Evolution of the Immune Repertoire with and without Somatic DNA Recombination

Katsuhisa Takumi and Pauline Hogeweg

Theoretical Biology and Bioinformatics, Utrecht University, Padualaan 8, 3584 CH, Utrecht, The Netherlands

Repertoire of an immune system is a set of antigen receptors each having a unique specificity to bind an antigen. In many vertebrate species, antigen receptors are produced via combinatorial arrangements of DNA segments in specialized immune cells. Due to this molecular mechanism, repertoire of vertebrate species is potentially very large. The diversity of repertoire is thought to guarantee recognition of most ill-causing micro-organisms. In invertebrate species however, similar editing of DNA segments has not been demonstrated to take place. Immune system of invertebrate species therefore seems to operate in a distinct manner from that of vertebrate species.

Using an evolutionary model in which organisms struggle to fight infections, we attempt to understand why some species use a more diverse set of antigen receptors than others. Individuals in our model either use somatic DNA recombination to produce antigen receptors (as in vertebrates) or do not use such a mechanism (as in invertebrates). We found that individuals having an invertebrate-like immune system came to employ only a few antigen receptors to recognize a set of pathogens whereas those with a vertebrate-like immune system use a larger set of more specific antigen receptors to recognize the same set of pathogens. Our interpretation of this finding is that because the genetics of the immune system imposed different constraints on the evolutionary process, two distinct recognition strategies have been adapted by these species.

© 1998 Academic Press Limited

Introduction

We want to find out why vertebrate species have a larger repertoire of antigen receptors than invertebrate species. To our knowledge there are only two published papers which study repertoire diversity using a formal model (de Boer & Perelson, 1993; Percus *et al.*, 1993). Both groups believed that the immune system is the optimal solution to the problem of properly distinguishing foreign antigens from self-antigens, i.e. the self–non-self-discrimination problem. However the conclusion of one group almost contradicts the conclusion of the other because each group defined the problem slightly differently. One group assumed self-tolerance during ontogeny and concluded that the diversity reflects the number of self antigens rather than the number of pathogens (de Boer & Perelson, 1993). The other assumed self-tolerance during evolution and concluded that

the diversity reflects the number of pathogens and is relatively unaffected by the number of self-antigens (Percus *et al.*, 1993). Another study relates repertoire diversity to autoimmune diseases (J. A. M. Borghans & R. J. de Boer, 1998, in preparation) and reaches a different conclusion from the two previous studies.

Unlike the previous studies, our study investigates repertoire diversity by means of an evolutionary model. The model incorporates explicitly replicating individuals and the coding of receptors. Vertebrate species use somatic DNA recombination (Schatz *et al.*, 1992) but invertebrate species do not. In this study, evolution is not modeled as a global optimization process. Our study differs from the previous two studies in two other respects. First, the self–non-self-discrimination problem is not considered. Second, the immune repertoire is not taken to be a set of random specificities. We are interested

in the evolution of functional antigen receptors which recognize a fixed set of pathogens.

Our main finding is that the evolution of the immune repertoire is influenced by the somatic DNA recombination mechanism. In an immune system with the somatic DNA recombination, a large set of antigen receptors is employed to recognize the pathogens. In contrast, only a few antigen receptors are used for this purpose in the invertebrate immune system. This is not a trivial consequence of the recombination because in our model the potential repertoire is the same for both immune systems. Because in our model both immune systems adapt to the same set of pathogens, we conclude that the immune repertoire diversity need not reflect different requirements for species survival. The diversity could simply be a consequence of the evolutionary process constrained by the different genetics.

Model

We consider an evolutionary process in which organisms do their best to avoid infections in order to survive. Healthy organisms have a higher chance of reproducing offspring whereas poor organisms are replaced by these offspring. We assume a constant population size. We divide organisms into invertebrate or vertebrate according to the genetics that encode the antigen receptors. The simulations are carried out separately for the vertebrate and the invertebrate immune systems.

We use a 63 bit binary string to represent a pathogen. An antigen receptor is also represented by a 63 bit binary string. The invertebrate immune system is a set of up to 1000 antigen receptors, each 63 bits long. The vertebrate immune system is three sets of 10 gene segments, each 21 bits long. An antigen receptor is a combination of three gene segments from each set. Thus, there are 1000 possible combinations, each representing an antigen receptor. Thus the potential repertoires of the vertebrate and the invertebrate immune systems are the same. For invertebrates, however, we found that the actual repertoire used by the immune system to recognize the set of pathogens is largely independent of the initial, potential repertoire. Because choosing a smaller potential repertoire for the invertebrates did not affect the result, we could speed up the computation by reducing the potential repertoire for the invertebrates to 100.

