

UNIVERSIDAD DEL VALLE DE GUATEMALA
Facultad de Ciencias y Humanidades



**Análisis *in silico* de péptidos estructurales de
SARS-CoV-2. Estudio de prueba de concepto para el
diseño de pruebas antígeno-anticuerpo**

Trabajo de graduación en modalidad de Tesis presentado por
Sebastián Tobar Medrano
para optar al grado académico de Licenciado en Biotecnología Molecular

Guatemala,
2021

UNIVERSIDAD DEL VALLE DE GUATEMALA
Facultad de Ciencias y Humanidades

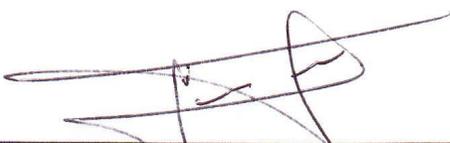


**Análisis *in silico* de péptidos estructurales de
SARS-CoV-2. Estudio de prueba de concepto para el
diseño de pruebas antígeno-anticuerpo**

Trabajo de graduación en modalidad de Tesis presentado por
Sebastián Tobar Medrano
para optar al grado académico de Licenciado en Biotecnología Molecular

Guatemala,
2021

Vo.Bo.:

(f) 
MSc. Julio Roberto Matute

Tribunal Examinador:

(f) 
MSc. Julio Roberto Matute

(f) 
MSc. Stefany Argueta

(f) 
Dra. Dalva Lau

Fecha de aprobación: Guatemala, 6 de diciembre de 2021.

El presente trabajo de graduación nació de la idea de que, con el apoyo de la bioinformática y el avance de los métodos de transformación de células, pueden plantearse proyectos viables para el desarrollo de productos biotecnológicos de alto impacto en Guatemala. Aprovechando la relevancia de la coyuntura actual, se planteó una propuesta orientada hacia la validación de epítomos de SARS-CoV-2 con posible uso en diagnósticos; sin embargo, este proyecto puede extrapolarse hacia otros enfoques que sean de igual o mayor importancia para el país.

Deseo expresar mis agradecimientos a mi asesor, Julio Matute, y a Pamela Pennington por haberme orientado durante el desarrollo de este proyecto. Asimismo, agradezco a Stefany Argueta por haber proveído las cepas y plásmidos que permitieron plantear esta propuesta. A Daniel Reyes y a Rodrigo Mejicanos por su apoyo con la preparación y ejecución de la parte experimental del proyecto; su ayuda fue muy valiosa y espero esta experiencia haya sido enriquecedora para su desarrollo personal.

Finalmente, quiero expresar mi profunda gratitud a mis padres por su apoyo incondicional en todos estos años de carrera. Todo el esfuerzo, conocimientos y logros obtenidos durante esta etapa, son un reflejo de sus enseñanzas y sacrificios que me han permitido llegar hasta este punto. Mención especial a mi persona, ya que fue a través de mi propio esfuerzo y dedicación lo que me permitió superar mis metas personales en un año lleno de retos... ¡Gracias Totales!

Prefacio	III
Lista de figuras	VI
Lista de cuadros	VII
Resumen	VIII
Abstract	IX
1. Introducción	1
2. Objetivos	2
2.1. Objetivo general	2
2.2. Objetivos específicos	2
3. Justificación	3
4. Marco teórico	4
4.1. SARS-CoV-2 y la pandemia de COVID-19	4
4.1.1. Estructura del SARS-CoV-2	4
4.2. Inmunidad adaptativa ante el SARS-CoV-2	6
4.3. Regiones antigénicas	6
4.4. Estrategias de control y diagnóstico de SARS-CoV-2	8
4.4.1. Desarrollo de pruebas rápidas	10
4.5. Bioinformática para evaluación y predicción de epítomos	11
4.6. Producción de proteínas recombinantes	11
4.7. Transformación de células	12
4.7.1. Plásmidos como vectores de expresión	12
4.7.2. Diseño de vectores de expresión	12
4.7.3. Sistemas de expresión	13
4.8. Levaduras como sistema de expresión	14
4.8.1. Tipos de plásmido para expresión en levaduras	14
5. Metodología	16
5.1. Identificación de epítomos lineales de SARS-Cov2 reportados con una alta probabilidad de generación de anticuerpos específicos	18
5.1.1. <i>Cluster Epitope Analysis</i>	18

5.1.2.	Predicción de regiones antigénicas basadas en accesibilidad al solvente	18
5.1.3.	Predicción de epítomos de células B	19
5.1.4.	Predicción de epítomos de células T	19
5.1.5.	Caracterización fisicoquímica de los péptidos	19
5.1.6.	Cobertura poblacional	19
5.1.7.	Alineamiento de proteínas	20
5.2.	Diseño <i>in silico</i> del vector de expresión	20
5.2.1.	Optimización de codones en <i>S.cerevisiae</i> (S228C)	20
5.2.2.	Selección del vector de expresión	20
5.2.3.	Ensamblaje del diseño en el mapa del vector	21
5.3.	Evaluación de SOP de transformación para <i>Saccharomyces cerevisiae</i>	21
5.3.1.	Preparación de células electrocompetentes	21
5.3.2.	Electroporación	22
5.3.3.	Visualización de transformantes	22
6.	Resultados	23
6.1.	Proteína estructurales (M,N y S), poseen mayor número de ensayos positivos	24
6.2.	Secuencias consenso poseen elevada accesibilidad al solvente	25
6.3.	Secuencias consenso poseen alta probabilidad de formar parte de epítomos de células B	26
6.4.	Secuencias de epítomos consenso y epítomos verdaderos demuestran antigenicidad frente a predictor de epítomos de células T	29
6.5.	Epítomos propuestos muestran una amplia cobertura para alelos HLA en población guatemalteca	30
6.6.	Epítomos candidatos se encuentran en regiones altamente conservadas entre variantes de SARS-CoV-2	31
6.7.	Diseños teóricos de los vectores se presentan con un enfoque de proteínas de fusión y constructos policistrónicos	32
6.8.	Obtención de transformantes mediante SOP de electroporación	33
7.	Análisis de resultados	36
8.	Conclusiones	39
9.	Recomendaciones	40
10.	Bibliografía	41
	Referencias	41
11.	Anexos	47

Lista de figuras

1.	Representación gráfica de las proteínas estructurales y accesorias del SARS-CoV-2	5
2.	Modelo 3D de la proteína estructural S de SARS-CoV-2	6
3.	Representación gráfica de epítomos lineales y conformacionales	7
4.	Concentraciones de anticuerpos séricos IgG e IgM durante el transcurso de la infección por SARS-CoV-2	10
5.	Esquema simplificado del funcionamiento de una prueba de flujo lateral convencional	10
6.	Metodología del análisis bioinformático para la selección de epítomos candidatos con alta probabilidad antigénica	17
7.	Visualización gráfica de los grupos de epítomos y principales secuencias consenso	24
8.	Regiones antigénicas de proteínas S,N y M con alta accesibilidad al solvente	25
9.	Resumen de puntajes del análisis Kolaskar y Tongaokar	26
10.	Representación gráfica de los puntajes de antigenicidad de Kolaskar y Tongaokar para la secuencia de la glucoproteína S de SARS-CoV-2	26
11.	Representación gráfica de los puntajes de antigenicidad de Kolaskar y Tongaokar para la secuencia de la nucleoproteína de SARS-CoV-2	27
12.	Representación gráfica de los puntajes de antigenicidad de Kolaskar y Tongaokar para la secuencia de la proteína de membrana de SARS-CoV-2	27
13.	Inmunogenicidad de epítomos de células T y descripciones fisicoquímicas de los péptidos candidatos para su uso en métodos de diagnóstico	29
14.	Probabilidad de cobertura de los epítomos candidatos en la población guatemalteca	30
15.	Alineamiento global de secuencias de las proteínas S, N y M de variantes de SARS-CoV-2	31
16.	Diagrama general del vector HO-Poly-KanMX4-HO y el casete de integración	32
17.	Diagrama del casete de integración HO-Poly-KanMX4-HO y constructos propuestos para la producción de epítomos	33
18.	Crecimiento de transformantes de S228C con el plásmido pKT175	33
19.	Observación de transformantes por microscopía de fluorescencia	34
20.	Re-aislamiento de transformantes	35
21.	Modelo 3D de regiones con alta sensibilidad al solvente en proteínas de SARS-CoV-2	47
22.	Secuencia del sistema de proteína de fusión epítomo-GFP	58
23.	Continuación de secuencia del sistema de proteína de fusión epítomo-GFP	58
24.	Secuencia del sistema policistrónico expresando dos epítomos con espaciador P2A del ERVB-1	59

Lista de cuadros

1.	Proteínas estructurales del SARS-CoV-2	5
2.	Comparativa de principales características de las pruebas de diagnóstico empleadas para SARS-CoV-2.	9
3.	Resumen de características de los principales sistemas de expresión.	13
4.	Vectores de expresión en levaduras	14
5.	Secuencias de epítomos de SARS-CoV-2 recuperadas de IEDB y empleadas para la generación de agrupamientos por medio del <i>Cluster analysis</i>	48
6.	Continuación 1 de secuencias de epítomos de SARS-CoV-2 recuperadas de IEDB y empleadas para la generación de agrupamientos por medio del <i>Cluster analysis</i>	49
7.	Continuación 2 de secuencias de epítomos de SARS-CoV-2 recuperadas de IEDB y empleadas para la generación de agrupamientos por medio del <i>Cluster analysis</i>	50
8.	Continuación 3 de secuencias de epítomos de SARS-CoV-2 recuperadas de IEDB y empleadas para la generación de agrupamientos por medio del <i>Cluster analysis</i>	51
9.	Continuación 4 de secuencias de epítomos de SARS-CoV-2 recuperadas de IEDB y empleadas para la generación de agrupamientos por medio del <i>Cluster analysis</i>	52
10.	Continuación 5 de secuencias de epítomos de SARS-CoV-2 recuperadas de IEDB y empleadas para la generación de agrupamientos por medio del <i>Cluster analysis</i>	53
11.	Continuación 6 de secuencias de epítomos de SARS-CoV-2 recuperadas de IEDB y empleadas para la generación de agrupamientos por medio del <i>Cluster analysis</i>	54
12.	Índice de Kolanskar y Tangaokar para regiones de aminoácidos 447-472 de la proteína S del SARS-CoV-2	54
13.	Índice de Kolanskar y Tangaokar para regiones de aminoácidos 103-139 de la proteína N del SARS-CoV-2	55
14.	Índice de Kolanskar y Tangaokar para regiones de aminoácidos 138-181 de la proteína M del SARS-CoV-2	56
15.	Secuencias consenso generadas a través del <i>epitope Cluster Analysis</i> para las proteínas S,N y M del SARS-CoV-2	57
16.	Frecuencias de alelos MHC utilizados para el análisis de cobertura poblacional	57

El desarrollo de pruebas rápidas basadas en las interacciones antígeno-anticuerpo se ha vuelto una necesidad que se ha incrementado con la pandemia de COVID-19. Actualmente son diversas entidades las que se dedican al desarrollo de estos métodos de detección. A pesar de la creciente necesidad, existe una escasa participación en Guatemala para el desarrollo de productos basados en la expresión de proteínas recombinantes, lo que representa un rezago en la creación de nuevos servicios y productos biotecnológicos. Por estos motivos, en el presente estudio se plantea una metodología bioinformática para la selección *in silico* de epítomos candidatos con potencial aplicación como método de diagnóstico. Asimismo, se propone el diseño de un vector que emplee *Saccharomyces cerevisiae* como sistema de expresión para la producción de epítomos de SARS-CoV-2, así como un SOP de transformación. Los resultados del estudio demuestran que los péptidos candidatos tienen altos índices de inmunogenicidad, conservación entre cepas y alta cobertura poblacional, por lo que resultan prometedores para su empleo como biomarcadores de diagnóstico. Adicionalmente, se demostró que es posible transformar *S. cerevisiae* con el SOP evaluado. Estos datos en conjunto presentan una propuesta viable para que en un futuro pueda ejecutarse el proceso de clonación, transformación y producción de epítomos recombinantes para que puedan utilizarse para la detección de anticuerpos séricos.

The development of rapid tests based on antigen-antibody interactions has become a need that has increased with the COVID-19 pandemic. Currently, various entities are dedicated to the development of these detection methods. Despite the growing need, there is little participation in Guatemala for the development of products based on the expression of recombinant proteins, which represents a lag in the creation of new biotechnological services and products. For these reasons, the present study proposes a bioinformatic methodology for the selection *in silico* of candidate epitopes with potential application as a diagnostic method. Likewise, the design of a vector is proposed that uses *Saccharomyces cerevisiae* as an expression system for the production of SARS-CoV-2 epitopes, as well as a transformation SOP. The results of the study show that the candidate peptides have high rates of immunogenicity, conservation between strains and a high population coverage, thus showing promise for their use as diagnostic biomarkers. In addition, it was shown that it is possible to transform *S. cerevisiae* with the SOP evaluated. These data together present a viable proposal so that in the future the process of cloning, transformation and production of recombinant epitopes to be used for the detection of serum antibodies can be carried out.

CAPÍTULO 1

Introducción

La pandemia de COVID-19, ha acrecentado exponencialmente la necesidad y demanda de desarrollar pruebas diagnósticas de bajo costo, altamente confiables y que puedan mostrar resultados en tiempos relativamente cortos (Pilarowski y Marquez, 2021). La costo efectividad y accesibilidad de las pruebas rápidas de diagnóstico, especialmente los sistemas de flujo lateral han demostrado ser una medida indispensable para poder aislar los casos positivos y así reducir las posibilidades de transmisión del virus (Seo y Kim, 2020). Adicionalmente, son una herramienta útil para llevar a cabo estudios epidemiológicos, ya que estas pruebas se basan en reacciones inmunoquímicas, específicamente en la interacción antígeno-anticuerpo. Asimismo, la categorización de epítomos inmunogénicos ha permitido el desarrollo de dichas pruebas rápidas y lograr porcentajes de especificidad y sensibilidad relativamente altos (Liu y Rusling, 2021).

Para desarrollar este tipo de productos biotecnológicos, se requieren de tecnologías de biología molecular, manipulación genética y análisis bioinformáticos. La utilización de levaduras para la producción de productos de interés biotecnológico es una actividad cada vez más extendida. Uno de los motivos es una mayor accesibilidad a las tecnologías de transformación que permiten modificar a estos microorganismos para lograr la producción de proteínas recombinantes como la producción de enzimas que tienen utilidad en diferentes industrias, producción de metabolitos, mejoramiento del valor nutricional para su utilización como suplemento alimenticio y más recientemente, la producción de antígenos para el diseño de métodos de diagnóstico (Parapouli y Hatziloukas, 2020).

En Guatemala, el desarrollo de tecnologías de diagnóstico, así como la generación de organismos recombinantes son aspectos que se encuentran rezagados respecto a otros países, a pesar de que estas tecnologías son accesibles y llevan varios años utilizándose. La escasez de antecedentes en estos temas limita considerablemente que dichas tecnologías sean empleadas de manera más rutinaria en el país. Por este motivo, en el presente estudio se propone en primer lugar una metodología para seleccionar y validar secuencias de epítomos lineales de SARS-CoV-2 por medio de diferentes análisis bioinformáticos que evalúen la antigenicidad y la cobertura poblacional de las secuencias. Seguidamente se propone un vector de expresión y dos constructos genéticos para la producción teórica de estos epítomos empleando *S.cerevisiae* como sistema de expresión. Por último se evalúa un SOP para la transformación por electroporación de *S.cerevisiae* S228C para determinar su aplicabilidad en dado caso se llevara a cabo una evaluación experimental de los constructos. La finalidad conjunta de estos objetivos es la de poder establecer antecedentes que sirvan de plataforma para desarrollar tecnologías de expresión recombinante en Guatemala.

2.1. Objetivo general

Proponer y ejecutar una metodología basada en Qamar *et al* (2020) y Paul *et al* (2021) para el análisis y selección de epítomos de SARS-CoV-2 con potencial aplicación en pruebas rápidas de diagnóstico mediante herramientas bioinformáticas de libre acceso y generar un constructo genético *in silico* para su expresión en *S.cerevisiae*.

2.2. Objetivos específicos

- Realizar un análisis bioinformático para seleccionar y validar la antigenicidad y grado de conservación de epítomos lineales de SARS-Cov-2 previamente reportados con potencial aplicación para diseño de pruebas rápidas.
- Proponer constructos genéticos que podría clonarse en un vector de expresión para la producción de los epítomos lineales en *S.cerevisiae*.
- Evaluar una versión modificada del protocolo de electroporación propuesto (Ruell, 2020), para poder ser aplicado al equipo BIO-RAD Gene Pulser II e implementarse como un método rápido y de bajo costo para la transformación de *S.cerevisiae*.

Según la OMS, el SARS-CoV-2, virus responsable de la enfermedad COVID-19 y causante de la pandemia que se vive en la actualidad, ha sido responsable de más de 110 millones de personas infectadas, de las cuales 2.44 millones han fallecido por complicaciones relacionadas con la enfermedad. Una de las estrategias tomadas desde el inicio de la pandemia para intentar mitigar una rápida expansión, ha sido el diagnóstico de los casos sospechosos para su posterior aislamiento (Guest y Sanchez, 2020). El método de detección estandarizado es la prueba de PCR que detecta regiones genómicas específicas del virus; sin embargo, por la alta demanda, baja disponibilidad y alto costo de estas, se han desarrollado pruebas rápidas de diagnóstico que se basan en la detección de anticuerpos capaces de unirse a partes antigénicas del virus en sistemas de flujo lateral (Penalver y Bottcher, 2020). La desventaja de estas pruebas es que tienen una probabilidad de falsos negativos mayor que las pruebas de PCR; es por este motivo que en caso de resultados negativos se recomienda realizar una prueba de PCR confirmatoria (Smyrlaki y Reinius, 2020).

El diseño y desarrollo de pruebas rápidas sigue siendo una necesidad para que países de escasos recursos puedan realizar diagnósticos masivos sin necesidad de contar con equipo de laboratorio tan costoso y sofisticado. Adicionalmente, resultan una herramienta bastante útil para estudios epidemiológicos y poder monitorear la evolución de contagios en una región. En la actualidad, son diversas empresas las que se dedican al desarrollo de estos métodos de detección. A pesar de la creciente necesidad, existe una escasa participación en Guatemala para el desarrollo de productos basados en la expresión de proteínas recombinantes. Esto representa un rezago en la creación de nuevas tecnologías para brindar servicios y productos biotecnológicos.

Por estos motivos, en el presente estudio se propone una aproximación bioinformática para el análisis de epítomos lineales de SARS-Cov-2 con potencial uso diagnóstico y el diseño *in silico* de un vector de expresión para la producción de epítomos en *Saccharomyces cerevisiae*. También se evaluará un protocolo de electroporación, el cual permitirá a futuro llevar a cabo ensayos de producción de epítomos para el desarrollo de pruebas rápidas. Con este proyecto se espera demostrar el proceso de diseño de un constructo genético para ser utilizado en un plásmido. Por otro lado, se mostrará la posibilidad de emplear *S. cerevisiae* para la producción de proteínas recombinantes y como con estas herramientas se puede contribuir al desarrollo de productos biotecnológicos que ofrezcan soluciones prácticas a problemáticas actuales en Guatemala.

4.1. SARS-CoV-2 y la pandemia de COVID-19

La pandemia de COVID-19 ha sido una de las adversidades ambientales más significativas de la historia humana contemporánea. La naturaleza de su origen aún no ha podido establecerse con total precisión; sin embargo, la hipótesis más aceptada sugiere que es el resultado de una zoonosis cuya fuente proviene de algún coronavirus de murciélago y un hospedero intermediario, tal como lo sustentan algunos estudios con evidencia en rinolófidios de reservorios de coronavirus con genomas estrechamente relacionados al SARS-CoV-2 (Zhou y Shi, 2021). Según la OMS, el punto focal donde dio inicio la pandemia se sitúa en la región de Wuhan, China; donde el primer caso de COVID-19 fue confirmado en diciembre de 2019. No obstante, estudios exploratorios y registros anecdóticos de neumonías atípicas, estiman que pudo haber surgido entre octubre y noviembre de ese mismo año. Aun así, se desconoce el momento exacto cuando ocurrió el salto zoonótico y los eventos de recombinación en humanos (Roberts, Rossman y Jaric, 2021).

4.1.1. Estructura del SARS-CoV-2

El SARS-CoV-2 se encuentra dentro del grupo IV en la clasificación de Baltimore, por lo que su genoma está constituido por una hebra de ARN monocatenario en sentido positivo. Su genoma posee una longitud de 29,891 bases y se encuentra relacionado con otros coronavirus humanos dentro del género beta-coronavirus como SARS-Cov y MERS-Cov, responsables de brotes epidémicos en las décadas pasadas (Gao y Cui, 2021). En los coronavirus, se describen cuatro principales proteínas estructurales que conforman la partícula vírica y son las responsables de los mecanismos de infección y replicación.

Comenzando por la proteína de espiga, se ha caracterizado como una glucoproteína en conformación de trímero formada por dos subunidades, la S1 en el extremo N-terminal y S2 ubicada en el extremo C-terminal. Esta es una de las más estudiadas, por que induce la fusión del virus por medio del reconocimiento de una región de la proteína S conocida como dominio de receptor de unión (RBD por sus siglas en inglés). El RBD reconoce la enzima convertidora de angiotensina II (ECA2) ubicada en el exterior de la membrana de células humanas (Peacock y cols., 2021). La ECA2 está presente en altas concentraciones en tejido epitelial respiratorio, tejido cardiaco, riñones, hígado e intestinos; por este motivo, el SARS-CoV-2 tiene un alto tropismo hacia dichas células causando la

sintomatología que ha caracterizado a esta enfermedad (Shenoy, Qi, Katovich, y Raizada, 2011).

