GRID Computing and Computational Immunology

Ferdinando Chiacchio and Francesco Pappalardo
University of Catania
Italy

1. Introduction

Biological function emerges from the interaction of processes acting across a range of spatio-temporal scales. Therefore understanding disease and developing potential therapeutic strategies requires studies that bridge across multiple levels. This requires a systems biology approach and the tools used must be based on effective mathematical algorithms and built by combining experimental and theoretical approaches, addressing concrete problems and clearly defined questions.

Technological revolutions in both biotechnology and information technology have produced enormous amounts of data and are accelerating the extension of our knowledge of biological systems. These advances are changing the way biomedical research, development and applications are done. Mathematical and computational models are increasingly used to help interpret data produced by high-throughput genomics and proteomics projects, and through advances in instrumentation. Advanced applications of computer models that enable the simulation of biological processes are used to generate hypotheses and plan experiments.

Computational modeling of immune processes has emerged as a major support area for immunology and vaccinology research. Computational models have been developed for the simulation of immune processes at the molecular, cellular, and system levels. Computer models are used to complement or replace actual testing or experimentation. They are commonly used in situations where experimentation is expensive, dangerous, or impossible to perform. Models of the immune system fall into two categories:

1. molecular interactions (such as peptide binding to receptors), and

2. system-level interactions (such as models of immune response, or cellular models of immune system).

Until recently, models of molecular level immune processes have been successful in supporting immunology research, such as antigen processing and presentation. Computational models of the complete immune system have mainly been developed in the domain of theoretical immunology and were used to offer possible explanations of the overall function of the immune system, but were usually not applied in practice.

Newer computational tools that focus on immune interactions include numerous methods used for the mapping of T-cell epitopes (targets of immune responses). Computational methods for the simulation of antigen processing include prediction of proteasomal cleavage and peptide binding to transporters associated with antigen processing (TAP). The basic methods for the prediction of antigen processing and presentation have been extended to more sophisticated computational models for identification of promiscuous

Human Leukocyte Antigens (HLA)-restricted T-cell epitopes, identification of Major Histocompatibility Complex (MHC) supermotifs and T-cell epitope hot-spots.

Computational models of cellular or higher level processes or interactions have a longer history than those focusing on molecular processes, but are also more complex. These include models of T- cell responses to viruses, analysis of MHC diversity under host-pathogen co-evolution, B-cell maturation, or even the dynamic model of the immune system that can simulate both cellular and humoral immune responses.

The main problems that prevented the use of these models in practical applications, such as design of vaccines and optimization of immunization regimens are: *a)* large combinatorial complexity of the human immune system that could not be supported by existing computational infrastructures, *b)* lack of understanding of specific molecular interactions that resulted in an idealization of representation of molecular interactions as binary strings, and *c)* lack of experimental model data and correlation of model parameters to real-life measurements. Recent developments provide remedies to these problems and we are in the position to address each of these issues.

Grid computing brought powerful computational infrastructure and the capacity that can match the complexity of the real human immune system. Models of molecular interactions have reached high accuracy and we are routinely using prediction methods of antigen processing and presentation to identify the best targets for vaccine constructs. Finally, experimental models of immune responses to tumors and infectious diseases have been successfully modeled computationally.

In this work, we present two different experiences in which we show successful stories in using computational immunology approaches that have been implemented using GRID infrastructure:

- modeling atherosclerosis, a disease affecting arterial blood vessels, that is one of most common disease of the developed countries;

- a biological optimization problem on the Grid, i.e. an optimal protocol search algorithm based on Simulated Annealing (SA) capable to suggest optimal Triplex vaccine dosage used against mammary carcinoma induced lung metastasis.

2. Modeling atherogenesis using GRID computing

Atherosclerosis is, in large part, due to the deposition of low density lipoproteins (LDLs), i.e., plasma proteins carrying cholesterol and triglycerides, that determine the formation of multiple plaques within the arteries (Hanson, 2002; Ross, 1999). The origin of atherosclerosis is still not fully understood. However there are risk factors which increase the probability of developing atherosclerosis in humans. Some of these risk factors are beyond a person's control (smoking, obesity), others seem to have genetic origin (familial hypercholesterolemia, diabetes, hypertension) (Romero-Corral et al., 2006). Common denominator in all the form of atherosclerosis is the elevated level of LDL, which is subject to oxidation becoming oxidized low density lipoproteins (ox-LDL), that promotes an inflammatory response and immune activation in the artery walls (Berliner et al., 1996). The formation of atherosclerotic plaques in the artery reduces both the internal diameter of vessels and the blood flux leading to a number of serious pathologies (Vinereanu, 2006). Early studies demonstrated that ox-LDL can induce activation of monocytes/macrophages, endothelial cells and T cells. Ox-LDLs engulfed by macrophages form the so called foam cells (Steinberg, 1997). These cells represent the nucleus

of the plaques formation. Ox-LDL promotes also immune activation of B cells inducing the production of specific anti ox-LDL antibody (OLAB).

Atherosclerosis and their anatomical consequences cause severe problems. Stenosis (narrowing) and aneurysm of the artery are chronic, slowly progressing and cumulative effects indicating the progression of atherosclerotic disease. In both case the result is an insufficient blood supply to the organ fed by the artery. Most commonly, soft plaque suddenly ruptures, causes the formation of a thrombus that will rapidly slow or stop blood flow, leading to death of the tissues fed by the artery. This catastrophic event is called infarction and is not predictable. The most common event is thrombosis of the coronary artery causing infarction (a heart attack): However, since atherosclerosis is a body wide process, similar events also occur in the arteries of the brain (stroke attack), intestines, kidneys, etc. Those atherosclerosis associated events often cause of dead or serious invalidating diseases and require preventive treatments. Vaccine research for atherosclerosis is a hot pharmaceutical topic.

Recently we proposed a model based on the Agent Based Model (ABM) paradigm (Pappalardo et al., 2008) which reproduces clinical and laboratory parameters associated to atherogenesis. The model and its computer implementation (SimAthero simulator) considers all the relevant variables that play an important role in atherogenesis and its induced immune response, i.e., LDL, ox-LDL, OLAB, chitotriosidase and the foam cells generated in the artery wall.

