VOL.7 NO.1 March 2018 - 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

Validation Tools for Predicted Linear B-Epitopes: Surface Flexibility

S.A.R. Abidi
M. Can

Faculty of Engineering and Natural Sciences,
International University of Sarajevo International University of Sarajevo,
Hrasnicka Cesta 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:
flexibility index,
antigenic regions,
epitopes

ABSTRACT: Protein structural flexibility is important for catalysis, and binding. Flexibility has been predicted from amino acid sequence with a sliding window averaging technique and applied primarily to epitope search. Karplus, and Shultz divided amino acids in two categories, rigid and flexible. Then, proposed three flexibility scales for residues with no rigid neighbors, residues with one rigid neighbor, and residues for which both neighbors are rigid. To use flexibility scores for validation of candidate b-epitopes, 337,259 b-epitopes whose lengths are larger than seven amino acids are downloaded from iedb database, and 103,590 non epitopes are created randomly. When computed by Karplus-Shultz technique, b-epitopes achieved an average score of 1.0159, while the average score for non epitops is 0.9752. It is seen that the Karplus, and Shultz flexibility computation technique can be used as a validation tool with the criteria that the peptide with higher score, is more likely an epitope.

1. INTRODUCTION

One of the physic-chemical properties of proteins that is important for catalysis, and binding is the protein structural flexibility (Gaspar, and Csermely, 2012). Peptide flexibility has been predicted from amino acid sequence with a sliding window averaging technique and applied primarily to epitope search.

Flexibility analysis can be used to search for the most mobile and thus possibly also the surface residues in a sequence, which are thought to represent epitopes. For vaccine production, it would be of great value to be able to predict the antigenic regions of a protein from its sequence. Therefore flexibility predictions have been used in searching for continuous epitopes from amino acid sequence. Other epitope prediction methods include hydrophilicity, beta turn propensity, surface accessibility, and antigenicity (Ragone, et. al. 1989; Radivojac, et. al., 2004). Karplus, and, Schulz, (Karplus, and, Schulz, 1985) proposed a flexibility prediction technique. They start by amino acid sequences of 31 proteins with a sliding window averaging technique.

After Karplus, and, Schulz, (1985), mainly three other techniques have been used for predicting protein flexibility from amino acid sequence. The methods of Bhaskaran and Ponnuswamy (Bhaskaran and Ponnuswamy , 1989) is based on parameters derived from three-dimensional structures. By the use of a differential equation model, they

computed fluctuation amplitudes of the amino acid residues in 19 protein molecules are and are analyzed statistically to get new information as to their stability, position dependent nature, distribution and group dynamical behavior. For each of the residue types, the symmetry of the distributions of amino acid residues in the protein molecules is described by introducing a parameter called the flexibility index. Ragone et al. (Ragone et al., 1989) base their approach on a combination of hydropathy predictions and amino acid volumes.

Then Vhinen et. al., (Vhinen et. al. 1994) derived new prediction parameters from 92 refined protein structures in an unbiased selection of the Protein Data Bank by developing further the method of Karplus and Schulz, 1985. They compared the accuracies of three previous flexibility prediction techniques with theirs. They found the optimized size of the prediction window for each method, which was 9.

2. KARPLUS AND SCHULZ METHOD

The flexibility prediction technique proposed by Karplus, and, Schulz, (1985) has three flexibility scales. The predicted relative flexibility at residue position n of a given amino acid were refined by a nearest-neighbor analysis. First, the 20 amino acid types were divided into 2 groups, rigid, A, L, H, V, Y, I, F, C,W, M, and flexible, others. Then, separate average Bnorm-values were listed for residues with no rigid neighbors as in D1 column of table 1, residues with one rigid neighbor in D2, and residues for which both neighbors are rigid in D3.

Table 1. Scales for average Bnorm-values for residues with no rigid neighbors in D1column, residues with one rigid neighbor in D2, and residues for which both neighbors are rigid in D3.

