



BENEMÉRITA UNIVERSIDAD AUTÓNOMA DE PUEBLA  
FACULTAD DE CIENCIAS QUÍMICAS  
CENTRO DE QUÍMICA-INSTITUTO DE CIENCIAS  
POSGRADO EN CIENCIAS QUÍMICAS

Clonación del gen M del virus de la diarrea epidémica porcina para su expresión en la levadura  
*Pichia pastoris*, generación del modelo 3D de la proteína y análisis teórico de sus  
propiedades inmunogénicas

TESIS PARA OBTENER EL GRADO DE MAESTRO EN CIENCIAS QUÍMICAS  
EN EL ÁREA DE BIOQUÍMICA Y BIOLOGÍA MOLECULAR

PRESENTA

Lic. En Biotecnología Alan Rodríguez Enríquez

DIRECTORA DE TESIS

Dra. Nora Hilda Rosas Murrieta  
ICUAP-BUAP

Mayo 2022

Puebla, Puebla.

## ÍNDICE

Resumen.....	6
Introducción.....	7
I.I Importancia de la industria porcina en el planeta y en México.....	7
I.II Principales infecciones que afectan la producción porcina.....	8
I.III Diarrea epidémica porcina.....	9
I.IV Virus de la diarrea epidémica porcina.....	10
I.V Estrategias de control para el vDEP.....	12
I.VI Proteína M.....	13
I.VII Expresión heteróloga de proteínas en <i>Pichia pastoris</i> .....	13
II. Planteamiento de problema, justificación e hipótesis.....	15
II.I Planteamiento del problema.....	15
II.II Justificación.....	15
II.III Hipótesis.....	15
III. Objetivos.....	16
III.I Objetivo general.....	16
III.II Objetivos particulares.....	16
IV. Materiales y métodos.....	17
IV.II Clonación del gen M del vDEP en el vector pPICZαB de <i>Pichia pastoris</i> .....	17
IV.II.I Amplificación de ADN del gen M mediante PCR.....	17
IV.II.II Gel de agarosa y electroforesis.....	17
IV.II.III Purificación de fragmentos de ADN a partir de gel de agarosa.....	17
IV.II.IV Extracción de ADN plasmídico.....	17
IV.II.V Transformación de <i>E. coli</i> DH5α y selección de cepas transformadas.....	18
IV.II.VI Digestiones enzimáticas.....	18
IV.II.VII Clonación de la secuencia codificante de M en el plásmido pJET.....	18
IV.II.VIII Subclonación de la secuencia codificante de M en el plásmido pPICZαB.....	18
IV.II.IX Orientación de los fragmentos clonados en pPICZαB.....	18
IV.II.X Transformación de los plásmidos en <i>Pichia pastoris</i> .....	18
IV.III Predicción de la estructura 3D de la proteína M.....	19
IV.III.I Secuencias aminoacídicas de la proteína M.....	19
IV.III.II Análisis <i>in silico</i> de la proteína M.....	19
IV.III.III Programas para la búsqueda de plantillas para la generación de modelos 3D.....	19

IV.III.IV Programas para la generación de modelos 3D de proteínas.....	19
IV.III.V Programas para la evaluación de modelos 3D de proteínas.....	20
IV.III.VI Modelos de la proteína M del SARS-CoV-2.....	22
IV.III.VII Superposición de modelos 3D de proteínas.....	22
IV. IV Identificación de epítomos en la proteína M.....	22
IV.IV.I Identificación de epítomos continuos de células B.....	22
IV.IV.II Identificación de epítomos discontinuos de células B.....	23
IV.IV.III Identificación de epítomos de células T.....	23
V. Resultados .....	24
V.I Clonación del gen M del vDEP en el vector pJET1.2/blunt.....	24
V.II Clonación del gen M del vDEP en el vector pPICZαB de <i>Pichia pastoris</i> .....	26
V.III Secuenciación de los plásmidos pPICZαB-M <sup>2-226</sup> y pPICZαB-M <sup>102-212</sup> .....	28
V.V Predicción de la estructura 3D de la proteína M.....	35
V.V.I Predicción del modelo 3D de la proteína M.....	39
V.VII Identificación de epítomos en la proteína M del vDEP.....	47
V.VII.I Identificación de epítomos de células B lineales en el modelo de proteína vDEP M...47	47
V.VII.II Predicción de epítomos discontinuos de células B del modelo de proteína M de vDEP...48	48
V.VII.III Predicción de epítomos de células T para la proteína M de vDEP.....	52
V.VIII Revisión bibliográfica del VDEP y del SARS-CoV-2 para demostrar las similitudes y diferencias de ambos coronavirus.....	57
VI. Discusión.....	65
VII. Conclusión .....	68
VIII. Perspectivas del Trabajo .....	68
IX. Bibliografía.....	69
X. ANEXO .....	80
X.I Publicación.....	80

## Índice de ilustraciones

Ilustración 1. Mapa de la distribución del vDEP en México. ....	10
Ilustración 2. Clonación de los fragmentos M <sup>2-226</sup> y M <sup>102-212</sup> en el plásmido pJET1.2/blunt. ....	24
Ilustración 3. Mapa de los plásmidos construidos.....	25
Ilustración 4. Clonación de los fragmentos M <sup>2-226</sup> y M <sup>102-212</sup> en el plásmido pPICZαB.....	25
Ilustración 5. Análisis en clustal omega de los resultados de secuenciación.....	28
Ilustración 6. Análisis de secuencia en Jalview. ....	29
Ilustración 7. Resultados de secuenciación del plásmido pPICZαB-M <sup>2-226</sup> . ....	30
Ilustración 8. Resultados de secuenciación del plásmido pPICZαB-M <sup>102-212</sup> .....	31
Ilustración 9. Resultados del alineamiento .....	32
Ilustración 10. Análisis de Clustal Omega de la secuencia aminoacídica de la proteína M <sup>2-226</sup> .....	33
Ilustración 11. Fragmentos de PCR amplificados a partir del genoma de las cepas de <i>P. pastoris</i> ..	34
Ilustración 12. Topologías y regiones transmembranales de la proteína M.....	34
Ilustración 13. Árbol filogenético de las secuencias utilizadas .....	35
Ilustración 14. Múltiples alineaciones de las secuencias de la proteína M de vDEP .....	36
Ilustración 15. Modelos 3D predichos de las proteínas M de CV777 y M de 2013MMV, .....	39
Ilustración 16. Evaluación del modelo 3D de proteína M CV777 y 2013MMV .....	41
Ilustración 17. Evaluación en web ProSA del modelo 3D de la proteína M CV777 y 2013MMV.....	43
Ilustración 18. Modelos de la proteína M del SARS-CoV-2. ....	44
Ilustración 19. Modelos de la superposición de la proteína M de SARS-CoV-2 .....	44
Ilustración 20. Epítomos discontinuos predichos en ElliPro .....	48
Ilustración 21. Epítomos discontinuos con DiscoTope.....	50
Ilustración 22. Patrones de unión de los epítomos de la proteína M con SLA-01: 0401. ....	54
Ilustración 23. Resumen de epítomos en la proteína M de CV777 y M de 2013MMV.....	56

## Índice de tablas

Tabla 1. Comparación de los mejores valores de QMEAN de los diferentes modelos de la proteína M realizados con diferentes programas.....	39
Tabla 2. Estadísticas de la gráfica de Ramachandran de los modelos M de CV777 y M de 2013MMV.....	40
Tabla 3. Evaluación usando el servicios MolProbity de la proteína M de CV777 y M de 2013MMV generados en Robetta.....	42
Tabla 4. Resumen de los valores de los programas utilizados para evaluar la calidad de los modelos 3D.....	45
Tabla 5. Epítomos continuos de células B de la proteína M de CV777.....	46
Tabla 6. Predicción de epítomos discontinuos con ElliPro.....	48
Tabla 7. Predicción de epítomos discontinuos con DiscoTope.....	50
Tabla 8. Epítomos continuos de células T de la proteína M de CV777 .....	52
Tabla 9. Energía de unión de los epítomos predichos para el alelo SLA-01: 0401 usando Autodock.....	53

## Resumen

El virus de la diarrea epidérmica porcina (vDEP) es un virus reemergente que se encuentra en México y en distintos países, ocasiona vómito y diarrea acuosa en cerdos y tiene una alta tasa de mortalidad en lechones. La proteína de membrana M del vDEP es el componente más abundante de la envoltura viral. La proteína M juega un papel central en la morfogénesis y ensamblaje del virus a través de interacciones proteicas de tipo M-M, M-Spiga (S) y M-nucleocápside (N). Se sabe que la proteína M induce anticuerpos protectores en cerdos y participa en la respuesta antagónica del sistema antiviral celular coordinado por la vía del interferón tipo I y tipo III. Por sus características de la proteína M en este trabajo se decidió clonar la proteína M en el vector de expresión pPICZ $\alpha$ -B para su posterior expresión en el sistema de levadura de *P. pastoris*. La estructura 3D de la proteína vDEP M aún se desconoce. El presente trabajo presenta un modelo 3D predicho de la proteína M generada usando el protocolo de Robetta. El modelo de proteína M está organizado en una región transmembranal y una globular. El modelo 3D obtenido de la proteína vDEP M se comparó con modelos 3D de la proteína SARS-CoV-2 M creados con redes neuronales y con modelos iniciales basados en aprendizaje automático creados con trRosetta. El modelo 3D del presente estudio predijo cuatro epítomos de células B lineales (los péptidos RSVNASSGTG y KHGDYSAVSNPSALT son dignos de mención), seis epítomos de células B discontinuos, cuarenta epítomos de células T de unión débil y catorce de unión fuerte en la proteína CV777 M. Se observó un alto grado de conservación de los epítomos predichos en la proteína M entre diferentes cepas de vDEP aisladas en distintos países. La energía de unión *in silico* entre los epítomos predichos de la proteína M de vDEP y las moléculas de MHC-I porcina osciló entre -7.8 y -5.5 kcal/mol. Los datos obtenidos sugieren que la proteína M podría ser una potencial candidata para el desarrollo de nuevos tratamientos o estrategias que activen mecanismos celulares protectores frente a enfermedades virales.

## Introducción

Los coronavirus (CoV) infectan una gran cantidad de organismos causando enfermedades respiratorias e intestinales tanto en humanos como en animales y tienen un alto impacto en la salud y en la economía en cada emergencia y reemergencia ocurrida. Tal es el caso del virus de la diarrea epidémica porcina (DEP) que se encuentra en diferentes países y de SARS-CoV-2 causante de la pandemia iniciada en 2019. En la búsqueda de la explicación de la alta capacidad de los coronavirus para invadir a varias especies, la evidencia filogenética ha demostrado que los murciélagos y los roedores sirven como fuente genética de la mayoría de los alfa-CoV y beta-CoV, mientras que las aves son el principal reservorio de gamma-CoV y delta-CoV, las cuales son especies que por sus características físicas, permiten la transferencia viral a otras especies (1). La mayor diversidad genética se encuentran en los coronavirus de murciélagos, con respecto a los de las demás especies, por lo que puede considerarse que son los reservorios naturales de esta clase de virus (2). La diversidad de coronavirus es el resultado de la poca fidelidad de la ARN polimerasa, la alta frecuencia de recombinación de ARN homólogo y los genomas grandes de los coronavirus.(3). Con el número creciente de coronavirus, se han observado, cada vez más, coronavirus estrechamente relacionados de animales con parientes más lejanos, que fueron el resultado de saltos interespecie y pueden ser la causa de brotes desastrosos de enfermedades zoonóticas (4). En este sentido, se ha demostrado que los CoV de murciélagos tienen un potencial zoonótico así como un alto potencial de recombinación lo que podría favorecer la aparición de nuevos CoV que sean más transmisibles o mortales en los seres humanos, animales domésticos o de crianza y que tengan un alto impacto en la economía (5). Actualmente el vDEP ha reemergido en diferentes países, entre ellos México, y, aunque se han desarrollado biológicos y medicamentos para su control, estos tratamientos parecen ser cada vez menos eficientes en el tratamiento contra el vDEP (6), por lo que actualmente la comunidad científica se encuentra en la investigación y desarrollo de nuevos procedimientos para el tratamiento de infecciones producidas por coronavirus. En este trabajo se investiga de manera *in silico* el potencial inmunológico de la proteína M para su futura investigación como biológico de una vacuna multivalente, además, de realizar un modelo 3D de la proteína para futuras investigaciones

### I.I Importancia de la industria porcina en el planeta y en México

La carne de cerdo es la más consumida en el mundo a pesar de algunas restricciones culturales y religiosas. La producción global de carne de cerdo se concentra en un 75% en solo 3 regiones, China, la Unión Europea y Estados Unidos. México representa solo el 1.1% de la producción mundial de carne de cerdo y es el segundo importador más grande del mundo, ya que representa el 9% de las transacciones de cerdo en el ámbito internacional (7). La carne de cerdo ha mantenido su costo en los últimos años, a pesar del incremento que se ha realizado sobre los precios a lo largo del tiempo, su costo no ha aumentado en una tasa tan alta como la carne de pollo y de res (OECD-FAO, 2018 OECD-FAO Agricultural Outlook 2017-2026 and 2018-2027). En 2017 el consumo mundial de carne per cápita se estimó en 34.3 kg, para el año 2026 se estima que sea de 34.6 Kg, para México el consumo per cápita es de 12 Kg. En México la producción de carne de cerdo representa una quinta parte de la producción nacional de carne, con un aumento anual del 2.1%; sin embargo, el consumo de esta carne también ha aumentado, lo que ha generado un aumento en la importación de carne, lo que representa un 45% del total de consumo de carne, Estados Unidos es el principal exportador de carne para México (COMECARNE, 2017 <https://comecarne.org/compendio-estadistico-2017/>). La producción nacional de carne de cerdo se concentra en solo 6 estados (Jalisco,

Sonora, Puebla, Yucatán, Veracruz y Guanajuato), que representa el 76% de la demanda y el consumo de carne se concentra en 5 estados (Estado de México, CDMX, Veracruz, Jalisco y Puebla) y representa el 51.3% del consumo total (7) (8). La producción de cerdos incluye las etapas de gestación, destete, pre-engorde y engorda, algunas granjas llevan a cabo las diferentes etapas en diferentes tiempos y sitios claramente diferenciados para garantizar una mejor eficiencia alimenticia y un mejor control de sanidad. Los indicadores de productividad en las diferentes etapas son desfavorables para México. Las carnicerías representan el 55% de las ventas de la producción nacional de carne, los mercados el 21% y los supermercados el 11% (9).

## I.II Principales infecciones que afectan la producción porcina

Existen diferentes patógenos que afectan a los cerdos en diferentes etapas de vida, provocan diferentes enfermedades y síntomas, con tasas de mortalidad variables y de ordenes diferentes, algunos de estos patógenos se han esparcido por casi todo el mundo (10). En la revisión bibliográfica realizada por Kimberly Vanderwaal y John Deen se reportaron hasta 40 patógenos de cerdo los cuales incluyen 16 virus, 15 bacterias, 8 parásitos y 1 protozooario (10). Entre los patógenos más importantes se encuentran el virus que provoca el síndrome respiratorio y reproductivo porcino (PRRS), una de las enfermedades porcinas más importantes desde el punto de vista económico, a la industria porcina de los Estados Unidos le cuesta al menos \$600 millones de dólares anuales (11). El agente etiológico de esta enfermedad es el virus del síndrome respiratorio y reproductivo porcino (PRRSV) es un virus de ARN de cadena positiva, que pertenece al orden *Nidovirales* y a la familia *Arteriviridae* (12). Otra enfermedad de importancia es la ascariasis, causada por *Ascaris suum*, un nematodo esparcido por todo el mundo es una de las enfermedades parasitarias más comunes de los cerdos transmitidas por el suelo. Esta enfermedad conduce a grandes pérdidas económicas para la industria porcina como consecuencia de la reducción de peso del cerdo y una baja eficiencia de conversión alimenticia en los cerdos (13). La peste porcina africana (ASF) es una enfermedad viral que afecta a los cerdos, conduce a una alta mortalidad en cerdos, causa importantes pérdidas económicas que son inevitables en ausencia de una vacuna efectiva y los métodos disponibles para el control de enfermedades son la cuarentena del área afectada y el sacrificio de los animales infectados, la ASF es causada por el virus ASF, un virus de ADN de doble cadena, único miembro de la familia *Asfarviridae* y el único virus de ADN transmitido por artrópodos (14). *Streptococcus suis* es uno de los patógenos más importantes en la industria porcina que causa septicemia y meningitis en los cerdos infectados, adicionalmente, es de relevancia este patógeno bacteriano debido a que es capaz de infectar a humanos, distribuido actualmente en casi todo el mundo y con miles de casos reportados hasta la fecha de personas infectadas por contacto con algún cerdo contagiado (15). *Senecavirus A* (SVA) es el único miembro del género *Senecavirus* dentro de la familia *Picornaviridae* descubierto en el año 2002. Este virus provoca erosiones y vesículas en la piel, la cavidad oral y el hocico. El virus posee un genoma de ARN de cadena sencilla positiva de aproximadamente 7.3 kb. Se ha reportado la mayoría de casos de estos cerdos en Estados Unidos y Brasil (16). *Circovirus porcino* tipo 2 (PVC2), *Virus de la influenza de cerdo* (SIV), *Actinobacillus pleuropneumoniae* (APP) y *Mycoplasma hyopneumoniae* (MHYO) son patógenos que infectan a cerdos ocasionando el complejo de enfermedad respiratoria porcina, una de las enfermedades mayormente responsables de pérdidas económicas a la industria porcina alrededor del mundo. La combinación de estos patógenos infectando simultáneamente al cerdo se ha visto que sufren el complejo

de enfermedad respiratoria, que aumenta la mortalidad de los cerdos enfermos (17)(18)(19). El deltacoronavirus porcino (PDCoV), el virus de la diarrea epidémica porcina (vDEP), el coronavirus respiratorio porcino (PRCV) y el virus de la gastroenteritis transmisible (TGEV), son todos coronavirus que infectan a cerdos, todos pertenecen a la familia de los *Coronaviridae*, el vDEP y el TGEV pertenecen al género de los *alfacoronavirus*, mientras que el PDCoV al género *deltacoronavirus*, todos poseen un genoma de ARN de cadena sencilla positiva. Estos virus se caracterizan por ser emergentes y reemergentes, afectan principalmente a lechones neonatales, causando diarrea acuosa, vómito, deshidratación y tienen una alta mortalidad, los coronavirus han afectado a la industria porcina en diferentes países y en diferentes épocas debido a su reemergencia. El PDCoV se detectó por primera vez en 2012 en Hong Kong y posteriormente en 2014 en Estados Unidos. El TGEV fue reportado por primera vez en 1964 en los Estados Unidos y posteriormente en Europa y Asia. El PRCV fue detectado en Bélgica, se consideraba que proviene del TGEV por su alto porcentaje de identidad nucleotídica, sin embargo, la eliminación de una secuencia significativa del gen de la proteína S entre el PRCV y el TGEV es lo que sirve para diferenciar entre estos dos virus (20) (21). El vDEP generó en Estados Unidos una reducción en la cantidad lechones entre los años 2013 y 2014 (22). El inventario anual de producción porcina de Estados Unidos sufrió cambios en el año del vDEP (2014), la totalidad de sacrificios de cerdos comerciales en el año 2014 tuvo un 4.64% (5201000 cerdos) de decremento en comparación con el año 2013, además, hubo un decremento del 2.68% (13385500 libras) en la cantidad de libras exportadas en el año 2014 en comparación del 2013 (22). En el año del vDEP la industria porcina de Estados Unidos perdió millones de dólares debido a los cambios económicos en esta industria generados por la aparición del vDEP (23).

### I.III Diarrea epidémica porcina

La diarrea epidémica porcina (DEP) es una enfermedad económicamente importante que afecta cerdos en todas las edades, esta enfermedad está caracterizada por presencia de diarrea, vómito y deshidratación, tiene una tasa más alta de mortalidad en lechones neonatales que en cerdos adultos (24), el agente causal de esta enfermedad es el virus de la diarrea epidémica porcina (vDEP), el cual es un virus de genoma de ARN de cadena sencilla positivo, perteneciente al género *alfacoronavirus* en la familia *Coronaviridae* del orden *Nidovirales*. El primer reporte de este virus fue en Inglaterra, realizado por Wood en 1977, en el que se describe un nuevo aparente síndrome de la diarrea epidémica porcina (25). En 1978 Pensaert & de Bouck, ubicados en Bélgica, fueron los primeros científicos que evidenciaron que el vDEP era un virus diferente de los otros dos coronavirus conocidos en esa época, el TGEV y el virus de encefalomiélitis hemaglutinante, desde entonces se utilizó el nombre de diarrea epidémica porcina vDEP (26). Durante la década de los ochenta, el virus apareció en diferentes países de Europa, en Países Bajos (27), Hungría (28), nuevamente en Inglaterra en 1999 (29), en Italia en 2005 (30). El primer reporte en Asia fue en 1982, el cual ha tenido desde ese entonces un gran impacto económico debido a su ocurrencia, y ha involucrado países como China (31), Tailandia (32), Japón (33) y Corea del sur (34). En 2013 el virus llegó a Estados Unidos, el cual se esparció rápidamente por todo el país causando pérdidas económicas importantes, infectando a cerdos de todas las edades, posteriormente el virus se esparció por Canadá y México (35). En Mayo de 2014 el Servicio Nacional de Sanidad, Inocuidad y Calidad Agroalimentaria (SENASICA) reconoció la presencia de vDEP en México, confirmación realizada mediante el aislamiento y detección molecular del virus en cerdos con señales de la enfermedad (36) (37). En México se han estudiado cepas de 15 estados del país (CDMX,

Guanajuato, Hidalgo, Jalisco, Michoacán, Morelos, Nuevo León, Oaxaca, Puebla, Querétaro, San Luis Potosí, Sonora, Tlaxcala y Veracruz (**Ilustración 1**), los cuales han reportado brotes entre 2013 y 2018 (38) (39).



**Ilustración 1. Mapa de la distribución del vDEP en México (38).**

#### I.IV Virus de la diarrea epidémica porcina

Ensayos sobre las propiedades biológicas y fisicoquímicas del vDEP demostraron que es estable entre 4 y 50°C, el virus es completamente inactivado a pH menor a 4 y mayor a 9 (40). Pospischil y colaboradores demostraron que diferentes desinfectantes son capaces de inactivar al vDEP, entre ellos se encuentran Virkon S un agente oxidante, detergentes iónicos, cloroformo, entre otros (41). Para el aislamiento del virus mediante cultivo celular se han utilizado dos líneas celulares de riñón de mono verde africano (células Vero y células MARC-145), estas líneas celulares permiten el aislamiento y la propagación serial del vDEP, los medios son suplementados con la proteasa tripsina, la cual juega un papel importante en la entrada celular y liberación del virión del vDEP (42). Los efectos citopáticos derivados de la infección de vDEP consisten en la vacuolación y formación de sincitios (40). El genoma del vDEP se completó en el 2001, consiste de 28 Kb de largo, dos tercios del genoma codifican para proteínas no estructurales y un tercio para proteínas estructurales (43). En años posteriores se completaron genomas de diferentes cepas vDEP aisladas en distintos países como China (44), (45) y Corea (46). Alineamientos de genomas de las diferentes cepas de vDEP secuenciados hasta ese momento demostraron que estaban mayormente relacionados con el alfa-coronavirus de murciélago, lo que sugiere una transmisión inter-especies entre el murciélago y el cerdo (47), (37). La transmisión del vDEP se da mediante la ruta fecal-oral, además, las partículas virales contenidas en el aire de un lugar con cerdos positivos a vDEP pueden infectar a otros cerdos sin llegar por la ruta fecal, lo que posibilita aún más la transmisión de este virus (48). Excreciones diarreas, vómito y otros fómites contaminados como remolques de transporte (49) y comida (50) son medios de transmisión del virus. Estudios en los cuales han detectado el ARN del vDEP en muestras de leche contaminadas de cerdas madres sugieren que podría ser otro medio por el cual puede transmitirse el virus (45). El vDEP tiene tropismo por los enterocitos vellosos del intestino delgado del cerdo, este tejido se caracteriza por que expresan grandes cantidades de aminopeptidasa N (APN), una proteína transmembranal glicosilada, la cual se ha identificado como receptor celular para el vDEP (46). El vDEP es citolítico y causa necrosis aguda en los enterocitos infectados, lo que lleva a una atrofia vellosa marcada en el intestino delgado, que provoca la diarrea acuosa

característica de los cerdos infectados (51). También se ha demostrado que el vDEP es capaz de infectar y de replicarse en macrófagos alveolares porcinos (52), esto proporciona nuevos conocimientos sobre la patogénesis y abre la pauta para pensar si existen algún otro órgano donde el virus pueda infectar y reproducirse. La membrana viral de los coronavirus contiene tres grandes proteínas estructuras, la glicoproteína S o espícula, que forma los peplómeros de 15 a 20 nm, encontrados en la envoltura viral, la glicoproteína transmembranal M y la proteína de la nucleocápside fosforilada interna N. Existe además una pequeña proteína transmembranal E y una proteína de envoltura adicional con funciones de hemaglutinación y esterasa HE (53). Los coronavirus son pleomórficos frecuentemente esféricos de 80 a 220 nm. El genoma de los coronavirus es de ARN de cadena sencilla positiva, con un *cap* en el extremo 5', mientras que el extremo 3' se encuentra poliadenilado. La expresión de genes individuales ocurre a través de la producción de conjuntos de ARNm anidados, todos comparten la misma secuencia del extremo 5'. El genoma de los coronavirus posee entre 7 y 14 marcos de lectura abiertos. Desde el extremo 5', el gen 1, que ocupa dos tercios del genoma completo del virus, cerca de 20 a 22 kb de largo, el cual se compone de dos marcos de lectura traslapados, el 1a y el 1b que funcionan colectivamente para la expresión de la ARN polimerasa viral, mientras que el resto del genoma codifica a las proteínas estructurales S, E, M y N (54). Los ORF 1a y 1b son traducidos en dos poliproteínas, idénticas en su extremo N terminal pero diferentes en el extremo C terminal debido a la extensión generada por un cambio de marco, estas poliproteínas son precursores de proteínas utilizadas en el proceso de transcripción y replicación. Todos los coronavirus producen una proteasa similar a la quimotripsina llamada M<sup>pro</sup> o 3CL<sup>pro</sup> debido a su parecido con la proteasa 3C de *picornavirus*. Estas proteasas procesan la poliproteína para producir hasta 16 proteínas no estructurales con diferentes funciones. El ORF 1b codifica la ARN polimerasa dependiente de ARN y una helicasa (55). Existe otro marco de lectura abierto que codifica una proteína estructural. El ORF 3a codifica una proteína O-glicosilada capaz de unirse a las proteínas N, M y S, lo que sugiere un papel en la biogénesis viral (56).

La primera interacción de la partícula viral con la célula hospedera es mediante la proteína S cuyo dominio de unión al receptor se encuentra en la mitad del extremo N-terminal, mientras que la mitad del extremo C-terminal de esta proteína posee actividad fusogénica y está anclada a la membrana celular, permite la fusión de la envoltura viral con la membrana, posteriormente, la nucleocápside viral se libera en el citoplasma (57). El primer paso de la replicación viral consiste en la traducción de los dos marcos de lectura abiertos largos, los ORFs 1a y 1b, a partir del genoma de ARN viral que sirve como ARNm, cada ORF codifica unidades para la ARN polimerasa dependiente de ARN, una vez ensamblada la ARN polimerasa, esta genera toda la cadena de ARN positiva en una cadena negativa de ARN. A partir de la cadena negativa ARN, la ARN polimerasa viral transcribe toda la cadena sencilla positiva de ARN, además, transcribe un conjunto anidado de ARNm subgenómicos superpuestos en el extremo 3'-coterminar (58). Cada ARNm difiere en longitud pero comparte secuencias comunes en el extremo 3' y alrededor de 70 nucleótidos en la secuencia líder 5', solo la región que no comparten entre los grupos de ARNm es la región que se traducirá en proteína viral (43). Para el proceso de maduración y ensamblaje del virus, la proteína S es insertada cotraduccionalmente en el retículo endoplásmico rugoso (RER) y glicosilado con N-glicanos, la proteína S forma trímeros antes de ser exportada fuera del retículo endoplasmático y es entonces cuando interactúa con las proteínas M y E para migrar al sitio de ensamblaje viral (59).

Una característica de los coronavirus es el ensamblaje del virión y la gemación en el lumen del retículo endoplásmico ER y el compartimento intermedio de Golgi (ERGIC), seguido de su liberación por exocitosis (60). Las proteínas M, S y E deben de ser retenidas en el compartimento ERGIC para un ensamblaje eficiente del virión. La proteína M tiene un papel fundamental en la formación de la envoltura viral, diferentes coronavirus requieren solo la proteína M o combinaciones con las demás proteínas estructurales para la formación de la partícula viral, (61), (62). Es generalmente aceptado que la proteína S de los coronavirus es la responsable de la unión e internalización del virus a la célula. También se ha demostrado que es necesaria la presencia de heparán sulfato en la superficie celular como un factor de unión (63), (64), (65), (66). Recientemente, se ha descrito para el coronavirus humano NL63 que la proteína M es la responsable de la interacción con el receptor, no de la internalización, pero si una proteína clave para el proceso de infección permitiendo una interacción y unión eficiente entre el virus y la célula (67). La proteína M (~230 aminoácidos, 22-25 kDa) es la proteína más abundante de la envoltura viral de los coronavirus, consiste en un ectodominio N-terminal corto glicosilado, tres dominios transmembranales (TM) desde el N-terminal y un dominio largo de cola citoplasmática (CT) C-terminal. La topología de la proteína M consiste en un ectodominio N-terminal y un endodominio C-terminal (N<sup>exo</sup>-C<sup>endo</sup>) (68)(69) (70); sin embargo, existen diferencias topológicas de la proteína M entre algunos coronavirus como es el caso del TGEV, en el cual existe la topología N<sup>exo</sup>-C<sup>endo</sup> y N<sup>exo</sup>-C<sup>exo</sup>. Los tres dominios hidrofóbicos de la proteína M pueden funcionar como secuencia de señal de anclaje, cada dominio permite individualmente la integración estable dentro de la membrana, aunque se modifica} o elimina los dominios hidrofóbicos dentro de la membrana se puede alterar la orientación de la proteína en la membrana (71). La proteína M sigue la ruta de síntesis del retículo endoplasmático-Golgi-ERGIC (72). La proteína M es N-glicosilada y adquiere N-glicanos complejos que se generan en el Golgi (72), este proceso ocurre en coronavirus como el TGEV, FIPV y CoV-229E; sin embargo, en algunos coronavirus como el MHV y el CoV-OC43 la proteína M es O-glicosilada (73). La proteína M del TGEV es responsable de la inducción de IFN- $\alpha$  en leucocitos (59), además, se ha demostrado para la proteína M del MHV que, el efecto inductivo de la proteína M, depende del estado de glicosilación en el que se encuentre, ya que la N- glicosilación es más interferogénico que la O-glicosilación o la nula glicosilación de la proteína M.

