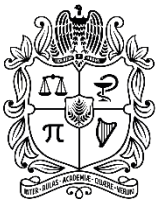


UNIVERSIDAD
NACIONAL
DE COLOMBIA

Modelo *in-silico* para la predicción de péptidos con restricción HLA-I candidatos a vacuna en SARS-CoV-2

Holman Alexander Hernandez Nieto

Universidad Nacional de Colombia

Facultad de ingeniería, Departamento de ingeniería de sistemas e industrial

Bogotá, Colombia

2023

Modelo *in-silico* para la predicción de péptidos con restricción HLA-I candidatos a vacuna en SARS-CoV-2

Holman Alexander Hernandez Nieto

Tesis o trabajo de investigación presentada(o) como requisito parcial para optar al título de:

Maestría en Bioinformática

Director:

Luis Fernando Niño Vásquez Ph.D.

Codirector:

Carlos Alberto Parra Lopez MD., Ph.D.

Línea de Investigación:

Tecnologías computacionales en Bioinformática

Grupo de Investigación:

Laboratorio de investigación en sistemas inteligentes (LISI);

Grupo de Inmunología y Medicina Traslacional (I&MT)

Universidad Nacional de Colombia

Facultad de ingeniería, Departamento de ingeniería de sistemas e industrial

Bogotá, Colombia

2023

Dedicatoria A:

A mis padres, que siempre han estado ahí para mí. Gracias por su amor, apoyo y aliento. No sería la persona que soy hoy sin ustedes.

Y mi hermana y novia que durante todo estos años me han dado ánimos y aliento para continuar.

“La disciplina tarde o temprano vencerá la inteligencia”. Anónimo

Declaración de obra original

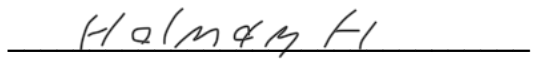
Yo declaro lo siguiente:

He leído el Acuerdo 035 de 2003 del Consejo Académico de la Universidad Nacional. «Reglamento sobre propiedad intelectual» y la Normatividad Nacional relacionada al respeto de los derechos de autor. Esta disertación representa mi trabajo original, excepto donde he reconocido las ideas, las palabras, o materiales de otros autores.

Cuando se han presentado ideas o palabras de otros autores en esta disertación, he realizado su respectivo reconocimiento aplicando correctamente los esquemas de citas y referencias bibliográficas en el estilo requerido.

He obtenido el permiso del autor o editor para incluir cualquier material con derechos de autor (por ejemplo, tablas, figuras, instrumentos de encuesta o grandes porciones de texto).

Por último, he sometido esta disertación a la herramienta de integridad académica, definida por la universidad.



Holman Alexander Hernandez Nieto

Fecha 01/08/2023

Agradecimientos

Quiero agradecer a mis tutores Luis Fernando Niño Vásquez Ph.D. y Carlos Alberto Parra Lopez MD., Ph.D. por su guía y apoyo durante este proceso. Su orientación ha sido invaluable, y sus comentarios me han ayudado a mejorar mi investigación y escritura. Estoy muy agradecido por su tiempo y dedicación.

Quiero agradecer a mis colegas, Daniel Alzate Gutierrez, Laura Camila Martinez Enriquez, Manuela Llano Leon, David Santiago Padilla Fino, Diego Amaya y demás colegas de los grupos de investigación LISI (Laboratorio de investigación en sistemas inteligentes) y I&MT (Grupo de Inmunología y Medicina Traslacional) por todo su apoyo. Sus comentarios y consejos me han ayudado a crecer como investigador.

Quiero agradecer a mi familia y amigos por su apoyo incondicional. Su aliento me ha dado la fuerza para seguir adelante, incluso cuando las cosas se han vuelto difíciles. Estoy muy agradecido por su presencia en mi vida

Finalmente, quiero agradecer a todos los que han contribuido de alguna manera a este proyecto. Sin su apoyo, no habría sido posible completarlo. Estoy muy agradecido por su ayuda.

Resumen

Modelo in-silico para la predicción de péptidos con restricción HLA-I candidatos a vacuna en SARS-CoV-2

En este trabajo se resalta la importancia de las herramientas computacionales en el diseño de vacunas contra el SARS-CoV-2, virus que desde su descubrimiento en Wuhan China en diciembre de 2019, entre los años 2020 y 2022, generó una pandemia a nivel mundial que causó cerca de siete millones de muertes y cerca de ochocientos millones de casos. El SARS-CoV-2 pertenece a la familia Coronaviridae y consta de múltiples proteínas, siendo la proteína de espícula (S) importante motivo de estudio para el desarrollo de vacunas. Las células presentadoras desempeñan un papel vital en este proceso, ya que son responsables de procesar y presentar los antígenos a las células T, lo cual desencadena la activación y regulación de la respuesta inmunitaria adaptativa mediada por las células T. Este mecanismo de presentación de antígenos es esencial para el funcionamiento adecuado del sistema inmunológico contra patógenos y el cáncer. En este trabajo, tiene que ver con los procesos de procesamiento y presentación de antígenos en la superficie de células presentadoras de antígeno en el contexto de moléculas del MHC clase I o II necesario para su reconocimiento por parte de los Receptores para el Antígenos de las Células T (TCR) requisito fundamental para la generación de una respuesta inmune eficaz de las células T contra el antígeno. Se destaca la imperiosa necesidad de impulsar constantemente el desarrollo y mejora de herramientas bioinformáticas con el objetivo de identificar el universo de péptidos que se unen de manera altamente estable a las moléculas MHC I y II, como herramienta útil para para el rápido diseño de nuevas vacunas contra patógenos emergentes como el SARS-CoV-2. En este contexto es necesario avanzar en el refinamiento de herramientas bioinformáticas en la identificación de fragmentos de proteínas de los patógenos que presentados en moléculas MHC, estimulan eficientemente a los linfocitos T, resulta de vital importancia para el ámbito clínico, ya que estas tienen un impacto significativo en la

rápida identificación de fragmentos procesados de los patógenos o de los tumores importante para el diseño de nuevas vacunas.

El avance de las ciencias ómicas y métodos de secuenciación de última generación, han permitido no solo un análisis más detallado y completo de la información genética y proteómica relacionada con los péptidos y el MHC sino mejorar el desempeño de herramientas bioinformáticas para la predicción de epítopos inmunogénicos (fragmentos de los patógenos o de los tumores presentados en moléculas MHC eficientemente reconocidos por los linfocitos T). Esto, a su vez, facilita la identificación de antígenos específicos presentados por el MHC, lo que es fundamental para comprender cómo el sistema inmunológico detecta y responde a distintos tipos de amenaza, como lo son las infecciones el cáncer y las enfermedades autoinmunes.

El perfeccionamiento continuo de las herramientas bioinformáticas para seleccionar de forma más precisas posibles antígenos útiles como vacuna, fortalece la posibilidad de diseñar vacunas sintéticas basadas en péptidos que por su inmunogenicidad y simplicidad de producción son una importante alternativa para el diseño racional de vacunas contra patógenos emergentes. La identificación de péptidos presentados eficientemente por moléculas MHC va a contribuir al desarrollo de nuevas vacunas más efectivas y a refinar estrategias de inmunoterapia dirigidas contra el cáncer, agentes infecciosos, y enfermedades autoinmunes, estrategias en las que los linfocitos T juegan un papel fundamental. En este trabajo, haciendo uso de herramientas de predicción se desarrolló un pipeline bioinformático para la predicción de epítopos candidatos a vacuna contra el SARS-CoV-2 teniendo en cuenta las moléculas MHC-I expresadas por la población colombiana. Cuando se comparó la inmunogenicidad para el sistema inmune de pacientes con SARS-CoV-2 del universo de péptidos identificados en el proteoma del virus utilizando la herramienta diseñada, con la inmunogenicidad de estos péptidos reportados en la literatura científica por otros autores revela que los péptidos predichos por nosotros merecen ser considerados como nuevos candidatos a vacuna contra el SARS-CoV-2 para ser utilizada en la población colombiana.

Palabras clave: (Pipeline Bioinformático, Péptido, Epítopo, SARS-CoV-2, Espícula, Vacuna, HLA).

Abstract

In-silico model for the prediction of HLA-I restricted peptides vaccine candidates in SARS-CoV-2

In this work, the importance of computational tools in the design of vaccines against SARS-CoV-2 is emphasized. Since its discovery in Wuhan, China, in December 2019, the virus caused a global pandemic between 2020 and 2022, resulting in nearly seven million deaths and around eight hundred million cases. SARS-CoV-2 belongs to the Coronaviridae family and consists of multiple proteins, with the spike protein (S) being a significant focus of study for vaccine development. Crucial to this process are antigen-presenting cells, responsible for processing and presenting antigens to T cells, triggering the activation and regulation of adaptive immune responses mediated by T cells. This mechanism of antigen presentation is essential for the proper functioning of the immune system against pathogens and cancer.

This work primarily deals with the antigen processing and presentation on the surface of antigen-presenting cells in the context of MHC class I or II molecules, which is necessary for recognition by T Cell Receptor (TCR) and is a fundamental requirement for generating an effective T cell immune response against the antigen. It highlights the urgent need to continually advance and improve bioinformatic tools to identify the universe of peptides that bind highly stably to MHC I and II molecules as a useful resource for the rapid design of new vaccines against emerging pathogens such as SARS-CoV-2. In this context, the refinement of bioinformatic tools in the identification of protein fragments from pathogens presented on MHC molecules, effectively stimulating T lymphocytes, becomes of vital importance in the clinical setting, as they significantly impact the rapid identification of processed pathogen or tumor fragments, crucial for vaccine design.

The progress in omics sciences and next-generation sequencing methods has not only allowed for a more detailed and comprehensive analysis of genetic and proteomic information related to peptides and MHC but also improved the performance of bioinformatic tools for predicting immunogenic epitopes (fragments of pathogens or tumors presented on MHC molecules and efficiently recognized by T cells). This, in turn,

facilitates the identification of specific antigens presented by MHC, which is essential for understanding how the immune system detects and responds to various threats such as infections, cancer, and autoimmune diseases.

The continuous improvement of bioinformatic tools to more accurately select potential vaccine antigens strengthens the possibility of designing synthetic peptide-based vaccines, which are immunogenic and easy to produce, making them a crucial alternative for the rational design of vaccines against emerging pathogens. The identification of efficiently presented peptides by MHC molecules will contribute to the development of more effective vaccines and refined immunotherapy strategies targeting cancer, infectious agents, and autoimmune diseases, where T cells play a fundamental role. In this work, a bioinformatic pipeline for predicting vaccine candidate epitopes against SARS-CoV-2 was developed using prediction tools, considering MHCI molecules expressed in the Colombian population. When comparing the immunogenicity for the immune system of SARS-CoV-2 patients from the universe of peptides identified in the virus's proteome using the designed tool with the immunogenicity of these peptides reported in the scientific literature by other authors, it reveals that the peptides predicted by us deserve consideration as new vaccine candidates against SARS-CoV-2 for use in the Colombian population.

Keywords: (Pipeline, Peptide, Epitope, SARS-CoV-2, Spike, Vaccine, HLA).

Contenido

Lista de figuras

Pág.

Lista de tablas

Pág.

Lista de Símbolos y abreviaturas

Abreviaturas

Abreviatura	Término
ADN	Ácido desoxirribonucleico
ARN	Ácido ribonucleico
OMS	Organización Mundial de la Salud
MHC	Complejo Mayor de Histocompatibilidad
MHC-II	Complejo Mayor de Histocompatibilidad de Clase II
hACE2	Enzima Transformadora de Angiotensina Humana
M.tb	Mycobacterium tuberculosis
ANN	redes neuronales artificiales
HMM	modelos ocultos de Markov
nM	nanomolar

Introducción

Desde su descubrimiento en diciembre de 2019 en Wuhan, China, el SARS-CoV-2 fue catalogado como pandemia por la Organización Mundial de la Salud (OMS) que generó un impacto en la salud mundial sin precedentes [1-3]. La generación rápida y efectiva de vacunas ha demostrado ser crucial en la lucha contra la enfermedad y las herramientas computacionales han desempeñado un papel importante en el tiempo de diseño y la agilización de procesos [12-16]. Aunque existen herramientas bioinformáticas de acceso libre con excelente rendimiento, el uso combinado de varias de estas es importante para obtener universos de péptidos más precisos y omitir su evaluación experimental en el laboratorio. La bioinformática permite cotejar y analizar datos biológicos a gran escala, lo cual ofrece una valiosa perspectiva para predecir y seleccionar con mayor eficiencia péptidos candidatos a vacuna en cualquier patógeno que pueden ser relevantes para simplificar fases de investigación experimental tanto *in-vivo* o *in-vitro* empleados habitualmente para el descubrimiento de nuevas vacunas. Al emplear múltiples herramientas bioinformáticas para la selección de péptidos que promueven una respuesta inmune contra los patógenos, es posible mediante el uso de estas herramientas identificar, seleccionar y validar *in-silico* de manera rápida, péptidos de los patógenos altamente inmunogénicos candidatos a vacuna (que generan respuesta inmune), acelerando y simplificando el proceso de diseño de una vacuna. El enfoque integrado de múltiples herramientas bioinformáticas permite reducir la necesidad de realizar experimentos innecesarios en el laboratorio. Al seleccionar previamente los péptidos más prometedores y relevantes, se disminuye la cantidad de trabajo experimental requerido, lo que a su vez reduce los tiempos y los costos involucrados en el proceso. Esto resulta especialmente valioso en investigaciones que implican el tamizaje de grandes conjuntos de péptidos, lo que podría ser extremadamente laborioso y costoso si se llevara a cabo sin la guía de las herramientas bioinformáticas [34-36].

El SARS-CoV-2 es un Betacoronavirus que puede infectar a los humanos y ha dado lugar a múltiples variantes, de las cuales al menos cinco se han clasificado como variantes de preocupación [5-6]. El virus causante de la enfermedad se ha propagado a más de 768,2 millones de personas en todo el mundo y ha causado la muerte de más de 6,9 millones [3].

La proteína S del SARS-CoV-2 es una de las proteínas estructurales más importantes del virus y ha sido el principal objetivo de estudio para el desarrollo de vacunas. Esta proteína se compone de 1273 aminoácidos y se divide en dos subunidades: S1 y S2. Aunque el virus consta de múltiples proteínas estructurales, las proteínas de espícula (S), membrana (M), envoltura (E) y nucleocápside (N) son consideradas las más importantes en el contexto del SARS-CoV-2 por varias razones, incluyendo su función esencial en el ciclo de vida viral y su papel en la respuesta inmunitaria del huésped[8].

La proteína S (Spike) del virus SARS-CoV-2 es una proteína de superficie que permite la entrada del virus a la célula huésped. La proteína S reconoce el dominio 2 de la enzima transformadora de angiotensina humana (hACE2) presente en la membrana celular. La proteína S se divide en dos subunidades, S1 y S2, donde S1 funciona para reconocer el receptor de la célula receptora facilitando así la unión del virus a la célula y S2 se fusiona con la membrana de la célula receptora.

El estudio de los péptidos presentes en la proteína S, también conocidos como epítopos, es de suma importancia en el contexto de la respuesta inmune. Los epítopos son los fragmentos de proteínas que se unen al MHC y desencadenan una respuesta inmunitaria de las células T. Estos péptidos permiten que las células T detecten antígenos derivados de patógenos y también detecten la presencia de antígenos anormales expresados por células cancerosas. Una vez que se han reconocido estos antígenos (epítopo), las células T se multiplican y forman una población efectiva capaz de detectar el mismo epítopo en otras células infectadas o malignas. Además, se generan poblaciones de células T de memoria de larga duración, lo que permite al sistema inmunológico mantener un recuerdo inmunológico específico de ese epítopo particular. Esta memoria inmunológica es vital, ya que garantiza que si el organismo vuelve a encontrarse con el mismo epítopo en el futuro, la respuesta inmune será rápida y eficiente, ayudando a

controlar la infección o el crecimiento celular anormal de manera más efectiva. La identificación de epítopos que generen eficientemente la expansión de células de memoria son la base estructural y funcional de las vacunas.

Para la identificación de epítopos a nivel experimental existen al menos tres formas o tipos de ensayo que evalúan tres mecanismos; (i) ensayos que miden la unión del epítopo a moléculas MHC *in-vitro*, (ii) los ensayos que detectan péptidos de MHC presentados en células mediante su detección por espectrometría de masas, y (iii) los ensayos que miden la capacidad efectora de los linfocitos tras el reconocimiento del epítopo antigénico. En cuanto a la predicción de epítopos, se han desarrollado múltiples enfoques que tienen como objetivo producir una puntuación cuantitativa, relacionada entre otros parámetros con la afinidad o con la probabilidad de unión del epítopo a las moléculas mHC. Los enfoques *in-silico* más populares son los modelos de aprendizaje automático, los modelos no lineales como redes neuronales artificiales (ANN), modelos ocultos de Markov (HMM) y los modelos de regresión basados en afinidad. Existen varias herramientas de acceso libre para la predicción de epítopos con cada vez mejor capacidad predictiva.

1. Definición del Problema y Objetivos

Si bien el proteoma del SARS-CoV-2 no es tan extenso como el de algunos otros virus, tan solo la proteína S alberga una cantidad significativa de epítopos potencialmente inmunogénicos. Sin embargo, el proceso de selección de estos epítopos para los linfocitos T presenta desafíos debido a la diversidad y complejidad de las interacciones entre los péptidos y las moléculas del complejo mayor de histocompatibilidad (MHC) en los receptores HLA del sistema inmunitario. Las técnicas tradicionales para la identificación de epítopos, que implican pruebas *in-vitro* o *in-vivo*, son costosas en términos de tiempo e insumos, lo que limita la capacidad de evaluar un gran número de candidatos.

Por lo que nos planteamos la siguiente pregunta ¿Cómo implementar un pipeline bioinformático que permita integrar un grupo de herramientas útiles para identificar epítopos potencialmente inmunogénicos en SARS-CoV-2, con el objetivo de filtrar de una manera más precisa los candidatos?

Y con lo que planteamos un serie de objetivos para dar respuesta a esta pregunta:

1.1 Objetivo General

1.1.1 Desarrollar un modelo in-silico para identificar y priorizar epítopos clase I potencialmente inmunogénicas en SARS-CoV-2, con el fin de contribuir a la identificación de epítopos potencialmente útiles para vacunas profilácticas.

1.2 Objetivos Específicos

1.2.1 Seleccionar un grupo de herramientas bioinformáticas que permitan la predicción de epítopos potencialmente inmunogénicos

1.2.2 Desarrollar un pipeline bioinformático que permita priorizar los epítopos predichos haciendo uso del procesamiento integrativo de herramientas bioinformáticas

1.2.3 Predecir el universo de epítopos candidatos haciendo uso del pipeline desarrollado

1.2.4 Identificar epítopos altamente inmunogénicos dentro de los epítopos predichas mediante validación con inmunología reversa

2. SARS-CoV-2 y otros patógenos

2.1 SARS-CoV-2

El SARS-CoV-2, responsable de la enfermedad del coronavirus 2019 (COVID-19), ha tenido un impacto sin precedentes en la salud pública mundial desde su aparición. Este virus pertenece a la especie de coronavirus relacionados con el síndrome respiratorio agudo severo (SARS-CoV-2), y está estrechamente relacionado con el virus SARS-CoV-1, que causó el brote de síndrome respiratorio agudo severo (SARS) en 2002-2004 [53]. Desde su primer brote en Wuhan, China, en diciembre de 2019, el SARS-CoV-2 se ha propagado rápidamente por todo el mundo, provocando una pandemia global que ha afectado a millones de personas y ha ejercido una enorme presión sobre los sistemas de salud y la sociedad en general.

