Vaccine xxx (2015) xxx–xxx

Contents lists available at ScienceDirect

Vaccine

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

A bioinformatics tool for epitope-based vaccine design that accounts for human ethnic diversity: Application to emerging infectious diseases

Q1 Patricio Oyarzun[a,b,*], Jonathan J. Ellis[a], Faviel F. Gonzalez-Galarza[c], Andrew R. Jones[c], Derek Middleton[d], Mikael Boden[a,e], Bostjan Kobe[a,**]

[a] School of Chemistry and Molecular Biosciences, Institute for Molecular Bioscience and Australian Infectious Diseases Research Centre, University of Queensland, Australia
[b] Biotechnology Centre, Universidad San Sebastián, Concepción, Chile
[c] Institute of Integrative Biology, University of Liverpool, United Kingdom
[d] Transplant Immunology Laboratory, Royal Liverpool University Hospital & School of Infection and Host Defence University of Liverpool, United Kingdom
[e] School of Information Technology and Electrical Engineering, University of Queensland, Queensland 4072, Australia

ARTICLE INFO

Article history:
Received 1 October 2014
Received in revised form
11 December 2014
Accepted 14 January 2015
Available online xxx

Keywords:
Emerging infectious diseases
Immunodominance
Lassa, Nipah and Hendra viruses
MHC (HLA) class II proteins
Multi-epitope peptide vaccination

ABSTRACT

Background: Peptide vaccination based on multiple T-cell epitopes can be used to target well-defined ethnic populations. Because the response to T-cell epitopes is restricted by HLA proteins, the HLA specificity of T-cell epitopes becomes a major consideration for epitope-based vaccine design. We have previously shown that CD4+ T-cell epitopes restricted by 95% of human MHC class II proteins can be predicted with high-specificity.
Methods: We describe here the integration of epitope prediction with population coverage and epitope selection algorithms. The population coverage assessment makes use of the Allele Frequency Net Database. We present the computational platform Predivac-2.0 for HLA class II-restricted epitope-based vaccine design, which accounts comprehensively for human genetic diversity.
Results: We validated the performance of the tool on the identification of promiscuous and immunodominant CD4+ T-cell epitopes from the human immunodeficiency virus (HIV) protein Gag. We further describe an application for epitope-based vaccine design in the context of emerging infectious diseases associated with Lassa, Nipah and Hendra viruses. Putative CD4+ T-cell epitopes were mapped on the surface glycoproteins of these pathogens and are good candidates to be experimentally tested, as they hold potential to provide cognate help in vaccination settings in their respective target populations.
Conclusion: Predivac-2.0 is a novel approach in epitope-based vaccine design, particularly suited to be applied to virus-related emerging infectious diseases, because the geographic distributions of the viruses are well defined and ethnic populations in need of vaccination can be determined ("ethnicity-oriented approach"). Predivac-2.0 is accessible through the website http://predivac.biosci.uq.edu.au/.

1. Introduction

Q3 Emerging infectious diseases (EIDs) caused by major families of viruses are increasing in frequency, causing a high disease burden and mortality world-wide [1,2]. Epitope-based vaccines (EVs) make use of short antigen-derived peptide fragments that are administered to be presented either to T-cells (as T-cell epitopes in association with HLA molecules), or B-cells (as B-cell epitopes) [3]. While CD8+ cytotoxic T-cells generally recognize intracellular peptides displayed by HLA class I molecules, CD4+ T-helper cells generally recognize peptides from the extracellular space, displayed by HLA class II molecules (CD4+ T-cell epitopes). Traditional vaccines against EIDs are difficult to produce due to the need for culturing pathogenic viruses *in vitro*. By contrast, EVs have a number of advantages: (i) biosafety: no *in vitro* culturing requirement; (ii) bioprocessing: large-scale production can be carried out economically and rapidly; (iii) selectivity: precise activation of immune response by selecting conserved or immunodominant epitopes, and epitopes

* Corresponding author at: School of Chemistry and Molecular Biosciences, Institute for Molecular Bioscience and Australian Infectious Diseases Research Centre, Q2 University of Queensland, Australia. Tel.: +61 7 3365 2132; fax: +61 7 3365 4699.
** Corresponding author. Tel.: +61 7 3365 2132; fax: +61 7 3365 4699.
E-mail addresses: patricio.oyarzun@uss.cl (P. Oyarzun), b.kobe@uq.edu.au (B. Kobe).

http://dx.doi.org/10.1016/j.vaccine.2015.01.040
0264-410X/

triggering predominantly cellular or humoral responses, and (iv) multivalency: multiple determinants from several pathogens [3]. For EIDs, the geographic distributions of the viruses are often well defined and the ethnic populations in need of vaccination can be determined [4].

The inclusion of CD4+ T-cell epitopes in vaccine formulations is a necessary condition to provide cognate help and thus to induce a vigorous immune response, with optimal CD8+ cytotoxic T-cell responses and neutralizing antibodies [5,6]. The challenge for CD4+ T-cell epitope prediction is that HLA class II proteins (allotypes) are encoded by the most polymorphic genes in the human genome; 2825 HLA class II alleles associated with three classical loci (1649 DR, 716 DQ and 460 DP alleles; IMGT/HLA Database [7], July 2014, Release 3.17.0). This huge diversity presents serious problems for vaccine design, as HLA alleles are expressed at different frequencies in different ethnicities. Individuals that display a different set of alleles, with potentially different binding specificities (HLA restriction), are likely to react to a different set of peptides from a given pathogen. For example, a recent study concluded that the lack of response to a recombinant vaccine designed to induce clade-specific neutralizing antibodies to HIV-1 in Thailand was associated with the presence of certain HLA class II alleles [8].

