

Review Article

Fundamentals and Methods for T- and B-Cell Epitope Prediction

Jose L. Sanchez-Trincado, Marta Gomez-Perosanz, and Pedro A. Reche

Laboratory of Immunomedicine, Faculty of Medicine, Complutense University of Madrid, Ave Complutense S/N, 28040 Madrid, Spain

Correspondence should be addressed to Pedro A. Reche; parecheg@med.ucm.es

Received 27 July 2017; Revised 22 November 2017; Accepted 27 November 2017; Published 28 December 2017

Academic Editor: Senthami R. Selvan

Copyright © 2017 Jose L. Sanchez-Trincado et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Adaptive immunity is mediated by T- and B-cells, which are immune cells capable of developing pathogen-specific memory that confers immunological protection. Memory and effector functions of B- and T-cells are predicated on the recognition through specialized receptors of specific targets (antigens) in pathogens. More specifically, B- and T-cells recognize portions within their cognate antigens known as epitopes. There is great interest in identifying epitopes in antigens for a number of practical reasons, including understanding disease etiology, immune monitoring, developing diagnosis assays, and designing epitope-based vaccines. Epitope identification is costly and time-consuming as it requires experimental screening of large arrays of potential epitope candidates. Fortunately, researchers have developed *in silico* prediction methods that dramatically reduce the burden associated with epitope mapping by decreasing the list of potential epitope candidates for experimental testing. Here, we analyze aspects of antigen recognition by T- and B-cells that are relevant for epitope prediction. Subsequently, we provide a systematic and inclusive review of the most relevant B- and T-cell epitope prediction methods and tools, paying particular attention to their foundations.

1. Introduction

The immune system is typically divided into two categories, innate and adaptive. Innate immunity involves nonspecific defense mechanisms that act immediately or within hours after a microbe appearance in the body. All multicellular beings exhibit some kind of innate immunity. In contrast, adaptive immunity is only present in vertebrates and it is highly specific. In fact, the adaptive immune system is able to recognize and destroy invading pathogens individually. Moreover, the adaptive immune system remembers the pathogens that fights, acquiring a pathogen-specific long-lasting protective memory that enables stronger attacks each time the pathogen is encountered [1]. Nonetheless, innate and adaptive immune mechanisms work together and adaptive immunity elicitation is contingent on prior activation of innate immune responses [1].

Adaptive immunity is articulated by lymphocytes, more specifically by B- and T-cells, which are responsible for the

humoral and cell-mediated immunity. B- and T-cells do not recognize pathogens as a whole, but molecular components known as antigens. These antigens are recognized by specific receptors present in the cell surface of B- and T-cells. Antigen recognition by these receptors is required to activate B- and T-cells but not enough, as second activation signals stemming from the activation of the innate immune system are also needed. The specificity of the recognition is determined by genetic recombination events that occur during lymphocyte development, which lead to generating millions of different variants of lymphocytes in terms of the antigen-recognizing receptors [1]. Antigen recognition by B- and T-cells differ greatly.

B-cells recognize solvent-exposed antigens through antigen receptors, named as B-cell receptors (BCR), consisting of membrane-bound immunoglobulins, as shown in Figure 1. Upon activation, B-cells differentiate and secrete soluble forms of the immunoglobulins, also known as antibodies, which mediate humoral adaptive immunity.

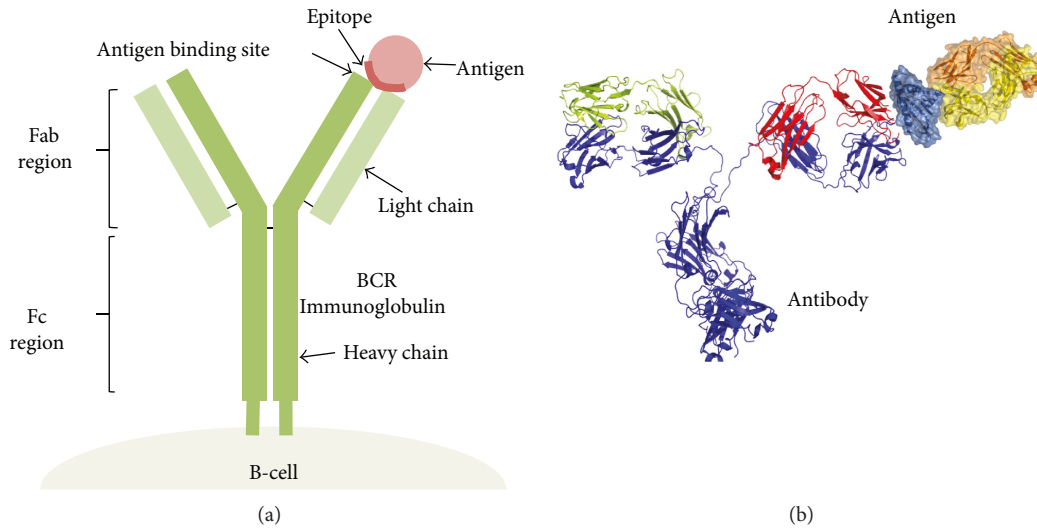


FIGURE 1: B-cell epitope recognition. B-cell epitopes are solvent-exposed portions of the antigen that bind to secreted and cell-bound immunoglobulins. (a) B-cell receptors encompass cell-bound immunoglobulins, consisting of two heavy chains and two light chains. The different chains and regions are annotated. (b) Molecular representation of the interaction between an antibody and the antigen. Antibodies are secreted immunoglobulins of known specificity.

Antibodies released by B-cells can have different functions that are triggered upon binding their cognate antigens. These functions include neutralizing toxins and pathogens and labeling them for destruction [1].

A B-cell epitope is the antigen portion binding to the immunoglobulin or antibody. These epitopes recognized by B-cells may constitute any exposed solvent region in the antigen and can be of different chemical nature. However, most antigens are proteins and those are the subjects for epitope prediction methods.

On the other hand, T-cells present on their surface a specific receptor known as T-cell receptor (TCR) that enables the recognition of antigens when they are displayed on the surface of antigen-presenting cells (APCs) bound to major histocompatibility complex (MHC) molecules. T-cell epitopes are presented by class I (MHC I) and II (MHC II) MHC molecules that are recognized by two distinct subsets of T-cells, CD8 and CD4 T-cells, respectively (Figure 2). Subsequently, there are CD8 and CD4 T-cell epitopes. CD8 T-cells become cytotoxic T lymphocytes (CTL) following T CD8 epitope recognition. Meanwhile, primed CD4 T-cells become helper (Th) or regulatory (Treg) T-cells [1]. Th cells amplify the immune response, and there are three main subclasses: Th1 (cell-mediated immunity against intracellular pathogens), Th2 (antibody-mediated immunity), and Th17 (inflammatory response and defense against extracellular bacteria) [2].

Identifying epitopes in antigens is of great interest for a number of practical reasons, including understanding disease etiology, immune monitoring, developing diagnosis assays, and designing epitope-based vaccines. B-cell epitopes can be identified by different methods including solving the 3D structure of antigen-antibody complexes, peptide library screening of antibody binding or performing functional assays in which the antigen is mutated and the interaction antibody-antigen is evaluated [3, 4]. On the other hand,

experimental determination of T-cell epitopes is carried out using MHC multimers and lymphoproliferation or ELISPOT assays, among others [5, 6]. Traditional epitope identification has depended entirely upon experimental techniques, being costly and time-consuming. Thereby, scientists have developed and implemented epitope prediction methods that facilitate epitope identification and decrease the experimental load associated with it. Here, we will first analyze aspects of antigen recognition by T- and B-cells that are relevant for a better understanding of the topic of epitope prediction. Subsequently, we will provide a systematic and inclusive review of the most important prediction methods and tools, paying particular attention to their foundations and potentials. We will also discuss epitope prediction limitations and ways to overcome them. We will start with T-cell epitopes.

2. T-Cell Epitope Prediction

T-cell epitope prediction aims to identify the shortest peptides within an antigen that are able to stimulate either CD4 or CD8 T-cells [7]. This capacity to stimulate T-cells is called immunogenicity, and it is confirmed in assays requiring synthetic peptides derived from antigens [5, 6]. There are many distinct peptides within antigens and T-cell prediction methods aim to identify those that are immunogenic. T-cell epitope immunogenicity is contingent on three basic steps: (i) antigen processing, (ii) peptide binding to MHC molecules, and (iii) recognition by a cognate TCR. Of these three events, MHC-peptide binding is the most selective one at determining T-cell epitopes [8, 9]. Therefore, prediction of peptide-MHC binding is the main basis to anticipate T-cell epitopes and we will review it next.

2.1. Prediction of Peptide-MHC Binding. MHC I and MHC II molecules have similar 3D-structures with bound peptides sitting in a groove delineated by two α -helices overlying a

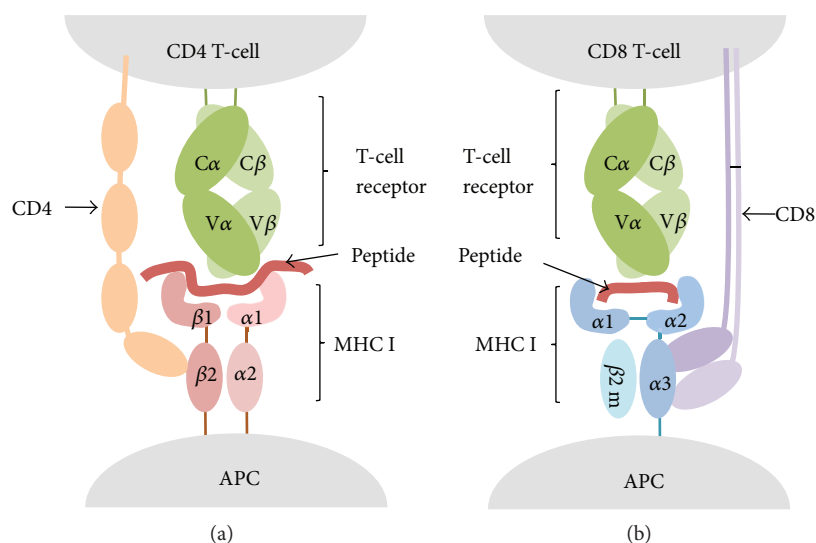


FIGURE 2: T-cell epitope recognition. T-cell epitopes are peptides derived from antigens and recognized by the T-cell receptor (TCR) when bound to MHC molecules displayed on the cell surface of APCs. (a) CD4 T-cells express the CD4 coreceptor, which binds to MHC II, and recognize peptides presented by MHC II molecules. (b) CD8 T-cells express the CD8 coreceptor, which binds to MHC I, and recognize peptides presented by MHC I molecules.

