Pharmacological Research 92 (2015) 40–45

Contents lists available at ScienceDirect

Pharmacological Research

journal homepage: www.elsevier.com/locate/yphrs

Review

Computational modelling approaches to vaccinology

Francesco Pappalardo[a,*], Darren Flower[b], Giulia Russo[a], Marzio Pennisi[c], Santo Motta[c]

[a] Dipartimento di Scienze del Farmaco, University of Catania, Catania, Italy
[b] School of Life and Health Sciences Aston University, Aston Triangle, Birmingham, UK
[c] Dipartimento di Matematica e Informatica, University of Catania, Catania, Italy

ARTICLE INFO

Article history:
Received 29 July 2014
Received in revised form 4 August 2014
Accepted 18 August 2014
Available online 16 September 2014

Keywords:
Vaccine research
Modelling
Computational vaccinology
Immune system
Epitopes
Simulations

ABSTRACT

Excepting the Peripheral and Central Nervous Systems, the Immune System is the most complex of somatic systems in higher animals. This complexity manifests itself at many levels from the molecular to that of the whole organism. Much insight into this confounding complexity can be gained through computational simulation. Such simulations range in application from epitope prediction through to the modelling of vaccination strategies. In this review, we evaluate selectively various key applications relevant to computational vaccinology: these include technique that operates at different scale that is, from molecular to organisms and even to population level.

Contents

Introduction 40
Epitopes 41
PAMPS and adjuvants 41
Higher order systems 42
Agent based models in computational vaccinology 42
Mathematical models in computational vaccinology 43
Conclusions 44
References 44

Introduction

Despite its overwhelming and often confounding complexity, the immune system is ultimately a collection of parts working together to effect defence against pathogens and many other homeostatic functions. The problem, of course, when one tries to understand the immune system, is the remarkable level of emergent behaviour we observe-at many levels-from the formation of supramolecular complexes at the Immune synapse; through the action of specific immune cells, such as dendritic cells and T-cells; to organs; and, finally, whole organisms. Beyond even the whole animal, effects such as herd immunity and infectivity patterns manifest themselves only in large, interacting pseudo-social networks. As we see, or allude to, elsewhere, much of this can be modelled, and modelled with some success. Yet despite the daunting emergent, higher-level behaviour we see, much can still be learned from attempting to understand and model the underlying molecular components that comprise the immune system.

Nowadays biological systems are analyzed and managed by means of new emerging technologies that are revolutionizing biotechnology and information technology, producing a huge amounts of data. This data needs to be integrated and is quickening the process of knowledge discovery, enabling the study of biological systems at various levels i.e., from molecules to organisms and even to the population level.

* Corresponding author. Tel.: +39 095 7383073.
E-mail addresses: francesco.pappalardo@unict.it, fp@francescopappalardo.net (F. Pappalardo).

http://dx.doi.org/10.1016/j.phrs.2014.08.006
1043-6618/

The human activity entailing the representation, the manipulation and the communication of real-world daily life objects is known as modelling. Mathematical and computational models are gradually used to assist deduce biomedical data produced by high-throughput genomics and proteomics endeavours. The application of advanced computer models allowing the simulation of complex biological processes produces hypotheses and proposes experiments. Computational models are set to exploit the wealth of data stored on biomedical databases through text mining and knowledge discovery methods.

The first immunoinformatics tools for vaccine design were developed in the 1980s by DeLisi and Berzofsky and others [56]. Chief among vaccine design informatics tools are epitope-mapping algorithms. A new era of vaccine research began in 1995, when the complete genome of Haemophilus influenzae (a pathogenic bacterium) was published [58]. In parallel with advances in molecular biology and sequencing technology, bioinformatics analysis of microbial genome data has allowed in silico selection of vaccine targets. Further advances in the field of immunoinformatics have led to the development of hundreds of new vaccine design algorithms. This novel approach for developing vaccines has been named reverse vaccinology [59] or immunome-derived vaccine design [60]. Pharmaceutical companies are starting to use models to optimize/predict therapeutic effects at the organism level, suggesting that computational biology can effectively play a key role in this field [57].

Along with these techniques, the simulation of the immune system in a detailed way to reproduce and predict the effects of artificial immunity elicited by vaccines represents a challenge that several people are attempting with success. The immune system represents one of the most complex biological system. It is, in fact, an adaptive learning system which operates at multiple levels (molecules, cells, organs, organisms, and groups of organisms). Immunological research, both basic and applied, needs to deal with this complexity [4].

