold (dependent on the antigen dose). This is in contrast to infection with replicating antigen, in which the gradually increasing antigen stimulates high-affinity clones first and gives them a time advantage over the lower-affinity clones. If these high-affinity clones clear the infection quickly, then low-affinity clones receive insufficient antigenic stimulation to

be recruited into the response. This time advantage is not a factor in infection with non-replicating antigen, in which the sharp spike in antigenic stimulation caused clones of different affinities to peak simultaneously (Fig. 5.14a). Because the model features antigen-independent proliferation, the high-affinity clones do not interfere with the proliferation of low-affinity clones that have already been stimulated. Therefore, non-replicating antigen creates a flatter distribution of high- and low-affinity clones, with the average affinity being dependent on the antigen dose. The decay phase of antigen provides a period during which high- and low-affinity clones receive different degrees of stimulation. That is, as antigen levels progressively decline, only high-affinity cells are stimulated. This occurs for both replicating and non-replicating antigen, as both undergo a decay phase. However, this effect probably makes only a small contribution to differentiating high- and low-affinity cells for two reasons: (i) it might occur during the phase of antigen-independent proliferation, and (ii) if antigen decay is very rapid, there is little time difference between when the stimulation thresholds of high- and low-affinity cells are crossed.

A variety of experiments suggests that a higher-affinity response can be recruited with lower doses of antigen (Rees et al., 1999; Alexander-Miller, 2000; Walter et al., 2003). Presumably, low doses of antigen cannot stimulate low-affinity clones, but can stimulate high-affinity ones. These high-affinity clones appear to be better for infection control (Alexander-Miller et al., 1996; Derby et al., 2001). To investigate this phenomenon, I simulated inoculation with a smaller dose of 4×10^4 viral units of non-replicating antigen. Fewer clones responded to the low dose (Fig. 5.14b) than the high dose (Fig. 5.14a). The low dose produced memory cells with a higher average affinity for antigen than the high dose. However, because the low dose recruited small numbers of T cells, systematic differences in affinities recruited by the different antigen doses were sometimes obscured by stochastic effects. When used as a vaccine, the smaller antigen dose afforded less protection against subsequent infection by virus, allowing the virus to peak at levels three times higher than in the trial with the larger antigen dose. The large number of memory cells of various affinities formed in response to the high-dose vaccine provided better protec-

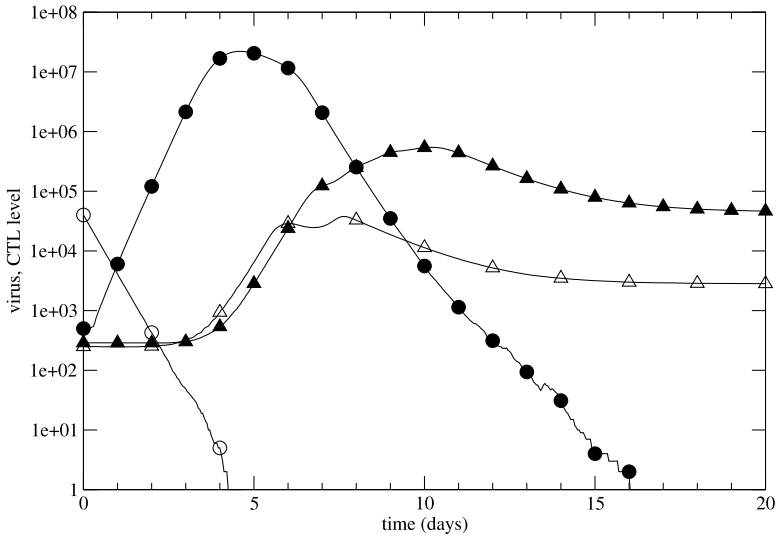


Figure 5.15: Comparison of responses to replicating and non-replicating virus challenges. For the replicating virus infection, the virus levels are indicated by • and the total number of CTLs by ▲. For the non-replicating antigen, the antigen levels are indicated by ○ and the total number of responding CTLs by △. The data in this figure are drawn from the experiments shown in Figs 5.11 and 5.14b.

tion than the small number of high-affinity cells from the low-dose vaccine. The lack of increased protection using low doses might be because the simulation does not include direct competition between clones. Thus, the same set of high-affinity clones are stimulated with high- and low-dose antigens in the simulation and grow equally well, while in an animal vaccinated with a low dose, these high-affinity clones may expand more due to a lack of competition with low-affinity clones for resources.

Comparing the dynamics of the CTL responses to replicating and non-replicating virus infection yielded results similar to those found in mice responding to a killed bacteria vaccine (Lefrançois et al., 2003). In both the computer model and the mouse experiments, the CTL levels in the replicating and non-replicating virus scenarios were indistinguishable on day 5 (Fig. 5.15). However, the responses soon diverged, with the response to the replicating virus peaking days later while the response to the non-replicating declined.

	number of clones			distance to self peptide		
	resolved	unres.	p-value	resolved	unres.	p-value
Hamming	27.13	24.57	5.07×10^{-10}	18.93	19.12	2.63×10^{-2}
xor	23.97	21.18	2.20×10^{-16}	76.75	69.36	2.82×10^{-16}
$L1'$	29.16	25.36	2.20×10^{-16}	9.76	8.50	2.20×10^{-16}

Table 5.3: Differences between repertoires that cleared infection and those that did not. For each of the metrics, the CTL simulation was run 1,000 times, and the infection was considered to be resolved if there was no virus present four weeks after infection. The second and third columns show the average number of CTL clones instantiated per epitope in the resolved and unresolved cases. The fifth and sixth columns show the antigenic distance between the epitope and the nearest self peptide in the resolved and unresolved cases. The p-values are from two-sample t-tests.

The final memory cell level induced by the replicating virus infection was about an order of magnitude larger than that from the non-replicating antigen.

5.3.4 The number of CTL clones per epitope

The CTL response does not always resolve infections in the model. Because the naïve repertoires are generated stochastically, some are more capable of eliminating particular antigens than others. I compared the repertoires that were able to clear an infection and those that could not. I ran the simulation 1,000 times for each of the three distance metrics (Hamming, xor, and $L1'$) using an initial virus dose of 1,000. The virus dynamics parameters are from Table 5.2, and infected cells express a single epitope. I considered the infection to be resolved if there was no virus present four weeks after infection. For the Hamming distance trials, 80% of the infections were resolved. For the xor and $L1'$ trials, 61% and 69% were resolved, respectively. The differences in these rates is not a property of the metrics themselves, but reflect the difficulty of calibrating the different metrics to produce similar results.

For each epitope instantiated during a simulation, a set of naïve CTL clones is created

using lazy evaluation (Section 4.2.2). All of these clones have some affinity for infected cells expressing this epitope, although not all of them participate in the response because of stochastic effects and indirect competition among clones for antigen. A higher number of instantiated clones correlates with a higher probability of pathogen clearance in the model. For all three distance metrics, the average number of clones instantiated was higher for cases in which the antigen was cleared than in those in which it was not (Table 5.3). Somewhat surprisingly, the antigenic distance from the epitope to the nearest self peptide also had an effect for two of the three distance metrics. The average distance was higher in the resolved cases for the xor and $L1'$ trials (Table 5.3). It appears that this is because the less similar an epitope is from all self peptides, the larger the number of clones generated for that epitope by the simulation for the xor and $L1'$ cases (Figure 5.16). This is probably because negative selection eliminates pre-selection CTLs that are too "close" to self peptides. For Hamming distance, the distance between the epitope and the nearest self peptide did not correlate with either the resolution of infection (Table 5.3) or the number of CTL clones instantiated (Figure 5.16a). These results are consistent with those in Section 5.1.5, in which the Hamming distance between an epitope and the nearest self peptide had no effect on CTL peptide binding degeneracy.

5.4 Summary

In this chapter, I tested the CTL model described in Chapter 3 under a wide range of conditions. Most of the model's results agree with the experimental literature, and the model makes a large number of predictions for laboratory experiments that have not yet been performed. Some of the results probe the composition of the naïve CTL repertoire and its ability to detect antigen. The remaining results explore the dynamics and efficacy of the CTL response.

The results of my model reveals effects of thymic selection that are different than pre-

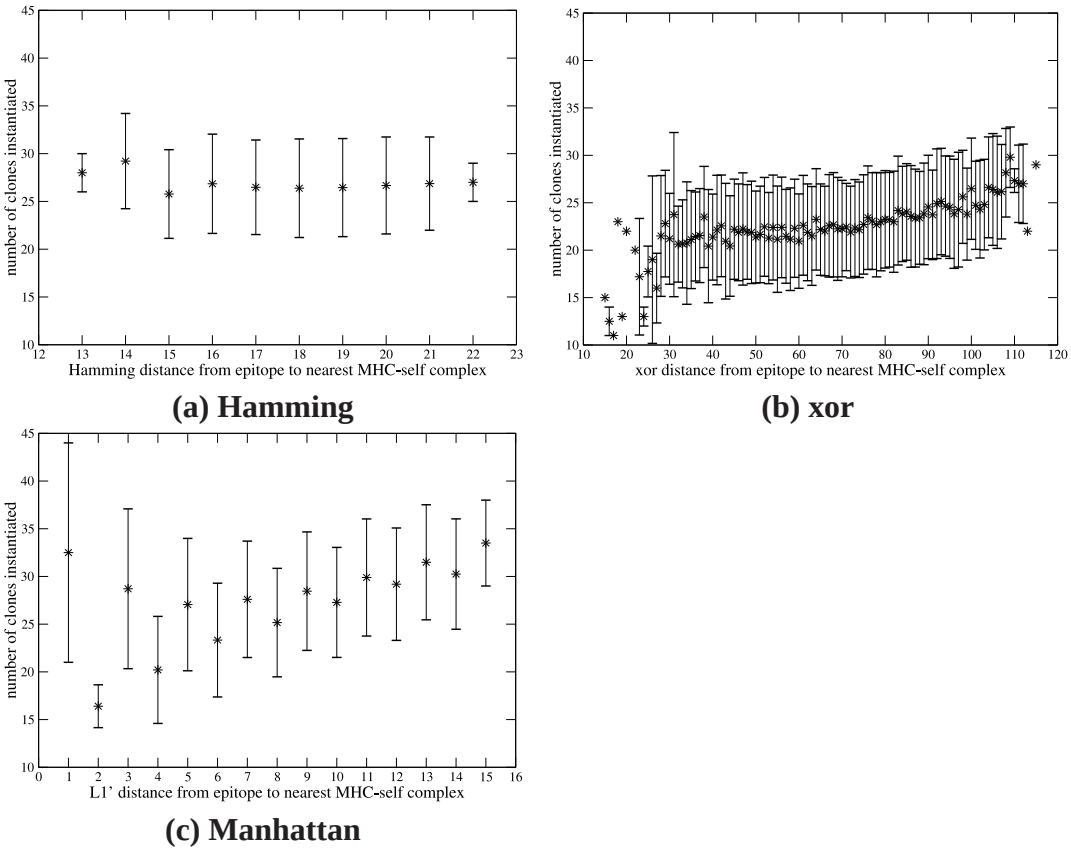


Figure 5.16: Number of clones vs. distance between an epitope and its closest self peptide. 10,000 random epitopes were generated, and the distances between these epitopes and their nearest self peptides were measured. A new CTL repertoire was created for each of these epitopes, and the average and standard deviation of the number of CTL clones is plotted against the epitope to selecting peptide distance. The results shown are from simulations using three distance metrics: (a) Hamming, (b) xor, and (c) modified Manhattan distances.

viously believed. The standard view of immunologists is that the role of negative selection in the thymus is to eliminate CTLs that respond to self peptides and the role of positive selection is to eliminate CTLs that can not bind MHC. The results from the model reveal more complex effects. Because TCRs bind to both peptide and MHC, negative selection also affects the CTL repertoire's affinity for MHC and positive selection affects its affinity for self peptide. In the model, negative selection eliminates both CTLs that are have high

affinity for self peptides and those that have high affinity for MHC. Positive selection not only eliminates CTLs that have low affinity for MHC, but also those with low affinity for all self peptides. I have found evidence of these effects in the literature. An implication of these findings is that thymic selection does not only eliminate CTLs that would be detrimental or useless in an immune response, but those that are inefficient and suboptimal. Thus, the process of thymic selection can be cast as an optimization problem.

The dynamic behavior of the model reproduces many observations of CTL responses in mice. In normal infections, the secondary response is larger and faster than the primary because of the larger pool of CTL and the shorter delay in the response of memory cells. One consequence of the greater magnitude of secondary responses is that they are more consistent among identical individuals than the primary response. The immune system can be manipulated in the laboratory to produce different behaviors. For example, antigen can be eliminated by massive doses of antibiotics or the immune response can be boosted by injecting extra T cells. The behavior of the model agrees with real-world experiments in these situations, and this agreement indicates that the model could be used to predict the outcome of similar laboratory experiments. One surprising result is that the immune response could clear an infection that reaches high viral loads faster than an infection that does not reach high levels. When a virus is too successful and infects most of the body's cells, the immune system can eliminate it quickly. If the infection is less extensive, then the infection is prolonged because the virus spreads to uninfected cells even as the immune system is eliminating infected cells. Thus, it is possible that a less effective immune response could result in a better outcome for the organism.

Other benefits of using a computer model instead of mouse models include the ability to make extremely detailed observations and to reproduce experiments exactly. For example, the affinity of responding cells may determine the body's ability to eliminate an infection, and the model makes predictions about the composition of the response that are supported by real-world observations. Low doses of vaccine recruit only the highest-

affinity CTLs, while high doses recruit a broad range of affinities.

Three versions of the model, each using a different definition of antigenic distance, were tested. The versions using xor and modified Manhattan distance produced results consistent with each other, while the Hamming distance version yielded different results in Sections 5.1.5 and 5.3.4. Hamming distance might be fundamentally different than the other two distance metrics, or these discrepancies could simply be due to the choice of parameters used in the models. It is not surprising that the xor and modified Manhattan distance versions agreed—they produce similar affinity distributions in the model (Section 4.4). The Hamming metric differs because it is coarse—the distance between digits is either 0 or 1, while the distance between digits in the other two metrics covers a range of values. This makes it difficult to calibrate the Hamming version of the model to match the others. It is possible that the differences in the metrics would disappear if longer string lengths were used to represent the TCRs and the peptides, but the number of possible peptides and TCRs would be unrealistically large and the model would be difficult to run. For reasonable string lengths, these differences will undoubtedly affect the model's results.

Chapter 6

Immune exhaustion and mutating pathogens

La perfection est atteinte non quand il ne reste rien à ajouter, mais quand il ne reste rien à enlever.
[You know you've achieved perfection in design, not when you have nothing more to add, but when you have nothing more to take away.]

—Antoine de Saint Exupery

So little of what could happen does happen.