#### **I.V Estrategias de control para el vDEP**

Desde el brote de vDEP en Estados Unidos en 2013 la SAGARPA, a través del SENASICA, restringió la importación de cerdos vivos de Estados Unidos de América y extrema las acciones de vigilancia epidemiológica. Se extremaron las acciones de vigilancia epidemiológica en zonas porcícolas, a fin de identificar el incremento anormal en la mortalidad de cerdos (74). El SENASICA y las organizaciones de porcicultores integraron un Grupo Técnico de Sanidad Porcina para dar seguimiento a las acciones de prevención y control de la Diarrea Epidémica Porcina en México (75). A fin de informar a los productores puntualmente, la dependencia de la SAGARPA desarrolló material de difusión para mejorar las medidas de bioseguridad y motivarlos a que apliquen en sus granjas las buenas prácticas de producción pecuaria. El SENASICA informó a los productores que es preciso incrementar y reforzar las medidas de bioseguridad en las granjas, entre las que se encuentran la delimitación de las unidades de producción por medio de bardas, cercas o muros perimetrales; el control de accesos, así como contar con arcos, vados y tapetes sanitarios. Entre las medidas que es indispensable adoptar, se encuentran restringir la entrada de personas, material,

equipo y enseres, así como en las aéreas de embarcaderos y salida de cerdos. Utilizar ropa y calzado exclusivo para el personal de la granja; lavar y desinfectar botas y calzado, así como evitar la entrada de camiones y vehículos que transporten cerdos que no vengan limpios y desinfectados, entre otras (75).

Existen diferentes compuestos con actividad antiviral que han sido aplicados contra el vDEP como la narasina (No. De patente WO2016/077282 A1), compuestos como triclorometiazida, D-(+) biotona y el glutatión han demostrado inhibir la replicación viral del vDEP (76). Otros compuestos como el LiCl, la lectina griffithsina de la alga roja y la quercetina 7-O-ramnosido han demostrado ser capaces de inhibir al vDEP *in vitro* (77)(78)(79). Los productos de fermentación de *Bacillus licheniformis* han logrado un efecto antiviral contra el vDEP (80). En cuestión de vacunas se han creado una variedad de estas contra el vDEP a lo largo de estos años en diferentes países debido a la variación de cepas que se han presentado y la evolución genética a través del tiempo de estas mismas. La primera vacuna para el vDEP en Estados Unidos fue desarrollada por Harrisvaccines™ iPED vaccine, basada en una versión truncada de la proteína S del vDEP (81). Se han creado otras vacunas con base en la proteína S del vDEP (82)(83). Otras vacunas se han formulado con base en cepas de vDEP atenuadas e inactivadas (84) de una cepa vDEP atenuada mediante propagación serial en células Vero (85)(86)(87). También se han formulado vacunas recombinantes utilizando probióticos como sistemas de liberación (88), o vectores como baculovirus (89) y adenovirus (90).

#### **I.VI Proteína M**

La proteína M se encuentra embebida en la membrana del virus, es la más abundante de todas las proteínas estructurales presentes. Es una proteína de membrana de triple hélice con una orientación N<sup>exo</sup>-C<sup>endo</sup>. La proteína M está relacionada con la inhibición de la respuesta inmune innata a través del bloqueo de la vía de interferón de tipo I y tipo III en la célula (91). Se ha reportado que la proteína M induce la producción de anticuerpos en el cerdo (92)(93)(94), en diferentes cepas de vDEP y en cultivo celular se conoce que la secuencia codificante de la proteína M del vDEP presenta menor variabilidad genética que la proteína S (95). La proteína M cumple varias funciones en la replicación viral, las cuales deben realizarse mediante diferentes regiones de la proteína (96). Existen diferentes proteínas cristalizadas del vDEP reportadas en la base de datos PDB, tal como la proteína S (97), la proteasa 3CLpro (98), NSP9 (99) y NSP1. Hasta el momento no se conoce la estructura tridimensional de la proteína M. Debido a la relevancia de la proteína M en el ciclo de infección viral, y a la falta de la estructura cristalizada de la proteína M del vDEP en el PDB, en el presente trabajo se generó un modelo 3D de la proteína M.

#### **I.VII Expresión heteróloga de proteínas en *Pichia pastoris***

La levadura *Pichia pastoris* también llamada *Komagataella pastoris*, es una especie de levadura metilotrófica (capaz de crecer en metanol como única fuente de carbono), se utiliza con frecuencia como un sistema de expresión para la producción de proteínas (100). Tiene una alta tasa de crecimiento, también cuenta con la capacidad de crecer en un medio simple y económico, por ello se vuelve adecuada para la producción a pequeña y gran escala. Esta levadura es similar a la bien estudiada *Saccharomyces cerevisiae* con condiciones de crecimiento y tolerancias similares (101). Es utilizada tanto para investigación básica como para fines industriales, debido a su fácil manipulación genética, a los altos niveles de producción de la

proteína de interés, la capacidad de realizar modificaciones postraduccionales similares a las de los organismos eucariontes superiores (glicosilación, puentes disulfuro y procesamiento proteolítico) que generan un plegamiento correcto de la proteína (102). Esta levadura como sistema de expresión cuenta con una amplia gama de vectores con diferentes características de producir proteínas de interés. Emplea vectores de integración al cromosoma de la levadura. En estos vectores se puede integrar un factor que señala a las proteínas recombinantes para ser expulsadas al medio extracelular, lo que facilita su posterior purificación (103). La producción de proteínas recombinantes en la levadura *Pichia pastoris* tiene grandes ventajas sobre otros sistemas de expresión procariotas y eucariotas, tasa de crecimiento rápido, acoplado a una fácil fermentación celular de alta densidad, por la generación de biomasa, altos niveles de expresión en un medio casi libre de proteínas, debido a la integración de etiquetas como el  $\alpha$  factor, eliminación de la contaminación con endotoxinas y bacteriófagos, fácil manipulación genética y disponibilidad para la caracterización de vectores de expresión de levaduras, el desconocimiento de patogenicidad humana, puede realizar diversas modificaciones postraduccionales que incluyen: plegamiento del polipéptido, glicosilación, metilación, acilación, regulación proteolítica y orientación a los espacios subcelulares, capacidad de secretar proteínas que pueden ser purificadas en el medio de cultivo libre de células (104).

## II. Planteamiento de problema, justificación e hipótesis

### II.I Planteamiento del problema

El virus de la diarrea epidémica porcina es un coronavirus endémico, emergente y reemergente cuyas estrategias de control dirigidas a la contención del virus tienen fallas que permiten la contaminación cruzada entre cerdos incrementa así el número de animales contagiados y hacen difícil la eliminación del patógeno. Las vacunas que actualmente se comercializan están dirigidas a cepas que afectan al ganado al momento de su producción, pero debido a que existen cambios genéticos, varios países han optado por montar líneas de investigación para caracterizar genética y estructuralmente las cepas del vDEP con el fin de desarrollar vacunas específicas basadas en la proteína S del coronavirus que tiene una alta tasa de variación genética. México es un país afectado por el vDEP, y al igual que en Estados Unidos, diferentes cepas de vDEP se han encontrado y caracterizado para el desarrollo de nuevas vacunas contra el vDEP. El análisis comparativo entre proteínas virales ha mostrado que la proteína M es genéticamente más estable que la proteína S, y la proteína estructural con mayor presencia del vDEP, resultaría interesante considerar su uso en la producción de una vacuna alternativa, la cual generaría una mayor protección contra un mayor número de cepas de vDEP. Se debe investigar el potencial de la proteína M como producto antigénico que brinde la protección inmunológica a los cerdos contra el vDEP y de la producción de los epítomos necesarios para conseguir la inmunización de los animales.

### II.II Justificación

Debido a las pérdidas económicas, en México la infección por el virus de la diarrea epidémica porcina es controlada mediante el empleo de diversas vacunas con el antígeno S, no obstante, existe la necesidad de generar nuevas herramientas para obtener opciones preventivas de bajo costo, seguras, y adaptables a las variantes antigénicas del vDEP. Estudios recientes muestran que existe una prevalencia serológica importante de vDEP en diferentes regiones del país, en la que circulan cepas con diferentes variaciones genéticas, lo que sugiere que el uso de una vacuna con un solo tipo de antígeno puede no ser eficaz para el control de la enfermedad. Las proteínas recombinantes provenientes de sistemas de expresión como las levaduras pueden generar los mejores resultados gracias a su fácil manipulación y, teniendo en cuenta la relevancia de la proteína M para la adhesión celular y la formación del virión, su producción en sistemas de expresión heterólogos y caracterización podría dar origen al diseño de un posible nuevo recurso preventivo de nueva generación.

### II.III Hipótesis

La proteína M del virus de la diarrea epidémica porcina expresada en la levadura *P. pastoris* mantiene sus propiedades inmunogénicas con respecto a la predicción realizada e induce anticuerpos protectores de la infección viral en cerdos. El potencial inmunogénico de la proteína M puede valorarse a partir del modelo 3D de la proteína M.

### III. Objetivos

#### III.I Objetivo general

Generar un vector de expresión de la proteína M para la levadura *Pichia pastoris* mediante la clonación del gen M del virus de la diarrea epidémica porcina, generar el modelo 3D y analizar *in silico* las características inmunogénicas de la proteína.

#### III.II Objetivos particulares

1. Construir los vectores recombinantes pPICZ $\alpha$ B-M<sub>2-226</sub> (proteína M transmembranal completa del vDEP) y pPICZ $\alpha$ B-M<sub>102-212</sub> (dominio globular intraviral de la proteína M del vDEP).
2. Obtener cepas de *Pichia pastoris* X-33 transformadas con los vectores pPICZ $\alpha$ B-M<sub>2-226</sub> y pPICZ $\alpha$ B-M<sub>102-212</sub>.
3. Generar el modelo 3D de la proteína M del vDEP.
4. Analizar *in silico* las propiedades inmunogénicas de la proteína M.
5. Realizar una revisión bibliográfica del vDEP contrastándola con información de SARS-CoV-2.

## IV. Materiales y métodos

### IV.I Material biológico

Se utilizó la cepa de *E. coli* transformada con el vector pETSumo-M que contiene el gen M codificante de la proteína M del vDEP de la cepa PEDV/MEX/MICH/01/2013 (GenBank: MH006957.1) aislada en Michoacán, México en 2013, donada por el Dr. José Francisco Rivera Benítez del INIFAP. La cepa X-33 de la levadura *Pichia pastoris* (Invitrogen) se empleó para la transformación con los vectores construidos en el presente trabajo para la expresión de la proteína M.

### IV.II Clonación del gen M del vDEP en el vector pPICZαB de *Pichia pastoris*

#### IV.II.I Amplificación de ADN del gen M mediante PCR

La amplificación de ADN se realizó por PCR con el sistema comercial GoTaq G2 Green Master Mix de Promega (Cat. Nos M7822). Se amplificaron dos productos de PCR, la secuencia completa de la proteína M<sup>226</sup> y la región globular de la proteína M<sup>102-212</sup> a partir del marco de lectura abierto del gen M clonado en el plásmido pETSUMO. Los oligonucleótidos utilizados se diseñaron con un sitio de corte para la enzima de restricción EcoRI en el oligonucleótido sentido y con un sitio *Xba*I en el oligonucleótido antisentido. Los oligonucleótidos específicos diseñados son MFecolarga: 5'-CCTGCGCGAATTCTATCTAACGGTTTTATTCCCGTTG-3' y MRxbalarga: 5'-GCCGACGTCTAGAAGGACTAAATGAAGCACTTTCTCAC-3', y MFecorta: 5'-CGTCTTATGAATTCGGTTGTGGCGCAGGACAC-3' y MRxbacorta 5'-:GCACTGCATCTAGAGCATTACTCACAGCTGAGTAGTCGC-3'. El tamaño de los productos de PCR para la secuencia completa y la región globular fue de 707 pb y 358 pb, respectivamente.

#### IV.II.II Gel de agarosa y electroforesis

Para el corrimiento electroforético de los productos de PCR, la purificación de plásmidos y las digestiones enzimáticas se utilizaron geles a una concentración del 0.7% de agarosa. Las condiciones de corrimiento fueron de 80 voltios por 1.5 horas para plásmidos y digestiones enzimáticas y de 80 voltios por 1 h para los productos de PCR. Todas las electroforesis se realizaron utilizando el amortiguador TBE (Tris-borato-EDTA) al 0.5%.

#### IV.II.III Purificación de fragmentos de ADN a partir de gel de agarosa

El ADN de los productos de PCR, plásmidos y digestiones enzimáticas se purificaron a partir del gel de agarosa con el sistema comercial Wizard SV Gel and PCR Clean-Up System de Promega (Cat No. A9280) de acuerdo con las instrucciones del fabricante.

#### IV.II.IV Extracción de ADN plasmídico

La extracción de ADN plasmídico se realizó con el sistema comercial de GeneJet Plasmid Miniprep Kit (Thermo Scientific Cat. No. K0502). 1 µl de la solución de purificación del plásmido se utilizó para cuantificar el plásmido purificado utilizando un equipo Nanodrop.

#### IV.II.V Transformación de *E. coli* DH5 $\alpha$ y selección de cepas transformadas

Las cepas de *E. coli* DH5 $\alpha$  se transformaron mediante CaCl<sub>2</sub> y choque térmico. Las cepas transformadas se cultivaron en medios LB con antibiótico, zeocina para el plásmido pPICZ $\alpha$ B y ampicilina para el plásmido pJET. Las colonias formadas se recuperaron cultivándolas individualmente en los medios LB-zeocina y LB-ampicilina.

#### IV.II.VI Digestiones enzimáticas

Los productos de PCR y el plásmido pPICZ $\alpha$ B se digirieron con las enzimas *EcoRI* y *XbaI*.

#### IV.II.VII Clonación de la secuencia codificante de M en el plásmido pJET

Para la clonación de los fragmentos de la proteína M<sup>2-226</sup> y la región globular de la proteína M<sup>102-212</sup>M se utilizó el sistema comercial de ligación CloneJET PCR Cloning Kit de ThermoFisher (Cat. Nos K1231). El producto amplificado fue clonado en el plásmido pJET1.2/blunt Cloning Vector (CloneJET PCR Cloning Kit) de acuerdo con las instrucciones del fabricante. Posteriormente el plásmido pJET1.2-M se introdujo en *E. coli* DH5 $\alpha$  y cultivó en medio LB-con ampicilina.

#### IV.II.VIII Subclonación de la secuencia codificante de M en el plásmido pPICZ $\alpha$ B

Para la subclonación en el plásmido pPICZ $\alpha$ B se realizó la digestión de los fragmentos de la proteína M<sup>2-226</sup> y la región globular de la proteína M<sup>102-212</sup> a partir del plásmido pJET- M<sup>2-226</sup> y pJET- M<sup>102-212</sup>. La ligación con el plásmido pPICZ $\alpha$ B se realizó con la enzima T4 ligasa. El plásmido pPICZ $\alpha$ B-M se introdujo en *E. coli* DH5 $\alpha$  y cultivó en medio LB-zeocina. El plásmido pPICZ $\alpha$ B-M se secuenció mediante el método de Sanger en la Unidad de secuenciación del LANGEBIO-CINVESTAV. Se utilizaron los mismos oligonucleótidos en la amplificación por PCR para la secuenciación.

#### IV.II.IX Orientación de los fragmentos clonados en pPICZ $\alpha$ B

Para comprobar la orientación de los fragmentos clonados en el plásmido pPICZ $\alpha$ B- M<sup>2-226</sup> y pPICZ $\alpha$ B- M<sup>102-212</sup> se realizaron diferentes combinaciones de digestiones enzimáticas con las enzimas *EcoRI* y *BamHI*.

#### IV.II.X Transformación de los plásmidos en *Pichia pastoris*

Las células competentes de *P. pastoris* X33 se prepararon acorde a las instrucciones de la compañía Invitrogen. Las células se cultivaron en medio YPD como precultivo, se inocularon 50 ml de medio YPD y se incubó a 30°C con agitación a 200 rpm, se recuperaron las células por centrifugación a 1500xg por 10 min a temperatura ambiente, el paquete celular se lavó con 25 ml de agua destilada, se centrifugó y las células se resuspendieron en 1 ml de 100 mM de LiCl, finalmente las células se empaquetaron a velocidad máxima por 15 segundos para remover el LiCl y volver a resuspender en 400  $\mu$ l de LiCl 100 mM. Las células se alicuotaron en tubos con 50  $\mu$ l de células para cada transformación. Para la transformación, se removió el LiCl de cada alícuota por centrifugación y las células se resuspendieron en 240  $\mu$ l de PEG 50%, después se añadieron 36  $\mu$ l de LiCl 1 M, 25  $\mu$ l de ADN de esperma de salmón (2 mg/ml) y el vector pPICZ $\alpha$ B-M<sub>2-226</sub> y pPICZ $\alpha$ B-M<sub>102-112</sub> linearizado (-10  $\mu$ g). la mezcla se incubó a 30°C 30 min sin agitación y posteriormente se realizó un choque térmico a 42°C 25 min. Las células se recuperaron por centrifugación a 6000 rpm 5

min y se resuspendieron en 1 ml de medio YPD, se incubaron a 30°C por 1 h con agitación moderada. Las transformantes fueron seleccionadas en placas de agar YPD con 200 µg/ml de Zeocina después de 3 días de incubación a 30°C.

#### **IV.III Predicción de la estructura 3D de la proteína M**

##### **IV.III.I Secuencias aminoacídicas de la proteína M**

Se utilizó la secuencia de la proteína M del vDEP de la cepa prototipo CV777 con Uniprot ID: P59771 en la búsqueda de las plantillas necesarias para la realización del modelo 3D de la proteína y como secuencia a modelar en los programas de predicción de la estructura 3D y evaluación de los modelos 3D. La secuencia de la proteína M del SARS-CoV-2 de los modelos de AlphaFold y de FeigLab corresponden al Uniprot ID: PODTC5. Diferentes secuencias de la proteína M se recuperaron de la página Taxonomy Browser de NCBI (<https://www.ncbi.nlm.nih.gov/Taxonomy/Browser/wwwtax.cgi>), las secuencias pertenecen a diferentes cepas aisladas en Estados Unidos (37), Europa (105) y Asia (106). El alineamiento de secuencias se realizó en Clustal Omega y se utilizaron para generar en el mismo servidor un árbol filogenético.

##### **IV.III.II Análisis *in silico* de la proteína M**

Para la predicción de las regiones transmembranales se utilizaron los programas, TMPred (<https://bio.tools/tmpred>), MEMSAT (<http://www.sacs.ucsf.edu/cgi-bin/memsat.py>), HMMTOP (<http://www.enzim.hu/hmmtop/>), TMHMM (<https://services.healthtech.dtu.dk/service.php?TMHMM-2.0>), Philius (<https://bio.tools/philius>) y Phobius (<https://phobius.sbc.su.se/>).

##### **IV.III.III Programas para la búsqueda de plantillas para la generación de modelos 3D**

Para la búsqueda de plantillas se realizó un análisis en BLAST (<https://blast.ncbi.nlm.nih.gov/Blast.cgi>) con los algoritmos BLASTP y PSI-BLAST, PSI-BLAST, que permiten ampliar la búsqueda de plantillas necesarias para la construcción del modelo 3D (107). Adicionalmente, se utilizó el servidor web SWISS MODEL (<https://swissmodel.expasy.org/>) para la búsqueda de plantillas y la generación de un modelo 3D de la proteína M del vDEP de la cepa CV777. SWISS MODEL funciona a través de la búsqueda de estructuras proteicas relacionadas evolutivamente que funcionan como plantillas para la generación del modelo, para cada plantilla un modelo 3D es generado automáticamente mediante la transferencia de las coordenadas de los átomos conservados definido por el alineamiento con la plantilla, las coordenadas correspondientes a inserciones o eliminaciones en el alineamiento son generados con el modelamiento de bucles, finalmente a cada modelo se le asigna un valor de calidad mediante QMEAN (108).

##### **IV.III.IV Programas para la generación de modelos 3D de proteínas**

Para la generación del modelo 3D de la proteína M del vDEP, se utilizaron los servidores en línea SWISS MODEL, SPARKS X, RAPTOR X, intFOLD5 y Robetta, todos formando parte del proyecto CAMEO, SWISS MODEL construye un modelo de proteína atomista dada una estructura de plantilla y una alineación de secuencia objetivo-plantilla, las inserciones y eliminaciones, definido por el alineamiento, se resuelven buscando fragmentos de candidatos viables en una base de datos estructurales, si no se pueden encontrar

fragmentos adecuados, se realiza una búsqueda de espacio conformacional utilizando un muestreo de Monte Carlo y finalmente se realiza una minimización de energía para mejorar la calidad de la estereoquímica del modelo (108). SPARKS X se basa en métodos de reconocimiento de un solo pliegue (SPARKS, SP2, SP3, SP4 y SP5) dentro de estos pliegues se basa la predicción de estructuras secundarias predichas vs. secundarias reales, área superficial accesible al solvente predicha vs. real y ángulos diedros predichos vs. reales, esta combinación de perfiles genera un modelo 3D (109). Phyre 2 genera modelos de proteínas a través de una combinación de modelamiento de múltiples plantillas y simulaciones de plegamiento *ab initio*. La alineación de secuencia múltiple se realiza median HHblits, el alineamiento múltiple resultante se utiliza para predecir la estructura secundaria con PSIPRED (110) y, tanto el alineamiento como la predicción de la estructura secundaria, se combinan en un modelo de Markov oculto (HMM) (111) y posteriormente se compara contra una base de datos de HMM de proteínas de estructura conocida. Los alineamientos con la puntuación más alta se utilizaron para construir modelos crudos únicamente de cadena principal, los *indels* (inserciones y deleciones) en estos modelos se corrigen mediante el modelado de asas y finalmente se agregan las cadenas laterales de aminoácidos para generar el modelo Phyre2 final. Las regiones no cubiertas por las plantillas se construyen por el método *ab initio* denominado Poing (112). RAPTOR X toma en consideración el número de homólogos no redundantes disponibles para la secuencia a modelar, utiliza campos neuronales condicionales (CNF), para integrar una variedad de características biológicas específicas en una función de puntuación probabilística, finalmente, implementa un procedimiento de enhebrado de múltiples plantillas (MTT) que permite modelar una única secuencia utilizando la distribución de probabilidad de orientación y distancia entre residuos/átomos para reconstruir modelos de proteínas en 3D (113). intFOLD5, este servidor web utiliza diferentes métodos de alineamiento múltiple en búsqueda de plantillas para el modelamiento, corrige los errores globales y locales por residuo mediante un modelamiento iterativo, los modelos se construyen mediante I-TASSER (114) y HHpred (115). Por último, el servidor Robetta, mediante el modelado comparativo de RosettaCM, se pueden obtener modelos 3D, RosettaCM busca las plantillas mediante PSI-BLAST y Hhsearch, posteriormente la secuencia de interés se enhebrada en cada una de las plantillas y los modelos parciales enhebrados resultantes son alineados en un único alineamiento global. Posteriormente, el modelado de la cadena completa se genera mediante una trayectoria larga de Monte Carlo empezando a partir de una plantilla seleccionada aleatoriamente, la trayectoria implementa la inserción de fragmentos en regiones desalineadas, el remplazamiento de segmentos elegidos aleatoriamente con estructuras de plantillas diferentes y una minimización del espacio cartesiano usando una versión más noble que la versión de Rosetta, la función de energía de centroide, finalmente, se realiza la optimización de todos los átomos, complementado con restricciones de distancia de las estructuras de la plantilla y una penalización por separación en el espacio de residuos adyacentes en la secuencia (116).

#### IV.III.V Programas para la evaluación de modelos 3D de proteínas

Se utilizaron los servidores en línea de QMEAN, QMEANDisCo, QMEANBrane, PROCHECK, ModFOLD6, ProSA-web y MolProbity para evaluar la calidad de los modelos generados con los servidores mencionados anteriormente. QMEAN permite evaluar diferentes propiedades geométricas y proporciona un valor que propone un nivel de calidad tanto global como local o por residuo sobre un modelo. El puntaje Z de QMEAN es un valor que puede interpretarse como el "grado de natividad" de las características estructurales

presentes en el modelo, comparable con valores de estructuras experimentales de tamaño similar. Las puntuaciones Z de QMEAN cercanas a cero indican una buena relación entre la estructura del modelo y las estructuras experimentales de tamaño similar, las puntuaciones de -4 a 0, o por debajo de -4, son una indicación de modelos con baja calidad. QMEANDisCo permite el mejoramiento en la precisión de las puntuaciones locales de QMEAN, evaluando la consistencia de las distancias interatómicas del modelo evaluado con información extraída de estructuras de proteínas determinadas experimentalmente que son homólogas a la secuencia de interés, los residuos o modelos que presentan una puntuación inferior a 0.6 son de baja calidad. QMEANBrane, utiliza una combinación de potenciales estadísticos enfocados en la evaluación de la calidad local de los modelos de proteínas de membrana en su estado oligomérico natural, se identifica la región transmembranal utilizando un modelo de solvatación y la similitud o diferencia entre un modelo y una estructura de referencia se puede expresar en forma de medidas sin superposición, como la puntuación de la prueba de diferencia de distancia local (IDDT) (117). PROCHECK, con este servidor se puede obtener una estimación de la calidad estereoquímica de las cadenas de proteínas en un modelo 3D, el análisis proporciona las distribuciones observadas de los valores phi-psi, chi1-chi2, chi-1, chi-3, chi-4 y omega para cada uno de los 20 tipos de aminoácidos y se comparan con estructuras de proteínas obtenidas experimentalmente (118). ModFOLD, programa que se encuentra dentro del proyecto CAMEO, utiliza diferentes métodos para realizar la evaluación de los modelos 3D de proteínas, entre ellos figuran la evaluación de los contactos entre residuos, la evaluación de la predicción de estructuras secundarias y residuos en regiones desordenadas y compara de cada modelo con un conjunto de referencia de 130 modelos construidos por IntFOLD, programa parte de su mismo servidor (119). Otro servidor web utilizado para la evaluación de la calidad de los modelos 3D de proteínas es ProSA-web, el cual se basa en el uso de los potenciales basados en el conocimiento de la fuerza media para evaluar la precisión del modelo, las estructuras tridimensionales de proteínas contienen una enorme cantidad de información sobre las fuerzas que estabilizan los pliegues nativos en solución, estas fuerzas se extraen de una base de datos de estructuras conocidas en forma de potenciales de fuerza media y el campo de fuerza para una proteína particular de estructura conocida o desconocida se obtiene mediante una recombinación de estos potenciales en función de la secuencia de aminoácidos, aplicando este enfoque al plegamiento de proteínas, utilizando el campo de fuerza se puede distinguir los pliegues nativos de regiones mal plegadas (120), esta característica permite el reconocimiento de estructuras incorrectas, dejando a un lado las violaciones de los principios estéricos básicos, ProSA web se enfoca en la disposición correcta de las cadenas de una proteína en tres dimensiones (121). Finalmente, MolProbity realiza una evaluación principalmente sobre la estereoquímica del modelo 3D de proteínas, analiza el nivel de choque o colisiones entre átomos que se superponen por más de 0.4 Å por cada mil átomos, adicionalmente, realiza un análisis de los rotámeros en el modelo y los compara con una base de datos permitiendo definir áreas validas o permitidas de rotámeros similar a los cuadros de Ramachandran, MolProbity también considera los ángulos phi y psi y los coloca en una distribución con los criterios de Ramachandran comparándolos con una base de datos y calculando el porcentaje de ángulos en zonas favorables y permitidas, esta combinación de criterios, colisión de átomos, rotámeros y criterios de Ramachandran permiten generar un valor de calidad global para el modelo, MolProbity score (122).

#### **IV.III.VI Modelos de la proteína M del SARS-CoV-2**

Se utilizaron los modelos 3D de la proteína M del SARS-CoV-2 realizados por FeigLab y AlphaFold para comparar las características de estos modelos con el modelo de la proteína M del vDEP realizado con Robetta. AlphaFold, es parte de un grupo más grande de investigación en inteligencia artificial llamado DeepMind, enfocados en solucionar problemas utilizando esta nueva herramienta, y debido al interés sobre el plegamiento de las proteínas en su estructura terciaria, se han enfocado a predecir estas estructuras, debido a la pandemia global del SARS-CoV-2, realizaron mediante AlphaFold la predicción de 5 estructuras del virus, la proteína de membrana, Nsp2, Nsp4, Nsp6 y el dominio C terminal de la proteinasa similar a la papaína, proteínas sin modelos o estructuras similares disponibles en bases de datos necesarias para realizar el modelamiento (<https://deepmind.com/research/open-source/computational-predictions-of-protein-structures-associated-with-COVID-19>). AlphaFold utiliza un alineamiento múltiple de secuencia para generar fragmentos del modelo mediante modelado basado en plantillas y lo combina con un modelado libre de plantillas para las regiones sin alguna plantilla adecuada, mediante una red neuronal profunda se predicen la distribución de distancias y torsiones y finalmente se realiza un gradiente descendente de potenciales (123). FeigLab, una plataforma web perteneciente a la Universidad de Michigan, dirigido por Michael Feig, se dedica al estudio de la biología desde escalas moleculares hasta celulares con el objetivo de unir la biología de estructuras y sistemas utilizando una variedad de métodos computacionales <https://feig.bch.msu.edu/>. Este grupo de investigación realizó el modelamiento de 10 proteínas del SARS-CoV-2, incluyendo a la proteína de membrana. Las proteínas fueron modeladas mediante el uso de machine-learning, utilizándola en la predicción de la distancia y orientaciones entre residuos, y posteriormente realizaron un refinamiento basado en simulaciones de dinámica molecular (124).

#### **IV.III.VII Superposición de modelos 3D de proteínas**

Para la superposición de modelos 3D se utilizó el programa USCF Chimera, el cual también se utilizó para obtener el valor de RMSD.

### **IV. IV Identificación de epítomos en la proteína M**

#### **IV.IV.I Identificación de epítomos continuos de células B**

Para la identificación de los epítomos continuos en la secuencia aminoacídica de la proteína M del vDEP se utilizaron diferentes herramientas pertenecientes al dominio de Immune Data Base (<https://www.iedb.org/>), se utilizó la herramienta Bepipred Linear Epitope Prediction 1.0 y 2.0 (<http://tools.iedb.org/bcell/>), Bepipred 2.0 predice epítomos de células B a partir de una secuencia de aminoácidos, utilizan el algoritmo Random Forest entrenado en epítomos y aminoácidos no epítomos determinados a partir de estructuras cristalinas (125), con una línea de corte de 0.5 (sensibilidad = 0.58564, especificidad = 0.57158). La herramienta Bepipred-1.0 Linear Epitope Prediction predice la ubicación de los epítomos lineales de células B mediante una combinación de un modelo de Markov oculto y un método de escala de propensión (126), con una línea de corte de 0.35 (sensibilidad = 0.49, especificidad = 0.75). Los epítomos predichos se evalúan mediante Chou & Fasman Beta-Turn Prediction, Emini Surface Accessibility

Prediction, Kolaskar and Tongaonkar Antigenicity, y Parker Hydrophilicity Prediction, usando una línea de corte de 1.0 adicionalmente. Para el análisis de conservación se utilizó la herramienta Epitope Conservancy Analysis de IEDB (<http://tools.iedb.org/conservancy/>).

