Journal of Biomedical Informatics 53 (2015) 405–414

Contents lists available at ScienceDirect

Journal of Biomedical Informatics

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

Methodological Review

An overview of bioinformatics tools for epitope prediction: Implications on vaccine development

Ruth E. Soria-Guerra [a], Ricardo Nieto-Gomez [b], Dania O. Govea-Alonso [b], Sergio Rosales-Mendoza [b,*]

[a] Laboratorio de Ingeniería de Biorreactores, Facultad de Ciencias Químicas, Universidad Autónoma de San Luis Potosí, Av. Dr. Manuel Nava 6, SLP 78210, Mexico
[b] Laboratorio de Biofarmacéuticos Recombinantes, Facultad de Ciencias Químicas, Universidad Autónoma de San Luis Potosí, Av. Dr. Manuel Nava 6, SLP 78210, Mexico

ARTICLE INFO

Article history:
Received 25 February 2014
Accepted 3 November 2014
Available online 10 November 2014

Keywords:
Antigenic determinants
Databases
Hypervariable pathogens
Immunological software
Immunoprotection

ABSTRACT

Exploitation of recombinant DNA and sequencing technologies has led to a new concept in vaccination in which isolated epitopes, capable of stimulating a specific immune response, have been identified and used to achieve advanced vaccine formulations; replacing those constituted by whole pathogen-formulations. In this context, bioinformatics approaches play a critical role on analyzing multiple genomes to select the protective epitopes *in silico*. It is conceived that cocktails of defined epitopes or chimeric protein arrangements, including the target epitopes, may provide a rationale design capable to elicit convenient humoral or cellular immune responses. This review presents a comprehensive compilation of the most advantageous online immunological software and searchable, in order to facilitate the design and development of vaccines. An outlook on how these tools are supporting vaccine development is presented. HIV and influenza have been taken as examples of promising developments on vaccination against hypervariable viruses. Perspectives in this field are also envisioned.

1. Introduction

Exploitation of vaccination as a tool in fighting wide spread diseases has resulted in substantial strides in the combat against many infectious diseases such as influenza, smallpox, varicella, pertussis, diphtheria, tetanus, polio, hepatitis, and rotavirus [1,2]. The conventional vaccines, which include attenuated or killed agents, might take up to 15 years of development; this includes cultivation of the desired microorganism at a larger scale and under proper conditions as well as an effective inactivation with a subsequent evaluation of vaccine immunogenicity. Although this kind of vaccines has saved countless lives, it can have unfavorable consequences as adverse effects, induce the disease or, in some instances, even death [3–5].

Bioinformatics is a field of science in which several disciplines such as biology, computing, and information technology converge to organize and store large amounts of biological information driven by advances generated in genetics, molecular biology, and biotechnology [6]. One goal of bioinformatics is to streamline and interpret, effectively and timely, information from the genome, transcriptome, and/or proteome [7]. This discipline aims to promote health benefits including the area of vaccines.

The development of bioinformatics tools along with advances in recombinant DNA technology (rDNA) and the knowledge on the host immune response and the genetic background of the pathogen will lead to new vaccines against diseases that currently have few or no control measures in just 1 or 2 years through computer *in silico* predictions to define targets [8] see Fig. 1. The vaccines developed through rDNA technologies are designed to be safer, more efficacious, and/or less expensive than traditional vaccines. In order to achieve these aims, a thorough understanding of the disease agent, particularly, critical epitopes to induce the appropriate immunological reaction is required [9–11].

While the availability of the complete genome sequence permits the identification of all potential protein products, this information could be not sufficient to allow for the identification of the subset of proteins that are in fact expressed at any stage of the life

Abbreviations: HIV, Human Immunodeficiency Virus; rDNA, recombinant DNA technology; MHC, major histocompatibility complex; HLA, human leukocyte antigen; PSSM, Position Specific Scoring Matrices; ANN, Artificial Neuronal Networks; QM, quantitative matrices; KISS, Kernel-based Inter-allele peptide binding prediction SyStem; SVM, support vector machine; WAPP, Whole Antigen Processing Pathway; CTL, cytotoxic T cell; PI, Protrusion Index; gp120, envelop glycoprotein gp120; gag, structural polyprotein; mAb, monoclonal antibody; MCC, Matthews correlation coefficient; HA, hemagglutinin; NA, neuraminidase; TAP, transporter associated protein; APC, antigen presenting cells; SEPPA, Spatial Epitope Prediction of Protein Antigens; ATP, adenosine triphosphate; AIDS, acquired immunodeficiency syndrome; RV144, Thai HIV phase III prime/boost vaccine trial.

* Corresponding author. Fax: +52 444 826 2440.
E-mail address: rosales.s@fcq.uaslp.mx (S. Rosales-Mendoza).

http://dx.doi.org/10.1016/j.jbi.2014.11.003
1532-0464/

of the pathogen. Proteomics also play an important role in this field as it can serve as a complementary strategy to the genomic-based approaches using immunomics techniques to identify and characterize immunogenic proteins. Vaccinomics, which consists on the characterization of host response to immunization, provides valuable information on pathogen–host cell interaction to validate candidate antigens. The information obtained from these disciplines also speeds the identification and characterization of new antigens [12].

Proteomic experiments conducted on bacterial species have not only been verified with data obtained from genome sequencing and bioinformatic analyses but have also lead to the discovery of new proteins, which could be potential new vaccine candidates [12–14]. An extensive review focused on the impact of proteomics on the development of antibacterial and antiviral vaccines was published by Adamczyk-Poplawska [12].

It is important to note that the standard bioinformatics web tools presented here, are not enough for detailed analysis of whole genomes. The immunoinformatics is a discipline whose main objective is to convert large-scale immunological data, using computational and mathematical approaches, to understand and organize these large scale data to obtain immunologically meaningful interpretations [15,16]. The tools in this field are based on statistical and machine learning system and are used for studies in modeling molecular interactions (such as antigen processing and presentation) and also plays a role in defining new hypotheses related to understand the immune system mechanisms [17,18].

This review is intended to provide a gateway to some of the most useful online immunological softwares and searchable databases for genomes analyses, based in our own experience and a laborious search in the literature and web databases, providing an outlook on how these tools have aided on the vaccine development field particularly on the development of epitope-based vaccines.

2. Epitope-based vaccines

Epitopes are of particular interest to both clinical and basic biomedical researchers as they hold huge potential for vaccine design, disease prevention, diagnosis, and treatment. Using rDNA technologies, we can isolate specific epitopes which can replace the whole pathogen in a vaccine. However, within the diversity of epitopes in a pathogen, it is important to notice that not all of the epitopes, even those that seem to be dominant, are equal in their ability to elicit antibody production [19–21].

Besides producing particular immunogens instead whole pathogens, rDNA has allowed for a rational vaccine design comprising the production of chimeric proteins that opens a wide number of possibilities for immunogen design; including the conception of multiepitopic vaccines having advantages such as: several immunoprotective epitopes are included in a single molecule, immunodominant but non-protective epitopes are discarded, and epitopes exerting adjuvant effects such as promiscuous T cell epitopes can be included to enhance immunogenicity [22]. These features offer the possibility of designing multitarget, highly efficient vaccines. However a requisite for the design of such immunogens consists on the discovery of the immunoprotective epitopes and the variants when genetic variability is of relevance for a particular pathogen.

The epitope-driven vaccine is an attractive concept that is being successfully pursued in a large number of research groups, especially to the development of vaccines targeting conserved epitopes in variable or rapidly mutating pathogens [23,24].

The selected epitopes in a vaccine should ideally be conserved across different stages of the pathogen and its variants. Furthermore it should be taken into consideration the desired immune response. Cytotoxic T cell-mediated response is elicited by a pathway comprising intracellular antigen processing with linear epitopes as predominant targets [21]. In this regard, the epitopes selected for a vaccine must have binding affinity with more than one major histocompatibility complex (MHC) allele and must cover a major population [25,26].

The proteins that contain many epitopes recognized by the common MHC alleles are known as promiscuous binders [26]. The human leukocyte antigen (HLA) supertype refers to a set of HLA alleles with overlapping peptide binding specificities. The alleles in the given HLA super type often represent the same epitope, which refers to the region on the surface of an antigen capable of eliciting an immune response for T cell recognition [25,27].

On the other hand, elicitation of humoral responses relies on the recognition of linear epitopes and conformational epitopes. The latter constitute a challenge for chimeric vaccine design as they must retain their native conformation to be functional [28]. Therefore, knowledge on the whole antigen structure is necessary to aid on the rational design of vaccines targeting conformational B cell epitopes [27].

