

## USING GENETIC ALGORITHMS TO STUDY THE EVOLUTION OF PARATOPES AND ANTIBODIES IN ARTIFICIAL IMMUNE SYSTEMS

**Grazziela P. Figueredo**

*gpfigueredo@gmail.com*

Programa de Engenharia Civil, COPPE, Universidade Federal do Rio de Janeiro, Ilha do Fundão CEP: 21945-970, Caixa Postal: 68501, Rio de Janeiro - RJ - Brazil

**Luis A. V. de Carvalho**

*meucorreioeletronico@gmail.com*

Programa de Engenharia de Sistemas e Computação, COPPE, Universidade Federal do Rio de Janeiro, Ilha do Fundão CEP: 21945-970, Caixa Postal: 68501, Rio de Janeiro - RJ - Brazil

**Helio J. C. Barbosa**

*hcbm@lncc.br*

Laboratório Nacional de Computação Científica/MCT, Av. Getulio Vargas 333 CEP: 25651 070 Petrópolis RJ, Brazil

**Nelson F. F. Ebecken**

*nelson@ntt.ufRJ.br*

Programa de Engenharia Civil, COPPE, Universidade Federal do Rio de Janeiro, Ilha do Fundão CEP: 21945-970, Caixa Postal: 68501, Rio de Janeiro - RJ - Brazil

**Abstract.** *Two binary-encoded models describing some aspects of the evolution in an artificial immune system have been proposed and analyzed. The first model has focused on the evolution of a paratope's population, considering a fixed group of epitopes, to simulate a hypermutation mechanism and observe how the system would self-adjust to cover the epitopes. In the second model, the evolution involves a group of antibodies adapting to a given antigen molecules population. A genetic algorithm was used to form adaptive niching inspired on Coevolutionary Shared Niching strategy ideas taken from an economic model of monopolistic competition where "businessmen" locate themselves among geographically distributed "clients" so as to maximize their profit. Numerical experiments with the two models are presented.*

**Keywords:** *Artificial modelling, Artificial life, Genetic algorithms, Artificial immune systems, Niching strategies*

## 1. INTRODUCTION

The Immune System (IS) is able to protect us from a number of pathogens. It also monitors the organism, searching and destroying anomalous cells. To perform such duties, the IS must recognize a great variety of different compounds and distinguish, among them, those which can remain in the organism and those that are to be eliminated. It is believed that the IS identifies about  $10^{16}$  foreign molecules (Inman, 1978), which means that it can identify any molecule (Forrest et al., 1993).

The IS recognizing pattern is performed through surface receptor molecules of T and B cells. The identification of antigens in both these types of lymphocytes occurs differently. B cells recognize antigens through immune globulins from its cell surface. The T cells can recognize only antigens presented by Antigen Presenting Cells (APCs). The creation of these receptors and their capability to cover all antigens have their origin in a very sophisticated genetic mechanism. During the receptor's formation process, the variation is caused by the combinatorial associations among the receptors codifying genes and the hypermutation mechanism. The hypermutations occur in the lymph nodes' germinative centers. Thus, when an APC penetrates the lymph node and shows an antigen to a T cell, or when a B cell finds a pathogen and identifies it, it means that the combination was well succeeded. After the recognizing pattern is established, the lymphocyte becomes activated, cloning itself. Clones with high capacity of recognizing a certain antigen tend to proliferate. On the other hand, clones with low recognizing capacity disappear and are replaced by others with higher efficiency (Janeway, 2001). The analogy between the clonal selection and the Darwinian natural selection (Darwin, 1859) is clearly seen here .

After recognizing an antigen by a B cell receptor, followed by a sequence of other events, a formation of plasma cells clones responsible for the secretion of the same receptor in its soluble form takes place, so it can bind the antigen. This bind is between the paratope of the antibody and the antigen's correspondent epitope. A paratope that presents a strong bind with the epitope has a greater capacity of neutralizing the antigen. In order to the IS defend the organism efficiently, the paratopes produced by an antibody must be well adapted to the given epitope. This adaptation is refined by somatic mutations.

These concepts will be the building blocks of the models presented in this work which simulates the dynamics between paratopes and epitopes, and between antigens and antibodies inside the organism along the evolution of a species.

