

**IDENTIFICACIÓN DE PÉPTIDOS INMUNODOMINANTES DE LAS CELULAS TCD4  
EN LA RESPUESTA INMUNOLOGICA FRENTE AL SARS CoV 2: UNA  
APROXIMACIÓN BIOINFORMÁTICA**

Andrea Juliana Casanova Barajas

Juan Camilo Chaustre Flórez

Isabella Daza Núñez

Susana María Victoria Jiménez De León

Mauricio Alfonso Uribe Rodríguez

**Universidad El Bosque  
Facultad de Medicina  
Pregrado en Medicina  
Bogotá  
2022**

**IDENTIFICACIÓN DE PÉPTIDOS INMUNODOMINANTES DE LAS CELULAS TCD4  
EN LA RESPUESTA INMUNOLOGICA FRENTE AL SARS CoV 2: UNA  
APROXIMACIÓN BIOINFORMÁTICA**

Andrea Juliana Casanova Barajas

Juan Camilo Chaustre Flórez

Isabella Daza Nuñez

Susana María Victoria Jiménez De León

Mauricio Alfonso Uribe Rodríguez

**Director**

Miguel Hernando Parra Avila

**Asesores metodológicos**

Ricardo Duarte  
Catalina Valencia

**Trabajo de grado para optar por el título de Médico Cirujano**

**Universidad El Bosque  
Facultad de Medicina  
Pregrado en Medicina  
Bogotá  
2022**



**La Universidad El Bosque no se hace responsable de los conceptos emitidos por los investigadores en su trabajo, solo velará por el rigor científico, metodológico y ético del mismo en aras de la búsqueda de la verdad y la justicia**

## **Agradecimientos**

En primer lugar, agradecemos al Dr. Miguel Parra por aceptar ser nuestro director y ayudarnos con el trabajo, por su paciencia, colaboración y transmitirnos todo su conocimiento.

Agradecemos también a la Universidad el Bosque por la oportunidad que nos da para estudiar esta carrera profesional y las oportunidades que tiene.

Finalmente, queremos agradecer a Dios y a nuestras familias por siempre estar y apoyarnos en este sueño.

## Resumen

*Introducción* El COVID-19 puede presentarse como un síndrome respiratorio severo causado por el coronavirus de tipo 2 (SARS-CoV-2). Este virus surgió en China a finales del 2019 y se extendió rápidamente a todas las regiones del mundo, por lo cual la organización mundial de la salud (OMS) declaró la pandemia el 11 de marzo de 2020. Por lo tanto, comprender cómo funciona la memoria inmune frente al SARS-CoV-2 es fundamental para conocer la capacidad protectora del sistema inmune frente a este nuevo virus.

*Objetivo* Identificar los posibles péptidos específicos derivados del virus SARS-CoV-2 presentados por las moléculas de MHC-II a los linfocitos T CD4.

*Método* El presente estudio corresponde a un análisis bioinformático usando los algoritmos de las bases de datos Syfpeity, P9 (Tetitope) y NetMHCIIpan con las proteínas del virus SARS-CoV-2 variante de Wuhan en el formato FASTA obtenidas del IEDB para de esta manera seleccionar los posibles péptidos específicos del SARS-CoV-2 que cumplieran con los requisitos previamente establecidos.

*Resultados y conclusiones* Con el presente estudio se espera realizar la predicción de los posibles péptidos derivados de las proteínas del virus que se unen a ciertos alelos de HLA-II y que son reconocidos por los linfocitos T, siendo importante para el entendimiento de la inmunología frente al virus y el desarrollo de futuras vacunas frente a este. Se analizaron 9585 péptidos, se seleccionaron 8.918 péptidos que cumplieron con los requisitos y finalmente 25 péptidos fueron reconocidos por los tres alelos.

*Palabras clave:* epitopes, inmunodominancia, SARS-CoV-2, inmunología, linfocitos T, TCD4, COVID-19

## Abstract

*Introduction* COVID-19 can present as a severe respiratory syndrome caused by coronavirus type 2 (SARS-CoV-2). This virus emerged in China at the end of 2019 and quickly spread to all regions of the world, for which the world health organization (WHO) declared a pandemic on March 11, 2020. Therefore, understanding how the immune memory against SARS-CoV-2 is essential to know the protective capacity of the immune system against this new virus.

*Objective* Identify possible specific peptides derived from the SARS-CoV-2 virus presented by MHC-II molecules to CD4 T lymphocytes.

*Methods* The present study corresponds to a bioinformatic analysis using the databases Syfpeity, P9 (Tetitope) and NetMHCIIpan with the SARS-CoV-2 Wuhan variant proteins in FASTA format obtained from the IEDB in order to find the possible specific peptides of SARS-CoV-2 that met the previous established requirements.

*Results and conclusions* With the present study, it is expected to predict the possible peptides derived from the virus proteins that bind to certain alleles of HLA-II and that are recognized by T lymphocytes, being important for the understanding of the immunology against the virus and the development of future vaccines against it. 9585 peptides were analyzed, 8918 peptides that met the requirements were selected and finally 25 peptides were recognized by the three alleles.

*Key words:* epitopes, immunodominance, SARS-CoV-2, immunology, T lymphocytes, TCD4, COVID-19

## Tabla de contenido

<b>PREGUNTA DE INVESTIGACIÓN</b> .....	<b>8</b>
<b>PLANTEAMIENTO DEL PROBLEMA</b> .....	<b>9</b>
<b>JUSTIFICACIÓN</b> .....	<b>11</b>
<b>OBJETIVOS</b> .....	<b>12</b>
OBJETIVO GENERAL .....	12
OBJETIVOS ESPECÍFICOS.....	12
<b>MARCO TEÓRICO</b> .....	<b>13</b>
VIRUS DEL SARS-CoV-2 Y SUS PROTEÍNAS .....	13
EPIDEMIOLOGÍA, SIGNOS Y SÍNTOMAS DE LA INFECCIÓN.....	14
PAPEL DE LOS LINFOCITOS EN LA INFECCIÓN POR SARS-CoV-2 .....	15
MEMORIA INMUNOLÓGICA CONTRA EL SARS-CoV-2.....	16
PREDICCIÓN DE PÉPTIDOS USANDO ALGORITMOS BIOINFORMÁTICOS .....	17
<b>METODOLOGÍA</b> .....	<b>19</b>
1. SELECCIÓN DE SECUENCIAS DE AMINOÁCIDOS DE LAS PROTEÍNAS DEL SARS-CoV-2 .....	19
2. PREDICCIÓN DE UNIÓN DE LOS PÉPTIDOS AL HLA-DR1 CLASE I USANDO PLATAFORMAS BIOINFORMÁTICAS: .....	19
3. ANÁLISIS Y SELECCIÓN DE LAS SECUENCIAS CON MEJOR CAPACIDAD DE UNIÓN:.....	20
4. DETERMINACIÓN DE LOS PÉPTIDOS ESPECÍFICOS DE UNIÓN.....	20
<b>RESULTADOS</b> .....	<b>21</b>
ANÁLISIS DE LOS PÉPTIDOS EN LAS PLATAFORMAS BIOINFORMÁTICAS.....	21
ESTUDIOS DE SELECCIÓN Y PREDICCIÓN.....	22
PÉPTIDOS EN COMÚN DE LOS ALELOS .....	33
<b>DISCUSIÓN DE RESULTADOS</b> .....	<b>37</b>
<b>CONCLUSIONES</b> .....	<b>41</b>
<b>RECOMENDACIONES</b> .....	<b>42</b>
<b>REFERENCIAS</b> .....	<b>43</b>
<b>SIGLAS</b> .....	<b>48</b>
<b>GLOSARIO</b> .....	<b>49</b>

### **Pregunta de investigación**

¿Cuáles son los posibles péptidos específicos del SARS-CoV-2 que son presentados por las moléculas del MHC-II a los linfocitos T CD4?



## Planteamiento del problema

El COVID-19 puede presentarse como un síndrome respiratorio severo causado por el coronavirus de tipo 2 (SARS-CoV-2). Este virus surgió en China a finales del 2019 y se extendió rápidamente a todas las regiones del mundo, por lo cual la organización mundial de la salud (OMS) declaró la pandemia el 11 de marzo de 2020 (1,2). Para el 24 de marzo de 2022 según el centro de recursos de Coronavirus de Johns Hopkins se han confirmado en el mundo 476.754.225 casos de COVID-19 incluyendo 6.108.494 decesos. De estos, los más afectados son Estados Unidos (79.879.839), India (43.014.687), Brasil (29.772.873), Francia (24.823.603), Reino Unido (20.767.521), Alemania (19.265.563) y Rusia (17.433.393) (3).

En Colombia, según el ministerio de salud, para el 24 de marzo de 2022 se han reportado 6.082.226 casos confirmados los cuales 5.290 son casos activos, 5.914.578 recuperados y 139.508 fallecidos (4). Hoy en día en Colombia, los departamentos y distritos con más casos confirmados son Bogotá (1.764.880), Antioquia (917.534), Valle Del Cauca (540.099), Atlántico (401.955) y Cundinamarca (317.831). Estos casos confirmados son principalmente en pacientes en edades entre 30 a 39 años, a diferencia de los fallecidos que están en edades entre 70 a 79 años. La distribución por sexo es mayor en las mujeres con un 53.21% para los casos confirmados y para la mortalidad es mayor en el sexo masculino con un 60.77% (5).

Comprender cómo funciona la memoria inmune frente al SARS-CoV-2 es fundamental para conocer la capacidad protectora del sistema inmune frente a este nuevo virus. La inducción de la memoria inmunológica se puede generar por infección primaria o por inmunización. Una comprensión de la memoria inmunológica requiere de la evaluación de sus diferentes componentes, especialmente de las células T. Varios trabajos han mostrado que las células T CD4 están asociadas con una reducción de la enfermedad, lo que sugiere la importancia de las células T CD4 en el control y la

resolución de la infección por SARS-CoV-2 (6). Hasta el momento distintos estudios han mostrado la importancia de las células T CD4, en dichos trabajos las células TCD4 están asociadas con una reducción de la enfermedad, lo que sugiere que las respuestas de estas células pueden ser importantes para el control y la resolución de la infección por SARS-CoV-2 (6).

Por otro lado, se ha observado que muchos pacientes con COVID-19 presentan una respuesta marcada de TCD4+ y TCD8+ a diferentes proteínas virales en las cuales están incluidas la S, la de membrana (M), la nucleocápside (N), las proteínas no estructurales (NSP). Los antígenos derivados de estas proteínas pueden ser detectados hasta semanas después del inicio de los síntomas (7). Debido a esto, para poder definir la respuesta inmunológica mediada por las células T al SARS-CoV-2, es necesario responder la siguiente pregunta ¿Cuáles son los posibles péptidos específicos del SARS-CoV-2 que son presentados por las moléculas del complejo mayor de histocompatibilidad (MHC) -II a los linfocitos T CD4?

## Justificación

Las células T CD4 + son parte fundamental de la respuesta inmune para resolver infecciones virales, y en diversos reportes se ha identificado una respuesta clara de células T CD4+ dirigida a componentes estructurales y no estructurales del SARS-CoV-2 (6,7). Peng y colaboradores identificaron las regiones del virus que son reconocidas por las células T, utilizando péptidos que abarcan todas las proteínas del virus. (7,8)

Tras la infección viral por SARS-CoV-2, se activan los protocolos de procesamiento y presentación antigénica por parte de las células APC que terminan con el reconocimiento y presentación de los péptidos derivados de las proteínas del virus por medio del complejo mayor de histocompatibilidad, desencadenando la activación de los linfocitos T CD4+ (10). La identificación de los péptidos específicos derivados de las proteínas del virus SARS-CoV-2 que pueden ser reconocidos por los TCD4+ es clave para poder entender adecuadamente la respuesta inmunitaria T.

Es por esto, que el objetivo de este proyecto es realizar la predicción de los posibles péptidos derivados de las proteínas del virus que se unen a ciertos alelos de HLA-II (12) y son reconocidos por los linfocitos T, mediante la combinación de los resultados obtenidos al usar los algoritmos bioinformáticos de las plataformas SYFPEITHI, TETITOPE (P9) y NetMHCIIpan (23, 24).

## Objetivos

### *Objetivo general*

Identificar los posibles péptidos específicos derivados del virus SARS-CoV-2 presentados por las moléculas de MHC-II a los linfocitos T CD4.

### *Objetivos específicos*

- Identificar las secuencias de todas las proteínas del SARS-CoV-2.
- Determinar los péptidos específicos de la molécula MHC de clase II.

