We are IntechOpen, the world's leading publisher of Open Access books
Built by scientists, for scientists

4,800
Open access books available

122,000
International authors and editors

135M
Downloads

154
Countries delivered to

Our authors are among the

TOP 1%
most cited scientists

12.2%
Contributors from top 500 universities

CLARIVATE ANALYTICS
BOOK CITATION INDEX
INDEXED

WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

The Impact of Bioinformatics on Vaccine Design and Development

Ribas-Aparicio Rosa María,
Castelán-Vega Juan Arturo, Jiménez-Alberto Alicia,
Monterrubio-López Gloria Paulina and
Aparicio-Ozores Gerardo

Additional information is available at the end of the chapter

Abstract

Vaccines are the pharmaceutical products that offer the best cost-benefit ratio in the prevention or treatment of diseases. In that a vaccine is a pharmaceutical product, vaccine development and production are costly and it takes years for this to be accomplished. Several approaches have been applied to reduce the times and costs of vaccine development, mainly focusing on the selection of appropriate antigens or antigenic structures, carriers, and adjuvants. One of these approaches is the incorporation of bioinformatics methods and analyses into vaccine development. This chapter provides an overview of the application of bioinformatics strategies in vaccine design and development, supplying some successful examples of vaccines in which bioinformatics has furnished a cutting edge in their development. Reverse vaccinology, immunoinformatics, and structural vaccinology are described and addressed in the design and development of specific vaccines against infectious diseases caused by bacteria, viruses, and parasites. These include some emerging or re-emerging infectious diseases, as well as therapeutic vaccines to fight cancer, allergies, and substance abuse, which have been facilitated and improved by using bioinformatics tools or which are under development based on bioinformatics strategies.

Keywords: reverse vaccinology, immunoinformatics, structural vaccinology, computational strategies, vaccine

1. Introduction

The success of vaccination is reflected in its worldwide impact by improving human and veterinary health and life expectancy. It has been asserted that vaccination, as well as clean water, has had such a major effect on mortality reduction and population growth [1, 2]. In addition to the invaluable role of traditional vaccines to prevent diseases, the society has observed remarkable scientific and technological progress since the last century in the improvement of these vaccines and the generation of new ones. This has been possible by the fusion of computational technologies with the application of recombinant DNA technology, the fast growth of biological and genomic information in database banks, and the possibility of accelerated and massive sequencing of complete genomes [3–5]. This has aided in expanding the concept and application of vaccines beyond their traditional immunoprophylactic function of preventing infectious diseases, and also serving as therapeutic products capable of modifying the evolution of a disease and even cure it [3]. Vaccines are the pharmaceutical products that offer the best cost-benefit ratio in the prevention or treatment of diseases. In that it is a pharmaceutical product, a vaccine development and production are costly and it takes years for this to be accomplished. Several approaches have been applied to reduce the times and costs of their development, mainly focusing on the selection of appropriate antigens or antigenic structures, carriers, and adjuvants [6]. One of these approaches is the incorporation of bioinformatics methods and analyses into vaccine development. At present, there are many alternative strategies to design and develop effective and safe new-generation vaccines, based on bioinformatics approaches through reverse vaccinology, immunoinformatics, and structural vaccinology [7]. This chapter provides an overview of the application of bioinformatics strategies in vaccine design and development, supplying some successful examples of vaccines in which bioinformatics has furnished a cutting edge in their development.

2. Reverse vaccinology

Reverse vaccinology is a methodology that uses bioinformatics tools for the identification of structures from bacteria, virus, parasites, cancer cells, or allergens that could induce an immune response capable of protecting against a specific disease [7].

This approach possesses many advantages over traditional vaccinology: it reduces time and cost in vaccine development; refines the number of proteins to be studied, facilitating the selection process; can identify antigens present in small amounts or expressed only at certain stages, which would hinder or prevent their purification; and allows for the study of noncultivable or risky microorganisms [3]

An important requirement for utilizing this methodology is the availability of genomic information of the pathogen under study and, in some instances, even the human or animal cell genome must be known (i.e., DNA vaccines and therapeutic vaccines). Once the genome sequence is obtained, it is possible to identify all likely proteins that could be expressed. For this purpose,

several software systems and programs identify all open reading frames (ORFs) that constitute the sequences expressing the majority of proteins [8–10].

The next step in reverse vaccinology is to determine several antigenic and physicochemical properties that have been associated with good antigens. These characteristics must be analyzed for each protein in the proteome under study, employing different bioinformatics approaches to select the protein(s) with the best properties for testing through *in vitro* and *in vivo* assays, in order to demonstrate its safety and immunogenicity. With the best vaccine candidates, different types of vaccines can be designed and developed, for example: subunit, recombinant, and nucleic acid vaccines [11].

The first application of reverse vaccinology was to study *Neisseria meningitidis* to obtain a new subunit vaccine based on the genome study of this microorganism by means of bioinformatics tools [12]. Thereafter, this technology has been used to study pathogenic agents including eukaryotic organisms and those involved in diseases transmitted by vectors [13], to design and obtain not only vaccines for humans but also for animals [5]. The majority of new vaccines against infectious diseases that have been developed with this technology are currently found in preclinical or clinical trial. However, it is important to mention that in some instances, the vaccine candidate obtained by this technology could fail as a good vaccine antigen, because it is identified based solely on computational probabilistic studies, and there are other factors that could interfere when this antigen is administered in a complete organism. In addition, vaccine candidates identified by this technology are restricted to proteins or lipoproteins, in that they are encoded in the genome. By reverse vaccinology, it is impossible to identify carbohydrate or lipid antigenic molecules [3, 14].

Some of the important properties to detect good vaccine candidates are described as follows:

2.1. Protein cellular localization

Proteins are localized in different parts of the cell: in the cytoplasm, the cell membrane, or they can be secreted out of the cell and become extracellular. Molecules localized on the cell membrane or extracellularly are better antigens because they are more exposed to host cells, specifically to those related to the immune system; thus, they have a greater probability of generating a protective response [15]. In addition to the software that can predict these characteristics, there are protein databases that generate information about protein subcellular localization, such as LOCATE, LocDB, and eSLDB.

2.2. Adhesin properties

In an infectious process, the first contact of the microorganism with the host cells is through adhesins. Molecules with adhesin properties are vaccine candidates [16]. The probability of identifying an adhesin is calculated based on the frequency of amino acids, dipeptides, or homopolymers present in the protein, and the physicochemical characteristics of each amino acid that constitutes a protein: acidic, basic, neutral, hydrophilic, or hydrophobic. There are programs that analyze all of these characteristics, comparing them with those of adhesins that have been previously proven experimentally [17].

2.3. Antigenicity

There are known sequences of antigens with good *in vivo* and *in vitro* immunologic inductions that are compared with each sequence of the proteome under study in order to search for similarities. In this case, it is probable that two proteins with similar sequences have comparable antigenic effects. Moreover, predictions of independent antigenicity alignment exist based on the physicochemical properties of amino acids [18].

2.4. Similarity

It is important to study the similarity between the sequences under study with molecules from the host that will receive the vaccine, as well as between the related etiological agents. Molecules with a high degree of similarity could generate two different effects: the first is undesirable because the antigen could cause autoimmune reactions; on the other hand, if the molecules are similar between other etiological agents, the vaccine could induce cross-protection [19]. In the case of a vaccine against cancer, it is important to select molecules present in cancer cells but absent in healthy cells. The similarity analysis can also be utilized to search for molecules with the same function, providing an idea of antigenicity and virulence [20]. It is important to predict these values because the main characteristic of a vaccine must be innocuous; in this way, if it is inferred that a protein can be antigenic but also toxic, the better course is not to use it.

2.5. Transmembrane helix

A transmembrane helix is a protein segment of 17–25 amino acids that conforms an α-helix structure that spans through the membrane cell. Most of the time, vaccine candidates are expressed in biological systems that are different from the original source; in that case, the three-dimensional (3D) structure of the protein could be changed or difficult to purify if it has a transmembrane helix, due to differences in membrane structure [21]. The low transmembrane helix number is a major characteristic for the selection of a vaccine candidate.

According to the etiology of the disease under study, protein cellular localization, adhesin properties, antigenicity, lack of homology with human proteins to avoid the induction of a potential autoimmune response, and low or null transmembrane helix structures are the main properties that should be identified. This can be addressed by utilizing several computer programs to analyze each of these properties and by bioinformatics tools for the screening and selection of vaccine candidates, according to their top feature values.

