Human Immunology xxx (2015) xxx–xxx

Contents lists available at ScienceDirect

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

# Major histocompatibility complex linked databases and prediction tools for designing vaccines

Satarudra Prakash Singh [a,*], Bhartendu Nath Mishra [b]

[a] Amity Institute of Biotechnology, Amity University Uttar Pradesh (Lucknow Campus), Malhaur, Gomti Nagar Extension, Lucknow 226028, India
[b] Institute of Engineering and Technology, U.P. Technical University, Lucknow 226021, India

ARTICLE INFO

*Article history:*
Received 30 April 2015
Revised 29 August 2015
Accepted 9 November 2015
Available online xxxx

*Keywords:*
Epitope
Immunoinformatics
MHC
Vaccine
Prediction
Supertype

ABSTRACT

Presently, the major histocompatibility complex (MHC) is receiving considerable interest owing to its remarkable role in antigen presentation and vaccine design. The specific databases and prediction approaches related to MHC sequences, structures and binding/nonbinding peptides have been aggressively developed in the past two decades with their own benchmarks and standards. Before using these databases and prediction tools, it is important to analyze why and how the tools are constructed along with their strengths and limitations. The current review presents insights into web-based immunological bioinformatics resources that include searchable databases of MHC sequences, epitopes and prediction tools that are linked to MHC based vaccine design, including population coverage analysis. In T cell epitope forecasts, MHC class I binding predictions are very accurate for most of the identified MHC alleles. However, these predictions could be further improved by integrating proteasome cleavage (in conjugation with transporter associated with antigen processing (TAP) binding) prediction, as well as T cell receptor binding prediction. On the other hand, MHC class II restricted epitope predictions display relatively low accuracy compared to MHC class I. To date, pan-specific tools have been developed, which not only deliver significantly improved predictions in terms of accuracy, but also in terms of the coverage of MHC alleles and supertypes. In addition, structural modeling and simulation systems for peptide–MHC complexes enable the molecular-level investigation of immune processes. Finally, epitope prediction tools, and their assessments and guidelines, have been presented to immunologist for the design of novel vaccine and diagnostics.



## 1. Introduction

In the post-genomic era, vaccine research has attempted to extract the maximum out of the big data available in the field of immunomics. The conventional approaches, such as killed/live-attenuated pathogens of vaccine development, have been found to be time consuming, costly, and inefficient. Further, such approaches are not particularly feasible for pathogens, which are antigenically diverse, not cultivable in the laboratory, lack suitable animal models of infection, and/or are controlled by major histocompatibility complex (MHC) dependent immune responses [1,2]. However, these limitations can be overcome by exploiting the concept of reverse vaccinology and/or reverse immunology, where safer and more efficient vaccines can be developed. These approaches have been found to be effective against various life threatening pathogens (e.g., *Plasmodium falciparum*, *Mycobacterium tuberculosis*, etc.). Among these high-throughput strategies, reverse vaccinology involves the computational screening of novel gene products as antigens, based on the genome/proteome of the pathogen [3,4]. On the other hand, reverse immunology is based upon the prediction and identification of immunogenic peptides from the antigens [5,6]. These facts suggest that the induction of MHC mediated adaptive immune responses as an appealing method for the design of successful and effective vaccines [6,7].

Specificity and memory are the two most important features associated with adaptive immune responses. These antibody- and cell-mediated responses are carried out by B and T cells, respectively. Interestingly, epitopes representing specific parts of antigens that are recognized by either B or T cells receptors result in adaptive responses. B cell epitopes can be linear (contiguous amino acids), or conformational (discontinuous amino acids that are spatially brought together in folded proteins). Contrary to this, T cell epitopes are short, linear peptides, which are primarily derived

* Corresponding author.
*E-mail addresses:* spsingh5@amity.edu, sprakashsingh@gmail.com (S.P. Singh).

http://dx.doi.org/10.1016/j.humimm.2015.11.012
0198-8859/

through antigen processing and presentation pathways. The two types of T cell, cytotoxic T lymphocyte (CTL) and helper T lymphocyte (HTL), are differentiated by the expression of CD8+ and CD4+ co-receptor proteins, respectively. The primary role of CTLs is cytolytic activity, resulting in apoptosis of infected cells. However, the primary function of HTLs is to produce cytokines that regulate the rest of the immune responding cells. Vaccines developed through conventional approaches rely on inducing serum antibodies. It is well established that neutralizing antibodies cannot be efficiently produced without the help of HTLs that play key roles in B cell expansion and differentiation, including class switching and affinity maturation [7].

A critical literature review revealed the complexity of the vaccine design resulting from two primary sources – pathogen variability and human immune system diversity. Considering these, the design of broadly protective vaccines involves the identification and selection of conserved and promiscuous peptide based vaccine targets. These vaccine targets are cross-reactive to pathogen strains as well as protective for a large human population (Fig. 1). Thus, bioinformatics involving vaccine design is receiving immense attention due to its rapid, safe, efficient and economical nature, with the capability to design a small set of key experiments that can be followed by validation.

