A Domain Model of Experimental Autoimmune Encephalomyelitis

Mark Read[1], Jon Timmis[1,2], Paul S. Andrews[1], and Vipin Kumar[3]

[1] Department of Computer Science, University of York, UK.
{markread,jtimmis,psa}@cs.york.ac.uk
[2] Department of Electronics, University of York, UK.
[3] Laboratory of Autoimmunity, Torrey Pines Institute for Molecular Studies, CA, USA.

Abstract. Experimentation with simulations of complex systems can be used to gain insights into those systems' nature and operation. Such *in silico* experimentation represents a valuable tool that can complement conventional *in vivo* experimentation. Validation of a simulation's representation of the real world system remains an open question in complex systems research. As the engineer of a complex system simulation, demonstrating one's understanding of the complex system through the creation of models which can be validated by a domain expert affords some degree of confidence that the results obtained through *in silico* experimentation are representative of the real world system. As a precursor to the creation of simulations of experimental autoimmune encephalomyelitis, a complex autoimmune disease in mice, we present here a model of the disease. The models are expressed using UML, and their construction has afforded insight into UML's expressive capabilities when applied to complex system modelling.

1 Introduction

Experimental Autoimmune Encephalomyelitis (EAE) [14, 15] is an autoimmune disease in mice that serves as a model for multiple sclerosis in humans. The disease, and its subsequent spontaneous recovery, is complex. A large number of immune system cells interact with one another across several bodily compartments to mediate both EAE autoimmunity and its recovery. As is the case with many complex systems EAE is difficult to understand through a reductionist scientific approach alone. The construction of models and simulations of the disease can afford insights into the disease's behaviour and can guide wet-lab experimentation to

points of interest. Experiments that would be difficult to engineer *in vivo* can be engineered into a simulation with relative ease. Simulations permit the investigation of hypotheses concerning the disease's operation within the context of known biological data.

A major concern for *in silico* experimentation using a simulation is the simulation's validity. It is important that one can demonstrate that the results obtained through experimentation with a simulation are representative of the *in vivo* system. How to demonstrate the validity of a simulation remains an open question in complex systems research. The CoSMoS project[4] [1] purposes to develop general principles for the creation and validation of models and simulations of complex systems [2]. The project is developing "the CoSMoS process", an approach to the engineering of complex systems that proposes the construction of models of the complex system as a prerequisite to the construction of any simulations of it.

It is critical that the developers of a complex system simulation possess a decent[5] understanding of how the system works. By developing models of the complex system this understanding can be demonstrated, and can be validated by a domain expert. The construction of such models frequently necessitates examination of the complex system from angles that might not otherwise be considered, and can raise further questions of its operation. The models themselves can form a specification for the construction of a complex system simulation. Validation of a simulation's specification (models) by a domain expert can go some way towards instilling confidence that the simulation is representative of the real world system.

The unified modelling language (UML) is a collection of diagrammatic tools that are were designed for the purpose of specifying software systems. It has been suggested that UML holds potential for modelling biological systems [6]. In this paper we present a model of EAE expressed using UML. By employing UML in this fashion we have identified several strengths and weaknesses of the language when expressing complex biological systems, some of which are outlined in [19]. We find that although UML incorporates several mechanisms that are useful in expressing EAE, such as activity diagrams and state machine diagrams, there are aspects of the biological system that UML cannot satisfactorily communicate. For example, feedback mechanisms that manifest through populations

[4] The CoSMoS project, EPSRC grants EP/E053505/1 and EP/E049419/1, http://www.cosmos-reseach.org.

[5] A "complete" or "detailed" understanding is not always possible, since attempting to resolve uncertainty concerning the system's nature is one motivation for developing the simulation in the first place.

of cells that amplify and counteract each other's operation. Through constructing the present domain model of EAE we have identified some general principles and approaches for applying UML to modelling complex systems. The models we present here serve as an example of how to create a domain model of a complex biological system.

Section 2 details the CoSMoS approach to developing models and simulations of complex systems. Section 3 provides a detailed account of EAE and its recovery. Section 4 introduces UML. In Section 5 we present the model of EAE, highlighting the assumptions we have made in its creation, and how UML has been employed in the generation of the models. Section 6 concludes the paper.

2 CoSMoS Process and the Domain Model

The CoSMoS project is concerned with the development of a modelling and simulation infrastructure that facilitates the design and analysis of complex systems [3]. Ongoing work within CoSMoS seeks to develop a "minimal process" for the development of models and simulations of complex systems. The creation of accurate models and simulations is non-trivial, and demonstrating that they are representative of the real complex systems that they attempt to capture is critical to their use in research. The minimal process represents a first step towards building *validated* models of complex systems.

Figure 1 shows the minimal process as it currently stands. The domain represents the real-world system of interest, in this case EAE. The models and simulations constructed attempt to capture behaviours and properties exhibited by this real world system.

The domain model details the current understanding of the biological domain as held by the modeller. It captures the behaviours present in the biological domain that the modelling and simulation process hopes to investigate. A domain model may span multiple levels of abstraction, from the high level depiction of perceived emergent behaviours exhibited at a system-wide level, to the low-level entities of the system and how they interact with one another. The model should be free from any implementation-specific bias. Validation of the domain model is important, and is carried out by a domain expert. If the domain model is invalid, then the understanding that the modeller has of the system is most likely incorrect, and any simulation built upon that understanding is unlikely to be representative of the real biological domain.

The software model is constructed from the concepts captured in the domain model. It is tailored toward the design and implementation of the

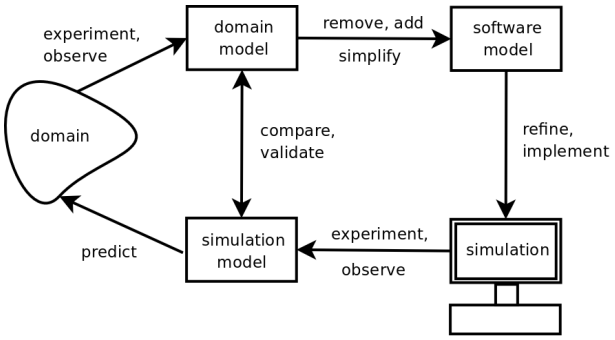

Fig. 1. The CoSMoS minimal process for the development of complex systems simulations [4].

simulator itself; explicit notions of emergent properties and behaviours are removed, and implementation specific concepts may be introduced.

The simulation model is derived through observations of and experimentation with the simulation. The simulation model is to the simulation what the domain model is the domain. Validation can be performed between the domain model and the simulation model; if the simulation correctly captures the desired behaviours present in the domain, then the simulation model should closely resemble the domain model.

The nature of complex systems research dictates that there will exist unknown aspects of the system that no domain expert can be sure of. One of the motivations in creating models and simulations of the system is to bring these areas to light, but where no certain answer currently exists an assumption must be made. Likewise, it is infeasible to model and hence reason about every aspect of a complex system system; abstracting away from complexity believed not to be integral to the system's behavioural dynamics is essential. Whenever an abstraction is made to simplify the system there entails an implicit assumption that the abstraction will not compromise the simulation/model's capture of the behaviours present in the real-world system. These assumptions should be recorded and validated by the domain expert as being appropriate and sensible. If the simulation fails to properly capture the behaviours of interest, then it is likely that an assumption was inappropriate, and it should be readdressed. Thus, the minimal process is iterative.

It is important to recognise and document the questions and issues that one hopes to address through modelling and simulating a complex system. The nature of the assumptions and abstractions that are made in constructing models and simulations of a complex system are moti-

vated by what those models and simulations are to be used for. The assumptions and abstractions must be appropriate for the problem at hand.

In the event that a simulation fails to capture a complex system's emergent properties, yet all abstractions and transitions between the minimal process's models were deemed just and appropriate by a domain expert, then it may hold that some higher level hypothesis upon which the simulation was built is incorrect. It is essential that the simulation's world be properly delineated. There exists a huge variety of interacting elements in a biological system, and accurately simulating all of them is impractical. At its highest level the domain model assumes a hypothesis over which elements in the complex system are responsible for the manifestation of some target abstract behaviour, and thus which elements it will represent. It is plausible that this hypothesis itself will prove to be incorrect, hence the importance of documenting it.

3 Experimental Autoimmune Encephalomyelitis

In this section we provide a detailed description of experimental autoimmune encephalomyelitis and its subsequent recovery. The information provided here provides the basis for the domain model that follows.

3.1 Autoimmunity

Experimental autoimmune encephalomyelitis (EAE) is an autoimmune disease in mice that serves as a model for multiple sclerosis in humans [14, 15]. The disease constitutes the body's immune system attacking myelin, an insulator material that covers the neurons of the central nervous system (CNS) and is essential to their function. Damage to the CNS through demyelination can lead to paralysis and death [16].

