VOL.6 NO.2 September 2017 - ISSN 2233 – 1859

Operations Research Society in Bosnia and Herzegovina

Southeast Europe Journal of Soft Computing

Available online: http://scjournal.ius.edu.ba

IUS Soft Computing Research Group

On the Accuracies of Sequence Based Linear B Cell Epitope Predictors

A. Abidi
M. Can

Faculty of Engineering and Natural Sciences,
Department of Biological Sciences and Bioengineering
International University of Sarajevo International University of Sarajevo,
HrasnickaCesta 15, Ilidža 71210 Sarajevo,
Bosnia and Herzegovina
azraabidi34@yahoo.com
mcan@ius.edu.ba

Article Info

Article history:
Article received on 17 June 2017
Received in revised form 17 August .2017

Keywords:

Antigens; B-cell epitopeprediction;accuracy; reliability

Abstract

The accuracy of online tools employed in attempts to predict B-cell epitopes based on sequence are very poor. In order to improve the accuracy of these predictions it is essential to design algorithms to benefit from the features achieved in wet lab in vivo experimental models. To shed some light on accuracy and reliability of these online tools, we set an insilico experiment on five selected online tools using five antigens whose b-cell epitopes are known through wet lab experiments. To evaluate successes of online tools, we defined two measures, accuracy, and reliability. To the findings of this experiment, the most accurate tool is ABCpred with a score of 43.59 %. That is peptides that are predicted as b-epitopes, cover in average 43.59 % of the wet lab listed b-epitopes. The most reliable predictors are BCpred and AAPred with scores of 52.54%, and 52.60% respectively, which means that in average around half of the peptides that are predicted as a b-epitope by these predictors have a chance to be a real b-epitope. Combining several predictors to get better predictors is not an advisable technique. From this experiment it is concluded that the accuracy and reliability of online tools still are far away being satisfactory.

1. INTRODUCTION

The human body is a wonderful machine, which has the ability, to fight off attacking microorganisms and adapt to a variety of situations and to repair itself. Very often, microorganisms or foreign particles enter the body and attempt to infect and harm the body causing diseases. Disease is a condition where an organism experiences impaired function often with detrimental symptoms. According to the World Health Organization, as of 2017 there were 7,186 classified different diseases and health-related ailments.[1] Virus is one of these organisms that can cause severe health issues. They invade living, normal cells and hijacks their normal molecular balance

[1] http://www.who.int/classifications/ICD11January2017Newsletter.pdf?ua=1

and use those cells to multiply and produce other viruses like themselves. For most viral infections, treatments can only help with symptoms while you wait for your immune system to fight off the virus.

One of the sources of modern medicine is Bioinformatics, a hybrid science that joins biological data with techniques for information storage, distribution, and analysis to support multiple areas of scientific research, including biomedicine. Bioinformatics is a field with high-through-put data-generating experiments, including genomic sequence determinations and measurements of gene expression patterns. Omics-based technologies allow the analysis of a large number of molecular features collectively and in one go through. This field has rapidly evolved and is now technologically mature. our main purpose is to introduce role of different online available tools and their role in biological analysis and databases and provide guidance on how to use them effectively. Identification of epitopes which invoke strong humoral responses is an essential issue in the field of immunology. In order to facilitate the insilico design and development of vaccines, bioinformatics approaches can play a critical role on analyzing multiple genomes and to predict potential predicted epitopes, one can define epitopes or chimerical epitopes. It is conceived that with the help of these protein arrangements, including target epitopes, that may provide a rationale to design peptides that are capable to elicit convenient humoral or cellular immune responses. For this purpose, after a comprehensive compilation of the online immunological software, five advantageous ones are chosen, i.e. ABCpred, BCpred, K&T, Bepipred and AAP.

2. LINEAR B-EPITOPE PREDICTION TOOLS

At least for three decades computational methods have been developed for facilitating epitope recognition. Previously, the majority of the insilico methods were focused on continuous linear epitopes. Most of these approaches are sequence-based and use amino acid-based propensity scales, related to physico-chemical properties such as hydrophilicity, solvent accessibility, secondary structure and flexibility; a score derived from the propensity scales is assigned to each residue, and the whole sequence is examined for high-scoring window fragments, which are then predicted as epitopes [1-7]. In recent years some artificial intelligence tools also started to emerge as prediction tools [8-10].

Table 1. Online tools created to predict linear b- epitopes with their websites and references.