There are Websites and downloadable software that can be useful for a particular reverse vaccinology analysis, for example, NERVE, Vaxign, Jenner-predict server, and Vacceed. In some cases, the proteome-of-interest can be uploaded, and in others the organism in a specific database needed to be chosen; for this analysis, some characteristics about the agent and the host are required. In addition, there are databases with vaccine candidates already identified or with complete information about vaccines, for example VIOLIN and MycobacRV (**Table 1**).
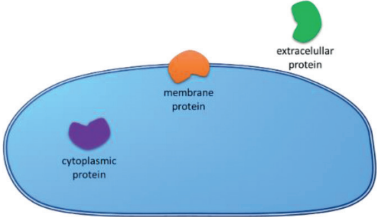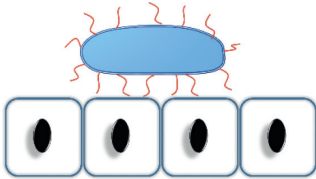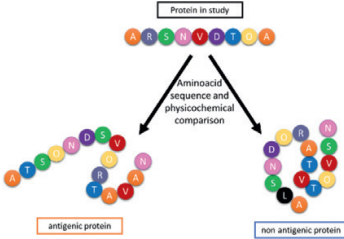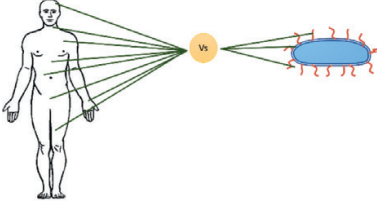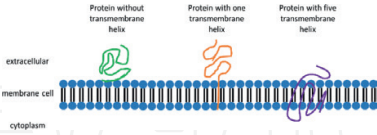
Characteristic	Description	Software
Protein cellular localization		• Psortb • CELLO • TargetP • Cell-PLoc • LocDB • LocTree 2/3 • MultiLoc2
Adhesin properties		• SPAAN • FungalRV • MAAP
Antigenicity		• VaxiJen • Protegen • EpiToolKit • SVMTriP
Similarity		• BLAST
Transmembrane helix		• TMHMM • TMpred • THGS • Sidekick • HMMTOP • SPLIT • DAS • Phobius • CCTOP • TMPad

Table 1. Main characteristics considered for vaccine candidate selection by reverse vaccinology.

3. Immunoinformatics

The immunological system can be classified as cellular or humoral and, depending on the disease, it can be induced the expected immune response. If a vaccine that induces a cellular response is needed, for example a tuberculosis vaccine [22] or a parasite vaccine against leishmaniasis [23], the software must search for antigens that can be recognized by the major histocompatibility complex (MHC) molecules present in T lymphocytes [4]. Software for this purpose include TEpredict, CTLPred, nHLAPred, ProPred-I, MAPPP, SVMHC, GPS-MBA, PREDIVAC, NetMHC, NetCTL, MHC2 Pred, IEDB, BIMAS, SVMHC, POPI, Epitopemap, iVAX, FRED2, Rankpep, BIMAS, PickPocket, KISS, and MHC2MIL. At their Websites, there are several options for search for MHC molecules as follows: for a specific species; type I or II, or even the allele(s) that will be employed for the prediction. The latter use different algorithms and some of these analyze the genome of the organism-under-study in order to identify new, probable MHC molecules.

On the other hand, if a humoral response is required, the software needs to identify antigens for B cells, for example, in the case of influenza virus or HIV [24, 25]. There is software that specifically searches for sequential epitopes for B cells, including BCPREDS, BepiPred, BEpro or PEPITO, ABCpred, Bcepred, IgPred, and BCEP. In addition, there are also Websites that, utilizing the 3D structure of a protein, can predict conformational epitopes for B cells, including the CEP, SEPPA, and DiscoTope Websites.

These software packages are based on computer training with the epitopes and nonepitopes previously identified, in order to provide values for new proteins and to predict whether or not it is an epitope. There are different techniques for this machine learning: position-specific scoring matrices (PSSMs), support vector machines (SVMs), hidden Markov models (HMMs), or artificial neural networks (ANNs). Each technique possesses different advantages and accuracy levels [26].

To achieve an analysis, the "immunome" of an organism is required; this includes all of the genes and proteins of cells that take part in its immune response. The study of all of the reactions that take part in the immune response is known as "immunomics" and it is specific for each organism; therefore, it is important to perform the study with information of the recipient organism. There have been many advances in the knowledge of immunomics using molecular biology and other throughput techniques, in order to understand the mechanisms of the immune system [27].

When immunomics and bioinformatics merged, a new science-denominated immunoinformatics was created, with the purpose of analyzing all of the information of an organism's immunomics and of making predictions of immune responses against specific molecules [28]. Websites already exist that present databases with antigens, with their epitopes identified in several organisms, and other immunological information, for example, IEDB, SIFPEITHI, IMGT, MHCBN, AntiJen, Dana-Farber Repository, and AgAbDb.

Once an antigen with the expected response has been identified, immunoinformatics can predict whether a region of an antigen, which usually is a protein, can generate a best stimulus by itself. If a protein has one epitope, this can be employed in a subunit vaccine and can be

combined with other epitopes of different organisms in order to generate a polyvalent vaccine, reducing the cost of the formulation. The epitopes can be synthesized artificially or obtained with molecular biology tools. This renders a vaccine safer, not only in its formulation but also in its production process, because there is no risk of the presence of infectious organisms [29].

With the purpose of determining epitopes, the proteins are analyzed to identify hydrophilic regions. The tertiary structure of a protein is based on the interactions between the amino acids and the medium, that is, the region with hydrophilic amino acids is exposed to the exterior. In the opposite case, the hydrophobic amino acids are located in the center of the structure. If this protein interacts with immune cells, it is more probable that contact will be generated with the hydrophilic region, a place localized in the epitope [28].

An additional step can be added, that is the prediction of the stability of peptide binding to MHC, because some epitopes can be attached with greater force and affinity, making activation of the immune system more probable. For this purpose, software has been created such as NetMHCStab, which utilizes artificial networks for the analysis [30].

In the case of cancer vaccines, antigens present in B cell have been developed that can help in the cancer cell elimination process. Additionally, antibodies against regulatory T-cells have been found with aid in the regression process of the tumor [9, 31]. The latter opens the way in the search for epitopes that could be used in vaccines, allowing better and faster elimination of the disease. For an allergy vaccine, other predictors, such as Allermatch and AlgPred, can be employed with the purpose of identifying proteins with potential allergenicity.

Other software developers have addressed the analysis of the complete immune response against specific antigens, such as C-ImmSim. In this case, the software uses different algorithms for each step; at the end, a series of graphic representations of each cell type can supply an idea of whether the response is sufficient to protect against a disease [32]. However, the general panorama is limited because this analysis implies the interaction of many cells and molecules and, in many cases, we do not yet know how these can interact with each other in a specific disease.

4. Structural vaccinology

Structural vaccinology focuses on the conformational features of macromolecules, mainly proteins that make them good candidate antigens. This approach to vaccine design has been used mainly to select or design peptide-based vaccines or cross-reactive antigens with the capability of generating immunity against different antigenically divergent pathogens. The initial stage in bioinformatics analyses involves linear epitope prediction, taking hydrophilicity as the major characteristic for locating epitopes. However, considering these predictions as the sole factor in determining the potential of a sequence to be immunogenic is risky. For example, the predicted epitopes could be sterically hindered by nearby amino acids, or if a peptide vaccine is being developed, the resulting peptide could adopt a conformation that differs from the peptide within the context of a whole protein, resulting in different conformational epitopes. In fact, available structures from nonoclonal antibodies (Mab) complexed to proteins have

demonstrated that, in the majority of cases, Mab recognize conformational rather than linear epitopes [33].

Many epitope-based vaccines attempt to elicit an antibody-mediated immune response that could neutralize the activity of toxins or pathogen receptors. Currently, there are many bioinformatics programs that predict protein epitopes. However, the majority of these programs rely only on the hydrophobicity or the hydrophilicity of amino acids. The main drawbacks in this are that many predicted epitopes are buried within the protein; thus, they would not be detected by the antibodies. In addition, the predicted epitopes are linear, leaving out conformational epitopes. In these cases, structural information can be helpful for selecting the epitopes that are exposed to the solvent and that are proximal to functional sites of the target protein, such as catalytic pockets or receptor binding pockets, or for detecting conformational epitopes on the surface of the target protein. Structural information is utilized to map antigenic epitopes to detect conformational features that could affect immunogenicity, such as the structural stability of proteins or the solvent exposure of candidate peptides, and to select antigenic regions shared by proteins of different pathogens that otherwise (i.e., by multiple alignments or epitope mapping) could not be evident. The approach that has been employed to develop vaccines is to perform several bioinformatics analyses at both at the sequence and structure level. For example, Cornick et al. [34] developed universal vaccine candidates against serotype 1 *Streptococcus pneumoniae* considering epitope prediction and structure modeling.

