



CENTRO DE INVESTIGACIONES BIOLÓGICAS
DEL NOROESTE, S.C.

Programa de Estudios de Posgrado

**DESARROLLO Y EVALUACIÓN *IN SILICO* DE UN PROTOTIPO
DE VACUNA CONTRA EL SARS-CoV-2 (COVID-19).**

T E S I S

Que para obtener el grado de

Maestro en Ciencias

Uso, Manejo y Preservación de los Recursos Naturales
(Orientación en Agricultura Sustentable)

P r e s e n t a

Imanol Fuentes Corona

La Paz, Baja California Sur, marzo de 2023.

ACTA DE LIBERACIÓN DE TESIS

En la Ciudad de La Paz, B. C. S., siendo las 11:42 horas del día 15 del Mes de febrero del 2023, se procedió por los abajo firmantes, miembros de la Comisión Revisora de Tesis avalada por la Dirección de Estudios de Posgrado y Formación de Recursos Humanos del Centro de Investigaciones Biológicas del Noroeste, S. C., a liberar la Tesis de Grado titulada:

"Desarrollo y evaluación *in silico* de un prototipo de vacuna contra SARS-CoV-2 (COVID-19)"

Presentada por el alumno:

Imanol Fuentes Corona

Aspirante al Grado de MAESTRO EN CIENCIAS EN EL USO, MANEJO Y PRESERVACIÓN DE LOS RECURSOS NATURALES CON ORIENTACIÓN EN **Agricultura Sustentable**

Después de intercambiar opiniones los miembros de la Comisión manifestaron su **APROBACIÓN DE LA TESIS**, en virtud de que satisface los requisitos señalados por las disposiciones reglamentarias vigentes.

LA COMISIÓN REVISORA



Dr. Carlos Eliud Angulo Valadez
Co-Director de Tesis




Dra. Martha Candelaria Reyes Becerril
Co-Directora de Tesis



Dra. Elizabeth Monreal Escalante
Co-Tutora de Tesis



Dra. Luvia Enid Sánchez Torres
Co-Tutora de Tesis



Dra. Gracia Alicia Gómez Anduro,
Directora de Estudios de Posgrado y
Formación de Recursos Humanos.

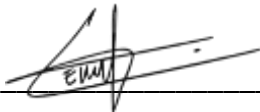
La Paz, Baja California Sur, a 15 de febrero de 2023.

Los miembros del comité de tesis del estudiante Imanol Fuentes Corona del Programa de Maestría en Uso, Manejo y Preservación de los Recursos Naturales, revisamos el contenido de la tesis y otorgamos el Vo.Bo. dado que la tesis no representa un plagio de otro documento como lo muestra el reporte de similitud realizado:


- Herramienta antiplagio:
iThenticate
- Filtros utilizados:
Citas y bibliografía excluida
- Porcentajes de similitud:
8%

	Citas excluidas Bibliografía excluida	8% SIMILAR
Resumen de Coincidencias		
1	Internet 140 palabras hdl.handle.net	1%
2	Internet 128 palabras Copiado el 05-Ago-2022 revistasojs.ucaldas.edu.co	<1%
3	Internet 65 palabras Copiado el 09-Feb-2023 www.researchgate.net	<1%
4	Internet 59 palabras Copiado el 17-Oct-2022 cibnor.repositorioinstitucional.mx	<1%
5	Internet 52 palabras Copiado el 14-Nov-2022 repositorio.uchile.cl	<1%
6	Internet 52 palabras Copiado el 18-Ago-2021 www.tandfonline.com	<1%
7	Internet 49 palabras Copiado el 18-Nov-2020 patents.google.com	<1%

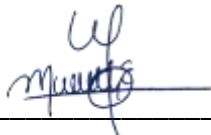
Firmas del comité



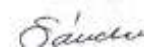
Dr. Carlos Eliud Angulo Valadez



Dra. Martha Candelaria Reyes Becerril



Dra. Elizabeth Monreal Escalante



Dra. Luvia Enid Sánchez Torres

Conformación de Comités

Comité Tutorial

Dr. Carlos Eliud Angulo Valadez
Centro de Investigaciones Biológicas del Noroeste, S.C.
Co-Director de Tesis

Dra. Martha Candelaria Reyes Becerril
Centro de Investigaciones Biológicas del Noroeste, S.C.
Co-Directora de Tesis

Dra. Elizabeth Monreal Escalante
Centro de Investigaciones Biológicas del Noroeste, S.C.
Co-Tutora de Tesis

Dra. Luvia Enid Sánchez Torres
Escuela Nacional de Ciencias Biológicas Instituto Politécnico Nacional
Co-Tutora de Tesis

Comité Revisor de Tesis

Dr. Carlos Eliud Angulo Valadez
Dra. Martha Candelaria Reyes Becerril
Dra. Elizabeth Monreal Escalante
Dra. Luvia Enid Sánchez Torres

Jurado de Examen

Dra. Martha Candelaria Reyes Becerril
Dra. Elizabeth Monreal Escalante
Dra. Luvia Enid Sánchez Torres

Suplente

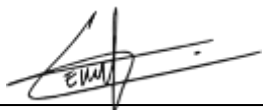
Dr. Carlos Eliud Angulo Valadez

Resumen

La pandemia de COVID-19 provocada por el virus SARS-CoV-2 ha causado millones de muertes. Las variantes emergentes del virus han reducido la eficacia de las vacunas por lo que se investigan nuevas alternativas. El SARS-CoV-2 posee proteínas estructurales (S, N, M y E) que han mostrado alta inmunogénicidad como antígenos candidatos para el diseño de vacunas. Mediante el uso de herramientas inmunoinformáticas se han generado propuestas de vacunas multiepitópicas basadas en estas proteínas. Por lo anterior, en el presente trabajo se planteó el diseño de vacunas mutiepitópicas contra SARS-CoV-2 basadas en las proteínas estructurales, que tuvieran una alta cobertura poblacional, fueran inmunogénicas y no presenten alta tasa de mutación. Con programas de predicción se predijeron y seleccionaron epítomos para células T CD8+ y T CD4+ de alta afinidad a los complejos principales de histocompatibilidad I y II (MHC I y II). Las regiones obtenidas fueron refinadas para mantener la mayor cobertura poblacional posible con epítomos antigénicos y que no contengan mutaciones de las variantes de interés y preocupación. La predicción de epítomos lineales para células B se realizó con 4 programas diferentes y se seleccionaron los sugeridos por 2 o más programas, que fueran antigénicos, tuvieran menos de 50% de residuos escondidos y que no contengan residuos en los que se hayan reportado mutaciones y sitios glicosilados. Las vacunas fueron diseñadas probando diferentes enlazadores y se mantuvieron los candidatos más antigénicos. Se adicionó el adyuvante PADRE, una etiqueta de histidinas, una señal de retención al retículo endoplásmico y un péptido señal. Las proteínas quiméricas se modelaron *ab initio* y posteriormente se refinaron. El factor general de calidad (ERRAT) y el porcentaje de residuos en regiones favorables de Ramachandran sirvieron para la selección y validación de los modelos tridimensionales. Se obtuvieron 4 diseños vacunales, cada uno basado en una proteína estructural. El diseño basado en la proteína S destacó por sobre los demás al obtener una cobertura poblacional completa con los epítomos utilizados. También, se incluyeron epítomos B lineales hacia regiones importantes de esta proteína como los dominios RBD y HR2. Para la proteína N se obtuvo una cobertura poblacional final casi completa para MHC I y >85% para MHC II. La proteína M tuvo una cobertura poblacional >80% con epítomos T CD8+, pero baja con epítomos T CD4+ (<60%). La proteína que tuvo menos cobertura poblacional para ambos MHC fue la proteína E, debido principalmente a su corta longitud. Las coberturas poblacionales finales se mantuvieron intactas en todos los casos al tomar en consideración las mutaciones de las nuevas variantes Ómicron (BA.2.12.2, BA.2.75, BQ.1 y XBB). Las propiedades fisicoquímicas y estructurales de los modelos tridimensionales sugirieron una buena calidad en las moléculas predichas para SARS2_E y SARS2_N y una calidad media para SARS2_M y SARS2_S. Los diseños multiepitópicos generados en esta investigación se predicen con una buena respuesta humoral y celular en la población mundial y tienen potencial de mantenerse vigentes ante el surgimiento de nuevas variantes del virus.

Palabras clave: SARS-CoV-2, vacuna multiepitópica, inmunoinformática

ORCID: 0000-0002-4467-431X



Dr. Carlos Eliud Angulo Valadez
Co-Director de Tesis



Dra. Martha Candelaria Reyes Becerril
Co-Directora de Tesis

Summary

COVID-19 pandemic provoked by the SARS-CoV-2 virus has caused millions of deaths. The emerging variants of the virus have reduced the efficacy of the vaccines. Because of this, new alternatives are needed. SARS-CoV-2 possess structural proteins (S, N, M and E), which have showed high immunogenicity as antigenic candidates for vaccine design. Using immunoinformatic tools, multiepitope vaccine proposals based on these proteins have been generated. Based on this, it's proposed to design antigenic multiepitope vaccines against SARS-CoV-2 based on its structural proteins, with high population coverage and avoiding variant mutations. T CD4+ and T CD8+ epitopes were selected with data retrieved from major histocompatibility complexes (MHC) binding affinity prediction programs. The obtained regions were refined for maintaining the highest possible population coverage with antigenic epitopes, avoiding mutated residues present in variants of interest and of concern. The linear B-cell epitope prediction was done in 4 different programs. Antigenic epitopes predicted by 2 or more programs that had less than 50% of buried residues and that avoided mutations and glycosylated sites were selected. The vaccines were designed to test different linker and select the most antigenic combination. The PADRE adjuvant amongst other vaccine components were added. The chimeric proteins were *ab initio* modeled and subsequently refined. The overall quality factor (ERRAT) and the percentage of Ramachandran favored residues were used for model selection and validation. The design based on the S protein stand out among the rest by obtaining complete population coverage against both MHC. It also includes B-cell epitopes from interest regions like RBD and HR2 domains. For the N protein, a total MHCI coverage was achieved with also >85% of MHCII coverage. M protein had a >80% with the T CD8+ epitopes and a low coverage (>60%) with the T CD4+ epitopes. The protein with the lowest population coverage for both MHC was E protein, due mainly to its small length. Final population coverages of all designs were intact after a new analysis considering new Omicron variants mutations (BA.2.12.1, BA.2.75, BQ.1 and XBB). The physicochemical and structural properties from the models suggested that SARS2_E and SARS2_N have a good quality, while SARS2_E and SARS2_S have average quality. The generated multiepitope designs are predicted to have a good humoral and cellular immune response in human population and have the potential to stay updated through the emergence of new variants.

Key words: SARS-CoV-2, multiepitope vaccine, immunoinformatic

ORCID: 0000-0002-4467-431X



Dr. Carlos Eliud Angulo Valadez
Co-Director de Tesis



Dra. Martha Candelaria Reyes Becerril
Co-Directora de Tesis

Dedicatoria

A mi mamá, papá y hermanita

A mis Rosas

A Mariana

A Billy y Mobu

A la Baja

A Mario

Agradecimientos

Al Consejo Nacional de Ciencia y Tecnología (CONACyT) por la beca de manutención otorgada No.: 1081842 que me permitió estudiar esta Maestría.

Al Centro de Investigaciones Biológicas del Noroeste, S.C. (CIBNOR), por permitirme realizar mis estudios y crecer profesional y personalmente.

A la Dirección de Estudios de Posgrado y Formación de Recursos Humanos y las personas que laboran ahí, por siempre atenderme con una sonrisa y apoyarme en lo que pudieran.

Al Dr. Carlos Eliud Angulo Valadez, por sus enseñanzas y apoyo durante todo este proceso. Por dejarme explorar libremente este mundo en el que me adentré, siempre guiando el camino.

A la Dra. Martha C. Reyes Becerril, por involucrarme en Inmuno-peques, lo que sin duda fue de lo más bonito de esta maestría. También por su gran apoyo en la parte final de este proceso, ha cumplido con creces lo que me dijo en la posada.

A la Dra. Elizabeth Monreal Escalante, por estar siempre disponible y por su paciencia. Por enseñarme tanto en el laboratorio, llegué sin saber nada y ahora sé que puedo ir a cualquier lado y podré con lo que sea.

A la Dra. Luvia Enid Sánchez Torres, por involucrarse en esto a pesar de la distancia. Por sus preguntas difíciles que siempre me hicieron cuestionarme lo que sé y buscar ampliar mis conocimientos.

Al M. C. Margarito Rodríguez, por la excelente organización del laboratorio y siempre estar al pendiente de las necesidades.

Al M. C. René Rebollar, M. C. Ángel Carrillo y la M. C. Delia Rojas, por siempre estar en la mejor disposición de permitirme usar los equipos y espacios a su cargo.

Al GIV, todos mis compañeros y compañeras del lab y de Inmuno-peques. Abel, Miriam, Cristian, Kevyn, Vero, Alonso, Armando, Sean, Eduardo, Lety, Omar, Marcio. Su presencia hizo más amenos mis días y sin saberlo me motivaron a seguir.

A Paul, por su apoyo y compañerismo durante todas las materias y luego en el laboratorio.

A Amy, por ser la mejor compañera de laboratorio. Por enseñarme, aprender y fallar conmigo.

A todas mis profesoras y profesores de la maestría, por compartir su conocimiento conmigo y esforzarse por hacerlo aún durante una pandemia.

A todo el personal de limpieza, técnico, choferes, seguridad y administrativo del CIBNOR, por hacer nuestro día a día más sencillo y mantener la excelencia de esta institución.

A mi papá y mamá, por hacerme la persona que soy hoy, por dar todo de ustedes para que yo tenga las herramientas que necesito en la vida. Por todo su amor incondicional. Les amo.

A mi hermana, porque, aunque somos muy diferentes siempre estamos el uno para el otro. Estoy muy orgulloso de la persona que eres, te amo.

A Pink, por ser mi segunda mamá, por siempre cuidarme, quererme y apapacharme.

A mi abuelita Rosa, por ser la abuela más chida que hay, por tu comida deliciosa y tu cariño.

A mi abuelito Nayo, por ser un pilar tan fuerte en la familia, por todo su apoyo y amor.

A Mariana, porque si no te hubiera conocido no hubiera vivido esta gran aventura y porque sin ti, no hubiera salido adelante. Por todo lo que haces por mí y por nosotros. Por estar en los peores momentos y motivarme a seguir. Por darme los mejores momentos y hacerme feliz todos los días, eres la mejor compañera. Te amo.

A Billy y Mobula, por darme tanto amor perruno. Por ser un soporte emocional enorme sin darse cuenta ni esperar nada a cambio. Por darle tranquilidad y locura a mi vida.

A Isa y Sisi, por ser unas perritas maravillosas y amorosas que, a pesar de ahora no verlas seguido, siempre me reciben con amor. Por cuidar a mi familia junto con Lolo.

A mis tías Mané y Gina, por siempre estar al pendiente de mí. Por darme tanto amor y buenos momentos siempre que las veo.

A Ethan, Dylan, José Carlos y Maya, porque más que primos somos hermanos. Por siempre estar para mí y por todas las risas.

A Brenda, Tlatoani, Metzery, Anibal y Meztli, por ser mis carnalitos y carnalitas. Por todas las locuras, desmadres y regaños que compartimos.

A Daniel, por ser el hermano que escogí. Por siempre estar para mí a pesar de la distancia. Por todos esos momentos y pláticas trasnochadas en el callejón desde que estábamos morros.

A Mario, por todo lo que me diste. Por todo el amor que me demostrabas, por mostrarme lados de mí que no conocía. Por crecer conmigo. Porque, aunque sé que jamás leerás esto y no verás la persona que soy y seré, siempre vivirá una parte de ti en mí.

A Uvas, por ser una inspiración enorme y por los buenos chistes y refranes.

A mi demás familia de Apan, por todas las porras que me echan y el amor que me tienen.

A mis tías Marité, Gugui, Rosa, Charito y Chela, por mandarme siempre su amor.

A mi tío Aldo, por las visitas tardías y las buenas pláticas que nos echamos.

A mis tíos Anibal, Arturo, Memo, Armando, a mis tías Lulú, Claudia, Anahí y Thelma y a mi madrina Lety, por sus consejos, pláticas, amor y apoyo cuando les necesité.

A Mariana, Vivi, Rodrigo y Pancho por ser los mejores compañeros de aventuras por la Baja.

A Felipe, Víctor, Rodrigo, Daniel y Ricardo por las buenas reuniones para jugar, beber, pescar, comer o simplemente compartir un buen rato y unas buenas risas.

A los House of Dragon (Citlali, Andrea, Majo, Vivi, Mariana, Rodrigo, Víctor, y Felipe), por las buenas cenitas de los domingos, noches de juegos, fiestas y demás buenos momentos que pasamos juntos.

A las demás personas chingonas que he conocido en La Paz (Valeria, Abel, Isis, Pancho, Mafe, Walter, Checo, Clara, Cristóbal, Andrea, Martha y muchos y muchas más), por hacer tan disfrutable la vida aquí.

A Aldo, Javier, Jorge y Nitzia, que cada que los veo parece que el tiempo no ha pasado y disfruto cada momento.

A Iván, por la buena amistad que tenemos a pesar de los años y la distancia.

A Imanol, por no darte por vencido, por enfrentarte a la adversidad y salir adelante, por recordar lo que vales y lo especial que eres.

A La Paz y la Baja, por llenarme el corazón y todos mis sentidos con su belleza.

A todas las personas que desde sus frentes combatieron la pandemia, porque gracias a ellas seguimos aquí, espero hacer honor a su esfuerzo con este trabajo.

Contenido

Resumen	i
Summary	ii
Dedicatoria	iii
Agradecimientos	iv
Contenido	vii
Lista de figuras	ix
Lista de tablas	xiii
1. INTRODUCCIÓN	1
1.1 Pandemia COVID-19	1
1.2 Virus SARS-CoV-2	1
1.3 Proteínas estructurales del SARS-CoV-2.....	2
1.3.1 Proteína S.....	2
1.3.2 Proteína N	4
1.3.3 Proteína M	5
1.3.4 Proteína E.....	5
1.4 Vacunas basadas en proteínas contra SARS-CoV-2 (COVID-19).....	5
2. ANTECEDENTES	7
2.1 Diseño inmunoinformático de la vacuna multiepitópica usando las proteínas estructurales.....	7
2.2 Importancia de las variantes de preocupación de SARS-CoV-2 para el diseño de vacunas .	9
3. JUSTIFICACIÓN	11
4. HIPÓTESIS	12
5. OBJETIVOS	13
5.1 Objetivo general	13
5.2 Objetivos particulares.....	13
6. MATERIAL Y MÉTODOS	14
6.1 Selección de epítomos.....	14
6.1.1 Análisis de epítomos para células T CD8+.....	14
6.1.2 Análisis de epítomos de células T CD4+	17
6.1.3 Análisis epítomos de células B	18
6.2 Refinamiento de las regiones	19
6.3 Diseños finales	20
6.3.1 Adición de enlazadores	20
6.3.2 Adición de elementos finales del diseño de la vacuna	21
6.3.3 Diseño de primers específicos	22
6.4 Caracterización <i>in silico</i> de las proteínas quiméricas	23
6.4.1 Evaluación de seguridad y caracterización de propiedades fisicoquímicas	23
6.4.2 Modelaje 3D.....	23
6.4.3 Selección de péptidos para producción de sueros hiperinmunes.....	24
7. RESULTADOS	25
7.1 Selección de epítomos y refinamiento	25
7.1.1 Análisis proteína E.....	25

7.1.2 Análisis proteína M	30
7.1.3 Análisis proteína N	34
7.1.4 Análisis proteína S.....	37
7.2 Análisis final de las regiones multiepitópicas seleccionadas	41
7.2.1 Cobertura poblacional final promedio.....	41
7.2.2 Cobertura poblacional por proteína	42
7.2.3 Afectaciones de nuevas variantes Ómicron	49
7.3 Diseños finales	50
7.3.1 Construcciones finales	50
7.3.2 Diseño de primers específicos	52
7.4 Caracterización <i>in silico</i> de las proteínas quiméricas	54
7.4.1 Evaluación de seguridad y caracterización de propiedades fisicoquímicas	54
7.4.2 Modelaje 3D.....	55
7.4.2.1 Modelaje de la Proteína E	55
7.4.2.2 Modelaje de la Proteína M.....	58
7.4.2.3 Modelaje de la Proteína N	61
7.4.2.4 Modelaje de la Proteína S	64
7.4.3 Selección de péptidos para la producción de sueros hiperinmunes	67
8. DISCUSIÓN	68
8.1 Selección y refinamiento de las regiones multiepitópicas	68
8.2 Diseños finales	74
8.3 Caracterización <i>in silico</i> de las proteínas quiméricas	76
9. CONCLUSIONES	80
10. LITERATURA CITADA	82
11. ANEXOS	91
Anexo A. Listas de epítomos finales	91
Anexo B. Construcción genética en los vectores pAlgevir y pBI121 y clonación	110
Anexo C. Otros productos y participaciones	117

Lista de figuras

- Figura 1.** Regiones de la proteína E seleccionadas para la construcción de la vacuna. Zonas amarillas indican donde hay exclusivamente epítomos T CD8+. Zonas azules indican donde hay exclusivamente epítomos T CD4+. Zonas verdes indican regiones en donde hay epítomos T CD8+ y T CD4+. Zona roja indica región que se añadió a la construcción sin cumplir todos los criterios de selección. Cuadrado violeta indica zona de epítomos lineales B. Los números unidos a las regiones marcan las posiciones que son extremos de estas. Números dentro de los cuadros violetas indican las posiciones de las regiones de epítomos B. Número al extremo izquierdo indica la posición en la que inicia ese segmento. 25
- Figura 2.** Regiones de la proteína M seleccionadas para la construcción de la vacuna. Zonas amarillas indican donde hay exclusivamente epítomos T CD8+. Zonas azules indican donde hay exclusivamente epítomos T CD4+. Zonas verdes indican regiones en donde hay epítomos T CD8+ y T CD4+. Zonas rojas indican regiones que se añadieron a la construcción sin cumplir todos los criterios de selección. Cuadrado violeta indica zona de epítomos lineales B. Los números unidos a las regiones marcan las posiciones que son extremos de estas. Números dentro de los cuadros violetas indican las posiciones de las regiones de epítomos B. Números al extremo izquierdo indican la posición en la que inicia ese segmento. 30
- Figura 3.** Regiones de la proteína N seleccionadas para la construcción de la vacuna. Zonas amarillas indican donde hay exclusivamente epítomos T CD8+. Zonas azules indican donde hay exclusivamente epítomos T CD4+. Zonas verdes indican regiones en donde hay epítomos T CD8+ y T CD4+. Zona roja indica región que se añadió a la construcción sin cumplir todos los criterios de selección. Los números unidos a las regiones marcan las posiciones que son extremos de estas. Números al extremo izquierdo indican la posición en la que inicia ese segmento..... 34
- Figura 4.** Regiones de la proteína S seleccionadas para la construcción de la vacuna. Zonas amarillas indican donde hay exclusivamente epítomos T CD8+. Zonas azules indican donde hay exclusivamente epítomos T CD4+. Zonas verdes indican regiones en donde hay epítomos T CD8+ y T CD4+. Zonas rojas indican regiones que se añadieron a la construcción sin cumplir todos los criterios de selección. Cuadrado violeta indica zona de epítomos lineales B. Los números unidos a las regiones marcan las posiciones que son extremos de estas. Números dentro de los cuadros violetas indican las posiciones de las regiones de epítomos B. Números al extremo izquierdo indican la posición en la que inicia ese segmento. 37
- Figura 5.** Representación gráfica de la cantidad de epítomos T CD8+ obtenidos de la proteína E que cumplieron los criterios de selección para cada alelo de MHCI. Las barras representan el promedio de la cantidad de epítomos de los tres programas utilizados que cubren el alelo correspondiente. Las barras de error en cada barra representan la desviación estándar de la misma. A) Alelos del grupo HLA-A, B) Alelos del grupo HLA-C, C) Alelos de grupo HLA-C, D) Alelos de los grupos HLA-E y HLA-G..... 43
- Figura 6.** Representación gráfica de la cantidad de epítomos T CD8+ obtenidos de la proteína M que cumplieron los criterios de selección para cada alelo de MHCI. Las barras representan el promedio de la cantidad de epítomos de los tres programas utilizados que cubren el alelo correspondiente. Las barras de error en cada barra representan la desviación estándar de la misma. A) Alelos del grupo HLA-A, B) Alelos del grupo HLA-C, C) Alelos de grupo HLA-C, D) Alelos de los grupos HLA-E y HLA-G..... 44

Figura 7. Representación gráfica de la cantidad de epítomos T CD8+ obtenidos de la proteína N que cumplieron los criterios de selección para cada alelo de MHC I. Las barras representan el promedio de la cantidad de epítomos de los tres programas utilizados que cubren el alelo correspondiente. Las barras de error en cada barra representan la desviación estándar de la misma. A) Alelos del grupo HLA-A, B) Alelos del grupo HLA-C, C) Alelos de grupo HLA-C, D) Alelos de los grupos HLA-E y HLA-G. 45

Figura 8. Representación gráfica de la cantidad de epítomos T CD8+ obtenidos de la proteína S que cumplieron los criterios de selección para cada alelo de MHC I. Las barras representan el promedio de la cantidad de epítomos de los tres programas utilizados que cubren el alelo correspondiente. Las barras de error en cada barra representan la desviación estándar de la misma. A) Alelos del grupo HLA-A, B) Alelos del grupo HLA-C, C) Alelos de grupo HLA-C, D) Alelos de los grupos HLA-E y HLA-G. 46

Figura 9. Representación gráfica de la cantidad de epítomos T CD4+ obtenidos de la proteína E que cumplieron los criterios de selección para cada alelo de MHC II. Las barras representan el promedio de la cantidad de epítomos de los tres programas utilizados que cubren el alelo correspondiente. No se grafica la desviación estándar por falta de un tercer programa que provea datos. A) Alelos del grupo HLA-DRB, B) Alelos de los grupos HLA-DP y HLA-DQ. 47

Figura 10. Representación gráfica de la cantidad de epítomos T CD4+ obtenidos de la proteína M que cumplieron los criterios de selección para cada alelo de MHC II. Las barras representan el promedio de la cantidad de epítomos de los tres programas utilizados que cubren el alelo correspondiente. No se grafica la desviación estándar por falta de un tercer programa que provea datos. A) Alelos del grupo HLA-DRB, B) Alelos de los grupos HLA-DP y HLA-DQ. 47

Figura 11. Representación gráfica de la cantidad de epítomos T CD4+ obtenidos de la proteína N que cumplieron los criterios de selección para cada alelo de MHC II. Las barras representan el promedio de la cantidad de epítomos de los tres programas utilizados que cubren el alelo correspondiente. No se grafica la desviación estándar por falta de un tercer programa que provea datos. A) Alelos del grupo HLA-DRB, B) Alelos de los grupos HLA-DP y HLA-DQ. 48

Figura 12. Representación gráfica de la cantidad de epítomos T CD4+ obtenidos de la proteína S que cumplieron los criterios de selección para cada alelo de MHC II. Las barras representan el promedio de la cantidad de epítomos de los tres programas utilizados que cubren el alelo correspondiente. No se grafica la desviación estándar por falta de un tercer programa que provea datos. A) Alelos del grupo HLA-DRB, B) Alelos de los grupos HLA-DP y HLA-DQ. 49

Figura 13. Modelo tridimensional inicial y refinado de la proteína SARS2_E. A) Modelo sugerido por programa I-TASSER, C-score: -2.60. B) Modelo sugerido con programa GalaxyRefine2. 55

Figura 14. Gráfico de Ramachandran de SARS2_E generado por PROCHECK. Zonas rojas son “regiones favorables”. Zonas amarillas son “regiones adicionales permitidas”. Zonas color beige son “regiones extendidas permitidas”. Zonas blancas son “regiones no permitidas”. A/a/~a indican estructuras de hélices α dextrógiras. B/b/~b indican estructuras de láminas β . L/l/~l indican estructuras de hélices α levógiras. p/~p Indican estructuras de hélices ϵ . ■ Residuos que no son prolina. ▲ Prolina. ▫ Residuos en regiones no permitidas y extendidas permitidas, descritos con su respectiva abreviación. 57

Figura 15. Modelo tridimensional inicial y refinado de la proteína SARS2_M. A) Modelo sugerido por programa I-TASSER, C-score: -2.68. B) Modelo sugerido con programa GalaxyRefine2. 58

- Figura 16.** Gráfico de Ramachandran de SARS2_M generado por PROCHECK. Zonas rojas son “regiones favorables”. Zonas amarillas son “regiones adicionales permitidas”. Zonas color beige son “regiones extendidas permitidas”. Zonas blancas son “regiones no permitidas”. A/a/~a indican estructuras de hélices α dextrógiras. B/b/~b indican estructuras de láminas β . L/l/~l indican estructuras de hélices α levógiras. p/~p Indican estructuras de hélices ϵ . ■ Residuos que no son prolina.▲Prolina. ■ Residuos en regiones no permitidas y extendidas permitidas, descritos con su respectiva abreviación. 60
- Figura 17.** Modelo tridimensional inicial y refinado de la proteína SARS2_N. A) Modelo sugerido por programa I-TASSER, C-score: -4.81. B) Modelo sugerido con programa GalaxyRefine2. 61
- Figura 18.** Gráfico de Ramachandran de SARS2_N generado por PROCHECK. Zonas rojas son “regiones favorables”. Zonas amarillas son “regiones adicionales permitidas”. Zonas color beige son “regiones extendidas permitidas”. Zonas blancas son “regiones no permitidas”. A/a/~a indican estructuras de hélices α dextrógiras. B/b/~b indican estructuras de láminas β . L/l/~l indican estructuras de hélices α levógiras. p/~p Indican estructuras de hélices ϵ . ■ Residuos que no son prolina.▲Prolina. ■ Residuos en regiones no permitidas y extendidas permitidas, descritos con su respectiva abreviación. 63
- Figura 19.** Modelo tridimensional inicial y refinado de la proteína SARS2_S. A) Modelo sugerido por programa I-TASSER, C-score: -2.68. B) Modelo sugerido con programa GalaxyRefine. 65
- Figura 20.** Gráfico de Ramachandran de SARS2_S generado por PROCHECK. Zonas rojas son “regiones favorables”. Zonas amarillas son “regiones adicionales permitidas”. Zonas color beige son “regiones extendidas permitidas”. Zonas blancas son “regiones no permitidas”. A/a/~a indican estructuras de hélices α dextrógiras. B/b/~b indican estructuras de láminas β . L/l/~l indican estructuras de hélices α levógiras. p/~p Indican estructuras de hélices ϵ . ■ Residuos que no son prolina.▲Prolina. ■ Residuos en regiones no permitidas y extendidas permitidas, descritos con su respectiva abreviación. 66
- Figura 21.** Digestión genes CIB-GIVac en pUC57 de células *E. coli* TOP 10 transformadas para su inserción en pAlgevir. Tamaño de fragmentos: SARS2_E = 333pb; SARS2_M = 657pb; SARS2_N = 759pb; SARS2_S = 999pb. Carril 1: Marcador 100pb; Carril 2: SARS2_E; Carril 4: SARS2_M; Carril 6: SARS2_N; Carril 8: Mezcla de reacción 1; Carril 9: Marcador 100pb; Carril 10: SARS2_S; Carril 11: Mezcla de reacción 2. 112
- Figura 22.** Digestión genes CIB-GIVac en pUC57 de células de *E. coli* TOP 10 transformadas para su inserción en pBI121. Tamaño de fragmentos: SARS2_E = 333pb; SARS2_M = 657pb; SARS2_N = 759pb; SARS2_S = 999pb. Carril 1: Marcador 100pb; Carril 2: Mezcla de reacción1; Carril 3: SARS2_M; Carril 4: SARS2_N; Carril 5: Marcador 100pb; Carril 6: Mezcla de reacción 2; Carril 7: SARS2_E; Carril 8: SARS2_S. 113
- Figura 23.** Calidad de ADN de células TOP 10 de *E. coli* transformadas con genes CIB-GIVac en vectores pAlgevir y pBI121. Tamaños esperados: pAlgevir > 17,000; pBI121 > 15,000. Carril 1: Marcador 1Kb Plus; Carril 2: SARS2_E Algevir; Carril 3: SARS2_E pBI121; Carril 5: SARS2_M Algevir; Carril 6: SARS2_M pBI121; Carril 8: SARS2_N Algevir; Carril 9: SARS2_N pBI121; Carril 11: SARS2_S Algevir; Carril 12: SARS2_S pBI121. 114
- Figura 24.** PCR para confirmación de cultivos de *E. coli* TOP 10 transformada. A) SARS2_N. Tamaño de amplicón: 652pb. Carril 1: Marcador 100pb; Carril 2: Mezcla de reacción; Carril 3: Muestra S2NA (Algevir); Carril 4: Muestra S2NB (pBI121). B) SARS2_M y SARS2_S. Tamaños de amplicón: 504pb (SARS2_M) y 615pb (SARS2_S) . Carril 1: Marcador 100pb; Carril 2: Mezcla de

reacción M; Carril 3: Muestra S2MA (pAlgevir); Carril 4: Muestra S2MB (pBI121); Carril 5: Blanco; Carril 6: Mezcla de reacción S; Carril 7: Muestra S2SA (Algevir); Carril 8: Muestra S2SB (pBI121).	115
Figura 25. Participaciones Inmuno-Peques. A) Taller “COVID-19: ¿En dónde vamos con las variantes?” en la Semana de Seguridad Laboral 2022 del CIBNOR. B) 1er Taller virtual de actualización para docentes “Vacunas: Un enfoque hacia las vacunas contra COVID-19” (39 horas). C) 2do Taller virtual de actualización para docentes “Vacunas y consejos prácticos para prevenir la COVID-19”.	117
Figura 26. Artículo de divulgación “¿Podríamos anticiparnos a otras pandemias? Una Salud: Una estrategia integral de prevención y control”. Coautor en revista “Recursos Naturales y Sociedad”, 2021. Vol. 7 (3): 129-145. https://10.18846/renaysoc.2021.07.07.03.0010	117
Figura 27. Participación semanas de posgrado en Baja California Sur. A) XX Semana de Posgrado (2021). B) XXI Semana de Posgrado (2022).....	118
Figura 28. Curso “Requisitos para el registro sanitario de nuevos medicamentos (vacunas) ante COFEPRIS”. Curso en línea de 20 h, 2021.....	118

Lista de tablas

Tabla 1. Alelos utilizados en el análisis de epítomos para células T CD8+.	15
Tabla 2. Variantes de SARS-CoV-2 consideradas en este estudio.	16
Tabla 3. Alelos utilizados en el análisis de epítomos para células T CD4+.	17
Tabla 4. Cantidad de epítomos obtenidos dentro de las regiones seleccionadas de la proteína E.	25
Tabla 5. Epítomos para células B sugeridos por al menos 2 programas en la proteína E.	26
Tabla 6. Valoración de epítomos B lineales sugeridos por al menos 2 programas por valor de antigenicidad y porcentaje de residuos expuestos en la proteína E.	27
Tabla 7. Porcentajes de coberturas poblacionales de alelos MHCI de la secuencia seleccionada para la proteína E.	28
Tabla 8. Porcentajes de coberturas poblacionales de alelos MHCII iniciales y mejoradas con el refinamiento de la secuencia seleccionada para la proteína E.	29
Tabla 9. Cantidad de epítomos obtenidos dentro de las regiones seleccionadas de la proteína M.	30
Tabla 10. Epítomos para células B sugeridos por al menos 2 programas en la proteína M.	31
Tabla 11. Valoración de epítomos B lineales sugeridos por al menos 2 programas por valor de antigenicidad y porcentaje de residuos expuestos en la proteína M.	32
Tabla 12. Porcentajes de coberturas poblacionales de alelos MHCI iniciales y mejoradas con el refinamiento de la secuencia seleccionada para la proteína M.	33
Tabla 13. Porcentajes de coberturas poblacionales de alelos MHCII de la secuencia seleccionada para la proteína M.	34
Tabla 14. Cantidad de epítomos obtenidos dentro de las regiones seleccionadas de la proteína N.	35
Tabla 15. Porcentajes de coberturas poblacionales de alelos MHCI iniciales y mejoradas con el refinamiento de la secuencia seleccionada para la proteína N.	36
Tabla 16. Porcentajes de coberturas poblacionales de alelos MHCII iniciales y mejoradas con el refinamiento de la secuencia seleccionada para la proteína N.	37
Tabla 17. Cantidad de epítomos obtenidos dentro de las regiones seleccionadas de la proteína S.	38
Tabla 18. Epítomos para células B que cumplen todos los requerimientos en la proteína S con sus respectivos valores de antigenicidad y porcentaje de residuos escondidos.	39
Tabla 19. Coberturas poblacionales de alelos MHCI iniciales y mejoradas con el refinamiento de la secuencia obtenida para la proteína S.	40
Tabla 20. Porcentajes de coberturas poblacionales finales de alelos MHCII de la secuencia seleccionada para la proteína S.	41
Tabla 21. Porcentajes de coberturas poblacionales finales promedio para MHCI.	41
Tabla 22. Porcentajes de coberturas poblacionales finales promedio para MHCII.	42
Tabla 23. Cantidad de epítomos afectados por cada mutación de las nuevas variantes Ómicron.	50
Tabla 24. Comparativa entre epítomos totales bajo los criterios de selección planteados y el nuevo total afectado por las nuevas variantes Ómicron.	50
Tabla 25. Diseños finales generados para producir la vacuna CIB-GIVac.	51

Tabla 26. Primers elegidos para PCR y qPCR de los genes con sus correspondientes datos tomados de Primer3 y Primer-BLAST	53
Tabla 27. Propiedades fisicoquímicas de los cuatro diseños finales para CIB-GIVac.	55
Tabla 28. Valores provistos por GalaxyRefine2 para modelos inicial y refinado de SARS2_E.	56
Tabla 29. Desglose de distribución de los residuos de SARS2_E en las regiones del gráfico de Ramachandran y factor general de calidad.	58
Tabla 30. Valores provistos por GalaxyRefine2 para modelos inicial y refinado de SARS2_M.	59
Tabla 31. Desglose de distribución de los residuos de SARS2_M en las regiones del gráfico de Ramachandran y factor general de calidad.	61
Tabla 32. Valores provistos por GalaxyRefine2 para modelos inicial y refinado de SARS2_N.	62
Tabla 33. Desglose de distribución de los residuos de SARS2_N en las regiones del gráfico de Ramachandran y factor general de calidad.	64
Tabla 34. Valores provistos por GalaxyRefine para modelos inicial y refinado de SARS2_S.	65
Tabla 35. Desglose de distribución de los residuos de SARS2_S en las regiones del gráfico de Ramachandran y factor general de calidad.	67
Tabla 36. Péptidos diseñados para la producción de sueros hiperinmunes.	67
Tabla 37. Epítomos para células T CD8+ seleccionados en la proteína E.	91
Tabla 38. Epítomos para células T CD4+ seleccionados en la proteína E.	92
Tabla 39. Epítomos para células T CD8+ seleccionados en la proteína M.	94
Tabla 40. Epítomos para células T CD4+ seleccionados en la proteína M.	96
Tabla 41. Epítomos para células T CD8+ seleccionados en la proteína N.	98
Tabla 42. Epítomos para células T CD4+ seleccionados en la proteína N.	100
Tabla 43. Epítomos para células T CD8+ seleccionados en la proteína S.	105
Tabla 44. Epítomos para células T CD4+ seleccionados en la proteína S.	107
Tabla 45. Epítomos para células B sugeridos por al menos dos programas que contienen el residuo T376 en la proteína S.	110
Tabla 46. Secuenciación de productos de PCR obtenidos de ADN de células de <i>E. coli</i> TOP 10 transformadas con los vectores pBI121 y pAlgevir.	116

1. INTRODUCCIÓN

1.1 Pandemia COVID-19

El 12 de diciembre de 2019, el primer paciente de lo que se convertiría en una pandemia global fue hospitalizado (Wu *et al.*, 2020; Zhou *et al.*, 2020). Días después, el 1ro de enero de 2020, la Organización Mundial de la Salud (OMS) contactó a las autoridades de la República Popular China para solicitar información y apoyo sobre los casos de neumonía atípica reportados en Wuhan, provincia de Hubei, China central (World Health Organization [WHO], 2020a). Más tarde, el 11 de febrero de 2020, la OMS nombró a esta enfermedad COVID-19 (WHO, 2020b).

El patógeno causante de esta enfermedad fue identificado como un virus de la familia *Coronaviridae* tentativamente llamado 2019-nCov y posteriormente renombrado como SARS-CoV-2 (Coronaviridae Study Group of the International Committee on Taxonomy of Viruses *et al.*, 2020). Este virus se propaga a través de gotas de saliva provenientes de la persona infectada y puede tomar hasta 14 días en desarrollar síntomas. Los síntomas más comunes son fiebre, tos seca y cansancio, aunque también pueden presentarse dolores de cuerpo, cabeza y garganta, diarrea, conjuntivitis, pérdida del olfato y gusto, y decoloración de los dedos. Dificultades para respirar, dolor o presión en pecho y pérdida del aliento o movilidad son síntomas serios que requieren atención médica inmediata (WHO, 2020c). Al 30 de enero de 2023, en el mundo ha habido más de 753 millones de casos y más de 6.8 millones de personas han fallecido (WHO, 2023a). En México, se han confirmado más de 7.6 millones de casos y más de 346,000 defunciones (Secretaría de Salud México, 2023). Además, las consecuencias de esta pandemia en los aspectos social, político, económico y muchos otros son imposibles de cuantificar.

1.2 Virus SARS-CoV-2

SARS-CoV-2 está clasificado dentro de la familia *Coronaviridae*, subfamilia *Orthocoronaviridae*, género *Betacoronavirus*, subgénero *Sarbecovirus*, que contiene a los coronavirus relacionados con el síndrome respiratorio agudo severo (International Committee on Taxonomy of Viruses, 2023). El origen exacto del virus permanece sin resolver, pero existe una identidad del 96.2% con el virus de murciélago RaTG13 (Zhou *et al.*, 2020). También, el Pangolin-CoV tiene una

identidad del 91.02% y dentro del dominio RBD sólo hay un aminoácido de diferencia (500H/500Q), que no es ninguno de los seis aminoácidos clave que interaccionan con el receptor humano de la enzima convertidora de la angiotensina 2 (ACE2) reportados por Wan et al. (2020). Estos descubrimientos sugieren que el Pangolin-CoV podría ser el origen común del SARS-CoV-2 y el RaTG13, aunque esto continúa en debate (Zhang *et al.*, 2020).

Su genoma está constituido por una cadena de ARN en sentido positivo (Wu *et al.*, 2020). La longitud del genoma es mayor a las 30 kb (Khailany *et al.*, 2020). ORF1a y ORF1b codifican 11 y 5 proteínas no estructurales (nsp1-nsp11; nsp12-nsp16) respectivamente codificadas en un sólo polipéptido denominado ORF1ab y cuenta con otros 6 ORF que se traducen en proteínas no estructurales (3a, 6, 7a, 7b, 8 and 10) (National Center for Biotechnology Information [NCBI], 2020). ORF3a (Konno *et al.*, 2020), ORF9b (Davidson *et al.*, 2020) y ORF14 (Jungreis *et al.*, 2020) son marcos de lectura abiertos (ORF) hipotéticos no considerados en la secuencia de referencia (NCBI: NC_0045512.2). Las proteínas estructurales que componen este virus son: espícula (S), envoltura (E), membrana (M) y nucleocápside (N) (Nakagawa y Miyazawa, 2020). La estructura consiste en el ARN interactuando con las proteínas N, rodeado del envoltorio conformado por las proteínas E y M y a donde están sujetas las proteínas S que sobresalen de la cápside, dando la característica de forma de corona al virus (Jackson *et al.*, 2022).