No	Name	Web	Ref
1	AAP	http://ailab.ist.psu.edu/bcpred/	[11]
2	ABCpred	http://osddlinux.osdd.net/raghava/abcpred/index.html	[9]
3	Bayesb	http://www.immunopred.org/bayesb/index.html. (NA)	[12]
4	BCpreds	http://ailab.ist.psu.edu/bcpred/	[13]
5	Bcepred	http://webs.iiitd.edu.in/raghava/bcepred/	[14]
6	BepiPred	http://www.cbs.dtu.dk/services/BepiPred-1.0/	[15]
7	BEST	http://biomine.ece.ualberta.ca/BEST/ (Desktop)	[16]
8	BROracle	https://sites.google.com/site/oracleclassifiers/ (Desktop)	[17]
9	COBEpro	http://scratch.proteomics.ics.uci.edu/	[18]
10	CBTOPE	http://crdd.osdd.net/raghava//cbtope/ (Conformational)	[19]
11	Discotope	http://www.cbs.dtu.dk/services/DiscoTope/	[20]
12	EPCES	http://sysbio.unl.edu/EPCES/	[21]
13	EPMLR	http://www.bioinfo.tsinghua.edu.cn/epitope/EPMLR/ (NA)	[22]
14	EPIC	http://saphire.usask.ca/saphire/epic/	[23]
15	EpiPred	http://www.stats.ox.ac.uk/research/proteins/resources (NA)	[24]
16	Epitopia	http://epitopia.tau.ac.il/	[25-26]
17	FBCPred	http://ailab.ist.psu.edu/bcpred/ (NA)	[27-29]
16	IgPred	http://crdd.osdd.net/raghava/igpred/	[30]
17	Kolaskar&T	http://tools.immuneepitope.org/bcell/	[31]
19	LBtope	http://crdd.osdd.net/raghava/lbtope/	[32]
20	LBEEP	https://github.com/brsaran/LBEEP/blob/master/LBEEP (Desktop)	[33]
21	LEP-LP	http://biotools.cs.ntou.edu.tw/lepd_antigenicity. php (NA)	[34]
22	LEPS	http://leps.cs.ntou.edu.tw/	[35]
23	PEASE	www.ofranlab.org/PEASE.	[36]
24	PEOPLE	Desktop	[37]
25	PREDITOP	Desktop	[38]
26	SEPIa	https://github.com/SEPIaTool/	[39]
27	SVMTriP	http://sysbio.unl.edu/SVMTriP	[40]

In this research, information about online tools created to predict linear b-epitopes are searched by a comprehensive literature search around thirty web sites are visited. In Table 1, some of these online tools are listed alongside with their websites and references.

2.1 The Five Linear B-Epitope Prediction Tools Investigated

To use insilico techniques to predict linear B-epitopes of antigenic proteins, first we must have an idea about their accuracy and reliability. For this reason five most prominent online tools are chosen.

2.1.1 ABCpred

ABCpred server is used to predict linear B cell epitope regions in an antigen sequence. Saha and Raghava (2006a) have made available the data sets used to train and evaluate ABCpred. We have selected 20mers data set to analyze its performance. ABCpred uses recurrent artificial neural networks for predicting linear B-cell epitopes B-cell epitopes play a vital and significant role in the development of peptide vaccines. This is the first server developed based on recurrent neural network (machine based technique). Users can select window length of 10, 12, 14, 16 and 20 as predicted epitope lengths. ABCpred generates datasets of fixed length patterns by eliminating or adding residues at the terminal ends of the peptides. It presents the results in graphical and tabular frame. In case of graphical frame, this server plots the epitopes in blue color along protein backbone (black color), which assist the users in rapid visualization of B-cell epitope on protein. The tabular output is in the form of a table, which will provide the amino acids length from N-terminal to C-terminal in a protein predicted by the server. Experimental methods used for characterizing epitopes are time consuming and demand large resources. The availability of epitope prediction method(s) can rapidly help experimenters in simplifying and modifying this problem. In ABCpred all calculations done by using artificial neural network. This server will assist in locating and finding of epitope regions that are useful in selecting synthetic vaccine candidates, disease diagnosis, prediction of potential vaccine candidates and also in allergy research. ABCpred server is available at the web site [9]:

http://osddlinux.osdd.net/raghava/abcpred/index.html

2.1.2 BCpreds

The BCpreds server allows users to choose the method for predicting B-cell epitopes among several developed prediction methods. The current implementation of BCPREDS allows the user to select among three prediction methods: (i) implementation of AAP method [Chen et al., 2007]; (ii) BCPred [EL-Manzalawy et al., 2008]; (iii) FBCPred (EL-Manzalawy et al., 2008b). The way to handle is that users provide an antigen sequence and optionally can specify and notify desired epitope length and specificity thresholds also. Results are returned in several user-friendly formats. BCPred server allows choosing prediction method among amino acids pair scaling method (AAP), BCPred, and FBCPred. AAP approach is based on the finding that particular amino acid pairs occur more frequently in epitope than nonepitope sequence. Combination of AAP propensity scale with turns, accessibility, antigenicity, hydrophilicity, and flexibility propensity scales improved the accuracy (72.5%). BCPred method employs subsequence kernel-based SVM classifier and was trained on homology-reduced dataset of linear B-cell epitopes (with <80% sequence identity) derived from dataset previously used to test ABCpred. BCpred server is available at the web site [13]:

http://ailab.ist.psu.edu/bcpred/

2.1.3 Kolaskar&Tongaonkar

The quality of Kolaskar&Tongaonkar is that it Analyses of the data from experimentally determined antigenic sites on proteins has revealed that the hydrophobic residues Cys, Leu and Val, if they occur on the surface of a protein, are more likely to be a part of antigenic sites. A semi-empirical method which makes use of physicochemical properties of amino acid residues and their frequencies of occurrence in experimentally known segmental epitopes was developed to predict antigenic determinants on proteins. Application of this method to a large number of proteins has shown that our method can predict antigenic determinants with about 75% accuracy which is better than most of the known methods. This method is based on a single parameter and thus very simple to use. The B cell epitope prediction was performed using the program Predicting Antigenic Peptides available online. The software for the detection of antigenic peptides is based on Kolaskar and Tongaonkar method previously described (Kolaskar and Tongaonkar, 1990). Kolaskar&Tongaonkar Antigenicity is a semi-empirical method for the prediction of antigenic regions including information of surface accessibility and flexibility. Antigenicity was determined on the basis of well-balanced algorithmic method developed by Kolaskar and Tongaonkar for the prediction of antigenic determinants on protein antigens. This method is based on physicochemical properties of amino acid residues and their frequencies of occurrence in experimentally known segmental epitopes. Kolaskar&Tongaonkar online prediction tool is available at the web site [31]:

