INTERNATIONAL JOURNAL OF
IMMUNOGENETICS

Review

doi: 10.1111/iji.12214

Computer-aided design of T-cell epitope-based vaccines: addressing population coverage

P. Oyarzun* & B. Kobe†

Summary

Epitope-based vaccines (EVs) make use of short antigen-derived peptides corresponding to immune epitopes, which are administered to trigger a protective humoral and/or cellular immune response. EVs potentially allow for precise control over the immune response activation by focusing on the most relevant – immunogenic and conserved – antigen regions. Exper imental screening of large sets of peptides is time-consuming and costly; therefore, *in silico* methods that facilitate T-cell epitope mapping of protein antigens are paramount for EV development. The prediction of T-cell epitopes focuses on the peptide presentation process by proteins encoded by the major histocompatibility complex (MHC). Because different MHCs have different specificities and T-cell epitope repertoires, individuals are likely to respond to a different set of peptides from a given pathogen in genetically heterogeneous human populations. In addition, protective immune responses are only expected if T-cell epitopes are restricted by MHC proteins expressed at high frequencies in the target population. Therefore, without careful consideration of the specificity and prevalence of the MHC proteins, EVs could fail to adequately cover the target population. This article reviews state-of-the-art algorithms and computational tools to guide EV design through all the stages of the process: epitope prediction, epitope selection and vaccine assembly, while optimizing vaccine immunogenicity and coping with genetic variation in humans and pathogens.

* Biotechnology Centre, Facultad de Ingeniería y Tecnología, Universidad San Sebastián, Concepción, Chile and † School of Chemistry and Molecular Biosciences, Institute for Molecular Bioscience and Australian Infectious Diseases Research Centre, University of Queensland, Brisbane, QLD, Australia

Received 19 March 2015; revised 28 April 2015; accepted 25 May 2015

Correspondence: Patricio Oyarzun, Biotechnology Centre, Facultad de Ingeniería y Tecnología, Universidad San Sebastián, Campus Las Tres Pascualas, Lientur 1457, Concepción, Chile. Tel: +56 41 2487962; Fax: +56 41 400002; E-mail: patricio.oyarzun@uss.cl
Bostjan Kobe, School of Chemistry and Molecular Biosciences, University of Queeland, Brisbane, Qld 4072, Australia. Tel: +61 7 3365 2132; Fax: +61 3365 4699; E-mail: b.kobe@uq.edu.au

Introduction

The adaptive immune system comprises two arms: the humoral immune response, which is mediated by secreted antibodies produced in B cells, and the cell-mediated immune response, mainly orchestrated by T cells. Epitopes are antigenic determinants that are recognized by B cells (B-cell epitopes) or T cells (T-cell epitopes). B-cell receptors recognize epitopes exposed directly on the surface of native protein antigens, while T-cell receptors (TCRs) recognize epitopes that are associated with major histocompatibility complexes (MHCs; called human leucocyte antigen (HLA) in humans) in a phenomenon known as MHC restriction. The interaction of the peptide: MHC complexes (pMHCs) with TCRs activates T cells, leading to the stimulation of the adaptive immune response. Two subpopulations of T cells are involved in epitope recognition: T-cells bearing coreceptor molecules CD8 (cytotoxic T lymphocytes – CTLs) and CD4 (helper T lymphocytes – HTLs). CTLs recognize intracellular peptides presented by MHC class I molecules (CTL epitopes) and HTL recognize peptides from the extracellular space that are displayed by MHC class II molecules (HTLs epitopes). The knowledge of epitopes recognized by CTLs and HTLs is critical for the development of effective vaccination strategies against infectious diseases (Rosa *et al.*, 2010).

Mass immunization strategies based on most prevalent pathogenic serotypes have been historically successful; however, more sophisticated approaches are needed to effectively deal with genetic variation both in pathogens and in humans. Vaccine design in the context of genetically heterogeneous human populations faces two major problems: first, individuals displaying a different set of alleles, with potentially different binding specificities, are likely to react with a different set of peptides from a given pathogen; and second, alleles are expressed at dramatically different frequencies in different ethnicities. HLA genes are the most polymorphic in the human genome, with a year-to-year growing list of HLA alleles, according to the IMGT/HLA database (Robinson *et al.*, 2015). A fraction of the HLA alleles contains synonymous mutations or amino acid differences occurring in

noncoding regions, leading to the same gene product. Figure 1 illustrates both the total number of HLA alleles and the number of distinct HLA proteins for each locus. This vast diversity represents a significant challenge for EV design, as demonstrated for example for HIV (Goulder & Watkins, 2008).

Computational tools can be valuable in dealing with these issues in vaccine design. Available computational methods for T-cell epitope vaccine design mostly focus on the stage of epitope prediction of peptide binding to MHCs. A lesser number of tools and algorithms have been developed to guide the selection of putative epitopes, either by maximizing coverage in the target population and/or in terms of pathogen diversity, and to optimize the design of polypeptide vaccine constructs.

The epitope-based approach

This approach focuses on the administration of synthetic antigen-derived peptides containing minimal epitopes with the potential to trigger humoral (B-cell epitopes) and/or cellular (T-cell epitopes) immune responses. Epitope-based vaccines (EVs) offer several potential advantages over traditional vaccines, including the ability to: (i) focus the immune response on conserved and immunogenic antigen regions, affording multistrain protection against rapidly mutating pathogens; (ii) use multiple epitopes from one or several pathogens (multivalency), allowing the induction of large repertoires of immune specificities; and (iii) precisely control the immune response activation, by stimulating different subpopulations of lymphocytes that lead selectively to humoral and/or cellular immune responses (Purcell *et al.*, 2007). Thus far, EVs have not reached the pharma market; however, a significant number of clinical trials involving peptide-based strategies are underway (in phase I (270), phase II (224) and phase III (12) stages in mid-2014) (Li *et al.*, 2014).

