International Journal of Biomedical and Clinical Sciences
Vol. 4, No. 1, 2019, pp. 29-32
http://www.aiscience.org/journal/ijbcs

Linear B-cell Epitope in *Mycobacterium tuberculosis* Using a Similarity Index

Raúl Isea[1, *], Karl Erik Lonngren[2]

[1]Fundación Instituto de Estudios Avanzados, Hoyo de la Puerta, Baruta, Venezuela
[2]Department of Electrical and Computer Engineering, University of Iowa, Iowa City, USA

Abstract

At the present time, almost one third of the human population resides in regions where one could contract Tuberculosis and the standard Bacille Calmette-Guerin (BCG) vaccine is not able to protect adults. For this reason, a new vaccine against this disease must be developed. This paper identifies all the known epitopes of B cells in order to assist in the development of such a new vaccine against this disease. In order to do that, the epitopes of the B cell present in the genome of M. tuberculosis will be presented using six different computer programs. To decrease the number of epitopes predicted computationally, a restriction based on using the Similarity Index will be imposed. With this limitation, the number of epitopes is reduced to 198 epitopes. Using this result, it should be possible to perform laboratory studies to validate each of these epitopes which would allow one to develop a more useful vaccine against tuberculosis.

Keywords

Tuberculosis, Similarity Index, B Epitope, *Mycobacterium*

Received: February 5, 2019 / Accepted: March 29, 2019 / Published online: April 17, 2019

1. Introduction

Actually, it has been estimated that 1.3 million people died from Tuberculosis (TB) in 2017 and almost 1/3 of the world's population has been infected with this disease. It is one of the main causes of mortality in people infected with HIV [1]. TB is caused by the bacterium *Mycobacterium tuberculosis* that is transmitted in air through the inhalation of droplets introduced into the atmosphere by infected individuals when coughing or talking. This in turn would introduce the bacterium into the lungs.

At the moment, the Bacilla Calmette-Guérin (BCG) is a vaccine that has been used against TB although this vaccine is infrequently used in the United States. It is given to infants and young children in countries where TB is more although BCG may not always protect adults against this disease. For this reason, it is necessary to develop a new vaccine. This vaccine could arise from computational methodologies that examine new antigenic proteins present in the genome and obtaining the necessary epitopes.

The goal of this work is tofind the linear Bcell epitopes in *Mycobacterium* by using the *Similarity Index*. With this information it may be possibleto design new vaccines against this disease.

2. Material and Methods

We identify all proteinsthat are present in the *Mycobacterium tuberculosis* genome available in NCBI's resources (https://www.ncbi.nlm.nih.gov/). The next step is to determine the lineal B cell epitopes using the following computer programs: BePiPred [2], Emini Surface Accessibility Prediction [3], Kolaskar and Tongaonkar Antigenicity [4], ABCpred [5], SVMTriP [6] and Bcpred [7]. Finally, we obtained a consensus B linear epitope according to the procedure described by Isea et al [8-11]. To do this, we

* Corresponding author
E-mail address: raul.isea@gmail.com (R. Isea)

employed a script developed in Python that is capable of overlapping all the B-cell epitopes for each antigen where we select the B cell epitope with three or more coincidences that will be found. In view of the high number of epitope predictions that have been obtained, we will only select those epitomes that are at least 5 *mer* in length and had a value of the Similarity Index that is greater than 3.Remember that the higher the value obtained from the Similarity Index should be the best candidate for the design of vaccines against this disease as previously explained in [12].

3. Results

Table 1 shows the results of this study in two columns. The first column presents the antigen obtained from the *Mycobacterium tuberculosis* genome, and the second column presents the consensus of linearB cell epitopes according to the Similarity Index. This table has been arranged so that each antigen is in decreasing order. For example, we obtained five linear B cell epitopes for the 14 kDa antigen. Of all of them, the epitope KAERTEQKDFDGRSE should be the best selection because it is the one with the highest Similarity Index value (4.53) with respect to the epitope PGVDPDKDV that was 3. It is a way to be able to quantitatively select the results and therefore, we can have a selection criterion between epitope results.