http://tools.immuneepitope.org/bcell/

2.1.4 Bepipred

For each input sequence, BepiPred outputs not only a prediction score for every single residue but also it predicts the sequence of potential epitope. The positions of the linear B-cell epitopes are predicted to be located at the residues with the highest scores. Bepipred uses a hidden Markov model-based method along with amino acid propensity scales for accessibility, hydrophilicity, flexibility and polarity trained on a dataset of curated B cell epitopes. Lastly, the ABCpred prediction tool is an artificial neural network-based B cell epitope prediction server that recognizes that B cell epitopes have varying lengths (5 to 20 residues). In this paper we used IEDB Analysis Bepipred tool. The BepiPred method applies a hidden Markov model which takes two propensity scores as its inputs. A number of machine learning-based model were recently developed, from decision trees and k-nearest neighbor that utilized a combination of multiple propensities and sequence complexity as inputs, to neural network-based ABCpred that performs predictions directly from protein chain. The later method is designed to recognize epitopic peptides with 20 or fewer (i.e., 10,12,14,16 and 20) amino acids (AAs). The newest sequence-based predictors of continuous B-cell epitopes exclusively use support vector machine (SVM) models. BepiPred (http://tools.iedb.org/bcell/) is a combination method, produced by combining the predictions of a Hidden Markova model and the propensity scale by Parker et al. This method assigns a score value to each protein residue. Threshold was set at −0.2 (to obtain the sensitivity of 75% and specificity of 50%, similar to those of SVMTriP) or at 0.35 (the default). BepiPred online prediction tool is available at the web site [15]:

http://www.cbs.dtu.dk/services/BepiPred-1.0/

2.1.5 AAP

The AAP antigenicity scale can reflect some special sequence-coupled feature in the prediction of B-cell epitopes, which is the essence why the new approach is superior to the existing ones. It is anticipated that with the continuous increase of the known epitope data, the power of the AAP antigenicity scale approach will be further enhanced and improved. AAP method 20 maps each peptide sequence into a set of fixed length numeric features and therefore it can be trained using datasets of flexible length sequences. However, the performance of this method had been reported using a dataset of 20-mer peptides. Identification of antigenic sites on proteins is of vital importance for developing synthetic peptide vaccines, immunodiagnostic tests and antibody production. Currently, most of the prediction algorithms rely on amino acid propensity scales using a sliding window approach. These methods are oversimplified and sometime yield poor predicted results in practice. In this paper, a novel scale, called the amino acid pair (AAP) antigenicity scale, is proposed that is based on the finding that B-cell epitopes favor particular AAPs. It is demonstrated that, using SVM (support vector machine) classifier, the AAP antigenicity scale approach has much better performance than the existing scales based on the single amino acid propensity. The path toward a balanced portfolio of ABCpred is a capable, safe, and transparent artificial intelligence based system which draw on a broad spectrum of computing ideas and principles. ABCpred along with other online prediction tools are likely to become a driver for new advances in computing. AAP online prediction tool is available at the web site [11]:

http://ailab.ist.psu.edu/bcpred/

3. IEDB DATABASE FOR EPITOPES

The National Institute of Allergy and Infectious Diseases has funded a free resource namely IEDB. Which provides easy searching of experimental data characterizing antibody, B and T cell epitopes studied in humans, non-human primates, and other animal species. Epitopes involved in different infectious disease, allergy, autoimmunity, and transplant are included. The summary Metrics of IEDB shows that how comprehensive and effective this tool is. It contains 385,452Peptidic Epitopes, 2,551Non-Peptidic Epitopes, 324,017T Cell Assays, 403,574B Cell Assays,788,676MHC Ligand Assays,3,632Epitope Source Organisms,753Restricting MHC Alleles and18,836References.

3.1 The Five Antigens with Known Linear B-Epitopes

To have an idea about accuracies and reliabilities of the above mentioned online tools, we have chosen five antigens, whose linear b-epitopes are known and compiled through wet lab experiments. i.e. Plasmodium Falciparum, Human Polio Virus Sabin Strain, Meningitis, Plasmodium Vivax and Mycobacterium Tuberculosis.

3.1.1 Plasmodium Falciparum:

Plasmodium falciparum is a protozoan parasite that causes an infectious disease known as malaria. P. falciparum is the most severe strain of the malaria species correlated with almost every malarial death. The other 3 species that cause malaria include: P. vivax, P. ovale, and P. malariae. Humans become infected by a female Anopheles mosquito which, transfers a parasitic vector through its saliva into the blood stream. The parasite then infects the liver and undergoes asexual reproduction followed by insertion into red blood cells where an additional round of replication takes place. P. falciparum changes the surface of an infected red blood cell causing it to adhere to blood vessels, cytoadherence, as well as to other red blood cells. In severe cases this leads to obstructions of micro circulation resulting in dysfunction of many organs. Symptoms depend on severity of infection and can present a range of signs such as flulike symptoms, vomiting

diarrhea, shock, kidney failure, coma, and death. Plasmodium falciparum mostly infects children under the age of 5 as well as pregnant women. An important virulence property of P. falciparum is the expression of parasite-derived antigens on the surface of IEs, generally known as variant surface antigens, and its strong propensity to adhere in the vasculature.

