MULTI-SCALE MODELING OF ANTIGEN PRESENTATION WITH APPLICATIONS TO TUBERCULOSIS

by

Stewart T. Chang

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Bioinformatics)
in The University of Michigan
2007

Doctoral Committee:

Associate Professor Denise E. Kirschner, Co-Chair
Professor Jennifer J. Linderman, Co-Chair
Professor Daniel M. Burns
Associate Professor Malini Raghavan

To My Parents

ACKNOWLEDGMENTS

I would first like to thank my friends in Michigan and in other places. I cannot even begin to count how many times I depended on them and was glad to have them in my life.

I would like to thank the teachers I had growing up in Texas, particularly Mr. Brad Newton, Mrs. Sandra Morris, and Mrs. Colleen Webb. Their passion for teaching and zeal for subject matter and for life continue to inspire me.

I would also like to thank my advisors, Dr. Denise Kirschner and Dr. Jennifer Linderman, who intrepidly took me on and trained me here at Michigan. They allowed me the freedom to explore niches in the experimental literature where mathematical modeling could provide new insights, even when the territory was unfamiliar to all of us. In addition, I would like to thank the other members of my dissertation committee, Dr. Daniel Burns and Dr. Malini Raghavan, and other faculty members at Michigan with whom I have collaborated, particularly Dr. Debashis Ghosh and Dr. Noah Rosenberg, for their encouragement and many helpful discussions.

Finally I would like to thank my mother, my father, and my brother for seeing me off to school with a hot breakfast, taking me to the library on Saturday, and playing baseball with me in the backyard, respectively. And for a million other things.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABSTRACT

Antigen presentation is the process by which cells of the immune system display peptides from pathogens on their surface after binding the peptides to major histocompatibility complex (MHC) molecules. T helper cells recognize peptides from pathogens in this context then secrete cytokines that activate other cells, initiating an immune response. Antigen presentation is therefore a requisite for immunity to several pathogens including *Mycobacterium tuberculosis* (*Mtb*). To approach questions related to antigen presentation and disease, I represented antigen presentation at different scales using a series of mathematical and statistical models. At the molecular scale, I asked whether heterogeneity in peptide length affects binding to MHC class II, the class of MHC responsible for binding peptides from bacteria such as *Mtb*. By developing statistical models of peptide-MHC binding, I found that length has a nonlinear effect on binding affinity and that this information, or a more accurate representation of register shifting, could improve the accuracy of binding prediction. At the cellular scale, I asked why *Mtb* possesses multiple mechanisms to inhibit antigen presentation on the cell surface. My mathematical model shows that these mechanisms may be acting on different timescales and therefore complementary rather than merely redundant. Finally, at the multi-cellular level, I asked how polymorphisms in multiple genes related to antigen presentation might affect T cell response and susceptibility to infectious diseases such as tuberculosis. Using a multi-scale model representing both an antigen-presenting cell and T cell, I found that polymorphisms in two different genes may exert the same influence on the output, potentially canceling out their effects. Future work with these models may

include evaluation of candidate peptide-based vaccines to ensure high-affinity binding, T cell response, and broad efficacy in diverse populations.

CHAPTER 1

Introduction

Antigen presentation is the process by which proteins from pathogens are partially degraded and then displayed (i.e., presented) on the surfaces of cells in complex with major histocompatibility complex (MHC) molecules. Once bound to MHC molecules, peptides can be recognized by cognate T cells which then respond by either killing the original antigen-presenting cell (APC) or activating the APC and other cells. Proteins that elicit an immune response are known as antigens. While antigen presentation may appear to be dictated by events at the molecular and sub-cellular scales of the APC, events at other scales also affect the outcome (Fig. 1.1). For instance, once peptides have been bound by MHC molecules and trafficked to the surface of the APC, additional events are required by the T cell to result in a functional response, starting with the engagement of peptide-MHC (pMHC) complexes by T cell receptors (TCR) on the T cell surface. Activated TCRs then initiate a signaling cascade within the T cell, resulting in the elicitation of either cytotoxic molecules or activating cytokines including interleukin-2 (IL-2) and interferon-γ (IFN-γ). Therefore, at a minimum, a representation of the events of antigen presentation requires consideration of molecular and sub-cellular events occurring in APCs and T cells. In addition, the larger tissue- or organ system-scale context may also be important in determining the outcome of antigen presentation, as this

environment may enhance or interfere with the ability of APCs to interact with T cells (Fig. 1.1).

How best to represent events at these different scales using the tools of mathematical and statistical modeling and how then to apply these tools to the study of infectious diseases remain open questions. A vast amount of experimental research has been done, and continues to be done, to elucidate the steps involved in antigen presentation. At the same time, computational models of one or more steps in antigen presentation have been developed. I review previous efforts in both experimental and computational areas in this chapter and then describe my efforts to improve on computational models and apply them to one infectious disease, tuberculosis, in subsequent chapters.

1.1 Overview of Antigen Presentation

Two main pathways for antigen presentation exist, depending on the source of the antigen. All nucleated human (and mammalian) cells perform antigen presentation to some extent by expressing one type of MHC molecule, MHC class I, which binds peptides derived primarily from proteins in the cytoplasm (Yewdell 2007). However, some cells are also capable of presenting peptides from proteins found in the extracellular medium by expressing another type of MHC molecule, MHC class II (Trombetta and Mellman 2005). These specialized cells include macrophages, dendritic cells, and B cells, constituting the professional APC. [A recently discovered pathway for presenting lipid antigens, the CD1 pathway, will not be considered here. See Mahanty *et al.* (2003) for a review of this topic.]

For the most part the two pathways for antigen presentation remain distinct within the cell. Proteins found in the cytoplasm, including those produced by most viruses, are considered endogenous and degraded (i.e., processed) into shorter peptides by the main

protein turnover machinery of the cell, the proteasome. A subset of these peptides is transported into the endoplasmic reticulum (ER) by the transporter associated with antigen processing (TAP). Within the ER peptides of a limited range of lengths (8-10 amino acids) bind MHC class I molecules, and the resulting complexes are trafficked to the cell surface.

In contrast antigens from pathogens that do not reside in the cytoplasm, including most bacteria and parasites, are considered exogenous and taken up and processed in the endosomal pathway of the APC. Within the endosomal pathway cathepsin proteases become activated by the increasingly acidified environment and cleave proteins into shorter peptides. Peptides of various lengths (often greater than 9 amino acids) then bind MHC class II molecules later in the pathway in a specialized vacuole known as the MHC class II compartment (MIIC). Peptide-MHC class II complexes are then trafficked to the cell surface as in the case of MHC class I. In both cases the final stage is recognition of the pMHC complexes by TCRs found on the surfaces of $CD8^+$ cytotoxic T cells and $CD4^+$ helper T cells which are specific for MHC class I and MHC class II, respectively.

Although the two pathways are for the most part distinct, exceptions have been found. For instance, during cross-presentation, exogenous antigens gain access to MHC class I molecules within APCs and yield peptide-MHC class I complexes on the APC surface that stimulate $CD8^+$ T cells. Some aspects of cross-presentation remain controversial such as the degree to which it occurs and the instances in which it may be important to the immune response (see Rock and Shen 2005 for a recent review).

1.2 Biology of the Molecular Scale: Peptide-MHC Binding

One theme that arises from this cursory overview of the two antigen presentation pathways is the centrality of the peptide-MHC binding event. MHC class I and MHC class II molecules bind peptides in a similar manner, and this similarity can be traced to

similarities in the structures of the two molecules (Jones 1997). Both MHC class I and MHC class II molecules comprise heterodimers of polypeptides. In the case of MHC class I, a single transmembrane polypeptide (designated the α chain) is coupled to a smaller accessory protein (β_2 microglobulin), and the peptide-binding groove is formed between two domains of the transmembrane polypeptide (α_1 and α_2). In the case of MHC class II, two transmembrane polypeptides of similar size are coupled (designed α and β chains), and the peptide-binding groove is formed by overlap of the membrane-distal domains of the two polypeptides (α_1 and β_1).

In addition to overall structural similarities, the peptide-binding grooves of the two classes of MHC molecule are also similar (Jones 1997). In both cases, eight β-pleated sheets and two α-helices form the bottom and walls of the peptide-binding groove, respectively. Pockets within both grooves bind amino acids at select positions within the peptides which are similarly extended in an α-helical conformation in both cases. However, in the case of MHC class I, the termini of the peptide attach at both ends of the peptide-binding groove, restricting peptide length to between 8 and 10 amino acids. In contrast, the ends of the MHC class II peptide-binding groove are open, allowing peptides to be of a greater range of lengths.

1.2.1 Quantification of Peptide-MHC Affinity

While peptide-MHC binding was formerly viewed as a dichotomous event, either occurring or not, a more quantitatively continuous view has accompanied advances in the ability to measure affinity. In the most commonly used assay, the strength of binding is assessed by titrating reporter peptide-MHC complexes with increasing concentrations of the peptide of interest (Southwood *et al.* 1998). The concentration at which 50% of the reporter peptide is displaced then yields the 50% inhibitory concentration (IC_{50}) which can be shown to approximate the equilibrium dissociation constant (K_D) of the peptide of

interest with the MHC molecule. High-affinity binding, associated with smaller values of K_D and IC_{50}, is expected to yield more pMHC complexes on the APC surface and in turn facilitate T cell activation. 500 nM is commonly used as an upper limit for K_D values that result in binding (Sette *et al.* 1994). According to one survey, the majority of functional pMHC complexes have K_D values in the range of 10-100 nM (McFarland and Beeson 2002).

1.2.2 Genetic Polymorphism within the MHC

Because affinity is specific to each combination of peptide and MHC, variations in either peptide or MHC sequence may affect binding. In humans the genes encoding the MHC molecules, known as the human leukocyte antigens (HLA), are particularly variable. For example, among the three sets of genes encoding MHC class II molecules in humans, HLA-DR, -DP, and -DQ, 875 alleles have currently been identified (Robinson *et al.* 2003). In most cases, variation occurs in positions associated with peptide-binding or TCR-binding and can therefore be assumed to be functional. Recently attempts have been made to categorize alleles on the basis of their peptide-binding characteristics, thereby defining MHC supertypes and effectively reducing the number of alleles (Ou *et al.* 1998, Lund *et al.* 2004, Doytchinova and Flower 2005).

1.2.3 Peptide-MHC Affinity and Disease Susceptibility

The greater significance of peptide-MHC binding can also be discerned from the epidemiological literature. Various MHC alleles have been correlated with increased susceptibility to autoimmune and infectious diseases such as HLA-DRB1*1501 (a variant of the MHC class II β polypeptide) with tuberculosis (Vukmanovic *et al.* 2003). Other diseases for which MHC alleles have been found to affect susceptibility include type I diabetes, rheumatoid arthritis, and malaria. The mechanism behind these associations has

not yet been elucidated, though several hypotheses exist (Vukmanovic *et al.* 2003, Rajagopalan and Long 2005, Thorsby and Lie 2005). One possibility is that MHC variants associated with disease are deficient in their ability to bind key peptides that enable recognition of the pathogen and activation of the immune response. Consistent with this hypothesis, a correlation between peptide-MHC affinity and magnitude of response at the cellular level has been demonstrated, though the relevance to disease remains speculative (Geluk *et al.* 1998, Hill *et al.* 2003).

1.3 Biology of the Sub-Cellular Scale: Events within APCs

Peptide-MHC binding is by no means the only step that is regulated in the antigen presentation pathway. Other steps are controlled dynamically – by the amount of antigen available, for instance, or the constantly changing cytokine environment surrounding the APC – and allow antigen presentation to be fine-tuned. We describe these steps and others relevant to the MHC class II-mediated pathway in more detail below but refer the reader to a review for a full treatment (Bryant and Ploegh 2004).

1.3.1 Antigen Processing

Antigens for the MHC class II-mediated pathway are generally internalized by one of three routes before converging on the endosomal pathway: phagocytosis, fluid-phase pinocytosis, and receptor-mediated endocytosis. Internalized antigens then progress through the endosomal pathway where they encounter cathepsin proteases that degrade the antigens into peptides (Honey and Rudensky 2003). Peptides then either bind MHC class II molecules or are directed to lysosomes for degradation. Questions remain regarding antigen processing, such as how the many cathepsin proteases differ in function and whether it might be possible that proteins are first bound by MHC and then processed into peptides (Villadangos and Ploegh 2000).

1.3.2 MHC Expression

MHC class II expression normally occurs at low levels in resident APCs but can be up-regulated by cytokines such as IFN-γ and TNF (reviewed by van den Elsen *et al.* 2004). After IFN-γ binds to its receptor on the APC surface, a signal is propagated through the JAK-STAT pathway increasing the level of class II transactivator (CIITA) in the cell. CIITA acts as the master regulator of MHC class II transcription, and increased levels of CIITA lead to parallel increases in MHC class II expression several hours after exposure to IFN-γ. Nascent MHC class II molecules enter the ER and are coupled to another protein, invariant chain (Ii). The luminal domain of Ii binds the peptide-binding groove of MHC class II, protecting it from proteases, while the cytoplasmic domain of Ii directs the paired molecules to the endosomal pathway. After reaching the endosomal pathway MHC class II molecules retain a remnant of Ii, the class II invariant peptides (CLIP), until released by the enzymatic activity H2-/HLA-DM (Denzin and Cresswell 1995). Antigenic peptides then compete for binding to MHC class II with self peptides that are present at high levels and may bind greater than 80% of the available MHC class II in the absence of exogenous peptides (Adorini *et al.* 1988, Chicz *et al.* 1993). In complex with either self or exogenous peptides, MHC class II molecules then traffic to the cell surface where they may remain stably for days until they are recognized by $CD4^+$ T cells or internalized and degraded.

1.3.3 Differences among APC Types

Macrophages, dendritic cells (DCs), and B cells constitute the professional APCs, expressing not only MHC class II molecules but also co-stimulatory and adhesion molecules necessary to engage T cells. Both macrophages and DCs derive from a common precursor, the monocyte, which differentiates into one of the two cell types based on environmental cues (Chomarat *et al.* 2000, Chomarat *et al.* 2003), while B cells

are derived from hematopoietic cells in the bone marrow (Bryant and Ploegh 2004). Macrophages and DCs are found in overlapping distributions within the body in areas such as the lymph nodes.

Differences between macrophages and DCs occur in the rates at which they perform processes related to antigen presentation. DCs express 10-100 times the number of MHC class II molecules expressed by macrophages and also perform antigen uptake at generally increased rates (Inaba and Steinman 1985, Inaba *et al.* 1997). Consistent with these findings, fewer DCs are required to activate T cells than macrophages (Inaba and Steinman 1985). Though both DCs and macrophages perform antigen presentation, their roles in the development of the immune response are thought to be distinct. DCs take up antigen at the site of infection and migrate to the nearest lymph node to present antigen to naïve T cells, while macrophages present antigen primarily at the infection site to re-stimulate T cells (Reinhardt *et al.* 2001).

1.4 Biology of the Cellular Scale: T Cells

1.4.1 T Cell Receptor

pMHC complexes on the APC surface provide signals to T cells when engaged by TCRs and the co-stimulatory molecules CD4 and CD8 found on the T cell surface. Each TCR comprises two trans-membrane subunits of approximately equal size (designated α and β) (reviewed by Rudolph *et al.* 2006). The membrane-distal domain of each subunit resembles the immunoglobulin variable (V) domain and engages portions of both peptide and MHC molecule in the pMHC complex. A membrane-proximal domain resembles the immunoglobulin constant (C) domain and connects the membrane-distal domain to the transmembrane region and a short cytoplasmic tail. The CD3 molecule is associated with the cytoplasmic tail and plays an integral role in the signal transduction that follows TCR activation.

pMHC and TCR bind in a specialized structure that forms between the APC and T cell known as the immunological synapse. The structure of the synapse has been the subject of intense investigation and revealed to comprise concentric zones: the central supramolecular activation cluster (cSMAC), the peripheral supramolecular activation cluster (pSMAC), and the distal supramolecular activation cluster (dSMAC) (Cemerski and Shaw 2006). Within each zone is found a distinct complement of molecules contributed by the APC and T cell. For instance, pMHC and TCR are found within the cSMAC of the mature synapse, along with co-stimulatory molecules B7 and CD28 on APC and T cell surfaces, respectively. In the pSMAC surrounding the cSMAC is a palisade of structural molecules – complexes of intercellular adhesion molecule 1 (ICAM1) and leukocyte function associated antigen 1 (LFA1) contributed by APC and T cell, respectively – along with other co-stimulatory molecules such as B7 and CD28. Finally, in the recently defined dSMAC surrounding the pSMAC additional signaling between pMHC and TCR has recently been observed, though the contribution of signaling in the dSMAC to the overall T cell response is still unknown (Yokosuka *et al.* 2005, Varma *et al.* 2006).

1.4.2 Signal Transduction

Several intracellular signaling events follow engagement of pMHC by TCR. Briefly, these events include recruitment of kinases, activation of intermediate signaling molecules, and activation of transcription factors responsible for expression of molecules with effector function such as IL-2 and IFN-γ (reviewed in Samelson 2002, Liu 2005, and Weil and Israel 2006). Of the molecules involved, the most notable include the Src family kinases such as Lck, which phosphorylates the CD3 molecule after TCR activation, and ZAP-70, which is recruited to phosphorylated CD3 and phosphorylates the adapter protein LAT. In turn, phosphorylated LAT recruits other adapter proteins including

PLCγ1 and Grb2. Activation of several distinct signaling pathways follows which has several effects including a sustained increase in intracellular calcium and the activation of the transcription factors NF-AT, NF-κB, and AP-1. These transcription factors are ultimately responsible for the expression of molecules associated with T cell activation such as IL-2 and IFN-γ.

1.5 Tuberculosis and Antigen Presentation

1.5.1 Pathogenesis of *Mycobacterium tuberculosis*

Mycobacterium tuberculosis, the causative agent of tuberculosis, infects nearly a third of the human population and results in nearly two million deaths per year (WHO 2007). *M. tuberculosis* bacilli are inhaled on droplets, enter the lungs, and are phagocytosed by resident macrophages or dendritic cells (reviewed in Fenton 1998, Russell 2001). Residing in specialized phagosomes that do not fuse with lysosomes, the bacilli become dormant or slowly replicate. Meanwhile, monocytes are recruited to the site of infection and differentiate into macrophages which, together with infected macrophages and multinucleate macrophages (known as giant cells), form the center of a specialized structure known as a granuloma wherein the infection is contained. Surrounding the center are T cells which define the edge of the granuloma. When contained in a granuloma, the *M. tuberculosis* bacilli cannot be transmitted and the host has no signs of infection; this state called latency may persist for the lifetime of the host. In approximately one-tenth of these cases, however, the granuloma fails to continue containing the *M. tuberculosis* bacilli, resulting in active disease.

1.5.2 *M. tuberculosis* Inhibition of Antigen Presentation

Many pathogens including *M. tuberculosis* interfere with antigen presentation to evade immune surveillance and effect their own survival. To avoid immune surveillance,

M. tuberculosis inhibits antigen presentation in chronically infected macrophages. This inhibition can be demonstrated *in vitro*. When cultured macrophages are infected with *M. tuberculosis*, their capacity to present model antigen to T cell hybridoma is greatly reduced compared to uninfected controls (Gercken *et al.* 1994). Though the mechanisms by which *M. tuberculosis* achieves this inhibition have not been completely elucidated, several hypotheses have been proposed (e.g., in Moreno *et al.* 1988, Hmama *et al.* 1998, Noss *et al.* 2000). I discuss these hypotheses in detail in Chapter 2.

1.6 Models of Peptide-MHC Binding

Peptide-MHC binding is a prerequisite for antigen presentation and the event most likely to be affected by polymorphisms that exist within the MHC of human populations. From a clinical perspective these polymorphisms may distinguish individuals who succumb to a particular infectious disease from those who remain healthy, and significant effort has been expended to assess whether binding occurs between relevant peptide-MHC combinations. However, the sheer numbers of possible peptides (20^9 or ~10^{11} peptides of length nine) and MHC molecules (more than 2200 known HLA alleles) make this task all but impossible for anything more than a small sampling of the peptide-MHC combination space.

To circumvent this difficulty computational algorithms have been developed to predict whether binding occurs between particular combinations of peptide and MHC. In general these algorithms have the same aim as other algorithms in bioinformatics: to identify patterns in sequences that are known to either possess or not possess a particular trait. In this case the trait is binding to a particular MHC molecule. Statistical methods of varying degrees of complexity have been applied to peptide-MHC binding prediction. The simplest algorithms were based on the identification of motifs within peptides binding particular MHC (Rammensee 1995). An example of such a motif is the

requirement for a hydrophobic amino acid at the N-terminus (position 1) of a 9mer to bind MHC of the DR1 serotype, a guideline still generally followed today (Southwood *et al.* 1998). The advent of competitive binding assays allowed a more nuanced view of binding. Motifs that required certain amino acids to be present in MHC-binding peptides were superseded by matrices scoring amino acids at each position within the peptide. Different statistical methods could be used to generate the elements of the matrix, including nonlinear and linear programming (Parker *et al.* 1994, Murugan and Dai 2005), stepwise discriminant analysis (Mallios 1999, Mallios 2001), and partial least squares (Doytchinova and Flower 2002, Doytchinova and Flower 2003). One simplifying assumption made in many of these algorithms is that binding of each amino acid within the peptide to the MHC molecule occurs independently of adjacent as well as more distal amino acids. Though this assumption was largely confirmed by available crystal structures, algorithms were also developed that did not rely on this assumption based on machine learning methods. Several machine learning methods have now been incorporated into prediction algorithms including artificial neural networks (Brusic *et al.* 1998, Honeyman *et al.* 1998, Milik *et al.* 1998, Buus *et al.* 2003), hidden Markov models (Noguchi *et al.* 2002), and support vector machines (Zhao *et al.* 2003, Bhasin and Raghava 2004). A different approach has been to predict the structure of the peptide-MHC complex and attempt to calculate the free energy change (Altuvia *et al.* 1997, Schueler-Furman *et al.* 2000, Altuvia and Margalit 2004, Bui *et al.* 2006, Fagerberg *et al.* 2006). Structure-based prediction may someday supplant statistical- or machine learning-based algorithms but is currently hampered by the limited availability of solved structures and high computational costs. For a more comprehensive review of algorithms, the reader is referred elsewhere (Yu *et al.* 2002, Brusic *et al.* 2004).

1.6.1 Scoring Peptide-MHC Binding Prediction

An obvious question to ask about the preceding list of algorithms is how well each one performs compared to the others. To gauge prediction accuracy an algorithm is typically trained on one set of data comprising peptide sequences and their affinities for a particular MHC molecule and used to make predictions on a test set of peptides for which the affinities are also known. Algorithm output is then compared to the known affinities using one of several possible scoring measures. However, this task is complicated by differences in the nature of the available binding data and algorithm outputs. In some cases, affinity is measured directly as a continuous variable, other times only indirectly as a discrete variable (binding or non-binding). Some databases provide only lists of peptides that either bind or do not bind particular MHC variants (Rammensee 1999) while other databases provide a direct measure of affinity such as IC_{50} (Toseland *et al.* 2005). The appropriate scoring measure therefore differs according to whether known and predicted affinities are both continuous (Pearson correlation coefficient), both discrete (Matthews correlation coefficient), or discrete and continuous, respectively (area under receiver operating characteristic curve, or A_{ROC}). Both correlation coefficients vary between -1 and 1 while A_{ROC} ranges from 0.5 to 1.0. In both cases higher scores indicate more accurate predictions. Continuous data can be converted into discrete data by assuming that a certain threshold affinity is required for binding such as an IC_{50} of 500 nM (Sette *et al.* 1994) allowing some overlap between performance measures. Examples of scores obtained for several algorithms are provided in MHCBench (Singh and Raghava 2001). For example, using binding data for HLA-DRB1*0401 from which homologous sequences were removed, twelve algorithms were found to produce A_{ROC} scores between 0.57 and 0.76.

1.7 Models of the APC

Peptide-MHC binding is only one step of many that constitute the antigen presentation pathway, and other steps confer additional specificity to or alter the dynamics of which peptides are ultimately presented. In both MHC class I- and MHC class II-mediated antigen presentation, antigens are acquired (from either intracellular or extracellular sources), degraded into peptides (i.e., processed), and trafficked to the cell surface after binding MHC. At the same time MHC molecules are synthesized, trafficked between cellular compartments, and degraded. Many of these steps are subject to regulation by the cytokine environment and feedback signals. The peptides found to bind a particular MHC variant may therefore only provide a rough, static approximation of peptides that are ultimately presented in a dynamic fashion.

Models of antigen presentation must therefore account for more than peptide-MHC binding. In the case of MHC class I-mediated antigen presentation, at least two additional events are known to confer selectivity: proteasomal cleavage and TAP transport. Algorithms have been developed to predict which peptides progress through these stages, and only recently have they been linked with algorithms of peptide-MHC binding to represent antigen presentation *in toto* (Petrovsky and Brusic 2004, Donnes and Kohlbacher 2005). The result is a more accurate but still static picture of the peptides encountered by $CD8^+$ T cells.

1.7.1 ODE Models

In contrast, previous models of MHC class II-mediated antigen presentation have focused on its dynamic aspects (e.g., times required for certain steps to be completed and levels of pMHC presented on the APC surface) but not necessarily its specificity. These previous models were based on a mathematical representation known as ordinary differential equations (ODEs) in which each variable represents the level of a different

molecular species at a particular time, e.g., the number of pMHC complexes appearing on the APC surface a certain number of hours after the APC encounters antigen. Processes affecting the level of each species are then represented as terms in the equation for each variable. For instance, in the case of surface pMHC complexes, one term might represent the trafficking of pMHC complexes from the interior of the APC to its surface. The equation itself would then represent the rate of change in this variable as the sum total of these constituent processes.

ODEs are commonly used to represent systems that are both continuous and deterministic. One assumption inherent in the use of ODEs is that the represented entities exist as well-mixed populations, allowing their interactions to be approximated by the law of mass action. For MHC class II-mediated antigen presentation the available data validate this assumption. Baseline estimates of the number of MHC class II molecules expressed by APCs are on the order of 10^5 and antigen is typically present at high concentrations, at least *in vitro* ($>10^{12}$ peptides per cell in Hmama *et al.* 1998 and Noss *et al.* 2000). Furthermore, precedent for using ODEs has been provided by models of receptor-ligand systems of which peptide-MHC could be considered one instance (Lauffenburger and Linderman 1993).

Previous models have used ODEs to represent MHC class II-mediated antigen presentation with increasing levels of detail. The first published model included only those intracellular processes thought to be essential to antigen presentation (antigen uptake and processing, peptide-MHC binding, and MHC trafficking and recycling) but was sufficient to generate realistic time courses of peptide-MHC levels on the APC surface (Singer and Linderman 1990). Parameters that would have been difficult to manipulate experimentally, e.g., the rate of antigen uptake, were easily varied in the model, allowing the relationship between such parameters and the number of pMHC on the APC surface to be studied without concerns of inhibitor toxicity, etc. In later versions of this model, additional molecular species such as self peptides and TCR were included,

expanding the range of questions that could be approached (Singer and Linderman 1991, Agrawal and Linderman 1996).

1.7.2 Sensitivity Analysis

The creation of a mathematical model for antigen presentation or any other phenomenon entails the estimation of parameter values. In most cases these parameters represent rate constants of chemical processes, initial values for numbers of molecules, or probabilities of an event. Parameter values may be estimated in one of several ways, including direct experimental determination, fitting such that model output matches experimental observation, or constraints based on known relationships to other parameters. Each of these cases involves some degree of uncertainty which leads to uncertainty in the output of the model.

The effect of uncertainty in model parameter values on model output can be determined for any given model using sensitivity analysis. Different methods for sensitivity analysis exist, but all involve the correlation of variance in parameter values to variance in model output. For example, in the Latin hypercube sampling (LHS) algorithm, each parameter is first assigned a distribution, typically uniform or normal and centered on a baseline or estimated value, allowing the effect of under- and over-estimation to be examined. The entire range of each distribution is then sampled to generate a set of values for each parameter, and parameter values for each simulation are chosen to cover the entire parameter space efficiently. The extent to which each parameter affects the output can then be quantified using one of several metrics such as the partial rank correlation coefficient (PRCC). PRCC, like the more familiar Pearson correlation coefficient, varies between -1 and 1 indicating strongly negative and positive associations, respectively. A PRCC of 0 indicates no association. PRCC values can also be calculated at different time points of the simulation allowing the relative importance of

a particular parameter in determining model output to be examined over time. In addition a confidence interval can be determined for each PRCC, and differences between PRCCs can be tested for statistical significance (Meng *et al.* 1992). This allows parameters to be ranked in order of effect on output by PRCC magnitude.

Sensitivity analysis identifies processes that are important to the behavior of the system. These processes may represent potential targets for therapeutic intervention; that is, one could target a pathway to which cell behavior is sensitive as identified by sensitivity analysis.

1.8 Models of the T Cell

Models of the APC provide a useful view of the first stages of the immune response but are limited if T cell response is not also considered. T cells provide functional responses to the appearance of pMHC on the APC surface, and several models have sought to capture different aspects of T cell activation.

Some models have focused on the level of receptor-ligand engagement and how the kinetics of pMHC-TCR binding influence downstream events. For instance, in the model of Coombs *et al.* (2002), pMHC-TCR complexes are depicted in discrete, progressive states of activation, culminating in fully activated TCR that can either be internalized or return to a basal, inactivated state. Three zones of the surfaces of APCs and T cells are represented: a contact area, transition region, and remainder of the cell surfaces. pMHC-TCR binding occurs only within the contact area, though activated forms of the TCR are allowed to persist in the transition region. To represent pMHC and TCR, partial differential equations (PDEs) are used, allowing the level of each molecular species to be tracked with respect to a given distance away from the center of the contact area as well as a given time. TCR internalization serves as the output of this model,

occurring only after TCR are fully activated and free from pMHC in either the contact area or the transition region.

Using this model, Coombs *et al.* (2002) found that an optimal half-life for the pMHC-TCR complex might exist resulting in maximal TCR internalization. This optimal half-life occurs as a result of competition between serial engagement (the ability of a single pMHC to activate multiple TCRs, favoring a short half-life) and kinetic proofreading (the requirement for a pMHC-TCR complex to remain bound long enough to result in activation, favoring a long half-life). Subsequent iterations of this model recapitulated other experimental phenomena including the dissipation of an optimal half-life at high initial pMHC densities (Gonzalez *et al.* 2005, Utzny *et al.* 2006).

Other models have focused on signaling occurring within the T cell after TCR activation. For example, in the ODE model of Chan *et al.* (2004), two signaling molecules, a kinase (such as Lck) and a phosphatase, are depicted generically in inactivated and activated forms. pMHC-TCR binding leads to activation of the kinase which in turn promotes activation of the phosphatase as well as further activation of itself. The activated phosphatase then returns the kinase to its inactivated form. Two feedback loops are therefore represented, one positive (the kinase on itself) and the other negative (the phosphatase on the kinase). Like the model of Coombs *et al.* (2002), this model, which uses amount of activated kinase as its readout, shows how T cell response is sensitive to the kinetics of pMHC-TCR binding. Altan-Bonnet and Germain (2005) later created a model with similar feedback structure but focused on the role of one signaling pathway in particular, the MAP kinase cascade.

1.9 Motivation and Goals

Antigen presentation traverses several spatial and temporal scales in its mechanisms and its effects (Fig. 1.1). At the molecular scale, peptide-MHC binding must

occur but is by no means automatic. Variability exists in peptide sequence as well as in MHC sequence. At the cellular scale, a number of processes contribute to the appearance of pMHC complexes on the APC surface but many of these may be disrupted by pathogens. Finally, at the multi-cellular scale, the pMHC signal must be delivered to the T cell to attain a response, but the kinetics of pMHC-TCR binding or other, intracellular processes may determine whether this occurs.

Therefore, a model of antigen presentation must account for events occurring at each of these scales. In the following chapters I describe how I developed computational models at each scale and applied them to questions, initially scale-specific and then more multi-scale in nature. I also describe how I applied the models to the study of tuberculosis and its causative pathogen, *M. tuberculosis*.

Figures

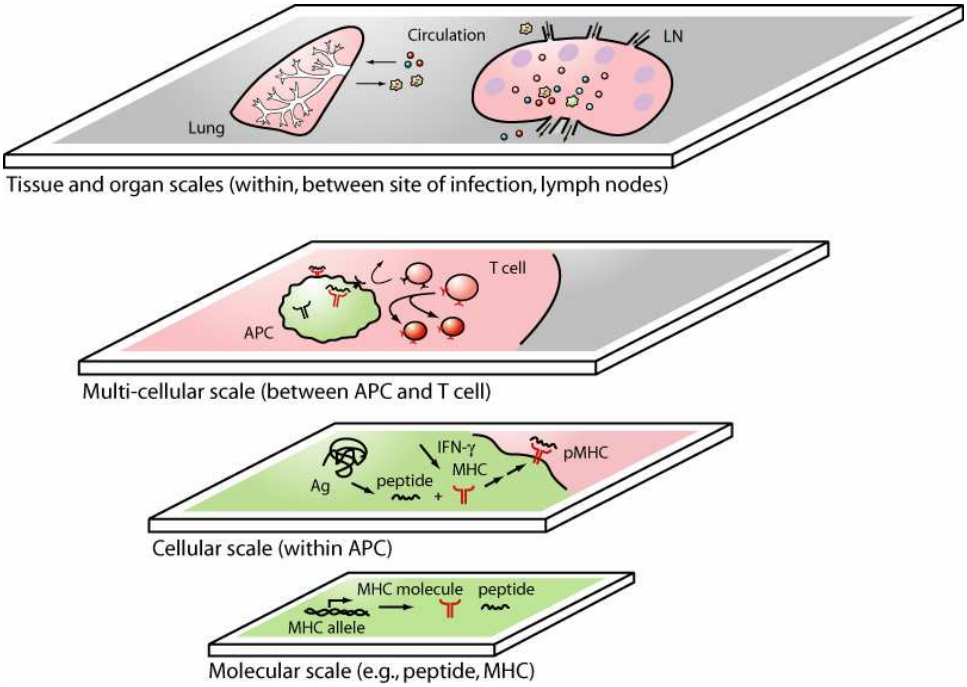

Figure 1.1. Multiple scales involved in antigen presentation.

1.10 References

Adorini, L., Muller, S., Cardinaux, F., Lehmann, P.V., Falcioni, F. and Nagy, Z.A. (1988) *In vivo* competition between self peptides and foreign antigens in T-cell activation. *Nature*, **334**, 623–625.

Agrawal, N.G. and Linderman, J.J. (1996) Mathematical modeling of helper T lymphocyte/antigen-presenting cell interactions: analysis of methods for modifying antigen processing and presentation. *J. Theor. Biol.*, **182**, 487–504.

Altan-Bonnet, G. and Germain, R.N. (2005) Modeling T cell antigen discrimination based on feedback control of digital ERK responses. *PLoS Biol.*, **3**, e356.

Altuvia, Y., Sette, A., Sidney, J., Southwood, S. and Margalit, H. (1997) A structure-based algorithm to predict potential binding peptides to MHC molecules with hydrophobic binding pockets. *Hum. Immunol.*, **58**, 1-11.

Altuvia, Y. and Margalit, H. (2004) A structure-based approach for prediction of MHC-binding peptides. *Methods*, **34**, 454–459.

Bhasin, M. and Raghava, G.P. (2004) Prediction of CTL epitopes using QM, SVM and ANN techniques. *Vaccine*, **22**, 3195–3204.

Brusic, V., Rudy, G., Honeyman, G., Hammer, J. and Harrison L. (1998) Prediction of MHC class II binding peptides using an evolutionary algorithm and artificial neural network. *Bioinformatics*, **14**, 121–130.

Brusic, V., Bajic, V.B. and Petrovsky, N. (2004) Computational methods for prediction of T-cell epitopes – a framework for modelling, testing, and applications. *Methods*, **34**, 436–443.

Bryant, P. and Ploegh, H. (2004) Class II MHC peptide loading by the professionals. *Curr. Opin. Immunol.*, **16**, 96–102.

Bui, H.H., Schiewe, A.J., von Grafenstein, H. and Haworth, I.S. (2006) Structural prediction of peptides binding to MHC class I molecules. *Proteins*, **63**, 43–52.

Buus, S., Lauemoller, S.L., Worning, P., Kesmir, C., Frimurer, T., Corbet, S., Fomsgaard, A., Hilden, J., Holm, A. and Brunak, S. (2003) Sensitive quantitative predictions of peptide-MHC binding by a 'Query by Committee' artificial neural network approach. *Tissue Antigens*, **62**, 378–384.

Cemerski, S. and Shaw, A. (2006) Immune synapses in T-cell activation. *Curr. Opin. Immunol.*, **18**, 298-304.

Chan, C., Stark, J. and George, A.J. (2004) Feedback control of T-cell receptor activation. *Proc. Biol. Sci.*, **271**, 931-9.

Chicz, R.M., Urban, R.G., Gorga, J.C., Vignali, D.A., Lane, W.S. and Strominger, J.L. (1993) Specificity and promiscuity among naturally processed peptides bound to HLA-DR alleles. *J. Exp. Med.*, **178**, 27–47.

Chomarat, P., Banchereau, J., Davoust, J. and Palucka, A.K. (2000) IL-6 switches the differentiation of monocytes from dendritic cells to macrophages. *Nat. Immunol.*, **1**, 510–514.

Chomarat, P., Dantin, C., Bennett, L., Banchereau, J. and Palucka, A.K. (2003) TNF skews monocyte differentiation from macrophages to dendritic cells. *J. Immunol.*, **171**, 2262–2269.

Coombs, D., Kalergis, A.M., Nathenson, S.G., Wofsy, C. and Goldstein, B. (2002) Activated TCRs remain marked for internalization after dissociation from pMHC. *Nat. Immunol.*, **3**, 926-31.

Denzin, L.K. and Cresswell, P. (1995) HLA-DM induces CLIP dissociation from MHC class II alpha-beta dimers and facilitates peptide loading. *Cell*, **82**, 155–165.

Donnes, P. and Kohlbacher, O. (2005) Integrated modeling of the major events in the MHC class I antigen processing pathway. *Protein Sci.*, **14**, 2132–2140.

Doytchinova, I.A. and Flower, D.R. (2002) Quantitative approaches to computational vaccinology. *Immunol. Cell Biol.*, 80, 270–279.

Doytchinova, I.A. and Flower, D.R. (2003) Towards the in silico identification of class II restricted T-cell epitopes: a partial least squares iterative self-consistent algorithm for affinity prediction. *Bioinformatics*,**19**, 2263–2270.

Doytchinova, I.A. and Flower, D.R. (2005) In silico identification of supertypes for class II MHCs. *J. Immunol.*, **174**, 7085-95.

Fagerberg, T., Cerottini, J.C. and Michielin, O. (2006) Structural prediction of peptides bound to MHC class I. *J. Mol. Biol.*, **356**, 521–546.

Fenton, M.J. Macrophages and tuberculosis. (1998) *Curr. Opin. Hematol.*, **5**, 72–78.

Geluk, A., Taneja, V., van Meijgaarden, K.E., Zanelli, E., Abou-Zeid, C., Thole, J.E., de Vries, R.R., David, C.S. and Ottenhoff, T.H. (1998) Identification of HLA class II-restricted determinants of Mycobacterium tuberculosis-derived proteins by using HLA transgenic, class II-deficient mice. *Proc. Natl. Acad. Sci. USA*, **95**, 10797–10802.

Gercken, J., Pryjma, J., Ernst, M., Flad, H.D. 1994. Defective antigen presentation by *Mycobacterium tuberculosis*-infected monocytes. *Infect. Immun.*, **62**, 3472-3478.

Gonzalez, P.A., Carreno, L.J., Coombs, D., Mora, J.E., Palmieri, E., Goldstein, B., Nathenson, S.G. and Kalergis, A.M. (2005) T cell receptor binding kinetics required for T cell activation depend on the density of cognate ligand on the antigen-presenting cell. *Proc. Natl. Acad. Sci. USA*, **102**, 4824-9.