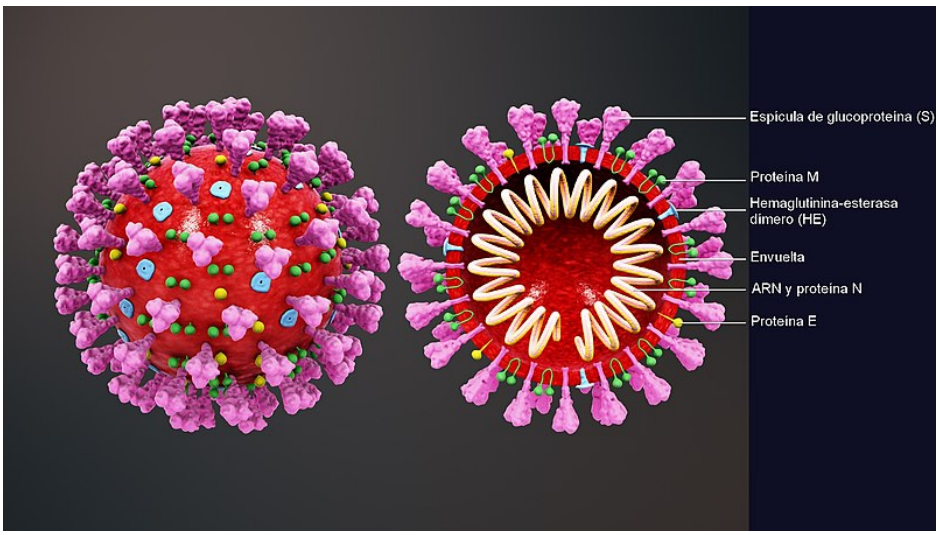


Figura 1. *Representación estructural SARS-CoV-2 [51]*

Una característica destacada de los virus es su capacidad de cambio y adaptación, y en el caso particular del SARS-CoV-2, esta capacidad es aún más notable debido a su naturaleza como virus de ARN. Los virus de ARN tienen una tasa de mutación más alta en comparación con los virus de ADN, lo que significa que pueden experimentar cambios genéticos más rápidos y frecuentes. A lo largo del tiempo, se han observado múltiples

mutaciones en el genoma del virus, y algunos de estos cambios pueden tener implicaciones significativas en la propagación del virus, la gravedad de la enfermedad y la eficacia de las intervenciones terapéuticas y preventivas, como vacunas y medicamentos [54]. Es esencial realizar un monitoreo constante de estas mutaciones por parte de las autoridades sanitarias, para comprender mejor el virus y adaptar las estrategias de respuesta en consecuencia.

Los síntomas asociados al SARS-CoV-2 pueden variar desde leves hasta graves, y pueden manifestarse en un período de incubación de 2 a 14 días después de la exposición al virus. Entre los síntomas más comunes se incluyen fiebre o escalofríos, tos, dificultad para respirar, fatiga, dolores musculares y corporales, dolor de cabeza y pérdida reciente del olfato o el gusto [55]. Sin embargo, es importante tener en cuenta que algunas personas pueden ser portadoras asintomáticas y, sin saberlo, transmitir el virus a otros individuos, lo que dificulta el control de su propagación.

La principal vía de transmisión del SARS-CoV-2 es a través de las gotas respiratorias expulsadas por personas infectadas al toser, estornudar, hablar o respirar. Estas gotas pueden ser inhaladas por personas cercanas o depositarse en las superficies, donde pueden permanecer viables durante un período de tiempo. Además, la transmisión también puede ocurrir al tocar superficies contaminadas y luego llevarse las manos a la cara, especialmente los ojos, la nariz o la boca [56]. Por lo tanto, es fundamental seguir las medidas preventivas recomendadas por las autoridades sanitarias, como el uso de mascarillas, el distanciamiento social y la higiene adecuada de las manos, para reducir el riesgo de contagio.

El SARS-CoV-2 ha demostrado ser un virus altamente transmisible y ha causado una pandemia global con consecuencias devastadoras para la salud pública y la sociedad en general. A medida que el virus continúa evolucionando, es crucial mantener una vigilancia constante de sus mutaciones y adaptar las estrategias de prevención y control en respuesta a los cambios observados. La investigación científica continua y el seguimiento de la propagación del virus son fundamentales para comprender mejor sus características y abordar eficazmente esta y futuras crisis sanitarias.

2.1.1 La proteína Spike

La proteína Spike del SARS-CoV-2, también conocida como proteína S, desempeña un papel fundamental en el proceso de reconocimiento del receptor y fusión de la membrana celular. Esta proteína está compuesta por dos subunidades, S1 y S2, que trabajan en conjunto para permitir la entrada del virus en las células huésped [57].

La subunidad S1 contiene un dominio de unión al receptor que se une al receptor de la enzima convertidora de angiotensina 2 (ACE2) presente en las células del huésped. Esta interacción entre la proteína Spike y el receptor ACE2 es crucial para que el virus pueda ingresar en las células [57]. Una vez que la proteína Spike se une al receptor ACE2, la subunidad S2 facilita la fusión de la membrana celular viral con la membrana celular huésped, permitiendo así que el material genético del virus penetre en la célula [57].

Debido a la importancia de la proteína Spike en el proceso de entrada del virus, se ha convertido en un objetivo prioritario para el desarrollo de medicamentos antivirales y vacunas. Al bloquear la interacción entre la proteína Spike y el receptor ACE2, es posible prevenir la infección por el virus [58]. En la actualidad, se han realizado intensas investigaciones científicas que han dado lugar a notables avances en el desarrollo de medicamentos y vacunas efectivas contra el SARS-CoV-2 [57]. Empresas farmacéuticas como Pfizer, AstraZeneca, Moderna, Johnson & Johnson, entre otras, han logrado desarrollar y producir vacunas altamente eficaces que han sido autorizadas para uso de emergencia y han sido administradas a millones de personas en todo el mundo.

El estudio de la proteína Spike y su papel en la infección viral es crucial para comprender mejor la biología del SARS-CoV-2 y para desarrollar estrategias efectivas de prevención y tratamiento. Además, el conocimiento detallado de esta proteína puede ayudar a identificar posibles variantes del virus que puedan surgir y a evaluar su impacto en la transmisión y gravedad de la enfermedad.

La proteína Spike del SARS-CoV-2 es un componente clave en el proceso de entrada del virus en las células huésped. Su interacción con el receptor ACE2 permite la unión y fusión de la membrana celular, lo que facilita la infección viral. El estudio de esta proteína es esencial para el desarrollo de medicamentos antivirales y vacunas efectivas contra el

virus, y su comprensión brinda información valiosa para abordar esta pandemia global de manera más eficaz.

2.2 *Mycobacterium tuberculosis*

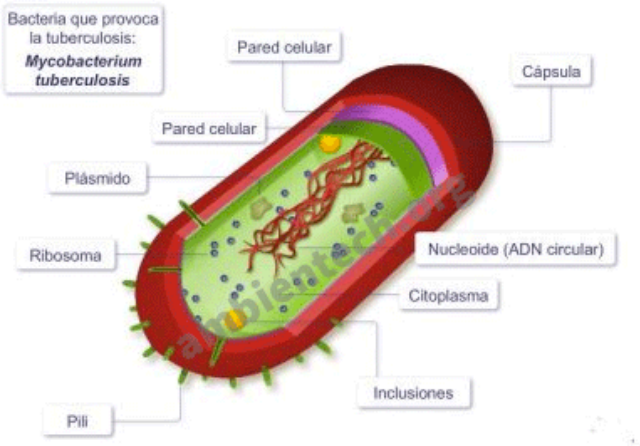


Figura 2. Representación estructural del Mycobacterium tuberculosis [52]

Mycobacterium tuberculosis (*M.tb*), bacteria responsable de la tuberculosis, es un patógeno altamente contagioso que puede afectar diversos órganos del cuerpo humano, siendo los pulmones la ubicación más común de la infección. Sin embargo, la TB también puede afectar otros órganos, como los huesos, el cerebro y la columna vertebral [59].

La transmisión de *M.tb* ocurre a través de las gotas infectadas que son liberadas al aire cuando una persona infectada tose, estornuda o habla. Estas gotas contienen las bacterias y pueden ser inhaladas por personas cercanas, lo que lleva a la propagación de la enfermedad. Es importante destacar que la exposición prolongada a una persona infectada aumenta el riesgo de adquirir la infección [59].

Las personas con sistemas inmunológicos debilitados tienen un mayor riesgo de contraer la tuberculosis. Esto incluye a personas con enfermedades crónicas como diabetes,

enfermedad renal, cáncer y VIH/SIDA, ya que su sistema inmunológico no puede combatir eficazmente la infección [60].

Afortunadamente, la tuberculosis es una enfermedad curable y prevenible. La vacuna más comúnmente utilizada para prevenir la TB es la vacuna BCG (bacilo de Calmette-Guérin). Esta vacuna es más efectiva en niños y su eficacia en adultos varía. A pesar de que no es completamente protectora contra todas las formas de tuberculosis, la vacuna BCG puede ayudar a reducir la gravedad de la enfermedad en caso de infección [60].

El tratamiento de la tuberculosis generalmente implica una combinación de medicamentos y cuidados adecuados; la infección de tuberculosis latente puede usar diferentes esquemas posológicos con medicamentos como isoniacida (INH), rifapentina (RPT) o rifampina (RIF). Se recomienda prescribir los tratamientos más cortos y convenientes siempre que sea posible, ya que esto aumenta la probabilidad de que los pacientes completen el tratamiento. Si el paciente ha estado en contacto con una persona con tuberculosis resistente a los medicamentos, el tratamiento debe ser modificado y se debe consultar a un experto en tuberculosis en esos casos. Las opciones de tratamiento varían en duración y frecuencia, y se deben tener en cuenta consideraciones como el VIH/sida, el embarazo y la resistencia a ciertos medicamentos al elegir el esquema adecuado. Es importante tener en cuenta las recomendaciones de los CDC sobre la combinación de rifampina (RIF) y piracinamida (PZA) debido a los informes de lesiones graves en el hígado y muertes asociadas a esta combinación [4]. Para determinar el tratamiento más adecuado, se realiza un cultivo de la bacteria *Mycobacterium tuberculosis* y se analizan sus sensibilidades a los antibióticos. El tratamiento suele requerir una combinación de diferentes medicamentos que se deben tomar durante varios meses para asegurar la erradicación completa de las bacterias [59].

En casos de tuberculosis resistente a los medicamentos, se necesitan terapias de segunda línea más potentes o incluso antibióticos intravenosos. La tuberculosis multirresistente, que es más difícil de tratar, puede requerir un enfoque terapéutico aún más complejo y prolongado [59].

La detección temprana, el diagnóstico preciso y el tratamiento oportuno son fundamentales para controlar y prevenir la propagación de la tuberculosis. Además, es importante fomentar medidas de prevención, como la educación sobre higiene respiratoria, el uso de mascarillas en casos necesarios y el mantenimiento de un sistema inmunológico saludable, especialmente, en personas con factores de riesgo.

A pesar de años de intensa investigación y de uso de la vacuna BCG, esta sigue siendo la única vacuna autorizada con una eficacia muy variable. La BCG brinda protección contra la tuberculosis infantil, pero no es eficaz para prevenir la tuberculosis pulmonar del adulto por lo cual la tuberculosis continúa siendo la enfermedad infecciosa principal causa de muerte en todo el mundo, provocando alrededor de 1,6 millones de muertes cada año. Esta situación se ha vuelto más complicada por la aparición de cepas de *Mycobacterium tuberculosis* resistente a los medicamentos y a la coinfección de los individuos adultos con el VIH. Por estas razones la infección con el bacilo ha empeorado significativamente el pronóstico, tratamiento y el control de la tuberculosis a nivel mundial. Si bien hay muchas vacunas candidatas nuevas en desarrollo clínico y muchas más en ensayos preclínicos que tienen como objetivo reemplazar o reforzar la vacuna BCG, la disponibilidad de una nueva vacuna que disminuya la mortalidad y prevenga los casos nuevos de tuberculosis sigue siendo para los investigadores un verdadero desafío.

3. Presentación de antígenos y respuesta inmune

3.1 Presentación de antígenos

La presentación de antígenos es un proceso crucial en la respuesta inmunitaria que permite que los linfocitos T reconozcan y respondan a los antígenos. Las células presentadoras de antígenos (CPA) desempeñan un papel central en este proceso al procesar y presentar péptidos derivados de los antígenos en su superficie en asociación con moléculas del MHC [61].

Existen dos tipos principales de presentación de antígenos: la presentación de antígenos por moléculas MHC de clase I y la presentación de antígenos por moléculas MHC de clase II. La presentación de antígenos por moléculas MHC de clase I ocurre en todas las células nucleadas del cuerpo y permite la presentación de péptidos derivados de antígenos intracelulares a los linfocitos T CD8 citotóxicos. Esto es especialmente relevante en el contexto de las infecciones virales, donde las células infectadas pueden ser reconocidas y destruidas por los linfocitos T CD8, evitando así la propagación viral [61].

Por otro lado, la presentación de antígenos por moléculas MHC de clase II ocurre principalmente en células presentadoras de antígenos profesionales como células dendríticas, macrófagos y células B. Estas células son capaces de capturar antígenos extracelulares y presentarlos a los linfocitos T helper CD4. La interacción entre los linfocitos T CD4 y las células presentadoras de antígenos desencadena una respuesta inmunitaria coordinada que incluye la activación de

otros tipos de células inmunitarias y la producción de anticuerpos específicos para el antígeno [61].

La presentación de antígenos es un componente esencial en la respuesta inmunitaria frente a diferentes patógenos, incluyendo el SARS-CoV-2 y *Mycobacterium tuberculosis*. En el caso del SARS-CoV-2, las células infectadas por el virus pueden presentar péptidos derivados de antígenos virales en su superficie a través de moléculas MHC de clase I, lo que activa una respuesta de linfocitos T CD8 citotóxicos para combatir la infección. Por otro lado, en el caso de *Mycobacterium tuberculosis*, las células presentadoras de antígenos profesionales como las células dendríticas y los macrófagos presentan péptidos derivados de antígenos bacterianos en moléculas MHC de clase II, lo que activa una respuesta de linfocitos T helper CD4 para controlar la infección [61].

Comprender los mecanismos de presentación de antígenos y su relación con la respuesta inmunitaria es fundamental para el desarrollo de estrategias terapéuticas y vacunas eficaces contra diversos patógenos, incluyendo el SARS-CoV-2 y *Mycobacterium tuberculosis*. Un enfoque importante en la investigación científica actual es identificar y desarrollar vacunas que estimulen una respuesta inmunitaria protectora a través de la presentación de antígenos adecuada, mejorando así la capacidad del sistema inmunitario para combatir estas infecciones y ayudar a prevenirlas.

3.1.1 Linfocitos T CD8 citotóxicos

Los linfocitos T CD8 citotóxicos, también conocidos como células T citotóxicas son un tipo de célula inmunitaria que desempeña un papel crucial en la respuesta inmune adaptativa. Estas células son fundamentales para la eliminación de células infectadas por virus y células tumorales, ya que reconocen y atacan

específicamente a las células que presentan antígenos en su molécula MHC de clase I.

El proceso de activación de los linfocitos T CD8 citotóxicos comienza cuando una célula presentadora de antígenos, como una célula dendrítica o un macrófago, presenta un péptido derivado de un antígeno intracelular en su molécula MHC de clase I. Este complejo antígeno-MHC de clase I es reconocido por los receptores para el antígeno de los linfocitos T CD8 citotóxicos, conocidos como el receptor de células T (TCR), que sumada a la activación de receptors y sus respectivos ligandos expresados tanto en la superficie de las células T como de las células presentadoras del antígeno (moléclas coestimuladoras), generan la activación de este linfocito T estimulado [62].

Una vez que los linfocitos T CD8 citotóxicos han sido activados, son capaces de reconocer y destruir las células que presentan el mismo péptido en la molécula CMH de clase I. Estos linfocitos T liberan sustancias como la perforina, que forma poros en la membrana de la célula diana, y las granzimas, que se introducen en la célula a través de estos poros e inducen la muerte celular programada o apoptosis en la célula objetivo [62].

Además de su capacidad de matar directamente a las células infectadas o tumorales, los linfocitos T CD8 citotóxicos también secretan citocinas, como el interferón gamma (IFN-γ), que desempeñan un papel importante en la regulación y coordinación de la respuesta inmunitaria. El IFN-γ puede estimular a otras células inmunitarias, como los macrófagos, para que aumenten su actividad fagocítica y destruyan patógenos, y también modula eficientemente la respuesta de otros linfocitos T como los linfocitos T CD4 y a las células dendríticas que son las células presentadoras del antígeno profesionales [62].

La función de los linfocitos T CD8 citotóxicos en la respuesta inmune adaptativa es esencial para la eliminación eficiente de células infectadas y la prevención de la propagación de la infección. Su capacidad de reconocer y eliminar selectivamente a las células que presentan antígenos específicos en su molécula MHC de clase I los convierte en una defensa poderosa contra las infecciones virales y el desarrollo de tumores. Estos linfocitos T son clave en la inmunidad antiviral y juegan un papel importante en la respuesta inmunitaria contra el SARS-CoV-2, donde la activación de los linfocitos T CD8 citotóxicos contribuye a la eliminación del virus y la recuperación del individuo infectado. Su capacidad para secretar citocinas también ayuda a modular y coordinar la respuesta inmunitaria general, mejorando así la eficacia de la respuesta inmune adaptativa y la capacidad de combatir las enfermedades infecciosas y el cáncer.

3.1.2 Linfocitos T helper CD4

Los linfocitos T CD4+, también conocidos como células T helper o linfocitos T colaboradores, desempeñan un papel esencial en la respuesta inmunitaria contra infecciones y enfermedades. Estas células son responsables de coordinar y regular la respuesta inmunitaria, interactuando con otras células del sistema inmunitario para combatir los antígenos y promover una respuesta eficaz [61].

Una de las características distintivas de los linfocitos T CD4+ es la expresión de la molécula CD4 en su superficie. Al igual que los linfocitos T CD8+ los linfocitos T CD4+ expresan un receptor específico para el antígeno o TCR que permite la interacción y el reconocimiento de los antígenos presentados en las moléculas del complejo mayor de histocompatibilidad de clase II (MHC-II) en las células presentadoras de antígenos profesionales, como las células dendríticas, los macrófagos y las células B. A través de esta interacción, los linfocitos T CD4+ desencadenan una serie de respuestas inmunológicas, estas respuestas inmunológicas coordinadas por los linfocitos T CD4+ ayudan a eliminar patógenos

invasores, coordinar la respuesta inmunitaria y mantener la homeostasis del sistema inmunológico [62].

La principal función de los linfocitos T CD4+ es la de proporcionar ayuda y apoyo a otras células del sistema inmunitario, como los linfocitos B y los linfocitos T CD8+. Estas células T colaboradoras secretan citocinas, como interleucinas y interferones, que regulan y modulan la respuesta inmunitaria. Estas señales ayudan a activar y potenciar la función de los linfocitos B para que produzcan anticuerpos específicos contra los antígenos, así como a activar y mejorar la capacidad de los linfocitos T CD8+ para destruir células infectadas [62].

Es importante destacar que los linfocitos T CD4+ desempeñan un papel crítico en la respuesta inmunitaria contra las infecciones por el virus de la inmunodeficiencia humana (VIH). El VIH específicamente infecta y destruye los linfocitos T CD4+, lo que debilita progresivamente el sistema inmunitario y lleva a la inmunodeficiencia. Como resultado, las personas infectadas con VIH tienen un mayor riesgo de contraer infecciones oportunistas y desarrollar cánceres relacionados con la inmunosupresión [62]. La disminución de los linfocitos T CD4+ y la disfunción inmunológica asociada a la infección por VIH han sido un área de gran interés en la investigación médica, y el desarrollo de terapias antirretrovirales ha sido fundamental para controlar la progresión de la enfermedad y mejorar la calidad de vida de las personas afectadas.

Los linfocitos T CD4+ son células clave en la respuesta inmunitaria adaptativa. Su capacidad para coordinar y regular la respuesta inmunitaria, así como para proporcionar ayuda a otras células del sistema inmunitario, los convierte en actores fundamentales en la defensa contra infecciones y enfermedades. Estos linfocitos tienen una relevancia crucial en diversas infecciones, incluyendo el VIH, la COVID-19 y la tuberculosis, cada una de las cuales presenta desafíos específicos para el sistema inmunitario.

En el contexto del VIH, los linfocitos T CD4+ son las células que el virus infecta y destruye progresivamente, debilitando el sistema inmunitario del organismo. Como resultado, las personas con infección por VIH tienen una mayor susceptibilidad a infecciones oportunistas y enfermedades graves debido a su inmunidad comprometida. La disminución de los niveles de linfocitos T CD4+ es un marcador clave de la progresión de la infección por VIH y es utilizado para monitorear la salud inmunológica de los pacientes.

En el caso de la COVID-19, los linfocitos T CD4+ también desempeñan un papel central en la respuesta inmunitaria contra el virus SARS-CoV-2. Estas células T ayudan a coordinar la acción de otras células del sistema inmunitario, como los linfocitos T CD8+ y las células B, para combatir la infección. Además, la memoria inmunológica generada por los linfocitos T CD4+ después de la infección o la vacunación es fundamental para una respuesta más rápida y efectiva en futuros encuentros con el virus.