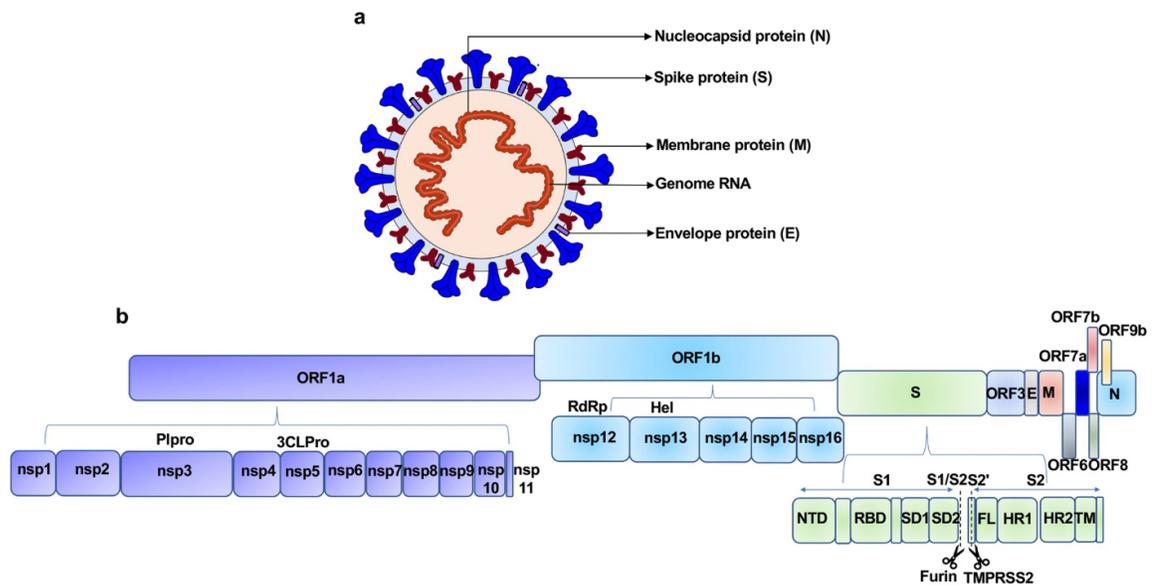
Cuadro 1: Proteínas estructurales del SARS-CoV-2

Proteína viral	Dominios	Función
Proteína de espiga (S)	Subunidades S1 y S2	Responsable de inducir la fusión del virus por medio del reconocimiento del receptor ACE2
Nucleoproteína (N)	N1b, N2b, N1a, N2a	Brinda estructura al ARN y regula la transcripción
Proteína de membrana (M)	n.a	Rol crucial en morfogénesis viral y ensamblaje de otros componentes
Proteína de envoltura (E)	n.a	Maduración del virión, estructura y posible virulencia

(Zhu y Sun, 2020)

Seguidamente, la nucleocápside, conformada también por dos dominios importantes es una proteína multifuncional que se une de manera no específica al genoma de ARN viral y permite la formación del núcleo ribo-nucleoprotéico. Además de su función estructural, también contribuye a la regulación de la transcripción y la replicación viral (Zeng y Jin, 2020). Por otro lado, la proteína E y M intervienen en el ensamblaje, tráfico intracelular, maduración y proceso de *budding* del virión (Shepley-McTaggart y Harty, 2021). En el caso de la proteína M no se conoce con certeza todas sus funciones, pero se hipotetiza que interactúa de forma activa con las proteínas N, E y S e incluso interviene en la glucosilación de esta última (Thomas, 2020).

Figura 1: Representación gráfica de las proteínas estructurales y accesorias del SARS-CoV-2



Descripción: El genoma del SARS-CoV-2 codifica para (a) proteínas estructurales y (b) proteínas accesorias que están involucradas tanto en los procesos de replicación y ensamblaje como el proceso de invasión. (Zhang, Yamanaka, Tang, y Moore, 2019)

Adicionalmente, el genoma viral posee un repertorio de proteínas no estructurales con funciones específicas y que requieren de modificaciones postraduccionales para poder alcanzar su forma activa. Por ejemplo, en la Figura 1 se observa la disposición de otras proteínas como la ORF3A y proteasas

Plpro (Kandeel y Abdelazim, 2021).

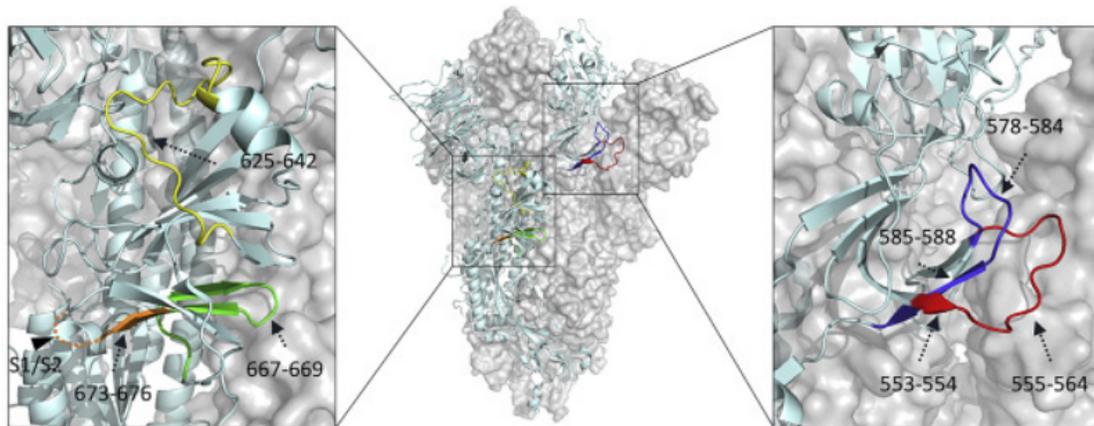
4.2. Inmunidad adaptativa ante el SARS-CoV-2

El proceso de respuesta del organismo ante el SARS-Cov-2 ha sido objeto de un intenso estudio no solo para poder desarrollar tratamientos efectivos sino para para comprender el motivo por el cual en ciertas personas la infección llega a ser mortal mientras que en otras se presentan síntomas muy leves e incluso completamente pueden llegar a ser asintomáticos. Se conoce que los casos graves de infección no se dan por un proceso generado directamente por el virus sino a una respuesta exacerbada del propio sistema inmune en un fenómeno conocido como tormenta de citoquinas (Song y You, 2020). Específicamente, se ha estudiado el rol de la interleucina-6 (IL-6) que activa a los linfocitos CD14+ CD16 y también induce la producción del factor de necrosis tumoral (TNF-alfa) (Moderbacher y Crotty, 2020).

Investigadores han determinado tres componentes fundamentales que participan en la respuesta inmune adaptativa frente al SARS-Cov-2: Células B, la fuente de producción de inmunoglobulinas; células CD4 y células T CD8. En conjunto tienen un rol en el proceso de infección viral y su capacidad neutralizante contribuye tanto en los casos leves como en los que se requieren de hospitalización (Sette y Crotty, 2021). El grado de reacción de la respuesta inmune adaptativa limita la severidad de la infección por COVID-19. No obstante, una respuesta inmune múltiple coordinada resulta más efectiva que respuestas parciales, así también células T inmaduras aumentan el riesgo de desarrollar una infección severa (Hu y Yin, 2020).

4.3. Regiones antigénicas

Figura 2: Modelo 3D de la proteína estructural S de SARS-CoV-2



Descripción: Se muestran dos regiones altamente inmunogénicas específicamente, S1-93, aa 553–564, rojo; S1-97, aa 577–588, azul; S1-105/106, aa 625–642, amarillo; S1-111, aa 661–672, verde; and S1-113, S1-113 y aa 673–684, anaranjado. Se han caracterizado potentes anticuerpos afines a dichas regiones (Z. Li y cols., 2020).

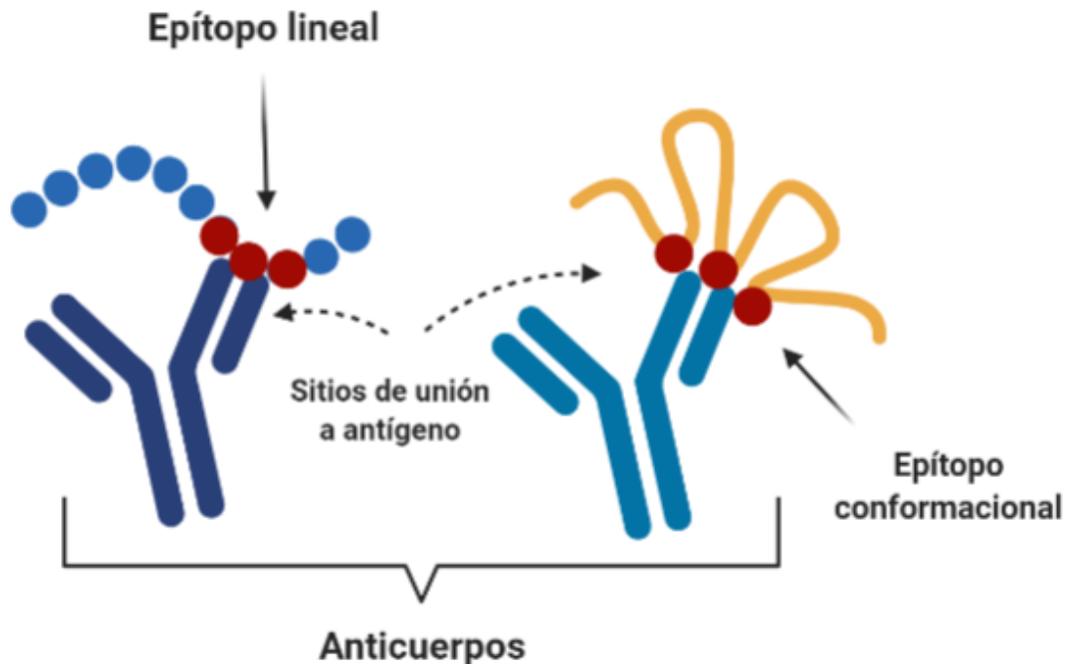
El reconocimiento y presentación de antígenos es un paso determinante para el desarrollo de la inmunidad adaptativa. Desde el inicio de la pandemia, los investigadores se dieron a la tarea de identificar las principales regiones inmunogénicas del SARS-Cov-2. Esto es altamente útil no solo para el desarrollo de las vacunas sino también para el diseño de pruebas de diagnóstico más

precisas. Se han reportado diversas regiones en diferentes proteínas virales capaces de producir una fuerte respuesta inmune adaptativa. Dependiendo de la composición de aminoácidos en cada una de estas regiones se ha visto afectado considerablemente el grado de afinidad y avidéz por parte de los anticuerpos en diferentes inmunoensayos (Bauer, 2021). Otra peculiaridad de los anticuerpos generados contra estas regiones es que se han reportado en algunos casos una respuesta de inmunidad cruzada frente a otros coronavirus humanos, por lo que son un punto central de investigación (van der Heide, 2020).

La mayoría de los anticuerpos que se han identificado son afines principalmente a regiones provenientes de la proteína S; específicamente a la RBD de la S1. Cerca de esta secuencia se ha descrito un *hot spot* capaz de generar anticuerpos diversos y potentes contra el virus (Z. Li y cols., 2020). Esto se correlaciona con el rol que tiene esta proteína en el mecanismo de infección. También se han estudiado posibles dominios antigénicos en secuencias de la nucleoproteína N, poliproteína replicasa 1ab, proteína ORF3a y proteínas de membrana (Alagarasu y Team, 2020). De estas, la proteína N es una de las más abundantes y que mayor generación de anticuerpos induce seguido de la proteína S (Terry y Geiss, 2021).

Para comprender mejor la interacción antígeno anticuerpo se deben analizar los epítomos responsables de dicha antigenicidad. Un epítomo se define como un grupo de aminoácidos o cualquier otra molécula expuesta en la superficie del antígeno (generalmente una proteína) que va a interactuar por medio de complementariedad eléctrica con el sitio de unión a antígeno de los anticuerpos (Diamandis y Christopoulos, 1996). Cuando el epítomo se trata de una secuencia de aminoácidos se pueden dividir en dos categorías: Epítomos lineales y conformacionales como se muestra en la Figura 3.

Figura 3: Representación gráfica de epítomos lineales y conformacionales



Descripción: En el caso de secuencias de aminoácidos consecutivos reconocidas por el sitio de unión a antígeno, se les denomina epítomos lineales o continuos. En contraparte, los epítomos conformacionales están constituidos por aminoácidos cercanos, pero no consecutivos. Fuente propia.

Epítomos continuos

Estos epítomos están conformados por una secuencia de aminoácidos consecutiva; es decir, que la interacción con los anticuerpos va a estar guiada estrictamente por la estructura primaria de un conjunto de aminoácidos continuos (Van Regenmortel, 2009).

Epítomos lineales

Los epítomos conformacionales o discontinuos, al contrario de los anteriores, dependen de residuos no consecutivos; pero que residiendo en la estructura terciaria de la proteína se encuentran bastante cercanos entre sí. En este caso los aminoácidos sí contribuyen a la conformación estructural del epítomo, por lo que al desnaturalizarse pierden su antigenicidad. El estudio de epítomos conformacionales es preferible para aplicaciones que requieran utilizar la proteína completa como en el caso de desarrollo de terapias; mientras que los epítomos lineales se prefieren en condiciones en las cuales la muestra requiere ser desnaturalizada como ensayos de *Western blot* (Forsstrom y Uhlen, 2015).

4.4. Estrategias de control y diagnóstico de SARS-CoV-2

El control epidemiológico de pandemias y epidemias es un reto cada vez mayor, ya que se deben enfrentar las condiciones de la época actual; mayor globalización e interconectividad. En este contexto, la globalización puede resultar beneficiosa, pero al mismo tiempo contraproducente. Las nuevas tecnologías y alta interacción internacional aceleran los esfuerzos para el control de las infecciones y el desarrollo de tratamientos como vacunas; sin embargo, también se incrementa la velocidad y el alcance de los nuevos brotes infecciosos que puedan surgir (Sreerag, Sreehari, Srinivas, y Variar, 2021). Los métodos de control epidemiológico se han tenido que refinar y aplicar de una manera más intensiva para poder lograr un control efectivo. Una de las primeras estrategias tomadas fue la aplicación de toques de queda y la limitación del contacto social. Si bien esta medida resulta controvertida, algunos estudios demuestran una tendencia en la disminución de casos en periodos de aplicación de esta medida (Andronico y cols., 2021). La desventaja de esto es que produce daños a nivel micro y macroeconómico, viéndose ya reflejados estos efectos en la mayoría de los países de Latinoamérica y el mundo (García y Atun, 2020).

Actualmente, se está apostando por la vacunación; otra estrategia de control epidemiológico que ha mostrado resultados beneficiosos en la reducción del número de casos. Si bien, la vacuna no es infalible en la totalidad de los casos, aún las vacunaciones parcialmente efectivas proveen una solución importante a la pandemia de SARS-CoV-2. Aunque esto no controle por completo la transmisión, reduce significativamente el riesgo de infección y el de mortalidad (Petersen y Gokengin, 2020). Aunado a lo anterior, se ha demostrado que es necesario implementar medidas de vigilancia serológica para proyectar las trayectorias epidémicas y así predecir posibles focos de infección (Larremore y Grad, 2021). La vigilancia serológica va unida al desarrollo y empleo de métodos de diagnóstico de infecciones. Durante el transcurso de la pandemia, el diagnóstico temprano y masivo ha sido una de las principales estrategias utilizadas para mitigar los efectos que produce el contagio exponencial, debido a que una detección temprana permite colocar a los pacientes en cuarentena y evitar que se siga propagando el virus (Alvargonzalez y Cao, 2020).

Al inicio de la pandemia, uno de los métodos utilizados para el diagnóstico de COVID-19 era por medio de imágenes del tórax (Ahammed y Islam, 2020); sin embargo, rápidamente se recomendó no utilizar este método debido a que los pacientes que podían ser diagnosticado con dicha técnica eran únicamente aquellos que presentaban infecciones graves, dejando fuera a los casos moderados y asintomáticos. Adicionalmente, solo era efectivo si se encontraba en un estado más o menos avanzado la infección (Pokhrel y Hanbin, 2020). Gracias a las nuevas técnicas de secuenciación, el genoma

Cuadro 2: Comparativa de principales características de las pruebas de diagnóstico empleadas para SARS-CoV-2.

Especificaciones	PCR	Antígeno	ELISA
Tiempo	3-5 hrs	10-15 min	48 hrs
Precio	Alto	Bajo	Medio
Sensibilidad	96-99 %	70-87 %	90 %

(Ahammed y Islam, 2020)

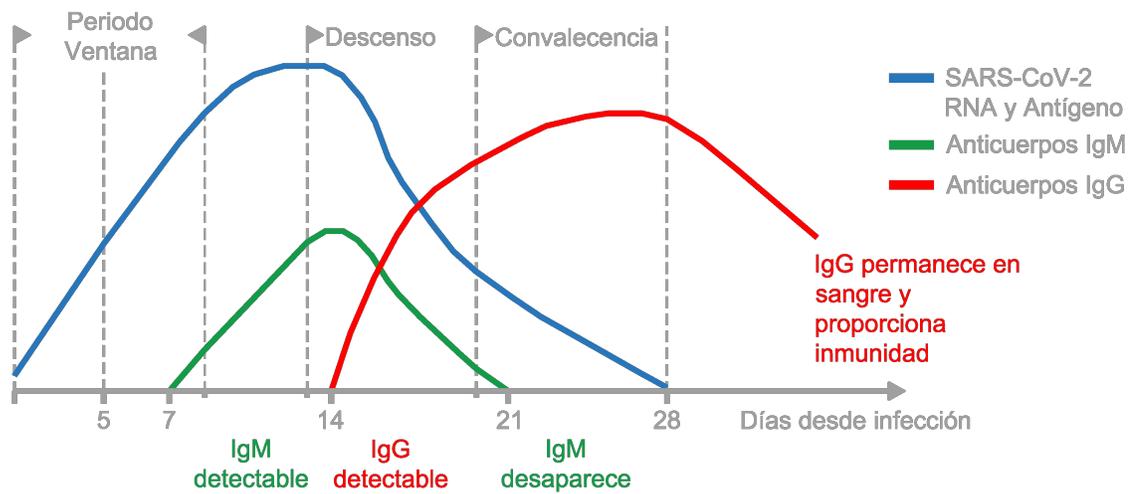
completo del virus estuvo disponible con acceso libre al público general e investigadores alrededor del mundo. Esto permitió que rápidamente se desarrollaron pruebas moleculares basadas en la detección de material genético.

A corto plazo, los kits de diagnóstico molecular basados en q(RT)-PCR estuvieron disponibles de forma comercial. Este método es altamente fiable, ya que permite la detección del virus, a pesar de que este se encuentre en etapas tempranas del proceso infeccioso; es decir en un bajo número de copias virales. Se basa en la detección de regiones del genoma del virus que codifican para secuencias que participan en funciones fundamentales del mismo. Se seleccionaron segmentos que codifican para la proteína S o de espiga; también la secuencia para la nucleocápside, que es altamente conservada. En ocasiones, se utiliza una combinación de genes para llevar a cabo la detección, aumentando así el poder de la prueba y reduciendo así la probabilidad de obtener falsos positivos o negativos (Calvache y Paucar, 2020).

Otro método para diagnosticar el virus es por medio de inmunoensayos. Estos son métodos muy bien establecidos para la detección de una alta variedad de enfermedades. Se basa en la detección de proteínas antigénicas o anticuerpos séricos en pacientes que han estado expuestos al virus. En cuanto a la detección de anticuerpos, estos ensayos pueden detectar IgG, IgM e IgA. Una ventaja de las pruebas basadas en IgG e IgM es que se puede visualizar la seroconversión entre ambas inmunoglobulinas como se observa en la Figura 4, por lo que se puede no solo diagnosticar la presencia del virus sino también determinar el estado de la infección (Liu y Rusling, 2021). Tradicionalmente, se han utilizado ELISAs *Enzyme-Linked ImmunoSorbent Assay* para la identificación inmunoquímica de anticuerpos o partículas antigénicas; sin embargo, por la necesidad de un rápido diagnóstico y muestreo masivo se han desarrollado numerosos kits de pruebas rápidas basados en el mismo principio inmunoquímico, pero empleando el sistema de flujo lateral similar al de las pruebas de embarazo. En estos sistemas, la partícula ya sea anticuerpo o antígeno marcado se encuentran embebidos en una matriz generalmente de nitrocelulosa (Seo y Kim, 2020).

La posibilidad de tener diversos métodos de detección permite ampliar la capacidad de diagnóstico. No obstante, es necesario analizar el fin principal y las necesidades que se tienen para seleccionar uno de los métodos, ya que cada uno conlleva ciertas ventajas y desventajas. En el caso de la prueba de PCR es el estándar por excelencia cuando se desea tener un diagnóstico certero, pero presenta la desventaja de su alto costo, requiere de personal entrenado y el resultado no es inmediato. Por otro lado, las pruebas de antígeno-anticuerpo de flujo lateral muestran un resultado en cuestión de minutos, pero presenta una mayor probabilidad de mostrar falsos negativos, lo cual puede resultar riesgoso cuando se trata de diagnosticar al mayor número posible de personas para colocar en aislamiento (Tono y Mahecha, 2020). Algunos autores recomiendan el uso de estas pruebas con fines diagnósticos únicamente si el resultado es positivo. En caso de que la prueba resulte negativa, es necesario realizar una prueba confirmatoria con PCR. Por otro lado, para estudios epidemiológicos y de rastreo, las pruebas rápidas han demostrado ser una herramienta muy costo efectiva (Langa y Díez, 2021).

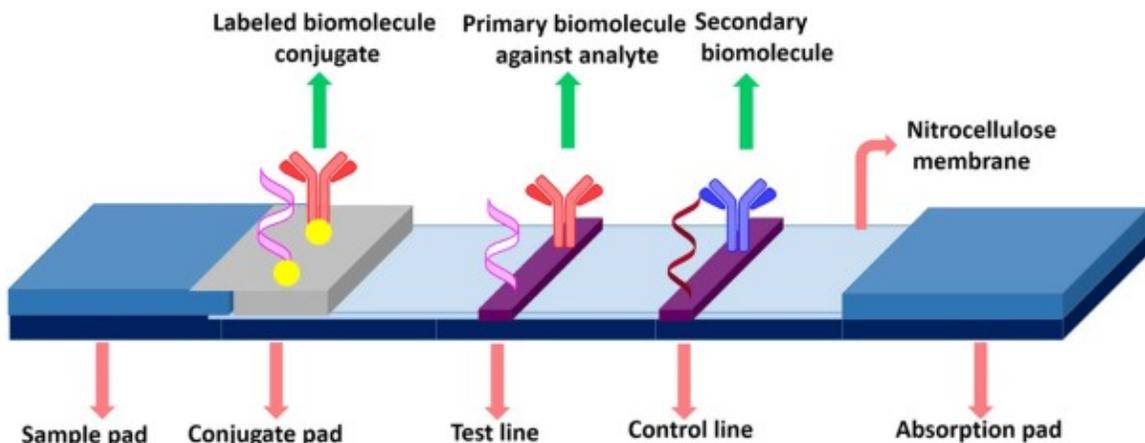
Figura 4: Concentraciones de anticuerpos séricos IgG e IgM durante el transcurso de la infección por SARS-CoV-2



Descripción: A lo largo del proceso infeccioso se produce una seroconversión entre anticuerpos IgM e IgG. Al final de la infección, las IgG son las que proporcionan la inmunidad de memoria (Alvarado Orozco, Rangel Aguilera, Barba Hernández, García Baeza, y Sánchez Navarro, 2021).

4.4.1. Desarrollo de pruebas rápidas

Figura 5: Esquema simplificado del funcionamiento de una prueba de flujo lateral convencional



Descripción: La migración de los analitos ocurre en dirección de izquierda a derecha en donde interactúan con los conjugados marcados (Bahadır y Sezgintürk, 2016).