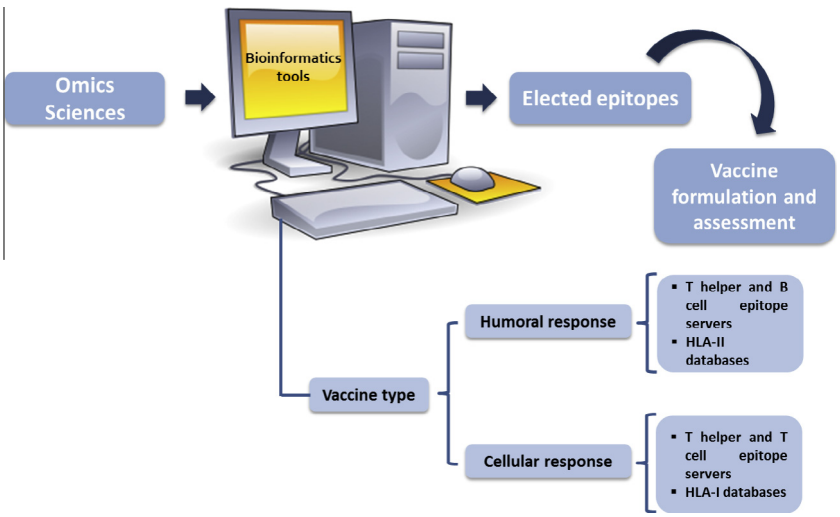

Fig. 1. Schematic representation of the workflow to identify epitopes for vaccine development.

In this context, bioinformatics approaches can contribute to the design of epitope-based vaccines. Using these tools, an appropriate *in silico* selection of epitopes can be accomplished [29,30].

3. Bioinformatics tools to predic potential T cell binding-epitopes

The first step on applying bioinformatics to vaccine development consists on discriminating epitopes that are potentially immunoprotective from epitopes that are not. Since T-cell epitopes are bound in a linear form to MHCs, the interface between ligands and T-cells can be modeled with accuracy [31]. It is currently well-known that epitopes link together into the binding groove of MHC Class I and Class II molecules through interactions between their R group side chains and pockets located on the floor of the MHC [32–34]. Based on this knowledge, a large number of T-cell epitope-mapping algorithms have been established and used to develop tools to rapidly identify putative T-cell epitopes [31,35,36].

MHC-I binding predictors are currently very efficient and have wide allelic coverage, a prediction accuracy in the range of 90–95% positive prediction value has been estimated [29,37,38]. Among the numerous servers for MHC-I alleles is included RANKPEP, which predicts peptide binders to MHC-I and MHC-II molecules from protein sequences or sequence alignments using Position Specific Scoring Matrices (PSSMs). In addition, it predicts those MHC-I ligands whose C-terminal end is likely to be the result of proteasomal cleavage [39]. This is a friendly platform who offers the most wide allelic coverage to MHC-I and MHC-II alleles (118 and 67 alleles, respectively) for human and mouse. To search epitopes sequences for MHC-I ligands using PSSMs, a dynamic algorithm written in Python is used; which scores all protein segments with the length of the PSSM width and sorts them accordingly. Scoring starts at the beginning of each sequence and the PSSM is slid over the sequence one residue at a time until reaching the end of the sequence. Furthermore, to narrow down the potential binders from the list of ranked peptides, a binding threshold is defined as the score value that includes 90% of the peptides within the PSSM. This binding threshold is built into each matrix, delineating the range of putative binders among the top scoring peptides [39].

The IEDB Analysis Resource database uses NetMHCpan as prediction method since 2011. This method generates a quantitative prediction of the affinity of any peptide-MHC class I interaction, covering HLA-A and HLA-B for humans as well as chimpanzee, macaque, gorilla, cow, pig and mouse. This constitutes one of the few databases that include this variety of organisms [37].

nHLAPred is another comprehensive tool for the prediction of MHCI binding peptides for 67 MHC alleles. The prediction of alleles is based on Artificial Neural Networks (ANNs) and quantitative matrices (QM). The predicted MHC binders are filtered to potential CTL epitopes by using proteasomal matrices. Although this server offers two options (Compred and ANNPred), the most wide-ranging is Compred; based on the hybrid approach of artificial neural networks and quantitative matrices [40].

NetMHC server predicts binding of peptides to a number of different HLA alleles using ANNs. ANNs have been trained for 78 different human MHC (HLA) alleles representing all 12 HLA-A and -B Supertypes. Furthermore predictions for 41 animals like monkey, cattle, pig, and mouse alleles are available [38].

Kernel-based Inter-allele peptide binding prediction SyStem (KISS) predicts whether 9-mer peptides will bind an MHC-I molecule for 64 alleles using a support vector machine (SVM) multitask Kernel to leverage the available training information across the alleles, which improves its accuracy especially for the alleles with few known epitopes. The predictor is trained on databases which contain known epitopes from SYFPEITHI, MHCBN, LANL, and IEDB databases.

Although there are other servers available to identify MHC-I binding predictors, the servers described above are the most complete in terms of allelic coverage and the identification of alleles in other organism besides human, however a more detailed list of MHC-I binding predictors available on-line is presented in Table 1.

TAP is a transporter associated with the MHC class I restricted antigen processing. TAP is heterodimeric transporter that belongs to the family of ABC transporters and uses the energy provided by ATP hydrolysis to translocate the peptides across the endoplasmic reticulum membrane. The transporter is composed of two proteins named TAP-1 and TAP-2. The subset of this transported peptide will bind MHC class I molecules. These MHC-peptide complexes are translocated on the surface of antigen presenting cells (APCs), with a subsequent potential of mounting T cell immune responses [41–43]. The TAP binding prediction softwares available include TAPPred, EpiJen, and WAPP (Table 2). TAPPred is an on-line tool to predict binding affinity of peptides toward the TAP transporter. The prediction of TAP binding peptides is crucial in identifying the MHC-I restricted T cell epitopes. The prediction is

Table 1
Comprehensive list of T cell epitope prediction servers.

Server name	Link	Predictive server for		Predictive method
		MHC I	MHC II	
EpiJen	http://www.ddg-pharmfac.net/epijen/EpiJen/EpiJen.htm	24		Multi-step algorithm
SYFPEITHI	http://www.syfpeithi.de/bin/MHCServer.dll/EpitopePrediction.htm	42	7	Published motifs
ANNPRED	http://www.imtech.res.in/raghava/nhlapred/neural.html	30		ANN-regression
BIMAS	http://www-bimas.cit.nih.gov/molbio/hla_bind/	41		Published coefficient tables
ProPred I	http://www.imtech.res.in/raghava/propred1/	47		Quantitative matrix
ProPred	http://www.imtech.res.in/raghava/propred/		51	Quantitative matrix
MHCPred	http://www.ddg-pharmfac.net/mhcpred/MHCPred/	14	11	Additive method
MHC2Pred	http://www.imtech.res.in/raghava/mhc2pred/		42	SVM-based method
NetMHC	http://www.cbs.dtu.dk/services/NetMHC/	57		ANN based method
PREDEP	http://margalit.huji.ac.il/Teppred/mhc-bind/index.html	13		Published coefficient tables
RANKPEP	http://bio.dfci.harvard.edu/RANKPEP/	118	62	PSSM
SVMHC	http://abi.inf.uni-tuebingen.de/Services/SVMHC	33	51	SVM-based method
IEDB binding	http://tools.immuneepitope.org/analyze/html/mhc_processing.html	77		ANN and SMM method
EpiVax	http://www.epivax.com/	6	8	Epimatrix algorithm
MMBPred	http://www.imtech.res.in/raghava/mmbpred/	46		Quantitative matrix
NetCTL	http://www.cbs.dtu.dk/services/NetCTL	12		ANN-regression
nHLAPred	http://www.imtech.res.in/raghava/nhlapred/	67		Artificial Neural Networks
KISS	http://cbio.ensmp.fr/kiss/	64		SVM based method
SVRMHC	http://svrmhc.biolead.org/	36	6	SVM-basedmethod
IMTECH	http://www.imtech.res.in/raghava/mhc		3	Quantitative matrix

Table 2
Predictive server for TAP binding epitopes and CTL.

Server name	Link	Description	Predictive method
EpiJen	http://www.ddg-pharmfac.net/epijen/EpiJen/EpiJen.htm	Predictive server for TAP binding epitopes	Multi-step algorithm
TAP Pred	http://www.imtech.res.in/raghava/tappred/	Predictive server for TAP binding epitopes	SVM method
WAPP	http://abi.inf.uni-tuebingen.de/Services/WAPP/information	Predictive server for TAP binding epitopes	SVM method
CTLPred	http://www.imtech.res.in/raghava/ctlpred/	Predictive server for CTL	SVM and ANN method
NetCTLPan	http://www.cbs.dtu.dk/services/NetCTLpan/	Predictive server for CTL	Multi-step algorithm

based on cascade SVM, using sequence and properties of the amino acids [44,45]. EpiJen server offers not only a TAP binding prediction but also a proteasome cut off. EpiJen uses an additive method which assumes that each substituent makes an additive and independent contribution to the biological activity. Their additive method considers the interactions between specific amino acids and the binding site [46].

