

Review

Predicting Immunogenicity Risk in Biopharmaceuticals

Nikolet Doneva , Irini Doytchinova  and Ivan Dimitrov * 

Faculty of Pharmacy, Medical University of Sofia, 1000 Sofia, Bulgaria; ndoneva@pharmfac.mu-sofia.bg (N.D.); idoytchinova@pharmfac.mu-sofia.bg (I.D.)

* Correspondence: idimitrov@pharmfac.mu-sofia.bg

Abstract: The assessment of immunogenicity of biopharmaceuticals is a crucial step in the process of their development. Immunogenicity is related to the activation of adaptive immunity. The complexity of the immune system manifests through numerous different mechanisms, which allows the use of different approaches for predicting the immunogenicity of biopharmaceuticals. The direct experimental approaches are sometimes expensive and time consuming, or their results need to be confirmed. In this case, computational methods for immunogenicity prediction appear as an appropriate complement in the process of drug design. In this review, we analyze the use of various *In silico* methods and approaches for immunogenicity prediction of biomolecules: sequence alignment algorithms, predicting subcellular localization, searching for major histocompatibility complex (MHC) binding motifs, predicting T and B cell epitopes based on machine learning algorithms, molecular docking, and molecular dynamics simulations. Computational tools for antigenicity and allergenicity prediction also are considered.

Keywords: immunogenicity; *In silico*; bioinformatics; machine learning; epitope



Citation: Doneva, N.; Doytchinova, I.; Dimitrov, I. Predicting Immunogenicity Risk in Biopharmaceuticals. *Symmetry* **2021**, *13*, 388. <https://doi.org/10.3390/sym13030388>

Academic Editors: Jan Awrejcewicz and Dumitru Baleanu

Received: 4 February 2021

Accepted: 23 February 2021

Published: 27 February 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Biopharmaceuticals are drug therapy products of biological origins such as living cells or viruses, usually derived with the methods of biotechnology. One of the main obstacles in using biological drugs for new therapies is the unwanted immunogenicity. On the other hand, immunogenicity is a keystone in vaccine development. When a foreign substance is exposed for the first time to the immune systems, it induces a specific immune response. This ability is known as immunogenicity [1]. The development of methods for the assessment of immunogenicity requires deep understanding of the mechanisms of immune response activation.

The immune system can be divided in two main branches: innate and adaptive immunity. The main difference between them is in how they recognize the pathogens. The innate immunity system uses nonspecific defense mechanisms and is represented by eosinophils, monocytes, macrophages, natural killer cells, complement system. The chemical properties of the antigens activate the innate immune response, which induces inflammation on the site.

The adaptive immunity is induced by the activation of cellular and/or humoral responses [1,2]. The T cell activation is a decisive step in both responses. The T cell receptor (TCR) is a specific receptor, which enables the recognition of antigens bounded to major histocompatibility complex (MHC) molecules. The MHC molecules present antigens (intracellular or extracellular) on the surface of the antigen-presenting cells (APCs). MHC proteins in humans are extremely polymorphic. There are three gene loci that encode the MHC class I proteins in humans: HLA-A, HLA-B and HLA-C, and other three loci that encode the MHC class II proteins: HLA-DR, HLA-DQ and HLA-DP. The MHC class I molecules bind to intracellular antigens and present them on the surface of all nucleated cells [3–5]. The MHC class II molecules present extracellular antigens and can be found on

macrophages, dendritic cells and B cells that have phagocytosed pathogens (e.g., parasites, bacteria, toxins, allergens, etc.) [3,6–8].

The studies on the mechanisms of immune functions, immunological interactions, and related disease pathogenesis generate a huge amount of data that can be managed thanks to the immunoinformatics and computational immunology [9]. The immunological databases and prediction software have inevitable roles in the immunological research, allowing the researchers to identify the molecules involved in immune response in an easier and faster way [10]. The recognition of the immune responses is accelerated and facilitated by the usage of computational tools, of which development is impacted by the implementation of immunological databases. In the past, vaccine development has relied on experiments, both immunological and biochemical, cultivating microorganisms, which were time-consuming and cost-effective. These disadvantages have been largely circumvented using *In Silico* analysis and immunoinformatics. A major contribution in the efforts for vaccine design, disease prevention, diagnosis and treatment has been the immunogenic epitopes predictions and the virtual immunogenicity assessments [11]. However, the *In Silico* modelling could not fully replace the *In Vitro* and *In Vivo* confirmation of immunogenicity during the drug development pathway.

The predictive immunogenicity screening has its strengths and weaknesses. For relative immunogenicity ranking, most useful are the *In Vivo* models. However, due to the basic differences between human and animal immune systems, it is very difficult to use *In Vivo* models for the prediction of immunogenicity incidence rate. Compared with *In Vivo* testing, *In Vitro* systems and computational tools are more cost effective and less time consuming. The *In Vitro* assays can be used for investigation, identification (e.g., immunogenic epitopes) and evaluation (e.g., MHC affinity) of the cellular immune mechanism of response. However, the attempt to replicate the immune system *Ex Vivo* is one of the main challenges in the *In Vitro* immunogenicity prediction. For example, it is difficult to predict the 2D epitopes and 3D protein structure interplay and how the protein will be digested by the APC in order to present the epitopes via MHC. *In silico* programs are used for the identification of immunogenic regions to narrow down epitope candidates [12]. The *In Silico* methods are relatively low-cost and high-throughput, which are significant advantages in using them to predict both B and T cell epitopes and assist in the selection of lead proteins (based on predicted immunogenicity). Moreover, in the early stages of drug development, the *In Silico* analysis can be used as an aid for the rational design of novel protein therapeutics, avoiding the B and T cell epitopes [13,14].

Currently, the gold standard for the *In Silico* prediction of immunogenicity is the usage of machine learning algorithms. The machine learning methods turn the collected input data into useful information. The main goal is to analyze the input data in order to predict what the response is going to be for future new/unknown input variables. Moreover, the machine learning algorithms aim to extract information about the process relationship between the input and the response variables. The machine learning approaches used in the *In Silico* analysis are divided in two main categories: supervised learning and unsupervised learning. In the supervised learning, we need “training data”, which are pairs of input and output variables. The task is to “learn” the model to predict the output value for new, previously unseen data after having “seen” a number of training examples. Some of the most common algorithms used in supervised learning are linear regression, partial least squares, neural networks, support vector machines, and random forest. In the unsupervised learning, the main task is to determine how the data are organized—find relationships, correlations, cluster the information, and find hidden patterns. Some of the most common algorithms used in unsupervised learning are clustering (hierarchical and *k*-means) and principal component analysis.

2. Classification of the Methods for Immunogenicity Prediction

There are two main groups of *In Silico* methods used for the creation and development of prediction models. According to the available data, the *In Silico* methods are classified as

structure-based and sequence-based. Structure-based methods use the data from the three-dimensional (3D) structure of the proteins. Sequence-based methods analyze the amino acid (aa) sequence. The list of approaches includes QSAR (quantitative structure–activity relationship) analysis, matrix-driven methods, protein threading, homology modeling, identification of structural binding motifs, docking techniques, and the application of machine-learning algorithms and tools [5].

2.1. Sequence-Based Methods

The first computational methods used for sequence-based binding affinity prediction were sequence motifs [15] and quantitative matrices [16]. They were quickly replaced by the gold standard—machine learning algorithms [17–19]. Sequence-based methods have some limitations. In the statistical learning methods, in order to train the model, an experimental (training) dataset is required. The composition of the dataset is very important because it can bias the predictions [20,21]. Therefore, it is more reliable to predict MHC variants (i.e., alleles) with larger datasets available for training than to predict less studied alleles.

2.2. Structure-Based Methods

A training dataset is not required, because these methods rely on the peptide–MHC (pMHC) interaction and the biochemical properties of the amino acids involved in it [20]. The structure-based methods can be used to analyze the impact on the binding affinity and immunogenicity of MHC-binding peptides after phosphorylation [22], citrullination [23], and glycosylation [24]. The latter are post-translational amino acid modifications, which may affect the binding affinity. In addition, the structural basis of TCR/pMHC interactions can be detailed. One of the main challenges in the structure-based methods is the obtaining and processing of three-dimensional structural data. In order to conduct personalized structure-based analyses, the structural prediction and/or modeling should be performed using computational methods and tools. However, from a computational perspective, the pMHC modeling and structure-based binding affinity prediction is challenged by the size and flexibility of the ligands involved [25]. In order to perform the required structural analyses in an efficient way, we should rely on the expert knowledge and/or experimental data that we already have [20,25].

