



Revista Internacional de Investigación e Innovación Tecnológica

Página principal: www.riit.com.mx

Diseño de una proteína de fusión para expresarse en tomate contra el SARS-CoV-2: un enfoque bioinformático

Design of a fusion protein to be expressed in tomato against SARS-CoV-2: a bioinformatics approach

Gallegos-López, J.A., Garza-García, D.M., Martínez-Villalobos, J.M., Viader-Salvadó, J.M., Guerrero-Olazarán, M., Galán-Wong, L.J.

Instituto de Biotecnología; Facultad de Ciencias Biológicas; Universidad Autónoma de Nuevo León; 66455; San Nicolás de los Garza, N.L., México.

juan.gallegosp@uanl.edu.mx; dani_rex_12@hotmail.com; juan.martinezvbs@uanl.edu.mx;
jmviader@yahoo.com.mx; mguerrer58@yahoo.com.mx; ljpgw55@gmail.com

Innovación tecnológica: Radica en el empleo de la bioinformática para diseñar una vacuna contra el SARS CoV-2.

Área de aplicación industrial: La innovación tecnológica presentada en este artículo podría ser aplicada en la industria farmacéutica.

Recibido: 29 junio 2021

Aceptado: 01 febrero 2022

Abstract

SARS-CoV-2 is transmitted to humans through the respiratory tract with the ability to infect lung and intestinal tissue, causing COVID-19 disease, which can be fateful. The aim of this study was to design *in silico* a fusion protein (FP) carrying conserved epitopes of the SARS-CoV-2 glycoprotein S (gpS), and its nucleotide sequence (NS) optimized to be expressed in the tomato plant *Solanum lycopersicum* as a candidate vaccine against the SARS-CoV-2 and its variants of concern (VOC). Because of this, a multiple sequence alignment of VOC gpS was calculated, epitopes were predicted and its conservation was verified. The predicted epitopes were used to design the FP sequence. FP sequence was analyzed to predict its physicochemical properties, allergenicity, and toxicity. In addition, its NS was optimized with the codon usage bias of tomato. FP showed 6 epitopes (3 of them conserved), 226 residues, an isoelectric point of 6.29, a molecular weight of 23.08 kDa, it was soluble, stable, was not considered allergenic and it was classified as non-toxic. The *in silico* designed FP for expressing tomato plants has the potential to stimulate an

immune response against SARS-CoV-2 and its current and future VOC. The area of the industry where this innovation could be applied is in the pharmaceutical industry.

Key words: *Agrobacterium tumefaciens*, *in silico*, *Solanum lycopersicum*, SARS-CoV-2, fusion protein.

Resumen

El SARS-CoV-2 se transmite a seres humanos a través de las vías respiratorias con la capacidad de infectar el tejido pulmonar e intestinal, causando la enfermedad COVID-19, la cual puede ser fatal. La finalidad de este estudio fue diseñar *in silico* una proteína de fusión (PF) portadora de epítomos conservados de la glicoproteína S (gpS) del SARS-CoV-2, y su secuencia nucleotídica (SN) optimizada para expresarse en la planta de tomate (*S. lycopersicum*) como vacuna candidata contra del SARS-CoV-2 y sus variantes de preocupación (VDP). Para esto, se calculó un alineamiento múltiple con secuencias de la gpS de las VDP, se determinaron las regiones conservadas y se predijeron epítomos. Estos epítomos se emplearon para diseñar la secuencia de la PF. La PF se analizó para predecir sus propiedades fisicoquímicas, alergenicidad y toxicidad. Además, su secuencia nucleotídica (SN) fue optimizada con los codones preferenciales del tomate. La PF mostró 6 epítomos (3 de los cuales conservados), 226 residuos, un punto isoeléctrico de 6.29, un peso molecular de 23.08 kDa, fue soluble, estable, no se consideró alergénica y se clasificó como no tóxica. La PF diseñada *in silico* para expresarse en la planta de tomate posee potencial para estimular una respuesta inmune contra el SARS-CoV-2 y sus VDP actuales y futuras. El área de la industria donde podría aplicarse esta innovación es en la industria farmacéutica.

Palabras clave: *Agrobacterium tumefaciens*, *in silico*, *Solanum lycopersicum*, SARS-CoV-2, proteína de fusión.

Introducción

A finales de diciembre de 2019, las autoridades de salud de Wuhan, China, informaron de un brote de neumonía letal de etiología desconocida. Para identificar el agente etiológico de la enfermedad se utilizó la técnica de secuenciación de ARN y muestras de pacientes infectados. Los resultados mostraron una secuencia nucleotídica con una identidad del 89.1% con respecto a la secuencia nucleotídica (GenBank: MG772933) del coronavirus (CoV) que causa el Síndrome Respiratorio Agudo Severo (SARS, por sus siglas en inglés) previamente reportado en murciélagos en China (Wu *et al.*, 2020). Adicionalmente, el análisis filogenético realizado a partir de 83 secuencias permitió clasificar al virus como

SARS-CoV-2 y la enfermedad que produce como COVID-19: "CO" corresponde a corona, "VI" a virus y "D" a enfermedad (por sus siglas en inglés) y 19 por el año (2019) en el que fue descubierto (Gorbalenya *et al.*, 2020). Los síntomas que pueden presentarse durante la COVID-19 son fatiga, tos, fiebre, pérdida del olfato, dolor de cabeza, dolor de garganta, dolor muscular, vómito, diarrea y disnea. En algunos casos puede evolucionar a neumonía, disminuyendo el nivel la saturación de oxígeno pudiendo tener un desenlace fatal <https://www.cdc.gov/spanish/>. El SARS-CoV-2 pertenece la familia *Coronaviridae* y al género *Betacoronavirus* (Walls *et al.*, 2020). Es un virus envuelto de aproximadamente 60 a 140 nm de diámetro

(Singhal, 2020), pose un genoma de ARN de cadena sencilla de sentido positivo de alrededor de 30 kb, que codifica 29 proteínas: 16 no estructurales, 9 accesorias y 4 estructurales (Satarker & Nampoothiri, 2020). Entre las proteínas estructurales se encuentra la glicoproteína de espina (S, spike), proteína de membrana (M, membrane), proteína de envoltura (E, envelope) y fosfoproteína de nucleocápside (N, nucleocapsid). El receptor ACE2 de la Enzima Tipo 2 Convertidora de Angiotensina (ACE2, por sus siglas en inglés) se expresa en la superficie de células del sistema respiratorio, del intestino delgado, de los riñones, del corazón, de los vasos sanguíneos, de los testículos, del páncreas, del hígado, de la glándula tiroidea, de la vesícula seminal, y de la placenta. La gpS se encuentra en la superficie del SARS CoV-2, cuya unión con el receptor ACE2 permite la entrada del virus a la célula huésped del sistema respiratorio, pero también está reportado que puede infectar enterocitos del intestino delgado (Lamers *et al.*, 2020; Wrapp *et al.*, 2020; Yan *et al.*, 2020). Por lo anterior, la gpS se ha utilizado para el desarrollo de anticuerpos y posibles vacunas (Dong *et al.*, 2020; Lee *et al.*, 2021).

