

UNIVERSIDAD DEL VALLE DE GUATEMALA
Facultad de Ciencias y Humanidades



Análisis estructural y proteómico de la glucoproteína *spike* del SARS-CoV-2 en Centroamérica: el presente y futuro del virus que conquistó al mundo

Trabajo de graduación presentado por Álvaro Bautista Hurtarte para optar al grado de Licenciado en Microbiología y Bioquímica

Guatemala,

2022

UNIVERSIDAD DEL VALLE DE GUATEMALA
Facultad de Ciencias y Humanidades



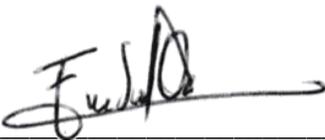
Análisis estructural y proteómico de la glucoproteína *spike* del SARS-CoV-2 en Centroamérica: el presente y futuro del virus que conquistó al mundo

Trabajo de graduación presentado por Álvaro Bautista Hurtarte para optar al grado de Licenciado en Microbiología y Bioquímica

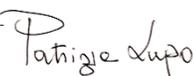
Guatemala,

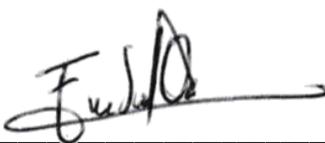
2022

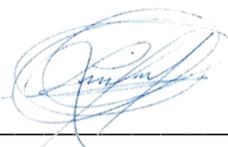
Vo. Bo.:

(f) 
M.Sc. Elizabeth Solorzano

Tribunal examinador

(f) 
PhD. Patrizia Lupo

(f) 
M.Sc. Elizabeth Solorzano

(f) 
M.Sc. Augusto Franco

Fecha de aprobación: Guatemala 08 de diciembre del 2021

ÍNDICE

Listado de cuadros	vi
Listado de figuras.....	vii
Resumen	viii
I. Introducción	1
II. Objetivos.....	3
III. Justificación	4
IV. Marco teórico	6
A. SARS-CoV-2	7
B. GLucoproteína <i>spike</i>	19
C. Efecto de las modificaciones en proteínas	22
V. Metodología.....	32
A. Alineamiento y filtrado de calidad del genoma	32
B. Modelaje molecular y análisis mutagenómico.....	33
C. Análisis predictivo del panorama de glicosilación	34
D. Análisis de recombinación	34
E. Ruta bioinformática.....	35
VI. Resultados.....	35
A. Análisis mutagenómico	36
B. Modelaje molecular y análisis de acoplamiento molecular	42
C. Análisis predictivo del panorama de glicosilación	43
D. Análisis de recombinación	46
VII. Análisis de resultados.....	47
A. Análisis mutagenómico	48
B. Análisis de acoplamiento molecular	49
C. Análisis predictivo del panorama de glicosilación	50
D. Análisis de recombinación	51
VIII. Conclusiones.....	51
IX. Recomendaciones.....	52
X. Literatura citada	53

XI. Anexos 68

LISTADO DE CUADROS

Cuadro 1: Función de las proteínas accesorias del SC2	12
Cuadro 2: Función de las 16 proteínas no estructurales codificadas por el genoma viral del SC2	14
Cuadro 3: Secuencias utilizadas para el estudio de SARS-CoV-2 obtenidas de GISAID	32
Cuadro 4: Frecuencia de mutaciones en las secuencias de SARS-CoV-2 en Centroamérica.....	36
Cuadro 5: El efecto de la sustitución de aminoácidos en la proteína S del SARS-CoV-2 sobre la estabilidad de su microambiente	37
Cuadro 6: Análisis de acoplamiento molecular realizado con HADDOCK	43
Cuadro 7: Frecuencia de glicosilaciones ligadas a O en la proteína S de las secuencias de Centroamérica de SARS-CoV-2.	44
Cuadro 8: Secuencias de SARS-CoV-2 de la proteína S de Centroamérica que no presentan glicosilaciones ligadas a O en las posiciones 673, 678 y/o 686.....	44
Cuadro 9: Presencia de glicosilaciones ligadas a O en las secuencias de la proteína S del SARS-CoV-2 en Centroamérica	45
Anexo 1: Lista completa de frecuencia de mutaciones en la proteína S en las secuencias de Centroamérica del SARS-Cov-2.	68
Anexo 2: Frecuencia de mutaciones en las secuencias de SARS-CoV-2 en Belice.....	70
Anexo 3: Frecuencia de mutaciones en las secuencias de SARS-CoV-2 en El Salvador	71
Anexo 4: Frecuencia de mutaciones en las secuencias de SARS-CoV-2 en Guatemala.....	71
Anexo 5: Frecuencia de mutaciones en las secuencias de SARS-CoV-2 en Honduras	73
Anexo 6: Frecuencia de mutaciones en las secuencias de SARS-CoV-2 en Panamá	73
Anexo 7: Frecuencia de mutaciones en las secuencias de SARS-CoV-2 en Costa Rica.....	74
Anexo 8: Presencia de glicosilaciones ligadas a O en las secuencias de Centroamérica	76
Anexo 9: Distribución de las mutaciones en los distintos dominios de Subunidad S1 la proteína S del SARS-Cov2 en las secuencias de Centroamérica	82
Anexo 10: Distribución de las mutaciones en los distintos dominios de Subunidad S2 la proteína S del SARS-Cov2 en las secuencias de Centroamérica.....	84
Anexo 11: Análisis de acoplamiento molecular realizado con HADDOCK	85

LISTADO DE FIGURAS

Figura 1: Esquema del genoma viral y estructura del SC2-----	11
Figura 2: Tipos de N-glucanos -----	24
Figura 3: Núcleos de glucanos Tipos de O-GalNAc-----	26
Figura 4: Equilibrio entre proteínas plegadas y desplegadas y su relación con la estabilidad proteica en proteínas mutadas y tipo WT-----	30
Figura 5: Flujo de trabajo para análisis bioinformático de las secuencias de Centroamérica de SARS-CoV-2-----	35
Figura 6: Representación molecular de las 20 posiciones donde ocurren más frecuentemente mutaciones en las secuencias de Centroamérica de una cadena de la proteína S del SARS-CoV-2 glicosilada utilizando el modelo de referencia 6VXX de Woo y compañía-----	42
Figura 8: Gráfico de similitud utilizando la secuencia de referencia NC_045512 como comparación entre cada país-----	47

RESUMEN

El SARS-CoV2 (SC2) es un virus de ARN con gran infectividad en humanos y el agente etiológico de la enfermedad denominada COVID-19. Se transmite utilizando la glicoproteína de membrana S que usa como receptor la enzima convertidora de angiotensina 2 (ACE2), llegando a ser activada por corte proteolítico por distintas proteasas (TMPRSS2, furina, etc.). En este estudio se evaluó el efecto de las mutaciones, modificaciones postraduccionales y recombinación genética en la estructura de la proteína S y su afinidad para formar el complejo RBD-ACE2. Utilizando análisis *in silico* se realizó una predicción de sitios potenciales de glicosilaciones ligadas a N y a O, se determinó si existía o no evidencia de recombinación genética, por último y se evaluó si las mutaciones encontradas, afectaban la estructura primaria de la proteína S y su afinidad por ACE2, a través de la evaluación de secuencias genómicas de SARS-CoV-2 procedentes de Centroamérica (Belice, Guatemala, El Salvador, Honduras, Panamá y Costa Rica), depositadas en el repositorio GISAID. Pareciera que hay una menor energía libre en aquellas mutaciones con baja frecuencia en la población a comparación de aquellas con alta frecuencia, sin embargo, esto no quiere decir que las mutaciones con un $\Delta\Delta G$ cercano a cero no afecten la estructura tridimensional. La alta afinidad de V503F, D614G y T1117I por ACE2 podrían deberse a una sinergia, posiblemente creando una cepa superinfectiva. Adicionalmente, S459F eleva la afinidad por ACE2, así como las mutaciones del linaje B.1.1.328 y B.1.575 denotando nuevamente que a pesar del leve efecto desestabilizante podría afectar la estructura y función de la proteína S. Se pudo observar un panorama modificado de glicosilaciones ligadas a N y a O, causado posiblemente la presencia de mutaciones desestabilizante y afectando la disponibilidad por glicosiltransferasas que perturban la antigenicidad de la proteína S y su interacción con ACE2. No se detectaron eventos de recombinación en las secuencias de Centroamérica, dado por una alta similitud entre las mismas. En ese análisis realizado se observa una baja frecuencia de mutaciones, pero eso se debe a que el número de datos disponibles en la CA no son suficientes para describir el comportamiento del SARS-CoV-2 en el tiempo. Los países en la región Centro Americana deberían de implementar la vigilancia genómica en apoyo a la vigilancia epidemiológica con el fin de ayudar a mitigar amenazas y controlar la dispersión del virus.

I. INTRODUCCIÓN

El SC2 es un virus de ARN (+) de una sola hebra causante de la enfermedad denominada COVID-19. El primer brote detectado fue en la ciudad de Wuhan, China y actualmente se encuentra diseminado en todos los continentes. Un evento clave que ayudó a la diseminación inicial del virus, fue la coincidencia del brote inicial con la proximidad del Año Nuevo lunar, por lo que los viajes entre ciudades facilitaron su transmisión en China y posteriormente el mundo (Izda *et al.*, 2021). A pesar de la severidad al inicio de la pandemia en el continente asiático y europeo, la región de América es la región más afectada a la fecha, seguida de Europa, sudeste de Asia hasta el 15 noviembre (Elflein, 2021; Hu *et al.*, 2021).

Este coronavirus cuenta con un genoma de 29881bp que codifican proteínas estructurales (S, E, M y N) y no estructurales (proteasas y polimerasa dependiente de ARN, RdRp) (Y. Huang *et al.*, 2020). Una proteína clave es la glicoproteína *spike* (S), que es utilizada por el virus para mediar su entrada a la célula huésped al interactuar con su receptor, ACE2. La proteína sufre un cambio conformacional al ser escindida proteolíticamente por distintas proteasas (furina, proteasa transmembrana serina 2 (TMPRSS2), la tripsina y la catépsina L) (Samudrala *et al.*, 2020).

Una característica del virus que afectó la evolución de la pandemia es que su material genético es de ARN, lo que significa que muta con gran facilidad. Se estima que el SARS-CoV-2 tiene una tasa de mutación genómica de 3.7×10^{-6} unidad de nucleótidos por ciclo de infección, para el linaje con la proteína S descrita originalmente y de 2.9×10^{-6} nt-1 ciclo⁻¹ para un linaje con la mutación D614G. La tasa de mutación en el genoma de este virus no es homogénea, para la proteína S se tiene una tasa de 16×10^{-6} nt-1 por ciclo (Borges *et al.*, 2021). Las mutaciones que se puedan dar en la proteína S son claves y muy importantes, ya que puede afectar su estructura y por ende su función, logrando evadir el sistema inmune, aumentando la virulencia del virus e incluso pueden disminuir la eficiencia o dejar obsoletos los métodos de diagnóstico molecular enfocados en el gen que codifica para esta proteína (Asghari *et al.*, 2020).

Además de las mutaciones puntuales, un proceso que afecta de manera radical a la proteína S es la recombinación genética. Esta involucra dos moléculas separadas de ADN o ARN, que intercambian regiones homólogas en el genoma, a partir de un proceso de entrecruzamiento. En virus, la recombinación lleva a la creación de moléculas quiméricas a partir de genomas parentales de diferente origen puede causar expansión en los huéspedes virales, aumento en la virulencia, evasión de la inmunidad del huésped y la evolución de la resistencia a los antivirales (Fravev, 2020). Sin embargo, este mecanismo es un poco más difícil de detectar, ya que involucra la infección de

dos virus con genomas suficientemente distintos entre sí, en una misma célula, pero indicaría un salto enorme en la evolución viral (Walls *et al.*, 2020). Sin embargo, para el SARS-CoV-2 se ha reportado con estudios recientes que hay niveles bajos de virus recombinantes a nivel mundial, siendo 2.7% de 1,6 millones de secuencias (Turkahia *et al.*, 2021).

Además de los mecanismos mencionados anteriormente, las modificaciones postraduccionales ocurren en las últimas etapas de la biosíntesis de las proteínas y aportan cambios químicos reversibles o irreversibles en la proteína (Samudrala *et al.*, 2020). Dentro de estos cambios están las glicosilaciones ligadas a N y a O, que afectan la estructura y la función del péptido, así como la estabilidad de la proteína y regulación de la bioactividad (Andersen *et al.*, 2020).

Por otro lado, los virus no cuentan con maquinaria celular, por lo que utilizan la maquinaria del huésped para su replicación; estos procesos pueden inducir cambios que podrían afectar la estabilidad y estructura de la proteína S. Cambios como la recombinación y mutaciones puntuales son fuerzas evolutivas, que incluso podrían volver más mortal o infeccioso al virus.

Al ser un virus previamente desconocido el que causa esta pandemia, aún hay muchos vacíos de conocimiento en cuanto a su evolución y la patología que causa. Es por ello por lo que es necesaria una vigilancia genómica activa del virus, sin embargo, la falta de recursos económicos y recurso humano formado en las áreas de genómica y bioinformática en regiones como Centroamérica ha creado una desigualdad en cuanto a la información disponible por país.

Es por ello por lo que en este estudio se pretende evaluar cómo los dichos cambios anteriormente descritos dirigidos a la proteína S pueden afectar su estructura y la formación del complejo RBD-ACE2. Para ello se utilizarán análisis *in silico* para determinar si existe recombinación genética y evaluar el panorama de glicosilación ligado a N y a O en las secuencias de Centroamérica (Belize, Guatemala, El Salvador, Honduras, Panamá y Costa Rica), disponibles en el repositorio GISAID. Adicionalmente, se construirán modelos moleculares para evaluar cómo las mutaciones afectan la afinidad de la proteína S a su receptor, lo cual se evaluará por medio de acoplamiento molecular.

A. PREGUNTA DE INVESTIGACIÓN

¿Qué efectos han producido las mutaciones, recombinación genética y modificaciones postraduccionales en la estructura de la proteína S del SARS-CoV-2 y formación del complejo RBD-ACE2, en secuencias provenientes de Centroamérica?

1. Hipótesis

Las modificaciones nucleotídicas en las secuencias provenientes de Centroamérica de la proteína *Spike* afectan su estructura y función en la formación del complejo RBD-ACE2

2. Variables y definiciones

- **Mutaciones:** variaciones en el código a causa de daño en el genoma o un error en la replicación. Estas pueden llegar a crear péptidos inservibles o proteínas con características adicionales (ventajas evolutivas como mayor afinidad a cierto sustrato) (Allison, 2007). Estas pueden ser identificadas al comparar un genoma en cuestión contra uno de referencias (detectando mutaciones puntuales, SNPs), además de evaluar el grado en que afectan a la estructura terciaria midiendo estabilidad del producto, midiendo si afecta su unión con un sustrato específico por modelaje molecular.
- **Presencia recombinación genética:** la recombinación genética es un mecanismo que permite el intercambio de información entre dos hebras distintas de material genético (ADN o ARN) (Andersen *et al.*, 2020). Este puede ser medido al comparar distintas secuencias contra una de referencia, mediante análisis *in silico* (Pang & Wilkins, 2018).
- **Modificaciones postraduccionales:** estas son modificaciones como fosforilaciones, glicosilaciones, acetilaciones que dan después de que se produce el péptido con fin de alterar la función de la proteína, llevando consigo posibles ventajas (Pang & Wilkins, 2018). Estas se pueden medir al analizar sitios candidatos que presentarían estas modificaciones y mediante parámetros se determina si en condiciones normales esta secuencia podría tener las modificaciones postraduccionales en cuestión (Uversky, 2013).

II. OBJETIVOS

A. OBJETIVO GENERAL

Determinar *in silico* el efecto que tienen sobre la estructura y función de la glucoproteína S, las mutaciones, recombinación genética y modificaciones postraduccionales en las secuencias de SARS-CoV-2 de Centroamérica (Belice, Guatemala, El Salvador, Honduras, Panamá y Costa Rica).

B. OBJETIVOS ESPECÍFICOS

1. Determinar si la estabilidad de la glucoproteína *spike* del SARS-CoV-2 es afectada por mutaciones en su genoma basándose en escaneo mutacional profundo.
2. Determinar si las mutaciones en la glucoproteína *spike* del SARS-CoV-2 en Centroamérica afectan la formación del complejo RBD-ACE2 a través de análisis de acoplamiento molecular.
3. Evaluar si las mutaciones afectan al panorama de glicosilación ligado a N y O en la glucoproteína *spike* de las secuencias de SARS-CoV-2 provenientes de Centroamérica
4. Analizar la presencia de recombinación genética en las secuencias de SARS-CoV-2 en Centroamérica.

III. Justificación

El SARS-CoV-2 (SC2) es el coronavirus causante de la pandemia actual, y hasta la fecha 5.2 millones de personas han fallecido a causa de esta enfermedad (Center for Disease Control and Prevention, 2020b). Este virus cuenta con un material genético de ARN (pertenece al grupo IV de Baltimore), lo que hace que mute con mayor facilidad que otros virus, dado que la ARN polimerasa dependiente de ARN no puede corregir los errores al azar que comete (Hu *et al.*, 2021). Se estima que el virus cuenta con una tasa de mutación genómica de 3.7×10^{-6} unidad de nucleótidos por ciclo para el linaje con la glucoproteína *spike* (S) descrita originalmente y de 2.9×10^{-6} nt⁻¹ ciclo⁻¹ para un linaje con la mutación D614G (Borges *et al.*, 2021). Las mutaciones pueden ocurrir a lo largo del genoma, pero son de gran importancia aquellas que afectan a S, ya que esta es la encargada de realizar el contacto con la célula humana (mediado por la enzima convertidora de angiotensina 2, ACE2) (Ou *et al.*, 2020). Al afectar la secuencia de dicha proteína se puede llegar a modificar su estabilidad, afectando así la transmisión del virus (Duchene *et al.*, 2020). Las implicaciones que tienen las mutaciones son grandísimas, pudiendo afectar la respuesta del sistema inmune e incluso las estrategias actuales para monitoreo y prevención, ya que gran parte están basadas en esta proteína (Q. Li *et al.*, 2020). Al verse afectada a futuro la respuesta contra el virus, se resalta la importancia de mantener un control y vigilancia de este, así como determinar qué efecto pueden tener sobre S (Eaaswarkhanth *et al.*, 2020).

Además de la diversidad que generan las mutaciones puntuales y las deleciones e inserciones, otra forma en la que la estructura de la proteína puede ser alterada drásticamente es a través de la recombinación genética. La recombinación es el proceso en donde dos moléculas separadas de ADN o ARN intercambian regiones (normalmente homologas) en el genoma (Etherington *et al.*, 2005). Dicho proceso puede producir moléculas quiméricas, expansión de huéspedes virales, mayor virulencia, evasión de la inmunidad del huésped y la evolución de la resistencia a los antivirales. Este mecanismo es un impulso importante en la evolución del virus, por lo que es necesario evaluarlo (Griffiths *et al.*, 2000; Simon-Lorier & Holmes, 2011). Adicionalmente, la proteína también puede ser afectada por modificaciones postraduccionales (MPT), que le confiere otras características (Allison, 2007). De estas, las glicosilaciones son de las MPT más importantes, y se pueden generar nuevos sitios por mutaciones puntuales (Freeze & Ng, 2013).

Además de la escasa cantidad de secuencias provenientes de Centroamérica en los repositorios internacionales, dichas secuencias también han sido poco estudiadas, es por ello por lo que resulta necesario el desarrollo de análisis *in silico* de la información disponible para la región.

Este estudio evaluará a través de métodos bioinformáticos los efectos de las modificaciones mencionadas sobre la proteína S, con un soporte estadístico a partir de la evaluación de la formación del complejo RBD-ACE2 (modelaje molecular), análisis de recombinaciones genética, y escaneo mutacional profundo.

IV. MARCO TEÓRICO

A. SARS-CoV-2

El SC2 es el agente etiológico del COVID-19, que desde el 2019 ha llevado a la muerte de casi 5 millones de personas, dando un aproximado de 226 millones de casos activos. El primer brote fue reportado en noviembre del 2019 en la ciudad de Wuhan, China. Sacudió al mundo y, al ser altamente transmisible, se extendió rápidamente. Ha superado al SARS y al MERS en términos tanto del número de personas infectadas como del rango espacial de las áreas epidémicas (Hu *et al.*, 2021).

Utilizando la secuenciación del ARN metagenómico y el aislamiento del virus, los científicos chinos identificaron el agente etiológico de la enfermedad como un nuevo betacoronavirus. El brote de infecciones de este nuevo virus coincidió con la proximidad del Año Nuevo lunar, por lo que los viajes entre ciudades facilitaron su transmisión en China. El brote alcanzó un pico epidémico en febrero y la propagación internacional de COVID-19 se aceleró. El 11 de marzo de 2020, la OMS caracterizó oficialmente el brote mundial de COVID-19 como una pandemia (Izda *et al.*, 2021). En abril de 2020, Estados Unidos se convirtió en el país más afectado por el virus, seguido de España, Italia, Alemania, Francia y China (Canedo-Marroquín *et al.*, 2020).

El valor del número de reproducción básico (R_0) varía entre los estudios y entre épocas. Esto se da gracias a las distintas mutaciones que pueden ocurrir que llegan a afectar la virulencia del patógeno (Bogoch *et al.*, 2020; Easwarkhanth *et al.*, 2020; Read *et al.*, 2020; Riou & Althaus, 2020). Según distintos estudios, el R_0 del nuevo coronavirus parece estar entre 2-4, lo que indica una alta tasa de infectividad (Read *et al.*, 2020). Este virus presenta una mayor transmisibilidad incluso que otros coronavirus, tales como el SARS-CoV o el MERS-CoV (W. Li *et al.*, 2003). Es por ello por lo que es necesario monitorear a lo largo del tiempo este virus, logrando determinar medidas efectivas contra las distintas variantes de interés.

1. Características clínicas de la COVID-19

El virus SC2 afecta principalmente al sistema respiratorio, aunque también están involucrados otros sistemas orgánicos. Síntomas relacionados con la infección del tracto respiratorio inferior que incluyen fiebre, tos seca y disnea. Además, se observa dolor de cabeza, mareos, mialgias, anosmia, ageusia, debilidad generalizada, vómitos y diarrea (Hu *et al.*, 2021; Izda *et al.*, 2021; Yuki *et al.*, 2020). A pesar de que la mayoría de las personas presentan una sintomatología leve, una infección puede progresar a neumonía severa, fallo multisistémico y la muerte (Sternberg & Naujokat, 2020). Los estudios epidemiológicos han demostrado que la

mortalidad es alta en la población anciana y la incidencia es mucho menor en los niños y se cree que el 30% de los pacientes son verdaderamente asintomáticos (Yuki *et al.*, 2020).

Uno de los primeros síntomas, fiebre, puede persistir hasta 12 días, seguido pronto por disnea y tos (Yuki *et al.*, 2020). Tanto el ácido nucleico viral como el virus con capacidad de replicación son detectables al inicio de los síntomas en frotis nasales. Los títulos de virus con capacidad de replicación disminuyen y la infección viral activa se detiene después de 10 a 15 días (Canedo-Marroquín *et al.*, 2020). Las enfermedades pulmonares más graves, incluido el síndrome de dificultad respiratoria aguda (SDRA) y el SRC, tienden a ocurrir en la segunda semana de infección y pueden requerir soporte ventilatorio mecánico (hasta el 20% de los pacientes). Además, las lesiones cardíacas y renales agudas pueden aparecer en un período de 10 a 20 días (Hu *et al.*, 2021; Yuki *et al.*, 2020).

La edad es un factor determinante que aumenta en gran medida la posibilidad de desarrollar una afección más grave. Al inicio de la pandemia se estimaba que el grupo de riesgo primario eran las personas mayores de 45 años, particularmente aquellas con múltiples condiciones médicas crónicas preexistentes (Center for Disease Control and Prevention, 2020b, 2020a). Sin embargo, informes recientes reportan que las personas mayores a 60 años tienen un riesgo mayor en desarrollar una sintomatología severa (World Health Organization, 2021). Inicialmente, la mayoría de los informes de muertes fueron de este grupo y, a medida que avanza la pandemia, ha sucumbido un número cada vez mayor de adultos jóvenes (20 a 35 años). Esta diferencia entre mortalidad es causada por el sistema inmune del huésped. Mientras la edad avanza, el cuerpo se cansa y las defensas contra patógenos serán más débiles, llegando a facilitar que el virus se reproduzca (Yuki *et al.*, 2020).

Al comienzo del brote, los casos se observaron principalmente entre las personas mayores (65+ años) y, a medida que continuó, el número aumentó, pero también se observó cierto aumento entre los niños (<18 años) (Canedo-Marroquín *et al.*, 2020; Hong *et al.*, 2020). Un estudio con pacientes de 65 años en Wuhan mostró que la mayoría de los pacientes murieron por insuficiencia multiorgánica como insuficiencia respiratoria, shock y SDRA (en el 94%, 81% y 74% de los casos, respectivamente). También se consideraron factores de riesgo la presencia de enfermedad arterial coronaria, diabetes e hipertensión (Chilamakuri & Agarwal, 2021). El sexo parecía ser una diferencia clave, pero a medida que avanzaba la pandemia, no se observó diferencia significativa (Canedo-Marroquín *et al.*, 2020).

Desde el inicio, el porcentaje de niños (que representan el 20% de la población total) dentro del total de pacientes con COVID-19 fue pequeño (Canedo-Marroquín *et al.*, 2020; Viner *et al.*, 2021). Sin embargo, esto puede ser una subestimación de la incidencia real en la población pediátrica si se realizaron menos pruebas en niños debido a menos síntomas (Hong *et al.*, 2020). Dado que los niños a menudo se resfrían, su sistema inmunológico podría estar preparado para brindarles cierta protección contra el COVID-19. También es posible que el sistema inmunológico de los niños interactúe con el virus de manera diferente que el sistema inmunológico de los adultos (Centers for Disease Control and Prevention, 2020). Los adultos se enferman porque su sistema inmunológico reacciona de forma exagerada, causando más daño, lo que puede ser menos probable que ocurra en los niños (Viner *et al.*, 2021).

Además de afectar los pulmones, corazón, riñón, hígado e intestinos, el SC2 llega al sistema nervioso central (SNC), provocando una amplia gama de manifestaciones clínicas (Garg *et al.*, 2021; Yuki *et al.*, 2020). Se ha sugerido que las neuronas son un objetivo potencial de SC2 porque estas células expresan el receptor ACE2. El 36,4% de los pacientes puede presentar varios síntomas neurológicos, como cefalea, encefalitis, encefalopatía, alteración de la conciencia e incluso síndrome de Guillain-Barré (Canedo-Marroquín *et al.*, 2020; Yuki *et al.*, 2020). Las posibles rutas de entrada del virus al SNC son los nervios periféricos y la barrera hematoencefálica (BBB). Se cree que otro coronavirus, el HCoV-OC43, ingresa al SNC a través de la ruta hematogena como un transporte de células inmunitarias, y se sugiere un mecanismo similar para el SC2. La proteína S parece ser capaz de alterar la BBB y existe evidencia de la presencia del virus en el líquido cefalorraquídeo, lo que demuestra que el SC2 puede llegar al SNC (Canedo-Marroquín *et al.*, 2020).

Curiosamente, hallazgos recientes sugieren que, a pesar de la recuperación y el alta del hospital, los pacientes con COVID-19 pueden tener algunas secuelas de salud a largo plazo, como la inducción de diabetes, urticaria, lesiones en la piel y también se ha visto daño vascular en el cerebro, esterilidad en hombres, daño renal y cardíaco (Callard & Perego, 2021; Canedo-Marroquín *et al.*, 2020; Yuki *et al.*, 2020). Esta patología es referida como COVID prolongado. Además, los cambios microestructurales cerebrales que ocurren después de la recuperación explican en parte los informes generalizados de anosmia prolongada y "niebla mental" o dificultad con diversas tareas cognitivas (Izda *et al.*, 2021; National Health Institute, 2021). A medida que avanza la pandemia y se recopilan más datos, se identificarán más y más efectos y consecuencias a largo plazo del COVID-19.

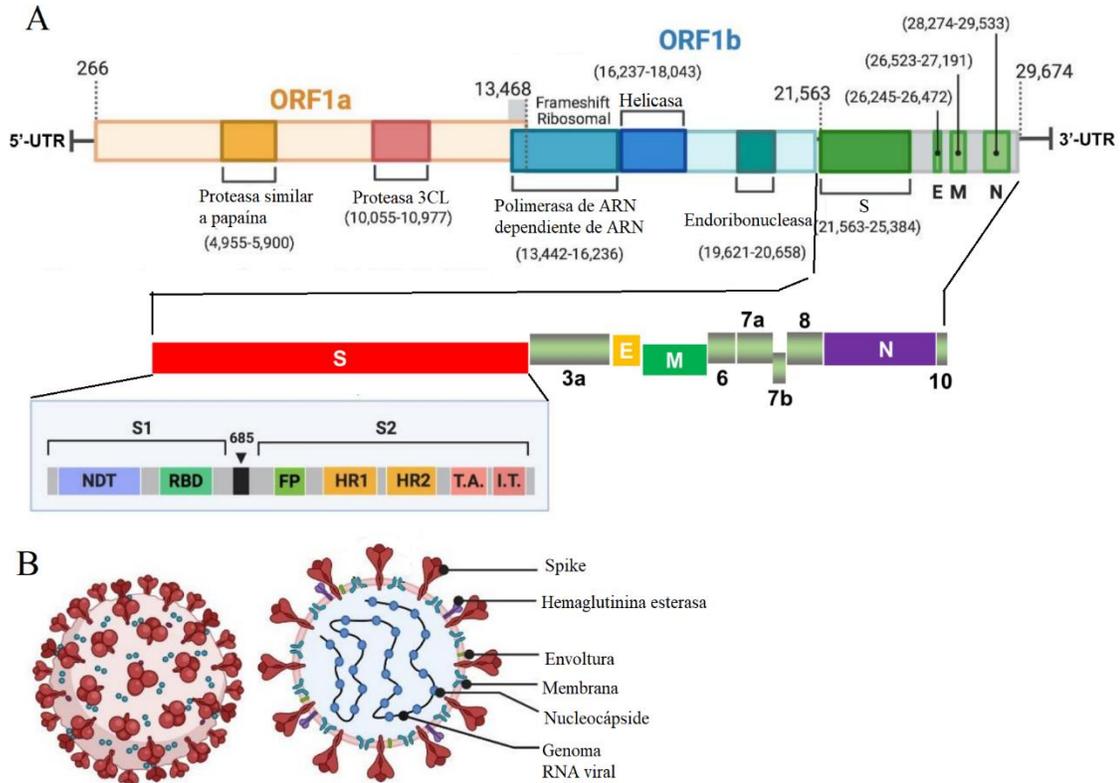
2. Genoma y estructura

El agente etiológico del COVID-19 es un beta coronavirus no segmentado y encapsulado perteneciente al grupo IV de Baltimore (ssRNA+). Cuenta con un genoma de 29881bp (uno de los genomas de ARN monocatenario más grandes) que codifican para 9860aa que dan lugar a proteínas estructurales (S, E, M y N) así como no estructurales (proteasas y RdRp) (Y. Huang *et al.*, 2020). Como es un virus de ARN, que muta con gran facilidad. Se estima que el SARS-CoV-2 tiene una tasa de mutación genómica de 3.7×10^{-6} unidad de nucleótidos por ciclo para el linaje con la proteína S descrita originalmente y de 2.9×10^{-6} nt-1 ciclo- 1 para un linaje con la mutación D614G. La tasa de mutación en el genoma de este virus no es heterogénea, para la proteína S se tiene una tasa de 16×10^{-6} nt-1 por ciclo de infección. Tiene una cabeza de tapa de 5' y una cola de poli (A) de 3'. El extremo 5' del genoma contiene una secuencia líder y una región no traducida (UTR) con varias estructuras de tallo-bucle que son necesarias para replicar y transcribir el ARN. La UTR 3' también incluye estructuras de ARN que son esenciales para la replicación y síntesis del ARN viral (Asghari *et al.*, 2020). Las partículas del virus son de forma esférica o pleomórfica, con un diámetro de aproximadamente 60 a 140 nm. La organización del genoma del SARS-CoV-2 se compone principalmente de marcos de lectura abiertos (ORF) (Chilamakuri & Agarwal, 2021). Aproximadamente el 67% del genoma codifica a ORF1a/b y 16 poliproteínas no estructurales (nsp1-16), mientras que el 33% restante codifica proteínas accesorias (3a, 3b, 6, 7a, 7b, 8a, 8b, 9b, y ORFs) y proteínas estructurales (Figura 1) (Asghari *et al.*, 2020; Chilamakuri & Agarwal, 2021).

a. Proteínas no estructurales y accesorias

Los genomas y subgenomas de coronavirus codifican seis ORF. La mayor parte del extremo 5' está ocupada por ORF1 a/b (Dhama *et al.*, 2020). ORF1a y ORF1b contienen un desplazamiento de marco que produce dos polipéptidos, pp1a y pp1ab. La proteasa similar a la papaína (PL-pro) o la proteasa similar a la quimotripsina (3CLpro) procesan estos dos polipéptidos en 16 proteínas no estructurales (NSPs) (Asghari *et al.*, 2020; Chen *et al.*, 2020). Todos estos NSP tienen funciones específicas (Cuadro 1), aunque algunos NSP no ha sido posible elucidar su función hasta el momento (Chen *et al.*, 2020). El complejo de replicación-transcripción (RTC) se forma en vesículas de doble membrana (DMV) por proteínas no estructurales (nsps), codificadas por el gen de la poliproteína. Posteriormente, el RTC sintetiza un conjunto anidado de ARN subgenómicos (sgRNA) a través de una transcripción discontinua. La terminación de la transcripción y la posterior adquisición de un ARN líder ocurre en las secuencias reguladoras de la transcripción (TRS), ubicadas entre los ORF. Estos ARN subgenómicos de hebra negativa sirven como plantillas para producir ARNm subgenómicos (Chen *et al.*, 2020; Dhama *et al.*, 2020; Hu *et al.*, 2021).