Evolutionary Computation has been chosen to implement the models studied in this article. A genetic algorithm (GA) was used to form adaptive niching based on the ideas of Goldberg and Wang (Goldberg, Wang, 1998). In that work, in contrast with fixed shared schemes, a niching formation strategy named *Coevolutionary Shared Niching* (CSN) was proposed to allow for the adaptation of the location and the radius of each niche . CSN was inspired by Tullock's (Tullock, 1967) economic model of monopolistic competition where "businessmen" locate themselves among geographically distributed "clients" so as to maximize their profit.

Two models are proposed in Sections 2 and 3, respectively, which also present numerical experiments. The paper ends with Section 4 which discusses the results of our ongoing work and presents the future steps.

## 2. THE FIRST MODEL

In the first model proposed here, which represents a simplification of what happens in biological immune systems, there is a group of paratopes that have to be adapted through the generations, so they can optimize the coverage of a fixed given group of epitopes. The aim is

to analyze the capability of adaptation of the system in an environment full of aggressive elements, and its behavior due to pattern identification within the epitopes structure as well. That's why the simulations are initially performed with a number of paratopes smaller than the number of epitopes.

It is known that the antibodies are antigen specific, meaning that it is assumed that there is just one paratope able to bind itself to a particular epitope of the antigen molecule. In this model, however, epitopes with slight structural differences can be inactivated by the same paratope.

## 2.1 The algorithm for the first model

This section will give details of the first model. The corresponding pseudo-code is shown in "Fig.1". After the algorithm, each part of it is explained in details.

```

Begin
  Initialize paratopes population
  Initialize fixed epitopes population
  Divide niches according to a distance function
  for i = 1 to maximum number of epoches do
    for j = 1 to size of paratopes population do
      if random() < probability of mutation then
        repeat
          Make mutation in the paratope's j-th chromosome
          Redistribute niches
          if the fitness of paratope j increases
            then keep the mutation
            else discard the mutation and redistribute niches
          end if
        until maximum number of mutations is reached
      end if
    end for
    Evaluate the overall state of the system
  end for
End

```

Figure 1- Pseudo-code for the first model.

**Encoding.** In real biological systems, it is known that the constituting regions of epitopes and paratopes are formed by complex chains of organic compounds. Nevertheless, in this artificial model, epitopes and paratopes are represented by binary chains, following (Forrest et al., 1993) and (Farmer, 1986). Therefore, in the GA context (Holland, 1975) (Goldberg, 1989), phenotypes and genotypes will be the same.

**Initialization.** The initialization of paratopes can be made entirely at random or, according to (Forrest et al., 1993), inserting some pre-defined binary blocks in the chromosome.

**Niche distribution.** The number of niches is always the same as the paratope population size. In terms of the CSN, paratopes and epitopes play the roles of the businessmen and

clients, respectively. The distribution of epitopes among the niches is determined by the smallest distance between them and the paratopes. Each epitope is compared to a paratope, in order to establish which paratope is the closest one and, consequently, which niche the epitope will belong to. The individuals in the  $j^{th}$  niche are the epitopes that the  $j^{th}$  paratope is more apt to neutralize among the current paratope population. The capability of a paratope to neutralize an epitope is measured by means of a distance computation. Here, distance is understood as a function which compares the epitope and paratope chromosomes, also known as matching function.

There are various types of matching functions (Perelson, 1989), however, in this model, the one believed to be most faithful to biological systems was chosen. The chromosomes are compared bitwise, and the matching value is determined by the longest complementary chain between them, as it can be seen in this following example.

*Epitope:* 00000000001111110111

*Paratope:* 11111111111111110111

*Matching value:* 10

The complementary chains represent the molecular bind between a paratope and an epitope. The objective here is to reduce such distance along the evolution. The distance will be given by the formula in Eq. (1):

$$Distance = Epitope's Chromosome Size - Matching Value \quad (1)$$

**Mutation.** The genetic operator used was the classical mutation for binary GAs. The mutation to an individual is retained only when its fitness improves. This procedure makes the search similar to **Hill Climbing** algorithms. It was chosen because there was not an explicit objective function in the system capable to determine gradients to drive the search. Thus, it is the evaluation of a paratope mutation that guarantees a bias to increase performance through the generations, and allows the system to organize itself in the best way to defend the organism.

**System's general state evaluation.** An important question for this model is how to determine the efficacy of the generated system after a number of generations, or, in other words, if it is capable to combat the given epitopes. During the paratope's population evolution, there will be, at least, one minimum site of bind to all epitopes. Nevertheless, it is prudent to say that weak binds are not able to produce efficient neutralizations, since they could break when in contact with other molecules or under slight environmental variation. Concerning this problem, a performance measurement has been established to determine the efficiency rate in fighting aggressors. This parameter was named **Inefficiency Limit**, and its value corresponds to the minimum percentage of an epitope that must be recognized by a paratope so that the latter can be considered inactive.

