



UOIuBIH
ORSinBIH
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

Validation Tools for Predicted Linear B-Epitopes: Antigenicity

S.A.R. Abidi

M. Can

Faculty of Engineering and Natural Sciences,
International University of Sarajevo International University of Sarajevo,
Hrasnicka Cesta 15, Ildža 71210 Sarajevo,
Bosnia and Herzegovina
mcan@ius.edu.ba
azraabidi34@yahoo.com

Article Info

Article history:

Article received on 10 May 2018

Received in revised form 17 June 2018

Keywords:

Antigenic determinant; Protein;
Prediction; Semi-empirical method;
Sequential antigenic site

ABSTRACT: A important step in designing peptide vaccines involves the identification of antigenic regions in a protein. It will be helpful to synthesize peptides which may elicit antibodies reactive with the intact protein. Earliest method used by Levitt (1976) was folding simulations, other pioneer works by Hopp and Woods (1981), Parker et. al. (1986) are based on the assumption that antigenic regions are primarily hydrophilic regions at the surface of the protein molecule. On the other hand the method presented by Welling, et. al., (1985) is based on the comparison of propensities of amino acids in known antigenic regions in 20 proteins, with that of 314 proteins. While Kolaskar, and Tongaonkar (1990) experimentally observed that if hydrophobic residues occur on the surface of a protein, they are more likely a part of the antigenic sites. In this article, the same method of Welling, et. al. is applied with relatively huge data of 80,592 non redundant proteins of PDB Database, and 344,121 linear b-epitopes of iadb database. The success of antigenicity scale obtained is compared by other five scales on five antigens with known linear b-epitopes.

1. INTRODUCTION

Antibodies are sensitive to certain parts of native proteins that are called antigenic residues. To be accessible, these sites are most probably on the surfaces of a antigenic proteins. Furthermore these regions are possibly more mobile than interior regions, and are hydrophilic. Indeed, from seventies on, scores for hydrophilicity, flexibility, and surface accessibility have been used to predict antigenic regions.

Recently, new tools are emerged to predict the antigenic sites, which are called epitopes of antigenic proteins. Prediction of immunogenic epitopes using bioinformatics tools is a challenging task because of the inherent complexity of antigen recognition.

2. SCORE TABLES TO PREDICT ANTIGENIC DETERMINANTS

The identification of B cell epitopes on protein antigens has attracted the attention of many scientists. This would be useful for diagnostic purposes and also in the development of peptide vaccines. To save time and money in wet labs experiments, Levitt, (Levitt, 1976) started a tradition to create score tables to predict antigenic determinants. Hopp and Woods (1981) followed.

Parker et al., (1986) modified the approach of Hopp and Woods taking into account that antigenic sites are on the surface of the protein. They used three parameters - hydrophilicity, accessibility and flexibility simultaneously.

On the other hand, Welling et al. (1985) calculated the antigenicity value for each amino acid from its frequency of occurrence in antigenic regions in 20 proteins, with that of 314 proteins, and used these values to predict epitopes. Then in this article similar approach is used database used by these workers is very small and consists of only 606 amino acids from 20 proteins. In this article the same approach is used for a relatively big data of 80,592 non redundant proteins of PDB, and 344,121 linear b-epitopes of iadb database. Kolaskar, and Tongaonkar derived a score table using experimental antigenic determinant data and physicochemical properties of amino acids.

2.1 Lewitt Scores o Predict Antigenic Determinants

Levitt (1976) showed how the concept of time-averaged forces, can be used to simplify conformational energy calculations on globular proteins. Folding simulations are done under a variety of conditions, and the relevance of such calculations to the actual in vitro folding process is discussed at some length. These same techniques have many potential applications including enzyme-substrate binding, changes in protein tertiary and quaternary structure, and protein-protein interactions. Using binding energies in kcal/mol, he succeeded to give a score table of solvent parameter values as in Table 1.