COMPUTATION OF FITNESS

By definition, the fitness of an organism is determined by the ability of the immune system to resist the pathogen most dangerous to the organism. This is the pathogen to which all antigen receptors in the immune system have poor affinities. To compute the fitness, we need a well-defined relationship between any antigen receptor and any pathogen. We employ an NK landscape formalism to define relationships. The metaphor using the NK landscape and its interpretation as the affinity has been used previously to model affinity maturation of antibodies via somatic hyper-mutations (Kauffman & Levin, 1987; Kauffman & Weinberger, 1989; Macken & Perelson, 1989).

Let A and B denote two arbitrary binary strings, one representing an antigen receptor and the other a pathogen. To specify positions in a string, we write $A = \{a_1, a_2, \ldots, a_{63}\}$ and $B = \{b_1, b_2, \ldots, b_{63}\}$. For each position a_i in the string A, define an "internal connectivity" I_i to be a set of s distinct positions in the string A, including a_i. Similarly we define an "external connectivity" E_i to be a set of t distinct positions in the string B, including b_i. Then we define a set $C_i = I_i \cup E_i$ to be the connectivity of the position a_i. Note that the connectivity C_i can take 2^{s+t} distinct forms. We assign a random number drawn uniformly between 0 and 1 to each of 2^{s+t} distinct forms. Repeating the procedure for each position a_i of the string A, we complete the landscape. Given any two aligned bit-strings A and B, this landscape gives a well-defined real number which indicates how well the antigen receptor A recognizes the pathogen B. This number can be considered as the affinity of the antigen receptor. We use this landscape as a crude model of receptor protein folding and of interactions with pathogens.

We consider the case in which the organisms use all antigen receptors to combat all pathogens. The antigen receptor having the maximum affinity to each pathogen can be determined from the NK landscape. We take the poorest of all the best antigen receptors in order to define the fitness of the organism. For a similar approach to the question of repertoire diversity, we refer to Hightower et al. (1995).

SELECTION AND REPRODUCTION

One time step is defined as a selection of one variant in the population of 100 individuals, followed by reproduction of one offspring by the selected individual. All variants in the population are ranked according to their fitness values. The probability that a variant will be selected is given by

$$\frac{e^{-i}}{\sum_{j=1}^{100} e^{-j}}, \quad i = 1, \ldots, 100, \tag{1}$$
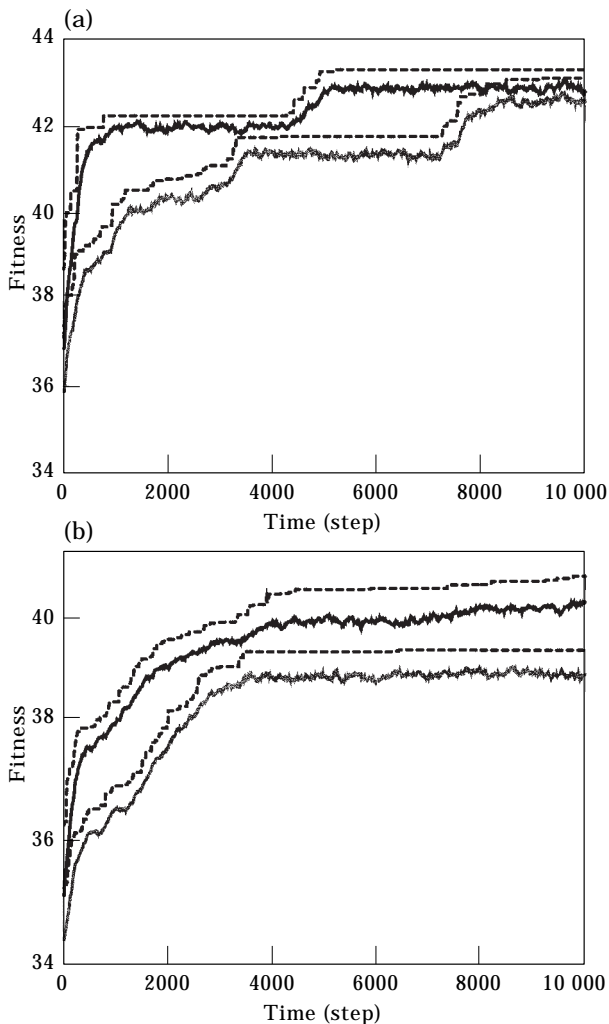
FIG. 1. Time course of the mean and the maximum fitness values. The invertebrate and vertebrate species were challenged by either only 10 pathogens (a) or 600 pathogens (b). In each panel, the upper two lines are for the vertebrates and the lower two lines for the invertebrates. (– – –) indicates the maximum fitness values, and the solid lines indicate the mean fitness values. The host population size is 100. The mutation rate is 0.005 per bit per reproduction. The integral connectivity is 3. The external connectivity is 1.