We present three different situations over a time scale of two years. The standard normal patients where no foam cells are formed; patients having high level of LDL but who delay to apply appropriate treatments and finally patients who may have many events of high level of LDL but takes immediately appropriate treatments.

2.1 Description of the model
2.1.1 The biological scenario

Exogenous and endogenous factors induce in humans a very small, first oxidative process of blood circulating native LDLs (minimally modified LDLs or mm-LDLs). In endothelium mm-LDLs are extensively oxidized from intracellular oxidative products and then recognized by the macrophage scavenger receptor. High level and persistent in time LDLs lead to macrophages engulfment and their transformation in foam cells. Contrary, low level of LDLs and their oxidized fraction, lead to the internalization of the oxidized low density lipoproteins and subsequent presentation by major histocompatibility complex class II at the macrophages surface. Recognition of ox-LDL by macrophages and naive B cells, leads, by T helper lymphocytes cooperation, to the activation of humoral response and production of OLAB. When the OLAB/ox-LDL immune complexes are generated in the vascular wall, the macrophages catch them by the Fc receptor or via phagocytosis and destroy ox-LDL in the lysosome system. During this process, the activated macrophage releases chitotriosidase enzyme, that is then used as a marker of macrophage activation.

2.1.2 The model

To describe the above scenario one needs to include all the crucial entities (cells, molecules, adjuvants, cytokines, interactions) that biologists and medical doctors recognize as relevant in the game. The model described in (Pappalardo et al., 2008) contains entities and interactions which both biologist and MD considered relevant to describe the process.

Atherosclerosis is a very complex phenomenon which involves many components some of them not fully understood. In the present version of the simulator we considered only in

the immune system processes that control the atherogenesis. These processes may occur in immune system organs like lymph nodes or locally in the artery endothelium. To describe the Immune processes we considered both cellular and molecular entities.

Cellular entities can take up a state from a certain set of suitable states and their dynamics is realized by means of state-changes. A state change takes place when a cell interacts with another cell or with a molecule or both of them. We considered the relevant lymphocytes that play a role in the atherogenesis-immune system response, B lymphocytes and helper T lymphocytes. Monocytes are represented as well and we take care of macrophages. Specific entities involved in atherogenesis are present in the model: low density lipoproteins, oxidized low density lipoproteins, foam cells, auto antibodies anti oxidized low density lipoproteins and chitotriosidase enzyme. Cytotoxic T lymphocytes are not taken into consideration because they are not involved in the immune response (only humoral response is present during atherogenesis).

Molecular entities The model distinguishes between simple small molecules like interleukins or signaling molecules in general and more complex molecules like immunoglobulins and antigens, for which we need to represent the specificity. We only represent interleukin 2 that is necessary for the development of T cell immunologic memory, one of the unique characteristics of the immune system, which depends upon the expansion of the number and function of antigen-selected T cell clones. For what is related to the immunoglobulins, we represent only type immunoglobulins of class G (IgG). This just because at the actual state we don't need to represent other classes of Ig and because IgG is the most versatile immunoglobulin since it is capable of carrying out all of the functions of immunoglobulins molecules. Moreover IgG is the major immunoglobulin in serum (75% of serum Ig is IgG) and IgG is the major Ig in extra vascular spaces.

The actual model does not consider multi-compartments processes and mimics all processes in a virtual region in which all interactions take place. Our physical space is therefore represented by a 2D domain bounded by two opposite rigid walls and left and right periodic boundaries. This biological knowledge is represented using an ABM technique. This allows to describe, in a defined space, the immune system entities with their different biological states and the interactions between different entities. The system evolution in space and in time is generated from the interactions and diffusion of the different entities. Compared to the complexity of the real biological system our model is still very naive and it can be extended in many aspects. However, the model is sufficiently complete to describe the major aspects of the atherogenesis-immune system response phenomenon.

The computer implementation of the model (SimAthero hereafter) has two main classes of parameters: the first one refers to values known from standard immunology literature (Abbas et al., 2007; Celada et al., 1996; Goldspy et al., 2000; Klimov et al., 1999); the second one collects all the parameters with unknown values which we arbitrarily set to plausible values after performing a series of tests (*tuning phase*).

The simulator takes care of the main interactions that happens during an immune response against atherogenesis.

Physical proximity is modeled through the concept of lattice-site. All interactions among cells and molecules take place within a lattice-site in a single time step, so that there is no correlation between entities residing on different sites at a fixed time. The simulation space is represented as a $L \times L$ hexagonal (or triangular) lattice (six neighbors), with periodic boundary conditions to the left and right side, while the top and bottom are represented by rigid walls. All entities are allowed to move with uniform probability between neighboring lattices in the

grid with equal diffusion coefficient. In the present release of the simulator chemotaxis is not implemented.

LDLs values can be fixed in order to simulate different patients both in normolipidic condition and in hypercholesterolemic condition. The same applies to ox-LDLs. However human habits change with time and personal life style. A normolipidic patient can change its attitude becoming an hypercholesterolemic one and vice versa. For this reason we allow the simulator to accept varying life style conditions and preventive actions to decrease risk factors.

2.2 Results

The model described include the possibility of mimicking biological diversity between patients. The general behavior of a class of virtual patients arise from the results of a suitable set of patients, i.e., the mean values of many runs of the simulator of different patients under the same conditions. The class of virtual patients described by the model were tuned against *human* data data collected by (Brizzi et al., 2003; 2004) where different conditions, normal and hypercholesterolemic diabetic patients were analyzed.

In this section we analyze the behavior of the same patients in three broad class of clinical conditions to show how SimAthero could be used in order to analyze and predict the effects of various LDL levels in blood. The normal patient simulation is used as control experiment for the other simulations. The differences among these four clinical conditions depend on the LDL level and the time interval which occurs between the time in which concentration of LDL rise above normal level and the time in which the patient take appropriate treatments (lifestyle o drug) to reduce it to normal level.

