

ImmunoGrid: towards agent-based simulations of the human immune system at a natural scale

Mark Halling-Brown, Francesco Pappalardo, Nicolas Rapin, Ping Zhang, Davide Alemanni, Andrew Emerson, Filippo Castiglione, Patrice Duroux, Marzio Pennisi, Olivo Miotto, Daniel Churchill, Elda Rossi, David S. Moss, Clare E. Sansom, Massimo Bernaschi, Marie-Paule Lefranc, Søren Brunak, Ole Lund, Santo Motta, Pier-Luigi Lollini, Annalisa Murgu, Arianna Palladini, Kaye E. Basford, Vladimir Brusich and Adrian J. Shepherd

Phil. Trans. R. Soc. A 2010 **368**, 2799-2815
doi: 10.1098/rsta.2010.0067

References

This article cites 43 articles, 13 of which can be accessed free
<http://rsta.royalsocietypublishing.org/content/368/1920/2799.full.html#ref-list-1>

Rapid response

Respond to this article
<http://rsta.royalsocietypublishing.org/letters/submit/roypta;368/1920/2799>

Subject collections

Articles on similar topics can be found in the following collections

[computational biology](#) (117 articles)

Email alerting service

Receive free email alerts when new articles cite this article - sign up in the box at the top right-hand corner of the article or click [here](#)

To subscribe to *Phil. Trans. R. Soc. A* go to:
<http://rsta.royalsocietypublishing.org/subscriptions>

ImmunoGrid: towards agent-based simulations of the human immune system at a natural scale†

BY MARK HALLING-BROWN¹, FRANCESCO PAPPALARDO², NICOLAS RAPIN³,
PING ZHANG⁴, DAVIDE ALEMANI⁵, ANDREW EMERSON⁶,
FILIPPO CASTIGLIONE⁷, PATRICE DUROUX⁸, MARZIO PENNISI²,
OLIVO MIOTTO⁹, DANIEL CHURCHILL¹⁰, ELDA ROSSI⁶, DAVID S. MOSS¹,
CLARE E. SANSOM¹, MASSIMO BERNASCHI⁷, MARIE-PAULE LEFRANC⁸,
SØREN BRUNAK³, OLE LUND³, SANTO MOTTA², PIER-LUIGI LOLLINI¹¹,
ANNALISA MURGO¹¹, ARIANNA PALLADINI¹¹, KAYE E. BASFORD⁵,
VLADIMIR BRUSIC^{5,12} AND ADRIAN J. SHEPHERD^{1,*}

¹*Institute of Structural and Molecular Biology, Department of Biological Sciences, Birkbeck College, University of London, Malet Street, London WC1E 7HX, UK*

²*Department of Mathematics and Computer Science, University of Catania, Catania, Italy*

³*Centre for Biological Sequence Analysis, Danish Technical University, Lyngby, Denmark*

⁴*Faculty of Health Sciences and Medicine, Bond University, Robina, Queensland, Australia*

⁵*School of Land, Crop and Food Sciences, University of Queensland, Brisbane, Australia*

⁶*High Performance Systems Division, CINECA, Casalecchio di Reno, Italy*

⁷*Institute for Applied Computing, Italian National Research Council (CNR), Rome, Italy*

⁸*Laboratoire d'ImmunoGénétique Moléculaire, Institut de Génétique Humaine, CNRS, Montpellier, France*

⁹*Institute of Systems Science, National University of Singapore, Singapore*

¹⁰*Division of Information and Technology Studies, University of Hong Kong, Pokfulam, Hong Kong*

¹¹*Section of Cancer Research, University of Bologna, Italy*

¹²*Cancer Vaccine Center, Dana-Farber Cancer Institute, Boston MA, USA*

The ultimate aim of the EU-funded ImmunoGrid project is to develop a natural-scale model of the human immune system—that is, one that reflects both the diversity and the relative proportions of the molecules and cells that comprise it—together with the

*Author for correspondence (a.shepherd@mail.cryst.bbk.ac.uk).

†The authors of this paper belong to the ImmunoGrid Consortium.

One contribution of 13 to a Theme Issue ‘The virtual physiological human: computer simulation for integrative biomedicine I’.

grid infrastructure necessary to apply this model to specific applications in the field of immunology. These objectives present the ImmunoGrid Consortium with formidable challenges in terms of complexity of the immune system, our partial understanding about how the immune system works, the lack of reliable data and the scale of computational resources required.

In this paper, we explain the key challenges and the approaches adopted to overcome them. We also consider wider implications for the present ambitious plans to develop natural-scale, integrated models of the human body that can make contributions to personalized health care, such as the European Virtual Physiological Human initiative.

Finally, we ask a key question: How long will it take us to resolve these challenges and when can we expect to have fully functional models that will deliver health-care benefits in the form of personalized care solutions and improved disease prevention?

Keywords: systems biology; agent-based simulation; immunoinformatics; Grid computing; vaccine discovery

1. Introduction

The ImmunoGrid Consortium was funded by the European Commission in 2006 through the Framework 6 programme with the aim of developing a natural-scale model of the human immune system together with the Grid infrastructure necessary to apply this model to specific applications in the field of immunology (Pappalardo *et al.* 2009). The Consortium brings together researchers from several countries (Denmark, France, Italy, the UK and Australia) with expertise in the areas of immunoinformatics, Grid technologies and experimental models of cancer immunotherapy.

From its inception, it was clear that the Consortium would face formidable challenges. The immune system is a biological system of extreme complexity, there is a lack of reliable data about many of its constituent cells and molecules, and any simulation that aims to model the immune system at a natural scale will inevitably require substantial computational resources. Arguably, ImmunoGrid has the task of modelling one of the most challenging components of the Virtual Physiological Human (VPH).¹ In this paper, we explain the key challenges faced by the ImmunoGrid Consortium and the approaches we are adopting to overcome them.

Ultimately, the wider challenges of modelling the human immune system are shared with the current ambitious plans to develop an integrated natural-scale model of the human body. This framework comprises a set of models integrated within the VPH that are aimed to be descriptive, integrative and predictive (Fenner *et al.* 2008). Observations from nature, clinics and experiments are collected, catalogued, organized, combined and shared through the VPH framework, allowing descriptions of the systems, processes and entities of the human organism to be formulated. The VPH framework allows multiple observations to be integrated and analysed collaboratively, by experts from multiple fields enabling postulation of systemic hypotheses and testing them.

