Immunoinformatics

Dr Filippo Castiglione discusses the emerging field of immunoinformatics and how machine learning has benefited his most recent project that has advanced the immune system simulator, C-ImmSim

Could you explain how you became interested in immunoinformatics and discuss your research objectives?

I am a computer scientist, and many years ago, I happened to meet an immunologist who was the co-author of a computer model called the IMMune system SIMulator (IMMSIM), which simulates the immune response to antigens. This inspired my work on developing my own version of that simulation system, and I successively conducted a number of studies that aimed to understand the immune dynamics in pathological conditions such as AIDS, allergies and cancer. My interest in bioinformatics came after that and as the availability of free genomic and proteomic data grew. Alongside statistical analysis of this kind of data, I became interested in combining this information with my simulation of the immune system.

Your most recent project combines systems biology with data-driven prediction models as a new approach for studying the immune system. What was the inspiration behind this work?

I have been involved in an EU project with the Center for Biological Sequence Analysis at the Technical University of Denmark in Copenhagen, a top class institute for molecular prediction algorithms. This has given me the chance to discover a new area that is not only interesting, but beyond my usual reach. I had the idea to combine algorithms that predict antigenic peptides with my agent-based simulation system, so I pushed for that and received the support of experts in machine-learning bioinformatics predictions.

What advantage does C-ImmSim have over existing simulation tools?

The advantage of C-ImmSim is that it is one of the first and most complete 'model of models'. It is not an ad-hoc model to answer this or that question, but rather one entailing a number of immunological assumptions, theories, conjectures and known facts. It reproduces a virtual immune system from a combination of many ideas.

How did you use machine-learning techniques in this study, and why is machine learning so valuable for projects like this?

Machine learning provides a method for assessing molecular binding in the simulation. What was previously achieved with binary strings is now possible with the primary molecular structure: the linear sequences of amino acids. This is extremely valuable because it allows the simulation to move a step away from the realm of mathematics towards that of biologists. The language has changed and the way of interpreting results has changed, meaning that a closer biological definition of the pathogen and haplotype of the virtual patient is now possible.

A recent modification of the C-ImmSim model represents molecules as strings of letters representing the 20 amino acids, in place of a binary string. What were the challenges of implementing this change? Are there benefits to this approach?

There are technical challenges in using an alphabet of 20 symbols rather than the binary alphabet. The most obvious is the increase in the computational requirements of the simulation tool; however, large-memory and high-performance computers can manage this. The real challenge was replacing the matching bit-string algorithm with a much more complex affinity estimation. Ultimately, immunoinformatics approaches helped to realise this. The benefit of replacing the binary alphabet with a string of letters is that the tool is much more intuitive for biologists and the readouts have a closer relationship with reality.

Has a multidisciplinary approach proved important to the success of the project?

Having a multidisciplinary approach has been absolutely necessary for the success of most projects I've been involved with so far. You need to learn about the topic that you want to look at before you can model it.

When you bring back some new insight to your fellow collaborator, it is important to see how they realise the potential of the multidisciplinary collaboration. Great things usually come after that.

Could you explain your hopes for the field of immunoinformatics?

There are probably two areas in which I'd like to see improvement. The first, of course, is advances in prediction algorithms. The second regards the standardisation of databases. It can be very time consuming to find useful information if you're not comfortable or don't feel at home with specialised bioinformatics databases. Often you even find fragmented information that is contradictory, which is frustrating.

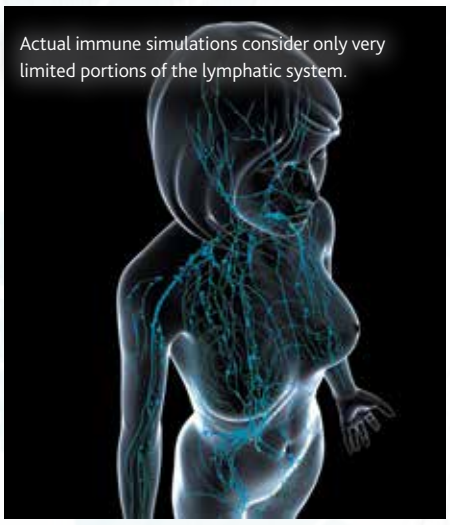
Improving **immune system** simulations

A multidisciplinary team of researchers in the **Institute of Applied Mathematics 'Mauro Picone'** at the National Research Council has developed the first immune system simulator that integrates predictive algorithms

THE IMMUNE SYSTEM is fundamental in protecting the body against disease, yet understanding how it works is one of the most complex topics in biology. In the past, immunology research typically relied on studying *in vitro* or *in vivo* animal models, but since the mid-1990s, scientists have been using computational (*in silico*) models as alternatives to performing experiments in the laboratory.

