

# Computational approaches for vaccine designing

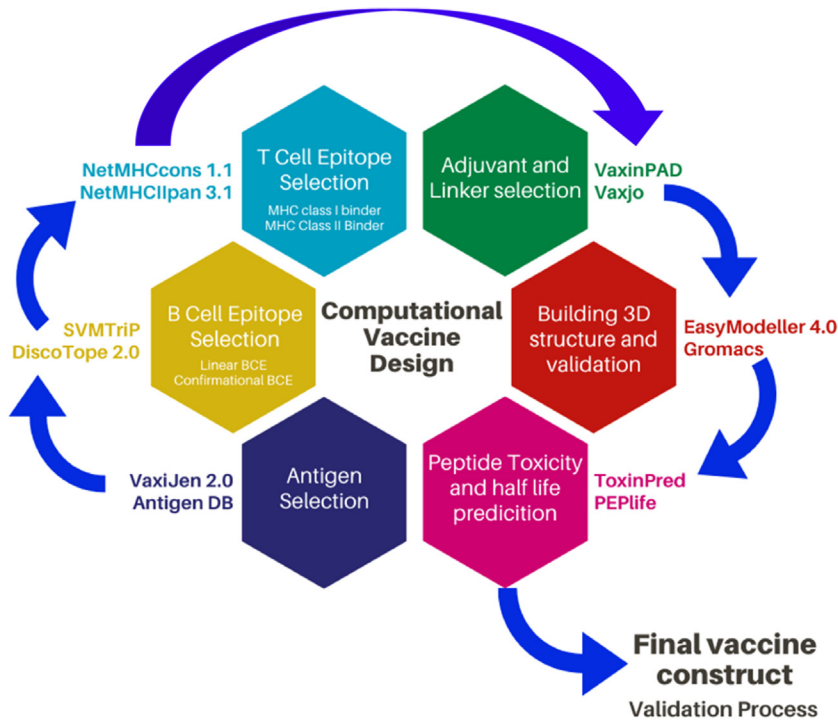
Animesh Awasthi<sup>1,2</sup>, Gaurav Sharma<sup>1</sup> and Piyush Agrawal<sup>3</sup>

<sup>1</sup>Institute of Bioinformatics and Applied Biotechnology (IBAB), Bengaluru, India, <sup>2</sup>Department of Biotechnology, Indian Institute of Technology Kharagpur, Kharagpur, India, <sup>3</sup>Centre for Computational Biology, Indraprastha Institute of Information Technology Delhi (IIITD), Okhla Phase III, New Delhi, India

## 20.1 Introduction

The recent emergence of several pathogens and their associated diseases has made us realize the importance of developing proper treatment and preventive measures for infectious disease and how devastating disease could be without one. Amongst all medical interventions invented, vaccination stands tallest in preventing so many infectious diseases overall. Vaccines have been used over a long period as an efficient and cost-effective method to prevent and cure diseases. Any vaccine candidate's success can be implicated via its global impact on human/animal healthiness and life expectancy. It must be noted that the traditional vaccine designing for fast-evolving pathogens, for example, Influenza A virus, has not been entirely successful as they lack the broad spectrum and does not provide long-lasting protection. Conventional vaccine preparation methods include the use of living, killed, or attenuated strains of the pathogens (McVey & Shi, 2010; Unnikrishnan, Rappuoli, & Serruto, 2012). These methods usually involve culturing the coveted microorganisms at a large scale followed by effective inactivation and verification of the vaccine's ability to induce an immune response (Abd-Elghaffar, Rashed, Ali, & Amin, 2020). This process is expensive and usually takes around 5–15 years to complete. During the last three decades, with the evolution of next-generation sequencing (NGS) technologies along with the participation of sequence analysis, structural modeling, and machine learning, the field of vaccinology has become more interdisciplinary (Jorge & Dellagostin, 2017). With the availability of numerous bacterial, viral, archaeal, and eukaryotic genomes, identifying potential vaccine antigens (Ags) has become rapid, converging, and cheap. Structural biology-based advances in epitope mapping and modeling-based prediction have provided essential understandings of immunogenic Ags rational optimization. Machine learning, statistical approaches, and mathematical models have helped immensely in understanding the meaning of the high-throughput data followed by the identification of the best antigenic candidates from a large pool of candidates (He, Rappuoli, De Groot, & Chen, 2010). Therefore *in silico* techniques for vaccine design, being both cost- and time-effective, can substantially assist the vaccine development process. These revolutionary technologies allowed scientists to move from traditional vaccine-making strategies to computation-based rationally designed vaccines, evolving a new approach and terminology known as “reverse vaccinology” (RV) (Rappuoli, 2000).

The first step in developing any vaccine is the identification of an epitope, which is an immunogenic segment of the Ag to which the complementary antibody or T-cell receptors (TCRs) bind (Flower, Macdonald, Ramakrishnan, Davies, & Doytchinova, 2010). Designing vaccine constructs with multiple epitopes offers better immunogenicity over single epitope (Livingston et al., 2001). Epitope-based vaccines provide a remarkable advantage among the different methods used for vaccine development. This approach allows us to develop highly specific vaccines that do not trigger an undesirable immune response (Mahanty, Prigent, & Garraud, 2015). They are cost-effective and generate prolonged immunity in the host (Ahmad, Eweida, & Sheweita, 2016). Building a multiepitope vaccine involves the identification of suitable epitopes with the ability to induce desirable humoral and cellular immune responses. A humoral immunity, often referred as antibody-mediated immunity, is responsible for differentiating pathogenic Ags circulating inside the host's body. The cellular response (cell-mediated immunity) activates cytotoxic T-cell lymphocytes (CTLs) and triggers



**FIGURE 20.1** Broad overview of the reverse vaccinology-based vaccine development process.

the release of cytotoxins in response to abnormal MHC markers. Such markers usually originate from pathogen-infection or by the presence of tumor cells (Akinola et al., 2019).

RV takes advantage of the genomic information sequenced from infectious agents and employs computational tools to identify and examine immunogenic Ags (Rappuoli, 2000). RV provides an advantage over conventional vaccine development methods mainly because it is a nonculture-based approach to identify Ags in a pathogen, making it fast and inexpensive (Dalsass, Brozzi, Medini, & Rappuoli, 2019). Advancements in high-throughput sequencing technology and its accessibility at low-cost have led to the development of numerous web-based resources for *in silico* vaccine designing. This chapter enlists the computational tools used in the vaccine development process, starting with Ag selection methods, immunological databases, and RV pipelines. Furthermore, it describes the methods used for B-cell epitope (BCE) prediction (both linear and conformational), T-cell epitope (TCE) prediction (MHC class I and MHC class II binders), adjuvant and linker selection followed by computational tools used for structural analysis of the vaccine construct, toxicity test, peptide half-life, etc. In the end, we also converse the role of NGS in vaccine design and the success paradigms of computer-aided vaccines. In all sections, we have mentioned several reported web servers and databases. However, special emphasis has been given to the popular ones having more practical relevance. We have summarized our broad view of the vaccine development process in Fig. 20.1.

## 20.2 Antigen selection and immunological databases

An Ag is a foreign substance with the ability to induce an immune response in the body. These are usually parts of proteins derived from pathogens. Discrimination of self-Ags from the nonself-Ags and MHC molecules is achieved by the elimination of immature T cells with receptors that will potentially bind to self-Ags and MHC molecules (von Boehmer & Kisielow, 1990). This process occurs in the thymus gland and prevents such aberrant T cells from instigating an auto-immune response. Ags can broadly be classified into three groups:

### 20.2.1 Exogenous antigens

Extracellular Ags (e.g., fungi, viruses, bacteria) are known as exogenous Ags. These Ags are taken up by Ag-presenting cells (APCs), which include B lymphocytes (B cells) and phagocytic cells (macrophages, dendritic cells, etc.) by endocytosis.

### 20.2.2 Endogenous antigens

Ags that originate within the cell (intracellular) due to the degradation of the viral infection of cells, tumor cells, or defective Ags are known as endogenous Ags. They might be categorized into autologous, heterologous (xenogenic), and homologous (allogenic). These Ags are presented by MHC class I molecules (Rock, Reits, & Neefjes, 2016). They are processed by macrophage cells, followed by cytotoxic T cells.

### 20.2.3 Autoantigens

These are simple or complex molecules (such as nucleoproteins, nucleic acid, and carbohydrates), which under normal conditions do not work as an Ag; however, in case of specific autoimmune diseases, the immune system of those patients recognized them as Ags.

Ag selection is the first and essential step in epitope-based vaccine development. We must determine the specific conformational and structural characteristics of an Ag, sufficient to induce immune response without the risk of cytokine storm or tolerance of immune function. Multiple RV programs have been developed to address this issue. They can be used to predict Ags from the pathogen's genome/proteome. Based on the approach they use, RV programs can be classified into two types: classifying or filtering (Dalsass et al., 2019). Filtering programs are based on decision tree algorithms, whereas classifying programs are based on machine-learning algorithms, such as support vector machines (SVM), logistic regression, etc. VaxiJen (Doytchinova & Flower, 2007) was the first server to be developed following an alignment-free approach to predict protective Ags. This server is based on a machine-learning strategy and can be used to predict bacterial, viral, and tumor Ags. Vaxign (He, Xiang, & Mobley, 2010) is another RV tool used to predict and analyze vaccine targets based on genome sequences. It is a decision tree software. The Vaxign server includes two programs: (1) Vaxign query, which allows the user to explore precomputed Vaxign results, and (2) dynamic Vaxign analysis, which enables the user to dynamically execute Vaxign for a given input sequence and visualize the results.

Besides building various Ag prediction tools, several immunological databases have been generated in the last three decades. These databases contain different types of information related to the human immune system and pathogenic Ags. AntigenDB (Ansari, Flower, & Raghava, 2010) is one such repository that comprehensively maintains information about experimentally-validated pathogenic Ags. Each entry constitutes details, such as their chemical structure, protein sequence, and origin. Immune Epitope Database and Analysis Resource (IEDB) (Fleri et al., 2017; Vita et al., 2019) is one of the largest and freely available (funded by NIAID) databases containing information about B-cell assays, T-cell assays, peptidic as well as nonpeptidic epitopes, and MHC ligand assays along with source organisms of these epitopes in the context of diverse infectious diseases, autoimmunity, transplantation, and allergy. The first version of this database was established in 2011, and now it consists of approximately two million experiments gathered from >20,000 publications. Likewise, we have MHCBN 4.0, which maintains information about the MHC/TAP-binding peptides and TCEs (Lata, Bhasin, & Raghava, 2009), BciPep maintaining BCEs (Saha, Bhasin, & Raghava, 2005). Apart from generalized databases, scientists across the world have developed several resources, specific to certain diseases. MalVac (Chaudhuri, Ahmed, Ansari, Singh, & Ramachandran, 2008) is a database that maintains information related to Malaria vaccine candidates. Similarly, MycobacRV (Chaudhuri, Kulshreshtha, Raghunandan, & Ramachandran, 2014) is a database for Mycobacterial vaccine candidates, MtbWeb is a web portal that is specifically created for designing vaccines against mycobacterium tuberculosis and its various strains (Dhanda et al., 2016), and FungalRV (Chaudhuri, Ansari, Raghunandan, & Ramachandran, 2011) is a database for human fungal pathogens. Additional details about other immunological databases and Ag-predicting tools are provided in Table 20.1.

## 20.3 *In silico* method for B-cell epitope prediction

BCE is the antigenic part that will bind to the antibody or immunoglobulin (Ig). Most of these epitopes are proteinaceous in nature, and B cells recognize them. The conventional epitope identification methods include X-ray cocrystallography, tandem mass spectroscopy, enzyme-linked immunosorbent spot assay, etc. However, these techniques are expensive and time-consuming. Hence, computational tools to predict epitopes propose an excellent alternative. B cells have receptors known as Igs or antibodies. Prediction of BCE plays a vital role in the epitope vaccine development. BCE can be divided into two types: continuous (or linear) and discontinuous (or conformational) epitopes (Zhang, Zhao, Sun, Gao, & Ma, 2014). Over the years, several computational tools have been developed to predict BCE. These tools can be classified based on structured-based methods or sequence-based methods.