La caracterización de epítomos inmunogénicos ha sido relevante para el desarrollo de pruebas rápidas, ya que, dependiendo del grado de afinidad y avidéz hacia estas regiones antigénicas, se verá afectado significativamente el porcentaje de sensibilidad y especificidad de la prueba. Esta es una característica fundamental que todo desarrollo de kits de diagnóstico busca. Otra característica importante que se debe tomar en cuenta es que debe ser de bajo costo y fácil manufactura (Sajid, Kawde, y Daud, 2015). En la Figura 5 se presenta el funcionamiento básico de las pruebas de flujo lateral. Como se mencionó antes, la base generalmente consiste en una matriz de celulosa o nitrocelulosa, debido a que esta tiene una baja afinidad por las proteínas. La primer almohadilla contiene

biomoléculas conjugadas normalmente con nanopartículas de oro. Si el analito es un anticuerpo son entonces los antígenos los que se conjugan con las nanopartículas. Luego de esto, migran hacia las líneas de prueba y de control donde se encuentran con anticuerpos primarios afines a los conjugados, produciendo así una agregación de partículas. Los resultados se visualizan gracias a un cambio colorimétrico resultante de la agregación de nanopartículas (Bahadır y Sezgintürk, 2016)

4.5. Bioinformática para evaluación y predicción de epítomos

El desarrollo de las herramientas bioinformáticas en las últimas décadas ha permitido acelerar la investigación científica y creación de nuevas tecnologías. Los algoritmos de predicción poseen cada vez un mayor poder computacional, lo que reduce los recursos y experimentos necesarios para llevar a cabo un proyecto. En el contexto de la inmunología, la *Immune Epitope Database* (iedb.org) es un repositorio en el cual se recopila información experimental de epítomos inmunogénicos y datos de estudios que incluyen ensayos de afinidad de anticuerpos, células T, asociación a MHC, entre otros. Dicha información abarca alérgenos y patógenos convencionales y emergentes como el SARS-CoV-2 (Vita y cols., 2019). Para ejemplificar la utilidad de estas herramientas, se puede hacer referencia a estudios en los cuales mediante el análisis bioinformáticos de epítomos de células B y T evaluaron y validaron experimentalmente las predicciones realizadas por los algoritmos para proponer epítomos a partir de una librería de secuencias de SARS-CoV-2. Estos epítomos demostraron tanto *in silico* como *in vitro* la capacidad de inducir potentes anticuerpos neutralizantes. El resultado fue un listado de péptidos, los cuales se describieron como potenciales componentes para la creación de una vacuna contra el COVID-19 (Lim, Lim, Jazayeri, Poppema, y Poh, 2021).

4.6. Producción de proteínas recombinantes

Las proteínas recombinantes son aquellas cuyo gen es clonado para producirse en una línea celular distinta a la del organismo original. El primer ejemplo exitoso de la producción de una proteína recombinante a nivel industrial es la insulina, a en la segunda mitad de los 1900s (Gunby, 1978). Desde entonces, con el avance de la ingeniería genética y la biología molecular, se ha acelerado el diseño y producción de proteínas. En la actualidad, son una herramienta indispensable para la biotecnología y tienen como finalidad ser un instrumento altamente eficiente y costo-efectivo, además de poderse producir en cantidades y calidad deseables (Anné, Maldonado, Van Impe, Van Mellaert, y Bernaerts, 2012). La presencia de estas proteínas es mayoritaria en el área de desarrollo de medicamentos y es crucial tanto para la producción de nuevas proteínas, así como para la determinación de dianas terapéuticas (Overton, 2014). Las terapias anticáncer son las que mayor impulso han tenido empleando estas tecnologías (Grupta, 2014), pero es posible ver cada vez más aplicaciones en distintas áreas como en la agroindustria, industria de alimentos para producción de proteínas de alto valor alimenticio o enzimas que faciliten procesos; bioconversión de moléculas con valor agregado como la generación de biocombustibles y desarrollo de biomateriales, por mencionar algunos ejemplos (Chang y He, 2017), (Trono, 2019) y (Ghio y Campos, 2018).

En el contexto de la pandemia, el uso de tecnologías de proteína recombinante ha permitido la creación de diversos candidatos a vacunas basados principalmente en la expresión de fragmentos de la proteína S del SARS-CoV-2 (Zang y cols., 2021). En adición, muchas de las pruebas rápidas de diagnóstico también emplean fragmentos del RBD para la detección principalmente de IgM e IgG (Y. Li y cols., 2021). Estos métodos poseen una alta flexibilidad ante cualquier cambio sustancial por mutaciones en el genoma del virus, ya que la producción de estas proteínas se basa en las secuenciaciones que constantemente se están actualizando de la cepa original y las nuevas variantes que están surgiendo. No obstante, aún no han sido aprobadas vacunas ni pruebas de diagnóstico específicos para alguna de las nuevas variantes, pero las ya aprobadas continúa mostrando un buen

porcentaje de efectividad (Khateeb y Zhang, 2021).

4.7. Transformación de células

La transformación genética es especialmente útil para la producción de proteínas recombinantes. Es un proceso en el cual se incorpora ADN exógeno a una célula para que esta lo retenga en forma de un replicón y así poder transcribir y traducir la información de dicha secuencia (Granados y Giraldo, 2012). Existen diversos métodos por el cual se puede lograr este proceso, pero se deben tomar en cuenta consideraciones previas para poder lograr una transformación efectiva. En primer lugar, se debe tomar en cuenta el tipo de célula que se desea transformar. En algunos casos es más difícil incorporar el ADN exógeno debido a la presencia de elementos extra-membranales como una pared celular (Ruvalcaba-Hidrogo y Espino-Solís, 2019). Seguidamente es necesario dar un pretratamiento a las células para que estas sean aptas para recibir el ADN exógeno. Esto se logra principalmente con métodos químicos como el cloruro de calcio o el acetato de litio que permeabilizan la membrana celular. A este procedimiento se le denomina generación de células competentes (Serrano-Rivero y Fando-Calzada, 2013).

4.7.1. Plásmidos como vectores de expresión

Para los procesos de transformación es necesario que el ADN exógeno sea acarreado en una forma que pueda ser asimilada por la célula. Por este motivo se emplean los vectores de expresión que consisten en constructos ya sea en forma de plásmidos o vectores virales que, al ingresar a la célula, esta con sus mecanismos normales de transcripción y traducción pueden reconocer las secuencias del vector y expresar el gen de interés (Zhang y cols., 2019). Aunque originalmente los plásmidos fueron descubiertos en procariotas, en la actualidad se han desarrollado numerosos plásmidos sintéticos con orígenes de replicación, promotores, terminadores y otros factores de transcripción optimizados para organismos no procariotas como lo son las levaduras, células vegetales y células animales (Y. N. S. Kang Lim y Jung, 2019). Esto abre muchas posibilidades para la expresión de genes en diferentes organismos según sea la necesidad. La técnica para construir estos vectores es el mismo utilizado para insertar un gen de interés en un plásmido en un proceso denominado clonación. Para esto es necesario contar con enzimas de restricción las cuales son endonucleasas que escinden en patrones específicos de nucleótidos. Los métodos de clonación se encuentran altamente estandarizados para diferentes propósitos (Buckhout-White y Goldman, 2018).

4.7.2. Diseño de vectores de expresión

El diseño de los vectores de expresión para la producción de proteínas recombinantes o expresión de cualquier gen en general requiere de ciertas consideraciones especiales. Si bien con las actuales tecnologías resulta un proceso mucho más simplificado se requiere de una planificación *in silico* previa para asegurar que el trabajo experimental resulte efectivo no solo para ahorrar tiempo sino para ahorrar recursos. Para esto se siguen los principios básicos de los mecanismos de transcripción de genes y el dogma central de la biología molecular. El primer elemento importante para mencionar es el promotor. Los promotores son secuencias reguladoras que favorecen la unión de las ARN polimerasas a la hebra de ADN para iniciar la transcripción (Mostafa, Kanrai, Ziebuhr, y Pleschka, 2013). El estudio de promotores ha sido muy importante y ha permitido encontrar diferentes niveles de expresión en diferentes organismos. Es determinante que al momento de seleccionar un vector de expresión se cuente con un promotor que sea afín al organismo, ya que, de lo contrario, la expresión de los genes de interés se verá limitado o incluso inhibido (Rytter y cols., 2014).

Adicionalmente se debe hacer una diferenciación entre promotores constitutivos e inducibles.

Los promotores constitutivos son aquellos que regulan la expresión de un gen *downstream* de forma constante. Por otro lado, los promotores inducibles requieren de algún factor que generalmente es la presencia de una molécula o alguna condición específica como determinada temperatura para activar la unión de la ARN polimerasa a la región del promotor. Generalmente este tipo de promotor inducible se utiliza cuando el producto de la expresión es tóxico para la célula hospedera, por lo que es necesario regular más los niveles de expresión. En contraparte, un promotor constitutivo se recomienda cuando se desea obtener el máximo rendimiento de un producto (Sun, Jiang, Zhao, Pu, y Bao, 2021). Otro factor importante para considerar al momento de diseñar un vector de expresión es el marco abierto de lectura, que es la secuencia que comprende entre un codón de inicio y un codón de terminación. Cuando se diseña un vector, la inserción de la secuencia en un sitio de clonación múltiple debe cuidarse que no se altere el marco de lectura y que obedezca la transcripción de codones a partir del codón de inicio (Torrance y Lydall, 2018)

Adicional a tomar en cuenta es el proceso de optimización de codones. El código genético es degenerado, es decir que existe más de un codón o secuencia de nucleótidos que codifica para un mismo aminoácido. Sin embargo, se ha demostrado que para cada organismo existe cierta preferencia a alguna de las posibles combinaciones para codificar un aminoácido. El proceso de optimización de codones permite convertir una secuencia que originalmente pertenece a otro organismo diferente a una secuencia con los codones más utilizados por el organismo hospedero. Lo que se logra con esto es aumentar la tasa de transcripción de un gen recombinante (Yang, Lyu, Zhao, y Liu, 2021). Finalmente se debe considerar factores como tamaño del transcrito de interés que mientras más pares de bases tenga, menos eficiente será la transcripción; el terminador también es fundamental para tener transcritos correctamente estructurados; por último, escoger un marcador de selección que permita diferenciar entre las células transformantes y aquellas que no lograron acoplar correctamente el vector de expresión. Si se siguen todos estos principios para el modelaje del vector se aumenta la probabilidad de éxito (Morse, Gopal, Wagner, y Alper, 2017).

4.7.3. Sistemas de expresión

Cuadro 3: Resumen de características de los principales sistemas de expresión.

Sistema de expresión	Rendimiento	Costo de producción	Velocidad de producción	MPT
Bacteria	Alto	Bajo	Alto	Ninguna
Levadura	Alto	Bajo	Alto	Simples
Línea celular (mamífera)	Bajo	Alto	Bajo	Complejas

(Dana, 2018)

Descripción: Se describe un breve resumen de las características que influyen en la costo-efectividad de emplear determinado sistema de expresión. Donde MPT se refiere a modificaciones post-traduccionales

Un sistema de expresión está conformado por el vector de expresión y un organismo hospedero, el cual contiene los elementos necesarios para llevar a cabo los procesos de transcripción y traducción de los genes de interés. Casi cualquier organismo puede llegar conformar un sistema de expresión, por lo cual, la selección de este dependerá de la finalidad de la transformación y los recursos que se tengan disponibles. En el caso de las bacterias, en general son fácilmente transformables y por su corto tiempo de generación se recuperan altos rendimientos. Como ejemplo, uno de los sistemas que más se utilizan para investigación es el de *Escherichia coli* (Chen, 2012). La principal desventaja de estos sistemas es que los organismos procariotas no son capaces de realizar modificaciones postraduccionales, por lo que no son aptos para la producción de ciertas proteínas eucariotas. Esta desventaja puede ser superada realizando modificaciones a ciertas rutas metabólicas para lograr dichos cambios

postraduccionales; sin embargo, esto sigue continua siendo un proceso poco eficiente en la actualidad (Valderrama-Rincon y cols., 2012).

En el caso de las líneas celulares mamíferas como las células CHO (*chinese hámster ovary*), son ampliamente utilizadas tanto en investigación científica como para la producción de muchas proteínas terapéuticas derivadas de secuencias humanas. La preferencia por estas líneas celulares se debe justamente a la habilidad de producir modificaciones postraduccionales complejas, no obstante, son muy complicadas de mantener y reproducir; los costos de producción son altos y la recuperación de producto tiene generalmente bajos rendimientos (Johari y cols., 2021). En contraparte las levaduras poseen la particularidad de ser organismos eucariotas unicelulares, por lo que combina en cierta medida las ventajas de los sistemas bacterianos y de células mamíferas. Con las levaduras se pueden obtener altos rendimientos de transformación y los costos de producción son bajos de manera similar a las bacterias. Adicionalmente, en condiciones óptimas pueden obtenerse elevados rendimientos de producción (Kim, Yoo, y Kang, 2015).

4.8. Levaduras como sistema de expresión

Cuadro 4: Vectores de expresión en levaduras

Tipo	YIp	YE _p	YC _p	YR _p
Mecanismo	Integración al genoma	Replicación autónoma	Replicación autónoma empleando CEN	Replicación autónoma empleando ARS
Copias/célula	variado	20-50	1-2	>100
Eficiencia de transformación	moderada	alta	alta	alta

(Gnügge y Rudolf, 2017)

Descripción: Existen cuatro principales vectores con un mecanismo diferente de replicación y cuyas funciones son específicas para trabajar en levaduras.

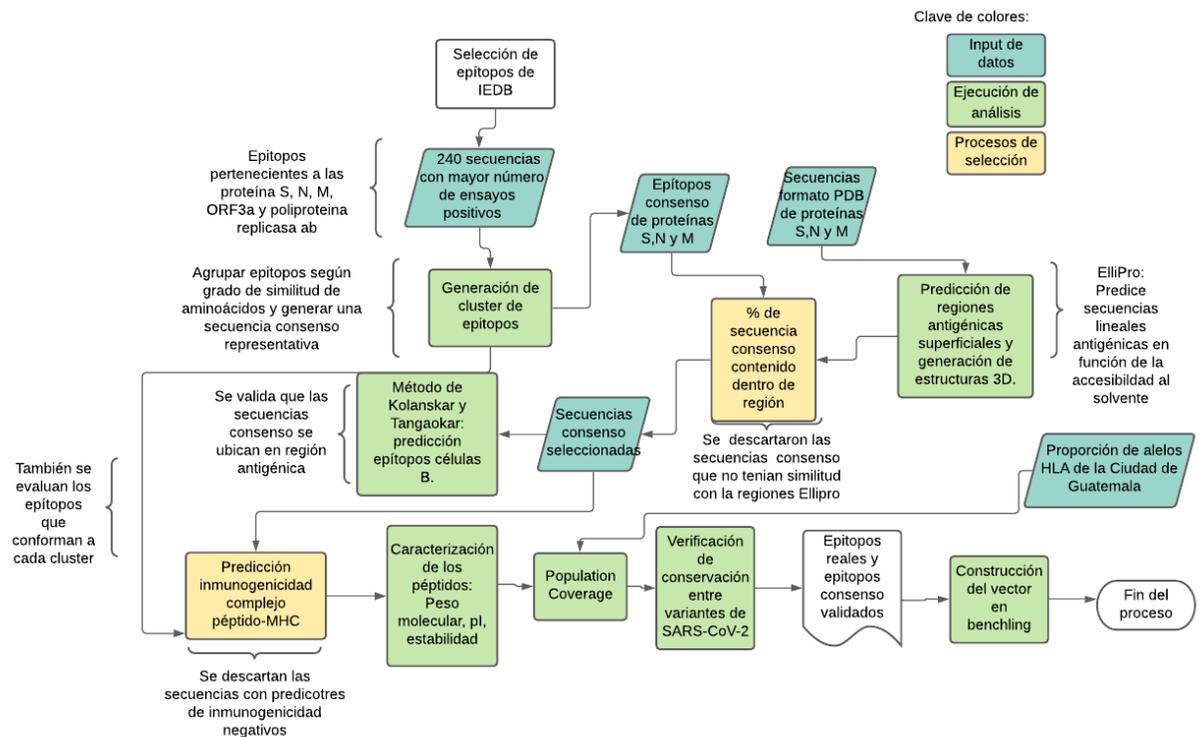
En las últimas décadas, uno de los principales sistemas de expresión ha sido *Escherichia coli* gracias a su alta docilidad a la edición genética, resiliencia ante cambios ambientales y estímulos, además de que se conoce su genoma completo. No obstante, a pesar de su versatilidad y fiabilidad, el uso de organismos procariontes para expresión de genes recombinantes posee ciertas desventajas dentro de las cuales se puede mencionar como principal característica no deseada y es la carencia de modificaciones postraduccionales, por lo que no son modelos recomendados si se desean producir proteínas glucosiladas o con alguna otra modificación. Esto puede afectar significativamente la función de la proteína. Otra desventaja es la formación de cuerpos de inclusión, los cuales son agregados proteicos que al encontrarse en este estado no pueden cumplir su rol o función original (Baghban y cols., 2019).

4.8.1. Tipos de plásmido para expresión en levaduras

Las técnicas para la construcción de vectores sintéticos han permitido diseñar diferentes plásmidos de expresión en levaduras con características variadas que afectan principalmente al grado de estabilidad y el número de copias. Con base en esto se ha establecido una clasificación que permite describir cada una de estas variaciones como plásmidos integrativos (YIp), episomales (YE_p) y centroméricos (YC_p) (Vieira Gomes, Souza Carmo, Silva Carvalho, Mendonça Bahia, y Parachin, 2018).

- Ylp: Su nombre proviene del hecho de que estos plásmidos carecen de un origen de replicación en levaduras, por lo que la única manera en la cual se pueden conservar es por medio de su integración al genoma de la célula. Para lograr esto se debe diseñar un casete genético que esté flanqueado por secuencias homólogas a un locus específico donde se quiere inducir la integración. Estos plásmidos son del tipo suicida, ya que si no se integran por recombinación de las regiones homólogas para poder conservar una característica favorable se pierden durante los eventos mitóticos. Estos plásmidos son preferidos por su estabilidad y número de copias moderado (Ergün, Hücetoğulları, Öztürk, Çelik, y Çalık, 2019).
- YCp: Los vectores centroméricos son considerados altamente estables, pero poseen un bajo número de copias (1-2 copias/célula). Estos incorporan secuencias de replicación autónoma (ARS por sus siglas en inglés) en conjunto con el elemento centromérico CEN. Ambos elementos le permiten al plásmido replicarse como si se tratara de un cromosoma independiente. Esto les otorga una alta estabilidad sin necesidad de integración. Una variante de este vector son los plásmidos replicativos que poseen únicamente las secuencias ARS, lo que les permite un alto número de copias, pero baja estabilidad (Piva, De Marco, Moraes, Reis, y Torres, 2020).
- YEp: Estos plásmidos poseen un elemento 2μ , un origen de replicación que se encuentra naturalmente en plásmidos de levaduras. Debido a esto, el vector se comporta de manera más parecida a los plásmidos bacterianos, ya que al tener su propio origen de replicación se pueden obtener alto número de copias por células (5-50 copias). Aunque su estabilidad es más reducida en comparación a un plásmido integrativo o centromérico; bajo una presión selectiva constante se tiene una conservación del plásmido moderada (Ergün y cols., 2019).

Figura 6: Metodología del análisis bioinformático para la selección de epítopos candidatos con alta probabilidad antigénica



Descripción: La metodología está basada en una versión modificada de los procedimientos propuestos por (Qamar y cols., 2020) y (Paul y cols., 2021). Se combinan diferentes herramientas predictorias en IEDB y algunos *softwares* externos.

5.1. Identificación de epítomos lineales de SARS-Cov2 reportados con una alta probabilidad de generación de anticuerpos específicos

Para aumentar las probabilidades de seleccionar epítomos altamente inmunogénicos, se optó por utilizar péptidos que ya hayan sido evaluados previamente en laboratorio. A través del repositorio de *Immune Epitope Database Analysis Resource* (IEDB) se realizó una revisión en el cual se filtró una búsqueda de epítomos con mayor número de ensayos inmunológicos que obtuvieron resultados positivos. La mayoría de estas secuencias han sido caracterizadas en estudios con ensayos de células T, células B y ligandos de MHC. Se evaluaron un total de 240 secuencias de péptidos cortos de entre 9-28 mer. Del total de secuencias 81 pertenecían a regiones de la proteína S, 26 secuencias a la nucleoproteína, 22 a la proteína M, 20 a la poliproteína replicasa 1ab y 10 a la ORF3a.

Los epítomos candidatos demostraron tener índices óptimos en los análisis de accesibilidad al solvente y predicción de epítomos de células B y T. El éxito de este análisis recae en la aproximación de la metodología, la cual no busca predecir epítomos desde cero, sino que plantea aprovechar la alta cantidad de información que se actualiza constantemente por medio de las bases de datos de péptidos previamente caracterizados y de estos conjuntos determinar los epítomos óptimos para fines específicos. Esto no significa que los algoritmos posean baja eficiencia; de hecho, algunos estudios reportan que no existe cambio significativo entre emplear datos predichos en su totalidad por un algoritmo y emplear datos experimentales (Paul y cols., 2021), pero si es reconocido que aún existen limitaciones en el poder computacional, principalmente para la predicción de epítomos de células B (Isea, Cardinale, y Ortega, 2017).

5.1.1. *Cluster Epitope Analysis*

En ocasiones, el MHC puede procesar los péptidos virales de distintas maneras. Esto ocasiona que dentro de una región específica del antígeno se puedan obtener diferentes epítomos cuyas secuencias se intercalan entre sí. El objetivo es generar una secuencia consenso que sea representativa de un grupo de péptidos cortos. Se empleó esta herramienta para generar agrupaciones de epítomos basados en su cercanía y a la identidad de su secuencia con un porcentaje de similitud mayor a 70 %. Las secuencias dentro de los conjuntos fueron alineadas para comparar superposiciones y generar una secuencia consenso representativa de cada grupo (Dhanda y cols., 2018). Con este análisis, se identificaron las regiones o residuos de mayor frecuencia en los epítomos caracterizados experimentalmente. Se generaron conjuntos para todas las proteínas, pero para el resto del análisis solo se tomaron en cuenta aquellas agrupaciones con más de 3 secuencias relacionadas.

5.1.2. Predicción de regiones antigénicas basadas en accesibilidad al solvente

La accesibilidad al solvente es una propiedad utilizada para describir que tan expuesta se encuentra una región específica de las proteínas ante interacciones con el medio externo. Es importante estudiar estas regiones en los antígenos, ya que a mayor accesibilidad, mayor probabilidad de que el MHC interactúe con esa zona específica (McCombs y Owen, 2015). *ElliPro: Antibody Epitope Prediction* es una herramienta que permite predecir epítomos continuos o discontinuos basándose en la propiedad de accesibilidad al solvente y a la flexibilidad de los enlaces peptídicos (Ponomarenko y cols., 2008). Para este análisis se emplearon las secuencias de las proteínas S (Klein y cols., 2020), N (S. Kang y cols., 2020) y M (Mahtarin y cols., 2020). Se seleccionaron aquellas secuencias con puntaje más alto de accesibilidad. Posteriormente se utilizaron las secuencias consenso generadas en el paso anterior para determinar cuánto porcentaje de cada secuencia consenso estaba contenida

dentro de una de las regiones predichas por este análisis. Esta comparación se utilizó como criterio para descartar aquellas secuencias consenso que tenían un bajo porcentaje de similitud con alguna de las regiones, es decir no se encontraban contenidas y al mismo tiempo permitió seleccionar aquellas secuencias consenso con mayor accesibilidad al solvente.

5.1.3. Predicción de epítomos de células B

La propiedad inmunogénica de las secuencias consenso fue validada por medio del predictor de epítomos de células B, a través de la escala de antigenicidad de Kolaskar y Tongaonkar. Este es un método semi-empírico que se basa en las propiedades fisicoquímicas de los aminoácidos tomando como base la ocurrencia de residuos en epítomos caracterizados en diferentes experimentos. Se emplea para determinar la probabilidad de una secuencia para ser un epítomo de células B (Kolaskar y Tongaonkar, 1990). Se ingresaron las secuencias en formato FASTA de las proteínas S, N y M con sus respectivos número de acceso: P0DTC2, P0DTC9 y P59596. Para este análisis se consideró que un puntaje de probabilidad de 0.5 era aceptable y por arriba de 1.0 óptimo. Se generaron gráficos de puntaje de probabilidad en función de la secuencia y se determinaron las regiones del gráfico donde se ubicaban las secuencias consenso y con esto validar que se encontraban en regiones altamente antigénicas.