3. Methods for Immunogenicity Prediction

The complexity of the immune system allows the application of different approaches for immunogenicity predictions.

3.1. Sequence Alignment Algorithms

The sequence alignment algorithms compare two or more sequences trying to highlight similarities between them. These similarities can be a result of different relationships between the compared sequences—structural, functional, or evolutionary. The aim of the sequence alignments is to organize, visualize, search, and analyze the sequences. The sequence alignment algorithms are applied for immunogenicity prediction by screening databases and comparison of proteins or peptides with known antigens.

The application of sequence alignment algorithms for the identification of antigens is problematic for several reasons and may produce ambiguous results or fail. For example, some proteins which do not have obvious sequence similarity and are formed through evolution (divergent or convergent), may share similar properties and structure [26]. Moreover, in some cases the antigenicity (as a property) may not be available for direct identification using sequence alignment methods because it is encoded in a subtle sequence.

3.2. Predicting Subcellular Localization

One of the basic assumptions for the search of vaccine antigen candidates is that a potential antigen is located at the bacterial surface where it is available for antibody recogni-

tion. The N-terminal of a protein contains an indication of its cellular destination. A “leader sequence” is an indication that the protein will be exported to extracytoplasmic compartments. Additional signature is when the cleavage site is just after the leader peptide—a sign that the protein in Gram-positive bacteria will be released to the extracellular space, or in Gram-negative bacteria into the periplasmic space [27].

A few tools for the prediction of the subcellular localization exist:

- PsortB (<https://www.psort.org/psortb/> (accessed on 27 February 2021)) consists of multiple analytical modules, each of which analyzes one biological feature known to influence or be characteristic of subcellular localization. This tool provides predictions of protein subcellular localization for both Gram-positive and Gram-negative bacteria [28];
- SignalP (<http://www.cbs.dtu.dk/services/SignalP/> (accessed on 27 February 2021)) is a neural network-based method, which can discriminate signal peptides from trans membrane regions. It predicts signal peptide cleavage sites, their presence and location in the amino acid sequence [29];
- PSLpred (<http://crdd.osdd.net/raghava/pslpred/> (accessed on 27 February 2021)) is a web server for predicting the subcellular localization of Gram-negative bacterial proteins based on support vector machine modules [30].

3.3. T and B Cell Epitopes Prediction

The epitope is a part of an antigen, comprising a few amino acids residues and can be recognized by the host immune system (B or T cells). The epitopes can elicit an immune response, which is why it is important to identify or predict them [31].

The main goal when predicting the epitopes is to investigate their binding affinity to the MHC molecules, which is a prerequisite for peptide immunogenicity [2]. The complex of peptide and MHC proteins is involved in the known mechanisms of both cellular and humoral immune response. For the contemporary identification of epitopes, several In Silico methodologies have been created and constantly improved, because the experimental techniques were found to be time-consuming, difficult, and expensive. There are different databases storing the immunological data, which is necessary for the development of computational methods. Some of the most reliable and up-to-date are presented in Table 1.

Table 1. Databases containing immunological information.

Database	Content	Reference
MHCBN http://crdd.osdd.net/raghava/mhcbn/ (accessed on 27 February 2021)	Binding, non-binding and T cell epitopes.	[32]
AntiJen http://www.ddg-pharmfac.net/antijen/AntiJen/antijenhomepage.htm (accessed on 27 February 2021)	Data for thermodynamics and kinetics. Functional and cellular data in the context of vaccinology and immunology.	[33]
IEDB (Immune Epitope Database) www.iedb.org (accessed on 27 February 2021)	B and T cell epitopes which are characterized experimentally, as well as major histocompatibility complex (MHC) binding data and MHC ligand elution experimental data.	[34]
IMGT (International ImMunoGeneTics information system) http://www.imgt.org/ (accessed on 27 February 2021)	Immunoglobulins, T cell receptors, MHC proteins of human and other vertebrate species.	[35]
ImmPort https://www.import.org/shared (accessed on 27 February 2021)	Archival repository and dissemination vehicle for clinical and molecular datasets.	[36]
BCIPEP http://crdd.osdd.net/raghava/bcipep/ (accessed on 27 February 2021)	B cell epitope database with 3031 peptide entries.	[37]
SYFPEITHI http://www.syfpeithi.de/ (accessed on 27 February 2021)	Contains more than 7000 peptide sequences known to bind class I and class II MHC molecules.	[15]

3.3.1. T Cell Epitope Prediction

The MHC class II binding groove is open-ended, and peptides of varying sizes can accommodate in it. Therefore, it is considered important the influence of flanking residues (p. 1, p. 10, and p. 11) which are positioned outside of the core 9-mer [38,39].

The potency of a peptide to stimulate a T cell response In Vivo is not determined by the affinity of a peptide to MHC class II, although a limited number of models have demonstrated a correlation between predicted MHC class II binding and T cell epitopes [40]. All T cell epitopes can bind to MHC class II, however not all MCH binders are T cell epitopes. The In Silico methods are a useful aid to identify peptide sequences which potentially are MHC class II binders. The number of actual T cell epitopes in a certain protein sequence are constantly over-predicted by the In Silico methods. Therefore, normally an Ex Vivo or In Vivo T cell activation analysis is required in order to determine which are the actual T cell epitopes and to rank the epitopes based on their potency to induce a T cell response [41].

There are many obstacles for the high prediction accuracy of MHC-II-binding epitopes: insufficient or low-quality data, the limitation of the binding core length and difficulty in its identification, the influence of the residues flanking the core, and the permissiveness of the binding groove [42,43].

Sequence-Based Methods

Some of the web tools for MHC binding and T cell epitope prediction are listed in Table 2.

Table 2. Web tools for MHC binding and T cell epitope prediction.

Tool	Description	Method	Reference
SYFPEITHI http://www.syfpeithi.de/ (accessed on 27 February 2021)	Scores the peptides and evaluates their immunogenicity.	Motif based	[15]
RankPep http://imed.med.ucm.es/Tools/rankpep.html (accessed on 27 February 2021)	Ranks peptides using the position-specific scoring matrix coefficients. The method is validated with proteins known to bear MHC class I T cell epitopes. Performance and result analysis show that more than 80% of these epitopes are among the top 2% of scoring peptides.	Motif-based	[44]
PREDIVAC http://predivac.biosci.uq.edu.au/ (accessed on 27 February 2021)	T cell epitope mapping. Covers binding to 95% of HLA class II protein from DR locus.	Motif-based	[45]
PEPVAC http://imed.med.ucm.es/PEPVAC/ (accessed on 27 February 2021)	Predicts binders of HLA class I molecules with similar peptide-binding specificity (supertypes) and identifies conserved MHC ligands, as well as those with a C-terminus resulting from proteasomal cleavage.	Motif-based	[46]
Propred http://crdd.osdd.net/raghava/propred/ (accessed on 27 February 2021)	Predicts binders to HLA DR alleles.	Quantitative matrices-driven (QM)	[47]
TepitopePan http://datamining-iip.fudan.edu.cn/service/TEPITOPEpan/TEPITOPEpan.html (accessed on 27 February 2021)	Identifies peptides binding to HLA DR alleles and recognizes binding cores.	QM	[48]
EpiJen http://www.ddg-pharmfac.net/epijen/EpiJen/EpiJen.htm (accessed on 27 February 2021)	Predicts T cell epitopes through multi-step algorithm proteasome cleavage, TAP transport, binding to MHC class I, and epitope selection.	QM	[49]
MMBpred http://www.intech.res.in/raghava/mmbpred/ (accessed on 27 February 2021)	Identifies binders to 47 MHC class I alleles by introducing mutations in an antigenic sequence.	QM	[50]
NetMHCII http://www.cbs.dtu.dk/services/NetMHCII-2.0/ (accessed on 27 February 2021)	Uses SMM align method in order to predict the MHC binding affinities to 14 HLA-DR and 3 mouse H2-IA alleles.	QM	[51]
NetCTL http://www.cbs.dtu.dk/services/NetCTL/ (accessed on 27 February 2021)	Integrates prediction of binding to 12 MHC class I alleles and proteasomal C-terminal cleavage using artificial neural networks (ANN) and TAP transport efficiency using weight matrix.	ANN	[52]

Table 2. Cont.