Sickle cell individuals have shown to rarely contract malaria. Research has shown that this is partially due to weakened binding of parasite-infested sickle cell erythrocytes to micro vascular endothelial cells when compared to normal hemoglobin parasite erythrocytes binding. The virulence factor PfEMP1 that normally conducts cytoadherence is altered creating a weekend attachment between it and the epithelial wall. Due to the ability to attach lacking, sequestration would also not occur limiting the severe malarial response. The mechanism for how this is done is still unknown and needs further research.

The 26 wet Lab reported linear B Cell epitopes of Plasmodium Falciparum are as follows [41]:

AEENVEENVEEVEENVEENV
AEENVEH
DDEHVEEPTVA
DDEHVEEPTVADDEHVEEPTVA
EENVEENV
EENVEENVEENVEENV
EENVEENVEENVEENVEENVEENV
EENVEHDA
EENVEHDAEENVEENV
EENVEHDAEENVEHDA
EENVEHDAEENVEHDAEENVEENV
EENVEHDAEENVEHDAEENVEHDA
EENVEHDAEENVEHDAEENVEHDAEENVEHDA
EKVDNLGRSGGDIIK
EKVDNLGRSGGDIIKKMQTLWDEIMDINKRK
IVGYIMHGISTINTEMK
LFDYNEKVDNLGRSG
LGRSGGDIIKKMQTL
MQTLWDEIMDINKRK
NADMNEITERYFKLAENYY
PTVAEEH
SLRWIFKHVAKTHLK
TVAEEHV
TVAEEHVEEPTVAEE
VAEEHVE
VEENVEE

In our insilico research project we have taken wet lab results as standard parameter for Plasmodium Falciparum and used all five above mentioned online prediction tools, these tools predicted different sequences that can be the potential epitopes. All these predicted epitopes were analyzed on the basis of wet lab results. Our parameter was minimum 70% match of predicted sequence with wet lab results, also we have chosen sequence length of 20 amino acids. Different predictors suggested some similar and different results, the sequences predicted by at least 4 of the predicted tools are:

DDEHVEEPTVA
DDEHVEEPTVADDEHVEEPTVA
PTVAEEH, TVAEEHV
TVAEEHVEEPTVAEE
VAEEHVE

3.1.2 Human Polio Virus

Poliovirus, the causative agent of paralytic poliomyelitis, is an enterovirus spread by the oral route. The principal infection associated with the poliovirus is enteritis with the prodromal illness of fever, headache, arthralgia, vomiting, and diarrhea lasting 3–4 days. About half of the patients do not develop paralytic manifestations. In the remaining, a biphasic course evolves. As the initial enteritis subsides, the paralysis begins. Severe back and limb pain, headache, and meningismus develop, accompanied by severe and disabling muscle spasms. Paralysis tends to occur in a patchy, multifocal distribution. Weakness of individual muscles comes on rapidly over days and typically reaches a maximum within 1 week. The virus has a specific tropism for the motor neurons, resulting in motor neuron death. Virtually any of the skeletal muscles, including bulbar, limb, and respiratory muscles, can be affected. The time from being infected with the virus to developing symptoms of disease (incubation) ranges from 5 - 35 days (average 7 - 14 days). Most people do not develop symptoms. Outbreaks can still occur in the developing world, usually in groups of people who have not been vaccinated. Some victims develop neurological complications, including stiffness of the neck and back, weak muscles, pain in the joints, and paralysis of one or more limbs or respiratory muscles. In severe cases it may be fatal, due to respiratory paralysis. Despite the eradication of acute poliomyelitis, there remains a large population of patients with significant motor deficits who were infected before the onset of the vaccination programs.

The World Health Organization has now eradicated wild-type polio from all but four countries limited to central Africa. It is hoped that if mass vaccination programs are allowed to continue in central Africa, eradication there will be complete within a few more years.

In our insilico research project we have taken FASTA OF Human Polio Virus Sabin strain from GenBank with identification number AAN85444.1 polyprotein [Human poliovirus 3]. 64 linear B-epitopes are reported among them are [42]:

KEVPALTAVETGAT
ALTLSLPKQQDSLPDTKA
ATNPLAPSDTVQTRHVVQ
DNEQPTTRAQKLFAM
KEVPALTAVETGATNPLA

KHVRVWCPRPPRAVPYYG
NGHALNQVYQIMYIPPGA
QKLFAMWRITYKDTV

We have chosen wet lab results as standard parameter for Polio Virus Sabin Strain and used all five above mentioned online prediction tools, all wet lab results were taken by Hindawi Publishing Corporation Journal of Immunology Research, and these tools predicted different sequences that can be the potential epitopes. All these predicted epitopes were analyzed on the basis of wet lab results. Our parameter was minimum 70% match of predicted sequence with wet lab results, also we have chosen sequence length of 20 amino acids. Different predictors suggested some similar and different results, the sequences predicted by at least 4 of the predicted tools are:

ILVAYAPPGAQPPT
PPGAQPPTSRKEAM
AYAPPGAQPPTSRK
VRVWCPRPPRAVPY
WCPRPPRAVPYYGP
RPPRAVPYYGPGVD
NPSIFYTYGAAPAR
IFYTYGAAPARISV
AHSKEVPALTAVET
YIPPGAPTPKSWDDYTWQ

