Associative Memory in an Immune-Based System

C.J. Gibert* and T.W. Routen
Department of Computing Science
De Montfort University
Leicester LE1 9BH, United Kingdom
twr@dmu.ac.uk

**Current Address: Chemin de la Monnerie, Route de la Ville es Blais, 44 380 Pornichet, France.*

Abstract

The immune system offers to be a rich source of metaphors to guide the exploration of the notion of an adaptive system. We might define a class of systems which are inspired by, but diverge from, descriptions of the immune system, and refer to them as *immune-based* systems. The research reported here is motivated by a desire to explore the possibilities of such systems. Specifically, we attempt to construct an associative memory using immune system modelling as a starting point.

1. Introduction

The immune system and more particularly the immune network and the immune response have been compared with neural networks (e.g. Hoffmann, 1986) and classifier systems (e.g. Farmer et al., 1986). Although the immune system presents similarities with these systems, it also has interesting differences from both (Gibert, 1993). Farmer (1991) described some of these similarities and differences in identifying the computation performed by immune networks as a species of connectionism. One can indeed view the immune system as displaying parallel distributed computation. The coherency of the overall behaviour of the system is an emergent property of many local interactions. Farmer further suggested that aspects of the immune system make it more complicated than its sister connectionisms. It would seem therefore, that the immune system holds great potential for machine learning and offers to be a rich source of metaphors to guide the exploration of the notion of an adaptive system.

An impressive amount of work has been done in theoretical immunology to model faithfully the behaviour of the immune system, or some of its components. These models are computationally very complex, although still simplified from an immunological point of view. For the purposes of artificial intelligence, we are not constrained by biological actuality just as those working on neural networks are not constrained by properties of real neurones. Since the immune system has interesting properties, the models of the immunologists might offer a good starting point from which to construct computational models which embody some of those properties and yet which are of practical use. We might define a class of systems which are inspired by, but can diverge from, descriptions of the immune system, and refer to them as *immune-based systems*. We believe that this class currently has few, if any, members. The research reported here is motivated by a desire to begin to explore the possibilities of such systems.

The implicit role of the immune system is to defend the host from external entities which may lead to disease (called *pathogens*). It operates by being able to discriminate between inner (endogenous) and foreign (exogenous) entities. When an entity is recognised as foreign, several mechanisms leading to its destruction are triggered. The end of the *immune response* is normally marked by the absence of the foreign agent, or its reduction to a harmless quantity. Upon presentation of the same pathogen, a *secondary response* is normally generated. A secondary response is characterised by a speedier obliteration of the infectious agent. It is apparent from the differential responses that initial contact with the pathogen leads the immune system to adapt, in order better to be able to deal with the same pathogen subsequently. Thus, it is a form of memory, and it is a form of content-addressable memory since the secondary response can be elicited from a pathogen which is similar, although not identical, with the original one which established the memory (this is known as *cross-reactivity*).

2. Immune system

Many different cells and molecules are involved in the immune response. For the sake of simplicity in this exercise, we only present and model some of these.

2.1. *Recognition*

The basis of the immune system is in cellular and molecular interactions. Certain cells (B-lymphocytes, henceforth *B cells*) synthesise and carry on their surface molecules called *antibodies*. Molecules are three-dimensional structures with uneven surfaces made of projections and indentations. They therefore have shape[1], which is referred to as *specificity*. If two molecules have complementary specificities, they bind to each other (in a chemical reaction); the strength of the bond depending on the degree of complementarity. A fundamental operation of the immune system is the binding of antibodies with other molecules (which are, in that case, called *antigen*) which serves to tag them for destruction by other cells. This process is referred to as *antigen recognition.*

Antigen specificities are characteristic of the B cells which produce them, in the sense that all antibodies produced by any particular B cell have the same specificity. Therefore, we can speak of an antibody recognising the antigen, or of the cell recognising it. It is important, as will become evident later, that, in addition to antigen-antibody binding, there are also reactions between endogenous entities. Antibodies can themselves be 'recognised' by other antibodies.

2.2. *B cells and the immune network*

When a B cell recognises an antigen, it may be *stimulated*, in which case it becomes enlarged and starts replicating (producing identical copies of itself). We refer to a set of identical cells as a *clone*. Since all cells in a clone are identical, a clone can be said to have a specificity. A clone also has a *size*, which is the number of cells in the clone.

Whether a B cell actually *is* stimulated or not, depends on its affinity with present antigens and also with other clones in the system, and their respective sizes. The network formed by clones recognising other clones in the system is referred to as the *immune network* which was postulated by Jerne (Jerne 1973; 1974). This relationship is often formally expressed in terms of the *field* of the clone in many models in the literature.

