

ImmunoGrid, an integrative environment for large-scale simulation of the immune system for vaccine discovery, design and optimization

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

Submitted: 23rd December 2008; Received (in revised form): 17th February 2009

Corresponding author. Vladimir Brusic, Cancer Vaccine Center, Dana-Farber Cancer Institute, 77 Avenue Louis Pasteur, HIM 401, Boston, MA 02115, USA. Tel: +1 617 632 3824; Fax: +1 617 632 3351; E-mail: vladimir_brusic@dfci.harvard.edu

Francesco Pappalardo earned his MSc and PhD degrees from University of Catania. He is a postdoctoral fellow and has a temporary professorial appointment in Mathematics at the University of Catania. His major research area is on modeling immune system responses involved in several pathologies, including viral infections, tumors, multiple sclerosis and atherosclerosis.

Mark D. Halling-Brown has research interests in the areas of Bioinformatics, Computational Immunology, Chemogenomics and Grid Computing. Originally trained as a computer scientist, he holds a masters and PhD in Bioinformatics and is currently involved with developing techniques for improvement of drug discovery for cancer.

Nicolas Rapin holds a masters degree in Food Engineering and Computational Molecular Biology, and a PhD in Bioinformatics and Bio-simulation. His research interests focus on merging computational modeling and machine learning methods to explain experimental data in Biology. He is now a postdoctoral fellow in the Pharmaceutical Institute at the University of Copenhagen, Denmark.

Ping Zhang has a masters and a PhD in Information Technology and 12 years work experience in Electronics. Her current research interests include computational and statistical modeling, pattern recognition and Bioinformatics. She works as a Bioinformatician at the Diamantina Institute for Cancer, Immunology and Metabolic Medicine at the University of Queensland, Australia.

Davide Alemani graduated from the University of Milan, Italy, in 2001 with a masters degree in Physics. He obtained a PhD in Computational Science at the Department of Chemistry, University of Geneva in 2007. He received a scholarship by the Swiss National Foundation to spend a year at the University of Queensland, Australia, working on computational modeling of tumor growth. He is currently pursuing a postdoctoral research in the field of molecular dynamics at the EPFL in Lausanne, Switzerland.

Andrew Emerson holds a degree and a PhD in Chemistry from the University of Southampton, UK. He has worked in several fields including Materials Science, Bioinformatics, Systems Biology and Molecular Modeling. His interests include the uses of grid and high-performance computing in Computational Science. He currently works as an application scientist at the CINECA Supercomputing Centre in Bologna, Italy.

Paola Paci was trained as a theoretical physicist and she worked over 10 years on the field of Computer Modeling. She holds a degree in Physics from the University of Rome 'La Sapienza' and a PhD in Physics from the University of Pavia 'A. Volta'. Her research interests include Physics applied to Medical Immunology. She holds a postdoctoral position at the Institute for Computing Application of the CNR in Rome.

Patrice Duroux holds a PhD in Informatics from University Montpellier 2, France. He works as a research engineer at CNRS and he is in charge of computer management, database and software developments at IMGT[®], the international ImMunoGeneTics information system[®] (<http://www.imgt.org>).

Marzio Pennisi graduated in Computer Science in March, 2006, and is pursuing PhD studies at the University of Catania. His project focuses on studying Modeling, Mathematical Methods and Computer Simulation of the immune system. In particular, he is working on models of mammary tumor growth, diffusion of metastasis and therapy. He uses evolutionary algorithms, and constrained and unconstrained optimization methods for simulation of the immune system.

Arianna Palladini is completing her PhD in Cellular and Molecular Biotechnology at the University of Bologna. She is a recipient of a fellowship of the Italian Foundation for Cancer Research. She has been working in the Laboratory of Immunology and Biology of Metastasis where she is involved in the preclinical development of cancer vaccines.

Olivo Miotto is a Senior Informatics Fellow at Oxford University. He is based in Bangkok, where he works on the Genomic Epidemiology of malaria parasites. His research interests focus on the computational study of pathogen diversity and the aggregation of knowledge for large-scale bioinformatics. He holds a BSc from Imperial College London, and is completing a PhD at the National University of Singapore.

Daniel Churchill is an assistant professor in the Division of Information and Technology Studies at the Faculty of Education, the University of Hong Kong. He specializes in the areas of educational technology, e-learning and multimedia design. His current research effort focuses on Web 2.0, mobile learning and learning objects design and application.

Elda Rossi has a degree in Chemistry from the University of Modena. Since 1984, she has been working at CINECA, where she has been involved in computational applications in fields related to Chemistry and Biology. Her expertise is in research support using high-performance computing tools and in writing and integrating scientific programs.

Adrian J. Shepherd has been doing Bioinformatics research for over 12 years. He has a PhD in Computer Science from University College London. His research interests include biomedical text mining, modeling uncertainty in biological data and developing tools that support distributed biomedical applications. He is currently senior lecturer in Computational Biology at Birkbeck College, University of London.

David S. Moss holds BSc and PhD degrees from the University of London. He was Head of the School of Crystallography at Birkbeck for 9 years and is currently Professor of Biomolecular Structure. His research interests include bacterial toxins, computational immunology and computational approaches to protein–ligand interactions.

Filippo Castiglione is a researcher at the Institute for Computing Applications of the National Research Council of Italy. He holds a bachelor degree in Computer Science from the University of Milan and a PhD in Scientific Computing from the University of Cologne. He was a visiting research scientist at the IBM Watson Research Center, the Molecular Biology Department of the Princeton University and the Department of Cell Biology of the Harvard Medical School. His main interests range from mathematical modeling to computational biology and bioinformatics.

Massimo Bernaschi spent 10 years with IBM working in the field of Parallel and Distributed Computing. Currently he works at the National Research Council of Italy (CNR) as the Chief Technology Officer of the Institute for Computing Applications. He is an adjunct professor of Computer Science in 'La Sapienza' University in Rome.