floor comprised of eight antiparallel β -strands. However, there are also key differences between MHC I and II binding grooves that we must highlight for they condition peptide-binding predictions (Figure 3). The peptide-binding cleft of MHC I molecules is closed as it is made by a single α chain. As a result, MHC I molecules can only bind short peptides ranging from 9 to 11 amino acids, whose N- and C-terminal ends remain pinned to conserved residues of the MHC I molecule through a network of hydrogen bonds [10, 11]. The MHC I peptide-binding groove also contains deep binding pockets with tight physicochemical preferences that facilitate binding predictions. There is a complication however. Peptides that have different sizes and bind to the same MHC I molecule often use alternative binding pockets [12]. Therefore, methods predicting peptide-MHC I binding require a fixed peptide length. However, since most MHC I peptide ligands have 9 residues, it is generally preferable to predict peptides with that size. In contrast, the peptide-binding groove of MHC II molecules is open, allowing the N- and C-terminal ends of a peptide to extend beyond the binding groove [10, 11]. As a result, MHC II-bound peptides vary widely in length (9–22 residues), although only a core of nine residues (peptide-binding core) sits into the MHC II binding groove. Therefore, peptide-MHC II binding prediction methods often target to identify these peptide-binding cores. MHC II molecule binding pockets are also shallower and less demanding than those of MHC I molecules. As a consequence, peptide-binding prediction to MHC II molecules is less accurate than that of MHC I molecules.

Given the relevance of the problem, there are numerous methods to predict peptide-MHC binding. The most relevant with free online use are collected on Table 1. They can be divided in two main categories: data-driven and structure-based methods. Structure-based approaches generally rely

on modeling the peptide-MHC structure followed by evaluation of the interaction through methods such as molecular dynamic simulations [8, 13]. Structure-based methods have the great advantage of not needing experimental data. However, they are seldom used as they are computationally intensive and exhibit lower predictive performance than data-driven methods [14].

Data-driven methods for peptide-MHC binding prediction are based on peptide sequences that are known to bind to MHC molecules. These peptide sequences are generally available in specialized epitope databases such as IEDB [15], EPIMHC [16], Antijen [17, 18]. Both MHC I and II binding peptides contain frequently occurring amino acids at particular peptide positions, known as anchor residues. Thereby, prediction of peptide-MHC binding was first approached using sequence motif (SM) reflecting amino acid preferences of MHC molecules at anchor positions [19]. However, it was soon shown that nonanchor residues also contribute to the capacity of a peptide to bind to a given MHC molecule [20, 21]. Subsequently, researchers developed motif matrices (MM), which could evaluate the contribution of each and all peptide positions to the binding with the MHC molecule [22–25]. The most sophisticated form of motif matrices consists of profiles [24–26] that are similar to those used for detecting sequence homology [27]. We would like to remark that motif matrices are often mistaken with quantitative affinity matrices (QAMs) since both produce peptide scores. However, MMs are derived without taking in consideration values of binding affinities and, therefore, resulting peptide scores are not suited to address binding affinity. In contrast, QAMs are trained on peptides and corresponding binding affinities, and aim to predict binding affinity. The first method based on QAMs was developed by Parker et al. [28] (Table 1). Subsequently, various approaches were developed to obtain QAMs from peptide

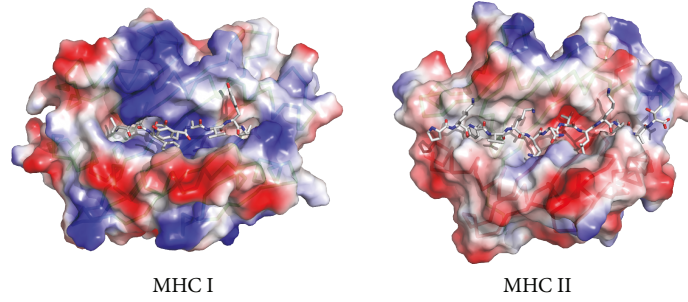


FIGURE 3: MHC molecule binding groove. The figure depicts the molecular surface as seen by the TCR of representative MHC I and II molecules. Note how the binding groove of the MHC I molecule is closed but that of MHC II is open. As a result, MHC I molecules bind short peptides (8–11 amino acids), while MHC II molecules bind longer peptides (9–22 amino acids). The figure was prepared from PDB files 1QRN (MHC I) and 1FYT (MHC II) using PyMol.

TABLE 1: Selected T-cell epitope prediction tools available online for free public use.

Tool	URL	Method ¹	MHC class	A	S	T	P	Ref.
EpiDOCK	http://epidock.ddg-pharmfac.net	SB	II	—	—	—	—	[86]
MotifScan	https://www.hiv.lanl.gov/content/immunology/motif_scan/motif_scan	SM	I and II	—	X	—	—	—
Rankpep	http://imed.med.ucm.es/Tools/rankpep.html	MM	I and II	—	—	—	X	[26]
SYFPEITHI	http://www.syfpeithi.de/	MM	I and II	—	—	—	—	[23]
MAPPP	http://www.mpiib-berlin.mpg.de/MAPPP/	MM	I	—	X	—	X	[87]
PREDIVAC	http://predivac.biosci.uq.edu.au/	MM	II	—	—	—	—	[88]
PEPVAC	http://imed.med.ucm.es/PEPVAC/	MM	I	—	X	—	X	[63]
EPISOPT	http://bio.med.ucm.es/episopt.html	MM	I	—	X	—	—	[64]
Vaxign	http://www.violinet.org/vaxign/	MM	I and II	—	—	—	—	[89]
MHCPred	http://www.ddg-pharmfac.net/mhcpred/MHCPred/	QSAR	I and II	X	—	—	—	[34]
EpiTOP	http://www.pharmfac.net/EpiTOP	QSAR	II	X	—	—	—	[90]
BIMAS	https://www.bimas.cit.nih.gov/molbio/hla_bind/	QAM	I	X	—	—	—	[28]
TEPITOPE	http://datamining-iip.fudan.edu.cn/service/TEPITOPEpan/TEPITOPEpan.html	QAM	II	X	—	—	—	[32]
Propred	http://www.imtech.res.in/raghava/propred/	QAM	II	X	X	—	—	[91]
Propred-1	http://www.imtech.res.in/raghava/propred1/	QAM	I	X	X	—	X	[92]
EpiJen	http://www.ddg-pharmfac.net/epijen/EpiJen/EpiJen.htm	QAM	I	X	—	X	X	[82]
IEDB-MHCI	http://tools.immuneepitope.org/mhci/	Combined	I	X	—	—	—	[93]
IEDB-MHCII	http://tools.immuneepitope.org/mhcii/	Combined	II	X	—	—	—	[93]
IL4pred	http://webs.iitd.edu.in/raghava/il4pred/index.php	SVM	II	—	—	—	—	[67]
MULTIPRED2	http://cvc.dfci.harvard.edu/multipred2/index.php	ANN	I and II	—	X	—	—	[62]
MHC2PRED	http://www.imtech.res.in/raghava/mhc2pred/index.html	SVM	II	—	—	—	—	[38]
NetMHC	http://www.cbs.dtu.dk/services/NetMHC/	ANN	I	X	—	—	—	[49]
NetMHCII	http://www.cbs.dtu.dk/services/NetMHCII/	ANN	II	X	—	—	—	[30]
NetMHCpan	http://www.cbs.dtu.dk/services/NetMHCpan/	ANN	I	X	—	—	—	[54]
NetMHCIIpan	http://www.cbs.dtu.dk/services/NetMHCIIpan/	ANN	II	X	—	—	—	[55]
nHLApred	http://www.imtech.res.in/raghava/nhlapred/	ANN	I	—	—	—	X	[94]
SVMHC	http://abi.inf.uni-tuebingen.de/Services/SVMHC/	SVM	I and II	—	—	—	—	[95]
SVRMHC	http://us accurascience.com/SVRMHCdb/	SVM	I and II	X	—	—	—	[46]
NetCTL	http://www.cbs.dtu.dk/services/NetCTL/	ANN	I	X	X	X	X	[83]
WAPP	https://abi.inf.uni-tuebingen.de/Services/WAPP/index_html	SVM	I	—	—	X	X	[37]

¹Method used for prediction of peptide-MHC binding. Keys for methods: SM: sequence motif; SB: structure-based; MM: motif matrix; QAM: quantitative affinity matrix; SVM: support vector machine; ANN: artificial neural network; QSAR: quantitative structure-activity relationship model; combined: tool uses different methods including ANN and QAM, selecting the more appropriate method for each distinct MHC molecule. The table also indicates whether the tools predict quantitative binding affinity (A), supertypes (S), TAP binding (T), and proteasomal cleavage (P); marked with an X in the affirmative case.

affinity data and predict peptide binding to MHC I and II molecules [29–32].

QAMs and motif matrices assume an independent contribution of peptide side chains to the binding. This assumption is well supported by experimental data but there is also evidence that neighboring peptide residues interfere with others [33]. To account for those interferences, researchers introduced quantitative structure-activity relationship (QSAR) additive models wherein the binding affinity of peptides to MHC is computed as the sum of amino acid contributions at each position plus the contribution of adjacent side chain interactions [34]. However, machine learning (ML) is the most popular and robust approach introduced to deal with the nonlinearity of peptide-MHC binding data [8]. Researchers have used ML for two distinct problems: the discrimination of MHC binders from nonbinders and the prediction of binding affinity of peptides to MHC molecules.