Protein flexibility can lead to vaccine failure due to high conformational variations that can avoid recognition by cell receptors or antibodies; for example, the failure of vaccines aimed at the HIV has been attributed to high flexibility of the globular head of gp120 [33, 35]. This is a concern, especially with peptides, which are usually more flexible and disordered than when they are found in a complete protein context. Bioinformatics predictions of flexibility can be attained from amino acid sequences (through structural alphabets) or from a 3D structure. High-performance bioinformatics tools such as molecular dynamics (MD) simulations can be employed to predict the stability of proteins or peptides [36]. This tool can be used to select the appropriate size of a peptide in order to render its stability and to introduce stabilizing mutations or chemical modifications that minimize flexibility, hence yielding better vaccine candidates than simple peptides.

Molecular docking is another bioinformatics tool that can be utilized in the selection and design of target antigens. It consists of complexing two molecules (protein-protein or protein-ligand) with best shape complementarity and minimal binding energy. In the field of structural vaccinology, molecular docking can be employed to predict the binding of epitopes to antibodies or to MHC receptors. Candidate antigens can be evaluated through the binding energy of the complex, and even mutations can be introduced to improve binding, but maintaining the specificity of the immune response [37].

Alam et al. [38], in a preliminary report, designed peptides as vaccine candidates against the Zika virus. They predicted MHC-I restricted epitopes, and then performed docking of these peptides with human leukocyte antigen (HLA) receptors to confirm their predictions. Toxicity analyses included allergenicity prediction. Another study proposed a multivalent vaccine

with fused peptides against *Staphylococcus aureus*. Again, epitope prediction was followed by peptide structure prediction, docking with TLR2, molecular dynamics simulations to assess the stability of the complexes, and finally, allergenicity prediction [39].

Care should be taken while designing peptide-based vaccines because the resulting peptide could be toxic or allergenic. Several bioinformatics studies perform toxicity or allergenicity prediction on peptide candidates to rule out adverse effects in the resulting candidate vaccine [38, 39].

Bioinformatics analyses have been performed to improve the functionality of antibodies. One study modified the Fc portion of antibodies to increase binding of proteins to the antibodies' Fc. This approach is relevant to improve the functionality of designed antibodies, to study immune response evasion by some pathogens, and in biotechnology to purify antibodies or proteins [37].

One premise of bioinformatics is to detect epitopes that can be recognized by antibodies, but modeling antibody-antigen complexes has been difficult because of the mobility of protein loops in the Fab region of antibodies [40]. One way to avoid this drawback is the strategy presented by Koivuniemi et al., which involved homology modeling to deduce the structure of the antigen and the antibody, docking, and molecular dynamics simulations [41] (**Figure 1**).
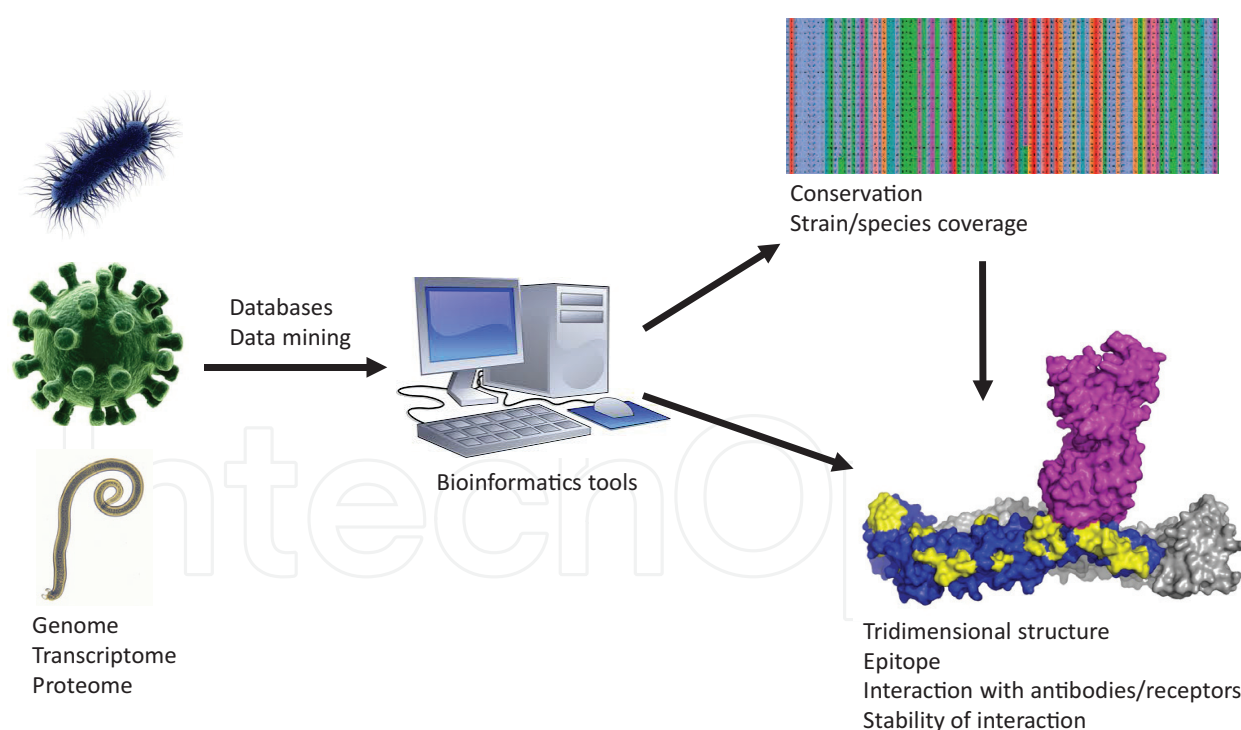
Figure 1. Path to antigen selection and validation. Databanks are created with experimental data from pathogens that can originate in the lab or be gathered through databases. Protein or nucleic acid sequences can be aligned to detect conservation and strain or species coverage. Three-dimensional (3D) structure information can be obtained from databases or inferred from bioinformatics analysis. Several predictions can be mapped into the structure, such as epitope prediction or amino acid conservation. Molecular docking tools can be used to establish interaction between two or more molecules (antibodies and cell receptors). Finally, the stability of these interactions can be assessed through energy calculations or molecular dynamics simulations.

5. Special cases: vaccines against infectious and noninfectious diseases

5.1. Vaccines against infectious diseases

5.1.1. Tuberculosis

Tuberculosis is an infectious disease caused by *Mycobacterium tuberculosis*, which is the most virulent and transmissible bacterium of the genus; however, it is a microorganism that is difficult to study because of its requirements and slow growth. The number of new cases worldwide rose to 10.4 million [42]; this high incidence rate is based on several factors, and one of the most important factors is the ineffectiveness of the vaccine used at present: the BCG. Thus, why many working groups are investigating new vaccines that can improve the level of protection against this disease, and one of the tools utilized is reverse vaccinology [10].

One strategy applied for vaccine design is to identify the structures present only in *M. tuberculosis* and absent in *Mycobacterium bovis* BCG [43]. In addition, the vaccine candidates studied presented the characteristics described previously, such as nonhuman homology, adhesins [44], secreted or membrane structures [45, 46] with low transmembrane helix, and in addition, the proteins expressed in the latent or active state of the microorganism [47]. The immunity sought is a protective response that is cellular. Therefore, immunoinformatics is based on the study of T-cell epitopes [22, 48–50].

Several candidates and epitopes have been found with different software. Some of these have been expressed and proven *in vitro* and *in vivo*, demonstrating their immunogenicity and protective effect. Among these are highlighted the ESAT-6, PE and PPE protein family group [51], and the Ag85 protein family, which obtained better immune response than the BCG vaccine in an animal model [43].

5.1.2. Influenza

The design of influenza vaccines is challenging due to the influenza virus's antigenic plasticity. Influenza viruses evade the immune response through antigenic drift and antigenic shift [52], rendering a long-lasting immune response very difficult. Current influenza vaccines contain hemagglutinin (HA) and neuraminidase (NA) as main antigenic components, usually having one type-B strain, and one H1 and one H3 subtype strain [53, 54]. Predicting the composition of next-year's vaccines relies on epidemiological data, although evolutionary models can aid in predicting antigenic drift, improving vaccine design [55].

Influenza HA recognizes cell receptors and mediates membrane fusion between the virus and the target cell. The globular head of HA contains the receptor binding site and the majority of the antigenic sites; consequently, this region is also the most variable. The stem region contains the fusion peptide and, although it previously was not considered a target for vaccine development, the discovery of neutralizing antibodies aimed at this region revealed its potential in vaccine design [52, 56]. Several conserved regions have been described in the stem region of HA [57], which make a universal vaccine a possibility. It has been found that neutralizing antibodies can bind to intact trimers, confirming the possibility of a universal

vaccine aimed at the HA stem. In fact, engineered HA stem antigens have been shown to elicit immune responses against heterosubtypic challenge models and serve as a proof-of-concept that these vaccines work [58].