Marie-Paule Lefranc is a professor at the University Montpellier 2, senior member of the Institut Universitaire de France, Chair of Immunogenetics and Immunoinformatics, Head of the Laboratoire d'ImmunoGénétique Moléculaire (LIGM) at Institut de Génétique Humaine (IGH, CNRS). She works on molecular immunogenetics of immunoglobulins, T-cell receptors, antibody engineering and in immunoinformatics. She is the founder and director of IMGT[®], the international ImMunoGeneTics information system[®] (<http://www.imgt.org>), since 1989.

Søren Brunak holds a masters degree in Physics and a PhD in Computational Biology. He is the Professor of Disease Systems Biology at the Novo Nordisk Foundation Center for Protein Research, University of Copenhagen, and Professor of Bioinformatics at the Department of Systems Biology at the Technical University of Denmark. He is the founding Director of the Center for Biological Sequence Analysis, since 1993. Dr Brunak contributed new computational strategies for analysis of data in molecular biology, medicine and biotechnology. His current research interests focus on systems biology and data integration focused on disease phenotypes.

Santo Motta obtained 'Laurea' in Physics at the University of Catania in 1970 and MSc in Applied Mathematics at the University of London, Queen Mary College in 1971. He is an associate professor of Mathematical Physics at the Faculty of Pharmacy and he is a member of the Department of Mathematics and Computer Science of the University of Catania. He works on computational models of the cancer immune system competition induced by an *in vivo* tested immune prevention vaccine. Its present scientific interests are BioMaths, BioComputing and BioInformatics, Immunomics.

Pier-Luigi Lollini is the professor of General Pathology at the University of Bologna, Italy. He is a founding member of the Laboratory of Immunology and Biology of Metastasis. His research is focused on the prevention of non-infectious cancer by immunological means, in particular vaccines and biological immunomodulators.

Kaye E. Basford is the professor of Biometry and Head of the School of Land, Crop and Food Sciences at the University of Queensland. Her research is at the interface between statistics and quantitative genetics, particularly in the design and analysis of multi-environment plant breeding trials, the application of data mining techniques (such as pattern analysis) and bioinformatics.

Vladimir Brusic holds masters degrees in Biomedical Engineering, Information Technology and Business Administration; and a PhD in Bioinformatics. His research interests focus on applications of machine learning, statistical methods and computational modeling in immunology and vaccine research. He is the Director of Bioinformatics at Cancer Vaccine Center, Dana-Farber Cancer Institute, Boston, MA, USA.

Abstract

Vaccine research is a combinatorial science requiring computational analysis of vaccine components, formulations and optimization. We have developed a framework that combines computational tools for the study of immune function and vaccine development. This framework, named ImmunoGrid combines conceptual models of the immune system, models of antigen processing and presentation, system-level models of the immune system, Grid computing, and database technology to facilitate discovery, formulation and optimization of vaccines. ImmunoGrid modules share common conceptual models and ontologies. The ImmunoGrid portal offers access to educational simulators where previously defined cases can be displayed, and to research simulators that allow the development of new, or tuning of existing, computational models. The portal is accessible at igrid-ext.cryst.bbk.ac.uk/immunogrid.

Keywords: *computational modeling; Grid computing; immunoinformatics; simulations; vaccine discovery*

INTRODUCTION

Vaccination is an immunologic intervention that produced the largest single positive impact on control of infectious disease and saving of human lives over the last 200 years. Some 50 successful vaccines have been developed during this period to control almost 30 infectious diseases [1]. Two major classes of vaccinations exist: prophylactic vaccines to prevent future infections or diminish their effects, and therapeutic vaccines to treat established disease. Advances in vaccine development come in waves produced by technological advances. The historic technological breakthroughs in the vaccine field include pathogen attenuation, pathogen inactivation, cell culture of viruses, genetic engineering and induction of cellular immunity. The current biotechnological revolution brings combination vaccines, advanced adjuvants, genomics and proteomics, nanotechnology-based delivery systems, immunomodulation, rational vaccine design and computational simulations to improve our understanding of both the makings of human organism and immune system and of pathogens [2, 3]. These latest technological advances also offer great hope for immunologic control of noninfectious disease including cancer, autoimmunity, allergies and post-transplantation complications. These interventions are specially promising in cases where alternatives (such as surgery, chemotherapy or others) are limited, ineffective, or non-existent [4].

Unfortunately, for most diseases variation in host and pathogens make a universal vaccine a remote possibility [5]. Thanks to advances in genomics and sequencing of entire genomes of multiple strains of pathogenic microorganisms, the issue of pathogen diversity can be addressed by targeting pan-genomes or total gene repertoire for a given microorganism [6]. The diversity of the human immune system is enormous; the number of different products of the immune system is five orders of magnitude larger than some 10^6 non-immune system products encoded by the human genome. Vaccines must contain at minimum two antigenic epitopes: one to induce specific B cell or cytotoxic T-cell responses, and one to provide T-cell help [7]. Broadly protective vaccines, against multiple strains of pathogen require cocktails of target antigens or epitopes [8, 9]. In addition, immunomodulatory components (adjuvants and various suppressors and

enhancers) are added to enable desired immune responses. The development of combinations of adjuvant and immunomodulatory components requires considering the route and kinetics of immunization; the resulting combinatorial space of possible adjuvants is huge [10]. The efficiency of vaccines depends on multiple factors: components and formulation, the administration route [11], vaccine delivery systems [12, 13], the dose of vaccine [14], the number of vaccinations and vaccination schedules [15]. Vaccination induces a broad spectrum of regulatory cytokine responses [16]. The typical effect of vaccination is the promotion of protective immune responses; however vaccines are also known to induce immunosuppression [17]. Therefore, the combination of bioinformatics and high-throughput immunologic assays is essential for screening potential vaccine targets [4].

The large number of possible vaccine formulations makes computational and biomathematical approaches essential for projects aiming to perform systematic exploration and optimization of vaccine formulations. We have developed a framework that combines computational tools for the study of immune function and vaccine development with experimental validation. This framework, named ImmunoGrid combines conceptual models of the immune system, models of antigen processing and presentation, system-level immune system models, Grid computing, and database technology to facilitate vaccine discovery, formulation and optimization (Figure 1). Currently ImmunoGrid has selected simulators of infection, cancer and atherosclerosis, and allows simulation of antigen interactions in lymph nodes. ImmunoGrid is an example of collaborative environment in which well defined modeling and simulation tools are integrated with the expansive computational infrastructure. In future, ImmunoGrid will include models, simulators and tools that address various further aspects of the immune system. In principle there is no limit as to what models can be included. Thanks to Grid infrastructure these tools can be distributed across different physical sources while the portal serves only as the common interface. The criteria for inclusion of tools into the ImmunoGrid framework are the accuracy and the utility of the tools, and compatibility with the conceptual framework.

