

Immune system simulation online

Nicolas Rapin¹, Ole Lund² and Filippo Castiglione^{3,*}

¹Biotech Research & Innovation Centre (BRIC) and Bioinformatics Centre, University of Copenhagen, DK-2200 Copenhagen N, ²Center for Biological Sequence Analysis (CBS), Technical University of Denmark, DK-2800 Lyngby, Denmark and ³Istituto per le Applicazioni del Calcolo “M. Picone” of the National Research Council of Italy, 00185 Rome, Italy

Associate Editor: Trey Ideker

ABSTRACT

Motivation: The recognition of antigenic peptides is a major event of an immune response. In current mesoscopic-scale simulators of the immune system, this crucial step has been modeled in a very approximated way.

Results: We have equipped an agent-based model of the immune system with immuno-informatics methods to allow the simulation of the cardinal events of the antigenic recognition, going from single peptides to whole proteomes.

The recognition process accounts for B cell-epitopes prediction through Parker-scale affinity estimation, class I and II HLA peptide prediction and binding through position-specific scoring matrices based on information from known HLA epitopes prediction tools, and TCR binding to HLA-peptide complex calculated as the averaged sum of a residue-residue contact potential.

These steps are executed for all lymphocytes agents encountering the antigen in a wide-reaching Monte Carlo simulation.

Availability: <http://www.cbs.dtu.dk/services/C-ImmSim-10.1/>

Contact: f.castiglione@iac.cnr.it

Received and revised on April 6, 2011; accepted on May 27, 2011

1 INTRODUCTION

Computational methods for the study of the immune system have grown to embed a great level of sophistication. Usually the level of description adopted is that of cells. Discrete mathematical models like cellular automata or agent based have shown to be more suitable for representing the details of the immunological processes. Models of the immune system include lymphocytes and antigens like viruses or bacteria but they usually keep the description of the molecular recognition events at a very simplified level. This however, is the triggering event in the onset of the immune response.

Seeking for an adequate solution to this problem, we have enhanced our agent-based model of the immune system Celada and Seiden (1992) by embedding different *in silico* methods for epitope recognition based on statistical methods derived from machine learning and trained on experimental data (Rapin *et al.*, 2010).

The overall simulation is based on three events: (i) B-cell epitopes binding, (ii) class I and II HLA epitope binding and (iii) TCR binding to HLA-peptide complexes. These events are independently executed by cells represented by agents and populating a given simulated biological volume.

The outcome is a fully detailed tracking of the dynamics of cell populations together with molecules involved in the epitope recognition process including those immune complexes exposed on the surface of antigen-processing cells.

2 METHODS

In the computational model described in Rapin *et al.* (2010), the original idea of representing the shape space of lymphocytes' receptors in the binary strings has been substituted with the representation of molecules through their amino acid sequence, and the prediction of binding events relies on *in silico* epitope discovery and binding prediction methods.

For example, the interaction between the host T cells and pathogens relies on position-specific scoring matrices developed in Rapin *et al.* (2008). Other immuno-informatics methods have also been incorporated, namely B-cell epitopes prediction, and a general protein-protein potential measure aimed at estimating the binding affinity between molecules.

The *C-ImmSim* simulator consists of >18000 lines of ANSI C code. Thanks to dynamic memory allocation and other optimization techniques, it is possible to represent large simulated volumes and, most importantly, a substantial immune repertoire. The high degree of complexity of the underlying model makes it suitable to simulate the immune response to pathogens.

The *immune system simulation server* presented here consists of a simple interface to the simulation method mentioned above where the user can specify the antigen injection schedule, the haplotype of the host and input a few other parameters such as the simulation volume and duration (see Fig. 1).

The most important feature is the possibility to specify the antigen in terms of its constituting proteins. The user can submit the antigen protein sequence and specify the schedule of its injection as well as the host haplotype. Finally, other input fields allow the user to specify the random initial repertoire, the simulation duration and the simulated volume.

