HAL open science

Quantitative Computational Prediction of the Consensus B-cell Epitopes of 2019-nCoV

Raúl Isea

To cite this version:

Raúl Isea. Quantitative Computational Prediction of the Consensus B-cell Epitopes of 2019-nCoV. Journal of Current Viruses and Treatment Methodologies, 2020, 1 (1), pp.42 - 47. 10.14302/issn.2691-8862.jvat-20-3278 . hal-03519489

HAL Id: hal-03519489

https://hal.science/hal-03519489v1

Submitted on 10 Jan 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

OPEN ACCESS
Freely Available Online

JOURNAL OF CURRENT VIRUSES AND TREATMENT METHODOLOGIES

ISSN NO: 2691-8862

Research

DOI: 10.14302/issn.2691-8862.jvat-20-3278

Quantitative Computational Prediction of the Consensus B-cell Epitopes of 2019-nCoV

Raúl Isea[1,*]

[1]Fundación IDEA, Hoyo de la Puerta, Baruta, Venezuela

Abstract

The goal of this paper is to obtain the numerical consensus of B cell epitopes from the three-dimensional structure of the prefusion spike glycoprotein of the new betacoronavirus that could lead to the development of a vaccine to 2019-nCoV. In order to do that, we first calculated the B-cell epitopes that are predicted using fourteen different mathematical algorithms. Later, we obtained the consensus of B-cell epitopes according to the Similarity Index, and finally selecting the best candidates according to the results of a function called <F> which is evaluated for the glycoprotein. The best candidates that we obtained in order to design a vaccine are SSANNCT, PLQSYGFQPT, TESNKKFLP, NNSYEC, AENS, LPDPSK and YDPLQPE.

Corresponding author: Raúl Isea, Fundación IDEA, Hoyo de la Puerta, Baruta, Venezuela. Email: raul.isea@gmail.com

Keywords: Coronavirus, 2019-nCoV,Covid-19, Epitope, B cell, vaccine

Received: Mar 18, 2020 **Accepted:** Apr 10, 2020 **Published:** Apr 11, 2020

Editor: Qianqian Song, Wake Forest School of Medicine, Wake Forest Baptist Comprehensive Cancer Center, Medical Center Boulevard, Winston-Salem, NC 27157

Introduction

On March 11 a new epidemic of an unusual pneumonia appearing to originate in the Huanan Seafood Market in Wuhan (China) was eventually declared to be a pandemic. More than 150,000 cases of 2019-nCoV infected until March 10, 2020, and almost 6000 deaths in almost 140 countries in the world. Unfortunately there currently is no vaccine to fight this disease.

In January 2020, Chen *et al* obtained the sequence of the virus from a sample obtained from bronchoalveolar lavage fluids from an infected patient in December 2019 [1]. They concluded that it is a new human coronavirus, and called it 2019-nCoV (or Covid-19). Later, the team of Wrapp *et al* determined the three-dimensional structure by X-rays (resolution 3,46 Å) of one of the key proteins for the development of a vaccine against this disease, the spike (S) glycoprotein [2].

With these results, it is possible to determine the B-cell epitopes that could be the basis for developing a vaccine, it means, it is basically residues present on the surface of an antigen that stimulates humoral immune responses [3, 4]. In fact, the goal of epitope prediction is to design the minimal immune unit that invoke strong humal responses in human body [4], Therefore, the goal of this paper is determine *in silico* the consensus of linear and conformational B-cells epitopes obtained from the spike glycoprotein. Since these *in silico* methods obtain a large number of results, structural and energetic properties are considered from a function called <F> as will be explained in the next section.

Function <F>

In 2015, a computational methodology was published that allowed quantifying the quantification of the B cell epitopes employing a function called <F>, which is based on structural and energy factors evaluated in this glycoprotein which includes the degree of exposure to the solvent, the mobility, and the Gibbs free energy. Only the first of the three factors have been previously used in the scientific literature [5]. The second factor is the mobility in order to identify the displacement of the amino acid in the time, and the last factor is the Gibbs free energy because it tells us how likely that an amino acid can mutate.

Therefore, the function <F> will be defined as follows:

$$\langle F \rangle = \langle Q \rangle \cdot \delta\Delta G / \langle R \rangle$$

where <Q> is the average value obtained from the Similarity Index; δΔG is the sum of all the values obtained from the free energy changes, and <R> is the average value of the mobility of the amino acids. In the next section we will explain how these values are calculated.

An important aspect is to verify the degree of solvent exposure of the region from where the epitopes come, and for to calculate it, we average values obtained from two different computer programs. Therefore, the work will calculated the consensus of B-cell epitopes derived from the spike glycoprotein of the 2019-nCoV, filtered according to the values of the function <F>, and the degree of solvent exposure.