where i is the rank of the variant. The selected variant reproduces one offspring. The fidelity is determined by the copying error rate of 0.005% per bit in the sequence representing the variant. The vertebrates reproduce asexually. We allowed the invertebrate to reproduce sexually because we thought that sexual recombination might help the invertebrates to evolve better. However, sexual reproduction did not significantly change essential properties of the model. When reproduction occurs sexually, one parent is chosen according to eqn (1), and the other with the probability decreasing linearly with the rank. Eight time out of 10, the offspring then copies genes from the fitter parent.

Results

We present the results of four simulations. In all four cases, the same landscape was used to model the antigen receptors that fold and bind the pathogens. We first investigate fitness curves over time. Then we examine immune repertoires after they have adapted to the pathogens. In our reported result, we set the number of internal connectivity to 3 and the external connectivity to 1. Simulations using other setups are summarized in the section "Robustness of the model".

INVERTEBRATES ARE AS HEALTHY AS VERTEBRATES

Figure 1(a) shows the two mean (solid lines) and the two maximum (dashed lines) fitness curves from the simulations, one for the vertebrate evolution (upper lines) and another for the invertebrate evolution (lower lines). We used a fixed set of 10 pathogens. At the beginning, the vertebrate and the invertebrate immune systems record the mean fitness values 36 to 37. Offspring often have better antigen receptors, and the fitness of both the invertebrate and the vertebrate immune systems increase sharply. Both immune systems finally reach approximately equal fitness values at the 10000 time step.

Because there are only 10 pathogens in the simulation, we take the final fitness value at day 10000 to be a reference value. In the next section, we use the reference value to estimate to what extent the fitness increases within the same time period. If the pathogen population is increased, we expect that the fitness values will be lower than the reference value. We define the reference value because the theoretical maximum is very difficult to obtain for a given genetic algorithm, except in a special case.

Figure 1(b) shows the mean and the maximum fitness curves obtained from a similar simulation setup except that we have now increased the number of pathogens to 600. The final fitness values are lower than those in the previous simulations, as we expected. At day 10000 the vertebrate immune system records a somewhat higher final fitness value than does the invertebrate immune system. From a number of simulation runs, we found that the final fitness value of the invertebrate immune system did not change significantly with the size of the potential repertoire. We conducted a few simulations in which each invertebrate immune
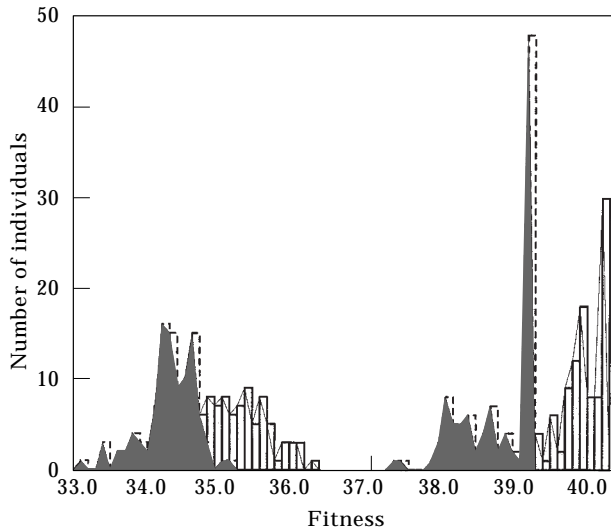
FIG. 2. Initial and the final fitness distributions of the two species are shown for comparison. The distributions of the vertebrate population are shaded in a light color. Those of the invertebrate population are shaded in a dark color. On the left half of the panel, the two initial distributions at time step zero are shown. The final distributions at 10000 time step are shown on the right half of the panel. The data refer to the simulation shown in Fig. 1(b) using 600 pathogens.

system has the capacity to store up to 3000 receptor genes. The results were always the same: the invertebrates reached almost the same fitness values in all cases. These fitness values were always somewhat lower than the fitness values of the vertebrate species.