Hill, J.A., Wang, D., Jevnikar, A.M., Cairns, E. and Bell, D.A. (2003) The relationship between predicted peptide-MHC class II affinity and T-cell activation in a HLA-DRβ1*0401 transgenic mouse model. *Arthritis Res. Ther.*, **5**, R40–R48.

Honeyman, M., Brusic, V., Stone, N. and Harrison, L. (1998) Neural network-based prediction of candidate T-cell epitopes. *Nat. Biotechnol.*, **16**, 966-969.

Hmama, Z., Gabathuler, R., Jefferies, W.A., de Jong, G. and Reiner, N.E. 1998. Attenuation of HLA-DR expression by mononuclear phagocytes infected with *Mycobacterium tuberculosis* is related to intracellular sequestration of immature class II heterodimers. *J. Immunol.*, **161**, 4882-4893.

Honey, K. and Rudensky, A.Y. Lysosomal cysteine proteases regulate antigen presentation. *Nat. Rev. Immunol.* **3**, 472–482.

Inaba, K. and Steinman, R.M. (1985) Protein-specific helper T-lymphocyte formation initiated by dendritic cells. *Science*, **229**, 475–479.

Inaba, K., Pack, M., Inaba, M., Sakuta, H., Isdell, F. and Steinman, R.M. (1997) High levels of a major histocompatibility complex II-self peptide complex on dendritic cells from the T cell areas of lymph nodes. *J. Exp. Med.*, **186**, 665–672.

Jones, E.Y. (1997) MHC class I and class II structures. *Curr. Opin. Immunol.*, **9**, 75-9.

Lauffenburger, D.A. and Linderman, J.L. (1993) *Receptors: Models for Binding, Trafficking and Signaling.* New York: Oxford University Press.

Liu, J.O. (2005) The yins of T cell activation. *Sci STKE*, **265**, re1.

Lund, O., Nielsen, M., Kesmir, C., Petersen, A.G., Lundegaard, C., Worning, P., Sylvester-Hvid, C., Lamberth, K., Roder, G., Justesen, S., Buus, S. and Brunak, S. (2004) Definition of supertypes for HLA molecules using clustering of specificity matrices. *Immunogenetics*, **55**, 797-810.

Mahanty, S., Hutchinson, K., Agarwal, S., McRae, M., Rollin, P.E. and Pulendran, B. (2003) Cutting edge: impairment of dendritic cells and adaptive immunity by Ebola and Lassa viruses. *J. Immunol.*, **170**, 2797–2801.

Mallios, R.R. (1999) Class II MHC quantitative binding motifs derived from a large molecular database with a versatile iterative stepwise discriminant analysis meta-algorithm. *Bioinformatics*, **15**, 432–439.

Mallios, R.R. (2001) Predicting class II MHC/peptide multi-level binding with an iterative stepwise discriminant analysis meta-algorithm. *Bioinformatics*, **17**, 942–948.

McFarland, B.J. and Beeson, C. (2002) Binding interactions between peptides and proteins of the class II major histocompatibility complex. *Med. Res. Rev.*, **22**, 168-203.

Meng, X.L., Rosenthal, R. and Rubin, D.B. (1992) Comparing correlated correlation coefficients. *Psychol. Bull.*, **111**, 172–175.

Milik, M., Sauer, D., Brunmark, A.P., Yuan, L., Vitiello, A., Jackson, M.R., Peterson, P.A., Skolnick, J. and Glass, C.A. (1998) Application of an artificial neural network to predict specific class I MHC binding peptide sequences. *Nat. Biotechnol.*, **16**, 753–756.

Moreno C., Mehlert A. and Lamb, J. 1988. The inhibitory effects of mycobacterial lipoarabinomannan and polysaccharides upon polyclonal and monoclonal human T cell proliferation. *Clin. Exp. Immunol.*, **74**, 206-210.

Murugan, N. and Dai, Y. 2005. Prediction of MHC class II binding peptides based on an iterative learning model. *Immunome Res.*, **1**, 6.

Noss, E.H., Harding, C.V. and Boom, W.H. 2000. *Mycobacterium tuberculosis* inhibits MHC class II antigen processing in murine bone marrow macrophages. *Cell. Immunol.*, **201**, 63-74.

Ou, D., Mitchell, L.A. and Tingle, A.J. (1998) A new categorization of HLA DR alleles on a functional basis. *Hum. Immunol.*, **59**, 665-76.

Parker, K.C., Bednarek, M.A. and Coligan, J.E. (1994) Scheme for ranking potential HLA-A2 binding peptides based on independent binding of individual peptide side-chains. *J. Immunol.*, **152**, 163-175.

Petrovsky, N. and Brusic, V. (2004) Virtual models of the HLA class I antigen processing pathway. *Methods*, **34**, 429–435.

Rajagopalan, S. and Long, E.O. (2005) Understanding how combinations of HLA and KIR genes influence disease. *J. Exp. Med.*, **201**, 1025–1029.

Rammensee, H. (1995) Chemistry of peptides associated with MHC class I and class II molecules. *Curr. Opin. Immunol.*, **7**, 85-96.

Rammensee, H., Bachmann, J., Emmerich, N.P., Bachor, O.A. and Stevanovic, S. (1999) SYFPEITHI: database for MHC ligands and peptide motifs. *Immunogenetics*, **50**, 213–219.

Reinhardt, R.L., Khoruts, A., Merica, R., Zell, T. and Jenkins, M.K. (2001) Visualizing the generation of memory CD4 T cells in the whole body. *Nature*, **410**, 101–105.

Robinson, J., Waller, M.J., Parham, P., de Groot, N., Bontrop, R., Kennedy, L.J., Stoehr, P. and Marsh, S.G.E. (2003) IMGT/HLA and IMGT/MHC: sequence databases for the study of the major histocompatibility complex. *Nuc. Acids Res.*, **31**, 311-314

Noguchi, H., Kato, R., Hanai, T., Matsubara, Y., Honda, H., Brusic, V. and Kobayashi, T. (2002) Hidden markov model based prediction of antigenic peptides that interact with MHC class II molecules. *J. Biosci. Bioeng.*, **94**, 264–270.

Rock, K.L. and Shen, L. (2005) Cross-presentation: underlying mechanisms and role in immune surveillance. *Immunol. Rev.*, **207**, 166-83.

Rudolph, M.G., Stanfield, R.L. and Wilson, I.A. (2006) How TCRs bind MHCs, peptides, and coreceptors. *Annu. Rev. Immunol.*, **24**, 419-66.

Samelson, L.E. (2002) Signal transduction mediated by the T cell antigen receptor: the role of adapter proteins. *Annu. Rev. Immunol.*, **20**, 371-394.

Schueler-Furman, O., Altuvia, Y., Sette, A. and Margalit, H. (2000) Structure-based prediction of binding peptides to MHC class I molecules: application to a broad range of MHC alleles. *Protein Sci.*, **9**, 1838-1846.

Sette, A., Vitiello, A., Reherman, B., Fowler, P., Nayersina, R., Kast, W.M., Melief, C.J., Oseroff, C., Yuan, L., Ruppert, J., Sidney, J., del Guercio, M.F., Southwood, S., Kubo, R.T., Chesnut, R.W., Grey, H.M. and Chisari, F.V. (1994) The relationship between class I binding affinity and immunogenicity of potential cytotoxic T cell epitopes. *J. Immunol.*, **153**, 5586–5592.

Singer, D.F. and Linderman, J.J. (1990) The relationship between antigen concentration, antigen internalization, and antigenic complexes: modeling insights into antigen processing and presentation. *J. Cell. Biol.*, **111**, 55-68.

Singer, D.F. and Linderman, J.J. (1991) Antigen processing and presentation: how can a foreign antigen be recognized in a sea of self proteins? *J. Theor. Biol.*, **151**, 385-404.

Singh, H. and Raghava, G.P. (2001) ProPred: prediction of HLA-DR binding sites. *Bioinformatics*, **17**, 1236-7.

Southwood, S., Sidney, J., Kondo, A., del Guercio, M.F., Appella, E., Hoffman, S., Kubo, R.T., Chesnut, R.W., Grey, H.M. and Sette, A. (1998) Several common HLA-DR types share largely overlapping peptide binding repertoires. *J. Immunol.*, **160**, 3363–3373.

Thorsby, E. and Lie, B.A. (2005) HLA associated genetic predisposition to autoimmune diseases: genes involved and possible mechanisms. *Transpl. Immunol.*, **14**, 175–182.

Toseland, C.P., Clayton, D.J., McSparron, H., Hemsley, S.L., Blythe, M.J., Paine, K., Doytchinova, I.A., Guan, P., Hattotuwagama, C.K. and Flower, D.R. (2005) AntiJen: a quantitative immunology database integrating functional, thermodynamic, kinetic, biophysical, and cellular data. *Immunome Res.*,**1**, 4.

Trombetta, E.S. and Mellman, I. (2005) Cell biology of antigen processing *in vitro* and *in vivo*. *Annu. Rev. Immunol.*, **23**, 975-1028.

Utzny, C., Coombs, D., Muller, S. and Valitutti, S. (2006) Analysis of peptide/MHC-induced TCR downregulation: deciphering the triggering kinetics. *Cell Biochem. Biophys.*, **46**, 101-11.

van den Elsen, P.J., Holling, T.M., Kuipers, H.F. and van der Stoep, N. (2004) Transcriptional regulation of antigen presentation. *Curr. Opin. Immunol.*, **16**, 67-75.

Varma, R., Campi, G., Yokosuka, T., Saito, T. and Dustin, M.L. (2006) T cell receptor-proximal signals are sustained in peripheral microclusters and terminated in the central supramolecular activation cluster. *Immunity*, **25**, 117-27.

Villadangos, J.A. and Ploegh, H.L. (2000) Proteolysis in MHC class II antigen presentation: who's in charge? *Immunity*, **12**, 233-9.

Vukmanovic, S., Neubert, T.A. and Santori, F.R. (2003) Could TCR antagonism explain associations between MHC genes and disease? *Trends Mol. Med.*, **9**, 139–146.

Weil, R. and Israel, A. (2006) Deciphering the pathway from the TCR to NF-kappaB. *Cell Death Differ.*, **13**, 826-33.

World Health Organization. (2007) Tuberculosis, Fact Sheet No. 104, www.who.int/mediacentre/factsheets/fs104/en/.

Yewdell, J.W. (2007) Plumbing the sources of endogenous MHC class I peptide ligands. *Curr. Opin. Immunol.*, **19**, 79-86.

Yokosuka, T., Sakata-Sogawa, K., Kobayashi, W., Hiroshima, M., Hashimoto-Tane, A., Tokunaga, M., Dustin, M.L. and Saito, T. (2005) Newly generated T cell receptor microclusters initiate and sustain T cell activation by recruitment of Zap70 and SLP-76. *Nat. Immunol.*, **6**, 1253-62.

Yu, K., Petrovsky, N., Schonbach, C., Koh, J.Y. and Brusic, V. (2002) Methods for prediction of peptide binding to MHC molecules: a comparative study. *Mol. Med.*, **8**, 137–148.

Zhao, Y., Pinilla, C., Valmori, D., Martin, R. and Simon, R. (2003) Application of support vector machines for T-cell epitopes prediction. *Bioinformatics*, **19**, 1978–1984.

CHAPTER 2

Why *M. tuberculosis* Has Multiple Mechanisms to Inhibit Antigen Presentation

2.1 Introduction

Macrophages play dual roles during tuberculosis (TB) infection (Fenton 1998). On the one hand, they serve as the preferred host for *Mycobacterium tuberculosis* (*Mtb*), the intracellular pathogen that causes TB. On the other hand, they also help to alert the immune system to the presence of *Mtb* and, if activated, can eliminate it directly. Activation depends on the presentation of antigenic peptide-MHC class II (pMHC) complexes on the macrophage surface that can bind T cell receptors (TCRs) on cognate $CD4^+$ T helper cells. pMHC-TCR binding induces $CD4^+$ T helper cells to secrete IFN-γ which stimulates macrophages to produce molecules capable of killing *Mtb* such as nitric oxide (Chan *et al.* 2001). This process constitutes an important arm of cell-mediated immunity and may determine infection outcome (Kaufmann 1999).

The fact that *Mtb* inhibits antigen presentation in macrophages is now well established (Pancholi *et al.* 1993). Initial studies showed that fewer macrophages infected with mycobacteria express detectable levels of antigen on their surface compared to uninfected macrophages (Kaye *et al.* 1986, Mshana *et al.* 1988). Functional assays later confirmed that infected macrophages are deficient in their ability to signal $CD4^+$ T helper cells by measuring T cell response. The magnitude of T cell response is in turn proportional to pMHC levels, assuming a lower threshold number of pMHC complexes has been exceeded (Bekkhoucha *et al.* 1984, Demotz *et al.* 1990). Using such an assay

Gercken *et al.* (1994) found that monocytes co-cultured with *Mtb* for six days exhibit a three- to ten-fold reduction in their ability to stimulate T cell proliferation compared to uninfected controls. Furthermore, higher numbers of *Mtb* bacilli, e.g., MOI (multiplicity of infection) of 50 versus MOI of 10, correlated with lower T cell response levels. Subsequent studies provided further evidence that an inverse relationship exists between *Mtb* infectious dose and T cell response (Noss *et al.* 2000, Mazzaccaro *et al.* 1996).

After it was established that *Mtb* inhibits antigen presentation in macrophages, several intracellular mechanisms were proposed (reviewed in Harding *et al.* 2003). Moreno *et al.* (1988) observed that macrophages co-cultured with the *Mtb* cell wall component lipoarabinomannan (LAM) fail to present antigen from whole inactivated virus though presentation of synthesized epitope is unimpaired. This observation led to the hypothesis that *Mtb* inhibits antigen presentation at the stage of antigen processing, a hypothesis also made by Noss *et al.* (2000). Later, based on the observation that *Mtb*-infected monocytes do not produce stable pMHC complexes and do not localize labeled MHC class II molecules and antigens to the same intracellular compartment, Hmama *et al.* (1998) proposed that *Mtb* affects MHC class II at a post-translational stage such as maturation (delivery to the MIIC endosome or Ii processing) or peptide loading. Finally, based on the observation that infected macrophages express lower levels of MHC class II mRNA than uninfected macrophages, Noss *et al.* (2000) proposed that *Mtb* inhibits MHC class II mRNA synthesis.

The goal of the present study is to investigate why multiple mechanisms have been proposed to explain how *Mtb* inhibits antigen presentation. In particular, we address three issues using a mathematical model: (1) what purpose multiple mechanisms may serve, (2) if experimental protocols may have favored the detection of some mechanisms over others, and (3) if alternative mechanisms exist that may be used to guide future experiments. Our immediate motivation stems from conflicting data in the literature regarding these mechanisms. Specifically, we refer to the observation by Hmama *et al.*

(1998) that MHC class II mRNA levels were unchanged in infected cells and the observation by Noss *et al.* (2000) that MHC class II mRNA levels decreased in infected cells. Because the two studies differed with respect to experimental conditions (e.g., macrophage cell type, *Mtb* strain, and degree of IFN-γ-induced activation), it is unclear if the conclusions hold in general. We seek to help clarify these observations with our model.

2.2 Methods

2.2.1 Model Overview

Our mathematical model comprises a set of ordinary differential equations representing the major intracellular processes that contribute to antigen presentation within the context of a single macrophage (Fig. 2.1). These processes relate to MHC class II expression (at both mRNA and protein levels), antigen processing, and peptide-MHC binding and trafficking and include the processes hypothesized to be targeted by *Mtb*. Our model also accounts for the effects of IFN-γ which is typically added to cultured macrophages during studies on antigen presentation (Hmama *et al.* 1998, Noss *et al.* 2000).

To represent these processes we use ordinary differential equations which allow large numbers of molecules to be tracked. For each molecular species we derive an equation for the rate of change using the law of mass action and estimate parameter values using published experimental data. In total our model uses 16 equations and 30 parameters to simulate antigen presentation within the context of a single macrophage. Equations and parameter values, as well as details of how equations were derived and parameter values estimated, can be found in Supporting Information.

2.2.2 Simulations Using the Mathematical Model

The baseline model comprises Equations 1-16, the parameters in Table 2.3, and the initial conditions in Table 2.4 (Supporting Information). Protocol-specific parameter values and initial conditions can also be found in Supporting Information. To generate simulations using the mathematical model, we use the NDSolve function of Mathematica v4.2 (Wolfram Research, Inc.) as well as our own differential equation solver coded in C and run on Sun UNIX machines for confirmation of numerical results. We analyze model output in terms of major features such as relative changes in numbers of molecules and times at which highest levels are reached. As a marker for antigen presentation, we generally use the number of surface-localized exogenous peptide-bound MHC class II molecules (Equation 16, Supporting Information).

2.2.3 Representation of the Inhibitory Effects of *Mtb* on Intracellular Processes

To represent the inhibitory effect that *Mtb* is hypothesized to have on an intracellular process, we decrease the corresponding parameter in the model by a factor proportional to experimental infectious dose. We assume that the number of *Mtb* bacilli does not change significantly on the timescales of the protocols being simulated based on the observation that the doubling time of *Mtb* is on the order of days (Dunn and North 1995). We also assume that the inhibitory effect exerted by *Mtb* on any given intracellular process saturates at high levels of bacilli. Therefore, we represent the inhibitory effect as a multiplicative factor having a value between 0 and 1 (corresponding to complete inhibition and no inhibition, respectively) that approaches 0 as the number of bacilli increases. Further details are provided in Supporting Information.

2.2.4 Sensitivity Analysis

The goal of sensitivity analysis is to correlate variances in parameter values to variances in model output and is useful when parameter values are not known with certainty. Sampling-based sensitivity analysis entails specifying a distribution for each parameter from which values are selected at random and used in model simulations (Helton and Davis 2001). In particular we use Latin hypercube sampling which allows several parameters to be analyzed simultaneously in a computationally efficient manner. To quantify the correlation of model output with each parameter, we calculate a partial rank correlation coefficient (PRCC) value. PRCC values vary between -1 and 1, corresponding to perfect negative and positive correlations, respectively, and can be further differentiated based on p values derived from Student's *t*-tests. We use the algorithm of Blower and Dowlatabadi (1994) implemented in both Mathematica and our own differential equation solver. In general, we specify a uniform distribution for each parameter with a range of 10% and 190% of the baseline value, allowing us to examine the effects of both decreases and increases in each parameter.

2.3 Results

2.3.1 Baseline Characteristics

In the absence of IFN-γ and antigen, conditions that we used as a negative control, seven molecular species in the model were present in non-zero quantities: free IFN-γ receptors, MHC class II mRNA, free intracellular and surface MHC class II molecules, self peptides, and intracellular and surface self peptide-MHC class II complexes. These results are consistent with the finding that cultured macrophages constitutively express several molecules relevant to antigen presentation at basal levels including IFN-γ receptors and MHC class II molecules (Hume 1985, Celada *et al.* 1985).

2.3.2 Dynamics of IFN-γ Response

As one positive control we simulated the addition of IFN-γ to macrophages and compared dynamics of the response to experimental observations. In response to IFN-γ treatment, CIITA mRNA levels in the model increased immediately and reached a maximum approximately 14 hours later, while MHC class II mRNA levels increased more gradually and continued to increase for the first 24 hours (Fig. 2.2A). Pai *et al.* measured levels of CIITA and MHC class II mRNAs 6, 12, and 24 hours after adding IFN-γ and observed highest levels at the 12- and 24-hour time points, respectively, in agreement with our model (Fig. 2.2B, Pai *et al.* 2002). We also compared the coupled dynamics of MHC class II mRNA and protein expression from our model to experimental data. In our simulations highest MHC class II mRNA and protein levels were attained approximately 45 and 60 hours after IFN-γ treatment, respectively (Fig. 2.2C). In comparison, highest MHC class II mRNA and protein levels were observed experimentally 48 and 72 hours after IFN-γ treatment, respectively (Fig. 2.2D, Cullell-Young *et al.* 2001). Although MHC class II protein expression reaches its highest levels in the model in less time than observed experimentally, this apparent difference may be attributable to the sparseness of experimental time points.

2.3.3 Dynamics of Antigen Presentation

In the presence of exogenous antigen the number of surface pMHC complexes in our model rapidly increases, reaches a maximum approximately 3 hours later, and then decreases over the course of several hours (Fig. 2.2E). Antigen presentation by macrophages not pretreated with IFN-γ has been found to exhibit similar dynamics experimentally (Fig. 2.2F, Buus and Werdelin 1986, Ziegler and Unanue 1981). In such cases antigen presentation can be detected by T cell hybridoma assay minutes after the addition of antigen (Buus and Werdelin 1986, Ziegler and Unanue 1981). These

macrophages elicit maximal responses after one to four hours and remain capable of eliciting responses at the same or slightly decreased levels for several more hours (Buus and Werdelin 1986, Ziegler and Unanue 1981). Another feature of our model is dose-dependence between exogenous antigen concentration and maximum number of resultant surface pMHC complexes (data not shown) which has also been observed experimentally with T cell responses (Demotz *et al.* 1990, Reske-Kunz *et al.* 1984).

2.3.4 Increases in Antigen Presentation Due to IFN-γ Pretreatment

Experimental studies on antigen presentation by macrophages typically use both IFN-γ and exogenous antigen. Timing of IFN-γ treatment may be important, as studies in which IFN-γ is added prior to antigen show that pretreated macrophages are capable of eliciting T cell responses at levels several fold higher than untreated macrophages (Delvig *et al.* 2002). We simulated the addition of IFN-γ 16 h prior to exogenous antigen and observed a two-fold increase in surface pMHC levels compared to untreated levels (Fig. 2.2G). This result is consistent with T cell proliferation data from Delvig *et al.* (2002) (Fig. 2.2H). In subsequent simulations we avoided the issue of pretreatment timing by using the simultaneous addition of IFN-γ and antigen unless stated otherwise.

2.3.5 Simulations of *Mtb* and Its Hypothesized Mechanisms

After testing the model under the preceding conditions, we used the model to simulate the inhibition of various intracellular processes targeted by *Mtb*. These processes included: antigen processing (Moreno *et al.* 1998), MHC class II protein maturation (Hmama *et al.* 1998), MHC class II peptide loading (Hmama *et al.* 1998), and MHC class II mRNA synthesis which we consider MHC class II transcription (Noss *et al.* 2002); we designate these hypotheses as H_1, H_2, H_3, and H_4, respectively. We then simulated the simultaneous addition of IFN-γ and antigen and recorded surface pMHC levels at time

points spanning four orders of magnitude (0.1, 1.0, 10, and 100 hours). These results were compared to results from the baseline model in which no processes were inhibited.

In general, inhibiting any particular intracellular process had either immediate or delayed effects on antigen presentation (Table 2.1). When antigen processing (H_1) or MHC class II peptide loading (H_3) was inhibited, surface pMHC levels were immediately affected as indicated at the earliest time point, 0.1 h. The deviation from baseline levels was reduced at intermediate 1 h and 10 h time points and then increased by the final 100 h time point. In contrast, inhibition of MHC class II maturation (H_2) or MHC class II transcription (H_4) resulted in negligible reductions in surface pMHC levels at the 0.1 h time point. However, these levels increasingly deviated from baseline levels at 1 h, 10 h, and 100 h time points. Both H_2 and H_4 targeted MHC class II expression and required a delay of at least 10 h to have substantial effects (greater than 25% change in surface pMHC levels). We also simulated the inhibition of pairs of intracellular processes to determine the effect multiple mechanisms may have on antigen presentation when acting together (cf. H_1+H_4 and H_2+H_3 in Table 2.1). Inhibitory mechanisms were synergistic and decreased antigen presentation levels to a greater extent in pairs than singly. In these simulations each intracellular process was inhibited to the same degree. In a separate set of simulations we used varying degrees of inhibition, further differentiating mechanisms targeting MHC class II expression from other mechanisms (Fig. 2.5, Supporting Information).

2.3.6 Simulations of Previous Experimental Protocols

To determine if previous experimental protocols may have favored the detection of some mechanisms over others and if any of the four previously hypothesized mechanisms could account for all of the observed changes in macrophages infected with *Mtb*, we simulated two different experimental protocols under each hypothesized

mechanism (Hmama *et al.* 1998, Noss *et al.* 2000). These protocols differed in several ways that could be accounted for in our model, including duration for which cells were exposed to IFN-γ and *Mtb* as well as concentration of IFN-γ and number of *Mtb* bacilli used (Fig. 2.3A, B). These protocols also differed with respect to the macrophage cell line and *Mtb* strain used, but these factors fell outside the scope of our model and were not considered.

In our simulations of the experimental protocol of Hmama *et al.* (1998), we found that only an inhibition of MHC class II protein maturation (H_2) was consistent with all of their observations. In the absence of *Mtb* the levels of several molecules rose over baseline levels during the course of this protocol including CIITA mRNA, MHC class II mRNA, and MHC class II protein (Fig. 2.3C, D, E). Only H_2 and another hypothesized mechanism, inhibition of MHC class II transcription (H_4), led to reductions in surface MHC class II expression of the same magnitude as those observed by Hmama *et al.* (1998): 42% and 86% using heat-killed and live *Mtb* bacilli, respectively (Fig. 2.3E). However, H_4 also led to a significant reduction in MHC class II mRNA levels which was not observed by Hmama *et al.* (1998) and could therefore be ruled out as a possible mechanism (Fig. 2.3D).

When we simulated the experimental protocol of Noss *et al.* (2000), we found that only an inhibition of MHC class II transcription (H_4) was capable of producing substantial changes in the levels of all three molecules they monitored. In our simulations this mechanism reduced levels of MHC class II mRNA, total MHC class II protein, and surface pMHC by 54%, 31%, and 31%, respectively (Fig. 2.3F, G, H). Another mechanism, inhibition of MHC class II protein maturation (H_2), reduced levels of these molecules by 0%, 55%, and 55%, respectively (Fig. 2.3F, G, H). In comparison, Noss *et al.* (2000) measured reductions of 80%, 30%, and between 40% and 80%, respectively, consistent with H_2 but not H_4. Interestingly, in our simulations of this protocol neither

inhibition of antigen processing (H_1) nor inhibition of MHC class II peptide loading (H_3) had any significant effect on surface pMHC levels (Fig. 2.3H).

2.3.7 Sensitivity to Changes in Other Intracellular Processes

While many intermediates of the antigen presentation pathway have been monitored in macrophages following *Mtb* infection *in vitro* (Hmama *et al.* 1998), assays for other processes represented in our model have either not been developed or not been applied to this context. To determine what effect changes in these processes might have on antigen presentation, we varied all of the corresponding rates, rate constants, and scaling factors as well as experimental conditions in the model over a defined range and tracked surface pMHC levels over time. We then calculated the correlation between these levels and specific parameter values at 1 h, 10 h, and 100 h time points.

We found that surface pMHC levels correlated significantly with a number of different intracellular processes, including several not previously considered (Table 2.2). In particular, at times less than 10 hours following exposure to IFN-γ and antigen, surface pMHC levels correlated positively with rate constants for antigen uptake by pinocytosis and MHC class II trafficking to the cell surface as well as with the concentration of exogenous antigen. When the concentration of exogenous antigen was sufficiently low, other processes correlated strongly with surface pMHC levels on this timescale including delivery of antigen to lysosomes and self peptide production (data not shown). At times greater than 10 hours following exposure to IFN-γ and antigen, surface pMHC levels correlated with factors affecting MHC class II expression including CIITA transcription and translation and the concentration of IFN-γ in the medium as well as MHC class II transcription and protein maturation.

2.4 Discussion

Multiple hypotheses have been offered to explain how *Mtb* inhibits antigen presentation in macrophages to escape immune surveillance. These hypotheses stem from different experimental protocols that appear in at least one instance to have led to conflicting results. In this study we address why several mechanisms have been hypothesized by formulating a mathematical model of antigen presentation that accounts for different experimental conditions and can be used to simulate each mechanism.

2.4.1 *Mtb* Mechanisms Differ in Timing of Effect

We found that hypothesized *Mtb* mechanisms generally fall into one of two categories: those having an immediate effect on the ability of the cell to present antigen and those requiring a delay of approximately 10 hours to have an effect. The first subset of mechanisms targets intracellular processes involved in the initial formation of pMHC complexes including antigen processing and MHC class II peptide loading. In our simulations the effectiveness of these mechanisms in inhibiting antigen presentation decreased after an intermediate length of time (at 1 h and 10 h) and later increased (at 100 h). The intermediate decrease resulted from new rounds of pMHC binding resulting from prolonged exposure to IFN-γ and increasing numbers of free MHC class II. The second subset of mechanisms targets intracellular processes necessary for the continued supply of nascent MHC class II molecules including MHC class II transcription and protein maturation. In our simulations the effect of these mechanisms on antigen presentation steadily increased over time as a greater proportion of surface pMHC complexes involved nascent MHC class II.

These results are consistent with the intuitive notion that disruptions at different points along the antigen presentation pathway, or any multi-enzymatic pathway, require different lengths of time to manifest in the end product. These results are also consistent

with the interpretation of the experimental data of Noss *et al.* (2000) given by Heldwein and Fenton (2002), that substantial inhibition of MHC class II expression requires prolonged (> 18 h) incubation with *Mtb*. The requirement of a delay of greater than 10 hours for inhibition of MHC class II expression to affect antigen presentation was also evident in our sensitivity analysis.

The fact that these four hypothesized mechanisms appear to impair the same cellular function, antigen presentation, raises the question: do these mechanisms serve the same purpose and act redundantly or do they serve subtly different purposes? Our results suggest that these mechanisms act on different timescales and therefore serve different purposes. As demonstrated in our simulations of pairs of mechanisms, having mechanisms that operate on both shorter and longer timescales may allow *Mtb* to exert continuous inhibition on antigen presentation despite external sources of IFN-γ. In contrast, having only a single mechanism or multiple mechanisms that act on the same timescale may result in an inhibitory effect that either abates with time (if MHC class II expression increases) or is delayed.

Nascent and recycling MHC class II molecules may have distinct roles in antigen-presenting cells (Pinet and Long 1998), and *Mtb* may have evolved mechanisms to undermine both sources of MHC class II. T cells require at least 2 to 4 hours of stimulation to become fully activated (Weiss *et al.* 1987), and mechanisms acting on timescales of both minutes and hours may be physiologically relevant. A recent study by Huppa *et al.* shows that signaling between an antigen-presenting cell and a T cell has a cumulative effect over 10 hours and is sensitive to disruptions that occur even several hours after initial contact (Huppa *et al.* 2003).

2.4.2 Previous Protocols Favor Detection of Mechanisms Targeting MHC Expression

Our simulations of previous experimental protocols suggest that *Mtb* mechanisms targeting MHC class II expression may have been responsible for most of the changes observed in levels of various molecules. Specifically, in simulations of the protocols of Hmama *et al.* (1998) and Noss *et al.* (2000), only mechanisms targeting processes associated with MHC class II expression were found to produce changes of the same magnitude as those observed. While no single mechanism was found to account for all of the observations, these results do support the individual conclusions of Hmama *et al.* (1998) and Noss *et al.* (2000) who implicated inhibition of MHC class II protein maturation and MHC class II mRNA synthesis, respectively.

Why did Noss *et al.* (2000) observe a decrease in MHC class II mRNA levels but not Hmama *et al.* (1998)? Noss *et al.* (2000) attribute this discrepancy to differences in macrophage cell lines, macrophage activation, and infection lengths and methods. Our model accounts for some of these factors, including one aspect of macrophage activation (IFN-γ-stimulated MHC class II expression) and one consequence of infection length (inhibition of particular intracellular processes), as well as experimental differences in duration of IFN-γ stimulation and amount of IFN-γ used. In our model, none of these factors accounted for the observed discrepancy in MHC class II mRNA levels.

Hmama *et al.* (1998) and Noss *et al.* (2000) also hypothesized that *Mtb* inhibits either MHC class II peptide loading or antigen processing. Our simulations show that neither of these mechanisms could have accounted for the observed changes in levels of molecules given the experimental protocols that were used. On the timescales of both protocols MHC class II expression is expected to be the limiting factor on antigen presentation as suggested by the half life of MHC class II and our sensitivity analysis. Indeed, in the protocol used by Noss *et al.* (2000), we predict that the high level of MHC

class II expression masks whatever reductions in antigen presentation may result from an inhibition of antigen processing or MHC class II peptide loading.

Because these experimental protocols may have favored the detection of mechanisms targeting MHC class II expression, the actual contribution of mechanisms targeting other processes to the overall ability of *Mtb* to inhibit antigen presentation may not have been accurately assessed. Without experimental evidence to the contrary, the possibility even exists that mechanisms targeting antigen processing and MHC class II peptide loading are incidental to *Mtb* infection and do not significantly affect the ability of macrophages to present antigen in the presence of IFN-γ. While an experiment using an *Mtb* mutant specifically unable to inhibit either intracellular process would quickly answer this question, such a mutant is not yet available to our knowledge.

Therefore, we propose an alternative experimental protocol to determine whether mechanisms targeting intracellular processes besides MHC class II expression actually contribute to the ability of *Mtb* to inhibit antigen presentation (Fig. 2.4A). In this protocol macrophages are infected with *Mtb in vitro* and treated with IFN-γ for varying durations prior to assaying for antigen presentation using model antigen and T cell hybridoma. If mechanisms targeting MHC class II expression are the only means by which *Mtb* inhibits antigen presentation, the difference in the levels of T cell response (e.g., IL-2 production) elicited by uninfected and infected macrophages should increase as the duration of IFN-γ stimulation increases (Fig. 2.4B). On the other hand, if mechanisms targeting other intracellular processes play a significant role in the inhibition of antigen presentation, the difference in T cell response should be apparent even with short durations of IFN-γ stimulation and remain relatively constant as the duration of IFN-γ stimulation increases.

2.4.3 Additional Mechanisms May Target Other Processes Affecting Presentation

As part of our analysis we also identified all of the parameters in our model that strongly correlate with the number of pMHC complexes on the macrophage surface. These parameters represent intracellular processes likely to affect antigen presentation if perturbed and may serve as attractive targets to pathogens that evade immune surveillance such as *Mtb*. Other processes related to MHC class II expression, besides those already considered by previous hypotheses, strongly correlated with surface pMHC levels at long timescales. Recent evidence indicates that one of these processes, CIITA transcription, may be targeted by *Mtb* (Kincaid and Ernst 2003, Pai *et al.* 2003). It would be interesting to test experimentally whether *Mtb* also affects any other candidate process such as IFN-γ receptor-ligand binding.

We found that several intracellular processes also negatively correlated with antigen presentation. In contrast to positively correlated processes such as those in Table 2.1, these processes are expected to inhibit antigen presentation if up-regulated rather than down-regulated. In the presence of low levels of exogenous antigen, one such process is the delivery of antigens (both self and exogenous) and derived peptides to MHC class II-inaccessible lysosomes. Conceivably, an intracellular pathogen such as *Mtb* could decrease the availability of its own antigens by increasing the rate at which this process occurs, though benefit to the pathogen may be somewhat offset by a concurrent decrease in competing self antigens (Chicz *et al.* 1993, Rosloniec *et al.* 1990). Nevertheless, the possibility that some pathogens up-regulate delivery to lysosomes cannot be ruled out since the rate of this process and the concentration of self peptide have not been carefully measured.

Most of the experimental data on which we base our model originates from studies using murine cell lines. Therefore, the dynamics of human macrophages infected with *Mtb* may differ somewhat from those observed in our simulations. However, based

on our sensitivity analysis, we believe that our results are robust and can be generalized to the human host.

2.5 Supporting Information

2.5.1 IFN-γ Receptor-Ligand Binding

IFN-γ binds to receptors on the macrophage surface, initiating a cascade of events leading to an increase in MHC class II expression (Steimle *et al.* 1994). We assume that this increase is the primary mechanism by which IFN-γ facilitates antigen presentation. To represent IFN-γ receptor-ligand binding, we use the general reaction scheme ligand + receptor ↔ complex. Other processes are likely to affect the number of IFN-γ receptor-ligand complexes on the time scales of IFN-γ treatment used experimentally, 20-36 hours (Hmama *et al.* 1998, Noss *et al.* 2000). Celada *et al.* (1984) observed that IFN-γ levels in solution decrease 13% and 83% after 4 h in the absence and presence of macrophages, respectively, indicating that appreciable levels of IFN-γ both degrade in solution and are taken up by macrophages. Therefore, in addition to representing IFN-γ receptor-ligand binding, we also represent degradation of IFN-γ in solution and within the macrophage following uptake (Eqns. 2.1-3).

$$dG/dt = (-k_{on\text{-}IFN\text{-}\gamma} G\, R + k_{off\text{-}IFN\text{-}\gamma} C)\, [n_{cells} / (N_A\, v_{rxn})] - k_{deg\text{-}IFN\text{-}\gamma} G \qquad [2.1]$$

$$dR/dt = -k_{on\text{-}IFN\text{-}\gamma} G\, R + k_{off\text{-}IFN\text{-}\gamma} C + k_{recyc}\, C \qquad [2.2]$$

$$dC/dt = k_{on\text{-}IFN\text{-}\gamma} G\, R - k_{off\text{-}IFN\text{-}\gamma} C - k_{recyc}\, C \qquad [2.3]$$

where G is the molar concentration of IFN-γ in the medium and R and C are the numbers of free IFN-γ receptors and IFN-γ receptor-ligand complexes on the surface of each macrophage, respectively. Values for the parameters n_{cells}, the number of macrophages to which IFN-γ is added, and v_{rxn}, the volume of the medium containing both IFN-γ and macrophages, depend on the protocol being simulated, and N_A is Avogadro's number. Values for $k_{on\text{-}IFN\text{-}\gamma}$ and $k_{off\text{-}IFN\text{-}\gamma}$, the association and dissociation rate constants of the IFN-

γ receptor-ligand complex, can be found in the literature (Sadir *et al.* 1998), while a value for $k_{deg\text{-}IFN\text{-}\gamma}$, the rate constant for the degradation of IFN-γ in solution, can be derived from the observed decrease in IFN-γ levels when macrophages are not present if first-order decay is assumed (Celada *et al.* 1984). We estimate a value for k_{recyc}, the rate constant for receptor internalization and recycling, to match the observed decrease in IFN-γ levels when macrophages are present (Celada *et al.* 1984), given the experimental conditions of that study. Celada *et al.* (1985) also found that the total number of IFN-γ receptors on the surface of the macrophage, R_{tot}, does not change over time in the presence of IFN-γ. Therefore, we assume that R_{tot} is constant, allowing either Eqn. **2.2** or Eqn. **2.3** to be eliminated when the formula $R_{tot} = R + C$ is used to derive an expression for either R or C. In all of our simulations, we set the initial conditions for R and C to R_{tot} and 0, respectively.