**TABLE 20.1** List of repositories or databases developed for maintaining antigenic protein/peptide vaccine candidates.

Name	Description	References
<b>Database maintaining immunogenic/antigenic protein/peptide vaccine candidate</b>		
AntigenDB 2.0	A database of Ags	Ansari et al. (2010)
Vaxijen 2.0	A tool for protective Ags and subunit vaccines prediction	Doytchinova and Flower (2007)
Epitome	Database of structurally inferred antigenic epitopes in proteins	Schlessinger, Ofran, Yachdav, and Rost (2006)
SYFPEITHI	A database of MHC class I and II ligands and peptide motifs	Rammensee, Bachmann, Emmerich, Bachor, and Stevanović (1999)
MHCBN 4.0	A curated database comprising information of MHC class I- and II-binding peptides, nonbinding peptides, and T-cell epitopes	Lata et al. (2009)
IPD-IMGT/HLA	The database provides information on human MHC sequences	Robinson et al. (2020)
SDAP	Structural database of allergenic proteins	Ivanciuc, Schein, and Braun (2003)
EpIC	Pipeline for epitope immunogenicity characterization	Marciniuk, Trost, and Napper (2015)
PRRDB	A database comprising of pattern recognition receptor	Lata and Raghava (2008)
PRRDB 2.0	An updated version of PRRDB	Kaur, Patiyl, Sharma, Usmani, and Raghava (2019)
IEDB	A repository maintaining Immune epitopes	Fleri et al. (2017)
Antijen	A database maintaining quantitative data of various immune epitopes	McSparron, Blythe, Zygouri, Doytchinova, and Flower (2003)
BciPep	A repository of B-cell epitopes	Saha et al. (2005)
<b>Organism-specific potential vaccine candidate database</b>		
MycobacRV	Database of mycobacterial vaccine candidates	Chaudhuri et al. (2014)
FungalRv	Immunoinformatic portal and adhesion prediction for fungal pathogens	Chaudhuri et al. (2011)
AgAbDb	A repository of antigen–antibody interactions	Kulkarni-Kale, Raskar-Renuse, Natekar-Kalantre, and Saxena (2014)
InnateDB	A knowledgebase resource of innate immunity interactions and pathways	Breuer et al. (2013)
MalVac	A repository of malarial vaccine candidates	Chaudhuri et al. (2008)
<b>Pipelines for developing reverse vaccinology vaccine candidates</b>		
Vacceed (RV)	<i>In silico</i> vaccine candidate for eukaryotic pathogens discovery pipeline	Goodswen, Kennedy, and Ellis (2014)
VacSol (RV)	Pipeline for reverse vaccinology. It screens out therapeutic targets, such as proteins/genes from the microbial genome	Rizwan et al. (2017)

Ags, Antigens; IEDB, Immune Epitope Database and Analysis Resource; RV, reverse vaccinology.

### 20.3.1 Prediction of conformational B-cell epitopes

Conformational BCE consists of patches of solvent-exposed atoms present proximally in the 3-D functional protein structure. However, it may or may not be continuous in the protein sequence. Conformational BCEs are more abundant (about 90%) as compared to the linear BCEs (Van Regenmortel, 2009). However, the length of the antigenic peptide, which contributes to the binding of the paratope site of the antibodies is still less understood. Multiple tools have been developed to address this issue, as indicated in Table 20.2. DiscoTope 2.0 (Kringelum, Lundegaard, Lund, & Nielsen, 2012) is one such tool that precisely predicts conformational epitopes by differentiating epitope and non-epitope residues based on amino acid statistics, spatial neighborhood, and surface availability. CBTOPE (Ansari & Raghava, 2010) is a SVM-based tool used to predict conformational BCEs. PEPOP 2.0 (Demolombe et al., 2019) is another such tool that predicts discontinuous epitopes by computing different combinations of amino acids using 34 different algorithms.

**TABLE 20.2** List of computational methods developed for predicting conformational B-cell epitopes.

Name	Description	References
DiscoTope 2.0	A structure-based antibody prediction tool	<a href="#">Kringelum et al. (2012)</a>
PEPOP 2.0	A tool for predicting noncontinuous epitopes	<a href="#">Demolombe et al. (2019)</a>
SEPPA 3.0	A tool for spatial epitope prediction for protein Ags, especially N-linked glycoproteins	<a href="#">Zhou et al. (2019)</a>
BEpro	Discontinuous B-cell epitope prediction tool	<a href="#">Sweredoski and Baldi (2008)</a>
CBTOPE	SVM-based method for predicting conformational B-cell epitope	<a href="#">Ansari and Raghava (2010)</a>
SEPla	A standalone method for predicting conformational B-cell epitopes	<a href="#">Dalkas and Rومان (2017)</a>
ElliPro	A tool for predicting B-cell epitopes using solvent accessibility and flexibility information	<a href="#">Ponomarenko et al. (2008)</a>

Ags, Antigens; SVM, support vector machine.

**TABLE 20.3** List of computational methods developed for predicting linear B-cell epitopes.

Name	Description	References
ABCPred	B-cell epitope prediction server based on Artificial Neural Network	<a href="#">Saha and Raghava (2006a)</a>
BcePred	B-cell epitope prediction server based on sequence physiochemical properties	<a href="#">Saha and Raghava (2004)</a>
SVMTriP	A tool for linear antigenic epitopes prediction	<a href="#">Yao et al. (2012)</a>
LBtope	A tool for linear B-cell epitopes prediction	<a href="#">Singh, Ansari, and Raghava (2013)</a>
BepiPred-2.0	A tool for sequential B-cell epitope prediction	<a href="#">Jespersen, Peters, Nielsen, and Marcatili (2017)</a>
IgPred	A tool for antibody-specific B-cell epitope prediction	<a href="#">Gupta, Ansari, Gautam, and Raghava (2013)</a>
LBEEP	A standalone for linear B-cell exact epitope prediction	<a href="#">Saravanan and Gautham (2015)</a>
COBEpro	A tool for linear B-cell epitope prediction	<a href="#">Sweredoski and Baldi (2009)</a>
IEDB	A tool for B-cell epitope prediction	<a href="#">Fleri et al. (2017)</a>
Epitopia	A web server for predicting B-cell epitope	<a href="#">Rubinstein, Mayrose, Martz, and Pupko (2009)</a>

IEDB, Immune Epitope Database and Analysis Resource.

### 20.3.2 Prediction of linear B-cell epitopes

Continuous BCEs account for nearly 10% of the reported BCEs. Continuous (or linear) BCE prediction is more challenging as compared to the discontinuous (or conformational) epitopes. This might be due to the variable length of linear BCEs, making them difficult to predict. ABCPred ([Saha & Raghava, 2006a](#)) is one of the early linear BCE prediction servers developed based on Recurrent Neural Network architectures. Other servers have also been created, such as BcePred ([Saha & Raghava, 2006a](#)) and SVMTriP ([Yao, Zhang, Liang, & Zhang, 2012](#)), which predict linear BCEs based on SVM algorithm. More details about BCE predicting methods are provided in [Table 20.3](#).

BCE prediction tools have been used by multiple research groups in the past to identify pathogenic epitopes. Baral et al. in their research work used BepiPred 2.0, BCPREDS, and BcePred servers independently to predict BCEs of Lassa Virus (LASV) based on its glycoprotein's amino acid sequence ([Baral, Pavadai, Gerstman, & Chapagain, 2020](#)). They also performed structure-based BCE prediction for LASV glycoprotein using ElliPro, Epitopia, and DiscoTope. The final potential BCE was selected by taking a consensus of all the hits from sequence- and structure-based tools.

Likewise, in another study, linear BCE was predicted using the BepiPred 1.0 tool for the HSP83.1 protein of *Leishmania braziliensis* revealing potential targets for the serodiagnosis of human tegumentary visceral leishmaniasis and canine visceral leishmaniasis (Menezes-Souza et al., 2015).

Similarly, another study used ABCPred server and BepiPred 1.0 server to predict linear BCEs of spike glycoprotein of infectious bronchitis virus (Ding et al., 2015). A recent research used IEDB and BCPRED to forecast the BCE in spike (S) protein of Middle East respiratory syndrome coronavirus (MERS-COV) (Tahir ul Qamar et al., 2019). In this study, VaxiJen 2.0 server was used to identify the antigenicity of the predicted epitopes. Linear BCE was predicted by the BepiPred 1.0 method, and discontinuous BCE was predicted using the DiscoTope server.

## 20.4 *In silico* method for T-cell epitope prediction

TCEs are peptide sequences that are attained from Ags and recognized by TCRs, when they are exhibited on the exterior surface of APCs. TCEs are presented by MHC class I and class II molecules, which are, respectively, recognized by two different types of T cells, that is, CD8 and CD4 cells. Extensive research has been going on for the past three decades for TCE prediction. TCEs can be predicted based on direct or indirect methods. Direct method includes analysis of structure and sequence of TCE (Stille, Thomas, Reyes, & Humphreys, 1987) whereas indirect methods include prediction of MHC binders. Due to high precision and specificity, MHC binder prediction tools are widely used rather than the use of TCE prediction tools. MHC binders can be classified into two types: (1) MHC class I binders (CD8 + T cells), which are responsible for endogenous Ag processing (Table 20.4), and (2) MHC class II binders (CD4 + T cells or helper T cells) for exogenous Ag processing (Table 20.5).

**TABLE 20.4** List of computational tools developed for predicting MHC class I binders and cytotoxic T-cell epitopes.

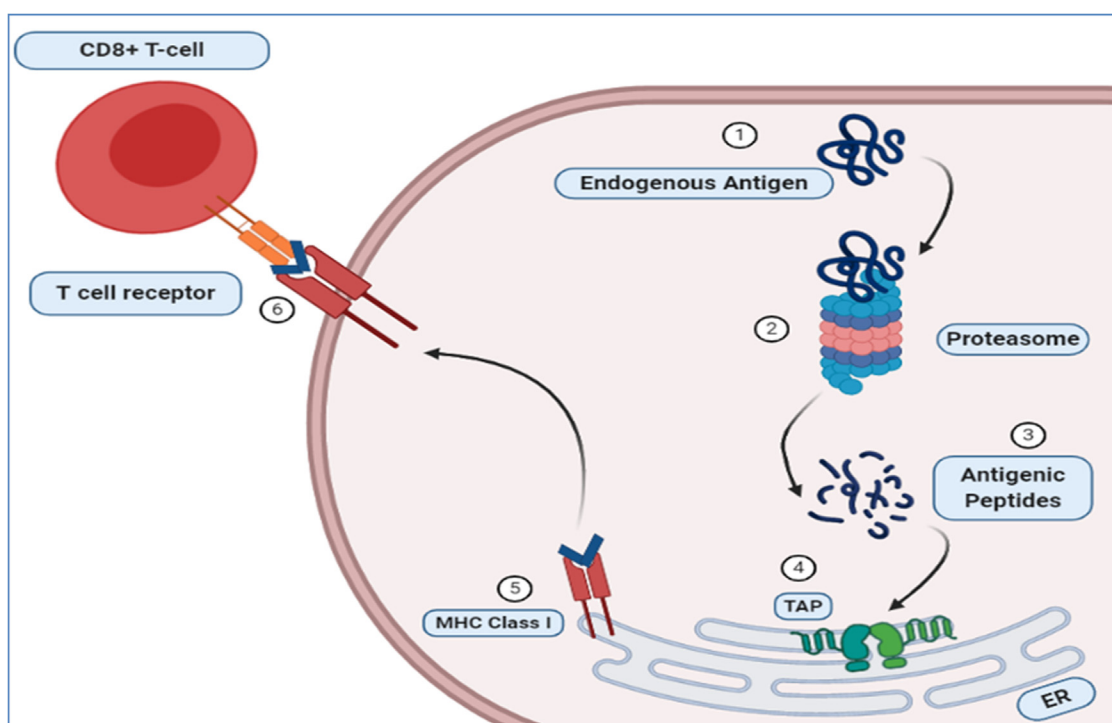
Name	Description	References
PCPS	A prediction server for proteasome cleavage	Gomez-Perosanz, Ras-Carmona, and Reche (2020)
Pcleavage	SVM-based tool for predicting proteasome cleavage	Bhasin and Raghava (2005)
NetChop 3.1	A neural network-based method for predicting proteasome cleavage sites	Nielsen et al. (2005)
TAPPred	A web server for predicting peptide-binding affinity toward TAP transporter	Bhasin and Raghava (2004a)
NetMHCcons 1.1	An ANN-based method for predicting MHC class I-binding peptides	Karosiene et al. (2012)
NetMHC	An ANN-based method for predicting MHC class I-binding peptides	Andreatta and Nielsen (2016)
NetMHCpan 4.1	A tool for predicting peptide binding to any MHC molecule of known sequence	Reynisson et al. (2020)
NetCTLpan 1.1	A tool for predicting CTL epitopes in a query sequence	Stranzl, Larsen, Lundegaard, and Nielsen (2010)
PickPocket 1.1	A tool for predicting peptide binding to any MHC molecule using position-specific weight matrices	Zhang et al. (2009)
nHLAPred	A tool for predicting MHC class I restricted T-cell epitopes	Bhasin and Raghava (2007)
CTLPred	SVM- and ANN-based method for predicting CTL epitopes in a query sequence	Bhasin and Raghava (2004b)
MMBPred	A web server for predicting mutated high-affinity MHC class I-binding peptides	Bhasin and Raghava (2003)
NetTepi 1.0	A method for predicting T-cell epitopes from protein sequence. It predicts for 13 different human MHC alleles	Trolle and Nielsen (2014)
ProPred 1.0	A tool for predicting MHC class I-binding regions in antigen	Singh and Raghava (2001)
TAPREG	SVM-based tool for predicting peptide affinity to TAP	Diez-Rivero, Chenlo, Zuluaga, and Reche (2010)
FRED2	A standalone for predicting T-cell epitopes	Schubert et al. (2016)
IEDB Consensus	A tool for predicting MHC class I-binding peptides	Moutaftsi et al. (2006)

CTL, Cytotoxic T-lymphocyte; IEDB, Immune Epitope Database and Analysis Resource; SVM, support vector machine.

