

## Predicción de la inmunogenicidad de la proteína del SARS-CoV-2 responsable de la infección COVID-19 en humanos

### Immunogenicity prediction of the SARS-CoV-2 protein responsible of COVID-19 infection in humans

Orlando Rafael Serrano-Barrera<sup>1,2</sup> 

<sup>1</sup>Hospital General Docente "Dr. Ernesto Guevara de la Serna", Las Tunas. <sup>2</sup>Universidad de Ciencias Médicas de Las Tunas. Facultad de Ciencias Médicas "Dr. Zoilo Enrique Marinello Vidaurreta". Las Tunas, Cuba. **Correspondencia a:** Orlando Rafael Serrano-Barrera, correo electrónico: orlandosb@infomed.sld.cu

Recibido: 28 de abril de 2020

Aprobado: 5 de mayo de 2020

#### RESUMEN

**Fundamento:** las características estructurales y funcionales de la glicoproteína S del virus SARS-CoV-2 la convierten en un antígeno relevante y una diana deseable para el desarrollo de vacunas y fármacos, pues su bloqueo o neutralización impediría el ciclo infeccioso.

**Objetivo:** modelar, con el empleo de herramientas bioinformáticas, la inmunogenicidad de la glicoproteína S, a partir de la identificación de péptidos que puedan ser reconocidos por linfocitos T y B humanos.

**Métodos:** se alinearon ocho secuencias diferentes de la glicoproteína S por medio de CLUSTAL O, para determinar la conservación del dominio de unión al receptor. Se seleccionaron las moléculas HLA-A\*0201, HLA-DRB1\*0301 y HLA-DRB1\*0701 para la predicción de péptidos T, con los algoritmos SYFPEITHI, IEDB y NetMHC/NetMHCII. En el caso de los epítopes B lineales se utilizaron los algoritmos BepiPred-2.0, ABCpred, BcePred y BepiPred-IEDB.

**Resultados:** las secuencias empleadas para comparar el grado de conservación de la estructura de la molécula, particularmente el dominio de unión al receptor, mostraron una homología completa. La región comprendida entre los residuos 207-222 contiene epítopes B y T CD4+, al tiempo que en las posiciones 133-147 y 407-425 se concentran epítopes B y T CD8+; todas están ubicadas en el segmento externo de la proteína. En el motivo de unión al receptor se localizan varios epítopes B y residuos de interacción directa entre la glicoproteína S y ACE2.

**Conclusiones:** la glicoproteína S del SARS-CoV-2 tiene un potencial inmunogénico, con el que puede generar respuesta T y B.

**Palabras clave:** EPÍTOPES T; EPÍTOPES B; INMUNOGENICIDAD; SARS; SARS-COV-2; COVID-19; GLICOPROTEÍNA S; VACUNAS.

**Descriptor:** INMUNOGENICIDAD VACUNAL; SÍNDROME RESPIRATORIO AGUDO GRAVE; VIRUS DEL SRAS; INFECCIONES POR CORONAVIRUS; SIALILTRANSFERASAS; VACUNAS.

#### ABSTRACT

**Background:** the structural and functional characteristics of the SARS-CoV-2 S glycoprotein make it a relevant antigen, and a desirable target for vaccine and drug development, since its blockade or neutralization would hinder the whole infectious cycle.

**Objective:** to model, with the aid of bioinformatic tools, the immunogenicity of S glycoprotein, by identifying the peptides that could be recognized by human T and B lymphocytes.

**Methods:** 8 different sequences of S glycoprotein were aligned by using CLUSTAL O, in order to assess the conservation of the receptor-binding domain. HLA-A\* 0201, HLA-DRB1\* 0301 and HLA-DRB1\* 0701 alleles were selected for T-cell epitope prediction, with the use of SYFPEITHI, IEDB and NetMHC/NetMHCII algorithms. In the case of linear B-cell epitopes, BepiPred-2.0, ABCpred, BcePred and BepiPred/IEDB were the chosen tools.

**Results:** a complete homology was found among the sequences used to compare the degree of conservation of the molecular structure, particularly in the receptor-binding domain. The region spanning from residues 207 to 222 contains B-cell and CD4+ T-cell epitopes, while within the positions 133-147 and 407-425, B-cell and CD8+ T-cell epitopes were abundant; all such regions are located in the external domain of the protein. In the receptor-binding motif, there are several B-cell epitopes and residues of direct interaction between S glycoprotein and ACE2.



Citar como: Serrano-Barrera OR. Predicción de la inmunogenicidad de la proteína del SARS-CoV-2 responsable de la infección COVID-19 en humanos. Revista Electrónica Dr. Zoilo E. Marinello Vidaurreta. 2020; 45(3). Disponible en: <http://revzoiilomarinellosld.cu/index.php/zmv/article/view/2270>

**Conclusions:** the SARS-CoV-2 S glycoprotein has an immunogenic potential with which T and B responses can be generated.

**Keywords:** T-CELL EPITOPES; B-CELL EPITOPES; IMMUNOGENICITY; SARS; SARS-COV-2; COVID-19; S GLYCOPROTEIN; VACCINES.

**Descriptors:** IMMUNOGENICITY, VACCINE; SEVERE ACUTE RESPIRATORY SYNDROME; SARS VIRUS; CORONAVIRUS INFECTIONS; SIALYLTRANSFERASES; VACCINES.

## INTRODUCCIÓN

La familia de los coronavirus, a la cual pertenece SARS-CoV-2, agente causal de la actual pandemia de COVID-19, tiene entre sus características la presencia de una proteína de superficie, llamada espiga (spike, en inglés), que protruye de la envoltura para dar, al microscopio electrónico, el peculiar aspecto por el cual han sido llamados. <sup>(1)</sup> La glicoproteína S, como también se le denomina, es el elemento estructural clave para los dos primeros eventos del ciclo de la infección: la unión a su receptor, la molécula ACE2 humana, y la fusión con la membrana celular para iniciar la penetración. <sup>(1,2)</sup>

La secuencia de la glicoproteína S se organiza en tres regiones: extracelular (del aminoácido 13 al 1213), transmembrana (1214-1234) y un pequeño dominio intracelular (1235-1273). <sup>(3)</sup> El dominio externo, a su vez, se compone de dos unidades, S1, que contiene el sitio de unión al receptor (en inglés receptor-binding domain, RBD) y S2, que media la fusión de las membranas viral y de la célula diana. <sup>(1,2,4-6)</sup> Dentro del RBD se ha descrito un motivo de unión al receptor (receptor-binding motif, RBM), que es la secuencia de interacción directa con ACE2, ubicada entre los residuos 437 y 507. <sup>(7)</sup> Esta proteína es, por tanto, la que determina cuáles son los hospederos que puede infectar SARS-CoV-2 y las células susceptibles en ellos.

