

Systems biology

Optimal vaccination schedules using simulated annealing

Marzio Pennisi¹, Roberto Catanuto¹, Francesco Pappalardo^{1,2,*} and Santo Motta^{1,2}¹Department of Mathematics and Computer Science, University of Catania and ²Faculty of Pharmacy, University of Catania, Catania, Italy

Received on March 31, 2008; revised and accepted on June 3, 2008

Advance Access publication June 5, 2008

Associate Editor: Limsoon Wong

ABSTRACT

Summary: Since few years the problem of finding optimal solutions for drug or vaccine protocols have been tackled using system biology modeling. These approaches are usually computationally expensive. Our previous experiences in optimizing vaccine or drug protocols using genetic algorithms required the use of a high performance computing infrastructure for a couple of days. In the present article we show that by an appropriate use of a different optimization algorithm, the simulated annealing, we have been able to downsize the computational effort by a factor 10^2 . The new algorithm requires computational effort that can be achieved by current generation personal computers.

Availability: Software and additional data can be found at <http://www.immunomics.eu/SA/>

Contact: francesco@dmi.unict.it

1 INTRODUCTION

Drugs and vaccines design and development is an inherently laborious process. The availability of animal models and in particular of humanized animal models has contributed greatly to their experimentation. The refinement of a vaccine protocol or the trials to improve a drug efficacy could lead to an endless series of long-term wet-lab experiments. In such a framework, computational and mathematical models in system biology could represent a valuable help to biologists and medical researchers. Some interesting applications of system biology models and optimization techniques can be found in recent literature (Davies and Flower, 2007). Mathematical continuous models have been proposed to describe cancer therapy (Kirschner and Panetta, 1998; Nani and Freedman, 2000; Vainstein *et al.*, 2006), to optimize immunotherapy (Castiglione and Piccoli, 2007) and chemotherapy (Agur *et al.*, 2006). Agent-based models (ABM) like SimTriplex (Pappalardo *et al.*, 2005) and C-ImmSim (Bernaschi and Castiglione, 2001) have been proposed for reproducing important pathologies like mammary carcinoma and AIDS. Drug and vaccine scheduling optimization using ABM has been also performed (Castiglione *et al.*, 2007; Lollini *et al.*, 2006a). We only note here that, from the immunological point of view, the major difference between continuous and discrete models is that the latter ones take into consideration the recognition phase of the immune response while the former only consider the effector cells. Continuous models

are appropriate to describe the general behavior of the system while discrete models provide a detailed description of the immune processes. An extensive survey of ‘Tumor—Immune System’ competition can be found in Adam and Bellomo (1997) and Preziosi (2003).

We briefly recall here the vaccine schedule optimization problem to introduce its major computational aspect. A vaccination schedule is usually designed empirically using a combination of immunological knowledge, vaccinological experience from previous endeavors and practical constraints. In subsequent trials, the schedule of vaccinations is then renewed on the basis of the protection elicited in the first batch of subjects and their immunological responses e.g. kinetics of antibody titers, cell mediated response, etc. The problem of defining optimal schedules is particularly important in cancer immunopreventive approaches, which requires a sequence of vaccine administrations to keep a high level of protective immunity against the constant generation of cancer cells over very long periods, ideally for the entire lifetime of the host.

The Triplex vaccine represents a clear example of such immunopreventive approaches. It has been designed (Lollini *et al.*, 2006b) to improve the efficacy of existing immunopreventive treatments against mammary carcinoma. Triplex showed to be effective in preventing the *carcinoma in situ* (CIS) formation in HER-2/neu mice using a *Chronic schedule* in a follow-up time between 52 and 57 weeks.

However it is not known if Chronic schedule really is the minimal set of vaccinations able of assuring complete, long-term protection against mammary carcinoma. Shorter heuristic protocols failed, in *in vivo* experiments, in fulfilling this requirement, but between the Chronic and the shorter schedules there is still a huge number of possibilities which remain yet unexplored.