**TABLE 20.5** List of computational tools developed for predicting MHC class II binders.

Name	Description	References
ProPred	A tool for predicting MHC class II-binding regions in antigen	Singh and Raghava (2001)
IEDB Consensus 2.22	A tool for predicting MHC class II-binding peptides	Wang et al. (2008)
NetMHCIIpan 3.1	A neural network-based method for predicting MHC class II-binding peptides	Andreatta et al. (2015)
HLA-Dr4Pred	SVM- and ANN-based method for predicting peptides interacting with HLA-DRB1*0401	Bhasin and Raghava (2004c)
EpiTOP3	QSAR-based method for predicting MHC class II-binding peptides	Dimitrov, Garnev, Flower, and Doytchinova (2010)
MHC2Pred	SVM-based method for the prediction of MHC class II binders	Lata, Bhasin, and Raghava (2007)
PREDIVAC	A method for MHC class II-binding prediction based on specificity determining residue (SDR) concept	Oyarzún et al. (2013)
IEDB CD4episcore	A method for predicting CD4 T-cell immunogenicity	Dhanda et al. (2018)

IEDB, Immune Epitope Database and Analysis Resource; SVM, support vector machine.



**FIGURE 20.2** Schematic representation of the endogenous antigen-processing pathway.

#### 20.4.1 MHC class I binder prediction

Endogenous or intracellular Ag processing involves multiple steps (Fig. 20.2). Initially, Ags are recognized and cleaved by the proteasome to small antigenic peptides. MAPPP (Hakenberg et al., 2003) and NetChop 3.1 (Keşmir, Nussbaum, Schild, Detours, & Brunak, 2002; Nielsen, Lundegaard, Lund, & Keşmir, 2005) are specialized web servers developed to predict proteasomal cleavage sites. The transportation of peptide residues to the endoplasmic reticulum (ER) is mediated by transporters associated with Ag-processing (TAP) protein complexes. The binding affinity (BA) of antigenic peptides toward TAP can be predicted using computational tools, such as TAPPred (Bhasin & Raghava, 2004a), which

is based on cascade SVM (C-SVM), with the use of amino acid features and their sequence. The antigenic peptides are then presented to cytotoxic T cells via MHC class I molecules. Multiple computational tools have been developed to predict MHC class I peptide binders. NetMHCcons 1.1 (Karosiene, Lundegaard, Lund, & Nielsen, 2012) is a web server used to predict peptide binding to known MHC class I molecules. It is an amalgam of three web-based tools to provide a precise prediction of MHC class I binders. NetMHC 4.0 is based on allele-specific Artificial Neural Network (ANN) method, which has been trained using 81 different human MHC alleles (NetMHCcons 1.1 server uses version 3.4) (Andreatta & Nielsen, 2016). NetMHCpan 4.1 is a web server that can be used to predict the binding of peptides to any known sequence of MHC molecules (Reynisson, Alvarez, Paul, Peters, & Nielsen, 2020). It is based on the pan-specific ANN method, which has been trained on more than 850,000 BA and Eluted Mass-Spectrometry Ligands peptides combined (NetMHCcons 1.1 server uses version 2.8). PickPocket 1.1 is an online server that uses position-specific weight matrices to predict the binding of peptides to any known MHC molecule (Zhang, Lund, & Nielsen, 2009). It has been trained on more than 150,000 quantitative binding data comprising more than 150 different MHC molecules.

Peptides bound to MHC molecules are presented to the cell surface, which is then recognized by CD8 + T cells. CTLpred (Bhasin & Raghava, 2004b) is a computational tool developed to predict the binding of MHC-bound peptides to CD8 + T cells. It is based on machine-learning techniques, such as SVM and ANN.

### 20.4.2 MHC class II binder prediction

In exogenous Ag processing, Ags are broken down into peptides by the *endo*-lysosomal system (Fig. 20.3). These peptides bind to MHC class II molecules, which are presented to CD4 + T helper cells. MHC class II-binding peptides can be predicted by several computational tools (Table 20.5). EpiDOCK (Atanasova, Patronov, Dimitrov, Flower, & Doytchinova, 2013) is the first structure-based MHC class II-binding prediction server. It employs homology modeling and molecular docking approaches to predict the BA of antigenic peptides to MHC class II molecules.

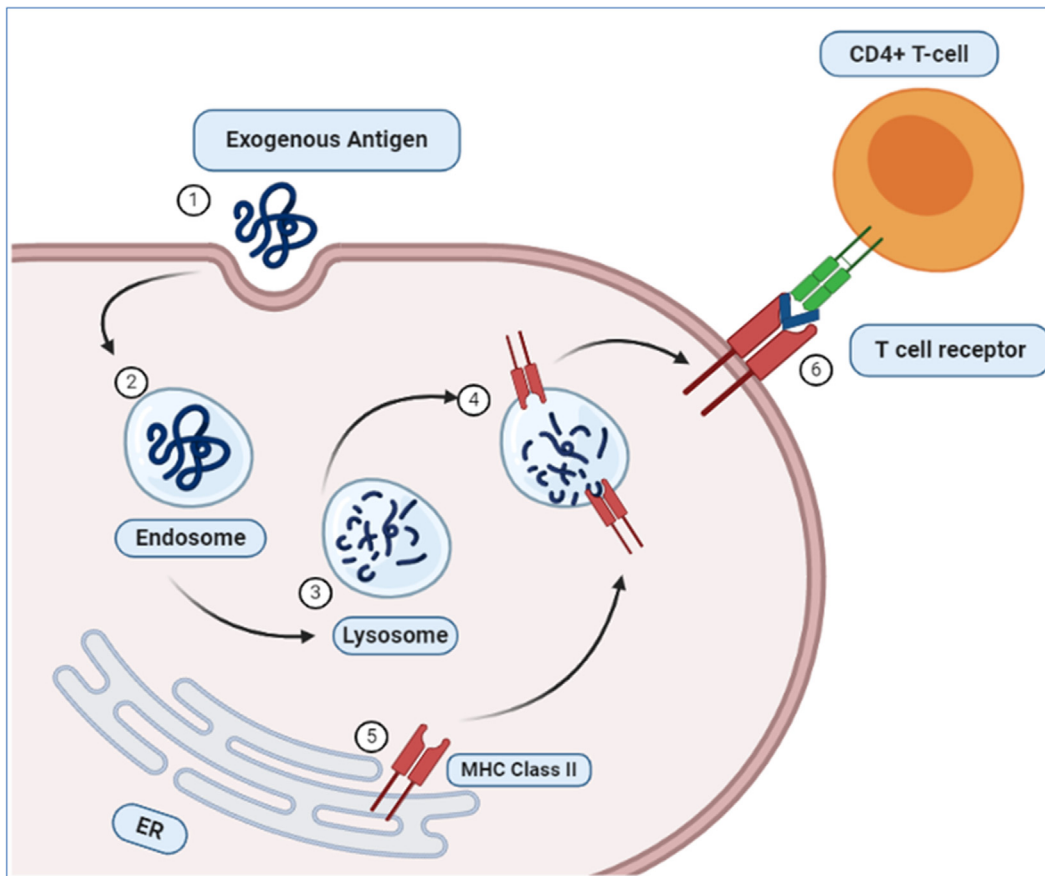


FIGURE 20.3 Schematic representation of exogenous antigen-processing pathway.

ProPred (Singh & Raghava, 2001) is another web-based tool that utilizes quantitative matrices to predict MHC class II-binding regions in an Ag sequence. NetMHCIIpan 3.1 (Andreatta et al., 2015) is a state-of-the-art quantitative method to predict the binding of peptides to any MHC class II molecule of known sequence in humans and mice. The computational tools mentioned above are based on indirect methods for predicting TCE. Recently direct methods have been developed to predict TCE. PREDIVAC (Oyarzún, Ellis, Bodén, & Kobe, 2013) comprehensively covers HLA class II alleles and predicts CD4 + Helper T cells with high precision.

The tools mentioned above have been very resourceful in the prediction and validation of TCEs and have been used extensively by various research groups in the past. This research (Baral et al., 2020) used ProPred-1, CTLPred, and NetCTL1.2 methods for the LASV glycoprotein sequence for MHC class I TCE prediction; however, in contrast, MHC class II TCE was predicted using ProPred, NetMHCII 2.3, and EpiTOP 3.0 methods. Another research (Simoneau et al., 2008) predicted the proteasomal cleavage sites in Tyrosine Phosphate SHP-1 using Pcleavage server. SHP-1 is expressed in the intestinal epithelial cells. They identified that the 208th residue tyrosine and 591st residue serine play a significant role in SHP-1 proteolysis (Simoneau et al., 2008). In another prominent work (Tahir ul Qamar et al., 2019), TCEs were predicted for the MERS-CoV spike protein using ProPred1 for MHC class I whereas ProPred tool was used for MHC class II binder. These studies show the significance of computational tools in predicting potential vaccine candidates.

### 20.4.3 MHC gene diversity and its importance in T-cell epitope prediction

Genes encoded by the MHC molecules are essential for adaptive immune responses in vertebrates. These genes encode receptors that recognize peptides obtained by intracellular and extracellular pathogen processing (Minias, Pikus, & Anderwald, 2019). Peptides (obtained by processing of intracellular proteins) of length 8–11 amino acids are presented by MHC class I for activating CD8 + CTLs. In contrast, MHC class II presents peptides (obtained by processing of extracellular proteins) of a more variable length to activate CD4 + T helper cells (Neeffjes, Jongtsma, Paul, & Bakke, 2011). However, all the presented peptides are not immunogenic, and the MHC molecules bind to a limited number of foreign peptides. The number of pathogens that can be recognized and killed depends upon the number of MHC molecules expressed within the individual organism. The diversity of the MHC alleles in an individual is primarily determined by the number of duplicate MHC loci (O'Connor, Strandh, Hasselquist, Nilsson, & Westerdahl, 2016). In humans, there are three functional MHC class I loci, responsible for presenting pathogenic peptides replicating in the cytoplasm to CD8 + T cells, and three to four functional MHC class II loci, responsible for presenting pathogenic peptides replicating in extracellular space to CD4 + T cells. According to databases, such as Immuno Polymorphism Database (IPD-IMGT/HLA), which is a part of the international ImmunoGeneTics project (IMGT), there are more than 17,000 MHC alleles in humans. However, as mentioned above, the number of functional loci (3–4 in number) allows only 12–14 alleles in an individual (fully heterozygous). One of the possible explanations for this is TCRs' depletion. TCRs play an essential role in identifying peptide-MHC complexes mostly via complementary determining region 3 (CDR3) loops (Kolev & Kemper, 2017; Zhong et al., 2013). TCR CDR3 regions are very diversified because of somatic V(D)J recombination or somatic hypermutation in rare cases. These diversified regions are also known as TCR repertoire and undergo positive/negative thymic selection. The positive selection allows retaining of the T cells interacting moderately with the MHC-self peptide complex, whereas negative selection leads to the deletion of T cells having strong avidity to MHC-self peptide complex. Hence, the relationship between the number of TCR repertoire and MHC genes is very crucial and requires a more detailed understanding for designing immunodominant T-cell epitopes.