**Evaluation.** The evaluation is obtained by observing individually the paratope; it is a self-organized system in which what is expected is the individual action of each paratope leading to an efficient global defense system.

## 2.2 Experiments

**The first example.** This first example shows the adaptation of the paratopes according to the fixed epitopes population. It considers an epitope population greater than the paratope population and explores the capacity of the model in recognizing patterns and grouping the epitopes into niches. Experiments have shown that mutation probabilities ranging from 60% to 85% do not interfere on the evolution of the paratopes. The results presented correspond to a typical run with a mutation rate of 85%. The Inefficiency Limit was set to 10%.

As it can be seen in the graph of “Fig. 2”, at the start of the evolution of the paratopes, there is no good performance of the system in recognizing and neutralizing the epitopes. That is why the curve starts indicating a small number of recognized epitopes (vertical axis).

However, along the evolution course (horizontal axis), the artificial hypermutation mechanism enables paratopes to cover the epitopes population. Another graph, presented in “Fig. 3”, shows the improvement of the fitness of the paratopes along the generations. In this experiment, the number of epoches was set to 200 (this number was used in all experiments), chromosome size 65, paratopes population size is 10, and epitopes population size is 300, which is the maximum value to be reached by the paratopes population as shown in the graph of “Fig. 2”.

The results observed show a great similarity to real immune systems. Those who were able to adapt to new pathogens survived and multiplied. In this model, one can simulate a system inefficient in recognizing the epitopes by increasing the Inefficiency Limit parameter.

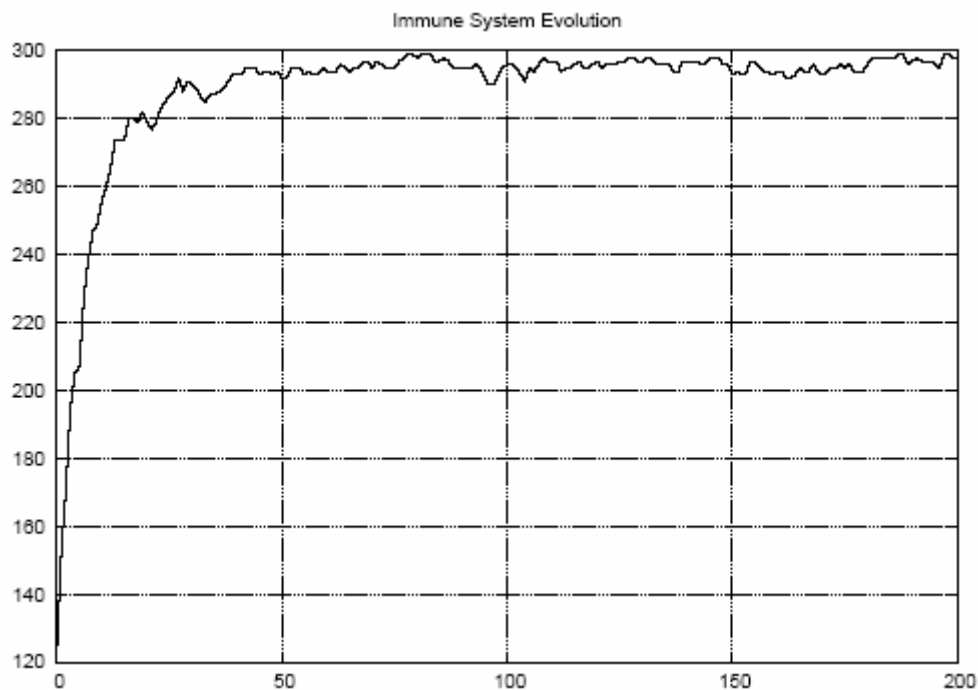


Figure 2- The evolution of the IS considering populations of epitopes and paratopes in different sizes.