—Salvador Dalí

The model presented in Chapter 3 used to produce the results in Chapter 5 assumes that viral infections are resolved quickly. When the immune system can not clear an infection quickly or if the body is repeatedly exposed to antigens, CTLs can behave differently. During long-term infections, CTLs can die from over-stimulation, a phenomenon known as *exhaustion*. To explore long-term dynamics in the model, I added exhaustion to the

model (Section 6.1). Without this feature, the model can produce unrealistically large primary responses to antigens, and this effect is compounded if the infection is not cleared by the primary response, leading to wild oscillations in CTL and virus levels. Adding exhaustion to the model not only affects the ability of the immune system to clear an infection (described in Section 6.2), but it also allows one to observe a response that lasts for longer periods of simulated time. As an illustration of a prolonged CTL response, I record the effects of a rapidly mutating virus in Section 6.3.

The results described in this chapter should be interpreted with caution. Experimentalists have studied acute responses more thoroughly than the long-term dynamics of the immune system. This is due, in part, to their reliance on mouse models. Not only do mice have short life spans, but immunological assays often require the mouse to be killed. Instead of tracking individual mice over time in longitudinal studies, researchers usually begin with a cohort of identical mice and sacrifice them at different times to simulate a time series. This approach is problematic if the mice are not identical or if stochastic effects play a significant role in the immune response.

Long-term dynamics are also difficult for modelers. For some infections, one can assume that the CTL response is so fast and effective that the contribution of other components of the immune system is minor. However, if the infection is not resolved quickly, then the roles of other immune cells, such as helper T cells, B cells, and macrophages, cannot be ignored. Therefore, realistic models of long-term infections need to be considerably more complex. The mechanism for immune exhaustion proposed in this chapter is intended to be biologically plausible but not complete.

6.1 Implementation

When over-stimulated by antigen, CTLs can become anergic or even die, a phenomenon known as *exhaustion*. Prolonged exposure to antigen appears to cause effector CTLs to become progressively more impaired, eventually leading to T cell deletion (Fuller and Zajac, 2003; Wherry et al., 2003). High doses of antigen or moderate doses of antigens that express excessively high epitope levels can also induce apoptosis in CTLs (Moskophidis et al., 1993; Wherry et al., 1999, 2002). Exhaustion might be a *peripheral tolerance* mechanism to eliminate self-reactive T cells. If thymic selection does not eliminate all T cells that react to healthy cells (i.e., central tolerance fails), then these cells will react to the extremely high constant levels of self peptides in the body. Exhaustion might prevent self-reactive T cells from effecting a sustained response against healthy tissue (Anderton et al., 2001).

I assume that exhaustion is induced by antigenic stimulation because high-affinity T cells are preferentially eliminated by high doses of antigen (Anderton et al., 2001) and those that respond to immunodominant epitopes appear to be more susceptible to exhaustion than those that respond to subdominant epitopes (Aichele et al., 1997; Zajac et al., 1998; Slifka et al., 2003). It has also been found that death of effector T cells by antigenic stimulation in vitro is dose-dependent (Iezzi et al., 1998).

To add CTL exhaustion to the model, I introduce an additional effector cell death term based on the level of antigenic stimulation that it receives. Recall that stimulation was defined in Equation 3.13:

$$\text{Stimulation} = \frac{\sum \frac{e_i I_i}{K_i}}{1 + \sum \frac{e_i I_i}{K_i}} \tag{6.1}$$

Stimulation determines the rate at which naïve and memory cells are recruited into the response. If this level of exposure to antigen recruits cells, then I assume that a higher level will cause these cells to die. Therefore, I introduce a new term, "over-stimulation,"

which is identical to stimulation except that the dissociation constants K are multiplied by 25:

$$\text{Over-stimulation} = \frac{\sum \frac{e_i I_i}{25K_i}}{1 + \sum \frac{e_i I_i}{25K_i}} \tag{6.2}$$

Like the original stimulation term, over-stimulation is a function that saturates at high levels of antigen, except that it requires higher levels of antigen for it to reach its maximum. In addition to their normal death rates, effector CTLs are also subjected to a death rate of $\delta_s = 5$ day^{-1} times the "over-stimulation." This rate needs to be high to counter the high proliferation rate of effector CTLs.

Overexposure to antigen also appears to impair the formation of immunological memory. Memory T cells formed during a persistent infection can be unresponsive to antigen (Masopust et al., 2004; Wherry and Ahmed, 2004). Therefore, the effector cells that are in the process of converting to a memory phenotype die at a rate of the stimulation (not the over-stimulation) times 1.0 day^{-1}. The effector cells at the end of their programmed division cycles (see Section 3.2.4) also die at this rate.

In summary, naïve CTLs convert to effector cells upon exposure to antigen. If effector cells are exposed to levels of antigen much higher than that necessary to recruit them, they die of over-stimulation. If there is antigen present at the end of the primary response, most effectors die without converting to memory cells.

6.2 Viral dynamics and viral clearance

It has been observed that LCMV infection (see Section 2.1.4) is more likely to be chronic if the virus is administered at high doses or if a more virulent strain is used (Moskophidis et al., 1993). I used the model to simulate infections with slow- and fast-replicating viruses administered at low (500 units) and high (10,000,000 units) doses. For the slow-replicating virus, I set $\pi = 65$ and $\beta = 1 \times 10^{-7}$ and for the fast-replicating

virus, I set $\pi = 100$ and $\beta = 2 \times 10^{-7}$. Section 3.1 defines these parameters. Typical runs are shown in Figure 6.1. In many cases, the initial CTL response does not resolve the infection. When the primary response fails to eliminate the virus, the viral load rebounds, and the memory cells formed by the first effector CTLs eliminate the virus a few weeks later in a secondary response (e.g., Figures 6.1c and d).

For the slow-replicating virus, the immune system could clear the infection more easily for low- than high-dose exposures (Table 6.1). For the fast-replicating virus, the immune system could rarely clear the infection quickly, regardless of the initial dose (Table 6.1). These results are qualitatively consistent with LCMV observations. However, the dynamics of a prolonged infection do not agree with laboratory observations.

In the lab, a persistent virus can survive at low levels in an apparent dynamic equilibrium with the immune response. In the model, viral replication and CTL clearance never find an equilibrium value, and the viral load declines during a CTL response and rapidly recovers when the programmed response ends. Additions to the model could allow the virus to persist at low levels. One extension would be the introduction of spatial compartments to the model. In the current model, the immune cells are assumed to be evenly mixed throughout the body, so the entire body is under the same level of immune surveillance. If the body in the model were compartmentalized, then spatial heterogeneity would give the virus the ability to temporarily evade the immune response by moving to new compartments. The virus might be able to survive at low levels under these circumstances. Adding complexity to the CTL response would also change the kinetics of viral clearance. The model assumes that an effector CTL will eliminate infected cells at the maximum rate until its programmed response ends or the cell is killed by exhaustion. In fact, real effector cells gradually lose effector functions (Fuller and Zajac, 2003; Wherry et al., 2003), and they can even recover functions after losing them (Schwartz, 2003). Adding a wider range of CTL response levels to the model could allow the virus to persist at an equilibrium level.

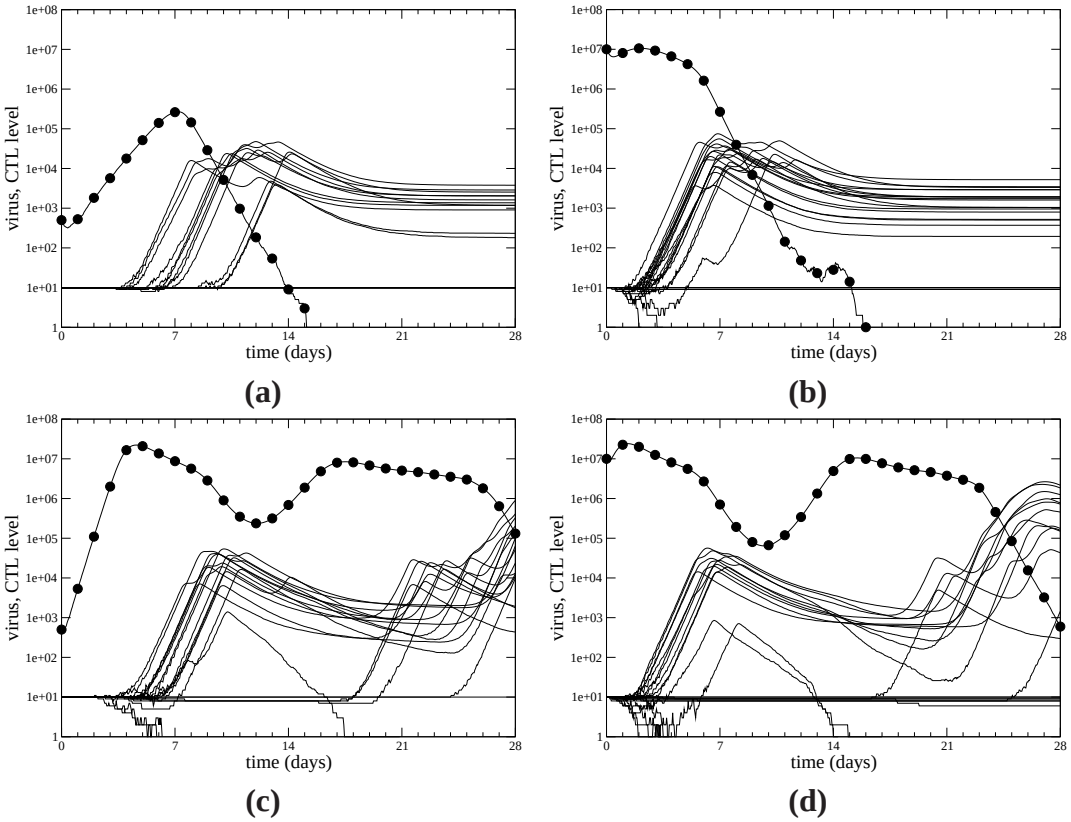


Figure 6.1: The resolution of infection with slow- and fast-replicating virus strains introduced at low and high doses. The low dose was 500 viral units, and the high dose was 10,000,000. The plots show representative runs initialized with (a) low-dose of slow-replicating virus, (b) high-dose of slow-replicating virus, (c) low-dose of fast-replicating virus, and (d) high-dose of fast-replicating virus. The virus loads are indicated by lines marked by • and the numbers of CTLs from each clone are plotted as unmarked lines.

6.3 Immune escape

Some antigens can alter their epitopes in order to evade the immune response. If the immune response targets a particular immunodominant epitope, infected cells expressing variants of this epitope might partially or fully avoid detection by effector cells. The immune system subjects pathogens to evolutionary pressure. Rapidly mutating pathogens can generate thousands of new antigenic strains within a single host, and the successful

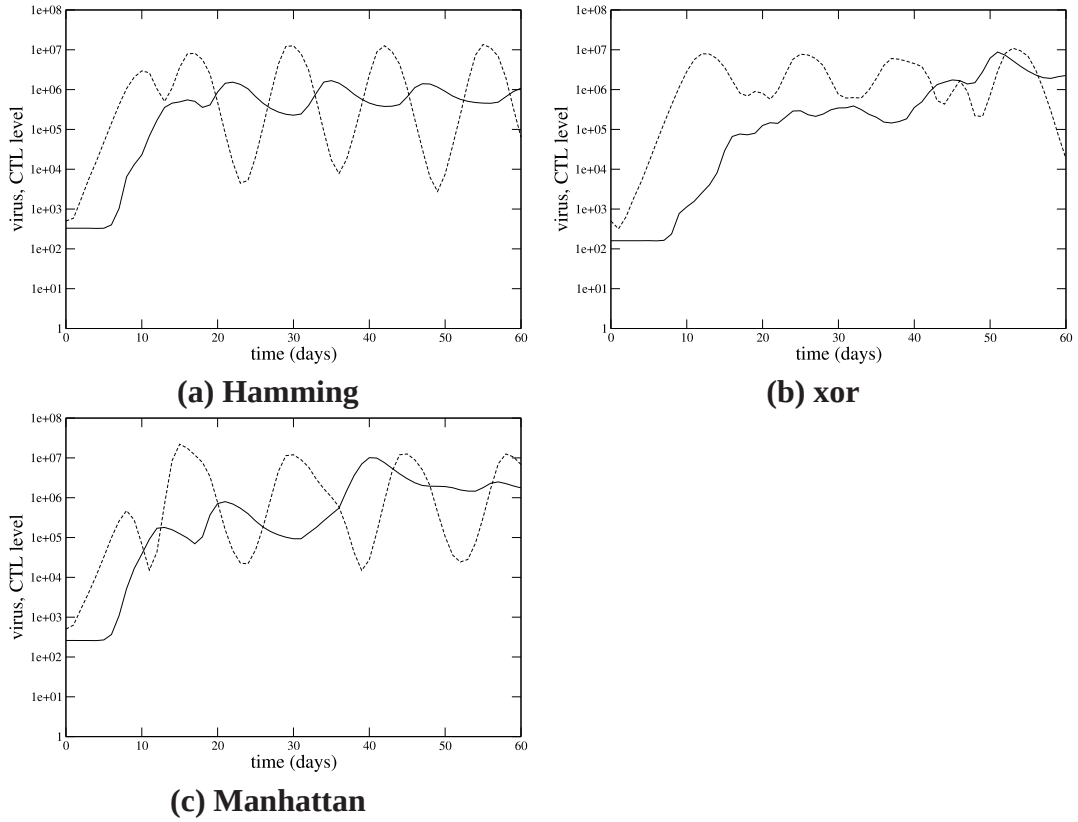
	slow		fast	
	low-dose	high-dose	low-dose	high-dose
Hamming	99%	99%	2%	8%
xor	41%	20%	0%	0%
$L1'$	64%	52%	1%	0%

Table 6.1: The resolution of infection with slow- and fast-replicating virus strains introduced at low and high doses. The low dose was 500 virus particles, and the high dose was 10,000,000 virus particles. For each set of parameters, the CTL simulation was run 100 times, and the numbers in the table indicate the percentage of these runs in which the infection was resolved, which I define to be the absence of virus four weeks after infection.

mutants can quickly spread, while the unsuccessful ones disappear. The immune system is also highly adaptable, and the host and pathogen must compete for control of the body.

I ran simulations with mutating pathogens that express a single epitope and have the same growth parameters as the slow-replicating virus described in Section 6.2. The non-mutating version of this virus, when administered at low doses, could usually be cleared within four weeks (Table 6.1). By adding a mutation rate of 10^{-5} mutations per virus replication (mutation is defined in Section 4.3), the infection was rarely cleared. Mutant strains are assigned the same replication and mutation rates as their parent strains. The total virus loads oscillated, peaking about every two weeks. Typical runs are shown in Figure 6.2.

The progression of individual virus strains is plotted in Figure 6.3. The dark vertical bands in these plots correspond to the simultaneous peaks of multiple strains. The strains that comprise each peak are not necessarily created at the same time; one can see that the lines that represent each strain can begin at different time points but still peak at the same time. Therefore, I assume that the CTL response periodically suppresses the spread of most strains, which proliferate after the response diminishes. Most strains do not survive these purges. The peaks in the viral loads are followed by bursts of new mutants, which appear in in Figure 6.3. as groups horizontal lines that start at the same point on the x axis.