En el caso de la tuberculosis, los linfocitos T CD4+ también juegan un papel crucial en la defensa contra la bacteria *Mycobacterium tuberculosis*. Estas células T ayudan a activar a los macrófagos para que puedan fagocitar y destruir a la bacteria, limitando así su propagación. La capacidad de los linfocitos T CD4+ para formar granulomas, estructuras que encapsulan a las bacterias, es esencial para mantener la infección bajo control y evitar su diseminación.

Su papel en la coordinación y regulación de la respuesta inmunitaria, así como su capacidad para brindar ayuda a otras células del sistema inmunitario, los convierte en una pieza clave en la protección del organismo contra patógenos y enfermedades. El estudio y comprensión de la función de los linfocitos T CD4+ en distintos contextos infecciosos son fundamentales para el desarrollo de estrategias terapéuticas y vacunas efectivas que fortalezcan nuestra capacidad de combatir estas enfermedades.

3.2 Respuesta inmune

La respuesta inmune es un proceso vital del organismo para protegerse contra agentes invasores, como patógenos (virus, bacterias, hongos, parásitos), sustancias extrañas y potencialmente dañinas y contra las células tumorales. Esta respuesta se activa cuando el sistema inmunitario reconoce la presencia de microorganismos, células tumorales o toxinas que pueden causar enfermedades e infecciones.

La respuesta inmune se divide en dos tipos principales: la respuesta inmune innata y la respuesta inmune adaptativa. La respuesta inmune innata es la primera línea de defensa del cuerpo y actúa de manera rápida y generalizada ante una amenaza. Incluye barreras físicas, como la piel y las mucosas, que previenen la entrada de microorganismos y otras sustancias extrañas al cuerpo. Además, la respuesta inmune innata cuenta con células y proteínas especializadas que pueden reconocer y atacar a los invasores. Estas células, como los macrófagos y los neutrófilos, son capaces de fagocitar y destruir a los microorganismos invasores [63].

Por otro lado, la respuesta inmune adaptativa es más específica y se desarrolla a medida que el sistema inmunitario se encuentra con antígenos específicos. Los antígenos son sustancias extrañas que desencadenan una respuesta inmune. La respuesta inmune adaptativa involucra la participación de células especializadas, como los linfocitos B y los linfocitos T, que pueden reconocer y responder a antígenos específicos. Los linfocitos B producen y liberan anticuerpos, que son proteínas capaces de unirse a los antígenos y neutralizarlos, mientras que los linfocitos T pueden destruir células infectadas o tumorales que presenten esos antígenos [50].

Una característica fundamental de la respuesta inmune adaptativa es su capacidad de generar memoria inmunológica. Esto significa que una vez que el

sistema inmunitario ha encontrado y reconocido un antígeno, se produce una respuesta más rápida y eficiente en encuentros futuros con el mismo antígeno. Esta memoria inmunológica es la base de la protección a largo plazo contra enfermedades infecciosas, ya que el sistema inmunitario puede montar una respuesta más efectiva para controlar la infección antes de que cause daño significativo y la generación de linfocitos de memoria son la razón de las vacunas [50].

La interacción entre la respuesta inmune innata y adaptativa es crucial para una protección completa del organismo contra enfermedades. La respuesta inmune innata brinda una respuesta rápida y no específica ante una amenaza, mientras que la respuesta inmune adaptativa se encarga de la generación de células de memoria que favorezcan la rápida eliminación específica de los patógenos o de las células tumorales. Ambos componentes trabajan en conjunto para asegurar una protección efectiva y duradera contra las sustancias extrañas y patógenos.

4. Herramientas bioinformáticas en el diseño de vacunas

Las herramientas bioinformáticas han desempeñado un papel importante en el diseño de vacunas en los últimos años. Una de las metodologías más importantes en este campo es la vacunología inversa, qué es una metodología *in-silico* que estudia diferentes características de los agentes infecciosos para identificar antígenos que sean candidatos a vacunas omitiendo algunas fases de las fases de selección y evaluación de candidatos a vacuna que demandan trabajo experimental en el laboratorio [49].

La vacunología inversa es una de estas estrategias que se basa en el uso de herramientas bioinformáticas y tecnologías ómicas, como la genómica, transcriptómica, proteómica, interactómica, metabolómica y biología de sistemas, para identificar posibles dianas vacunales y farmacológicas. Estas herramientas permiten analizar y comprender de manera integral el complejo funcionamiento del sistema inmunológico y las interacciones entre proteínas y moléculas en el contexto de una enfermedad o infección [48].

A través de la vacunología inversa, los científicos pueden utilizar la información genómica y estructural de los patógenos o agentes infecciosos para identificar proteínas o componentes específicos que puedan ser utilizados como blancos para desarrollar vacunas. Al analizar la estructura tridimensional de las proteínas diana, se pueden diseñar vacunas de manera más precisa y efectiva, lo que aumenta la probabilidad de generar una respuesta inmunitaria adecuada.

Las estrategias bioinformáticas combinadas con los últimos diseños de vacunas de nueva generación permiten la selección de candidatos a vacunas en un corto período de tiempo, lo cual es muy importante en el desarrollo de nuevas vacunas contra patógenos

con potencial pandémico [49]. Los avances en tecnologías de producción de vacunas han sido igualmente significativos. Las plataformas de producción basadas en ARN mensajero (ARNm) han demostrado ser particularmente prometedoras, ya que permiten una producción más rápida y flexible de vacunas. Esta tecnología innovadora ha sido clave en el desarrollo de vacunas COVID-19 altamente efectivas en un tiempo récord. Además, las plataformas de producción de vectores virales y proteínas recombinantes también han experimentado mejoras significativas, lo que ha permitido una producción más escalable y eficiente de vacunas.

Las herramientas bioinformáticas son utilizadas en el diseño de vacunas para identificar antígenos que sean promisorios candidatos a vacunas. Esto se logra mediante el análisis de grandes conjuntos de datos genómicos, transcriptómicos y proteómicos para identificar genes y proteínas relacionados con la patogenicidad, virulencia o inmunogenicidad de un agente infeccioso.

Una vez que se han identificado los antígenos candidatos, las herramientas bioinformáticas también pueden ser utilizadas para diseñar y optimizar la secuencia de aminoácidos de la proteína antigénica para mejorar su inmunogenicidad y estabilidad. Además, las herramientas bioinformáticas pueden ser utilizadas para predecir la estructura tridimensional de la proteína antigénica y para diseñar epítopos que puedan ser reconocidos por el sistema inmune.

En pocas palabras, el uso de herramientas bioinformáticas en el diseño de vacunas ha transformado la forma en que se identifican y seleccionan los antígenos candidatos. Estas herramientas permiten un enfoque más rápido, preciso y personalizado en el desarrollo de vacunas, lo que representa un avance significativo en la lucha contra enfermedades infecciosas y la preparación ante posibles pandemias.

5. Pipeline bioinformático

Un *pipeline* bioinformático es una serie de pasos computacionales automatizados que se utilizan para analizar datos biológicos. Un *pipeline* típico puede incluir la recopilación de datos, el preprocesamiento, el análisis y la visualización de los resultados.

Los *pipelines* bioinformáticos son ampliamente utilizados en la investigación genómica y proteómica para analizar grandes conjuntos de datos. Por ejemplo, un *pipeline* de secuenciación del genoma puede incluir pasos para el ensamblaje de lecturas de secuenciación, la anotación de genes y la identificación de variantes genéticas. Los *pipelines* bioinformáticos son útiles porque automatizan tareas repetitivas y reducen el riesgo de errores humanos. También pueden ser diseñados para ser escalables y para manejar grandes conjuntos de datos.

En el año 1989, Sette y sus colegas describieron un programa de computadora para alelos MHC que identifica ligandos potenciales en una secuencia de proteínas [25]. Desde entonces se han desarrollado varios enfoques que tienen como objetivo producir una puntuación cuantitativa, relacionada con la afinidad de unión predicha o con la probabilidad de unión; los enfoques más populares se encuentran en los modelos de aprendizaje automático [24], los modelos no lineales como redes neuronales artificiales (ANN), modelos ocultos de Markov (HMM) y los modelos de regresión basados en afinidad [22-24]. Basándose en esto, se han desarrollado múltiples herramientas de acceso libre con cada vez mejor capacidad predictiva de epítopos [27]; la mayoría de estas herramientas se han entrenado utilizando datos de ensayos de afinidad, aunque muchas de ellas también incorporan péptidos identificados por análisis de unión de HLA [28]; todo esto con el objetivo de optimizar el proceso de selección de péptidos, en vista de que ya se ha comprobado que con la identificación de estos epítopos inmunogénicos dentro de una proteína objetivo, se puede informar y guiar mejor el diseño de las vacunas

[29]. Desde el inicio de la pandemia por COVID-19 y debido a su importancia se han presentado gran cantidad de publicaciones muy relevantes que proponen múltiples enfoques para afrontar esta problemática, como Prachar 2020 [27] quien identificó y validó 174 epítopos los cuales predijeron y validaron *in-vitro.* Orsburn 2020 [30] describió un método para la selección rápida *in-silico* de péptidos descritos y aplicados a SARS-CoV-2, partiendo de una lista de 496 péptidos, terminó con un listado de 24 candidatos; Sitthiyotha 2020, [31] con el uso de Rosseta para mejorar la afinidad de unión de SPB25 al SARS-CoV-2-RBD y evitar interrumpir las interacciones favorables mediante el uso de residuos que no se han informado, logró identificar cinco péptidos diseñados. Todos estos péptidos reportados por diferentes autores y metodologías permiten realizar una comparación de los resultados obtenidos por el *pipeline* desarrollado.

6. Metodología de desarrollo

Con el fin de cumplir los objetivos se planteó la metodología de desarrollo que se ilustra en la Figura 3 y que se describe a continuación.

Figura 3. Representación gráfica de la metodología de desarrollo implementada.

- Selección de herramientas: Se planteó realizar una búsqueda de diferentes *benchmarks* en la literatura sobre herramientas disponibles para HLA-1. Esta búsqueda debía incluir herramientas de diferentes fuentes, como bases de datos públicas, artículos científicos y sitios web de desarrolladores.

  Los datos recolectados se utilizaron para realizar un metanálisis, que es una técnica estadística que permite combinar los resultados de diferentes estudios

para obtener una estimación más precisa del efecto de una intervención. En este caso, el efecto de interés era la precisión y eficiencia de las diferentes herramientas para la predicción de epítopes HLA-1.

- Selección de los alelos más prevalentes en la población colombiana: Se planteó realizar una búsqueda en literatura, o base de datos especializada de los alelos. La selección de los alelos más prevalentes en la población colombiana era importante para garantizar que los péptidos candidatos a vacunas sean presentados por la mayoría de las personas.

Figura 4. Representación metodología ágil SCRUM [87].

- Planeación y desarrollo del *pipeline* bioinformático: El pipeline bioinformático es un conjunto de herramientas y procesos que se utilizan para predecir epítopes HLA-1. El pipeline debe ser diseñado para ser eficiente y preciso.

  En este caso, se propuso utilizar una arquitectura de desarrollo ágil para el desarrollo del pipeline. El desarrollo ágil es un enfoque iterativo para el desarrollo de software que permite entregar rápidamente los resultados a los usuarios, que incluye las siguientes etapas:

  - Análisis: En esta etapa, se recopilan los requisitos del software, que son las características y funciones que el software debe tener. Los requisitos se pueden recopilar de diversas fuentes, como los clientes, los usuarios, los expertos en el dominio y los propios desarrolladores.

  - Diseño: En esta etapa, se crea un modelo del software que describe su estructura, comportamiento y datos. El diseño debe ser lo suficientemente detallado para que los desarrolladores puedan implementar el software, lo suficientemente flexible para adaptarse a los cambios en los requisitos.
  - Desarrollo: Es la etapa en la que se crea el código del software. En esta etapa, los desarrolladores traducen el diseño en código. El código debe ser legible, mantenible y eficiente.
  - Testing: Es la etapa en la que se verifica que el software cumple con los requisitos. En esta etapa, se ejecutan pruebas para encontrar errores y problemas de rendimiento. Las pruebas se pueden realizar manualmente o utilizando herramientas automatizadas.
  - Implementación: Es la etapa en la que se pone en funcionamiento el software. En esta etapa, el software se instala en los sistemas de los usuarios y se pone a disposición de los usuarios finales.

- Selección del universo de péptidos candidatos a vacunas: Se realizó utilizando el proteoma curado de SARS-CoV-2, es un conjunto de proteínas que se han identificado y secuenciado. El uso de un proteoma curado garantiza que los péptidos candidatos a vacunas sean representativos de las proteínas que se encuentran en el virus. Esto para ejecutar el pipeline desarrollado y obtener el universo de péptidos candidatos a vacunas.

- Revisión en literatura de péptidos reportados por diferentes autores (Inmunología reversa): Se planteó realizar una exhaustiva búsqueda en la literatura de publicaciones sobre péptidos con y sin sustento experimental, esto con la finalidad de poder validar los resultados obtenidos con el *pipeline* desarrollado

- Comparación de resultados: La comparación de resultados es una forma de validar la efectividad de precisión del *pipeline*. En este caso, se propone utilizar otras herramientas similares como pVACtools para comparar los péptidos generados por el *pipeline* con los obtenidos de la literatura.

7. Resultados y Discusión

Durante el desarrollo del proyecto, se implementaron varios *pipelines* bioinformáticos con el objetivo de simplificar y mejorar la eficiencia de diversos procesos. El primer *pipeline* se diseñó con el fin de pre-procesar y generar los archivos y formatos necesarios para ser procesados por el *pipeline* de pVACseq (*personalized Variant Antigens by Cancer Sequencing*) es una plataforma de inmunoterapia contra el cáncer utilizada para la identificación de Antígenos Variantes Personalizados mediante Secuenciación de células tumorales. A partir de este primer *pipeline*, se estableció el proceso base para el *pipeline* de SARS-CoV-2 que se basó en gran parte de los procesos internos del *pipeline* de pVACseq ajustado y aplicado en la identificación de epítopos candidatos a vacunas [7].

7.1.1 Pipeline pVACseq

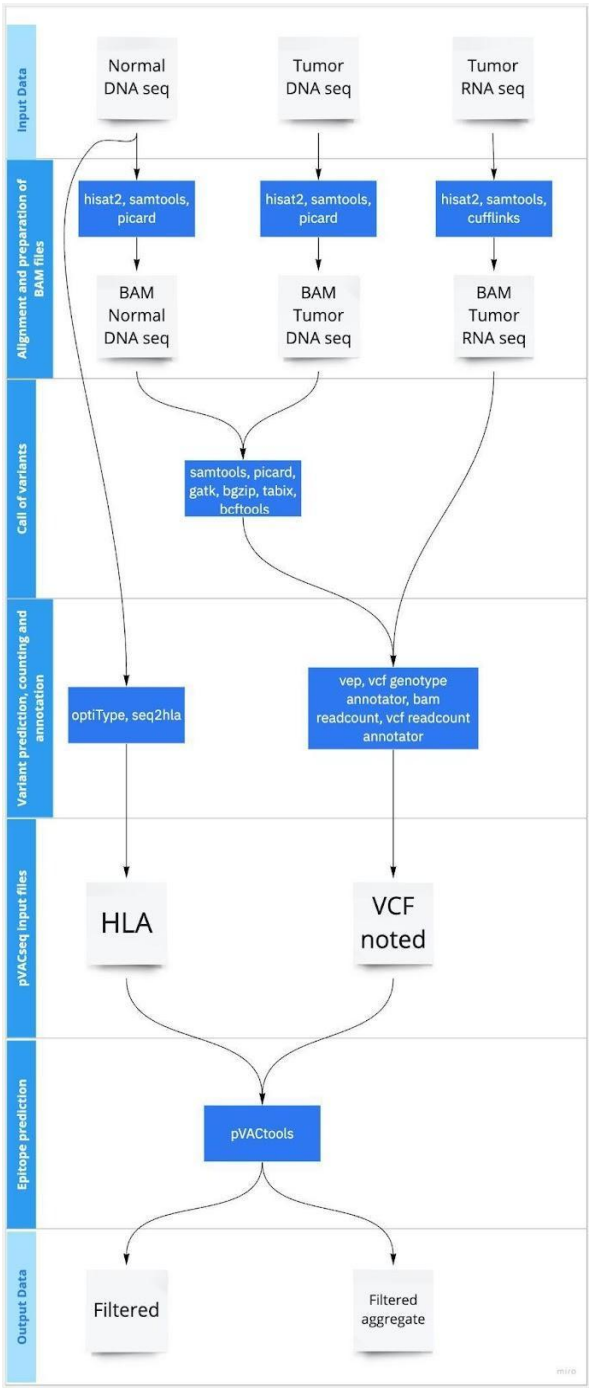


Figura 5. Representación del *pipeline* pVACseq [7]: Representa paso a paso del pipeline bioinformático de insumos, formatos y resultados

El *pipeline* diseñado se ejecuta de manera secuencial en cinco niveles distintos, como se ilustra en la Figura 5. Cada nivel desempeña un papel específico en el procesamiento y análisis de los datos genómicos y de expresión.

En el primer nivel del *pipeline*, se alinean y preparan los archivos BAM (*Binary Aligned Mapped*) utilizando herramientas como hisat2, samtools y picard [9-11]. Estas herramientas procesan los datos de entrada, que incluyen DNA normal, DNA tumoral y RNA tumoral, para generar archivos BAM alineados y mapeados correctamente.

En el segundo nivel, se realiza el llamado de variantes a partir de los archivos BAM utilizando herramientas como samtools, picard, gatk, bgzip, tabix y bcftools [10-11][17-21]. Estas herramientas permiten identificar y registrar las variantes genéticas presentes en los datos de secuenciación, generando archivos VCF (*Variant Call Format*) que contienen información detallada sobre las variantes encontradas [37].

En el tercer nivel, se lleva a cabo la precisión, conteo y anotación de las variantes detectadas en el nivel anterior. Para esto, se utilizan herramientas como optiType, seq2hla, vep (*Variant Effect Predictor*), vcf genotype annotator, bam readcount y vcf readcount annotator [38-41]. Estas herramientas permiten determinar la presencia de variantes específicas en los genes HLA, realizar un análisis detallado de los efectos de las variantes en la estructura y función de las proteínas, y anotar las variantes con información relevante para su interpretación.

En el cuarto nivel, se obtienen los insumos necesarios para la predicción de epítopos. Esto incluye la información de los alelos HLA del paciente y un archivo VCF anotado con las variantes detectadas. Estos datos se utilizan como entrada para la herramienta pVACseq (*personalized Variant Antigens by Cancer Sequencing*), que se encarga de realizar la predicción de los epítopos específicos de cada paciente de cuyo tumor proviene la información.

En el quinto y último nivel del *pipeline*, se procesan y obtienen los epítopos predichos por la herramienta pVACseq. Estos epítopos representan las secuencias peptídicas potenciales que podrían ser reconocidas por el sistema inmunológico en el contexto del

cáncer, y desencadenar una respuesta inmune específica tras el reconocimiento y activación de un linfocito T que presente un receptor de célula T (TCR) con afinidad hacia el complejo MHC-péptido expresado por el haplotipo de HLA del paciente.

En conjunto, este *pipeline* secuencial permite realizar un análisis exhaustivo de los datos genómicos y de expresión, desde la alineación y llamado de variantes hasta la predicción de epítopos personalizados. Cada nivel del *pipeline* cumple una función crucial en el procesamiento y análisis de los datos, brindando resultados que podrían llegar a ser muy relevantes para la comprensión y tratamiento del cáncer mediante la implementación de vacunas personalizadas basadas en los neoantígenos tumorales identificados en el tumor de cada paciente.

7.2 Selección de herramientas

Para la selección de las herramientas se tuvieron en cuenta análisis de *benchmarks* en los cuales se comparaban sus métodos de cálculo para la puntuación de predicción, los algoritmos empleados, las estrategias de evaluación y las funcionalidades del software. Además, se ha evaluado el rendimiento de predicción de las herramientas revisadas en función de un conjunto de datos de validación independiente, que contiene 21.101 ligandos verificados experimentalmente en 19 alomorfos de HLA-I [35].