Las características estructurales y funcionales de la glicoproteína S la convierten en un antígeno relevante y una diana deseable para el desarrollo de vacunas o fármacos, pues su bloqueo o neutralización impediría el ciclo viral. <sup>(1,5)</sup> Por el escaso tiempo transcurrido desde la emergencia de la pandemia de COVID-19, queda mucho por conocer en términos de la respuesta inmune humana frente a SARS-CoV-2, lo que incluye dos aspectos críticos: la inmunogenicidad de los componentes de la partícula viral, y los mecanismos más efectivos en el control y eliminación de la infección; también, las tasas de mutación que lleven a la aparición de cuasiespecies o variantes de escape. <sup>(8)</sup> El conocimiento obtenido de otros coronavirus, entre ellos los de potencial epidémico como SARS y MERS, puede ser de utilidad, por las similitudes estructurales y antigénicas: se ha reportado una identidad del 76 % entre las secuencias de la glicoproteína S de SARS y SARS-CoV-2, que llega al 90 % en el caso de la región S2. <sup>(1,9)</sup>

Aunque otras proteínas y varios coronavirus han sido objeto de estudio en busca de candidatos vacunales, <sup>(10)</sup> un número significativo de los esfuerzos se ha dirigido a la glicoproteína S, como se discutirá más

adelante en este reporte, y el SARS-CoV-2 es el centro actual de las investigaciones, debido a su rápida expansión y letalidad creciente. En ese sentido, las herramientas bioinformáticas pueden agilizar la identificación de péptidos con potencial inmunogénico, con lo que se minimiza el riesgo biológico y se reduce el tiempo que demanda todo el proceso desde la experimentación preclínica hasta la evaluación clínica. <sup>(11)</sup>

En este reporte se describe el análisis de los posibles epítopes T y B de la glicoproteína S de SARS-CoV-2, para el que se incluyen moléculas de histocompatibilidad reportadas como frecuentes en Cuba, para ajustarlo en alguna medida a la inmunología de la población cubana. También se consideró la ubicación de los epítopes identificados en relación con los sitios críticos de la molécula objeto de estudio: el RBM, los aminoácidos que interactúan directamente con la molécula receptora en el humano, ACE2, así como el péptido de fusión (FP, en inglés fusion peptide).

## MATERIALES Y MÉTODOS

En la base de datos de proteínas del Centro Nacional de Información Biotecnológica de Estados Unidos (NCBI, por su sigla en inglés, <https://www.ncbi.nlm.nih.gov/protein/>), se buscaron las secuencias disponibles de la glicoproteína S de SARS-CoV-2, empleando como término de búsqueda "Spike glycoprotein SARS-CoV-2". De las 20 entradas recuperadas, que incluían la proteína completa o el RBM, se seleccionaron 8, tomando como criterios que fueran reportadas por autores diferentes, en fechas distintas. Al menos dos de las secuencias empleadas en la alineación de secuencias de este reporte, 6M17 y 6VSB, pertenecen a estructuras cuyas conformaciones tridimensionales han sido dilucidadas y reportadas en el Protein DataBank. <sup>(7)</sup> Se alinearon por medio de CLUSTAL Omega (1.2.4), disponible en <http://www.ebi.ac.uk/Tools/msa/clustalo/>; para representar la alineación múltiple, las secuencias se editaron para mostrar solo la región del RBM.

Para la predicción de la inmunogenicidad se seleccionó la secuencia de referencia para la glicoproteína S de SARS-CoV-2 en el NCBI (YP\_009724390.1, disponible en [https://www.ncbi.nlm.nih.gov/protein/YP\\_009724390.1](https://www.ncbi.nlm.nih.gov/protein/YP_009724390.1)); también se obtuvo información estructural y funcional adicional en la entrada correspondiente a la misma molécula en Uniprot. <sup>(3)</sup>

Para la modelación de la presentación de epítopes T, se seleccionaron tres moléculas de MHC humano,

reportadas entre las más frecuentes en Cuba: <sup>(12,13)</sup> la molécula de clase I HLA-A\*0201 para la presentación antigénica a células T CD8+, y los alelos HLA-DRB1\*0301 y HLA-DRB1\*0701, de clase II para linfocitos T CD4+. Se fijó en 9 y 14 el número de aminoácidos, como longitud de los péptidos a obtener, para MHC-I y -II, respectivamente. Se emplearon tres algoritmos diferentes para la predicción de péptidos T: SYFPEITHI (<http://www.syfpeithi.de/>), IEDB (<http://www.iedb.org/>) y NetMHC (<http://www.cbs.dtu.dk/services/NetMHC-4.0/>) o NetMHCII (<http://www.cbs.dtu.dk/services/NetMHCII/>). Se seleccionaron los 10 péptidos de más alta puntuación en cada caso y se muestra la posición de cada uno de ellos en la secuencia primaria de la proteína nativa.

En la predicción de epítopes B lineales se emplearon cuatro algoritmos: BepiPred-2.0 (<http://www.cbs.dtu.dk/services/BepiPred-2.0/>), ABCpred (<https://webs.iitd.edu.in/raghava/abcpred/index.html>), BcePred (<https://webs.iitd.edu.in/raghava/bcepred/>) y BepiPred-IEDB (<http://tools.iedb.org/bcell/>). Cada

uno de los epítopes B predichos por cada herramienta se representó sobre la secuencia primaria de la proteína.

En todos los casos, los aminoácidos son representados por el código de una letra. Las posiciones de los péptidos se fijaron a partir de su localización en la secuencia de referencia YP\_009724390.1.

**RESULTADOS**

En la **imagen 1** se observa la total similitud entre las secuencias alineadas de las ocho moléculas de glicoproteína S, o sus fragmentos, seleccionados en el presente estudio, particularmente en la región correspondiente al RBD, sombreado en color amarillo. Las coincidencias se extienden más allá de la zona sombreada, pero se establecieron límites de acuerdo con la longitud de los dos fragmentos más pequeños de la proteína (6W41 y 6M17). Se destacaron en color azul los residuos que interactúan directamente con ACE2, de acuerdo con reportes previos. <sup>(1,8)</sup>

**IMAGEN 1. Alineación de secuencias de glicoproteína S de SARS-CoV-2, editada para mostrar solo el dominio de unión al receptor humano**