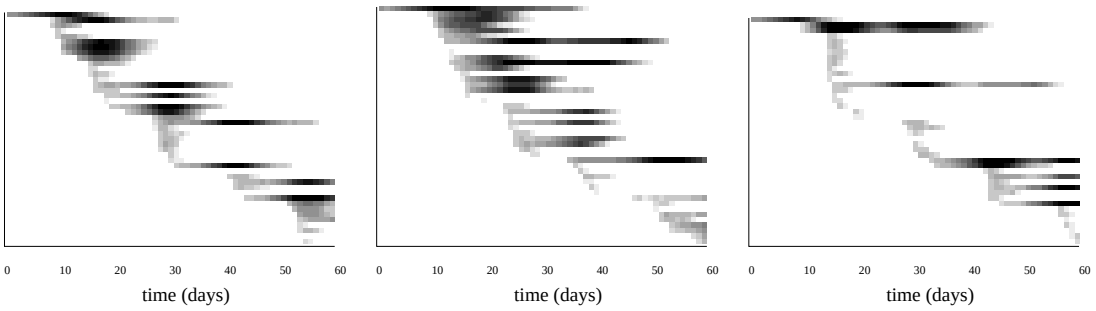


(a) Hamming **(b) xor**

(c) Manhattan

Figure 6.2: The viral load of a slowly mutating pathogen. The mutation rate of the virus is 1 mutation for every 10^5 replications. Solid lines represent CTL levels, while the dashed lines represent the total viral load of all strains. Versions of the model using different definitions of antigenic distance were used: (a) Hamming, (b) xor, and (c) modified Manhattan distance.

From these observations, I assume that the original virus replicates until it reaches high levels and generates variants. The CTL response eliminates the original strain and most of its descendents, but some of the strains will survive, possibly because their epitopes escape immune system detection or possibly just by luck. The growth of these survivors is temporarily suppressed by the cross-reactive CTL response, which eliminates cells infected with the original virus and similar strains. Because the programmed response of effector cells lasts for only a few days, each response will eventually end. When the initial response is over, the surviving strains soon generate a second peak in viral load, which

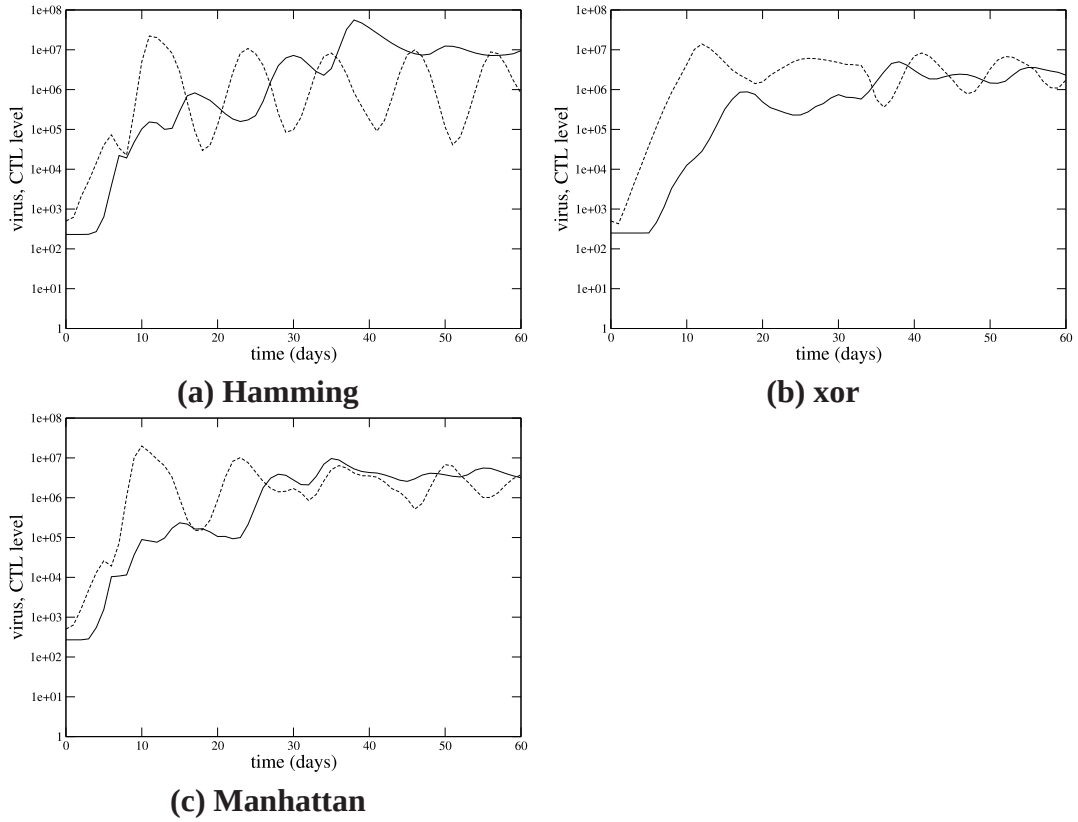


(a) Hamming **(b) xor** **(c) Manhattan**

Figure 6.3: Emergence of strains using slowly mutating pathogens. The individual strains are displayed from the same trials shown in Figure 6.2. The virus load of each distinct strain over a period of sixty days is represented as a row of gray dots, and the darkness of each dot is proportional to the log of the viral load of that strain at a time point. Time runs along the x-axis. Versions of the model using different definitions of antigenic distance were used: (a) Hamming, (b) xor, and (c) modified Manhattan distance.

recruits a new set of CTLs. A new set of strains is created during this peak.

The viral load peaks of multiple strains do not remain synchronized. If a new mutant virus has an epitope that is not recognized by effector CTLs, then it can replicate before the response to its parental strain is over. This can be observed when the mutation rate was raised to 10^{-4} mutations per virus replication. Figure 6.4 plots the total virus loads of representative runs, and Figure 6.5 shows the progress of the individual strains in these runs. Although the viral load peaks are synchronized at first, they quickly go out of phase and the peaks become less coherent over time (Figure 6.4). The faster mutation rate quickly generates mutants that express epitopes that evade the immune response. The responses to these different strains is independent, thus their dynamics are not synchronized by cross-reactive CTL responses. This effect is less apparent when Hamming distance is used to define antigenic distance (Figures 6.4a and 6.5a). The viral loads of the various strains appear to remain synchronized. In the Hamming distance version of the model, a single mutation in an epitope never allows a virus to escape CTL detection (Section 4.3). Therefore, the cross-reactive CTL response can keep a strain and its mutants synchronized,



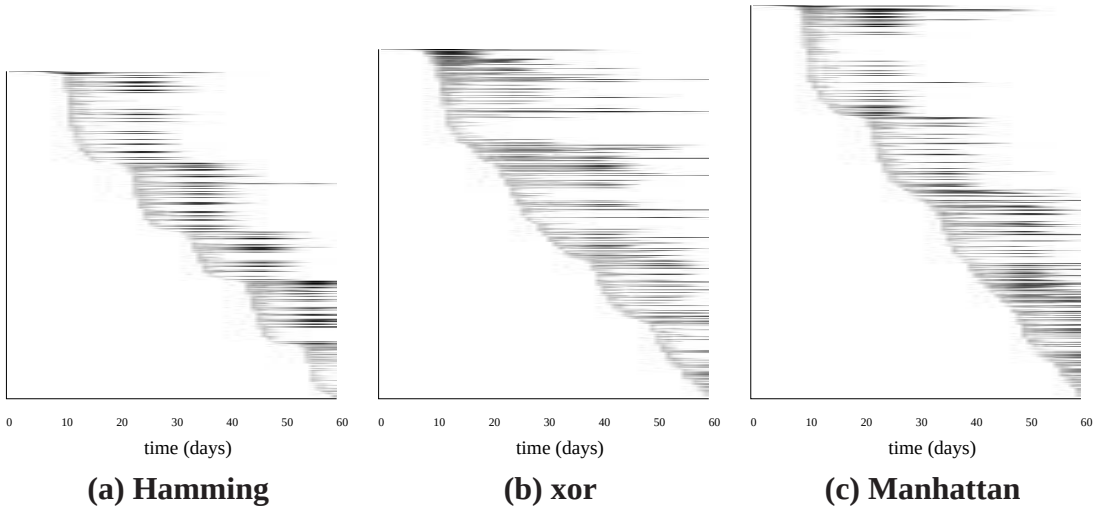
(a) Hamming **(b) xor**

(c) Manhattan

Figure 6.4: The viral load of a quickly mutating pathogen. The mutation rate of the virus is 10^{-4} mutations per replication. Solid lines represent CTL levels, while the dashed lines represent the total viral load of all strains. Versions of the model using different definitions of antigenic distance were used: (a) Hamming, (b) xor, and (c) modified Manhattan distance.

perhaps until strains accumulate multiple mutations.

The mutation rate does not significantly affect the rate at which virus strains can accumulate mutations. It is unlikely that a virus will produce a mutant strain until it reaches high population levels, at which point it can produce many new strains that differ from itself by exactly one mutation. The CTL response to the parent strain keeps these new viruses at low levels or eliminates them altogether. Therefore, these new strains will not replicate widely enough to generate their own mutant strains (which would differ from the parent by two mutations) until the previous CTL response diminishes. Thus, the number



(a) Hamming **(b) xor** **(c) Manhattan**

Figure 6.5: Emergence of strains using quickly mutating pathogens. The individual strains are displayed from the trials shown in Figure 6.4, Each distinct strain is represented as a row of gray dots, and the darkness of the dot is proportional to the log of the viral load of that strain at a time point. Time runs along the x-axis. Versions using three different definitions of antigenic distance were used: (a) Hamming, (b) xor, and (c) modified Manhattan distance.

of mutations that can accumulate within a single lineage is limited to approximately one mutation every two weeks, regardless of the virus's mutation rate.

Many pathogens produce periodic "bursts" of new strains in their hosts. Some examples (many of which are reviewed in Deitsch et al. (1997)) are *Plasmodium falciparum* (which causes malaria) (Roberts et al., 1992), *Trypanosoma brucei* (sleeping sickness) (Barry, 1986; Vickerman, 1989), and *Anaplasma marginale* (rickettsemia) (French et al., 1999). Although some of these pathogens are controlled by the B cell response, the primary mechanisms that create the viral load oscillations in my T cell model (i.e., cross-reactivity and a delayed immune response) also apply to other forms of immune response. This phenomenon has been modeled by others (e.g., Nowak et al. (1995); Antia et al. (1996); Haraguchi and Sasaki (1997); Recker et al. (2004)). These models include separate "strain-specific" responses that can eliminate only a particular

antigenic strain and "cross-reactive" responses that are effective against all variants. In my model, these "responses" are not separate. The CTLs in my model eliminate infected cells at a rate based on the similarity between their receptors and the MHC–epitope complex. My CTL model also differs from previous work by including a programmed CTL response. The programmed response causes the frequency of oscillations in viral load to be determined solely by CTL kinetics, not viral kinetics. The result is that the period between peaks in viral load are the same for any virus, regardless of its replication rate. Data in the literature to support or contradict this result is difficult to find because the pathogens that produce periodic peaks are generally parasites, not viruses, so they will not be controlled by CTLs.

6.4 Summary

I added immune exhaustion to the CTL model by incorporating an additional effector cell death term. With this extended model, I found that increasing the initial dose or the growth rate of a virus makes it more difficult for the immune system to eliminate it. Adding exhaustion also made the model's prolonged CTL responses more realistic, so I tested the effects of infecting the system with a mutating pathogen. The virus level oscillated, with peaks about every two weeks. These oscillations were caused by the regular creation of new viral strains and their elimination by CTLs. A higher mutation rate decreased the coherence of these oscillations but did not increase the rate at which a single strain could accumulate mutations over time. The model's response to a mutating pathogen highlighted another difference among the different distance metrics. The CTL response was highly cross-reactive when the Hamming distance metric was used, so the CTL responses to various mutant strains remained synchronized. For the other metrics were used, a single mutation could sometimes cause a new variant to temporarily evade the CTL response. The set of naïve CTLs responding to this variant would behave independently of the CTLs

responding to the parent strain, so the oscillations in total viral load become less well-defined over time.

Chapter 7

Conclusion

> *I had worked hard for nearly two years, for the sole purpose of infusing life into an inanimate body. For this I had deprived myself of rest and health. I had desired it with an ardour that far exceeded moderation; but now that I had finished, the beauty of the dream vanished, and breathless horror and disgust filled my heart…*
>
> —Mary Shelley, *Frankenstein*

I developed a computer model of the cytotoxic T lymphocyte (CTL) response to viral infection. Using this model, I explored the composition of the naïve CTL repertoire and the dynamics of the CTL response. The model revealed a mechanism by which the immune system can shape the naïve CTL repertoire to detect foreign peptides with efficiency using positive and negative selection. This mechanism is both biologically plausible and explains several somewhat surprising results in the literature. Further experiments are necessary to refine the hypothesis. If the model's results prove to be correct, then the model could be used to develop "thymic vaccination" therapies, in which peptides are introduced directly into the thymus to shape the immune cell repertoire (Fridkis-Hareli et al., 2004).

The model also replicates the dynamics of the CTL response to viral infection under a

wide variety of conditions, including natural infection, vaccination with a non-replicating antigen, administration of antibiotics, and the injection of massive amounts of additional CTLs. The typical approach to immunological modeling has been to create a minimal model for each of these scenarios, which makes the models easier to construct but restricted in scope. By creating a single model that accommodates many different immunological phenomena, one can use it to test new *combinations* of vaccination and other treatment strategies for preventing or controlling viral infection.

I have demonstrated that modeling can be used to enhance our understanding of immunology in different ways. Some of the results described in this work took advantage of the ease with which one can perform experiments with a computer model. For example, Section 5.1's analysis of the CTL repertoire before and after thymic selection would be extremely expensive and time-consuming to perform in the laboratory. Computer models can be used to run experiments before more costly animal tests are used. I performed other experiments in an attempt to explain known CTL behavior. Although the model contains a simple representation of the CTL life cycle, the magnitude and composition of the simulated responses reported in Sections 5.2 and 5.3 resemble those observed in mouse experiments. Thus, the elements of CTL behavior included in the model could be the primary factors governing short-term responses. In Chapter 6, I use the model to explore immunological phenomena that are not well-understood. Because models can be easily modified, they provide convenient vehicles for hypothesis testing.

A major task in constructing the model was the representation and implementation of CTL–infected cell binding. Because it was infeasible to implement a simulation of the molecular binding interactions between a TCR and an MHC–peptide complex, I used a highly abstract digit string representation for the TCR and the complex and a string distance rule to define the binding strength between them. The distance calculation had to be computationally efficient but still capture some aspect of the binding interactions. Because there is no single "correct" rule for this purpose, I implemented three different ones and

ran the model using each of them. When there is only one epitope in the system, then the choice of rule is less important—all three distance rules that I used produced similar affinity distributions. In experiments involving only a single epitope, I arbitrarily chose to use the xor metric, which was the most efficiently implemented. When studying the foreign peptide coverage of a CTL repertoire or simulating exposure to mutating antigens or more than one kind of pathogen, the choice of distance metric becomes important; the metric defines how cross-reactive responses behave in the model, which play a major role in these situations. I assumed that a result produced by all three versions of the model was robust to the metric's definition. If an effect only occurred using one of the metrics, then I had less confidence in the result. In such cases, one needs to consider why an effect depends on the definition of the metric and what properties do the molecular binding events that occur in the immune system share with the metric. Each metric probably has unique characteristics that reflect some of the properties of the "real" binding events that occur in the body, so each could be valid in different situations.

