Predicting epitopes for vaccine development using bioinformatics tools

Therapeutic Advances in Vaccines and Immunotherapy
2022, Vol. 10: 1–12
DOI: 10.1177/25151355221100218

Valentina Yurina and Oktavia Rahayu Adianingsih

***Abstract*:** Epitope-based DNA vaccine development is one application of bioinformatics or *in silico* studies, that is, computational methods, including mathematical, chemical, and biological approaches, which are widely used in drug development. Many *in silico* studies have been conducted to analyze the efficacy, safety, toxicity effects, and interactions of drugs. In the vaccine design process, *in silico* studies are performed to predict epitopes that could trigger T-cell and B-cell reactions that would produce both cellular and humoral immune responses. Immunoinformatics is the branch of bioinformatics used to study the relationship between immune responses and predicted epitopes. Progress in immunoinformatics has been rapid and has led to the development of a variety of tools that are used for the prediction of epitopes recognized by B cells or T cells as well as the antigenic responses. However, the *in silico* approach to vaccine design is still relatively new; thus, this review is aimed at increasing understanding of the importance of *in silico* studies in the design of vaccines and thereby facilitating future research in this field.

Keywords: antibody, B cells, epitope, immunoinformatics, T cells, tools

Received: 21 January 2022; revised manuscript accepted: 14 April 2022.

Introduction

The development of new drugs is time-consuming and costly. For example, the prerequisite average costs for the development of an active substance accepted by the Food and Drug Administration (FDA) are $648.0 million (US dollars; range: $157.3–1950.8 million) and the average time required to develop a new drug is 7.3 years (range: 5.8–15.2 years).[1] Furthermore, around 53% of newly developed drugs fail to reach the preclinical phase, mainly due to intolerable side effects, unacceptable toxicological effects, and unpredictable drug interactions. For these reasons, the development of new drugs is not a top priority for the pharmaceutical industry, which instead largely focuses on the development and production of pre-existing active compounds to increase profits and reduce losses. Nevertheless, many effective treatments have yet to be discovered, especially for multifactorial diseases such as degenerative diseases.[2,3]

When vaccines were first created, they were intended to prevent diseases caused by infectious agents. However, as vaccine technology has developed, vaccines have been expanded in an effort to combat noninfectious diseases such as autoimmune diseases,[4,5] cancer,[6–8], and degenerative diseases.[9–11]

Conventional vaccines are produced by inactivating or attenuating some part of an infectious agent and exposing it to the body's immune system. This approach has been so successful that vaccines are considered one of the major successes of the modern world. In particular, this approach has been effective against infectious agents with low antigen variations such as polio, smallpox, measles, and rubella.[12,13] However, for diseases with mechanisms involving complex immune reactions, this approach is often ineffective; thus, new strategies for vaccine development are required.[14] Conventional vaccination methods also often trigger side effects including fever and

Correspondence to:
Valentina Yurina
Department of Pharmacy, Medical Faculty, Universitas Brawijaya, Jalan Veteran, Malang 65145, East Java, Indonesia.
v_yurina@ub.ac.id

Oktavia Rahayu Adianingsih
Department of Pharmacy, Medical Faculty, Universitas Brawijaya, Malang, Indonesia

hypersensitivity reactions. Therefore, it is necessary to develop a new generation of vaccines, such as epitope-based vaccines, with high effectiveness and minimal side effects.

The evolution of epitope-based vaccines is one of the most promising developments to arise from bioinformatics-based research.[15] Bioinformatics or *in silico* studies, that is, computational methods that include mathematical, chemical, and biological approaches, are widely used in drug development. For example, *in silico* studies are often utilized to analyze the bioavailability of drug compounds,[16,17] pharmacokinetic–pharmacodynamic processes,[18] interactions among drug compounds, and the toxic effects of drug compounds.[19,20]

During vaccine design, *in silico* studies are performed to predict epitopes that can trigger both T-cell and B-cell reactions, which in turn produce cellular and humoral immune responses.[21] Immunoinformatics is the branch of bioinformatics in which the relationship between immune responses and predicted epitopes is studied.[22] The rapid development of immunoinformatics has been characterized by the creation of tools used for the prediction of epitopes recognized by B cells and T cells and for antigen responses.[23,24] Nevertheless, the *in silico* approach to vaccine design is still relatively new; therefore, it is necessary to conduct in-depth studies that will increase general understanding of the importance of *in silico* research to the vaccine design process.

In silico approach for vaccine design

The aim of vaccination is to stimulate the memory of the adaptive immune system to ensure that it responds immediately to the next antigen exposure. The adaptive immune system consists of two classes, namely the humoral immune system mediated by antibodies produced by B lymphocyte cells and the cellular immune system mediated by T lymphocytes. The humoral and cellular immune systems are stimulated when a receptor recognizes a particular part of an antigen known as an epitope.[21,25]

Conventional vaccination approaches that use weakened or activated antigens are ineffective against several types of diseases, especially those involving complex immunity such as HIV, tuberculosis, cancer, and atherosclerosis. Thus, to increase specificity, effectiveness, and safety, bioinformatics methods are used during epitope-based vaccine development.[10,26–28] This approach minimizes resource use and time costs because the initial screening is conducted *in silico* to increase the efficiency of the vaccine candidate search.[23] In comparison to conventional vaccines, epitope-based vaccines are typically well-tolerated and have fewer side effects.[15,24] At present the approach has been demonstrated to be effective designing vaccine, including for COVID-19 vaccines[29–31] and other infectious diseases.[32,33]