Given the high cooperation, hence availability, of influenza viral protein sequences, there are open databases such as OpenFluDB [59] or the Influenza Research Database [60] that help in the designing of influenza vaccines. EpiCombFlu is a database that aids in defining conserved epitopes across influenza strains that can be combined to maximize strain coverage. Analysis of these sequences has led to the identification of conserved motifs among influenza strains that can be targets in vaccine or inhibitor design [61].

5.1.3. Chikungunya fever

For CHIKungunya Virus (CHIKV), there are some vaccine candidates in clinical trials, but there is no licensed vaccine to date. Efforts include the development of vaccines of inactivated virus, live attenuated virus (LAV), and virus-like particles (VLPs). In preclinical studies, LAV and VLP vaccines have been promising, but during clinical trials, they have shown inadequate immunogenicity and residual virulence, for example, the risk of production of chronic rheumatism seen for LAV [62]. However, vaccines should be able to induce high levels of neutralizing antibodies, ideally with only one dose, LAV remain good candidates for which attenuation strategies are of central importance.

Because the CHIKV E2 glycoprotein is thought to interact with cellular receptors and has demonstrated to elicit neutralizing antibodies, generating protection against lethal challenge in mice [63], it has been extensively studied. Kam et al. [64] mapped its epitope-containing sequences using experimentally infected macaque antibodies. Their results revealed that one of four recognized regions mapped onto the surface of E2, that the majority of the epitopes clustered in the middle of the protein, and that antibody recognition of E2 changes throughout the disease course in experimentally infected macaques may be due to the spatial positions of the B-cell epitopes on the native form of the E1/E2 glycoprotein complex. As part of the study, these authors included computational modeling utilizing the structural data of the E2 retrieved from PDB and visualizing the results using UCSF CHIMERA software.

In the design of an LAV for CHIKV, Gardner et al. [65] considered three known facts: that the substitution for positively charged residues in E2 that confer enhanced, Heparan sulfate (HS)-dependent infectivity *in vitro* is a common phenomenon among cell culture-passaged strains of some CHIKV-related viruses; that these mutations can be selected from within only a few serial passages *in vitro*, and that viruses whose *in vitro* infectivity is enhanced by artificial HS attachment/entry are typically attenuated/avirulent *in vivo*. In the case of CHIKV, an LAV candidate, attenuated by serial passages in MRC-5 fibroblasts, the authors predicted an amino acid substitution at E2 position 82, which was highly dependent upon ionic interaction with HS for infectivity. Afterward, this mutation demonstrated the attenuation two strains of CHIKV *in vivo*. Based on this fact [59], E2 mutations were selected that confer HS dependence on infectivity by serial passage of wild-type CHIKV-LR on different cell types *in vitro*. Then they introduced these mutations individually into CHIKV and identified a panel of E2 mutations that confer reduced virulence in a murine model. In this work, computational modeling

played an important role because it helped to explain the effect of the single amino acid mutations on altering the electrostatic profile of the E2 glycoprotein and increasing net positive charge in two exposed regions.

5.1.4. Zika virus disease

Zika virus, a positive single-stranded RNA virus transmitted by mosquito bites, is currently spreading worldwide and there is no available commercial vaccine. Several candidates are undergoing preclinical and clinical studies, and some platforms being investigated include inactivated, subunit/peptide, DNA-based, live-attenuated, and vectored vaccines. For a vaccine against this pathogen, multiple bioinformatics strategies are being exploited as an essential tool; the majority of studies involve *in silico* predictions to find the best epitopes. Dikhit et al. [66] found nine promiscuous highly conserved class I restricted epitopes among capsid 1, the envelope, and NS2A, NS4B, and NS5 viral proteins. Then, the tertiary structure of the selected epitopes was modeled using PEPstr and finally there was docking to HLA calculation with PatchDock.

Dar et al. [67] utilized ProPred1 to predict antigenic epitopes for HLA class I, as well as 48 antigenic epitopes for HLA class II employing ProPred immunoinformatics algorithms. These authors found 21% of MHC class I binding epitopes among NS5 viral proteins, followed by the envelope (17%). For MHC class II, NS5 contained 19% of predicted epitopes, and 17% were in the envelope, 17% in NS1, and 17% in NS2. Additionally, they obtained the antigenicity score for each predicted epitope using the VaxiJen 2.0 tool. Ashfaq and Ahmed are other researchers who used ProPred1 and ProPred, but focused in the envelope protein, finding two highly antigenic candidates among T-cell epitopes. They also performed a molecular docking to study the interactions of B-cell epitopes with HLA-B7 [68].

Another bioinformatics-based study is that of Mirza et al. [69], in which the authors predicted antigenic B-cell (IEDB) and CTL epitopes (NetCTL.1.2 server). They determined, by *in silico* studies, surface accessibility, surface flexibility, hydrophilicity, homology modeling (MODELLER ver. 9.12, CHARMM, WhatIF, PROCHECK, Verify 3D), and structure-based epitope prediction for E protein, NS3, and NS5. They performed molecular docking of the ZIKV-E protein with HLA-A0201, of the ZIKV-NS3 protein with HLA-B2705, and of the ZIKV-NS5 protein with HLA-C0801 (PatchDock rigid-body docking server, FireDock server). Finally, these authors investigated the stability of the docked peptide-MHC I protein complexes by performing Molecular Dynamics (MD) simulations (AMBER 12 simulation package) [69].

An important aspect in the design of a vaccine is the study of the virus's molecular biology, its proteome, and the genotypes. Sun et al. reported such data, to our knowledge for the first time, using new computational methods for annotation of mature peptide proteins, genotypes, and recombination events for all ZIKV genomes [70]. In an effort to aid in the development of vaccines and therapeutic drugs, an integrative multi-omics platform, ZikaVR (http://bioinfo.imtech.res.in/manojk/zikavr/) was created by Gupta el at.. This platform contains genomic, proteomic, and therapeutic information about the Zika virus [71].

5.2. Vaccines against noninfectious diseases

5.2.1. Vaccines to treat addictions

In the search for a vaccine to fight drug abuse, cocaine, nicotine, and methamphetamines are some of the main targets; however, to date there are, to our knowledge, no US: Federal Drug Administration (FDA)-approved vaccines. The development of such products has been hindered by the need of a carrier protein and an adjuvant to combine with haptens of the drugs to elicit the necessary antibody levels expected to interfere with the transport of the drug to the Central Nervous System (CNS), thus with the expected effect [72].

Kimishima et al. have explored tetanus toxoid (TT), the bacterial flagellin FliC, alum, and CpG (cytosine-phosphate-guanine oligodeoxynucleotide) in the development of an anticocaine vaccine. TT is used as a carrier; FliC acts as a carrier protein, and additionally it has been demonstrated that it stimulates toll-like receptor 5 (TLR5), therefore inducing myeloid differentiation factor 88 (MyD88), which renders a TH2 response to predominant production of IgG1 and no cytotoxic T lymphocytes (CTL). CpG (a B-class OligoDeoxyNucleotide [ODN]) motifs can be used as activators of TLR9 to promote a TH1-type immune response, stimulating B-cell immune responses to generate IgG2a and CTL [73].

Lockner et al., in a first attempt, conjugated GNE (a cocaine hapten) with a recombinant FliC, utilized *in silico* modeling and computational analysis of the recombinant protein to ensure its structural integrity and conservation of the binding to TLR5; by Modeler, they studied the homology of the recombinant flagellin, as well as the number of lysines per domain and relative solvent accessibility with and without GNE cocaine haptens present. Their computational results agreed with those used for experimentation since then in a TLR5 reporter assay: the modified flagellin protein still activated TLR5 when the hapten density was <10 GNE per FliC. Finally, the authors showed that cocaine-flagellin conjugates induced, in a dose-dependent model, the production of anticocaine antibodies in mice, improving the response with the adjuvant alum [73, 74].

On the other hand, as they observed in prior experiments in which they conjugated GNE (a cocaine hapten) with FliC, TLR5 activation was attenuated at higher hapten densities (i.e., above ~10 GNE per flagellin). Consequently, they induced a mutation in the flagellin gene (*mFliC*), which could protect the TLR5 binding interface against covalent modification with the bulky GNE hapten, thus potentially preserving the ability of the modified flagellin to activate TLR5 independently of hapten densities. *mFliC* consisted of a mutation of the 10 lysine residues within the D0 and D1 domains of wild-type *FliC* (as well as one additional lysine residue previously introduced through cloning) to arginine residues [73]. Again, bioinformatics was necessary to assess the secondary structure and MHC-II binding predictions for FliC and mFliC, employing the PSIPRED (http://bioinf. cs.ucl.ac.uk/psipred/) method and the external software from IEDB (http://www.immuno-epitope.org/), respectively [74].