3.1.3 Mycobacterium Tuberculosis

Members of the genus Mycobacterium are characterized by a very complex cell wall envelope that is responsible for the remarkable low permeability of their cells as well as the characteristic differential staining procedure (known as Zhiel-Neelsen acid-faststain), which specifically stains all members of the genera. Both features are due to the presence of long chain a-alkyl, β-hydroxy fatty acids in their cell wall. The Mycobacterium genus is usually separated into two major groups on the basis of their growth rate. Tuberculosis remains the most devastating bacterial cause of human mortality (1). Despite improved diagnosis, surveillance, and treatment regimens, the incidence of TB increases annually (2). The ability to combat this deadly pathogen hinges on the dissection and understanding of the mechanisms of pathogenesis for Mycobacterium tuberculosis. Central to the ability of the microbe to cause disease is the capability to survive and replicate within macrophages by avoiding lysosomal fusion with the mycobacteria-containing phagosome. M. tuberculosis interacts with and invades various human and animal epithelial cells in culture and appears to possess multiple mechanisms of entry into macrophages. Furthermore, the specific bacterial adhesins involved in the complex interplay between M. tuberculosis and the human host are largely unknown. For Mycobacterium Tuberculosis 13 linear B-epitopes are reported [43]:

MTEQQWNFAGIEAAA
NFAGIEAAASAIQGN
ASAIQGNVTSIHSLL
NVTSIHSLLDEGKQS
SLLDEGKQSLTKLAA
KQSLTKLAAAWGGSG
AAWGGSGSEAYQGVQ
GSEAYQGVQQKWDAT
QQKWDATATELNNAL
TATELNNALQNLART
ALQNLARTISEAGQA
TISEAGQAMASTEGN
QAMASTEGNVTGMFA

In our insilico research project we have taken wet lab results as standard parameter for Mycobacterium Tuberculosis and used all five above mentioned online prediction tools, all wet lab results were taken by Department of Infectious Immunology, Statens Serum Institute, Copenhagen, Denmark by www. jimmunol. org/cgi/, these tools predicted different sequences that can be the potential epitopes. All these predicted epitopes were analyzed on the basis of wet lab results. Our parameter was minimum 70% match of predicted sequence with wet lab results, also we have chosen sequence length of 20 amino acids. Different predictors suggested some similar and different results, the sequences predicted by at least 3 of the predicted tools are:

NFAGIEAAASAIQGN,
AAWGGSGSEAYQGVQ
GSEAYQGVQQKWDAT
QQKWDATATELNNAL
TISEAGQAMASTEGN

3.1.4 Meningitis

Viral meningitis is contagious and infectious disease in which there is an inflammation of the membranes and cerebrospinal fluid (CSF). The membranes and cerebrospinal fluid (CSF) encase and bath the brain and spinal cord. Viral meningitis is the most common type of meningitis. Bacterial meningitis is less common. Viral meningitis is also sometimes called aseptic meningitis. Meningitis is by far the most common neurological manifestation of mumps virus infection. Before widespread immunization, mumps was a common cause of meningitis, which occurred in 15% of patients with mumps. Mumps meningitis can precede or follow the parotid swelling, and 50% of cases occur in the absence of parotitis. Meningitis is more common in male than female patients. Diagnostic tests include a lumbar puncture, also called a spinal tap. A lumbar puncture involves withdrawing a small sample of cerebrospinal fluid (CSF) from the spine with a needle. The sample of CSF is tested to rule-out bacterial meningitis and diagnose viral meningitis. Meningitis may be accompanied by mucocutaneous manifestations of enterovirus infection,

including localized vesicles such as in hand, foot, and mouth disease; herpangina; and generalized maculopapular rash. Most cases that present clinically with meningitis are self-limiting and carry a good prognosis. Nevertheless, enteroviral meningitis causes considerable morbidity, with moderate or high fever despite antipyretics and several days of severe headache warranting opiate analgesia. Abrupt deterioration in mental status or seizures may be caused by progression from meningitis to meningo encephalitis. No specific antiviral treatment is available, and management is conservative. Immunoglobulin replacement has a role in patients with hypogamma globulinemia, who are prone to severe and chronic enteroviral disease.]. For meningitis 9 linear B-epitopes are reported [44];

VDNQKQQHGALRNQGSRFHIKATHNFGD
ARTRTTETGKGVKTEKEKSVGVGLRVYF
FGDGFYAQGYLETRFVTKASENGSDNFGD
FGDITSKYAYVTLGNKAFGEVKLGRAKT
GEKTREQAVLFGVDHKLHKQLL
GVLATLGYRFSDLGLLVSLDSGYAKT
LSIIAEQSNSTVDNQK
YAKTKNYKIKHEKRYFVSPGFQYEL
YELMEDTNVYGNFKYERTSVDQGEKTR

In our insilico research project we have taken wet lab results as standard parameter for Meningitis and used all five above mentioned online prediction tools, these tools predicted different sequences that can be the potential epitopes. All these predicted epitopes were analyzed on the basis of wet lab results. Our parameter was minimum 70% match of predicted sequence with wet lab results, also we have chosen sequence length of 20 amino acids. Different predictors suggested some similar and different results, the sequences predicted by at least 3 of the prediction tools are:

