

Systems biology

Optimization of HAART with genetic algorithms and agent-based models of HIV infection

F. Castiglione^{1,*}, F. Pappalardo², M. Bernaschi¹ and S. Motta¹

¹Institute for Computing Applications 'M. Picone', Consiglio Nazionale delle Ricerche (CNR), V.le del Policlinico, 137, 00161 Rome and ²Department of Mathematics and Computer Science, University of Catania, V.le A. Doria 6-I, 95125 Catania Italy

Received on May 3, 2007; revised and accepted on August 7, 2007

Advance Access publication October 17, 2007

Associate Editor: Limsoon Wong

ABSTRACT

Motivation: Highly Active AntiRetroviral Therapies (HAART) can prolong life significantly to people infected by HIV since, although unable to eradicate the virus, they are quite effective in maintaining control of the infection. However, since HAART have several undesirable side effects, it is considered useful to suspend the therapy according to a suitable schedule of Structured Therapeutic Interruptions (STI).

In the present article we describe an application of genetic algorithms (GA) aimed at finding the optimal schedule for a HAART simulated with an agent-based model (ABM) of the immune system that reproduces the most significant features of the response of an organism to the HIV-1 infection.

Results: The genetic algorithm helps in finding an optimal therapeutic schedule that maximizes immune restoration, minimizes the viral count and, through appropriate interruptions of the therapy, minimizes the dose of drug administered to the simulated patient.

To validate the efficacy of the therapy that the genetic algorithm indicates as optimal, we ran simulations of opportunistic diseases and found that the selected therapy shows the best survival curve among the different simulated control groups.

Availability: A version of the C-IMM-SIM simulator is available at <http://www.iac.cnr.it/~filippo/c-immSim.html>

Contact: f.castiglione@iac.cnr.it

1 INTRODUCTION

Highly active antiretroviral therapies (HAART) are based on the combined use of extremely powerful antiretroviral agents in the treatment of HIV infections. A HAART is composed of multiple antiviral drugs, and it is commonly prescribed to most HIV-positive patients, even before they start to present any symptom of the disease (FDA, 1999). The therapy includes at least two components: a Protease Inhibitor (PI) and a reverse Transcription Inhibitor (TI). Nowadays, more complex cocktails of antiretroviral drugs include a nucleotide analog (DNA chain terminator), a second nucleotide analog (nucleoside) or a Non-Nucleotide Reverse Transcription Inhibitor (NNRTI).

Although most studies agree that a HAART is not able to eradicate the virus, it is quite effective in maintaining the infection under control. An indirect confirmation comes from the fact that the number of deaths caused by AIDS has dropped significantly in the countries where HAART has been extensively used (FDA, 1999).

However there are, at least, two issues related to HAART: (i) as most of effective drugs, it has some unpleasant side effects (e.g. *lipodystrophy*); (ii) drugs have to be administered over many years (actually 'forever'), which demands an enormous degree of discipline from patients.

To address, at least in part, these problems, several researchers and clinicians proposed to schedule interruptions of treatment and possibly to delay the start of the HAART (Structured Treatment Interruption, STI). Another reason to suspend periodically the therapy is to give the virus the chance to proliferate up to the point that the immune system recognizes it and begins its response. This in turn could reduce the chance of the emergence of resistant HIV strains.

To be more precise, the HAART therapy has a 2-fold effect on the viral/immune system dynamics. On the one hand, it causes a decline in the number of free HIV virions in patients. Plasma viremia declines in two phases: within 2 weeks, there is 99% of the decline. This phase is followed by a slower phenomenon, lasting for 8 weeks, after which HIV viral titers decline below the limit of detection of the assay. On the other hand, it determines an 'immune reconstitution' (or restoration), a sort of inverse process of the immune system damage caused by the HIV. This is one of the primary goals of HAART. Reconstitution consists of an increase in the number of functional CD4 T helper cells that are pivotal for the immune response against all pathogens (HIV included). The increase in CD4 cell count happens in two stages: a rapid increase in the number of circulating CD4 cell lymphocytes occurs within 1–2 weeks from the beginning of the HAART and continues over the first 2–3 months. A second, slower phase of CD4 cell expansion persists over subsequent months and may continue for years in HAART regimen (Autran *et al.*, 1997, 1999; Li *et al.*, 1998). In numbers, after 2–4 years of HAART, mean increase in CD4 cell counts is approximately 200–400 10^6 cells/l

*To whom correspondence should be addressed.