For developing discrimination models, ML algorithms are trained on data sets consisting of peptides that either bind or do not bind to MHC molecules. Relevant examples of ML-based discrimination models are those based on artificial neural networks (ANNs) [35, 36], support vector machines (SVMs) [37–39], decision trees (DTs) [40, 41], and Hidden Markov models (HMMs), which can also cope with nonlinear data and have been used to discriminate peptides binding to MHC molecules. However, unlike other ML algorithms, they have to be trained only on positive data. Three types of HMMs have been used to predict MHC-peptide binding: fully connected HMMs [42], structure-optimized HMMs [43], and profile HMMs [43, 44]. Of these, only fully connected HMMs (fcHMMs) and structure-optimized HMMs (soHMMs) can recognize different patterns in the peptide binders. In fact, profile HMMs that are derived from sets of ungapped alignments (the case for peptides binding to MHC) are nearly identical to profile matrices [45] (Table 1).

With regard to predicting binding affinity, ML algorithms are trained on datasets consisting of peptides with known affinity to MHC molecules. Both SVMs and ANNs have been used for such purpose. SVMs were first applied to predict peptide-binding affinity to MHC I molecules [46] and later to MHC II molecules [47] (Table 1). Likewise, ANNs were also applied first to the prediction of peptide binding to MHC I [48, 49] and later to MHC II molecules [50] (Table 1). Benchmarking of peptide-MHC binding prediction methods appears to indicate that those based on ANNs are superior to those based on QAMs and MMs. However, the differences between the distinct methods are marginal and vary for different MHC molecules [51]. Moreover, it has been shown that the performance of peptide-MHC predictions is improved by combining several methods and providing consensus predictions [52].

A major complication for predicting T-cell epitopes through peptide-MHC binding models is MHC polymorphism. In humans, MHC molecules are known as human leukocyte antigens (HLAs), and there are hundreds of allelic variants of class I (HLA I) and class II (HLA II) molecules. These HLA allelic variants bind distinct sets of peptides [53] and require specific models for predicting peptide-

MHC binding. However, peptide-binding data is only available for a minority of HLA molecules. To overcome this limitation, some researchers have developed pan-MHC-specific methods by training ANNs on input data combining MHC residues that contact the peptide with peptide-binding affinity that are capable of predicting peptide-binding affinities to uncharacterized HLA alleles [54, 55].

HLA polymorphism also hampers the development of worldwide covering T-cell epitope-based vaccines as HLA variants are expressed at vastly variable frequencies in different ethnic groups [56]. Interestingly, different HLA molecules can also bind similar sets of peptides [57, 58] and researchers have devised methods to cluster them in groups, known as HLA supertypes, consisting of HLA alleles with similar peptide-binding specificities [59–61]. The HLA-A2, HLA-A3, and HLA-B7 are relevant examples of supertypes; 88% of the population expresses at least an allele included in these supertypes [25, 57, 58]. Identification of promiscuous peptide-binding to HLA supertypes enables the development of T-cell epitope vaccines with high-population coverage using a limited number of peptides. Currently, several web-based methods allow the prediction of promiscuous peptide-binding to HLA supertypes for epitope vaccine design including MULTIPRED [62] and PEPVAC [63] (Table 1). A method to identify promiscuous peptide-binding beyond HLA supertypes was developed and implemented by Molero-Abraham et al. [64] with the name of EPI-SOPT. EPI-SOPT predicts HLA I presentation profiles of individual peptides regardless of supertypes and identifies epitope combinations providing a wider population protection coverage.

Prediction of peptide binding to MHC II molecules readily discriminate CD4 T-cell epitopes, but cannot tell their ability to activate the response of specific CD4 T-cell subsets (e.g., Th1, Th2, and Treg). However, there is evidence that some CD4 T-cell epitopes appear to stimulate specific subsets of Th cells [65, 66]. Distinguishing the ability of MHC II-restricted epitopes to elicit distinct responses is clearly relevant for epitope vaccine development and has prompted researchers' attention. A relevant example is the work by Dhanda et al. [67] who generated classifiers capable of predicting potential peptide inducers of interleukin 4 (IL-4) secretion, typical of Th2 cells, by training SVM models on experimentally validated IL4 inducing and noninducing MHC class II binders (Table 1).

2.2. Prediction of Antigen Processing and Integration with Peptide-MHC Binding Prediction. Antigen processing shapes the peptide repertoire available for MHC binding and is a limiting step determining T-cell epitope immunogenicity [68]. Subsequently, computational modeling of the antigen processing pathway provides a mean to enhance T-cell epitope predictions. Antigen presentation by MHC I and II molecules proceed by two different pathways. MHC II molecules present peptide antigens derived from endocytosed antigens that are degraded and loaded onto the MHC II molecule in endosomal compartments [69]. Class II antigen degradation is poorly understood, and there is lack of good prediction algorithms yet [70]. In contrast, MHC I molecules

present peptides derived mainly from antigens degraded in the cytosol. The resulting peptide antigens are then transported to the endoplasmic reticulum by TAP where they are loaded onto nascent MHC I molecules [69] (Figure 4). Prior to loading, peptides often undergo trimming by ERAAP N-terminal amino peptidases [71].

Proteasomal cleavage and peptide-binding to TAP have been studied in detail and there are computational methods that predict both processes. Proteasomal cleavage prediction models have been derived from peptide fragments generated *in vitro* by human constitutive proteasomes [72, 73] and from sets of MHC I-restricted ligands mapped onto their source proteins [74–76]. On the other hand, TAP binding prediction methods have been developed by training different algorithms on peptides of known affinity to TAP [77–80]. Combination of proteasomal cleavage and peptide-binding to TAP with peptide-MHC binding predictions increases T-cell epitope predictive rate in comparison to just peptide-binding to MHC I [37, 77, 81–83]. Subsequently, researchers have developed resources to predict CD8 T-cell epitopes through multistep approaches integrating proteasomal cleavage, TAP transport, and peptide-binding to MHC molecules [26, 37, 82–85] (Table 1).

3. Prediction of B-Cell Epitopes

B-cell epitope prediction aims to facilitate B-cell epitope identification with the practical purpose of replacing the antigen for antibody production or for carrying structure-function studies. Any solvent-exposed region in the antigen can be subject of recognition by antibodies. Nonetheless, B-cell epitopes can be divided in two main groups: linear and conformational (Figure 5). Linear B-cell epitopes consist of sequential residues, peptides, whereas conformational B-cell epitopes consist of patches of solvent-exposed atoms from residues that are not necessarily sequential (Figure 5). Therefore, linear and conformational B-cell epitopes are also known as continuous and discontinuous B-cell epitopes, respectively. Antibodies recognizing linear B-cell epitopes can recognize denatured antigens, while denaturing the antigen results in loss of recognition for conformational B-cell epitopes. Most B-cell epitopes (approximately a 90%) are conformational and, in fact, only a minority of native antigens contains linear B-cell epitopes [3]. We will review both, prediction of linear and conformational B-cell epitopes.

3.1. Prediction of Linear B-Cell Epitopes. Linear B-cell epitopes consist of peptides which can readily be used to replace antigens for immunizations and antibody production. Therefore, despite being a minority, prediction of linear B-cell epitopes have received major attention. Linear B-cell epitopes are predicted from the primary sequence of antigens using sequence-based methods. Early computational methods for the prediction of B-cell epitopes were based on simple amino acid propensity scales depicting physicochemical features of B-cell epitopes. For example, Hopp and Wood applied residue hydrophilicity calculations for B-cell epitope prediction [96, 97] on the assumption that hydrophilic regions are predominantly located on the protein surface and are potentially

antigenic. We know now, however, that protein surfaces contain roughly the same number of hydrophilic and hydrophobic residues [98]. Other amino acid propensity scales introduced for B-cell epitope prediction are based on flexibility [99], surface accessibility [100], and β -turn propensity [101]. Current available bioinformatics tools to predict linear B-cell epitopes using propensity scales include PREDITOP [102] and PEOPLE [103] (Table 2). PREDITOP [102] uses a multiparametric algorithm based on hydrophilicity, accessibility, flexibility, and secondary structure properties of the amino acids. PEOPLE [103] uses the same parameters and in addition includes the assessment of β -turns. A related method to predict B-cell epitopes was introduced by Kolarik and Tongaonkar [104], consisting on a simple antigenicity scale derived from physicochemical properties and frequencies of amino acids in experimentally determined B-cell epitopes. This index is perhaps the most popular antigenicity scale for B-cell epitope prediction, and it is actually implemented by GCG [105] and EMBOSS [106] packages. Comparative evaluations of propensity scales carried out in a dataset of 85 linear B-cell epitopes showed that most propensity scales predicted between 50 and 70% of B-cell epitopes, with the β -turn scale reaching the best values [101, 107]. It has also been shown that combining the different scales does not appear to improve predictions [102, 108]. Moreover, Blythe and Flower [109] demonstrated that single-scale amino acid propensity scales are not reliable to predict epitope location.

The poor performance of amino acid scales for the prediction of linear B-cell epitopes prompted the introduction of machine learning- (ML-) based methods (Table 2). These methods are developed by training ML algorithms to distinguish experimental B-cell epitopes from non-B-cell epitopes. Prior to training, B-cell epitopes are translated into feature vectors capturing selected properties, such as those given by different propensity scales. Relevant examples of B-cell epitope prediction methods based on ML include BepiPred [110], ABCpred [111], LBtope [112], BCPREDS [113], and SVMtrip [114]. Datasets, training features, and algorithms used for developing these methods differ. BepiPred is based on random forests trained on B-cell epitopes obtained from 3D-structures of antigen-antibody complexes [110]. Both BCPREDS [113] and SVMtrip [114] are based on support vector machines (SVM) but while BCPREDS was trained using various string kernels that eliminate the need for representing the sequence into length-fixed feature vectors, SVMtrip was trained on length-fixed tripeptide composition vectors. ABCpred and LBtope methods consist on artificial neural networks (ANNs) trained on similar positive data, B-cell epitopes, but differ on negative data, non-B-cell epitopes. Negative data used for training ABCpred consisted on random peptides while negative data used for LBtope was based on experimentally validated non-B-cell epitopes from IEDB [15]. In general, B-cell epitope prediction methods employing ML-algorithm are reported to outperform those based on amino acid propensity scales. Nevertheless, some authors have reported that ML algorithms show little improvement over single-scale-based methods [115].