#### **IV.IV.II Identificación de epítomos discontinuos de células B**

Se utilizaron los servidores en línea DiscoTope y ElliPro pertenecientes al dominio de Immune Data Base. ElliPro con un puntaje mínimo de 0.5 y una distancia máxima de 6 Å. Para DiscoTope se utilizó la versión 1.1 con una línea de corte de -7.7 (sensibilidad = 0.47, especificidad = 0.75).

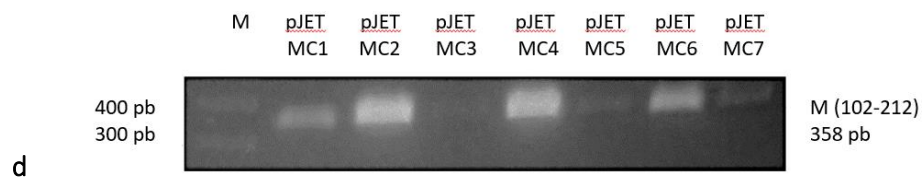
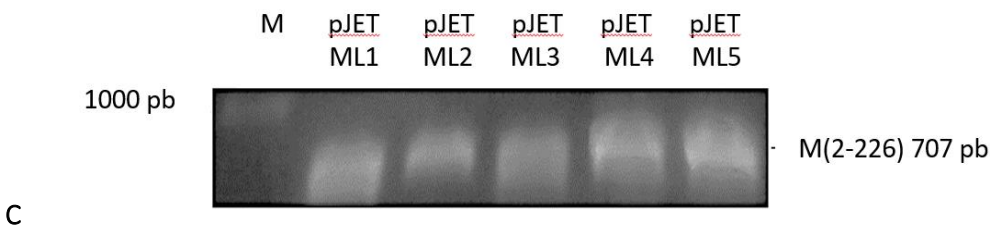
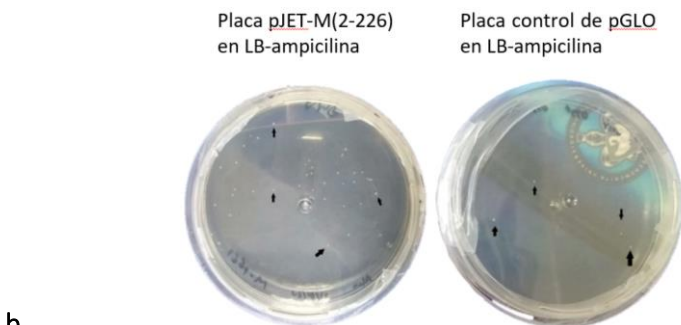
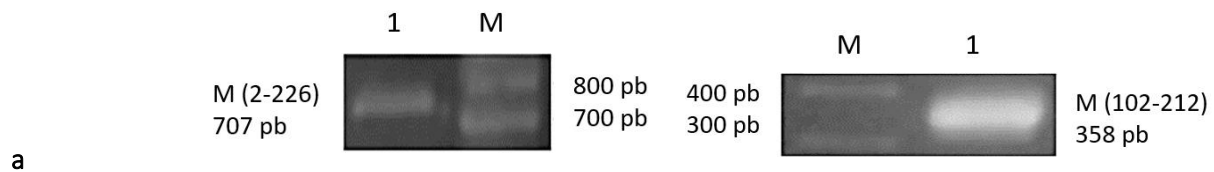
#### **IV.IV.III Identificación de epítomos de células T**

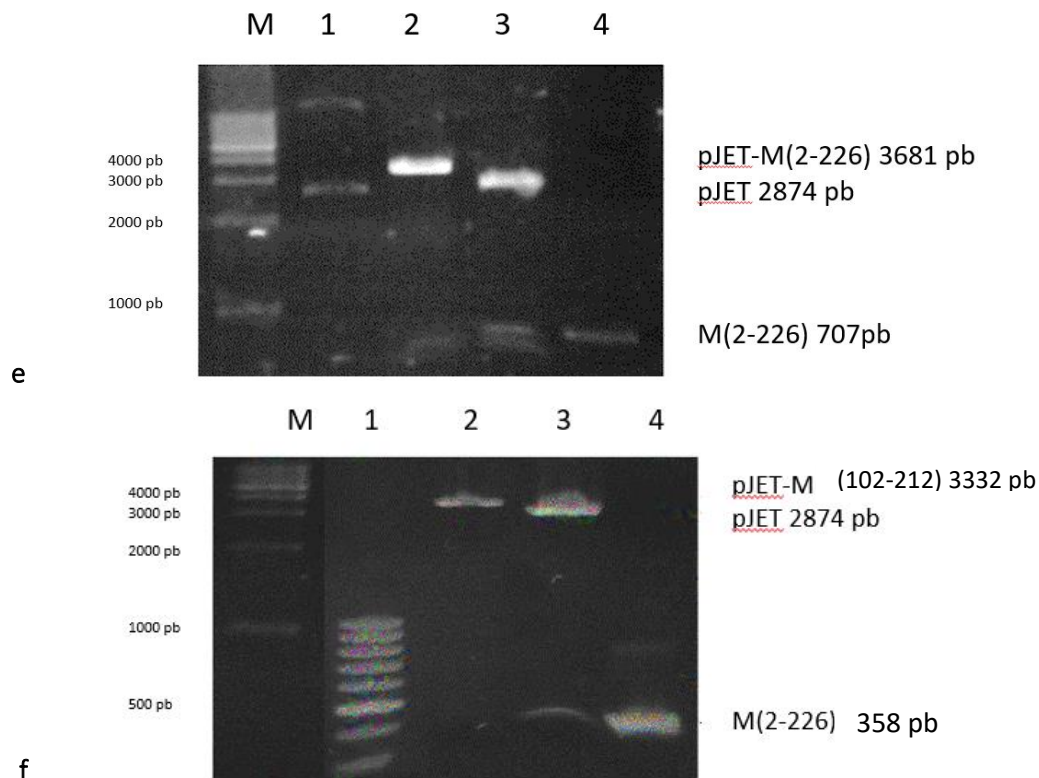
Los epítomos de células T se identificaron con el servidor NetMHCpan 4.1, el cual realiza la predicción de péptidos que se unen a cualquier molécula de MHC mediante el uso de redes neuronales, utiliza más de 850000 datos cuantitativos de afinidad de unión de péptidos y datos de péptidos de ligandos eluidos (EL) por espectrometría de masas (127). Se utilizaron los alelos leucocitarios porcinos SLA-1:0101; SLA-1:0401; SLA-1:0801, los cuales están ampliamente distribuidos entre la población porcina (128). La búsqueda de péptidos se estableció en noámeros para todos los epítomos, usando 0.5% como línea de corte para una unión fuerte y 2% para uniones débiles. Para el estudio de simulación de acoplamiento se utiliza la estructura cristalizada de la molécula SLA-1:0401 (PDB ID 3QQ3) (129). El epítomo de la influenza, el cual forma un complejo en el surco de unión de SLA-1:0401, fue removido con AutoDock Tools. La simulación de acoplamiento se realizó mediante AutoDock Vina (130).

## V. Resultados

### V.I Clonación del gen M del vDEP en el vector pJET1.2/blunt

Para cumplir con el primer objetivo del trabajo, se amplificaron los fragmentos  $M^{2-226}$  y  $M^{102-212}$  a partir del plásmido pETSUMO-M de 1065 pb (678 pb del gen M (**ilustración 2a**)). Los productos de PCR para los insertos  $M^{2-226}$  y  $M^{102-212}$  se purificaron e insertaron en el vector pJET1.2/blunt. Las construcciones pJET- $M^{2-226}$  y pJET- $M^{102-212}$  se transformaron en la cepa *E. coli* DH5 $\alpha$  y se cultivaron en placas de LB-ampicilina (**ilustración 2b**). Las clonas positivas se identificaron por PCR en colonia (**ilustración 2c**). A partir de dos clonas positivas se purificaron los plásmidos y se confirmó la presencia de los fragmentos  $M^{2-226}$  y  $M^{102-212}$  (**ilustración 2d y 2e**) mediante digestión con las enzimas *EcoRI* y *XbaI*. Los fragmentos esperados de 685 pb y 340 pb, correspondientes a  $M^{2-226}$  y  $M^{102-212}$  se purificaron para la subclonación en el plásmido pPICZ $\alpha$ B entre los sitios de restricción *EcoRI* y *XbaI*.

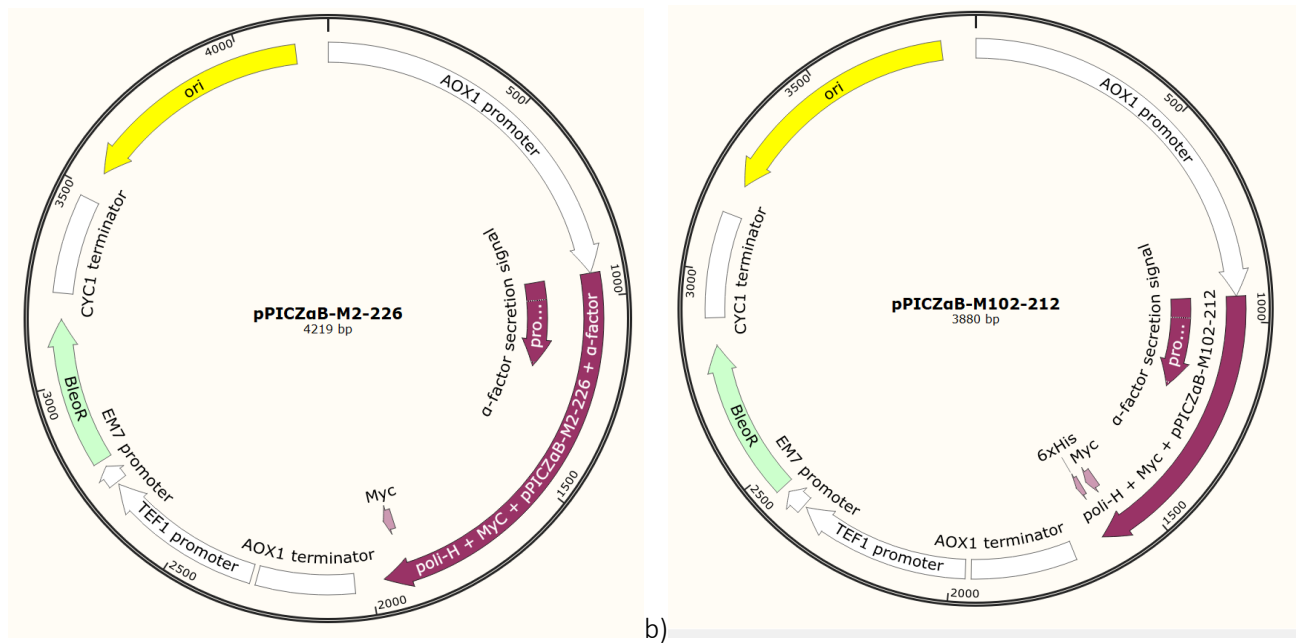




**Ilustración 2. Clonación de los fragmentos  $M^{2-226}$  y  $M^{102-212}$  en el plásmido pJET1.2/blunt.** a) Del lado izquierdo, amplificación mediante PCR del producto  $M^{2-226}$ , carril 1: producto de PCR del fragmento  $M^{2-226}$  de un tamaño de 707 pb. Carril M: marcador de peso molecular de 100 pb. Del lado derecho, amplificación mediante PCR del producto  $M^{102-212}$ , carril M, marcador de peso molecular de 100 pb. Carril 1: producto de PCR del fragmento  $M^{102-212}$  de un tamaño de 358 pb. b) Del lado izquierdo, placa de las colonias transformadas de *E. coli* DH5 $\alpha$  con el plásmido pJET-M<sup>2-226</sup> cultivadas en medio LB-ampicilina. Del lado derecho, placa de las colonias transformadas de *E. coli* DH5 $\alpha$  con el plásmido pGLO utilizado como control positivo cultivadas en medio LB-ampicilina. c y d) En la parte superior, amplificación mediante PCR en colonia del producto  $M^{2-226}$  a partir de las cepas de *E. coli* DH5 $\alpha$  transformadas con el plásmido pJET-M<sup>2-226</sup>, carril M: marcador de peso molecular de 1 kb, carril 1 al 5, cepas pJET-ML1-5. En la parte inferior, amplificación mediante PCR en colonia del producto  $M^{102-212}$  a partir de las cepas de *E. coli* DH5 $\alpha$  transformadas con el plásmido pJET-M<sup>102-212</sup>, carril M: marcador de peso molecular de 100 pb, carril 1 al 7, cepas pJET-MC1-7. e) Digestiones del plásmido pJET-M<sup>2-226</sup>. Carril M: marcador de peso molecular de 100 pb. Carril 1: plásmido pJET-M<sup>2-226</sup>, carril 2: plásmido pJET-M<sup>2-226</sup> linealizado con EcoRI de un tamaño de 3681 pb, carril 3: plásmido digerido con EcoRI y XbaI da un producto de 2974 pb correspondiente al plásmido pJET1.2/blunt vacío más un producto de 685 pb correspondiente al fragmento  $M^{2-226}$ , carril 4: producto de PCR correspondiente al fragmento  $M^{2-226}$  de 707 pb. f) Digestiones del plásmido pJET-M<sup>102-212</sup>. Carril M1: marcador de peso molecular de 1 kb, carril M2: marcador de peso molecular de 100 pb, carril 1: plásmido pJET-M<sup>102-212</sup> de un tamaño de 3332 pb, carril 2: plásmido digerido con EcoRI y XbaI da un producto de 2974 pb correspondientes al plásmido pJET vacío y un producto de 340 pb correspondientes al fragmento  $M^{102-212}$ , carril 3: producto de PCR correspondiente al fragmento  $M^{102-212}$  de 358 pb.

## V.II Clonación del gen M del vDEP en el vector pPICZαB de *Pichia pastoris*

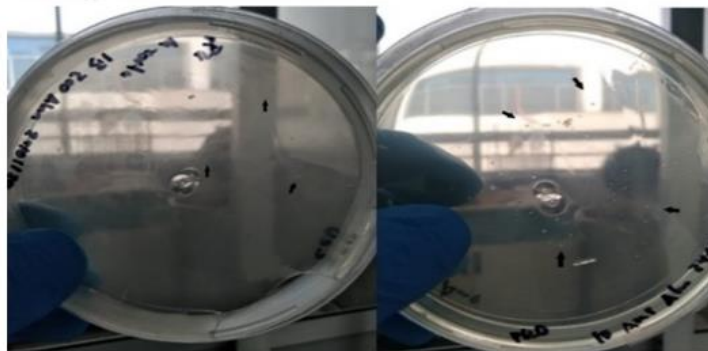
Las construcciones pPICZαB-M<sup>2-226</sup> y pPICZαB-M<sup>102-212</sup> (ilustración 3) se transformaron en la cepa *E. coli* DH5α y se cultivaron en placas de LB-zeocina (ilustración 4a y 4b). Las clonas positivas a los insertos M<sup>2-226</sup> y M<sup>102-212</sup> fueron identificadas por PCR en colonia (ilustración 4b). El ADN plasmídico de las colonias positivas seleccionadas se purificó y digirió con las enzimas EcoRI, XbaI y BamHI para confirmar la presencia de los fragmentos y su orientación en el plásmido (ilustración 4c y 4d). Con los datos expuestos se confirmó la clonación exitosa de los fragmentos M<sup>2-226</sup> y M<sup>102-212</sup> para expresar a las versiones completa y corta de las proteínas M.



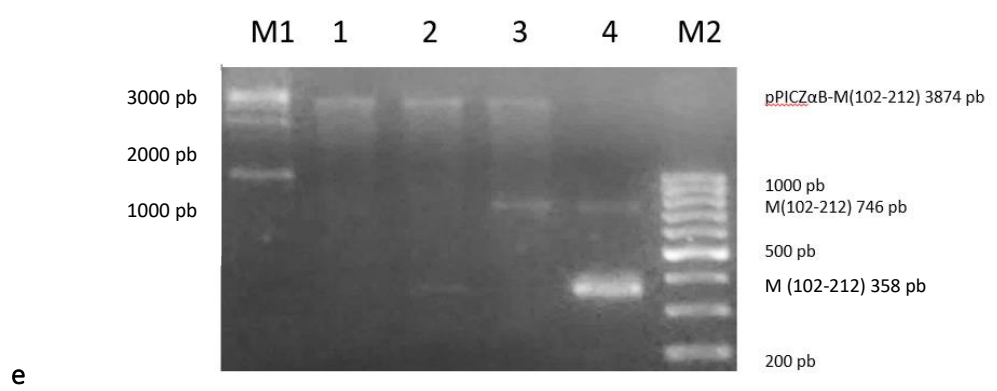
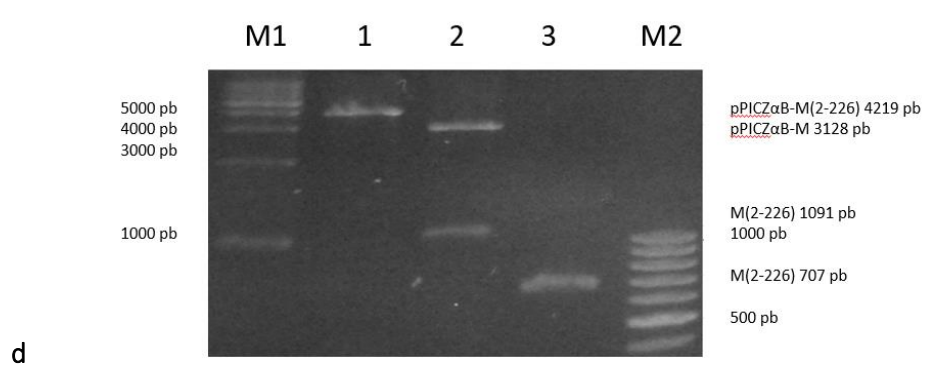
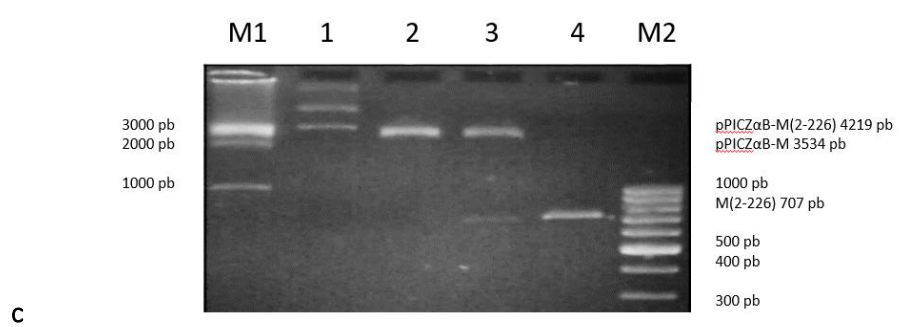
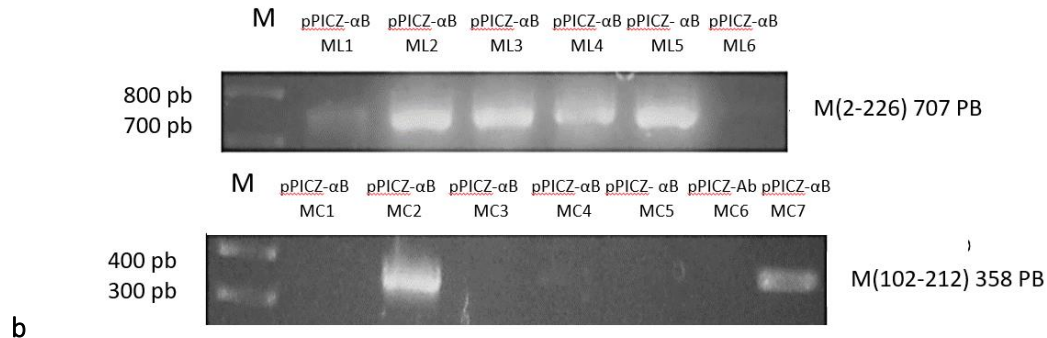
a) Ilustración 3. Mapa de los plásmidos construidos a) Plásmido pPICZαB-M<sub>2-226</sub> y b) pPICZαB-M<sub>102-212</sub>.

Placa de pPICZ-αB-M (2-226) en zeocina

Placa control de pGLO en LB-ampicilina



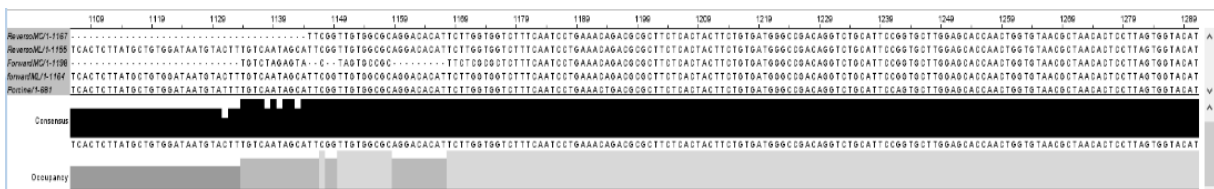
a



**Ilustración 4. Clonación de los fragmentos M<sup>2-226</sup> y M<sup>102-212</sup> en el plásmido pPICZαB.** a) Del lado izquierdo, placa de las colonias transformadas de *E. coli* DH5α con el plásmido pPICZαB-M<sup>2-226</sup> cultivadas en medio LB-zeocina. Del lado derecho, placa de las colonias transformadas de *E. coli* DH5α con el plásmido pGLO utilizado como control positivo cultivadas en medio LB-ampicilina. b) En la parte superior, amplificación mediante PCR en colonia del producto M<sup>2-226</sup> a partir de las cepas de *E. coli* DH5α transformadas con el plásmido pPICZαB-M<sup>2-226</sup>, carril M: marcador de peso molecular de 100 pb, carril 1 al 6, cepas pPICZαB-ML1-6. En la parte inferior, amplificación mediante PCR en colonia del producto M<sup>102-212</sup> a partir de las cepas de *E. coli* DH5α transformadas con el plásmido pPICZαB-M<sup>102-212</sup>, carril M: marcador de peso molecular de 100 pb, carril 1 al 7, cepas pPICZαB-MC1-7. c) Digestiones del plásmido pPICZαB-M<sup>2-226</sup>. Carril M: marcador de peso molecular de 1 kb. Carril 1: plásmido pPICZαB-M<sup>2-226</sup>, carril 2: plásmido pPICZαB-M<sup>2-226</sup> linealizado con EcoRI de un tamaño de 4219 pb, carril 3: plásmido digerido con EcoRI y XbaI da un producto de 3534 pb correspondiente al plásmido pPICZαB vacío más un producto de 685 pb correspondiente al fragmento M<sup>2-226</sup>, carril 4: producto de PCR correspondiente al fragmento M<sup>2-226</sup> de 707 pb. d) Digestiones del plásmido pPICZαB-M<sup>2-226</sup>. Carril M1: marcador de peso molecular de 1 kb, carril 1: plásmido pPICZαB-M<sup>2-226</sup> de un tamaño de 4219 pb, carril 2: plásmido digerido con EcoRI y BamHI da un producto de 3128 pb correspondientes al plásmido pPICZαB vacío y un producto de 1091 pb correspondientes al fragmento M<sup>102-212</sup>, carril 3: producto de PCR correspondiente al fragmento M<sup>2-226</sup> de 707 pb, carril M2: marcador de peso molecular de 100 pb. e) Digestiones del plásmido pPICZαB-M<sup>102-212</sup>. Carril M1: marcador de peso molecular de 1 kb, carril 1: plásmido pPICZαB-M<sup>102-212</sup> de un tamaño de 3874 pb, carril 2: plásmido digerido con EcoRI y XbaI da un producto de 3534 pb correspondientes al plásmido pPICZαB vacío y un producto de 340 pb correspondientes al fragmento M<sup>102-212</sup>, carril 3: plásmido digerido con EcoRI y BamHI da un producto de 3128 pb correspondientes al plásmido pPICZαB vacío y un producto de 746 pb correspondientes al fragmento M<sup>102-212</sup>, carril 4: producto de PCR correspondiente al fragmento M<sup>102-212</sup> de 358 pb, carril M2: marcador de peso molecular de 100 pb.

### V.III Secuenciación de los plásmidos pPICZαB-M<sup>2-226</sup> y pPICZαB-M<sup>102-212</sup>

Se realizó un análisis en Clustal Omega de los 4 resultados de secuenciación para obtener la secuencia consenso de los fragmentos M<sup>2-226</sup> y M<sup>102-212</sup> en los plásmidos pPICZαB-M<sup>2-226</sup> y pPICZαB-M<sup>102-212</sup> (ilustración 5).



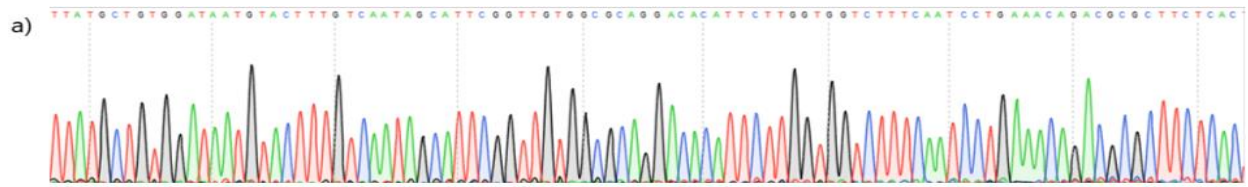
**Ilustración 5. Análisis en clustal omega de los resultados de secuenciación.** Se hizo el alineamiento de las 4 secuencias obtenidas del proceso de secuenciación. “ReversoMC” secuencia obtenida del plásmido pPICZαB-M<sup>102-212</sup> con el oligonucleótido antisentido. “ForwardMC” secuencia obtenida del plásmido pPICZαB-M<sup>102-212</sup> con el oligonucleótido sentido. “Forward ML” secuencia obtenida del plásmido pPICZαB-M<sup>2-226</sup> con el oligonucleótido sentido. “ReversoML” secuencia obtenida del plásmido pPICZαB-M<sup>2-226</sup> con el oligonucleótido antisentido.

Utilizando el programa Jalview se obtuvo la secuencia consenso para los fragmentos M<sup>2-226</sup> y M<sup>102-212</sup> (ilustración 6).

ReversoMC	GGAA-----	844	ReversoMC	GCTACGGACGTGTTGGTCGTTCAAGTCAATGCTTCATCTGGCACTGGTTGGGCTTCTAT	1140
ForwardMC	-----	0	ForwardMC	GCTACGGACGTGTTGGTCGTTCAAGTCAATGCTTCATCTGGCACTGGTTGGGCTTCTAT	294
ForwardML	-----CCCAAAAGTAGGGTATTGAACACCTTAGAAATCGG	37	ForwardML	GCTACGGACGTGTTGGTCGTTCAAGTCAATGCTTCATCTGGCACTGGTTGGGCTTCTAT	577
ReversoML	GGAAATCTATCTAAAGGTTTATCCCGTGTGAGGGTATTGAACACCTTAGAAATCGG	547	ReversoML	GCTACGGACGTGTTGGTCGTTCAAGTCAATGCTTCATCTGGCACTGGTTGGGCTTCTAT	1087
ReversoMC	-----	844	ReversoMC	T--CGTCCAAACACAGCAAAACCCCC--G-----	1167
ForwardMC	-----	0	ForwardMC	GTCGGTCCAAACACGGCGACTACTAGCTGTGAGTAATGC-----	335
ForwardML	AAATTCACATGGAATATCATCTACGACACTACTCTGTAGTGCCTTCAATATGGCAATTAC	97	ForwardML	GTCGGTCCAAACACGGCGACTACTAGCTGTGAGTAATGCAGTTCTGCAGAT	637
ReversoML	AAATTCACATGGAATATCATCTACGACACTACTCTGTAGTGCCTTCAATATGGCAATTAC	607	ReversoML	GTCGGTCCAAACACGGCGACTACTAGCTGTGAGTAATCCGAGCGGT--CTCACAGAT	1145
ReversoMC	-----	844	ReversoMC	-----TCTAGAACAAAACTCATCTAGAAAGGATCTG	1167
ForwardMC	-----	0	ForwardMC	AGTGAAGGATGCTTCATTAGTCCCTTAGAACAAAACTCATCTAGAAAGGATCTG	369
ForwardML	AAATTCACATGGAATATCATCTACGACACTACTCTGTAGTGCCTTCAATATGGCAATTAC	157	ForwardML	AGTGAAGGATGCTTCATTAGTCCCTTAGAACAAAACTCATCTAGAAAGGATCTG	697
ReversoML	AAATTCACATGGAATATCATCTACGACACTACTCTGTAGTGCCTTCAATATGGCAATTAC	667	ReversoML	ACTAATCAAT-----	1155
ReversoMC	-----	844	ReversoMC	-----	1167
ForwardMC	-----	0	ForwardMC	AAATAGCGCCGTGACCATCATCATCATCATTGAGTTGTAGCCTTAGACATGACTGT	429
ForwardML	GTGTTAGCACTGCTCACTTTTGTAGCATGGGCTAGCTTTCAGGTCAATTGGGCTTTTTT	217	ForwardML	AAATAGCGCCGTGACCATCATCATCATCATTGAGTTGTAGCCTTAGACATGACTGT	757
ReversoML	GTGTTAGCACTGCTCACTTTTGTAGCATGGGCTAGCTTTCAGGTCAATTGGGCTTTTTT	727	ReversoML	-----	1155
ReversoMC	-----	844	ReversoMC	-----	1167
ForwardMC	-----TGCTAG	7	ForwardMC	TCCTCAGTTCAAGTTGGGCACTTACGGAAGACCGGCTTGTCTAGATTCTAATCAAGAGG	489
ForwardML	GCCTTCAGCATCCTTAGGCTTGCATCACTCTTATGCTGTGGATAATGTACTTTGTCAAT	277	ForwardML	TCCTCAGTTCAAGTTGGGCACTTACGGAAGACCGGCTTGTCTAGATTCTAATCAAGAGG	817
ReversoML	GCCTTCAGCATCCTTAGGCTTGCATCACTCTTATGCTGTGGATAATGTACTTTGTCAAT	787	ReversoML	-----	1155
ReversoMC	---TTGCGTGTGGCCAGGACACTTCTGGTGGCTTTCACTCTGAAACAGACGCG	900	ReversoMC	-----	1167
ForwardMC	AGTA--C--TAGTGGCCG-----TTCTCGCGCTTTCACTCTGAAACAGACGCG	54	ForwardMC	ATGTCAGAAATGCCATTTGGCTGAGAGATGCAAGGCTTCAATTTTGTACTTTTTTATTGT	549
ForwardML	AGCATTGGTGTGGCCAGGACACTTCTGGTGGCTTTCACTCTGAAACAGACGCG	337	ForwardML	ATGTCAGAAATGCCATTTGGCTGAGAGATGCA--GCTTCAATTTTGTACTTTTTTATTGT	876
ReversoML	AGCATTGGTGTGGCCAGGACACTTCTGGTGGCTTTCACTCTGAAACAGACGCG	847	ReversoML	-----	1155
ReversoMC	CTTCTCACTACTTCTGTGATGGGCCGACAGGCTGCAATTCGGGTCTTGGAGCACCAACT	960	ReversoMC	-----	1167
ForwardMC	CTTCTCACTACTTCTGTGATGGGCCGACAGGCTGCAATTCGGGTCTTGGAGCACCAACT	114	ForwardMC	AACTATATAGTATAGGATTTTTTGTCTTTTGTCTTTTCTTCTGACGAGCTTGTCTCT	689
ForwardML	CTTCTCACTACTTCTGTGATGGGCCGACAGGCTGCAATTCGGGTCTTGGAGCACCAACT	397	ForwardML	AACTATATAGTATAGGATTTTTTGTCTTTTGTCTTTTCTTCTGACGAGCTTGTCTCT	936
ReversoML	CTTCTCACTACTTCTGTGATGGGCCGACAGGCTGCAATTCGGGTCTTGGAGCACCAACT	907	ReversoML	-----	1155
ReversoMC	GGGTAAACGCTAACACTCCCTTAGTGGTACATTCCTGTAGAGGGCTAAGGTTGCTACT	1020	ReversoMC	-----	1167
ForwardMC	GGGTAAACGCTAACACTCCCTTAGTGGTACATTCCTGTAGAGGGCTAAGGTTGCTACT	174	ForwardMC	GATCAGCCTATCTCGAGCTGATGAATATCTTGTGGTAG--GGTTTGGGAAAATATTG	668
ForwardML	GGGTAAACGCTAACACTCCCTTAGTGGTACATTCCTGTAGAGGGCTAAGGTTGCTACT	457	ForwardML	GATCAGCCTATCTCGAGCTGATGAATATCTTGTGGTAG--GGTTTGGGAAAATATTG	995
ReversoML	GGGTAAACGCTAACACTCCCTTAGTGGTACATTCCTGTAGAGGGCTAAGGTTGCTACT	967	ReversoML	GATCAGCCTATCTCGAGCTGATGAATATCTTGTGGTAG--GGTTTGGGAAAATATTG	1155
ReversoMC	GGCGTACAGGTAAGTCAATTACCTAATTTGCTCACAGTGGCCAAAGCCACTACAACAATT	1080	ReversoMC	-----	1167
ForwardMC	GGCGTACAGGTAAGTCAATTACCTAATTTGCTCACAGTGGCCAAAGCCACTACAACAATT	234	ForwardMC	AGTTTGTGTTTT--TCTGGTATTTCCCACTCTCTCAGAGTACAGAGAAATAGTGA	727
ForwardML	GGCGTACAGGTAAGTCAATTACCTAATTTGCTCACAGTGGCCAAAGCCACTACAACAATT	517	ForwardML	AGTTTGTGTTTTTCTGGTATACCTCC--C--TCTCAGAGTACAGAGATA--AGTAG	1051
ReversoML	GGCGTACAGGTAAGTCAATTACCTAATTTGCTCACAGTGGCCAAAGCCACTACAACAATT	1027	ReversoML	-----	1155

**Ilustración 6. Análisis de secuencia en Jalview.** El análisis de la secuencia confirmó la presencia de los fragmentos M<sup>2-226</sup> y M<sup>102-212</sup> así como su correcta orientación.