(Scott-Algara *et al.*, 2001; Mezzaroma *et al.*, 1999). The increase of CD4 cells occurs mainly in the first 2 years.

In the present work, we look for an optimal STI schedule of the therapy by using a simulator of the disease progression based on a detailed model of the HIV infection. Here, 'optimal' means a therapeutic schedule able to keep the infection under control even during the suspension period. At the same time, the optimal therapy should allow immune restoration, i.e. it should favor the rebound of the CD4 T helper lymphocytes' count.

There are a number of mathematical models aimed at describing the immune system response against HIV. Few of them also account for the use of a therapy (Nowak and May, 2000; Perelson and Nalson, 1999; Simson and Ho, 2003). Most of these models work in continuous-time where optimal control techniques can be applied in order to find the best time-allocation strategy for the therapy (Kwon, 2007). The techniques are similar to those used in cancer therapy (Swan, 1990).

2 APPROACH

In the present study, we adopt an approach that differs in two ways. First, the model is not described by differential equations in continuous-time but it is a generalization of a discrete-time stochastic cellular automaton belonging to the family of so-called agent-based models (ABM) (e.g. similar approaches can be found in Zorzenon dos Santos (1998); Pandey and Stauffer (1990); Pandey (1991); Zorzenon dos Santos and Coutinho (2001); Hershberg *et al.* (2001); and Ruskin *et al.* (2002). Second, since there is no analytical formulation of the model, the application of classic optimal control techniques is not a viable alternative and therefore, other, often heuristic, solutions must be pursued. One of these heuristic approaches are the genetic algorithms (GA) that are introduced below (Corne *et al.*, 1999; Mitchell, 1996). We have applied a GA-based optimization technique to find out the optimal schedule of HAART administration/interruption in order to achieve both HIV control and immune system recovering.

Since immune restoration should result in a stronger immune system, to verify the efficacy of the therapy found by optimization, we simulated the occurrence of opportunistic infections by introducing bacteria in the virtual patients at the end of the therapeutic period.

One of the most common opportunistic bacterial infections in immune deficient individuals is caused by the *Mycobacterium tuberculosis* (Mtb). Developing from infection to active tuberculosis is extremely rapid in HIV-infected individuals and, usually, it occurs within weeks or months from infection. The progression of the disease in individuals with immune deficiency caused by HIV depends on the extent of the CD4 cell count reduction and therefore can vary greatly. Based on these considerations, we set the replication rate of the 'virtual' bacterium and the limiting value so to have that the time from Mtb infection to active illness is 1–2 'months'.

For the comparison we used three other simulated control groups. In the first case, ('full therapy') the drugs were administered every day for the simulated therapeutic period; in the second case, ('random therapy') the total amount of drug

was equal to that of the optimal therapy but the administration within the therapeutic period was at random and, finally, in the third case there was no therapy ('void therapy').

3 METHODS

The model of immune system response we employ has been quite extensively described in Bernaschi and Castiglione (2001); and Castiglione *et al.* (2004). In short, it consists of a 3-dimension stochastic cellular automaton in which we represent the major classes of cells of the lymphoid lineage (T Helper lymphocytes or TH, Cytotoxic T lymphocytes or CT, B lymphocytes and antibody-producer plasma cells, PLB) and some of the myeloid lineage. All these entities interact each other following a set of 'rules' that describe the different phases of the recognition and response of the immune system against a pathogen. In order to represent the special features of the HIV infection, the model has been enhanced with the description of HIV replication inside infected lymphocytes, T production impairment, specific response against HIV strains, HIV mutation and evolution. The evolution of HIV in untreated patients has been studied and described in Castiglione and Bernaschi (2005).

For the study of the effects of a HAART on the dynamics of the HIV disease, we modified the model to account for the action of two classes of inhibitors:

- Reverse Transcriptase Inhibitors (RTIs), a class of antiretroviral drugs used to inhibit activity of reverse transcriptase, a viral DNA polymerase enzyme that HIV needs to reproduce. In detail, when the HIV infects a cell, reverse transcriptase copies the viral single-stranded RNA genome into a double-stranded viral DNA. The viral DNA is then integrated into the host chromosomal DNA which then allows host cellular processes, such as transcription and translation to reproduce the virus. RTIs block reverse transcriptase's enzymatic function and prevent completion of synthesis of the double-stranded viral DNA thus preventing HIV from reproducing. The first of these competitive reverse transcriptase inhibitors to be approved by the US Food and Drug Administration for treatment of HIV was Zidovudine (also called AZT).
- Protease inhibitors (PIs) are a class of medications used to treat or prevent infections by viruses (in general). PIs prevent viral replication by blocking the activity of protease, an enzyme used by viruses to cleave nascent proteins for final assembly of new virions. In the case of HIV/AIDS, examples of antiretroviral protease inhibitors are Saquinavir, Ritonavir, Indinavir, Nelfinavir, just to mention some.