A pesar de que este nuevo coronavirus está relacionado directamente al SARS-CoV (77.5% similitud; Kim et al., 2020), la inserción de 12 nucleótidos en la unión S1-S2 lo proveen de un sitio de escisión polibásico de furina (Walls *et al.*, 2020) que juega un rol crítico en la infección y patogénesis (Johnson *et al.*, 2021). El número estimado de sustituciones en el genoma del SARS-CoV-2 de acuerdo con Nexstrain el 29 de enero de 2023 es de 28.434 sustituciones por año, con una velocidad de evolución de 0.8×10^{-4} sustituciones/sitio/año, la cual es menor comparada con otros virus de ARN como los de la influenza estacional (Nextstrain, 2023a).

1.3 Proteínas estructurales del SARS-CoV-2

1.3.1 Proteína S

La proteína estructural S es la que le da la característica forma de corona a los coronavirus. En el SARS-CoV-2, el gen que codifica esta proteína tiene un tamaño de alrededor de 3,821pb, lo que

se traduce generalmente en 1,273aa(GISAID, 2020).La conformación cuaternaria es un homotrímero de la proteína S (Chen *et al.*, 2020) Cada monómero es una glicoproteína de membrana de tipo 1 compuesto por dos subunidades: S1 y S2. La subunidad S2 es transmembranal y regula la entrada celular, mientras que la región S1 se encarga de la interacción con el receptor ACE2 (Hasöksüz *et al.*, 2020). A diferencia de otros *Betacoronavirus*, SARS-CoV-2 cuenta con un sitio de escisión de furina, una proteasa de membrana, entre S1 y S2, que potencia el procesamiento de la proteína posterior a la unión al receptor y hace que sea menos dependiente de otras proteasas como TMPRSS2 o las catepsinas L o B (Shang *et al.*, 2020). Esto, aunado a que la furina es abundantemente expresada en los pulmones y este virus infecta por medio de las vías respiratorias, podría explicar porque su transmisión ha sido tan exitosa (Coutard *et al.*, 2020).

Esta proteína se encuentra altamente glicosilada, a la fecha se han identificado al menos 22 sitios de posibles N-glicosilaciones de los cuales se han reportado ocupados como mínimo 17, además de la ocupación de otros 5 sitios de O-glicosilaciones (Watanabe *et al.*, 2020; Shajahan *et al.*, 2020).Estas glicosilaciones actúan a manera de escudo, cubriendo alrededor del 40% de su superficie, impidiendo la unión de anticuerpos neutralizantes (Grant *et al.*, 2020). Sin embargo, este porcentaje es menor a otros virus de su familia como el MERS-CoV, lo que indica que la respuesta humoral hacia este antígeno podría ser mejor que en otros virus de la familia (Gong *et al.*, 2021).

Durante del proceso de infección, la proteína se une mediante su dominio RBD al receptor ACE2 de las células (Yan *et al.*, 2021). Tras este paso, se procesa la proteína mediante proteasas que cortan ciertas regiones de la subunidad S-1 de la proteína (Shang *et al.*, 2020). Posterior a esto, la conformación de la estructura de la proteína cambia, formando una estructura intermedia similar a una lanza con el péptido fusión interaccionando con la membrana celular. Después del péptido fusión se encuentra la región HR1, luego la CH y finalmente la HR2 en la zona más cercana a la membrana del virus, y estas regiones no se encuentran glicosiladas (Mercurio *et al.*, 2021). La región HR2 jala la membrana del virus para unirse a la membrana celular a modo de que esta región quede al mismo nivel que la región HR1, las membranas se fusionan, se libera el péptido fusión y se inserta el material genético del virus al citoplasma celular (Cai *et al.*, 2020).

Es importante resaltar que, es en esta proteína donde se han presentado la mayor cantidad de mutaciones y es la que principalmente se ha utilizado para la distinción entre los diferentes linajes del virus. También, es el mayor inmunógeno utilizado para el desarrollo de vacunas (Samrat *et al.*, 2020). Todo esto en conjunto ha generado desafíos en el diseño de vacunas y presenta un camino para el desarrollo de nuevas generaciones de vacunas enfocadas en estructuras del virus con una menor tasa de mutación o basadas en las secuencias de los nuevos linajes circulantes como el Ómicron.

1.3.2 Proteína N

La única proteína estructural que no se encuentra embebida en la membrana es la proteína N. Esta tiene una longitud de 419aa, traducida de un gen que mide alrededor de 1,259pb (GISAID, 2020). Esta proteína se compone principalmente de dos dominios: N-terminal (NTD) y C-terminal (CTD), unidas por una región de enlace (LKR) que contiene una región rica en serina y arginina (SRD) (Zeng *et al.*, 2020). El dominio NTD interacciona con el ARN del virus, mientras que el dominio CTD interacciona con sus homólogos de otras proteínas para formar dímeros, tetrámeros y hexámeros (Chang *et al.*, 2013). El material genético junto con el oligómero de esta proteína forma un complejo de ribonucleoproteína, el cual interacciona con la proteína M para su empaquetamiento (Ye *et al.*, 2020).

Esta proteína también ha mostrado ser un antígeno importante tanto para la respuesta inmune como para ser usada en pruebas de diagnóstico. Tal como demuestran Zeng *et al.* (2020), existen respuestas de inmunoglobulinas (Ig) de tipo A, G y M en sueros de pacientes confirmados con COVID-19. Por otro lado, a lo largo de los meses en los que ese ha desarrollado la pandemia, las mutaciones se han hecho presentes en esta proteína. Los trabajos de (Rahman *et al.*, 2021) y Ye *et al.* (2020) mostraron que la región LKR es la más propensa a desarrollar mutaciones y se repiten en gran medida en los residuos 203 y 204. Hasta hoy, se han reportado más de 25 mutaciones presentes en las diferentes variantes de interés y preocupación clasificadas a lo largo de la pandemia para esa proteína (Nextstrain, 2023b).

1.3.3 Proteína M

El gen que codifica a la proteína M tiene una longitud de alrededor de 668pb, generando un péptido de 222aa (GISAID, 2020). Esta es la proteína más abundante en la envoltura y, con análisis *in silico*, se ha descrito que tiene una estructura transmembranal de triple hélice con una región corta en el extremo N-terminal hacia el exterior del virus y su región más larga al interior (Thomas, 2020). La proteína M tiene la característica de ser la única proteína estructural del virus que puede interactuar con las otras tres proteínas estructurales, siendo su función de estabilización del complejo proteína N-ARN probablemente la más importante (Naqvi *et al.*, 2020). También, es una de las proteínas involucradas en el proceso de ensamblaje del virus (Bianchi *et al.*, 2020). Si bien, la función de esta proteína no está entendida en su totalidad, se hipotetiza que, debido a su similitud con proteínas transportadoras de azúcares, podría transportar sacarosa hacia dentro del fagosoma que lo contiene dentro de la célula hospedera, propiciando la liberación del virus a la célula. También, podría estar relacionada con la alta glicosilación que se presenta en la proteína S (Thomas, 2020).

1.3.4 Proteína E

Con una longitud de 75aa, la proteína E es la más pequeña de las cuatro proteínas estructurales del SARS-CoV-2, siendo producto de un gen de 228pb (GISAID, 2020). Esta proteína se organiza principalmente en pentámeros concéntricos formando un poro en la envoltura del virus que puede estar orientado tanto con los dominios N-terminal como C-terminal al exterior de la partícula viral (Bianchi *et al.*, 2020). Este poro tiene su principal función en realizar intercambios iónicos entre el interior y exterior del virus (Mandala *et al.*, 2020). Dentro de la célula infectada, la proteína E se encuentra principalmente en el retículo endoplásmico (RE) y el aparato de Golgi participando en el proceso de unión, ensamblaje y tráfico de viriones infecciosos (Schoeman y Fielding, 2019).

1.4 Vacunas basadas en proteínas contra SARS-CoV-2 (COVID-19)

Ante la emergencia que representó la pandemia de COVID-19, la comunidad científica y médica se dio a la tarea de desarrollar vacunas para combatir el SARS-CoV-2. Gracias a los avances

tecnológicos y médicos de los últimos años, estas vacunas se han desarrollado en tiempo récord. Las vacunas de nueva generación poco a poco han sustituido a las tradicionales, como por ejemplo de virus atenuados. Esta nueva generación de vacunas requiere de la incorporación de antígenos específicos de los patógenos en vez de usar el organismo completo. Estas nuevas tecnologías pueden dividirse en vacunas basadas en proteínas y vacunas de vectores virales (Wang *et al.*, 2020).

Las vacunas basadas en proteínas que se encuentran en fase 3 actualmente utilizan diferentes tecnologías. Por ejemplo, la vacuna EpiVacCorona, desarrollada en Rusia, es una construcción de antígenos peptídicos de las proteínas del SARS-CoV-2 unidas a una proteína acarreadora adsorbida en un adyuvante de hidróxido de aluminio, aplicada en dos dosis separadas de 21 a 28 días (ClinicalTrials.gov, 2021). Anhui Zhifei desarrolló la vacuna ZF2001 que contiene un dímero del dominio RBD y que se aplica en tres dosis separadas por 30 días (ClinicalTrials.gov, 2020a). NVX-CoV2373 es la vacuna desarrollada por Novavax que contiene a la proteína S recombinante con el adyuvante Matrix-M1™ y que es suministrada en dos dosis separadas por 21 días (ClinicalTrials.gov, 2020b). Clover Pharmaceuticals ha creado la vacuna SCB-2019, que consta de una construcción trimérica de la conformación natural de la proteína S adyuvada con CpG1018/Aluminio, administrada en dos dosis separadas por 21 días (ClinicalTrials.gov, 2020c). El Instituto Finlay en Cuba ha desarrollado la vacuna SOBERANA-02, siendo esta un conjugado monovalente del dominio RBD a toxoide tetánico (CECMED, 2020) con un esquema de dos dosis separadas por 28 días (Registro Público Cubano de Ensayos Clínicos, 2021). Este mismo instituto también ha desarrollado una vacuna de subunidad basada en el dominio RBD y producida en la levadura *Pichia pastoris* llamada ABDALA (Hernández-Bernal *et al.*, 2022; Más-Bermejo *et al.*, 2022). Esta vacuna se ha administrado a la población cubana y actualmente es administrada como dosis de refuerzo en territorio mexicano (Díaz-Ortega, 2022).

2. ANTECEDENTES

2.1 Diseño inmunoinformático de la vacuna multiepitópica usando las proteínas estructurales

Se ha demostrado que la respuesta inmune de anticuerpos contra el virus SARS-CoV ha sido efectiva, pero de corta duración en pacientes infectados (Ahmed *et al.*, 2020; Achiron *et al.*, 2021). Por otro lado, la prevalencia de la respuesta de células T contra el género *Betacoronavirus* se ha encontrado duradera. Existen estudios, como el de Le Bert *et al.* (2020), que han probado la prevalencia de respuesta de células T contra algunos antígenos del SARS-CoV-2 en pacientes previamente infectados con el virus SARS-CoV. Se determinó que, en pacientes positivos a COVID-19 que estuvieron infectados con SARS, hay respuesta de células T contra la proteína N y otras proteínas no estructurales del virus hasta 17 años después de la infección con SARS-CoV.

Channappanavar *et al.* (2014) probaron que las células T de memoria de ratones inmunizados con epítomos de la proteína S del virus del SARS generan una respuesta efectiva de protección ante una dosis letal del virus, 8 a 10 meses post-inmunización. Por otro lado, los títulos de anticuerpos IgG duraron 4 meses y los IgM e IgA estuvieron presentes hasta el sexto mes. Tras el reto infeccioso 10-11 meses después de la inmunización se comprobó la funcionalidad de las células T CD8+ por la producción de las citocinas efectoras IFN- γ y TNF- α . Además, la capacidad citotóxica de los linfocitos T CD8+ se confirmó con la expresión de la granzima B, una molécula citotóxica producida por estas células inmunes. En dicho estudio subrayan la importancia de la respuesta de células T CD4+ y de anticuerpos para generar una protección completa ante la infección.

Las herramientas inmunoinformáticas nos permiten realizar diseños de vacunas, reduciendo sustancialmente los costos y tiempo. Se han obtenido resultados de análisis *in silico* que indican que epítomos provenientes de la proteína S, N, M y E son inmunogénicos. También, se han generado una buena cantidad de diseños de vacunas utilizando epítomos provenientes de estas proteínas.

En el trabajo realizado por Cun *et al.* (2021) se analizaron las dos secuencias de las proteínas S y N del virus SARS-CoV-2 en búsqueda de epítomos afines al MHC I. Se predijeron los resultados

para 70 diferentes alelos pertenecientes a los sistemas del antígeno leucocitario humano (HLA) de tipos A, B y C. Todos estos presentes en al menos el 1% de la población humana y representando a los polimorfismos presentes en las moléculas del MHC. Se obtuvieron 90 epítomos para la proteína N y 106 para la proteína S que cubrían el 95.91 y 96.14% de los alelos analizados. Finalmente, se tamizaron para quedarse con 14 epítomos de alta afinidad que estén presentes en el SARS-CoV-2 y otros 12 coronavirus que afectan a humanos. Esta selección dio como resultado una cobertura poblacional del 89.6% y podría servir como una opción de vacuna contra otros coronavirus como el SARS-CoV o el MERS-CoV.

Se han generado prototipos de vacunas utilizando la proteína M como base. En un estudio se diseñó una vacuna multiepitópica con epítomos obtenidos para MHCI, MHCII y epítomos lineales de células B. Se analizó la secuencia de esta proteína en búsqueda de epítomos de alta afinidad con un programa para cada tipo de epítomo. Además, se predijo su antigenicidad, alergenicidad y toxicidad. Los epítomos que cumplieron los criterios planteados se utilizaron para construir la vacuna a la cual se le añadieron los adyuvantes β -defensina y PADRE. La primera estimula la respuesta inmune reclutando células T y dendríticas inmaduras interaccionando con receptores como los TLR o CCR6 (Lei *et al.*, 2019). Para PADRE se ha reportado el aumento de la respuesta de células T cooperadoras (Sarkar *et al.*, 2022). Cada epítomo y adyuvante fue unido mediante alguno de los siguientes enlazadores: KK, EAAAK, GPGPG y AAY. Los epítomos que contiene esta vacuna generan una cobertura poblacional del 94.06%. El modelo tridimensional de la vacuna mostró interacciones con los receptores TLR3. La secuencia optimizada fue clonada *in silico* al vector pET-28a (+). Finalmente, la vacuna mostró respuestas inmunes primarias, secundarias y terciarias en la simulación inmune *in silico* (Ayyagari *et al.*, 2020).

Dagur (2020) realizaron un estudio *in silico* de la proteína E del SARS-CoV-2 en busca de epítomos candidatos para el desarrollo de vacunas. En dicho trabajo se buscaron epítomos lineales de células B y de MHCI y II utilizando las herramientas de la Base de Datos de Epítomos Inmunes (IEDB, por sus siglas en inglés). Los epítomos obtenidos se caracterizaron en su antigenicidad, alergenicidad, toxicidad, capacidad de ser digeridos y varios parámetros bioquímicos. Se obtuvieron 3 epítomos para células B, 2 de MHCI y 1 de MHCII que cumplieron los criterios de selección y seguridad. Los autores destacan la posibilidad de aumentar la

especificidad de las vacunas multiepitópicas al enfocarse en epítomos que se encuentren expuestos fuera de la membrana del virus.

2.2 Importancia de las variantes de preocupación de SARS-CoV-2 para el diseño de vacunas

Durante el transcurso de la pandemia de COVID-19, diferentes variantes han circulado entre la población, producto de las mutaciones que el virus ha presentado. Organismos internacionales como la OMS, y los Centros para el Control y Prevención de Enfermedades de Estados Unidos (CDC) y Europa (ECDC) se dieron a la tarea de identificar y catalogar dichas variantes (Center for Disease Control and Prevention [CDC], 2023; European Centre for Disease Prevention and Control [ECDC], 2023; WHO, 2023b). A partir de esto, se les pudo dar un seguimiento y llevar un registro de cuales fueron más abundantes durante el tiempo. Este sistema de clasificación las ha encasillado en cuatro categorías principales: (1) variantes bajo monitoreo, (2) variantes de interés, (3) variantes de preocupación, y (4) variantes de alta consecuencia (ECDC, 2023; WHO, 2023b). Esta clasificación está basada en cómo se comportan las variantes en transmisibilidad, severidad y resistencia a tratamientos o vacunación (CDC, 2023). Al empezar a encontrarse más cantidad de variantes se empezó a hacer referencia de estas por el país en donde se identificaban inicialmente. Posteriormente, la OMS decidió nombrarlas con letras griegas para facilitar la comunicación y evitar actos de xenofobia y discriminación hacia las personas procedentes de los países en donde estas se habían originado (WHO, 2023b).

Muchas de las variantes que emergieron fueron consideradas como de interés inicialmente, donde se mantuvieron bajo vigilancia por los organismos internacionales. Sin embargo, pocas pasaron a ser consideradas de preocupación y afortunadamente ninguna ha pasado a ser de preocupación mayor. Históricamente, las variantes que han sido consideradas de interés son: (1) Alfa, (2) Beta, (3) Gamma, (4) Delta y (5) Ómicron. Durante el final del 2020 y la primera mitad del 2021, las variantes Alfa, Beta y Gamma fueron las predominantes en la población mundial (CDC, 2023). La variante Delta poco a poco se fue abriendo camino hasta convertirse en la predominante en el mundo debido a su capacidad de contagiar entre 5 a 8 personas por cada persona infectada, comparadas con las 1.4 a 2.4 del SARS-CoV-2 original (Achaiah y Subbarajsetty, 2020; Chavda y Apostolopoulos, 2022). En noviembre del 2021, el curso de la

pandemia cambió con la identificación de la primera variante Ómicron (linaje 21K), la cual fue directamente considerada como de preocupación (CDC, 2023; WHO, 2023b).

Ómicron rápidamente se propagó gracias a que es dos y media veces más rápida su transmisión que Delta, convirtiéndose en la predominante en tan sólo un mes ((Wu *et al.*, 2022). Esta variante ha cambiado con el tiempo, dividiéndose en varios sublinajes (ocho al 21 de enero del 2023), los cuales han proliferado en diferente medida desde entonces (Nextstrain, 2023a). Gracias a la aparición de estas variantes, y al avance en la vacunación mundial, las que anteriormente eran consideradas de interés se han visto desplazadas. Para marzo del 2022, las variantes Alfa, Beta y Gamma se dejaron de considerar como de interés y en junio del mismo año, Delta sufrió el mismo cambio (WHO, 2023b). A partir de entonces, los diferentes linajes de Ómicron han sido los predominantes en el mundo.

La proliferación de estas variantes, y la diferencia que estas tienen con el virus original, han preocupado a la comunidad científica y población en general. La principal preocupación, a raíz de que se ha comprobado que la severidad de la enfermedad causada por Ómicron es menor a Delta (Wolter *et al.*, 2022), es que las vacunas y tratamientos específicos no funcionen contra ésta. Algunas de las vacunas disponibles han sido evaluadas en su eficacia contra las variantes Delta y Ómicron para poder determinar si siguen protegiendo. En algunos casos, como la vacuna BNT162b2, se ha mostrado una neutralización menor ante estas variantes comparadas con el virus silvestre (Nemet *et al.*, 2022). Esto ha llevado a nuevas propuestas de vacunas, entre las que destacan las vacunas bivalentes de Pfizer/BioNTech y Moderna que contienen su formulación original junto con la correspondiente a las variantes Ómicron BA.4/BA.5 y que ya se aplican como refuerzo en los Estados Unidos (Swanson, 2022; Rosenblum *et al.*, 2022).

3. JUSTIFICACIÓN

La pandemia de COVID-19 alteró por completo el orden mundial, trayendo consigo inconmensurables consecuencias en todos los aspectos de la sociedad. Lastimosamente, esta enfermedad ha causado millones de muertes alrededor del mundo y le ha dejado con secuelas a miles de personas más. Gracias a los esfuerzos de la comunidad científica y médica se han logrado producir vacunas en tiempo récord, lo que ha servido para disminuir enormemente la tasa de letalidad y la severidad de la COVID-19. A la par de estos esfuerzos, el virus ha mutado, dando pie a nuevas variantes con mayor transmisibilidad. Estas mutaciones han generado disminución en la eficacia de las vacunas disponibles actualmente, ya que estas fueron diseñadas con base en el linaje original del SARS-CoV-2.

En los últimos años se han dado avances sustanciales en el campo de la inmunoinformática, los cuales permiten el desarrollo y caracterización *in silico* de diseños vacunales. El uso de estas herramientas computacionales reduce significativamente los costos de desarrollo y el tiempo para el diseño de vacunas, ya que permiten predecir las respuestas inmunes que podría tener el organismo hacia ciertos antígenos. También, disminuyen el gasto de insumos de laboratorio y reducen la cantidad de pruebas *in vivo* en comparación con el desarrollo tradicional de vacunas. Estas herramientas son una alternativa clave tanto para el desarrollo de prototipos vacunales ante emergencias como la causada por esta enfermedad, como para la actualización de vacunas ante las mutaciones de los patógenos.

Por esto, las herramientas *in silico* son importantes para el desarrollo de nuevos modelos de vacunas contra el SARS-CoV-2 que se mantengan vigentes ante las nuevas variantes del virus y las futuras variantes que se desarrollen. Además, investigaciones de este tipo con nuevos enfoques y metodologías aportan al conocimiento de este campo. Finalmente, los prototipos de vacunas ante esta enfermedad tan vigente podrían ser de interés para la industria farmacéutica, la cual tiene un gran crecimiento en el país.

4. HIPÓTESIS

Si las proteínas estructurales S, N, M y E del SARS-CoV-2 son inmunogénicas y las variantes emergentes del virus contienen mutaciones que afectan la capacidad de protección de las vacunas ante éste, entonces, diseños de vacunas multiepitópicas basadas en cada una de estas proteínas que tomen en cuenta las mutaciones de estas variantes podrán predecir una respuesta inmune tanto humoral como citotóxica con una amplia cobertura poblacional mundial que se mantenga vigente ante las variantes.

¿Es posible el diseño de diferentes vacunas multiepitópicas contra el SARS-CoV-2 basadas en cada una de sus cuatro proteínas estructurales S, N, M y E que predigan una respuesta inmune tanto humoral como citotóxica con una amplia cobertura poblacional mundial?

5. OBJETIVOS

5.1 Objetivo general

Diseñar mediante herramientas inmunoinformáticas vacunas mutiepitópicas contra SARS-CoV-2 basadas en las proteínas estructurales S, N, M y E que tengan una amplia cobertura poblacional, sean antigénicas y eviten mutaciones de las variantes del virus.

5.2 Objetivos particulares

1. Seleccionar los epítomos para MHCI, MHCII y células B para las proteínas estructurales S, M, N y E de SARS-CoV-2.
2. Refinar las regiones de epítomos seleccionados para evitar afectaciones que se generen por baja antigenicidad y por mutaciones de las variantes circulantes.
3. Diseñar y caracterizar *in silico* las vacunas multiepitópicas de amplia cobertura poblacional.

6. MATERIAL Y MÉTODOS

6.1 Selección de epítomos

El diseño de esta vacuna consiste en epítomos para células T y células B obtenidos de las proteínas estructurales S, N, M y E del SARS-CoV-2. Para la proteína S se obtuvo la secuencia referencia y se modificó con la mutación D614G, debido a que todas las variantes de interés y preocupación, que han sido designadas por la OMS, CDC y ECDC la contienen. Sumado a esto, se tomaron en cuenta todas las mutaciones contenidas en las variantes de interés y preocupación. Para las proteínas N, M y E se obtuvieron las secuencias de referencia y se utilizaron sin modificaciones. Secuencia de referencia SARS-CoV-2 (GenBank no. NC_045512.2) (NCBI, 2020).

6.1.1 Análisis de epítomos para células T CD8+

La selección de alelos para antígenos leucocitarios humanos (HLA) se realizó con base en lo planteado por Cun et al. (2021), tomando los alelos con frecuencia mayor a 1% en todas las poblaciones o presentes en al menos 1% de la población humana que se encuentren en la Base de Datos de Epítomos Inmunes (IEDB) (<http://tools.iedb.org/processing/help/>). Además, se evaluó el set de 27 alelos que dan una cobertura poblacional > 97% que propone la IEDB basándose en el trabajo de Weiskopf et al. (2013). De estos últimos, 26 alelos están incluidos en los 77 antes seleccionados, por lo que exclusivamente el alelo A*33:01 fue añadido al grupo final, siendo analizados un total de 78 alelos de los grupos A, B, C, E y G desglosados en la Tabla 1.

Tabla 1. Alelos utilizados en el análisis de epítomos para células T CD8+.

Grupo de Alelos	Alelos Analizados				
Alelos A	A*01:01	A*02:01	A*02:06	A*03:01	A*11:01
	A*23:01	A*24:02	A*25:01	A*26:01	A*29:02
	A*30:01	A*30:02	A*31:01	A*32:01	A*33:01
	A*33:03	A*68:01	A*68:02	A*74:01	
Alelos B	B*07:02	B*08:01	B*13:01	B*13:02	B*14:02
	B*15:01	B*15:02	B*15:25	B*18:01	B*27:02
	B*27:05	B*35:01	B*35:03	B*37:01	B*38:01
	B*39:01	B*40:01	B*40:02	B*44:02	B*44:03
	B*46:01	B*48:01	B*49:01	B*50:01	B*51:01
	B*52:01	B*53:01	B*55:01	B*56:01	B*57:01
	B*58:01	B*58:02			
Alelos C	C*01:02	C*02:02	C*02:09	C*03:02	C*03:03
	C*03:04	C*04:01	C*05:01	C*06:02	C*07:01
	C*07:02	C*07:04	C*08:01	C*08:02	C*12:02
	C*12:03	C*14:02	C*15:02	C*16:01	C*17:01
Alelos E	E*01:01	E*01:03			
Alelos G	G*01:01	G*01:02	G*01:03	G*01:04	G*01:06

Nota: Total = 78 alelos

Se seleccionaron epítomos para células T CD8+ de 8-14aa de longitud utilizando el programa NetMHCpan 4.1b en su servidor online (<https://services.healthtech.dtu.dk/service.php?NetMHCpan-4.1>) con los siguientes parámetros basados en el puntaje del rango calculado con las herramientas de espectrometría de masas-ligandos eluidos: afinidad fuerte (AF) < 0.5; afinidad débil (AD) > 0.5 y < 2. Se utilizaron las secuencias previamente obtenidas para el análisis y se capturaron los resultados en una hoja de cálculo. Con esta base de datos se generaron dos tablas dinámicas: una que ordena los epítomos por número de alelos cubiertos y otra que muestra la cantidad de alelos cubiertos por posición de inicio del epítomo en la secuencia. Con la primera se pudieron identificar rápidamente los epítomos con una mayor promiscuidad por los alelos analizados. Éstos se unieron en un solo constructo, juntando todos los alelos a los cuales tuvieron una afinidad fuerte y se calculó la cobertura hacia los alelos analizados y la longitud del posible péptido. La segunda tabla permite

la identificación de las que se llamarán de ahora en adelante “regiones de alta afinidad”, las cuales son regiones dentro de la secuencia analizada que contienen, en su mayoría, varios epítomos con afinidad fuerte a los alelos analizados. Las regiones de alta afinidad recibieron un tratamiento igual al anterior para posteriormente unir diferentes regiones en busca de un aumento de la cantidad de alelos que tengan epítomos de afinidad fuerte hacia éstos con la secuencia más corta posible. Tras determinar la combinación de regiones o epítomos que fueran afines a la mayor cantidad de alelos y tuvieran la menor longitud, se descartaron los epítomos que abarcaron alguno de los residuos con posibilidad de mutación basándose en la CDC, ECDC, OMS y el servidor CoVariants (<https://covariants.org/shared-mutations>) de las variantes enlistadas en la Tabla 2 (Nextstrain, 2023b). Epítomos con afinidad fuerte a los alelos sin cubrir fueron identificados en los datos y se seleccionaron tomando en cuenta si alguna parte de ellos estaba incluida dentro de las regiones de alta afinidad, considerando su longitud, la cantidad de alelos faltantes abarcados y el no presentar residuos con mutaciones. Finalmente, se realizó una reducción a la secuencia, identificando los alelos que sólo estuvieran cubiertos por un epítomo y considerando eliminar algunos de los epítomos que abarcaron alelos que tuvieran más de un epítomo con afinidad fuerte hacia éstos (Ahmed *et al.*, 2020).

Tabla 2. Variantes de SARS-CoV-2 consideradas en este estudio.

Variantes de Interés	Variantes de Preocupación	Variantes Ómicron
Épsilon (B.1.427/B.1.429)	Alfa (B.1.1.7)	BA.1
Eta (B.1.525)	Beta (B.1.351)	BA.2
Iota (B.1.526)	Gamma (P.1)	BA.4
Kappa (B.1.617.1)	Delta (B.1.617.2)	BA.5
Zeta (P.2)	Ómicron ¹	
Mu (B.1.621/B.1.621.1)		

Nota: ¹ Todas las variantes Ómicron han sido consideradas de preocupación.

Al terminar la selección de las regiones con base en el programa NetMHCpan 4.1b, se procedió a analizar las mismas proteínas con el servidor de predicción para MHC-I de la IEDB (v2.24, <http://tools.iedb.org/mhci/>), el cual funciona con el mismo motor que el primero, pero con diferencias en las bases de datos. Se corroboró que los alelos estuvieran cubiertos dentro de las mismas regiones y se tomó la decisión de añadir algunos aminoácidos a las regiones para aumentar el porcentaje de cobertura poblacional de la secuencia en los resultados de este programa. De igual manera, se realizó el procedimiento anterior para el programa MHCflurry utilizando Python ya que este no cuenta con servidor online.

6.1.2 Análisis de epítomos de células T CD4+

En la selección de epítomos para células T CD4+ se analizaron 27 alelos HLA que la IEDB reporta que cubren a más del 99% de la población con base en el trabajo de Greenbaum et al. (2011). Los alelos pertenecen a los grupos DP, DQ y DR y están desglosados en la tabla 3. El análisis se basó en la búsqueda de péptidos de longitud entre 12-18aa en los programas NetMHCIIpan 4.0 (<https://services.healthtech.dtu.dk/service.php?NetMHCIIpan-4.0>) y IEDB MHC-II Binding Prediction (<http://tools.iedb.org/mhcii/>) con el mismo mecanismo de ligandos eluidos que NetMHCIIpan 4.0. Se consideraron como epítomos de afinidad fuerte a los epítomos dentro del rango del 2% y de afinidad débil entre el 2-10% en los servidores NetMHCIIpan 4.0 e IEDB.

Tabla 3. Alelos utilizados en el análisis de epítomos para células T CD4+.

Grupo de Alelos	Alelos Analizados		
Alelos DP	A1*02:01/B1*01:01	A1*01:03/B1*02:01	A1*01:03/B1*04:01
	A1*03:01/B1*04:02	A1*02:01/B1*05:01	A1*02:01/14:01
Alelos DQ	A1*05:01/B1*02:01	A1*05:01/B1*03:01	A1*03:01/B1*03:02
	A1*04:01/B1*04:02	A1*01:01/B1*05:01	A1*01:02/B1*06:02
Alelos DR	B1*04:05	B1*07:01	B1*08:02
	B1*09:01	B1*11:01	B1*12:01
	B1*13:02	B1*15:01	B3*01:01
	B3*02:02	B4*01:01	B5*01:01

Nota: Total = 27 alelos

Para seleccionar los epítomos, se analizaron inicialmente los alelos contenidos dentro de las regiones de alta afinidad para células T CD8+ y se tomó la decisión de añadir algunos aminoácidos y regiones nuevas para aumentar la cobertura poblacional al máximo posible. Así como con los epítomos de MHCI, se tomaron en cuenta las mutaciones del virus al momento de la selección.

6.1.3 Análisis epítomos de células B

Se realizó la predicción de epítomos lineales para células B para las secuencias completas de las proteínas S, M y E, utilizando los servidores BepiPred 2.0 (<https://services.healthtech.dtu.dk/service.php?BepiPred-2.0>), IEDB Antibody Epitope Prediction (el cual utiliza el mecanismo de predicción de BepiPred 2.0, <http://tools.iedb.org/bcell/>), iBCE-EL (<http://thegleelab.org/iBCE-EL/iBCE.html>) y FBCPRED (<https://omictools.com/bcpreds-tool>). En todos los programas se buscaron epítomos de entre 12-25aa.

Para la proteína S, se buscaron epítomos en la sección de la proteína que se encuentra al exterior del virus, que va del aa 1 al 1,213, con especial atención en el dominio RBD, así como en las regiones HR1, CH y HR2. Además, se tomaron en cuenta los sitios de posibles glicosilaciones descritos por Shajahan et al. (2020), la accesibilidad para los anticuerpos (Grant *et al.*, 2020) y los residuos que presentan mutaciones en las variantes de interés y de preocupación que circulan en la actualidad hasta el linaje BA.5 de la variante Ómicron (Nextstrain, 2023b).

Por otro lado, para la obtención de los epítomos lineales de las proteínas M y E no se tomaron en cuenta sitios de glicosilación ni de mutaciones. Se analizaron las secciones abarcadas por los residuos 1-19 de la proteína M y por los aminoácidos 1-8 y 38-75 para la proteína E.

Se consideraron como posibles epítomos a los residuos que tuvieron un resultado mayor a 0.5 en los programas BepiPred 2.0 y la IEDB. Por su parte, en el programa iBCE-EL, se dividieron a los epítomos sugeridos por clases dependiendo del valor de probabilidad calculado por el programa y se consideraron epítomos con valores igual o mayores a 0.7. En el servidor FBCPRED, se dividieron los epítomos por clases de porcentaje de especificidad, considerando los epítomos de la clase del 75% y superiores. Con estos parámetros en cuenta, se procedieron a evaluar los

epítomos que cumplieran los criterios correspondientes en 2/4 servidores utilizados. Posteriormente, se tomaron en cuenta los resultados de posible exposición para cada residuo que provee el programa BepiPred 2.0, siendo el valor 0.3 el mínimo para considerarse como residuo expuesto. Los epítomos que contuvieran menos del 50% de sus residuos expuestos, se descartaron ya que, por tratarse de epítomos lineales para células B, lo más deseable es tener secuencias con la mayor cantidad posible de epítomos expuestos, para permitir la unión de anticuerpos neutralizantes. Cabe resaltar que es deseable que los epítomos se seleccionen parcial o completamente dentro de las regiones de alta afinidad de epítomos para células T CD8+, sin embargo, esto no limita la posibilidad de seleccionar epítomos de otros sitios.

Se seleccionaron las regiones que contengan epítomos lineales que cumplan con los criterios de tamaño, número de programas, residuos expuestos y que no se vean interrumpidas por algún sitio de posible glicosilación o de mutación registrada.

Finalmente, se analizaron todos los posibles epítomos en el programa VaxiJen 2.0 (<http://www.ddg-pharmfac.net/vaxijen/VaxiJen/VaxiJen.html>) y se consideraron inmunogénicos los que tuvieron un resultado mayor a 0.4, basado en lo que sugiere el programa para secuencias de virus. Los que no cumplieron el criterio fueron descartados o reportados con la especificación correspondiente en caso de verse contenidos dentro de las regiones de alta afinidad obtenidas previamente.

6.2 Refinamiento de las regiones

Dada la actualidad de la pandemia de COVID-19, el SARS-CoV-2 ha sufrido una gran cantidad de mutaciones durante el tiempo en que se desarrolló este trabajo. Debido a esto, algunas mutaciones como las presentes en los diferentes linajes de las variantes Delta y Ómicron se han tenido que tomar en cuenta después del proceso inicial de selección de los epítomos. Además, los valores de antigenicidad no fueron tomados en cuenta en la selección inicial de epítomos de MHC I y II, por lo que algunos de los epítomos seleccionados inicialmente no se predijeron antigénicos por el programa Vaxijen 2.0.

Por estas razones se llevó a cabo un proceso de refinamiento de las secuencias. El proceso consistió en primeramente identificar los epítomos que estuvieran afectados por mutaciones o

que tuvieran una antigenicidad menor a 0.4. Posterior a esto se evaluó la cobertura poblacional que resultaría de eliminar esos epítomos afectados. Después, se buscaron epítomos que cubrieron los alelos que se vieron afectados. Se añadieron los epítomos que aumentaron en mayor medida la cobertura poblacional buscando añadir la menor cantidad de residuos posibles y que cumplieran con los criterios previamente establecidos. Finalmente, se buscó optimizar la secuencia eliminando residuos o epítomos completos que ya no fueran requeridos para dar cobertura a algún alelo.

A las regiones de alta afinidad que resultaron de los procesos de selección y refinamiento se les nombró “regiones multiepitópicas”, término con el cual se les referirá de aquí en adelante.

Con las secuencias finales ya seleccionadas se reportaron las coberturas poblacionales finales por programa y en promedio para cada tipo de respuesta, tomando en cuenta todas las mutaciones reportadas en variantes de interés y preocupación hasta las variantes Ómicron enlistadas en la Tabla 2 y la resultante de sólo tomar en cuenta variantes Delta, Ómicron o ambas.

Finalmente, se realizó una evaluación de la cobertura poblacional tomando en cuenta las nuevas variantes Ómicron BA.2.12.1, BA.2.75, BQ.1 y XBB. Para esto se eliminaron los epítomos afectados por estas mutaciones y se calculó nuevamente la cobertura poblacional hacia los alelos de MHCI y II. También se evaluaron las afectaciones hacia los epítomos B lineales.

6.3 Diseños finales

6.3.1 Adición de enlazadores

Teniendo las regiones multiepitópicas seleccionadas, se procedió a analizar la adición de enlazadores. Debido al tamaño de las secuencias obtenidas, se optó por diseñar cuatro genes, uno para cada proteína, para su expresión y producción de manera independiente.

Primeramente, se obtuvieron todas las diferentes permutaciones posibles en el orden de las regiones de alta afinidad para cada construcción. Posteriormente, se analizó la adición de enlazadores en medio de las regiones. Los enlazadores probados fueron AAY, KK, GPGPG, EAAAK y GGGS, además de evaluar una opción sin añadir enlazadores y en algunos casos la

opción de unir dos regiones con los aminoácidos originales de la secuencia entre estas, debido a la cercanía entre las mismas. En un primer paso, se añadió cada enlazador en todos los espacios entre las regiones multiepitópicas y se evaluó su antigenicidad con Vaxijen 2.0. Cabe resaltar que debido a la cantidad de regiones obtenidas para la proteína S, el número de permutaciones era demasiado grande para la capacidad de procesamiento, por lo que se dividió esta proteína en dos grupos de regiones. Se identificaron las mejores dos o tres permutaciones para cada enlazador en los constructos de las proteínas N, M y E. Para la proteína S se realizó el mismo procedimiento para cada grupo, luego se procedió a combinar las opciones seleccionadas de los grupos y se añadieron los enlazadores a probar entre los dos grupos, para nuevamente evaluar la antigenicidad. Las construcciones con la antigenicidad más alta se eligieron para probar la adición del adyuvante.

6.3.2 Adición de elementos finales del diseño de la vacuna

A las secuencias seleccionadas se les añadió el adyuvante PADRE (AKFVAAWTLKAA) de dos maneras: 1) al inicio y 2) al final de la secuencia. Entre el adyuvante y la secuencia se adicionó alguno de los enlazadores probados previamente. Se evaluó la antigenicidad y se seleccionaron los diseños de cada proteína con la antigenicidad más alta. Posteriormente, se añadieron el péptido señal (MKMKVLFVATILVAWQCHA) antes de la secuencia y la señal de retención al retículo endoplásmico (SEKDEL) al final de ésta. Se añadió una etiqueta de 6 histidinas (HHHHHH) de dos maneras: 1) entre el péptido señal y la secuencia, y 2) entre la secuencia y la señal de retención. Finalmente, se evaluó la antigenicidad de las posibles secuencias y se seleccionó la que tuviera el valor mayor en cada uno de los cuatro diseños.

Se realizó un alineamiento múltiple de secuencias de las construcciones con el genoma de *Homo sapiens*, *Agrobacterium tumefaciens*, *Escherichia coli*, *Schizochytrium* sp., *Nicotiana tabacum*, *Medicago sativa* y *Daucus carota* con la herramienta del NCBI (https://blast.ncbi.nlm.nih.gov/Blast.cgi?PROGRAM=blastp&PAGE_TYPE=BlastSearch&LINK_LOC=blasthome). Esto para corroborar que no haya proteínas similares en el ser humano por seguridad y en los demás organismos porque son o podrían ser parte del proceso de producción de los antígenos.

Se les añadió a las construcciones finales el codón de paro (tga) en el extremo 3' de la secuencia. También, se añadieron los sitios de corte para enzimas de restricción *SmaI* en el extremo 5' y para *BamHI* y *SacI* en el extremo 3' después del codón de paro. Esta construcción fue enviada a la empresa Gene Universal Inc. (www.geneuniversal.com), donde la secuencia de nucleótidos generada fue optimizada para su expresión en Tabaco (*Nicotiana tabacum*) y en microalga (*Schizochytrium* sp.). Tras la optimización, se corroboró que se mantuviera correctamente la traducción de la proteína y que los sitios de corte indicados estuvieran en el sitio y no se repitieran al interior de la secuencia. Finalmente, se solicitó la síntesis de los genes con esta misma empresa.

6.3.3 Diseño de primers específicos

Las secuencias optimizadas por el fabricante fueron insertadas en el servidor Primer3web v4.1.0 (<https://primer3.ut.ee/>), excluyendo los sitios de corte y el codón de paro. Se buscaron primers específicos de entre 18 y 22 bases de longitud, con un porcentaje de GC entre 40-60%, que terminaran en guanina (G) o citocina (C), con una Tm de entre 58 y 62°C y una diferencia de 1°C o menos. Se buscó que de preferencia los oligos tuvieran un valor de temperatura para unión a sitios no específicos de la secuencia tanto en 5' como en 3' de 0 °C, lo que indicaría que no hay sitios de este tipo. Para los pares de primers para PCR punto final, se buscó que amplificaran el mayor porcentaje posible de la secuencia. En el caso de los oligos para PCR tiempo real (qPCR) se buscó que los pares amplificaran entre 100-200 bases, de preferencia en el centro de la secuencia.

Los primers que cumplieron los requerimientos se evaluaron en el servidor RNAfold (<http://rna.tbi.univie.ac.at/cgi-bin/RNAWebSuite/RNAfold.cgi>) para predecir su estructura secundaria con el mínimo de energía libre y la estructura centroide. Si ambas estructuras eran circulares, se tomaron en cuenta y se registró el porcentaje de frecuencia de dicha estructura. De cumplir dicha condición, se evaluaron en el servidor Oligo Evaluator de Sigma-Aldrich (<http://www.oligoevaluator.com/LoginServlet>). Si la predicción de estructura secundaria y dímeros de los oligos era negativa, se pasaron a la última fase del proceso. Los primers junto con la secuencia se analizaron en el servidor Primer-BLAST

(<https://www.ncbi.nlm.nih.gov/tools/primer-blast/>) para buscar inespecificidades en la base de datos del GenBank. Se buscó que fueran específicos para la secuencia, aunque si no lo eran, se evaluó a que organismo eran específicos y se ponderó el seleccionarlos si no afectaba en ninguna parte del proceso de producción ni prueba de la vacuna.

6.4 Caracterización *in silico* de las proteínas quiméricas

6.4.1 Evaluación de seguridad y caracterización de propiedades fisicoquímicas

Para corroborar la seguridad de la administración de los diseños se analizó la alergenicidad y toxicidad. La alergenicidad se evaluó con los programas AllerTOP v2.0 (<https://www.ddg-pharmfac.net/AllerTOP/>) y Allergen FP v1.0 (<https://ddg-pharmfac.net/AllergenFP/>). La toxicidad se analizó para cada epítipo seleccionado con el servidor ToxinPred (<https://webs.iitd.edu.in/raghava/toxinpred/index.html>) con el método SVM (TrEMBL) + Motif Based.

El servidor ProtParam (<https://web.expasy.org/protparam/>) se utilizó para el cálculo del tamaño, peso molecular, punto isoeléctrico, tiempo de vida media, índice de estabilidad, índice alifático y promedio general de hidropatía (GRAVY).