It is advantageous for EVs to prime immune responses against epitopes that bind to many HLA molecules and are recognized by more than one T-cell clone (here termed promiscuous epitopes) [9]. An established approach to select promiscuous epitopes is based on the concept of supertypes, i.e., clusters of HLA molecules that share overlapping peptide repertories [10,11]. Drawbacks of this approach include the potential skewing of epitope selection to major alleles [12], poor specificity characterization for many alleles within a supertype [13] and lack of agreement on supertype classification [14–16].

Bioinformatics tools are an essential component of a high-throughput pipeline for *in silico* mapping of thousands of potential epitopes, helping reduce the time and cost involved in the experimental testing of such peptides [17]. A computational method for EV design must implement algorithms for epitope discovery (prediction) and selection, and determine the population coverage potentially afforded by a vaccine based on these peptides. Bioinformatics tools for epitope prediction focus on peptide binding to HLA proteins, assuming that T cells with the required specificity will be present in the T cell repertoire. Methods range from approaches entirely based on binding data (data-driven methods) to those based on structural principles and molecular modeling [18]. A group of so-called "pan-specific approaches" has emerged recently, which extend the scope of the prediction toward HLA class II allotypes for which no experimental data are available, including Predivac [19], TEPITOPEpan [20], NetMHCIIpan-3.0 [21] and MultiRTA [22]. Current computational tools to estimate the fraction of individuals that would be protected by putative T-cell epitopes are listed in Table S1.

We have previously developed Predivac [19], a pan-specific bioinformatics tool for CD4+ T-cell epitope prediction that affords almost full coverage of HLA class II proteins associated with the DRB loci. Here, we describe an extension of Predivac for world-wide and ethnicity-specific HLA class II-restricted EV design (Predivac-2.0), which implements the three algorithms required for EV design (peptide binding prediction and selection, population coverage prediction and an optimization of population coverage) into a web-interfaced computational platform. The ability of Predivac-2.0 to pick promiscuous and immunogenic CD4+ T-cell epitopes in virus antigens was confirmed using the Gag protein of HIV. To demonstrate the utility of the tool, we investigated putative CD4+ T-cell epitopes for vaccine design against EIDs caused by Lassa (LASV), Nipah (NiV) and Hendra (HeV) viruses.

2. Methods

2.1. T-cell epitope mapping algorithm

Predivac-2.0 predicts CD4+ T-cell epitopes based on the specificity-determining residue (SDR) approach [19,23,24]. Details are provided in Supplementary materials.

2.2. Promiscuous epitope prediction

The ability of Predivac-2.0 to identify promiscuous and immunodominant regions in antigens was tested using CD4+ T-cell epitope maps of the HIV Gag polyprotein, available in the Los Alamos HIV Molecular Immunology Database (http://www.hiv.lanl.gov/content/immunology/) [25]. Details are provided in Supplementary materials.

2.3. Population coverage algorithm

Predivac-2.0 determines the fraction of individuals that would be potentially covered by the selected epitopes by processing HLA class II allele frequency data retrieved from the "The Allele Frequency Net Database" (AFND; http://www.allelefrequencies.net/) [26], which is the most comprehensive repository of immune gene frequencies of world-wide populations. It defines a target population at four levels: world, geographic regions, countries and ethnicities, consistent with the AFND. Details are provided in Supplementary materials.

2.4. Epitope selection

Predivac-2.0 implements two methods to select CD4+ T-cell epitopes based on population coverage: "simple search" and "optimized search". Details are provided in Supplementary materials. Optimized search is potentially more accurate than simple search. In addition, splitting the global search into the ethnicities making up the target population allows the algorithm to explore a greater number of peptide combinations in order to maximize population coverage. However, this calculation can be substantially slower for populations with a significant mix of ethnicities. The user is given the option of having the results returned via email.

2.5. Study cases

EV design was performed on Lassa and henipavirus surface glycoproteins. Details are provided in Supplementary materials.

3. Results

3.1. Promiscuous epitope prediction

As shown in Fig. 1, Predivac-2.0 can perform the prediction of promiscuous CD4+ T-cell epitopes for one or many pathogen proteins over almost the entire set of HLA class II alleles occurring in any target population. The degree of epitope promiscuity varies depending on the threshold set for the MHC class II allele binding prediction. We suggest setting the threshold at 3% of top scoring peptides, as this is the threshold under which Predivac-2.0 identifies ~75% of immunodominant epitopes [19]. Below the 5% threshold, the percentage of immunodominant epitopes identified only increases marginally, while above 5% the number of predicted epitopes becomes too large to be useful.

Predivac-2.0 predicted three CD4+ T-cell epitopes in the Gag polyprotein sequence, which together have the potential of covering ~75% of the United States population, and correspond to

Please cite this article in press as: Oyarzun P, et al. A bioinformatics tool for epitope-based vaccine design that accounts for human ethnic diversity: Application to emerging infectious diseases. Vaccine (2015), http://dx.doi.org/10.1016/j.vaccine.2015.01.040