2.5.2 MHC Class II Transcription

The formation of IFN-γ receptor-ligand complexes on the macrophage surface activates the Jak-Stat signaling pathway, increasing CIITA expression over its basal level (Darnell *et al.* 1994, Steimle *et al.* 1994). Because CIITA expression may be delayed by as much as two hours in response to IFN-γ (Morris *et al.* 2002) and this delay may contribute to the longer delay observed prior to an increase in MHC class II expression (Cullell-Young *et al.* 2001), we represent both CIITA and MHC class II at the mRNA and protein levels explicitly in the model (Eqns. 2.4-6 and 11). To represent transcription and translation, we use the same basic formulation as Maynard Smith (1968). Nascent MHC class II molecules undergo several posttranslational events, including coupling of constituent subunits to invariant chain (Ii), transport through the transGolgi network, and degradation of Ii into class II-associated invariant chain peptide (CLIP) (Hudson and Ploegh 2002). The presence of low levels of mature MHC class II molecules in

unstimulated macrophages suggests that these processes occur constitutively (Hume 1985). We do not distinguish these processes in our model but refer to them collectively as MHC class II protein maturation. In practice, we consider MHC class II protein maturation to be part of MHC class II translation which we represent in our model explicitly. There is evidence that IFN-γ up-regulates MHC class II translation independent of its effect on MHC class II transcription (Cullell-Young *et al.* 2001). Therefore, we represent IFN-γ receptor-ligand complexes as having an effect on both processes (Eqns. 2.4 and 11).

$$dT_1/dt = k_{txn1} (1 + \alpha C / R_{tot}) - k_{deg\text{-}mRNA1} T_1 \quad [2.4]$$

$$dP/dt = k_{tsl1} T_1 - k_{deg\text{-}P} P \quad [2.5]$$

$$dT_2/dt = k_{txn2} P - k_{deg\text{-}mRNA2} T_2 \quad [2.6]$$

where T_1, P, and T_2 are levels of CIITA mRNA, CIITA protein, and MHC class II mRNA per macrophage, respectively. k_{txn1}, k_{tsl1}, and k_{txn2} are rate constants for CIITA transcription, CIITA translation, and MHC class II transcription, whereas $k_{deg\text{-}mRNA1}$, $k_{deg\text{-}P}$, and $k_{deg\text{-}mRNA2}$ are rate constants for degradation of CIITA mRNA, CIITA protein, and MHC class II mRNA, respectively. We assume that unstimulated macrophages possess steady-state levels of CIITA mRNA, CIITA protein, and MHC class II mRNA (i.e., $T_1' = 0$ h^{-1}, $T_1 = T_{1,0}$, $P' = 0$ h^{-1}, $P = P_0$, $T_2' = 0$, and $T_2 = T_{2,0}$ when $C = 0$ mol L^{-1}), allowing values for k_{txn1}, k_{tsl1}, and k_{txn2} to be estimated from known values of $k_{deg\text{-}mRNA1}$, $k_{deg\text{-}P}$, and $k_{deg\text{-}mRNA2}$. The quantity C/R_{tot} represents the fraction of surface IFN-γ receptors occupied at any given time, while α is a scaling factor for CIITA transcription. We assume that the rate of CIITA transcription increases linearly with the fraction of occupied IFN-γ receptors based on receptor occupation theory (Kenakin 1987) and observed correlation between IFN-γ receptor occupancy and tumoricidal activity in macrophages (Celada and Schreiber 1987). The scaling factor α allows MHC class II mRNA levels in the model to match increases observed experimentally in response to IFN-γ. To find a value for α, we simulate the experimental conditions used in two studies (Cullell-Young *et al.* 2001, Pai

et al. 2002) and approximate values for α that most closely yield the observed maximal increases in MHC class II mRNA.

2.5.3 Exogenous Antigens

Exogenous antigens generally enter macrophage endosomes by pinocytosis, phagocytosis, or receptor-mediated endocytosis. Because soluble model antigens such as hen egg lysozyme are typically used to assess the ability of macrophages to present antigen *in vitro* (compare Gercken *et al.* 1994, Hmama *et al.* 1998, Noss *et al.* 2000), we represent only pinocytosis in the model (Eqns. 2.7 and 8). We assume that endocytosed antigens either undergo partial degradation resulting in the production of MHC class II-binding peptides, i.e., antigen processing, or are transported to lysosomes and degraded. Although a small number of exogenous antigens may also be shunted to the MHC class I pathway (Yewdell *et al.* 1999), we do not consider the loss of antigen due to this pathway in the current model. We assume that peptides resulting from antigen processing then either bind MHC class II molecules or are transported to lysosomes and degraded (Eqn. 2.9). The portions of our model representing antigen processing as well as peptide-MHC class II binding are similar to those used in a simpler model by Singer and Linderman (Singer and Linderman 1990).

$$dA^*/dt = -(k_{pino}\, n_{cells} / v_{rxn})\, A^* - k_{deg\text{-}A^*}\, A^* \qquad [2.7]$$

$$dA/dt = (k_{pino} / v_{MIIC})\, A^* - k_{deg\text{-}A}\, A - k_{lys}\, A \qquad [2.8]$$

$$dE/dt = k_{deg\text{-}A}\, A - k_{on\text{-}MHC}\, M\, E + k_{off\text{-}MHC}\, M_e - k_{lys}\, E \qquad [2.9]$$

where A^*, A, and E are molar concentrations of native antigen in the medium, native antigen in the endosomal compartments of each macrophage, and antigen-derived peptide in the endosomal compartments of each macrophage, respectively. Values for the average rate of pinocytic uptake, k_{pino}, and the total volume of the MHC class II-accessible endosomal compartments, v_{MIIC}, can be found in the literature (Dean 1979, Marsh *et al.*

1986). The parameters n_{cells} and v_{rxn} are the same as those found in Eqn. **2.1**. We assume that the rate constant for the degradation of native antigen in the medium, $k_{deg\text{-}A^*}$, has the same value as the rate constant for the degradation of IFN-γ in solution, $k_{deg\text{-}IFN\text{-}\gamma}$. The rate constant for antigen processing, $k_{deg\text{-}A}$, represents what is likely a group of reactions, including the unfolding of native antigen and proteolytic degradation by one or more cathepsin proteases. We derive a value for $k_{deg\text{-}A}$ based on the length of time required for macrophages to degrade 50% of internalized mannosylated BSA (Diment and Stahl 1985), assuming that processing of most antigens yields only one peptide capable of binding MHC class II. We also assume that all soluble materials in the endosomal lumen are delivered to MHC class II-inaccessible lysosomes with the same kinetics and that therefore a single rate constant for this process, k_{lys}, is sufficient. We derive a value for k_{lys} based on the length of time required for receptor degradation (Lauffenburger *et al.* 1987). The rate constants $k_{on\text{-}MHC}$ and $k_{off\text{-}MHC}$ represent association and dissociation of pMHC complexes, respectively, and are described in more detail below.

2.5.4 Self Peptides

Macrophages constitutively produce a population of self-peptides capable of binding MHC class II molecules within endosomes (Chicz *et al.* 1993). In the absence of exogenous antigens, these peptides may bind 80% or more of available MHC class II molecules (Chicz *et al.* 1993). MHC class II-binding self-peptides are derived predominantly from transmembrane proteins including several MHC-related proteins (Chicz *et al.* 1992). In our model we consider both MHC-derived and non-MHC-derived self-peptides as a single population (Eqn. 2.10). We treat self-peptides similarly to peptides derived from exogenous antigen and assume that they either bind MHC class II molecules or are transported to lysosomes and degraded. Our treatment of self-peptides is similar to that used in a previous model by Singer and Linderman (1991).

$$dS/dt = k_{source} + [k_{deg\text{-}MHC} (M_s + M_s^*) - k_{on\text{-}MHC} M S + k_{off\text{-}MHC} M_s] [1 / (N_A v_{MIIC})] - k_{lys} S \quad [2.10]$$

where S is the molar concentration of self-peptides within macrophage endosomes and M, M_s, and M_s^* are the numbers of free intracellular MHC class II molecules, intracellular self-peptide-MHC class II complexes, and surface self-peptide-MHC class II complexes per macrophage, respectively. We assume that the rate of self-peptide synthesis, k_{source}, for which we did not find a value in the literature, is equal to the rate of self-peptide degradation in resting macrophages, $k_{lys} S_0$. An additional source term, $k_{deg\text{-}MHC} M_s [1 / (N_A v_{MIIC})]$, is used to represent the replenishment of MHC-derived self-peptides that are ultimately lost when pMHC complexes are degraded. For the initial value of the endosomal self-peptide concentration, S_0, we use the steady-state value which we did not find in the literature but approximate to be 6×10^{-4} mol L^{-1} by solving Eqn. 2.10 when $S' = 0$ mol L^{-1} h^{-1}, $M_s = M_{s,0}$, and $M = M_0$. During simulations of hypothesis H_3 (i.e. when the value of $k_{on\text{-}MHC}$ was changed) the values of S_0 and k_{source} were recalculated accordingly. However, during simulations to determine PRCC values, all rate constants were changed independently and the values of S_0 and k_{source} were not recalculated.

2.5.5 MHC Class II Translation and Peptide-MHC Class II Binding

We assume that the reaction scheme peptide + MHC ↔ peptide-MHC complex is accurate on the timescales of most *in vitro* experimental protocols allowing us to forego more complicated representations (e.g., those in Beeson and McConnell 1995). We also assume that the enzyme HLA-DM is expressed at sufficiently high levels within endosomes so that dissociation of CLIP from MHC class II is not rate limiting and does not require explicit representation. In addition, because the signal sequence that localizes MHC class II to endosomes is found in the cytoplasmic domain of Ii and removed from mature forms of MHC class II, we assume that all forms of MHC class II in our model

are free to be transported to and from the plasma membrane. Consistent with this assumption, peptide-free MHC class II molecules have been detected on the surface of antigen-presenting cells (Germain and Hendrix 1991, Santambrogio *et al.* 1999). In our model, we represent MHC class II using six variables to distinguish between intracellular and surface localizations as well as free, self-peptide-bound, and exogenous peptide-bound forms (Eqns. 2.11-16).

$$dM/dt = k_{tsl2}\,(1 + \beta\, C / R_{tot})\, T - k_{on\text{-}MHC}\, M\, S + k_{off\text{-}MHC}\, M_s - k_{on\text{-}MHC}\, M\, E + k_{off\text{-}MHC}\, M_e - k_{out}\, M + k_{in}\, M^* - k_{deg\text{-}MHC}\, M \quad [2.11]$$

$$dM^*/dt = k_{out}\, M - k_{in}\, M^* - k_{deg\text{-}MHC}\, M \quad [2.12]$$

$$dM_s/dt = k_{on\text{-}MHC}\, M\, S - k_{off\text{-}MHC}\, M_s - k_{out}\, M_s + k_{in}\, M_s^* - k_{deg\text{-}MHC}\, M_s \quad [2.13]$$

$$dM_s^*/dt = k_{out}\, M_s - k_{in}\, M_s^* - k_{deg\text{-}MHC}\, M_s^* \quad [2.14]$$

$$dM_e/dt = k_{on\text{-}MHC}\, M\, P - k_{off\text{-}MHC}\, M_e - k_{out}\, M_e + k_{in}\, M_e^* - k_{deg\text{-}MHC}\, M_e \quad [2.15]$$

$$dM_e^*/dt = k_{out}\, M_e - k_{in}\, M_e^* - k_{deg\text{-}MHC}\, M_e^* \quad [2.16]$$

where M, M_s, and M_e are the numbers of free MHC class II proteins, self-peptide-MHC class II complexes, and exogenous peptide-MHC class II complexes within the endosomal compartments of each macrophage, respectively, and M^*, M_s^*, and M_e^* are the numbers of the same MHC class II species on the surface of each macrophage. We did not find a measurement in the literature for the rate constant representing MHC class II translation, k_{tsl2}, but derive a value by assuming that unstimulated macrophages maintain a constant total number of MHC class II proteins in the absence of exogenous antigen [i.e., $(M + M^* + M_s + M_s^*)' = 0$ mol L^{-1} h^{-1} when $G_0 = 0$ mol L^{-1} and $E_0 = 0$ mol L^{-1}]. Therefore, given Eqns. 2.11-14, k_{tsl2} is equal to the combined rates of MHC class II protein degradation, $k_{deg\text{-}MHC}$ $(M_0 + M^*_0 + M_{s,0} + M_s^*_0)$. We also assume that all MHC class II proteins are degraded with the same rate constant, $k_{deg\text{-}MHC}$, whose value we derive from the half-life of MHC class II proteins on the surface of cultured macrophages (Poutsiaka *et al.* 1985). We estimate a value for the translation scaling factor, β, in a manner similar to that used for the transcription scaling factor, α. That is, we simulate the

experimental conditions used by Cullell-Young *et al.* (2001) who observed a maximum increase of 37-fold in the MHC class II protein levels of macrophages incubated with IFN-γ ($n_{cells} = 5 \times 10^5$, $v_{rxn} = 1 \times 10^{-3}$ L, $G_0 = 3 \times 10^2$ units/mL $\approx 2 \times 10^{-9}$ mol L^{-1}) and approximate a value for β that matches this output. We derive a value for the rate constant of MHC class II protein transport from endosomes to the plasma membrane, k_{out}, based on the length of time this process takes in cultured macrophages (between 5 and 15 min, Harding and Geuze 1993), assuming that 50% of the proteins are transported during this time. Cultured macrophages retain approximately one-third of their MHC class II proteins intracellularly (Harding and Unanue 1989), a ratio we define as p_{in}. Based on this ratio we derive a value for the rate constant of MHC class II protein internalization from the plasma membrane, k_{in}, by assuming that $(M^*_0 + M_s^*_0) / M_{tot} = p_{in}$ where $M_{tot} = (M_0 + M^*_0 + M_{s,0} + M_s^*_0)$ and that therefore $(M^*_0 + M_s^*_0) = [(1 - p_{in}) / p_{in}] (M_0 + M_{s,0})$. We solve Eqns. 2.12 and 14 for the steady-state values of M^* and M_s^* which we use as initial conditions, set their sum, $k_{out} (M_0 + M_{s,0}) / (k_{in} + k_{deg\text{-}MHC})$, equal to the expression for $(M^*_0 + M_s^*_0)$ above, and solve for k_{in}.

Values for the rate constants of peptide-MHC class II association and dissociation, $k_{on\text{-}MHC}$ and $k_{off\text{-}MHC}$, vary widely in the literature depending on the particular peptide being used. For example, complexes with peptides derived from OVA and myelin basic protein (MBP) dissociate in solution at rates of 3×10^{-6} s^{-1} and 4×10^{-4} s^{-1}, respectively (Buus *et al.* 1986, Mason and McConnell 1994). By using the formula $t_{95\%} = -\ln(0.05) / [k_d (1 + L_0 / K_D)]$ where $t_{95\%}$ is the time required to reach 95% of equilibrium binding, k_d the dissociation rate constant, L_0 the initial ligand concentration, and K_D the equilibrium dissociation constant (Lauffenburger and Linderman 1993), we estimate $t_{95\%}$ values to be on the order of 100 h and 1 h for OVA and MBP peptides, respectively, when $L_0 = K_D$. Considering that the length of time between the administration of exogenous antigen and the assay for surface pMHC complexes is on the order of 1 h in the experimental protocols of interest (Hmama *et al.* 1998, Noss *et al.* 2000), these values

for $t_{95\%}$ suggest that, at least in some cases, the number of complexes does not reach equilibrium. If this is true, the choice of $k_{on\text{-}MHC}$ and $k_{off\text{-}MHC}$ values would be important to the outcome of the simulations. However, the presence of the enzyme HLA-DM increases the dissociation rate constant of pMHC complexes by 10^4-fold (Weber *et al.* 1996), resulting in $t_{95\%}$ values of 1×10^{-2} and 1×10^{-4} h for OVA and MBP peptides, respectively, when $L_0 = K_D$. In both cases, the number of pMHC complexes is expected to reach equilibrium well before the conclusion of the assay. Therefore, we assume that the choice of peptide-specific $k_{on\text{-}MHC}$ and $k_{off\text{-}MHC}$ values from the literature does not significantly affect the outcome of the simulations due to the enzymatic activity of HLA-DM.

In all simulations, we set the initial conditions for the variables representing the different MHC class II species (i.e., M_0, M^*_0, $M_{s,0}$, $M_s{}^*_0$, $M_{e,0}$, and $M_e{}^*_0$) based on two ratios, p_{in} and p_{bound}, the fractions of all MHC class II that are intracellular and bound to self-peptide, respectively, in unstimulated macrophages when exogenous antigen is not present. We assume that p_{in} and p_{bound} apply to both free and peptide-bound MHC class II, so that $M_0 / (M_0 + M^*_0) = M_{s,0} / (M_{s,0} + M_s{}^*_0) = p_{in}$ and $M_{s,0} / (M_0 + M_{s,0}) = p_{bound}$. We express M^*_0, $M_s{}^*_0$, and $M_{s,0}$ in terms of M_0, the number of free endosomal MHC class II proteins, sum M_0, M^*_0, $M_s{}^*_0$, and $M_{s,0}$ to the known total number of MHC class II molecules in unstimulated macrophages (M_{tot}, Harding and Unanue 1989), and solve for each value.

2.5.6 Inclusion of *Mtb* and Its Inhibitory Effect on Intracellular Processes

We simulate the inhibitory effect of *Mtb* on various intracellular processes by multiplying the corresponding rate constant in the baseline model by the quantity $[1 - B / (K_M + B)]$ where B is the multiplicity of infection (moi, or bacteria-to-macrophage ratio) used *in vitro* and K_M is the moi needed to inhibit a process by 50%. For simplicity we use

a single value for K_M in all of our simulations and derive this value from the data of Noss *et al.* (2000) who found that MHC class II transcription decreases by 20% and 80% when the infectious dose of *Mtb* is 5 and 40, respectively. We fit these data to the function $k_{inf} = k_{uninf}\,[1 - B / (K_M + B)]$ where k_{inf} and k_{uninf} are rate constants for a given process in infected and uninfected macrophages, respectively, resulting in $K_M \approx 18$. To measure the effect on antigen presentation, we calculate $(M_e{*}_{uninf} - M_e{*}_{inf}) / M_e{*}_{uninf}$ where $M_e{*}_{uninf}$ and $M_e{*}_{inf}$ are surface exogenous peptide-MHC class II levels (M_e*) using k_{uninf} and k_{inf}, respectively.

2.5.7 Parameters and Initial Conditions

Model simulations generating the figures and tables in the main text used the following parameters and initial conditions in place of the baseline parameters and initial conditions in Tables 3 and 4.

For Figure 2.2. *A* and *B*, $n_{cells} = 4 \times 10^6$, $v_{rxn} = 8 \times 10^{-3}$ L, $G_0 = 2 \times 10^{-6}$ mol L^{-1}, $A{*}_0 = 0$ mol L^{-1}, Pai *et al.* 2002; α and k_{pino} were set to 200 and 1×10^{-12} L h^{-1}, respectively. *C* and *D*, $n_{cells} = 5 \times 10^5$, $v_{rxn} = 1 \times 10^{-3}$ L, $G_0 \approx 2 \times 10^{-9}$ mol L^{-1}, $A{*}_0 = 0$ mol L^{-1}, Cullell-Young *et al.* (2001); α and k_{pino} were set to 30 and 1×10^{-12} L h^{-1}, respectively. *E* and *F*, $n_{cells} = 5 \times 10^6$, $v_{rxn} = 1 \times 10^{-3}$ L, $G_0 = 0$ mol L^{-1}, $A{*}_0 \approx 1 \times 10^{-8}$ mol L^{-1}, Buus and Werdelin (1986); $\alpha = 30$, $k_{pino} = 1 \times 10^{-12}$ L h^{-1}. *G* and *H*, $n_{cells} = 9 \times 10^4$, $v_{rxn} = 4 \times 10^{-4}$ L, $G_0 = 0$ mol L^{-1} or $G_0 \approx 6 \times 10^{-11}$ mol L^{-1}, $A{*}_0 = 0$ mol L^{-1}, $A{*}_{16} = 2 \times 10^{-6}$ mol L^{-1}, Delvig *et al.* (2002); $\alpha = 30$, $k_{pino} = 1 \times 10^{-12}$ L h^{-1}.

For Table 2.1. $n_{cells} = 1 \times 10^5$, $v_{rxn} = 1 \times 10^{-3}$ L, $G_0 = 1 \times 10^{-9}$ mol L^{-1}, $A{*}_0 = 1 \times 10^{-4}$ mol L^{-1}; $\alpha = 30$, $k_{pino} = 1 \times 10^{-12}$ L h^{-1}. $B = 40$. H_1, H_2, H_3, and H_4 correspond to model parameters $k_{deg\text{-}A}$, k_{tsl2}, $k_{on\text{-}MHC}$, and k_{txn2}, respectively.

For Figure 2.3. *C*, *D*, and *E*, $n_{cells} = 1 \times 10^5$, $v_{rxn} = 1 \times 10^{-3}$ L, $B = 50$, $G_0 = 0$ mol L^{-1}, $A{*}_0 = 0$ mol L^{-1}, and $G_{24} \approx 1.3 \times 10^{-9}$ mol L^{-1} where subscript *n* refers to a condition

at the n^{th} hour of the experiment; α and k_{pino} were set to 30 and 1×10^{-12} L h^{-1}, respectively. *F*, *G*, and *H*, $n_{cells} = 5 \times 10^4$, $v_{rxn} = 3.7 \times 10^{-4}$ L, $B = 40$, $G_0 = 1.3 \times 10^{-10}$ mol L^{-1}, $A^*_0 = 0$ mol L^{-1}, $G_{22} = 0$ mol L^{-1}, $G_{24} = 1.3 \times 10^{-10}$ mol L^{-1}, $G_{46} = 0$ mol L^{-1}, and $A^*_{46} = 2.0 \times 10^{-1}$ mol L^{-1}; α and k_{pino} were set to 30 and 1×10^{-12} L h^{-1}, respectively.

For Table 2.2. $n_{cells} = 1 \times 10^5$, $v_{rxn} = 1 \times 10^{-3}$ L, $G_0 = 1 \times 10^{-9}$ mol L^{-1}, $A^*_0 = 1 \times 10^{-4}$ mol L^{-1}; $\alpha = 30$, $k_{pino} = 1 \times 10^{-12}$ L h^{-1} were used as baseline values. MHC class II export, antigen concentration in medium, antigen uptake, MHC class II protein maturation, and IFN-γ stimulation of MHC class II translation correspond to model parameters k_{out}, A^*_0, k_{pino}, k_{tsl2}, and β, respectively. IFN-γ receptor-ligand binding, IFN-γ concentration in medium, MHC class II transcription, CIITA translation, and CIITA transcription correspond to model parameters $k_{on\text{-}IFN\text{-}\gamma}$, G_0, k_{txn2}, k_{tsl1}, and k_{txn1}, respectively. IFN-γ stimulation of CIITA transcription, IFN-γ degradation in solution, MHC class II degradation, CIITA protein degradation, CIITA mRNA degradation, and IFN-γ receptor-ligand dissociation correspond to model parameters α, $k_{deg\text{-}IFN\text{-}\gamma}$, $k_{deg\text{-}MHC}$, $k_{deg\text{-}P}$, $k_{deg\text{-}mRNA1}$, and $k_{off\text{-}IFN\text{-}\gamma}$, respectively.

For Figure 2.4. $n_{cells} = 1 \times 10^5$, $v_{rxn} = 1 \times 10^{-3}$ L, $B = 40$, $G_0 = 0$ mol L^{-1}, $A^*_0 = 0$ mol L^{-1}, $G_{24} \approx 1.3 \times 10^{-10}$ mol L^{-1}, and $A^*_{t+24} = 1 \times 10^{-9}$ mol L^{-1} where subscript *n* refers to conditions at the n^{th} hour of the experiment and *t* is variable; α and k_{pino} are set to 30 and 1×10^{-12} L h^{-1}, respectively.

Figures

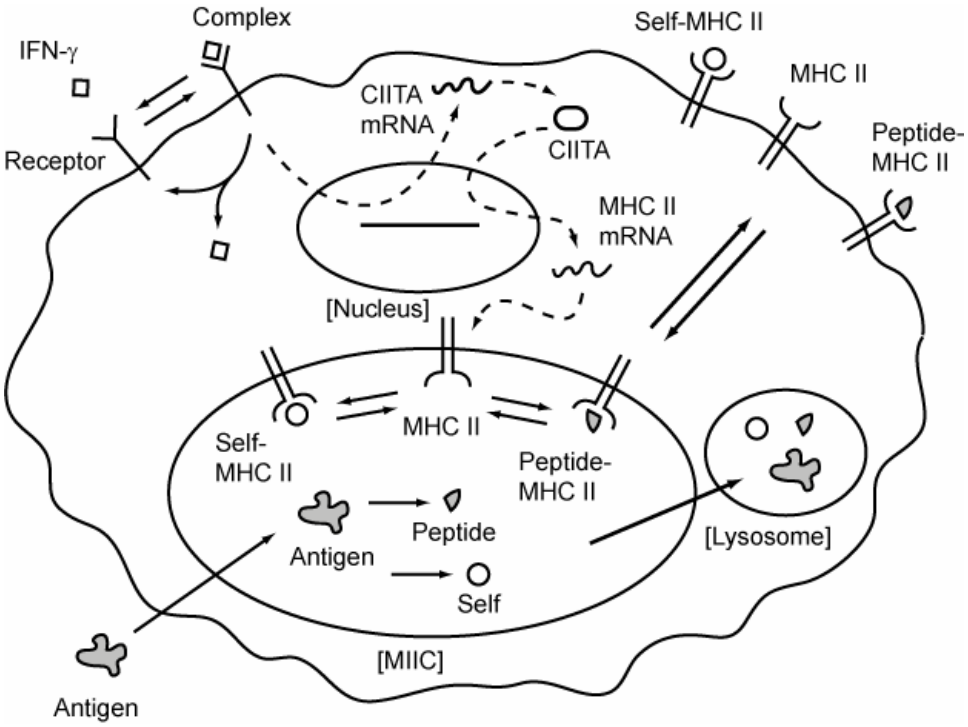

Figure 2.1. Model schematic. Molecular species represented in the model include: extracellular IFN-γ (*G*), IFN-γ receptors (free: *R*, bound: *C*), CIITA (mRNA: T_1, protein: *P*), MHC class II mRNA (T_2), exogenous antigen (extracellular: *A**, intracellular: *A*, peptide: *E*), self peptide (*S*), free MHC class II molecules (intracellular: *M*, surface: *M**), self peptide-bound MHC class II molecules (intracellular: M_s, surface: M_s*), and exogenous peptide-bound MHC class II molecules (intracellular: M_e, surface: M_e*). Solid arrows indicate one-step reactions and dashed arrows indicate regulatory interactions. Degradation is represented in the model for the following molecules but not shown: *G*, T_1, *P*, T_2, *A**, *M*, *M**, M_s, M_s*, M_e, M_e*. Up-regulation of *M* by *C* directly and contribution of M_s and M_s* to *S* are also included in the model but not shown.

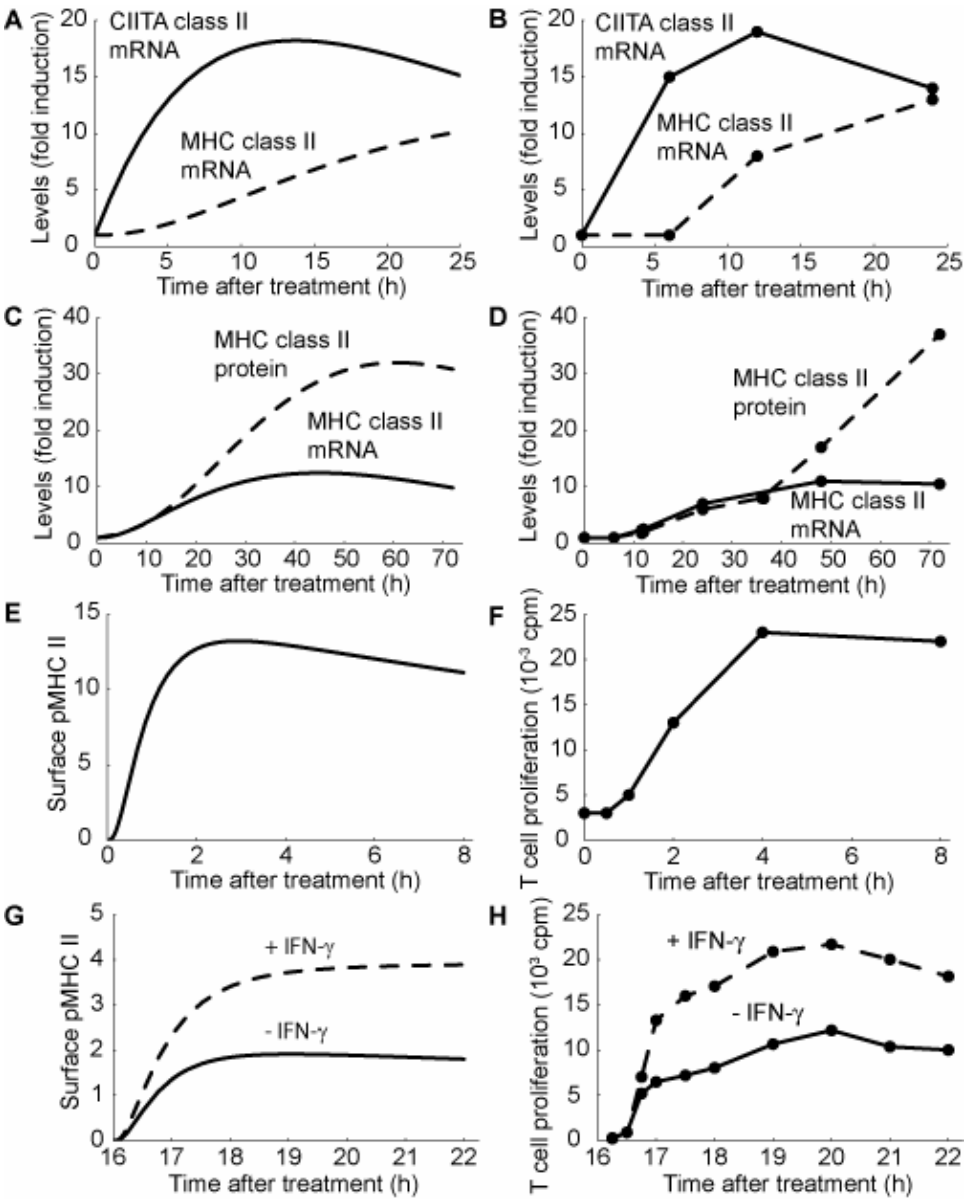

Figure 2.2. Model testing using various controls. *A* and *B*, Simulation results and experimental data for levels of CIITA mRNA (solid lines) and MHC class II mRNA (dashed lines) in IFN-γ-treated macrophages from Pai *et al.* (2002). *C* and *D*, Simulation results and experimental data for levels of MHC class II mRNA (solid lines) and MHC class II protein (dashed lines) in IFN-γ-treated macrophages from Cullell-Young *et al.* (2001). *E* and *F*, Simulation results for surface pMHC levels (in arbitrary units) and experimental data for T cell response in non-IFN-γ-treated macrophages exposed to antigen from Buus and Werdelin (1986). *G* and *H*, Simulation results for surface pMHC levels (in arbitrary units) and experimental data for T cell response in non-IFN-γ-treated macrophages (solid lines) and IFN-γ-treated macrophages (dashed lines) exposed to antigen from Delvig *et al.* (2002). 16 h pretreatment with medium or IFN-γ is not shown; hence, the x-axis is enumerated from 16 h onward (i.e. when antigen is present).

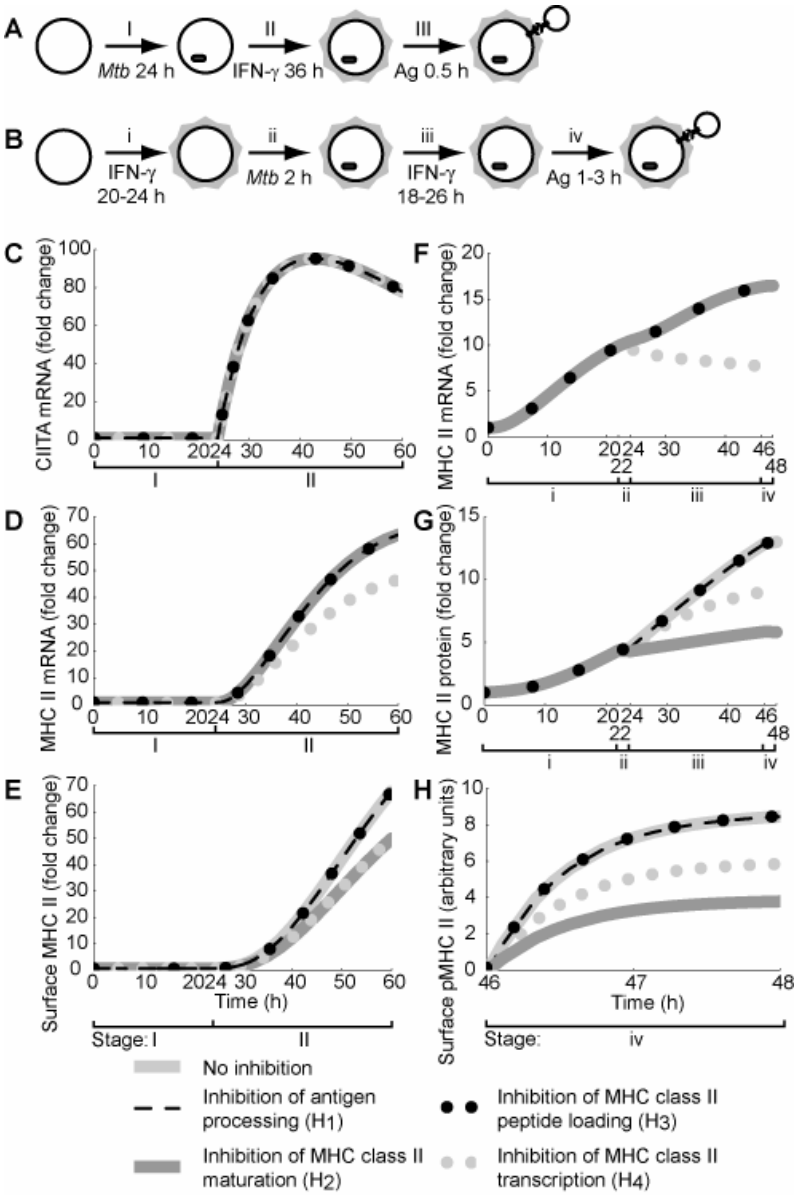

Figure 2.3. Simulation results of two *in vitro* experimental protocols using four published hypotheses. In *A* and *B*, large circles represent macrophages, highlighted circles IFN-γ-treated macrophages, and small circles T cell hybridomas. *A*, Protocol of Hmama *et al.* (1998). 10^5 monocytes were infected with *Mtb* at MOI 50 for 24 h, treated with 200 U/ml IFN-γ for 36 h, pulsed with 1 mg/ml BSA for 0.5 h, and chased for 0.5 h, 1 h, or 4 h. *B*, Protocol of Noss *et al.* (2000). $5 \cdot 10^4$ macrophages were treated with 2 ng/ml IFN-γ for 20-24 h, infected with *Mtb* at MOI 40 for 2 h, treated with 2 ng/ml IFN-γ for an additional 18-26 h, and pulsed with 0-100 μg/ml hen egg lysozyme or 0-1000 μg/ml RNase for 1-3 h. *C*, *D*, and *E*, Simulation results using the protocol of Hmama *et al.* (1998) for levels of CIITA mRNA, MHC class II mRNA, and surface MHC class II protein, respectively. *F*, *G*, and *H*, Simulation results using the protocol of Noss *et al.* (2000) for levels of MHC class II mRNA, total MHC class II protein, and surface pMHC, respectively.

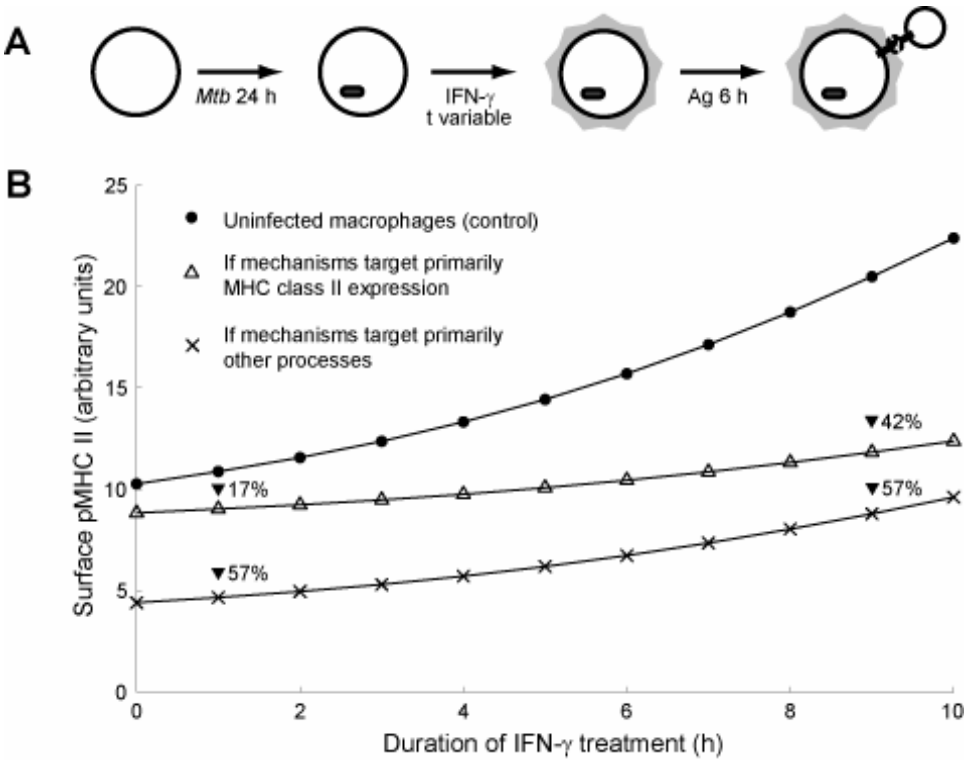

Figure 2.4. Proposed experimental protocol to determine the contribution of different mechanisms to *Mtb* of antigen presentation. *A.* Protocol schematic using representations of Fig. 2.3*A* and *B*. *B*, Surface pMHC levels expected in uninfected macrophages, *Mtb*-infected macrophages if mechanisms target primarily MHC class II expression (in this case, MHC class II transcription), *Mtb*-infected macrophages if mechanisms target primarily other processes (in this case, antigen processing). Percentage reductions in infected macrophages (relative to uninfected controls) are also shown.

Tables

Table 2.1. Changes in surface pMHC levels following inhibition of various intracellular processes hypothesized to be affected by *M. tuberculosis.*[a]

Hypothesis: Affected process	0.1 h	1.0 h	10 h	100 h
H_1: Antigen processing	▼47%	▼8.4%	▼7.2%	▼43%
H_2: MHC class II maturation	▼1.4%	▼8.2%	▼49%	▼69%
H_3: MHC class II peptide loading	▼44%	▼11%	▼12%	▼57%
H_4: MHC class II transcription	▼0.0026%	▼0.16%	▼26%	▼66%
$H_1 + H_4$	▼47%	▼8.6%	▼31%	▼81%
$H_2 + H_3$	▼45%	▼18%	▼55%	▼86%

[a] Identical experimental conditions were used in each simulation, and comparisons were made to the baseline model, i.e. when no processes were inhibited.

Table 2.2. Additional intracellular processes significantly correlated with surface pMHC levels.[a]

Time: 1.0 h	10 h	100 h
Description (Correlation coefficient)	Description (Correlation coefficient)	Description (Correlation coefficient)
MHC class II export (0.79)	*MHC class II protein maturation (0.72)*	*MHC class II transcription (0.57)*
Antigen concentration in medium (0.41)	IFN-γ stimulation of translation[b] (0.62)	*MHC class II protein maturation (0.56)*
Antigen uptake (0.40)	MHC class II export (0.55)	CIITA translation (0.56)
MHC class II protein maturation (0.38)	IFN-γ receptor-ligand binding (0.52)	CIITA transcription (0.53)
IFN-γ stimulation of translation[b] (0.33)	IFN-γ concentration in medium (0.52)	IFN-γ concentration in medium (0.51)
	MHC class II transcription (0.49)	IFN-γ receptor-ligand binding (0.49)
	CIITA translation (0.44)	IFN-γ stimulation of transcription[c] (0.47)
	IFN-γ stimulation of transcription[c] (0.36)	Antigen concentration in medium (0.36)
	CIITA transcription (0.36)	Antigen uptake (0.33)
		MHC class II export (0.32)
		IFN-γ degradation in solution (-0.87)
		MHC class II degradation (-0.56)
		CIITA protein degradation (-0.53)
		CIITA mRNA degradation (-0.49)
		IFN-γ receptor-ligand dissociation (-0.48)

[a] 1000 simulation runs were performed using different sampled parameter values. PRCC values determined to be significant ($p \leq 10^{-30}$) are shown in parentheses. Intracellular processes considered in previous hypotheses (H_1-H_4) are italicized.
[b] MHC class II translation
[c] CIITA transcription

Table 2.3. Parameters used in the APC model.