4. Conclusion

To date, many scientific publications analyze one epitope without having developed a criteria to choose a particular epitope. This work allows one to quantitatively select those epitopes derived from several predictors according to the Similarity Index. In fact, the use of the Similarity Index concept allows one to quantifytheresults of theprediction of B cellepitopesobtained from an analysis of the results of several computational predictions.

It is surprising that the genome of *Mycobacterium* yielded only 198 epitopes that were obtained from 30 differentantigens in thestudy. Only two epitopes are the same in two different antigens, NRDRNYT and EGHNY, but all the results obtained in this work are different from each other. Moreover, it is possible to select those that have a high value of the *Similarity Index* which should be the best candidate in order to design an experimental vaccine.

On the other hand, thanks to this methodology, the epitopes with the highest Similarity Index value should be selected. It is impossible to know if antigenically that it is the best candidate for the development of a vaccine. For example, it is impossible to know which of the two epitopes: QDPEG and AAATGNDKTDRDDAN is the best candidate for the development of a vaccine without laboratory confirmation although the values of the Similarity Index are 6.00 and 5.00, respectively.

Table 1. The first column is the antigen and the second column is the consensus epitope and the Similarity Index [12].

Antigen	Epitope, Similarity Index
14 kDaAntigen	KAERTEQKDFDGRSE, 4.53 EQKDFDG, 4.00 EGKPTEK, 3.14 EDEMKEG, 3.00 PGVDPDKDV, 3.00
19 KdaLipoproteinAntigenPrecursor LPQH	DGNPPEVKSVGL, 3.91 GKDQNVTGSV, 3.20 GTGQGN, 3.00 ANPMSPVN, 3.00 TNNDDP, 4.00
3-Oxoacyl-[Aacyl-Carrierprotein] Synthase 2 KASB	PFDRDRD, 4.00 DEGKQSL, 3.57 SPEVDTNR, 3.50 SEAYQGVQQKWD, 3.25 TDAETT, 3.17 DDMRARG, 3.14 GEPRPGN, 3.14 ETEEHA, 3.00 PDPNGE, 3.00
6 kDaEarlySecretory antigenic target	DEGKQSL, 3.57 SEAYQGVQQKWD, 3.25 KEVETKEQ, 4.00 IAYDEEARRG, 3.80 DKPEKEKA, 3.50
60 kDaChaperonin 2 (Protein Cpn60-2)	VTDPERQEAVLEDPY, 3.47 RQEIENSDSDYDREKLQ, 3.18 TDDVA, 3.00 AEDVEGE, 3.00 KDETT, 3.00
Acyl-CoA synthetase	PGQPPE, 4.00 VARRPGESPPS, 3.73 NGYWQKPD, 3.63 IEVDLLDL, 3.63 KKRGDSDQDA, 3.30 GYTFKEDEYPS, 3.19 TDDDMA, 3.00 PSPGTPE, 3.00 QGGVTDERSDS, 3.00 GRNHSP, 3.00 SRKIRAW, 4.29 GPSSDPA, 3.71
Antigen 85-B	GKAGCQTYKW, 3.70 ERNDPTQQI, 3.56 DPSQGMG, 3.43 GGNNSP, 3.00 MPVGGQ, 3.00
CutinasePrecursor CFP21	SDDYRASAS, 4.00 GRSIGV, 3.50 AAATGNDKTDRDDAN, 5.00 DGTATGAADPR, 4.73 PDHSRDQARRN, 4.64
Cyclopropane-fatty-acyl-phospholipid Synthase	TAGADGGGRAHQDTGDQTVATSSSGG AAMTVETSQTPSAA, 4.38 VEQQRLA, 3.57 YRDVDGQY, 3.50 DDNQP = 4.60
DnaK	TAKDKGTGKENTIRI, 3.60 TVDADKNP, 3.50 KFVKEQREA, 3.44 HAEEDRKRREEADVRN, 3.00