While efficacy in traditional vaccination is usually correlated with neutralizing antibodies, the combination of epitopes targeting HLA molecules can yield a more prominent role for cell-mediated immune responses, especially important for viruses (Guo *et al.*, 2011) and intracellular bacteria (Chaitra *et al.*, 2007). HLA molecules are cell-surface heterodimeric glycoproteins. Despite differences in domain organization, HLA class I and class II molecules shape a peptide-binding groove that interacts with antigenic peptides for display to T cells. The groove has several distinct pockets, in which polymorphic residues give rise to differential peptide-binding specificities that determine the antigenic peptide repertoire (Matsui *et al.*, 1994). Unlike HLA class I, the groove in HLA class II molecules lacks interactions that close it at both ends. Consequently, its ligands are longer (9–22 mers) than HLA class I ligands (8–11 mers), although a core of nine residues fits into the groove. Peptides can interact through different nonameric sequences (registers), imposing nontrivial complexity to the determination of binding cores and to the development of algorithms predicting HTL epitopes.

Immunodominance

When individuals are immunized with a complex antigen, only a small fraction of the peptides – the immunodominant epitopes – are able to elicit T-cell responses. The immunodominance hierarchies result from a complex combination of factors, including T-cell repertoire, antigen processing and presentation. The unpredictability in the immunodominance hierarchy leads to significant impediments in the ability to rationally design efficient vaccines (Weaver *et al.*, 2008).

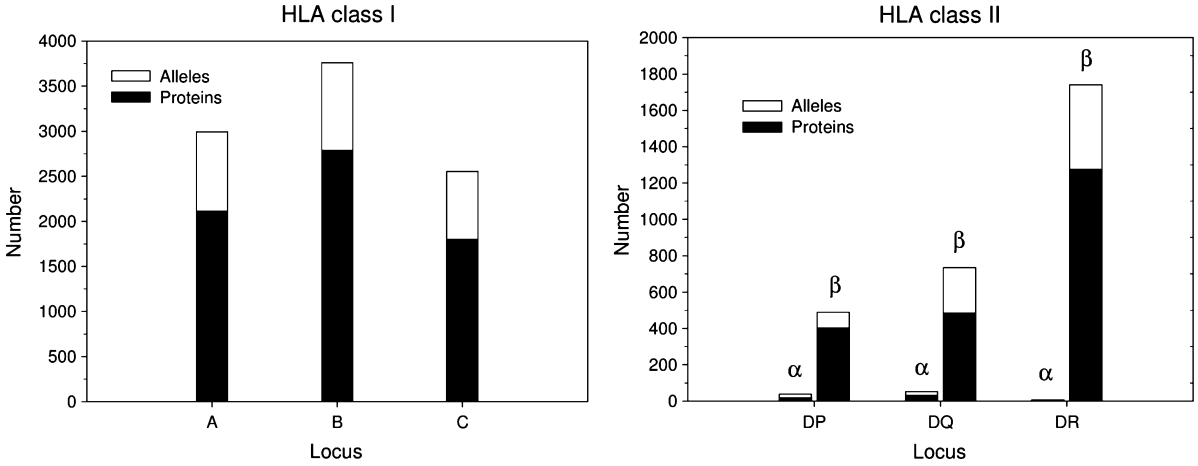

Figure 1. Distribution of the gene polymorphism for the HLA classical loci, both for HLA class I (a, b and c) and HLA class II (DP, DQ and DR). Stacked bars indicate the number of HLA allelic variants and the number of distinct HLA proteins (in black) according to the IMGT/HLA database release version 3.19.0 (Robinson *et al.*, 2015).

The main processing pathway for HLA class I ligands involves cytosolic degradation of proteins by the proteasome, followed by transport of the resulting peptides to the endoplasmic reticulum (ER) by the transporter associated with antigen processing (TAP), where they can be loaded onto nascent HLA class I molecules. The proteasomes are multisubunit ATP-dependent proteases, with a central role in establishing the immunodominance hierarchy of CTLs (Chen *et al.*, 2001). Proteasomes cleave preferentially at the C-terminal end of the epitopes, but in the ER, the peptides undergo N-terminal trimming. A number of methods have been developed for prediction of antigen-processing steps preceding HLA class I binding, including proteasomal cleavage patterns (Saxova *et al.*, 2003) and TAP transport efficiency (Brusic *et al.*, 1999; Diez-Rivero *et al.*, 2010). The NetChop tool predicts proteasomal cleavage sites based on artificial neural networks (ANNs) (Nielsen *et al.*, 2005). Likewise, Peters *et al.* (2003) developed a matrix-based method to predict TAP transport efficiency. The method was subsequently integrated with proteasomal cleavage prediction and MHC class I peptide-binding affinity prediction in the web-based computational tool NetCTL (Larsen *et al.*, 2005). NetCTLpan is a 'pan-extension' of NetCTL, allowing its broad coverage of MHC class I alleles (Stranzl *et al.*, 2010). The NetChop tool predicts proteasomal cleavage sites based on artificial neural networks (ANNs) (Nielsen *et al.*, 2005).

The MHC class II processing pathway involves antigen protein degradation by the lysosomal–endosomal apparatus, binding of the resulting peptides to MHCs, and subsequent transport of the complex to the cell surface. It has been suggested that antigen properties related to its processing are the major determinants of immunodominance, including the antigen tertiary structure (Carmicle *et al.*, 2007) and protease susceptibility of the sequences flanking the peptide (Godkin *et al.*, 2001). An alternative point of view, the peptide-intrinsic model, suggests that spontaneous kinetic stability of the pMHC class II complex and peptide loading onto MHCs dictate immunodominance hierarchies and the resulting peptide repertoire (Sant *et al.*, 2007).

Epitope-based vaccine design

Focusing epitope selection on most prevalent and well-characterized HLA molecules allows broad worldwide coverage in theory (Gulukota & DeLisi, 1996). However, these proteins are not necessarily prevalent among minor ethnic populations that are in need of new vaccination initiatives (Mehra & Kaur, 2003). Thus, the rational consideration of these phenomena in the EV design process offers the prospect to improve population coverage in well-defined ethnic populations (Shu *et al.*, 2014). EV design pipeline involves three computational steps (Toussaint & Kohlbacher, 2009b): (i) epitope prediction: given a set of antigens, candidate T-cell epitopes are identified with respect to a set of target HLA alleles; (ii) epitope selection: the most suitable subset of epitopes is selected out of the set of candidate epitopes; and (iii) vaccine assembly: a polypeptide vaccine construct is assembled from the selected epitopes (Fig. 2).