Un fragmento del electroferograma de los resultados de secuenciación para el plásmido pPICZαB-M<sup>2-226</sup> se muestra en la **ilustración 7a**. El análisis comparativo de la secuencia obtenida en la plataforma BLAST demuestra un 98% de identidad con la secuencia nucleotídica de la proteína M de la cepa CV777 del vDEP (**ilustración 7b**) e interesantemente un 99% con una cepa de vDEP mexicana en Puebla (**ilustración 7c**) y otras cepas mexicanas (dato no mostrado).



**b) NC\_003436.1:25682-26362 Porcine epidemic diarrhea virus, complete genome**  
Sequence ID: Query\_17293 Length: 681 Number of Matches: 1

Range 1: 25 to 678 [Graphics](#) [Next Match](#) [Previous Match](#)

Score	Expect	Identities	Gaps	Strand
1142 bits(618)	0.0	642/654(98%)	0/654(0%)	Plus/Plus
Query 8	GATGAGGTGATTGAACACCTTAGA	67		
Sbjct 25	GATGAGGTGATTGAACACCTTAGA	84		
Query 68	CTACTTGTAGTCTTCAGTATGCC	127		
Sbjct 85	CTACTTGTAGTCTTCAGTATGCC	144		
Query 128	ATGGCTATTCTATGGATACCTTT	187		
Sbjct 145	ATGGCTATTCTATGGATACCTTT	204		
Query 188	GCTAGCTTTCAGGTCAATTTGG	247		
Sbjct 205	GCTAGCTTTCAGGTCAATTTGG	264		
Query 248	CTTAGTCTGGGATAATGTACTT	307		
Sbjct 265	CTTAGTCTGGGATAATGTACTT	324		
Query 308	TGGTGGCTTTCAATCTGAAAC	367		
Sbjct 325	TGGTGGCTTTCAATCTGAAAC	384		
Query 368	GTCCTGATCCCGGTCTGGAGC	427		
Sbjct 385	GTCCTGATCCCGGTCTGGAGC	444		
Query 428	TTGCTTGTAGAGGGCTATAAG	487		
Sbjct 445	TTGCTTGTAGAGGGCTATAAG	504		
Query 488	GTACAGTCCCAAGGCCACTACA	547		
Sbjct 505	GTACAGTCCCAAGGCCACTACA	564		
Query 548	GCTTCATCTGGCAGTGGTGG	607		
Sbjct 565	GCTTCATCTGGCAGTGGTGG	624		
Query 608	GTGAGTAATCCGAGTCCGGT	661		
Sbjct 625	GTGAGTAATCCGAGTCCGGT	678		

**c) Porcine epidemic diarrhea virus strain PEDV/MEX/PUE/01/2016, complete genome**  
Sequence ID: MH006963.1 Length: 27685 Number of Matches: 1

Range 1: 25541 to 26194 [GenBank](#) [Graphics](#) [Next Match](#) [Previous Match](#)

Score	Expect	Identities	Gaps	Strand
1197 bits(648)	0.0	652/654(99%)	0/654(0%)	Plus/Plus
Query 8	GATGAGGTGATTGAACACCTTAGA	67		
Sbjct 25541	GATGAGGTGATTGAACACCTTAGA	25600		
Query 68	CTACTTGTAGTCTTCAGTATGCC	127		
Sbjct 25601	CTACTTGTAGTCTTCAGTATGCC	25600		
Query 128	ATGGCTATTCTATGGATACCTTT	187		
Sbjct 25661	ATGGCTATTCTATGGATACCTTT	25720		
Query 188	GCTAGCTTTCAGGTCAATTTGG	247		
Sbjct 25721	GCTAGCTTTCAGGTCAATTTGG	25780		
Query 248	CTTAGTCTGGGATAATGTACTT	307		
Sbjct 25781	CTTAGTCTGGGATAATGTACTT	25840		
Query 308	TGGTGGCTTTCAATCTGAAAC	367		
Sbjct 25841	TGGTGGCTTTCAATCTGAAAC	25900		
Query 368	GTCCTGATCCCGGTCTGGAGC	427		
Sbjct 25901	GTCCTGATCCCGGTCTGGAGC	25960		
Query 428	TTGCTTGTAGAGGGCTATAAG	487		
Sbjct 25961	TTGCTTGTAGAGGGCTATAAG	26020		
Query 488	GTACAGTCCCAAGGCCACTACA	547		
Sbjct 26021	GTACAGTCCCAAGGCCACTACA	26080		
Query 548	GCTTCATCTGGCAGTGGTGG	607		
Sbjct 26081	GCTTCATCTGGCAGTGGTGG	26140		
Query 608	GTGAGTAATCCGAGTCCGGT	661		
Sbjct 26141	GTGAGTAATCCGAGTCCGGT	26194		

**Ilustración 7. Resultados de secuenciación del plásmido pPICZαB-M<sup>2-226</sup>.** a) Sección del electroferograma de la secuencia del plásmido pPICZαB-M<sup>2-226</sup>. b) Resultado del alineamiento de la secuenciación del plásmido pPICZαB-M<sup>2-226</sup> con la secuencia nucleotídica de la proteína M de la cepa CV777 del vDEP. c) Resultado del alineamiento de la secuenciación del plásmido pPICZαB-M<sup>2-226</sup> con la secuencia nucleotídica de la proteína M de la cepa VDEP/MEX/PUE/01/2016.

En la **ilustración 8** se expone el resultado de la secuenciación para el plásmido pPICZαB-M<sup>102-212</sup>. Un fragmento del electroferograma se muestra en la **ilustración 8a**. El análisis comparativo de la secuencia obtenida en la plataforma BLAST demuestra un 98% de identidad con la secuencia nucleotídica de la proteína M de la cepa CV777 del vDEP (**ilustración 8b**) y un 99% de identidad con una cepa de vDEP mexicana, en esta ocasión con una cepa de Sonora (**ilustración 8c**) además de otras cepas mexicanas (dato no mostrado).



**Ilustración 8. Resultados de secuenciación del plásmido pPICZαB-M<sup>102-212</sup>.** a) Sección del electroferograma de la secuencia del plásmido pPICZαB-M<sup>102-212</sup>. b) Resultado del alineamiento de la secuenciación del plásmido pPICZαB-M<sup>102-212</sup> con la secuencia nucleotídica de la proteína M de la cepa CV777 del vDEP. c) Resultado del alineamiento de la secuenciación del plásmido pPICZαB-M<sup>102-212</sup> con la secuencia nucleotídica de la proteína M de la cepa VDEP/MEX/SON/01/2015.

Las secuencias obtenidas analizadas y editadas se alinearon con la secuencia de la proteína M en el vector pETSUMO-M, el análisis realizado con el programa EMBOSS Needle [https://www.ebi.ac.uk/Tools/psa/emboss\\_needle/](https://www.ebi.ac.uk/Tools/psa/emboss_needle/), demuestra que existen diferencias también con respecto a la plantilla de donde se obtuvo la secuencia. **Ilustración 9.**

SecuenciaM	1	-----tctaacgggtttttattcccgttGATGAGGTGATTGAACACCTTAG	44
pETSUMO-M	1	GGCATGCTAACGGTTTTATTCCCGTTGATGAGGTGATTGAACACCTTAG	50
SecuenciaM	45	AAACTGGAATTCACATGGAATATCATACTGACGATACTACTTGTAGTGC	94
pETSUMO-M	51	AAACTGGAATTCACATGGAATATCATACTGACGATACTACTTGTAGTGC	100
SecuenciaM	95	TTCAGTATGGCCATTACAAGTACTCTGCGTCTTGATGGTGTCAAGATG	144
pETSUMO-M	101	TTCAGTATGGCCATTACAAGTACTCTGCGTCTTGATGGTGTCAAGATG	150
SecuenciaM	145	GCTATTCATGGATACTTTGGCCTCTTGTTAGCACTGTCACTTTTGA	194
pETSUMO-M	151	GCTATTCATGGATACTTTGGCCTCTTGTTAGCACTGTCACTTTTGA	200
SecuenciaM	195	TGCATGGGCTAGCTTTCAGGTCAATTGGGCTTTTTTGGCTTTCAGCATCC	244
pETSUMO-M	201	TGCATGGGCTAGCTTTCAGGTCAACTGGGCTTTTTTCGCTTTCAGCATCC	250
SecuenciaM	245	TTATGGCTTGATCACCTTATGCTGTTGGATAATGACTTTGTCAATAGC	294
pETSUMO-M	251	TTATGGCTTGATCACCTTATGCTGTTGGATAATGACTTTGTCAATAGC	300
SecuenciaM	295	ATTCGGTTGTGGCCAGGACACATTCTTGGTGGTCTTTCATCTGAAAC	344
pETSUMO-M	301	ATTCGGTTGTGGCCAGGACACATTCTTGGTGGTCTTTCATCTGAAAC	350
SecuenciaM	345	AGACGCGTTCTCACTACTTCTGTGATGGCCGACAGGCTGCATTCCGG	394
pETSUMO-M	351	TGACGCGCTTCTCACTACTTCTGTGATGGCCGACAGGCTGCATTCCAG	400
SecuenciaM	395	TGCTTGAGCACC AAC TGGTGT AACGCTAACACTCCTTAGTGGTACATTG	444
pETSUMO-M	401	TGCTTGAGCACC AAC TGGTGT AACGCTAACACTCCTTAGTGGTACATTG	450
SecuenciaM	445	CTTGTAGAGGGCTATAAGGTTGCTACTGGCGTACAGGTAAGTCAATTACC	494
pETSUMO-M	451	TTTGTAGAGGGCTATAAGGTTGCTACTGGCGTACAGGTAAGTCAATTGCC	500
SecuenciaM	495	TAATTTTCGT CACAGTCGCCAAGGCCACTACAACAATTGCTACGGACGTG	544
pETSUMO-M	501	TGATTTTCGT CACAGTCGCCAAGGCCACTACAACAATTGCTATGGACGTG	550
SecuenciaM	545	TTGGTCGTT CAGTCAATGCTTCATCTGGCACTGGTGGGCTTTCATATGC	594
pETSUMO-M	551	TTGGTCGTT CAGTCAATGCTTCATCTGGCACTGGTGGGCTTTCATATGC	600
SecuenciaM	595	CGGTCAAACACGGCGACTACTCAGCTGTGAGTAATCCGAGTTCGGTTCT	644
pETSUMO-M	601	CGGTCAAACACGGCGACTACTCAGCTGTGAGTAATCCGAGTTCGGTTCT	650
SecuenciaM	645	CACAGATAGTGAGAAAGTCTTCATTTAGTCTTC TAGAACAAAACTC-	693
pETSUMO-M	651	CACAGATAGTGAGAAAGTCTTCATTTAGTCTAATAAGCGCGGCATGTCT	700

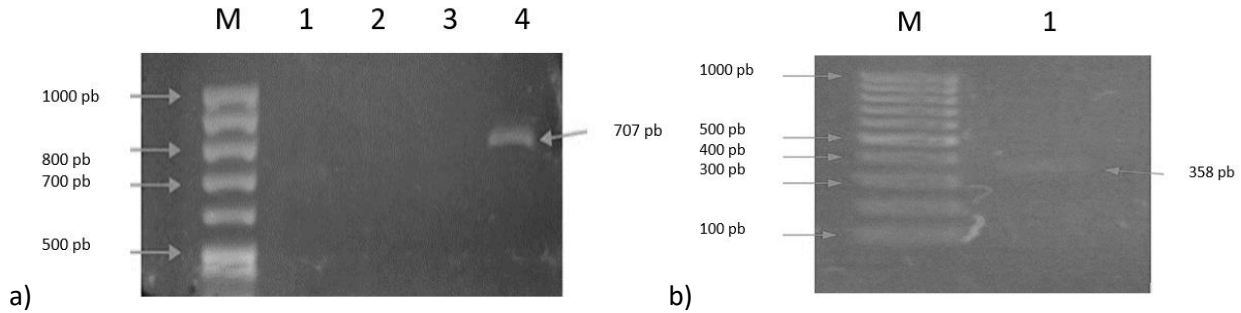
**Ilustración 9.** Resultados del alineamiento de la secuencia consenso con la secuencia nucleotídica de la proteína M del vector pETSUMO-M de INIFAP. “SecuenciaM” secuencia consenso. “SecuenciaDrFr” secuencia del plásmido pETSUMO-M.

Se realizó la traducción *in silico* con el programa translate de ExPasy de la secuencia consenso M, de la secuencia nucleotídica del plásmido pETSUMO-M del INIFAP y de la secuencia nucleotídica disponible en NCBI de la cepa de VDEP VDEP/MEX/PUE/01/2016 (GenBank: MH006963.1) y se compararon con la secuencia aminoacídica de la proteína M de la cepa CV777 mediante Clustal Omega (**ilustración 10**). Se observaron 5 mutaciones de la proteína M<sup>2-226</sup> con respecto a la proteína M de la cepa CV777, 4 mutaciones con respecto a la proteína M del plásmido pETSUMO-M y solo 2 mutaciones con respecto a la proteína M de la cepa VDEP/MEX/PUE/01/2016.

CV777	MSNGSIPVDEVIEHLRNWNFTWNIILTILLVVLQYGHYKYSVFLYGVKMAILWILWPLVL	60
PEDV/MEX/MICH01/2013-V	MSNGFIPVDEVIEHLRNWNFTWNIILTILLVVLQYGHYKYSVFLYGVKMAILWILWPLVL	180
secuenciafinal	-SNGFIPVDEVIEHLRNWNFTWNIILTILLVVLQYGHYKYSVFLYGVKMAILWILWPLVL	59
pPICZaB2-226	-SNGSIPVDEVIQHLRNWNFTWNIILTILLVVLQYGHYKYSVFLYGVKMAILWILWPLVL	59
	*** *****.*****.*****.*****.*****.*****.*****	
CV777	ALSFLDAWASFQVNWVFFAFSILMACITLMLWIMYFVNSIRLWRRTHSWWSFNPETDALL	120
PEDV/MEX/MICH01/2013-V	ALSFLDAWASFQVNWVFFAFSILMACITLMLWIMYFVNSIRLWRRTHSWWSFNPETDALL	240
secuenciafinal	ALSFLDAWASFQVNWVFFAFSILMACITLMLWIMYFVNSIRLWRRTHSWWSFNPETDALL	119
pPICZaB2-226	ALSFLDAWASFQVNWVFFAFSILMACITLMLWIMYFVNSIRLWRRTHSWWSFNPETDALL	119
	*****.*****.*****.*****.*****.*****.*****	
CV777	TTSVMGRQVCIPVLGAPTGVTLTLLSGTLFVEGYKVATGVQVSQLPNFVTVAKATTTIVY	180
PEDV/MEX/MICH01/2013-V	TTSVMGRQVCIPVLGAPTGVTLTLLSGTLFVEGYKVATGVQVSQLPDPFVTVAKATTTIVY	300
secuenciafinal	TTSVMGRQVCIPVLGAPTGVTLTLLSGTLFVEGYKVATGVQVSQLPNFVTVAKATTTIVY	179
pPICZaB2-226	TTSVMGRQVCIPVLGAPTGVTLTLLSGTLFVEGYKVATGVQVSQLPNFVTVAKATTTIVY	179
	*****.*****.*****.*****.*****.*****.*****	
CV777	GRVGRSVNASSGTGWAFYVRSKHGDYSAVSNPSSVLTSEKVLHLV-----	226
PEDV/MEX/MICH01/2013-V	GRVGRSVNASSGTGWAFYVRSKHGDYSAVSNPSSVLTSEKVLHLV-----	346
secuenciafinal	GRVGRSVNASSGTGWAFYVRSKHGDYSAVSNPSSVLTSEKVLHLVLEQKLTISEEDLNS	239
pPICZaB2-226	GRVGRSVNASSGTGWAFYVRSKHGDYSAVSNPSSVLTSEKVLHLV-----	225
	*****.*****.*****.*****.*****.*****.*****	

**Ilustración 10.** Análisis de Clustal Omega de la secuencia aminoacídica de la proteína M<sup>2-226</sup>, la secuencia pETSUMO-M y la proteína M de la cepa CV777. “secuenciafinal” corresponde a la secuencia traducida *in silico* de la secuencia nucleotídica consenso de la proteína M. “CV777” corresponde a la secuencia aminoacídica de la proteína M de la cepa CV777. “PEDV/MEX/MICH01/2013-V.” corresponde a la secuencia traducida *in silico* de la secuencia nucleotídica del plásmido pETSUMO-M.

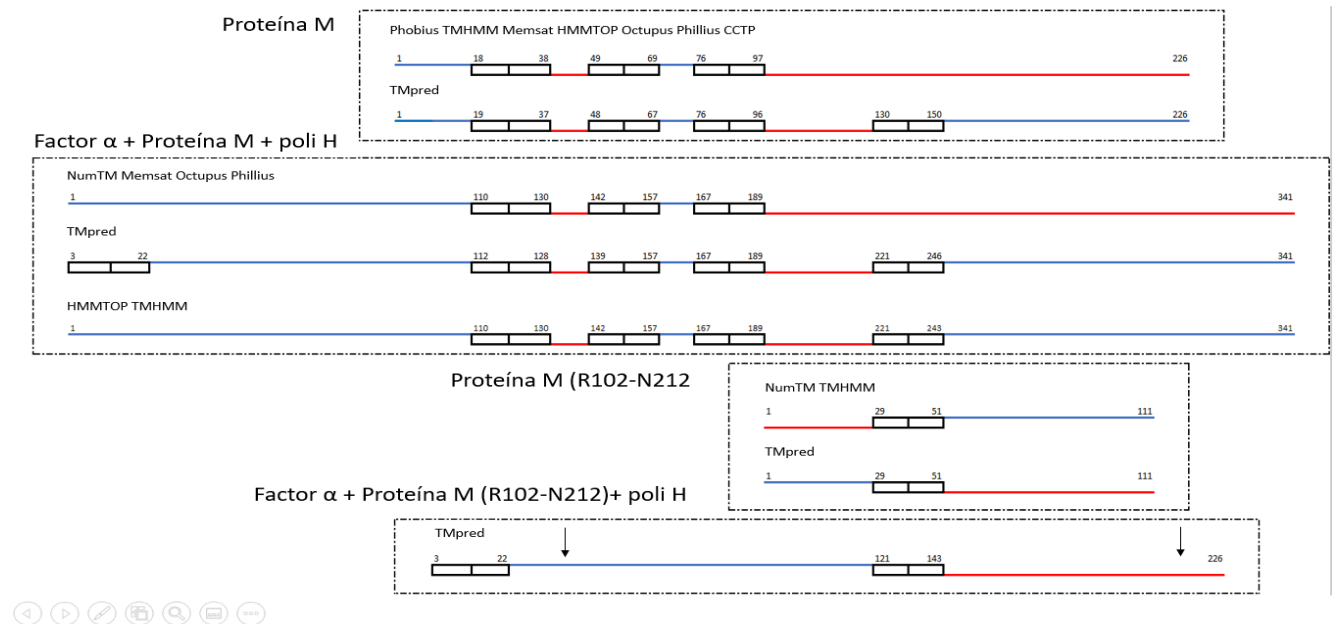
Se realizó la transformación de los plásmidos de pPICZαB-M<sub>2-226</sub> y pPICZαB-M<sub>102-212</sub> en levaduras *Pichia pastoris*. Los vectores pPICZαB-M<sub>2-226</sub> y pPICZαB-M<sub>102-212</sub> fueron linearizados con *PmeI* para integrarlo por recombinación homóloga al genoma de *P. pastoris* en su promotor nativo de AOX1. El vector linearizado se incorporó por transformación de la cepa X33 de *P. pastoris*, para ello se llevó a cabo por choque térmico usando el método de cloruro de litio y después de transcurridos los días necesarios, se observó el crecimiento de colonias en medio AGAR YPD + Zeocina. Se seleccionaron 4 cepas para extraer ADN cromosómico y realizar una PCR para amplificar el inserto correspondiente a M<sub>2-226</sub> y M<sub>102-212</sub> (Ilustración 11), usando los iniciadores MFecolarga, MRxbalarga, MFecorta y MRxbacorta, para finalmente corroborar la incorporación del vector al cromosoma de la levadura, estas cepas fueron denominadas X33-pPICZαB-M<sub>2-226</sub> y X33-pPICZαB-M<sub>102-212</sub>.



**Ilustración 11. Fragmentos de PCR amplificados a partir del genoma de las cepas de *P. pastoris* transformadas con los plásmidos a) pPICZαB-M<sub>2-226</sub> y b) pPICZαB-M<sub>102-212</sub>.**

#### V.IV Análisis *in silico* de la proteína M

Para cumplir con el objetivo relacionado con el análisis *in silico* de la proteína M se realizó la traducción *in silico* de las secuencias nucleotídicas obtenidas inicialmente, se realizó un análisis de hidrofobicidad para conocer las regiones transmembranales de la proteína, con los programas CCTOP, TMpred, MEMSAT, HMMTOP, TMHMM, OCTOPUS, Philius y Phobius (ilustración 12). Adicionalmente se usaron los programas WoLFPSORT y PSORT II, la construcción de la proteína M<sup>102-212</sup> en el plásmido pPICZαB es la proteína con mayor probabilidad de encontrarse en el exterior.



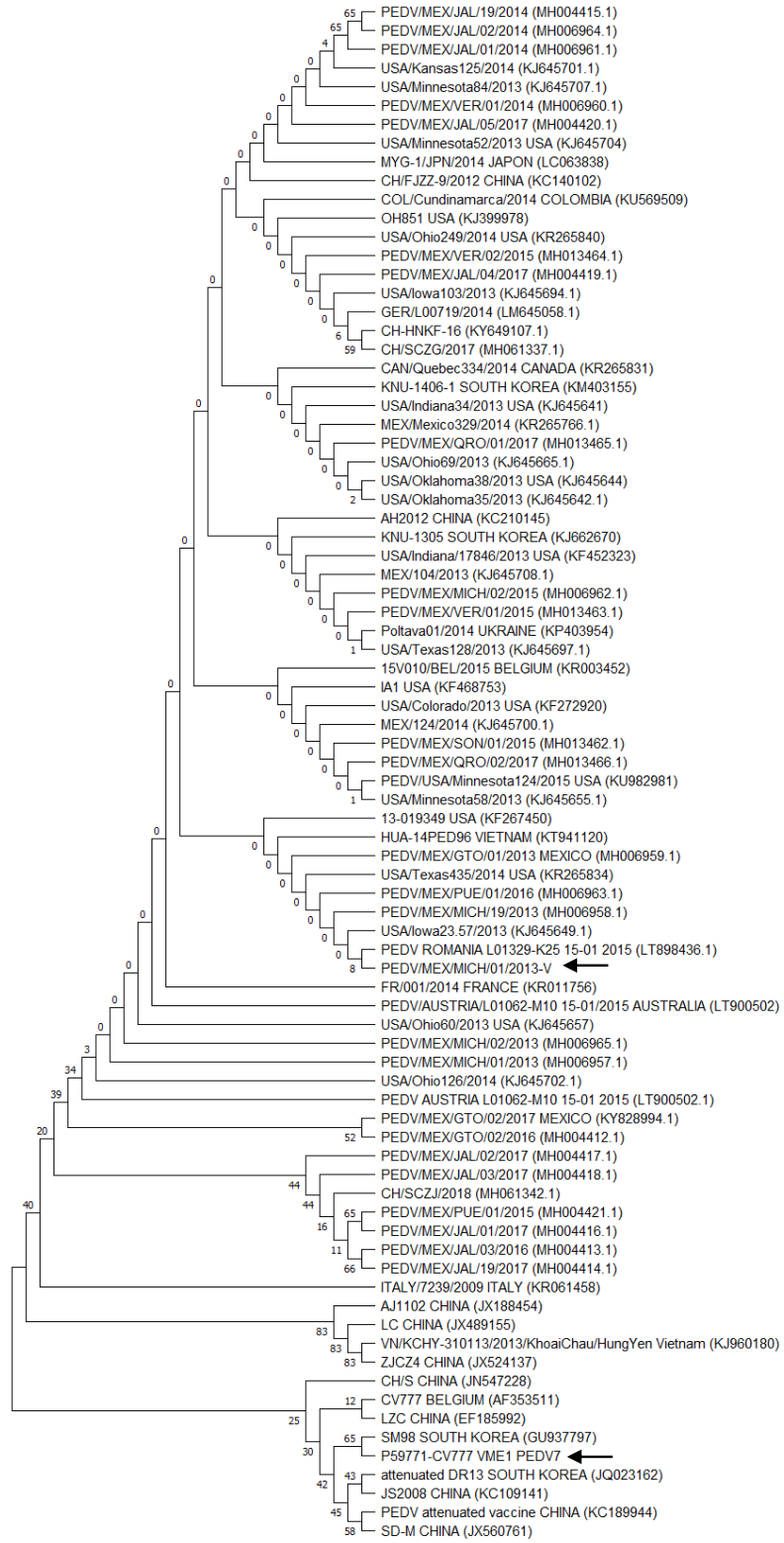
**Ilustración 12. Topología y regiones transmembranales de la proteína M y sus variantes.** Los rectángulos representan las regiones transmembranales. La línea azul representa la fracción de la proteína que se encuentra en el exterior del virus mientras que la línea roja la región interna. Las ilustraciones se encuentran alineadas con respecto a la secuencia nativa de la proteína M. La primera flecha limita el factor α y la segunda la proteína M, por último, se encuentra la cola de histidina.

## V.V Predicción de la estructura 3D de la proteína M

Para generar el modelo 3D de la estructura de la proteína VDEP M, se llevó a cabo una comparación de secuencias de aminoácidos. Los resultados mostraron una identidad del 97.35 al 99.56% entre las cepas de México, Estados Unidos, Europa y Asia. Las secuencias de aminoácidos de la proteína M se agruparon en tres grupos (**Ilustración 13**). Se observó un alto grado de conservación entre 87 cepas de vDEP de diferentes latitudes cuando se utilizó una secuencia de proteína M de la cepa prototipo CV777 como referencia. Sin embargo, se observaron 9 diferencias de aminoácidos, 8 en la región transmembrana (**Ilustración 14**). Por tanto, el modelo 3D de la proteína M predicha se construyó basándose en la secuencia de la proteína M de la cepa vDEP CV777 (denominada proteína M CV777 en este trabajo, UniProt ID P59771).

**Ilustración 13. Árbol filogenético de las secuencias utilizadas en el alineamiento.**

La historia evolutiva se infirió mediante el método de unión de vecinos (131). Se muestra el árbol óptimo con la suma de la longitud de la rama = 0.04452215. El porcentaje de árboles replicados en los que los taxones asociados se agruparon en la prueba de arranque (1000 réplicas) se muestra junto a las ramas (132). Las distancias evolutivas se calcularon utilizando el método de corrección de Poisson (133) y están en las unidades del número de sustituciones de aminoácidos por sitio. Este análisis involucró 43 secuencias de aminoácidos. Todas las posiciones ambiguas se eliminaron para cada par de secuencias (opción de eliminación por pares). Hubo un total de 226 posiciones en el conjunto de datos final. Los análisis evolutivos se realizaron en MEGA X (134). Las flechas señalan la secuencia CV777, y PEDV/MEX/MICH01/2013-V.



