CIENCIA EN REVOLUCIÓN

Isea y Vega, Vol. 6, 7–13, enero–abril, 2020.

Article:

The consensus B-cell epitopes of SARS coronavirus spike glycoprotein

Raúl Isea[1]*, Cristóbal Vega[2][3]
[1]Fundación Instituto de Estudios Avanzados, Hoyo de la Puerta, Baruta, Venezuela
[2]Laboratorio de Química Computacional, Centro de Química,
Instituto Venezolano de Investigaciones Científicas, Altos del Pipe, Venezuela.
[3]Laboratorio de Procesos Estocásticos, Instituto de Matemáticas y Cálculo Aplicado
Facultad de Ingeniería, Universidad de Carabobo, Valencia, Venezuela.

Recibido: marzo, 2020,
Aceptado: abril, 2020.
Autor para correspondencia: R. Isea e–mail: raul.isea@gmail.com
DOI: https://doi.org/10.5281/zenodo.3930513

Abstract

The goal of this paper is to obtain the consensus B-cell epitopes calculated from the spike glycoprotein of SARS-CoV, after an analysis of 946 epitopes predicted by thirteen different computer programs. To reduce and select the best epitopes, we define a function called $< F >$ that includes energy and structural factors obtained from the changes in the Gibbs free energy, the mobility and the degree of solvent exposure. With this information, it was possible to obtain eight of the twenty-four consensus B-cell epitopes could be useful in the design of a vaccine. These epitopes are PNYTQHT, STMNNKSQSV, SKPMGTQT, DVSEKSGN, KYDENGTIT, PSSKRFQPFQQF, FTDSVRDPKTSE, YVPSQERNFT.
Keywords: Glycoprotein; SARS; Coronavirus; CoV; Vaccine; Epitope; B cell; $< F >$.

Artículo:

El consenso de los epítopos de células B de la glicoproteína espiga del coronavirus del SARS

Resumen

El objetivo del trabajo es determinar el consenso de los epítopos de las células B derivados de la glicoproteína espícular responsable del SARS-CoV, tras analizar los resultados obtenidos de trece programas computacionales diferentes. Se obtuvieron 946 epítopos de células B, y para seleccionar los mejores candidatos entre todos ellos, se definió una función llamada $< F >$ que considera factores de estructura y de energía obtenidos del análisis de esta glicoproteína. Con esta información es posible seleccionar ocho consensos que podrían ser útiles para el diseño de una vacuna, o un método de diagnóstico contra el SARS-CoV, siendo los mismos PNYTQHT, STMNNKSQSV, SKPMGTQT, DVSEKSGN, KYDENGTIT, PSSKRFQPFQQF, FTDSVRDPKTSE, YVPSQERNFT.
Palabras clave: Glicoproteína; SARS; Coronavirus; CoV; Vacuna; Epítopo; Células B; $< F >$.

Isea y Vega, Vol. 6, 7–13, enero–abril, 2020.

1. Introducción

Entre los años 2002 y 2003 hubo en registro un brote de neumonía por un coronavirus llamado SARS-CoV, donde fueron infectadas más de 8.000 personas distribuidas en más de 30 países [1]. El mundo nuevamente afronta otra epidemia desde diciembre 2019 causada por otro tipo de coronavirus llamada 2019-nCoV (o COVID-19) infectando más de 1.000.000 personas distribuidos en más de 130 países diferentes, razón por la cual fue declarada una pandemia el 11 de marzo de 2020 por la Organización Mundial de la Salud.

Sin embargo, no existe una vacuna contra el SARS-CoV, y mucho menos contra el nuevo coronavirus 2019-nCoV, de modo que es necesario desarrollar nuevas metodologías computacionales que nos permitan identificar los epítopos de las células B que jueguen un papel importante para el desarrollo de una vacuna contra este tipo de enfermedad.