Actualmente, ya se cuenta con diferentes tipos de vacunas aprobadas contra el SARS-CoV-2 (Dong *et al.*, 2020): vacunas de ARN mensajero (de las compañías Moderna y Pfizer-BioNTech), vacunas basadas en antígenos peptídicos (Centro de Ingeniería Genética y Biotecnología de Cuba, Instituto Vector e Instituto Finlay), vacunas de virus inactivado (Bharat Biotech, Sinopharm y Sinovac Biotech), vacunas de vector viral (CanSino Biologics, Instituto Gamaleya, Oxford-Astra Zeneca y Janssen-J&J). Sin embargo, debido a que el SARS Cov-2 es propenso a mutar esto genera nuevas variantes. Entre estas variantes se encuentran las VDP Alfa (B.1.1.7), Beta (B.1.351), Gamma (P.1), Delta (B.1.617.2), Ómicron

BA.1 y Ómicron BA.2, llamadas así por ser más transmisibles, más virulentas y resistentes a la protección de las vacunas. El día 28 de enero del 2022 se notificaron 3.84 millones de contagios por la variante Ómicron BA.1 rompiendo con todas las marcas de contagios impuestas en olas anteriores. De acuerdo con el Instituto Johns Hopkins (<https://coronavirus.jhu.edu/map.html>), al día 1 de febrero del 2022 se han reportado más de 378 millones en casos acumulados por COVID-19 y más de 5.6 millones de muertes acumuladas a nivel mundial por dicha enfermedad. Las vacunas actuales han perdido cierto grado de protección ante las VDP. Por tal motivo, es necesario desarrollar una vacuna que sea efectiva contra las VDP actuales y futuras.

Otra problemática de las vacunas aprobadas contra el SARS CoV-2 es que requieren de una cadena de frío para su preservación. Las vacunas expresadas en plantas no requieren de refrigeración, lo que las hace atractivas para las naciones que se encuentran en vías de desarrollo (Parvathy, 2020). Otras ventajas del sistema de expresión de plantas es la realización de modificaciones postraduccionales, su seguridad, su alta producción y bajo costo (Shim *et al.*, 2019; He *et al.*, 2006).

Así mismo, en una pandemia es necesario desarrollar las vacunas en tiempo récord, la vacunología inversa es un enfoque que permite predecir mediante bioinformática epítomos a partir de la secuencia del genoma de microorganismos patógenos, permitiendo desarrollar vacunas más rápido que los métodos tradicionales (Masignani *et al.*, 2019).

En el presente trabajo, se diseñó *in silico* una PF portadora de epítomos conservados de la gpS del SARS-CoV-2 para expresarse en la planta de tomate *S. lycopersicum*, como una vacuna candidata contra las VDP del SARS-CoV-2.

Material

Se utilizó el sistema operativo: Windows 10 Profesional de 64 bits ejecutado en una computadora Lenovo Ideapad 330-15IKB con un procesador Intel® Core™ i3-8130U de 2.21GHz con 4 Gb de RAM.

Métodos experimentales

Obtención de secuencias de la gpS del SARS-CoV-2

A partir de la base de datos del Centro Nacional para la Información Biotecnológica (NCBI, por sus siglas en inglés), se obtuvieron las secuencias aminoacídicas de la gpS del SARS-CoV-2 (no. de acceso: QHO62877.1), y de las VDP Alfa B.1.1.7, (no. de acceso QWE88920), Beta B.1.351 (no. de acceso QRN78347), Gamma P.1 (no. de acceso QVE55289), Delta B.1.617.2 (no. de acceso QWK65230), Ómicron BA.1 (no. de acceso UFO69279.1) y Ómicron BA.2 (<https://www.ncbi.nlm.nih.gov/datasets>).

Predicción de la topología de la gpS del SARS Cov-2

Adicionalmente, se predijo la topología de la secuencia aminoacídica de la gpS del SARS-CoV-2 utilizando el servidor CCTOP (<http://cctop.enzim.ttk.mta.hu/>), el cual predice usando una base de datos experimental.

Predicción de epítomos de células B

Se analizó la secuencia aminoacídica de la gpS del SARS-CoV-2 con herramientas para predecir epítomos de células B (<https://www.iedb.org/>) de la Base de Datos Epítomo Inmune (IEDB, por sus siglas en inglés). La herramienta Bepipred se empleó para predecir regiones inmunogénicas, utilizando un umbral de 0.35. Las regiones antigénicas se identificaron con la herramienta de antigenicidad de Kolaskar y Tongaonkar (umbral: 1.041). Las regiones hidrofílicas se predijeron con la herramienta Parker Hydrophilicity (umbral: 1.238). Las

regiones accesibles se verificaron con el programa Emini de Predicción de Accesibilidad de Superficie (umbral: 1.000). El programa de predicción de flexibilidad Karplus y Schulz (umbral: 0.993), se utilizó para determinar regiones flexibles. La herramienta de Chou y Fasman, se utilizó para identificar péptidos en giros β (umbral: 0.997).

Alineamiento múltiple de secuencias de la gpS del SARS CoV-2 y sus VDP

Para verificar si las secuencias de los epítomos de la PF identificados en la secuencia de gpS del SARS CoV-2 se encuentran en regiones conservadas o en regiones susceptibles a mutaciones, dichas secuencias se alinearon con el programa Jalview 2.11.1.5 (<https://www.jalview.org/>).