In our model, the HAART modify the HIV life cycle. In untreated patients the life cycle of the virus can be described as follows: (i) when the virus infects a cell it enters the cytoplasm; (ii) then its RNA is transcribed into DNA of the host cell; (iii) the virus remains at rest until an event activates the host cell. In that case the virus replicates (it is said to *actively replicate*); (iv) a replicating virus also buds from cell membrane. Correctly assembled virions are able to infect other cells so that the *life cycle* starts again (see Fig. 1 for further details).

The application of a HAART composed by a transcriptase and a protease inhibitor is described in the model as follows: the transcriptase inhibitor does not allow a virus inside an infected cell to go from stage i to stage ii the protease inhibitor acts in stage iv because, by inhibiting the assembly of the virus, it produces virions that are at most defective and therefore unable to infect other cells.

In the model, the effect of interruptions of the therapy can be simulated in a very simple way under the assumption that the efficacy drops to zero during an interruption.

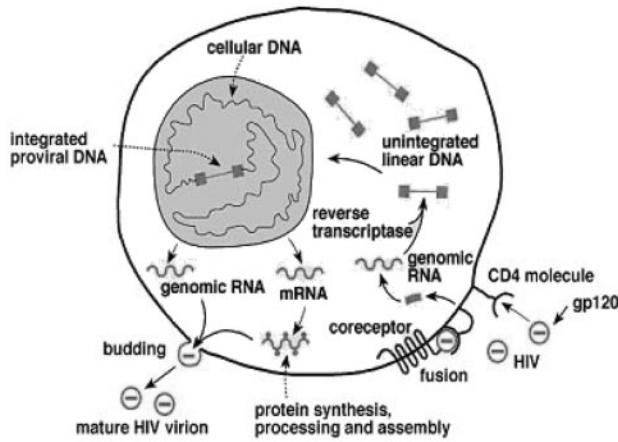


Fig. 1. HIV replication cycle from National Institute of Allergy and Infectious Disease (part of NIH) (<http://www.niaid.nih.org/>). The dynamics can be summarized as follows: (a) HIV attaches to CD4 cell-surface receptor molecules by its gp120 molecule; (b) the virus and cell membrane fuse, and the virion core (containing viral RNA and other proteins) enters the cell; (c) the core releases its content in the cell cytoplasm; (d) the viral RNA genome is converted into double-stranded DNA through an enzyme unique to viruses, reverse transcriptase; (e) The double-stranded viral (linear) DNA moves into the cell nucleus; (f) using the viral enzyme called integrase, the viral DNA is integrated into the cellular DNA; (g) Viral RNA is synthesized by the cellular enzyme RNA polymerase II using integrated viral DNA as a template. Two types of RNA transcripts mRNA and genomic RNA are produced; (h) mRNAs are transported to the cytoplasm and used for the production of several viral proteins that are then modified in the Golgi apparatus of the cell; Genomic RNA is transported to the cytoplasm as it is; (i) a new virion is assembled and then buds off.

The CD4 T-cell replenishment damage due to the HIV is not an ‘emergent property’ in the current version of our model. We represent this phenomenon from a functional point of view since there is no full understanding of the microscopic mechanisms leading to the progressive loss of peripheral CD4+ T cells [the mechanisms of CD4+ T-cell loss in HIV-1 infection have been attributed to viral-induced CD4+ T-cell death (Ho *et al.*, 1995; Wei *et al.*, 1995) and to progressive destruction of immune cell’s micro environment (Rosenberg *et al.*, 1998; Wolthers *et al.*, 1998)].

In our model we describe hematopoiesis (the process of cell replenishment by production of new cells in the bone marrow) as a mean reverting process at a baseline cell count: if cells are below baseline then cells are created; if the count is above (because of a clone expansion due to an immune response) then cells death is accelerated. In either case, the cell count relaxes to the baseline level of cells. We model the damage due to the HIV by decreasing this baseline level, whereas immune reconstitution is modeled by increasing it. The parameters of this process are set so as to reproduce, in absence of therapy, a decrease of CD4 cell count in line with clinical expectations of time to AIDS (Castiglione *et al.*, 2004; Munoz and Xu, 1996) whereas immune reconstitution due to a HAART is set to achieve an average CD4 T-cell increase of about 200 cells in a microliter in 6 months (Mezzaroma *et al.*, 1999; Scott-Algara *et al.*, 2001).