VDNQKQQHGALRNQGSRFHIKATHNFGD
ARTRTTETGKGVKTEKEKSVGVGLRVYF
YELMEDTNVYGNFKYERTSVDQGEKTR

3.1.5 Plasmodium Vivax

Plasmodium vivaxis a protozoal parasite and a human pathogen. The most frequent and widely distributed cause of recurring (Benign tertian) malaria, P. vivax is one of the six species of malaria parasites that commonly infect humans. It is less virulent than Plasmodium falciparum, the deadliest of the six, but vivax malaria can lead to severe disease and death due to splenomegaly (a pathologically enlarged spleen). P. vivax is carried by the female Anopheles mosquito, since it is only the female of the species that bite.Plasmodium vivax malaria is prevalent in many regions of the world. It accounts for more than half of all malaria cases in Asia and Latin America. Despite the high prevalence of disease caused by this parasite, research into its effects has lagged disproportionately Organ dysfunction seen in P. falciparum malaria is not seen in P. vivax infections. Thus, severe malaria is reported with P. falciparum but not with P. vivax infection. 26 linear b cell epitopes are reported [45];

AYFLLGPVVKTLFNK
EGGSEFSERIGNSLS
EVIGNELADNIANEIVSSLQK
FDVKTQLKATAKKVL
FNKVEDVLHKPIPDT
KVLVEALKAALEPTE
LALFCFVNVLSLRGK
LEEEEAEDEFSDELLD
LKAALEPTEKIVAST
LKATAKKVLVEALKA
LQKDSASFLQSGFDV
MHLFNKPPKGKMNKV
NEIVSSLQKDSASFL
NKVNRVSIICAFLALFCFVNV
PDTIWEYESKGSLEE
PPKGKMNKVNRVSII
PTEKIVASTIKPPRVSEDAYFLLGPVV
PVVKTLFNKVEDVLH
SERIGNSLSSFLSES
SESASLEVIGNELAD
SFLQSGFDVKTQLKA
SLSSFLSESASLEVI
TASSSLEGGSEFSER
VLHKPIPDTIWEYES
VNVLSLRGKSGSTAS
YESKGSLEEEEAEDE

In our insilico research project we have taken wet lab results as standard parameter for Meningitis and used all five above mentioned online prediction tools, all wet lab results were taken by, these tools predicted different sequences that can be the potential epitopes. All these predicted epitopes were analyzed on the basis of wet lab results. Our parameter was minimum 70% match of predicted sequence with wet lab results, also we have chosen sequence length of 20 amino acids. Different predictors suggested some similar and different results, the sequences predicted by at least 3 of the predicted tools are:

FDVKTQLKATAKKVL
FNKVEDVLHKPIPDT
LALFCFVNVLSLRGK
LEEEEAEDEFSDELLD
LKAALEPTEKIVAST
PDTIWEYESKGS
PTEKIVASTIKPPRVSEDAYFLLGPVV
SESASLEVIGNELAD
TASSSLEGGSEFSER
YESKGSLEEEEAEDE

4 ACCURACY AND RELIABILITY OF ONLINE TOOLS

For the success of predictors, we define two measures, accuracy and reliability.

ACC=Number of correctly predicted epitopes by an online tool/ Number of epitopes in wet lab report (1)

REL=Number of correctly predicted epitopes by an online tool/ Number of all epitopes predicted by this online tool (2)

Table 2. Columns 2-6, and rows 2-6 contain the average accuracies of the five predictors in the first column on the five antigens in the first row computed along the above formula (1). The average accuracies of each of the predictors on five antigens are shown on the last column.

Antg	1	2	3	4	5	Acc%
ABC	46.15	50.00	11.11	46.88	46.15	40.06
BC	15.38	34.62	11.11	31.25	11.54	20.78
K&T	7.69	26.92	22.22	40.63	19.23	23.34
BEPI	23.08	53.85	44.44	28.13	11.54	32.21
AAP	23.08	42.31	11.11	29.69	19.23	25.08
					Overall Average %	28.29

From Table 2 it is seen that the highest accuracy 40.06 % is obtained by the predictor ABCpred. ABCpred correctly reported around 40 % of wet lab reported B-epitopes correctly.

Table 3. Columns 2-6, and rows 2-6 contain the average reliabilities of the five predictors in the first column on the five antigens in the first row computed along the above formula (2). Average reliabilities of each of the five predictors on five antigens are shown on the last column.

	1	2	3	4	5	Acc%
ABC	100.00	13.68	3.45	15.46	80.00	42.52
BC	100.00	32.14	11.11	44.44	75.00	52.54
K&T	33.33	33.33	20.00	48.15	100.00	46.96
BEPI	60.00	51.85	36.36	24.00	42.86	43.01
AAP	100.00	35.48	12.50	31.67	83.33	52.60
					Overall Average	47.53

From Table 3 it is seen that the highest reliability 100 % is achieved by several predictors on several antigens. These high reliabilities are due to small numbers of predicted epitopes. Indeed around 1/3 of linear B-epitopes reported by wet labs, are not predicted by any of the five predictors.

Another measure for the success of predictors may be the number of common true positives predicted by predictors.

Table 4. The percentage of predicted common B-epitopes, when the two preditors correctly predict wet lab reported B-epitopes.

Tool	ABC	BC	K&T	BEPI	AAP
ABC	-	15.22	13.04	13.77	18.84
BC	15.22	-	7.25	10.87	15.94
K&T	13.04	7.25	-	7.25	10.87
BEPI	13.77	10.87	7.25	-	13.04
AAP	18.84	15.94	10.87	13.04	-

Although 70% match threshold is used to decide the coincidence, the percentages of common epitopes are very low as seen in Table 4.

5 CONCLUSION

Computational insilico methods are favorable for their low cost and high speed compared to the use of experimental methods which are expensive in cost, and long in time to predict B-cell epitopes. Recently, many insilico tools are created, some of which are listed in Table 1. The tools employed in attempts to predict B-cell epitopes based on sequence and/or structural data have varying degrees of accuracy shown in Tables 1, 2, and 3.. In order to improve the accuracy of these predictions it is essential to "train" the algorithms by using the features achieved in wet lab in vivo experimental models. The shed some light on accuracy and reliability of these online tools, we set an insilico experiment on five antigens using five selected online tools.