## Marco Teórico

### *Virus del SARS-CoV-2 y sus proteínas*

El SARS-CoV-2 pertenece a la familia de los Coronaviridae, estos virus se caracterizan por ser envueltos y poseer un genoma de ARN monocatenario de sentido positivo, además adoptan una morfología similar a una corona cuando se observan al microscopio electrónico y por eso reciben su nombre (13). El genoma del SARS-CoV-2 tiene información para la producción de las proteínas estructurales denominadas Spike (S), envoltura (E), proteína de membrana (M) nucleocápside (N) y un complejo de replicasa. Adicionalmente, tiene varios marcos de lectura abiertos (ORF, open reading frame, nombre en inglés), ORF1a, ORF3b, ORF6, ORF7a, ORF7b, ORF8, ORF9b, ORF9c y ORF10, que producen varias proteínas no estructurales, algunas con funciones aún desconocidas. Adicionalmente, en su estructura este genoma del virus posee dos sitios no codificantes. (UTR, untranslated region, nombre en inglés) y una cola de poli A localizados en los extremos de 5' a 3'. (13,14)

El SARS-CoV-2 utiliza como receptor celular a la enzima convertidora de angiotensina 2 (ACE2), una proteína transmembranal glicosilada que se encuentra en dos isoformas en el organismo. La proteína ACE2 se expresa en varios tejidos como el epitelio respiratorio, pulmones, riñones, intestinos, testículos (células de Sertoli y Leydig), útero, vagina, endotelio y corazón. La subunidad S1 (S1) de la proteína spike del virus se une al ACE2 de las células, acto seguido mediante la acción de la serin-proteasa TMPRSS2 del hospedador, se activa la proteína S viral permitiendo la unión de las membranas viral y celular favoreciendo la penetración del virus en la célula blanco. El proceso infeccioso del SARS-CoV-2 desencadena una compleja respuesta inmunológica, que se cree está relacionada con las diferentes manifestaciones clínicas más severas de la enfermedad. (13,15)

*Epidemiología, signos y síntomas de la infección*

Desde que se reportó la presencia del SARS-CoV-2 en China a finales del 2019, el virus ha mostrado una alta infectividad y patogenicidad. La infección por SARS-CoV-2 puede ocasionar cambios histopatológicos en los pulmones de los pacientes al unirse a las células epiteliales del tracto respiratorio (13), provocando una lesión tisular en las vías respiratorias debido a un proceso inflamatorio severo, alteraciones en la microcirculación, entre otras diversas complicaciones. Dentro de las manifestaciones clínicas encontradas en los contagiados por este virus tenemos entre las más comunes: fiebre, tos, disnea y mialgias, dentro de las menos comunes encontramos: confusión, cefalea, odinofagia, rinorrea, dolor torácico, diarrea, náuseas y emesis. Estos pacientes también pueden cursar con neumonía, linfocitopenia, síndrome de dificultad respiratoria aguda (SDRA) o también llamado síndrome de distrés respiratorio agudo y algunos pueden terminar en insuficiencia orgánica múltiple (16). La infección en personas mayores con comorbilidades tales como diabetes mellitus, enfermedad renal, hipertensión arterial, asma, enfermedad pulmonar obstructiva crónica (EPOC), enfermedad pulmonar intersticial, fibrosis quística, hipertensión pulmonar, insuficiencia cardíaca, enfermedad coronaria, miocardiopatías, sobrepeso u obesidad, enfermedad hepática, embarazo, enfermedad de células falciformes, entre otras, tienen un mayor riesgo de sufrir una enfermedad grave que cursa con insuficiencia respiratoria y puede llevar al paciente a la muerte (17); mientras que en niños y adultos jóvenes la infección cursa generalmente de forma asintomática. La mortalidad observada por la infección del SARS-CoV-2, especialmente en las poblaciones susceptibles, llevó a la Organización Mundial de la Salud (OMS) a declarar un estado de emergencia de salud pública de carácter internacional (13). Hoy en día se han reportado 476.754.225 casos a nivel mundial, siendo Estados Unidos (79.879.839), India (43.014.687), Brasil (29.772.873), Francia (24.823.603) y Reino Unido (20.767.521) los países con mayor número de casos. Colombia ocupa el doceavo puesto en números de casos con un total de 6.082.226, siendo activos un total de 5.290 y

139.508 muertes. Bogotá tiene el mayor número de casos acumulados con 1.764.880, seguido por el departamento de Antioquia con 917.534 casos (3,4).

### *Papel de los linfocitos en la infección por SARS-CoV-2*

Estudios realizados en pacientes que han experimentado la infección por SARS-CoV-2 en diferentes escalas de severidad, han mostrado que los pacientes que sufren infecciones graves muestran una linfocitopenia más marcada que aquellos que presentan infecciones con signos clínicos moderados. Las células que más se ven afectadas por este fenómeno son las poblaciones TH1, Treg y en especial las células T CD8, que además muestran un perfil de agotamiento (18). Por el contrario, a pesar de que hay una disminución en el número de T CD4, estas células muestran un perfil de activación en los pacientes evaluados, hecho que soporta la hipótesis de que la respuesta inmune contra SARS-CoV-2 es dependiente de linfocitos T CD4, células que fomentan la producción de anticuerpos por parte de los linfocitos B y la actividad citotóxica de los T CD8, estas últimas encargadas del aclaramiento viral (19).

La falta de una respuesta celular eficiente conduce a una incapacidad para eliminar al virus y a las células infectadas, prolongando de forma patológica el proceso inflamatorio que lleva a los pacientes a desarrollar una tormenta de citocinas que se manifiesta principalmente por el desarrollo de Síndrome de Dificultad Respiratoria Aguda (SDRA) y algunas consecuencias sistémicas, como la coagulación intravascular diseminada (18). En casos graves de infección por SARS-CoV-2 se ha observado un incremento significativo de IL-2, IL-2R, IL-6, IL-7, IL-8 IL-10, IP10, MIP1A y TNF-alfa, en comparación con pacientes que sufren infecciones leves o moderadas (21).

A pesar de que aún no es claro el papel que desempeñan los linfocitos T CD4 en la infección por SARS-CoV-2, se ha observado que en casos de infecciones graves la inmunidad mediada por los linfocitos T se ve disminuida de forma significativa, de hecho, se ha determinado una correlación

negativa entre la funcionalidad de los linfocitos T CD4 y la severidad de la enfermedad (20). Algunos estudios señalan que en pacientes que sobrevivieron a la infección, la funcionalidad de los linfocitos T CD4 se incrementa de manera significativa durante y después de la infección (21). En pacientes asintomáticos se evidenció un incremento en la proliferación de los linfocitos TCD4 (21); a pesar de que los linfocitos T CD8 son importantes en el aclaramiento viral, se ha observado una predominancia de la función de los linfocitos TCD4+ sobre los TCD8+ ante la respuesta al virus SARS-CoV-2 (22).

### *Memoria inmunológica contra el SARS-CoV-2*

Cómo y por cuánto tiempo se puede mantener la memoria inmunológica contra el SARS-CoV-2, sigue siendo una de las preguntas a resolver para poder entender el comportamiento de este virus. De acuerdo con un estudio publicados por Wang et al (2021), es posible detectar células de memoria contra el virus hasta 3-5 meses después de la infección (22). En dicho estudio (22), se observó que la memoria inmune estaba dirigida principalmente contra la glicoproteína Spike del SARS-CoV-2, ya que los pacientes recuperados de COVID 19 mostraron una activación mucho eficiente de los linfocitos TCD4+ frente a péptidos derivados de la proteína S en comparación con los pacientes no expuestos al virus. Se encontró además la presencia de anticuerpos neutralizantes contra el virus en pacientes después de 4 meses post infección, en comparación con los de los pacientes no expuestos al virus que mostraron anticuerpos con una baja capacidad de neutralización (22). Incluso al analizar los contactos cercanos de pacientes infectados, se observó una respuesta células significativa de las células T a pesar de que esos individuos no habían presentado signos de infección, ni resultados de pruebas moleculares positivas (22).

Debido a que también es posible encontrar una memoria inmune a infecciones previas por otras cepas de coronavirus, también se ha evaluado el papel de estas células en el desarrollo de la infección por SARS-CoV-2. Los resultados de Wang et Al (2021), mostraron que algunos pacientes no infectados con respuestas de memoria contra otros coronavirus humanos tenían células reactivas



contra péptidos derivados de las proteínas de SARS-CoV-2, especialmente contra antígenos diferentes a la proteína S (22). Estos resultados muestran que se debe hacer un análisis más profundo sobre la presencia de estas células en el desarrollo de la infección por SARS-CoV-2.

### *Predicción de péptidos usando algoritmos bioinformáticos*

Para estudiar la respuesta de las células T es crucial hacer la identificación de los posibles péptidos presentados por las diferentes moléculas del HLA. Diferentes aproximaciones bioinformáticas han sido desarrolladas durante los últimos años con el ánimo de reconocer estos epítopes y disminuir, a números razonables, la cantidad de epítopes posibles en los ensayos funcionales (23). Para identificar estos péptidos se han usado dos estrategias, la primera es evaluar la unión de los péptidos a las moléculas del HLA y la segunda identificar la posibilidad de que esos péptidos sean procesados y presentados en el contexto de infecciones naturales (24,25). Recientemente otras aproximaciones que usan megadatos y aplican el uso de inteligencia artificial también se han incorporado en este tipo de estrategias (25). Usando la propuesta hecha por Calvo-Calle, se combinan tres estrategias para determinar los péptidos por estas herramientas, la matriz de Syfpeithi para la predicción del procesamiento y presentación de los péptidos, la matriz de P9 que evalúa la unión al péptido a las moléculas del HLA, y la matriz NetMHCIIpan que trabaja con inteligencia artificial (25,26).

Syfpeithi es una base de datos que contiene una colección de ligandos de MHC clase I y clase II y patrones peptídicos tanto de humanos como de otras especies, de esta manera, funciona como un algoritmo basado en el análisis de la base de datos de secuencias peptídicas que fueron procesadas naturalmente y se asocian a las proteínas MHC particulares, por consiguiente, puede predecir el procesamiento y la presentación de antígenos (26, 27). P9 es un algoritmo que utiliza resultados de ensayos cualitativos de unión de péptidos para poder predecir las afinidades relativas de unión de MHC-péptido. Hay varios bolsillos encargados de la especificidad del alelo a la interacción HLA-ligando, los cuales van a estar formados por residuos polimórficos de la cadena de la molécula de

HLA (25, 26). NetMHCIIpan predice la unión de péptidos de cualquier longitud a cualquier molécula MHC II de secuencia conocida utilizando redes neuronales artificiales. Contiene un gran conjunto de datos de más de 500.000 mediciones de afinidad de ligandos y espectrometría de masas de ligando eluido, que cubre los tres isotipos humanos de clase II del MHC HLA-DR, HLA-DQ, HLA-DP, así como las moléculas de ratón [H-2](28).

## Metodología

Tipo de estudio: búsqueda sistemática de proteínas en bases bioinformáticas

### *1. Selección de secuencias de aminoácidos de las proteínas del SARS-CoV-2*

El primer paso para el desarrollo de este trabajo es la obtención de las secuencias de aminoácidos que componen las proteínas estructurales y no estructurales del SARS-CoV-2, para predecir los epítopes presentados por las moléculas de HLA-DR1 clase II (DRB1\*01:01, DRB1\*01:04, DRB1\*01:07). Se consultaron las secuencias de las proteínas en la base de datos del NCBI (<https://www.ncbi.nlm.nih.gov/guide/dna-rna/>) y finalmente se seleccionaron las secuencias en formato FASTA de las proteínas E (YP\_009724392.1), M(YP\_009724393.1), N(YP\_009724397.2), S(YP\_009724390.1), ORF1ab (YP\_009724389.1), ORF3a(YP\_009724391), ORF6(YP\_009724394.1), ORF7a (YP\_009724395), ORF8 (YP\_009724396.1), ORF 10 (YP\_009725255.1), de acuerdo a lo reportado por Griffoni (12) TABLA 1.

### *2. Predicción de unión de los péptidos al HLA-DR1 clase I usando plataformas bioinformáticas:*

Las secuencias de las proteínas del virus se analizaron en 3 plataformas bioinformáticas: Syfpeithi ([www.syfpeithi.de](http://www.syfpeithi.de)), P9 ([www.iedb.org](http://www.iedb.org)) y NetMHCIIpan 4.0 BA ([www.iedb.org](http://www.iedb.org)). En cada una de las bases de datos se calificaron las secuencias de los posibles péptidos derivados de las proteínas (En secuencias de 15 kmers sobrelapadas en 14 aminoácidos) para cada uno de los alelos del HLA-DR1 clase II (DRB1\*01:01, DRB1\*01:04, DRB1\*01:07). Los valores numéricos obtenidos en cada una de las 3 plataformas bioinformáticas fueron tabulados de forma individual y luego analizados en conjunto.

### *3. Análisis y selección de las secuencias con mejor capacidad de unión:*

Cada una de las secuencias de 15 kmers (de las proteínas señaladas) fue valorada por las tres plataformas bioinformáticas. Para seleccionar las secuencias con mejor capacidad de unión a las moléculas HLA-DR1 clase II, se usaron los parámetros descritos por Calvo-Calle (25). En resumen, se seleccionaron aquellos péptidos que cumplían las siguientes parámetros:  $>-1$  en P9;  $>18$  en SYFPEITHI (26,27) y en NetMHCIIpan un  $IC50 <50$  (26, 27). Las secuencias que fueron finalmente seleccionadas fueron las que obtuvieron el puntaje requerido en las 3 plataformas para cada alelo.

### *4. Determinación de los péptidos específicos de unión.*

Luego de obtener las secuencias que cumplen con los puntajes requeridos en Syfpeithi ([www.syfpeithi.de](http://www.syfpeithi.de)), P9 ([www.iedb.org](http://www.iedb.org)) y NetMHCIIpan 4.0 BA ([www.iedb.org](http://www.iedb.org)) para cada alelo, se seleccionaron cuáles son los péptidos que se unen a los tres tipos de alelos del HLA-DR1 clase II (DRB1\*01:01, DRB1\*01:04, DRB1\*01:07), siendo estos alelos escogidos debido a que son los que más comúnmente se encuentran en la población colombiana (29, 30).