6.4.2 Modelaje 3D

Se realizó una predicción *ab initio* de la estructura tridimensional de los diseños con el programa I-TASSER (<https://zhanggroup.org/I-TASSER/>). Este programa arroja los mejores cinco resultados. Los cinco modelos se sometieron a refinamiento en los servidores GalaxyRefine (<https://galaxy.seoklab.org/cgi-bin/submit.cgi?type=REFINE>) para el diseño de la proteína S, obteniendo cinco resultados y sus respectivos valores de GDT-HA, MolProbity, Clash Score, rotámeros deficientes, porcentaje de residuos favorables y raíz de la desviación cuadrática media (RMSD). Para el resto de los diseños se utilizó el servidor GalaxyRefine2 (<https://galaxy.seoklab.org/cgi-bin/submit.cgi?type=REFINE2>), obteniendo 10 resultados con los mismos parámetros excepto GDT-HA y con la adición del Galaxy Score.

Todos los modelos refinados se sometieron a validación en el servidor SAVES v.6 (<https://saves.mbi.ucla.edu/>). En este, se utilizaron los programas ERRAT para evaluar el factor

general de calidad y PROCHECK para evaluar la estructura con el gráfico de Ramachandran. Fue seleccionado el modelo refinado de cada diseño que tuviera el mayor factor general de calidad y porcentaje de residuos en regiones favorables y menor cantidad de residuos en regiones no favorables.

6.4.3 Selección de péptidos para producción de sueros hiperinmunes

Para poder en un futuro producir sueros hiperinmunes para la cuantificación de los antígenos proteicos se realizó la selección de péptidos incluidos en la proteína para una futura inmunización y obtención de anticuerpos policlonales. Los epítomos previamente seleccionados con mayores valores de antigenicidad de entre 14-18aa fueron evaluados. Se analizaron con la herramienta “Peptide Analyzing Tool to Assist Peptide Design” de GeneScript (<https://www.genscript.com/tools/peptide-analyzing-tool>) y se buscó que los péptidos tuvieran buena solubilidad en agua. Los péptidos con los mayores valores de antigenicidad que tuvieran buena solubilidad en agua fueron seleccionados.

7. RESULTADOS

7.1 Selección de epítomos y refinamiento

7.1.1 Análisis proteína E

Se obtuvo una secuencia final de 58aa para la proteína E que representa el 77.33% de la secuencia de la proteína contenido en la selección multiepitópica (Fig. 1).

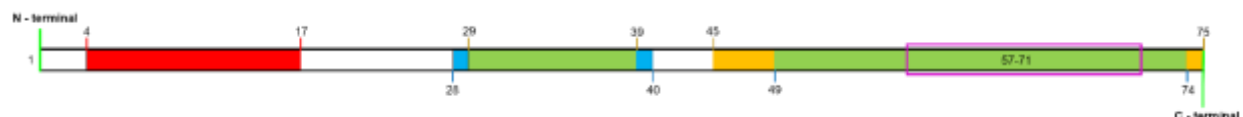


Figura 1. Regiones de la proteína E seleccionadas para la construcción de la vacuna. Zonas amarillas indican donde hay exclusivamente epítomos T CD8+. Zonas azules indican donde hay exclusivamente epítomos T CD4+. Zonas verdes indican regiones en donde hay epítomos T CD8+ y T CD4+. Zona roja indica región que se añadió a la construcción sin cumplir todos los criterios de selección. Cuadrado violeta indica zona de epítomos lineales B. Los números unidos a las regiones marcan las posiciones que son extremos de estas. Números dentro de los cuadros violetas indican las posiciones de las regiones de epítomos B. Número al extremo izquierdo indica la posición en la que inicia ese segmento.

La selección final de los epítomos contiene 3 regiones multiepitópicas en las cuales se encuentran 24 epítomos para células T CD8+, 10 epítomos para células B y 61 epítomos para células T CD4+, que cumplen los requerimientos de selección, distribuidos como se muestra en la Tabla 4. Cabe resaltar que dos epítomos más para células B no cumplieron únicamente el requisito de residuos expuestos, pero se encuentran dentro de la secuencia.

Tabla 4. Cantidad de epítomos obtenidos dentro de las regiones seleccionadas de la proteína E.

Región	Posición	Epítomos T CD8+	Epítomos B	Epítomos T CD4+
FVSEETGLIVNSV	4-17	0	0	0
LVTLAILTALRLC	28-40	4	0	2
NIVNVSLVKPSFYVYSRVKLNLSRVPDLLV	45-75	20	10	59
Epítomos totales:		24	10	61

Nota: Posición indica residuos de inicio y final en la secuencia de referencia de la proteína para la región correspondiente.

De las regiones expuestas de la proteína, se obtuvieron 12 epítomos para células B sugeridos por al menos dos de los cuatro programas utilizados para la predicción. De estos, 10 fueron sugeridos por tres de los programas, 9 de estos epítomos fueron sugeridos por el programa FBCPRED con una especificidad del 90% y uno con el 75%. Dos epítomos fueron sugeridos por dos programas, uno con una puntuación mayor a 0.8 en el programa iBCE-EL y sugerido con una especificidad del 90% en el programa FBCPRED, mientras que el otro tuvo una puntuación mayor a 0.75 y era sugerido con una especificidad del 85% (Tabla 5).

Tabla 5. Epítomos para células B sugeridos por al menos 2 programas en la proteína E.

Epítomos	Bepipred 2.0	IEDB	iBCE-EL	FBCPRED (%)
YVYSRVKLNLS	✓	✓	-	75
YVYSRVKLNSSR	✓	✓	-	90
YVYSRVKLNSSRV	✓	✓	-	90
YVYSRVKLNSSRVP	✓	✓	-	90
VYSRVKLNSSR	✓	✓	-	90
VYSRVKLNSSRV	✓	✓	-	90
VYSRVKLNSSRVP	✓	✓	-	90
YSRVKLNSSRV	✓	✓	-	90
YSRVKLNSSRVP	✓	✓	-	90
SRVKLNSSRVP	✓	✓	-	90
NIVNVSLVKPSFYVYSR	-	-	>0.8	90
IVNVSLVKPSFYVYSRV	-	-	>0.75	85

Los 12 epítomos sugeridos por al menos dos programas cumplieron el criterio de antigenicidad, pero dos de ellos no cumplieron el criterio de porcentaje de residuos escondidos, por lo que fueron descartados. De los 10 epítomos restantes, 4 están afectados por la mutación P71L, presente en la variante Beta del SARS-CoV-2 (Tabla 6). Estos epítomos afectados no fueron eliminados ya que forman parte de las regiones de alta afinidad seleccionadas para epítomos T CD8+ y T CD4+.

Tabla 6. Valoración de epítomos B lineales sugeridos por al menos 2 programas por valor de antigenicidad y porcentaje de residuos expuestos en la proteína E.

Epítomos	Longitud (aa)	Posición	Valor Antigenicidad	Residuos Escondidos (%)
YVYSRVKLNLS	12	57-68	0.6906	41.67
YVYSRVKLNLSR	13	57-69	0.5246	38.46
YVYSRVKLNLSRV	14	57-70	0.5405	35.71
YVYSRVKLNLSRVP ¹	15	57-71	0.4492	33.33
VYSRVKLNLSR	12	58-69	0.5867	33.33
VYSRVKLNLSRV	13	58-70	0.6008	30.77
VYSRVKLNLSRVP ¹	14	58-71	0.4983	28.57
YSRVKLNLSRV	12	59-70	0.5788	25.00
YSRVKLNLSRVP ¹	13	59-71	0.4745	23.08
SRVKLNLSRVP ¹	12	60-71	0.5572	16.67
NIVNVSLVKPSFYVYSR ²	17	45-61	0.8337	70.59
IVNVSLVKPSFYVYSRV ²	17	46-62	0.6711	76.47

Notas: ¹ Epítomos afectados por mutaciones de variantes que no son Ómicron ni Delta; ² Epítomos que no fueron seleccionados por su alto porcentaje de residuos escondidos; Posición indica residuos de inicio y final en la secuencia de referencia de la proteína para la región correspondiente.

Se obtuvieron coberturas poblacionales por alta afinidad para los alelos de MHCI seleccionados del 74.4% para el servidor NetMHCpan 4.1b, 82.1% para el servidor de la IEDB y 69.2% para el programa MHCflurry. Tanto mutaciones como baja antigenicidad disminuyen esta cobertura. Debido a la falta de opciones de epítomos de alta afinidad sugeridos por los programas, no se realizó ningún ajuste y estas coberturas poblacionales no pudieron mejorarse. Bajo los criterios propuestos en este trabajo, las coberturas finales son de 66.7%, 71.8% y 60.3% para cada uno de los programas (Tabla 7).

Tabla 7. Porcentajes de coberturas poblacionales de alelos MHCI de la secuencia seleccionada para la proteína E.

Cobertura	T CD8+		
	NetMHCpan 4.1b	IEDB	MHCflurry
Alta afinidad	74.36	82.05	69.23
Sin epítomos de baja antigenicidad	66.67	71.79	60.26
Sin epítomos afectados por mutaciones	66.67	71.79	58.97
Sin epítomos afectados por mutaciones Ómicron	66.67	71.79	60.26
Sin epítomos de baja antigenicidad y afectados por mutaciones	66.67	71.79	58.97
Sin epítomos de baja antigenicidad y afectados por mutaciones Ómicron	66.67 ¹	71.79 ¹	60.26 ¹

Nota: ¹ Coberturas poblacionales finales bajo los criterios de selección.

Los epítomos para células T CD4+ generaron una cobertura poblacional por alta afinidad del 63% para el servidor NetMHCIIpan 4.0 y del 66.7% para el servidor de la IEDB. Las afectaciones por baja antigenicidad provocan la misma disminución de cobertura poblacional que las afectaciones por la mutación T9I de la variante Ómicron. Las afectaciones por la mutación del aminoácido P71L, presente en la variante Beta, generan una mayor afectación al porcentaje de cobertura poblacional. La adición de los aminoácidos 28L, 29V y 40C mejora la cobertura poblacional en el servidor de la IEDB a un 63%, al incluir epítomos de alta afinidad hacia los alelos DRB1*01:01 y DRB1*12:01, los cuales son antigénicos y no son afectados por mutaciones. (Tabla 8).

Tabla 8. Porcentajes de coberturas poblacionales de alelos MHCII iniciales y mejoradas con el refinamiento de la secuencia seleccionada para la proteína E.

Cobertura	NetMHCIIpan 4.0		IEDB	
	Inicial	Mejorada	Inicial	Mejorada
Alta afinidad	62.96		66.67	
Sin epítomos de baja antigenicidad	59.26	59.26	59.26	62.96 ¹
Sin epítomos afectados por mutaciones	48.15	48.15	51.85	59.26 ¹
Sin epítomos afectados por mutaciones Ómicron	59.26	59.26	59.26	62.96 ¹
Sin epítomos de baja antigenicidad y afectados por mutaciones	48.15	48.15	51.85	59.26 ¹
Sin epítomos de baja antigenicidad y afectados por mutaciones Ómicron	59.26	59.26	59.26	62.96 ¹²

Notas: ¹ Coberturas poblacionales mejoradas tras el refinamiento; ² Coberturas poblacionales finales bajo los criterios de selección.

7.1.2 Análisis proteína M

Para la proteína M se construyó una secuencia de 167aa, representando el 75.22% de la secuencia de la proteína (Fig. 2).

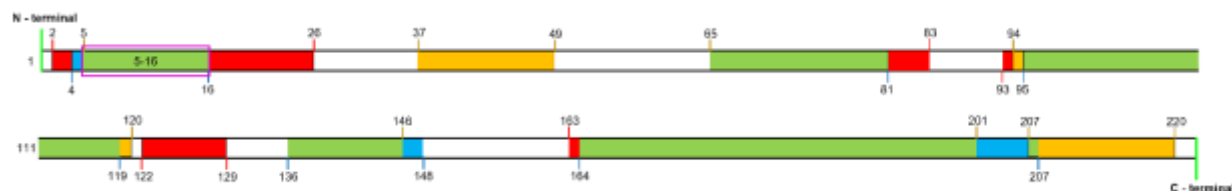


Figura 2. Regiones de la proteína M seleccionadas para la construcción de la vacuna. Zonas amarillas indican donde hay exclusivamente epítomos T CD8+. Zonas azules indican donde hay exclusivamente epítomos T CD4+. Zonas verdes indican regiones en donde hay epítomos T CD8+ y T CD4+. Zonas rojas indican regiones que se añadieron a la construcción sin cumplir todos los criterios de selección. Cuadrado violeta indica zona de epítomos lineales B. Los números unidos a las regiones marcan las posiciones que son extremos de estas. Números dentro de los cuadros violetas indican las posiciones de las regiones de epítomos B. Números al extremo izquierdo indican la posición en la que inicia ese segmento.

La secuencia construida consta de 7 regiones multiepitópicas, que contienen 75 epítomos para células T CD8+, 1 epítomo para células B y 93 epítomos para células T CD4+ que cumplen los criterios de selección (Tabla 9).

Tabla 9. Cantidad de epítomos obtenidos dentro de las regiones seleccionadas de la proteína M.

Región	Posición	Epítomos T CD8+	Epítomos B	Epítomos T CD4+
ADSNGTITVEELKKLLEQWNLVIGF	2-26	5	1	2
FAYANRRNFLYII	37-49	6	0	0
FVLAAYRINWITGGIAIAMA	65-85	6	0	11
LSYFIASFRLFARTRSMWSFNPETNILL	93-120	18	0	16
VPLHGTIL	122-129	0	0	0
SELVIGAVILRGH	136-148	6	0	3
DLPKEITVATSRTLSTYKLGASQRVAGDS GFAAYSRYRIGNYKLNTDHSSSDNIALL	163-220	34	0	61
Epítomos totales:		75	1	93

Nota: Posición indica residuos de inicio y final en la secuencia de referencia de la proteína para la región correspondiente.

Se obtuvieron 10 epítomos para células B sugeridos por al menos dos de los cuatro programas utilizados para la predicción, todos por los servidores Bepipred 2.0 y la IEDB (Tabla 10).

Tabla 10. Epítomos para células B sugeridos por al menos 2 programas en la proteína M.

Epítomos	Bepipred 2.0	IEDB	iBCE-EL	FBCPRED
NGTITVEELKKL	✓	✓	-	-
NGTITVEELKKLL	✓	✓	-	-
NGTITVEELKKLLE	✓	✓	-	-
NGTITVEELKKLLEQ	✓	✓	-	-
GTITVEELKKLL	✓	✓	-	-
GTITVEELKKLLE	✓	✓	-	-
GTITVEELKKLLEQ	✓	✓	-	-
TITVEELKKLLE	✓	✓	-	-
TITVEELKKLLEQ	✓	✓	-	-
ITVEELKKLLEQ	✓	✓	-	-

De los 10 sugeridos por al menos dos programas, todos cumplieron el criterio de porcentaje de residuos escondidos, pero sólo uno cumplió el criterio de antigenicidad. De los 10 epítomos analizados, 3 están afectados por la mutación Q19E, presente en diferentes variantes de Ómicron, pero ninguno de estos es el epítomo seleccionado por cumplir todos los criterios. Sin embargo, los 9 epítomos para células B descartados están incluidos en la secuencia por estar contenidos en zonas con epítomos T CD8+ y T CD4+ (Tabla 11).

Tabla 11. Valoración de epítomos B lineales sugeridos por al menos 2 programas por valor de antigenicidad y porcentaje de residuos expuestos en la proteína M.

Epítomos	Longitud (aa)	Posición	Valor Antigenicidad	Residuos Escondidos (%)
NGTITVEELKKL	12	5-16	0.5412	41.67%
NGTITVEELKKLL ¹	13	5-17	0.1884	46.15%
NGTITVEELKKLLE ¹	14	5-18	0.1395	42.86%
NGTITVEELKKLLEQ ^{1 2}	15	5-19	0.0039	40.00%
GTITVEELKKLL ¹	12	6-17	0.2136	50.00%
GTITVEELKKLLE ¹	13	6-18	0.1572	46.15%
GTITVEELKKLLEQ ^{1 2}	14	6-19	0.0073	42.86%
TITVEELKKLLE ¹	12	7-18	-0.0122	50.00%
TITVEELKKLLEQ ^{1 2}	13	7-19	-0.1613	46.15%
ITVEELKKLLEQ ¹	12	8-19	-0.2163	50.00%

Notas: ¹ Epítomos que no fueron seleccionados por su bajo valor de antigenicidad; ² Epítomos afectados por mutaciones de variante Ómicron; Posición indica residuos de inicio y final en la secuencia de referencia de la proteína para la región correspondiente.

La secuencia de la proteína M, obtuvo coberturas poblacionales por alta afinidad del 98.7% para el servidor NetMHCpan 4.1b, 100% para el servidor de la IEDB y 94.9% para el programa MHCflurry. Tanto mutaciones como baja antigenicidad disminuyen esta cobertura poblacional, sin embargo, si se toman en cuenta exclusivamente las mutaciones presentes en la variante Ómicron no hay afectaciones a la cobertura poblacional en el programa MHCflurry. Se añadieron los aminoácidos L46, Y47, I48, I49, F65, V66, W110, S111 y F112 para mejorar las afectaciones a los porcentajes de cobertura poblacional, obteniendo mejoras en todos los casos para todos los programas. No se logró una mejora total de los porcentajes de cobertura poblacional, a excepción de los casos de afectaciones por mutaciones incluyendo las variantes Delta y Ómicron en el programa MHCflurry, donde sí se mejoró la cobertura en su totalidad. Las coberturas finales con los criterios de selección fueron de 83.3% para NetMHCpan 4.1b, 94.9% para IEDB y 87.2 para MHCflurry (Tabla 12).

Tabla 12. Porcentajes de coberturas poblacionales de alelos MHCI iniciales y mejoradas con el refinamiento de la secuencia seleccionada para la proteína M.

T CD8+						
Cobertura	NetMHCpan 4.1b		IEDB		MHCflurry	
Alta afinidad	98.72		100		94.87	
	Inicial	Mejorada	Inicial	Mejorada	Inicial	Mejorada
Sin epítomos de baja antigenicidad	76.92	85.9 ¹	92.31	97.44 ¹	80.77	87.18 ¹
Sin epítomos afectados por mutaciones	92.31	94.87 ¹	97.44	98.72 ¹	93.59	94.87 ²
Sin epítomos afectados por mutaciones Delta + Ómicron	92.31	94.87 ¹	97.44	98.72 ¹	93.59	94.87 ²
Sin epítomos afectados por mutaciones Ómicron	92.31	94.87 ¹	97.44	98.72 ¹	94.87 ³	94.87
Sin epítomos de baja antigenicidad y afectados por mutaciones	74.36	83.33 ¹	89.74	94.87 ¹	75.64	87.18 ¹
Sin epítomos de baja antigenicidad y afectados por mutaciones Delta + Ómicron	74.36	83.33 ^{1 4}	89.74	94.87 ^{1 4}	75.64	87.18 ^{1 4}
Sin epítomos de baja antigenicidad y afectados por mutaciones Ómicron	74.36	83.33 ¹	89.74	94.87 ¹	80.77	87.18 ¹

Notas: ¹ Coberturas poblacionales mejoradas tras el refinamiento; ² Coberturas poblacionales mejoradas completamente tras el refinamiento; ³ Condiciones que no afectaron la cobertura poblacional original del programa; ⁴ Coberturas poblacionales finales bajo los criterios de selección.

Se obtuvieron coberturas poblacionales de los alelos de MHCII seleccionados por alta afinidad del 85.2% para el servidor NetMHCIIpan 4.0 y 92.6% para el servidor de la IEDB. Se presentaron afectaciones en la cobertura poblacional debido a la baja antigenicidad y las mutaciones, aunque las mutaciones exclusivamente de la variante Ómicron afectan en menor medida que si se toman en cuenta también las de otras variantes, incluyendo la variante Delta. Bajo los criterios de selección, las coberturas poblacionales finales fueron de 55.6% para NetMHCIIpan 4.0 y 66.7% para IEDB (Tabla 13).

Tabla 13. Porcentajes de coberturas poblacionales de alelos MHCII de la secuencia seleccionada para la proteína M.

Cobertura	NetMHCIIpan 4.0	IEDB
Alta afinidad	85.19	92.59
Sin epítomos de baja antigenicidad	66.67	77.78
Sin epítomos afectados por mutaciones	74.07	77.78
Sin epítomos afectados por mutaciones Delta + Ómicron	74.07	77.78
Sin epítomos afectados por mutaciones Ómicron	77.78	81.48
Sin epítomos de baja antigenicidad y afectados por mutaciones	55.56	66.67
Sin epítomos de baja antigenicidad y afectados por mutaciones Delta + Ómicron	55.56 ¹	66.67 ¹
Sin epítomos de baja antigenicidad y afectados por mutaciones Ómicron	59.26	70.37

Nota: ¹ Coberturas poblacionales finales bajo los criterios de selección.

7.1.3 Análisis proteína N

La secuencia seleccionada para la proteína N tiene una longitud de 200aa, lo cual representa el 47.73% del tamaño total de la proteína (Fig. 3).

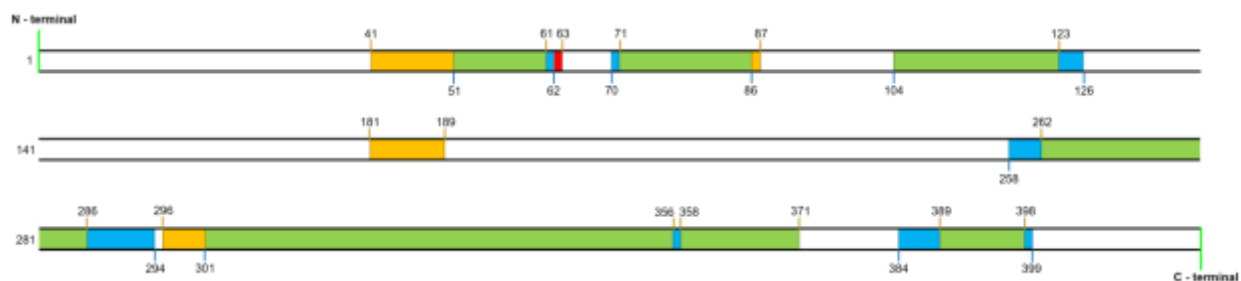


Figura 3. Regiones de la proteína N seleccionadas para la construcción de la vacuna. Zonas amarillas indican donde hay exclusivamente epítomos T CD8+. Zonas azules indican donde hay exclusivamente epítomos T CD4+. Zonas verdes indican regiones en donde hay epítomos T CD8+ y T CD4+. Zona roja indica región que se añadió a la construcción sin cumplir todos los criterios de selección. Los números unidos a las regiones marcan las posiciones que son extremos de estas. Números al extremo izquierdo indican la posición en la que inicia ese segmento.

Se seleccionaron 7 regiones multiepitópicas, conteniendo 89 epítomos para células T CD8+ y 229 epítomos para células T CD4+ que cumplen los criterios de selección. No se seleccionaron epítomos para células B ya que la proteína se encuentra en su totalidad al interior del virus (Tabla 14).

Tabla 14. Cantidad de epítomos obtenidos dentro de las regiones seleccionadas de la proteína N.

Región	Posición	Epítomos T CD8+	Epítomos T CD4+
RPQGLPNNTASWFTALTQHGKED	41-63	5	1
QGVPIINTNSSPDDQIGYY	70-87	11	6
LSPRWYFYLLGTGPEAGLPYGAN	104-126	5	30
QASSRSSSR	181-189	1	0
PRQKRTATKAYNVTQAFGRRGPEQTQGNFGDQELIRQ	258-294	10	45
TDYKHWPQIAQFAPSASAFFGMSRIGMEVTPSGTWLTYTGAIKL	296-371	46	137
DDKDPNFKDQVILLNKHIDAYKTFPPTPKKD			
QRQKKQQTVTLPAAD	384-399	11	10
Epítomos totales:		89	229

Nota: Posición indica residuos de inicio y final en la secuencia de referencia de la proteína para la región correspondiente.

Las coberturas poblacionales obtenidas por alta afinidad para los alelos de MHC I seleccionados son del 97.4% para el servidor NetMHCpan 4.1b, 100% para el servidor de la IEDB y 94.9% para el programa MHCflurry. Las mutaciones por sí mismas no generaron ninguna afectación a la cobertura poblacional en ninguno de los programas, caso contrario a la antigenicidad que afectó en los tres casos. Es importante remarcar que ninguna mutación de las variantes Ómicron utilizadas como criterio de selección (BA.1, BA.2, BA.4 y BA.5) afecta las regiones seleccionadas. En todos los casos, se pudo mejorar totalmente la cobertura poblacional en el servidor NetMHCpan 4.1b y el de la IEDB. Se logró cubrir el alelo A*01:01 en el programa MHCflurry, el cual estaba afectado por baja antigenicidad, pero no se logró mejorar en mayor medida la cobertura poblacional. Los porcentajes de cobertura poblacional finales bajo los criterios de selección fueron de 97.4%, 100% y 89.7% (Tabla 15).

Tabla 15. Porcentajes de coberturas poblacionales de alelos MHCI iniciales y mejoradas con el refinamiento de la secuencia seleccionada para la proteína N.

T CD8+						
Cobertura	NetMHCpan 4.1b		IEDB		MHCflurry	
Alta afinidad	97.44		100		94.87	
	Inicial	Mejorada	Inicial	Mejorada	Inicial	Mejorada
Sin epítomos de baja antigenicidad	93.59	97.44 ¹	97.44	100 ¹	88.46	89.74 ²
Sin epítomos afectados por mutaciones	97.44 ³	97.44	100 ³	100	94.87 ³	94.87
Sin epítomos afectados por mutaciones Delta	97.44 ³	97.44	100 ³	100	94.87 ³	94.87
Sin epítomos de baja antigenicidad y afectados por mutaciones	93.59	97.44 ¹	97.44	100 ¹	88.46	88.46
Sin epítomos de baja antigenicidad y afectados por mutaciones Delta	93.59	97.44 ^{1 4}	97.44	100 ^{1 4}	88.46	89.74 ^{2 4}

Notas: ¹ Coberturas poblacionales mejoradas completamente tras el refinamiento; ² Coberturas poblacionales mejoradas tras el refinamiento; ³ Condiciones que no afectaron la cobertura poblacional original del programa; ⁴ Coberturas poblacionales finales bajo los criterios de selección.

Se eliminó la región “LQLPQGTTL” en el proceso de refinamiento ya que las adiciones de regiones de epítomos para células T CD4+ compensaron con epítomos de afinidad fuerte para los alelos de MHCI que esta región cubría. Además, los epítomos para células T CD8+ dentro de esta región no son antigénicos, por lo que no aportarían cobertura en la selección final.

En la Tabla 16 se puede observar que la cobertura poblacional obtenida inicialmente en los servidores NetMHCIIpan 4.0 y la IEDB es del 92.6%. Los epítomos de baja antigenicidad y las mutaciones de las variantes de interés y de preocupación por sí mismas no generaron afectaciones en la cobertura poblacional. Al tomar en cuenta tanto la baja antigenicidad como las mutaciones sí se generaron afectaciones, siendo la mayor disminución en la cobertura poblacional la de baja antigenicidad combinada con las mutaciones de todas las variantes contempladas. Se logró mejorar la cobertura poblacional en los casos que involucran baja antigenicidad y las variantes, pero en ningún caso se mejoró en su totalidad. La variante Ómicron no generó disminución de la cobertura poblacional en ningún caso. Considerando todos los criterios de selección planteados (alta afinidad + antigenicidad + ausencia de

mutaciones Delta y Ómicron), los porcentajes de cobertura poblacional final fueron de 88.9% para ambos programas.

Tabla 16. Porcentajes de coberturas poblacionales de alelos MHCII iniciales y mejoradas con el refinamiento de la secuencia seleccionada para la proteína N.

Cobertura	NetMHCIIpan 4.0		IEDB	
Alta Afinidad	92.59		92.59	
	Inicial	Mejorada	Inicial	Mejorada
Sin epítomos de baja antigenicidad	92.59 ¹	92.59	92.59 ¹	92.59
Sin epítomos afectados por mutaciones	92.59 ¹	92.59	92.59 ¹	92.59
Sin epítomos afectados por mutaciones Delta	92.59 ¹	92.59	92.59 ¹	92.59
Sin epítomos de baja antigenicidad y afectados por mutaciones	77.78	85.19 ²	81.48	88.89 ²
Sin epítomos de baja antigenicidad y afectados por mutaciones Delta	81.48	88.89 ^{2 3}	81.48	88.89 ^{2 3}

Notas: ¹ Condiciones que no afectaron la cobertura poblacional original del programa; ² Coberturas poblacionales mejoradas tras el refinamiento; ³ Coberturas poblacionales finales bajo los criterios de selección.

7.1.4 Análisis proteína S

Para la proteína S se seleccionaron 261aa para componer la secuencia multiepitópica, que representa el 20.5% del tamaño total de la proteína (Fig. 4).

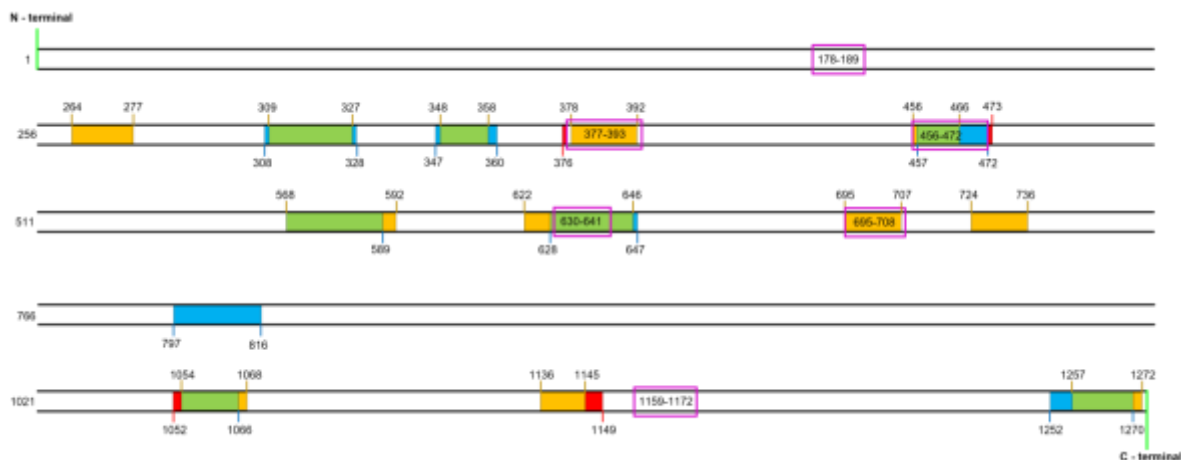


Figura 4. Regiones de la proteína S seleccionadas para la construcción de la vacuna. Zonas amarillas indican donde hay exclusivamente epítomos T CD8+. Zonas azules indican donde hay exclusivamente epítomos T CD4+. Zonas verdes indican regiones en donde hay epítomos T CD8+ y T CD4+. Zonas rojas indican regiones que se añadieron a la construcción sin cumplir todos los criterios de selección. Cuadrado violeta indica zona de epítomos lineales B. Los números unidos a las regiones marcan las posiciones que son extremos de estas. Números dentro de los cuadros violetas indican las posiciones de las regiones de epítomos B. Números al extremo izquierdo indican la posición en la que inicia ese segmento.

Para la proteína S se seleccionaron 15 regiones multiepitópicas conteniendo: 88 epítomos para células T CD8+, 43 epítomos para células B y 106 epítomos para células T CD4+ que cumplen los criterios de selección (Tabla 17).

Tabla 17. Cantidad de epítomos obtenidos dentro de las regiones seleccionadas de la proteína S.

Región	Posición	Epítomos T CD8+	Epítomos B	Epítomos T CD4+
DLEGKQGNFKNL	178-189	0	1	0
AYYVGYLQPRTFLL	264-277	14	0	0
VEKGIYQTSNFRVQPTESIVR	308-328	4	0	40
FASVYAWNRRKRISN	347-360	4	0	5
TFKCYGVSPTKLNDLCFT	376-393	7	16	0
FRKSNLKPFERDISTEIIY	456-473	3	13	8
DIADTTDAVRDPQTLIEDITPCSF	568-592	13	0	4
VAIHADQLTPTWRVYSTGSNVFQTRA	622-647	11	1	17
YTMSLGAENSVAYS	695-708	5	6	0
TEILPVSMTKTSV	724-736	7	0	0
FGGFNFSQILPDPSKPSKRS	797-816	0	0	17
FPQSAPHGVVFLHVTYV	1052-1068	7	0	1
TVYDPLQPELDSFK	1136-1149	4	0	0
HTSPDVDLGDISGI	1159-1172	0	6	0
SCCKFDEDDSEPVKGVKLHY	1252-1272	9	0	14
Epítomos totales:		88	43	106

Nota: Posición indica residuos de inicio y final en la secuencia de referencia de la proteína para la región correspondiente.

De los 44 epítomos para células B que cumplen los requerimientos de selección, 39 fueron señalados como epítomos por los servidores BepiPred 2.0 y la IEDB. Los 5 restantes fueron reconocidos por los servidores BepiPred 2.0 e iBCE-EL. El programa FBCPRED sólo sugirió uno de los epítomos que cubrió el resto de los requerimientos con una especificidad del 90%. De los epítomos seleccionados, 30 se ubican dentro del dominio RBD y 6 se ubican en la región HR2 (Tabla 18).

Tabla 18. Epítotos para células B que cumplen todos los requerimientos en la proteína S con sus respectivos valores de antigenicidad y porcentaje de residuos escondidos.

Posición	Epítoto	Bepipred 2.0	IEDB	iBCE-EL	FBCPRED (%)	Valor Antigenicidad	Residuos Escondidos (%)
178-189	DLEGKQGNFKNL	✓	✓	-	-	1.1816	41.67
377-388	FKCYGVSPTKLN ¹	✓	✓	-	-	1.5509	50.00
377-389	FKCYGVSPTKLN ¹	✓	✓	-	-	1.441	46.15
377-390	FKCYGVSPTKLN ¹	✓	✓	-	-	1.4748	50.00
378-389	KCYGVSPTKLN ¹	✓	✓	-	-	1.3195	41.67
378-390	KCYGVSPTKLN ¹	✓	✓	-	-	1.3683	46.15
378-391	KCYGVSPTKLN ¹	✓	✓	-	-	1.643	50.00
378-393	KCYGVSPTKLN ¹	✓	✓	0.85	-	2.0395	50.00
378-394	KCYGVSPTKLN ¹	✓	-	0.85	-	1.887	47.06
378-395	KCYGVSPTKLN ¹	✓	-	0.85	-	1.8687	50.00
379-390	CYGVSPTKLN ¹	✓	✓	-	-	1.4808	50.00
379-394	CYGVSPTKLN ¹	✓	-	0.85	-	2.0134	50.00
380-391	YGVSPKLN ¹	✓	✓	-	-	1.8347	50.00
380-393	YGVSPKLN ¹	✓	✓	-	-	2.2554	50.00
381-392	GVSPTKLN ¹	✓	✓	-	-	2.4277	50.00
381-393	GVSPTKLN ¹	✓	✓	-	-	2.4422	46.15
382-393	VSPTKLN ¹	✓	✓	-	-	2.4268	50.00
456-467	FRKSNLKPFRD ¹	✓	✓	-	-	0.7053	50.00
456-469	FRKSNLKPFRDIS ¹	✓	✓	-	-	0.6091	50.00
457-468	RKSNLKPFRDI ¹	✓	✓	-	-	0.6927	50.00
457-469	RKSNLKPFRDIS ¹	✓	✓	-	-	0.7514	46.15
457-470	RKSNLKPFRDIST ¹	✓	✓	-	-	0.4607	42.86
457-471	RKSNLKPFRDISTE ¹	✓	✓	-	-	0.4847	46.67
457-472	RKSNLKPFRDISTE ¹	✓	✓	-	-	0.4298	50.00
458-469	KSNLKPFRDIS ¹	✓	✓	-	-	0.7736	41.67
458-470	KSNLKPFRDIST ¹	✓	✓	-	-	0.444	38.46
458-471	KSNLKPFRDISTE ¹	✓	✓	-	-	0.4643	42.86
458-472	KSNLKPFRDISTE ¹	✓	✓	-	-	0.4014	46.67
460-471	NLKPFRDISTE ¹	✓	✓	-	-	0.5155	41.67
460-472	NLKPFRDISTE ¹	✓	✓	-	-	0.4346	46.15
630-641	TPTWRVYSTGSN	✓	-	0.85	-	0.4496	41.67
695-706	YTMSLGAENVA	✓	✓	-	-	0.5856	25.00
695-707	YTMSLGAENSVAY	✓	✓	-	-	0.5713	23.08
695-708	YTMSLGAENSVAYS	✓	✓	-	90%	0.633	21.43

696-707	TMSLGAENSVAY	✓	✓	-	-	0.5896	16.67
696-708	TMSLGAENSVAYS	✓	✓	-	-	0.6581	15.38
697-708	MSLGAENSVAYS	✓	✓	-	-	0.8473	16.67
1159-1170	HTSPDVDLGDIS ²	✓	✓	-	-	1.6408	41.67
1159-1171	HTSPDVDLGDISG ²	✓	✓	-	-	1.0906	30.77
1159-1172	HTSPDVDLGDISGI ²	✓	✓	-	-	1.0779	35.71
1160-1171	TSPDVDLGDISG ²	✓	✓	-	-	1.0857	25.00
1160-1172	TSPDVDLGDISGI ²	✓	✓	-	-	1.0715	30.77
1161-1172	SPDVDLGDISGI ²	✓	✓	-	-	1.0184	33.33

Notas: ¹ Epítomos ubicados dentro del dominio RBD; ² Epítomos ubicados dentro de la región HR2; Posición indica residuos de inicio y final en la secuencia de referencia de la proteína para la región correspondiente.

Se obtuvieron coberturas poblacionales por alta afinidad del 100% en los servidores NetMHCpan 4.1b y de la IEDB, mientras que en el programa MHCflurry se obtuvo el 97.4%. Las mutaciones de Ómicron tomadas en cuenta no afectaron las coberturas poblacionales hacia los alelos MHCI en el programa NetMHCpan 4.1b y la IEDB. Por su parte, la baja antigenicidad de algunos epítomos afectó la cobertura poblacional en todos los programas. Combinado con las mutaciones Ómicron, la antigenicidad afectó en mayor medida la cobertura poblacional en el programa NetMHCpan 4.1b. Después del refinamiento de la secuencia, las coberturas poblacionales mejoraron en su totalidad en todos los casos, obteniendo coberturas poblacionales finales de 100%, 100% y 97.4% (Tabla 19).

Tabla 19. Coberturas poblacionales de alelos MHCI iniciales y mejoradas con el refinamiento de la secuencia obtenida para la proteína S.

Cobertura	T CD8+					
	NetMHCpan 4.1b		IEDB		MHCflurry	
Alta afinidad	100		100		97.44	
	Inicial	Mejorada	Inicial	Mejorada	Inicial	Mejorada
Sin epítomos de baja antigenicidad	87.18	100 ¹	91.03	100 ¹	87.18	97.44 ¹
Sin epítomos afectados por mutaciones	98.72	100 ¹	100 ²	100	94.87	97.44 ¹
Sin epítomos afectados por mutaciones Ómicron	100 ²	100	100 ²	100	94.87	97.44 ¹
Sin epítomos de baja antigenicidad y afectados por mutaciones	85.9	100 ¹	91.03	100 ¹	87.18	97.44 ¹
Sin epítomos de baja antigenicidad y afectados por mutaciones Ómicron	85.9	100 ^{1 3}	91.03	100 ^{1 3}	87.18	97.44 ^{1 3}

Notas: ¹ Coberturas poblacionales mejoradas completamente tras el refinamiento; ² Condiciones que no afectaron la cobertura poblacional original del programa; ³ Coberturas poblacionales finales bajo los criterios de selección.

Ambos servidores de epítomos para MHCII permitieron la selección de epítomos que cubrieron el 100% de los alelos analizados. No hubo afectaciones a las coberturas poblacionales, ya que la antigenicidad y las mutaciones se tomaron en cuenta para la selección (Tabla 20).

Tabla 20. Porcentajes de coberturas poblacionales finales de alelos MHCII de la secuencia seleccionada para la proteína S.

Cobertura	NetMHCpan 4.0	IEDB
Final	100	100

7.2 Análisis final de las regiones multiepitópicas seleccionadas

7.2.1 Cobertura poblacional final promedio

En las Tablas 21 y 22 se puede observar que, en todas las proteínas exceptuando la proteína S, se obtuvo una cobertura menor en la considerada “mejorada” que la indicada de “alta afinidad” bajo los criterios de selección. No se pudo mejorar en ningún sentido la cobertura poblacional para MHCI de la proteína E y para MHCII de la proteína M. La mayor afectación es en estos últimos, ya que se perdió un 27.8% de la cobertura por alta afinidad, llevándola a ser la menor de todos los casos, junto con la cobertura de MHCII de la proteína E. Los epítomos seleccionados de la proteína E obtuvieron en su conjunto las coberturas poblacionales finales más bajas. Los mejores resultados se obtuvieron con la proteína S, logrando cubrir 99.15% de los alelos analizados para MHCI y todos los analizados para MHCII.

Tabla 21. Porcentajes de coberturas poblacionales finales promedio para MHCI.

Cobertura poblacional promedio	Proteína E	Proteína M	Proteína N	Proteína S
Alta afinidad	75.21	97.86	97.44	99.15
Afectada por baja antigenicidad y mutaciones Delta y Ómicron	66.24	79.91	93.16	88.04
Mejorada	66.24	88.46	95.73	99.15 ¹

Nota: ¹ Coberturas poblacionales mejoradas completamente tras el refinamiento.

Tabla 22. Porcentajes de coberturas poblacionales finales promedio para MHCII.

Cobertura poblacional promedio	Proteína E	Proteína M	Proteína N	Proteína S
Alta afinidad	64.82	88.89	92.59	100
Afectada por baja antigenicidad y mutaciones Delta y Ómicron	59.26	61.11	81.48	100 ¹
Mejorada	61.11	61.11	88.89	100

Nota: ¹ Condiciones que no afectaron la cobertura poblacional original del programa.

7.2.2 Cobertura poblacional por proteína

Para la proteína E, al menos uno de los tres programas sugirió cuando menos un epítipo que cumpliera todos los criterios de selección para 17 de los 19 alelos de HLA-A; y también sugirieron para 16 de los 32 alelos de HLA-B considerados. Para el grupo HLA-C, 17 de los 20 alelos fueron cubiertos. Finalmente, se cubrieron los 2 alelos de HLA-E y los 5 de HLA-G. Esto indica que, bajo estas condiciones, 57 de los 78 alelos analizados están cubiertos con al menos uno de los tres programas. Esto representa el 73.07% de cobertura poblacional (Fig. 5).