Figure 2 shows how fitness values are distributed among populations of the invertebrate species (dark color) and of the vertebrate species (light color). The initial distributions at time zero (those on the left) do not overlap with the fitness distributions after 10000 time steps (on the right). Therefore we can rule out the possibility that the number of fittest individuals in the initial population has increased. Then, it must be that innovative receptor genes have caused the fitness level to rise. The highest dark bar indicating the fittest invertebrates at 10000 time step make up 48% of the total population. The high percentage indicates that there has been no major innovation in the receptor genes for a long time. It also indicates that under the given mutation rate and the given selection scheme in the model the existing genes are inheritable for generations. The initial fitness distribution in Fig. 2 is somewhat lower for the invertebrate species because we speed up the computation by reducing the potential repertoire of the invertebrates. However, as we have already explained, the end result is almost the

same even if we did not reduce the potential repertoire.

ADAPTING DIFFERENT RECOGNITION STRATEGIES

In order to reveal the evolutionary process underling the fitness curves shown in Fig. 1(b), we study the repertoire of the vertebrate and the invertebrate immune systems. We found that the invertebrate [Fig. 3(a)] and the vertebrate [Fig. 3(b)] immune systems have developed different ways of using the antigen receptors in order to recognize the same set of 600 pathogens. The horizontal distribution of the circles indicates the repertoire diversity. We see that in Fig. 3(b) the circles are distributed more widely that in Fig. 3(a). Thus the vertebrate repertoire is more diverse than the invertebrate repertoire.

The vertebrate immune system used 76 antigen receptors. On the other hand, only six antigen receptors were used in the invertebrate immune system, although the potential repertoire is much larger. The three vertical chains of circles in Fig. 3(a) show that the three antigen receptors are heavily used to recognize hundreds of pathogens. To confirm the generality of the result, we ran several simulations with different numbers of genes, with different initial host populations, and with different selection criteria. All simulations showed the same trend: invertebrate immune systems used only a few antigen receptors. Therefore the trend for the invertebrate species to adapt few antigen receptors appears to be a robust property of the model.