Model-building requires one to make many simplifications, assumptions, and compromises. By necessity, a model is a simplified representation of the system of interest. The simplification process demands that many aspects of the system be ignored or drastically reduced. Some of the decisions I made while implementing the model were based computational efficiency rather than biological fidelity. For example, because little is known about how the birth and death rates of effector CTLs change during the course of a response, I assumed that they were constant. Such assumptions made both the implementation and calibration of the model easier. As more quantitative measurements are made of the CTL response to antigen, the model should be extended to include them. I chose to adapt the scope of my model to match data available in the literature. However, the literature grows each day, and our ability to observe immune cells is improving so rapidly that many parts of the model will require revision within a few years. In particular, quantifying the effects of interactions among different kinds of immune cells, such as the innate immune system and helper T cells, will provide modelers with new opportunities. These

interactions will advance the state of immunological modeling immensely. Many immune cell types are involved in responses, and restricting a model to a single type severely limits its applicability. Therefore, model-building should be a dialog between the modeler and the disciplinary scientist. I have used the published results of immunologists to construct a model, and I hope that immunologists will be able to use this model to guide future experiments.

Appendix A

Calibrating the distance metrics

There are three versions of the CTL model, each using a different definition of string distance: Hamming, xor, and modified Manhattan. This appendix defines these metrics and summarizes their calibration in the context of the model using the procedure described in Section 4.2. Each of the following three sections consists of six parts:

1. *A formal definition of one of the distance metrics.*

2. *The alphabet size and string lengths for the peptides and TCRs.* The alphabet size and string lengths must be determined for the receptors and ligands in the CTL model. One constraint is that the ratio of the length of a TCR's MHC-binding portion to its peptide-binding portion must be 2:3 (Detours et al., 1999). Another constraint is that the number of possible TCR strings that the model can create should be at least 10^{15}, which is the estimated number of different TCRs that a human can generate (Davis and Bjorkman, 1988). The number of possible TCR strings for strings of length L using an alphabet of size k is k^L.

3. *The thymic selection thresholds.* Using the definition of the metric, the string lengths, and the alphabet size, I apply the algorithm described in Detours et al.

(1999) to find the distance distribution of random pre-selection TCRs from their selecting peptides. This distribution is used to compute the positive and negative thymic selection thresholds according to the procedure described in Section 4.2.1. The positive and negative selection thresholds are found to satisfy observed constraints of thymic selection in murine systems.

4. *The size of the pre-selection and naïve repertoires and the average number of responding clones per epitope.* Because the CTL model is calibrated using data from mouse experiments, mouse estimates are used to determine the size of the naïve repertoire and the number of naïve clones per epitope in the model. $10^6 - 10^7$ is the estimated number of clones in a mouse's naïve repertoire (Pannetier et al., 1993; Casrouge et al., 2000), and the number of responding naïve clones per epitope has been estimated to be 10–20 in mice (Blattman et al., 2002). Both of these values depend on the size of the pre-selection repertoire. The number of clones in the naïve repertoire size is the number in the pre-selection repertoire times the fraction of clones that survive selection (computed in part 3). The *foreign peptide response frequency*, which is the fraction of naïve clones that respond to a random peptide presented on self MHC, is computed using a procedure from Detours et al. (1999). The average number of clones that respond to an epitope is the foreign peptide response frequency, which has been experimentally observed to be between 10^{-6} and 10^{-4} (Stockinger et al., 1980; Zinkernagel, 1996), multiplied by the naïve repertoire size.

5. *The distribution of distances between two random strings.* This distribution is used by the lazy evaluation procedure described in Section 4.2.2. To lazily generate the CTLs for each MHC–epitope complex, the simulation generates the appropriate number of pre-selection TCRs that are at each distance from 0 to one less than the cross-reactive cutoff. To do this, I first determine the distribution of distances from random strings to a reference string. This distribution gives the proportion of

pre-selection TCRs that are at each distance from the MHC–epitope complex. One can compute the expected number of clones at each distance by multiplying this distribution by the total number of clones in the full pre-selection repertoire, which is about 10^8. During a simulation, to compute the actual number of clones at each distance d, a random number is drawn from the binomial distribution, using the size of the full pre-selection repertoire and the proportion of clones at d as parameters.

6. *An algorithm for generating a random string at the desired distance from a given string.* After the number of strings to generate at distance d from an MHC–epitope complex string is determined, the TCR strings are generated. These new strings form the pre-selection repertoire for a particular MHC–epitope pair.

In the sections that follow, I use the following notation: strings are sequences of L digits, digits are drawn from an alphabet of size k, U is the universe of possible strings, $|U|$ is the number of different strings in U, $I(x,y)$ is the distance between two digits x and y, $D(a,b)$ is the distance between two strings a and b, and $\Pr\{X = z\}$ is the probability that random variable X is equal to z.

A.1 Calibrating Hamming distance

1. *A formal definition of Hamming distance.* The Hamming distance between two strings is the number of positions in which they differ (Hamming, 1950). The Hamming distance I_H between two digits x and y is:

$$I_H(x,y) = \begin{cases} 1 & \text{if } x = y \\ 0 & \text{if } x \neq y \end{cases} \tag{A.1}$$

The Hamming distance D_H between two strings a and b is

$$D_H(a,b) = \sum_i I_H(a_i, b_i) \tag{A.2}$$

where a_i and b_i are the ith digits of the strings a and b.

2. *The alphabet size and string lengths for the peptides and TCRs.* Hamming distance is "coarse" because the distance between two digits is a boolean, rather than scalar, value (Equation A.1). The distance between two strings can only take values between 0 and L, the length of the strings. If one chose to use small string lengths to represent the peptide and MHC strings (such as 6 digits and 4 digits as suggested in Detours et al. (1999)), it would be difficult to find positive and negative selection thresholds such that 1%–3% of pre-selection clones survive thymic selection. Therefore, the string lengths must be longer. Long strings L can lead to an unreasonably large universe of TCRs, which is k^L where k is the alphabet size. Therefore, I chose a small alphabet size to reduce the number of possible TCRs. For reasons discussed in Kanerva (1988) and Smith et al. (1997), 2 might not be a suitable alphabet size, so I set the alphabet size to be 3. I test many combinations of MHC length and peptide length such that their ratio is 4:6 to match the values in Detours et al. (1999).

3. *The thymic selection thresholds.* The distribution of distances between a TCR and its selecting MHC–self peptide complex can be derived using Equation A.2 and the procedure in Detours et al. (1999). I used this distribution to find string lengths for the MHC- and peptide-binding portions of the TCR that satisfy the 2:3 length ratio, contain an appropriate-sized selection window of about 1–3% of pre-selection TCRs, and are of moderate length. I found that an MHC length of 32 and a peptide length of 48 yielded several appropriately sized selection windows; the distances of 34, 35, or 36 would all be plausible thymic selection windows (Figure A.1). For all of these candidate windows, the positive and negative selection thresholds are equal, so only TCRs that are exactly 34, 35, or 36 away from the nearest MHC–self peptide complex survive thymic selection. I chose to set the positive and negative selection thresholds to clones at distance 34 for the Hamming distance version of the model. About 3.96% of clones survive thymic selection.

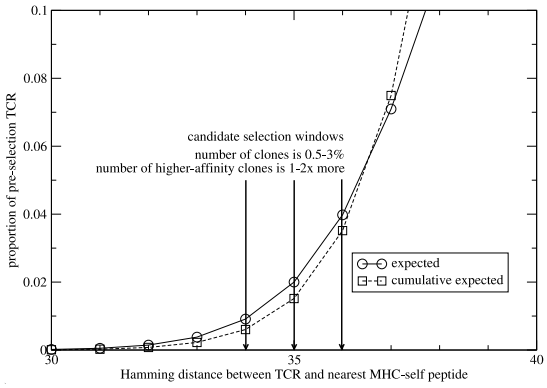


Figure A.1: Setting the thymic selection window using Hamming distance. The thymic selection window should cover about 1-3% of the possible strings, so the strings at distance 34, 35, or 36 satisfy this constraint (indicated by the arrows). About 1–2 times more strings should be of higher affinity than those in the window, so the cumulative distribution (in this case, the sum of the number of strings up to but not including the current distance) should be about 1–2 times larger than the number of strings in the selection window. The strings at distance 34, 35, or 36 each satisfy this constraint.

4. *The size of the pre-selection and naïve repertoires and the average number of responding clones per epitope.* When the size of the pre-selection repertoire is set to 8×10^7 clones, the naïve repertoire size is 3.17×10^6, which agrees with observations in mice. The expected number of responding clones per epitope is simply the naïve repertoire size multiplied by the foreign peptide response frequency. I calculated the foreign peptide response frequency to be 8.40×10^{-6}, which falls within the range observed in laboratory experiments. Using these values, the number of responding clones per epitope is about 27.

5. *The distribution of distances between two random strings.* The lazy evaluation algorithm requires that the expected number of strings that are a given distance away from a reference string be computed. In Kanerva (1988), it was proven that the proportion of all strings that are distance d away from a reference string is defined by

the binomial:

$$\Pr\{D_H(x,y)=d\} = \mathcal{B}(L,\frac{k-1}{k},d) \quad (A.3)$$

$$= \binom{L}{d}\left(\frac{k-1}{k}\right)^d\left(1-\frac{k-1}{k}\right)^{L-d} \quad (A.4)$$

where $\Pr\{D_H(x,y)=d\}$ is the probability that random strings x and y are exactly Hamming distance d apart, k is the alphabet size, and L is the length of the string.

6. *An algorithm for generating a random string at the desired distance from a given string.* I generate random strings at Hamming distance d from a reference string by randomly choosing d digits to differ from the reference string. These digits are set to random digits that are not equal to the other string's, while the remaining digits are copied from the reference string.

A.2 Calibrating xor distance

1. *A formal definition of xor distance.* The xor distance, D_{xor}, is the sum of the bitwise xors between the corresponding digits of two strings (Detours et al., 1999). For two one-bit numbers, the xor operation ($\oplus$) is defined to be 1 if the numbers are different and 0 if they are the same. This operation is extended to multi-bit numbers by decomposing each number x into a sum of bits $x_0 \ldots x_i$, each multiplied by 2 raised to a power: $x = \sum_i x_i 2^i$. The xor distance between two digits, I_{xor}, is:

$$I_{xor}(x,y) = x \oplus y \quad (A.5)$$

$$= \sum_i (x_i \oplus y_i) 2^i \quad (A.6)$$

where x_i and y_i are the ith bits of the digits x and y. The xor distance D_{xor} between two strings is

$$D_{xor}(a,b) = \sum_i I_{xor}(a_i,b_i) \quad (A.7)$$

where a_i and b_i are the ith digits of the strings a and b. For example, the xor distance between 3 and 5 is 6 because 3 can be decomposed into the sum $0\times2^2+1\times2^1+1\times2^0$ and 5 is $1\times2^2+0\times2^1+1\times2^0$, and the sum of the xors of the coefficients of the powers of 2 is $1\times2^2+1\times2^1+0\times2^0$, which is 6:

$$\begin{array}{rl} & 3{=}0\times2^2{+}1\times2^1{+}1\times2^0 \\ + & 5{=}1\times2^2{+}0\times2^1{+}1\times2^0 \\ \hline & 6{=}1\times2^2{+}1\times2^1{+}0\times2^0 \end{array}$$

2. *The alphabet size and string lengths for the peptides and TCRs.* I use values derived in Detours et al. (1999) for the lengths of the MHC and peptide strings: 4 digits for the MHC strings and 6 for the peptides. However, I reduced the alphabet size from 256 to 128 for computational efficiency and to reduce the size of the universe of possible TCRs from 1.2×10^{24} to 1.2×10^{21}.

3. *The thymic selection thresholds.* The expected distribution of distances from a random TCR to its selecting MHC–self peptide complex is plotted in Figure A.2. This distribution was verified by generating 3 random MHC strings and 30,000 random self peptide strings then computing the distance between 100,000 random TCR strings and the nearest MHC–self peptide complex string. The expected and observed distributions agree except in the low-affinity (large distance) tail, which are eliminated by positive selection (Figure A.2). The thymic selection window consists of strings at distances from 140 to 149 from the selecting MHC–self peptide complex. This range covers 0.807% of the random TCR strings, and 61% of the TCRs that survive positive selection are killed by negative selection.

4. *The size of the pre-selection and naïve repertoires and the average number of responding clones per epitope.* I set the size of the pre-selection repertoire to 2.5×10^8 clones, and after thymic selection the repertoire is reduced to 2×10^6. The foreign peptide response frequency is 1.27×10^{-5}, so about 25 naïve clones respond to each

epitope. Note that the number of pre-selection clones is higher than that used in the Hamming version (Section A.1) because a smaller fraction of clones happens to survive selection in the xor version.

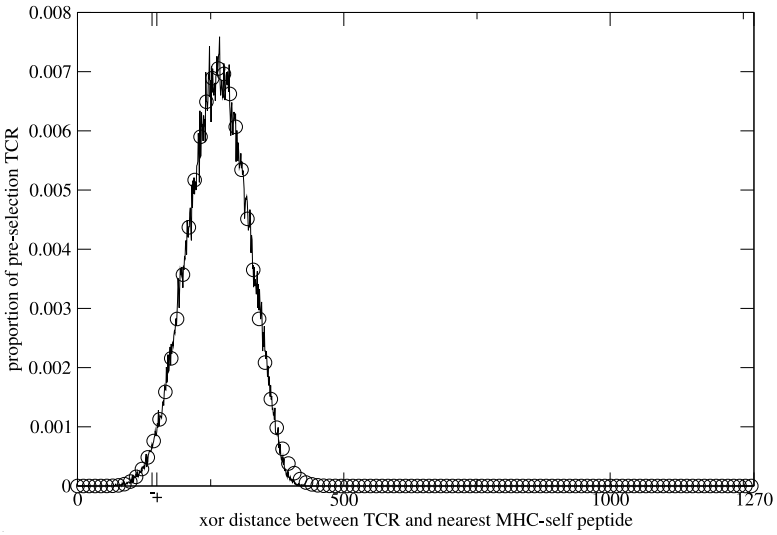


Figure A.2: The distribution of xor distances between a pre-selection TCR and the nearest MHC–self complex. The negative selection threshold is 140 and the positive is 149 (indicated on the x axis by the "-" and "+"). This results in a window size containing 0.807% of all possible TCRs, with about 1.29% of pre-selection clones eliminated by negative selection. The plot shows the expected results (○) and empirical results from generating 100,000 random TCR strings (solid line). The computations were performed using 3 MHC types, 10000 self peptides per MHC type, MHC length of 4 digits, peptide length of 6 digits, and an alphabet size of 128.