During vaccine design, the bioinformatics approach is based on the availability of data and epitope predictions.[34] In broad terms, the steps involved in vaccine design using the *in silico* method include searching antigen protein databases, analyzing protein interactions, characterizing the epitopes recognized by both T cells and B cells, and analyzing antigenicity and homology. This approach usually requires massive amounts of reliable antigenic protein data. The prediction of epitopes recognized by B cells and T cells is based on sequences and structure and not on pathological mechanisms.[35]

Protein databases and protein interaction analysis

Massive, reliable protein databases are required for the design of epitope-based vaccines. Currently, several such protein databases can be easily accessed including those provided by the National Center of Biotechnology Information (NCBI), UniProt, or Protein Data Bank (PDB).

From the NCBI website (www.ncbi.nlm.nih.gov), various data related to proteins can be accessed. Some programs for analysis can also be operated from this website including a search and retrieval system that provides users with integrated access to sequence, mapping, taxonomy, and structural data. Moreover, the website provides sequence similarity search tools and can be used to identify genes and genetic features.[36]

The PDB (www.rcsb.org) provides access to protein-related data including protein sources, crystallographic data, chemical structures, peptide sequences, and protein structures based on nuclear magnetic resonance (NMR). In addition, this

database enables various protein visualizations and protein analyses such as sequencing, prediction of protein structure, and protein symmetry analysis. Furthermore, the PDB provides a domain-based structural alignment method. It also includes structure depositions that have been determined using several techniques including macromolecular crystallography, three-dimensional (3D) electron microscopy (EM), powder diffraction, and fiber diffraction.[34,37,38]

UniProt (www.uniprot.org) provides almost complete information on proteins including data on functions, names and taxonomies, subcellular locations, related pathologies, post-translational modifications (protein processing), expressions, interactions, structures, sequences, families and domains, reference information, and similar proteins. Moreover, Uniprot includes improved metagenomic assembly and binning tools that provide high-quality metagenomic assembled genomes. In addition, Uniprot provides the UniRef databases, which cluster sequence sets at various levels of sequence identity, and the UniProt Archive (UniParc), which delivers a complete set of known sequences.[39,40]

Many new databases and tools have been developed as accessible repositories for storing and analyzing large amounts of immunology-related biological data. Most of these databases have been listed as public repositories to make it easier for researchers to find the databases they need. These public repositories provide access to up-to-date annotated lists of immunoinformatic resources, ensuring the quality and relevance of these databases and tools.[41] The three main public repositories containing information on available databases and tools related to immunoinformatics are (1) Nucleic Acids Research Database Annual Issue, (2) Canadian Bioinformatics Links Directory, and (3) Immune Epitope Database and Analysis Resources (IEDB).

Antigenicity and epitope prediction

After protein data is analyzed, 3D protein structure analysis or modeling can be conducted. In general, homology modeling or comparative modeling methods are performed because not all studied proteins have known 3D structures. Using these methods, protein structure can be predicted based on alignment results with one or more other proteins for which the structure is known. The program largely applied in protein modeling is MODELER, in which information from an input target-template alignment is used to create a series of homology-derived spatial restraints that act on the atoms of the 3D protein model. Sigma values of homology-derived distance restraints define the acceptable amount of conformational freedom for the model based on its templates.[42]

Analysis of protein interactions is performed using protein docking, which predicts the formation of protein complexes or ligands based on binding models and surface free energy. Protein docking can be divided into two processes: sampling and scoring. Sampling is a method used to determine which parts of a protein are relevant to conformational or binding orientations. The sampling process can involve the use of a binding orientation algorithm (rigid-body sampling) or may be based on protein conformation (conformational sampling). Once sampling is completed, scoring is conducted to assess each binding model. Moreover, each binding model is sorted, with the binding model possessing the highest score suggested as a protein complex formation model.[43]

Molecular docking is then performed to ensure that a candidate epitope vaccine could generate a stable immune response. This is achieved by measuring interactions between the candidate and target immune cell receptors such as Toll-like receptor 2 (TLR2), TLR3, and TLR4.[44–46] Several studies have included molecular docking of vaccines with the human leukocyte antigen molecule or major histocompatibility complex I (MHC I) and MHC II receptors.[45,47] After molecular docking is completed, the protein–protein interactions among docked molecules are also analyzed.

Antigenicity prediction

Antigenicity prediction is used to determine the peptides that have high antigenicity and can be developed as vaccine candidates. The many tools used to predict antigenicity are based on various antigenicity determination methods, for example, the Kolaskar–Tongaonkar method[48] and Welling method.[49]

The Kolaskar–Tongaonkar method is based on experimental research indicating that hydrophobic residues, such as cysteine, leucine, and valine, on the surface of proteins tend to have antigenic characteristics. Based on these experimental data, a semi-empirical method was developed to assess whether a peptide is or is not antigenic. The method has an accuracy of around 75% and has the advantage of being simple to use as it only requires one parameter.[48] The Welling method is used to determine an antigenicity value based on a comparison between the percentage of specific amino acids on the antigenic side and the percentage of these amino acids in the protein.[49]