Evolutionary algorithms have been applied with satisfactory results to a very long list of hard combinatorial problems. A detailed description or enumeration of such results is beyond the scope of the present article but the interested reader can find useful reviews in Corne *et al.* (1999) and Mitchell (1996).

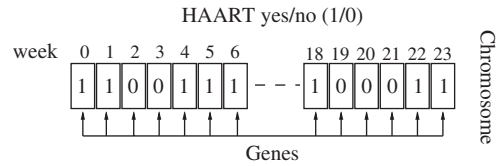


Fig. 2. Figure describing the HAART schedule as a chromosome of the genetic algorithm. A chromosome represents one possible drug administration schedule. It is a binary string of 24 bits where the j th bit when set equal to one means that during the j th week of the therapeutic period, the patient takes the antiretroviral drugs (both RTI and PI). If the chromosome j th bit is set equal to zero, then no HAART is applied.

Genetic algorithms, proposed by Holland in 1960, are a particular class of evolutionary algorithms that use techniques inspired by evolutionary biology such as inheritance, mutation, selection and crossover (also called recombination). A GA is a method for moving from one population of chromosomes to a new population by using a kind of natural selection together with the genetic inspired operators of crossover, mutation and so on. Each chromosome consists of genes (e.g. bits), each gene being an instance of a particular allele (e.g. 0 or 1). The selection operator chooses those chromosomes in the population that will be allowed to reproduce, and on average the fitter chromosomes produce more offspring than the less fit ones. A common application of GAs is global optimization, where the goal is to find a set of parameter values that maximizes a complex multi-parameter function.

The approach we present in this article differs from ‘classic’ GA applications since we use a *simulator* to evaluate the fitness function. To the best of our knowledge, very few applications (Lollini *et al.*, 2006; Pappalardo *et al.*, 2006) use a complete simulator in a genetic algorithm and, for all we know, no application of this type exists in HIV-related studies.

The therapy is applied to our virtual patient for a period of 6 months. Since we simulate the infection of the virtual patient at time zero of the simulation, we start HAART after (about) 7.5 years of simulated life, i.e. during the chronic phase of the disease progress.

The HAART consists in administering daily a cocktail of drugs. Therefore, we could optimize the therapy on a daily basis but we quickly discarded this option since it is practically impossible for a human being to follow, on daily basis, a schedule that is 6-month long. Hence, in more practical terms, we defined our schedule on a weekly basis. Another important choice, this time dictated by medical practice, is that reverse transcriptase and protease inhibitors are always administered at the same time.

In GA terms we defined a chromosome to represent a therapy schedule. A chromosome is a binary string of 24 bits (i.e. weeks in 6 months). The j th bit ($j=0, \dots, 23$) indicates that during the j th week, the simulated patient takes the antiretroviral drugs (Fig. 2).

There are many possible selection schemes for genetic algorithms, each one with different characteristics. An ideal selection scheme is simple to code, and efficient for both serial and parallel computing platforms. Furthermore, a selection scheme should be able to adjust its selection pressure so as to tune its performance for different domains. *Tournament selection* is increasingly being used as a GA selection scheme because it satisfies all the above-mentioned criteria. Significant progress was made some years ago in understanding the convergence rates of various selection schemes, including tournament selection (Goldberg, 1991). All of these factors contributed to our choice to implement this operator in our GA. Reproduction uses *uniform crossover*. Given two chromosomes selected by the selection mechanism, they will produce two offsprings. Genes in offspring are selected

from parents with the same probability. Mutation acts as follows: first, a gene subject to mutation is chosen with probability $p = 1/g_n$, where g_n represents the number of chromosome's genes; second, the selected gene is set equal to 1 with probability $p = 1/p_s$ and to 0 with probability $p = 1 - (1/p_s)$, where p_s represents the population size. Elitism is used on two specific elements of the population: (i) best fitness element; (ii) second best fitness element.

As already mentioned, to determine the quality (i.e. the fitness) of each chromosome of the population we used the C-IMMSIM simulator (Castiglione and Bernaschi, 2005; Castiglione *et al.*, 2004) that takes a non-negligible amount of time to run. That is why we developed a parallel implementation of the GA.