Figura 1: Esquema del genoma viral y estructura del SC2. (A) Diagrama esquemático de la organización genómica de SC2. Las regiones genómicas o los marcos de lectura abierta (ORF), el pico (S), la envoltura (E), la membrana (M) y las proteínas nucleocápsidas (N), así como las proteínas no estructurales traducidas de ORF 1a y ORF 1b y proteínas accesorias. Se indican 3a, 6, 7a, 7b, 8 y 10. (B) Estructura esquemática del virión de SC2 y sus principales proteínas estructurales



(Asghari *et al.*, 2020; N. Wang *et al.*, 2020)

Los virus de ARN tienen tasas de mutación más altas que los virus de ADN, por lo que los genomas de los virus de ARN suelen tener menos de 10 K nucleótidos, pero el tamaño del genoma SC2 es de ~ 30 kb (el más grande entre todos los virus de ARN) (Dhama *et al.*, 2020). El mantenimiento del genoma gigante podría estar relacionado con características específicas del CoV RTC, que contiene varias enzimas de procesamiento de ARN como la exoribonucleasa 3'-5' de nsp14. Esto ha demostrado funcionar como proofreading del RTC (Asghari *et al.*, 2020; Chen *et al.*, 2020).

SC2 también codifica proteínas accesorias 3a, 6, 7a, 7b, 8 y 10 en los ORF, cuyas funciones están brevemente descritas en el Cuadro 1. Se llaman así porque no son esenciales para la replicación o la estructura, sino para la patogenicidad viral o la propagación mediante la

modulación de las vías de señalización del IFN del huésped. Realmente no son estudiadas (a pesar de que pueden llegar a funcionar como biomarcador) (Michel *et al.*, 2020).

b. Proteínas estructurales

Dentro de su genoma, SC2 codifica cuatro proteínas estructurales: proteína *spike* (S), proteína de membrana (M), proteína de envoltura (E) y proteína de nucleocápside (N). Estas proteínas estructurales están codificadas por genes S, M, E, N en los ORF 10 y 11 en un tercio del genoma

Cuadro 1: Función de las proteínas accesorias del SC2. Estas no forman parte de la estructura del virus, en su mayoría tienen una función de defensa o evasión contra el sistema inmune del huésped.

Proteína accesoria	Función	Cita(s)
3a	Canal iónico, componente estructural y puede provocar la activación de NF-κB y JNK	(Kanzawa <i>et al.</i> , 2006)
6	Se asocia con las membranas celulares, mejora la replicación viral, la evasión inmune	(Tangudu <i>et al.</i> , 2007)
7a	Induce apoptosis dependiente de caspasa y vías de apoptosis independientes de MAPK, proteína estructural menor, interacción virus-huésped	(Fielding <i>et al.</i> , 2006; C. Huang <i>et al.</i> , 2006; Pekosz <i>et al.</i> , 2006)
7b	Puede tener un rol en la replicación	(Pekosz <i>et al.</i> , 2006)
8	Media actividades únicas de evasión y supresión inmunológica	(Flower <i>et al.</i> , 2021)
10	No comparte similitud de secuencia con ninguna otra proteína conocida y no puede asociarse con una función.	(Pancer <i>et al.</i> , 2020)

cerca del extremo 3'. Estas proteínas estructurales maduras son responsables del mantenimiento y la replicación viral. El origen del nombre coronavirus está dado por la proteína S, el cual se encuentra en la superficie del virus y contribuye a formar una corona (Asghari *et al.*, 2020). Los virus se nombran en función de su estructura genética para facilitar el desarrollo de pruebas de diagnóstico, vacunas y medicamentos. Los virólogos y la comunidad científica en general hacen este trabajo, por lo que los virus son nombrados por el Comité Internacional de Taxonomía de Virus (ICTV). En febrero de 2020, la ICTV anunció que "coronavirus 2 del síndrome respiratorio agudo severo (SARS-CoV-2)" como el nombre del nuevo virus que causa el COVID-19. Se eligió este

nombre porque el virus está genéticamente relacionado con el coronavirus responsable del brote de SARS de 2003 (World Health Organization, 2020).

La proteína S es una proteína transmembrana viral de clase I de 150 kDa con 1237 aminoácidos. Estos actúan como proteínas homotrímeras y secuencias señal N-terminal (NT) para 20 glucanos ligados a asparagina en el retículo endoplásmico (ER) (Dhama *et al.*, 2020). S está anclada a la envoltura del virus para unir los receptores del coronavirus y la internalización. Está dividida en dos subunidades, S1 y S2. El primero, S1, ayuda en la unión del receptor del huésped, mientras que el segundo, S2, representa la fusión (Chilamakuri & Agarwal, 2021). S2 está compuesta por estructuras en forma de bobinas enrolladas que constan de tres hélices α etiquetadas como repetición de heptada (HR1), además de tres cadenas conocidas como HR2. La complejidad entre HR1 y HR2 los hace resistentes a las proteasas celulares durante la posfusión (N. Wang *et al.*, 2020). Funcionalmente, S es necesaria para la entrada de las partículas del virión infeccioso en la célula a través de la interacción con varios receptores celulares del huésped (Chilamakuri & Agarwal, 2021). HR1 y HR2 están compuestos por un heptapéptido repetitivo: HPPHCPC, donde H es un residuo hidrofóbico o tradicionalmente voluminoso, P es un residuo polar o hidrofílico y C es otro residuo cargado. HR1 y HR2 forman el haz de seis hélices (6-HB), que es esencial para la fusión viral y la función de entrada de la subunidad S2 (Bosch *et al.*, 2003). HR1 está ubicado en el extremo C de un FP hidrófobo y HR2 está ubicado en el extremo N del dominio TM. El dominio TM downstream ancla la proteína S a la membrana viral y la subunidad S2 termina en una cola CT (Duan *et al.*, 2020).

La proteína E tiene una longitud de 74 a 109 aminoácidos y de 8,4 a 109 kDa y desempeña un papel multifuncional en la patogénesis, el ensamblaje y la liberación del virus (M.-Y. Wang *et al.*, 2020). Es un pequeño polipéptido integral de membrana que actúa como viroporina (canal iónico). La inactivación o ausencia de esta proteína está relacionada con la virulencia alterada de los coronavirus debido a cambios en la morfología y el tropismo (Asghari *et al.*, 2020). La proteína E consta de tres dominios: un extremo amino hidrófilo corto, un dominio transmembrana hidrófobo grande y un dominio C-terminal eficaz. Tiene un residuo de prolina conservado que está protegido en la región CT del motivo β -espiral- β (Asghari *et al.*, 2020; Chen *et al.*, 2020; Hu *et al.*, 2021). Desempeña un papel clave en la maduración de la proteína en el aparato de Golgi. También hay un motivo de unión a PDZ (PBM) en la región CT, que participa en la patogenicidad al interferir en la señalización celular. Estas características estimulan una tormenta de citocinas, desencadenan el inflammasoma y, posteriormente, aumentan el edema en los pulmones. Este proceso causa el

síndrome de dificultad respiratoria aguda y la muerte en las infecciones por SARS-CoV1 y SARS-CoV-2 (C. Huang *et al.*, 2011).

Cuadro 2: Función de las 16 proteínas no estructurales codificadas por el genoma viral del SC2

Proteína no estructural	Función	Cita
NSP1	Inhibe la degradación celular del ARNm, inhibiendo la señalización de IFN	(Graham <i>et al.</i> , 2005)
NSP2	Desconocido	(Lei <i>et al.</i> , 2018)
NSP3	PLP, escisión de polipéptidos, bloqueo de la respuesta inmune innata del huésped, promoción de la expresión de citocinas	(Gadlage <i>et al.</i> , 2010)
NSP4	Formación de DMV	(Zhu <i>et al.</i> , 2017)
NSP5	3CLpro, Mpro, escisión de polipéptidos, inhibición de la señalización de IFN	(Cottam <i>et al.</i> , 2014)
NSP6	Restricción de la expansión de autofagosomas, formación de DMV	(Zhai <i>et al.</i> , 2005)
NSP7	Cofactor con nsp8 y 12	(Graham <i>et al.</i> , 2005)
NSP8	Cofactor con nsp7 y 12, primasa	
NSP9	Dimerización y unión de ARN	(Fang <i>et al.</i> , 2008)
NSP10	Proteína de andamio para nsp14 y nsp16	(Jia <i>et al.</i> , 2019)
NSP11	Desconocido	
NSP12	RdRp dependiente de cebador	(Bhardwaj <i>et al.</i> , 2006)
NSP13	ARN helicasa, 5' trifosfatasa	
NSP14	Exoribonucleasa N7-MTase	(Shi <i>et al.</i> , 2019)
NSP15	Endoribonucleasa, evasión de sensores de dsRNA	
NSP16	2'-O-MTase; evitando el reconocimiento de MDA5, regulando negativamente la inmunidad innata	(Arndt <i>et al.</i> , 2010)

La proteína M es una glicoproteína transmembrana tipo III, y la proteína viral más abundante presente en la partícula del virión, dando una forma definida a la envoltura viral. M contiene ~230 aminoácidos de longitud con 25 a 35 kDa, lo que la convierte en la más pequeña de todas las proteínas estructurales (M.-Y. Wang *et al.*, 2020). Se une a la nucleocápside y actúa como organizador central del ensamblaje del coronavirus, está compuesta por tres dominios transmembrana flanqueados por un extremo amino corto fuera del virión y un extremo carboxilo

largo dentro del virión (Synowiec *et al.*, 2021a). Se puede suponer que esta proteína juega un papel en la entrada viral del virus en la célula huésped y la maduración de los virus ARN, pero se necesitan más estudios (Chen *et al.*, 2020; Synowiec *et al.*, 2021a). En general, el armazón viral se mantiene mediante la interacción M-M; los dominios NT y C-terminal (se une a ARN) son dominio ecto y endo, respectivamente. Esta proteína es traducida por ribosomas, que están unidos al ER rugoso (co-traduccional). La glicosilación ligada a N en α , δ y la glicosilación ligada a O en β se puede encontrar en esta glicoproteína, lo que conduce a un papel antigénico (Chen *et al.*, 2020; Masotti *et al.*, 2009).

La proteína N tiene 43 a 50 kD y es una proteína estructural de la nucleocápside helicoidal que puede unirse al ARN con varios aminoácidos lisina y arginina. Esta proteína contiene NTD (sitio de unión al ARN o una región enlazadora [LKR]), CTD (dominio de dimerización) y carece de una estructura tridimensional fija u ordenada (Chen *et al.*, 2020). El LKR cargado es rico en serina y arginina y es responsable de la señalización celular, el antagonismo de los interferones (IFN) y la interferencia del ARN (Asghari *et al.*, 2020; Samudrala *et al.*, 2020). La evidencia sugiere que puede acumularse en el citoplasma de las células contaminadas, provocando una detención del ciclo celular en la fase G2 / M. Esta proteína es multipropósito, une y ensambla el genoma del ARN viral a una estructura o matriz de nucleocápside helicoidal larga de la ribonucleoproteína. También facilita la interacción de la proteína M necesaria durante el ensamblaje del virión y mejora la eficiencia de transcripción del virus. Además, es rico en serina y arginina para la fosforilación en posiciones específicas, lo que puede alterar funciones tales como distinguir el ARN viral y no viral, y alterar la unión del anticuerpo monoclonal a la superficie del virus, además de la maduración y ensamblaje del virus (Canedo-Marroquín *et al.*, 2020).

3. Respuesta inmune del huésped y SARS-CoV-2

SC2 es un virus que no solo puede llegar a invadir las células del sistema inmune, sino utilizarlas contra el mismo. Estas características pueden ser medidas por distintos marcadores moleculares (como lo es la elevada concentración de IL-6). Sin embargo, además de observarse una elevada producción de citoquinas y quimioquinas, también se tiene una linfopenia. En conjunto, estos dos llevan a una respuesta exacerbada, pero corta que puede llegar a conducir a la muerte (M.-Y. Wang *et al.*, 2020).

a. Respuesta inmune innata

El sistema inmunológico humano utiliza receptores de reconocimiento de patrones (PRR) para reconocer patrones moleculares asociados a patógenos (PAMP) de microorganismos (Canedo-

Marroquín *et al.*, 2020; Prompetchara *et al.*, 2020; Rabi *et al.*, 2020). Cuando se produce la fusión viral, el genoma viral, así como las proteínas virales, pueden ser reconocidos por varios PRR. Los principales receptores para este trabajo son los toll-like (TLR) 7 y 8, que desencadenan la respuesta primaria de diferenciación mielóide (MyD88) y la vía TLR4-MyD88. Esto conduce a la activación de los factores de transcripción NF- κ B y la vía de las proteínas quinasas activadas por patógenos (MAPK) para inducir proteínas proinflamatorias (Canedo-Marroquín *et al.*, 2020; Portela Sousa & Brites, 2020). Todas estas vías promueven la expresión de los IFN de tipo I y III (Canedo-Marroquín *et al.*, 2020; Prompetchara *et al.*, 2020). Alternativamente, la activación de receptores endosomales como TLR-3 y TLR que podrían reconocer el genoma viral conduce al reclutamiento de la proteína adaptadora TRIF directamente. Posteriormente, TRIF activa los factores de transcripción IRF3 y NF- κ B para inducir citocinas proinflamatorias como el interferón- α y el TNF- β (Synowiec *et al.*, 2021b). El INF tipo I formará complejos con sus receptores, IFNAR, y posteriormente activará las vías JAK-STAT. Las quinasas JAK1 y TYK2 fosforilan aún más STAT 1/2 una vez que esto se une con IRF9, y juntas migran al núcleo para iniciar la transcripción de genes estimulados por IFN (ISG). El resultado es una supresión de la replicación viral y previene la gravedad de la enfermedad (Astuti & Ysrafil, 2020). Sin embargo, liberación excesiva de citocinas proinflamatorias como IFN- α , IFN- γ , IL-1 β , IL-6, IL-12, IL-18, IL-33, TNF- α , TGF β y quimiocinas CCL2, CCL3, CCL5, CXCL8, CXCL9, CXCL10 por parte de las células efectoras inmunes causa hiperinflamación que eventualmente conducirá al síndrome de dificultad respiratoria aguda (SDRA) (Astuti & Ysrafil, 2020; Canedo-Marroquín *et al.*, 2020). También se ha demostrado que SC2 puede causar una respuesta deficiente al IFN y una disminución en la secreción de IFN-g (asociado con un aumento de la mortalidad del paciente debido al deterioro de la respuesta inmune antiviral) (Canedo-Marroquín *et al.*, 2020). Aunque la vía TLR4 implica la activación de la quinasa de unión a tanque 1 (TBK1), los estudios han demostrado que esta quinasa no se activa durante una infección *in vitro*. Lo que esto da como resultado es la inhibición de la vía del IFN (Synowiec *et al.*, 2021b). Adicionalmente SC2 es capaz de activar el sistema del complemento dando lugar a depósitos del complejo de ataque a la membrana (C5b-9), C3, C4 y serina proteasa 2 asociada a lectina de unión a manosa (MBL) (MASP2) (Prompetchara *et al.*, 2020).

Los estudios hematológicos han demostrado un aumento en la cantidad de neutrófilos en la sangre periférica, niveles bajos de eosinófilos y menor presencia de células asesinas naturales (NK) (Walls *et al.*, 2020). Esto podría estar asociado con niveles más altos de ligando de quimiocinas (CXCL) 9 y 16 (Canedo-Marroquín *et al.*, 2020). Anteriormente se ha demostrado que el virus infecta a los macrófagos alveolares, así como a los macrófagos positivos que expresan los

receptores para ACE2 y CD68, e induce la secreción de interleucina-6 (IL-6) (Astuti & Ysrafil, 2020; Canedo-Marroquín *et al.*, 2020). Como mencionado anteriormente, también es capaz de activar a los macrófagos y las células dendríticas para que liberen citocinas proinflamatorias, lo que conduce a una hiperinflamación sistémica ("tormenta de citocinas"). Esto se asocia con un pronóstico negativo, incluida la muerte. En casos graves consiste en la reducción de las células T y la difusión del daño de las vías respiratorias debido a la infiltración de células inmunes y la formación de membranas hialinas (Astuti & Ysrafil, 2020; Canedo-Marroquín *et al.*, 2020; Y. Yi *et al.*, 2020). Los casos graves de COVID-19 muestran niveles aumentados de varias citocinas y quimiocinas, como IL-2, IL-7, factor estimulante de colonias de granulocitos (G-CSF), proteína 10 inducida por interferón gamma (IP-10), proteína quimioatrayente de monocitos. 1 (MCP-1), proteína inflamatoria de macrófagos 1A (MIP-1A), proteína quimiotáctica de monocitos-3 (MCP-3), factor de necrosis tumoral alfa (TNF- α), IL-6 e IL-1RA. Además, los niveles de IL-6 se correlacionan directamente con las cargas virales y se detectan niveles aumentados antes de la intubación para ventilación mecánica. Los niveles de IL-6 están relacionados con niveles elevados de IP-10, MCP-3 e IL-1RA y están asociados con desenlaces fatales (Shah *et al.*, 2020).

b. Respuesta inmune adaptativa

Al inicio de la infección por SARS-CoV, las células B provocan una respuesta inicial contra la proteína N, mientras que los anticuerpos contra la proteína S podrían detectarse después de 4 a 8 días desde la aparición de los síntomas iniciales (Guo *et al.*, 2020; Shah *et al.*, 2020). Aunque la proteína N es más pequeña que la proteína S, es altamente inmunogénica y la ausencia de sitios de glicosilación en ella da como resultado la producción de anticuerpos neutralizantes específicos de N en una etapa temprana de la infección aguda (Iwasaki & Yang, 2020). Estudios han demostrado que los anticuerpos SC2 IgA, IgG e IgM se detectan después de la aparición de los síntomas, pero solo la IgG persiste durante un período más prolongado, mientras que los niveles de IgM comienzan a disminuir después de 3 meses. Un estudio reveló que hay títulos de anticuerpos IgG e IgM más altos en pacientes graves. Sin embargo, los respondedores débiles para el anticuerpo IgG tuvieron un aclaramiento viral más alto que los respondedores fuertes. Esto sugiere que la respuesta de anticuerpos robusta conduce a la gravedad de la enfermedad, mientras que la respuesta débil se asocia con la eliminación del virus (Shah *et al.*, 2020). Curiosamente, los anticuerpos del SARS-CoV previamente existentes pueden promover la infección viral en las células que expresan FcR y no hay reactividad cruzada de dichos anticuerpos con SC2. Esto mejora la diseminación viral por los macrófagos y, en consecuencia, la inflamación y la lesión tisular por la activación de las células mieloides (Iwasaki & Yang, 2020). Adicionalmente, las células T, tras ser presentados antígenos

virales por el MHCI, se promueve la liberación de citocinas y la actividad citotóxica de las células T CD8 + (Guo *et al.*, 2020). Alternativamente, se encuentra que el MHC de clase II presenta péptidos SC2 a las células T CD4 +. Además, la expresión de MHC se reduce durante la infección debido a modificaciones epigenéticas de moléculas posteriores (Astuti & Ysrafil, 2020). La mayoría de los epítomos de células T presentados por el complejo MHC se derivan de proteínas estructurales como las proteínas S y N del coronavirus, mientras que los NSP tienen efectos reguladores sobre la cascada de señalización. Las células T pueden ser estimuladas por 14 epítomos, la mayoría de los cuales se encuentran localizados en ORF3 y la proteína S (Yang *et al.*, 2020).

En comparación con los niveles elevados en los casos leves, los casos graves presentan subconjuntos más bajos de células T como CD4 +, CD8 + y células T de memoria. Esto indica que después de la infección hay una respuesta antiviral inadecuada en casos graves (Thevarajan *et al.*, 2020). También hay un aumento progresivo de las células auxiliares foliculares T circulantes (cTfh) después de 9 días después de la admisión durante una enfermedad leve y permanecen en la sangre periférica durante la convalecencia, de acuerdo con las células secretoras de anticuerpos (ASC). Estos niveles están ligeramente elevados en comparación con los pacientes sanos, incluso después de 20 días de la aparición de los síntomas. La función de las células Tfh es esencial para el desarrollo de células B de memoria y la producción de anticuerpos de alta afinidad. De acuerdo con esto, se ha encontrado una relación de células B / T más alta en pacientes con COVID-19 (Shah *et al.*, 2020).

Se ha demostrado que los niveles de IgG en pacientes asintomáticos se redujeron en comparación con los sintomáticos, además de una reducción en la actividad de los anticuerpos neutralizantes. Solo el 19,3% de los pacientes muestran un nivel robusto de actividad de neutralización y que disminuyen hasta 28 días después de su recuperación, lo que sugiere que los anticuerpos neutralizantes tienen una vida media corta (Canedo-Marroquín *et al.*, 2020). Un estudio informa que una proporción alta de neutrófilos / linfocitos (NRL) y niveles de IgG conducen a una reacción grave (relacionada positivamente con la muerte) mientras que las proporciones bajas no requieren ventilación mecánica. Los resultados indican que el NRL puede considerarse una medida razonable y útil para definir la gravedad de la enfermedad en los pacientes (Astuti & Ysrafil, 2020).

c. Inmunoevasión del SARS-CoV-2

SC2, como otros patógenos, necesita superar las defensas del huésped. Esa es la única forma en que podría sobrevivir y reproducirse. Durante el proceso de reconocimiento, SC2 puede formar vesículas dobles en el exterior de la célula. Esto provoca el reconocimiento de escudo de los PRR citosólicos a dsRNA como un producto intermedio del virus de replicación (C. Huang *et al.*, 2011).

Además, posee 8 proteínas capaces de esquivar el sistema inmunológico mediante el bloqueo de INF. Nsp1 inactiva la maquinaria de traducción del hospedador, la degradación del ARN-hospedador e inhibe la fosforilación del factor de transcripción STAT1 (Astuti & Ysrafil, 2020). Esto puede llevar a cabo una enfermedad más severa al inducir la replicación y diseminación de virus en una etapa temprana (Bhardwaj *et al.*, 2006; Shi *et al.*, 2019). SC2 tiene una casquete 5' (nucleósido de 7-metilguanina en el extremo 5' del ARNm) menos que el ARN de la célula huésped, por lo que es fácil de reconocer. Para evitar esto, utiliza NSP14 para iniciar la formación de casquete y con NSP16 modifica el casquete de los ARN virales para evitar el reconocimiento de PRR (Fehr *et al.*, 2015).

SC2 también cuenta con NSP3, que codifica macrodominios y PL-pro, que parecen emplearse para evadir la respuesta inmune. En un estudio se utilizaron ratones BALB / c infectados por macrodominios que carecen de SARS-CoV, y reveló un aumento en la expresión de IFN tipo I, ISG15, CXCL10, IL-6 y TNF y ningún desarrollo de patología pulmonar (Lei *et al.*, 2018). PLpro utiliza una actividad de enzima desubiquitinante (DUB) para antagonizar la respuesta de IFN (Nelemans & Kikkert, 2019). ORF3 y ORF6 también pueden ayudar a evadir el sistema inmunológico. El segmento del gen ubicado en ORF3b puede antagonizar la vía de señalización de INF y causar la inhibición de la cascada de activación de células efectoras para la erradicación e inhibición de la replicación viral. La proteína codificada en ORF6 podría inhibir la vía de señalización JAK-STAT al unirse a la carioferina- α 2 y amarrar la carioferina- β 1 en las membranas internas para bloquear la translocación nuclear de STAT1 (Q. Li *et al.*, 2020).

B. GLUCOPROTEÍNA *SPIKE*

1. Qué es la proteína S

En su forma nativa, S está presente como un trímero en la superficie del virión, con las subunidades S1 y S2 formando el tallo extracelular y la "corona" (Walls *et al.*, 2020). La corona sufre cambios conformacionales en forma de bisagra entre una conformación cerrada/ descendente y una conformación abierta/ ascendente menos estable. En la conformación abierta/ ascendente, el RBD es accesible para unirse al receptor ACE2; mientras que, en la conformación cerrada/ descendente, el RBD no puede interactuar con el receptor ACE2 (Cai *et al.*, 2020; Walls *et al.*, 2020; C. Wang *et al.*, 2020). Los picos están recubiertos con moléculas de polisacáridos para camuflarlos, evitando la vigilancia del sistema inmunológico del huésped durante la entrada. La proteína S es un péptido de 1273 aminoácidos compuesta por dos subunidades, S1 (aminoácidos 14-685) y S2 (aminoácidos 686-1273), que son responsables de la unión del receptor y la fusión de la membrana con la célula huésped, respectivamente, precedidas por un péptido señal corto

(aminoácidos 1-13). La subunidad S1 consta de un dominio N-terminal (NTD; aminoácidos 14-305) y un dominio de unión al receptor (RBD; aminoácidos 319-541) (Duan *et al.*, 2020). La subunidad S2 está compuesta por una secuencia de repetición heptapéptido 1 (HR1, 912-984 aa), HR2 (1163-1213 aa), dominio transmembranal (TM, 1213-1237 aa), un péptido de fusión (FP, 788-806 aa) y un dominio citoplásmico (1237-1273 aa) (Bosch *et al.*, 2003; Duan *et al.*, 2020; Pancera *et al.*, 2014; Wrapp *et al.*, 2020). S1 contribuye a la estabilización del estado de prefusión de la subunidad S2 anclada a la membrana que contiene la maquinaria de fusión (Duan *et al.*, 2020). Los trímeros de la proteína S forman un halo bulboso en forma de corona que rodea la partícula viral, en el que las subunidades S1 y S2 forman la región de la cabeza bulbosa y el tallo. En el estado nativo, existe como un precursor inactivo, y durante la infección, las proteasas lo activan por escisión en subunidades S1 y S2 (Cai *et al.*, 2020; Duan *et al.*, 2020).

2. Biosíntesis de la proteína S

La proteína S se sintetiza como precursora en el retículo endoplásmico rugoso (RER). Alberga una secuencia de señal del retículo endoplásmico (RE) en el extremo N, que lo dirige a la membrana RER y es eliminado por las peptidasas de señal celular en el lumen del RE. Una única secuencia que atraviesa la membrana de parada-transferencia en el extremo C evita que se libere completamente en el lumen del RE y la subsecuente secreción de la célula infectada (Bosch *et al.*, 2003; Duan *et al.*, 2020). Co-traduccionalmente durante la síntesis, cadenas laterales de oligosacáridos con alto contenido de manosa (glicosilaciones ligadas a N) (Aebi, 2013). Poco después de la síntesis, los monómeros se trimerizan y, una vez en el complejo de Golgi, la mayoría de las cadenas laterales de oligosacáridos con alto contenido de manosa se modifican a formas más complejas y también se añaden cadenas laterales de oligosacáridos O-unidos (X. Zhang & Wang, 2016).

En la red trans-Golgi, S se escinde proteolíticamente por furina celular o proteasas similares a furina. Después de la escisión, una señal de recuperación de ER (ERRS) en el extremo C terminal asegura que la proteína madura se acumule cerca del compartimento intermedio ER-Golgi (ERGIC) (Duan *et al.*, 2020). Aquí, impulsada por interacciones con otra proteína estructural, M, la proteína S participa en el ensamblaje de las partículas del virus y se incorpora a la envoltura del virus (Bosch *et al.*, 2003). Alternativamente, una fracción de proteínas S maduras viaja a través de la vía secretora hacia la membrana plasmática, donde pueden mediar la fusión de células infectadas con no infectadas para formar células gigantes multinucleadas (sincitios). Este proceso puede permitir la propagación directa entre las células y potencialmente alterar la virulencia del SARS-CoV-2 (Hoffmann *et al.*, 2020).

3. Invasión a la célula del huésped y replicación viral

Como se mencionó anteriormente, SCE se une al receptor ACE2 con S, luego S sufre un cambio conformacional que resulta en la accesibilidad de los sitios de escisión S1 y S2 a las proteasas del huésped. Una gama de diferentes proteasas puede escindir y activar S. SC2 contiene una secuencia "RPPA" en sus subunidades S1 y S2 que introduce sitios de escisión de furina. Además, la proteasa transmembrana serina 2 (TMPRSS2), la tripsina y la catepsina L también pueden activar la proteína S (Samudrala *et al.*, 2020). Los dos pasos para activar S mediante la escisión con proteasas son: (1) escisión en el sitio de escisión S1 / S2 para cebado y (2) escisión en el sitio S'2 (adyacente a un péptido de fusión dentro de S2, formando FP y S2') para activación. Después de (1), las subunidades S1 y S2 permanecen unidas de forma no covalente, y S1 distal contribuye a la estabilización de S2 anclado a la membrana en el estado de prefusión. (2) activa el pico para la fusión de la membrana a través de cambios conformacionales irreversibles. Estos cambios se producen insertando FP en la membrana de la célula diana, exponiendo la bobina en espiral del dominio HR1 y provocando la interacción entre el dominio HR2 y el trímero HR1 para formar 6-HB, lo que acerca la envoltura viral y la membrana celular para fusión y entrada viral (Alejandra Tortorici *et al.*, 2019; Cai *et al.*, 2020; Guillén *et al.*, 2005; Reguera *et al.*, 2014; Sainz *et al.*, 2005; L. Wang & Xiang, 2020).

El ARNm viral, junto con la nucleocápside, ingresa al citoplasma de la célula huésped a través de endocitosis mediada por receptor o directamente a través de la membrana de la célula huésped después de la unión y fusión de S. Como se mencionó anteriormente, pp1a y pp1ab, se escinden para formarse dentro de la membrana de RER el RTC (Chen *et al.*, 2020). El genoma se transcribe en ARN antisentido (3' a 5') con la ayuda de RdRp. Este se replica y se transforma en una hebra positiva de ARNg de longitud completa con la ayuda de la actividad replicasa de la RdRp viral. Más tarde, se utiliza para sintetizar varios ARNm anidados (subgenómicos) de diferentes tamaños a través de un proceso de transcripción discontinuo por RdRp (Y. Huang *et al.*, 2020; L. Wang & Xiang, 2020). La traducción de este ARNm subgenómico da lugar a proteínas estructurales virales. E, M y S se introducen en la membrana ER o la membrana de Golgi. El ARN genómico viral, combinado con N, forma nucleocápsidas. Esto brota en el complejo del compartimento intermedio ER o Golgi. Las vesículas maduras que contienen virión / virus se liberan de la célula por exocitosis o por fusión con la membrana plasmática (Duan *et al.*, 2020; L. Wang & Xiang, 2020).

