Vaccine 33 (2015) 1267–1273

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

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 25 January 2015

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

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

* Corresponding author at: Biotechnology Centre, Facultad de Ingeniería y Tecnología, Universidad San Sebastián, Concepción, Chile. Tel.: +56 41 2487962; fax: +56 41 2400002.
** Corresponding author at: School of Chemistry and Molecular Biosciences, University of Queensland, Cooper Road, Brisbane, Queensland 4072, Australia. 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/

by selecting conserved or immunodominant epitopes, and epitopes 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.

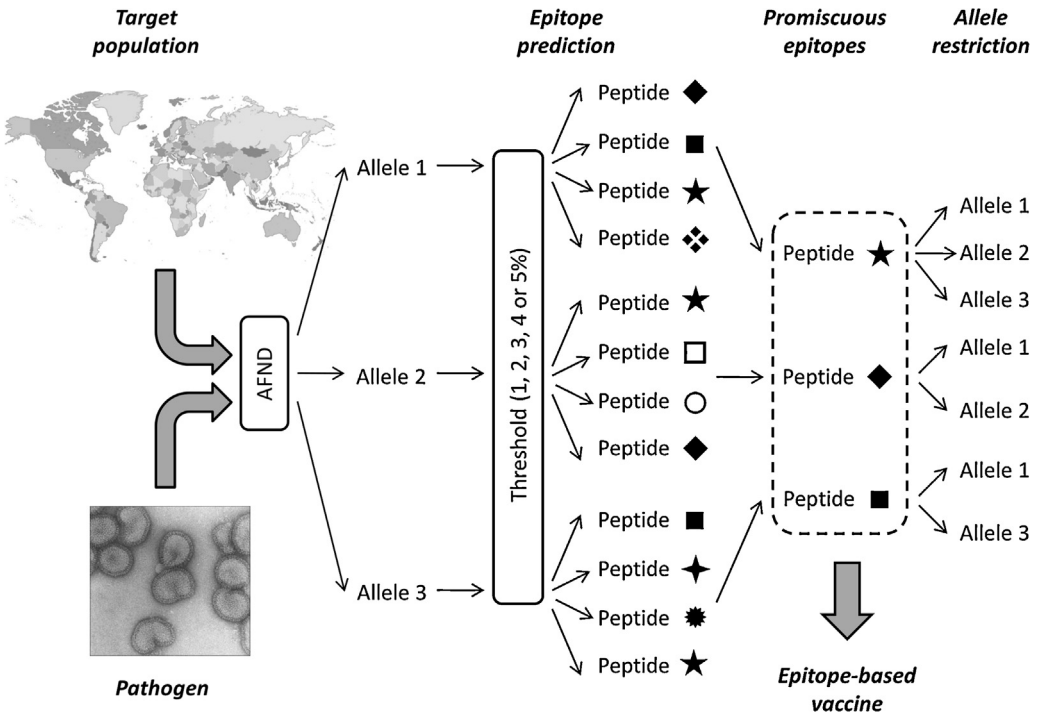

Fig. 1. Schematic illustration of the steps followed by Predivac-2.0 to select promiscuous CD4+ T-cell epitopes for EV design. Upon the user setting the target geographic region, the program retrieves from the AFND all the HLA class II allele frequency data available for population samples occurring in this region and searches the input proteins for promiscuous epitopes restricted to those alleles. In this example, the peptide with the symbol star is the most promiscuous, as it is presented by the three alleles that are expressed in the target population.

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 the most immunodominant regions of the p17 and p24 proteins. One of the epitopes (FAVNPGLLE) was predicted in the center of the promiscuous and immunodominant region 37-ASRELERFAVNPGLLETSEGCR-58 of the p17 protein [27,28], and was not predicted by Multipred2 (Figs. S1 and S3). The four remaining CD4+ T-cell epitopes were predicted in the p24 protein (VQNIQGQMV, IVRMYSPTS, YKTLRAEQA and ILKALGPAA), two of them within the most promiscuous and immunodominant regions of the Gag protein (125-PVGEIYKRWIILGLNKIVRMYSPTSI-150 and 164-YVDRFYKTLRAEQASQEV-182) [27,29,30,28] (Figs. S2 and S4). Multipred2 identified three of these epitopes in p24, although LKALGPAAT was shifted by one residue to yield ILKALGPAA. Predivac-2.0 did not retrieve any potential epitope in the p15 region, while Multipred2 predicted two overlapped nonamers in a non-promiscuous region (Fig. S5).