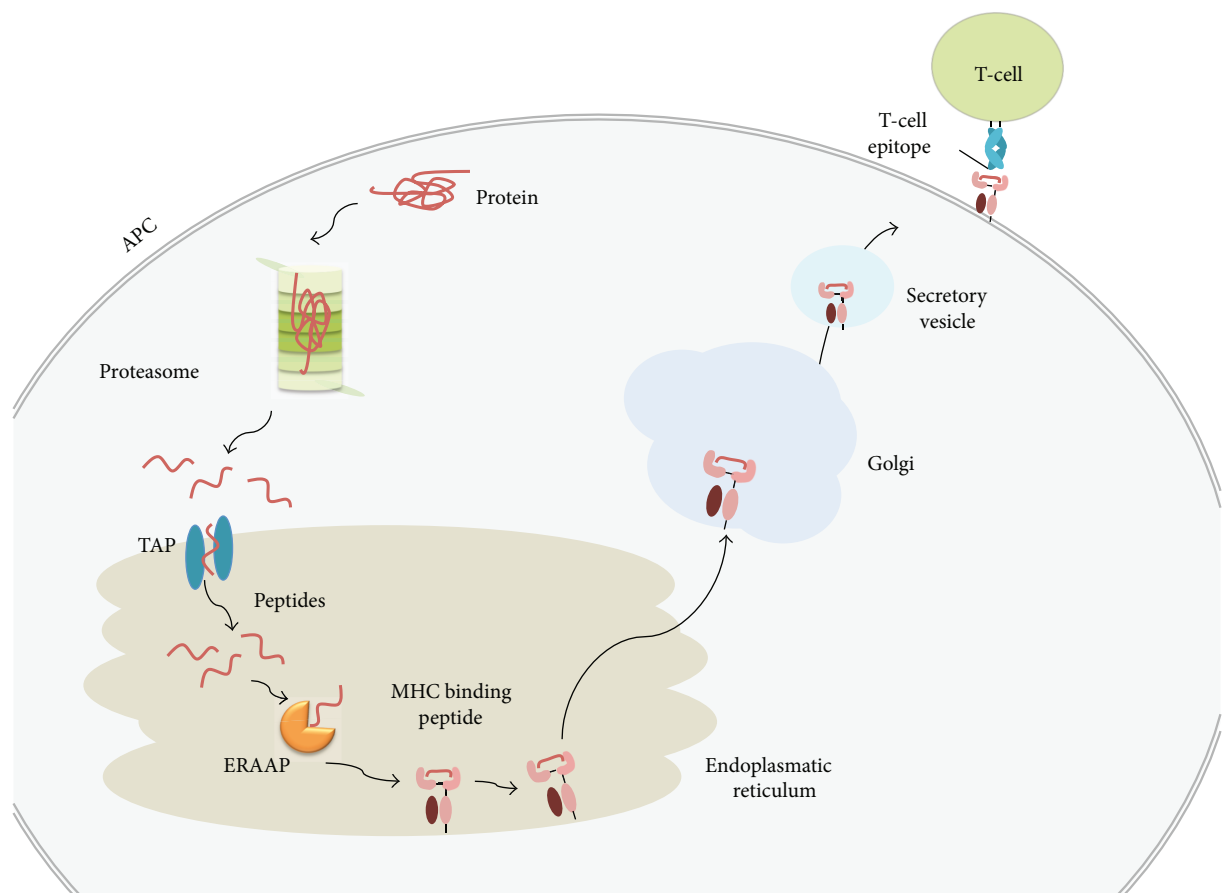


FIGURE 4: Class I antigen processing. The figure depicts the major steps involved in antigen presentation by MHC I molecules. Proteins are degraded by the proteasome and peptide fragments transported to the endoplasmic reticulum (ER) by TAP where they are loaded onto nascent MHC I molecules. TAP transports peptides ranging from 8 to 16 amino acids. Long peptides cannot bind MHC I molecules but often become suitable for binding after N-terminal trimming by ERAAP.

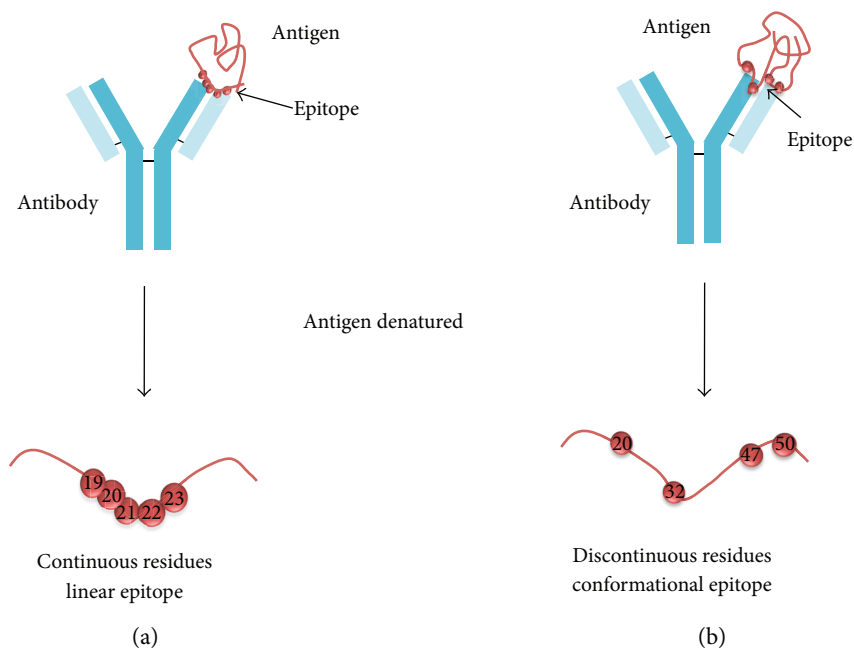


FIGURE 5: Linear and conformational B-cell epitopes. Linear B-cell epitopes (a) are composed of sequential/continuous residues, while conformational B-cell epitopes (b) contain scattered/discontinuous residues along the sequence.

TABLE 2: Selected B-cell epitope prediction methods available for free online use.

Tool	Method	Server (URL)	Ref.
<i>Linear B cell epitope</i>			
PEOPLE	Propensity scale method	http://www.iedb.org/	[103]
BepiPred	ML (DT)	http://www.cbs.dtu.dk/services/BepiPred/	[110]
ABCpred	ML (ANN)	http://www.imtech.res.in/raghava/abcpred/	[111]
LBtope	ML (ANN)	http://www.imtech.res.in/raghava/lbtope/	[112]
BCPREDS	ML (SVM)	http://ailab.ist.psu.edu/bcpred/	[113]
SVMtrip	ML (SVM)	http://sysbio.unl.edu/SVMTriP/prediction.php	[114]
<i>Conformational B-cell epitope</i>			
CEP	Structure-based method (solvent accessibility)	http://bioinfo.ernet.in/cep.htm	[118]
DiscoTope	Structure-based method (surface accessibility and propensity amino acid score)	http://tools.iedb.org/discotope/	[119]
Ellipro	Structure-based method (geometrical properties)	http://tools.iedb.org/ellipro/	[121]
PEPITO	Structure-based method (physicochemical properties and geometrical structure)	http://pepito.proteomics.ics.uci.edu/	[122]
SEPPA	Structure-based method (physicochemical properties and geometrical structure)	http://lifecenter.sgsg.cn/seppa/	[123]
EPITOPIA	Structure-based method (ML-naïve Bayes)	http://epitopia.tau.ac.il/	[125]
EPSVR	Structure-based method (ML-SVR)	http://sysbio.unl.edu/EPSVR/	[126]
EPIPREP	Structure-based method (ASEP, Docking)	http://opig.stats.ox.ac.uk/webapps/sabdab-sabpred/EpiPred.php	[129]
PEASE	Structure-based method (ASEP, ML)	http://www.ofranlab.org/PEASE	[130]
MIMOX	Mimotope	http://immunet.cn/mimox/helps.html	[131]
PEPITOPE	Mimotope	http://pepitope.tau.ac.il/	[132]
EpiSearch	Mimotope	http://curie.utmb.edu/episearch.html	[133]
MIMOPRO	Mimotope	http://informatics.nenu.edu.cn/MimoPro	[134]
CBTOPE	Sequence based (SVM)	http://www.imtech.res.in/raghava/cbtope/submit.php	[136]

Antibodies elicited in the course of an immune response are generally of a given isotype that determines their biological function. A recent advance in B-cell epitope prediction is the development of a method by Gupta et al. [116] that allows the identification of B-cell epitopes capable of inducing specific class of antibodies. This method is based on SMVs trained on a dataset that includes linear B-cell epitopes known to induce IgG, IgE, and IgA antibodies.

3.2. Prediction of Conformational B-Cell Epitopes. Most B-cell epitopes are conformational and yet, prediction of conformational B-cell epitopes has lagged behind that of linear B-cell epitopes. There are two main practical reasons for that. First of all, prediction of conformational B-cell epitopes generally requires the knowledge of protein three-dimensional (3D) structure and this information is only available for a fraction of proteins [117]. Secondly, isolating conformational B-cell epitopes from their protein context for selective antibody production is a difficult task that requires suitable scaffolds for epitope grafting. Thereby, prediction of conformational B-cell prediction is currently of little relevance for epitope vaccine design and antibody-based technologies. Nonetheless, prediction of conformational B-cell epitopes is interesting for carrying structure-function studies involving antibody-antigen interactions.