On the bases of our findings, the most accurate tool is ABCpred with a score of 43.59 %, which means that the number of correctly predicted epitopes by this tool is 43.59 % of the all epitopes reported by wet lab experiments. While the most reliable predictors are BCpred and AAPred with scores of 52.54%, and 52.60% respectively, which means that the number of correctly predicted epitopes by these tools are of 52.54%, and 52.60% of all peptides they reported as b-epitopes. That is more than half of their predictions are correct b-epitopes of the antigens. What will be the percentage of true positives if we get the union of the predicted peptides by the two prediction tools? From Table 1, the sum of percentages of their true positives is 20.78+25.08 = 45.86 %. Subtracting half of the 15.94 % common true positives seen in From Table 3, one gets 45.84-15.94/2=37.87, which is slightly less than true positives of ABCpred. On the other hand the sum of the numbers of peptides reported by BCpred and AAPred is 26+26=52, while the numbers of peptides reported by ABCpred alone is 64. Although it is not a promising technique, combining several predictors one may get better predictions but still much less a satisfactory accuracy.

REFERENCES:

[1] Hopp, T.P., and Woods, K.R., Prediction of protein antigenic determinants from amino acid sequences, Proc. Nati. Acad. Sci. USA, Immunology, Vol. 78, No. 6, pp. 3824-3828, June 1981.

[2] Parker PJ, Coussens L, Totty N, Rhee L, Young S, Chen E, Stabel S, Waterfield MD, Ullrich A., The complete primary structure of protein kinase C--the major phorbol ester receptor, Science. 1986 Aug 22; 233 (4766): 853-9.

[3] Pellequer JL, Westhof E, Van Regenmortel MH. Predicting location of continuous epitopes in proteins from their primary structures. Methods Enzymol. 1991;203:176-201.

[4] Emini, EA, Hughes, JV, Perlow,DS and Boger, J. Induction of hepatitis A virus-neutralizing antibody by a virus-specific synthetic peptide. J Virol. 1985 Sep; 55(3): 836–839.

[5] Karplus, PA., Schultz, G. Prediction of chain flexibility in proteins. Naturwissenschaften (1985) 72, 212-213.

[6] [21a] Kolaskar, AS., and Tongaonkar, PC., A semi-empirical method for prediction of antigenic determinants on protein antigens, Volume 276, number 1,2, 172-174 FEBS 09210 December 1990

[7] Welling GW, Weijer WJ, van der Zee R, Welling-Wester S. Prediction of sequential antigenic regions in proteins. FEBS Lett. 1985 Sep 2;188(2):215-8.

[8] Bhasin, M., and Raghava, G. P. S. , ESLpred: SVM-based method for subcellular localization of eukaryotic proteins using dipeptide composition and PSI-BLAST, W414–W419 Nucleic Acids Research, 2004, Vol. 32, Web Server issue.

[9] Saha S, Raghava GP. Prediction of continuous B-cell epitopes in an antigen using recurrent neural network. Proteins. 2006 Oct 1;65(1):40-8.

[10] Lundegaard C, Lund O, Nielsen. Prediction of epitopes using neural network based methods. J Immunol Methods 2011, 374: 26–34

[11] Chen J, Liu H, Yang J, Chou KC (2007) Prediction of linear B-cell epitopes using amino acid pair antigenicity scale. Amino Acids 33: 423–428.

[12] Lawrence JK Wee1,2, Diane Simarmata1 ,Yiu-Wing Kam1 , Lisa FP Ng1,3, JooChuan Tong, SVM-based prediction of linear B-cell epitopes using Bayes Feature Extraction, Wee et al. BMC Genomics 2010, 11(Suppl 4):S21

[13] El-Manzalawy Y, Dobbs D, Honavar V (2008) Predicting linear B-cell epitopes using string kernels. Journal of Molecular Recognition 21: 243–255.

[14] Saha.S and Raghava G.P.S. BcePred: Prediction of Continuous B-Cell Epitopes in Antigenic Sequences Using Physico-chemical Properties. In G.Nicosia, V.Cutello, P.J. Bentley and J.Timis (Eds.) ICARIS 2004, LNCS 3239, 197-204, Springer,2004.

[15] J. Larsen, O. Lund, and M. Nielsen, “Improved method for predicting linear B-cell epitopes,” Immunome Research, vol. 2, article 2, 2006. W24–W29

[16] Gao J, Faraggi E, Zhou Y, Ruan J, Kurgan L (2012) BEST: Improved Prediction of B-Cell Epitopes from Antigen Sequences. PLoS ONE 7(6): e40104.

[17] Wang Y, Wu W, Negre NN, White KP, Li C, Shah PK (2011) Determinants of antigenicity and specificity in immune response for protein sequences. BMC Bioinformatics, 12:251

[18] Sweredoski MJ, Baldi P (2009) COBEpro: a novel system for predicting continuous B cell epitopes. Protein EngDesSel, 22(3):113-120

[19] Ansari HR1, Raghava GP. , Identification of conformational B-cell Epitopes in an antigen from its primary sequence. Immunome Res. 2010 Oct 20;6:6.

[20] Andersen, PH., Nielsen, M., and Lund, O., Prediction of residues in discontinuous B cell epitopes using protein 3D structures, Protein Science, 15:2558-2567, 2006.