Algo clave que hace a S tan inmunogénica es que contiene sitios de corte polibásicos (RPPA), lo que contribuye a la alta virulencia de SARS-CoV-2. Las formas de virus de baja patogenicidad, como las del virus de la influenza H1N1, tienen un sitio de escisión por proteasas similares a la

tripsina (Sun *et al.*, 2010). Por el contrario, SC2 tiene un sitio de escisión de furina escindido por proteasas de tipo furina. Las proteasas de tipo tripsina suelen tener una distribución tisular estrecha (TMPRSS11D se expresa solo en el esófago). Por el contrario, las proteasas de tipo furina son ubicuas. Por lo tanto, si SC2 necesitaba ser escindido por TMPRSS11D, entonces su entrada celular se limita al esófago. Al tener un sitio de escisión de la furina, esta restricción se elimina porque esta es ubicua en los tejidos humanos, lo que da como resultado un ensanchamiento espectacular del tropismo de la célula huésped (Coutard *et al.*, 2020). Esto porque las proteasas necesarias para la activación proteolítica están expresadas ubicuamente en humanos, lo que demuestra un tropismo extendido del patógeno (Livesey & Marsh, 2020; Reeb *et al.*, 2020). Adicionalmente, S2 puede interferir con la producción de una respuesta humoral contra el virus, así como con el reclutamiento de anticuerpos neutralizantes. Adicionalmente, los RBDs de S1 contienen glucanos ligados a N y glucanos ligados a O que obstruyen la obtención de anticuerpos neutralizantes tras la exposición inmunitaria (Reeb *et al.*, 2020).

C. EFECTO DE LAS MODIFICACIONES EN PROTEÍNAS

Las modificaciones en las proteínas pueden ocurrir durante varias etapas de la síntesis de esta. En primera instancia, la regulación genética (*splicing*, regulación de transcripción, etc) y modificaciones postraduccionales (MPT) son los mecanismos principales que pueden afectar a la estabilidad y función de una proteína. Ambas son contrarias en cierto sentido. Esto es dado a que la regulación genética ocurre previo a la síntesis de la proteína en sí (va más a controlar el material genético), mientras que las MPT ocurren posterior a la síntesis (Allison, 2007).

Las MPT afectan ya son realizadas para modificar la funcionalidad de la proteína al integrarle distintos grupos funcionales (ThermoFisher Scientific, 2008). A comparación, la regulación genética permite la formación de isoformas proteicas que pueden ser utilizadas en distintas reacciones en distintos tejidos del cuerpo (llegando a tener funciones similares, pero no iguales) (Allison, 2007). Si bien, ambas son importantes en el producto final de la proteína, realmente las que afectan a las proteínas virales serían y propiamente las MPT. Ambas aumentan la complejidad del proteoma, pero las MPT propiamente pueden crear efectos indeseados para los huéspedes virales (ThermoFisher Scientific, 2008). Esto está dado ya que la regulación génica evita la producción de proteínas ajenas (evitar la replicación viral), mientras que las con las MPT, ya existen las proteínas y solo se trabaja sobre ellas (Carter & Saunders, 2007; Uversky, 2013).

Adicional a las MPT y regulación genética, también existe un tercer tipo de modificación que puede ocurrir en una proteína: las mutaciones. Las mutaciones son un mecanismo muy importante en la evolución viral (Allison, 2007; Carter & Saunders, 2007). Estas permiten crear variantes más

patogénicas y/o más infectivas. Las mutaciones pueden ocurrir por algún daño en el genoma (químicos, radiación, etc.) o en la replicación (en el caso de SC2 estaría dado por la RdRp) (Allison, 2007; Hu *et al.*, 2021). Las mutaciones crean formas alternas de la proteína dada, llegando a afectar la estabilidad, llegando incluso a formar péptidos que no son útiles. Sin embargo, estas no son importantes clínicamente, ya que si se dan no habría un efecto sobre la población general. Esto es dado a que ya no sería funcional, inhibiendo la replicación del virus y así evitando que llegue a causar algún efecto. Entonces, no podrían ser apreciadas, pero las que si se observan son las que sí llevan a un producto funcional. Este podría tener una afinidad mayor por un sustrato (o en este caso con un receptor, ACE2), lo que conduciría a una mayor infectividad (Allison, 2007; Carter & Saunders, 2007; Hu *et al.*, 2021).

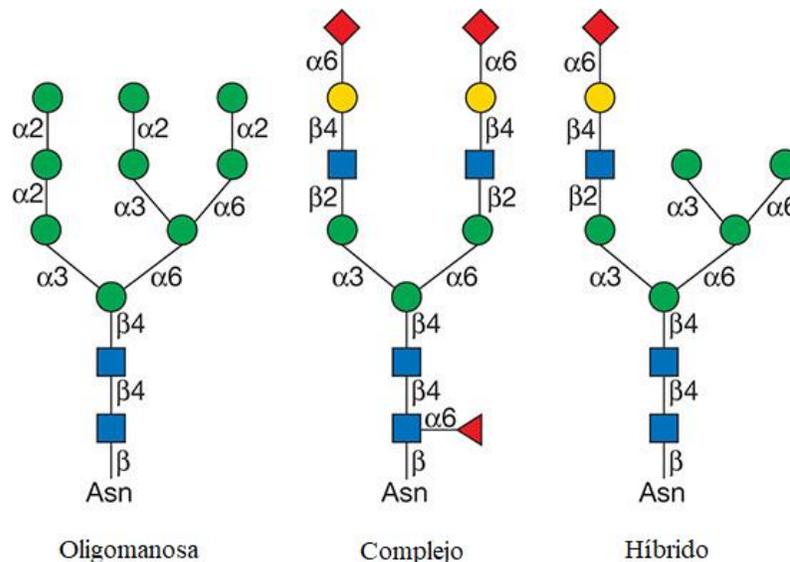
1. Modificaciones postraduccionales

Las modificaciones postraduccionales de las proteínas se refieren a una de las últimas etapas de la biosíntesis de las proteínas. En esta etapa ocurren los cambios químicos reversibles o irreversibles en la proteína. Estos cambios pueden ser los cortes enzimáticos a las uniones peptídicas de ciertos grupos (lípidos, carbohidratos, fosforilación e incluso cadenas de péptidos). Como resultado la proteína tiene distintas propiedades estructurales y funcionales (Y. Huang *et al.*, 2020). Las modificaciones postraduccionales de proteínas (MPT) aumentan la diversidad funcional del proteoma mediante la adición covalente de grupos funcionales o proteínas, la escisión proteolítica de subunidades reguladoras o la degradación de proteínas completas. Estas modificaciones incluyen fosforilación, glicosilación, ubiquitinación, nitrosilación, metilación, acetilación, lipidación y proteólisis e influyen en casi todos los aspectos de la biología y patogénesis celular normal (Samudrala *et al.*, 2020).

Las MPT que serán evaluadas en este estudio son únicamente dos: glicosilaciones ligadas a N y O. Las glicosilaciones ligadas a N se dan cuando un oligosacárido se une a un sequon, una secuencia de aminoácidos consecutivos en una proteína que puede servir como sitio de unión a un polisacárido, frecuentemente un glucano ligado a N. En este caso el sequon sería un nitrógeno del grupo amida de un residuo de asparagina en Asn-X-Ser / Thr (Bettenbaugh *et al.*, 2007; Varki *et al.*, 2017). Este tipo de modificación afecta la estructura y la función del péptido. Todos los N-glucanos eucariotas comparten una secuencia central común, $\text{Man}\alpha 1-3 (\text{Man}\alpha 1-6) \text{Man}\beta 1-4\text{GlcNAc}\beta 1-4\text{GlcNAc}\beta 1 - \text{Asn-X-Ser / Thr}$, y se clasifican en tres tipos: (1) oligomanosa (o también denominados altos en manosa), en la que solo residuos de manosa extienden el núcleo; (2) complejo, en el que "antenas" iniciadas por GlcNAc extienden el núcleo; e (3) híbrido, en el que la manosa extiende el brazo $\text{Man}\alpha 1-6$ del núcleo y uno o dos GlcNAcs extienden el brazo $\text{Man}\alpha 1-3$

(Figura 2). Aproximadamente el 70% de los sequones de Asn-X-Ser / Thr están glicosilados. Ocasionalmente, pero los N-glicanos también pueden aparecer ocasionalmente en Asn-X-Cys y rara vez con un aminoácido diferente en la tercera posición. La identidad de "X" puede reducir la eficacia de la glicosilación, como cuando "X" es ácido (Asp o Glu), o mejorar la eficacia, como cuando Phe está en un giro inverso adyacente. Las glicoproteínas a menudo tienen una gama de N-glicanos diferentes en un período de N-glicosilación de Asn-X-Ser / Thr particular, lo que conduce a la heterogeneidad de los glicanos en cada sitio. Cuando hay más de un sequon por molécula, diferentes moléculas en una población pueden tener diferentes subconjuntos de N-glicanos en diferentes sequones, lo que conduce a una microheterogeneidad de glicoproteínas. La variación en los N-glicanos de una glicoproteína puede deberse a la conformación de la proteína que afecta la disponibilidad de sustrato para las glicosidasas o glicosiltransferasas de Golgi, el metabolismo de los azúcares nucleotídicos, la velocidad de transporte de la glicoproteína a través de la luz del RE y Golgi, y la proximidad de un Asn-X-Sequon Ser / Thr a un dominio transmembrana. Además, la localización de las glicosiltransferasas dentro de los subcompartimientos del Golgi puede determinar qué enzimas encuentran aceptores de N-glicanos (Varki *et al.*, 2017).

Figura 2: Tipos de N-glicanos. Los N-glicanos en las secuencias Asn-X-Ser / Thr en las glicoproteínas eucariotas son de tres tipos generales: oligomanosa, complejo e híbrido. Cada N-glicano contiene el núcleo común Man3GlcNAc2Asn. Los N-glicanos complejos pueden tener hasta seis ramas iniciadas por GlcNAc y cada una puede alargarse con repeticiones de Gal β 1-4GlcNAc (LacNAc).



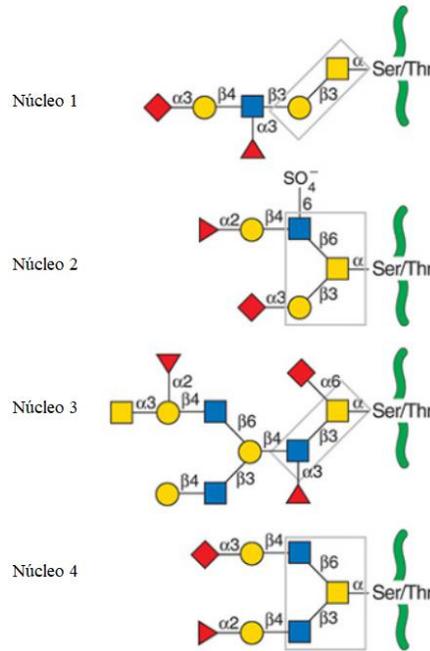
(Varki *et al.*, 2017).

Muchas glicoproteínas llevan glicanos iniciados por GalNAc (N-Acetilgalactosamina) unidos al hidroxilo de residuos de Ser o Thr. Las O-GalNAc están compuestas por azúcar unida a un oxígeno de una serina o treonina. Los azúcares que se encuentran en los glicanos O-GalNAc

incluyen GalNAc, Gal, GlcNAc, Fuc y Sia, mientras que los residuos Man, Glc o Xyl no están representados. La longitud de los glucanos O-GalNAc puede variar desde una única GalNAc hasta más de 20 residuos de azúcar y puede incluir grupos sanguíneos y otros epítomos de glucanos. Las glicoproteínas pueden contener cientos de glucanos O-GalNAc heterogéneos unidos al armazón proteico. GalNAc O unido a Ser / Thr es el azúcar iniciador de los glucanos O-GalNAc y generalmente se extiende para formar una de las cuatro estructuras centrales comunes. Cada núcleo se puede extender posteriormente para dar un glucano O-GalNAc maduro lineal o ramificado. La extensión de O-GalNAc por un β 1-3Gal forma el glucano O-GalNAc más común (núcleo 1). El núcleo 2 se forma mediante la adición de β 1-6GlcNAc a la GalNAc del núcleo 1. El núcleo 3 es menos común y se forma cuando se agrega β 1-3GlcNAc a O-GalNAc. El núcleo 4 consiste en un núcleo 3 ramificado mediante la adición de β 1-6GlcNAc () (Varki *et al.*, 2017). Afectan a la proteína de diferentes maneras al cambiar la estabilidad de la proteína y regular la actividad de la proteína; además de ser importante para muchas enfermedades (Andersen *et al.*, 2020). Los N-glucanos afectan muchas propiedades de las glicoproteínas, incluida su conformación, solubilidad, antigenicidad, actividad y reconocimiento por las proteínas de unión a glucanos. La introducción de un sitio de N-glucano se usa como método para localizar u orientar una glicoproteína o para seguir su movimiento a través de la célula (Varki *et al.*, 2017).

El proteoma es dinámico y cambia en respuesta a los estímulos, y las MPT se emplean comúnmente para regular la actividad celular y pueden ocurrir en cualquier momento de la vida de la proteína. La mayoría de las veces están mediadas por actividad enzimática (Pang & Wilkins, 2018). Se estima que el 5% del proteoma comprende enzimas que realizan más de 200 tipos de MPT. Estas enzimas incluyen quinasas, fosfatasas, transferasas y ligasas. Muchas proteínas también pueden modificarse a sí mismas utilizando dominios autocatalíticos, como los dominios autocinasa y autoprotolíticos. Se sabe que todas las cadenas laterales de aminoácidos experimentan una diversificación química debido a varios MPT (Korber *et al.*, 2020). Sin embargo, la mayoría de las veces, las proteínas MPT se encuentran en las cadenas laterales que pueden actuar como nucleófilos fuertes (C, M, S, T, Y, K, H, R, D, E) o débiles (N, Q), mientras que los residuos restantes (P, G, L, I, V, A, W, F) rara vez participan en modificaciones covalentes de sus cadenas laterales (Uversky, 2013).

Figura 3: Núcleos de glucanos Tipos de O-GalNAc. Unido a las líneas verdes (proteínas) hay glucanos O-GalNAc complejos con diferentes núcleos identificados por recuadros grises. Las cuatro estructuras centrales (en recuadros) pueden extenderse, ramificarse y terminarse con determinantes antigénicos Fuc, Sia o del grupo sanguíneo.



(Varki *et al.*, 2017).

Los MPT juegan un papel fundamental en la regulación del plegamiento de proteínas, su orientación a compartimentos subcelulares específicos, su interacción con ligandos u otras proteínas y su estado funcional, como la actividad catalítica. Algunas proteínas requieren múltiples tipos distintos de MPT para su función. La glicosilación y fosforilación [y otras MPT] combinan modificaciones que están asociadas principalmente con proteínas y regiones intrínsecamente desordenadas (es decir, proteínas o regiones biológicamente activas que no poseen una estructura espacial única) (ThermoFisher Scientific, 2008; Uversky, 2013). Se basan en las interacciones de unión de baja afinidad y alta especificidad entre una enzima específica y un sustrato. Las quinasas ilustran la importancia del trastorno intrínseco para la acción de las enzimas que catalizan diferentes PTM. Se estima que las funciones de un tercio de las proteínas eucariotas se controlan mediante ciclos de fosforilación-desfosforilación. Los sitios de fosforilación se encuentran frecuentemente dentro de dominios proteicos funcionalmente importantes (Pang & Wilkins, 2018).

En el sitio de unión de la subunidad S1 y S2 hay un sitio de escisión polibásico, lo que facilita una escisión eficaz de la proteína S por furina y otras proteasas y conduce a la adición de glucanos unidos a O a S673, T678 y S686, que flanquean este sitio (Korber *et al.*, 2020; Q. Li *et al.*, 2020). Adicionalmente, un estudio reciente demostró la presencia de nuevas glicosilaciones ligadas a O

en los sitios Thr323 y Ser325 (Shajahan *et al.*, 2020). Cada monómero de la proteína S está fuertemente glicosilado y juega un papel crítico en la provocación de la respuesta inmune del huésped. S1 y S2 juntos poseen 22 sitios potenciales de N-glicosilación. Los N-glucanos de la proteína S juegan un papel importante en el plegamiento y cebado adecuados de la proteína por las proteasas del huésped. Como los glucanos pueden proteger los residuos de aminoácidos y otros epítomos del reconocimiento de células y anticuerpos, la glicosilación puede permitir que el coronavirus evite las respuestas inmunes tanto innatas como adaptativas (Sanda *et al.*, 2021; X.-Y. Zhang *et al.*, 2020). Los N-glucanos de esta proteína de tipo complejo están presentes en las posiciones 17, 74, 149, 165, 282, 331, 343, 616, 657, 1098, 1134, 1158, 1173 y 1194; los híbridos en 61, 122, 603, 717, 801 y 1074; y los que tienen alto contenido de manosa en 234 y 709 (Shajahan *et al.*, 2020).

2. Recombinación genética

La recombinación es el proceso donde dos moléculas separadas de ADN o ARN intercambian regiones (normalmente homologas) en el genoma. Esto hace que sea un mecanismo muy importante que impulsa el cambio evolutivo (Andersen *et al.*, 2020; Plante *et al.*, 2020). En virus la recombinación lleva a la creación de moléculas quiméricas a partir de genomas parentales de diferente origen. La recombinación es un mecanismo que se asocia con la expansión de la gama de huéspedes virales, el aumento de la virulencia, la evasión de la inmunidad del huésped y la evolución de la resistencia a los antivirales (Fratev, 2020).

Es difícil de medir en la actual pandemia dado a que hay muy poca variación entre variantes. Además, para que esto ocurra tiene que haber una infección con más de un virus diferente y estas tienen que ser suficientemente distintas para que se pueda detectar (Fratev, 2020). Además, las tasas de recombinación varían drásticamente entre los virus de ARN, que pueden variar desde clonales hasta altamente recombinogénicos (Walls *et al.*, 2020). Una vez ingresado el material genómico a la célula huésped, el ARN viral se libera en el citoplasma y el ORF1 se traduce para producir RdRp. Esto conlleva que se produzcan ARNm subgenómicos por transcripción discontinua, que es un proceso característico de esta RdRp y que favorece la recombinación genética (Ortega *et al.*, 2020). En otras palabras, se crean genotipos quiméricos entre dos cepas virales que infectan la misma célula. La RdRp deja de replicar prematuramente el primer genotipo antes de reensamblar y reanuda la replicación con el segundo genotipo como plantilla. El resultado final es la desvinculación de mutaciones en todo el genoma, creando nuevas combinaciones de mutaciones existentes. La relevancia clínica y epidemiológica de estas nuevas combinaciones es

sustancial, ya que tienen el potencial de crear genotipos con características únicas de virulencia y transmisibilidad (VanInsberghe *et al.*, 2021).

Las mediciones de la frecuencia de recombinación entre coronavirus en cultivo celular sugieren que es muy común (LR & MM, 1991). Además, ha habido intentos para poder detectar y medir la magnitud de la recombinación entre los genomas del SARS-CoV-2 que circulan naturalmente. Huiguang Yi de la Universidad de Ciencia y Tecnología del Sur se basó en cuatro polimorfismos de un solo nucleótido (SNP) para poder detectar genomas recombinantes entre los primeros 85 genomas secuenciados del SARS-CoV-2 y fue capaz de demostrar la capacidad de este virus por recombinar (H. Yi, 2020). Recientemente se demostró que la exoribonucleasa con capacidad correctora de coronavirus (nsp14-ExoN) es necesaria para mantener las tasas y los loci de recombinación generados durante la infección. Estos datos sugieren que la recombinación y la exoribonucleasa del coronavirus se conservan y son determinantes importantes de la replicación que pueden ser objeto de inhibición y atenuación (Gribble *et al.*, 2021).

A inicios de la pandemia, Xiaojun Li y compañía mostraron evidencia que demuestra que todo el RBD del SARS-CoV-2 se introdujo mediante la recombinación con coronavirus de pangolines, posiblemente un paso crítico en la evolución de la capacidad del SARS-CoV-2 para infectar a los humanos. Una selección de purificación similar en diferentes especies de huéspedes, junto con la recombinación frecuente entre coronavirus, sugiere un mecanismo evolutivo común que podría conducir a nuevos coronavirus humanos emergentes (X. Li *et al.*, 2020a). Un estudio realizado a principios del 2021 demuestra que existen genotipos recombinantes y que contienen sustituciones preocupantes por una transmisibilidad elevada y una menor eficacia de la vacuna, incluidos D614G, N501Y, E484K y L452R. Aunque, el análisis filogenético no reveló ninguna evidencia de *hotspots* de recombinación en el genoma del SARS-CoV-2. Ellos estimaron que como máximo el 5% de los virus circulantes en los EE. UU. Y el Reino Unido son recombinantes (VanInsberghe *et al.*, 2021). El muestreo de 279,000 virus del SARS-CoV-2 de finales de 2020 y principios de 2021 en el Reino Unido se utilizó como evidencia de múltiples orígenes independientes de genomas recombinantes. En cuatro casos, hubo evidencia de transmisión progresiva de un virus de origen recombinante, incluido un grupo de transmisión de 45 casos secuenciados en el transcurso de 2 meses. Las ubicaciones genómicas inferidas de los puntos de corte de la recombinación sugieren que cada virus recombinante transmitido por la comunidad heredó su región de pico de un virus parental B.1.1.7, de acuerdo con una ventaja de transmisión para el conjunto de mutaciones de B.1.1.7 (Jackson *et al.*, 2021). Un estudio más reciente, realizado en agosto de 2021, utilizó un árbol de filogenia con una muestra de 1,6 millones e identificó 606 eventos de recombinación. Es decir, aproximadamente

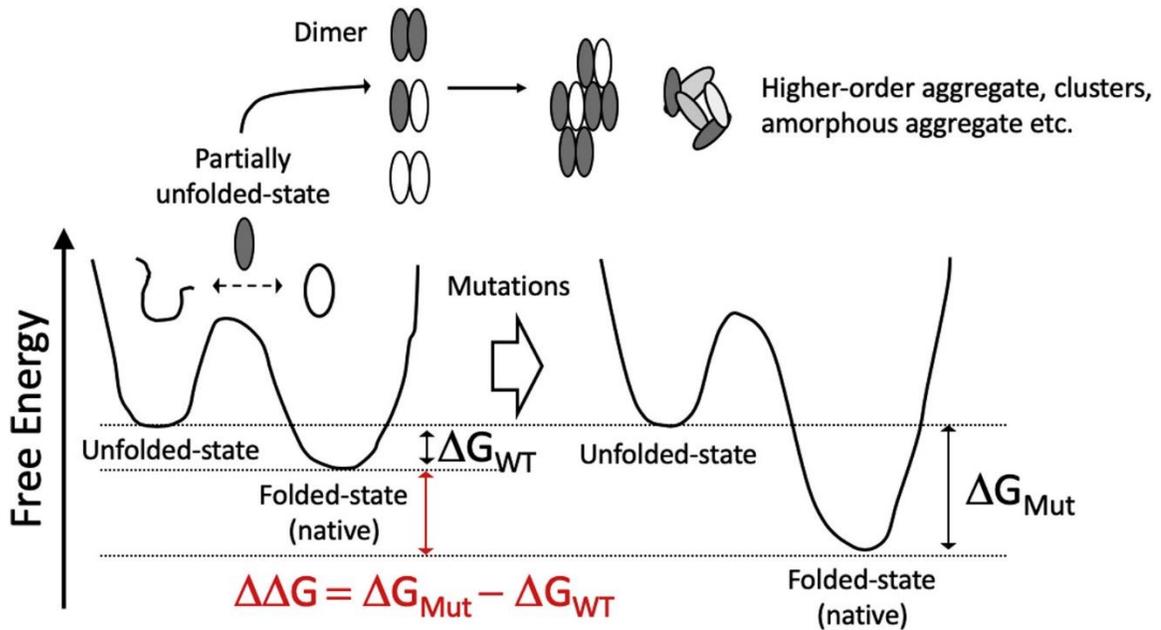
el 2,7% de los genomas secuenciados del SARS-CoV-2 [en ese momento] tienen ascendencia recombinante (Turkahia *et al.*, 2021). En conjunto la evidencia demuestra la capacidad de este virus de poder recombinar y, aunque presente en niveles bajos, si es una amenaza real que impulsa la evolución viral.

3. Efecto de los cambios en la estabilidad de una proteína en su función

La estabilidad es el potencial de un patrón para sobrevivir en el tiempo y, por lo tanto, es parte integral de la comprensión de los sistemas biológicos y su evolución. Los cambios en el "patrón" de una molécula de proteína incluyen procesos como el despliegue, desnaturalización, degradación, cambio conformacional, modificación enzimática y escisión proteolítica de la proteína. Estas transformaciones se analizan con respecto a la integridad de la estructura primaria y conformacional de la proteína completamente plegada. En biología estructural, los cambios en la estructura primaria, secundaria, terciaria o cuaternaria de una proteína son una medida de su estabilidad (Deller *et al.*, 2016). Ahora, la estabilidad conformacional de las proteínas se define como la diferencia de energía libre (ΔG) entre los estados plegado y desplegado, es decir, $\Delta G = G_{Desplegado} - G_{Plegado}$. Para estabilizar una proteína, uno debe estabilizar el estado plegado o desestabilizar el estado desplegado de la proteína desplazando el equilibrio plegado-desplegado hacia la estabilidad de plegamiento. Por otro lado, el cambio de energía libre causado por mutaciones ($\Delta\Delta G$), es decir $\Delta\Delta G = \Delta G_{mutante} - \Delta G_{proteina\ WT}$, es el cambio causado en la estabilidad conformacional de una proteína con mutaciones y una *WildType* (WT) (Figura 4) (Kazlauskas, 2018; Kuroda & Tsumoto, 2020).

Los PTM pueden modificar la estructura primaria de la proteína de varias formas, lo que da como resultado la alteración de la estructura y función de una proteína. Estas pueden afectar tanto a la estabilidad de la composición, ya que las modificaciones pueden ser no uniformes o incompletas, como a la estabilidad conformacional, ya que las modificaciones pueden afectar el desorden y la dinámica de las proteínas (Deller *et al.*, 2016; Freeze & Ng, 2013; Varki *et al.*, 2017). Un claro ejemplo de esto son las glicoproteínas, como S, que a menudo no están glicosiladas de manera uniforme en todos los posibles sitios de glicosilación, lo que conduce a una heterogeneidad de composición (Varki *et al.*, 2017). Además, las cadenas de hidrocarburos complejas tienden a tener un mayor grado de libertad conformacional. Esta libertad conformacional da como resultado un aumento del desorden en la superficie de la proteína, mientras que al mismo tiempo protege los residuos polares o cargados en la superficie de la proteína (L. Wang & Xiang, 2020). Aunque la heterogeneidad de la glicosilación tiende a perjudicar la cristalización, su variabilidad puede tener importantes implicaciones funcionales para una proteína (Deller *et al.*, 2016).

Figura 4: Equilibrio entre proteínas plegadas y desplegadas y su relación con la estabilidad proteica en proteínas mutadas y tipo WT (Kuroda & Tsumoto, 2020).



La estructura secundaria de la proteína (SS) es la estructura tridimensional localizada de la cadena polipeptídica. Este nivel se describe por el enlace de hidrógeno entre los átomos de amida H y los átomos de carbonilo O y por la estereoquímica adoptada por la cadena principal (Nick Pace *et al.*, 2014). La hélice α está presente predominantemente en la estructura SS, lo que hace \sim un tercio de todos los SS. Un residuo específico, la prolina, tiene un efecto desestabilizador sobre la hélice α . Esta desestabilización es el resultado de la falta del átomo de H de la amida de la columna vertebral, lo que evita que la prolina participe en la estabilización de los enlaces de hidrógeno. La glicina tiene la siguiente propensión más baja a formar hélices α debido a la flexibilidad conformacional mejorada al plegarse para formar una hélice α (Deller *et al.*, 2016; Nick Pace *et al.*, 2014; Sticke *et al.*, 1992). Es importante señalar que muchas de estas propensiones de estructura secundaria dependen en gran medida del contexto. Esto respalda los hallazgos de que la estabilidad de la proteína plegada viene dictada por la composición de aminoácidos (Deller *et al.*, 2016).

La estructura terciaria es básicamente la forma, o pliegue, adoptado por la cadena polipeptídica. Muchos factores afectan el proceso de plegamiento de proteínas, incluida la estabilidad conformacional y composicional, temperatura y pH, estructura primaria y secundaria, solvatación, enlaces de hidrógeno, puentes salinos, efectos hidrofóbicos, fuerzas de van der Waals (vdW), unión de ligandos, unión de cofactores, iones encuadernación, acompañantes y PTM, y más (Deller *et al.*, 2016). La estabilidad conformacional de la cadena polipeptídica da como resultado una penalización entrópica significativa ($-T\Delta S \gg 0$), y en condiciones celulares normales una proteína

plegada es solo marginalmente estable. Para superar esto, todos los demás factores que influyen en el plegamiento deben superar la pérdida de entropía conformacional. Los estudios sugieren que el efecto hidrófobo contribuye aproximadamente en un 60% a la estabilidad de la proteína y que los enlaces de hidrógeno contribuyen en aproximadamente un 40%. La contribución neta de los enlaces de hidrógeno a la estabilidad total de la proteína es $\sim 1,1$ kcal / mol y es independiente del tamaño de la proteína. Sin embargo, por el contrario, las interacciones hidrófobas suelen contribuir menos a la estabilidad de las proteínas pequeñas (Deller *et al.*, 2016; Nick Pace *et al.*, 2014; Pace *et al.*, 2011, 2014; Sticke *et al.*, 1992).

V. Metodología

A. ALINEAMIENTO Y FILTRADO DE CALIDAD DEL GENOMA

Se descargaron secuencias de genomas completos de Belice, Guatemala, El Salvador, Honduras, Panamá y Costa Rica de SARS-COV-2 depositadas en la base de datos de GISAID, hasta el día 16 de septiembre del 2021 en formato FASTA, para un total de 3014 secuencias. Se determinó la calidad y las variaciones de las secuencias mediante la herramienta Nexclade de Nexstrain. Nexclade, la cual clasifica las secuencias como de buena calidad, mediocre y mala, de acuerdo con el número de caracteres N (datos faltantes), el número de caracteres no ACGTN, el número de caracteres "-" (espacios), el número total de cambios inesperados del marco de lectura en la secuencia y el número de codones *stop* prematuros inesperados en las secuencias, para los análisis subsecuentes se conservaron únicamente las secuencias clasificadas como de buena calidad y mediocre. Posteriormente se realizó un alineamiento múltiple de las secuencias utilizando el servidor en línea del programa MAFFT v7 utilizando los parámetros por defecto. Una vez alineadas al genoma de referencia, las secuencias fueron recortadas para conservar únicamente la región genómica correspondiente de las bases 118 a la 28740 de la secuencia de referencia (NC_045512) dado a que estas regiones se ensamblan de manera inconsistente y aumentan la demanda de recursos e incertidumbre en los siguientes pasos.

Luego del recorte, se usó PRINSEQ para eliminar redundancias y conservar únicamente aquellas secuencias que cuenten con menos de un 1% de Ns dando así 1722 secuencias (Cuadro 3) (Schmieder & Edwards, 2011). Para los análisis que utilizan únicamente la proteína S (obtenida de Nexclade de Nexstrain), las secuencias fueron alineadas nuevamente utilizando los parámetros por defecto del paquete MAFFT v7 en línea (debido a que PRINSEQ desalinea las secuencias) y nuevamente procesadas con PRINSEQ para excluir duplicados para obtener un total de 422 secuencias (Schmieder & Edwards, 2011).