There are several available methods to predict conformational B-cell epitopes (Table 2). The first to be introduced was CEP [118], which relied almost entirely on predicting patches of solvent-exposed residues. It was followed by DiscoTope [119], which, in addition to solvent accessibility, considered amino acid statistics and spatial information to predict conformational B-cell epitopes. An independent evaluation of these two methods using a benchmark dataset of 59 conformational epitopes revealed that they did not exceed a 40% of precision and a 46% of recall [120]. Subsequently, more methods were developed, like ElliPro [121] that aims to identify protruding regions in antigen surfaces and PEPITO [122] and SEPPA [123] that combine single physicochemical properties of amino acids and geometrical structure properties. The reported area under the curve (AUC) of these methods is around 0.7, which is indicative of a poor discrimination capacity yet better than random. Though, in an independent evaluation, SEPPA reached an AUC of 0.62 while all the mentioned methods had an AUC around 0.5 [124]. ML has also been applied to predict conformational B-cell epitopes in 3D-structures. Relevant examples include EPITOPIA [125] and EPSVR [126] which are based on naïve Bayes classifiers and support vector regressions, respectively, trained on feature vectors combining different scores. The reported AUC of these two methods is around 0.6.

The above methods for conformational B-cell epitope prediction identify generic antigenic regions regardless of antibodies, which are ignored [127]. However, there are also methods for antibody-specific epitope prediction. This approach was pioneered by Soga et al. [128] who defined an antibody-specific epitope propensity (ASEP) index after analyzing the interfaces of antigen-antibody 3D-structures. Using this index, they developed a novel method for predicting epitope residues in individual antibodies that worked by narrowing down candidate epitope residues predicted by conventional methods. More recently, Krawczyk et al. [129] developed EpiPred, a method that uses a docking-like approach to match up antibody and antigen structures, thus identifying epitope regions on the antigen. A similar approach is used by PEASE [130], adding that this method utilizes the sequence of the antibody and the 3D-structure of the antigen. Briefly, for each pair of antibody sequence and antigen structure, PEASE uses a machine learning model trained on properties from 120 antibody-antigen complexes to identify pair combination of residues from complementarity-determining regions (CDRs) of the antibody and the antigen that are likely to interact.

Another approach to identify conformational B-cell epitopes in a protein with a known 3D-structure is through mimotope-based methods. Mimotopes are peptides selected from randomized peptide libraries for their ability to bind to an antibody raised against a native antigen. Mimotope-based methods require to input antibody affinity-selected peptides and the 3D-structure of the selected antigen. Examples of bioinformatics tools for conformational B-cell epitope prediction using mimotopes include MIMOX [131], PEPITOPE [132], EPISEARCH [133], MIMOPRO [134], and PEPMAPPER [135] (Table 2).

As remarked before, methods for conformational B-cell epitope prediction generally require the 3D-structure of the antigen. Exceptionally, however, Ansari and Raghava [136] developed a method (CBTOPE) for the identification of conformational B-cell epitope from the primary sequence of the antigen. CBTOPE is based on SVM and trained on physico-chemical and sequence-derived features of conformational B-cell epitopes. CBTOPE reported accuracy was 86.6% in crossvalidation experiments.

4. Concluding Remarks

Currently, T-cell epitope prediction is more advanced and reliable than that of B-cell prediction. However, while it is possible to confirm experimentally the predicted binding to MHC molecules of most peptides predicted, only ~10% of those are shown to be immunogenic (able to elicit a T-cell response) [68]. Such a low T-cell epitope discovery rate is due to the fact that we do not have adequate models for predicting antigen processing yet [68]. The economic toll of low T-cell epitope discovery rate can be overcome, at least in part, by prioritizing protein antigens for epitope prediction [137–139]. For T-cell epitope vaccine development, researchers can also resort to experimentally known T-cell epitopes, available in epitope databases, selecting through immunoinformatics those that provide maximum

population protection coverage [64, 140, 141]. In any case, T-cell epitope prediction remains an integral part of T-cell epitope mapping approaches. In contrast, B-cell epitope prediction utility is currently much more limited. There are several reasons to that. First of all, prediction of B-cell epitopes is still unreliable for both linear and conformational B-cell epitopes. Secondly, linear B-cell epitopes do usually elicit antibodies that do not crossreact with native antigens. Third, the great majority of B-cell epitopes are conformational and yet predicting conformational epitopes have few applications, as they cannot be isolated from their protein context. Under this scenario, the key is not only to improve current methods for B-cell epitope prediction but also to develop novel approaches and platforms for epitope grafting onto suitable scaffolds capable of replacing the native antigen.

To conclude, we wish to make two final remarks that are relevant for epitope vaccine design. First of all, it is that epitope prediction methods can provide potential epitopes from any given protein query but not all the antigens are equally relevant for vaccine development. Therefore, researchers have also developed tools to identify vaccine candidate antigens [142, 143], those likely to induce protective immunity, which can then be targeted for epitope prediction and epitope vaccine design. Second, it should be borne in mind that epitope peptides exhibit little immunogenicity and need to be used in combination with adjuvants, which increase immunogenicity by inducing strong innate immune responses that enable adaptive immunity [144–146]. Consequently, the discovery of new adjuvants is particularly relevant for epitope-based vaccines [146] and to that end, Nagpal et al. [147] developed a pioneered method that can predict the immunomodulatory activity of RNA sequences.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Authors' Contributions

Jose L. Sanchez-Trincado and Marta Gomez-Perosanz contributed equally to this work.

Acknowledgments

The authors wish to thank *Inmunotek, SL* and the Spanish Department of Science at MINECO for supporting the Immunomedicine group research through Grants SAF2006: 07879, SAF2009:08301, and BIO2014:54164-R to Pedro A. Reche. The authors also wish to thank Dr. Esther M. Lafuente for critical reading and corrections.

References

- [1] W. E. Paul, *Fundamental Immunology*, Lippincott Williams & Wilkins, 2012.
- [2] B. Sun and Y. Zhang, "Overview of orchestration of CD4+ T cell subsets in immune responses," *Advances in Experimental Medicine and Biology*, vol. 841, pp. 1–13, 2014.
- [3] M. H. Van Regenmortel, "What is a B-cell epitope?," *Methods in Molecular Biology*, vol. 524, pp. 3–20, 2009.

- [4] J. Ponomarenko and M. Van Regenmortel, "B-cell epitope prediction," in *Structural Bioinformatics*, pp. 849–879, John Wiley & Sons, Inc, 2009.
- [5] T. A. Ahmad, A. E. Eweida, and L. H. El-Sayed, "T-cell epitope mapping for the design of powerful vaccines," *Vaccine Reports*, vol. 6, pp. 13–22, 2016.
- [6] L. Malherbe, "T-cell epitope mapping," *Annals of Allergy, Asthma & Immunology*, vol. 103, no. 1, pp. 76–79, 2009.
- [7] R. K. Ahmed and M. J. Maeurer, "T-cell epitope mapping," *Methods in Molecular Biology*, vol. 524, pp. 427–438, 2009.
- [8] E. M. Lafuente and P. A. Reche, "Prediction of MHC-peptide binding: a systematic and comprehensive overview," *Current Pharmaceutical Design*, vol. 15, no. 28, pp. 3209–3220, 2009.
- [9] P. E. Jensen, "Recent advances in antigen processing and presentation," *Nature Immunology*, vol. 8, no. 10, pp. 1041–1048, 2007.
- [10] L. J. Stern and D. C. Wiley, "Antigenic peptide binding by class I and class II histocompatibility proteins," *Structure*, vol. 2, no. 4, pp. 245–251, 1994.
- [11] D. R. Madden, "The three-dimensional structure of peptide-MHC complexes," *Annual Review of Immunology*, vol. 13, no. 1, pp. 587–622, 1995.
- [12] D. R. Madden, D. N. Garboczi, and D. C. Wiley, "The antigenic identity of peptide-MHC complexes: a comparison of the conformations of five viral peptides presented by HLA-A2," *Cell*, vol. 75, no. 4, pp. 693–708, 1993.
- [13] D. V. Desai and U. Kulkarni-Kale, "T-cell epitope prediction methods: an overview," *Methods in Molecular Biology*, vol. 1184, pp. 333–364, 2014.
- [14] A. Patronov and I. Doytchinova, "T-cell epitope vaccine design by immunoinformatics," *Open Biology*, vol. 3, no. 1, article 120139, 2013.
- [15] R. Vita, J. A. Overton, J. A. Greenbaum et al., "The immune epitope database (IEDB) 3.0," *Nucleic Acids Research*, vol. 43, D1, pp. D405–D412, 2015.
- [16] M. Molero-Abraham, E. M. Lafuente, and P. Reche, "Customized predictions of peptide-MHC binding and T-cell epitopes using EPIMHC," *Methods in Molecular Biology*, vol. 1184, pp. 319–332, 2014.
- [17] C. P. Toseland, D. J. Clayton, H. McSparron et al., "AntiJen: a quantitative immunology database integrating functional, thermodynamic, kinetic, biophysical, and cellular data," *Immunome Research*, vol. 1, no. 1, p. 4, 2005.
- [18] S. P. Singh and B. N. Mishra, "Major histocompatibility complex linked databases and prediction tools for designing vaccines," *Human Immunology*, vol. 77, no. 3, pp. 295–306, 2016.
- [19] J. D'Amaro, J. G. A. Houbiers, J. W. Drijfhout et al., "A computer program for predicting possible cytotoxic T lymphocyte epitopes based on HLA class I peptide-binding motifs," *Human Immunology*, vol. 43, no. 1, pp. 13–18, 1995.
- [20] M. Bouvier and D. Wiley, "Importance of peptide amino and carboxyl termini to the stability of MHC class I molecules," *Science*, vol. 265, no. 5170, pp. 398–402, 1994.
- [21] J. Ruppert, J. Sidney, E. Celis, R. T. Kubo, H. M. Grey, and A. Sette, "Prominent role of secondary anchor residues in peptide binding to HLA-A2.1 molecules," *Cell*, vol. 74, no. 5, pp. 929–937, 1993.
- [22] M. Nielsen, C. Lundegaard, P. Worning et al., "Improved prediction of MHC class I and class II epitopes using a novel Gibbs sampling approach," *Bioinformatics*, vol. 20, no. 9, pp. 1388–1397, 2004.
- [23] H. G. Rammensee, J. Bachmann, N. P. N. Emmerich, O. A. Bachor, and S. Stevanovic, "SYFPEITHI: database for MHC ligands and peptide motifs," *Immunogenetics*, vol. 50, no. 3–4, pp. 213–219, 1999.
- [24] P. A. Reche, J. P. Glutting, and E. L. Reinherz, "Prediction of MHC class I binding peptides using profile motifs," *Human Immunology*, vol. 63, no. 9, pp. 701–709, 2002.
- [25] P. A. Reche and E. L. Reinherz, "Definition of MHC super-types through clustering of MHC peptide-binding repertoires," *Methods in Molecular Biology*, vol. 409, pp. 163–173, 2007.
- [26] P. A. Reche, J. P. Glutting, H. Zhang, and E. L. Reinherz, "Enhancement to the RANKPEP resource for the prediction of peptide binding to MHC molecules using profiles," *Immunogenetics*, vol. 56, no. 6, pp. 405–419, 2004.
- [27] M. Gribskov and S. Veretnik, "Identification of sequence pattern with profile analysis," *Methods in Enzymology*, vol. 266, pp. 198–212, 1996.
- [28] K. C. Parker, M. A. Bednarek, and J. E. Coligan, "Scheme for ranking potential HLA-A2 binding peptides based on independent binding of individual peptide side-chains," *The Journal of Immunology*, vol. 152, no. 1, pp. 163–175, 1994.
- [29] H. H. Bui, J. Sidney, B. Peters et al., "Automated generation and evaluation of specific MHC binding predictive tools: ARB matrix applications," *Immunogenetics*, vol. 57, no. 5, pp. 304–314, 2005.
- [30] M. Nielsen, C. Lundegaard, and O. Lund, "Prediction of MHC class II binding affinity using SMM-align, a novel stabilization matrix alignment method," *BMC Bioinformatics*, vol. 8, no. 1, p. 238, 2007.
- [31] B. Peters and A. Sette, "Generating quantitative models describing the sequence specificity of biological processes with the stabilized matrix method," *BMC Bioinformatics*, vol. 6, no. 1, p. 132, 2005.
- [32] T. Sturniolo, E. Bono, J. Ding et al., "Generation of tissue-specific and promiscuous HLA ligand databases using DNA microarrays and virtual HLA class II matrices," *Nature Biotechnology*, vol. 17, no. 6, pp. 555–561, 1999.
- [33] B. Peters, W. Tong, J. Sidney, A. Sette, and Z. Weng, "Examining the independent binding assumption for binding of peptide epitopes to MHC-I molecules," *Bioinformatics*, vol. 19, no. 14, pp. 1765–1772, 2003.
- [34] P. Guan, I. A. Doytchinova, C. Zygouri, and D. R. Flower, "MHCpred: a server for quantitative prediction of peptide-MHC binding," *Nucleic Acids Research*, vol. 31, no. 13, pp. 3621–3624, 2003.
- [35] M. Milik, D. Sauer, A. P. Brunmark et al., "Application of an artificial neural network to predict specific class I MHC binding peptide sequences," *Nature Biotechnology*, vol. 16, no. 8, pp. 753–756, 1998.
- [36] V. Brusic, G. Rudy, G. Honeyman, J. Hammer, and L. Harrison, "Prediction of MHC class II-binding peptides using an evolutionary algorithm and artificial neural network," *Bioinformatics*, vol. 14, no. 2, pp. 121–130, 1998.
- [37] P. Donnes and O. Kohlbacher, "Integrated modeling of the major events in the MHC class I antigen processing pathway," *Protein Science*, vol. 14, no. 8, pp. 2132–2140, 2005.
- [38] M. Bhasin and G. P. S. Raghava, "SVM based method for predicting HLA-DRB1*0401 binding peptides in an antigen sequence," *Bioinformatics*, vol. 20, no. 3, pp. 421–423, 2004.