The Whole Antigen Processing Pathway (WAPP) server includes tools for predicting proteosomal cleavage, TAP transport, and MHC-peptide binding. This server offers an integrated prediction for these three aspects. Prediction of proteasomal cleavage is based on experiments performed on the Enolase and Prion proteins. Sequences around experimental cleavage sites are used to construct a weight matrix, while a regression form of SVMs is used for prediction of TAP affinity [47]. Predictions of the MHC class I pathway can be improved by predictions of proteasomal cleavage, TAP transport efficiency, and MHC class I binding affinity [48,49].

Nevertheless while good performance has been achieved for MHC class I predictions, there is still limited success in predicting MHC-II-binding epitopes [50,51]. The low prediction accuracy of MHC-II binding epitopes is due to several factors including the insufficient or low-quality training data, difficulty on identifying 9-mer binding cores within longer peptides used for training and lack of consideration of the influence of flanking residues, and the relative permissiveness of the binding groove of MHC-II molecules which limits the binding stringency [29,50].

ProPred1 is a server that predicts MHC Class-II binding regions in an antigen sequence using quantitative matrices. The server will assist in locating promiscuous binding regions that are useful in selecting vaccine candidates covering 51 alleles [52]. The SVMHC server enables prediction of both MHC class I and MHC class II binding peptides, however the most widely coverage is for MHC-II (51 alleles). The graphical output displayed in this software also allows for simple identification of promiscuous epitopes. SVMHC uses the matrices developed by the TEPITOPE software [47].

The MHC2Pred is a SVM based method for prediction of promiscuous MHC class II binder. The average accuracy of SVM based method for 42 alleles is ~80%. The performance of the method was poorer for few alleles due to a smaller size of dataset. This server will be useful in cellular immunology, vaccine design, immunodiagnostics, immunotherapeutics, and molecular understanding of autoimmune susceptibility [52]. These last three servers along with RANKPEP described above, are the most complete servers with the broadest allelic coverage to predict epitopes binding to MHC-II and it also has a friendly interface. Other resources to predict MHC Class-II binding epitopes are described in Table 1.

There is another direct method for prediction of CTL epitopes, which are crucial in subunit vaccine design; examples of these servers are CTLPred and NetCTLPan. CTLPred uses the T cell epitope patterns instead of MHC binders. The method is based on techniques such as ANNs and SVM. The methods also allow the consensus and combined prediction based on these two approaches [53].

NetCTLpan server predicts CTL epitopes in protein sequences and use artificial neural networks. The method has been updated to include the newest MHC allele releases from the IMGT/HLA and IPD-MHC databases (for non-human primates and pig). Predictions in this software can be made for 8-11mer peptides, since most HLA molecules have a strong preference for binding 9mers [Table 2 and 48].

The epitope discovery tools described above can be readily applied to most pathogens, although certain approaches are more suitable than others depending on their characteristics and limitations. However, one should keep in mind that this prediction is an indicator of potential function but is not a criterion for function assignment. Consequently these kind of *in silico* analyses should be combined with other pieces of evidence, including experimental data, to assign function.

4. Bioinformatics tools for predicting potential B cell binding-epitopes

B cell epitopes are recognized by B cell receptors or antibodies in their native structure. Continuous B cell epitope prediction is very similar to T cell epitope prediction, which has mainly been based on the amino acid properties such as hydrophilicity, charge, exposed surface area and secondary structure. Discontinuous B cell epitope prediction requires 3D structure of the antigen [54–56].

Table 3
Comprehensive list of B cell epitope prediction servers.

Server name	Link	Type
Bcepred	http://www.imtech.res.in/raghava/bcepred/	Prediction of continuous B-cell epitopes
BepiPred	http://www.cbs.dtu.dk/services/BepiPred/	Prediction of continuous B-cell epitopes
ABCPred	http://www.imtech.res.in/raghava/abcpred/	Prediction of continuous B-cell epitopes
BEST	http://biomine.ece.ualberta.ca/BEST/	Prediction of continuous B-cell epitopes
EPCES	http://sysbio.unl.edu/services/EPCES/	Prediction of discontinuous B-cell epitopes
Discotope	http://www.cbs.dtu.dk/services/DiscoTope/	Prediction of discontinuous B-cell epitopes
BEPro (PEPITO)	http://pepito.proteomics.ics.uci.edu/	Prediction of discontinuous B-cell epitopes
SEPPA	http://lifecenter.sgst.cn/seppa/index.php	Prediction of discontinuous B-cell epitopes
EpiSearch	http://curie.utmb.edu/episearch.html	Prediction of discontinuous B-cell epitopes
MimoPro	http://informatics.nenu.edu.cn/MimoPro	Prediction of discontinuous B-cell epitopes
MIMOX	http://immunet.cn/mimox/	Prediction of discontinuous B-cell epitopes
Pep-3D-Search	http://kyc.nenu.edu.cn/Pep3DSearch	Prediction of discontinuous B-cell epitopes
Epitopia	http://epitopia.tau.ac.il/	Prediction of continuous and discontinuous B-cell epitopes
PepSurf	http://pepitope.tau.ac.il	Prediction of continuous and discontinuous B-cell epitopes
ElliPro	http://tools.immuneepitope.org/tools/ElliPro/iedb_input	Prediction of continuous and discontinuous B-cell epitopes

SVM: support vector machine. *ANN:* artificial neural networks. *PSSM:* position-specific scoring matrix.

To date, some specific resources to predict continuous or discontinuous B-cell epitopes are available on the Web (Table 3).

To predict linear B-cell epitopes, the Bcepred tool is based on physicochemical properties such as hydrophilicity, flexibility, polarity, and exposed surface on a non-redundant dataset. The dataset consists of 1029 B-cell epitopes obtained from Bcipep database and an equal number of non-epitopes obtained randomly from Swiss-Prot database. The prediction accuracy for models based on these properties varies from 52.92% to 57.53% [57]. The ABCpred server, which is based on neural networks, has an estimated accuracy of 65.93% [54]. Another server called BepiPred predicts the location of linear B-cell epitopes using a combination of a hidden Markov model and a propensity scale method [58]. The servers mentioned above are easy to use and properly organized.

Among the tools used to predict discontinuous B cell epitopes we can mention DiscoTope, which uses the three dimensional structure of proteins to determinate the surface accessibility and a novel epitope propensity amino acid score. The final scores are calculated by combining the propensity scores of residues in spatial proximity and the contact numbers. This server also predicts epitopes in complexes of multiple chains [59]. This tool along with BEpro (formerly known as PEPITO) and SEPPA (Spatial Epitope Prediction of Protein Antigens) requires a 3-D structure as input, specifically, in PDB format [60]. Using SEPPA, each residue in the query protein will be given a score according to information from its neighborhood residues. The higher score corresponds to the higher probability of the residue to be involved in an epitope [61].

One of the most complete tools in this field is ElliPro. This server predicts linear and discontinuous epitopes based on a protein antigen's 3-D structure. ElliPro associates each predicted epitope with a score, defined as a PI (Protrusion Index) value. Compared with databases mentioned earlier, in ElliPro the input is a protein sequence. A 3-D structure will be predicted for the input protein sequence by homology modeling based on user-selected structural template. Afterwards, linear and discontinuous epitopes will be computed based on the predicted protein structure. Some other bioinformatics tools to predict continuous and discontinuous B cell epitopes are included in Table 3.

All of these integrative tools represent an opportunity for the development of new vaccines in special those that aim at the elicitation of humoral responses.

5. Bioinformatics strategies for emergent peptide-based vaccines against hypervariable viruses

Historically, most of the known successful vaccines have been developed empirically. However the emergence of highly sophisticated viruses, such as HIV and influenza characterized by having a high degree of genetic and antigenic diversity, has impeded the development of effective, broad-coverage vaccines using traditional methods. The rapid emergence of these viral pathogens underscores the need for improved and accelerated processes to develop and produce vaccines, a need that can be addressed by the methods described above allowing a rapid, *in silico*-based approach to formulate vaccine candidates.

This section briefly discusses some approaches developed for the case of the human immunodeficiency (HIV) and influenza viruses as examples on how successful candidate vaccine design can be achieved in the case of hypervariable viruses using bioinformatic tools.