## 20.5 Adjuvant and linker selection

Epitope-based vaccines usually require a few additional components in the vaccine construct to maximize the effectiveness of the vaccine. These additional components, known as adjuvants, are molecular complexes that enhance the immunogenicity of vaccine Ags. No matter how precise the prediction of BCEs and TCEs is, without a proper selection of adjuvants, the peptide-based vaccine may result in low immunogenicity. Vaxjo (Sayers, Ulysse, Xiang, & He, 2012) is a web-based central database maintaining information about vaccine adjuvants and their usage in vaccine development. VaxinPAD (Nagpal, Chaudhary, Agrawal, & Raghava, 2018) is a web-based tool used to predict immunomodulatory peptides (APC epitopes), which can potentially act as vaccine adjuvants.

Similarly, linkers are also an essential component of peptide-based vaccines. Linkers are also called “spacers” and are responsible for linking the three main components of an epitope-based vaccine, that is, TCEs, BCEs, and intramolecular adjuvants. Linkers are selected based on multiple properties, such as spacer length, hydrophobicity, peptide

arrangement, hydrophobicity, secondary structure, and possible interaction of the spacer with the rest of the vaccine construct. SynLinker (Liu, Chin, & Lee, 2015) and Linker (Craστο & Feng, 2000) are two such tools developed for linker selection. However, as per our prior search history, we found the websites of these tools nonfunctional.

## 20.6 Building 3D model and validation of fusion vaccine construct

Structures of antibodies and their targets have shown promising results in designing new vaccine candidates (Kanekiyo, Ellis, & King, 2019). The complex of antibody–target allows the approaches aimed at eliciting desired immune responses, in developing the immunogens. However, different pathogens use different strategies to evade immune responses. Therefore it is essential to understand how protective antibodies work. Ags might adopt multiple conformations throughout the pathogen life cycle. This could be due to its natural ability or because of the pathogen's response while evading the immune system. Therefore it is of utmost necessary to have a stabilized antigenic state, which could be addressed by the genetic fusion events. Genetic fusion of such Ags with the scaffolds of the multimeric proteins has offered stabilized quaternary stabilization. Structural knowledge could be advantageous in identifying the desired and undesired structural states of such Ags. This information could be further used for protein engineering of target proteins.

For *in silico* prediction of the structural stability of any vaccine constructed, it is necessary to build a 3D model of the vaccine sequence. One of the most prevalent approaches to create a 3D model is through homology modeling. Multiple computational tools are available to perform modeling of protein, such as EasyModeller 4.0 (Kuntal, Aparoy, & Reddanna, 2010), which is a recently developed GUI for MODELLER (Martí-Renom et al., 2000), SWISS-MODEL (Waterhouse et al., 2018), Phyre2 (Kelley, Mezulis, Yates, Wass, & Sternberg, 2015), and Rosetta (Lauck, Smith, Friedland, Humphris, & Kortemme, 2010). The structural qualities of the constructed vaccine should be verified using structural validation tools. Molecular docking analysis is usually performed to identify the BA of predicted vaccines with Toll-like receptors (TLRs) and MHC class I and II receptors. TLRs play a significant role in the identification of conserved pathogen-associated molecular patterns (PAMPs) activating the innate immune system. One of the most popular computational tools used for molecular docking is AutoDock 4.2 and AutoDock Vina (Morris et al., 2009; Trott & Olson, 2010). These tools provide several valuable information, such as docking score, RMSD value, and binding energy. MGL tools (<http://mgltools.scripps.edu/>) can be used to assist in structural analysis, docking studies, and implementing AutoDock tools. It can be used to visualize protein structures, prepare receptor protein, perform docking, etc. Molecular dynamic (MD) simulation is also a popular approach used to refine the 3D model of the vaccine construct. This approach is broadly used to predict the behavior of atoms. It also predicts structural stability and helps in understanding the orientational changes at an atomic scale. Commonly used computational tools available to perform MD simulations include NAMD (Phillips et al., 2005), GROMACS (GROningen MACHine for Chemical Simulations) (Abraham et al., 2015), and LAMMPS (Lykov, Li, Lei, Pivkin, & Karniadakis, 2015).

## 20.7 Miscellaneous properties

There are various known properties, such as toxicity, half-life, and delivery, which govern the suitability of a peptide being a potential vaccine candidate. These properties, as well as the resources developed regarding this, are discussed below (Table 20.6).

### 20.7.1 Peptide toxicity

One of the main challenges associated with peptide-based vaccines is their toxicity, which could be immune toxicity, cytotoxicity, or hemolysis. Peptides are among the most important immune response inducers; however, their toxicity limits their uses. Several tools and resources have been developed to address this very issue. For example, ToxinPred (Gupta, Kapoor et al., 2013) is an *in silico* method that predicts peptide/protein toxicity. It can also be used to design a peptide with the least toxicity and determine toxic regions in a peptide. Apart from peptide toxicity, it is necessary to predict whether the peptide is allergic to the body or not. Researchers have developed several tools that address the issue of allergenicity of the proteins/peptides, for example, AlgPred (Saha & Raghava, 2006b), AlgPred 2.0 (Sharma et al., 2020), and AllerTOP version 2.0 (Dimitrov, Bangov, Flower, & Doytchinova, 2014). Other databases, such as Hemolytik (Gautam et al., 2014), comprise peptides that are hemolytic in nature. Likewise, HemoPI database predicts the hemolysis property of a given peptide or number of peptides (Chaudhary et al., 2016). AHTPDB is a repository that comprises peptides that are antihypertensive in nature (Kumar, Chaudhary, Sharma et al., 2015). AHTpin

**TABLE 20.6** List of miscellaneous tools and resources for vaccine design.

Name	Description	References
VaxinPAD	SVM-based method for designing A-cell epitopes	Nagpal et al. (2018)
Vaxjo	A method for analyzing vaccine adjuvants	Sayers et al. (2012)
Vaxgen	A database comprising of vaccine-related pathogens and host genes and proteins	Xiang et al. (2008)
Protegen	A database comprising information related to protective Ags	Yang, Sayers, Xiang, and He (2011)
VirmugenDB	A repository of virulent genes used for developing live attenuated vaccines	Racz, Chung, Xiang, and He (2013)
DNAVaxDB	A database of DNA vaccines	Racz, Li, Patel, Xiang, and He (2014)
Vaxvec	A database comprising information related to recombinant vaccine vectors and vaccines	Deng et al. (2015)
ProInflam	A method for predicting proinflammatory peptides	Gupta, Madhu, Sharma, and Sharma (2016)
ToxinPred	SVM-based method for predicting the toxicity of peptides	Gupta, Kapoor et al. (2013)
AlgPred	A method for mapping IgE epitopes and predicting allergenic proteins	Saha and Raghava (2006b)
AlgPred 2.0	An updated version of AlgPred	Sharma et al. (2020)
AntiCP 2.0	A tool for predicting anticancer peptides	Agrawal, Bhagat, Mahalwal, Sharma, and Raghava (2020)
SATPdb	A database of structurally annotated peptides	Singh et al. (2016)
CPPsite 2.0	A database of natural and chemically modified cell-penetrating peptides	Agrawal et al. (2016)
IL4pred	A method for predicting interleukin-4 (IL-4)-inducing peptides	Dhanda, Gupta, Vir, and Raghava (2013)
IL6pred	A method for predicting IL-6-inducing peptides	Dhall, Patiyal, Sharma, Usmani, and Raghava (2020)
IL-10Pred	A method for predicting IL-10-inducing peptides	Nagpal et al. (2017)
IL17escan	A method for identification of IL-17-inducing peptides	Gupta, Mittal, Madhu, and Sharma (2017)
IFNepitope	A method for predicting interferon- $\gamma$ -inducing peptides	Dhanda, Vir, and Raghava (2013)

Ags, Antigens; SVM, support vector machine.

(Kumar, Chaudhary, Singh Chauhan et al., 2015) and PAAP (Win, Schaduangrat, Prachayasittikul, Nantasenamat, & Shoombuatong, 2018) are the *in silico* methods for predicting the antihypertensive nature of the peptides.

### 20.7.2 Half-life or stability

Another essential consideration for a good peptide-based vaccine is its stability in the body fluid or its “half-life.” Various chemical modifications, amino acid substitution with D-form, cyclization, etc., have been done to increase the half-life of therapeutic peptides. PEPLife is one such database that maintains the information of peptide half-life along with chemical modifications, and sequence information (Mathur et al., 2016). PlifePred (Mathur, Singh, Mehta, Agrawal, & Raghava, 2018) is an *in silico* method for predicting the half-life of a natural or chemically modified peptide.

### 20.7.3 Delivery methodology

Intracellular delivery of vaccine candidates to their suitable target or location is one of the critical and challenging tasks faced by the researchers. Most of these vaccine candidates get ruled out in the final stage due to their low bioavailability or poor delivery. Several methods have been developed in the past to address this issue; however, cell-penetrating

peptides (CPPs) have shown a very promising role in drug delivery. CPPs are short cationic peptides (10–30 residues), water-soluble, less toxic, and amphipathic in nature. CPPsite 2.0 database (Agrawal et al., 2016) provides a large amount of information on such peptides. It not only stores information about natural CPPs but also chemically modified CPPs. Besides these databases, machine-learning-based methods have been developed, which can predict as well as design novel CPPs. Some of the widely used methods are CellPPD (Gautam et al., 2013), CellPPDMod (Kumar et al., 2018), and CPPred-RF (Wei et al., 2017).

## 20.8 Role of next-generation sequencing technology in vaccine design

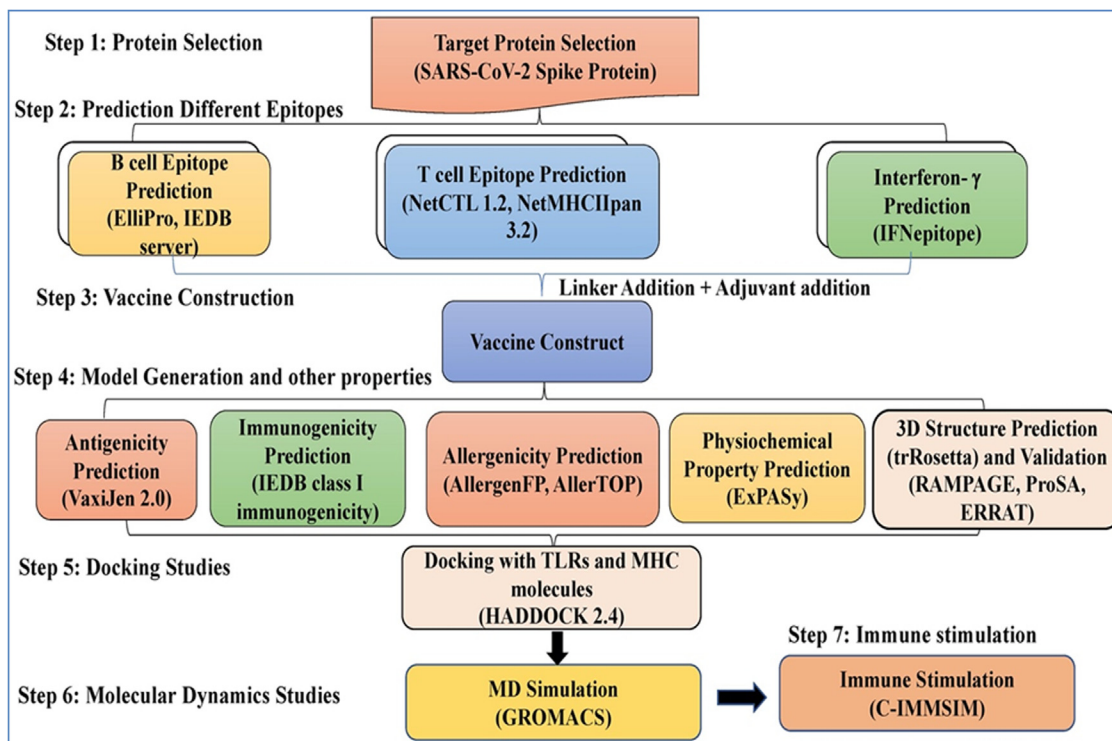
Currently, there are many vaccines available in the market against infection caused by influenza, diphtheria, pneumococcus, measles, mumps, rubella, etc.; however, there are several other diseases against which we do not have any relevant vaccines (Rappuoli, Miller, & Falkow, 2002). One of the main reasons that we do not have vaccines against these pathogenic organisms is the variability in their genomes across various regions during the pandemic. Also, we do not have a proper understanding of the mechanism by which these organisms evade host immune responses. Among these pathogenic organisms, RNA viruses form a significant class that is of major concern. These viruses are the common pathogens of humans and animals. Some of the examples include coronavirus, hepatitis, HIV-1, influenza A H1N1, and the dengue virus.