Tool	Description	Method	Reference
NetMHC http://www.cbs.dtu.dk/services/NetMHC-4.0/ (accessed on 27 February 2021)	Predicts binding affinity to 81 different MHC class I alleles, including HLA-A, -B, -C and -E as well as 41 animal (monkey, cattle, pig, and mouse) alleles, using gapped sequence alignment and ANN.	ANN	[53]
NetChop http://www.cbs.dtu.dk/services/NetChop/ (accessed on 27 February 2021)	Predicts the cleavage sites of the human proteasome.	ANN	[54]
NetMHCII http://www.cbs.dtu.dk/services/NetMHCII/ (accessed on 27 February 2021)	Predicts binding of peptides to 25 HLA-DR alleles, 20 HLA-DQ, 9 HLA-DP, and 7 mouse H2 class II alleles.	ANN	[55]
NetMHCIIpan http://www.cbs.dtu.dk/services/NetMHCIIpan/ (accessed on 27 February 2021)	Predicts peptide binding to any MHC II molecule of known sequence, including human MHC class II isotypes HLA-DR, HLA-DQ, HLA-DP, and mouse alleles (H-2).	ANN	[56]
TAPPRED http://crdd.osdd.net/raghava/tappred/ (accessed on 27 February 2021)	Predicts the binding affinity of TAP protein. A cascade support vector machine (SVM; which uses amino acid features along with sequence) was used to improve the reliability.	SVM	[57]
Pcleavage http://crdd.osdd.net/raghava/pcleavage/ (accessed on 27 February 2021)	Predicts the immunoproteasome and constitutive cleavage sites in antigenic sequences.	SVM	[58]
SVRMHC http://c1.accurascience.com/SVRMHCdb/ (accessed on 27 February 2021)	Uses a quantitative SVM regression approach. The similarities in the amino acids physicochemical properties are taken into account.	SVM	[59]
WAPP https://kohlbacherlab.org/ (accessed on 27 February 2021)	Combines matrix-based prediction of proteasomal cleavage, MHC binding and TAP transport into a single prediction system. In order to predict peptides transported by TAP a SVM regression method is used.	SVM	[60]
MHCPRED http://www.ddg-pharmfac.net/mhcpred/MHCPred (accessed on 27 February 2021)	Composed of a number of allele specific quantitative structure–activity relationship (QSAR) models created using partial least squares. An additive method is used in order to predict the binding affinity between MHC class I and II and the TAP.	QSAR	[61]
EpiTOP http://www.pharmfac.net/EpiTOP (accessed on 27 February 2021)	Predicts MHC class II binding affinity based on proteochemometrics (a QSAR approach for ligands binding to several HLA alleles). Predicts the binding affinity to the most common HLA alleles by partial least squares derived models.	QSAR	[62]
CTLpred http://crdd.osdd.net/raghava/ctlpred/ (accessed on 27 February 2021)	Predicts MHC class I restricted T cell epitopes.	Combined QM, SVM and ANN	[63]
Deimmunization http://tools.iedb.org/deimmunization/ (accessed on 27 February 2021)	Identifies immunodominant regions in a given therapeutically important protein and suggests amino acid substitutions in order to create non-immunogenic versions of the proteins.	Combined	[64]
Class I immunogenicity http://tools.iedb.org/immunogenicity/ (accessed on 27 February 2021)	Predicts immunogenicity of pMHC by using amino acid properties and their position within the peptide.	QM	[65]
C-ImmSim http://www.cbs.dtu.dk/services/C-ImmSim-10.1/ (accessed on 27 February 2021)	Calculates the immune response. The lymphocyte receptors and the pathogens are represented by their amino acid sequences.	ANN	[66]

Motif Search-Based Approach

A motif is a preferred amino acid combination at some of the peptide binding positions. Motifs can be searched in the peptide sequence by using a motif library [41]. If we know and compare the binders and non-binders for a given peptide, we can identify the MHC-binding motifs [67]. The motif-based algorithm's accuracy is about 60–70%. The reason for this relatively low accuracy is an absence of recognizable motifs in some binding peptides [68]. Quantitative matrices-driven methods (QMs) look like an extended motif,

where for every amino acid in a peptide a coefficient is imputed [69]. The estimation of QM is made by specific binding profiles, which are derived experimentally.

Prediction by Statistical and Machine Learning Algorithms

The development of computer science and the need to analyze an increased amount of data has led to the development of machine learning algorithms. The main challenges in working with proteins as sequence data are to represent amino acid sequence using appropriate numerical descriptors, and how sequences with different lengths should be compared. Once there is success with these challenges, a huge arsenal of machine learning methods can be applied in order to find reliable models.

Artificial Neural Networks (ANN)

The ANNs provide a convenient method to find relationships and describe nonlinear data [70]. ANNs are based on a simple model of a neuron: dendrites that collect inputs from other neurons, a soma that performs a nonlinear processing step, and an axon that transmit the signal to another neuron. The base for the learning process is the connections between the elements: dendrites (summation of input), soma (threshold, nonlinearity), and axon (transmission of output). The peptide length can be highly variable when we are applying it for epitope prediction. A specific anchor position is assigned to the sequences which are included in the training set. The model prediction for the MHC I model is less challenging than those for MHC class II; the reason is that for MHC I, the difference in the peptide length is negligible, while for MHC class II the length variability is larger [5]. In order to have reliable models based on ANNs, huge amounts of data are required for proper training of the neural network, this is one major limitation in the application of ANNs.

Support Vector Machines (SVMs)

SVMs, or support vector machines, handle non-linearity by using a kernel function to map the input variables into a very high-dimensional (possibly infinite-dimensional) space in which a hyperplane can be used to perform separation. The goal in SVM modeling is an optimal hyperplane, which ensures that the clusters are separated in such a way that the observations from two different classes are correctly separated between the two sides of the plane [71].

Structure-Based Methods

Structure-based drug design relies on the 3D structure of the biological target, which can be obtained with X-ray crystallography or nuclear magnetic resonance (NMR) spectroscopy [72].

Molecular Docking

The most frequently used method among the structure-based methods is molecular docking. Molecular docking allows the interaction between a protein and a small molecule (ligand) to be modeled at atomic level. The molecular docking process includes two main steps: the ligand conformation is sampled in the active site of the protein, and then the obtained conformations are ranked using a scoring function. Ideally, the experimental binding mode would be reproduced by the sampling algorithms, and among all generated conformations the scoring function would rank the highest [73].

A central challenge in molecular docking is the ligand flexibility—more flexible bonds mean more conformations that can be adopted. Consequently, a good docking method must consider not only the position and orientation of the ligand in the receptor's binding cleft but also all possible conformations [74]. The drug-like ligands usually have fewer than 10 flexible bonds; however, the MHC-binding peptides in most of the cases have more than 30 flexible bonds (reaching to more than 50 for MHC class II). The increased number of flexible bonds requires the usage of more computational resources. These limitations of the molecular docking have triggered the development of protein-peptide docking techniques.

Protein–peptide docking techniques include template-based docking; local and global docking [75].

In template-based docking methods, known structures (templates) are used as scaffolds in order to build a model of the complex. GalaxyPepDock (<http://galaxy.seoklab.org/cgi-bin/submit.cgi?type=PEPDOCK> (accessed on 27 February 2021)) searches for templates based on structure and interaction similarity [76].

Local docking methods apply a search for a peptide binding pose in the proximity of a user-defined binding site. Usually, the criteria for improvement of the initial model are the backbone—root-mean-square deviation (RMSD) from the experimental model. Rosetta FlexPepDock (<http://flexpepdock.furmanlab.cs.huji.ac.il/cite.php> (accessed on 27 February 2021)) [77], and PepCrawler (<http://bioinfo3d.cs.tau.ac.il/PepCrawler/php.php> (accessed on 27 February 2021)) [78] requires the user to prepare an initial model of the complex. HADDOCK (<http://milou.science.uu.nl/services/HADDOCK2.2/haddock.php> (accessed on 27 February 2021)) [79] automatically places the peptide in the proximity of the binding site defined by a user-provided list of interface residues. Some tools such as AutoDock Vina [80], Gold [81], Glide [82] or Surflex-Dock [83], used for the docking of small molecules can be used for the docking of peptides (up to a few amino acids). For those tools, there is also a requirement for manual placement of the peptide conformation within the binding site. DINC (<http://dinc.kavrakilab.org> (accessed on 27 February 2021)) [84] divides the peptide into segments of increasing length, and in this way overcomes the short peptide limitation. DINC is based on AutoDock protocol. The receptor structure remains rigid during the docking procedure.