Epitope prediction

High-throughput T-cell assays are costly and time-consuming. Thus, *in silico* methods for epitope discovery have become the core subject of immunoinformatics, focusing on the peptide: MHC binding to predict T-cell reactivity. These methods can be divided into two groups: (i) data-driven methods, based on sequence information; and (ii) structure-based methods, making use of protein 3D structural information.

Data-driven computational methods have evolved thanks to the huge amount of experimental peptide binding data held in public repositories, such as the Immune Epitope Database (IEDB) (Vita *et al.*, 2014). Earlier computational methods are allele-specific, not capable of addressing HLA polymorphism. By contrast, a few so-called pan-specific methods are capable of extrapolating predictions to experimentally uncharacterized MHC molecules, making them useful for EV design (Zhang *et al.*, 2012b). A milestone in this category is the 'pocket profile' method (TEPITOPE program) (Sturniolo *et al.*, 1999), based on the assumption of an additive effect of specificities of individual pockets. Pocket profiles are matrices based on experimental testing of the effects of different amino acid side chains on specific pockets. The method has been made web-accessible (Propred) (Singh & Raghava, 2001). More recently, TEPITOPEpan (Zhang *et al.*, 2012a) introduced a pan extension in terms of HLA class II allele coverage by extrapolating the pocket profiles to other allotypes sharing similar pockets. Another variant is the 'virtual pocket' method, which characterizes pockets in terms of amino acid environments through energy calculations (Zhao *et al.*, 2003). MultiRTA is based on thermodynamic principles (Bordner & Mittelmann, 2010). Predivac is based on the specificity-determining residue (SDR) approach and covers 95% of HLA class II allotypes (Oyarzun *et al.*, 2013). The ANN-driven NetMHCPan-3.0 (Lundegaard *et al.*, 2008) and NetMHCIIPan-3.0 (Karosiene *et al.*, 2013) are state-of-the-art methods for MHC class I and class II peptide-binding prediction, respectively, which reach full HLA allelic coverage (Nielsen *et al.*, 2007).

Structure-based methods infer the physicochemical compatibility between the MHC molecules and putative peptide binders. These methods potentially enable predictions for any MHC allotype based on universal physical principles. However, they depend on experimentally determined pMHC structures or high-quality homology models. Atomistic molecular dynamics simulations are CPU-intensive and too time-consuming to be applicable for high-throughput screening of T-cell

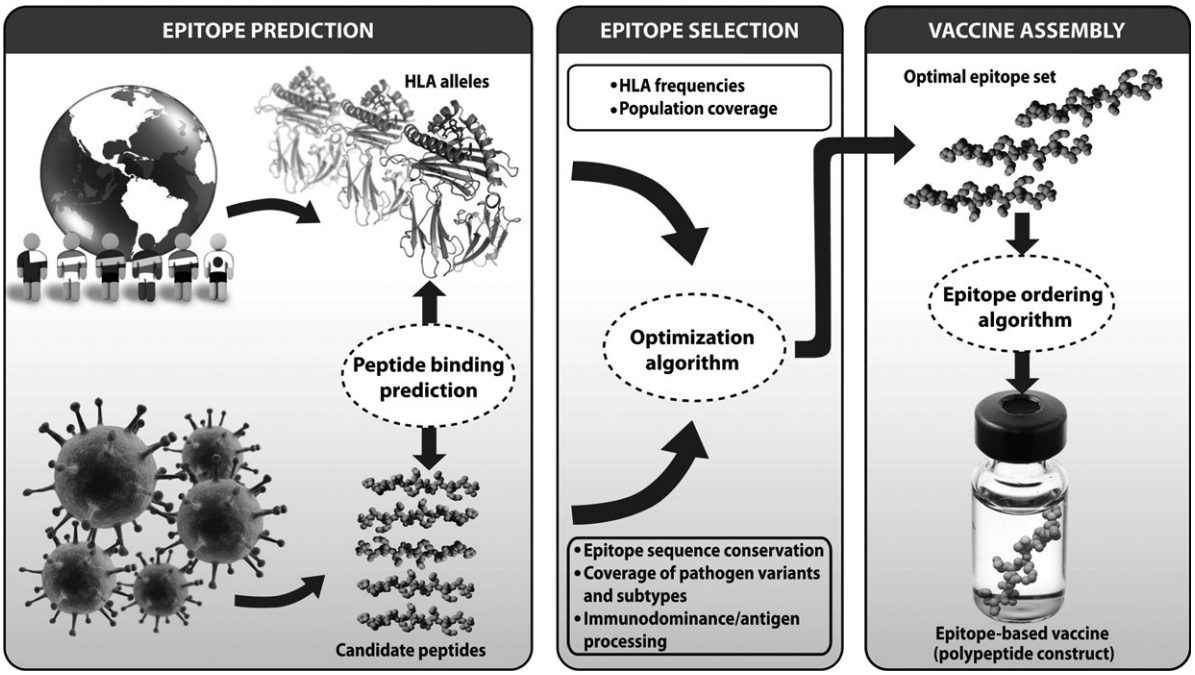

Figure 2. Diagram of the T-cell epitope-based vaccine design process. Starting from the target population and antigens, a peptide: HLA binding prediction method retrieves a set of candidate epitopes (epitope prediction stage), which together with additional information, such as HLA frequencies, population coverage, immunodominance/antigen processing, epitope conservation and coverage of the pathogen variants and subtypes, is all fed into an algorithm that selects an optimal set of epitopes (epitope selection stage). Finally, an epitope ordering algorithm assembles the selected peptides into a polypeptide construct (vaccine assembly stage).

epitopes (Flower *et al.*, 2010). Molecular docking programs balance computational time with accuracy by employing scoring functions to estimate binding affinities. The EpiDock method (Logean & Rognan, 2002) was developed for high-throughput prediction of CTL epitopes from viral genomes using the empirical scoring function Fresno (Rognan *et al.*, 1999). Khan & Ranganathan (2010) developed pDOCK that can be applied to any MHC allotype through homology modelling. Another docking-based protocol extrapolates the predictions to different MHC allotypes by incorporating a machine learning classifier trained on the docking solutions (Bordner, 2010). However, a recent analysis showed that the structure-based methods, while better than random, are inferior to state-of-the-art data-driven approaches (Zhang *et al.*, 2010).