Since the sequencing of the complete human genome in 2001, NGS technologies have revolutionized biological research in humans as well as microbes. It has allowed the generation of large sequencing datasets at affordable cost and time. Some of the key advantages of NGS are the generation of a large volume of transcriptomics data (RNA-seq), histone modifications data (NGS methylation), DNA–protein interaction data (CHIP-Seq), etc. Researchers are utilizing these datasets for vaccine development. NGS has allowed analyzing many genomes in a brief period. With the use of different types of omics data, it has been easy to analyze the variability in the genome, gene expression changes, methylation level, host–pathogen interaction, pathway enrichment after the pathogen infection, etc. The final goal is to decipher the mechanisms by which pathogens protect themselves from the host immune response. A detailed description of various NGS technologies and their association with vaccine development has been well-documented in this review (Luciani, Bull, & Lloyd, 2012).

Utilizing these evolved technologies, several *in silico* tools have been developed over the years by scientists. Some of the important resources developed in this field include various genomics sequencing projects, such as TCGA (Weinstein et al., 2013), ENCODE (Davis et al., 2018), ICGC (Zhang et al., 2019), and NCBI. TCGA and ICGC provide the different types of omics data (transcriptomics, genomics, epigenomics, etc.) of nearly 33 cancer tissue types. ENCODE database offers insights about the functional elements in the human and mouse genomes. Other than these databases, we have several genome sequencing projects against various pathogenic organisms, such as NCBI viral genome projects (Aw et al., 2014; Bao et al., 2004; Imai et al., 2018; Pillay et al., 2020), bacterial sequencing projects like the Integrated Microbial Genome database (Markowitz et al., 2012), RefSeq Microbial Genomes database (Tatusova, Ciufu, Fedorov, O’Neill, & Tolstoy, 2014), and TIGR Microbial database (Peterson, Umayam, Dickinson, Hickey, & White, 2001). These resources could be beneficial in gaining insight for vaccine designing and addressing the major challenges faced by scientists earlier.

## 20.9 Computer-aided vaccine development example

Here, we will briefly discuss this latest research (Kar et al., 2020) where they have utilized diverse computational methods (explained here) to develop a potential multiepitope vaccine against SARS-CoV-2 utilizing its tail spike protein. Fig. 20.4 shows the stepwise progression used for designing the potential vaccine candidate. In this study, the authors predicted the multiepitope vaccine candidates using viral spike glycoprotein. In general, the target proteins vary in the case of different diseases caused by an organism and is based on their pathogenicity. The spike protein was screened for potential TCEs (MHC class I and II binders), BCEs, and epitopes that can induce the activation of interferon- $\gamma$ . These epitopes were joined using linkers and an adjuvant was also attached to boost the immune response. The 3D structure of the final vaccine was constructed and validated using various *in silico* tools. Once the model was generated, other properties, such as immunogenicity, allergenicity, and physiochemical properties, were measured. Docking studies with various TLRs and MHC molecules were performed, and BA was computed. This was followed by reverse translation and codon optimization study. Next, MD simulation studies showed that the vaccine construct was stable. Lastly, immune stimulation studies validate the expression of the vaccine in the bacterial host and the ability of the vaccine to



**FIGURE 20.4** Flowchart describing the *in silico* methods used for constructing potential vaccine candidate for SARS-CoV-2. For every process, the name of the tools used has been mentioned in the brackets (step 2 onward).

stimulate an immune response. Fig. 20.4 provides a schematic representation of the various processes used in this study for the multiepitope vaccine construct. It also elucidates the various computational tools used for each step used by the authors in this study. This strategy and tools could be useful for researcher working in the vaccine design field.

## 20.10 Conclusion

In the 20th century modern world, life expectancy and mortality rate in any country are correlated with the availability of vaccines for diverse infectious diseases to its resident population. Traditional approaches used for vaccine development has been expensive, laborious, and time-consuming; moreover, for antigenically diverse pathogens, it is almost impossible. Revolution in the NGS technologies, recombinant DNA technology, and structural modeling have exceedingly catalyzed the vaccine development process, starting from epitope selection to the vaccine validation process. Nowadays, the genomic, transcriptomic, proteomic, and metabolomic exploration of any putative pathogen allow generating innovative vaccines for age-long diseases associated with those pathogens. Novel ways of systematic studies of diverse physio-chemo-genomic-biological aspects are enabling the researchers to correlate, extrapolate, and overlap multiple research aspects altogether to develop a safe, fast, and accurate vaccine. Immunoinformatics, along with computational and mathematical methods, are providing valuable information on the determination of immunogenicity of all putative Ags and molecular interactions during Ag presentation and processing. With the availability of numerous databases, prediction web servers, advanced algorithms, and powerful computational machines to analyze the high-throughput data, vaccine development soon will become more accurate, reliable, safe, and effective.

## Acknowledgments

GS is supported by funding from the Department of Science and Technology-INSPIRE (DST-INSPIRE) program, Government of India. This work was partially supported by the Department of Electronics, IT, BT, and S&T of the Government of Karnataka, India. PA would like to thank Indraprastha Institute of Information Technology, Delhi (IIIT-Delhi), for the support. The views expressed in this chapter are those of the authors, and not necessarily those of either funding agency or any other institution.

## Conflict of interest

The authors declare no conflict of interest to disclose.

## References

- Abd-Elghaffar, A. A., Rashed, M. E., Ali, A. E., & Amin, M. A. (2020). In-vitro inactivation of sabin-polioviruses for development of safe and effective polio vaccine. *Vaccines*, *8*, 601. Available from <https://doi.org/10.3390/vaccines8040601>.
- Abraham, M. J., Murtola, T., Schulz, R., Páll, S., Smith, J. C., Hess, B., & Lindahl, E. (2015). GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX*, *1–2*, 19–25. Available from <https://doi.org/10.1016/j.softx.2015.06.001>.
- Agrawal, P., Bhagat, D., Mahalwal, M., Sharma, N., & Raghava, G. P. S. (2020). AntiCP 2.0: An updated model for predicting anticancer peptides. *Briefings in Bioinformatics*. Available from <https://doi.org/10.1093/bib/bbaa153>.
- Agrawal, P., Bhalla, S., Usmani, S. S., Singh, S., Chaudhary, K., Raghava, G. P. S., & Gautam, A. (2016). CPPsite 2.0: A repository of experimentally validated cell-penetrating peptides. *Nucleic Acids Research*, *44*, D1098–D1103. Available from <https://doi.org/10.1093/nar/gkv1266>.
- Ahmad, T. A., Eweida, A. E., & Sheweita, S. A. (2016). B-cell epitope mapping for the design of vaccines and effective diagnostics. *Trials in Vaccinology*, *5*, 71–83. Available from <https://doi.org/10.1016/j.trivac.2016.04.003>.
- Akinola, F., Muloiwa, R., Hussey, G. D., Dirix, V., Kagina, B., & Amponsah-Dacosta, E. (2019). Assessment of humoral and cell-mediated immune responses to pertussis vaccination: A systematic review protocol. *BMJ Open*. Available from <https://doi.org/10.1136/bmjopen-2018-028109>.
- Andreatta, M., Karosiene, E., Rasmussen, M., Stryhn, A., Buus, S., & Nielsen, M. (2015). Accurate pan-specific prediction of peptide-MHC class II binding affinity with improved binding core identification. *Immunogenetics*, *67*, 641–650. Available from <https://doi.org/10.1007/s00251-015-0873-y>.
- Andreatta, M., & Nielsen, M. (2016). Gapped sequence alignment using artificial neural networks: Application to the MHC class I system. *Bioinformatics (Oxford, England)*, *32*, 511–517. Available from <https://doi.org/10.1093/bioinformatics/btv639>.
- Ansari, H. R., Flower, D. R., & Raghava, G. P. S. (2010). AntigenDB: An immunoinformatics database of pathogen antigens. *Nucleic Acids Research*, *38*, D847–D853. Available from <https://doi.org/10.1093/nar/gkp830>.
- Ansari, H. R., & Raghava, G. P. S. (2010). Identification of conformational B-cell epitopes in an antigen from its primary sequence. *Immunome Research*, *6*, 6. Available from <https://doi.org/10.1186/1745-7580-6-6>.
- Atanasova, M., Patronov, A., Dimitrov, I., Flower, D. R., & Doytchinova, I. (2013). EpiDOCK: A molecular docking-based tool for MHC class II binding prediction. *Protein Engineering, Design & Selection*, *26*, 631–634. Available from <https://doi.org/10.1093/protein/gzt018>.
- Aw, P. P. K., de Sessions, P. F., Wilm, A., Hoang, L. T., Nagarajan, N., Sessions, O. M., & Hibberd, M. L. (2014). Next-generation whole genome sequencing of dengue virus. *Methods in Molecular Biology*, *1138*, 175–195. Available from [https://doi.org/10.1007/978-1-4939-0348-1\\_12](https://doi.org/10.1007/978-1-4939-0348-1_12).
- Bao, Y., Federhen, S., Leippe, D., Pham, V., Resenchuk, S., Rozanov, M., . . . Tatusova, T. (2004). National center for biotechnology information viral genomes project. *Journal of Virology*, *78*, 7291–7298. Available from <https://doi.org/10.1128/JVI.78.14.7291-7298.2004>.
- Baral, P., Pavada, E., Gerstman, B. S., & Chapagain, P. P. (2020). In-silico identification of the vaccine candidate epitopes against the Lassa virus hemorrhagic fever. *Scientific Reports*, *10*, 7667. Available from <https://doi.org/10.1038/s41598-020-63640-1>.
- Bhasin, M., & Raghava, G. P. S. (2003). Prediction of promiscuous and high-affinity mutated MHC binders. *Hybridoma and Hybridomics*, *22*, 229–234. Available from <https://doi.org/10.1089/153685903322328956>.
- Bhasin, M., & Raghava, G. P. S. (2004a). Analysis and prediction of affinity of TAP binding peptides using cascade SVM. *Protein Science: A Publication of the Protein Society*, *13*, 596–607. Available from <https://doi.org/10.1110/ps.03373104>.
- Bhasin, M., & Raghava, G. P. S. (2004b). Prediction of CTL epitopes using QM, SVM and ANN techniques. *Vaccine*, *22*, 3195–3204. Available from <https://doi.org/10.1016/j.vaccine.2004.02.005>.
- Bhasin, M., & Raghava, G. P. S. (2004c). SVM based method for predicting HLA-DRB1\*0401 binding peptides in an antigen sequence. *Bioinformatics (Oxford, England)*, *20*, 421–423. Available from <https://doi.org/10.1093/bioinformatics/btg424>.
- Bhasin, M., & Raghava, G. P. S. (2005). Pcleavage: An SVM based method for prediction of constitutive proteasome and immunoproteasome cleavage sites in antigenic sequences. *Nucleic Acids Research*, *33*, W202–W207. Available from <https://doi.org/10.1093/nar/gki587>.
- Bhasin, M., & Raghava, G. P. S. (2007). A hybrid approach for predicting promiscuous MHC class I restricted T cell epitopes. *Journal of Biosciences*, *32*, 31–42. Available from <https://doi.org/10.1007/s12038-007-0004-5>.
- Breuer, K., Foroushani, A. K., Laird, M. R., Chen, C., Sribnaia, A., Lo, R., . . . Lynn, D. J. (2013). InnateDB: Systems biology of innate immunity and beyond—recent updates and continuing curation. *Nucleic Acids Research*, *41*, D1228–D1233. Available from <https://doi.org/10.1093/nar/gks1147>.
- Chaudhary, K., Kumar, R., Singh, S., Tuknait, A., Gautam, A., Mathur, D., . . . Raghava, G. P. S. (2016). A web server and mobile app for computing hemolytic potency of peptides. *Scientific Reports*, *6*, 22843. Available from <https://doi.org/10.1038/srep22843>.
- Chaudhuri, R., Ahmed, S., Ansari, F. A., Singh, H. V., & Ramachandran, S. (2008). MalVac: Database of malarial vaccine candidates. *Malaria Journal*, *7*, 184. Available from <https://doi.org/10.1186/1475-2875-7-184>.
- Chaudhuri, R., Ansari, F. A., Raghunandan, M. V., & Ramachandran, S. (2011). FungalRV: Adhesin prediction and immunoinformatics portal for human fungal pathogens. *BMC Genomics*, *12*, 192. Available from <https://doi.org/10.1186/1471-2164-12-192>.
- Chaudhuri, R., Kulshreshtha, D., Raghunandan, M. V., & Ramachandran, S. (2014). Integrative immunoinformatics for Mycobacterial diseases in R platform. *Systems and Synthetic Biology*, *8*, 27–39. Available from <https://doi.org/10.1007/s11693-014-9135-9>.