SimTriplex is an ABM that describes all the relevant processes of the ‘immune system—mammary carcinoma’ competition by means of rules derived from biological experiences. This model and simulator includes all the most relevant entities (immune system cells, cancer cells, vaccine cells) and processes (recognition, differentiation, duplication) needed to reproduce the immune response induced by the Triplex vaccine. A detailed description can be found in (Pappalardo *et al.*, 2005). A validated simulator will reasonably reproduce, in the validation range, the immune response activated by a vaccination protocol, thus one can reproduce *in silico* different vaccination schedules and search for the schedules with a minimal number of vaccine administrations still capable to prevent CIS formation.

*To whom correspondence should be addressed.

Combining SimTriplex and genetic algorithms (GA) approach, an optimal vaccination schedule capable to guarantee *in silico survival* has been found (Lollini *et al.*, 2006a). Optimal search strategy was biologically guided. Considering that Chronic proved to be effective for tumor control, the optimal search tried to find a protocol with minimum number of vaccine administrations able to reproduce, *in silico*, the time evolution of the chronic schedule. This strategy was used in the GA optimal search. It is well known that GA are slowly converging algorithms; the GA optimal search required about 3 days using 32 nodes of an high performance computing infrastructure.

The application of such techniques in a clinical environment will require to downsize the requested computational resources and computational time. We note here that model complexity increases with biological knowledge and physical scale. Thus, even if computer power will continue to grow following Moore's law, complexity of models will grow too, rendering useless the availability of the added power. Therefore better performing approaches for the *optimal protocol finder* are always desirable. To this end we decided to investigate the applicability of **simulated annealing** (SA), a global optimization algorithm, widely tested and known for its computational speed and ability to achieve optimal solutions. (Interested readers can find an extended description of SA algorithm in van Laarhoven and Aaris, 1997). In the present article we show how the combination of the SA algorithm with biologically driven heuristic strategies, leads to a much faster algorithm and better results for the optimal vaccination schedule problem for Triplex vaccine.

2 IMPLEMENTATION OF THE SA ALGORITHM

The translation of the biological concept of *vaccine effectiveness* in mathematical-computational terms has been shown in Lollini *et al.* (2006a). Here we remark that *in silico* experiments, like the *in vivo* ones, consider a randomly chosen sample to represent a larger population. In our *in silico* experiment, we select a population of 200 virtual mice and a simple random sample of $k=8$ mice. The SA is a computational method which reproduces the slow cooling process of a solid. Physically this leads to a sequence of semi-equilibrium configurations toward the equilibrium configuration of minimum energy and temperature. The use of this optimization method for the optimal protocol problem requires: (i) the definition of the SA relevant concepts in terms of vaccine protocol, (ii) a description of the optimal protocol problem in terms of a cooling process.

In this context the relevant concepts in SA are: the initial solid configuration (and its subsequent perturbations), the temperature, the energy and the semi-equilibrium condition.

The relevant concepts of a vaccine administration protocol are: the number of injections, the mean survival age of the selected population of mice and the time distribution of injections of a given protocol. The optimal protocol is the one which provides the maximum *survival* of the sample with the minimum number of vaccine administrations. Since we want to model an *in vivo* experiment which runs for 57 weeks performing a maximum of two vaccine administrations per week, we describe a candidate protocol with a bit-string, i.e. a binary vector \mathbf{P} of cardinality $V=114$, where the bits position represents the administration time t and the bit value 1/0 represent vaccine, or no-vaccine administration

at time t . The total number of vaccine administrations n and \mathcal{M} , the total number of possible schedules with n vaccine administrations, $\mathbf{P}_n^l, l=1, \dots, \mathcal{M}$ are, respectively, given by: $n = \sum_{j=1}^V P_n^l(j)$ and $\mathcal{M} = V!/[n!(V-n)!]$.

The initial configuration. The initial distribution is an equilibrium configuration defined by \mathbf{P}_{in}, n_{in} and the initial energy E_{in} as defined later on. To obtain an initial equilibrium configuration we use the Metropolis algorithm (Metropolis *et al.*, 1953).

Temperature. The temperature of a semi-equilibrium represents, in the SA, the control parameter of the algorithm. The temperature is slowly but constantly lowered to reach the minimum of energy. In our case the most suitable entity to represent this feature is n , the *number of vaccine administration* of a semi-equilibrium configuration.