Epitope selection

The identification of HLA allele-specific T-cell epitopes is insufficient for vaccine design, because different alleles are expressed at dramatically different frequencies in different populations (Fig. 3); people from different ethnic backgrounds are likely to react with a different set of peptides from a given pathogen. A minimal set of promiscuous epitopes that yields the strongest and broadest immune response must be selected (Ribeiro *et al.*, 2010).

Current computational methods to guide the EV design process and to estimate the fraction of individuals potentially protected by putative T-cell epitopes are listed in Table 1. Two classes can be distinguished: supertype-based and allele-based epitope selection methods. Supertypes are clusters of HLA molecules sharing overlapping peptide repertories. The targeting of supertypes for promiscuous epitopes provides a way to ensure broad population coverage (Sette & Sidney, 1998). For EV development, only representative alleles for each supertype have to be considered during the selection process, drastically decreasing the complexity of the problem. HLA class I supertypes described by Sette & Sidney (1999) corresponded to nine groups covering most of the HLA-A and HLA-B polymorphisms. Classifications are generally consistent for HLA class I, but lack agreement for HLA class II molecules. Several drawbacks have been reported for the supertype approach, including skewing of epitope selection to major alleles and poor performance of supertype-based selection strategies in population contexts with diverse HLA backgrounds (Schubert *et al.*, 2013).

Available bioinformatics tools for EV design based on the supertype concept include Pepvac (Reche & Reinherz, 2005) and Multipred2 (Zhang *et al.*, 2011). Both methods retrieve population coverage exclusively for HLA class I alleles, which have been precomputed according to the HLA allele and haplotype frequencies for five major ethnic groups in the United States population (Black, Caucasian, Hispanic, North American Natives and Asian) (Cao *et al.*, 2001). The

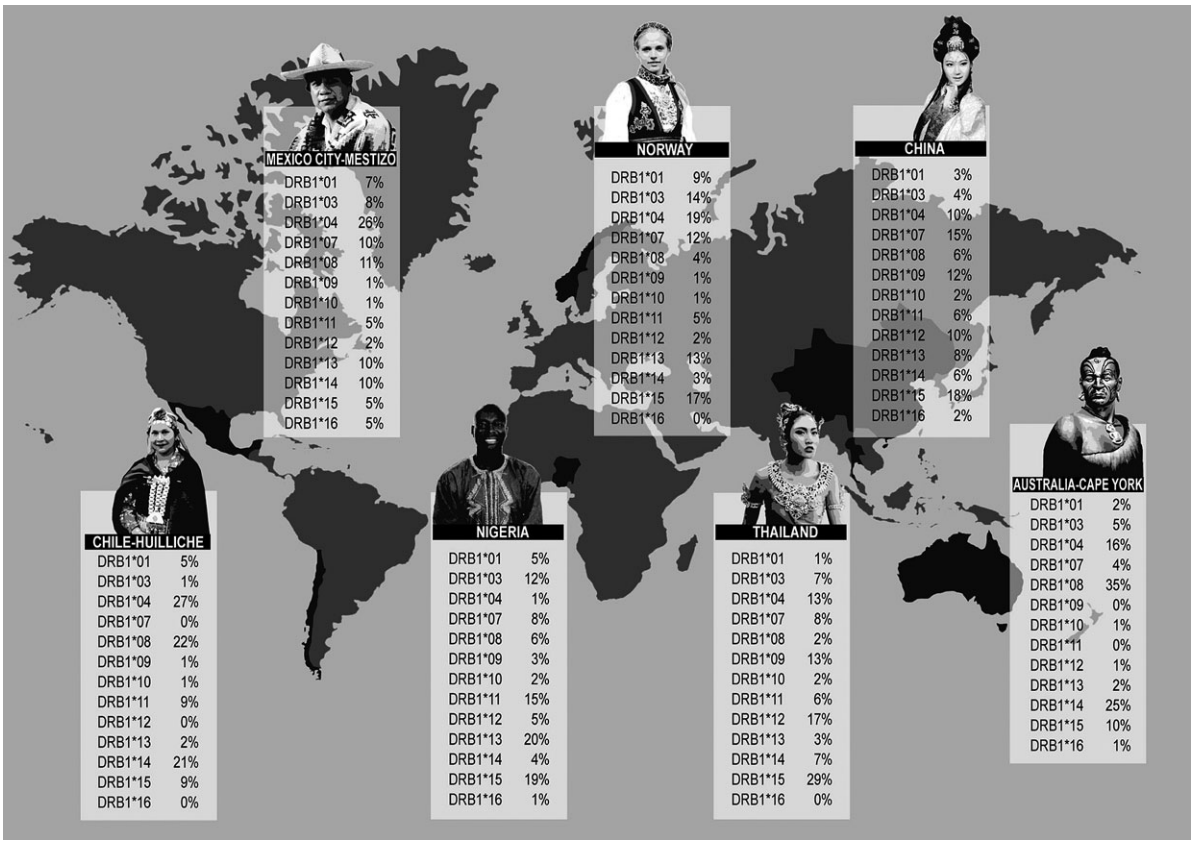

Figure 3. Differences in the frequency of various HLA-DR alleles in seven selected populations worldwide The population samples from Chile (Huilliche ethnicity) and Australia (Cape York ethnicity) are examples of HLA allele frequencies of indigenous populations, while the data from Mexico, Norway, China, Thailand and Nigeria represent the frequency of mixed populations [allele frequency data from the AFND (Gonzalez-Galarza *et al.*, 2011)].

population coverage depends on the combination of supertypes predicted to be targeted. Pepvac implements the profile-based method Rankpep (Reche *et al.*, 2002) for CTL epitope prediction, while Multipred2 employs NetMHCpan-2.8 (Nielsen *et al.*, 2007) and NetMHCIIpan-3.0 (Karosiene *et al.*, 2013) as prediction engines for CTL and HTL epitopes, respectively. Multipred2 retrieves a list of putative promiscuous T-cell epitopes, which correspond to those peptides predicted to bind >50% of the alleles in a given supertype.