Figure 2 presents an informal depiction of how EAE is induced through immunisation with MBP, a myelin derivative. The immunisation is accompanied by complete Freund's adjuvant (CFA) and pertussis toxin, both immunopotentiators which stimulate the immune system. The immunisation occurs subcutaneously. The phagocytosis of MBP by dendritic cells (DCs) resident in the periphery leads to its presentation as MHC-I-MBP and MHC-II-MBP molecules on the DCs. The CFA and pertussis toxin stimulate the DCs, and they up-regulate co-stimulatory molecule expression and migrate to the secondary lymphoid organs. There populations of naive autoimmune MBP-reactive CD4Th1 and CD4Th2 cells bind with MHC-I-MBP as expressed on the immigrant DCs and derive signal 1. The high level of co-stimulatory molecule

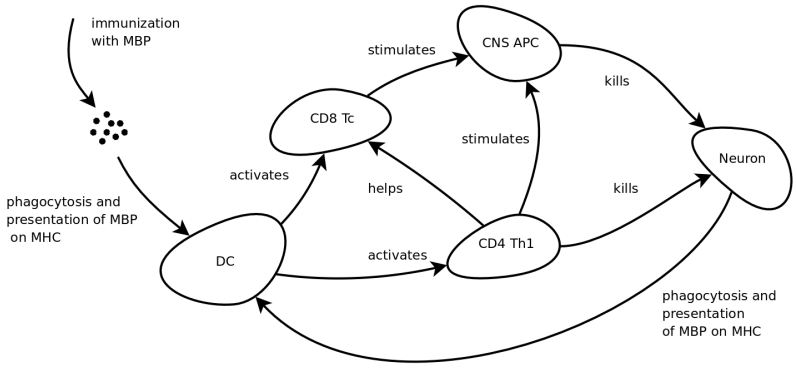

Fig. 2. An informal depiction of how EAE is induced.

expression by these DCs delivers signal 2 to the CD4Th1 and CD4Th2 cells resulting in their activation. MHC-I-MBP molecules, as expressed by these same DCs, are bound by MBP reactive CD8 cytotoxic T (Tc) cells. With help of the MBP-reactive CD4Th1 cells these Tc cells become fully activated.

The now activated CD4Th1, CD4Th2, and CD8Tc cells migrate to the CNS compartment. The CD4Th1 and CD8Tc cells secrete pro-inflammatory type 1 cytokines such as IL-2, INF-γ, and TNF-β [13]. These cytokines represent an inflammatory context to resident antigen presenting cells (APCs) such as macrophages and microglia which become stimulated. When stimulated these CNS APCs secrete TNF-α, reactive oxygen species (ROS), and nitric oxide (NO), all of which are toxic to neurons in high doses [11, 18, 22]. Neurons contain MBP, and those that are killed in this manner are subsequently phagocytosed by CNS APCs, which then express MHC-I-MBP and MHC-II-MBP. The inflammatory conditions in the CNS prompt these APCs to upregulate co-stimulatory molecules, and hence induce the full activation of naive CD4Th1, CD4Th2, and CD8Tc cells that result from proliferation in the CNS.

The CD4Th2 cells secrete type 2 cytokines such as IL-4, IL-5, and IL-10. Type 1 cytokines suppress Th2 cell activity, and type 2 cytokines suppress that of Th1 cells, reducing the cells' proliferative and differentiation capabilities [13]. During the course of EAE autoimmunity the Th1 cell population is dominant, they have a higher affinity for MHC-I-MBP (bindings are stronger and last longer, resulting in less failings to receive signal 1 before the bindings are broken) and proliferate more quickly.

3.2 Regulation-mediated recovery

A network of immune cell interactions can mediate recovery from EAE, and is depicted (at an abstract level) in Figure 3. The physiological turnover of CD4Th1 cells results in their apoptotic death, and subsequent phagocytosis by APCs (such as the dendritic cell) in the CNS draining lymph nodes. Two regions of the T cell receptor (TCR) of MBP-reactive CD4Th1 cells form peptides that are presented on MHC molecules by the APC to prime two populations of regulatory T cell (Treg). These two regions are complementarity determining region 1/2 (CDR1/2) and Fr3, which are presented on non-classical MHC-I (Qa-1) and MHC-II respectively. Binding of MHC-II-Fr3 by Fr3-reactive CD4Treg cells leads to their receipt of signal 1. Molecules generated by the inflammation in the CNS drain into the draining lymph nodes and stimulate the APCs that reside there to upregulate their expression of co-stimulatory molecules. This upregulated expression of co-stimulatory molecules delivers signal 2 to the CD4Treg cells. When activated, and upon binding with MHC-II-Fr3, CD4Treg cells secrete INF-γ, which is required for the processing and presentation of CDR1/2 on non-classical MHC-I (Qa-1) molecules by the APC [23]. This phenomenon is called "licensing" of the APC by the CD4Treg.

CD8Treg cells bind with MHC-I-CDR1/2 as expressed on APCs resident in the CNS's draining lymph nodes and derive signal 1. The high level of co-stimulatory molecule expression on these APCs allows the CD8Treg cells to derive signal 2, becoming fully activated. For a short period of time, around eight hours, following their initial activation CD4Th1 cells express MHC-I-CDR1/2. If this is bound by a fully activated CD8Treg cell the Treg cell can induce the apoptotic death of the CD4Th1 cell.

On a population-wide scale this rise in CD8Treg cell population number leads to a reduction of CD4Th1 number. This occurs in the circulatory system and lymphoid organs such as the spleen; however, Treg cells have not been identified in the CNS, hence this regulation is not assumed to occur there. The transient expression of MHC-I-CDR1/2 by CD4Th1 cells renders them susceptible to regulation for only a short period of time into their full activation. Once this period of time has passed these cells are still susceptible to death through the Fas-FasL pathway. Once activated T cells begin to upregulate their expression of both Fas and FasL on their cell membranes. Sufficient bindings between these two types of molecule can induce a cell's death. Apoptotic death though the Fas-FasL pathway is called activation induced cell death (AICD). The decline in CD4Th1 population number through regulation results in a global reduction of type 1 cytokines being produced. This reduction al-

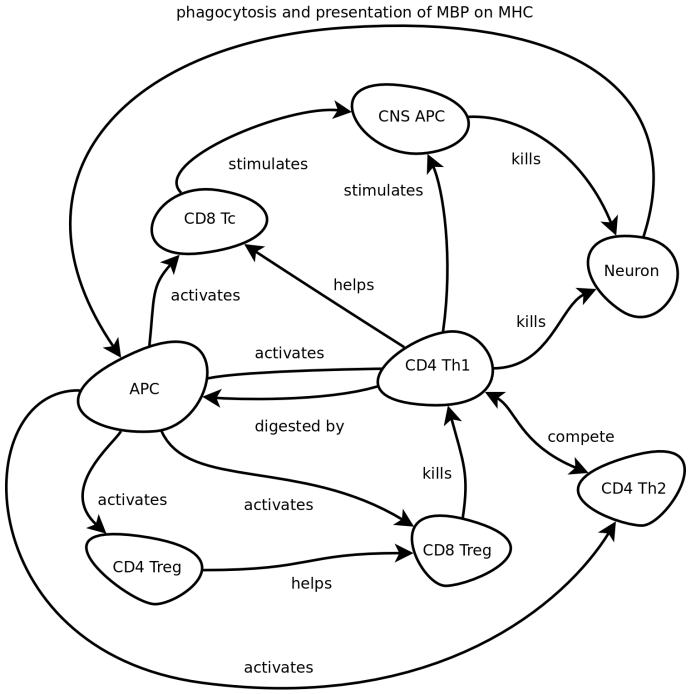

Fig. 3. An informal depiction of all the cells involved in EAE and its regulation-mediated recovery, and their relations to one another.

leviates the suppression of the CD4Th2 population, which then expands and assumes dominant status. The activity of MBP-reactive CD4Th2 cells is not toxic to neurons, and their population expansion does not result in EAE.

4 The Unified Modelling Language

The unified modelling language (UML) [17] is a collection of diagrammatic modelling tools designed to aid the specification and construction of software systems. The diagrammatic tools incorporated within UML provide a wide range of specification scopes, from the relationships held across an entire software system, to full low-level expression of a single system entity. UML diagrams can represent both static and dynamic views of a system. Static views depict the relationships that system entities may hold with one another, whilst dynamic views express the collab-

orations between system entities and the changes to their internal states (which influence their external behaviours).

This multi-viewed approach to specifying systems has made UML a popular modelling tool, and it finds application outside of the software domain within which it was conceived. There have been numerous attempts to model biological systems using UML, for example, [2, 6, 9, 20] [7, 12][6].

In total UML describes 13 different modelling notations [8]. In the domain model presented here we make use of class diagrams, activity diagrams, and state machine diagrams. An assessment of the use of various UML diagram notations in creating the present domain model is to appear in ICARIS 2009 [19].

Class digrams depict the static relationships between entities in the system. Relationships can be assigned a role name, and cardinalities at both ends of the relationship indicate how many instances of each entity may partake in a relationship at any one point in time.

Activity diagrams represent a dynamic view of a system, and indicate an ordering of events between instances of system entities that occur within a particular scenario. The events (called activities) depicted in an activity diagram may be any abstract concept.

State machine diagrams are a dynamic view of individual entity types in the system. All instances of a particular entity follow the dynamics defined for their type. State machine diagrams describe the states that an entity may exist in. An entity's state determines which events and interactions it is capable of partaking in. States can be mutually exclusive, orthogonal, and hierarchical.

5 Domain Model of EAE

This section presents the domain model of EAE and its regulation. As detailed in section 2, the abstractions and assumptions that are made in arriving at a simulation of a complex system are heavily dependent on the intended purposes of the simulation, and so we document these purposes here.