- Craeto, C. J., & Feng, J. (2000). LINKER: A program to generate linker sequences for fusion proteins. *Protein Engineering, Design & Selection*, 13, 309–312. Available from <https://doi.org/10.1093/protein/13.5.309>.
- Dalkas, G. A., & Rooman, M. (2017). SEPIa, a knowledge-driven algorithm for predicting conformational B-cell epitopes from the amino acid sequence. *BMC Bioinformatics*, 18, 95. Available from <https://doi.org/10.1186/s12859-017-1528-9>.
- Dalsass, M., Brozzi, A., Medini, D., & Rappuoli, R. (2019). Comparison of open-source reverse vaccinology programs for bacterial vaccine antigen discovery. *Frontiers in Immunology*, 10, 113. Available from <https://doi.org/10.3389/fimmu.2019.00113>.
- Davis, C. A., Hitz, B. C., Sloan, C. A., Chan, E. T., Davidson, J. M., Gabdank, I., ... Cherry, J. M. (2018). The Encyclopedia of DNA elements (ENCODE): Data portal update. *Nucleic Acids Research*, 46, D794–D801. Available from <https://doi.org/10.1093/nar/gkx1081>.
- Demolombe, V., de Brevern, A. G., Felicori, L., NGuyen, C., Machado de Avila, R. A., Valera, L., ... Moreau, V. (2019). PEPOP 2.0: New approaches to mimic non-continuous epitopes. *BMC Bioinformatics*, 20, 387. Available from <https://doi.org/10.1186/s12859-019-2867-5>.
- Deng, S., Martin, C., Patil, R., Zhu, F., Zhao, B., Xiang, Z., & He, Y. (2015). Vaxvec: The first web-based recombinant vaccine vector database and its data analysis. *Vaccine*, 33, 6938–6946. Available from <https://doi.org/10.1016/j.vaccine.2015.07.113>.
- Dhall, A., Patiyal, S., Sharma, N., Usmani, S. S., & Raghava, G. P. S. (2020). Computer-aided prediction and design of IL-6 inducing peptides: IL-6 plays a crucial role in COVID-19. *Briefings in Bioinformatics*. Available from <https://doi.org/10.1093/bib/bbaa259>.
- Dhanda, S. K., Gupta, S., Vir, P., & Raghava, G. P. S. (2013). Prediction of IL4 inducing peptides. *Clinical & Developmental Immunology*, 2013, 263952. Available from <https://doi.org/10.1155/2013/263952>.
- Dhanda, S. K., Karosiene, E., Edwards, L., Grifoni, A., Paul, S., Andreatta, M., ... Sette, A. (2018). Predicting HLA CD4 immunogenicity in human populations. *Frontiers in Immunology*, 9, 1369. Available from <https://doi.org/10.3389/fimmu.2018.01369>.
- Dhanda, S. K., Vir, P., & Raghava, G. P. S. (2013). Designing of interferon-gamma inducing MHC class-II binders. *Biology Direct*, 8, 30. Available from <https://doi.org/10.1186/1745-6150-8-30>.
- Dhanda, S. K., Vir, P., Singla, D., Gupta, S., Kumar, S., & Raghava, G. P. S. (2016). A web-based platform for designing vaccines against existing and emerging strains of Mycobacterium tuberculosis. *PLoS One*, 11, e0153771. Available from <https://doi.org/10.1371/journal.pone.0153771>.
- Diez-Rivero, C. M., Chenlo, B., Zuluaga, P., & Reche, P. A. (2010). Quantitative modeling of peptide binding to TAP using support vector machine. *Proteins*, 78, 63–72. Available from <https://doi.org/10.1002/prot.22535>.
- Dimitrov, I., Bangov, I., Flower, D. R., & Doytchinova, I. (2014). AllerTOP v.2—a server for in silico prediction of allergens. *Journal of Molecular Modeling*, 20, 2278. Available from <https://doi.org/10.1007/s00894-014-2278-5>.
- Dimitrov, I., Garnev, F., Flower, D. R., & Doytchinova, I. (2010). EpiTOP—a proteochemometric tool for MHC class II binding prediction. *Bioinformatics (Oxford, England)*, 26, 2066–2068. Available from <https://doi.org/10.1093/bioinformatics/btq324>.
- Ding, M., Wang, H., Cao, H., Fan, W., Ma, B., Xu, P., ... Yang, X. (2015). Development of a multi-epitope antigen of S protein-based ELISA for antibodies detection against infectious bronchitis virus. *Bioscience, Biotechnology, and Biochemistry*, 79, 1287–1295. Available from <https://doi.org/10.1080/09168451.2015.1025692>.
- Doytchinova, I. A., & Flower, D. R. (2007). VaxiJen: A server for prediction of protective antigens, tumour antigens and subunit vaccines. *BMC Bioinformatics*, 8, 4. Available from <https://doi.org/10.1186/1471-2105-8-4>.
- Fleri, W., Paul, S., Dhanda, S. K., Mahajan, S., Xu, X., Peters, B., & Sette, A. (2017). The immune epitope database and analysis resource in epitope discovery and synthetic vaccine design. *Frontiers in Immunology*, 8, 278. Available from <https://doi.org/10.3389/fimmu.2017.00278>.
- Flower, D. R., Macdonald, I. K., Ramakrishnan, K., Davies, M. N., & Doytchinova, I. A. (2010). Computer aided selection of candidate vaccine antigens. *Immunome Research*, 6, S1. Available from <https://doi.org/10.1186/1745-7580-6-S2-S1>.
- Gautam, A., Chaudhary, K., Kumar, R., Sharma, A., Kapoor, P., Tyagi, A., & Raghava, G. P. S. (2013). In silico approaches for designing highly effective cell penetrating peptides. *Journal of Translational Medicine*, 11, 74. Available from <https://doi.org/10.1186/1479-5876-11-74>.
- Gautam, A., Chaudhary, K., Singh, S., Joshi, A., Anand, P., Tuknait, A., ... Raghava, G. P. S. (2014). Hemolytik: A database of experimentally determined hemolytic and non-hemolytic peptides. *Nucleic Acids Research*, 42, D444–D449. Available from <https://doi.org/10.1093/nar/gkt1008>.
- Gomez-Perosanz, M., Ras-Carmona, A., & Reche, P. A. (2020). PCPS: A web server to predict proteasomal cleavage sites. *Methods in Molecular Biology*, 2131, 399–406. Available from [https://doi.org/10.1007/978-1-0716-0389-5\\_23](https://doi.org/10.1007/978-1-0716-0389-5_23).
- Goodswen, S. J., Kennedy, P. J., & Ellis, J. T. (2014). Vacceed: A high-throughput in silico vaccine candidate discovery pipeline for eukaryotic pathogens based on reverse vaccinology. *Bioinformatics (Oxford, England)*, 30, 2381–2383. Available from <https://doi.org/10.1093/bioinformatics/btu300>.
- Gupta, S., Ansari, H. R., Gautam, A., & Raghava, G. P. S. (2013). Identification of B-cell epitopes in an antigen for inducing specific class of antibodies. *Biology Direct*, 8, 27. Available from <https://doi.org/10.1186/1745-6150-8-27>.
- Gupta, S., Kapoor, P., Chaudhary, K., Gautam, A., Kumar, R., Consortium, O. S. D. D., & Raghava, G. P. S. (2013). In silico approach for predicting toxicity of peptides and proteins. *PLoS One*, 8, e73957.
- Gupta, S., Madhu, M. K., Sharma, A. K., & Sharma, V. K. (2016). ProInflam: A webserver for the prediction of proinflammatory antigenicity of peptides and proteins. *Journal of Translational Medicine*, 14, 178. Available from <https://doi.org/10.1186/s12967-016-0928-3>.
- Gupta, S., Mittal, P., Madhu, M. K., & Sharma, V. K. (2017). IL17eScan: A tool for the identification of peptides inducing IL-17 response. *Frontiers in Immunology*, 8, 1430. Available from <https://doi.org/10.3389/fimmu.2017.01430>.
- Hakenberg, J., Nussbaum, A. K., Schild, H., Rammensee, H.-G., Kuttler, C., Holzhütter, H.-G., ... Mollenkopf, H.-J. (2003). MAPPP: MHC class I antigenic peptide processing prediction. *Applied Bioinformatics*, 2, 155–158.
- He, Y., Rappuoli, R., De Groot, A. S., & Chen, R. T. (2010). Emerging vaccine informatics. *Journal of Biomedicine & Biotechnology*, 2010, 218590. Available from <https://doi.org/10.1155/2010/218590>.