G2a

G2b

G1

ITALY/7239/2009\_ITALY\_(KR061458) MSNGSTPVDEV----IQLRNNFTNWTITLLVVLQGHYYSFVLYGVMAIWLTLWPLVLAISLFDAAASFQNNVFFAFSLMACITLMLIMYFVNSILWRRTSHKSFNPE 116  
SM98\_SOUTH\_KOREA\_(G0937797) MSNGSTPVDEV----IHLRNNFTNWTITLLVVLQGHYYSFVLYGVMAIWLTLWPLVLAISLFDAAASFQNNVFFAFSLMACITLMLIMYFVNSILWRRTSHKSFNPE 120  
L2C\_CHINA\_(F7187992) MSNGSTPVDEV----IHLRNNFTNWTITLLVVLQGHYYSFVLYGVMAIWLTLWPLVLAISLFDAAASFQNNVFFAFSLMACITLMLIMYFVNSILWRRTSHKSFNPE 116  
CH/S\_CHINA\_(3X642128) MSNGSTPVDEV----IHLRNNFTNWTITLLVVLQGHYYSFVLYGVMAIWLTLWPLVLAISLFDAAASFQNNVFFAFSLMACITLMLIMYFVNSILWRRTSHKSFNPE 116  
A1102\_CHINA\_(3X188454) MSNGSTPVDEV----IHLRNNFTNWTITLLVVLQGHYYSFVLYGVMAIWLTLWPLVLAISLFDAAASFQNNVFFAFSLMACITLMLIMYFVNSILWRRTSHKSFNPE 116  
LC\_CHINA\_(3X489155) MSNGSTPVDEV----IHLRNNFTNWTITLLVVLQGHYYSFVLYGVMAIWLTLWPLVLAISLFDAAASFQNNVFFAFSLMACITLMLIMYFVNSILWRRTSHKSFNPE 116  
VN/KCHY-318113/2013/KhoaiChau/HungVen\_Vietnam\_(K3960180) MSNGSTPVDEV----IHLRNNFTNWTITLLVVLQGHYYSFVLYGVMAIWLTLWPLVLAISLFDAAASFQNNVFFAFSLMACITLMLIMYFVNSILWRRTSHKSFNPE 116  
Z3C24\_CHINA\_(3X524137) MSNGSTPVDEV----IHLRNNFTNWTITLLVVLQGHYYSFVLYGVMAIWLTLWPLVLAISLFDAAASFQNNVFFAFSLMACITLMLIMYFVNSILWRRTSHKSFNPE 116  
attenuated\_DR13\_SOUTH\_KOREA\_(J0623162) MSNGSTPVDEV----IHLRNNFTNWTITLLVVLQGHYYSFVLYGVMAIWLTLWPLVLAISLFDAAASFQNNVFFAFSLMACITLMLIMYFVNSILWRRTSHKSFNPE 116  
J52808\_CHINA\_(KC189141) MSNGSTPVDEV----IHLRNNFTNWTITLLVVLQGHYYSFVLYGVMAIWLTLWPLVLAISLFDAAASFQNNVFFAFSLMACITLMLIMYFVNSILWRRTSHKSFNPE 116  
P59771-CV777\_VME1\_PDEV7 MSNGSTPVDEV----IHLRNNFTNWTITLLVVLQGHYYSFVLYGVMAIWLTLWPLVLAISLFDAAASFQNNVFFAFSLMACITLMLIMYFVNSILWRRTSHKSFNPE 116  
CV777\_BELGIUM\_(AF353511) MSNGSTPVDEV----IHLRNNFTNWTITLLVVLQGHYYSFVLYGVMAIWLTLWPLVLAISLFDAAASFQNNVFFAFSLMACITLMLIMYFVNSILWRRTSHKSFNPE 116  
Sequence\_66 MSNGSTPVDEV----IHLRNNFTNWTITLLVVLQGHYYSFVLYGVMAIWLTLWPLVLAISLFDAAASFQNNVFFAFSLMACITLMLIMYFVNSILWRRTSHKSFNPE 116  
PDEV/MEX/ATL/01/2013-V MSNGSTPVDEV----IHLRNNFTNWTITLLVVLQGHYYSFVLYGVMAIWLTLWPLVLAISLFDAAASFQNNVFFAFSLMACITLMLIMYFVNSILWRRTSHKSFNPE 116  
SD-MI\_CHINA\_(3X560761) MSNGSTPVDEV----IHLRNNFTNWTITLLVVLQGHYYSFVLYGVMAIWLTLWPLVLAISLFDAAASFQNNVFFAFSLMACITLMLIMYFVNSILWRRTSHKSFNPE 116  
PDEV/MEX/PUE/01/2015\_(MH004421.1) MSNGSTPVDEV----IHLRNNFTNWTITLLVVLQGHYYSFVLYGVMAIWLTLWPLVLAISLFDAAASFQNNVFFAFSLMACITLMLIMYFVNSILWRRTSHKSFNPE 116  
PDEV/MEX/JAL/01/2017\_(MH004416.1) MSNGSTPVDEV----IHLRNNFTNWTITLLVVLQGHYYSFVLYGVMAIWLTLWPLVLAISLFDAAASFQNNVFFAFSLMACITLMLIMYFVNSILWRRTSHKSFNPE 116  
PDEV/MEX/JAL/06/2017\_MEXICO MSNGSTPVDEV----IHLRNNFTNWTITLLVVLQGHYYSFVLYGVMAIWLTLWPLVLAISLFDAAASFQNNVFFAFSLMACITLMLIMYFVNSILWRRTSHKSFNPE 116  
PDEV/MEX/JAL/03/2016\_(MH004413.1) MSNGSTPVDEV----IHLRNNFTNWTITLLVVLQGHYYSFVLYGVMAIWLTLWPLVLAISLFDAAASFQNNVFFAFSLMACITLMLIMYFVNSILWRRTSHKSFNPE 116  
PDEV/MEX/JAL/19/2017\_(MH004414.1) MSNGSTPVDEV----IHLRNNFTNWTITLLVVLQGHYYSFVLYGVMAIWLTLWPLVLAISLFDAAASFQNNVFFAFSLMACITLMLIMYFVNSILWRRTSHKSFNPE 116  
PDEV/MEX/JAL/02/2017\_(MH004417.1) MSNGSTPVDEV----IHLRNNFTNWTITLLVVLQGHYYSFVLYGVMAIWLTLWPLVLAISLFDAAASFQNNVFFAFSLMACITLMLIMYFVNSILWRRTSHKSFNPE 116  
PDEV/MEX/JAL/03/2017\_(MH004418.1) MSNGSTPVDEV----IHLRNNFTNWTITLLVVLQGHYYSFVLYGVMAIWLTLWPLVLAISLFDAAASFQNNVFFAFSLMACITLMLIMYFVNSILWRRTSHKSFNPE 116  
CH/SC23/2018\_(MH061342.1) MSNGSTPVDEV----IHLRNNFTNWTITLLVVLQGHYYSFVLYGVMAIWLTLWPLVLAISLFDAAASFQNNVFFAFSLMACITLMLIMYFVNSILWRRTSHKSFNPE 116  
PDEV/MEX/MICH/01/2013-V MSNGSTPVDEV----IHLRNNFTNWTITLLVVLQGHYYSFVLYGVMAIWLTLWPLVLAISLFDAAASFQNNVFFAFSLMACITLMLIMYFVNSILWRRTSHKSFNPE 116  
CH-INKF-16\_(3X642107.1) MSNGSTPVDEV----IHLRNNFTNWTITLLVVLQGHYYSFVLYGVMAIWLTLWPLVLAISLFDAAASFQNNVFFAFSLMACITLMLIMYFVNSILWRRTSHKSFNPE 116  
CH/SC26/2017\_(MH061337.1) MSNGSTPVDEV----IHLRNNFTNWTITLLVVLQGHYYSFVLYGVMAIWLTLWPLVLAISLFDAAASFQNNVFFAFSLMACITLMLIMYFVNSILWRRTSHKSFNPE 116  
USA/Oklahoma38/2013\_USA\_(K3645644) MSNGSTPVDEV----IHLRNNFTNWTITLLVVLQGHYYSFVLYGVMAIWLTLWPLVLAISLFDAAASFQNNVFFAFSLMACITLMLIMYFVNSILWRRTSHKSFNPE 116  
Poltava01/2014\_UKRAINE\_(K403954) MSNGSTPVDEV----IHLRNNFTNWTITLLVVLQGHYYSFVLYGVMAIWLTLWPLVLAISLFDAAASFQNNVFFAFSLMACITLMLIMYFVNSILWRRTSHKSFNPE 116  
PDEV/USA/Minnesota124/2015\_USA\_(KU982981) MSNGSTPVDEV----IHLRNNFTNWTITLLVVLQGHYYSFVLYGVMAIWLTLWPLVLAISLFDAAASFQNNVFFAFSLMACITLMLIMYFVNSILWRRTSHKSFNPE 116  
PDEV/MEX/JAL/OR/2014\_MEXICO MSNGSTPVDEV----IHLRNNFTNWTITLLVVLQGHYYSFVLYGVMAIWLTLWPLVLAISLFDAAASFQNNVFFAFSLMACITLMLIMYFVNSILWRRTSHKSFNPE 116  
PDEV/MEX/JAL/01/2014\_(MH006961.1) MSNGSTPVDEV----IHLRNNFTNWTITLLVVLQGHYYSFVLYGVMAIWLTLWPLVLAISLFDAAASFQNNVFFAFSLMACITLMLIMYFVNSILWRRTSHKSFNPE 116  
PDEV/MEX/JAL/19/2014\_(MH004415.1) MSNGSTPVDEV----IHLRNNFTNWTITLLVVLQGHYYSFVLYGVMAIWLTLWPLVLAISLFDAAASFQNNVFFAFSLMACITLMLIMYFVNSILWRRTSHKSFNPE 116  
PDEV/MEX/JAL/02/2014\_(MH006964.1) MSNGSTPVDEV----IHLRNNFTNWTITLLVVLQGHYYSFVLYGVMAIWLTLWPLVLAISLFDAAASFQNNVFFAFSLMACITLMLIMYFVNSILWRRTSHKSFNPE 116  
13-019349\_USA\_(KF267458) MSNGSTPVDEV----IHLRNNFTNWTITLLVVLQGHYYSFVLYGVMAIWLTLWPLVLAISLFDAAASFQNNVFFAFSLMACITLMLIMYFVNSILWRRTSHKSFNPE 116  
FV938/BEI/2011\_USA\_(KR003452) MSNGSTPVDEV----IHLRNNFTNWTITLLVVLQGHYYSFVLYGVMAIWLTLWPLVLAISLFDAAASFQNNVFFAFSLMACITLMLIMYFVNSILWRRTSHKSFNPE 116  
AH2012\_CHINA\_(KC120105) MSNGSTPVDEV----IHLRNNFTNWTITLLVVLQGHYYSFVLYGVMAIWLTLWPLVLAISLFDAAASFQNNVFFAFSLMACITLMLIMYFVNSILWRRTSHKSFNPE 116  
CAN/Quebec134/2014\_CANADA\_(KR265831) MSNGSTPVDEV----IHLRNNFTNWTITLLVVLQGHYYSFVLYGVMAIWLTLWPLVLAISLFDAAASFQNNVFFAFSLMACITLMLIMYFVNSILWRRTSHKSFNPE 116  
CH/F22-9/2012\_CHINA\_(KC148102) MSNGSTPVDEV----IHLRNNFTNWTITLLVVLQGHYYSFVLYGVMAIWLTLWPLVLAISLFDAAASFQNNVFFAFSLMACITLMLIMYFVNSILWRRTSHKSFNPE 116  
COL/Cundinamarca/2014\_COLOMBIA\_(KU569509) MSNGSTPVDEV----IHLRNNFTNWTITLLVVLQGHYYSFVLYGVMAIWLTLWPLVLAISLFDAAASFQNNVFFAFSLMACITLMLIMYFVNSILWRRTSHKSFNPE 116  
FR/001/2014\_FRANCE\_(KR11756) MSNGSTPVDEV----IHLRNNFTNWTITLLVVLQGHYYSFVLYGVMAIWLTLWPLVLAISLFDAAASFQNNVFFAFSLMACITLMLIMYFVNSILWRRTSHKSFNPE 116  
HUA-14PDE96\_VIETNAM\_(KT941120) MSNGSTPVDEV----IHLRNNFTNWTITLLVVLQGHYYSFVLYGVMAIWLTLWPLVLAISLFDAAASFQNNVFFAFSLMACITLMLIMYFVNSILWRRTSHKSFNPE 116  
IA1\_USA\_(KF468753) MSNGSTPVDEV----IHLRNNFTNWTITLLVVLQGHYYSFVLYGVMAIWLTLWPLVLAISLFDAAASFQNNVFFAFSLMACITLMLIMYFVNSILWRRTSHKSFNPE 116  
KNU-1305\_SOUTH\_KOREA\_(K3662670) MSNGSTPVDEV----IHLRNNFTNWTITLLVVLQGHYYSFVLYGVMAIWLTLWPLVLAISLFDAAASFQNNVFFAFSLMACITLMLIMYFVNSILWRRTSHKSFNPE 116  
KNU-1406-1\_SOUTH\_KOREA\_(KM403155) MSNGSTPVDEV----IHLRNNFTNWTITLLVVLQGHYYSFVLYGVMAIWLTLWPLVLAISLFDAAASFQNNVFFAFSLMACITLMLIMYFVNSILWRRTSHKSFNPE 116  
L00721/GER/2014 MSNGSTPVDEV----IHLRNNFTNWTITLLVVLQGHYYSFVLYGVMAIWLTLWPLVLAISLFDAAASFQNNVFFAFSLMACITLMLIMYFVNSILWRRTSHKSFNPE 116  
WIS-1/3PM/2014\_JAPAN\_(LC063838) MSNGSTPVDEV----IHLRNNFTNWTITLLVVLQGHYYSFVLYGVMAIWLTLWPLVLAISLFDAAASFQNNVFFAFSLMACITLMLIMYFVNSILWRRTSHKSFNPE 116  
MH51\_USA\_(K3399978) MSNGSTPVDEV----IHLRNNFTNWTITLLVVLQGHYYSFVLYGVMAIWLTLWPLVLAISLFDAAASFQNNVFFAFSLMACITLMLIMYFVNSILWRRTSHKSFNPE 116  
PDEV/AUSTRIA/L01062-M10\_15-01/2015\_AUSTRALIA\_(L7900502) MSNGSTPVDEV----IHLRNNFTNWTITLLVVLQGHYYSFVLYGVMAIWLTLWPLVLAISLFDAAASFQNNVFFAFSLMACITLMLIMYFVNSILWRRTSHKSFNPE 116  
PDEV/MEX/GTO/01/2013\_MEXICO\_(MH006959.1) MSNGSTPVDEV----IHLRNNFTNWTITLLVVLQGHYYSFVLYGVMAIWLTLWPLVLAISLFDAAASFQNNVFFAFSLMACITLMLIMYFVNSILWRRTSHKSFNPE 116  
PDEV/MEX/MICH/OR/2013\_MEXICO MSNGSTPVDEV----IHLRNNFTNWTITLLVVLQGHYYSFVLYGVMAIWLTLWPLVLAISLFDAAASFQNNVFFAFSLMACITLMLIMYFVNSILWRRTSHKSFNPE 116  
USA/Colorado/2013\_USA\_(KF272920) MSNGSTPVDEV----IHLRNNFTNWTITLLVVLQGHYYSFVLYGVMAIWLTLWPLVLAISLFDAAASFQNNVFFAFSLMACITLMLIMYFVNSILWRRTSHKSFNPE 116  
USA/Indiana/17846/2013\_USA\_(KF452323) MSNGSTPVDEV----IHLRNNFTNWTITLLVVLQGHYYSFVLYGVMAIWLTLWPLVLAISLFDAAASFQNNVFFAFSLMACITLMLIMYFVNSILWRRTSHKSFNPE 116  
USA/Indiana34/2013\_USA\_(K3645641) MSNGSTPVDEV----IHLRNNFTNWTITLLVVLQGHYYSFVLYGVMAIWLTLWPLVLAISLFDAAASFQNNVFFAFSLMACITLMLIMYFVNSILWRRTSHKSFNPE 116  
USA/Minnesota25/2013\_USA\_(K3645704) MSNGSTPVDEV----IHLRNNFTNWTITLLVVLQGHYYSFVLYGVMAIWLTLWPLVLAISLFDAAASFQNNVFFAFSLMACITLMLIMYFVNSILWRRTSHKSFNPE 116  
USA/Ohio249/2014\_USA\_(KR265840) MSNGSTPVDEV----IHLRNNFTNWTITLLVVLQGHYYSFVLYGVMAIWLTLWPLVLAISLFDAAASFQNNVFFAFSLMACITLMLIMYFVNSILWRRTSHKSFNPE 116  
USA/Ohio60/2011\_USA\_(K3645657) MSNGSTPVDEV----IHLRNNFTNWTITLLVVLQGHYYSFVLYGVMAIWLTLWPLVLAISLFDAAASFQNNVFFAFSLMACITLMLIMYFVNSILWRRTSHKSFNPE 116  
USA/Texas435/2014\_USA\_(KF265834) MSNGSTPVDEV----IHLRNNFTNWTITLLVVLQGHYYSFVLYGVMAIWLTLWPLVLAISLFDAAASFQNNVFFAFSLMACITLMLIMYFVNSILWRRTSHKSFNPE 116  
MEX/124/2014\_(K3645700.1) MSNGSTPVDEV----IHLRNNFTNWTITLLVVLQGHYYSFVLYGVMAIWLTLWPLVLAISLFDAAASFQNNVFFAFSLMACITLMLIMYFVNSILWRRTSHKSFNPE 116  
MEX/104/2013\_(K3645708.1) MSNGSTPVDEV----IHLRNNFTNWTITLLVVLQGHYYSFVLYGVMAIWLTLWPLVLAISLFDAAASFQNNVFFAFSLMACITLMLIMYFVNSILWRRTSHKSFNPE 116  
MEX/Mexico329/2014\_(KR265766.1) MSNGSTPVDEV----IHLRNNFTNWTITLLVVLQGHYYSFVLYGVMAIWLTLWPLVLAISLFDAAASFQNNVFFAFSLMACITLMLIMYFVNSILWRRTSHKSFNPE 116  
PDEV/MEX/JAL/05/2017\_(MH004420.1) MSNGSTPVDEV----IHLRNNFTNWTITLLVVLQGHYYSFVLYGVMAIWLTLWPLVLAISLFDAAASFQNNVFFAFSLMACITLMLIMYFVNSILWRRTSHKSFNPE 116  
PDEV/MEX/VER/02/2015\_(MH013464.1) MSNGSTPVDEV----IHLRNNFTNWTITLLVVLQGHYYSFVLYGVMAIWLTLWPLVLAISLFDAAASFQNNVFFAFSLMACITLMLIMYFVNSILWRRTSHKSFNPE 116  
PDEV/MEX/MICH/02/2013\_(MH006965.1) MSNGSTPVDEV----IHLRNNFTNWTITLLVVLQGHYYSFVLYGVMAIWLTLWPLVLAISLFDAAASFQNNVFFAFSLMACITLMLIMYFVNSILWRRTSHKSFNPE 116  
PDEV/MEX/PUE/01/2016\_(MH006963.1) MSNGSTPVDEV----IHLRNNFTNWTITLLVVLQGHYYSFVLYGVMAIWLTLWPLVLAISLFDAAASFQNNVFFAFSLMACITLMLIMYFVNSILWRRTSHKSFNPE 116  
PDEV/MEX/SON/01/2015\_(MH013463.1) MSNGSTPVDEV----IHLRNNFTNWTITLLVVLQGHYYSFVLYGVMAIWLTLWPLVLAISLFDAAASFQNNVFFAFSLMACITLMLIMYFVNSILWRRTSHKSFNPE 116  
PDEV/MEX/MICH/02/2015\_(MH006962.1) MSNGSTPVDEV----IHLRNNFTNWTITLLVVLQGHYYSFVLYGVMAIWLTLWPLVLAISLFDAAASFQNNVFFAFSLMACITLMLIMYFVNSILWRRTSHKSFNPE 116  
PDEV/MEX/QRO/01/2017\_(MH013465.1) MSNGSTPVDEV----IHLRNNFTNWTITLLVVLQGHYYSFVLYGVMAIWLTLWPLVLAISLFDAAASFQNNVFFAFSLMACITLMLIMYFVNSILWRRTSHKSFNPE 116  
PDEV/MEX/VER/01/2014\_(MH006960.1) MSNGSTPVDEV----IHLRNNFTNWTITLLVVLQGHYYSFVLYGVMAIWLTLWPLVLAISLFDAAASFQNNVFFAFSLMACITLMLIMYFVNSILWRRTSHKSFNPE 116  
PDEV/MEX/JAL/04/2017\_(MH004419.1) MSNGSTPVDEV----IHLRNNFTNWTITLLVVLQGHYYSFVLYGVMAIWLTLWPLVLAISLFDAAASFQNNVFFAFSLMACITLMLIMYFVNSILWRRTSHKSFNPE 116  
PDEV/MEX/MICH/01/2013\_(MH006957.1) MSNGSTPVDEV----IHLRNNFTNWTITLLVVLQGHYYSFVLYGVMAIWLTLWPLVLAISLFDAAASFQNNVFFAFSLMACITLMLIMYFVNSILWRRTSHKSFNPE 116  
PDEV/MEX/MICH/19/2013\_(MH006958.1) MSNGSTPVDEV----IHLRNNFTNWTITLLVVLQGHYYSFVLYGVMAIWLTLWPLVLAISLFDAAASFQNNVFFAFSLMACITLMLIMYFVNSILWRRTSHKSFNPE 116  
PDEV/MEX/QRO/02/2017\_(MH013466.1) MSNGSTPVDEV----IHLRNNFTNWTITLLVVLQGHYYSFVLYGVMAIWLTLWPLVLAISLFDAAASFQNNVFFAFSLMACITLMLIMYFVNSILWRRTSHKSFNPE 116  
PDEV/MEX/VER/01/2015\_(MH013464.1) MSNGSTPVDEV----IHLRNNFTNWTITLLVVLQGHYYSFVLYGVMAIWLTLWPLVLAISLFDAAASFQNNVFFAFSLMACITLMLIMYFVNSILWRRTSHKSFNPE 116  
USA/Ohio69/2013\_(K3645665.1) MSNGSTPVDEV----IHLRNNFTNWTITLLVVLQGHYYSFVLYGVMAIWLTLWPLVLAISLFDAAASFQNNVFFAFSLMACITLMLIMYFVNSILWRRTSHKSFNPE 116  
USA/Minnesota184/2013\_(K3645707.1) MSNGSTPVDEV----IHLRNNFTNWTITLLVVLQGHYYSFVLYGVMAIWLTLWPLVLAISLFDAAASFQNNVFFAFSLMACITLMLIMYFVNSILWRRTSHKSFNPE 116  
USA/Iowa103/2013\_(K3645694.1) MSNGSTPVDEV----IHLRNNFTNWTITLLVVLQGHYYSFVLYGVMAIWLTLWPLVLAISLFDAAASFQNNVFFAFSLMACITLMLIMYFVNSILWRRTSHKSFNPE 116  
USA/Ohio126/2014\_(K3645702.1) MSNGSTPVDEV----IHLRNNFTNWTITLLVVLQGHYYSFVLYGVMAIWLTLWPLVLAISLFDAAASFQNNVFFAFSLMACITLMLIMYFVNSILWRRTSHKSFNPE 116  
USA/Iowa23.57/2013\_(K3645649.1) MSNGSTPVDEV----IHLRNNFTNWTITLLVVLQGHYYSFVLYGVMAIWLTLWPLVLAISLFDAAASFQNNVFFAFSLMACITLMLIMYFVNSILWRRTSHKSFNPE 116  
USA/Minnesota58/2013\_(K3645655.1) MSNGSTPVDEV----IHLRNNFTNWTITLLVVLQGHYYSFVLYGVMAIWLTLWPLVLAISLFDAAASFQNNVFFAFSLMACITLMLIMYFVNSILWRRTSHKSFNPE 116  
USA/Texas128/2013\_(K3645697.1) MSNGSTPVDEV----IHLRNNFTNWTITLLVVLQGHYYSFVLYGVMAIWLTLWPLVLAISLFDAAASFQNNVFFAFSLMACITLMLIMYFVNSILWRRTSHKSFNPE 116  
USA/Oklahoma35/2013\_(K3645642.1) MSNGSTPVDEV----IHLRNNFTNWTITLLVVLQGHYYSFVLYGVMAIWLTLWPLVLAISLFDAAASFQNNVFFAFSLMACITLMLIMYFVNSILWRRTSHKSFNPE 116  
USA/Kansas125/2014\_(K3645701.1) MSNGSTPVDEV----IHLRNNFTNWTITLLVVLQGHYYSFVLYGVMAIWLTLWPLVLAISLFDAAASFQNNVFFAFSLMACITLMLIMYFVNSILWRRTSHKSFNPE 116  
GER/L00719/2014\_(LM645058.1) MSNGSTPVDEV----IHLRNNFTNWTITLLVVLQGHYYSFVLYGVMAIWLTLWPLVLAISLFDAAASFQNNVFFAFSLMACITLMLIMYFVNSILWRRTSHKSFNPE 116  
PDEV\_AUSTRIA\_L01062-M10\_15-01\_2015\_(L7900502.1) MSNGSTPVDEV----IHLRNNFTNWTITLLVVLQGHYYSFVLYGVMAIWLTLWPLVLAISLFDAAASFQNNVFFAFSLMACITLMLIMYFVNSILWRRTSHKSFNPE 116  
PDEV\_ROMANIA\_L01329-K25\_15-01\_2015\_(K7898436.1) MSNGSTPVDEV----IHLRNNFTNWTITLLVVLQGHYYSFVLYGVMAIWLTLWPLVLAISLFDAAASFQNNVFFAFSLMACITLMLIMYFVNSILWRRTSHKSFNPE 116  
PDEV/MEX/GTO/02/2013\_MEXICO\_(KY828994.1) MSNGSTPVDEV----IHLRNNFTNWTITLLVVLQGHYYSFVLYGVMAIWLTLWPLVLAISLFDAAASFQNNVFFAFSLMACITLMLIMYFVNSILWRRTSHKSFNPE 116  
PDEV/MEX/GTO/02/2016\_(MH004412.1) MSNGSTPVDEV----IHLRNNFTNWTITLLVVLQGHYYSFVLYGVMAIWLTLWPLVLAISLFDAAASFQNNVFFAFSLMACITLMLIMYFVNSILWRRTSHKSFNPE 116





## V.V.I Predicción del modelo 3D de la proteína M

Se realizó una búsqueda de proteínas homólogas de estructura proteica 3D conocida. BLASTP identificó la proteína cadherin 1 de *Mus musculus*, que mostró una identidad del 46.43%, pero solo una cobertura del 12% (28 aminoácidos) de la secuencia de la proteína M de longitud completa. PSI-BLAST identificó la secuencia mencionada anteriormente y 12 secuencias adicionales con porcentajes de identidad inferiores al 40% y una cobertura del 18-39% de la región carboxilo terminal de la proteína M. SWISS-MODEL generó tres modelos de la proteína CV777 M; los modelos se generaron a partir de NAD quinasa 1, NSP10 de SARS-CoV-2 / SARS-CoV y una proteína miembro de la superfamilia 9 del ligando del factor de necrosis tumoral, con porcentajes de identidad del 15, 27 y 24% respectivamente. La cobertura fue inferior al 34%, y el mejor valor QMEAN de todos los modelos fue -3.70 (QMEAN o Análisis energético del modelo cualitativo, es una función de puntuación compuesta que estima la calidad del modelo local y global al describir los principales aspectos geométricos de las estructuras de las proteínas) (135). Por lo tanto, en el presente trabajo la construcción del modelo 3D predicho de la proteína VDEP M se realizó utilizando el proyecto CAMEO (Continuous Automated Model EvaluatiOn). Inicialmente, se utilizaron los siguientes programas: IntFOLD5, SPARKS X, Phyre 2 y Raptor X. Sin embargo, algunos de los modelos construidos no presentaban estructura secundaria definida o estructura alguna en las regiones transmembranal y globular de la proteína M de los coronavirus, y los valores de QMEAN oscilaron entre -14.60 a -5.91. En consecuencia, esos modelos 3D no se consideraron (**Ilustración S1,A-D**).

Los siguientes modelos se construyeron utilizando el servidor Robetta (parte del proyecto CAMEO), con modelado comparativo RosettaCM. Se obtuvieron cinco modelos diferentes de la proteína CV777 M, con valores de QMEAN entre -0.86 y -3.41 (**Ilustración S2.pdf**). El mejor valor de QMEAN correspondió al modelo número 1. Este modelo se comparó con los modelos creados con los programas mencionados anteriormente (**Tabla 1**). También se creó un modelo 3D predicho de la proteína M de PEDV/MEX/MICH/01/2013 (llamado 2013MMV en este trabajo). Se trata de una cepa mexicana aislada en 2013 en Michoacán, México; tiene 6 diferencias de aminoácidos en comparación con la proteína CV777 M. De los 5 modelos generados a partir de ambas proteínas M, el modelo 1 se seleccionó en función de su valor QMEAN y su arquitectura proteica predicha, incluida la de los dominios transmembranal y globular (**Ilustración S3**). Las **ilustraciones 15A y B** muestran los modelos 3D predichos de la proteína CV777 M y la proteína 2013MMV M, respectivamente. Los modelos tienen 3 hélices  $\alpha$  que atravesarían la membrana, con regiones de bucle dentro y fuera de la envoltura viral, y un dominio globular formado principalmente por cadenas  $\beta$  antiparalelas. Los modelos 3D generados en Robetta se construyeron a partir de varias plantillas, la más importante de las cuales fue la proteína ORF3 del SARS-CoV-2 (6xdc) dato obtenido por el servidor Dali (**Tabla S1 y Ilustración S4**).

Tabla 1. Comparación de los mejores valores de QMEAN de los diferentes modelos de la proteína M realizados con diferentes programas.

SWISS Model	SPARKS X	Phyre 2	Raptor X	Robetta	IntFOLD
-3.51	-5.91	-14.60	-8.22	-0.86	-6.62

### V.VII Validación del modelo 3D predicho de la proteína M

Los modelos 3D predichos de la proteína M de CV777 y la proteína M de 2013MMV se evaluaron con QMEANDisCo y QMEANBrane, que arrojaron valores de QMEANDisCo de 0.59 y 0.57 respectivamente (Modelo 1 en ambas proteínas) (Ilustración S2, S3). El modelo 3D de la proteína M de CV777 y el modelo 3D de la proteína M de 2013MMV se superpusieron usando el programa UCSF Chimera, mostrando que el valor de RMSD era 12.195 Å y que había diferencias en la topología de la proteína correspondiente a la región transmembranal (1-99 aminoácidos). Las principales diferencias estaban en la región globular, que involucra a los últimos 26 aminoácidos de la proteína M de 2013MMV, que se plegaron con una orientación diferente (Ilustración 15C).