¹<http://www.vph-noe.eu/>

Finally, it allows the interconnection of predictive models that are defined at multiple scales (molecules, cells, tissues and organs, body-wide systems, the whole organism, and collection of organisms) into systemic networks. These networks can address systemic hypotheses and can help validate the hypotheses by combining predictive modelling, clinical observations and experimentation. The VPH is currently developed through several initiatives, including the European VPH (Fenner *et al.* 2008), and these are expected to enable an integrative and analytical approach to the study of medicine and physiology and to drive the paradigm shift in health care. The key benefits that the VPH aims to deliver are: a holistic approach to medicine, personalized care solutions, a reduced need for animal experiments and a preventative approach to the treatment of disease. Models of the immune system and immune responses provide key links for the study of human disease (including infectious disease, cancer, allergy and autoimmunity) as well as medical interventions such as immunotherapy, vaccination, transplantation and others.

2. The human immune system

In the whole of biology, the immune system is arguably second only to the central nervous system in terms of complexity and our knowledge about how it works, while expanding rapidly, remains incomplete.

Even a concise description of the immune system lies outside the scope of this paper and the reader requiring further information is referred to an introductory book, such as *Janeway's immunobiology* (Murphy *et al.* 2008). Here, we list the main points that summarize the complexities of the immune system:

- *Systemic complexity.* One essential function of the immune system is to destroy invading infectious agents. These agents gain access to the body through wounds and lesions, or through the respiratory, gastrointestinal and urogenital tracts, where they first encounter the *mucosal immune system* and where most pathogens are usually quickly destroyed before they can cause clinical symptoms. If this defence is penetrated, pathogens meet the *systemic immune system* whose organs, such as the thymus, spleen and lymph nodes, are linked by the vascular and lymphatic systems. Cells of the immune system (and among them the B and T cells that are specific to the adaptive immune system of vertebrates) are derived from stem cells in the bone marrow, and plasma cells (which differentiate from B cells having recognized an antigen) return to the bone marrow to produce antibodies. Antigen recognition and proliferation of reactive T and B cells takes place in specialized areas of the spleen and in more than 500 lymph nodes distributed throughout the body.
- *Complexities of scale.* The human immune system is hierarchical and operates at molecular, cellular, organ/tissue, organism and group of organism levels. For example, at the molecular level, major histocompatibility complex (MHC) proteins that present on the cell surface peptides cleaved from antigen proteins are integral to the process of antigen recognition by cytotoxic and helper T cells, whereas soluble cytokines are crucial for intercellular signalling. At the cellular level, cytotoxic T cells

contribute to the neutralization of intracellular pathogens and potential cancers by eliminating the infected or malfunctioning cells, whereas plasma cells derived from B cells contribute to the neutralization of extracellular pathogens through the production of antibodies. At the organ level, the thymus has an essential role in the maturation of T cells and the elimination of self-reactive T cells, while the lymphatic system provides an essential mechanism for transporting immune response cells and molecules to sites of infection.

- *Spatial complexities.* The immune system is highly distributed, involving signalling and diffusion of cells and molecules throughout the body. There is no centralized control; rather, the response to a given infection emerges from the combined actions of vast numbers of molecules and cells. The location and extent of a given infection or cancerous growth are factors that are relevant to an individual's prognosis; hence, a uniform model for the whole human body would not capture important aspects of observed behaviour.
- *Diversity of molecular and cellular entities.* Variability of gene and protein sequence presents considerable challenges to natural-scale modelling of immune responses. Each individual has a unique repertoire of molecular entities such as antigen receptors, immunoglobulins (Igs) on the surface of B cells or secreted by the plasma cells, and T cell receptors (TCRs) on the surface of T cells, together with the polymorphic MHC (in humans, human leucocyte antigen, HLA) proteins. The repertoire of antigen receptors within a given individual changes constantly and the immune profile varies over that person's lifetime, predominantly as a response to his/her interactions with the environment (e.g. the micro-organisms encountered, the immunizations undergone). Each individual will exhibit approximately 4×10^{12} different B and T cell antigen receptors produced by somatic gene rearrangements, mechanisms of junctional diversity and, for the Ig, somatic hypermutations (that are unique features of the adaptive immune system). Furthermore, MHC genetic polymorphism gives rise to (as one example) more than 1500 different class I alleles in the human population (Sette *et al.* 2005), which means that, depending on their haplotypes, individuals will differ in the presentation of antigenic peptides. It can be possible to use subsets to model natural-scale populations of cells or molecules in an efficient and effective way; for example, Smith *et al.* (1998) performed realistic simulations of B and T cell populations using subsets of less than 0.001 per cent of the total repertoire. However, not all scenarios allow this simplification: approximations are particularly difficult to apply to rapidly evolving systems, such as the HIV (Castiglione *et al.* 2004) and influenza A viruses (Handel *et al.* 2009) under selective pressure from the immune system. A further layer of complexity comes from the differentiation of specialized cell subsets, such as helper, cytotoxic, regulatory (suppressive) and memory T cells, with novel functions and new cell subsets constantly being added by advanced research.
- *Temporal complexities.* The immune system operates on a wide range of time scales, ranging from seconds (to take account of intracellular signal transduction after receptor engagement, for example) to years (the effects of memory cells on repeat infections).

It is worth noting that, in terms of the disparate spatial and temporal scales, the immune system presents approximately the same order of modelling challenges as those associated with the human body as a whole.

3. Modelling the immune system

A significant number of tools and simulators have been developed for predicting aspects of the adaptive immune system, although much less attention has been paid to modelling innate immunity. Various tools have been developed for predicting key molecular aspects of adaptive immunity, notably T cell and B cell epitopes, proteasomal cleavage and peptide binding to TAP (the transporter associated with antigen processing), from protein sequences—see [Lundegaard *et al.* \(2007\)](#) for a useful summary. The prediction of class I T-cell epitopes has proved particularly successful, with the tools developed at the Center for Biological Sequence Analysis (CBS) in Denmark ([Larsen *et al.* 2005](#)) proving to be the most accurate in a recent independent evaluation ([Lin *et al.* 2008](#)).