## Resultados

### *Análisis de los péptidos en las plataformas bioinformáticas*

Las proteínas seleccionadas fueron ingresadas en las plataformas Syfpeithi, NetMHCIIpan y P9 para analizar los péptidos. De todas las proteínas se evaluaron 9585 péptidos distintos distribuidos de la siguiente manera: para la proteína S se analizaron 1259 péptidos, para la proteína E se analizaron 61 péptidos, la proteína M se analizaron 208 péptidos, para la proteína N se analizaron 405 péptidos. En las proteínas no estructurales se obtuvo, para la proteína ORF1ab se analizaron 7082 péptidos, para la proteína ORF3a se analizaron 261 péptidos, para la proteína ORF6 se analizaron 47 péptidos, para la ORF7a se analizaron 107 péptidos, para la ORF8 107 107 péptidos y finalmente para la ORF10 se analizaron 24 péptidos, todos estos péptidos estaban formados por en 15 aminoácidos sobrelapados en 14. (Tabla 1).

PROTEÍNA	TAMAÑO aa	CÓDIGO IEDB	FUNCIÓN	PÉPTIDOS ANALIZADOS
<b>E</b>	75	YP_009724392.1	Estructural de Envoltura	<b>61</b>
<b>M</b>	222	YP_009724393.1	Estructural de membrana	<b>208</b>
<b>N</b>	419	YP_009724397.2	Estructural de nucleocápside	<b>405</b>
<b>ORF1ab</b>	7096	YP_009724389.1	No estructural	<b>7082</b>
<b>ORF3a</b>	275	YP_009724391	No estructural	<b>261</b>
<b>ORF6</b>	61	YP_009724394.1	No estructural	<b>47</b>
<b>ORF7a</b>	121	YP_009724395	No estructural	<b>107</b>
<b>ORF8</b>	121	YP_009724396.1	No estructural	<b>107</b>
<b>ORF10</b>	38	YP_009725255.1	No estructural	<b>24</b>
<b>TOTAL</b>				<b>9585</b>

Tabla 1. Proteínas con total de péptidos analizados. (Creado por grupo de investigación)

*Estudios de selección y predicción*

Siguiendo los criterios de selección previamente descritos de acuerdo con los trabajos de Calvo-Calle (25), se seleccionaron péptidos que cumplieran estos requisitos, primero se hicieron con las proteínas estructurales y posteriormente con las no estructurales.

Para la Proteína S se identificaron para el DRB1\*0101 37 péptidos, para el DRB1\*0401 se seleccionaron 5 péptidos y para el DRB1\*0701 16 péptidos cumplieron con los requisitos. Para la Proteína E, para el DRB1\*0101 salieron 8 péptidos, mientras que para el DRB1\*0401 se identificaron 2 péptidos y para el DRB1\*0701 solamente salió 1 péptido. Para la Proteína M se seleccionaron 13 péptidos para el DRB1\*0101, 2 péptidos para el DRB1\*0401 y 6 péptidos para el DRB1\*0701. Para la última proteína estructural, la Proteína N, salieron 6 péptidos para el DRB1\*0101 y 4 péptidos para el DRB1\*0701, mientras que para el DRB1\*0401 no se identificaron péptidos que cumplieran con los requisitos previamente establecidos.

Para las proteínas no estructurales empezando por la Proteína ORF 1 ab, para el DRB1\*0101 se se seleccionaron 58 péptidos, 42 péptidos para el DRB1\*0401 y 123 péptidos se seleccionaron para el DRB1\*0701. En la Proteína ORF3a, para el DRB1\*0101 se identificaron 16 péptidos y para el DRB1\*0701 7 péptidos, no se encontraron péptidos que cumplieran con los requisitos pre establecidos para el DRB1\*0401. Siguiendo con el análisis del ORF6, para el DRB1\*0101 salieron 6 péptidos, 3 péptidos para el DRB1\*0401 y otros 3 para el DRB1\*0701. Proteína ORF7a para el DRB1\*0101 salieron 3 péptidos, y no se identificaron péptidos que cumplieran con los requisitos ni para el DRB1\*0401 ni para el DRB1\*0701. De igual manera, para la proteína ORF8 solamente se identificaron 6 péptidos para el DRB1\*0101, pero no se identificaron péptidos que cumplieran los requisitos para el DRB1\*0401 ni para DRB1\*0701. Tampoco se seleccionaron péptidos derivados del

ORF 10 que cumplieran los requisitos previamente descritos para los alelos DRB1\*0401 y DRB1\*0701. Solamente se identificaron 2 péptidos para el alelo DRB1\*0101. (Tabla 2)

PROTEINA S						
POS	SECUENCIA	CORE PEPTIDE P9	CORE PEPTIDE NETPAN	SYFPEITHI	P9	IC50
<b>PREDICCIÓN DE EPÍTOPES DE ALTA UNIÓN HLA-DRB1*0101</b>						
512	VLSFELLHAPATVCG	FELLHAPAT	FELLHAPAT	27	1.87	5
895	QIPFAMQMAYRFNGI	FAMQMAYRF	FAMQMAYRF	33	1.60	6
540	NFNFNGLTGTGVLTE	FNGLTGTGV	FNGLTGTGV	38	1.99	6
855	FNGLTVLPPLLTDEM	LTVLPPLLT	LTVLPPLLT	23	3.000	10
854	KFNGLTVLPPLLTDE	LTVLPPLLT	LTVLPPLLT	20	3.000	11
235	ITRFQTLALHRSYL	FQTLALHR	FQTLALHR	26	1.40	11
852	AQKFNGLTVLPPLLT	LTVLPPLLT	FNGLTVLPP	27	3.000	12
236	TRFQTLALHRSYLT	FQTLALHR	FQTLALHR	18	1.40	13
314	QTSNFRVQPTESIVR	FRVQPTESI	FRVQPTESI	21	1.70	13
362	VADYSVLYNSASFST	VLYNSASF	YSVLYNSAS	26	3.400	14
264	AYYVGYLQPRTFLLK	VGYLQPRTF	YLQPRTFLL	31	6.900	15
885	GWTFGAGAALQIPFA	FGAGAALQI	FGAGAALQI	26	8.000	16
115	QSLLVNNTATNVVIK	LIVNNTATNV	LIVNNTATNV	25	01.03	21
1057	PHGVVFLHVTVVPAQ	VVFLHVTVV	VVFLHVTVV	27	02.09	21
364	DYSVLYNSASFSTFK	VLYNSASF	VLYNSASF	24	3.400	22
238	FQTLALHRSYLTTPG	FQTLALHR	LALHRSYLT	24	1.40	23
1005	QTYVTQQLIRAAEIR	YVTQQLIRA	YVTQQLIRA	24	8.000	24
883	TSGWTFGAGAALQIP	FGAGAALQI	WTFGAGAAL	28	8.000	24
960	NLTVKQLSSNFGAIS	LVKQLSSNF	VKQLSSNFG	23	0	24
1004	LQTYVTQQLIRAAEI	YVTQQLIRA	YVTQQLIRA	25	8.000	25
1058	HGVVFLHVTVVPAQE	VVFLHVTVV	VVFLHVTVV	23	02.09	25
116	SLLVNNTATNVVIKV	LIVNNTATNV	LIVNNTATNV	23	01.03	25
959	LNTLVKQLSSNFGAI	LVKQLSSNF	VKQLSSNFG	31	0	27
1007	YVTQQLIRAAEIRAS	YVTQQLIRA	QQLIRAAEI	23	8.000	27
486	FNCYFPLQSYGFQPT	YFPLQSYGF	YFPLQSYGF	34	1.50	29
4	FLVLLPLVSSQCVNL	FLVLLPLVS	LLPLVSSQC	22	2.20	29
1059	GVVFLHVTVVPAQEK	VVFLHVTVV	FLHVTVVPA	24	02.09	30
198	DGYFKIYSKHTPINL	YFKIYSKHT	YFKIYSKHT	22	3.000	31
715	PTNFTISVTTEILPV	FTISVTTEI	FTISVTTEI	29	3.900	35
5	LVLLPLVSSQCVNLT	LVSSQCVNL	LLPLVSSQC	29	1.80	38
2	FVFLVLLPLVSSQCV	FVFLVLLPL	LLPLVSSQC	26	2.60	39
1044	GKGYHLMSFPQSAPH	YHLMSFPQS	YHLMSFPQS	25	2.700	42
716	TNFTISVTTEILPVS	FTISVTTEI	FTISVTTEI	20	3.900	43
948	LQDVVNQNAQALNTL	VVNQNAQAL	VVNQNAQAL	31	2.40	43
183	QGNFKNLREFVFKNI	FKNLREFVF	FKNLREFVF	27	0	46
492	LQSYGFQPTNGVGYQ	YGFQPTNGV	YGFQPTNGV	33	6.000	46
450	NYLYRFLFRKSNLKP	YRFLFRKSNL	YRFLFRKSNL	21	1.58	47

PREDICCIÓN DE EPÍTOPES DE ALTA UNIÓN HLA-DRB1*0401						
564	QFGRDIADTTDAVRD	FGRDIADTT	DIADTTDAV	18	1.10	1.1
1015	AAEIRASANLAATKM	IRASANLAA	IRASANLAA	20	4.00	36
116	SLLIVNNTATNVVIVK	IVNNTATNVV	IVNNTATNVV	20	4.30	41
362	VADYSVLYNSASFST	YSVLYNSAS	YSVLYNSAS	28	2.80	44
114	TQSLIVNNTATNVVI	IVNNTATNVV	LLIVNNTATN	26	4.30	46

PREDICCIÓN DE EPÍTOPES DE ALTA UNIÓN HLA-DRB1*0701						
632	TWRVYSTGNSVVFQTR	VYSTGNSVVF	VYSTGNSVVF	30	5.60	19
201	FKIYSKHTPINLVRD	YSKHTPINL	YSKHTPINL	18	5.10	24
339	GEVFNATRFASVYAW	FNATRFASV	FNATRFASV	26	3.40	26
893	ALQIPFAMQMAYRFN	LQIPFAMQM	FAMQMAYRF	20	4.30	27
630	TPTWRVYSTGNSVVFQ	VYSTGNSVVF	VYSTGNSVVF	18	5.60	28
309	EKGIYQTSNFRVQPT	IYQTSNFRV	IYQTSNFRV	30	4.80	29
690	QSIIAYTMSLGAENS	IAYTMSLGA	IAYTMSLGA	28	4.00	30
717	NFTISVTTEILPVSM	FTISVTTEI	FTISVTTEI	22	7.22	32
266	YVGYLQPRTFLLKYN	YLQPRTFLL	YLQPRTFLL	26	5.22	36
639	GSNVFQTRAGCLIGA	VFQTRAGCL	VFQTRAGCL	24	4.60	38
450	NYLYRFRKSNLKP	YRFRKSNL	YRFRKSNL	18	6.10	41
688	ASQSIIAYTMSLGAE	IAYTMSLGA	IAYTMSLGA	22	4.00	42
640	SNVVFQTRAGCLIGAE	VFQTRAGCL	VFQTRAGCL	18	4.60	46
116	SLLIVNNTATNVVIVK	VNNTATNVVI	IVNNTATNVV	24	4.70	46
106	FGTTLDSKTQSLIV	FGTTLDSKT	DSKTQSLI	22	2.000	49
365	YSVLYNSASFSTFKC	LYNSASFST	LYNSASFST	22	5.00	51

PROTEINA E

POS	SECUENCIA	CORE PEPTIDE P9	CORE PEPTIDE NETPAN	SYFPEITHI	P9	IC50
PREDICCIÓN DE EPÍTOPES DE ALTA UNIÓN HLA-DRB1*0101						
59	YSRVKNLNSSRVPDL	VKNLNSSRV	VKNLNSSRV	31	1.00	14
54	PSFYVYSRVKNLNS	YVYSRVKNL	YVYSRVKNL	20	1.40	14
56	FVYVYSRVKNLNSRV	YVYSRVKNL	YSRVKNLNS	24	1.40	14
1	MYSFVSEETGTLIVN	FVSEETGTL	FVSEETGTL	26	1.000	25
28	LVTLAILTALRLCAY	LAILTALRL	LAILTALRL	35	2.20	27
2	YSFVSEETGTLIVNS	FVSEETGTL	FVSEETGTL	25	1.000	36
25	VFLLVTLAILTALRL	LAILTALRL	LAILTALRL	25	2.20	44
26	FLLVTLAILTALRLC	LAILTALRL	LAILTALRL	19	2.20	48

PREDICCIÓN DE EPÍTOPES DE ALTA UNIÓN HLA-DRB1*0401						
7	EETGTLIVNSVLLFL	LIVNSVLLF	LIVNSVLLF	18	2.10	1.1
56	FVYVYSRVKNLNSRV	YVYSRVKNL	YSRVKNLNS	28	3.30	19

PREDICCIÓN DE EPÍTOPES DE ALTA UNIÓN HLA-DRB1*0701						
54	PSFYVYSRVKNLNS	YVYSRVKNL	YVYSRVKNL	24	8,4	16.01

PROTEINA M

POS	SECUENCIA	CORE PEPTIDE P9	CORE PEPTIDE NETPAN	SYFPEITHI	P9	IC50
-----	-----------	-----------------	---------------------	-----------	----	------



PREDICCIÓN DE EPÍTOPES DE ALTA UNIÓN HLA-DRB1*0101						
175	TLSYYKLGASQRVAG	YKLGASQRV	YKLGASQRV	29	1	4,4
176	LSYYKLGASQRVAGD	YKLGASQRV	YKLGASQRV	26	1	4,9
70	VYRINWITGGIAIAM	YRINWITGG	INWITGGIA	31	1	13
71	YRINWITGGIAIAMA	YRINWITGG	INWITGGIA	20	1	14
100	FRLFARTRSMWSFNP	FRLFARTRS	FARTRSMWS	24	1	20
34	LLQFAYANRNRFYI	LQFAYANRN	LQFAYANRN	18	10.000	20
109	MWSFNPETNILLNVP	FNPETNILL	FNPETNILL	19	7.000	22
68	AAVYRINWITGGIAI	YRINWITGG	INWITGGIA	18	1	22
97	IASFRLFARTRSMWS	FRLFARTRS	FRLFARTRS	22	1	36
118	ILLNVPLHGTILTRP	ILLNVPLHG	NVPLHGTIL	26	3.400	37
29	LTWICLLQFAYANRN	LQFAYANRN	LQFAYANRN	25	10.000	37
107	RSMWSFNPETNILLN	FNPETNILL	FNPETNILL	20	7.000	37
62	LACFVLAAYRINWI	FVLAAYVRI	FVLAAYVRI	20	3	50