Parameter	Value (ref.)	Parameter	Value (ref.)
$k_{on\text{-}IFN\text{-}\gamma}$	2.6×10^{10} mol^{-1} liter h^{-1} (Sadir *et al.* 1998)	$k_{deg\text{-}MHC}$	3×10^{-1} h^{-1} (Poutsiaka *et al.* 1985)
n_{cells}	Varies by experiment	v_{MIIC}	4×10^{-16} L (Marsh *et al.* 1986)
v_{rxn}	Varies by experiment	$k_{on\text{-}MHC}$	4×10^{5} mol^{-1} L h^{-1} (Buus *et al.* 1986)
$k_{off\text{-}IFN\text{-}\gamma}$	1.8×10^{1} h^{-1} (Sadir *et al.* 1998)	$k_{off\text{-}MHC}$	6×10^{3} h^{-1} (Buus *et al.* 1986)
$k_{deg\text{-}IFN\text{-}\gamma}$	3.5×10^{-2} h^{-1} (Celada *et al.* 1984)	k_{lys}	6 h^{-1} (Lauffenburger *et al.* 1987)
k_{recyc}	1 h^{-1} (Gercken *et al.* 1994)	k_{pino}	5×10^{-14} - 1×10^{-12} L h^{-1} (Dean 1979, Selby *et al.* 1995)
k_{txn1}	$k_{deg\text{-}mRNA1}\ T_{1,0} = 0.12$ h^{-1}	$k_{deg\text{-}A^*}$	3.5×10^{-2} h^{-1} (Celada *et al.* 1984)
α	30 - 200 (Cullell-Young *et al.* 2001, Pai *et al.* 2002)	$k_{deg\text{-}A}$	4 h^{-1} (Diment and Stahl 1985)
$k_{deg\text{-}mRNA1}$	0.12 h^{-1} (Pai *et al.* 2002)	k_{tsl2}	$k_{deg\text{-}MHC}\ (M_0 + M^*_0 + M_{s,0} + M_s^*{}_0) \approx 2 \times 10^{-2}$ h^{-1}
k_{tsl1}	$k_{deg\text{-}mRNA1}\ P_0 / T_{1,0} = 1.4$ h^{-1}	β	10 (Cullell-Young *et al.* 2001)
$k_{deg\text{-}P}$	1.4 h^{-1} (Schnappauf *et al.* 2003)	k_{out}	4 h^{-1} (Harding and Geuze 1998)
k_{txn2}	$k_{deg\text{-}mRNA2}\ T_{2,0} \approx 4 \times 10^{3}$ h^{-1}	p_{in}	1/3 (Harding and Unanue 1989)
R_{tot}	1×10^{4} (Celada *et al.* 1984)	k_{in}	$[p_{in}/(1\text{-}p_{in})]\ k_{out} - k_{deg\text{-}MHC} \approx 1.97$ h^{-1}
$k_{deg\text{-}mRNA2}$	4×10^{-2} h^{-1} (Cullell-Young	p_{bound}	4/5 (Chicz *et al.* 1993)

	et al. 2001)		
k_{source}	$k_{lys}\ S_0 \approx 3.6 \times 10^{-3}$ mol L^{-1} h^{-1}	M_{tot}	2×10^5 (Harding and Unanue 1989)

Table 2.4. Initial conditions used in the APC model.

Parameter	Value (ref.)	Parameter	Value (ref.)
G	Varies by experiment	A	0 mol L^{-1}
R	1×10^4 (Celada *et al.* 1985)	E	0 mol L^{-1}
C	0	M	$p_{in}\ (1 - p_{bound})\ M_{tot} \approx 1.3 \times 10^4$
T_1	1	M^*	$[(1 - p_{in}) / p_{in}]\ M_0 \approx 2.7 \times 10^4$
P	1	M_s	$[p_{bound} / (1 - p_{bound})]\ M_0 \approx 5.3 \times 10^4$
T_2	1	M_s^*	$[(1 - p_{in}) / p_{in}]\ M_{s,0} \approx 1.1 \times 10^5$
S	6×10^{-4} mol L^{-1}	M_e	0
A^*	Varies by experiment	M_e^*	0

References

Beeson, C. and McConnell, H.M. (1995) Reactions of peptides with class II proteins of the major histocompatibility complex. *J. Am. Chem. Soc.* **117**, 10429-10433

Bekkhoucha, F., Naquet, P., Pierres, A., Marchetto, S. and Pierres, M. (1984) Efficiency of antigen presentation to T cell clones by (B cell X B cell lymphoma) hybridomas correlates quantitatively with cell surface ia antigen expression. *Eur. J. Immunol.*, **14**, 807-814.

Blower, S. M. and Dowlatabadi, H. (1994) Sensitivity and uncertainty analysis of complex models of disease transmission: an HIV model, as an example. *Internat. Stat. Rev.*, **62**, 229-243.

Buus, S., Sette, A., Colon, S. M., Jenis, D.M., and Grey, H.M. (1986) Isolation and characterization of antigen-Ia complexes involved in T cell recognition. *Cell.* **47**, 1071-1077.

Buus, S. and Werdelin, O. (1986) Large, but not small, antigens require time- and temperature-dependent processing in accessory cells before they can be recognized by T cells. *Acta. Pathol. Microbiol. Immunol. Scand. [C]*, **94**, 17-24.

Celada, A., Gray, P.W., Rinderknecht, E. and Schreiber, R.D. (1984) Evidence for a gamma-interferon receptor that regulates macrophage tumoricidal activity. *J. Exp. Med.* **160**, 55-74.

Celada, A., Allen, R., Esparza, I., Gray, P.W. and Schreiber, R.D. (1985) Demonstration and partial characterization of the interferon gamma receptor on human mononuclear phagocytes. *J. Clin. Invest.*, **76**, 2196-2205.

Celada, A. and Schreiber, R.D. (1987) Internalization and degradation of receptor-bound interferon-γ by murine macrophages. Demonstration of receptor recycling. *J. Immunol.* **139**, 147-153.

Chan, E.D., Chan, J. and Schluger, N.W. (2001) What is the role of nitric oxide in murine and human host defense against tuberculosis? Current knowledge. *Am. J. Respir. Cell. Mol. Biol.*, **25**, 606-612.

Chicz, R.M., Urban, R.G., Lane, W.S., Gorga, J.C., Stern, L.J., Vignali, D.A. and Strominger, J.L. (1992) Predominant naturally processed peptides bound to HLA-DR1 are derived from MHC-related molecules and are heterogeneous in size. *Nature.* **358**, 764-768.

Chicz, R.M., Urban, R.G., Gorga, J.C., Vignali, D.A., Lane, W.S. and Strominger, J.L. (1993) Specificity and promiscuity among naturally processed peptides bound to HLA-DR alleles. *J. Exp. Med.*, **178**, 27-47.

Cullell-Young, M., Barrachina, M., Lopez-Lopez, C., Gonalons, E., Lloberas, J., Concepcio, S. and Celada, A. (2001) From transcription to cell surface expression, the induction of MHC class II I-Aα by interferon-γ in macrophages is regulated at different levels. *Immunogen.*, **53**, 136-144.

Darnell, J.E., Kerr, I.M. and Stark, G.R. (1994) Jak-STAT pathways and transcriptional activation in response to IFNs and other extracellular signaling proteins. *Science.* **264**, 1415-1421

Dean, R.T. (1979) Effects of cytochalasin B on the pinocytosis and degradation of proteins by macrophages *Biochem. Soc. Trans.* **7**, 362-364.

Delvig, A.A., Lee, J.J., Chrzanowska-Lightowlers, Z.M.A. and Robinson, J.H. (2002) TGF-β1 and IFN-γ cross-regulate antigen presentation to CD4 T cells by macrophages. *J. Leukoc. Biol.*, **72**, 163-166.

Demotz, S., Grey, H.M. and Sette, A. (1990) The minimal number of class II MHC-antigen complexes needed for T cell activation. *Science,* **249**, 1028-1030.

Diment, S. and Stahl, P. (1985) Macrophage endosomes contain proteases which degrade endocytosed protein ligands. *J. Biol. Chem.* **260**, 15311-15317.

Dunn, P. L. and North, R.J. (1995) Virulence ranking of some *Mycobacterium tuberculosis* and *Mycobacterium bovis* strains according to their ability to multiply in the lungs, induce lung pathology, and cause mortality in mice. *Infect. Immun.*, **63**, 3428-3437.

Fenton, M.J. (1998) Macrophages and tuberculosis. *Curr. Opin. Hematol.*, **5**, 72-78.

Gercken, J., Pryjma, J., Ernst, M. and Flad, H.D. (1994) Defective antigen presentation by *Mycobacterium tuberculosis*-infected monocytes. *Infect. Immun.*, **62**, 3472-3478.

Germain, R.N. and Hendrix, L.R. (1991) MHC class II structure, occupancy and surface expression determined by post-endoplasmic reticulum antigen binding. *Nature.* **353**, 134-139.

Harding, C.V. and Unanue, E.R. (1989) Antigen processing and intracellular Ia. Possible roles of endocytosis and protein synthesis in Ia function. *J. Immunol.* **142**, 12-19

Harding, C.V. and Geuze, H.J. (1993) Immunogenic peptides bind to class II MHC molecules in an early lysosomal compartment. *J. Immunol.* **151**, 3988-3998

Harding, C.V., Ramachandra, L. and Wick, M.J. (2003) Interaction of bacteria with antigen presenting cells: influences on antigen presentation and antibacterial immunity. *Curr. Opin. Immunol.*, **15**, 112-119.

Heldwein, K.A. and Fenton, M.J. (2002) The role of Toll-like receptors in immunity against mycobacterial infection. *Microbes Infect.*, **4**, 937-944.

Helton, J.C. and Davis, F.J. (2001) Sampling-based methods. In Saltelli, A., Chan, K. and Scott, E.M., *Sensitivity Analysis*. Wiley, New York, pp. 101-153.

Hmama, Z., Gabathuler, R., Jefferies, W.A., de Jong, G. and Reiner, N.E. (1998) Attenuation of HLA-DR expression by mononuclear phagocytes infected with *Mycobacterium tuberculosis* is related to intracellular sequestration of immature class II heterodimers. *J. Immunol.*, **161**, 4882-4893.

Hudson, H.W. and Ploegh, H.L. (2002) The cell biology of antigen presentation. *Exp. Cell Res.* **272**, 1-7.

Hume, D.A. (1985) Immunohistochemical analysis of murine mononuclear phagocytes that express class II major histocompatability antigens. *Immunobiol.*, **170**, 381-389.

Huppa, J.B., Gleimer, M., Sumen, C. and Davis, M.M. (2003) Continuous T cell receptor signaling required for synapse maintenance and full effector potential. *Nat. Immunol.*, **4**, 749-755.

Kaufmann, S.H.E. (1999) Cell-mediated immunity: dealing a direct blow to pathogens. *Curr. Biol.*, **9**, R97-R99.

Kaye, P.M., Sims, M. and Feldmann, M. (1986) Regulation of macrophage accessory cell activity by mycobacteria. II. *In vitro* inhibition of Ia expression by *Mycobacerium microti*. *Clin. Exp. Immunol.*, **64**, 28-34.

Kenakin, T.P. (1987) *Pharmacologic analysis of drug-receptor interaction*, Kenakin, T. P. (Raven Press, New York), pp. 1-42.

Kincaid, E.Z. and Ernst, J.D. (2003) Continuous T cell receptor signaling required for synapse maintenance and full effector potential. *Nat. Immunol.*, **171**, 2042-2049.

Lauffenburger, D.A., Linderman, J. and Berkowitz, L. (1987) Analysis of mammalian cell growth factor receptor dynamics. *Ann. N. Y. Acad. Sci.* **506**, 147-162.

Lauffenburger, D.A. and Linderman, J.J. (1993) *Receptors: Models for binding, trafficking, and signaling* (Oxford University Press, New York), pp. 9-72.

Marsh, M., Griffiths, G., Dean, G.E., Mellman, I. and Helenius, A. (1986) Three-dimensional structure of endosomes in BHK-21 cells *Proc. Natl. Acad. Sci. USA.* **83**, 2899-2903.

Mason, K. and McConnell, H.M. (1994) Short-lived complexes between myelin basic protein peptides and IA^k. *Proc. Natl. Acad. Sci. USA.* **91**, 12463-12466.

Maynard Smith, J. (1968). *Mathematical Ideas in Biology*, Maynard Smith, J. (Cambridge University Press, New York), pp. 98-116

Mazzaccaro, R.J., Gedde, M., Jensen, E.R., van Santen, H.M., Ploegh, H.L., Rock, K.L. and Bloom, B.R. (1996) Major histocompatibility class I presentation of soluble antigen facilitated by *Mycobacterium tuberculosis* infection. *Proc. Natl. Acad. Sci. USA.*, **93**, 11786-11791.

Moreno, C., Mehlert, A. and Lamb, J. (1988) The inhibitory effects of mycobacterial lipoarabinomannan and polysaccharides upon polyclonal and monoclonal human T cell proliferation. *Clin. Exp. Immunol.,* **74**, 206-210.

Morris, A.C., Beresford, G.W., Mooney, M.R. and Boss, J.M. (2002) Kinetics of a gamma interferon response: expression and assembly of CIITA promoter IV and inhibition by methylation. *Mol. Cell. Biol.* **22**, 4781-4791

Mshana, R.N., Hastings, R.B. and Krahenbuhl, J.L. (1988) Infection with live mycobacteria inhibits *in vitro* detection of Ia antigen on macrophages. *Immunobiol.*, **177**, 40-54.

Noss, E.H., Harding, C.V. and Boom, W.H. (2000) *Mycobacterium tuberculosis* inhibits MHC class II antigen processing in murine bone marrow macrophages. *Cell. Immunol.*, **201**, 63-74.

Pai, R.K., Askew, D., Boom, W.H. and Harding, C.V. (2002) Regulation of class II MHC expression in APCs: roles of types I, III, and IV class II transactivator. *J. Immunol.*, **169**, 1326-1333.

Pai, R.K., Convery, M., Hamilton, T.A., Boom, W.H. and Harding, C.V. (2003) Inhibition of IFN-γ-induced class II transactivator expression by a 19-kDa lipoprotein from *Mycobacterium tuberculosis*: a potential mechanism for immune evasion. *J. Immunol.*, **171**, 175-184.

Pancholi, P., Mirza, A., Bhardwaj, N. and Steinman, R.M. (1993) Sequestration from immune $CD4^+$ T cells of mycobacteria growing in human macrophages. *Science*, **260**, 984-986.

Pinet, V.M. and Long, E.O. (1998) Peptide loading onto recycling HLA-DR molecules occurs in early endosomes. *Eur. J. Immunol.*, **28**, 799-804.

Poutsiaka, D.D., Yogeeswaran, G., Taylor, D.D. and Black, P.H. (1985) Turnover and fate of I-Ak antigen on the murine macrophage cell surface. *Biochem. Biophys. Res. Commun.* **131**, 216-221

Reske-Kunz, A.B., Spaeth, E., Reske, K., Lohmann-Matthes, M.L. and Rude, E. (1981) Induction of anamnestic T cell proliferation by antigen-pulsed, bone marrow-derived macrophages. *Eur. J. Immunol.*, **11**, 745-750.

Rosloniec, E.F., Vitez, L.J., Buus, S. and Freed, J.H. (1990) MHC class II-derived peptides can bind to class II molecules, including self molecules, and prevent antigen presentation. *J. Exp. Med.*, **171**, 1419-1430.

Sadir, R., Forest, E. and Lortat-Jacob, H. (1998) The heparin sulfate binding sequence of interferon-γ increased the on rate of the interferon-γ-interferon-γ receptor complex formation. *J. Biol. Chem.* **273**, 10919-10925.

Santambrogio, L.K., Fischer, F.R., Dorf, M.E. and Stern, L.J. (1999) Abundant empty class II MHC molecules on the surface of immature dendritic cells. *Proc. Natl. Acad. Sci. USA.* **96**, 15050-15055.

Schnappauf, F., Hake, S.B., Carvajal, M.M., Bontron, S., Lisowska-Grospierre, B. and Steimle, V. (2003) N-terminal destruction signals lead to rapid degradation of the major histocompatibility complex class II transactivator CIITA *Eur. J. Immunol.* **33**, 2337-2347.

Selby, D.M., Singer, D.F., Anderson, R.W., Coligan, J.E., Linderman, J.J. and Nairn, R. (1995) Antigen-presenting cell lines internalize peptide antigens via fluid-phase endocytosis. *Cell Immunol.* **163**, 47-54.

Singer, D.F. and Linderman, J.J. (1990) The relationship between antigen concentration, antigen internalization, and antigenic complexes: modeling insights into antigen processing and presentation. *J. Cell. Biol.* **111**, 55-68.

Singer, D.F. and Linderman, J.J. (1991) Antigen processing and presentation: how can a foreign antigen be recognized in a sea of self proteins? *J. Theor. Biol.* **151**, 385-404.

Steimle, V., Siegrist, C., Mottet, A., Lisowska-Grospierre, B. and Mach, B. (1994) Regulation of MHC class II expression by interferon-gamma mediated by the transactivator gene CIITA. *Science.* **265**, 106-109.

Weber, D.A., Evavold, B.D. and Jensen, P.E. (1996) Enhanced dissociation of HLA-DR-bound peptides in the presence of HLA-DM. *Science.* **274**, 618-620.

Weiss, A., Shields, R., Newton, M., Manger, B. and Imboden, J. (1987) Ligand-receptor interactions required for commitment to the activation of the interleukin 2 gene. *J. Immunol.*, **138**, 2169-2176.

Yewdell, J.W., Norbury, C.C. and Bennink, J.R. (1999) Mechanisms of exogenous antigen presentation by MHC class I molecules *in vitro* and *in vivo*: implications for generating CD8+ T cell responses to infectious agents, tumors, transplants, and vaccines. *Adv. Immunol.* **73**, 1-77.

Ziegler, K. and Unanue, E.R. (1981) Identification of a macrophage antigen-processing event required for I-region-restricted antigen presentation to T lymphocytes. *J. Immunol.*, **127**, 1869-1875.

CHAPTER 3

How Peptide Length Affects Binding to MHC Class II

3.1 Introduction

Major histocompatibility complex (MHC) molecules, also known as human leukocyte antigens (HLA), are a vital component to the development of the immune response to pathogens (Kaufmann 2005). These molecules act as receptors for peptides derived from foreign antigens as well as self peptides and enable the long-term display of antigens on the cell surface. T cells recognize antigenic peptides in the context of MHC, and depending on the class of MHC involved, recognition can lead to the death of the presenting cell or its activation. In either case peptide-MHC binding is an important prerequisite event and has far-reaching consequences to the ensuing response.

Prediction of peptide-MHC binding therefore represents an important goal in bioinformatics, particularly as applied to immunology, and a number of computational approaches have been developed (reviewed in Buus 1999; see also Robinson *et al.* 2003 for other MHC-specific bioinformatics tools). The simplest are based on motifs, i.e. requirements for particular amino acids at positions within the peptide as determined from pool sequencing of eluted peptides (Falk *et al.* 1991, Rammensee 1995 and references therein). Such approaches have largely been superseded by algorithms using matrices to score the relative contribution of amino acids at each position within the peptide (Parker *et al.* 1994, Davenport *et al.* 1995, Marshall *et al.* 1995). Machine learning methods including hidden Markov models and artificial neural networks have

also been applied, with peptide sequence serving as input and binding/non-binding as output (Brusic and Harrison 1994, Honeyman *et al.* 1998, Mamitsuka 1998). More recently, attempts have been made to predict the structure of the peptide-MHC complex and free energy changes associated with binding (Altuvia *et al.* 1997, Rognan *et al.* 1999, Schueler-Furman *et al.* 2000, Davies *et al.* 2003, Schafroth and Floudas 2004; for a review of current structural information and nomenclature see Kaas and Lefranc 2005). It is also possible to combine some of these approaches, as Sturniolo *et al.* (1999) did using matrices to represent each pocket lining the peptide-binding groove.

Continued progress in the development of these algorithms faces a number of challenges including how to handle differences between the two classes of MHC. Most prediction algorithms were first developed in the context of peptide-MHC class I binding which involves peptides of a narrow range of lengths, usually 8-10 amino acids. These algorithms were then applied to peptide-MHC class II binding, typically with little or no modification.

Despite the fact that both classes of MHC share superficial similarities and bind a core of nine amino acids within peptides (Jones 1997), important differences exist. In particular the open-ended nature of MHC class II peptide-binding groove allows for a wide range of peptide lengths (Brown *et al.* 1993). Peptides binding MHC class II usually vary between 13 and 17 amino acids in length, though shorter or longer lengths are not uncommon (Chicz *et al.* 1992, Sercarz and Maverakis 2003). As a result peptides are hypothesized to shift within the MHC class II peptide-binding groove, changing which 9mer window (register) sits directly within the groove at any given time. In contrast the capped nature of the MHC class I peptide-binding groove does not allow variation in length or such register shifting.

Variation in peptide length may have important consequences for the binding and function of antigenic peptides (Malcherek *et al.* 1994, Vogt *et al.* 1994). For instance, Srinivasan *et al.* (1993) found that a 23mer peptide derived from cytochrome c was 32

times more immunogenic than a 10mer peptide containing the same putative binding core. A direct relationship between peptide length and binding affinity has been observed for some MHC class II alleles, but whether this holds true for most alleles remains unknown, as does an explanation for why this relationship exists (Bartnes *et al.* 1999, Fleckenstein *et al.* 1999, Arnold *et al.* 2002, Sercarz and Maverakis 2003). In addition to having more binding registers, longer peptides also possess peptide-flanking residues (PFR) which lie outside of the peptide-binding groove and may interact with the MHC class II molecule at more distal locations (Sercarz and Maverakis 2003). Whether information regarding peptide length, or any other peptide property lost by considering only 9mers, may aid prediction also remains unknown.

In this study we address several issues related to peptide length and binding to MHC class II. Using aggregate data that are now available from online databases, we first examine whether a relationship exists between length and affinity for several MHC class II alleles. We then attempt to incorporate length into two existing binding algorithms in a number of ways, including using regression to pre-process the data, treating length as an additional variable within the algorithms, and deriving a formula to more accurately represent register shifting (Fig. 3.1). We show that improvements to more than one current algorithm for predicting peptide-MHC class II binding are possible with relatively simple amendments. We also comment on which mechanisms are likely to be affecting binding as peptide length increases.

3.2 Methods

3.2.1 Data Sources

Peptide data sets used in this study are available from the AntiJen database (http://www.jenner.ac.uk, Blythe *et al.* 2002) and can be downloaded using the perl `LWP::Simple` module. Other peptide-MHC databases listing affinities are also

available, including the Immune Epitope Database (currently in beta version at http://www.immuneepitope.org, Peters *et al.* 2005), but were not used in this study. Our data sets comprised the sequences and IC_{50} values of peptides binding the MHC class II alleles HLA-DRB1*0101, -DRB1*0401, and -DRB1*1501 from AntiJen. IC_{50} refers to the concentration of peptide required to inhibit 50% of reporter peptide-MHC binding. When more than one IC_{50} measurement was available for a given peptide-MHC complex, the first measurement listed was used, unless otherwise indicated. IC_{50} values were converted into pIC_{50} using the formula $pIC_{50} = -\log IC_{50}$ where IC_{50} has units of molar. Homologous sequences and their IC_{50} measurements were removed using UniqueProt (Mika and Rost 2003). Other algorithms for removing homologous sequences are also available, including Hobohm 1 and Hobohm 2 (Hobohm *et al.* 1992), but were not used in this study. The data sets were of the following sizes (before/after filtering by UniqueProt): DRB1*0101 (464/303), DRB1*0401 (606/414), DRB1*1501 (343/213). Two additional data sets were used to assess the effect of data set size, those for DRB1*0404 (81/54) and DRB1*0405 (116/102). To assess the role of data quality in determining algorithm performance, data sets for DRB1*0101, DRB1*0301, DRB1*0401, DRB1*1101, DRB1*1501, and A*0201 were obtained from AntiJen, and data points for which the concentration of reporter peptide was unavailable were excluded. Data sets are available as part of the online Supplementary Data at http://malthus.micro.med.umich.edu/Bioinformatics/.

3.2.2 Regression of Binding Affinity Versus Peptide Length

Both parametric and nonparametric fits were made to plots of affinity vs. length in the data. Parametric fits were made with one, two, and three fitted parameters (linear, quadratic, and cubic, respectively) using the open-source statistical program R (http://www.R-project.org, R Development Core Team 2005) and the function `lm`.

Nonparametric local regression fits were made using the R function `loess` with default settings (Cleveland and Devlin 1988). To evaluate fit quality, analysis of variance was performed using the R function `anova`. An F statistic was generated which we used to compare linear with nonlinear parametric fits (Motulsky and Christopoulos 2004).

Nonparametric local regression fits were evaluated using a permutation test. In this test each pIC_{50} value was reassigned to a different peptide sequence at random, and a loess fit was re-derived for the shuffled values. This was repeated 1000 times, and the smallest 25 (2.5%) and largest 25 (2.5%) fitted values at each length were excluded. The local regression fit to the original, non-shuffled data set was then compared to the remaining 95% of permuted values at each length and was determined to be significant if it fell outside of this interval.

3.2.3 Simulations of Register Shifting

To simulate the effects of register shifting on peptide-MHC class II binding affinity over a range of peptide lengths, we derived a formula for the expected value of the affinity of a single hypothetical peptide with multiple registers:

$$E[K(X)] = \sum K(x_i)\, p(x_i) \qquad [3.1]$$

where $K(X)$ is the equilibrium association constant, or affinity, of a peptide X, $K(x_i)$ is the affinity of a complex with a single register x_i, and $p(x_i)$ is the probability of register x_i occurring. We assume that $p(x_i)$ can be approximated by the proportion of complexes having register x_i:

$$p(x_i) = N(x_i) / \sum N(x_i) \qquad [3.2]$$

where $N(x_i)$ denotes the number of complexes having register x_i and the sum is taken over all possible registers. Belmares and McConnell (2001) found that the kinetics of shifting between two registers could be accurately represented as $x_1 \leftrightarrow P + M \leftrightarrow x_2$ where P and

M are peptide and MHC, respectively. Based on this result, at equilibrium $[x_1] = K(x_1)[P][M]$ and $[x_2] = K(x_2)[P][M]$. Because both x_1 and x_2 exist in the same solution, it follows that:

$$N(x_1) / [\, N(x_1) + N(x_2) \,] = K(x_1) / [\, K(x_1) + K(x_2) \,]. \quad [3.3]$$

More generally,

$$N(x_i) / \sum N(x_i) = K(x_i) / \sum K(x_i). \quad [3.4]$$

Combining Equations 3.1, 3.2, and 3.4, we obtain the following result for the expected affinity of a given complex when multiple registers are available:

$$E[K(X)] = \sum K(x_i)^2 / \sum K(x_i) \quad [3.5]$$

This result can also be applied to log-transformed measures of affinity such as log $K(X)$. Henceforth we refer to Equation 3.5 or its log-transformed counterpart as the equilibrium-based formula for reconciling multiple registers.

We assume that every overlapping 9mer window within a peptide can result in binding to MHC and therefore set the lower and upper limits of summation at 1 and $l - 8$, respectively, where l represents peptide length and is varied between 9 and 25, the shortest and longest lengths typically observed in our data sets. $K(x_i)$ was generated from a lognormal distribution with mean $10^{7.5}$ and standard deviation $10^{0.5}$, based on the observation that most values for the equilibrium dissociation constant K_D of peptide-MHC binding fall in the range of 10^{-7}-10^{-8} M (McFarland and Beeson 2002). Moreover, a lognormal distribution was chosen based on the equation for standard free energy change, $\Delta G° = -RT \ln (1/K_D)$ where R and T are the gas constant and temperature, respectively (Eisenberg and Crothers 1979), and the assumption that free energy change for peptide-MHC binding is normally distributed. For each value of l between 9 and 25, a set number of values were generated (in our case, either 10 or 100), resulting in a scatter plot of

simulated pIC_{50} values versus length. A curve was then fit to this plot using local regression (the `loess` function in R) with default settings.

3.2.4 Peptide-MHC Binding Affinity Prediction

Two algorithms were selected to generate baseline predictions against which the effects of modifications based on length could be compared. One of these algorithms was the iterative self-consistent (ISC) partial-least-squares (PLS) algorithm of Doytchinova and Flower (2003). We implemented this matrix-based algorithm for predicting peptide-MHC binding affinity in perl and R. Briefly, this algorithm uses partial least squares regression to identify underlying factors (also known as latent variables) relating multiple predictor variables to an outcome variable. In the case of peptide-MHC binding, 180 predictor variables were used to denote the presence or absence of the 20 possible amino acids within each 9mer window, and the outcome variable was binding affinity as pIC_{50}.

The initial steps of the algorithm were performed using perl scripts: splitting each data set into training and test sets; generating all possible 9mers for each training set peptide; selecting only those 9mers having position 1 anchor residues (F, I, L, M, V, W, and Y); and converting 9mers thus selected into bit strings. PLS regression was then performed in R using the bit-encoded 9mers and their corresponding pIC_{50} values. PLS is available for R as the `pls.pcr` library (available at http://cran.r-project.org) and was called from within a perl script using the `IPC::Open2` module. Default settings were used for PLS; however, some options in the commercial software used by Doytchinova and Flower (2003) were not available in R, namely scaling method and column filtering. Subsequent steps in the algorithm were performed using additional perl scripts: selecting those 9mers in the training set yielding predicted pIC_{50} values closest to experimental pIC_{50} values during cross-validation and repeating the algorithm until the selected set of 9mers matched the previously selected set, i.e. when self-consistency was achieved. For

computational expediency we limited the number of PLS iterations for any given peptide to 10. At that point the final PLS model was extracted and used to generate predictions on the test set.

For test set peptides having more than one 9mer with an anchor residue in position 1, multiple predictions were generated and a rule was needed to make a final prediction. One option is to assume only one register predominates and to take the highest score from among the predictions. More complicated rules are also possible such as the combination rule of Doytchinova and Flower (2003) whereby the mean of the pIC_{50} predictions is chosen if they fall within a one log range; otherwise, the highest is chosen.

To measure the performance of the algorithm we used five-fold cross-validation (5x-CV), setting aside one-fifth of each data set to use as a test set and using the other four-fifths as the training set. This process was repeated on the same data set four additional times until a prediction was made for each peptide in the data set and complete coverage was achieved. (This instance of cross-validation was independent of the leave-one-out-cross validation used in the ISC-PLS algorithm.) The accuracy of each set of predictions was scored by calculating the area under receiver operating characteristic curve (A_{ROC}). This calculation can be done in R using the `prediction` and `performance` functions of the `ROCR` library. By repeating each 5x-CV multiple times, we were able to calculate the standard error of the A_{ROC} scores which could then be used to determine whether two mean A_{ROC} scores significantly differed by Student's *t* test. Pearson correlation coefficients between predicted and experimentally determined pIC_{50} values were also used to score performance and are provided in the online Supplementary Data (Lund *et al.* 2005).

A second algorithm that was selected was the TEPITOPE algorithm of Sturniolo *et al.* (1999). In this algorithm amino acid-binding profiles are generated for each pocket within the peptide-binding groove, and these profiles are combined according to MHC

sequence. We did not regenerate these matrices but rather used the matrices available on the ProPred website (http://www.imtech.res.in/raghava/propred, Singh and Raghava 2001). Using the appropriate matrix a sum was calculated for each peptide in a selected AntiJen data set. To this value we added an approximation of the binding affinity of an all-alanine 9mer (pIC_{50} = 6.169, Doytchinova and Flower 2003) generating a final prediction. Performance was scored by calculating the A_{ROC}.

3.2.5 Incorporating Length into Existing Prediction Algorithm

Peptide length was incorporated into the ISC-PLS algorithm using one of three modifications. In Modification 1 (Mod. 1) a local regression fit was first made to the peptide lengths and pIC_{50} measurements in each training set. (In the event that the pIC_{50} value for either the shortest or the longest length peptide was excluded from the training set but included in the test set, a local regression fit at that length could not be generated; instead, we assigned the average fitted values at the remaining lengths.) The value of the fit was then subtracted from the original pIC_{50} value for each peptide, and the resulting difference, i.e. the residual, was then used in place of the original pIC_{50} value. The ISC-PLS algorithm was performed as described earlier providing initial predictions on the test set. To these predictions the value of the regression fit was added yielding final predictions. Alternatively, in Alternative Modification 1 (Alt. 1), peptide length was appended as the 181st predictor variable to the bit-encoded training set and test set 9mers. The remainder of the algorithm was then performed as described earlier. Finally, in Modification 2 (Mod. 2) the formula derived to represent register shifting (Equation 5) was used to reconcile predictions made on multiple candidate 9mers, i.e. registers, within a test set peptide. This modification occurred at the last stage of the ISC-PLS algorithm and was used in place of the combination rule described above.

Only Mod. 2 was used to incorporate length into the TEPITOPE/ProPred algorithm. When TEPITOPE/ProPred is applied to peptides with multiple registers, the highest score among the different registers is typically taken to be the score of the entire peptide (Brusic *et al.* 1998, Nielsen *et al.* 2004, Murugan and Dai 2005). We reconciled individual register scores using the equilibrium-based formula (Equation 5) but did not regenerate the pocket profiles and therefore did not apply Mod. 1 or Alt. 1 in this case.

3.2.6 Data Filtering by Experimental Parameters

To gauge the sensitivity of binding affinity measurements to variations in experimental parameters, we used the Cheng-Prusoff equation (1973) which relates the equilibrium dissociation constant of a peptide-MHC complex (K_D) to its observed IC_{50} value:

$$K_D = IC_{50} (1 + L_r/K_r)^{-1} \quad [3.6]$$

Here L_r and K_r represent the concentration of the reporter peptide and the equilibrium dissociation constant of the reporter peptide-MHC complex, respectively. These parameters frequently vary by protocol, L_r explicitly so and K_r by virtue of being specific to each combination of peptide and MHC (Sette 1989, Roche 1990, Southwood 1998). Other experimental parameters are also likely to affect IC_{50} measurements, including temperature and pH, but were not considered explicitly.

As a measure of peptide-MHC binding affinity, K_D has the benefit of not being dependent on the identity or concentration of the reporter peptide used (Kenakin 1997). Additionally, K_D is directly proportional to the change in standard free energy of a reaction $\Delta G°$ when log transformed as indicated by the equation $\Delta G° = -RT \ln (1/K_D)$ where R and T are the gas constant and temperature, respectively (Eisenberg 1979).

Because the Cheng-Prusoff equation (Equation 3.6) shows that the ratio L_r/K_r distinguishes IC_{50} from K_D, we used this ratio to estimate the degree to which IC_{50}

measurements are affected by reporter peptide variation in experimental protocols. The value of L_r depended on what was reported in the database. A single value was used if provide; otherwise, the reported range became the lower and upper limits of a uniform distribution. To approximate values for K_r which is usually unreported, we used a lognormal distribution based on the equation for standard free energy change and the assumption that changes in free energy for peptide-MHC binding reactions are normally distributed. From the observation that most K_D values for peptide-MHC binding fall in the range of 10^{-7}-10^{-8} M (McFarland 2002), we assigned the lognormal distribution a mean of $10^{-7.5}$ and a standard deviation of $10^{0.5}$. For each peptide we generated 1000 values of L_r/K_r and calculated the mean and standard error of the mean. Filtered data subsets were created by excluding IC_{50} values associated with mean L_r/K_r exceeding either 1 or 9, representing deviations from K_D of 2- or 10-fold, respectively.

3.3 Results

3.3.1 Peptide Length Affects Binding Affinity to MHC Class II

To determine the nature of the relationship between peptide length and peptide-MHC class II binding affinity, we derived a number of regression fits to binding data for several MHC class II alleles from the AntiJen database. In all cases homologous sequences were first removed from the data sets using a pre-filtering algorithm, UniqueProt (Mika and Rost 2003). Parametric fits were then made based on polynomials with one, two, or three fitted parameters (linear, quadratic, and cubic, respectively). Analysis of variance from these fits showed that for these MHC class II alleles the nature of the relationship was most likely nonlinear (Table 3.1). A quadratic or cubic fit resulted in a significant reduction in sum of squares in all three cases at the 0.05 level.

To better characterize the apparent nonlinearities in the length-affinity data we then made nonparametric fits to the data and analyzed the fits. Local regression was used

to make nonparametric fits, and analysis was done using a permutation test. In this test binding affinities were reshuffled among peptide lengths to create 1000 new data sets, and a local regression fit was re-derived for each data set. If the fit to the original data fell outside of the middle 95% of permutation fits at any particular length, the nonlinearity at that length was determined to be significant. In each data set we found that the nonlinearity between length and affinity was significant at one or more lengths (Fig. 3.2). Lengths associated with strongest affinity could be identified, as could lengths associated with weakest affinity. For example, for DRB1*0401 affinity appeared strongest for peptides of 12 amino acids and weakest for peptides of 20 amino acids. When the data sets were combined and the local regression fits were regenerated, the same trends were seen (Fig. 3.2D): shorter peptide lengths, of approximately 12 amino acids, were associated with higher affinity, while longer peptide lengths, of approximately 20 amino acids, were associated with lower affinity.

Nonlinearities may have been present in the length-affinity data for several reasons, including the ability of peptides to shift registers within the MHC class II peptide-binding groove. To simulate the effect of register shifting on the mean affinity observed for peptides of different lengths, we used a simple statistical model based on two assumptions: first, that longer peptides are likely to contain more registers than shorter peptides, and secondly, that the measured affinity of a given peptide-MHC complex approximates the weighted average of the affinities of all the registers in a peptide (Equation 5). For a simulated peptide of a given length l, the affinities of $l - 8$ registers were generated and averaged. This process was repeated until the average affinities of either 10 or 100 peptides at each length (i.e. each value of l) were obtained, resulting in data sets of two sizes (one of the same magnitude as those typically obtained from databases, the other an order of magnitude larger). At this point a regression curve was derived (Fig. 3.3). For the larger sized data set the fitted curve was nonlinear and monotonically increasing (Fig. 3.3A). The same trend was seen in the smaller data set; in

this case, however, deviations were also possible, resulting in maxima at mid-length peptides (Fig. 3.3B). Together these results suggest that register shifting may be one mechanism behind the nonlinearities in the length-affinity relationship from experimental data sets.

We also estimated the lengths of the amino- and carboxyl-terminal portions of each peptide extending outside of the MHC class II peptide-binding groove to determine if particular lengths at either end of the peptide were favorable or unfavorable for binding. 9mer cores were identified by position 1 anchor residues (F, I, L, M, V, W, Y), and the lengths remaining at each end were calculated. Local regression fitting and permutation testing were done as with overall peptide length. In most cases fits to amino- and carboxyl-terminal peptide extensions were determined to be significant at one or more lengths (Fig. 3.4 and additional data not shown). In comparing fits we found that extensions of 2-4 amino acids at the amino terminus and extensions of 1-2 at the carboxyl terminus generally appeared favorable for binding (Fig. 3.4 and additional data not shown). Likewise, longer extensions (8 and 10 amino acids at the amino and carboxyl termini, respectively) generally appeared unfavorable for binding (Fig. 3.4 and additional data not shown). We also found that in at least some cases fits to overall peptide length could be decomposed into amino- and carboxyl-terminal contributions. For example binding to DRB1*0401 was strongest when amino and carboxyl termini were 2 and 1 amino acids, respectively (Fig. 3.4). Together with the 9mer core, these lengths sum to match the overall length associated with strongest binding, 12 amino acids (Fig. 3.2B).