Table 1 Antigenicity scales

aa	Lew	Hopp	Park	Welli	Kolas	Mcan
A	-0.5	-0.5	2.1	0.115	1.06	0.05
R	3.0	3.0	4.2	0.058	0.87	0.02
N	0.2	0.2	7.0	0.077	0.78	0.02
D	2.5	3.0	10.0	0.065	0.87	-0.02
C	-1.0	-1.0	1.4	-0.12	1.41	0.06
E	2.5	3.0	6.0	-0.071	0.85	0.32
Q	0.2	0.2	7.8	-0.011	1.02	0.22
G	0.0	0.0	5.7	-0.184	0.87	0.02
H	-0.5	-0.5	2.1	0.312	1.11	-0.49
I	-1.8	-1.8	-8.0	0.292	1.15	-0.17
L	-1.8	-1.8	-9.2	0.075	1.25	-0.08
K	3.0	3.0	5.7	0.206	0.93	-0.01
M	-1.3	-1.3	-4.2	0.385	0.83	-0.11
F	-2.5	-2.5	-9.2	-0.141	1.09	-0.13
P	-1.4	0.0	2.1	0.053	1.06	0.12
S	0.3	0.3	6.5	0.026	1.01	0.10
T	-0.4	-0.4	5.2	0.045	0.91	0.08
W	-3.4	-3.4	-10.0	-0.114	0.89	-0.13
Y	-2.3	-2.3	-1.9	0.013	1.16	-0.17
V	-1.5	-1.5	-3.7	0.013	1.38	-0.04

2.2. Hopp and Woods Score Table to Predict Antigenic Determinants

Hopp and Woods (1981) presented an antigenicity score list fo 20 amino acids for locating protein antigenic determinants. This is accomplished by assigning each

amino acid their hydrophilicity value and then repetitively averaging these values along the peptide chain. They claim that the point of highest local average hydrophilicity is located in, or immediately adjacent to, an antigenic determinant. The method was developed using 12 proteins for which extensive immunochemical analysis has been carried out and the score table in Table 2. subsequently used to predict antigenic determinants for several proteins.

2.3. Parker et. al. Score Table to Predict Antigenic Determinants

Parker et. al. (1986) score table to predict antigenic determinants of a protein is indeed a set of hydrophilicity high-performance liquid chromatography (HPLC) parameters. These parameters were derived from the retention times of 20 model synthetic peptides, Hydrophilicity parameters have been used extensively in algorithms to predict which amino acid residues

are antigenic, they compared the profiles generated by nine other sets of parameters. Generally, it is found that the parameters obtained by Parker et. al. (1986) in Table 2., correlated with antigenicity. In addition, it was shown that a combination of the three best parameters for predicting antigenicity further improved the predictions. The hydrophilic, accessible, and flexible regions were then correlated to the known antigenic sites from immunological studies and accessible sites determined by X-ray crystallographic data for several proteins.

2.4. Welling et. al. Algorithm to Predict Antigenic Determinants

Some of the previously studied methods are based on the assumption that antigenic regions are primarily hydrophilic at the surface of the antigenic protein. The method by Welling et. al. (1985) is based on the amino acid composition of known antigenic regions in 20 proteins which is compared with that of 314 proteins. Antigenicity values were derived from the differences between the two data sets. The score table in Table 2. is applied to some antigenic proteins and a good correlation between the predicted regions and antigenic regions previously determined in wet lab.

2.5. Kolaskar, and Tongaonkar Algorithm to Predict Antigenic Determinants

Kolaskar, and Tongaonkar (1990) observed in data from experimentally determined antigenic sites on proteins that if hydrophobic residues occur on the surface of a protein, they are more likely to be a part of antigenic sites. They developed a semi-empirical method which uses physic chemical properties of amino acid residues and their frequencies of occurrence in wet lab reported epitopes to predict antigenic determinants on proteins. They claim that the method can predict antigenic determinants with about 75%. The algorithm consists of one score table as seen in Table 2.

2.6. Can Antigenicity Score

To obtain antigenicity scores of twenty amino acids, the same method of Welling, et. al. (1985) is applied with relatively huge data the propensity vector of amino acids in 80,592 non redundant proteins downloaded from PDB Database, and the propensity vector of amino acids in 344,121 linear b-epitopes from iadb database are obtained as in the second and third columns of the Table 2. The third column is the division of the second and third columns, which is the relative abundance of amino acids in epitopes with respect to their abundance in proteins. The last column is the logarithm of the third column with the base 10. This column is the antigenicity scores of the 20 amino acids in Table 1.