Although, as we shall see, such molecular tools can make a useful contribution to large-scale simulation of the immune system, they do not attempt to capture the dynamics of the immune system as a whole. As [Louzoun \(2007\)](#) points out in an interesting paper about the evolution of mathematical modelling in the field of immunology, there has been a significant shift in the types and scope of mathematical models of immune systems over the past decade, with interest moving increasingly from classical mathematical models based primarily on ordinary differential equations (ODEs) to other paradigms, including Monte Carlo simulations. Agent-based models, of which those developed through ImmunoGrid are examples, are becoming more popular ([Bauer *et al.* 2009](#)). These are stochastic models that describe populations of interacting agents, such as the molecules and cells of the immune system, using a system of simple rules. However, modelling the immune system is still more or less in its infancy; all models of the immune system, including our own, are relatively primitive. For a recent concise review of computational models of the immune system, see [Pappalardo *et al.* \(2008b\)](#); for a comparison of agent-based models with ODEs, as applied to the specific problem of the optimization of vaccine protocols, see [Pappalardo *et al.* \(2010\)](#).

On the ImmunoGrid project, we have developed two core immune system simulators—one a revised version of C-IMMSIM, the other SIMTRIPLEX—with a common code base derived from the original C-IMMSIM ([Castiglione *et al.* 1997](#)). Both simulators use lattice-gas cellular automata (LGCA) as their core mathematical approach, a model originally developed for immune system simulation by [Celada & Seiden \(1992\)](#). Each lattice position may contain multiple agents (representing different types of cells or molecules) that interact with each other probabilistically. Each interaction phase alternates with a diffusion phase that allows the agents to move across the lattice. As part of the ImmunoGrid project, various refinements have been undertaken, including ([Pappalardo *et al.* 2008a](#)): physical models of tumour growth based on nutrient or oxygen starvation (using the lattice Boltzmann method ([He & Luo 1997](#))); three-dimensional lattice models of lymph nodes ([Baldazzi *et al.* 2009](#)); and a model of chemotaxis that uses partial differential equations (PDEs).

Such an agent-based approach has several potential advantages. By directly representing individual agents, we are able to model the life history of individual cells, rather than merely their average behaviour. LGCA simulations are inherently stochastic, in terms of both the initial state of the simulator and the interactions that take place at a given time step. Hence, it is possible to model the distribution of behaviours within a population of identical individuals, or to assess the multiple possible outcomes for a single individual given a particular infection or disease scenario. For non-specialists, the behaviour of an LGCA simulation is also much more intuitive than that of an equation-based model. Finally, LGCA simulators are easily extensible; new cell types or new states for an existing cell type can easily be added. This is particularly appealing in the context of immune system simulation, where the critical issue is what level of detail is necessary for the simulator to exhibit realistic behaviour.

C-IMMSIM has been used to model HIV infection (Baldazzi *et al.* 2006; Castiglione *et al.* 2007b), Epstein–Barr virus infection (Castiglione *et al.* 2007a), and cancer immunotherapy (Castiglione *et al.* 2005). SIMTRIPLEX has been used to model immunoprevention vaccines (Motta *et al.* 2005; Pappalardo *et al.* 2005, 2006; Lollini *et al.* 2006) and atherosclerosis (Pappalardo *et al.* 2008a). Rather than describe the work covered in the preceding papers, here we will consider the modelling and computational challenges that we have faced.

The ultimate goal of ImmunoGrid is to develop a natural-scale model of the immune system. By *natural scale* we mean a model that captures both the diversity of the immunological (cellular and molecular) repertoire and the scale of the natural immune system, most notably in terms of the true population of its cellular components. Capturing the diversity of the immunological repertoire is necessary both to study the immune response of diseases such as influenza A that are subject to continuous mutation, and to understand the diversity of responses among different individuals within a population. Modelling the scale of the real immune system may be necessary to capture some of its emergent properties. In a recent paper, Chavali *et al.* (2008) suggest that the ability of agent-based models and cellular automata to display emergent behaviour makes them particularly appropriate for modelling immunological processes. Developing a model that captures both the diversity and scale of the immune system has not, as far as we are aware, been attempted before.

4. Descriptive versus predictive models

There is a fundamental distinction between descriptive models and predictive models. The descriptive model of a system is designed to characterize its nature in a way that is consistent with its known behaviour, whereas the predictive model aims to predict its future behaviour. Intuitively, it is much easier to design a descriptive model than a predictive one. Major obstacles to the development of both types of model include lack of available data about the components of the system and how they behave over time, and lack of knowledge of how the various components of the system interact; the bar is higher for predictive models than descriptive ones. Different models, both descriptive and predictive, require different quantities of data. How much data is required depends in part on how sensitive the system is to its state at a given time. Indeed, for chaotic

systems there are theoretical limits to what may be predicted however much data is available. For example, accurate weather forecasting beyond a relatively modest limit of a few weeks is considered impossible in theory owing to the chaotic nature of the Earth's atmosphere (Sneyers 1997).

The extent to which the immune system is predictable remains an open question. What is clear is that the development of predictive models of the immune system is exceedingly challenging given the relative paucity of some types of available data. Given the combinatorial nature of the human immune system, large number of cell types, and variability of pathogens, experimentally obtained immunological data represent only a tiny fraction of possible situations. To date, the ImmunoGrid Consortium has developed descriptive models of HIV infection, Epstein–Barr virus infection and cancer immunotherapy using the generic C-IMMSIM simulator; and predictive models of an immunopreventive vaccine using the SIMTRIPLEX simulator. It is the latter that we will discuss here, as it gives us insights into the challenges faced by the VPH initiative regarding the development of predictive models that will ultimately be useful in the treatment of human patients.

The SIMTRIPLEX simulator was developed to predict the effects of different vaccination schedules for the Triplex vaccine, which had previously been shown to prevent the onset of highly aggressive mammary carcinomas in HER-2/neu transgenic mice when applied chronically (one intraperitoneal vaccination every 3–4 days for two weeks followed by two weeks of rest starting at six weeks of age and continuing for the entire duration of the experiment) (Lollini *et al.* 2005). Given a range of vaccination schedules to evaluate, the aim was to use the simulator to identify one that required significantly fewer injections than the chronic schedule but which nevertheless produced a high survival rate (Lollini *et al.* 2006; Pappalardo *et al.* 2006). The behaviour predicted by the SIMTRIPLEX simulator has recently been validated experimentally by the Lollini group (one of the ImmunoGrid partners) using a small population of mice.