El rendimiento de diferentes herramientas se evaluó en términos de cinco métricas de uso común, a saber, AUC, Sn, Sp, Acc y MCC, utilizando el conjunto de datos de validación como entrada. Cabe señalar que los conjuntos de datos de entrenamiento de algunas herramientas revisadas no están disponibles, mientras que otras herramientas se han actualizado con un conjunto de datos de entrenamiento ampliado desde su primer lanzamiento. Por lo tanto, podría haber cierta superposición entre los conjuntos de datos utilizados para desarrollar algunas herramientas y nuestro conjunto de datos de validación. Siempre que fue posible, descargamos los conjuntos de datos de entrenamiento de estas herramientas y eliminamos las entradas superpuestas con las de nuestro conjunto de datos de validación. Luego, enviamos las secuencias de péptidos de 9, 10 y 11 meros específicas para cada alotipo HLA-I en el conjunto de datos de validación a las herramientas. Para esta evaluación, los parámetros de las herramientas se establecieron en las configuraciones recomendadas en las publicaciones correspondientes o en los valores predeterminados, si no se dieron recomendaciones.

MixMHCpred 2.0.1 logró el mejor rendimiento entre todas las herramientas basadas en funciones de puntuación, ya que logró los valores de AUC más altos entre casi todos los 19 alotipos, mientras que NetNHCpan 4.0 obtuvo el mejor rendimiento entre todas las herramientas basadas en aprendizaje automático y NetMHCcons 1.0 logró un mejor rendimiento que IEDB-ARConsensus en la categoría de consenso.

Si bien ninguna herramienta logró universalmente el mejor rendimiento para todos los alotipos de HLA-I en el conjunto de datos de prueba independiente, MixMHCpred 2.0.1 se desempeñó mejor para la mayoría de los alotipos de HLA-I examinados. Conjeturamos que una razón por la que MixMHCpred 2.0.1 logra el mejor rendimiento es porque es una herramienta publicada recientemente que se entrenó con fuentes de datos públicas de péptidos HLA de 40 líneas celulares y también con datos internos de experimentos de purificación de inmunoafinidad que involucraron 10 líneas celulares adicionales. Además, MixMHCpred 2.0.1 aplicó estrategias de aprendizaje automático totalmente no supervisadas y semi-supervisadas para identificar un total de 52 motivos de unión específicos del alotipo HLA-I [35].

También se tuvo en cuenta el análisis realizado en el artículo de la publicación de MCHflurry donde el enfoque de evaluación comparativa utilizado para evaluar el rendimiento de MHCflurry, respecto a NetMHC y NetMHCpan utiliza dos conjuntos de datos:

El conjunto de datos ABELIN, que contiene péptidos eluidos de la superficie celular del MHC e identificados por espectrometría de masas, y el conjunto de datos del VPH, que contiene mediciones de afinidad generadas a través de un proyecto de desarrollo de vacunas contra el VPH. Todos los registros en ambos puntos de referencia son distintos de los registros en el conjunto de datos TRAIN.

Para crear el punto de referencia ABELIN, se muestrearon secuencias no observadas (señuelos) de las transcripciones codificadoras de proteínas que contenían los péptidos identificados (aciertos) en función de las secuencias de proteínas en el proteoma UCSC hg19 y las cuantificaciones de transcripciones de la secuenciación de ARN de la línea celular relevante. Para cada alelo con n aciertos, se muestrearon 100n señuelos, ponderando las transcripciones por el número de aciertos y muestreando un número igual de señuelos de cada longitud 8-15. Esto produjo 2,045,100 señuelos para 20,451 aciertos, de los cuales se eliminaron 118 (0.005%) entradas, también presentes en el conjunto de datos TRAIN, para un punto de referencia de 20,361 aciertos y 2,045,072 señuelos. La precisión de cada predictor para diferenciar los aciertos de los señuelos se evaluó en términos de valor predictivo positivo (PPV). Para calcular el PPV para un alelo con n aciertos, se clasificaron los n + 100n aciertos y señuelos de mayor a menor afinidad de unión prevista y se calculó la fracción de los n péptidos principales que eran aciertos.

Además del MHCflurry, NetMHC y NetMHCpan, se consideraron siete variaciones de la arquitectura de MHCflurry en el punto de referencia ABELIN. Las variantes cambiaron uno o dos aspectos de la arquitectura o los datos de entrenamiento y, por lo demás, fueron idénticos a MHCflurry 0.9.1. Para cada arquitectura, se evaluó tanto un predictor único como un conjunto de ocho modelos.

El conjunto de datos de referencia del VPH consta de 194 mediciones de afinidad en siete alelos. Los péptidos derivados de las proteínas E6 y E7 del VPH16 se analizaron mediante un ensayo de unión competitiva basado en células. La precisión se evaluó en este punto de referencia utilizando tres métricas bien conocidas: área bajo la curva característica de funcionamiento del receptor (AUC), F1 y coeficiente de correlación de rango de Kendall (tau de Kendall). La puntuación AUC estima la probabilidad de que un péptido de unión fuerte (afinidad medida de 500 nM -nanomolar- o menos) tenga una afinidad prevista más fuerte que un péptido de unión débil o no vinculante (afinidad medida superior a 500 nM). Nanomolar es una unidad de concentración molar que se utiliza comúnmente en biología y química. Un nanomolar es una mil millonésima parte de un mol. La puntuación F1 resume la precisión y la recuperación de un predictor al clasificar los péptidos según tengan una afinidad menor o mayor a 500 nM . La puntuación tau de Kendall mide la correlación en rango cuando los péptidos se ordenan por afinidad medida o prevista; no utiliza ningún punto de corte y evalúa la concordancia en todo el espectro de afinidad [44].

7.2.1 *NetMHC-4.0*

NetMHC-4.0 es una herramienta bioinformática que se utiliza para predecir la unión de péptidos a moléculas de clase I del complejo mayor de histocompatibilidad (MHC).

NetMHC-4.0 utiliza un algoritmo de aprendizaje automático para predecir la afinidad de unión de un péptido a una molécula MHC-I. El algoritmo se ha entrenado en un conjunto de datos de péptidos conocidos por unirse a moléculas MHC-I. Utiliza redes neuronales artificiales (ANNs, por su sigla en inglés) para predecir la unión de péptidos a moléculas MHC-I. Las ANNs han sido entrenadas para 81 alelos HLA diferentes, incluyendo HLA-A, -B, -C y -E. También están disponibles predicciones para 41 alelos de animales (monos, ganado, cerdos y ratones). Se pueden hacer predicciones para péptidos de cualquier longitud. Sin embargo, hay que tener en cuenta que la mayoría de las moléculas HLA tienen una fuerte preferencia por la unión de péptidos de 9 aminoácidos.

NetMHC-4.0 es una herramienta ampliamente utilizada en investigación y desarrollo. Se utiliza para el desarrollo de vacunas, el diagnóstico de enfermedades y la investigación básica sobre el sistema inmune [42].

7.2.2 *NetMHCpan-4.1*

NetMHCpan-4.1 es una herramienta bioinformática que se utiliza para predecir la unión de péptidos a moléculas de clase I del complejo mayor de histocompatibilidad (MHC).

NetMHCpan-4.1 utiliza un algoritmo de aprendizaje automático para predecir la afinidad de unión de un péptido a una molécula MHC-I. El algoritmo utiliza redes neuronales artificiales (ANNs) para predecir la unión de péptidos a moléculas MHC-I de clase I de cualquier alelo de secuencia conocida. El método está entrenado en una combinación de más de 180.000 datos de afinidad de unión y ligandos MHC eluidos por MS. Los datos de afinidad de unión cubren 172 moléculas MHC de humanos (HLA-A, B, C, E), ratones (H-2), ganado (BoLA), primates (Patr, Mamu, Gogo) y cerdos (SLA). Los datos de ligandos MHC eluidos por MS cubren 55 alelos HLA y de ratón. Además, el usuario puede obtener predicciones para cualquier molécula MHC-I de clase I personalizada cargando una secuencia de proteína MHC de longitud completa.

NetMHCpan-4.1 es una herramienta ampliamente utilizada en investigación y desarrollo. Se utiliza para el desarrollo de vacunas, el diagnóstico de enfermedades y la investigación básica sobre el sistema inmune [43].

7.2.3 MHCflurry-3.1

MHCflurry-3.1 es una herramienta bioinformática que se utiliza para predecir la unión de péptidos a moléculas de clase I del complejo mayor de histocompatibilidad (MHC). Está implementada en Python y utiliza la biblioteca Keras. MHCflurry se entrena en un gran conjunto de datos de mediciones de afinidad de unión de péptidos, que incluye datos cuantitativos y cualitativos. La arquitectura del modelo consta de dos capas locales conectadas, una capa completamente conectada y una salida sigmoidea.

El conjunto de datos de entrenamiento para el modelo MHCflurry se construyó a partir de una instantánea de la base de datos IEDB de ligandos de MHC descargada el 17 de mayo de 2017, aumentada con el conjunto de datos BD2013. Se eliminaron las entradas de IEDB con alelos de clase I no específicos, mutantes o no analizables (*unparseable*), así como aquellas con péptidos que contenían modificaciones postraduccionales o aminoácidos no canónicos. Esto dio lugar a un conjunto de datos de IEDB de 147.716 mediciones de afinidad cuantitativas y 43.704 cualitativas. Se asignaron afinidades nanomolares a las mediciones cualitativas de la siguiente manera: alta positiva, 50; intermedia positiva, 500; baja positiva, 5000; positiva, 100; negativa, 50000.

De las 179.692 mediciones del conjunto de datos BD2013 publicado en la referencia, 55.473 no estaban también presentes en el conjunto de datos IEDB. Tras seleccionar péptidos de longitud 8-15 y descartar alelos con menos de 200 mediciones, el conjunto de datos de entrenamiento combinado consta de 235.597 mediciones en 101 alelos.

MHCflurry-3.1 utiliza un algoritmo de aprendizaje automático para predecir la afinidad de unión de un péptido a una molécula MHC-I. El algoritmo se ha entrenado en un conjunto de datos de péptidos conocidos por unirse a moléculas MHC-I.

MHCflurry-3.1 es una herramienta ampliamente utilizada en investigación y desarrollo. Se utiliza para el desarrollo de vacunas, el diagnóstico de enfermedades y la investigación básica sobre el sistema inmune [44].

7.2.4 NetMHCstab-1.0

NetMHCstab es un método de predicción de la estabilidad de complejos peptídicos-MHC de clase I. Utiliza un modelo estadístico entrenado con datos de espectrometría de masas para predecir la probabilidad de que un péptido se disocie de la molécula de MHC. NetMHCstab es una herramienta útil para el diseño de vacunas peptídicas, ya que puede ayudar a seleccionar péptidos que sean estables en el complejo MHC y, por lo tanto, tengan más probabilidades de ser reconocidos por los linfocitos T. El método se entrena con más de 25.000 datos de estabilidad cuantitativos que cubren 75 moléculas HLA diferentes. El usuario puede cargar secuencias completas de proteínas MHC y

hacer que el servidor prediga péptidos restringidos por MHC a partir de cualquier proteína de interés. Los valores de predicción se dan en valores de semivida en horas y como %-Rango en un conjunto de 200.000 péptidos naturales aleatorios.

NetMHCstab 1.0 está disponible como una herramienta web gratuita en el sitio web del Barcelona Supercomputing Center (BSC). También está disponible como un paquete de software que se puede instalar en un ordenador local.

Para utilizar NetMHCstab 1.0, los usuarios deben proporcionar la secuencia del péptido y el alelo MHC del que desean predecir la estabilidad. NetMHCstab calculará entonces la probabilidad de que el péptido se disocie de la molécula de MHC. Un valor de probabilidad más alto indica una mayor estabilidad.

NetMHCstab 1.0 es una herramienta poderosa para la predicción de la estabilidad de complejos peptídicos-MHC de clase I. Se utiliza ampliamente en la investigación inmunológica y el desarrollo de vacunas peptídicas [45].

7.2.5 MHCnuggets-1.0

MHCnuggets puede predecir la unión para alelos comunes o raros de MHC de clase I o II con una sola arquitectura de red neuronal. MHCnuggets también es más rápido que otros métodos y produce un aumento de cuatro veces en el valor predictivo positivo en datos independientes.

Utiliza dos tipos de funciones de pérdida (MSE y BCE) dependiendo de si los datos son de afinidad continua o binarios. Se utiliza la pérdida de error cuadrático medio (MSE, por su sigla en inglés) para entrenar redes con datos de afinidad continua. Para datos binarios de HLAp, se emplea la pérdida de entropía cruzada binaria (BCE, por su sigla en inglés). Se utiliza la técnica de retropropagación (backpropagation) con el optimizador Adam. La tasa de aprendizaje se fija en 0.001.

La arquitectura MHCnuggets puede manejar péptidos de cualquier longitud. En la práctica, se elige una longitud máxima (15 para clase I y 30 para clase II). Los péptidos más cortos se rellenan al final con el carácter "Z", que no está en el alfabeto de aminoácidos. Se utilizan datos del IEDB 2018 con medidas de afinidad química para 241,553 pares péptido-alelo en clase I y 96,211 pares en clase II. Se incluyen datos de 16 inmunopeptidomas de líneas celulares B monoalélicas clase I.

Los autores aplicaron MHCnuggets a 26 tipos de cáncer en el Atlas del Genoma del Cáncer y encontraron que las mutaciones *missense* inmunogénicas predichas (IMM, por su sigla en inglés) estaban significativamente asociadas con un aumento de la infiltración de células inmunitarias, incluidas las células T CD8+. Solo el 0,16 % de las IMM predichas se observaron en más de 2 pacientes, y el 61,7 % de ellas se derivaron de mutaciones impulsoras [46].

7.2.6 *Pipeline* SARS-CoV-2

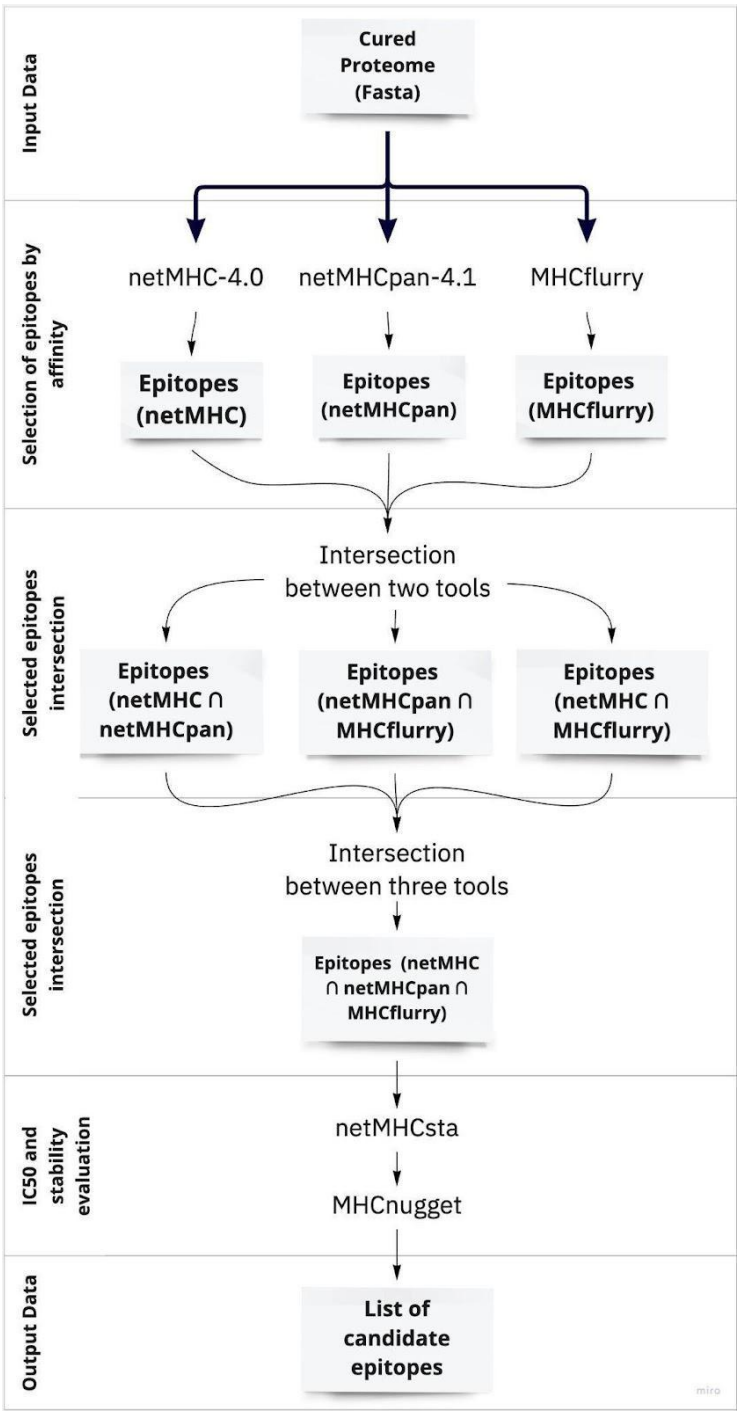


Figura 6. Representación del *pipeline* SARS-CoV-2 Clase I: Representa paso a paso del pipeline bioinformático de insumos, formatos y resultados

El pipeline diseñado se ejecuta de manera secuencial en cuatro niveles distintos, como se ilustra en la Figura 6. En el primer nivel, se emplean las herramientas NetMHC-4.0, NetMHCpan-4.1 y MHCflurry [42-44] para seleccionar los péptidos dentro del proteoma completo del virus, que se proporciona como entrada en formato FASTA. Esta selección se realiza con base en la mejor afinidad reportada por cada una de las herramientas.

Para algunos casos, debido a la gran cantidad de péptidos seleccionados y las limitaciones de memoria de los servidores, fue necesario implementar un *pipeline* adicional que separa y agrupa los péptidos según su longitud de 8, 9, 10, 11 y 12. Esto se hizo con el propósito de ajustarse a las restricciones de memoria requeridas por las herramientas bioinformáticas utilizadas en el análisis.

Al realizar esta división y agrupación de péptidos, se logró mejorar la eficiencia y el rendimiento del análisis, permitiendo un mejor manejo de los recursos computacionales disponibles. Esta estrategia ha sido de gran utilidad para superar las limitaciones técnicas y avanzar en la identificación y caracterización de los péptidos seleccionados en el estudio.

La adaptación del *pipeline* en función de las especificidades de cada conjunto de péptidos ha sido crucial para garantizar resultados precisos y confiables en la investigación, optimizando así el uso de las herramientas bioinformáticas y permitiendo el análisis de grandes volúmenes de datos de manera eficiente.

En el segundo nivel, se reciben las listas de péptidos predichos en formato de texto plano y se generan archivos de salida adicionales mediante el análisis de las listas de péptidos por pares de herramientas, es decir, (NetMHC-4.0 ∩ NetMHCpan-4.1), (NetMHC-4.0 ∩ MHCflurry) y (NetMHCpan-4.1 ∩ MHCflurry). Estos archivos de salida proporcionan información adicional sobre los péptidos identificados mediante la combinación de las diferentes herramientas.

En el tercer nivel, se reciben estas listas de texto plano y se genera un nuevo archivo de salida que contiene la lista de epítopos seleccionados por las tres herramientas combinadas (NetMHC-4.0 ∩ NetMHCpan-4.1 ∩ MHCflurry). Esto permite identificar los

epítopos que han sido considerados de alta afinidad por todas las herramientas utilizadas en el análisis.

Posteriormente, en el cuarto nivel, la lista de epítopos seleccionados por las tres herramientas se evalúa utilizando dos herramientas adicionales: netMHCstab y MHCnuggets [45-46]. Estas herramientas generan el archivo de salida final, que contiene valores de estabilidad y IC50, así como otros parámetros asignados por cada herramienta a cada epítopo. Estos valores y parámetros resultan fundamentales para la selección final de los epítopos candidatos a vacunas.

En este trabajo, se adoptó un enfoque riguroso y sistemático para la selección de epítopos candidatos a vacunas en el contexto del SARS-CoV-2. Este enfoque secuencial y estructurado ha permitido obtener resultados confiables y precisos en la identificación de las mejores opciones para la generación de una vacuna efectiva contra este virus.

7.2.7 Interfaz de usuario

Para desarrollar la interfaz de los *pipelines*, se seleccionó el lenguaje de programación Python con el Framework Django debido a su amplia disponibilidad de bibliotecas y su facilidad de uso.