3.3.2 Incorporating Peptide Length Improves Algorithm Performance

We incorporated peptide length into two peptide-MHC class II binding prediction algorithms in one of three ways. First, as a pre-processing event (Mod. 1 in Fig. 3.1) a local regression fit was made for affinity vs. length in the training/fitting data and the

value of the fit was subtracted from each affinity measurement. The resulting residuals were used in place of the original pIC_{50} values in the training set. After the algorithm was used to make initial predictions for the target set peptides, the value of the regression fit for each target set peptide length was added to yield final predictions. Alternatively (Alt. 1 in Fig. 3.1) length was also incorporated directly into the existing algorithm as an additional variable (in the case of ISC-PLS, as the 181st variable). Training/fitting was then performed as published, and predictions were made on test set peptide sequences and peptide lengths. Lastly we used a formula derived from the equilibrium-based statistical model to reconcile predictions made by existing algorithms on multiple registers within the peptide (Mod. 2 in Fig. 3.1). We point out that Mod. 1 and Alt. 1 are similar modifications that both consider peptide length directly (by fitting length as a discrete variable); in contrast Mod. 2 considers binding registers (i.e. 9mers with a valid position 1 anchor) and the relationship among them. Therefore, Mod. 1 and Alt. 1 are not used together, although either can be used with Mod. 2.

Incorporating peptide length by one or more modifications into the ISC-PLS algorithm improved the performance of the algorithm for all alleles examined (Table 3.2). Performance was measured by area under receiver operating characteristic curves (A_{ROC}) when a threshold of 500 nM was used to differentiate binding from non-binding affinities (Sette *et al.* 1994). The performance of ISC-PLS in conjunction with a combination rule (mean if less than one order range; highest otherwise) to reconcile register predictions was used as a baseline (Doytchinova and Flower 2003). Taking the highest scoring register to be representative of the entire peptide was also done as a reference. In general using any of three modifications resulted in increases in algorithm performance. However the modification resulting in the greater increase differed by MHC class II allele. In the case of DRB1*0101, deriving a regression fit (Mod. 1) resulted in significantly greater improvements than either using length as an additional variable or using the equilibrium-based formula to reconcile register predictions. In the case of DRB1*0401, all three

modifications resulted in the same magnitude of increase in performance. Finally in the case of DRB1*1501 only an application of both the regression fit (Mod. 1) and the equilibrium-based formula (Mod. 2) resulted in the greatest increase in performance. Differences in which modifications resulted in the greatest increase in performance may be suggestive of allele- or data set-specific mechanisms behind the length-affinity relationships.

We also incorporated peptide length into the TEPITOPE/ProPred algorithm (Sturniolo *et al.* 1999) and without re-deriving the pocket-specific matrices that define that algorithm found that increases in performance could be obtained by use of the equilibrium-based formula alone (Table 3.3). Typically in applications of TEPITOPE/ProPred to MHC class II, predictions on multiple registers are reconciled by taking the highest scoring register to be representative of the whole peptide (Brusic *et al.* 1998, Nielsen *et al.* 2004, Murugan and Dai 2005). We therefore used this rule to generate baseline predictions against which we could compare the performance of the equilibrium-based formula. Applying the formula for register shifting increased algorithm performance for all three data sets examined.

We also investigated whether our modifications might be applied to alleles for which fewer data exist. In analyzing the data for two other alleles, DRB1*0404 and DRB1*0405, we found no significant nonlinearities in regression fits of length versus affinity (Supplementary Data). Consistent with the results of these fittings, we observed no increase in performance after applying either Mod. 1 or Alt. 1 to the ISC-PLS algorithm when training sets were derived from these data sets (Supplementary Data). An increase in performance was observed, however, for the larger of the two data sets using Mod. 2 (Supplementary Data). These results suggest that our proposed modifications, like matrix-based prediction algorithms, are subject to limitations based on the size of the training set.

3.3.3 Experimental Variation in Data Sets Does Not Affect Algorithm Performance

In addition to the nonrandom effect of peptide length, another source of variation in the peptide-MHC class II binding data may be the use of multiple experimental protocols. Two parameters that frequently vary among protocols designed to measure peptide-MHC IC_{50} are the concentration of reporter peptide L_r and the equilibrium dissociation constant of the reporter peptide-MHC complex K_r. We made estimates of the ratio L_r/K_r for IC_{50} measurements in several data sets and used this ratio as the basis of classifying the data (Fig. 3.5).

Specifically, estimates of L_r/K_r were used to create two pIC_{50} data subsets that varied in their degree of deviation from pK_D and subsequently $\Delta G°$. L_r/K_r values of 1 and 9 were used as cutoffs, filtering out pIC_{50} values that differed from pK_D by more than approximately 2- and 10-fold, respectively. 29% of the total non-overlapping data had mean L_r/K_r values of greater than 1, while 5% had mean L_r/K_r values of greater than 9.

To determine whether variability in these experimental parameters affects data quality, we compared the accuracy of binding predictions using filtered and unfiltered data sets (Table 3.4). Successive rounds of leave-one-out cross-validation were performed such that predictions were made for each peptide in a given data set. Prediction accuracy was scored by calculating the Pearson correlation coefficient between experimental and predicted pIC_{50} values, and differences between scores were evaluated for statistical significance. Because data set size is known to affect prediction accuracy, random data subsets of the same sizes as the filtered subsets were used as controls.

In general, filtering data based on reporter peptide-specific parameters did not significantly improve the accuracy of prediction at either level of filtering (Table 3.4). While taking random subsets of the same data sets often resulted in only small changes to prediction accuracy, filtering sometimes resulted in significant degradation, particularly

for the smallest data sets (DRB1*0301, e.g.). An alternative approach, using the mean estimated L_r/K_r to convert pIC_{50} to pK_D, also failed to improve prediction accuracy (data not shown).

The effect of other experimental parameters on IC_{50} measurements could be observed indirectly. When the reporter peptide was identical to the peptide of interest, K_r was equal to K_D, allowing K_D to be calculated directly from the Cheng-Prusoff equation (Equation 3.6) as $K_D = IC_{50} - L_r$. In the DRB1*0401 data set, this condition was fulfilled for multiple IC_{50} measurements of the peptide PKYVKQNTLKLAT (HA 307-319), two of which were also made at the same temperature and using the same method of reporter peptide labeling (Hansen 1998, Consogno 2003). For these measurements, K_D was calculated to be 41 and 700, indicating that variation in parameters unrelated to temperature and reporter peptide concentration and affinity to MHC could also lead to variation in IC_{50} measurements.

3.4 Discussion

Information is typically lost during the prediction of peptide-MHC class II binding because most algorithms focus exclusively on 9mers within the peptide. An underlying assumption is that properties of the parent peptides that cannot be captured in their 9mers are irrelevant. This assumption may be true for MHC class I binding which involves peptides of nine amino acids almost exclusively but may not be true for MHC class II binding. Peptides that bind MHC class II are variable in length and may contain segments that extend past the ends of the peptide-binding groove, also known as peptide-flanking residues or PFR (Brown *et al.* 1993). PFR-MHC interactions may in turn affect peptide-MHC binding in a manner that is consistent and useful to prediction. Longer peptides also allow for register shifting, i.e. the ability of peptides to bind MHC using different core 9mers. PFR-MHC interactions and register shifting represent two possible mechanisms by which variability in peptide length affects affinity to MHC class II.

In this study we found that nonlinear relationships exist between peptide length and peptide-MHC class II binding affinity in a number of aggregate data sets available

online. When these nonlinearities were examined in more detail, they were found to be significant at several lengths, suggesting some lengths were more favorable for binding than others. This is consistent with the data from a number of experimental studies (Malcherek *et al.* 1994, Vogt *et al.* 1994, Bartnes *et al.* 1999, Fleckenstein *et al.* 1999). In these studies affinity was generally found to increase with length up to the longest lengths examined, typically between 15 and 17 amino acids. In our simulations register shifting was found to be one mechanism that could account for the direct relationship between length and binding affinity. However, our analysis of aggregate data sets suggests that additional mechanisms also contribute to the effect of length on affinity. For example, register shifting alone cannot explain why certain lengths at the amino and carboxyl termini are advantageous or disadvantageous for binding DRB1*0401. In this case other mechanisms such as hypothesized PFR-MHC interactions that are either attractive or repulsive may also be playing a role (Sercarz and Maverakis 2003).

Incorporating peptide length into existing binding prediction algorithms by one or more of our modifications consistently improved performance for multiple MHC class II alleles. Three modifications were used—one at the level of the training set data (Mod. 1), another within the algorithm itself (Alt. 1), and the last after 9mer predictions were generated (Mod. 2)—and all resulted in performance gains over reference algorithms ISC-PLS and TEPITOPE/ProPred. Baseline A_{ROC} scores for two different algorithms varied between 0.57 and 0.73. By comparison A_{ROC} scores for modified algorithms varied between 0.68 and 0.77, consistent with the range of scores listed in MHCBench (http://www.imtech.res.in/raghava/mhcbench/). The modification resulting in the largest performance increase differed by allele, and this may in part reflect differences in the mechanisms by which length affects affinity. For DRB1*0401, for example, using the formula for register shifting resulted in performance gains that were statistically indistinguishable from those obtained using other modifications. For DRB1*0101, however, modifications based on regression modeling resulted in significantly greater

performance increases. These data therefore support roles for both register shifting and other mechanisms.

Previous studies have provided indirect evidence that accounting for variability in peptide length could improve prediction. Godkin *et al.* (1998), for example, found that matrices based on 15mers generally outperformed matrices based on shorter lengths, showing the usefulness of considering information outside of the core 9mer. Likewise, Bui *et al.* (2005) have proposed deriving a separate matrix for each length of peptide (Bui *et al.* 2005). Despite the suggestion that explicit consideration of peptide length could improve binding prediction (McFarland and Beeson 2002), to our knowledge no previous study has implemented this idea. Our results affirm the use of peptide length in binding prediction. In addition our modifications are sufficiently general that they could be incorporated into other current algorithms based on scoring 9mers.

Thus far experimental evidence of either register shifting or PFR-MHC interactions has involved only a small sampling of MHC class II alleles and been of indeterminate generality. For example, register shifting has been demonstrated to occur with alleles I-A^d and I-A^u in mice and DR2 in humans (McFarland *et al.* 1999, Li *et al.* 2000, Seamons *et al.* 2003, Bankovich *et al.* 2004). Solved structures exist for a somewhat wider array of alleles, including I-A^d and I-A^k in mice and DR1, DR3, and DR4 in humans (see McFarland and Beeson 2002 for a review). Although these structures show the presence of PFRs in peptide-MHC class II complexes, they fail to capture the dynamics of either register shifting or PFR-MHC interactions.

Our analysis of regression fits to different aggregate binding data sets suggests that longer PFRs (i.e. in peptides longer than approximately 16 amino acids) may generally be deleterious to binding. At the same time, however, PFRs of a certain minimum length increase the probability of a peptide having multiple binding registers which, our simulations show, increases overall binding affinity. An optimal peptide length for binding each MHC class II variant may therefore exist. Further computational

analysis of aggregate data sets may provide a complement to more direct, observation-based studies in continuing to elucidate the role of peptide length in MHC class II binding. In addition these findings may be of use to the design of peptide vaccines which often comprise only short segments of disease-relevant protein antigens (Larche and Wraith 2005). Including PFRs of optimal lengths may help to ensure efficacious binding to MHC.

Figures

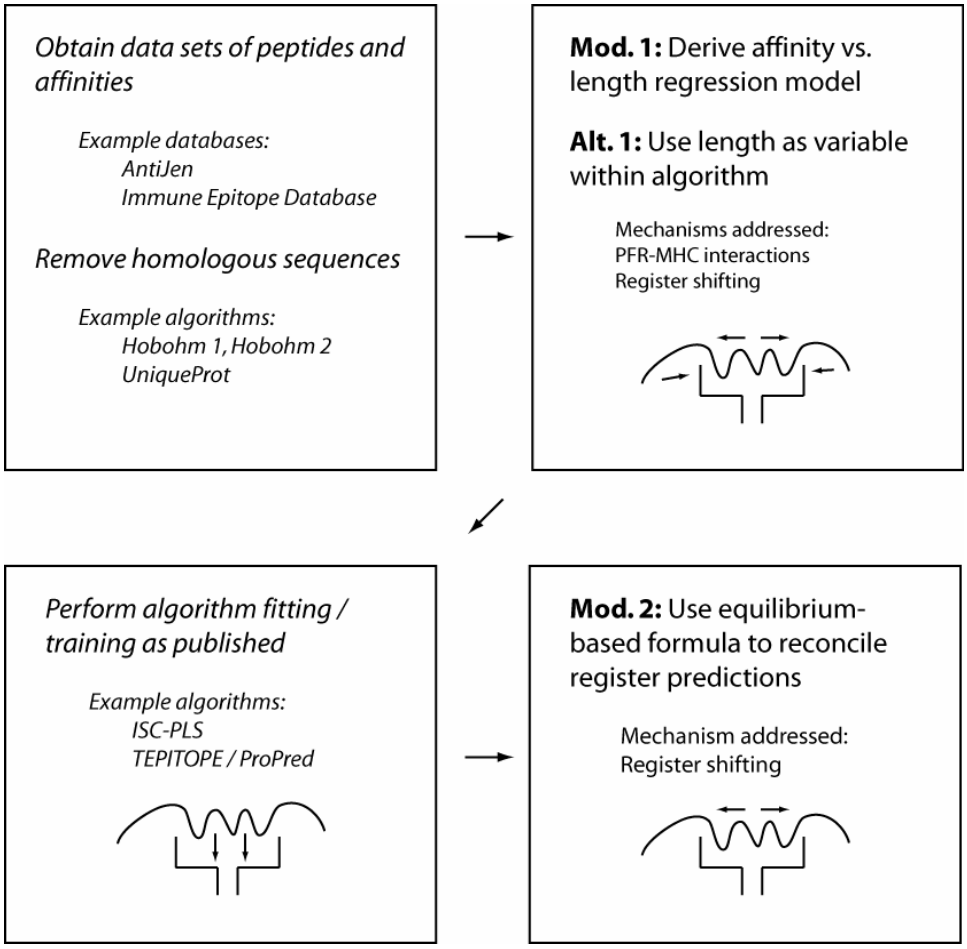

Figure 3.1. Schematic of modifications made to existing algorithms to incorporate peptide length. Modification 1, Alternative Modification 1, and Modification 2 are abbreviated Mod. 1, Alt. 1, and Mod. 2. Also shown are examples of sources of data, algorithms used to remove homologous sequences from data, and algorithms to predict peptide-MHC class II binding.

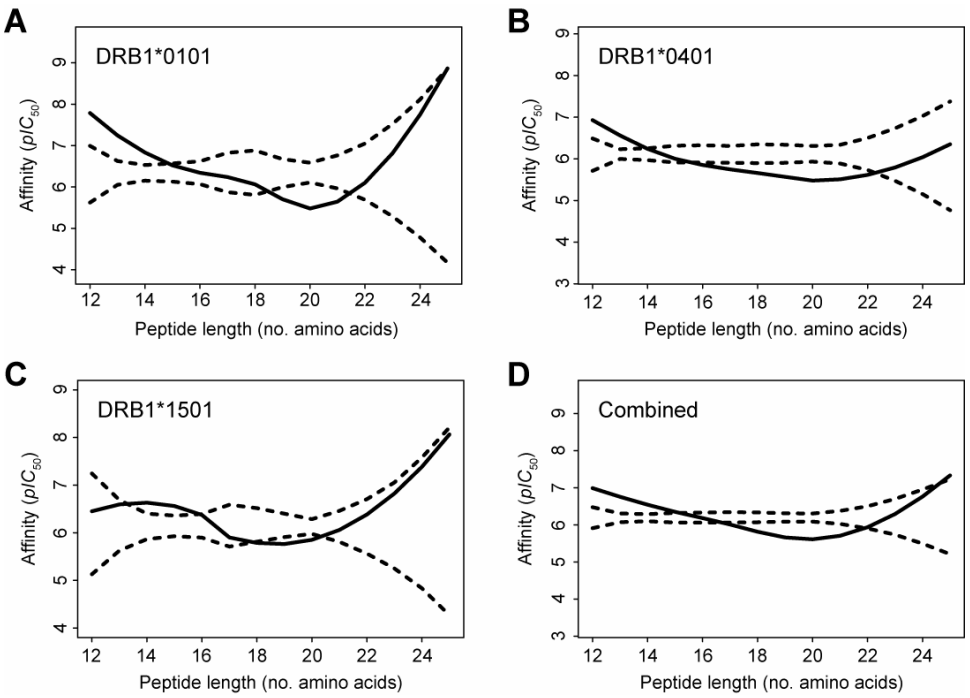

Figure 3.2. Local regression fits of peptide-MHC class II binding affinity versus peptide length for three HLA data sets: *A*, DRB1*0101; *B*, DRB1*0401; *C*, DRB1*1501; and *D*, the three data sets combined. 95% boundaries of permutation distributions are shown (dotted) with fits to the original, non-shuffled data (solid).

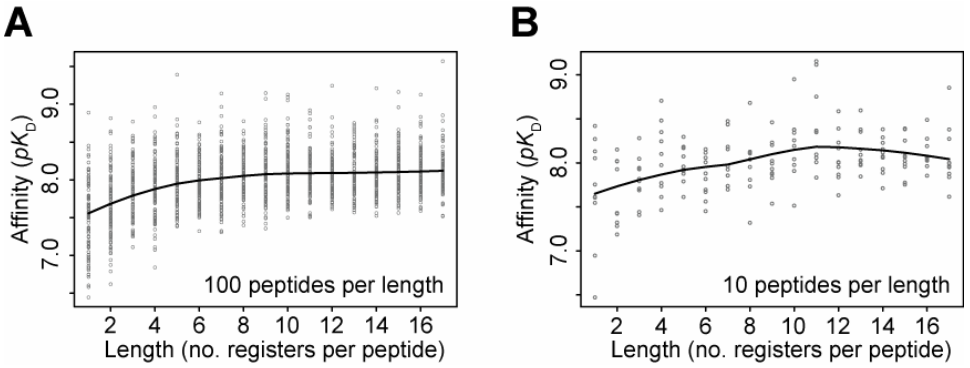

Figure 3.3. Statistical simulations of the effects of register shifting on MHC class II binding affinity over a range of peptide lengths: A, for a 1700-peptide data set; B, for a 170-peptide data set. Curve fits by local regression are shown overlaid.

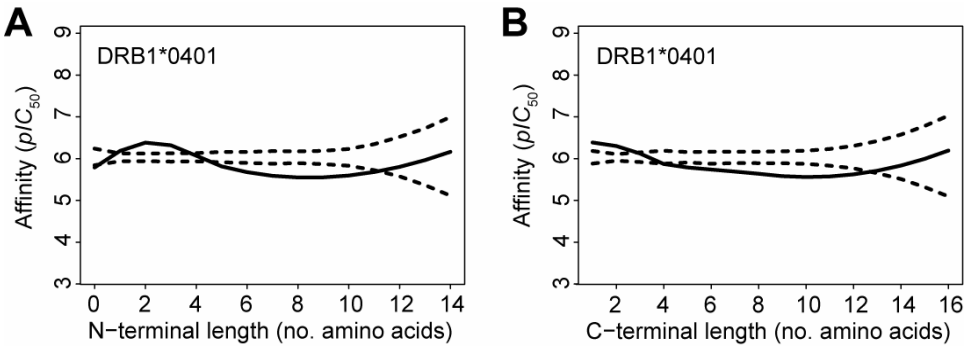

Figure 3.4. Local regression fits of peptide-MHC class II binding affinity versus lengths of portions of the peptide extending outside of the peptide-binding groove for the HLA-DRB1*0401 data set: *A*, at the N terminus and *B*, at the C terminus. 95% boundaries of permutation distributions are shown (dotted) with fits to the original, non-shuffled data (solid).

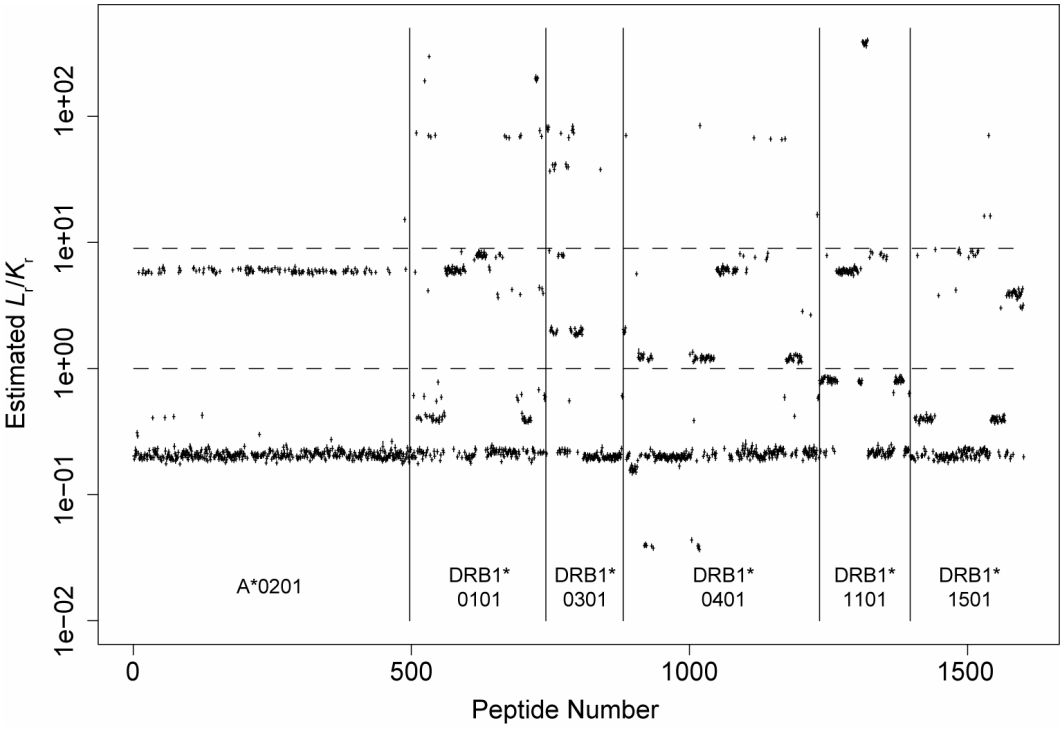

Figure 3.5. Variance among experimental parameters used in measuring peptide-MHC binding affinity. Mean estimates of the ratio L_r/K_r for the IC_{50} measurements in six data sets are shown with error bars. Dotted lines indicate L_r/K_r cutoffs of 1 and 9. Data for the DR4 supertype are not shown.

Tables

Table 3.1. Evidence of non-linear relationships in length-affinity data for several MHC class II alleles.

	DRB1*0101	DRB1*0401	DRB1*1501
Quadratic, F	11.745 (<0.001)	8.575 (0.004)	3.670 (0.057)
Cubic, F	5.849 (0.016)	0.708 (0.401)	4.871 (0.028)

F statistics are shown for analysis of variance results with p values in parentheses.

Table 3.2. Evidence of non-linear relationships in length-affinity data for several MHC class II alleles.

DRB1*0101	ISC-PLS	Mod. 1: Regression fit	Alt. 1: Length as variable
Combination rule	0.615 ± 0.009[1]	0.754 ± 0.009	0.690 ± 0.013
Highest scoring register	0.652 ± 0.008	0.758 ± 0.006	0.705 ± 0.013
Mod. 2: Equilibrium formula	0.709 ± 0.005	0.770 ± 0.009	0.752 ± 0.003
DRB1*0401	ISC-PLS	Mod. 1: Regression fit	Alt. 1: Length as variable
Combination rule	0.730 ± 0.007[1]	0.741 ± 0.010	0.749 ± 0.009
Highest scoring register	0.732 ± 0.015	0.750 ± 0.006	0.751 ± 0.005
Mod. 2: Equilibrium formula	0.757 ± 0.008	0.757 ± 0.004	0.754 ± 0.008
DRB1*1501	ISC-PLS	Mod. 1: Regression fit	Alt. 1: Length as variable
Combination rule	0.574 ± 0.009[1]	0.596 ± 0.015	0.584 ± 0.020
Highest scoring register	0.575 ± 0.021	0.626 ± 0.014	0.603 ± 0.011
Mod. 2: Equilibrium formula	0.609 ± 0.019	0.677 ± 0.014	0.609 ± 0.018

Five-fold cross-validation (5x-CV) was used and repeated five times. Mean AROC scores between predicted and experimentally determined pIC_{50} values are shown with standard errors of the mean. Highest scores are shown in bold with multiple scores in bold if pair-wise differences were not statistically significant. A threshold of 500 nM (Sette *et al.* 1994) was used to distinguish binding from non-binding peptides. [1]The ISC-PLS algorithm with combination rule (Doytchinova and Flower 2003) was used as a baseline prediction.

Table 3.3. Binding prediction accuracy of ProPred algorithm for different MHC class II alleles when peptide length was incorporated.

	ProPred: DRB1*0101	ProPred: DRB1*0401	ProPred: DRB1*1501
Combination rule	0.685	0.741	0.669
Highest scoring register	0.667[1]	0.754[1]	0.635[1]
Mod. 2: Equilibrium formula	0.702	0.764	0.680

Matrices were obtained from the ProPred website and used to calculate a score for each register within a peptide. To each score the approximate affinity of an all-alanine 9mer to MHC was added (pIC_{50} = 6.169, Doytchinova and Flower 2003). A_{ROC} scores between predicted and experimentally determined pIC_{50} are shown, using a threshold of 500 nM (Sette *et al.* 1994) to distinguish binding from non-binding peptides. [1]Highest ProPred-predicted scores from all eligible registers were used as baseline predictions following recent precedents (Brusic *et al.* 1998, Nielsen *et al.* 2004, Murugan and Dai 2005).

Table 3.4. Accuracy of peptide-MHC binding affinity predictions when made using data filtered on the basis of reporter peptide-specific conditions and random data subsets.

	A*0201	DRB1* 0101	DRB1* 0301	DRB1* 0401	DR4	DRB1* 1101	DRB1* 1501
Complete data set	0.5534 (496)	0.4161 (245)	0.4583 (139)	0.5028 (353)	0.4707 (524)	0.6235 (163)	0.5672 (205)
Exclude $L_r/K_r \geq 9$	0.5507 (495)	0.4844 (227)	0.3837 (120)	0.4839 (346)	0.5203 (480)	0.6731 (153)	0.5758 (202)
Random subset	0.5546 (495)	0.4275 (227)	0.4863 (120)	0.4621 (346)	0.4690 (480)	0.6124 (153)	0.5957 (202)
Exclude $L_r/K_r \geq 1$	0.4861 (383)	0.3958 (161)	0.1516 (86)	0.3170 (237)	0.3704 (367)	0.4482 (104)	0.6793 (161)
Random subset	0.5129 (383)	0.3616 (161)	0.3462 (86)	0.5411 (237)	0.4990 (367)	0.5815 (104)	0.6581 (161)

Pearson correlation coefficients between predicted and experimental pIC_{50} values are shown along with data set sizes in parentheses.

References

Altuvia, Y., Sette, A., Sidney, J., Southwood, S. and Margalit, H. (1997) A structure-based algorithm to predict potential binding peptides to MHC molecules with hydrophobic binding pockets. *Hum. Immunol.*, **58**, 1-11.

Arnold, P.Y., La Gruta, N.L., Miller, T., Vignali, K.M., Adams, P.S., Woodland, D.L. and Vignali, D.A. (2002) The majority of immunogenic epitopes generate $CD4^+$ T cells that are dependent on MHC class II-bound peptide flanking residues. *J. Immunol.* **169**, 739-749.

Bankovich, A.J., Girvin, A.T., Moesta, A.K. and Garcia, K.C. (2004) Peptide register shifting within the MHC groove: theory becomes reality. *Mol. Immunol.*, **40**, 1033-1039.

Bartnes, K., Leon, F., Briand, J.P., Travers P.J. and Hannestad K. (1999) N-terminal elongation of a peptide determinant beyond the first primary anchor improves binding to H-2 I-Ad and HLA-DR1 by backbone-dependent and aromatic side chain-dependent interactions, respectively. *Eur. J. Immunol.*, **29**, 189-195.

Belmares, M.P. and McConnell, H.M. (2001) Kinetics of registry selection of chimeric peptides binding to MHC II. *Biochemistry*, **40**, 10284-10292.

Blythe, M.J., Doytchinova, I.A. and Flower, D.R. (2002) JenPep: a database of quantitative functional peptide data for immunology. *Bioinformatics*, **18**, 434-439.

Brown, J.H., Jardetzky, T.S., Gorga, J.C., Stern, L.J., Urban, R.G., Strominger, J.L. and Wiley, D.C. (1993) Three-dimensional structure of the human class II histocompatiblity antigen HLA-DR1. *Nature*, **364**, 33-39.

Brusic, V. and Harrison, L. (1994) Prediction of MHC binding peptides using artificial neural networks. In Stonier, R.J. and Yu, X.S. (eds) *Complex Systems: Mechanism of Adaptation.* IOS Press, Amsterdam, pp. 253-260.

Brusic, V., Rudy, G., Honeyman, M., Hammer, J. and Harrison, L. (1998) Prediction of MHC class II-binding peptides using an evolutionary algorithm and artificial neural network. *Bioinformatics*, **14**, 121-130.

Bui, H.H., Sidney, J., Peters, B., Sathiamurthy, M., Sinichi, A., Purton, K.A., Mothe, B.R., Chisari, F.V., Watkins, D.I. and Sette, A. (2005) Automated generation and evaluation of specific MHC binding predictive tools: ARB matrix applications. *Immunogenetics*, **57**, 304-314.

Buus, S. (1999) Description and prediction of peptide-MHC binding: the 'human MHC project'. *Curr. Opin. Immunol.*, **11**, 209-213.

Cheng, Y. and Prusoff, W.H. (1973) Relationship between the inhibition constant (K1) and the concentration of inhibitor which causes 50 per cent inhibition (I50) of an enzymatic reaction. *Biochem. Pharmacol.*, **22**, 3099-3108.

Chicz, R.M., Urban, R.G., Lane, W.S., Gorga, J.C., Stern, L.J., Vignali, D.A.A. and Strominger, J.L. (1992) Predominant naturally-processed peptides bound to HLA-DR1 are derived from MHC-related molecules and are heterogeneous in size. *Nature*, **358**, 764-768.

Cleveland, W.S. and Devlin, S.J. (1988) Locally weighted regression: an approach to regression analysis by local fitting. *J. Amer. Stat. Assoc.*, **83**, 596-610.

Consogno, G., Manici, S., Facchinetti, V., Bachi, A., Hammer, J., Conti-Fine, B.M., Rugarli, C., Traversari, C. and Protti, M.P. (2003) Identification of immunodominant regions among promiscuous HLA-DR-restricted CD4+ T-cell epitopes on the tumor antigen MAGE-3. *Blood*, **101**, 1038-1044.

Davenport, M.P., Ho Shon, I.A. and Hill, A.V. (1995) An empirical method for the prediction of T-cell epitopes. *Immunogenetics*, **42**, 392-397.

Davies, M.N., Sansom, C.E., Beazley C. and Moss, D.S. (2003) A novel predictive technique for the MHC class II peptide-binding interaction. *Mol. Med.*, **9**, 220-225.

Doytchinova, I.A. and Flower, D.R. (2003) Towards the *in silico* identification of class II restricted T-cell epitopes: a partial least squares iterative self-consistent algorithm for affinity prediction. *Bioinformatics*, **19**, 2263-2270.

Eisenberg, D. and Crothers, D. (1979) *Physical Chemistry with Applications to the Life Sciences*. Benjamin/Cummings, Menlo Park.

Falk, K., Rotzschke, O., Stevanovic, S., Jung, G. and Rammensee, H.G. (1991) Allele-specific motifs revealed by sequencing of self-peptides eluted from MHC molecules. *Nature*, **351**, 290-296.

Fleckenstein, B., Jung, G. and Wiesmuller, K.H. (1999) Quantitative analysis of peptide-MHC class II interaction. *Semin. Immunol.*, **11**, 405-416.

Godkin, A.J., Davenport, M.P., Willis, A., Jewell, D.P. and Hill, A.V.S. (1998) Use of complete eluted peptide sequence data from HLA-DR and -DQ molecules to predict T cell epitopes, and the influence of nonbinding terminal regions of ligands in epitope selection. *J. Immunol.*, **161**, 850-858.

Hansen, B.E., Andersson, E.C., Madsen L.S., Engberg, J., Sondergaard, L., Svejgaard, A., and Fugger, L. (1998) Functional characterization of HLA-DRA1*0101/DRB1*0401 molecules expressed in *Drosophila melanogaster* cells. *Tissue Antigens*, **51**, 119-128.

Hobohm, U., Scharf, M., Schneider, R. and Sander, C. (1992) Selection of representative protein data sets. *Protein Sci.*, **1**, 409-417.

Honeyman, M., Brusic, V., Stone, N. and Harrison, L. (1998) Neural network-based prediction of candidate T-cell epitopes. *Nat. Biotechnol.*, **16**, 966-969.

Jones, E.Y. (1997) MHC class I and class II structures. *Curr. Opin. Immunol.*, **9**, 75-79.

Kass, Q. and Lefranc, M.P. (2005) T cell receptor/peptide/MHC molecular characterization and standardized pMHC contact sites in IMGT/3Dstructure-DB. *In Silico Biol.*, **5**, 505-528.

Kaufmann, S.H. and Schaible, U.E. (2005) Antigen presentation and recognition in bacterial infections. *Curr. Opin. Immunol.*, **17**, 79-87.

Larche, M. and Wraith, D.C. (2005) Peptide-based therapeutic vaccines for allergic and autoimmune diseases. *Nat. Med.*, **11**, S69-S76.

Li, Y., Li, H., Martin, R. and Mariuzza, RA. (2000) Structural basis for the binding of an immunodominant peptide from myelin basic protein in different registers by two HLA-DR2 proteins. *J. Mol. Biol.*, **304**, 177-188.

Lund, O., Nielsen, M., Lundegaard, C., Kesmir, C. and Brunak, S. (2005) Methods applied in immunological bioinformatics. In Lund, O., Nielsen, M., Lundegaard, C., Kesmir, C. and Brunak, S., *Immunological Bioinformatics*. MIT Press, Cambridge, pp. 69-102.

Malcherek, G., Gnau, V., Stevanovic, S., Rammensee, H.G., Jung, G. and Melms, A. (1994) Analysis of allele-specific contact sites of natural HLA-DR17 ligands. *J. Immunol.*, **153**, 1141-1149.

Mamitsuka, H. (1998) MHC molecules using supervised learning of hidden Markov models. *Proteins: Structure, Function and Genetics*, **33**, 460-474.

Marshall, K.W., Wilson, K.J., Liang, J., Woods, A., Zaller, D. and Rothbard, J.B. (1995) Prediction of peptide affinity to HLA DRB1*0401. *J. Immunol.*, **154**, 5927-5933.

McFarland, B.J., Sant, A.J., Lybrand, T.P. and Beeson, C. (1999) Ovalbumin(323-339) peptide binds to the major histocompatibility complex class II I-A^d protein using two functionally distinct registers. *Biochemistry*, **38**, 16663-16670.

McFarland, B.J. and Beeson, C. (2002) Binding interactions between peptides and proteins of the class II major histocompatibility complex. *Med. Res. Rev.*, **22**, 168-203.

Mika, S. and Rost, B. (2003) UniqueProt: creating representative protein-sequence sets. *Nucleic Acids Res.*, **31**, 3789-3791.

Motulsky, H. and Christopoulos, A. (2004) *Fitting Models to Biological Data Using Linear and Nonlinear Regression: A Practical Guide to Curve Fitting.* Oxford University Press, Oxford.

Murugan, N. and Dai, Y. 2005. Prediction of MHC class II binding peptides based on an iterative learning model. *Immunome Res.*, **1**, 6.

Nielsen, M., Lundegaard, C., Worning, P., Hvid, C.S., Lamberth, K., Buus, S., Brunak, S. and Lund, O. (2004) Improved prediction of MHC class I and class II epitopes using a novel Gibbs sampling approach. *Bioinformatics*, **20**, 1388-1397.

Parker, K.C., Bednarek, M.A. and Coligan, J.E. (1994) Scheme for ranking potential HLA-A2 binding peptides based on independent binding of individual peptide side-chains. *J. Immunol.*, **152**, 163-175.

Peters, B., Sidney, J., Bourne, P., Bui, H.H., Buus, S., Doh, G., Fleri, W., Kronenberg, M., Kubo, R., Lund, O., Nemazee, D., Ponomarenko, J.V., Sathiamurthy, M., Schoenberger, S., Stewart, S., Surko, P., Way, S., Wilson, S. and Sette, A. (2005) The immune epitope database and analysis resource: from vision to blueprint. *PLoS Biol.*, **3**, e91.

R Development Core Team. (2005) *R: A Language and Environment for Statistical Computing.* R Foundation for Statistical Computing, Vienna.

Rammensee, H. (1995) Chemistry of peptides associated with MHC class I and class II molecules. *Curr. Opin. Immunol.*, **7**, 85-96.

Robinson, J., Waller, M.J., Parham, P., de Groot, N., Bontrop, R., Kennedy, L.J., Stoehr, P., and Marsh, S.G.E. (2003) IMGT/HLA and IMGT/HLA: sequence databases for the study of the major histocompatibility complex. *Nuc. Acids Res.*, **31**, 311-314.

Roche, P.A. and Cresswell, P. (1990) High-affinity binding of an influenza hemagglutinin-derived peptide to purified HLA-DR. *J. Immunol.*, **144**, 1849-1856

Rognan, D., Lauemoller, S.L., Holm, A., Buus, S. and Tschinke, V. (1999) Predicting binding affinities of protein ligands from three-dimensional models: application to peptide binding to class I major histocompatibility proteins. *J. Med. Chem.*, **42**, 4650-4658.

Schafroth, H.D. and Floudas, C.A. (2004) Predicting peptide binding to MHC pockets via molecular modeling, implicit solvation, and global optimization. *Proteins*, **54**, 534-556.

Schueler-Furman, O., Altuvia, Y., Sette, A. and Margalit, H. (2000) Structure-based prediction of binding peptides to MHC class I molecules: application to a broad range of MHC alleles. *Protein Sci.*, **9**, 1838-1846.

Seamons, A., Sutton, J., Bai, D., Baird, E., Bonn, N., Kafsack, B.F., Shabanowitz, J., Hunt, D.F., Beeson, C. and Goverman, J. (2003) Competition between two MHC binding registers in a single peptide processed from myelin basic protein influences tolerance and susceptibility to autoimmunity. *J. Exp. Med.*, **197**, 1391-1397.

Sercarz, E.E. and Maverakis, E. (2003) MHC-guided processing: binding of large antigen fragments. *Nat. Rev. Immunol.*, **3**, 621-629.

Sette, A., Vitiello, A., Reherman, B., Fowler, P., Nayersina, R., Kast, W.M., Melief, C.J., Oseroff, C., Yuan, L. and Ruppert, J. (1994) The relationship between class I binding affinity and immunogenicity of potential cytotoxic T cell epitopes. *J. Immunol.*, **153**, 5586-5592.

Singh, H. and Raghava, G.P.S. (2001) ProPred: prediction of HLA-DR binding sites. *Bioinformatics*, **17**, 1236-1237.

Srinivasan, M., Domanico, S.Z., Kaumaya, P.T.P. and Pierce, S.K. (1993) Peptides of 23 residues or greater are required to stimulate a high affinity class II-restricted T cell response. *Eur. J. Immunol.*, **23**, 1011-1016.

Southwood, S., Sidney, J., Kondo, A., del Guercio, M.F., Appella, E., Hoffman, S., Kubo, R.T., Chesnut, R.W., Grey, H.M., and Sette, A. (1998) Several common

HLA-DR types share largely overlapping peptide binding repertoires. *J. Immunol.*, **160**, 3363-3373.

Sturniolo, T., Bono, E., Ding, J., Raddrizzani, L., Tuereci, O., Sahin, U., Braxenthaler, M., Gallazzi, F., Protti, M.P., Sinigaglia, F. and Hammer, J. (1999) Generation of tissue-specific and promiscuous HLA ligand databases using DNA microarrays and virtual HLA class II matrices. *Nat. Biotechnol.*, **17**, 555-561.