The parallel implementation works on any cluster of systems. The only requirement is the availability of a shared filesystem in order to exchange data among different computational nodes.

The implementation adopts a quite simple master/slave scheme. The master node reads the parameters in input, generates the initial random population of chromosomes (i.e. the different therapies) and dispatches to slaves a subset of all therapies constituting the chromosome population. Then, master and slaves set up all files required for the simulation and run. At the end of the simulation of all assigned therapies, the master node computes the fitness of each therapy by looking at the dynamics of HIV and CD4 cell count generated by the simulations. Then, it applies typical operations of genetic algorithms (i.e. selection, crossover, mutation and elitism) on the original population and dispatches to slave nodes the newly generated population. This process is repeated for a predefined number of generations.

The workload of the master for the computation of the fitness function is negligible compared to running a simulation. Moreover, since the communication between master and slaves is reduced to the dispatch of the chromosome population, and also because the simulations are completely independent each other, this parallel implementation scales very well with the number of processors.

One of the crucial elements of a successful genetic search is a careful definition of the fitness function that measures how good a solution is for the optimization problem under study. In the present case, each potential solution (represented by a possible therapy) receives a score based on the results of a simulation of the dynamics of HIV and CD4 cell count during the therapeutic period made by means of the C-IMMSIM simulator. It is reasonable to assume that a therapy resulting in (i) a lower level of HIV (both free and viral), (ii) a higher CD4 count and (iii) a reduced number of pills to be administered, should be given a high score (fitness).

Therefore, we compute the fitness F_i^g of the therapy i at generation g as the sum of three terms:

- HIV load, computed by summing the free and proviral load divided by viral set point in the beginning of therapy;
- CD4 fitness, given by the average of the ratio between the CD4 in the beginning of the simulation (i.e. before the infection, which corresponds to the normal CD4 count) and the actual CD4 count during the therapeutic period;
- The number of active therapy days divided by the total number of days in the therapeutic period.

Hereafter t_s indicates the time the therapy starts; t_e the time the therapy ends; for each therapy i at generation g , $H_i^g(t)$ is the sum of free virions and proviral HIV in infected cells; $D_i^g(t)$ the CD4 T helper cell count in the simulated space and finally $z_i^g(t)$ is a function such that $z_i^g(t) = 1$ if the therapy is active at time t , zero otherwise.

The fitness of the i th individual (therapy) at generation g can be written as follows:

$$F_i^g = H_i^g + D_i^g + z_i^g$$

where

$$H_i^g = \frac{1}{(t_e - t_s)} \sum_{t=t_s}^{t_e} H_i^g(t),$$

$$D_i^g = \frac{1}{(t_e - t_s)} \sum_{t=t_s}^{t_e} D_i^g(t),$$

$$z_i^g = \frac{1}{t_e - t_s} \sum_{t=t_s}^{t_e} z_i^g(t).$$

The optimization consists in finding the best individual j (therapy) corresponding to the minimal value of the affinity function, i.e. $j = \min_j F_j^g$.

4 DISCUSSION

As already mentioned, we set $t_s = 7.5$ years from infection and $t_e = t_s + 6$ months, i.e. the therapeutic horizon is 6 months. Simulated space is 5 cubic millimeters of a lymph node.

The genetic algorithm population is composed by individuals with identical initial settings that, as a consequence, show a similar (but not identical, due to the stochastic components of the model) dynamics. Individual variations account for a different repertoire and a different major Histocompatibility complex (MHC) that determine a different wild-type immunogenicity. In the GA optimization, a population of 32 individuals is used corresponding to 32 different therapeutic schedules. For each therapy, a set of 16 virtual patients was used to account for differences among patients (this avoids that the optimal therapy depends too much on the specific features of a single patient). Therefore, the population size is $32 \times 16 = 512$.

The GA ran on 60 generations. Although population size and number of generations could appear limited, it is important to note that the number of simulations required to obtain the optimal therapy is equal to $32 \times 16 \times 60 = 30720$. Since each simulation requires on, average, 30 mins on a high-end PC, we resorted to a cluster of 128 PC to carry out this GA optimization. Note that, by employing uniform crossover and tournament selection in the genetic search we compensated the limited population size that we were forced to adopt for computational reasons.

The optimal therapy found after 60 generations on a population of $N = 32 \times 16$ individuals, is displayed in Figure 3.