Django es un marco de trabajo de desarrollo web de alto nivel que se utiliza para crear aplicaciones web en Python. Django fue diseñado para facilitar el desarrollo rápido y limpio de aplicaciones web, y se enfoca en la reutilización del código y la eliminación de la redundancia. Django es gratuito y de código abierto, y cuenta con una amplia comunidad de desarrolladores que contribuyen a su desarrollo y mantenimiento. Django proporciona una amplia variedad de herramientas y características para el desarrollo web, incluyendo un sistema de administración de bases de datos, un sistema de autenticación, un sistema de plantillas, un sistema de enrutamiento, y mucho más. Django también es altamente escalable y se puede utilizar para crear aplicaciones web complejas y escalables [84].

También, se aprovechó el sistema de gestión de paquetes CONDA para la instalación de las herramientas bioinformáticas.

Conda es un sistema de gestión de paquetes, entornos y dependencias que se utiliza principalmente en el ámbito de la ciencia de datos, aprendizaje automático y desarrollo de software. Conda permite instalar, actualizar y desinstalar paquetes de software. Los paquetes pueden contener bibliotecas, herramientas y otros recursos necesarios para diversas tareas. Conda facilita la creación y gestión de entornos virtuales. Estos entornos son espacios aislados donde puedes instalar y ejecutar software específico sin interferencias con otros proyectos. Esto es útil para evitar conflictos de dependencias [85].

Además, se optó por utilizar el sistema de contenedores Docker para generar un sistema multiplataforma, lo que permite una mayor flexibilidad y portabilidad del *pipeline* en diferentes entornos.

Docker es una plataforma de software que permite a los desarrolladores crear, compartir y ejecutar aplicaciones en contenedores. Los contenedores son una forma de virtualización que permite a las aplicaciones ejecutarse en un entorno aislado y portátil. Docker es una herramienta popular para la creación y gestión de contenedores, lo que facilita el desarrollo y la implementación de aplicaciones en diferentes entornos. Docker proporciona una *suite* de herramientas de desarrollo, servicios, contenido confiable y automatizaciones, que se utilizan individualmente o juntos, para acelerar la entrega de aplicaciones seguras [86].

Inicialmente, se configuró un contenedor Docker con el sistema operativo Linux Ubuntu 20 LTS, que proporciona un entorno estable y confiable para el desarrollo y ejecución del *pipeline*. Dentro del contenedor, se instalaron y configuraron las herramientas seleccionadas, las cuales abarcan una variedad de enfoques computacionales, incluyendo métodos basados en aprendizaje automático e inteligencia artificial. Estas herramientas son esenciales para el análisis de los datos y la generación de resultados relevantes.

Una vez validado el correcto funcionamiento del *pipeline* inicial, se procedió a desarrollar una interfaz web utilizando el framework de Python Django. Esta interfaz web tiene como objetivo simplificar la ejecución del *pipeline* y brindar una experiencia más amigable para

los investigadores. A través de la interfaz web, los usuarios pueden cargar fácilmente sus archivos de entrada y realizar consultas sobre el estado del proceso. Además, se implementó un sistema de gestión de usuarios y accesos controlados, lo que garantiza que solo los usuarios autorizados puedan utilizar el *pipeline*.

La interfaz web desarrollada en Django proporciona una forma intuitiva de interactuar con el *pipeline* y ofrece funcionalidades adicionales, como la visualización de los resultados generados y la posibilidad de programar ejecuciones automáticas. Esto agiliza el proceso de análisis y permite a los investigadores obtener rápidamente los resultados deseados. Estas mejoras contribuyen a la accesibilidad y usabilidad del *pipeline*, brindando una herramienta eficaz y práctica para el análisis de datos en el ámbito de la investigación científica.

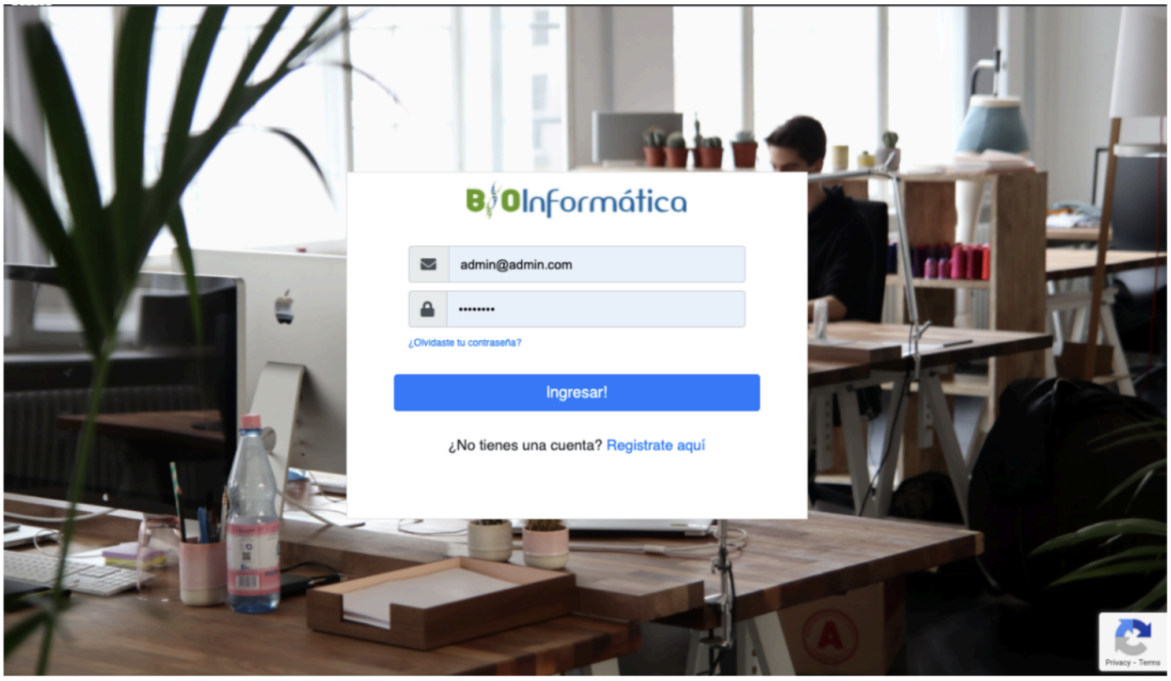


Figura 7. Panel de ingreso: control de ingreso a la plataforma Usuario/Contraseña.

En esta sección se controla el acceso de los usuarios a la plataforma mediante un usuario y contraseña, junto con un sistema anti-bots.

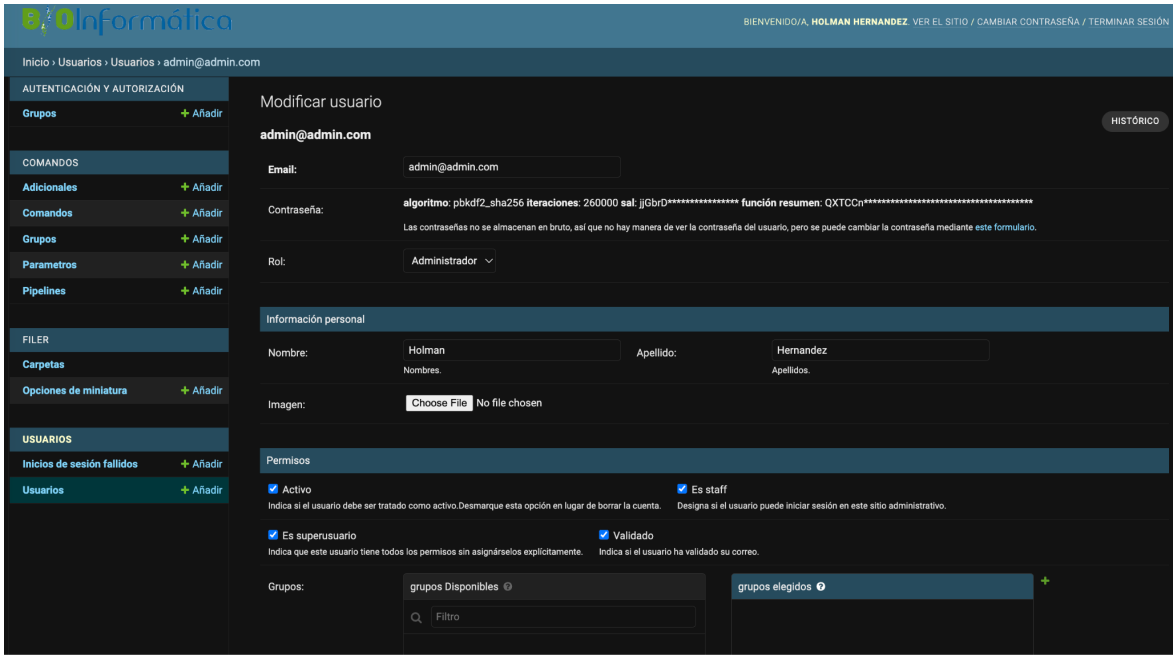


Figura 8. Administrador de usuarios Django y Gestión de permisos: Interfaz para el registro de usuarios nuevos de la plataforma y Permisos de usuarios de la plataforma Eje. Subir archivos, ejecutar pipelines, etc.

En esta sección se permite la creación de nuevos usuarios con los campos de email, contraseña, rol, nombres, apellidos, imagen y permisos específicos de acceso. Y la configuración de grupos de permisos sobre tablas de manera controlada.

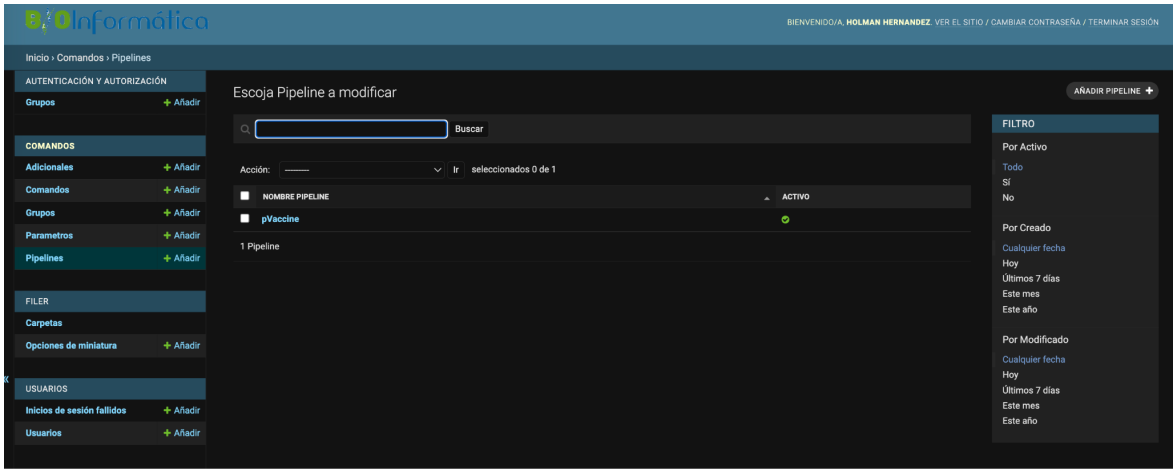


Figura 9. Administrador de generación de pipelines: Interfaz que permite al perfil de bioinformático crear los pipelines que verán los usuarios finales.

Esta sección es especialmente para los bioinformáticos, donde se tienen múltiples niveles para crear el *pipeline*, las secciones del *pipeline*, los comandos específicos y la asociación de variables de configuración de los comandos.

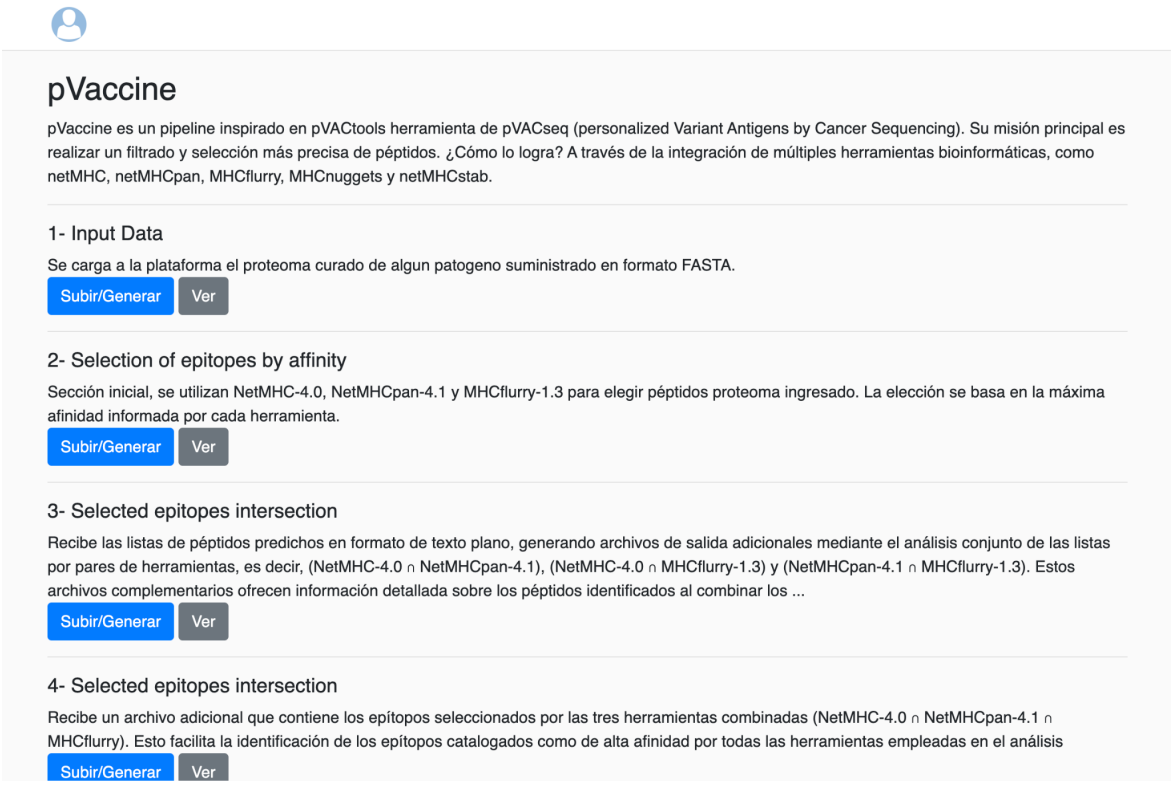


Figura 10. Interfaz del *pipeline* general: Permite a los usuarios finales ver los pasos del pipeline programado por el bioinformático con su descripción general.

En esta sección se ve la interfaz de los *pipelines* creados por los bioinformáticos y permite ser ejecutado con una descripción de cada paso del pipeline.

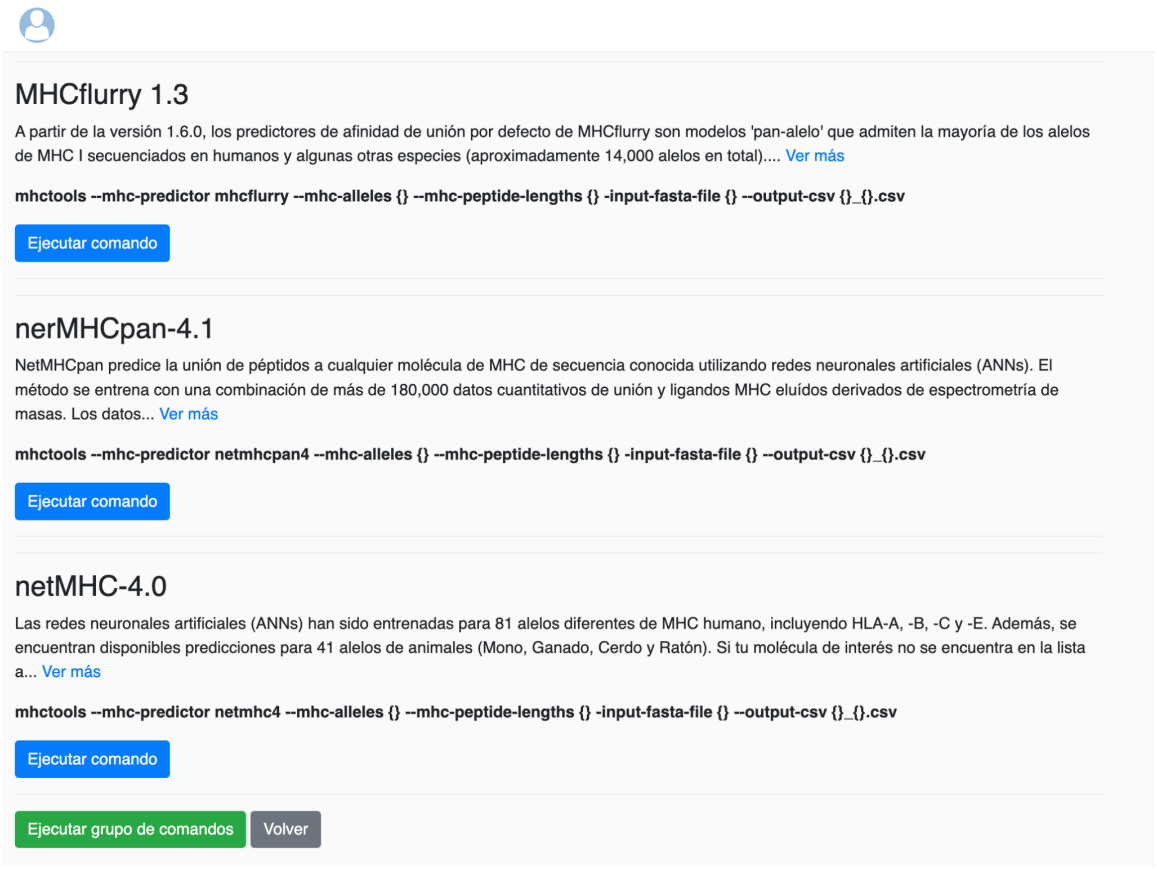


Figura 11. Interfaz del *pipeline* específica: Al ingresar a cada paso del *pipeline* permite ver el comando generar y modificar parámetros en caso de ser necesario.

Esta sección especifica muestra los comandos de cada sección del *pipeline*, que se pueden ejecutar o configurar de ser necesario y estar habilitado por parte del programador del *pipeline*.

Una vez completada la configuración del *pipeline* para establecer la plataforma de predicción de epítopos *in silico* en el modelo de SARS-CoV-2, se realizaron ajustes al *script* con el objetivo de mejorar la organización y la presentación de los resultados. En particular, se implementó un proceso para agrupar los datos de salida en tablas que asocian los conjuntos compartidos de péptidos entre las diferentes herramientas utilizadas en el análisis.

El proceso de agrupamiento se llevó a cabo mediante una clasificación de los péptidos de unión fuerte (SB: Strong Binders) generados por cada herramienta. Esta clasificación permitió identificar los epítopos con mayor afinidad y potencial inmunogénico. Posteriormente, se realizó una separación de las predicciones de cada herramienta, lo que facilitó la visualización y el análisis individualizado de los resultados.

Una vez completada esta etapa, se obtuvo una lista de péptidos de cierta longitud, específicamente epítopos de 9, 10 y 11, que fueron considerados de interés en el contexto del estudio. Para complementar esta información, se aplicó la herramienta NetMHCstabpan [47], que permite predecir el tiempo de estabilidad media del complejo péptido-MHC. Esta información resulta relevante para evaluar la viabilidad y la duración de la interacción entre los epítopos y las moléculas MHC. Estos procesos adicionales permitieron obtener una visión más completa y detallada de los epítopos de interés en el modelo de SARS-CoV-2.

7.3 Alelos más prevalentes en la población colombiana

Con el fin de focalizar el pipeline en una población en concreto, se realizó una búsqueda sistemática de los alelos HLA más prevalentes en la población colombiana. El listado de alelos presentados en la Tabla 1 (ver además el archivo Suplementaria 2), corresponde a aquellos más prevalentes en la población Colombiana de acuerdo a un rango entre 0 y 1 (frecuencia alélica), donde 1 representa un alelo presente en el 100% de la población. Así fue posible adaptar las predicciones de epítopos a la diversidad genética de la población colombiana como población objetivo en este trabajo. Esta consideración es crucial para garantizar que las vacunas diseñadas sean efectivas y cubran las variantes genéticas relevantes en la población de interés (ver Tabla 1).