The first input field is used to get the molecular definition of the antigen/s chosen to be injected either in FASTA format or as UniProt (The Universal Protein Resource, <http://www.uniprot.org/>) protein accession codes or PDB (The Protein Data Bank, <http://www.pdb.org/>), structure code or PDB ID (i.e. primary identifier) by which entries can be retrieved from the UniProt and Protein Data Bank servers automatically.

The second input field allows the user to specify the antigen injection schedule with the following simple syntax:

$$[A_1, T_1, D_n]; [A_2, T_2, D_2]; \dots; [A_n, T_n, D_n]$$

to specify the n injections of different compounds with identifier A_k indicating the protein name in the fasta file field or UniProt/PDB accession number, T_k is the time step of the injection (1 simulation step = 8 h) and D_k is the antigen dose in units per micro liters of volume.

Having specified the antigen, the user can chose the host haplotype by means of the corresponding drop-down menus; two for HLA A, two for HLA B and two for HLA DR. These choices determine the position-specific scoring matrices used for epitope discovery during the simulation.

*To whom correspondence should be addressed.



Fig. 1. The *C-ImmSim* online server allows the user to define the antigen to be injected as a list of UniProt accession numbers, or PDB primary identifiers, or, as a multiprotein FASTA text. The haplotype is defined by drop-down menus. Other parameters are the simulation time and the volume.

Once the input is set and the user launches the simulation, a new page informs that the simulation has started and ask whether the user wants to be warned upon completion. Depending on the simulated volume chosen, the simulation can take from minutes up to several hours. The injection schedule and the complexity of the pathogen itself also have an influence on the simulation time.

The output page of the simulator shows a set of files corresponding to population data (both total number of lymphocytes and the division between clonotypes, cytokines and antibody concentrations per lattice point). Finally, the user can also download a compressed archive containing the pictures displayed on the page.

3 DISCUSSION

The immune system simulation server described here offers the chance to test the overall immunogenicity of a generic protein sequence in the form of its amino acids sequence. It is based on

an agent-based model of the immune system coupled with epitope prediction tools derived from the field of immuno-informatics and novel methods to assess protein–protein binding affinity on the basis of the work performed by Miyazawa and Jernigan on protein potentials (Miyazawa and Jernigan, 1996) that provides us with a method for assessing the chances of direct interactions among proteins in the simulation.

The proposed methods for assessing epitope discovery and binding are embedded in the architecture in a modular way and as such can be upgraded or substituted by more accurate methods should one become available.

4 CONCLUSION

We have implemented as a web service a tool to calculate the immune response to a generic pathogenic secondary structure. The user inputs a sequence or a set of sequences comprising the interested pathogen as FASTA sequences or by means of Uniprot identifiers, then specifies the injection schedule in terms of time and dosage. The simulator will generate a virtual injection schedule and the corresponding immune response (if any) to the peptides identified in the secondary sequences specified in input. Other parameters are customizable such as the simulated volume and the random seed determining the specificity of the initial population of lymphocytes. Since the method is of the Monte Carlo type, a larger volume will result in more accurate data. Moreover, for the same reason, more simulations will allow for better statistical power.

Funding: EC contract (FP6-2004-IST-4, No.028069) (ImmunoGrid) in part.

Conflict of Interest: none declared.

REFERENCES

Celada,F. and Seiden,P.E. (1992) A computer model of cellular interactions in the immune system. *Immunol. Today*, **13**, 56–62.
Miyazawa,S. and Jernigan,R.L. (1996) Residue-residue potentials with a favorable contact pair term and an unfavorable high packing density term, for simulation and threading. *J. Mol. Biol.*, **256**, 623–644.
Rapin,N. et al. (2008) MHC motif viewer. *Immunogenetics*, **60**, 759–765.
Rapin,N. et al. (2010) Computational immunology meets bioinformatics: the use of prediction tools for molecular binding in the simulation of the immune system. *PLoS ONE*, **5**, e9862.