Allele-based selection methods include OptiTope (Toussaint & Kohlbacher, 2009a), Episopt (Molero-Abraham *et al.*, 2013) and Predivac-2.0 (Oyarzun *et al.*, 2015). These tools define promiscuous epitopes as those restricted to many HLA alleles in the target population, regardless of supertype classification. The fraction of individuals potentially covered by the epitopes is determined as a function of specific allele frequency distributions. OptiTope implements a pipeline for HLA class I-restricted EV design, based on methods for CTL epitope prediction and an optimization algorithm based on integer linear programing to maximize population coverage (Toussaint *et al.*, 2008). Episopt employs Rankpep as CTL epitope prediction method and population coverage prediction is delivered for five major ethnic groups in the US population, based on HLA allele frequencies described by Cao *et al.* (2001). Predivac-2.0 enables HLA class II-restricted EV design, based on a previously described method for HTL epitope prediction (Oyarzun *et al.*, 2013) and implementing a genetic algorithm to explore epitope combinations maximizing population coverage. Predivac-2.0 accounts comprehensively for human genetic diversity through integration with the Allele Frequency Net Database (AFND) (Gonzalez-Galarza *et al.*, 2011), allowing 'ethnicity-oriented' applications.

A second source of diversity comes from the pathogen genomic variation (Sirskyj *et al.*, 2011). A few algorithms have been reported to explore T-cell epitope combinations that simultaneously optimize HLA allele and antigen coverage. Vider-Shalit *et al.* (2007) reported an evolutionary algorithm to find optimal combinations of CTL epitopes covering all HLA alleles and all viral proteins. PopCover (Buggert *et al.*, 2012) is an iterative procedure that selects a given number of putative epitopes in a way that maximizes the

Table 1. Computational tools and algorithms available for EV design

Method	Web implementation	Epitopes	Epitope prediction	Epitope selection	Epitope optimization	Population coverage
Tools						
Predivac-2.0	http://predivac.biosci.uq.edu.au/	HTL	Predivac	Yes	Yes (genetic algorithm)	Yes (worldwide/AFND)
OptiTope	http://etk.informatik.uni-tuebingen.de/optitope	CTL	Bimas, SYFPEITHI SVMHC	Yes	Yes (integer linear programing)	Yes (worldwide/dbMHC)
Episopt	http://bio.med.ucm.es/episopt.html	CTL	Rankpep	Yes	No	Yes (5 major ethnicities in the USA)
Pepvac	http://imed.med.ucm.es/PEPVAC/	CTL	Supertypes/Rankpep	Yes	No	Yes (5 major ethnicities in the USA)
Multipred2	http://cvc.dfci.harvard.edu/multipred2/	HTL	Supertypes/NetMHCIIpan	Yes	No	No
		CTL	Supertypes/NetMHCIpan	Yes	No	Yes (5 major ethnicities in the USA)
Algorithms						
Vider-Shalit *et al.* (2007)	No	CTL	Any/Bimass[a]	Yes	Yes (genetic algorithm)	Yes (set of target alleles)
Toussaint *et al.* (2011)	No	CTL	Any/Bimass[a]	Yes	Yes (integer linear programing)	Yes (set of target alleles)
PopCover	No	HTL	Any/NetMHC[a]/NetMHCpan[a]	Yes	Yes	Yes (set of target alleles)
		CTL	Any/NetCTL[a]	Yes	Yes	Yes (set of target alleles)

CTL: cytotoxic T-cell epitope, HTL: helper T-cell epitope.
[a]Methods reported in the original publications.

coverage of the viral strain with the smallest number of T-cell epitopes. Toussaint *et al.* (2011) reported a general mathematical formalism for CTL epitope selection to design universal EVs for highly variable viruses. These algorithms are not implemented as computational tools. PopCover focuses on HLA binding and epitope conservation, while the algorithms reported by Vider-Shalit *et al.* and Toussaint *et al.* incorporate additional variables such as the probability of predicted CTL epitopes to result from proteasomal cleavage of the source antigen. For HTL epitopes, the prediction of protease cleavage sites is less developed and has not been incorporated into algorithms for EV design. A different approach to cover pathogen diversity is the Mosaic method (Fischer *et al.*, 2007). It is a genetic algorithm-based method to generate artificial composite protein sequences (mosaic proteins), which are optimized to include a maximal diversity of putative T-cell epitopes from a set of viral proteins. A recent benchmark showed that supertype-based selection had a poorer performance compared to allele-based selections, suggesting it should only be used in populations where supertype clusters are prevalent (Schubert *et al.*, 2013).

Vaccine assembly

The administration of peptide cocktails containing short synthetic epitopes has been reported (Nehete *et al.*, 2001). However, the rapid clearance of the peptides from the bloodstream due to enzymatic degradation and the consequent short-lasting activity is a serious limitation of this strategy. In addition, nonameric epitopes lack flanking residues that are potentially important for HLA peptide-binding, presentation and T-cell recognition. A preferable strategy is to concatenate T-cell epitopes into linear polypeptide constructs composed of contiguous epitope sequences, with or without spacer sequences.