5. *The distribution of distances between two random strings.* For the lazy evaluation algorithm, I compute the distribution of distances between two random strings. The distance between two random strings is the sum of the distances between their corresponding digits, and the distribution of the sum of independent random variables is the convolution of their individual distributions. The probability distribution of distances between two random digits, $\Pr\{I_{xor}(x,y) = d\}$, is uniform for $d = 0 \ldots k-1$ (where k is the alphabet size) and zero elsewhere. Therefore, the probability distri-

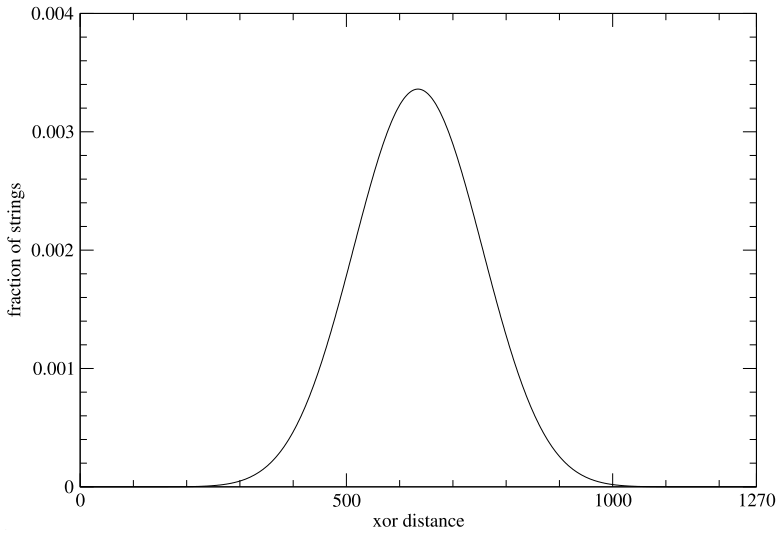


Figure A.3: The probability distribution of xor distances between two random strings of length 10 using an alphabet size of 128.

bution of xor distances between two strings is:

$$\Pr\{D_{xor}(a,b) = d\} = \Pr\{I_{xor}\}^n \tag{A.8}$$

where $\Pr\{I_{xor}\}^n$ is the probability distribution $\Pr\{I_{xor}\}$ convolved with itself $n-1$ times. An example of this distribution using the string parameters used by the CTL model (string length of 10 and alphabet size of 128) is plotted in Figure A.3.

6. *An algorithm for generating a random string at the desired distance from a given string.* To generate strings at a given distance from a reference string, I use the bit composition sets algorithm, described in Detours et al. (2001). This technique is computationally efficient but requires a large amount of memory.

A.3 Calibrating a modified Manhattan distance

1. *A formal definition of modified Manhattan distance (L_1').* I use a modified version of the first-order Minkowski metric, L_1. The standard first-order Minkowski distance is the sum of the absolute values of the differences between the digits of two strings:

$$D_{L1}(a,b) = \sum_i |a_i - b_i| \tag{A.9}$$

In two dimensions, it can represent the number of blocks one needs to travel to go between two points in a city if one must travel along a grid of streets. Therefore, L_1 is also known as Manhattan distance, city-block distance, and taxicab distance (Krause, 1987).

I use a modified Manhattan metric, which I denote L_1', in which the dimensions have cyclic boundaries so that the space "wraps around" (Figure A.4). The distance between two digits is:

$$I_{L1'}(x,y) = \begin{cases} |x-y| & \text{if } |x-y| \leq k/2 \\ k-|x-y| & \text{otherwise} \end{cases} \tag{A.10}$$

The distance between two strings is the sum of the distances between their corresponding digits:

$$D_{L1'}(a,b) = \sum_i I_{L1'}(a_i,b_i) \tag{A.11}$$

In standard L_1 space, the positions near the edges of the space have truncated neighborhoods. The term *neighborhood* is the set of strings that are at or within a given distance of a reference string. For example, on a line, a point at position 2 has two neighbors that are distance 1 away (at 1 and 3), while a point at 0 has only one neighbor that is distance 1 away (at 1). Using L_1' space with cyclic boundaries, the point at 0 has two neighbors at distance 1: the points at 1 and n, where n is the maximum

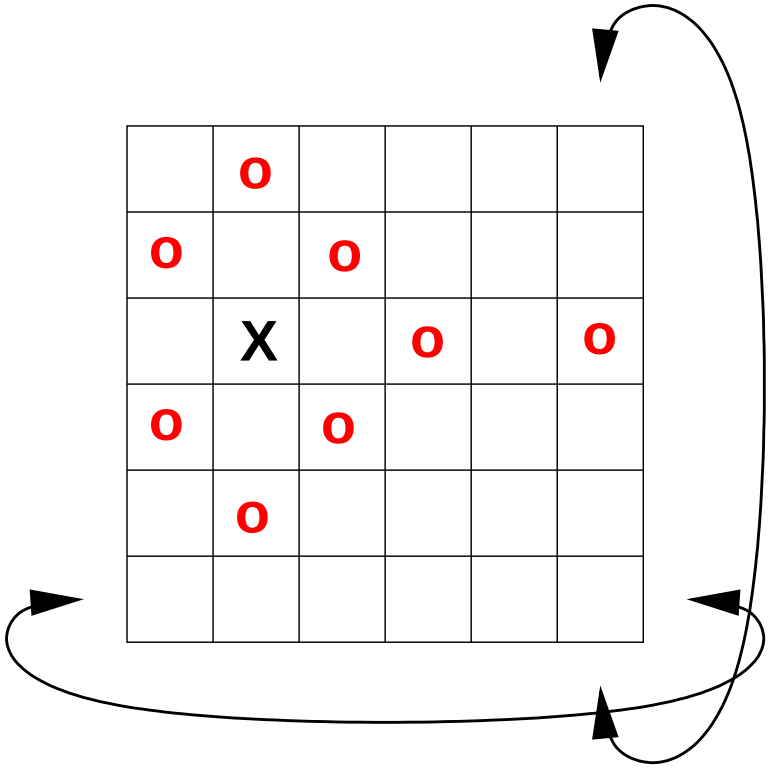


Figure A.4: The modified Manhattan distance (L_1') in 2 dimensions. It is the same as the standard L_1 distance except that boundaries are cyclic so the space "wraps around." In this figure, all o's are distance 2 away from the x.

value on the line. Thus, the neighborhoods of all points in L_1' have the same size and shape.

2. *The alphabet size and string lengths for the peptides and TCRs.* For the modified Manhattan distance L_1', I use the lengths of the MHC and peptide strings from Detours et al. (1999): 4 digits for the MHC strings and 6 for the peptides. I chose an alphabet of size 32, so the number of possible TCRs is $32^{10} = 1.13 \times 10^{15}$.

3. *The thymic selection thresholds.* Using the derivation from Detours et al. (1999), I found that the thymic selection window consists of the clones at distance 19 from their selecting peptides, resulting in 0.778% of the pre-selection repertoire surviv-

ing thymic selection (Figure A.5). About 70% of the clones that survive positive selection are killed by negative selection.

4. *The size of the pre-selection and naïve repertoires and the average number of responding clones per epitope.* I set the size of the pre-selection repertoire to 2.5×10^8 clones, and after thymic selection the repertoire is reduced to 2×10^6. The foreign peptide response frequency is 1.43×10^{-5}, resulting in about 28 responding clones per epitope.

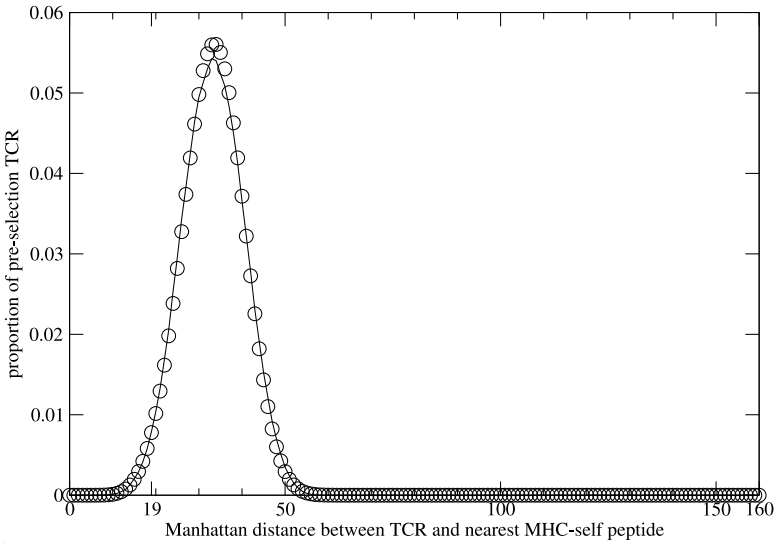


Figure A.5: The distribution of modified Manhattan distances between a random pre-selection TCR and the nearest MHC–self complex. The thymic selection window consists of clones at distance 19. This results in a window size of 0.778%, with about 1.78% of pre-selection clones eliminated by negative selection. The expected distribution is denoted by $\circ$ and a trial using 100,000 random TCR strings by the solid line. The computations were performed using 3 MHC types, 10000 self peptides per MHC, MHC length of 4 digits, peptide length of 6 digits, and an alphabet size of 32.

5. *The distribution of distances between two random strings.* As was the case for xor distance (Section A.2), the distribution of distances between two random strings is the convolution of the distribution of distances between two random digits. The

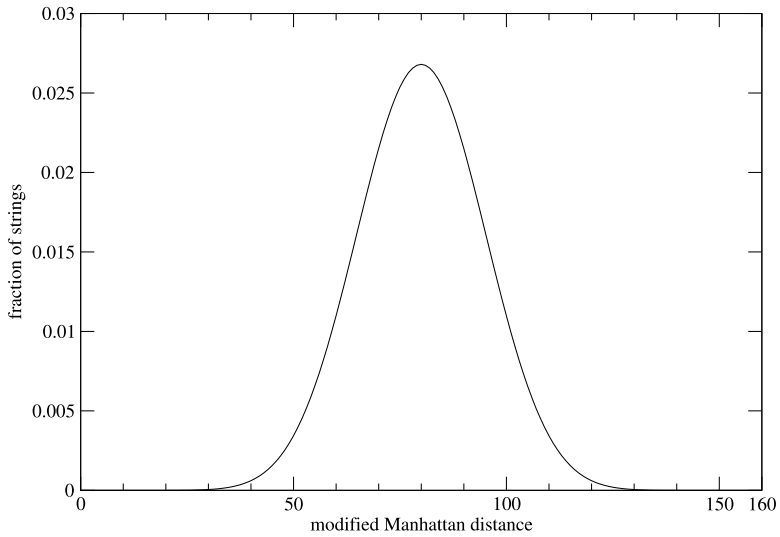


Figure A.6: The probability distribution of modified Manhattan distances (L_1') between two random strings of length 10 using an alphabet size of 32.

probability distribution of inter-digit distances for L_1' is:

$$\Pr\{I_{L1'}(x,y) = d\} = \begin{cases} 1/k & \text{if } d = 0 \text{ or } d = k/2 \\ 2/k & \text{if } d < k/2 \\ 0 & \text{otherwise} \end{cases} \tag{A.12}$$

where k is the alphabet size. Note that d can only equal $k/2$ when k is even. The distribution of inter-string distances is $\Pr\{I_{L_1'}\}^L$. An example of this distribution using the values used by the model (string length of 10 and alphabet size of 32) is plotted in Figure A.6.

6. *An algorithm for generating a random string at the desired distance from a given string.* I have implemented an algorithm to generate random strings at a given L_1' distance from a reference string. One part of this algorithm calculates probability distributions required to generate these strings. The recursive routine takes two parameters, n and d, which are the length of the string and the desired distance from the reference string. For all valid combinations of n and d, the routine computes and

stores the probability distribution of distances for the first digit of the string. For example, if half of all possible strings of length n and distance d from the origin begin with the digits 1 or $k-1$ (both distance 1 from the origin), then the distribution of distances for the first digit would be 0.5 for distance 1. For a string of length 1 (i.e., $n = 1$), the distribution that describes the number of strings whose first (and only) digit is distance d from the origin is:

$$|U|\Pr\{I_{L1'}(a_1,b_1)=r\} = \begin{cases} 1 & \text{if } d=0 \text{ and } r=0 \text{ or} \\ & \text{if } d=k/2 \text{ and } r=k/2 \\ 2 & \text{if } d<k/2 \text{ and } r=d \\ 0 & \text{otherwise} \end{cases} \tag{A.13}$$

From this distribution, the distributions for strings of length d can be computed recursively:

$$|U|\Pr\{I_{L1'}(a_d,b_d)=r\} = \begin{cases} |U|\Pr\{I_{L1'}(a_{d-1},b_{d-1})=d-r\} & \text{if } r=0 \text{ or} \\ & \text{if } r=k/2 \\ 2|U|\Pr\{I_{L1'}(a_{d-1},b_{d-1})=d-r\} & \text{otherwise} \end{cases} \tag{A.14}$$

The digit distributions are normalized to sum to 1 to obtain probability distributions.

The digit distance distribution is used to generate new random strings at distance d from a reference string. The first digit is chosen by randomly drawing a digit distance d_1 from the distribution $\Pr\{I_{L1'}(a_L,b_L)=r\}$. This random value is converted to a digit by either adding or subtracting (with equal probability) it from the digit in the reference string. The distance d_2 for the second digit is drawn from the distribution $\Pr\{I_{L1'}(a_{L-1},b_{L-1})=d-d_1\}$. Again, this distance is either added to or subtracted from the second digit in the reference string. This procedure is repeated to compute the remaining digits of the new string.

Appendix B

Alternative biological assumptions

The model presented in this work is not intended to be comprehensive. My intent was to create a computationally efficient model of the CTL response to infection that would elucidate issues of repertoire selection and the dynamics of the response. If one is interested in other aspects of CTLs, the model would need to be extended. In addition, some features of the T cell response are incompletely or possibly incorrectly understood, so I often had to choose among competing hypotheses. The purpose of this appendix is to list a few alternatives to the assumptions used by my CTL model.

In Section 3.2.2, the representation of affinity in the model has been simplified to exclude the phenomenon of *serial triggering*, in which a single MHC–peptide complex can stimulate multiple TCRs (Valitutti et al., 1995). An MHC–peptide complex that has a low dissociation rate with a CTL's TCRS stimulates only a small number of TCRs because each binding interaction takes a long time. A complex with a higher dissociation rate would have the opportunity to bind to more TCRs per unit time because after disengaging with one TCR it could bind to another. Thus both the affinity of the interactions and the dissociation rate determine the stimulation that a CTL receives. The model could be extended by adding dissociation rates to the CTL recruitment process.

In the same section, I assert that memory cells have the same antigenic stimulation requirements as naïve cells (Bachmann et al., 1999) and incorporate this fact into the model. Some studies have found that memory cells are more sensitive to antigen (Pihlgren et al., 1996). However, this effect is not consistently supported in the literature, and it is certainly not well quantified. Memory cells in the model can respond to lower levels of antigen without requiring lower stimulation thresholds because they are usually present in larger numbers than naïve cells and require less time to begin their initial rounds of proliferation.

