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 Accessibility

S.A.R. Abidi
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

Article Info

Article history:
Article received on 10 January 2018
Received in revised form 27 February 2018

Keywords:
accessible surface; B-cell epitopes; Janin's accessible surface probabilities; validation

ABSTRACT: Identifying B-cell epitopes plays an important role in vaccine design, immunodiagnostic tests, and antibody production. Therefore, computational tools for reliably predicting B-cell epitopes are highly desirable. In this article the possibility of usage of accessible surface scores of peptides as a validation tool is studied. Janin et al. determined empirical amino acid accessible surface probabilities of twenty amino acids. With these fractional surface probabilities for amino acids, a surface probability (S) at sequence position n can be found using a formula given by Emini et. al. When a peptide is uploaded to the Emini Surface Accessibility Prediction in iedb Analysis Resource, prediction tool separates residues into two groups buried, and surface according to the average of S_ns. To create a criterion to decide whether a given peptide is a linear b-epitope or not, for 344,121 b-epitopes downloaded from iedb database, average buried and exposed probabilities, as well as the ratio ρ of averages for these b-epitopes are computed. The same is done for 111,306 artificially created non epitopes. It is seen that for b-epitopes, the ratio ρ is significantly larger than the ratio ρ for the non epitopes. Therefore if ρ is larger, the peptide is more likely is a b-epitope, and this property can be used as to rank peptides while choosing the most probable linear b-epitope from a long list.

1. INTRODUCTION

Quantitative estimations of the protein-water interactions is very difficult. The role of the solvent is very important in processes involving proteins and other biological molecules. The physics of substances in liquid state, especially of liquid water, is very complex. To study the physics of liquids a number of approaches have been proposed, including molecular dynamics, and Monte Carlo simulations of water surrounding protein molecules.

However, these calculations are very cumbersome even for the smallest proteins. The use of the accessible surface areas simplifies the problem. Assumption of that their free energy is proportional to the surface of contact between the protein molecule and water, and other procedures developed removes the major difficulties linked to the geometric constructions used previously in accessibility measurements. In those procedures accessible surface areas are calculated simply and efficiently as a combination of pairwise interactions between residues.

2. MATERIALS AND METHODS

Surface accessibility of a protein to the solvent is estimated from the accessible surface area of each residue in the protein structure, Janin et al. (1978). Lee, and Richards (1971) defined this area as a surface over which a water molecule can be placed so that it makes contact with an atom of the residue without penetrating any other atom of the structure (Richmond, 1984). Janin et al. (1978), used a program of Levitt to compute accessible surface area from X-ray co-ordinates (Levitt, and Greer 1977).

Residues are then classed as:

1) Buried if their accessible surface area A is smaller than 20 $Å^2$,
2) Intermediate if A is between 20 and 60 $Å^2$,
3) Exposed if A is larger than 60 $Å^2$.

This classification is based on average values of A calculated by Chothia (1976).

Relying on these computations, Janin et al. (1978), determined empirical amino acid accessible surface probabilities of twenty amino acids as seen in Table 1.

Table 1. Empirical amino acid accessible surface probabilities of amino acids

	A Acid	Buried %	δ_n	Exposed %
1	Ala	51	49	15
2	Arg	5	95	67
3	Asn	22	78	4 9
4	Asp	19	81	50
5	Cys	74	26	5
6	Glu	16	84	55
7	Gln	16	84	35
8	Gly	52	48	10
9	His	34	66	34
10	Ile	66	34	13
11	Leu	60	40	16
12	Lys	3	97	85
13	Met	52	48	20
14	Phe	58	42	10
15	Pro	25	75	4 5
16	Ser	35	65	32
17	Thr	30	70	32
18	Trp	49	51	17
19	Tyr	24	76	41
20	Val	64	36	14

Emini et. al. 1985 proposed a formula to find a surface probability (S) at sequence position n. The calculation is based on surface accessibility scale on a product instead of an addition within the window. The accessibility profile is obtained using the formulae

$$S_n = 0.37^{-6} \prod_{i=1}^{6} \delta_{n-4+i} \quad (1)$$

Where S_n is the surface probability, δ_n is the fractional surface probability value, and i vary from 1 to 6.

3. IMMUNE EPITOPE DATABASE AND ANALYSIS RESOURCE IEDB

IEDB is funded by a the National Institute of Allergy and Infectious Diseases, which traces its origins to a small laboratory established in 1887 at the Marine Hospital on Staten Island, New York. It offers easy searching of experimental data characterizing antibody and B, and T cell epitopes studied in humans, non-human primates, and other animal species. Epitopes involved in infectious disease, allergy, autoimmunity, and transplant are included.

The IEDB also hosts tools to assist in the prediction and analysis of B cell and T cell epitopes. Analysis Resource implements among others the method of Emini surface accessibility scale.