Figure 4 shows the average fitness $\langle F^g \rangle = \frac{1}{N} \sum_i F_i^g$ computed on all individuals at each generation g . Note that the average fitness is not strictly monotone decreasing since this is not the fitness of best adapted individual but the average among all individuals. However, it shows clearly a plateau indicating a lack of (or very small) improvement for generations greater than ~ 50 .

Among the three components of F_i^g , it is z_i^g that results to be the most decisive in determining the efficacy of the therapy. This is in line with the evaluation of the best therapy (described below), which shows that the optimal therapy is able to achieve a performance comparable with the 'full' therapy but with reduced drug dosage.

To show that the 'optimal' therapy resulting from the GA-based optimization really achieves immune reconstitution and it does so with 'minimal' drug administration, we performed four sets of experiments in which, after the therapy,

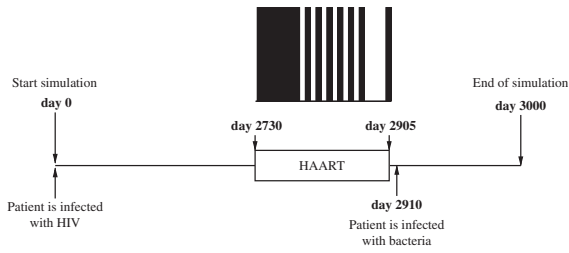


Fig. 3. Simulated patients are infected at time zero and start therapy at day 2730 after infection (7.5 years). The best therapy found by genetic optimization is shown on top. After therapy has terminated, an opportunistic infection challenges the simulated patients. The immune response to the antigen is monitored until the end of the simulation. If the antigen count reaches the limiting value of 40×10^6 copies/microliter then the virtual patient is declared death and the simulation stops. Efficient therapies achieve an immune restoration sufficient for containing the antigen growth.

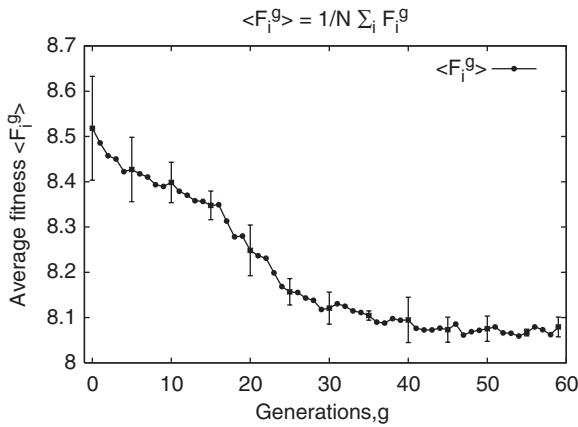


Fig. 4. Average fitness function during the GA optimization. The average fitness is computed as $\frac{1}{N} \sum_i F_i^g$ where $N=32 \times 16$ is the population size. The plot shows the standard deviation of the fitness as error-bar for some significant points. Note its decrease in correspondence of the plateau for $g > 50$.

the simulated patient is challenged by four injections of bacteria. In normal conditions (i.e. for a healthy organism) the bacteria are eliminated whereas, under immunodeficient conditions, the bacteria grow undisturbed due to the lack of a strong immune response. By counting, on a set of 200 runs, the number of survivors, we computed the survival curves for each of the following four groups: for the first group of simulated patients the optimal therapy is applied; for the second group a ‘full’ therapy is applied, in which drugs are administered every day during the simulated therapeutic period; for the third group a ‘random’ therapy is applied in which the total number of therapy days is equal to that of the optimal therapy but the distribution is at random within the therapeutic period and, finally, in the fourth group there is no therapy (‘void therapy’), i.e. no drug is administered to the virtual patient. Both full- and void-therapy need to be considered as possible control groups since the efficacy of any therapy must, obviously, remain within the values of efficacy shown by these two extreme protocols.