Jobs were launched using the SimAthero simulator on the COMETA Grid. The submission process was done through the web interface of the ImmunoGrid project (http://www.immunogrid.eu).

A patient with a LDL level of roughly 950-970 ng/μl of blood is considered normal in clinical practice and he has with very low risk of atheroslerotique plaque. The results of SimAthero for a virtual normal patient (Figure 1) show that he will not support the formation of foam cells and, as a consequence, the beginning of atherogenesis process is absent.

We then simulated a scenario in which a patient, due to several reasons (diet, life style, oxidative agents and so on, so forth) leads its LDL level at 1300 ng/μl, taking it up to 1700 ng/mul. Looking at figure 2 one can observe about 12 foam cells per μl at the end of in silico follow up. This leads to a small atherogenesis process due to the high level of LDL.

Lastly (figure 3), we analyzed a virtual patient that initially takes its LDL level to small peaks, causing no damage. After that, he takes its LDL level to a hypercholesterolemic behavior, generating a small damage, as shown. This shows that small LDL alteration are completely taken under control by the normal behavior of the organism, but high LDL peaks lead to foam cells formation and then to the beginning of the atherogenesis process.

2.3 Remarks on atherosclerosis modeling using GRID computing

Atherosclerosis is a pathology where the immune control plays a relevant role. We presented studies on the increased atherosclerosis risk using an ABM model of atherogenesis and its induced immune system response in humans. Very few mathematical models (Cobbold et al., 2002; Ibragimov et al., 2005) and (to our best knowledge) no computational models of atherogenesis have been developed to date.

It is well known that the major risk in atherosclerosis is persistent high level of LDL concentration. However it is not known if short period of high LDL concentration can cause
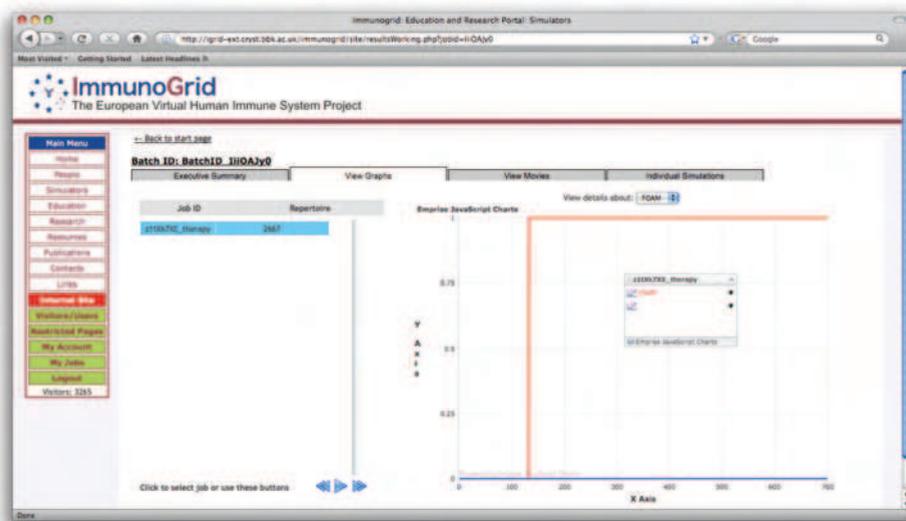
Fig. 1. Simulation results of a virtual patient with level of LDL considered normal. The follow-up period is two years. The figure shows that foam cells formation is absent in this patient.
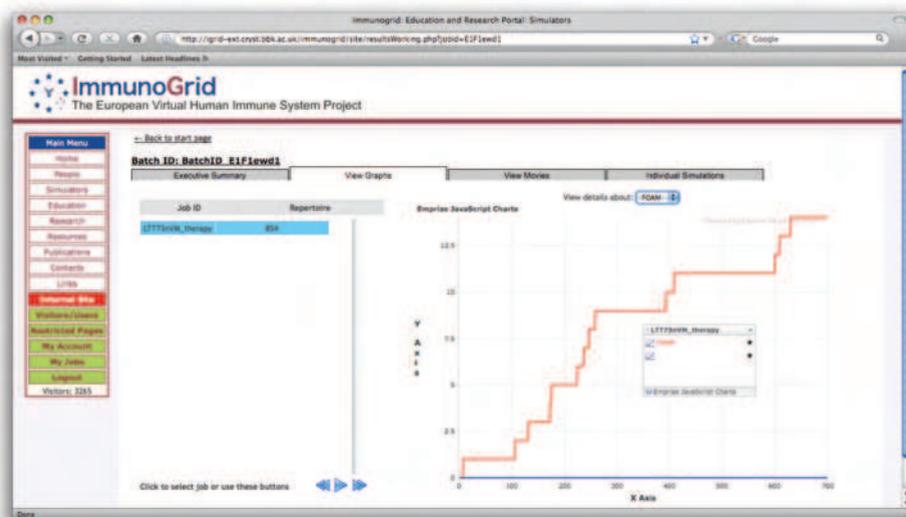

Fig. 2. Simulation results of a virtual patient with level of LDL considered at high risk. The follow-up period is two years. The figure shows that foam cells formation is present, leading to an atherogenesis process.
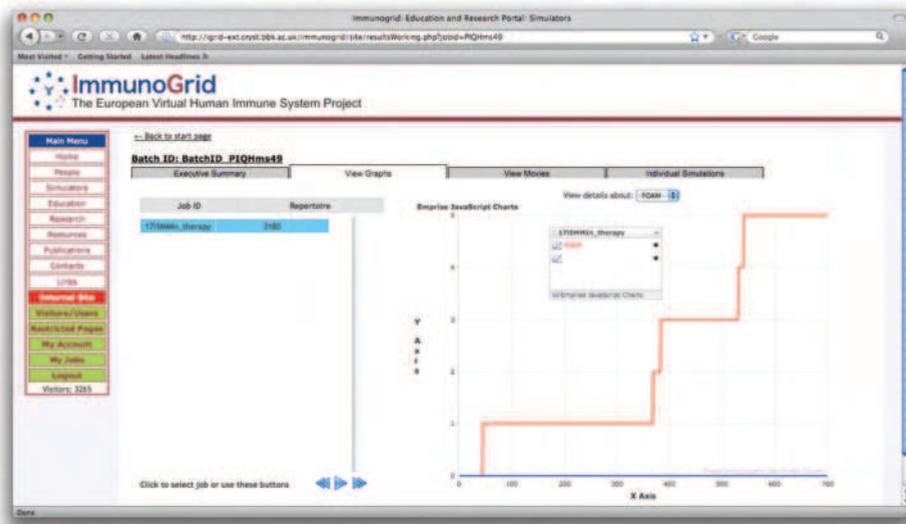
Fig. 3. Simulation results of a virtual patient with level of LDL considered quasi-normal at the beginning and then at high risk. The follow-up period is two years. The figure shows that foam cells formation is negligible in the first time, but becomes important soon after.