Vogt, A.B., Kropshofer, H., Kalbacher, H., Kalbus, M., Rammensee, H.G., Coligan, J.E. and Martin, R. (1994) Ligand motifs of HLA-DRB5*0101 and DRB1*1501 molecules delineated from self-peptides. *J. Immunol.*, **153**, 1665- 1673.

CHAPTER 4

How Multiple Host Polymorphisms Affect Immune Response to Tuberculosis

4.1 Introduction

Tuberculosis (TB) continues to pose a global health problem. An estimated one-third of the human population is infected with the pathogen *Mycobacterium tuberculosis* (*Mtb*), and approximately two million individuals succumb to the disease annually (WHO, www.who.int/mediacentre/factsheets/fs104). However, only a fraction of infected individuals ever progress to disease (Small and Fujiwara 2001). What distinguishes those who are able to control the infection from those who are not? In addition to environmental factors such as nutrition and HIV co-infection, host genetics are likely to play a role, and identifying polymorphisms in genes that predispose individuals to TB continues to be an area of active research (reviewed in Bellamy 2005, Fernando and Britton 2006, Hill 2006).

Identification of genetic polymorphisms that affect susceptibility to TB proceeds primarily on the basis of epidemiological data from association studies (Casanova and Abel 2002). In association studies the frequency of a polymorphism-based allele is compared in patients and healthy controls. If the allele is found to be over-represented in patients, it can then be hypothesized to encode a protein variant that renders its bearer more susceptible to TB.

Association studies do not always yield consistent results, however, and an allele that is correlated with TB in one study might not be correlated with TB in another study

(Fernando and Britton 2006). Several reasons might explain this inconsistency (Alcais *et al.* 2001). Small samples sizes (i.e., insufficient numbers of subjects) may lead to unreliable detection of low-frequency alleles. Alternatively, alleles may have been detected using indirect, serological tests that are unable to distinguish between closely related alleles, unlike more direct, sequence-based tests.

Even more difficult to account for is the problem of genetic heterogeneity. Because the number of polymorphisms affecting immune function is vast, and many have yet to be discovered, epidemiological studies necessarily fail to assay all polymorphisms. In most studies only a single polymorphism is assayed. Other polymorphisms may compound or counteract the effect of a single polymorphism, and such interactions can be difficult to detect without additional testing. For instance, a polymorphism X_1 in gene X may render the human host more susceptible to TB in the presence of polymorphism Y_1 in gene Y but not in the presence of an alternative polymorphism Y_2. Such interactions underlie the difficulty in comparing studies done on different populations (Alcais *et al.* 2001), even when the same allele is being studied and study design is identical.

Antigen presentation requires the contribution of several genes, and many of these genes bear polymorphisms that have been associated with TB (c.f. Bellamy 2005, Fernando and Britton 2006, Hill 2006; Table 4.2 in Supplementary Information). During antigen presentation, cells bind peptides from pathogens to receptors known as MHC (major histocompatibility complex) molecules and display the resulting peptide-MHC (pMHC) complexes on their surface (Fig. 4.1A). Two classes of MHC molecule exist: class I, primarily for peptides found in the cytoplasm, and class II, primarily for peptides found in endosomal compartments. Cells known as antigen-presenting cells (APCs) express both classes of MHC, allowing them to display peptides from a variety of pathogens.

A T cell response to antigen presentation begins when a T cell scans the surface of the APC (Fig. 4.1A, reviewed in Santana and Esquivel-Guadarrama 2006). If T cell

receptors (TCRs) expressed by a T cell bind pMHC complexes displayed on the APC surface with sufficient affinity, and co-stimulatory molecules such as B7 and CD28 are present, a signaling cascade is initiated that ultimately results in T cell activation. Within minutes of APC-T cell contact, intracellular calcium levels rise and TCRs from the surface are internalized. After several hours cytokines including IL-2 and IFN-γ are produced by T cells leading to T cell proliferation and activation of other cell types. Antigen presentation therefore involves events occurring on timescales of seconds to hours and at molecular, cellular, and multi-cellular length scales.

Among the antigen presentation-specific polymorphisms associated with TB, perhaps the best studied are those in genes for HLA (human leukocyte antigen), the human form of MHC. Of the two classes of HLA, HLA class II is particularly relevant to TB because HLA class II molecules bind peptides from antigens in endosomal compartments where *Mtb* resides. Over 800 HLA class II alleles have been identified, among which are several that have been associated with increased or decreased susceptibility to TB (Robinson *et al.* 2003, Bellamy 2005). In particular, the DRB1*1501 allele has been associated with susceptibility to TB in numerous studies (Mehra *et al.* 1995, Ravikumar *et al.* 1999, Teran-Escandon *et al.* 1999, Sriram *et al.* 2001; Table 4.2 in Supplementary Information). Generally polymorphisms in HLA map to the peptide-binding regions of the molecule and can therefore be assumed to affect function (Rammensee 1995). A general mechanism explaining how HLA polymorphisms affect the immune response to TB, however, has not been established. Other polymorphisms affect HLA expression rather than the peptide-binding properties of HLA; to our knowledge, none of these polymorphisms has yet been tested for TB association (Louis *et al.* 1994, Cowell *et al.* 1998).

MHC expression occurs constitutively in APCs but is subject to the up-regulatory effects of signals from the extracellular environment including the cytokine IFN-γ. Polymorphisms in the IFN-γ gene have also been associated with TB susceptibility. In

particular, the +874A allele was found to be significantly over-represented in TB patients, suggesting +874A increases susceptibility to TB (Lio *et al.* 2002, Lopez-Maderuelo *et al.* 2003, Rossouw *et al.* 2003; Table 4.2 in Supplementary Information). PBMCs (peripheral blood mononuclear cells which include APCs and T cells) from +874A individuals produce significantly less IFN-γ *in vitro* than PBMCs from +874T individuals after antigenic stimulation (Pravica *et al.* 1999, Lopez-Maderuelo *et al.* 2003). In addition to its up-regulatory effects on MHC expression, IFN-γ has a number of other functions including activation of macrophages and NK cells and inhibition of the TH2 phenotype in T cells (Maher *et al.* 2007), but exactly which of these functions is undermined in +874A individuals has not been determined.

Finally polymorphisms affecting antigen processing (the partial degradation of proteins into peptides) have been associated with susceptibility to TB, though thus far only for MHC class I-mediated antigen presentation. The transporter associated with antigen processing (TAP) is responsible for translocating peptides from the cytoplasm into the endoplasmic reticulum where they can be bound by MHC class I molecules. A polymorphism in TAP2, one of the subunits that constitute TAP, was found over-represented in TB patients (Rajalingam *et al.* 1997, Gomez *et al.* 2006; Table 4.2 in Supplementary Information). Other enzymes perform analogous functions for the MHC class II-mediated pathway, namely the cathepsin proteases responsible for antigen processing. Polymorphisms are known to exist in genes for cathepsins (Taggart 1992), but to our knowledge, none has been tested for TB association.

How polymorphisms in HLA, IFN-γ, and other genes interact to ultimately determine genetic susceptibility to TB remains an open question. Mathematical modeling can help to provide a unifying framework with which to consider these polymorphisms. Such a framework would ideally have immunologically relevant readouts such as cytokine production and allow the effect of each polymorphism to be simulated and observed, both individually and with other polymorphisms. Several questions could then

be approached theoretically. For example, could a polymorphism that up-regulates IFN-γ expression compensate for a polymorphism that results in HLA deficient in binding *Mtb* peptides? How might T cell response differ among individuals with different combinations of polymorphisms?

To approach these questions we developed a multi-scale mathematical model of antigen presentation that represents both APCs and T cells and tracks events from the molecular scale to cellular and multi-cellular scales. Particular detail was given to pathways involving MHC and IFN-γ, allowing polymorphisms affecting both pathways to be simulated. The extent to which a polymorphism in one gene compounded or counteracted the effect of a polymorphism in another gene could be observed, allowing us to determine whether the presence of multiple polymorphisms could be a confounding factor in association studies.

4.2 Methods

The multi-scale model comprises three models that were developed separately (Fig. 4.1): an APC model representing the events leading up to the appearance of pMHC on the APC surface, a T cell model representing the events leading up to TCR internalization, and an intracellular T cell signaling model representing the events leading up to cytokine (IFN-γ) production. We provide an overview of the three models here.

4.2.1 APC Model

We describe the APC model elsewhere in detail (Singer and Linderman 1990, Singer and Linderman 1991, Agrawal and Linderman 1996, Chang *et al.* 2005). Briefly, we represented the major events leading up to antigen presentation within APCs, e.g., macrophages, using ordinary differential equations (ODEs). These events include *de novo* synthesis of MHC, the up-regulatory effect of IFN-γ, uptake and processing of

extracellular antigens, formation of peptide-MHC (pMHC) complexes, and trafficking of pMHC to and from the APC surface (Fig. 4.1B). Equations are provided in Supplementary Information, and parameters in Tables 4.3 and 4.4.

4.2.2 T Cell Model

To represent T cell response to antigen presentation, we developed a T cell model and linked it to the APC model. The T cell model was based on a model originally published in Coombs *et al.* (2002), further developed in Gonzalez *et al.* (2005), which represents two important features of T cell signaling, namely kinetic proofreading (i.e., the requirement for pMHC-TCR engagement to persist for a certain duration to result in TCR activation) and serial triggering (i.e., the ability of one pMHC to engage multiple TCR). In the Coombs model as well as in our model, the events following APC-T cell contact are represented, specifically engagement of pMHCs by TCRs, progression of pMHC-TCR complexes through various states of activation, and finally internalization of fully activated TCRs as a marker of T cell activation (Fig. 4.1C). Co-stimulatory molecules such as B7, CD28, ICAM-1, and LFA-1 were assumed to be present in non-limiting quantities. In order to be internalized, TCRs in the model were required to be fully activated in either free or pMHC-bound forms; the contribution of constitutively recycling TCRs to the pool of internalized TCRs was excluded. Only the contact zone between the APC and T cell was considered. The degree of TCR internalization occurring in the contact zone was therefore assumed to be representative of the degree of TCR internalization occurring elsewhere on the T cell surface. The T cell model comprises a set of ODEs separate from the ODEs of the APC model (equations provided in Supplementary Information, parameters in Supplementary Tables 4.5 and 4.6).

4.2.3 Cytokine Secretion Model

To provide an additional, longer-term readout of T cell activation, we extended the T cell model to include signaling events that follow TCR activation culminating in the production of cytokines, particularly IFN-γ. These events include the recruitment of kinases such as Lck and ZAP-70, the activation of intermediate signaling molecules such as phospholipase C and calcineurin, and ultimately the activation of transcription factors NF-AT, NF-κB, and AP-1 (reviewed in Liu 2005). We developed a simplified model of these events representing transcription factor activation, cytokine gene expression, and cytokine production (Fig. 4.1D). More detailed features of T cell signaling such as the synthesis and breakdown of transcription factor intermediates were assumed to have a negligible effect on long-term (> 12 h) responses and therefore not considered in our model; such features are considered in other models (Hoffmann *et al.* 2002, Fisher *et al.* 2006). The cytokine production model comprises a third set of ODEs in addition to the sets of ODEs constituting the APC and T cell models (equations provided in Supplementary Information, parameters in Supplementary Tables 4.5 and 4.6).

4.2.4 Solving the Multi-Scale Model

Together these three models constitute the multi-scale model of antigen presentation. The models were run in three sequential phases: (1) exposure of APC to IFN-γ in the absence of exogenous antigen for 24 h (APC model only); (2) exposure of APC to exogenous antigen in the absence of IFN-γ for 4 h (APC model only); and (3) exposure of APC to T cell for 24 h (T cell model and cytokine production models only). Information was passed between APC and T cell models in the form of number of pMHC on the APC surface appearing 4 h after APC exposure to antigen, i.e., in a feed-forward manner. Feedback from T cell to APC, in the form of IFN-γ that could increase MHC expression, was assumed to be negligible on the timescales being simulated (≤ 24 h) and

therefore not represented; such feedback may be easily accommodated by the model if longer timescales are investigated. The three variables serving as outputs were the number of pMHC complexes on the APC surface 4 h after exposure to exogenous antigen (APC model), fraction of TCR internalized 5 h after APC-T cell contact (T cell model), and concentration of cytokine IFN-γ produced 24 h after APC-T cell contact (cytokine production model), which we consider indicative of short-, medium-, and long-term responses, respectively. These outputs have been considered intermediate indicators of cellular response in the experimental literature (Valitutti *et al.* 1995, Hemmer *et al.* 1998, Itoh *et al.* 1999). A minimum threshold of approximately 200-350 pMHC on the APC surface is needed to elicit a T cell response, though numbers may vary between tens and thousands (Demotz *et al.* 1990, Harding and Unanue 1990). Internalization of approximately 10-90% of TCRs from the T cell surface occurs within hours of APC contact depending on the amount of antigen initially present (Valitutti *et al.* 1995, Itoh *et al.* 1999). No threshold level of TCR internalization required for T cell activation has been determined, though a correlation with other responses such as T cell proliferation has been observed (Itoh *et al.* 1999). Cytokine IFN-γ production by T cells in response to antigenic stimulation varies over several logs, though pM amounts *in vitro* are typical (Hemmer *et al.* 1998, Laaksonen *et al.* 2003, Listvanova *et al.* 2003). The ODEs constituting the multi-scale model were solved using the NDSolve function of Mathematica 4.2 (Wolfram Research, Inc.) using default options, and the model was tested against experimental dose-response data available for each of the three outputs (Valitutti *et al.* 1995, Hemmer *et al.* 1998, Itoh *et al.* 1999).

4.2.5 Sensitivity Analysis of the Multi-Scale Model

We determined how variability in the processes represented in the model affects the three model outputs using sensitivity analysis. Briefly, we varied parameter values in

the model, generated an output for each different set of parameter values, and then determined the degree of correlation between each parameter and the output. A particular parameter was varied if a genetic polymorphism was known to exist for the process represented or to ensure that each of three constituent models had approximately the same number of parameters represented during the analysis. A total of 16 parameters were varied (Table 4.7 in Supplementary Information). Parameters were assigned log-uniform distributions. That is, minimum and maximum values were assigned to each parameter, and sampling was done uniformly on a range defined by the log-transform of these values. When several biological values were available in the literature the approximate range of these values was used: for pMHC dissociation, 10^{-6}-10^{-2} s^{-1} (Rothbard and Gefter 1991, McFarland and Beeson 2002); for pMHC-TCR dissociation, 10^{-3}-10^{0} s^{-1} (Davis *et al.* 1998); for IFN-γ dose, 10^{-12}-10^{-6} M (c.f. Lin *et al*. 1996); and for antigen dose, 10^{-9}-10^{-4} M. For all other cases, a range of one order of magnitude above and below the baseline value was specified (Supplementary Information). 500 values for each parameter were generated by a Latin hypercube sampling scheme (LHS, Helton and Davis 2000), resulting in 500 different sets of parameter values. An equivalent number of output values were then derived, and correlations between output values and parameter values were then quantified by using partial rank correlation coefficients (PRCC, Blower and Dowlatabadi 1994). Significance was assigned based on a Bonferroni-corrected α value of 0.05 (Bland and Altman 1995).

4.2.6 Experimental Scenarios Simulated

Using the multi-scale model of antigen presentation, we were able to simulate *in vitro* protocols intended to test responsiveness of host cells to particular antigens (as used in Katial *et al.* 1998, e.g.). In such a protocol peripheral blood mononuclear cells (PBMCs) are isolated from patient blood, exposed to antigens such as purified protein

derivative (PPD) from *Mtb*, and then assayed for response either indirectly (by measuring tritiated thymidine uptake as a marker of proliferation) or directly (by enzyme-linked immunsorbent assay, or ELISA, for cytokine). Because monocytes, a precursor of macrophages, serve as APCs in PBMC and there is likely little to no IFN-γ present in the blood, the model can simulate PBMC protocols when the amount of IFN-γ initially present is set to zero. The model can also simulate the *in vivo* scenario of antigen presentation at a site of infection where macrophages and activated T cells are present. In this case, the amount of IFN-γ initially present is set at a non-zero value in the model. Both of these scenarios were examined during sensitivity analysis.

4.2.7 Trade-Off Plots

In addition to performing sensitivity analysis by varying several parameters concurrently, we also examined the relationship between processes represented in the model in a pair-wise manner, by varying two parameters at a time. Pairs of parameter values that yielded approximately the same target output value were compiled and plotted. Because such plots show how a change in one parameter is able to compensate for a change in another parameter, we refer to such plots as trade-off plots. When values for both parameters are plotted as log-transforms, regions in which the plots are diagonal (slope approximately 1 or -1) identify conditions under which a compensatory relationship exists. That is, a one-log change in one parameter is able to compensate for a one-log change in another parameter to maintain a given output value. In contrast, regions in which the plots are horizontal or vertical identify conditions under which one parameter has dominant effect on the output over the other parameter. In such regions the output is relatively insensitive to changes in one of the two parameters. Parameters that were chosen to generate trade-off plots were either from the same-scale sub-model (the intra-model case) or from different-scale sub-models (the inter-model case). A

sufficiently wide range of values was assigned to each parameter in generating each trade-off plot to capture the full range of behaviors in each curve. Biologically realistic values were then overlaid on each plot, shown as boxes whose edges represent the range of values from *in vitro* measurements when available. When *in vitro* measurements were not available, a range of one order of magnitude above and below the baseline value was specified (Tables 4.3-4.6 in Supplementary Information). When pMHC dissociation rate constant or pMHC-TCR dissociation rate constant was varied, parameter values were plotted as their respective equilibrium dissociation constants K_D, assuming invariant association rate constants (Kasson *et al.* 2000). We chose different values for the three outputs to serve as target output values, generally accepting pairs of parameter values that resulted in output values between 80% and 120% of the target output values. Target output values were 100, 500, or 1000 pMHC on the APC surface; 10%, 40%, or 80% internalization of total TCR; and 0.1 pM, 1 pM, or 5 pM IFN-γ production corresponding to ~2, ~20, and ~200 pg/ml IFN-γ. To assist visualization of plots, curve-fitting was done using the SplineFit function (Bezier option) of the NumericalMath library in Mathematica 4.2 (Wolfram Research, Inc.), except in cases where more than one *y*-value mapped to the same *x*-value (pMHC-TCR affinity vs. pMHC affinity plots and TCR internalization vs. pMHC affinity plots) when curves were fit by hand.

4.3 Results

To relate genetic polymorphisms to changes in APC and T cell responses, we developed a multi-scale model of antigen presentation that traverses several biological and temporal scales (from molecular to multi-cellular and seconds to hours). This model represents several different immunological processes that could potentially vary due to genetic polymorphisms and allows us to examine the effect of multiple polymorphisms occurring simultaneously.

Initially we tested the model by comparing three model outputs to expected behaviors from experimental data. As a negative control we checked baseline results of the model: In the absence of exogenous antigen, no exogenous peptide-MHC complexes were formed, no TCRs were internalized, and no IFN-γ was produced (data not shown). As a positive control, we examined outputs of the model when exogenous antigen was present. The dynamics of pMHC display, TCR internalization, and IFN-γ production approximated experimentally observed time courses (Fig. 4.2A-C). Specifically, the number of pMHC on the APC surface peaked within four hours of antigen exposure (Fig. 4.2A, Harding and Unanue 1990); the majority of TCR internalization occurred within the first two hours of T cell exposure to APC (Fig. 4.2B, Valitutti *et al.* 1995); and IFN-γ production continued to rise through the first 24 h of T cell exposure to APC (Fig. 4.2C, Listvanova *et al.* 2003). The model also recapitulated dose-response data available for the various outputs (Fig. 4.2D-F).

4.3.1 Sensitivity of T Cell Response to Genetically Variable Processes

To determine how biological variability due to genetic polymorphism or other causes might affect APC and T cell responses, we simulated variability in the multi-scale model and correlated changes in output variables to changes in input parameters (Table 4.1). These outputs were found in either the same sub-model as the parameter being varied (the intra-model case) or in a different sub-model (the inter-model case). Two scenarios were simulated, the absence and presence of IFN-γ initially, scenarios representing antigen presentation during PBMC protocols and at the site of infection, respectively.

Multiple parameters were found to correlate significantly with model outputs, identifying biological processes that may positively or negatively govern antigen

presentation and T cell response. Variability in a number of these processes is known to exist and in some cases has been associated with susceptibility to TB. For example, both peptide-MHC binding affinity and IFN-γ expression (as IFN-γ dose) correlated significantly with all three outputs (Table 4.1). Other processes bear polymorphisms that may affect their level of expression or function but have not previously been associated with TB susceptibility. Antigen processing correlated significantly with all three outputs but more strongly at early time points than later time points (Table 4.1). Likewise, MHC expression correlated significantly with all three outputs, more strongly at early time points than at later time points, but only in the absence of IFN-γ, a scenario resembling PBMC protocols rather than infection *in vivo*, illustrating the overlapping effects of changes in IFN-γ expression and MHC expression (Table 4.1).

Most parameters displayed a similar degree of influence on both T cell responses of TCR internalization and IFN-γ production (Table 4.1). One exception was the rate constant for the internalization of free, activated TCR which correlated positively with TCR internalization and negatively with IFN-γ production. In the model internalized TCRs are incapable of initiating signal transduction and therefore cease to contribute to cytokine production.

4.3.2 Possible Confounding Effects Among Multiple Polymorphisms

Sensitivity analysis demonstrated that multiple processes, including several that may vary due to genetic polymorphisms, govern the dynamics of antigen presentation and subsequent T cell responses. To examine interactions between polymorphisms in more detail, we varied parameters in a pair-wise manner and determined the extent to which one parameter could compensate for another in governing the dynamics of antigen presentation and T cell response.

IFN-γ expression and HLA binding polymorphisms can be compensatory. The polymorphisms in the antigen presentation pathway most commonly associated with TB susceptibility affect the level of IFN-γ expressed by T cells and peptide-binding by MHC (Bellamy 2005). The consequences of polymorphisms in two genes acting simultaneously on antigen presentation have not been examined either experimentally or theoretically to our knowledge. To simulate these polymorphisms, we varied parameters for IFN-γ levels and peptide-MHC binding affinity in the APC model and plotted those pairs of parameter values resulting in approximately the same output levels (Fig. 4.3A-C).

In the trade-off plots, three distinct regions can be discerned (described here for Fig. 4.3B, the case of TCR internalization). First, at low IFN-γ concentrations ($< 10^{-10}$ M), TCR internalization is determined almost entirely by pMHC affinity and is invariant to small changes in IFN-γ concentration, apparent as nearly vertical lines on the plots. Under these conditions few of the IFN-γ receptors are bound, and small changes in IFN-γ concentrations do not alter MHC expression. Secondly, at intermediate IFN-γ concentrations (between 10^{-10} M and 10^{-6} M), IFN-γ has an effect on TCR internalization nearly equal to the effect of pMHC affinity, apparent as diagonal lines on the plots. In this region, for example, 80% TCR internalization can be achieved by pairing either 10^{-9} M IFN-γ and 10^{-9} M pMHC binding affinity (as K_D) or 10^{-8} M IFN-γ and 10^{-8} M pMHC binding affinity. Finally, at high IFN-γ concentrations ($> 10^{-6}$ M), TCR internalization is again determined almost entirely by pMHC affinity, apparent as nearly vertical lines on the plots. Under these conditions most of the IFN-γ receptors are bound, and small changes in IFN-γ concentrations do not affect near-maximal increases in MHC expression.

Superimposing experimental data on these plots allows realistic regions to be defined. IFN-γ expression in PBMC from individuals with +874A and +874T alleles have been measured and found to differ by as much as 3-fold, in the range of 10^{-10}-10^{-11} M (Pravica *et al.* 1999, Lopez-Maderuelo *et al.* 2003, I. Aguilar-Delfin, personal

communication). A wider range of IFN-γ concentrations (between 10^{-9} and 10^{-12} M) is typically applied *in vitro*; this wider range is shown as boxes in Fig. 4.3A-C. Likewise, the affinities of different peptide-MHC class II complexes have been measured and found to vary largely between 10^{-9} and 10^{-6} M (Rothbard and Gefter 1991, Peters *et al.* 2005). These measurements make it possible to define a region on the trade-off plots in which realistic combinations of parameters could be expected. The plots show that at realistic levels of IFN-γ expression, pMHC affinity has a stronger effect on all outputs, from number of pMHC displayed through amount of cytokine produced.

HLA expression and HLA binding polymorphisms can be compensatory. Though polymorphisms in HLA promoters have been identified, none have yet been associated with susceptibility to TB (Louis *et al.* 1994, Cowell *et al.* 1998). One reason may be the difficulty involved in measuring the total level of expression of a particular HLA class II variant within and on the surface of an APC simultaneously. Another reason may be the difficulty involved in attributing an association with TB to the HLA promoter and not the HLA coding sequence with which it is likely in linkage disequilibrium. In the model HLA expression and binding affinity are separate parameters and were found to exert a nearly equivalent influence on output values (Fig. 4.3D-F). For instance, a pMHC affinity of 10^{-9} M when 10^5 MHC molecules were present resulted in nearly the same degree of TCR internalization (~80%) as a weaker pMHC affinity of 10^{-8} M when more (10^6) MHC molecules were present (Fig. 4.3E). At lower levels of MHC expression ($< 10^5$ MHC molecules per APC), however, pMHC affinity becomes much more determinative of T cell response. The possibility that higher levels of expression might compensate for lower affinity binding has been raised previously in non-human studies (Kaufman and Salomonsen 1997, Wegner *et al.* 2006).

Antigen processing and HLA binding polymorphisms can be compensatory. Like polymorphisms affecting MHC expression, polymorphisms affecting antigen processing have also been identified, though none have yet been associated with susceptibility to TB

(Taggart 1992). Polymorphisms affecting antigen processing can be expected to either increase or decrease the availability of antigenic peptides available to bind MHC and therefore affect antigen presentation and subsequent T cell responses. In the model variability in antigen processing was found to compensate for variability in pMHC affinity (Fig. 4.3G-I). For example, to elicit 5 pM IFN-γ production, an increase in the rate constant for antigen processing (from 10^0 to 10^1 h^{-1}) could be coupled with a decrease in pMHC binding affinity (from 10^{-9} to 10^{-8} M, Fig. 4.3I). The extent to which polymorphisms in antigen processing cathepsins affect enzymatic activity is not known (Taggart 1992), but within a one-log range of the level of activity observed *in vitro*, the trade-off plots show that variability in cathepsin activity may affect antigen processing and T cell response to the same extent as variability in pMHC affinity.

Optimal pMHC-TCR affinity affects TCR internalization, not IFN-γ secretion. The binding affinity of the pMHC-TCR tri-molecular complex has been shown to be an important quantity in determining T cell response (Matsui *et al.* 1994). We examined trade-offs between peptide-MHC and pMHC-TCR affinities in eliciting different responses (Fig. 4.4A-C). Because the parameter for pMHC-TCR affinity does not occur in the APC model, variability in pMHC-TCR affinity does not affect pMHC numbers. This lack of effect is apparent as vertical lines on the trade-off plot for this output (Fig. 4.4A). Coombs *et al.* (2002) and Gonzalez *et al.* (2005) showed that under certain conditions an optimal half-life for pMHC-TCR interaction exists resulting in maximal TCR internalization. Because our model of the T cell was based on the model of Coombs *et al.* (2002), it was not surprising to see an optimal binding affinity for pMHC-TCR appear on the trade-off plot for TCR internalization (Fig. 4.4B). However, the peak showing this optimal affinity was greatly lessened at lower pMHC affinities, particularly when IFN-γ production was considered the output (Fig. 4.4C). Indeed, at biological values (10^{-9}-10^{-6} pMHC K_D), pMHC affinity was more determinative of T cell response

than pMHC-TCR affinity, apparent as vertical lines on the plots (Fig. 4.4B and 4.4C, boxed regions).

Internalization of activated TCR is oppositely correlated with different T cell responses. We also examined trade-offs between peptide-MHC affinity and the rate constant for TCR internalization (Fig. 4.4D-F). While most parameters in the model were correlated consistently (i.e., either positively or negatively) with the three different responses, the parameter for internalization of free, activated TCR differed in that it was positively correlated with one response, TCR internalization, and negatively correlated with another, IFN-γ production (Fig. 4.4E-F, c.f. Table 4.1). This effect persisted up to a certain threshold value for the internalization rate constant (~1 h^{-1}), above which other processes such as pMHC binding became limiting. These results were obtained under the assumption that internalized TCR do not continue to signal. This assumption has previously been challenged for TCRs as well as for other receptors (Luton *et al.* 1997, Burke *et al.* 2001). If internalized TCRs are assumed to continue signaling in the model, then vertical trade-off plots with pMHC affinity are observed and TCR internalization has little effect on IFN-γ production (data not shown).

4.4 Discussion

A large body of epidemiological literature links polymorphisms in various host genes to increased susceptibility to TB (c.f. Bellamy 2005, Fernando and Britton 2006, Hill 2006). Mechanistic explanations are still lacking, however, for how the polymorphisms identified in the epidemiological literature increase susceptibility to TB. We posed a fundamental question: how do polymorphisms in multiple genes acting simultaneously affect immune functions such as antigen presentation? For example, considering that IFN-γ up-regulates MHC expression, could an allele of IFN-γ increase

the number of MHC molecules per APC enough to offset deficiencies in binding exhibited by a particular HLA allele to elicit the same T cell response?

To approach these questions, we developed a multi-scale model of antigen presentation that links molecular and intracellular events to cellular and multi-cellular outcomes. By varying parameters for IFN-γ expression, peptide-MHC binding, and other processes, we were able to simulate changes in different processes that might occur due to genetic variation or other causes and then analyze the sensitivity of antigen presentation and T cell responses to these changes. Sensitivity analysis showed that many of the processes in the model exerted strong and comparable influences on the outputs. For instance, both IFN-γ expression (as represented by the amount of IFN-γ to which APC were initially exposed) and peptide-MHC binding were found to significantly affect all outputs in the model, both at the same scale (within the APC, intra-scale) and at different scales (within the T cell, inter-scale). These outputs included the number of pMHC appearing on the APC surface, the degree of TCR internalization, and the amount of cytokine produced by T cells (Table 4.1).

We then analyzed interactions between genetically variable processes in more detail using trade-off plots and found that changes in these processes may compensate for one another. Furthermore, we determined conditions under which such compensatory relationships may exist. For instance, within a certain range of concentrations (10^{-10}-10^{-6} M), alterations in the amount of IFN-γ to which APCs were exposed affected T cell response as strongly as alterations in pMHC affinity (Fig. 4.3B and 4.3C). Outside of this range, however, pMHC affinity had a more dominant effect on T cell response, minimizing the contribution of IFN-γ. In primary cultures of PBMC re-stimulated with antigen, IFN-γ has been detected at concentrations of 10^{-11} to 10^{-10} M. At these concentrations polymorphisms in HLA may mask the effects of polymorphisms in IFN-γ. This interaction may account for inconsistencies in the epidemiological association data. The +874A IFN-γ polymorphism results in decreased IFN-γ expression and has been

associated with susceptibility to TB in some but not all studies (e.g., Moran *et al.* 2007). Could variability in peptide-binding exhibited by different HLA alleles be masking the effect of IFN-γ polymorphisms in such studies? Jepson *et al.* (1997) found that variability in the immune response to TB antigens was the result of variability in both non-HLA genes (such as IFN-γ) and HLA genes. Given the significant presence of HLA polymorphisms in human populations, our study suggests that the accuracy and consistency of association studies could be increased by comparing the frequencies of concurrent pairs of polymorphisms (such as IFN-γ +874A / HLA-DRB1*1501) in TB patients rather single polymorphisms alone.

We also found that polymorphisms need not affect the same cell or the same timescale (intra-scale) to be compensatory. Parameters affecting different scales (inter-scale) may be compensatory as well. For instance, peptide-MHC affinity and pMHC-TCR affinity have a compensatory relationship, though the first affects APCs while the second affects the interface between APC and T cell. Because TCRs are generated by somatic recombination, TCRs do not exist in the human population as alleles, though an individual can be expected to express a diverse set of TCRs, each differing in its affinity for a given pMHC ligand (Davis *et al.* 1998). The importance of pMHC-TCR affinity in determining T cell response has been demonstrated experimentally (Matsui *et al.* 1994, McMahan *et al.* 2006). Previous models have suggested that trade-offs can exist between pMHC affinity and pMHC-TCR affinity, but the conditions under which changes in one of these processes can compensate for changes in the other process has not been previously defined (Eberl *et al.* 1995, Agrawal and Linderman 1996). Verification of our results awaits measurements made on pMHC-TCR tri-molecular complexes.

In the future we hope to consider additional questions regarding the dynamic interplay between host and pathogen. A threshold minimum number of pMHC exists to elicit a T cell response (Demotz *et al.* 1990, Harding and Unanue 1990), and presumably many combinations of parameter values (e.g., for antigen dose, IFN-γ expression, and

pMHC affinity) yield this threshold. Plotting these combinations of values would produce a surface (shown in gray in Fig. 4.5), above which all parameter values would yield a successful T cell response (Fig. 4.5, point 1). Within the span of an infection, either host or pathogen may alter one or more processes underlying these parameters. The pathogen, for example, might produce less antigen thereby providing a temporary advantage (Fig. 4.5, point 2). The host might then respond, increasing the rate of another process, leading to a point placed on the other side of the surface (Fig. 4.5, point 3). This dynamic interplay would resemble the “cycle of antigen frustration” hypothesized to occur during TB (Murray 1999). A path traced by these points on both sides of the surface would represent this cycle, and the final point of this path, the resolution of the cycle, resulting in either a successful immune response to TB or disease.

Figures

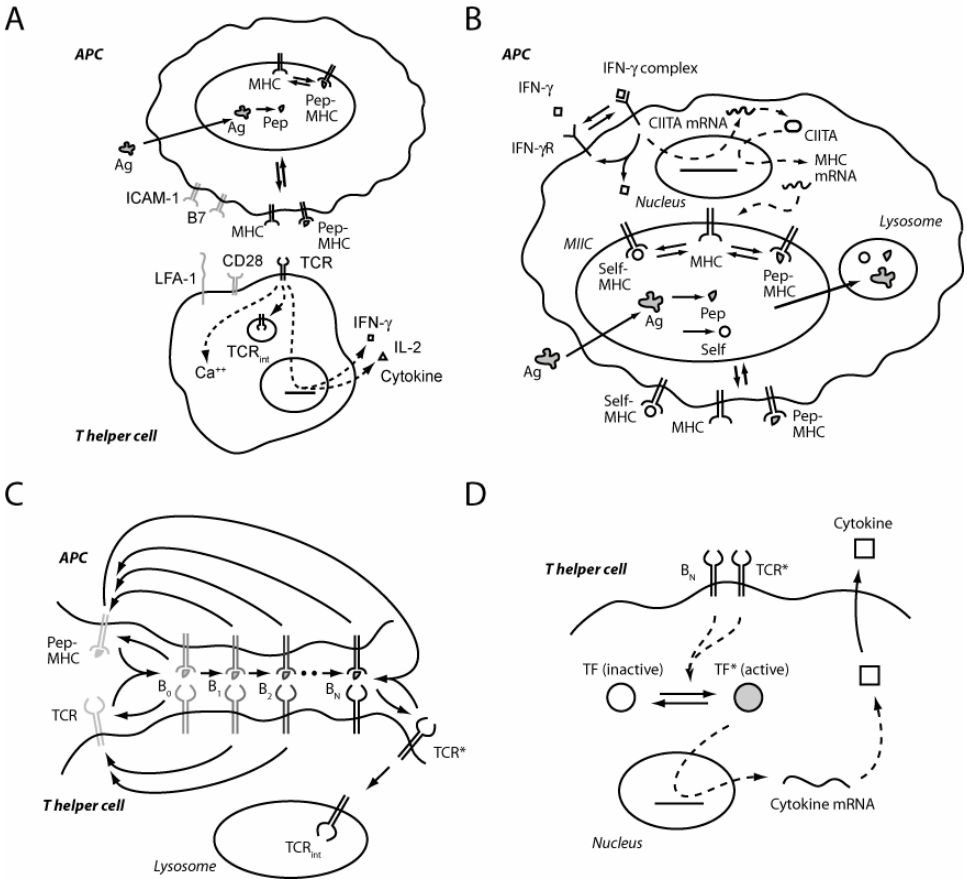

Figure 4.1. Schematic of multi-scale model of antigen presentation and T cell response. (A) Overview of antigen presentation by APC and T cell response. (B) APC model (input: IFN-γ, exogenous antigen; output: surface pMHC). (C) T cell model (input: surface pMHC from APC model; outputs: activated TCR, internalized TCR). (D) Cytokine production model (input: activated TCR; output: cytokine, specifically IFN-γ). Abbreviations are as follows: Ag for antigen, pep for exogenous peptide, self for self peptide, B with subscripts 0 through N for pMHC-TCR complexes in different stages of activation, and TF for transcription factor. Direct (mechanistic) reactions in the model are indicated by solid arrows, while indirect (regulatory) interactions in the model are indicated by dotted arrows. The names of cellular compartments are italicized.

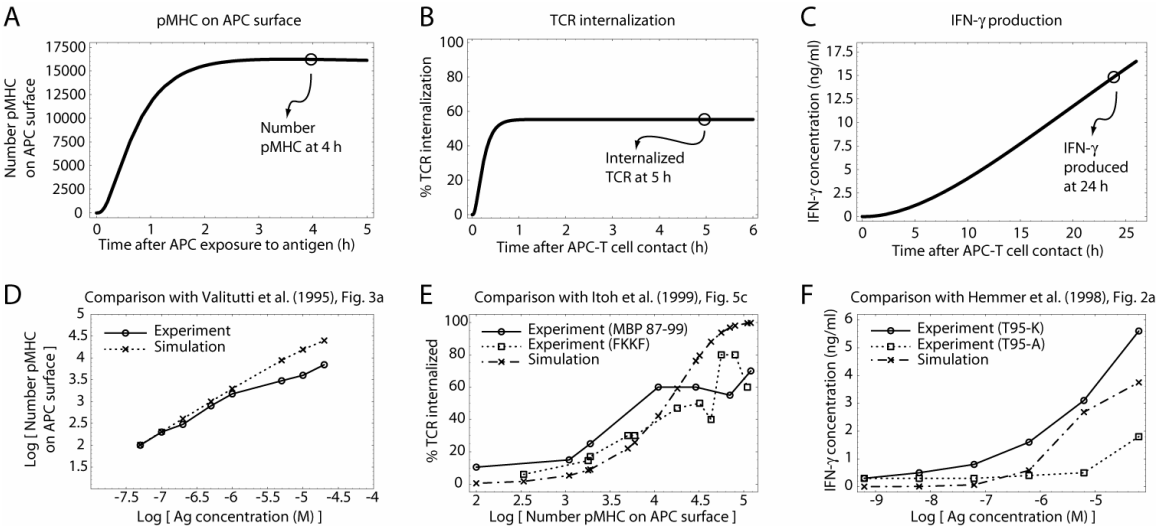

Figure 4.2. The multi-scale model conforms to expected behaviors during testing. (A) Time course of pMHC on APC surface in the model. (B) Time course of TCR internalization in T cells in the model. (C) Time course of IFN-γ production in the model. (D) Dose-response curve for pMHC as antigen concentration is varied in the model and experimental data. (E) Dose-response curve for TCR internalization as the number of peptide-MHC on the APC surface is varied in the model and experimental data. (F) Dose-response curve for IFN-γ production as antigen concentration is varied in the model and experimental data. Parameter values for each simulation are given in Supplementary Information. When more than one curve was available from the experimental data, the highest and lowest non-zero experimental curves were selected and are shown (E and F).

