VOL.7 NO.2September 2018 - ISSN 2233 – 1859: DOI: 10.21533/scjournal.v7i2.144

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: Beta Turns

S.A.R. Abidi
Faculty of Engineering and Natural Sciences,
International University of Sarajevo International University of Sarajevo,
HrasnickaCesta 15, Ilidža 71210 Sarajevo,
Bosnia and Herzegovina
azraabidi34@yahoo.com

Article Info

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

Keywords:
hydrophilicity/hydrophobicity of B-cell epitopes; Parker's scale, Cornette scale, Doolittle scale

ABSTRACT: It is claimed thatamino acid replacements in surface loops usually do notperturb the three-dimensional structure of the protein,since surface loops are relatively flexible (Saunders andBaker 2002). Thus, the conservation variability of epitopesmight be biased by the abundance of loops inepitopes. These results imply that epitopes do not tendto overlap functional regions, but rather cover separateregions. Pellequera, et. al., (1993), developed new turn scales based on the occurrence of amino acids at each of the four positions of a turn using a structural database comprised of 87 proteins. It is found that the scales correctly predicted a fraction of the turn regions in proteins with approximately 80% confidence. They used the turn scales for predicting the location of antigenic sites in proteins. The method was developed with the specific aim of predicting only a few peaks for each protein. They found that it leads to a high level of accurate prediction around 70% of correct prediction of known epitopes. In this article we update turn scales using large numbers of proteins and epitopes. Improved method will be more helpful in selecting protein regions to be synthesized in order to produce anti-peptide antibodies cross-reacting with the parent protein.

1. INTRODUCTION

Antigenic sites are residues of a native protein that are those recognized by antibodies. It is most probably that these sites are accessible or on the surface ofanantigenic proteins. Furthermore, these regions are possibly more mobile thaninterior regions. Since these sites are on the surface, they are probably hydrophilic. Indeed, scores for hydrophilicity, flexibility, and accessibility have been used to predict antigenicity.

At least for the past three decades, 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(Flower,

2003).At the beginning, to predict continuous B-cell epitopes, a number of algorithms have been developed based on physico-chemical properties of amino acids (Pellequer, et. Al. 1991), but their accuracies were very low. The commonly used properties are hydrophilicity (Parker, et. Al., 1986) flexibility (Karplus, and Schulz1985), accessibility (Emini, et. al 1985), and beta turns (Pellequer, et.al., 1993).

2. CHOU FASMAN SCORE TABLES TO PREDICT BETA TURNS

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 i.e, 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 iedb database. Kolaskar, and Tongaonkar derived a score table using experimental antigenic determinant data and physicochemical properties of amino acids.

Table 4. Modified scales of Vhinen et. al for average Bnorm-values for residueswith no rigid neighbors inD1column, residues with one rigid neighbor inD2, and residues for which both neighbors are rigid in D3.

No	A Acid	C&F	No	A Acid	C&F
1	A	0.66	11	L	0.59
2	R	0.95	12	K	1.01
3	N	1.56	13	M	0.60
4	D	1.46	14	F	0.60
5	C	1.19	15	P	1.52
6	E	0.74	16	S	1.43
7	Q	0.98	17	T	0.96
8	G	1.56	18	W	0.96
9	H	0.95	19	Y	1.14
10	I	0.47	20	V	0.50

3. MATERIALS AND METHODS

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 causes 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), andAbidi, 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 verycomplex cell wall envelope that is responsible for the remarkablelow permeability of their cells as well as the characteristicdifferential staining procedure (known as Zhiel-Neelsen acid-faststain), which specifically stains all members of the genera. Bothfeatures are due to the presence of long chain a-alkyl, β-hydroxyfatty acids in their cell wall. TheMycobacterium genus is

usuallyseparated into two major groups on the basis of their growthrate.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) andAbidi, and Can (2017).

3.4 Meningitis

Viral meningitis is contagious and infectious disease in which there is an inflammation of themembranes of 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 males than in females. 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 of the high prevalence of disease caused by this parasite, research regarding its effects has lagged disproportionately.Organ dysfunction which is 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), andAbidi, and Can (2017).

4. PREDICTING ANTIGENIC DETERMINANTS USING SCORE TABLES

In this section as ademonstration, 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 formula

$$a_i = \frac{1}{7}\sum_{j=1}^{7} c_{i+j-4}, \quad i = 4, \ldots, 358$$

Where $c_k, k = 1, \ldots, 361$ is the betaturn 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.

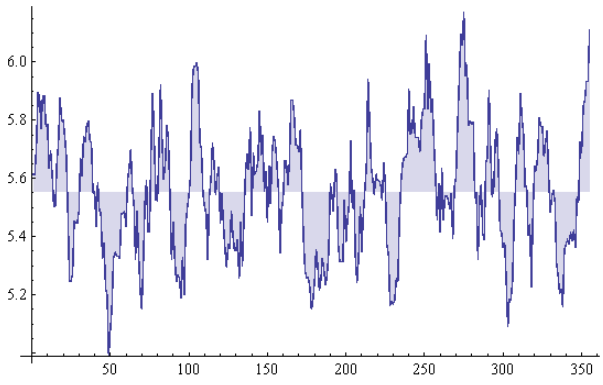

Figure 1. Beta turn profile of the antigenic protein.