5.1.4. Predicción de epítomos de células T

Debido al mecanismo de activación de células T, también es necesario realizar una predicción de epítomos de células T. *Class I immunogenicity* se basa en propiedades fisicoquímicas de los aminoácidos, así como su posición dentro de un péptido para predecir la inmunogenicidad en un complejo pMHC (Calis y cols., 2013). En este paso se ingresaron tanto las secuencias consenso seleccionadas como los epítomos verdaderos que las integran, dando un resultado de 42 epítomos evaluados. Se analizaron los índices de inmunogenicidad individuales de cada péptido. Mientras más alto un valor, mayor la probabilidad de formar una unión con MHC clase I. Se consideraron los valores positivos como aceptables mientras que los negativos se descartaron.

5.1.5. Caracterización fisicoquímica de los péptidos

Para procesos de producción, extracción y purificación es necesario conocer las propiedades fisicoquímicas de los péptidos con los cuales se está trabajando. Con *Expasy tool*, se introdujeron las 42 secuencias peptídicas y se calculó el peso molecular (PM), el punto isoelectrico teórico (pI) y la estabilidad del péptido. Valores superiores a 40 en el índice de estabilidad indican que el péptido es probablemente inestable en un tubo de ensayo (Gasteiger y cols., 2005).

5.1.6. Cobertura poblacional

Los alelos de los antígenos leucocitarios humanos (HLA) son altamente polimórficos y estos influenciarán en gran medida el que un péptido pueda ser reconocido como un epítomo. Debido a que dependiendo de la región geográfica cambiará la ocurrencia de los diferentes alelos es necesario tomar en consideración que este tipo de pruebas deben diseñarse para poblaciones específicas. Con el repositorio *Allele frequency net database* (Gonzalez-Galarza y cols., 2020), pueden observarse alelos HLA más comunes de cada región geográfica. Se tomaron en consideración los alelos más frecuentes para Guatemala. La ocurrencias de estos alelos pueden verse más a detalle en apéndice 3. Se evaluaron las secuencias consenso y epítomos verdaderos presentados en la Figura 13 y se

determinó el porcentaje de cobertura teórico que tendrían los epítomos propuestos con los alelos evaluados.

5.1.7. Alineamiento de proteínas

Con el fin de verificar si existe conservación de las secuencias consenso entre las cepas emergentes de SARS-CoV-2 se realizó un alineamiento múltiple por medio de Clustal omega (Sievers y cols., 2011). Las secuencias de SARS-CoV-2 fueron obtenidas de la iniciativa global para almacenamiento de datos de influenza y COVID-19, GSAID (<https://www.gisaid.org>) (Shu y McCauley, 2017). Se emplearon los siguientes códigos de acceso: Variante alpha (EPI-ISL-2264978), beta (EPI-ISL-2266083), delta (EPI-ISL-2123970) y gamma (EPI-ISL-1568558).

5.2. Diseño *in silico* del vector de expresión

5.2.1. Optimización de codones en *S.cerevisiae* (S228C)

Se seleccionaron los siguientes epítomos consenso para proponer los diseños teóricos de los vectores de expresión: proteína S(NYNYLYRLFRKSNLKPFFERDISTEIQQA); proteína M (GAVILRGHLRIAGHHLGRCDIKDLP). Para la generación del constructo es necesario contar con la secuencia de ADN; sin embargo, de manera inicial se contaba únicamente con la secuencia de aminoácidos que conformaban a cada péptido. Por este motivo se realizó la conversión de los péptidos a su respectiva secuencia de ADN junto con una optimización de codones para *S.cerevisiae*. Se utilizó el algoritmo de <http://www.jcat.de> (Grote y cols., 2005) para la traducción de aminoácido a bases nitrogenadas. Posteriormente, se realizó de manera automatizada la optimización de uso de codones específicamente para *S. cerevisiae*. Se propuso esta levadura por fácil transformación, mantenimiento económico y generalmente suelen tener producciones de proteína recombinante con altos rendimientos (Mattanovich y cols., 2012).

5.2.2. Selección del vector de expresión

Se realizó una búsqueda en el repositorio de plásmidos Addgene para determinar vectores que contaran con promotores y marcadores de selección adecuados para *S.cerevisiae*. Se propuso el vector HO-poly-KanMX4-HO, un plásmido integrativo, el cual está diseñado para su expresión en *S. cerevisiae*. Este vector posee un tamaño de 6063 pb y tiene un origen de replicación derivado de ColE1. Este origen de replicación ha sido utilizado también en vectores derivados de pMB1, pBR322 y pUC, por lo que produce un alto número de copias. No se tienen reportes detallados sobre el rango de hospederos sobre los cuales funciona este origen de replicación, por lo que se recomienda su utilización en *E. Coli*, especialmente la cepa DH5alpha (Camps, 2010). Además contiene un gen de resistencia hacia la ampicilina para la selección de transformantes. Cuenta con un caset de integración al locus HO. El gen de HO (endonucleasa de conmutación homotálica) está involucrado en la interconversión del *mating type* celular. Este gen ha sido utilizado generalmente como sitio de integración para genes recombinantes debido a que se ha demostrado que no afecta el desarrollo ni el crecimiento celular (Baganz, Hayes, Marren, Gardner, y Oliver, 1997). Dentro de este caset se encuentra el promotor HO, el sitio de clonación múltiple y el gen de resistencia a kanamicina KanMX4.

Adicionalmente se empleó la secuencia para la codificación de poli-histidinas. Esta es una etiqueta que se puede fusionar a la proteína de interés y por medio de kits comerciales de separación obtener una proteína recombinante de alta pureza. Se pueden diseñar constructos para generar tanto proteínas etiquetadas en el extremo N-terminal como en el extremo C-terminal (Funakoshi y Hochstrasser,

2009).

5.2.3. Ensamblaje del diseño en el mapa del vector

Se ingresó el código de referencia del plásmido HO-poly-KanMX4-HO (Addgene No. 51662) en el sitio de Benchling (<https://benchling.com>) para generar los mapas de los plásmidos. Otros códigos de acceso para el plásmido son: GenBank(AF324728) y ATCC (87804). Se propusieron dos posibles diseños del vector. El primer diseño consiste en un constructo policistrónico cuyo fin es el de poder producir varios epítomos empleando un único promotor y terminador. Se propuso para este diseño utilizar una secuencia P2A (*2A self-cleaving peptides*) del virus de rinitis equina B (ERBV-1). Este péptido induce saltos ribosómicos durante la traducción de la proteína y ha demostrado una eficiencia de escisión del 90% en *S.cerevisiae* (Souza-Moreira y cols., 2018). El segundo diseño está basado en la utilización de una proteína de fusión GFP-epítomo para ser detectada por medio de anti-GFP (Wang, Zhang, Richardson, y Wu, 2021). Se utilizó la secuencia de la *yeast Enhanced Green Fluorescent Protein* (Hochrein, Machens, Messerschmidt, y Mueller-Roeber, 2017).

Para el diseño se seleccionaron las enzimas BamHI y BglII, ya que el plásmido cuenta con un único sitio de restricción natural para cada una de esas enzimas y además producen extremos pegajosos que facilitarían un proceso de clonación. Se propone entonces la inserción teórica de los sitios de restricción (G/GATCC) en el extremo cercano a 5' y (A/GATCT) cercano a 3' para ambos diseños. Se verificó además la posición del codón de inicio para asegurar que se cumpliera con un marco de lectura abierto capaz de transcribir por completo los constructos.

5.3. Evaluación de SOP de transformación para *Saccharomyces cerevisiae*

5.3.1. Preparación de células electrocompetentes

La preparación de células electrocompetentes se basó en un protocolo propuesto anteriormente por otro autor Russell con modificaciones en cuanto al número de pasos de centrifugación y a la sustitución del ditioneitol (DTT) por beta-mercaptoetanol. Todos los medios de cultivo, soluciones de lavado y soluciones tampón fueron preparadas con agua ultrapura. El tamaño del rotor, volúmenes y tiempos de centrifugación no son determinantes siempre y cuando se cumpla con dos objetivos: Obtener una alta densidad celular y una baja conductividad eléctrica (Ruell, 2020).

Se preparó un cultivo *overnight* de *S. cerevisiae* S228C en 200 mL de medio YPD (1% extracto de levadura, 2% peptona, y 2% glucosa disueltos en agua desionizada). La incubación se llevó a cabo a 250 rpm de agitación y una temperatura de 30 °C. Pasadas las 24 horas se midió la densidad óptica por medio de la absorbancia a 600 nm empleando espectrofotómetro UV-vis. El cultivo fue refrigerado en baño de hielo por 15 minutos para ralentizar la actividad celular. Para no superar la capacidad máxima de la centrifugadora disponible, se separaron los 200 mL del cultivo en cuatro tubos cónicos de 50 mL y se centrifugaron a 4000 xg por 5 min y 4°C. El sobrenadante fue decantado y los precipitados se resuspendieron con pequeños volúmenes de agua ultrapura sin llegar a la capacidad máxima.

Posteriormente se trasvasaron todos a un mismo tubo de 50 mL. Al final se llevó el volumen de agua a 36 mL y se añadieron 4 mL 10 X buffer TE (concentración final 1X= 10 mM Tris, 1 mM EDTA; pH. 8.0). Los tubos se incubaron por 45 min a 30°C y 85 rpm. Se añadió 1 mL de beta-mercaptoetanol (1M) e incubó por 15 min a 30 °C 85 rpm. El beta-mercaptoetanol puede ser utilizado en sustitución al DTT ya que al igual que el segundo, interactúa con las proteínas de la

pared celular desestabilizándola y aumentando la eficiencia de transformación (Ito, Fukuda, Murata, y Kimura, 1983).

El cultivo se centrifugó aplicando los mismos parámetros mencionados anteriormente; se decantó el sobrenadante y se resuspendió nuevamente en 50 mL de agua ultrapura. Seguidamente, se centrifugó una vez más y se resuspendió en 12 mL de buffer sH (sorbitol 1 M, HEPES 20 mM). Una vez más se centrifugó y decantó el sobrenadante. Finalmente, se resuspendió el precipitado en 0.5 mL de buffer sH y luego de homogenizar se llevó a un volumen final de 1.5 mL. Se midió nuevamente el OD600. A partir de este punto, es determinante que las células estén refrigeradas, por lo que se mantuvieron en baño de hielo hasta su utilización.

5.3.2. Electroporación

Por cada muestra para ser transformada se preparó 1 mL de sorbitol 1 M en tubos estériles y se mantuvieron en baño de hielo, al igual que las celdas de electroporación de 0.4 cm. Se añadió 1 μ L del plásmido pKT-0175 resuspendido en buffer TE, el cual se encontraba a una concentración de 435 ng ADN/ μ L. Luego, se tomó una alícuota de 80 μ L de células competentes y se combinó con la alícuota del plásmido. Se mezcló delicadamente por pipeteo y se incubó en baño de hielo por 5 minutos. Por último, esta mezcla fue transferida a la celda de electroporación previamente refrigerada y rápidamente se colocó en el sistema electroporador Biorad Gene Pulser II. Se accionó el botón para activar el programa de electroporación previamente establecido. Los siguientes parámetros de electroporación fueron utilizados para la celda de 0.4 cm empleando una descarga eléctrica con decaimiento exponencial:

Capacitancia = 25 μ F; Resistencia= 200 ohm; Voltaje= 1.5kV o 7.5 kV/cm; t= 5 ms

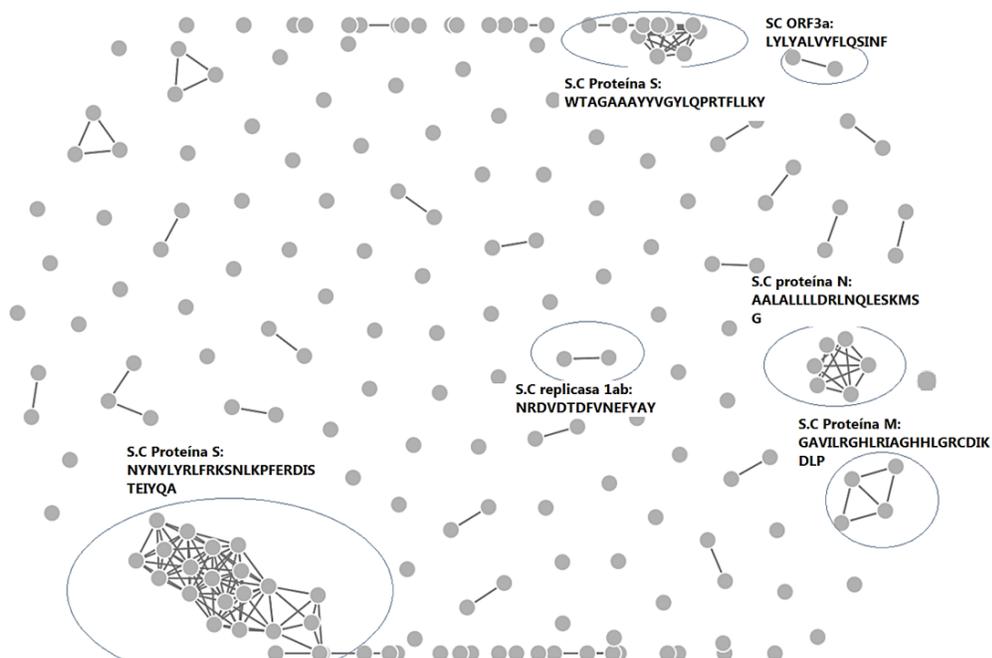
Inmediatamente después de la electroporación se agregó a la celda 1 mL de sorbitol 1 M previamente enfriado. Diferentes alícuotas de 100 μ L fueron inoculadas en placas con agar YPD y se incubaron por 72 horas a 30 °C.

5.3.3. Visualización de transformantes

Se buscaron transformantes mediante la observación de colonias que emitían fluorescencia bajo luz ultravioleta. Las placas fueron colocadas bajo un trans-iluminador a 356 nm. Posteriormente se realizó el reaislado en agar YPD de una colonia que presentaba fluorescencia y se incubó a 30 °C por 48 horas. Luego de este tiempo se visualizaron las nuevas placas de igual manera bajo el trans-iluminador para poder observar la prevalencia del plásmido. Adicionalmente se preparó un frotis y se observó empleando microscopía de fluorescencia. Se compararon bajo los mismo parámetros frotis de los transformantes como de levadura *wild type*.

6.1. Proteína estructurales (M,N y S), poseen mayor número de ensayos positivos

Figura 7: Visualización gráfica de los grupos de epítomos y principales secuencias consenso



Descripción: Los conjuntos se generaron a partir de 240 secuencias. Se presentan las secuencias consenso de los principales conjuntos formados para las proteínas S, N y M. Puntos individuales representan en su mayoría a epítomos de la ORF3A y la replicasa 1ab.

En la Figura 7 se presenta una visualización gráfica de los grupos de epítomos generados cuyos alineamientos pueden observarse a mayor detalle en anexos (cuadros 5,6,7,8,9,10 y 11). De las 240 secuencias de epítomos analizadas se obtuvieron dos conjuntos densos que representan un alto número de estudios con secuencias relacionadas entre sí, principalmente para la proteína S. Seguidamente la nucleoproteína N y la proteína de membrana M parecen ser las más estudiadas dentro del conjunto evaluado. El conjunto disperso que conforman el resto de los péptidos indica que han sido analizadas muchas secuencias independientes, principalmente para la poliproteína 1ab y la ORF3A. Esto sugiere que tanto la poliproteína 1ab como ORF3A son candidatos para considerar como blanco para métodos de detección; sin embargo, para facilitar el análisis y aumentar la probabilidad de seleccionar epítomos más inmunogénicos, se tomaron en cuenta únicamente las secuencias consenso de las proteínas S, N y M para los análisis posteriores.

6.2. Secuencias consenso poseen elevada accesibilidad al solvente

Figura 8: Regiones antigénicas de proteínas S,N y M con alta accesibilidad al solvente

Secuencia consenso	Región antigénica predicha	IP	porcentaje coincidencia	Región
NYNYLYRLFRKSNLKP FERDISTEIQYQA	VIAWNSNNLDSKGNINYLYRLFRKSNLKP PLQSYGFQPTNVGYQPYRVVLSFELLHAPATVCGPKK STNLVKNK	0.765	92.86%	S
VGGNYNYLYRLFRKS NLKP			89.47%	S
NCTFEYVSQPFLMDL E	FASTEKSNIIRGWIFGTTLDSTQSLIVNATNVVIVKVC EFQFCNDPFLGVNCTFEYVSFKNREF	0.788	31.25%	S
LSPRWYFYLLGTGPE AGL	YLGTGPEAGLPYGANKDGIWVATEGALNTPKDHIGTR NP	0.761	89.47%	N
NKDGIIWVATEGALN TPK			100%	N
GAVILRGHLRIAGHHL GRCDIKDLP	RCDIKDLPKE	0.661	100%	M
	LRGH	0.608	100%	M
SELVIGAVILRGHLR			100%	M
ATSRTLSSYYKLGASQ	TLSYYKLGASQRVAGDSG	0.698	20%	M

Descripción: Índices de protrusión (IP) cercanos a 1 presentan mayor accesibilidad al solvente. En gris se muestran las secuencias consenso que fueron descartadas para análisis posteriores, debido a poseían bajos porcentajes de coincidencia con las regiones accesibles.

El índice de protrusión (IP) es un puntaje asignado a cada péptido generado por la herramienta ElliPro y que está relacionado con la accesibilidad al solvente. Esta característica es importante, ya que se requiere para facilitar los mecanismos de procesamiento y reconocimiento de antígenos. Por este motivo, de las secuencias consenso generadas en el análisis anterior, se seleccionaron aquellas que se encontraban contenidas en al menos un 70 % dentro de estas regiones accesibles. En la Figura 8 se presentan las regiones con un IP por encima de 0.5. Puntajes mayores a este valor, indican una alta accesibilidad al solvente.

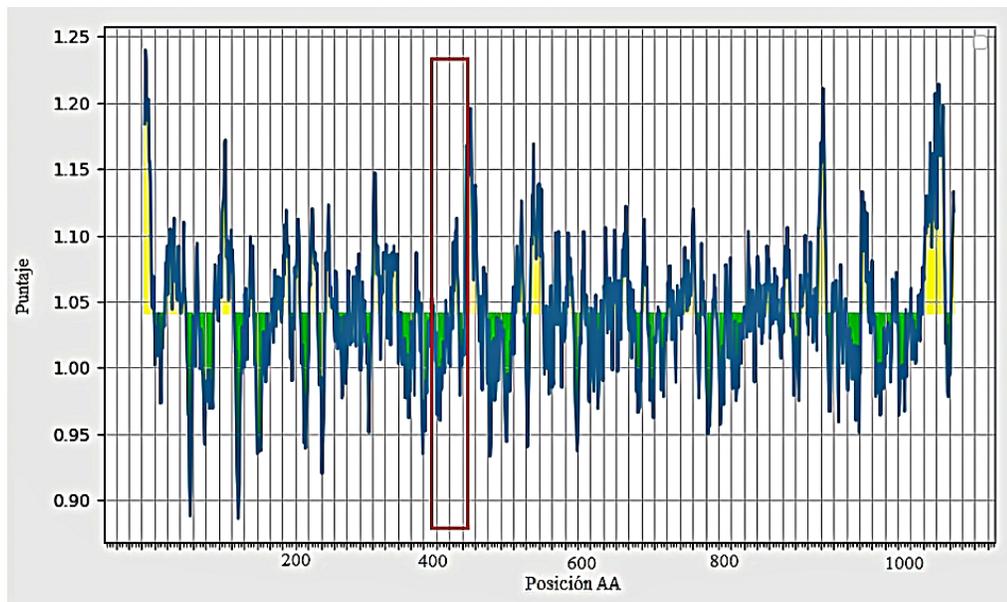
6.3. Secuencias consenso poseen alta probabilidad de formar parte de epítomos de células B

Figura 9: Resumen de puntajes del análisis Kolaskar y Tongaokar

Proteína	Puntaje promedio	Mínimo	Máximo
S	1.041	0.866	1.261
N	1.053	0.904	1.235
M	1.066	0.903	1.244

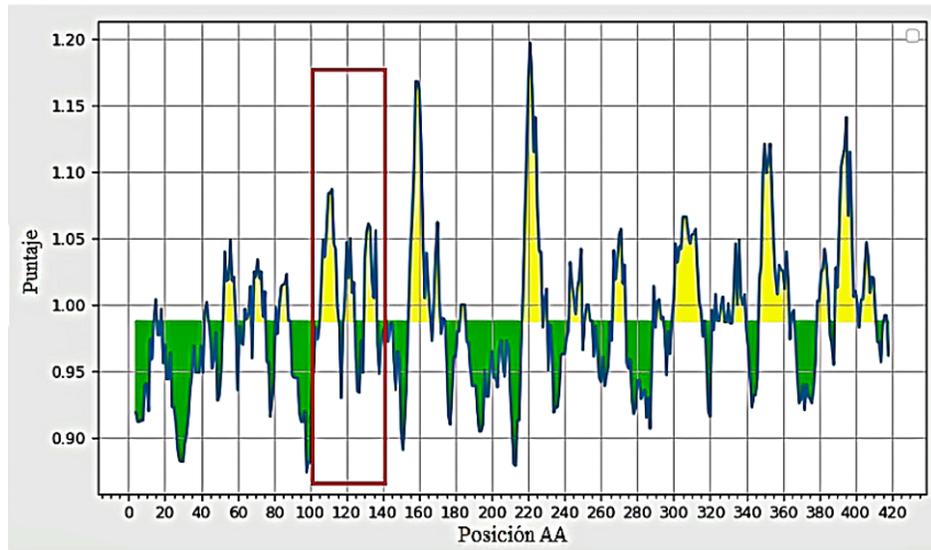
Descripción: Proteínas se consideran antigénicas si poseen valores superiores a 0.5. Adicionalmente, valores por encima de 1.04 se consideran como altamente antigénicos.

Figura 10: Representación gráfica de los puntajes de antigenicidad de Kolaskar y Tongaokar para la secuencia de la glucoproteína S de SARS-CoV-2



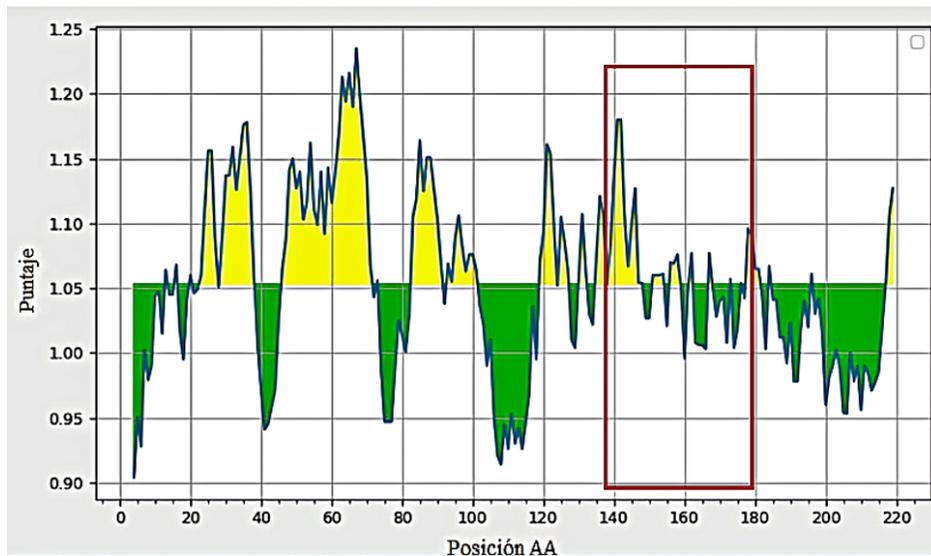
Descripción: Se muestra en el recuadro rojo los aminoácidos que abarcan los epítomos consenso de la proteína S. En amarillo se indican regiones con un puntaje de antigenicidad que supera el umbral de 1.04.