Whereas neural net models most commonly use threshold activation functions, considerable evidence (Coutinho, 1989) suggests that proliferation of a B cell is well approximated as a bell-shaped function of its general field, mutated by two thresholds, lower and upper, beyond and below which the cell is activated and proliferates. Below the lower threshold, the cell does not respond because too few of its antibody receptors are cross-linked, beyond the higher threshold, the cell stops responding (a phenomenon referred to as *high zone tolerance*); the cell is said to be *suppressed*. Since antibody binding depends on the affinity with antigen or other antibodies, B cells with high affinity are suppressed at lower concentrations than low affinity B cells (Male et al., 1991).

2.3. *Mutation*

During an immune response, dividing B cells are subject to replication "errors". These genetic errors, e.g. gene recombination and somatic mutation, generate cells which produce different antibody specificities. Some of these specificities are not functional in that they will not be able to bind to the present antigen, but others may have an even higher affinity with it. For an interesting exploration of the consequences of different mutation rates see (Weinand, 1990).

2.4. *Meta-dynamics of the system*

The immune system is in a state of constant flux. New clones are produced by the bone marrow continuously. Populations of B cells show high turnover rates, of order of 15%-30% of the total pool per day (Kinkade, 1987; Coutinho, 1989). Since the total population of B cells is almost constant in the immune system a great number of them die each day. When a new clone is created by the bone marrow, if its affinity with other clones present in the immune network is not zero, the clone can proliferate and possibly survive longer that other clones. The immune network is self-organising, since it determines the survival of newly created clones. It also determines its own size (for a detailed discussion see De Boer and Perelson, 1991). This is referred to as the *meta-dynamics* of the system (Stewart and Varela, 1991; Bersini and Varela, 1991).

---

[1] This depends also on other factors such as electrostatic forces, hydrogen bonding, hydrophobic groups and Van der Waals forces. In the remainder of this text, the word shape will refer to the geometrical shape together with all these factors; some authors use the phrase *generalised shape* (Perelson, 1989).

2.5. *Memory*

Immunological research offers two main classes of hypotheses concerning the maintenance of the memory of a pathogen by the immune system. Firstly, cells which participate in a primary response acquire "memory cell" characteristics which distinguish them from other "virgin", cells. Memory cells are thought to decay more slowly or have an infinite life. They are not suppressed by high dose of antigen and proliferate faster.

An alternative view is based on the immune network hypothesis, due to Jerne (1973,1974). As Farmer (1991) postulates "In an immune network a memory can potentially be modelled by a fixed point of the network. The concentrations at the fixed point are held constant through the feedback of one type to another type." (p. 174). The dynamic maintenance of memory in immune network would seem to constitute an attractive approach for an adaptive system, since a particularity of memory in general is its limited capacity; adaptive systems need to be able selectively to forget. Localised memories due to network interactions are hypothesised to be stable memory states of the network (De Boer and Hogeweg, 1989b; Weisbuch, 1990; Weisbuch et al., 1990).

3. Engineering associative memory

Thus, we have the bare bones of a possible model of the immune system. The task we set ourselves was to harness this (albeit crude) model in an attempt to create a content-addressable auto-associative memory.

Inputs to the system are black and white pictures of 64 by 64 pixels and are analogous to antigen. Our aim is to present these 'antigen', initiate a 'primary response' which creates the memory of the antigen, and then be able to observe the existence of the memory by prompting a secondary response via either a further injection of the same, or similar, antigen.

3.1. *Equations*

The differential equations we use are discussed and justified in the immunology literature. The generic equation for computing the field f_i, of a clone i, in a system containing n clones is as follows:-

$$f_i = \sum_{j=1}^{n} a_{ij} x_j \quad (1)$$

where a_{ij} is the affinity with which clone i interacts with clone j and x_j is the size of the clone j. It is worthy of note that this equation is isomorphic with the familiar weighted sum of artificial neural networks, where the affinities correspond with connection weights and the sizes correspond with activation levels.

The bell-shaped activation function was first proposed by De Boer and Hogeweg (1989b) and can be produced as the product of two sigmoid functions as in equation (2):-

$$f(h) = \frac{h}{\theta_1 + h} \times \frac{\theta_2}{\theta_2 + h} \quad (2)$$

This function has been studied extensively (e.g. De Boer and Hogeweg, 1989a, b; Weisbuch, 1990; De Boer and Perelson, 1991). If $\theta_1 << \theta_2$ is chosen, the maximum value of the function is almost equal to unity, at each threshold the value is almost equal to 0.5.