PREDICCIÓN DE EPÍTOPES DE ALTA UNIÓN HLA-DRB1*0401						
100	FRLFARTRSMWSFNP	FARTRSMWS	FARTRSMWS	22	4	24
98	ASFRLFARTRSMWSF	FARTRSMWS	FARTRSMWS	18	4	38

PREDICCIÓN DE EPÍTOPES DE ALTA UNIÓN HLA-DRB1*0701						
100	FRLFARTRSMWSFNP	FARTRSMWS	FARTRSMWS	26	4,2	16
98	ASFRLFARTRSMWSF	FARTRSMWS	FARTRSMWS	22	4,2	18
176	LSYYKLGASQRVAGD	YKLGASQRV	YKLGASQRV	26	4,7	29
167	EITVATSRTLSSYYKL	ITVATSRTL	VATSRTLSSY	28	4,2	33
34	LLQFAYANRNRFYI	FAYANRNRF	FAYANRNRF	22	3,2	49

PROTEINA N						
POS	SECUENCIA	CORE PEPTIDE P9	CORE PEPTIDE NETPAN	SYFPEITHI	P9	IC50
PREDICCIÓN DE EPÍTOPES DE ALTA UNIÓN HLA-DRB1*0101						
389	QQTVTLLPAADLDDF	VTLLPAADL	VTLLPAADL	34	1.40	14
107	RWYFYLLGTGPEAGL	YYLGTGPEA	YYLGTGPEA	25	9.700	16
108	WYFYLLGTGPEAGLP	YYLGTGPEA	YYLGTGPEA	27	9.700	18
154	NAAIVLQLPQGTTLTLP	IVLQLPQGT	IVLQLPQGT	24	2.00	23
106	PRWYFYLLGTGPEAG	YYLGTGPEA	YYLGTGPEA	29	9.700	24
156	AIVLQLPQGTTLPKG	IVLQLPQGT	LQLPQGTTL	26	2.00	40
PREDICCIÓN DE EPÍTOPES DE ALTA UNIÓN HLA-DRB1*0701						
326	PSGTWLTGTGAIKLD	LTYTGAIKL	LTYTGAIKL	24	7.00	9,3
328	GTWLTGTGAIKLDDK	LTYTGAIKL	LTYTGAIKL	24	7.00	12
304	IAQFAPSASAFFGMS	FAPSASAFF	FAPSASAFF	30	4.50	16
302	PQIAQFAPSASAFFG	FAPSASAFF	FAPSASAFF	18	4.50	19

PROTEINA ORF 1AB						
POS	SECUENCIA	CORE PEPTIDE P9	CORE PEPTIDE NETPAN	SYFPEITHI	P9	IC50
PREDICCIÓN DE EPÍTOPES DE ALTA UNIÓN HLA-DRB1*0101						
2210	SKLINIIWFLLSV	FNYLKSPNF	FYILPSIIS	26	2	3,8

6998	MILSLLSKGRLI IRE	FLIGCNYLG	LRIMASLVL	18	4.000	4
2908	TNVLEGSVAYESLRP	FATSACVLA	VTALRANSA	24	0	4,8
2228	SLIYSTAALGVMSN	WFLLSVCL	FYILPSIIS	18	3	5,3
4827	DKYFDCYDGGCINAN	FFAQDGNAA	YLKLRSDVL	26	7.000	6,4
2561	KMFDAYVNTFSSTFN	YSQLMCQPI	FKLSYGIAT	20	2	7,1
4845	VNLDKSAGFPFNKW	YRYNLPTMC	YLKLRSDVL	26	3	7,3
5673	LPAPRTLLTKGTLEP	YVFCTVNAL	FLVQAGNVQ	19	2	9,5
5679	LLTKGTLEPEYFNSV	YVFCTVNAL	FLVQAGNVQ	23	2	9,8
2567	VNTFSSTFNVPMEKL	ILLDDQALV	YGIATVREV	18	4.000	10
3508	QNGMNGRTILGSALL	VDILGPLSA	LYFIKGLNN	19	2.000	10
4769	FYDFAVSKGFFKEGS	MHAASGNLL	YNSQNAVAS	18	1.000	10
3777	EFRYMNSQGLLPKKN	IMLVYCFLG	YRSLPGVFC	26	5.500	11
6118	VHGNAHVASCDAIMT	FVKIGPERT	FDVLKSEDA	20	2	14
4143	LSLDQLDKWARFPKS	LRQMCAAG	YSHVVAFNT	24	1	14
263	FDTFNGECPNFVFPL	FNGECPNFV	FNICQAVTA	25	8.000	15
1408	FYFYTSKTTVASLIN	IKIQEGVVD	LRVSSSKL	25	2.000	16
2234	AALGVMSNLGMPSY	LLLSVCLGS	FYILPSIIS	18	3.000	16
3580	ENAFLPFAMGIIAMS	LTILTSLLV	VLGSLAATV	18	1	17
3077	YSVIYLYLTFYLTND	LLFLMSFTV	VMPLSAPTL	23	2	17
297	MGRIRSVYPVASPNE	IRSVYPVAS	YVYANGGKG	30	2	17
2299	SFKWDLTAFGLVAEW	YPSLETIQI	YHNESGLKT	19	7.000	20
3565	VLVQSTQWSLFFFLY	FQSAVKRTI	MAYITGGVV	20	1.900	20
3083	YLTFYLTNDVSFLAH	LLFLMSFTV	VMPLSAPTL	19	2	20
2864	VGNICYTPSKLIEYT	VVPGLPGTI	ARYMRSLKV	23	1.000	21
5556	KSHFAIGLALYPSA	VMPLSAPTL	YVMHANYIF	21	2	21
5322	GCDVTDVTLQYLGGM	LQAVGACVL	LYKLMGHFA	23	6.500	21
6386	NVAFNVNKGHFDGQ	YVPLKSATC	LSAQTGIIV	25	3	21
2346	SNSWLMWLIINLVQM	MQLFFSYFA	YTKVDGVDV	19	1	22
3903	ASEFSSLPYAAFAT	FEKMSVLLS	IQPIGALDI	32	3.000	22
6522	GVLITEGSVKGLQPS	ILNNLGVDI	FKEGSSVEL	20	2	23
556	QNSVRVLQKAAITIL	VRVLQKAAI	IISVTSNYS	26	1	25
1335	CKSAFYILPSIISNE	YTVVEAKTV	MEKCLKLVA	20	2.000	27
3636	KLKDCVMYASAVVLL	LFLPLSLAT	LHLLIGLAK	22	1	27
4227	FCAFAVDAKAYKDY	LYFIKGLNN	FTVLCCLTPV	19	0	28
2352	WLIINLVQMAPISAM	VHFISNSWL	YTKVDGVDV	24	3.000	28
3646	AVVLLIILMTARTVYD	MVYMPASWV	LQAGNATEV	25	3	29
3553	HWLLLTILTSLLVLV	VVRQCSGVT	MAYITGGVV	28	0	29
3104	FTPLVPFWITIAAYII	YLYLTFYLT	FIETISLAG	18	9.000	29
2807	DTWFSQRGGSYTNDK	YKAIDGGVT	YRVTKNSKV	19	0	32
3220	VLYQPPQTSITSAVL	YREAAACHL	FVLALLSDL	20	2	32
575	QYSLRLIDAMMFTSD	LRLIDAMMF	YAYLRKHFS	33	2	33
5542	MQKYSTLQGGPPTGK	YFVLTSHTV	VYANLGERV	26	2	33
3604	TVAYFNMVYMPASWV	FLYENAFLP	LGSLAATVR	24	6.000	33
736	PKEIIFLEGETLPTE	IIFLEGETL	IFFITGNTL	34	1	33

3867	EKMVSLLSVLLSMQG	VVLLSVLQQ	LTNMFPLI	25	6.000	34
3583	FLPFAMGIIAMSAFA	LTILTSLLV	YITGGVVQL	19	1	34
5387	DYILANTCTERLKLKLF	YLGGMSSYYC	ITGLHPTQA	22	4.000	34
3223	QPPQTSITSAVLQSG	YREAACCHL	VRSIFSRTL	21	2	36
4584	VDSYYSLLMPILTLT	MRNAGIVGV	LTLAVPYNM	26	8.000	37
3873	LSVLLSMQGAVDINK	LRVSSSKL	FTPLIQPIG	22	1	38
4180	GPKVKYLYFIKGLNN	FVLALLSDL	FCDLKGKYV	18	1	40
3652	LMTARTVYDDGARRV	MVYMPASWV	LQAGNATEV	18	3	42
737	KEIIFLEGETLPTEV	IIFLEGETL	VASIKNFKS	19	1	44
3677	YKVYYGNALDQAISM	VMYASAVVL	IIPARARVE	18	2	45
5622	QYVFCTVNALPETTA	IVYTACSHA	WRNTNPIQL	24	7.000	47
1347	SNEKQEILGTVSWNL	FYILPSIIS	LKTLVATAE	23	3	49
4977	PNMLRIMASLVLARK	VVIGTSKIFY	YKYFSGAMD	33	2.000	49
<b>PREDICCIÓN DE EPÍTOPES DE ALTA UNIÓN HLA-DRB1*0401</b>						
6215	AYKIEELFYSYATHS	WTIEYPIIG	AGDTRTITV	20	2	1
2369	YVHVVDGCNSSTCMM	LVQMAPIA	WTNAGDYIL	20	3	1,1
3638	KDCVMYASAVVLLIL	LFLPLSLAT	YGQQFGPTY	20	2	1,1
1930	LTGYKKPASRELKVT	FVCDNIKFA	YEQAVANGD	22	3	1,3
2593	AKNVSLDNVLSSTFIS	YVNTFSSTF	VVGESEGL	20	3	1,7
3557	LTILTSLLVLVQSTQ	VVRQCSGVT	VDALCEKAL	20	1	1,8
5627	TVNALPETTADIVVF	IVYTACSHA	LISMMGFKM	18	2	14
2210	SKLINIIIWFLLSV	FNYLKSPNF	FYILPSIIS	20	1,7	15
2427	CDTFCAGSTFISDEV	VRRSFYVYA	FYLITPVHV	22	1,28	16
6886	QQKALGGSVAIKIT	WLPTGTLV	LKSIAATRG	20	3,9	20
1054	GGGVAGALNKATNNA	VVNAANVY	FRYMNSQGL	26	2,1	20
274	VFPLNSIIKTIQPRV	VFPLNSIIK	LILMTARTV	20	1,68	24
3586	FAMGIIAMSAFAMMF	LTSLLVLVQ	VLGSLAATV	18	2,1	24
2459	SSYIVDSVTVKNGSI	FISDEVARD	LKLFAAETL	20	1,9	25
2228	SLIYSTAALGVLSMN	LLLSVCLGS	FYILPSIIS	22	3	26
5673	LPAPRTLLTKGTLEP	YVFCTVNAL	LVQAGNVQL	18	3	27
5777	AVFISPYNQNAVAS	LVYDNKLKA	IRHVRWIG	20	2	27
6828	LNDFVSDADSTLIGD	YGDSATLPK	LKYAISAKN	22	2	29
2561	KMFDAYVNTFSSTFN	LMCQPILL	FKLSYGIAT	18	5.000	29
5937	GYPNMFITREEAIRH	LHPTQAPTH	LTLTRALTA	18	1.800	30
2870	TPSKLIEYTDFA TSA	ILRTTNGDF	ARYMRS LKV	18	4.000	31
3580	ENAF LPFAMGIIAMS	WLLLTILTS	VLGSLAATV	22	4	31
2890	ECTIFK DASGKVPY	ICYTPSKLI	VTALRANSA	20	9.000	32
749	TEVLTEEVVLKTGDL	VLTEEVVLK	VASIKNFKS	20	2	33
2567	VNTFSSSTFNVPEK L	LLLDQALVS	YGIATVREV	22	4	34
1931	TGYKKPASRELKVTF	IKFADDLNQ	VLSVNPYVC	18	1	35
1660	CYLATALLT LQQIEL	VNGLTSIKW	LSMQGAVDI	18	8.800	36
6616	HIVYGD FSHS QL GGL	LIGEAVKTQ	IKASMP TTI	28	3	37
4880	YTKRNV I P TITQ MNL	IVNNLDKSA	LVQQESPFV	18	1	38
1072	ESDDYIATNGPLKVG	LNKATNNAM	FRYMNSQGL	18	1	39