Tabla 1

Allele HLA-I	Allelic Frequency
A*01:01	0.061
A*02:01	0.161
A*03:01	0.061
A*24:02	0.208
A*29:02	0.045
A*68:01	0.052
B*07:02	0.050
B*35:01	0.045
B*35:43	0.086
B*40:02	0.084
B*44:03	0.056
B*51:01	0.056
C*01:02	0.114
C*03:04	0.082
C*04:01	0.149
C*05:01	0.051
C*06:02	0.051
C*07:01	0.089
C*07:02	0.097
C*16:01	0.051

Se tomo como punto de corte que los alelos seleccionados tuvieran una representacion mayor al 4% de la poblacion colombiana. Entre los que podemos destacar los Alelos A*02:01, A*24:02, C*01:02 y C*04:01 como los mas prevalentes en la poblacion.

7.4 pVaccine

Aquí se presenta un enfoque computacional integrativo de algoritmos bioinformáticos orientados a predecir epítopos a partir del proteoma de SARS-CoV-2 útiles para el diseño de vacunas para la población colombiana. Este enfoque se basa en la modificación y personalización de pVacTools, reconocida por su capacidad para el diseño de vacunas personalizadas contra el cáncer basadas en neoantígenos tumorales [7]. Mediante la integración de diversas herramientas bioinformáticas, fue diseñado un *pipeline* que permite predecir epítopos MHCI candidatos a vacuna útil para la población colombiana partiendo del proteoma de SARS-CoV-2.

El enfoque propuesto es adecuado para predecir epítopos de células T restringidos por un conjunto específico de alelos HLA altamente expresados por la población colombiana, ejercicio que resulta valioso para el diseño de candidatos a vacunas contra otros patógenos. Para seleccionar las herramientas bioinformáticas en el *pipeline*, se realizaron rigurosas evaluaciones de comparaciones (*benchmarks*) no solo considerando las tasas de falsos descubrimientos (FDR) [35], sino también teniendo en cuenta las tasas de aciertos y los algoritmos utilizados en cada herramienta. Además, se consideró cuidadosamente la disponibilidad y relevancia de cada herramienta en el campo de la bioinformática y la inmunología, asegurándose de utilizar aquellas que cuentan con un sólido respaldo científico y una amplia comunidad de usuarios. Asimismo, se evaluó el tiempo de la última versión de cada herramienta para garantizar que se empleen las versiones más actualizadas, lo que es esencial para obtener resultados precisos y confiables en el análisis.

Para validar el rendimiento del *pipeline*, se evaluó su capacidad para identificar epítopos inmunogénicos de SARS-CoV-2 previamente identificados *ex-vivo* o *in-vitro* por otros investigadores [72-83]. Inicialmente, se realizaron varias ejecuciones del *pipeline* desde el proteoma de referencia del virus SARS-CoV-2 utilizando el alelo HLA: A02:01 como

modelo del MHC y para un primer acercamiento como prueba de concepto. Los resultados, obtenidos tras ejecutar el pipeline en cinco ocasiones bajo los mismos parámetros, demostraron una alta reproducibilidad y consistencia. En cada una de las cinco ejecuciones, se generó consistentemente una lista de péptidos relevantes (Ver Tabla Suplementaria 1 - pVacTools), lo que refuerza la robustez y confiabilidad de los hallazgos. La repetición exitosa del análisis en múltiples ocasiones bajo condiciones idénticas proporciona una sólida evidencia de la estabilidad y coherencia de los resultados obtenidos, lo que valida aún más la efectividad y precisión del enfoque bioinformático propuesto.

El proceso de configuración del *pipeline* incluyó ajustes en el *script* para agrupar los datos de salida en tablas que contienen conjuntos compartidos de péptidos entre las diferentes herramientas. Se realizó un proceso de ordenamiento o clasificación computacional de los péptidos de unión fuerte con cada herramienta (SB: Strong Binders), seguido de un proceso de separación (*split*) para obtener listas de péptidos con longitudes requeridas por la herramienta netstabpan, que predice el tiempo medio de estabilidad del complejo MHC-péptido. La optimización del flujo de trabajo (*workflow*) permitió obtener péptidos de SARS-CoV-2 con información relevante como IC50, vida media, estabilidad y probabilidad de procesamiento proteasomal (mhcflurry). Esta configuración permitió generar tablas comparativas con los péptidos potencialmente inmunogénicos seleccionados con cada herramienta, así como los conjuntos o intersecciones de péptidos con predicciones coincidentes entre varias herramientas bioinformáticas.

Es importante destacar que, aunque los modelos computacionales son herramientas valiosas para predecir epítopos inmunogénicos, es necesario realizar experimentos adicionales para validar y confirmar la inmunogenicidad de los péptidos seleccionados como ensayos de unión MHC-péptido, estimulación de células T, citotoxicidad mediada por células T, respuesta de anticuerpos, etc.

El *pipeline* ha sido adaptado específicamente para la población colombiana y muestra resultados prometedores en la selección de epítopos potencialmente inmunogénicos. En primer lugar, se basa en la plataforma pVacSeq/pVacTools, que ha demostrado ser

altamente efectiva en el diseño de vacunas personalizadas contra el cáncer basadas en neoantígenos tumorales [7], Al adaptar esta plataforma para predecir epítopos MHCI específicos para la población colombiana, hemos aprovechado su robustez y capacidad de generar resultados precisos.

En segundo lugar, el *pipeline* diseñado integra múltiples herramientas bioinformáticas ampliamente reconocidas, seleccionadas mediante evaluaciones de comparaciones y tasas de falsos descubrimientos. Esto asegura que se están utilizando las herramientas más confiables y validadas en el campo de la predicción de epítopos [35].

El rendimiento del *pipeline* se ha validado mediante la comparación de los epítopos predichos con aquellos previamente identificados como inmunogénicos por otros investigadores. Esta validación proporciona confianza en la capacidad del enfoque para identificar de manera precisa y confiable los epítopos relevantes para el diseño de vacunas.

Inicialmente para la prueba de concepto se realizó una comparación entre dos *pipelines* bioinformáticos, pVaccine (el desarrollado en este trabajo) y pVACtools (nuestro referente [7], para evaluar lado lado su rendimiento en la predicción de epítopos altamente inmunogénicos en la proteína SP del patógeno SARS-CoV-2. Se utilizaron diferentes algoritmos (NetMHCpan, NetMHC y MHCflurry) en ambos *pipelines* para predecir epítopos HLA-A*02:01 restringidos en SP.

Los resultados mostraron que pVaccine tuvo un mejor rendimiento en la predicción de un mayor número de epítopos HLA-A*02:01 en SP en comparación con pVACtools. Además, muchos de los epítopos detectados por pVaccine también fueron identificados en el proteoma completo, lo que resalta la eficacia y versatilidad de esta herramienta.

Es importante anotar que para comparar este desempeño, los algoritmos NetMHCpan, NetMHC y MHCflurry se utilizaron de manera individual en cada *pipeline*. Los resultados presentados (ver Figura 12, Parte Inferior) muestran que mientras que pVaccine seleccionó en el proteoma completo 424; 341 y 1340 epítopos HLA-A*02:01, respectivamente, pVacTools seleccionó sólo 17 epítopos (Figura 12, Parte Superior).

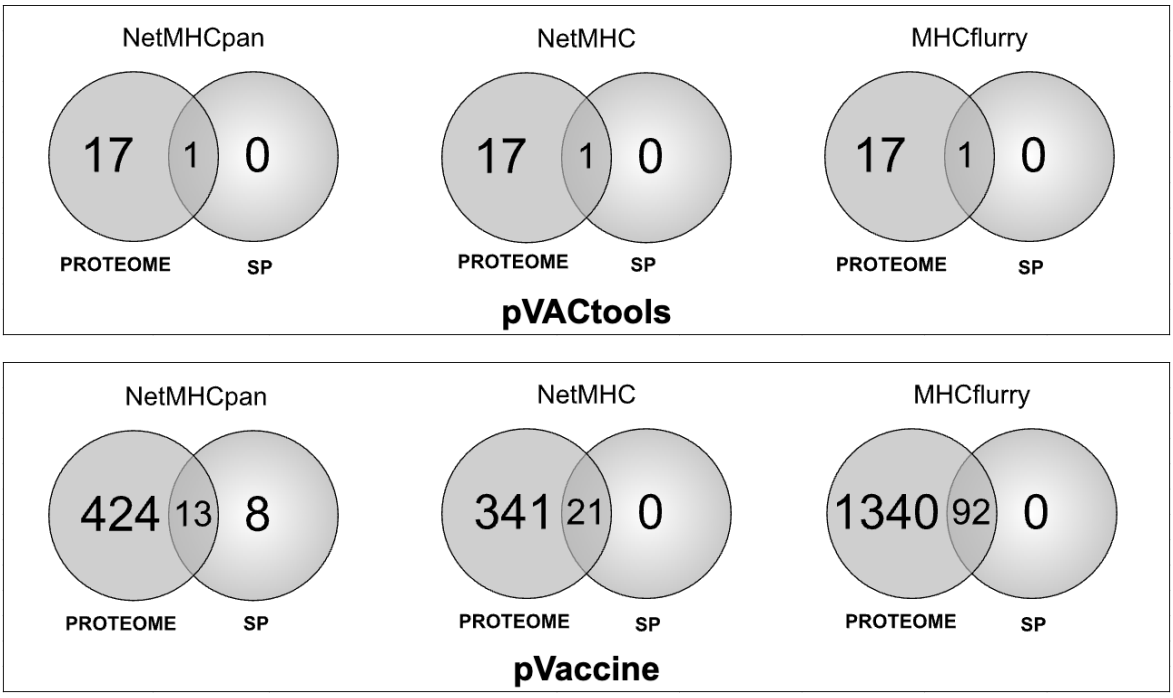


Figura 12. Universos de péptidos seleccionados por pVACtools para SARS-CoV-2 con HLA-A*02:01 en proteoma completo y Spike (panel superior). Universos de péptidos seleccionados por pVaccine para SARS-CoV-2 con HLA-A*02:01 en proteoma completo y Spike (panel inferior)

Está limitada capacidad de pVACtools para seleccionar con cada algoritmo epítopos HLA-A*02:01 en el proteoma completo de SARS-CoV-2 fue aún más evidente cuando se comparó la capacidad de los dos pipelines para predecir epítopos HLA-A*02:01 de SP utilizando el proteoma en comparación con SP como entrada. Notablemente, este análisis reveló que pVACtools utilizando los algoritmos NetMHCpan, NetMHC y MHCflurry detectó el unico epitopo en SP (el epítopo YLQPRTFLL) la cual también fue

identificada tanto en el proteoma como en SP por pVACtools utilizando los tres algoritmos.

Un análisis detallado reveló que pVACtools sólo pudo detectar un único epítopo en SP, mientras que pVaccine identificó 21, 21 y 92 epítopos en diferentes regiones de la proteína. Estos resultados respaldan la superioridad de pVaccine en la selección de epítopos relevantes y su capacidad para identificarlos tanto en la proteína SP como en el proteoma completo.

Además, se llevó a cabo una comparación exhaustiva de los resultados de ambas herramientas y se encontraron 61 péptidos compartidos entre pVaccine y pVACtools. Tras aplicar un filtro, se seleccionaron 26 epítopos que fueron validados por al menos un autor (Ver Tabla 2). En esta tabla, los péptidos seleccionados por el pipeline pVaccine utilizando el proteoma completo: Contiene los péptidos seleccionados y filtrados, la proteína a la que pertenece cada péptido y un mapa de calor representando la afinidad relativa de unión y tiempo medio de la unión. Los autores, herramientas y péptidos encontrados por pVaccine y pVACtools se encuentran en la tabla amplia de los datos suplementarios (ver Tabla 2 Suplementaria).

Tabla 2.

Peptido	*Proteína*	*netMHC*	*netMHCpan*	*mhcflurry*	*netMHCstabpan (Horas)*
LLLDRLNQL	**N**	**14,81**	**1,64**	**21,10**	**4,05**
LLYDANYFL	**ORF3a**	**3,06**	**2,80**	**10,66**	**12,45**
YLQPRTFLL	**Spike**	**5,36**	**28,65**	**12,30**	**8,04**
RLNEVAKNL	**Spike**	**940,92**	**300,00**	**109,47**	**1,27**
YLYALVYFL	**ORF3a**	**2,67**	**19,55**	**11,05**	**19,08**
ALSKGVHFV	**ORF3a**	**7,25**	**0,11**	**11,50**	**15,56**
FIAGLIAIV	**Spike**	**10,29**	**1,21**	**16,06**	**5,11**
FLLPSLATV	**ORF1ab**	**2,81**	**9,44**	**10,24**	**32,26**
GMSRIGMEV	**NPC-10H ORF**	**50,61**	**93,83**	**36,20**	**1,78**
ILFTRFFYV	**ORF1ab**	**3,15**	**0,26**	**11,91**	**13,07**
VLNDILSRL	**Spike**	**33,57**	**37,00**	**19,49**	**10,69**
VVFLHVTYV	**Spike**	**36,56**	**57,39**	**19,92**	**5,94**
ALNTLVKQL	**Spike**	**1.633,23**	**1,13**	**139,26**	**1,33**
ALWEIQQVV	**ORF1ab**	**7,85**	**1,49**	**19,26**	**11,35**
FGDDTVIEV	**ORF1ab**	**328,47**	**255,00**	**90,94**	**0,58**
KIADYNYKL	**Spike**	**36,12**	**94,36**	**16,76**	**3,42**
KLDDKDPNF	**NPC-10H ORF**	**1.512,02**	**45,25**	**316,93**	**0,59**
KLNDLCFTNV	**Spike**	**15,27**	**34,68**	**34,93**	**17,27**
KLPDDFTGCV	**Spike**	**77,11**	**23,85**	**78,02**	**3,21**
LLFNKVTLA	**Spike**	**202,00**	**8,06**	**22,43**	**5,61**
MIAQYTSAL	**Spike**	**452,98**	**6,29**	**29,50**	**1,51**
NLIDSYFVV	**ORF1ab**	**5,93**	**37,07**	**13,76**	**4,74**
RLQSLQTYV	**Spike**	**16,66**	**62,35**	**19,11**	**11,16**
SIIAYTMSL	**Spike**	**207,35**	**72,49**	**23,56**	**1,96**
YLATALLTL	**ORF1ab**	**5,07**	**126,22**	**11,80**	**9,01**
YLDAYNMMI	**ORF1ab**	**14,77**	**29,23**	**21,29**	**11,00**

Con el fin de validar la importancia inmunológica de los péptidos seleccionados mediante la herramiento pVaccine, se llevó a cabo una búsqueda sistemática y consolidación de información reportada en la literatura en la que hubiese información acerca del reconocimiento inmunológico por el sistema inmune de pacientes infectados con SAESpCoV-2, de las epítipes seleccionadas mediante pVaccines, lo cual permitiese validad la selección de péptidos candidatos para la vacuna contra el virus SARS-CoV-2

seleccionados con la herramienta. Dicha selección fue realizada analizando las publicaciones y literatura científica de diferentes autores, cuya estrategia se basó principalmente en enfoques de predicción de péptidos de SARS-CoV-2 desde el inicio de la pandemia en marzo de 2020 hasta agosto de 2021 (ver Tabla 3). En esta tabla se resume el número de péptidos comunes reportados por otros utilizando análisis inmunológicos y distintos algoritmos predictivos.

Tabla 3

Autor	Año	Número de péptidos reportados
Daouda	2021	48
Grifoni	2020	573
Liu	2020	295
Sohail	2020	379
Campbell	2020	379
Fren	2020	810
Mishra	2020	399
Nathan	2021	208

Con el propósito de contrastar y validar los resultados obtenidos por estos autores, se compararon los péptidos seleccionados con los obtenidos mediante el *pipeline* bioinformático pVaccines centrando este análisis en los péptidos que se unen a los alelos más prevalentes en la población colombiana (seleccionados de la base de datos Allele Frequency Net (http://www.allelefrequencies.net/)). Para facilitar este proceso, se desarrolló un *script* adicional que contrastaba la información de las bases de los péptidos seleccionados de la literatura con nuestras predicciones. El *script* se encarga de comparar los archivos de texto plano que contienen los péptidos reportados por los otros autores con los péptidos predichos por las herramientas utilizadas por pVaccines en nuestro trabajo. Esta comparación permitió validar la coincidencia y la concordancia entre los péptidos reportados por otros autores y los predichos *in-silico* utilizando pVaccines, lo que es esencial para corroborar la precisión y confiabilidad de las predicciones realizadas por esta herramienta.

Este análisis evidenció que nuestro *pipeline* bioinformático generó un porcentaje de superposición de universos de péptidos predichos por otros grupos, empleando enfoques

predictivos e inmunológicos. Estos resultados son presentados en un documento de gráficos comparativos, donde se contrastaron los porcentajes de superposición de nuestra aplicación con los universos de péptidos reportados por los autores en la literatura científica (ver archivo Suplementaria 3) [64-71].

Además, se implementó una estrategia de inmunología inversa (IR) basada en una exhaustiva búsqueda de información reportada en la literatura. Esta estrategia permitió la selección y consolidación de un conjunto de péptidos candidatos para la vacuna del virus SARS-CoV-2, los cuales contaban con apoyo experimental por parte de diferentes autores (ver Tabla 4). Este apoyo experimental incluía técnicas como CBA, Elisa, INF EliSpot, espectrometría de masas en tándem, ensayos de competencia y ensayos de estabilidad con gradientes de temperatura, y se enfoca en las restricciones HLA de los péptidos. La tabla 4, señala el número de péptidos reportados como inmunogénicos en individuos humanos con SARS-CoV-2 por diferentes autores entre 2020 y 2021 como resultado del uso de diferentes tipos de ensayos inmunológicos y sustento experimental.

Tabla 4

Autor	Año	Número de péptidos reportados
Agerer	2021	9
Campbell	2020	9
Daouda	2021	54
Kared	2020	74
Mallajosy	2021	89
Nathan	2021	107
Peng	2020	5
Prachar	2020	174
Quadeer	2021	19
Schulien	2021	37
Sohail	2021	33
Weingarten Gabbay	2020	30

Nuevamente, se compararon estos péptidos con los resultados obtenidos mediante nuestro *pipeline* bioinformático, utilizando los 14 alelos más prevalentes en la población colombiana. Esta comparación permitió establecer la superposición de universos de péptidos compartidos, la cual se registra en el archivo Suplementaria 4 [72-83]. La

verificación de la superposición entre los universos de péptidos determinados por la estrategia de inmunología inversa y nuestro *pipeline* demostró la existencia, procesamiento y presentación de péptidos en SARS-COV-2 predichos computacionalmente obtenidos mediante pVaccines cuya inmunogenicidad había sido evidenciada experimentalmente por otros autores. Estas pruebas sistemáticas evidenciaron que los algoritmos predictivos han sido refinados para generar péptidos con inmunogenicidad comprobada, tal como se muestra en el archivo Suplementaria 5.

Además, se desarrolló un *script* para identificar el número de péptidos que informaron los autores con apoyo experimental y que también fueron determinados con nuestro *pipeline*; el *script* se encarga de comparar los archivos de texto plano que contienen los péptidos reportados con sustento experimental por los autores con los péptidos predichos por nuestra herramienta. A partir de esta información, se seleccionaron los universos de péptidos cuyas predicciones fueron encontradas por más de un autor (ver archivo Suplementaria 6) o por más de tres autores (ver archivo Suplementaria 7). Estos resultados fueron sometidos a un proceso de eliminación de duplicados y consolidación de todos los parámetros de las predicciones de las herramientas integradas en nuestro pipeline. De esta manera, obtuvimos los péptidos finales (ver Tabla 5) que contaban con apoyo experimental y poseían características *in-sílico*, *ex-vivo* e *in-vivo* necesarias para considerarse como candidatos a ser parte de una vacuna contra el SARS-CoV-2 para individuos que expresan los alelos más prevalentes de la población colombiana.