[21] Liang, S., Liu, S., Zhang, C., and Zhou, Y. `A simple reference state makes a significant improvement in near-native selections from structurally refined docking decoys., Proteins 69, 244--253 (2007)

[22] Yao Lian, Y., Ge, M., and Pan, X-M., EPMLR: sequence-based linear B-cell epitope prediction method using multiple linear regression. BMC Bioinformatics. 2014; 15(1): 414.

[23] Marciniuk, K., Trost, B., and Napper, S., EpIC: a rational pipeline for epitope immunogenicity characterization, Bioinformatics. 2015 Jul 15; 31(14): 2388–2390.

[24] Krawczyk, K., Liu, X., Baker, T., Shi, J., and Deane, C.M., Improving B-cell epitope prediction and its application to global antibody-antigen docking. Structural Bioinformatics, Vol. 30 no. 16 2014, pages 2288–2294.

[25] Rubinstein ND, Mayrose I, Martz E, Pupko T (2009) Epitopia: a web-server for predicting B cell epitopes. BMC Bioinformatics, 10:287

[26] Rubinstein ND, Mayrose I, Pupko T (2009) A machine-learning approach for predicting B cell epitopes. Mol Immunol, 46: 840¬847

[27] El-Manzalawy Y, Dobbs D, Honavar V (2008) Predicting flexible length linear B cell epitopes. ComputSyst Bioinformatics Conf, 7:121-132

[28] Chen J, Liu H, Yang J, Chou K (2007) Prediction of linear B-cell epitopes using amino acid pair antigenicity scale. Amino Acids 33: 423-428.

[29] EL-Manzalawy Y, Dobbs D, Honavar V (2008) Predicting linear B-cell epitopes using string kernels. J MolRecognit 21: 243-255

[30] Gupta, s., Ansari, HR., Gautam, A., and Raghava, GPS. Identification of B-cell epitopes in an antigen for inducing specific class of antibodies. Biol Direct. 2013, 8:27

[31] Kolaskar, AS., and Tongaonkar, PC., A semi-empirical method for prediction of antigenic determinants on protein antigens, Volume 276, number 1,2, 172-174 FEBS 09210 December 1990

[32] 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

[33] Saravanan V., Gautham N., Harnessing Computational Biology for Exact Linear B-Cell Epitope Prediction: A Novel Amino Acid Composition-Based Feature Descriptor, OMICS. 2015 Oct;19(10):648-58.

[34] Chang HT, Liu CH, Pai TW (2008) Estimation and extraction of B cell linear epitopes predicted by mathematical morphology approaches. J MolRecognit, 21(6):431-441

[35] Wang HW, Lin YC, Pai TW, Chang HT (2011) Prediction of B cell linear epitopes with a combination of support vector machine classification and amino acid propensity identification. J Biomed Biotechnol, 2011:432830

[36] (Sela-Culang I., Ashkenazi, S., Peters, B., and Ofran, Y., et al., 2015) PEASE: predicting B-cell epitopes utilizing antibody sequence. Bioinformatics, Volume 31, Issue 8, 15 April 2015, Pages 1313–1315

[37] Alix AJ (1999) Predictive estimation of protein linear epitopes by using the program PEOPLE. Vaccine, 18:311-314

[38] Pellequer JL, Westhof E (1993) PREDITOP: a program for antigenicity prediction. J Mol Graph, 11:191-202

[39 Dalkas, GA., and Rooman, M., SEPIa, a knowledge-driven algorithm for predicting conformational B-cell epitopes from the amino acid sequence, Dalkas and Rooman BMC Bioinformatics (2017) 18:95.

[40] Yao B, Zhang L, Liang S, Zhang C (2012) SVMTriP: A Method to Predict Antigenic Epitopes Using Support Vector Machine to Integrate Tri-Peptide Similarity and Propensity. PLoS ONE 7(9): e45152.

[41] Raul Isea. Quantitative Prediction of Linear B-Cell Epitopes. Biomedical Statistics and Informatics. Vol. 2, No.1, 2017, pp.1-3.

[42] Nomoto A., Omata T., Toyoda H., Kuge S., Horie H., Kataoka Y., Genba Y., Nakano Y., Imura N., Complete nucleotide sequence of the attenuated poliovirus Sabin 1 strain genome." Proc. Natl. Acad. Sci. U.S.A. 79:5793-5797(1982)

http://www.uniprot.org/uniprot/P03301#sequences

[43] Yang, Y., Chen, H., Liu, Z., Ma, H., Qin, L., Jin, J., Zheng, R., Feng, Y., Cui, Z., Wang, J., Liu, J., Hu,.Z, A Novel B-Cell Epitope Identified within Mycobacterium tuberculosis CFP10/ESAT-6 Protein, Published: January 7, 2013https://doi.org/10.1371/

[44] Chandra, S., and Singh, TR., Linear B cell epitope prediction for epitope vaccine design against meningococcal disease and their computational validations through physicochemical properties, Netw Model Anal Health Inform Bioinforma (2012) 1:153–159.

[45] Caro-Aguilar, I., Rodríguez, A., Calvo-Calle, JM., Guzmán, F., De la Vega, P., Patarroyo, ME., Galinski, MR., and Moreno, A. Plasmodium vivax Promiscuous T-Helper Epitopes Defined and Evaluated as Linear Peptide Chimera Immunogens, INFECTION AND IMMUNITY, July 2002, p. 3479–3492 Vol. 70, No. 7