3445	AAVINGDRWFLNRFT	FVDRQTAQA	LRLIDAMMF	20	8.000	39
4069	LSEISMNDSPNLAWP	VIPDYNTYK	FLALCADSI	26	3	39
6360	GFSLWVYKQFDTYNL	FVNLKQLPF	YAADPAMHA	20	2	39
4095	VKLQNNELSPVALRQ	WEIQQVVDA	FRRAFGEYS	18	3	40
5742	MFYKGVITHDVSSAI	YFNSVCRML	LRGTAVMSL	18	5.000	41
6403	EVVPSIINNTVYTKV	VCRHHANEY	LKTLATHGL	20	4.000	41
2839	PGLPGTILRRTNGDF	YTNDKACPL	LASFSASTS	18	9.000	42
4578	GSGVVPVDSYYSLLM	VQFCDAMRN	LTLAVPYNM	26	4.000	43
1420	LINTLNDLNETLVTM	FYTSKTTVA	LRVESSSKL	18	4	46
3909	LPSYAAFATAQEAYE	VSLLSVLLS	IQPIGALDI	22	3	46
3416	FYGFVDRQTAQAAG	FCYMHMEL	FDNLKTLLS	18	1.000	47
4135	GGRFVLALLSDLQDL	VKLQNNELS	IIIGGAKLK	22	4	48
<b>PREDICCIÓN DE EPÍTOPES DE ALTA UNIÓN HLA-DRB1*0701</b>						
832	EVQGYKSVNITFELD	VQGYKSVNI	VYCFGLGYFC	18	5,1	1,1
2210	SKLINIIIWFLLLSV	FNYLKSPNF	FYILPSIIS	22	3,3	3,8
4934	GATVVIGTSKIFYGGW	YAISAKNRA	FVMSAPPA	22	3,3	3,8
1828	KETLYCIDGALLTKS	YQCGHYKHI	FSSLPSYAA	22	3	3,9
3417	YGFVDRQTAQAAGT	FCYMHMEL	YVLMDSGSI	22	4,8	4,1
1834	IDGALLTKSSEYKGP	YKHITSKET	FSSLPSYAA	22	3,9	4,1
2427	CDTFCAGSTFISDEV	VRRSFYVYA	FYLITPVHV	18	4	4,3
1054	GGVAGALNKATNNA	VVNAANVYL	FRYMNSQGL	20	6	4,9
2210	SKLINIIIWFLLLSV	FNYLKSPNF	FYILPSIIS	22	3,3	3,8
2198	EASFNYLKSPNFSKL	VKSVGKFC	FYILPSIIS	24	5,6	5,8
4827	DKYFDCYDGGCINAN	VELKHFFFA	YKLRSDVL	18	1.000	6,4
5	VPGFNEKTHVQLSLP	FNEKTHVQL	FYRLANCA	18	2,5	6,9
1072	ESDDYIATNGPLKVG	LNKATNNAM	FRYMNSQGL	20	2,6	7,3
4845	VNLDKAGSAGPFNKW	YNLPTMCDI	YKLRSDVL	20	3,2	7,3
3759	NRYFRLTLGVYDYL	VEYCIFFI	YRSLPGVFC	18	4,9	7,4
4775	SKGFFKEGSSVELKH	LLDKRTTCF	YNSQNAVAS	18	3.200	7,6
2789	TRDIASDTCFANKH	ITPVHVMK	YRVTKNSKV	28	1	9
2882	TSACVLAAECTIFKD	FLPRVFSAV	ARYMRSLKV	20	1,8	9,8
6888	KLALGGSVAIKITEH	LVSDSLNDF	LKSIAATRG	22	3,9	10
3826	TSVLLSVLQQLRVE	LPPKNSIDA	YLAGGGQPI	22	5.000	10
2567	VNTFSSTFNVPEKEL	LLLQALVS	YGIATVREV	32	1,3	10
1465	SPDAVTAYNGYLTSS	YMRSLKVPA	YELQTPFEI	20	3,8	10
2136	VSTTTNIVTRCLNRV	YAKPFLNKV	FTRLQSLN	22	1,3	11
6886	QKALGGSVAIKIT	WLPTGTLV	LKSIAATRG	24	7	11
2124	IANYAKPFLNKVVST	LAAVNSVPW	FTRLQSLN	18	2,9	11
6709	ISFMLWCKDGHVETF	FELEDIFIM	FSKLINII	22	3,3	12
2952	DSEYCRHGTCESEA	YVLMDSGSI	FRNARNGVL	18	5,9	12
4069	LSEISMNDSPNLAWP	VVIPDYNTY	FLALCADSI	24	2,9	12
1625	DPSFLGRYMSALNHT	LRVEAFEYY	YEAMYPHT	18	1,5	12
6947	IDGYVMHANYIFWRN	FFTYICGFI	FFYVGLGAA	22	3,3	13
17	SLPVLQVRDVLVRGF	VRDVLVRGF	LANECAQVL	22	3,7	13

1582	PTYLDGADVTKIKPH	MSMTYQQQF	FEKMVSLLS	22	4,1	14
3586	FAMGIIAMSAFAMMF	LVLVQSTQW	VLGSLAATV	18	5,3	14
263	FDTFNGECPNFVFPPL	FNGECPNFV	FNICQAVTA	18	2,5	15
4841	NQVIVNNLDKSAGFP	YYRYNLPMT	ILGLPTQTV	18	1,6	15
959	PLEFGATSAALQPEE	FGATSAALQ	FCSQHTMLV	18	10.000	15
6616	HIVYGFDFSHSQLGGL	IGEAVKTQF	IKASMPPTI	18	2.000	15
6682	DDFVEIIKSQDLSVV	FEHIVYGDF	FNYLKS PNF	30	2,5	16
6403	EVVPSIINNTVYTKV	VCRHHANEY	LKTLATHGL	26	2	16
6828	LNDFVSDADSTLIGD	YGDSATLPK	LKYAISAKN	24	2	16
6813	LRQWLPTGTLVSDSD	YKMQRMLLE	YLNSTNVTI	34	2	16
4787	LKHFFFAQDGNAAIS	LTNNVAFQT	YNSQNAVAS	26	9.000	16
2118	SVPWDTIANYAKPFL	LKTLATHGL	FTRLQSLN	22	4	16
5020	VCRLMKTIGPDMFLG	LRIMASLVL	YNLLKDCPA	24	9	16
3580	ENAFLPFAMGIIAMS	LLLTILTSL	VLGSLAATV	18	6	17
3077	YSVIYLYLTFYLTND	VVAFNTLLF	VMPLSAPTL	24	6	15.01
2559	AVKMFDAVNTFSST	LMCQPILL	YKAIDGGVT	20	2.1	17
2459	SSYIVDSVTVKNGSI	ISDEVARDL	LKLFAAETL	22	1.000	17
4304	KGKYVQIPTTCANDP	ITVTPEANM	YYRARAGEA	28	4.000	18
3648	VLLILMTARTVYDDG	VYMPASWVM	LHLLIGLAK	22	3.1	18
958	KPLEFGATSAALQPE	FGATSAALQ	FRLTLGVYD	22	10.000	19
6363	LWVYKQFDTYNLWNT	LKQLPFFYY	YAADPAMHA	26	31.12	19
2299	SFKWDLTAFGLVAEW	YPSLETIQI	YHNESGLKT	26	3.1	20
3909	LPSYAAFATAQEAYE	MVSLLSVLL	IQPIGALDI	18	6	20
3083	YLTFYLTNDVSLFLAH	LMSFTVLCL	VMPLSAPTL	30	6	20
2864	VGNICYTPSKLIEYT	VVPGLPGTI	ARYMRS LKV	32	3	21
5051	EQYVFCTVNALPETT	VLSEMVMCG	FNVAITRAK	22	1.000	21
2346	SNSWMLWLIINLVQM	FSYFAVHFI	YTKVDGVDV	18	5	20.01
5108	LEQYVFCTVNALPET	YVRNLQHRL	YKHITSKET	18	5	22
3820	MSDVKCTSVVLLSVL	FRYMNSQGL	YLASGGQPI	24	8	22
3903	ASEFSSLP SYAAFAT	FEKMVSLLS	IQPIGALDI	22	2	22
2151	CTNYMPYFFTL LQL	IVTRCLNRV	LTCATTRQV	18	2,5	23
6755	LQNYGDSATLPGIM	VEI IKSQDL	LIYSTAALG	24	6,8	23
5057	NVAITRAKVGILCIM	VMCGGSLYV	VAITRAKVG	22	3,8	23
4529	GERVRQALLKTVQFC	LKEILVTYN	YKGVITHDV	22	5.000	23
938	EEEFEPSTQYEGTE	FEPSTQY EY	HEVLLAPLL	22	3,6	24
6703	TIDYTEISFMLWCKD	IGLAKRFKE	FSKLINII	22	1,2	24
3640	CVMYASAVLLILMT	LATVAYFNM	VRLQAGNAT	24	3	26
6124	VASCDAIMTRCLAVH	IGPERTCCL	LKSEDAQGM	22	1,72	26
2624	VECLKLSHQSDIEVT	LDNVLSTFI	FYFYTSKTT	22	4	26
6406	VSIINNTVYTKVDGV	VCRHHANEY	LKTLATHGL	24	2	26
581	IDAMMFTSDLATNNL	MMFTSDLAT	LRKHFSMMI	22	2	26
3915	FATAQEAYEQAVANG	LSMQGAVDI	IQPIGALDI	20	4	27
5114	QAPTHLSVDTKFKTE	LQHRLYECL	YKHITSKET	18	3	27
6619	YGDFSHSQLGGLHLL	IGEAVKTQF	IKASMPPTI	24	2.000	27

5045	GDAVVYRGTTTTYKLN	LANECAQVL	FNVAITRAK	22	3	27
1602	KTFYVLPNDTLRVE	IKPHNSHEG	FVNLKQLPF	24	1,4	27
2777	EIIGYKAIDGGVTRD	FVAAIIFYLI	YRVTKNSKV	20	6,4	28
6354	NMMISAGFSLWVYKQ	FDKSAFVNL	LLVYAADPA	24	4,4	29
3553	HWLLLTILTSLLVLV	FTPFDVVRQ	MAYITGGVV	30	1,3	29
1950	NGDVVAIDYKHYTPS	LKVTFFPDL	YLDAYNMMI	20	4,4	29
2858	PRVFSAVGNICYTPS	ITREVGFEV	ARYMRSLKV	22	1,8	29
5026	VTGLFKDCSKVITGL	LVLARKHTT	YNLLKDCPA	24	2,3	30
1253	SATLVSDIDITFLKK	LYIDINGNL	FVSLAIDAY	20	3	30
5039	ACVLCNSQTSLRCA	FYRLANECA	FNVAITRAK	22	1	30
4746	FSVAALTNNAFQTV	VNLHSSRLS	WLPTGTLV	20	2	30
6701	KVTIDYTEISFMLWC	LLIGLAKRF	I IAMSAFAM	24	3	31
3855	DILLAKDTEAFEKEM	VATVQSKMS	LTNMFPLI	22	7.000	31
4071	EISMDNSPNLAWPLI	VVIPDYNTY	YFMRFRRA	26	3	29.01
2340	FAVHFISNSWLMWLI	LGLAAIMQL	YTKVDGVDV	22	5	31
5063	SHRFYRLANCAQVL	YVKPGGTSS	VAITRAKVG	20	9.000	32
1238	NLLLYIDINGNLHPD	LEETKFLTE	FLEMKSEKQ	22	4.000	32
575	QYSLRLIDAMMFTSD	LRLIDAMMF	YAYLRKHFS	20	7	33
3222	YQPQTSITSAVLQS	YREAACCHL	YKKVDGVVQ	22	4	33
3968	TQMYKQARSEDKRAK	VANGDSEVV	YMSALNHTK	24	2	33
4135	GGRFVLALLSDLQDL	LQNNELSPV	I IIGGAKLK	24	1	34
3223	QPPQTSITSAVLQSG	YREAACCHL	VRSIFSRTL	22	4	36
2133	NKVVSTTNIIVTRCL	IANYAKPFL	FKLTCATTR	22	7.000	36
754	EEVVLKTDLQPLEQ	LKTGDLQPL	IFFITGNTL	22	2,6	36
4584	VDSYSLMLPILTLT	MRNAGIVGV	LTLAVPYNM	18	4,8	37
428	SANIGNHTGVVGE	IGCNHTGVV	VVNAANVYL	24	3.200	37
1615	VEAFEYHHTDPSFL	LPNDTLRV	FVFPLNSII	18	1,52	37
403	TIAFGGCVFSYVGCH	FGGCVFSYV	IKRSDARTA	18	2,3	37
2565	AYVNTFSSTFNVPME	LLLDQALVS	YKAIDGGVT	22	1,3	37
5102	FGLYKNTCVGSDNVT	IADKYVRNL	YKHITSKET	18	6.000	38
2883	SACVLAAECTIFKDA	FSAVGNICY	VVPGLPGTI	20	3,5	39
6498	CAPLTVFFDGRVDGQ	LPVNAFEL	YDFAVSKGF	22	3,4	39
6820	GTLLVSDLDNFVSD	LEKCDLQNY	MNLKYAISA	20	4.000	40
4569	FGDFIQTPGSGVPV	VRQALLKTV	YIICISTKH	18	2,8	40
4789	HFFFAQDGNAAISDY	LTNNVAFQT	YCALAPNMM	24	9.000	40
4496	LVTYNCCDDYFNKK	MVPHISRQR	YLFQHANLD	22	2	40
3691	MWALII SVTSNYSGV	ILMTARTVY	YEKLKPVLD	28	3,7	42
803	APNMMVTNNTFTLKG	MVTNNTFTL	LYYQNNVFM	22	5,42	42
2253	REGYLNSTNVTIATY	LMSNLGMPS	I ALKGGKIV	30	1.000	43
28	VRGFGDSVEEVLSEA	FGDSVEEVL	YFNMVYMPA	30	2,5	43
6670	SKCVCSVIDLLDDF	IERYKLEGY	FCLEASFNY	22	0	44
3139	FNGVSFSTFEEAALC	ITIAIYICI	KSILSPLYA	18	5,8	44
3935	LKKLKKSLNVAKSEF	LQAIASEFS	ISMATNYDL	20	2,5	44
6697	SKVVKVTIDYTEISF	LLIGLAKRF	LKSPNFSKL	22	3	45