	A Acid	D1	D2	D3
1	A	1.041	0.946	0.892
2	R	1.038	1.028	0.901
3	N	1.117	1.006	0.93
4	D	1.033	1.089	0.932
5	C	0.96	0.878	0.925
6	E	1.094	1.036	0.933
7	Q	1.165	1.028	0.885
8	G	1.142	1.042	0.923
9	H	0.982	0.952	0.894
10	I	1.002	0.892	0.872
11	L	0.967	0.961	0.921
12	K	1.093	1.082	1.057
13	M	0.947	0.862	0.804
14	F	0.93	0.912	0.914
15	P	1.055	1.085	0.932
16	S	1.169	1.048	0.923
17	T	1.073	1.051	0.934
18	W	0.925	0.917	0.803
19	Y	0.961	0.93	0.837
20	V	0.982	0.927	0.913

2.1 KS Computation of Flexibility Scores

Karplus, and, Schulz, (Karplus, and, Schulz, 1985) proposed a formula to find a flexibility score (F) for a residue at sequence position n. The calculation is based on three flexibility scales in Table 1 on a weighted average with weights

$$w=\{0.25,0.50,0.75,1.,\ 0.75,0.50,0.25\} \quad (1)$$

within the window of size seven. The flexibility profile is obtained using the formulae

$$F_n = \sum_{i=1}^{7} w_i D_{k,n-4+i} \quad (2)$$

Where F_n is the flexibility score for a residue at sequence position n, $D_{k,n-4+i}$ is the flexibility scale value for a residue at sequence position $n-4+i$. $k = 1,2,3,$ depending on the neighbors of the residue at the sequence position n w_i is weight attached to the residue at the window position i.

3. TESTING KARPLUS AND SCHULZ VALIDATION TOOL ON FIVE ANTIGENS

To have an idea about the success of using several antigicity 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 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).

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 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). For Mycobacterium Tuberculosis 13 linear B-epitopes are reported Young et. al., (2013)), and Abidi, and Can (2017)

3.4 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. For meningitis 9 linear B-epitopes are reported Chandra, and Singh (2012)), Abidi, and Can (2017).

3.5 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 Caro-Aguilar, et. al., (2002)), and Abidi, and Can (2017)..

3.6 Predicting Antigenic Determinants Using Karplus And Schulz flexibility scales

In this section as a demonstration, the antigenic residues of Meningitis are found. The FASTA of the antigen consists of 361 residues:

MKKTLAALIVGAFAASAANAAVVYNNEGTNVELG
GRLSIIAEQSNSTVDNQKQQHGALRNQGSRFHIKAT
HNFGDGFYAQGYLETRFVTKASENGSDNFGDITSKY
AYVTLGNKAFGEVKLGRAKTIADGITSAEDKEYGV
LNNSDYIPTSGNTVGYTFKGIDGLVLGANYLLAQKR
EGAKGENKRPNDKAGEVRIGEINNGIQVGAKYDAN
DIVAKIAYGRTNYKYNESDEHKQQLNGVLATLGYR
FSDLGLLVSLDSGYAKTKNYKIKHEKRYFVSPGFQY
ELMEDTNVYGNFKYERTSVDQGEKTREQAVLFGVD
HKLHKQLLTYIEGAYARTRTTETGKGVKTEKEKSV
GVGLRVYF

Using the Equation (2) in the above, the flexibility scores of residues are obtained. When these scores are plotted, we get a surface accessibility profile of the antigenic protein Menengitis as in Figure 1.

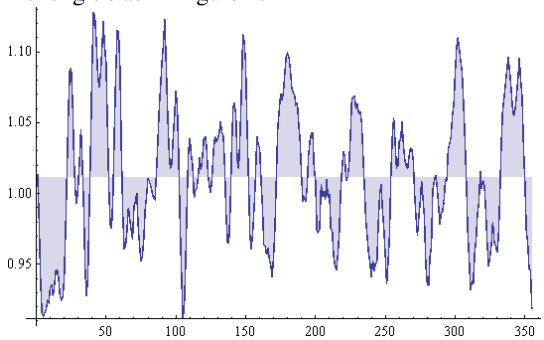

Figure 1. Surface accessibility profile of the antigenic protein Meningitis.

The average surface accessibility of the antigenic protein is found to be 1.01194, 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 high flexibility regions of the antigen that can be taken as the antigenic regions. The starting and ending addresses of these regions, when regions shorter than 6 residues are eliminated are

{{39,51},{86,102},{116,123},{126,135},{140,152},{172, 189},{224,234},{254,270},{295,307},{333,349}}

which correspond to the strings

IIAEQSNSTVDNQ,
RFVTKASENGSDNFGDI,
FGEVKLGR,
TIADGITSAE,
GVLNNSDYIPTSG,
LLAQKREGAKGENKRPND,
NYKYNESDEH,
LVSLDSGYAKTKNYKIK,
FKYERTSVDQGEK,
YARTRTTETGKGVKTEK