6VYB A	NGTITDAVDCALDPLSETKCTLKSFTVEKGIYQTSNFRVQPTESIVRFPNITNLCPPFGEV	360
6VSB A	NGTITDAVDCALDPLSETKCTLKSFTVEKGIYQTSNFRVQPTESIVRFPNITNLCPPFGEV	341
6W41 C	-----RVQPTESIVRFPNITNLCPPFGEV	23
QIA20044.1	NGTITDAVDCALDPLSETKCTLKSFTVEKGIYQTSNFRVQPTESIVRFPNITNLCPPFGEV	341
YP_009724390.1	NGTITDAVDCALDPLSETKCTLKSFTVEKGIYQTSNFRVQPTESIVRFPNITNLCPPFGEV	341
6M17 E	-----RVQPTESIVRFPNITNLCPPFGEV	23
QHR63250.2	NGTITDAVDCALDPLSETKCTLKSFTVEKGIYQTSNFRVQPTESIVRFPNITNLCPPFGEV	341
QIC53213.1	NGTITDAVDCALDPLSETKCTLKSFTVEKGIYQTSNFRVQPTESIVRFPNITNLCPPFGEV	341
6VYB A	FNATRFASVYAWNRKRISNCVADYSLVLYNSASFSTFKCYGVSPTKLNDLCFTNYYADSFV	420
6VSB A	FNATRFASVYAWNRKRISNCVADYSLVLYNSASFSTFKCYGVSPTKLNDLCFTNYYADSFV	401
6W41 C	FNATRFASVYAWNRKRISNCVADYSLVLYNSASFSTFKCYGVSPTKLNDLCFTNYYADSFV	83
QIA20044.1	FNATRFASVYAWNRKRISNCVADYSLVLYNSASFSTFKCYGVSPTKLNDLCFTNYYADSFV	401
YP_009724390.1	FNATRFASVYAWNRKRISNCVADYSLVLYNSASFSTFKCYGVSPTKLNDLCFTNYYADSFV	401
6M17 E	FNATRFASVYAWNRKRISNCVADYSLVLYNSASFSTFKCYGVSPTKLNDLCFTNYYADSFV	83
QHR63250.2	FNATRFASVYAWNRKRISNCVADYSLVLYNSASFSTFKCYGVSPTKLNDLCFTNYYADSFV	401
QIC53213.1	FNATRFASVYAWNRKRISNCVADYSLVLYNSASFSTFKCYGVSPTKLNDLCFTNYYADSFV	401
6VYB A	IRGDEVQRQIAPGQTGKIADYNYKLPDDFTGCVIAWNSNNLDSKVGGNYYNLYRLFRKSNL	480
6VSB A	IRGDEVQRQIAPGQTGKIADYNYKLPDDFTGCVIAWNSNNLDSKVGGNYYNLYRLFRKSNL	461
6W41 C	IRGDEVQRQIAPGQTGKIADYNYKLPDDFTGCVIAWNSNNLDSKVGGNYYNLYRLFRKSNL	143
QIA20044.1	IRGDEVQRQIAPGQTGKIADYNYKLPDDFTGCVIAWNSNNLDSKVGGNYYNLYRLFRKSNL	461
YP_009724390.1	IRGDEVQRQIAPGQTGKIADYNYKLPDDFTGCVIAWNSNNLDSKVGGNYYNLYRLFRKSNL	461
6M17 E	IRGDEVQRQIAPGQTGKIADYNYKLPDDFTGCVIAWNSNNLDSKVGGNYYNLYRLFRKSNL	143
QHR63250.2	IRGDEVQRQIAPGQTGKIADYNYKLPDDFTGCVIAWNSNNLDSKVGGNYYNLYRLFRKSNL	461
QIC53213.1	IRGDEVQRQIAPGQTGKIADYNYKLPDDFTGCVIAWNSNNLDSKVGGNYYNLYRLFRKSNL	461
6VYB A	KPFFERDISTEIQAGSTPCNGVEGFNCYFPLQSYGFQPTNIGVGYQPYRVVVLFSPELLHAP	540
6VSB A	KPFFERDISTEIQAGSTPCNGVEGFNCYFPLQSYGFQPTNIGVGYQPYRVVVLFSPELLHAP	521
6W41 C	KPFFERDISTEIQAGSTPCNGVEGFNCYFPLQSYGFQPTNIGVGYQPYRVVVLFSPELLHAP	203
QIA20044.1	KPFFERDISTEIQAGSTPCNGVEGFNCYFPLQSYGFQPTNIGVGYQPYRVVVLFSPELLHAP	521
YP_009724390.1	KPFFERDISTEIQAGSTPCNGVEGFNCYFPLQSYGFQPTNIGVGYQPYRVVVLFSPELLHAP	521
6M17 E	KPFFERDISTEIQAGSTPCNGVEGFNCYFPLQSYGFQPTNIGVGYQPYRVVVLFSPELLHAP	203
QHR63250.2	KPFFERDISTEIQAGSTPCNGVEGFNCYFPLQSYGFQPTNIGVGYQPYRVVVLFSPELLHAP	521
QIC53213.1	KPFFERDISTEIQAGSTPCNGVEGFNCYFPLQSYGFQPTNIGVGYQPYRVVVLFSPELLHAP	521
6VYB A	ATVCGPKKSTNLVKNKCVNFFNFGLTGTGVLTESNKKFLPFQQFGRDIADTTDAVRDPQT	600
6VSB A	ATVCGPKKSTNLVKNKCVNFFNFGLTGTGVLTESNKKFLPFQQFGRDIADTTDAVRDPQT	581
6W41 C	ATVCGPKKSTNLVKNKCVNFFNFGLTGTGVLTESNKKFLPFQQFGRDIADTTDAVRDPQT	231
QIA20044.1	ATVCGPKKSTNLVKNKCVNFFNFGLTGTGVLTESNKKFLPFQQFGRDIADTTDAVRDPQT	581
YP_009724390.1	ATVCGPKKSTNLVKNKCVNFFNFGLTGTGVLTESNKKFLPFQQFGRDIADTTDAVRDPQT	581
6M17 E	ATVCGPKKSTNLVKNKCVNFFNFGLTGTGVLTESNKKFLPFQQFGRDIADTTDAVRDPQT	223
QHR63250.2	ATVCGPKKSTNLVKNKCVNFFNFGLTGTGVLTESNKKFLPFQQFGRDIADTTDAVRDPQT	581
QIC53213.1	ATVCGPKKSTNLVKNKCVNFFNFGLTGTGVLTESNKKFLPFQQFGRDIADTTDAVRDPQT	581



La **tabla 1** contiene los epítopes T de la glicoproteína S de SARS-CoV-2 que pueden ser reconocidos por linfocitos T en el contexto de la molécula HLA-A\*0201, muy frecuente en la población cubana, de acuerdo con los algoritmos empleados. Como puede apreciarse, existen coincidencias en la modelación de la presentación antigénica entre las tres herramientas, como es el caso de los nonúmeros YLQPRTFLL, ubicado en la posición 269-277, y FIAGLIAIV (1220-1228); también, la región 996-1008,

en la que se superponen los péptidos LITGRLQSL y RLQSLQTYV. Otros nonúmeros son propuestos por al menos dos de los algoritmos: FQFCNDPFL (133-141), KIADYNYKL (417-425), SIIAYTMSL (691-699), FTISVTTEI (718-726), VLNDILSRL (976-984), RLDKVEAEV (983-991) y HLMSFPQSA (1048-1056). Con la excepción de FIAGLIAIV, el resto de los epítopes T CD8+ se ubican en la porción extracelular de la proteína, de acuerdo con lo descrito en la entrada correspondiente a la molécula en UniProt. <sup>(3)</sup>

**TABLA 1. Predicción de epítopes T CD8+ de la glicoproteína S de SARS-CoV-2 para HLA-A\*0201**

Nº	SYFPEITHI		IEDB		NetMHC	
	Epítope	Pos.	Epítope	Pos.	Epítope	Pos.
1	FIAGLIAIV	1220	YLQPRTFLL	269	YLQPRTFLL	269
2	SALEPLVDL	221	FIAGLIAIV	1220	FQFCNDPFL	133
3	SIIAYTMSL	691	FVFLVLLPL	2	FIAGLIAIV	1220
4	ALNTLVKQL	958	KIADYNYKL	417	SIIAYTMSL	691
5	VLNDILSRL	976	RLQSLQTYV	1000	RLQSLQTYV	1000
6	LITGRLQSL	996	FTISVTTEI	718	FTISVTTEI	718
7	NLNESLIDL	1192	RLDKVEAEV	983	LLFNKVTLA	821
8	YLQPRTFLL	269	HLMSFPQSA	1048	HLMSFPQSA	1048
9	KIADYNYKL	417	FQFCNDPFL	133	RLDKVEAEV	983
10	TLDSKTQSL	109	VLNDILSRL	976	VVFLHVTYV	1060

Leyenda: Pos.: posición del primer aminoácido del péptido en la secuencia primaria de la proteína

De manera similar, el péptido PINLVRDLPQGFS (209-222) fue identificado por los tres algoritmos empleados en la predicción de epítopes a ser reconocidos por los linfocitos T CD4+, al ser presentados por la molécula HLA-DRB1\*0301 (**tabla**

2). En general, la región entre los aminoácidos 207-223 fue señalada por todas las herramientas bioinformáticas, mientras dos de ellas identificaron la secuencia ubicada en las posiciones 44-47.