irreversible damage and if reduction of the LDL concentration (either by lfe style or drug) can drastically or partially reduce the already acquired risk.
Using an ABM cellular model describing the initial phase of plaque formation (atherogenesis) we are able to simulate the effect of life style which increases the risk of atherosclerosis.

3. Vaccine dosage optimization using GRID

As a second application, we show an example of how a *physiological* model in conjunction with some optimization techniques can be used to speed-up the research of an opitmal vaccination schedule for an immunopreventive vaccine, using the powerful of the GRID computing paradigm.
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 to raise the immune response against the breast cancer for the prevention of the mammary carcinoma formation in HER-2/Neu mouse models using a *Chronic schedule* in a follow up time between 52 and 57 weeks.
However it is not known if the *Chronic schedule* schedule is minimal, i.e. if it can guarantee survival for the mice population avoiding unnecessary/redundant vaccine administrations.

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.

The SimTriplex () is a *physiological* computational model developed with the aim to answer to this question. It demonstrated able to reproduce *in silico* the *in vivo* Immune System (IS) - breast cancer competition elicted by the Triplex vaccine.

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 several days using 32 nodes of an High Performance Computing infrastructure.

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 found an extended description of SA algorithm in (Van Laarhoven at al., 1987)). The combination of the Simulated Annealing algorithm with biologically driven heuristic strategies, leads to a much faster algorithm and better results for the optimal vaccination schedule problem for Triplex vaccine. In this context we remark how the COMETA grid infrastructure demostrated an excellent framework for protocol search and validation. We first executed the SA algorithm on a subset of the virtual mice population using MPI jobs. We therefore checked the protocol quality calculating (with simple jobs) the survivals of the entire population.

3.1 The algorithm

The work done by Kirkpatrick (Kirkpatrick et al., 1983) opened the path to a deep analogy between Statistical Mechanics (the behavior of systems with many degrees of freedom in thermal equilibrium at a finite temperature) and Combinatorial Optimization (the method of finding the minimum, if any, of a given function with respect to many parameters). There is a close similarity, indeed, between the procedure of *annealing* in solids and the framework required for optimization of complex systems.

3.1.1 The optimal vaccination schedule search problem

The SimTriplex model (Pappalardo et al., 2005) has been created with the aim to mimic the behavior of the immune system stimulated by the Triplex vaccine. It simulates all the major interactions of cells and molecules of the immune system in vaccinated as well as naive HER-2/neu mice. *In silico* experiments showed an excellent agreement with the *in vivo* ones.

As previously said, a protocol is said to be optimal if it can maintain efficacy with a minimum number of vaccine administrations. As in standard drug administration, the vaccination protocol must assure survival for a high percentages of patients. Schedule design is usually achieved using *medical consensus*, i.e. a public statement on a particular aspect of medical knowledge available at the time it was written, and that is generally agreed upon as the evidence-based, state-of-the-art (or state-of-science) knowledge by a representative group of experts in that area. Our goal is to improve medical consensus, helping biologists in design vaccine protocols with simulators and optimization techniques.

Let us consider a time interval $[0, T]$, in which we study the action of the vaccine on a set of virtual mice S. We then discretize the given time interval in $N - 1$ equally spaced subintervals of width Δt (=8 hours), i.e. $\{t_1 = 0, t_2, \ldots, t_i, \ldots, t_N = T\}$.

Let $\mathbf{x} = \{x_1, x_2, \ldots, x_i, \ldots x_N\}$ be a binary vector representing a vaccine schedule, where $x_i = 0/1$ means respectively administration/no administration of the vaccine at time t_i. The number of vaccine administrations is given by $n = \sum_{i=1}^{N} x_i$. With $T = 400$ days, and $\Delta t = 8$ hours the search space D has cardinality 2^{400}, excluding any exhaustive search.

One of the wet biologists requirements imposes no more than two administrations a week (monday and thursday) because this is already considered a very intensive schedule from an immunological point of view. This reduces the cardinality of the search space D, from 2^{400} ($\sim 10^{120}$) to 2^{114} ($\sim 10^{34}$). We still have no chance for an exhaustive search.

The time of the carcinoma in situ (CIS) formation is computed through SimTriplex simulator. It is defined by $\tau(\mathbf{x}, \lambda_j)$, which is a function of the vaccination schedule $\mathbf{x}$ administered to the mouse $j \in S$ and a parameter λ_j which represents the biological diversity. The vaccine will be obviously effective if $\tau \geq T$.

As pointed out in § 1, any optimal protocol should try to reproduce , *in silico*, the chronic time evolution of cancer cells. This leads us to the need to use two thresholds on the allowed maximum number of cancer cells.

We also note here that, due to biological variability, a schedule found for a single mouse is not immunologically effective as it usually does not reveal able to protect high percentages of the treated patients. Having this in mind, we formulate our optimization problem as follows.