1281	VLTAVVIPTKKAGGT	YIVGDVVQE	FSSTFNVPM	18	1,9	46
788	LMLEIKDTEKYCAL	LEIKDTEKY	LHVVGPNVN	22	1,22	46
1883	DNYYKKDNSYFTEQP	LDGVVCTEI	VISTSHKLV	18	3	48
4977	PNMLRIMASLVLARK	VVIGTSKIFY	YKYFSGAMD	24	4	49

PROTEINA ORF 3A						
POS	SECUENCIA	CORE PEPTIDE P9	CORE PEPTIDE NETPAN	SYFPEITHI	P9	IC50
<b>PREDICCIÓN DE EPÍTOPES DE ALTA UNIÓN HLA-DRB1*0101</b>						
66	KKRWQLALSKGVHFV	WQLALSKGV	WQLALSKGV	28	1,7	4,8
25	PSDFVRATATIPIQA	FVRATATIP	FVRATATIP	28	1,79	8,9
26	SDFVRATATIPIQAS	FVRATATIP	VRATATIPI	24	1,79	10
208	TSDYYQLYSTQLSTD	YYQLYSTQL	YYQLYSTQL	33	1,79	11
157	CIPYNSVTSSIVITS	YNSVTSSIV	YNSVTSSIV	31	4.500	11
209	SDYYQLYSTQLSTDT	YYQLYSTQL	YYQLYSTQL	25	2	11
68	RWQLALSKGVHFVCN	WQLALSKGV	WQLALSKGV	25	2	12
90	VYSHLLVAAGLEAP	LLVAAGLEA	LLVAAGLEA	27	5.000	16
91	YSHLLVAAGLEAPF	LLVAAGLEA	LLVAAGLEA	23	5.000	16
92	SHLLVAAGLEAPFL	LLVAAGLEA	LLVAAGLEA	24	5.000	23
4	FMRIFTIGTVTLKQG	FMRIFTIGT	IFTIGTVTL	29	2	26
2	DLFMRIFTIGTVTLK	FMRIFTIGT	IFTIGTVTL	21	2	38
93	HLLVAAGLEAPFLY	LLVAAGLEA	LLVAAGLEA	18	5.000	42
112	VYFLQSIN FVRIIMR	VYFLQSINF	LQSIN FVRI	22	1	42
109	YALVYFLQSIN FVRI	VYFLQSINF	YFLQSIN FV	31	1	47
110	ALVYFLQSIN FVRII	VYFLQSINF	YFLQSIN FV	25	1	48
<b>PREDICCIÓN DE EPÍTOPES DE ALTA UNIÓN HLA-DRB1*0701</b>						
66	KKRWQLALSKGVHFV	WQLALSKGV	WQLALSKGV	32	5,3	7
26	SDFVRATATIPIQAS	VRATATIPI	VRATATIPI	32	8,32	11
158	IPYNSVTSSIVITSG	YNSVTSSIV	YNSVTSSIV	24	4,9	11
157	CIPYNSVTSSIVITS	YNSVTSSIV	YNSVTSSIV	26	4,9	12
68	RWQLALSKGVHFVCN	WQLALSKGV	LALSKGVHF	20	5,3	13
156	YCIPYNSVTSSIVIT	YNSVTSSIV	YNSVTSSIV	22	4,9	13
52	LLAVFQSASKIITLK	VFQSASKII	VFQSASKII	30	5,8	14

PROTEINA ORF 6						
POS	SECUENCIA	CORE PEPTIDE P9	CORE PEPTIDE NETPAN	SYFPEITHI	P9	IC50
<b>PREDICCIÓN DE EPÍTOPES DE ALTA UNIÓN HLA-DRB1*0101</b>						
11	IAEILLIIMRTFKVS	LIIMRTFKV	LIIMRTFKV	24	1.80	6,6
12	AEILLIIMRTFKVSI	LIIMRTFKV	LIIMRTFKV	18	1.80	7
13	EILLIIMRTFKVSIW	LIIMRTFKV	LIIMRTFKV	32	1.80	7,9

14	ILLIIMRTFKVSIWN	LIIMRTFKV	LIIMRTFKV	18	1.80	13
29	LDYIINLIKLNLSKS	YIINLIKLN	IINLIKLN	25	1.400	33
30	DYIINLIKLNLSKSL	YIINLIKLN	IINLIKLN	24	1.400	35
<b>PREDICCIÓN DE EPÍTOPES DE ALTA UNIÓN HLA-DRB1*0401</b>						
12	AEILLIIMRTFKVSI	LIIMRTFKV	LIIMRTFKV	26	2.70	13
13	EILLIIMRTFKVSIW	LIIMRTFKV	LIIMRTFKV	20	2.70	14
31	YIINLIKLNLSKSLT	YIINLIKLN	IKNLSKSL	18	02.08	49
<b>PREDICCIÓN DE EPÍTOPES DE ALTA UNIÓN HLA-DRB1*0701</b>						
11	IAEILLIIMRTFKVSI	LIIMRTFKV	LIIMRTFKV	20	5.22	6,6
13	EILLIIMRTFKVSIW	IMRTFKVSI	LIIMRTFKV	24	6.10	7,9
15	LLIIMRTFKVSIWNL	IMRTFKVSI	LIIMRTFKV	24	6.10	16

PROTEINA ORF 7A						
POS	SECUENCIA	CORE PEPTIDE P9	CORE PEPTIDE NETPAN	SYFPEITHI	P9	IC50
<b>PREDICCIÓN DE EPÍTOPES DE ALTA UNIÓN HLA-DRB1*0101</b>						
71	VKHVYQLRARSVSPK	VYQLRARSV	VYQLRARSV	33	1,19	7,7
72	KHVYQLRARSVSEPKL	VYQLRARSV	VYQLRARSV	19	1,19	9,4
68	PDGVKHVYQLRARSV	VYQLRARSV	VYQLRARSV	23	1,19	14

PROTEINA ORF 8						
POS	SECUENCIA	CORE PEPTIDE P9	CORE PEPTIDE NETPAN	SYFPEITHI	P9	IC50
<b>PREDICCIÓN DE EPÍTOPES DE ALTA UNIÓN HLA-DRB1*0101</b>						
43	SKWYIRVGARKSAPL	YIRVGARKS	YIRVGARKS	35	1.54	7,6
42	YSKWYIRVGARKSAP	YIRVGARKS	YIRVGARKS	21	1.54	9,3
44	KWYIRVGARKSAPLI	YIRVGARKS	YIRVGARKS	19	1.54	9,7
4	LVFLGIITTVAAAFHQ	LVFLGIIT	FLGIITVA	31	1.30	26
3	FLVFLGIITTVAAAFH	FLVFLGIIT	FLGIITVA	31	1.68	32
1	MKFLVFLGIITTVAA	FLVFLGIIT	FLGIITVA	19	1.68	48

PROTEINA ORF 10						
POS	SECUENCIA	CORE PEPTIDE P9	CORE PEPTIDE NETPAN	SYFPEITHI	P9	IC50
<b>PREDICCIÓN DE EPÍTOPES DE ALTA UNIÓN HLA-DRB1*0101</b>						
10	PFTIYSLLLCRMNSR	YSLLLCRMN	YSLLLCRMN	24	4.900	35.02
11	FTIYSLLLCRMNSRN	YSLLLCRMN	YSLLLCRMN	20	4.900	33.95

Tabla 2. Selección de péptidos en cada proteína (Creado por grupo de investigación)



*Péptidos en común de los alelos*

Finalmente, se identificaron los péptidos en común para los tres alelos. Las proteínas que presentaron estos péptidos comunes fueron: Para la proteína S se encontró un péptido en común entre DRB1\*0101, DRB1\*0401 y DRB1\*0701 (diagrama 1). Para la proteína E se encontró 1 péptido en común entre DRB1\*0101 y DRB1\*0701 y 1 péptido en común entre DRB1\*0101 y DRB1\*0401 (diagrama 2). La proteína M tuvo en común 1 péptido entre DRB1\*0101, DRB1\*0401 y DRB1\*0701 (diagrama 3). Para la proteína ORF 1ab se encuentran 4 péptidos en común entre DRB1\*0101, DRB1\*0401 Y DRB1\*0701 (diagrama 4). En la proteína ORF3a se encontraron 4 péptidos en común solamente entre DRB1\*0101 y DRB1\*0701 (diagrama 5) y por último la proteína ORF 6 se encontró 1 péptido en común entre DRB1\*0101, DRB1\*0401 y DRB1\*0701 (diagrama 6). Por otro lado, las proteínas a las cuales no se le encontraron ningún péptido en común fueron la proteína N, la proteína ORF7a, la proteína ORF8 y la proteína ORF 10.

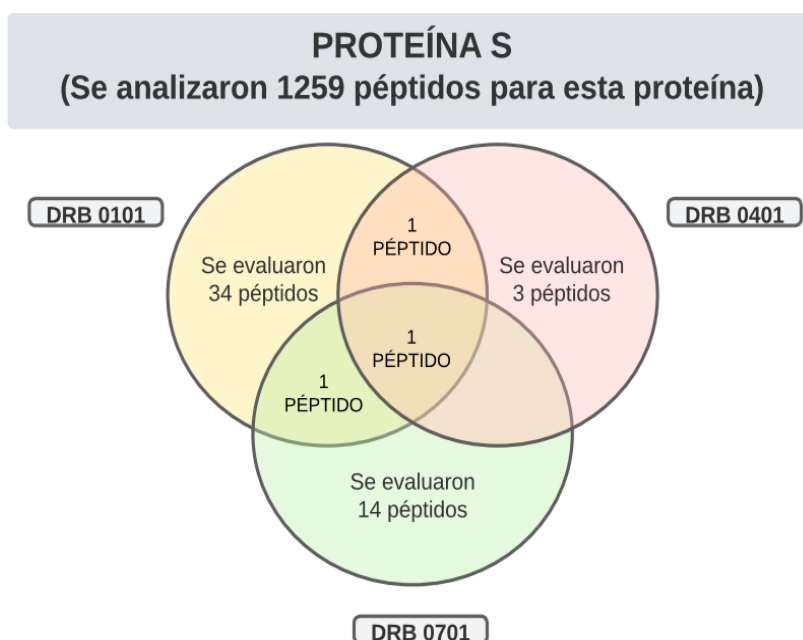


Diagrama 1. Péptidos en común de la Proteína S (Creado por grupo de investigación)

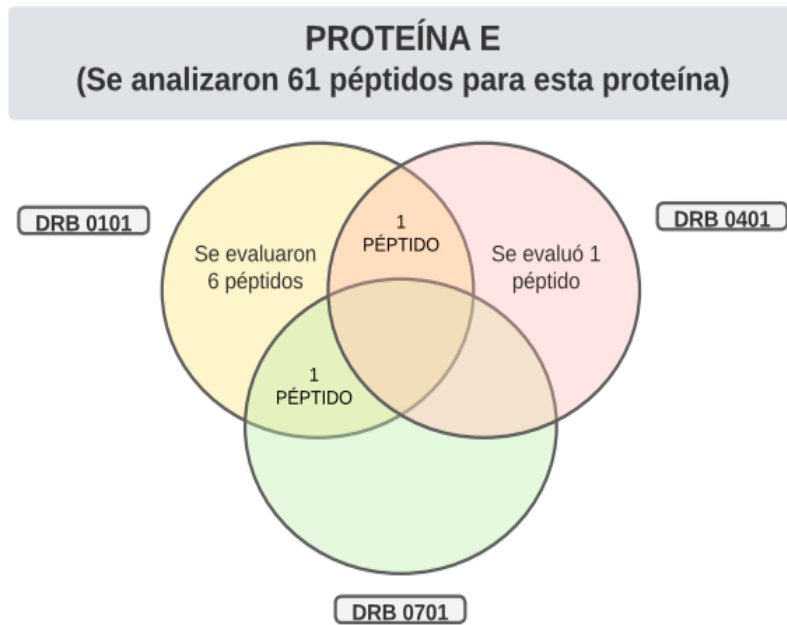


Diagrama 2. Péptidos en común de la Proteína E (Creado por grupo de investigación)

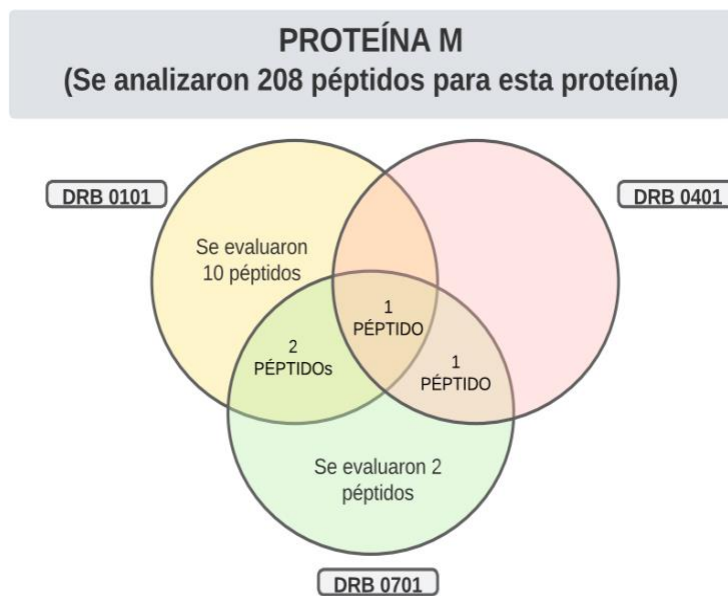


Diagrama 3. Péptidos en común de la Proteína M (Creado por grupo de investigación)