5.1. HIV

Considering its pandemic importance worldwide, successful anti-HIV vaccine is an immediate need. This development constitutes a major challenge for researchers as HIV defeats immune system intended to neutralize it. In addition, genetic material can remain in dormant form and thus escapes the immune system [62]. Although antiretroviral drugs can control HIV/AIDS progression in many patients, they only succeed in reducing viral loads without completely eliminating the virus [63]. Therefore, the development of an effective HIV-1 vaccine represents the optimal solution for the control of the pandemic HIV.

One advanced proof of concept has been accomplished by the group of Diaz-Mitoma and coworkers [64] whom developed the Variosite-based HIV-1 vaccine. This candidate vaccine comprised a pool of 176 peptides representing five hypervariable epitopes of gp120 envelope and two variable epitopes of gag. The potency and coverage of this polyvalent vaccine was tested against a panel of heterologous HIV-1 subtypes in a non-human primate model. Specific CD8+ T-cell immune responses against HIV-1 subtypes A–F were detected, which is remarkable as HIV-1 sequences within a subtype differ by up to 20%, and between subtypes by up to 35%. Binding antibody titer and neutralizing activity were also characterized in the immunized animals, observing a substantial level of IgG antibody titer to variant gp120 proteins in all animals. In addition three out of six immunized macaques developed neutralization activity against two primary HIV-1 [64].

On the other hand, Huang et al. [65] reported one method for the prediction of conformational B-cell epitopes. In order to define whether or not certain regions in specific peptides may constitute B-cell epitopes, phage-displayed random peptide libraries were applied as a powerful tool in identifying mimotopes; which are selected by binding to a given monoclonal antibody (mAb) in a similar pattern to the native epitope. These mimotopes can be considered as functional epitope mimics. This kind of methods can predict not only linear but also conformational epitopes and thus this approach represent an important strategy in the field. This method is designated as Pep-3D-Search, it relies on the 3-D structure of a specific antigen and a set of mimotopes (or a motif sequence derived from the set of mimotopes), and can be used in two modalities: mimotope or motif. In order to evaluate the capacity to predict epitopes from a set of mimotopes, 10 epitopes defined by crystallography were compared with the predicted results from a Pep-3D-Search. Compared with other available prediction algorithms Pep-3D-Search showed comparable Matthews correlation coefficient (MCC), specificity and precision, and could provide novel, rational results. On the other hand, authors verified the capability of Pep-3D-Search to align a motif sequence to a 3-D structure in order to predict epitopes. Six test cases were analyzed including three HIV proteins, demonstrating a superior performance to other similar programs. In addition, the program is capable of quickly localize the epitope regions mimicked by longer isotopes. Although this promising tool provides a powerful approach to localize the surface region mimicked by the mimotopes, it is necessary to evaluate the immunoprotective capacity of these identified epitopes.

In the context of atomic-level structure of the antibody-antigen complex, structure determination in many cases may be impractical. Recently Georgiev and colleagues [66] describe an efficient computational method to predict antibody-specific HIV-1 envelope (Env) epitopes at the residue level. This method consists on assessing neutralization potency data over a set of diverse viral strains representing the antigen; enhanced accuracy could be achieved by incorporating information from the unbound structure of the antigen. In particular, 19 HIV-1 Env antibodies were evaluated in neutralization panels comprising 181 diverse viral strains and available antibody-antigen complex structures were considered in the analyses. The prediction efficiency was 8-fold higher than a random prediction. In addition when used to prospectively predict epitope residues for two HIV-1 antibodies, 8ANC131 and

8ANC195, this tool allowed for successful prediction that was validated experimentally. This procedure is inherently applicable to antigens that exhibit sequence diversity, displaying an accuracy that correlates inversely with epitope sequence conservation. These insights show how data derived from a neutralization panel and unbound antigen structure can be utilized for residue-level prediction of antibody epitopes, representing an important approach to generate efficacious vaccines against hypervariable pathogens [66].

Table 4 concentrates a representative view of those efforts based on reverse vaccinology approaches and reflects the feasibility to overcome the aforementioned obstacles under this focus. Therefore bioinformatic tools offer a powerful resource that lead to prediction of vaccine targets. Additional cases are provided in Table 4.

5.2. Influenza

Influenza is a highly contagious, airborne respiratory tract infection associated with a significant disease burden. New influenza subtypes periodically emerge to which no immunity exists in the human population and thus these may cause global pandemics. Outbreaks of new influenza subtypes exemplify how this pathogen can evolve pandemic [67]. The search for an effective universal broad-coverage influenza vaccine is one discouraging task because of the rapidly mutation rate, the variable and divergent characteristics of the virus, along with extremely long protocols for vaccine development [68].

Some studies have demonstrated that using immuno-bioinformatics predictor allows for a faster and global screening of epitopes to design candidate vaccines (see Table 5). In 2007, Wang et al. [69] identified ten novel influenza epitopes and confirmed three previously known restricted to cytotoxic T cell (CTL) by the use of the SYFPEITHI software. These epitopes are conserved in different isolates of the highly pathogenic H5N1 influenza virus and all of these are also present in the emerging bird flu isolates. The immunogenicity of the predicted peptides was evaluated by ELISPOT assay. These epitopes could be used to detect influenza specific CTL responses in patients. In addition their results would have important implications for the rational design of vaccines, using individual epitopes or epitopes fused as polytopes, applicable to all ethnic groups.

In a similar approach, Cheung et al. [70] targeted the influenza A virus (H5N1 strain) using the SYFPEITHI software achieving the identification of nine potential immunogenic peptides. *In vitro* assays were conducted to determine the immunogenic potential of these peptides. Their findings highlight a promising potential for epitopes AMDSNTLEL and QGRGVFEL. This novel cytotoxic T-cell epitopes constitute relevant information for the development of a human H5N1 vaccine.

On the other hand, using EpiMatrix (a T-cell epitope prediction and comparison tool), De Groot et al. [30] compared the sequences of the three hemagglutinin (HA) and neuraminidase (NA) proteins of influenza A virus. *In silico* analysis revealed sixteen promiscuous helper T-cell epitopes contained in the HA sequence, nine of which were 100% conserved in the 2008–2009 influenza vaccine strain. In a subsequent study conducted by the same group, a biological study was performed with the selected epitopes using peripheral blood mononuclear cell from human donors. IFN-γ ELISPOT and CD4+ T cell stimulation assays allowed to evaluate primary and post-boost T cell responses, observing a correlation between the *in silico* predictions with the observed responses with an 80–90% accurate prediction for CD4+ T cell epitopes. These findings reflect the robustness of this computational tool [71], with implications on the formulation of new vaccines against hypervariable pathogens. Table 5 shows other bioinformatic-based approaches applied in the development of new influenza vaccines, which taken together demonstrate that potential epitopes can be identified

Table 4
Potential candidate peptide-based vaccines against HIV designed with the aid of bioinformatic tools.

Database/resource	HIV proteins analized	Prediction output	Outcome	Immunological assay validation	Refs.
Propred1, IEDB consensus method, MODPROPEP	Gag, Nef, gp120, gp41, p31-integrase, p51RT, protease, rev, tat, vif, vpr, and vpu	Prediction of potential MHC binding (HLA-B*27:05)HIV epitopes	Fourteen peptides were identified to interact strongly with HLA-B*27:05	None	[107]
SYFPEITHI, BIMAS, Immunepitope	Gag, pol	Prediction of the cryptic HIV epitopes of gag and pol	Six affinity cryptic HIV epitopes presented by HLA-A*0201	ELISA, splenocyte proliferation assays, cytotoxic assays	[108]
BIMAS	Env, gag, nef, pol, rev, tat, vif, vpr, vpu	Prediction of MHC epitope binders	Select 30 epitope with 92% of coverage of all alleles	None	[109]
Conservatrix, EpiMatrix	Env, gag, pol,rev, tat, vif, vpr, vpu	Prediction of epitopes conserved across HIV-1 clades	Fifteen epitopes stimulated gamma-interferon release	ELISpot assays	[110]
Epi-Assembler, VaccineCAD, EpiMatrix EpiVax	Env, gag, pol, vpr, vif, tat, vpu, nef	Prediction of highly immunogenic conserved HLA-class II restricted epitopes	The 50% of the HIV epitopes tested induces specific immune tesponses in mice	Murine immunization studies and ELISpot assays	[111]
EpiMatrix	Env, gag, pol, vpr, vif, tat, vpu, nef	Prediction of broadly conserved T cell epitopes	The 45% of epitopes tested stimulated gamma-interferon release	ELISpot assays	[112]
Conservatrix, Epimatrix,	Env, pol, gag, vif, nef, tat, vpr	Prediction of CTL epitopes	Twenty-seven HLA-A3 epitopes are conserved across time, clades and geography	ELISA and ELISpot assays	[113]
Variosite	gp120 and gp41	Prediction of B cell neutralizing epitopes	Six macaques developed neutralization activity against two primary HIV-1	Proliferation assays, ELISPOT assays, intracellular cytokine staining, ELISA, neutralization assays	[64]
Pep-3D	gp120, p24 and nef	Prediction of B cell neutralizing epitopes and mimotope design	Quickly localize the epitope regions mimicked by longer isotopes	No specified	[65]
Unnamed method	Envelope (Env) epitopes	Prediction of B cell neutralizing epitopes	Neutralization-based method in combination with structural information allow an accurate prediction of B cell epitopes	ELISA	[66]

Table 5
Potential candidate peptide-based vaccines against influenza designed with the aid of bioinformatic tools.