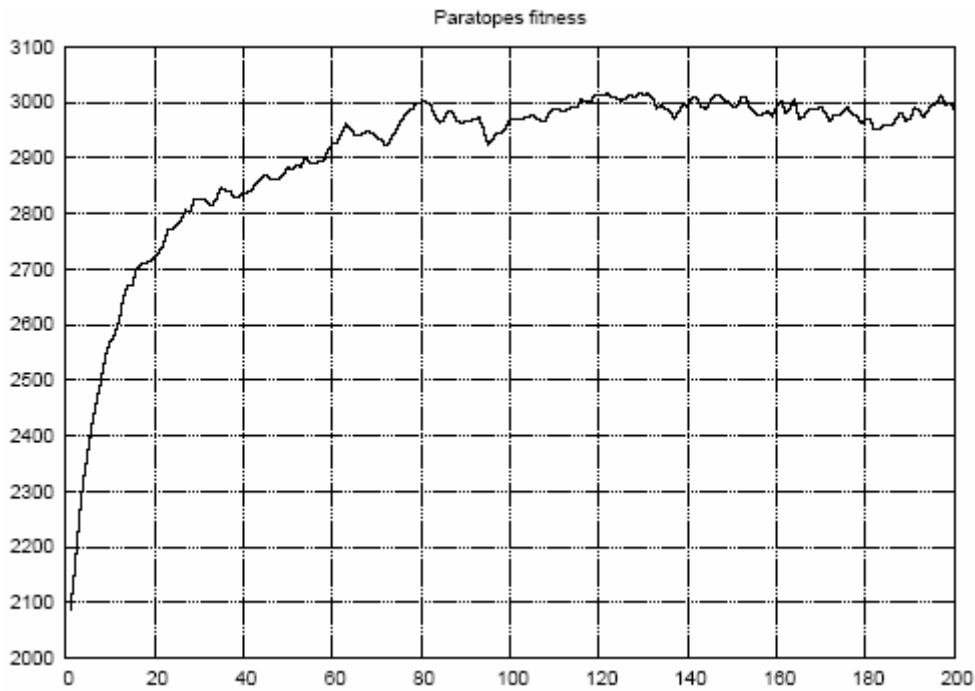


Figure 3- The evolution of the sum of the paratopes fitness – the populations of epitopes and paratopes have different sizes.

**The second example.** In this second example, shown in “Fig. 4”, paratopes and epitopes populations have the same size. Now the model is closer to the real biological systems. What is examined here is the capacity of the system in neutralizing the given epitopes. The Inefficiency Limit was increased to 50%. Both epitope's and paratope's population have a size of 100, and chromosomes 25-bit long. Two hundred epoches were performed. Another graph, presented in “Fig. 5”, shows the improvement of paratopes fitness along the evolution.

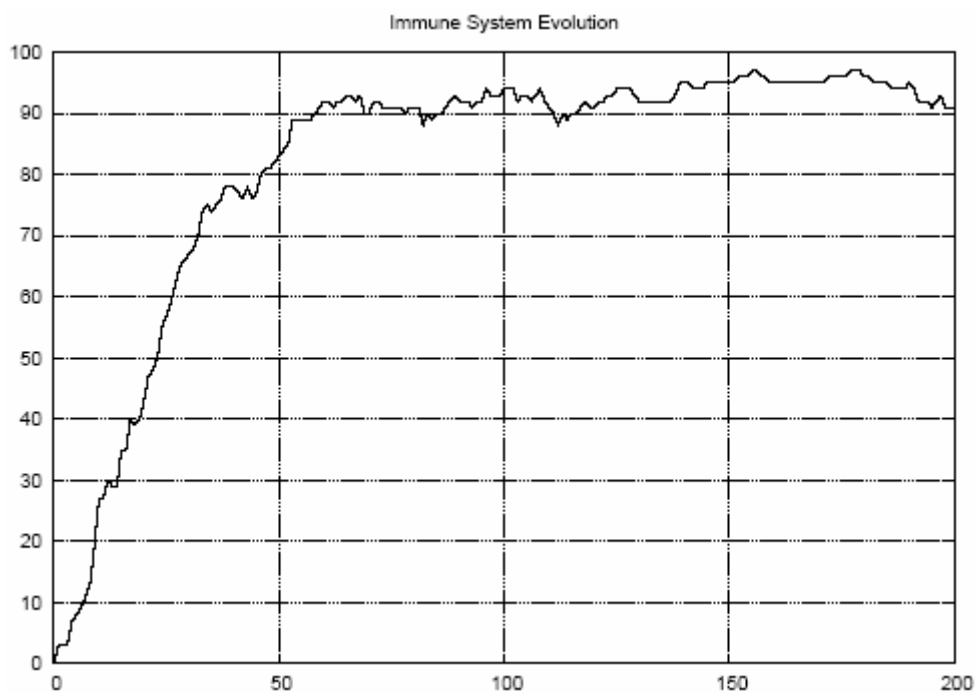


Figure 4- The evolution of the IS considering populations of the same size.