Tabla 5

Peptidos:	affinity_netMHC	ic50_netMHC	netMHCstabpan_netMHC	affinity_netMHCpan	ic50_netMHCpan	affinity_mhcflurry-predict-scan	processing_score_mhcflurry-predict-scan	Alelo
EEAIRHVRAW	0.69	28.7	0.6	0.94	27.01	55.1	0.68	B4403
EEIAIILASF	0.73	19.65	0.93	0.82	21.7	47.1	0.56	B4403
NYMPYFFTL	0.74	17.44	0.16	0.78	16.15	67.4	0.96	C0702
QEYADVFHLY	0.78	10.73	4.25	0.99	9.78	37.1	0.74	B4403
TMLFTMLRK	0.80	9.07	1.06	0.79	11.05	31.5	0.01	A0301
TSFGPLVRK	0.72	20.71	0.98	0.94	41.89	36.4	0.27	A0301
TTIKPVTYK	0.70	25.27	1.77	0.96	30.29	29.3	0.57	A0301
VLSGHNLAK	0.69	28.49	7.09	0.88	19.69	31.0	0.07	A0301
VTNNTFTLK	0.70	24.67	6	0.91	22.63	27.6	0.84	A0301
YTMADLVYA	0.82	7.34	5.9	0.45	6.42	20.9	0.20	A0201
YYTSNPTTF	0.77	11.58	21.72	0.99	14.03	28.4	0.76	A2402
YYTSNPTTF	0.73	18.94	0.16	0.89	23.75	33.6	0.76	C0702

Finalmente, con el fin de evaluar la versatilidad, adaptabilidad y utilidad potencial de pVaccines como herramienta para predecir epítopos útiles para diseño de vacunas contra otros patógenos para la población colombiana, se llevó a cabo un estudio utilizando el proteoma de referencia de *Mycobacterium tuberculosis (M.tb* H37Rv), un patógeno de gran relevancia debido a sus impactos en la salud pública a nivel mundial y su alta peligrosidad en poblaciones inmunocomprometidas [33].

Dado que existen notables diferencias en cuanto a la complejidad y tamaño entre los proteomas de referencia de SARS-CoV-2 y *M.tb,* fue necesario subdividir el proteoma de referencia de *M.tb* en 27 secciones distintas. Se realizaron ajustes en el *pipeline* bioinformático para garantizar que no se alterase la integración final de las predicciones de todas las herramientas ni los formatos de salida esperados. Con el objetivo de validar el rendimiento de nuestro *pipeline*, se realizó una prueba de concepto centrada en la predicción de epítopos inmunogénicos restringidos por HLA:02:01 dentro del proteoma de referencia de *M.tb,* empleando tres herramientas bioinformáticas predictivas integradas (NetMHC-4.0 ∩ NetMHCpan-4.1 ∩ MHCflurry).

En la fase de análisis final, se aplicó un proceso de filtrado exhaustivo para seleccionar aquellos péptidos con la mayor probabilidad global de procesamiento proteasomal (el

proteosoma es un complejo multimolecular que degrada proteínas y genera los péptidos que son presentados en moléculas MHCI), los mejores valores de afinidad y los valores más bajos de IC50, con el fin de consolidar la comparación de las predicciones. De esta manera, se obtuvieron los resultados de los péptidos candidatos generados por cada herramienta, así como aquellos compartidos por péptidos predichos con más de una herramienta bioinformática, explorando todas las posibles comparaciones.

Finalmente, tras combinar los resultados de las diferentes herramientas, se logró identificar dos epítopos restringidos por HLA-A* 02:01 en Mt: YVIGDDVEV (CopG family DNA-binding protein), con un IC50 de 38.04 (mhcflurry) y una estabilidad de 2.22 horas (predicha por netstabpan); y YVHSAPWSV (Ig-like domain-containing protein), con un IC50 de 15.69 y una estabilidad de 2.31 horas (netstabpan) (ver Tabla Suplementaria 8).

A pesar de que el proteoma de *Mycobacterium tuberculosis (M.tb H37Rv)* es muy grande en comparación con el del SARS-CoV-2 se obtuvo un número muy reducido de epítopes predichos entre las diferentes herramientas; esto puede obedecer a varios aspectos entre los que se encuentran la alta complejidad del genoma, que incluye secuencias altamente repetitivas y elementos reguladores difíciles de analizar mediante herramientas *in-silico*; La escasa representación en bases de datos, las bases de datos utilizadas para entrenar las herramientas de predicción *in-silico* pueden también tener una limitada representación de secuencias específicas de *Mycobacterium tuberculosis.* La alta diversidad de cepas: *Mycobacterium tuberculosis* presenta una alta diversidad genética entre diferentes cepas y regiones geográficas. Las herramientas *in-silico* pueden no ser lo suficientemente sensibles para capturar todas las variantes presentes en diferentes cepas, lo que puede conducir a una reducción en el número de péptidos identificados.

Si bien la identificación de epítopos a nivel experimental sigue siendo la prueba definitiva en términos de confiabilidad y precisión, el alto costo en términos de tiempo y recursos ha impulsado el desarrollo de modelos computacionales que buscan reducir el conjunto de posibles epítopos a ser verificados experimentalmente. Es importante destacar que, si bien el enfoque computacional es una herramienta valiosa para predecir epítopos inmunogénicos, se requiere en fases ulteriores la necesidad de llevar a cabo estudios experimentales adicionales para validar y confirmar la inmunogenicidad de los péptidos

seleccionados mediante herramientas bioinformáticas. Los ensayos *in-vitro* e *in-vivo* son necesarios para evaluar la capacidad de los epítopos identificados para inducir respuestas inmunitarias específicas y generar una protección efectiva contra el SARS-CoV-2 y otros patógenos de interés en salud pública.

Nuestros resultados confirman que el enfoque computacional integrativo implementado en este trabajo presenta una metodología prometedora para el diseño de vacunas contra el SARS-CoV-2 y otros patógenos que pueden afectar a la población colombiana. Al combinar la integración de herramientas bioinformáticas y la consideración de la diversidad genética de la población objetivo, utilizando pVaccine se obtuvieron resultados confiables en la identificación de epítopos inmunogénicos en dos tipos de microbio con complejidades genéticas distintas: un virus y en una bacteria lo que sugiere que este *pipeline* tiene el potencial de ser aplicado en otros patógenos de interés epidemiológico, lo que lo convierte en una herramienta versátil y valiosa en la lucha contra enfermedades infecciosas. Sin embargo, es importante tener en cuenta que el desarrollo de vacunas eficaces es un proceso complejo que requiere una combinación de enfoques computacionales y experimentales para garantizar su seguridad y eficacia. Además, como producto del arduo trabajo realizado, se ha generado un borrador para un documento de publicación, el cual se encuentra adjunto en los anexos de esta tesis.

8. Conclusiones y recomendaciones

8.1 Conclusiones

Aunque el proteoma de SARS-CoV-2 no es tan extenso como el de algunos otros virus, las herramientas de predicción de epítopos *in-silico* en la proteína S pueden generar listas de miles de péptidos potencialmente inmunogénicos para los linfocitos T CD8+. Sin embargo, la selección de epítopos adecuados para el diseño de vacunas no se basa únicamente en la afinidad entre el complejo MHC-péptido, sino que también se deben considerar otros parámetros biofísicos, como la estabilidad del complejo y otros aspectos relacionados con la inmunogenicidad.

Los parámetros biofísicos, como la estabilidad del complejo MHC-péptido, pueden afectar la presentación de antígenos a las células T de varias maneras. Por ejemplo, un complejo MHC-péptido que es inestable puede ser más susceptible a la degradación por las enzimas del sistema inmunitario, lo que puede reducir su capacidad para estimular las células T. Además, un complejo MHC-péptido que es inestable puede ser más susceptible a la interacción con otras moléculas, como los anticuerpos, lo que puede interferir con su presentación a las células T. Por lo tanto, es importante considerar los parámetros biofísicos al seleccionar epítopos para el diseño de vacunas. Las herramientas bioinformáticas pueden ayudar a identificar epítopos que tengan una afinidad favorable para el MHC y que sean estables a las mutaciones. Sin embargo, es importante validar experimentalmente estos epítopos para confirmar su inmunogenicidad y estabilidad.

En este trabajo, hemos abordado el desafío de optimizar la selección de epítopos inmunogénicos al integrar múltiples herramientas *in-silico* en un *pipeline* bioinformático. Además de utilizar la afinidad entre el complejo MHC-péptido, hemos incorporado otros parámetros relevantes para la predicción de la inmunogenicidad, lo que nos ha permitido identificar un universo de péptidos con potencial inmunogénico para el diseño de vacunas contra el SARS-CoV-2 y otros patógenos emergentes de interés epidemiológico como *Mycobacterium tuberculosis*.

La integración de la inmunología inversa fue una estrategia valiosa en este enfoque, ya que nos permitió seleccionar péptidos candidatos basados en información reportada en la literatura y comparar los resultados obtenidos con las predicciones bioinformáticas. Esto nos brinda una mayor confianza en los listados de epítopos seleccionadas como probables epítopos prometedores para hacer parte de una vacuna contra SARS-CoV-2.

Es importante destacar que, si bien el enfoque bioinformático es una herramienta valiosa y eficiente para predecir epítopos inmunogénicos, la validación experimental sigue siendo esencial. Los ensayos *in-vitro* e *in-vivo* son necesarios para evaluar la capacidad de los epítopos seleccionados para inducir respuestas inmunitarias específicas y generar una protección efectiva contra el SARS-CoV-2. Continuar explorando y refinando estas estrategias bioinformáticas y experimentales nos acerca a la creación de vacunas más seguras y eficaces para hacer frente a las amenazas de patógenos emergentes actuales y futuros.

Las pruebas *in-vitro* son importantes para identificar los efectos preliminares de un nuevo producto o tratamiento. Son relativamente rápidas y económicas, y permiten a los investigadores evaluar una amplia gama de parámetros, como la toxicidad, la actividad biológica y la biodisponibilidad. Las pruebas *in-vitro* tienen algunas limitaciones. Los resultados de las pruebas *in-vitro* no siempre se correlacionan con los resultados de las pruebas *in-vivo*. Además, las pruebas *in-vitro* pueden no tener en cuenta todos los factores que pueden afectar la seguridad y eficacia de un nuevo producto o tratamiento en un organismo vivo.

8.2 Recomendaciones

Como recomendaciones para futuros trabajos, es importante destacar que el diseño de vacunas basado en la predicción de epítopos inmunogénicos mediante enfoques bioinformáticos representa una estrategia prometedora en la lucha contra enfermedades infecciosas, como el SARS-CoV-2. En el transcurso de esta investigación, hemos explorado el desarrollo de un enfoque computacional integrativo utilizando herramientas bioinformáticas para predecir epítopos del proteoma de SARS-CoV-2 y, en particular, su relevancia para la población colombiana.

Ahora, con el objetivo de mejorar y avanzar en este campo, se presentan las siguientes recomendaciones que podrían servir como pautas para futuros trabajos:

- Ampliar la evaluación experimental: Aunque la predicción bioinformática es una herramienta valiosa, es esencial realizar experimentos adicionales para validar y confirmar la inmunogenicidad de los péptidos seleccionados. Esto puede incluir ensayos *in-vitro* e *in-vivo* para evaluar la capacidad de los epítopos para inducir respuestas inmunitarias específicas.

- Incluir otros patógenos: El enfoque bioinformático utilizado en este estudio ha demostrado ser prometedor en el diseño de vacunas contra el SARS-CoV-2. Sería interesante explorar su aplicación en otros patógenos de importancia epidemiológica, como virus de la influenza, dengue o zika, entre otros. Esto ampliará el alcance, validez y la utilidad de la metodología desarrollada.

- Mejorar la selección de parámetros biofísicos: Aunque se han utilizado diferentes parámetros biofísicos en la selección de epítopos, es importante seguir refinando y mejorando estos criterios. Investigar y evaluar la relevancia y peso de los parámetros, como la estabilidad del complejo MHC-péptido, la probabilidad de procesamiento proteasomal y la respuesta inmune previa en diferentes poblaciones, podría mejorar aún más la precisión de las predicciones.

- Actualización constante de las herramientas bioinformáticas: El campo de la bioinformática está en constante evolución y se desarrollan nuevas herramientas y algoritmos de manera permanente. Mantenerse actualizado con los avances en este campo y utilizar las herramientas más actualizadas puede mejorar la precisión y confiabilidad de las predicciones de epítopos inmunogénicos. La versatilidad y adaptabilidad de las herramientas bioinformáticas utilizadas en el pipeline son fundamentales para mantener su vigencia en un campo en constante evolución. Gracias a esta característica, el *pipeline* puede adaptarse fácilmente a los avances y actualizaciones en los algoritmos predictivos, lo que mejora la precisión y confiabilidad de las predicciones de epítopos inmunogénico; debido la naturaleza dinámica de la bioinformática requiere una continua actualización y mejora de las herramientas y algoritmos utilizados. La capacidad del *pipeline* para incorporar nuevas versiones y mejoras en las herramientas garantiza que pueda seguir siendo una herramienta relevante y efectiva en la identificación de epítopos inmunogénicos a medida que la investigación avanza.

Estas recomendaciones pueden ayudar a guiar futuros trabajos y contribuir al desarrollo de vacunas más efectivas y personalizadas no solo contra patógenos virales como el SARS-CoV-2, sino también contra otros patógenos de interés epidemiológico y enfermedades como el cáncer. La combinación de enfoques bioinformáticos, validación experimental y colaboración interdisciplinaria es clave para avanzar en este campo y hacer frente a las amenazas virales actuales y futuras, así como a los desafíos relacionados con enfermedades no infecciosas.

En el caso de enfermedades infecciosas, como el VIH, la influenza, el virus del Zika y otros patógenos emergentes o reemergentes, la aplicación de herramientas bioinformáticas y análisis genómicos puede permitir la identificación rápida de variantes virales y la selección de los epítopos más inmunogénicos para el diseño de vacunas. La personalización de las vacunas basadas en el perfil genético del patógeno y la población afectada puede mejorar significativamente la efectividad y la adaptabilidad de las estrategias de inmunización.

A. Anexo: Información suplementaria

En la carpeta de google Drive (https://drive.google.com/drive/folders/1gu3hSuwPIA0wDa4Yg5X76qrGG-zYc8Xl?usp=sharing) se encuentra la información suplementaria de los siguientes archivos.

- Suplementaria 1: Resultado SARS-CoV-2 pVACtools.
- Suplementaria 2: HLAs más prevalentes en la población Colombiana
- Suplementaria 3: Listado de péptidos publicados sin sustento experimental
- Suplementaria 4: Listado de péptidos publicados con sustento experimental
- Suplementaria 5: Péptidos reportados por autor y comparados con predichos por las herramientas
- Suplementaria 6: Péptidos seleccionados por uno o más autores y las herramientas
- Suplementaria 7: Péptidos seleccionados por tres o más autores y las herramientas
- Suplementaria 8: Péptidos seleccionados para Mycobacterium tuberculosis (M.tb H37Rv)

B. Anexo: Borrador documento de publicación

En la carpeta de google Drive (https://drive.google.com/drive/folders/19XhztchfUq0_749gW0Cj-4ppYhAx0I6v?usp=sharing) contiene el borrador del artículo de publicación, los anexos y figuras correspondientes al artículo.

Bibliografía

[1] Zhu, N. et al. A novel coronavirus from patients with pneumonia in China, 2019. N. Engl. J. Med. 382, 727–733 (2020).

[2] Riou, J. & Althaus, C. L. Pattern of early human-to-human transmission of Wuhan 2019 novel coronavirus (2019-nCoV), December 2019 to January 2020. Eurosurveillance 25, 2000058 (2020).

[3] Coronavirus disease (COVID-19). (n.d.). Retrieved December 11, 2021, from https://www.who.int/emergencies/diseases/novel-coronavirus-2019

[4] SARS-CoV-2 Variant Classifications and Definitions. (n.d.). Retrieved December 11, 2021, from https://www.cdc.gov/coronavirus/2019-ncov/variants/variant-classifications.html?CDC_AA_refVal=https%3A%2F%2Fwww.cdc.gov%2Fcoronavirus%2F2019-ncov%2Fvariants%2Fvariant-info.html

[5] Classification of Omicron (B.1.1.529): SARS-CoV-2 Variant of Concern. (n.d.). Retrieved December 11, 2021, from https://www.who.int/news/item/26-11-2021-classification-of-omicron-(b.1.1.529)-sars-cov-2-variant-of-concern

[6] Esquemas posológicos para el tratamiento de la infección de tuberculosis latente | Tratamiento | TB | CDC. (n.d.). Retrieved July 30, 2023, from https://www.cdc.gov/tb/esp/topic/treatment/ltbi.htm

[7] Hundal, J., Carreno, B. M., Petti, A. A., Linette, G. P., Griffith, O. L., Mardis, E. R., & Griffith, M. (2016). pVAC-Seq: A genome-guided in silico approach to identifying tumor neoantigens. Genome Medicine, 8(1), 1–11. https://doi.org/10.1186/S13073-016-0264-5/FIGURES/3

[8] V'kovski, P., Kratzel, A., Steiner, S., Stalder, H. & Thiel, V. Coronavirus biology and replication: implications for SARS-CoV-2. Nat. Rev. Microbiol. 19, 1–16 (2020).

[9] Zhang, Y., Park, C., Bennett, C., Thornton, M., & Kim, D. (2021). Rapid and accurate alignment of nucleotide conversion sequencing reads with HISAT-3N. Genome Research, 31(7), 1290–1295. https://doi.org/10.1101/GR.275193.120

[10] Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., & Durbin, R. (2009). The Sequence Alignment/Map format and SAMtools. Bioinformatics, 25(16), 2078–2079. https://doi.org/10.1093/BIOINFORMATICS/BTP352

[11] Scholak, T., Schucher, N., & Bahdanau, D. (2021). PICARD: Parsing Incrementally for Constrained Auto-Regressive Decoding from Language Models. EMNLP 2021 - 2021 Conference on Empirical Methods in Natural Language Processing, Proceedings, 9895–9901. https://doi.org/10.18653/v1/2021.emnlp-main.779

[12] Craik, DJ , Fairlie, DP , Liras, S. y Price, D. (2013). El futuro de los fármacos basados en péptidos . Chemical Biology & Drug Design , 81 (1), 136 - 147 . https://doi.org/10.1111/cbdd.12055

[13] Uhlig, T. , Kyprianou, T. , Martinelli, FG , Oppici, CA , Heiligers, D. , Hills, D. ,… Verhaert, P. (2014). La aparición de péptidos en el negocio farmacéutico: de la exploración a la explotación . Proteómica Abierta EuPA , 4 , 58 - 69 . https://doi.org/10.1016/j.euprot.2014.05.003

[14] Amaya Ramirez - Niño - Parra Lopez. (n.d.). Implementación de una estrategia in-silico para la predicción de epítopes potencialmente inmunogénicas en tumores de pacientes con cáncer (mama).https://extension.unal.edu.co/fileadmin/recursos/proyectos-importancia-institucional/medicina-traslacional/docs/Una_estrategia_in-silico_para_prediccion_de_neoantigenos.pdf

[15] Rezaei, S., Sefidbakht, Y., & Uskoković, V. (2021). Tracking the pipeline: immunoinformatics and the COVID-19 vaccine design. Briefings in Bioinformatics, 22(6), 1–20. https://doi.org/10.1093/bib/bbab241

[16] Fosgerau, K. y Hoffmann, T. (2015). Terapéutica de péptidos: estado actual y direcciones futuras . Drug Discovery Today , 20 (1), 122 - 128 . https://doi.org/10.1016/j.drudis.2014.10.003

[17] bgzip(1) manual page. (n.d.). Retrieved July 28, 2023, from http://www.htslib.org/doc/bgzip.html

[18] Troubleshooting GATK-SV – GATK. (n.d.). Retrieved July 28, 2023, from https://gatk.broadinstitute.org/hc/en-us/articles/5334566940699-Troubleshooting-GATK-SV

[19] bgzip(1) manual page. (n.d.). Retrieved July 28, 2023, from http://www.htslib.org/doc/bgzip.html

[20] tabix(1) manual page. (n.d.). Retrieved July 28, 2023, from http://www.htslib.org/doc/tabix.html

[21] bcftools(1). (n.d.). Retrieved July 28, 2023, from https://samtools.github.io/bcftools/bcftools.html

[22] Zhang, C., Bickis, M. G., Wu, F. X., & Kusalik, A. J. (2006). Optimally-connected hidden markov models for predicting MHC-binding peptides. Journal of Bioinformatics and Computational Biology, 4(5), 959–980. https://doi.org/10.1142/S0219720006002314

*[23] Doytchinova, I. A., & Flower, D. R. (2001). Toward the Quantitative Prediction of T-Cell Epitopes: CoMFA and CoMSIA Studies of Peptides with Affinity for the Class I MHC Molecule HLA-A*0201. Journal of Medicinal Chemistry, 44(22), 3572–3581. https://doi.org/10.1021/JM010021J*

[24] Kotsiantis, S. B. (2007). Supervised Machine Learning: A Review of Classification Techniques. Informatica, 31, 249–268.