We can characterise the motion of a clone in the system in the following terms:-

$$\frac{dx_i}{dt} = m_i + x_i(1 + b(f_i) + b(A \times a_{iA}) - d) \quad (3)$$

Here, x_i is the size of clone i, m^i is the daily production of cells in this clone by the bone marrow, d represents the decay of the cells in the system, b is the bell-shaped function (2) and f_i is the field "seen" by the clone i at instant t. A describes the concentration of antigen, while a_{iA} is the affinity of the clone i for the antigen (Farmer, 1991; Weisbuch, 1990; De Boer and Perelson, 1991).

3.2. *Entities and affinity*

A basic requirement for an immune-based system is the development of a way of modelling the entities of the system along with a means for computing the affinities between them. In fact, we are interested not in representing cells and antibodies in all their complexity, but only representing those aspects relevant from the point of view of their interaction (more realistic models have been developed, e.g. Inman, 1978 as described by Stewart and Varela, 1991 or Weinand, 1990). In immunological terms, we are interested in representing only their combining regions. These are called paratopes and epitopes and are best thought of as keys and locks. A cell has a key of a certain shape (its paratope) which can fit a lock of a certain shape (epitope) held by certain antibodies, and by other cells.

We follow Seiden and Celada (1992; Celada and Seiden, 1992) who suggest an extremely simple fixed-length binary representation of the shape of paratopes and epitopes; a form of representation common to many investigations in current machine learning. The affinity a_{ij}, between a paratope and an epitope is then determined

by the number of complementary bits. When $a_{ij} = a_{ji}$, the network is said to be symmetrical. Several models have used symmetrical networks (e.g. De Boer and Hogeweg, 1989a; Weisbuch, 1990) although this represents a simplification since non-symmetrical interactions occur in the immune system (De Boer and Perelson, 1991). Interactions are not symmetrical either when the affinity function does not produce the same affinity between reciprocal interactions, or when paratopes and epitopes are separately represented. Non-symmetrical networks are less stable that symmetrical ones (Hoffmann, 1986).

3.3. *Output: defining the winner*

As Farmer et al. (1986) note that there is no clear analogue for output in the actual immune system, which simply seeks to remove antigen. We reasoned as follows. If there were an output of the immune system then it would surely be related somehow to the successful termination of the immune response. A natural assumption to make is that, if we can obtain interesting output from the system, it must surely be related to the clone which is, in some sense, the *most relevant effectors of the destruction* of the antigen.

We shall refer to this clone as the *winner* of the response, and define output in terms of it. We shall see later that it is not straightforward to define the winner. This is not necessarily a disadvantage; many possibilities present themselves and suggest interesting avenues of investigation.

3.4. *Simulation*

A constant and continuous production of clones by the bone marrow is simulated. At each time step, a certain number of randomly generated clones are inserted into the system. If a new clone has the same paratope and same epitope as an existing one, the size of the existing clone is increased by the size of the new one. The initial size of a clone is chosen so that no network activity is initiated. Since clones decay, the number of clones and the total population of cells are constant in the system. The parameters are chosen so that the number of clones is large enough for the probability of recognition of an antigen to be equal to unity. Therefore, the *repertoire* is complete, as in the real immune system. When an antigen (i.e. a pattern to be remembered by the system) is injected, all the clones present in the system follow equation (3). The clones which recognise the antigen start expanding. When the sizes of these clones reach a sufficient level, network activity is triggered. Anti-idiotypic clones are stimulated, proliferate, and then stimulate their own anti-idiotypic partners and so on.

3.5. *Artificial idiotope assignment*

We experimented initially with a non-symmetrical model in which epitopes and paratopes were represented separately. Our idea was to to **force** the memorisation of an antigen. During the period when the antigen is present, the clones directed against that antigen (referred to as Ab1s) expand, exciting their anti-idiotypic partners (Ab2s). In the system, the idiotope of each Ab2 present in the system is assigned the same shape as the antigen and therefore represents an internal image of the antigen.

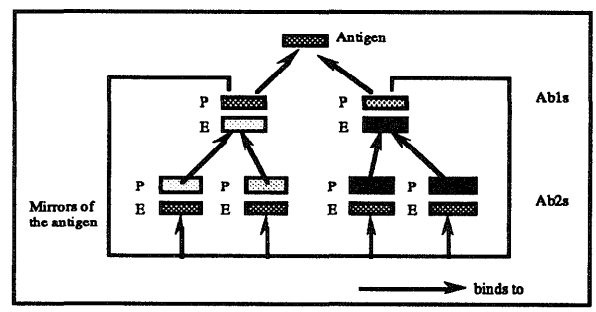

Figure 1. The shape of the antigen becomes the idiotope of Ab2s present in the system. Therefore a loop is created between Ab1s and Ab2.