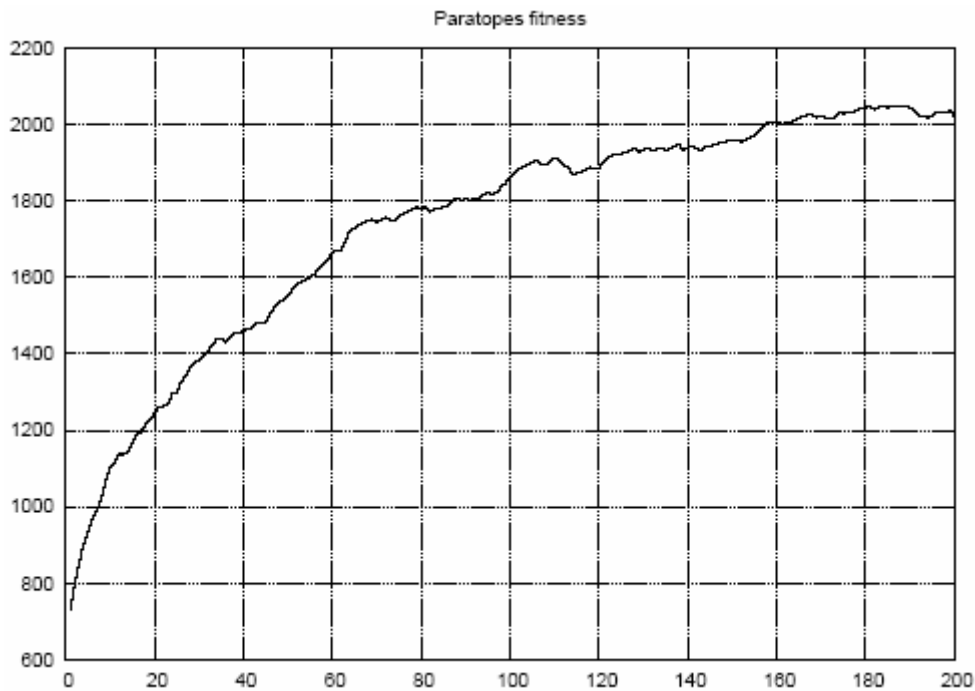


Figure 5- The evolution of the sum of the paratopes fitness –the populations of epitopes and paratopes have equal sizes.

### 3. THE SECOND MODEL

The first model was the simplest prototype elaborated in this work and it presents some limitations. The main limitation is that it broadens the specific antigenic restriction for each antibody. During the implementation and the analysis of this model, an upgrade of the first system was presented as a second version in which the behavioral patterns would be more faithful to real immune systems.

It is known that molecules have to be large, rigid and chemically complex to be considered antigenic. Pathogenic organisms - such as bacteria, anomalous cells or erythrocytes - can start up an immune response because their structure has a complex compound of various molecules that alone are taken as antigens.

As a result, a bacteria could be seen as an antigenic region with a multiple bind site for antibodies, for example. Each site stands for a different antigen. After the study of these notions, it was possible to establish new parameters to improve the model.

The new version does not deal with epitopes and paratopes, but antibodies and antigenic regions. This notion was adopted to make possible the implementation of bind sites of pathogens' molecules, where antibodies could match. Now the antigenic regions are represented by longer bit chains, and antibodies of shorter bit sequences have to bind to antigen sub-chains. These sub-chains represent the antigenic determinants for the molecule.

“Figure 6” shows an example of the new model. Only one antigenic molecule was considered, and a possible configuration of antibodies to neutralize it are also shown in the bottom of the figure.