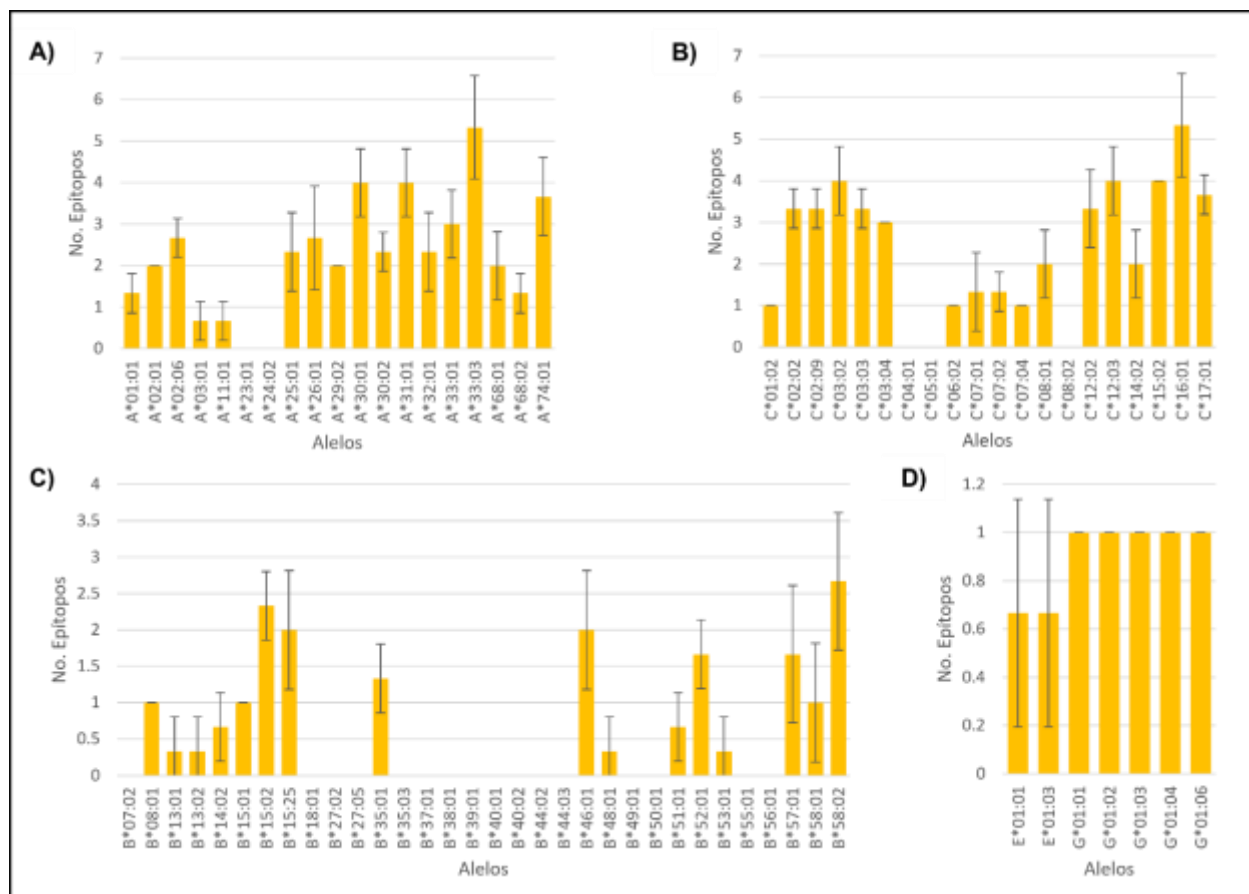


Figura 5. Representación gráfica de la cantidad de epítomos T CD8+ obtenidos de la proteína E que cumplieron los criterios de selección para cada alelo de MHC I. Las barras representan el promedio de la cantidad de epítomos de los tres programas utilizados que cubren el alelo correspondiente. Las barras de error en cada barra representan la desviación estándar de la misma. A) Alelos del grupo HLA-A, B) Alelos del grupo HLA-C, C) Alelos de grupo HLA-C, D) Alelos de los grupos HLA-E y HLA-G.

En la proteína M, al menos un epítomo que cumpliera todos los criterios de selección fue sugerido por cuando menos un programa utilizado para todos los alelos de los grupos HLA-A, HLA-C, HLA-E y HLA-G. Además, sugirieron al menos un epítomo para 30 de los 32 alelos de HLA-B considerados que cumplió con los criterios. Bajo estas condiciones, 76 de los 78 alelos analizados están cubiertos con al menos uno de los tres programas. Esto representa el 97.43% de cobertura poblacional (Fig. 6).

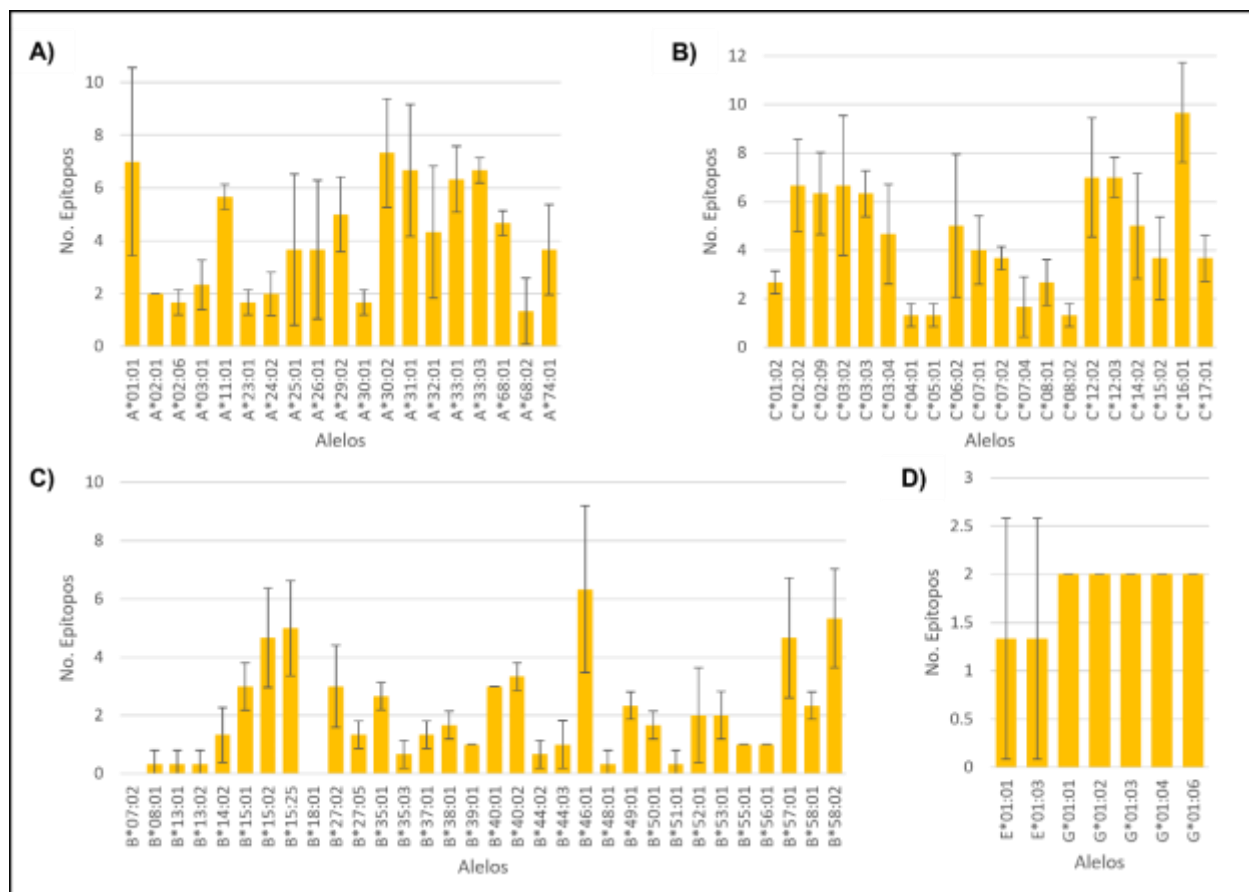


Figura 6. Representación gráfica de la cantidad de epítomos T CD8+ obtenidos de la proteína M que cumplieron los criterios de selección para cada alelo de MHC I. Las barras representan el promedio de la cantidad de epítomos de los tres programas utilizados que cubren el alelo correspondiente. Las barras de error en cada barra representan la desviación estándar de la misma. A) Alelos del grupo HLA-A, B) Alelos del grupo HLA-C, C) Alelos de grupo HLA-C, D) Alelos de los grupos HLA-E y HLA-G.

En la proteína N, al menos un epítomo que cumpliera todos los criterios de selección fue sugerido por cuando menos un programa utilizado para todos los alelos considerados. Esto se traduce en un 100% de cobertura poblacional (Fig. 7).

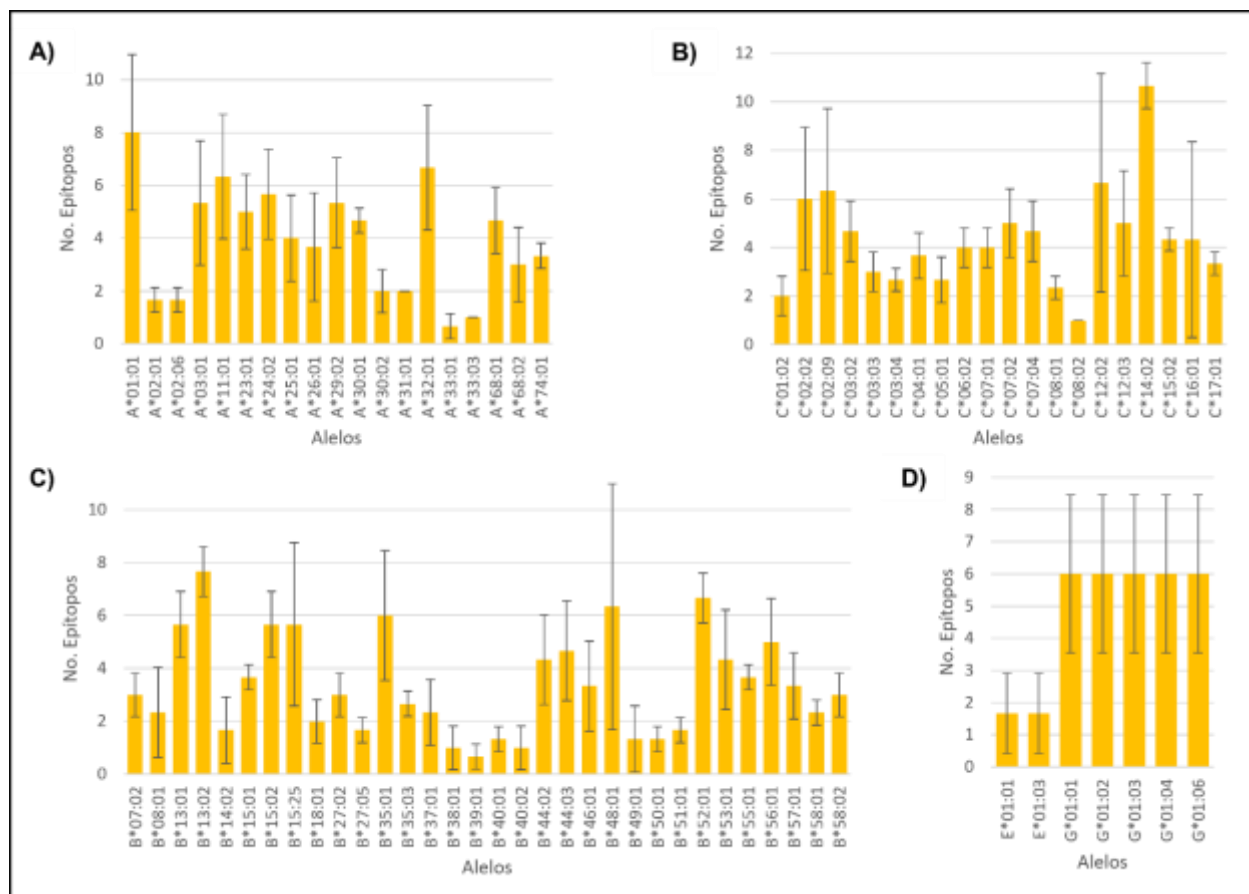


Figura 7. Representación gráfica de la cantidad de epítomos T CD8+ obtenidos de la proteína N que cumplieron los criterios de selección para cada alelo de MHC I. Las barras representan el promedio de la cantidad de epítomos de los tres programas utilizados que cubren el alelo correspondiente. Las barras de error en cada barra representan la desviación estándar de la misma. A) Alelos del grupo HLA-A, B) Alelos del grupo HLA-C, C) Alelos de grupo HLA-C, D) Alelos de los grupos HLA-E y HLA-G.

En la proteína S, al menos un epítomo que cumpliera todos los criterios de selección fue sugerido por cuando menos un programa utilizado para todos los alelos considerados. Esto se traduce en un 100% de cobertura poblacional (Fig. 8).

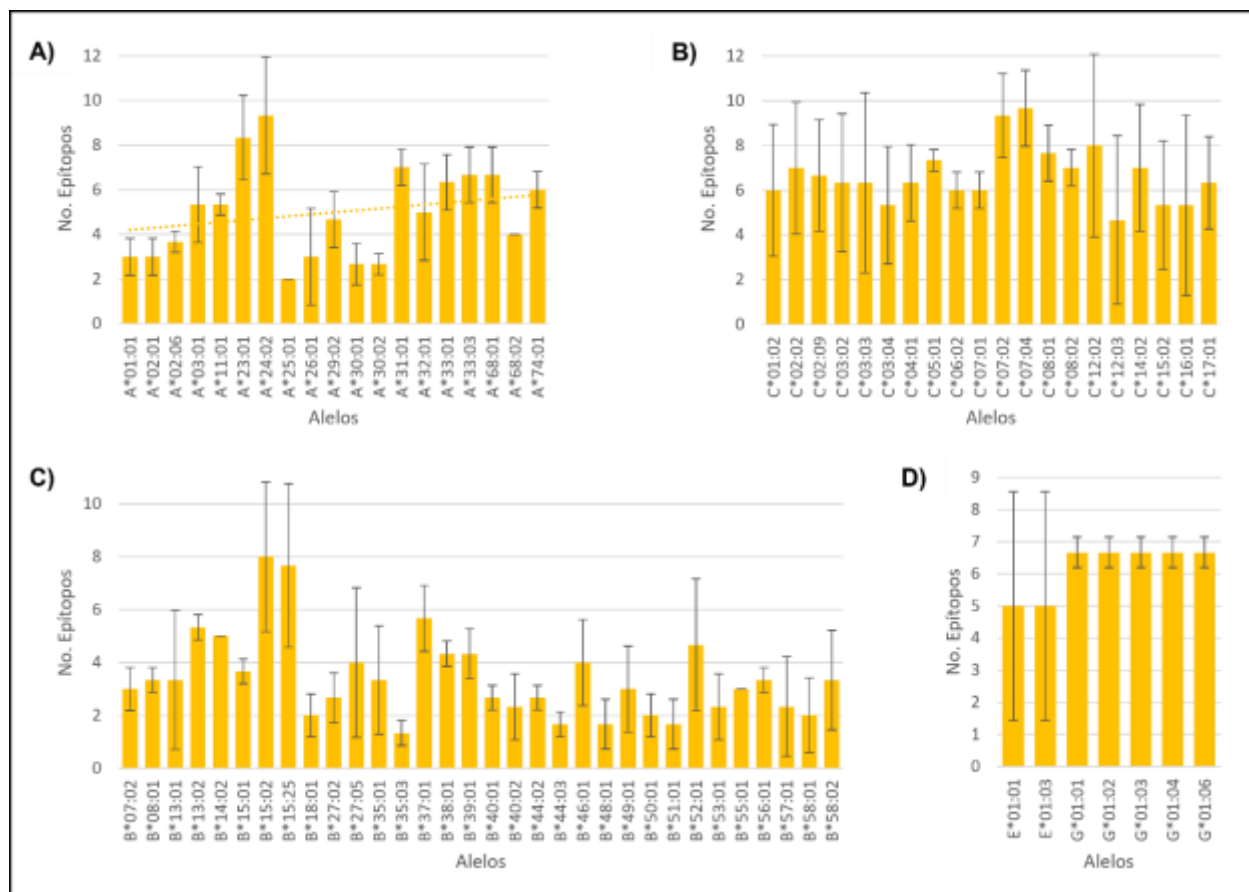


Figura 8. Representación gráfica de la cantidad de epítomos T CD8+ obtenidos de la proteína S que cumplieron los criterios de selección para cada alelo de MHC I. Las barras representan el promedio de la cantidad de epítomos de los tres programas utilizados que cubren el alelo correspondiente. Las barras de error en cada barra representan la desviación estándar de la misma. A) Alelos del grupo HLA-A, B) Alelos del grupo HLA-C, C) Alelos de grupo HLA-C, D) Alelos de los grupos HLA-E y HLA-G.

Para la proteína E, hay al menos un epítomo que cumpla los criterios de selección sugerido por cuando menos un programa que cubra 11 de los 15 alelos considerados del grupo HLA-DRB. Por otro lado, los 6 alelos DP están cubiertos y ninguno de los 6 alelos DQ lo están. En conjunto, 17 de los 27 alelos están cubiertos para un 63% de cobertura poblacional (Fig. 9).

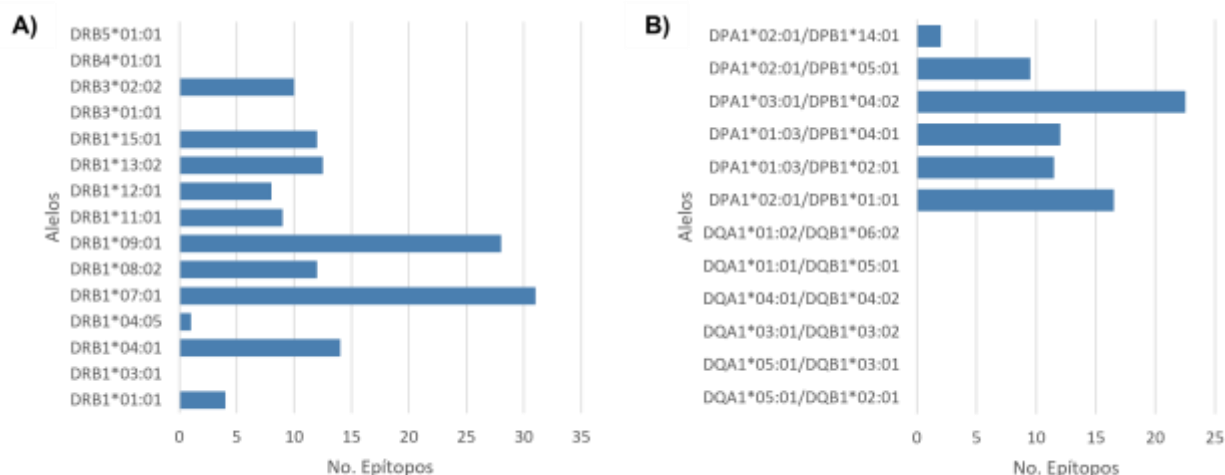


Figura 9. Representación gráfica de la cantidad de epítomos T CD4+ obtenidos de la proteína E que cumplieron los criterios de selección para cada alelo de MHCII. Las barras representan el promedio de la cantidad de epítomos de los tres programas utilizados que cubren el alelo correspondiente. No se grafica la desviación estándar por falta de un tercer programa que provea datos. A) Alelos del grupo HLA-DRB, B) Alelos de los grupos HLA-DP y HLA-DQ.

Para la proteína M, se cubren 12 de los 15 alelos considerados del grupo HLA-DRB con al menos un epítomo que cumpla los criterios de selección sugerido por cuando menos un programa. Por otro lado, 4 de los 6 alelos HLA-DP están cubiertos y 2 de los 6 alelos HLA-DQ lo están. En conjunto, 17 de los 27 alelos están cubiertos para un 66.7% de cobertura poblacional (Fig. 10).

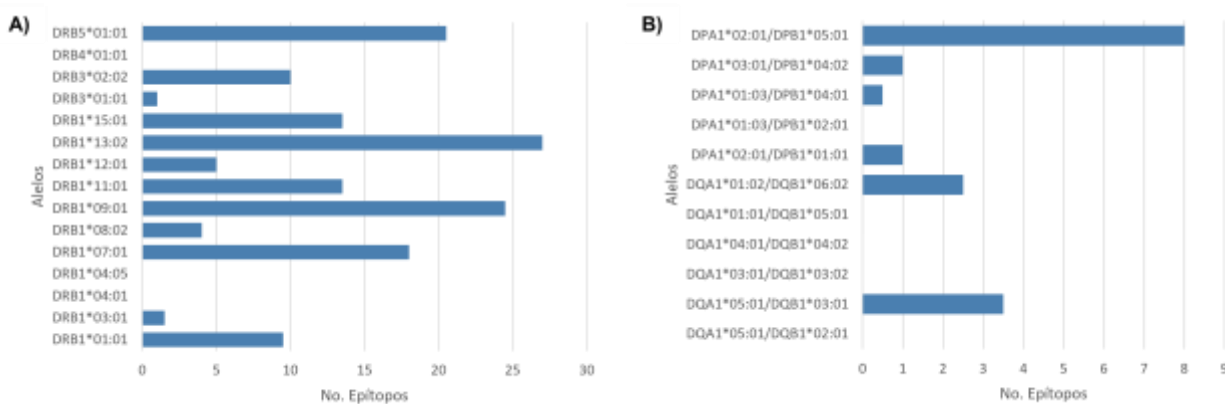


Figura 10. Representación gráfica de la cantidad de epítomos T CD4+ obtenidos de la proteína M que cumplieron los criterios de selección para cada alelo de MHCII. Las barras representan el promedio de la cantidad de epítomos de los tres programas utilizados que cubren el alelo correspondiente. No se grafica la desviación estándar por falta de un tercer programa que provea datos. A) Alelos del grupo HLA-DRB, B) Alelos de los grupos HLA-DP y HLA-DQ.

Para la proteína N, se cubren los 15 alelos considerados del grupo HLA-DRB con al menos un epítipo que cumpla los criterios de selección sugerido por cuando menos un programa. También, 4 de los 6 alelos HLA-DP y 5 de los 6 alelos HLA-DQ están cubiertos. En este caso, 24 de los 27 alelos están cubiertos para un 88.9% de cobertura poblacional (Fig. 11).

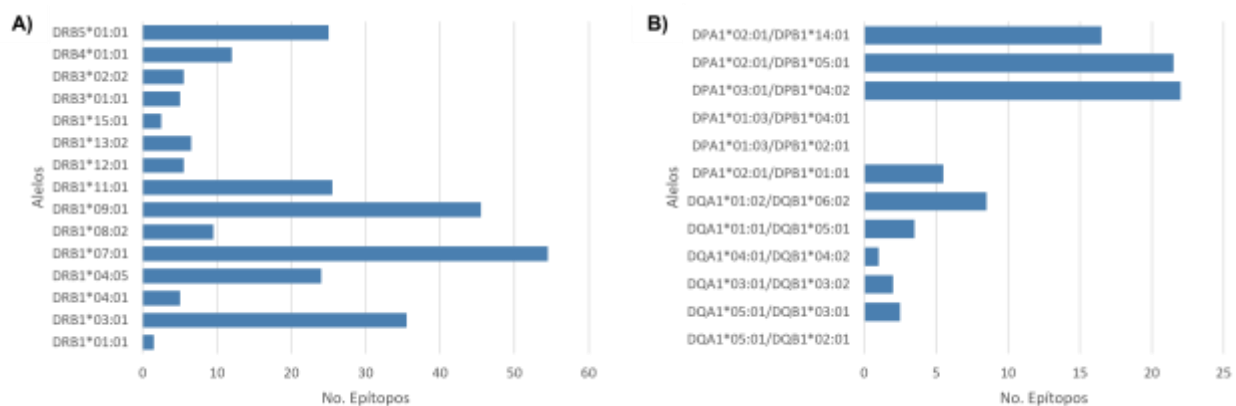


Figura 11. Representación gráfica de la cantidad de epítipos T CD4+ obtenidos de la proteína N que cumplieron los criterios de selección para cada alelo de MHCII. Las barras representan el promedio de la cantidad de epítipos de los tres programas utilizados que cubren el alelo correspondiente. No se grafica la desviación estándar por falta de un tercer programa que provea datos. A) Alelos del grupo HLA-DRB, B) Alelos de los grupos HLA-DP y HLA-DQ.

Para la proteína S, se cubren los 15 alelos considerados del grupo HLA-DRB con al menos un epítipo que cumpla los criterios de selección sugerido por cuando menos un programa. Además, los 6 alelos HLA-DP y 6 alelos HLA-DQ están cubiertos. En esta proteína los 27 alelos están cubiertos para un 100% de cobertura poblacional (Fig. 12).

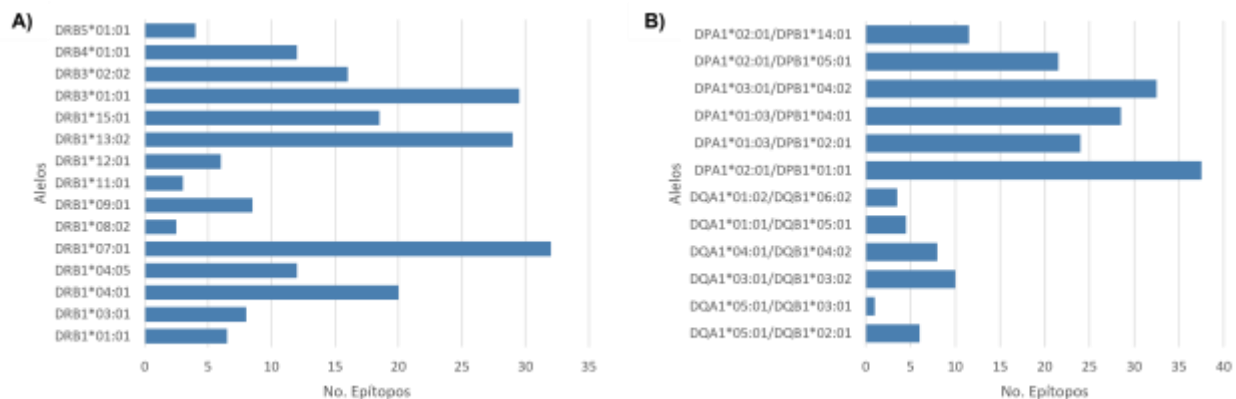


Figura 12. Representación gráfica de la cantidad de epítomos T CD4+ obtenidos de la proteína S que cumplieron los criterios de selección para cada alelo de MHCII. Las barras representan el promedio de la cantidad de epítomos de los tres programas utilizados que cubren el alelo correspondiente. No se grafica la desviación estándar por falta de un tercer programa que provea datos. A) Alelos del grupo HLA-DRB, B) Alelos de los grupos HLA-DP y HLA-DQ.

7.2.3 Afectaciones de nuevas variantes Ómicron

Para la selección de las regiones multiepitópicas se tomaron en cuenta las mutaciones de las variantes Ómicron BA.1, BA.2, BA.4 y BA.5. Con la aparición de las nuevas variantes Ómicron, se analizaron las afectaciones que las mutaciones de estas tienen sobre las regiones multiepitópicas seleccionadas. Se tomaron en cuenta las variantes BA.2.12.1, BA.2.75, BQ.1 y XBB.

De todas las nuevas mutaciones presentes en estas variantes, solamente 3 están presentes en las regiones multiepitópicas seleccionadas, y todas se encuentran en regiones de la proteína S. La mutación Q183E se presenta en la variante XBB. Las variantes BA.2.75, BQ.1 y XBB presentan la mutación N460K. Finalmente, la variante BA.2.12.1 presenta la mutación S704L. En el resto de las proteínas se repiten mutaciones tomadas en cuenta anteriormente o las nuevas mutaciones no están dentro de las regiones multiepitópicas.

Es importante resaltar que estas nuevas mutaciones no ocasionaron afectaciones a las coberturas poblacionales en ninguno de los programas de MHC I o MHC II utilizados. Sin embargo, sí ocasionaron una disminución en la cantidad de epítomos totales de cada tipo en la proteína S. En la Tabla 23 se enlistan la cantidad de epítomos afectados por cada una de las

mutaciones, siendo los B lineales los más afectados, con 20 epítomos. De estos, los 13 afectados por la mutación N460K se encuentran dentro del dominio RBD. En conjunto, todas estas nuevas mutaciones disminuyen a 72 el número de epítomos T CD8+, a 24 los B lineales y a 90 los T CD4+ (Tabla 24).

Tabla 23. Cantidad de epítomos afectados por cada mutación de las nuevas variantes Ómicron.

Mutación	Epítomos T CD8+	Epítomos B	Epítomos T CD4+
Q183E	-	1	-
N460K	3	13 ¹	8
S704L	5	6	-
Total	8	20	8

Nota: ¹ Cantidad de epítomos B lineales afectados que se encuentran en el dominio RBD.

Tabla 24. Comparativa entre epítomos totales bajo los criterios de selección planteados y el nuevo total afectado por las nuevas variantes Ómicron.

	Epítomos T CD8+	Epítomos B	Epítomos T CD4+
Epítomos totales iniciales	80	44	98
Epítomos totales actualizados	72	24	90

7.3 Diseños finales

7.3.1 Construcciones finales

Se generaron cuatro diseños basados en cada una de las proteínas analizadas para crear la vacuna CIB-GIVac. Cada uno con las regiones de alta afinidad obtenidas, el adyuvante PADRE, el péptido señal 'MKMKVLFFVATILVAWQCHA', una etiqueta de histidinas, la secuencia de retención al retículo endoplásmico, el codón de paro, los sitios de corte para las enzimas *SmaI*, *BamHI* y *SacI* y los enlazadores necesarios. Los genes que codifican para estas construcciones se nombraron como muestra la primera columna de la Tabla 25. La construcción generada a partir de la proteína S tuvo la mayor antigenicidad con un valor de 0.8858, mientras que la construcción SARS2_M tuvo la menor con un valor de 0.5729. La secuencia de mayor tamaño fue SARS2_S con 326aa (999pb) y la de menor tamaño la de la proteína E con 104aa (333pb). El uso de enlazadores entre las regiones de alta afinidad no fue necesario más que en la construcción SARS2_S, en donde se utilizó el enlazador KK siete veces y el GPGPG una vez. El enlazador KK fue seleccionado para unir la secuencia de epítomos al adyuvante PADRE después de las regiones de alta afinidad en tres de las cuatro construcciones, siendo SARS2_N la única

sin este enlazador. También, en todos los diseños se colocó al inicio el péptido señal y al final la cola de histidinas seguida de la secuencia de retención al retículo endoplásmico, ya que esta conformación fue la que obtuvo el mayor valor de antigenicidad.

Tabla 25. Diseños finales generados para producir la vacuna CIB-GIVac.

Diseño	Construcción	Tamaño secuencia (aa)	pb	Antigenicidad
SARS2_E	<p>cccgggMKMKVLFFVATILVAWQCHALVTLAILTA LRLCFVSEETGTLIVNSVNIIVNSLVKPSFYVYS RVKNLNSSRVPDLLVKKAKFVAAWTLKAAHHH HHHSEKDELtgaaggatccgagctc</p>	104	333	0.7327
SARS2_M	<p>cccgggMKMKVLFFVATILVAWQCHALSIFYASF RLFARTRSMWSFNPETNILLADSNGTITVEELK KLEQWNLVIGFVPLHGTILSELVIGAVILRGHDL PKEITVATSRTLSTYKLGASQRVAGDSGFAAYS RYRIGNYKLNTHSSSDNIALLFAYANRNRFL YIIFVLAAYRINWITGGIAIAMAkkAKFVAAWTL KAAHHHHHHSEKDELtgaaggatccgagctc</p>	212	657	0.5729
SARS2_N	<p>cccgggMKMKVLFFVATILVAWQCHAQGVPIINTN SSPDDQIGYYQRQKKQQTVTLLPAADLSPRWY FYFLGTGPEAGLPYGANPRQKRTATKAYNVTQ AFGRRGPEQTQGNFGDQELIRQRPQGLPNNT ASWFTALTQHGEDQASSRSSRTDYKHWPQ IAQFAPSASAFFGMSRIGMEVTPSGTWLTYTG AIKDDKDPNFKDQVILLNKHIDAYKTFPPTPEPK KDAKFVAAWTLKAAHHHHHHSEKDELtgaaggatc cgagctc</p>	246	759	0.7108
SARS2_S	<p>cccgggMKMKVLFFVATILVAWQCHAFGGFNFS QILPDPSPSKRSkkFPQSAPHGVVFLHVTYVkk kVAIHADQLTPTWRVYSTGSNVFQTRAKKSCC KFDEDDSEPVKGVKLHYkkYTMSLGAENSV YSkkTVYDPLQPELDSFKkkHTSPDVLGDISG IkkTEILPVSMKTSVGGPGDLEGKQGNFKNL DIADTTDAVRDPQTLEILDITPCSFTFKCYGVSP TKLNDLCFTFRKSNLKPFERDISTEIFYASVYAW NRKRISNAYYVGYLQPRTFLLVEKGIYQTSNFR VQPTESIVRkkAKFVAAWTLKAAHHHHHHSEK DELtgaaggatccgagctc</p>	326	999	0.8858

Notas: Letras resaltadas en verde indican sitio de corte *Sma*I; Letras subrayadas indican el péptido señal; Letras resaltadas en amarillo indican enlazadores; Letras resaltadas en cian indican cola de histidinas; Letras resaltadas en rojo indican señal de retención al retículo endoplásmico; Letras en rojo y negritas indican codón de paro; Letras resaltadas en magenta indican sitio de corte *Bam*HI; Letras resaltadas en azul indican sitio de corte *Sac*I; aa = aminoácidos; pb = pares de bases.

7.3.2 Diseño de primers específicos

Se presenta la selección de primers específicos para la amplificación de los genes que se utilizarán para la producción de la vacuna CIB-GIVac, tanto para productos de PCR como qPCR (Tabla 26). Todos cumplen con las especificaciones planteadas en la metodología. Cabe resaltar que, para el gen SARS2_M, se utilizó el mismo primer forward en las reacciones de PCR y qPCR.

Los pares de primers para ambos tipos de PCR para el gen SARS_E no son específicos exclusivamente a esta secuencia. El par diseñado para PCR es inespecífico con el mRNA que se traduce en la proteína FZD7 de *Pogona vitticeps*, dragón barbudo (NCBI No.: XM_020814479.1). También lo es para el transcrito perteneciente al pez *Salarias fasciatus* que traduce a la proteína LOC115387537 en su variante 1 y 2 (GenBank no. XM_030090280.1 y XM_030090281.1, respectivamente). Para qPCR, el par de oligos podría unirse al mRNA que traduce la proteína LOC105565014 de la hormiga *Vollenhovia emeryi* (GenBank no. XM_012017826.1). El par diseñado para qPCR del gen SARS2_S podría dar una reacción cruzada con parte del mRNA que genera una policétido sintasa (BGW36DRAFT_4236069) del hongo *Talaromyces proteolyticus* (GenBank no. XM_046219854.1). Dado que ninguno de estos organismos está involucrado en ninguno de los procesos para la producción de la vacuna, sus inespecificidades no son determinantes para evitar la selección de los primers.

Tabla 26. Primers elegidos para PCR y qPCR de los genes con sus correspondientes datos tomados de Primer3 y Primer-BLAST

Gen	Forward/ Reverse	Secuencia	Longitud	Posición Inicio	Tm	%GC	RNAfold (%)	Producto	Especificidad
SARS2_E PCR	Forward	TCCTGTTCTTCGTGGCCA	18	14	58.44	55.56	42.51	276	No específicos
	Reverse	GATGATGATGATGGGCCGC	19	289	58.82	57.89	88.85		
SARS2_E qPCR	Forward	TCAACGTCAGCCTGGTCAAG	20	149	60.25	55	73.81	142	No específicos
	Reverse	TGATGATGATGATGGGCCGC	20	290	60.61	55	88.18		
SARS2_M PCR	Forward	TACTTCATCGCCTCGTTCCG	20	67	59.9	55	91.37	504	Específicos
	Reverse	CTTCGCTTTCTTCGCCATCG	20	570	59.97	55	52.78		
SARS2_M qPCR	Reverse	TTCTTCGACGGTGATGGTGC	20	177	60.67	55	28.25	111	Específicos
SARS2_N PCR	Forward	GCGTCCCTATCAACACCAATTC	22	65	59.65	55	91.86	652	Específicos
	Reverse	TGATGATGATGATGGGCCGC	20	716	60.61	55	88.18		
SARS2_N qPCR	Forward	ATAACGTGACCCAGGCCTTC	20	263	59.75	55	86.21	153	Específicos
	Reverse	CCTGATCTTCCTTGCCATGC	20	415	58.97	55	92.41		
SARS2_S PCR	Forward	TCGTCTTCTGCATGTCACC	20	152	60.04	55	84.13	615	Específicos
	Reverse	ACACCGAGGCGAAGTAGATC	20	766	59.26	55	77.34		
SARS2_S qPCR	Forward	AAAGACCGTGTATGATCCGC	20	381	58.06	50	65.28	170	No específicos
	Reverse	GCTTGCCTTCCAGATCACC	19	550	58.51	57.89	85.05		

7.4 Caracterización *in silico* de las proteínas quiméricas

7.4.1 Evaluación de seguridad y caracterización de propiedades fisicoquímicas

El análisis de toxicidad predijo solamente 3 péptidos tóxicos entre todos los epítomos seleccionados de todas las proteínas, correspondientes a epítomos T CD4+ de la proteína S (Tabla 44). La eliminación de los residuos S1252 y C1253 se consideró, pero debido a la posible afectación a la cobertura poblacional y dado que no todos los epítomos que contienen a esos residuos son identificados como tóxicos, se decidió mantener la región intacta.

El análisis de alergenicidad se realizó para los cuatro constructos finales con los programas AllerTOP v2.0 (<https://www.ddg-pharmfac.net/AllerTOP/>) y Allergen FP v1.0 (<https://ddg-pharmfac.net/AllergenFP/>). Todas las proteínas fueron predichas como “No alergénicas” en ambos programas.

Con el programa ProtParam (<https://web.expasy.org/protparam/>) se calcularon las propiedades fisicoquímicas de los cuatro diseños finales para CIB-GIVac. El diseño de mayor tamaño y peso molecular fue el SARS2_S, con un peso de 36.94 kDa, mientras que el SARS2_E fue el más pequeño con un peso de 11.74 kDa. Los puntos isoeléctricos de todas las moléculas estuvieron cercanos al 9.5, por lo que todas son básicas. El programa indicó que 3 de las 4 moléculas son estables, siendo SARS2_N la única marcada como inestable. Sin embargo, para todos los diseños se calcularon tiempo de vida media en diferentes plataformas de expresión. Para células de mamífero *in vitro* fueron 30 h, para levaduras *in vivo* más de 20 h y para *E. coli in vivo* más de 10 h. La molécula SARS2_N fue la que tuvo el menor índice alifático y el menor promedio general de hidropatía, mientras que SARS2_E fue la que tuvo los valores más altos en ambos casos (Tabla 27).

Tabla 27. Propiedades fisicoquímicas de los cuatro diseños finales para CIB-GIVac.

Diseño	Tamaño (aa)	Peso molecular (kDa)	Punto isoeléctrico	Índice de inestabilidad	Índice alifático	GRAVY ¹
SARS2_E	104	11.74	9.64	33.48 / ESTABLE	118.94	0.417
SARS2_M	212	23.92	9.71	30.18 / ESTABLE	105.0	0.196
SARS2_N	246	27.78	9.46	43.88 / INESTABLE	60.0	-0.709
SARS2_S	326	36.94	9.39	35.19 / ESTABLE	73.53	-0.419

Nota: ¹ Promedio general de hidropatia (siglas en inglés).

7.4.2 Modelaje 3D

7.4.2.1 Modelaje de la Proteína E

El modelo predicho seleccionado para el refinamiento tuvo un C-score de -2.6. En el mismo se observan 3 regiones de hélices α y algunas regiones de espiral. La diferencia más clara con el modelo refinado son las posiciones que toman las regiones (Fig. 13). Tras el refinamiento, aumentaron los valores de RMSD y porcentaje en regiones favorables de Ramachandran. Por su parte, disminuyeron los valores de MolProbity, Clash score, rotámeros deficientes y energía GALAXY. El valor de rotámeros deficientes disminuyó en su totalidad (Tabla 28).

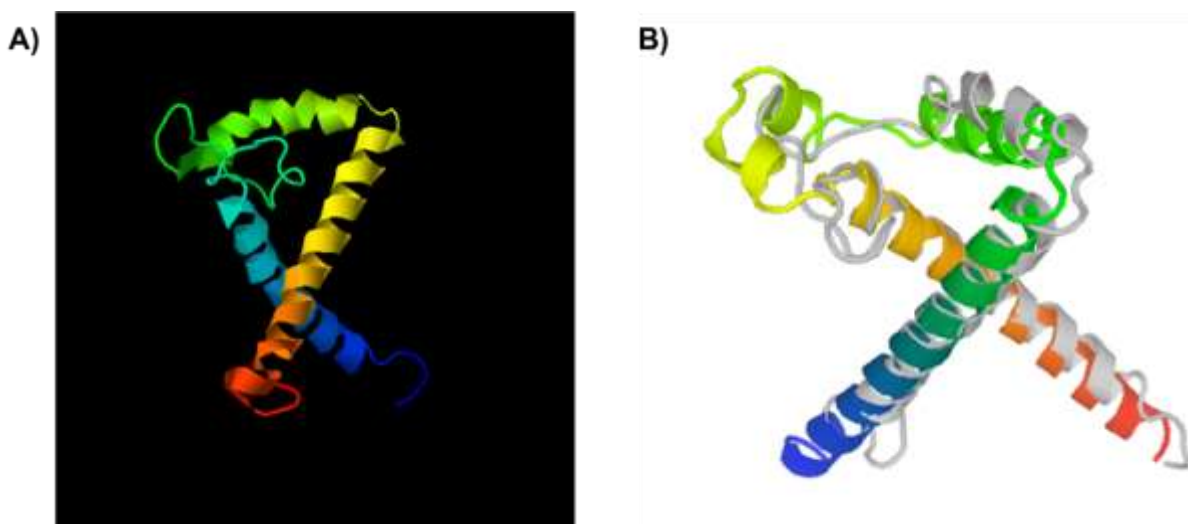


Figura 13. Modelo tridimensional inicial y refinado de la proteína SARS2_E. A) Modelo sugerido por programa I-TASSER, C-score: -2.60. B) Modelo sugerido con programa GalaxyRefine2.

Tabla 28. Valores provistos por GalaxyRefine2 para modelos inicial y refinado de SARS2_E.

Modelo	RMSD	MolProbity	Clash score	Rotámeros deficientes	Rama favorables (%)	Energía GALAXY
Inicial	0.000	3.008	6.8	11.8	72.5	-161.76
Refinado	4.567	1.142	2.1	0.0	97.1	-2167.14

Dos de los aminoácidos de la molécula se encuentran en las regiones no favorables del gráfico de Ramachandran, siendo la serina de la posición 71 y la asparagina de la posición 51 (Fig. 14). El 94.9% de los aminoácidos se encuentran en las regiones favorables y el 97.9% en conjunto con las regiones adicionales permitidas. El factor general de calidad calculado con ERRAT fue de 95.83% (Tabla 29).