The computational results for MHC-II binding and hapten presentation revealed that the FliC conjugate was better than mFliC; these results indirectly correlated with those conducted by enzyme-linked immunosorbent assays (ELISA) and radioimmunoassays (RIA). However,

because FliC and mFliC exhibited poor efficacy as carrier proteins when comparing two for-mulations, GNE-FliC + CpG and GNE-TT + CpG, through a hyperlocomotion test and analy-sis of cocaine in blood, where GNE-TT + CpG had best efficacy, the authors proposed the investigation of monomers of FliC instead of the polymeric form utilized [74].

5.2.1.1. Allergies

Allergies comprise another area where vaccine (specific immunotherapy (SIT)) investigation is conferred due to the association of allergy with asthma and anaphylaxis. Some common allergies are caused by cat, peanut, and cockroach allergens, with the specific immunother-apy (SIT) effective, but sometimes associated with IgE-dependent adverse events. In allergies, computational approaches have been applied to find T-cell epitopes to target allergen-specific T cells, thus improving the safety of the immunotherapy.

In 2011, Worm et al. performed a clinical study administering the ToleroMune cat vaccine (short synthetic peptide sequences from the major cat allergen Fel d 1) to 66 subjects with cat allergy. The authors identified each peptide-MHC interaction by using physical binding assays and analyzed these *in silico* with the immune epitope database (www.immuneepitope.org/); *in vitro*, the individual peptides and the vaccine were at least 1000-fold less able to induce basophil histamine release associated with adverse effects than the native allergen. The vaccine administered intradermally (i.d.) or subcutaneously (s.c.) showed no serious adverse events (SAEs) during the study and no subject withdrew from the latter due to an adverse event. Thus, the vaccine was safe and well tolerated [75].

Another example of research to improve safety comprises the work of Pascal et al. for the treatment of peanut allergy, which presents symptoms ranging from mild oropharyngeal pruritus to life-threatening anaphylaxis, considerably compromising the patient's quality of life. Ara h 1, Ara h 2, and Ara h 3 include the three major peanut allergens, although IgE antibodies to Ara h 2 correlate most closely with clinical reactivity, and *in vitro* Ara h 2 and its homologue, Ara h 6, are more potent inducers of basophil degranulation than Ara h 1 and Ara h 3. Because conventional s.c. immunotherapy with crude peanut extract entertains a high risk of anaphylaxis and since peptides have been successful in the desensitization of patients to cat-allergy and bee venom-allergy, an alternative is the use of peptide fragments that retain immunogenicity, but that are of insufficient length to cross-link allergen-specific IgE on mast cells and basophils. In addition to proliferation assays utilizing peripheral blood mononuclear cells (PBMCs) from peanut-allergic children and Ara h 2 peptides, Pascal and colleagues predicted, to our knowledge for the first-time, epitopes in a food-allergy through the artificial neural network-based alignment (NN-align) method NetMHCIIpan-2.0. Their objective was to analyze additional theoretical peptides that are not included in the prolif-eration assays, finding that both strategies, *in vitro* and *in silico,* rendered consistent results; therefore, they were able to select peptide candidates for the development of a peanut allergy vaccine [76].

Regarding allergy to cockroaches, there are some research studies that have followed the *in silico* prediction of B-cell, T-cell, and IgE-binding epitopes in a first stage to propose a vaccine

formulation. Chen et al., Yang et al., and Tong et al. are members of a workgroup that studied this allergy by means of *in vitro* and *in silico* approaches. The allergens analyzed were Per a 6 and Bla g (found in *Periplaneta americana* and *Blattella germanica*, respectively) [77–79].

Chen et al. employed three immunoinformatics tools: the Protean™ system (DNAStar, Inc., Madison, WI, USA); the bioinformatics predicted antigenic peptides (BPAP) system (http://imed.med.ucm.es/Tools/antigenic.pl), and the BepiPred 1.0 server (http://www.cbs.dtu.dk/services/BepiPred/), which utilizes four properties, including hydrophilicity, flexibility, accessibility, and antigenicity as parameters for the prediction of B-cell epitopes. After a consensus of the three bioinformatics tools, these authors selected the final potential epitope regions (regions whose consensus epitope result was 67 or 100%) to develop a vaccine. Additionally, through the NN-align method NetMHCIIpan-2.0 (http://www.cbs.dtu.dk/services/NetMHCIIpan/) for HLA-DR alleles and NetMHCII-2.2 (http://www.cbs.dtu.dk/services/NetMHCII/) for HLA-DQ alleles, they found strong and weak binders [77]. In 2016, Yang et al. and Tong et al. predicted, using the same strategy, B- and T-cell peptides belonging to Per a 9 and Per a 10 (two major allergens as assessed by enzyme-linked immunosorbent assays (ELISA) but, in order to obtain substantial quantities of these allergens for use in functional studies, they cloned and expressed them in an *Escherichia coli* system [78, 79]

5.2.1.2. Cancer

Since T cells educated in the thymus do not recognize mutated antigens expressed in cancer cells, there is no negative selection, and these neoantigens are ideal targets for therapeutic vaccination; furthermore, they are not present in healthy tissue. On the other hand, advances in next-generation sequencing (NGS) permit the sequencing of genomes, exomes, or transcriptomes within hours. Therefore, they investigated the mutanome (the tens-to-hundreds of somatic nonsynonymous mutations) in order to select the specific targets for the recognition by cytotoxic and helper T cells with antitumor activity. The complexity of some experimental tools such as mass spectrometry hampers its usefulness in the selection of targets in a clinical setting where personalized therapy is needed. In this context, because it is not possible to analyze all of the mutations, bioinformatics addresses this problem and has become important in the selection of targets and in their prioritization [80].

An example of the success of *in silico* predicted mutations is the study of Castle et al., where the authors, applying thresholds for MHC class II binding prediction and mRNA expression levels, without further validation by immunogenicity testing, were able to enrich immunogenic MHC class II-restricted epitopes. They obtained efficient and sustained control of advanced tumors in mice [81].

Although there are successful *in vitro* and preclinical studies that initiated by utilizing computational approaches, the majority of algorithms predict the affinity of peptide binding to MHC molecules, which may not correlate well with their immunogenicity or may not predict peptides that are not generated and presented. Moreover, some immunogenic ligands may escape detection. Additionally, in general *in silico* prediction of ligands for MHC II is less

accurate than for MHC molecules. Because the immunogenicity of predicted peptides has been reported to correlate better with peptide-MHC complex stability, the use has been proposed of biochemical methods to reduce the number of *in silico* predicted MHC ligands and to generate data that helps in the training of prediction algorithms to validate peptide binding predictions. Some biochemical methods include peptide rebinding (referred to as iTopia), peptide-rescuing, and refolding for MHC I peptide binding validation, and peptide-driven refolding for MHC II [82].

Another approach to circumvent the limitations of the binding prediction for MHC molecules is molecular docking, a structure-based method that has been tested on both peptide-MHC class I and II complexes. This method can be applied to previously predicted peptides and is expected to improve prediction accuracy in order to identify the best MHC class I and II binders. Following this strategy, in a research for vaccine candidates against breast cancer, predicted discontinuous B-cell epitope peptides using PEPOP for the first time, then the 3D structure of epitope-based peptides by PEP-FOLD server, and their theoretical physicochemical properties utilizing the Prot Param algorithm, and finally, with.pdb files of two class I and seven class II MHC-peptide complexes from the protein data bank, perform molecular docking through the genetic optimization for ligand docking (GOLD) 5.4. After virtual screening, they confirmed a predicted peptide agreement between their docked results and previous experimental results (i.e., the immunogenicity of this peptide was confirmed *in vivo* studies), thus proposing molecular docking as an additional technique to improve the selection of peptide candidates for cancer vaccines [83].