Table 2. The third column is the relative abundance of amino acids in epitopes. The last column is the logarithm of the third column with the base 10, which is the antigenicity scores of the 20 amino acids.

aa	In E %	In P %	R Abon.	Log10
A	8.88	7.94	1.1184	0.0486
R	5.27	4.99	1.0561	0.0237
N	4.47	4.28	1.0444	0.0189
D	5.48	5.73	0.9564	-0.0194
C	1.6	1.39	1.1511	0.0611
E	7.87	3.8	2.0711	0.3162
Q	4.49	2.7	1.6630	0.2209
G	7.73	7.37	1.0489	0.0207
H	2.12	6.61	0.3207	-0.4939
I	3.72	5.51	0.6751	-0.1706
L	7.47	9.05	0.8254	-0.0833
K	5.8	5.88	0.9864	-0.0059
M	1.73	2.23	0.7758	-0.1103
F	2.92	3.95	0.7392	-0.1312
P	6.15	4.65	1.3226	0.1214
S	7.96	6.4	1.2438	0.0947
T	6.72	5.56	1.2086	0.0823
W	1.03	1.39	0.7410	-0.1302
Y	2.34	3.48	0.6724	-0.1724
V	6.24	6.89	0.9057	-0.0430

3. MATERIALS AND METHODS

To have an idea about the success of using several antigenicity tables to predict the linear b-epitopes of antigenic peptides, a sample of five antigens, Plasmodium Falciparum, Human Polio Virus Sabin Strain, Meningitis, Plasmodium Vivax and Mycobacterium Tuberculosis are considered.

3.1 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 microcirculation 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 given in Isea, R. (2017), and Abidi, and Can (2017).

3.2 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.

FASTA OF Human Polio Virus Sabin strain is down loaded from GenBank with identification number AAN85444.1 polyprotein [Human poliovirus 3]. 64 linear B-epitopes are reported (Nomoto, et. al., 1982; Kanduc, et. al., 2015; Abidi, and Can (2017).

3.3 Mycobacterium Tuberculosis

Members of the genus *Mycobacterium* are characterized by a very complex cell wall envelope that irresponsible for the remarkable low permeability of their cells as well as the characteristic differential staining procedure (known as Zhiel-Neelsen acid-fast stain), which specifically stains all members of the genera. Both features are due to the presence of long chain α -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 Young et. al., (2013), and Abidi, and Can (2017).

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 meningoencephalitis. No specific antiviral treatment is available, and management is conservative. Immunoglobulin replacement has a role in patients with hypogammaglobulinemia, who are prone to severe and chronic enteroviral disease. J. For meningitis 9 linear B-epitopes are reported Chandra, and Singh (2012), and Abidi, and Can (2017).

3.5 Plasmodium Vivax

Plasmodium vivax is 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 Caro-Aguilar, et. al., (2002), and Abidi, and Can (2017).

4. PREDICTING ANTIGENIC DETERMINANTS USING SCORE TABLES

In this section as a demonstration, the antigenic residues of *Mycobacterium Tuberculosis* are found. The fasta of the antigen consists of 95 residues:

```
MTEQQWNFAGIEAAASAIQGNVTSIHSLLDDEGKQSL
TKLAAAWGGSGSEAYQGVQKWDATATELNNAL
QNLARTISEAGQAMASTE GNV TGMFA
```

Using the formula

$$a_i = \frac{1}{7} \sum_{j=1}^7 c_{i+j-4}, \quad i = 4, \dots, 92$$

Where $c_k, k = 1, \dots, 95$ is the antigenicity score from Table 2. of the amino acid at the position k of the sequence.

When the local average antigenicity scores of residues are plotted, we get an antigenicity profile of the antigenic protein as in Figure 1.

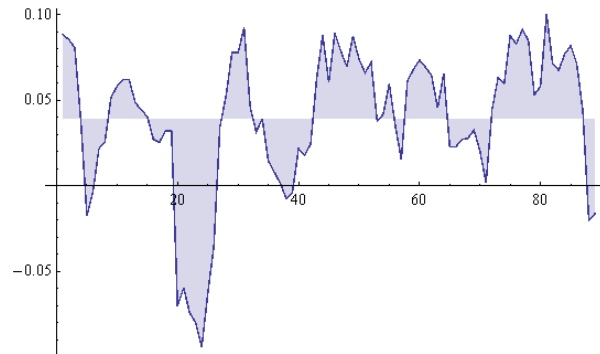


Figure 1. Antigenicity profile of the antigenic protein.