Cuadro 3: Secuencias utilizadas para el estudio de SARS-CoV-2 obtenidas de GISAID. La totalidad de secuencias fueron obtenidas el día 16 de septiembre de 2021. Al no contarse con datos para nicaragua, este país no será tomado para los análisis

País	Código
Belice	EPI_ISL_ 509711-14 y 1822737- 84
Guatemala	EPI_ISL_ 509695-710, 815398, y 99, 837551, 52, y 79-96, 1533981, 82, y 84-89, 2014355, 2825071- 136, 2826912- 7036, y 7791- 7867, 3098763, 3156041-44, 3275178- 248, 3319031-53, 3332930, 44, y 73, 3002, 3, 44, 91, 121, 175, y 176, 3347507, y 8, 3347863, y 954, 3446662, 67, 69, 70, 72, 75, 77, 80, 82, 84, 85, 87, 90, 91, 93, 98, 99, 700, 01, 05, 07, 08, 10, 11, 13, 15, 20-25, y 27, 3527791-829, 3801279-93, 3835359-415, 4029265- 360, 4051043- 103 y 4195008
El Salvador	EPI_ISL_671974, 671978, 672012, 70, 72, y 73, 3275254- 347, 3298771, 3299050, 3299422 y 3299814
Honduras	EPI_ISL_2157344-60, 378, 551-55, 77, 79, 80, y 83-85, 2648242-62, 2650523-38, 3375855-60, y 4029392- 4029435
Panamá	EPI_ISL_415152, 496603-915, 906306, 1001457, y 8, 1225330-583 y 1502818-3142

País	Código
Costa Rica	EPI_ISL_434533-36, y 38-40, 458236, 480312-28, 491437-57, 512653-75, 512846, 527739-60, 682235-77, 769986-770032, 914794-840, 1067587-625, 1196421-438, 1201437-40, 1379392-434, 1517387-434, 1527011-28, 1712378-421, 1811224-41, 1827503-47, 2103368-419, 2272958-3004, 2502718-62, 2657669, y 74-88, 2658264-85, 2827961-8015, 3026021-33, 3037793, 95-829, 31, 32, 34-36, 38-42, 44, y 45, 3274346-68, 3298304-73, 3464502-49, 3638747-49, 52, 63, 75, 84, 93, 803, 07, 12, 16, 20, 24, 35, 46, 56, 60, 64, 69, 74, 78, 88, 98, 902, 07, 17, 26, 35, 39, 43, 47, 51, 55, 59, 70, 74, 78, 86, 94, 98, 9001, 05, 17, 30, 33, 43, 48, 52, 56, 60, 64, 75, 78, 82, 85, 96, 107, 11, 14, 20, 34, 39, 43, 48, 51, 56, 66, 75, 78, 83, y 94, 3761768-811 y 3948477-548

B. MODELAJE MOLECULAR Y ANÁLISIS MUTAGENÓMICO

1. Efecto de las mutaciones en la estructura de la proteína S

Se utilizó como referencia la estructura cristalina 6VXX de la proteína S y utilizando los paquetes en línea de DeepDDG V2021 (que utiliza métodos de redes neuronales) (Cao *et al.*, 2019) y mCMS *Protein Stability* V2021 (que se basa en firmas basadas en gráficos que codifican patrones de distancia entre átomos para representar el entorno de residuos de proteínas) (Pires *et al.*, 2014) se determinó si las mutaciones encontradas con la herramienta de Nexclade de Nexstrain tienen un efecto desestabilizante o estabilizante sobre estructura de la proteína S con base en los cambios de energía libre en el estado sin mutar y el estado mutante ($\Delta\Delta G$).

Además, utilizando el paquete en línea de SuSPect V2021 (Yates *et al.*, 2014) se determinó el nivel de conservación de cada mutación por medio de una matriz de puntuación específica de la posición (PSSM), que es una forma de representación y predicción de motivos o patrones en secuencias que describen las variaciones intrínsecas en sus patrones (Gromiha, 2010).

2. Modelaje molecular

En este análisis, para reducir la cantidad de recursos computacionales, se recortaron las secuencias de la proteína S obtenidas con Nextclade de Nexstrain para obtener únicamente la región RBD (residuos 319-541). Para ello se utilizó la traducción de Nextclade de dichas secuencias, posteriormente se eliminó la redundancia utilizando PRINSEQ y se conservaron aquellas secuencias que contaban con un largo de 223 aminoácidos, dando así un total de 35 secuencias (Y. Huang *et al.*, 2020; Schmieider & Edwards, 2011). La secuencia de referencia de la proteína S del SARS-CoV-2 (NC_045512) se descargó de GISAID y modelos basados en homología fueron creados utilizando las secuencias de Centroamérica y la herramienta en línea de Phyre2 (Kelley *et al.*, 2015).

3. Análisis de acoplamiento molecular

La estructura cristalina de la región del receptor ACE2 humano (código PDB: 1R42) se descargó del repositorio de Protein Data Bank. Se utilizó HDOCK V1.1 (Y *et al.*, 2017; Yan *et al.*, 2017, 2020) para determinar los residuos que interactúan en la unión del complejo RBD-ACE2. Luego, se realizó el acoplamiento entre la región RBD de la proteína S (utilizando el recorte en la sección pasada) y su receptor mediante el servidor de HADDOCK (Honorato *et al.*, 2021; van Zundert & Bonvin, 2014), donde se realizó los análisis de acoplamiento para cada mutante de la proteína de S con ACE2 y se obtuvo la energía de van der Waals, energía de desolvatación, energía

de restricción, energía de afinidad y desviación cuadrática media (RMSD) de átomos de C α utilizando los ajustes predeterminados. Los modelos de la región RBD acoplado con ACE2 se eligieron de acuerdo con el valor Z y HADDOCK score.

C. ANÁLISIS PREDICTIVO DEL PANORAMA DE GLICOSILACIÓN

Para realizar un análisis predictivo del panorama de glicosilación se utilizaron las traducciones de la proteína S obtenidas de la herramienta de Nextclade de Nexstrain y dado a que algunas secuencias cuentan con N's (lo cual puede afectar el marco de lectura en algunos programas, causando resultados falsos positivos o negativos), se eliminaron todas aquellas secuencias que cuentan con un largo menor de 1237 aminoácidos. Esto dio como resultado un total de 206 secuencias (contando la secuencia de referencia).

La predicción de glicosilaciones ligadas a N se analizó utilizando el paquete en línea de N-Glycosite (M. Zhang *et al.*, 2004), que evalúa dos patrones distintivos de variación de sequones, con los parámetros predefinidos. El primer patrón (fijo) describe sitios fácilmente alineados que simplemente están presentes o ausentes. El segundo patrón (desplazamiento) se refiere a sitios incrustados en regiones de variación de longitud local extrema y se caracteriza por cambios en términos de posición relativa y densidad local de sequones. La predicción de sitios de glicosilaciones ligadas a O se realizó utilizando el paquete en línea de NetOGlyc 4.0 (Steentoft *et al.*, 2013), que utiliza redes neuronales artificiales.

D. ANÁLISIS DE RECOMBINACIÓN

La presencia de recombinación en las secuencias de Centroamérica se evaluó utilizando las 1722 secuencias traducidas de genoma completo de Nextclade de Nexstrain mediante Simplot 3.5.1 (Lole *et al.*, 1999). Simplot, el cual es una herramienta para detectar recombinación entre las secuencias analizadas, a través del desarrollo de gráficos de similitud. Es una de las herramientas bioinformáticas más utilizadas para detectar recombinación, y en general, proporciona un medio muy eficaz para detectar recombinación. Las regiones recombinantes potenciales entre las secuencias analizadas se identificaron deslizando una ventana de 400 pb en un paso de 50 pb a través de la alineación utilizando el modelo de 2 parámetros de Kimura (X. Li *et al.*, 2020b).

E. RUTA BIOINFORMÁTICA

Figura 5: Flujo de trabajo para análisis bioinformático de las secuencias de Centroamérica de SARS-CoV-2



VI. Resultados

A. ANÁLISIS MUTAGENÓMICO

1. Mutaciones presentes en la región S de secuencias del SC2

Como era de esperarse, la mutación más prevalente es la D614G, que está diseminada a nivel mundial desde inicios de la pandemia (Korber *et al.*, 2020) (Cuadro 4). En Belice la mutación más frecuente fue D614G (N=29), seguido de W152R (N=7) (Anexo 2); para El Salvador D614G (N=53) y L452R (N=32) (Anexo 3); Guatemala D614G (N=329) y L452R (N=198) (Anexo 4); Honduras D614G (N=40) y T1027I (N=22) (Anexo 5); Panamá D614G (N=266) y L452R (N=121) (Anexo 7); y Costa Rica D614G (N=799) y L452R (N=253) (Anexo 7). Algunas mutaciones encontradas en este estudio no pudieron ser utilizadas para los análisis debido a que algunos residuos no pudieron ser resueltos en la estructura cristalina de referencia, 6VXX (anexo 1), por lo que no todas serán incluidas en los análisis posteriores. Cabe notar que en las secuencias está presente más de una mutación a la vez, razón por la cual el porcentaje sumaría más de 100%. Los porcentajes serían relativos a la presencia individual en el total de secuencias

Cuadro 4: Frecuencia de mutaciones en las secuencias de SARS-CoV-2 en Centroamérica. Aquí están incluidos únicamente las 20 mutaciones más frecuentes. Para ver los resultados completos ver anexo 1.

Mutación	Frecuencia	Porcentaje
D614G	1516	87.8
L452R	611	35.4
D215Y	377	21.8
N501Y	231	13.4
P681H	205	11.9
T478K	161	9.3
D950N	158	9.1
T19R	151	8.7
P681R	149	8.6
T95I	146	8.5
T1117I	133	7.7
E484K	131	7.6
T716I	128	7.4
R158G	124	7.2
A570D	121	7.0
S982A	121	7.0
D1118H	121	7.0
T1027I	111	6.4
G142D	106	6.1
H655Y	81	4.7

2. Efecto de las mutaciones en la estabilidad de la proteína S

De aquellas mutaciones que presentan una gran prevalencia en la población de secuencias Centroamericanas (las 20 mayores), D614G, L452R, N501Y, D950N, T95I, T1117I, A570D, S982A, D1118H, G142D, H655Y, D138Y, R190S, K417T y S735A presentan un efecto desestabilizante. Por otro lado, únicamente D215Y, T716I, T1027I son estabilizantes en la proteína S (Cuadro 5). Pocas mutaciones presentan incongruencia entre los programas utilizados para evaluar el cambio de $\Delta\Delta G$, en su mayoría el análisis desarrollado en los dos programas muestra

resultados que confirmaban el cambio de estabilidad en la estructura. En total 146 de 194 mutaciones muestran concordancias entre los programas (MCP). En total, las mutaciones MCP D614G, L452R, N501Y, D950N, T95I, T1117I, A570D, S982A, D1118H, G142D, R190S, K417T, S735A, R346K, F1052Y, W64R, A263S, A701V, Q779K, T732A, N440Y, K417N, S98F, G1124V, T859N, L452Q, V120L, D936Y, S494P, G769R, L948I, V503I, A67V, V213A, Q675H, A879S, E583D, L560Q, N658D, S46L, S939F, F888L, A899S, A27V, D80G, L141Y, A348S, A522V, A1078S, D950H, N1074S, N188K, N30H, V308L, E96G, T547I, P384L, V143F, T307I, K558N, R214G, L212A, R346S, G769V, L226V, V1122L, T618I, V860I, Q613H, I714V, K786Q, V1104L, V407I, L141F, V367L, A419S, T307S, F59S, V1133F, K1073N, A1070V, K1073M, I210T, V367F, Y266I, V171D, A93T, F32L, P82H, L118I, F490L, Q52H, A67D, Y265N, A411S, E554G, F306L, V143S, G142S, V503F, F220L, E309Q, V227L, Q218H, G142L, P812S, E340G, L270H, F238C, N211S, A67S, A1022S, V534I, G142P, L242F, S112L, E309A, V143G, E583Q, Y508H, V512I, K378R, T274I, T95S, D1118Y, V511I, P812L, S659L, N658T, V1129L, A647S, E654V, V608I, I1114L, S689I, S940F, K304R, A942V, A1080S, S929N, E654Q, L582F, A1087S, I896T y T778A presentan un efecto desestabilizante en la proteína, mientras que una sola MCP, T791I, tiene efecto estabilizante en la proteína (Cuadro 5).

Las mutaciones G35C, N439K, T29I, K97E, F342L, D215G, D215H, D215Y, R408T, Q957R, D796H, T719I, D1127H, D80Y, D287H, Q52R, H49Y, D88H, D80H, D80A, P209R, N211R, D138V, H1101Y, H1058Y, A222V, N440K, S221L, N1074D, Q1071L, E324K, H655Y, D138Y, D936N, H69Y, T716I, D138N, F490S, T572I, V289I, T1027I, A570S, T859I, V911I, S704L, S50L, T573I y D215N presentan una discordancia entre los dos programas, en el sentido de que uno de los resultados es positivo, mientras que el otro es negativo (Cuadro 5). Es por ello por lo que estos no fueron considerados. Cabe notar que en su mayoría están presentes únicamente en un porcentaje menor a 1%, a excepción de D215Y (21.8%), H655Y (4.7%), D138Y (4.4%), T716I (7.4%) y T1027I (6.4%). A pesar de que D215Y si está significativamente presente, los resultados de $\Delta\Delta G$ son muy diferentes, variando 1.30 kcal/mol entre sí. Adicionalmente, la mutación D215N no puede ser tomada en cuenta dado que no es congruente con el resto de las mutaciones o entre programas el resultado de DeepDGG, 508 kcal/mol en comparación con -0.07 kcal/mol en mCSM.

Cuadro 5: El efecto de la sustitución de aminoácidos en la proteína S del SARS-CoV-2 sobre la estabilidad de su microambiente. DeepDGG y mCSM miden el cambio en la estabilidad por energía libre en kcal/mol, dando una acción desestabilizante (D) o estabilizante (E). El resultado resume si existe una concordancia entre ambos programas, dando un Sí o No.

Mutación	Cambio en $\Delta\Delta G$ (kcal/mol)		Predicción		Resultado
	DeepDGG	mCSM	DeepDGG	mCSM	
D614G	-0.13	-0.23	D	D	Sí
L452R	-0.83	-0.78	D	D	Sí
D215Y	-0.71	2.01	D	E	No
N501Y	-0.73	-0.48	D	D	Sí
D950N	-0.28	-0.242	D	D	Sí
T95I	-0.85	-0.689	D	D	Sí
T1117I	-0.13	0.00	D	D	Sí
T716I	-0.02	0.08	D	E	No
A570D	-1.49	-1.40	D	D	Sí
S982A	-0.20	-0.62	D	D	Sí
D1118H	-0.26	-0.47	D	D	Sí
T1027I	0.09	-0.40	E	D	No
G142D	-0.84	-0.198	D	D	Sí

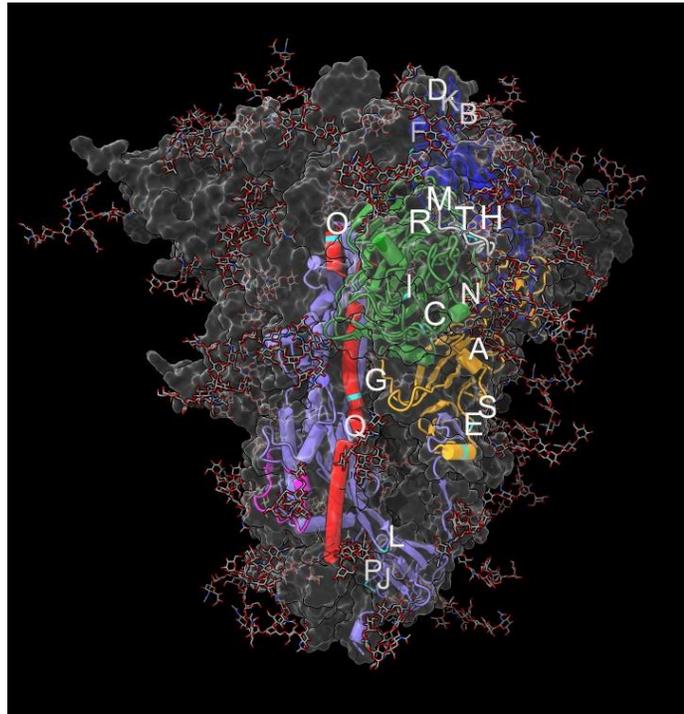
Mutación	Cambio en $\Delta\Delta G$ (kcal/mol)		Predicción		Resultado
	DeepDGG	mCSM	DeepDGG	mCSM	
H655Y	-0.14	1.21	D	E	No
D138Y	-0.14	0.05	D	E	No
R190S	-2.01	-2.33	D	D	Sí
K417T	-0.62	-0.92	D	D	Sí
S735A	-0.53	-0.76	D	D	Sí
R346K	-0.07	-0.04	D	D	Sí
F1052Y	-4.39	-1.09	D	D	Sí
W64R	-3.50	-1.35	D	D	Sí
A263S	-0.77	-1.568	D	D	Sí
A701V	-0.15	-0.38	D	D	Sí
D215G	-0.85	0.60	D	E	No
Q779K	-0.24	-0.184	D	D	Sí
T732A	-0.08	-0.70	D	D	Sí
N440Y	-0.32	-0.56	D	D	Sí
K417N	-0.43	-0.68	D	D	Sí
D80A	-0.35	0.22	D	E	No
S98F	-0.23	-0.79	D	D	Sí
G1124V	-1.12	-0.20	D	D	Sí
T859N	-3.04	-0.18	D	D	Sí
A222V	-0.25	0.249	D	E	No
L452Q	-2.62	-0.98	D	D	Sí
F490S	0.03	-0.93	E	D	No
V120L	-1.08	-0.391	D	D	Sí
D936Y	-0.19	-0.48	D	D	Sí
S494P	-0.31	-0.24	D	D	Sí
N440K	-0.22	0.223	D	E	No
G769R	-0.24	-0.63	D	D	Sí
L948I	-0.35	-1.15	D	D	Sí
V503I	-0.03	-0.24	D	D	Sí
A67V	-0.18	-0.21	D	D	Sí
V213A	-0.93	-0.617	D	D	Sí
Q675H	-0.04	-0.79	D	D	Sí
A879S	-0.49	-1.69	D	D	Sí
E583D	-0.88	-0.90	D	D	Sí
L560Q	-0.62	-1.47	D	D	Sí
S221L	-0.21	0.07	D	E	No
T573I	1.28	-0.40	E	D	No
N658D	-0.23	-0.90	D	D	Sí
Q52R	-0.41	0.17	D	E	No
N439K	-2.14	0.24	D	E	No
S46L	-0.20	-0.253	D	D	Sí
T572I	0.07	-0.24	E	D	No
S939F	-0.17	-1.23	D	D	Sí
F888L	-1.56	-1.47	D	D	Sí
A899S	-0.05	-1.93	D	D	Sí
A27V	-0.84	-0.31	D	D	Sí
D80G	-0.34	-0.12	D	D	Sí
D80Y	-0.42	1.13	D	E	No

Mutación	Cambio en $\Delta\Delta G$ (kcal/mol)		Predicción		Resultado
	DeepDGG	mCSM	DeepDGG	mCSM	
L141Y	-0.79	-0.48	D	D	Sí
A348S	-1.28	-1.124	D	D	Sí
A522V	-0.48	-0.683	D	D	Sí
A1078S	-1.27	-1.83	D	D	Sí
D950H	-0.78	-0.85	D	D	Sí
N1074S	-0.23	-0.28	D	D	Sí
N188K	-0.55	-0.49	D	D	Sí
H49Y	-0.40	0.84	D	E	No
N30H	-1.87	-1.12	D	D	Sí
T29I	-1.42	0.20	D	E	No
V308L	-1.14	-0.77	D	D	Sí
E96G	-4.54	-2.05	D	D	Sí
T547I	-0.76	-0.60	D	D	Sí
P384L	-1.03	-0.47	D	D	Sí
V143F	-0.82	-0.66	D	D	Sí
D80H	-0.36	1.035	D	E	No
T307I	-0.44	-0.128	D	D	Sí
K558N	-0.27	-0.20	D	D	Sí
R214G	-0.65	-0.13	D	D	Sí
L212A	-1.69	-1.364	D	D	Sí
R346S	-0.27	-0.19	D	D	Sí
G769V	-0.64	-0.385	D	D	Sí
H1101Y	-0.26	1.25	D	E	No
D936N	-0.06	0.28	D	E	No
L226V	-0.44	-0.569	D	D	Sí
T719I	-0.50	0.09	D	E	No
V1122L	-1.34	-0.435	D	D	Sí
T618I	0.00	-0.052	D	D	Sí
V860I	-0.44	-0.323	D	D	Sí
Q613H	-2.84	-0.935	D	D	Sí
I714V	-0.97	-1.382	D	D	Sí
K786Q	-0.39	-0.207	D	D	Sí
V1104L	-2.47	-0.696	D	D	Sí
N1074D	-0.21	0.30	D	E	No
V407I	-0.87	-0.53	D	D	Sí
L141F	-0.43	-0.659	D	D	Sí
V367L	-0.29	-0.26	D	D	Sí
A419S	-0.78	-1.82	D	D	Sí
T307S	-0.90	-0.41	D	D	Sí
F59S	-1.96	-2.131	D	D	Sí
V1133F	-1.32	-1.03	D	D	Sí
H1058Y	-0.26	1.14	D	E	No
K1073N	-0.81	-0.88	D	D	Sí
A1070V	-0.27	-0.68	D	D	Sí
K1073M	-0.29	-0.22	D	D	Sí
I210T	-2.00	-1.88	D	D	Sí
V367F	-0.36	-0.68	D	D	Sí
Y266I	-4.91	-1.59	D	D	Sí

Mutación	Cambio en $\Delta\Delta G$ (kcal/mol)		Predicción		Resultados
	DeepDGG	mCSM	DeepDGG	mCSM	
V171D	-3.01	-0.62	D	D	Sí
A93T	-3.06	-0.85	D	D	Sí
F32L	-1.56	-1.24	D	D	Sí
P82H	-1.10	-1.08	D	D	Sí
L118I	-0.92	-0.43	D	D	Sí
F490L	-0.34	-0.50	D	D	Sí
Q52H	-0.39	-0.50	D	D	Sí
A67D	-2.34	-0.718	D	D	Sí
D138N	0.01	-0.132	E	D	No
D88H	-0.36	0.367	D	E	No
P209R	-0.33	0.039	D	E	No
Y265N	-5.38	-2.88	D	D	Sí
A411S	-1.70	-1.48	D	D	Sí
E554G	-1.01	-0.43	D	D	Sí
V911I	0.24	-0.34	E	D	No
F306L	-1.21	-1.13	D	D	Sí
V143S	-1.13	-0.41	D	D	Sí
G142S	-0.91	-0.72	D	D	Sí
V503F	-0.39	-0.59	D	D	Sí
F220L	-2.33	-1.166	D	D	Sí
E309Q	-0.21	-0.25	D	D	Sí
D138V	-0.27	0.05	D	E	No
V227L	-0.44	-0.31	D	D	Sí
Q218H	-0.39	-0.10	D	D	Sí
G35C	-2.68	0.38	D	E	No
H69Y	-0.06	0.02	D	E	No
G142L	-0.49	-0.229	D	D	Sí
N211R	-0.32	0.204	D	E	No
P812S	-0.23	-0.57	D	D	Sí
E340G	-1.11	-0.46	D	D	Sí
L270H	-4.13	-2.84	D	D	Sí
F238C	-3.22	-1.26	D	D	Sí
N211S	-0.33	-0.06	D	D	Sí
E324K	-0.17	0.14	D	E	No
A570S	0.11	-1.254	E	D	No
A67S	-0.65	-0.793	D	D	Sí
S704L	0.43	-0.17	E	D	No
T859I	0.22	-0.26	E	D	No
A1022S	-1.89	-1.92	D	D	Sí
F342L	-0.87	0.57	D	E	No
V534I	-0.05	-0.54	D	D	Sí
S50L	0.87	-0.02	E	D	No
G142P	-0.56	-0.21	D	D	Sí
L242F	-1.64	-1.29	D	D	Sí
S112L	-0.71	-0.339	D	D	Sí
E309A	-0.49	-0.437	D	D	Sí
V143G	-1.90	-0.402	D	D	Sí
D796H	-0.51	0.91	D	E	No

Mutación	Cambio en $\Delta\Delta G$ (kcal/mol)		Predicción		Resultado
	DeepDGG	mCSM	DeepDGG	mCSM	
E583Q	-0.04	-0.73	D	D	Sí
D215N	508.00	-0.07	E	D	No
Y508H	-4.73	-2.50	D	D	Sí
V512I	-0.79	-0.05	D	D	Sí
K378R	-0.64	-0.396	D	D	Sí
T274I	-1.21	-0.303	D	D	Sí
K97E	-0.94	0.225	D	E	No
T95S	-2.38	-1.634	D	D	Sí
D1118Y	-0.28	-0.26	D	D	Sí
V511I	-1.27	-0.07	D	D	Sí
P812L	-0.28	-0.40	D	D	Sí
S659L	-0.23	-0.43	D	D	Sí
N658T	-0.18	-0.08	D	D	Sí
V1129L	-1.73	-0.43	D	D	Sí
D1127H	-0.44	0.417	D	E	No
A647S	-0.93	-1.465	D	D	Sí
E654V	-0.42	-0.17	D	D	Sí
V608I	-1.17	-0.70	D	D	Sí
I1114L	-0.09	-0.53	D	D	Sí
Q1071L	-0.20	0.15	D	E	No
R408T	-0.70	0.121	D	E	No
S689I	0.00	-0.37	D	D	Sí
D287H	-0.42	1.182	D	E	No
D215H	-0.72	1.525	D	E	No
S940F	-0.01	-1.05	D	D	Sí
V289I	0.07	-0.111	E	D	No
Q957R	-0.68	0.192	D	E	No
K304R	-0.63	-0.283	D	D	Sí
A942V	-0.13	-0.13	D	D	Sí
T791I	0.04	0.05	E	E	Sí
A1080S	-2.88	-1.539	D	D	Sí
S929N	-0.02	-0.496	D	D	Sí
E654Q	-0.46	-0.427	D	D	Sí
L582F	-0.08	-0.896	D	D	Sí
A1087S	-3.33	-2.15	D	D	Sí
I896T	-2.05	-3.053	D	D	Sí
T778A	-1.65	-1.15	D	D	Sí

Figura 6: Representación molecular de las 20 posiciones donde ocurren más frecuentemente mutaciones en las secuencias de Centroamérica de una cadena de la proteína S del SARS-CoV-2 glicosilada utilizando el modelo de referencia 6VXX de Woo y compañía (Woo *et al.*, 2020). La representación está separada por colores, siendo gris el péptido de señal, Naranja la subunidad S1, verde el dominio amino terminal (NTD) y, azul el dominio de unión al receptor (RDB); en Morado subunidad S2, en magenta el péptido de fusión (FP) y en rojo la secuencia repetida de heptapéptidos 1 (HR1). Las letras (A-T) representan las posiciones con las mutaciones más frecuentes (en color cian) en orden descendente, 614, 452, 215, 501, 681, 478, 950, 19, 95, 1117, 484, 716, 158, 570, 982, 1118, 1027, 142, 655, 18, respectivamente. Para ver una lista completa de las mutaciones y su ubicación ver Anexo 19 y 20



B. MODELAJE MOLECULAR Y ANÁLISIS DE ACOPLAMIENTO MOLECULAR

Los resultados de los complejos creados de las secuencias EPI_ISL_509713, EPI_ISL_914803, EPI_ISL_1533989, EPI_ISL_2103385, EPI_ISL_1533987, EPI_ISL_1503045, EPI_ISL_1502972, EPI_ISL_1225487, EPI_ISL_1502847, EPI_ISL_1225554, EPI_ISL_4029413, EPI_ISL_4029275, EPI_ISL_3761785 y EPI_ISL_3275315 no pueden ser considerados como resultados representativos entre los análisis de acoplamiento dado al valor en angstroms de RMSD (Anexo 11). Este valor es una medida de similitud en la estructura tridimensionales; aquellos modelos con un valor muy grande ($>3 \text{ \AA}$) describe estructuras que son diferentes y cero significa que son idénticas en conformación con respecto a la pose de unión verdadera. Un valor 4 \AA no es mejor que uno de 6 \AA , ninguno fue realmente significativo (Kufareva & Abagyan, 2012; Mayorov & Crippen, 1994). Es por ello por lo que no estarán presentes en los análisis de interacciones moleculares y afinidad de unión.

De los complejos creados, únicamente EPI_ISL_1517411, EPI_ISL_1225513, EPI_ISL_1811226, EPI_ISL_1196430 presentan una menor energía de afinidad por ACE2 que la secuencia de referencia. Por los valores tan similares que presenta, EPI_ISL_3319047 también está

dentro de este grupo. El resto de los complejos son más estables o presentan valores mayores en los resultados de acoplamiento. Cabe notar que todos estos modelos muestran resultados bastante cercanos a 0 en cuando a RMSD, lo que indica que los modelos se acoplan muy bien a la realidad.

Cuadro 6: Análisis de acoplamiento molecular realizado con HADDOCK. Se presenta el promedio de la desviación cuadrática media (RMSD) de átomos de Ca en Angstroms, energía de Van der Waals (EVdW) en kcal/mol, energía electrostática (EE) en kcal/mol, energía de desolvatación (ED) en kcal/mol, y Energía de Unión (EU) en kcal/mol. NC_045512 es la secuencia de referencia.

Código	RMSD (Å)	EVdW (kcal/mol)	EE (kcal/mol)	ED (kcal/mol)	EU (kcal/mol)
NC_045512	0.8± 0.6	-67.4± 5.4	-371.3± 30.5	-17.3± 4.7	-378.024
EPI_ISL_527757	1.3± 1	-67.1± 10.6	-274.1± 28.6	-20.3± 4.3	-317.011
EPI_ISL_1196430	0.4± 0.3	-50.4± 5.3	-402.5± 33.5	-18.1± 1.6	-383.841
EPI_ISL_3275184	0.5± 0.3	-60.5± 6	-398.6± 32.5	-17.6± 2.7	-372.941
EPI_ISL_3275178	0.6± 0.4	-67.8± 6.9	-299.7± 44.9	-11.4± 1.9	-334.931
EPI_ISL_2157577	0.9± 0.6	-75.6± 12.7	-182.3± 18.5	-14.4± 3.6	-209.313
EPI_ISL_1822780	1.4± 0.8	-82.5± 8.1	-149.8± 38	-21.9± 3.9	-157.344
EPI_ISL_1811226	0.8± 0.6	-60.9± 12.4	-404.5± 49.2	-13.8± 2.8	-388.744
EPI_ISL_1527015	0.7± 0.5	-72.8± 1.5	-238.8± 31.4	-19.8± 1.6	-242.608
EPI_ISL_1517411	0.7± 0.4	-82.3± 3.6	-217.4± 48.3	-29.4± 7.2	-32849.7
EPI_ISL_1502946	0.8± 0.5	-74.7± 9	-188.3± 43.1	-21.6± 3.1	-178.145
EPI_ISL_1502939	0.8± 0.4	-63.2± 8.5	-395.2± 51.8	-17.4± 3.9	-337.129
EPI_ISL_1502908	0.7± 0.6	-74.6± 3.6	-237.5± 20.4	-16.5± 3.3	-245.882
EPI_ISL_1225513	0.6± 0.4	-60.9± 3.6	-407.8± 38	-15.2± 4.8	-429.147
EPI_ISL_4029280	0.4± 0.2	-63.4± 2.4	-304.1± 29.7	-19.6± 3.5	-264.91
EPI_ISL_4029266	0.6± 0.4	-72.5± 4.9	-357.4± 57.5	-10.7± 2.9	-360.401
EPI_ISL_3835382	0.5± 0.4	-63.1± 7.5	-363.6± 28.8	-18.1± 3.8	-313.92
EPI_ISL_3319047	0.8± 0.5	-72.2± 4.3	-324.6± 28.1	-25.8± 4.4	-377.856
EPI_ISL_3275319	1.2± 1.1	-81.2± 3	-147.4± 17.2	-22.8± 4.1	-199.402
EPI_ISL_3275226	1.4± 0.8	-77.6± 7	-256.3± 41.4	-13.2± 4.7	-209.647

C. ANÁLISIS PREDICTIVO DEL PANORAMA DE GLICOSILACIÓN

Aún no está claro el panorama de glicosilaciones ligadas a O (GLO) de la proteína S del nuevo coronavirus, pero en este estudio realizó una predicción de los 3 sitios potenciales de GLO, 673, 678 y 686, que concuerdan con la literatura y que están presentes en la mayoría de las secuencias (Wand *et al.*, 2021; Xu *et al.*, 2020; Zhao *et al.*, 2021) (Cuadro 7 y 9, Anexo 8). Adicionalmente, fue posible identificar sitios de potenciales de GLO adicionales, que en su mayoría están en la posición 29 (n=45), mientras que el resto de secuencias, 19, 22, 26, 250, 688 y 689, están presentes únicamente una o dos veces (Cuadro 7, 9, y Anexo). Cabe notar que no todas las secuencias cuentan con las tres sitios potenciales GLO de referencia (Cuadro 8).

Cuadro 7: Frecuencia de glicosilaciones ligadas a O en la proteína S de las secuencias de Centroamérica de SARS-CoV-2.