The models and simulations we are constructing are for the purposes of conducting *in silico* experimentation. Through construction of a simulation of EAE that intergrates known biological data about the disease we hope to extrapolate the potential values of otherwise unknown biological parameters. This is elaborated upon in section 5.2. Once constructed,

[6] These works used state charts [10] as their modelling medium. State charts are very similar to the state machine diagrams of UML.

it is trivial to remove or alter the nature of entities in the simulation. By observing the altered dynamic of the system we hope to ascertain the importance of those entities to the system's behaviour. As an example, we can experiment with the length of time that a recently activated CD4Th1 cell expresses MHC-I-CDR1/2 for, and observe how the system's behavioural dynamics are affected by this change. Such experimentation is extremely challenging to engineer into an *in vivo* system, yet is relatively trivial to perform through simulation. By performing *in silico* experimentation we hope to highlight areas of significance within the system that can then be used to direct wet-lab experimentation to points of interest.

5.1 Delineating the system

As detailed in section 2, it is essential to delineate the system of interest; modelling and simulating an entire biological system is intractable. Figure 4 denotes the observable phenomena of the real-world domain. Argument over these phenomena is deemed to be outside the scope of this modelling work. "Autoimmunity" is an overloaded term which immunologists may disagree over the exact origins of; there is more than one form of autoimmunity. This diagram delineates the system we intend to model in exact terms, both the physical entities within it and the behaviours we expect them to manifest, omitting overloaded definitions such as "autoimmunity". Note that Figure 4 does not conform to any UML notation.

The diagram explicitly depicts several levels of hypothesis that the model and simulation will incorporate. The transitions across the dotted line depict our hypotheses concerning those abstract behaviours/events that we believe to be responsible for the observable phenomenon. These transitions delineate the outer bounds of our investigations; we will not attempt to investigate whether anything other than our "expected behaviours" are responsible for the observable phenomenon. Our investigations are scoped within the context of these expected behaviours, as indicated on the diagram. It is these expected behaviours that we are attempting to capture in our models and simulations. The transitions over the dotted line indicate how the work carried out with our simulations fits into the wider context of study on EAE and autoimmunity.

Further hypotheses are detailed in the links between the expected behaviours and the real physical entities of the system. These links indicate which entities in the real-world system we believe to be responsible for manifesting the expected behaviours, and will thus find explicit representation within our system. The "expected behaviours" are so named because we expect these system-wide behaviours to manifest from the

lower level entities and their interactions, as depicted on the diagram (albeit at an abstract level).

In our case the observed phenomena are that mice experience paralysis from EAE, and that the immune system is regarded as being responsible for carrying out damage to the central nervous system. Mice can recover from EAE spontaneously. Following recovery mice are typically insusceptible to further attempts to induce EAE autoimmunity in them.

We hypothesise that "autoimmunity against the CNS" is caused by the behaviour of immune cells harming central nervous system (CNS) cells. This behaviour manifests through the actions of several immune cells. Dendritic cells activate auto-reactive CD4 Th1 cells, which in turn facilitate the activation of CD8 Tc cells, which together stimulate CNS Macrophages into secreting molecules that are toxic to neurons (CNS Cells). These actions are all quite abstract, and are expanded upon in other diagrams, as discussed below.

Of note is that one of the observable phenomena is not linked to an expected behaviour. We are unsure as to what is responsible for "protection against subsequent attempts to induce autoimmunity against CNS". Two possibilities include the establishment of an equilibrium between the rise of CD4Th1 cells and their apoptotic death through regulation, or the action of memory Treg cells that efficiently subvert the onset of autoimmunity before significant damage is caused.

Figure 4 is an alternative to another technique of expressing the expected behaviours or emergent properties of a system; Garnett *et al.* [9] have represented the emergent property that their simulations attempted to capture as a first class entity on a class diagram. We have found this technique unsuitable for EAE; a class labelled "autoimmunity" or "regulation" cannot be instantiated in the same manner that class labelled "dendritic cell" can be, yet their representation as such would imply same semantic behaviour. Instead, through Figure 4, we have captured the system-wide behaviours of autoimmunity and regulation as unique abstract entities and linked them to the physical components in the system responsible for their manifestation.

5.2 Validating models and simulations

A central issue for validation of a simulation of a complex system, and results obtained thereof, is identifying how well it captures the behaviours exhibited by the real world system. In the case of EAE there is no available metric to measure how well "autoimmunity" has been captured. *In vivo* experimentation defines a scale based upon the degree of paralysis experienced by a subject. Since modelling and simulating the entire mouse is unmanageable, a parallel of this metric for the simulation is not

possible. It is unknown how much "damage" to a central nervous system in terms of neuron death corresponds to a particular degree of paralysis, so although neurons can find explicit representation within a simulation the extent to which they are attacked by the immune system cannot be used as a metric either.

Through interaction with a domain expert a timeline of significant events that can be observed within the *in vivo* system can be devised. In the case of EAE these events are depicted in Table 1. They correspond with observations made at the cell population level. This timeline can potentially be used to validate a simulation's capture of EAE; if the population dynamics within the simulation match those of the timeline, and if the the behaviours of entities represented within the simulation are validated by the domain expert, then some level of confidence that the simulation is representative of EAE can be obtained.

The nature of current immunological research dictates that not all biological parameters are known, and some will be subject to controversy within the field. This is the case with EAE, and presents a problem for any simulation that attempts to capture it. The present domain model details which biological parameters are and are not known; see Table 2. Those that are known can be incorporated into a simulation, whilst those that are not will be subject to experimentation. Given the timeline of *in vivo* EAE, and that the behavioural dynamics of the cells that mediate it are validated by a domain expert, correct values for the unknown biological parameters should recreate the timeline within the simulation. This hinders on the assumptions that have been made in arriving at the simulation being appropriate and valid, as indicated by a domain expert. Documenting these assumptions is critical for determining the validation of models and simulations. Appendix A captures the assumptions made in the present domain model.

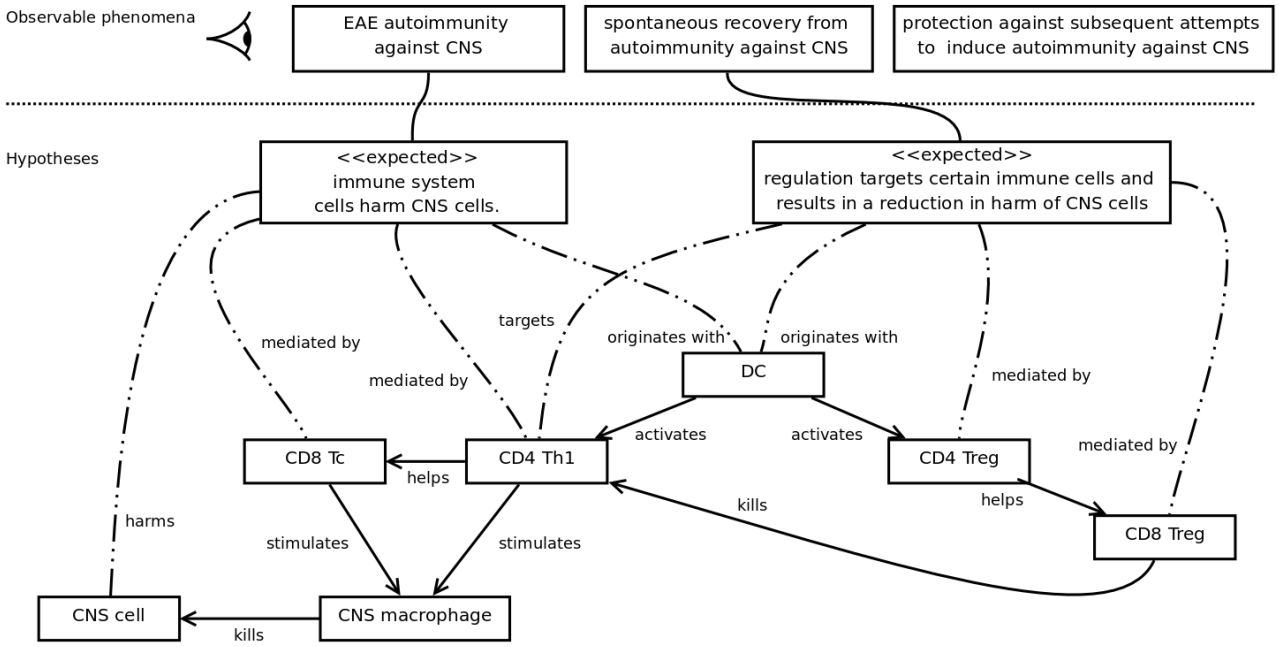

Fig. 4. This diagram details: the observable phenomenon of the biological domain; the behaviours that we hypothesise to be responsible for those phenomenon; and, at an abstract level, which physical entities of the real-world biological domain we believe to be responsible for manifesting those behaviours. Note that there are not formal semantics attached to this diagram.

Time	Event
0 days	Immunisation with MBP, CFA, and pertussis toxin in the periphery
3-5 days	Detectable proliferation of CD4Th1, CD4Th2, and CD8Tc cells in the secondary lymphoid organs
5-7 days	Detectable proliferation of CD4Th1, CD4Th2, and CD8Tc cells in the CNS
10-15 days	Visible paralysis of mouse
10 days	Detectable proliferation of CD4Treg and CD8Treg cells in secondary lymphoid organs
30-40	Recovery from EAE.

Table 1. The key population level events in EAE and its regulation-mediated recovery.