The computational challenge for T-cell epitope vaccine assembly is to optimize the ordering of the epitopes in the final polypeptide to ensure favourable cleavage of the peptides and to minimize junctional 'neoepitopes' (Sette *et al.*, 2002). Poly-CTL-epitopes should provide proteasomal cleavage sites at their C-termini and their N-termini optimized for TAP binding. Junctional epitopes interfere with vaccine function by suppressing immunogenicity of authentic epitopes. A commonly reported peptidic spacer for genetic constructs is the GPGPG sequence (Livingston *et al.*, 2002). The computational algorithm CANVAC II has been developed to aid the design of junctional epitope-free polyepitopes (Lee *et al.*, 2010). Given two authentic epitopes, a target set of MHC-binding motifs and a range of desired linker lengths, CANVAC predicts spacer sequences with optimal length and composition. GAIA, a multiepitope peptide vaccine against HIV, was developed using the

proprietary tools EpiAssembler® and VaccineCAD®. The former allowed the design of extensions of nonameric HTL epitopes to contain multiple overlapping conserved HLA class II binding regions, and the latter to align the epitopes into a 'beads-on-a-string construct' that minimizes junctional immunogenicity (De Groot *et al.*, 2005). PolyCTLDesigner has been developed for rational design of polyepitope T-cell antigens with a special focus on optimizing both amino acid spacer sequences and the ordering of the epitopes within the polyepitope (Antonets & Bazhan, 2013). Given a set of CTL epitopes, the tool predicts binding affinity to TAP and adds up to three N-terminal flanking amino acid residues, following Peters *et al.* (2003). Subsequently, it selects optimal spacer sequences for each peptide pair in terms of proteasomal cleavage sites at the C-terminus of the first peptide according to Toes *et al.* (2001), while providing the least number of nontarget junctional epitopes.

Conclusions

HLA polymorphism and the consequent population coverage are major issues for EV design in the context of genetically heterogeneous human populations. Peptide vaccination can take into consideration ethnic-level variations in immune responsiveness. Several algorithms and computational tools have been developed to aid the discovery and selection of T-cell epitopes with the potential to induce broad immune responses in target populations. A myriad of computational methods focuses on T-cell epitope prediction, but only partially accounts for immunodominance and antigen processing. In addition, epitope selection and methods for vaccine assembly are gaining interest in the scientific community and gradually becoming the focus of future EV development efforts.

Acknowledgements

BK is a National Health and Medical Research Council (NHMRC, Australia) Research Fellow (1003325).

References

Antonets, D.V. & Bazhan, S.I. (2013) Polyctldesigner: a computational tool for constructing polyepitope T-cell antigens. *BMC Research Notes*, **6**, 407.

Bordner, A.J. (2010) Towards universal structure-based prediction of class II MHC epitopes for diverse allotypes. *PLoS One*, **5**, e14383.

Bordner, A.J. & Mittelmann, H.D. (2010) Multirta: a simple yet reliable method for predicting peptide binding affinities for multiple class II MHC allotypes. *BMC Bioinformatics*, **11**, 482.

Brusic, V., van Endert, P., Zeleznikow, J., Daniel, S., Hammer, J. & Petrovsky, N. (1999) A neural network model approach to the study of human tap transporter. *In Silico Biology*, **1**, 109.

Buggert, M., Norstrom, M.M., Czarnecki, C., Tupin, E., Luo, M., Gyllensten, K. *et al.* (2012) Characterization of HIV-specific CD4+ T cell responses against peptides selected with broad population and pathogen coverage. *PLoS One*, **7**, e39874.

Cao, K., Hollenbach, J., Shi, X., Shi, W., Chopek, M. & Fernandez-Vina, M.A. (2001) Analysis of the frequencies of HLA-A, B, and C alleles and haplotypes in the five major ethnic groups of the United States reveals high levels of diversity in these loci and contrasting distribution patterns in these populations. *Human Immunology*, **62**, 1009.

Carmicle, S., Steede, N.K. & Landry, S.J. (2007) Antigen three-dimensional structure guides the processing and presentation of helper T-cell epitopes. *Molecular Immunology*, **44**, 1159.

Chaitra, M.G., Shaila, M.S. & Nayak, R. (2007) Evaluation of T-cell responses to peptides with MHC class I-binding motifs derived from pe_pgrs 33 protein of *Mycobacterium tuberculosis*. *Journal of Medical Microbiology*, **56**, 466.

Chen, W., Norbury, C.C., Cho, Y., Yewdell, J.W. & Bennink, J.R. (2001) Immunoproteasomes shape immunodominance hierarchies of antiviral CD8(+) T cells at the levels of T cell repertoire and presentation of viral antigens. *The Journal of Experimental Medicine*, **193**, 1319.

De Groot, A.S., Marcon, L., Bishop, E.A., Rivera, D., Kutzler, M., Weiner, D.B. & Martin, W. (2005) HIV vaccine development by computer assisted design: the GAIA vaccine. *Vaccine*, **23**, 2136.

Diez-Rivero, C.M., Chenlo, B., Zuluaga, P. & Reche, P.A. (2010) Quantitative modeling of peptide binding to TAP using support vector machine. *Proteins*, **78**, 63.

Fischer, W., Perkins, S., Theiler, J., Bhattacharya, T., Yusim, K., Funkhouser, R. *et al.* (2007) Polyvalent vaccines for optimal coverage of potential T-cell epitopes in global HIV-1 variants. *Nature Medicine*, **13**, 100.

Flower, D.R., Phadwal, K., Macdonald, I.K., Coveney, P.V., Davies, M.N. & Wan, S. (2010) T-cell epitope prediction and immune complex simulation using molecular dynamics: state of the art and persisting challenges. *Immunome Research*, **6**(Suppl 2), S4.

Godkin, A.J., Smith, K.J., Willis, A., Tejada-Simon, M.V., Zhang, J., Elliott, T. & Hill, A.V. (2001) Naturally processed HLA class II peptides reveal highly conserved immunogenic flanking region sequence preferences that reflect antigen processing rather than peptide-MHC interactions. *Journal of Immunology*, **166**, 6720.

Gonzalez-Galarza, F.F., Christmas, S., Middleton, D. & Jones, A.R. (2011) Allele frequency net: a database and online repository for immune gene frequencies in worldwide populations. *Nucleic Acids Research*, **39**, D913.

Goulder, P.J. & Watkins, D.I. (2008) Impact of MHC class I diversity on immune control of immunodeficiency virus replication. *Nature Reviews Immunology*, **8**, 619.