Let $\{j_1, j_2, \ldots, j_m\} \subset S$, with $m = 8$, a random chosen subset of *in silico* mice, the problem is defined as:

$$
\begin{cases}
\tau(\bar{\mathbf{x}}, \lambda_{j_1}) = \max(\tau(\mathbf{x}, \lambda_{j_1})) \\
\tau(\bar{\mathbf{x}}, \lambda_{j_2}) = \max(\tau(\mathbf{x}, \lambda_{j_2})) \\
\qquad \vdots \qquad \vdots \\
\tau(\bar{\mathbf{x}}, \lambda_{j_m}) = \max(\tau(\mathbf{x}, \lambda_{j_m})) \\
n(\bar{\mathbf{x}}) = \min(n(\mathbf{x})) \\
\text{subject to:} \\
M_1(\bar{\mathbf{x}}) \leq \gamma_1, \ t \in [0, T_{in}] \\
M_2(\bar{\mathbf{x}}) \leq \gamma_2, \ t \in [T_{in}, T]
\end{cases}
\tag{1}
$$

where $M_1(\mathbf{x})$ and $M_2(\mathbf{x})$ are the maximum number of cancer cells in $[0, T_{in}]$ (cellular-mediated controlled phase) and in $[T_{in}, T]$ (humoral-mediated controlled phase) respectively, and $T_{in} \sim T/3$, while γ_1 and γ_2 represent cancer cells threshold in $[0, T_{in}]$ and in $[T_{in}, T]$, respectively.

We deal with a multi-objective discrete and constrained optimization problem.

We modified this last formulation of the problem grouping all the $\tau(\mathbf{x}, \lambda_{j_h})$ ($h = 1, \ldots, m$) by a proper statistical indicator. We chose the harmonic mean H of the survivals:

$$
H(\mathbf{x}, \lambda_{j_1}, \ldots, \lambda_{j_m}) = \frac{m}{\displaystyle\sum_{i=1}^{m} \frac{1}{\tau(\mathbf{x}, \lambda_{j_i})}}
\tag{2}
$$

since it is very frequently used when statistic measurements of time are involved.

Therefore, the system (1) translates as:

$$
\begin{cases}
H(\bar{\mathbf{x}}, \vec{\lambda}) = \max\left(H(\mathbf{x}, \vec{\lambda})\right) \\
n(\bar{\mathbf{x}}) = \min(n(\mathbf{x})) \\
\text{subject to:} \\
M_1(\bar{\mathbf{x}}) \leq \gamma_1, \ t \in [0, T_{in}] \\
M_2(\bar{\mathbf{x}}) \leq \gamma_2, \ t \in [T_{in}, T]
\end{cases}
\tag{3}
$$

with $\vec{\lambda} = (\lambda_{j_1}, \dots, \lambda_{j_m})$.

3.1.2 Simulated annealing

An acclaimed Monte Carlo method, commonly referred as the *Metropolis criterion*, has been designed by Metropolis (Metropolis et al., 1953) with the aim to mimic the evolution of the complex systems towards equilibrium at a fixed temperature. This method randomly perturbates the position of the particles of a solid modifying its *configuration*. If the energy difference, ΔE, between the unperturbed and perturbed configurations is negative, the new configuration has lower energy and it's considered as the new one. Otherwise the probability of acceptance of the new configuration is given by the *Boltzmann factor* (which expresses the "probability" of a state with energy E relative to the probability of a state of zero energy).

After a large number of perturbations the probability distribution of the states should approach the Boltzmann distribution.

The Metropolis algorithm can also be used in combinatorial optimization problems to generate sequences of configurations of a system using a *cost function C* and control parameter *c* respectively as the energy and temperature in the physical annealing.

The SA algorithm can be represented as a sequence of Metropolis algorithms evaluated at a sequence of decreasing values of the control parameter *c*. A generalization of the method is given as follows: a generation mechanism is defined so that, given a configuration *i*, another configuration *j* can be obtained by choosing at random an element in the neighborhood of *i*. If $\Delta C_{ij} = C(j) - C(i) \leq 0$, then the probability that the next configuration is *j* is given by 1; if $\Delta C_{ij} > 0$ the probability is given by $e^{-\Delta C_{ij}/c}$ (Metropolis criterion).

This process is continued until the probability distribution, *P*, of the configurations approaches the Boltzmann distribution, which translates as:

$$P\{configuration = i\} = \frac{1}{Q(c)} \cdot e^{-C(i)/c} \qquad (4)$$

where $Q(c)$ is a normalization constant depending on the control parameter *c*.

The control parameter is then lowered in steps, with the system being allowed to reach equilibrium by generating a chain of configurations at each step. The algorithm stops for some fixed value of the control parameter *c* where virtually no deteriorations can be accepted anymore.

At the end the final *frozen* configuration is assumed as a solution of the problem.

3.1.3 Implementation

In our *in silico* experiment, we select a population of 200 virtual mice and a simple random sample of $k = 8$ mice. To use the SA algorithm for the optimal vaccine schedule search problem we tried to define the SA relevant concepts (the solid configuration, the temperature, the energy and the semi-equilibrium condition) in terms of vaccine protocol and to describe the main protocol elements (the number of injections, the mean survival age, the time distribution of injections) in terms of a cooling process.

As previously said, we describe any candidate protocol as a binary vector **x** of cardinality $V = 114$. The total number of vaccine administrations *n* and the total number of possible schedules with *n* vaccine administrations $\mathcal{M}$ are given by:

$$n = \sum_{j=1}^{V} x_n^l(j)$$

$$\mathcal{M} = V! / [n! \, (V - n)!]$$

The configuration distribution is defined by $\mathbf{x}_{in}$, n_{in}, and the initial energy E_{in} as defined later on.

The *temperature* is slowly but constantly lowered to reach a state with minimum energy. We coupled this entity with n, the *number of vaccine administrations* of a semi-equilibrium configuration.

At a given temperature, a semi-equilibrium configuration is reached when its *Energy* is minimal. Since we want to maximize survival times of a mice sample set, the concept of Energy can be easily associated with the *harmonic mean H* of the survival times τ_i, $H(\bar{\mathbf{x}}, \vec{\lambda})$ (i.e. $E \propto H$). As a matter of fact H decreases when the cumulative survival time of the sample increases, in perfect accord with the energy definition.

The perturbation of a protocol has been initially implemented as random *1* bits reallocation (Pennisi et al., 2008b). This perturbation has been heuristically improved using *biological* knowledge. As we want to optimize mice survival, scheduling many vaccine administrations after the death of a mouse makes nonsense. So, in moving from x_i^j to x_i^{j+1}, we improve random bits reallocation also moving some "1" at a suitable time $t < min\{\tau_i\}$, $i = 1, k$.