Global docking methods perform a coupled search for peptide binding sites and the respective pose. The protein and the ligand conformations are represented as rigid and usually an exhaustive rigid-body docking is performed [75]. However, some docking tools allows flexibility in ligands or proteins by using different approaches: PeppAttract (<http://www.attract.ph.tum.de/services/ATTRACT/peptide.html> (accessed on 27 February 2021)) [85] performs global rigid-body docking of peptide structures within binding pockets with subsequent scoring and flexible refinement of models; CABS-dock (<http://biocomp.chem.uw.edu.pl/CABSdock> (accessed on 27 February 2021)) [86] generates peptide conformations combined with global docking in one explicit simulation. ClusPro PeptiDock (<https://cluspro.org/peptide/index.php> (accessed on 27 February 2021)) [87] predicts the peptide conformation based on motifs, followed by rigid-body docking, structural clustering used for scoring; and subsequent final structure minimization.

Although the great variety of methods and tools for protein–ligand docking, there are only a few MHC protein–ligand docking tools. Hlaffy (<http://proline.biochem.iisc.ernet.in/HLaffy> (accessed on 27 February 2021)) [88] estimates the binding strengths of peptide–MHCs and predicts the epitopes for any MHC class I allele. The assessment of the binding strength of peptide–MHCs is achieved through learning pair-potentials, which are important for the peptide binding and provide an estimation of the ligand space (total) for each allele. DockTope (<http://tools.iedb.org/docktope> (accessed on 27 February 2021)) [89] is a web-based tool for pMHC class I modeling. It is based on crystal structures of peptide complexes with two human and three mice MHC class I alleles DB. EpiDOCK (<http://epidock.ddg-pharmfac.net> (accessed on 27 February 2021)) [90] predicts the binding affinity to the most frequent human MHC class II alleles. Docking score-based quantitative matrices represent the prediction models.

Molecular Dynamics Simulations

Molecular dynamic (MD) simulation is another structure-based method. MD analyzes the physical movements of atoms and molecules. The evolution of a molecular system is tracked over time by using a potential energy function. MD simulations require huge computational resources. The MD simulations are more appropriate to study the binding between TCR, epitope and MHC protein in detail, rather than for pMHC interaction

predictions. The T cell epitope engineering and the rational design to fit to the peptide binding cleft of new MHC inhibitors may be facilitated by using MD simulation [91].

3.3.2. B Cell Epitope Prediction

The B cell receptors and/or antibodies recognize the B cell epitopes in their native structure. B cell epitopes are classified into linear (continuous) and conformational (discontinuous). Linear epitopes are short peptides, composed of sequential amino acids in the antigen. The greatest part of B cell epitopes are discontinuous (90% of B-cell epitopes). The amino acids of an epitope may not be contiguous in the primary sequence and the protein folding can bring them into spatial proximity. The continuous B cell epitopes are predicted using amino acid properties such as secondary protein structure, amino acid charge, hydrophilicity, and exposed surface area. The discontinuous B cell epitopes are predicted using the 3D structure of the antigen [92,93].

The main methodologies of B cell epitope prediction involve sequence-based methods, machine-learning methods, and structure-based methods (Table 3). The epitope surface accessibility to antibody binding is generally used by the sequence-based methods [94]. These B cell prediction methods can be widely used because the antigen sequence can be obtained easily; however, the predicted epitopic residues are not grouped into the corresponding epitopes [93]. The machine learning-based linear B cell epitope prediction method principle includes several steps: collection of a dataset with comprehensive and clean data; extraction of antigen features of the sequences (e.g., amino acid composition, physicochemical properties, evolutionary information); and training the model using machine learning algorithms [92]. Structure-based methods for conformational B cell epitope prediction identify epitopes by antigen structure and epitope-related propensity scales, including specific physicochemical properties and geometric attributes.

Table 3. Web tools for the prediction of B cells epitopes.

Tool	Description	Type of B Cell Epitope	Reference
Bcepred http://crdd.osdd.net/raghava/bcepred/ (accessed on 27 February 2021)	Prediction is based on different physicochemical properties: flexibility, hydrophilicity, exposed surface, and polarity.	Linear	[95]
BepiPred http://www.cbs.dtu.dk/services/BepiPred/index.php (accessed on 27 February 2021)	Based on a random forest algorithm, trained on epitopes annotated from antibody-antigen protein structures.	Linear	[96]
ABCPred http://crdd.osdd.net/raghava/abcpred/ (accessed on 27 February 2021)	Based on machine learning algorithm using ANN.	Linear	[95]
COBEpro http://scratch.proteomics.ics.uci.edu/ (accessed on 27 February 2021)	Use a two-step system for predicting: short peptide fragments are predicted using SVM. Then, epitope propensity scores for standalone peptide fragments and residues within an antigen sequence are calculated	Linear	[97]
DiscoTope http://www.cbs.dtu.dk/services/DiscoTope/ (accessed on 27 February 2021)	The surface accessibility and novel epitope propensity amino acid score are determined using the 3D structure of proteins. The combination of propensity residues scores in spatial proximity and the contact numbers is used to calculate the final score.	Discontinuous	[98]
BEPro (former PEPITO) http://pepito.proteomics.ics.uci.edu/ (accessed on 27 February 2021)	Requires the tertiary structure of the antigen and uses a combination of amino acid propensity scores and half sphere exposure values at multiple distances.	Discontinuous	[99]
SEPPA https://www.biosino.org/seppa3/ (accessed on 27 February 2021)	Based on logistic regression algorithm, it allows assessment of the glycoprotein antigens.	Discontinuous	[100]
EpiSearch http://curie.utmb.edu/episearch.html (accessed on 27 February 2021)	Maps conformational epitopes on the antigen protein surface. Then performs patch analysis to identify the spatial contiguous clusters of residues on the antigen surface with similar physico-chemical properties, as found in the phage display sequences.	Discontinuous	[101]

Table 3. Cont.

Tool	Description	Type of B Cell Epitope	Reference
Epitopia http://epitopia.tau.ac.il/ (accessed on 27 February 2021)	Based on machine-learning algorithm, which is trained to distinguish antigenic features within a given protein.	Both linear and discontinuous	[102]
PepSurf http://pepitope.tau.ac.il/ (accessed on 27 February 2021)	Uses a mimotope based methods. Maps each of the affinity selected peptides onto the surface of antigen and aims to find interacting interfaces by only considering solvent-exposed residues.	Both linear and discontinuous	[103]
ElliPro http://tools.iedb.org/elliPro/ (accessed on 27 February 2021)	Implements three algorithms: approximation of the protein shape as an ellipsoid, calculation of the residue protrusion index; and clustering of neighboring residues based on their protrusion index values.	Both linear and discontinuous	[104]
EPSVR http://sysbio.unl.edu/EPSVR/ (accessed on 27 February 2021)	Based on support vector regression method in order to integrate six scoring terms: residue epitope propensity, conservation, side chain energy score, contact number, secondary structure protein composition and surface planarity score.	Both linear and discontinuous	[105]

Mimotope-based methods use peptides called mimotopes which are usually obtained by random peptide library screening using a monoclonal antibody [106]. The genuine epitopes can be mimicked by the mimotopes on the corresponding antigen. The high sequential similarity of the selected mimotope implies existence of key binding motifs and physicochemical preferences [107].

3.4. Antigenicity Prediction

Antigenicity of immunogens is among the most important criteria for efficient protein design. In order to have better immunization results in the empirical phase of immunological studies, we need a protein with a high antigenicity score. In silico antigenicity prediction can be the first step in the search of the potential vaccine candidates. The employment of In Silico methods and bioinformatics in the vaccinology led to the definition of the term reverse vaccinology, invented by Rino Rappuoli [108] and applied for the first time in the development of a vaccine against serogroup B meningococcus [109].

A few computational tools for antigenicity prediction are available online:

- AntigenPro (<http://scratch.proteomics.ics.uci.edu/> (accessed on 27 February 2021)) uses machine learning methods to predicts the protein likelihood to be protective antigen. The predictive models are built using feature sets from six microbial species and their known antigens and non-antigens [110];
- VaxiJen (<http://www.ddg-pharmfac.net/vaxijen/VaxiJen/VaxiJen.html> (accessed on 27 February 2021)) is an alignment-independent method based on the auto cross covariance (ACC) transformation of protein sequences into uniform equal-length vectors and a partial least squares model for antigen prediction [111];
- Vaxign (<http://www.violinet.org/vaxign/> (accessed on 27 February 2021)) integrates open-source tools and internally developed programs to predict different biological properties: subcellular localization, number of transmembrane helices, adhesin probability, similarity to host (human, mouse, pig) proteins and MHC I and MHC II epitope binding [112];
- NERVE (www.bio.unipd.it/molbinform/ (accessed on 27 February 2021)) is used for analysis and comparison of proteins, using multiple robust algorithms. It is built on data from five microbial species which are potential vaccine candidates [113].