The idea is forcibly to create recognition loops in the network to enable the maintenance by the network of the clones responding to the antigen. After the removal of the antigen, the size of Ab1s will be large (due to the immune response), so will that of Ab2s (due to stimulation from the expanded Ab1s). Ab1s and Ab2s suppress each other and start decaying until an equilibrium is reached, with Ab1 clones maintaining some Ab2 clones and vice-versa. Depending on their respective sizes at the end of the response, the equilibrium can be reached in a stimulatory state, where the size of the clones oscillate or in a suppressed state where they do not change. This is over-simplified and the memory state may not be stable, depending on the reciprocal affinity of Ab1s and Ab2s, and the presence of other cross-reactive clones. However, no stable state would be possible without the bell-shaped proliferation function. For a more detailed discussion, see Weisbuch (1990) or Weisbuch et al. (1990).

In this model, the winner is defined as the largest clone directed against the antigen upon its removel. The output of the system is taken as the idiotope of the largest anti-idiotypic clone of the winner.

Patterns presented to the system were remembered through the maintenance, within the system, of the clones directed against them by the interactions in the network. However, the system was not stable. Should the field of a clone fall between the activation thresholds, it would proliferate continuously. Since the probability for inserting a new clone which would suppress the clone was small, the system would collapse. Suppression was not the dominant influence, as it is hypothesised to be in the real immune system.

We have investigated ways of stabilising the dynamics of the model. If the field of a clone is modified according the size of the clone, the clone can be prevented from expanding. For example, if, in equation (3), f_i becomes:-

$$\frac{\theta^n}{\theta^n + x_i^n} f_i \quad (4)$$

the system becomes stable. In equation (4), θ is a constant, x_i is the size of the clone i, f_i is the field of the clone i and n is another constant. The constants are chosen according to the other parameters of the system.

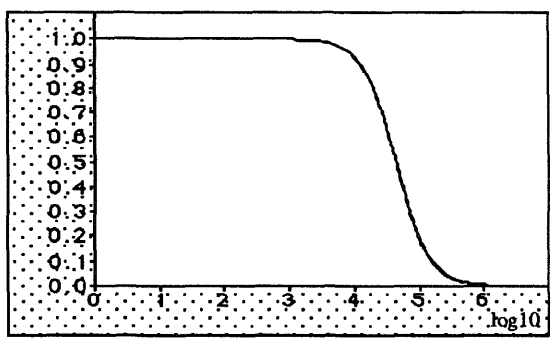

Figure 2 According to equation (4), the function is plotted for θ = 10000, and n = 2.

Equation (4) can be interpreted as follows: when a clone grows large, the field seen by that clone becomes less and less significant. Experimentally, this seemed to stabilise the immune network, *apparently* without modifying the properties of the system. It is not certain that there exist immunological interpretations for equation (4), although it is conceivable that antibody feedback and other regulatory effects within the immune system may have something like this effect.

3.6. Symmetrical network model

Because symmetrical networks are known to be more stable, we decided to experiment with one. Since network interactions are symmetrical, it is assumed that, for a clone responding to an antigen, at least one anti-idiotypic clone is present in the system and therefore will proliferate as its partner proliferates. These interactions should maintain memory cells in the system.

This time, in defining the winner of the response, we took into account the affinity of each clones with the antigen. The winner was chosen to be the clone for whom the product of size with affinity was greatest. This was thought desirable since we had observed that, although a high affinity clone proliferates faster than lower affinity clones, should its size have been low at the presentation of the antigen, it is not certain that its growth during the immune response would be sufficient to ensure that it be the largest clone at the termination of the response. This time, the output was chosen to be the complement of the shape of the winner (since we expect high affinity clones to win, and high affinity clones represent a close or perfect reproduction of the input pattern).

In this model, most clones in the network were suppressed. Memory cells were naturally maintained by the interactions in the immune network, but did tend to dissipate slowly and eventually disappear. The model provided stable network behaviour, but this is not really a surprise (e.g. see Weisbuch, 1990).

However, the system did not show good quality outputs, particularly after secondary responses. Although memory cells expand faster than virgin cells, the output after a secondary response is of relatively poor quality.