- [39] L. Jacob and J. P. Vert, "Efficient peptide-MHC-I binding prediction for alleles with few known binders," *Bioinformatics*, vol. 24, no. 3, pp. 358–366, 2008.
- [40] S. Zhu, K. Udaka, J. Sidney, A. Sette, K. F. Aoki-Kinoshita, and H. Mamitsuka, "Improving MHC binding peptide prediction by incorporating binding data of auxiliary MHC molecules," *Bioinformatics*, vol. 22, no. 13, pp. 1648–1655, 2006.
- [41] C. J. Savoie, N. Kamikawaji, T. Sasazuki, and S. Kuhara, "Use of BONSAI decision trees for the identification of potential MHC class I peptide epitope motifs," *Pacific Symposium on Biocomputing*, vol. 4, pp. 182–189, 1999.
- [42] H. Mamitsuka, "Predicting peptides that bind to MHC molecules using supervised learning of hidden Markov models," *Proteins*, vol. 33, no. 4, pp. 460–474, 1998.
- [43] C. Zhang, M. G. Bickis, F. X. Wu, and A. J. Kusalik, "Optimally-connected hidden Markov models for predicting MHC-binding peptides," *Journal of Bioinformatics and Computational Biology*, vol. 04, no. 05, pp. 959–980, 2006.
- [44] M. Lacerda, K. Scheffler, and C. Seoighe, "Epitope discovery with phylogenetic hidden Markov models," *Molecular Biology and Evolution*, vol. 27, no. 5, pp. 1212–1220, 2010.
- [45] R. Durbin, S. Eddy, A. Krogh, and G. Mitchison, *Biological Sequence Analysis: Probabilistic Models of Proteins and Nucleic Acids*, Cambridge University Press, Cambridge, 1998.
- [46] W. Liu, X. Meng, Q. Xu, D. R. Flower, and T. Li, "Quantitative prediction of mouse class I MHC peptide binding affinity using support vector machine regression (SVR) models," *BMC Bioinformatics*, vol. 7, no. 1, p. 182, 2006.
- [47] D. R. Jandrlic, "SVM and SVR-based MHC-binding prediction using a mathematical presentation of peptide sequences," *Computational Biology and Chemistry*, vol. 65, pp. 117–127, 2016.
- [48] S. Buus, S. L. Lauemoller, P. Worning et al., "Sensitive quantitative predictions of peptide-MHC binding by a 'query by committee' artificial neural network approach," *Tissue Antigens*, vol. 62, no. 5, pp. 378–384, 2003.
- [49] M. Nielsen, C. Lundegaard, P. Worning et al., "Reliable prediction of T-cell epitopes using neural networks with novel sequence representations," *Protein Science*, vol. 12, no. 5, pp. 1007–1017, 2003.
- [50] M. Nielsen and O. Lund, "NN-align. An artificial neural network-based alignment algorithm for MHC class II peptide binding prediction," *BMC Bioinformatics*, vol. 10, no. 1, p. 296, 2009.
- [51] K. Yu, N. Petrovsky, C. Schonbach, J. Y. Koh, and V. Brusic, "Methods for prediction of peptide binding to MHC molecules: a comparative study," *Molecular Medicine*, vol. 8, no. 3, pp. 137–148, 2002.
- [52] P. Wang, J. Sidney, C. Dow, B. Mothe, A. Sette, and B. Peters, "A systematic assessment of MHC class II peptide binding predictions and evaluation of a consensus approach," *PLoS Computational Biology*, vol. 4, no. 4, article e1000048, 2008.
- [53] P. A. Reche and E. L. Reinherz, "Sequence variability analysis of human class I and class II MHC molecules: functional and structural correlates of amino acid polymorphisms," *Journal of Molecular Biology*, vol. 331, no. 3, pp. 623–641, 2003.
- [54] M. Nielsen, C. Lundegaard, T. Blicher et al., "NetMHCpan, a method for quantitative predictions of peptide binding to any HLA-A and -B locus protein of known sequence," *PLoS One*, vol. 2, no. 8, article e796, 2007.
- [55] M. Nielsen, C. Lundegaard, T. Blicher et al., "Quantitative predictions of peptide binding to any HLA-DR molecule of known sequence: NetMHCIIpan," *PLoS Computational Biology*, vol. 4, no. 7, article e1000107, 2008.
- [56] P. I. Terasaki, "A brief history of HLA," *Immunologic Research*, vol. 38, no. 1-3, pp. 139–148, 2007.
- [57] A. Sette and J. Sidney, "HLA supertypes and supermotifs: a functional perspective on HLA polymorphism," *Current Opinion in Immunology*, vol. 10, no. 4, pp. 478–482, 1998.
- [58] A. Sette and J. Sidney, "Nine major HLA class I supertypes account for the vast preponderance of HLA-A and -B polymorphism," *Immunogenetics*, vol. 50, no. 3-4, pp. 201–212, 1999.
- [59] I. A. Doytchinova and D. R. Flower, "In silico identification of supertypes for class II MHCs," *The Journal of Immunology*, vol. 174, no. 11, pp. 7085–7095, 2005.
- [60] J. Greenbaum, J. Sidney, J. Chung, C. Brander, B. Peters, and A. Sette, "Functional classification of class II human leukocyte antigen (HLA) molecules reveals seven different supertypes and a surprising degree of repertoire sharing across supertypes," *Immunogenetics*, vol. 63, no. 6, pp. 325–335, 2011.
- [61] O. Lund, M. Nielsen, C. Kesmir et al., "Definition of supertypes for HLA molecules using clustering of specificity matrices," *Immunogenetics*, vol. 55, no. 12, pp. 797–810, 2004.
- [62] G. L. Zhang, D. S. DeLuca, D. B. Keskin et al., "MULTI-PRED2: a computational system for large-scale identification of peptides predicted to bind to HLA supertypes and alleles," *Journal of Immunological Methods*, vol. 374, no. 1-2, pp. 53–61, 2011.
- [63] P. A. Reche and E. L. Reinherz, "PEPVAC: a web server for multi-epitope vaccine development based on the prediction of supertypic MHC ligands," *Nucleic Acids Research*, vol. 33, Supplement 2, pp. W138–W142, 2005.
- [64] M. Molero-Abraham, E. M. Lafuente, D. R. Flower, and P. A. Reche, "Selection of conserved epitopes from hepatitis C virus for pan-population stimulation of T-cell responses," *Clinical and Developmental Immunology*, vol. 2013, Article ID 601943, 10 pages, 2013.
- [65] S. L. Constant and K. Bottomly, "Induction of Th1 and Th2 CD4⁺ T cell responses: the alternative approaches," *Annual Review of Immunology*, vol. 15, no. 1, pp. 297–322, 1997.
- [66] E. M. Janssen, A. J. M. van Oosterhout, A. J. M. L. van Rensen, W. van Eden, F. P. Nijkamp, and M. H. M. Wauben, "Modulation of Th2 responses by peptide analogues in a murine model of allergic asthma: amelioration or deterioration of the disease process depends on the Th1 or Th2 skewing characteristics of the therapeutic peptide," *The Journal of Immunology*, vol. 164, no. 2, pp. 580–588, 2000.
- [67] S. K. Dhanda, S. Gupta, P. Vir, and G. P. Raghava, "Prediction of IL4 inducing peptides," *Clinical and Developmental Immunology*, vol. 2013, Article ID 263952, 9 pages, 2013.
- [68] W. Zhong, P. A. Reche, C. C. Lai, B. Reinhold, and E. L. Reinherz, "Genome-wide characterization of a viral cytotoxic T lymphocyte epitope repertoire," *The Journal of Biological Chemistry*, vol. 278, no. 46, pp. 45135–45144, 2003.
- [69] J. S. Blum, P. A. Wearsch, and P. Cresswell, "Pathways of antigen processing," *Annual Review of Immunology*, vol. 31, no. 1, pp. 443–473, 2013.
- [70] E. Hoze, L. Tsaban, Y. Maman, and Y. Louzoun, "Predictor for the effect of amino acid composition on CD4 + T cell