Computational Methodology

From the sequence and three dimensional structure of the spike (S) glycoprotein obtained from PDB, we calculated the linear and conformational B-cell epitopes with the following computational programs: BePiPred [6], Emini Surface Accessibility Prediction [7], Kolaskar and Tongaonkar Antigenicity [8], ABCpred [9], BCPred based on flexibility, accessibility, exposed solvents and hydrophobicity [10], ElliPro [11], DiscoTope [12], CBTope [13], SEPPA [14], COBEpro [15] , and SVMTriP [16].

The next step was to determine the consensus of the B-cell epitopes according to the procedure described by Isea *et al* [17-22]. To do this, we employed a script in Python that allows us to calculate the overlap of B-cell epitopes where only those epitopes with a length equal to or greater than four was used (cutoff of 5.0) [23].

In parallel, the contributions of the Gibbs free energy was calculated with the PoPMuSiC program [24]. The values of δΔG will be equal to the sum of the different values obtained from these amino acids where we have assumed that small values represent a low probability that their amino acids can mutate.

On the other hand, the mobility associated with each epitope, abbreviated as <R>, was obtained with

the elNémo program [25]. This value was calculated according to the normal modes of the protein and indicates the gross displacement of the amino acids of the protein.

It is necessary to verify that these epitopes come from a region exposed to the solvent, and for this reason, it was determined from the average value obtained from two different computer programs: PoPMuSiC [24] and Polyview [26]. Logically, these results must be normalized in order to compare these results.

Thus, the best candidates to select a vaccine against this disease should be those with the lowest value of the <F> function and the highest degree of solvent exposure. These last two criteria have been set in the present work and must be verified with experimental results

Results

We have obtained 373 B-cell epitopes according to results obtained from the sequence and the three-dimensional structure of the spike glycoprotein (PDB ID) 6VSB as detailed below: 24 linear B-cell epitopes (the threshold is 0,35) yields with the BePiPred 2.0 program. The ABCPred predicted 49 epitopes of the length of 10 amino acids whose score is equal to or greater than 0,66. The BCPred program based on the hydrophilicity, flexible, accessibility and exposed surface were predicted 15, 14, 30 and 8, respectively. The Emili procedure generated 22 epitopes with a threshold of 1 and a window size of 6; and Kolasar antigenic were 37 (threshold 1,044 and windows size = 7). Discotope 1.1 (threshold = -7,7) yields 19, and ElliPro was 14 (minimum score = 0,5 and maximum distance = 6 Å). CBTope (threshold = -0.3) predicted 31, COBEpro (threshold = 0 .69) was 59, SEPPA 3.0 was 41 and finally SVMTripP was 10 (windows size = 10 and threshold = 0.325).

The next step was to calculate the consensus of B cell epitopes and the function <F>. To visualize this calculation, we let us focus on the region between 276 to 287 amino acids as show in Table 1.

In the first and second columns of Table 1, the position (abbreviated Pos) and their corresponding amino acid sequence of spike glycoprotein are shown. The third column indicates the value of Q, that is, the Similitud Index calculated by Isea *et al*. The consensus B cell epitope will be one whose values of Q are greater than 4 (with a minimum extension of four amino acids). As seen in Table 1, the consensus B cell epitope will be KYNENGT (highlighted in bold for easy viewing). The value <Q> will be simply the average of these Q values, which is (6+5+5+6+6+6+6) / 7 = 5,71.

The value $\delta\Delta G$ is the sum of the Gibbs energy contributions, it means, 0,04+2,51+0,63+0,51+0,81+2,24+0,88=7,62; while the average mobility value (<R>) is equal to (3.22+3.75+4.19+4.94+4.95+4.49+3.99)/7=4,22.

Therefore the value of the function <F> obtained from the glycoprotein and extrapolate to the consensus of B cell epitope KYNENGT is equal to 10,31 (*ie.,* 5,71 ·7,62/4.22).

Table 2 shows the values obtained according to the procedure described above. However, it is important to verify that the regions are exposed to the solvent. For this reason, we calculated the average value that was obtained from two different programs from the data shown in Table 1. From the previous example, the degree of solvent exposure obtained with the PoPMusic program is simply (34,98+2,18+39,07+33,60+48,86+2,49+29,27)/7 =27,21 (see the values in Table 1); while the average value with the Polivew program is 2,29. To compare these values, they must be normalized it, taking into account that the maximum values are 100 and 9 for PoPMusIc and Polyview, respectively. The average value of the degree of exposure to the solvent is 0,26 (*ie.*, 27,21/100+2,29/9). This last result indicates that this epitope (theoretically) is 26% exposed to the solvent. The rest of the results are shown directly in Table 2.