Event	Time
Delay in phagocytosis of substance to its appearance on MHC	1-2days
Delay in upregulation of co-stimulatory molecules following stimulation of APC	∼2 hours
Persistence of Qa-1-CDR1/2 on CD4Th1 cell following activation	∼8 hours
CD4Th2 cell dies from AICD after initial activation after	∼8 days
Stimulated, activated CD4Th2 cell proliferates every	∼3 days
Other T cells die from AICD after initial activation after	∼5 days
Other activated T cells, given sufficient stimulation, proliferate every	∼2 days
Persistence of MHC-peptide on APC membrane	unknown
Lifetime of DC	unknown
Half life of cytokine (this is the case for all cytokines in the domain model)	unknown

Table 2. The time taken for key events within EAE to occur. Some of these biological parameters are not known.

5.3 Modelling expected behaviours

The expected behaviours of autoimmunity ("Immune system cells harm CNS cells") and regulation, as depicted on Figure 4, represent system-wide behaviours that manifest from low-level interactions between cells. Activity diagrams represent a powerful medium in which to express how these scenarios occur. The activity diagrams in Figures 5 and 6 depict the order in which events at the individual cell level occur for autoimmunity and regulation to manifest. Figure 7 shows how the single cell events in regulation (Figure 6) translate to deviation from autoimmunity at the system-wide level.

The events depicted as activities in these activity diagrams are abstract concepts. They do not themselves specify the behavioural dynamics of individual cells given a range of scenarios; that is accomplished through use of state machine diagrams, discussed below. The activity diagrams are very effective at showing how the individual cell-level dynamics expressed in state machine diagrams integrate to constitute a system-wide dynamic. EAE contains many cascades of events, and the top level behaviours manifest from the interactions of population dynamics which themselves manifest through the concurrent actions of many individual cells. Of all the diagrams defined within UML, activity diagrams are the most expressive in terms of depicting a break down of system-wide dynamics.

From the activity diagrams that depict scenarios within the system, class diagrams that represent a static perspective of the scenario can be created. Figures 8 and 9 represent class diagrams of EAE and regulation respectively. Class diagrams are concerned with expressing the relationships that entities in the system hold with one another, and the number of entities that take part in those relationships at any particular point in time. These diagrams are somewhat informative for EAE and its recovery; however, reasoning about the system in a static manner is not as informative to its operation as is examination from a dynamic viewpoint. Ordinarily there is little to constrain the number of biological entities that attempt (either successfully or not) to partake in a particular interaction at a time, and this can potentially manifest in "0..*" cardinalities on class diagrams (note that in diagrams 8 and 9 this is not the case, due to assumption 1). "0..*" cardinalities appearing all over a diagram can perhaps convey that the system is complicated, but they do not highlight how the system operates. Furthermore, in biology there is relatively little to stop anything from attempting to interact with anything else, and many such interactions produce effects. This can lead to highly connected class diagrams that are difficult to interpret in a meaningful manner. The approach taken in Figures 8 and 9 has been to

represent *partial* class diagrams that depict a subset of the entire system; in this case delineated by how low level entities manifest different high level system-wide behaviours.

5.4 Capturing low-level dynamics of system entities

State machine diagrams are used to depict low-level behavioural dynamics, and are constructed for all entities that either actively change the state of the system, or those that play important roles in mediating system dynamics.

Correctly capturing the dynamics of a complex biological entity, such as a cell, on a two dimensional diagram can prove challenging. The dynamics of a cell exhibit high dimensionality, and the dimensions are not necessarily completely independent. As an example, Figure 10 shows the state machine diagram for a CD4Th1 cell. The locations in which the CD4Th1 cell may reside are depicted as a mutually exclusive set of states that are orthogonal to the rest of the cell's behaviour; however this is not really the case. The state transitions that depend on binding with MHC-II-MBP complexes can only occur when the cell is in the SLO or CNS compartments, since these are the only locations in the model where APCs reside. Depicting this diagrammatically would make the diagram very cluttered. The use of guards for relationships such as this would add to the complexity of the diagram. This particular example is covered by the case that the state machine diagrams of the dendritic cell and CNS macrophage (Figures 15 and 16) indicate where they can reside. However, the purpose of these models is to be informative and transparent; excess complexity should be avoided.

There are behavioural aspects that are impossible to represent through conventional use of state machine diagrams. For example, Figure 15 presents the state machine diagram for a dendritic cell. The levels of MHC-I-peptide presentation are dependent on: being licensed by a CD4 T cell; the quantity of peptide available within the cell for presentation; the level of stimulation within the cell, which is itself dependent on the perception of CFA and type 1 cytokines. Exactly how these variables interact with one another to dictate MHC-I-peptide presentation is not completely clear; resolution of these unknowns will require experimentation with the simulation and interaction with the domain expert. However, even if this were not the case, state machine diagrams incorporate no way for us to represent these complex relationships and variable values without the use of equations or significant quantities of text.

Several of the state machine diagrams presented here incorporate a single state with no transitions to alternative states that is orthogonal to the others, for example "express MHC-II-peptides" on Figure 16. It

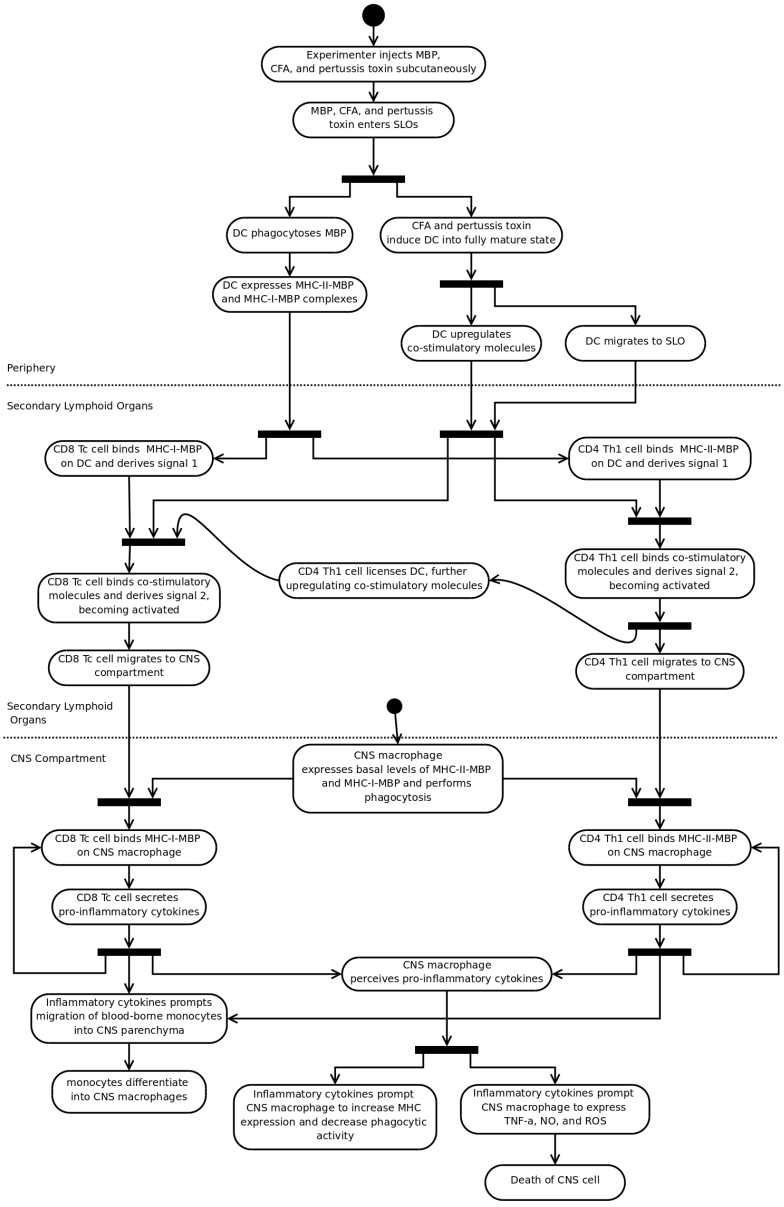

Fig. 5. Activity diagram representing the order of low-level inter-cellular events that lead to EAE.