In Emini Surface Accessibility Prediction in iedb Analysis Resource[1], $\delta_n = 100 - Buried$ in table 1. For each amino acid of an uploaded peptide, surface probability S_n is computed. Then a normalized surface accessibility score S_{0n} is obtained by the formula

$$S_{0n} = \frac{S_n}{\bar{S}} \quad (2)$$

Where $\bar{S}$ is the mean of the surface probabilities S_n in the peptide. A residue at the position n of the peptide with S_n greater than 1.0 indicates an increased chance for being found on the surface of the peptide.

As an example when the peptide with the sequence

VDNQKQQHGALRNQGSRFHIKATHNFGD

is uploaded to the Emini Surface Accessibility Prediction Analysis Resource in iedb, it returns Figure 1.

[1] http://tools.iedb.org/bcell/

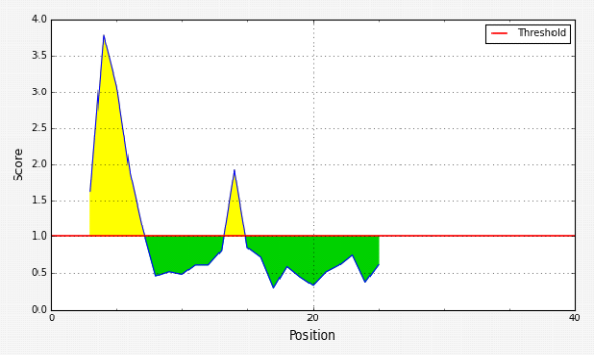

Figure 1. Emini Surface Accessibility Prediction results for the given peptide: Residues with S_{0n} greater than 1.0 indicates an increased probability of these residues for being found on the surface.

The results of Emini Surface Accessibility Prediction tool is a relative information. Whichever peptide you upload to the prediction tool, it separates residues into two groups buried, and surface according the threshold 1, which corresponds to the average of S_ns.

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 (Demir, and Can 2018).

4. PREDICTING ANTIGENIC DETERMINANTS USING EMINI SURFACE ACCESSABILITY SCORE TABLE

In this section as a demonstration, the antigenic residues of Plasmodium Vivax are found. The FASTA of the antigen consists of 196 residues:

MHLFNKPPKGKMNKVNRVSIICAFLALFCFVNVLSL
RGKSGSTASSSLEGGSEFSERIGNSLSSFLSESASLEVI
GNELADNIANEIVSSLQKDSASFLQSGFDVKTQLKA
TAKKVLVEALKAALEPTEKIVASTIKPPRVSEDAYFL
LGPVVKTLFNKVEDVLHKPIPDTIWEYESKGSLEEEE
AEDEFSDELLD

Using the formula

$$a_i = \frac{1}{7}\sum_{j=1}^{7} c_{i+j-4}, \quad i = 4, \ldots, 193$$

Where $c_k, k = 1, \ldots, 193$ is the antigenicity score from Table 1. of the amino acid at the position k of the sequence.

When the local average Emini surface accessibility scores of residues are plotted, we get a surface accessibility profile of the antigenic protein Plasmodium Vivax as in Figure 1.

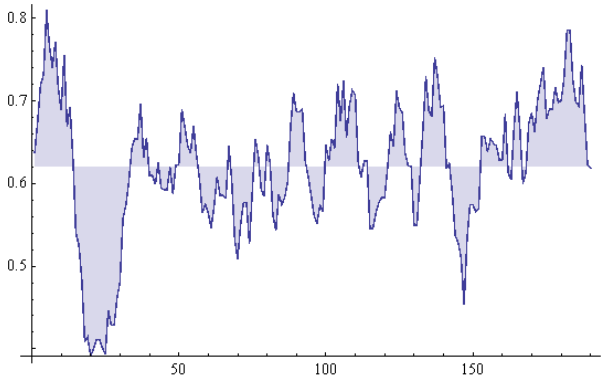

Figure 1. Surface accessibility profile of the antigenic protein Plasmodium Vivax.

The average surface accessibility of the antigenic protein is found to be 0.621, 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 surface accessibility regions of the antigen that can be taken as the antigenic regions on the protein. The starting and ending addresses of these regions, when regions shorter than 6 residues are eliminated are

{{1,14},{49,56},{100,111},{122,129},{133,140},{153,161},{169,189}}

which correspond to the strings

MHLFNKPPKGKMNK,EGGSEFSE,QSGFDVKT
QLKA,KAALEPTE,ASTIKPPR,VKTLFNKV,
PDTIWEYESKGSLEEEEAEDE

The wet lab reported antigenic sites of this antigenic protein are

AYFLLGPVVKTLFNK,**EGGSEFSE**RIGNSLS,
EVIGNELADNIANEIVSSLQK,FDVKTQLKAT
AKKVL,FNKVEDVLHKPIPDT,KVLVEAL**KAA
LEPTE**,LALFCFVNVLSLRGK,LEEEEAEDEF
SDELLD,LKAALEPTEKIVAST,LKATAKKVL
VEALKA,LQKDSASFL**QSGFDV**,**MHLFNKPPK
GKMNK**V,NEIVSSLQKDSASFL,NKVNRVSII
CAFLALFCFVNV,PDTIWEYESKGSLEE,PPK
GKMNKVNRVSII,PTEKIV**ASTIKPPR**VSEDA
YFLLGPVV,PV**VKTLFNKV**EDVLH,SERIGNS
LSSFLSES,SESASLEVIGNELAD,SFL**QSGF
DVKTQLKA**,SLSSFLSESASLEVI,TASSSLE
GGSEFSER,VLHKPI**PDTIWEYES**,VNVLSLR
GKSGSTAS,**YESKGSLEEEEAEDE**