The average antigenicity of the antigenic protein is found to be 0.039, and as seen as the boundary of the shaded regions. We claim that the projections of shaded regions upper side of the average on the horizontal coordinate axis are antigenic regions on the protein. The starting and ending addresses of these regions, when regions shorter than 6 residues are eliminated are

{ {9,15}, {43,52}, {58,64}, {72,87} }

which correspond to the strings

```
AGIEAAA, WGGSGSEAYQ, WDATATE, LARTI
SEAGQAMASTE
```

The wet lab reported antigenic sites of this antigenic protein are

```
MTEQQWNFAGIEAAA, NFAGIEAAASAIQGN,
ASAIQGNVTSIHSLLD, NVTSIHSLLDDEGKQS,
SLLDEGKQSLTKLAA, KQSLTKLAAAWGGSG,
AAWGGSGSEAYQGVQ, GSEAYQGVQKWDAT,
QQKWDATATELNNAL, TATELNNALQNLART,
ALQNLARTISEAGQA, TISEAGQAMASTE GN,
QAMASTE GNV TGMFA
```

The numbers of residues in red are

{7,7,2,1,2,5,10,6,7,4,11,13,7}

The numbers of residues in web lab reported peptides are

{15,15,15,15,15,15,15,15,15,15,15,15}

In three predicted regions, the number of correctly predicted residues are

{10, 11, 13}.

These antigenic regions are accepted as correctly predicted, since more than half of residues are correctly anticipated.

This calculation is repeated for each of five antigens and six antigenicity scores.

5. CONCLUSION

When the calculation in Section 4. is repeated for each of five antigens and six antigenicity scores, we get the following Table 3.

Table 3. When calculation in Section 4. is repeated for each of five antigens and six antigenicity scores, we get the table of correctly predicted antigenic regions.

Antigen	Lab	Lew	Mc	Hop	Pa	We	KK
Menin	9	3	3	3	2	2	2
Falcip	26	6	13	19	9	7	4
Polio	49	15	11	16	20	49	33
Esat6	13	5	3	3	4	8	6
Vivax	26	5	5	10	9	8	10
Total	123	34	35	51	44	74	55

From Table 3. It is seen that the antigenicity scores of Welling et. al. (1985) performs best. Although antigenicity scores list is based on observations of the amino acid composition of known antigenic regions of 20 proteins and other 314 proteins. Almost 2/3 of the antigenic regions are correctly predicted. On the other hand, in this research a data of 80,592 non redundant proteins of PDB Database, and 344,121 linear b-epitopes of iadb database are used with the same technique, and the resulted antigenicity score list could predict only 1/3 of antigenic regions. The abundance of the information weakens the efficiency.

REFERENCES

Abidi, S.A.R., Can, M. (2017) On the Accuracies of Sequence Based Linear B Cell Epitope Predictors, Southeast Europe Journal of Soft Computing Vol.6 No.2 September, pp.1-11

Caro-Aguilar, I., Rodríguez, A., Calvo-Calle, JM., Guzmán, F., De la Vega, P., Patarroyo, ME., Galinski, MR., and Moreno, A. (2002) 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

Chandra, S., and Singh, TR. (2012) 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 1:153–159.

Hopp, T.P. and Woods, K.R. (1981) Prediction of protein antigenic determinants from amino acid sequences, Proc. Natl. Acad. Sci. USA 78, 3824-3828.

Isea, R. (2017) Quantitative Prediction of Linear B-Cell Epitopes. Biomedical Statistics and Informatics. Vol. 2, No.1, pp.1-3.

Kanduc, D., Fasano, C., Capone, G., Delfino, AP., Calabrò, M., and Polimeno, M. (2015) Applying the Concept of Peptide Uniqueness to Anti-Polio Vaccination, Journal of Immunology Research, Volume 2015, Article ID 541282, 8 pages

Kolaskar AS, Tongaonkar PC. (1990). A semi-empirical method for prediction of antigenic determinants on protein antigens. FEBS Lett 276:172-174.

Levitt, M. (1976) A simplified representation of protein conformations for rapid simulation of protein folding, J. Mol. Biol. 104, 59-107.

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

Parker, J. M. R., Guo, D., and Hodges, R. S. (1986) New Hydrophilicity Scale Derived from High-Performance Liquid Residues with Antigenicity and X-ray-Derived Accessible Sites? Biochemistry, 25, 5425-5432.

Welling, G.W., Weijer, W.J., van der Zee, R. and Welling-Wester, S. (1985) Prediction of sequential antigenic regions in proteins, FEBS Lett. 188, 215-218.

Yang, Y., Chen, H., Liu, Z., Ma, H., Qin, L., Jin, J., Zheng, R., Feng, Y., Cui, Z., Wang, J., Liu, J., Hu, Z. (2013) A Novel B-Cell Epitope Identified within