Evaluation of the outcome of the *in vivo* experiment is still ongoing, although a significant level of agreement between predictions and experiment has been demonstrated for the first 52 weeks. However, there are several important characteristics of this approach that we believe have wider implications:

- *The use of a model organism.* SIMTRIPLEX simulates the immune system of mice, rather than humans. Although the development of human simulators is our ultimate goal, the kind of validation experiment that was undertaken would not have been possible with human subjects, as it is currently impracticable (as well as perhaps unethical) to validate a simulator by testing vaccine schedules on large populations of cancer patients. Furthermore, animal (particularly rodent) models are regularly used in vaccine design, and immune protection in appropriate animal models has been used as a proxy for a human immune response in preparing a dataset of vaccine antigens for testing computational models (Mayers *et al.* 2003). During the development phase, the use of a model organism has been invaluable. Yet, there remains a crucial issue: How relevant are results from mouse experiments to the future development of human therapies? This is a difficult question to answer and one that is likely to remain controversial for a long time.

- *The limited feedback provided by small-scale experiments.* The amount and range of data collected during the *in vivo* experiment, though extremely valuable, were relatively modest. This is because it is very costly and time-consuming to run large numbers of such experiments with large populations of mice. In order to provide measurements of the state of the immune system that are both detailed and accurate, it would inevitably be necessary for many mice to be killed before the end of the experiment. Hence *in vivo* experiments that aim to validate *in silico* models are often expensive and difficult to design. One further consequence of the experiment's modest scale is that the amount of noise in the data generated is relatively high, which inevitably makes the subsequent assessment of simulator accuracy less precise.
- *The time-consuming nature of in vivo experiments.* *In vivo* experiments typically take much longer to run than *in silico* ones; indeed, this is one motivation for the development of accurate *in silico* models. In the case of the SIMTRIPLEX validation experiment, this ran for over a year, and significant additional time was required to plan and set up the experiment. An obvious consequence is that the iterative loop of model refinement is a long one.

Bearing these factors in mind, it is clear that the refinement of our models is a long-term enterprise, which will require a combination of modelling, experimental validation and model refinement in incremental fashion.

Another important point to note is that the SIMTRIPLEX simulator is designed to address a single disease scenario, with specific disease-related enhancements to the C-IMMSIM simulator upon which it was based. These include the addition of new cell types (cancer cells and vaccine cells) and molecules (tumour-associated antigens, interleukin 12 (IL-12) and allogenic MHC-I (alloMHC-I)) (Motta *et al.* 2005; Pappalardo *et al.* 2005). Ultimately, the reason why ImmunoGrid is not attempting to develop a single, generic simulator is a pragmatic one and relates to the computational issues addressed below. Different disease-related scenarios require contrasting aspects of the immune system to be modelled in greater or lesser detail. (For example, modelling tumour growth is integral to our work with the SIMTRIPLEX simulator, but irrelevant in most other contexts.) Hence, the scope for simplification is inevitably higher if a model addresses a single disease rather than all aspects of the immune system simultaneously.

5. Computational challenges and solutions

From a computational perspective, the ImmunoGrid simulators require significant and increasing resources. Indeed, the availability of computational resources is arguably as significant a limiting factor on what we are able to achieve as the lack of accurate data about the state of the immune system under a significant range of conditions. A rough calculation of the potential requirements of a notional 'full-complexity model' of the human immune system makes it clear that such a model is infeasible to run, both currently and for the foreseeable future. Take, for example, the simulation of interactions between peptides and MHC molecules; using NAMD (NANoscale Molecular Dynamics) ABF (Adaptive

Biasing Force) software (Darve & Pohorille 2001), it takes approximately 2 h to simulate a single peptide–MHC interaction (which in reality lasts for fractions of a second) using an eight-node computer cluster; yet, millions of such interactions are occurring within a single individual at any given time. In contrast, a single prediction of MHC binding from peptide sequence takes a fraction of a second. Consequently, gross simplification is an inevitable characteristic of any current immune system simulator.

In this section, we address two points: the pragmatic choices we have made about simplifying our model in order to limit the computational costs of running our simulators; and the Grid framework we have developed to maximize the resources available to run them.

(a) *Minimizing the computational costs of our model*

The single most important contribution to reducing the computational costs of our simulators comes from simplification. Indeed, arguably this is one of the points of doing modelling; only in the worst-case scenario does a model have to be as complex as the reality it describes. Ultimately, we gain important insights by understanding what level of detail is essential for a model to behave in a realistic way (descriptive modelling) or to anticipate the behaviour of the real system with a high level of accuracy (predictive modelling).

Decisions about what to simplify are driven by a combination of necessity and intuition. The key simplifications that have been made in the ImmunoGrid simulators are: the discretization of time and space; the use of binary strings to represent peptides (thereby ignoring their three-dimensional structure); modelling the concentration of molecules (notably antigens, antibodies and cytokines), rather than representing them individually; not simulating processes that occur within cells (transport, cleavage, presentation); and ignoring certain entities completely (for example, not all the simulators model natural killer cells). With respect to time, space and the size of the molecular repertoire (as defined by the length of the binary strings), user-defined parameters determine the complexity of the simulator's underlying model during a given run. A hybrid approach in which agent-based cellular models were successfully combined with linear representations of molecular concentrations has been developed by Guo *et al.* (2008).

The extent to which the simplifications used in the ImmunoGrid simulators are consistent with effective descriptive and/or predictive models largely remains an open question. However, to date many gross behaviours of the human immune system have been successfully replicated (C-IMMSIM), and (as discussed above) there has been some success in predicting the development of mammary carcinoma in genetically susceptible mice treated according to different immunopreventive vaccination protocols (SIMTRIPLEX).

A second way in which we have reduced the computational cost of simulations at run-time is by precalculating key interactions between peptides and T cells using the tools developed at CBS (Larsen *et al.* 2005). This approach is feasible for many bacterial and viral infections, but is problematic for species such as HIV that are associated with long-term diseases and that rely on mutation as a way of circumventing the immune system. In the latter case, it is not obvious in advance what set of peptides will be involved in binding.