The numbers of residues in red are

{7,8,2,9,5,8,2,9,8,3,6,14,2,3,15,8,8,9,2,3,12,3,8,9,2,15}

The numbers of residues in web lab reported peptides are

{15, 15, 21, 15, 15, 15, 15, 16, 15, 15, 15, 15, 15, 21, 15, 15, 27, 15, 15, 15, 15, 15, 15, 15, 15, 15}

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

{8,9,8,,9,8,14,15,8,8,9,12,8,9,15}.

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

When the calculation 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 Emini surface accessibility score, we get the table of correctly predicted antigenic regions.

Antigen	Lab	Emini
Menin	9	4
Falcip	26	1
Polio	49	18
Esat6	13	7
Vivax	26	13
Total	123	43

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

5. A B-EPITOPE VALIDATION TOOL BY RANKING VIA SURFACE ACCESSIBILITY SCORES

A b-epitope ranking tool is created by the use of buried and exposed propensities of amino acids in Table 1.

$$E_n = \prod_{i=1}^{6} e_{n-4+i} \tag{3}$$

$$B_n = \prod_{i=1}^{6} b_{n-4+i} \tag{4}$$

Scores of amino acids in a peptide is computed. Here e_k and b_k are the exposed and buried propensities of amino acids in Table 1. Then overall surface accessibility ratio ρ of the peptide is computed by the formula

$$\rho = \frac{\sum_n E_n}{\sum_n B_n} \tag{5}$$

If this score ρ is larger, the more residues of the peptides are exposed, and more likely the peptide is an epitope.

To compute average exposed and buried amino acids, and an overall ρ value for 344,121 b-epitopes downloaded from iedb database, and for 111,306 non epitopes created and filtered randomly are used.

Table 4. Although average buried residues surface accessibility score of b-epitopes is almost the same as the non-epitope peptides, average exposed residues of surface accessibility score of b-epitopes is much higher than non-epitope peptides. The ratio ρ for epitopes is also significantly larger than the ratio ρ for the non epitopes

	Epitopes		Nonepitopes	
	Exposed	Buried	Exposed	Buried
Mean	0.075	0.003	0.056	0.003
STDV	0.056	0.005	0.005	0.004
Skewness	2.103	4.866	4.866	3.917
Kurtosis	10.066	34.837	34.837	29.240
	Rho	27.320	Rho	18.456

From the Table 4. we conclude that if ρ is larger for a peptide among others, this peptide is more likely is a b-epitope, and this property can be used as to rank peptides while choosing the most probable linear b-epitope from a long list.

6. CONCLUSION

Emini Surface Accessibility Prediction tool in iedb Analysis Resource yields information about antigenic sites of antigenic proteins. On five antigens, it is used to predict linear b-epitopes. The results are tabulated in Table 3. It is seen that around 1/3 of wet lab reported b-epitopes are predicted through Emini surface accessibility scores.To create a criterion to decide whether a given peptide is a linear b-epitope or not, for 344,121 b-epitopes downloaded from iedb database, average buried and exposed probabilities, as well as the ratio ρ of averages for these b-epitopes are computed. The same is done for 111,306 non epitopes. It is seen that for b-epitopes, the ratio ρ is significantly larger than the ratio ρ for the non epitopes. Therefore if ρ is larger, the peptide is more likely is a b-epitope, and this property can be used as to rank peptides while choosing the most probable linear b-epitope from a long list.

REFERENCES

Abidi, A., 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

Abidi, S. A. R., and Can, M. (2018) Validation Tools for Predicted Linear B-Epitopes: Surface Flexibility, Southeast Europe Journal of Soft Computing Vol.7 No.1 March, pp.32-37.

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

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

Janin, J., S. Wodak, M. Levitt, and B. Maigret. (1978) Conformation of amino acid side-chains in proteins. J. Mol. Biol. 125:357-386.

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,

Lee B, Richards FM. (1971) The interpretation of protein structures: estimation of static accessibility, J Mol Biol. 14;55(3):379-400.

Levitt M, Greer J., (1977) Automatic identification of secondary structure in globular proteins, J Mol Biol. 5;114(2):181-239.

Richmond, T.J. (1984)Solvent accessible surface area and excluded volume in proteins: Analytical equations for overlapping spheres and implications for the hydrophobic effect, Journal of Molecular Biology Volume 178, Issue 1, 5 Pages 63-89.