3.2. Population coverage

A central feature that makes Predivac-2.0 useful for EV design is its capability to predict CD4+ T-cell epitopes for 1147 out of 1211 (95%) HLA class II DR proteins (IMGT/HLA Database release 3.17.0 [7]). Predivac-2.0 takes advantage of this capability by integrating allele frequency information from the AFND, which provides data over 856 population samples (Table S2). Each population sample from the AFND is assigned to a particular ethnic group, but a given ethnic group can be potentially associated with many population samples. Because not all the population samples are reported with high-resolution data (four digits; DRB1*xx:xx), low resolution data (two digits; e.g. DRB1*xx) were assigned by default to the most prevalent allele subtype 01 (e.g. DRB1*xx:01) [31]. Together, the wide coverage of HLA class II DRB alleles and frequencies across numerous populations with different ethnic backgrounds enables Predivac-2.0 to account comprehensively for human genetic diversity in EV design. Future development of the method to include DQ and DP loci would potentially lead to an improvement in population coverage, as the current tool only accounts for the DR locus.

As shown in Fig. 2, Predivac-2.0 enables either the prediction of CD4+ T-cell epitopes restricted to one particular allele, or EV design for a given geographic region, including countries or ethnic groups. The program retrieves from the AFND the HLA class II allele frequencies belonging to the population samples occurring in the geographic region set by the user, and predicts promiscuous epitopes. The user is prompted to define whether the search for epitopes will be optimized (genetic algorithm) or not (simple search; the default option). Epitope validation is an optional step that can be used when the allele restriction of the epitopes has been experimentally tested.

3.3. Applications

EV design was performed using the surface glycoproteins of LASV and henipaviruses, chosen based on their ability to induce resistance to viral infections and elicit potent neutralizing antibodies against these pathogens. The analysis focused on the geographic regions associated with the respective viruses (target population; Table 1).

The envelope glycoproteins in LASV include GP1, a peripheral membrane protein, and GP2, a transmembrane protein with an ectodomain, both of which assemble into homotrimers [32]. It has been demonstrated in primates, using vaccinia [33] and vesicular stomatitis virus vectors [34], that LASV glycoproteins (GP1 and GP2) protect against lethal challenge. Data from human and animal studies indicate a dominant role of T-cells over antibodies in controlling acute LASV infection and providing immunity to re-infection [35]. A key protective effect of CD4+ T cells has been described in other arenaviruses (e.g. lymphocytic choriomeningitis virus) in sustaining optimal CD8+ T-cell activity, containing viremia [36] and

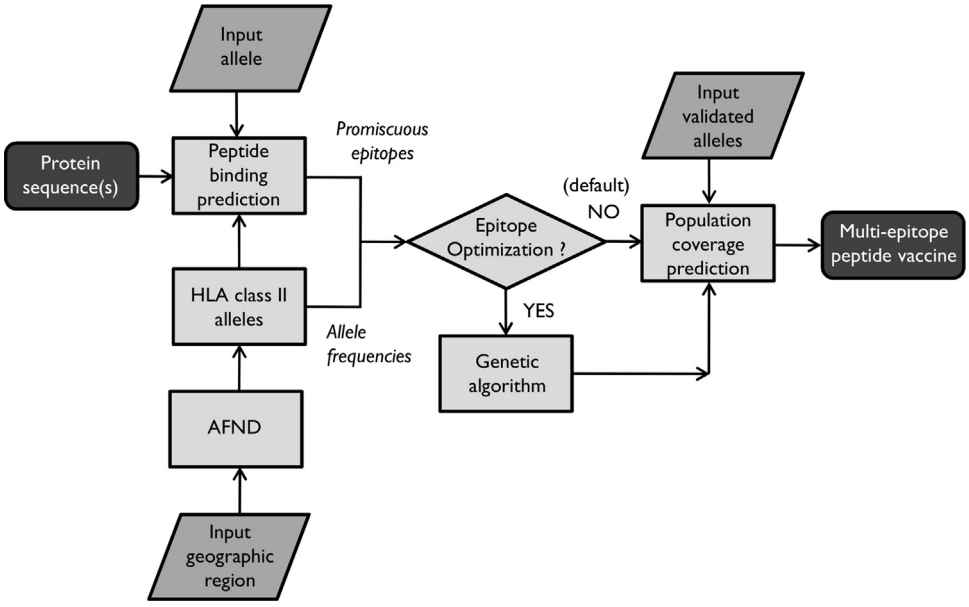

Fig. 2. Flow-chart representing algorithms that Predivac 2.0 implements to perform epitope-based vaccine design. The process involves the following three algorithms: (i) peptide binding prediction; (ii) population coverage prediction and (iii) optimization of population coverage. The two conditional steps are: (i) epitope optimization and (ii) epitope validation. By default epitopes are selected without optimization (simple search). Optionally, information about HLA restriction of the epitopes can be entered in order to calculate a more accurate population coverage assessment associated with the vaccine.

providing direct cytotoxic function [37]. It has been also shown that cross-reactive GP2-derived epitopes are involved in human CD4+ T lymphocyte responses against LASV [38].

The GP2 glycoprotein of LASV was screened in search of promiscuous CD4+ T-cell epitopes that potentially afford protection for the West African population at risk of contracting Lassa fever (Table 1). The HLA class II allele frequency distribution of the region formed by these populations was calculated by Predivac-2.0 (Table S3). Fig. 3 shows that using the simple and optimized search, respectively, ~97% and ~99% of population coverage can be achieved, respectively, with four epitopes in each case (Table S4, Fig. S6).