Posición	Frecuencia	Porcentaje
673	200	97.1
678	174	84.5
686	200	97.1
19	1	0.5
22	5	2.4
26	1	0.5
29	45	21.8
250	1	0.5
688	1	0.5
689	2	1.0

Cuadro 8: Secuencias de SARS-CoV-2 de la proteína S de Centroamérica que no presentan glicosilaciones ligadas a O en las posiciones 673, 678 y/o 686 basándose en un puntaje menor a 0.5 de NetOGlyc.

Secuencia	Posición	Puntaje	Secuencia	Posición	Puntaje	Secuencia	Posición	Puntaje
EPI_ISL_1822780	678	0.311	EPI_ISL_1067606	673	0.488	EPI_ISL_3037838	678	0.328
EPI_ISL_1502946	673	0.481	EPI_ISL_1067606	686	0.482	EPI_ISL_2502729	673	0.475
EPI_ISL_3835382	673	0.491	EPI_ISL_4029345	686	0.464	EPI_ISL_2657677	678	0.317
EPI_ISL_4029280	686	0.492	EPI_ISL_914828	678	0.318	EPI_ISL_1712390	678	0.351
EPI_ISL_3835393	678	0.416	EPI_ISL_4029329	678	0.316	EPI_ISL_2657688	678	0.315
EPI_ISL_2157344	678	0.424	EPI_ISL_4029304	678	0.322	EPI_ISL_1712384	673	0.473
EPI_ISL_4029269	678	0.395	EPI_ISL_4029320	678	0.423	EPI_ISL_1712384	678	0.269
EPI_ISL_2157345	678	0.421	EPI_ISL_4029328	678	0.325	EPI_ISL_1712384	686	0.464
EPI_ISL_2157583	678	0.484	EPI_ISL_3275328	678	0.406	EPI_ISL_3274367	678	0.331
EPI_ISL_2157350	678	0.440	EPI_ISL_1196430	678	0.325	EPI_ISL_3274352	678	0.337
EPI_ISL_4029287	678	0.443	EPI_ISL_1067620	678	0.319	EPI_ISL_3274351	678	0.331

Secuencia	Posición	Puntaje	Secuencia	Posición	Puntaje	Secuencia	Posición	Puntaje
EPI_ISL_ 402927 4	678	0.315	EPI_ISL_ 106762 2	673	0.485	EPI_ISL_ 327435 8	678	0.317
EPI_ISL_ 402927 5	678	0.443	EPI_ISL_ 106762 2	686	0.478	EPI_ISL_ 327435 6	678	0.321
EPI_ISL_ 402930 9	678	0.397	EPI_ISL_ 282799 5	678	0.329	EPI_ISL_ 327435 5	678	0.312
EPI_ISL_ 106758 9	678	0.323	EPI_ISL_ 227296 7	686	0.473	EPI_ISL_ 303783 8	678	0.328

Cuadro 9: Presencia de glicosilaciones ligadas a O en las secuencias de la proteína S del SARS-CoV-2 en Centroamérica basándose en un puntaje mayor a o igual a 0.5 de NetOGlyc. Aquí se encuentran únicamente aquellas secuencias con predicciones de glicosilaciones adicionales a las de las posiciones 673, 678 y 686. Para ver los resultados de dichas posiciones ver anexo 14.

Secuencia	Posición	Puntaje	Secuencia	Posición	Puntaje	Secuencia	Posición	Puntaje
EPI_ISL_ 150298 7	22	0.540	EPI_ISL_ 327518 4	29	0.530	EPI_ISL_ 402926 6	29	0.519
EPI_ISL_ 122548 6	29	0.565	EPI_ISL_ 402926 7	29	0.519	EPI_ISL_ 402934 5	29	0.500
EPI_ISL_ 150296 5	29	0.500	EPI_ISL_ 327517 8	29	0.509	EPI_ISL_ 682243	29	0.500
EPI_ISL_ 122548 8	29	0.500	EPI_ISL_ 215755 1	22	0.500	EPI_ISL_ 671974	29	0.500
EPI_ISL_ 150312 8	29	0.528	EPI_ISL_ 215734 4	29	0.500	EPI_ISL_ 402933 4	29	0.500
EPI_ISL_ 150290 8	29	0.500	EPI_ISL_ 402926 9	29	0.500	EPI_ISL_ 402929 6	26	0.589
EPI_ISL_ 150284 7	29	0.500	EPI_ISL_ 402927 6	29	0.500	EPI_ISL_ 402932 0	19	0.526
EPI_ISL_ 150290 0	29	0.523	EPI_ISL_ 402926 8	29	0.528	EPI_ISL_ 327532 8	29	0.500
EPI_ISL_ 182278 0	29	0.500	EPI_ISL_ 402927 3	29	0.519	EPI_ISL_ 171242	29	0.510
EPI_ISL_ 150289 1	29	0.500	EPI_ISL_ 215736 0	29	0.500	EPI_ISL_ 0	29	0.510

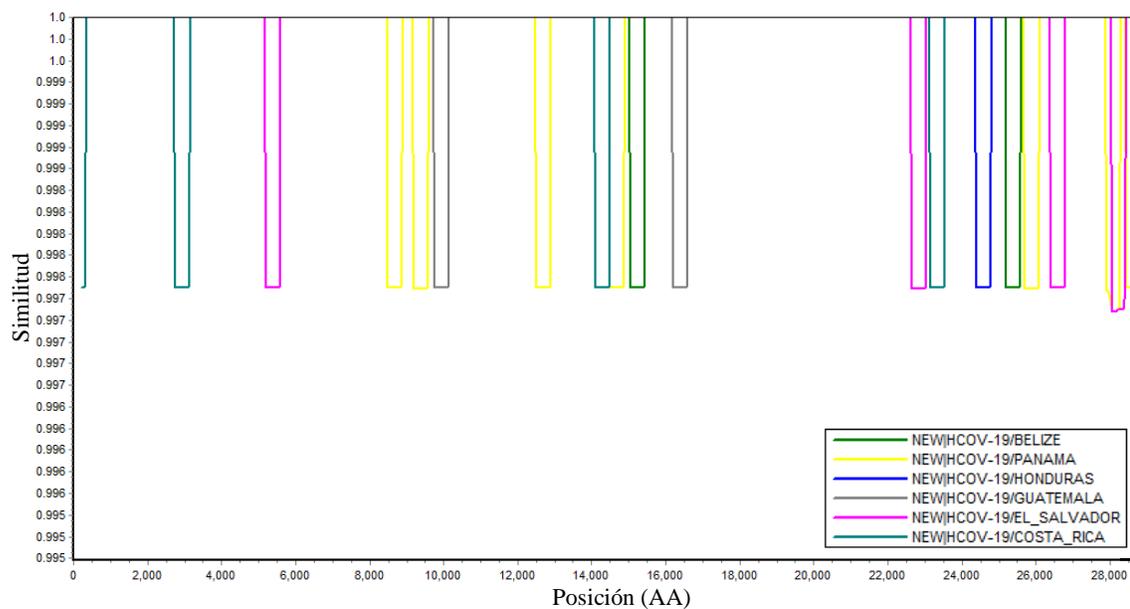
Secuencia	Posición	Puntaje	Secuencia	Posición	Puntaje	Secuencia	Posición	Puntaje
EPI_ISL_150294	6	0.509	EPI_ISL_402928	4	0.500	EPI_ISL_227300	4	0.541
EPI_ISL_122547	8	0.500	EPI_ISL_153398	7	0.500	EPI_ISL_151741	1	0.500
EPI_ISL_150285	7	0.541	EPI_ISL_327523	1	0.528	EPI_ISL_120143	9	0.506
EPI_ISL_215755	2	0.500	EPI_ISL_215735	0	0.511	EPI_ISL_106762	5	0.500
EPI_ISL_837584	688	0.511	EPI_ISL_402927	0	0.528	EPI_ISL_182751	7	0.500
EPI_ISL_402926	5	0.526	EPI_ISL_150308	5	0.500	EPI_ISL_171239	0	0.500
			EPI_ISL_327522	0	0.500	EPI_ISL_152701	5	0.506
			EPI_ISL_327522	6	0.514			

Todas las secuencias analizadas presentan los 22 sitios potenciales de glicosilaciones ligadas a N (GLN) que muestra la secuencia de referencia y la literatura. De las 206 secuencias, tres no presentan el sequón en la posición 17 (EPI_ISL_4029275, EPI_ISL_4029287 y EPI_ISL_3275319), dos en la posición 616 (EPI_ISL_4029285 y EPI_ISL_4029286) y una en la posición 657 (EPI_ISL_1502965) y una en la 1074 (EPI_ISL_496834), respectivamente. Fue posible identificar dos sequones potenciales adicionales para la posición 20 (EPI_ISL_4029413, EPI_ISL_3275214, EPI_ISL_3275241, EPI_ISL_3319047, EPI_ISL_3527791, EPI_ISL_4029319, EPI_ISL_1712404, EPI_ISL_2103368, EPI_ISL_2272980, EPI_ISL_2273001, EPI_ISL_2273004, EPI_ISL_2502747, EPI_ISL_3037810 y EPI_ISL_3037826) y la posición 188 (EPI_ISL_4029413, EPI_ISL_3275214, EPI_ISL_3275241, EPI_ISL_3319047, EPI_ISL_3527791, EPI_ISL_3835385, EPI_ISL_1712404, EPI_ISL_2103368, EPI_ISL_2272980, EPI_ISL_2273001, EPI_ISL_2273004, EPI_ISL_2502747, EPI_ISL_3037810 y EPI_ISL_3037826).

D. ANÁLISIS DE RECOMBINACIÓN

Los resultados de SimPlot no mostraron evidencia de algún evento de recombinación entre las secuencias centroamericanas utilizadas en este estudio. Esto se debe al alto nivel de similitud de todas las secuencias (>99.7%), por lo que, aunque eventos de entrecruzamiento y recombinación hubiesen sucedido, estos quedarían enmascarados. Las diferencias entre las secuencias de genoma completo (los picos de distintos colores, ver clave) se deben a las pocas mutaciones en las secuencias y no por eventos de recombinación.

Figura 7: Gráfico de similitud utilizando la secuencia de referencia NC_045512 como comparación entre cada país. Todos los grupos muestran una alta similitud entre sí, sin siquiera salir del 99%. En el eje X se muestra la posición en aminoácidos de las secuencias y en el eje Y el porcentaje de similitud entre las mismas.



VII. Análisis de resultados

A. ANÁLISIS MUTAGENÓMICO

1. Efecto de las mutaciones en la estabilidad de la proteína S

Dentro de las MPCs encontradas, el mayor efecto desestabilizante es dado por la mutación Y265N con un valor de -5.38 y -2.88 kcal/mol para DeepDDG y mCSM, respectivamente (Cuadro 5). Sin embargo, no está presente de manera significativa en la población (0.1%) (Anexo 1), lo cual concuerda con los hallazgos del PSSM que demuestran que N es menos favorable que Y. De manera similar, Y266I, Y508H, E96G, F1052Y, L270H, W64R, A1087S, F238C, T859N, V171D, A1080S, Q613H, L452Q, V1104L, T95S, A67D, F220L, I896T, F59S, A93T, I210T, W64R y R190S cuentan con valores de $\Delta\Delta G$ elevados (-2 a -4.5 kcal/mol) (Cuadro 5), pero no están presentes de forma significativa en la población (<1%) (Anexo 1) y según el PSSM es más probable que ocurra otro aminoácido o que ocurra el aminoácido original. La excepción sería A93T y I210T, que según el PSSM es más probable que ocurra el residuo mutado que el original. Este patrón nos podría indicar que las mutaciones más desestabilizante son las que menor efecto tienen sobre la población, dado a que no están presentes de una forma significativa en la población. A pesar de afectar la estructura tridimensional de la proteína S, realmente no presentan un peligro por la baja frecuencia en la población. Sin embargo, esto no se puede afirmar ya que solo reflejan que está en baja proporción dentro de la población de estudio dado por la baja disponibilidad de secuencias de SARS-CoV-2 en la región. Esto pone en evidencia que es necesario aumentar los esfuerzos por mantener una vigilancia genómica activa.

De manera similar a lo discutido anteriormente, frecuencias ≥ 4 y <10% cuentan con un $\Delta\Delta G$ <-1,5 y ≥ 0 kcal/mol, lo que indica que se tiene un efecto levemente desestabilizante. Estas mutaciones serían A570D, T95I, G142D, K417T, D950N, D1118H, S982A y T1117I y el PSSM demuestra básicamente lo mismo que antes: otros residuos son más favorables o el residuo original es más probable que ocurra, lo que concuerda con la frecuencia dentro de la población. Una excepción es K417T (región RBD, Anexo 9) que el PSSM indica alta conservación, pero está únicamente en 70 secuencias, por lo que, si pudiera llegar a verse de una manera más frecuente, pero para ello se necesita una muestra de secuencias más representativa.

Las tres MPCs más frecuentes, L452R (35.4%), N501Y (13.4%) y D614G (87.8%), se tienen efectos levemente desestabilizante (<-1 kcal/mol), pero la D614G es mundialmente dominante, la N501Y es mutación de alta preocupación y L452 ha sido descrita que tiene la capacidad de ayudar a SC2 a evadir el sistema inmune y aumenta la afinidad por ACE2 al estar en la región RBD (Leung *et al.*, 2021; Motozono *et al.*, 2021; Nelson *et al.*, 2021; Ozono *et al.*, 2021). Estas tres mutaciones tienen la capacidad de hacer la proteína S más afín a ACE2, indicando que a pesar de ser levemente desestabilizante si afectan a la estructura de manera significativa, por lo que es necesario estar alerta por la aparición de estas mutaciones en las secuencias.

En conjunto estos resultados muestran que a pesar de que los resultados muestran que algunas mutaciones tienen un $\Delta\Delta G$ levemente negativo si puede llegar a afectar de manera significativa la estructura tridimensional, llegando a alterar la función de la proteína S. Adicionalmente, la baja frecuencia en la población de estudio no refleja realmente el comportamiento del virus en la región, dejando escaso el análisis espacio temporal y llamando a aumentar los esfuerzos para tener un programa de vigilancia genómica activa.

B. ANÁLISIS DE ACOPLAMIENTO MOLECULAR

Los complejos que pertenecen a las secuencias EPI_ISL_1517411, EPI_ISL_1225513, EPI_ISL_1811226, EPI_ISL_1196430 tienen menor energía de afinidad que el complejo de referencia. Es importante notar que EPI_ISL_1517411 presenta una EU casi 87 veces menor que la de referencia. Adicionalmente, este complejo es el más estable de todos, teniendo una ED de -29.4 kcal/mol, a comparación de la referencia (Cuadro 6).

La mutación en EPI_ISL_1517411, V503F, es levemente desestabilizante (>1 kcal/mol) (Cuadro 5), siendo menos favorable que se produzca la mutación a comparación de otros aminoácidos según el PSSM. Esto es confirmado al estar presente únicamente en una sola secuencia (siendo el 0.1% de la población) (Anexo 1). Según la herramienta de CoV-Glue este cambio si afecta la estructura, ya que F es más grande que V e introduce aromaticidad, pero junto con los resultados de GISAIID nos indica que está en baja frecuencia (0.01% de las secuencias) (CoV-Glue, n.d.). Adicional a dicha mutación, esta secuencia también cuenta con D614G y T1117I que D614G se sabe que puede llegar a aumentar la afinidad por ACE2, llevando a una elevada infectividad (Ozono *et al.*, 2021). Ambas presentan un efecto levemente desestabilizante (Cuadro 5), lo que podría explicar la elevada afinidad por ACE2, llegando a afectar la estructura tridimensional de la proteína S de forma que alteran su conformación. En conjunto estas tres mutaciones parecieran funcionar en sinergia para elevar su afinidad por el receptor, y a pesar de ser alarmante estas mutaciones no constituyen por sí mismas una razón para clasificarlas como variante de interés (VdI). La mutación D614G es mundialmente dominante, pero la V503F se ve muy raramente, sin embargo, es necesario mantener en vigilancia la aparición de esta mutación, porque no se puede afirmar o negar si la elevada afinidad es a causa de una sinergia entre estas tres sustituciones o por la mutación V503F por si sola.

EPI_ISL_1225513 presenta la mutación en RBD S459F, pero dicha mutación no pudo ser encontrada en la estructura cristalina 6VXX, por lo que no pudo ser evaluada el cambio de estabilidad (Anexo 1). Sin embargo, no cuenta con otra mutación en la proteína S, lo cual significa que estos efectos de afinidad elevada se deben exclusivamente a la mutación, y a pesar de ser una mutación de peligro, no está presente en la población de una manera significativa (0.1%) (Anexo 1). De igual manera, a nivel mundial solo ha sido registrada en un total de 1930 secuencias según GISAIID, indicando que es de bajo peligro. Sin embargo, no es debe tomarse a la ligera, ya que si cuenta con mayor afinidad por ACE2, por lo que sería de tomar en cuenta a la hora de realizar una vigilancia genómica

EPI_ISL_1811226 cuenta con la mutación en RBD E484Q, que al igual que la mutación anteriormente mencionada, no pudo ser evaluada (ocurre lo mismo con otra mutación en su secuencia, A845S) (Anexo 1). Sin embargo, también cuenta con D614G que se ha demostrado que presenta una mayor afinidad por ACE2, y es levemente desestabilizante; lo que es parte de la causa del efecto de mayor afinidad por ACE2, pero muy posiblemente es a causa de una sinergia de estas tres mutaciones (Cuadro 5) (Ozono *et al.*, 2021). Por otro lado, esta secuencia es del linaje B.1.1.328 (de mediados del 2020) está en su mayoría en Costa Rica, pero realmente no es de preocupación, dado a que a nivel mundial se han detectado 46 secuencias (Emily Scher, n.d.; Rambaut *et al.*, 2020).

EPI_ISL_1196430 además de la mutación levemente desestabilizante S494P (>1 kcal/mol) (Cuadro 5), cuenta con D614G, P681H y T716I. De estas, es levemente desestabilizante (>1 kcal/mol) (Cuadro 5), P681H no está en el estudio (Anexo 1), y T716I presenta descoordinación

entre DeepDgg (-0.02 kcal/mol) y mCSM (0.08 kcal/mol). Sin embargo, el efecto del cambio de estabilidad de esta última es tan bajo en cualquiera de los programas que realmente no podría ser significativo en la elevada afinidad sobre ACE2. Es por ello por lo que la mejorada afinidad por ACE2 podría deberse a un efecto sinérgico de S494P y D614G (y P681H, pero no fue posible determinar su efecto en la estructura). Muy probablemente esta mejora en la afinidad se debe a la presencia de prolina en la posición 494, llegando a afectar la estructura secundaria (la hélice α). Esta desestabilización es el resultado de la falta del átomo de H de la amida en su estructura principal, lo que evita que la prolina participe en la estabilización de los enlaces de hidrógeno (Deller *et al.*, 2016; Nick Pace *et al.*, 2014; Sticke *et al.*, 1992). Esta secuencia pertenece al linaje B.1.575, pero realmente no es de importancia en la región, siendo identificada únicamente 6 veces (solamente en Costa Rica) y 3 veces en este estudio (Anexo 1) (Emily Scher, n.d.; Rambaut *et al.*, 2020). A pesar de la elevada afinidad por ACE2 que el complejo de referencia, no hay necesidad de mantener una vigilancia de este linaje en Centroamérica.

Al tratarse de EVdW, EPI_ISL_4029266, EPI_ISL_1502939, EPI_ISL_3275178, EPI_ISL_527757, EPI_ISL_3835382, EPI_ISL_4029280, EPI_ISL_1502908, EPI_ISL_1527015, EPI_ISL_3275226, EPI_ISL_2157577, EPI_ISL_3275319, EPI_ISL_1502946 y EPI_ISL_1822780 se muestran resultados menores al complejo. En estos casos la EE es mayor, lo que indica que las atracciones más débiles juegan un papel mayor en estos complejos que las fuerzas hidrofóbicas, creando un posible efecto desestabilizante. Un caso similar, es posible observarse con EPI_ISL_3275184, EPI_ISL_4029266, EPI_ISL_1502939, EPI_ISL_3275178, EPI_ISL_527757, EPI_ISL_3835382, EPI_ISL_4029280, EPI_ISL_1502908, EPI_ISL_1527015, EPI_ISL_3275226, EPI_ISL_2157577, EPI_ISL_3275319, EPI_ISL_1502946 y EPI_ISL_1822780, donde tienen menor EE y mayor EVdW, pero en su mayoría son más inestables (estando las diferencias dentro del rango de incertidumbre). De igual modo, estos últimos complejos todos tienen una mayor EU que el complejo de referencia, posiblemente por este efecto de inestabilidad dentro de la estructura.

C. ANÁLISIS PREDICTIVO DEL PANORAMA DE GLICOSILACIÓN

A excepción de 6 secuencias, todas presentan GLO en las posiciones 673 y 686 (97.1%, Cuadro 10). Sin embargo, la tercera posición de referencia 673 está presente únicamente un 84.5%. Esto coincide con lo que reporta la literatura, donde no siempre todos los sequones presentes estarán glicosilados, dando un panorama de microheterogeneidad en la proteína S. No es posible determinar qué fue lo que causó esta falta de GLO, debido a que no siempre se presentan mutaciones próximas a los sequones. Una causa probable es la presencia de mutaciones desestabilizantes en la proteína (Cuadro 7), afectando la conformación del péptido y así afectando la disponibilidad superficial hacia las glicosiltransferasas. Adicionalmente, podría ser efecto del metabolismo propio de cada célula lo que afectó el panorama de glicosilación, dando un efecto deletéreo de las GLO (Freeze & Ng, 2013; Varki *et al.*, 2017). Esto podría afectar de forma negativa al reconocimiento celular de esta proteína, posiblemente llegando a evadir el sistema inmune. Dicho fenómeno es observado para los sitios adicionales predichos 19, 22, 26, 29, 250, 688 y 689 (Cuadro 12). De manera similar como con las GLO, las GLN también fueron afectadas, dando 203/206 sequones de referencia glicosilados. Adicionalmente, se detectaron como positivos 2 sequones, en la posición 20 y 188, dando así un patrón modificado de GLN en la proteína S, aunque en menor grado que GLO. Sin embargo, estas PTM juegan un rol muy importante en la función de la proteína S, dado a que está altamente glicosilada (en mayor grado GLN que GLO) puede afectar su interacción a nivel extracelular e intracelular (Walls *et al.*, 2020; Zhao *et al.*, 2021).

Como es evidente, las mutaciones no solo influyen directamente en la afinidad con ACE2, también influyen en el patrón de glicosilación de la proteína S. En gran medida el principal efecto que tiene las mutaciones es afectar la conformación de la proteína al desestabilizarla llevando a cambios en torción, área disponible para interacción, y cambiando la disponibilidad de interacción de algunos residuos al afectar al estructura secundaria (Kazlauskas, 2018; Kuroda & Tsumoto, 2020; Nick Pace *et al.*, 2014). Esto tiene grandes implicaciones, ya que los oligosacáridos contribuyen al plegamiento de la proteína S, inciden en el cebado de las proteasas del huésped y regulan el reconocimiento de anticuerpos (Shajahan *et al.*, 2021; Walls *et al.*, 2020; Yuan *et al.*, 2020). Los glucanos también desempeñan diversas funciones en las superficies de las células huésped y los virus. Aunque varios mecanismos de unión virus-hospedador incluyen directamente la interacción proteína-proteína, los carbohidratos sirven de manera similar como receptores primarios o correceptores que contribuyen al tropismo celular y la restricción del virus por parte del hospedador (Raman *et al.*, 2016). Esto quiere decir que las interacciones entre superficies enriquecidas por glucanos complejos y también son reconocidas como proteínas de unión a glucanos (GBP), jugando un papel sustancial en la infección por varios virus (Lin *et al.*, 2020). Se ha informado de una glicosilación extensa en la interfaz del complejo RBD-ACE2, lo que demuestra que los glicanos desempeñan un papel en la modulación de las interacciones. La entrada viral en las células hospedadoras se inicia mediante la interacción del virión con factores de unión a la superficie celular y los glucanos de la célula hospedadora a menudo sirven como las líneas frontales de los contactos iniciales para facilitar la unión de alta afinidad siguiente a los receptores específicos del virus (Koehler *et al.*, 2020). Esto expresa la preocupación por los cambios en la modulación del patrón de glicosilación afectados en parte por las mutaciones encontradas.

D. ANÁLISIS DE RECOMBINACIÓN

Los resultados de los gráficos de similitud demuestran que hay evidencia de recombinación en las secuencias de Centroamérica utilizadas en esta estudio. No existe realmente una disimilitud significativa, siendo en más de 99.7% similares entre sí. Esto concuerda con la literatura, donde estudios recientes de SC2 muestran niveles bajos de recombinación. Un estudio del 2021 con gran parte de las secuencias disponibles en la base de datos de GISAID encontró eventos de recombinación a nivel mundial, pero en únicamente 606 secuencias. Ellos utilizaron 1.6 millones de secuencias, y si se usa esta razón para predecir recombinación en la población de Centroamérica, se tendría un total de 0.7 secuencias (Turkahia *et al.*, 2021; VanInsberghe *et al.*, 2021). En otras palabras, no sería detectable de ninguna manera, por lo que se puede decir no existe evidencia de que haya eventos de recombinación detectables. Sin embargo, dado a la baja cantidad de secuencias disponibles de alta calidad no se puede decir con certeza esto, por lo que es necesario tener más información genética de cada país de manera continua, para tener así un panorama completo en el tiempo.

VIII. CONCLUSIONES

- A. La mutación que presenta un mayor efecto desestabilizante se encuentra en muy baja frecuencia en la población. De manera inversa, las mutaciones con un $\Delta\Delta G$ cercano a 0 están en mayor proporción. Sin embargo, esto no significa que no puedan afectar a la estructura tridimensional, sino que pueden tener un efecto de manera más leve en la región de Centroamérica, pero como mencionado anteriormente, los resultados pueden ser afectados por la cantidad de secuencias disponibles por país.
- B. Al analizar el comportamiento de las mutaciones a nivel de interacción proteína-proteína fue posible identificar que si hay modelos que presentan mayor afinidad al receptor ACE2. Evidenciado por los análisis de acoplamiento, la combinación de V503F, D614G y T1117I tienen un efecto significativo sobre la proteína S, llevando a 87 veces menor tener energía de unión la secuencia de referencia de SARS-CoV-2. Nuevamente, se encuentra en baja proporción, pero no significa que no esté presente en la región, sino que dentro de las muestras utilizadas para secuenciación estuvo una única vez. Adicional a esto, se puede decir que S459F eleva la afinidad por ACE2, dado por la falta de otras mutaciones en la secuencia. Sin embargo, está en baja proporción en la población, pero como mencionado anteriormente, no significa que no esté presente en la región. Las mutaciones en el linaje B.1.1.328 llevaron a una unión más estable y fuerte con ACE2, pero se han detectado pocas secuencias, por lo que no es de preocupación. Adicionalmente, el linaje B.1.575 también es más afín por ACE2, denotando nuevamente que a pesar de que las mutaciones tienen un efecto levemente desestabilizante pueden afectar la estructura y función de la proteína S.
- C. Se pudo observar un panorama modificado de glicosilaciones ligadas a N y a O, dando patrón de microheterogeneidad en la proteína S. Una causa probable es la presencia de mutaciones desestabilizantes en la proteína, lo cual podría afectar la disponibilidad por glicosiltransferasas. Estos cambios son preocupantes, dado a los posibles efectos sobre la antigenicidad de la proteína S y su interacción con ACE2.
- D. No fue posible identificar ningún evento de recombinación en las secuencias de Centroamérica, dado por una alta similitud entre las mismas. Esto concuerda con los hallazgos mundiales, que revelan que las tasas de recombinación son bajas y difíciles de detectar.

IX. RECOMENDACIONES

Múltiples resultados llevaban a la misma conclusión, el efecto que se observó se encuentra en baja proporción, pero no refleja que no esté de manera significativa en la región, sino que pocas secuencias muestran este efecto. La cantidad de secuencias de SARS-CoV-2 generadas para la región centroamericana es baja, por lo que no permite realizar una evaluación significativa del patrón de comportamiento del SARS-CoV-2 en el tiempo. Es un desafío para la región aumentar los esfuerzos en la vigilancia genómica. Sin embargo, los resultados de este estudio refuerzan la importancia de mantener un programa de vigilancia genómica de manera activa, que genere información epidemiológica que ayude a mitigar amenazas y controlar la dispersión del virus.