Figura 11: Representación gráfica de los puntajes de antigenicidad de Kolaskar y Tongaokar para la secuencia de la nucleoproteína de SARS-CoV-2



Descripción: Se muestra en el recuadro rojo los aminoácidos que abarcan los epítomos consenso de la proteína N. En amarillo se indican regiones con un puntaje de antigenicidad que supera el umbral de 1.04.

Figura 12: Representación gráfica de los puntajes de antigenicidad de Kolaskar y Tongaokar para la secuencia de la proteína de membrana de SARS-CoV-2



Descripción: Se muestra en el recuadro rojo los aminoácidos que abarcan los epítomos consenso de la proteína M. En amarillo se indican regiones con un puntaje de antigenicidad que supera el umbral de 1.04.

El método de kolaskar y tongaonkar permite determinar la probabilidad de una secuencia peptídica de conformar parte de un epítipo verdadero. Los resultados de este análisis para las secuencias de las proteínas S, N y M arrojan datos interesantes que respaldan el interés de esas regiones por su alta antigenicidad. El puntaje promedio de antigenicidad para las tres proteínas virales se muestra en la Figura 9 y reflejan como para estas estructuras se encuentran valores por superiores a 1.0. El método sugiere establecer un umbral en el puntaje de 0.5 para considerar un péptido antigénico y un umbral de 1.04 para considerarlo un antígeno óptimo. Un detalle notable es que la nucleoproteína posee el mayor promedio, seguido por la proteína S y M. No obstante, la proteína S es la que posee regiones con los puntajes mayores de las tres. Dicha información se traduce en una alta diversidad antigénica a lo largo de la secuencia de la proteína S. Asimismo, se muestra que las tres proteínas son altamente eficientes para la conformación epítipos de células B.

En la Figura 10 se muestran los valores de antigenicidad a lo largo de toda la secuencia de la glucoproteína de espiga del SARS-CoV-2. De la misma manera se pueden observar las regiones antigénicas tanto para la proteína N como M representadas en las Figuras 11 y 12. El listado de secuencias, puntajes y posiciones de los aminoácidos con los cuales se construyeron las figuras ya presentadas, se muestran con mayor detalle en anexos (cuadros 12, 13 y 14). Los rangos marcados en rojo se obtuvieron al comparar las secuencias generadas por el análisis y seleccionar aquellas que pertenecieran en al menos un aminoácido a las secuencias consenso evaluadas hasta este momento. Con esto se determinó que las secuencias consenso NYNYLYRLFRKSNLKPFERDISTEIQYQA y VGGNYNYLYRLFRKSNLKP comprendían desde el aminoácido 447 hasta el 478 de la proteína S.

Al verificar la naturaleza de esta región se encontró que pertenece al dominio de unión al receptor (RBD por sus siglas en inglés). Cuando se compara con la Figura 11 se muestra como esta región es altamente antigénica. Se debe tomar en consideración que según este análisis no es la más antigénica y existen varias regiones prometedoras a lo largo de la proteína S, pero se tratan únicamente de probabilidades. Además, esta sección ha sido altamente estudiada y caracterizada por su fuerte antigenicidad. Asimismo, se identificó que los epítipos consenso de N y M se encontraban en dominios antigénicos. Las secuencias de N (LSPRWYFYLLGTGPEAGL Y NKDGIIWVATEGALNTPK) en las posiciones 103-133 y las secuencias de M (GAVILRGHLRIAGHHLGRCDIKDLP y SELVIGAVILRGHLR en las posiciones 138-187. Todos estos resultados en conjunto demuestran que no solo las proteínas S, N y M tienen alto número de posibles epítipos, sino que las secuencias consenso evaluadas se encuentran dentro de regiones fuertemente antigénicas y que han sido validadas por diversos estudios.

6.4. Secuencias de epítomos consenso y epítomos verdaderos demuestran antigenicidad frente a predictor de epítomos de células T

Figura 13: Inmunogenicidad de epítomos de células T y descripciones fisicoquímicas de los péptidos candidatos para su uso en métodos de diagnóstico

Estructura	Péptido	Mer	Puntaje	Peso molecular (g/mol)	pl teórico	Estabilidad
Proteína S	LKPFERDISTEIYQA	15	0.44198	1810.04	4.68	si
	LKPFERDISTEI	12	0.34883	1447.65	4.68	no
	KPFERDISTEI	11	0.27588	1334.49	4.68	no
	NLKPFERDISTE	12	0.21068	1448.6	4.68	si
	NLKPFERD	8	0.17077	1018.14	6.07	si
	NYNLYRLFRK	11	0.14878	1549.8	9.99	si
	FERDISTEI	9	0.10425	1109.2	4.14	no
	YNLYRLFR	9	0.0918	1307.52	9.7	si
	NYNLYRLF	9	0.0171	1265.43	8.5	so
	NYNLYRLFRKSNLKPFERDISTEIYQA	28	0.01656	3543	9.31	si
	RKSNLKPFERDISTE	15	3.00E-05	1820.04	8.59	si
	Nucleoproteína	NKDGIIWATEGALNTPK	18	0.81029	1927.19	6.07
NKDGIIWATEGALN		15	0.78824	1600.79	4.37	si
LSPRWYFYLTGTGPEAGL		18	0.59665	2090.37	6	no
SPRWYFYLYL		9	0.34101	1294.47	8.22	no
ATEGALNTPK		10	0.11409	1001.1	6.05	si
Proteína de membrana	GAVILRGHLRIAGHHLGRCDIKDLP	25	0.53526	2718.22	9.5	si
	GAVILRGHLRIAGHH	15	0.50495	1606.9	12	si
	SELVIGAVILRGHLR	15	0.50076	1632.97	9.34	no
	SELVIGAVIL	10	0.34997	1013.24	4	si
	HLRIAGHHLGR	11	0.29421	1266.47	12	si
	IAGHHLGRCDIKDLP	15	0.05559	1644.91	6.91	si

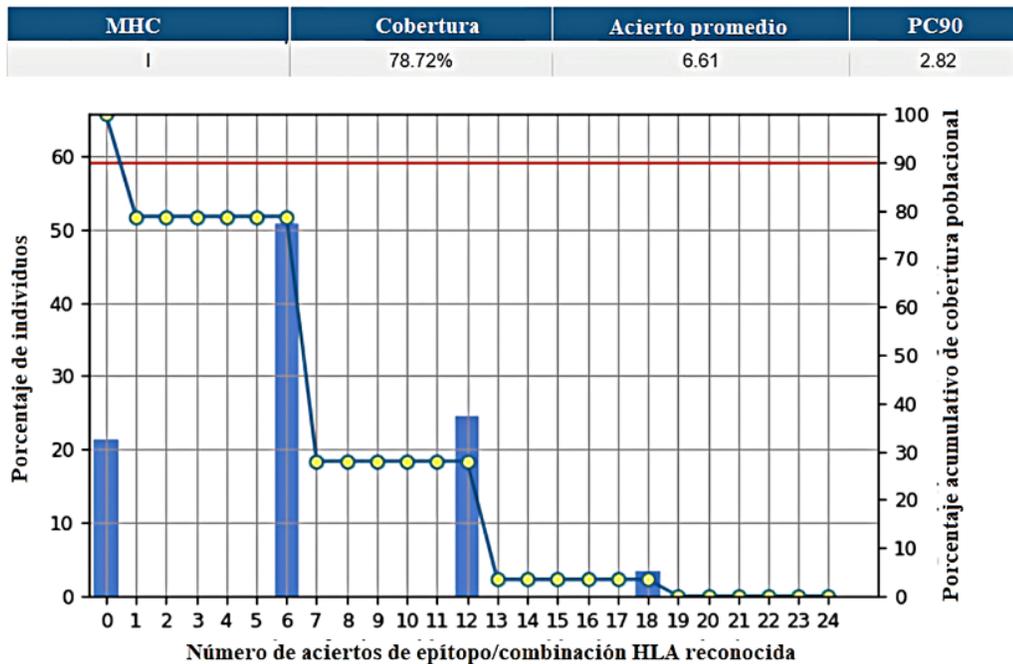
Descripción: En gris se presentan las secuencias consenso que fueron generadas con la herramienta de agrupamiento y validadas con el resto del análisis. El resto de las secuencias son los epítomos verdaderos que derivan de los consensos.

Las secuencias consenso seleccionadas y los epítomos verdaderos fueron sometidos a un predictor, el cual considera un índice positivo como una alta probabilidad de cada una de las determinadas secuencias de ser un epítomo de células T. A continuación, se muestran en la Figura 13 estas secuencias. Los péptidos que se proponen poseen longitudes mínimas de 8 mer y máximas de 28 mer. Adicionalmente se muestran los puntos isoeléctricos teóricos.

Esta información es relevante al momento de proponer métodos de extracción y purificación de los péptidos. Se observa que los pI son variables y abarcan tanto pHs ácidos como alcalinos. Otro dato importante es el índice de estabilidad obtenido por medio de Expsy, el cual indica si el péptido es estable o inestable en condiciones *in vitro*. Si bien no se descartaron las secuencias que arrojaron resultados de inestabilidad, es necesario tomar en cuenta esta característica al momento de seleccionar uno de los epítomos que se proponen.

6.5. Epítomos propuestos muestran una amplia cobertura para alelos HLA en población guatemalteca

Figura 14: Probabilidad de cobertura de los epítomos candidatos en la población guatemalteca



Descripción: A medida que aumenta el número de aciertos disminuye el porcentaje de individuos; sin embargo, sigue aumentando el porcentaje acumulado.

Los alelos HLA son altamente polimórficos, por lo que es necesario realizar análisis predictivos más específicos dependiendo de la ocurrencia de alelos en determinadas poblaciones. Por este motivo, el análisis de cobertura resulta una herramienta valiosa para el desarrollo de diagnósticos basados en antígenos. Para este análisis se emplearon 4 alelos de mayor ocurrencia dentro de la población guatemalteca. Los alelos HLA-A*02, HLA-A*24, HLA-A*68 y HLA-B*35 se muestran más a detalle con sus respectivas frecuencias en anexos (cuadro 16). La figura 14 refleja el número de aciertos de epítomos reconocidos por al menos una de las combinaciones de HLA. En al menos el 50% de la

población podrían ser reconocidas 6 combinaciones. A medida que se aumenta el número de aciertos, se observa que la probabilidad de reconocer más combinaciones a la vez disminuye. También vemos que el porcentaje de individuos que no reconocería ninguna de estas combinaciones es del 20 %.

Esta cifra podría parecer alta, pero se deben de tener en consideración otros parámetros de la prueba. Por ejemplo, el porcentaje acumulativo de cobertura población es del 78.72%; en otras palabras, se esperaría que con esa combinación de alelos, los epítotos sean reconocidos por las moléculas de MHC en el porcentaje poblacional ya mencionado. Adicionalmente el PC90 indica el número mínimo de aciertos de epítoto/ combinación HLA, el cual es de 2.82 aciertos.

6.6. Epítotos candidatos se encuentran en regiones altamente conservadas entre variantes de SARS-CoV-2

Figura 15: Alineamiento global de secuencias de las proteínas S, N y M de variantes de SARS-CoV-2

Gamma_SARS-CoV-2_Espiga	IAWNSNNLDSKVGGNYYNLYRLFRKSNLKPFERDISTEIQAGSTPCNGVKGFNCYFPLQ	7495
Delta_SARS-CoV-2_Espiga	IAWNSNNLDSKVGGNYYNLYRLFRKSNLKPFERDISTEIQAGSKPCNGVEGFNCYFPLQ	7489
Alfa_SARS-CoV-2_Espiga	IAWNSNNLDSKVGGNYYNLYRLFRKSNLKPFERDISTEIQAGSTPCNGVEGFNCYFPLQ	7491
Beta_SARS-CoV-2_Espiga	IAWNSNNLDSKVGGNYYNLYRLFRKSNLKPFERDISTEIQAGSTPCNGVEGFNCYFPLQ	7505

Gamma_SARS-CoV-2_Np	TQHGKEDLKFPRGQGVPIINTSSPDDQIGYYRRATRRIRGGDGKMKDLSPRWYFYLLGTG	9282
Delta_SARS-CoV-2_Np	TQHGKEGLKFPRGQGVPIINTSSPDDQIGYYRRATRRIRGGDGKMKDLSPRWYFYLLGTG	9274
Alfa_SARS-CoV-2_Np	TQHGKEDLKFPRGQGVPIINTSSPDDQIGYYRRATRRIRGGDGKMKDLSPRWYFYLLGTG	9279
Beta_SARS-CoV-2_Np	TQHGKEDLKFPRGQGVPIINTSSPDDQIGYYRRATRRIRGGDGKMKDLSPRWYFYLLGTG	9292

Gamma_SARS-CoV-2_Np	PEAGLPYGANKHGIIWVATEGALNTPKDHIGTRNPANNAIIVLQLPQGTTLPKGFYAEGS	9342
Delta_SARS-CoV-2_Np	PEAGLPYGANKDGIWVATEGALNTPKDHIGTRNPANNAIIVLQLPQGTTLPKGFYAEGS	9334
Alfa_SARS-CoV-2_Np	PEAGLPYGANKDGIWVATEGALNTPKDHIGTRNPANNAIIVLQLPQGTTLPKGFYAEGS	9339
Beta_SARS-CoV-2_Np	PEAGLPYGANKDGIWVATEGALNTPKDHIGTRNPANNAIIVLQLPQGTTLPKGFYAEGS	9352

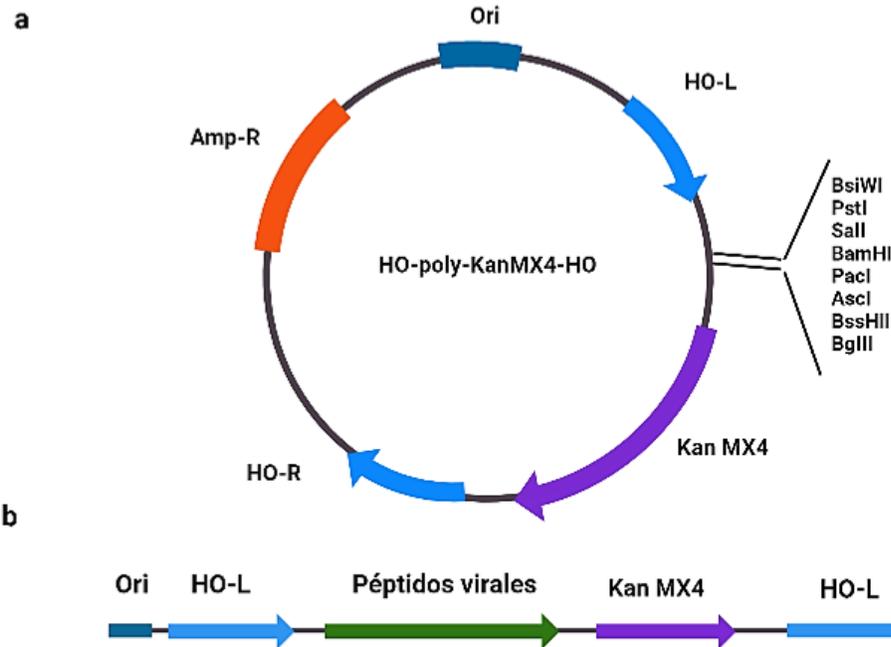
Gamma_SARS-CoV-2_M	.VIGAVILRGHLRIAGHHLGRCDIKDLPKEITVATSRTLSTYKLGASQRVAGDSGFAAYS	3540
Delta_SARS-CoV-2_M	.VIGAVILRGHLRIAGHHLGRCDIKDLPKEITVATSRTLSTYKLGASQRVAGDSGFAAYS	3540
Alfa_SARS-CoV-2_M	.VIGAVILRGHLRIAGHHLGRCDIKDLPKEITVATSRTLSTYKLGASQRVAGDSGFAAYS	3540
Beta_SARS-CoV-2_M	.VIGAVILRGHLRIAGHHLGRCDIKDLPKEITVATSRTLSTYKLGASQRVAGDSGFAAYS	3540

Descripción: Se muestran los alineamientos parciales donde se encuentran resaltadas las secuencias consenso. Los aminoácidos marcados en rojo representan un cambio respecto al resto de variantes.

En la Figura 15 se presentan los alineamientos parciales de SARS-CoV-2 (alfa, beta, delta y gamma) en donde se puede observar un alto grado de conservación en las regiones donde se encuentran ubicados los epítotos consenso de las proteínas S, N y M. Las únicas excepciones que se observan son un cambio de residuo L por R para la variante delta en el epítoto de la proteína S y un cambio de residuo D por H en la nucleoproteína de la variante gamma. En los casos del epítoto de la proteína M y el segundo de la proteína N, comparten los mismo residuos entre todas las variantes.

6.7. Diseños teóricos de los vectores se presentan con un enfoque de proteínas de fusión y constructos policistrónicos

Figura 16: Diagrama general del vector HO-Poly-KanMX4-HO y el casete de integración

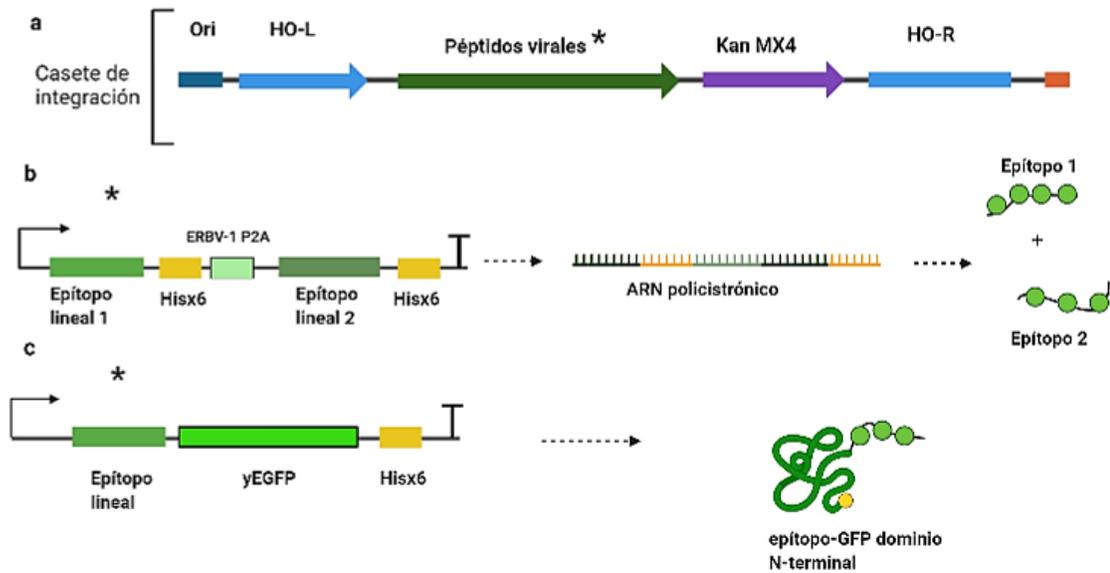


Descripción:(a) Se observa el mapa general donde se presenta el gen de resistencia a ampicilina para su selección en *E.coli*. El casete de integración posee secuencias afines al locus HO y un sitio de clonación múltiple justo antes del gen de resistencia a kanamicina.(b) Se observa una ampliación del casete de integración.

Las secuencias completas de los circuitos propuestos se encuentran con más detalle en los cuadros 22, 23 y 24 de anexos. El vector HO-Poly-KanMX4-HO es un plásmido integrativo que cuenta con la ventaja de poder clonarse en *E. coli*. Además, posee un marcador de selección por kanamicina que facilita la separación de transformantes, a diferencia de los métodos de selección auxotróficos que requieren de cepas y medios sintéticos muy específicos. Para el primer constructo que se muestra en la Figura 17a se propone un circuito policistrónico que funciona bajo el promotor de HO. Esto se logra empleando las secuencias de péptidos autoescindibles que inducen el salto ribosómico. Además, los epítomos con una etiqueta de poli histidinas para facilitar su extracción y purificación. La ventaja de este constructo es que permite producir múltiples péptidos bajo un mismo promotor en cantidades aproximadamente equimolares. Esto facilitaría establecer una producción con péptidos heterogéneos y así aumentar la sensibilidad de la prueba.

Para el segundo diseño que se muestra en la Figura 17b, se propone una proteína quimérica empleando el mismo vector base. El objetivo de esto es poder plantear una prueba de flujo lateral con un principio similar al de las nanopartículas de oro. En este caso, el funcionamiento sería mediante la fijación del epítomo-GFP en la matriz de celulosa. Al añadir una muestra con anticuerpos afines a los antígenos, estos se unirán y migrarán hacia una línea de prueba en la matriz que contenga anti-Ig(G,M o A) y una segunda línea de control con anti-GFP.

Figura 17: Diagrama del casete de integración HO-Poly-KanMX4-HO y constructos propuestos para la producción de epítomos



Descripción: (a) Diagrama general del casete de integración. (b) Constructo policistrónico para la producción de múltiples epítomos bajo un mismo promotor. (c) La secuencia optimizada del péptido NYNYLYRLFRKSNLKPFRDISTEIIYQA de la proteína S se encuentra en el extremo N-terminal de la GFP. El marcador poli histidina permite la purificación de la proteína de fusión por medio de columnas cromatográfica.

6.8. Obtención de transformantes mediante SOP de electroporación

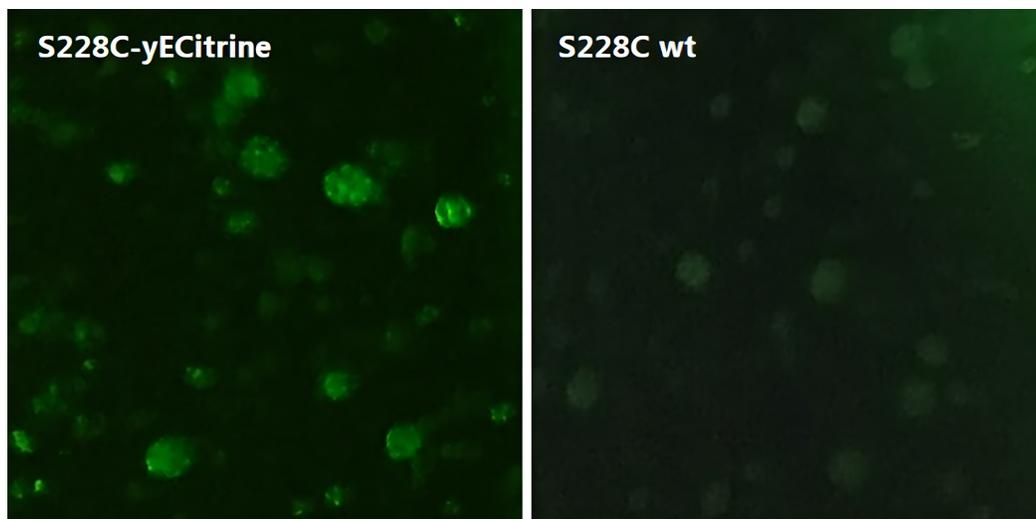
Figura 18: Crecimiento de transformantes de S228C con el plásmido pKT175



Descripción: Colonias de S228C expuestas a luz ultravioleta en trans-iluminador. A la izquierda se presentan levaduras nativas. A la derecha se observan colonias transformadas emitiendo fluorescencia.