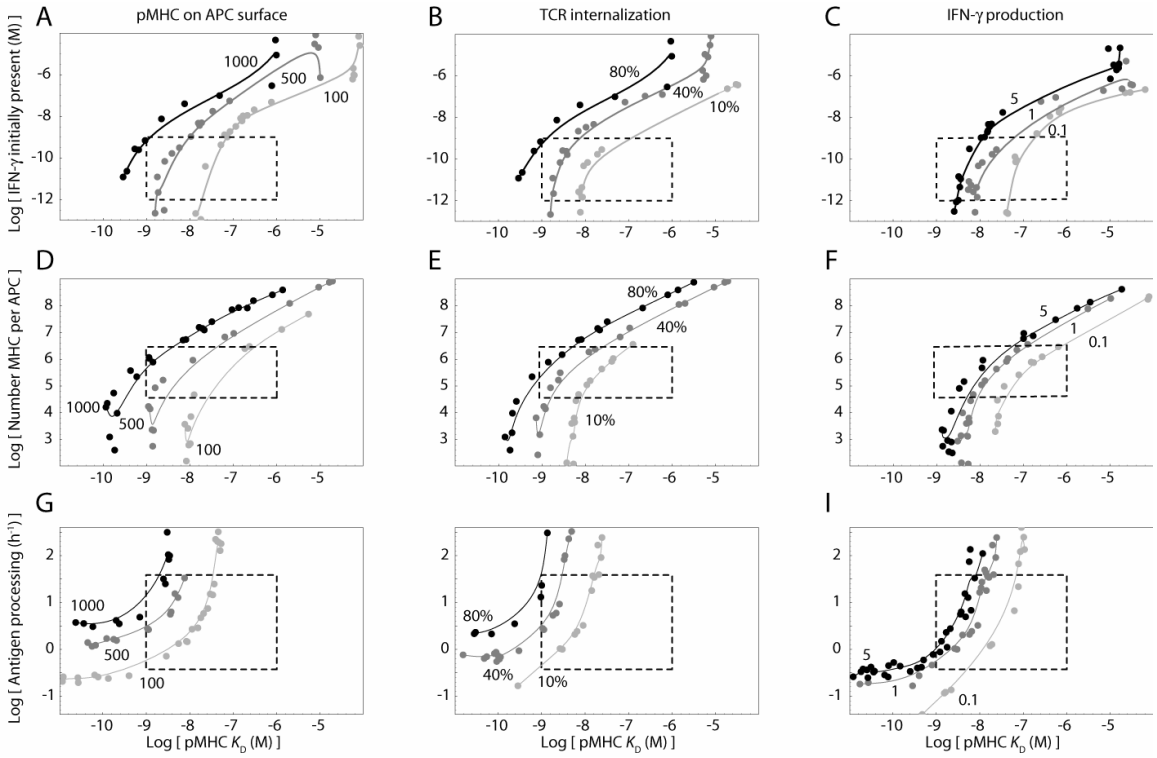

Figure 4.3. Trade-off plots show multiple polymorphisms can compensate for polymorphisms in pMHC binding to achieve the same response. Pairs of parameter values resulting in the same target output value are plotted. (A-C) IFN-γ expression (as amount initially present) vs. peptide-MHC binding. (D-F) MHC expression vs. peptide-MHC binding. (G-I) Antigen processing vs. peptide-MHC binding. Target output values were 100, 500, or 1000 pMHC on the APC surface; 10%, 40%, or 80% internalization of total TCR; and 0.1 pM, 1 pM, or 5 pM IFN-γ production corresponding to ~2, ~20, and ~200 pg/ml IFN-γ. Boxes delineate realistic biological ranges. Values of other parameters used during simulations are provided in Supplementary Information.

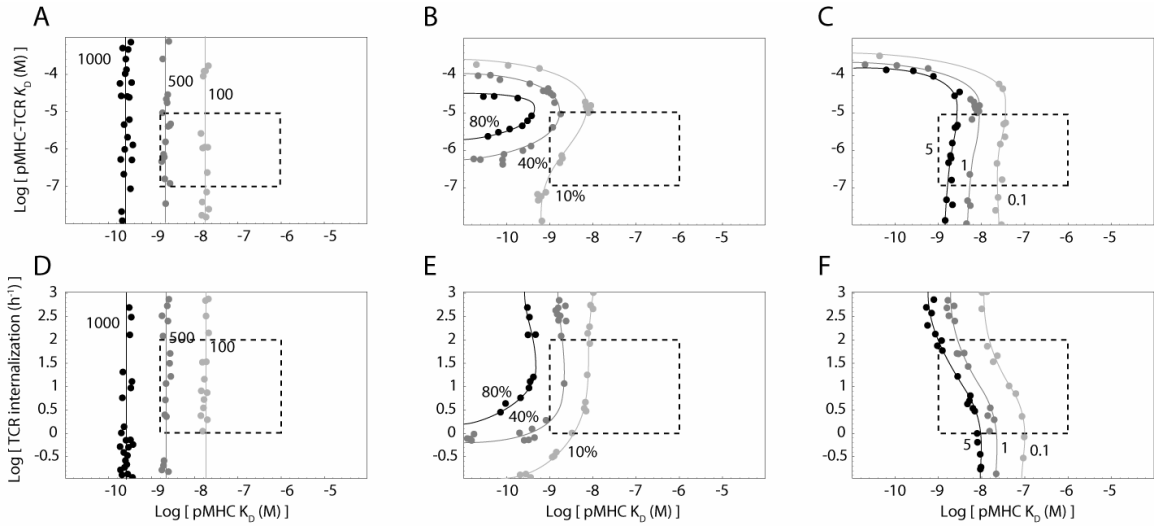

Figure 4.4. Trade-off plots show compensatory relationships between APC and T cell (inter-model) parameters. (A-C) pMHC-TCR affinity vs. peptide-MHC binding. (D-F) Internalization of free, activated TCR vs. peptide-MHC binding. Target output values were 100, 500, or 1000 pMHC on the APC surface; 10%, 40%, or 80% internalization of total TCR; and 0.1 pM, 1 pM, or 5 pM IFN-γ production corresponding to ~2, ~20, and ~200 pg/ml IFN-γ. Boxes delineate reasonable biological ranges. Values of other parameters used during simulations are provided in Supplementary Information.

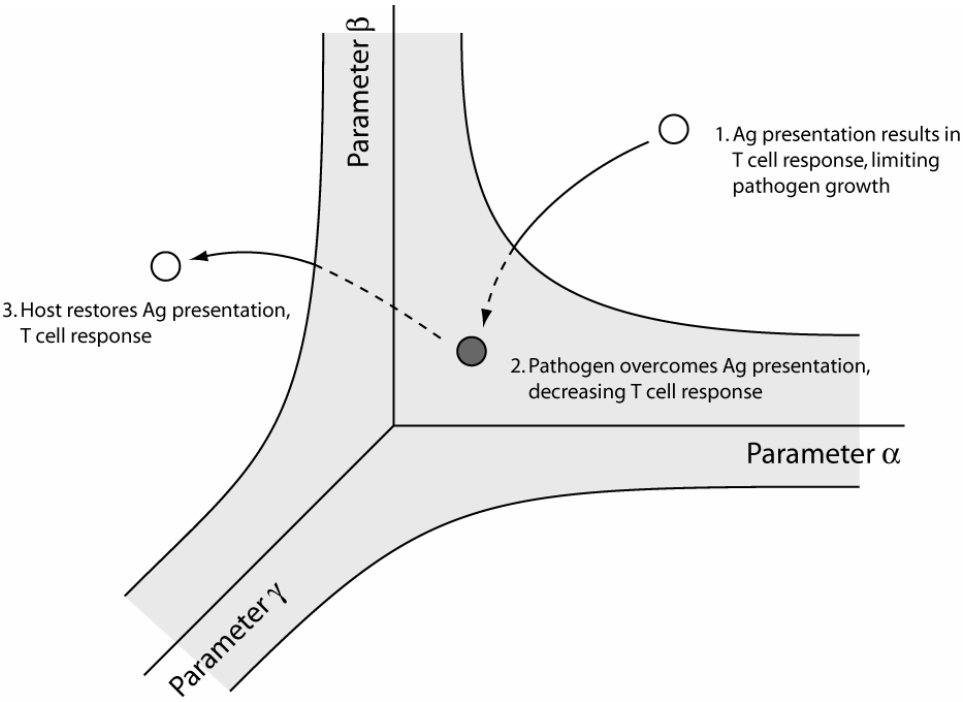

Figure 4.5. Conceptualized multi-dimensional trade-off plot showing how host and pathogen respond to efforts by the other to circumvent and bolster antigen presentation, respectively, during the course of an infection. The surface shown in gray represents all parameters that lead to a threshold number of pMHC on the APC surface or its corresponding T cell response. Points represent combinations of parameter values measured at time points throughout an infection, with points above and below the surface representing a successful immune response and disease, respectively.

Tables

Table 4.1. Sensitivity analysis of the multi-scale model.

	No IFN-γ initially present			IFN-γ initially present		
	pMHC[1]	TCR[2]	IFN-γ[3]	pMHC[1]	TCR[2]	IFN-γ[3]
IFN-γ dose[4]	N/A	N/A	N/A	0.64	0.14	0.15
MHC expression[5]	0.41	0.19	0.15	0.29	(0.07)	(0.05)
pMHC affinity[6]	-0.80	-0.44	-0.40	-0.65	-0.29	-0.28
Ag dose	0.97	0.70	0.68	0.97	0.71	0.72
Ag processing[7]	0.66	0.17	0.16	0.62	0.21	0.24
TCR expression	N/A	0.55	0.42	N/A	0.55	0.34
pMHC-TCR affinity[8]	N/A	-0.58	-0.60	N/A	-0.56	-0.60
pMHC-TCR activation[9]	N/A	0.51	0.49	N/A	0.46	0.46
Activ'd free TCR internaliz.[10]	N/A	(0.08)	-0.24	N/A	(0.07)	-0.23
IFN-γ signaling[11]	N/A	N/A	0.56	N/A	N/A	0.66

PRCC values measuring sensitivity of model outputs to parameter variability when IFN-γ is either not present initially or present initially, akin to PBMC (monocyte-T cell) and infection (macrophage-T helper cell) scenarios, respectively ($n_{cells}=1\cdot10^{6}$, $v_{medium}=1\cdot10^{-3}$ L, $k_{off\text{-}MHC}=2\cdot10^{-3}$ s^{-1}, $k_{on\text{-}TCR}=1\cdot10^{-5}$ $molecule^{-1}s^{-1}$). Shown are PRCC values when 16 parameters in the model were varied (see Methods for detail). Parameters corresponding to processes in which genetic polymorphisms have been observed are indicated in bold. Non-significant PRCC values (α = 0.05, Bonferroni-adjusted) are in parentheses. [1]Number of pMHC on the APC surface 4 h after Ag exposure. [2]Number of TCR internalized by the T cell 5 h after APC-T cell contact. [3]Amount of IFN-γ produced by the T cell 24 h after APC-T cell contact. [4]Amount of IFN- γ to which APCs are exposed 24 h prior to Ag exposure. [5]Number of MHC molecules initially expressed on the cell. [6]As pMHC K_A when pMHC dissociation rate constant was varied. [7]Rate constant for antigen processing. [8]As pMHC-TCR K_A when pMHC-TCR dissociation rate constant was varied. [9]Rate constant for progressive activation of pMHC-TCR complexes. [10]Rate constant for TCR-induced IFN-γ transcription.

4.5 Supplementary Information

4.5.1 APC Model Equations

Equations 4.1-4.16 constitute the APC model and are identical to the equations presented in Chang *et al.* (2005).

$$dG/dt = [-k_{\text{on-IFN-}\gamma}\, G\, R_G + k_{\text{off-IFN-}\gamma} C_G]\, [n_{\text{cells}} / (N_A\, v_{\text{medium}})] - k_{\text{deg-IFN-}\gamma} G \quad [4.1]$$

$$dR_G/dt = -k_{\text{on-IFN-}\gamma}\, G\, R_G + k_{\text{off-IFN-}\gamma}\, C_G + k_{\text{recyc}}\, C_G \quad [4.2]$$

$$dC_G/dt = k_{\text{on-IFN-}\gamma}\, G\, R_G - k_{\text{off-IFN-}\gamma}\, C_G - k_{\text{recyc}}\, C_G \quad [4.3]$$

$$dC_{2m}/dt = k_{\text{txn-C2}}\, (1 + \alpha_{C2}\, C_G) - k_{\text{deg-C2m}}\, C_{2m} \quad [4.4]$$

$$dC_2/dt = k_{\text{tsl-C2}}\, C_{2m} - k_{\text{deg-C2}}\, C_2 \quad [4.5]$$

$$dM_m/dt = k_{\text{txn-M}}\, C_2 - k_{\text{deg-Mm}}\, M_m \quad [4.6]$$

$$dA^*/dt = -(k_{\text{pino}}\, n_{\text{cells}} / v_{\text{medium}})\, A^* - k_{\text{deg-A}^*}\, A^* \quad [4.7]$$

$$dA/dt = (k_{\text{pino}} / v_{\text{MIIC}})\, A^* - k_{\text{deg-A}}\, A - k_{\text{lys}}\, A \quad [4.8]$$

$$dE/dt = k_{\text{deg-A}}\, A + (-k_{\text{on-MHC}}\, M\, E + k_{\text{off-MHC}}\, M_E)\, [1 / (N_A\, v_{\text{MIIC}})] - k_{\text{lys}}\, E \quad [4.9]$$

$$dS/dt = k_{\text{source}} + [k_{\text{deg-MHC}}\, (M_S + M_S^*) - k_{\text{on-MHC}}\, M\, S + k_{\text{off-MHC}}\, M_S]\, [1 / (N_A\, v_{\text{MIIC}})] - k_{\text{lys}}\, S \quad [4.10]$$

$$dM/dt = k_{\text{tsl-M}}\, (1 + \alpha_M\, C_G)\, M_m - k_{\text{on-MHC}}\, M\, S + k_{\text{off-MHC}}\, M_S - k_{\text{on-MHC}}\, M\, E + k_{\text{off-MHC}}\, M_E - k_{\text{out}}\, M + k_{\text{in}}\, M^* - k_{\text{deg-MHC}}\, M \quad [4.11]$$

$$dM^*/dt = k_{\text{out}}\, M - k_{\text{in}}\, M^* - k_{\text{deg-MHC}}\, M^* \quad [4.12]$$

$$dM_S/dt = k_{\text{on-MHC}}\, M\, S - k_{\text{off-MHC}}\, M_S - k_{\text{out}}\, M_S + k_{\text{in}}\, M_S^* - k_{\text{deg-MHC}}\, M_S \quad [4.13]$$

$$dM_S^*/dt = k_{\text{out}}\, M_S - k_{\text{in}}\, M_S^* - k_{\text{deg-MHC}}\, M_S^* \quad [4.14]$$

$$dM_E/dt = k_{\text{on-MHC}}\, M\, E - k_{\text{off-MHC}}\, M_E - k_{\text{out}}\, M_E + k_{\text{in}}\, M_E^* - k_{\text{deg-MHC}}\, M_E \quad [4.15]$$

$$dM_E^*/dt = k_{\text{out}}\, M_E - k_{\text{in}}\, M_E^* - k_{\text{deg-MHC}}\, M_E^* \quad [4.16]$$

Descriptions of the terms in each equation are provided in Chang *et al.* (2005). Variables and parameters are defined, and values provided, in Table 4.3 and 4.4.

4.5.2 TCR Internalization Model Equations

Equations 4.17-4.27 constitute the T cell model and are approximated from the PDEs of Coombs *et al.* (2002) that pertain to the contact zone.

$$dM_E^C/dt = -k_{on\text{-}B}\,(T^C + T_{activ}^C)\,M_E^C + k_{off\text{-}B}\,(B_0 + B_1 + B_2 + B_3 + B_4 + B_5 + B_N) - \lambda_B\,B_N - k_{deg\text{-}MHC,C}\,M_E^C \quad [4.17]$$

$$dT^C/dt = -k_{on\text{-}B}\,T^C\,M_E^C + k_{off\text{-}B}\,(B_0 + B_1 + B_2 + B_3 + B_4 + B_5 + B_N) \quad [4.18]$$

$$dB_0/dt = k_{on\text{-}B}\,T^C\,M_E^C - (k_{off\text{-}B} + k_p)\,B_0 \quad [4.19]$$

$$dB_1/dt = k_p\,B_0 - (k_{off\text{-}B} + k_p)\,B_1 \quad [4.20]$$

$$dB_2/dt = k_p\,B_1 - (k_{off\text{-}B} + k_p)\,B_2 \quad [4.21]$$

$$dB_3/dt = k_p\,B_2 - (k_{off\text{-}B} + k_p)\,B_3 \quad [4.22]$$

$$dB_4/dt = k_p\,B_3 - (k_{off\text{-}B} + k_p)\,B_4 \quad [4.23]$$

$$dB_5/dt = k_p\,B_4 - (k_{off\text{-}B} + k_p)\,B_5 \quad [4.24]$$

$$dB_N/dt = k_{on\text{-}B}\,T_{activ}^C\,M_E^C + k_p\,B_5 - k_{off\text{-}B}\,B_N \quad [4.25]$$

$$dT_{activ}^C/dt = -k_{on\text{-}B}\,T_{activ}^C\,M_E^C + k_{off\text{-}B}\,B_N - \lambda_T\,T_{activ}^C \quad [4.26]$$

$$dT_{int}/dt = \lambda_T\,(T_{activ} + T_{activ}^T) + \lambda_B\,B_N \quad [4.27]$$

Briefly, Equations 4.17-4.19 describe the processes by which free pMHC complexes on the APC surface and free TCRs on the T cell surface bind and form pMHC-TCR tri-molecular complexes. (Superscript C represents molecular species occurring in the contact zone between the APC and T cell.) Equations 4.20-4.25 describe the progressive activation of pMHC-TCR tri-molecular complexes that occurs during kinetic proofreading. Finally, Equations 4.26 and 4.27 describe the association and dissociation of fully activated TCRs to and from pMHC-TCR tri-molecular complexes and the internalization of activated TCR in free or bound forms. In this model, only the contact zone of Coombs et al. (2002) was represented, and therefore terms representing diffusion between the contact zone and other zones in the Coombs model were excluded.

Variables and parameters are defined, and parameter values provided, in Tables 4.5 and 4.6. The model recapitulated major features of the model of Coombs *et al.* (2002) and Gonzalez *et al.* (2005) such as the existence of an optimal pMHC-TCR half-life for TCR internalization (data not shown).

4.5.3 Cytokine Production Model Equations

Equations 4.28-4.31 describe the cytokine production portion of the T cell model.

$$dF_{activ}/dt = k_{resp} (T_{activ} + B_N) F - k_{decay} F_{activ} \quad [4.28]$$

$$F = 1 - F_{activ} \quad [4.29]$$

$$dG_m/dt = k_{txn\text{-}IFN\text{-}\gamma} F_{activ} - k_{deg\text{-}Gm} G_m \quad [4.30]$$

$$dG_2/dt = k_{tsl\text{-}IFN\text{-}\gamma} G_m [n_{cells} / (N_A v_{medium})] - k_{deg\text{-}IFN\text{-}\gamma} G_2 \quad [4.31]$$

Briefly, Equation 4.28 represents the first-order activation and deactivation of a transcription factor for cytokines produced by the T cell, e.g., NF-κB, in units of fraction total transcription factor. Equation 4.29 represents the pool of un-activated transcription factor. Equation 4.30 represents the first-order synthesis (i.e., transcription) and degradation of cytokine mRNA and in particular the absolute dependence of the synthesis of cytokine mRNA on the presence of activated transcription factor. Finally Equation 4.31 represents the first-order synthesis (i.e., translation) and degradation of cytokine protein. Like cytokine mRNA, cytokine protein in the model is completely dependent on the presence of its activator, cytokine mRNA. Variables and parameters are defined, and parameter values provided, in Tables 4.5 and 4.6.

4.5.4 Parameters for Figures and Tables

Parameter values and initial conditions used in solving Equations 4.1-4.31 of the model were as provided in Tables 4.3-4.6 with the following exceptions:

For Fig. 4.2: (a)-(c) $n_{cells}=1 \cdot 10^6$, $v_{medium}=1 \cdot 10^{-3}$ L, $G_0=0$, $A_0=1 \cdot 10^{-5}$ M, $k_{off\text{-}MHC}=2 \cdot 10^{-3}$ s^{-1}, $k_{on\text{-}TCR}=1 \cdot 10^{-6}$ $molecule^{-1}s^{-1}$. (d) $n_{cells}=8 \cdot 10^6$, $v_{medium}=1 \cdot 10^{-3}$ L. (e) $k_{on\text{-}TCR}=1 \cdot 10^{-6}$ $molecule^{-1}s^{-1}$. (f) $n_{cells}=2 \cdot 10^4$, $r_{vol}=2 \cdot 10^{-4}$ L.

For Table 4.1: $n_{cells}=1 \cdot 10^6$, $v_{medium}=1 \cdot 10^{-3}$ L, $k_{off\text{-}MHC}=2 \cdot 10^{-3}$ s^{-1}, $k_{on\text{-}TCR}=1 \cdot 10^{-5}$ $molecule^{-1}s^{-1}$. During LHS parameters that were defined in terms of other parameters (Tables 4.4 and 4.6) were re-derived, with the exception of k_{txn}, the MHC transcription rate constant. Instead, k_{txn} was varied during LHS, and $k_{deg\text{-}Mm}$, the MHC mRNA degradation rate constant, was re-derived.

For Fig. 4.3: $n_{cells}=1 \cdot 10^6$, $v_{medium}=1 \cdot 10^{-3}$ L, $G_0=0$, $A_0=1 \cdot 10^{-5}$ M, $k_{off\text{-}MHC}=2 \cdot 10^{-3}$ s^{-1}, $k_{on\text{-}TCR}=1 \cdot 10^{-6}$ $molecule^{-1}s^{-1}$.

For Fig. 4.4: $n_{cells}=1 \cdot 10^6$, $v_{medium}=1 \cdot 10^{-3}$ L, $G_0=0$, $A_0=1 \cdot 10^{-5}$ M, $k_{off\text{-}MHC}=2 \cdot 10^{-3}$ s^{-1}, $k_{on\text{-}TCR}=1 \cdot 10^{-6}$ $molecule^{-1}s^{-1}$.

For Table 4.7: $n_{cells}=1 \cdot 10^6$, $v_{medium}=1 \cdot 10^{-3}$ L, $k_{off\text{-}MHC}=2 \cdot 10^{-3}$ s^{-1}, $k_{on\text{-}TCR}=1 \cdot 10^{-5}$ $molecule^{-1}s^{-1}$.

Table 4.2. Polymorphisms in antigen presentation affecting susceptibility to TB.

Gene	Allele	Odds ratio	References
HLA class II	DR2 (serotype)	1.8-2.7	Bothamley *et al.* 1989, Brahmajothi *et al.* 1991, Rajalingam *et al.* 1996
	DRB1*1501 (DR2 subtype)	2.7–7.9	Mehra *et al.* 1995, Ravikumar *et al.* 1999, Teran-Escandon *et al.* 1999, Sriram *et al.* 2001
	DQB1*0503	N/A[1]	Goldfeld *et al.* 1998
IFN-γ	+874A	1.6-3.8	Lio *et al.* 2002, Lopez-Maderuelo *et al.* 2003, Rossouw *et al.* 2003
TAP	TAP2*0201	2.4-4.3	Rajalingam *et al.* 1997, Gomez *et al.* 2006

[1]This polymorphism was not detected in the control population.
A more complete list of polymorphisms associated with TB susceptibility can be found elsewhere (Bellamy 2005, Fernando and Britton 2006, Hill 2006). Odds ratio presents a measure of the relative risk associated with each allele.

Table 4.3. Initial values in the APC model.

Variable	Description	Initial value[(1)]
G	IFN-γ concentration in medium	Varies by experiment
R_G	Free IFN-γ receptors per cell	1×10^3
C_G	IFN-γ complexes per cell	0
C_{2m}	CIITA mRNA as fraction of basal level	1
C_2	CIITA protein as fraction of basal level	1
M_m	MHC[(2)] mRNA per cell	1×10^5
A^*	Antigen concentration in medium	Varies by experiment
A	Antigen concentration within MIIC	0
E	Peptide concentration within MIIC	0
S	Self peptide concentration within MIIC	4×10^{-4} M[(3)]
M	Free intracellular MHC per cell	$p_{in}\ (1 - p_{bound})\ M_{tot} \approx 6.7 \times 10^3$
M^*	Free surface MHC per cell	$[(1 - p_{in}) / p_{in}]\ M_0 \approx 1.3 \times 10^4$
M_S	Intracellular self-MHC complexes per cell	$[p_{bound} / (1 - p_{bound})]\ M_0 \approx 2.7 \times 10^4$
M_S^*	Surface self-MHC complexes per cell	$[(1 - p_{in}) / p_{in}]\ M_{S,0} \approx 5.3 \times 10^4$
M_E	Intracellular peptide-MHC complexes per cell	0
M_E^*	Surface peptide-MHC complexes per cell	0

[(1)]When used in the definition of another parameter or variable, the subscript 0 refers to the initial value of a particular variable such that, e.g., M_0 refers to the initial value of M. Units are numbers of molecules per cell (APC or T cell) unless otherwise indicated.
[(2)]MHC in this and following entries refers to MHC class II.
[(3)]This value was estimated from $[k_{\text{deg-Mm}} (M_{\text{S},0} + M_{\text{S},0^*}) + k_{\text{off-MHC}} M_{\text{S},0}] / k_{\text{on-MHC}} M_0$ (Chang *et al.* 2005).

Table 4.4. Parameters in the APC model.

Parameter	Description	Value[(1)]
$k_{\text{on-IFN-}\gamma}$	IFN-γ / IFN-γR association rate constant	3×10^{9} M^{-1} h^{-1}
$k_{\text{off-IFN-}\gamma}$	IFN-γ / IFN-γR dissociation rate constant	7×10^{-1} h^{-1}
n_{cells}	Number of APC in medium	Varies by experiment
v_{medium}	Volume of culture medium	Varies by experiment
$k_{\text{deg-IFN-}\gamma}$	IFN-γ degradation rate constant	1×10^{-2} h^{-1}
k_{recyc}	IFN-γ receptor recycling rate constant	1×10^{1} h^{-1}
$k_{\text{txn-C2}}$	CIITA transcription rate constant	$k_{\text{deg-C2m}}\ C_{2\text{m},0} = 2 \times 10^{-1}$ h^{-1}
α_{C2}	IFN-γ-dependent CIITA scaling factor	1×10^{-1}
$k_{\text{deg-C2m}}$	CIITA mRNA degradation rate constant	2×10^{-1} h^{-1}
$k_{\text{tsl-C2}}$	CIITA mRNA translation rate constant	$k_{\text{deg-C2m}}\ C_2 / C_{2\text{m},0} = 1.4 \times 10^{0}$ h^{-1}
$k_{\text{deg-C2}}$	CIITA degradation rate constant	1.4×10^{0} h^{-1}
$k_{\text{txn-M}}$	MHC transcription rate constant	$k_{\text{deg-Mm}}\ M_{\text{m},0} \approx 4 \times 10^{3}$ h^{-1}
$k_{\text{deg-Mm}}$	MHC mRNA degradation rate constant	4×10^{-2} h^{-1}
k_{pino}	Pinocytosis rate	1×10^{-13} L h^{-1}
$k_{\text{deg-A*}}$	Antigen degradation rate constant in medium	1×10^{-2} h^{-1}
v_{MIIC}	Volume of MIIC compartment	4×10^{-16} L
$k_{\text{deg-A}}$	Antigen processing rate constant	4×10^{0} h^{-1}
k_{lys}	Lysosomal degradation rate constant	6×10^{0} h^{-1}
k_{source}	Self peptide synthesis rate constant	$k_{\text{lys}}\ S_0 \approx 2.4 \times 10^{-3}$ M^{-1} h^{-1}

$k_{\text{deg-MHC}}$	MHC degradation rate constant	2×10^{-2} h^{-1}
$k_{\text{on-MHC}}$	Peptide-MHC association rate constant	7.2×10^{8} M^{-1} h^{-1}
$k_{\text{off-MHC}}$	Peptide-MHC dissociation rate constant	7.2×10^{4} h^{-1}
$k_{\text{tsl-M}}$	MHC mRNA translation rate constant	$k_{\text{deg-MHC}}$ $(M_0 + M^*_0 + M_{S,0} + M_S{}^*_0) \approx 2 \times 10^{-2}$ h^{-1}
α_M	IFN-γ-dependent MHC scaling factor	1×10^{-1}
k_{out}	MIIC-to-surface trafficking rate constant	4×10^{0} h^{-1}
k_{in}	Surface-to-MIIC trafficking rate constant	$[p_{\text{in}}/(1-p_{\text{in}})]$ k_{out} $- k_{\text{deg-MHC}} \approx 2 \times 10^{0}$ h^{-1}
p_{in}	Proportion of MHC intracellular at time 0	1/3
p_{bound}	Proportion of MHC bound to self at time 0	4/5
M_{tot}	Total number of MHC per cell	1×10^{5}

[(1)]When used in the definition of another parameter or variable, the subscript 0 refers to the initial value of a particular variable such that, e.g., M_0 refers to the initial value of M.

Table 4.5. Initial values in the T cell models.

Variable	Description[(1)]	Initial value
M_E^C	Peptide-MHC complexes within contact zone	0
T^C	Free TCR within contact zone, inactive	$(\sigma_C/\sigma_{tot\text{-}Tcell})\ T_{tot} \approx 4.2 \times 10^3$
B_0	Peptide-MHC-TCR complex, inactive	0
B_1	Peptide-MHC-TCR complex, state 1	0
B_2	Peptide-MHC-TCR complex, state 2	0
B_3	Peptide-MHC-TCR complex, state 3	0
B_4	Peptide-MHC-TCR complex, state 4	0
B_5	Peptide-MHC-TCR complex, state 5	0
B_N	Peptide-MHC-TCR complex, activated	0
T_{activ}^C	Free TCR within contact zone, activated	0
T_{int}	Internalized TCR	0
F	Inactive NF-κB, fraction of total NF-κB	1
F_{activ}	Activated NF-κB, fraction of total NF-κB	0
G_m	IFN-γ mRNA	0
G_2	IFN-γ secreted	0

[(1)]Units are numbers of molecules per cell (APC or T cell) unless otherwise indicated.

Table 4.6. Parameters in the T cell models.

Parameter	Description	Value[(1)]
σ_C	Surface area of APC-T cell contact zone	7×10^{-11} m^2
$\sigma_{tot\text{-}APC}$	Total surface area of APC	5×10^{-10} m^2
$k_{on\text{-}B}$	pMHC-TCR association rate constant	3.6×10^{-2} h^{-1} molecule^{-1}
$k_{off\text{-}B}$	pMHC-TCR dissociation rate constant	3.6×10^{1} h^{-1}
$\sigma_{tot\text{-}Tcell}$	Total surface area of T cell	5×10^{-10} m^2
μ	TCR deactivation rate constant	0 h^{-1}
k_p	TCR activation rate constant	9×10^{2} h^{-1}
λ_T	Free TCR internalization rate constant	1.08×10^{1} s^{-1}
λ_B	Bound TCR internalization rate constant	1.08×10^{0} s^{-1}
k_{resp}	NF-κB activation rate constant	5×10^{-3} h^{-1} molecule^{-1}
k_{decay}	NF-κB deactivation rate constant	1×10^{-1} h^{-1}
$k_{txn\text{-}IFN\text{-}\gamma}$	IFN-γ transcription rate constant	$k_{deg\text{-}Gm}\, G_{m,0} \approx 1 \times 10^{2}$ h^{-1}
$k_{deg\text{-}Gm}$	IFN-γ mRNA degradation rate constant	1×10^{-2} h^{-1}
$k_{tsl\text{-}IFN\text{-}\gamma}$	IFN-γ translation rate constant	6×10^{1}
T_{tot}	Total number of TCR per cell	3×10^{4}

[(1)]When used in the definition of another parameter or variable, the subscript 0 refers to the initial value of a particular variable such that, e.g., M_0 refers to the initial value of M.
The values of most parameters are identical to the parameters in Coombs et al. (2002), including surface areas of the APC and T cells, surface area of the contact zone, TCR activation and de-activation rate constants, and TCR internalization rate constants. Association and dissociation rate constants for the pMHC-TCR complex were estimated from values measured *in vitro* (reviewed in Davis *et al.* 1998). The NF-κB activation rate constant was estimated by summing constituent rate constants d4, d5, d6, r4, r5, and r6 from Hoffmann *et al.* (2002). The NF-κB de-activation rate constant was estimated by fitting the time course of activated NF-κB in the model to an experimentally observed peak in NF-κB levels occurring approximately 1 h after activation (Hoffmann *et al.*

2002). The IFN-γ transcription rate constant and mRNA degradation rate constant were estimated by fitting the time course of cytokine IFN-γ mRNA to match an experimentally observed peak in expression approximately 20 hours after exposure to APC (Listvanova *et al.* 2003). The IFN-γ translation rate constant was estimated by fitting the time course of cytokine IFN-γ to match an experimentally observed peak in protein levels detected by ELISA approximately 96 hours after exposure to APC (Listvanova *et al.* 2003).

Table 4.7. PRCC values for all 16 parameters that were varied during sensitivity analysis.

Biological process/factor	No IFN-γ initially present			IFN-γ initially present		
	pMHC[1]	TCR[2]	IFN-γ[3]	pMHC[1]	TCR[2]	IFN-γ[3]
IFN-γ dose[4]	N/A	N/A	N/A	0.64	0.14	0.15
MHC expression[5]	0.41	0.19	0.15	0.29	(0.07)	(0.05)
pMHC affinity[6]	-0.80	-0.44	-0.40	-0.65	-0.29	-0.28
Ag dose	0.97	0.70	0.68	0.97	0.71	0.72
Ag processing[7]	0.66	0.17	0.16	0.62	0.21	0.24
pMHC export to surface	0.53	(0.06)	(0.08)	0.16	(0.05)	(0.06)
pMHC deg. within contact	N/A	-0.26	-0.20	N/A	-0.25	-0.20
TCR expression	N/A	0.55	0.42	N/A	0.55	0.34
pMHC-TCR affinity[8]	N/A	-0.58	-0.60	N/A	-0.56	-0.60
pMHC-TCR activation[9]	N/A	0.51	0.49	N/A	0.46	0.46
Act'd, freeTCR internal.[10]	N/A	(-0.10)	-0.15	N/A	(0.07)	(0.01)
Act'd, bound TCR internal.[11]	N/A	(0.08)	-0.24	N/A	(0.07)	-0.23
IFN-γ signaling[12]	N/A	N/A	0.56	N/A	N/A	0.66
Trans. factor deactivation	N/A	N/A	(-0.04)	N/A	N/A	(-0.07)
IFN-γ mRNA synthesis	N/A	N/A	0.56	N/A	N/A	0.66
IFN-γ mRNA degradation	N/A	N/A	(0.03)	N/A	N/A	(0.03)

Parameters corresponding to processes in which genetic polymorphisms have been observed are indicated in bold. Non-significant PRCC values (α=0.05, Bonferroni-adjusted) are shown in parentheses. N/A is indicated for parameters representing

processes that occur later in the antigen presentation pathway than the indicated output and therefore do not affect output value.

[1]Number of pMHC on the APC surface 4 h after Ag exposure
[2]Number of TCR internalized by the T cell 5 h after APC-T cell contact
[3]Amount of IFN-γ produced by the T cell 24 h after APC-T cell contact
[4]Amount of IFN-γ to which APCs are exposed 24 h prior to Ag exposure
[5]Number of MHC molecules initially expressed on the APC
[6]As pMHC K_D when peptide-MHC dissociation rate constant was varied
[7]Rate constant for antigen processing
[8]As pMHC-TCR K_D when pMHC-TCR dissociation rate constant was varied
[9]Rate constant for progressive activation of pMHC-TCR complexes
[10]Rate constant for internalization of bound, activated TCR
[11]Rate constant for internalization of free, activated TCR
[12]Rate constant for TCR-induced IFN-γ transcription

4.6 References

Agrawal, N.G. and Linderman, J.J. (1996) Mathematical modeling of helper T lymphocyte/antigen-presenting cell interactions: analysis of methods for modifying antigen processing and presentation. *J. Theor. Biol.* **182**, 487-504.

Alcais, A., Remus, N., Abel, L. and Casanova, J.L. (2001) Genetic susceptibility to tuberculosis: from monogenic to polygenic inheritance. *Sepsis* **4**, 237-246.

Bellamy, R. (2005) Genetic susceptibility to tuberculosis. *Clin. Chest Med.* **26**, 233-246.

Bland, J.M. and Altman, D.G. (1995) Multiple significance tests: the Bonferroni method. *BMJ* **310**, 170.

Blower, S.M. and Dowlatabadi, H. (1994) Sensitivity and uncertainty analysis of complex models of disease transmission: an HIV model, as an example. *Int. Stat. Rev.* **2**, 229-243.

Bothamley, G.H., Beck, J.S., Schreuder, G.M., D'Amaro, J., de Vries, R.R., Kardjito, T. and Ivanyi, J. (1989) Association of tuberculosis and *M. tuberculosis*-specific antibody levels with HLA. *J. Infect. Dis.* **159**, 549-555.

Brahmajothi, V., Pitchappan, R.M., Kakkanaiah, V.N., Sashidhar, M., Rajaram, K., Ramu, S., Palanimurugan, K., Paramasivan, C.N. and Prabhakar, R. (1991) Association of pulmonary tuberculosis and HLA in south India. *Tubercle* **72**, 123-132.

Burke, P., Schooler, K. and Wiley, H.S. (2001) Regulation of epidermal growth factor receptor signaling by endocytosis and intracellular trafficking. *Mol. Biol. Cell.* **12**, 1897-910.

Casanova, J.L. and Abel, L. (2002) Genetic dissection of immunity to mycobacteria: the human model. *Annu. Rev. Immunol.* **20**, 581-620.

Chang, S.T., Linderman, J.J. and Kirschner, D.E. (2005) Multiple mechanisms allow *Mycobacterium tuberculosis* to continuously inhibit MHC class II-mediated antigen presentation by macrophages. *Proc. Natl. Acad. Sci. USA* **102**, 4530-4535.

Coombs, D., Kalergis, A.M., Nathenson, S.G., Wofsy, C. and Goldstein, B. (2002) Activated TCRs remain marked for internalization after dissociation from pMHC. *Nat. Immunol.* **3**, 926-931.

Cowell, L.G., Kepler, T.B., Janitz, M., Lauster, R. and Mitchison, N.A. (1998) The distribution of variation in regulatory gene segments, as present in MHC class II promoters. *Genome Res.* **8**, 124-134.

Davis, M.M., Boniface, J.J., Reich, Z., Lyons, D., Hampl, J., Arden, B. and Chien, Y. (1998) Ligand recognition by alpha beta T cell receptors. *Annu. Rev. Immunol.* **16**, 523-544.

Demotz, S., Grey, H.M. and Sette, A. (1990) The minimal number of class II MHC-antigen complexes needed for T cell activation. *Science* **249**, 1028-1030.

Eberl, G., Roggero, M.A. and Corradin, G. (1995) A simple mathematical model for the functional peptide/MHC/TCR interactions. *J. Immunol.* **154**, 219-225.

Fernando, S.L. and Britton, W.J. (2006) Genetic susceptibility to mycobacterial disease in humans. *Immunol. Cell Biol.* **84**, 125-137.

Fisher, W.G., Yang, P.C., Medikonduri, R.K. and Jafri, M.S. (2006) NFAT and NFkappaB activation in T lymphocytes: a model of differential activation of gene expression. *Ann. Biomed. Eng.* **34**, 1712-1728.

Goldfeld, A.E., Delgado, J.C., Thim, S., Bozon, M.V., Uglialoro, A.M., Turbay, D., Cohen, C. and Yunis, E.J. (1998) Association of an HLA-DQ allele with clinical tuberculosis. *JAMA*, **279**, 226-8.

Gomez, L.M., Camargo, J.F., Castiblanco, J., Ruiz-Narvaez, E.A., Cadena, J. and Anaya, J.M. (2006) Analysis of IL1B, TAP1, TAP2 and IKBL polymorphisms on susceptibility to tuberculosis. *Tissue Antigens* **67**, 290-296.