B-cell epitope mapping

Structural epitope mapping could be conducted using X-ray crystallography, nuclear magnetic resonance (NMR),[50] EM,[51] or cryoelectron microscopy (CM).[52,53] X-ray crystallography is believed to be the most precise method for structural epitope mapping. However, the quality of cocrystals and the antibody's electron density limit X-ray crystallography.[54] NMR offers peptide mapping based on the difference in the NMR signal of the free antigen or the antibody-bound antigen to determine the epitopes. NMR epithelial mapping provides more detailed information than mutagenesis or peptide mapping and can be much faster than X-ray crystallography.[55,56] CM is another technique used to determine macromolecular structures with resolution comparable to X-ray crystallography. Because the samples are flash frozen in CM, crystallization is not required. Typically, fewer samples are required, but they must still be relatively homogeneous in purity. CM provides higher resolution information for larger molecules and less information for smaller molecules.[57]

B-cell epitope prediction

B cells mediate the humoral immune system through antibody secretion that neutralizes antigens. B cells are stimulated when the antigen receptor, which is part of the paratope, recognizes antigenic epitopes. Most available epitope mapping methods (structural and functional approaches) are costly, time-consuming, and frequently fail to detect all epitopes. The protein structure including residues in direct contact with an antibody is interpreted using structural epitope mapping methods, although these methods frequently fail to identify the role of amino acids in binding strength. The goal of functional epitope mapping techniques is to identify and characterize residues critical for binding within structurally specified antigenic determinants.[24]

Epitopes recognized by B cells can be classified into two types: continuous and discontinuous epitopes. Continuous epitopes (also referred to as linear or sequential epitopes) are short peptide fragments (about 15 amino acids in size) of an antigen protein that are specifically identified by certain antibodies. Discontinuous epitopes consist of amino acid residues that are not sequential in their primary structure but involve a folding mechanism that forms into a region that is close together. However, the folding mechanism increases the complexity of epitope prediction; the classification is not rigid because several continuous epitopes could form certain conformations that are recognized by antibodies and discontinuous epitopes can also contain several sequential linear peptide sequences.[58] Because of their complexity, the prediction of B-cell epitopes is often less accurate than the simpler prediction of T-cell epitopes.

Linear epitope prediction

Sequence-based prediction can be used to predict continuous epitopes based on the propensity scale method, which is used to assess and compare the tendency of amino acids to become epitopes recognized by B cells relative to amino acids that form antigens. To determine the propensity value for a residue, i, a central residue in a window chosen with size n, we would use the formula $i-(n-1)/2$. The value of residue i is the average value for amino acids in a predetermined window range. In general, 5–7 amino acids are used to determine an epitope. The assessment is based on the physical characteristics of these amino acids, for example, hydrophilicity, flexibility, solvent accessibility, or protein helixes.[59,60]

Hydrophilic scores are determined based on amino acid retention times in high-performance liquid chromatography in the reverse phase column. In such assessments, a window consisting of seven amino acids has been used, in which for the fourth amino acid residue value is determined

from the average hydrophilicity value of the seven residues.[61]

The flexibility assessment by Karplus and Schulz is based on the mobility of protein segments at a factor B temperature of carbon α for 31 proteins with known structures. This flexibility calculation uses the first amino acid from a window span consisting of six amino acids.[59]

Solvent accessibility scores are determined based on the probability of an amino acid being exposed to the X-ray structure of 28 proteins. The surface probability (*Sn*) is determined using the following formula:

$$Sn = \left[\prod_{i=1}^{6} \sigma\, n+4-i\right] * (0.37)^{-6}$$

where *Sn* is surface probability, δ*n* is the fractional probability value of the surface, and the *i* value varies from 1 to 6. A hexapeptide with $Sn > 1$ indicates an increased probability that an amino acid will be on the surface.[62] Moreover, an assessment by Chou and Fasman is based on the probability that a certain range of residues are part of a β-turn structure.[59]

Studies have shown that predictions made using a single physicochemical characteristic cannot accurately predict B-cell epitopes. Therefore, some tools simultaneously use a combination of physicochemical character assessments, such as PREDITOP, PEOPLE, and BEPITOPE, to improve the accuracy of predictions.[21,63,64] The machine learning method is a computational method that uses a train classifier to distinguish epitope and non-epitope antigenic structures based on data related to structural differences and physicochemical characteristics.[64]

Sequence-based prediction has the advantage of not requiring any understanding of the target antigen's 3D structure. To determine the 3D structure of a target antigen, data from X-ray crystallography studies are required; however, not all target antigens have known 3D structures. In contrast, the disadvantage of sequence-based prediction of B-cell epitopes is their relatively low accuracy, which on average is 60–70% because the majority of epitopes recognized by B cells are naturally in a discontinuous condition.[60]

Furthermore, newer methods have been developed to predict continuous B-cell epitopes. The SVMTriP service predicts continuous B-cell epitopes using the support vector machine algorithm (SVM) and the Tri-peptide similarity and propensity score (SVMTriP) in order to improve prediction accuracy.[65] Another approach is using validated B-cell epitopes as well as non B-cell epitopes from Immune Epitope Database which is resulting in two types of datasets called Lbtope_Variable and Lbtope_Fixed length.[66]

Confirmed epitope prediction

The prediction of a confirmed epitope was developed because 90% of the epitopes recognized by B cells are in a discontinuous condition or form specific conformations. The first discontinuous epitope prediction method developed was the conformational epitope prediction, which can predict continuous or discontinuous epitopes using 3D protein structures. The underlying algorithm employs solvent accessibility data based on the Voronoi polyhedron. Continuous epitopes are determined based on the presence of at least three sequential residues, whereas discontinuous epitopes are determined by decreasing continuous epitopes with Cα within 6Å.[24]