The results in Table 2 have been ordered according to the position of their amino acid sequence in the spike glycoprotein. It is interesting to comment that 12 of the 22 consensus epitopes obtained in this work have a value of <F> less than 6, which would imply that they could be possible candidates for the design of a vaccine. However of these 12, only 7 have an average value of exposure to the solvent greater than 40%. So the best candidates for the development of a vaccine appear to be: SSANNCT, PLQSYGFQPT, TESNKKFLP, NNSYEC, AENS, LPDPSK and YDPLQPE.

Table 1. Region selected in the spike glycoprotein to visualize the procedure for calculating the consensus of B cell epitope and function <F> (see text for more details).

Pos.	Amino Acid	Q	R	Mutation	Polyview	PoPMuSiC solvent
276	L	2	2.73	1.68	0	0.27
277	L	2	2.96	2.33	0	0
278	**K**	**6**	3.22	0.04	3	34.98
279	**Y**	**5**	3.75	2.51	0	2.18
280	**N**	**5**	4.19	0.63	3	39.07
281	**E**	**6**	4.94	0.51	3	33.60
282	**N**	**6**	4.95	0.81	4	48.86
283	**G**	**6**	4.49	2.24	0	2.49
284	**T**	**6**	3.99	0.88	3	29.27
285	I	3	3.39	1.72	0	2.38
286	T	4	3.26	0.75	3	37.94
287	D	4	2.79	0.36	4	39.51

Table 2. Values obtained from the consensus of B-cell epitopes obtained in the present work ordered according to the position of the sequence in the spike glycoprotein. See text for more details.

Position	**Amino acids**	**<Q>**	**<F>**	**<Solv>**
39-46	PDKVFRSS	5,75	9,38	0,23
108-115	TTLDSKTQ	7,13	8,19	0,34
161-167	**SSANNCT**	5,71	3,74	0,46
172-175	SQPF	5,50	3,60	0,29
207-217	HTPINLVRDLP	6,00	15,48	0,37
278-284	KYNENGT	5,71	10,31	0,26
307-326	TVEKGIYQTSNFRVQPTESI	5,40	28,36	0,33
351-362	YAWNRKRISNCV	6,92	4,87	0,33
404-431	GDEVRQIAPGQTGKIADYNYKLPDDFTG	6,79	4,06	0,28
491-500	**PLQSYGFQPT**	6,10	0,67	0,43
517-532	LLHAPATVCGPKKSTN	6,50	4,55	0,36
553-561	**TESNKKFLP**	7,33	4,74	0,45
602-608	TNTSNQV	5,86	9,06	0,48
657-662	**NNSYEC**	5,33	4,77	0,53
701-704	**AENS**	5,50	1,15	0,52
706-709	AYSN	6,00	5,29	0,39
772-779	VEQDKNTQ	6,63	6,57	0,23
786-796	KQIYKTPPIKD	6,82	6,30	0,34
806-815	**LPDPSK**	9,33	5,77	0,51
983-989	RLDPPEA	6,00	2,84	0,26
1035-1039	GQSKR	6,00	15,26	0,17
1138-1144	**YDPLQPE**	5,86	2,59	0,45

Conclusions

The present work calculated the consensus linear and conformational B-cell epitopes from a new coronavirus 2019-nCoV obtained from spike glycoprotein. With this information, it may be possible to design a vaccine for this disease. We employed a function called <F> that considers energy factors and the structure of the glycoprotein for selected the best candidates of B cell epitopes. Finally, the next step is to begin to validate these results with *in vivo* experiments.

Acknowledgment

The author wishes to thank Karl E. Lonngren for his suggestions in this work.