X. LITERATURA CITADA

- Aebi, M. (2013). N-linked protein glycosylation in the ER. In *Biochimica et Biophysica Acta - Molecular Cell Research* (Vol. 1833, Issue 11, pp. 2430–2437). Biochim Biophys Acta. <https://doi.org/10.1016/j.bbamcr.2013.04.001>
- Alejandra Tortorici, M., Walls, A. C., Lang, Y., Wang, C., Li, Z., Koerhuis, D., Boons, G. J., Bosch, B. J., Rey, F. A., de Groot, R. J., & Veesler, D. (2019). Structural basis for human coronavirus attachment to sialic acid receptors. *Nature Structural and Molecular Biology*, 26(6), 481–489. <https://doi.org/10.1038/s41594-019-0233-y>
- Allison, L. A. (2007). *Fundamental Molecular Biology*. Blackwell Publishing Inc.
- Andersen, K. G., Rambaut, A., Lipkin, W. I., Holmes, E. C., & Garry, R. F. (2020). The proximal origin of SARS-CoV-2. In *Nature Medicine* (Vol. 26, Issue 4, pp. 450–452). Nature Research. <https://doi.org/10.1038/s41591-020-0820-9>
- Arndt, A. L., Larson, B. J., & Hogue, B. G. (2010). A Conserved Domain in the Coronavirus Membrane Protein Tail Is Important for Virus Assembly. *Journal of Virology*, 84(21), 11418–11428. <https://doi.org/10.1128/jvi.01131-10>
- Asghari, A., Naseri, M., Safari, H., Saboory, E., & Parsamanesh, N. (2020). The Novel Insight of SARS-CoV-2 Molecular Biology and Pathogenesis and Therapeutic Options. *DNA and Cell Biology*, 39(10), 1741–1753. <https://doi.org/10.1089/dna.2020.5703>
- Astuti, I., & Ysrafil. (2020). Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2): An overview of viral structure and host response. *Diabetes and Metabolic Syndrome: Clinical Research and Reviews*, 14(4), 407–412. <https://doi.org/10.1016/j.dsx.2020.04.020>
- Betenbaugh, M. J., Tomiya, N., & Narang, S. (2007). Glycoengineering: Recombinant Glycoproteins. *Comprehensive Glycoscience: From Chemistry to Systems Biology*, 4–4, 607–642. <https://doi.org/10.1016/B978-044451967-2/00121-5>
- Bhardwaj, K., Sun, J., Holzenburg, A., Guarino, L. A., & Kao, C. C. (2006). RNA Recognition and Cleavage by the SARS Coronavirus Endoribonuclease. *Journal of Molecular Biology*, 361(2), 243–256. <https://doi.org/10.1016/j.jmb.2006.06.021>
- Bogoch, I. I., Watts, A., Thomas-Bachli, A., Huber, C., Kraemer, M. U. G., & Khan, K. (2020). Potential for global spread of a novel coronavirus from China. *Journal of Travel Medicine*, 27(2), 1–3. <https://doi.org/10.1093/jtm/taaa011>
- Borges, V., Alves, M. J., Amicone, M., Isidro, J., Zé-Zé, L., Duarte, S., Vieira, L., Guiomar, R., Gomes, J. P., & Gordo, I. (2021). Mutation rate of SARS-CoV-2 and emergence of mutators during experimental evolution. *BioRxiv*, 2021.05.19.444774. <https://doi.org/10.1101/2021.05.19.444774>

- Bosch, B. J., van der Zee, R., de Haan, C. A. M., & Rottier, P. J. M. (2003). The Coronavirus Spike Protein Is a Class I Virus Fusion Protein: Structural and Functional Characterization of the Fusion Core Complex. *Journal of Virology*, 77(16), 8801–8811. <https://doi.org/10.1128/jvi.77.16.8801-8811.2003>
- Cai, Y., Zhang, J., Xiao, T., Peng, H., Sterling, S., Walsh, R., Rawson, S., Rits-Volloch, S., & Chen, B. (2020). Distinct conformational states of SARS-CoV-2 spike protein. *Science*, 2020.05.16.099317. <https://doi.org/10.1101/2020.05.16.099317>
- Callard, F., & Perego, E. (2021). How and why patients made Long Covid. *Social Science and Medicine*, 268, 113426. <https://doi.org/10.1016/j.socscimed.2020.113426>
- Canedo-Marroquín, G., Saavedra, F., Andrade, C. A., Berrios, R. v., Rodríguez-Guilarte, L., Opazo, M. C., Riedel, C. A., & Kalergis, A. M. (2020). SARS-CoV-2: Immune Response Elicited by Infection and Development of Vaccines and Treatments. In *Frontiers in Immunology* (Vol. 11, p. 3259). Frontiers Media S.A. <https://doi.org/10.3389/fimmu.2020.569760>
- Cao, H., Wang, J., He, L., Qi, Y., & Zhang, J. Z. (2019). DeepDDG: Predicting the Stability Change of Protein Point Mutations Using Neural Networks. *Journal of Chemical Information and Modeling*, 59(4), 1508–1514. <https://doi.org/10.1021/ACS.JCIM.8B00697>
- Carter, J., & Saunders, V. (2007). *Virology: Principles and Applications*. John Wiley & Sons, Ltd. <https://books.google.com.gt/books?id=EKRgZCdr-74C&printsec=frontcover&dq=virology&hl=es&sa=X&ved=2ahUKEwijuL-G8MHrAhUE01kKHZw0AUsQ6AEwBXoEAcQAg#v=onepage&q=virology&f=false>
- Center for Disease Control and Prevention. (2020a). *How Coronavirus Spreads*. Coronavirus Disease 2019 (COVID-19). <https://www.cdc.gov/coronavirus/2019-ncov/prevent-getting-sick/how-covid-spreads.html>
- Center for Disease Control and Prevention. (2020b). *Symptoms of Coronavirus*. COVID-19. <https://www.cdc.gov/coronavirus/2019-ncov/symptoms-testing/symptoms.html>
- Centers for Disease Control and Prevention. (2020, December 30). *Information for Pediatric Healthcare Providers*. Healthcare Workers. <https://www.cdc.gov/coronavirus/2019-ncov/hcp/pediatric-hcp.html>
- Chen, Y., Liu, Q., & Guo, D. (2020). Emerging coronaviruses: Genome structure, replication, and pathogenesis. In *Journal of Medical Virology* (Vol. 92, Issue 4, pp. 418–423). John Wiley and Sons Inc. <https://doi.org/10.1002/jmv.25681>
- Chilamakuri, R., & Agarwal, S. (2021). COVID-19: Characteristics and Therapeutics. In *Cells* (Vol. 10, Issue 2). NLM (Medline). <https://doi.org/10.3390/cells10020206>

- Cottam, E. M., Whelband, M. C., & Wileman, T. (2014). Coronavirus NSP6 restricts autophagosome expansion. *Autophagy*, *10*(8), 1426–1441. <https://doi.org/10.4161/auto.29309>
- Coutard, B., Valle, C., de Lamballerie, X., Canard, B., Seidah, N. G., & Decroly, E. (2020). The spike glycoprotein of the new coronavirus 2019-nCoV contains a furin-like cleavage site absent in CoV of the same clade. *Antiviral Research*, *176*. <https://doi.org/10.1016/j.antiviral.2020.104742>
- CoV-Glue. (n.d.). *S replacement V503F*. Retrieved November 30, 2021, from <http://cov-glue.cvr.gla.ac.uk/#/project/replacement/S:V:503:F>
- Deller, M. C., Kong, L., & Rupp, B. (2016). Protein stability: a crystallographer's perspective. *Acta Crystallographica. Section F, Structural Biology Communications*, *72*(Pt 2), 72. <https://doi.org/10.1107/S2053230X15024619>
- Dhama, K., Khan, S., Tiwari, R., Sircar, S., Bhat, S., Malik, Y. S., Singh, K. P., Chaicumpa, W., Bonilla-Aldana, D. K., & Rodriguez-Morales, A. J. (2020). Coronavirus disease 2019–COVID-19. *Clinical Microbiology Reviews*, *33*(4), 1–48. <https://doi.org/10.1128/CMR.00028-20>
- Duan, L., Zheng, Q., Zhang, H., Niu, Y., Lou, Y., & Wang, H. (2020). The SARS-CoV-2 Spike Glycoprotein Biosynthesis, Structure, Function, and Antigenicity: Implications for the Design of Spike-Based Vaccine Immunogens. In *Frontiers in Immunology* (Vol. 11, p. 2593). Frontiers Media S.A. <https://doi.org/10.3389/fimmu.2020.576622>
- Duchene, S., Featherstone, L., Haritopoulou-Sinanidou, M., Rambaut, A., Lemey, P., & Baele, G. (2020). Temporal signal and the phylodynamic threshold of SARS-CoV-2. *Virus Evolution*, *6*(2). <https://doi.org/10.1093/ve/veaa061>
- Eaaswarkhanth, M., al Madhoun, A., & Al-Mulla, F. (2020). Could the D614G substitution in the SARS-CoV-2 spike (S) protein be associated with higher COVID-19 mortality? *International Journal of Infectious Diseases*, *96*, 459–460. <https://doi.org/10.1016/j.ijid.2020.05.071>
- Elflein, J. (2021, November 16). *Number of novel coronavirus (COVID-19) confirmed cases and deaths as of November 15, 2021, by region*. State of Health. <https://www.statista.com/statistics/1101373/novel-coronavirus-2019ncov-mortality-and-cases-worldwide-by-region/>
- Emily Scher. (n.d.). *SARS-CoV-2 lineages*. Retrieved October 16, 2021, from https://cov-lineages.org/lineage_list.html
- Etherington, G. J., Dicks, J., & Roberts, I. N. (2005). Recombination Analysis Tool (RAT): a program for the high-throughput detection of recombination. *Bioinformatics*, *21*(3), 278–281. <https://doi.org/10.1093/bioinformatics/bth500>

- Fang, S. G., Shen, H., Wang, J., Tay, F. P. L., & Liu, D. X. (2008). Proteolytic processing of polyproteins 1a and 1ab between non-structural proteins 10 and 11/12 of Coronavirus infectious bronchitis virus is dispensable for viral replication in cultured cells. *Virology*, 379(2), 175–180. <https://doi.org/10.1016/j.virol.2008.06.038>
- Fehr, A. R., Athmer, J., Channappanavar, R., Phillips, J. M., Meyerholz, D. K., & Perlman, S. (2015). The nsp3 Macrodomain Promotes Virulence in Mice with Coronavirus-Induced Encephalitis. *Journal of Virology*, 89(3), 1523–1536. <https://doi.org/10.1128/jvi.02596-14>
- Fielding, B. C., Gunalan, V., Tan, T. H. P., Chou, C. F., Shen, S., Khan, S., Lim, S. G., Hong, W., & Tan, Y. J. (2006). Severe acute respiratory syndrome coronavirus protein 7a interacts with hSGT. *Biochemical and Biophysical Research Communications*, 343(4), 1201–1208. <https://doi.org/10.1016/j.bbrc.2006.03.091>
- Flower, T. G., Buffalo, C. Z., Hooy, R. M., Allaire, M., Ren, X., & Hurley, J. H. (2021). Structure of SARS-cov-2 ORF8, a rapidly evolving immune evasion protein. *Proceedings of the National Academy of Sciences of the United States of America*, 118(2). <https://doi.org/10.1073/pnas.2021785118>
- Fratev, F. (2020). The N501Y and K417N mutations in the spike protein of SARS-CoV-2 alter the interactions with both hACE2 and human derived antibody: A Free energy of perturbation study. *BioRxiv*. <https://doi.org/10.1101/2020.12.23.424283>
- Freeze, H. H., & Ng, B. G. (2013). Glycomics, Glycobiology, and Glyco-Medicine. In *Genomic and Personalized Medicine* (Vol. 1, pp. 173–191). Elsevier Inc. <https://doi.org/10.1016/B978-0-12-382227-7.00015-X>
- Gadlage, M. J., Sparks, J. S., Beachboard, D. C., Cox, R. G., Doyle, J. D., Stobart, C. C., & Denison, M. R. (2010). Murine Hepatitis Virus Nonstructural Protein 4 Regulates Virus-Induced Membrane Modifications and Replication Complex Function. *Journal of Virology*, 84(1), 280–290. <https://doi.org/10.1128/jvi.01772-09>
- Garg, R. K., Paliwal, V. K., & Gupta, A. (2021). Encephalopathy in patients with COVID-19: A review. In *Journal of Medical Virology* (Vol. 93, Issue 1, pp. 206–222). John Wiley and Sons Inc. <https://doi.org/10.1002/jmv.26207>
- Graham, R. L., Sims, A. C., Brockway, S. M., Baric, R. S., & Denison, M. R. (2005). The nsp2 Replicase Proteins of Murine Hepatitis Virus and Severe Acute Respiratory Syndrome Coronavirus Are Dispensable for Viral Replication. *Journal of Virology*, 79(21), 13399–13411. <https://doi.org/10.1128/jvi.79.21.13399-13411.2005>
- Gribble, J., Stevens, L. J., Agostini, M. L., Anderson-Daniels, J., Chappell, J. D., Lu, X., Pruijssers, A. J., Routh, A. L., & Denison, M. R. (2021). The coronavirus proofreading exoribonuclease mediates extensive viral recombination. *PLOS Pathogens*, 17(1), e1009226. <https://doi.org/10.1371/JOURNAL.PPAT.1009226>

- Griffiths, A., Miller, J., Suzuki, D., Lewontin, R. C., & Gelbart, W. M. (2000). Recombination. In *An Introduction to Genetic Analysis*. (7th ed.). W. H. Freeman. <https://www.ncbi.nlm.nih.gov/books/NBK21889/>
- Gromiha, M. M. (2010). Protein Sequence Analysis. *Protein Bioinformatics*, 29–62. <https://doi.org/10.1016/B978-8-1312-2297-3.50002-3>
- Guillén, J., Pérez-Berná, A. J., Moreno, M. R., & Villalaín, J. (2005). Identification of the Membrane-Active Regions of the Severe Acute Respiratory Syndrome Coronavirus Spike Membrane Glycoprotein Using a 16/18-Mer Peptide Scan: Implications for the Viral Fusion Mechanism. *Journal of Virology*, 79(3), 1743–1752. <https://doi.org/10.1128/jvi.79.3.1743-1752.2005>
- Guo, L., Ren, L., Yang, S., Xiao, M., Chang, D., Yang, F., dela Cruz, C. S., Wang, Y., Wu, C., Xiao, Y., Zhang, L., Han, L., Dang, S., Xu, Y., Yang, Q. W., Xu, S. Y., Zhu, H. D., Xu, Y. C., Jin, Q., ... Wang, J. (2020). Profiling early humoral response to diagnose novel coronavirus disease (COVID-19). *Clinical Infectious Diseases*, 71(15), 778–785. <https://doi.org/10.1093/cid/ciaa310>
- Hoffmann, M., Kleine-Weber, H., & Pöhlmann, S. (2020). A Multibasic Cleavage Site in the Spike Protein of SARS-CoV-2 Is Essential for Infection of Human Lung Cells. *Molecular Cell*, 78(4), 779–784.e5. <https://doi.org/10.1016/j.molcel.2020.04.022>
- Hong, H., Wang, Y., Chung, H. T., & Chen, C. J. (2020). Clinical characteristics of novel coronavirus disease 2019 (COVID-19) in newborns, infants and children. In *Pediatrics and Neonatology* (Vol. 61, Issue 2, pp. 131–132). Elsevier (Singapore) Pte Ltd. <https://doi.org/10.1016/j.pedneo.2020.03.001>
- Honorato, R. v., Koukos, P. I., Jiménez-García, B., Tsaregorodtsev, A., Verlato, M., Giachetti, A., Rosato, A., & Bonvin, A. M. J. J. (2021). Structural Biology in the Clouds: The WeNMR-EOSC Ecosystem. *Frontiers in Molecular Biosciences*, 0, 708. <https://doi.org/10.3389/FMOLB.2021.729513>
- Hu, B., Guo, H., Zhou, P., & Shi, Z. L. (2021). Characteristics of SARS-CoV-2 and COVID-19. In *Nature Reviews Microbiology* (Vol. 19, Issue 3, pp. 141–154). Nature Research. <https://doi.org/10.1038/s41579-020-00459-7>
- Huang, C., Ito, N., Tseng, C.-T. K., & Makino, S. (2006). Severe Acute Respiratory Syndrome Coronavirus 7a Accessory Protein Is a Viral Structural Protein. *Journal of Virology*, 80(15), 7287–7294. <https://doi.org/10.1128/jvi.00414-06>
- Huang, C., Lokugamage, K. G., Rozovics, J. M., Narayanan, K., Semler, B. L., & Makino, S. (2011). SARS Coronavirus nsp1 Protein Induces Template-Dependent Endonucleolytic Cleavage of mRNAs: Viral mRNAs Are Resistant to nsp1-Induced RNA Cleavage. *PLoS Pathogens*, 7(12), e1002433. <https://doi.org/10.1371/journal.ppat.1002433>

- 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*, *41*(9), 1141–1149. <https://doi.org/10.1038/s41401-020-0485-4>
- Iwasaki, A., & Yang, Y. (2020). The potential danger of suboptimal antibody responses in COVID-19. In *Nature Reviews Immunology* (Vol. 20, Issue 6, pp. 339–341). Nature Research. <https://doi.org/10.1038/s41577-020-0321-6>
- Izda, V., Jeffries, M. A., & Sawalha, A. H. (2021). COVID-19: A review of therapeutic strategies and vaccine candidates. In *Clinical Immunology* (Vol. 222, p. 108634). Academic Press Inc. <https://doi.org/10.1016/j.clim.2020.108634>
- Jackson, B., Boni, M. F., Bull, M. J., Colleran, A., Colquhoun, R. M., Darby, A. C., Haldenby, S., Hill, V., Lucaci, A., McCrone, J. T., Nicholls, S. M., O'Toole, Á., Pacchiarini, N., Poplawski, R., Scher, E., Todd, F., Webster, H. J., Whitehead, M., Wierzbicki, C., ... Rambaut, A. (2021). Generation and transmission of interlineage recombinants in the SARS-CoV-2 pandemic. *Cell*, *184*(20), 5179–5188.e8. <https://doi.org/10.1016/J.CELL.2021.08.014>
- Jia, Z., Yan, L., Ren, Z., Wu, L., Wang, J., Guo, J., Zheng, L., Ming, Z., Zhang, L., Lou, Z., & Rao, Z. (2019). Delicate structural coordination of the Severe Acute Respiratory Syndrome coronavirus Nsp13 upon ATP hydrolysis. *Nucleic Acids Research*, *47*(12), 6538–6550. <https://doi.org/10.1093/nar/gkz409>
- Kanzawa, N., Nishigaki, K., Hayashi, T., Ishii, Y., Furukawa, S., Niino, A., Yasui, F., Kohara, M., Morita, K., Matsushima, K., Le, M. Q., Masuda, T., & Kannagi, M. (2006). Augmentation of chemokine production by severe acute respiratory syndrome coronavirus 3a/X1 and 7a/X4 proteins through NF- κ B activation. *FEBS Letters*, *580*(30), 6807–6812. <https://doi.org/10.1016/j.febslet.2006.11.046>
- Kazlauskas, R. (2018). Engineering more stable proteins. *Chemical Society Reviews*, *47*(24), 9026–9045. <https://doi.org/10.1039/C8CS00014J>
- Kelley, L. A., Mezulis, S., Yates, C. M., Wass, M. N., & Sternberg, M. J. E. (2015). The Pyre2 web portal for protein modeling, prediction and analysis. *Nature Protocols*, *10*(6), 845–858. <https://doi.org/10.1038/nprot.2015.053>
- Koehler, M., Delguste, M., Sieben, C., Gillet, L., & Alsteens, D. (2020). Initial Step of Virus Entry: Virion Binding to Cell-Surface Glycans. *Annual Review of Virology*, *7*(1), 143–165. <https://doi.org/10.1146/ANNUREV-VIROLOGY-122019-070025>
- Korber, B., Fischer, W. M., Gnanakaran, S., Yoon, H., Theiler, J., Abfalterer, W., Hengartner, N., Giorgi, E. E., Bhattacharya, T., Foley, B., Hastie, K. M., Parker, M. D., Partridge, D. G., Evans, C. M., Freeman, T. M., de Silva, T. I., Angyal, A., Brown, R. L., Carrilero, L., ... Montefiori, D. C. (2020). Tracking Changes in SARS-CoV-2 Spike: Evidence that D614G

- Increases Infectivity of the COVID-19 Virus. *Cell*, 182(4), 812-827.e19. <https://doi.org/10.1016/j.cell.2020.06.043>
- Kufareva, I., & Abagyan, R. (2012). Methods of protein structure comparison. *Methods in Molecular Biology (Clifton, N.J.)*, 857, 231. https://doi.org/10.1007/978-1-61779-588-6_10
- Kuroda, D., & Tsumoto, K. (2020). Engineering Stability, Viscosity, and Immunogenicity of Antibodies by Computational Design. *Journal of Pharmaceutical Sciences*, 109(5), 1631–1651. <https://doi.org/10.1016/J.XPHS.2020.01.011>
- Lei, J., Kusov, Y., & Hilgenfeld, R. (2018). Nsp3 of coronaviruses: Structures and functions of a large multi-domain protein. In *Antiviral Research* (Vol. 149, pp. 58–74). Elsevier B.V. <https://doi.org/10.1016/j.antiviral.2017.11.001>
- Leung, K., Shum, M. H., Leung, G. M., Lam, T. T., & Wu, J. T. (2021). Early transmissibility assessment of the N501Y mutant strains of SARS-CoV-2 in the United Kingdom, October to November 2020. *Eurosurveillance*, 26(1), 2002106. <https://doi.org/10.2807/1560-7917.ES.2020.26.1.2002106>
- Li, Q., Wu, J., Nie, J., Li, X., Huang, W., Correspondence, Y. W., Zhang, L., Hao, H., Liu, S., Zhao, C., Zhang, Q., Liu, H., Nie, L., Qin, H., Wang, M., Lu, Q., Li, X., Sun, Q., Liu, J., ... Wang, Y. (2020). The Impact of Mutations in SARS-CoV-2 Spike on Viral Infectivity and Antigenicity. *Cell*, 182, 1284-1294.e9. <https://doi.org/10.1016/j.cell.2020.07.012>
- Li, W., Moore, M. J., Vasllieva, N., Sui, J., Wong, S. K., Berne, M. A., Somasundaran, M., Sullivan, J. L., Luzuriaga, K., Greeneugh, T. C., Choe, H., & Farzan, M. (2003). Angiotensin-converting enzyme 2 is a functional receptor for the SARS coronavirus. *Nature*, 426(6965), 450–454. <https://doi.org/10.1038/nature02145>
- Li, X., Giorg, E. E., Marichannegowda, M. H., Foley, B., Xiao, C., Kong, X. P., Chen, Y., Gnanakaran, S., Korber, B., & Gao, F. (2020a). Emergence of SARS-CoV-2 through recombination and strong purifying selection. *Science Advances*, 6(27). <https://doi.org/10.1126/SCIADV.ABB9153>
- Li, X., Giorg, E. E., Marichannegowda, M. H., Foley, B., Xiao, C., Kong, X. P., Chen, Y., Gnanakaran, S., Korber, B., & Gao, F. (2020b). Emergence of SARS-CoV-2 through recombination and strong purifying selection. *Science Advances*, 6(27), eabb9153. <https://www.science.org>
- Lin, B., Qing, X., Liao, J., & Zhuo, K. (2020). Role of Protein Glycosylation in Host-Pathogen Interaction. *Cells* 2020, Vol. 9, Page 1022, 9(4), 1022. <https://doi.org/10.3390/CELLS9041022>
- Livesey, B. J., & Marsh, J. A. (2020). Using deep mutational scanning to benchmark variant effect predictors and identify disease mutations. *Molecular Systems Biology*, 16(7). <https://doi.org/10.15252/msb.20199380>

- Lole, K. S., Bollinger, R. C., Paranjape, R. S., Gadkari, D., Kulkarni, S. S., Novak, N. G., Ingersoll, R., Sheppard, H. W., & Ray, S. C. (1999). Full-length human immunodeficiency virus type 1 genomes from subtype C-infected seroconverters in India, with evidence of intersubtype recombination. *Journal of Virology*, *73*(1), 152–160. <https://doi.org/10.1128/JVI.73.1.152-160.1999>
- LR, B., & MM, L. (1991). Random nature of coronavirus RNA recombination in the absence of selection pressure. *Virology*, *185*(1), 441–445. [https://doi.org/10.1016/0042-6822\(91\)90795-D](https://doi.org/10.1016/0042-6822(91)90795-D)
- Maiorov, V. N., & Crippen, G. M. (1994). Significance of root-mean-square deviation in comparing three-dimensional structures of globular proteins. *Journal of Molecular Biology*, *235*(2), 625–634. <https://doi.org/10.1006/JMBI.1994.1017>
- Masotti, A., Caputo, V., da Sacco, L., Pizzuti, A., Dallapiccola, B., & Bottazzo, G. F. (2009). Quantification of small non-coding RNAs allows an accurate comparison of miRNA expression profiles. *Journal of Biomedicine & Biotechnology*, *2009*, 659028. <https://doi.org/10.1155/2009/659028>
- Michel, C. J., Mayer, C., Poch, O., & Thompson, J. D. (2020). Characterization of accessory genes in coronavirus genomes. *Virology Journal*, *17*(1), 131. <https://doi.org/10.1186/s12985-020-01402-1>
- Motozono, C., Toyoda, M., Zahradnik, J., Saito, A., Nasser, H., Tan, T. S., Ngare, I., Kimura, I., Uriu, K., Kosugi, Y., Yue, Y., Shimizu, R., Ito, J., Torii, S., Yonekawa, A., Shimono, N., Nagasaki, Y., Minami, R., Toya, T., ... Sato, K. (2021). SARS-CoV-2 spike L452R variant evades cellular immunity and increases infectivity. *Cell Host and Microbe*, *29*(7), 1124–1136.e11. <https://doi.org/10.1016/J.CHOM.2021.06.006/ATTACHMENT/DD789B27-4E96-4989-9997-E1396688BFDE/MMC7.XLSX>
- National Health Institute. (2021). *Long-term effects of coronavirus (long COVID)* . Coronavirus (COVID-19). <https://www.nhs.uk/conditions/coronavirus-covid-19/long-term-effects-of-coronavirus-long-covid/>
- Nelemans, T., & Kikkert, M. (2019). Viral innate immune evasion and the pathogenesis of emerging RNA virus infections. In *Viruses* (Vol. 11, Issue 10). MDPI AG. <https://doi.org/10.3390/v11100961>
- Nelson, G., Buzko, O., Patricia, S., Niazi, K., Rabizadeh, S., & Soon-Shiong, P. (2021). Molecular dynamic simulation reveals E484K mutation enhances spike RBD-ACE2 affinity and the 1 combination of E484K, K417N and N501Y mutations (501Y.V2 variant) induces conformational 2 change greater than N501Y mutant alone, potentially resulting in an escape mutant. *BioRxiv*, 2021.01.13.426558. <https://doi.org/10.1101/2021.01.13.426558>

- Nick Pace, C., Martin Scholtz, J., & Grimsley, G. R. (2014). Forces Stabilizing Proteins. *FEBS Letters*, 588(14), 2177. <https://doi.org/10.1016/J.FEBSLET.2014.05.006>
- Ortega, J. T., Zambrano, J. L., Jastrzebska, B., Liprandi, F., Rangel, H. R., & Pujol, F. H. (2020). Understanding Severe Acute Respiratory Syndrome Coronavirus 2 Replication to Design Efficient Drug Combination Therapies. In *Intervirolgy* (Vol. 63, Issues 1–6, pp. 2–9). S. Karger AG. <https://doi.org/10.1159/000512141>
- Ou, X., Liu, Y., Lei, X., Li, P., Mi, D., Ren, L., Guo, L., Guo, R., Chen, T., Hu, J., Xiang, Z., Mu, Z., Chen, X., Chen, J., Hu, K., Jin, Q., Wang, J., & Qian, Z. (2020). Characterization of spike glycoprotein of SARS-CoV-2 on virus entry and its immune cross-reactivity with SARS-CoV. *Nature Communications*, 11(1), 1–12. <https://doi.org/10.1038/s41467-020-15562-9>
- Ozono, S., Zhang, Y., Ode, H., Sano, K., Tan, T. S., Imai, K., Miyoshi, K., Kishigami, S., Ueno, T., Iwatani, Y., Suzuki, T., & Tokunaga, K. (2021). SARS-CoV-2 D614G spike mutation increases entry efficiency with enhanced ACE2-binding affinity. *Nature Communications* 2021 12:1, 12(1), 1–9. <https://doi.org/10.1038/s41467-021-21118-2>
- Pace, C. N., Fu, H., Fryar, K. L., Landua, J., Trevino, S. R., Schell, D., Thurlkill, R. L., Imura, S., Scholtz, J. M., Gajiwala, K., Sevcik, J., Urbanikova, L., Myers, J. K., Takano, K., Hebert, E. J., Shirley, B. A., & Grimsley, G. R. (2014). Contribution of hydrogen bonds to protein stability. *Protein Science: A Publication of the Protein Society*, 23(5), 652. <https://doi.org/10.1002/PRO.2449>
- Pace, C. N., Fu, H., Fryar, K. L., Landua, J., Trevino, S. R., Shirley, B. A., Hendricks, M. M. N., Imura, S., Gajiwala, K., Scholtz, J. M., & Grimsley, G. R. (2011). Contribution of Hydrophobic Interactions to Protein Stability. *Journal of Molecular Biology*, 408(3), 514. <https://doi.org/10.1016/J.JMB.2011.02.053>
- Pancer, K., Milewska, A., Owczarek, K., Dabrowska, A., Kowalski, M., Łabaj, P. P., Branicki, W., Sanak, M., & Pyrc, K. (2020). The SARS-CoV-2 ORF10 is not essential in vitro or in vivo in humans. *PLOS Pathogens*, 16(12), e1008959. <https://doi.org/10.1371/journal.ppat.1008959>
- Pancera, M., Zhou, T., Druz, A., Georgiev, I. S., Soto, C., Gorman, J., Huang, J., Acharya, P., Chuang, G. Y., Ofek, G., Stewart-Jones, G. B. E., Stuckey, J., Bailer, R. T., Joyce, M. G., Louder, M. K., Tumba, N., Yang, Y., Zhang, B., Cohen, M. S., ... Kwong, P. D. (2014). Structure and immune recognition of trimeric pre-fusion HIV-1 Env. *Nature*, 514(7253), 455–461. <https://doi.org/10.1038/nature13808>
- Pang, C. N. I., & Wilkins, M. R. (2018). Protein post-translational modification prediction. In *Encyclopedia of Bioinformatics and Computational Biology: ABC of Bioinformatics* (Vols. 1–3, pp. 15–27). Elsevier. <https://doi.org/10.1016/B978-0-12-809633-8.20265-3>
- Pekosz, A., Schaecher, S. R., Diamond, M. S., Fremont, D. H., Sims, A. C., & Baric, R. S. (2006). Structure, expression, and intracellular localization of the SARS-CoV accessory proteins 7a

- and 7b. *Advances in Experimental Medicine and Biology*, 581, 115–120. https://doi.org/10.1007/978-0-387-33012-9_20
- Pires, D. E. v., Ascher, D. B., & Blundell, T. L. (2014). mCSM: predicting the effects of mutations in proteins using graph-based signatures. *Bioinformatics*, 30(3), 335–342. <https://doi.org/10.1093/BIOINFORMATICS/BTT691>
- Plante, J. A., Liu, Y., Liu, J., Xia, H., Johnson, B. A., Lokugamage, K. G., Zhang, X., Muruato, A. E., Zou, J., Fontes-Garfias, C. R., Mirchandani, D., Scharton, D., Bilello, J. P., Ku, Z., An, Z., Kalveram, B., Freiberg, A. N., Menachery, V. D., Xie, X., ... Shi, P. Y. (2020). Spike mutation D614G alters SARS-CoV-2 fitness. *Nature*, 1–6. <https://doi.org/10.1038/s41586-020-2895-3>
- Portela Sousa, C., & Brites, C. (2020). Immune response in SARS-CoV-2 infection: the role of interferons type I and type III. In *Brazilian Journal of Infectious Diseases* (Vol. 24, Issue 5, pp. 428–433). Elsevier Editora Ltda. <https://doi.org/10.1016/j.bjid.2020.07.011>
- Prompetchara, E., Ketloy, C., & Palaga, T. (2020). Immune responses in COVID-19 and potential vaccines: Lessons learned from SARS and MERS epidemic. *Asian Pacific Journal of Allergy and Immunology*, 38(1), 1–9. <https://doi.org/10.12932/AP-200220-0772>
- Rabi, F., al Zoubi, M., Kasasbeh, G., Salameh, D., & Al-Nasser, A. (2020). SARS-CoV-2 and Coronavirus Disease 2019: What We Know So Far. *Pathogens*, 9(3), 231. <https://doi.org/10.3390/pathogens9030231>
- Raman, R., Tharakaraman, K., Sasisekharan, V., & Sasisekharan, R. (2016). Glycan–protein interactions in viral pathogenesis. *Current Opinion in Structural Biology*, 40, 153–162. <https://doi.org/10.1016/J.SBI.2016.10.003>
- Rambaut, A., Holmes, E. C., O’Toole, Á., Hill, V., McCrone, J. T., Ruis, C., du Plessis, L., & Pybus, O. G. (2020). A dynamic nomenclature proposal for SARS-CoV-2 lineages to assist genomic epidemiology. *Nature Microbiology* 2020 5:11, 5(11), 1403–1407. <https://doi.org/10.1038/s41564-020-0770-5>
- Read, J. M., Bridgen, J. R. E., Cummings, D. A. T., Ho, A., & Jewell, C. P. (2020). Novel coronavirus 2019-nCoV: Early estimation of epidemiological parameters and epidemic predictions. In *medRxiv* (p. 2020.01.23.20018549). medRxiv. <https://doi.org/10.1101/2020.01.23.20018549>
- Reeb, J., Wirth, T., & Rost, B. (2020). Variant effect predictions capture some aspects of deep mutational scanning experiments. *BMC Bioinformatics*, 21(1), 107. <https://doi.org/10.1186/s12859-020-3439-4>
- Reguera, J., Mudgal, G., Santiago, C., & Casasnovas, J. M. (2014). A structural view of coronavirus-receptor interactions. *Virus Research*, 194, 3–15. <https://doi.org/10.1016/j.virusres.2014.10.005>

- Riou, J., & Althaus, C. L. (2020). Pattern of early human-to-human transmission of Wuhan 2019 novel coronavirus (2019-nCoV), December 2019 to January 2020. *Eurosurveillance*, 25(4), 2000058. <https://doi.org/10.2807/1560-7917.ES.2020.25.4.2000058>
- Sainz, B., Rausch, J. M., Gallaher, W. R., Garry, R. F., & Wimley, W. C. (2005). Identification and Characterization of the Putative Fusion Peptide of the Severe Acute Respiratory Syndrome-Associated Coronavirus Spike Protein. *Journal of Virology*, 79(11), 7195–7206. <https://doi.org/10.1128/jvi.79.11.7195-7206.2005>
- Samudrala, P. K., Kumar, P., Choudhary, K., Thakur, N., Wadekar, G. S., Dayaramani, R., Agrawal, M., & Alexander, A. (2020). Virology, pathogenesis, diagnosis and in-line treatment of COVID-19. *European Journal of Pharmacology*, 883. <https://doi.org/10.1016/j.ejphar.2020.173375>
- Sanda, M., Morrison, L., & Goldman, R. (2021). N-and O-Glycosylation of the SARS-CoV-2 Spike Protein. *Analytical Chemistry*, 93(4), 2003–2009. <https://doi.org/10.1021/acs.analchem.0c03173>
- Schmieder, R., & Edwards, R. (2011). Quality control and preprocessing of metagenomic datasets. *Bioinformatics*, 27(6), 863–864. <https://doi.org/10.1093/bioinformatics/btr026>
- Shah, V. K., Firmal, P., Alam, A., Ganguly, D., & Chattopadhyay, S. (2020). Overview of Immune Response During SARS-CoV-2 Infection: Lessons From the Past. In *Frontiers in Immunology* (Vol. 11). Frontiers Media S.A. <https://doi.org/10.3389/fimmu.2020.01949>
- Shajahan, A., Archer-Hartmann, S., Supekar, N. T., Gleinich, A. S., Heiss, C., & Azadi, P. (2021). Comprehensive characterization of N- and O- glycosylation of SARS-CoV-2 human receptor angiotensin converting enzyme 2. *Glycobiology*, 31(4), 410–424. <https://doi.org/10.1093/GLYCOB/CWAA101>
- Shajahan, A., Supekar, N. T., Gleinich, A. S., & Azadi, P. (2020). Deducing the N- And O-glycosylation profile of the spike protein of novel coronavirus SARS-CoV-2. *Glycobiology*, 30(12), 981–988. <https://doi.org/10.1093/glycob/cwaa042>
- Shi, P., Su, Y., Li, R., Liang, Z., Dong, S., & Huang, J. (2019). PEDV nsp16 negatively regulates innate immunity to promote viral proliferation. *Virus Research*, 265, 57–66. <https://doi.org/10.1016/j.virusres.2019.03.005>
- Simon-Loriere, E., & Holmes, E. C. (2011). Why do RNA viruses recombine? In *Nature Reviews Microbiology* (Vol. 9, Issue 8, pp. 617–626). Nature Publishing Group. <https://doi.org/10.1038/nrmicro2614>
- Steentoft, C., Vakhrushev, S. Y., Joshi, H. J., Kong, Y., Vester-Christensen, M. B., Schjoldager, K. T. B. G., Lavrsen, K., Dabelsteen, S., Pedersen, N. B., Marcos-Silva, L., Gupta, R., Paul Bennett, E., Mandel, U., Brunak, S., Wandall, H. H., Levery, S. B., & Clausen, H. (2013).