Se ha demostrado que la glicoproteína espícular es la proteína blanco necesaria para la selección de epítopos como se ha documentado recientemente en la literatura científica [2, 3]. De hecho, la estructura tridimensional del SARS-CoV fue resuelta recientemente [4], permitiendo reconocer las regiones claves que permitan diseñar estrategias *in silico* para combatir esta enfermedad.

Por ello, el presente trabajo determina cuantitativamente el consenso de los epítopos de las células B obtenidos de la glicoproteína espícular responsable del SARS-CoV, al estar mejor documentada que el 2019-nCoV. Este análisis considera factores de estructura y energía a partir de una función que es denominada $< F >$, como se explicará en la Sección 2.

Para finalizar, se debe garantizar que los consensos provengan de las regiones de la glicoproteína que estén expuestas al solvente, y por ende, se predecirá *in silico* dicha condición al promediar el resultado obtenido de tres programas computacionales diferentes.

2. La función $< F >$

Una función llamada $< F >$ será definida para cuantificar numéricamente el consenso de los epítopos de células B a partir de factores derivados de la energía y la estructura de la glicoproteína, es decir, el grado de exposición al solvente, la movilidad y los cambios de la energía libre de Gibbs. Únicamente el primero de ellos ha sido empleado en el procedimiento para seleccionar los mismos [5], mientras que los otros dos factores se introducen en este trabajo, y por ende, deberán corroborarse con futuros estudios experimentales.

La función $< F >$ será definida por la Ecuación 1

$$< F >=< Q > \cdot \frac{(\delta \Delta G)}{< R >}, \qquad (1)$$

donde $< Q >$ es el valor promedio del Índice de Similitud calculado por la metodología desarrollada por Isea *et al.* [6], $\delta\Delta$G es la suma de todas las contribuciones de la energía libre de Gibbs y $< R >$ es el valor promedio del desplazamiento de los aminoácidos obtenidos del cálculo proveniente de los modos de vibración como se detallará en la Sección 4.

3. Antecedentes

El enfoque de Kazi *et al.* [7] permite la selección de regiones inmunogénicas de los genomas del patógeno. Las regiones ideales podrían desarrollarse como posibles candidatos a vacunas para activar respuestas inmunes protectoras en los huéspedes. En la actualidad, las vacunas basadas en epítopos son conceptos atractivos que se han seguido con éxito para desarrollar vacunas que se dirigen a patógenos que mutan rápidamente. Los autores proporcionan una visión general del progreso actual de la inmunoinformática y sus aplicaciones en el diseño de vacunas, el modelado del sistema inmunológico y la terapéutica.

El objetivo de Ebrahimi *et al.* [8] fue identificar los antígenos más conservados e inmunogénicos de S. pyogenes mediante ABCpred, que puedan ser candidatos potenciales para el diseño de vacunas en el futuro. Analizaron ocho proteínas de superficie importantes. Mediante diferentes servidores de predicción seleccionaron los epítopos más fuertes que tenían la capacidad de estimular el sistema inmune humoral y celular.

Manavalan *et al* [9] afirman que la identificación de epítopos de células B es un paso fundamental

Isea y Vega, Vol. 6, 7–13, enero–abril, 2020.

para el desarrollo de vacunas basadas en epítopos, la producción de anticuerpos y la prevención y diagnóstico de enfermedades. Luego, es esencial desarrollar un método computacional automatizado para permitir la identificación rápida y precisa de nuevos epítopos de células B dentro de un gran número de proteínas y péptidos candidatos. Buscan un método confiable mediante EPMLR.

4. Metodología computacional

A partir de la secuencia y la estructura tridimensional de la glicoproteína espícular depositada en la base de datos de proteína (www.pdb.org), se determinaron todos los epítopos lineales y conformacionales de células B obtenidos con los siguientes programas computacionales: BePiPred [10], Emini Surface Accessibility Prediction [11], Kolaskar and Tongaonkar Antigenicity [12], ABCpred [13], EPMLR [14], BCPred based on flexibility, accessibility and hydrophobicity [15], ElliPro [16], DiscoTope [17], SEPPA [18], COBEpro [19] y EPCSE [20].