Se llevó a cabo la transformación de *S.cerevisiae* S228C empleando el plásmido pKT175 el cual codifica para el gen de la yECitrine, una versión modificada de la proteína verde fluorescente. Este plásmido contiene un marcador de selección auxotrófico constituido por el gen *ura3*. Por lo tanto, para realizar una selección efectiva es necesario contar con cepas *ura3*-; sin embargo, la cepa S228C disponible no poseía estas características, por lo que se procedió con el SOP empleando medio YPD no selectivo. El medio YPD es un medio completo que posee todos los elementos necesarios para que la mayoría de las levaduras se desarrollen.

Figura 19: Observación de transformantes por microscopía de fluorescencia

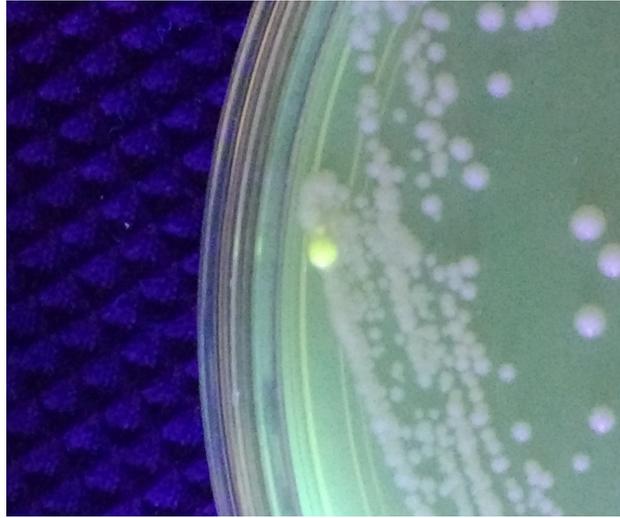


Descripción: Se presentan células de S228C transformadas (S228C-yECitrine) y levaduras nativas (wt) bajo un aumento de 1000X y una longitud de onda de excitación de 325 nm.

En la Figura 18 se observa una placa con S228C nativa (izquierda) y S228C-pKT175 (derecha). Es posible distinguir la fluorescencia de las colonias transformadas al ser excitadas bajo la luz UV de un trans-iluminador. A pesar de no tener un medio de selección apropiado se determinó una eficiencia de transformación de 7.2×10^{-10} UFC/ μ g. Por el mismo motivo de no tener un medio de selección se llevó a cabo una prueba confirmatoria de la transformación mediante microscopía de fluorescencia. Esto se observa en la Figura 19 donde se ve *S.cerevisiae* presentando fluorescencia a una longitud de onda de excitación de 325 nm. Es posible observar incluso algunas células en proceso de gemación. Pese a no existir presión selectiva, las células manifiestan una expresión de la yECitrine considerable. Cabe resaltar que estos resultados son meramente cualitativos y no son aptos para determinar grado de eficiencia en la transformación.

El plásmido pKT175 es un plásmido integrativo, por lo que posee una estabilidad media y tiende a perderse entre generaciones en ausencia de presión selectiva. Adicionalmente, al no contar con un sitio de integración específico para este diseño no es posible que ocurra el evento de recombinación homóloga. Para evaluar la pérdida del plásmido en ausencia de estos dos factores se hizo un re-aislamiento de una de las colonias transformadas mostradas en la Figura 18. Dicho re-aislamiento se muestra en la Figura 20 donde es notable la pérdida del plásmido, ya que solamente se observa una colonia expresando la yECitrine al ser iluminada bajo luz UV. Probablemente el plásmido fue capaz de mantenerse activo en el citoplasma de las células de esta colonia específica; sin embargo es de esperar que en las siguientes generaciones se pierda por completo.

Figura 20: Re-aislamiento de transformantes



Descripción: Se muestra transformante expresando yECitrine en re-aislado. La ausencia de presión selectiva y de secuencias homólogas para la recombinación acelera la pérdida del plásmido.

Análisis de resultados

En el presente estudio se proponen un total de 22 secuencias peptídicas que constituyen epítos lineales de SARS-CoV-2 con un potencial uso como biomarcadores para pruebas de antígeno-anticuerpo de flujo lateral. De estas 22 secuencias, 5 son un consenso que agrupan al resto de epítos verdaderos que han sido reportados previamente en estudios experimentales. Los epítos provienen de las proteínas S, N y M, las cuales han sido de mayor interés por parte de los científicos debido al rol fundamental que tienen en los procesos de invasión, procesamiento del genoma viral y auto-ensamblaje de los viriones (Kalita, Padhi, Zhang, y Tripathi, 2020). Otras proteínas accesorias con funciones igual de importantes como la ORF3A y la poliproteína replicasa 1ab, han demostrado ser candidatos viables con un alto número de epítos reportados con reacciones positivas frente a anticuerpos (Tan y cols., 2021); sin embargo, en las secuencias evaluadas había poca relación entre cada uno de estos epítos y para propósitos de la metodología planteada se descartaron, ya que se buscaba obtener aquellos epítos o regiones con mayor número de estudios.

Los epítos candidatos demostraron tener índices óptimos en los análisis de accesibilidad al solvente y predicción de epítos de células B y T. El éxito de este análisis recae en la aproximación de la metodología, la cual no busca predecir epítos desde cero, sino que plantea aprovechar la alta cantidad de información que se actualiza constantemente por medio de las bases de datos de péptidos previamente caracterizados y de estos conjuntos determinar los epítos óptimos para fines específicos. Esto no significa que los algoritmos posean baja eficiencia; de hecho, algunos estudios reportan que no existe cambio significativo entre emplear datos predichos en su totalidad por un algoritmo y emplear datos experimentales (Paul y cols., 2021), pero si es reconocido que aún existen limitaciones en el poder computacional, principalmente para la predicción de epítos de células B (Isea y cols., 2017).

Otra complicación que se debe tomar en cuenta con los análisis de predicción, es la dificultad de predecir epítos de células T por medio de modelos a causa de las cientos de variantes alélicas del MHC. Cada una de estas variantes se unen a diferentes grupos de epítos (Reche, Flower, Fridkis-Hareli, y Hoshino, 2018). Debido a esto, es necesario respaldar la inmunogenicidad del epítoto con su probabilidad de unirse a cada una de estas variantes. El resultado de cobertura poblacional arrojó un resultado prometedor al mostrar una cobertura acumulada del 78.2% de la población. No obstante, se debe tomar en consideración dos debilidades de este análisis: La información sobre la ocurrencia de alelos de la región de Guatemala es limitada, ya que dichas frecuencias obtenidas de la base de

datos se determinaron con un tamaño muestral $n = 127$.

La segunda debilidad recae en el algoritmo, ya que al emplearse una interfaz gráfica se pierde poder computacional y dificulta el análisis con muchas variantes alélicas a la vez; además, no permite analizar cada epítipo individual, sino que evalúa un agrupamiento completo. A pesar de estas limitantes, se considera que, al aumentar el número de variantes alélicas, también lo haría la probabilidad de los epítipos de ser reconocidos, por lo que los resultados siguen siendo prometedores.

A pesar de las limitaciones a considerar, una sorpresa fue que al momento de realizar una revisión de literatura se encontraron al menos tres estudios en los cuales caracterizaron experimentalmente el péptido NYNYLYRLFRKSNLKPFERDISTEI de la proteína S del SARS-CoV-2. El mismo no se encuentra registrado en IEDB, pero se describe como un posible candidato para la creación de vacunas y bioindicadores. Este exhibe una producción positiva para IFN- γ , así como una alta cobertura poblacional HLA (Feng y cols., 2020) (Lim y cols., 2021). Dicho péptido resulta casi idéntico a la secuencia consenso validada por el presente estudio (NYNYLYRLFRKSNLKPFERDISTEIQ), difiriendo únicamente en la inclusión de los residuos Tyr-Gln-Ala al final de la secuencia. Esta evidencia brinda soporte adicional a la validez de la metodología empleada para encontrar candidatos altamente inmunogénicos, así como la antigenicidad de la proteína de espiga. La accesibilidad al solvente también se sustenta en estos estudios y se demuestra el fácil acceso a esta región en específico como se observa en la Figura 21 de anexos.

De manera similar, se han reportado *hotspots* de epítipos lineales capaces de ser reconocidos por fuertes uniones antígeno-anticuerpo en la nucleoproteína y proteína de membrana (Grifoni y cols., 2020). Por lo tanto, en este estudio se propusieron candidatos para las tres proteínas estructurales. No obstante, existen diferencias en cuanto a la abundancia y especificidad que necesitarían ser determinadas. Generalmente se ha reportado que la nucleocápside y proteína de membrana son las proteínas virales más abundantes en una infección por SARS-CoV-2. Esto significa que son más fáciles de detectar en cuanto a grado de sensibilidad. En contraparte, la proteína de espiga es menos abundante, pero es la estructura más divergente entre coronavirus, por lo que en términos de especificidad ofrece una ventaja respecto al resto de estructuras (Poh y cols., 2020). Por tal circunstancia, en casos que se emplean epítipos lineales se recomienda utilizar una combinación de estos para ampliar el rango y la probabilidad de detección (Amrun y cols., 2020).

Un reto que existe al momento de validar epítipos para diagnóstico es la alta tasa de mutación que poseen los hCoV. El riesgo de que la secuencia de un péptido cambie entre cepas es alto, principalmente para la proteína de espiga (De Maio y cols., 2021). Los resultados de alineamiento demuestran que existe un alto grado de conservación de las secuencias consenso entre las cuatro principales variantes. En los casos en los que la conservación no era del 100%, únicamente un residuo de aminoácido era el que se veía afectado. A pesar de esto, cabe destacar que no es estricto que un anticuerpo reconozca la totalidad del péptido, sino solo una fracción como se observa en los diversos epítipos verdaderos que derivan de estas secuencias. Por este motivo no se descarta que cuando existan cambios de residuo mínimos, aún pueda ocurrir una reacción positiva.

Para afrontar el reto de producir múltiples epítipos, se propuso un constructo genético policistrónico, el cual emplea un péptido auto escindible P2A. Este péptido ha demostrado poseer una alta tasa de escisión específicamente en *S. cerevisiae* (Souza-Moreira y cols., 2018). Si bien esta secuencia no ha sido evaluada para traducir más de tres péptidos bajo un mismo promotor, generalmente se utiliza para proteínas de gran tamaño, por lo que evaluar su comportamiento para péptidos cortos aportaría información valiosa. Lo prometedor de este diseño está en que existe la posibilidad de que pudieran producirse múltiples péptidos cortos bajo un mismo promotor, en un mismo vector. Se hipotetiza que facilitaría y reduciría significativamente los costos y esfuerzos para un proceso de producción y purificación; no obstante, sería una línea de investigación que debe evaluarse más a profundidad. El otro constructo propone fusionar al péptido antigénico con un bioindicador, en este caso la GFP para que sea reconocido por una anti-GFP. Esto permitiría sustituir el mecanismo de agregación de nanopartículas de oro y así reducir costos. La desventaja de esto es que únicamente

se produce un epítopo a la vez; sin embargo, no se descarta la posibilidad de poder emplear la secuencia P2A para producir múltiples epítopo-GFP. En cuanto a la selección del vector para el diseño teórico, se optó por un plásmido integrativo debido a su estabilidad y capacidad de controlar el número de copias. Esto se logra mediante la selección de un locus para su integración. En este caso el locus HO es un candidato que ha sido muy utilizado como sitio de integración debido a que este no afecta el desarrollo ni la tasa de crecimiento celular (Voth, Richards, Shaw, y Stillman, 2001a). Otra ventaja del plásmido integrativo es la capacidad de poder realizar múltiples inserciones, lo que podría aumentar el grado de transcripción de los genes de interés. En este caso con el gen HO se produce una sola inserción; no obstante, es posible realizar inserciones repetidas o bien para futuros diseños considerar un sitio de inserción en tándem (Gnügge, Liphardt, y Rudolf, 2016).

En cuanto a la transformación de *S.cerevisiae*, bajo los parámetros de preparación y electroporación, se obtuvieron colonias que expresaban la yECitrine. Una desventaja del plásmido empleado fue que no se contaba con el medio de selección auxotrófico, por lo que no se obtuvieron transformantes totalmente aislados. Al momento de realizar un reaislado, únicamente se logró recuperar un transformante. Esto pudo ocurrir debido a que el vector pKT175 es un plásmido integrativo, es decir que se integra al genoma por recombinación homóloga. Estos plásmidos poseen una estabilidad media y requieren de presión selectiva para incentivar la integración al genoma. Al no haber presión selectiva es probable que conforme se reproducen las células, el plásmido se pierda (Voth, Richards, Shaw, y Stillman, 2001b). Sin embargo, el factor más determinante es la carencia de un sitio de recombinación para inducir la integración de este, por lo que la pérdida del plásmido era bastante predecible (Bai Flagfeldt, Siewers, Huang, y Nielsen, 2009). Aun así, los resultados cumplen con el objetivo de determinar si este protocolo de electroporación podría emplearse para la transformación de *S.cerevisiae* y así utilizarse como medio para la producción de los epítopos. Sin embargo, se debe tomar en cuenta que estos resultados son estrictamente cualitativos y se requieren de estudios adicionales con medios de selección adecuados para cuantificar la eficiencia de transformación y poder llevar a cabo una optimización del protocolo.

Aunado a la situación anterior, cabe destacar que el principal motivo por el cual se empleó el vector pKT175 está relacionado estrictamente a la disponibilidad que se tenía de este al momento de llevar a cabo el presente estudio. Además, al igual que el vector HO-poly-KanMX4-HO, es un plásmido integrativo que cuenta con un marcador de resistencia a ampicilina para su clonación en *E.coli* y también podría emplearse para la producción de epítopos añadiendo sitios de recombinación. No obstante, es importante mencionar que si bien el protocolo fue probado con el vector pKT175 es necesario para futuros estudios realizar validaciones para el vector teórico HO-poly-HO. Esto se debe a que es necesario determinar la eficiencia de integración y posiblemente el efecto más determinante, el tamaño del vector teórico. Este plásmido difiere en aproximadamente 1.1 kb más respecto al pKT175. Es de conocimiento general que el tamaño de un vector puede tener impactos significativos en la eficiencia de transformación, entendiéndose que mientras mayor sea el tamaño del ADN, menor será la eficiencia. Este efecto se ha estudiado con diferentes sistemas de expresión (Green y Sambrook, 2021). Sin embargo, se ha reportado también que en determinados organismos no existe una diferencia significativa en la eficiencia de transformación para vectores en un rango de 3.7-12.6 kb. Por lo tanto, a pesar de la diferencia de tamaño, únicamente evaluando el vector se podrá determinar el cambio en la eficiencia (Ohse, Takahashi, Kadowaki, y Kusaoke, 1995)

Para culminar la discusión, un último punto que amerita ser destacado es que la metodología evaluada en este estudio posee el valor añadido de que puede extrapolarse del diseño de pruebas de diagnóstico al desarrollo de vacunas basadas en péptidos cortos. En este caso se le denominaría vacunología reversa, ya que se parten de análisis bioinformáticos para encontrar secuencias candidatas a vacunas. Si bien la vacunología reversa es un campo relativamente nuevo, estudios han demostrado que las vacunas de epítopos son capaces de proveer un efecto terapéutico óptimo con efectos secundarios reducidos. Además, con estas herramientas bioinformáticas se ha sido capaz de obtener péptidos que contienen tanto epítopos de células B como T y que pueden interactuar (hasta un 99%) con los alelos HLA más prevalentes de determinadas regiones (Feng y cols., 2020).

El presente estudio propone 22 péptidos de las proteínas S,N y M del SARS-CoV-2 con alta antigenicidad y alta cobertura poblacional. Estas características les otorgan a los candidatos una alta probabilidad para poder ser empleados como antígenos en pruebas de diagnóstico. Además, la metodología utilizada para la selección y validación de epítomos muestra poseer una ventaja sobre métodos convencionales de predicción, ya que los resultados bioinformáticos se sustentan con los datos experimentales provenientes de la cooperación científica. Esto representa una reducción en el esfuerzo experimental y en los costos, lo que permitiría concentrar recursos a la parte del desarrollo biotecnológico. Las características que proveen los análisis bioinformáticos resultan beneficiosas principalmente en países como Guatemala en los cuales existe limitada infraestructura e inversión para la investigación. Por otro lado, los constructos genéticos diseñados presentan una alternativa para facilitar la producción de múltiples péptidos por medio del modelo policistrónico, o bien la producción de un péptido biomarcador para facilitar el mecanismo de detección. Por otra parte, el funcionamiento del SOP de electroporación abre campo para estudiar la posibilidad de emplear *S.cerevisiae* para la producción de estos péptidos recombinantes mediante su transformación y podría resultar una técnica que requiere de poca preparación e inversión de tiempo. Sin embargo, para poder validar esto es necesario en primer lugar probar el funcionamiento del plásmido y seguidamente demostrar que *S.cerevisiae* es un candidato adecuado para la expresión óptima de estos epítomos aplicando criterios como la costo-efectividad de la producción.

Recomendaciones

Una característica que se puede reforzar de la metodología es el uso de interfaces gráficas, las cuales reducen el poder computacional y limitan el número de datos que pueden analizarse a la vez. Por lo tanto, la principal recomendación para mejorar el proceso de análisis es por medio del diseño de un algoritmo que permita automatizar la manipulación e interpretación de datos. Dentro de esto se incluye aumentar el número de péptidos analizados, ya que en el presente estudio se evaluó un total de 240 secuencias, pero idealmente se debería estudiar la totalidad de las secuencias disponibles. Adicionalmente es recomendable aumentar el número de alelos HLA evaluados para el análisis de la cobertura poblacional. Finalmente, antes de llevar a cabo la construcción *in vitro* del vector para la expresión en levaduras, se debe optimizar la transformación por electroporación con el vector propuesto para asegurar la mayor cantidad de transformantes en el proceso. Posteriormente se tendría que evaluar la efectividad del caset de integración en cuanto a estabilidad y eficiencia de transcripción para así poder validar si *S. cerevisiae* resulta un sistema óptimo para la producción de los epítomos.

Referencias

- Ahammed, A. R., Satu, y Islam. (2020). Early detection of coronavirus cases using chest x-ray images employing machine learning and deep learning approaches. *medRxiv*, 10(1), 6.
- Alagarasu, L. A. P., Choudhary, y Team, N. (2020). Evaluation of rdrp y orf-1b-nsp14-based real-time rt-pcr assays for confirmation of sars-cov-2 infection: An observational study. *The Indian Journal of Medical Research*, 151(5), 483.
- Alvarado Orozco, M., Rangel Aguilera, J. M., Barba Hernández, B. A., García Baeza, L. C., y Sánchez Navarro, M. d. R. (2021). Seroprevalencia y seguimiento de anticuerpos igg contra sars-cov-2 en personal del laboratorio estatal de salud pública de guanajuato laesap. *Nova scientia*, 13(SPE), 0–0.
- Alvargonzalez, y Cao. (2020). Pooling for sars-cov-2 control in care institutions. *BMC infectious diseases*, 20(1), 1-6.
- Amrun, S. N., Lee, C. Y.-P., Lee, B., Fong, S.-W., Young, B. E., Chee, R. S.-L., ... others (2020). Linear b-cell epitopes in the spike and nucleocapsid proteins as markers of sars-cov-2 exposure and disease severity. *EBioMedicine*, 58, 102911.
- Andronico, A., Kiem, C. T., Paireau, J., Succo, T., Bosetti, P., Lefrancq, N., ... others (2021). Evaluating the impact of curfews and other measures on sars-cov-2 transmission in french guiana. *Nature communications*, 12(1), 1–8.
- Anné, J., Maldonado, B., Van Impe, J., Van Mellaert, L., y Bernaerts, K. (2012). Recombinant protein production and streptomycetes. *Journal of biotechnology*, 158(4), 159–167.
- Baganz, F., Hayes, A., Marren, D., Gardner, D. C., y Oliver, S. G. (1997). Suitability of replacement markers for functional analysis studies in *saccharomyces cerevisiae*. *Yeast*, 13(16), 1563–1573.
- Baghban, R., Farajnia, S., Rajabibazl, M., Ghasemi, Y., Mafi, A., Hoseinpoor, R., ... Aria, M. (2019). Yeast expression systems: overview and recent advances. *Molecular biotechnology*, 61(5), 365–384.
- Bahadır, E. B., y Sezgintürk, M. K. (2016). Lateral flow assays: Principles, designs and labels. *TrAC Trends in Analytical Chemistry*, 82, 286–306.
- Bai Flagfeldt, D., Siewers, V., Huang, L., y Nielsen, J. (2009). Characterization of chromosomal integration sites for heterologous gene expression in *saccharomyces cerevisiae*. *Yeast*, 26(10), 545–551.
- Bauer. (2021). The variability of the serological response to sars-corona virus-2: Potential resolution of ambiguity through determination of avidity (functional affinity). *Journal of Medical Virology*, 93(1), 311-322.

- Buckhout-White, M., Person, y Goldman. (2018). Restriction enzymes as a target for dna-based sensing and structural rearrangement. *ACS omega*, 3(1), 495-502.
- Calis, J. J., Maybeno, M., Greenbaum, J. A., Weiskopf, D., De Silva, A. D., Sette, A., ... Peters, B. (2013). Properties of mhc class i presented peptides that enhance immunogenicity. *PLoS computational biology*, 9(10), e1003266.
- Calvache, M., Rodríguez, y Paucar. (2020). Utilidad de pruebas de cadena de polimerasa, pruebas rápidas y tomografías en pacientes con covid-19. *Journal of America health*, 3(2), 32-39.
- Camps, M. (2010). Modulation of cole1-like plasmid replication for recombinant gene expression. *Recent Patents on DNA & Gene Sequences (Discontinued)*, 4(1), 58-73.
- Chang, W., Guo, y He. (2017). Extracellular expression of alkali tolerant xylanase from bacillus subtilis lucky in e. coli and application for xylooligosaccharides production from agro-industrial waste. *International journal of biological macromolecules*, 96(1), 249-256.
- Chen, R. (2012). Bacterial expression systems for recombinant protein production: E. coli and beyond. *Biotechnology advances*, 30(5), 1102-1107.
- Dana, R. (2018). Advances and innovations in recombinant protein expression technology. *J. Adv. Pharm. Educ. Res.*, 89.
- De Maio, N., Walker, C. R., Turakhia, Y., Lanfear, R., Corbett-Detig, R., y Goldman, N. (2021). Mutation rates and selection on synonymous mutations in sars-cov-2. *Genome Biology and Evolution*, 13(5), evab087.
- Dhanda, S. K., Vaughan, K., Schulten, V., Grifoni, A., Weiskopf, D., Sidney, J., ... Sette, A. (2018). Development of a novel clustering tool for linear peptide sequences. *Immunology*, 155(3), 331-345.
- Diamandis, E. P., y Christopoulos, T. K. (1996). *Immunoassay*. Academic Press.
- Ergün, B. G., Hücetoğulları, D., Öztürk, S., Çelik, E., y Çalık, P. (2019). Established and upcoming yeast expression systems. *Recombinant Protein Production in Yeast*, 1-74.
- Feng, Y., Qiu, M., Zou, S., Li, Y., Luo, K., Chen, R., ... others (2020). Multi-epitope vaccine design using an immunoinformatics approach for 2019 novel coronavirus in china (sars-cov-2). *BioRxiv*.
- Forsstrom, R. D. B., Axnas, y Uhlen. (2015). Dissecting antibodies with regards to linear and conformational epitopes. *PLOS ONE*, 10(3), 3-20.
- Funakoshi, M., y Hochstrasser, M. (2009). Small epitope-linker modules for pcr-based c-terminal tagging in saccharomyces cerevisiae. *Yeast*, 26(3), 185-192.
- Gao, C. Z. H.-W., Qin, y Cui. (2021). Crystal structure of sars-cov-2 papain-like protease. *Pharmaceutica Sinica*, 11(1), 237-245.
- Garcia, B. B. G.-R., Alarcon, y Atun. (2020). Covid-19 response in latin america. *The American Journal of Tropical Medicine and Hygiene*, 103(5), 1765.
- Gasteiger, E., Hoogland, C., Gattiker, A., Wilkins, M. R., Appel, R. D., Bairoch, A., y cols. (2005). Protein identification and analysis tools on the expasy server. *The proteomics protocols handbook*, 571-607.
- Ghio, P. V. T.-G., Ontañon, y Campos. (2018). Paenibacillus sp. a59 gh10 and gh11 extracellular endoxylanases: application in biomass bioconversion. *BioEnergy Research*, 11(1), 174-190.
- Gnügge, R., Liphardt, T., y Rudolf, F. (2016). A shuttle vector series for precise genetic engineering of saccharomyces cerevisiae. *Yeast*, 33(3), 83-98.
- Gnügge, R., y Rudolf, F. (2017). Saccharomyces cerevisiae shuttle vectors. *Yeast*, 34(5), 205-221.
- Gonzalez-Galarza, F. F., McCabe, A., Santos, E. J. M. d., Jones, J., Takeshita, L., Ortega-Rivera, N. D., ... others (2020). Allele frequency net database (afnd) 2020 update: gold-standard data classification, open access genotype data and new query tools. *Nucleic acids research*, 48(D1), D783-D788.
- Granados, D., y Giraldo, C. (2012). Métodos de transformación genética de plantas. , 1(1), 1-8.
- Green, M. R., y Sambrook, J. (2021). Cloning and transformation with plasmid vectors. *Cold Spring Harbor Protocols*, 2021(11), pdb-top101170.
- Grifoni, A., Sidney, J., Zhang, Y., Scheuermann, R. H., Peters, B., y Sette, A. (2020). A sequence homology and bioinformatic approach can predict candidate targets for immune responses to sars-cov-2. *Cell host & microbe*, 27(4), 671-680.