[25] Sette, A., Buus, S., Appella, E., Smith, J. A., Chesnut, R., Miles, C., Colon, S. M., & Grey, H. M. (1989). Prediction of major histocompatibility complex binding regions of protein antigens by sequence pattern analysis. Proceedings of the

National Academy of Sciences of the United States of America, 86(9), 3296–3300. https://doi.org/10.1073/PNAS.86.9.3296

[26] Agerer, B., Koblischke, M., Gudipati, V., Montaño-Gutierrez, L. F., Smyth, M., Popa, A., Genger, J.-W., Endler, L., Florian, D. M., Mühlgrabner, V., Graninger, M., Aberle, S. W., Husa, A.-M., Shaw, L. E., Lercher, A., Gattinger, P., Torralba-Gombau, R., Trapin, D., Penz, T., … Bergthaler, A. (2021). SARS-CoV-2 mutations in MHC-I-restricted epitopes evade CD8 + T cell responses. Science Immunology, 6(57). https://doi.org/10.1126/sciimmunol.abg6461

[27] Prachar, M., Justesen, S., Steen-Jensen, D.B. et al. Identification and validation of 174 COVID-19 vaccine candidate epitopes reveals low performance of common epitope prediction tools. Sci Rep 10, 20465 (2020). https://doi.org/10.1038/s41598-020-77466-4

[28] Harndahl, M. et al. Peptide binding to HLA class I molecules: homogenous, high-throughput screening, and affinity assays. J. Biomol. Screen. 14, 173–180 (2009).

[29] Polyiam, K., Phoolcharoen, W., Butkhot, N. et al. Immunodominant linear B cell epitopes in the spike and membrane proteins of SARS-CoV-2 identified by immunoinformatics prediction and immunoassay. Sci Rep 11, 20383 (2021). https://doi.org/10.1038/s41598-021-99642-w

[30] Orsburn, B., Jenkins, C., Miller, S. M., Neely, B. A., & Bumpus, N. M. (2020). In silico - Approach Toward the Identification of Unique Peptides from Viral Protein Infection: Application to COVID-19. SSRN Electronic Journal. https://doi.org/10.2139/SSRN.3589835

[31] Sitthiyotha, T., & Chunsrivirot, S. (2020). Computational Design of 25-mer Peptide Binders of SARS-CoV-2. Journal of Physical Chemistry B, 124(48), 10930–10942.

https://doi.org/10.1021/ACS.JPCB.0C07890/SUPPL_FILE/JP0C07890_SI_001.PDF

[32] Peng, Y., Mentzer, A. J., Liu, G., Yao, X., Yin, Z., Dong, D., Dejnirattisai, W., Rostron, T., Supasa, P., Liu, C., López-Camacho, C., Slon-Campos, J., Zhao, Y.,

Stuart, D. I., Paesen, G. C., Grimes, J. M., Antson, A. A., Bayfield, O. W., Hawkins, D. E. D. P., … Dong, T. (2020). Broad and strong memory CD4+ and CD8+ T cells induced by SARS-CoV-2 in UK convalescent individuals following COVID-19. Nature Immunology, 21(11), 1336–1345. https://doi.org/10.1038/s41590-020-0782-6

[33] Machuca, I., Vidal, E., de la Torre-Cisneros, J., & Rivero-Román, A. (2018). Tuberculosis in immunosuppressed patients. Enfermedades Infecciosas y Microbiologia Clinica (English Ed.), 36(6), 366–374. https://doi.org/10.1016/J.EIMC.2017.10.009

[34] Peters, B., Nielsen, M. & Sette, A. T cell epitope predictions. Annu. Rev. Immunol. https://doi.org/10.1146/annurev-immunol-082119 (2019).

[35] Mei, S. et al. A comprehensive review and performance evaluation of bioinformatics tools for HLA class I peptide-binding prediction. Brief. Bioinform. 21, 1119–1135 (2020).

[36] Saethang, T. et al. EpicCapo: epitope prediction using combined information of amino acid pairwise contact potentials and HLA-peptide contact site information. BMC Bioinform. 13, 313 (2012).

[37] The Variant Call Format (VCF) Version 4.2 Specification. (2022).

[38] Szolek, A, Schubert, B, Mohr, C, Sturm, M, Feldhahn, M, and Kohlbacher, O (2014). OptiType: precision HLA typing from next-generation sequencing data Bioinformatics, 30(23):3310-6.

[39] McLaren W, Gil L, Hunt SE, Riat HS, Ritchie GR, Thormann A, Flicek P, Cunningham F. The Ensembl Variant Effect Predictor. Genome Biology Jun 6;17(1):122. (2016) https://doi:10.1186/s13059-016-0974-4

[40] Lank, S. M., Golbach, B. A., Creager, H. M., Wiseman, R. W., Keskin, D. B., Reinherz, E. L., Brusic, V., & O'Connor, D. H. (2012). Ultra-high resolution HLA genotyping and allele discovery by highly multiplexed cDNA amplicon pyrosequencing. BMC Genomics, 13(1). https://doi.org/10.1186/1471-2164-13-378

[41] McLaren, W., Gil, L., Hunt, S. E., Riat, H. S., Ritchie, G. R. S., Thormann, A., Flicek, P., & Cunningham, F. (2016). The Ensembl Variant Effect Predictor.

Genome Biology, 17(1), 1–14. https://doi.org/10.1186/S13059-016-0974-4/TABLES/8

[42] NetMHC - 4.0 - Services - DTU Health Tech. (n.d.). Retrieved April 27, 2022, from https://services.healthtech.dtu.dk/service.php?NetMHC-4.0

[43] NetMHCpan - 4.0 - Services - DTU Health Tech. (n.d.). Retrieved April 28, 2022, from https://services.healthtech.dtu.dk/service.php?NetMHCpan-4.0

[44] O'Donnell, T. J., Rubinsteyn, A., & Laserson, U. (2020). MHCflurry 2.0: Improved Pan-Allele Prediction of MHC Class I-Presented Peptides by Incorporating Antigen Processing. Cell Systems, 11(1), 42-48.e7. https://doi.org/10.1016/J.CELS.2020.06.010

[45] NetMHCstabpan - 1.0 - Services - DTU Health Tech. (n.d.). Retrieved April 28, 2022, from https://services.healthtech.dtu.dk/service.php?NetMHCstabpan-1.0

[46] Karchin Lab Johns Hopkins University SCHISM. (n.d.). Retrieved April 28, 2022, from https://karchinlab.org/apps/appMHCnuggets.html

[47] NetMHCstabpan - 1.0 - Services - DTU Health Tech. (n.d.). Retrieved April 28, 2022, from https://services.healthtech.dtu.dk/service.php?NetMHCstabpan-1.0

[48] Vacunas y fármacos biotecnológicos (uab.cat)

[49] [Reverse vaccinology: strategy against emerging pathogens] - PubMed (nih.gov)

[50] Componentes celulares del sistema inmunitario - Inmunología y trastornos alérgicos - Manual MSD versión para profesionales (msdmanuals.com)

[51] (2021-12-11) Representación tridimensional en la que se muestran las cuatro proteínas de superficie del virus: E, S, M, HE. tomado de: https://www.scientificanimations.com

[52] Imagen tomada de https://ambientech.org/mycobacterium-tuberculosis

[53] Gorbalenya, A. E., Baker, S. C., Baric, R. S., de Groot, R. J., Drosten, C., Gulyaeva, A. A., Haagmans, B. L., Lauber, C., Leontovich, A. M., Neuman, B. W., Penzar, D., Perlman, S., Poon, L. L. M., Samborskiy, D. v., Sidorov, I. A., Sola, I., & Ziebuhr, J. (2020). The species Severe acute respiratory syndrome-related coronavirus: classifying 2019-nCoV and naming it SARS-CoV-2. Nature Microbiology, 5(4), 536. https://doi.org/10.1038/S41564-020-0695-Z

[54] Tracking SARS-CoV-2 variants. (n.d.). Retrieved July 30, 2023, from https://www.who.int/activities/tracking-SARS-CoV-2-variants/

[55] Síntomas del COVID-19 | CDC. (n.d.). Retrieved July 30, 2023, from https://espanol.cdc.gov/coronavirus/2019-ncov/symptoms-testing/symptoms.html

[56] Cómo se propaga el coronavirus | CDC. (n.d.). Retrieved July 30, 2023, from https://espanol.cdc.gov/coronavirus/2019-ncov/prevent-getting-sick/how-covid-spreads.html

[57] Huang, Y., Yang, C., Xu, X. feng, Xu, W., & Liu, S. wen. (2020). Structural and functional properties of SARS-CoV-2 spike protein: potential antivirus drug development for COVID-19. Acta Pharmacologica Sinica 2020 41:9, 41(9), 1141–1149. https://doi.org/10.1038/s41401-020-0485-4

[58] Walls, A. C., Park, Y. J., Tortorici, M. A., Wall, A., McGuire, A. T., & Veesler, D. (2020). Structure, Function, and Antigenicity of the SARS-CoV-2 Spike Glycoprotein. Cell, 181(2), 281-292.e6. https://doi.org/10.1016/J.CELL.2020.02.058

[59] Tuberculosis. (n.d.). Retrieved July 30, 2023, from https://www.who.int/news-room/fact-sheets/detail/tuberculosis

[60] Basic TB Facts | TB | CDC. (n.d.). Retrieved July 30, 2023, from https://www.cdc.gov/tb/topic/basics/default.htm

[61] Junttila, I. S. (2018). Tuning the cytokine responses: An update on interleukin (IL)-4 and IL-13 receptor complexes. Frontiers in Immunology, 9(JUN). https://doi.org/10.3389/FIMMU.2018.00888/FULL

[62] Yarmohammadi, H., & Cunningham-Rundles, C. (2017). Idiopathic CD4 lymphocytopenia: Pathogenesis, etiologies, clinical presentations and treatment strategies. Annals of Allergy, Asthma and Immunology, 119(4), 374–378. https://doi.org/10.1016/j.anai.2017.07.021

[63] Generalidades sobre el sistema inmunitario - Inmunología y trastornos alérgicos - Manual Merck versión para profesionales. (n.d.). Retrieved July 30, 2023, from https://www.merckmanuals.com/es-us/professional/inmunolog%C3%ADa-y-trastornos-al%C3%A9rgicos/biolog%C3%ADa-del-sistema-inmunitario/generalidades-sobre-el-sistema-inmunitario

[64] Grifoni, A., Weiskopf, D., Ramirez, S. I., Mateus, J., Dan, J. M., Moderbacher, C. R., Rawlings, S. A., Sutherland, A., Premkumar, L., Jadi, R. S., Marrama, D., de Silva, A. M.,

Frazier, A., Carlin, A. F., Greenbaum, J. A., Peters, B., Krammer, F., Smith, D. M., Crotty, S., & Sette, A. (2020). Targets of T Cell Responses to SARS-CoV-2 Coronavirus in Humans with COVID-19 Disease and Unexposed Individuals. Cell, 181(7), 1489-1501.e15. https://doi.org/10.1016/J.CELL.2020.05.015

[65] Mateus, J., Grifoni, A., Tarke, A., Sidney, J., Ramirez, S. I., Dan, J. M., Burger, Z. C., Rawlings, S. A., Smith, D. M., Phillips, E., Mallal, S., Lammers, M., Rubiro, P., Quiambao, L., Sutherland, A., Yu, E. D., da Silva Antunes, R., Greenbaum, J., Frazier, A., … Weiskopf, D. (2020). Selective and cross-reactive SARS-CoV-2 T cell epitopes in unexposed humans. Science (New York, N.Y.), 370(6512). https://doi.org/10.1126/SCIENCE.ABD3871

[66] Zhang, H., Deng, S., Ren, L., Zheng, P., Hu, X., Jin, T., & Tan, X. (2021). Profiling CD8 + T cell epitopes of COVID-19 convalescents reveals reduced cellular immune responses to SARS-CoV-2 variants. Cell Reports, 36(11). https://doi.org/10.1016/J.CELREP.2021.109708

[67] Moise, L., Gutierrez, A., Kibria, F., Martin, R., Tassone, R., Liu, R., Terry, F., Martin, B., & de Groot, A. S. (2015). iVAX: An integrated toolkit for the selection and optimization of antigens and the design of epitope-driven vaccines. Human Vaccines & Immunotherapeutics, 11(9), 2312–2321. https://doi.org/10.1080/21645515.2015.1061159

[68] Liu, G., Carter, B., Bricken, T., Jain, S., Viard, M., Carrington, M., & Gifford, D. K. (2020). Computationally Optimized SARS-CoV-2 MHC Class I and II Vaccine Formulations Predicted to Target Human Haplotype Distributions. Cell Systems, 11(2), 131. https://doi.org/10.1016/J.CELS.2020.06.009

[69] Kared, H., Redd, A. D., Bloch, E. M., Bonny, T. S., Sumatoh, H., Kairi, F., Carbajo, D., Abel, B., Newell, E. W., Bettinotti, M. P., Benner, S. E., Patel, E. U., Littlefield, K., Laeyendecker, O., Shoham, S., Sullivan, D., Casadevall, A., Pekosz, A., Nardin, A., … Quinn, T. C. (2020). CD8+ T cell responses in convalescent COVID-19 individuals target epitopes from the entire SARS-CoV-2 proteome and show kinetics of early differentiation. BioRxiv : The Preprint Server for Biology. https://doi.org/10.1101/2020.10.08.330688

[70] Grifoni, A., Sidney, J., Zhang, Y., Scheuermann, R. H., Peters, B., & 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.e2. https://doi.org/10.1016/J.CHOM.2020.03.002

[71] Finkel, Y., Mizrahi, O., Nachshon, A., Weingarten-Gabbay, S., Morgenstern, D., Yahalom-Ronen, Y., Tamir, H., Achdout, H., Stein, D., Israeli, O., Beth-Din, A., Melamed, S., Weiss, S., Israely, T., Paran, N., Schwartz, M., & Stern-Ginossar, N. (2020). The coding capacity of SARS-CoV-2. Nature 2020 589:7840, 589(7840), 125–130. https://doi.org/10.1038/s41586-020-2739-1

[72] Agerer, B., Koblischke, M., Gudipati, V., Montaño-Gutierrez, L. F., Smyth, M., Popa, A., Genger, J.-W., Endler, L., Florian, D. M., Mühlgrabner, V., Graninger, M., Aberle, S. W., Husa, A.-M., Shaw, L. E., Lercher, A., Gattinger, P., Torralba-Gombau, R., Trapin, D., Penz, T., … Bergthaler, A. (2021). SARS-CoV-2 mutations in MHC-I-restricted epitopes evade CD8 + T cell responses. Science Immunology, 6(57). https://doi.org/10.1126/sciimmunol.abg6461

[73] Campbell, K. M., Steiner, G., Wells, D. K., Ribas, A., & Kalbasi, A. (2020). Prioritization of SARS-CoV-2 epitopes using a pan-HLA and global population inference approach. BioRxiv : The Preprint Server for Biology. https://doi.org/10.1101/2020.03.30.016931

[74] Daouda, T., Dumont-Lagacé, M., Feghaly, A., & Villani, A.-C. (2021). Codon arrangement modulates MHC-I peptides presentation: implications for a SARS-CoV-2 peptide-based vaccine. BioRxiv, 2021.02.04.429819. https://doi.org/10.1101/2021.02.04.429819

[75] Kared, H., Redd, A. D., Bloch, E. M., Bonny, T. S., Sumatoh, H., Kairi, F., Carbajo, D., Abel, B., Newell, E. W., Bettinotti, M. P., Benner, S. E., Patel, E. U., Littlefield, K., Laeyendecker, O., Shoham, S., Sullivan, D., Casadevall, A., Pekosz, A., Nardin, A., … Quinn, T. C. (2020). CD8+ T cell responses in convalescent COVID-19 individuals target epitopes from the entire SARS-CoV-2 proteome and show kinetics of early differentiation. BioRxiv : The Preprint Server for Biology. https://doi.org/10.1101/2020.10.08.330688

[76] Mallajosyula, V., Ganjavi, C., Chakraborty, S., McSween, A. M., Pavlovitch-Bedzyk, A. J., Wilhelmy, J., Nau, A., Manohar, M., Nadeau, K. C., & Davis, M. M. (2021). CD8+ T cells specific for conserved coronavirus epitopes correlate with milder disease in COVID-19 patients. Science Immunology, 6(61). https://doi.org/10.1126/sciimmunol.abg5669

[77] Nathan, A., Rossin, E. J., Kaseke, C., Park, R. J., Khatri, A., Koundakjian, D., Urbach, J. M., Singh, N. K., Bashirova, A., Tano-Menka, R., Senjobe, F., Waring, M. T., Piechocka-Trocha, A., Garcia-Beltran, W. F., Iafrate, A. J., Naranbhai, V., Carrington, M.,

Walker, B. D., & Gaiha, G. D. (2021). Structure-guided T cell vaccine design for SARS-CoV-2 variants and sarbecoviruses. Cell, 184(17), 4401-4413.e10. https://doi.org/10.1016/j.cell.2021.06.029

[78] Peng, Y., Mentzer, A. J., Liu, G., Yao, X., Yin, Z., Dong, D., Dejnirattisai, W., Rostron, T., Supasa, P., Liu, C., López-Camacho, C., Slon-Campos, J., Zhao, Y., Stuart, D. I., Paesen, G. C., Grimes, J. M., Antson, A. A., Bayfield, O. W., Hawkins, D. E. D. P., … Dong, T. (2020). Broad and strong memory CD4+ and CD8+ T cells induced by SARS-CoV-2 in UK convalescent individuals following COVID-19. Nature Immunology, 21(11), 1336–1345. https://doi.org/10.1038/s41590-020-0782-6

[79] Prachar, M., Justesen, S., Steen-Jensen, D. B., Thorgrimsen, S., Jurgons, E., Winther, O., & Bagger, F. O. (2020). Identification and validation of 174 COVID-19 vaccine candidate epitopes reveals low performance of common epitope prediction tools. Scientific Reports, 10(1), 20465. https://doi.org/10.1038/s41598-020-77466-4

[80] Quadeer, A. A., Ahmed, S. F., & McKay, M. R. (2021). Landscape of epitopes targeted by T cells in 852 individuals recovered from COVID-19: Meta-analysis, immunoprevalence, and web platform. Cell Reports Medicine, 2(6), 100312. https://doi.org/10.1016/j.xcrm.2021.100312

[81] Schulien, I., Kemming, J., Oberhardt, V., Wild, K., Seidel, L. M., Killmer, S., Sagar, Daul, F., Salvat Lago, M., Decker, A., Luxenburger, H., Binder, B., Bettinger, D., Sogukpinar, O., Rieg, S., Panning, M., Huzly, D., Schwemmle, M., Kochs, G., … Neumann-Haefelin, C. (2021). Characterization of pre-existing and induced SARS-CoV-2-specific CD8+ T cells. Nature Medicine, 27(1), 78–85. https://doi.org/10.1038/s41591-020-01143-2

[82] Sohail, M. S., Ahmed, S. F., Quadeer, A. A., & McKay, M. R. (2021). In silico T cell epitope identification for SARS-CoV-2: Progress and perspectives. Advanced Drug Delivery Reviews, 171, 29–47. https://doi.org/10.1016/j.addr.2021.01.007

[83] Weingarten-Gabbay, S., Klaeger, S., Sarkizova, S., Pearlman, L. R., Chen, D.-Y., Bauer, M. R., Taylor, H. B., Conway, H. L., Tomkins-Tinch, C. H., Finkel, Y., Nachshon, A., Gentili, M., Rivera, K. D., Keskin, D. B., Rice, C. M., Clauser, K. R., Hacohen, N., Carr, S. A., Abelin, J. G., … Sabeti, P. C. (2020). SARS-CoV-2 infected cells present HLA-I peptides from canonical and out-of-frame ORFs. BioRxiv : The Preprint Server for Biology. https://doi.org/10.1101/2020.10.02.324145

[84] The web framework for perfectionists with deadlines | Django. (n.d.). Retrieved October 31, 2023, from https://www.djangoproject.com/

[85] Conda — conda documentation. (n.d.). Retrieved October 31, 2023, from https://docs.conda.io/en/latest/

[86] Docker: Accelerated Container Application Development. (n.d.). Retrieved October 31, 2023, from https://www.docker.com/

[87] *Scrum y las metodologías ágiles en construcción - miguelgarcia.me*. (n.d.). Retrieved November 1, 2023, from https://miguelgarcia.me/scrum-y-las-metodologias-agiles-en-construccion/