The SA algorithm for protocol optimization. *i*) starts from a randomly chosen initial vaccine distribution and finds the initial semi-equilibrium configuration n_{in}, $\mathbf{x}_{n_{in}}^{l_{in}}$, $E_{n_{in}}^{l_{in}}$

ii) Decrease the number of injections of 1 unit; *iii*) find a semi-equilibrium configuration $\mathbf{x}_i$, 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.

3.2 Computational results and conclusions

In silico optimal protocol search is a two-step process: *search* and *validation*. During the search step, the optimization tecnique tries to find an optimal protocol. As pointed out in section 3.1.1, optimal search stragies have to be executed on a representative subset of the population in order to guarantee significative survival percentages for the population.

The search technique therefore needs to test simultaneously every candidate protocol on the mice subset to compute its fitness function. This process usually requires a relatively small number of nodes with high communication throughputs, representing a typical massive parrallel application. In our case it has been implemented using the MPI (Message passing Interface) libraries.

Validation represents a completely different process. Here the protocol found by the search technique is tested over the entire population to compute mean survival rates, requiring an high number of CPUs with almost no need of communication. The "*search and validation*" process is represented in figure 4.

In this context the Cometa grid revealed itself as an excellent tool for our needs. It demonstrated to be highly flexible, giving us the opportunity to execute these so-different processes on the same infrastructure. We only needed to define the first process as an *MPI*
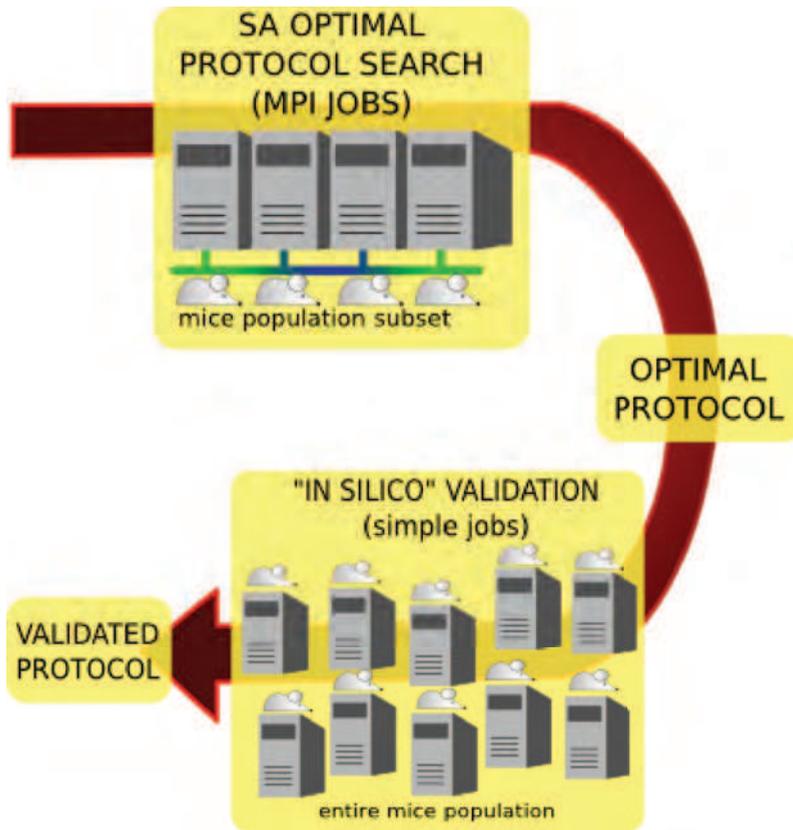
Fig. 4. The "in silico" *search* and *validation* optimal protocol representation.

job and the second as a sequence of *simple jobs* in a trasparent way, without worrying of the hidden the architectures behind the job submission interface.

To compare the results with those obtained using GA optimization technique (Lollini et al., 2006), we executed the SA algorithm on the same 8 random selected virtual mice sample used by GA. The protocol was then validated on the same population set (200 virtual mice).

The SA *in silico* tumor free percentages of the mice population show no substantial difference with GA results (87% for GA vs 86,5% for SA). Figure 5 shows the mean number of cancer cells, computed on the 200-mice set, for the GA-protocol (up lhs) and the SA-protocol (down lhs). Only the SA-protocol is able to totally fulfill the safety threshold conditions (shown in red).

Moreover the SA algorithm required a computational effort of about 2 hrs on a 8 processor unit to find a protocol with 37 vaccine administrations, showing speed-up factor of $\sim 1.4 \cdot 10^2$ in respect to previous GA experiments (Pennisi et al., 2008a).
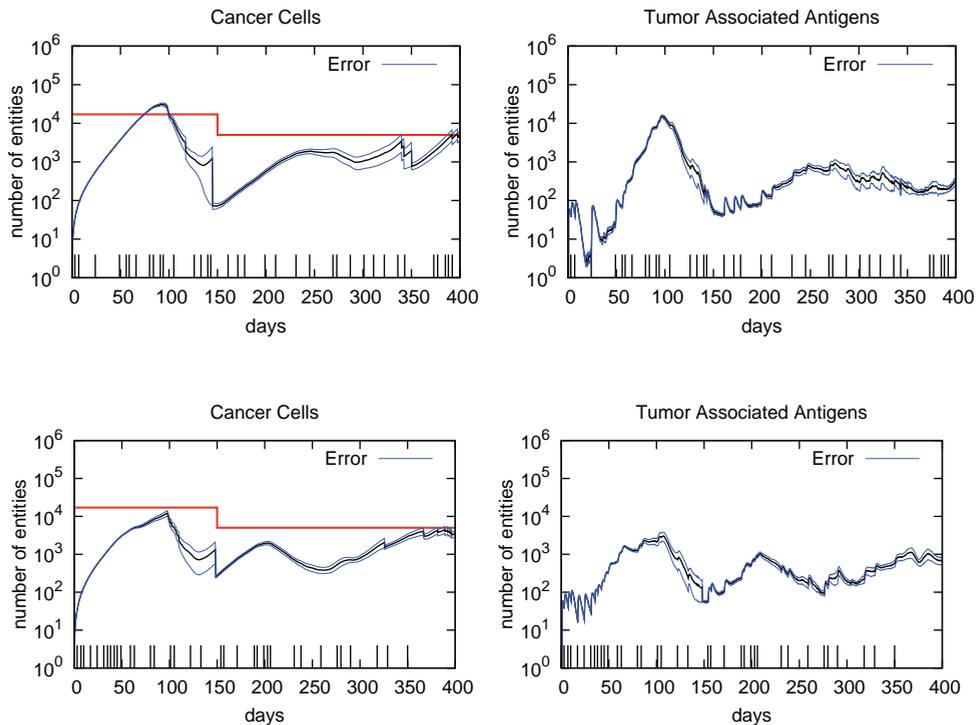
Fig. 5. Cancer cells behaviors and thresholds in GA (top) and SA (bottom). Small vertical bars on the x-axis represent vaccine administration times. Broken-line graphs on the lhs represent safety thresholds.