- epitopes preprocessing," *Journal of Immunological Methods*, vol. 391, no. 1-2, pp. 163-173, 2013.
- [71] G. E. Hammer, F. Gonzalez, M. Champsaur, D. Cado, and N. Shastri, "The aminopeptidase ERAAP shapes the peptide repertoire displayed by major histocompatibility complex class I molecules," *Nature Immunology*, vol. 7, no. 1, pp. 103-112, 2006.
 - [72] A. K. Nussbaum, C. Kuttler, K. P. Haderl, H. G. Rammensee, and H. Schild, "PAPProC: a prediction algorithm for proteasomal cleavages available on the WWW," *Immunogenetics*, vol. 53, no. 2, pp. 87-94, 2001.
 - [73] H. G. Holzthutter, C. Frommel, and P. M. Kloetzel, "A theoretical approach towards the identification of cleavage-determining amino acid motifs of the 20s proteasome," *Journal of Molecular Biology*, vol. 286, no. 4, pp. 1251-1265, 1999.
 - [74] M. Nielsen, C. Lundegaard, O. Lund, and C. Kesmir, "The role of the proteasome in generating cytotoxic T-cell epitopes: insights obtained from improved predictions of proteasomal cleavage," *Immunogenetics*, vol. 57, no. 1-2, pp. 33-41, 2005.
 - [75] C. M. Diez-Rivero, E. M. Lafuente, and P. A. Reche, "Computational analysis and modeling of cleavage by the immunoproteasome and the constitutive proteasome," *BMC Bioinformatics*, vol. 11, no. 1, p. 479, 2010.
 - [76] M. Bhasin and G. P. S. Raghava, "Pcleavage: an SVM based method for prediction of constitutive proteasome and immunoproteasome cleavage sites in antigenic sequences," *Nucleic Acids Research*, vol. 33, Supplement 1, pp. W202-W207, 2005.
 - [77] M. Bhasin and G. P. Raghava, "Analysis and prediction of affinity of TAP binding peptides using cascade SVM," *Protein Science*, vol. 13, no. 3, pp. 596-607, 2004.
 - [78] C. M. Diez-Rivero, B. Chenlo, P. Zuluaga, and P. A. Reche, "Quantitative modeling of peptide binding to TAP using support vector machine," *Proteins*, vol. 78, no. 1, pp. 63-72, 2010.
 - [79] S. Daniel, V. Brusic, S. Caillat-Zucman et al., "Relationship between peptide selectivities of human transporters associated with antigen processing and HLA class I molecules," *The Journal of Immunology*, vol. 161, no. 2, pp. 617-624, 1998.
 - [80] V. Brusic, P. van Endert, J. Zeleznikow, S. Daniel, J. Hammer, and N. Petrovsky, "A neural network model approach to the study of human TAP transporter," *In Silico Biology*, vol. 1, no. 2, pp. 109-121, 1999.
 - [81] S. Tenzer, B. Peters, S. Bulik et al., "Modeling the MHC class I pathway by combining predictions of proteasomal cleavage, TAP transport and MHC class I binding," *Cellular and Molecular Life Sciences*, vol. 62, no. 9, pp. 1025-1037, 2005.
 - [82] I. A. Doytchinova, P. Guan, and D. R. Flower, "EpiJen: a server for multistep T cell epitope prediction," *BMC Bioinformatics*, vol. 7, no. 1, p. 131, 2006.
 - [83] M. V. Larsen, C. Lundegaard, K. Lamberth et al., "An integrative approach to CTL epitope prediction: a combined algorithm integrating MHC class I binding, TAP transport efficiency, and proteasomal cleavage predictions," *European Journal of Immunology*, vol. 35, no. 8, pp. 2295-2303, 2005.
 - [84] B. Schubert, H. P. Brachvogel, C. Jurges, and O. Kohlbacher, "EpiToolKit—a web-based workbench for vaccine design," *Bioinformatics*, vol. 31, no. 13, pp. 2211-2213, 2015.
 - [85] B. Schubert, M. Walzer, H. P. Brachvogel, A. Szolek, C. Mohr, and O. Kohlbacher, "FRED 2: an immunoinformatics framework for Python," *Bioinformatics*, vol. 32, no. 13, pp. 2044-2046, 2016.
 - [86] M. Atanasova, A. Patronov, I. Dimitrov, D. R. Flower, and I. Doytchinova, "EpiDOCK: a molecular docking-based tool for MHC class II binding prediction," *Protein Engineering, Design and Selection*, vol. 26, no. 10, pp. 631-634, 2013.
 - [87] J. Hakenberg, A. K. Nussbaum, H. Schild et al., "MAPPP: MHC class I antigenic peptide processing prediction," *Applied Bioinformatics*, vol. 2, no. 3, pp. 155-158, 2003.
 - [88] P. Oyarzun, J. J. Ellis, M. Boden, and B. Kobe, "PREDIVAC: CD4+ T-cell epitope prediction for vaccine design that covers 95% of HLA class II DR protein diversity," *BMC Bioinformatics*, vol. 14, no. 1, p. 52, 2013.
 - [89] Y. He, Z. Xiang, and H. L. T. Mobley, "Vaxign: the first web-based vaccine design program for reverse vaccinology and applications for vaccine development," *Journal of Biomedicine and Biotechnology*, vol. 2010, Article ID 297505, 15 pages, 2010.
 - [90] I. Dimitrov, P. Garnev, D. R. Flower, and I. Doytchinova, "EpiTOP—a proteochemometric tool for MHC class II binding prediction," *Bioinformatics*, vol. 26, no. 16, pp. 2066-2068, 2010.
 - [91] H. Singh and G. P. S. Raghava, "ProPred: prediction of HLA-DR binding sites," *Bioinformatics*, vol. 17, no. 12, pp. 1236-1237, 2001.
 - [92] H. Singh and G. P. Raghava, "ProPred1: prediction of promiscuous MHC class-I binding sites," *Bioinformatics*, vol. 19, no. 8, pp. 1009-1014, 2003.
 - [93] Q. Zhang, P. Wang, Y. Kim et al., "Immune epitope database analysis resource (IEDB-AR)," *Nucleic Acids Research*, vol. 36, Web Server issue, pp. W513-W518, 2008.
 - [94] M. Bhasin and G. P. S. Raghava, "A hybrid approach for predicting promiscuous MHC class I restricted T cell epitopes," *Journal of Biosciences*, vol. 32, no. 1, pp. 31-42, 2007.
 - [95] P. Donnes and A. Elofsson, "Prediction of MHC class I binding peptides, using SVMHC," *BMC Bioinformatics*, vol. 3, no. 1, p. 25, 2002.
 - [96] T. P. Hopp and K. R. Woods, "Prediction of protein antigenic determinants from amino acid sequences," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 78, no. 6, pp. 3824-3828, 1981.
 - [97] T. P. Hopp and K. R. Woods, "A computer program for predicting protein antigenic determinants," *Molecular Immunology*, vol. 20, no. 4, pp. 483-489, 1983.
 - [98] L. Lins, A. Thomas, and R. Brasseur, "Analysis of accessible surface of residues in proteins," *Protein Science*, vol. 12, no. 7, pp. 1406-1417, 2003.
 - [99] P. A. Karplus and G. E. Schulz, "Prediction of chain flexibility in proteins: a tool for the selection of peptide antigen," *Naturwissenschaften*, vol. 72, no. 4, pp. 212-213, 1985.
 - [100] E. A. Emini, J. V. Hughes, D. S. Perlow, and J. Boger, "Induction of hepatitis A virus-neutralizing antibody by a virus-specific synthetic peptide," *Journal of Virology*, vol. 55, no. 3, pp. 836-839, 1985.
 - [101] J. L. Pellequer, E. Westhof, and M. H. V. Van Regenmortel, "Correlation between the location of antigenic sites and the prediction of turns in proteins," *Immunology Letters*, vol. 36, no. 1, pp. 83-99, 1993.