Antigen	Epitope, Similarity Index
Esterase LipC	VTNVD = 3.00
	RHMGS = 3.00
	RRAAGSTG, 4.00
	WEDRSTPERA, 4.00
	DPQYQAE, 3.86
	EVRRRDRAS, 3.56
	TIDRH, 3.00
Immunogenic protein MPT63 (Antigen MPT63/MPB63)	VSQFNARTA, 3.78
	PQGEQSTGK, 3.44
	MTDTV, 3.00
	SDLKSS, 3.00
Immunogenic protein MPT64 (Antigen MPT64/MPB64)	ATSSTPREAP, 4.61
	APKTYCEELKGTDTG, 4.33
Integral Membrane rotein	ETRARRQ, 4.00
	DPPRKSRTKASEQELASPYRG, 3.71
	SDHVGSR, 3.00
	DNEGSRY, 3.00
Iron- Regulated HeparinBinding HemaglutininHBHA (Adhesin)	ENSNIDDIKA, 3.90
	ERLRSQQSFEEVSARAEGYV, 3.55
	EETRTDTRSRVE, 3.00
Lipoprotein lppZ	NDAQSQP, 4.00
	KTAGDAEKL, 3.44
	TKNSEVST, 3.25
	FTTEPELRPQPSSTPPPPPP, 3.00
Lipoprotein LpqG	MNQTNDRQQ, 3.89
	QYSNPEPAGT, 3.50
MajorSecreted Immunogenic protein MPT70	AAANP, 3.00
	STIDELKTN, 3.00
	ETDGAEKGPTYNKVRGD, 4.53
Malate Synthase G	NRDRNYT, 4.00
	NRDRNYT, 4.00
	AQQPNG, 4.00
	DKWHRRRVIEPID, 3.38
	AAVDKDGTA, 3.00
	LINSR, 3.00
	GKRRAT, 3.00
	DEIREE, 3.00
	EKPAPSDRAG, 3.00
	QFRGEFT, 4.00
MCE-FamliyAaily protein MCE1A	SARPS, 4.00
	MTTPGKLNKARVPPYK, 3.00
	SATATTG, 3.00
	RQVGDNT, 3.00
	ETGDQ, 4.20
	HPTYEDIVGLERLYAENPSAR, 3.57
	RSEKV, 3.40
	TEYNNKAAL, 3.33
	AKSKGSQQEDEA, 3.00
Mtb48	TEDRA, 3.00
	DNMRE, 3.00
	RANEVE, 3.00
	ATALDNDG, 3.00
	GGDSSAE, 3.00
	DNWEGD, 3.00
	VNPPKPPP, 3.00
	DPPPPPQE, 3.00
	RVGDPDE, 4.14
	TGQRSR, 4.00
	AYGPTRN, 4.00
	GTSERDERDR, 4.00
MycocerosicAcidSynthase	DGTRG, 3.80
	KPHDSGERAN, 3.70
	PWPNGN, 3.67
	IDDDGQEQRAQGACT, 3.67
	SEREATS, 3.57

Antigen	Epitope, Similarity Index
Nicotinamide-nucleotide Adenylyltransferase	RGTATNQDGRTET, 3.54
	DRWDADDYYDPE, 3.33
	DEHTDV, 3.33
	VPRSGRRS, 3.25
	TEEPERH, 3.14
	TVETNRD, 3.00
	LGHEERR, 3.00
	IDDREPQ, 3.00
	TPEQAP, 3.00
	PNPKARQ, 3.00
	AEPDTA, 3.00
	LVRRSP, 3.00
	VPGRSVS, 3.00
	PESTGKTT, 3.13
	ANENPQRP, 3.00
PeriplasmicPhosphate-Binding Lipoprotein PSTS1 (PBP-1) (PSTS1)	QDPEG, 6.00
	GSKPPSGSPET, 3.27
	LHRSDGSGDTFLFTQYL, 3.06
PeriplasmicPhosphate-binding Lipoprotein PSTS2 (PBP-2) (PSTS2)	RSDKSGTSDNFQKYLDGASNG, 4.67
	NPSTGQPDRSAERCGS, 4.06
	STAQENAMEQ, 3.30
	YTLDYNANGSGAG, 3.23
	WNDPQIQALNS, 3.00
	GKGASET, 3.00
	GGTNSSSSG, 3.00
	DLVLDT, 3.00
PeriplasmicPhosphate-Binding Lipoprotein PSTS3 (PBP-3) (PSTS3) (PHOS1)	RSDESGTTDNFQRY, 3.64
	LSKDEAAAAQ, 3.40
	WNNPAIQALNR, 3.00
	QAASNGA, 3.00
	AGKSFQGGVG, 3.00
	ARGNDG, 3.00
	KYPDSQ, 3.00
	NDDNV, 3.00
PPE family protein	ATTGQASA, 3.00
	NYTAN, 3.00
	GAQQTTLS, 3.63
	LPPEI, 3.00
	AELRASALSY, 3.00
	EEWH, 3.00
	AEQAGAQAE, 3.00
	QTTNA, 3.00
	DNGTRRT, 5.00
	ARSNHGTC, 4.13
	SPSRPQSSS, 4.00
	PVDECKDKDMT, 4.00
	STLEKDN, 4.00
	HDDAN, 4.00
	YRMRKF, 4.00
	LTADEEDRH, 3.89
	PPTKES, 3.67
	DARDT, 3.60
	RSGEPF, 3.50
RNA polymerase beta-subunit	RESAAERGDVNP, 3.42
	DSRQSKTAAS, 3.30
	VEYVP, 3.00
	NMQRQ, 3.00
	EGHNY, 3.00
	REPLE, 3.00
	EGHNY, 3.00
	KEKRYD, 3.00
	ERMTTQD, 3.00
	LIGSLS, 3.00
	EDEDE, 3.00
	RKISDGD, 3.00
	REGEDEDLE, 3.00
Secreted L-Alanine	SSAYEL, 4.00