References

1. L. Chen, W. Liu, Q. Zhang, K. Xu, G. Ye, W. Wu, Z. Sun, F. Liu, K. Wu, B. Zhong, Y. Mei, W. Zhang, Y. Chen, Y. Li, M. Shi, K. Lan, Y. Liu (2020). RNA based mNGS approach identifies a novel human coronavirus from two individual pneumonia cases in 2019 Wuhan outbreak. Emerg Microbes Infect. Vol 9 (1):313-319.
2. D. Wrapp, N. Wang, K.S. Corbett, J.A. Goldsmithm, C-L Hsieh, O. Abiola, B.S. Graham, J.S. McLellan (2020). Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation. Science, Vol 2507.
3. J. Sarkander, S. Hojyo, and K. Tokoyoda (2016). Vaccination to gain humoral immune memory. Clin Transl Immunology. Vol 5(12): e120.
4. Y. El-Manzalawy and V. Honavar (2010). Recent advances in B-cell epitope prediction methods. Immunome Res. Vol 6(Suppl 2): S2.
5. J.L. Sanchez-Trincado, M. Gomez-Perosanz, and P.A. Reche (2017). Fundamentals and methods for T- and B-cell epitope prediction. J. Immun Rese. Vol 2017: 2680160.
6. X. Yang andX. Yu (2009). An introduction to epitope prediction method and software. Rev Med Virol. Vol 19(29): 77-96.
7. M.C. Jespersen, B. Peters, M. Nielsen and P. Marcatili (2017). BepiPred-2.0: improving sequence-based B-cell epitope prediction using conformational epitopes. Nucleic Acids Res. Vol. 45 (Web Server issue): W24–W29.
8. E.A. Emini, J.V. Hughes, D.S. Perlow and J. Boger (1985). Induction of hepatitis A virus-neutralizing antibody by a virus-specific synthetic peptide. J Virol. Vol 55(3):836-839.
9. A.S. Kolaskar and P.C. Tongaonkar (1990). A semi-empirical method for prediction of antigenic determinants on protein antigens. FEBS Lett. Vol 276(1-2):172-174.
10. S. Saha and G.P.S Raghava (2006). Prediction of Continuous B-cell Epitopes in an Antigen Using Recurrent Neural Network. Proteins. Vol 65(1): 40-48.
11. S. Saha and G.P.S. Raghava (2004). BcePred: Prediction of Continuous B-Cell Epitopes in Antigenic Sequences Using Physico-chemical Properties. In G.Nicosia, V.Cutello, P.J. Bentley and J.Timis (Eds.) ICARIS, LNCS 3239, 197-204, Springer.
12. J.V. Ponomarenko, H. Bui, W. Li, N. Fusseder, P.E. Bourne, A. Sette A andB. Peters (2008). ElliPro: a new structure-based tool for the prediction of antibody epitopes. BMC Bioinformatics. Vol 9:514
13. P.H. Andersen, M. Nielsen and O. Lund (2006). Prediction of residues in discontinuous B cell epitopes using protein 3D structures. Protein Science. Vol 15: 2558-2567.
14. C. Zhou, Z. Chen, L. Zhang, D. Yan, T. Mao, K. Tang, T. Qiu and Z. Cao (2019). SEPPA 3.0—enhanced spatial epitope prediction enabling glycoprotein antigens. Nucleic Acids Res. Vol 47 (W1): W388–W394.
15. M.J. Sweredoski and P. Baldi (2009). COBEpro: a novel system for predicting continuous B-cell epitopes. Protein Eng Des Sel. Vol 22(3): 113–120.
16. B. Yao, L. Zhang, S.Liang and C.Zhang (2012). SVMTriP: a method topredict antigenic epitopes using support vector machine to integrate tripeptie similarity and propensity. PloS One, Vol 7(9): e45152.
17. R. Isea (2017). Quantitative Prediction of Linear B-cell Epitopes. Biomedical Statistics and Informatics. Vol. 2(1): 1.
18. R. Isea (2013). Predicción de epítopos consensos de células B lineales en Plasmodium falciparum 3D7. VacciMonitor, Vol. 22(1): 43.

19. R. Isea (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): 25-29.
20. R. Isea (2010). Identificación de once candidatos vacunales potenciales contra la malaria por medio de la Bioinformática. VacciMonitor. Vol. 19(3): 15.
21. R. Isea (2013). Designing a peptide-dendrimer for use as a synthetic vaccine against *Plasmodium falciparum*. Am. J. Bioinform Comput Biolm. Vol 1 (1), 1-8
22. R. Isea, R. Mayo-García and S. Restrepo (2016). Reverse Vaccinology in *Plasmodium falciparum 3D7*. Journal of Immunological Techniques in Infectious Diseases. Vol. 5: 3.
23. R. Isea (2015). Predicción computacional cuantitativa de eítopos de células B. VacciMonitor. Vol 24(5): 93-97.
24. G.D Rooman (2000). PoPMuSiC, an algorithm for predicting protein mutant stability changes. Applications to prion proteins. Protein Eng. Vol 13 (12):849-856.
25. V. Frappier and R.J. Najmanovich (2014). A coarse-grained elastic network atom contact model and its use in the simulation of protein dynamics and the prediction of the effect of mutations. PloS Comput Biol. Vol 10(4): e1003569.
26. Polyview-MM: web-based platform for animation and analysis Nucleic acid research. Vol 38: W663-W666.