3.5. Allergenicity Prediction

Allergen is a protein or a glycoprotein that produces an abnormally vigorous immune response and cause an allergic reaction. The potential immunogenicity of partly foreign biopharmaceuticals can develop allergic reactions. The immune response reaction and the

severity to a biological drug may range from life-threatening (e.g., anaphylactic) to severe or effects with no clinical significance [114].

Several online tools for allergenicity prediction exist (Table 4). Allergenicity prediction tools can be divided into two groups according to the used approach: assessment according to the rules of the guidance of *Codex Alimentarius*, created by the Food and Agriculture Organization of the United Nations [115], and assessment based on machine learning methods. According to the *Codex Alimentarius* guidance, a novel protein can be considered as a putative allergen if it shares greater than 35% sequence identity with a known allergen over a sliding window of at least 80 amino acids [115].

Table 4. Computational tools for allergenicity prediction.

Tool	Description	Reference
Allermatch http://www.allermatch.org/ (accessed on 27 February 2021)	Assess allergenicity of proteins according to the rules of the guidance of <i>Codex Alimentarius</i> .	[116]
AlgPred http://crdd.osdd.net/raghava/algpred/ (accessed on 27 February 2021)	Provides six different methods for allergenicity assessment: IgE epitopes mapping, motif search, SVM, and sequence alignment.	[117]
AllerTop http://www.ddg-pharmfac.net/AllerTOP/ (accessed on 27 February 2021)	A machine learning-based tool. The amino acids of the query proteins are represented by descriptors—size, residue, abundance, hydrophobicity, helix- and β -strand forming propensities.	[118]
AllergenFP http://ddg-pharmfac.net/AllergenFP (accessed on 27 February 2021)	A four-step algorithm: a protein sequence is transformed into numerical strings, the strings are converted by auto- and cross-covariance (ACC) transformation into vectors with equal length, the vectors are transformed into binary fingerprints, and the binary fingerprints are compared using Tanimoto similarity coefficient.	[119]

4. Conclusions

In silico methods for predicting antigenicity, allergenicity and MHC binding have become an imminent part of the process of discovery and development of new biopharmaceuticals. The In Silico methods play a crucial role in drug design and development, being safe and resource- and time-efficient. They significantly reduce the subsequent experimental work but cannot yet fully replace and be an alternative of the In Vivo animal testing. The robustness of the In Silico models depends on the quality of the biological data used in the models development; therefore, the compilation and assembling of consistent biological data is of utmost importance. It is expected that the In Silico techniques will influence the search for new biopharmaceuticals as much as they have influenced the design and discovery of new drugs and small molecules.

Author Contributions: Conceptualization, I.D. (Ivan Dimitrov) and N.D.; methodology, N.D.; resources, N.D.; data curation, N.D.; writing—original draft preparation, N.D.; writing—review and editing, I.D. (Irina Doytchinova) and I.D. (Ivan Dimitrov); supervision, I.D. (Ivan Dimitrov); project administration, I.D. (Irina Doytchinova); funding acquisition, I.D. (Irina Doytchinova). All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Grant No BG05M2OP001-1.001-0003, financed by the Science and Education for Smart Growth Operational Program (2014–2020) and co-financed by the European Union through the European Structural and Investment funds. The APC was covered by the Bulgarian National Roadmap for Research Infrastructure (2017–2023), Grant No. D01-271/2019.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Flower, D.R. Immunoinformatics and the In Silico Prediction of Immunogenicity: An Introduction. In *Immunoinformatics*; Humana Press: Totowa, NJ, USA, 2007.
2. Stevanović, S. Structural basis of immunogenicity. *Transpl. Immunol.* **2002**, *10*, 133–136. [[CrossRef](#)]
3. Abbas, A.K.; Lichtman, A.H.; Pillai, S. *Cellular and Molecular Immunology*, 8th ed.; Saunders: Philadelphia, PA, USA, 2014.
4. Lautscham, G.; Rickinson, A.; Blake, N. TAP-independent antigen presentation on MHC class I molecules: Lessons from Epstein-Barr virus. *Microbes Infect.* **2003**, *5*, 291–299. [[CrossRef](#)]
5. Patronov, A.; Doytchinova, I. T-cell epitope vaccine design by immunoinformatics. *Open Biol.* **2013**, *3*, 120139. [[CrossRef](#)]
6. Kovaltsuk, A.; Krawczyk, K.; Galson, J.D.; Kelly, D.F.; Deane, C.M.; Trück, J. How B-Cell Receptor Repertoire Sequencing Can Be Enriched with Structural Antibody Data. *Front. Immunol.* **2017**, *8*, 1753. [[CrossRef](#)] [[PubMed](#)]
7. Tonegawa, S. Somatic generation of antibody diversity. *Nature* **1983**, *302*, 575–581. [[CrossRef](#)]
8. D'Angelo, S.; Ferrara, F.; Naranjo, L.; Erasmus, M.F.; Hrabec, P.; Bradbury, A.R. Many routes to an antibody heavy-chain cdr3: Necessary, yet insufficient, for specific binding. *Front. Immunol.* **2018**, *9*, 1–13. [[CrossRef](#)] [[PubMed](#)]
9. Bansal, A.K. Bioinformatics in microbial biotechnology—A mini review. *Microb. Cell Fact* **2005**, *4*, 19. [[CrossRef](#)]
10. Korber, B.; LaBute, M.; Yusim, K. Immunoinformatics comes of age. *PLoS Comput. Biol.* **2006**, *2*, e71. [[CrossRef](#)]
11. De Groot, A.S. Immunomics: Discovering new targets for vaccines and therapeutics. *Drug Discov. Today* **2006**, *11*, 203–209. [[CrossRef](#)]
12. Gokemeijer, J.; Jawa, V.; Mitra-Kaushik, S. How Close Are We to Profiling Immunogenicity Risk Using In Silico Algorithms and In Vitro Methods?: An Industry Perspective. *Aaps. J.* **2017**, *19*, 1587–1592. [[CrossRef](#)] [[PubMed](#)]
13. Holgate, R.G.; Baker, M.P. Circumventing immunogenicity in the development of therapeutic antibodies. *IDrugs* **2009**, *12*, 233–237.
14. Jones, T.D.; Crompton, L.J.; Carr, F.J.; Baker, M.P. Deimmunization of monoclonal antibodies. *Methods Mol. Biol.* **2009**, *525*, 405–423. [[PubMed](#)]
15. Rammensee, H.; Bachmann, J.; Emmerich, N.P.; Bachor, O.A.; Stevanović, S. SYFPEITHI: Database for MHC ligands and peptide motifs. *Immunogenet* **1999**, *50*, 213–219. [[CrossRef](#)]
16. Parker, K.C.; Bednarek, M.A.; Coligan, J.E. Scheme for ranking potential HLA-A2 binding peptides based on independent binding of individual peptide side-chains. *J. Immunol.* **1994**, *152*, 163–175.
17. Lundegaard, C.; Lamberth, K.; Harndahl, M.; Buus, S.; Lund, O.; Nielsen, M. NetMHC-3.0: Accurate web accessible predictions of human, mouse and monkey MHC class I affinities for peptides of length 8–11. *Nucleic Acids Res.* **2008**, *36*, W509–W512. [[CrossRef](#)] [[PubMed](#)]
18. Nielsen, M.; Lund, O. NN-align. An artificial neural network-based alignment algorithm for MHC class II peptide binding prediction. *BMC Bioinform.* **2009**, *10*, 296. [[CrossRef](#)] [[PubMed](#)]
19. Lundegaard, C.; Lund, O.; Nielsen, M. Prediction of epitopes using neural network based methods. *J. Immunol. Methods* **2011**, *374*, 26–34. [[CrossRef](#)] [[PubMed](#)]
20. Koch, C.P.; Pillong, M.; Hiss, J.A.; Schneider, G. Computational resources for MHC ligand identification. *Mol. Inf.* **2013**, *32*, 326–336. [[CrossRef](#)]
21. Wang, S.; Bai, Z.; Han, J.; Tian, Y.; Shang, X.; Wang, L. Improving the prediction of HLA class I-binding peptides using a supertype-based method. *J. Immunol. Methods* **2014**, *405*, 109–120. [[CrossRef](#)]
22. Mohammed, F.; Stones, D.H.; Zarling, A.L.; Willcox, C.R.; Shabanowitz, J.; Cummings, K.L. The antigenic identity of human class I MHC phosphopeptides is critically dependent upon phosphorylation status. *Oncotarget* **2017**, *8*, 54160–54172. [[CrossRef](#)]
23. Durrant, L.G.; Metheringham, R.L.; Brentville, V.A. Autophagy, citrullination and cancer. *Autophagy* **2016**, *12*, 1055–1056. [[CrossRef](#)]
24. Galli-Stampino, L.; Meinjohanns, E.; Frische, K.; Meldal, M.; Jensen, T.; Werdelin, O. T-cell recognition of tumor-associated carbohydrates: The nature of the glycan moiety plays a decisive role in determining glycopeptide immunogenicity. *Cancer Res.* **1997**, *57*, 3214–3222. [[PubMed](#)]
25. Antunes, D.A.; Devaurs, D.; Moll, M.; Lizée, G.; Kaviraki, L.E. General prediction of peptide-MHC binding modes using incremental docking: A proof of concept. *Sci. Rep.* **2018**, *8*, 4327. [[CrossRef](#)]
26. Petsko, G.A.; Ringe, D. *Protein Structure and Function*; Blackwell Publishing: London, UK, 2004.
27. Curtiss, R. *Vaccine Design. Innovative Approaches and Novel Strategies*; Caister Academic Press: Norfolk, UK, 2011; ISBN 978-1-904455-74-5.
28. Yu, N.Y.; Wagner, J.R.; Laird, M.R.; Melli, G.; Rey, S.; Lo, R.; Dao, P.; Sahinalp, S.C.; Ester, M.; Foster, L.J.; et al. PSORTb 3.0: Improved protein subcellular localization prediction with refined localization subcategories and predictive capabilities for all prokaryotes. *Bioinformatics* **2010**, *26*, 1608–1615. [[CrossRef](#)]
29. Petersen, T.N.; Brunak, S.; von Heijne, G.; Nielsen, H. SignalP 4.0: Discriminating signal peptides from transmembrane regions. *Nat. Methods* **2011**, *8*, 785–786. [[CrossRef](#)]
30. Bhasin, M.; Garg, A.; Raghava, G.P.S. PSLpred: Prediction of subcellular localization of bacterial proteins. *Bioinformatics* **2005**, *21*, 2522–2524. [[CrossRef](#)]
31. Desai, D.V.; Kulkarni-Kale, U. T-cell epitope prediction methods: An overview. *Methods Mol. Biol.* **2014**, *1184*, 333–364. [[CrossRef](#)] [[PubMed](#)]