Database/resource	Prediction output	Outcome	Immunological assay validation	Refs.
SYFPEITHI	Prediction of new immunogenic HLA class I restricted cytotoxic T cell epitopes	Eighty-nine peptides were confirmed as HLA-I binders, and 13 were confirmed as CTL targets	HLA class I binding assay, IFN-γ ELISPOT assay,	[69]
NetCTL	Prediction of HLA-restricted binding epitopes	Fifty five conserved sequences were predicted to have immune relevance as T-cell epitopes	HLA binding assay, T-cell assay, ELISPOT assay	[114]
EpiMatrix	Prediction of CD4(+) T-cell epitopes	Sixteen T-cell epitopes were 100% conserved in the 2008-2009 influenza strain. The pre-existing CD4(+) T cells can elicit cross-reactive effector responses against influenza	CD4+ T cell culture assay, IFN-γ ELISPOT assay, CD4+ T cell stimulation assay, intracellular cytokine staining	[30,68]
SYFPEITHI	Prediction of HLA binding peptides	Two novel HLA-A*0201 restricted epitopes	T2-cell binding assay, citotoxiticy assay, ELISpot assay	[70]
NetCTL	Prediction HLA-I restricted cytotoxic T cell epitopes	One hundred and thirty one peptides have affinities for the HLA-I supertypes and only 21 were found to induce T cell responses	IFN-γ ELISPOT assay, biochemical assay	[71]
BIMAS, SYFPEITHI, NetCTL,	Prediction of immunogenic HLA-A24 restricted CTL epitopes	Of 35 CTL predicted peptides, six peptides exhibited remarkable cytotoxic activity *in vivo*	Mice were subcutaneously vaccinated with the selected epitopes and they survived lethal influenza virus challenge	[115]

rapidly by an *in silico* genome analysis; followed by confirmation conducting experimental evaluations.

6. Perspectives

Bioinformatics tools have enabled the capability of selecting potential epitopes without running the risks involved in cultivating the pathogen of interest. This kind of methodology represents a huge advantage over conventional vaccinology techniques, including faster outputs and lower costs. The application of 'omics' technologies to this field has also revolutionized the way in which potential vaccine candidates can be identified. Proteomics and transcriptomics have been used as complementary approaches to genomics and are often more useful in identifying surface proteins during host–pathogen interaction.

Despite that numerous epitope prediction methods are available, developing a systematic assessment of different methods on standard benchmark datasets is still a need. Launching a Critical Assesment of Techniques for Epitope Prediction will indeed benefit the field. It has been proposed that computational methods will be used to performed blinded de novo epitope prediction from query proteins previously screened experimentally [72,73]. Comparison of different methods is yet a complex task due to many aspects including the following: (i) inadequate documentation of datasets and prediction methods, (ii) unavailability of the benchmark dataset used to evaluate the methods, (iii) unavailability of the code that implements the method, (iv) the lack of a unified output format, which complicates the process of combining the results of several servers in order to obtain consensus predictions [74,75]. Therefore it is necessary to develop standardized data representations; this will enable the evaluation of different prediction methods on a standardized benchmark datasets in order to compare the methods and develop meta-servers combining the predictions of multiple prediction tools [55].

For example, Epitopes Toolkit (EpiT) is a platform design to develop epitope prediction methods. This allows other researchers to use the developed predictor on their own machines, rebuild the predictor on other datasets, or combine predictor with other predictors in order to obtain a customized hybrid or consensus predictor. EpiT comprises two components: (a) model builder, an application for building and evaluating epitope predictors and serializing these models in a binary format; (b) predictor, an application for applying a model to test data. Although EpiT was designed for developing epitope prediction tools, some of EpiT components can be used in different sequence classification tasks. Moreover, some data pre-processors in EpiT can be applied to a subcellular localization, and other protein sequence classifications [74–76]. This platform is implemented in Java and can be freely downloaded from the project web site at http://ailab.cs.iastate.edu/epit. The web site also offers a rich resource for the developers of epitope prediction tools and for EpiT users. The useful resources include: Epit documentation, an expanding Repository of Epitope Predictors and a Repositore of Epitope Datasets [74].

Other limitation is related to training an epitope predictor using a dataset, in which some epitope residues are incorrectly labeled as non-epitope residues and thus performance of the predictors on such dataset tend to exaggerate the number of false positives. Many authors [77–80] have attempted to improve this limitation by using non-epitope residues extracted form random sample of the protein sequences in SwissProt [81].

In addition, the comparative analyses are indispensable in order to improve the limitations of different epitope prediction methods, speeding up the development of enhanced methods. The critical assessment for epitope prediction methods has been demonstrated to be appropriate in other areas [82]. The use of large, non-redundant, and experimentally well-characterized datasets can help to increase the accuracy of the cross-validation-based estimates of the performance of epitope predictors.

All the bioinformatics tools described above provide a comprehensive list of epitopes to design a candidate vaccine. How to decide which of this large number of epitopes candidates are selected for clinical trials still remains a challenge. Although genomic based-approaches have tremendously accelerated the identification of vaccine targets, vaccinologists still have to go through the tedious and slow process of validation.

Thus the novel candidate epitope-based vaccines identified using these tools have to be subjected to the existing standard confirmatory *in vivo* studies (e.g. animal protection experiments) as one would do for vaccine antigens identified using conventional methods.

Different databases are available for predicting T- and B-cell epitopes. However, among the tools for predicting B cell epitopes, further developments are still needed since an ideal tool should

take in account the composition biases in datasets [83,84]. A realistic prediction of antigenic epitopes that are preferentially recognized critically depends of the protein composition. The majority of available epitope prediction methods are based upon the amino acid properties including hydrophilicity [85,86], solvent accessibility [84], secondary structure [87], flexibility [88] and antigenicity [83]. In 2005, Blythe and Flower demonstrated that, using single-scale amino acid propensity profiles is not enough to predict epitope location reliably, whereas in 2007, Greenbaum et al. showed that, using a combination of more than one amino acid propensity scale and a machine learning algorithms could improve prediction accuracy. The implication of such for studies obviously has immunological implications in epitope prediction [55,89].

Underperformance of *in silico* epitope prediction can be identified in a number of reports. As an example, Blythe and Flower examined the correlation between the predictor B-cell epitopes and the epitope location from 50 proteins for polyclonal responses. They found that the single-scale amino acid propensity profiles cannot be used to predict epitope location reliably as no correlation exists between sequence profile generated and the location of known linear epitopes. The appropriate combination of several algorithms, and critical evaluation of the information generated, are essential for the successful selection of antigenic epitopes. Therefore vaccinologists should be aware on the fact that outputs are highly dependent on the criteria used for tool selection. This fact highlights the complexicity of the task and motivates the development of more sophisticated and specific tools to improve performance [55].

In a report from Wang et al. MHC-II prediction tools did not perform as well as noted for class I predictions, which can be due to the fact that MHC class II requires a highly specific match. The prediction ability was initially assessed using 9-mer peptide cores revealed in crystal structures. From the eight algorithms assessed, a poor performance was observed except for PROPRED and SYFPEITHI, which reflect the difficulty to identify the correct binding core. In order to improve these limitations, a consensus prediction tool was develop by combining these two class II binding prediction methods, observing an improved performance when compared to individual outputs as 10 of the 14 candidates bound MHCII [90].

One interesting example has been published by Resende et al. who reported the evaluation of the performance of five algorithms for epitope prediction (NetCTL, NetMHC, BepiPred and AAP12) and three algorithms for subcellular localization prediction having as target trypanosomatid genomes. The comparison between algorithms was made in the basis of AUC (area under a ROC curve) values, which represent the probability that a randomly selected positive epitope will score higher than a randomly selected negative epitope. AUC data indicated that only a little difference among the NetCTL and Net MHC (0.66 and 0.60 respectively) performances exists. Considering that MHC-I prediction methods have achieved AUC values in the range 0.95–0.99, NetCTL and Net MHC were determined as low performance approaches. On the other hand, the evaluation of AUC values for B-cell epitopes algorithms (BepiPred and AAP12) revealed inferior performances than those observed for previous algorithms [91,92].