El próximo paso será calcular el consenso de los epítopos de células B de acuerdo a la metodología desarrollada por Isea *et at.* [6], [21]–[26], a partir de los valores Q tras desarrollar un algoritmo escrito en Python para determinar el solapamiento de los epítopos de células B, cuya longitud sea superior a cuatro aminoácidos y un valor de cutoff igual a 5.

En paralelo, fueron determinados los cambios de energía libre de Gibbs (abreviado como $\delta\Delta G$) con ayuda de los resultados del programa PoPMuSiC [27], mientras que el desplazamiento promedio ($< R >$) fue evaluado con el algoritmo implementado en elNémo [28].

El valor promedio de acceso al solvente será predicho de acuerdo al valor promedio de los resultados de tres programas computacionales diferentes: UPSAR [29], PoPMuSiC [27]y Polview [30]. Estos resultados fueron normalizados para poder comparar los resultados entre sí.

Finalmente, se van a normalizar los resultados obtenidos de la función $< F >$, identificando los mejores candidatos para una vacuna (o un método de diagnóstico) siempre y cuando el valor de $< F >$ sea superior a 0.2, y además presenten un valor promedio de exposición al solvente superior al 40 %.

5. Resultados

El identificador en la base de datos de proteínas de la glicoproteína espícular del SARS-CoV es 6ACC, donde centramos todos los cálculos en la cadena A, en la región comprendida entre los aminoácidos 28 y 1190, obteniendo una longitud total de 1162 aminoácidos.

A partir de la secuencia y la estructura tridimensional de esta glicoproteína, se determinaron 946 epítopos de células B obtenidos con trece programas computacionales diferentes (omitiendo los detalles por el gran número de resultados). En paralelo, se calculo el desplazamiento promedio de los distintos aminoácidos con el programa elNémo, así como los cambios en la energía libre de Gibbs con ayuda del programa PoPMuSiC. Con esta información fue posible determinar numéricamente el consenso de los epítopos de células B a partir de la función $< F >$.

A modo de ejemplo, se muestra una pequeña región comprendida entre los aminoácidos 141 y 148 para visualizar el cálculo de la función $< F >$, donde los datos están indicados en la Tabla 1.

Se muestran en la primera y segunda columna de la Tabla 1, la posición y el aminoácido correspondientes de la glicoproteína de la región comprendida entre 141 y 148 aa. En la tercera columna se indican los valores obtenidos del Índice de Similitud (representado con la letra Q) calculados de acuerdo al procedimiento descrito por Isea et al [6]. Con esta información fue posible calcular el valor promedio, $< Q >$, es decir $(6+6+8+8+6+6+7+6)/8 = 6.625$.

Con los resultados mostrados en la cuarta columna, se determinó la contribución correspondiente a la energía libre de Gibbs ($\delta\Delta G$), el cual consiste en sumar los distintos valores obtenidos en la glicoproteína, y posteriormente extrapolado al epítopo correspondiente. El valor correspondiente a la movilidad promedio de los aminoácidos presentes en el consenso ($< R >$) se obtendrá al promediar todas las contribuciones correspondiente a la glicoproteína, y extrapolando sus resultados al epítopo correspondiente.

Tabla 1: Resultados obtenidos del Índice de Similitud (representados con la letra Q), los cambios en la energía libre de Gibbs (ΔG), el desplazamiento de los aminoácidos etiquetados con la letra R, y los valores predichos del grado de exposición al solvente evaluados con los programas UPSAR, PoPMuSiC y Polyview.