Diseño de la proteína de fusión

Los péptidos sugeridos por las herramientas del IEDB y la secuencia aminoacídica de la gpS del SARS-CoV-2, se alinearon con el programa Clustal Omega (<https://www.ebi.ac.uk/Tools/msa/clustalo/>). Posteriormente, los péptidos de la gpS que fueron accesibles, inmunogénicas, antigénicas, hidrofílicas, flexibles y que estuvieron en giro beta fueron seleccionados como epítomos. Los epítomos seleccionados de la gpS del SARS-CoV-2, se unieron mediante el enlazador helicoidal EAAAKEAAAKEAAAKEAAK. Además, se predijo la estructura secundaria de la PF con el servidor SOPMA (https://npsa-prabi.ibcp.fr/cgi-bin/npsa_automat.pl?page=npsa_sopma.html), que emplea una base de datos experimental. Así mismo, la estructura terciaria de la PF se modeló y se seleccionó el mejor modelo con el servidor I-Tasser (<https://zhanggroup.org/I-TASSER/>), que utiliza una base de datos experimental. Adicionalmente, se realizó el refinamiento de la estructura terciaria con el servidor

GalaxyRefine (<http://galaxy.seoklab.org/>), que predice usando una base de datos experimental. Además, el mejor modelo obtenido de la PF fue evaluado a través del análisis del gráfico de Ramachandran (<https://saves.mbi.ucla.edu/>), el cual emplea una base de datos experimental. Así mismo, para validar la estructura de la PF se usó el servidor ProSA-web, el cual usa una base de datos experimental (<https://prosa.services.came.sbg.ac.at/prosa.php>). Finalmente, la estabilidad del modelo de la PF se calculó mediante el análisis de minimización de energía con el programa Swiss-PDB Viewer (<https://spdbv.unil.ch/>).

Propiedades fisicoquímicas de la proteína de fusión

Las propiedades fisicoquímicas de la PF se evaluaron con la herramienta ProtParam del Instituto Suizo de Bioinformática (SIB, por sus siglas en inglés, <https://web.expasy.org/protparam/>). La solubilidad de la PF se calculó utilizando el programa Protein-Sol (<https://protein-sol.manchester.ac.uk/>), que utiliza una base de datos experimental y maneja una escala de 0 a 1.0. Adicionalmente, la alergenicidad de la PF se predijo con los servidores AllerTop v.2.0 (<https://www.ddgpharmfac.net/AllerTOP/>). También, se evaluó la toxicidad de la PF con el servidor ToxinPred (<https://webs.iitd.edu.in/raghava/toxinpred/algorithm.php>) que utiliza una base de datos experimental y maneja una escala de 0.80 a 1.0.

Optimización de la secuencia nucleotídica codificante de la PF

La SN de la PF se optimizó con los codones de uso de *S. lycopersicum* utilizando el programa GenSmart™ Codon Optimization Tool (<https://www.genscript.com/tools/genSMART-codon-optimization>) que emplea una base de

datos experimental. Además, el sitio de restricción *BamHI*, la secuencia Kozak y el codón de inicio de la traducción se agregaron al extremo 5' de la SN. Adicionalmente, la secuencia codificante del péptido señal SEKDEL, el codón de paro (TAA) de la traducción y el sitio de restricción *SacI*, se agregaron al extremo 3' de la SN. La secuencia resultante se clonó *in silico* en el plásmido binario pBI121 de *Agrobacterium tumefaciens*, río abajo del promotor 35S CaMV.

Discusión de resultados

Predicción de la topología de la gpS del SARS-CoV-2

Para este estudio se seleccionó la gpS del SARS-CoV-2 debido a que su unión con el receptor ACE2 permite la entrada del virus a la célula huésped, lo que la convierte en un blanco óptimo para el diseño de vacunas (Dwivedi *et al.*, 2016).

El análisis de topología de la secuencia de la gpS del SARS-CoV-2, con el servidor CCTOP, mostró que la proteína posee una región expuesta, que abarca de la posición 1 a la 1,215 de la secuencia; una región dentro de la membrana del virus, que va de la posición 1,216 a la 1,241; y una región en el interior del virus, que va de la posición 1,242 a la 1,273 de la secuencia.

Predicción de epítomos de células B

El análisis de la secuencia aminoacídica de la gpS del SARS-CoV-2 con herramientas de predicción de epítomos del IEDB predijo 152 péptidos: 30 accesibles, 34 inmunogénicos, 11 hidrofílicos, 46 antigénicos, 13 flexibles y 10 en giro β . El alineamiento de estos péptidos y la secuencia aminoacídica de la gpS del SARS-CoV-2 permitieron identificar siete epítomos (Tabla I). Adicionalmente, el programa Swiss-PDB Viewer verificó que los epítomos seleccionados se encuentran accesibles en la superficie de la estructura

tridimensional de la gpS del SARS-CoV-2, lo que permitiría la interacción entre antígeno-anticuerpo, importante para desencadenar una respuesta inmunológica (Hughes & Gilleland Jr, 1995). Así mismo, los epítomos seleccionados mostraron ser inmunogénicos y antigénicos, lo que permitiría inducir una respuesta inmune (Fishman *et al.*, 2015).

Los péptidos inmunogénicos obtenidos en este trabajo están en concordancia con los

reportados por Bhattacharya y colaboradores (Bhattacharya *et al.*, 2020), quienes en el 2020 predijeron 34 péptidos lineales de células B a partir del análisis de la secuencia de la gpS del SARS CoV-2, con la herramienta Bepipred Linear Epitope Prediction del IEDB, para desarrollar una vacuna basada en epítomo contra el SARS CoV-2 mediante un enfoque bioinformático.

Tabla 1. Epítomos identificados en la estructura de la gpS del SARS-CoV-2 mediante las herramientas de predicción de epítomos de células B del IEDB. Además, se muestran los péptidos inmunogénicos, antigénicos, hidrofílicos, accesibles, flexibles y en giro β , que conforman los epítomos, su posición en la secuencia proteica y su ubicación en la gpS del virus.