Gulukota, K. & DeLisi, C. (1996) HLA allele selection for designing peptide vaccines. *Genetic Analysis*, **13**, 81.

Guo, H., Santiago, F., Lambert, K., Takimoto, T. & Topham, D.J. (2011) T cell-mediated protection against lethal 2009 pandemic H1N1 influenza virus infection in a mouse model. *Journal of Virology*, **85**, 448.

Karosiene, E., Rasmussen, M., Blicher, T., Lund, O., Buus, S. & Nielsen, M. (2013) 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*, **65**, 711.

Khan, J.M. & Ranganathan, S. (2010) pDOCK: a new technique for rapid and accurate docking of peptide ligands to Major Histocompatibility Complexes. *Immunome Research*, **6**(Suppl 1), S2.

Larsen, M.V., Lundegaard, C., Lamberth, K., Buus, S., Brunak, S., Lund, O. & Nielsen, M. (2005) An integrative approach to CTL epitope prediction: a combined algorithm integrating MHC class I binding, TAP transport efficiency, and

proteasomal cleavage predictions. *European Journal of Immunology*, **35**, 2295.
Lee, Y., Ferrari, G. & Lee, S.C. (2010) Estimating design space available for polyepitopes through consideration of major histocompatibility complex binding motifs. *Biomedical Microdevices*, **12**, 207.
Li, W., Joshi, M.D., Singhania, S., Ramsey, K.H. & Murthy, A.K. (2014) Peptide vaccine: progress and challenges. *Vaccines*, **2**, 515.
Livingston, B., Crimi, C., Newman, M., Higashimoto, Y., Appella, E., Sidney, J. & Sette, A. (2002) A rational strategy to design multiepitope immunogens based on multiple Th lymphocyte epitopes. *Journal of Immunology*, **168**, 5499.
Logean, A. & Rognan, D. (2002) Recovery of known T-cell epitopes by computational scanning of a viral genome. *Journal of Computer-Aided Molecular Design*, **16**, 229.
Lundegaard, C., Lamberth, K., Harndahl, M., Buus, S., Lund, O. & Nielsen, M. (2008) NetMHC-3.0: accurate web accessible predictions of human, mouse and monkey MHC class I affinities for peptides of length 8-11. *Nucleic Acids Research*, **36**, W509.
Matsui, M., Moots, R.J., McMichael, A.J. & Frelinger, J.A. (1994) Significance of the six peptide-binding pockets of HLA-A2.1 in influenza a matrix peptide-specific cytotoxic T-lymphocyte reactivity. *Human Immunology*, **41**, 160.
Mehra, N.K. & Kaur, G. (2003) MHC-based vaccination approaches: progress and perspectives. *Expert Reviews in Molecular Medicine*, **5**, 1.
Molero-Abraham, M., Lafuente, E.M., Flower, D.R. & Reche, P.A. (2013) Selection of conserved epitopes from hepatitis C virus for pan-populational stimulation of T-cell responses. *Clinical & Developmental Immunology*, **2013**, 601943.
Nehete, P.N., Chitta, S., Hossain, M.M., Hill, L., Bernacky, B.J., Baze, W., Arlinghaus, R.B. & Sastry, K.J. (2001) Protection against chronic infection and AIDS by an HIV envelope peptide-cocktail vaccine in a pathogenic SHIV-rhesus model. *Vaccine*, **20**, 813.
Nielsen, M., Lundegaard, C., Lund, O. & Kesmir, C. (2005) The role of the proteasome in generating cytotoxic T-cell epitopes: insights obtained from improved predictions of proteasomal cleavage. *Immunogenetics*, **57**, 33.
Nielsen, M., Lundegaard, C., Blicher, T., Lamberth, K., Harndahl, M., Justesen, S. *et al.* (2007) NetMHCpan, a method for quantitative predictions of peptide binding to any HLA-A and -B locus protein of known sequence. *PLoS One*, **2**, e796.
Oyarzun, P., Ellis, J.J., Boden, M. & Kobe, B. (2013) Predivac: CD4+ T-cell epitope prediction for vaccine design that covers 95% of hla class II DR protein diversity. *BMC Bioinformatics*, **14**, 52.
Oyarzun, P., Ellis, J.J., Gonzalez-Galarza, F.F., Jones, A.R., Middleton, D., Boden, M. & Kobe, B. (2015) A bioinformatics tool for epitope-based vaccine design that accounts for human ethnic diversity: application to emerging infectious diseases. *Vaccine*, **33**, 1267.
Peters, B., Bulik, S., Tampe, R., Van Endert, P.M. & Holzhutter, H.G. (2003) Identifying MHC class I epitopes by predicting the TAP transport efficiency of epitope precursors. *Journal of Immunology*, **171**, 1741.
Purcell, A.W., McCluskey, J. & Rossjohn, J. (2007) More than one reason to rethink the use of peptides in vaccine design. *Nature Reviews Drug Discovery*, **6**, 404.
Reche, P.A. & Reinherz, E.L. (2005) Pepvac: a web server for multi-epitope vaccine development based on the prediction of supertypic MHC ligands. *Nucleic Acids Research*, **33**, W138.
Reche, P.A., Glutting, J.P. & Reinherz, E.L. (2002) Prediction of MHC class I binding peptides using profile motifs. *Human Immunology*, **63**, 701.
Ribeiro, S.P., Rosa, D.S., Fonseca, S.G., Mairena, E.C., Postol, E., Oliveira, S.C., Guilherme, L., Kalil, J. & Cunha-Neto, E. (2010) 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*, **5**, e11072.
Robinson, J., Halliwell, J.A., Hayhurst, J.D., Flicek, P., Parham, P. & Marsh, S.G. (2015) The IPD and IMGT/HLA database: allele variant databases. *Nucleic Acids Research*, **43**, D423.
Rognan, D., Lauemoller, S.L., Holm, A., Buus, S. & Tschinke, V. (1999) Predicting binding affinities of protein ligands from three-dimensional models: application to peptide binding to class I major histocompatibility proteins. *Journal of Medicinal Chemistry*, **42**, 4650.
Rosa, D.S., Ribeiro, S.P. & Cunha-Neto, E. (2010) CD4+ T cell epitope discovery and rational vaccine design. *Archivum immunolgiae et therapiae experimentalis*, **58**, 121.
Sant, A.J., Chaves, F.A., Krafcik, F.R., Lazarski, C.A., Menges, P., Richards, K. & Weaver, J.M. (2007) Immunodominance in CD4 T-cell responses: implications for immune responses to influenza virus and for vaccine design. *Expert Review of Vaccines*, **6**, 357.
Saxova, P., Buus, S., Brunak, S. & Kesmir, C. (2003) Predicting proteasomal cleavage sites: a comparison of available methods. *International Immunology*, **15**, 781.
Schubert, B., Lund, O. & Nielsen, M. (2013) Evaluation of peptide selection approaches for epitope-based vaccine design. *Tissue Antigens*, **82**, 243.
Sette, A. & Sidney, J. (1998) HLA supertypes and supermotifs: a functional perspective on HLA polymorphism. *Current Opinion in Immunology*, **10**, 478.
Sette, A. & Sidney, J. (1999) Nine major HLA class I supertypes account for the vast preponderance of HLA-A and -B polymorphism. *Immunogenetics*, **50**, 201.
Sette, A., Newman, M., Livingston, B., McKinney, D., Sidney, J., Ishioka, G., Tangri, S., Alexander, J., Fikes, J. & Chesnut, R. (2002) Optimizing vaccine design for cellular processing, MHC binding and TCR recognition. *Tissue Antigens*, **59**, 443.
Shu, J., Fan, X., Ping, J., Jin, X. & Hao, P. (2014) Designing peptide-based HIV vaccine for Chinese. *Biomed Research International*, **2014**, 272950, doi:10.1155/2014/272950.
Singh, H. & Raghava, G.P. (2001) Propred: prediction of HLA-DR binding sites. *Bioinformatics*, **17**, 1236.
Sirskyj, D., Diaz-Mitoma, F., Golshani, A., Kumar, A. & Azizi, A. (2011) Innovative bioinformatic approaches for developing peptide-based vaccines against hypervariable viruses. *Immunology and Cell Biology*, **89**, 81.
Stranzl, T., Larsen, M.V., Lundegaard, C. & Nielsen, M. (2010) NetCTLpan: pan-specific MHC class I pathway epitope predictions. *Immunogenetics*, **62**, 357.
Sturniolo, T., Bono, E., Ding, J., Raddrizzani, L., Tuereci, O., Sahin, U. *et al.* (1999) Generation of tissue-specific and promiscuous HLA ligand databases using DNA microarrays and virtual HLA class II matrices. *Nature Biotechnology*, **17**, 555.
Toes, R.E., Nussbaum, A.K., Degermann, S., Schirle, M., Emmerich, N.P., Kraft, M. *et al.* (2001) Discrete cleavage motifs of constitutive and immunoproteasomes revealed by quantitative analysis of cleavage products. *The Journal of Experimental Medicine*, **194**, 1.
Toussaint, N.C. & Kohlbacher, O. (2009a) Optitope–a web server for the selection of an optimal set of peptides for epitope-based vaccines. *Nucleic Acids Research*, **37**, W617.
Toussaint, N.C. & Kohlbacher, O. (2009b) Towards *in silico* design of epitope-based vaccines. *Expert Opinion on Drug Discovery*, **4**, 1047.
Toussaint, N.C., Donnes, P. & Kohlbacher, O. (2008) A mathematical framework for the selection of an optimal set of