- He, Y., Xiang, Z., & Mobley, H. L. T. (2010). Vaxign: The first web-based vaccine design program for reverse vaccinology and applications for vaccine development. *Journal of Biomedicine & Biotechnology*, 2010, 297505. Available from <https://doi.org/10.1155/2010/297505>.
- Imai, K., Tamura, K., Tanigaki, T., Takizawa, M., Nakayama, E., Taniguchi, T., . . . Maeda, T. (2018). Whole genome sequencing of influenza A and B viruses with the MinION sequencer in the clinical setting: A pilot study. *Frontiers in Microbiology*, 9, 2748. Available from <https://doi.org/10.3389/fmicb.2018.02748>.
- Ivanciuc, O., Schein, C. H., & Braun, W. (2003). SDAP: Database and computational tools for allergenic proteins. *Nucleic Acids Research*, 31, 359–362. Available from <https://doi.org/10.1093/nar/gkg010>.
- Jespersen, M. C., Peters, B., Nielsen, M., & Marcotilli, P. (2017). BepiPred-2.0: Improving sequence-based B-cell epitope prediction using conformational epitopes. *Nucleic Acids Research*, 45, W24–W29. Available from <https://doi.org/10.1093/nar/gkx346>.
- Jorge, S., & Dellagostin, O. A. (2017). The development of veterinary vaccines: A review of traditional methods and modern biotechnology approaches. *Biotechnology Research and Innovation*, 1, 6–13. Available from <https://doi.org/10.1016/j.biori.2017.10.001>.
- Kanekiyo, M., Ellis, D., & King, N. P. (2019). New vaccine design and delivery technologies. *The Journal of Infectious Diseases*, 219, S88–S96. Available from <https://doi.org/10.1093/infdis/jiy745>.
- Kar, T., Narsaria, U., Basak, S., Deb, D., Castiglione, F., Mueller, D. M., & Srivastava, A. P. (2020). A candidate multi-epitope vaccine against SARS-CoV-2. *Scientific Reports*, 10, 10895. Available from <https://doi.org/10.1038/s41598-020-67749-1>.
- Karosiene, E., Lundegaard, C., Lund, O., & Nielsen, M. (2012). NetMHCcons: A consensus method for the major histocompatibility complex I predictions. *Immunogenetics*, 64, 177–186. Available from <https://doi.org/10.1007/s00251-011-0579-8>.
- Kaur, D., Patiyal, S., Sharma, N., Usmani, S. S., & Raghava, G. P. S. (2019). PRRDB 2.0: A comprehensive database of pattern-recognition receptors and their ligands. *Database*, 2019. Available from <https://doi.org/10.1093/database/baz076>.
- Kelley, L. A., Mezulis, S., Yates, C. M., Wass, M. N., & Sternberg, M. J. E. (2015). The Phyre2 web portal for protein modeling, prediction and analysis. *Nature Protocols*, 10, 845–858. Available from <https://doi.org/10.1038/nprot.2015.053>.
- Keşmir, C., Nussbaum, A. K., Schild, H., Detours, V., & Brunak, S. (2002). Prediction of proteasome cleavage motifs by neural networks. *Protein Engineering, Design & Selection*, 15, 287–296. Available from <https://doi.org/10.1093/protein/15.4.287>.
- Kolev, M., & Kemper, C. (2017). Keeping it all going-complement meets metabolism. *Frontiers in Immunology*, 8, 1. Available from <https://doi.org/10.3389/fimmu.2017.00001>.
- Kringelum, J. V., Lundegaard, C., Lund, O., & Nielsen, M. (2012). Reliable B cell epitope predictions: Impacts of method development and improved benchmarking. *PLoS Computational Biology*, 8, e1002829.
- Kulkarni-Kale, U., Raskar-Renuse, S., Natekar-Kalantre, G., & Saxena, S. A. (2014). Antigen-Antibody Interaction Database (AgAbDb): A compendium of antigen-antibody interactions. *Methods in Molecular Biology*, 1184, 149–164. Available from [https://doi.org/10.1007/978-1-4939-1115-8\\_8](https://doi.org/10.1007/978-1-4939-1115-8_8).
- Kumar, R., Chaudhary, K., Sharma, M., Nagpal, G., Chauhan, J. S., Singh, S., . . . Raghava, G. P. S. (2015). AHTPDDB: A comprehensive platform for analysis and presentation of antihypertensive peptides. *Nucleic Acids Research*, 43, D956–D962. Available from <https://doi.org/10.1093/nar/gku1141>.
- Kumar, R., Chaudhary, K., Singh Chauhan, J., Nagpal, G., Kumar, Rahul, Sharma, M., & Raghava, G. P. S. (2015). An in silico platform for predicting, screening and designing of antihypertensive peptides. *Scientific Reports*, 5, 12512. Available from <https://doi.org/10.1038/srep12512>.
- Kumar, V., Agrawal, P., Kumar, R., Bhalla, S., Usmani, S. S., Varshney, G. C., & Raghava, G. P. S. (2018). Prediction of cell-penetrating potential of modified peptides containing natural and chemically modified residues. *Frontiers in Microbiol.*, 9, 725. Available from <https://doi.org/10.3389/fmicb.2018.00725>.
- Kuntal, B. K., Aparoy, P., & Reddanna, P. (2010). EasyModeller: A graphical interface to MODELLER. *BMC Research Notes*, 3, 226. Available from <https://doi.org/10.1186/1756-0500-3-226>.
- Lata, S., Bhasin, M., & Raghava, G. P. S. (2007). Application of machine learning techniques in predicting MHC binders. *Methods in Molecular Biology*, 409, 201–215. Available from [https://doi.org/10.1007/978-1-60327-118-9\\_14](https://doi.org/10.1007/978-1-60327-118-9_14).
- Lata, S., Bhasin, M., & Raghava, G. P. S. (2009). MHCBN 4.0: A database of MHC/TAP binding peptides and T-cell epitopes. *BMC Research Notes*, 2, 61. Available from <https://doi.org/10.1186/1756-0500-2-61>.
- Lata, S., & Raghava, G. P. S. (2008). PRRDB: A comprehensive database of pattern-recognition receptors and their ligands. *BMC Genomics*, 9, 180. Available from <https://doi.org/10.1186/1471-2164-9-180>.
- Lauck, F., Smith, C. A., Friedland, G. F., Humphris, E. L., & Kortemme, T. (2010). RosettaBackrub—A web server for flexible backbone protein structure modeling and design. *Nucleic Acids Research*, 38, W569–W575. Available from <https://doi.org/10.1093/nar/gkq369>.
- Liu, C., Chin, J. X., & Lee, D.-Y. (2015). SynLinker: An integrated system for designing linkers and synthetic fusion proteins. *Bioinformatics (Oxford, England)*, 31, 3700–3702. Available from <https://doi.org/10.1093/bioinformatics/btv447>.
- Livingston, B. D., Newman, M., Crimi, C., McKinney, D., Chesnut, R., & Sette, A. (2001). Optimization of epitope processing enhances immunogenicity of multiepitope DNA vaccines. *Vaccine*, 19, 4652–4660. Available from [https://doi.org/10.1016/S0264-410X\(01\)00233-X](https://doi.org/10.1016/S0264-410X(01)00233-X).
- Luciani, F., Bull, R. A., & Lloyd, A. R. (2012). Next generation deep sequencing and vaccine design: Today and tomorrow. *Trends in Biotechnology*, 30, 443–452. Available from <https://doi.org/10.1016/j.tibtech.2012.05.005>.
- Lykov, K., Li, X., Lei, H., Pivkin, I. V., & Karniadakis, G. E. (2015). Inflow/outflow boundary conditions for particle-based blood flow simulations: Application to arterial bifurcations and trees. *PLoS Computational Biology*, 11, e1004410.
- Mahanty, S., Prigent, A., & Garraud, O. (2015). Immunogenicity of infectious pathogens and vaccine antigens. *BMC Immunology*, 16, 31. Available from <https://doi.org/10.1186/s12865-015-0095-y>.

- Marciniuk, K., Trost, B., & Napper, S. (2015). EpIC: A rational pipeline for epitope immunogenicity characterization. *Bioinformatics (Oxford, England)*, *31*, 2388–2390. Available from <https://doi.org/10.1093/bioinformatics/btv136>.
- Markowitz, V. M., Chen, I.-M. A., Palaniappan, K., Chu, K., Szeto, E., Grechkin, Y., ... Kyrpides, N. C. (2012). IMG: The integrated microbial genomes database and comparative analysis system. *Nucleic Acids Research*, *40*, D115–D122. Available from <https://doi.org/10.1093/nar/gkr1044>.
- Martí-Renom, M. A., Stuart, A. C., Fiser, A., Sánchez, R., Melo, F., & Šali, A. (2000). Comparative protein structure modeling of genes and genomes. *Annual Review of Biophysics and Biomolecular Structure*, *29*, 291–325. Available from <https://doi.org/10.1146/annurev.biophys.29.1.291>.
- Mathur, D., Prakash, S., Anand, P., Kaur, H., Agrawal, P., Mehta, A., ... Raghava, G. P. S. (2016). PEPLife: A repository of the half-life of peptides. *Scientific Reports*, *6*, 36617. Available from <https://doi.org/10.1038/srep36617>.
- Mathur, D., Singh, S., Mehta, A., Agrawal, P., & Raghava, G. P. S. (2018). In silico approaches for predicting the half-life of natural and modified peptides in blood. *PLoS One*, *13*, e0196829.
- McSparron, H., Blythe, M. J., Zygouri, C., Doytchinova, I. A., & Flower, D. R. (2003). JenPep: A novel computational information resource for immunobiology and vaccinology. *Journal of Chemical Information and Computer Sciences*, *43*, 1276–1287. Available from <https://doi.org/10.1021/ci030461e>.
- McVey, S., & Shi, J. (2010). Vaccines in veterinary medicine: A brief review of history and technology. *The Veterinary Clinics of North America. Small Animal Practice*, *40*, 381–392. Available from <https://doi.org/10.1016/j.cvsm.2010.02.001>.
- Menezes-Souza, D., Mendes, T. A., de, O., Gomes, M., de, S., Bartholomeu, D. C., & Fujiwara, R. T. (2015). Improving serodiagnosis of human and canine leishmaniasis with recombinant *Leishmania braziliensis* cathepsin L-like protein and a synthetic peptide containing its linear B-cell epitope. *PLoS Neglected Tropical Diseases*, *9*, e3426. Available from <https://doi.org/10.1371/journal.pntd.0003426>.
- Minias, P., Pikus, E., & Anderwald, D. (2019). Allelic diversity and selection at the MHC class I and class II in a bottlenecked bird of prey, the White-tailed Eagle. *BMC Evolutionary Biology*, *19*, 2. Available from <https://doi.org/10.1186/s12862-018-1338-3>.
- Morris, G. M., Huey, R., Lindstrom, W., Sanner, M. F., Belew, R. K., Goodsell, D. S., & Olson, A. J. (2009). AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. *Journal of Computational Chemistry*, *30*, 2785–2791. Available from <https://doi.org/10.1002/jcc.21256>.
- Moutafsi, M., Peters, B., Pasquetto, V., Tschärke, D. C., Sidney, J., Bui, H.-H., ... Sette, A. (2006). A consensus epitope prediction approach identifies the breadth of murine T(CD8+) cell responses to vaccinia virus. *Nature Biotechnology*, *24*, 817–819. Available from <https://doi.org/10.1038/nbt1215>.
- Nagpal, G., Chaudhary, K., Agrawal, P., & Raghava, G. P. S. (2018). Computer-aided prediction of antigen presenting cell modulators for designing peptide-based vaccine adjuvants. *Journal of Translational Medicine*, *16*, 181. Available from <https://doi.org/10.1186/s12967-018-1560-1>.
- Nagpal, G., Usmani, S. S., Dhanda, S. K., Kaur, H., Singh, S., Sharma, M., & Raghava, G. P. S. (2017). Computer-aided designing of immunosuppressive peptides based on IL-10 inducing potential. *Scientific Reports*, *7*, 42851. Available from <https://doi.org/10.1038/srep42851>.
- Neeffes, J., Jongsma, M. L. M., Paul, P., & Bakke, O. (2011). Towards a systems understanding of MHC class I and MHC class II antigen presentation. *Nature Reviews Immunology*, *11*, 823–836. Available from <https://doi.org/10.1038/nri3084>.
- Nielsen, M., Lundegaard, C., Lund, O., & Kesmir, C. (2005). The role of the proteasome in generating cytotoxic T-cell epitopes: Insights obtained from improved predictions of proteasomal cleavage. *Immunogenetics*, *57*, 33–41. Available from <https://doi.org/10.1007/s00251-005-0781-7>.
- O'Connor, E. A., Strandh, M., Hasselquist, D., Nilsson, J.-Å., & Wester Dahl, H. (2016). The evolution of highly variable immunity genes across a passerine bird radiation. *Molecular Ecology*, *25*, 977–989. Available from <https://doi.org/10.1111/mec.13530>.
- Oyarzún, P., Ellis, J. J., Bodén, M., & Kobe, B. (2013). PREDIVAC: CD4+ T-cell epitope prediction for vaccine design that covers 95% of HLA class II DR protein diversity. *BMC Bioinformatics*, *14*, 52. Available from <https://doi.org/10.1186/1471-2105-14-52>.
- Peterson, J. D., Umayam, L. A., Dickinson, T., Hickey, E. K., & White, O. (2001). The comprehensive microbial resource. *Nucleic Acids Research*, *29*, 123–125. Available from <https://doi.org/10.1093/nar/29.1.123>.
- Phillips, J. C., Braun, R., Wang, W., Gumbart, J., Tajkhorshid, E., Villa, E., ... Schulten, K. (2005). Scalable molecular dynamics with NAMD. *Journal of Computational Chemistry*, *26*, 1781–1802. Available from <https://doi.org/10.1002/jcc.20289>.
- Pillay, S., Giandhari, J., Tegally, H., Wilkinson, E., Chimukangara, B., Lessells, R., ... de Oliveira, T. (2020). Whole genome sequencing of SARS-CoV-2: Adapting illumina protocols for quick and accurate outbreak investigation during a pandemic. *Genes (Basel)*, *11*. Available from <https://doi.org/10.3390/genes11080949>.
- Ponomarenko, J., Bui, H.-H., Li, W., Fusseder, N., Bourne, P. E., Sette, A., & Peters, B. (2008). ElliPro: A new structure-based tool for the prediction of antibody epitopes. *BMC Bioinformatics*, *9*, 514. Available from <https://doi.org/10.1186/1471-2105-9-514>.
- Racz, R., Chung, M., Xiang, Z., & He, Y. (2013). Systematic annotation and analysis of “virulence factors”—Virulence factors whose mutants can be used as live attenuated vaccines. *Vaccine*, *31*, 797–805. Available from <https://doi.org/10.1016/j.vaccine.2012.11.066>.
- Racz, R., Li, X., Patel, M., Xiang, Z., & He, Y. (2014). DNAXDB: The first web-based DNA vaccine database and its data analysis. *BMC Bioinformatics*, *15*, S2. Available from <https://doi.org/10.1186/1471-2105-15-S4-S2>.
- Rammensee, H.-G., Bachmann, J., Emmerich, N. P. N., Bachor, O. A., & Stevanović, S. (1999). SYFPEITHI: Database for MHC ligands and peptide motifs. *Immunogenetics*, *50*, 213–219. Available from <https://doi.org/10.1007/s002510050595>.
- Rappuoli, R. (2000). Reverse vaccinology. *Current Opinion in Microbiology*, *3*, 445–450. Available from [https://doi.org/10.1016/s1369-5274\(00\)00119-3](https://doi.org/10.1016/s1369-5274(00)00119-3).
- Rappuoli, R., Miller, H. I., & Falkow, S. (2002). Medicine. The intangible value of vaccination. *Science*, *297*, 937–939. Available from <https://doi.org/10.1126/science.1075173>.