Análisis	Epítomo 1	Epítomo 2	Epítomo 3	Epítomo 4	Epítomo 5	Epítomo 6	Epítomo 7
Secuencia del Epítomo	TTLDSK TQSLLI VN	FQTLALHRS YLTPGDSSSG WTAG	KYNENGTI TDAVDCAL DPL	PGTNTSNQVAVLY QDVNCTEVPVAIH ADQL	YQTQTNSPRR ARSVASQSIIA YTMS	FSQILP DPSKPS KR	LKGCCSCGSCC KFDEDDSEPV L KGV
Posición en la secuencia	108-121	238-261	278-296	600-629	674-698	802-815	1244-1268
Localización en el virus	externa	externa	externa	externa	externa	externa	interna
Inmunogenicidad	LDSK	LTPGDSSSGW TAG	NGTITD	VITPGTNTSN	QTQTNSPRRA RSV	ILPDPS KPSKR S	FDEDDSEPV L
Antigenicidad	QSLLI N	FQTLALHRS YLTP	TDAVDCAL DPL	QVAVLYQDVNCT EVPVAIHADQL	ARSVASQSIIA YTMS	FSQILP DPS	LKGCCSCGSCC KF
Hidrofili	LDSKT	GDSSSS	GTI	TNTSN	YQTQTNSPRR	SKPS	FDEDDSS
Accesibilidad	LDSKT	TPGDSS	KYNENGT	GTNTSN	YQTQTNSPRR AR	DPSKPS KRSF	FDEDDSS
Giros-β	TLDSKT	YLTPGDSSSG WT	YNENGTI	TPGTNTSN	YQTQTNSPRR	LPDPS KPSKR	CGSCCKFDED DSE
Flexibilidad	TTLDSK TQS	LTPGDSSSGW	YNENGTIT D	TPGTNTSNQ	QTQTNSPRRA RSVASQS	LPDPS KPSKR S	FDEDDSEPV L KGV

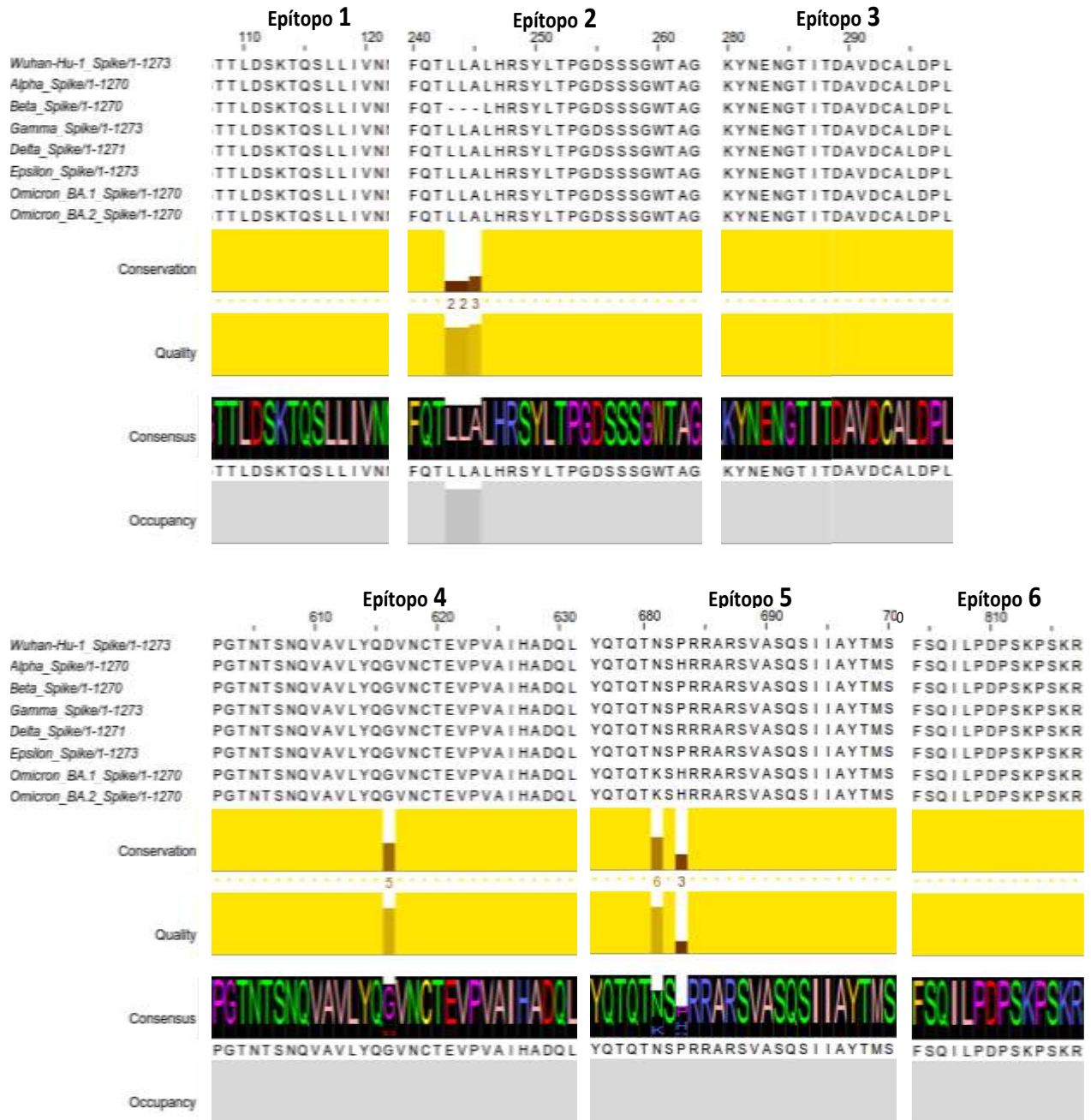


Figura 1. Alineamiento múltiple de las secuencias de los epítomos seleccionados de la gpS del SARS CoV-2 y de las VDP Alfa, Beta, Gamma, Delta y Ómicron (BA.1 y BA2) con el programa Jalview 2.11.1.5. Se muestran los epítomos 1, 3 y 6 en regiones conservadas. Así mismo, se muestra el epítomo 2 con deleciones de los residuos de aminoácidos L243, L244 y D245 en la variante Beta. Adicionalmente, se muestra la sustitución D616G en el epítomo 4 en las VDP. Además, se observó que la secuencia del epítomo 5 posee la sustitución N681K de la variante Ómicron (BA1 y BA.2), la sustitución P683H de la variante Alfa y la sustitución P683R de la variante Delta.