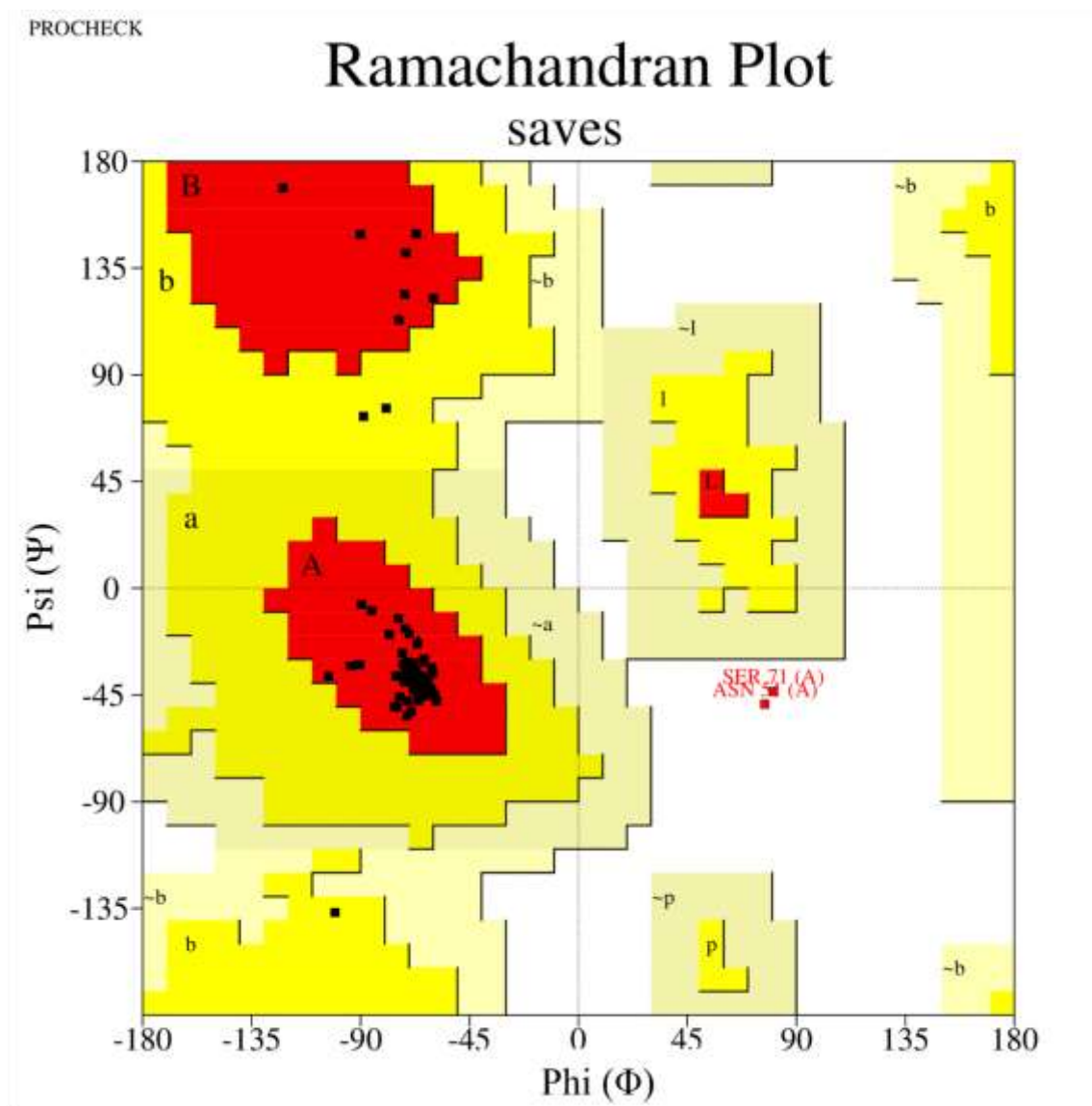


Figura 14. Gráfico de Ramachandran de SARS2_E generado por PROCHECK. Zonas rojas son “regiones favorables”. Zonas amarillas son “regiones adicionales permitidas”. Zonas color beige son “regiones extendidas permitidas”. Zonas blancas son “regiones no permitidas”. A/a/~a indican estructuras de hélices α dextrógiras. B/b/~b indican estructuras de láminas β . L/l/~l indican estructuras de hélices α levógiras. p/~p Indican estructuras de hélices ϵ . ■ Residuos que no son prolina. ▲ Prolina. ■ Residuos en regiones no permitidas y extendidas permitidas, descritos con su respectiva abreviación.

Tabla 29. Desglose de distribución de los residuos de SARS2_E en las regiones del gráfico de Ramachandran y factor general de calidad.

Porcentaje de residuos en regiones favorables	94.9
Porcentaje de residuos en regiones adicionales permitidas	3.0
Porcentaje de residuos en regiones extendidas permitidas	0.0
Porcentaje de residuos en regiones no permitidas	2.0
Factor general de calidad (%)	95.83

Notas: Distribución de los residuos obtenida en programa PROCHECK. Sólo se toman en cuenta los residuos que no son glicinas ni prolinas, en este caso 99 de los 104 totales. Factor general de calidad obtenido con programa ERRAT. Este indica el porcentaje de los residuos de la proteína que están debajo de límite de rechazo, idealmente debe ser igual o mayor a 95%.

7.4.2.2 Modelaje de la Proteína M

El modelo predicho seleccionado para el refinamiento tuvo un C-score de -2.68. En el mismo se observan 2 regiones de hélices α , 2 de láminas β y algunas regiones de espiral. La diferencia más clara con el modelo refinado son las posiciones que toman las regiones (Fig, 15). Tras el refinamiento, aumentaron los valores de RMSD y porcentaje en regiones favorables de Ramachandran. Por su parte, disminuyeron los valores de MolProbity, Clash score, rotámeros deficientes y energía GALAXY (Tabla 30).

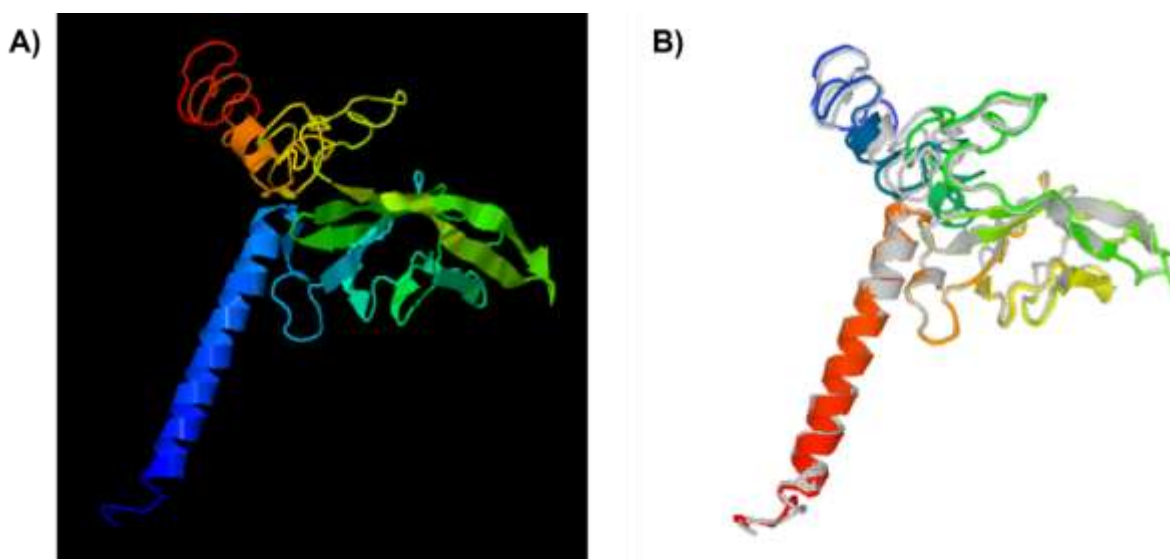


Figura 15. Modelo tridimensional inicial y refinado de la proteína SARS2_M. A) Modelo sugerido por programa I-TASSER, C-score: -2.68. B) Modelo sugerido con programa GalaxyRefine2.

Tabla 30. Valores provistos por GalaxyRefine2 para modelos inicial y refinado de SARS2_M.

Modelo	RMSD	MolProbity	Clash score	Rotámeros deficientes	Rama favorables (%)	Energía GALAXY
Inicial	0.000	4.012	52.5	17.6	66.2	816.39
Refinado	2.879	2.185	7.0	1.1	78.1	-3428.68

Dos de los aminoácidos de la molécula se encuentran en las regiones no favorables del gráfico de Ramachandran, siendo la leucina de la posición 84 y la treonina de la posición 178. Además, varios aminoácidos se encuentran en las regiones extendidas permitidas (Fig. 16). Sólo el 69.9% de los aminoácidos se encuentran en las regiones favorables, pero el 94.9% está en esas regiones en conjunto con las regiones adicionales permitidas. El factor general de calidad calculado con ERRAT fue de 69.31% (Tabla 31).

PROCHECK

Ramachandran Plot

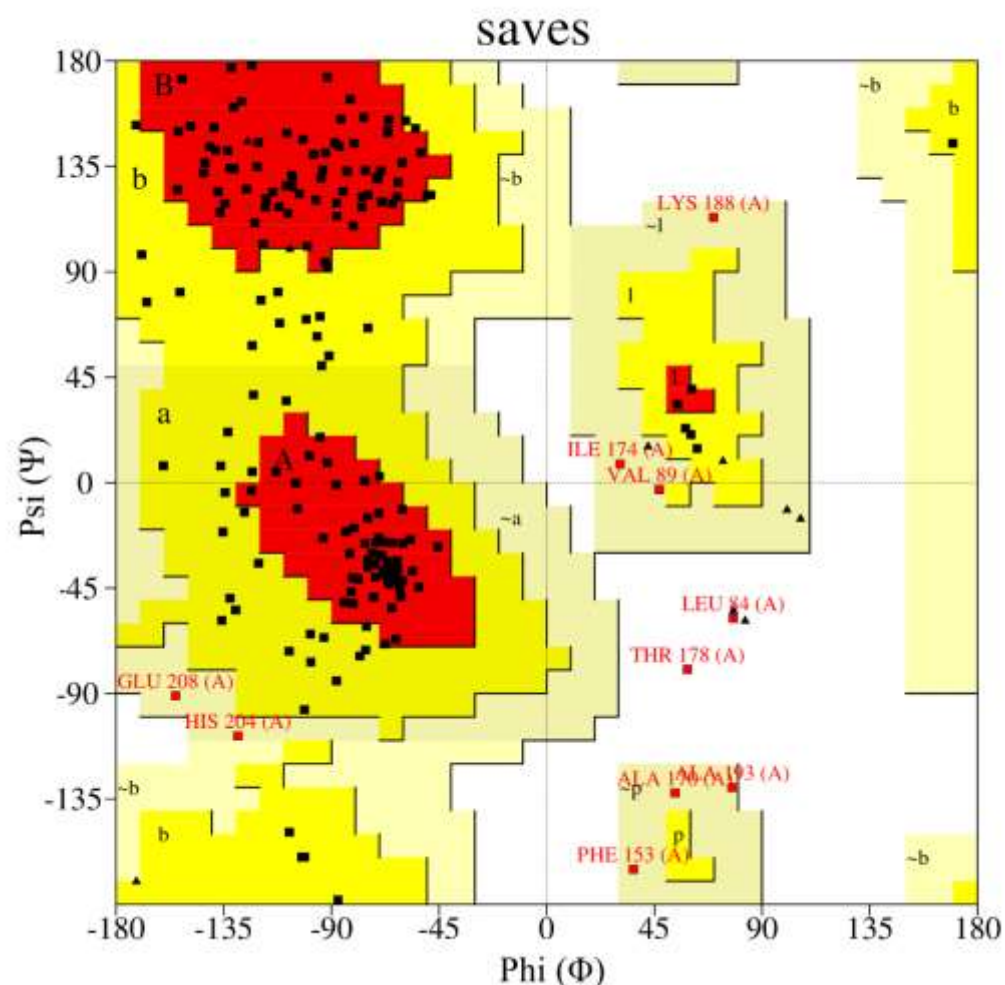


Figura 16. Gráfico de Ramachandran de SARS2_M generado por PROCHECK. Zonas rojas son “regiones favorables”. Zonas amarillas son “regiones adicionales permitidas”. Zonas color beige son “regiones extendidas permitidas”. Zonas blancas son “regiones no permitidas”. A/a/~a indican estructuras de hélices α dextrógiras. B/b/~b indican estructuras de láminas β . L/l/~l indican estructuras de hélices α levógiras. p/~p Indican estructuras de hélices ϵ . ■ Residuos que no son prolina.▲Prolina. ■ Residuos en regiones no permitidas y extendidas permitidas, descritos con su respectiva abreviación.

Tabla 31. Desglose de distribución de los residuos de SARS2_M en las regiones del gráfico de Ramachandran y factor general de calidad.

Porcentaje de residuos en regiones favorables	69.9
Porcentaje de residuos en regiones adicionales permitidas	25.0
Porcentaje de residuos en regiones extendidas permitidas	4.1
Porcentaje de residuos en regiones no permitidas	1.0
Factor general de calidad (%)	69.31

Notas: Distribución de los residuos obtenida en programa PROCHECK. Sólo se toman en cuenta los residuos que no son glicinas ni prolinas, en este caso 196 de los 212 totales. Factor general de calidad obtenido con programa ERRAT. Este indica el porcentaje de los residuos de la proteína que están debajo de límite de rechazo, idealmente debe ser igual o mayor a 95%.

7.4.2.3 Modelaje de la Proteína N

El modelo predicho seleccionado para el refinamiento tuvo un C-score de -4.81. En el mismo se observan una variedad de estructuras de todos los tipos. La diferencia más clara con el modelo refinado son las posiciones que toman las regiones (Fig. 17). Tras el refinamiento, aumentaron los valores de RMSD y porcentaje en regiones favorables de Ramachandran. Por su parte, disminuyeron los valores de MolProbity, Clash score, rotámeros deficientes y energía GALAXY (Tabla 32).

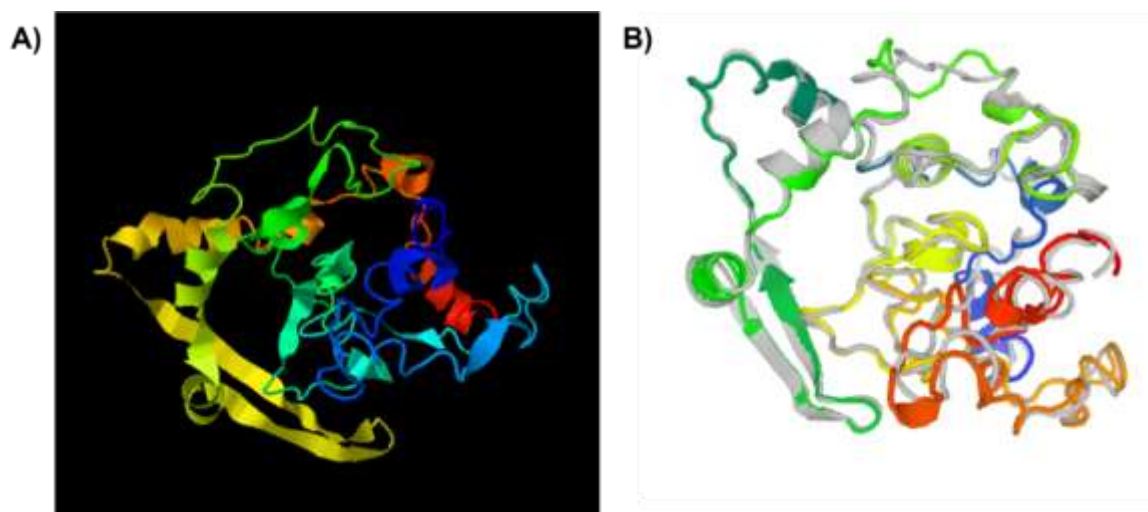


Figura 17. Modelo tridimensional inicial y refinado de la proteína SARS2_N. A) Modelo sugerido por programa I-TASSER, C-score: -4.81. B) Modelo sugerido con programa GalaxyRefine2.

Tabla 32. Valores provistos por GalaxyRefine2 para modelos inicial y refinado de SARS2_N.

Modelo	RMSD	MolProbity	Clash score	Rotámeros deficientes	Rama favorables (%)	Energía GALAXY
Inicial	0.000	3.632	18.6	17.0	57.4	261.19
Refinado	1.603	1.895	3.8	0.5	80.3	-3660.38

Cuatro de los aminoácidos de la molécula se encuentran en las regiones no favorables del gráfico de Ramchandran, siendo estos la fenilalanina de la posición 7, la glutamina de la posición 34, la prolina de la posición 116 y la leucina de la posición 231. Además de estos, sólo 5 residuos se encuentran en las regiones extendidas permitidas (Fig. 18). El 73.9% de los aminoácidos se encuentran en las regiones favorables y el 96.6% en conjunto con las regiones adicionales permitidas. El factor general de calidad calculado con ERRAT fue de 90.05% (Tabla 33).

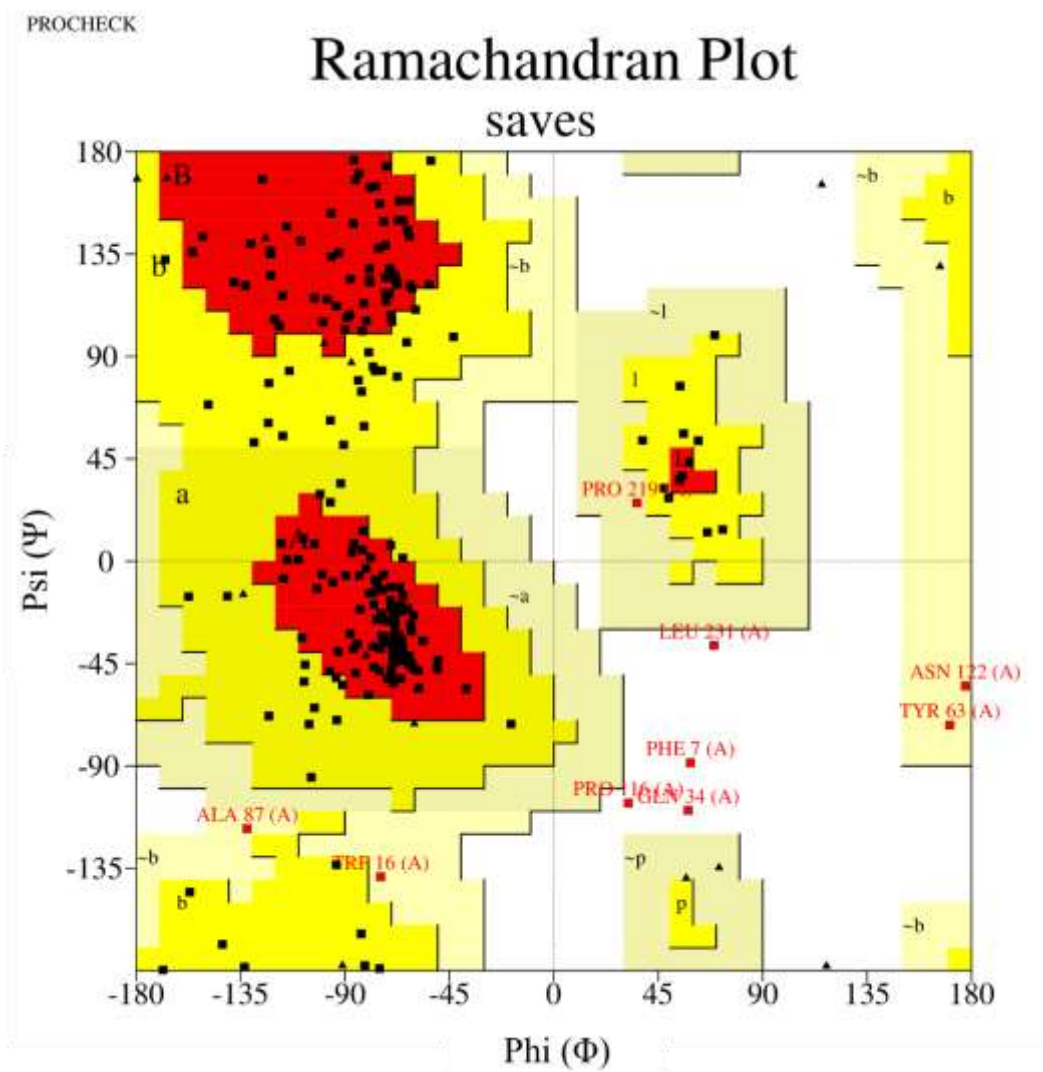


Figura 18. Gráfico de Ramachandran de SARS2_N generado por PROCHECK. Zonas rojas son “regiones favorables”. Zonas amarillas son “regiones adicionales permitidas”. Zonas color beige son “regiones extendidas permitidas”. Zonas blancas son “regiones no permitidas”. A/a/~a indican estructuras de hélices α dextrógiras. B/b/~b indican estructuras de láminas β . L/l/~l indican estructuras de hélices α levógiras. p/~p Indican estructuras de hélices ϵ . ■ Residuos que no son prolina.▲Prolina. ■ Residuos en regiones no permitidas y extendidas permitidas, descritos con su respectiva abreviación.

Tabla 33. Desglose de distribución de los residuos de SARS2_N en las regiones del gráfico de Ramachandran y factor general de calidad.

Porcentaje de residuos en regiones favorables	73.9
Porcentaje de residuos en regiones adicionales permitidas	22.7
Porcentaje de residuos en regiones extendidas permitidas	1.9
Porcentaje de residuos en regiones no permitidas	1.4
Factor general de calidad (%)	90.05

Notas: Distribución de los residuos obtenida en programa PROCHECK. Sólo se toman en cuenta los residuos que no son glicinas ni prolinas, en este caso 211 de los 246 totales. Factor general de calidad obtenido con programa ERRAT. Este indica el porcentaje de los residuos de la proteína que están debajo de límite de rechazo, idealmente debe ser igual o mayor a 95%.

7.4.2.4 Modelaje de la Proteína S

El modelo predicho seleccionado para el refinamiento tuvo un C-score de -2.68. En el mismo se observan una variedad de estructuras de todos los tipos. La diferencia más clara con el modelo refinado son las posiciones que toman las regiones y la disminución de regiones de láminas β (Fig. 19). Tras el refinamiento, aumentaron los valores de RMSD, porcentaje en regiones favorables de Ramachandran y Clash score. Por su parte, disminuyeron los valores de GDT-HA, MolProbity y rotámeros deficientes. Cabe destacar que no es deseable el aumento que se dio en el Clash score (Tabla 34).

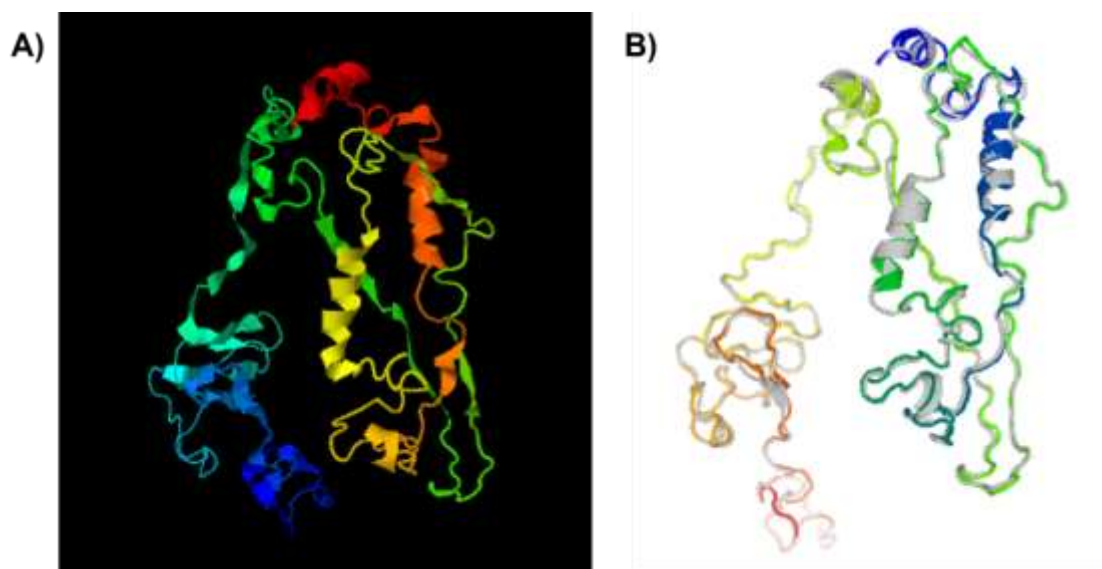


Figura 19. Modelo tridimensional inicial y refinado de la proteína SARS2_S. A) Modelo sugerido por programa I-TASSER, C-score: -2.68. B) Modelo sugerido con programa GalaxyRefine.

Tabla 34. Valores provistos por GalaxyRefine para modelos inicial y refinado de SARS2_S.

Modelo	GDT-HA	RMSD	MolProbity	Clash score	Rotámeros deficientes	Rama favorables (%)
Inicial	1.0000	0.000	3.505	14.9	14.4	54.9
Refinado	0.8351	0.692	2.383	16.4	0.7	84.6

En las regiones no favorables del gráfico de Ramchandran se encontraron 11 de los residuos de esta proteína, entre ellos dos prolinas, las cuales no se consideran para los porcentajes de distribución (Fig. 20). El 72.3% de los aminoácidos se encuentran en las regiones favorables y el 94.1% en conjunto con las regiones adicionales permitidas. Los 9 residuos considerados para el porcentaje de distribución representan el 3.1% del total, el mayor entre los cuatro diseños. El factor general de calidad ERRAT fue de 66.8%, el menor de los cuatro diseños (Tabla 35).

PROCHECK

Ramachandran Plot

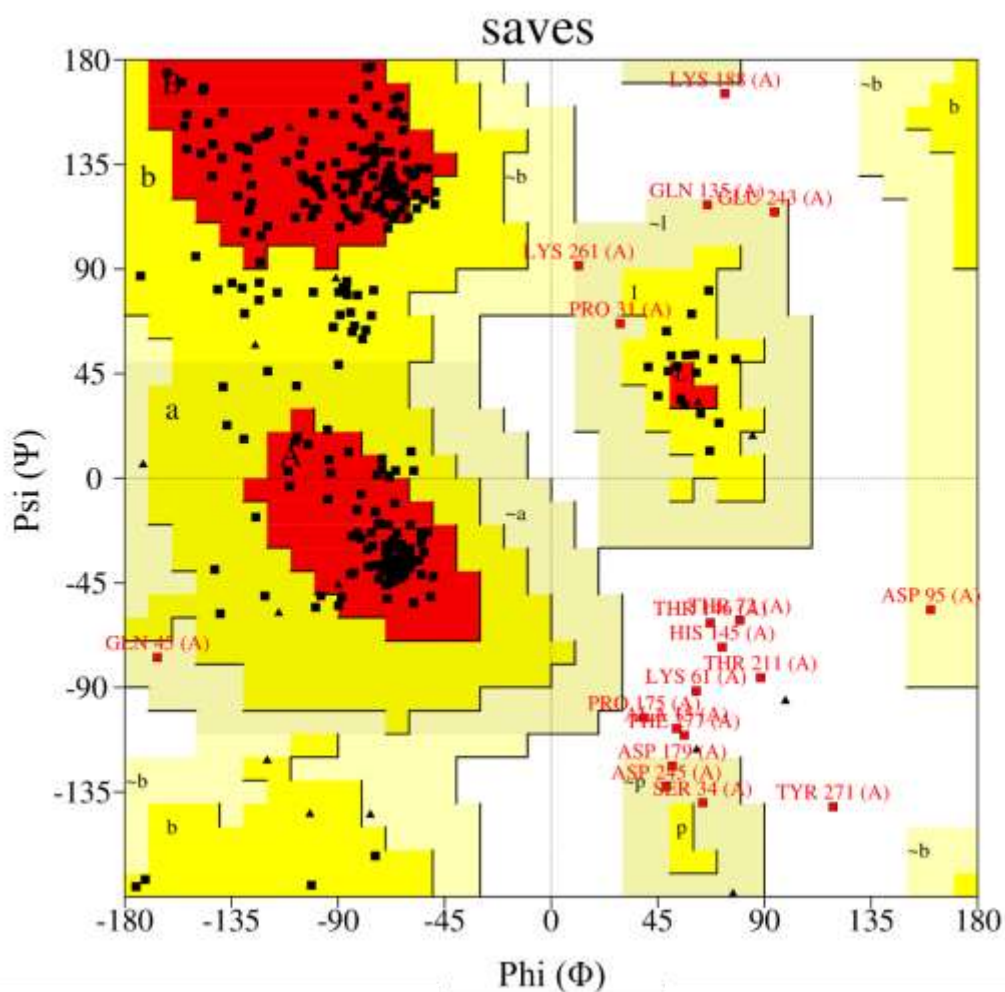


Figura 20. Gráfico de Ramachandran de SARS2_S generado por PROCHECK. Zonas rojas son “regiones favorables”. Zonas amarillas son “regiones adicionales permitidas”. Zonas color beige son “regiones extendidas permitidas”. Zonas blancas son “regiones no permitidas”. A/a/~a indican estructuras de hélices α dextrógiras. B/b/~b indican estructuras de láminas β . L/l/~l indican estructuras de hélices α levógiras. p/~p Indican estructuras de hélices ϵ . ■ Residuos que no son prolina. ▲ Prolina. ■ Residuos en regiones no permitidas y extendidas permitidas, descritos con su respectiva abreviación.

Tabla 35. Desglose de distribución de los residuos de SARS2_S en las regiones del gráfico de Ramachandran y factor general de calidad.

Porcentaje de residuos en regiones favorables	72.3
Porcentaje de residuos en regiones adicionales permitidas	21.8
Porcentaje de residuos en regiones extendidas permitidas	2.8
Porcentaje de residuos en regiones no permitidas	3.1
Factor general de calidad (%)	66.8

Notas: Distribución de los residuos obtenida en programa PROCHECK. Sólo se toman en cuenta los residuos que no son glicinas ni prolinas, en este caso 289 de los 326 totales. Factor general de calidad obtenido con programa ERRAT. Este indica el porcentaje de los residuos de la proteína que están debajo de límite de rechazo, idealmente debe ser igual o mayor a 95%.

7.4.3 Selección de péptidos para la producción de sueros hiperinmunes

En la tabla 36 se exponen los péptidos que fueron seleccionados para la inmunización de ratones para la producción de sueros hiperinmunes que sirvan para identificar la presencia y cuantificar las proteínas a producir con los genes diseñados.

Tabla 36. Péptidos diseñados para la producción de sueros hiperinmunes.

Gen	Secuencia Péptido	Longitud	Antigenicidad	Solubilidad en agua
SARS2_N	GAIKLDDKDPNFKDQ	15aa	1.9373	Buena
	ATKAYNVTQAFGRRG	15aa	0.7146	Buena
SARS2_M	PKEITVATSRTLSTY	14aa	0.8084	Buena
	GDSGFAAYSRYRIGNYK	17aa	0.7313	Buena
SARS2_E	VKPSFYVYSRVKLNLSR	18aa	0.8946	Buena
	RVKLNLSRVDPDLL	14aa	0.8807	Buena

8. DISCUSIÓN

8.1 Selección y refinamiento de las regiones multiepitópicas

Este diseño vacunal busca generar una respuesta inmune en las tres etapas del proceso de infección. Primero, antes de la unión del virus a la célula a través de la producción de anticuerpos neutralizantes hacia los epítomos B lineales de las regiones externas de las proteínas M, E y S (principalmente en el dominio RBD). Luego durante el proceso de fusión de membranas en el que se da un cambio conformacional de la proteína S liberando otras regiones como HR2 donde podría haber unión de anticuerpos neutralizantes (Huang *et al.*, 2020; Mercurio *et al.*, 2021). Finalmente, después de la internalización del virus en la célula con la presentación de epítomos por los complejos principales de histocompatibilidad.

Durante la selección se buscaron epítomos que sean reconocidos por los linfocitos T CD4+ y T CD8+ de alta afinidad a los MHC y antigénicos que cubrieran el mayor porcentaje de la población mundial. Sin embargo, el usar los valores de antigenicidad como criterio de selección de epítomos afecta en gran medida la cantidad de epítomos posibles a seleccionar (Bhattacharya *et al.*, 2020). Los resultados mostrados en este estudio soportan la idea de que los valores de afinidad predichos a MHCI, MHCII y epítomos lineales para células B no indican que estos epítomos sean antigénicos. Por lo que es necesario incluir uno o varios análisis de predicción de antigenicidad durante la selección de epítomos vacunales, algo que se logró en este estudio durante el proceso de refinamiento.

Además de esto, se ha buscado evitar sitios de glicosilaciones para la selección de epítomos B lineales de la proteína S, ya que estas protegen a los epítomos de los anticuerpos neutralizantes (Brun *et al.*, 2021). Por otro lado, la especificidad de las vacunas multiepitópicas aumenta al enfocarse en epítomos que se encuentren expuestos fuera de la membrana del virus (Dagur, 2020). Debido a esto, es que se decide el solamente analizar ciertas regiones de las proteínas membranales analizadas.

Como se ha descrito en la metodología de este trabajo, durante la selección de epítomos se tomaron en cuenta las diferentes mutaciones conocidas al momento. Al realizar esa parte del trabajo, las variantes de interés circulantes eran Alfa, Beta, Gamma y Delta. Sin embargo,

durante el proceso de refinamiento ya se había dado la aparición de la variante Ómicron. Para cuando esta se finalizó, las variantes que antes se consideraban de interés ya habían sido sustituidas por los linajes de Ómicron. Por ello, si bien se consideraron todas las mutaciones registradas en variantes de interés y preocupación a la fecha, en la toma de decisiones se le dio un menor valor a las variantes que sólo habían sido de interés, luego a las variantes Alfa, Beta y Gamma y el mayor valor a las variantes Delta y Ómicron. Las mutaciones provenientes de la variante Delta fueron aún tomadas en cuenta por dos razones: (1) aún era una variante de preocupación vigente al tiempo de realizar este paso y (2) la predominancia que esta variante tuvo por sobre otras durante un amplio periodo de tiempo (Chavda y Apostolopoulos, 2022). Debido a esto, se decidió sólo considerar a las variantes Delta y Ómicron (linajes BA.1, BA.2, BA.4 y BA.5) para el cálculo de las coberturas poblacionales. Los nuevos linajes de Ómicron han servido para evaluar la vigencia de los diseños vacunales generados en este trabajo.

Iniciando con la proteína E, las variantes Ómicron y Beta son las únicas variantes de preocupación en las que se han reportado mutaciones (Nextstrain, 2023b). La mutación P71L de la variante Beta afectaría la cobertura poblacional en el programa MHCflurry para MHCI si se tomara en consideración, así como en ambos programas del MHCII, eliminando epítomos antigénicos. Afortunadamente esta mutación no se ha vuelto a reportar en otras variantes, por lo que es posible que no se vuelva presentar.

Por otro lado, la mutación T9I está presente en todos los linajes de Ómicron, lo que genera afectaciones al momento de hacer el cálculo final de cobertura poblacional. Esta mutación afecta la región de alta afinidad 'FVSEETGTLIVNSV' (residuos 4-14), la cual presenta la particularidad de no contener ningún epítomo antigénico y está marcada en rojo en la Fig. 1. Con estos criterios, la decisión debería haber sido dejar fuera toda la región por no cumplir con los criterios de selección. Sin embargo, se optó por dejarla ya que el quitarla eliminaba por completo la cobertura poblacional de entre 6 y 8 alelos de MHCI dependiendo el programa y uno de MHCII en ambos programas. Si bien el objetivo de los criterios de selección es poder obtener los epítomos de mejor calidad que cubran la mayor cantidad de alelos posibles, al tener una proteína tan pequeña las opciones son muy reducidas y se complica el obtener un alto porcentaje de cobertura poblacional. Por ello, se optó por dejar la región con la idea de que

podiera ser presentada por los complejos principales de histocompatibilidad gracias a su alta afinidad y pudiera unirse a los receptores tipo Toll (TLRs), aunque la antigenicidad predicha no lo sugiera. Desgraciadamente, esta parece ser una mutación que perdurará mientras los linajes de Ómicron sean los predominantes. Incluso se ha reportado una nueva mutación en la zona, la T11A, en las variantes BA.2.75 y XBB, los cuales están en circulación actualmente (Nextstrain, 2023b). A pesar de esto, si tomamos en cuenta que cuando menos un epítipo de alta afinidad es reconocido por alelos con al menos uno de los programas, la cobertura poblacional sería del 73.07% para MHCI y 63% para MHCII. Cabe resaltar que los alelos que pudieran reconocer esta región no fueron detectados por ninguno de los programas, por lo que realmente son la única alternativa de una posible cobertura hacia estos grupos poblacionales. Si bien esto último es importante, si se realizara una revisión de este diseño conociendo que la mutación perdura y hay nuevas en la región, no se seleccionaría esta región. Esto podría ser benéfico para el diseño final ya que esta región posiblemente está disminuyendo la antigenicidad total del diseño, recordando que no hay epítopos antigénicos en ella. Esto es crucial ya que un valor bajo de antigenicidad indica una menor propensión a que haya una respuesta inmune hacia el epítipo.

En las regiones multiepitópicas de la proteína M encontramos seis secciones macadas en rojo, indicando que no cumplen con los criterios de selección (Fig. 2). A estas las podemos dividir en dos grupos de tres secciones cada uno: (1) afectados por antigenicidad y (2) afectados por mutaciones. En el primer caso tenemos a los residuos 93 y 163 junto con la región 'VPLHGTIL' (residuos 122-129). En estos casos tenemos que ningún epítipo que contenga los dos residuos o que esté dentro de la región tiene alta antigenicidad. El residuo 93 (L) no tiene ninguna afectación, por lo que directamente debería de haberse eliminado, pero por error humano en el procesamiento de datos se quedó en el diseño final. El residuo 163 (D) genera el único epítipo de alta afinidad hacia el alelo B*52:01 en el programa MHCflurry. Por su parte, la región 'VPLHGTIL' afecta el alelo B*51:01 en el programa MHCflurry. Si bien el programa Vaxijen v2.0 indica que los epítopos que contienen a este residuo y están en esta región no son antigénicos, se mantuvieron esperando que el procesamiento por MHCI sea suficiente para generar una respuesta inmune en este sector de la población. En una posible revisión al diseño, se debería de eliminar a la región 'VPLHGTIL' ya que podría estar afectando la antigenicidad total de la

molécula. Por su parte, el residuo D163 no debería de causar tantas afectaciones a la antigenicidad por ser un solo aminoácido, aunque manteniéndose apegados a los criterios de selección, también debería ser eliminado. Sumando a esto, si tomamos en cuenta la cobertura ilustrada en la Fig. 6, podríamos ver que estos dos alelos afectados en MHCflurry están cubiertos por al menos un epítipo sugerido por alguno de los otros dos programas que cumple los criterios.

En cuanto a mutaciones, se han reportado cuatro en las variantes de preocupación; (1) D3N, (2) Q19E, (3) A63T y (4) I82T. De estas, solamente A63T no se encuentra dentro de las regiones multiepitópicas. Como se menciona anteriormente, hay 3 secciones que se ven afectadas por estas mutaciones resaltadas en rojo en la Fig. 2. Las variantes Delta presentan la mutación I82T, afectando al alelo de MHCII DQA1*05:01/DQB1*02:01 en ambos programas, con la peculiaridad de solamente afectarlo en NetMHCIIpan si se toma en cuenta también la mutación D3N. Esta última mutación tiene la peculiaridad de estar presente en algunos de los linajes de Ómicron, pero no en todos. Esta mutación genera una de las secciones afectadas por mutaciones.

Al 22 de enero de 2023 se encuentran principalmente en circulación 5 variantes de Ómicron: (1) BA.2, (2) BA.5, (3) BA.2.75, (4) BQ.1 y (5) XBB; siendo BQ.1 la más abundante con un 50% de frecuencia (Nextstrain, 2023b). De éstas, BA.5 y BQ.1 presentan la mutación D3N y las restantes no. Si bien en el momento actual de la pandemia pareciera ser una mutación que prevalece por presentarse en el 67% de los casos, su comportamiento histórico señala que puede o no presentarse en las variantes de Ómicron. Por otro lado, es una sustitución conservativa, por lo que el procesamiento de los MHC podría no verse afectado (da Costa *et al.*, 2021). Esta mutación afecta a los alelos DQA1*03:01/DQB1*03:02 y DQA1*04:01/DQB1*04:02 y el DQA1*05:01/DQB1*02:01 combinado con la mutación I82T presente en Delta, además del alelo B*51:01. A pesar de esto, la decisión fue mantener este residuo porque es parte de epítopos de alta afinidad que cubren estos alelos afectados. Dado que la mutación es conservativa y no se encuentra en todas las variantes de Ómicron, es muy probable que se dé una respuesta inmune hacia los epítopos que contengan este aminoácido. Por lo que podríamos estar ante un 74.07% de cobertura al tomar en cuenta al menos un epítipo sugerido por alguno de los programas que cubra los criterios de selección, bastante mejor que el 66.7% que se encontró.

La mutación Q19E genera la última sección afectada por mutaciones de este diseño. El eliminar los epítomos que contienen a este residuo afecta al alelo B*13:01 de MCHI en los programas NetMHCpan y la IEDB y al B*18:01 en la IEDB. Esta mutación de nueva cuenta es conservativa, y aunque si se encuentra en todos los linajes de Ómicron, es posible que no afecte el procesamiento del MHC. Por lo que, estos epítomos deberían aportar a desencadenar una respuesta inmune en estos sectores de la población. Cabe destacar que, esta mutación es la que deja fuera al único epítomo que podría cubrir el alelo B*18:01 en alguno de los programas (Fig. 6). Tomando esto en cuenta, es posible que en lugar de tener cubiertos 76 alelos cubiertos por al menos un epítomo sugerido por alguno de los programas, se tendrían 77 de los 78 posibles por dejar esta sección en el diseño.

El diseño de la proteína N es el que finalmente tuvo menos zonas que no cumplieran los criterios (Fig. 3). El único residuo marcado de esta forma es el 63, afectado por la mutación D63G de las variantes Delta. Esta mutación es la misma que se presenta en la proteína S en el aminoácido 614, mutación que ha perdurado en todas las variantes que se han presentado (Korber *et al.*, 2020; Nextstrain, 2023b). Sin embargo, no ha sido el caso de esta mutación, ya que no se ha presentado en ninguna otra variante. Esta es una sustitución semiconservativa, que podría no reducir la afinidad hacia los MHC (da Costa *et al.*, 2021) Esta mutación afecta al alelo DQA1*05:01/DQB1*02:01 en ambos programas de MHCII. Por ello, se decidió dejar este residuo, con la idea de mantener una opción de cobertura para este grupo poblacional.

La proteína S contiene cuatro zonas que no cumplieron los criterios (Fig. 4). De éstas, tres son debido a la inclusión de residuos que no se encuentran en ningún posible epítomo antigénico. Estas zonas se encuentran en las posiciones: (1) Y473, (2) 1052-1053 (residuos 'FP') y (3) 1146-1149 (residuos 'DSFK'). En todos los casos no hay afectaciones a la cobertura final. Además, Y473 no se encuentra en epítomos lineales para células B antigénicos. Esto nos indica que se cometieron errores en estos casos. Una de las principales limitantes en este trabajo fue el manejo de datos. Si bien existen las herramientas para la predicción de epítomos, la complementariedad de estas se trabaja de maneras distintas en la comunidad científica, proponiendo diferentes metodologías (Bhattacharya *et al.*, 2020; Kumar *et al.*, 2020; Ayyagari *et al.*, 2020; Cun *et al.*, 2021; Sarkar *et al.*, 2022). En este caso, se diseñó una metodología nueva

que conglomerara diferentes herramientas inmunoinformáticas que se complementan unas a otras. Si bien para la toma de decisiones en el diseño metodológico y la selección de epítomos el factor humano es importante, este mismo limita la capacidad de procesamiento de datos. Sería conveniente el desarrollo de algoritmos computacionales que permitan el manejo integral de los datos recabados y que siga los criterios de selección propuestos para poder asegurar los mejores resultados posibles. Es posible que, debido al manejo manual de los datos, sitios que pudieran ser de interés se hayan pasado por alto. También, como es el caso en estas tres zonas, este manejo hace que los diseños no estén totalmente optimizados y se presenten sitios innecesarios en las regiones multiepitópicas. Considerando una revisión de este diseño, los residuos de estas zonas podrían ser eliminados sin consecuencias.

La zona faltante que no cumple los criterios corresponde al aminoácido T376. Esto se debe a que en todas las variantes Ómicron excepto BA.1, se presenta la mutación T376A. La sustitución treonina-alanina que se presenta es conservativa. Esta mutación no genera afectaciones a la cobertura poblacional. Sin embargo, su inclusión generaría 7 epítomos lineales B antigénicos, de los cuales 6 cumplirían con el porcentaje de residuos escondidos necesarios para su selección. Estos se incluyen en la Tabla 45 de la sección de anexos. Haciendo una revisión del diseño, podría proponerse el alterar la secuencia con esta mutación. Este cambio mantendría alta antigenicidad en estos epítomos y estarían más actualizados hacia el comportamiento de los linajes circulantes actuales. Además, al ser un sitio importante sólo para epítomos B y no para T CD8+ o T CD4+, no se vería comprometido el análisis de alta afinidad realizado basado en la secuencia de referencia modificada. Otra opción sería mantenerlo como está actualmente, esperando que los anticuerpos producidos hacia estos epítomos tengan efectos neutralizantes sobre el virus. La tercera y última opción sería eliminar este residuo. Sin embargo, dadas las afectaciones que se han generado sobre los epítomos B de este diseño debido a las nuevas mutaciones de las variantes emergentes, son mejores las opciones anteriores para tener una mayor cantidad de posibles epítomos.