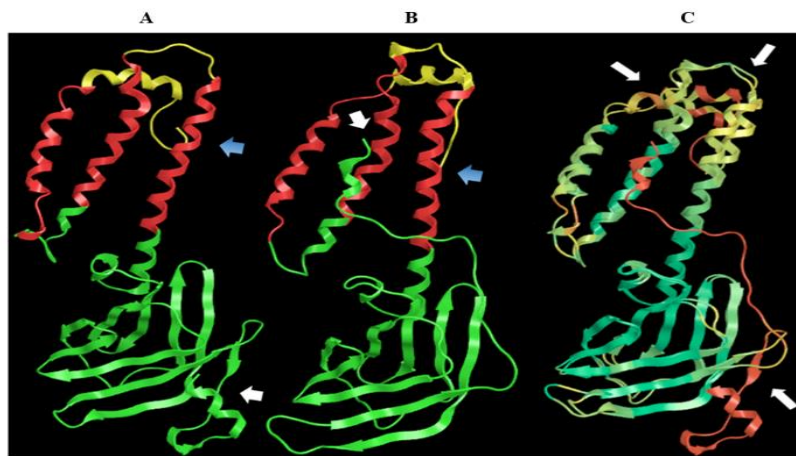


Ilustración 15. Modelos 3D predichos de las proteínas CV777 M y 2013MMV M, generados con Robetta. A) Modelo 1 de la proteína M de CV777 B) Modelo 1 de la proteína M de 2013MMV. La región transmembranal se muestra en rojo. Las regiones fuera de la membrana viral se muestran en amarillo. La región globular se muestra en verde. Las flechas azules apuntan al extremo N-terminal y las flechas blancas apuntan al extremo C-terminal. C) Superposición de los modelos 3D de las proteínas M de CV777 y M de 2013MMV. Las áreas rojas corresponden al valor RMSD más alto, mientras que las áreas verdes corresponden al valor RMSD más bajo. Las flechas blancas indican los sitios con el RMSD más alto, correspondientes a las zonas sin estructura definida y a los extremos N- y C-terminales.

La estereoquímica de ambos modelos se evaluó mediante el programa PROCHECK (**Tabla 2**). Un buen modelo predicho debería incluir más del 90% de los ángulos phi y psi en las regiones más favorecidas. El modelo de proteína M de CV777 mostró un valor de 89.2%, mientras que el modelo de proteína M de 2013MMV mostró un valor de 88.1%.

**Tabla 2.** Estadísticas de la gráfica de Ramachandran de los modelos M de CV777 y M de 2013MMV.

Gráfico estadístico de Ramachandran	Modelo 1 M CV777		Modelo 2013MMV	
	No. de residuos	%	No. de Residuos	%
Regiones más favorables	181	89.2*	178	88.1*
Regiones adicionales permitidas (a, b, l, p)	20	9.9	22	10.9
Regiones generosamente permitidas (~a, ~b, ~l, ~p)	1	0.5	0	0.0
Regiones no permitidas (XX)	1	0.5*	2	1*
Residuos no glicina y no prolina	203	100	202	100
residuos finales (excl.. Gly and Pro)	2		2	
Residuos de glicina	14		14	
Residuos de prolina	7		7	
Número total de residuos	226		225	
<b>G-Factors</b>				
Parámetros	Score	Promedio Score	Score	Promedio Score
Ángulos dihedros	-0.07		-0.08	
Phi-psi distribución	0.33		0.27	
Chi1-chi2 distribución	0.25		0.17	
Chi1 único	0.52		0.71	
Chi3 y chi4	0.05		0.07	
Omega		0.11		0.11
<b>Fuerzas covalentes de la cadena</b>				
Longitudes de enlace de la cadena	0.48		0.51	
Ángulos de enlace de la cadena	0.09		-0.18	
		0.25		0.11
<b>MEDIA GENERAL</b>		0.18		0.13

Los factores G proporcionan una medida de cuán inusual o fuera de lo común es una propiedad.

Valores por debajo de -0.5\* - inusuales

Valores por debajo de -1.0\*\* - muy inusuales

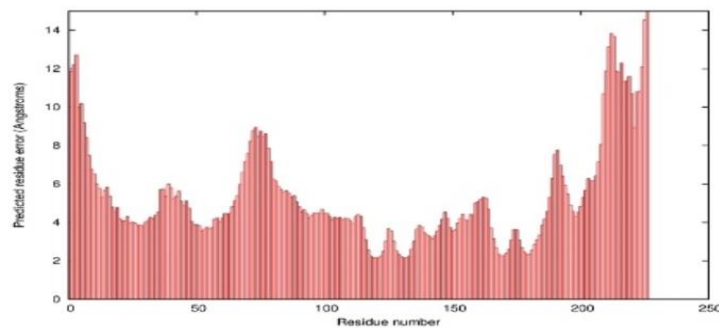
La media general es el promedio de los factores G analizados para determinar si una estructura es inusual o muy inusual.

Se utilizó el servidor web ModFOLD6 para estimar la calidad de los modelos 3D. El modelo de proteína M de CV777 tuvo una confiabilidad media, una probabilidad de 1/20 de ser un modelo incorrecto y una puntuación de calidad del modelo de 0.3855 (**ilustración 16**) El modelo de proteína M 2013MMV tuvo una confiabilidad media y un puntaje de calidad del modelo de 0.3931 (**ilustración 16**). Los puntajes de calidad del modelo global oscilan entre 0 y 1. En general, los puntajes inferiores a 0.2 indican que puede haber dominios modelados incorrectamente y los puntajes superiores a 0.4 generalmente indican modelos más completos y confiables (119).

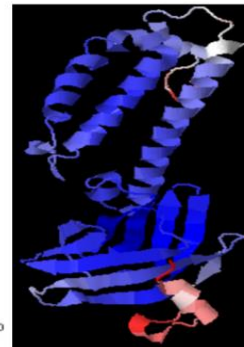
A

Modelo	Confianza y P-value	Puntuación de calidad del modelo global
Modelo 1 M CV777	Medio 1.161 E-2	0.3855
Modelo 1 2013MMV	Alto 8.473 E-3	0.3931

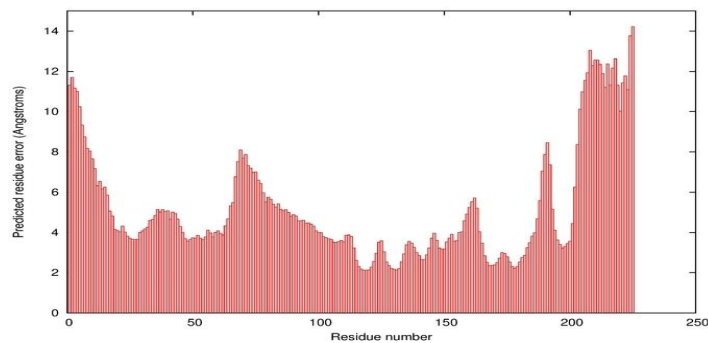
B



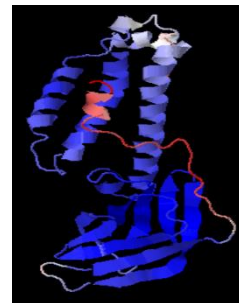
C



D



E



**Ilustración 16. Evaluación del modelo 3D de proteína M CV777 y proteína M 2013MMV utilizando el programa ModFOLD.** A) Resumen general de los resultados de ModFOLD, confiabilidad del modelo, valor p y valor global de la calidad del modelo. B) Tabla del error en Å por residuo predicho con respecto a los residuos del modelo de proteína M de CV777. C) Imagen del modelo 3D de proteína M de CV777 con vistas a la precisión por residuo. La alta precisión se resalta en azul y la baja precisión en rojo. D) Tabla de error en Å por residuo predicho con respecto a los residuos del modelo 3D de proteína M DE 2013MMV. E) Imagen del modelo 3D de proteína M de 2013MMV con vistas a la precisión por residuo. Alta precisión en azul y poca precisión en rojo.

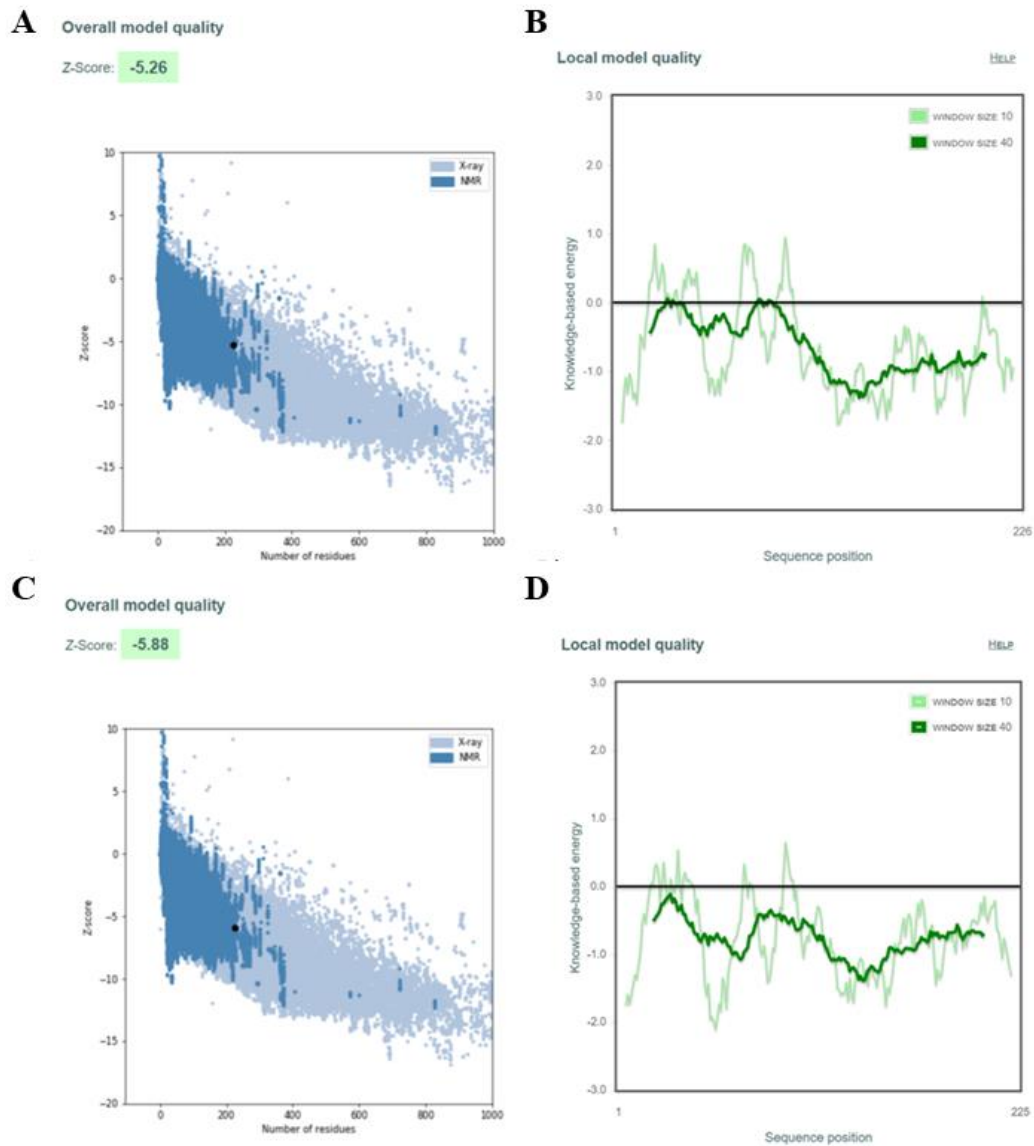
La evaluación de los modelos 3D de la proteína M de CV777 y la proteína M de 2013MMV en MolProbity (**Tabla 3**) arrojó puntuaciones de 1.34 y 1.37 respectivamente, lo que indica una estereoquímica aceptable (aceptable < 2.0, no aceptable > 2.0). Los valores obtenidos de choques estéricos o colisiones fueron 2.21 y 3.05, respectivamente, y ambos modelos tuvieron 0% de rotámeros malos. El modelo de proteína M de CV777 tenía 95.09% de sus enlaces en las áreas favorecidas en el gráfico de Ramachandran, mientras que el modelo de proteína M de 2013MMV tenía 95.96%. Sin embargo, el valor ideal es del 98%.

**Tabla 3. Evaluación usando el servidor MolProbity de las estructuras proteicas M de CV777 y M de MVM2013 generados en Robetta.**

Modelo de la proteína M CV777	Parámetros específicos	Valores obtenidos		Valores esperados
Contactos de todos los átomos	puntuación de choque	2.21		
Geometría de proteínas	rotámeros pobres	0	0.00%	Ideal: <0.3%
	rotámeros favoritos	194	98.48%	Ideal: >98%
	Valores atípicos de Ramachandran	1	0.45%	Ideal: <0.05%
	Valores favorecidos de Ramachandran	213	95.09%	Ideal: >98%
	Distribución Rama Z-score	0.63 +/- 0.55		Ideal: abs(Z score) <2
	Valor MolProbity	1.34		
	C $\beta$ desviaciones >0.25Å	0	0.00%	Ideal: 0
	Enlaces malos	4/1853	0.22%	Ideal: 0%
	Ángulos malos	7/2539	0.28%	Ideal: <0.1%
Modelo M 2013MMV	Parámetros específicos	Valores obtenidos		Valores esperados
Contactos de todos los átomos	puntuación de choque	3.05		
puntuación de choque	Rotámeros pobres	0	0.00%	Ideal: <0.3%
	Rotámeros favorecidos	195	99.49%	Ideal: >98%
	Valores atípicos de Ramachandran	1	0.45%	Ideal: <0.05%
	Valores favorecidos de Ramachandran	214	95.96%	Ideal: >98%
	Distribución Rama Z-score	-0.24 +/- 0.56		Ideal: abs(Z score) <2
	Valor MolProbity	1.37		
	C $\beta$ desviaciones >0.25Å	0	0.00%	Ideal: 0
	Enlaces malos	1/1846	0.05%	Ideal: 0%
	Ángulos malos	11/2530	0.43%	Ideal: <0.1%

Tanto el modelo de proteína M de CV777 como el modelo de proteína M 2013MMV se evaluaron en ProSA web para obtener una puntuación de calidad general. Los valores Z fueron -5,26 y -5,88, respectivamente (**ilustración 17A, 17C**). Estos valores están dentro del rango esperado para proteínas de 225-226 aminoácidos. El gráfico de energía muestra que en ambos modelos los primeros 100 aminoácidos en la región transmembranal tenían el nivel de energía más alto. Sin embargo, ninguno de los residuos tiene valores positivos elevados, lo que se ve penalizado por el programa (**ilustración 16B, 16D**), lo que sugiere que es tiene una calidad de modelado aceptable.

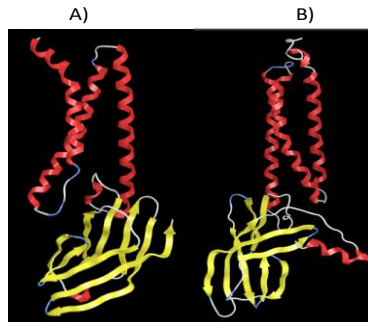
Ilustración 17. Evaluación en web ProSA del modelo 3D de la proteína M de CV777 y M 2013MMV. Los



modelos se generaron con Robetta. A) Tabla de los valores Z del modelo 3D de la proteína M de CV777. B) Tabla de energía del modelo 3D de proteína M CV777. C) Tabla de los valores Z del modelo 3D de la proteína MMV M de 2013. D) Tabla de energía del modelo 3D de proteína M de 2013MMV.

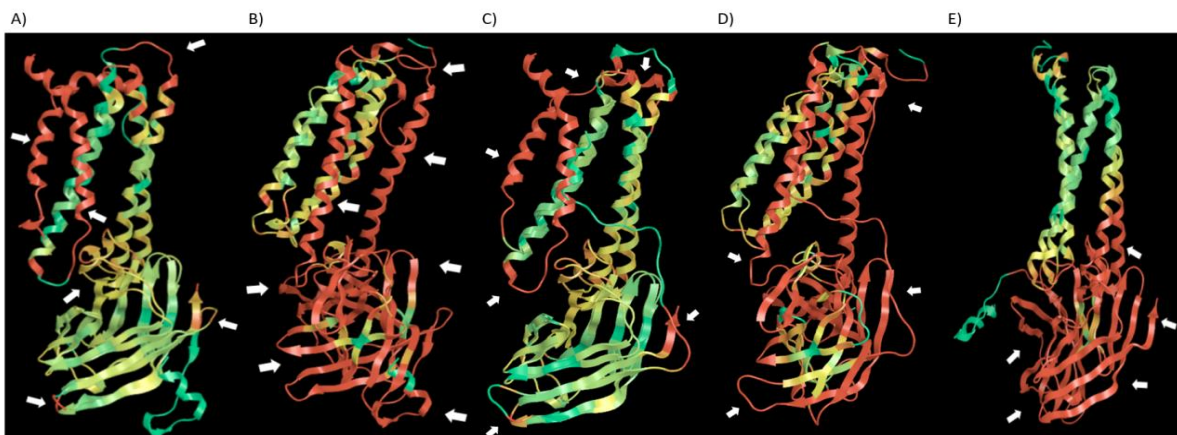
### V.VII.I Comparación de modelos 3D de proteína M predichos de modelos de proteína M vDEP y SARS-CoV-2

Como resultado de la pandemia de SARS-CoV-2, DeepMind (AlphaFold) y el laboratorio de Michael Feig (Feig lab) generaron modelos 3D de la proteína M del virus SARS-CoV-2 (ilustración 18).



**Ilustración 18. Modelos de la proteína M del SARS-CoV-2.** A) Modelo construido con AlphaFold. B) Modelo de la proteína M construido por FeigLab (136)(124).

AlphaFold utiliza una red neuronal para predecir las distancias entre pares de residuos en una proteína. Esta información permite construir un potencial de fuerza media que puede describir la forma de una proteína. El potencial resultante se optimiza mediante un algoritmo de descenso de gradiente simple para determinar las estructuras de las proteínas sin procedimientos de muestreo complejos [15]. El laboratorio de Feig Lab ha generado modelos iniciales basados en aprendizaje automático utilizando trRosetta. A continuación, los modelos se sometieron a un protocolo de refinamiento basado en simulación de dinámica molecular para maximizar la precisión del modelo [16]. En comparación, los modelos SARS-CoV-2 y los modelos vDEP mostraron patrones de plegado global similares (Ilustración 19).



**Ilustración 19. Modelos de la superposición de la proteína M de SARS-CoV-2. Superposición de los modelos de proteínas 3D M.** A) Modelo de proteína M de CV777 y M de SARS-CoV-2 M obtenido por AlphaFold. B) Modelo de proteína M de CV777 y M de SARS-CoV-2 M obtenido por FeigLab. C) Modelo de proteína M mexicana 2013MMV y M de SARS-CoV-2 obtenida por AlphaFold. D) Modelo de proteína M mexicana de 2013MMV y M de SARS-CoV-2 obtenida por FeigLab. E) Modelos de la proteína M de SARS-CoV-2 obtenidos por AlphaFold y FeigLab respectivamente.

La proteína M de SARS-CoV-2 tiene un 33% de identidad con la proteína CV777 M y un porcentaje de cobertura del 94% según BLAST. Los modelos de proteína M de SARS-CoV-2 se evaluaron con los programas web QMEAN, QMEANDisCo, MolProbity, ModFold, PROCHECK y ProSA. Los resultados se compararon con los de los modelos 3D de las proteínas M de CV777 y M de 2013MMV generados por Robetta (**Tabla 4**). El análisis mostró que los parámetros de calidad de los modelos de proteína vDEP M eran comparables a los parámetros de calidad de los modelos validados de la proteína M de SARS-CoV-2. Se observaron diferencias estructurales y topológicas entre los modelos de proteína M del vDEP generados por Robetta y los modelos de proteína M de SARS-CoV-2 del laboratorio AlphaFold y Feig cuando se superpusieron los modelos. Las imágenes superpuestas se pueden ver en la ilustración 17. Los valores de RMSD del modelo 3D de la proteína M de CV777 y los modelos de laboratorio AlphaFold y Feig fueron 6.0 y 14.1 Å respectivamente. Por el contrario, la superposición del modelo 3D de proteína M de 2013MMV con el modelo de proteína M de SARS-CoV-2 de AlphaFold y el modelo de proteína M de SARS-CoV-2 del laboratorio Feig mostró valores de RMSD de 8.120 y 15.716 Å respectivamente. La superposición del modelo AlphaFold con el modelo de laboratorio de Feig mostró un RMSD de 12.6 Å. La principal diferencia entre todos los modelos se observó en las hélices  $\alpha$  del dominio transmembranal. Aunque el patrón de plegamiento se conservó en la región globular, se observaron diferencias en la orientación. Los datos obtenidos demostraron que los modelos 3D de las proteínas M de CV777 y M de 2013MMV generados por Robetta son de buena calidad general. Los modelos de proteína M de VDEP generados por Robetta parecen ser de calidad confiable en comparación con los valores obtenidos en la evaluación de los modelos de proteína M de SARS-CoV-2 del laboratorio AlphaFold y Feig. Además, la plataforma Robetta proporciona una forma fácil y gratuita de generar modelos, sin la necesidad de instalar programas.

**Tabla 4. Resumen de los valores de los programas utilizados para evaluar la calidad de los modelos 3D.** Comparación entre los modelos 3D de proteína M de CV777 y proteína M de 2013MMV generada en Robetta, proteína M SARS-CoV-2 de AlphaFold y FeigLab.

	CV777 Robetta	2013MMV Robetta	SARS-CoV-2 AlphaFold	SARS-CoV-2 FeigLab
<b>QMEAN</b>	-1.58	-1.76	-0.86	-2.89
<b>QMEANDisCo</b>	0.59	0.57	0.49	0.45
<b>PROCHECK</b>	89.2	88.1	94.8	88.6
<b>ModFOLD</b>	Medio 1.161E-2 (0.3855)	Alto 9.549E-3 (0.3899)	Alto 7.113E-3 (0.3998)	Medio 4.804E-2 (0.3510)
<b>ProSA web</b>	-5.26	-5.88	-3.5	-3.87
<b>MolProbity</b>	1.34 (2.21, 0.00, 95.09)	1.37 (3.05, 0.00, 95.96)	1.32 (1.57, 2.45, 97.38)	0.74 (0.28, 0.53, 96.36)

## V.VII Identificación de epítomos en la proteína M del vDEP

### V.VII.I Identificación de epítomos de células B lineales en el modelo de proteína vDEP M

Los epítomos estructurales en el modelo 3D predicho de la proteína M de vDEP se analizaron y compararon con epítomos de células B lineales. Los epítomos de células B de la proteína M CV777 y M 2013MMV se predijeron utilizando la predicción del epítomo lineal Bepipred 1.0 y el predictor secuencial del epítomo de células b Bepipred 2.0. Ambos programas encontraron cuatro péptidos en el modelo de la proteína M de CV777; los péptidos RSVNASSGTG y KHGDYSAVSNPSAVLT fueron predichos por ambos programas (**Tabla 5**). En el modelo de proteína M de 2013MMV, los programas predijeron 3 y 4 péptidos respectivamente. Una evaluación adicional con otros programas mostró tres péptidos por encima del umbral de 1 para los modelos de proteínas M de CV777 y M de 2013MMV. Se realizó un análisis de conservación de epítomos (%) con 31 secuencias de proteínas M de diferentes cepas de vDEP aisladas en México, Estados Unidos y otros países. Entre las secuencias analizadas se observó un alto nivel de conservación de los epítomos identificados en las proteínas M de 2013MMV y CV777. Sin embargo, las secuencias de algunos epítomos de la proteína M de 2013MMV estaban más altamente conservadas que las de la proteína M de CV777 (**Tabla 5**).

**Tabla 5. Epítomos continuos de células B de la proteína M de CV777.**

Epítomos continuos B-cell	Posición	Largo	Chou	Emini	Karplus	Kolaskar	Parker	Conservación %	Identidad % Mínimo-máximo
<b>Modelo CV777</b>									
<b>Bepipred 1.0</b>									
MSNGSIPV	1-8	8	1.217	1.000	1.065	0.999	2.264	90.80	37.50-100.00
SFNPETDAL	111-119	9	1.108	1.000	1.051	0.951	3.781	100.00	100.00-100.00
RSVNASSGTG	185-194	10	1.207	1.000	1.038	0.993	4.561	94.25	90.00-100.00
KHGDYSAVSNPSAVLT	202-217	16	1.133	1.000	1.004	1.057	2.986	16.09	93.75-100.00
<b>Bepipred 2.0</b>									
SIPVDEVIEHLRNWNF	5-20	16	0.863	1.00	0.973	1.046	0.497	12.64	56.25-100.00
YKVATGVQVSQL	154-165	12	0.910	1.000	0.989	1.106	1.960	100.00	100.00-100.00
GRVGRSVNASSGTG	181-194	14	1.142	1.000	1.026	1.011	3.807	93.10	92.86-100.00
KHGDYSAVSNPSAVLTDSEKV	202-222	21	1.078	1.000	1.006	1.055	2.883	14.94	90.48-100.00
<b>Modelo 2013MMV</b>									
<b>Bepipred 1.0</b>									
SFNPETDAL	110-112	9	1.108	1.000	1.051	0.951	3.781	100.00	100.00-100.00
RSVNASSGTG	184-193	10	1.207	1.000	1.038	0.993	4.561	94.25	90.00-100.00

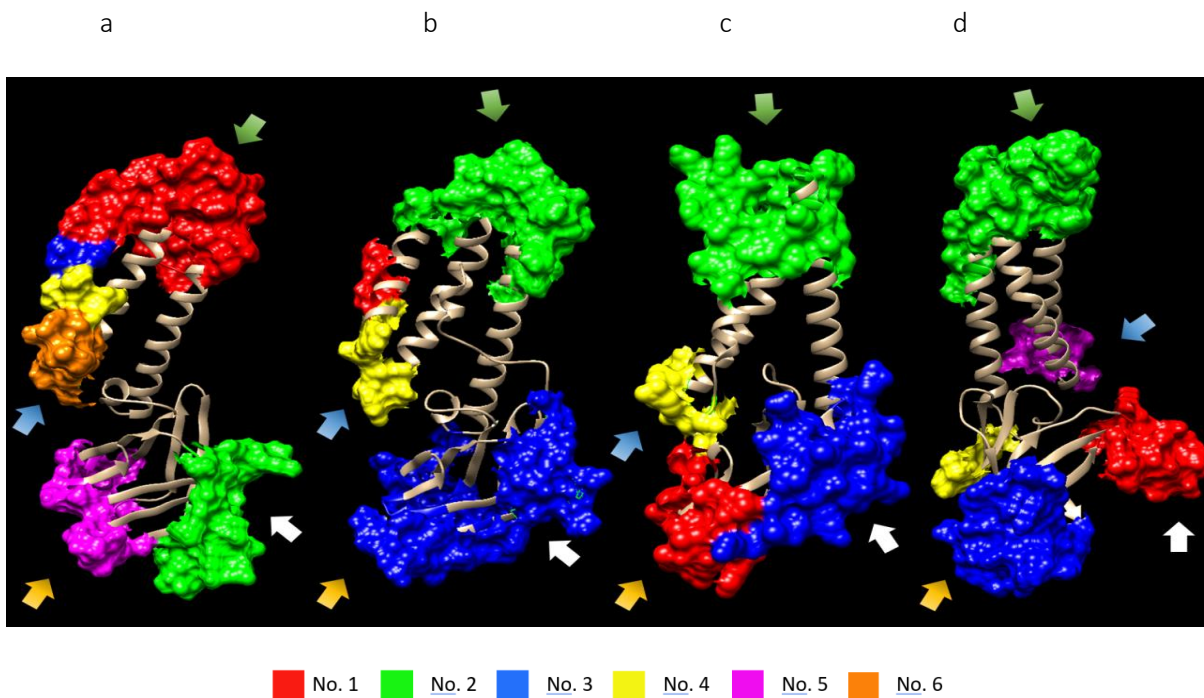
KHGDYSAVSNPS SVLTD	201-217	17	1.172	1.000	1.019	1.056	3.169	83.91	94.12- 100.00
Bepipred 2.0									
VDEVIQHLRNW NF	7-19	13	0.891	1.00	0.952	1.042	0.086	78.16	61.54- 100.00
YKVATGVQVSQL	153-164	12	0.910	1.000	0.989	1.106	1.960	100.00	100.00- 100.00
RVGRSVNASSGT G	181-193	13	1.153	1.000	1.026	1.007	3.965	94.25	92.31- 100.00
KHGDYSAVSNPS SVLTDSEKV	201-221	21	1.129	1.000	1.023	1.052	3.176	82.76	90.48- 100.00

Los epítomos sombreados son exclusivos de la proteína M 2013MMV

Chou, **Promedio:** 0.440 **Mínimo:** 0.228 **Máximo:** 0.637 **Línea límite:** 0.5000  
Emini, **Promedio:** 1.000 **Mínimo:** 0.083 **Máximo:** 4.505 **Línea límite:** 1.00  
Karplus, **Promedio:** 0.967 **Mínimo:** 0.846 **Máximo:** 1.090 **Línea límite:** 0.967  
Kolaskar, **Promedio:** 1.063 **Mínimo:** 0.873 **Máximo:** 1.242 **Línea límite:** 1.063  
Parker, **Promedio:** -0.429 **Mínimo:** -7.643 **Máximo:** 5.529, **Línea límite:** -0.429  
**Identidad % Línea límite: 100%**

#### V.VII.II Predicción de epítomos discontinuos de células B del modelo de proteína M de vDEP

La predicción de epítomos discontinuos a partir de modelos 3D de proteína M se realizó utilizando los programas ElliPro y DiscoTope. Se hizo la misma predicción con los modelos de proteína M de SARS-CoV-2 del laboratorio AlphaFold y Feig. ElliPro predijo 6 epítomos en el modelo de proteína M de CV777, 4 en el modelo de proteína M de 2013MMV, 4 en el modelo de proteína M de SARS-CoV2 de Alpha fold y 5 en el modelo de proteína M de SARS-CoV2 de Feig lab (**Tabla 6**) (**Ilustración 20**). Los epítomos predichos se ubicaron en áreas con estructura similar. En la Ilustración 18, las regiones coloreadas corresponden a epítomos discontinuos, ordenados desde la puntuación más alta (rojo) hasta la puntuación más baja (naranja). El programa DiscoTope predijo 15 residuos en el modelo CV777 y 14 residuos en el modelo de proteína M de 2013MMV con tendencia a formar parte de un epítomo, mientras que 14 y 24 residuos se predijeron en el modelo de proteína M de SARS-CoV-2 de AlphaFold y el Modelo de proteína M de SARS-CoV-2 del laboratorio Feig, respectivamente (**Tabla 7**). Las **ilustraciones 20 y 21** muestran los residuos en los modelos. Las flechas de colores indican residuos que se encuentran en regiones estructurales similares.