Acknowledgements

Author details

Ribas-Aparicio Rosa María[1]*, Castelán-Vega Juan Arturo[1], Jiménez-Alberto Alicia[1], Monterrubio-López Gloria Paulina[1] and Aparicio-Ozores Gerardo[2]

*Address all correspondence to: rribas233@yahoo.com

1 Laboratory of Production and Biological Control, Department of Microbiology, National School of Biological Sciences, National Polytechnic Institute, Mexico City, Mexico

2 Laboratory of Medical Bacteriology, Department of Microbiology, National School of Biological Sciences, National Polytechnic Institute, Mexico City, Mexico

References

[1]   Plotkin ST, Plotkin SA. A short history of vaccination. In: Plotkin SA, Orenstein WA, Offit PA, editors. Vaccines. 6th ed. Philadelphia, PA, USA: Elsevier/Saunders; 2013. pp. 1-13

[2]   Bulletin of the World Health Organization (WHO). Available from: http://www.who.int/bulletin/volumes/86/2/07-040089/en/

[3]   Rappuoli R. Reverse vaccinology. Current Opinion in Microbiology. 2000;**3**(5):445-450

[4]   He Y, Rappuoli R, De Groot AS, Chen RT. Emerging vaccine informatics. Journal of Biomedicine and Biotechnology. 2010;**2010**:218590. DOI: 10.1155/2010/218590

[5]   Welly BT, Miller MR, Stott JL, Blanchard MT, Islas-Trejo AD, O'Rourke SM, Young AE, Medrano JF, Van Eenennaam AL. Genome report: Identification and validation of antigenic proteins from *Pajaroellobacter abortibovis* using *de novo* genome sequence assembly and reverse vaccinology. G3 (Bethesda). 2017;**7**(2):321-331. DOI: 10.1534/g3.116.036673

[6]   Plotkin S. History of vaccination. Proceedings of the National Academy of Sciences of the United States. 2014;**111**(34):12283-12287. DOI: 10.1073/pnas.1400472111

[7]   Seib KL, Zhao X, Rappuoli R. Developing vaccines in the era of genomics: A decade of reverse vaccinology. Clinical Microbiology and Infection. 2012;**18**(Suppl 5):109-116. DOI: 10.1111/j.1469-0691.2012.03939.x

[8]   Vivona S, Bernante F, Filippini F. NERVE: New enhanced reverse vaccinology environment. BMC Biotechnology. 2006;**6**:35

[9]   Delany I, Rappuoli R, De Gregorio E. Vaccines for the 21st century. EMBO Molecular Medicine. 2014;**6**(6):708-720. DOI: 10.1002/emmm.201403876

[10] Monterrubio-López GP, González-Y-Merchand JA, Ribas-Aparicio RM. Identification of novel potential vaccine candidates against tuberculosis Based on reverse vaccinology. BioMed Research International. 2015;**2015**:483150. DOI: 10.1155/2015/483150

[11] Jaiswal V, Chanumolu SK, Gupta A, Chauhan RS, Rout C. Jenner-predict server: Prediction of protein vaccine candidates (PVCs) in bacteria based on host-pathogen interactions. BMC Bioinformatics. 2013;**14**:211. DOI: 10.1186/1471-2105-14-211

[12] Pizza M, Scarlato V, Masignani V, Giuliani MM, Aricò B, Comanducci M, Jennings GT, Baldi L, Bartolini E, Capecchi B, Galeotti CL, Luzzi E, Manetti R, Marchetti E, Mora M, Nuti S, Ratti G, Santini L, Savino S, Scarselli M, Storni E, Zuo P, Broeker M, Hundt E, Knapp B, Blair E, Mason T, Tettelin H, Hood DW, Jeffries AC, Saunders NJ, Granoff DM, Venter JC, Moxon ER, Grandi G, Rappuoli R. Identification of vaccine candidates against serogroup B meningococcus by whole-genome sequencing. Science. 2000;**287**(5459):1816-1820

[13] Lew-Tabor AE, Rodríguez-Valle M. A review of reverse vaccinology approaches for the development of vaccines against ticks and tick borne diseases. Ticks and Tick-borne Diseases. 2016;**7**(4):573-585. DOI: 10.1016/j.ttbdis.2015.12.012

[14] Goodswen SJ, Kennedy PJ, Ellis JT. A guide to *in silico* vaccine discovery for eukaryotic pathogens. Briefings in Bioinformatics. 2013;**14**(6):753-774. DOI: 10.1093/bib/bbs066

[15] Chaudhuri R, Kulshreshtha D, Raghunandanan MV, Ramachandran S. Integrative immunoinformatics for Mycobacterial diseases in R platform. Systems and Synthetic Biology. 2014;**8**(1):27-39. DOI: 10.1007/s11693-014-9135-9

[16] Chaudhuri R, Ramachandran S. Prediction of virulence factors using bioinformatics approaches. Methods in Molecular Biology. 2014;**1184**:389-400. DOI: 10.1007/978-1-4939-1115-8_22

[17] Sachdeva G, Kumar K, Jain P, Ramachandran S. SPAAN: A software program for prediction of adhesins and adhesin-like proteins using neural networks. Bioinformatics. 2005;**21**(4):483-491

[18] Doytchinova IA, Flower DR. VaxiJen: A server for prediction of protective antigens, tumour antigens and subunit vaccines. BMC Bioinformatics. 2007;**8**:4

[19] Argondizzo AP, da Mota FF, Pestana CP, Reis JN, de Miranda AB, Galler R, Medeiros MA. Identification of proteins in *Streptococcus pneumoniae* by reverse vaccinology and genetic diversity of these proteins in clinical isolates. Applied Biochemistry and Biotechnology. 2015;**175**(4):2124-2165. DOI: 10.1007/s12010-014-1375-3

[20] Altindis E, Cozzi R, Di Palo B, Necchi F, Mishra RP, Fontana MR, Soriani M, Bagnoli F, Maione D, Grandi G, Liberatori S. Protectome analysis: A new selective bioinformatics tool for bacterial vaccine candidate discovery. Molecular & Cellular Proteomics. 2015;**14**(2):418-429. DOI: 10.1074/mcp.M114.039362

[21] Meunier M, Guyard-Nicodème M, Hirchaud E, Parra A, Chemaly M, Dory D. Identification of novel vaccine candidates against *Campylobacter* through reverse vaccinology. Journal of Immunology Research. 2016;**2016**:5715790. DOI: 10.1155/2016/5715790

[22] Pandey K, Sharma M, Saarav I, Singh S, Dutta P, Bhardwaj A, Sharma S. Analysis of the DosR regulon genes to select cytotoxic T lymphocyte epitope specific vaccine candidates using a reverse vaccinology approach. International Journal of Mycobacteriology. 2016;**5**(1):34-43. DOI: 10.1016/j.ijmyco.2015.10.005

[23] Seyed N, Taheri T, Rafati S. Post-genomics and vaccine improvement for *Leishmania*. Frontiers in Microbiology. 2016;**7**:467. DOI: 10.3389/fmicb.2016.00467

[24] Rappuoli R, Bottomley MJ, D'Oro U, Finco O, De Gregorio E. Reverse vaccinology 2.0: Human immunology instructs vaccine antigen design. Journal of Experimental Medicine. 2016;**213**(4):469-481. DOI: 10.1084/jem.20151960

[25] Van Regenmortel MH. An outdated notion of antibody specificity is one of the major detrimental assumptions of the structure-based reverse vaccinology paradigm, which prevented it from helping to develop an effective HIV-1 vaccine. Frontiers in Immunology. 2014;**5**:593. DOI: 10.3389/fimmu.2014.00593

[26] Backert L, Kohlbacher O. Immunoinformatics and epitope prediction in the age of genomic medicine. Genome Medicine. 2015;**7**:119. DOI: 10.1186/s13073-015-0245-0

[27] Tomar N, De RK. Immunoinformatics: A brief review. Methods in Molecular Biology. 2014;**1184**:23-55. DOI: 10.1007/978-1-4939-1115-8_3

[28] Flower DR. Vaccines: Data driven prediction of binders, epitope and immunogenicity. In: Bioinformatics for Vaccinology. United Kingdom: Wiley-Blackwell; 2008. pp. 167-215

[29] Delany I, Rappuoli R, Seib KL. Vaccines, reverse vaccinology, and bacterial pathogenesis. Cold Spring Harbor Perspectives in Medicine. 2013;**3**(5):a012476. DOI: 10.1101/cshperspect.a012476

[30] Jørgensen KW, Rasmussen M, Buus S, Nielsen M. NetMHCstab- predicting stability of peptide-MHC-I complexes; impacts for cytotoxic T lymphocyte epitope discovery. Immunology. 2014;**141**(1):18-26. DOI: 10.1111/imm.12160

[31] Rappuoli R, Pizza M, Del Giudice G, De Gregorio E. Vaccines, new opportunities for a new society. Proceedings of the National Academy of Sciences of the United States. 2014;**111**(34):12288-12293. DOI: 10.1073/pnas.1402981111

[32] Rapin N, Lund O, Bernaschi M, Castiglione F. Computational immunology meets bioinformatics: The use of prediction tools for molecular binding in the simulation of the immune system. PLOS ONE. 2010;**5**(4):e9862. DOI: 10.1371/journal.pone.0009862