8.2 Diseños finales

Regularmente, la selección de los enlazadores (también conocidos como espaciadores) en el diseño de una vacuna multiepitópica se realiza de dos maneras. Una opción puede ser tomar en cuenta la idoneidad del enlazador para un enlace determinado. Por ejemplo, utilizar el enlazador rígido EAAAK para separar un adyuvante del resto de la construcción con pocas afectaciones a su funcionalidad (Ayyagari *et al.*, 2020). Por otro lado, también se pueden seleccionar evaluando propiedades fisicoquímicas de las moléculas generadas como lo son la estructura secundaria, flexibilidad, índice antigénico, estabilidad, solubilidad, entre otras (He *et al.*, 2021). En este trabajo se propone un enfoque diferente para la selección de los enlazadores, basada en un único parámetro inmunológico: la antigenicidad. Este valor representa la capacidad que tienen las moléculas de unirse o ser reconocidas por anticuerpos o por receptores de linfocitos T o B. Esto puede promover tanto la respuesta celular como la humoral y la producción de células de memoria (Mittal *et al.*, 2020). Sin embargo, para aumentar las posibilidades de que esto suceda se debe aumentar la inmunogenicidad de la vacuna con la adición de enlazadores y adyuvantes. En este trabajo se realizaron permutaciones de las regiones multiepitópicas y a las más antigénicas se les añadieron enlazadores entre cada una de las regiones. Al final, se seleccionaron las combinaciones con el valor de antigenicidad más alto tuvieran o no enlazadores entre las regiones.

El único constructo que contuvo enlazadores entre las regiones multiepitópicas fue SARS2_N. Se añadieron 7 enlazadores KK para separar a 8 de las regiones y un enlazador GPGPG que divide estas primeras regiones de las 7 restantes (Tabla 25). KK es también conocido como el enlazador de lisinas y es un sitio donde la Catepsina β procesa péptidos antigénicos que serán presentados en el complejo principal de histocompatibilidad II (Ayyagari *et al.*, 2020). Por su parte, GPGPG fue diseñado como un espaciador universal e induce la respuesta de los linfocitos T cooperadores (Livingston *et al.*, 2002). Ambos enlazadores incrementan la inmunogenicidad tanto de la vacuna como de los epítomos que enlazan (Li *et al.*, 2015). La respuesta de las células T CD4+ es importante en este diseño ya que podrán dirigir las respuestas celulares y humorales que se desean obtener con los epítomos T CD8+ y B lineales. Por otro lado, es también la respuesta más duradera de las tres (Channappanavar *et al.*, 2014; Le Bert *et al.*, 2020). Si bien

estas características no fueron tomadas en cuenta para la selección, su naturaleza añade mayores posibilidades de desencadenar la respuesta inmune deseada.

Para incrementar la inmunogenicidad de la vacuna se añadió la secuencia adyuvante PADRE, también conocida como epítipo pan-DR, la cual se ha probado segura en humanos (Sarkar *et al.*, 2022). Esta molécula potencia la respuesta de células T CD4+ cooperadoras (Ayyagari *et al.*, 2020). Se probó el añadir la secuencia a ambos extremos de las secuencias candidatas con y sin enlaces buscando la mayor antigenicidad posible. En los cuatro diseños se obtuvo el mejor resultado uniéndola al extremo C-terminal. El adyuvante KK fue seleccionado para unir a PADRE en 3 de 4 diseños, siendo el de la proteína N (SARS2_N) el único constructo para el cual no se añadió ningún enlace. Como ya se ha mencionado, el adyuvante KK ayuda en el procesamiento de los epítopos para su presentación en el MHCII. Por esto, su adición junto al adyuvante PADRE podría representar una sinergia importante que mejore ampliamente la respuesta T CD4+. Si bien ya se ha buscado la respuesta de este tipo con los epítopos afines a MHCII de cada uno de los diseños, la adición de este adyuvante prácticamente la asegura e incrementa las posibilidades de desencadenar la respuesta inmunológica humoral y celular deseada.

La adición de una etiqueta de histidinas es una práctica común en el diseño de vacunas recombinantes porque facilita la purificación y la solubilidad de las proteínas (Ki y Pack, 2020). También, es útil para la detección y cuantificación de los antígenos al utilizar anticuerpos específicos anti-His en técnicas como Western Blot y ELISA (Ramos-Vega *et al.*, 2021). Para estos diseños, se añadió una etiqueta de 6 histidinas (6x-His) para tener la posibilidad de estos beneficios en la futura producción de los diseños vacunales. Tras la evaluación de antigenicidad, se seleccionaron las construcciones con la 6x-His en el extremo C-terminal, entre la secuencia multiepitópica y la señal de retención al retículo endoplásmico (SEKDEL). Se ha evaluado que la utilización de una etiqueta de 6 histidinas puede inducir anticuerpos neutralizantes de mejor manera que cadenas de histidinas más cortas (Chen *et al.*, 2022). Por otro lado, también se han obtenido resultados que indican que una vacuna recombinante basada en el dominio RBD del SARS-CoV-2 sin una cola de histidinas indujo una respuesta de anticuerpos neutralizantes más rápida y potente que un antígeno que la contiene (Lin *et al.*, 2022). A pesar de esto, no es claro

que vaya a haber una afectación en la respuesta hacia un diseño multiepitópico que busca no sólo la respuesta de anticuerpos, sino también una respuesta celular. Incluso existen vacunas recombinantes en estudios clínicos que incluyen colas de histidinas, como la vacuna EuCorVac-19 que se encuentra en fase III (Lovell *et al.*, 2022). Por estas razones, la adición de este péptido para ayudar en el proceso de producción y detección es deseable.

Flanqueando a todos los diseños tenemos al péptido señal 'MKMKVLFVATILVAWQCHA' en el extremo N-terminal y a 'SEKDEL' en el C-terminal. El primero es un péptido de la proteína de almacenamiento vegetativo de *Glycine max* (soya) (Romero-Maldonado *et al.*, 2016). Esta proteína se produce en RE (Klauer y Franceschi, 1997). El péptido señal de esta proteína se utiliza para direccionar la proteína hacia el RE. Por otro lado, la secuencia SEKDEL es utilizada para evitar la secreción de la proteína recombinante (Ramos-Vega *et al.*, 2021). SEKDEL se incluyó por el potencial de poder utilizar alguna plataforma de expresión como *Schizochytrium* sp. que permita la administración de la vacuna sin la necesidad de purificación (Ramos-Vega *et al.*, 2018).

8.3 Caracterización *in silico* de las proteínas quiméricas

Los tamaños de las moléculas variaron entre 104-326 residuos, mientras que sus pesos moleculares estuvieron entre 11.74-36.94 kDa (Tabla 27). Esto se mantiene debajo de los pesos y tamaños máximos producidos mediante el sistema pAlgevir, por lo que no debería ser un problema para su producción (Bañuelos-Hernández, 2017). Los puntos isoeléctricos fueron similares, variando de 9.39-9.71, lo que indica que todas las proteínas tienen una naturaleza básica. Los tiempos de vida media son calculados por ProtParam con la regla N-terminal, la cual indica que hay una relación directa entre el tiempo que le toma desaparecer a una proteína dentro de una célula y el residuo en su extremo N-terminal (Gasteiger *et al.*, 2005). Para todas las moléculas se obtuvieron los valores más altos posibles (mamífero >30 h; levaduras >20 h; *E. coli* >10 h), esto se debe a que su aminoácido en la posición 1 es la metionina, ya que comparten la secuencia del péptido señal en el extremo N-terminal.

Los diseños SARS2_E, SARS2_M y SARS2_S obtuvieron índices de inestabilidad menores a 40, lo que indica que son moléculas estables, mientras que el de SARS2_N fue mayor, prediciéndola

inestable (Gasteiger *et al.*, 2005). La hidropatía (índice GRAVY) nos indica si una molécula es hidrofílica o hidrofóbica dependiendo si los valores son negativos o positivos, respectivamente (Rath *et al.*, 2021). El índice alifático es el volumen que ocupan las cadenas laterales de los aminoácidos alifáticos (Ala, Val, Ile y Leu) en una proteína y se ha encontrado una correlación entre la termoestabilidad de las moléculas y altos índices alifáticos (Ikai, 1980). Eso es posiblemente debido a que las interacciones de los residuos alifáticos aumentan con la temperatura (Sun *et al.*, 2020). Además, estas dos propiedades muestran una correlación ya que los residuos alifáticos son hidrofóbicos, por lo que un mayor índice alifático lleva a un aumento en la hidropatía y viceversa (Sun *et al.*, 2020). SARS2_E y SARS2_M tienen un índice alifático alto y valores GRAVY positivos, por lo que se predice que podrían ser estables a un amplio rango de temperaturas y que son hidrofóbicos. La proteína SARS2_S es también termoestable, aunque en menor medida que las dos anteriores, pero es hidrofílica, lo que podría indicar una buena solubilidad en agua. Finalmente, SARS2_N se predijo como la más hidrofílica y con el índice alifático más bajo. A pesar de esto, se encuentra en el límite bajo de los valores reportados como de buena termoestabilidad (Karimi *et al.*, 2020; Nnolim *et al.*, 2020; Rath *et al.*, 2021). Si bien se obtuvo un valor que indica que es una molécula inestable, su índice alifático se contrapone un poco a esto. Aun así, durante la producción y caracterización de este antígeno se deberían de considerar estas propiedades.

GalaxyRefine es un servidor que mejora la estructura de modelos de proteínas reconstruyendo las cadenas laterales y evaluando posibles choques y rotámeros (Heo *et al.*, 2013). GalaxyRefine2 es una versión mejorada del programa anterior (Lee *et al.*, 2019). Ambos evalúan la desviación del valor cuadrático medio (RMSD) en el que valores menores indican una mejor estabilidad. También se calcula el valor MolProbity, que indica mejores resoluciones del modelo cuando se presentan valores bajos. El Clash Score indica sobreposiciones estéricas (Yang *et al.*, 2021). Los rotámeros son las diferentes posiciones que los residuos de una proteína toman basado en sus isómeros rotacionales (Haddad *et al.*, 2019). Se enlista el número de rotámeros deficientes, buscando que sean los más cercanos a cero.

Para la validación de estructuras terciarias predichas se han utilizado diferentes enfoques, como por ejemplo tomar en cuenta el valor Z y la calidad de las estructuras mediante el gráfico de

Ramachandran (Patra *et al.*, 2020). El servidor ERRAT evalúa el porcentaje de residuos que se encuentran dentro de un intervalo de confianza comparando la estructura con datos de cristalografías de alta resolución (Mahmud *et al.*, 2021). Se buscan valores de 95% para estructuras de alta resolución y 91% para las de menor resolución. Sin embargo, es generalmente aceptado un valor mayor a 50% (Messaoudi *et al.*, 2013). En este trabajo se tomaron en cuenta la calidad de las estructuras y el índice ERRAT para la seleccionar el modelo final.

El gráfico de Ramachandran es una representación de las posibles conformaciones que puede tener un péptido sin que haya un impedimento estérico entre sus residuos gracias a los ángulos Φ (Phi) y Ψ (Psy). Se trata a los átomos como si fueran esferas sólidas con un radio equivalente al radio de van der Waals. Si un residuo no colisiona estará dentro de las regiones favorables. Se pueden tomar valores del radio de van der Waals más cortos para indicar regiones adicionales y extendidas permitidas, pensando en que los átomos pueden estar más cerca entre ellos. Si aún con estos radios hay algún choque estérico, el residuo estará en las regiones no permitidas (Choudhuri, 2014). Lo deseable en un modelo es tener arriba del 85% de la estructura en las regiones favorables (Yang *et al.*, 2021).

El modelo seleccionado para SARS2_E se obtuvo porcentajes de 94.9%, 3%, 0% y 2% para las regiones favorables, adicionales, extendidas y no permitidas, respectivamente. Se obtuvo un factor de calidad ERRAT de 95.83% y el valor MolProbity disminuyó, lo que indica que es una estructura de alta resolución. La disminución en los casi total de los choques estéricos y total de los rotámeros deficientes indican una estructura estable. Todo esto en conjunto indica que este modelo es de alta calidad.

La proteína SARS2_M tuvo porcentajes de 69.9%, 25%, 4.1% y 1% para las regiones favorables, adicionales, extendidas y no permitidas, respectivamente. Esto está debajo del valor deseado. El valor MolProbity disminuyó y el índice ERRAT es de 69.31%, debajo de lo deseable, pero sobre el 50% aceptado. Hubo una disminución considerable en los rotámeros deficientes y choques estéricos, indicando una amplia mejoría tras el proceso de refinamiento. Los valores indican una calidad media del modelo. Esta baja calidad respecto a los otros diseños se puede atribuir a las

limitantes que conlleva realizar un modelaje *ab initio* como el realizado en este caso en el programa I-TASSER (Adam, 2021).

Para SARS2_N se obtuvieron porcentajes de 73.9%, 22.7%, 1.9% y 1.4% para las regiones favorables, adicionales, extendidas y no permitidas, respectivamente. El valor MolProbity disminuyó de 3.632 a 1.865 y el índice ERRAT es de 95.05, indicando una alta resolución del modelo refinado. Los rotámeros deficientes y choques estéricos disminuyeron casi en su totalidad, indicando una posible estructura estable. A pesar de que el porcentaje de residuos en regiones favorables está debajo de lo deseable, los demás parámetros indican un modelo de buena calidad.

El modelo SARS2_S no pudo ser refinado en el servidor GalaxyRefine2 por su tamaño (326aa), ya que el servidor sólo permite estructuras de hasta 300aa. El valor GDT-HA dado en el programa GalaxyRefine indica la similitud entre dos proteínas (Yang *et al.*, 2021). Para este modelo el valor GDT-HA fue de 0.8351, indicando alta similitud con el modelo sin refinar. En este servidor, un valor de RMSD entre 0-1.2 es aceptable para sugerir una estructura estable (Yang *et al.*, 2021). El valor obtenido fue de 0.692, dentro del rango deseado. Los rotámeros deficientes fueron disminuidos casi en su totalidad por el refinamiento, pero los choques estéricos aumentaron. Estos tres factores en conjunto se contraponen, por lo que no se puede asegurar una estructura estable. Se obtuvieron porcentajes de 72.3%, 21.8%, 2.8% y 3.1% para las regiones favorables, adicionales, extendidas y no permitidas, respectivamente. El porcentaje de residuos en regiones favorables está debajo del 85% deseable. El valor ERRAT de 66.8% indica una resolución baja, pero mejor que la del modelo original ya que el valor de MolProbity disminuyó. En conjunto estos parámetros indican un modelo de calidad media. Esto puede ser atribuible a las limitantes que el servidor GalaxyRefine tiene frente a su versión más actualizada.

9. CONCLUSIONES

Los diseños multiepitópicos SARS2_M, SARS2_N y SARS2_S de la vacuna CIB-GIVac tuvieron una amplia cobertura poblacional y una inmunogenicidad humoral y celular predicha que podría mantenerse vigente ante las nuevas variantes del SARS CoV 2.

SARS2_S es el diseño que se predice más antigénico e inmunogénico. Este diseño tiene el potencial de cubrir a prácticamente toda la población mundial con epítomos para células T CD4+ y T CD8+ antigénicos y de alta afinidad a los MHC. También, contiene una gran variedad de epítomos lineales de células B en regiones de interés como el dominio RBD y el HR2 que tienen el potencial de inducir la producción de anticuerpos neutralizantes por su antigenicidad. Además, estos últimos evitan los residuos glicosilados de esta proteína, lo que incrementa las posibilidades de que los anticuerpos generados puedan unirse a la proteína S del SARS-CoV-2 sin impedimento. Esto indica un alto potencial de generar una respuesta humoral y celular ante el patógeno.

El diseño SARS2_N se predice como la segunda construcción más antigénica. Tiene el potencial de cubrir a casi toda la población mundial con epítomos para las células T CD8+ y a más del 85% con epítomos para T CD4+.

Tomando en cuenta los resultados de cobertura poblacional por proteína, SARS2_M tiene el potencial de cubrir a casi toda la población mundial con epítomos T CD8+, aunque su cobertura con epítomos T CD4+ es de las más bajas.

SARS2_E es el diseño con las coberturas poblacionales más bajas, pero son mayores al 60% en los dos casos. Esto es un indicador de que la metodología propuesta en este trabajo tiene la capacidad de aprovechar al máximo las posibles regiones antigénicas de las proteínas analizadas.

Todos los diseños han mantenido intactos sus porcentajes de cobertura poblacional ante las nuevas variantes Ómicron (BA.2.12.2, BA.2.75, BQ.1 y XBB).

Los diseños multiepitópicos basados en las proteínas estructurales del SARS-CoV-2 generados en esta investigación predicen una buena respuesta humoral y celular en la población mundial y tienen potencial de mantenerse vigentes ante el surgimiento de nuevas variantes del virus.

Todas estas construcciones están diseñadas para poder insertarse en los vectores de expresión pAlgevir y pBI121. Con esta tecnología estos diseños podrían producirse en plantas o microalgas tanto en cultivos individuales como todos en conjunto. Para esto se han diseñado los primers específicos para PCR punto final y tiempo real. También, se han diseñado péptidos sintéticos que servirán para la producción de sueros hiperinmunes a usarse en los ensayos de expresión y cuantificación de proteínas.

Los diseños generados están pensados para funcionar como vacunas individuales o mezclas conteniendo a todos o algunos de estos. El potencial inmunogénico de cada uno y las mezclas debería ser analizado mediante ensayos inmunológicos *in vivo* con parámetros de dosis, número de dosis y tiempo generales. Además, sería deseable medir la prevalencia de las respuestas humoral y celular tiempo después de los ensayos. Posteriormente, deberían hacerse ensayos con diferentes parámetros para obtener los mejores esquemas de inmunización e identificar una o varias vacunas candidatas. Idealmente, después estas vacunas deberían ser probadas en retos infecciosos para verificar la protección que proveen.

Ya han sido estandarizados los programas de PCR punto final para la amplificación de todos los genes. Se han confirmado por PCR y secuenciación diferentes cultivos de células de *E. coli* TOP 10 transformadas con los vectores pBI121 y pAlgevir conteniendo los genes de 3 de las proteínas diseñadas (SARS2_S, SARS2_N y SARS2_M) (Ver anexos).

10. LITERATURA CITADA

- Achaiah, N.C. y Subbarajasetty, S.B. (2020). R0 and Re of COVID-19: can we predict when the pandemic outbreak will be contained? *Indian Journal of Critical Care Medicine*, 24(11): 1125-1127. doi: 10.5005/jp-journals-10071-23649
- Achiron, A., Gurevich, M., Falb, R., Dreyer-Alster, S., Sonis, P. y Mandel, M. (2021). SARS-CoV-2 antibody dynamics and B-cell memory response over time in COVID-19 convalescent subjects. *Clinical Microbiology and Infection*, 27(9): 1349.e1–1349.e6. doi: 10.1016/j.cmi.2021.05.008
- Adam, K.M. (2021). Immunoinformatics approach for multi-epitope vaccine design against structural proteins and ORF1a polyprotein of severe acute respiratory syndrome coronavirus-2 (SARS-CoV-2). *Tropical Diseases, Travel Medicine and Vaccines*, 7(1): 22. doi: 10.1186/s40794-021-00147-1
- Ahmed, S.F., Quadeer, A.A. y McKay, M.R. (2020). Preliminary identification of potential vaccine targets for the COVID-19 coronavirus (SARS-CoV-2) based on SARS-CoV immunological studies. *Viruses*, 12(3): 254. doi: 10.3390/v12030254
- Ayyagari, V.S., Venkateswarulu, T.C., Karlapudi, AP. y Srirama, K. (2020). Design of a multi-epitope-based vaccine targeting M-protein of SARS-CoV2: an immunoinformatics approach. *Journal of Biomolecular Structure and Dynamics*, 40(7): 1-15. doi: 10.1080/07391102.2020.1850357
- Bañuelos-Hernández, B. (2017). Algevir: an expression system for microalgae based on viral vectors. *Frontiers in Microbiology*, 8. doi: 10.3389/fmicb.2017.01100
- Bhattacharya, M., Sharma, A.R., Patra, P., Ghosh, P., Sharma, G., Patra, B.C., Lee, S. y Chakraborty, C. (2020). Development of epitope-based peptide vaccine against novel coronavirus 2019 (SARS-COV-2): immunoinformatics approach. *Journal of Medical Virology*, 92(6): 618-631. doi: 10.1002/jmv.25736
- Bianchi, M., Benvenuto, D., Giovanetti, M., Angeletti, S., Ciccozzi, M. y Pascarella, S. (2020). Sars-CoV-2 envelope and membrane proteins: structural differences linked to virus characteristics? *BioMed Research International*, 2020: 1-6. doi: 10.1155/2020/4389089
- Brun, J., Vasiljevic, S., Gangadharan, B., Hensen, M., Chandran, A.V., Hill, ML., Kiappes, J.L., Dwek, R.A., Alonzi, D.S., Struwe, W.B. y Zitzmann, N. (2021). Assessing antigen structural integrity through glycosylation analysis of the SARS-CoV-2 viral spike. *ACS Central Science*, 7(4): 586-593. doi: 10.1021/acscentsci.1c00058
- Cai, Y., Zhang, J., Xiao, T., Peng, H., Sterling, S.M., Walsh, R.M., Rawson, S., Rits-Volloch, S. y Chen, B. (2020). Distinct conformational states of SARS-CoV-2 spike protein. *Science*, 369(6511): 1586-1592. doi: 10.1126/science.abd4251
- CECMED. (2020). Autorización inicio ensayo clínico: estudio fase I, Abierto, secuencial y adaptativo, para evaluar la seguridad, reactogenicidad, y explorar la inmunogenicidad, del candidato vacunal profiláctico FINLAY-FR-2 anti SARS-CoV-2. Disponible en: <https://www.cecmecd.cu/covid-19/aprobaciones/soberana-02>
- Center for Disease Control and Prevention. (2023). SARS-CoV-2 variant classifications and definitions. Disponible en: <https://www.cdc.gov/coronavirus/2019-ncov/variants/variant-classifications.html>

- Chang, C., Chen, C.-M.M., Chiang, M., Hsu, Y. y Huang, T. (2013). Transient oligomerization of the SARS-CoV N protein – implication for virus ribonucleoprotein packaging. *PLoS ONE*, 8(5): e65045. doi: 10.1371/journal.pone.0065045
- Channappanavar, R., Fett, C., Zhao, J., Meyerholz, D.K. y Perlman, S. (2014). Virus-specific memory CD8 T cells provide substantial protection from lethal severe acute respiratory syndrome coronavirus infection. *Journal of Virology*, 88(19): 11034-11044. doi: 10.1128/JVI.01505-14
- Chavda, V.P. y Apostolopoulos, V. (2022). Global impact of delta plus variant and vaccination. *Expert Review of Vaccines*, 21(5): 597-600. doi: 10.1080/14760584.2022.2044800
- Chen, L., Qi, X., Liang, D., Li, G., Peng, X., Li, X., Ke, B., Zheng, H., Liu, Z., Ke, C., Liao, G., Liu, L. y Feng, Q. (2022). Human Fc-conjugated receptor binding domain-based recombinant subunit vaccines with short linker induce potent neutralizing antibodies against multiple SARS-CoV-2 variants. *Vaccines*, 10(9): 1502. doi: 10.3390/vaccines10091502
- Chen, Y., Liu, Q. y Guo, D. (2020). Emerging coronaviruses: genome structure, replication, and pathogenesis. *Journal of Medical Virology*, 92(4): 418-423. doi: 10.1002/jmv.25681
- Choudhuri, S. (2014). Additional bioinformatic analyses involving protein sequences. En: *Bioinformatics for beginners: genes, genomes, molecular evolution, databases and analytical tools*, Academic Press, Londres, pp 183-207. doi: 10.1016/B978-0-12-410471-6.00008-6
- ClinicalTrials.gov. (2021). Study of the tolerability, safety, immunogenicity and preventive efficacy of the EpiVacCorona vaccine for the prevention of COVID-19. Disponible en: <https://www.clinicaltrials.gov/ct2/show/NCT04780035?term=vaccine%2C+protein&cond=Covid19&phase=2&draw=2>
- ClinicalTrials.gov. (2020a). A phase III clinical trial to determine the safety and efficacy of ZF2001 for prevention of COVID-19. Disponible en: <https://clinicaltrials.gov/ct2/show/record/NCT04646590?view=record>
- ClinicalTrials.gov. (2020b). A study to evaluate the efficacy, immune response, and safety of a COVID-19 vaccine in adults ≥ 18 years with a pediatric expansion in adolescents (12 to <18 years) at risk for SARS-CoV-2. Disponible en: <https://www.clinicaltrials.gov/ct2/show/NCT04611802?term=vaccine%2C+protein&cond=Covid19&phase=2&draw=2&rank=3>
- ClinicalTrials.gov. (2020c). A controlled phase 2/3 study of adjuvanted recombinant SARS-CoV-2 trimeric S-protein vaccine (SCB-2019) for the prevention of COVID-19 (SCB-2019). Disponible en: <https://www.clinicaltrials.gov/ct2/show/NCT04672395?term=vaccine%2C+protein&cond=Covid19&phase=2&draw=2&rank=1>
- Coronaviridae Study Group of the International Committee on Taxonomy of Viruses. (2020). The species *Severe acute respiratory syndrome-related coronavirus*: classifying 2019-nCoV and naming it SARS-CoV-2. *Nature Microbiology*, 5(4): 536-544. doi: 10.1038/s41564-020-0695-z

- Coutard, B., Valle, C., de Lamballerie, X., Canard, B., Seidah, N.G. y Decroly, E. (2020). The spike glycoprotein of the new coronavirus 2019-nCoV contains a furin-like cleavage site absent in CoV of the same clade. *Antiviral Research*, 176: 104742. doi: 10.1016/j.antiviral.2020.104742
- Cun, Y., Li, C., Shi, L., Sun, M., Dai, S., Sun, L., Shi, L. y Yao, Y. (2021). COVID-19 coronavirus vaccine T cell epitope prediction analysis based on distributions of HLA class I loci (HLA-A, -B, -C) across global populations. *Human Vaccines & Immunotherapeutics*, 17(4): 1097-1108. doi: 10.1080/21645515.2020.1823777
- da Costa, A.S., Fernandes, T.V.A., Bello, M.L. y de Souza, T.L.F. (2021). Evaluation of potential MHC-I allele-specific epitopes in Zika virus proteins and the effects of mutations on peptide-MHC-I interaction studied using *in silico* approaches. *Computational Biology and Chemistry*, 92: 107459. doi: 10.1016/j.compbiolchem.2021.107459
- Dagur, H.S. (2020). Epitope-Based Vaccine Design against Novel Coronavirus SARS-CoV-2 Envelope Protein. *Eurasian Journal of Medicine and Oncology*, 4(3): 201-208. doi: 10.14744/ejmo.2020.01978
- Davidson, A.D., Williamson, M.K., Lewis, S., Shoemark, D., Carroll, M.W., Heesom, K.J., Zambon, M., Ellis, J., Lewis, P.A., Hiscox, J.A. y Matthews, D.A. (2020). Characterisation of the transcriptome and proteome of SARS-CoV-2 reveals a cell passage induced in-frame deletion of the furin-like cleavage site from the spike glycoprotein. *Genome Medicine*, 12(1): 68. doi: 10.1186/s13073-020-00763-0
- Díaz-Ortega, J.L. (2022). Guía técnica para la aplicación de la vacuna ABDALA/CIGB-66 contra el virus SARS-CoV-2. Disponible en: https://vacunacovid.gob.mx/wp-content/uploads/2022/12/221213_GuiaTecnicaAbdala.pdf
- European Centre for Disease Prevention and Control. (2023). SARS-CoV-2 variants of concern as of 26 January 2023. Disponible en: <https://www.ecdc.europa.eu/en/covid-19/variants-concern>
- Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S., Wilkins, M.R., Appel, R.D. y Bairoch, A. (2005). Protein identification and analysis tools on the ExPASy server. En: *Walker JM (ed) The Proteomics Protocols Handbook*. Humana Press, Totowa, NJ, pp 571-607. doi: 10.1385/1-59259-890-0:571
- GISAID. (2020). Official hCoV-19 reference sequence. Disponible en: <https://www.gisaid.org/epiflu-applications/hcov-19-reference-sequence/>
- Gong, Y., Qin, S., Dai, L. y Tian, Z. (2021). The glycosylation in SARS-CoV-2 and its receptor ACE2. *Signal Transduction and Targeted Therapy*, 6(1): 396. doi: 10.1038/s41392-021-00809-8
- Grant, O.C., Montgomery, D., Ito, K. y Woods, R.J. (2020). Analysis of the SARS-CoV-2 spike protein glycan shield reveals implications for immune recognition. *Scientific Reports*, 10(1): 14991. doi: 10.1038/s41598-020-71748-7
- Greenbaum, J., Sidney, J., Chung, J., Brander, C., Peters, B. y Sette, A. (2011). Functional classification of class II human leukocyte antigen (HLA) molecules reveals seven different supertypes and a surprising degree of repertoire sharing across supertypes. *Immunogenetics*, 63(6): 325-335. doi: 10.1007/s00251-011-0513-0

- Haddad, Y., Adam, V. y Heger, Z. (2019). Rotamer dynamics: analysis of rotamers in molecular dynamics simulations of proteins. *Biophysical Journal*, 116(11): 2062-2072. doi: 10.1016/j.bpj.2019.04.017
- Hasöksüz, M., Kiliç, S. y Saraç, F. (2020). Coronaviruses and SARS-COV-2. *Turkish Journal of Medical Sciences*, 50(SI-1): 549-556. doi: 10.3906/sag-2004-127
- He, J., Huang, F., Zhang, J., Chen, Q., Zheng, Z., Zhou, Q., Chen, D., Li, J. y Chen, J. (2021). Vaccine design based on 16 epitopes of SARS-CoV-2 spike protein. *Journal of Medical Virology*, 93(4): 2115-2131. doi: 10.1002/jmv.26596
- Heo, L., Park, H. y Seok, C. (2013). GalaxyRefine: protein structure refinement driven by side-chain repacking. *Nucleic Acids Research*, 41(W1): W384–W388. doi: 10.1093/nar/gkt458
- Huang, Y., Yang, C., Xu, X., Xu, W. y Liu, S. (2020). Structural and functional properties of SARS-CoV-2 spike protein: potential antiviral drug development for COVID-19. *Acta Pharmacologica Sinica*, 41(9): 1141-1149. doi: 10.1038/s41401-020-0485-4
- Ikai, A. (1980). Thermostability and aliphatic index of globular proteins. *The Journal of Biochemistry*, 88(6): 1895-1898. PMID: 7462208.
- International Committee on Taxonomy of Viruses. (2023). Current ICTV taxonomy release. Disponible en: <https://talk.ictvonline.org/taxonomy/>
- Jackson, C.B., Farzan, M., Chen, B. y Choe, H. (2022). Mechanisms of SARS-CoV-2 entry into cells. *Nature Reviews Molecular Cell Biology*, 23(1): 3-20. doi: 10.1038/s41580-021-00418-x
- Johnson, B.A., Xie, X., Bailey, A.L., Kalveram, B., Lokugamage, K.G., Muruato, A., Zou, J., Zhang, X., Juelich, T., Smith, J.K., Zhang, L., Bopp, N., Schindewolf, C., Vu, M., Vanderheiden, A., Winkler, E.S., Swetnam, D., Plante, J.A., Aguilar, P., Plante, K.S., Popov, V., Lee, B., Weaver, S.C., Suthar, M.S., Routh, A.L., Ren, P., Ku, Z., An, Z., Debbink, K., Diamond, M.S., Shi, P.-Y., Freiberg, A.N. y Menachery, V.D. (2021). Loss of furin cleavage site attenuates SARS-CoV-2 pathogenesis. *Nature*, 591(7849): 293-299. doi: 10.1038/s41586-021-03237-4
- Jungreis, I., Sealfon, R. y Kellis, M. (2020). SARS-CoV-2 gene content and COVID-19 mutation impact by comparing 44 *Sarbecovirus* genomes. *Nature Communications*. 12(1): 2642. doi: 10.1101/2020.06.02.130955
- Karimi, M., Seyyed Tabaei, S.J., Ranjbar, M.M., Fathi, F., Jalili, A., Zamini, G., Javadi Mamaghani, A., Nazari, J., Roshani, D., Bagherani, N. y Khademerfan, M. (2020). Construction of a synthetic gene encoding the multi-epitope of *Toxoplasma gondii* and demonstration of the relevant recombinant protein production: a vaccine candidate. *Galen Medical Journal*, 9: e1708. doi: 10.31661/gmj.v9i0.1708
- Khailany, RA., Safdar, M. y Ozaslan, M. (2020). Genomic characterization of a novel SARS-CoV-2. *Gene Reports*, 19: 100682. doi: 10.1016/j.genrep.2020.100682
- Ki, M.-R. y Pack, S.P. (2020). Fusion tags to enhance heterologous protein expression. *Applied Microbiology and Biotechnology*, 104(6): 2411-2425. doi: 10.1007/s00253-020-10402-8
- Kim, J.-M., Chung, Y.-S., Jo, H.J., Lee, N.-J., Kim, M.S., Woo, S.H., Park, S., Kim, J.W., Kim, H.M. y Han, M.-G. (2020). Identification of coronavirus isolated from a patient in

- Korea with COVID-19. *Osong Public Health and Research Perspectives*, 11(1): 3-7. doi: 10.24171/j.phrp.2020.11.1.02
- Klauer, S.F. y Franceschi, V.R. (1997). Mechanism of transport of vegetative storage proteins to the vacuole of the paraveinal mesophyll of soybean leaf. *Protoplasma*, 200(3-4): 174-185. doi: 10.1007/BF01283293
- Konno, Y., Kimura, I., Uriu, K., Fukushi, M., Irie, T., Koyanagi, Y., Sauter, D., Gifford, R.J., Nakagawa, S. y Sato, K. (2020). SARS-CoV-2 ORF3b Is a potent interferon antagonist whose activity is increased by a naturally occurring elongation variant. *Cell Reports*, 32(12): 108185. doi: 10.1016/j.celrep.2020.108185
- Korber, B., Fischer, W.M., Gnanakaran, S., Yoon, H., Theiler, J., Abfalterer, W., Hengartner, N., Giorgi, E.E., Bhattacharya, T., Foley, B., Hastie, K.M., Parker, M.D., Partridge, D.G., Evans, C.M., Freeman, T.M., de Silva, T.I., McDanal, C., Perez, L.G., Tang, H., Moon-Walker, A., Whelan, S.P., LaBranche, C.C., Saphire, E.O., Montefiori, D.C., Angyal, A., Brown, R.L., Carrilero, L., Green, L.R., Groves, D.C., Johnson, K.J., Keeley, A.J., Lindsey, B.B., Parsons, P.J., Raza, M., Rowland-Jones, S., Smith, N., Tucker, R.M., Wang, D. y Wyles, M.D. (2020). Tracking changes in SARS-CoV-2 spike: evidence that D614G increases infectivity of the COVID-19 virus. *Cell*, 182(4): 812–827.e19. doi: 10.1016/j.cell.2020.06.043
- Kumar, A., Kumar, P., Saumya, K.U., Kapuganti, S.K., Bhardwaj, T. y Giri, R. (2020). Exploring the SARS-CoV-2 structural proteins for multi-epitope vaccine development: an in-silico approach. *Expert Review of Vaccines*, 19(9): 887-898. doi: 10.1080/14760584.2020.1813576
- Le Bert, N., Tan, A.T., Kunasegaran, K., Tham, C.Y.L., Hafezi, M., Chia, A., Chng, M.H.Y., Lin, M., Tan, N., Linster, M., Chia, W.N., Chen, M.I.-C., Wang, L.-F., Ooi, E.E., Kalimuddin, S., Tambyah, P.A., Low, J.G.-H., Tan, Y.-J. y Bertoletti, A. (2020). SARS-CoV-2-specific T cell immunity in cases of COVID-19 and SARS, and uninfected controls. *Nature*, 584(7821): 457-462. doi: 10.1038/s41586-020-2550-z
- Lei, Y., Zhao, F., Shao, J., Li, Y., Li, S., Chang, H. y Zhang, Y. (2019). Application of built-in adjuvants for epitope-based vaccines. *PeerJ Life and Environment*, 6: e6185. doi: 10.7717/peerj.6185
- Li, X., Guo, L., Kong, M., Su, X., Yang, D., Zou, M., Liu, Y. y Lu, L. (2015). Design and evaluation of a multi-epitope peptide of human metapneumovirus. *Intervirology*, 58(6): 403-412. doi: 10.1159/000445059
- Lin, T.-W., Huang, P.-H., Liao, B.-H., Chao, T.-L., Tsai, Y.-M., Chang, S.-C., Chang, S.-Y. y Chen, H.-W. (2022). Tag-free SARS-CoV-2 receptor binding domain (RBD), but not C-terminal tagged SARS-CoV-2 RBD, induces a rapid and potent neutralizing antibody response. *Vaccines*, 10(11): 1839. doi: 10.3390/vaccines10111839
- Livingston, B., Crimi, C., Newman, M., Higashimoto, Y., Appella, E., Sidney, J. y Sette, A. (2002). A rational strategy to design multiepitope immunogens based on multiple Th lymphocyte epitopes. *The Journal of Immunology*, 168(11): 5499-5506. doi: 10.4049/jimmunol.168.11.5499
- Lovell, J.F., Baik, Y.O., Choi, S.K., Lee, C., Lee, J.-Y., Miura, K., Huang, W.-C., Park, Y.-S., Woo, S.-J., Seo, S.H., Kim, J.-O., Song, M., Kim, C.-J., Choi, J.-K., Kim, J., Choo, E.J. y Choi, J.-H. (2022). Interim analysis from a phase 2 randomized trial of EuCorVac-19: a

- recombinant protein SARS-CoV-2 RBD nanoliposome vaccine. *BMC Medicine*, 20(1): 462. doi: 10.1186/s12916-022-02661-1
- Mahmud, S., Rafi, M.O., Paul, G.K., Promi, M.M., Shimu, M.S.S., Biswas, S., Emran, T. B., Dhama, K., Alyami, S.A., Moni, M.A. y Saleh, M.A. (2021). Designing a multi-epitope vaccine candidate to combat MERS-CoV by employing an immunoinformatics approach. *Scientific Reports*, 11(1): 15431. doi: 10.1038/s41598-021-92176-1
- Mandala, V.S., McKay, M.J., Shcherbakov, A.A., Dregni, A.J., Kolocouris, A. y Hong, M. (2020). Structure and drug binding of the SARS-CoV-2 envelope protein transmembrane domain in lipid bilayers. *Nature Structural & Molecular Biology*, 27(12): 1202-1208. doi: 10.1038/s41594-020-00536-8
- Márquez-Escobar, V.A., Bañuelos-Hernández, B. y Rosales-Mendoza, S. (2018). Expression of a Zika virus antigen in microalgae: towards mucosal vaccine development. *Journal of Biotechnology*, 282: 86-91. doi: 10.1016/j.jbiotec.2018.07.025
- Más-Bermejo, P.I., Dickinson-Meneses, F.O., Almenares-Rodríguez, K., Sánchez-Valdés, L., Guinovart-Díaz, R., Vidal-Ledo, M., Galbán-García, E., Olivera-Nodarse, Y., Morgado-Vega, I., Dueñas-Carrera, S., Pujol, M., Hernández-Bernal, F., Limonta-Fernández, M., Guillén-Nieto, G., Muzio-González, V.L. y Ayala-Ávila, M. (2022). Cuban Abdala vaccine: effectiveness in preventing severe disease and death from COVID-19 in Havana, Cuba; A cohort study. *The Lancet Regional Health – Americas*, 16. 100366. doi: 10.1016/j.lana.2022.100366
- Mercurio, I., Tragni, V., Busto, F., de Grassi, A. y Pierri, C.L. (2021). Protein structure analysis of the interactions between SARS-CoV-2 spike protein and the human ACE2 receptor: from conformational changes to novel neutralizing antibodies. *Cellular and Molecular Life Sciences*, 78(4): 1501-1522. doi: 10.1007/s00018-020-03580-1
- Messaoudi, A., Belguith, H. y Ben Hamida, J. (2013). Homology modeling and virtual screening approaches to identify potent inhibitors of VEB-1 β -lactamase. *Theoretical Biology and Medical Modelling*, 10(1): 22. doi: 10.1186/1742-4682-10-22
- Mittal, A., Sasidharan, S., Raj, S., Balaji, S.N. y Saudagar, P. (2020). Exploring the Zika genome to design a potential multiepitope vaccine using an immunoinformatics approach. *International Journal of Peptide Research and Therapeutics*, 26(4): 2231-2240. doi: 10.1007/s10989-020-10020-y
- Nakagawa, S. y Miyazawa, T. (2020). Genome evolution of SARS-CoV-2 and its virological characteristics. *Inflammation and Regeneration*, 40(1): 17. doi: 10.1186/s41232-020-00126-7
- Naqvi, A.A.T., Fatima, K., Mohammad, T., Fatima, U., Singh, I.K., Singh, A., Atif, S.M., Hariprasad, G., Hasan, G.M. y Hassan, M.I. (2020). Insights into SARS-CoV-2 genome, structure, evolution, pathogenesis and therapies: Structural genomics approach. *Biochimica et Biophysica Acta (BBA) - Molecular Basis of Disease*, 1866(10): 165878. doi: 10.1016/j.bbadis.2020.165878
- National Center for Biotechnology Information. (2020). Severe acute respiratory syndrome coronavirus 2 isolate Wuhan-Hu-1, complete genome. Disponible en: <https://www.ncbi.nlm.nih.gov/nucleotide/1798174254>
- Nemet, I., Kliker, L., Lustig, Y., Zuckerman, N., Erster, O., Cohen, C., Kreiss, Y., Alroy-Preis, S., Regev-Yochay, G., Mendelson, E. y Mandelboim, M. (2022). Third BNT162b2

- vaccination neutralization of SARS-CoV-2 Omicron infection. *New England Journal of Medicine*, 386(5): 492-494. doi: 10.1056/NEJMc2119358
- Nextstrain. (2023a). Genomic epidemiology of novel coronavirus - Global subsampling. Disponible en: <https://nextstrain.org/ncov/global?l=clock>
- Nextstrain. (2023b). CoVariants. Disponible en: <https://covariants.org/>
- Nnolim, N.E., Mpaka, L., Okoh, A.I. y Nwodo, U.U. (2020). Biochemical and molecular characterization of a thermostable alkaline metallo-keratinase from *Bacillus* sp. Nnolim-K1. *Microorganisms*, 8(9): 1304. doi: 10.3390/microorganisms8091304
- Patra, P., Ghosh, P., Patra, B.C. y Bhattacharya, M. (2020). Biocomputational analysis and *in silico* characterization of an angiogenic protein (RNase5) in Zebrafish (*Danio rerio*). *International Journal of Peptide Research and Therapeutics*, 26(4): 1687-1697. doi: 10.1007/s10989-019-09978-1
- Rahman, M.S., Islam, M.R., Alam, A.S.M.R.U., Islam, I., Hoque, M.N., Akter, S., Rahaman, M.M., Sultana, M. y Hossain, M.A. (2021). Evolutionary dynamics of SARS-CoV-2 nucleocapsid protein and its consequences. *Journal of Medical Virology*, 93(4): 2177-2195. doi: 10.1002/jmv.26626
- Ramos-Vega, A., Monreal-Escalante, E., Dumonteil, E., Bañuelos-Hernández, B. y Angulo, C. (2021). Plant-made vaccines against parasites: bioinspired perspectives to fight against Chagas disease. *Expert Review of Vaccines*, 20(11): 1373-1388. doi: 10.1080/14760584.2021.1893170
- Ramos-Vega, A., Rosales-Mendoza, S., Bañuelos-Hernández, B. y Angulo, C. (2018). Prospects on the use of *Schizochytrium* sp. to develop oral vaccines. *Frontiers in Microbiology*, 9: 2506. doi: 10.3389/fmicb.2018.02506
- Rath, S., Perikala, V., Jena, A.B. y Dandapat, J. (2021). Factors regulating dynamics of angiotensin-converting enzyme-2 (ACE2), the gateway of SARS-CoV-2: epigenetic modifications and therapeutic interventions by epidrugs. *Biomedicine & Pharmacotherapy*, 143: 112095. doi: 10.1016/j.biopha.2021.112095
- Registro Público Cubano de Ensayos Clínicos. (2021). SOBERANA 02 - Fase III. Disponible en: <https://rpcec.sld.cu/en/ensayos/RPCEC00000354-Sp>
- Romero-Maldonado, A., Monreal-Escalante, E. y Rosales-Mendoza, S. (2016). Expression in plants of two new antigens with implications in Alzheimer's disease immunotherapy. *Plant Cell, Tissue and Organ Culture (PCTOC)*, 126(2): 361-370. doi: 10.1007/s11240-016-0990-9
- Rosenblum, H.G., Wallace, M., Godfrey, M., Roper, L.E., Hall, E., Fleming-Dutra, K.E., Link-Gelles, R., Pilishvili, T., Williams, J., Moulia, D.L., Brooks, O., Talbot, H.K., Lee, G.M., Bell, B.P., Daley, M.F., Meyer, S., Oliver, S.E. y Twentyman, E. (2022). Interim recommendations from the advisory committee on immunization practices for the use of bivalent booster doses of COVID-19 vaccines — United States, October 2022. *MMWR Morbidity and Mortality Weekly Report*, 71(45): 1436-1441. doi: 10.15585/mmwr.mm7145a2
- Samrat, S.K., Tharappel, A.M., Li, Z. y Li, H. (2020). Prospect of SARS-CoV-2 spike protein: potential role in vaccine and therapeutic development. *Virus Research*, 288: 198141. doi: 10.1016/j.virusres.2020.198141