At the end of a primary response, the winner was usually the clone which showed the highest affinity for the antigen. Since many clones participated in the immune response, the winner, although having high affinity with the antigen, was not necessarily one which was well integrated in the network. Winners were not always maintained by the interactions of the network because of the absence of partners which could maintain them. Although the dynamics of this system showed better stability, it did not acquire the patterns we desired. There is possibly an analogy here with the ability of a neural network to 'train' apparently successfully, but in

fact have been learning a relationship other than the one intended by its designer.

4. Conclusions

In our experiments, we found two requirements which we found difficult to satisfy simultaneously: remembering patterns, while maintaining system stability. The first model proved unstable, but offered the possibility of forcing the insertion of memory cells into the network. The second model proved stable, but did not allow that forcing and did not systematically maintain the clones we desired.

It seems that interactions in the immune-based system can exhibit memory properties. Nevertheless, the clones which are maintained by those interactions still tend to disappear. It may well be that memory cells should have a slower decay rate, which would compensate for their observed decay. This solution should be experimented with, but we believe that it may well have consequences for the behaviour of the network, since interactions, although symmetrical in terms of affinity would be asymmetrical in terms of motion.

Although we failed to arrive at a satisfactory model, we found our investigations exciting since they took place in an extremely rich, yet relatively unexplored design space: immune-based modelling. We hope that this paper increases the prospects of a wider exploration of that space.

5. References

Bersini, H., and Varela, F. J. 1991. The immune recruitment mechanism: a selective evolutionary strategy. In: Belew, R. K. and Booker, L. B. (eds), *Proceedings of the Fourth International Conference on Genetic Algorithms*, San Mateo: Morgan-Kaufman, 520-526.

Celada, F., and Seiden, P. E. 1992. A computer model of cellular interactions in the immune system. In: *Immunology Today*, 13(2): 57-62.

Coutinho, A. 1989. Beyond clonal selection and network. In: *Immunological Reviews*, 110: 63-87.

De Boer, R. J., and Hogeweg, P. 1989a. Unreasonable implications of reasonable idiotypic network assumptions. In: *Bulletin of Mathematical Biology*, 51: 381-408.

De Boer, R. J., and Hogeweg, P. 1989b. Memory but no suppression in low- dimensional symmetric idiotypic networks. In: *Bulletin of Mathematical Biology*, 51: 223-246.

De Boer, R. J., and Perelson, A. S. 1991. Size and connectivity as emergent properties of a developing immune network. In: *Journal of Theoretical Biology*, 149: 381-424.

Farmer, J. D., and Packard, N. H. 1986. Evolution, games, and learning: models for adaptation in machine and nature. In: *Physica D*, 22: vii-xii.

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

Farmer, J. D. 1991. A Rosetta Stone for connectionism. In: Forrest S. (ed.), *Emergent Computation*, London: MIT Press, 153-187.

Gibert, C. J. 1993. The immune system and machine learning. *M.Sc. Human-Computer Systems Project Report* De Montfort University, Leicester.

Hoffmann, G. W. 1986. A neural network model based on the analogy with the immune system. In: *Journal of Theoretical Biology*, 122: 33-67.

Jerne, N. K. 1973. The immune system. In: *Scientific American*, 229: 52-60.

Jerne, N. K. 1974. Towards a network theory of the immune system. In: *Annales d'Immunologie (Institut Pasteur)*, 125C: 373-389.

Kinkade, P. W. 1987. Experimental models for understanding B lymphocyte formation. In: *Advances in Immunology*, 41: 181-267.

Male, D. Champion, B. Cooke, A. and Owen, M. 1991. *Advanced immunology*. Grower Medical Publishing: London, England.

Perelson, A. S. 1989. Immune network theory. In: *Immunological Review*, 110: 5-36.

Seiden, P. E., and Celada, F. 1992. A model for simulating cognate recognition and response in the immune system. In: *Journal of Theoretical Biology*, 158: 329-357.

Stewart, J. and Varela, F. J. 1991. Morphogenesis in shape-space. Elementary meta-dynamics in a model of the immune network. In: *Journal of Theoretical Biology*, 153: 477-498.

Weinand, R. G. 1990. Somatic mutation, affinity maturation and the antibody repertoire: a computer model. In: *Journal of Theoretical Biology*, 143: 343-382.

Weisbuch, G. 1990. A shape space approach to the dynamics of the immune system. In: *Journal of Theoretical Biology*, 143: 507-522.

Weisbuch, G., De Boer, R. J. and Perelson, A. S. 1990. Localized memories in idiotypic networks. In: *Journal of Theoretical Biology*, 146: 483-499.