32. Lata, S.; Bhasin, M.; Raghava, G.P. MHCBN 4.0: A database of MHC/TAP binding peptides and T-cell epitopes. *BMC Res. Notes* **2009**, *2*, 1–6. [[CrossRef](#)]
33. Toseland, C.P.; Clayton, D.J.; McSparron, H.; Hemsley, S.L.; Blythe, M.J.; Paine, K.; Doytchinova, I.A.; Guan, P.; Hattotuwigama, C.K.; Flower, D.R. AntiJen: A quantitative immunology database integrating functional, thermodynamic, kinetic, biophysical, and cellular data. *Immunome Res.* **2005**, *1*, 4. [[CrossRef](#)] [[PubMed](#)]
34. Vita, R.; Zarebski, L.; Greenbaum, J.A. The immune epitope database 2.0. *Nucleic Acids Res.* **2010**, *38*, D854–D862. [[CrossRef](#)]
35. Gorenshteyn, D.; Zaslavsky, E.; Fribourg, M. Interactive Big Data Resource to Elucidate Human Immune Pathways and Diseases. *Immunity* **2015**, *43*, 605–614. [[CrossRef](#)]
36. Bhattacharya, S.; Andorf, S.; Gomes, L. ImmPort: Disseminating data to the public for the future of immunology. *Immunol. Res.* **2014**, *58*, 234–239. [[CrossRef](#)] [[PubMed](#)]
37. Saha, S.; Raghava, G.P. Searching and mapping of B-cell epitopes in Bcipep database. *Methods Mol. Biol.* **2007**, *409*, 113–124. [[CrossRef](#)]
38. Conant, S.B.; Swanborg, R.H. MHC class II peptide flanking residues of exogenous antigens influence recognition by autoreactive T cells. *Autoimmun Rev.* **2003**, *2*, 8–12. [[CrossRef](#)]
39. O'Brien, C.; Flower, D.R.; Feighery, C. Peptide length significantly influences in vitro affinity for MHC class II molecules. *Immunome Res.* **2008**, *4*, 6. [[CrossRef](#)] [[PubMed](#)]
40. Hill, J.A.; Wang, D.; Jevnikar, A.M. The relationship between predicted peptide-MHC class II affinity and T-cell activation in a HLA-DRbeta1*0401 transgenic mouse model. *Arthritis. Res. Ther.* **2003**, *5*, R40–R48. [[CrossRef](#)] [[PubMed](#)]
41. Bryson, C.J.; Jones, T.D.; Baker, M.P. Prediction of immunogenicity of therapeutic proteins: Validity of computational tools. *BioDrugs* **2010**, *24*, 1–8. [[CrossRef](#)]
42. He, Y.; Rappuoli, R.; De Groot, A.S.; Chen, R. Emerging vaccine informatics. *J. Biomed. Biotechnol.* **2010**, *2010*, 1–26. [[CrossRef](#)]
43. Lin, H.H.; Zhang, G.L.; Tongchusak, S.; Reinherz, E.L.; Brusic, V. Evaluation of MHC-II peptide binding prediction servers: Applications for vaccine research. *BMC Bioinform.* **2008**, *9*, S22. [[CrossRef](#)]
44. Reche, P.A.; Glutting, J.P.; Reinherz, E.L. Prediction of MHC class I binding peptides using profile motifs. *Hum. Immunol.* **2002**, *63*, 701–709. [[CrossRef](#)]
45. Oyarzún, P.; Ellis, J.J.; Bodén, M.; Kobe, B. PREDIVAC: CD4+ T-cell epitope prediction for vaccine design that covers 95% of HLA class II DR protein diversity. *BMC Bioinform.* **2013**, *14*, 52. [[CrossRef](#)]
46. Reche, P.A.; Reinherz, E.L. PEPVAC: A web server for multi-epitope vaccine development based on the prediction of supertypic MHC ligands. *Nucleic Acids Res.* **2005**, *33*, W138–W142. [[CrossRef](#)] [[PubMed](#)]
47. Singh, H.; Raghava, G.P. ProPred: Prediction of HLA-DR binding sites. *Bioinformatics* **2001**, *17*, 1236–1237. [[CrossRef](#)]
48. Zhang, L.; Chen, Y.; Wong, H.S.; Zhou, S.; Mamitsuka, H.; Zhu, S. TEPITOPEpan: Extending TEPITOPE for peptide binding prediction covering over 700 HLA-DR molecules. *PLoS ONE* **2012**, *7*, e30483. [[CrossRef](#)] [[PubMed](#)]
49. Doytchinova, I.A.; Guan, P.; Flower, D.R. EpiJen: A server for multistep T cell epitope prediction. *BMC Bioinform.* **2006**, *7*, 131. [[CrossRef](#)]
50. Bhasin, M.; Raghava, G.P. Prediction of promiscuous and high-affinity mutated MHC binders. *Hybrid. Hybridomics* **2003**, *22*, 229–234. [[CrossRef](#)]
51. Nielsen, M.; Lundegaard, C.; Lund, O. Prediction of MHC class II binding affinity using SMM-align, a novel stabilization matrix alignment method. *BMC Bioinform.* **2007**, *8*, 238. [[CrossRef](#)]
52. Larsen, M.V.; Lundegaard, C.; Lamberth, K. An integrative approach to CTL epitope prediction: A combined algorithm integrating MHC class I binding, TAP transport efficiency, and proteasomal cleavage predictions. *Eur. J. Immunol.* **2005**, *35*, 2295–2303. [[CrossRef](#)]
53. Andreatta, M.; Nielsen, M. Gapped sequence alignment using artificial neural networks: Application to the MHC class I system. *Bioinformatics* **2016**, *32*, 511–517. [[CrossRef](#)] [[PubMed](#)]
54. Nielsen, M.; Lundegaard, C.; Lund, O.; Keşmir, C. The role of the proteasome in generating cytotoxic T-cell epitopes: Insights obtained from improved predictions of proteasomal cleavage. *Immunogenetics* **2005**, *57*, 33–41. [[CrossRef](#)]
55. Jensen, K.K.; Andreatta, M.; Marcatili, P. Improved methods for predicting peptide binding affinity to MHC class II molecules. *Immunology* **2018**, *154*, 394–406. [[CrossRef](#)]
56. Reynisson, B.; Barra, C.; Kaabinejadian, S.; Hildebrand, W.H.; Peters, B.; Nielsen, M. Improved Prediction of MHC II Antigen Presentation through Integration and Motif Deconvolution of Mass Spectrometry MHC Eluted Ligand Data. *J. Proteome Res.* **2020**, *19*, 2304–2315. [[CrossRef](#)]
57. Bhasin, M.; Raghava, G.P. Analysis and prediction of affinity of TAP binding peptides using cascade SVM. *Protein Sci.* **2004**, *13*, 596–607. [[CrossRef](#)] [[PubMed](#)]
58. Bhasin, M.; Raghava, G.P. Pcleavage: An SVM based method for prediction of constitutive proteasome and immunoproteasome cleavage sites in antigenic sequences. *Nucleic Acids Res.* **2005**, *33*, W202–W207. [[CrossRef](#)]
59. Liu, W.; Meng, X.; Xu, Q.; Flower, D.R.; Li, T. Quantitative prediction of mouse class I MHC peptide binding affinity using support vector machine regression (SVR) models. *BMC Bioinform.* **2006**, *7*, 182.
60. Dönnes, P.; Kohlbacher, O. Integrated modeling of the major events in the MHC class I antigen processing pathway. *Protein Sci.* **2005**, *14*, 2132–2140. [[CrossRef](#)] [[PubMed](#)]