Henipaviruses have spikes of F (fusion) glycoprotein trimers and G (attachment) glycoproteins within their lipid membranes. The native configuration of the G protein is predominantly tetrameric (dimer of dimers) with some dimeric species. In animal model vaccination studies using HeV recombinant soluble G glycoprotein ectodomain, both felines [39,40] and ferrets [41] survived HeV lethal challenges with no clinical signs of disease. The use of HeV-G as a potential subunit vaccine for preventing both HeV and NiV infection is underscored by the fact that immunization with this protein elicits a potent cross-reactive neutralizing antibody response against both henipaviruses [42]. The protective effects are in part humoral, as passive protection against HeV and NiV infection has been demonstrated [43]. HeV-G and NiV-G exhibit receptor-induced monoclonal antibody epitopes that cross-react with other henipavirus G proteins, suggesting they target a conserved functional epitope [44,45].

HeV-G and NiV-G were analyzed using Predivac-2.0 to find promiscuous CD4+ T-cell epitopes that are potentially protective against henipaviruses in the entire risk region, which comprises a number of countries in the South East Asia and the Pacific regions (Table 1). Predivac-2.0 built the target population based on 81 population samples occurring in this region, for which AFND contains HLA class II DRB allele frequency information (Table S5). The HeV and NiV proteins share 79% sequence identity (Fig. S7), therefore; a prediction was carried out to determine to what extent an EV with peptides from conserved regions of the G glycoproteins could provide protection. Fig. 4 shows that ~93% (simple search) and ~99% (optimized search; Table S6) population coverage can be reached with five epitopes whose sequence is conserved in HeV-G and NiV-G (Fig. S8).

A separate analysis was carried out for HeV and NiV by considering regions that are 100% conserved across all the strains (Fig. 5, Fig. S8). For NiV-G, considering the entire population at risk

Table 1
Geographic distribution of henipaviruses [50] and LASV [51] and the associated disease, mortality and biosafety level information.

Genus	Virus	Countries[a]	Geographic region	Disease	Mortality (%)	Biosafety level
Henipavirus	Hendra (HeV)	Australia	Australia	Respiratory and neurological disease	50–75	4
	Nipah (NiV)	Bangladesh, Malaysia, India, Singapore, Thailand[a], Indonesia[a], Cambodia[a], Papua New Guinea[a], Solomon, Islands[a], Timor-Leste[a]	South East Asia and the Pacific (excluding Australia)	Respiratory and neurological disease	50–75	4
Arenavirus	Lassa virus (LASV)	Sierra Leona[a], Nigeria[a], Liberia[a], Guinea-Bissau[a], Benin[a], Burkina Faso[a], Congo[a], Mali[a], Cote d'Ivoire[a], Ghana, Togo, Benin, Cameroon, Central African Republic	West Africa	Hemorrhagic fever (Lassa fever)	15–20	4

[a] Countries where either serological evidence or outbreaks have been detected.

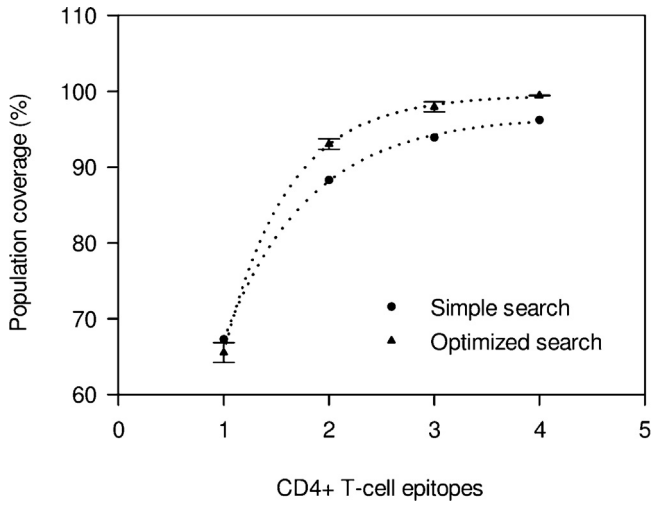

Fig. 3. Population coverage prediction for an epitope-based vaccine against Lassa virus. The target population is the West Africa's population at risk of contracting Lassa fever. Predictions were performed over six population samples from individuals with different ethnic backgrounds living in the risk area, for which the AFND contains HLA class II DR allele frequency information (from the Central African Republic, Cameroon, $n = 126$; Guinea Bissau, $n = 65$; Aka Pygmy, $n = 93$; and the ethnicities from Burkina Faso, Rimaibe, $n = 47$; Mossi, $n = 53$; and Fulni, $n = 49$). The x-axis corresponds to the number of epitopes to be potentially included in the multi-epitope peptide vaccine, using the simple (circles) and optimized (triangles) searches. For the optimized search, the predictions were run five times, indicating the average and standard deviation (bars). Epitopes were predicted based exclusively on 100% conserved regions of the GP2 glycoprotein for all the virus strains.