Another method, DiscoTope, uses a combination of amino acid statistics, spatial context, and amino acid surface accessibility to predict B-cell epitopes.[67] This method can detect 15.5% of the residues present in discontinuous epitopes with a specificity of up to 95%; at this level of specificity, Parker's hydrophilicity method can only detect 11% of residues in discontinuous epitopes.[24,26,67]

ElliPro is a tool based on the combination of the Thornton concept, the MODELER program, and Jmol viewer.[59] For the prediction of B-cell epitopes, ElliPro uses three steps: estimation of protein structure as an ellipsoid, calculation of the residual protrusion index (PI), and grouping of residues based on PI values. PI values are defined as the percentage of protein atoms inside the ellipsoid where the first residue is outside the ellipsoid.[59] The Antigenic Epitopes Prediction with Support Vector Regression server (EPSVR) manipulates vector regression to combine the same scores as EPCES and achieves an area under the curve (AUC) of 0.597.[68] Other tools that can be used to predict discontinuous B-cell

Table 1. B-cell epitope prediction tools.

Tools	Description	URL
ABCpred	Based on sequence with ANN	http://crdd.osdd.net/raghava/abcpred/
BEPITOPE	Based on sequence to predict continuous epitope	http://bepitope.ibs.fr/
BCPREDS	Predicting linear B-cell epitopes using the subsequence kernel	http://ailab-projects1.ist.psu.edu:8080/bcpred/index.html
Bepro	Based on antigen structure to predict discontinuous epitope	http://pepito.proteomics.ics.uci.edu/
CEP	Based on structure to predict continuous and discontinuous epitopes	http://bioinfo.ernet.in/cep.htm
COBEpro	Based on B-cell epitope primer sequence. Secondary structure and solvent accessibility are also responsible for increasing prediction accuracy	http://scratch.proteomics.ics.uci.edu/
DiscoTope	Based in sequence and structure for predicting continuous and discontinuous epitopes	http://www.cbs.dtu.dk/
Ellipro	Based on solvent accessibility and protein flexibility	http://tools.immuneepitope.org/tools/ElliPro/iedbinput
EMT	Based on phage display to predict continuous and discontinuous epitopes	elro@novozymes.com
EPCES	Prediction of discontinuous epitopes using support vector regression and multiple server	http://sysbio.unl.edu/EPCES/
EPIMAP	Based on phage display to predict continuous and discontinuous epitopes	mumey@cs.montana.edu
Epitopia	Based on linier sequence or 3D structure	http://epitopia.tau.ac.il
IEDB B-cell epitope tools	Based on amino acid scale for continuous epitope prediction and 3D structure for discontinuous epitope prediction	http://tools.immuneepitope.org/main/html/B-cell tools.html
LBtope	Using various techniques (e.g. SVM, IBk) on a large dataset of B-cell epitopes and non-epitopes	http://crdd.osdd.net/raghava/lbtope/
SVMTriP	Based on support vector machine (SVM) which is combining the tri-peptide similarity and propensity scores (SVMTriP)	http://sysbio.unl.edu/SVMTriP/

ANN, artificial neural network.

epitopes include Epitopia,[69] PEPOP,[70] EPIMAP,[60] and CBTOPE[71] (Table 1).

T-cell epitope prediction

Compared with B-cell epitope predictions, T-cell epitope predictions are generally easier and more accurate because the structures of epitopes identified by T cells are simpler, that is, short, linear peptides (9–15 amino acids in length). Epitopes are recognized by the T-cell receptors (TCRs) in the form that is presented by the MHCs, that is, MHC class I or class II. It is important to consider both epitope and MHC bonding and

epitope–MHC and TCR complex bonds during the prediction of epitopes recognized by T cells. Epitopes bind to specific parts of MHCs, known as grooves, which are usually formed from two α helices and one β sheet and are then presented to T cells.[72] Peptides are bound to MHCs through hydrogen bonds, electrostatic interactions, and van der Waals interactions. In general, peptides that bind to MHC class I have 8–11 amino acids sizes, whereas peptides that bind to MHC class II are 12–25 amino acids in length and protrude from the MHC groove but have at least 9 amino acids in the core.[73] However, other studies have shown that some larger peptides can also bind to the MHC but have a lower immunogenic potential.[23,74]

Some of the methods used to predict epitopes that are recognized by T cells are the motif-based system, matrix, SVM, empirical scoring, and molecular dynamics (MDs) methods.[73] The motif-based system was the first T-cell epitope prediction method developed. In this method, amino acid sequences that have a high tendency to bind to the MHC groove or so-called motif are predicted. The amino acid sequence is then compared with the data in a library motif, where the previously determined binding peptide sequence and the nonbinding MHC-binding motif are collected. The accuracy of this method can reach 60–70% because not all peptides have known motifs.[23]

Other motif-based system has been developed based on machine learning algorithms (MLAs). For instance, based on MLAs, peptide-binding motifs can be determined according to certain classifications, for example, a positive value for a peptide binder and a negative value for a nonpeptide binder. MLAs can also be used for several classifications at the same time. Artificial neural networks are one of the types of MLAs most widely used to determine the motifs for introducing peptides to MHCs.[22,75]

Prediction of T-cell epitopes can also be performed by simulating MDs, in which free binding energy is calculated for a molecular system. MDs can be used to explain the movement of atoms individually or collectively in a molecular system; thus, MDs provide a dynamic picture. The advantage of MDs relative to other methods is that they are not based on data alone but based on *de novo* predictions of all parameters that construct the structure of the receptor ligand complex.[76] A summary of the tools that can be used to predict epitopes recognized by T cells is shown in Table 2.