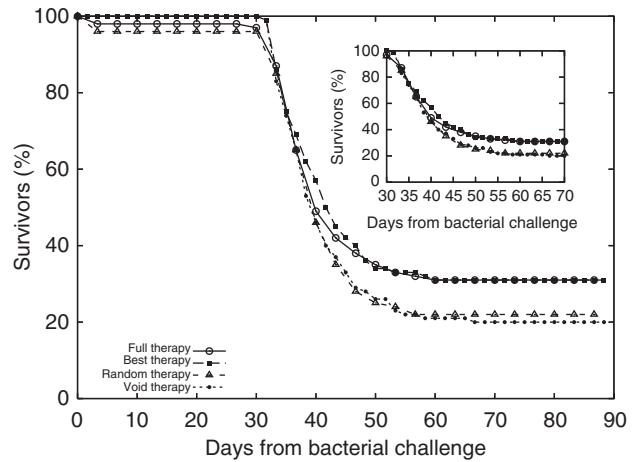


Fig. 5. Survival curve for four different protocols. Full therapy is the therapy applied each day. Optimal therapy is the therapy computed by the genetic algorithm. Random therapy is a therapy with a total number of therapy days equal to that of the optimal therapy but a random distribution within the therapeutic period. Void therapy is the case where no therapy is applied. Survival curves show a better performance of the full and the optimal therapies with respect to the random and the void therapies. Note how the optimal therapy shows an efficacy close to the full therapy whereas the random therapy an efficacy close to the case of absence of therapy. The inset plot shows a restricted range of the same plot.

The resulting survival curves are shown in Figure 5. As expected, the survival curve shows better performances for the full (34.11% survivals) and optimal (30.5%) therapies with respect to the random (21.86%) and void (20.25%) therapies. Moreover, the optimal therapy shows a performance close to the full therapy, that is close to the upper bound of efficacy, whereas the random therapy a performance close to the lower bound corresponding to the absence of therapy. Note also that in any case there is 100% of survival because complete immune restoration is never achieved, also in the full therapy case. On the other hand, the fact that the void therapy is able to guarantee (about) a 20% of survival can be explained by the fact that a weak immune system (low CD4 cell count) is not an *absent* immune system and therefore some survivors are possible due to the existence of minimal residual defense.

Finally, it is worth to mention that the optimal therapy uses $\sim 40\%$ less medicine than the full therapy although it achieves comparable survival results. This is particularly significant since we started from the reasonable assumptions that the full HAART has the best therapeutic effects but also the largest (undesirable) side effects (whereas, in absence of therapy there are no side effects at all but, obviously, not even therapeutic effects), and looked for an optimal interruptions schedule that could reduce undesirable side effects. The optimal schedule found by the genetic algorithm does not avoid completely the side effects of the therapy but reduces them maintaining at the same time a level of efficacy very close to the full therapy as shown by the survival curves in figure Figure 5. These results are summarized in Table 1.

Table 1. Comparison among the four different therapy schedules

Therapy	Number therapy weeks	Survival rate (%)
Full	25	34.11
Best	15	30.50
Random	15	21.86
Void	0	20.25

The optimal therapy shows a performance close to the full therapy, although it is administered for fewer weeks (i.e. 15 versus 25). With the same number of therapeutic weeks, the random therapy shows a performance close to the lower bound corresponding to the void therapy.

5 CONCLUSION

By means of genetic algorithms we find an optimal schedule of HAART based on structured interruptions. The effects of the therapy have been simulated using an agent-based model of the immune system that reproduces the most significant features of the response of an organism to the HIV-1 infection.

The resulting optimal schedule has been validated by running independent simulations with control groups. In particular, the optimal schedule achieves a performance comparable to a full-time therapy with a significant reduction of drug dosage.

Although these tests cannot be considered a validation protocol, they indicate that the underlying model is coherent and can be a viable alternative to animal models in the preliminary phase of designing *in vivo* experiments.

ACKNOWLEDGEMENTS

This work was supported under the EC contract FP6-2004-IST-4, No.028069 (ImmunoGrid). We thank the CINECA Super Computing Center for providing us with required computing time.

Conflict of Interest: none declared.

REFERENCES

Autran, B. *et al.* (1997) Positive effects of combined antiretroviral therapy on CD4+T cell homeostasis and function in advanced HIV disease. *Science*, **277**, 112–116.

Autran, B. *et al.* (1999) Restoration of the immune system with anti-retroviral therapy. *Immunol. Lett.*, **66**, 207–211.

Bernaschi, M. and Castiglione, F. (2001) Design and implementation of an immune system simulator. *Comput. Biol. Med.*, **31**, 303–331.

Castiglione, F. and Bernaschi, M. (2005) HIV-1 strategies of immune evasion. *Int. J. Mod. Phy. C*, **16**, 1869–1879.

Castiglione, F. *et al.* (2004) Mutation, fitness, viral diversity and predictive markers of disease progression in a computational model of HIV-1 infection. *AIDS Res. Hum. Retrovirus*, **20**, 1316–1325.