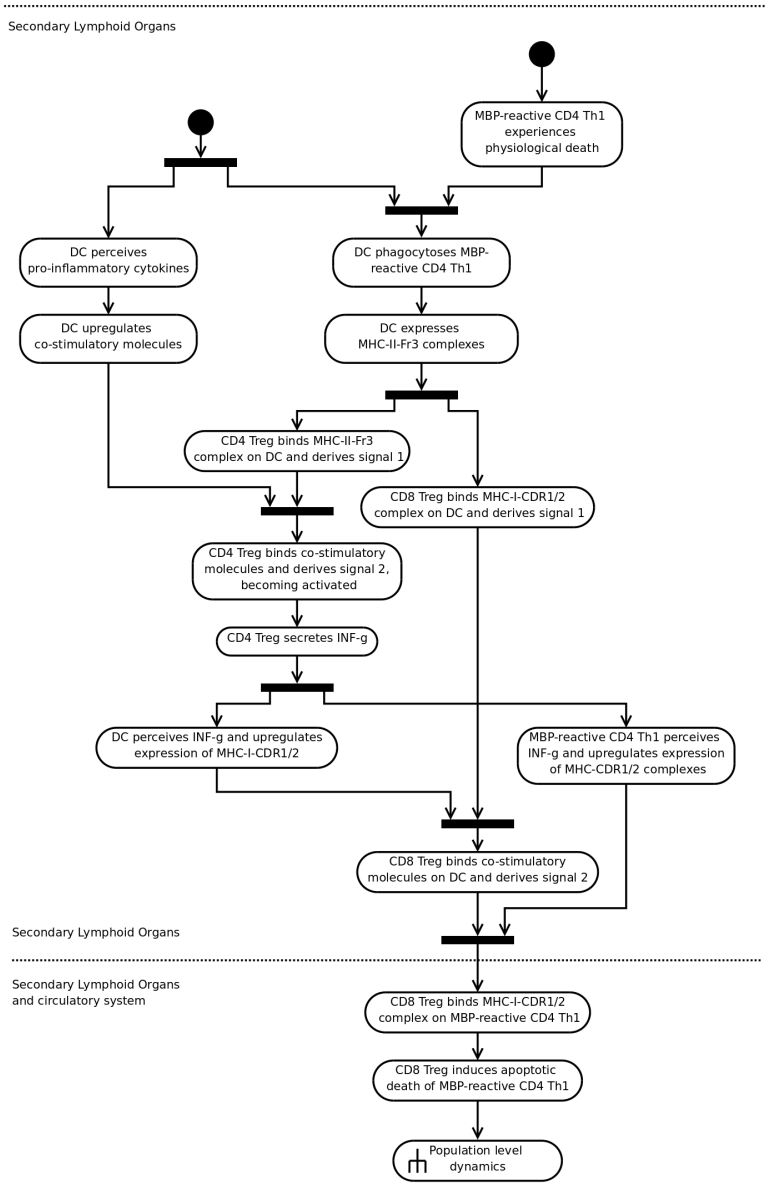

Fig. 6. Activity diagram representing the order of low-level inter-cellular events that lead to regulation mediated recovery from EAE. This diagram leads into Figure 7

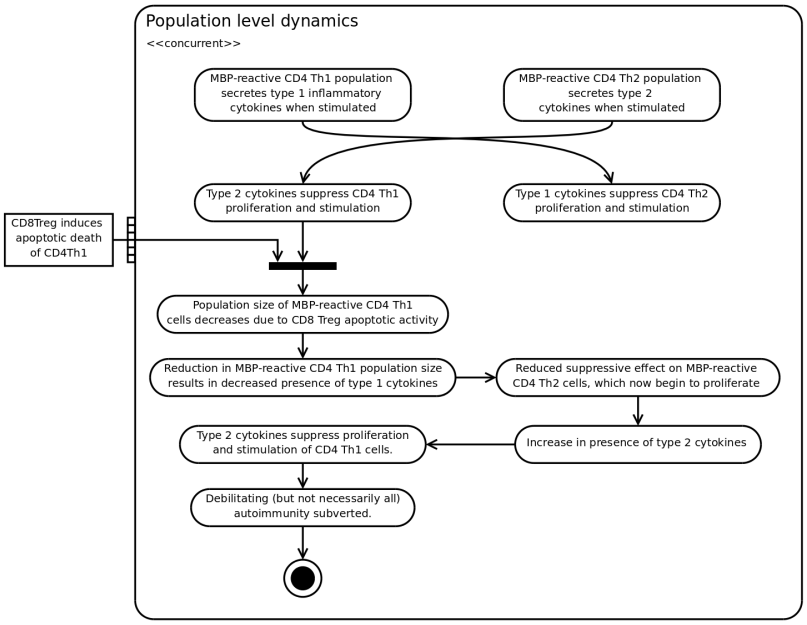

Fig. 7. Activity diagram representing regulation mediated recovery from EAE. This diagram follows from Figure 6.

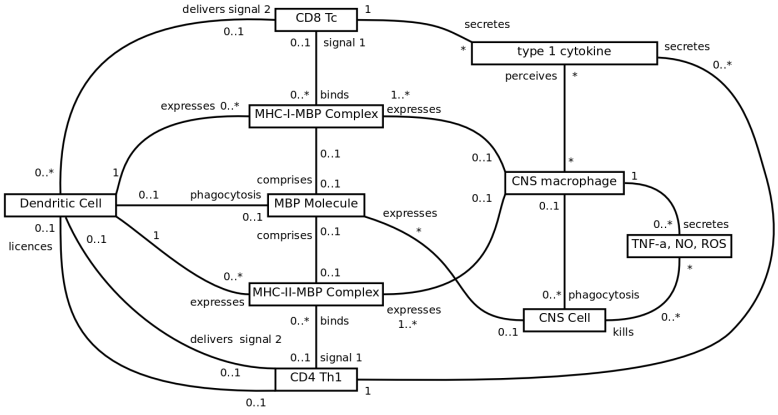

Fig. 8. Class diagram depicting the entities responsible for EAE.

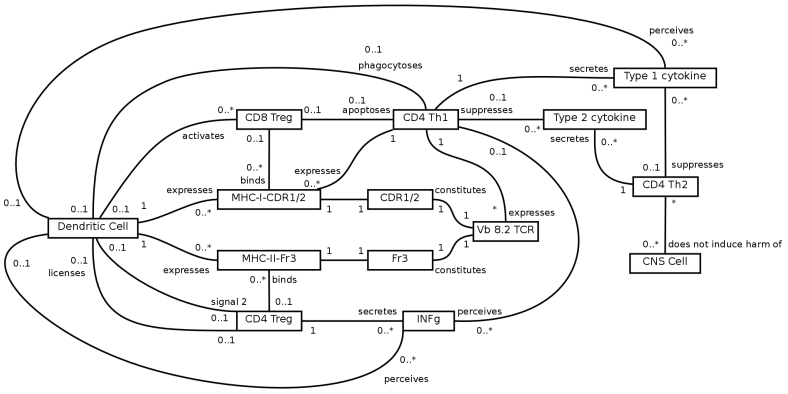

Fig. 9. Class diagram depicting the entities responsible for regulation mediated recovery from EAE.

is unconventional to express only one state in this manner, but the approach has been useful in depicting certain activities of cells.

It has proven useful to construct state machine diagrams of entities that do not necessarily carry state. For example, Figures 19 and 18 respectively show the locations in which MBP molecules may reside, and the effects that INF-γ has on cells the perceive them. These aspects do not necessarily comprise internal states of the molecules portrayed in the state machine diagrams. However, since these system elements mediate the actions of other elements that do carry state, then it can be informative to depict the system from their perspective.

Several of the state machine diagrams in this model depict the physical locations in the model where a cell may reside. It can be argued that a cell's physical location is not part of its internal state, but depicting it in this manner is informative. A similar approach has been used by [9].

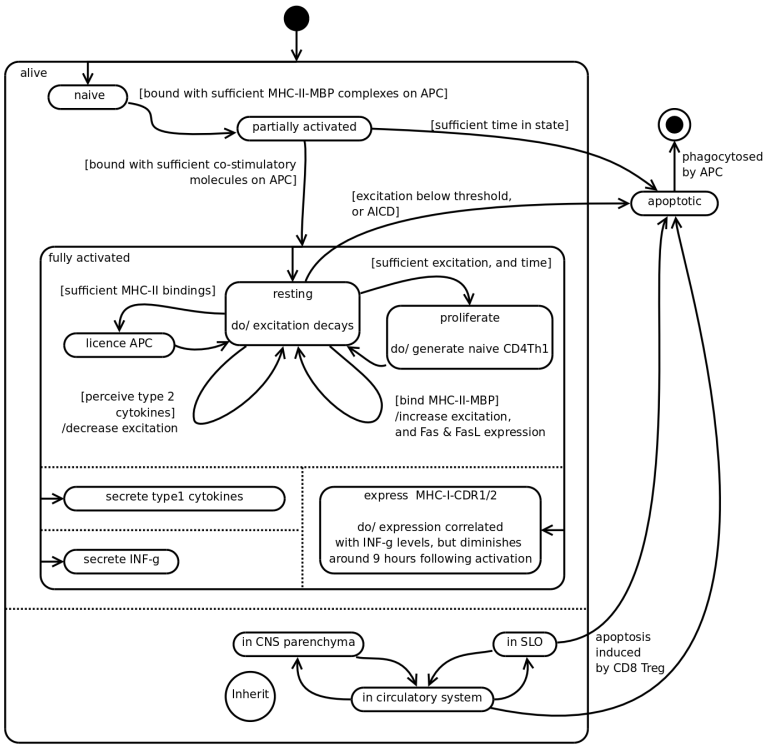

Fig. 10. State machine diagram of a CD4 Th1 cell.

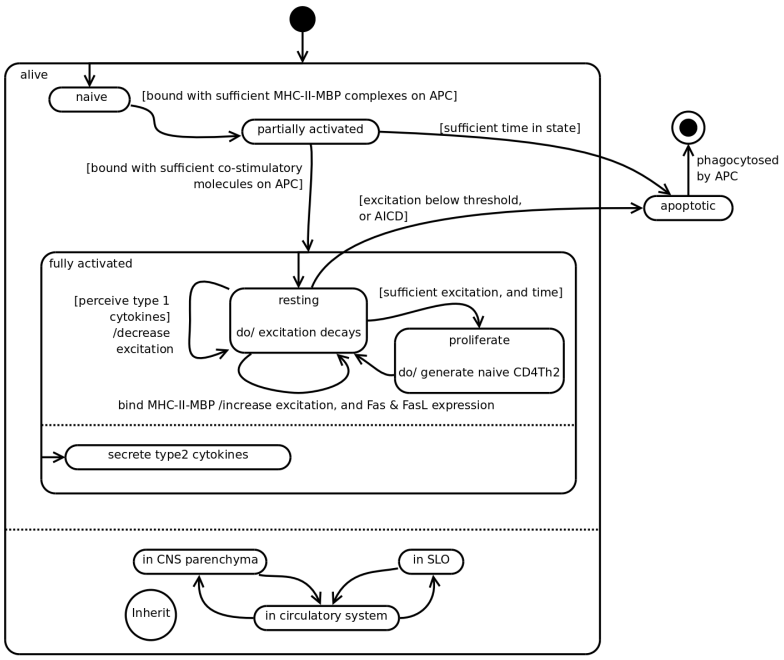

Fig. 11. State machine diagram of a CD4 Th2 cell.