Simultaneous responses to different epitopes expressed by the same infected cells are independent in the model (Section 3.2.3). This effect makes the modeling much easier and less computationally expensive, but it cannot be true. There must be competition for non-specific resources such as cytokines (Borghans et al., 1999) or even the surface of infected cells. However, it is not known how significant this competition is during the course of a typical response.

The model assumes that newly recruited effector cells have a constant death rate and divide for a fixed number of cycles before they stop replicating (Section 3.2.4), but the results from the CTL model described in Allan et al. (2004) indicate that the death rate for an effector cell should increase and the proliferation rate should decrease with each division. This is probably true, but current CFSE technology is not accurate enough to estimate the death and replication rates during the entire course of a response, so I have used the simpler assumption that these rates are constant.

The model uses a considerably simplified CTL life cycle (Section 3.2). CTLs perform many roles during an immune response, such as eliminating infected cells and producing cytokines, and the magnitude of infection might determine how many of these functions they adopt (Auphan-Anezin et al., 2003). These functions can also become progressively downregulated at the end of a response or during a chronic response (Fuller and Zajac, 2003; Wherry et al., 2003), as mentioned in Section 6.1. My model assumes an "all-or-nothing" activation of CTLs, while a more comprehensive model would allow for partial

activation.

All effector cells in the model have an equal probability of converting to memory throughout the course of the response (Section 3.2.5), but this assumption does not agree with recent findings that suggest that some effector cells are pre-determined to become memory cells. Some studies show that the subset of effectors that express the interleukin 7 receptor α-chain (IL-7Rα) early in the response become memory cells (Kaech et al., 2003; Huster et al., 2004). Expression of CD8αα receptors by effectors has also been found to correlate with conversion to memory (Madakamutil et al., 2004). However, it is not known if the effector cells that express IL-7Rα or CD8αα behave differently during the response than those that don't, so making this distinction in the model would not affect its behavior without this additional information.

Proliferation rates for naïve- and memory-derived (primary and secondary) effectors are the same in the model (Section 3.2.5), but in reality secondary effectors might have shorter division times (Rogers et al., 2000). Net population growth of secondary effectors happens to be higher in the model because they have lower death rates than primary effectors. Shorter division times would increase the secondary effector cell population growth even further.

The timing of the expression of epitopes probably plays a major role in immunodominance (van der Most et al., 2003). For example, CTLs responding to the LCMV epitopes expressed significantly earlier than others could dominate the response (Fuller-Pace and Southern, 1988). The CTL model does not consider timing effects, and all epitopes are immediately expressed upon the infection of healthy cells (Section 3.2.2). Only the expression levels of different epitopes differ in the model.

The model of thymic selection described in Section 4.2.1 makes many simplifying assumptions based on those used in Detours et al. (1999). Since that paper's publication, a handful of other thymic selection models have appeared in the literature.

van Den Berg et al. (2001) introduces antigen-presenting cells (APCs) that express a multitude of self peptides in the thymus. It also uses a more complex representation of CTL stimulation than that used in Detours et al. (1999). These features allow different self peptides to be presented at different levels, which could affect the naïve CTL repertoire (Laurie et al., 2004). The model in Faro et al. (2004) divides the thymus into two compartments: the cortex and the medulla. This division adds a new level of detail that could potentially be calibrated with experimental data. Another feature that could Finally, CTLs can modulate the number of TCRs they express in response to the environment in the thymus and the periphery (Grossman and Paul, 2001; Anderton and Wraith, 2002). This fact complicates the distinction between autoreactive and non-autoreactive CTLs, and models (such as van Den Berg and Rand (2004)) have been used to study these effects.

I assume that the cross-reactive cutoff is equal to the negative selection threshold used in thymic selection (Section 4.2.2), even though the the cross-reactive cutoff might be more stringent (Pircher et al., 1991). This would imply that a pre-selection cell is more sensitive to peptides and would respond to a larger range of peptides than a post-selection effector cell. Adding this effect to the model might affect a CTL model of autoimmunity, a phenomenon I do not address—uninfected cells in the model do not express self epitopes.

Finally, most of the data used in the model are based on murine experiments. Many modifications would be required to convert the model to accommodate human data. One could simply multiply many of the constants by 10,000 to represent the difference in mass between mice and humans, but one should probably use more realistic scaling laws, such as those described in Wiegel and Perelson (2004), to adapt the model to other organisms.

Glossary

affinity • The strength of the binding interaction between a single TCR and an MHC–peptide complex

anergy • A state of unresponsiveness in immune system cells.

antigen • An agent that stimulates an immune response, such as a virus or bacteria.

avidity • The total strength of the binding interactions between a CTL's TCRs and the MHC–peptide complexes expressed by a single target cell.

clone • A group of genetically identical cells derived from the same ancestor.

cross-reactivity • The ability of a single lymphocyte to respond to both an epitope and its variants.

cross-reactive cutoff • The maximum antigenic distance between a TCR's cognate peptide and another peptide recognized by the same TCR.

CTL • Cytotoxic T lypmohcyte. An immune cell that can eliminate infected cells by detecting abnormal peptides presented by MHC.

degeneracy • Peptide binding degeneracy is proportional to the size of the set of peptides to which a TCR can bind. Thus, a TCR that can bind to a large set of peptides has a high peptide-binding degeneracy.

effector T cell • A replicating T cell that eliminates infected cells.

epitope • The portion of an antigen that triggers an immune response. In the case of CTLs, a foreign peptide generated by virus or bacteria and presented by MHC.

foreign peptide response frequency • The fraction of clones that respond to a particular foreign peptide presented by MHC.

LCMV • Lymphocytic choriomeningitis virus. A non-cytopathic virus that infects mice. LCMV is often used to study the murine T cell response.

memory T cell • A quiescent T cell derived from an effector cell. These long-lived cells are created during T cell response and respond to antigen more quickly than naïve cells.

MHC • Major histocompatibility complex. MHC molecules present fragments of a cell's internal proteins on the cell's surface.

naïve T cell • A T cell that has survived thymic selection but has not yet been exposed to antigen.

negative selection • The phase of thymic selection that follows positive selection. Negative selection eliminates T cells that have high affinity for one or more MHC–self peptide complexes.

peptide • A short sequence of amino acids, a protein fragment.

positive selection • Positive selection eliminates T cells that have low affinity for all MHC–self peptide complexes expressed in the thymus.

programmed response • The pre-determined sequence of actions that a naïve T cell takes after antigenic stimulation. Even a brief exposure to antigen can cause a T cell to go through many rounds of division, adopt effector functions, then convert to memory cells.

repertoire • A set of T cells.

selecting peptide • During thymic selection, a CTL's selecting peptide is the MHC–self peptide complex to which the CTL has the highest affinity.

string distance metric • A function that takes two strings as input and returns a scalar value. The triangle inequality holds, so the distance between two strings A and B is less than or equal to the sum of the distances between A and a third string, C, and between B and C.

T cell • A type of lymphocyte.

TCR • T cell receptor. A CTL has many TCRs on its surface that bind to MHC–peptide complexes on other cells.

thymic selection • A process that takes place during the maturation of CTL precursors. CTLs that survive positive and negative selection exit the thymus and mature to become naïve cells.

References

Aichele, P., Brduscha-Riem, K., Oehen, S., Odermatt, B., Zinkernagel, R. M., Hengartner, H., and Pircher, H. Peptide antigen treatment of naive and virus-immune mice: antigen-specific tolerance versus immunopathology. *Immunity*, 6(5):519–29, 1997.

Alexander-Miller, M. A. Differential expansion and survival of high and low avidity cytotoxic T cell populations during the immune response to a viral infection. *Cell Immunol*, 201(1):58–62, 2000.

Alexander-Miller, M. A., Leggatt, G. R., and Berzofsky, J. A. Selective expansion of high- or low-avidity cytotoxic T lymphocytes and efficacy for adoptive immunotherapy. *Proc Natl Acad Sci U S A*, 93(9):4102–7, 1996.

Allan, M. J., Callard, R., Stark, J., and Yates, A. Comparing antigen-independent mechanisms of T cell regulation. *Journal of Theoretical Biology*, 228(1):81–95, 2004.

Altman, J. D., Moss, P. A., Goulder, P. J., Barouch, D. H., McHeyzer-Williams, M. G., Bell, J. I., McMichael, A. J., and Davis, M. M. Phenotypic analysis of antigen-specific T lymphocytes. *Science*, 274(5284):94–6, 1996.

Anderton, S. M., Radu, C. G., Lowrey, P. A., Ward, E. S., and Wraith, D. C. Negative selection during the peripheral immune response to antigen. *J Exp Med*, 193(1):1–11, 2001.

Anderton, S. M. and Wraith, D. C. Selection and fine-tuning of the autoimmune T-cell repertoire. *Nat Rev Immunol*, 2(7):487–98, 2002.

Antia, R., Bergstrom, C. T., Pilyugin, S. S., Kaech, S. M., and Ahmed, R. Models of CD8+ responses: 1. what is the antigen-independent proliferation program. *J Theor Biol*, 221(4):585–98, 2003.

Antia, R., Nowak, M. A., and Anderson, R. M. Antigenic variation and the within-host dynamics of parasites. *Proc Natl Acad Sci U S A*, 93(3):985–9, 1996.

Armstrong, C. and Lillie, R. Experimental lymphocytic choriomeningitis of monkeys and mice produced by a virus encountered in the studies of 1933 St Louis encephalitis epidemic. *Public Health Rep*, 49:1019–1027, 1934. LCMV.

Arstila, T. P., Casrouge, A., Baron, V., Even, J., Kanellopoulos, J., and Kourilsky, P. A direct estimate of the human $\alpha\beta$ T cell receptor diversity. *Science*, 286(5441):958–61, 1999.

Auphan-Anezin, N., Verdeil, G., and Schmitt-Verhulst, A. M. Distinct thresholds for CD8 T cell activation lead to functional heterogeneity: CD8 T cell priming can occur independently of cell division. *J Immunol*, 170(5):2442–8, 2003.

Bachmann, M. F., Barner, M., Viola, A., and Kopf, M. Distinct kinetics of cytokine production and cytolysis in effector and memory T cells after viral infection. *Eur J Immunol*, 29(1):291–9, 1999.

Bachmann, M. F., Speiser, D. E., and Ohashi, P. S. Functional maturation of an antiviral cytotoxic T-cell response. *J Virol*, 71(8):5764–8, 1997.

Badovinac, V. P., Porter, B. B., and Harty, J. T. Programmed contraction of $CD8^+$ T cells after infection. *Nat Immunol*, 3(7):619–26, 2002.

Bandeira, A. and Faro, J. Quantitative constraints on the scope of negative selection: robustness and weaknesses. *Trends Immunol*, 24(4):172–3, 2003.

Barber, D. L., Wherry, E. J., and Ahmed, R. Cutting edge: Rapid in vivo killing by memory CD8 T cells. *J Immunol*, 171(1):27–31, 2003.

Barchet, W., Oehen, S., Klenerman, P., Wodarz, D., Bocharov, G., Lloyd, A. L., Nowak, M. A., Hengartner, H., Zinkernagel, R. M., and Ehl, S. Direct quantitation of rapid elimination of viral antigen-positive lymphocytes by antiviral $CD8^+$ T cells in vivo. *Eur J Immunol*, 30(5):1356–63, 2000.

Barry, J. D. Antigenic variation during Trypanosoma vivax infections of different host species. *Parasitology*, 92:51–65, 1986.

Bell, G. I. Mathematical model of clonal selection and antibody production. *J Theor Biol*, 29(2):191–232, 1970.

Bernaschi, M. and Castiglione, F. Design and implementation of an immune system simulator. *Comput Biol Med*, 31(5):303–31, 2001.

Bevan, M. J. In thymic selection, peptide diversity gives and takes away. *Immunity*, 7(2):175–8, 1997.

Blackman, M., Kappler, J., and Marrack, P. The role of the T cell receptor in positive and negative selection of developing T cells. *Science*, 248(4961):1335–41, 1990.

Blattman, J. N., Antia, R., Sourdive, D. J., Wang, X., Kaech, S. M., Murali-Krishna, K., Altman, J. D., and Ahmed, R. Estimating the precursor frequency of naive antigen-specific CD8 T cells. *J Exp Med*, 195(5):657–64, 2002.

Blattman, J. N., Sourdive, D. J., Murali-Krishna, K., Ahmed, R., and Altman, J. D. Evolution of the T cell repertoire during primary, memory, and recall responses to viral infection. *J Immunol*, 165(11):6081–90, 2000.

Bocharov, G., Klenerman, P., and Ehl, S. Modelling the dynamics of LCMV infection in mice: II. compartmental structure and immunopathology. *J Theor Biol*, 221(3):349–78, 2003.

Bocharov, G. A. Modelling the dynamics of LCMV infection in mice: conventional and exhaustive CTL responses. *J Theor Biol*, 192(3):283–308, 1998.

Borghans, J. A., De Boer, R. J., and Segel, L. A. Extending the quasi-steady state approximation by changing variables. *Bull Math Biol*, 58(1):43–63, 1996.

Borghans, J. A., Taams, L. S., Wauben, M. H., and De Boer, R. J. Competition for antigenic sites during T cell proliferation: a mathematical interpretation of in vitro data. *Proc Natl Acad Sci U S A*, 96(19):10782–7, 1999.

Bousso, P., Casrouge, A., Altman, J. D., Haury, M., Kanellopoulos, J., Abastado, J. P., and Kourilsky, P. Individual variations in the murine T cell response to a specific peptide reflect variability in naive repertoires. *Immunity*, 9(2):169–78, 1998.

Bousso, P., Lemaitre, F., Bilsborough, J., and Kourilsky, P. Facing two T cell epitopes: a degree of randomness in the primary response is lost upon secondary immunization. *J Immunol*, 165(2):760–7, 2000.

Bousso, P., Levraud, J. P., Kourilsky, P., and Abastado, J. P. The composition of a primary T cell response is largely determined by the timing of recruitment of individual T cell clones. *J Exp Med*, 189(10):1591–600, 1999.

Busch, D. H., Pilip, I., and Pamer, E. G. Evolution of a complex T cell receptor repertoire during primary and recall bacterial infection. *J Exp Med*, 188(1):61–70, 1998a.

Busch, D. H., Pilip, I. M., Vijh, S., and Pamer, E. G. Coordinate regulation of complex T cell populations responding to bacterial infection. *Immunity*, 8(3):353–62, 1998b.

Butz, E. A. and Bevan, M. J. Massive expansion of antigen-specific $CD8^+$ T cells during an acute virus infection. *Immunity*, 8(2):167–75, 1998.

Byers, A. M., Kemball, C. C., Moser, J. M., and Lukacher, A. E. Cutting edge: Rapid in vivo CTL activity by polyoma virus-specific effector and memory $CD8^+$ T cells. *J Immunol*, 171(1):17–21, 2003.