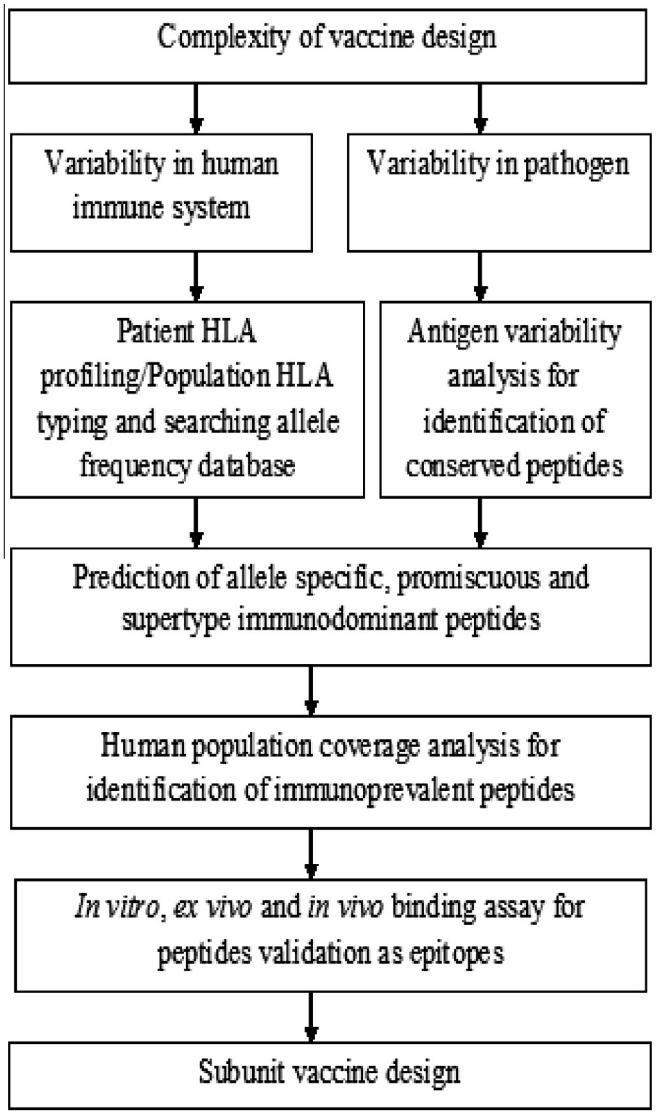


**Fig. 1.** Schematic flow diagram for conserved T cell epitope based vaccine design covering broader human population.

### 1.1. Cross antigen processing and presentation

Professional antigen presenting cells (APC), such as dendritic cells (DC) and macrophages, can activate T cells by efficiently processing antigens from both endogenous and exogenous sources, and presenting them at the plasma membrane.

Endogenous antigen processing machinery is essential for the generation of antigenic peptides and loading onto MHC class I molecules in the endoplasmic reticulum (ER). One of the major components of this pathway is the constitutive/immunoprotea some, which specifically cleaves ubiquitin-conjugated cytoplasmic proteins into small peptides. The transporter associated with antigen processing (TAP) protein mediates translocation of these peptides into the ER (where further trimming to nonameric peptides suitable for MHC class I binding occurs). Once folding is completed in the ER lumen, the peptide–MHC (pMHC) class I complex is exported through the default secretory pathway to the plasma membrane [8].

Exogenously derived antigenic peptides are generated in the endosomal/lysosomal pathway and loaded onto MHC class II molecules in specialized lysosomal antigen-loading compartments, and the pMHC class II complex is transported to the cell surface. Where MHC class I usually binds to 8–10 residue peptides, MHC class II binds to 13–20 residue peptides with a minimum 9-mer binding core [9].

In some cases, such as myeloid DCs, exogenous peptides are able to interact with MHC class I molecules via different processing pathways, a process known as cross-presentation [10]. Similarly, endogenous antigens can be transferred to the endosomal/lysosomal pathway for loading onto MHC class II molecules. Autophagy has been reported to play a role in this process by delivering cytoplasmic material for degradation in the lysosomes. The proteasome could also be involved by degrading cytoplasmic antigens before they enter the endosomal/lysosomal pathway [11]. Recent studies reveled that delivery of endogenous antigens to a degradative compartment by autophagy also contributes towards optimal CD8+ T cell activation. Hence, the contribution of autophagy to antigen processing and presentation on MHC class I molecules underlines the interplay between the different pathways of antigen presentation [12].

## 2. Databases of MHC alleles, binding peptides and T cell epitopes

As of October 2015, approximately 13,840 different human leukocyte antigen (HLA) molecules have been reported, including 10,297 HLA class I and 3543 HLA class II alleles [13]. A single HLA molecule may bind thousands of different peptides, requiring a compromise between high affinity and broad specificity. Antigenic peptide presentation as an epitope depends on interaction between MHC and T cell receptor (TCR). Several databases of MHC allele sequences, binding peptides, T cell epitopes, and related structural databases exist in the accessible public domain (Tables 1 and 2). However, only a few databases include authentic and experimentally validated peptides/epitopes with binding affinity ($IC_{50}$) values in nM (nmol/l). The immune epitope database (IEDB 3.1) is one such extensive database, which incorporates the largest number of epitopes, including 148,939 peptidic epitopes, 2298 non-peptidic epitopes, 275,074 T cell assays, 357,859 B cell assays, 343,010 MHC binding assays, 3450 epitope source organisms, 714 restricting MHC alleles, and 17,256 references [14].

Considering the above facts, a collection of authentic binders and nonbinders is crucial to build a reliable training and validation

Please cite this article in press as: S.P. Singh, B.N. Mishra, Major histocompatibility complex linked databases and prediction tools for designing vaccines, Hum. Immunol. (2015), http://dx.doi.org/10.1016/j.humimm.2015.11.012