Pos.	Aminoácido	Q	ΔG	R (elNémo)	UPSAR	PoPMuSiC (Solv)	Polyview
141	S	6	0	1.106	29.86	27.64	3
142	K	6	-2.12	1.147	62.23	56.75	6
143	P	8	-4.13	1.342	75.41	72.25	7
144	M	8	-3.14	1.298	70.96	70.13	7
145	G	6	0	1.098	15.25	13.07	1
146	T	6	0	1.129	61.28	55.18	6
147	Q	7	-1.99	0.996	17.36	15.66	1
148	T	6	-1.73	1.125	89.4	80.96	8

De modo que el valor de la función $< F >$ para el epítopo consenso SKPMGTQT es simplemente:

$$\begin{aligned} < F > |_{\text{SKPMGTQT}} &= 6.625 \cdot (-13.11)/1.155 \\ &= -75.198. \end{aligned}$$

Es importante determinar sí este epítopo proviene de una región de la glicoproteína que está expuesta al solvente. Para ello, se determina el grado de exposición al solvente al promediar los resultados obtenidos con tres programas diferentes, es decir, el valor promedio obtenido con el programa UPSAR es 52.719 (Tabla 1); mientras que los resultados obtenidos con los programas PoPMuSic y Polyview fueron 48.955 y 4.875, respectivamente. Es necesario normalizar estos resultados para poder compararlos entre sí, y tras realizar dicho cálculo se obtienen 0.96, 0.95 y 0.93 correspondientes a los programas UPSAR, PoPMuSiC y Polyview, respectivamente. Finalmente, el grado de exposición al solvente para este epítopo consenso (SKPMGTQT) es igual a (0.96+0.95+0.93)/3= 0.95, es decir, se predice que está 95 % expuesto al solvente. Todos los resultados se muestran en la última columna de la Tabla 2.

Es interesante señalar que el segundo epítopo consenso mostrado en la Tabla 2 (ie., AATEKSN) cumple la condición con respecto al valor de $< F >$, pero el grado de exposición al solvente es tan solo del 6 %, y por ende, debería descartarse sí se está diseñando una vacuna. Sin embargo, el siguiente epítopo (STMNNKSQSV) cumple con las dos condiciones fijadas en el trabajo, es decir, un valor de $< F >$ superior a 0.2, y un valor promedio de exposición al solvente superior al 40 %. Solo resta indicar se resaltaron en negritas en la Tabla 2, los distintos consensos de los epítopos de células B que pueden ser candidatos para diseñar una vacuna contra esta enfermedad.

6. Conclusiones

El presente trabajo determino los diferentes consensos de los epítopos de células B derivados de la glicoproteína espícular responsable del SARS-CoV. Los mismos fueron caracterizados teóricamente de acuerdos a factores energéticos y estructurales permitiendo seleccionar mejores candidatos para el diseño de una vacuna o un nuevo método de diagnóstico.

El próximo paso será evaluar experimentalmente la calidad de dichos resultados que han sido predichos teóricamente en este trabajo, y poder verificar que los mejores candidatos cumplan con las siguientes condiciones: un valor de $< F >$ superior a 0.2, así como un grado de exposición al solvente superior al 40 %.

Finalmente, una vez validado dicho procedimiento teórico, se podrán identificar los epítopos consensos del nuevo coronavirus 2019-nCoV que está afrontando el mundo entero.

Isea y Vega, Vol. 6, 7–13, enero–abril, 2020.

Tabla 2: Los consensos de los epítopos de células B ordenados de acuerdo a la posición en la glicoproteína espícular del SARS-CoV (ver el texto para más detalles)