**TABLA 2. Predicción de epítopes T CD4+ de la glicoproteína s de SARS-CoV-2 para HLA-DRB1\*0301**

Nº	SYFPEITHI		IEDB		NetMHCII	
	Epítope	Pos.	Epítope	Pos.	Epítope	Pos.
1	APAICHDGKAHFPR	1078	INLVRDLPQGFSAL	210	PVAIHADQLTPTWR	621
2	PINLVRDLPQGFS	209	PINLVRDLPQGFS	209	VAIHADQLTPTWRV	622
3	ISSVLNDILSRLDK	973	TPINLVRDLPQGFS	208	PQIITDNTFVSGN	1112
4	SQPFLMDLEKQGN	172	HTPINLVRDLPQGF	207	EPQIITDNTFVSG	1111
5	GTGVLTESNKKFLP	548	DSFVIRGDEVRQIA	398	PINLVRDLPQGFS	209
6	RSSVLHSTQDLFLP	44	EAEVQIDRLITGRL	988	VPVAIHADQLTPTW	620
7	GIAVEQDKNTQEVF	769	SFVIRGDEVRQIAP	399	TPINLVRDLPQGFS	208
8	GTTLDLSTQSLIV	107	AEVQIDRLITGRLQ	989	QIITDNTFVSGNC	1113
9	FQTLALHRSYLTP	238	VEAEVQIDRLITGR	987	INLVRDLPQGFSAL	210
10	QTGKIADYNYKLPD	414	RGVYYPDKVFRSSV	34	AIHADQLTPTWRVY	623

Leyenda: Pos.: posición del primer aminoácido del péptido en la secuencia primaria de la proteína



Al emplear el alelo HLA-DRB1\*0701, los tres algoritmos identificaron el epítipo PTNFTISVTTEILP (715-728), así como la región entre los residuos 712 y 731, como se observa en la **tabla 3**. Es de destacar la coincidencia de los péptidos

RSSLVHSTQDLFLPF (44-58) y GTTLDSKTQSLIV (107-120) entre los alelos de clase II seleccionados, tanto HLA-DRB1\*0301 como HLA-DRB1\*0701. También para los epítopes T mostrados, todos se ubicaron en la región extracelular de la proteína.

**TABLA 3. Predicción de epítopes T CD4+ de la glicoproteína S de SARS-CoV-2 para HLA-DRB1\*0701**

Nº	SYFPEITHI		IEDB		NetMHCII	
	Epítipo	Pos.	Epítipo	Pos.	Epítipo	Pos.
1	PPAYTNSFTRGVYY	25	IPNFTISVTTEIL	714	AIPNFTISVTTEI	712
2	PTNFTISVTTEILP	715	AIPNFTISVTTEI	713	IPNFTISVTTEIL	713
3	SSVLHSTQDLFLPF	45	TNFTISVTTEILPV	716	QSIIAYTMSLGAEN	689
4	EKGIYQTSNFRVQP	309	PTNFTISVTTEILP	715	PTNFTISVTTEILP	715
5	TWRVYSTGNSVFQT	632	NFTISVTTEILPVS	717	TNFTISVTTEILPV	715
6	SALLAGTITSGWTF	875	FTISVTTEILPVSM	718	ASQSIIAYTMSLGA	687
7	EPQIITDNTFVSG	1111	LYRLFRKSNLKPFE	452	SQSIIAYTMSLGAE	688
8	RTQLPPAYTNSFTR	21	YLYRLFRKSNLKPFE	451	SIIAYTMSLGAENS	690
9	LLIVNNATNVVIKV	17	YRLFRKSNLKPFE	453	NFTISVTTEILPVS	716
10	FRVYSSANNCTFEY	157	GTTLDSKTQSLIV	107	IIAYTMSLGAENSV	691

Leyenda: Pos.: posición del primer aminoácido del péptido en la secuencia primaria de la proteína

La **tabla 4** muestra la representación de epítopes B lineales, con letras en azul y subrayadas, en la secuencia primaria de la glicoproteína S de SARS-CoV-2, de acuerdo con la modelación de los cuatro algoritmos utilizados; se colorearon en rojo dos regiones claves en la patogenia viral: el RBM y el FP. Fueron sombreadas en amarillo aquellas regiones en las que hubo coincidencia de al menos tres de las herramientas empleadas, y sombreados en azul los residuos de interacción directa con ACE2 que quedaron incluidos en alguno de los epítopes B predichos.

Los epítopes B más extensos, con al menos siete aminoácidos, estuvieron comprendidos entre las posiciones 14-31, 71-81, 141-147, 207-214, 249-260, 374-385, 389-395, 407-420, 439-447, 522-535, 615-621, 656-665, 672-687, 788-797, 806-815 y 1137-1142. Como mismo sucedió con los epítopes T identificados anteriormente, los determinantes antigénicos reconocibles por los receptores de los linfocitos B se distribuyeron por la región extracelular, con la única excepción del último punto de coincidencia entre los tres algoritmos (EPVL).

Vale destacar, finalmente, que la región comprendida entre los residuos 207-222 contiene epítopes B y T CD4+, al tiempo que en las posiciones 133-147 y 407-425 se concentran epítopes B y T CD8+; todas están ubicadas en el segmento externo de la proteína. En el RBM se localizan varios epítopes B y, también, incluye a cinco de los residuos de interacción directa entre la glicoproteína S y ACE2; el epítipo LPDPSKPSKR forma parte del PF.

## DISCUSIÓN

La glicoproteína S de SARS-CoV-2 determina el tropismo viral, dada su capacidad de unión a su receptor en el humano, ACE2, y promover la fusión de las membranas para iniciar la invasión celular.<sup>(9)</sup> Su neutralización, o la inhibición de la unión y fusión, pueden ser parte de las opciones farmacológicas para la prevención o el tratamiento de la COVID-19.

Las secuencias empleadas para comparar el grado de conservación de la estructura de la molécula, particularmente el RBD, mostraron una homología completa, lo que traduce la importancia de ese dominio para la función de la proteína y, a su vez, para el ciclo de vida del virus, cuya capacidad infecciosa podría verse comprometida por los cambios que son, por esa razón, "aceptados" con mayor dificultad.<sup>(14)</sup> La estabilidad de esa estructura, al tiempo que es un mecanismo de supervivencia que mantiene y protege el tropismo viral, puede verse igualmente como una oportunidad, al convertirse en una diana sobre la cual dirigir las investigaciones para la erradicación del patógeno.

Si bien los virus ARN pueden tener altas tasas de mutación, en el caso de SARS-CoV-2, las cepas procedentes de lugares tan diversos como California, Brasil, Taiwán e India, se mantenían idénticas o muy similares, hasta el mes de marzo de 2020 y se correspondían con la secuencia de referencia YP\_009724390.1 hasta el 2 de abril de este año.<sup>(14)</sup> Para secuencias específicas dentro de la molécula, entre ellas los residuos de unión a ACE2, no se habían reportado cambios hasta el momento en que se escribió este reporte.<sup>(4,14)</sup> Tal estabilidad apoya la candidatura de la glicoproteína S, como una alternativa a considerar para conformar una vacuna.