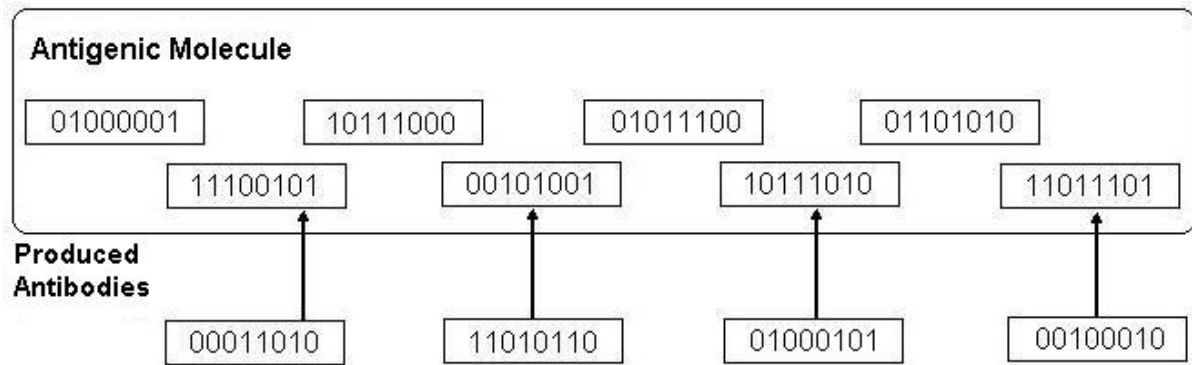


Figure 6- A graphic example of the second model.

Four different antibodies were generated. They were all able to recognize and neutralize antigens within a molecule. The matching rate for antigenic determinant identification was set to 100%. The following sub-sections explain in detail how this new model was implemented.

### 3.1 The algorithm for the second model

The algorithm corresponding to the first model underwent some changes in order to accommodate the additional requirements, as shown in "Fig. 7".

```

Begin
  Initialize antibodies population
  Initialize fixed antigenic molecules population
  Distribute niches according to the matching between antibody and
  antigenic molecule
  for i = 1 to maximum number of epoches do
    for j = 1 to size of antibodies population do
      if random() < probability of mutation then
        repeat
          Make mutation in the antibody's j-th chromosome
          if the antibody generated is unique in the current population
            then Redistribute niches
          end if
          if the fitness of paratope j increases and there's a new antibody
            then keep the mutation
          else discard the mutation and redistribute niches
          end if
        until the maximum number of mutations allowed is reached
      end if
    end for
    Evaluate the overall state of the system
  end for
End

```

Figure 7- Pseudo-code for the second model.

**Niches distribution.** In this new strategy a role inversion between antibodies and the antigenic molecules occurs. The niche owner – or “businessman” in the monopolistic

competition model – are now the antigenic molecules, and the antibodies are the “clients”. The decision to alter the original configurations emerged because the new system has the ability to determine, to each antigenic molecule, a new group of binding antibodies. In the model, every antigenic determinant represents an antigen and has a fixed part of the chromosome. As it could be seen in the above example, all parts have the same size, which is also the size of the antibody's chromosome.

The assignment of an antibody to a specific niche occurs when this antibody reaches a certain rate in the matching function when paired to some antigen in the molecule. It is possible to notice a peculiar situation derived from the model evolution. There will be times when an antibody will take part in more than one niche. This means that some niches will intersect. Some of the initial difficulties in obtaining the expected behavior from the model derived from this particular feature. In some examples that had been run in an intermediate model – between the first model and the algorithm shown above – large populations of antigenic molecules were used. This created various identical sub-chain gene patterns of chromosomes within the genotype. Consequently, the antibody population biased these more frequent sub-chains. At first, it seemed natural, for it's believed the greater the antigen number the more attention they draw from defense mechanisms. However, forming various identical antibodies within the population was not the objective of this model.

In the second model, each antibody in the population represents the whole group of antibodies secreted by a plasma cell clone. To solve the problem of identical antibodies, a strategy similar to the one used in CSN (Goldberg, Wang, 1998) was adopted. The algorithm accepts only mutations that generate different individuals from the ones that are already part of the population. This difference is given by the Hamming distance, which must be greater than zero.

***Evaluation.*** For each time the antibody equals the matching function to a given antigen, the antibody's fitness is increased by one point. If it happens that the matching rate is below the minimum value required to indicate a bind's location, a score lower than one is added to the antibody's fitness. The score value is found by dividing the matching rate by the smallest chromosome size. This method is used to avoid loss in combinations that could potentially excel in future generations.

***System's general state evaluation.*** In this second model, there's also the individual evaluation for the antibodies. This leads the system to self-adjust in order to increase its covering of the antigen group. Relevant to the system, nevertheless, is not only the improvement to the antibodies' fitness, but also the system's capability to maximize the neutralization of any antigen given. This characteristic is clearly derived from the improvement of the system.

## 3.2 Experiments

***The first example.*** This second model explores, basically, pattern recognition into the antigenic molecules. The following graphs show the evolution of two instances containing different numbers of antibodies. The first example represents a model with a small number of antibodies, whose mission is to find equal building blocks into the antigens and maximizing the neutralization of the whole antigens population.