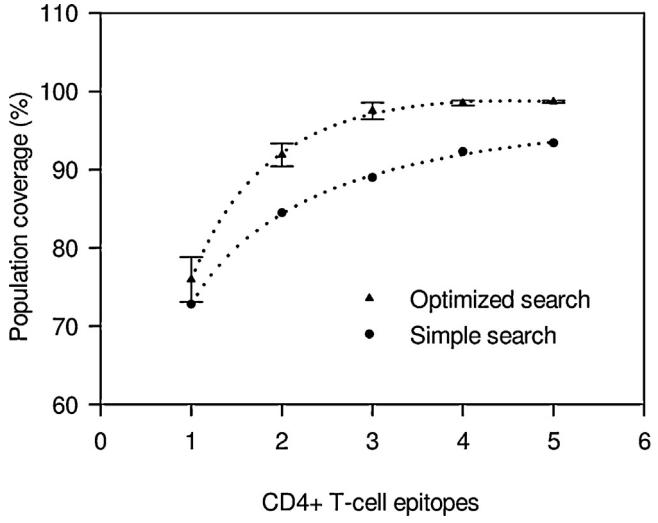

Fig. 4. Population coverage prediction for en epitope-based vaccine against henipaviruses. The target population is the whole region at risk of henipaviruses (HeV and NiV), comprised by the South East and Pacific countries, including Australia. The x-axis corresponds to the number of epitopes to be potentially considered in the multi-epitope peptide vaccine, using the simple (circle) and optimized (triangle) searches. For the optimized search, the predictions were run five times, indicating the average and standard deviation (bars). Epitopes were predicted exclusively on 100% conserved regions of the G glycoproteins for all the virus strains. CD4+ T-cell epitopes predicted with the simple search correspond to 50-FNTVIALLG-58, 67-IMIIQNYTR-75, 101-IGTEIGPKV-109, 120-IPANIGLLG-128 and 252-VGEVLDRGD-260. Four out of five of these peptides potentially protective against both henipaviruses are found in a densely antigenic region of the stalk domain of the G glycoprotein. This is the same region of the influenza virus HA glycoprotein that is being studied for the development of a universal influenza vaccine, owing to its high degree of sequence conservation [51]. Unlike the globular head of the HA, which undergoes significant antigenic variation, vaccination with headless HA has proven to be able to induce anti-stalk domain broadly neutralizing antibodies and confer protection against different virus sub-types [52].

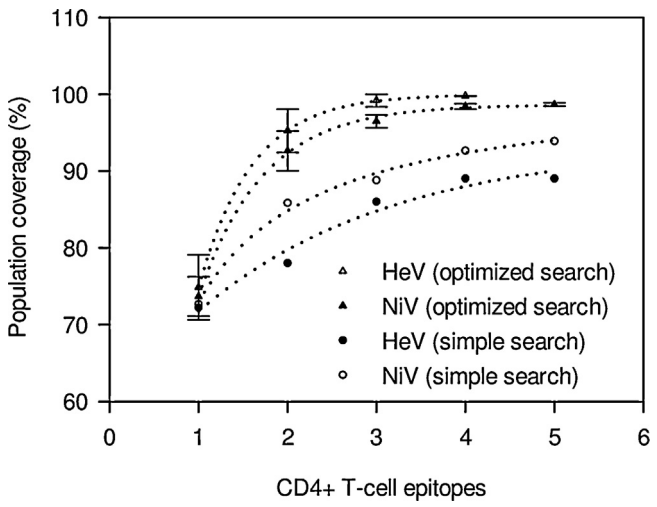

Fig. 5. Population coverage prediction for epitope-based vaccines against Hendra and Nipah viruses. The target populations correspond to: (i) the countries at risk of the Nipah virus in the South East and Pacific region (excluding Australia) (circles), and (ii) the region at risk of the Hendra virus (Australia) (triangles). The x-axis corresponds to the number of epitopes to be potentially considered in a multi-epitope peptide vaccine targeting these populations, using the simple (black) and optimized (white) searches. For the optimized search, the predictions were run five times, indicating the average and standard deviation (bars). Epitopes were predicted exclusively on 100% conserved regions of the G glycoprotein for all the virus strains. The putative epitopes predicted with the simple search are situated in the HeV-G stalk domain (7-LVSLNNNLS-15 and 50-FNTVIALLG-58) and in the globular domain (234-LEKIGSCTR-242 and 310-WTESLSLIR-318). The peptide 234-LEKIGSCTR-242 is next to a previously identified discontinuous B-cell epitope [53] (Fig. S8).

in the South-East Asia and Pacific regions (Table 1), five epitopes were predicted to cover ~94% (simple search) or ~99% (optimized search; Table S7) of the target population. The epitopes retrieved using the simple search are the same than those retrieved for both henipaviruses on the entire risk region. For HeV, Predivac-2.0 predicted a maximum coverage of ~89% (simple search) in the Australian population based on six peptides, while the optimized search retrieved four epitopes that would provide ~99% protection (Table S8).