Antigen	Epitope, Similarity Index
Dehydrogenase ALD	HLMRTQGGRG, 3.30
	VQTADGAL, 3.13
	PTETKNN, 3.00
	GEGSA, 3.00
	GSRPTTYD, 3.00
Transcriptional Regulatory protein AraC/XylS-family	AGVRDQDVGNYD, 4.08
	LSRPLNRDDV, 3.4
	PAAYRNQ, 3.0

Conflict of Interest

The authors declare no conflicts of interest in this article.

References

[1] Global tuberculosis report 2018 available at http://www.who.int/tb/publications/global_report/en/

[2] Larsen, E. Lund, O, and Nielsen, M. (2006). Improved method forpredicting linear B-cell epitopes. *Immunome Res*. Vol. 24, p. 2.

[3] E. A. Emini, E. A, Hughes, J. V, Perlow, D. S, Boger, J. J (1985). Induction of hepatitis A virus-neutralizing antibody by avirus-specific synthetic peptide. *Virol.* Vol. 55, pp. 836-839.

[4] Kolaskar A. S, and Tongaonkar P. C (1990). A semi-empirical method for prediction of antigenic determinants on protein antigens. *FEBS Lett.* Vol. 276 (1-2), pp. 172–174.

[5] Saha, S. and Raghava, G. P. S. (2004). Bce Pred: Prediction of continuous B-cell epitopes in antigenic sequences using physico-chemical properties. *Proteins*, Vol. 65, pp. 197-204.

[6] Yao B, Zhang L, Liang S, et al. (2012). SVMTriP: A method to predict antigenic epitopes using support vector machine to integrate tri-peptide similarity and propensity. *PLOSO*. Vol. 7 (9), pp. e45152.

[7] Chen, J, Liu H, Yang J, Chou K. C. (2007). Prediction of linear B-cellepitopes using amino acid pair antigenicity scale. *Amino Acids*. Vol. 33, pp. 423-428.

[8] Isea, R. (2017). Quantitative Prediction of Linear B-cell Epitopes. *Biomedical Statistics and Informatics*. Vol. 2 (1) pp: 1-3.

[9] Isea, R (2015). Predicción computacional cuantitativa de epítopos de células B. *Vaccimonitor*. Vol. 24 (2), pp: 93-97.

[10] Isea, R (2013). Mapeo computacional de epítopos de células B presentes en el virus del dengue. *Revista del Instituto Nacional de Higiene Rafael Rangel*. Vol. 44 (1), pp: 25-29.

[11] Isea, R (2010). Identificación de once candidatos vacunales potenciales contra la malaria por medio de la Bioinformática. *VacciMonitor*. Vol. 19 (3), pp. 15-19.

[12] Isea, R (2018). State of the art for Linear B cell epitope in Plasmodium falciparum according to a bibliographical investigation. *International Journal of Health and Clinical Research*. Vol 1 (1), pp: 1-8.