CONCEPTS, STANDARDS AND ONTOLOGIES

Standardized rules and formal concepts for the identification, description and classification of biological components and processes are used in the modeling of the Virtual Immune System (VIS) as a part of the Virtual Physiological Human (VPH). The VPH is a framework of methods and technologies that aim to enable the investigation of the human body as a single complex system [18]. Unique numbering and controlled vocabularies (ontologies) have been developed for the formal description of antigen receptors including antibodies or immunoglobulins (IG), T-cell receptors (TR) and the major histocompatibility complex (MHC). Human MHC is known as human leukocyte antigen (HLA). The formal definition is necessary for capturing the incredible diversity of antigen receptors amongst individuals (10^{12} IG, 10^{12} TR, or $>10^{20}$ possible combinations of HLA per person). These definitions (rules) are crucial for a standardized analysis of the interactions between receptors and ligands and between proteins and are used for molecular-level modes and system-level simulators. The concept definitions and ontologies are largely based on IMGT[®] [19, 20], which has become a widely accepted standard in immunology. IMGT focuses on antigen receptors and MHC at the molecular level and provides the standardization for immunogenetics data from genome, proteome, genetics and 3D structures [21–24]. The ImmunoGrid extensions, currently under development, provide the definitions of immune concepts at the cellular, organ and organism levels, and further provide formal definitions of various immune pathologies.

ImmunoGrid is a complex environment containing multiple development, application and dissemination components (Figure 1). Each of the components is connected directly, or indirectly, to the ‘Concepts’ and ‘Standards’ components, ensuring that a common conceptual model, terminology and standards are used.

MOLECULAR-LEVEL MODELS

A large body of mathematical and computational-modeling work has been done on modeling the adaptive immune system [25], but less so on modeling innate immunity. The adaptive immune

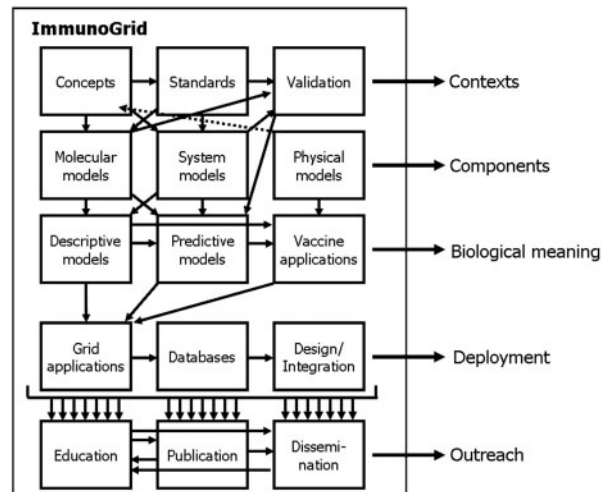


Figure 1: The components of ImmunoGrid and their relationships. There are five principal groups of components. The first three groups define the scientific background of the ImmunoGrid framework: conceptual background, individual components, models and applications. The fourth group defines the engineering background and the simulator of the immune system. The fifth group defines the means for reaching the broader community. The directed arrows in the figure show dependencies. For example concepts, standards and validation determine contexts. Concepts define standards, and standards are then used to define validations to be performed. Concepts and standards are used to define specific components, namely molecular and system models. Physical models such as tumor-susceptible mouse are external to ImmunoGrid and the observed experimental results are used to refine the concepts (the feedback arrow shown). Molecular- and system-level models can be descriptive or predictive and they are used for vaccine applications. These models and applications are implemented as Grid applications or databases and require integration of components. The results are then distributed to the researchers and the community.

system acts through its two arms: humoral immunity, mediated by antibodies and cellular immunity mediated by T cells. Antibodies, which are produced by B cells, neutralize pathogens and antigens outside the cells. T cells neutralize intracellular pathogens and cancers by eliminating infected or malfunctioning cells and they also provide regulatory help for both humoral and cellular immunity. The enormous repertoires of both B- and T-cell receptors (IG and TR) provide means for recognition, immune activation and acquiring of memory against

disease. B cell epitopes are 3D shapes on the antigen surface, mostly discontinuous parts of antigen sequences, recognized by antibodies. Molecular mechanisms involved in the IG synthesis result in a large diversity of B cell clones that are selected for specific antigens and mature into plasma cells producing highly-specific antibodies. T-cell epitopes are short fragments of antigenic peptides that are produced by antigen processing and presentation pathways. Cytotoxic CD8⁺ T-cells (CTLs) recognize and target cells that display foreign T-cell epitopes presented by class I pMHC complexes. Helper CD4⁺ T cells recognize class II pMHC complexes and provide regulatory signals (T-cell help) needed for activation of B cells and CTLs.

The combinatorial complexity of antigen processing and presentation makes the prediction and analysis of MHC-binding peptides and T-cell epitopes a problem suitable for large-scale screening and computational analysis. Vigorous research and development has produced a large number of computational models of antigen processing and presentation during the last 15 years. A large number of MHC ligands and T-cell epitopes have been stored in specialist databases [26–29]. A number of techniques, including motifs, quantitative matrices, artificial neural networks, hidden Markov models, support vector machines, molecular modeling and others have been developed [30, 31] and deployed as web servers. The antigenic epitope prediction, or ‘molecular-level models’, used in ImmunoGrid are principally the CBS tools [32–37], complemented by several other predictive models [38–40]. The predictive models and the modes of usage are judiciously selected using exhaustive validation with carefully selected experimental data. Because ImmunoGrid focuses on practical applicability to vaccine research, the models are applied in a manner that best suits vaccine discovery and the validation process (as shown in an example in Figure 2). Using this approach, we have analyzed 40 000 influenza proteins for peptides that bind to some 50 HLA alleles.