(b) Maximizing available resources

Several parallelized versions of the ImmunoGrid simulators have been produced, allowing them to run efficiently on multiprocessor machines and clusters. However, parallelization only addresses the computational requirements of individual simulations; when considering the broader requirements of the ImmunoGrid project, it is important to bear in mind that, in practice, we need to run very large numbers of simulations—thousands (at least) of simulations to scratch the surface of the simulator's parameter space during the development phase, and large numbers of simulations to examine how different individuals respond to a given clinical scenario. Broadly, we can define the computational requirements of ImmunoGrid as follows:

- To enable the most complex single simulations to be run, requiring access to a large cluster or supercomputer.
- To enable large sets of immune system simulations and epitope predictions to be carried out (to explore the parameter space of the simulators and to investigate the effects of clinical scenarios on multiple individuals).
- To support small-scale simulations, including runs of the ImmunoGrid educational simulators, for which standard workstations are sufficient.

As foreseen, when the project name was chosen, no single partner of ImmunoGrid could guarantee access to sufficient resources to meet these requirements. Hence, a Grid-based solution was a practical necessity. Neither could we guarantee uninterrupted access to one of the national or international production-quality Grids, such as the UK National Grid Service² (NGS). As a consequence, the development of our own Consortium Grid was the only practical solution.

A detailed description of our Grid-based solution is presented elsewhere (Halling-Brown *et al.* 2008). Here we focus on its main characteristics and their wider relevance for systems biology simulations.

- Our Grid maximized the range of resources that can be used, including: desktop personal computers; local clusters and supercomputers at a single institution; and national and international Grid services, including the UK National Grid Service, the European supercomputer Grid DEISA³ and the US TeraGrid.⁴
- Both developers and users are insulated from the complexities of the underlying middleware. This simplicity is vital, as individuals and organizations that have resources that could potentially be incorporated into the Grid will be deterred from doing so unless the addition of a new Grid node is as easy as possible.
- Our framework also allows resources to be accessed as Web services. A Web service provides an Application Programming Interface (API) that enables users to integrate a remotely hosted service seamlessly with other components of the applications they are developing. This

²National Grid Service; <http://www.grid-support.ac.uk/>

³Distributed European Infrastructure for Supercomputing Applications; <http://www.deisa.eu/>

⁴TeraGrid; <http://www.teragrid.org/>

approach is becoming increasingly popular in the field of bioinformatics, with many core services provided by organizations such as the European Bioinformatics Institute⁵ already being made available as Web services. For ImmunoGrid, instances of our simulators can be wrapped as Web services, deployed on a local machine and accessed via the Grid framework.

- A Web interface is built on top of the upper middleware. This hides the underlying complexity from the user, who (given relevant permissions) can run multiple simulations on diverse computational resources at various widely distributed sites in a completely transparent way.
- Given a set of available resources linked by the ImmunoGrid framework, specific resources are selected automatically by a simple job broker (by default), or manually (if so desired by the user).

The hallmarks of our framework are its flexibility, ease of installation and ease of use. Ultimately, we believe that our solution represents an effective compromise for a single, large project. Many of the characteristics of ImmunoGrid are shared by other systems biology projects: the involvement of multiple international partners (each bringing their own computational resources to the project); the need to run large numbers of computations, both large and small; and the need to provide an easy-to-use interface for a relatively non-technical user base. From this perspective, our approach can be viewed as a case study that demonstrates the relevance and effectiveness of our flexible, robust but sub-optimal solution to a much wider range of biological projects.

But what about integrating multiple biological simulators, where inter-communication between the distributed components at run-time is a fundamental requirement? This is what is proposed by the VPH initiative, and it presents a different order of challenge altogether.

6. Towards an integrated VPH

The fundamental aim of the EU VPH initiative, as stated in the final version of the VPH Roadmap⁶ published in 2007, is to develop a ‘methodological and technological framework’ that will enable the development of quantitative models that are predictive and can describe human life from genes to whole organism.

A key motivation is the belief that, to develop effective therapies or preventive strategies that address complex pathologies, it is necessary to consider the human body as a single integrated system. The VPH Roadmap explicitly recognizes the ambitious, long-term goal of the initiative; it argues that, given sufficient resources, the framework can be developed ‘over the next 10 years’. However, it acknowledges that a total model of a human being is unrealizable technically and may be, in principle, unrealizable, as the only complete model is the organism itself.

Although we broadly agree with this assessment, the latter point is arguably somewhat over-stated; as we have already argued, it is only in the worst-case scenario that a model has to be as complex as the reality it describes. Rather, the

⁵EBI; <http://www.ebi.ac.uk>

⁶STEP Consortium. *Seeding the EuroPhysiome: a roadmap to the virtual physiological human* [online 5 July 2007], <http://www.europhysiome.org/roadmap>

aim the VPH Roadmap is to develop ‘a realistic logical structure likely to enable practical results within a reasonably short time scale and that will remain flexible and open to continual revision, extension and collaboration on a worldwide scale’.

With respect to predictive models, the aim of the VPH is to enable the interconnection of predictive models that may be on different scales in order to be able to test more systemic hypotheses. A key issue, then, is how separately developed simulators of individual subsystems (such as the immune system and the heart) can be integrated. The challenge here is substantial owing to the diversity of the different models currently being developed. This diversity can be characterized in several ways:

- *Diversity of modelling paradigms.* The VPH Roadmap lists a range of simulation techniques that are (or may be) used to simulate different subsystems of the human body, including: ODEs and PDEs; discrete methods (such as cellular automata); hybrid methods (such as the combination of LGCA and PDEs used in the ImmunoGrid simulators); hierarchical models (with different levels of representation for global and local behaviours); and biostatistical models (such as those combining pharmacokinetics and pharmacodynamics).
- *Diversity of concepts and nomenclature.* Many biological subsystems have their own specialist concepts and nomenclature. For example, key molecules of the immune adaptive system (such as Igs, TCRs and the MHC) have distinctive sequence and structural characteristics not shared by other biomolecules. These characteristics can be described using IMGT-ONTOLOGY (Giudicelli & Lefranc 1999; Lefranc *et al.* 2004; Duroux *et al.* 2008) and captured in IMGT, the international ImMunoGeneTics information system developed by Lefranc *et al.* (2008, 2009), who are key partners in the ImmunoGrid Consortium. The concepts and nomenclature of IMGT-ONTOLOGY have been widely adopted and form the basis of international standards (such as those for monoclonal antibody definitions developed by the WHO International Nonproprietary Names (INN) Programme). However, only a limited mapping of IMGT-ONTOLOGY concepts to more generalized ontological resources such as the Gene Ontology (GO) (Gene Ontology Consortium 2006) has been possible. Given that manual mappings between specialized ontologies require a very significant input from domain experts, much current effort is being put into the development of automated mapping systems, although the coverage and accuracy of such systems when applied to biomedical ontologies is often rather poor—for the results of a recent evaluation see Euzenat *et al.* (2007).
- *Diversity of resolutions.* Different models are being developed with contrasting spatial and temporal resolutions. For example, not all subsystems require the same level of granularity with respect to the modelling of cells—indeed, for some subsystems it may not be necessary to model individual cells at all.