Energy. The configuration energy of SA is slowly stochastically decreasing. At a given temperature, a semi-equilibrium configuration is reached when the energy is minimal. To identify a suitable energy definition let us first consider the simple case of optimizing the schedule for a single mouse. In this case the semi-equilibrium will be reached when we find the maximum survival time τ of the mouse with the given number of vaccine administrations, i.e. $E \propto \tau$. However this would lead to finding a maximum (i.e. maximum energy) while SA is designed for finding a minimum. Taking into account that energy is always a positive quantity, we can define the energy $E \propto 1/\tau$.

The natural extension of this definition to the case in which we consider k mice is the arithmetic mean of the inverse of survival times, i.e. $E \propto 1/k \cdot \sum 1/\tau_i$. This simply represents the inverse of the *harmonic mean* H of the survival times τ_i .

The SA algorithm for protocol optimization. (i) starts from a randomly chosen initial vaccine distribution and find the initial semi-equilibrium configuration $n_{in}, \mathbf{P}_{n_{in}}^{l_{in}}, E_{n_{in}}^{l_{in}}$ using the Metropolis algorithm. The Metropolis algorithm finds a semi-equilibrium configuration building a finite sequence $\mathbf{P}_n^l, l=1, \dots, < \lambda$ (where λ is a predefined maximum number of iterations) with bits perturbed randomly and \mathbf{P}_n^{l+1} accepted stochastically according to the energy variation and temperature. (van Laarhoven and Aaris, 1997); (ii) Decrease the number of injections of 1 unit; (iii) find a semi-equilibrium configuration P_i, E_i according to Metropolis algorithm; (iv) cycle on (ii). The algorithm stops when, once the algorithm control parameter, i.e. the number of vaccine administrations, is decreased from n to $n-1$, the Metropolis algorithm is not able to find a semi-equilibrium configuration, i.e. an acceptable value of survivals, in λ iterations. The accepted protocol is the last found at temperature n .

Biologically inspired modification of Metropolis algorithm. Taking into account the energy definition, the Metropolis random bits perturbation can be heuristically improved using biological knowledge. As we want to optimize mice survival, in moving from \mathbf{P}_i^j to \mathbf{P}_i^{j+1} , we re-adjust random bits reallocation moving some '1' at a suitable time $t < \min\{\tau_i\}, i=1, k$.

Safety thresholds. To implement the optimal search strategy described in (1) we introduced, for each mouse i of the sample set, two safety thresholds on the number of cancer cells (CC). These safety thresholds bound to the *in silico* observed number of CC when a Chronic protocol is administered. If, at any time t , the number of CC is greater than any of the safety thresholds the survival time of that mouse i is $\tau_i = t$.

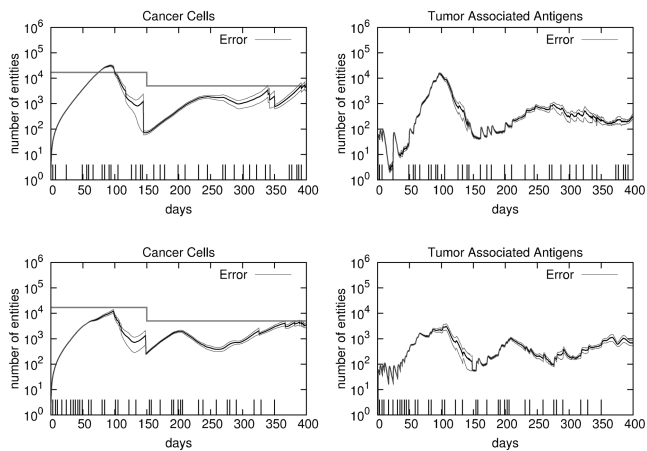


Fig. 1. Cancer cells behaviors and thresholds in GA (top) and SA (bottom). Small vertical bars on the x -axis represent vaccine administration times. Broken-line graphs on the lhs represent safety thresholds.