The testing and assessment of these models shows that the predictions of peptide binding to HLA class I molecules are of high accuracy and therefore directly applicable to identification of vaccine targets [41–43]. Furthermore, we have shown that these models are stable and their performance can be reproduced across several major HLA class I variants [44].

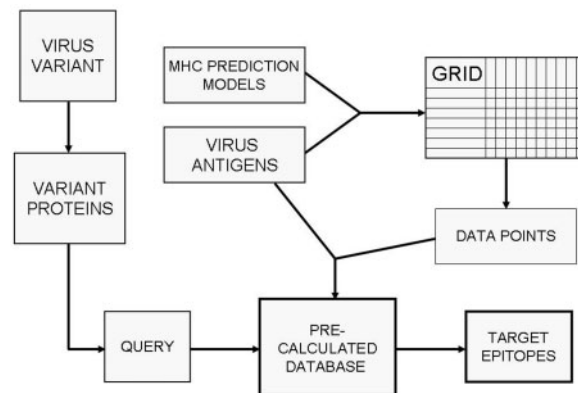


Figure 2: An example of integration of molecular-level models for discovery of target T-cell epitopes. A set of molecular models (multiple HLA variants) are used to analyze a complete set of viral antigens. Predictions are performed using Grid computing, to accommodate a large numbers of prediction jobs. Predicted vaccine targets are then stored in a relational database as annotated entries. Subsequent user requests for analysis of candidate T-cell epitopes from one, several, or all antigens can be performed by data-mining of the database of predicted targets.

Predictions of MHC-II class II ligands and T-cell epitopes are more complex: pMHC-II-binding predictions are of much lower accuracy than those of pMHC-I [45, 46]. The accuracy of predictions can be improved by using genetic algorithms [47], Gibbs sampling [48] or by combining predictions from multiple predictors [49]. For the time being, HLA class I predictions can be used for prediction of positives (binders and T-cell epitopes), while HLA class II predictions can be used for elimination of obvious negatives (non-binders, non-T-cell epitopes). Preliminary results from the combined analysis of computational prediction, that lead to selection of peptides used for T-cell-activation experiments, together with measurement of cytokine profiles from experimental validation studies of viral proteomes indicate that the concept of ‘T-cell epitope’ can not be linked to MHC binding only but rather to a larger panel of biological phenomena, including peptide processing, their availability in various tissues, T-cell repertoires, and others.

SYSTEM-LEVEL MODELS

Our knowledge of the immune system is incomplete and imprecise. Mathematical modeling provides a formal description of the underlying principles and organization of the immune system and the

relationships between its components. The immune system is a distributed system lacking central control, but it performs complex functions efficiently and effectively [50]. Mathematical modeling of the immune system requires understanding both the properties of types of mathematical models involved and understanding the source data used for modeling. These methodologies have strengths and limitations that need to be understood for effective development of mathematical models that use these data. For example the source data often need transformations (normalization, filtering, or other pre-processing) before it can be used for model development. Mathematical models of the immune system evolved from classical models using differential equations, difference equations, or cellular automata to model a small number of interactions, molecules, or cell types involved in immune responses. Principal types of mathematical models for immunological applications are described in [51, 52]. The key enabling technologies of genomics [53], proteomics [54, 55], bioinformatics [56] and systems biology (including such genomic pathway analysis) [57] in immunology have provided large quantities of data describing molecular profiles of both physiological and pathological states. Assays for immune monitoring (multiparametric flow cytometry, nanotechnology for quantitation of cytokine production, ELISPOT, intra-cytoplasmic cytokine staining and mRNA/micro-RNA-based assays) keep improving and expanding our ability to measure various immunoregulatory and modulatory molecules [58]. Laser scanning cytometry enables the measurement and analysis of effector function of individual cells *in situ* and thus quantification of molecular and cellular events in physiological and pathological states [59]. These emerging technologies provide data that previously were not available and that enable more detailed modeling of immune processes.

Mathematical modeling of the immune system has grown to include the extensive use of simulations [51] and iterative refinement of the immune system models at both molecular [60] and system level [61]. ImmunoGrid system-level models are based on the cellular automata model of the immune system originally developed by Celada and Seiden [62] and spatial physical models using partial differential equations to describe lymph nodes, chemotaxis, cell movement and diffusion. The two key models of the immune system are

C-ImmSim [63] and SimTriplex [15]. These models include components of both adaptive and innate immunity and the core function for both models is the modeling of adaptive immunity (both humoral and cellular arms are included). C-ImmSim is a generic model and it has been applied for the modeling of: primary and secondary immune responses, bacterial infection and viral infection. It has been applied to the descriptive studies of HIV infection [64, 65], Epstein-Barr virus infection [66] and cancer immunotherapy [67]. The first version of SimTriplex model was derived from C-ImmSim with a focus on predictive modeling, with the later versions evolving independently. The SimTriplex model has been applied to predictive modeling of immunoprevention vaccine [15, 61, 68, 69] and more recently to the descriptive modeling of atherosclerosis [70]. The most recent developments of models for inclusion in ImmunoGrid are physical models of tumor growth based on nutrient or oxygen starvation (based on the lattice Boltzmann method [71, 72]). These models have enabled the simulation of tumor growth for both benign and malignant tumors [manuscript in preparation]. ImmunoGrid also has a lymph node simulator offering a mechanistic view to the interactions of antigen and the immune system cells in the lymph node.

In ImmunoGrid, our main concern is the practical applicability of computational models to vaccine discovery, formulation, optimization and scheduling. The models must, therefore, be both biologically realistic as well as mathematically tractable [73]. The models used in ImmunoGrid are validated experimentally and incrementally refined (tuned). Specific applications that have been investigated include the study of mammary cancer immunoprevention; therapeutic approaches to melanoma and lung cancer; immunization with influenza epitopes; the study of HIV infection and HAART therapy; and modeling atherosclerosis.