In silico studies offer a new solution for cutting costs and time, which is important for drug development. By using *in silico* studies, we can predict the effectiveness of a drug compound and thereby design an 'ideal' drug. Indeed, with the aid of *in silico* technology, before a new drug is developed, its effectiveness, side effects, potential for contractions, and toxic effects can be determined in advance. Thus, the time and costs required for development can be reduced.

During vaccine development, *in silico* studies can provide huge benefits. Conventional vaccines that use all or part of a weakened or inactivated pathogen often cause severe side effects such as fever and hypersensitivity reactions. The imperfect inactivation process also allows active pathogens to enter the body and cause symptoms of the disease. The development of recombinant protein-based subunit vaccines requires a long time because it must include the most potent antigen screening process among many other antigen proteins.[23] The development of recombinant protein-based vaccines is expensive because the production process must be sterile. Recombinant protein stability is also relatively low; therefore, vaccines must be stored at a certain temperature, which increases difficulties with their distribution and storage. Therefore, epitope-based vaccines are considered to be a solution to the problems of conventional vaccines, including vaccines for infectious diseases,[28,29,77,78] and even metabolic disorders or inflammatory diseases.[10]

The development of informatics has given rise to new programs with respective advantages. The emergence of epitope prediction tools, antigenicity prediction, protein modeling, and docking analysis has made it possible to design epitope-based vaccines with maximum efficacy. In addition, the development of X-ray crystallization technology, NMR spectrophotometry, and CM has revealed the 3D structure of an increasing number of proteins, which in turn has facilitated the analysis of protein interactions. For proteins with unknown dimensions, 3D modeling and docking analysis methods have enabled

Table 2. T-cell epitope prediction tools.

Tool	Description	URL
EpiMatrix	Based on protein binding efficiency with MHC class I and II	http://www.epivax.com/
FRAGPREDICT	Based on proteasome cleavage site binding score	http://www.mpiib-berlin.mpg.de/MAPPP/cleavage.html
Immune Epitope Database and Analysis Resource (IEDB)	Prediction based on analysis of proteasomal processing, TAP transport, and MHC class I and II binding	http://www.immuneepitope.org/
MHCPred	Based on the binding value of MHC/peptide or TAP/peptideIC50	http://www.jenner.ac.uk/
MMBPred	Determination of high-affinity MHC binding peptide that undergoes mutations	http://www.imtech.res.in/raghava/mmbpred/
NetChop	Based on the immunoproteasome cleavage site	http://www.cbs.dtu.dk/services/NetChop/
NetCTL	Based on the combination of MHC subtype binding values, Tap transport and proteasome	http://www.cbs.dtu.dk/services/NetCTL/
NetMHC	Based on the binding propensity of peptides to different HLA alleles using ANN	http://www.cbs.dtu.dk/
ProPred-1	Based on peptide binding efficiency with MHC I	http://www.imtech.res.in/raghava/propred1
SYFPEITHI	Based on motif binding to MHC class I and II	http://www.syfpeithi.com/
TAPPred	Based on binding affinity with TAP protein	http://www.imtech.res.in/raghava/tappred/
RANKPEP	Predicts peptide binders to MHC I and MHC II molecules using position specific scoring matrices (PSSMs)	http://imed.med.ucm.es/Tools/rankpep.html
Epijen	Based on the immunoproteasome cleavage site and TAP binding affinity	http://www.ddg-pharmfac.net/epijen/EpiJen/EpiJen.htm
nHLAPred	Based on the hybrid approach of artificial neural networks (ANNs) and quantitative matrices (QMs)	http://crdd.osdd.net/raghava/nhlapred/

predictions of protein interactions including prediction of bonds between antibodies and antigens with the highest affinities.[79]

The application of *in silico* studies to the design of epitope-based vaccines is also relatively simple and does not require complex skills. The necessary tools used are also widely available for free and can be accessed easily. The immune epitope database (IEDB) provides tools for the prediction of epitopes that are recognized by B cells and T cells, as well as for analyzing epitope characteristics for more complete and reliable prediction results. This database and its associated tools have often been used in studies in which epitopes were predicted for vaccine development,[22,63] perhaps because the resource is easy to use. The main disadvantage of using *in silico* studies to develop epitope-based vaccines is that all predictions are computationally based on approaches involving mathematics, chemistry, and biology; thus, the accuracy never reaches 100%. In a

biological system, there can be unpredictable interactions since proteins are dynamic macromolecular complexes. Protein 3D conformations are prone to changes in the physical environment, such as changes in charge and pH, which disrupt the structure and activity of the protein including its binding with other proteins.[80] Antibodies are also proteins that are specific to certain antigens; changes in one residue alone prevent recognition by these antibodies. To improve the accuracy of epitope prediction, it is necessary to analyze MDs to validate the binding of antibodies to receptors. By improving accuracy in this manner, the effectiveness of vaccines is also expected to be improved.

Conclusion

Based on our reviews, the immunoinformatic tools are very valuable tools for predicting and evaluating the epitopes for vaccine candidate development. These tools undeniably are becoming the most informative and advantageous device for vaccine design.