IN SILICO MODELS

The advantages of *in silico* methods are numerous. They allow the study of hypothetical compounds, and generally, they are based on human data, ruling out the question of the transferability of results between species. Typically, scientists convert immunological data into a computational problem, solve it and interpret the results and their relevance.



Actual immune simulations consider only very limited portions of the lymphatic system.

Immunoinformatics programs have become an indispensable tool in such examinations, and the rise in availability of free genomic and proteomic data has allowed researchers to develop the programs continually.

Dr Filippo Castiglione currently works as a senior researcher at the Institute of Applied Mathematics 'Mauro Picone' at the National Research Council in Rome, Italy. He has been involved with immune system simulations for nearly 20 years, and continues to undertake research in this area. In 1995, he developed C-ImmSim, the C-language version of IMMSIM – the IMMune system SIMulator – that was created a few years previously. Now, Castiglione and his collaborators in Denmark are working to develop the program further, using a novel approach that combines computational immunology with machine-learning prediction techniques, in the hope that it will lead to a better understanding of pathogenic and patient-specific immune system dynamics.

THREE PREDICTIONS

Being able to predict which part of an invading pathogen will provoke an immune response is key to gaining a better understanding of immune-related diseases and developing better diagnostics and therapies. Studies investigating epitopes are particularly important for disease understanding, as they are the sites on pathogenic molecules that the immune system recognises.

The team's approach integrates three existing prediction algorithms into C-ImmSim to determine the epitopes of pathogens: the prediction of B-epitopes, which are sites antibodies of the immune system recognise; the prediction of major histocompatibility complex (MHC) and peptide interaction; and the affinity between the MHC-peptide complex and a T-cell receptor. The MHC is a set of molecules found on all nucleated cells that bind peptide fragments derived from pathogens and display them on the cell surface so the appropriate T-cells in the immune system can identify them. Class-1 MHC molecules display fragments of proteins from within the cell whereas class-2 MHC molecules display extracellular proteins. Not only do the algorithms powering C-ImmSim allow the prediction of these MHC-peptide complexes, but the B-epitope, MHC-peptide complex and T-cell affinity predictions can compute the likelihood that an immune response will be triggered because of this recognition.

BENCHMARKING THE SIMULATIONS

Although scientists have been using computational models to simulate immune system dynamics for many years, this is the first time a simulation has integrated prediction algorithms. Therefore, C-ImmSim's simulations needed to be benchmarked against classical immunology experiments, which Castiglione and his team achieved with encouraging results.

In one instance of benchmarking simulations, Castiglione's team examined a typical immunisation experiment in which the immune system develops memory. In a realistic scenario of this process, there are two responses – primary and secondary. Due to the development of memory cells, the secondary response occurs at a faster rate. To see how C-ImmSim's simulation measured up to a real-world situation, the team simulated the injection of an HIV-1 gag protein at time zero and again six months (of simulated time) later. C-ImmSim showed that

the dynamics of the model are consistent with a realistic immunisation process: both a primary and secondary immune response occurred, with the secondary response being much more rapid due to the immunological memory developed during the first response.

In another experiment designed to test the exactness of C-ImmSim's models, Castiglione looked at the immune system and its reaction to repeated exposures of the same antigen. When this situation happens, the body produces antibodies of successively greater affinities with the result that a secondary response can elicit antibodies with an even greater affinity than in a primary response. Castiglione's team mimicked this phenomenon by continuously injecting a certain amount of the HIV gag protein through the simulation period. As the simulation proceeded, antibodies with a higher affinity took over the original antibodies, evidencing the pathogen specificity of the simulation.

Through these tests, Castiglione's team also became confident that C-ImmSim soundly represents both the innate immune system, which recognises and responds to pathogens in a generic way, and the adaptive immune system, which is triggered by the presentation of antigen to T-cells.