Some of this diversity is both inevitable and desirable. Different subsystems of the human physiome are fundamentally different in terms of their structure and function, and it seems reasonable that each should be modelled on its own merits.

It has already been argued that a model should be as simple as is consistent with the aim of descriptive or predictive accuracy.

However, this diversity presents the VPH with one of its most daunting challenges—how to ensure effective intercommunication between contrasting models. The VPH Roadmap argues that this will ‘necessitate the development of software tools to facilitate model coupling’, but this work is as yet in its infancy. To make this work tractable, a loose coupling between components is arguably the only practical option. A given pair of components would communicate with each other via a well-defined interface that hides their respective modes of implementation. Given the diversity of models (outlined above) and the apparent lack of adherence to data and model interchange standards, even this apparently tractable approach to model integration poses significant challenges. It is worth noting here that, although the ImmunoGrid simulators do adhere to the IMGT data standards mentioned above, they do not currently employ any of the mark-up languages, such as CellML (Lloyd *et al.* 2004) and SBML (Hucka *et al.* 2003), advocated by the VPH Roadmap. In the long term, use of common mark-up, as well as common and widely accepted nomenclatures and ontologies, may be an essential prerequisite for the smooth interchange of data between models, however loosely coupled. Bauer *et al.* (2009) cite the potential heterogeneity of description and coding as one of the disadvantages of agent-based models, suggesting that it may be particularly important to maintain common mark-up and standards where this type of modelling is involved. Moreover, the computational resources that will be necessary to ensure effective intercommunication between distributed subsystem simulations at run-time are as yet unclear, but represent a potentially huge challenge.

But perhaps a more fundamental question is whether this loosely coupled approach is feasible in practice. Arguably, one of the key rationales for the development of the VPH, as stated in the VPH Roadmap, suggests that we should not expect the loosely coupled approach to take us very far because living things themselves are closely coupled.

To briefly explore this point further, let us consider a single scenario, where an immune system simulator is coupled with a model of the heart⁷ (Bassingthwaight *et al.* 2009; Niederer *et al.* 2009)—arguably, the most successful model in the whole of human systems biology—in order to simulate how an infection may affect the behaviour of this organ. In this scenario, the heart modellers would at the very least wish to know what concentration of antigens, antibodies and lymphocytes enter the heart (in fact, they may wish to have additional information about how these entities will interact with the cells in the heart). For the immune system simulator to be able to predict the concentration of those entities it would need to incorporate not only a model of lymphoid tissue (where lymphocytes are activated and antibodies created, and which is conventionally regarded as being within the remit of immune system modellers), but also a model of the cardiovascular system (or integrated but separate models of the heart and vascular systems).

This suggests, at the very least, a complex interplay between two or more models, with a combined ‘super-model’ that is significantly more complex than any of its individual components and that would require separate validation. But the situation is, in fact, more challenging than this description suggests. In

⁷The Wellcome Trust Heart Physiome Project; <http://heart.physiomeproject.org/index.html>

this scenario, the cardiovascular system has become part of the spatial context in which the immune system entities need to interact; the immune system and cardiovascular systems are not loosely coupled. In one sense, also, the innate and adaptive immune systems can be seen as being distinctive but tightly coupled, and, therefore, a complete immune system model being itself a ‘super-model’ requiring separate validation. Techniques developed for coupling models of different immune components should prove invaluable in the more challenging task of coupling an immune system model to models of other organs, including the heart.

7. Conclusion

The extent to which it will be possible for the VPH and related initiatives to meet the challenges discussed in this paper within the next decade remains an open question, but clearly there is a long way to go. In many respects, the modelling and computational challenges that are being faced by the VPH are exemplified by the immune system, which through its multiple subsystems and its complexities of scale and time encapsulates some of the most daunting obstacles within a single project. The ImmunoGrid project has made some important advances, but it is also making an important contribution by providing a reality check for some of the less realistic expectations of the wider systems biology community. Models of the immune system represent a convenient benchmark because of the nature of immunity: the immune system interacts intimately with all other systems in the human body. To its credit, the VPH Roadmap, although a visionary document, does not shy away from the scale of challenges that will be faced in the years to come.

So a final question may be posed: Given the scale of the complexities we are facing, when will we be able to provide a satisfactory solution? Again, the VPH Roadmap makes the important point that modelling efforts should be focused on what can have immediate impact and this will often require simplified models.

The need to develop effective educational tools is a recurrent theme in the VPH Roadmap, and the development of educational simulators and tools is an important role of the ImmunoGrid project. At the time of writing, four educational simulators can be run from the ImmunoGrid educational portal⁸ covering the topics of cancer vaccine scheduling, antigen processing and presentation, bacterial replication rates and atherogenesis. It is planned to expand this work in the near future to cover the analysis of T cell epitopes and immunological hot-spots, tumour growth and tumour regression. Each of the simulator-based educational tools is deliberately simplified, with a strong emphasis on visualization. For example, graphs are used to track the population of core cellular and molecular entities (T cells, B cells, cancer cells, antibodies, etc.) over time for each predefined clinical scenario. The graphs are updated in real time as the simulator is run. Movies showing the growth of cancer cells treated with the vaccination protocols simulated in SIMTRIPLEX are also made available as relatively small (approx. 1 Mb) and tractable .mpeg files. In this context, the available settings of the simulator are deliberately limited to ensure that the computational requirements are comparatively modest (thereby

⁸ImmunoGrid educational portal; <http://igrid-ext.cryst.bbk.ac.uk/immunogrid/site/education.php>

allowing us to offer open public access to this resource), and the clinical outcomes provide reasonable estimates in terms of, for example, the relative longevity of different patients.