3 RESULTS AND CONCLUSIONS

Results In previous experiments the GA optimization technique (Lollini et al., 2006a) used a sample of eight randomly selected virtually mice to compute its fitness function. This was needed for obtaining a vaccination protocol effective on a large population of individuals. The protocol was then tested on the entire population set (200 virtual mice) and the tumor free percentage of individuals was of 87%. In order to compare the SA results against GA results, we used the same 8-mice subset and the same mice population. The *in silico* tumor free percentages of the mice population show no substantial difference between the two methods (87% for GA versus 86.5% for SA). Moreover, Figure 1 shows the mean number of CC, computed on the 200-mice set, for the GA protocol (up lhs) and the SA-protocol (down lhs). Only the SA protocol totally fulfills the safety threshold conditions. The SA algorithm required a computational effort of about 2 h on a 8 processor unit to find a protocol with 37 vaccine administrations. Comparison with previous results obtained with the GA algorithm show a speedup factor of $\sim 1.4 \times 10^2$.

Conclusions We presented a new approach to search optimal vaccination schedules. The use of the SA algorithm reaches an optimal solution with a computational effort that can be achieved on computers available in the consumer market. We note here that the vaccine administration distribution is not the same in SA-protocol and GA protocol. This suggests that many protocols P_n which reach the same global survival goal may exist. Searching between those

protocols the one which better fulfills other clinical requirements implies a clinically guided optimal search. Work in this direction is in progress and results will be published in due course.

ACKNOWLEDGEMENTS

This work makes use of results produced by the PI2S2 Project managed by the Consorzio COMETA, a project co-funded by the Italian Ministry of University and Research (MIUR) within the Piano Operativo Nazionale ‘Ricerca Scientifica, Sviluppo Tecnologico, Alta Formazione’ (PON 2000-2006). More information is available at: <http://www.pi2s2.it> and <http://www.consorzio-cometa.it>. The authors would like to thank Prof. Salvatore Ingrassia for helpful discussion on SA technique.

Funding: This work was supported under the EC contract FP6-2004-IST- 4, No.028069 (ImmunoGrid).

Conflict of Interest: none declared.

REFERENCES

- Adam, J.A. and Bellomo, N. (eds) (1997) *A Survey of Models for Tumor-Immune System Dynamics*. Birkhauser, Boston.
- Agur, Z. et al. (2006) Optimizing chemotherapy scheduling using local search heuristics. *Oper. Res.*, **54**, 829–846.
- Bernaschi, M. and Castiglione, F. (2001) Design and implementation of an immune system simulator. *Comput. Biol. Med.*, **31**, 303–331.
- Castiglione, F. and Piccoli, B. (2007) Cancer immunotherapy, mathematical modeling and optimal control. *J. Theor. Biol.*, **247**, 723–732.
- Castiglione, F. et al. (2007) Optimization of HAART by means of genetic algorithms and agent based models of HIV infection. *Bioinformatics*, **23**, 3350–3355.
- Davies, M.N. and Flower, D.R. (2007) Harnessing bioinformatics to discover new vaccines. *Drug Discov. Today*, **12**, 389–395.
- Kirschner, D. and Panetta, J.C. (1998) Modeling immunotherapy of the tumor-immune interaction. *J. Math. Biol.*, **37**, 235–252.
- Lollini, P.-L. et al. (2006a) Discovery of cancer vaccination protocols with a genetic algorithm driving an agent based simulator. *BMC Bioinformatics*, **7**, doi:10.1186/1471-2105-7-352.
- Lollini, P.-L. et al. (2006b) Vaccines for tumour prevention. *Nat. Rev. Cancer*, **6**, 204–216.
- Metropolis, N. et al. (1953) Equations of state calculations by fast computing machines. *J. Chem. Phys.*, **21**, 1087–1092.
- Nani, F. and Freedman, S. (2000) A mathematical model of cancer treatment by immunotherapy. *Math. Biosci.*, **163**, 159–199.
- Pappalardo, F. et al. (2005) Modeling and simulation of cancer immunoprevention vaccine. *Bioinformatics*, **21**, 2891–2987.
- Preziosi, L. (ed.) (2003) *Cancer Modelling and Simulation*. Chapman & Hall/CRC, Boca Raton, Florida, USA.
- Vainstein, V. et al. (2006) Improving cancer therapy by doxorubicin and granulocyte colony-stimulating factors: insights from a mathematical model of human granulopoiesis. *Math. Model. Nat. Phenom.*, **1**, 2.
- van Laarhoven, P.J.M. and Aarts, E.H.L. (1997) *Simulated Annealing: Theory and Applications*. Reidel Publishing Company, Dordrecht, Holland.