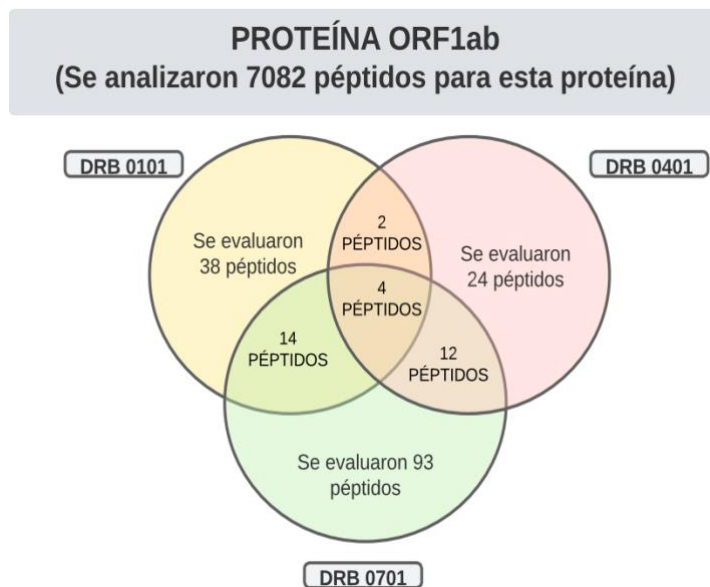


Diagrama 4. Péptidos en común de la Proteína ORF 1ab (Creado por grupo de investigación)

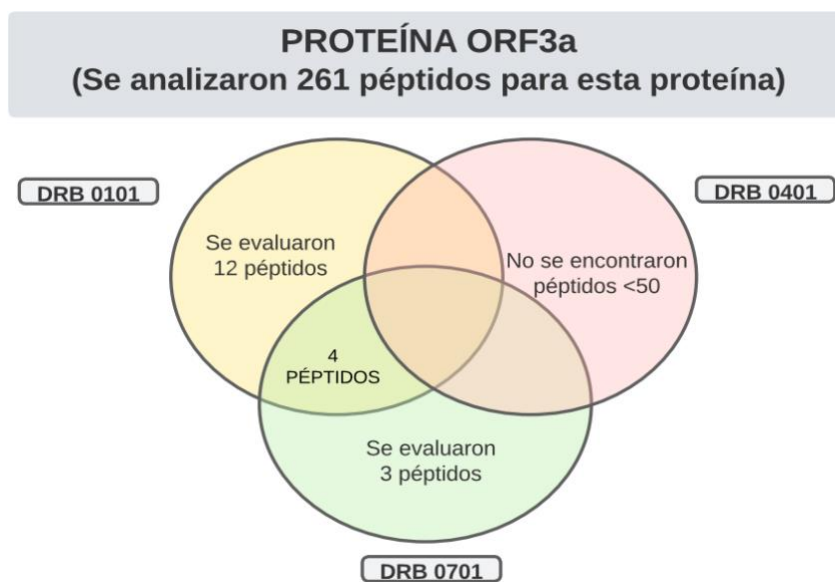


Diagrama 5. Péptidos en común de la Proteína ORF3a (Creado por grupo de investigación)

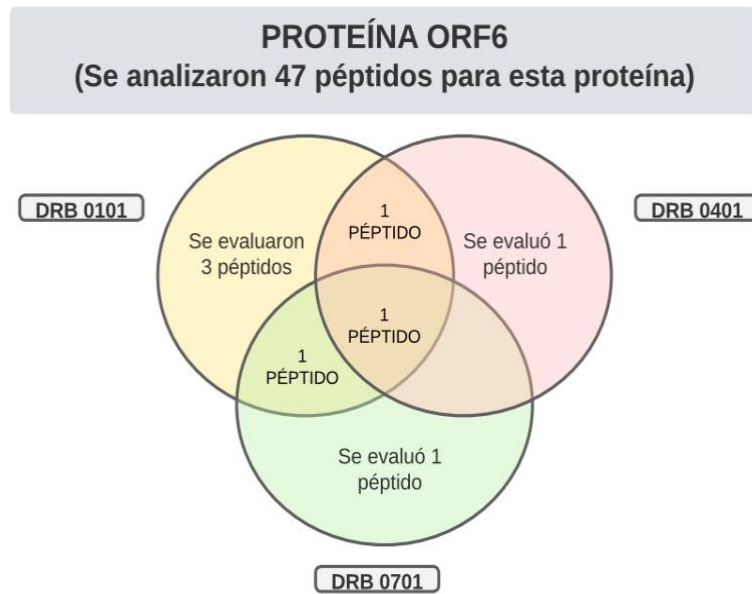


Diagrama 6. Péptidos en común de la Proteína ORF 6 (Creado por grupo de investigación)

## Discusión de resultados

La predicción y análisis de péptidos que se presentan a diferentes alelos del complejo mayor de histocompatibilidad se ha venido realizando hace un tiempo mediante programas y plataformas bioinformáticas y ha sido sugerido por autores como Nielsen et al. (31). El resultado de este análisis permite el reconocimiento de los posibles péptidos que van a ser presentados con una con una mayor probabilidad teniendo en cuenta su capacidad de unión evaluada mediante plataformas bioinformáticas que usan diferentes estrategias para poder hacer esas predicciones. Los péptidos identificados en este trabajo son los señalados en la Tabla 2. Además, se observó que la molécula de HLA que reconoce más probables péptidos de las proteínas del SARS-CoV-2 es HLA-DRB1\*0101 al cual se unieron mayor cantidad de péptidos, seguido del HLA-DRB1\*0701.

El enfoque bioinformático descrito en el estudio da la capacidad de identificar los epítopes más inmunodominantes los cuales proporcionan información sobre la probabilidad de unión a las células TCD4 + para poder producir la inmunidad. En total son 7 péptidos inmunodominantes que hacen parte de la proteína M, S, ORF6 y ORF1ab, siendo esta última la que mayores péptidos inmunodominantes tiene, ya que son 4 que se expresan y se pueden identificar tanto en DRB1\*0101 como en DRB1\*0401 y en DRB\*0701.

Para la correcta interpretación de los resultados expuestos se debe tener en cuenta que la confiabilidad del uso de plataformas bioinformáticas en la predicción de epítopes de moléculas MHC clase II no es muy alta, a pesar de que su desarrollo en los últimos años ha mejorado su capacidad de predicción. Esto es debido a la gran variabilidad de los bolsillos de unión, estimándose de esta manera un 50% de probabilidad de error (26). Sin embargo, los resultados obtenidos en estas plataformas son una primera muestra de la utilidad en futuros ensayos *in-vitro* o *ex-vivo*, destinados al desarrollo de vacunas contra el SARS- CoV-2.

En el estudio realizado por Grifoni y colaboradores, donde utilizando la plataforma IEDB buscan determinar los posibles blancos de la respuesta de células T y B contra la infección del Sars-CoV-2. El umbral de selección de los epítopes tuvo como límite técnico un percentil de consenso medio del 20 %. Dentro de los resultados obtenidos por los autores se destaca que para las células T los péptidos inmunodominantes son las proteínas S y la proteína N teniendo 23 y 10 epítopes que cumplían con los criterios de elección. Además, encontraron que las proteínas ORF 1ab, proteína M, y la proteína ORF3 tenían 8, 3 y 2 epítopes respectivamente (12).

A diferencia de los resultados entregados por Grifoni y colaboradores (12), en este estudio se constató que: la proteína ORF1ab fue la que más epítopes relacionados se encontraron, junto con la proteína M, la proteína S y la proteína ORF6a que tenían péptidos en común entre los 3 HLA y la proteína N no tuvo ningún péptido en común. Ambos estudios difieren principalmente en el carácter inmunodominante de la proteína S, ya que en este estudio se encontraron pocos epítopes de esta proteína que cumplieran con los criterios de selección. De esta manera se puede observar que utilizando umbrales de selección diferentes los resultados obtenidos también serán diferentes. No obstante, es importante aclarar el sesgo de selección que se presenta con trabajos de este tipo que únicamente se basan en algoritmos de predicción bioinformáticos, siendo de esta manera necesario complementar con técnicas in-vivo para la obtención de resultados veraces.

Tarker y colaboradores (32) por medio de estudios *ex-vivo* marcaron con positividad todos los epítopes que tuvieron recuentos netos de AIM+ >100 y un índice de estimulación (IS) > 2. Los resultados obtenidos recalcan la inmunodominancia de las proteínas S, ORF8, M, ORF3a, guardando cercanía con los resultados de este trabajo, pero discrepando en la superior inmunodominancia de la proteína S sobre las demás (31). Por otro lado, J mateus y colaboradores (33) realizaron un estudio en individuos no expuestos previamente al Sars-CoV-2 para establecer las células T CD4+ que reconocen los epítopes del virus mediante el estímulo in vitro PMBC (células mononucleares de

sangre periférica en un período de tiempo de 2 semanas con la estimulación in vitro de células mononucleares de sangre periférica (PBMC) durante 2 semanas con grupos de péptidos de 15 mer. Los resultados demuestran que un 54% de la respuesta positiva antigénica se asoció con epítomos derivados de la proteína Spike ( 11% para el dominio de unión al receptor DRB y 44% para la parte que no pertenece al receptor). Además de la gran afinidad que hubo a la proteína S, también hubo respuesta para la proteína ORF6, ORF3a, N, ORF8 y dentro de ORF1a/b. La proteína M no demostró una identificación de epítomos en individuos no expuestos al virus, pero en casos de infección por SARS-CoV-2 si se reconocen respuestas marcadas de las células T CD4+. En relación al presente trabajo, los resultados tienen en común la identificación y reconocimiento de epítomos tras la infección por el virus y su afinidad con ciertas proteínas como lo son la S, ORF6 y ORF1ab, aunque las metodologías difieren. La proteína M no mostró igualdad de resultados debido a la no exposición del virus en los pacientes en los cuales se realizó el estudio (33).

Los estudios ejemplifican resultados cercanos pero no similares a este, a pesar de ser realizados con diferente metodología, demostrando la inmunodominancia de los péptidos principalmente la proteína S, N, ORF1ab y ORF3a, siendo la proteína S la que se evidencia en los artículos y el presente trabajo (31-33).

En cuanto a las fortalezas evidenciadas durante el desarrollo de esta búsqueda sistemática evidenciamos que este estudio nos permite reconocer los posibles péptidos que van a ser presentados con una mayor probabilidad por las moléculas HLA-DR1 de clase II, adicionalmente el presente trabajo de investigación nos sirve como herramienta para identificar los epitopes más inmunodominantes los cuales proporcionan información sobre la probabilidad de unión de las células TCD4 para así producir la inmunidad, otra de las fortalezas evidenciadas durante el desarrollo del mismo es que los resultados obtenidos en estas plataformas son una primera muestra de la utilidad en futuros ensayos in-vitro o ex-vivo, destinados al desarrollo de vacunas contra el SARS.

Cov2, así mismo esta investigación nos demuestra la inmunodominancia de los péptidos respaldado por otros estudios que cuentan con resultados similares.

En cuanto a las limitaciones reflejadas durante la investigación encontramos que la confiabilidad de las plataformas utilizadas no es muy alta dado que es operador dependiente lo que nos podría arrojar falsos positivos, así mismo se evidenció la existencia de una gran variabilidad de bolsillos de unión, dando casi un 50% de probabilidad de error, por otro lado cabe mencionar dentro de las limitantes que no hay muchos estudios en Colombia que usen los mismos DR con las plataformas que se usan en el presente trabajo, por último es importante tener en cuenta que los estudios de comparación de esta investigación son realizados con diferentes metodologías, es por eso que los resultados de los diferentes estudios utilizados son cercanos, pero no similares a los nuestros.



## Conclusiones

Tras el análisis realizado podemos concluir que es importante profundizar un poco más en este tipo de estudios, llevando a cabo investigaciones en las cuales se haga uso de plataformas bioinformáticas como lo son Syfpeithi, P9 y NetMHCIIpan, ya que usando estas plataformas podemos filtrar mejor los péptidos y obtener resultados más específicos, todo esto con el fin de obtener diversidad de estudios los cuales nos arrojen información sobre los posibles péptidos derivados del SARS-CoV-2 presentados por las moléculas MHC-II a los T CD4, para así tener una información veraz en cuanto a resultados y de fácil acceso.

Adicionalmente, de la totalidad de proteínas analizadas en este estudio, aquellas en las que encontramos péptidos en común para los alelos son: la Proteína S, la Proteína E, la Proteína M, la proteína ORF 1ab, la Proteína ORF3a y la Proteína ORF 6, esto nos lleva a concluir que son estas las que nos darán inmunización por medio de las células T y que cobrarán gran importancia en nuestra respuesta inmunológica. A su vez encontramos que la proteína S al ser la más grande es la que más péptidos específicos podrá generar, teniendo así más afinidad con los linfocitos T CD4 y desempeñando un papel importante a nivel inmunitario.