- Sarkar, B., Ullah, M.A., Araf, Y., Islam, N.N. y Zohora, U.S. (2022). Immunoinformatics-guided designing and *in silico* analysis of epitope-based polyvalent vaccines against multiple strains of human coronavirus (HCoV). *Expert Review of Vaccines*, 21(12): 1851-1871. doi: 10.1080/14760584.2021.1874925
- Schoeman, D. y Fielding, B.C. (2019). Coronavirus envelope protein: current knowledge. *Virology Journal*, 16(1): 69. doi: 10.1186/s12985-019-1182-0
- Secretaría de Salud México. (2023). Covid-19 México. Disponible en: <https://datos.covid-19.conacyt.mx/>
- Shajahan, A., Supekar, N.T., Gleinich, A.S. y Azadi, P. (2020). Deducing the N- and O-glycosylation profile of the spike protein of novel coronavirus SARS-CoV-2. *Glycobiology*, 30(12): 981-988. doi: 10.1093/glycob/cwaa042
- Shang, J., Wan, Y., Luo, C., Ye, G., Geng, Q., Auerbach, A. y Li, F. (2020). Cell entry mechanisms of SARS-CoV-2. *Proceedings of the National Academy of Sciences*, 117(21): 11727. doi: 10.1073/pnas.2003138117
- Sun, S., Ao, C., Wang, D. y Dong, B. (2020). The frequencies of oppositely charged, uncharged polar, and β -branched amino acids determine proteins' thermostability. *IEEE Access*, 8: 66839-66845. doi: 10.1109/ACCESS.2020.2985737
- Swanson, K. (2022). Pfizer/BioNTech COVID-19 Omicron-modified bivalent vaccine. 202211(101007)Disponible en: <https://stacks.cdc.gov/view/cdc/120826>
- Thomas, S. (2020). The structure of the membrane protein of SARS-CoV-2 resembles the sugar transporter SemiSWEET. *Pathogens and Immunity*, 5(1): 342. doi: 10.20411/pai.v5i1.377
- Walls, A.C., Park, Y.-J., Tortorici, M.A., Wall, A., McGuire, A.T. y Veasley, D. (2020). Structure, function, and antigenicity of the SARS-CoV-2 spike glycoprotein. *Cell*, 181(2): 281–292.e6. doi: 10.1016/j.cell.2020.02.058
- Wan, Y., Shang, J., Graham, R., Baric, R.S. y Li, F. (2020). Receptor recognition by the novel coronavirus from Wuhan: an analysis based on decade-long structural studies of SARS coronavirus. *Journal of Virology*, 94(7). doi: 10.1128/JVI.00127-20
- Wang, J., Peng, Y., Xu, H., Cui, Z. y Williams, R.O. (2020). The COVID-19 vaccine race: challenges and opportunities in vaccine formulation. *AAPS PharmSciTech*, 21(6): 225. doi: 10.1208/s12249-020-01744-7
- Watanabe, Y., Allen, J.D., Wrapp, D., McLellan, J.S. y Crispin, M. (2020). Site-specific glycan analysis of the SARS-CoV-2 spike. *Science*, 369(6501): 330-333. doi: 10.1126/science.abb9983
- Weiskopf, D., Angelo, M.A., de Azeredo, E.L., Sidney, J., Greenbaum, J.A., Fernando, A.N., Broadwater, A., Kolla, R.V., de Silva, A.D., de Silva, A.M., Mattia, K.A., Doranz, B.J., Grey, H.M., Shrestha, S., Peters, B. y Sette, A. (2013). Comprehensive analysis of dengue virus-specific responses supports an HLA-linked protective role for CD8+ T cells. *Proceedings of the National Academy of Sciences*, 110(22): E2046–E2053. doi: 10.1073/pnas.1305227110
- Wolter, N., Jassat, W., Walaza, S., Welch, R., Moultrie, H., Groome, M., Amoako, D.G., Everatt, J., Bhiman, J.N., Scheepers, C., Tebeila, N., Chiwandire, N., du Plessis, M., Govender, N., Ismail, A., Glass, A., Mlisana, K., Stevens, W., Treurnicht, F.K., Makatini, Z., Hsiao, N., Parboosing, R., Wadula, J., Hussey, H., Davies, M.-A., Boulle, A., von

- Gottberg, A. y Cohen, C. (2022). Early assessment of the clinical severity of the SARS-CoV-2 Omicron variant in South Africa: a data linkage study. *The Lancet*, 399(10323): 437-446. doi: 10.1016/S0140-6736(22)00017-4
- World Health Organization. (2020a). Origin of SARS-CoV-2. Disponible en: <https://www.who.int/publications/i/item/origin-of-sars-cov-2>
- World Health Organization. (2020b). Timeline of WHO's response to COVID-19. Disponible en: <https://www.who.int/news/item/29-06-2020-covidtimeline>
- World Health Organization. (2020c). Coronavirus - Symptoms. Disponible en: https://www.who.int/health-topics/coronavirus#tab=tab_3
- World Health Organization. (2023a). WHO Coronavirus Disease (COVID-19) Dashboard. Disponible en: <https://covid19.who.int/>
- World Health Organization. (2023b). Tracking SARS-CoV-2 variants. Disponible en: <https://www.who.int/activities/tracking-SARS-CoV-2-variants>
- Wu, F., Zhao, S., Yu, B., Chen, Y.-M., Wang, W., Song, Z.-G., Hu, Y., Tao, Z.-W., Tian, J.-H., Pei, Y.-Y., Yuan, M.-L., Zhang, Y.-L., Dai, F.-H., Liu, Y., Wang, Q.-M., Zheng, J.-J., Xu, L., Holmes, E.C. y Zhang, Y.-Z. (2020). A new coronavirus associated with human respiratory disease in China. *Nature*, 579(7798): 265-269. doi: 10.1038/s41586-020-2008-3
- Wu, Y., Long, Y., Wang, F., Liu, W. y Wang, Y. (2022). Emergence of SARS-CoV-2 Omicron variant and strategies for tackling the infection. *Immunity, Inflammation and Disease*, 10(12). doi: 10.1002/iid3.733
- Yan, R., Zhang, Y., Li, Y., Ye, F., Guo, Y., Xia, L., Zhong, X., Chi, X. y Zhou, Q. (2021). Structural basis for the different states of the spike protein of SARS-CoV-2 in complex with ACE2. *Cell Research*, 31(6): 717-719. doi: 10.1038/s41422-021-00490-0
- Yang, Z., Bogdan, P. y Nazarian, S. (2021). An *in silico* deep learning approach to multi-epitope vaccine design: a SARS-CoV-2 case study. *Scientific Reports*, 11(1): 3238. doi: 10.1038/s41598-021-81749-9
- Ye, Q., West, A.M.V., Silletti, S. y Corbett, K.D. (2020). Architecture and self-assembly of the SARS-CoV-2 nucleocapsid protein. *Protein Science*, 29(9): 1890-1901. doi: 10.1002/pro.3909
- Zeng, W., Liu, G., Ma, H., Zhao, D., Yang, Y., Liu, M., Mohammed, A., Zhao, C., Yang, Y., Xie, J., Ding, C., Ma, X., Weng, J., Gao, Y., He, H. y Jin, T. (2020). Biochemical characterization of SARS-CoV-2 nucleocapsid protein. *Biochemical and Biophysical Research Communications*, 527(3): 618-623. doi: 10.1016/j.bbrc.2020.04.136
- Zhang, T., Wu, Q. y Zhang, Z. (2020). Probable pangolin origin of SARS-CoV-2 associated with the COVID-19 outbreak. *Current Biology*, 30(7): 1346-1351.e2. doi: 10.1016/j.cub.2020.03.022
- Zhou, P., Yang, X.-L., Wang, X.-G., Hu, B., Zhang, L., Zhang, W., Si, H.-R., Zhu, Y., Li, B., Huang, C.-L., Chen, H.-D., Chen, J., Luo, Y., Guo, H., Jiang, R.-D., Liu, M.-Q., Chen, Y., Shen, X.-R., Wang, X., Zheng, X.-S., Zhao, K., Chen, Q.-J., Deng, F., Liu, L.-L., Yan, B., Zhan, F.-X., Wang, Y.-Y., Xiao, G.-F. y Shi, Z.-L. (2020). A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature*, 579(7798): 270-273. doi: 10.1038/s41586-020-2012-7

11. ANEXOS

Anexo A. Listas de epítomos finales

Tabla 37. Epítomos para células T CD8+ seleccionados en la proteína E.

Epítopo	Posición	Antigenicidad	Toxicidad
VTLAILTALR	29-38	0.8404	No Tóxico
VTLAILTAL	29-37	0.614	No Tóxico
TLAILTALR	30-38	0.7223	No Tóxico
LAILTALRL	31-39	0.8872	No Tóxico
NIVNVSLVK	45-53	0.931	No Tóxico
NVSLVKPSFY	47-57	0.7279	No Tóxico
NVSLVKPSF	47-56	0.6836	No Tóxico
VSLVKPSFYVY	48-59	0.616	No Tóxico
VSLVKPSFY	48-57	0.7476	No Tóxico
VSLVKPSF	48-56	0.6901	No Tóxico
SLVKPSFYV	49-58	0.414	No Tóxico
SLVKPSFY	49-57	0.4174	No Tóxico
LVKPSFYVYSR	50-61	0.6211	No Tóxico
LVKPSFYVY	50-59	0.4213	No Tóxico
KPSFYVYSR	52-61	0.974	No Tóxico
SFYVYSRVK	54-63	0.8251	No Tóxico
FYVYSRVKNL	55-65	0.8286	No Tóxico
YVYSRVKNL	56-65	0.702	No Tóxico
VYSRVKNL	57-65	0.8466	No Tóxico
YSRVKNLNSSR	58-69	0.5512	No Tóxico
SRVKNLNSSR	59-69	0.6612	No Tóxico
RVKNLNSSRV	60-70	0.8936	No Tóxico
RVKNLNSSR	60-69	0.8998	No Tóxico
SSRVPDLLV	67-75	0.5455	No Tóxico

Tabla 38. Epítomos para células T CD4+ seleccionados en la proteína E.

Epítomo	Posición	Antigenicidad	Toxicidad
LVTLAILTALRL	28-39	0.7377	No Tóxico
VTLAILTALRLC	29-40	0.6089	No Tóxico
VSLVKPSFYVYSRVKLN	47-66	0.932	No Tóxico
SLVKPSFYVYSRVKLNLS	48-67	0.6192	No Tóxico
SLVKPSFYVYSRVKLN	48-66	0.8087	No Tóxico
LVKPSFYVYSRVKLNSS	49-68	0.716	No Tóxico
LVKPSFYVYSRVKLNLS	49-67	0.6582	No Tóxico
LVKPSFYVYSRVKLN	49-66	0.8657	No Tóxico
LVKPSFYVYSRVKLN	49-65	0.7311	No Tóxico
VKPSFYVYSRVKLNSSR	50-69	0.8946	No Tóxico
VKPSFYVYSRVKLNSS	50-68	1.0211	No Tóxico
VKPSFYVYSRVKLNLS	50-67	0.9811	No Tóxico
VKPSFYVYSRVKLN	50-66	1.2319	No Tóxico
VKPSFYVYSRVKLN	50-65	1.1188	No Tóxico
KPSFYVYSRVKLNSSRV	51-70	0.7431	No Tóxico
KPSFYVYSRVKLNSSR	51-69	0.7509	No Tóxico
KPSFYVYSRVKLNSS	51-68	0.878	No Tóxico
KPSFYVYSRVKLNLS	51-67	0.8229	No Tóxico
KPSFYVYSRVKLN	51-66	1.0831	No Tóxico
KPSFYVYSRVKLN	51-65	0.9472	No Tóxico
KPSFYVYSRVKN	51-64	0.9436	No Tóxico
PSFYVYSRVKLNSSRVP	52-71	0.5796	No Tóxico
PSFYVYSRVKLNSSRV	52-70	0.6651	No Tóxico
PSFYVYSRVKLNSSR	52-69	0.6667	No Tóxico
PSFYVYSRVKLNSS	52-68	0.7986	No Tóxico
PSFYVYSRVKLNLS	52-67	0.7301	No Tóxico
PSFYVYSRVKLN	52-66	1.0066	No Tóxico
PSFYVYSRVKLN	52-65	0.8491	No Tóxico
SFYVYSRVKLNSSRV	53-70	0.631	No Tóxico
SFYVYSRVKLNSSR	53-69	0.6291	No Tóxico
SFYVYSRVKLNSS	53-68	0.7713	No Tóxico
SFYVYSRVKLNLS	53-67	0.6919	No Tóxico
SFYVYSRVKLN	53-66	0.9948	No Tóxico
FYVYSRVKLNSSRVPDL	54-73	0.7442	No Tóxico
FYVYSRVKLNSSRVPD	54-72	0.6007	No Tóxico
FYVYSRVKLNSSRVP	54-71	0.5174	No Tóxico
FYVYSRVKLNSSRV	54-70	0.6103	No Tóxico
FYVYSRVKLNSSR	54-69	0.6062	No Tóxico
FYVYSRVKLNSS	54-68	0.7642	No Tóxico
FYVYSRVKLNLS	54-67	0.6751	No Tóxico
YVYSRVKLNSSRVPDLL	55-74	0.6135	No Tóxico

YVYSRVKKNLNSSRVPDL	55-73	0.7038	No Tóxico
YVYSRVKKNLNSSRVPD	55-72	0.5457	No Tóxico
YVYSRVKKNLNSSRVP	55-71	0.4492	No Tóxico
YVYSRVKKNLNSSRV	55-70	0.5405	No Tóxico
YVYSRVKKNLNSSR	55-69	0.5246	No Tóxico
YVYSRVKKNLNSS	55-68	0.6906	No Tóxico
VYSRVKKNLNSSRVPDLL	56-74	0.6646	No Tóxico
VYSRVKKNLNSSRVPDL	56-73	0.7648	No Tóxico
VYSRVKKNLNSSRVPD	56-72	0.5993	No Tóxico
VYSRVKKNLNSSRVP	56-71	0.4983	No Tóxico
VYSRVKKNLNSSR	56-69	0.5867	No Tóxico
YSRVKKNLNSSRVPDLL	57-74	0.6644	No Tóxico
YSRVKKNLNSSRVPDL	57-73	0.7705	No Tóxico
YSRVKKNLNSSRVPD	57-72	0.5896	No Tóxico
YSRVKKNLNSSRVP	57-71	0.4745	No Tóxico
SRVKKNLNSSRVPDLL	58-74	0.7404	No Tóxico
SRVKKNLNSSRVPDL	58-73	0.8624	No Tóxico
SRVKKNLNSSRVPD	58-72	0.675	No Tóxico
SRVKKNLNSSRVP	58-71	0.5572	No Tóxico
RVKNLNSSRVPD	59-72	0.8416	No Tóxico

Tabla 39. Epítomos para células T CD8+ seleccionados en la proteína M.

Epítopo	Posición	Antigenicidad	Toxicidad
NGTITVEELK	5-14	0.9557	No Tóxico
GTITVEELKK	6-15	0.7354	No Tóxico
GTITVEELK	6-14	1.0976	No Tóxico
GTITVEEL	6-13	1.097	No Tóxico
ITVEELKKL	8-16	0.4904	No Tóxico
FAYANRRNFLY	37-47	0.5547	No Tóxico
FAYANRRNFL	37-46	0.666	No Tóxico
FAYANRRNF	37-45	0.7785	No Tóxico
FAYANRRNR	37-44	1.2938	No Tóxico
NRNRFLYII	41-49	0.4868	No Tóxico
NRNRFLYI	41-48	0.6221	No Tóxico
FVLAAYRINW	65-75	1.258	No Tóxico
FVLAAYRI	65-73	0.5136	No Tóxico
VLAAYRINW	66-75	1.1808	No Tóxico
LAAYRINW	67-75	1.4322	No Tóxico
AAVYRINW	68-75	1.0625	No Tóxico
YRINWITGGIA	71-81	1.2743	No Tóxico
SYFIASFRL	94-102	0.4821	No Tóxico
SYFIASFR	94-101	0.479	No Tóxico
YFIASFRL	95-102	0.5814	No Tóxico
ASFRLFARTRSMW	98-110	0.5816	No Tóxico
ASFRLFARTR	98-107	0.6237	No Tóxico
SFRLFARTR	99-107	0.7038	No Tóxico
FRLFARTRSMW	100-110	0.5033	No Tóxico
FRLFARTRSM	100-109	0.4609	No Tóxico
RLFARTRSMWSF	101-112	0.5383	No Tóxico
LFARTRSMW	102-110	0.856	No Tóxico
LFARTRSM	102-109	0.8363	No Tóxico
FARTRSMWSF	103-112	0.9202	No Tóxico
ARTRSMWSF	104-112	1.2394	No Tóxico
RTRSMWSF	105-112	1.4716	No Tóxico
SMWSFNPETNIL	108-119	0.4106	No Tóxico
SMWSFNPET	108-116	0.7633	No Tóxico
WSFNPETNI	110-118	0.755	No Tóxico
SFNPETNILL	11-120	0.4676	No Tóxico
SELVIGAVILR	136-146	0.8265	No Tóxico
SELVIGAVIL	136-145	0.6521	No Tóxico
SELVIGAVI	136-144	0.6409	No Tóxico
SELVIGAV	136-143	0.9691	No Tóxico
ELVIGAVILR	137-146	0.9998	No Tóxico
LVIGAVILR	138-146	1.1027	No Tóxico

LPKEITVA	164-171	0.5063	No Tóxico
KEITVATSRTL	166-176	0.5818	No Tóxico
EITVATSRTL	167-178	0.8558	No Tóxico
EITVATSRTL	167-176	0.6074	No Tóxico
ITVATSRTL	168-178	0.8622	No Tóxico
ITVATSRTL	168-176	0.58	No Tóxico
TVATSRTL	169-179	0.6582	No Tóxico
TVATSRTL	169-178	0.8259	No Tóxico
TVATSRTL	169-176	0.4976	No Tóxico
VATSRTL	170-179	0.7297	No Tóxico
VATSRTL	170-178	0.9457	No Tóxico
ATSRTL	171-180	0.4317	No Tóxico
ATSRTL	171-179	0.6108	No Tóxico
ATSRTL	171-178	0.8451	No Tóxico
TSRTL	172-180	0.4811	No Tóxico
RTLSYKLGASQR	174-186	0.7034	No Tóxico
SYKLGASQR	177-186	0.7559	No Tóxico
YKLGASQR	178-186	0.9181	No Tóxico
YKLGASQRV	179-187	0.522	No Tóxico
RVAGDSGFAAYSR	186-198	0.533	No Tóxico
VAGDSGFAAY	187-196	0.589	No Tóxico
AGDSGFAAY	188-196	0.9095	No Tóxico
DSGFAAYSR	190-200	1.1195	No Tóxico
DSGFAAYSR	190-199	1.0869	No Tóxico
DSGFAAYSR	190-198	1.1733	No Tóxico
SGFAAYSR	191-200	1.0034	No Tóxico
SGFAAYSR	191-199	0.9471	No Tóxico
GFAAYSR	192-200	1.2734	No Tóxico
FAAYSR	193-201	0.8787	No Tóxico
AAYSR	194-201	0.6583	No Tóxico
NTDHSSSDNIAL	207-219	0.4648	No Tóxico
DHSSSDNIAL	209-220	0.4226	No Tóxico
DHSSSDNIAL	209-219	0.4106	No Tóxico
SSDNIAL	213-220	0.4351	No Tóxico

Tabla 40. Epítomos para células T CD4+ seleccionados en la proteína M.

Epítopo	Posición	Antigenicidad	Toxicidad
SNGTITVEELKKL	4-16	0.5137	No tóxico
NGTITVEELKKL	5-16	0.5412	No tóxico
FVLAAYRINWITGGIA	65-81	1.0081	No tóxico
VLAAYRINWITGGIA	66-81	0.9496	No tóxico
VLAAYRINWITGGI	66-80	0.9478	No tóxico
LAAYRINWITGGIA	67-81	1.0581	No tóxico
LAAYRINWITGGI	67-80	1.0671	No tóxico
AAVYRINWITGGIA	68-81	0.8848	No tóxico
AAVYRINWITGGI	68-80	0.8773	No tóxico
AAVYRINWITGG	68-79	0.8799	No tóxico
AVYRINWITGGIA	69-81	0.946	No tóxico
AVYRINWITGGI	69-80	0.9473	No tóxico
VYRINWITGGIA	70-81	0.9616	No tóxico
YFIASFRLFARTRSMWSF	95-112	0.5641	No tóxico
YFIASFRLFARTRSMW	95-110	0.421	No tóxico
FIASFRLFARTRSMWSF	96-112	0.5614	No tóxico
FIASFRLFARTRSMW	96-110	0.4072	No tóxico
IASFRLFARTRSMWSFN	97-113	0.6634	No tóxico
IASFRLFARTRSMWSF	97-112	0.629	No tóxico
IASFRLFARTRSMWS	97-111	0.4424	No tóxico
IASFRLFARTRSMW	97-110	0.4727	No tóxico
IASFRLFARTRSM	97-109	0.4431	No tóxico
IASFRLFARTRS	97-108	0.4473	No tóxico
ASFRLFARTRSMWSF	98-112	0.7304	No tóxico
ASFRLFARTRSMWS	98-111	0.5372	No tóxico
ASFRLFARTRSMW	98-110	0.5816	No tóxico
ASFRLFARTRSM	98-109	0.5577	No tóxico
SFRLFARTRSMW	99-110	0.6183	No tóxico
RSMWSFNPETNIL	107-119	0.457	No tóxico
SELVIGAVILRGH	136-148	0.649	No tóxico
SELVIGAVILRG	136-147	0.7605	No tóxico
ELVIGAVILRGH	137-148	0.7605	No tóxico
LPKEITVATSRTLSSYY	164-179	0.4597	No tóxico
LPKEITVATSRTLSSY	164-178	0.54	No tóxico
LPKEITVATSRTLSS	164-177	0.4228	No tóxico
PKEITVATSRTLSSYYKLG	165-182	0.7689	No tóxico
PKEITVATSRTLSSYYKL	165-181	0.59	No tóxico
PKEITVATSRTLSSYYK	165-180	0.5935	No tóxico
PKEITVATSRTLSSYY	165-179	0.7003	No tóxico
PKEITVATSRTLSSY	165-178	0.8084	No tóxico
PKEITVATSRTLSS	165-177	0.7044	No tóxico

PKEITVATSRTL	165-176	0.6166	No tóxico
KEITVATSRTLSYYK	166-180	0.5731	No tóxico
KEITVATSRTLSYY	166-179	0.6861	No tóxico
KEITVATSRTLSY	166-178	0.8025	No tóxico
KEITVATSRTLS	166-177	0.6843	No tóxico
ATSRTLSYYKLGASQRVA	171-188	0.6107	No tóxico
ATSRTLSYYKLGASQRV	171-187	0.5909	No tóxico
TSRTLSYYKLGASQRVAG	172-189	0.4645	No tóxico
TSRTLSYYKLGASQRVA	172-188	0.6398	No tóxico
TSRTLSYYKLGASQRV	172-187	0.6212	No tóxico
SRTLSYYKLGASQRVA	173-188	0.578	No tóxico
SRTLSYYKLGASQRV	173-187	0.5526	No tóxico
SRTLSYYKLGASQR	173-186	0.7061	No tóxico
RTLSYYKLGASQRVA	174-188	0.5644	No tóxico
RTLSYYKLGASQRV	174-187	0.5358	No tóxico
RTLSYYKLGASQR	174-186	0.7034	No tóxico
TLSYYKLGASQRVAG	175-189	0.4376	No tóxico
TLSYYKLGASQRVA	175-188	0.6573	No tóxico
TLSYYKLGASQRV	175-187	0.6352	No tóxico
TLSYYKLGASQR	175-186	0.8331	No tóxico
LSYYKLGASQRVAGDSGF	176-193	0.424	No tóxico
LSYYKLGASQRVAG	176-189	0.4264	No tóxico
LSYYKLGASQRVA	176-188	0.666	No tóxico
LSYYKLGASQRV	176-187	0.6421	No tóxico
SYKLGASQRVA	177-188	0.5528	No tóxico
ASQRVAGDSGFAAYS	183-193	0.479	No tóxico
SQRVAGDSGFAAYS	184-193	0.5226	No tóxico
RVAGDSGFAAYSRYRIGN	186-203	0.5429	No tóxico
VAGDSGFAAYSRYRIGNY	187-204	0.7097	No tóxico
VAGDSGFAAYSRYRIGN	187-203	0.6876	No tóxico
VAGDSGFAAYSRYRIG	187-202	0.7552	No tóxico
AGDSGFAAYSRYRIGNYK	188-205	0.6982	No tóxico
AGDSGFAAYSRYRIGNY	188-204	0.8472	No tóxico
AGDSGFAAYSRYRIGN	188-203	0.8348	No tóxico
AGDSGFAAYSRYRIG	188-202	0.9205	No tóxico
GDSGFAAYSRYRIGNYKL	189-206	0.6842	No tóxico
GDSGFAAYSRYRIGNYK	189-205	0.7313	No tóxico
GDSGFAAYSRYRIGNY	189-204	0.898	No tóxico
GDSGFAAYSRYRIGN	189-203	0.89	No tóxico
GDSGFAAYSRYRIG	189-202	0.9891	No tóxico
DSGFAAYSRYRIGNYKLN	190-207	0.7204	No tóxico
DSGFAAYSRYRIGNYKL	190-206	0.6463	No tóxico
DSGFAAYSRYRIGNYK	190-205	0.6929	No tóxico

DSGFAAYSRYRIGNY	190-204	0.8741	No tóxico
DSGFAAYSRYRIGN	190-203	0.8634	No tóxico
DSGFAAYSRYRIG	190-202	0.969	No tóxico
DSGFAAYSRYRI	190-201	0.9087	No tóxico
SGFAAYSRYRIGNYK	191-205	0.5817	No tóxico
SGFAAYSRYRIGNY	191-204	0.7738	No tóxico
SGFAAYSRYRIGN	191-203	0.7519	No tóxico
SGFAAYSRYRIG	191-202	0.855	No tóxico
GFAAYSRYRIGN	192-203	0.8975	No tóxico

Tabla 41. Epítomos para células T CD8+ seleccionados en la proteína N.

Epítomo	Posición	Antigenicidad	Toxicidad
RPQGLPNNTASW	41-52	0.4563	No Tóxico
RPQGLPNNTA	41-50	0.555	No Tóxico
RPQGLPNNT	41-49	0.576	No Tóxico
SWFTALTQHGK	51-61	0.413	No Tóxico
FTALTQHGK	53-61	0.851	No Tóxico
VPINTNSSP	72-80	0.4439	No Tóxico
NTNSSPDDQIGYY	75-87	0.6859	No Tóxico
NTNSSPDDQIGY	75-86	0.7734	No Tóxico
TNSSPDDQIGYY	76-87	0.682	No Tóxico
TNSSPDDQIGY	76-86	0.7761	No Tóxico
NSSPDDQIGYY	77-87	0.5757	No Tóxico
NSSPDDQIGY	77-86	0.663	No Tóxico
SSPDDQIGYY	78-87	0.4533	No Tóxico
SSPDDQIGY	78-86	0.526	No Tóxico
SPDDQIGYY	79-87	0.4863	No Tóxico
SPDDQIGY	79-86	0.5569	No Tóxico
LSPRWYFY	104-112	1.2832	No Tóxico
SPRWYFY	105-113	0.734	No Tóxico
YYLGTGPEAGLPY	111-123	0.4121	No Tóxico
YYLGTGPEAGL	111-121	0.56	No Tóxico
YYLGTGPEA	111-119	0.7969	No Tóxico
QASSRSSSR	181-189	0.8294	No Tóxico
RTATKAYNVTQAF	262-274	0.4965	No Tóxico
RTATKAYNV	262-270	0.6395	No Tóxico
ATKAYNVTQAF	264-274	0.5521	No Tóxico
TKAYNVTQAF	265-274	0.4932	No Tóxico
KAYNVTQAF	266-274	0.5669	No Tóxico
AYNVTQAF	267-274	0.514	No Tóxico
GRRGPEQTQGNF	275-286	0.898	No Tóxico
RRGPEQTQGNF	276-286	0.5763	No Tóxico
RGPEQTQGNF	277-286	0.4929	No Tóxico
GPEQTQGNF	278-286	0.7349	No Tóxico

TDYKHWPQI	296-304	0.7426	No Tóxico
DYKHWPQIAQF	297-307	0.6414	No Tóxico
HWPQIAQF	300-307	0.4161	No Tóxico
WPQIAQFAPSASA	301-313	0.4415	No Tóxico
WPQIAQFAPSA	301-311	0.4908	No Tóxico
WPQIAQFAP	301-309	0.611	No Tóxico
AQFAPSASAFF	305-315	0.4775	No Tóxico
AQFAPSASAF	305-314	0.5986	No Tóxico
AQFAPSASA	305-313	0.7468	No Tóxico
QFAPSASAFF	306-315	0.419	No Tóxico
QFAPSASAF	306-314	0.5495	No Tóxico
FAPSASAFFGM	307-317	0.488	No Tóxico
SAFFGMSRI	312-320	0.4244	No Tóxico
AFFGMSRIGM	313-322	0.9812	No Tóxico
FFGMSRIGM	314-322	0.9199	No Tóxico
GMSRIGMEV	316-324	0.6287	No Tóxico
MSRIGMEVTPSGTW	317-330	0.8252	No Tóxico
MSRIGMEV	317-324	0.9671	No Tóxico
SRIGMEVTPSGTW	318-330	0.9264	No Tóxico
RIGMEVTPSGTW	319-330	0.9171	No Tóxico
IGMEVTPSGTW	320-330	0.9722	No Tóxico
GMEVTPSGTW	321-330	0.6967	No Tóxico
MEVTPSGTWLTY	322-333	0.7571	No Tóxico
MEVTPSGTWL	322-331	0.6342	No Tóxico
MEVTPSGTW	322-330	0.755	No Tóxico
EVTPSGTWLTY	323-333	0.6364	No Tóxico
EVTPSGTWL	323-331	0.4548	No Tóxico
EVTPSGTW	323-330	0.5552	No Tóxico
LYTGAIKL	331-339	0.6524	No Tóxico
IKLDDKDPNF	337-346	2.6945	No Tóxico
KLDDKDPNFK	338-347	2.1298	No Tóxico
KLDDKDPNF	338-346	2.6591	No Tóxico
KDPNFKDQV	342-350	1.3338	No Tóxico
DPNFKDQVIL	343-352	1.551	No Tóxico
DPNFKDQVI	343-351	1.7367	No Tóxico
NFKDQVILL	345-353	1.1677	No Tóxico
NFKDQVIL	345-352	1.1403	No Tóxico
FKDQVILL	346-353	0.6457	No Tóxico
DQVILLNKH	348-356	0.5069	No Tóxico
DAYKTFPPTEPK	358-369	0.4302	No Tóxico
AYKTFPPTEPK	359-369	0.6958	No Tóxico
YKTFPPTEPKK	360-370	0.7617	No Tóxico
YKTFPPTEPK	360-369	0.7633	No Tóxico
KTFPPTEPKKD	361-371	0.69	No Tóxico
KTFPPTEPKK	361-370	0.7657	No Tóxico
KTFPPTEPK	361-369	0.7571	No Tóxico

QRQKKQQTV	384-392	0.5602	No Tóxico
RQKKQQTVTLL	385-395	0.6819	No Tóxico
RQKKQQTVTL	385-394	0.7542	No Tóxico
RQKKQQTV	385-392	0.6217	No Tóxico
QKKQQTVTLL	386-395	0.8129	No Tóxico
QKKQQTVTL	386-394	0.9285	No Tóxico
KKQQTVTLL	387-395	0.7709	No Tóxico
KKQQTVTL	387-394	0.8972	No Tóxico
KQQTVTLLP	388-396	0.8116	No Tóxico
KQQTVTLL	388-395	0.6199	No Tóxico
QTVTLLPAA	390-398	0.6871	No Tóxico

Tabla 42. Epítomos para células T CD4+ seleccionados en la proteína N.

Epítomo	Posición	Antigenicidad	Toxicidad
SWFTALTQHGKE	51-62	0.4008	No Tóxico
QGVPIINTNSSPDDQIG	70-85	0.6303	No Tóxico
QGVPIINTNSSPDDQI	70-84	0.4461	No Tóxico
GVPINTNSSPDDQIGY	71-86	0.5959	No Tóxico
GVPINTNSSPDDQIG	71-85	0.5606	No Tóxico
GVPINTNSSPDDQ	71-83	0.5195	No Tóxico
VPINTNSSPDDQ	72-83	0.4834	No Tóxico
LSPRWYFYLLGTGPEAGL	104-121	1.0484	No Tóxico
LSPRWYFYLLGTGPEAG	104-120	1.1414	No Tóxico
LSPRWYFYLLGTGPEA	104-119	1.2052	No Tóxico
SPRWYFYLLGTGPEAGLP	105-122	0.6261	No Tóxico
SPRWYFYLLGTGPEAGL	105-121	0.7518	No Tóxico
SPRWYFYLLGTGPEAG	105-120	0.8315	No Tóxico
SPRWYFYLLGTGPEA	105-119	0.8767	No Tóxico
SPRWYFYLLGTGPE	105-118	0.9616	No Tóxico
PRWYFYLLGTGPEAGLPY	106-123	0.5979	No Tóxico
PRWYFYLLGTGPEAGLP	106-122	0.5888	No Tóxico
PRWYFYLLGTGPEAGL	106-121	0.7219	No Tóxico
PRWYFYLLGTGPEAG	106-120	0.8083	No Tóxico
PRWYFYLLGTGPEA	106-119	0.8577	No Tóxico
PRWYFYLLGTGPE	106-118	0.9516	No Tóxico
RWYFYLLGTGPEAGLPYG	107-124	0.573	No Tóxico
RWYFYLLGTGPEAGLPY	107-123	0.6129	No Tóxico
RWYFYLLGTGPEAGLP	107-122	0.6049	No Tóxico
RWYFYLLGTGPEAGL	107-121	0.7505	No Tóxico
RWYFYLLGTGPEAG	107-120	0.8502	No Tóxico
RWYFYLLGTGPEA	107-119	0.911	No Tóxico
RWYFYLLGTGPE	107-118	1.0237	No Tóxico
WYFYLLGTGPEAGLP	108-122	0.7188	No Tóxico
WYFYLLGTGPEAGL	108-121	0.8873	No Tóxico
WYFYLLGTGPEAG	108-120	1.0132	No Tóxico

WYFYLLGTGPEA	108-119	1.1004	No Tóxico
YFYLLGTGPEAGLPYGAN	109-126	0.4941	No Tóxico
YFYLLGTGPEAGL	109-121	0.7502	No Tóxico
YFYLLGTGPEAG	109-120	0.8791	No Tóxico
FYYLGTGPEAGLPYGAN	110-126	0.4747	No Tóxico
FYYLGTGPEAGLPYGA	110-125	0.4724	No Tóxico
PRQKRTATKAYNVTQAFG	258-275	0.4021	No Tóxico
PRQKRTATKAYNVTQAF	258-274	0.4552	No Tóxico
PRQKRTATKAYNVTQA	258-273	0.5641	No Tóxico
PRQKRTATKAYNVTQ	258-272	0.4981	No Tóxico
PRQKRTATKAYNVT	258-271	0.5955	No Tóxico
PRQKRTATKAYNV	258-270	0.4904	No Tóxico
PRQKRTATKAYN	258-269	0.4178	No Tóxico
RQKRTATKAYNVTQAFG	259-275	0.4498	No Tóxico
RQKRTATKAYNVTQAF	259-274	0.5112	No Tóxico
RQKRTATKAYNVTQA	259-273	0.6318	No Tóxico
RQKRTATKAYNVTQ	259-272	0.5638	No Tóxico
RQKRTATKAYNVT	259-271	0.6739	No Tóxico
RQKRTATKAYNV	259-270	0.5657	No Tóxico
QKRTATKAYNVTQA	260-273	0.4767	No Tóxico
QKRTATKAYNVT	260-271	0.4832	No Tóxico
KRTATKAYNVTQAFGRRG	261-278	0.5342	No Tóxico
KRTATKAYNVTQA	261-273	0.4389	No Tóxico
RTATKAYNVTQAFGRRGP	262-279	0.5877	No Tóxico
RTATKAYNVTQAFGRRG	262-278	0.6601	No Tóxico
TATKAYNVTQAFGRRGPE	263-280	0.5899	No Tóxico
TATKAYNVTQAFGRRGP	263-279	0.6061	No Tóxico
TATKAYNVTQAFGRRG	263-278	0.6846	No Tóxico
ATKAYNVTQAFGRRGPEQ	264-281	0.4556	No Tóxico
ATKAYNVTQAFGRRGPE	264-280	0.6097	No Tóxico
ATKAYNVTQAFGRRGP	264-279	0.6281	No Tóxico
ATKAYNVTQAFGRRG	261-278	0.7146	No Tóxico
TKAYNVTQAFGRRGPEQT	262-282	0.4413	No Tóxico
TKAYNVTQAFGRRGPEQ	262-281	0.4181	No Tóxico
TKAYNVTQAFGRRGPE	262-280	0.5816	No Tóxico
TKAYNVTQAFGRRGP	262-279	0.5975	No Tóxico
TKAYNVTQAFGRRG	262-278	0.6877	No Tóxico
KAYNVTQAFGRRGPEQTQ	263-283	0.5415	No Tóxico
KAYNVTQAFGRRGPEQT	263-282	0.4542	No Tóxico
KAYNVTQAFGRRGPEQ	263-281	0.4307	No Tóxico
KAYNVTQAFGRRGPE	263-280	0.6104	No Tóxico
KAYNVTQAFGRRGP	263-279	0.6296	No Tóxico
KAYNVTQAFGRRG	263-278	0.7317	No Tóxico
AYNVTQAFGRRGPE	264-280	0.5798	No Tóxico
AYNVTQAFGRRGP	264-279	0.595	No Tóxico
AYNVTQAFGRRG	264-278	0.7025	No Tóxico

GPEQTQGNFGDQELIRQ	278-294	0.5705	No Tóxico
PEQTQGNFGDQELIRQ	279-294	0.558	No Tóxico
PEQTQGNFGDQELIR	279-293	0.627	No Tóxico
EQTQGNFGDQELIRQ	280-294	0.6565	No Tóxico
EQTQGNFGDQELIR	280-293	0.7417	No Tóxico
WPQIAQFAPSASAFFGMS	301-318	0.4822	No Tóxico
PQIAQFAPSASAFFGMS	302-318	0.4663	No Tóxico
PQIAQFAPSASAFFGM	302-317	0.4144	No Tóxico
QIAQFAPSASAFFGMS	303-318	0.4602	No Tóxico
QIAQFAPSASAFFGM	303-317	0.4032	No Tóxico
AQFAPSASAFFGMS	305-318	0.6688	No Tóxico
AQFAPSASAFFGM	305-317	0.6207	No Tóxico
AQFAPSASAFFG	305-316	0.5066	No Tóxico
APSASAFFGMSRIGMEVT	308-325	0.7431	No Tóxico
APSASAFFGMSRIGMEV	308-324	0.5903	No Tóxico
APSASAFFGMSRIGME	308-323	0.6112	No Tóxico
PSASAFFGMSRIGMEVTP	309-326	0.8585	No Tóxico
PSASAFFGMSRIGMEVT	309-325	0.7776	No Tóxico
PSASAFFGMSRIGMEV	309-324	0.6161	No Tóxico
PSASAFFGMSRIGME	309-323	0.6408	No Tóxico
SASAFFGMSRIGMEVTP	310-326	0.9123	No Tóxico
SASAFFGMSRIGMEVT	310-325	0.8291	No Tóxico
SASAFFGMSRIGMEV	310-324	0.6584	No Tóxico
SASAFFGMSRIGME	310-323	0.6882	No Tóxico
ASAFFGMSRIGMEVTP	311-326	0.9486	No Tóxico
ASAFFGMSRIGMEVT	311-325	0.862	No Tóxico
ASAFFGMSRIGMEV	311-324	0.6795	No Tóxico
ASAFFGMSRIGME	311-323	0.7144	No Tóxico
ASAFFGMSRIGM	311-322	0.7097	No Tóxico
SAFFGMSRIGMEVT	312-325	0.9543	No Tóxico
SAFFGMSRIGMEV	312-324	0.7641	No Tóxico
SAFFGMSRIGME	312-323	0.8127	No Tóxico
GMSRIGMEVTPSGTW	316-330	0.6835	No Tóxico
GMSRIGMEVTPSGT	316-329	0.893	No Tóxico
MSRIGMEVTPSGTW	317-330	0.8252	No Tóxico
MSRIGMEVTPSGT	317-329	1.0712	No Tóxico
SRIGMEVTPSGTW	318-330	0.9264	No Tóxico
SRIGMEVTPSGT	318-329	1.21	No Tóxico
EVTPSGTWLTYTGAIKLD	323-340	0.6186	No Tóxico
EVTPSGTWLTYTGAIKL	323-339	0.603	No Tóxico
VTPSGTWLTYTGAIKLDD	324-341	0.5608	No Tóxico
VTPSGTWLTYTGAIKLD	324-340	0.469	No Tóxico
VTPSGTWLTYTGAIKL	324-339	0.4404	No Tóxico
TPSGTWLTYTGAIKLDDK	325-342	0.758	No Tóxico
TPSGTWLTYTGAIKLDD	325-341	0.4711	No Tóxico
PSGTWLTGTGAIKLDDKD	326-343	0.9364	No Tóxico