61. Guan, P.; Hattotuwigama, C.K.; Doytchinova, I.A.; Flower, D.R. MHCpred 2.0: An updated quantitative T-cell epitope prediction server. *Appl. Bioinform.* **2006**, *5*, 55–61. [[CrossRef](#)] [[PubMed](#)]
62. Dimitrov, I.; Garnev, P.; Flower, D.R.; Doytchinova, I. EpiTOP—A proteochemometric tool for MHC class II binding prediction. *Bioinformatics* **2010**, *26*, 2066–2068. [[CrossRef](#)]
63. Bhasin, M.; Raghava, G.P.S. Prediction of CTL epitopes using QM, SVM and ANN techniques. *Vaccine* **2004**, *22*, 3195–3201. [[CrossRef](#)]
64. Dhanda, S.K.; Grifoni, A.; Pham, J. Development of a strategy and computational application to select candidate protein analogues with reduced HLA binding and immunogenicity. *Immunol* **2018**, *153*, 118–132. [[CrossRef](#)]
65. Calis, J.J.; Maybeno, M.; Greenbaum, J.A. Properties of MHC class I presented peptides that enhance immunogenicity. *PLoS Comput. Biol.* **2013**, *9*, e1003266. [[CrossRef](#)] [[PubMed](#)]
66. Rapin, N.; Lund, O.; Castiglione, F. Immune system simulation online. *Bioinformatics* **2011**, *27*, 2013–2014. [[CrossRef](#)]
67. Altuvia, Y.; Berzofsky, J.A.; Rosenfeld, R.; Margalit, H. Sequence features that correlate with MHC restriction. *Mol. Immunol.* **1994**, *31*, 1–19. [[CrossRef](#)]
68. Nussbaum, A.K.; Kuttler, C.; Tenzer, S.; Schild, H. Using the World Wide Web for predicting CTL epitopes. *Curr. Opin. Immunol.* **2003**, *15*, 69–74. [[CrossRef](#)] [[PubMed](#)]
69. Gulukota, K.; Sidney, J.; Sette, A.; DeLisi, C. Two complementary methods for predicting peptides binding major histocompatibility complex molecules. *J. Mol. Biol.* **1997**, *267*, 1258–1267. [[CrossRef](#)]
70. Beale, R.; Jackson, T. *Neural Computing: An introduction*; Adam, H., Ed.; Taylor & Francis Group: New York, NY, USA, 1990.
71. Wikberg, J.; Eklund, M.; Willighagen, E.L.; Spjuth, O.; Lapins, M.; Engkvist, O.; Alvarsson, J. *Introduction to Pharmaceutical Bioinformatics*; Oakleaf Academic: Stockholm, Sweden, 2011; ISBN 9789197940306.
72. Leach, A.R.; Harren, J. *Structure-Based Drug Discovery*; Springer: Berlin, Germany, 2007; ISBN 978-1-4020-4406-9.
73. Meng, X.Y.; Zhang, H.X.; Mezei, M.; Cui, M. Molecular docking: A powerful approach for structure-based drug discovery. *Curr. Comput. Aided Drug Des.* **2011**, *7*, 146–157. [[CrossRef](#)] [[PubMed](#)]
74. Guedes, I.A.; de Magalhães, C.S.; Dardenne, L.E. Receptor-ligand molecular docking. *Biophys. Rev.* **2014**, *6*, 75–87. [[CrossRef](#)]
75. Ciemny, M.; Kurcinski, M.; Kamel, K. Protein-peptide docking: Opportunities and challenges. *Drug Discov. Today* **2018**, *23*, 1530–1537. [[CrossRef](#)]
76. Hasup, L.; Lim, H.; Myeong, S.L.; Chaok, S. GalaxyPepDock: A protein-peptide docking tool based on interaction similarity and energy optimization. *Nucleic Acids Res.* **2015**, *43*, W431–W435. [[CrossRef](#)]
77. Nir, L.; Barak, R.; Eyal, C.; Guy, F.; Ora, S.-F. Rosetta FlexPepDock web server—high resolution modeling of peptide-protein interactions. *Nucleic Acids Res.* **2011**, *39*, W249–W253. [[CrossRef](#)]
78. Elad, D.; Wolfson, H.J. PepCrawler: A fast RRT-based algorithm for high-resolution refinement and binding affinity estimation of peptide inhibitors. *Bioinformatics* **2011**, *27*, 2836–2842. [[CrossRef](#)]
79. Van Zundert, G.C.P.; Rodrigues, J.P.G.L.M.; Trellet, M.; Schmitz, C.; Kastiris, P.L.; Karaca, E.; Melquiond, A.S.J.; van Dijk, M.; De Vries, S.J.; Bonvin, A.M.J.J. The HADDOCK2.2 webserver: User-friendly integrative modeling of biomolecular complexes. *J. Mol. Biol.* **2016**, *428*, 720–725. [[CrossRef](#)]
80. Trott, O.; Olson, A.J. AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization and multithreading. *J. Comput. Chem.* **2010**, *31*, 455–461. [[CrossRef](#)]
81. Jones, G.; Willett, P.; Glen, R.C.; Leach, A.R.; Taylor, R. Development and Validation of a Genetic Algorithm for Flexible Docking. *J. Mol. Biol.* **1997**, *267*, 727–748. [[CrossRef](#)]
82. Friesner, R.A.; Banks, J.L.; Murphy, R.B. Glide: A new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. *J. Med. Chem.* **2004**, *47*, 1739–1749. [[CrossRef](#)] [[PubMed](#)]
83. Spitzer, R.; Jain, A.N. Surflex-Dock: Docking benchmarks and real-world application. *J. Comput. Aided Mol. Des.* **2012**, *26*, 687–699. [[CrossRef](#)] [[PubMed](#)]
84. Antunes, D.A.; Moll, M.; Devaurs, D.; Jackson, K.R.; Lizée, G.; Kaviraki, L.E. DINC 2.0: A new protein-peptide docking webserver using an incremental approach. *Cancer Res.* **2017**, *77*, E55–E57. [[CrossRef](#)] [[PubMed](#)]
85. Schindler, C.E.M.; de Vries, S.J.; Zacharias, M. Fully Blind Peptide-Protein Docking with pepATTRACT. *Structure* **2015**, *23*, 1507–1515. [[CrossRef](#)]
86. Mateusz, K.; Michal, J.; Maciej, B.; Andrzej, K.; Sebastian, K. CABS-dock web server for the flexible docking of peptides to proteins without prior knowledge of the binding site. *Nucleic Acids Res.* **2015**, *43*, W419–W424. [[CrossRef](#)]
87. Porter, K.A.; Xia, B.; Beglov, D.; Bohnuud, T.; Alam, N.; Schueler-Furman, O.; Kozakov, D. ClusPro PeptiDock: Efficient global docking of peptide recognition motifs using FFT. *Bioinformatics* **2017**, *33*, 3299–3301. [[CrossRef](#)]
88. Mukherjee, S.; Bhattacharyya, C.; Chandra, N. HLaffy: Estimating peptide affinities for Class-1 HLA molecules by learning position-specific pair potentials. *Bioinformatics* **2016**, *32*, 2297–2305. [[CrossRef](#)]
89. Rigo, M.M.; Antunes, D.A.; Vaz de Freitas, M. DockTope: A Web-based tool for automated pMHC-I modelling. *Sci. Rep.* **2015**, *5*, 18413. [[CrossRef](#)]
90. Atanasova, M.; Patronov, A.; Dimitrov, I.; Flower, D.R.; Doytchinova, I. EpiDOCK: A molecular docking-based tool for MHC class II binding prediction. *Protein Eng. Des. Sel.* **2013**, *26*, 631–634. [[CrossRef](#)]
91. Rognan, D.; Scapozza, L.; Folkers, G.; Daser, A. Molecular dynamics simulation of MHC-peptide complexes as a tool for predicting potential T cell epitopes. *Biochemistry* **1994**, *33*, 11476–11485. [[CrossRef](#)]