Casrouge, A., Beaudoing, E., Dalle, S., Pannetier, C., Kanellopoulos, J., and Kourilsky, P. Size estimate of the $\alpha\beta$ TCR repertoire of naive mouse splenocytes. *J Immunol*, 164(11):5782–7, 2000.

Celada, F. and Seiden, P. E. A computer model of cellular interactions in the immune system. *Immunol Today*, 13(2):56–62, 1992.

Chao, D. L., Davenport, M. P., Forrest, S., and Perelson, A. S. Stochastic stage-structured modeling of the adaptive immune system. In *Proceedings of the IEEE Computer Society Bioinformatics Conference (CSB 2003)*, pages 124–131. IEEE Press, Los Alamitos, California, 2003.

Chao, D. L., Davenport, M. P., Forrest, S., and Perelson, A. S. Modelling the impact of antigen kinetics on T-cell activation and response. *Immunol Cell Biol*, 82(1):55–61, 2004a.

Chao, D. L., Davenport, M. P., Forrest, S., and Perelson, A. S. A stochastic model of cytotoxic T cell responses. *J Theor Biol*, 228(2):227–240, 2004b.

Claman, H. N., Chaperon, E. A., and Triplett, R. F. Thymus-marrow cell combinations. synergism in antibody production. *Proc Soc Exp Biol Med*, 122(4):1167–71, 1966.

Cohn, M., Langman, R. E., and Mata, J. J. A computerized model for the self-non-self discrimination at the level of the T(h) (Th genesis). I. the origin of 'primer' effector T(h) cells. *Int Immunol*, 14(10):1105–12, 2002.

Davenport, M. P., Fazou, C., McMichael, A. J., and Callan, M. F. Clonal selection, clonal senescence, and clonal succession: the evolution of the T cell response to infection with a persistent virus. *J Immunol*, 168(7):3309–17., 2002.

Davis, M. M. and Bjorkman, P. J. T-cell antigen receptor genes and T-cell recognition. *Nature*, 334(6181):395–402, 1988.

De Boer, R. J., Oprea, M., Antia, R., Murali-Krishna, K., Ahmed, R., and Perelson, A. S. Recruitment times, proliferation, and apoptosis rates during the $CD8^+$ T-cell response to lymphocytic choriomeningitis virus. *J Virol*, 75(22):10663–9, 2001.

De Boer, R. J. and Perelson, A. S. Towards a general function describing T cell proliferation. *J Theor Biol*, 175(4):567–76, 1995.

Deitsch, K. W., Moxon, E. R., and Wellems, T. E. Shared themes of antigenic variation and virulence in bacterial, protozoal, and fungal infections. *Microbiol Mol Biol Rev*, 61(3):281–93, 1997.

Derby, M., Alexander-Miller, M., Tse, R., and Berzofsky, J. High-avidity CTL exploit two complementary mechanisms to provide better protection against viral infection than low-avidity CTL. *J Immunol*, 166(3):1690–7, 2001.

Detours, V., Mehr, R., and Perelson, A. S. A quantitative theory of affinity-driven cell repertoire selection. *J Theor Biol*, 200(4):389–403, 1999.

Detours, V., Mehr, R., and Perelson, A. S. Deriving quantitative constraints on T cell selection from data on the mature T cell repertoire. *J Immunol*, 164(1):121–8, 2000.

Detours, V. and Perelson, A. S. Explaining high alloreactivity as a quantitative consequence of affinity-driven thymocyte selection. *Proc Natl Acad Sci U S A*, 96(9):5153–8, 1999.

Detours, V. and Perelson, A. S. The paradox of alloreactivity and self restriction: quantitative analysis and statistics. *Proc Natl Acad Sci U S A*, 97(15):8479–83, 2000.

Detours, V., Torney, D. C., and Perelson, A. S. A fast algorithm for simulating T cell immune response models with realistic size repertoires. 2001. (unpublished manuscript).

Dibrov, B. F., Livshits, M. A., and Volkenstein, M. V. Mathematical model of immune processes. *J Theor Biol*, 65(4):609–31, 1977.

Dutton, R. W., Bradley, L. M., and Swain, S. L. T cell memory. *Annu Rev Immunol*, 16:201–23, 1998.

Ehl, S., Klenerman, P., Zinkernagel, R. M., and Bocharov, G. The impact of variation in the number of $CD8^+$ T-cell precursors on the outcome of virus infection. *Cell Immunol*, 189(1):67–73, 1998.

Estcourt, M. J., Ramsay, A. J., Brooks, A., Thomson, S. A., Medveckzy, C. J., and Ramshaw, I. A. Prime-boost immunization generates a high frequency, high-avidity $CD8^+$ cytotoxic T lymphocyte population. *Int Immunol*, 14(1):31–7, 2002.

Farmer, J. D., Packard, N. H., and Perelson, A. S. The immune system, adaption and machine learning. *Physica D*, 22:187–204, 1986.

Faro, J., Velasco, S., González-Fernández, A., and Bandeira, A. The impact of thymic antigen diversity on the size of the selected T cell repertoire. *J Immunol*, 172(4):2247–55, 2004.

Fisher, R. A. and Yates, F. *Statistical Tables for Biological, Agricultural and Medical Research*. Oliver & Boyd, Edinburgh, 1938.

French, D. M., Brown, W. C., and Palmer, G. H. Emergence of Anaplasma marginale antigenic variants during persistent rickettsemia. *Infect Immun*, 67(11):5834–40, 1999.

Fridkis-Hareli, M., Reche, P. A., and Reinherz, E. L. Peptide variants of viral CTL epitopes mediate positive selection and emigration of Ag-specific thymocytes in vivo. *J Immunol*, 173(2):1140–50, 2004.

Fuller, M. J. and Zajac, A. J. Ablation of $CD8^+$ and $CD4^+$ T cell responses by high viral loads. *J Immunol*, 170(1):477–86, 2003.

Fuller-Pace, F. V. and Southern, P. J. Temporal analysis of transcription and replication during acute infection with lymphocytic choriomeningitis virus. *Virology*, 162(1):260–3, 1988.

Germain, R. N. The art of the probable: system control in the adaptive immune system. *Science*, 293(5528):240–5, 2001.

Gett, A. V. and Hodgkin, P. D. A cellular calculus for signal integration by T cells. *Nat Immunol*, 1(3):239–44, 2000.

Gibson, M. A. and Bruck, J. Efficient exact stochastic simulation of chemical systems with many species and many channels. *J Phys Chem*, 104:1876–1889, 2000.

Gillespie, D. T. Exact stochastic simulation of coupled chemical reactions. *J Phys Chem*, 81:2340–2361, 1977.

Goldrath, A. W. and Bevan, M. J. Selecting and maintaining a diverse T-cell repertoire. *Nature*, 402(6759):255–62, 1999.

Gray, P. M., Parks, G. D., and Alexander-Miller, M. A. High avidity $CD8^+$ T cells are the initial population elicited following viral infection of the respiratory tract. *J Immunol*, 170(1):174–81, 2003.

Grayson, J. M., Harrington, L. E., Lanier, J. G., Wherry, E. J., and Ahmed, R. Differential sensitivity of naive and memory $CD8^+$ T cells to apoptosis in vivo. *J Immunol*, 169(7):3760–70, 2002.

Grossman, Z. and Paul, W. E. Autoreactivity, dynamic tuning and selectivity. *Curr Opin Immunol*, 13(6):687–98, 2001.

Hamming, R. W. Error detecting and error correcting codes. *Bell System Technical Journal*, 29:147–160, 1950.

Haraguchi, Y. and Sasaki, A. Evolutionary pattern of intra-host pathogen antigenic drift: effect of cross-reactivity in immune response. *Philos Trans R Soc Lond B Biol Sci*, 352(1349):11–20, 1997.

Hemmer, B., Vergelli, M., Gran, B., Ling, N., Conlon, P., Pinilla, C., Houghten, R., McFarland, H. F., and Martin, R. Predictable TCR antigen recognition based on peptide scans leads to the identification of agonist ligands with no sequence homology. *J Immunol*, 160(8):3631–6, 1998.

Hildeman, D. A., Zhu, Y., Mitchell, T. C., Kappler, J., and Marrack, P. Molecular mechanisms of activated T cell death *in vivo*. *Curr Opin Immunol*, 14(3):354–9, 2002.

Ho, D. D., Neumann, A. U., Perelson, A. S., Chen, W., Leonard, J. M., and Markowitz, M. Rapid turnover of plasma virions and CD4 lymphocytes in HIV-1 infection. *Nature*, 373(6510):123–6, 1995.

Holland, J. H. Escaping brittleness: the possibilities of general-purpose learning algorithms applied to parallel rule-based systems. In R. S. Michalski, J. G. Carbonell, and T. M. Mitchell, editors, *Machine Learning, an artificial intelligence approach: Volume II*. Morgan Kaufmann, Los Alamos, CA, 1986.

Holler, P. D., Chlewicki, L. K., and Kranz, D. M. TCRs with high affinity for foreign pMHC show self-reactivity. *Nat Immunol*, 4(1):55–62, 2003.

Huseby, E. S., Crawford, F., White, J., Kappler, J., and Marrack, P. Negative selection imparts peptide specificity to the mature T cell repertoire. *Proc Natl Acad Sci U S A*, 100(20):11565–70. Epub 2003 Sep 22, 2003.

Huster, K. M., Busch, V., Schiemann, M., Linkemann, K., Kerksiek, K. M., Wagner, H., and Busch, D. H. Selective expression of IL-7 receptor on memory T cells identifies early CD40L-dependent generation of distinct $CD8^+$ memory T cell subsets. *Proc Natl Acad Sci U S A*, 101(15):5610–5, 2004.

Iezzi, G., Karjalainen, K., and Lanzavecchia, A. The duration of antigenic stimulation determines the fate of naive and effector T cells. *Immunity*, 8(1):89–95, 1998.

Jerne, N. K. Towards a network theory of the immune system. *Ann Immunol (Inst Pasteur)*, 125C:373–389, 1974.

Kaech, S. M. and Ahmed, R. Memory $CD8^+$ T cell differentiation: Initial antigen encounter triggers a developmental program in naïve cells. *Nat Immunol*, 2(5):415–22, 2001.

Kaech, S. M., Hemby, S., Kersh, E., and Ahmed, R. Molecular and functional profiling of memory CD8 T cell differentiation. *Cell*, 111:837–851, 2002.

Kaech, S. M., Tan, J. T., Wherry, E. J., Konieczny, B. T., Surh, C. D., and Ahmed, R. Selective expression of the interleukin 7 receptor identifies effector CD8 T cells that give rise to long-lived memory cells. *Nat Immunol*, 4(12):1191–8, 2003.

Kanerva, P. *Sparse Distributed Memory*. MIT Press, Cambridge, Massachusetts, 1988.

Kappler, J. W., Roehm, N., and Marrack, P. T cell tolerance by clonal elimination in the thymus. *Cell*, 49(2):273–80, 1987.

Kedzierska, K., Turner, S. J., and Doherty, P. C. Conserved T cell receptor usage in primary and recall responses to an immunodominant influenza virus nucleoprotein epitope. *Proc Natl Acad Sci U S A*, 101(14):4942–7, 2004.

Kersh, E. N., Kaech, S. M., Onami, T. M., Moran, M., Wherry, E. J., Miceli, M. C., and Ahmed, R. TCR signal transduction in antigen-specific memory CD8 T cells. *J Immunol*, 170(11):5455–63, 2003.

Kleinstein, S. H. and Seiden, P. E. Computer simulations: Simulating the immune system. *Computing in Science and Engineering*, 2(4):69–77, 2000.

Kleinstein, S. H. and Singh, J. P. Toward quantitative simulation of germinal center dynamics: biological and modeling insights from experimental validation. *J Theor Biol*, 211(3):253–75, 2001.

Kohler, B., Puzone, R., Seiden, P. E., and Celada, F. A systematic approach to vaccine complexity using an automaton model of the cellular and humoral immune system. *Vaccine*, 19(7–8):862–76, 2000.

Kraj, P., Pacholczyk, R., and Ignatowicz, L. αβTCRs differ in the degree of their specificity for the positively selecting MHC/peptide ligand. *J Immunol*, 166(4):2251–9, 2001.

Krause, E. F. *Taxicab Geometry: An Adventure in Non-Euclidean Geometry*. Dover Publications, New York, 1987.

Langman, R. E., Mata, J. J., and Cohn, M. A computerized model for the self-non-self discrimination at the level of the T(h) (Th genesis). II. the behavior of the system upon encounter with non-self antigens. *Int Immunol*, 15(5):593–609, 2003.

Laurie, K. L., Gruta, L. N. L., Koch, N., van Driel, I. R., and Gleeson, P. A. Thymic expression of a gastritogenic epitope results in positive selection of self-reactive pathogenic T cells. *J Immunol*, 172(10):5994–6002, 2004.

Lefkovitch, L. P. The study of population growth in organisms grouped by stages. *Biometrics*, 21:1–18, 1965.

Lefrançois, L., Marzo, A., and Williams, K. Sustained response initiation is required for T cell clonal expansion but not for effector or memory development in vivo. *J Immunol*, 171(6):2832–9, 2003.

Lehmann-Grube, F. Mechanism of recovery from acute virus infection. In H. Bauer, H.-D. Klenk, and C. Scholtissek, editors, *Modern Trends in Virology*, pages 49–64. Springer-Verlag, Berlin, 1988.

Lin, M. Y. and Welsh, R. M. Stability and diversity of T cell receptor repertoire usage during lymphocytic choriomeningitis virus infection of mice. *J Exp Med*, 188(11):1993–2005, 1998.

Lyons, A. B. and Parish, C. R. Determination of lymphocyte division by flow cytometry. *J Immunol Methods*, 171(1):131–7, 1994.

Madakamutil, L. T., Christen, U., Lena, C. J., Wang-Zhu, Y., Attinger, A., Sundarrajan, M., Ellmeier, W., von Herrath, M. G., Jensen, P., Littman, D. R., and Cheroutre, H. CD8$\alpha\alpha$-mediated survival and differentiation of CD8 memory T cell precursors. *Science*, 304(5670):590–3, 2004.

Manly, B. F. J. *Stage-structured populations: sampling, analysis, and simulation*. Chapman and Hall, London, 1990.

Marsh, S. G., Albert, E. D., Bodmer, W. F., Bontrop, R. E., Dupont, B., Erlich, H. A., Geraghty, D. E., Hansen, J. A., Mach, B., Mayr, W. R., Parham, P., Petersdorf, E. W., Sasazuki, T., Schreuder, G. M., Strominger, J. L., Svejgaard, A., and Terasaki, P. I. Nomenclature for factors of the HLA system, 2002. *Tissue Antigens*, 60(5):407–464, 2002.

Maryanski, J. L., Jongeneel, C. V., Bucher, P., Casanova, J. L., and Walker, P. R. Single-cell PCR analysis of TCR repertoires selected by antigen in vivo: a high magnitude CD8 response is comprised of very few clones. *Immunity*, 4(1):47–55, 1996.