PSGTWLYTGAIKLDDK	326-342	0.7717	No Tóxico
PSGTWLYTGAIKLDD	326-341	0.4631	No Tóxico
SGTWLYTGAIKLDDKDP	327-344	1.0942	No Tóxico
SGTWLYTGAIKLDDKD	327-343	1.1073	No Tóxico
SGTWLYTGAIKLDDK	327-342	0.9425	No Tóxico
SGTWLYTGAIKLDD	327-341	0.6215	No Tóxico
SGTWLYTGAIKLD	327-340	0.5069	No Tóxico
SGTWLYTGAIKL	327-339	0.4749	No Tóxico
GTWLYTGAIKLDDKDPN	328-345	1.1685	No Tóxico
GTWLYTGAIKLDDKDP	328-344	1.1493	No Tóxico
GTWLYTGAIKLDDKD	328-343	1.1681	No Tóxico
GTWLYTGAIKLDDK	328-342	0.9934	No Tóxico
GTWLYTGAIKLDD	328-341	0.646	No Tóxico
GTWLYTGAIKLD	328-340	0.5191	No Tóxico
GTWLYTGAIKL	328-339	0.483	No Tóxico
TWLYTGAIKLDDKDPNF	329-346	1.4926	No Tóxico
TWLYTGAIKLDDKDPN	329-345	1.2307	No Tóxico
TWLYTGAIKLDDKDP	329-344	1.2148	No Tóxico
TWLYTGAIKLDDKD	329-343	1.2416	No Tóxico
TWLYTGAIKLDDK	329-342	1.0566	No Tóxico
TWLYTGAIKLDD	329-341	0.6787	No Tóxico
TWLYTGAIKLD	329-340	0.5375	No Tóxico
WLYTGAIKLDDKDPNFK	330-347	1.4415	No Tóxico
WLYTGAIKLDDKDPNF	330-346	1.5679	No Tóxico
WLYTGAIKLDDKDPN	330-345	1.2914	No Tóxico
WLYTGAIKLDDKDP	330-344	1.2787	No Tóxico
WLYTGAIKLDDKD	330-343	1.3139	No Tóxico
WLYTGAIKLDDK	330-342	1.1147	No Tóxico
WLYTGAIKLDD	330-341	0.6967	No Tóxico
LTYTGAIKLDDKDPNFKD	331-348	1.656	No Tóxico
LTYTGAIKLDDKDPNFK	331-347	1.6031	No Tóxico
LTYTGAIKLDDKDPNF	331-346	1.7537	No Tóxico
LTYTGAIKLDDKDPN	331-345	1.4693	No Tóxico
LTYTGAIKLDDK	331-342	1.3268	No Tóxico
TYTGAIKLDDKDPNFKDQ	332-349	1.6068	No Tóxico
TYTGAIKLDDKDPNFKD	332-348	1.5973	No Tóxico
TYTGAIKLDDKDPNFK	332-347	1.5361	No Tóxico
TYTGAIKLDDKDPNF	332-346	1.6988	No Tóxico
TYTGAIKLDDKDPN	332-345	1.386	No Tóxico
YTGAIKLDDKDPNFKDQV	333-350	1.4556	No Tóxico
YTGAIKLDDKDPNFKDQ	333-349	1.6794	No Tóxico
YTGAIKLDDKDPNFKD	333-348	1.674	No Tóxico
YTGAIKLDDKDPNFK	333-347	1.6136	No Tóxico
YTGAIKLDDKDPNF	333-346	1.8035	No Tóxico
YTGAIKLDDKDPN	333-345	1.4717	No Tóxico
TGAIKLDDKDPNFKDQVI	334-351	1.5529	No Tóxico

TGAIKLDDKDPNFKDQV	334-350	1.5922	No Tóxico
TGAIKLDDKDPNFKDQ	334-349	1.8438	No Tóxico
TGAIKLDDKDPNFKD	334-348	1.8513	No Tóxico
TGAIKLDDKDPNFK	334-347	1.801	No Tóxico
TGAIKLDDKDPNF	334-346	2.0364	No Tóxico
TGAIKLDDKDPN	334-345	1.6967	No Tóxico
GAIKLDDKDPNFKDQVIL	335-352	1.5303	No Tóxico
GAIKLDDKDPNFKDQVI	335-351	1.6128	No Tóxico
GAIKLDDKDPNFKDQV	335-350	1.6587	No Tóxico
GAIKLDDKDPNFKDQ	335-349	1.9373	No Tóxico
GAIKLDDKDPNFKD	335-348	1.9535	No Tóxico
GAIKLDDKDPNFK	335-347	1.908	No Tóxico
GAIKLDDKDPNF	335-346	2.1926	No Tóxico
AIKLDDKDPNFKDQV	336-350	1.7616	No Tóxico
AIKLDDKDPNFKDQ	336-349	2.074	No Tóxico
AIKLDDKDPNFKD	336-348	2.103	No Tóxico
AIKLDDKDPNFK	336-347	2.0658	No Tóxico
LDDKDPNFKDQVILLNKH	339-356	1.3256	No Tóxico
DDKDPNFKDQVILLNKHI	340-357	1.0582	No Tóxico
DDKDPNFKDQVILLNKH	340-356	1.1991	No Tóxico
DKDPNFKDQVILLNKHID	341-358	1.0339	No Tóxico
DKDPNFKDQVILLNKHI	341-357	0.982	No Tóxico
DKDPNFKDQVILLNKH	341-356	1.1285	No Tóxico
DKDPNFKDQVILLNK	341-355	1.1508	No Tóxico
KDPNFKDQVILLNKHIDA	342-359	0.9275	No Tóxico
KDPNFKDQVILLNKHID	342-358	1.0343	No Tóxico
KDPNFKDQVILLNKHI	342-357	0.9792	No Tóxico
KDPNFKDQVILLNKH	342-356	1.1376	No Tóxico
KDPNFKDQVILLNK	342-355	1.1626	No Tóxico
DPNFKDQVILLNKHIDA	343-359	1.029	No Tóxico
DPNFKDQVILLNKHID	343-358	1.1532	No Tóxico
DPNFKDQVILLNKHI	343-357	1.1072	No Tóxico
DPNFKDQVILLNKH	343-356	1.2954	No Tóxico
DPNFKDQVILLNK	343-355	1.3437	No Tóxico
DPNFKDQVILLN	343-354	1.3832	No Tóxico
PNFKDQVILLNKHIDAYK	344-361	0.4894	No Tóxico
PNFKDQVILLNKHIDAY	344-360	0.6763	No Tóxico
PNFKDQVILLNKHIDA	344-359	0.8664	No Tóxico
PNFKDQVILLNKHI	344-357	0.9263	No Tóxico
PNFKDQVILLNKH	344-356	1.1181	No Tóxico
PNFKDQVILLNK	344-355	1.1559	No Tóxico
NFKDQVILLNKHIDAYKT	345-362	0.4646	No Tóxico
NFKDQVILLNKHIDAY	345-360	0.5719	No Tóxico
NFKDQVILLNKHIDA	345-359	0.7693	No Tóxico
FKDQVILLNKHIDA	346-359	0.4584	No Tóxico
KDQVILLNKHIDA	347-359	0.4381	No Tóxico

KDQVILLNKHID	347-358	0.5548	No Tóxico
KHIDAYKTFPPTEPKKD	355-371	0.402	No Tóxico
KHIDAYKTFPPTEPKK	355-370	0.4118	No Tóxico
HIDAYKTFPPTEPKK	356-370	0.4004	No Tóxico
QRQKKQQTVTLLPAAD	384-399	0.6161	No Tóxico
QRQKKQQTVTLLPAA	384-398	0.6824	No Tóxico
RQKKQQTVTLLPAAD	385-399	0.6555	No Tóxico
RQKKQQTVTLLPAA	385-398	0.7312	No Tóxico
QKKQQTVTLLPAAD	386-399	0.7176	No Tóxico
QKKQQTVTLLPAA	386-398	0.8106	No Tóxico
QKKQQTVTLLPA	386-397	0.8434	No Tóxico
KKQQTVTLLPAAD	387-399	0.6767	No Tóxico
KKQQTVTLLPAA	387-398	0.7764	No Tóxico
KQQTVTLLPAAD	388-399	0.5778	No Tóxico

Tabla 43. Epítomos para células T CD8+ seleccionados en la proteína S.

Epítomo	Posición	Antigenicidad	Toxicidad
AYYVGYLQPRTF	264-275	0.8536	No Tóxico
AYYVGYLQPR	264-273	1.3309	No Tóxico
YYVGYLQPRTFLL	265-277	0.6995	No Tóxico
YYVGYLQPRTF	265-275	0.8985	No Tóxico
YYVGYLQPR	265-273	1.4692	No Tóxico
YVGYLQPRTFLL	266-276	0.6882	No Tóxico
YVGYLQPRTF	266-275	1.0511	No Tóxico
VGYLQPRTFLL	267-276	0.797	No Tóxico
VGYLQPRTF	267-275	1.2441	No Tóxico
GYLQPRTFLL	268-277	0.7535	No Tóxico
GYLQPRTFLL	268-276	0.6082	No Tóxico
GYLQPRTF	268-275	1.1451	No Tóxico
YLQPRTFLL	269-277	0.4532	No Tóxico
LQPRTFLL	270-277	0.6653	No Tóxico
EKGIYQTSNFR	309-319	0.4125	No Tóxico
GIYQTSNFR	311-319	0.538	No Tóxico
FRVQPTESEI	318-326	0.9396	No Tóxico
RVQPTESEIV	319-327	0.599	No Tóxico
ASVYAWNRKR	348-357	0.6453	No Tóxico
ASVYAWNRK	348-356	0.5788	No Tóxico
SVYAWNRKR	349-357	0.765	No Tóxico
VYAWNRKRI	350-358	0.5003	No Tóxico
KCYGVSPTK	378-386	1.4199	No Tóxico
CYGVSPTKL	379-387	1.4263	No Tóxico
YGVSPTKL	380-387	1.4986	No Tóxico
GVSPKLNLDL	381-390	1.6525	No Tóxico
VSPKLNLDL	382-390	1.461	No Tóxico
SPTKLNLDL	383-392	2.2412	No Tóxico

SPTKLNLD	383-390	1.0358	No Tóxico
FRKSNLKPF	456-464	0.628	No Tóxico
RKSNLKPFER	457-466	0.8187	No Tóxico
KSNLKPFER	458-466	0.949	No Tóxico
DIADTTDAVR	568-577	0.8147	No Tóxico
DIADTTDAV	568-576	1.0904	No Tóxico
IADTTDAV	569-576	0.7546	No Tóxico
AVRDPQTLEIL	575-585	0.5643	No Tóxico
AVRDPQTLEI	575-584	0.7083	No Tóxico
VRDPQTLEIL	576-585	0.5446	No Tóxico
VRDPQTLEI	576-584	0.7025	No Tóxico
RDPQTLEIL	577-585	0.7101	No Tóxico
RDPQTLEI	577-584	0.9248	No Tóxico
LEILDITPC	582-590	1.639	No Tóxico
EILDITPCSF	583-592	1.2698	No Tóxico
ILDITPCSF	584-592	1.1835	No Tóxico
LDITPCSF	585-592	1.6304	No Tóxico
VAIHADQLTPTW	622-633	0.4051	No Tóxico
HADQLTPTWRVY	625-636	0.4457	No Tóxico
HADQLTPTWR	625-634	0.6323	No Tóxico
ADQLTPTWRV	626-635	0.5883	No Tóxico
ADQLTPTW	626-633	0.4563	No Tóxico
DQLTPTWRVY	627-636	0.6489	No Tóxico
DQLTPTWRV	627-635	0.7445	No Tóxico
QLTPTWRVY	628-636	1.2119	No Tóxico
TPTWRVYST	630-638	0.4605	No Tóxico
STGSNVFQTR	637-646	0.5411	No Tóxico
TGSNVFQTR	638-646	0.6003	No Tóxico
YTMSLGAENSVAY	695-707	0.5713	No Tóxico
TMSLGAENSVAY	696-707	0.5896	No Tóxico
MSLGAENSVAY	697-707	0.7925	No Tóxico
SLGAENSVAY	698-707	0.6175	No Tóxico
LGAENSVAY	699-707	0.4173	No Tóxico
TEILPVSMK	724-733	1.416	No Tóxico
TEILPVSM	724-732	1.5697	No Tóxico
TEILPVSM	724-731	1.2728	No Tóxico
EILPVSMK	725-733	1.6842	No Tóxico
LPVSMKTSV	727-736	1.2574	No Tóxico
LPVSMKTS	727-735	1.555	No Tóxico
VSMKTSV	729-736	1.2114	No Tóxico
QSAPHGVVFLHVTY	1054-1067	0.8073	No Tóxico
APHGVVFLHVTY	1056-1067	0.8472	No Tóxico
APHGVVFLHVT	1056-1066	0.7233	No Tóxico
HGVVFLHVTY	1058-1067	0.9811	No Tóxico
GVVFLHVTY	1059-1067	1.4104	No Tóxico
VVFLHVTYV	1060-1068	1.5122	No Tóxico

VVFLHVTY	1060-1067	1.4573	No Tóxico
TVYDPLQPEL	1136-1145	0.4913	No Tóxico
VYDPLQPEL	1137-1145	0.4525	No Tóxico
VYDPLQPE	1137-1144	0.5148	No Tóxico
YDPLQPEL	1138-1145	0.8519	No Tóxico
DEDDSEPVL	1257-1265	0.5104	No Tóxico
DSEPVLKGVKLHY	1260-1272	1.2964	No Tóxico
DSEPVLKGV	1260-1268	1.1314	No Tóxico
SEPVLKGVKLHY	1261-1272	1.1738	No Tóxico
SEPVLKGVKL	1261-1270	1.0654	No Tóxico
SEPVLKGV	1261-1268	0.8877	No Tóxico
EPVLKGVKLHY	1262-1272	1.3073	No Tóxico
EPVLKGVKL	1262-1270	1.2301	No Tóxico
VLKGVKLHY	1264-1272	1.2378	No Tóxico

Tabla 44. Epítomos para células T CD4+ seleccionados en la proteína S.

Epítomo	Posición	Antigenicidad	Toxicidad
VEKGIYQTSNFRVQPTES	308-325	0.7311	No Tóxico
VEKGIYQTSNFRVQPTE	308-324	0.8296	No Tóxico
VEKGIYQTSNFRVQPT	308-323	0.8907	No Tóxico
VEKGIYQTSNFRVQP	308-322	0.7959	No Tóxico
VEKGIYQTSNFRVQ	308-321	0.5508	No Tóxico
EKGIYQTSNFRVQPTESI	309-326	0.6643	No Tóxico
EKGIYQTSNFRVQPTES	309-325	0.749	No Tóxico
EKGIYQTSNFRVQPTE	309-324	0.8559	No Tóxico
EKGIYQTSNFRVQPT	309-323	0.9243	No Tóxico
EKGIYQTSNFRVQP	309-322	0.8229	No Tóxico
EKGIYQTSNFRVQ	309-321	0.5544	No Tóxico
KGIYQTSNFRVQPTESIV	310-327	0.6139	No Tóxico
KGIYQTSNFRVQPTESI	310-326	0.6741	No Tóxico
KGIYQTSNFRVQPTES	310-325	0.7664	No Tóxico
KGIYQTSNFRVQPTE	310-324	0.8838	No Tóxico
KGIYQTSNFRVQPT	310-323	0.962	No Tóxico
KGIYQTSNFRVQP	310-322	0.8542	No Tóxico
KGIYQTSNFRVQ	310-321	0.559	No Tóxico
GIYQTSNFRVQPTESIVR	311-328	0.4574	No Tóxico
GIYQTSNFRVQPTESIV	311-327	0.6888	No Tóxico
GIYQTSNFRVQPTESI	311-326	0.7593	No Tóxico
GIYQTSNFRVQPTES	311-325	0.8668	No Tóxico
GIYQTSNFRVQPTE	311-324	1.0038	No Tóxico
GIYQTSNFRVQPT	311-323	1.1031	No Tóxico
GIYQTSNFRVQP	311-322	0.9986	No Tóxico
IYQTSNFRVQPTESIVR	312-328	0.4239	No Tóxico

IYQTSNFRVQPTESIV	312-327	0.6706	No Tóxico
IYQTSNFRVQPT	312-324	1.0141	No Tóxico
IYQTSNFRVQPT	312-323	1.1283	No Tóxico
YQTSNFRVQPTESIVR	313-328	0.4863	No Tóxico
YQTSNFRVQPTESIV	313-327	0.7563	No Tóxico
QTSNFRVQPTESIVR	314-328	0.4885	No Tóxico
QTSNFRVQPTESIV	314-327	0.7811	No Tóxico
QTSNFRVQPTESI	314-326	0.8779	No Tóxico
TSNFRVQPTESIVR	315-328	0.4901	No Tóxico
TSNFRVQPTESIV	315-327	0.813	No Tóxico
TSNFRVQPTESI	315-316	0.9265	No Tóxico
SNFRVQPTESIVR	316-328	0.4991	No Tóxico
SNFRVQPTESIV	316-327	0.8573	No Tóxico
NFRVQPTESIVR	317-328	0.6511	No Tóxico
FASVYAWNRKRISN	347-360	0.4597	No Tóxico
FASVYAWNRKRIS	347-359	0.5876	No Tóxico
ASVYAWNRKRISN	348-360	0.4355	No Tóxico
ASVYAWNRKRIS	348-359	0.5771	No Tóxico
SVYAWNRKRISN	349-360	0.4858	No Tóxico
RKSNLKPFERDISTEI	457-472	0.4298	No Tóxico
RKSNLKPFERDISTE	457-471	0.4847	No Tóxico
KSNLKPFERDISTEI	458-472	0.4014	No Tóxico
KSNLKPFERDISTE	458-471	0.4643	No Tóxico
KSNLKPFERDIST	458-470	0.444	No Tóxico
KSNLKPFERDIS	458-469	0.7736	No Tóxico
NLKPFERDISTEI	460-472	0.4346	No Tóxico
NLKPFERDISTE	460-471	0.5155	No Tóxico
DIADTTDAVRDPQT	568-581	0.5256	No Tóxico
DIADTTDAVRDPQ	568-580	0.6991	No Tóxico
DIADTTDAVRDP	568-579	0.6079	No Tóxico
VRDPQTLILDITP	576-589	1.4571	No Tóxico
QLTPTWRVYSTGSNVFQT	628-645	0.6238	No Tóxico
QLTPTWRVYSTGSNVFQ	628-644	0.6086	No Tóxico
QLTPTWRVYSTGSNVF	628-643	0.6678	No Tóxico
LTPTWRVYSTGSNVFQTR	629-646	0.7438	No Tóxico
LTPTWRVYSTGSNVFQT	629-645	0.5474	No Tóxico
LTPTWRVYSTGSNVFQ	629-644	0.5257	No Tóxico
LTPTWRVYSTGSNVF	629-643	0.5831	No Tóxico
TPTWRVYSTGSNVFQTRA	630-647	0.4094	No Tóxico
TPTWRVYSTGSNVFQTR	630-646	0.4218	No Tóxico
TPTWRVYSTGSNV	630-642	0.4207	No Tóxico
PTWRVYSTGSNVFQTRA	631-647	0.4637	No Tóxico
PTWRVYSTGSNVFQTR	631-646	0.4819	No Tóxico

PTWRVYSTGSNV	631-642	0.5085	No Tóxico
TWRVYSTGSNVFQTRA	632-647	0.4092	No Tóxico
TWRVYSTGSNVFQTR	632-646	0.4253	No Tóxico
WRVYSTGSNVFQTRA	633-647	0.4133	No Tóxico
WRVYSTGSNVFQTR	633-646	0.4314	No Tóxico
FGGFNFSQILPDPSKPSK	797-814	0.4125	No Tóxico
FGGFNFSQILPDPSKPS	797-813	0.4125	No Tóxico
FGGFNFSQILPDPSK	797-811	0.4404	No Tóxico
FGGFNFSQILPDPS	797-810	0.5927	No Tóxico
FGGFNFSQILPDP	797-809	0.6564	No Tóxico
FGGFNFSQILPD	797-808	0.6291	No Tóxico
GGFNFSQILPDPSKPSKR	798-815	0.5239	No Tóxico
GGFNFSQILPDPS	798-810	0.5553	No Tóxico
GGFNFSQILPDP	798-809	0.6221	No Tóxico
GFNFSQILPDPSKPSKRS	799-816	0.6974	No Tóxico
GFNFSQILPDPSKPSKR	799-815	0.5358	No Tóxico
GFNFSQILPDPS	799-810	0.5713	No Tóxico
FNFSQILPDPSKPSKRS	800-816	0.7043	No Tóxico
FNFSQILPDPSKPSKR	800-815	0.5304	No Tóxico
NFSQILPDPSKPSKRS	801-816	0.505	No Tóxico
SQILPDPSKPSKRS	803-816	0.4761	No Tóxico
QILPDPSKPSKRS	804-816	0.5265	No Tóxico
QSAPHGVVFLHVT	1054-1066	0.7062	No Tóxico
SCCKFDEDDSEPVLKGVK	1252-1269	0.5409	No Tóxico
SCCKFDEDDSEPVLKGV	1252-1268	0.482	No Tóxico
SCCKFDEDDSEPVLKG	1252-1267	0.4347	Tóxico
CCKFDEDDSEPVLKGVK	1253-1269	0.6515	No Tóxico
CCKFDEDDSEPVLKGV	1253-1268	0.5963	No Tóxico
CCKFDEDDSEPVLKG	1253-1267	0.5557	Tóxico
CCKFDEDDSEPV	1253-1264	0.4603	Tóxico
CKFDEDDSEPVLKGVKL	1254-1270	0.5995	No Tóxico
CKFDEDDSEPVLKGVK	1254-1269	0.5198	No Tóxico
CKFDEDDSEPVLKGV	1254-1268	0.4486	No Tóxico
KFDEDDSEPVLKGVKL	1255-1270	0.7015	No Tóxico
KFDEDDSEPVLKGVK	1255-1269	0.6236	No Tóxico
KFDEDDSEPVLKGV	1255-1268	0.5556	No Tóxico
KFDEDDSEPVLKG	1255-1267	0.5066	No Tóxico

Tabla 45. Epítomos para células B sugeridos por al menos dos programas que contienen el residuo T376 en la proteína S.

Posición	Epítopo	Bepipred 2.0	IEDB	iBCE-EL	FBCPRED (%)	Valor Antigenicidad	Residuos Escondidos (%)
376-387	TFKCYGVSP TKL	✓	✓	0.85	-	1.4293	50.00
376-388	TFKCYGVSP TKLN	✓	✓	0.85	-	1.529	46.15
376-389	TFKCYGVSP TKLND	✓	✓	0.85	-	1.4312	42.86
376-390	TFKCYGVSP TKLN DL	✓	✓	0.85	-	1.4626	46.67
376-391	TFKCYGVSP TKLND LC	✓	✓	0.85	-	1.6787	50.00
376-392	TFKCYGVSP TKLND LCF ¹	✓	✓	0.85	-	1.9808	52.94
376-393	TFKCYGVSP TKLND LCF T	✓	✓	0.85	-	2.0228	50.00

Nota: ¹ Epítopo que no cumple el criterio de porcentaje de residuos escondidos

Anexo B. Construcción genética en los vectores pAlgevir y pBI121 y clonación

Los genes sintéticos enviados por la compañía Gene Universal, se insertaron en los vectores pAlgevir y pBI121 basándose en la metodología de Márquez-Escobar et al. (2018). Primeramente, células de *E. coli* previamente transformadas con otros constructos del vector pAlgevir y pBI121 se cultivaron en medio agar LB con 50 mg/L de kanamicina a 37°C por 24hrs. Las colonias aisladas se pasaron a medio líquido LB con 50 mg/L de kanamicina y se incubaron a 37°C por 24 h. Se realizó la extracción de ADN plasmídico con el kit de extracción de ADN PureLink® Quick Plasmid Miniprep Kit (K210011 Invitrogen), siguiendo la metodología del fabricante. Para corroborar la calidad de las extracciones se realizó una electroforesis en gel de agarosa al 1.2% a 120 V por 30 min y observado en el fotodocumentador Gel Doc® EZ Imager de BioRad.

Posteriormente, se realizó la digestión con las enzimas de digestión *Sma*I (No. Catálogo: R0141S) y *Bam*HI (R01361S) de New England Biolabs (NEB) para el vector pAlgevir y con *Sma*I y *Sac*I-HF (R3156M NEB) para el vector pBI121. Para pAlgevir se realizó la digestión con 400 µg de ADN plasmídico, 2 µL de buffer Cutsmart (B7204S NEB), 0.4 µL (8 U) de enzima *Sma*I y aforado a 20 µL con agua estéril. La reacción se incubó por 15min a temperatura ambiente y se inactivó la enzima a 65°C por 20 min. Posteriormente, se añadió la misma cantidad de enzima *Bam*HI y se incubó a 37°C por 15 min. Para el vector pBI121 se añadió 0.1 µL de las enzimas *Sac*I-HF y *Sma*I y

la reacción se incubó a temperatura ambiente por 15 min y luego a 37°C por 15 min. Se verificó el correcto corte con una electroforesis en gel de agarosa al 1.5% a 120 V por 30 min con el marcador de 1Kb Plus ADN (10787-018 Invitrogen), con el buffer de carga Gel Loading Dye Purple (6x) (B7024S NB) y observado en el Gel Doc® EZ Imager de BioRad. Una vez comprobada la correcta digestión, se procedió a realizar el corte de la banda correspondiente al vector pAlgevir (<17,000 pb) y pBI121 (<14,000 pb) con la ayuda de un transiluminador y bisturí. El gel recuperado fue purificado con el kit Wizard® SV Gel and PCR Clean-Up System (A9281 Promega) siguiendo las instrucciones del fabricante y recuperando 40µL de ADN diluido en el buffer de elución. Finalmente, se corroboró la purificación con una electroforesis de calidad comparando la muestra obtenida con ADN del mismo vector sin digerir en un gel de agarosa al 1.5% a 120V por 30min.

Los genes sintéticos se recibieron dentro del vector pUC57. Se resuspendieron en 10 µL de agua estéril. Para la propagación de los genes sintéticos se transformaron células competentes de *E. coli* TOP10 (Invitrogen™ C404010), Para esto se tomaron 3 µL de la suspensión de cada gen para insertarse en las células de *E. coli* por choque térmico. Los genes se añadieron a tubos con las células de *E. coli* conservadas en glicerol a -80°C, previamente descongeladas en baño de hielo. La mezcla se incubó por 15 min en baño de hielo, luego por 1 min a 42°C y finalmente se incubó 10min más en el baño de hielo. A esto se le añadieron 500 µL de caldo LB y se incubó a 37°C y 150 rpm por 4hrs. Las células transformadas se sembraron en medio agar LB con 50mg/mL de ampicilina y se incubaron a 37°C por 24 h; las colonias resistentes se propagaron en medio LB líquido con 50 mg/L de ampicilina y se incubaron a 37°C y 150 rpm por 24 h. El ADN plasmídico de las células de *E. coli* se extrajo con el kit ADN PureLink® Quick Plasmid Miniprep Kit (K210011 Invitrogen) siguiendo las instrucciones del fabricante y se realizó una electroforesis en gel de agarosa al 1.2% para corroborar la calidad. El ADN plasmídico se digirió, y se corroboró la digestión con una electroforesis en gel de agarosa al 1.2% a 120 V por 30 min y observando la banda del gen de interés en el fotodocumentador Gel Doc® EZ Imager de BioRad. Posteriormente se recuperó la banda correspondiente a cada gen y se purificó con los mismos procedimientos previos para la obtención de los vectores pAlgevir (Fig. 21) y pBI121 (Fig. 22), pero con la mitad del volumen de reacción (10 µL).

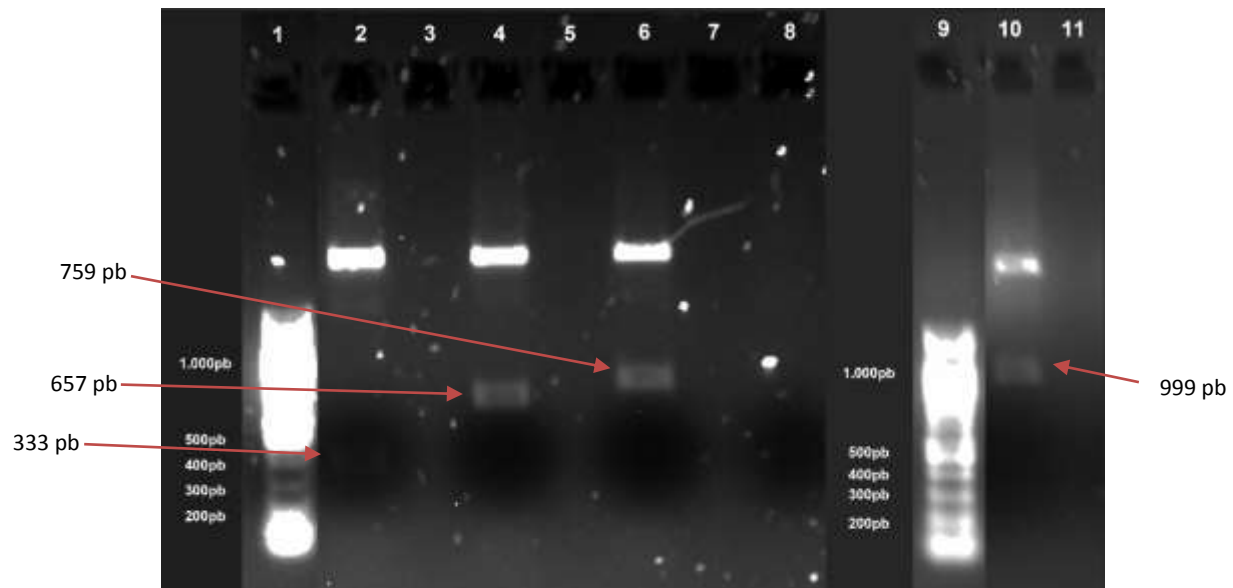


Figura 21. Digestión genes CIB-GIVac en pUC57 de células *E. coli* TOP 10 transformadas para su inserción en pAlgevir. Tamaño de fragmentos: SARS2_E = 333pb; SARS2_M = 657pb; SARS2_N = 759pb; SARS2_S = 999pb. Carril 1: Marcador 100pb; Carril 2: SARS2_E; Carril 4: SARS2_M; Carril 6: SARS2_N; Carril 8: Mezcla de reacción 1; Carril 9: Marcador 100pb; Carril 10: SARS2_S; Carril 11: Mezcla de reacción 2.

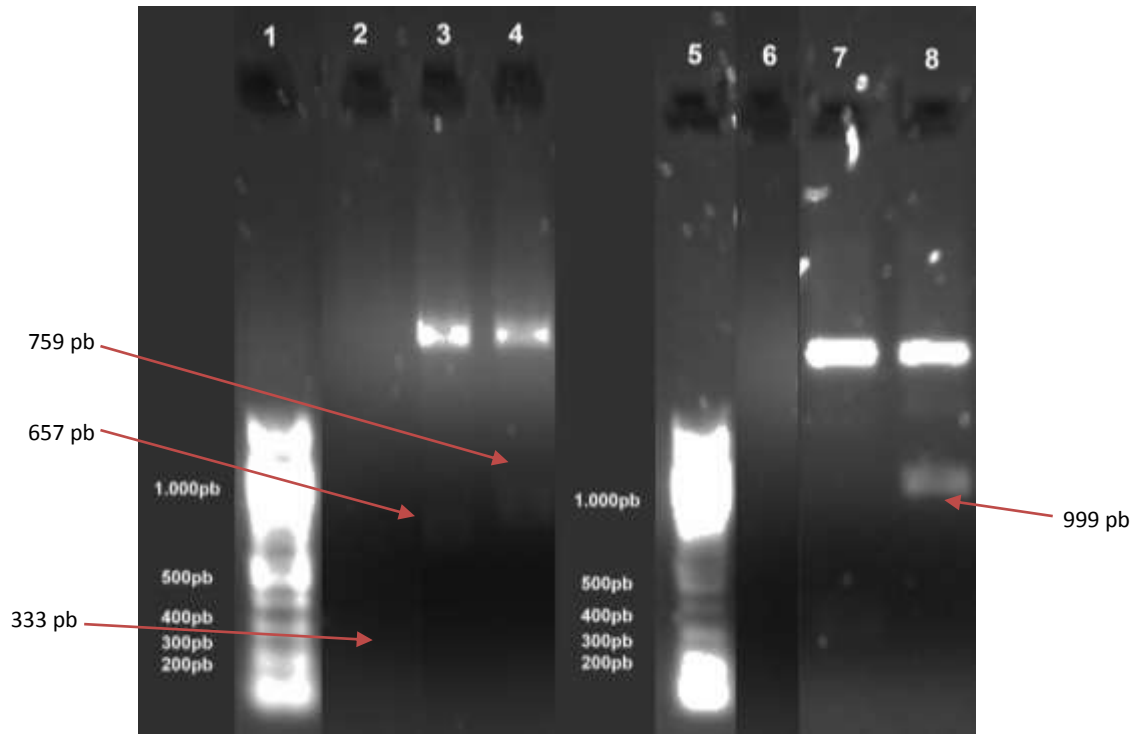


Figura 22. Digestión genes CIB-GIVac en pUC57 de células de *E. coli* TOP 10 transformadas para su inserción en pBI121. Tamaño de fragmentos: SARS2_E = 333pb; SARS2_M = 657pb; SARS2_N = 759pb; SARS2_S = 999pb. Carril 1: Marcador 100pb; Carril 2: Mezcla de reacción1; Carril 3: SARS2_M; Carril 4: SARS2_N; Carril 5: Marcador 100pb; Carril 6: Mezcla de reacción 2; Carril 7: SARS2_E; Carril 8: SARS2_S.

La reacción de ligación de los vectores con los genes obtenidos se realizó usando una relación 2:1 (gen:vector), con 2 μ L de buffer para ligasa (B0202S NEB) 0.5 μ L de enzima Hi-T4[®] DNA Ligase (M2622L NEB) y un volúmen final de 10 μ L con agua libre de endonucleasas (R0581 ThermoFisher Scientific), por 24 h a 4°C. La transformación de células competentes de *E. coli* se realizó siguiendo la misma metodología usada para la propagación de los genes sintéticos en pUC57, pero con la totalidad del volumen de la reacción de ligación.

El ADN plasmídico de las células de *E. coli* TOP 10 transformadas se extrajo con el kit ADN PureLink[®] Quick Plasmid Miniprep Kit (K210011 Invitrogen) siguiendo las instrucciones del fabricante y se realizó una electroforesis en gel de agarosa al 1.2% para corroborar la calidad (Fig. 23). Las clonas recombinantes se detectaron por PCR usando los primers específicos previamente diseñados. Las mezclas de reacción se realizaron con 2.5 μ L de buffer 10X PCR Rxn (EDWB1 Invitrogen), 1.5 μ L de MgCl₂ 25mM (Y02016b Invitrogen), 0.6 μ L de cada primer

específico (1.2 μL total), 0.25 μL de Platinum® *Taq* DNA Polymerase (10966-018 Invitrogen), 0.5 μL de dNTPs 10 mM (100004893 Invitrogen), 100 ng de ADN y se aforó a 25 μL con agua libre de endonucleasas (R0581 ThermoFisher Scientific). Se estandarizó el programa de PCR con los siguientes parámetros: (1) desnaturalización inicial de 5 min a 94°C, (2) 35 ciclos de 30 s de desnaturalización a 94°C, 40 s de alineamiento a 57°C y 50s de elongación a 72°C y (3) 5 min de elongación final a 72°C. Finalmente, la amplificación se corroboró con una electroforesis en un gel de agarosa al 1.5% a 120V por 30min, con el marcador 100bp DNA Ladder (N3231S NEB) y observado en el fotodocumentador Gel Doc® EZ Imager de BioRad (Fig. 24).

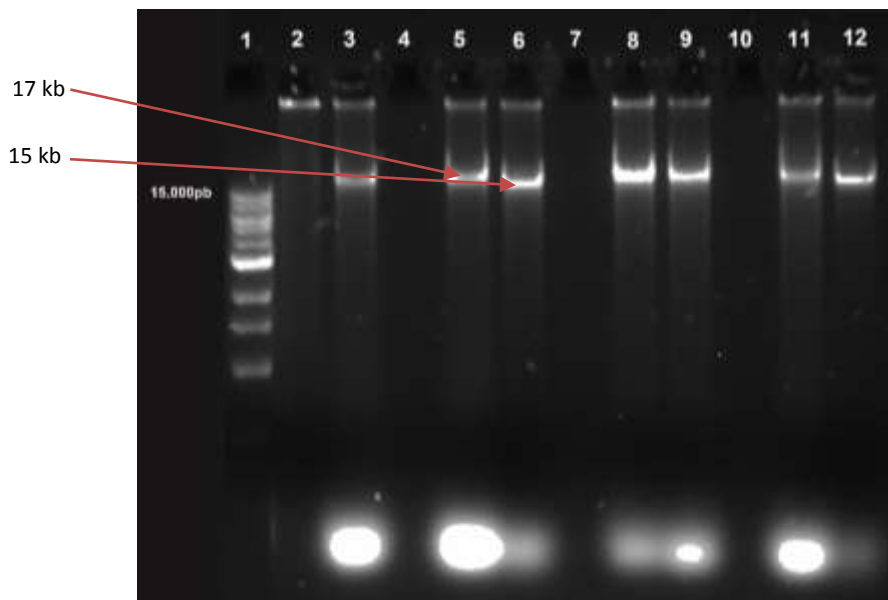


Figura 23. Calidad de ADN de células TOP 10 de *E. coli* transformadas con genes CIB-GIVac en vectores pAlgevir y pBI121. Tamaños esperados: pAlgevir > 17,000; pBI121 > 15,000. Carril 1: Marcador 1Kb Plus; Carril 2: SARS2_E Algevir; Carril 3: SARS2_E pBI121; Carril 5: SARS2_M Algevir; Carril 6: SARS2_M pBI121; Carril 8: SARS2_N Algevir; Carril 9: SARS2_N pBI121; Carril 11: SARS2_S Algevir; Carril 12: SARS2_S pBI121.

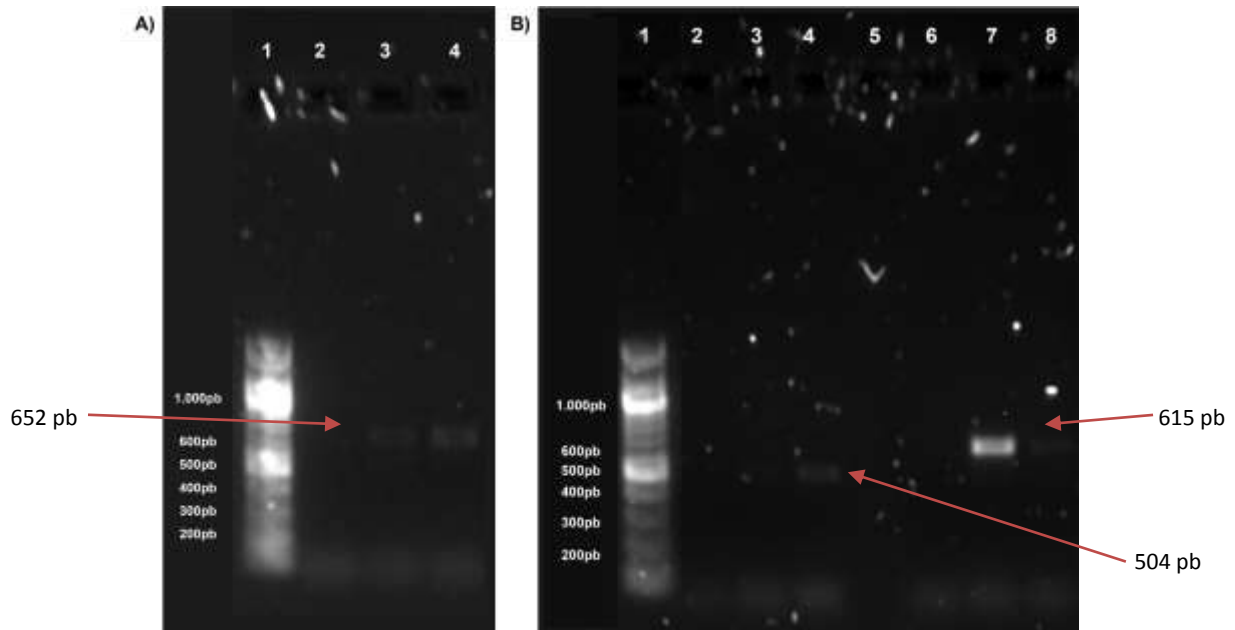


Figura 24. PCR para confirmación de cultivos de *E. coli* TOP 10 transformada. A) SARS2_N. Tamaño de amplicón: 652pb. Carril 1: Marcador 100pb; Carril 2: Mezcla de reacción; Carril 3: Muestra S2NA (Algevir); Carril 4: Muestra S2NB (pBI121). B) SARS2_M y SARS2_S. Tamaños de amplicón: 504pb (SARS2_M) y 615pb (SARS2_S) . Carril 1: Marcador 100pb; Carril 2: Mezcla de reacción M; Carril 3: Muestra S2MA (pAlgevir); Carril 4: Muestra S2MB (pBI121); Carril 5: Blanco; Carril 6: Mezcla de reacción S; Carril 7: Muestra S2SA (Algevir); Carril 8: Muestra S2SB (pBI121).

Las muestras con amplicones positivos se mandaron secuenciar a la empresa Genewiz® (<https://www.genewiz.com/>). Los resultados de la secuenciación fueron comprobados mediante alineamiento de secuencias en la herramienta Nucleotide BLAST del NCBI (https://blast.ncbi.nlm.nih.gov/Blast.cgi?PROGRAM=blastn&PAGE_TYPE=BlastSearch&BLAST_SP_EC=&LINK_LOC=blasttab&LAST_PAGE=blastp) contra las secuencias de los genes sintéticos (Tabla 46).

Tabla 46. Secuenciación de productos de PCR obtenidos de ADN de células de *E. coli* TOP 10 transformadas con los vectores pBI121 y pAlgevir.

Gen en Algevir	Porcentaje de identidad	Gen en PBI121	Porcentaje de identidad
SARS2_S	99.08	SARS2_S	99.8
SARS2_N	99.47	SARS2_N	99.66
SARS2_M	97.36	SARS2_M	97.52

Finalmente, para la transformación de *Agrobacterium tumefaciens*, se añadieron 200 µg/µL de ADN plasmídico de cada constructo y 50 µL de cultivo de células competentes de *A. tumefaciens* (cepa GV3103) a una celda de electroporación Gene Pulser/MicroPulser Electroporation Cuvettes de 0.1 cm (1652083 Biorad), se colocó en un electroporador BioRad configurado para bacterias y se realizaron 2 pulsos de 1.8 V. Las células fueron recuperadas con 400 µL de medio LB e incubadas a 28°C y 150 rpm por 4 h en oscuridad. Las células se concentraron a 8,000 rpm por 10 min y se descartó el sobrenadante. Los pellets de células se suspendieron en 100 µL de medio LB y se sembraron en placas de agar LB al 1.5% con 50 mg/L de kanamicina y se incubaron a 28°C por 48-72 h en oscuridad. Las colonias diferenciadas se pasaron a medio LB líquido con 50 mg/L de kanamicina y se incubaron a 28°C y 150rpm por 48-72 h.

Como se observa previamente, se han corroborado los genes SARS2_M, SARS2_N y SARS2_S insertados tanto en pAlgevir como en pBI121 por PCR y secuenciación. Dichos genes han sido insertados en células competentes de *A. tumefaciens* (cepa GV3101) por electroporación y se han cultivado en medio LB líquido las colonias aisladas. Para estos cultivos está pendiente la confirmación. En cuanto a SARS2_E, no se logró confirmar la transformación de *E. coli*.

Anexo C. Otros productos y participaciones



Figura 25. Participaciones Inmuno-Peques. A) Taller “COVID-19: ¿En dónde vamos con las variantes?” en la Semana de Seguridad Laboral 2022 del CIBNOR. B) 1er Taller virtual de actualización para docentes “Vacunas: Un enfoque hacia las vacunas contra COVID-19” (39 horas). C) 2do Taller virtual de actualización para docentes “Vacunas y consejos prácticos para prevenir la COVID-19”.



Figura 26. Artículo de divulgación “¿Podríamos anticiparnos a otras pandemias? Una Salud: Una estrategia integral de prevención y control”. Coautor en revista “Recursos Naturales y Sociedad”, 2021. Vol. 7 (3): 129-145. <https://10.18846/renaysoc.2021.07.07.03.0010>



Figura 27. Participación semanas de posgrado en Baja California Sur. A) XX Semana de Posgrado (2021). B) XXI Semana de Posgrado (2022).



Figura 28. Curso “Requisitos para el registro sanitario de nuevos medicamentos (vacunas) ante COFEPRIS”. Curso en línea de 20 h, 2021.