## Recomendaciones

Se recomienda a los investigadores realizar estudios funcionales en los cuales se pueda comprobar la veracidad de los resultados bioinformáticos de afinidad de las proteínas S, M, N, ORF1ab, ORF6 y ORF3a, sobre todo en la población colombiana para saber si los programas de vacunación son los adecuados. Igualmente, se recomienda usar las plataformas bioinformáticas, las cuales dan resultados sencillos y claros sobre la predicción de los péptidos y sugerimos usar criterios acorde a lo que la literatura específica menciona. La metodología de las plataformas que se usa en este trabajo sugerimos se utilice en más trabajos en Colombia, para poder hacer comparaciones más eficaces y pertinentes con respecto a las proteínas específicas. Finalmente, estos estudios deben complementarse con análisis de estimulación ex vivo como Elisspots o análisis de citometría de flujo para identificar si las predicciones se aproximan a lo que sucede en la realidad.

## Referencias

1. Yamayoshi S, Sakai-Tagawa Y, Koga M, Akasaka O, Nakachi I, Koh H, Maeda K, Adachi E, Saito M, Nagai H, Ikeuchi K, Ogura T, Baba R, Fujita K, Fukui T, Ito F, Hattori SI, Yamamoto K, Nakamoto T, Furusawa Y, Yasuhara A, Ujie M, Yamada S, Ito M, Mitsuya H, Omagari N, Yotsuyanagi H, Iwatsuki-Horimoto K, Imai M, Kawaoka Y. Comparison of Rapid Antigen Tests for COVID-19. *MDPI J Vir.* 2020 10;12(12):1420.
2. Umakanthan S, Sahu P, Ranade AV, Bukelo MM, Rao JS, Abrahao-Machado LF, Dahal S, Kumar H, Kv D. Origin, transmission, diagnosis and management of coronavirus disease 2019 (COVID-19). *Postgrad Med J.* 2020 96(1142):753-758.
3. Johns Hopkins University & Medicine. COVID-19 Map. Johns Hopkins Coronavirus Resource Center [Internet]. [Consultado 24 de marzo de 2022.] Disponible en: <https://coronavirus.jhu.edu/map.html>
4. Ministerio de Salud de Colombia. Coronavirus (COVID-19) Reportes de tableros y control [Internet]. [Consultado 24 de marzo de 2022]. Disponible en: [https://www.minsalud.gov.co/salud/publica/PET/Paginas/Covid-19\\_copia.aspx](https://www.minsalud.gov.co/salud/publica/PET/Paginas/Covid-19_copia.aspx)
5. Ministerio de Salud de Colombia. El Coronavirus en Colombia. [Consultado 24 de marzo de 2022] Disponible en: <https://coronaviruscolombia.gov.co/Covid19/index.html>
6. Dan JM, Mateus J, Kato Y, Hastie KM, Yu ED, Faliti CE, Grifoni A, Ramirez SI, Haupt S, Frazier A, Nakao C, Rayaprolu V, Rawlings SA, Peters B, Krammer F, Simon V, Saphire EO, Smith DM, Weiskopf D, Sette A, Crotty S. Immunological memory to SARS-CoV-2 assessed for up to 8 months after infection. *Science.* 2021. 5;371(6529)
7. Swadling L, Maini M. T cells in COVID-19 — united in diversity. *Nat Immunol.* 2020;21(11):1307-1308.

8. Peng, Y. et al. Broad and strong memory CD4+ and CD8+ T cells induced by SARS-CoV-2 in UK convalescent individuals following COVID-19 *Nat Immunol.* 2020. 1044-5323.
9. Jarjour NN, Masopust D, Jameson SC. T Cell Memory: Understanding COVID-19. *Immunity.* 2020. 12;54(1):14-18.
10. Kiyotani K, Toyoshima Y, Nemoto K, Nakamura, Y. Bioinformatic prediction of potential T cell epitopes for SARS-Cov-2. *J Hum Genet.* 2020.65, 569–575.
11. Sidney J, Peters B, Sette A. Epitope prediction and identification- adaptive T cell responses in humans. *Semin Immunol.* 2020. 50, 101418.
12. Grifoni A, Sidney J, Zhang Y, Scheuermann R, Peters B, Sette A. A Sequence Homology and Bioinformatic Approach Can Predict Candidate Targets for Immune Responses to SARS-CoV-2. *Cell Host Microbe.* 2020. 27, 671-680.
13. De Sanctis J, García A, Moreno D, Hajduch M. Coronavirus infection: An immunologists perspective. *Scand J Immunol.* 2021. 93, 6.
14. Mishra S, Tripathi T. One year update on the COVID-19 pandemic: Where are we now?. *Acta trop.* 2021. 214, 105778.
15. Ruyuan He, Zilong Lu, Lin Zhang, Tao Fan, Rui Xiong, Xiaokang Shen, Haojie Feng, Heng Meng, Weichen Lin, Wenyang Jiang, Qing Geng. The clinical course and its correlated immune status in COVID-19 pneumonia. *J Clin Virol.* 2020. 127, 104361.
16. Palacios M, Santos E, Velázquez M, León M. COVID-19, una emergencia de salud pública mundial. COVID-19, a worldwide public health emergency. *Rev Clin Esp.* 2020. 221, 1, 55–61.
17. Hu B, Guo H, Zhou P, Shi Z. Characteristics of SARS-CoV-2 and COVID-19. *Nat Rev Microbiol.* 2020. 19, 141–154
18. García, L. F. Immune Response, Inflammation, and the Clinical Spectrum of COVID-19. *Front Immunol.* 2020. 11, 144.

19. Li G, Fan Y, Lai Y, Han T, Li Z, Zhou P, Pan P, Wang W, Hu D, Liu X, Zhang Q, Wu J. Coronavirus infections and immune responses. *J Med Virol.* 2020. 94, 4, 424-432.
20. Chen G, Wu D, Guo W, Cao Y, Huang D, Wang H, Wang T, Zhang X, Chen H, Yu H, Zhang X, Zhang M, Wu S, Song J, Chen T, Han M, Li S, Luo X, Zhao J, Ning Q. Clinical and immunological features of severe and moderate coronavirus disease 2019. *J Clin Invest.* 2020. 130, 5, 2620-2629
21. Bordoni V, Sacchi A, Cimini E, Notari S, Grassi G, Tartaglia E, Casetti R, Giancola M, Bevilacqua N, Maeurer M, Zumla A, Locatelli F, De Benedetti F, Palmieri F, Marchioni L, Capobianchi M, D' Offizi G, Petrosillo N, Antinori A, Nicastrì E, Ippolito G, Agrati C. An Inflammatory Profile Correlates With Decreased Frequency of Cytotoxic Cells in Coronavirus Disease 2019. *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America.* 2020. 71, 16, 2272 - 2275
22. Wang Z, Yang X, Zhou Y, Sun J, Liu X, Zhang J, Mei X, Zhong J, Zhao J, Ran P. COVID-19 Severity Correlates with Weaker T-Cell Immunity, Hypercytokinemia, and Lung Epithelium Injury. *Am J Respir Crit Care Med.* 2020. 202, 4, 606–610.
23. Ansari A, Arya R, Sachan S, Nath Ja S, Kalia A, Lall A, Sette A, Grifoni A, Weiskopf D, Coshic P, Sharma A, Gupta N. Immune Memory in Mild COVID-19 Patients and Unexposed Donors Reveals Persistent T cell Responses After SARS-CoV-2 Infection. *Front Immunol.* 2021. 12, 636768.
24. Wang Z, Yang X, Zhong J, Zhou Y, Tang Z, Zhou H, He J, Mei X, Tang Y, Lin B, Chen Z, McCluskey J, Yang J, Corbett A, Ran P. Exposure to SARS-CoV-2 generates T-cell memory in the absence of a detectable viral infection. *Nat Commun.* 2021. 12, 1724.
25. Calvo-Calle J, Strug I, Nastke M, Baker S, Stern L. Human CD4+ T Cell Epitopes from Vaccinia Virus Induced by Vaccination or Infection. *PLoS Pathog.* 2007. 3, 10, e144.

26. Becerra-Artiles A, Dominguez-Amorocho O, Stern L, Calvo-Calle J. A Simple Proteomics-Based Approach to Identification of Immunodominant Antigens from a Complex Pathogen: Application to the CD4 T Cell Response against Human Herpesvirus 6B. *PLoS ONE*. 2015. 10, 11, e0142871.
27. Rammensee H, Bachmann J, Emmerich N, Bachor O, Stevanović S. SYFPEITHI: database for MHC ligands and peptide motifs. *Immunogenetics*. 1999. 50, 3-4, 213-9.
28. Reynisson B, Alvarez B, Paul S, Peters, B, Nielsen, M. NetMHCpan-4.1 and NetMHCIIpan-4.0: Improved predictions of MHC antigen presentation by concurrent motif deconvolution and integration of MS MHC eluted ligand data. *Nucleic Acids Res*. 2020. 48, W1, W449-W454.
29. Rodríguez L, Giraldo M, García N, Velásquez L, París S, Álvarez C, García L. Frecuencias alélicas, genotípicas y haplotípicas HLA-A, HLA-B, HLA-DRB1 en donantes fallecidos, Medellín, Colombia. *Biomédica*. 2007. 27, 4.
30. Ossa H, Manrique A, Quintanilla S, Peña A. Polimorfismos del sistema HLA (loci A\*, B\* y DRB1\*) en población colombiana. *Nova* 2007. 5, 7.
31. Nielsen M, Lundegaard C, Wornig P, Hvid CS, Lamberth K, Buus S, Brunak S, Lund O. Improved prediction of MHC class I and class II epitopes using a novel Gibbs sampling approach. *Bioinformatics*. 2004. 20, 9, 1388-1397.
32. Tarke A, Sidney J, Kidd C, Dan J, Ramirez S, Yu E, Mateus J, Antunes R, Moore E, Rubiro P, Methot N, Phillips E, Mallal S, Fraizer A, Rawlings S, Greenbaum J, Peters B, Smith D, Crotty S, Weiskopf D, Grifoni A, Sette A. Comprehensive analysis of T cell immunodominance and immunoprevalence of SARS-CoV-2 epitopes in COVID-19 cases. *Cell Rep Med*. 2021. 2, 2, 100204.
33. Mateus J, Grifoni A, Tarke A, Sidney J, Ramirez S, Dan J, Burger Z, Rawlings S, Smith D, Phillips E, Mallal S, Lammers M, Rubiro P, Quiambao L, Sutherland A, Yu E, Antunes R,

Greenbaum J, Frazier A, Markmann A, Premkumar L, de Silva A, Peters B, Crotty S, Sette A, Weiskopf D. Selective and cross-reactive SARS-CoV-2 T cell epitopes in unexposed humans. *Science*. 2020. 370, 6512, 89-94.

## Siglas

ACE 2: enzima convertidora de angiotensina 2

AIM: Activation Induced Marker

ARN: ácido ribonucleico

APC: presentadoras de antígenos profesionales

HLA: antígenos leucocitarios humanos (por sus siglas en inglés)

IEDB: inmune epitope database

IS: índice de estimulación

MHC-II: complejo mayor de histocompatibilidad de clase II

NSP: proteínas no estructurales

OMS: Organización Mundial de la Salud

ORF: Open Reading Frame (marcos de lectura abiertos)

Proteína S: proteína Spike

Proteína M: proteína de membrana

Proteína N: proteína de nucleocápside

Proteína E: proteína de envoltura

SARS-CoV-2: Coronavirus de tipo 2

SDRA: síndrome de dificultad respiratoria aguda

UTR: Untranslated Region (regiones no codificantes)



## Glosario

**Base de datos:** es una recopilación organizada de información o datos estructurados, que normalmente se almacena de forma electrónica en un sistema informático.

**Bioinformática:** la bioinformática, en relación con la genética y la genómica, es una subdisciplina científica que implica el uso de ciencias informáticas para recopilar, almacenar y analizar y diseminar datos e información biológicos, como secuencias de ADN y aminoácidos o anotaciones sobre esas secuencias.

**Células TCD4:** los linfocitos CD4 o conocidos también como T4, son glóbulos blancos que combaten infecciones y desempeñan un papel importante en el sistema inmunitario.

**COVID-19:** Enfermedad respiratoria muy contagiosa causada por el virus SARS-CoV-2.

**Coronavirus:** extensa familia de virus que causan infecciones respiratorias que pueden ir desde el resfriado común hasta enfermedades más graves.

**Estudios ex-vivo:** Afuera del cuerpo vivo. Se refiere a un procedimiento médico mediante el cual se extirpa un órgano, células o tejidos de un cuerpo vivo para un tratamiento o procedimiento, y luego se devuelven al cuerpo vivo

**Estudios in vitro:** Un ensayo realizado in vitro («en vidrio») quiere decir que se realiza fuera de un organismo vivo y normalmente en tejidos, órganos o células aislados.

**Epitope:** parte de una molécula que será reconocido por un anticuerpo al cual se unirá

**Infección:** invasión y multiplicación de gérmenes en el cuerpo. Los gérmenes pueden ser bacterias, virus, hongos u otros microorganismos.

**Inmunodominancia:** Característica que tienen algunos sitios del virus para desatar la réplica del organismo infectado, es decir estas porciones del virus dominan sobre otras partes del mismo, la acción de desencadenar la respuesta inmunológica, se dice entonces que tienen inmunodominancia.

**Péptido:** moléculas que están formadas por la unión de varios aminoácidos mediante enlaces peptídicos. Presentes en la naturaleza y son responsables de un gran número de funciones, muchas de las cuales aún no se conocen.

**Proteína:** Las proteínas son moléculas grandes y complejas que cumplen muchas funciones importantes en el cuerpo. Son vitales para la mayoría de los trabajos que realizan las células y son necesarias para mantener la estructura, función y regulación de los tejidos y órganos del cuerpo. Una proteína está formada por una o más cadenas largas, plegadas de aminoácidos (cada una llamada polipéptido), cuyas secuencias están determinadas por la secuencia de ADN del gen que codifica la proteína.