Alineamiento múltiple de secuencias de la gpS del SARS CoV-2 y sus VDP

El resultado del alineamiento múltiple de las secuencias de la gpS del SARS CoV-2 y de

las VDP Alfa, Beta, Gamma, Delta y Ómicron (BA.1 y BA2) con el programa Jalview 2.11.1.5 (Figura 1) mostró que las secuencias de los epítomos 1, 3 y 6, de la PF

se encuentran en regiones conservadas sin presentar sustituciones ni deleciones. Así mismo, este resultado mostró que la secuencia del epítipo 2 perdió los residuos de aminoácidos L243, L244 y D245 en la variante Beta. Además, se observó que las secuencias del epítipo 4 contienen la sustitución D616G en las variantes Alfa, Beta, Gamma, Delta y Ómicron (BA.1 y BA.2). Adicionalmente, se observó que la secuencia del epítipo 5 posee la sustitución N681K de la variante Ómicron (BA.1 y BA.2), la sustitución P683H de la variante Alfa y la sustitución P683R de la variante Delta. Esto sugiere que los epítopos 1, 3 y 6 de la PF son menos susceptibles a las mutaciones, por estar en regiones conservadas, lo que permitirá que la vacuna sea efectiva contra las VDP actuales y que no pierda su efectividad contra futuras VDP.

Diseño de la secuencia de la PF

La secuencia de la PF se diseñó con seis de los siete epítopos identificados en la secuencia de la gpS del SARS-CoV-2, ya que estos epítopos se encontraron en una región expuesta de la gpS del SARS-CoV-2, lo que los convierte en blancos idóneos para los anticuerpos neutralizantes (Dwivedi *et al.*, 2016). El epítipo 7 no se incluyó en la PF por encontrarse en el interior de la estructura viral. Además, los epítopos 1, 2, 3, 4, 5, y 6

de la PF, se unieron mediante enlazadores helicoidales (Figura 2), ya que este tipo de enlazadores controlan mejor la distancia y la interferencia entre los dominios de las proteínas de fusión (Arai *et al.*, 2001). Veinte residuos de aminoácidos de la secuencia del epítipo 2 de la PF, resultaron estar en concordancia con la secuencia del epítipo de células B ALHRSYLTPGDSSSGWTAGA, empleado previamente en la secuencia de una vacuna multiepítipo diseñada *in silico* contra nuevas variantes del SARS CoV-2 por Khan y colaboradores en el 2021 (Khan *et al.*, 2021).

De igual manera, trece residuos de aminoácidos de la secuencia del epítipo 2 empleado para el diseño de la PF fueron idénticos al epítipo LTPGDSSSGWTA usado por Kumar y colaboradores para diseñar una vacuna multiepítipo contra el SARS CoV-2 mediante un enfoque *in silico* (Kumar *et al.*, 2020). Así mismo, esta reportado el empleo de 16 residuos de aminoácidos de la secuencia del epítipo 2 de la PF, en el epítipo de células B (HRSYLTPGDSSSGWTA) de la gpS del SARS CoV-2, usado en la secuencia de una vacuna multiepítipo desarrollada mediante inmunoinformática por Sharma y colaboradores contra el SARS CoV-2 (Sharma *et al.*, 2021).


```

1 BamHI Kozak M P G T N T S N Q V A V L Y Q D V 16
1 5'ggatcc gcaaca atg cca gga act aat act agc aat caa gtt gct gtt ctt tat caa gat gtt 63
I 5'ggatcc gccacc atg cca gga aca aat act tct aac cag gtt gct gtt ctt tat cag gat gtt 63
17 N C T E V P V A I H A D Q L E A A A K E A 37
64 aat tgt act gaa gtt cca gtt gct atc cat gct gat caa ctt gaa gct gct gct aaa gaa gct 126
64 aac tgc aca gaa gtc cct gtt gct att cat gca gat caa ctt gag gca gca gca aaa gag gca 126
38 A A K E A A A K E A A A K F Q T L L A L H 58
127 gct gct aaa gaa gct gct gct aaa gaa gct gct gct aaa ttt caa act ctt ctt gct ctt cat 189
127 gca gca aaa gag gca gca gca aaa gag gca gca gca aaa ttt caa act tta ctt gct tta cat 189
59 R S Y L T P G D S S S G W T A G E A A A K 79
190 cgt agc tat ctt act cca gga gat agc agc agc gga tgg act gct gga gaa gct gct gct aaa 252
190 aga agt tat ttg act cct ggt gat tct tct tca ggt tgg aca gct ggt gag gca gca gca aaa 252
80 E A A A K E A A A K E A A A K T T L D S K 100
253 gaa gct gct gct aaa gaa gct gct gct aaa gaa gct gct gct aaa act act ctt gat agc aaa 315
253 gag gca gca gca aaa gag gca gca gca aaa gag gca gca gca aaa act act tta gat tgc aag 315
101 T Q S L L I V N E A A A K E A A A K E A A 121
316 act caa agc ctt ctt atc gtt aat gaa gct gct gct aaa gaa gct gct gct aaa gaa gct gct 378
316 acc cag tcc cta ctt att gtt aat gag gca gca gca aaa gag gca gca gca aaa gag gca gca 398
122 A K E A A A K Y Q T Q T N S P R R A R S V 142
379 gct aaa gaa gct gct gct aaa tat caa act caa act aat agc cca cgt cgt gct cgt agc gtt 441
379 gca aaa gag gca gca gca aaa tat cag act cag act aat tct cct cgg cgg gca cgt agt gta 441
143 A S Q S I I A Y T M S E A A A K E A A A K 163
442 gct agc caa agc atc atc gct tat act atg agc gaa gct gct gct aaa gaa gct gct gct aaa 504
442 gct agt caa tcc atc att gcc tac act atg tca gag gca gca gca aaa gag gca gca gca aaa 504
164 E A A A K E A A A K F S Q I L P D P S K F 184
505 gaa gct gct gct aaa gaa gct gct gct aaa ttt agc caa atc ctt cca gat cca agc aaa cca 567
505 gag gca gca gca aaa gag gca gca gca aaa ttt tca caa ata tta cca gat cca tca aaa cca 567
185 S K R E A A A K E A A A K E A A A K E A A 205
568 agc aaa cgt gaa gct gct gct aaa gaa gct gct gct aaa gaa gct gct gct aaa gaa gct gct 630
568 agc aag agg gag gca gca gca aaa gag gca gca gca aaa gag gca gca gca aaa gag gca gca 630
206 A K K Y N E N G T I T D A V D C A L D P I 226
631 gct aaa aaa tat aat gaa aat gga act atc act gat gct gtt gat tgt gct ctt gat cca ctt 693
631 gca aaa aaa tat aat gaa aat gga acc att aca gat gct gta gac tgt gca ctt gac cct ctc 693
227 S E K D E L 232
694 aga gaa aag gat gaa ctc taa gagctc 3' 720
694 cgc gag aag gac gag ctg taa gagctc 3' 720
Stop SacI