92. Sun, P.; Ju, H.; Liu, Z.; Ning, Q.; Zhang, J.; Zhao, X. Bioinformatics resources and tools for conformational B-cell epitope prediction. *Comput. Math Methods Med.* **2013**, *2013*, 943636. [[CrossRef](#)]
93. Yao, B.; Zheng, D.; Liang, S.; Zhang, C. Conformational B-cell epitope prediction on antigen protein structures: A review of current algorithms and comparison with common binding site prediction methods. *PLoS ONE* **2013**, *19*, e62249. [[CrossRef](#)] [[PubMed](#)]
94. Saha, S.; Raghava, G. Prediction of continuous B-cell epitopes in an antigen using recurrent neural network. *Proteins Struct. Funct. Bioinform.* **2006**, *65*, 40–48. [[CrossRef](#)] [[PubMed](#)]
95. Saha, S.; Raghava, G.P.S. BcePred: Prediction of Continuous B-Cell Epitopes in Antigenic Sequences Using Physico-chemical Properties. In *Artificial Immune Systems, ICARIS, Lecture Notes in Computer Science*; Cutello, G., Bentley, V., Timmis, P.J., Eds.; Springer: Berlin/Heidelberg, Germany, 2004.
96. Jespersen, M.C.; Peters, B.; Nielsen, M.; Marcatili, P. BepiPred-2.0: Improving sequence-based B-cell epitope prediction using conformational epitopes. *Nucleic Acids Res.* **2017**, *45*, W24–W29. [[CrossRef](#)] [[PubMed](#)]
97. Sweredoski, M.J.; Baldi, P. COBEpro: A novel system for predicting continuous B-cell epitopes. *Protein Eng. Des. Sel.* **2009**, *22*, 113–120. [[CrossRef](#)]
98. Kringelum, J.V.; Lundegaard, C.; Lund, O.; Nielsen, M. Reliable B cell epitope predictions: Impacts of method development and improved benchmarking. *PLoS Comput.* **2012**, *8*, 12. [[CrossRef](#)]
99. Sweredoski, M.J.; Baldi, P. PEPITO: Improved discontinuous B-cell epitope prediction using multiple distance thresholds and half sphere exposure. *Bioinformatics.* **2008**, *24*, 1459–1460. [[CrossRef](#)] [[PubMed](#)]
100. Zhou, C.; Chen, Z.; Zhang, L. SEPPA 3.0-enhanced spatial epitope prediction enabling glycoprotein antigens. *Nucleic Acids Res.* **2019**, *47*, W388–W394. [[CrossRef](#)] [[PubMed](#)]
101. Negi, S.S.; Braun, W. Automated detection of conformational epitopes using phage display Peptide sequences. *Bioinform. Biol. Insights* **2009**, *3*, 71–81. [[CrossRef](#)] [[PubMed](#)]
102. Rubinstein, N.D.; Mayrose, I.; Martz, E.; Pupko, T. Epitopia: A web-server for predicting B-cell epitopes. *BMC Bioinform.* **2009**, *10*, 287. [[CrossRef](#)] [[PubMed](#)]
103. Mayrose, I.; Shlomi, T.; Rubinstein, N.D. Epitope mapping using combinatorial phage-display libraries: A graph-based algorithm. *Nucleic Acids Res.* **2007**, *35*, 69–78. [[CrossRef](#)]
104. Ponomarenko, J.; Bui, H.H.; Li, W. ElliPro: A new structure-based tool for the prediction of antibody epitopes. *BMC Bioinform.* **2008**, *9*, 514. [[CrossRef](#)]
105. Liang, S.; Zheng, D.; Standley, D.M.; Yao, B.; Zacharias, M.; Zhang, C. EPSVR and EPMeta: Prediction of antigenic epitopes using support vector regression and multiple server results. *BMC Bioinform.* **2010**, *11*, 381. [[CrossRef](#)]
106. Geysen, H.M.; Rodda, S.J.; Mason, T.J. A priori delineation of a peptide which mimics a discontinuous antigenic determinant. *Mol. Immunol.* **1986**, *23*, 709–715. [[CrossRef](#)]
107. Mumei, B.; Angel, N.O.T. Filtering epitope alignments to improve protein surface prediction. In *ISPA*; Springer: Berlin/Heidelberg, Germany, 2006; pp. 648–657.
108. Rappuoli, R. Reverse Vaccinology. *Curr. Opin. Microbiol.* **2000**, *3*, 445–450. [[CrossRef](#)]
109. Pizza, M.; Scarlato, V.; Masignani, V.; Giuliani, M.M.; Arico, B.; Comanducci, M.; Jennings, G.T.; Baldi, L.; Bartolini, E.; Capecchi, B.; et al. Identification of Vaccine Candidates Against Serogroup B Meningococcus by Whole-Genome Sequencing. *Science* **2000**, *287*, 1816–1820. [[CrossRef](#)]
110. Magnan, C.N.; Zeller, M.; Kayala, M.A. High-throughput prediction of protein antigenicity using protein microarray data. *Bioinformatics* **2010**, *26*, 2936–2943. [[CrossRef](#)] [[PubMed](#)]
111. Doytchinova, I.A.; Flower, D.R. VaxiJen: A server for prediction of protective antigens, tumour antigens and subunit vaccines. *BMC Bioinform.* **2007**, *8*, 4. [[CrossRef](#)]
112. He, Y.; Xiang, Z.; Mobley, H.L. Vaxign: The first web-based vaccine design program for reverse vaccinology and applications for vaccine development. *J. Biomed. Biotechnol.* **2010**, *2010*, 297505. [[CrossRef](#)] [[PubMed](#)]
113. Vivona, S.; Bernante, F.; Filippini, F. NERVE: New enhanced reverse vaccinology environment. *BMC Biotechnol.* **2006**, *6*, 35. [[CrossRef](#)]
114. Gülsen, A.; Wedi, B.; Jappe, U. Hypersensitivity reactions to biologics (part I): Allergy as an important differential diagnosis in complex immune-derived adverse events. *Allergo J. Int* **2020**, *29*, 1–29. [[CrossRef](#)]
115. Food and Agriculture Organization of the United Nations. *Foods Derived from Modern Biotechnology*, 2nd ed.; Codex Alimentarius Commission: Rome, Italy, 2009.
116. Fiers, M.W.; Kleter, G.A.; Nijland, H.; Peijnenburg, A.A.; Nap, J.P.; van Ham, R.C. Allermatch, a webtool for the prediction of potential allergenicity according to current FAO/WHO Codex alimentarius guidelines. *BMC Bioinform.* **2004**, *5*, 133. [[CrossRef](#)] [[PubMed](#)]
117. Saha, S.; Raghava, G.P. AlgPred: Prediction of allergenic proteins and mapping of IgE epitopes. *Nucleic Acids Res.* **2006**, *34*, W202–W209. [[CrossRef](#)] [[PubMed](#)]
118. Dimitrov, I.; Bangov, I.; Flower, D.R.; Doytchinova, I. AllerTOP v.2—A server for in silico prediction of allergens. *J. Mol. Model.* **2014**, *20*, 2278. [[CrossRef](#)] [[PubMed](#)]
119. Dimitrov, I.; Naneva, L.; Doytchinova, I.; Bangov, I. AllergenFP: Allergenicity prediction by descriptor fingerprints. *Bioinformatics* **2014**, *30*, 846–851. [[CrossRef](#)] [[PubMed](#)]