Specificities of the antigen receptors shown in Fig. 3(a) and (b) were further analysed. Out of the six antigen receptors in the invertebrate repertoire [Fig. 3(a)], three receptors recognized broad ranges of the pathogens. Each of the three receptors recognized approximately 200 pathogens. According to our definition, an antigen receptor recognizes a pathogen when it recognizes the pathogen better than any other receptor in the repertoire. In the vertebrate immune system on the other hand, the most broadly reacting antigen receptor recognized 78 pathogens. Moreover, an analysis of the data revealed that about 50 antigen receptors in the vertebrate repertoire recognized fewer than five pathogens. The analysis also revealed that in both immune systems the affinities of a single antigen receptor to all 600 pathogens usually ranged between 30 and 45 (data not shown). Therefore no single antigen receptor could recognize all the pathogens. Since we thought that the vertebrate immune system might also be using only a few antigen receptors, we removed three to six of the most broadly reacting
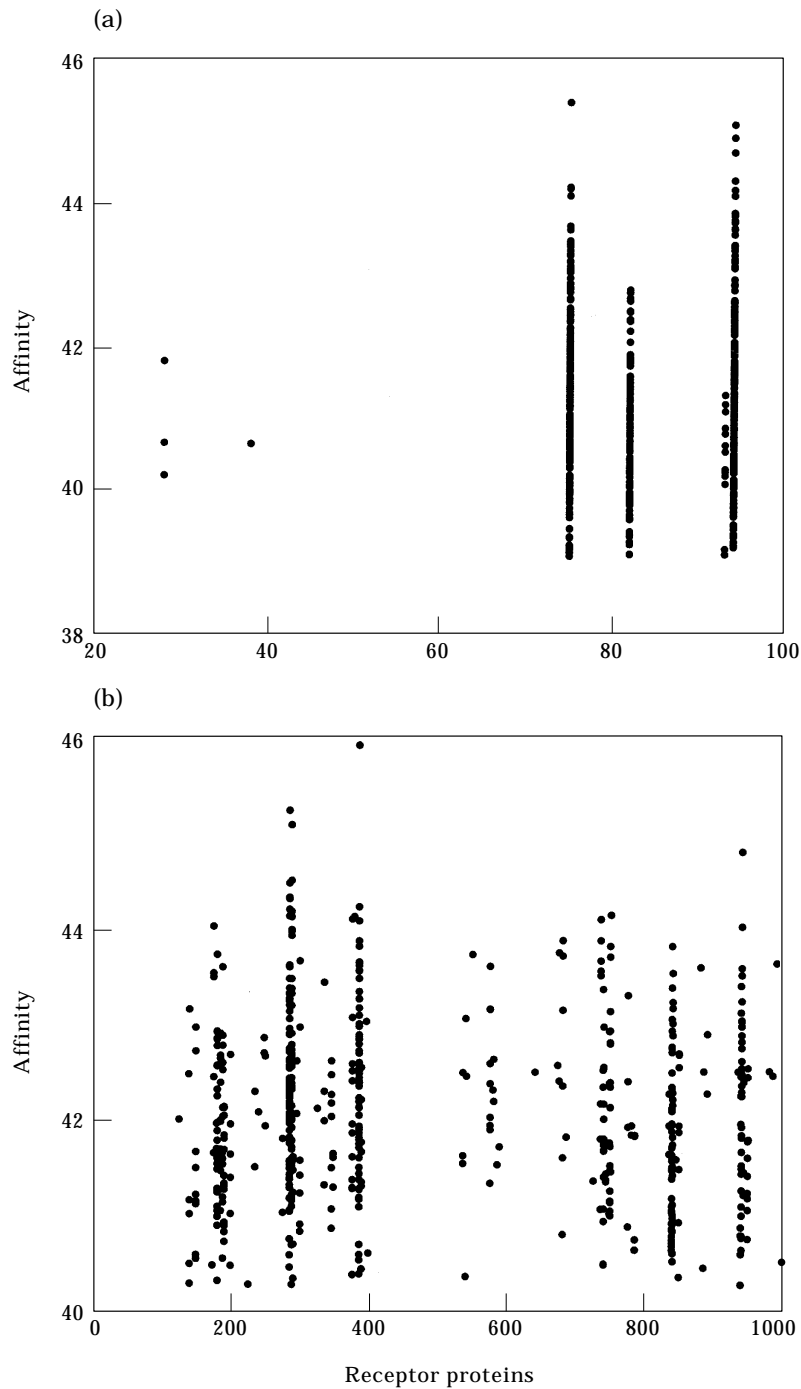
FIG. 3. Adapted repertoire of the invertebrate (a) and of the vertebrate (b) are shown for comparison. The spread of dots along the horizontal axis is a measure of repertoire diversity. At the end of the simulation shown in Fig. 1(b), the fittest individual was taken from each population for this analysis. To construct each panel, we selected for each of 600 pathogens used in Fig. 1(b), the antigen receptor having the highest affinity to the pathogen. Then using dots, we plotted the best antigen receptors (horizontal axis) vs. the highest affinities (vertical axis). Thus, if pathogens tend to be recognized by the same receptor, we will see a long vertical strip of circles as in (a). If pathogens tend to be recognized by different antigen receptors, we will see circles widely distributed along horizontal axis as in (b). Affinities were computed using the NK landscape as defined in the Model section.

antigen receptors from the repertoire. Using the modified repertoire, the vertebrate immune system recognized all 600 pathogens, and the fitness value did not change significantly. Since these very broadly reacting antigen receptors could potentially be auto-reactive, they are more likely to be deleted by some self tolerance process (Goodnow *et al.*, 1988; Nemazee & Burki, 1989). It is interesting that the remaining more specific antigen receptors could also recognize the pathogens after these very broadly reacting receptors had been deleted. Concerning the formation of a B cell repertoire when a self tolerance process is active, we refer to Takumi & de Boer (1996).

TRANSGENIC EXPERIMENTS IN SILICO

We want to demonstrate more convincingly that the two immune systems adapt different modes of recognition. We designed a vertebrate immune system which initially used a small repertoire. Then we simulated evolution to see whether the immune repertoire would become more diverse. We constructed the vertebrate immune system by transferring the four antigen receptor genes obtained from an invertebrate at time step 5000. Thus the vertebrate immune system can express $4^3 = 64$ antigen receptors via the artificial somatic recombination. We then continued the simulation for 5000 time steps. At the beginning, 27 antigen receptors out of the total of 64 are used to recognize the 600 pathogens. The diversification at this stage is an immediate consequence of the somatic recombination mechanism. After 5000 time steps, a descendant of the vertebrate used 93 antigen receptors. Thus, the vertebrate repertoire has diversified more than three times since the initial small repertoire. The diversification is accompanied by an increase from 39.1 to 40.1 in the mean fitness value. Thus, we conclude that the diversification of the repertoire is selected for during evolution.