In ImmunoGrid we use predictive models that are carefully validated, and are refined iteratively (predict-test-refine) using selected experiments. Modeling of immunoprevention vaccine against mammary carcinoma in genetically susceptible mice has been validated experimentally and shown to accurately reproduce immune responses for up to 52 weeks of age [15, 61, 68, 69] while additional model tuning is needed for later stages of the disease. The vaccination data indicate that for the older

(e.g. second year of mouse life) immune system, the rate of immune responses are lower than in the younger (e.g. first year of mouse life) immune system (P.L. Lollini, unpublished data) and, accordingly, the model needs additional tuning for older mice. Modeling of HIV-1 infection in untreated as well as patients receiving HAART has been tuned against data from literature as well as clinical observations [64, 65]. CBS tools have been assessed as best performing prediction systems for HLA binders and T-cell epitopes, specific details can be found in [43, 45].

The model for atherogenesis is descriptive and it was tuned to match published experimental data—the detailed descriptions are available in [70]. The descriptive simulator of a lymph node reproduces experimental data published in [74–76].

INTEGRATED SYSTEM AND GRID COMPUTING

Integrated systems for data management and laboratory automation are emerging as cyber infrastructure that is the new research environment that enables research using Internet resources [77, 78]. These systems mainly deal with the management and analysis of data addressing the bottleneck due to the huge and rapidly growing quantities of data and the need to automate the process of multi-step large-scale screening. Continuous developments of information and communication technologies and computational intelligence [79] have led to the concept of ‘Virtual Laboratory’. The Virtual Laboratory environment offers an integrated intelligent environment for systematic production of high-dimensional quality assured data, as opposed to the common approach where independent exploratory studies are typically combined in an *ad hoc* manner [80]. The ImmunoGrid environment provides a number of modules that can be suitably combined for addressing specific vaccine questions; these include Grid computing for large-scale tasks and sharing distributed data. Each vaccine study performed within the ImmunoGrid framework involves: conceptual modeling of the disease/pathology, followed by utilization of models for simulation of the immune system and processes. Descriptive models are tuned to the available data sets, while predictive models are developed incrementally, starting from simple solutions which are incrementally validated and expanded. The ImmunoGrid

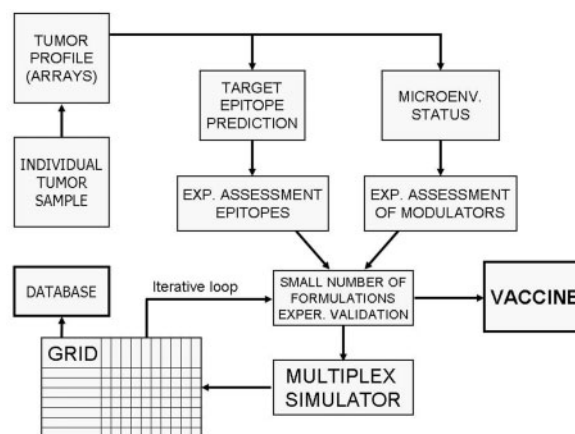


Figure 3: An example of integration of system-level models for discovery, formulation, or optimization of vaccines. First, a tumor sample is profiled for presence of tumor antigens and the profile of tumor microenvironment, including profiles of immunoregulatory molecules. T-cell epitopes are predicted using molecular-level models. A small number of experiments are followed by further computational simulations. The ‘Multiplex’ refers to the analysis of multiple components in the vaccines (SimTriplex contains three components). Massive scale simulations can be performed using Grid computing and the simulations can be stored in a database for future analysis. The iterative process of initial experiments → simulation → validation, ultimately leading to a vaccine can be used for formulation and optimization of formulation, dosage and scheduling. This approach has been used for optimization of immunoprevention vaccine scheduling using SimTriplex in mouse models of mammary cancer. Molecular-level models are linked with system-level models through predicted T-cell epitopes.

framework enables integration of molecular- and system-level models for vaccine research, an example is shown in Figure 3.

In ImmunoGrid, the meaning of the Grid is primarily that of a ‘virtual organization’, implying access to heterogeneous resources and administrative resources through a single point of access [81]. The main requirements for the ImmunoGrid simulator are processing power, data federation, visualization of results and user portals. The initial benchmarking of ImmunoGrid simulations showed that large-scale simulations of 1600 vaccine schedules in 100 virtual mice (representing 24 months of mouse life) can be accomplished within 26 h of wall time (time difference between job submission and receiving the results). Modeling mouse immune responses is a step towards modeling human immune system.

Overall, mouse system-level models are very similar to the human immune system, with an added advantage that experiments can be performed readily; they are normally used for vaccine research and also they are used as benchmarks for bioinformatics-driven vaccine research. ImmunoGrid development is based on the analysis of HLA as well as scaling-up of the mouse ImmunoGrid developments. In ImmunoGrid, Grid computing, therefore, enables robust and scalable solutions for immune system modeling.

Each model within ImmunoGrid simulator has integrated a set of user notes, accessible by clicking on the question marks. Also, tutorial notes are available on the web site.

EDUCATIONAL AND RESEARCH IMMUNOGRID PORTAL

The ImmunoGrid simulator can be accessed through the prototype portal web site [82]. The main aims of the ImmunoGrid simulator are to: provide tools to enhance our understanding of the immune system; support vaccine design, formulation and optimization; provide an interactive educational model of the human immune system; and apply computer modeling and Grid infrastructure to complex scientific applications. The modular architecture of ImmunoGrid enables content hierarchy and reusability of modules. Educational simulators are based on learning objects [83] and currently include the following simulation modules: cancer vaccine scheduling, antigen processing and presentation and bacterial replication rates. Learning objects enable students to select multiple scenarios and observe the effects of changing input to the outcome. This strategy promotes student-centered engagements where students learn through engagements such as querying and problem-solving. The learning objects provide information in highly visual and interactive formats supporting deep learning. Educational ImmunoGrid will be expanded to include the analysis of T-cell epitopes and immunological hot-spots, tumor growth, tumor regression and atherosclerosis, among others.