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

{{1,13},{16,22},{31,38},{81,87},{101,108},{141,147},{161,172},{235,257},{269,281},{308,315},{320,329}}

which correspond to the strings

MKKTLAALIVGAF,SAANAAV,VELGGRLS,GYLETRF,DITSKYAY,VLNNSDY,GIDGLVLGANYL,QQLNGVLATLGYRFSDLGLLVSL,IKHEKRYFVSPGF,TREQAVLF,KLHKQLLTYI

The wet lab reported antigenic sites of this antigenic protein are

VDNQKQQHGALRNQGSRFHIKATHNFGD,ARTRTT
ETGKGVKTEKEKSVGVGLRVYF,FGDGFYAQGYLE
TRFVTKASENGSDNFGD,FGDITSKYAYVTLGNKAF
GEVKLGRAKT,GEKTREQAVLFGVDHKLHKQLL,GV
LATLGYRFSDLGLLVSLDSGYAKT,LSIIAEQSNSTV
DNQK,YAKTKNYKIKHEKRYFVSPGFQYEL,YELME
DTNVYGNFKYERTSVDQGEKTR

The numbers of residues in red are

{2,2,7,8,8,19,2,13,2}

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

{19, 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. Whencalculation 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	C&F
Menin	9	2
Falcip	26	7
Polio	49	12
Esat6	13	4
Vivax	26	10
Total	123	35

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 iedb 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.

5. 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 scoresof 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	1.0091	0.9877
S. Deviation	0.1331	0.1098
Skewness	0.2275	0.0823
Kurtosis	2.9618	3.1126

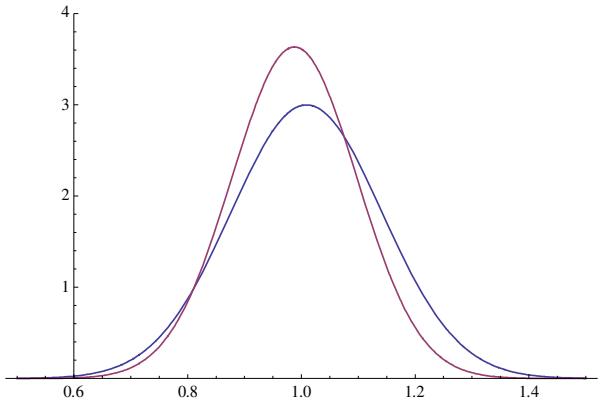

Figure 2 Average beta turn scores of epitopes, and non-epitopes distribute almost normally, with slightly different means 1.0091, and 0.9877 respectively.

4. CONCLUSION

Online Karplus & Schulz Flexibility Prediction tool in iedb Analysis Resource, yieldsflexibility 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 iedbto validate whether a given peptide is a linear b-epitope or not, one needs a threshold value.

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

For this purposeb-epitopes from iedb database are downloaded. Also, large number of non-epitopes 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 smallaverage flexibility scores, while keeping others.

REFERENCES

Aboderin A.A.(1971) Amino acid scale: Mobilities of amino acids on chromatography paper (RF). Int. J. Biochem. 2:537-544.

Abraham D.J., Leo A.J. (1987) Reference: Proteins: Structure, Function and Genetics 2:130-152.

Bull H.B., Breese K. (1974) Amino acid scale: Hydrophobicity (free energy of transfer to surface in kcal/mole).Arch. Biochem. Biophys. 161:665-670.

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.

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

Cheng, J., Tegge, AN., andBaldi, P. (2008) Machine Learning Methods for Protein Structure Prediction, Ieee Reviews In Biomedical Engineering, Vol. 1, 41.

Cornette, L, Cease, KB., Margalit, H., Spouge, JL., Berzofsky, JL., and DeLisi, C.(1987) Hydrophobicity scales and computational techniques for detecting amphipathic structures in proteins, Journal of Molecular Biology, Volume 195, Issue 3, Pages 659-685

Cowan R., Whittaker R.G.(1990) Amino acid scale: Hydrophobicity indices at ph 7.5 determined by HPLC. Peptide Research 3:75-80.

Eisenberg D., Schwarz E., Komarony M., Wall R.(1984) Amino acid scale: Normalized consensus hydrophobicity scale. J. Mol. Biol. 179:125-142**.**

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

Emini, EA, Hughes, JV, Perlow, DS, and Boger J. (1985) Induction of hepatitis A virus-neutralizing antibody by a virus-specific synthetic peptide. J. Virol. vol. 55 no. 3 836-839

Engelman, D. M., Steitz, T. A., and Goldman, A. (1986). Identifying nonpolar transbilayer helices in amino acid sequences of membrane proteins. Annu Rev BiophysBiophys Chem, 15:321-353.