peptides for epitope-based vaccines. *PLoS Computational Biology*, **4**, e1000246.
Toussaint, N.C., Maman, Y., Kohlbacher, O. & Louzoun, Y. (2011) Universal peptide vaccines – optimal peptide vaccine design based on viral sequence conservation. *Vaccine*, **29**, 8745.
Vider-Shalit, T., Raffaeli, S. & Louzoun, Y. (2007) Virus-epitope vaccine design: informatic matching the HLA-I polymorphism to the virus genome. *Molecular Immunology*, **44**, 1253.
Vita, R., Overton, J.A., Greenbaum, J.A., Ponomarenko, J., Clark, J.D., Cantrell, J.R. *et al.* (2015) The immune epitope database (IEDB) 3.0. *Nucleic Acids Research*, **43**, D405.
Weaver, J.M., Lazarski, C.A., Richards, K.A., Chaves, F.A., Jenks, S.A., Menges, P.R. & Sant, A.J. (2008) Immunodominance of CD4 T cells to foreign antigens is peptide intrinsic and independent of molecular context: implications for vaccine design. *Journal of Immunology*, **181**, 3039.
Zhang, H., Wang, P., Papangelopoulos, N., Xu, Y., Sette, A., Bourne, P.E., Lund, O., Ponomarenko, J., Nielsen, M. & Peters, B. (2010) Limitations of *ab initio* predictions of peptide binding to MHC class II molecules. *PLoS One*, **5**, e9272.
Zhang, G.L., DeLuca, D.S., Keskin, D.B., Chitkushev, L., Zlateva, T., Lund, O., Reinherz, E.L. & Brusic, V. (2011) Multipred2: a computational system for large-scale identification of peptides predicted to bind to HLA supertypes and alleles. *Journal of Immunological Methods*, **374**, 53.
Zhang, L., Chen, Y., Wong, H.S., Zhou, S., Mamitsuka, H. & Zhu, S. (2012a) TEPITOPEpan: extending TEPITOPE for peptide binding prediction covering over 700 HLA-DR molecules. *PLoS One*, **7**, e30483.
Zhang, L., Udaka, K., Mamitsuka, H. & Zhu, S. (2012b) Toward more accurate pan-specific MHC-peptide binding prediction: a review of current methods and tools. *Briefings in Bioinformatics*, **13**, 350.
Zhao, B., Mathura, V.S., Rajaseger, G., Moochhala, S., Sakharkar, M.K. & Kangueane, P. (2003) A novel MHCp binding prediction model. *Human Immunology*, **64**, 1123.