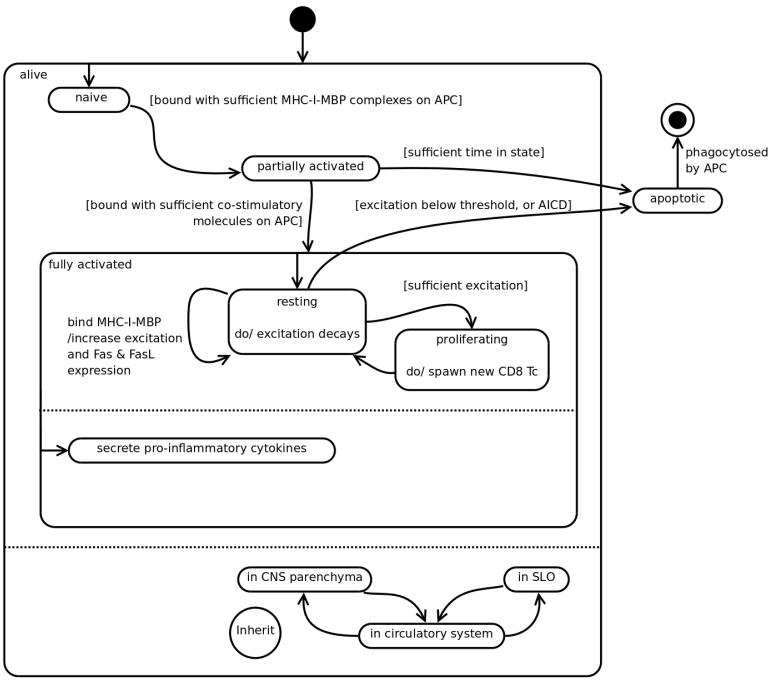

Fig. 12. State machine diagram of a CD8 Tc cell.

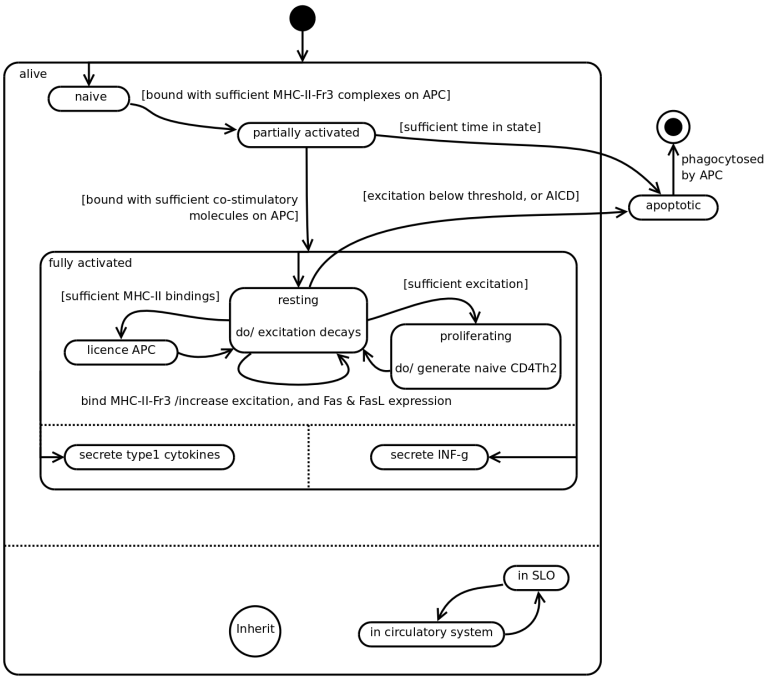

Fig. 13. State machine diagram of a CD4 Treg cell.

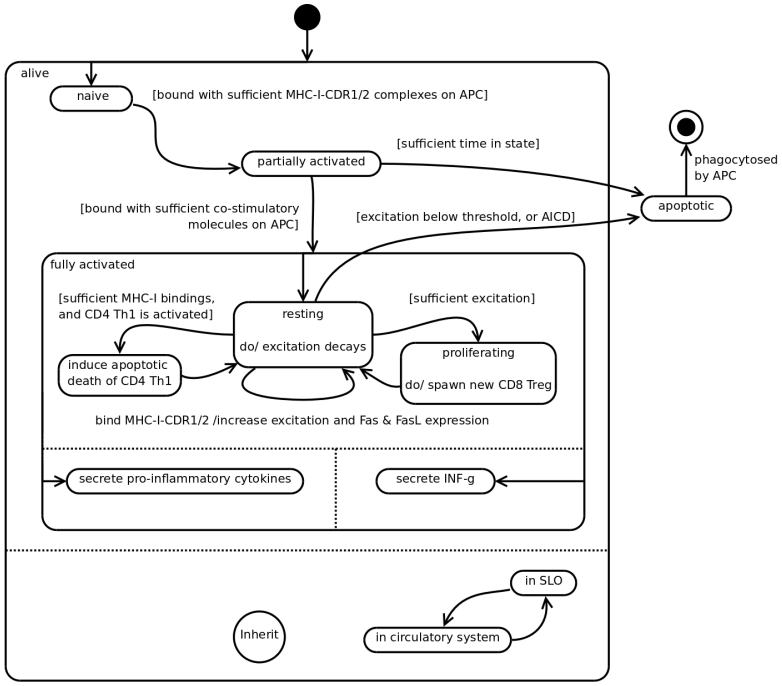

Fig. 14. State machine diagram of a CD8 Treg cell.

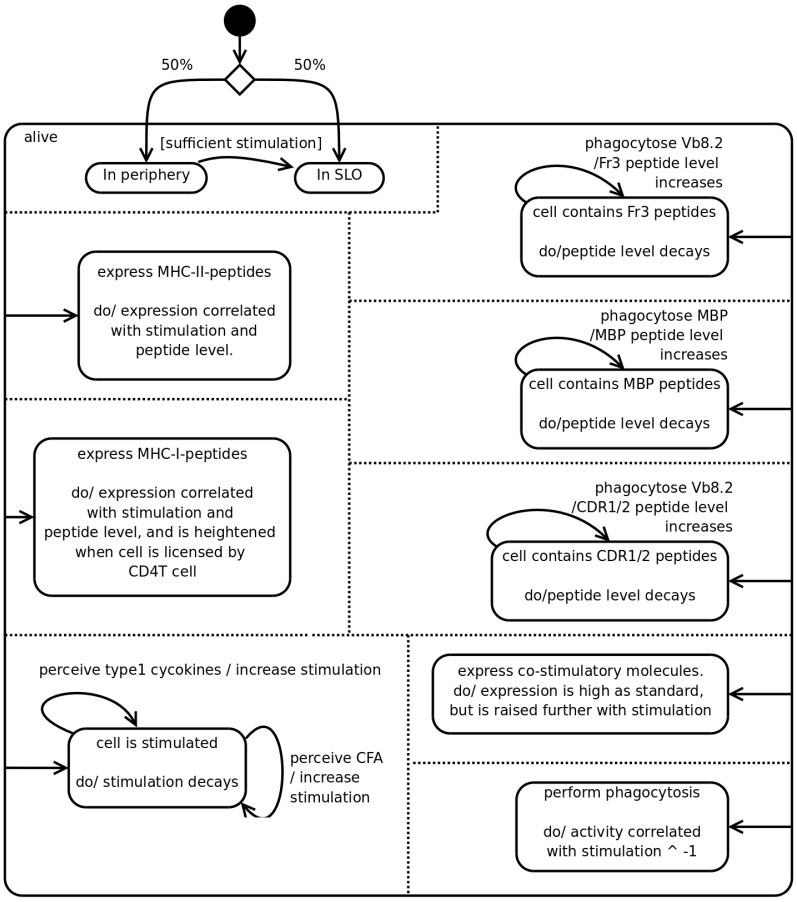

Fig. 15. State machine diagram of a dendritic cell.

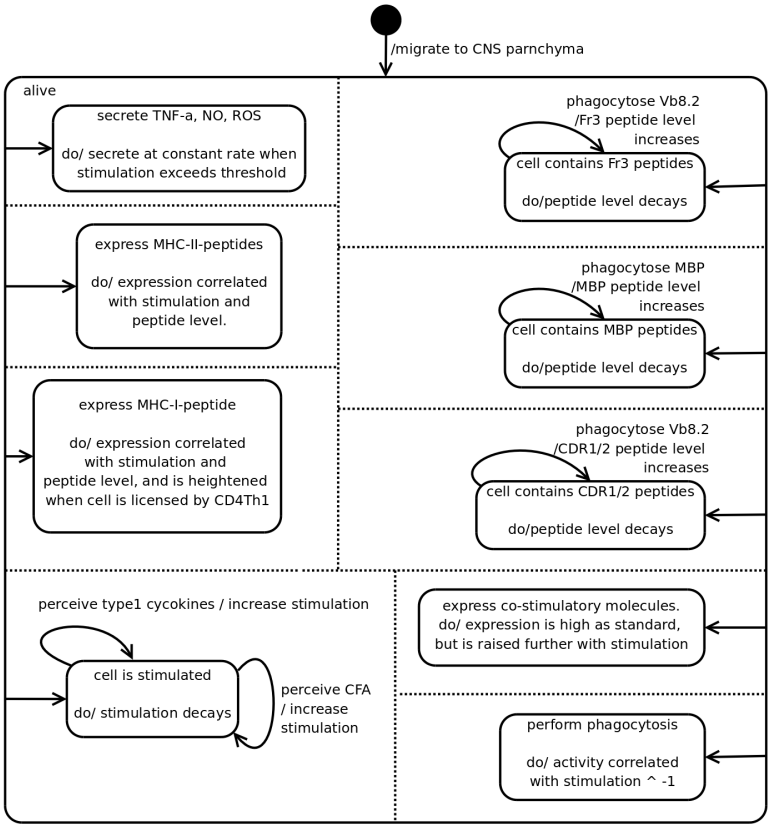

Fig. 16. State machine diagram of a CNS macrophage.