- Precision mapping of the human O-GalNAc glycoproteome through SimpleCell technology. *EMBO Journal*, 32(10), 1478–1488. <https://doi.org/10.1038/emboj.2013.79>
- Sternberg, A., & Naujokat, C. (2020). Structural features of coronavirus SARS-CoV-2 spike protein: Targets for vaccination. *Life Sciences*, 257, 118056. <https://doi.org/10.1016/j.lfs.2020.118056>
- Sticke, D. F., Presta, L. G., Dill, K. A., & Rose, G. D. (1992). Hydrogen bonding in globular proteins. *Journal of Molecular Biology*, 226(4), 1143–1159. [https://doi.org/10.1016/0022-2836\(92\)91058-W](https://doi.org/10.1016/0022-2836(92)91058-W)
- Sun, X., Tse, L. v., Ferguson, A. D., & Whittaker, G. R. (2010). Modifications to the Hemagglutinin Cleavage Site Control the Virulence of a Neurotropic H1N1 Influenza Virus. *Journal of Virology*, 84(17), 8683–8690. <https://doi.org/10.1128/jvi.00797-10>
- Synowiec, A., Szczepa, A., Barreto-Duran, E., Lie, L. K., & Pyrc, K. (2021a). *Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2): a Systemic Infection*. <https://doi.org/10.1128/CMR>
- Synowiec, A., Szczepa, A., Barreto-Duran, E., Lie, L. K., & Pyrc, K. (2021b). *Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2): a Systemic Infection*. <https://doi.org/10.1128/CMR>
- Tangudu, C., Olivares, H., Netland, J., Perlman, S., & Gallagher, T. (2007). Severe Acute Respiratory Syndrome Coronavirus Protein 6 Accelerates Murine Coronavirus Infections. *Journal of Virology*, 81(3), 1220–1229. <https://doi.org/10.1128/jvi.01515-06>
- ThermoFisher Scientific. (2008). *Overview of Post-Translational Modification*. Overview of Post-Translational Modifications (PTMs). [//www.thermofisher.com/ht/en/home/life-science/protein-biology/protein-biology-learning-center/protein-biology-resource-library/pierce-protein-methods/overview-post-translational-modification.html](http://www.thermofisher.com/ht/en/home/life-science/protein-biology/protein-biology-learning-center/protein-biology-resource-library/pierce-protein-methods/overview-post-translational-modification.html)
- Thevarajan, I., Nguyen, T. H. O., Koutsakos, M., Druce, J., Caly, L., van de Sandt, C. E., Jia, X., Nicholson, S., Catton, M., Cowie, B., Tong, S. Y. C., Lewin, S. R., & Kedzierska, K. (2020). Breadth of concomitant immune responses prior to patient recovery: a case report of non-severe COVID-19. In *Nature Medicine* (Vol. 26, Issue 4, pp. 453–455). Nature Research. <https://doi.org/10.1038/s41591-020-0819-2>
- Turkahia, Y., Thornlow, B., Hinrichs, A., McBroome, J., Ayala, N., Ye, C., Maio, N. de, Haussler, D., Lanfear, R., & Corbett-Detig, R. (2021). Pandemic-Scale Phylogenomics Reveals Elevated Recombination Rates in the SARS-CoV-2 Spike Region. *BioRxiv*, 2021.08.04.455157. <https://doi.org/10.1101/2021.08.04.455157>
- Uversky, V. N. (2013). Posttranslational Modification. In *Brenner's Encyclopedia of Genetics: Second Edition* (pp. 425–430). Elsevier Inc. <https://doi.org/10.1016/B978-0-12-374984-0.01203-1>

- van Zundert, G. C. P., & Bonvin, A. M. J. J. (2014). Modeling protein–protein complexes using the HADDOCK webserver “modeling protein complexes with HADDOCK.” *Methods in Molecular Biology*, *1137*, 163–179. https://doi.org/10.1007/978-1-4939-0366-5_12
- VanInsberghe, D., Neish, A. S., Lowen, A. C., & Koelle, K. (2021). Recombinant SARS-CoV-2 genomes are currently circulating at low levels. *BioRxiv*, 2020.08.05.238386. <https://doi.org/10.1101/2020.08.05.238386>
- Varki, A., Cummings, R. D., Esko, J. D., Stanley, P., Hart, G. W., Aebi, M., Darvill, A. G., Kinoshita, T., Packer, N. H., Prestegard, J. H., Schnaar, R. L., & Seeberger, P. H. (2017). Essentials of Glycobiology. In *Cold Spring Harbor (NY)* (3rd ed.). Cold Spring Harbor Laboratory Press. <https://www.ncbi.nlm.nih.gov/books/NBK310274/>
- Viner, R. M., Mytton, O. T., Bonell, C., Melendez-Torres, G. J., Ward, J., Hudson, L., Waddington, C., Thomas, J., Russell, S., van der Klis, F., Koirala, A., Ladhani, S., Panovska-Griffiths, J., Davies, N. G., Booy, R., & Eggo, R. M. (2021). Susceptibility to SARS-CoV-2 Infection among Children and Adolescents Compared with Adults: A Systematic Review and Meta-Analysis. In *JAMA Pediatrics* (Vol. 175, Issue 2, pp. 143–156). American Medical Association. <https://doi.org/10.1001/jamapediatrics.2020.4573>
- Walls, A. C., Park, Y. J., Tortorici, M. A., Wall, A., McGuire, A. T., & Velesler, 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>
- Wand, S., Xiao, F., He, L., Yang, J., & Wei, H. (2021). Altered N-/O-glycosylation sites on the receptor-binding domain(RBD) in COVID-19 are related to coronavirus infection and pathogenesis. *Research Square*. <http://www.cbs.dtu.dk/services/NetNGlyc/>
- Wang, C., Xu, H., Lin, S., Deng, W., Zhou, J., Zhang, Y., Shi, Y., Peng, D., & Xue, Y. (2020). GPS 5.0: An Update on the Prediction of Kinase-specific Phosphorylation Sites in Proteins. *Genomics, Proteomics and Bioinformatics*, *18*(1), 72–80. <https://doi.org/10.1016/j.gpb.2020.01.001>
- Wang, L., & Xiang, Y. (2020). Spike Glycoprotein-Mediated Entry of SARS Coronaviruses. In *Viruses* (Vol. 12, Issue 11). NLM (Medline). <https://doi.org/10.3390/v12111289>
- Wang, M.-Y., Zhao, R., Gao, L.-J., Gao, X.-F., Wang, D.-P., & Cao, J.-M. (2020). SARS-CoV-2: Structure, Biology, and Structure-Based Therapeutics Development. *Frontiers in Cellular and Infection Microbiology*, *10*. <https://doi.org/10.3389/fcimb.2020.587269>
- Wang, N., Shang, J., Jiang, S., & Du, L. (2020). Subunit Vaccines Against Emerging Pathogenic Human Coronaviruses. In *Frontiers in Microbiology* (Vol. 11, p. 298). Frontiers Media S.A. <https://doi.org/10.3389/fmicb.2020.00298>
- Woo, H., Park, S.-J., Choi, Y. K., Park, T., Tanveer, M., Cao, Y., Kern, N. R., Lee, J., Yeom, M. S., Croll, T. I., Seok, C., & Im, W. (2020). Developing a Fully Glycosylated Full-Length

- SARS-CoV-2 Spike Protein Model in a Viral Membrane. *The Journal of Physical Chemistry B*, 124(33), 7128–7137. <https://doi.org/10.1021/ACS.JPCB.0C04553>
- World Health Organization. (2021). *COVID-19: vulnerable and high risk groups*. Information for the Public. <https://www.who.int/westernpacific/emergencies/covid-19/information/high-risk-groups>
- Wrapp, D., Wang, N., Corbett, K. S., Goldsmith, J. A., Hsieh, C. L., Abiona, O., Graham, B. S., & McLellan, J. S. (2020). Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation. In *bioRxiv*. bioRxiv. <https://doi.org/10.1101/2020.02.11.944462>
- Xu, W., Wang, M., Yu, D., & Zhang, X. (2020). Variations in SARS-CoV-2 Spike Protein Cell Epitopes and Glycosylation Profiles During Global Transmission Course of COVID-19. *Frontiers in Immunology*, 0, 2222. <https://doi.org/10.3389/FIMMU.2020.565278>
- Y, Y., Z, W., X, W., & SY, H. (2017). Addressing recent docking challenges: A hybrid strategy to integrate template-based and free protein-protein docking. *Proteins*, 85(3), 497–512. <https://doi.org/10.1002/PROT.25234>
- Yan, Y., Tao, H., He, J., & Huang, S.-Y. (2020). The HDock server for integrated protein–protein docking. *Nature Protocols* 2020 15:5, 15(5), 1829–1852. <https://doi.org/10.1038/s41596-020-0312-x>
- Yan, Y., Zhang, D., Zhou, P., Li, B., & Huang, S. (2017). HDock: a web server for protein-protein and protein-DNA/RNA docking based on a hybrid strategy. *Nucleic Acids Research*, 45(W1), W365–W373. <https://doi.org/10.1093/NAR/GKX407>
- Yang, J., Wang, W., Chen, Z., Lu, S., Yang, F., Bi, Z., Bao, L., Mo, F., Li, X., Huang, Y., Hong, W., Yang, Y., Zhao, Y., Ye, F., Lin, S., Deng, W., Chen, H., Lei, H., Zhang, Z., ... Wei, X. (2020). A vaccine targeting the RBD of the S protein of SARS-CoV-2 induces protective immunity. *Nature*, 586(7830), 572–577. <https://doi.org/10.1038/s41586-020-2599-8>
- Yates, C. M., Filippis, I., Kelley, L. A., & Sternberg, M. J. E. (2014). SuSPect: Enhanced Prediction of Single Amino Acid Variant (SAV) Phenotype Using Network Features. *Journal of Molecular Biology*, 426(14), 2692–2701. <https://doi.org/10.1016/J.JMB.2014.04.026>
- Yi, H. (2020). 2019 Novel Coronavirus Is Undergoing Active Recombination. *Clinical Infectious Diseases*, 71(15), 884–887. <https://doi.org/10.1093/CID/CIAA219>
- Yi, Y., Lagniton, P. N. P., Ye, S., Li, E., & Xu, R. H. (2020). COVID-19: What has been learned and to be learned about the novel coronavirus disease. *International Journal of Biological Sciences*, 16(10), 1753–1766. <https://doi.org/10.7150/ijbs.45134>
- Yuan, M., Wu, N. C., Zhu, X., Lee, C. C. D., So, R. T. Y., Lv, H., Mok, C. K. P., & Wilson, I. A. (2020). A highly conserved cryptic epitope in the receptor binding domains of SARS-CoV-2

- and SARS-CoV. *Science*, 368(6491), 630–633.
https://doi.org/10.1126/SCIENCE.ABB7269/SUPPL_FILE/PAP.PDF
- Yuki, K., Fujiogi, M., & Koutsogiannaki, S. (2020). COVID-19 pathophysiology: A review. In *Clinical Immunology* (Vol. 215). Academic Press Inc.
<https://doi.org/10.1016/j.clim.2020.108427>
- Zhai, Y., Sun, F., Li, X., Pang, H., Xu, X., Bartlam, M., & Rao, Z. (2005). Insights into SARS-CoV transcription and replication from the structure of the nsp7-nsp8 hexadecamer. *Nature Structural and Molecular Biology*, 12(11), 980–986. <https://doi.org/10.1038/nsmb999>
- Zhang, M., Gaschen, B., Blay, W., Foley, B., Haigwood, N., Kuiken, C., & Korber, B. (2004). Tracking global patterns of N-linked glycosylation site variation in highly variable viral glycoproteins: HIV, SIV, and HCV envelopes and influenza hemagglutinin. *Glycobiology*, 14(12), 1229–1246. <https://doi.org/10.1093/GLYCOB/CWH106>
- Zhang, X., & Wang, Y. (2016). Glycosylation Quality Control by the Golgi Structure. In *Journal of Molecular Biology* (Vol. 428, Issue 16, pp. 3183–3193). Academic Press.
<https://doi.org/10.1016/j.jmb.2016.02.030>
- Zhang, X.-Y., Guo, J., Wan, X., Zhou, J.-G., Jin, W.-P., Lu, J., Wang, W.-H., Yang, A.-N., Liu, D. X., Shi, Z.-L., Yuan, Z.-M., Li, X.-G., Meng, S.-L., Duan, K., Wang, Z.-J., Yang, X.-M., & Shen, S. (2020). Biochemical and antigenic characterization of the structural proteins and their post-translational modifications in purified SARS-CoV-2 virions of an inactivated vaccine candidate. *Emerging Microbes & Infections*, 9(1), 2653–2662.
<https://doi.org/10.1080/22221751.2020.1855945>
- Zhao, X., Chen, H., & Wang, H. (2021). Glycans of SARS-CoV-2 Spike Protein in Virus Infection and Antibody Production. *Frontiers in Molecular Biosciences*, 0, 53.
<https://doi.org/10.3389/FMOLB.2021.629873>
- Zhu, X., Fang, L., Wang, D., Yang, Y., Chen, J., Ye, X., Foda, M. F., & Xiao, S. (2017). Porcine deltacoronavirus nsp5 inhibits interferon- β production through the cleavage of NEMO. *Virology*, 502, 33–38. <https://doi.org/10.1016/j.virol.2016.12.005>

XI. ANEXOS

Anexo 1: Lista completa de frecuencia de mutaciones en la proteína S en las secuencias de Centroamérica del SARS-Cov-2. * Estas secuencias no pudieron ser analizadas con las estructuras cristalinas disponibles por los distintos programas (algunos residuos no pudieron ser resueltos dentro de los modelos), es por ello por lo que no están presentes en el estudio.

Mutación	Frecuencia	Porcentaje	Mutación	Frecuencia	Porcentaje	Mutación	Frecuencia	Porcentaje
D614G	1516	87.8	A348S	3	0.2	V227L	1	0.1
L452R	611	35.4	A522V	3	0.2	Q218H	1	0.1
D215Y	377	21.8	A1078S	3	0.2	G35C	1	0.1
N501Y	231	13.4	D253N	3	0.2	H69Y	1	0.1
P681H*	205	11.9	E1195A	3	0.2	G142L	1	0.1
T478K	161	9.3	G1219V	3	0.2	N211R	1	0.1
D950N	158	9.1	D950H	3	0.2	L922F	1	0.1
T19R	151	8.7	N1074S	3	0.2	A626V	1	0.1
P681R	149	8.6	N188K	2	0.1	P812S	1	0.1
T95I	146	8.5	H49Y	2	0.1	E340G	1	0.1
T1117I	133	7.7	N30H	2	0.1	L270H	1	0.1
E484K*	131	7.6	T29I	2	0.1	F238C	1	0.1
T716I	128	7.4	P174S	2	0.1	G261V	1	0.1
R158G	124	7.2	T20I	2	0.1	N211S	1	0.1
A570D	121	7.0	T73I	2	0.1	E324K	1	0.1
S982A	121	7.0	V308L	2	0.1	N185D	1	0.1
D1118H	121	7.0	E96G	2	0.1	Y248C	1	0.1
T1027I	111	6.4	Y144D	2	0.1	K77T	1	0.1
G142D	106	6.1	T547I	2	0.1	A570S	1	0.1
H655Y	81	4.7	P9L	2	0.1	A67S	1	0.1
L18F*	80	4.6	F157L	2	0.1	G446S	1	0.1
V1176F*	79	4.6	P384L	2	0.1	S704L	1	0.1
P26S*	78	4.5	V11F	2	0.1	T859I	1	0.1
D138Y	76	4.4	E484Q	2	0.1	E1195Q	1	0.1
T20N*	71	4.1	S151I	2	0.1	E156A	1	0.1
R190S	70	4.1	V143F	2	0.1	A1022S	1	0.1
K417T	69	4.0	D80H	2	0.1	G181R	1	0.1
G1219C*	42	2.4	N1187S	2	0.1	E180V	1	0.1
A688V*	40	2.3	G446V	2	0.1	V622I	1	0.1
S13I*	39	2.3	T307I	2	0.1	F342L	1	0.1
W152R	33	1.9	A623S	2	0.1	V534I	1	0.1
M153T*	33	1.9	D1163Y	2	0.1	S50L	1	0.1
S735A	33	1.9	A852S	2	0.1	W258C	1	0.1
W152C*	31	1.8	K558N	2	0.1	G142P	1	0.1
K1191N*	29	1.7	R214G	2	0.1	L242F	1	0.1
R346K	27	1.6	L212A	2	0.1	S112L	1	0.1
F1052Y	26	1.5	A846V	2	0.1	E309A	1	0.1
W64R	24	1.4	A684V	2	0.1	V143G	1	0.1

Mutación	Frecuencia	Porcentaje	Mutación	Frecuencia	Porcentaje	Mutación	Frecuencia	Porcentaje
L5F	23	1.3	M153I	2	0.1	H146Y	1	0.1
A263S	23	1.3	R346S	2	0.1	D796H	1	0.1
A701V	23	1.3	Y145S	2	0.1	E583Q	1	0.1
Y145N	19	1.1	S155M	2	0.1	D215N	1	0.1
F157S	16	0.9	R246N	2	0.1	Y508H	1	0.1
S255P*	15	0.9	G769V	2	0.1	V512I	1	0.1
E1258D*	15	0.9	H1101Y	2	0.1	A688S	1	0.1
D215G	14	0.8	D936N	2	0.1	K378R	1	0.1
Q779K	14	0.8	H146N	2	0.1	T274I	1	0.1
T732A	14	0.8	C1250F	2	0.1	K97E	1	0.1
N440Y	14	0.8	L226V	2	0.1	T95S	1	0.1
Y144S	14	0.8	T719I	2	0.1	D1118Y	1	0.1
K417N	14	0.8	V1122L	2	0.1	V511I	1	0.1
D80A	13	0.8	M177I	2	0.1	P812L	1	0.1
S98F	13	0.8	T618I	2	0.1	P25L	1	0.1
Q677H	13	0.8	V860I	2	0.1	Y144V	1	0.1
R158S	12	0.7	Q613H	2	0.1	S659L	1	0.1
G1124V	12	0.7	C1236F	2	0.1	N658T	1	0.1
T859N	11	0.6	I714V	2	0.1	V1129L	1	0.1
D253G	10	0.6	K786Q	2	0.1	S477N	1	0.1
A222V	10	0.6	V1104L	2	0.1	D1127H	1	0.1
E156G	10	0.6	N1074D	1	0.1	A831V	1	0.1
G842S	9	0.5	Q1180L	1	0.1	H625R	1	0.1
Q677P	9	0.5	V407I	1	0.1	A647S	1	0.1
L452Q	9	0.5	M1L	1	0.1	E654V	1	0.1
F490S	9	0.5	C1253Y	1	0.1	G1267V	1	0.1
W258L	9	0.5	L141F	1	0.1	V1264M	1	0.1
V120L	8	0.5	G181A	1	0.1	V608I	1	0.1
T76I	7	0.4	V367L	1	0.1	T678I	1	0.1
V1228L	7	0.4	A419S	1	0.1	I1114L	1	0.1
D936Y	6	0.3	T470N	1	0.1	D1259Y	1	0.1
L18R	6	0.3	T307S	1	0.1	S680F	1	0.1
S494P	6	0.3	S71F	1	0.1	Q1071L	1	0.1
L176F	6	0.3	F59S	1	0.1	R408T	1	0.1
N148T	6	0.3	V1133F	1	0.1	S689I	1	0.1
N440K	6	0.3	C1235S	1	0.1	D1184Y	1	0.1
G769R	6	0.3	T19S	1	0.1	A852V	1	0.1
V1264L	6	0.3	E1188Q	1	0.1	D287H	1	0.1
L948I	5	0.3	H1058Y	1	0.1	D215H	1	0.1
T19K	5	0.3	K1073N	1	0.1	V483F	1	0.1
G75V	5	0.3	F565L	1	0.1	A845V	1	0.1
V503I	5	0.3	Q23L	1	0.1	L1224F	1	0.1
A67V	5	0.3	S459F	1	0.1	S940F	1	0.1
S254F	5	0.3	A1070V	1	0.1	K147H	1	0.1

Mutación	Frecuencia	Porcentaje	Mutación	Frecuencia	Porcentaje	Mutación	Frecuencia	Porcentaje
V213A	5	0.3	K1073M	1	0.1	Y145D	1	0.1
Q675H	5	0.3	R21T	1	0.1	V289I	1	0.1
A879S	5	0.3	I210T	1	0.1	Q957R	1	0.1
Y144T	5	0.3	V367F	1	0.1	K304R	1	0.1
E583D	5	0.3	Y266I	1	0.1	P1263L	1	0.1
F157V	5	0.3	V171D	1	0.1	L938F	1	0.1
L560Q	5	0.3	A93T	1	0.1	T1231A	1	0.1
S221L	4	0.2	F32L	1	0.1	A942V	1	0.1
V622F	4	0.2	P82H	1	0.1	T791I	1	0.1
T573I	4	0.2	L118I	1	0.1	A1080S	1	0.1
T22I	4	0.2	F490L	1	0.1	D1184E	1	0.1
V6F	4	0.2	Q52H	1	0.1	S929N	1	0.1
N658D	4	0.2	T19I	1	0.1	E654Q	1	0.1
Q52R	4	0.2	A67D	1	0.1	L582F	1	0.1
N439K	4	0.2	D138N	1	0.1	A846S	1	0.1
S12C	4	0.2	D88H	1	0.1	V1189I	1	0.1
Q23E	4	0.2	P209R	1	0.1	A1087S	1	0.1
P479S	4	0.2	G72E	1	0.1	Q1201H	1	0.1
S46L	4	0.2	Y265N	1	0.1	A1190S	1	0.1
N679K	4	0.2	A411S	1	0.1	K1157N	1	0.1
T572I	4	0.2	E554G	1	0.1	I896T	1	0.1
M1237I	4	0.2	V911I	1	0.1	T778A	1	0.1
S939F	4	0.2	Y144F	1	0.1	S1252F	1	0.1
F888L	4	0.2	V159I	1	0.1	I1130V	1	0.1
A899S	3	0.2	F306L	1	0.1	V1094I	1	0.1
A27V	3	0.2	V143S	1	0.1	E1151D	1	0.1
P26L	3	0.2	G142S	1	0.1	V1264A	1	0.1
R78K	3	0.2	G75A	1	0.1	A694T	1	0.1
D80G	3	0.2	G72V	1	0.1	A623V	1	0.1
D80Y	3	0.2	V503F	1	0.1	V1177I	1	0.1
G257V	3	0.2	Y248F	1	0.1	L828F	1	0.1
S12F	3	0.2	F220L	1	0.1	C1254F	1	0.1
L141Y	3	0.2	E309Q	1	0.1			
A845S	3	0.2	D138V	1	0.1			

Anexo 2: Frecuencia de mutaciones en las secuencias de SARS-CoV-2 en Belice. Tal como en el resto de los países, la mutación D614G es la dominante, aunque se presenta en siguiente lugar sería W152R.

Mutación	Frecuencia	Porcentaje
D614G	29	100.0
W152R	7	24.1
N188K	2	6.9
T732A	2	6.9
H655Y	1	3.4

Mutación	Frecuencia	Porcentaje
T478K	1	3.4
V407I	1	3.4
P681H	1	3.4

Anexo 3: Frecuencia de mutaciones en las secuencias de SARS-CoV-2 en El Salvador. Al igual que el resto de los países, la mutación más frecuente es D614G. La siguiente mutación más frecuente es L452R

Mutación	Frecuencia	Porcentaje	Mutación	Frecuencia	Porcentaje	Mutación	Frecuencia	Porcentaje
D614G	53	100.0	W64R	5	9.4	S982A	2	3.8
L452R	32	60.4	T19K	3	5.7	D1118H	2	3.8
W152C	15	28.3	L452Q	3	5.7	T29I	1	1.9
S13I	14	26.4	F490S	3	5.7	K1191N	1	1.9
M153T	9	17.0	T859N	3	5.7	E96G	1	1.9
P681H	9	17.0	T716I	2	3.8	D80Y	1	1.9
D215Y	7	13.2	N501Y	2	3.8	G446S	1	1.9
A688V	7	13.2	G75V	2	3.8	A623S	1	1.9
T1027I	7	13.2	A570D	2	3.8	S704L	1	1.9
F1052Y	7	13.2	S46L	2	3.8	A647S	1	1.9
S735A	6	11.3	T76I	2	3.8	A845V	1	1.9

Anexo 4: Frecuencia de mutaciones en las secuencias de SARS-CoV-2 en Guatemala. Al igual que los otros países la mutación D614G es la dominante. De igual modo que El Salvador, la siguiente mutación más frecuente es L452R.

Mutación	Frecuencia	Porcentaje	Mutación	Frecuencia	Porcentaje	Mutación	Frecuencia	Porcentaje
D614G	329	99.7	A348S	3	0.9	P209R	1	0.3
L452R	198	60.0	T76I	3	0.9	G72E	1	0.3
D215Y	95	28.8	L452Q	3	0.9	S254F	1	0.3
T478K	82	24.8	A522V	3	0.9	D80H	1	0.3
D950N	81	24.5	F490S	3	0.9	A845S	1	0.3
T19R	79	23.9	D253N	3	0.9	E324K	1	0.3
G142D	77	23.3	K1191N	2	0.6	N185D	1	0.3
P681R	77	23.3	P26L	2	0.6	Y248C	1	0.3
R158G	76	23.0	T573I	2	0.6	K77T	1	0.3
T95I	72	21.8	R78K	2	0.6	A570S	1	0.3
N501Y	35	10.6	T22I	2	0.6	T307I	1	0.3
P681H	30	9.1	D80G	2	0.6	A67S	1	0.3
E484K	29	8.8	Y144D	2	0.6	A879S	1	0.3
T1027I	29	8.8	V6F	2	0.6	A688S	1	0.3

Mutación	Frecuencia	Porcentaje	Mutación	Frecuencia	Porcentaje	Mutación	Frecuencia	Porcentaje
W152R	23	7.0	T19K	2	0.6	K417N	1	0.3
D138Y	23	7.0	P9L	2	0.6	A684V	1	0.3
A263S	22	6.7	F157S	2	0.6	K378R	1	0.3
P26S	21	6.4	P479S	2	0.6	T274I	1	0.3
H655Y	20	6.1	Q677H	2	0.6	K97E	1	0.3
L18F	19	5.8	S46L	2	0.6	T95S	1	0.3
V1176F	19	5.8	T1117I	2	0.6	A1078S	1	0.3
T20N	18	5.5	V1228L	2	0.6	S477N	1	0.3
R190S	17	5.2	G769V	2	0.6	D1127 H	1	0.3
K417T	17	5.2	M177I	2	0.6	A831V	1	0.3
Q779K	14	4.2	T618I	2	0.6	H625R	1	0.3
T716I	12	3.6	V860I	2	0.6	S939F	1	0.3
A570D	12	3.6	D950H	2	0.6	C1250F	1	0.3
S735A	12	3.6	Q613H	2	0.6	D1184 Y	1	0.3
S982A	12	3.6	C1236F	2	0.6	V1122L	1	0.3
D1118 H	12	3.6	K786Q	2	0.6	A852V	1	0.3
R158S	11	3.3	V1104L	2	0.6	D287H	1	0.3
A688V	11	3.3	S221L	1	0.3	D215H	1	0.3
S13I	11	3.3	A27V	1	0.3	V483F	1	0.3
M153T	9	2.7	S71F	1	0.3	G1219 V	1	0.3
L5F	9	2.7	F59S	1	0.3	Q957R	1	0.3
V120L	8	2.4	D936Y	1	0.3	K304R	1	0.3
S98F	6	1.8	V308L	1	0.3	P1263L	1	0.3
W64R	5	1.5	D215G	1	0.3	D1184E	1	0.3
L18R	5	1.5	F490L	1	0.3	S929N	1	0.3
G1219C	5	1.5	Q52H	1	0.3	E654Q	1	0.3
F1052Y	5	1.5	D80A	1	0.3	L582F	1	0.3
T859N	5	1.5	T547I	1	0.3	I714V	1	0.3
A701V	5	1.5	T19I	1	0.3	A1087S	1	0.3
D253G	4	1.2	G75V	1	0.3	A1190S	1	0.3
Q675H	4	1.2	A67D	1	0.3	K1157 N	1	0.3
N440K	4	1.2	D138N	1	0.3	I896T	1	0.3
R346K	4	1.2	F157L	1	0.3	I1130V	1	0.3
T732A	4	1.2	D88H	1	0.3	V1177I	1	0.3
W152C	3	0.9						

Anexo 5: Frecuencia de mutaciones en las secuencias de SARS-CoV-2 en Honduras. Tal como en el resto de los países, la mutación más frecuentes es D614G. Sin embargo, a comparación del resto, la siguiente sería T1027I.

Mutación	Frecuencia	Porcentaje	Mutación	Frecuencia	Porcentaje	Mutación	Frecuencia	Porcentaje
D614G	40	100.0	D138Y	2	5.0	D215G	1	2.5
T1027I	22	55.0	P26S	2	5.0	P82H	1	2.5
A688V	20	50.0	T19R	2	5.0	L118I	1	2.5
P681H	16	40.0	T95I	2	5.0	L18R	1	2.5
S735A	14	35.0	E484K	2	5.0	A570D	1	2.5
M153T	13	32.5	T20N	2	5.0	N211S	1	2.5
W64R	13	32.5	G142D	2	5.0	Q677H	1	2.5
F1052Y	13	32.5	R158G	2	5.0	Q675H	1	2.5
L18F	8	20.0	V1176F	2	5.0	N440K	1	2.5
L452R	7	17.5	R190S	2	5.0	T732A	1	2.5
N148T	6	15.0	K417T	2	5.0	N679K	1	2.5
G769R	6	15.0	P681R	2	5.0	D215N	1	2.5
D215Y	5	12.5	D950N	2	5.0	Y508H	1	2.5
T478K	3	7.5	R158S	1	2.5	V512I	1	2.5
N501Y	3	7.5	T716I	1	2.5	A701V	1	2.5
Q677P	3	7.5	T22I	1	2.5	S982A	1	2.5
H655Y	2	5.0	F32L	1	2.5	D1118H	1	2.5

Anexo 6: Frecuencia de mutaciones en las secuencias de SARS-CoV-2 en Panamá. Al igual que El Salvador y Guatemala las mutaciones más frecuentes son D614G y L452R.