We also performed the transgene experiment in reverse and found that the invertebrate immune repertoire decreased in diversity. Seventy-seven antigen receptors in a vertebrate immune system at time step 5000 were transferred to an invertebrate immune system. After 5000 time steps, the number of antigen receptors decreased to 39. Only six of these receptors were identical to the transferred genes. The mean fitness value at the beginning was 40.1 and decreased over the 5000 time steps to 39.5. In a longer simulation with up to 10000 time steps, we found that the number of antigen receptors still decreased, this

time to 34. This repertoire contained only two of the original receptors. We conclude that the invertebrate immune repertoire did indeed become less diverse.

The trend for the invertebrates to adapt only a few antigen receptors can be partially explained using a concept of error threshold (Eigen & Schuster, 1979). According to this concept, there is an upper bound (error threshold) on the length of replicating RNA strings, beyond which no useful genetic information can be transmitted over generations. Close to the error threshold, only a few individuals in the population of RNA strings have the fittest genome because most of their offsprings will be mutilated or severely damaged by mutation. Conversely, sufficiently below the error threshold, individuals having the fittest genome form a majority, and "quasi-species" are sharply defined centering around these fittest individuals. An example of such quasi-species is the population of the invertebrate species shown in Fig. 2 indicated by a high number of the fittest individuals (the dark shaded fitness distribution on the right). Since the error threshold is thought to be a general limitation on any replicating entity subjects to mutations, it should also apply to the species in our
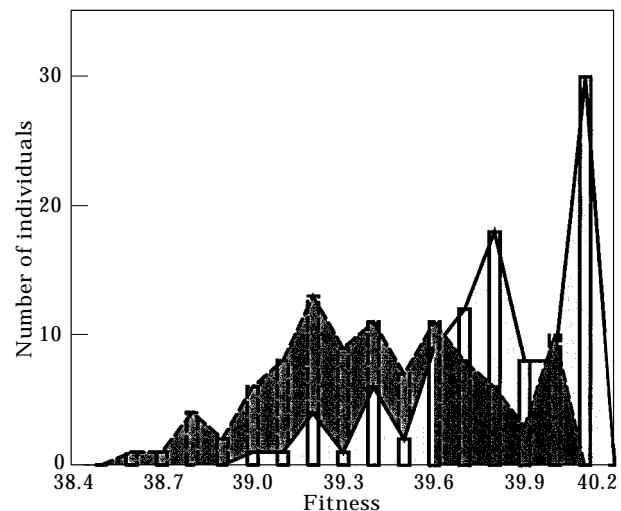

FIG. 4. These fitness distributions are supportive evidence for the error threshold based explanation. The distribution of the transgenic invertebrate population is shaded in a dark color and that of the control population in a light color. The vertebrate population at 10000 time step of the simulation shown in Fig. 1(b) was taken as the control. The genome of the transgenic invertebrates were obtained from the fittest vertebrate at 5000 time step in Fig. 1(b). A simulation using these transgenic invertebrates was continued for additional 5000 time steps. The dark shaded distribution shown in this panel was obtained at the end of this new simulation. During the simulation, the number of functional antigen receptors had decreased to 39 from initial 77.

model. Granted that the rate of mutation per bit per reproduction in our model is the same, the limit on the length of a bit-string representing an individual should be similar for both species. However, to encode the same number of useful antigen receptors, a bit-string representing the invertebrate will be much longer than that representing the vertebrate. For instance, the invertebrate with 77 antigen receptors is a string of about 5000 bits in our model, whereas the vertebrates kept the same number of receptors using only 630 bits. Consequently, much fewer numbers of antigen receptors can possibly be maintained in the genome of the invertebrates. Thus, by applying the concept of the error threshold, we can explain why the number of functional antigen receptors in the transgenic invertebrates decreased.