The access to the research portal is currently restricted to the ImmunoGrid consortium members and to selected projects. Projects of a very large scale, such as prediction of peptide binders for several hundred HLA alleles within 40 000 influenza sequences, include several steps: pre-calculation of

results using Grid infrastructure, storage of results in the database and access to the results by interrogating the database. Smaller-scale projects produce results by direct access to computational resources. Educational simulators are developed and implemented from representative examples from earlier projects. The ImmunoGrid portal is accessible at <igrid-ext.cryst.bbk.ac.uk/immunogrid>. At this time the educational simulator contains selected examples of simulations of cancer vaccine scheduling, cancer growth, viral infection (influenza, HIV and EBV), atherosclerosis simulator, a mechanistic lymph node model and antigen processing and presentation predictions.

CONCLUSIONS AND DISCUSSION

The ImmunoGrid simulator is a new generation of systems fitting into the framework of the VPH Initiative [84] enabling both descriptive and predictive modeling of the immune system in humans and in model animals. ImmunoGrid promotes a new, technology-driven, systematic approach to vaccine research. ImmunoGrid is neither a data management nor a laboratory automation framework—it is a distributed system containing databases and computational simulators of immune system and immune processes that enables vaccine development applications. ImmunoGrid enables combination of computational simulations and experimental validation with iterative improvement of *in silico* screening. It represents a new concept of integrated research environment where virtual experiments are tightly bound with laboratory experiments and clinical observation. The combination of these resources helps improve the economy of vaccine discovery and development through consideration of a large proportion of combinatorial space of vaccine targets—a quest not possible using traditional methods.

Two representative accomplishments show the utility of ImmunoGrid. First, ImmunoGrid is, to our knowledge, the first true Grid application on immune system modeling. Utilization of Grid infrastructure has enabled screening in parallel tens of thousands of viral sequence variants for thousands of HLA variants and peptide lengths that vary from 8–11mers (HLA class I) and 9–15mers (HLA class II). Such large-scale applications are now accessible as on-line resources. Second, the combination of computational modeling and experimentation of

immunoprevention vaccine in genetically susceptible mice provided a speed-up of the project and significant savings in search for optimal vaccination schedule. Hundreds of thousands of virtual experiments were used to select tens of key mouse experiments and a rough estimate is that the time and cost of experimentation was reduced by two thirds.

Future developments of the ImmunoGrid will include refinements of the existing models, additional models and simulators and extension to other diseases. Researchers who wish to contribute their models and tools to the ImmunoGrid and those who want to use the research ImmunoGrid should contact the corresponding author or contact person indicated at the ImmunoGrid Web site. Models considered for inclusion in ImmunoGrid must satisfy requirements of validation, concordance with well established experimental or clinical observations, and freely accessible source codes.

Key Points

- ImmunoGrid combines conceptual models of the immune system, models of antigen processing and presentation, system-level models of the immune system, Grid computing and database technology to facilitate vaccine discovery, formulation and optimization.
- Molecular level models address the combinatorial complexity of antigen processing and presentation. In combination with Grid computing, such prediction systems enable exhaustive parallel screening of thousands of pathogen proteomes for hundreds of MHC molecules.
- System level models enable simulations of millions of experiments. Years of experimentation can be simulated in hours, helping select 'best' or 'most informative' experiments for actual screening.

FUNDING

The ImmunoGrid project has been funded by the EC contract FP6-2004-IST-4, No. 028069.

References

1. Andre FE. Vaccinology: past achievements, present road-blocks and future promises. *Vaccine* 2003;**21**:593–5.
2. Plotkin SA. Six revolutions in vaccinology. *Pediatr Infect Dis J* 2005;**24**:1–9.
3. Kurstak E. Towards the new global vaccinology era in prevention and control of diseases. *Vaccine* 2003;**21**:580–1.
4. Brusic V, August JT, Petrovsky N. Information technologies for vaccine research. *Expert Rev Vaccines* 2005;**4**:407–17.
5. Brusic V, August JT. The changing field of vaccine development in the genomics era. *Pharmacogenomics* 2004;**5**:597–600.
6. Muzzi A, Masignani V, Rappuoli R. The pan-genome: towards a knowledge-based discovery of novel targets for vaccines and antibacterials. *Drug Discov Today* 2007;**12**:429–39.
7. Purcell AW, McCluskey J, Rossjohn J. More than one reason to rethink the use of peptides in vaccine design. *Nat Rev Drug Discov* 2007;**6**:404–14.
8. Heiny AT, Miotto O, Srinivasan KN, et al. Evolutionarily conserved protein sequences of influenza A viruses, avian and human, as vaccine targets. *PLoS ONE* 2007;**2**:e1190.
9. Nickle DC, Rolland M, Jensen MA, et al. Coping with viral diversity in HIV vaccine design. *PLoS Comput Biol* 2007;**3**:e75.
10. Fraser CK, Diener KR, Brown MP, Hayball JD. Improving vaccines by incorporating immunological adjuvants. *Expert Rev Vaccines* 2007;**6**:559–78.
11. Denis F, Alain S, Ploy MC. New routes of administration: epidermal, transcutaneous mucosal ways of vaccination. *Med Sci (Paris)* 2007;**23**:379–85.
12. Singh M, Chakrapani A, O'Hagan D. Nanoparticles and microparticles as vaccine-delivery systems. *Expert Rev Vaccines* 2007;**6**:797–808.
13. Liniger M, Zuniga A, Naim HY. Use of viral vectors for the development of vaccines. *Expert Rev Vaccines* 2007;**6**:255–66.
14. Frey SE, Newman FK, Cruz J, et al. Dose-related effects of smallpox vaccine. *N Engl J Med* 2002;**346**:1275–80.
15. Motta S, Castiglione F, Lollini P, Pappalardo F. Modelling vaccination schedules for a cancer immunoprevention vaccine. *Immunome Res* 2005;**1**:5.
16. Pinto LA, Castle PE, Roden RB, et al. HPV-16 L1 VLP vaccine elicits a broad-spectrum of cytokine responses in whole blood. *Vaccine* 2005;**23**:3555–64.
17. Pukhalsky AL, Shmarina GV, Bliacher MS, et al. Cytokine profile after rubella vaccine inoculation: evidence of the immunosuppressive effect of vaccination. *Mediators Inflamm* 2003;**12**:203–7.
18. The STEP Consortium. *Seeding the EuroPhysiome: A Roadmap to the Virtual Physiological Human*. [Online] 5 July 2007, <<http://www.europhysiome.org/roadmap>>.
19. Lefranc MP. IMGT, the international ImMunoGeneTics information system[®]: a standardized approach for immunogenetics and immunoinformatics. *Immunome Res* 2005;**1**:3. [imgt.cines.fr]
20. Lefranc MP, Giudicelli V, Duroux P. IMGT[®], a system and an ontology that bridge biological and computational spheres in bioinformatics. *Brief Bioinform* 2008;**9**:263–75.
21. Lefranc MP, Giudicelli V, Ginestoux C, et al. IMGT-ONTOLOGY for Immunogenetics and Immunoinformatics. *In Silico Biol* 2004;**4**:17–29.
22. Lefranc MP, Clément O, Kaas Q, et al. IMGT-Choreography for Immunogenetics and Immunoinformatics. *In Silico Biol* 2005;**29**:185–203.
23. Duroux P, Kaas Q, Brochet X, et al. IMGTKaleidoscope, the formal IMGT-ONTOLOGY paradigm. *Biochimie* 2008;**4**:570–83.