Therefore one can state that distinct prediction models vary in performance and thus a set of tools should be assessed for a specific target organism. It is then expected that this combinatorial prediction approach will provide solid advancements in the vaccine development field.

Once the selection of the target epitopes has been accomplished, designing vaccine formulation may be based in synthetic peptide mixtures that may include adjuvant sequences such as Th epitopes to enhance immunogenicity [for review see 93]. However synthetic peptides are costly for clinical applications. An attractive alternative consists on designing recombinant proteins bearing the set of target epitopes [94]. Additional sequences that play an important role in the chimeric protein are linkers, acting as spacers that facilitate folding to achieve a proper epitope display, and adjuvant sequences. Excellent reviews on vaccine chimeric protein designs were published by Thomas and Luxon [95] and Cozzi et al. [96].

Our group have produced and evaluated multiepitopic proteins as vaccines candidates against infectious diseases. Immunogenicity of these candidates in a mouse model has showed to be promising as humoral responses against the distinct targets were induced [97–99]. These designs comprised epitopes that were mainly described by traditional immunogenicity trials such as epitope mapping and serologic and functional assays. It is envisioned that the bioinformatic tools summarized in this review may lead to innovative chimeric designs against these pathogens.

Specific cases where these technologies would also provide crucial strides within the biomedical field include HIV/AIDS and influenza. Despite the global HIV/AIDS epidemics, the search for one effective vaccine was elusive after decades of focused research. However new hopes were generated from the RV144 Thai trial, where a modest protective potential was identified [100]. Epitopes associated with the observed immunoprotection are being identified and it is expected that further vaccine developments based in those epitopes would provide new promising candidates [101]. The development of vaccines consisting of protective selected epitopes is the ideal approach and thus represents a field of opportunity for the fight against HIV/AIDS.

In the case of influenza although protective vaccines are available, these cover certain variants with the risk of failing on protecting against new variants and thus new vaccines should be periodically generated to fight emerging variants. A universal vaccine is the ideal approach to provide broad-spectrum immunity against multiple strains. Among the candidates investigated thus far, the highly conserved ectodomain of the M2 protein of influenza A viruses and a stalk domain can be mentioned [102]. T cell epitope-based vaccines using nucleoprotein and M1 protein have also been tested. Multiepitopic designs comprising these promising epitopes augur new potential influenza vaccine as model for a universal intervention against this relevant hypervariable pathogen [103,104].

In conclusion, the extensive research accomplished on developing tools useful on vaccine rational design are considerable but further improvements on *in silico* analysis along with experimental evaluations will be critical to advance in the vaccine development field to eventually introduce to the market new vaccines derived from these technologies, especially for highly variable viral pathogens [105,106].

Acknowledgments

Current investigations from the group are supported by CONACYT Grants 151480 to RESG, PROMEP-2010 to Bioprocess CA. Omar González corrected the English version of the manuscript.

References

[1] WHO, UNICEF, World Bank. State of the world's vaccines and immunization, 3rd ed. Geneva, World Health Organization; 2009.
[2] Arinaminpathy N, Ratmann O, Koelle K, Epstein SL, Price GE, Viboud C, et al. Impact of cross-protective vaccines on epidemiological and evolutionary dynamics of influenza. PNAS 2012;109:3173–7.
[3] Jarząb A, Skowicki M, Witkowska D. Subunit vaccines- antigens, carriers, conjugation methods and the role of adjuvants. Postepy Hig Med Dosw 2013;67:1128–43.
[4] Bogdanos DP, Choudhuri K, Vergani D. Molecular mimicry and autoimmune liver disease: virtuous intentions, malign consequences. Liver 2001;21:225–32.
[5] Olson JK, Croxford JL, Calenoff MA, Dal Canto MC, Miller SD. A virus-induced molecular mimicry model of multiple sclerosis. J Clin Invest 2001;108:311–8.