As cross-reactivity with self-antigens may lead to deleterious immune responses, epitope sequences were evaluated for similarity to the human proteins by doing a BLAST analysis of the human proteome. It was found that none of the putative nonameric epitopes retrieved by the simple and optimized search (LASV and henipaviruses) shares more than seven identities with sequences in human proteins.

4. Discussion

The prediction of CD4+ T-cell epitopes by Predivac-2.0 in the HIV Gag protein confirmed the ability of the method to accurately locate promiscuous and immunodominant epitopes. As we have shown previously [19], the method relies on high-affinity peptide:MHC class II interactions, which contributes to its ability to identify highly immunogenic sequences. This result supports the correlation between high-affinity peptide binding, promiscuity and immunodominance, both for MHC class I [46] and class II proteins [47]. The immunodominant regions highlighted within the Gag protein were reported based on studies involving HIV-seropositive individuals from several ethnic backgrounds (Figs. S3 and S4). The fact that all of these sequences overlap around the same regions is a direct consequence of their wide promiscuity, as individuals expressing a different array of HLA class II allelic variants will be able to recognize these epitopes. The performance comparison with the state-of-the-art method Multipred2 showed

that despite their different underlying concepts, they delivered comparable predictions that matched the two most immunodominant regions of the p24 protein. See Supplementary materials for further discussion of promiscuous epitope prediction and the different approaches for population coverage prediction.

The default method that Predivac-2.0 uses for population coverage prediction (simple search) is faster computationally, but assumes that all the ethnicities making up the target population are breeding with one another, through taking into account all possible diploid combinations from the allele frequencies occurring in those ethnicities to obtain the allele distribution. The same assumption is inherent in the IEDB population coverage tool [48]. However, this assumption is not necessarily accurate when it comes to geographically distant ethnicities. Experimental evidence indicates that allele frequencies tend to vary substantially [49,50]. In this context, the optimized search is probably more accurate than the simple search, because it analyses every population in an independent manner, i.e. diploid combinations are calculated specifically for each particular ethnic group. Furthermore, this search mode enables a more detailed exploration of epitope combinations to maximize population coverage. Optimized search of CD4+ T-cell epitopes in surface glycoproteins of the three viruses analyzed delivered higher population coverage with a potentially smaller number of associated epitopes (Figs. 3–5). However, while the simple search calculates coverage over a single population, thus retrieving a unique set of putative epitopes, the optimized search does it simultaneously with many populations. This greater degree of freedom leads to more than one solution (epitope combination) for the same optimization problem.

To test the utility of our method, we applied Predivac-2.0 to the identification of potentially protective CD4+ T-cell epitopes in the GP2 protein of LASV, a highly pathogenic arenavirus that causes 300,000–500,000 cases of disease annually. This is, therefore, a serious public health problem that justifies the development of a prophylactic vaccination strategy that is particularly effective in targeting the ethnic populations living in this area. Although protective immunity against LASV is clearly associated with T-cell responses, only a few human T-cell epitopes have been experimentally characterized (Fig. S6). Thus, the five epitopes predicted by Predivac-2.0 are expected to be helpful to guide EV design targeting the West Africa's population. See Supplementary materials for further discussion of the epitopes.

For henipaviruses, the lack of vaccines is of particular concern due to its case fatality rate and highly contagious nature, properties that render them bioterrorism threats. The development of an effective EV against these viruses must ideally induce protection against both NiV and HeV. One strategy is to use promiscuous CD4+ T-cell epitopes that lie in conserved regions of the G glycoprotein, to decrease their likelihood of mutations to escape T-cell surveillance (so-called antigenic drift). According to our results (Fig. 4), both the simple and optimized searches predicted that at least 90% of population coverage could be achieved with five epitopes that target simultaneously the populations at risk of HeV and NiV. Another potentially effective approach for vaccine design against henipaviruses would be to combine, into a single formulation, the ten putative epitopes that were independently predicted using the optimized search for HeV-G (four epitopes) and NiV-G (six epitopes), which together would afford coverage of 99% of the respective risk population. This strategy of assembling an EV from epitopes accounting independently for conserved regions of different virus subgroups seems particularly appealing for hypervariable viruses having multiple serotypes or clades (such as HIV and influenza). For example, the fact that CD4+ T-cell epitopes predicted for NiV (simple search) over the entire risk region (southeast Asia and the Pacific, without Australia) are the same peptides as those predicted for both NiV and HeV over the same region plus Australia, is a consequence of the significantly higher variability of the NiV-G protein sequence in comparison to HeV. Thus, the few available conserved regions in both henipavirus G protein sequences are similar to those in the NiV G protein (Fig. S7). It is conceivable that new sequenced viral strains may present mutations in these regions of the G protein, making the design an EV that simultaneously covers the whole viral diversity no longer feasible. Thus far, Predivac-2.0 only accounts for the variability imposed by HLA molecules; the selection of conserved regions for epitope prediction was done manually. See Supplementary materials for a further discussion of the application of Predivac-2.0 to the EV design for Lassa and henipaviruses.