PATIENT SPECIFICITY

In later experiments, Castiglione tested C-ImmSim's ability to work with patient-specific data. Current research shows that individuals with a heterozygote haplotype – a combination of different genes – as opposed to homozygote haplotype, have a better immune response. Castiglione and his team tested over 500 simulations involving heterozygous and homozygous individuals responding to an antigen from the influenza A serotype H1N1. The results reflected reality and showed that heterozygote individuals were much more likely to recognise a pathogen compared to homozygote individuals. "This simulation evidences how important it is to be able to perform patient-specific simulations," explains Castiglione. "The result was expected, but it is another example of C-ImmSim being able to reproduce realistic data."

A NEW LANGUAGE

As well as the integration of prediction methods, Castiglione's version of C-ImmSim is the first immune system simulation model to represent molecules' structures as strings of letters corresponding to the 20 amino acids' linear sequences. Previous models used binary strings. This new approach allows biologists to use the tool more readily and in a 'language' closer to the one they use daily in their field. Moreover, they can interpret results in a more realistic way, as Castiglione underlines: "A closer biological definition of the pathogen and of the haplotype of the virtual patient is now possible".

FUTURE IMPROVEMENTS

"The accuracy of C-ImmSim can be enhanced, but considering it's difficult to improve prediction methods, I guess there is a long way to go before my simulation results can be refined," states Castiglione. Therefore, he suggests that one way to get around inexactness that may arise in C-ImmSim models due to the underlying prediction methods would be for scientists to bypass them and to pre-compute an affinity for peptides in the injected pathogen. Pre-calculating the affinities via other methods or using available lab data could be more reliable and improve the ability of the simulator to answer questions about the specificity of an immune response against a particular pathogen.

However, as new prediction methods arise, testing them in C-ImmSim will require minimal programming effort, according to Castiglione, as he has purposefully kept the computer code separate within the system. Therefore, he is confident that once the prediction algorithms are improved and acquire precision, C-ImmSim could be the computerised immune system of the future, aiding immunology research: "My personal hope regarding the simulation of the immune system is that it becomes an important tool used in the vaccine development pipeline".

COMPUTATIONAL IMMUNOLOGY MEETS BIOINFORMATICS: THE USE OF PREDICTION TOOLS FOR MOLECULAR BINDING IN THE SIMULATION OF THE IMMUNE SYSTEM

OBJECTIVES

- To integrate a simulation tool for immune system dynamics that is both patient and pathogen specific
- To provide molecular information and bioinformatics prediction methods to create a detailed computational model of the innate and adaptive immune response

KEY COLLABORATORS

Dr Nicolas Rapin, Biotech Research and Innovation Centre, University of Copenhagen, Denmark • **Professor Ole Lund**, Center for Biological Sequence Analysis, Department of Systems Biology, Technical University of Denmark, Denmark • **Professor Massimo Bernaschi**, Institute for Computing Applications, National Research Council, Italy

PARTNERS

Center for Biological Sequences Analysis (CBS), Technical University of Denmark (DTU), Denmark

FUNDING

European Commission (funded project: ImmunoGrid)

National Research Council (CNR), Italy

CONTACT

Dr Filippo Castiglione
Senior Researcher

Institute of Applied Mathematics 'Mauro Picone'
National Research Council (CNR)
Via dei Taurini n 19
00185 Rome
Italy

T +39 6771 6452
E f.castiglione@iac.cnr.it

www.iac.cnr.it/~filippo

www.plosone.org/article/info%3Adoi%2F10.1371%2Fjournal.pone.0009862

FILIPPO CASTIGLIONE completed a degree in Computer Science from the University of Milan and, in an academic capacity, briefly visited the IBM Thomas J Watson Research Center, USA, and the Department of Molecular Biology at Princeton University, USA. After a year at ST Microelectronics in Italy, he started his scientific career, completing a doctorate in Scientific Computing at the University of Cologne in Germany. He continued his progression, becoming a four-month visiting scholar at Harvard Medical School, USA, and then completing a one-year postdoctoral study in Israel. Castiglione is now a Senior Researcher at the National Research Council, studying the modelling of complex and biological systems with a particular interest in the immune system and related pathologies.