24. Lefranc M-P. IMGT-ONTOLOGY, IMGT® databases, tools and Web resources for Immunoinformatics. In: Schönbach C, Ranganathan S, Brusic V (eds). *Immunoinformatics. Immunomics Reviews*. New York, USA: Springer, 2008;1–18.
25. Lundegaard C, Lund O, Kecmir C, et al. Modeling the adaptive immune system: predictions and simulations. *Bioinformatics* 2007;**23**:3265–75.
26. Brusic V, Rudy G, Harrison LC. MHCPEP, a database of MHC-binding peptides: update 1997. *Nucleic Acids Res* 1998;**26**:368–71.
27. Rammensee H, Bachmann J, Emmerich NP, et al. SYFPEITHI: database for MHC ligands and peptide motifs. *Immunogenetics* 1999;**50**:213–9.
28. Toseland CP, Clayton DJ, McSparron H, et al. Antijen: a quantitative immunology database integrating functional, thermodynamic, kinetic, biophysical, and cellular data. *Immunome Res* 2005;**1**:4.
29. Sette A, Bui H, Sidney J, et al. The immune epitope database and analysis resource. *Lecture Notes Comput Sci* 2006;**4146**:126–32.
30. Brusic V, Bajic VB, Petrovsky N. Computational methods for prediction of T-cell epitopes – a framework for modelling, testing, and applications. *Methods* 2004;**34**:436–43.
31. Tong JC, Tan TW, Ranganathan S. Methods and protocols for prediction of immunogenic epitopes. *Brief Bioinform* 2007;**8**:96–108.
32. Nielsen M, Lundegaard C, Lund O, Kesmir C. The role of the proteasome in generating cytotoxic T-cell epitopes: insights obtained from improved predictions of proteasomal cleavage. *Immunogenetics* 2005;**57**:33–41.
33. Larsen MV, Lundegaard C, Lamberth K, et al. An integrative approach to CTL epitope prediction: a combined algorithm integrating MHC class I binding, TAP transport efficiency, and proteasomal cleavage predictions. *Eur J Immunol* 2005;**35**:2295–303.
34. Nielsen M, Lundegaard C, Lund O. Prediction of MHC class II binding affinity using SMM-align, a novel stabilization matrix alignment method. *BMC Bioinformatics* 2007;**8**:238.
35. Nielsen M, Lundegaard C, Blicher T, et al. NetMHCpan, a method for quantitative predictions of peptide binding to any HLA-A and -B locus protein of known sequence. *PLoS ONE* 2007;**2**:e796.
36. Larsen JE, Lund O, Nielsen M. Improved method for predicting linear B-cell epitopes. *Immunome Res* 2006;**2**:2.
37. Andersen PH, Nielsen M, Lund O. Prediction of residues in discontinuous B-cell epitopes using protein 3D structures. *Protein Sci* 2006;**15**:2558–67.
38. Reche PA, Glutting JP, Zhang H, Reinherz EL. Enhancement to the RANKPEP resource for the prediction of peptide binding to MHC molecules using profiles. *Immunogenetics* 2004;**56**:405–19.
39. Zhang GL, Khan AM, Srinivasan KN, et al. MULTIPRED: a computational system for prediction of promiscuous HLA binding peptides. *Nucleic Acids Res* 2005;**33**:W172–9.
40. Zhang GL, Bozic I, Kwok CK, et al. Prediction of supertype-specific HLA class I binding peptides using support vector machines. *J Immunol Meth* 2007;**320**:143–54.
41. Peters B, Bui HH, Frankild S. A community resource benchmarking predictions of peptide binding to MHC-I molecules. *PLoS Comput Biol* 2006;**2**:e65.
42. Larsen MV, Lundegaard C, Lamberth K, et al. Large-scale validation of methods for cytotoxic T-lymphocyte epitope prediction. *BMC Bioinformatics* 2007;**8**:424.
43. Lin HH, Ray S, Tongchusak S, et al. Evaluation of MHC class I peptide binding prediction servers: applications for vaccine research. *BMC Immunol* 2008;**9**:8.
44. You L, Zhang P, Bodén M, Brusic V. Understanding prediction systems for HLA-binding peptides and T-cell epitope identification. *Lect Notes Bioinform* 2007;**4774**:337–48.
45. Lin HH, Zhang GL, Tongchusak S, et al. Evaluation of MHC-II peptide binding prediction servers: applications for vaccine research. *BMC Bioinformatics* 2008;**9**(Suppl 12):S22.
46. Gowthaman U, Agrewala JN. *In silico* tools for predicting peptides binding to HLA-class II molecules: more confusion than conclusion. *J Proteome Res* 2008;**7**:154–63.
47. Rajapakse M, Schmidt B, Feng L, Brusic V. Predicting peptides binding to MHC class II molecules using multi-objective evolutionary algorithms. *BMC Bioinformatics* 2007;**8**:459.
48. Nielsen M, Lundegaard C, Worning P, et al. Improved prediction of MHC class I and class II epitopes using a novel Gibbs sampling approach. *Bioinformatics* 2004;**20**:1388–97.
49. Karpenko O, Huang L, Dai Y. A probabilistic meta-predictor for the MHC class II binding peptides. *Immunogenetics* 2008;**60**:25–36.
50. Motta S, Brusic V. Mathematical modeling of the immune system. In: Ciobanu G, Rozenberg G (eds). *Modelling in Molecular Biology*. Berlin: Natural Computing Series, Springer Verlag, 2004;193–218. ISBN 3-540-40799-5.
51. Louzoun Y. The evolution of mathematical immunology. *Immunol Rev* 2007;**216**:9–20.
52. Castiglione F, Liso A. The role of computational models of the immune system in designing vaccination strategies. *Immunopharmacol Immunotoxicol* 2005;**27**:417–32.
53. Falus A (ed). *Immunogenomics and Human Disease*. Hoboken, New Jersey: Wiley, 2006. ISBN: 0-470-01530-6.
54. Purcell AW, Gorman JJ. Immunoproteomics: Mass spectrometry-based methods to study the targets of the immune response. *Mol Cell Proteomics* 2004;**3**:193–208.
55. Brusic V, Marina O, Wu CJ, Reinherz EL. Proteome informatics for cancer research: from molecules to clinic. *Proteomics* 2007;**7**:976–91.
56. Schönbach C, Ranganathan S, Brusic V (eds). *Immunoinformatics*. Springer 2007. ISBN: 978-0-387-72967-1
57. Tegnér J, Nilsson R, Bajic VB, et al. Systems biology of innate immunity. *Cell Immunol* 2006;**244**:105–109.
58. Sachdeva N, Asthana D. Cytokine quantitation: technologies and applications. *Front Biosci* 2007;**12**:4682–95.
59. Harnett MM. Laser scanning cytometry: understanding the immune system in situ. *Nat Rev Immunol* 2007;**7**:897–904.
60. Brusic V, Bucci K, Schönbach C, et al. Efficient discovery of immune response targets by cyclical refinement of QSAR models of peptide binding. *J Mol Graph Model* 2001;**19**:405–11.