EIDs are becoming a significant burden on global economies. As seen in the 2014 Ebola outbreak in West Africa, EIDs can quickly spread worldwide due to increasing global travel. An important aspect involves focussing on tacking initial outbreaks (epidemics), rather than fighting diseases on a global scale (pandemics). We believe that Predivac-2.0 has great potential to contribute to such efforts by identifying epitopes protective for particular populations at risk of EIDs.

5. Conclusions

Population coverage and the associated HLA polymorphism are major issues for EV vaccine design, when the aim is to induce broad immune responses in genetically diverse human populations. Predivac-2.0 was developed to address this problem by implementing a computational framework for rational design of vaccines based on multiple CD4+ T-cell epitopes that target well-defined ethnic populations. We term this concept the "ethnicity-oriented approach". HLA polymorphism is handled by favoring the inclusion of broadly recognized immunodominant CD4+ T-cell epitopes having the ability to potentially trigger a broad CD4+ T-cell response in a high proportion of individuals expressing distinct HLA molecules. We demonstrated that Predivac-2.0 is capable of identifying CD4+ T-cell epitopes within highly promiscuous and immunodominant regions of HIV Gag protein. As the method accounts comprehensively for human ethnic diversity, we propose that it is particularly suited as a tool to aid EV design in the context of EIDs. Consequently, Predivac-2.0 was applied to map epitopes in surface glycoproteins of three emerging viruses, for the specific ethnic populations at risk of contracting the corresponding diseases. The predicted epitopes are suitable candidates to be experimentally tested, as they hold the potential to provide cognate help in vaccination settings in these particular geographic regions. Overall, the Predivac-2.0 method sets the basis for a novel approach in EV design that potentially overcomes some of the drawbacks associated with the supertype-based approach, particularly for infectious diseases associated with geographic regions where the ethnic background of the target population can be determined.

Author contributions

Conceived and designed the experiments: PO, JJE, BK. Performed the experiments: PO, JJE. Analyzed the data: PO, JJE, MB, BK. Contributed research tools: FFG, ARJ, DM. Wrote the paper: PO, BK.

Acknowledgment

We thank the members of the Kobe lab for valuable discussions.

Conflict of interest statement

The authors declare that no conflict of interest exists.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.vaccine.2015.01.040.