```

Figura 2. Se muestran las secuencias de los epítomos de la PF marcados en gris y unidos mediante enlazadores helicoidales de cuatro repeticiones de la secuencia EAAAK. Además, se observa la secuencia nucleotídica, que codifica para la PF, optimizada con los codones de uso de la planta de tomate *S. lycopersicum*. En el extremo 5' se muestran: el sitio de restricción de *BamHI*, así como el sitio de unión al ribosoma (kozak) y del codón de inicio de la traducción. Así mismo, en el extremo 3', se muestra la secuencia nucleotídica que codifica el péptido señal del retículo endoplásmico (SEKDEL), el codón de paro y el sitio de restricción *SacI*. Adicionalmente, se observa la secuencia nucleotídica nativa (en itálicas) que codifica la PF.

Análisis de estructura de la PF

El análisis de la PF con el servidor SOPMA mostró 84% de hélices α y 16% de otras (bobinas y vueltas). La validación del modelo refinado de la estructura tridimensional con ProSA-web evidenció un Z-Score de -6.51 . Así mismo, el análisis mediante el gráfico de Ramachandran evidenció que 96.7% (205 residuos) de la PF se encontraron en las

regiones más favorables, el 2.8% (6 residuos) en las regiones permitidas adicionales y el 0.5% (1 residuo) en las regiones no permitidas, lo que sugiere la buena calidad del modelo (figura 3). El análisis de minimización de energía del modelo de la PF con el programa Swiss-PDB Viewer fue de -10129.876 KJ/mol, lo que indica la estabilidad de la construcción diseñada.



Figura 3. Se observa el modelo teórico de la estructura tridimensional del PF, así mismo se observa el extremo amino, el extremo carboxilo, las hélices alfa en color rosa, las tiras beta en color amarillo y los giros en color gris.

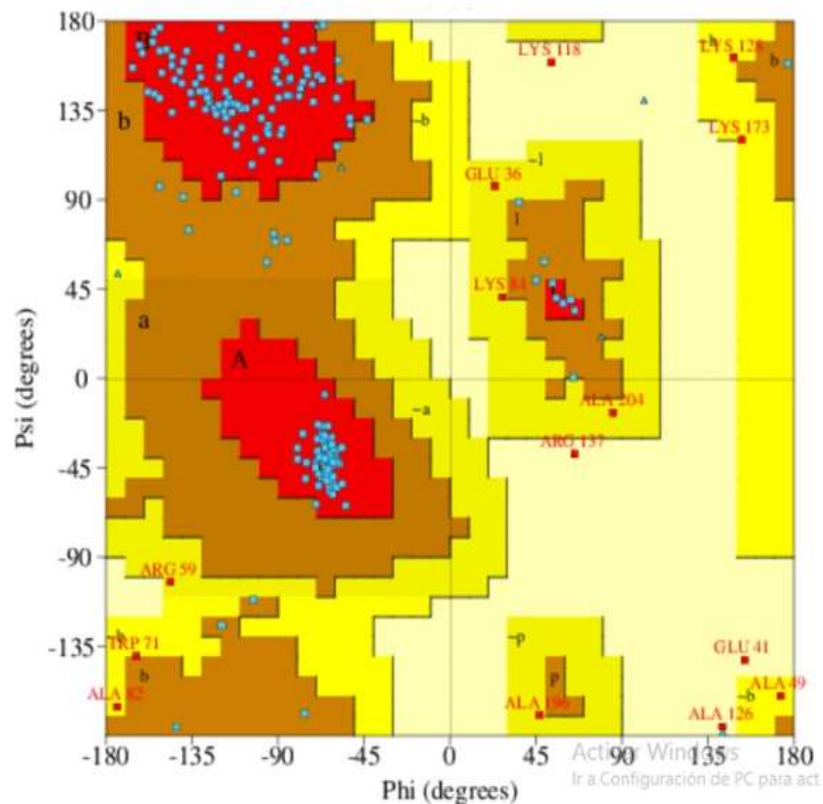


Figura 4. Muestra el gráfico de Ramachandran del modelo teórico de la estructura tridimensional de la PF con 205 residuos (96.7%) de la PF en las regiones más favorables, 6 residuos (2.8%) en las regiones permitidas adicionales y 1 residuo (0.5%) en las regiones no permitidas.

Propiedades fisicoquímicas de la PF

El análisis fisicoquímico de la PF mostró 226 residuos de aminoácidos, un punto isoeléctrico de 6.29 y un peso molecular de 23.08 kDa. La vida media estimada de la PF fue de 20 horas en reticulocitos de mamíferos y >20 horas en levadura. El índice de inestabilidad (II) fue de 32.54, lo que clasifica a la PF como estable. Los índices GRAVY y alifáticos resultaron ser -0.366 y 72.30, respectivamente. La solubilidad de la PF fue de 0.737, lo que indica que es altamente soluble. Así mismo, los servidores AllerTop v.2.0 y ToxinPred clasificaron a la PF como no alergénica y no tóxica, respectivamente, lo que sugiere que la probabilidad que la PF induzca alergia o toxicidad es baja.