4. Conclusions

Mathematical models and Cellular Automata are mostly used for cellular level simulations, while a range of statistical modeling applications are suitable for the analysis of sequences at molecular level of the immune system.

Grid computing technology brings the possibility of simulating the immune system at the natural scale. In our opinion, a Grid solution is only as good as the interface provided to the users. We presented two successful stories in which we have shown successful stories in using computational immunology approaches that have been implemented using GRID infrastructure.

We like to conclude by stressing the interdisciplinary nature of the experiences described above and by noting that the contribution of life scientists needs to go beyond the only data supply, as it is extremely important in defining the biological scenario and ultimately construct a robust and validated mathematical or computational model. Only through a common effort of life and computer scientists it is possible to turn software into a valuable tools in life sciences.

5. Acknowledgments

6. References

Abbas, A.K.; Lichtman, A.H.; Pillai, S. (2007) Cellular and molecular immunology, *Saunders*, 6th edition.

Artieda, M.; Cenarro, A.; Gañàn, A.; Jericó, I.; Gonzalvo, C.; Casado, J.M.; Vitoria, I.; Puzo, J.; Pocoví,M.; Civeira,F. (2003) Serum chitotriosidase activity is increased in subjects with atherosclerosis disease, *Arterioscler. Thromb. Vasc. Biol.*, 23, 1645-1652.

Artieda, M.; Cenarro, A.; Gañàn, A.; Lukic, A.; Moreno, E.; Puzo, J.; Pocoví, M.; Civeira, F. (2007) Serum chitotriosidase activity, a marker of activated macrophages, predicts new cardiovascular events independently of C-Reactive Protein, *Cardiology*, 108, 297-306.

Berliner, J.A.; Heinecke, J.W. (1996) The role of oxidized lipoproteins in atherogenesis, *Free Radic. Biol. Med.*, 20(5), 707-727.

Brizzi, P.; Isaja, T.; D'Agata, A.; Malaguarnera, A.; Malaguarnera, M.; Musumeci, S. (2002) Oxidized LDL antibodies (OLAB) in patients with beta-thalassemia major, *J. Atheroscler. Thromb.*, 9(3), 139-144.

Brizzi, P.; Tonolo, G.; Carusillo, F.; Malaguarnera, M.; Maioli, M.; Musumeci, S. (2003) Plasma Lipid Composition and LDL Oxidation, *Clin. Chem. Lab. Med.*, 41(1), 56-60.

Brizzi, P.; Tonolo, G.; Bertrand, G.; Carusillo, F.; Severino, C.; Maioli, M.; Malaguarnera, L.; Musumeci, S. (2004) Autoantibodies against oxidized low-density lipoprotein (ox-LDL) and LDL oxidation status, *Clin. Chem. Lab. Med.*, 42(2), 164-170.

Celada, F.; Seiden, P.E. (1996) Affinity maturation and hypermutation in a simulation of the humoral immune response, *Eur. J. Immunol.*, 26, 1350.

Cobbold, C.A.; Sherratt, J.A.; Maxwell, S.R.J. (2002) Lipoprotein Oxidation and its Significance for Atherosclerosis: a Mathematical Approach, *Bulletin of Mathematical Biology*, 64, 65-95.

Goldsby, R.A. et al. (2000) In Austen,K.F., Frank,M.M., Atkinson,J.P. and Cantor,H. (eds.) Kuby Immunology. W.H. Freeman and Company, New York.

Hanson, G.K. (2002) Inflammation, atherosclerosis, and coronary artery disease, *N. Engl. J. Med.*, 352(16), 1685-1695.

Ibragimov, A.I.; McNeal, C.J.; Ritter, L.R.; Walton,J.R. (2005) A mathematical model of atherogenesis as an inflammatory response, *Math Med Biol.*, 22(4), 305-333.

Klimov, A.N.; Nikul'cheva, N.G. (1999) Lipid and Lipoprotein Metabolism and Its Disturbances, St. Petersburg: Piter Kom.

Lollini,P.-L. (2008) Private communication.

Orem, C.; Orem, A.; Uydu, H.A.; Celik, S.; Erdol, C.; Kural,B.V. (2002) The effects of lipid-lowering therapy on low-density lipoprotein auto-antibodies: relationship with low-density lipoprotein oxidation and plasma total antioxidant status, *Coron. Artery Dis.*, 13(1), 56-71.

Pappalardo, F.; Musumeci, S.; Motta, S. (2008) Modeling immune system control of atherogenesis, *Bioinformatics*, 24:15, 1715-1721.

Romero-Corral, A.; Somers, V.K.; Korinek, J.; Sierra-Johnson, J.; Thomas, R.J.; Allison, T.G.; Lopez-Jimenez,F. (2006) Update in prevention of atherosclerotic heart disease: management of major cardiovascular risk factors, *Rev. Invest. Clin.*, 58(3), 237-244.