**TABLA 4. Predicción de epítopes B lineales de la glicoproteína S de SARS-CoV-2**

BepiPred-2.0	MFVFLVLLPLVSSQCVNLTTRTOLPPAYTNSFTRGVYYPDKVFRSSVLHSTQDLFLPFESNV
ABCpred	MFVFLVLLPLVSSQCVNLTTRTOLPPAYTNSFTRGVYYPDKVFRSSVLHSTQDLFLPFESNV
BcePred	MFVFLVLLPLVSSQCVNLTTRTOLPPAYTNSFTRGVYYPDKVFRSSVLHSTQDLFLPFESNV
Bepipred-IEDB	MFVFLVLLPLVSSQCVNLTTRTOLPPAYTNSFTRGVYYPDKVFRSSVLHSTQDLFLPFESNV
BepiPred-2.0	TWFHAIHVSGTNGTKRFDPNPLVPFNDDGVYFASTEKSNIRGWIFGTTLDLSDKTSLLIVNNA
ABCpred	TWFHAIHVSGTNGTKRFDPNPLVPFNDDGVYFASTEKSNIRGWIFGTTLDLSDKTSLLIVNNA
BcePred	TWFHAIHVSGTNGTKRFDPNPLVPFNDDGVYFASTEKSNIRGWIFGTTLDLSDKTSLLIVNNA
Bepipred-IEDB	TWFHAIHVSGTNGTKRFDPNPLVPFNDDGVYFASTEKSNIRGWIFGTTLDLSDKTSLLIVNNA
BepiPred-2.0	TNVVIVKCEFCQNDPFLGVYYHKNNKSWMESEFRVYSSANNCTFEYVSQPFMLDLEGGK
ABCpred	TNVVIVKCEFCQNDPFLGVYYHKNNKSWMESEFRVYSSANNCTFEYVSQPFMLDLEGGK
BcePred	TNVVIVKCEFCQNDPFLGVYYHKNNKSWMESEFRVYSSANNCTFEYVSQPFMLDLEGGK
Bepipred-IEDB	TNVVIVKCEFCQNDPFLGVYYHKNNKSWMESEFRVYSSANNCTFEYVSQPFMLDLEGGK
BepiPred-2.0	GNFKNLREFVFNKIDGYFKIYSKHTPINLVRDLPQGFSALEPLVDLPIGINITRFQTLALHR
ABCpred	GNFKNLREFVFNKIDGYFKIYSKHTPINLVRDLPQGFSALEPLVDLPIGINITRFQTLALHR
BcePred	GNFKNLREFVFNKIDGYFKIYSKHTPINLVRDLPQGFSALEPLVDLPIGINITRFQTLALHR
Bepipred-IEDB	GNFKNLREFVFNKIDGYFKIYSKHTPINLVRDLPQGFSALEPLVDLPIGINITRFQTLALHR
BepiPred-2.0	SYLTPGDSSSGWTAAGAAAYVGYLQPRFTLLKYNENGTITDAVDCALDPLSEKTKLTSFT
ABCpred	SYLTPGDSSSGWTAAGAAAYVGYLQPRFTLLKYNENGTITDAVDCALDPLSEKTKLTSFT
BcePred	SYLTPGDSSSGWTAAGAAAYVGYLQPRFTLLKYNENGTITDAVDCALDPLSEKTKLTSFT
Bepipred-IEDB	SYLTPGDSSSGWTAAGAAAYVGYLQPRFTLLKYNENGTITDAVDCALDPLSEKTKLTSFT
BepiPred-2.0	VEKGIYQTSNFRVQPTESIVRFPNITNLCPEGEVFNATRFASVYAWNRKRISNCVADYSVL
ABCpred	VEKGIYQTSNFRVQPTESIVRFPNITNLCPEGEVFNATRFASVYAWNRKRISNCVADYSVL
BcePred	VEKGIYQTSNFRVQPTESIVRFPNITNLCPEGEVFNATRFASVYAWNRKRISNCVADYSVL
Bepipred-IEDB	VEKGIYQTSNFRVQPTESIVRFPNITNLCPEGEVFNATRFASVYAWNRKRISNCVADYSVL
BepiPred-2.0	YNSASFSTFKCYGVSPTKLNLDLCTNVAADSFVIRGDEVROIAPGOTGKIADYNYKLPDDF
ABCpred	YNSASFSTFKCYGVSPTKLNLDLCTNVAADSFVIRGDEVROIAPGOTGKIADYNYKLPDDF
BcePred	YNSASFSTFKCYGVSPTKLNLDLCTNVAADSFVIRGDEVROIAPGOTGKIADYNYKLPDDF
Bepipred-IEDB	YNSASFSTFKCYGVSPTKLNLDLCTNVAADSFVIRGDEVROIAPGOTGKIADYNYKLPDDF
BepiPred-2.0	TGCVIAWNSNLDLSDKVGGNYYLYRFRKSNLKPFERDISTEIQAGSTPCNGVEGFNCY
ABCpred	TGCVIAWNSNLDLSDKVGGNYYLYRFRKSNLKPFERDISTEIQAGSTPCNGVEGFNCY
BcePred	TGCVIAWNSNLDLSDKVGGNYYLYRFRKSNLKPFERDISTEIQAGSTPCNGVEGFNCY
Bepipred-IEDB	TGCVIAWNSNLDLSDKVGGNYYLYRFRKSNLKPFERDISTEIQAGSTPCNGVEGFNCY
BepiPred-2.0	FPLQSYGFQPTNGVGYQPYRVVLSFELLHAPATVCGPKKSTNLVKNKCVNFMFNGLTGT
ABCpred	FPLQSYGFQPTNGVGYQPYRVVLSFELLHAPATVCGPKKSTNLVKNKCVNFMFNGLTGT
BcePred	FPLQSYGFQPTNGVGYQPYRVVLSFELLHAPATVCGPKKSTNLVKNKCVNFMFNGLTGT
Bepipred-IEDB	FPLQSYGFQPTNGVGYQPYRVVLSFELLHAPATVCGPKKSTNLVKNKCVNFMFNGLTGT
BepiPred-2.0	GVLTESNKKFLPFQOQGRDIADTTDAVRDPQTELEIDITPCSFGGVSVITPGTNTSNQAV
ABCpred	GVLTESNKKFLPFQOQGRDIADTTDAVRDPQTELEIDITPCSFGGVSVITPGTNTSNQAV
BcePred	GVLTESNKKFLPFQOQGRDIADTTDAVRDPQTELEIDITPCSFGGVSVITPGTNTSNQAV
Bepipred-IEDB	GVLTESNKKFLPFQOQGRDIADTTDAVRDPQTELEIDITPCSFGGVSVITPGTNTSNQAV
BepiPred-2.0	LYQDVNCTEVPVAIHADQLTPTWRVYSTGSNVQTRAGCLLIGAEHVNSNYECCDPIIGAGIC
ABCpred	LYQDVNCTEVPVAIHADQLTPTWRVYSTGSNVQTRAGCLLIGAEHVNSNYECCDPIIGAGIC
BcePred	LYQDVNCTEVPVAIHADQLTPTWRVYSTGSNVQTRAGCLLIGAEHVNSNYECCDPIIGAGIC
Bepipred-IEDB	LYQDVNCTEVPVAIHADQLTPTWRVYSTGSNVQTRAGCLLIGAEHVNSNYECCDPIIGAGIC
BepiPred-2.0	ASYQTOTNSPRRARSVASQSIIAYTMSLGAENSVAYSNNSIAIPTNFTISVTTEILPVSMTK
ABCpred	ASYQTOTNSPRRARSVASQSIIAYTMSLGAENSVAYSNNSIAIPTNFTISVTTEILPVSMTK
BcePred	ASYQTOTNSPRRARSVASQSIIAYTMSLGAENSVAYSNNSIAIPTNFTISVTTEILPVSMTK
Bepipred-IEDB	ASYQTOTNSPRRARSVASQSIIAYTMSLGAENSVAYSNNSIAIPTNFTISVTTEILPVSMTK
BepiPred-2.0	TSVDCTMYICGDSTECSNLLLOYGSFCTQLNRALTGIAVEQDKNTOEVFAQVKQIYKTPPI
ABCpred	TSVDCTMYICGDSTECSNLLLOYGSFCTQLNRALTGIAVEQDKNTOEVFAQVKQIYKTPPI
BcePred	TSVDCTMYICGDSTECSNLLLOYGSFCTQLNRALTGIAVEQDKNTOEVFAQVKQIYKTPPI
Bepipred-IEDB	TSVDCTMYICGDSTECSNLLLOYGSFCTQLNRALTGIAVEQDKNTOEVFAQVKQIYKTPPI
BepiPred-2.0	KDFGGFNFSQILPDPSPKSKRSFIEDLLFNKVTLADAGFIKOYGDCLGDIARDLCAQKFN
ABCpred	KDFGGFNFSQILPDPSPKSKRSFIEDLLFNKVTLADAGFIKOYGDCLGDIARDLCAQKFN
BcePred	KDFGGFNFSQILPDPSPKSKRSFIEDLLFNKVTLADAGFIKOYGDCLGDIARDLCAQKFN
Bepipred-IEDB	KDFGGFNFSQILPDPSPKSKRSFIEDLLFNKVTLADAGFIKOYGDCLGDIARDLCAQKFN
BepiPred-2.0	GLTVLPPLLTDEMIAQYTSALLAGTITSGWTFGAGAALQIPFAMQMAYRFNIGVTONVLY
ABCpred	GLTVLPPLLTDEMIAQYTSALLAGTITSGWTFGAGAALQIPFAMQMAYRFNIGVTONVLY
BcePred	GLTVLPPLLTDEMIAQYTSALLAGTITSGWTFGAGAALQIPFAMQMAYRFNIGVTONVLY
Bepipred-IEDB	GLTVLPPLLTDEMIAQYTSALLAGTITSGWTFGAGAALQIPFAMQMAYRFNIGVTONVLY
BepiPred-2.0	ENQKLIANQFNLSAIGKIQDLSSTASALGKLDVNVNOAQAALNTLVKQLSSNFGAISSVL
ABCpred	ENQKLIANQFNLSAIGKIQDLSSTASALGKLDVNVNOAQAALNTLVKQLSSNFGAISSVL
BcePred	ENQKLIANQFNLSAIGKIQDLSSTASALGKLDVNVNOAQAALNTLVKQLSSNFGAISSVL
Bepipred-IEDB	ENQKLIANQFNLSAIGKIQDLSSTASALGKLDVNVNOAQAALNTLVKQLSSNFGAISSVL
BepiPred-2.0	NDILSRDKVEAEVQIDRLITGRLOSLOTYVTOOLIRAAEIRASANLAATKMSECVLGOSK
ABCpred	NDILSRDKVEAEVQIDRLITGRLOSLOTYVTOOLIRAAEIRASANLAATKMSECVLGOSK
BcePred	NDILSRDKVEAEVQIDRLITGRLOSLOTYVTOOLIRAAEIRASANLAATKMSECVLGOSK
Bepipred-IEDB	NDILSRDKVEAEVQIDRLITGRLOSLOTYVTOOLIRAAEIRASANLAATKMSECVLGOSK
BepiPred-2.0	RVDFCGKGYHLMSPQSAFHGVVFLHVTVYPAQEKNFTTAPAICHGDKAHFREGVFSN
ABCpred	RVDFCGKGYHLMSPQSAFHGVVFLHVTVYPAQEKNFTTAPAICHGDKAHFREGVFSN
BcePred	RVDFCGKGYHLMSPQSAFHGVVFLHVTVYPAQEKNFTTAPAICHGDKAHFREGVFSN
Bepipred-IEDB	RVDFCGKGYHLMSPQSAFHGVVFLHVTVYPAQEKNFTTAPAICHGDKAHFREGVFSN
BepiPred-2.0	GTHWVFTQRNFYEPQIITDNTFVSGNCDVIGIVNNTVYDPLQPELDSFKEELDKYFKNH
ABCpred	GTHWVFTQRNFYEPQIITDNTFVSGNCDVIGIVNNTVYDPLQPELDSFKEELDKYFKNH
BcePred	GTHWVFTQRNFYEPQIITDNTFVSGNCDVIGIVNNTVYDPLQPELDSFKEELDKYFKNH
Bepipred-IEDB	GTHWVFTQRNFYEPQIITDNTFVSGNCDVIGIVNNTVYDPLQPELDSFKEELDKYFKNH
BepiPred-2.0	TSPDVLGDJSGINASVNNIQKEIDRLNEVAKNLNESLIDLQELGKYEQYIKWPWYIWLGF
ABCpred	TSPDVLGDJSGINASVNNIQKEIDRLNEVAKNLNESLIDLQELGKYEQYIKWPWYIWLGF
BcePred	TSPDVLGDJSGINASVNNIQKEIDRLNEVAKNLNESLIDLQELGKYEQYIKWPWYIWLGF
Bepipred-IEDB	TSPDVLGDJSGINASVNNIQKEIDRLNEVAKNLNESLIDLQELGKYEQYIKWPWYIWLGF
BepiPred-2.0	IAGLIAVMVTIMLCCMTSCCSCLKGCSCGSCCKFDEDDSEPVILKGVKLYHT
ABCpred	IAGLIAVMVTIMLCCMTSCCSCLKGCSCGSCCKFDEDDSEPVILKGVKLYHT
BcePred	IAGLIAVMVTIMLCCMTSCCSCLKGCSCGSCCKFDEDDSEPVILKGVKLYHT
Bepipred-IEDB	IAGLIAVMVTIMLCCMTSCCSCLKGCSCGSCCKFDEDDSEPVILKGVKLYHT