Ultimately, the immune system exemplifies within a single biological subsystem many of the wider challenges that will be faced by the VPH over the next decade. We believe that immune system modelling provides researchers with the opportunity to explore and address many of the difficulties, both conceptual and practical, that will be critical to the future of integrated, multiscale modelling of the human physiology. The findings from the ImmunoGrid project, however partial, are therefore likely to feed into more extensive, precise or complete models of human organs and systems during the VPH project lifetime. We conclude this paper with a final positive note. For years computational power has been doubling every 18 months, and our ability to make use of these resources is growing steadily (albeit at a slower rate). This growth is our best guarantee that a complete virtual model of human physiology and its constituent systems, such as the immune system, are a not-too-distant reality.

The authors of this paper belong to the ImmunoGrid Consortium. The ImmunoGrid project has been funded by the EC contract FP6-2004-IST-4, No. 028069.

References

- Baldazzi, V., Castiglione, F. & Bernaschi, M. 2006 An enhanced agent based model of the immune system response. *Cell Immunol.* **244**, 77–79. (doi:10.1016/j.cellimm.2006.12.006)
- Baldazzi, V., Paci, P., Bernaschi, M. & Castiglione, F. 2009 Modeling lymphocytes homing and encounters in lymph nodes. *BMC Bioinform.* **10**, 387. (doi:10.1186/1471-2105-10-387)
- Bassingthwaighe, J., Hunter, P. & Noble, D. 2009 The Cardiac Physiome: perspectives for the future. *Exp. Physiol.* **94**, 597–605. (doi:10.1113/expphysiol.2008.044099)
- Bauer, A. L., Beauchemin, C. A. A. & Perelson, A. S. 2009 Agent-based modelling of host–pathogen systems: the successes and challenges. *Inform. Sci.* **179**, 1379–1389. (doi:10.1016/j.ins.2008.11.012)
- Castiglione, F., Bernaschi, M. & Succi, S. 1997 Simulating the immune response on a distributed parallel computer. *Int. J. Mod. Phys. C* **8**, 527–545. (doi:10.1142/S0129183197000424)
- Castiglione, F., Poccia, F., D’Offizi, G. & Bernaschi, M. 2004 Mutation, fitness, viral diversity, and predictive markers of disease progression in a computational model of HIV type 1 infection. *AIDS Res. Hum. Retrov.* **20**, 1314–1323. (doi:10.1089/aid.2004.20.1314)
- Castiglione, F., Toschi, F., Bernaschi, M., Succi, S., Benedetti, R., Falini, B. & Liso, A. 2005 Computational modeling of the immune response to tumor antigens: implications for vaccination. *J. Theor. Biol.* **237**, 390–400. (doi:10.1016/j.jtbi.2005.04.024)
- Castiglione, F., Duca, K., Jarrah, A., Laubenbacher, R., Hochberg, D. & Thorley-Lawson, D. 2007a Simulating Epstein–Barr virus infection with C-ImmSim. *Bioinformatics* **23**, 1371–1377. (doi:10.1093/bioinformatics/btm044)
- Castiglione, F., Pappalardo, F., Bernaschi, M. & Motta, S. 2007b Optimization of HAART with genetic algorithms and agent-based models of HIV infection. *Bioinformatics* **23**, 3350–3355. (doi:10.1093/bioinformatics/btm408)
- Celada, F. & Seiden, P. E. 1992 A computer model of cellular interaction in the immune system. *Immunol. Today* **13**, 56–62. (doi:10.1016/0167-5699(92)90135-T)
- Chavali, A. K., Gianchandani, E. P., Tung, K. S., Lawrence, M. B., Peirce, S. M. & Papin, J. A. 2008 Characterizing emergent properties of immunological systems with multi-cellular rule-based computational modelling. *Trends Immunol.* **29**, 589–599. (doi:10.1016/j.it.2008.08.006)
- Darve, E. & Pohorille, A. 2001 Calculating free energies using average force. *J. Chem. Phys.* **115**, 9169–9183. (doi:10.1063/1.1410978)