The wet lab reported antigenic sites of this antigenic protein are

VDNQKQQHGALRNQGSRFHIKATHNFGD,
ARTRTTETGKGVKTEKEKSVGVGLRVYF,
FGDGFYAQGYLET**RFVTKASENGSDNFGD**,
FGDITSKYAYVTLGNKA**FGEVKLGR**AKT,
GEKTREQAVLFGVDHKLHKQLL,
GVLATLGYRFSDLGLLVSLDSGYAKT,
LS**IIAEQSNSTVDNQ**K,
YAKTKNYKIKHEKRYFVSPGFQYEL,
YELMEDTNVYGN**FKYERTSVDQGEK**TR

The numbers of residues in red (bold) are

{4,16,16,8,3,11,13,10,13}

The numbers of residues in web lab reported peptides are

{28,28,29,28,22,26,16,25,27}.

In three predicted regions, the number of correctly predicted residues are {16,16,13}

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

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

Antigen	Lab	K&S
Menin	9	3
Falcip	26	6
Polio	49	9
Esat6	13	2
Vivax	26	8
Total	123	43

From Table 3. It is seen that flexibility scores performs best on the antigen Plasmodium Vivax.

4. A THRESHOLD FOR B-EPITOPE VALIDATIONS

There is a belief that antigenic peptides are more flexible than others. To support the hypothesis more strongly, 337,259 b-epitopes of lengths 7-25 amino acids are downloaded from iedb database[1], average flexibility scores of b-epitopes are computed according to the formula (2).

On the other hand 103,590 non epitopes of length 7-25 amino acids are constructed artificially. Their average flexibility scores are also computed by the same formula (2). The descriptive statistics of the results are shown in Table 4.

Table 4. The descriptive statistics of the average flexibility scores of epitopes, and non epitopes computed by Karplus, and Shults scales.

Flexibility	Epitopes	Nonepitopes
Mean	1.0159	0.9752
S. Deviation	0.0429	0.0404
Skewness	-0.0071	0.2125
Kurtosis	2.6727	2.9164

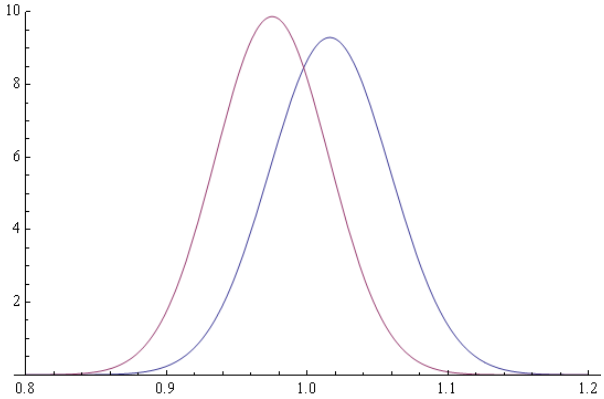

Figure 2 Average flexibility scores of epitopes, and non epitopes distribute almost normally, with slightly different means 1.0159, and 0.9752 respectively.

[1] http://www.iedb.org/database_export_v3.php

5. VHINEN SCALES

Vhinen et. al. method differs only in Table 1. They found that the optimal window size is nine, and weights for a window should be

w={0.25, 0.4375, 0.6250, 0.8225, 1., 0.8225, 0.6250, 0.4375, 0.25} (3)

To see flexibility scores computed using Whinen scales, 317,040 b-epitopes of lengths 9-25 amino acids are downloaded from iedb database, average flexibility scores of b-epitopes are computed according to the same formula (2) with window size nine..

On the other hand 83,654 non epitopes of length 9-25 amino acids are constructed artificially. Their average flexibility scores are also computed by a modified formula.

Table 5. Modified scales of Vhinen et. al for average Bnorm-values for residues with no rigid neighbors in D1column, residues with one rigid neighbor in D2, and residues for which both neighbors are rigid in D3.