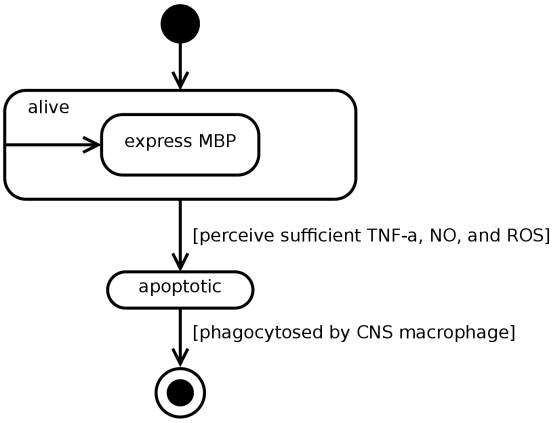

Fig. 17. State machine diagram of a CNS cell.

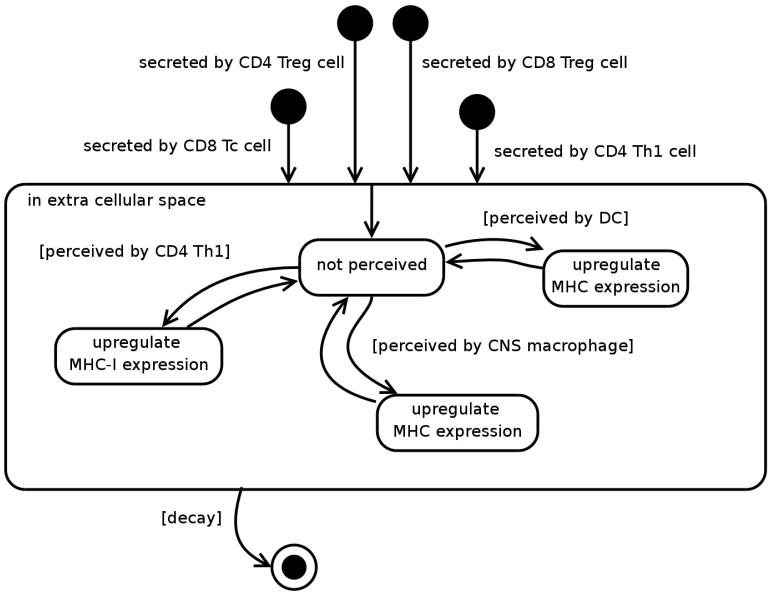

Fig. 18. State machine diagram of INF-γ, a cytokine.

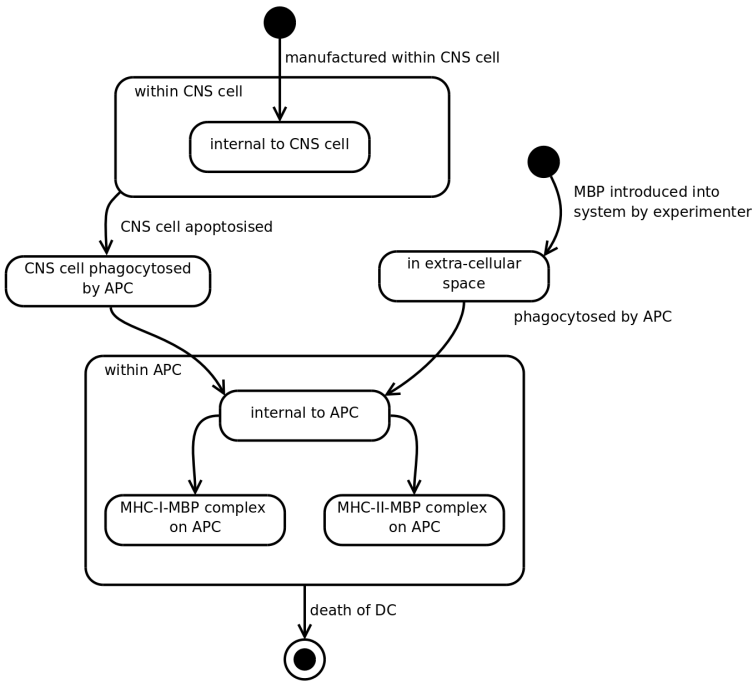

Fig. 19. State machine diagram of MBP.

5.5 Representing feedback

Activity diagrams can demonstrate the order in which critical interactions and events must take place for a high level behaviour to manifest. However they (incorrectly) imply that one activity stops and another starts. In reality the entity responsible for a preceding activity does not hand off control to that which follows, it continues and can potentially perform the same activity again. This concurrency amongst system elements can manifest in feedback, where an increasing number of elements engage in some activity.

To give two examples in the context of Figure 5, a fully activated CD4Th1 will not stimulate a single CNS macrophage and then stop, it will repeat this process. Likewise, the death of a CNS cell (through the secretion of TNF-α, ROS, and NO by a CNS macrophage) will lead to its phagocytosis by a CNS Macrophage, which then presents MBP to additional naive CD4Th1 cells, facilitating their activation and accelerating the progress of EAE. This latter feedback can further amplify the effect of the former. Relative population dynamics play a significant role in EAE (for example, consider the interplay between CD4Th1 and CD4Th2 cells) and it is important to communicate this information in the domain model. Depicting these feedbacks, and others like them, on the activity diagrams will significantly clutter it; as yet we have found no mechanism within UML to satisfactorily express these feedbacks and the interplay between them.

6 Conclusion

We believe that the modelling of a complex system is a necessary precursor to the implementation of a simulation of that complex system. It is necessary to demonstrate a detailed understanding of the complex system of interest and have this validated by a domain expert. If one's understanding of the system is incorrect then the simulation will not be representative of the real world system; uncovering such errors can be difficult and time consuming. A model of a complex system can serve as a specification for a simulation, and its validation by a domain expert can deliver some measure of confidence in the simulation's own validity.

We have presented here a domain model of experimental autoimmune encephalomyelitis (EAE), a complex autoimmune disease in mice, and its regulatory T cell mediated recovery. The models are expressed using UML, and the creation of the present domain model has afforded insights into UML's expressive capabilities when applied to complex systems.

Complex systems tend to exhibit many interactions between entities within the system. Attempting to capture all this interaction in one

diagram that fully describes the system renders the diagram cluttered and illegible. Since a primary purpose of creating a domain model is to communicate one's understanding of a complex system a balance must be struck between fully specifying the system where ever possible and maintaining informative diagrams. We have found it useful to identify the scenarios in EAE, being autoimmunity and recovery, and depict these separately using activity diagrams. Class diagrams of the entities that partake in a scenario have been constructed, but we have found their contribution to the model to be minor; diagrams that depict system dynamics and the order in which events that comprise a scenario take place have been more relevant than a static depiction of all the interactions that are possible. UML state machine diagrams of system entities that don't themselves carry state or instigate interactions, but do mediate interactions between other entities have been informative.

The dynamics of EAE are heavily dependent on the interplay between cell populations and feedback mechanisms. We have found no satisfactory method to use UML in expressing these aspects of the system. Biological cells incorporate many features that are subject to continuous domains, such as variable levels of stimulation or molecule expression. These aspects cannot be expressed through state transitions alone, and require either textual explanation or use of equations to specify. UML encompasses several mechanisms that have proven useful in modelling EAE; however, in its current form, UML alone is insufficient to fully specify the system.

The CoSMoS minimal process, the principled approach to complex system simulation development that we are following, is iterative. As we explore EAE through simulation and investigate alternative hypotheses concerning its operation our domain model may require amendment. Should we wish to investigate the effect that a cell previously not represented in the simulation has on EAE our domain model will be modified to reflect its incorporation into the system. Validation of the domain model by a domain expert is intended to provide some measure of confidence in the results of experimentation with a simulation. The model must be maintained to reflect the experiments we conduct, and changes must be validated.

The next stage in our work is to refactor the domain model into an implementation specific simulation model. The simulation model will form the specification for the construction of a simulation, and will be used to conduct *in silico* experimentation with the intention of gaining insights into EAE's nature and operation.

7 Acknowledgements

The work of Paul Andrews is supported by the CoSMoS project, EPSRC grants EP/E053505/1 and EP/E049419/1.

References

[1] Complex Systems Modelling and Simulation Infrastructure (CoSMoS) project homepage. *http://www.cosmos-research.org/*.

[2] Paul S. Andrews, Fiona Polack, Adam T. Sampson, Jon Timmis, Lisa Scott, and Mark Coles. Simulating biology: towards understanding what the simulation shows. In Stepney et al. [21], pages 93–124.