- [102] J. L. Pellequer and E. Westhof, "PREDITOP: a program for antigenicity prediction," *Journal of Molecular Graphics*, vol. 11, no. 3, pp. 204–210, 1993, 191–2.
- [103] A. J. P. Alix, "Predictive estimation of protein linear epitopes by using the program PEOPLE," *Vaccine*, vol. 18, no. 3–4, pp. 311–314, 1999.
- [104] A. S. Kolaskar and P. C. Tongaonkar, "A semi-empirical method for prediction of antigenic determinants on protein antigens," *FEBS Letters*, vol. 276, no. 1–2, pp. 172–174, 1990.
- [105] D. D. Womble, "GCG: the Wisconsin package of sequence analysis programs," *Methods in Molecular Biology*, vol. 132, pp. 3–22, 2000.
- [106] P. Rice, I. Longden, and A. Bleasby, "EMBOSS: the European molecular biology open software suite," *Trends in Genetics*, vol. 16, no. 6, pp. 276–277, 2000.
- [107] J. L. Pellequer, E. Westhof, and M. H. V. Van Regenmortel, "Predicting location of continuous epitopes in proteins from their primary structures," *Methods in Enzymology*, vol. 203, pp. 176–201, 1991.
- [108] M. Odorico and J. L. Pellequer, "BEPITOPE: predicting the location of continuous epitopes and patterns in proteins," *Journal of Molecular Recognition*, vol. 16, no. 1, pp. 20–22, 2003.
- [109] M. J. Blythe and D. R. Flower, "Benchmarking B cell epitope prediction: underperformance of existing methods," *Protein Science*, vol. 14, no. 1, pp. 246–248, 2005.
- [110] M. C. Jespersen, B. Peters, M. Nielsen, and P. Marcatili, "BepiPred-2.0: improving sequence-based B-cell epitope prediction using conformational epitopes," *Nucleic Acids Research*, vol. 45, Web Server issue, pp. W24–W29, 2017.
- [111] S. Saha and G. P. S. Raghava, "Prediction of continuous B-cell epitopes in an antigen using recurrent neural network," *Proteins*, vol. 65, no. 1, pp. 40–48, 2006.
- [112] H. Singh, H. R. Ansari, and G. P. S. Raghava, "Improved method for linear B-cell epitope prediction using antigen's primary sequence," *PLoS One*, vol. 8, no. 5, article e62216, 2013.
- [113] Y. El-Manzalawy, D. Dobbs, and V. Honavar, "Predicting linear B-cell epitopes using string kernels," *Journal of Molecular Recognition*, vol. 21, no. 4, pp. 243–255, 2008.
- [114] B. Yao, L. Zhang, S. Liang, and C. Zhang, "SVMTriP: a method to predict antigenic epitopes using support vector machine to integrate tri-peptide similarity and propensity," *PLoS One*, vol. 7, no. 9, article e45152, 2012.
- [115] J. A. Greenbaum, P. H. Andersen, M. Blythe et al., "Towards a consensus on datasets and evaluation metrics for developing B-cell epitope prediction tools," *Journal of Molecular Recognition*, vol. 20, no. 2, pp. 75–82, 2007.
- [116] S. Gupta, H. R. Ansari, A. Gautam, Open Source Drug Discovery Consortium, and G. P. Raghava, "Identification of B-cell epitopes in an antigen for inducing specific class of antibodies," *Biology Direct*, vol. 8, no. 1, p. 27, 2013.
- [117] M. Levitt, "Nature of the protein universe," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 106, no. 27, pp. 11079–11084, 2009.
- [118] U. Kulkarni-Kale, S. Bhosle, and A. S. Kolaskar, "CEP: a conformational epitope prediction server," *Nucleic Acids Research*, vol. 33, Web Server issue, pp. W168–W171, 2005.
- [119] P. Haste Andersen, M. Nielsen, and O. Lund, "Prediction of residues in discontinuous B-cell epitopes using protein 3D structures," *Protein Science*, vol. 15, no. 11, pp. 2558–2567, 2006.
- [120] J. V. Ponomarenko and P. E. Bourne, "Antibody-protein interactions: benchmark datasets and prediction tools evaluation," *BMC Structural Biology*, vol. 7, no. 1, p. 64, 2007.
- [121] J. Ponomarenko, H. H. Bui, W. Li et al., "ElliPro: a new structure-based tool for the prediction of antibody epitopes," *BMC Bioinformatics*, vol. 9, no. 1, p. 514, 2008.
- [122] M. J. Sweredoski and P. Baldi, "PEPITO: improved discontinuous B-cell epitope prediction using multiple distance thresholds and half sphere exposure," *Bioinformatics*, vol. 24, no. 12, pp. 1459–1460, 2008.
- [123] J. Sun, D. Wu, T. Xu et al., "SEPPA: a computational server for spatial epitope prediction of protein antigens," *Nucleic Acids Research*, vol. 37, no. suppl_2, Web Server issue, pp. W612–W616, 2009.
- [124] X. Xu, J. Sun, Q. Liu et al., "Evaluation of spatial epitope computational tools based on experimentally-confirmed dataset for protein antigens," *Chinese Science Bulletin*, vol. 55, no. 20, pp. 2169–2174, 2010.
- [125] N. D. Rubinstein, I. Mayrose, E. Martz, and T. Pupko, "Epitopia: a web-server for predicting B-cell epitopes," *BMC Bioinformatics*, vol. 10, no. 1, p. 287, 2009.
- [126] S. Liang, D. Zheng, D. M. Standley, B. Yao, M. Zacharias, and C. Zhang, "EPSVR and EPMeta: prediction of antigenic epitopes using support vector regression and multiple server results," *BMC Bioinformatics*, vol. 11, no. 1, p. 381, 2010.
- [127] I. Sela-Culang, Y. Ofra, and B. Peters, "Antibody specific epitope prediction — emergence of a new paradigm," *Current Opinion in Virology*, vol. 11, pp. 98–102, 2015.
- [128] S. Soga, D. Kuroda, H. Shirai, M. Kobori, and N. Hirayama, "Use of amino acid composition to predict epitope residues of individual antibodies," *Protein Engineering, Design and Selection*, vol. 23, no. 6, pp. 441–448, 2010.
- [129] K. Krawczyk, X. Liu, T. Baker, J. Shi, and C. M. Deane, "Improving B-cell epitope prediction and its application to global antibody-antigen docking," *Bioinformatics*, vol. 30, no. 16, pp. 2288–2294, 2014.
- [130] I. Sela-Culang, S. Ashkenazi, B. Peters, and Y. Ofra, "PEASE: predicting B-cell epitopes utilizing antibody sequence," *Bioinformatics*, vol. 31, no. 8, pp. 1313–1315, 2015.
- [131] J. Huang, A. Gutteridge, W. Honda, and M. Kanehisa, "MIMOX: a web tool for phage display based epitope mapping," *BMC Bioinformatics*, vol. 7, no. 1, p. 451, 2006.
- [132] I. Mayrose, O. Penn, E. Erez et al., "Pepitope: epitope mapping from affinity-selected peptides," *Bioinformatics*, vol. 23, no. 23, pp. 3244–3246, 2007.
- [133] S. S. Negi and W. Braun, "Automated detection of conformational epitopes using phage display peptide sequences," *Bioinformatics and Biology Insights*, vol. 3, pp. 71–81, 2009.
- [134] W. Chen, P. Sun, Y. Lu, W. W. Guo, Y. Huang, and Z. Ma, "MimoPro: a more efficient web-based tool for epitope prediction using phage display libraries," *BMC Bioinformatics*, vol. 12, no. 1, p. 199, 2011.
- [135] W. Chen, W. W. Guo, Y. Huang, and Z. Ma, "PepMapper: a collaborative web tool for mapping epitopes from affinity-selected peptides," *PLoS One*, vol. 7, no. 5, article e37869, 2012.
- [136] H. Ansari and G. P. S. Raghava, "Identification of conformational B-cell epitopes in an antigen from its primary sequence," *Immunome Research*, vol. 6, no. 1, p. 6, 2010.

- [137] C. M. Diez-Rivero and P. A. Reche, "CD8 T cell epitope distribution in viruses reveals patterns of protein biosynthesis," *PLoS One*, vol. 7, no. 8, article e43674, 2012.
- [138] D. R. Flower, I. K. Macdonald, K. Ramakrishnan, M. N. Davies, and I. A. Doytchinova, "Computer aided selection of candidate vaccine antigens," *Immunome Research*, vol. 6, Supplement 2, p. S1, 2010.
- [139] M. Molero-Abraham, J. P. Glutting, D. R. Flower, E. M. Lafuente, and P. A. Reche, "EPIPOX: immunoinformatic characterization of the shared T-cell epitome between variola virus and related pathogenic orthopoxviruses," *Journal of Immunology Research*, vol. 2015, Article ID 738020, 11 pages, 2015.
- [140] P. A. Reche, D. B. Keskin, R. E. Hussey, P. Ancuta, D. Gabuzda, and E. L. Reinherz, "Elicitation from virus-naïve individuals of cytotoxic T lymphocytes directed against conserved HIV-1 epitopes," *Medical Immunology*, vol. 5, no. 1, p. 1, 2006.
- [141] Q. M. Sheikh, D. Gatherer, P. A. Reche, and D. R. Flower, "Towards the knowledge-based design of universal influenza epitope ensemble vaccines," *Bioinformatics*, vol. 32, no. 21, pp. 3233–3239, 2016.
- [142] S. J. Goodswen, P. J. Kennedy, and J. T. Ellis, "Vacceed: a high-throughput *in silico* vaccine candidate discovery pipeline for eukaryotic pathogens based on reverse vaccinology," *Bioinformatics*, vol. 30, no. 16, pp. 2381–2383, 2014.
- [143] M. Rizwan, A. Naz, J. Ahmad et al., "VacSol: a high throughput *in silico* pipeline to predict potential therapeutic targets in prokaryotic pathogens using subtractive reverse vaccinology," *BMC Bioinformatics*, vol. 18, no. 1, p. 106, 2017.
- [144] H. Yang and D. S. Kim, "Peptide immunotherapy in vaccine development: from epitope to adjuvant," *Advances in Protein Chemistry and Structural Biology*, vol. 99, pp. 1–14, 2015.
- [145] A. Di Pasquale, S. Preiss, F. Tavares Da Silva, and N. Garçon, "Vaccine adjuvants: from 1920 to 2015 and beyond," *Vaccines*, vol. 3, no. 2, pp. 320–343, 2015.
- [146] F. Azmi, A. A. Ahmad Fuaad, M. Skwarczynski, and I. Toth, "Recent progress in adjuvant discovery for peptide-based subunit vaccines," *Human Vaccines & Immunotherapeutics*, vol. 10, no. 3, pp. 778–796, 2014.
- [147] G. Nagpal, K. Chaudhary, S. K. Dhanda, and G. P. S. Raghava, "Computational prediction of the immunomodulatory potential of RNA sequences," *Methods in Molecular Biology*, vol. 1632, pp. 75–90, 2017.