Fauchere J.-L., Pliska V.E.(1983)Amino acid scale: Hydrophobicity scale (pi-r). Eur. J. Med. Chem. 18:369-375.

Flower DR. (2003) Towards in silico prediction of immunogenic epitopes.Trends Immunol; 24:667–674.

Guy H.R. (1985) Amino acid scale: Hydrophobicity scale based on free energy of transfer (kcal/mole). Biophys J. 47:61-70.

Hessa T, Kim H, Bihlmaier K, Lundin C, Boekel J, Andersson H, Nilsson I, White SH, von Heijne G. (2005)Recognition of transmembrane helices by the endoplasmic reticulum translocon. Nature, 433(7024):377-81.

Hopp T.P., Woods K.R. (1981) Amino acid scale: Hydrophilicity. Proc. Natl. Acad. Sci. U.S.A. 78:3824-3828.

Janin, J. (1979). Surface and inside volumes in globular proteins. Nature, 277(5696):491-492.

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.

Karplus, PA, and Schulz, G.E. (1985) Prediction of chain flexibility in proteins – a tool for the selection of peptide antigens, Naturwissenschaft, vol 72, pp 212-213.

KavithaK V, Saritha R, Vinod Chandra S S, (2013) Computational Methods in Linear B-cell Epitope Prediction, International Journal of Computer Applications (0975 – 8887) Volume 63– No.12.

Kolaskar, AS., Tongaonkar, PC (1990) A semi-empirical method for prediction of antigenic determinants on protein antigens, FEBS Lett,vol 276,pp172-174.

Kyte J., Doolittle R.F.(1982) Amino acid scale: Hydropathicity. J. Mol. Biol. 157:105-132.

Larsen, J., Lund, O., and Nielsen, M.(2006) Improved method for predicting linear B-cell epitopes, Immunome Research, vol. 2, article 2, W24–W29

Lawrence JKW, Simarmata , D. Kam, YW. Ng, LFP, Tong, JC. (2010) SVM-based prediction of linear B-cell epitopes using Bayes Feature Extraction, Wee et al. BMC Genomics, 11(Suppl 4):S21.

Miyazawa S., Jernigen R.L. (1985) Amino acid scale: Hydrophobicity scale (contact energy derived from 3D data).Macromolecules 18:534-552.

Moon CP, Fleming KG. (2011) Side-chain hydrophobicity scale derived from transmembrane protein folding into lipid bilayers. Proc Natl Acad Sci USA. 108(25):10174-7.

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 JMD, Guo D, Hodges RS. (1986). New hydrophilicity scale derived from high-performance liquid chromatography peptide retention data: correlation of predicted surface residues with antigenicity and X-ray-derived accessible sites. Biochemistry;25: 5425–5432.

Pellequer, JL, Westhof', E., and Van Regenmortela MHV (1993)Correlation between the location of antigenic sites and the prediction of turns in proteins, Immunology Letters, 36 83-100.

Pellequer J-L, Westhof E, Regenmortel MHV. (1993) Correlation between the location of antigenic sites and the prediction of turns in proteins. Immunol Lett;36:83–99.

Pellequer JL, Westhof E, Regenmortel MHV. (1991) Predicting location of continuous epitopes in proteins from their primary structures. Methods Enzymol;203:176–201.

Rose, G. D. (1978) Prediction of chain turns in globular proteins on a hydrophobic basis, Nature 272:586–590.

Roseman M.A. (1988) Amino acid scale: Hydrophobicity scale (pi-r). J. Mol. Biol. 200:513-522.

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

Saunders CT1, Baker D. (2002) Evaluation of structural and evolutionary contributions to deleterious mutation prediction.J Mol Biol. ;322(4):891-901.

Sweet R.M., Eisenberg D.(1983) Amino acid scale: Optimized matching hydrophobicity (OMH). J. Mol. Biol. 171:479-488

Tanford C.(1962) Amino acid scale: Hydrophobicity scale (Contribution of hydrophobic interactions to the stability of the globular conformation of proteins). J. Am. Chem. Soc. 84:4240-4274.

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

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.

Wilson K.J., Honegger A., Stotzel R.P., Hughes G.J.(1981)Amino acid scale: Hydrophobic constants derived from HPLC peptide retention times. Biochem. J. 199:31-41.

Wimley WC, White SH. (1996) Experimentally determined hydrophobicity scale for proteins at membrane interfaces. Nat Struct Biol. (10):842-8. Attribute assignment filewwHydrophobicity.txt.

Wolfenden R.V., Andersson L., Cullis P.M., Southgate C.C.F. (1981) Amino acid scale: Hydration potential (kcal/mole) at 25øC. Biochemistry 20:849-855.

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 (2013) Novel B-Cell Epitope Identified within Mycobacterium tuberculosis CFP10/ESAT-6 Protein, https://doi.org/10.1371/journal.pone.0052848

Zhang, W., and Niu, Y. (2010) Predicting flexible length linear B-cell epitopes using pair wise sequence similarity, Third International Conference on Biomedical engineering and Informatics.