In this paper, we analyze and discuss several computational modelling techniques applied to vaccinology science.

Epitopes

Arguably, the simplest unambiguous component of the immune is the so-called epitope. The epitope at its most generally defined is very much the immunological quantum that lies central to immune responses and vaccination. It is the ability of the immune system to identify, respond to, and remember epitopes that powers natural immunity, and thus vaccination. Peptide epitopes are mediated primarily by their interaction with Major Histocompatibility Complexes (T-cell Epitopes, or TCEs) and antibodies (B-cell epitopes, or BCE).

Currently, commonly-used prediction of B cell epitopes often remains primitive, or depends on an elusive knowledge of protein structure, and both structure- [9] and data-driven [10] prediction of antibody-mediated epitopes have again been shown to be poor. Explaining such sub-optimality may point to a fundamental misinterpretation of extant epitope data. PEPSCAN is perhaps the most abundant data available currently but may not be what it seems. Experimentally derived epitopes are identified by assayed against pre-existing antibodies with affinity for whole antigens. If, for example, “epitopes” are mapped back to their original antigen structure, we find them randomly located through the structure rather than equating to obvious surface patches, as might be expected if they simply reproduced discontinuous epitopes identified by crystallography. In situ antigenic regions are often not exposed and thus accessible to binding by antibodies binding but rather completely buried. If we compare the conformation of antibody-bound peptides with those from the intact antigen, they are usually quite different. However, B-cell epitopes in isolated antigen and in whole antigen-antibody complexes are much more similar. Is it possible then, that the isolated peptide adopts a conformation which mimics the surface features of a discontinuous epitope or that the preformed antibody recognize denatured antigen in vivo.

Currently, prediction of T cell epitopes remains largely confined to predictions of varying accuracy of peptide binding to Major Histocompatibility Complex. Nonetheless, and compared to B-cell prediction, methods for predicting T cell epitopes show significant algorithmic sophistication. Prediction of the binding of peptides to class I MHCs, at least for well-studied alleles, such as HLA-A*0201, is now at useable accuracy [11]. However, comparative studies have shown recently that the prediction of class II MHC binding prediction T-cell epitopes is typically poor [12–15], and likewise for structure-driven prediction of class I and class II T-cell epitopes [16].

All epitope prediction methods remain severely constrained by the data used to construct them; this is particularly true of T-cell prediction. It has recently been shown that that T-cell epitopes, which were previously thought to be short peptides of 8–10 amino acids, can be up to 16 amino acids or perhaps even more. The existence of such longmer epitopes has significantly enlarged the repertoire of peptides open to inspection by T-cells [17]. Many of the cutting edge approaches to epitope discovery are trying to address these issues by inducing models of large numbers of alleles across many peptide lengths by making assumptions about how separable are the sub-sites in the peptide binding groove and how thee can be combined combinatorially to generate pseudo-binding profiles [18–20]. However, as is well-known, no data-driven method can go beyond the data used to train it; all methods are likewise much superior in their ability to interpolate than their ability to extrapolate.

Evidence exists that the responsiveness of the immune system to pathogenic proteins is only poorly correlated with the possession of T cell epitopes, and that many potential epitopes have been deleted in proteins regularly accessible to immune surveillance, perhaps as an evolutionary counter measure in the war between host and pathogen [21]. Such a deficit, and the significantly suboptimal prediction of both B-cell and T-cell epitopes described above have suggested that methods which rely solely on the possession of epitopes are unlikely to be effective at identifying antigens or immunogens. This conjecture is confirmed by what information there is, which indicates that there is little simple correspondence between antigens selected on this basis and experimentally verified antigenic or protective proteins. In turn this has led to the development of other approaches to predicting whole antigens within pathogen genomes, proteins likely to be antigenic and protective; of which there are three key approaches: subcellular location prediction, sequence similarity, and empirical statistical approaches, typified by VaxiJen [22,23] and expert systems such as nerve [24].

PAMPS and adjuvants

Other epitopes exist, notable the so-called Pathogen-Associated Molecular Pattern (or PAMP), highly conserved and typically complex molecular moieties recognized by pattern recognition receptors (or PRRs) of the innate immune system [25]. Many PAMPs, and molecules mimicking the recognition of PAMPs, form the basis of adjuvants. Adjuvants potentiate immune responses, reducing the dosing requirements needed to induce protective immunity, particularly for weakly immunogenic subunit vaccines. Few adjuvants are licensed for human use: principally alum, and squalene-based