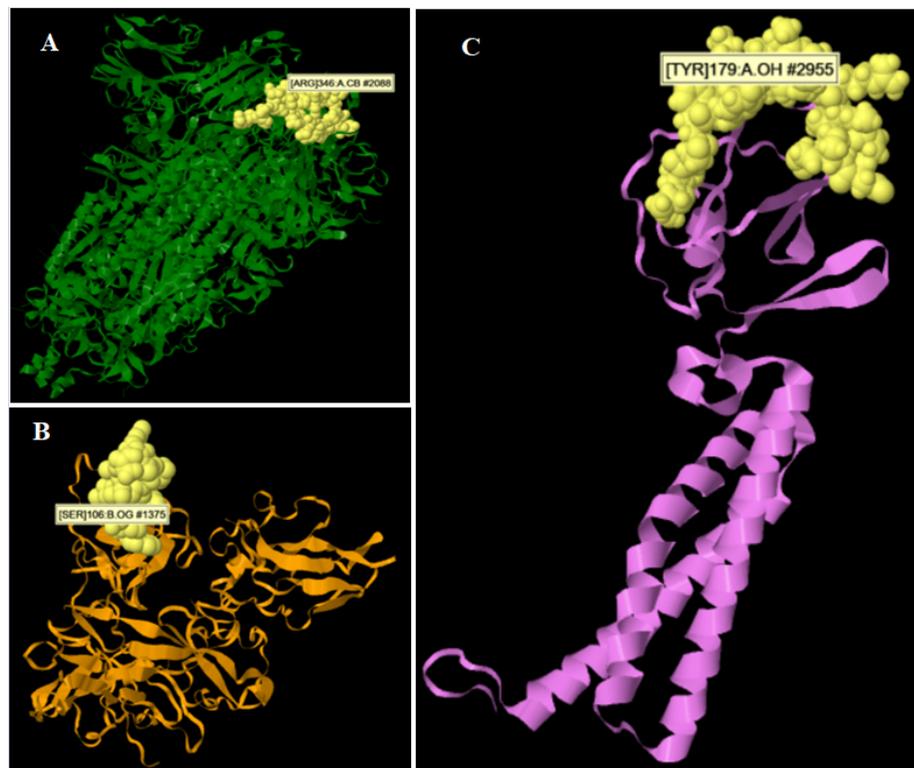
- Grote, A., Hiller, K., Scheer, M., Münch, R., Nörtemann, B., Hempel, D. C., y Jahn, D. (2005). Jcat: a novel tool to adapt codon usage of a target gene to its potential expression host. *Nucleic acids research*, 33(suppl_2), W526–W531.
- Gupta. (2014). *Drug discovery today*, 19(5), 590-601.
- Guest, D. R., y Sanchez. (2020). The three steps needed to end the covid-19 pandemic: bold public health leadership, rapid innovations, and courageous political will. *ACS nano*, 6(2), 1.
- Gunby. (1978). Bacteria directed to produce insulin in test application of genetic code. *JAMA*, 240(16), 1697-1698.
- Hochrein, L., Machens, F., Messerschmidt, K., y Mueller-Roeber, B. (2017). Phirex: a programmable and red light-regulated protein expression switch for yeast. *Nucleic acids research*, 45(15), 9193–9205.
- Hu, H., y Yin. (2020). The cytokine storm and covid 19. *Journal of medical virology*, 93(1), 250-256.
- Isea, R., Cardinale, Y., y Ortega, A. (2017). Predicción de epítosos consenso de células b en el virus de la influenza h5n1. *Conocimiento Libre y Licenciamiento (CLIC)*(15).
- Ito, H., Fukuda, Y., Murata, K., y Kimura, A. (1983). Transformation of intact yeast cells treated with alkali cations. *Journal of bacteriology*, 153(1), 163–168.
- Johari, Y. B., Jaffé, S. R., Scarrott, J. M., Johnson, A. O., Mozzanino, T., Pohle, T. H., ... others (2021). Production of trimeric sars-cov-2 spike protein by cho cells for serological covid-19 testing. *Biotechnology and bioengineering*, 118(2), 1013–1021.
- Kalita, P., Padhi, A. K., Zhang, K. Y., y Tripathi, T. (2020). Design of a peptide-based subunit vaccine against novel coronavirus sars-cov-2. *Microbial Pathogenesis*, 145, 104236.
- Kandeel, K. Y. F.-M. V.-K. I., Mahmoud, y Abdelazim. (2021). The emerging sars-cov-2 papain-like protease: Its relationship with recent coronavirus epidemics. *Journal of Medical Virology*, 93(3), 1581–1588.
- Kang, S., Yang, M., Hong, Z., Zhang, L., Huang, Z., Chen, X., ... others (2020). Crystal structure of sars-cov-2 nucleocapsid protein rna binding domain reveals potential unique drug targeting sites. *Acta Pharmaceutica Sinica B*, 10(7), 1228–1238.
- Kang, Y. N. S., Lim, y Jung. (2019). Synthetic auxotrophs for stable and tunable maintenance of plasmid copy number. *Metabolic engineering*, 48(1), 121-128.
- Khateeb, L., y Zhang. (2021). Emerging sars-cov-2 variants of concern and potential intervention approaches. *Critical Care*, 25(1), 1-8.
- Kim, H., Yoo, S. J., y Kang, H. A. (2015). Yeast synthetic biology for the production of recombinant therapeutic proteins. *FEMS yeast research*, 15(1), 1–16.
- Klein, S., Cortese, M., Winter, S. L., Wachsmuth-Melm, M., Neufeldt, C. J., Cerikan, B., ... Chlanda, P. (2020). Sars-cov-2 structure and replication characterized by in situ cryo-electron tomography. *Nature communications*, 11(1), 1–10.
- Kolaskar, A. S., y Tongaonkar, P. C. (1990). A semi-empirical method for prediction of antigenic determinants on protein antigens. *FEBS letters*, 276(1-2), 172–174.
- Langa, S., y Díez. (2021). Interpretación de las pruebas diagnósticas de la covid-19. *FMC-Formación Médica Continuada en Atención Pri*, 1(1), 1.
- Larremore, B. Z. K.-M., Fosdick, y Grad. (2021). Estimating sars-cov-2 seroprevalence and epidemiological parameters with uncertainty from serological surveys. *Elife*, 10(1), 509-514.
- Li, Y., Ma, M.-l., Lei, Q., Wang, F., Hong, W., Lai, D.-y., ... others (2021). Linear epitope landscape of the sars-cov-2 spike protein constructed from 1,051 covid-19 patients. *Cell reports*, 34(13), 108915.
- Li, Z., Yi, Y., Luo, X., Xiong, N., Liu, Y., Li, S., ... others (2020). Development and clinical application of a rapid igm-igg combined antibody test for sars-cov-2 infection diagnosis. *Journal of medical virology*, 92(9), 1518–1524.
- Lim, H. X., Lim, J., Jazayeri, S. D., Poppema, S., y Poh, C. L. (2021). Development of multi-epitope peptide-based vaccines against sars-cov-2. *Biomedical Journal*, 44(1), 18–30.
- Liu, y Rusling. (2021). Covid-19 antibody tests and their limitations. *ACS nano*, 6(3), 593-612.
- Mahtarin, R., Islam, S., Islam, M. J., Ullah, M. O., Ali, M. A., y Halim, M. A. (2020). Structure and

- dynamics of membrane protein in sars-cov-2. *Journal of Biomolecular Structure and Dynamics*, 1–14.
- Mattanovich, D., Branduardi, P., Dato, L., Gasser, B., Sauer, M., y Porro, D. (2012). Recombinant protein production in yeasts. *Recombinant gene expression*, 329–358.
- McCombs, J. R., y Owen, S. C. (2015). Antibody drug conjugates: design and selection of linker, payload and conjugation chemistry. *The AAPS journal*, 17(2), 339–351.
- Moderbacher, D. G. H.-W., Ramirez, y Crotty. (2020). Antigen-specific adaptive immunity to sars-cov-2 in acute covid-19 and associations with age and disease severity. *Cell*, 183(4), 996-1012.
- Morse, N. J., Gopal, M. R., Wagner, J. M., y Alper, H. S. (2017). Yeast terminator function can be modulated and designed on the basis of predictions of nucleosome occupancy. *ACS synthetic biology*, 6(11), 2086–2095.
- Mostafa, A., Kanrai, P., Ziebuhr, J., y Pleschka, S. (2013). Improved dual promotor-driven reverse genetics system for influenza viruses. *Journal of virological methods*, 193(2), 603–610.
- Ohse, M., Takahashi, K., Kadowaki, Y., y Kusaoke, H. (1995). Effects of plasmid dna sizes and several other factors on transformation of bacillus subtilis isw1214 with plasmid dna by electroporation. *Bioscience, biotechnology, and biochemistry*, 59(8), 1433–1437.
- Overton, T. W. (2014). Recombinant protein production in bacterial hosts. *Drug discovery today*, 19(5), 590–601.
- Parapouli, A., Vasileiadis, y Hatziloukas. (2020). Saccharomyces cerevisiae and its industrial applications. *ACS nano*, 6(1), 1.
- Paul, D., Sharif, I. H., Sayem, A., Ahmed, H., Saleh, A., y Mahmud, S. (2021). In silico prediction of a highly immunogenic and conserved epitope against zika virus. *Informatics in Medicine Unlocked*, 100613.
- Peacock, T. P., Goldhill, D. H., Zhou, J., Baillon, L., Frise, R., Swann, O. C., ... others (2021). The furin cleavage site in the sars-cov-2 spike protein is required for transmission in ferrets. *Nature Microbiology*, 1–11.
- Penalver, S. S. G.-S. K., Schmid, y Bottcher. (2020). *A ligand selection strategy identifies chemical probes targeting the proteases of sars-cov-2*. Angewandte Chemie International Edition.
- Petersen, y Gokengin. (2020). Sars-cov-2 epidemiology and control, different scenarios for turkey. *Turkish Journal of Medical Sciences*, 50(1), 509-514.
- Pilarowski, S. L. C., Lebel, y Marquez. (2021). Performance characteristics of a rapid severe acute respiratory syndrome coronavirus 2 antigen detection assay at a public plaza testing site in san francisco. *The Journal of Infectious Diseases*, 223(7), 1139-1144.
- Piva, L. C., De Marco, J. L., Moraes, L. M. P. d., Reis, V. C. B., y Torres, F. A. G. (2020). Construction and characterization of centromeric plasmids for komagataella phaffii using a color-based plasmid stability assay. *PLoS one*, 15(7), e0235532.
- Poh, C. M., Carissimo, G., Wang, B., Amrun, S. N., Lee, C. Y.-P., Chee, R. S.-L., ... others (2020). Two linear epitopes on the sars-cov-2 spike protein that elicit neutralising antibodies in covid-19 patients. *Nature communications*, 11(1), 1–7.
- Pokhrel, M., Hu, y Hanbin. (2020). Detecting the coronavirus (covid-19). *ACS Sensors*, 1(1), 1.
- Ponomarenko, J., Bui, H.-H., Li, W., Füsseder, N., Bourne, P. E., Sette, A., y Peters, B. (2008). ElliPro: a new structure-based tool for the prediction of antibody epitopes. *BMC bioinformatics*, 9(1), 1–8.
- Qamar, M. T. u., Shahid, F., Ashfaq, U. A., Aslam, S., Fatima, I., Fareed, M. M., ... Chen, L.-L. (2020). Structural modeling and conserved epitopes prediction against sars-cov-2 structural proteins for vaccine development.
- Reche, P., Flower, D. R., Fridkis-Hareli, M., y Hoshino, Y. (2018). *Peptide-based immunotherapeutics and vaccines 2017*. Hindawi.
- Ruell, L. (2020). High-efficiency transformation of yeast by electroporation.
- Ruvalcaba-Hidrogo, G.-F. J.-A. E.-D., Alvarado-González, y Espino-Solís. (2019). Comparación de la eficiencia de transformación entre diferentes cepas de e. coli. *TECNOCENCIA Chihuahua*, 13(2), 112-120.
- Rytter, J. V., Helmark, S., Chen, J., Lezyk, M. J., Solem, C., y Jensen, P. R. (2014). Synthetic

- promoter libraries for corynebacterium glutamicum. *Applied microbiology and biotechnology*, 98(6), 2617–2623.
- Sajid, M., Kawde, A.-N., y Daud, M. (2015). Designs, formats and applications of lateral flow assay: A literature review. *Journal of Saudi Chemical Society*, 19(6), 689–705.
- Seo, K.-B. C.-K., Lee, y Kim. (2020). Rapid detection of covid-19 causative virus (sars-cov-2) in human nasopharyngeal swab specimens using field-effect transistor-based biosensor. *ACS nano*, 14(4), 5135-5142.
- Serrano-Rivero, H.-G., y Fando-Calzada. (2013). Comparación de la eficiencia de transformación entre diferentes cepas de e. coli. *Ciencias Biológicas*, 44(2), 1.
- Sette, y Crotty. (2021). Adaptive immunity to sars-cov-2 and covid-19. *Cell*, 1(1), 1.
- Shenoy, V., Qi, Y., Katovich, M. J., y Raizada, M. K. (2011). Ace2, a promising therapeutic target for pulmonary hypertension. *Current opinion in pharmacology*, 11(2), 150–155.
- Shepley-McTaggart, O. R.-D. L., Sagum, y Harty. (2021). Sars-cov-2 envelope (e) protein interacts with pdz-domain-2 of host tight junction protein zo1. *Plos one*, 16(6), 1-11.
- Shu, Y., y McCauley, J. (2017). Gisaid: Global initiative on sharing all influenza data—from vision to reality. *Eurosurveillance*, 22(13), 30494.
- Sievers, F., Wilm, A., Dineen, D., Gibson, T. J., Karplus, K., Li, W., ... others (2011). Fast, scalable generation of high-quality protein multiple sequence alignments using clustal omega. *Molecular systems biology*, 7(1), 539.
- Smyrlaki, L. S. P.-V., Ekman, y Reinius. (2020). Massive and rapid covid-19 testing is feasible by extraction-free sars-cov-2 rt-pcr. *ACS nano*, 11(1), 1-12.
- Song, X. H., Li, y You. (2020). Cytokine storm induced by sars-cov-2. *Clinica chimica acta*, 509(1), 280-287.
- Souza-Moreira, T. M., Navarrete, C., Chen, X., Zanelli, C. F., Valentini, S. R., Furlan, M., ... Krivoruchko, A. (2018). Screening of 2a peptides for polycistronic gene expression in yeast. *FEMS yeast research*, 18(5), foy036.
- Sreerag, C., Sreehari, C., Srinivas, M., y Variar, K. S. (2021). Impact of globalization on ergodic behaviour of pandemic diseases outbreaks. *Journal of Statistics and Management Systems*, 1–22.
- Sun, W., Jiang, B., Zhao, D., Pu, Z., y Bao, Y. (2021). Integration of metabolic pathway manipulation and promoter engineering for the fine-tuned biosynthesis of malic acid in bacillus coagulans. *Biotechnology and Bioengineering*.
- Tan, T. H., Patton, E., Munro, C. A., Corzo-Leon, D. E., Porter, A. J., y Palliyil, S. (2021). Monoclonal human antibodies that recognise the exposed n and c terminal regions of the often-overlooked sars-cov-2 orf3a transmembrane protein. *Viruses*, 13(11), 2201.
- Terry, S. M. P.-S., Anderson, y Geiss. (2021). Development of a sars-cov-2 nucleocapsid specific monoclonal antibody. *Virology*, 558(1), 28-37.
- Thomas. (2020). The structure of the membrane protein of sars-cov-2 resembles the sugar transporter semisweet. *Pathogens and Immunity*, 5(1), 342.
- Tono, M. W., García, y Mahecha. (2020). Covid-19: generalidades, comportamiento epidemiológico y medidas adoptadas en medio de la pandemia en colombia. *Acta de Otorrinolaringología y Cirugía de Cabeza y Cuello*, 1(1), 4-13.
- Torrance, V., y Lydall, D. (2018). Overlapping open reading frames strongly reduce human and yeast stn1 gene expression and affect telomere function. *PLoS genetics*, 14(8), e1007523.
- Trono. (2019). Recombinant enzymes in the food and pharmaceutical industries. *In Advances in Enzyme Technology*, 1(1), 349-387.
- Valderrama-Rincon, J. D., Fisher, A. C., Merritt, J. H., Fan, Y.-Y., Reading, C. A., Chhiba, K., ... DeLisa, M. P. (2012). An engineered eukaryotic protein glycosylation pathway in escherichia coli. *Nature chemical biology*, 8(5), 434–436.
- van der Heide. (2020). Sars-cov-2 cross-reactivity in healthy donors. *Nature Reviews Immunology*, 20(7), 408-408.
- Van Regenmortel, M. H. (2009). What is a b-cell epitope? En *Epitope mapping protocols* (pp. 3–20). Springer.
- Vieira Gomes, A. M., Souza Carmo, T., Silva Carvalho, L., Mendonça Bahia, F., y Parachin, N. S.

- (2018). Comparison of yeasts as hosts for recombinant protein production. *Microorganisms*, 6(2), 38.
- Vita, R., Mahajan, S., Overton, J. A., Dhanda, S. K., Martini, S., Cantrell, J. R., ... Peters, B. (2019). The immune epitope database (iedb): 2018 update. *Nucleic acids research*, 47(D1), D339–D343.
- Voth, W. P., Richards, J. D., Shaw, J. M., y Stillman, D. J. (2001a). Yeast vectors for integration at the ho locus. *Nucleic acids research*, 29(12), e59–e59.
- Voth, W. P., Richards, J. D., Shaw, J. M., y Stillman, D. J. (2001b). Yeast vectors for integration at the ho locus. *Nucleic acids research*, 29(12), e59–e59.
- Wang, J. J., Zhang, N., Richardson, S. A., y Wu, J. V. (2021). Rapid lateral flow tests for the detection of sars-cov-2 neutralizing antibodies. *Expert Review of Molecular Diagnostics*, 21(4), 363–370.
- Yang, Q., Lyu, X., Zhao, F., y Liu, Y. (2021). Effects of codon usage on gene expression are promoter context dependent. *Nucleic acids research*, 49(2), 818–831.
- Zang, J., Zhu, Y., Zhou, Y., Gu, C., Yi, Y., Wang, S., ... others (2021). Yeast-produced rbd-based recombinant protein vaccines elicit broadly neutralizing antibodies and durable protective immunity against sars-cov-2 infection. *Cell Discovery*, 7(1), 1–16.
- Zeng, M. Z. Y.-L., Liu, y Jin. (2020). Biochemical characterization of sars-cov-2 nucleocapsid protein. *Biochemical and biophysical research communications*, 527(3), 618–623.
- Zhang, J. J., Yamanaka, K., Tang, X., y Moore, B. S. (2019). Direct cloning and heterologous expression of natural product biosynthetic gene clusters by transformation-associated recombination. *Methods in enzymology*, 621, 87–110.
- Zhou, C. B. L.-W., Ji, y Shi. (2021). Identification of novel bat coronaviruses sheds light on the evolutionary origins of sars-cov-2 and related viruses. *ACS nano*, 11(1), 1–12.
- Zhu, Z., Zhu, y Sun. (2020). Minireview of progress in the structural study of sars-cov-2 proteins. *Current research in microbial sciences*, 14(4), 5135–5142.

Figura 21: Modelo 3D de regiones con alta sensibilidad al solvente en proteínas de SARS-CoV-2



Descripción: Regiones con alta accesibilidad para (A)Glucoproteína S; (B)nucleoproteína; (C)proteína de membrana.