[6] Lesk AM. Introduction to bioinformatics. New York, NY: Oxford University Press; 2002. EEUU, 255.
[7] Brusic V, Flower D. Bioinformatics tools for identifying T-cell epitopes. Drug Disc Today: Biosilico 2004;2:8–23.
[8] Jackwood MW, Hickle L, Kapil S, Silva R. Vaccine development using recombinant DNA technology. Council Agric Sci Technol 2008;38:1–11.
[9] Streatfield SJ. Plant based vaccines for animal health. Rev Sci Techol 2005;24:189–99.
[10] Verma AK, Kumar A, Dhama K, Deb R, Rahal A, Mahima, et al. Leptospirosis-persistence of a dilemma: an overview with particular emphasis on trends and recent advances in vaccines and vaccination strategies. Pak J Biol Sci 2012;15:954–63.
[11] Ellis RW. New Technologies for making vaccines. Vaccine 1999;17:1596–604.
[12] Adamczyk-Poplawska M, Markowicz S, Jagusztyn-Krynicka EK. Proteomics for development of vaccine. J Proteomic 2011;74:2596–616.
[13] Klade CS. Proteomics approaches towards antigen discovery and vaccine development. Curr Opin Mol Ther 2002;4:216–23.
[14] Falisse-Poirrier N, Ruelle V, ElMoualij B, Zorzi D, Pierard O, Heinen E, et al. Advances in immunoproteomics for serological characterization of microbial antigens. J Microbiol Methods 2006;67:593–6.
[15] Brusic V. Immunoinformatics: tools for large-scale immunology. Genome Inform 2003;14:711–2.
[16] Korber B, LaBute M, Yusim K. Immunoinformatics comes of age. PLoS Comput Biol 2006;2:e71.
[17] Tong JC, Ren EC. Immunoinformatics: current trends and future directions. Drug Discov Today 2009;14:684–9.
[18] Tomar N, De RK. Immunoinformatics: an integrated scenario. Immunology 2010;131:153–68.
[19] Babai I, Samira S, Barenholz Y, Zakay-Rones Z, Kedar E. A novel influenza subunit vaccine composed of liposome-encapsulated haemagglutinin/neuraminidase and IL-2 or GM-CSF. I: Vaccine characterization and efficacy studies in mice. Vaccine 1999;17:1223–38.
[20] Nemchinov LG, Liang TJ, Rifaat MM, Mazyad HM, Hadidi A, Keith JM. Development of a plant-derived subunit vaccine candidate against hepatitis C virus. Arch Virol 2000;145:2557–73.
[21] Sun DX, Seyer JM, Kovari I, Sumrada RA, Taylor RK. Localization of protective epitopes within the pilin subunit of the Vibrio cholerae toxin-coregulated pilus. Infect Immunol 1991;59:114–8.
[22] Almeida RR, Rosa DS, Ribeiro SP, Santana VC, Kallás EG, Sidney J, et al. Broad and cross-clade CD4+ T-cell responses elicited by a DNA vaccine encoding highly conserved and promiscuous HIV-1 M-group consensus peptides. PLoS ONE 2012;7:e45267.
[23] Iurescia S, Fioretti D, Fazio VM, Rinaldi M. Epitope-driven DNA vaccine design employing immunoinformatics against B-cell lymphoma: a biotech's challenge. Biotechnol Adv 2012;30:372–83.
[24] Gershoni JM, Roitburd A, Siman DD, Tarnovitsk N, Weiss Y. Epitope mapping the first step in developing epitope-based vaccines. Biodrugs 2007;21:145–56.
[25] Foged C, Hansen J, Agger EM. License to kill: formulation requirements for optimal priming of CD8(+) CTL responses with particulate vaccine delivery systems. Eur J Pharm Sci 2012;45:482–91.
[26] Ling A, Whitton JL. A multivalent minigene vaccine, containing B cell, CTL, and Th epitopes from several microbes, induces appropriate responses in vivo and confers protection against more than one pathogen. J Virol 1997;71:2292–302.
[27] Purcell AW, Zeng W, Mifsud NA, Ely LK, Macdonald WA, Jackson DC. Dissecting the role of peptides in the immune response: theory, practice and the application to vaccine design. J Pept Sci 2003;9:255–81.
[28] Van Regenmortel MH. What is a B-cell epitope? Methods Mol Biol 2009;524:3–20.
[29] He Y, Rappuoli R, De Groot AS, Chen R. Emerging vaccine informatics. J Biomed Biotechnol 2010;2010:1–26.
[30] De Groot AS, McMurry J, Moise L, Martin B. Epitope-based Immunome-derived vaccines: a strategy for improved design and safety. Clinical Appl Immun 2009;2:39–69.
[31] DeLisi C, Berzofsky JA. T-cell antigenic sites tend to be amphipathic structures. PNAS 1985;82:7048–52.
[32] Zinkernagel RM, Doherty PC. The discovery of MHC restriction. Immunol Today 1997;18:14–7.
[33] Falk K, Rotzschke O, Stevanovic S, Jung G, Rammensee HG. Allele-specific motifs revealed by sequencing of self-peptides eluted from MHC molecules. Nature 1991;351:290–6.
[34] Rötzschke O, Falk K, Stevanović S, Jung G, Walden P, Rammensee HG. Exact prediction of a natural T cell epitope. Eur J Immunol 1991;21:2891–4.
[35] McMurry J, Sbai H, Gennaro ML, Carter EJ, Martin W, De Groot AS. Analyzing Mycobacterium tuberculosis proteomes for candidate vaccine epitopes. Tuberculosis 2005;85:95–105.
[36] De Groot AS, Moise L. New tools, new approaches and new ideas for vaccine development. Expert Rev Vaccine 2007;6:125–7.
[37] Hoof I, Peters B, Sidney J, Pedersen LE, Sette A, Lund O, et al. NetMHCpan, a method for MHC class I binding prediction beyond humans. Immunogenetics 2009;61:1–13.
[38] Lundegaard C, Lund O, Nielsen M. Accurate approximation method for prediction of class I MHC affinities for peptides of length 8, 10 and 11 using prediction tools trained on 9mers. Bioinformatics 2008;24:1397–8.
[39] Reche PA, Reinherz EL. Prediction of peptide-MHC binding using profiles. Methods Mol Biol 2007;409:185–200.
[40] Bhasin M, Raghava GP. A hybrid approach for predicting promiscuous MHC class I restricted T cell epitopes. J Biosci 2007;32:31–42.
[41] Peters B, Bulik S, Tampe R, Van Endert PM, Holzhütter HG. Identifying MHC class I epitopes by predicting the TAP transport efficiency of epitope precursors. J Immunol 2003;171:1741–9.
[42] Lankat-Buttgereit B, Tampe R. The transporter associated with antigen processing TAP: structure and function. FEBS Lett 1999;464:108–12.
[43] Hill A, Ploegh H. Getting the inside out: the transporter associated with antigen processing (TAP) and the presentation of viral antigen. PNAS 1995;92:341–3.
[44] Larsen MV, Lundegaard C, Lamberth K, Buus S, Brunak S, Lund O, et al. An integrative approach to CTL epitope prediction: a combined algorithm integrating MHC class I binding, TAP transport efficiency, and proteasomal cleavage predictions. Eur J Immunol 2005;35:2295–303.
[45] Bhasin M, Raghava GPS. Prediction of CTL epitopes using QM, SVM and ANN techniques. Vaccine 2004;22:3195–201.
[46] Doytchinova IA, Guan P, Flower DR. EpiJen: a server for multi-step T cell epitope prediction. BMC Bioinformatics 2006;7:131.
[47] Dönnes P, Elofsson A. Prediction of MHC class I binding peptides, using SVMHC. BMC Bioinformatics 2001;3:25.
[48] Stranzl T, Larsen MV, Lundegaard C, Nielsen M. NetCTLpan: pan-specific MHC class I pathway epitope predictions. Immunogenetics 2010;62:357–68.
[49] Larsen MV, Lundegaard C, Lamberth K, Buus S, Lund O, Nielsen M. Large-scale validation of methods for cytotoxic T-lymphocyte epitope prediction. BMC Bioinformatics 2007;8:e424.
[50] Lin HH, Zhang GL, Tongchusak S, Reinherz EL, Brusic V. Evaluation of MHC-II peptide binding prediction servers: applications for vaccine research. BMC Bioinformatics 2008;9:S22.
[51] Gowthaman U, Agrewala JN. In silico tools for predicting peptides binding to HLA-class II molecules: more confusion than conclusion. J Proteome Res 2008;7:154–63.
[52] Singh H, Raghava GPS. ProPred: Prediction of HLA-DR binding sites. Bioinformatics 2001;17:1236–7.
[53] Bhasin M, Raghava GPS. Analysis and prediction of affinity of TAP binding peptides using cascade SVM. Protein Sci 2004;13:596–607.
[54] Sun P, Ju H, Liu Z, Ning Q, Zhang J, Zhao X, et al. Bioinformatics resources and tools for conformational B-cell epitope prediction. Comput Math Methods Med 2013;2013:943636.
[55] Greenbaum JA, Andersen PH, Blythe M, Bui HH, Cachau RE, Crowe J, et al. Towards a consensus on datasets and evaluation metrics for developing B-cell epitope prediction tools. J Mol Recognit 2007;20:75–82.
[56] Yao B, Zheng D, Liang S, Zhang C. Conformational B-cell epitope prediction on antigen protein structures: a review of current algorithms and comparison with common binding site prediction methods. PLoS ONE 2013;19(8):e62249.
[57] Saha S, Raghava GPS. Prediction of continuous B-cell epitopes in an antigen using recurrent neural network. Proteins 2006;65:40–8.
[58] Pontoppidan-Larsen JE, Lund O, Nielsen M. Improved method for predicting linear B-cell epitopes. Immun Res 2006;2:2.
[59] Kringelum JV, Lundegaard C, Lund O, Nielsen M. Reliable B cell epitope predictions: impacts of method development and improved benchmarking. Plos Comput Biol 2012;8:12.
[60] Sweredoski MJ, Baldi P. PEPITO: improved discontinuous B-cell epitope prediction using multiple distance thresholds and half-sphere exposure. Bioinformatics 2008;24:1459–60.
[61] Sun J, Wu D, Xu T, Wang X, Xu X, Tao L, et al. SEPPA: a computational server for spatial epitope prediction of protein antigens. Nucl Acids Res 2009;37:W612–6.
[62] Chomont N, El-Far M, Ancuta P, Trautmann L, Procopio FA, Yassine B, et al. HIV reservoir size and persistence are driven by T cell survival and homeostatic proliferation. Nat Med 2009;15:893–900.
[63] Shen L, Peterson S, Sedaghat AR, McMahon MA, Callender M, Zhang H, et al. Dose-response curve slope sets class-specific limits on inhibitory potential of anti-HIV drugs. Nat Med 2008;14:762–6.
[64] Azizi A, Anderson DE, Torres JV, Ogrel A, Ghorbani M, Soare C. Induction of broad cross-subtype-specific HIV-1 immune responses by a novel multivalent HIV-1 peptide vaccine in cynomolgus macaques. J Immunol 2008;180:2174–86.
[65] Huang YX, Bao YL, Guo SY, Wang Y, Zhou CG, Li YX. Pep-3D-Search: a method for B-cell epitope prediction based on mimotope analysis. BMC Bioinformatics 2008;9:538.
[66] Chuang GY, Acharya P, Schmidt SD, Yang Y, Louder MK, Zhou T, et al. Residue-level prediction of HIV-1 antibody epitopes based on neutralization of diverse viral strains. J Virol 2013;87:10047–58.
[67] Chan PK. Outbreak of avian influenza A (H5N1) virus infection in Hong Kong in 1997. Clin Infect Dis 2002;34:S58–64.
[68] Schanen BC, De Groot AS, Moise L, Ardito M, Mcclaine E, Martin W, et al. Coupling sensitive in vitro and in silico techniques to assess cross-reactive CD4+ T cells against the swine-origin H1N1 influenza virus. Vaccine 2011;29:3299–309.
[69] Wang M, Lamberth K, Harndahl M, Røder G, Stryhn A, Larsen MV, et al. CTL epitopes for influenza A including the H5N1 bird flu; genome-, pathogen-, and HLA-wide screening. Vaccine 2007;25:2823–31.