Author contributions

Valentina Yurina: Conceptualization; Methodology; Writing – original draft; Writing – review & editing.

Oktavia Rahayu Adianingsih: Investigation; Writing – original draft.

ORCID iD

Valentina Yurina https://orcid.org/0000-0003-4319-942X

Funding

The authors received no financial support for the research, authorship, and/or publication of this article.

Conflict of interest statement

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

References

1. Prasad V and Mailankody S. Research and development spending to bring a single cancer drug to market and revenues after approval. *JAMA Intern Med* 2017; 177: 1569–1575.
2. Rifaioglu AS, Atas H, Martin MJ, *et al*. Recent applications of deep learning and machine intelligence on in silico drug discovery: methods, tools and databases. *Brief Bioinform* 2019; 20: 1878–1912.
3. Mottini C, Napolitano F, Li Z, *et al*. Computer-aided drug repurposing for cancer therapy: approaches and opportunities to challenge anticancer targets. *Semin Cancer Biol* 2021; 68: 59–74.
4. Ahmad B, Batool M, Kim MS, *et al*. Computational-driven epitope verification and affinity maturation of tlr4-targeting antibodies. *Int J Mol Sci* 2021; 22: 5989.
5. Zhang N and Nandakumar KS. Recent advances in the development of vaccines for chronic inflammatory autoimmune diseases. *Vaccine* 2018; 36: 3208–3220.
6. Safavi A, Kefayat A, Abiri A, *et al*. In silico analysis of transmembrane protein 31 (TMEM31) antigen to design novel multiepitope peptide and DNA cancer vaccines against melanoma. *Mol Immunol* 2019; 112: 93–102.
7. Lőrincz O, Tóth J, Molnár L, *et al*. In silico model estimates the clinical trial outcome of cancer vaccines. *Cells* 2021; 10: 3048.
8. Patra P, Bhattacharya M, Sharma AR, *et al*. Identification and design of a next-generation multi epitopes bases peptide vaccine candidate against prostate cancer: an in silico approach. *Cell Biochem Biophys* 2020; 78: 495–509.
9. Brisse M, Vrba SM, Kirk N, *et al*. Emerging concepts and technologies in vaccine development. *Front Immunol* 2020; 11: 583077.
10. Yurina V, Yudani T, Raras M, *et al*. Design and construction of DNA vaccine expressing lectin-like oxidize-LDL receptor-1 (LOX-1) as atherosclerosis vaccine candidate. *J Biotech Res* 2017; 1: 103–112.
11. Zhao B, Marciniuk K, Gibbs E, *et al*. Therapeutic vaccines for amyotrophic lateral sclerosis directed against disease specific epitopes of superoxide dismutase 1. *Vaccine* 2019; 37: 4920–4927.
12. Puig-Barberà J, Burtseva E, Yu H, *et al*. Influenza epidemiology and influenza vaccine effectiveness during the 2014-2015 season: annual report from the Global Influenza Hospital Surveillance Network. *BMC Public Health* 2016; 16: 757.
13. Vela Ramirez JE, Sharpe LA and Peppas NA. Current state and challenges in developing oral vaccines. *Adv Drug Deliv Rev* 2017; 114: 116–131.