[33] Van Regenmortel MH. Structure-based reverse vaccinology failed in the case of HIV because it disregarded accepted immunological theory. International Journal of Molecular Sciences. 2016;**17**(9):1591

[34] Cornick JE, Tastan Bishop Ö, Yalcin F, Kiran AM, Kumwenda B, Chaguza C, Govindpershad S, Ousmane S, Senghore M, du Plessis M, Pluschke G, Ebruke C, McGee L, Sigaùque B, Collard JM, Bentley SD, Kadioglu A, Antonio M, von Gottberg A, French N, Klugman KP, Heyderman RS, Alderson M, Everett DB; PAGe Consortium. The global distribution and diversity of protein vaccine candidate antigens in the highly virulent *Streptococcus pneumoniae* serotype 1. Vaccine. 2017;**35**(6):972-980. DOI: 10.1016/j.vaccine.2016.12.037

[35] Korkut A, Hendrickson WA. Structural plasticity and conformational transitions of HIV envelope glycoprotein gp120. PLOS ONE. 2012;**7**(12):e52170. DOI: 10.1371/journal.pone.0052170

[36] Zhu Z, Zhang C, Song W. Rational derivation, extension, and cyclization of self-inhibitory peptides to target TGF-β/BMP signaling in ONFH. Amino Acids. 2017;**49**(2):283-290. DOI: 10.1007/s00726-016-2358-1

[37] Agostino M, Mancera RL, Ramsland PA, Fernández-Recio J. Optimization of protein-protein docking for predicting Fc-protein interactions. Journal of Molecular Recognition. 2016;**29**(11):555-568. DOI: 10.1002/jmr.2555

[38] Alam A, Ali S, Ahamad S, Malik MZ, Ishrat R. From ZikV genome to vaccine: *In silico* approach for the epitope-based peptide vaccine against Zika virus envelope glycoprotein. Immunology. 2016;**149**(4):386-399. DOI: 10.1111/imm.12656

[39] Hajighahramani N, Nezafat N, Eslami M, Negahdaripour M, Rahmatabadi SS, Ghasemi Y. Immunoinformatics analysis and *in silico* designing of a novel multi-epitope peptide vaccine against *Staphylococcus aureus*. Infection, Genetics and Evolution. 2016;**48**:83-94. DOI: 10.1016/j.meegid.2016.12.010

[40] Bohnuud T, Luo L, Wodak SJ, Bonvin AM, Weng Z, Vajda S, Schueler-Furman O, Kozakov D. A benchmark testing ground for integrating homology modeling and protein docking. Proteins. 2017;**85**(1):10-16. DOI: 10.1002/prot.25063

[41] Koivuniemi A, Takkinen K, Nevanen T. A computational approach for studying antibody-antigen interactions without prior structural information: The anti-testosterone binding antibody as a case study. Proteins. 2017;**85**(2):322-331. DOI: 10.1002/prot.25226

[42] World Health Organization. Global Tuberculosis Report; 2016

[43] Mollenkopf HJ, Grode L, Mattow J, Stein M, Mann P, Knapp B, Ulmer J, Kaufmann SH. Application of mycobacterial proteomics to vaccine design: Improved protection by *Mycobacterium bovis* BCG prime-Rv3407 DNA boost vaccination against tuberculosis. Infection and Immunity. 2004;**72**(11):6471-6479

[44] Kumar S, Puniya BL, Parween S, Nahar P, Ramachandran S. Identification of novel adhesins of *M. tuberculosis* H37Rv using integrated approach of multiple computational algorithms and experimental analysis. PLOS ONE. 2013;**8**(7):e69790. DOI: 10.1371/journal.pone.0069790

[45] Vizcaíno C, Restrepo-Montoya D, Rodríguez D, Niño LF, Ocampo M, Vanegas M, Reguero MT, Martínez NL, Patarroyo ME, Patarroyo MA. Computational prediction and experimental assessment of secreted/surface proteins from *Mycobacterium tuberculosis* H37Rv. PLOS Computational Biology. 2010;**6**(6):e1000824. DOI: 10.1371/journal.pcbi.1000824

[46] Song H, Sandie R, Wang Y, Andrade-Navarro MA, Niederweis M. Identification of outer membrane proteins of *Mycobacterium tuberculosis*. Tuberculosis (Edinburgh). 2008;**88**(6):526-544. DOI: 10.1016/j.tube.2008.02.004

[47] Nguyen Thi le T, Sarmiento ME, Calero R, Camacho F, Reyes F, Hossain MM, González GS, Norazmi MN, Acosta A. Immunoinformatics study on highly expressed *Mycobacterium tuberculosis* genes during infection. Tuberculosis (Edinburgh). 2014;**94**(5):475-481

[48] Sundaramurthi JC, Brindha S, Shobitha SR, Swathi A, Ramanandan P, Hanna LE. *In silico* identification of potential antigenic proteins and promiscuous CTL epitopes in *Mycobacterium tuberculosis*. Infection, Genetics and Evolution. 2012;**12**(6):1312-1318. DOI: 10.1016/j.meegid.2012.03.023

[49] Vani J, Shaila MS, Chandra NR, Nayak R. A combined immuno-informatics and structure-based modeling approach for prediction of T cell epitopes of secretory proteins of Mycobacterium tuberculosis. Microbes and Infection. 2006;**8**(3):738-746

[50] Tang ST, van Meijgaarden KE, Caccamo N, Guggino G, Klein MR, van Weeren P, Kazi F, Stryhn A, Zaigler A, Sahin U, Buus S, Dieli F, Lund O, Ottenhoff TH. Genome-based *in silico* identification of new *Mycobacterium tuberculosis* antigens activating polyfunctional CD8+ T cells in human tuberculosis. Journal of Immunology. 2011;**186**(2):1068-1080. DOI: 10.4049/jimmunol.1002212

[51] Pajón R, Yero D, Lage A, Llanes A, Borroto CJ. Computational identification of beta-barrel outer-membrane proteins in *Mycobacterium tuberculosis* predicted proteomes as putative vaccine candidates. Tuberculosis (Edinburgh). 2006;**86**(3-4):290-302

[52] Bouvier NM, Palese P. The biology of influenza viruses. Vaccine. 2008;**26**(Suppl 4):D49-D53

[53] Lambach P, Álvarez AM, Hirve S, Ortiz JR, Hombach J, Verweij M, Hendriks J, Palkonyay L, Pfleiderer M. Considerations of strategies to provide influenza vaccine year round. Vaccine. 2015;**33**(47):6493-6498. DOI: 10.1016/j.vaccine.2015.08.037

[54] Alan H, Ian B, Nancy C, Ruben OD, Siddhivinayak H, Daniel J, Jacqueline K, John M, Fernando M, Takato O, Tam JS, Anthony W, Richard W, Thedi Z, Wenqing Z. Improving the selection and development of influenza vaccine viruses—Report of a WHO informal consultation on improving influenza vaccine virus selection, Hong Kong SAR, China, 18-20 November 2015. Vaccine. 2017;**35**(8):1104-1109. DOI: 10.1016/j.vaccine.2017.01.018

[55] Xu H, Yang Y, Wang S, Zhu R, Qiu T, Qiu J, Zhang Q, Jin L, He Y, Tang K, Cao Z. Predicting the mutating distribution at antigenic sites of the influenza virus. Scientific Reports. 2016;**6**:20239. DOI: 10.1038/srep20239

[56] Neu KE, Henry Dunand CJ, Wilson PC. Heads, stalks and everything else: How can antibodies eradicate influenza as a human disease? Current Opinion in Immunology. 2016;**42**:48-55. DOI: 10.1016/j.coi.2016.05.012

[57] Harris AK, Meyerson JR, Matsuoka Y, Kuybeda O, Moran A, Bliss D, Das SR, Yewdell JW, Sapiro G, Subbarao K, Subramaniam S. Structure and accessibility of HA trimers on intact 2009 H1N1 pandemic influenza virus to stem region-specific neutralizing antibodies. Proceedings of the National Academy of Sciences of the United States. 2013;**110**(12):4592-4597. DOI: 10.1073/pnas.1214913110

[58] Impagliazzo A, Milder F, Kuipers H, Wagner MV, Zhu X, Hoffman RM, van Meersbergen R, Huizingh J, Wanningen P, Verspuij J, de Man M, Ding Z, Apetri A, Kükrer B, Sneekes-Vriese E, Tomkiewicz D, Laursen NS, Lee PS, Zakrzewska A, Dekking L, Tolboom J, Tettero L, van Meerten S, Yu W, Koudstaal W, Goudsmit J, Ward AB, Meijberg W, Wilson IA, Radošević K. A stable trimeric influenza hemagglutinin stem as a broadly protective immunogen. Science. 2015;**349**(6254):1301-1306. DOI: 10.1126/science.aac7263