To test whether the error threshold is indeed the reason, we analysed the genome of the transgenic invertebrates and found that in the whole genome of about 5000 bits, approximately 6% had been mutated. Thus the population had drifted from the initial population over 5000 time steps. To confirm that this drift is correlated with the decreased mean fitness value, we additionally compared the fitness distribution of the transgenic invertebrates with that of a control. The population of the vertebrates at 10000 time steps was used as the control because the control shares the same ancestor and has the same evolutionary time-span of 10000 time steps. Figure 4 shows the comparison. We observe that in the distribution of the control population (shaded in a light color) the fittest individuals define a sharper peak. Conversely, for the transgenic population (dark

TABLE 1

The numbers of distinct functional antigen receptors, showing that the vertebrates tend to use more antigen receptors to recognize the same set of pathogens than do the invertebrates

Cases	Vertebrates	Invertebrates
A	5(2 4 1)	3 (3 3 3)
B	76 (5 6 8)	6 (6 6 6)
C	10 (5 3 6)	7 (7 7 7)
D	90 (6 4 10)	9 (9 8 9)
E	163 (6 10 6)	12 (12 12 12)

The data were obtained by analysing the genome of the fittest individuals after 10000 time steps in each simulations. These in parentheses are the numbers of unique gene segments in each of three blocks of antigen receptors. The cases A and B are as described in Fig. 1 (using 10 pathogens) and 2 (using 600 pathogens), respectively. In the cases C (10 pathogens) and D (600 pathogens), different NK landscape structure was used to test robustness of the model properties. In these simulations, both the internal and the external connectivities are equal to one. In case E (600 pathogens), we defined the fitness as the average of all the best antigen receptors

TABLE 2

Summary of maximum fitness values from several simulations using different set-ups

Cases	Vertebrates			Invertebrates		
	Initial	Final	Test	Initial	Final	Test
A	38.4	43.3	36.9	37.1	43.1	35.8
B	36.3	40.3	38.6	35.1	39.1	37.2
C	35.4	39.8	32.9	35.1	39.0	32.1
D	33.0	36.6	34.5	32.9	35.4	32.5
E	38.9	44.0	—	38.6	42.8	—

The cases are the same as those explained under Table 1. The columns "Initial" and "Final", respectively refer to zero and 10000 time steps of simulations. The column "Test" refers to the simulation using evolving pathogens. After each simulation, the fittest individual was further challenged by 5000 evolving pathogens. The fittest individual was taken to be no longer mutating. The pathogens evolved by point mutation until they can no longer lower the fitness value of the organism

color), fitness values are distributed more widely and are generally less than that of the control. We conclude that although the invertebrate species could maintain the highest fitness value, their genome drifted through the sequence space. This drift is towards a phenotype in which only few receptor genes are used. This phenotype reduces the risk that offspring often are mutilated by mutation.

TESTING THE IMMUNE SYSTEM WITH EVOLVING PATHOGENS

In our experiments we concentrated on the host dynamics. The simulations are concerned solely with the fixed population of pathogens. One would expect that a mutant arising from the pathogen population would make the immune systems completely useless. Therefore we tested the immune systems of the fittest individual in an environment of evolving pathogens. The resulting fitness values are summarized in Table 2 under the label "Test". Fitness values decreased in all cases as expected. However, the vertebrates that have been challenged by 600 pathogens (cases B and D) seem to resist better against the evolving pathogens. Because both the vertebrates in cases B and D have larger number of antigen receptors (see Table 1), the resistance seems to be correlated with the receptor diversity.

ROBUSTNESS OF THE MODEL

The results which we reported in this paper appear to be robust properties of the model. We have done a number of simulations using different settings. In all of these simulations, we generally observed behaviours similar to the reported case. Sexual recombination did not change the result significantly. The same is true for the selection criteria. In one

simulation, the definition of fitness was changed to the average of all the best antigen receptors. In addition, we have tested the model using a variety of parameter values. However, we always observed very similar behavior in those simulations. In particular, the repertoire of the vertebrate immune system was consistently larger than the repertoire of the invertebrate immune system. In Tables 1 and 2, we listed repertoire sizes and fitness values obtained from several simulations, including the reported case. We have also tested the following mutation rates: 0.0005, 0.005, 0.05, 0.1 and 0.2. The fitness values increased most rapidly when we used the mutation rate 0.005.

Discussion

We have demonstrated that the vertebrate and the invertebrate immune systems reflect two different courses of evolutionary adaptation in the environment of the same pathogens. Both adaptations seem to be equally good ways of recognizing the same set of pathogens. By analysing the data, the repertoire size is found to depend on the coding of the antigen receptors. The invertebrate immune system adapted so that it used only a few antigen receptors although potential repertoire was much larger. We explained this trend based on the error threshold. Such trend to decrease a functional repertoire should also operate in the real invertebrate immune system.