**Ilustración 20. Epítomos discontinuos predichos en ElliPro a partir de modelos de proteína CV777, 2013MMV y SARS-CoV-2 M.** Los modelos se generaron con AlphaFold y FeigLab. Los resultados de ElliPro muestran patrones de proteínas con epítomos discontinuos resaltados con diferentes colores. Las flechas con el mismo color indican epítomos discontinuos que pertenecen a la misma región marco en los modelos. A) Proteína M de CV777. B) Proteína M de 2013MMV. C) Proteína M de SARS-CoV-2 de AlphaFold. D) proteína M de SARS-CoV-2 de FeigLab. Los residuos coloreados para cada epítomo se indican en la tabla 6. **Los colores solo diferencian entre epítomos**

**Tabla 6. Predicción de epítomos discontinuos con ElliPro a partir del modelo 3D de proteína M de CV777, modelo 3D de proteína M de 2013MMV y modelos de proteína M de SARS-CoV-2 de AlphaFold y FeigLab. INDICAR AL FINAL DE LA TABLA EL INTERVALO DE VALORES SIGNIFICATIVOS, ¿Cuál es el valor significativo?**

Modelo 3D de la proteína M CV777			
No.	Residuos	Número de residuos	Valor
1	M1, S2, N3, G4, S5, I6, P7, V8, D9, E10, V11, I12, E13, H14, L15, R16, W18, N19, F20, T21, W22, N23, I24, L64, F65, D66, A67, W68, A69, S70, F71, Q72, V73, N74, W75, V76, F77, F78, A79, F80, S81, I82, L83, M84	45	0.741
2	A136, P137, T138, R200, S201, K202, H203, G204, D205, Y206, S207, A208, V209, S210, N211. P212, S213, A214, V215, L216, T217, D218, S219, E220, K221, V222, L223, H224, L225, V226	30	0.719
3	T27, I28, L29	3	0.698
4	V31, V32, Q29	3	0.684
5	S123, V124, M125, G126, R127, L144, L145, S146, G147, T148, A157, T158, G159, V160, Q161, V162, S163, Q164, L165, S186, V187, N188, A189, S190, S191, G192, T193	27	0.674
6	Y35, G36, H37, Y38, K39, Y40, S41, V42, F43, L44	10	0.645

Proteína M 2013MMV			
No.	Residuos	Número de residuos	Valor
1	I27, L28, V31	3	0.803
2	S1, N2,G3, F4,I5,P6, V7, D8, E9,V10, I11, Q12, H13, L14, R15, W17, N18, F19, T20,F64, D65, A66, W67, A68, S69,F70, Q71, V72, N73, W74, V75, F76, F77, A78, F79, L82, M83	37	0.747
3	H106, S122, V123, M124, G125, R126, Q127, G134, A136, P136, T137, G138, T142, L143, L144, S145, G146,T147, A156, T157, G158, V159, Q160, V161, S162, Q163, L164, P165, N166, F167, I177, V178, G180, R181, V182, G183, R184, S185, V186, N187, A188, S189, S190, G191, T192, G193,W194, A195, F196, V198, S200, K201, H202, G203, D 204, Y205, S206, A207, V208, S209.	60	0.699
4	L32, G35, Y37, K38, Y39, S40, A41, F42, L43.	9	0.548
Proteína M SARS-CoV-2 M de AlphaFold			
No.	Residuos	Número de residuos	Valor
1	V143, I144, G147, H148, L149, H155, G157, R158, C159, D160, I161, K162, D163, L164, R186, V187, A188, G189, D190, S191, G192, F193	22	0.699
2	E12, L13, K14, K15, L16, L17, E18, Q19, W20, N21, L22, V23, I24, G25, F26, L27, T30, A63, C64, F65, L67, A68. A69, V70, Y71, R72, I73, N74, W75, T77, G78, G79, I80, A81, I82, A83, M84, C86	28	0.676
3	R101, A104, R105, E135, S136, E137, L138, V139, E167, S173, R174, T175, L176, S177, Y178, K180, L181, G182, A183, S184, Q185, A195, S197, R198, Y199, R200, I201, G202, N203	29	0.676
4	A40, N41, N43, R44, F45, L46, I49	7	0.573
Proteína M SARS-CoV-2 de FeigLab			
No.	Residuos	Número de residuos	Valor
1	S173, R174, K205, L206, N207, T208, D209, H210, S211, S212, S213, S214, D215, N216, I217, A218, L219, L220, V221, Q222	20	0.747
2	M1, A2, D3, S4, N5, G6, T7, I8, T9, V10, E11, E12, L13, K14, K15, L16. L17, E18. Q19, W20, V66, L67, A68, A69, V70, Y71, R72, I73, N74, W75, I76, T77, G78, G79, I80, A81, I82, A83, C86, L87, L90	41	0.739
3	I144, L145, R146, G147, H148, G157, R158, C159, D160, I161, K162, D163, L164, P165, K166, E167, K180, L181, G182, A183, S184, Q185, R186, V187, A188, D190, S191, G192, F193, A194	31	0.703
4	R107, L124, H125, G126, T127	5	0.686
5	Q36, F37, Y39, A40, N41, N43	6	0.665

Puntaje mínimo 0.5

Distancia máxima (Angstrom) 6

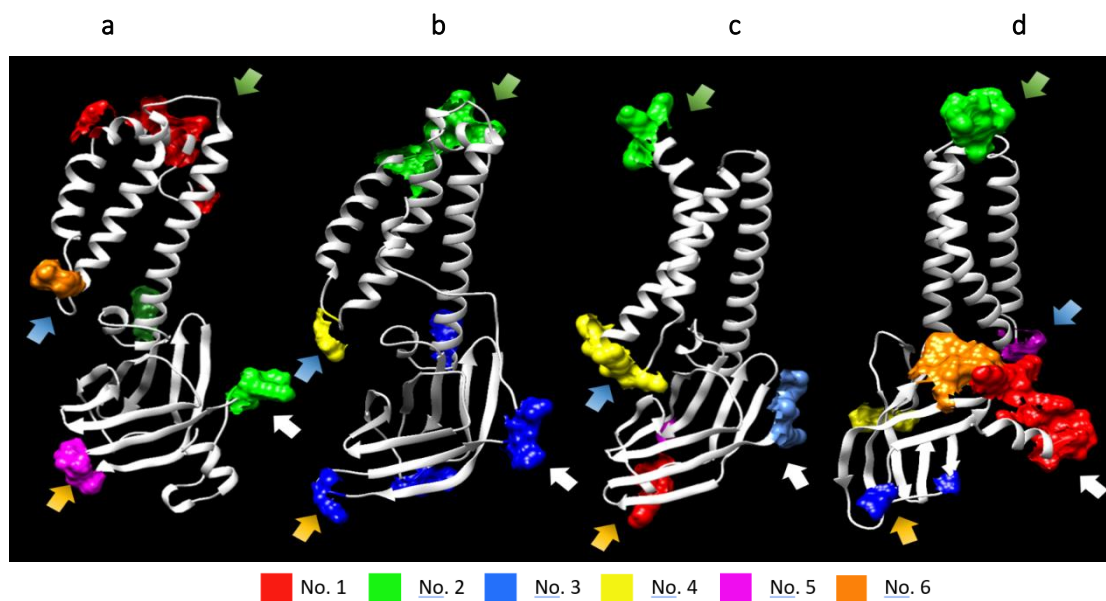


Ilustración 21. Epítomos discontinuos predichos con DiscoTope de los modelos de proteína M CV777 y 2013MMV y proteína M del SARS-CoV-2. Los últimos modelos son del laboratorio AlphaFold y Feig. Los patrones de proteínas con epítomos discontinuos, resaltados con diferentes colores, corresponden a los residuos en los resultados de DiscoTope para cada patrón. Los residuos resaltados con diferentes colores corresponden a los epítomos discontinuos en los resultados de ElliPro para cada modelo. Las flechas con el mismo color indican epítomos discontinuos que pertenecen a la misma región marco en los modelos. Los colores adicionales muestran residuos que no pertenecen a algún epítopo discontinuo en los resultados de ElliPro. A) Proteína M de CV777. B) Proteína M 2013MMV. C) Proteína M de SARS-CoV-2 de AlphaFold. D) Proteína M de SARS-CoV-2 del laboratorio Feig. Los residuos coloreados para cada epítopo se indican en la tabla 7.

Tabla 7. Predicción de epítomos discontinuos con DiscoTope de proteína M CV777 y proteína M en modelos de proteína 2013MMV, SARS-CoV-2 M de AlphaFold y FeigLab.

Modelo CV777					Modelo 2013MMV				
Residuo ID	Nombre de residuo	Número de contacto	Puntaje de propensión	DiscoTope Score	Residuo ID	Nombre de residuo	Número de contacto	Puntaje de propensión	DiscoTope Score
5	SER	11	-0.488	-5.988	12	GLN	13	-0.452	-5.988
7	PRO	11	-0.269	-5.769	15	ARG	13	-0.242	-5.769
9	ASP	9	-0.726	-5.226	16	ASN	9	1.494	-5.226
10	GLU	11	-0.492	-5.992	17	TRP	12	-0.702	-5.992
13	GLU	12	-0.083	-6.083	18	ASN	12	-1.154	-6.083
20	PHE	12	-1.389	-7.389	39	TYR	10	-1.852	-7.389
37	HIS	10	-1.703	-6.703	103	ARG	12	1.41	-6.703
38	TYR	9	-1.208	-5.708	162	SER	9	-2.048	-5.708
104	ARG	13	0.851	-5.649	163	GLN	10	-2.42	-5.649
189	ALA	9	-2.024	-6.524	189	SER	8	-1.891	-6.524
190	SER	7	-1.6	-5.1	190	SER	8	-2.125	-5.1
191	SER	9	-2.805	-7.305	202	HIS	10	-0.463	-7.305
202	LYS	9	-0.019	-4.519	203	GLY	7	0.388	-4.519
203	HIS	9	0.273	-4.227	204	ASP	7	0.211	-4.227
205	ASP	14	-0.578	-7.578					
Modelo AlphaFold					Modelo FeigLab				

Residuo ID	Nombre de residuo	Número de contacto	Puntaje de propensión	DiscoTope Score	Residuo ID	Nombre de residuo	Número de contacto	Puntaje de propensión	DiscoTope Score
11	GLU	6	-0.824	-3.824	1	MET	12	-1.188	-7.188
12	GLU	8	-1.146	-5.146	3	ASP	9	-0.396	-4.896
13	LEU	9	-1.484	-5.984	4	SER	9	-0.396	-4.896
41	ASN	10	-0.801	-5.801	5	ASN	8	-0.133	-4.133
42	ARG	10	-0.801	-5.801	6	GLY	9	-0.396	-4.896
43	ASN	9	-0.694	-5.194	7	THR	12	-1.491	-7.491
45	PHE	10	-2.508	-7.508	42	ARG	10	-0.801	-5.801
126	GLY	8	-2.206	-6.206	43	ASB	11	-1.314	-6.814
162	LYS	11	-1.441	-6.941	115	GLU	14	1.057	-5.943
163	ASP	11	-1.052	-6.552	125	HIS	8	-2.206	-6.206
200	ARG	13	-0.069	-6.569	126	GLY	7	-1.762	-5.262
201	ILE	14	-0.161	-7.161	157	GLY	10	-2.308	-7.308
202	GLY	10	0.468	-4.532	163	ASP	8	-0.444	-4.444
203	ASN	8	0.669	-3.331	200	ARG	14	-0.345	-7.345
					201	ILE	15	0.429	-7.071
					202	GLY	13	1.418	-5.082
					203	ASN	13	1.329	-5.171
					204	TYR	12	0.106	-5.894
					205	LYS	13	-0.664	-7.164
					209	ASP	11	0.823	-4.677
					210	HIS	10	0.938	-4.062
					211	SER	10	0.938	-4.062
					212	SER	11	0.301	-5.199
					213	SER	10	-0.201	-6.201

Número de contacto: El número de contacto del residuo es el número de átomos de C $\alpha$  en el antígeno dentro de una distancia de 10 Å del átomo de C $\alpha$  del residuo. Un número de contacto bajo se correlaciona con la localización del residuo cerca de la superficie o en regiones sobresalientes de las estructuras del antígeno.

Puntaje de propensión: este puntaje le informa sobre la probabilidad o tendencia de ser parte de un epítipo para ese residuo en particular. La propensión se refleja en los índices de posibilidades logarítmicas de los epítipos de aminoácidos, que se calcularon sobre un conjunto de 75 antígenos. La puntuación de propensión se calcula promediando secuencialmente los índices de probabilidades de epítipos dentro de una ventana de 9 residuos. Luego, las puntuaciones se suman en función de la proximidad en la estructura 3D del antígeno. Para cualquier residuo dado, las puntuaciones de probabilidades logarítmicas promediadas secuencialmente de todos los residuos dentro de 10 Å se suman para dar la puntuación de propensión.

Puntaje de DiscoTope: este puntaje se calcula combinando los números de contacto con el puntaje de propensión. La puntuación de DiscoTope por encima del valor de umbral indica predicciones positivas y por debajo del valor de umbral indica predicciones negativas. Valor umbral -7.7

### V.VII.III Predicción de epítipos de células T para la proteína M de vDEP.

Se realizó un acoplamiento de epítipos de células T con el alelo del complejo MHC SLA-1:0401 para complementar el perfil inmunogénico teórico de la proteína M del vDEP. La predicción de los epítipos de células T se realizó utilizando el servidor NetMHCpan 4.1 (DTU Bioinformatics). Se utilizaron tres alelos para la predicción, SLA-1: 0101, SLA-1: 0401 y SLA-1: 0801. Con respecto a la proteína M de CV777, se identificaron 40 péptidos de unión débil y 14 de unión fuerte (**Tabla 8**). Con respecto a la proteína M de 2013MMV, se obtuvieron 36 péptidos de unión débil y 16 de unión fuerte. La inmunogenicidad de los péptidos de unión fuerte se predijo utilizando el servidor de inmunogenicidad IEDB clase 1, que identificó 10 epítipos, 2 de ellos repetidos en diferentes alelos. Los 8 péptidos se utilizaron para una simulación de acoplamiento con el alelo SLA-1: 0401 (PDB ID 3QQ3) para un perfil inmunológico. Los umbrales para aglutinantes fuertes y débiles se expresan en términos de %Rank, que es el percentil de la afinidad de unión predicha en comparación con la distribución de afinidades calculada en un conjunto de péptidos naturales

aleatorios. El péptido se identificará como un aglutinante fuerte si se encuentra entre los mejores x% de los péptidos predichos, donde x% es el umbral especificado para aglutinantes fuertes (por defecto, 0.5%). El péptido se identificará como un aglutinante débil si el % de rango está por encima del umbral de los aglutinantes fuertes pero por debajo del umbral especificado para los aglutinantes débiles (por defecto 2%).

**Tabla 8. Epítomos continuos de células T de la proteína M de CV777. Indicar al final de la tabla que significan %Rank EL %Rank BA Afinidad Nm Inmunogenicidad clase 1, indicar el intervalo de valores de importancia.**

Alelo SLA-1:0101								
Péptidos de unión débil	%Rank EL	%Rank BA	Afinidad Nm	Péptidos de unión fuerte	%Rank EL	%Rank BA	Afinidad Nm	Inmunogenicidad clase 1
GSIPVDEVI	1.379	1.825	7527.49	PVDEVIEHL	0.135	1.608	7076.59	0.37683
ILWPLVLAL**	1.278	1.294	6391.21	LGAPTGVTL**	0.165	0.384	3530.72	0.12876
SLFDWASF**	0.795	0.464	3876.20	APTGVTLTL**	0.360	1.183	6126.86	0.13676
SFNPETDAL**	0.938	2.037	7930.56	SQLPNFVTY**	0.055	0.048	1310.84	0.14666
TDALLTTSV**	1.285	1.210	6194.39	PVDEVIQHL*	0.105	1.758	7386.78	0.19457
QVSQLPNFV**	1.499	2.900	9428.25	AVSNPSAVL	0.218	0.411	3653.72	-0.1696
TVAKATTI**	0.505	0.761	4916.88	SAVSNPSSV*	0.476	0.682	4683.27	-0.40609
ASSGTGWAF**	0.787	0.490	3991.62	AVSNPSSVL*	0.245	0.746	4875.56	-0.34224
YSAVSNPSA	1.404	0.869	5261.96					
SAVSNPSAV	0.621	0.580	4323.47					
AVLTDSEKV	1.561	2.739	9167.12					
VLTDSEKVL	1.988	8.192	15472.46					
TDSEKVLHL**	0.946	8.053	15354.38					
DSEKVLHLV**	1.600	5.414	12759.76					
SAFLYGVKM*	1.695	2.275	8362.81					
SVLTDSEKV*	1.383	2.249	8314.10					
VLTDSEKVL*	1.988	8.192	15472.46					
Alelo SLA-1:0401								
Péptidos de unión débil	%Rank EL	%Rank BA	Afinidad Nm	Péptidos de unión fuerte	%Rank EL	%Rank BA	Afinidad Nm	Inmunogenicidad clase 1
PVDEVIEHL	1.470	4.053	4846.79	VLQYGHYKY**	0.211	0.270	238.88	-0.10699
EVIEHLRNW	1.011	1.352	1420.59	SLFDWASF**	0.189	0.164	148.60	0.2429
TILLVVLQY**	0.972	2.606	3009.49	ASSGTGWAF**	0.226	0.126	110.35	0.2599
SVFLYGVKM	1.859	2.498	2882.55	AVSNPSSVL*	0.481	0.763	751.51	-0.34224
ASFQNVWF**	1.431	1.567	1700.35					
QVNWVFFAF**	1.889	0.978	987.97					
RLWRRTHSW**	0.689	0.582	552.91					
LTLLSGTLF	0.953	0.933	936.80					
TVAKATTI**	1.186	1.485	1599.22					
SSGTGWAFY**	0.863	0.382	339.85					
AVSNPSAVL	0.553	0.692	679.83					
LTDSEKVLH**	0.794	1.551	1682.18					
VQVSQLPNF*	1.454	2.016	2249.84					
PVDEVIQHL*	0.726	2.470	2842.62					
EVIQHLRNW*	1.597	2.071	2314.73					
Alelo SLA-1:0801								
Péptidos de unión débil	%Rank EL	%Rank BA	Afinidad Nm	Péptidos de unión fuerte	%Rank EL	%Rank BA	Afinidad Nm	Inmunogenicidad clase 1
EVIEHLRNW	0.837	1.907	6253.17	TILLVVLQY**	0.409	1.346	4685.44	-0.1274
HLRNWNFTW**	1.228	0.856	3057.97	VLQYGHYKY**	0.126	0.214	850.54	-0.10699
LVVLQYGHY**	1.508	0.767	2751.72	SLFDWASF**	0.039	0.011	117.99	0.24329
SVFLYGVKM	0.944	1.175	4166.01	RLWRRTHSW**	0.451	0.445	1647.27	0.14156
ILWPLVLAL**	1.053	0.874	3124.42	VQVSQLPNF**	0.441	0.678	2445.83	-0.28945
ASFQNVWF**	0.827	0.776	2780.60	ASSGTGWAF**	0.322	0.188	762.6	0.2599
QVNWVFFAF**	1.312	0.513	1889.30					
TLMLWIMYF**	1.126	0.569	2071.38					
LTLLSGTLF	0.969	0.649	2342.52					

SQLPNFVTV**	0.864	1.650	5612.88					
AKATTTIVY**	0.599	1.520	5225.76					
SSGTGWAFY**	1.186	0.635	2296.34					
YVRSKHGDY**	1.263	1.539	5282.61					
AVSNPSAVL	0.702	1.251	4399.64					
SAFLYGVKM*	1.087	1.256	4413.65					
AVSNPSSVL*	0.645	1.440	4974.67					
EVIQHRLRNW*	1.293	2.977	8876.16					
IQHLRNWNF*	1.807	2.405	7434.73					

Péptidos de afinidad de unión (BA) y ligandos eluidos por espectrometría de masas (EL), afinidad Nm, la afinidad nanomolar predicha se toma como la media geométrica de los resultados del modelo individual. Inmunogenicidad de clase 1, esta herramienta utiliza las propiedades de los aminoácidos, así como su posición dentro del péptido, para predecir la inmunogenicidad de un complejo peptídico MHC (pMHC). Un valor de puntaje predictivo superior a 0,5 se considera inmunogenicidad (positivo-alto), un puntaje entre 0.4 y 0.5 se considera positivo-bajo y un puntaje inferior a 0.4 se considera negativo-alto.

Se calculó la energía de unión entre los epítomos y el alelo SLA-01: 0401 para cada péptido (Tabla 9). se puede observar los epítomos de color verde dentro del alelo color blanco (Ilustración 22), los epítomos se observan en diferentes posiciones dentro del surco del alelo, lo que indica diferentes interacciones con respecto a cada epítomo, lo que resulta en diferentes valores de energía de interacción, siendo el epítomo SLFDAWASF el epítomo con la menor energía (-7.8 kcal / mol), lo que representaría la interacción más fuerte entre todos los epítomos y el alelo SLA-01: 0401, lo que significa que existe la posibilidad de que los epítomos de la proteína M del vDEP sean reconocidos por los alelos de células T.

**Tabla 9. Energía de unión de los epítomos predichos para el alelo SLA-01: 0401 usando AutoDock. Indicar al final de la tabla qué significan  $\sigma$  (kcal/mol) Conservación % Identidad %, así como el intervalo de valores de importancia. Para que pueda interpretarse.**

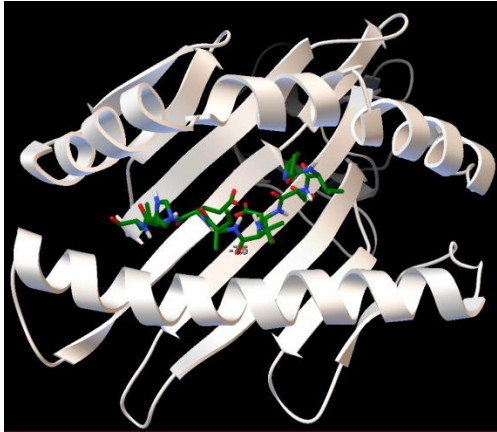
Secuencia de péptidos	Energía de unión (kcal/mol)	$\sigma$ (kcal/mol)	Conservación %	Identidad %
PVDEVIEHL	-7.5	0.410	14.94	66.67-100.00
LGAPTGVTL	-5.5	0.161	100.00	100.00-100.00
APTGVTLTL	-7.0	0.150	100.00	100.00-100.00
SQLPNFVTV	-7.0	0.181	95.40	88.89-100.00
PVDEVIQHL	-6.4	0.206	78.16	66.67-100.00
SLFDAWASF	-7.8	0.300	98.85	77.78-100.00
ASSGTGWAF	-6.3	0.259	95.40	88.89-100.00
RLWRRTHSW	-6.9	0.238	100.00	100.00-100.00

Desviación estándar  $\sigma$  (kcal/mol) es la desviación de las energías resultantes en el programa.

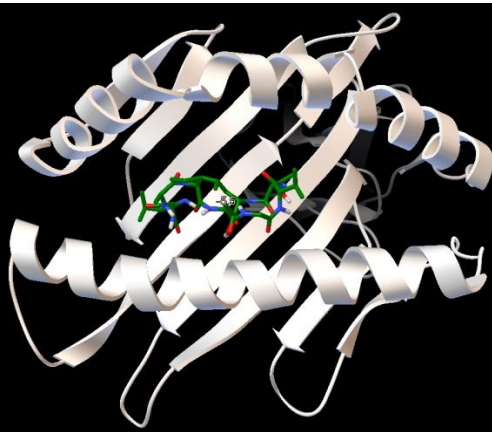
El porcentaje de conservación se refiere a la conservación que tiene la secuencia peptídica analizada con respecto al total de secuencias de la proteína M de diferentes cepas analizadas.

El porcentaje de identidad representa los valores de identidad (mínimo y máximo) que existe entre el péptido analizado y todas las secuencias de la proteína M de diferentes cepas analizadas.

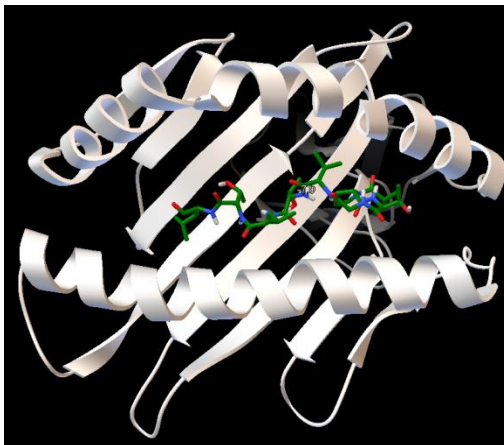
A



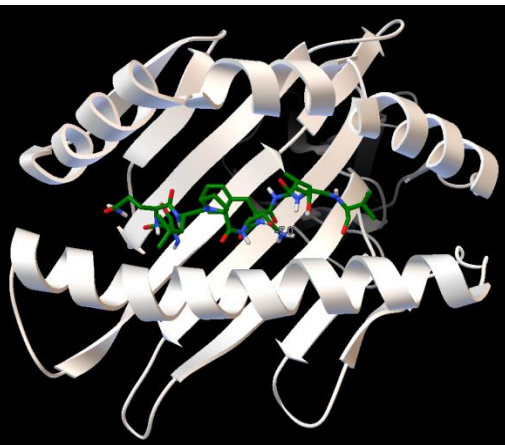
B



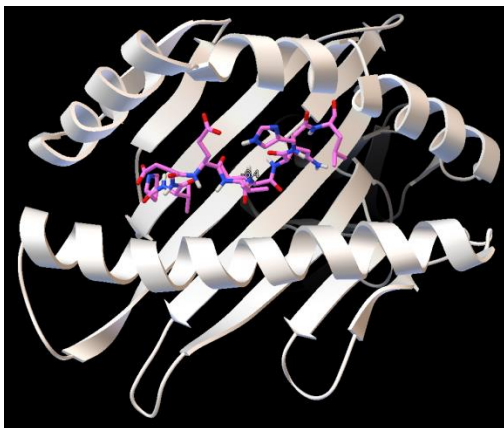
C



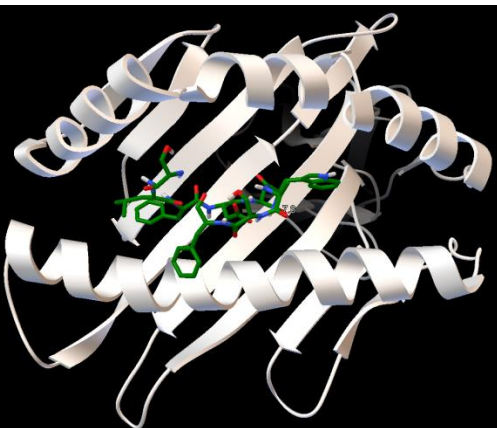
D



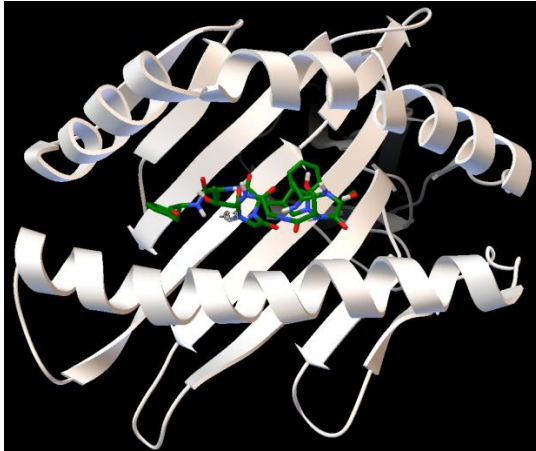
E



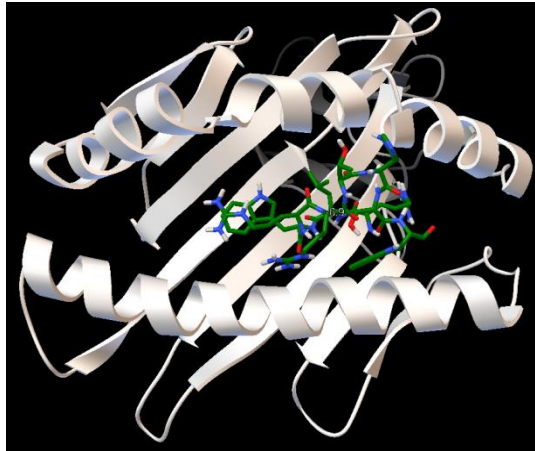
F



G

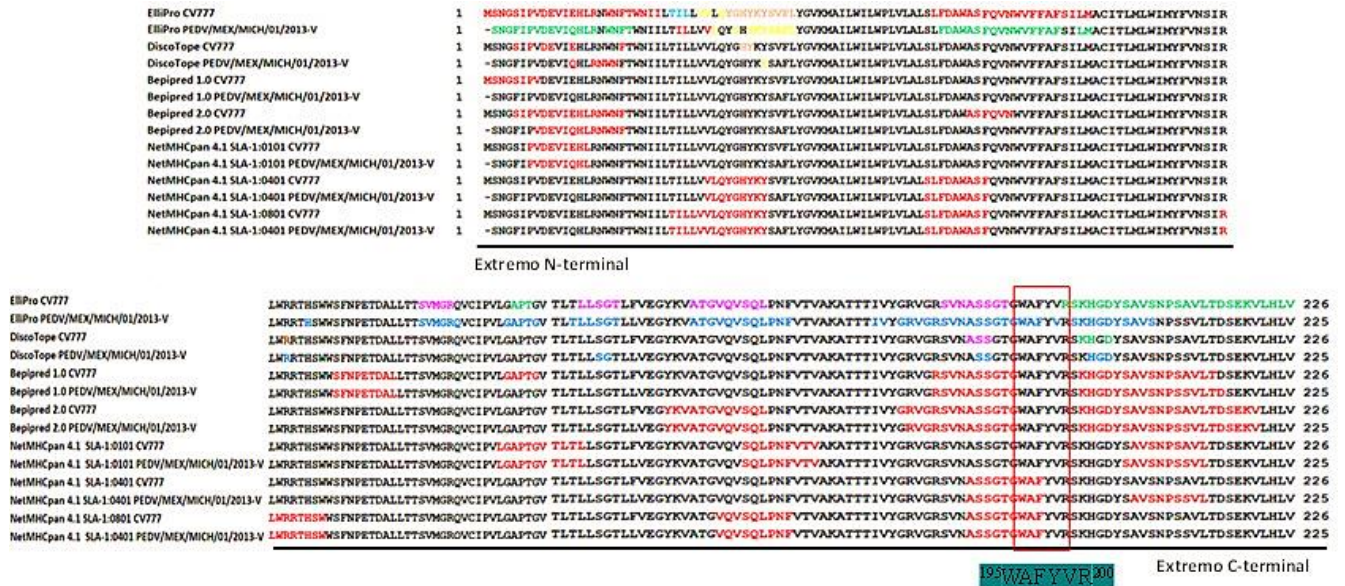


H



**Ilustración 22. Patrones de unión de los epítomos de la proteína M con SLA-01: 0401.** Los epítomos predichos se muestran como líneas coloreadas. El alelo SLA-01: 0401 se muestra en blanco. Los epítomos de la proteína M son A) PVDEVIEHL, B) LGAPTGVTL, C) APTGVTLTL, D) SQLPNFVTV, E) PVDEVIQHL, es el único epítomo de la secuencia 2013MMV, en color rosa, F) SLFDAWASF, G) ASSGTGWAF y H) RLWRRTHSW.