[3] Paul S. Andrews, Adam T. Sampson, John Markus Bjrndalen, Susan Stepney, Jon Timmis, Douglas N. Warren, and Peter H. Welch. Investigating patterns for the process-oriented modelling and simulation of space in complex systems. In *ALife XI, Winchester, UK, September 2008*, pages 17–24. MIT Press, 2008.

[4] Paul S. Andrews, Adam T. Sampson, Fiona Polack, Susan Stepney, and Jon Timmis. Cosmos development lifecycle, version 0 (in preparation). Technical report, The University of York, 2008.

[5] Etty N. Benveniste. Role of macrophages/microglia in multiple sclerosis and experimental allergic encephalomyelitis. *Journal of Molecular Medicine*, 75(3):165–173, March 1997.

[6] Hugues Bersini. Immune system modeling: The OO way. In *ICARIS*, pages 150–163, 2006.

[7] Sol Efroni, David Harel, and Irun R. Cohen. Toward rigorous comprehension of biological complexity: Modeling, execution, and visualization of thymic T-cell maturation. *Genome Research*, 13:2485–2497, 2003.

[8] Martin Fowler. *UML Distilled*. Addison-Wesley, 3rd edition, 2004.

[9] Philip Garnett, Susan Stepney, and Ottoline Leyser. Towards an Executable Model of Auxin Transport Canalisation. [21], pages 63–91.

[10] David Harel. Statecharts: A visual formalism for complex systems. *Science of Computer Programming*, 8(3):231–274, June 1987.

[11] Jerome J. A. Hendriks, Charlotte E. Teunissen, Helga E. de Vries, and Christine D. Dijkstra. Macrophages and neurodegeneration. *Brain Research Reviews*, 48(2):185–195, April 2005.

[12] Na'aman Kam, Irun R. Cohen, and David Harel. The Immune System as a Reactive System: Modeling T Cell Activation with Statecharts. In *Proceedings of Visual Languages and Formal Methods (VLFM'01), part of IEEE Symposium on Human-Centric Computing*, pages 15–22, 2001.

[13] Thomas J. Kindt, Richard A. Goldsby, and Barbara A. Osbourne. *Kuby Immunology*. W. H. Freeman and Company, 6th edition, 2007.

[14] Vipin Kumar. Homeostatic control of immunity by TCR peptide-specific Tregs. *The Journal of Clinical Investigation*, 114(9):1222–1226, November 2004.

[15] Vipin Kumar and Eli Sercarz. An integrative model of regulation centered on recognition of TCR peptide/MHC complexes. *Immunological Reviews*, 182:113–121, 2001.

[16] Loui Thomas Madakamutil, Igor Maricic, Eli E. Sercarz, and Vipin Kumar. Immunodominance in the TCR repertoire of a TCR peptide-specific CD4+ Treg population that controls experimental autoimmune encephalomyelitis. *The Journal of Immunology*, 180(1):4577–4585, April 2008.

[17] Object Management Group. Maintainer of the unified modelling language standards. *http://www.uml.org*.

[18] Gennadij Raivich and Richard Banati. Brain microglia and blood-derived macrophages: molecular profiles and functional roles in multiple sclerosis and animal models of autoimmune demyelinating disease. *Brain Research Reviews*, 48(3):261–281, November 2004.

[19] Mark Read, Paul S. Andrews, and Jon Timmis. Using UML to Model EAE and its Regulatory Network. To appear in ICARIS '09, 2009.

[20] Avital Sadot, Jasmin Fisher, Dan Barak, Yishai Admanit, Michael J. Stern, E. Jan Albert Hubbard, and David Harel. Toward Verified Biological Models. *IEEE/ACM transactions on Computational Biology and Bioinformatics*, 5(2):223–234, April-June 2008.

[21] Susan Stepney, Fiona Polack, and Peter Welch, editors. *Proceedings of the 2008 Workshop on Complex Systems Modelling and Simulation, York, UK, September 2008*. Luniver Press, 2008.

[22] Bart R. Tambuyzer, Peter Ponsaerts, and Etienne J. Nouwen. Microglia: gatekeepers of central nervous system immunology. Journal of Leukocyte Biology, published online. doi:jlb.0608385, November 2008.

[23] Xiaolei Tang, Igor Maricic, Nikunj Purohit, Berge Bakamjian, Lisa M Read-Loisel, Tara Beeston, Peter Jensen, and Vipin Kumar. Regulation of immunity by a novel population of Qa-1-restricted CD8$\alpha\alpha^+$tcr$\alpha\beta^+$ T cells. *The Journal of Immunology*, 117:7645–7655, 2006.

A Domain Model Assumptions

This appendix details the assumptions that have been made in creating the present domain model of EAE. Assumptions are labelled with the cell or phenomenon that they correspond to, and are numbered.

CD4Th1-1. All CD4Th1 cells considered in this domain model are specific for MHC-II-MBP complexes only, though their individual affinities for the complex may vary. An implication of this assumption is that the spatial/binding events brought about by Th cells of other specificities are absent.

CD4Th2-1. All CD4Th2 cells considered in this domain model are specific for MHC-II-MBP complexes only, though their individual affinities for the complex may vary.

CD4Th2-2. CD4Th2 cells do not license APCs in this model, and they do not provide "help" to any cell (the model contains no B cells).

CD4Treg-1. All CD4Treg cells considered in this domain model are specific for MHC-II-Fr3 complexes only, though their individual affinities for the complex may vary.

CD8Tc-1. All CD8 Tc cells considered in this domain model are specific for MHC-I-MBP complexes only, though their individual affinities for the complex may very.

CD8Tc-2. A single CD8 Tc cell can induce apoptosis in at most one CNS cell at any specific point in time.

CD8Treg-1. All CD8 Treg cells considered in this domain model are specific for MHC-I-CDR1/2 complexes only, though their individual affinities for the complex may vary.

CD8Treg-2. A CD8 Treg can induce the apoptotic death of at most one CD4Th1 cell at any point in time.

CD8Treg-3. CD8Treg cells as represented in this model cannot induce apoptosis in APCs that express MHC-I-CDR1/2, although this is potentially possible *in vivo*.

TCell-1. Cytokine secretion by a T cell is assumed to have only two rates of secretion: none at all, or a steady rate of secretion. There is no notion of variable secretion based on a cell's stimulation, or any other effect.

TCell-2. Activated T cells have associated with them an "excitation" level. This is an abstraction of the cell's internal metabolic activity.

TCell-3. Signal 1 and 2, delivered through bindings with MHC and co-stimula-tory molecules, can only be derived through *simultaneously* binding sufficient such molecules; there is no notion of how many molecules were "recently" bound. *In vivo* this may not necessarily be the case.

TCell-4. Where applicable, a CD4 T cell cannot simultaneously license an APC and proliferate.

Cell-1. A cell can interact with at most one other cell at any point in time. For example, a T cell can bind with molecules expressed on only one APC at a time. *In vivo* these aspects are dictated by the physical space surrounding a cell, and what occupies that space.

CNS-1. The "CNS Cell" of this domain model is an abstract represention of the various MBP-expressing cells of the *in vivo* central nervous system.

CNS-2. An apoptotic CNS cell can be phagocytosed by only a single dendritic cell.
CNS-3. CNS cells do not reproduce/divide.
CNS-4. CNS cells do not incur natural death.
CNS-5. Upon phagocytosis by an APC only MBP is received by that APC. No other molecules that might stimulate the APC are derived from phagocytosis of a CNS cell.

CNSMacrophage-1. CNS Macrophage is an abstraction of microglia and macrophages that reside within the CNS during EAE. A study of the literature has revealed that there is currently no consensus from which the functions of macrophages and microglia can be distinguished within the context of EAE [5, 11, 18, 22]
CNSMacrophage-2. Secretion of TNF-α, ROS, and NO by CNS macrophages is at a constant rate, and occurs only when the cell is heavily stimulated.
CNSMacrophage-3. CNS macrophages in this model do not secrete any cytokines, other than TNF-γ.

Cytokine-1. This domain model represents all type 1 and pro-inflammatory cytokines as one cytokine abstraction, called "type1". Where a specific cytokine (for example INF-γ) exhibits some function that is not well represented by this abstraction, that specific cytokine is explicitly represented, but only to serve the concerned function.
Cytokine-2. The model represents all type 2 cytokines as one cytokine abstraction, called "type2".
Cytokine-3. Despite being a pro-inflammatory type 1 cytokine, INF-γ is not depicted in this model to suppress CD4Th2 cell activity, since that is already handled by assumption 1.

DC-1. A dendritic cell can provide signal 2 to only a single T cell at a time.
DC-2. A dendritic cell in this domain model will never die, though its expression of MHC-peptide levels is variable.
DC-3. Dendritic cells in this model do not secrete any cytokines.

MHC-1. The only MHC-peptide complexes considered in this domain model are: MHC-I-MBP; MHC-II-MBP; MHC-I-CDR1/2; MHC-II-Fr3. No other MHC-peptide complexes are considered integral to EAE or its recovery.

Co-stimulatory-1. CD4Th1, CD4Th2, and CD8Tc cells all require equal quantities of co-stimulatory molecule bindings to derive signal 2.

Apoptosis-1. We have omitted any notion of an "anergy" state, since anergic cells can be rescued though receipt of signal 2. T cells that spend sufficient time in a "partially activated" state will become apoptotic.

Apoptosis-2. *In vivo*, interaction between an anergic T cell and an APC can have a regulatory effect on the APC. This is not represented in this domain model.