14. Grimm SK and Ackerman ME. Vaccine design: emerging concepts and renewed optimism. *Curr Opin Biotechnol* 2013; 24: 1078–1088.
15. Correia BE, Bates JT, Loomis RJ, *et al*. Proof of principle for epitope-focused vaccine design. *Nature* 2014; 507: 201–206.
16. Paixão P, Gouveia LF, Morais J, *et al*. Prediction of the human oral bioavailability by using in vitro and in silico drug related parameters in a physiologically based absorption model. *Int J Pharm* 2012; 429: 84–98.
17. Olasupo SB, Uzairu A, Shallangwa GA, *et al*. Unveiling novel inhibitors of dopamine transporter via in silico drug design, molecular docking, and bioavailability predictions as potential antischizophrenic agents. *Futur J Pharm Sci* 2021; 7: 63.
18. Moroy G, Martiny VY, Vayer P, *et al*. Toward in silico structure-based ADMET prediction in drug discovery. *Drug Discov Today* 2012; 17: 44–55.
19. Kazmi SR, Jun R, Yu MS, *et al*. In silico approaches and tools for the prediction of drug metabolism and fate: a review. *Comput Biol Med* 2019; 106: 54–64.
20. Ntie-Kang F, Lifongo LL, Mbah JA, *et al*. In silico drug metabolism and pharmacokinetic profiles of natural products from medicinal plants in the Congo basin. *In Silico Pharmacol* 2013; 1: 12.
21. Sanchez-Trincado JL, Gomez-Perosanz M and Reche PA. Fundamentals and methods for T- and B-cell epitope prediction. *J Immunol Res* 2017; 2017: 2680160.
22. Jørgensen KW, Rasmussen M, Buus S, *et al*. NetMHCstab – predicting stability of peptide-MHC-I complexes; impacts for cytotoxic T lymphocyte epitope discovery. *Immunology* 2014; 141: 18–26.
23. Patronov A and Doytchinova I. T-cell epitope vaccine design by immunoinformatics. *Open Biol* 2013; 3: 120139.
24. Potocnakova L, Bhide M and Pulzova LB. An introduction to B-cell epitope mapping and in silico epitope prediction. *J Immunol Res* 2016; 2016: 6760830.
25. Topuzoğullari M, Acar T, Pelit Arayici P, *et al*. An insight into the epitope-based peptide vaccine design strategy and studies against COVID-19. *Turk J Biol* 2020; 44: 215–227.
26. Sitompul LS, Widodo N, Djati MS, *et al*. Epitope mapping of gp350/220 conserved domain of Epstein Barr virus to develop nasopharyngeal carcinoma (NPC) vaccine. *Bioinformation* 2012; 8: 479–482.
27. Kawakami R, Nozato Y, Nakagami H, *et al*. Development of vaccine for dyslipidemia targeted to a proprotein convertase subtilisin/kexin type 9 (PCSK9) epitope in mice. *PLoS ONE* 2018; 13: e0191895.
28. Adianingsih OR and Kharisma VD. Study of B cell epitope conserved region of the Zika virus envelope glycoprotein to develop multi-strain vaccine. *J Appl Pharm Sci* 2019; 9: 98–103.
29. Feng Y, Jiang H, Qiu M, *et al*. Multi-epitope vaccine design using an immunoinformatic approach for SARS-CoV-2. *Pathogens* 2021; 10: 737.
30. Yurina V. Coronavirus epitope prediction from highly conserved region of spike protein. *Clin Exp Vaccine Res* 2020; 9: 169–173.
31. Shehata MM, Mahmoud SH, Tarek M, *et al*. In silico and in vivo evaluation of SARS-CoV-2 predicted epitopes-based candidate vaccine. *Molecules* 2021; 26: 6182.
32. Majidiani H, Dalimi A, Ghaffarifar F, *et al*. Multi-epitope vaccine expressed in *Leishmania tarentolae* confers protective immunity to *Toxoplasma gondii* in BALB/c mice. *Microb Pathog* 2021; 155: 104925.
33. Abdollahi S, Raoufi Z and Fakoor MH. Physicochemical and structural characterization, epitope mapping and vaccine potential investigation of a new protein containing Tetratrico Peptide Repeats of *Acinetobacter baumannii*: an in-silico and in-vivo approach. *Mol Immunol* 2021; 140: 22–34.
34. Rose PW, Bi C, Bluhm WF, *et al*. The RCSB Protein Data Bank: new resources for research and education. *Nucleic Acids Res* 2013; 41: D475–D482.
35. Fink K. Can we improve vaccine efficacy by targeting T and B cell repertoire convergence? *Front Immunol* 2019; 10: 110.
36. Agarwala R, Barrett T, Beck J, *et al*. Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res* 2018; 46: D8–D13.
37. Burley SK, Berman HM, Bhikadiya C, *et al*. Protein Data Bank: the single global archive for 3D macromolecular structure data. *Nucleic Acids Res* 2019; 47: D520–D528.
38. Rose PW, Beran B, Bi C, *et al*. The RCSB Protein Data Bank: redesigned web site and web services. *Nucleic Acids Res* 2011; 39: D392–D401.