Corne, D. *et al.* (1999) New Ideas in Optimization. *Advanced Topics in Computer Science*, McGraw-Hill, UK.

FDA Consumer magazine July-August 1999. Vol. 33. http://www.fda.gov/FDAC/features/1999/499_aids.html

Goldberg, D.E. (1991) A comparative analysis of selection schemes used in genetic algorithms. In Rawlins, G. (ed.) *Foundations of Genetic Algorithms*. Morgan Kaufmann Publishers, San Mateo, CA, pp. 69–93.

Hershberg, U., Louzoun, Y., Atlan, H. and Solomon, S. (2001) HIV time hierarchy: Winning the war while losing all the battles. *Physica A*, **289**, 178–190.

Ho, D.D. *et al.* (1995) Rapid turnover of plasma virions and CD4 lymphocytes in HIV-1 infection. *Nature*, **373**, 123–126.

Kwon, H.-D. (2007) Optimal treatment strategies derived from a HIV model with drug-resistant mutants. *Appl. Math. Comput.*, **188**, 1193–1204.

Li, T.S. *et al.* (1998) Long-lasting recovery in CD4 T-cell function and viral-load reduction after highly active antiretroviral therapy in advanced HIV-1 disease. *Lancet*, **351**, 1682–1686.

Lollini, P.-L. *et al.* (2006) Discovery of cancer vaccination protocols with a genetic algorithm driving an agent based simulator. *BMC Bioinformatics*, **7**, 352. doi:10.1186/1471-2105-7-352.

Mezzaroma, I. *et al.* (1999) Long-term evaluation of T-cell subsets and T-cell function after HAART in advanced stage HIV-1 disease. *AIDS*, **13**, 1187–1193.

Mitchell, M. (1996) *An Introduction to Genetic Algorithm*. The MIT Press, USA.

Muñoz, A. and Xu, J. (1996) Models for the incubation of AIDS and variations according to age and period. *Stat. Med.*, **15**, 2459–2473.

Nowak, M. and May, R.M. (2000) *Virus Dynamics: Mathematical Principles of Immunology and Virology*. Oxford University Press, USA.

Pandey, R.B. (1991) Cellular automata approach to interacting cellular network models for the dynamics of cell population in an early HIV infection. *Physica A*, **179**, 442–470.

Pandey, R.B. and Stauffer, D. (1990) Metastability with probabilistic cellular automata in an HIV infection. *J. Stat. Phys.*, **61**, 235–240.

Pappalardo, F. *et al.* (2006) Genetic Algorithm against Cancer. *Lect. Notes Comput. Sci.*, **3849**, 223–228.

Perelson, A.S. and Nelson, P.W. (1999) Mathematical analysis of HIV-1 dynamics in vivo. *SIAM Rev.*, **41**, 3–44.

Rosenberg, Y.J. *et al.* (1998) HIV-induced decline in blood CD4/CD8 ratios: viral killing or altered lymphocyte trafficking? *Immunol. Today*, **19**, 10–17.

Ruskin, H.J., Pandey, R.B. and Liu, Y. (2002) Viral load and stochastic mutation in a Monte Carlo simulation of HIV. *Physica A*, **311**, 212–220.

Scott-Algara, D. *et al.* (2001) CD4 T cell recovery is slower in patients experiencing viral load rebounds during HAART. *Clin. Exp. Immunol.*, **126**, 295–303.

Simon, V. and Ho, D.D. (2003) HIV-1 dynamics in vivo: implications for therapy. *Nat. Rev. Microbiol.*, **1**, 181–190.

Swan, G.W. (1990) Role of optimal control theory in cancer chemotherapy. *Math. Biosci.*, **101**, 237–284.

Wei, X. *et al.* (1995) Viral dynamics in human immunodeficiency virus type I infection. *Nature*, **373**, 117–122.

Wolthers, K.C., Shuitemaker, H. and Miedema, F. (1998) Rapid CD4+T-cell turnover in HIV-1 infection: a paradigm revisited. *Immunol. Today*, **19**, 44–48.

Zorzenon dos Santos, R.M. (1998) Immune responses: getting close to experimental results with cellular automata models. In Stauffer, D. (ed.) *Annual Reviews on Computational Physics*. Vol. IV. World Scientific, Singapore, p. 159.

Zorzenon dos Santos, R.M. and Coutinho, S. (2001) Dynamics of HIV infection: a cellular automata approach. *Phys. Rev. Lett.*, **87**, 168102.