Mutación	Frecuencia	Porcentaje	Mutación	Frecuencia	Porcentaje	Mutación	Frecuencia	Porcentaje
D614G	266	56.7	G181A	1	3.4	E96G	1	3.4
L452R	121	417.2	V367L	1	3.4	T22I	1	3.4
D215Y	107	369.0	K1191N	1	3.4	V171D	1	3.4
G842S	9	31.0	A419S	1	3.4	A93T	1	3.4
H655Y	5	17.2	T470N	1	3.4	P384L	1	3.4
L948I	5	17.2	T307S	1	3.4	L922F	1	3.4
S221L	3	10.3	T73I	1	3.4	A626V	1	3.4
V1176F	3	10.3	T573I	1	3.4	A845S	1	3.4
N30H	2	6.9	M153T	1	3.4	P812S	1	3.4
P174S	2	6.9	A688V	1	3.4	Q677P	1	3.4
T20I	2	6.9	E1188Q	1	3.4	E340G	1	3.4
V622F	2	6.9	H1058Y	1	3.4	L270H	1	3.4

Mutación	Frecuencia	Porcentaje	Mutación	Frecuencia	Porcentaje	Mutación	Frecuencia	Porcentaje
A27V	2	6.9	K1073N	1	3.4	F238C	1	3.4
L5F	2	6.9	F565L	1	3.4	G261V	1	3.4
N1187S	2	6.9	Q23L	1	3.4	K558N	1	3.4
G446V	2	6.9	T716I	1	3.4	D796H	1	3.4
T572I	2	6.9	S459F	1	3.4	A846V	1	3.4
H49Y	1	3.4	A1070V	1	3.4	E583Q	1	3.4
N1074			K1073			A1078S		
D	1	3.4	M	1	3.4		1	3.4
Q1180L	1	3.4	R21T	1	3.4	S659L	1	3.4
M1L	1	3.4	I210T	1	3.4	N658T	1	3.4
A899S	1	3.4	V367F	1	3.4	V1129		
T29I	1	3.4	R78K	1	3.4	L	1	3.4
C1253Y	1	3.4	Y266I	1	3.4	T719I	1	3.4
L141F	1	3.4				S689I	1	3.4

Anexo 7: Frecuencia de mutaciones en las secuencias de SARS-CoV-2 en Costa Rica. Al igual que El Salvador y Guatemala las mutaciones más frecuentes son D614G y L452R.

Mutación	Frecuencia	Porcentaje	Mutación	Frecuencia	Porcentaje	Mutación	Frecuencia	Porcentaje
D614G	799	99.1	W152R	3	0.4	D80H	1	0.1
L452R	253	31.4	G257V	3	0.4	N211R	1	0.1
N501Y	191	23.7	S12F	3	0.4	A845S	1	0.1
D215Y	163	20.2	L141Y	3	0.4	N440K	1	0.1
P681H	149	18.5	L452Q	3	0.4	A263S	1	0.1
T1117I	131	16.3	N679K	3	0.4	T307I	1	0.1
T716I	112	13.9	F490S	3	0.4	A623S	1	0.1
A570D	106	13.2	E1195A	3	0.4	T859I	1	0.1
S982A	106	13.2	S939F	3	0.4	E1195Q	1	0.1
D1118			T859N			E156A		
H	106	13.2		3	0.4		1	0.1
E484K	100	12.4	N1074S	3	0.4	A1022S	1	0.1
T478K	75	9.3	A899S	2	0.2	G181R	1	0.1
D950N	75	9.3	V622F	2	0.2	E180V	1	0.1
T95I	72	8.9	V6F	2	0.2	V622I	1	0.1
T19R	70	8.7	G75V	2	0.2	F342L	1	0.1
P681R	70	8.7	D80Y	2	0.2	K558N	1	0.1
P26S	55	6.8	V11F	2	0.2	V534I	1	0.1
V1176F	55	6.8	E484Q	2	0.2	S50L	1	0.1

Mutación	Frecuencia	Porcentaje	Mutación	Frecuencia	Porcentaje	Mutación	Frecuencia	Porcentaje
H655Y	53	6.6	S151I	2	0.2	W258C	1	0.1
L18F	53	6.6	P479S	2	0.2	G142P	1	0.1
T1027I	53	6.6	V143F	2	0.2	L242F	1	0.1
D138Y	51	6.3	T76I	2	0.2	S112L	1	0.1
T20N	51	6.3	D1163Y	2	0.2	E309A	1	0.1
R190S	51	6.3	A852S	2	0.2	V143G	1	0.1
K417T	50	6.2	R214G	2	0.2	H146Y	1	0.1
R158G	46	5.7	L212A	2	0.2	A846V	1	0.1
G1219C	37	4.6	T572I	2	0.2	A684V	1	0.1
G142D	27	3.3	M153I	2	0.2	F1052Y	1	0.1
K1191N	25	3.1	R346S	2	0.2	D1118Y	1	0.1
R346K	23	2.9	Y145S	2	0.2	A1078S	1	0.1
Y145N	19	2.4	S155M	2	0.2	V511I	1	0.1
A701V	17	2.1	R246N	2	0.2	P812L	1	0.1
S255P	15	1.9	H1101Y	2	0.2	P25L	1	0.1
E1258D	15	1.9	D936N	2	0.2	Y144V	1	0.1
S13I	14	1.7	H146N	2	0.2	S735A	1	0.1
F157S	14	1.7	L226V	2	0.2	E654V	1	0.1
N440Y	14	1.7	G1219V	2	0.2	G1267V	1	0.1
Y144S	14	1.7	H49Y	1	0.1	V1264M	1	0.1
W152C	13	1.6	T73I	1	0.1	V608I	1	0.1
K417N	13	1.6	P26L	1	0.1	T678I	1	0.1
L5F	12	1.5	T573I	1	0.1	I1114L	1	0.1
D215G	12	1.5	M153T	1	0.1	D1259Y	1	0.1
D80A	12	1.5	V1133F	1	0.1	S680F	1	0.1
G1124V	12	1.5	C1235S	1	0.1	Q1071L	1	0.1
Q677H	10	1.2	V308L	1	0.1	C1250F	1	0.1
A222V	10	1.2	T19S	1	0.1	R408T	1	0.1
E156G	10	1.2	A688V	1	0.1	T719I	1	0.1
W258L	9	1.1	W64R	1	0.1	V1122L	1	0.1
S98F	7	0.9	D80G	1	0.1	L1224F	1	0.1
T732A	7	0.9	T547I	1	0.1	S940F	1	0.1
S494P	6	0.7	F157L	1	0.1	K147H	1	0.1
L176F	6	0.7	Y265N	1	0.1	Y145D	1	0.1

Mutación	Frecuencia	Porcentaje	Mutación	Frecuencia	Porcentaje	Mutación	Frecuencia	Porcentaje
D253G	6	0.7	A411S	1	0.1	V289I	1	0.1
V1264L	6	0.7	E554G	1	0.1	D950H	1	0.1
D936Y	5	0.6	V911I	1	0.1	L938F	1	0.1
V503I	5	0.6	Y144F	1	0.1	T1231A	1	0.1
A67V	5	0.6	V159I	1	0.1	A942V	1	0.1
V213A	5	0.6	P384L	1	0.1	T791I	1	0.1
Q677P	5	0.6	F306L	1	0.1	A1080S	1	0.1
Y144T	5	0.6	V143S	1	0.1	A846S	1	0.1
E583D	5	0.6	G142S	1	0.1	V1189I	1	0.1
V1228L	5	0.6	G75A	1	0.1	I714V	1	0.1
F157V	5	0.6	G72V	1	0.1	Q1201H	1	0.1
L560Q	5	0.6	V503F	1	0.1	T778A	1	0.1
N658D	4	0.5	Y248F	1	0.1	S1252F	1	0.1
Q52R	4	0.5	F220L	1	0.1	V1094I	1	0.1
N439K	4	0.5	E309Q	1	0.1	E1151D	1	0.1
S254F	4	0.5	D138V	1	0.1	V1264A	1	0.1
S12C	4	0.5	V227L	1	0.1	A694T	1	0.1
Q23E	4	0.5	Q218H	1	0.1	A623V	1	0.1
A879S	4	0.5	G35C	1	0.1	L828F	1	0.1
M1237I	4	0.5	H69Y	1	0.1	C1254F	1	0.1
F888L	4	0.5	G142L	1	0.1			

Anexo 8: Presencia de glicosilaciones ligadas a O en las secuencias de Centroamérica. Aquí se encuentran únicamente aquellas secuencias con predicciones de en las posiciones 673, 678 y/o 686.

Secuencia	Posición	Puntaje	Secuencia	Posición	Puntaje	Secuencia	Posición	Puntaje
			EPI_ISL_			EPI_ISL_		
EPI_ISL_	673	0.500	327517	686	0.567	402933	686	0.532
509713			8			6		
	678	0.557	EPI_ISL_	673	0.538	EPI_ISL_	673	0.586
	686	0.525	215755	678	0.634	914803	678	0.635
	673	0.574	1	686	0.529		686	0.563
EPI_ISL_	678	0.616	EPI_ISL_	673	0.595	EPI_ISL_	673	0.531
122554			215734			914828		
8	686	0.519	4	686	0.575		686	0.537
	673	0.588	EPI_ISL_	673	0.592	EPI_ISL_	673	0.562
EPI_ISL_	678	0.665	327521	678	0.678	402933	678	0.597
150297			4	686	0.632	3	686	0.517
2	686	0.603						

Secuencia	Posición	Puntaje	Secuencia	Posición	Puntaje	Secuencia	Posición	Puntaje
EPI_ISL_182276	673	0.574	EPI_ISL_402926	673	0.580	EPI_ISL_673	673	0.603
3	678	0.617	9	686	0.544	402929	678	0.628
EPI_ISL_150298	686	0.519	EPI_ISL_215734	673	0.595	6	686	0.527
7	673	0.571	5	686	0.572	EPI_ISL_434536	673	0.530
EPI_ISL_122554	678	0.606	EPI_ISL_402927	673	0.562	678	678	0.618
0	686	0.516	6	686	0.597	EPI_ISL_106760	686	0.612
EPI_ISL_122548	673	0.629	EPI_ISL_402926	673	0.578	5	673	0.630
6	678	0.666	8	678	0.578	EPI_ISL_402932	678	0.671
EPI_ISL_122551	686	0.603	EPI_ISL_215758	686	0.620	9	686	0.612
4	673	0.547	3	686	0.513	EPI_ISL_769997	673	0.513
EPI_ISL_122539	678	0.625	EPI_ISL_402927	673	0.546	678	686	0.523
7	686	0.536	3	686	0.500	EPI_ISL_914799	673	0.584
EPI_ISL_150296	673	0.574	EPI_ISL_402927	673	0.556	678	678	0.630
5	678	0.617	3	678	0.612	EPI_ISL_106758	686	0.536
EPI_ISL_122548	686	0.521	EPI_ISL_215736	686	0.527	8	673	0.640
8	673	0.528	0	686	0.558	EPI_ISL_402931	678	0.652
EPI_ISL_150296	678	0.625	EPI_ISL_402928	678	0.574	9	686	0.546
5	686	0.526	4	686	0.550	EPI_ISL_402930	673	0.640
EPI_ISL_122548	673	0.516	EPI_ISL_402928	673	0.571	4	678	0.670
8	678	0.566	5	678	0.607	8	686	0.572
EPI_ISL_122543	686	0.500	EPI_ISL_402928	686	0.516	9	673	0.599
9	673	0.593	5	686	0.537	EPI_ISL_402931	678	0.672
EPI_ISL_122548	678	0.593	EPI_ISL_215758	678	0.581	9	686	0.625
7	686	0.681	0	686	0.581	0	673	0.516
EPI_ISL_150291	673	0.632	EPI_ISL_215758	673	0.519	4	686	0.525
3	678	0.531	0	678	0.538	EPI_ISL_402932	673	0.594
EPI_ISL_150312	678	0.617	EPI_ISL_837588	686	0.506	8	686	0.574
8	686	0.611	678	686	0.592	EPI_ISL_402932	673	0.530
EPI_ISL_150291	673	0.574	EPI_ISL_344672	673	0.592	8	686	0.514
3	678	0.616	2	678	0.636	EPI_ISL_491438	673	0.500
EPI_ISL_150312	686	0.521	EPI_ISL_327532	673	0.527	8	678	0.555
8	673	0.514	8	686	0.550	EPI_ISL_120143	686	0.525
EPI_ISL_150312	678	0.563	7	686	0.573	7	673	0.591
8	686	0.500	EPI_ISL_215757	673	0.500	7	686	0.556
EPI_ISL_150312	673	0.578	7	673	0.570	7	673	0.592
8	678	0.619	7	686	0.609	7	678	0.671
EPI_ISL_150312	686	0.513	7	686	0.525	7	678	0.671

Secuencia	Posición	Puntaje	Secuencia	Posición	Puntaje	Secuencia	Posición	Puntaje
EPI_ISL_182277	673	0.575	EPI_ISL_153398	673	0.574		686	0.604
9	678	0.616	7	678	0.608	EPI_ISL_303779	673	0.662
EPI_ISL_150290	686	0.517	EPI_ISL_215757	686	0.523	8	678	0.713
8	673	0.561	9	673	0.575		686	0.648
EPI_ISL_150284	678	0.595		678	0.617	EPI_ISL_181122	673	0.574
7	686	0.509	EPI_ISL_402927	686	0.522	6	678	0.617
			2	673	0.575		686	0.521
	678	0.598		678	0.617	EPI_ISL_152701	673	0.641
	686	0.516		686	0.517	7	678	0.655
EPI_ISL_150294	673	0.629	EPI_ISL_327523	673	0.551		686	0.548
1	678	0.666	1	678	0.579	EPI_ISL_171239	673	0.574
EPI_ISL_122534	686	0.603	EPI_ISL_402928	686	0.515	8	678	0.617
5	673	0.574	6	673	0.565		686	0.515
	678	0.587		678	0.620	EPI_ISL_137940	673	0.577
	686	0.511		686	0.555	5	678	0.617
EPI_ISL_150290	673	0.588	EPI_ISL_383538	673	0.590		686	0.520
0	678	0.632	5	678	0.672	EPI_ISL_171240	673	0.646
	686	0.520		686	0.634	3	678	0.713
	673	0.500	EPI_ISL_215735	673	0.603		686	0.640
EPI_ISL_182276	678	0.557	0	678	0.635	EPI_ISL_210336	673	0.590
9	686	0.530	EPI_ISL_402927	686	0.578	8	678	0.672
EPI_ISL_122555	673	0.586	0	678	0.619		686	0.632
2	678	0.623		686	0.513		673	0.641
	686	0.528	EPI_ISL_402928	673	0.541	EPI_ISL_914833	678	0.594
			7	686	0.564		686	0.522
EPI_ISL_150293	673	0.591	EPI_ISL_150308	673	0.551	EPI_ISL_137940	673	0.589
9	678	0.633	5	678	0.574	1	678	0.621
	686	0.535		686	0.500		686	0.516
EPI_ISL_182276	673	0.574	EPI_ISL_327522	673	0.562	EPI_ISL_119643	673	0.532
0	678	0.616	0	678	0.606	0	678	0.546
	686	0.521		686	0.513	EPI_ISL_171239	686	0.592
EPI_ISL_122545	673	0.592	EPI_ISL_153398	673	0.575	3	673	0.631
9	678	0.606	9	678	0.617		686	0.534
	686	0.531		686	0.519	EPI_ISL_303781	673	0.588
EPI_ISL_150294	673	0.588	EPI_ISL_331904	673	0.601	0	678	0.665
9	678	0.626	7	678	0.687		686	0.624
	686	0.531		686	0.626	EPI_ISL_106762	673	0.521
EPI_ISL_122549	673	0.609	EPI_ISL_402927	673	0.513	0	678	0.514
2	678	0.647	4	673	0.513		686	0.514
	686	0.599		686	0.525		673	0.585

Secuencia	Posición	Puntaje	Secuencia	Posición	Puntaje	Secuencia	Posición	Puntaje
EPI_ISL_496764	673	0.531	EPI_ISL_327522	673	0.565	EPI_ISL_171240	678	0.646
	678	0.620	6	678	0.623	4	686	0.567
	686	0.615		686	0.519		673	0.635
EPI_ISL_182278	673	0.516		673	0.573	EPI_ISL_151739	678	0.666
0	686	0.521	EPI_ISL_402926	678	0.621	5	686	0.568
			6					
EPI_ISL_122552	673	0.574		686	0.524	EPI_ISL_106762	678	0.511
6	678	0.616	EPI_ISL_352779	673	0.601	2	673	0.603
	686	0.520	1	678	0.686	171242	678	0.656
	673	0.500		686	0.626	0	686	0.602
EPI_ISL_122542	678	0.557	EPI_ISL_402927	673	0.528		673	0.576
3	686	0.531	5	686	0.564	EPI_ISL_120143	678	0.617
						8	686	0.520
EPI_ISL_150284	673	0.604	EPI_ISL_914810	673	0.589		673	0.580
4	678	0.632		678	0.645	EPI_ISL_227300	678	0.640
	686	0.533		686	0.548	1	686	0.556
	673	0.589	EPI_ISL_402930	673	0.645		673	0.500
NC_0455	678	0.631	9	686	0.580	EPI_ISL_210338	678	0.556
12_2	686	0.577		673	0.585	5	686	0.518
	673	0.574	EPI_ISL_914807	678	0.631		673	0.577
EPI_ISL_122538	678	0.587		686	0.536	EPI_ISL_171238	678	0.653
2	686	0.511		673	0.584	3	686	0.563
	673	0.530	EPI_ISL_914804	678	0.630		673	0.531
EPI_ISL_150289	678	0.570		686	0.536	EPI_ISL_282799	678	0.513
1	686	0.500	EPI_ISL_491439	673	0.576	5	686	0.513
				678	0.624		673	0.538
EPI_ISL_150290	678	0.713		686	0.524	EPI_ISL_227300	678	0.686
7	686	0.645	EPI_ISL_106760	673	0.574	4	686	0.750
	673	0.576	7	678	0.616		673	0.600
EPI_ISL_122543	678	0.617		686	0.520	EPI_ISL_303782	678	0.680
7	686	0.520	EPI_ISL_106758	673	0.524	6	686	0.627
			9	686	0.538		673	0.586
EPI_ISL_150294	678	0.539	EPI_ISL_106760	678	0.533	EPI_ISL_250274	678	0.672
6	686	0.530	6			7		
	673	0.592		673	0.514		686	0.632
EPI_ISL_122551	678	0.631	EPI_ISL_769988	678	0.563	EPI_ISL_227296	673	0.571
7	686	0.534		686	0.500	7	678	0.578

Secuencia	Posición	Puntaje	Secuencia	Posición	Puntaje	Secuencia	Posición	Puntaje
EPI_ISL_122547	673	0.575	EPI_ISL_770024	673	0.596	EPI_ISL_119642	673	0.635
8	678	0.617		678	0.655	5	678	0.666
	686	0.517		686	0.558		686	0.570
EPI_ISL_496834	673	0.590	EPI_ISL_402934	673	0.536	EPI_ISL_265767	673	0.641
	678	0.631	5	678	0.573	8	678	0.594
	686	0.577	EPI_ISL_106760	673	0.635		686	0.518
EPI_ISL_496621	673	0.574	3	678	0.666	EPI_ISL_151741	673	0.581
	678	0.587		686	0.570	1	678	0.639
	686	0.511	EPI_ISL_527753	673	0.574		686	0.583
EPI_ISL_122551	673	0.613		678	0.616	EPI_ISL_182750	673	0.612
3	678	0.670		686	0.518	9	678	0.640
	686	0.627	EPI_ISL_914811	673	0.563		686	0.595
EPI_ISL_122551	673	0.588		678	0.607	EPI_ISL_303783	673	0.527
5	678	0.609		686	0.520	8	686	0.513
	686	0.539	EPI_ISL_682236	673	0.574		673	0.597
EPI_ISL_150292	673	0.592		678	0.616	EPI_ISL_227298	678	0.664
1	678	0.631		686	0.520	0	686	0.624
	686	0.534	EPI_ISL_527746	673	0.512		673	0.629
EPI_ISL_122550	673	0.574		678	0.560	EPI_ISL_152701	678	0.666
8	678	0.617		686	0.526	3	686	0.603
	686	0.521	EPI_ISL_914813	673	0.582		673	0.536
EPI_ISL_122554	673	0.586		678	0.649	EPI_ISL_120143	678	0.575
6	678	0.630		686	0.568	9	686	0.518
	686	0.534	EPI_ISL_682243	673	0.579		673	0.574
EPI_ISL_150285	673	0.511		678	0.632	EPI_ISL_914838	678	0.616
7	678	0.556		686	0.580		686	0.515
	686	0.525	EPI_ISL_434533	673	0.574	EPI_ISL_210338	673	0.576
EPI_ISL_122540	673	0.597		678	0.617	3	678	0.617
5	678	0.633		686	0.521		686	0.520
	686	0.572	EPI_ISL_327531	673	0.551	EPI_ISL_106762	673	0.590
EPI_ISL_215755	673	0.533	9	678	0.608	5	678	0.654
2	678	0.632		686	0.505		686	0.617
	686	0.534	EPI_ISL_527754	673	0.574	EPI_ISL_250272	678	0.540
EPI_ISL_837584	673	0.540		678	0.616	9	686	0.535
	678	0.555		686	0.521		673	0.589
	686	0.554	EPI_ISL_512668	673	0.590	EPI_ISL_182751	678	0.619
EPI_ISL_402941	673	0.590		678	0.626	7	686	0.528
3	678	0.672		686	0.533		673	0.592
	686	0.632	EPI_ISL_671974	673	0.604	EPI_ISL_171240	678	0.636
	673	0.561		678	0.672	5	686	0.533

Secuencia	Posición	Puntaje	Secuencia	Posición	Puntaje	Secuencia	Posición	Puntaje
EPI_ISL_402926	678	0.597		686	0.609	EPI_ISL_673	673	0.525
5	686	0.515		673	0.584	265767	686	0.526
EPI_ISL_327524	673	0.599	EPI_ISL_914814	678	0.630	7	673	0.553
1	678	0.672		686	0.536	171239	686	0.544
EPI_ISL_383538	686	0.625	EPI_ISL_327529	673	0.640	0	673	0.520
2_	678	0.537	1	678	0.670	265768	686	0.528
EPI_ISL_150308	673	0.561		686	0.573	8	673	0.640
8	678	0.596	EPI_ISL_682276	673	0.581	152701	678	0.695
EPI_ISL_402928	686	0.511		678	0.620	5	686	0.615
0	673	0.516	EPI_ISL_402934	686	0.518	EPI_ISL_210337	673	0.500
EPI_ISL_327523	678	0.552	4	673	0.570	2	678	0.554
8	686	0.562		678	0.607		686	0.525
EPI_ISL_327518	673	0.562	EPI_ISL_682267	686	0.530	EPI_ISL_137942	673	0.647
4	678	0.606		673	0.500	4	678	0.605
EPI_ISL_402928	686	0.519	EPI_ISL_682253	678	0.557	151741	686	0.526
3	673	0.597		686	0.525	2	673	0.621
EPI_ISL_402929	678	0.597	EPI_ISL_914802	673	0.510	151740	678	0.632
4	678	0.627		678	0.558	4	686	0.500
EPI_ISL_402926	673	0.567	EPI_ISL_402933	686	0.530	EPI_ISL_327436	673	0.574
7	678	0.604	4	673	0.640	7	678	0.617
EPI_ISL_327524	686	0.510		678	0.653	327435	686	0.521
5	673	0.596	EPI_ISL_527757	686	0.547	2	673	0.556
EPI_ISL_383539	678	0.596		673	0.579	327435	686	0.527
3	686	0.640	EPI_ISL_402933	678	0.579	1	673	0.538
EPI_ISL_327517	673	0.528	4	678	0.620	327435	686	0.517
8	678	0.573		686	0.517	2	686	0.517
	678	0.524	EPI_ISL_402933	673	0.563	EPI_ISL_327435	673	0.552
	686	0.563	4	678	0.600	1	686	0.527
	673	0.563		686	0.519	EPI_ISL_327435	673	0.517
	678	0.600	EPI_ISL_527757	673	0.584	8	686	0.515
	686	0.519		678	0.630	EPI_ISL_327435	673	0.514
	673	0.594		686	0.536	6	686	0.527
	686	0.562	EPI_ISL_402933	673	0.564	EPI_ISL_327435	673	0.518
	673	0.587	6	678	0.606	5	686	0.515
	678	0.701						

Anexo 9: Distribución de las mutaciones en los distintos dominios de Subunidad S1 la proteína S del SARS-Cov2 en las secuencias de Centroamérica. La mayoría de las mutaciones en las secuencias se encuentran presentes en el dominio del péptido de señal (SP, 8), amino terminal (NTD, 141) o la región de unión al receptor (RBD, 42), mientras que únicamente 44 mutaciones están presentes fuera de la subunidad S1 (FDS1).

Estructura	Posición	Mutación	Estructura	Posición	Mutación	Estructura	Posición	Mutación
SP	1	M1L	NTD	144	Y144V	RBD	384	P384L
	5	L5F		145	Y145N		407	V407I
	6	V6F		145	Y145S		408	R408T
	9	P9L		145	Y145D		411	A411S
	11	V11F		146	H146N		417	K417T
	12	S12C		146	H146Y		417	K417N
	12	S12F		147	K147H		419	A419S
	13	S13I		148	N148T		439	N439K
	18	L18F		151	S151I		440	N440Y
	18	L18R		152	W152R		440	N440K
	19	T19R		152	W152C		446	G446V
	19	T19K		153	M153T		446	G446S
	19	T19S		153	M153I		452	L452R
	19	T19I		155	S155M		452	L452Q
	20	T20N		156	E156G		459	S459F
	20	T20I		156	E156A		470	T470N
	21	R21T		157	F157S		477	S477N
	22	T22I		157	F157V		478	T478K
	23	Q23E		157	F157L		479	P479S
	23	Q23L		158	R158G		483	V483F
25	P25L	158	R158S	484	E484K			
26	P26S	159	V159I	484	E484Q			
26	P26L	171	V171D	490	F490S			
27	A27V	174	P174S	490	F490L			
29	T29I	176	L176F	494	S494P			
30	N30H	177	M177I	501	N501Y			
32	F32L	180	E180V	503	V503I			
35	G35C	181	G181A	503	V503F			
46	S46L	181	G181R	508	Y508H			
49	H49Y	185	N185D	511	V511I			
50	S50L	188	N188K	512	V512I			
52	Q52R	190	R190S	522	A522V			
52	Q52H	209	P209R	534	V534I			
59	F59S	210	I210T	306	F306L			
64	W64R	211	N211R	307	T307I			
67	A67V	211	N211S	307	T307S			
67	A67D	212	L212A	308	V308L			
67	A67S	213	V213A	309	E309Q			
69	H69Y	214	R214G	309	E309A			
71	S71F	215	D215Y	547	T547I			
						FDS1		

Estructura	Posición	Mutación	Estructura	Posición	Mutación	Estructura	Posición	Mutación
	72	G72E		215	D215G		554	E554G
	72	G72V		215	D215N		558	K558N
	73	T73I		215	D215H		560	L560Q
	75	G75V		218	Q218H		565	F565L
	75	G75A		220	F220L		570	A570D
	76	T76I		221	S221L		570	A570S
	77	K77T		222	A222V		572	T572I
	78	R78K		226	L226V		573	T573I
	80	D80A		227	V227L		582	L582F
	80	D80G		238	F238C		583	E583D
	80	D80Y		242	L242F		583	E583Q
	80	D80H		246	R246N		608	V608I
	82	P82H		248	Y248F		613	Q613H
	88	D88H		248	Y248C		614	D614G
	93	A93T	NTD	253	D253G		618	T618I
	95	T95I		253	D253N		622	V622F
	95	T95S		254	S254F		622	V622I
	96	E96G		255	S255P		623	A623S
	97	K97E		257	G257V		623	A623V
NTD	98	S98F		258	W258L	FDS1	625	H625R
	112	S112L		258	W258C		626	A626V
	118	L118I		261	G261V		647	A647S
	120	V120L		263	A263S		654	E654V
	138	D138Y		265	Y265N		654	E654Q
	138	D138N	266	Y266I		655	H655Y	
	138	D138V	270	L270H		658	N658D	
	141	L141Y	274	T274I		658	N658T	
	141	L141F	287	D287H		659	S659L	
	142	G142D	289	V289I		675	Q675H	
	142	G142S	304	K304R		677	Q677H	
	142	G142L	RBD	324	E324K		677	Q677P
	142	G142P		340	E340G		678	T678I
	143	V143F		342	F342L		679	N679K
	143	V143S		346	R346K		680	S680F
	143	V143G		346	R346S		681	P681H
	144	Y144S		348	A348S		681	P681R
	144	Y144T		367	V367L		684	A684V
	144	Y144D		367	V367F			
	144	Y144F		378	K378R			

Anexo 10: Distribución de las mutaciones en los distintos dominios de Subunidad S2 la proteína S del SARS-Cov2 en las secuencias de Centroamérica. A comparación de S1, la mayoría de las mutaciones en las secuencias se encuentran presentes fuera de la subunidad S2 (FDS2, 63), mientras que solo 2 en el péptido de fusión (FP), 13 en secuencia repetida de heptapéptidos 1 (HR1), 14 en HR2, 8 en el dominio transmembranal (TM) y 11 en el dominio citoplásmico

Estructura	Posición	Mutación	Estructura	Posición	Mutación	Estructura	Posición	Mutación
FP	791	T791I	CD	1250	C1250F	FDS2	852	A852V
	796	D796H		1252	S1252F		859	T859N
	922	L922F		1253	C1253Y		859	T859I
929	S929N	1254		C1254F	860		V860I	
936	D936Y	1258		E1258D	879		A879S	
936	D936N	1259		D1259Y	888		F888L	
938	L938F	1263		P1263L	896		I896T	
939	S939F	1264		V1264L	899		A899S	
HR1	940	S940F		1264	V1264M		911	V911I
	942	A942V		1264	V1264A		1022	A1022S
HR2	948	L948I		1267	G1267V		1027	T1027I
	950	D950N		688	A688V		1052	F1052Y
	950	D950H		688	A688S		1058	H1058Y
	957	Q957R	689	S689I	1070	A1070V		
	982	S982A	694	A694T	1071	Q1071L		
	1163	D1163Y	701	A701V	1073	K1073N		
	1176	V1176F	704	S704L	1073	K1073M		
	1177	V1177I	714	I714V	1074	N1074S		
	1180	Q1180L	716	T716I	1074	N1074D		
	1184	D1184Y	719	T719I	1078	A1078S		
	1184	D1184E	732	T732A	1080	A1080S		
	1187	N1187S	735	S735A	1087	A1087S		
	1188	E1188Q	769	G769R	1094	V1094I		
	1189	V1189I	769	G769V	1101	H1101Y		
	1190	A1190S	778	T778A	1104	V1104L		
	1191	K1191N	779	Q779K	1114	I1114L		
	1195	E1195A	786	K786Q	1117	T1117I		
1195	E1195Q	812	P812S	1118	D1118H			
1201	Q1201H	812	P812L	1118	D1118Y			
1219	G1219C	828	L828F	1122	V1122L			
1219	G1219V	831	A831V	1124	G1124V			
1224	L1224F	842	G842S	1127	D1127H			
TM	1228	V1228L	845	A845S	1129	V1129L		
	1231	T1231A	845	A845V	1130	I1130V		
	1235	C1235S	846	A846V	1133	V1133F		
	1236	C1236F	846	A846S	1151	E1151D		
	1237	M1237I	852	A852S	1157	K1157N		

Anexo 11: Análisis de acoplamiento molecular realizado con HADDOCK. Se presenta el promedio de la desviación cuadrática media (RMSD) de átomos de C α en Angstroms de aquellas secuencias cuyos complejos no pueden ser utilizados por la baja significancia que tendrían los resultados.

Código	RMSD (Å)
EPI_ISL_509713	12± 0.5
EPI_ISL_914803	11± 0.2
EPI_ISL_2272967	10.8± 0.6
EPI_ISL_2103385	4.4± 0.2
EPI_ISL_1533989	13.6± 0.7
EPI_ISL_1533987	15.7± 0.2
EPI_ISL_1503045	11.8± 0.1
EPI_ISL_1502972	16.3± 0.2
EPI_ISL_1502847	10.4± 0.4
EPI_ISL_1225554	12.4± 0.2
EPI_ISL_1225487	12± 0.2
EPI_ISL_4029413	12.3± 0.2
EPI_ISL_4029275	8.9± 0.1
EPI_ISL_3761785	5.5± 0.2
EPI_ISL_3275315	4± 0.1