[59] Liechti R, Gleizes A, Kuznetsov D, Bougueleret L, Le Mercier P, Bairoch A, Xenarios I. OpenFluDB, a database for human and animal influenza virus. Database (Oxford). 2010;**2010**:baq004. DOI: 10.1093/database/baq004

[60] Zhang Y, Aevermann BD, Anderson TK, Burke DF, Dauphin G, Gu Z, He S, Kumar S, Larsen CN, Lee AJ, Li X, Macken C, Mahaffey C, Pickett BE, Reardon B, Smith T, Stewart L, Suloway C, Sun G, Tong L, Vincent AL, Walters B, Zaremba S, Zhao H, Zhou L, Zmasek C, Klem EB, Scheuermann RH. Influenza research database: An integrated bioinformatics resource for influenza virus research. Nucleic Acids Research. 2017;**45**(D1):D466-D474. DOI: 10.1093/nar/gkw857

[61] ElHefnawi M, Alaidi O, Mohamed N, Kamar M, El-Azab I, Zada S, Siam R. Identification of novel conserved functional motifs across most Influenza A viral strains. Virology Journal. 2011;**8**:44. DOI: 10.1186/1743-422X-8-44

[62] Singh P, Chhabra M, Mittal V, Sharma P, Rizvi MA, Chauhan LS, Rai A. Current research and clinical trials for a vaccine against chikungunya virus. Vaccine: Development and Therapy. 2013;**3**:35-46. DOI: https://doi.org/10.2147/VDT.S25513

[63] van den Doel P, Volz A, Roose JM, Sewbalaksing VD, Pijlman GP, Middelkoop I, Duiverman V, Wetering E, Sutter G, Osterhaus ADME, Martina BEE. Recombinant modified vaccinia virus ankara expressing glycoprotein E2 of chikungunya virus protects AG129 mice against lethal challenge. PLOS Neglected Tropical Diseases. 2014;**8**(9):e3101. DOI: 10.1371/journal.pntd.0003101

[64] Kam Y-W, Lee WWL, Simarmata D, Le Grand R, Tolou H, Merits A, Roques P, Ng LFP. Unique epitopes recognized by antibodies induced in chikungunya virus-infected non-human primates: Implications for the study of immunopathology and vaccine development. PLOS ONE. 2014;**9**(4):e95647. DOI: 10.1371/journal.pone.0095647

[65] Gardner CL, Hritz J, Sun C, Vanlandingham DL, Song TY, Ghedin E, Higgs S, Klimstra WB, Ryman KD. Deliberate attenuation of chikungunya virus by adaptation to Heparan sulfate-dependent infectivity: A model for rational arboviral vaccine design. PLOS Neglected Tropical Diseases. 2014;**8**(2):e2719. DOI: 10.1371/journal.pntd.0002719

[66] Dikhit MR, Ansari MY, Vijaymahantesh, Kalyani, Mansuri R, Sahoo BR, Dehury B, Amit A, Topnof RK, Sahoo GC, Ali V, Bimal S, Das P. Computational prediction and analysis of potential antigenic CTL epitopes in Zika virus: A first step towards vaccine development. Infection, Genetics and Evolution. 2016;**45**:187-197. DOI: org/10.1016/j.meegid.2016.08.037 DOI: 10.1016/j.meegid.2016.08.037

[67] Dar H, Zaheer T, Rehman MT, Ali A, Javed A, Khan GA, Babar MM, Waheed Y. Prediction of promiscuous T-cell epitopes in the Zika virus polyprotein: An *in silico* approach. Asian Pacific Journal of Tropical Disease. 2016;**9**(9):844-850. DOI: org/10.1016/j.apjtm.2016.07.004

[68] Ashfaq UA, Ahmed B. *De novo* structural modeling and conserved epitopes prediction of Zika virus envelope protein for vaccine development. Viral Immunology. 2016;**29**(7):436-443. DOI: 10.1089/vim.2016.0033

[69] Mirza MU, Rafique S, Ali A, Munir M, Ikram N, Manan A, Salo-Ahen OMH, Idrees M. Zika virus: Immunoinformatics combined with molecular dynamics simulations to predict antigenic epitopes of Zika viral proteins. Scientific Reports. 2016;**6**:37313. DOI: 10.1038/srep37313

[70] Sun G, Larsen CN, Baumgarth N, Klem EB, Scheuermann RH. Comprehensive annotation of mature peptides and genotypes for Zika virus. PLOS ONE. 2017;**12**(1):e0170462. DOI: 10.1371/journal.pone.0170462

[71] Gupta AK, Kaur K, Rajput A, Dhanda SK, Sehgal M, Khan MdS, Monga I, Dar SA, Singh S, Nagpal G, Usmani SS, Thakur A, Kaur G, Sharma S, Bhardwaj A, Qureshi A, Raghava GPS, Kumar M. ZikaVR. An integrated Zika virus resource for genomics, proteomics, phylogenetic and therapeutic analysis. Scientific Reports. 2016;**6**:32713. DOI: 10.1038/srep32713

[72] National Institute on Drug Abuse. NIH, USA. Available from: https://www.drugabuse.gov/

[73] Kimishima A, Wenthur CJ, Eubanks LM, Sato S, Janda KD. Cocaine vaccine development: Evaluation of carrier and adjuvant combinations that activate multiple Toll-Like receptors. Molecular Pharmaceutics. 2016;**13**:3884-3890

[74] Lockner JW, Eubanks LM, Choi JL, Lively JM, Schlosburg JE, Collins KC, Globisch D, Rosenfeld-Gunn RJ, Wilson IA, Janda KD. Flagellin as carrier and adjuvant in cocaine vaccine development. Molecular Pharmaceutics. 2015;**12**:653-662. DOI: 10.1021/mp500520r

[75] Worm M, Lee H-H, Kleine-Tebbe J, Hafner RP, Laidler P, Healey D, Buhot C, Verhoef A, Maillere B, Kay AB, Larché M. Development and preliminary clinical evaluation of a peptide immunotherapy vaccine for cat allergy. The Journal of Allergy and Clinical Immunology. 2011;**127**:89-97. DOI: 10.1016/j.jaci.2010.11.029

[76] Pascal M, Konstantinou GN, Masilamani M, Lieberman J, Sampson HA. *In silico* prediction of Ara h 2 T cell epitopes in peanut-allergic children. Clinical & Experimental Allergy. 2012;**43**:116-127. DOI: 10.1111/cea.12014

[77] Chen H, Yang H-W, Wei JF, Tao A-L. *In silico* prediction of the T-cell and IgE-binding epitopes of Per a 6 and Bla g 6 allergens in cockroaches. Molecular Medicine Reports. 2014;**10**:2130-2136. DOI: 10.3892/mmr.2014.2399

[78] Yang H, Chen H, Jin M, Xie H, He S, Wei J-F. Molecular cloning, expression, IgE binding activities and *in silico* epitope prediction of Per a 9 allergens of the American cockroach. International Journal of Molecular Sciences. 2016;**38**:1795-1805. DOI: 10.3892/ijmm.2016.2793

[79] Tong X, Guo M, Jin M, Chen H, Li Y, Wei J-F. *In silico* epitope prediction, expression and functional analysis of Per a 10 allergen from the American cockroach. International Journal of Molecular Sciences. 2016;**38**:1806-1814. DOI: 10.3892/ijmm.2016.2790

[80] Vormehr M, Schrörs B, Boegel S, Löwer M, Türeci Ö, Sahin U. Mutanome engineered RNA immunotherapy: Towards patient-centered tumor vaccination. Journal of Immunology Research. 2015;**2015**:1-6. DOI: 10.1155/2015/595363

[81] Castle JC, Kreiter S, Diekmann J, Löwer M, Roemer NVD, Graaf JD, Selmi A, Diken M, Boegel S, Paret C, Koslowski M, Kuhn AN, Britten CM, Huber C, Türeci Ö, Sahin U. Exploiting the mutanome for tumor vaccination. Cancer Research. 2012;**72**(5):1081-1091. DOI: 10.1158/0008-5472.CAN-11-3722

[82] Gfeller D, Bassani-Sternberg M, Schmidt J, Luescher IF. Current tools for predicting cancer-specific T cell immunity. Oncoimmunology. 2016;**5**(7):e1177691-1e1177691-9. DOI: 10.1080/2162402X.2016.1177691

[83] Mahdavi M, Moreau V. *In silico* designing breast cancer peptide vaccine for binding to MHC class I and II: A molecular docking study. Computational Biology and Chemistry. 2016;**65**:110-116. DOI: 10.1016/j.compbiolchem.2016.10.007