References

[1] Jones KE, Patel NG, Levy MA, Storeygard A, Balk D, Gittleman JL, et al. Global trends in emerging infectious diseases. Nature 2008;451:990–3.
[2] Morens DM, Folkers GK, Fauci AS. The challenge of emerging and re-emerging infectious diseases. Nature 2004;430:242–9.
[3] Purcell AW, McCluskey J, Rossjohn J. More than one reason to rethink the use of peptides in vaccine design. Nat Rev Drug Discov 2007;6:404–14.
[4] Daszak P, Cunningham AA, Hyatt AD. Emerging infectious diseases of wildlife – threats to biodiversity and human health. Science 2000;287:443–9.
[5] Ribeiro SP, Rosa DS, Fonseca SG, Mairena EC, Postol E, Oliveira SC, et al. A vaccine encoding conserved promiscuous HIV CD4 epitopes induces broad T cell responses in mice transgenic to multiple common HLA class II molecules. PLoS ONE 2010;5:e11072.
[6] Rosa DS, Ribeiro SP, Cunha-Neto E. CD4+ T cell epitope discovery and rational vaccine design. Arch Immunol Ther Exp (Warsz) 2010;58:121–30.
[7] Gonzalez-Galarza FF, Lawless C, Hubbard SJ, Fan J, Bessant C, Hermjakob H, et al. A critical appraisal of techniques, software packages, and standards for quantitative proteomic analysis. Omics 2012;16:431–42.
[8] Paris R, Bejrachandra S, Thongcharoen P, Nitayaphan S, Pitisuttithum P, Sambor A, et al. HLA class II restriction of HIV-1 clade-specific neutralizing antibody responses in ethnic Thai recipients of the RV144 prime-boost vaccine combination of ALVAC-HIV and AIDSVAX((R)) B/E. Vaccine 2012;30:832–6.
[9] Khan AM, Miotto O, Heiny AT, Salmon J, Srinivasan KN, Nascimento EJ, et al. A systematic bioinformatics approach for selection of epitope-based vaccine targets. Cell Immunol 2006;244:141–7.
[10] Sette A, Sidney J. HLA supertypes and supermotifs: a functional perspective on HLA polymorphism. Curr Opin Immunol 1998;10:478–82.
[11] Sette A, Sidney J. Nine major HLA class I supertypes account for the vast preponderance of HLA-A and -B polymorphism. Immunogenetics 1999;50:201–12.
[12] Lundegaard C, Lund O, Kesmir C, Brunak S, Nielsen M. Modeling the adaptive immune system: predictions and simulations. Bioinformatics 2007;23:3265–75.
[13] Hillen N, Mester G, Lemmel C, Weinzierl AO, Muller M, Wernet D, et al. Essential differences in ligand presentation and T cell epitope recognition among HLA molecules of the HLA-B44 supertype. Eur J Immunol 2008;38:2993–3003.
[14] Doytchinova IA, Flower DR. In silico identification of supertypes for class II MHCs. J Immunol 2005;174:7085–95.
[15] Greenbaum J, Sidney J, Chung J, Brander C, Peters B, Sette A. Functional classification of class II human leukocyte antigen (HLA) molecules reveals seven different supertypes and a surprising degree of repertoire sharing across supertypes. Immunogenetics 2011;63:325–35.
[16] Saha I, Mazzocco G, Plewczynski D. Consensus classification of human leukocyte antigen class II proteins. Immunogenetics 2013;65:97–105.
[17] Sirskyj D, Diaz-Mitoma F, Golshani A, Kumar A, Azizi A. Innovative bioinformatic approaches for developing peptide-based vaccines against hypervariable viruses. Immunol Cell Biol 2011;89:81–9.
[18] Tsurui H, Takahashi T. Prediction of T-cell epitope. J Pharmacol Sci 2007;105:299–316.
[19] Oyarzun P, Ellis JJ, Boden M, Kobe B. PREDIVAC: CD4+ T-cell epitope prediction for vaccine design that covers 95% of HLA class II DR protein diversity. BMC Bioinform 2013;14:52.
[20] Zhang L, Chen Y, Wong HS, Zhou S, Mamitsuka H, Zhu S. TEPITOPEpan: extending TEPITOPE for peptide binding prediction covering over 700 HLA-DR molecules. PLoS ONE 2012;7:e30483.
[21] Karosiene E, Rasmussen M, Blicher T, Lund O, Buus S, Nielsen M. NetMHCIIpan-3.0, a common pan-specific MHC class II prediction method including all three human MHC class II isotypes, HLA-DR, HLA-DP and HLA-DQ. Immunogenetics 2013;65:711–24.
[22] Bordner AJ, Mittelmann HD. MultiRTA: a simple yet reliable method for predicting peptide binding affinities for multiple class II MHC allotypes. BMC Bioinform 2010;11:482.
[23] Brinkworth RI, Breinl RA, Kobe B. Structural basis and prediction of substrate specificity in protein serine/threonine kinases. Proc Natl Acad Sci U S A 2003;100:74–9.
[24] Kobe B, Boden M. Computational modelling of linear motif-mediated protein interactions. Curr Top Med Chem 2012;12:1553–61.
[25] Korber B, Brander C, Walker B, Koup R, Moore J, Haynes B, et al. HIV molecular immunology database. Los Alamos, NM: Los Alamos National Laboratory, Theoretical Biology and Biophysics; 1995 [LA-UR 95-4371].
[26] Gonzalez-Galarza FF, Christmas S, Middleton D, Jones AR. Allele frequency net: a database and online repository for immune gene frequencies in worldwide populations. Nucleic Acids Res 2011;39:D913–9.
[27] Ramduth D, Day CL, Thobakgale CF, Mkhwanazi NP, de Pierres C, Reddy S, et al. Immunodominant HIV-1 Cd4+ T cell epitopes in chronic untreated clade C HIV-1 infection. PLoS ONE 2009;4:e5013.
[28] Kaufmann DE, Bailey PM, Sidney J, Wagner B, Norris PJ, Johnston MN, et al. Comprehensive analysis of human immunodeficiency virus type 1-specific CD4 responses reveals marked immunodominance of gag and nef and the presence of broadly recognized peptides. J Virol 2004;78:4463–77.