La modelación de la presentación antigénica a células T se realizó incluyendo moléculas de histocompatibilidad reportadas como frecuentes en Cuba y otros países, <sup>(11)</sup> y se combinaron tres algoritmos bioinformáticos diferentes, en busca de mayor consistencia, cuyos resultados fueron comparados con la intención de sugerir los mejores péptidos para un posible candidato de vacuna multiepitópica o similar. Ya se había sugerido que la glicoproteína S, junto a la nucleoproteína de SARS-CoV-2, contienen epítopes T, como RGDEVRQIAPGQTGKIAD (posición 403 de la secuencia de referencia), <sup>(13)</sup> con el que superponen los epítopes T KIADYNYKL, QTGKIADYNYKLPD y el epítipo B VRQIAPGQTGKIAD, todos reportados en el presente trabajo. En esta región se localizan las asas 1a y 1d, de probable interacción directa con ACE2, <sup>(13)</sup> por lo que puede tratarse de una diana adecuada para bloquear la infección viral.

Algunos péptidos derivados de la glicoproteína S del virus SARS, que dio origen en el año 2002 a un brote de menor magnitud que la actual pandemia, tienen homología con epítopes de SARS-CoV-2 <sup>(11)</sup> y coinciden igualmente con lo aquí obtenido:

- ALNTLVKQL, predicho por SYFPEITHI para clase I, pos. 958, y por dos de los algoritmos para péptidos B;
- NLNESLIDL, predicho por SYFPEITHI, pos. 1192;
- VLNDILSRL, predicho por SYFPEITHI e IEDB, pos. 976, como clase I, y por SYFPEITHI como clase II.

En la evaluación de la inmunogenicidad de péptidos derivados del proteoma de SARS-CoV-2, por medio de un algoritmo diferente, iPred, <sup>(11)</sup> se encontraron epítopes que coinciden con varios de los hallados en la presente investigación:

- SIIAYTMSL, predicho ahora por SYFPEITHI y NetMHCII, como epítipo de clase II;
- AGCLIGAEHVNSYECDI, contiene péptido VNNS-E, predicho como epítipo B;
- DEVRQI, predicho por IEDB como de clase II, y contiene epítipo B reconocido por tres de los algoritmos empleados;
- TLDSKTQSL, predicho como epítipo T, CD4+ y CD8+, para los tres alelos evaluados;
- YLQPRTFLL, predicho por los tres algoritmos, como epítipo T de clase I.

De manera similar, la modelación realizada para obtener secuencias con potencial como epítopes B, que empleó cuatro herramientas bioinformáticas, arrojó resultados que se corresponden con hallazgos de otros investigadores. La glicoproteína S, junto a la fosfoproteína de la nucleocápsida y la glicoproteína de membrana, ya habían sido identificadas como dianas de la respuesta adaptativa humoral contra otros coronavirus; como cabría esperar, hubo coincidencia con las siete regiones identificadas en la molécula, por otro grupo de investigadores con el uso de BepiPred 2.0, uno de los mismos algoritmos del presente trabajo: 491-505, 558-562, 703-704, 793-794, 810, 914 y 1140-1146, <sup>(7)</sup> comparables a lo mostrado en la **tabla 4**. Otras cinco regiones fueron propuestas por los mismos autores, asociadas a altas

tasas de respuesta inmune: 274-306, 510-586, 587-628, 784-803 y 870-893; <sup>(7)</sup> al menos tres de ellas comparten residuos con los epítopes B de al menos siete aminoácidos mencionados en los resultados de este trabajo.