No	Epítopo consensus	Posición	$< Q >$	$< R >$	$\delta\Delta G$	$< F >$	$< Solv >$
1	**PNYTQHT**	28-34	5.43	0.35	-3.39	0.28	0.87
2	AATEKSN	90-96	7.14	0.62	-4.37	0.27	0.06
3	**STMNNKSQSV**	105-114	6.20	0.59	-8.62	0.49	0.45
4	**SKPMGTQT**	141-148	6.63	1.16	-13.11	0.41	0.95
5	**DVSEKSGN**	171-178	8.38	0.97	-5.01	0.23	0.87
6	KGYQPIDVVRD	198-208	6.36	0.66	-19.18	1.00	0.02
7	**KYDENGTIT**	265-273	7.44	0.89	-10.55	0.48	0.47
8	AVDCSQ	275-280	5.50	0.44	-0.62	0.04	0.18
9	AWERKKISNC	339-348	5.10	1.29	-5.69	0.12	0.46
10	YKLPD	410-414	5.20	1.73	-2.59	0.04	0.39
11	TRNIDATSTGNYN	425-437	7.00	2.68	-7.95	0.11	0.84
12	YLRHGKLRPFERDISNVPFSPDGKPCTPPA	442-471	6.07	3.36	-16.14	0.16	0.67
13	GIGY	488-491	5.50	2.60	-1.71	0.02	0.88
14	**PSSKRFQPFQQF**	540-551	8.00	1.17	-8.72	0.32	0.65
15	**FTDSVRDPKTSE**	558-569	7.67	0.84	-13.17	0.66	0.49
16	IAYSN	687-691	5.60	1.97	-1.62	0.02	0.89
17	EQDRNTR	755-761	5.29	2.01	-9.47	0.13	0.37
18	QVKQ	766-769	5.25	2.93	-1.42	0.01	0.72
19	PDPLKPTKR	789-797	7.56	2.64	-10.31	0.16	1.00
20	YENQKQI	899-905	5.86	1.52	-5.68	0.12	0.57
21	GQSKR	1017-1021	6.00	1.04	-2.90	0.09	0.19
22	**YVPSQERNFT**	1049-1058	7.10	0.87	-11.05	0.49	0.46
23	HEGKAYFPRE	1065-1074	6.30	2.01	-5.71	0.10	0.34
24	FVSGNC	1103-1108	5.00	2.70	-0.10	0.00	0.45

Referencias

[1] P.A. Rota, M.S. Oberste, S. S. Monroe, W. A. Nix, and E. Campganoli. Characterization of a novel coronavirus associated with severe acute respiratory syndrome. *Science*, 300: 1394–1399, 2003.

[2] F. Li. Structure, Function, and Evolution of Coronavirus Spike Proteins. *Annual Review of Virology*. 3(1): 237–261, 2016.

[3] J.L. Nieto-Torres, M.L. DeDiego, C. Verdiá-Báguena, J.M. Jimenez-Guardeño, J.A. Regla-Nava, R. Fernandez-Delgado, C. Castaño-Rodriguez, A. Alcaraz, J. Torres, V.M. Aguilella, L. Enjuanes. Severe Acute Respiratory Syndrome Coronavirus Envelope Protein Ion Channel Activity Promotes Virus Fitness and Pathogenesis. *PLoS Pathogens*, 10(5): e1004077, 2014.

[4] W. Song, M. Gui, X. Wang, Y. Xiang. Cryo-EM structure of the SARS coronavirus spike glycoprotein in complex with its host cell receptor ACE2. *PLoS Pathogens*. 14: 07236, 2018.

[5] J. Sarkander, S. Hojyo, K. Tokoyoda. Vaccination to gain humoral immune memory. *Clinical & Translational Immunology*. 5(12): e120, 2016.

[6] R. Isea, R. Mayo-García, S. Restrepo. Reverse Vaccinology in Plasmodium falciparum 3D7. *Journal of Immunological Techniques in Infectious Diseases*. 5: 3, 2016.

[7] A. Kazi, C. Chuah, A.B. Abdul M., C. Herng L., B. Huat L., and C. Yee L. Current progress of immunoinformatics approach harnessed for cellular- and antibody-dependent vaccine design, *Pathogens and Global Health*. 112(3): 123-131, 2018.

[8] S. Ebrahimi, H. Mohabatkar, and M. Behbahani, Predicting Promiscuous T Cell Epitopes for Designing a Vaccine Against Streptococcus pyogenes. Applied Biochemistry Biotechnoly 187: 90–100, 2019.

[9] B. Manavalan, RG. Govindaraj, TH. Shin, MO. Kim and G. Lee, iBCE-EL: A New Ensemble Learning Framework for Improved Linear B-Cell Epitope Prediction. *Frontiers Immunology*. 9: 1695, 2018.