Ross,R. (1999) Atherosclerosis–an inflammatory disease, *N. Engl. J. Med.*, 340(2), 115-126.

Tonolo,G. (2008) Private communication.

Shaw, P.X.; Hörkkö, S.; Tsimikas, S.; Chang, M.K.; Palinski, W.; Silverman, G.J.; Chen, P.P.; Witztum, J.L. (2001) Human-derived anti-oxidized LDL autoantibody blocks uptake of oxidized LDL by macrophages and localizes to atherosclerotic lesions in vivo, *Arterioscler. Thromb. Vasc. Biol.*, 21(8), 1333-1339.

Shoji, T.; Nishizawa, Y.; Fukumoto, M.; Shimamura, K.; Kimura, J; Kanda, H.; Emoto, M.; Kawagishi, T.; Morii,H. (2000) Inverse relationship between circulating oxidized low density lipoprotein (oxLDL) and anti-oxLDL antibody levels in healthy subjects, *Atherosclerosis*, 148(1), 171-177.

Steinberg,D. (1997) Low density lipoprotein oxidation and its pathobiological significance, *J. Biol. Chem.*, 272(34), 20963-20966.

Tinahones, F.J.; Gomez-Zumaquero, J.M.; Rojo-Martinez, G.; Cardona, F.; Esteva de Antonio, I.E.; Ruiz de Adana, M.S.; Soriguer, F.K. (2002) Increased levels of anti-oxidized low-density lipoprotein antibodies are associated with reduced levels of cholesterol in the general population, *Metabolism.*, 51(4), 429-431.

Tinahones, F.J.; Gomez-Zumaquero, J.M.; Garrido-Sanchez, L.; Garcia-Fuentes, E.; Rojo-Martinez, G.; Esteva, I.; Ruiz de Adana, M.S.; Cardona, F.; Soriguer,F. (2005) Influence of age and sex on levels of anti-oxidized LDL antibodies and anti-LDL immune complexes in the general population, *J. Lipid Res.*, 46(3), 452-457.

Vinereanu,D. (2006) Risk factors for atherosclerotic disease: present and future, *Herz*, Suppl. 3, 5-24.

Metropolis, N.; Rosenbluth, A.W.; Rosenbluth, M.N.; Teller, A.H.; Teller, E. (1953) "Equation of State Calculations by Fast Computing Machines", *Journal of Chemical Physics*, 21, 1087-1092.

Kirkpatrick, S.; Gelatt, C. D.; Vecchi, Jr., M. P. Optimization by simulated Annealing Science, Vol. 220, No. 4598, 671-680, 1983

Van Laarhoven, P.J.M.; Aarts, E.H.L. (1987) Simulated Annealing: Theory and Applications, *Springer Edition*.

Kirschner, D.; Panetta, J.C. (1998) Modeling immunotherapy of the tumor-immune interaction. *J. Math. Biol.*, (37), 235-252.

Nani,F.,Freedman,S. (2000) A mathematical model of cancer treatment by immunotherapy, *Math. Biosci.*, 163, 159-199.

Pappalardo, F.; Lollini, P.L.; Castiglione, F; Motta, S. Modeling and simulation of cancer immunoprevention vaccine. Bioinformatics. 21, 2891-2897

Agur, Z.; Hassin, R.; Levy, S. Optimizing chemotherapy scheduling using local search heuristics, Operations Research, 54:5, 829-846.

Kumar, N.; Hendriks, B.S.; Janes, K.A.; De Graaf, D.; Lauffenburger D.A. Applying computational modeling to drug discovery and development Drug Discovery Today Volume 11, Issues 17-18, September 2006, Pages 806-811

Lollini, P.L.; Motta, S.; Pappalardo, F. Discovery of cancer vaccination protocols with a genetic algorithm driving an agent based simulator. BMC. Bioinformatics. 7, 352.

Davies, M.N.; Flower, D.R. (2007) Harnessing bioinformatics to discover new vaccines, *Drug Discovery Today*, 12(9-10), 389-395.

Castiglione, F.; Piccoli, B. (2007) Cancer immunotherapy, mathematical modeling and optimal control, *J. Theo. Biol.*, 247, 723-732.

Pennisi, M.; Catanuto, R.; Pappalardo, F.; Motta, S. (2008) Optimal vaccination schedules using Simulated Annealing, *Bioinformatics*, 24:15, 1740-1742.

Pennisi, M.; Catanuto, R.; Mastriani, E.; Cincotti, A.; Pappalardo,F.; Motta,S. (2008) Simulated Annealing And Optimal Protocols, Journal of Circuits Systems, and Computers, 18:8, 1565-1579.
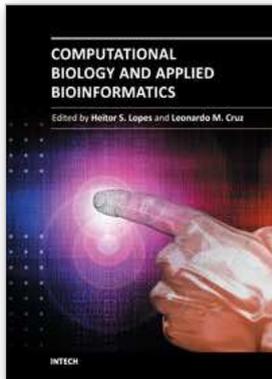
Computational Biology and Applied Bioinformatics

Edited by Prof. Heitor Lopes

Nowadays it is difficult to imagine an area of knowledge that can continue developing without the use of computers and informatics. It is not different with biology, that has seen an unpredictable growth in recent decades, with the rise of a new discipline, bioinformatics, bringing together molecular biology, biotechnology and information technology. More recently, the development of high throughput techniques, such as microarray, mass spectrometry and DNA sequencing, has increased the need of computational support to collect, store, retrieve, analyze, and correlate huge data sets of complex information. On the other hand, the growth of the computational power for processing and storage has also increased the necessity for deeper knowledge in the field. The development of bioinformatics has allowed now the emergence of systems biology, the study of the interactions between the components of a biological system, and how these interactions give rise to the function and behavior of a living being. This book presents some theoretical issues, reviews, and a variety of bioinformatics applications. For better understanding, the chapters were grouped in two parts. In Part I, the chapters are more oriented towards literature review and theoretical issues. Part II consists of application-oriented chapters that report case studies in which a specific biological problem is treated with bioinformatics tools.

How to reference

In order to correctly reference this scholarly work, feel free to copy and paste the following:

INTECH
open science | open minds