[29] Vingert B, Perez-Patrigeon S, Jeannin P, Lambotte O, Boufassa F, Lemaitre F, et al. HIV controller CD4+ T cells respond to minimal amounts of Gag antigen due to high TCR avidity. PLoS Pathog 2010;6:e1000780.
[30] Ondondo BO, Rowland-Jones SL, Dorrell L, Peterson K, Cotten M, Whittle H, et al. Comprehensive analysis of HIV Gag-specific IFN-gamma response in HIV-1- and HIV-2-infected asymptomatic patients from a clinical cohort in The Gambia. Eur J Immunol 2008;38:3549–60.
[31] Luoni G, Verra F, Arca B, Sirima BS, Troye-Blomberg M, Coluzzi M, et al. Anti-malarial antibody levels and IL4 polymorphism in the Fulani of West Africa. Genes Immun 2001;2:411–4.
[32] Eschli B, Quirin K, Wepf A, Weber J, Zinkernagel R, Hengartner H. Identification of an N-terminal trimeric coiled-coil core within arenavirus glycoprotein 2 permits assignment to class I viral fusion proteins. J Virol 2006;80:5897–907.
[33] Fisher-Hoch SP, Hutwagner L, Brown B, McCormick JB. Effective vaccine for Lassa fever. J Virol 2000;74:6777–83.
[34] Geisbert TW, Jones S, Fritz EA, Shurtleff AC, Geisbert JB, Liebscher R, et al. Development of a new vaccine for the prevention of Lassa fever. PLoS Med 2005;2:e183.
[35] Fisher-Hoch SP, McCormick JB. Towards a human Lassa fever vaccine. Rev Med Virol 2001;11:331–41.
[36] Matloubian M, Concepcion RJ, Ahmed R. CD4+ T cells are required to sustain CD8+ cytotoxic T-cell responses during chronic viral infection. J Virol 1994;68:8056–63.
[37] Jellison ER, Kim SK, Welsh RM. Cutting edge: MHC class II-restricted killing in vivo during viral infection. J Immunol 2005;174:614–8.
[38] Meulen J, Badusche M, Satoguina J, Strecker T, Lenz O, Loeliger C, et al. Old and New World arenaviruses share a highly conserved epitope in the fusion domain of the glycoprotein 2, which is recognized by Lassa virus-specific human CD4+ T-cell clones. Virology 2004;321:134–43.
[39] Mungall BA, Middleton D, Crameri G, Bingham J, Halpin K, Russell G, et al. Feline model of acute nipah virus infection and protection with a soluble glycoprotein-based subunit vaccine. J Virol 2006;80:12293–302.
[40] McEachern JA, Bingham J, Crameri G, Green DJ, Hancock TJ, Middleton D, et al. A recombinant subunit vaccine formulation protects against lethal Nipah virus challenge in cats. Vaccine 2008;26:3842–52.
[41] Pallister J, Middleton D, Wang LF, Klein R, Haining J, Robinson R, et al. A recombinant Hendra virus G glycoprotein-based subunit vaccine protects ferrets from lethal Hendra virus challenge. Vaccine 2011;29:5623–30.
[42] Bossart KN, Crameri G, Dimitrov AS, Mungall BA, Feng YR, Patch JR, et al. Receptor binding, fusion inhibition, and induction of cross-reactive neutralizing antibodies by a soluble G glycoprotein of Hendra virus. J Virol 2005;79:6690–702.
[43] Guillaume V, Contamin H, Loth P, Grosjean I, Courbot MC, Deubel V, et al. Antibody prophylaxis and therapy against Nipah virus infection in hamsters. J Virol 2006;80:1972–8.
[44] Aguilar HC, Ataman ZA, Aspericueta V, Fang AQ, Stroud M, Negrete OA, et al. A novel receptor-induced activation site in the Nipah virus attachment glycoprotein (G) involved in triggering the fusion glycoprotein (F). J Biol Chem 2009;284:1628–35.
[45] Zhu Z, Dimitrov AS, Bossart KN, Crameri G, Bishop KA, Choudhry V, et al. Potent neutralization of Hendra and Nipah viruses by human monoclonal antibodies. J Virol 2006;80:891–9.
[46] Eisen HN, Hou XH, Shen C, Wang K, Tanguturi VK, Smith C, et al. Promiscuous binding of extracellular peptides to cell surface class I MHC protein. Proc Natl Acad Sci U S A 2012;109:4580–5.
[47] Weaver JM, Lazarski CA, Richards KA, Chaves FA, Jenks SA, Menges PR, et al. Immunodominance of CD4T cells to foreign antigens is peptide intrinsic and independent of molecular context: implications for vaccine design. J Immunol 2008;181:3039–48.
[48] Bui HH, Sidney J, Dinh K, Southwood S, Newman MJ, Sette A. Predicting population coverage of T-cell epitope-based diagnostics and vaccines. BMC Bioinform 2006;7:153.
[49] Gonzalez-Galarza FF, Mack SJ, Hollenbach J, Fernandez-Vina M, Setterholm M, Kempenich J, et al. 16(th) IHIW: extending the number of resources and bioinformatics analysis for the investigation of HLA rare alleles. Int J Immunogenet 2013;40:60–5.
[50] Middleton D, Gonzalez F, Fernandez-Vina M, Tiercy JM, Marsh SG, Aubrey M, et al. A bioinformatics approach to ascertaining the rarity of HLA alleles. Tissue Antigens 2009;74:480–5.
[51] Steel J, Lowen AC, Wang TT, Yondola M, Gao Q, Haye K, et al. Influenza virus vaccine based on the conserved hemagglutinin stalk domain. MBio 2010;1, pii:e00018-10.
[52] Ekiert DC, Friesen RH, Bhabha G, Kwaks T, Jongeneelen M, Yu W, et al. A highly conserved neutralizing epitope on group 2 influenza A viruses. Science 2011;333:843–50.
[53] White JR, Boyd V, Crameri GS, Duch CJ, van Laar RK, Wang LF, et al. Location of, immunogenicity of and relationships between neutralization epitopes on the attachment protein (G) of Hendra virus. J Gen Virol 2005;86:2839–48.