Gonzalez, P.A., Carreno, L.J., Coombs, D., Mora, J.E., Palmieri, E., Goldstein, B., Nathenson, S.G. and Kalergis, A.M. (2005) T cell receptor binding kinetics required for T cell activation depend on the density of cognate ligand on the antigen-presenting cell. *Proc. Natl. Acad. Sci. USA* **102**, 4824-4829.

Harding, C.V. and Unanue, E.R. (1990) Quantitation of antigen-presenting cell MHC class II/peptide complexes necessary for T-cell stimulation. *Nature* **346**, 574-576.

Helton, J.C. and Davis, F.J. (2002) Illustration of sampling-based methods for uncertainty and sensitivity analysis. *Risk Analysis* **22**, 591–622.

Hemmer, B., Stefanova, I., Vergelli, M., Germain, R.N. and Martin, R. (1998) Relationships among TCR ligand potency, thresholds for effector function elicitation, and the quality of early signaling events in human T cells. *J. Immunol.* **160**, 5807-5814.

Hill, A.V. (2006) Aspects of genetic susceptibility to human infectious diseases. *Annu. Rev. Genet.* **40**, 469-486.

Hoffmann, A., Levchenko, A., Scott, M.L. and Baltimore, D. (2002)The IkappaB-NF-kappaB signaling module: temporal control and selective gene activation. *Science* **298**, 1241-1245.

Itoh, Y., Hemmer, B., Martin, R. and Germain, R.N. (1999) Serial TCR engagement and down-modulation by peptide:MHC molecule ligands: relationship to the quality of individual TCR signaling events. *J. Immunol.* **162**, 2073-2080.

Jepson, A., Banya, W., Sisay-Joof, F., Hassan-King, M., Nunes, C., Bennett, S. and Whittle, H. (1997) Quantification of the relative contribution of major histocompatibility complex (MHC) and non-MHC genes to human immune responses to foreign antigens. *Infect. Immun.* **65**, 872-876.

Kasson, P.M., Rabinowitz, J.D., Schmitt, L., Davis, M.M. and McConnell, H.M. (2000) Kinetics of peptide binding to the class II MHC protein I-Ek. *Biochemistry* **39**, 1048-1058.

Katial, R.K., Sachanandani, D., Pinney, C. and Lieberman, M.M. (1998) Cytokine production in cell culture by peripheral blood mononuclear cells from immunocompetent hosts. *Clin. Diagn. Lab Immunol.* **5**, 78-81.

Kaufman, J. and Salomonsen, J. (1997) The "minimal essential MHC" revisited: both peptide-binding and cell surface expression level of MHC molecules are polymorphisms selected by pathogens in chickens. *Hereditas* **127**, 67-73.

Laaksonen, K., Waris, M., Makela M.J., Terho, E.O. and Savolainen, J. (2003) In vitro kinetics of allergen- and microbe-induced IL-4 and IFN-γ mRNA expression in PBMC of pollen-allergic patients. *Allergy* **58**, 62-66.

Lin, Y., Zhang, M., Hofman, F.M., Gong, J. and Barnes, P.F. (1996) Absence of a prominent Th2 cytokine response in human tuberculosis. *Infect. Immun.* **64**, 1351-1356.

Lio, D., Marino, V., Serauto, A., Gioia, V., Scola, L., Crivello, A., Forte, G.I., Colonna-Romano, G., Candore, G. and Caruso, C. (2002) Genotype frequencies of the +874T-->A single nucleotide polymorphism in the first intron of the interferon-gamma gene in a sample of Sicilian patients affected by tuberculosis. *Eur. J. Immunogenet.* **29**, 371-374.

Listvanova, S., Temmerman, S., Stordeur, P., Verscheure, V., Place, S., Zhou, L., Locht, C. and Mascart, F. (2003) Optimal kinetics for quantification of antigen-induced cytokines in human peripheral blood mononuclear cells by real-time PCR and by ELISA. *J. Immunol. Methods* **281**, 27-35.

Liu, J.O. (2005) The yins of T cell activation. *Sci. STKE* **265**, 1-8.

Lopez-Maderuelo, D., Arnalich, F., Serantes, R., Gonzalez, A., Codoceo, R., Madero, R., Vazquez, J.J. and Montiel, C. (2003) Interferon-gamma and interleukin-10 gene polymorphisms in pulmonary tuberculosis. *Am. J. Respir. Crit. Care Med.* **167**, 970-975.

Louis, P., Pinet, V., Cavadore, P., Kerlan-Candon, S., Clot, J. and Eliaou, J.F. (1994) Differential expression of HLA-DRB genes according to the polymorphism of their regulatory region. *C. R. Acad. Sci. III* **317**, 161-166.

Luton, F., Legendre, V., Gorvel, J.P., Schmitt-Verhulst, A.M. and Boyer, C. (1997) Tyrosine and serine protein kinase activities associated with ligand-induced internalized TCR/CD3 complexes. *J. Immunol.* **158**, 3140-3147.

Maher, S.G., Romero-Weaver, A.L., Scarzello, A.J. and Gamero, A.M. (2007) Interferon: cellular executioner or white knight? *Curr. Med. Chem.* **14**, 1279-89.

Matsui, K., Boniface, J.J., Steffner, P., Reay, P.A. and Davis, M.M. (1994) Kinetics of T-cell receptor binding to peptide/I-Ek complexes: correlation of the dissociation rate with T-cell responsiveness. *Proc. Natl. Acad. Sci. USA* **91**, 12862-12866.

McFarland, B.J. and Beeson, C. (2002) Binding interactions between peptides and proteins of the class II major histocompatibility complex. *Med. Res. Rev.* **22**, 168-203.

McMahan, R.H., McWilliams, J.A., Jordan, K.R., Dow, S.W., Wilson, D.B. and Slansky, J.E. (2006) Relating TCR-peptide-MHC affinity to immunogenicity for the design of tumor vaccines. *J. Clin. Invest.* **116**, 2543-2551.

Mehra, N.K., Rajalingam, R., Mitra, D.K., Taneja, V. and Giphart, M.J. (1995) Variants of HLA-DR2/DR51 group haplotypes and susceptibility to tuberculoid leprosy and pulmonary tuberculosis in Asian Indians. *Int. J. Lepr. Other Mycobact. Dis.* **63**, 241-248.

Moran, A., Ma, X., Reich, R.A. and Graviss, E.A. (2007) No association between the +874T/A single nucleotide polymorphism in the IFN-gamma gene and susceptibility to TB. *Int. J. Tuberc. Lung Dis.* **11**, 113-115.

Murray, P.J. (1999) Defining the requirements for immunological control of mycobacterial infections. *Trends Microbiol.* **7**, 366-372.

Peters, B., Sidney, J., Bourne, P., Bui, H.H., Buus, S., Doh, G., Fleri, W., Kronenberg, M., Kubo, R., Lund, O., Nemazee, D., Ponomarenko, J.V., Sathiamurthy, M., Schoenberger, S., Stewart, S., Surko, P., Way, S., Wilson, S. and Sette, A. (2005) The immune epitope database and analysis resource: from vision to blueprint. *PLoS Biol.* **3**, e91.

Pravica, V., Asderakis, A., Perrey, C., Hajeer, A., Sinnott, P.J. and Hutchinson, I.V. (1999) In vitro production of IFN-gamma correlates with CA repeat polymorphism in the human IFN-gamma gene. *Eur. J. Immunogenet.* **26**, 1-3.

Rajalingam, R., Mehra, N.K., Jain, R.C., Myneedu, V.P. and Pande, J.N. (1996) Polymerase chain reaction-based sequence-specific oligonucleotide hybridization analysis of HLA class II antigens in pulmonary tuberculosis: relevance to chemotherapy and disease severity. *J. Infect. Dis.* **173**, 669-676.

Rajalingam, R., Singal, D.P. and Mehra, N.K. (1997) Transporter associated with antigen-processing (TAP) genes and susceptibility to tuberculoid leprosy and pulmonary tuberculosis. *Tissue Antigens* **49**, 168-172.

Rammensee, H. (1995) Chemistry of peptides associated with MHC class I and class II molecules. *Curr. Opin. Immunol.* **7**, 85-96.

Ravikumar, M., Dheenadhayalan, V., Rajaram, K., Lakshmi, S.S., Kumaran, P.P., Paramasivan, C.N., Balakrishnan, K. and Pitchappan, R.M. (1999) Associations of HLA-DRB1, DQB1 and DPB1 alleles with pulmonary tuberculosis in south India. *Tuber. Lung Dis.* **79**, 309-317.

Robinson, J., Waller, M.J., Parham, P., de Groot, N., Bontrop, R., Kennedy, L.J., Stoehr, P. and Marsh, S.G. (2003) IMGT/HLA and IMGT/MHC: sequence databases for the study of the major histocompatibility complex. *Nucleic Acids Res.* **31**, 311-314.

Rossouw, M., Nel, H.J., Cooke, G.S., van Helden, P.D. and Hoal, E.G. (2003) Association between tuberculosis and a polymorphic NFkappaB binding site in the interferon gamma gene. *Lancet* **361**, 1871-1872.

Rothbard, J.B. and Gefter, M.L. (1991) Interactions between immunogenic peptides and MHC proteins. *Annu. Rev. Immunol.* **9**, 527-565.

Santana, M.A. and Esquivel-Guadarrama, F. (2006) Cell biology of T cell activation and differentiation. *Int. Rev. Cytol.* **250**, 217-274.

Singer, D.F. and Linderman, J.J. (1990) The relationship between antigen concentration, antigen internalization, and antigenic complexes: modeling insights into antigen processing and presentation. *J. Cell Biol.* **111**, 55-68.

Singer, D.F. and Linderman, J.J. (1991) Antigen processing and presentation: how can a foreign antigen be recognized in a sea of self proteins? *J. Theor. Biol.* **151**, 385-404.

Small, P.M. and Fujiwara, P.I. (2001) Management of tuberculosis in the United States. *N. Engl. J. Med.* **345**, 189-200.

Sriram, U., Selvaraj, P., Kurian, S.M., Reetha, A.M. and Narayanan, P.R. (2001) HLA-DR2 subtypes and immune responses in pulmonary tuberculosis. *Indian J. Med. Res.* **113**, 117-124.

Taggart, R.T. (1992) Genetic variation of human aspartic proteinases. *Scand. J. Clin. Lab. Invest. Suppl.* **210**, 111-119.

Teran-Escandon, D., Teran-Ortiz, L., Camarena-Olvera, A., Gonzalez-Avila, G., Vaca-Marin, M.A., Granados, J. and Selman, M. (1999) Human leukocyte antigen-associated susceptibility to pulmonary tuberculosis: molecular analysis of class II

alleles by DNA amplification and oligonucleotide hybridization in Mexican patients. *Chest* **115**, 428-433.

Valitutti, S., Muller, S., Cella, M., Padovan, E. and Lanzavecchia, A. (1995) Serial triggering of many T-cell receptors by a few peptide-MHC complexes. *Nature* **375**, 148-151.

Wegner, K.M., Kalbe, M., Rauch, G., Kurtz, J., Schaschl, H. and Reusch, T.B. (2006) Genetic variation in MHC class II expression and interactions with MHC sequence polymorphism in three-spined sticklebacks. *Mol. Ecol.* **15**, 1153-1164.

CHAPTER 5

Conclusions

5.1 Summary

Antigen presentation, one of the key events in the initiation and maintenance of the immune response, offers many possible avenues of research for both experimentalists and theoreticians. In the preceding chapters I described my efforts to pursue three such avenues – how best to represent events at the cellular scale (tracking the number of pMHC on the APC surface), then at the molecular scale (predicting binding between peptides and MHC class II molecules), and finally at the multi-cellular scale (between APC and T cell leading up to T cell activation). After describing how each model was developed, I showed how I applied each model to answer questions related to the immune response to infection, particularly with *M. tuberculosis*.

5.1.1 APC Model

In the case of the cellular-scale model, I asked why *M. tuberculosis* had been found to use multiple mechanisms to inhibit antigen presentation, each targeting a different intracellular process (Fig. 5.1). The APC model suggested that these mechanisms differed in their effects. Some mechanisms were effective at stymieing antigen presentation almost immediately but could be overcome by the provision of an activating cytokine, IFN-γ, by other cells. Targeting the ability of macrophages to process antigens into smaller peptides could be considered one such mechanism. Other

mechanisms required a delay of several hours to have an effect on antigen presentation but were more effective at longer timescales. Mechanisms targeting MHC class II expression, for example, required at least ten hours to have an effect, consistent with the length of time required for protein synthesis. *M. tuberculosis* may therefore employ multiple mechanisms to a complementary rather than merely redundant effect. Furthermore, the application of external IFN-γ in experimental protocols may interfere with the ability to detect *M. tuberculosis* mechanisms that do not target MHC class II expression.

5.1.2 Peptide-MHC Binding Model

At the molecular scale, I asked whether differences in peptide length affected binding affinity to MHC class II, and if so, whether incorporating these differences into existing binding prediction algorithms could improve algorithm performance (Fig. 5.1). After analyzing binding data from currently available databases, I found that a significant and nonlinear relationship existed between length and affinity. This relationship was allele-specific and often revealed an optimal length for maximizing binding affinity, a finding that may have implications for the study of antigens from pathogens such as *M. tuberculosis* or for vaccine design (described in more detail below). Furthermore, incorporating information about peptide length into binding prediction algorithms consistently improved performance, whether through the use of these relationships or alternatively through the use of a statistical method for reconciling predicted affinities made on multiple binding registers (i.e., 9mer windows that engage the MHC class II molecule directly) within longer peptides.

5.1.3 T Cell Model

Finally, at the multi-cellular scale, I asked whether multiple polymorphisms affecting different steps in the antigen presentation pathway might interact, either canceling or intensifying their effects on the outcome. To answer this question, I extended the APC model to include the T cell responses of TCR down-regulation and cytokine secretion and then determined whether and under what conditions a polymorphism in one gene might compensate for a polymorphism in another gene (Fig. *5*.5.1). For example, polymorphisms have been observed in both IFN-γ non-coding and MHC class II coding sequences, affecting IFN-γ expression and pMHC affinity, respectively. In the model, increased IFN-γ expression could compensate for decreased pMHC affinity to maintain the same level of cytokine secretion but only when IFN-γ levels exceeded a lower threshold. Below this threshold, changes in pMHC affinity had a much stronger effect on the level of cytokine secretion. The finding that polymorphisms can have similar effects on antigen presentation could explain discrepancies in the epidemiological literature where some polymorphisms (e.g., the MHC class II allele DRB1*1501) are inconsistently associated with disease susceptibility.

5.2 Future Work: Additions to the Models

5.2.1 More Detailed Representation of *M. tuberculosis*

While antigen presentation is largely the product of host processes, the pathogen ultimately plays a role by supplying antigenic peptides. In each of the sub-models constituting the multi-scale model, *M. tuberculosis* has been represented as a static quantity. For instance, in the APC model, the degree of inhibition of a particular process is assumed to be proportional to the initial number of bacilli to which the APC is exposed.

Adding *M. tuberculosis* as a dynamic variable would allow presentation to be more realistically represented by the model (Fig. 5.1C). Because the number of *M. tuberculosis* bacilli infecting any given APC is discrete and likely to be low but the ODE representation of the APC model assumes continuous variables, the most effective way to incorporate *M. tuberculosis* into the model would be to use some continuously variable quantity as a proxy for intracellular bacterial number. One such quantity might be the amount of a particular antigen secreted by the bacillus.

Adding *M. tuberculosis* as a dynamic variable into the model would also allow feedback to be more realistically represented by the model. Since T cell secretion of IFN-γ was represented in the model and IFN-γ-specific parameter values were used, it is possible to use the output of the T cell model as an additional input for the APC model. In doing so, the model would depict the increased capacity of macrophages at the site of infection to present antigen after being exposed to effector T cells that have been re-stimulated by other macrophages.

However, in addition to its effects on MHC expression, IFN-γ also increases the killing capacity of macrophages. The effect that killing *M. tuberculosis* bacilli has on the availability of antigen is currently unknown. While killing may result in the liberation of proteins formerly contained within the bacilli and increase the antigen pool, killing may also reduce the quantity of proteins that were formerly being secreted and decrease the antigen pool. In either case, the sum effect of additional exposure to IFN-γ may be a higher capacity to present antigen but a changed antigen pool.

By representing the growth state of *M. tuberculosis* and availability of antigen as dynamic quantities, it may be possible to simulate antigen presentation on longer timescales. Currently the multi-scale model represents events to approximately one day. However, since the doubling time of the *M. tuberculosis* bacillus and the lifespan of macrophages are both on the scale of days, antigen presentation may extend to longer timescales than currently represented. Murray (1999) hypothesized that during

tuberculosis a “cycle of antigen frustration” occurs in which periods of greater antigen availability alternate with periods of lesser antigen availability. Each period of greater antigen availability induces a T cell response and antimicrobial activity leading to a period of lesser antigen availability during which the bacilli recover undetected by immune surveillance. A model representing *M. tuberculosis* as a dynamic variable would allow virtual experiments of the hypothesis of Murray (1999) to be performed, such as the addition of inhibitors of mycobacterial growth that would presumably slow the period of oscillations in the cycle. The results could then be used to design new protocols for *in vitro* experiments.

5.2.2 More Detailed Representation of Particular Processes

In the APC and T cell models, most processes were represented with mass-action or first-order kinetics, depending on whether the process involved two molecular species in the model or only one, respectively. While more detailed mechanistic representations are possible for nearly all of the processes in these models, it was assumed that on the timescales of interest (typically hours after exposure to antigen) the differences between such representations and the ones actually used in the models would be negligible. For instance, Witt and McConnell (1992) proposed that peptide-MHC binding is more accurately described using a two-step binding mechanism. However, this mechanism would be expected to produce the same output as the simpler, one-step mechanism hours after the binding reaction, when pMHC on the APC surface was read as output. Likewise, peptide-MHC binding to TCR has also been hypothesized to involve a two-step binding mechanism (Wu *et al.* 2002), but the additional complexity in this model is not expected to yield any difference in output on the timescale of hours. In addition, the data supporting more complicated binding mechanisms has sometimes been controversial, as

in the case of the two-step peptide-MHC binding mechanism which has been disputed (Berezhkovskiy 1998).

Nevertheless, in some cases it may be possible to gain additional insights from the model using more complicated representations, especially when different mechanisms of regulation are possible (Fig. 5.1C). For instance, in the APC model endocytosis was represented in a general way, involving only one intracellular compartment, the MHC class II compartment (MIIC). However, different compartments may play distinct roles during antigen presentation, particularly during infection with *M. tuberculosis*. While *M. tuberculosis* resides in specialized phagosomes where access to MHC class II molecules is impaired, a subset of antigens is transferred to endosomes accessible to MHC class II molecules (Beatty and Russell 2000, Gehring *et al.* 2003). Therefore, additional selectivity is likely conferred by the path that antigens of *M. tuberculosis* traverse within the macrophage, and some antigens may be presented more quickly than others. Additional steps may be introduced into the APC model to account for these disease-specific differences in endocytosis.

5.3 Future Work: Integration with Other Models

5.3.1 Models of Antigen Presentation by MHC Class I

Exceptions to the rule that MHC class I binds antigens from the cytoplasm (i.e., endogenous antigens) while MHC class II binds antigens from the extracellular space (i.e., exogenous antigens) have been identified, and a lack of complete knowledge regarding the mechanisms involved presents an opportunity for models to provide insights.

During cross-presentation, exogenous antigens gain access to MHC class I molecules and are presented as peptide-MHC class I complexes on the APC surface (Rock and Shen 2005). While a detailed mechanism has yet to be elucidated, cross-

presentation appears to follow the internalization of exogenous antigens via phagosomes that later fuse with the endoplasmic reticulum (ER) (Guermonprez *et al.* 2003, Houde *et al.* 2003). Once in the ER exogenous antigens are presumably treated like endogenous antigens and then exported into the cytoplasm, processed by the proteasome, reintroduced into the ER by the transporter associated with antigen presentation (TAP), and bound to MHC class I molecules. Cross-presentation has been shown to occur for antigens from several bacterial pathogens including *M. tuberculosis*.

Antigens from *M. tuberculosis* are therefore likely to be presented on the surface of APCs with both MHC class I and MHC class II molecules. To what extent are antigens presented via one type of pMHC complex versus the other? To answer this question, the APC model may be extended to account for the loss of antigens due to the cross-presentation pathway (Fig. 5.1C). This may be accomplished most directly with additional first-order loss terms in the ODE model, assuming that a constant proportion of antigen is shunted away from MHC class II-accessible compartments.

However, it may be interesting to first determine whether the antigens lost to cross-presentation and the MHC class I pathway are capable of being bound by MHC class II. The two pathways may be non-competing if antigens destined for the cross-presentation pathway are enriched in MHC class I-binding sequences but deficient in MHC class II-binding sequences. Because the antigens of *M. tuberculosis* are normally not considered candidates for binding MHC class I, a study has not been done to identify possible MHC class I-binding sequences within the *M. tuberculosis* proteome, though a similar study has been done to identify MHC class II-binding sequences (McMurry *et al.* 2005). Models that represent selectivity at the steps of TAP binding and pMHC binding are available and could be applied to *M. tuberculosis* proteome (Petrovsky and Brusic 2004).

5.3.2 Larger-Scale Models of the Immune System

Of the three main types of professional APC that are commonly distinguished (dendritic cells, macrophages, and B cells), macrophages are most closely represented by the APC model. The majority of the parameters in the APC model were derived from *in vitro* murine macrophage data, and initial testing was performed against other *in vitro* macrophage data. Because the steps involved in MHC class II-mediated antigen presentation are common to all APC cell types, however, the structure and mathematical representation in the model are likely to be correct for other cell types as well. Indeed, the APC and T cell models were shown to reproduce both time course and dose-response data from a variety of experiments. Dendritic cells have been found to differ from macrophages with respect to two antigen presentation-relevant parameters, the level of MHC expression and rate of antigen uptake (Inaba and Steinman 1985, Inaba *et al.* 1997). Using different values for these parameters may be sufficient to distinguish these two cell types in the APC model as well as in the multi-scale model.

Accounting for differences between dendritic cells and macrophages may help extend the model to two different instances of antigen presentation during the course of an immune response: antigen presentation to naïve T cells in the lymph node by dendritic cells and antigen presentation to effector T cells at a site of infection by macrophages. Both instances are likely to play important roles during the immune response to *M. tuberculosis*, especially given the additional role of macrophages as the preferred host cells to the *M. tuberculosis* bacillus. In conjunction with a more realistic representation of *M. tuberculosis* growth and antigen availability (discussed above, in Section 5.2.1), distinct models of the dendritic cell and the macrophage would allow additional, stage-specific questions to be approached. For example, are the mechanisms by which *M. tuberculosis* inhibits antigen presentation more likely to be effective on the timescales of the initial instance of antigen presentation (when dendritic cells commingle with naïve T

cells shortly after infection) or the timescales of later instances of antigen presentation (when macrophages encounter and receive stimulation from effector T cells)? Likewise, are differences in peptide-MHC affinity resulting from differences between MHC alleles more likely to play a role in initial or later instances of antigen presentation? Comparisons of the results from dendritic cell and macrophage versions of the APC model and larger multi-scale model could provide answers to these questions and experimentally testable predictions.

In addition, representation of additional cell types would help the current model to be integrated with other larger-scale models of antigen presentation (Fig. 5.1C). Models of both the site of infection during tuberculosis and the lymph node have been developed by Kirschner and colleagues (Segovia-Juarez *et al.* 2004, unpublished data). In both of these models, APCs and T cells are represented as discrete objects and their interaction is probabilistic. An APC that displays more pMHC complexes on its surface would presumably either have a higher probability of a successful interaction with a T cell or be able to elicit a stronger T cell response following a successful interaction (Bekkhoucha *et al.* 1984). The probability of a successful APC-T cell interaction is a static quantity in these larger-scale models, but using the output of the multi-scale model of the APC and T cell presented herein (Fig. 5.1B), this probability could be set as a dynamic quantity that more accurately reflects the infection scenario.

5.4 Applications to Disease

5.4.1 Mechanisms of HLA-Disease Association

Several alleles of MHC class II have been found to be either over-represented or under-represented in patients with certain diseases relative to healthy controls and thereby associated with either susceptibility or resistance to these diseases, respectively. For instance, in the case of tuberculosis, MHC class II alleles of the HLA-DR2 and HLA-

DR3 serotypes have been associated with susceptibility and resistance, respectively. Although the mechanisms underlying these associations are not known, several hypotheses exist (Vukmanovic *et al.* 2003). In the case of associations with infectious diseases, the most direct hypotheses concern the ability of the MHC variants to bind peptides from the antigens of the pathogen and elicit an immune response.

Here I show preliminary data for how two possible hypotheses for MHC-TB associations may be tested using statistical analyses and models (Fig. 5.1C). According to one hypothesis, an MHC variant associated with susceptibility may bind a broad range of peptides with lower affinity than MHC variants associated with resistance or other MHC variants in the population. If this were the case then the average affinity for the susceptibility allele, as calculated from a database of pMHC affinities, would be expected to be lower than the average affinity for the resistance allele. In the case of HLA-DR2 and HLA-DR3, the opposite was found to be true: in fact, peptides binding HLA-DR2 do so with a significantly higher average affinity than peptides that bind HLA-DR3, even when affinities for repeated and highly sequence-similar peptides have been removed (Fig. 5.2).

Another hypothesis is that the allele associated with susceptibility, HLA-DR2, binds only disease-relevant peptides with lower affinities than the allele associated with resistance, HLA-DR3. To investigate this possibility, I predicted affinities for all possible 9mers from one *M. tuberculosis* antigen found to be secreted by the bacillus at high levels, Ag85B, with the two MHC alleles, HLA-DR2 and HLA-DR3. The predictions were generated using a previously published binding algorithm, ISC-PLS (Doytchinova and Flower 2003). The average affinity of each allele for the same set of peptides could then be compared statistically as well as visualized on a plot. The second hypothesis, that an MHC allele associated with susceptibility binds disease-relevant peptides with lower affinity than a non-associated MHC allele, also turned out to be false in this case: HLA-

DR2 was not predicted to bind 9mers from *M. tuberculosis* Ag85B with lower affinity than HLA-DR3 (Fig. 5.3).

Nevertheless, these preliminary data show how statistical models of peptide-MHC binding may be used to test hypotheses regarding HLA-disease associations. As part of future work, the model used here and other models predicting pMHC affinity may be applied to a wider range of *M. tuberculosis* antigens and HLA alleles.

5.4.2 Vaccine Design: rBCG30

Several new vaccines to tuberculosis are currently in development (reviewed in Martin 2005). Among these one strategy has been to engineer the only vaccine in current use, the BCG strain of *M. bovis*, to over-express particular antigens from *M. tuberculosis* to elicit a more lasting and more directed immune response. In one of these candidate vaccines, rBCG30, *M. bovis* BCG has been engineered to express Ag85B from *M. tuberculosis* (Horwitz 2005). Currently in phase I clinical trials, this candidate vaccine and others like it offer an opportunity for models of antigen presentation and the immune response to answer key questions.

One question that might arise, particularly if rBCG30 progress to phase II clinical trials and a larger test population, is how well this candidate vaccine performs in a heterogeneous population displaying polymorphisms in many genes. Some of these polymorphisms (such as IFN-γ +874T leading to increased IFN-γ expression) may predispose individuals to a stronger immune response, while other polymorphisms may have the opposite effect. Different MHC variants present in the population are likely to vary in their ability to bind peptides from the antigen being over-expressed, and these differences in binding affinity may also affect the magnitude of the immune response. With the refinements discussed above, such as a dynamic representation of *M. tuberculosis*, the multi-scale model could incorporate information such as the observed

allele frequencies for antigen presentation-relevant genes in a population and the dose at which a vaccine is administered and predict the range of T cell responses and the frequency of each response that might be expected. In this way the multi-scale model may offer not only a tool to understand *in vitro* results in the laboratory but also a tool to help develop new treatments to diseases such as tuberculosis.

Figures

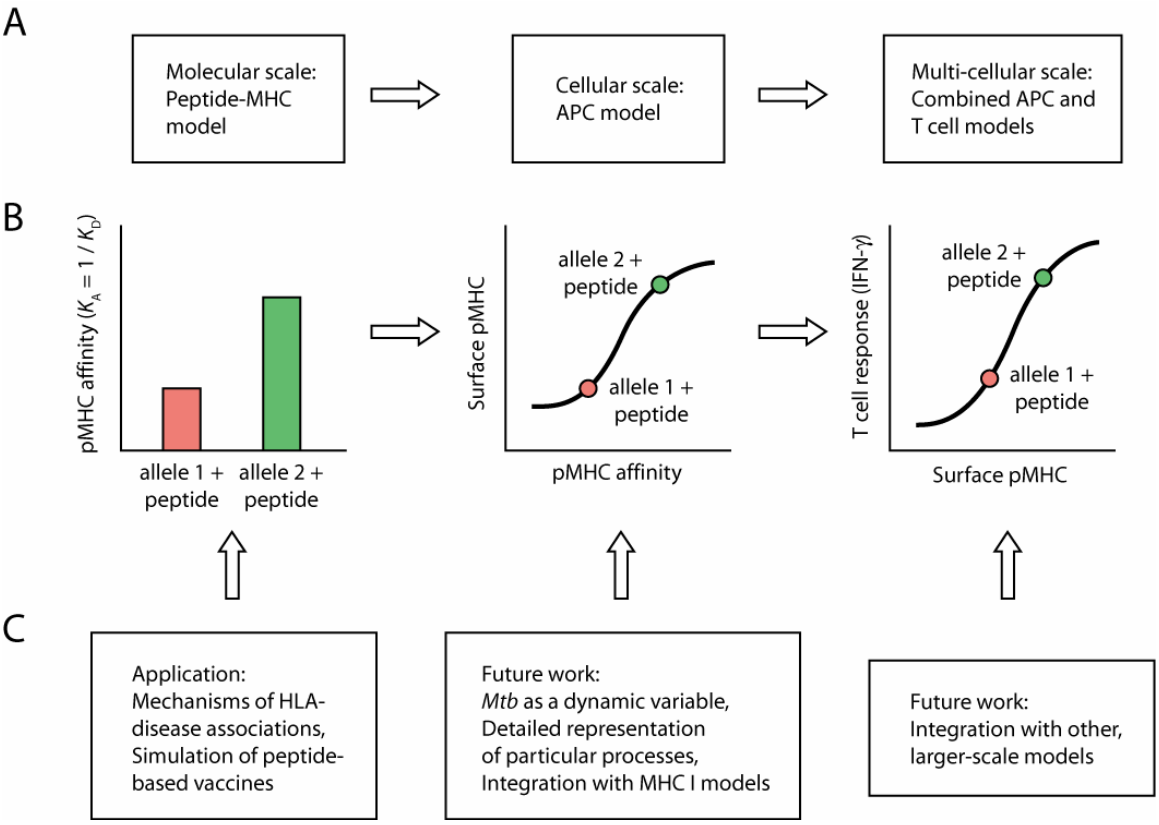

Figure 5.1. Multi-scale model of antigen presentation. *A*, Overview of the three models. *B*, Schematized output of the three models. *C*, Overview of applications and future work to be done with the multi-scale model.

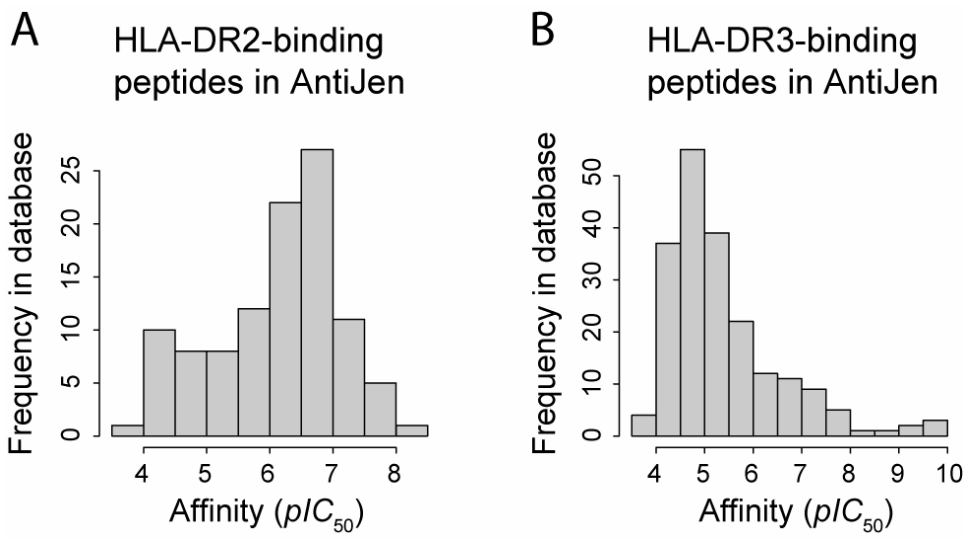

Figure 5.2. Analysis of affinities for TB-associated HLA alleles. *A*, Affinities of peptides for HLA-DR2, associated with TB susceptibility. *B*, Affinities of peptides for HLA-DR3, associated with TB resistance. In both cases, affinities for homologous peptides were first removed using UniqueProt (Mika and Rost 2003). pIC_{50} = -log IC_{50} approximates association equilibrium constant K_A.

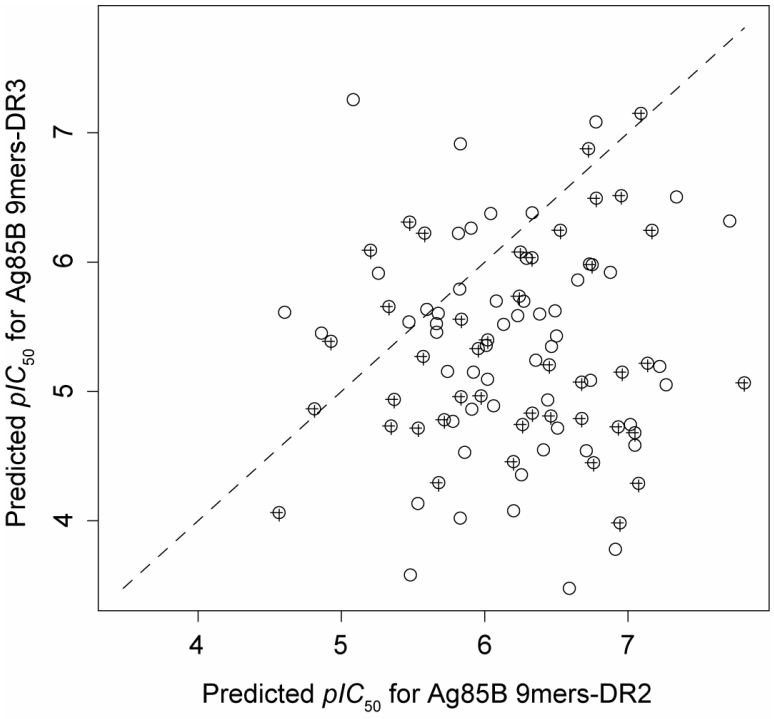

Figure 5.3. Predicted affinities of TB-associated HLA alleles for Ag85B 9mers. 9mers with human homology are indicated in crosshairs. The ISC-PLS algorithm (Doytchinova and Flower 2003) was used to make predictions after being trained on DR2- and DR3-binding peptides from AntiJen from which homologous peptides had been removed using UniqueProt (Mika and Rost 2003). The protein sequence of Ag85B was obtained from the NCBI Entrez Protein database (www.ncbi.nlm.nih.gov/sites/entrez, accession #AAO62005), and predictions were made on all possible 9mers within Ag85B. Ag95B 9mers with human homology were identified using blastp with default parameter settings on the NCBI website (www.ncbi.nlm.nih.gov/blast/) to non-redundant protein sequences from human.

5.5 References

Beatty, W.L. and Russell, D.G. (2000) Identification of mycobacterial surface proteins released into subcellular compartments of infected macrophages. *Infect. Immun.*, **68**, 6997-7002.

Bekkhoucha, F., Naquet, P., Pierres, A., Marchetto, S. and Pierres, M. (1984) Efficiency of antigen presentation to T cell clones by (B cell X B cell lymphoma) hybridomas correlates quantitatively with cell surface Ia antigen expression. *Eur. J. Immunol.*, **14**, 807-814.

Berezhkovskiy, L.M. (1998) On the kinetics of peptide binding to MHC proteins. *Biophys. Chem.*, **71**, 1-8.

Doytchinova, I.A. and Flower, D.R. (2003) Towards the *in silico* identification of class II restricted T-cell epitopes: a partial least squares iterative self-consistent algorithm for affinity prediction. *Bioinformatics*, **19**, 2263-2270.

Gehring, A.J., Rojas, R.E., Canaday, D.H., Lakey, D.L., Harding, C.V. and Boom, W.H. (2003) The *Mycobacterium tuberculosis* 19-kilodalton lipoprotein inhibits gamma interferon-regulated HLA-DR and Fc gamma R1 on human macrophages through Toll-like receptor 2. *Infect. Immun.*, **71**, 4487-97.

Guermonprez, P., Saveanu, L., Kleijmeer, M., Davoust, J., Van Endert, P. and Amigorena, S. ER-phagosome fusion defines an MHC class I cross-presentation compartment in dendritic cells. *Nature*, **425**, 397-402.

Horwitz, M.A. (2005) Recombinant BCG expressing Mycobacterium tuberculosis major extracellular proteins. *Microbes Infect.*, **7**, 947-54.

Houde, M., Bertholet, S., Gagnon, E., Brunet, S., Goyette, G., Laplante, A., Princiotta, M.F., Thibault, P., Sacks, D. and Desjardins, M. (2003) Phagosomes are competent organelles for antigen cross-presentation. *Nature*, **425**, 402-6.

Inaba, K. and Steinman, R.M. (1985) Protein-specific helper T-lymphocyte formation initiated by dendritic cells. *Science*, **229**, 475–479.

Inaba, K., Pack, M., Inaba, M., Sakuta, H., Isdell, F. and Steinman, R.M. (1997) High levels of a major histocompatibility complex II-self peptide complex on dendritic cells from the T cell areas of lymph nodes. *J. Exp. Med.*, **186**, 665–672.

Martin, C. (2005) The dream of a vaccine against tuberculosis; new vaccines improving or replacing BCG? *Eur. Respir. J.*, **26**, 162-7.

McMurry, J., Sbai, H., Gennaro, M.L., Carter, E.J., Martin, W. and De Groot, A.S. (1999) Analyzing *Mycobacterium tuberculosis* proteomes for candidate vaccine epitopes.*Tuberculosis*, **85**, 95-105.

Mika, S. and Rost, B. (2003) UniqueProt: creating representative protein-sequence sets. *Nucleic Acids Res.*, **31**, 3789-3791.

Murray, P.J. (1999) Defining the requirements for immunological control of mycobacterial infections. *Trends Microbiol.*, **7**, 366-72.

Rock, K.L. and Shen, L. (2005) Cross-presentation: underlying mechanisms and role in immune surveillance. *Immunol. Rev.*, **207**, 166-83.

Segovia-Juarez, J.L., Ganguli, S. and Kirschner, D. (2004) Identifying control mechanisms of granuloma formation during *M. tuberculosis* infection using an agent-based model. *J. Theor. Biol.*, **231**, 357-76.

Witt, S.N. and McConnell, H.M. (1993) The kinetics of peptide reactions with class II major histocompatibility complex membrane proteins. *Acc. Chem. Res.*, **26**, 442-448.

Wu, L.C., Tuot, D.S., Lyons, D.S., Garcia, K.C. and Davis, M.M. (2002) Two-step binding mechanism for T-cell receptor recognition of peptide MHC. *Nature*, **418**, 552-6.

Yee, L.J. and Thursz, M.R. (2004) Genetic diversity in the major histocompatability complex and the immune response to infectious diseases. In Bellamy, R. *Susceptibility to infectious diseases: the importance of host genetics.* Cambridge, Cambridge University Press.