[70] Cheung YK, Cheng SC, Ke Y, Xie Y. Two novel HLA-A*0201 T-cell epitopes in avian H5N1 viral nucleoprotein induced specific immune responses in HHD mice. VetRes 2010;41:24.
[71] Wang M, Larsen MV, Nielsen M, Harndahl M, Justesen S, Dziegiel MH, et al. HLA Class I binding 9mer peptides from influenza A virus induce CD4+ T cell responses. PLoS ONE 2010;5:e10533.
[72] Lafuente EM1, Reche PA. Prediction of MHC-peptide binding: a systematic and comprehensive overview. Curr Pharm Des 2009;15(28):3209–20.
[73] Ostell JM, Wheelan SJ, Kans JA. The NCBI data model. In: Baxevanis AD, Ouellette BFF, editors. Bioinformatics: A Practical Guide to the Analysis of Genes and Proteins. New York: John Wiley & Sons Publishing; 2001. p. 19–44.
[74] El-Manzalawy Y, Honavar V. A framework for developing epitope prediction tools. In: Proceedings of the first ACM international conference on bioinformatics and computational biology, ACM; 2010a. p. 660–2.
[75] EL-Manzalawy Y, Honavar V. Recent advances in B-cell epitope prediction methods. Immun Res 2010;6:1–9.
[76] EL-Manzalawy Y, Honavar V. Building classifier ensembles for B-cell epitopes prediction. In: De Rajat K, Tomar N, editors. Immunoinformatics, methods in molecular biology. New York: Springer; 2014. p. 285–94.
[77] Chen J, Liu H, Yang J, Chou K. Prediction of linear B-cell epitopes using amino acid pair antigenicity scale. Amino Acids 2007;33:423–8.
[78] EL-Manzalawy Y, Dobbs D, Honavar V. Predicting flexible length linear Bcell epitopes. In: 7th International conference on computational systems bioinformatics; 2008. p. 121–31.
[79] Saha S, Raghava G. Prediction of continuous B-cell epitopes in an antigen using recurrent neural network. Proteins 2006;65:40–8.
[80] Söllner J, Mayer B. Machine learning approaches for prediction of linear B-cell epitopes on proteins. J Mol Recognit 2006;19:200–8.
[81] Bairoch A, Apweiler R. The SWISS-PROT protein sequence database and its supplement TrEMBL in 2000. Nucl Acids Res 2000;28:45–8.
[82] Moult J. A decade of CASP: progress, bottlenecks and prognosis in protein structure prediction. Curr Opin Struct Biol 2005;15(3):285–9.
[83] Emini EA, Hughes JV, Perlow DS, Boger J. Induction of hepatitis A virus-neutralizing antibody by a virus-specific synthetic peptide. J Virol 1985;55:836–9.
[84] Kolaskar AS, Tongaonkar PC. A semi-empirical method for prediction of antigenic determinants on protein antigens. FEBS Lett 1990;276:172–4.
[85] Parker JM, Guo D, Hodges RS. New hydrophilicity scale derived from high-performance liquid chromatography peptide retention data: correlation of predicted surface residues with antigenicity and X-ray-derived accessible sites. Biochemistry 1986;25:5425–32.
[86] Hopp TP, Woods KR. Prediction of protein antigenic determinants from amino acid sequences. Proc Natl Acad Sci USA 1981;78:3824–8.
[87] Pellequer JL, Westhof E, Van Regenmortel MH. Correlation between the location of antigenic sites and the prediction of turns in proteins. Immunol Lett 1993;36:83–99.
[88] Karplus PA, Schulz GE. Prediction of chain flexibility in proteins a -tool for the selection of peptide antigens. Naturwissenschaften 1985;72:212–3.
[89] Blythe MJ, Flower DR. Benchmarking B cell epitope prediction: underperformance of existing methods. Protein Sci 2005;14:246–8.
[90] Wang P, Sidney J, Dow C, Mothé B, Sette A, Peters B. A systematic assessment of MHC class II peptide binding predictions and evaluation of a consensus approach. PLoS Comput Biol 2008;4(4):e1000048.
[91] Larsen JE, Lund O, Nielsen M. Improved method for predicting linear B-cell epitopes. Immun Res 2006;2:2.
[92] Resende DM1, Rezende AM, Oliveira NJ, Batista IC, Corrêa-Oliveira R, Reis AB, et al. An assessment on epitope prediction methods for protozoa genomes. BMC Bioinformatics 2012;13:309.
[93] Rosendahl Huber S, van Beek J, de Jonge J, Luytjes W, van Baarle D. T cell responses to viral infections-opportunities for peptide vaccination. Front Immunol 2014;5:171.
[94] Zeinalzadeh N, Salmanian AH, Ahangari G, Sadeghi M, Amani J, Bathaie SZ, et al. Design and characterization of a chimeric multiepitope construct containing CfaB, heat-stable toxoid, CssA, CssB, and heat-labile toxin subunit B of enterotoxigenic Escherichia coli: a bioinformatic approach. Biotechnol Appl Biochem 2013. http://dx.doi.org/10.1002/bab.1196.
[95] Thomas S, Luxon BA. Vaccines based on structure-based design provide protection against infectious diseases. Expert Rev Vaccines 2013;12:1301–11.
[96] Cozzi R, Scarselli M, Ferlenghi I. Structural vaccinology: a three-dimensional view for vaccine development. Curr Top Med Chem 2013;13:2629–37.
[97] Soria-Guerra RE, Rosales-Mendoza S, Moreno-Fierros L, López-Revilla R, Alpuche-Solís AG. Oral immunogenicity of tomato-derived sDPT polypeptide containing Corynebacterium diphtheriae, Bordetella pertussis and Clostridium tetani exotoxin epitopes. Plant Cell Rep 2011;30:417–24.
[98] Govea-Alonso DO, Rubio-Infante N, García-Hernández AL, Varona-Santos JT, Korban SS, Moreno-Fierros L, et al. Immunogenic properties of a lettuce-derived C4(V3)6 multiepitopic HIV protein. Planta 2013;238:785–92.
[99] Rosales-Mendoza S, Rubio-Infante N, Elizabeth Monreal-Escalante, Govea-Alonso DO, García-Hernández AL, Salazar-González JA, et al. Chloroplast expression of an HIV envelop-derived multiepitope protein: towards a multivalent plant-based vaccine. PCTOC 2014;116:111–23.
[100] Excler JL, Tomaras GD, Russell ND. Novel directions in HIV-1 vaccines revealed from clinical trials. Curr Opin HIV AIDS 2013;8:421–31.
[101] Rao M, Peachman KK, Kim J, Gao G, Alving CR, Michael NL, et al. HIV-1 variable loop 2 and its importance in HIV-1 infection and vaccine development. Curr HIV Res 2013;11:427–38.
[102] Shaw A. New technologies for new influenza vaccines. Vaccine 2012;30:4927–33.
[103] Pica N, Palese P. Toward a universal influenza virus vaccine: prospects and challenges. Annu Rev Med 2013;64:189–202.
[104] Goodman AG, Heinen PP, Guerra S, Vijayan A, Sorzano CO, Gomez CE, et al. A human multi-epitope recombinant vaccinia virus as a universal T cell vaccine candidate against influenza virus. PLoS ONE 2011;6:e25938.
[105] Oberg AL, Kennedy RB, Li P, Ovsyannikova IG, Poland GA. Systems biology approaches to new vaccine development. Curr Opin Immunol 2011;23:436–43.
[106] Andersen-Nissen E, Heit A, McElrath MJ. Profiling immunity to HIV vaccines with systems biology. Curr Opin HIV AIDS 2012;7:32–7.
[107] Sundaramurthi JC, Ramanathan VD, Hanna LE. HLA-B*27:05-specific cytotoxic T lymphocyte epitopes in Indian HIV type 1C. AIDS Res Hum Retroviruses 2013;29:47–53.
[108] Li Y, Huang Y, Liang J, Xu Z, Shen Y, Zhang N, et al. Immune responses induced in HHD mice by multiepitope HIV vaccine based on cryptic epitope modification. Mol Biol Rep 2013;40:2781–7.
[109] Toussaint NC, Maman Y, Kohlbacher O, Louzoun Y. Universal peptide vaccines – optimal peptide vaccine design based on viral sequence conservation. Vaccine 2011;29:8745–53.
[110] De Groot AS, Bosma A, Chinai C, Frost J, Jesdale BM, Gonzalez MA, et al. From genome to vaccine: in silico predictions, ex vivo verification. Vaccine 2001;19:4385–95.
[111] De Groot AS, Marcon L, Bishop EA, Rivera D, Kutzler M, Weiner DB, et al. HIV vaccine development by computer assisted design: the GAIA vaccine. Vaccine 2005;23:2136–48.
[112] De Groot AS, Jesdale B, Martin W, Saint Aubin C, Sbai H, Bosma A, et al. Mapping cross-clade HIV-1 vaccine epitopes using a bioinformatics approach. Vaccine 2003;21:4486–504.
[113] De Groot AS, Levitz L, Ardito MT, Skowron G, Mayer KH, Buus S, et al. Further progress on defining highly conserved immunogenic epitopes for a global HIV vaccine: HLA-A3-restricted GAIA vaccine epitopes. Hum Vaccin Immunother 2012;8:987–1000.
[114] Heiny AT, Miotto O, Srinivasan KN, Khan AM, Zhang GL, Brusic V, et al. Evolutionarily conserved protein sequences of influenza a viruses, avian and human, as vaccine targets. PLoS ONE 2007;2:e1190.
[115] Ichihashi T, Yoshida R, Sugimoto C, Takada A, Kajino K. Cross-protective peptide vaccine against influenza A viruses developed in HLA-A*2402 human immunity model. PLoS ONE 2011;6:e24626.