61. Pappalardo F, Motta S, Lollini PL, Mastriani E. Analysis of vaccine's schedules using models. *Cell Immunol* 2006;**244**:137–40.
62. Celada F, Seiden PE. A computer model of cellular interaction in the immune system. *Immunol Today* 1992;**13**:56–62.
63. Castiglione F, Bernaschi M, Succi S. Simulating the immune response on a distributed parallel computer. *Int J Mod Phys C* 1997;**8**:527–45.
64. Castiglione F, Poccia F, D'Offizi G, Bernaschi M. Mutation, fitness, viral diversity and predictive markers of disease progression in a computational model of HIV-1 infection. *AIDS Res Hum Retroviruses* 2004;**20**:1316–25.
65. Baldazzi V, Castiglione F, Bernaschi M. An enhanced agent based model of the immune system response. *Cell Immunol* 2006;**244**:77–9.
66. Castiglione F, Duca K, Jarrah A, et al. Simulating Epstein-Barr virus infection with C-ImmSim. *Bioinformatics* 2007;**23**:1371–7.
67. Castiglione F, Toschi F, Bernaschi M, et al. Computational modeling of the immune response to tumor antigens: implications for vaccination. *J Theo Biol* 2005;**237**:4:390–400.
68. Lollini PL, Motta S, Pappalardo F. Discovery of cancer vaccination protocols with a genetic algorithm driving an agent based simulator. *BMC Bioinformatics* 2006;**7**:352.
69. Pappalardo F, Lollini PL, Castiglione F, Motta S. Modeling and simulation of cancer immunoprevention vaccine. *Bioinformatics* 2005;**21**:2891–7.
70. Pappalardo F, Musumeci S, Motta S. Modeling immune system control of atherogenesis. *Bioinformatics* 2008;**24**:1715–21.
71. He X, Luo L. Theory of the lattice Boltzmann method: from the Boltzmann equation to the lattice Boltzmann equation. *Phys Rev E* 1997;**56**:6811–7.
72. Ferreira SCJr, Martins ML, Vilela MJ. Morphology transitions induced by chemotherapy in carcinomas in situ. *Phys Rev E* 2003;**67**:051914.
73. Yates A, Chan CC, Callard RE, et al. An approach to modelling in immunology. *Brief Bioinform* 2001;**2**:245–57.
74. Catron DM, Itano AA, Pape KA, et al. Visualizing the first 50 hr of the primary immune response to a soluble antigen. *Immunity* 2004;**21**:341–7.
75. Garside P, Ingulli E, Merica RR, et al. Visualization of specific B and T lymphocyte interactions in the lymph node. *Science* 1998;**281**:96–9.
76. Mempel TR, Henrickson SE, Von Andrian UH. T-cell priming by dendritic cells in lymph nodes occurs in three distinct phases. *Nature* 2004;**427**:154–9.
77. Zhang C, Crasta O, Cammer S, et al. An emerging cyberinfrastructure for biodefense pathogen and pathogen-host data. *Nucleic Acids Res* 2008;**36**:D884–91.
78. Laghaee A, Malcolm C, Hallam J, Ghazal P. Artificial intelligence and robotics in high throughput post-genomics. *Drug Discov Today* 2005;**10**:1253–9.
79. Fogel G. Computational Intelligence approaches for pattern discovery in biological systems. *Brief Bioinform* 2008;**9**:307–16.
80. Rauwerda H, Roos M, Hertzberger BO, Breit TM. The promise of a virtual lab in drug discovery. *Drug Discov Today* 2006;**11**:228–36.
81. Emerson A, Rossi E. ImmunoGrid - the virtual human immune system project. *Stud Health Technol Inform* 2007;**126**:87–92.
82. ImmunoGrid – The European Virtual Human Immune System Project: Educational and Research Portal. <http://igrid-ext.cryst.bbk.ac.uk/immunogrid/> (22 March 2009, date last accessed).
83. Churchill D. Towards a useful classification of learning objects. *Educ Technol Res Dev* 2007;**55**:479–97.
84. Hunter PJ, Crampin EJ, Nielsen PMF. Bioinformatics, multiscale modeling and the IUPS Physiome Project. *Brief Bioinform* 2008;**9**:333–43.