The graph in “Fig. 8” shows the evolution of the system producing antibodies able to adapt to the antigenic molecules given. On the vertical axis it is shown the antigenic molecules recognized by the system. The horizontal axis are the epoches. The first example

paratope's population size was set to 50, epitopes 300. The paratope's chromosome size was 8 and epitope's, 64.

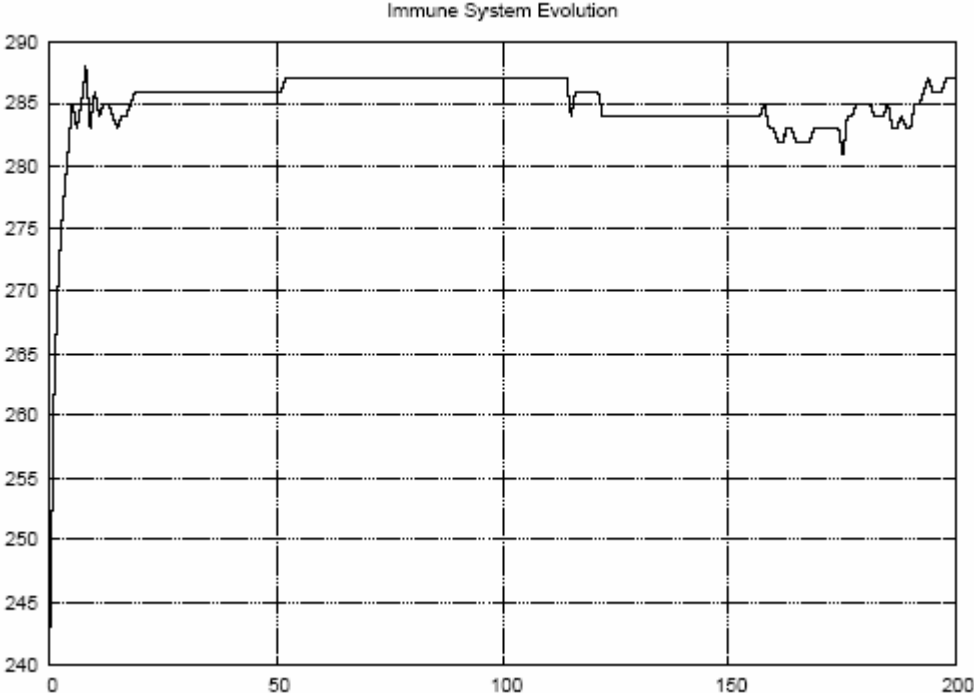


Figure 8- The evolution of the IS in the first example.

**The Second Example.** The second example, shown in “Fig. 9” introduces a greater number of antibodies and shows how this could improve the performance of the immune system. Paratope's population size was set to 50, epitopes, 200.

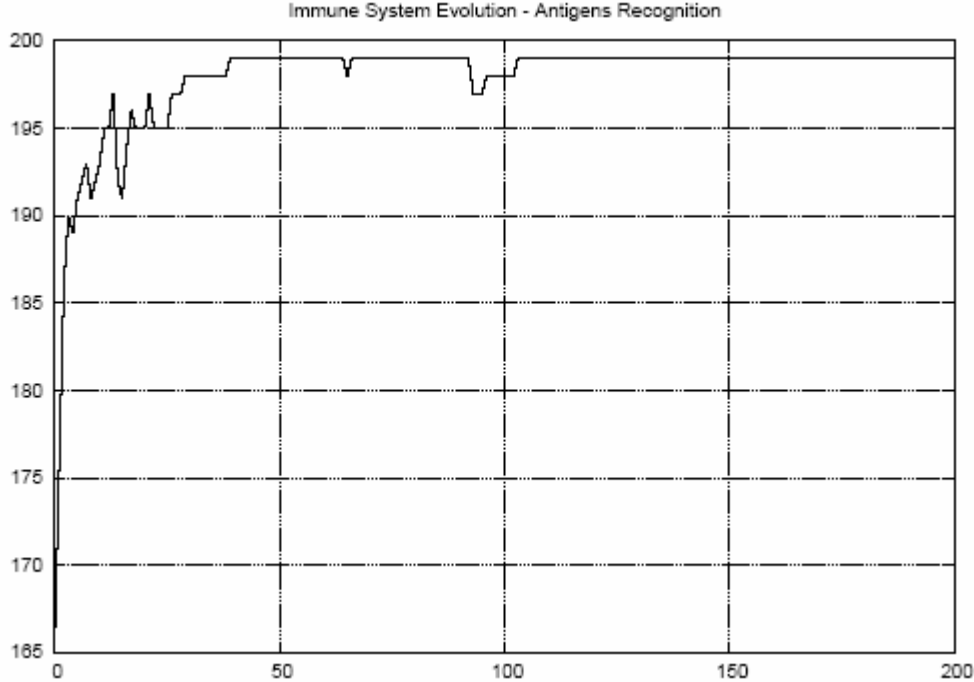


Figure 9-The evolution of the IS in the second example.