[10] M.C. Jespersen, B. Peters, M. Nielsen, P. Marcatili. BepiPred-2.0: improving sequence-based B-cell epitope prediction using conformational epitopes. *Nucleic Acids Res*. 45(Web Server issue): W24–W29, 2017.

[11] E.A. Emini, J.V. Hughes, D.S. Perlow, J. Boger. Induction of hepatitis A virus-neutralizing antibody by a virus-specific synthetic peptide. *Journal of Virology*. 55(3):836-839, 1985.

[12] A.S. Kolaskar, P.C. Tongaonkar. A semi-empirical method for prediction of antigenic determinants on protein antigens. *FEBS Letters*. 276(1-2):172-174, 1990.

[13] S. Saha, G.P.S Raghava. Prediction of Continuous B-cell Epitopes in an Antigen Using Recurrent Neural Network. *Proteins*. 65(1): 40-48, 2006.

[14] S. Saha, G.P.S. Raghava. *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, 2004.

[15] J.V. Ponomarenko, H. Bui, W. Li, N. Fusseder, P.E. Bourne, A. Sette A, B. Peters. ElliPro: a new structure-based tool for the prediction of antibody epitopes. *BMC Bioinformatics*. 9:514, 2008

[16] P.H. Andersen, M. Nielsen, O. Lund. Prediction of residues in discontinuous B cell epitopes using protein 3D structures. *Protein Science*. 15: 2558-2567, 2006.

[17] C. Zhou, Z. Chen, L. Zhang, D. Yan, T. Mao, K. Tang, T. Qiu, Z. Cao. SEPPA 3.0—enhanced spatial epitope prediction enabling glycoprotein antigens. *Nucleic Acids Research*. 47(W1): W388–W394, 2019.

[18] M.J. Sweredoski, P. Baldi. COBEpro: a novel system for predicting continuous B-cell epitopes. *Protein Engineering Design and Selection*. 22(3): 113–120, 2009.

[19] B. Yao, L. Zhang, S.Liang, C.Zhang. SVMTriP: a method topredict antigenic epitopes using support vector machine to integrate tripeptie similarity and propensity. *PloS One*, 7(9): e45152, 2012.

[20] S. Liang, D. Zheng, D.M. Standley, B. Yao, M. Zacharias, C. Zhang. EPSVR and EPMeta: prediction of antigenic epitopes using support vector regression and multiple server results. *BMC Bioinformatics*, 11: 381, 2010.

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

[22] R. Isea. Predicción de epítopos consensos de células B lineales en Plasmodium falciparum 3D7. *VacciMonitor*, 22(1): 43-49, 2013.

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

[24] R. Isea. Identificación de once candidatos vacunales potenciales contra la malaria por medio de la Bioinformática. *VacciMonitor*. 19(3): 15-18, 2010.

[25] R. Isea. Designing a peptide-dendrimer for use as a synthetic vaccine against Plasmodium falciparum. *American Journal of Bioinformatics and Computational Biology* . 1(1): 1-8, 2013.

[26] R. Isea. Predicción computacional cuantitativa de epítopos de células B. *VacciMonitor*. 24(5): 93-97, 2015.

[27] G.D Rooman. PoPMuSiC, an algorithm for predicting protein mutant stability changes.

Applications to prion proteins. *Protein Engineering*. 13(12):849-856, 2000.

[28] V. Frappier, R.J. Najmanovich. 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 Computational Biology*. 10(4): e1003569, 2014.

[29] R. Murphy, J. Casper, J. Hyams, M. Micire, B. Minten, Mobility and sensing demands in USAR, 2000 26th *Annual Conference of the IEEE Industrial Electronics Society. IECON 2000. 2000 IEEE International Conference on Industrial Electronics, Control and Instrumentation. 21st Century Technologies*, Nagoya, Japan, pp. 138-142 vol.1, 2000.

[30] A. Porollo, R. Adamczak, J. Meller. POLYVIEW: A Flexible Visualization Tool for Structural and Functional Annotations of Proteins. *Bioinformatics*, 20: 2460-2462, 2004.