Cuadro 5: Secuencias de epítomos de SARS-CoV-2 recuperadas de IEDB y empleadas para la generación de agrupamientos por medio del *Cluster analysis*

No. Cluster	No. Péptido	Alineamiento	Descripción
1.1	Consenso	NYNLYRLFRKSNLKPFFERDISTEIQQA	-
1.1	1	NYNLYRLFRK—————	glucoproteína S
1.1	2	NYNLYRLF—————	glucoproteína S
1.1	3	-YNLYRLFRKSNLKP————	glucoproteína S
1.1	4	-YNLYRLFR—————	glucoproteína S
1.1	5	—YLYRLFRKSNLKPFE————-	glucoproteína S
1.1	6	—YRLFRKSNLKPFFERD————	glucoproteína S
1.1	7	—RLFRKSNLKPFE————-	glucoproteína S
1.1	8	—RLFRKSNLKP—————	glucoproteína S
1.1	9	—RLFRKSNLK—————	glucoproteína S
1.1	10	—-LFRKSNLKPFFERDIS————	glucoproteína S
1.1	11	—FRKSNLKPFFERDIST———	glucoproteína S
1.1	12	—FRKSNLKPFFERD—————	glucoproteína S
1.1	13	—RKNLKPFFERDISTE——-	glucoproteína S
1.1	14	—-KSNLKPFFER—————	glucoproteína S
1.1	15	—NLKPFFERDISTE——-	glucoproteína S
1.1	16	—NLKPFFERD—————	glucoproteína S
1.1	17	—-LKPFFERDISTEIQQA	glucoproteína S
1.1	18	—-LKPFFERDISTEI———	glucoproteína S
1.1	19	—KPPFERDISTEI———	glucoproteína S
1.1	20	—-FERDISTEI———	glucoproteína S

Cuadro 6: Continuación 1 de secuencias de epítomos de SARS-CoV-2 recuperadas de IEDB y empleadas para la generación de agrupamientos por medio del *Cluster analysis*

No. Cluster	No. Péptido	Alineamiento	Descripción
1.2	Singleton	VGGNYNYLYRLFRKSNLKP	glucoproteína S
2.1	Consenso	WTAGAAAYYVGYLQPRTFLLKY	-
2.1	1	WTAGAAAYYVGY-----	glucoproteína S
2.1	2	WTAGAAAYY-----	glucoproteína S
2.1	3	--GAAAYYVGYLQPRTF--	glucoproteína S
2.1	4	--GAAAYYVGY-----	glucoproteína S
2.1	5	--AAYYVGYLQPRTF--	glucoproteína S
2.1	6	---YYVGYLQPRTFLLKY	glucoproteína S
2.1	7	---YYVGYLQPRTFLL-	glucoproteína S
2.1	8	---YYVGYLQPRTF--	glucoproteína S
2.1	9	-----GYLQPRTFLL-	glucoproteína S
2.1	10	-----GYLQPRTFL-	glucoproteína S
2.1	11	-----YLQPRTFLL-	glucoproteína S
3.1	Consenso	AALALLLLDRLNQLESKMSG	-
3.1	1	AALALLLLDRLNQLE---	Nucleoproteína
3.1	2	-LALLLLDRL-----	Nucleoproteína
3.1	3	---LLLLDRLNQLESKMS-	Nucleoproteína
3.1	4	---LLLLDRLNQL---	Nucleoproteína
3.1	5	---LLLDRLNQLESKMSG	Nucleoproteína
3.1	6	---LLLDRLNQL---	Nucleoproteína
4.1	Consenso	NFGDQELIRQGTDYKHWPQIAQFAPS-	-
4.1	1	NFGDQELIRQGTDYKHWPQ---	Nucleoproteína
4.1	2	---LIRQGTDYKHWPQIA---	Nucleoproteína
4.1	3	-----TDYKHWPQIAQFAPS	Nucleoproteína
4.1	4	-----DYKHWPQIAQF--	Nucleoproteína
5.1	Consenso	GAVILRGHLRIAGHHLGRCDIKDLP	-
5.1	1	GAVILRGHLRIAGHH-----	Proteína M
5.1	2	---RGHLRIAGHHLGRCD---	Proteína M
5.1	3	---HLRIAGHHLGR---	Proteína M
5.1	4	-----IAGHHLGRCDIKDLP	Proteína M
6.1	Consenso	HIDAYKTFPPTEPKKDKKKK	-

Cuadro 7: Continuación 2 de secuencias de epítomos de SARS-CoV-2 recuperadas de IEDB y empleadas para la generación de agrupamientos por medio del *Cluster analysis*

No. Cluster	No. Péptido	Alineamiento	Descripción
6.1	1	HIDAYKTFPPTPEPKK—	Nucleoproteína
6.1	2	—KTFPPTPEPKKDKKKK	Nucleoproteína
6.1	3	—KTFPPTPEPKK—	Nucleoproteína
6.1	4	—KTFPPTPEPK—	Nucleoproteína
7.1	Consenso	LSRLDKVEAEVQIDRLITGR	-
7.1	1	LSRLDKVEAEVQIDRLITGR	glucoproteína S
7.1	2	LSRLDKVEAEVQIDR—	glucoproteína S
7.1	3	—AEVQIDRLI—	glucoproteína S
8.1	Consenso	GMEVTPSGTWLTYTG	-
8.1	1	GMEVTPSGTWLTYTG	Nucleoproteína
8.1	2	-MEVTPSGTWL--	Nucleoproteína
8.1	3	-MEVTPSGTW—	Nucleoproteína
9.1	Consenso	ASAFFGMSRIGMEVT	-
9.1	1	ASAFFGMSRIGMEVT	Nucleoproteína
9.1	2	ASAFFGMSR—	Nucleoproteína
9.1	3	—GMSRIGMEV-	Nucleoproteína
10.1	Consenso	NCTFEYVSQPFLMDLE	-
10.1	1	NCTFEYVSQPFLMDL-	glucoproteína S
10.1	2	-CTFEYVSQPFLMDLE	glucoproteína S
11.1	Consenso	DISTEIQAGSTPCNGVEGFNCY FPLQSYGFQPTNGVGYQPYRVVVL	-
11.1	1	DISTEIQAGSTPCNGVEGFNCYFP LQSYGFQPTNGVGYQPYRVVVL	glucoproteína S
11.1	2	—IYQAGSTPCNGVEGFNCYFPLQSY —	glucoproteína S
12.1	Consenso	NLVRDLPQGFSALEP	-
12.1	1	NLVRDLPQGFSALEP	glucoproteína S
12.1	2	—LPQGFSAL-	glucoproteína S
13.1	Consenso	LSPRWYFYLLGTGPEAGL	-
13.1	1	LSPRWYFYLLGTGPEAGL	Nucleoproteína
13.1	2	-SPRWYFYLL—	Nucleoproteína
14.1	Consenso	LYLYALVYFLQSINF	-
14.1	1	LYLYALVYFLQSINF	ORF3a
14.1	2	—VYFLQSINF	ORF3a
15.1	Consenso	PWYIWLGFIAGLIAIV	-
15.1	1	PWYIWLGFIAGLIAI-	glucoproteína S
15.1	2	—-FIAGLIAIV	glucoproteína S

Cuadro 8: Continuación 3 de secuencias de epítomos de SARS-CoV-2 recuperadas de IEDB y empleadas para la generación de agrupamientos por medio del *Cluster analysis*

No. Cluster	No. Péptido	Alineamiento	Descripción
16.1	Consenso	NKDGIIWVATEGALNTPK	-
16.1	1	NKDGIIWVATEGALN—	Nucleoproteína
16.1	2	——ATEGALNTPK	Nucleoproteína
17.1	Consenso	RSVASQSIIAYTMSL	-
17.1	1	RSVASQSIIAYTMSL	glucoproteína S
17.1	2	——SIIAYTMSL	glucoproteína S
18.1	Consenso	PSKPSKRSFIEDLLFNKV	-
18.1	1	PSKPSKRSFIEDLLFNKV	glucoproteína S
18.1	2	-KPSKRSFIEDLLFNK-	glucoproteína S
19.1	Consenso	SELVIGAVILRGHLR	-
19.1	1	SELVIGAVILRGHLR	Proteína M
19.1	2	SELVIGAVIL——	Proteína M
20.1	Consenso	LSYYKLGASQRVAGDS	-
20.1	1	LSYYKLGASQRVAGD-	Proteína M
20.1	2	-SYYKLGASQRVAGDS	Proteína M
21.1	Consenso	ATSRTLSTYYKLGASQ	-
21.1	1	ATSRTLSTYYKLGASQ	Proteína M
21.1	2	ATSRTLSTYYK——	Proteína M
22.1	Consenso	NRDVDTDFVNEFYAY	-
22.1	1	NRDVDTDFVNEFYAY	replicasa 1ab
22.1	2	——DTDFVNEFY-	replicasa 1ab
23.1	Consenso	DDSEPVKLGVKLHYT	-
23.1	1	DDSEPVKLGVKLHYT	glucoproteína S
23.1	2	-SEPVKLGVKL——	glucoproteína S
24.1	Consenso	EFVFKNIDGYFKIYS	-
24.1	1	EFVFKNIDGYFKIYS	glucoproteína S
24.1	2	-VFKNIDGYF——	glucoproteína S
25.1	Consenso	YLDAYNMMISAGFSLW	-
25.1	1	YLDAYNMMI——-	replicasa 1ab
25.1	2	-LDAYNMMISAGFSLW	replicasa 1ab
26.1	Consenso	ILLNKHIDAYKTFPP	-
26.1	1	ILLNKHIDAYKTFPP	Nucleoproteína
26.1	2	ILLNKHIDA——	Nucleoproteína
27.1	Consenso	CVADYSVLYNSASFS	-
27.1	1	CVADYSVLYNSASFS	glucoproteína S
27.1	2	CVADYSVLY——	glucoproteína S
28.1	Consenso	ISGINASVVNIQKEIDR	-
28.1	1	ISGINASVVNIQKEI-	glucoproteína S

Cuadro 9: Continuación 4 de secuencias de epítomos de SARS-CoV-2 recuperadas de IEDB y empleadas para la generación de agrupamientos por medio del *Cluster analysis*

No. Cluster	No. Péptido	Alineamiento	Descripción
28.1	Consenso	GINASVVNIQKEIDR	glucoproteína S
29.1	Consenso	WPQIAQFAPSASAFF	-
29.1	1	WPQIAQFAPSASAFF	Nucleoproteína
29.1	2	—FAPSASAFF	Nucleoproteína
30.1	Consenso	VYMPASWVMRI	-
30.1	1	VYMPASWVM-	replicasa 1ab
30.1	2	-MPASWVMRI	replicasa 1ab
31.1	Singleton	NLLQYGSFCTQLNR	seq21
32.1	Singleton	ITRFQTLALHRSYL	seq26
33.1	Singleton	QPYRVVVLSEFELLHAPATVC	seq39
34.1	Singleton	CAQKFNGTLVLPPLL	seq51
35.1	Singleton	QGTTLPKGFYAEGRGGS	seq130
36.1	Singleton	LGVDYLVSTQEFYR	seq225, seq235
37.1	Singleton	QYIKWPWYI	seq2, seq8
38.1	Singleton	NRNRFLLYIKLIFLW	seq168
39.1	Singleton	WICLLQFAYANRNRF	seq169
40.1	Singleton	VLAAYRINWITGGI	seq180
41.1	Singleton	LMIERFVSLAIDAYP	seq194
42.1	Singleton	LTDEMIAQY	seq5
43.1	Singleton	TQLNRALTGIAVEQD	seq32
44.1	Singleton	VTQNVLYENQKLIAN	seq33
45.1	Singleton	TESNKKFLPFQGFGRDIA	seq44
46.1	Singleton	DTTDAVRDPQTLEIL	seq92, seq101
47.1	Singleton	NFSQILPDPSKPSKR	seq6, seq9
48.1	Singleton	VLNDILSRL	seq7, seq13
49.1	Singleton	FKIYSKHTPINLVRD	seq24
50.1	Singleton	QSINFVRIIMRLWLC	seq110
51.1	Singleton	SFNPETNILLNVPLH	seq184, seq189
52.1	Singleton	YLTFFYLTNDVSFLAH	seq210
53.1	Singleton	AVKLQNNELSPVALR	seq218, seq233
54.1	Singleton	RLQSLQTYV	seq12
55.1	Singleton	YYQLYSTQL	seq105
56.1	Singleton	FTSDYYQLY	seq106
57.1	Singleton	TTDPSFLGRY	seq195, seq197
58.1	Singleton	FLLPSLATV	seq219
59.1	Singleton	QEYADVFLHY	seq228
60.1	Singleton	RPDTRYVLM	seq229
61.1	Singleton	TMADLVYAL	seq230
62.1	Singleton	RLNEVAKNL	seq98
63.1	Singleton	ALWEIQQVV	seq203, seq208
64.1	Singleton	LLDDFVEI	seq214, seq226
65.1	Singleton	KCYGVSPTK	seq17
66.1	Singleton	LLHAPATVCGPKKST	seq94
67.1	Singleton	ALNTLVKQL	seq18
68.1	Singleton	EDLLFNKVTADAGF	seq41, seq87

Cuadro 10: Continuación 5 de secuencias de epítomos de SARS-CoV-2 recuperadas de IEDB y empleadas para la generación de agrupamientos por medio del *Cluster analysis*

No. Cluster	No. Péptido	Alineamiento	Descripción
69.1	Singleton	TVTLLPAADLDDFSK	seq119
70.1	Singleton	AIVLQLPQGTTLPKG	seq125, seq134
71.1	Singleton	QKKQQTVTLLPAADL	seq143
72.1	Singleton	EELKKLLEQWNLVIG	seq174
73.1	Singleton	HWFVTQRNFYEPQII	seq19
74.1	Singleton	EHVTFFIYNKIVDEP	seq114
75.1	Singleton	LPPAYTNSF	seq46
76.1	Singleton	IKDLPKEITVATSRT	seq185
77.1	Singleton	VYIGDPAQL	seq196, seq200
78.1	Singleton	KLSYGIATV	seq213
79.1	Singleton	KQFDTYNLW	seq223
80.1	Singleton	KLNVGDYFV	seq221
81.1	Singleton	YIFFASFYY	seq232, seq239
82.1	Singleton	IAIVMVTIMLCCMTS	seq93, seq103
83.1	Singleton	MWLSYFIASFRLFAR	seq167
84.1	Singleton	CLVGLMWLSYFIASF	seq173
85.1	Singleton	FIASFRLFARTRSMW	seq177
86.1	Singleton	IPRRNVATL	seq198, seq205
87.1	Singleton	FVSLAIDAYPLTKHP	seq212
88.1	Singleton	IPFAMQMAYRFNGIG	seq25
89.1	Singleton	FPRGQGVPI	seq128, seq135
90.1	Singleton	TSNQVAVLY	seq50
91.1	Singleton	FLLNKEMYL	seq204
92.1	Singleton	TNFTISVTTEILPVS	seq30
93.1	Singleton	VYYPDKVFRSSVLHS	seq34
94.1	Singleton	NFRVQPTESIVRFPN	seq35
95.1	Singleton	VVFLHVITYV	seq42
96.1	Singleton	TLDSKTQSL	seq49
97.1	Singleton	CPFGEVFNATRFASV	seq52
98.1	Singleton	AENSVAYSNNIAIP	seq53
99.1	Singleton	SCGSCCKFDEDDSEP	seq91, seq99
100.1	Singleton	ALSKGVHFV	seq108, seq109
101.1	Singleton	LLYDANYFL	seq112
102.1	Singleton	EPIYDEPTTTTSVPL	seq115
103.1	Singleton	KAYNVTQAF	seq120
104.1	Singleton	NTNSSPDDQIGYY	seq121
105.1	Singleton	PSGTWLTYYTGAIKLD	seq136

Cuadro 11: Continuación 6 de secuencias de epítomos de SARS-CoV-2 recuperadas de IEDB y empleadas para la generación de agrupamientos por medio del *Cluster analysis*

No. Cluster	No. Péptido	Alineamiento	Descripción
106.1	Singleton	GKGQQQQGQTVTKKS	seq146
107.1	Singleton	RRGPEQTQGNFGDQE	seq150
108.1	Singleton	TRSMWSFNPETNILL	seq172
109.1	Singleton	RVAGDSGFAAY	seq175, seq188
110.1	Singleton	TNILLNVPLHGTILT	seq176
111.1	Singleton	RLFARTRSMWSFNPE	seq178
112.1	Singleton	RYRIGNYKL	seq193
113.1	Singleton	CTDDNALAYY	seq199
114.1	Singleton	PTDNYITTY	seq202, seq207
115.1	Singleton	IQWMVMFTPLVPFWI	seq209
116.1	Singleton	YYTSNPTTF	seq211
117.1	Singleton	ATVVIGTSK	seq217
118.1	Singleton	FLPRVFSAV	seq220, seq234
119.1	Singleton	KLWAQCVQL	seq222
120.1	Singleton	VVYRGTTTTYK	seq231, seq237
121.1	Singleton	SEFSSLPSY	seq236
122.1	Singleton	LLWPVTLACFVLAAY	seq183
123.1	Singleton	KPRQKRTAT	seq137
124.1	Singleton	YRINWITGGIAIAMA	seq190
125.1	Singleton	ASMPPTIAK	seq216
126.1	Singleton	VTKKSAEASKKPRQ	seq151
127.1	Singleton	LIFLWLLWPVTLACF	seq182

Cuadro 12: Índice de Kolanskar y Tangaokar para regiones de aminoácidos 447-472 de la proteína S del SARS-CoV-2

Residuo central	Inicio	Fin	Péptido	Puntaje K y T
N	447	453	GNYNYLY	1.023
Y	448	454	NYNYLYR	1.023
L	449	455	YNYLYRL	1.09
Y	450	456	NYLYRLF	1.08
R	451	457	YLYRLFR	1.094
L	452	458	LYRLFRK	1.061
F	453	459	YRLFRKS	1.027
K	455	461	LFRKSNL	1.026
T	467	473	DISTEY	1.015
I	469	475	STEYQA	1.023
Y	470	476	TEYQAG	1.004
Q	471	477	EYQAGS	1.018
A	472	478	EYQAGST	1.027

Cuadro 13: Índice de Kolanskar y Tangaokar para regiones de aminoácidos 103-139 de la proteína N del SARS-CoV-2

Residuo central	Inicio	Final	Péptido	Puntaje K y T
P	103	109	DLSPRWY	1.017
R	104	110	LSPRWYF	1.049
W	105	111	SPRWYFY	1.036
Y	106	112	PRWYFYF	1.058
F	107	113	RWYFYFL	1.084
Y	108	114	WYFYFLG	1.084
Y	109	115	YFYFLGT	1.087
L	110	116	FYFLGTG	1.046
G	111	117	YFLGTGP	1.042
A	116	122	GPEAGLP	1.006
G	117	123	PEAGLPY	1.047
L	118	124	EAGLPYG	1.02
I	127	133	KDGIIWV	1.036
I	128	134	DGIIWVA	1.055
W	129	135	GIIWVAT	1.061
V	130	136	IIWVATE	1.058
T	132	138	WVATEGA	1.005
E	133	139	VATEGAL	1.056

Cuadro 14: Índice de Kolanskar y Tangaokar para regiones de aminoácidos 138-181 de la proteína M del SARS-CoV-2

Residuo central	Inicio	Final	Péptido	Puntaje K y T
G	138	144	LVIGAVI	1.18
A	139	145	VIGAVIL	1.18
V	140	146	IGAVILR	1.107
I	141	147	GAVILRG	1.067
L	142	148	AVILRGH	1.1
R	143	149	VILRGHL	1.127
G	144	150	ILRGHLR	1.054
H	145	151	LRGHLRI	1.054
L	146	152	RGHLRIA	1.027
R	147	153	GHLRIAG	1.027
I	148	154	HLRIAGH	1.06
A	149	155	LRIAGHH	1.06
G	150	156	RIAGHHL	1.06
H	151	157	IAGHHLG	1.061
H	152	158	AGHHLGR	1.021
L	153	159	GHHLGRC	1.07
G	154	160	HHLGRCD	1.069
R	155	161	HLGRCDI	1.076
C	156	162	LGRCDIK	1.051
I	158	164	RCDIKDL	1.05
A	168	174	ITVATSR	1.043
T	169	175	TVATSRT	1.008
S	170	176	VATSRTL	1.057
R	171	177	ATSRTLS	1.004
T	172	178	TSRTLSY	1.018
L	173	179	SRTLSYY	1.054
S	174	180	RTLSYYK	1.042
Y	175	181	TLSYYKL	1.096
K	177	183	SYYKLG A	1.065
L	178	184	YYKLGAS	1.065
G	179	185	YKLGASQ	1.044
A	180	186	KLGASQR	1.003
S	181	187	LGASQRV	1.067
S	181	187	LGASQRV	1.067

Cuadro 15: Secuencias consenso generadas a través del *epitope Cluster Analysis* para las proteínas S, N y M del SARS-CoV-2

Proteína S
NYNYLYRLFRKSNLKPFERDISTEIQ
VGGNYNYLYRLFRKSNLKP
WTAGAAAYYVGYLQPRTFLLKY
LSRLDKVEAEVQIDRLITGR
NCTFEYVSQPFLMDLE
NLVRDLPQGFSALEP
PWYIWLGFIAGLIAIV
PSKPSKRSFIEDLLFNKV
DDSEPVKGVKLHYT
EFVFKNIDGYFKIYS
DISTEIQAGSTPCNGVEGFNCYF
PLQSYGFQPTNGVGYQPYRVVVL
RSVASQSIAYTMSL
Proteína N
AALALLLDRLNQLESKMSG
NFGDQELIRQGTDYKHWPQIAQFAPS
HIDAYKTFPPTEPKKDKKKK
GMEVTPSGTWLTYTG
ASAFFGMSRIGMEVT
LSPRWYFYLLGTGPEAGL
NKDGIIWVATEGALNTPK
Proteína M
GAVILRGHLRIAGHHLGRCDIKDLP
SELVIGAVILRGHLR
LSYYKLGASQRVAGDS
ATSRTLSTYYKLGASQ

Cuadro 16: Frecuencias de alelos MHC utilizados para el análisis de cobertura poblacional

MHC	MHC Locus	Alelo MHC	P. asendencia maya	P. asendencia mixta
I	HLA-A	HLA-A*02	0.363	0.2677
I	HLA-A	HLA-A*24	0.231	0.1614
I	HLA-A	HLA-A*68	0.158	0.1654
I	HLA-B	HLA-B*35	0.459	0.3386
I	HLA-DQB1	HLA-DQB1*03	0.481	0.3937

Figura 22: Secuencia del sistema de proteína de fusión epítipo-GFP



Figura 23: Continuación de secuencia del sistema de proteína de fusión epítipo-GFP

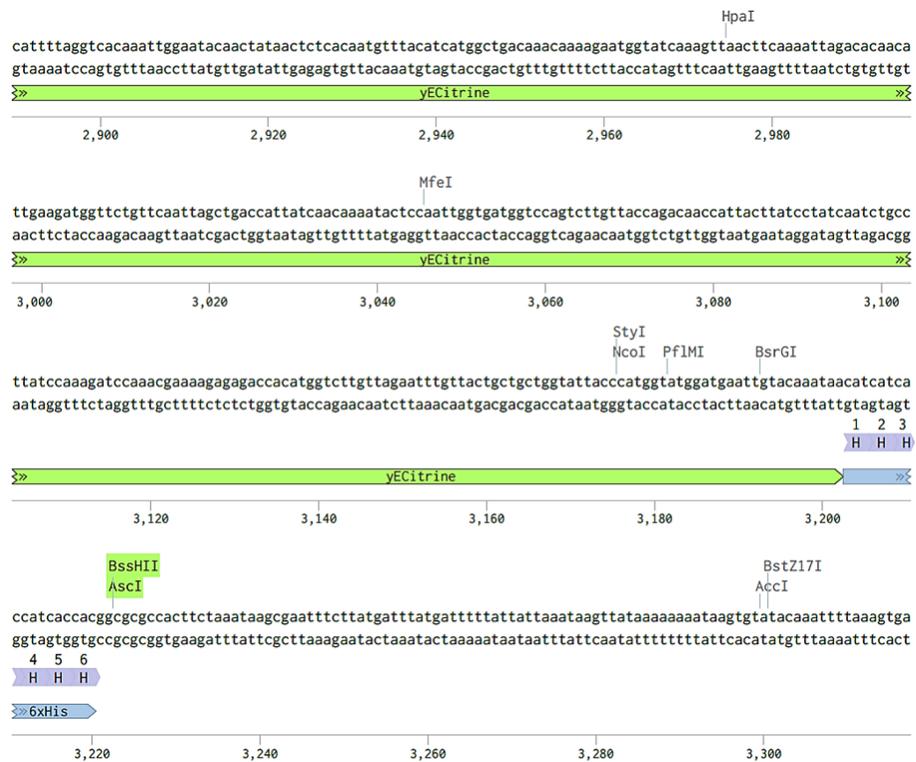


Figura 24: Secuencia del sistema policistrónico expresando dos epítomos con espaciador P2A del ERVB-1