Algunos de los epítopes B, encontrados en la presente modelación y coincidentes con la literatura, son sitios de interés, debido a la probabilidad de que contengan los residuos que interactúan con ACE2, el receptor de SARS-CoV-2 en el humano: <sup>(14)</sup>

- LDSKVGG (contiene el asa 2g, pos. 441, ubicada en el RBM y predicho por tres de los algoritmos empleados);
- YQAGST (asa 3a, pos. 473 del RBM, por tres algoritmos);
- FNCYFPL (asa 3b, pos. 486 del RBM, por tres algoritmos);
- FQPT (asa 3f, pos. 497 del RBM, por tres algoritmos); y
- GFNCYFPLQSYGFQPT, que contiene varios de los residuos de interacción directa con ACE2 (Q498, T500, N501) y ha sido propuesto como un péptido análogo. <sup>(14)</sup>

Otra modelación de epítopes B, que utilizó la herramienta IEDB, encontró 13 péptidos en la glicoproteína S: <sup>(15)</sup> VLLPLVSSQCVNLTTRTQLPPAYTN, VTWFHAIHVSNGTKRFDN, VRFPNITNLCPFGE, GCVIA WNSNNLDSKVGG, LKPFERDISTEIQAGSTPCNGVEGF N, GFQPTNGVGYQPY y PDPSKPSKRSFI. Todos ellos contienen o comparten secuencias identificadas en los resultados mostrados en la **tabla 4**.

Hasta aquí, ha habido correspondencia entre lo modelado con el diseño in silico empleado en el presente trabajo y lo realizado por otros colegas. Hay que añadir que otras evidencias experimentales in vivo apoyan la validez de la simulación computacional realizada para la presentación antigénica de la glicoproteína S de SARS-CoV-2. Un anticuerpo neutralizante, denominado CR3022, obtenido de un paciente convaleciente de SARS, pero con reactividad cruzada con el RBD de SARS-CoV-2, ha sido estudiado por cristalografía a una resolución de 3,1 Å, lo que permitió identificar los residuos que median la interacción entre CR3022 y el RBD de SARS-CoV-2: A372, F374, F377, K378, Y380, V382, L390, F392, T430, F515 y L517. <sup>(2)</sup> He aquí un soporte experimental importante que apoya la validez de la modelación realizada, pues esos residuos se localizan en regiones que contienen epítopes B predichos por los algoritmos empleados: YNSA-FSTFKCYGVSPTK--DL-F (pos. 369), DDFT (428-431) y EL-H (515-518).

La interacción de ese mismo anticuerpo, CR3022, con la glicoproteína S de SARS-CoV-2, fue estudiada a una resolución de 1,95 Å, la más alta jamás lograda hasta la fecha. <sup>(7)</sup> Resulta reconfortante comprobar que en el epítipo identificado, catalogado como críptico, porque se encuentra inaccesible en la configuración nativa de la glicoproteína S y solo se muestra en una conformación "abierta", <sup>(7)</sup> aparece la secuencia FKCYGVSPTK, que fue reconocida como

un epítipo B, por los algoritmos utilizados en la presente investigación. Ello habla de la fortaleza de la modelación computacional de la identificación de péptidos con potencial como epítipos.

La secuencia KRSFIEDLLFNKV, que se ubica en el FP de la glicoproteína S (pos. 814), está altamente conservado, no solo en el SARS-CoV-2, sino entre coronavirus menos relacionados, por lo que se ha especulado que podría tener una función importante para el patógeno, ya sea en la activación proteolítica, la unión celular u otras interacciones; ha sido propuesto para una vacuna que, incluso, brindaría protección cruzada contra cepas o especies de coronavirus diferentes.<sup>(14)</sup> Esa secuencia comparte algunos aminoácidos con LPDPSKPSKR, epítipo B (pos. 806) predicho en la modelación realizada para esta investigación.

Este reporte identifica un grupo de péptidos que pueden ser considerados epítipos a ser reconocidos por los linfocitos T y B humanos en la respuesta inmune adaptativa frente al SARS-CoV-2. Se describieron tanto epítipos T como B, así como regiones en las que confluyen ambos tipos de péptidos, que pueden tener una mayor influencia en la inmunogenicidad deseada, al aportar la estructura que serviría de blanco a los efectores de la respuesta adaptativa y la estimulación de la necesaria de la cooperación por linfocitos T CD4+. Se localizaron aquellos que se ubican en regiones críticas de la glicoproteína S, tanto en el sitio de unión a su receptor específico como en el dominio responsable de la fusión y la penetración al interior celular; ello podría conducir a un mayor efecto neutralizante o inhibitorio.

Todos los elementos hasta aquí abordados pueden ser vistos desde la perspectiva, en primer lugar, de la utilidad de la metodología y los algoritmos empleados para la predicción de epítipos y la identificación de regiones inmunogénicas en la glicoproteína S de SARS-CoV-2 que, al mismo tiempo, puedan ser evaluados para los diversos usos posibles: el diseño de candidatos vacunales, fármacos y sistemas diagnósticos; la caracterización de la respuesta inmune humana, su expresión poblacional y su relación con el espectro clínico de la enfermedad, entre otras. Algunos de los péptidos aquí identificados podrían ser empleados como antagonistas de la interacción con ACE2; su potencial terapéutico se basa en la experiencia de otras moléculas pequeñas, en uso o investigación, de menor toxicidad que los fármacos tradicionales y con características propias como la resistencia a la biodegradación.<sup>(14)</sup>

Estos resultados, no obstante, son propuestos no como solución única y definitiva, sino para ser integrados como parte de los criterios a considerar en el estudio de la respuesta al nuevo coronavirus y sus potenciales aplicaciones. Las interacciones moleculares entre la glicoproteína S y su receptor en el humano, ACE2, podrían no ser las únicas implicadas en el proceso de infección viral, sus dinámicas deben ser completamente dilucidadas, al

igual que su influencia en la respuesta inmune. En el mismo sentido, el mecanismo de neutralización del virus SARS por el anticuerpo CR3022, no depende de la unión directa al receptor, en tanto las cadenas monoméricas que forman el homotrímero de la glicoproteína S, tienen cierta flexibilidad, por un efecto similar al de una bisagra, que hace que la molécula tenga disposiciones espaciales variables, que influirían en la interacción efectiva con ACE2 u otros ligandos.<sup>(2)</sup>

La glicoproteína S no es el único antígeno viral y queda por ver qué lugar ocupa entre todas las estructuras del patógeno, en términos de la estimulación de los efectores de la inmunidad. En el mismo sentido, cómo influyen las modificaciones postraduccionales, algunas de las cuales pueden predecirse desde la secuencia primaria. Tal es el caso de la glicosilación, que promueve cambios en el plegamiento de la proteína, la interacción con ligandos y el acceso y susceptibilidad a las proteasas y los anticuerpos;<sup>(4)</sup> se ha considerado que puede inactivar a un epítipo B.<sup>(14)</sup> A la N-glicosilación en el residuo T370, por ejemplo, se le han atribuido las diferencias en la inmunogenicidad entre los RBD de SARS y SARS-CoV-2.<sup>(2)</sup>

También, los epítipos identificados pueden tener inmunodominancia o cripticidad diversas y se debe definir su relación en la protección efectiva frente al virus, así como la inducción de memoria y su duración, su promiscuidad en términos de afinidad por uno o varios alelos HLA, entre algunos aspectos claves de la respuesta inmune específica. Desde el punto de vista poblacional, se debe considerar la variabilidad de la respuesta y la identificación de los respondedores, hiper-respondedores y malos respondedores, tanto frente al virus como a los candidatos vacunales. Desde el punto de vista clínico, qué diferencias hay en la respuesta frente a estos u otros epítipos, entre los asintomáticos y los pacientes de evolución tórpida. Se debe pensar en la posibilidad de que los péptidos identificados puedan ofrecer protección cruzada frente a otros coronavirus, dada la conservación de secuencias entre ellos y algunas evidencias experimentales en animales,<sup>(2,4)</sup> lo que ampliaría la aplicación de las vacunas en que se incluyan.