Optimización de la secuencia nucleotídica que codifica para la PF

En este estudio se diseñó una PF para expresarse en la planta de tomate, como sistema de expresión, ya que esta reportado la producción exitosa de vacunas en el fruto de la planta (Pogrebnyak *et al.*, 2005; Jiang *et al.*, 2007; Beihaghi *et al.*, 2018). Por lo tanto, la SN que codifica para la PF se optimizó con los codones de uso del tomate, y se agregó la secuencia Kozak y la secuencia SEKDEL a los extremos 5' y 3', respectivamente, para mejorar la producción de la PF en la planta de tomate, tal como se describió anteriormente por Gil *et al.*, (2001). El porcentaje promedio de G+C en la PF fue de 46.01%, el cual se encuentra dentro umbral idóneo (30-70%). La SN optimizada mostró una longitud de 720 pb (Figura 2) y el nuevo plásmido portador de la SN optimizada, llamado pBI121-PF, mostró un tamaño de 13,572 pb (Figura 5).

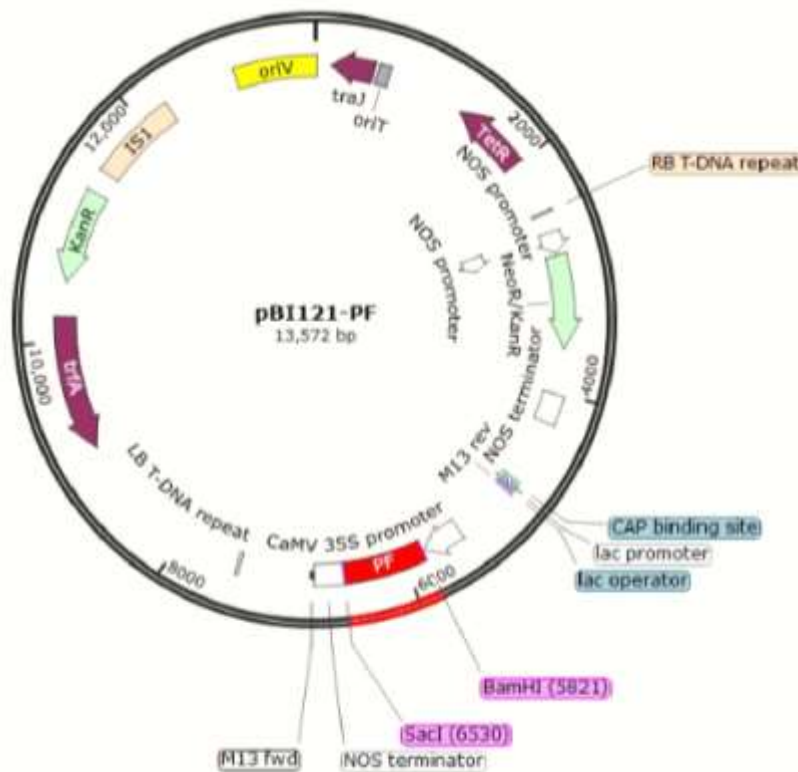


Figura 5. Mapa del plásmido pBI121-PF portador de la secuencia nucleotídica que codifica para la PF (rojo), en marco de lectura con el promotor CaMV 35S y el terminador de la transcripción NOS.

Conclusión

Los resultados obtenidos en este trabajo indican a que la PF diseñada *in silico* para su expresión en *S. lycopersicum*, posee el potencial para activar una respuesta inmunológica contra el SARS-CoV-2 y las VDP actuales y futuras. No obstante, se necesitarán más ensayos *in vitro* e *in vivo* para verificar la efectividad y la seguridad de la PF.

Agradecimientos

Al LBG Jesús Eduardo Pizaña Alvarado por la elaboración de las figuras y la edición del texto del presente manuscrito.

Referencias

Arai, R., Ueda, H., Kitayama, A., Kamiya, N., & Nagamune, T. (2001). Design of the linkers which effectively separate domains of a bifunctional fusion protein. *Protein engineering*, *14*(8), 529-532. <https://doi.org/10.1093/protein/14.8.529>

Bhattacharya, M., Sharma, A. R., Patra, P., Ghosh, P., Sharma, G., Patra, B. C., ... & Chakraborty, C. (2020). Development of epitope-based peptide vaccine against novel coronavirus 2019 (SARS-COV-2): Immunoinformatics approach. *Journal of medical virology*, *92*(6), 618-631. <https://doi.org/10.1002/jmv.25736>

Beihaghi, M., Marashi, H., Bagheri, A., & Sankian, M. (2018). Transient expression of CCL21as recombinant protein in tomato. *Biotechnology reports*, *17*, 10-15. <https://doi.org/10.1016/j.btre.2017.11.007>

COVID-19 Map. (2021, 7 noviembre). Johns Hopkins Coronavirus Resource Center. <https://coronavirus.jhu.edu/map.html>

Dong, Y., Dai, T., Wei, Y., Zhang, L., Zheng, M., & Zhou, F. (2020). A systematic review of SARS-CoV-2 vaccine candidates. *Signal transduction and targeted therapy*, *5*(1), 1-

14. <https://doi.org/10.1038/s41392-020-00352-y>

Dwivedi, P., Alam, S. I., & Tomar, R. S. (2016). Secretome, surfome and immunome: emerging approaches for the discovery of new vaccine candidates against bacterial infections. *World Journal of Microbiology and Biotechnology*, *32*(9), 1-9. [10.1007/s11274-016-2107-3](https://doi.org/10.1007/s11274-016-2107-3)

Fishman, J. M., Wiles, K., & Wood, K. J. (2015). The acquired immune system response to biomaterials, including both naturally occurring and synthetic biomaterials. In *Host response to biomaterials* (pp. 151-187). Academic Press. <https://doi.org/10.1016/B978-0-12-800196-7.00008-6>

Gil, F., Brun, A., Wigdorovitz, A., Catalá, R., Martínez-Torrecuadrada, J. L., Casal, I., ... & Escribano, J. M. (2001). High-yield expression of a viral peptide vaccine in transgenic plants. *FEBS letters*, *488*(1-2), 13-17. [https://doi: 10.1016/S0014-5793\(00\)02405-4](https://doi.org/10.1016/S0014-5793(00)02405-4).