## 4. CONCLUSIONS

Understanding how immune systems in mammals have evolved to their present configuration is challenging but it also may be the key to figure out more details of their mechanisms. This paper has proposed two binary encoded models describing aspects of the evolution in an artificial immune system with some characteristics similar to the real biological systems. The first model has focused on the evolution of a paratope's population considering a fixed group of epitopes. The objective of this first experiment was to simulate a hypermutation mechanism and observe how the system would self-adjust to cover the epitopes. This covering capacity is the measure of how well the system could protect an artificial specie along its evolution. The results of this first experiment showed that, at the beginning of the evolution, the paratopes available were not well adapted to the epitopes. However, as the evolution proceeded, the paratopes were becoming much more adapted to the environment presented, being able to recognize almost all epitopes given. The analogy with real immune systems evolution here becomes clear.

The improvement of the first model produced a second model with characteristics more similar to real immune systems. Instead of paratopes and epitopes, now the evolution involves a group of antibodies adapting to a given antigen molecules population. The results of this second experiment has also shown an improvement of the adaptation curve, as in the first model. It has also brought a recognizing pattern apparatus able to find equal blocks into a population of bit strings. This algorithm could be enhanced in order to be applied to more complex systems. As a next step to this modelling work (Figueredo et al. 2005), the concept of gene libraries will be adopted to produce the repertoire of a species antibody together with the simulation of a coevolution between antigens and antibodies.

### *Acknowledgements*

The authors acknowledge the support received from CNPq (301233/86-1 and 475398/2001-7) and CAPES.

The authors would also like to thank the reviewers for the corrections and suggestions which helped improve the quality of the paper.

## REFERENCES

- Darwin, C, 1859. *On the Origin of Species by Means of Natural Selection*. Murray, London. 6<sup>th</sup> ed. (final).
- Farmer, J. D., Packard, N. H., & Perelson, A. S., 1986. The immune system, adaptation, and machine learning. *Physica*, 22D, pp. 187–204.
- Figueredo, G. P., de Carvalho, L. A., & Barbosa, H. J .C., 2005. Coevolutionary genetic algorithms to simulate the immune system's gene libraries evolution. *The First International Conference on Natural Computation*. Wang, L., Chen, K., & Ong, Y. S., eds. Springer-Verlag Berlin Heidelberg. LNCS 3611, pp. 936–939.
- Forrest, S., Javornik, B., Smith R. E., & Perelson , A. S., 1993. Using genetic algorithms to explore pattern recognition in the immune system. *Evolutionary Computation*, vol. 1, n. 3, pp.191-211

- Goldberg, D.E., 1989. *Genetic Algorithms in Search, Optimization, and Machine Learning*, Addison-Wesley, Reading, Mass.
- Goldberg, D. E., Wang L., 1998. Adaptive niching via coevolutionary sharing, *Genetic Algorithms and Evolution Strategy in Engineering and Computer Science*, Wiley, J, & Sons, Chichester. Quagliarella D., Périaux, J., Poloni C. and Winter, G., eds, pp. 21-38.
- Holland, J. H., 1975. *Adaptation in Natural and Artificial Systems*, University of Michigan Press, Ann Arbor, MI.
- Inman, J.K, 1978. The antibody combining region: Speculations on the hypothesis of general multiespecificity. In Bell, G. I., Perelson, A. S. & Pimbley, J. G. H., eds, *Theoretical immunology*, 5<sup>th</sup> ed, pp. 243–278. NY- Marcel Dekker.
- Janeway, C.A., Travers, P. Walport, M., and Capra, J.D., 2001. *Immunobiology: The immune system in health and disease*, . Current Biology/Garland, London.
- Perelson, A.S., Immune network theory, 1989. *Immunol. Rev.*, vol. 110, pp. 5-36.
- Tullock, G., *Towards a mathematics of politics*, 1967. Ann Arbor, The University of Michigan Press.