- Reynisson, B., Alvarez, B., Paul, S., Peters, B., & Nielsen, M. (2020). NetMHCpan-4.1 and NetMHCIIpan-4.0: Improved predictions of MHC antigen presentation by concurrent motif deconvolution and integration of MS MHC eluted ligand data. *Nucleic Acids Research*, 48, W449–W454. Available from <https://doi.org/10.1093/nar/gkaa379>.
- Rizwan, M., Naz, A., Ahmad, J., Naz, K., Obaid, A., Parveen, T., ... Ali, A. (2017). VacSol: A high throughput in silico pipeline to predict potential therapeutic targets in prokaryotic pathogens using subtractive reverse vaccinology. *BMC Bioinformatics*, 18, 106. Available from <https://doi.org/10.1186/s12859-017-1540-0>.
- Robinson, J., Barker, D. J., Georgiou, X., Cooper, M. A., Flicek, P., & Marsh, S. G. E. (2020). IPD-IMGT/HLA database. *Nucleic Acids Research*, 48, D948–D955. Available from <https://doi.org/10.1093/nar/gkz950>.
- Rock, K. L., Reits, E., & Neefjes, J. (2016). Present yourself! By MHC class I and MHC class II molecules. *Trends in Immunology*, 37, 724–737. Available from <https://doi.org/10.1016/j.it.2016.08.010>.
- Rubinstein, N. D., Mayrose, I., Martz, E., & Pupko, T. (2009). EpiToptia: A web-server for predicting B-cell epitopes. *BMC Bioinformatics*, 10, 287. Available from <https://doi.org/10.1186/1471-2105-10-287>.
- Saha, S., Bhasin, M., & Raghava, G. P. S. (2005). Bcipep: A database of B-cell epitopes. *BMC Genomics*, 6, 79. Available from <https://doi.org/10.1186/1471-2164-6-79>.
- Saha, S., & Raghava, G. P. S. (2004). BcePred: Prediction of continuous B-cell epitopes in antigenic sequences using physico-chemical properties. In G. Nicosia, V. Cutello, P. J. Bentley, & J. Timmis (Eds.), *Artificial immune systems* (pp. 197–204). Heidelberg, Berlin: Springer.
- Saha, S., & Raghava, G. P. S. (2006a). Prediction of continuous B-cell epitopes in an antigen using recurrent neural network. *Proteins: Structure, Function, and Bioinformatics*, 65, 40–48. Available from <https://doi.org/10.1002/prot.21078>.
- Saha, S., & Raghava, G. P. S. (2006b). AlgPred: Prediction of allergenic proteins and mapping of IgE epitopes. *Nucleic Acids Research*, 34, W202–W209. Available from <https://doi.org/10.1093/nar/gkl343>.
- Saravanan, V., & Gautham, N. (2015). Harnessing computational biology for exact linear B-cell epitope prediction: A novel amino acid composition-based feature descriptor. *OMICS*, 19, 648–658. Available from <https://doi.org/10.1089/omi.2015.0095>.
- Sayers, S., Ulysse, G., Xiang, Z., & He, Y. (2012). Vaxjo: A web-based vaccine adjuvant database and its application for analysis of vaccine adjuvants and their uses in vaccine development. *Journal of Biomedicine & Biotechnology*, 2012, 831486. Available from <https://doi.org/10.1155/2012/831486>.
- Schlessinger, A., Ofra, Y., Yachdav, G., & Rost, B. (2006). Epitome: Database of structure-inferred antigenic epitopes. *Nucleic Acids Research*, 34, D777–D780. Available from <https://doi.org/10.1093/nar/gkj053>.
- Schubert, B., Walzer, M., Brachvogel, H.-P., Szolek, A., Mohr, C., & Kohlbacher, O. (2016). FRED 2: An immunoinformatics framework for Python. *Bioinformatics (Oxford, England)*, 32, 2044–2046. Available from <https://doi.org/10.1093/bioinformatics/btw113>.
- Sharma, N., Patiyal, S., Dhali, A., Pande, A., Arora, C., & Raghava, G. P. S. (2020). AlgPred 2.0: An improved method for predicting allergenic proteins and mapping of IgE epitopes. *Briefings in Bioinformatics*. Available from <https://doi.org/10.1093/bib/bbaa294>.
- Simoneau, M., Boulanger, J., Coulombe, G., Renaud, M.-A., Duchesne, C., & Rivard, N. (2008). Activation of Cdk2 stimulates proteasome-dependent truncation of tyrosine phosphatase SHP-1 in human proliferating intestinal epithelial cells. *The Journal of Biological Chemistry*, 283, 25544–25556. Available from <https://doi.org/10.1074/jbc.M804177200>.
- Singh, H., Ansari, H. R., & Raghava, G. P. S. (2013). Improved method for linear B-cell epitope prediction using antigen's primary sequence. *PLoS One*, 8, e62216.
- Singh, H., & Raghava, G. P. S. (2001). ProPred: Prediction of HLA-DR binding sites. *Bioinformatics (Oxford, England)*, 17, 1236–1237. Available from <https://doi.org/10.1093/bioinformatics/17.12.1236>.
- Singh, S., Chaudhary, K., Dhanda, S. K., Bhalla, S., Usmani, S. S., Gautam, A., ... Raghava, G. P. S. (2016). SATPdb: A database of structurally annotated therapeutic peptides. *Nucleic Acids Research*, 44, D1119–D1126. Available from <https://doi.org/10.1093/nar/gkv1114>.
- Stille, C. J., Thomas, L. J., Reyes, V. E., & Humphreys, R. E. (1987). Hydrophobic strip-of-helix algorithm for selection of T cell-presented peptides. *Molecular Immunology*, 24, 1021–1027. Available from [https://doi.org/10.1016/0161-5890\(87\)90068-X](https://doi.org/10.1016/0161-5890(87)90068-X).
- Stranzl, T., Larsen, M. V., Lundegaard, C., & Nielsen, M. (2010). NetCTLpan: Pan-specific MHC class I pathway epitope predictions. *Immunogenetics*, 62, 357–368. Available from <https://doi.org/10.1007/s00251-010-0441-4>.
- Sweredoski, M. J., & Baldi, P. (2008). PEPITO: Improved discontinuous B-cell epitope prediction using multiple distance thresholds and half sphere exposure. *Bioinformatics (Oxford, England)*, 24, 1459–1460. Available from <https://doi.org/10.1093/bioinformatics/btn199>.
- Sweredoski, M. J., & Baldi, P. (2009). COBEpro: A novel system for predicting continuous B-cell epitopes. *Protein Engineering, Design & Selection*, 22, 113–120. Available from <https://doi.org/10.1093/protein/gzn075>.
- Tahir ul Qamar, M., Saleem, S., Ashfaq, U. A., Bari, A., Anwar, F., & Alqahtani, S. (2019). Epitope-based peptide vaccine design and target site depiction against Middle East Respiratory Syndrome Coronavirus: An immune-informatics study. *Journal of Translational Medicine*, 17, 362. Available from <https://doi.org/10.1186/s12967-019-2116-8>.
- Tatusova, T., Ciufu, S., Fedorov, B., O'Neill, K., & Tolstoy, I. (2014). RefSeq microbial genomes database: New representation and annotation strategy. *Nucleic Acids Research*, 42, D553–D559. Available from <https://doi.org/10.1093/nar/gkt1274>.
- Trolle, T., & Nielsen, M. (2014). NetTepi: An integrated method for the prediction of T cell epitopes. *Immunogenetics*, 66, 449–456. Available from <https://doi.org/10.1007/s00251-014-0779-0>.
- Trott, O., & Olson, A. J. (2010). AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *Journal of Computational Chemistry*, 31, 455–461. Available from <https://doi.org/10.1002/jcc.21334>.
- Unnikrishnan, M., Rappuoli, R., & Serruto, D. (2012). Recombinant bacterial vaccines. *Current Opinion in Immunology*, 24, 337–342. Available from <https://doi.org/10.1016/j.coi.2012.03.013>.

- Van Regenmortel, M. H. V. (2009). What is a B-cell epitope? In M. Schutkowski, & U. Reineke (Eds.), *Epitope Mapping Protocols* (2nd ed., pp. 3–20). Totowa, NJ: Humana Press. Available from [https://doi.org/10.1007/978-1-59745-450-6\\_1](https://doi.org/10.1007/978-1-59745-450-6_1).
- Vita, R., Mahajan, S., Overton, J. A., Dhanda, S. K., Martini, S., Cantrell, J. R., . . . Peters, B. (2019). The Immune Epitope Database (IEDB): 2018 update. *Nucleic Acids Research*, *47*, D339–D343. Available from <https://doi.org/10.1093/nar/gky1006>.
- von Boehmer, H., & Kisielow, P. (1990). Self-nonsel self discrimination by T cells. *Science*, *248*, 1369 LP–1361373. Available from <https://doi.org/10.1126/science.1972594>.
- Wang, P., Sidney, J., Dow, C., Mothé, B., Sette, A., & Peters, B. (2008). A systematic assessment of MHC class II peptide binding predictions and evaluation of a consensus approach. *PLoS Computational Biology*, *4*, e1000048. Available from <https://doi.org/10.1371/journal.pcbi.1000048>.
- Waterhouse, A., Bertoni, M., Bienert, S., Studer, G., Tauriello, G., Gumienny, R., . . . Schwede, T. (2018). SWISS-MODEL: Homology modelling of protein structures and complexes. *Nucleic Acids Research*, *46*, W296–W303. Available from <https://doi.org/10.1093/nar/gky427>.
- Wei, L., Xing, P., Su, R., Shi, G., Ma, Z. S., & Zou, Q. (2017). CPPred-RF: A sequence-based predictor for identifying cell-penetrating peptides and their uptake efficiency. *Journal of Proteome Research*, *16*, 2044–2053. Available from <https://doi.org/10.1021/acs.jproteome.7b00019>.
- Weinstein, J. N., Collisson, E. A., Mills, G. B., Shaw, K. R. M., Ozenberger, B. A., Ellrott, K., . . . Stuart, J. M. (2013). The Cancer Genome Atlas Pan-Cancer analysis project. *Nature Genetics*, *45*, 1113–1120. Available from <https://doi.org/10.1038/ng.2764>.
- Win, T. S., Schaduangrat, N., Prachayasittikul, V., Nantasenamat, C., & Shoombuatong, W. (2018). PAAP: A web server for predicting antihypertensive activity of peptides. *Future Medicinal Chemistry*, *10*, 1749–1767. Available from <https://doi.org/10.4155/fmc-2017-0300>.
- Xiang, Z., Todd, T., Ku, K. P., Kovacic, B. L., Larson, C. B., Chen, F., . . . He, Y. (2008). VIOLIN: Vaccine investigation and online information network. *Nucleic Acids Research*, *36*, D923–D928. Available from <https://doi.org/10.1093/nar/gkm1039>.
- Yang, B., Sayers, S., Xiang, Z., & He, Y. (2011). Protegen: A web-based protective antigen database and analysis system. *Nucleic Acids Research*, *39*, D1073–D1078. Available from <https://doi.org/10.1093/nar/gkq944>.
- Yao, B., Zhang, L., Liang, S., & Zhang, C. (2012). SVMTriP: A method to predict antigenic epitopes using support vector machine to integrate tripeptide similarity and propensity. *PLoS One*, *7*, e45152.
- Zhang, H., Lund, O., & Nielsen, M. (2009). The PickPocket method for predicting binding specificities for receptors based on receptor pocket similarities: Application to MHC-peptide binding. *Bioinformatics (Oxford, England)*, *25*, 1293–1299. Available from <https://doi.org/10.1093/bioinformatics/btp137>.
- Zhang, J., Bajari, R., Andric, D., Gerthoffert, F., Lepsa, A., Nahal-Bose, H., . . . Ferretti, V. (2019). The International Cancer Genome Consortium Data Portal. *Nature Biotechnology*. Available from <https://doi.org/10.1038/s41587-019-0055-9>.
- Zhang, J., Zhao, X., Sun, P., Gao, B., & Ma, Z. (2014). Conformational B-cell epitopes prediction from sequences using cost-sensitive ensemble classifiers and spatial clustering. *BioMed Research International*, *2014*, 689219. Available from <https://doi.org/10.1155/2014/689219>.
- Zhong, S., Malecek, K., Johnson, L. A., Yu, Z., Vega-Saenz de Miera, E., Darvishian, F., . . . Krogsgaard, M. (2013). T-cell receptor affinity and avidity defines antitumor response and autoimmunity in T-cell immunotherapy. *Proceedings of the National Academy of Sciences of the United States of America*, *110*, 6973–6978. Available from <https://doi.org/10.1073/pnas.1221609110>.
- Zhou, C., Chen, Z., Zhang, L., Yan, D., Mao, T., Tang, K., . . . Cao, Z. (2019). SEPPA 3.0-enhanced spatial epitope prediction enabling glycoprotein antigens. *Nucleic Acids Research*, *47*, W388–W394. Available from <https://doi.org/10.1093/nar/gkz413>.