- Duroux, P., Kaas, Q., Brochet, X., Lane, J., Ginestoux, C., Lefranc, M.-P. & Giudicelli, V. 2008 IMGT-Kaleidoscope, the formal IMGT-ONTOLOGY paradigm. *Biochimie* **90**, 570–583. (doi:10.1016/j.biochi.2007.09.003)
- Euzenat, J., Isaac, A., Meilicke, C., Shvaiko, P., Stuckenschmidt, H., Šváb, O., Svátek, V., van Hage, W. R. & Yatskevich, M. 2007 Results of the ontology alignment evaluation initiative. In *Proc. 2nd ISWC Int. Workshop on Ontology Matching (OM), Busan, South Korea, November 2007* (eds P. Shvaiko, J. Euzenat, F. Giunchiglia & B. He), pp. 96–132.
- Fenner, J. W. et al. 2008 The EuroPhysiome, STEP and a roadmap for the virtual physiological human. *Phil. Trans. R. Soc. A* **366**, 2979–2999. (doi:10.1098/rsta.2008.0089)
- Gene Ontology Consortium. 2006 The Gene Ontology (GO) project in 2006. *Nucleic Acids Res.* **34**, D322–D326. (doi:10.1093/nar/gkj021)
- Giudicelli, V. & Lefranc, M.-P. 1999 Ontology for immunogenetics: IMGT-ONTOLOGY. *Bioinformatics* **15**, 1047–1054. (doi:10.1093/bioinformatics/15.12.1047)
- Guo, Z., Sloot, P. M. A. & Tay, J. C. 2008 A hybrid agent-based approach for modelling microbiological systems. *J. Theor. Biol.* **255**, 163–175. (doi:10.1016/j.jtbi.2008.08.008)
- Halling-Brown, M. D., Moss, D. S. & Shepherd, A. J. 2008 Towards a lightweight generic computational grid framework for biological research. *BMC Bioinform.* **9**, 407. (doi:10.1186/1471-2105-9-407)
- Handel, A., Longini Jr, I. M. & Antia, R. 2009 Antiviral resistance and the control of pandemic influenza: the roles of stochasticity, evolution and model details. *J. Theor. Biol.* **256**, 117–125. (doi:10.1016/j.jtbi.2008.09.021)
- He, X. & Luo, L. 1997 Theory of the lattice Boltzmann method: from the Boltzmann equation to the lattice Boltzmann equation. *Phys. Rev. E* **56**, 6811–6817. (doi:10.1103/PhysRevE.56.6811)
- Hucka, M. et al. 2003 The Systems Biology Markup Language (SBML): a medium for representation and exchange of biochemical network models. *Bioinformatics* **19**, 524–531. (doi:10.1093/bioinformatics/btg015)
- Larsen, M. V., Lundegaard, C., Lamberth, K., Buus, S., Brunak, S., Lund, O. & Nielsen, M. 2005 An integrative approach to CTL epitope prediction: a combined algorithm integrating MHC class I binding, TAP transport. *Eur. J. Immunol.* **35**, 2295–2303. (doi:10.1002/eji.200425811)
- Lefranc, M.-P. et al. 2004 IMGT-ONTOLOGY for immunogenetics and immunoinformatics. *In silico Biol.* **4**, 17–29.
- Lefranc, M.-P., Giudicelli, V., Regnier, L. & Duroux, P. 2008 IMGT, a system and an ontology that bridge biological and computational spheres in bioinformatics. *Brief. Bioinform.* **9**, 263–275. (doi:10.1093/bib/bbn014)
- Lefranc, M.-P. et al. 2009 IMGT, the international ImMunoGeneTics information system. *Nucleic Acids Res.* **37**, D1006–D1012. (doi:10.1093/nar/gkn838)
- Lin, H. H., Ray, S., Tongchusak, S., Reinherz, E. L. & Brusic, V. 2008 Evaluation of MHC class I peptide binding prediction servers: applications for vaccine research. *BMC Immunol.* **9**(Suppl. 12), S22. (doi:10.1186/1471-2172-9-8)
- Lloyd, C., Halstead, M. & Nielsen, P. 2004 CellML: its future, present and past. *Prog. Biophys. Mol. Biol.* **85**, 433–450. (doi:10.1016/j.pbiomolbio.2004.01.004)
- Lollini, P. L., De Giovanni, C., Pannellini, T., Cavallo, F., Forni, G. & Nanni, P. 2005 Cancer immunoprevention. *Future Oncol.* **1**, 57–66. (doi:10.1517/14796694.1.1.57)
- Lollini, P. L., Motta, S. & Pappalardo, F. 2006 Discovery of cancer vaccination protocols with a genetic algorithm driving an agent based simulator. *BMC Bioinform.* **7**, 352. (doi:10.1186/1471-2105-7-352)
- Louzoun, Y. 2007 The evolution of mathematical immunology. *Immunol. Rev.* **216**, 9–20. (doi:10.1111/j.1600-065X.2006.00495.x)
- Lundegaard, C., Lund, O., Kecmir, C., Brunak, S. & Nielsen, M. 2007 Modeling the adaptive immune system: predictions and simulations. *Bioinformatics* **23**, 3265–3275. (doi:10.1093/bioinformatics/btm471)
- Mayers, C., Duffield, M., Rowe, S., Miller, J., Lingard, B., Hayward, S. & Titball, R. W. 2003 Analysis of known bacterial protein vaccine antigens reveals biased physical properties and amino acid composition. *Compar. Funct. Genom.* **4**, 468–478. (doi:10.1002/cfg.319)

- Motta, S., Castiglione, F., Lollini, P. & Pappalardo, F. 2005 Modelling vaccination schedules for a cancer immunoprevention vaccine. *Immunome Res.* **1**, 5. (doi:10.1186/1745-7580-1-5)
- Murphy, K., Travers, P. & Walport, M. 2008 *Janeway's immunobiology*. New York, NY: Garland.
- Niederer, S. A., Fink, M., Noble, D. & Smith, N. P. 2009 A meta-analysis of cardiac electrophysiology computational models. *Exp. Physiol.* **94**, 486–495. (doi:10.1113/expphysiol.2008.044610)
- Pappalardo, F., Lollini, P. L., Castiglione, F. & Motta, S. 2005 Modeling and simulation of cancer immunoprevention vaccine. *Bioinformatics* **21**, 2891–2897. (doi:10.1093/bioinformatics/bti426)
- Pappalardo, F., Motta, S., Lollini, P. L. & Mastriani E. 2006 Analysis of vaccine's schedules using models. *Cell. Immunol.* **244**, 137–140. (doi:10.1016/j.cellimm.2007.03.002)
- Pappalardo, F., Musumeci, S. & Motta, S. 2008a Modeling immune system control of atherogenesis. *Bioinformatics* **24**, 1715–1721. (doi:10.1093/bioinformatics/btn306)
- Pappalardo, F., Zhang, P., Halling-Brown, M., Basford, K., Scalia, A., Shepherd, A., Moss, D., Motta, S. & Brusica, V. 2008b Computational simulations of the immune system for personalized medicine: state of the art and challenges. *Curr. Pharmacogenom. Pers. Med.* **6**, 260–271. (doi:10.2174/187569208786733839)
- Pappalardo, F. *et al.* 2009 ImmunoGrid, an integrative environment for large-scale simulation of the immune system for vaccine discovery, design and optimization. *Brief Bioinform.* **10**, 330–340. (doi:10.1093/bib/bbp014)
- Pappalardo F., Pennisi M., Castiglione F. & Motta S. 2010 Vaccine protocols optimization: *in silico* experiences. *Biotechnol. Adv.* **28**, 82–93. (doi:10.1016/j.biotechadv.2009.10.001)
- Sette, A., Fleri, W., Peters, B., Sathiamurthy, M., Bui, H. H. & Wilson, S. 2005 A roadmap for the immunomics of category A–C pathogens. *Immunity* **22**, 155–161. (doi:10.1016/j.immuni.2005.01.009)
- Smith, D. J., Forrest, S., Ackley, D. H. & Perelson, A. S. 1998 Using lazy evaluation to simulate realistic-size repertoires in models of the immune system. *Bull. Math. Biol.* **60**, 647–658. (doi:10.1006/bulm.1997.0035)
- Sneyers, R. 1997 Climate chaotic instability: statistical determination and theoretical background. *Environmetrics* **8**, 517–532. (doi:10.1002/(SICI)1099-095X(199709/10)8:5%3C517::AID-ENV267%3E3.0.CO;2-L)