39. Magrane M UniProt Consortium. UniProt Knowledgebase: a hub of integrated protein data. *Database (Oxford)* 2011; 2011: bar009.
40. The UniProt Consortium. Erratum: UniProt: the universal protein knowledgebase (*Nucleic Acids Res* 2017; 45: D158–D169). *Nucleic Acids Res* 2018; 46: 2699.
41. Kazi A, Chuah C, Majeed ABA, *et al*. Current progress of immunoinformatics approach harnessed for cellular- and antibody-dependent vaccine design. *Pathog Glob Health* 2018; 112: 123–131.
42. Janson G, Grottesi A, Pietrosanto M, *et al*. Revisiting the 'satisfaction of spatial restraints' approach of MODELLER for protein homology modeling. *PLoS Comput Biol* 2019; 15: e1007219.
43. Smith GR and Sternberg MJ. Prediction of protein-protein interactions by docking methods. *Curr Opin Struct Biol* 2002; 12: 28–35.
44. Naz A, Shahid F, Butt TT, *et al*. Designing multi-epitope vaccines to combat emerging coronavirus disease 2019 (COVID-19) by employing immuno-informatics approach. *Front Immunol* 2020; 11: 1663.
45. Kar T, Narsaria U, Basak S, *et al*. A candidate multi-epitope vaccine against SARS-CoV-2. *Sci Rep* 2020; 10: 1–24.
46. Ashfaq UA, Saleem S, Masoud MS, *et al*. Rational design of multi epitope-based subunit vaccine by exploring MERS-COV proteome: reverse vaccinology and molecular docking approach. *PLoS ONE* 2021; 16: e0245072.
47. Hossain MS, Hossan MI, Mizan S, *et al*. Immunoinformatics approach to designing a multi-epitope vaccine against Saint Louis Encephalitis Virus. *Informatics Med Unlocked* 2021; 22: 100500.
48. Kolaskar AS and Tongaonkar PC. A semi-empirical method for prediction of antigenic determinants on protein antigens. *FEBS Lett* 1990; 276: 172–174.
49. Welling GW, Weijer WJ, van der Zee R, *et al*. Prediction of sequential antigenic regions in proteins. *FEBS Lett* 1985; 188: 215–218.
50. Zuniga A, Rassek O, Vrohlings M, *et al*. An epitope-specific chemically defined nanoparticle vaccine for respiratory syncytial virus. *NPJ Vaccines* 2021; 6: 85.
51. Walls AC, Tortorici MA, Frenz B, *et al*. Glycan shield and epitope masking of a coronavirus spike protein observed by cryo-electron microscopy. *Nat Struct Mol Biol* 2016; 23: 899–905.
52. Toole EN, Dufresne C, Ray S, *et al*. Rapid highly-efficient digestion and peptide mapping of adeno-associated viruses. *Anal Chem* 2021; 93: 10403–10410.
53. Gallagher JR, McCraw DM, Torian U, *et al*. Characterization of hemagglutinin antigens on influenza virus and within vaccines using electron microscopy. *Vaccines* 2018; 6: 31.
54. Ahmad TA, Eweida AE and Sheweita SA. B-cell epitope mapping for the design of vaccines and effective diagnostics. *Trials Vaccinol* 2016; 5: 71–83.
55. Bardelli M, Livoti E, Simonelli L, *et al*. Epitope mapping by solution NMR spectroscopy. *J Mol Recognit* 2015; 28: 393–400.
56. Valente AP and Manzano-Rendeiro M. Mapping conformational epitopes by NMR spectroscopy. *Curr Opin Virol* 2021; 49: 1–6.
57. Vishweshwaraiah YL and Dokholyan NV. Toward rational vaccine engineering. *Adv Drug Deliv Rev* 2022; 183: 114142.
58. Lundegaard C, Lund O, Keşmir C, *et al*. Modeling the adaptive immune system: predictions and simulations. *Bioinformatics* 2007; 23: 3265–3275.
59. Ponomarenko J, Bui HH, Li W, *et al*. ElliPro: a new structure-based tool for the prediction of antibody epitopes. *BMC Bioinformatics* 2008; 9: 514.
60. El-manzalawy Y and Honavar V. Recent advances in B-cell epitope prediction methods. *Immunome Res* 2010; 6: S2.
61. Parker JM, Guo D and 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: 5425–5432.
62. Emini EA, Hughes JV, Perlow DS, *et al*. Induction of hepatitis A virus-neutralizing antibody by a virus-specific synthetic peptide. *J Virol* 1985; 55: 836–839.
63. Jespersen MC, Peters B, Nielsen M, *et al*. BepiPred-2.0: improving sequence-based B-cell epitope prediction using conformational epitopes. *Nucleic Acids Res* 2017; 45: W24–W29.
64. Rubinstein ND, Mayrose I and Pupko T. A machine-learning approach for predicting B-cell epitopes. *Mol Immunol* 2009; 46: 840–847.
65. Yao B, Zhang L, Liang S, *et al*. SVMTriP: a method to predict antigenic epitopes using

support vector machine to integrate tri-peptide similarity and propensity. *PLoS ONE* 2012; 7: e45152.

66. Singh H, Ansari HR and Raghava GP. Improved method for linear B-cell epitope prediction using antigen's primary sequence. *PLoS ONE* 2013; 8: e62216.
67. Haste Andersen P, Nielsen M and Lund O. Prediction of residues in discontinuous B-cell epitopes using protein 3D structures. *Protein Sci* 2006; 15: 2558–2567.
68. Liang S, Zheng D, Standley DM, *et al*. EPSVR and EPMeta: prediction of antigenic epitopes using support vector regression and multiple server results. *BMC Bioinformatics* 2010; 11: 381.
69. Rubinstein ND, Mayrose I, Martz E, *et al*. Epitopia: a web-server for predicting B-cell epitopes. *BMC Bioinformatics* 2009; 10: 287.
70. Moreau V, Fleury C, Piquer D, *et al*. PEPOP: computational design of immunogenic peptides. *BMC Bioinformatics* 2008; 9: 71.
71. Ansari HR and Raghava GPS. Identification of conformational B-cell Epitopes in an antigen from its primary sequence. *Immunome Res* 2010; 6: 6.
72. Peters B, Nielsen M and Sette A. T cell epitope predictions. *Annu Rev Immunol* 2020; 38: 123–145.
73. Tsurui H and Takahashi T. Prediction of T-cell epitope. *J Pharmacol Sci* 2007; 105: 299–316.
74. Burrows SR, Rossjohn J and McCluskey J. Have we cut ourselves too short in mapping CTL epitopes. *Trends Immunol* 2006; 27: 11–16.
75. Lafuente EM and Reche PA. Prediction of MHC-peptide binding: a systematic and comprehensive overview. *Curr Pharm Des* 2009; 15: 3209–3220.
76. Flower DR, Phadwal K, Macdonald IK, *et al*. T-cell epitope prediction and immune complex simulation using molecular dynamics: state of the art and persisting challenges. *Immunome Res* 2010; 6: S4.
77. Raza MT, Mizan S, Yasmin F, *et al*. Epitope-based universal vaccine for Human T-lymphotropic virus-1 (HTLV-1). *PLoS ONE* 2021; 16: e0248001.
78. Michel-Todó L, Reche PA, Bigey P, *et al*. In silico design of an epitope-based vaccine ensemble for Chagas disease. *Front Immunol* 2019; 10: 2698.
79. Huang SY. Search strategies and evaluation in protein–protein docking: principles, advances and challenges. *Drug Discov Today* 2014; 19: 1081–1096.
80. Dalvi H, Bhat A, Iyer A, *et al*. Armamentarium of cryoprotectants in peptide vaccines: mechanistic insight, challenges, opportunities and future prospects. *Int J Pept Res Ther* 2021; 27: 2965–2982.