Masopust, D., Kaech, S. M., Wherry, E. J., and Ahmed, R. The role of programming in memory T-cell development. *Curr Opin Immunol*, 16(2):217–25, 2004.

McNally, J. M., Zarozinski, C. C., Lin, M. Y., Brehm, M. A., Chen, H. D., and Welsh, R. M. Attrition of bystander $CD8^+$ T cells during virus-induced T-cell and interferon responses. *J Virol*, 75(13):5965–76, 2001.

Mercado, R., Vijh, S., Allen, S. E., Kerksiek, K., Pilip, I. M., and Pamer, E. G. Early programming of T cell populations responding to bacterial infection. *J Immunol*, 165(12):6833–9, 2000.

Miller, J. F. Immunological function of the thymus. *Lancet*, 2:748–9, 1961.

Miller, M. J., Wei, S. H., Parker, I., and Cahalan, M. D. Two-photon imaging of lymphocyte motility and antigen response in intact lymph node. *Science*, 296(5574):1869–73, 2002.

Morpurgo, D., Serentha, R., Seiden, P. E., and Celada, F. Modelling thymic functions in a cellular automaton. *Int Immunol*, 7(4):505–16, 1995.

Moskophidis, D., Lechner, F., Pircher, H., and Zinkernagel, R. M. Virus persistence in acutely infected immunocompetent mice by exhaustion of antiviral cytotoxic effector T cells. *Nature*, 362(6422):758–61, 1993.

Müller, V. and Bonhoeffer, S. Quantitative constraints on the scope of negative selection. *Trends Immunol*, 24(3):132–5, 2003.

Murali-Krishna, K., Altman, J. D., Suresh, M., Sourdive, D. J., Zajac, A. J., Miller, J. D., Slansky, J., and Ahmed, R. Counting antigen-specific $CD8^+$ T cells: a reevaluation of bystander activation during viral infection. *Immunity*, 8(2):177–87, 1998.

Murali-Krishna, K., Lau, L. L., Sambhara, S., Lemonnier, F., Altman, J., and Ahmed, R. Persistence of memory CD8 T cells in MHC class I-deficient mice. *Science*, 286(5443):1377–81, 1999.

Neumann, A. U., Lam, N. P., Dahari, H., Gretch, D. R., Wiley, T. E., Layden, T. J., and Perelson, A. S. Hepatitis C viral dynamics in vivo and the antiviral efficacy of interferon-α therapy. *Science*, 282(5386):103–7, 1998.

Nowak, M. A. and Bangham, C. R. Population dynamics of immune responses to persistent viruses. *Science*, 272(5258):74–9, 1996.

Nowak, M. A. and May, R. M. *Virus Dynamics*. Oxford University Press, Oxford, 2000.

Nowak, M. A., May, R. M., Phillips, R. E., Rowland-Jones, S., Lalloo, D. G., McAdam, S., Klenerman, P., Köppe, B., Sigmund, K., Bangham, C. R. M., and McMichael, A. J. Antigenic oscillations and shifting immunodominance in HIV-1 infections. *Nature*, 375(6532):606–11, 1995.

Ochsenbein, A. F., Sierro, S., Odermatt, B., Pericin, M., Karrer, U., Hermans, J., Hemmi, S., Hengartner, H., and Zinkernagel, R. M. Roles of tumour localization, second signals and cross priming in cytotoxic T-cell induction. *Nature*, 411(6841):1058–64, 2001.

Oehen, S. and Brduscha-Riem, K. Differentiation of naive CTL to effector and memory CTL: correlation of effector function with phenotype and cell division. *J Immunol*, 161(10):5338–46, 1998.

Opferman, J. T., Ober, B. T., and Ashton-Rickardt, P. G. Linear differentiation of cytotoxic effectors into memory T lymphocytes. *Science*, 283(5408):1745–8, 1999.

Pannetier, C., Cochet, M., Darche, S., Casrouge, A., Zoller, M., and Kourilsky, P. The sizes of the CDR3 hypervariable regions of the murine T-cell receptor beta chains vary as a function of the recombined germ-line segments. *Proc Natl Acad Sci U S A*, 90(9):4319–23, 1993.

Perelson, A. S. Modelling viral and immune system dynamics. *Nat Rev Immunol*, 2(1):28–36, 2002.

Perelson, A. S., Neumann, A. U., Markowitz, M., Leonard, J. M., and Ho, D. D. HIV-1 dynamics in vivo: virion clearance rate, infected cell life-span, and viral generation time. *Science*, 271(5255):1582–6, 1996.

Perelson, A. S. and Weisbuch, G. Modeling immune reactivity in secondary lymphoid organs. *Bull Math Biol*, 54(4):649–72, 1992.

Pihlgren, M., Dubois, P. M., Tomkowiak, M., Sjogren, T., and Marvel, J. Resting memory $CD8^+$ T cells are hyperreactive to antigenic challenge in vitro. *J Exp Med*, 184(6):2141–51, 1996.

Pircher, H., Rohrer, U. H., Moskophidis, D., Zinkernagel, R. M., and Hengartner, H. Lower receptor avidity required for thymic clonal deletion than for effector T-cell function. *Nature*, 351(6326):482–5, 1991.

Přikrylová, D., Jílek, M., and Waniewski, J. *Mathematical modeling of the immune response*. CRC Press, Boca Raton, Florida, 1992.

Recker, M., Nee, S., Bull, P. C., Kinyanjui, S., Marsh, K., Newbold, C., and Gupta, S. Transient cross-reactive immune responses can orchestrate antigenic variation in malaria. *Nature*, 429(6991):555–8, 2004.

Rees, W., Bender, J., Teague, T. K., Kedl, R. M., Crawford, F., Marrack, P., and Kappler, J. An inverse relationship between T cell receptor affinity and antigen dose during $CD4^+$ T cell responses *in vivo* and *in vitro*. *Proc Natl Acad Sci U S A*, 96(17):9781–6, 1999.

Reilly, D. L., Cooper, L. N., and Elbaum, C. A neural model for category learning. *Biological Cybernetics*, 45:35–41, 1982.

Roberts, D. J., Craig, A. G., Berendt, A. R., Pinches, R., Nash, G., Marsh, K., and Newbold, C. I. Rapid switching to multiple antigenic and adhesive phenotypes in malaria. *Nature*, 357(6380):689–92, 1992.

Rocha, B., Dautigny, N., and Pereira, P. Peripheral T lymphocytes: expansion potential and homeostatic regulation of pool sizes and $CD4^+/CD8^+$ ratios in vivo. *Eur J Immunol*, 19(5):905–11, 1989.

Rogers, P. R., Dubey, C., and Swain, S. L. Qualitative changes accompany memory T cell generation: faster, more effective responses at lower doses of antigen. *J Immunol*, 164(5):2338–46, 2000.

Salfi, R. A long-period random number generator with application to permutations. *COMPSTAT 1974: Proceedings in Computational Statistics*, pages 28–35, 1974.

Savage, P. A., Boniface, J. J., and Davis, M. M. A kinetic basis for T cell receptor repertoire selection during an immune response. *Immunity*, 10(4):485–92, 1999.

Schwartz, R. H. T cell anergy. *Annu Rev Immunol*, 21:305–34, 2003.

Seiden, P. E. and Celada, F. A model for simulating cognate recognition and response in the immune system. *J Theor Biol*, 158(3):329–57, 1992.

Selin, L. K., Vergilis, K., Welsh, R. M., and Nahill, S. R. Reduction of otherwise remarkably stable virus-specific cytotoxic T lymphocyte memory by heterologous viral infections. *J Exp Med*, 183(6):2489–99, 1996.

Shi, Y. F., Sahai, B. M., and Green, D. R. Cyclosporin A inhibits activation-induced cell death in T-cell hybridomas and thymocytes. *Nature*, 339(6226):625–6, 1989.

Shortman, K., Egerton, M., Spangrude, G. J., and Scollay, R. The generation and fate of thymocytes. *Semin Immunol*, 2(1):3–12, 1990.

Slifka, M. K., Blattman, J. N., Sourdive, D. J., Liu, F., Huffman, D. L., Wolfe, T., Hughes, A., Oldstone, M. B., Ahmed, R., and von Herrath, M. G. Preferential escape of subdominant $CD8^+$ T cells during negative selection results in an altered antiviral T cell hierarchy. *J Immunol*, 170(3):1231–9, 2003.

Smith, D. J. *The cross-reactive immune response: Analysis, Modeling, and Application to Vaccine Design*. Ph.D. thesis, University of New Mexico, 1997.

Smith, D. J., Forrest, S., Ackley, D. H., and Perelson, A. S. Using lazy evaluation to simulate realistic-size repertoires in models of the immune system. *Bull Math Biol*, 60(4):647–58, 1998.

Smith, D. J., Forrest, S., Ackley, D. H., and Perelson, A. S. Variable efficacy of repeated annual influenza vaccination. *Proc Natl Acad Sci U S A*, 96(24):14001–6, 1999.

Smith, D. J., Forrest, S., Hightower, R. R., and Perelson, A. S. Deriving shape space parameters from immunological data. *J Theor Biol*, 189(2):141–50, 1997.

Smith, J. A. and Martin, L. Do cells cycle? *Proc Natl Acad Sci U S A*, 70(4):1263–7, 1973.

Sourdive, D. J., Murali-Krishna, K., Altman, J. D., Zajac, A. J., Whitmire, J. K., Pannetier, C., Kourilsky, P., Evavold, B., Sette, A., and Ahmed, R. Conserved T cell receptor repertoire in primary and memory CD8 T cell responses to an acute viral infection. *J Exp Med*, 188(1):71–82, 1998.

Stewart, J. J., Agosto, H., Litwin, S., Welsh, J. D., Shlomchik, M., Weigert, M., and Seiden, P. E. A solution to the rheumatoid factor paradox: pathologic rheumatoid factors can be tolerized by competition with natural rheumatoid factors. *J Immunol*, 159(4):1728–38, 1997.

Stockinger, H., Pfizenmaier, K., Hardt, C., Rodt, H., Rollinghoff, M., and Wagner, H. H-2 restriction as a consequence of intentional priming: T cells of fully allogeneic chimeric mice as well as of normal mice respond to foreign antigens in the context of H-2 determinants not encountered on thymic epithelial cells. *Proc Natl Acad Sci U S A*, 77(12):7390–4, 1980.

Tough, D. F. and Sprent, J. Turnover of naive- and memory-phenotype T cells. *J Exp Med*, 179(4):1127–35, 1994.

Traub, E. A filtrable virus recovered from white mice. *Science*, 81:298–299, 1935. LCMV.

Usher, M. B. A matrix approach to the management of renewable resources, with special reference to selection forests. *Journal of Applied Ecology*, 3:355–367, 1966.

Valitutti, S., Muller, S., Cella, M., Padovan, E., and Lanzavecchia, A. Serial triggering of many T-cell receptors by a few peptide-MHC complexes. *Nature*, 375(6527):148–51, 1995.

van Den Berg, H. A. and Rand, D. A. Dynamics of T cell activation threshold tuning. *J Theor Biol*, 228(3):397–416, 2004.

van Den Berg, H. A., Rand, D. A., and Burroughs, N. J. A reliable and safe T cell repertoire based on low-affinity T cell receptors. *J Theor Biol*, 209(4):465–86, 2001.

van der Most, R. G., Murali-Krishna, K., Lanier, J. G., Wherry, E. J., Puglielli, M. T., Blattman, J. N., Sette, A., and Ahmed, R. Changing immunodominance patterns in antiviral CD8 T-cell responses after loss of epitope presentation or chronic antigenic stimulation. *Virology*, 315(1):93–102, 2003.

van Stipdonk, M. J. B., Lemmens, E. E., and Schoenberger, S. P. Naïve CTLs require a single brief period of antigenic stimulation for clonal expansion and differentiation. *Nat Immunol*, 2(5):423–9, 2001.

Veiga-Fernandes, H., Walter, U., Bourgeois, C., McLean, A., and Rocha, B. Response of naive and memory $CD8^+$ T cells to antigen stimulation in vivo. *Nat Immunol*, 1(1):47–53, 2000.

Vickerman, K. Trypanosome sociology and antigenic variation. *Parasitology*, 99 Suppl:S37–47, 1989.

Vijh, S., Pilip, I. M., and Pamer, E. G. Effect of antigen-processing efficiency on in vivo T cell response magnitudes. *J Immunol*, 160(8):3971–7, 1998.

Vijh, S., Pilip, I. M., and Pamer, E. G. Noncompetitive expansion of cytotoxic T lymphocytes specific for different antigens during bacterial infection. *Infect Immun*, 67(3):1303–9, 1999.

Walter, S., Herrgen, L., Schoor, O., Jung, G., Wernet, D., Buhring, H. J., Rammensee, H. G., and Stevanović, S. Cutting edge: Predetermined avidity of human CD8 T cells expanded on calibrated MHC/anti-CD28-coated microspheres. *J Immunol*, 171(10):4974–8, 2003.

Wei, X., Ghosh, S. K., Taylor, M. E., Johnson, V. A., Emini, E. A., Deutsch, P., Lifson, J. D., Bonhoeffer, S., Nowak, M. A., Hahn, B. H., Saag, M. S., and Shaw, G. M. Viral dynamics in human immunodeficiency virus type 1 infection. *Nature*, 373(6510):117–22, 1995.

Welsh, R. M. and Selin, L. K. No one is naive: the significance of heterologous T-cell immunity. *Nat Rev Immunol*, 2(6):417–26, 2002.

Wherry, E. J. and Ahmed, R. Memory CD8 T-cell differentiation during viral infection. *J Virol*, 78(11):5535–45, 2004.

Wherry, E. J., Blattman, J. N., Murali-Krishna, K., Van Der Most, R., and Ahmed, R. Viral persistence alters CD8 T-cell immunodominance and tissue distribution and results in distinct stages of functional impairment. *J Virol*, 77(8):4911–27, 2003.

Wherry, E. J., McElhaugh, M. J., and Eisenlohr, L. C. Generation of $CD8^+$ T cell memory in response to low, high, and excessive levels of epitope. *J Immunol*, 168(9):4455–61, 2002.

Wherry, E. J., Puorro, K. A., Porgador, A., and Eisenlohr, L. C. The induction of virus-specific CTL as a function of increasing epitope expression: responses rise steadily until excessively high levels of epitope are attained. *J Immunol*, 163(7):3735–45, 1999.

Wiegel, F. W. and Perelson, A. S. Some scaling principles for the immune system. *Immunol Cell Biol*, 82(2):127–31, 2004.

Zajac, A. J., Blattman, J. N., Murali-Krishna, K., Sourdive, D. J., Suresh, M., Altman, J. D., and Ahmed, R. Viral immune evasion due to persistence of activated T cells without effector function. *J Exp Med*, 188(12):2205–13, 1998.

Zinkernagel, R. M. Immunology taught by viruses. *Science*, 271(5246):173–8, 1996.

Zinkernagel, R. M. and Doherty, P. C. Restriction of in vitro T cell-mediated cytotoxicity in lymphocytic choriomeningitis within a syngeneic or semiallogeneic system. *Nature*, 248(450):701–2, 1974.