La glicoproteína S del SARS-CoV-2 tiene un potencial inmunogénico, con el que puede generar respuesta T y B. Los resultados de este reporte pueden ser de utilidad en la selección de péptidos para diseño de vacunas y sistemas diagnósticos, terapias basadas en péptidos, además de aplicaciones inmunoepidemiológicas, entre otras.

## AGRADECIMIENTOS

A la Dra. Jenny de la Caridad Hernández Betancourt, por la revisión crítica del manuscrito. Los resultados presentados forman parte del proyecto institucional "Creación de una unidad de investigaciones moleculares en el Hospital General Docente Dr. Ernesto Guevara de Las Tunas."




**REFERENCIAS BIBLIOGRÁFICAS:**

1. Chen Y, Guo Y, Pan Y, Zhao ZJ. Structure analysis of the receptor binding of 2019-nCoV. *Biochemical and Biophysical Research Communications* [revista en internet]. 2020 [citado 28 de abril 2020]; 525: 135-140. Disponible en: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7092824/>.
2. Yuan M, Wu NC, Zhu X, Lee CCD, So RTY, Lv H, et al. A highly conserved cryptic epitope in the receptor-binding domains of SARS-CoV-2 and SARS-CoV. *Science* [revista en internet]. 2020 [citado 28 de abril 2020]; doi: 10.1126/science.abb7269. Disponible en: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7164391/>.
3. S Glycoprotein. UniProt, 2020 [citado 28 de abril 2020]. Disponible en: <https://covid-19.uniprot.org/uniprot/kb/P0DTC2>.
4. Walls AC, Park YJ, Tortorici MA, Wall A, McGuire AT, Veesler D. Structure, Function, and Antigenicity of the SARS-CoV-2 Spike Glycoprotein. *Cell* [revista en internet]. 2020 [citado 28 de abril 2020]; 180: 1-12. Disponible en: <https://www.sciencedirect.com/science/article/pii/S0092867420302622>.
5. Su Z, Wu Y. A Multiscale and Comparative Model for Receptor Binding of 2019 Novel Coronavirus and the Implication of its Life Cycle in Host Cells. *BioRxiv*. 2020 [citado 28 de abril 2020]; doi: <https://doi.org/10.1101/2020.02.20.958272>. Disponible en: <https://www.biorxiv.org/content/10.1101/2020.02.20.958272v1.full>.
6. Yan R, Zhang Y, Li Y, Xia L, Guo Y, Zhou Q. Structural basis for the recognition of SARS-CoV-2 by full-length human ACE2. *Science* [revista en internet]. 2020 [citado 6 de abril 2020]; 367: 1444-1448. Disponible en: <https://science.sciencemag.org/content/367/6485/1444.full>.
7. Joyce MG, Sankhala RS, Chen WH, Choe M, Bai H, Hajduczek A, et al. A Cryptic Site of Vulnerability on the Receptor Binding Domain of the SARS-CoV-2 Spike Glycoprotein. *BioRxiv*. 2020 [citado 28 de abril 2020]; doi: <https://doi.org/10.1101/2020.03.15.992883>. Disponible en: <https://www.biorxiv.org/content/10.1101/2020.03.15.992883v1.full>.
8. Grifoni A, Sidney J, Zhang Y, Scheuermann RH, Peters B, Sette A. A sequence homology and bioinformatic approach can predict candidate targets for immune responses to SARS-CoV-2. *Cell Host Microbe* [revista en internet]. 2020 [citado 28 de abril 2020]; 27(4): 671-680.e2. Disponible en: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7142693/>.
9. Jaimes JA, André NM, Chappie JS, Millet JK, Whittaker GR. Phylogenetic Analysis and Structural Modeling of SARS-CoV-2 Spike Protein Reveals an Evolutionary Distinct and Proteolytically Sensitive Activation Loop. *Journal of Molecular Biology* [revista en internet]. 2020 [citado 28 de abril 2020]; S0022-2836(20)30287-4. doi: <https://doi.org/10.1016/j.jmb.2020.04.009>. Disponible en: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7166309/>.
10. Basu A, Sarkar A, Maulik U. Strategies for vaccine design for corona virus using Immunoinformatics techniques. *BioRxiv*. 2020 [citado 28 de abril 2020]; doi: <https://doi.org/10.1101/2020.02.27.967422>. Disponible en: <https://www.biorxiv.org/content/10.1101/2020.02.27.967422v1.abstract>.
11. Lee CHJ, Koohy H. In silico identification of vaccine targets for 2019-nCoV. *F1000Research* [revista en internet]. 2020 [citado 28 de abril 2020]; 9: 145. Disponible en: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7111504.1/>.
12. Paradoa ML, Middleton D, Acosta A, Sarmiento ME, Leyva J. HLA genes in a sample of the Cuban population. *Vaccinmonitor* [revista en internet]. 2000 Sep [citado 18 de abril 2020]; 9(3): 1-5. Disponible en: [http://scielo.sld.cu/scielo.php?script=sci\\_arttext&pid=S1025-028X2000000300001&lng=es](http://scielo.sld.cu/scielo.php?script=sci_arttext&pid=S1025-028X2000000300001&lng=es).
13. Ferrer A, Nazábal M, Companioni O, de Cossío MEF, Camacho H, Cintado A, et al. HLA class I polymorphism in the Cuban population. *Human Immunology* [revista en internet]. 2007 [citado 28 de abril 2020]; 68(11): 918-927. Disponible en: <https://www.sciencedirect.com/science/article/abs/pii/S0198885907004363>.
14. Robson B. COVID-19 Coronavirus spike protein analysis for synthetic vaccines, a peptidomimetic antagonist, and therapeutic drugs, and analysis of a proposed Achilles' heel conserved region to minimize probability of escape mutations and drug resistance. *Computers in Biology and Medicine* [revista en internet]. 2020 [citado 28 de abril 2020]; 103749. doi: 10.1016/j.combiomed.2020.103749. Disponible en: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7151553/>.
15. Rasheed MA, Raza S, Zohaib A, Yaqub T, Rabbani M, Riaz MI, et al. In silico Identification of novel B Cell and T cell epitopes of Wuhan coronavirus (2019-nCoV) for effective multi epitope-based peptide vaccine production. *Preprints* [revista en internet]. 2020 [citado 28 de abril 2020]; 2020020359. doi: 10.20944/preprints202002.0359.v1. Disponible en: <https://www.preprints.org/manuscript/202002.0359/v1>



### Declaración de autoría

Orlando Rafael Serrano-Barrera

 <https://orcid.org/0000-0002-2605-6999>

Realizó la concepción y diseño del artículo. Seleccionó las bases de datos y herramientas bioinformáticas a utilizar. Realizó las modelaciones, el análisis y discusión de los resultados. Confeccionó el manuscrito.

Copyright Revista Electrónica Dr. Zoilo E. Marinello Vidaurreta. Este artículo está bajo una [licencia de Creative Commons Reconocimiento 4.0 Internacional](#), los lectores pueden realizar copias y distribución de los contenidos por cualquier medio, siempre que se mantenga el reconocimiento de sus autores.