We separated in our model the evolution of the immune repertoire and the self–non-self-discrimination problem. The ideal adaptation may then seem to be a single universal antigen receptor that has come to recognize everything non-specifically. However, this is not what we observed in our model. Implicit in this intuition is the assumption that the ideal adaptation can indeed be realized in evolution. This assumption is not true in general because evolution is always constrained. Interestingly, our model demonstrates that the antigen receptors were far more specific in the vertebrate immune system than they needed to be according to the NK landscape (as seen in the invertebrate immune system). Self–non-self-discrimination was therefore not necessary for the evolution of the specificity. Moreover, the evolved repertoire is not random and recognizes new pathogens far better than does a random repertoire.

An interesting open question is whether there is a model of receptor–pathogen interactions (the NK landscape was our choice) which causes the fitness values to be significantly different for invertebrates and vertebrate immune systems. Another open question is whether the two coding schemes residing in the same immune system can coexist over an evolutionary time scale. The human immune system does indeed use the two coding schemes, one for innate immunity and the other for adaptive immunity. In fact, it has been shown that innate immunity plays a critical role in inducing the adaptive immune response (Medzhitov *et al.*, 1997).

In summary, the message of this paper is that repertoire diversity reflects the different ways in which organisms adapts to the same environment. For a similar result regarding adapting strategies, we refer to Pagie & Hogeweg (1998). In our investigation, we did not assume self–non-self-discrimination. Two distinct genetic organizations were all that were necessary to demonstrate the discrepancy in the repertoire diversity of the vertebrate and the invertebrate immune systems. We propose that this bias to a higher repertoire diversity due to encoding may have set the stage for further evolutionary processes.

REFERENCES

DE BOER, R. J. & PERELSON, A. S. (1993). How diverse should be immune system be. *Proc. R. Soc. Lond. B. Biol. Sci.* **252,** 171–175.

EIGEN, M. & SCHUSTER, P. (1979). *The Hypercycle—a Principle of Natural Self-organization.* Berlin: Springer.

GOODNOW, C. C., CROSBIE, J., ADELSTEIN, S., LAVOIE, T. B., SMITH-GILL, S. J., BRINK, R. A., PRITCHARD-BRISCOE, H., WOTHERSPOON, J. S., LOBLAY, R. H., RAPHAEL, K. *et al.* (1988). Altered immunoglobulin expression and functional silencing of self-reactive B lymphocytes in transgenic mice. *Nature* **334,** 676–682.

HIGHTOWER, R. R., FORREST, S. & PERELSON, A. S. (1995). The evolution of emergent organization in immune system gene libraries. In: *Proc. of the Sixth Int. Conf. on Genetic Algorithms* (Eshelman, L. J., ed.), pp. 344–350. Morgan Kaufmann.

KAUFFMAN, S. & LEVIN, S. (1987). Towards a general theory of adaptive walks on rugged landscapes. *J. theor. Biol.* **128,** 11–45.

KAUFFMAN, S. A. & WEINBERGER, E. D. (1989). The NK model of rugged fitness landscapes and its application to maturation of the immune response. *J. theor. Biol.* **141,** 211–245.

MACKEN, C. A. & PERELSON, A. S. (1989). Protein evolution on rugged landscapes. *Proc. Natl. Acad. Sci. U.S.A.* **86,** 6191–6195.

MEDZHITOV, R., PRESTON-HURLBURT, P. & JANEWAY, C. A. (1997). A human homologous of the Drosophila toll protein signals activation of adaptive immunity. *Science* **388,** 394–397.

NEMAZEE, D. A. & BURKI, K. (1989). Clonal deletion of B lymphocytes in a transgenic mouse bearing anti-MHC class I antibody genes. *Nature* **337,** 562–566.

PAGIE, L. & HOGEWEG, P. (1998). Evolving adaptability due to coevolving targets (submitted).

PERCUS, J. K., PERCUS, O. E. & PERELSON, A. S. (1993). Predicting the size of the T-cell receptor and antibody combining region from consideration of efficient self-nonself discrimination. *Proc. Natl. Acad. Sci. U.S.A.* **90,** 1691–1695.

SCHATZ, D. G., OETTINGER, M. A. & SCHLISSEL, M. S. (1992). V(D)J recombination: molecular biology and regulation. *Annu. Rev. Immunol.* **10,** 359–383.

TAKUMI, K. & DE BOER, R. J. (1996). Self-assertion modeled as a network repertoire of multi-determinant antibodies. *J. theory Biol.* **183,** 55–66.