	A Acid	D1	D2	D3
1	A	1.315	0.994	0.783
2	R	1.31	1.026	0.807
3	N	1.38	1.022	0.799
4	D	1.372	1.022	0.822
5	C	1.196	0.939	0.785
6	E	1.376	1.052	0.826
7	Q	1.342	1.041	0.817
8	G	1.382	1.018	0.784
9	H	1.279	0.967	0.777
10	I	1.241	0.977	0.776
11	L	1.234	0.982	0.783
12	K	1.367	1.029	0.834
13	M	1.269	0.963	0.806
14	F	1.247	0.934	0.774
15	P	1.342	1.05	0.809
16	S	1.381	1.025	0.811
17	T	1.324	0.998	0.795
18	W	1.186	0.938	0.796
19	Y	1.199	0.981	0.788
20	V	1.235	0.968	0.781

The descriptive statistics of the results obtained with scales in Table 5. are shown in Table 6.

Table 6. The descriptive statistics of the average flexibility scores of epitopes, and non epitopes obtained by Vihinen scales.

Flexibility	Epitopes	Nonepitopes
Mean	1.1134	0.9752
S. Deviation	0.1090	0.0849
Skewness	0.1247	0.3960
Kurtosis	2.6912	3.1131

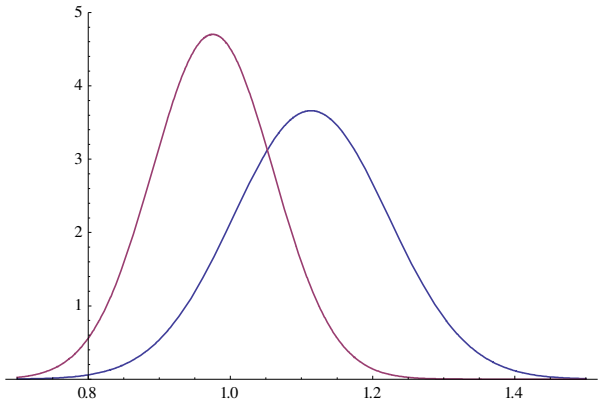

Figure 3 Average flexibility scores of epitopes, and non epitopes for Vihinen scales distribute also almost normally, with slightly different means 1.1134, and 0.9752 respectively.

.

4. CONCLUSION

Online Karplus & Schulz Flexibility Prediction tool in iedb Analysis Resource, yields flexibility scores of residues of peptides that are epitope candidates. It is possible also to compute average flexibility scores of these peptides. To derive a validation tool from this residue flexibility prediction tool in iedb to validate whether a given peptide is a linear b-epitope or not, one needs a threshold value. For this purpose b-epitopes from iedb database are downloaded. Also large number of nonepitopes are created.

Using both Karplus, and Shults, and Vhinen et. al. flexibility scales, average flexibility scores of epitopes, and non epitopes are computed. The percentage difference between average flexibility scores of epitopes and non epitopes is found to be less than 0.03%

Therefore if average flexibility scores of a peptide is larger, the peptide is more likely is a b-epitope, and this property can be used as validation tool to eliminate some of the peptides with small average flexibility scores, while keeping others.

REFERENCES

Bhaskaran, R., Ponnuswamy, P.K. (1988) Positional flexibilities of amino acid residues in globular proteins. Int. J. Peptide Prot. Res. 32:241-255.

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, 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.

Gaspar, ME., and Csermely, P., (2012) Rigidity and flexibility of biological networks, Briefings in Functional Genomics. Vol 11. No 6. 443- 456.

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

Karplus, P. A,, Schulz, G. E. (1985) Prediction of chain flexibility in proteins, Naturwissenschaften 72:212-213.

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.

Radivojac, P., Obradovic, Z., Smith, DK., Zhu, G., Vucetic, S., Brown, CJ., Lawson, JD., And Dunker, AK., (2004) Protein flexibility and intrinsic disorder, Protein Science 13:71–80.

Ragone, R., Facchiano, F., Facchiano, A,, Facchiano, A. M., Colonna, G. (1989) Flexibility plot of proteins. Prot. Eng. 2:497-504

Udhayakumar V1, Anyona D, Kariuki S, Shi YP, Bloland PB, Branch OH, Weiss W, Nahlen BL, Kaslow DC, Lal AA., (1995) Identification of T and B cell epitopes recognized by humans in the C-terminal 42-kDa domain of the Plasmodium falciparum merozoite surface protein (MSP)-1. J Immunol. Jun 1;154(11):6022-30.

Vihinen, M., Torkkila, E., and Riikonen, P., (1994) Accuracy of Protein Flexibility Predictions, PROTEINS: Structure, Function, and Genetics 19:141-149.

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 Mycobacterium tuberculosis CFP10/ESAT-6 Protein, Published: January 7,

https://doi.org/10.1371/journal.pone.00528