Gorbalenya, A., Baker, S., Baric, R., de Groot, R., Drosten, C., Gulyaeva, A., ... & Ziebuhr, J. (2020). The species Severe acute respiratory syndrome-related coronavirus: classifying 2019-nCoV and naming it SARS-CoV-2. *Nature microbiology*, *5*(4), 536. [https://doi:10.1038/s41564-020-0695-z](https://doi.org/10.1038/s41564-020-0695-z)

He, Y., Li, J., Heck, S., Lustigman, S., & Jiang, S. (2006). Antigenic and immunogenic characterization of recombinant baculovirus-expressed severe acute respiratory syndrome coronavirus spike protein: implication for vaccine design. *Journal of virology*, *80*(12), 5757-5767. <https://doi.org/10.1128/JVI.00083-06>

Hughes, E. E., Gilleland, L. B., & Gilleland Jr, H. E. (1992). Synthetic peptides representing epitopes of outer membrane protein F of *Pseudomonas aeruginosa* that

elicit antibodies reactive with whole cells of heterologous immunotype strains of *P. aeruginosa*. *Infection and immunity*, 60(9), 3497-3503.

<https://doi.org/10.1128/iai.60.9.3497-3503.1992>

Jiang, X. L., He, Z. M., Peng, Z. Q., Qi, Y., Chen, Q., & Yu, S. Y. (2007). Cholera toxin B protein in transgenic tomato fruit induces systemic immune response in mice. *Transgenic research*, 16(2), 169-175. DOI 10.1007/s11248-006-9023-5

Khan, A., Khan, S., Saleem, S., Nizam-Uddin, N., Mohammad, A., Khan, T., ... & Wei, D. Q. (2021). Immunogenomics guided design of immunomodulatory multi-epitope subunit vaccine against the SARS-CoV-2 new variants, and its validation through in silico cloning and immune simulation. *Computers in biology and medicine*, 133, 104420.

<https://doi.org/10.1016/j.compbiomed.2021.104420>

Lamers, M. M., Beumer, J., van der Vaart, J., Knoop, K., Puschhof, J., Breugem, T. I., ... & Clevers, H. (2020). SARS-CoV-2 productively infects human gut enterocytes. *Science*, 369(6499), 50-54. <https://doi.org/10.1126/science.abc1669>

Lee, P., Kim, C. U., Seo, S. H., & Kim, D. J. (2021). Current status of COVID-19 vaccine development: focusing on antigen design and clinical trials on later stages. *Immune network*, 21(1). 10.4110/in.2021.21.e4

Masignani, V., Pizza, M., & Moxon, E. R. (2019). The development of a vaccine against meningococcus B using reverse vaccinology. *Frontiers in immunology*, 10, 751. <https://doi.org/10.3389/fimmu.2019.00751>

Parvathy, S. T. (2020). Engineering plants as platforms for production of vaccines. *American Journal of Plant Sciences*, 11(5), 707-735. 10.4236/ajps.2020.115052

Pogrebnyak, N., Golovkin, M., Andrianov, V., Spitsin, S., Smirnov, Y., Egolf, R., & Koprowski, H. (2005). Severe acute respiratory syndrome (SARS) S protein production in plants: development of recombinant vaccine. *Proceedings of the National Academy of Sciences*, 102(25), 9062-9067. 10.1073/pnas.0503760102

Ryzhikov, A. B., Ryzhikov, E. A., Bogryantseva, M. P., Usova, S. V., Danilenko, E. D., Nechaeva, E. A., ... & Maksyutov, R. A. (2021). A single blind, placebo-controlled randomized study of the safety, reactogenicity and immunogenicity of the “EpiVacCorona” Vaccine for the prevention of COVID-19, in volunteers aged 18–60 years (phase I–II). *Russian Journal of Infection and Immunity*, 11(2), 283-296: <http://dx.doi.org/10.15789/2220-7619-ASB-1699>

Satarker, S., & Nampoothiri, M. (2020). Structural proteins in severe acute respiratory syndrome coronavirus-2. *Archives of medical research*, 51(6), 482-491. <https://doi.org/10.1016/j.arcmed.2020.05.012>

Sharma, A., Pal, S., Panwar, A., Kumar, S., & Kumar, A. (2021). In-silico immunoinformatic analysis of SARS-CoV-2 virus for the development of putative vaccine construct. *Immunobiology*, 226(5), 152134. <https://doi.org/10.1016/j.imbio.2021.152134>

Shim, B. S., Hong, K. J., Maharjan, P. M., & Choe, S. (2019). Plant factory: new resource for the productivity and diversity of human and veterinary vaccines. *Clinical and experimental vaccine research*, 8(2), 136-139. <https://doi.org/10.7774/cevr.2019.8.2.136>

Singhal, T. (2020). A review of coronavirus disease-2019 (COVID-19). *The indian journal of pediatrics*, 87(4), 281-286. <https://doi.org/10.1007/s12098-020-03263-6>

Thye, A. Y. K., Law, J. W. F., Pusparajah, P., Letchumanan, V., Chan, K. G., & Lee, L. H. (2021). Emerging SARS-CoV-2 Variants of Concern (VOCs): An Impending Global Crisis. *Biomedicines*, 9(10), 1303.

Walls, A. C., Park, Y. J., Tortorici, M. A., Wall, A., McGuire, A. T., & Veesler, D. (2020). Structure, function, and antigenicity of the SARS-CoV-2 spike glycoprotein. *Cell*, 181(2), 281-292. <https://doi.org/10.1016/j.cell.2020.02.058>

Wrapp, D., Wang, N., Corbett, K. S., Goldsmith, J. A., Hsieh, C. L., Abiona, O., ... & McLellan, J. S. (2020). Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation. *Science*, 367(6483), 1260-1263. <https://10.1126/science.abb2507>

Wu, A., Peng, Y., Huang, B., Ding, X., Wang, X., Niu, P., ... & Jiang, T. (2020). Genome composition and divergence of the novel coronavirus (2019-nCoV) originating in China. *Cell host & microbe*, 27(3), 325-328. <https://doi.org/10.1016/j.chom.2020.02.001>.

Yan, R., Zhang, Y., Li, Y., Xia, L., Guo, Y., & Zhou, Q. (2020). Structural basis for the recognition of SARS-CoV-2 by full-length human ACE2. *Science*, 367(6485), 1444-1448. <https://doi.org/10.1126/science.abb2762>