La **Ilustración 23** muestra y resume el perfil inmunológico de la proteína M del VDEP. Los epítomos de células B lineales y los epítomos discontinuos predichos en ElliPro y Discotope, así como los epítomos de células T, que se predijeron en este estudio, se muestran alineados y representados, en la secuencia de la proteína M, con diferentes colores, los cuales corresponden a los programas utilizados y a cada epítome predicho, se pueden observar áreas con mayor potencial inmunológico, es decir, con mayor presencia de sitios inmunogénicos con base en los análisis realizados en este trabajo, entre los primeros 35 aminoácidos de la proteína y los últimos 70 (156-226) donde se observa mayor presencia de secuencias coloreadas. El estudio proporciona datos sobre la viabilidad de la proteína M para formar parte de una vacuna multivalente sobre la viabilidad de la proteína M para formar una vacuna candidata multivalente.



**Ilustración 23. Resumen de epítomos en la proteína M de CV777 y M de 2013MMV en un formato de alineación múltiple.** Los epítomos de ElliPro y DiscoTope se tiñen siguiendo el formato usado en la ilustración 10 y 11. Los epítomos de células B predichos en Bepipred 1.0 y 2.0, así como los epítomos de células T predichos con NetMHCpan 4.1 están coloreados de rojo. En el recuadro rojo se indica el único epítomo identificado experimentalmente (6).

Debido a la pandemia de SARS-CoV-2 no fue posible continuar con el desarrollo experimental de la tesis, por lo cual se optó por realizar una revisión bibliográfica que comparará las similitudes y diferencias entre el vDEP y el SARS-CoV-2, ambos coronavirus de actual interés e indagar sobre posibles panoramas de investigación para ambos virus.

### V.VIII Revisión bibliográfica del vDEP y del SARS-CoV-2 para demostrar las similitudes y diferencias de ambos coronavirus

El virus de la diarrea epidémica porcina o vDEP, un alfa-coronavirus de importancia económica debido a las pérdidas millonarias en la industria porcina, presenta muchas similitudes con el actual virus que azota a la humanidad, causante de la enfermedad COVID-19, el beta-coronavirus SARS-CoV-2. En este trabajo se presenta en forma resumida las similitudes que comparten estos dos virus, además, de posibles escenarios futuros para el SARS-CoV-2 con base en la bibliografía recopilada del vDEP y herramientas o estrategias desarrolladas en la actualidad para el SARS-CoV-2 que serían de utilidad para la investigación del vDEP. Una revisión rápida de los datos arrojados en la plataforma Scopus revela que existen 1,150 documentos relacionados con la palabra clave “vDEP” hasta la actualidad (19/05/2021) mientras que, utilizando la palabra clave “SARS-CoV-2” existen 54,509 documentos relacionados, una diferencia astronómica que refleja el trabajo colectivo de la comunidad científica para un mismo fin, además, una diferencia aún más significativa si consideramos que el primer artículo publicado relacionado con la palabra clave “vDEP” aparece desde el año 1981, lo que significaría casi 40 años de estudio del vDEP, a diferencia del SARS-CoV-

2, el cual aparece a finales del año 2019 y tiene un poco menos de dos años de estudio; sin embargo, la facilidad y rapidez con la que la comunidad científica respondió a este nuevo virus es gracias también a los años de estudio de otros coronavirus humanos como el SARS y el MERS. Con base en lo anterior, se puede inferir que el comportamiento que tiene y tendrá el SARS-CoV-2 y, probablemente, nuevos coronavirus infecciosos emergentes, pueden ser previsto por los datos e investigaciones ya realizados en otros coronavirus.

#### V.VIII.I VDEP y SARS-CoV-2

Dos enfermedades, en principio diferentes, ocasionadas por estos dos coronavirus son la enfermedad de la diarrea epidémica porcina causada por el virus vDEP, el cual causa vomito, deshidratación, diarrea acuosa y pérdida de peso, con tasas de mortalidad hasta del 100% en lechones (137), y la enfermedad COVID-19 causada por el virus SARS-CoV-2, el cual tiene un espectro de manifestaciones clínicas respiratorias que van desde fiebre, tos seca y disnea hasta neumonía, edema pulmonar, síndrome de dificultad respiratoria aguda y fallas orgánicas múltiples que conducen a la muerte en casos graves (138). En esta revisión se observará que comparten características tanto en la ruta de transmisión como en los síntomas generados. Sobre la ruta de transmisión, se considera que la ruta fecal-oral es la principal vía de transmisión del VDEP, esto debido a las características clínicas que presenta la enfermedad, sin embargo, existe la posibilidad de que las partículas virales de VDEP se dispersen a través de corrientes de aire (139), la importancia de este fenómeno radica en que se ha encontrado hasta  $3.5 \times 10^8$  copias de ARN/m<sup>3</sup> del virus vDEP (140) a una distancia de hasta 10 millas a partir de una granja infectada por VDEP naturalmente (48). Algunos estudios han demostrado la presencia del material genético de VDEP en la cavidad nasal de cerdos que no tuvieron contacto directo o indirecto con cerdos negativos a vDEP (141), también se ha encontrado que las células subepiteliales T CD3<sup>+</sup> median la invasión de vDEP a través de la cavidad nasal en lechones recién nacidos para poder llegar al intestino mediante circulación en sangre (142), estos datos apoyan la idea de la transmisión del vDEP mediante el aire, sumando una ruta alternativa a la ruta fecal-oral por la cual normalmente se transmite el vDEP. En comparación con el virus SARS-CoV-2, la primera hipótesis sobre la transmisión y diseminación del virus, debido a sus características clínicas, es la de transmitirse mediante contacto de persona a persona, incluyendo la transmisión directa como tos, estornudo, transmisión por inhalación de gotas, contacto con las membranas mucosas oral, nasal y ocular y aerosoles (143)(144). Se ha demostrado que el virus puede permanecer viable en aerosoles hasta por 3h (tiempo de duración del experimento) (145), basándonos en la evidencia sobre la probable capacidad del vDEP de transmitirse por aire, era evidente que el virus SARS-CoV-2 tuviera la misma capacidad de diseminación en el ambiente, por lo que, la implementación del uso de cubrebocas desde etapas tempranas de la pandemia por SARS-CoV-2 podría haber ayudado a la disminución en la curva de contagios. Al mismo tiempo, estudios recientes han detectado la presencia de SARS-CoV-2 en muestras orales y anales, lo cual indica que posiblemente la transmisión del virus, si bien es mediante la vía respiratoria, pudiese transmitirse también mediante la ruta fecal-oral (146)(147)(148)(149), de la misma forma que lo haría el vDEP, por lo que incrementar las medidas de higiene con esta información ayudaría también a la disminución del número de contagios.

Otra característica importante en el estudio del SARS-CoV-2 se ha observado en la presencia de virus en muestras anales, la cual es más persistente en estadios tardíos de la infección que en muestras orales, lo que implica que probablemente un paciente con resultados negativos a las pruebas para detectar al virus a partir de muestras orales sea un falso negativo (150)(151)(146). Aunado a esto, la presencia del SARS-CoV-2 en muestras fecales se ha relacionado con pacientes que presentan síntomas característicos de enfermedades gastrointestinales como diarrea, anorexia y náuseas (152)(153)(154)(155). Estudios clínicos han demostrado una incidencia de diarrea en un rango del 2 al 50% de los casos de COVID-19 (156). Se ha detectado, por lo menos en un paciente grave de COVID-19, ARN de SARS-CoV-2 en muestras de esófago, estómago y el recto (152). Adicionalmente, algunos experimentos han comprobado que el virus SARS-CoV-2 tiene la capacidad de infectar y replicarse en células epiteliales intestinales humanas (157)(158); sin embargo, el virus pudiese ser inactivado por los fluidos colónicos humanos, limitando el número de partículas virales infecciosas presentes en las heces (159). Comparando esta información con la conocida en el vDEP, este virus, durante las primeras horas de la infección viral en cerdos, la región media del yeyuno e íleon son las primeras regiones del tracto entérico en ser infectadas, seguidamente del yeyuno proximal y distal y del duodeno (160). Los enterocitos de las vellosidades del intestino son infectadas frecuentemente, generando la necrosis aguda y exfoliación de la lámina propia, ocasionando el vómito y diarrea acuosa característica de la infección por vDEP (161). Es necesario realizar mayores investigaciones para conocer el mecanismo por el cual el SARS-CoV-2 ocasiona las enfermedades gastrointestinales en pacientes infectados y como llega a los tejidos gastrointestinales. Interesantemente, el vDEP también se ha detectado en el tejido pulmonar, principalmente en las células epiteliales y en macrófagos alveolares; a pesar de su detección, no se observaron cambios patológicos significativos (162). Además, se ha comprobado la capacidad del vDEP de infectar también a células dendríticas derivadas de monocitos porcinos; sin embargo, no son el sitio de replicación de este (163), aparentemente el vDEP utiliza las células dendríticas para atravesar la barrera epitelial de la cavidad nasal (164), un fenómeno que se ha identificado en varios virus que aprovechan las células dendríticas para superar la barrera epitelial, similar a un "Caballo de Troya" para evadir las respuestas inmunes antivirales y diseminarse hacia la capa submucosa (165), tal fenómeno se comprobó que utiliza el vDEP para diseminarse a partir del sitio de entrada, en el epitelio nasal, a la mucosa intestinal (164). Por otra parte, la detección del material genético de vDEP en semen de cerdos infectados con el virus abre la posibilidad de cuestionarse acerca de otros métodos de diseminación de la enfermedad (166). Son necesarias mayores investigaciones para conocer si el SARS-CoV-2 es capaz de utilizar el mismo método que el vDEP para llegar a los tejidos intestinales o si utiliza algún mecanismo diferente y poder generar nuevas estrategias que ayuden a disminuir los síntomas y la severidad la enfermedad COVID-19.

#### **V.VIII.II Tropicismo celular y receptores**

La proteína S es la que limita el tropismo de los virus vDEP y el SARS-CoV-2 a cierto tipo celular o tejido debido a su interacción con algún receptor. La proteína S es una glicoproteína de membrana tipo 1, que en la envoltura viral forma las proyecciones con una apariencia similar a una corona en las micrografías electrónicas (167)(168). La proteína S del vDEP y del SARS-Cov-2 contienen un péptido señal en el extremo N-terminal, una región larga extracelular, un dominio transmembranal único y una pequeña región citoplasmática. El ectodominio de la proteína S consiste en una subunidad de unión a receptor, S1, y de una

subunidad de fusión de membrana, S2. La subunidad S1 contiene dos dominios, el dominio N-terminal (S1-NTD) y el dominio C-terminal (S1-CTD), ambos potencialmente funcionales como dominios de unión a receptor (RBD). La proteína aminopeptidasa N (APN) es, presuntamente, el receptor del virus vDEP (169), proteína altamente expresada en tejidos gastrointestinales (170), que explican el tropismo del vDEP por este tejido y su patología. En el caso del virus SARS-CoV-2, la enzima convertidora de angiotensina (ACE2) es el receptor utilizado para la infección viral (171) (172), la cual es altamente expresada en diferentes tejidos, como el intestino delgado, testículos, riñones, corazón, tiroides, tejido adiposo, enterocitos, vesícula biliar, cardiomiocitos, células reproductoras masculinas, trofoblastos placentarios, células ductales y de ojos y pulmones (173)(174), lo que podría relacionarse con el daño multiorgánico presente en los pacientes infectados por el virus (175)(176)(177)(178). Entre los tejidos con menores niveles de expresión se encuentran el bazo, médula ósea, cerebro, vasos sanguíneos y músculos e, interesantemente, muestra niveles medios de expresión se encuentran el colon, hígado, vejiga, glándula suprarrenal (173)(174). En otro estudio se ha observado una expresión baja o nula de ACE2 en el sistema respiratorio (174); sin embargo, se ha encontrado una alta expresión en células secretoras transitorias bronquiales (179), en epitelio nasal y bronquial (180)(181) y en células epiteliales alveolares tipo II (182)(183)(184). Ambas proteínas, APN y ACE2 tienen la característica de ser peptidasas, particularidad distintiva de los receptores de diferentes coronavirus (185). El dominio S1-CTD de la proteína S del vDEP es el responsable de la interacción con APN (186); sin embargo, se ha demostrado la capacidad del vDEP de entrar en células y tejidos APN-*knockout* (187)(188), lo que demuestra que APN no es esencial para la entrada del vDEP a la célula, a pesar de esto, la actividad proteasa de APN promueve la replicación del vDEP (189). En relación a esto, cerdos APN-*knockout* han sido susceptibles a la infección por vDEP (190)(191), lo que demuestra nuevamente que APN no es necesaria para la infección por vDEP, manteniendo la interrogante del receptor para vDEP (192). Acrecentando la duda sobre el receptor, se ha demostrado que células de humano, mono y murciélago son susceptibles a la infección por vDEP (193) lo que podría indicar que existe algún otro receptor utilizado por vDEP, además, de una potencial transmisión zoonótica. En el caso del SARS-CoV-2, el mismo dominio es responsable de la interacción con ACE2 (97)(172). A diferencia del vDEP, no hay duda en la identidad del receptor, pero requiere de la participación de proteasas como la serin proteasa TMPRSS2, necesaria para la entrada del SARS-CoV-2 a la células (159). Además, la presencia del SARS-CoV-2 en el tejido intestinal y las características clínicas de una enfermedad gastrointestinal presentes en los pacientes infectados de SARS-CoV-2 se puede explicar por la presencia de ACE2 en estos tejidos y células, análisis del tipo *single-cell* RNAseq han demostrado que ACE2 es altamente expresado en células alveolares de tipo II de pulmón y también en enterocitos absorbentes del íleon, colon y estómago (182)(194)(184). También se ha observado a través de tinciones la presencia del virus SARS-CoV-2 en el citoplasma de las células epiteliales glandulares gástricas, duodenales y del recto (148), lo que demuestra que la presencia de ACE2 en diferentes tejidos y células permite la infección por SARS-CoV-2. Esta flexibilidad de la proteína S y de los virus para infectar diferentes líneas celulares y de diferentes organismos demuestra que aún faltan investigaciones que permitan entender el origen de esta capacidad infectiva para poder generar mejores fármacos o tratamientos para evitar la enfermedad. Una característica importante en ambas proteínas, la S de vDEP y S de SARS-CoV-2, es la de tener un sitio de escisión por proteasas diferentes a TMPRSS2 como la proteasa furina (195)(196)(197), característica que está relacionada con el procesamiento proteolítico de la proteína S, el cual es necesario para la infección viral, se ha demostrado que la proteína S del vDEP requiere un procesamiento proteolítico para comenzar con su función de fusión, este suceso ocurre únicamente en

el estado de unión a receptor (57). Para la propagación de vDEP en cultivos celulares es necesario suplementar el medio con tripsina (57). Este requisito puede ser complementado *in vivo* por las proteasas del intestino. Las proteasas gástricas y pancreáticas, además de las proteasas expresadas localmente por las células epiteliales intestinales, pueden realizar este proceso proteolítico para la infección por vDEP, esta característica limitaría el tropismo del virus al tracto entérico. En el caso del SARS-CoV-2 la proteasa TMPRSS2 realiza el procesamiento proteolítico en la proteína S (159), y se ha comprobado que a través del inhibidores de estas proteasas privan la entrada del virus (198), interesantemente esta proteasa de TMPRSS2 también se encuentra altamente expresada en tejidos intestinales. Investigaciones sobre el receptor de vDEP han demostrado que la proteína S, en específico, el dominio S1-NTD, posee actividad de unión a azúcares como al ácido siálico ácido 5'-N-acetil 9-O-acetil neuramínico o Neu5,9Ac2, ácido 5-N-glicolil neuramínico o Neu5Gc y ácido 5-N-acetil neuramínico o Neu5Ac proteínas mucinas (186) teniendo la más alta afinidad por el ácido N-acetil neuramínico (193), tales carbohidratos funcionan como correceptores. El colesterol también se ha descubierto como una molécula necesaria para la infección viral del vDEP, el secuestro farmacológico de esta molécula bloquea eficazmente tanto la unión como la internalización del virus (199). Adicionalmente la proteína de unión estrecha ocludina y la integrina  $\alpha\beta3$  mejoran la entrada y replicación de vDEP (200)(201). La alta expresión de la proteína receptor 1 de transferrina (TfR1) en el tejido apical de las vellosidades intestinales promueven la entrada del vDEP a la célula, mediante la interacción entre la proteína S1 con la región extracelular de TfR1 (202). Proteínas similares a estas presentes en humanos podrían interactuar de la misma manera con la proteína S del SARS-CoV-2, y podría ser una razón del por qué la severidad de la enfermedad es mayor en un grupo específico de personas.

### V.VIII.III Clasificación y variantes de virus

Existen diferentes genogrupos-del vDEP, G1 y G2 (cepa clásica y cepas variantes, los cuales se pueden dividir en G1a, G1b, G2a y G2b. En el genogrupo-G1a incluyen cepas clásicas como la cepa prototipo CV777 y otras cepas antiguas de vDEP. El G1b contiene cepas aisladas en Asia que presentan eliminaciones en diferentes proteínas, principalmente en la proteína nsp3 y ORF3, que poseen una similitud alta a una cepa atenuada que se usó como vacuna (cepa DR13) y otras cepas clásicas de China y South Korea (37). El genogrupo G2 se diferencia del G1 principalmente por los cambios aminoacídicos en el dominio NTD de la proteína S. Dentro del genogrupo G2a se encuentran cepas circulantes asiáticas reportadas como casos reemergentes de vDEP y en el genogrupo G2b se encuentran cepas circulantes relacionadas con el brote en Estados Unidos y otras partes del mundo (203). Si bien la proteína S es la proteína que determina el tropismo celular y la interacción con el receptor celular, se ha reportado una alta tasa de mutaciones en la proteína S (204)(39)(205)(206), debido a esto, existen una variante de cepa conocida como vDEP S-INDEL, inicialmente llamada variante estadounidense "US variant", que se caracteriza por presentar inserciones y eliminaciones (INDEL) nucleotídicas y aminoacídicas en la proteína S, principalmente en el dominio S1, similares a la cepa clásica de vDEP CV777, con una virulencia reducida en comparación con otras cepas sin deleciones, como la cepa prototipo US vDEP, la cual tiene un origen relacionado con una cepa vDEP AH2012 circulante en China durante los años 2011-2012 y que se encuentra circulando alrededor del mundo (207)(208)(209). Actualmente las cepas vDEP S-INDEL y non-S-INDEL se encuentran circulando en varias partes del mundo (210). Adicionalmente se ha analizado el genoma completo de varias cepas de vDEP observando mutaciones

que ocurren con frecuencia en ciertas regiones, denominándolas regiones hipervariables, dentro de estas se encuentran algunas regiones del ORF1 como la proteína nsp2, 3, el dominio S1 y S2 de la proteína S, ORF3 y la proteína N (210)(211). En contraste, también se ha propuesto realizar las reconstrucciones filogenéticas del vDEP utilizando la proteína N y ORF3. Los análisis con estas proteínas agrupan a las cepas en dos grandes grupos, G1 y G2, a su vez G2 está compuesta por tres subgrupos G2-1, G2-2 y G2-3 (212)(213). En comparación con el SARS-CoV-2, un análisis de 12,343 genomas de cepas de SARS-CoV-2 aislados de pacientes y depositados en el Global Initiative on Sharing Avian Influenza Data (GISAID) se encontraron más de 1000 mutaciones de las cuales 131 tienen una frecuencia mayor al 10%. Dentro de las más relevantes se encuentran la mutación D614G en la proteína S y la P4715L en la proteína Nsp12 (214), la cual es importante para la replicación del ARN viral. La mutación D614G se ha encontrado ampliamente distribuida en Europa (215) y se ha hipotetizado, se relaciona con una mayor virulencia y mortalidad (216)(217). La mutación P4715L se ha demostrado estar relacionada con una mayor mortalidad, aunque el significado funcional de la mutación no está esclarecido. Varios análisis filogenéticos se han realizado con genomas de diferentes cepas de SARS-CoV-2 aisladas de diferentes pacientes alrededor del mundo y en diferentes momentos de tiempo; sin embargo, por la magnitud de la pandemia, no es posible clasificar con tanta facilidad a las diferentes cepas circulantes (218)(219). A pesar de esto, diferentes estudios han tratado de clasificarlos como en el realizado por Foster y colaboradores en el cual ha clasificado a las secuencias del virus en tres variantes principales; A, B y C. En el grupo A se encuentran los virus ancestrales las cuales están relacionados con coronavirus de murciélagos de los cuales probablemente emergieron, encontrando este grupo, junto con el C, ampliamente distribuido por Europa y América. El grupo B, el más común circulante en el este de Asia, el cual mutó a la variante B a partir de las cepas ancestrales sin salir de Asia (220). Se ha desarrollado también un sistema de clasificación filogenética basado en “linajes” del virus que depende de varios factores y que permitiría identificar a las variantes que más contribuyen a la propagación activa (221). Nextstrain o proyecto de seguimiento en tiempo real de la evolución de patógenos a partir de datos del genoma de patógenos, también ha realizado la relación filogenética de clados entre las cepas de SARS-CoV-2, organizándolos en tres grandes grupos, 19A y 19B, emergentes en Wuhan y las primeras relacionadas con el brote, 20A, las cuales emergen a partir de 19A y dominantes en el brote de Europa en marzo y esparcida globalmente, y en el último, 20B y 20C un grupo de grandes subclados genéticamente distintos 20A (222). Gracias a toda la comunidad científica que está compartiendo los datos de secuenciación de genomas de SARS-CoV-2, en GISAID (<https://www.gisaid.org/>), se pueden encontrar más de 1 millón de secuencias del virus, además, es posible observar un árbol filogenético de hasta 3000 mil genomas del SARS-CoV-2 reportados entre diciembre del 2019 a mayo del 2021, se puede clasificar por regiones, fechas y el análisis puede realizarse a partir del genoma entero o por gen. GISAID utiliza un sistema de nomenclatura para los principales clados, desarrollado por Sebastian Maurer-Stroh y colaboradores basado en mutaciones de marcador dentro de 6 agrupaciones filogenéticas de alto nivel de la división temprana de S y L, a la posterior evolución de L en V y G y más tarde de G en GH y GR (223)(224)(225). Este último sistema podría ser utilizado para la clasificación filogenética del vDEP y otros coronavirus lo que facilitaría el monitoreo casi en tiempo real de las variantes genéticas del virus.

#### **V.VIII.IV Recombinación**

Existen diferentes estudios donde se ha observado la existencia de eventos de recombinación entre cepas de vDEP (226)(227)(228)(229)(230), estos sucesos de recombinación han hecho que la clasificación filogenética se extienda a un grupo adicional, el G2c (231)(232), el cual está integrado por cepas con regiones del genoma resultantes de los eventos de recombinación entre cepas de los demás genogrupos, más interesante aún y, potencialmente peligroso, se ha comprobado la existencia de cepas de vDEP resultado de la recombinación entre una cepa de vDEP y una del virus de la gastroenteritis transmisible o TGEV, tanto en Rusia como en Italia (233) (234). También se ha reportado la presencia de una cepa recombinante de vDEP entre una cepa altamente virulenta y una atenuada, generando una cepa nueva e igualmente virulenta llamada CN/Liaoning25/2018 (235). La gran variabilidad del genoma del vDEP puede adjudicarse a la posibilidad de eventos de recombinación en el dominio S1 del vDEP, lo que originaría las diferencias entre las cepas circulantes y la mutabilidad en la proteína S (236)(210). Eventos de recombinación que tendrían alguna posibilidad de ocurrir entre cepas de SARS-CoV-2 o en algún momento, con otros coronavirus humanos, así como cambios en la proteína S del vDEP por mutaciones o eventos de recombinación han generado dificultades para el control de la enfermedad, así como cepas reemergentes y la baja efectividad de las estrategias de control, entre ellas la vacunación. Por lo que, si bien se están desarrollando vacunas para el SARS-CoV-2, con diferentes tecnologías y plataformas, la efectividad de estas podría verse comprometida si estos eventos de recombinación o mutaciones llegaran a suceder, la idea de generar vacunas anual o bianualmente, similar a lo que ocurre con el virus de la gripe, es una posibilidad que no debería descartarse.

#### **V.VIII.V Vacunas**

Actualmente se formulan diferentes vacunas basadas en la tecnología de ARNm, ADN, vectores virales, proteínas recombinantes, así como de cepas atenuadas e inactivadas. Entre las más utilizadas alrededor del mundo para el control de la infección por SARS-CoV-2 están las formulaciones de Pfizer y BioNTech, las cuales se basan en una vacuna de ARNm llamada BNT162b1, en esta una nanopartícula contiene el ARNm que codifica para el RBD de la proteína S (237). La compañía, Moderna, produce una vacuna para SARS-CoV-2 llamada ARNm-1273, que se compone de un ARNm que codifica a la proteína S estabilizada en un estado de prefusión, el ARNm se encuentra encapsulado en una nanopartícula lipídica (238). La empresa Inovio Pharmaceuticals generó una vacuna de ADN, INO-4800, una secuencia dentro del vector pGX9501 que codifica para la proteína S (239). La empresa CanSino Biological Inc. ha desarrollado una vacuna empleando un virus incompleto como vector, la vacuna llamada Ad5-nCoV que utiliza el vector Ad5 de adenovirus, el cual contiene el gen codificante de la proteína S (240). La misma tecnología es usada por Janssen Vaccines & Prevention, la cual desarrolló la vacuna Ad26 basada en adenovirus que incluye el marco de lectura (241). Otra de las vacunas que utilizan esta tecnología es la vacuna ChAdOx1-nCoV-19 ó AZD-1222, de la Universidad de Oxford y AstraZeneca, que emplea un adenovirus incompleto de chimpancé (ChAd)-que incorpora-al gen S con codones optimizados (242). La vacuna Sputnik-V, desarrollada por Gamaleya Research Institute of Russia, combina a los vectores de adenovirus Ad26 y Ad5 que codifican a la proteína S (243). Entre las vacunas basadas en proteínas recombinantes se encuentra la vacuna NVX-CoV2373 de la empresa Novavax, que utiliza la proteína S completa recombinante en forma trimérica expresada y purificada en células de insecto (244). Como vacunas inactivadas para el SARS-CoV-2, existen las

desarrolladas por el Wuhan Institute of Biological Products (245) y por el Beijing Institute of Biological Products (246). La vacuna de un virus inactivado de PiCoVacc, es una vacuna multivalente desarrollada por Sinovac, que ha demostrado generar anticuerpos con actividad neutralizante para 9 aislados diferentes de SARS-CoV-2 representativos de las cepas circulantes mundialmente (247).

En relación con el desarrollo de vacunas de virus atenuados, se han detectado o desarrollado cepas con una menor virulencia y que mantienen una capacidad replicativa; entre estas investigaciones se encuentran las desarrolladas por el Chungnam National University (248), y la Duke-NUS Medical School of Singapore que han empleado la cepa  $\Delta$ 382. En lo que respecta al control de la infección por el vDEP, existen vacunas de virus inactivados desarrolladas en la década de los 90's en Asia (249); sin embargo, estas vacunas utilizan cepas que ahora son clasificadas como las cepas clásicas o ancestrales, y que no tienen efectividad con las cepas circulantes en la actualidad (250). A partir del brote de 2010 en China, se crearon vacunas de virus atenuados basados en las cepas Zj08 la 83P-5 (95), y DR-13 (85). A partir del brote del 2013 en Estados Unidos se inició el desarrollo de vacunas para el vDEP. Harrisvaccines fue una de las empresas pioneras que produjo la vacuna iPED, basada en la versión truncada del gen que codifica a la proteína S del vDEP producida en la plataforma de tecnología de partículas de ARN SirraVax<sup>SM</sup> (81). Posteriormente, la misma empresa creó la segunda generación de este tipo de vacunas a través de la llamada iPED plus, que genera un fragmento más largo de la proteína S con codones optimizados. Otras vacunas fueron las creadas por Zoetis (251), basada en un virus completo inactivado y la desarrollada por Vaccine and Infectious Disease Organization-InterVac en Canada, basada en la producción recombinante del dominio S1 de la proteína S en células humanas (82).

Una vacuna alternativa se generó en el modelo de *Lactobacillus* que expresa a las proteínas N y S del vDEP, para estimular la producción de inmunoglobulina A (IgA) e IgG específicas de la proteína en la mucosa intestinal porcina (252)(253). También se han desarrollado vacunas basadas en ADN, en los que se ha utilizado un plásmido eucariota recombinante que expresa el gen S de vDEP, que ha demostrado niveles altos de anticuerpos anti-vDEP e inmunidad celular en los animales vacunados (254). Otro estudio desarrolló la vacuna de ADN SL7207 que expresa a la proteína S del vDEP y del TGEV; su inoculación en cerdos demostró una respuesta inmune celular y humoral alta, y particularmente, la generación de anticuerpos neutralizantes (255). Se ha investigado también el potencial de desarrollar una vacuna basada en virus como vector, con el sistema de vector adenoviral Ad5Max y la proteína S de la cepa de vDEP CH/HBXT/2018 se obtuvo al adenovirus recombinante rAD-vDEP-S, el cual indujo una alta respuesta inmune humoral en cerdos inoculados y una protección parcial en cerdos al ser infectados con vDEP (90).

Si bien el surgimiento del SARS-CoV-2 ha originado la llegada de la tecnología de ARNm como nueva plataforma para el desarrollo de vacunas, aún falta camino para la consolidación de las vacunas con tecnología similar en desarrollo para el vDEP. Por otra parte, la aparición de cepas virulentas de vDEP con material genético proveniente de cepas atenuadas que fungieron como vacunas (235), sugiere la necesidad de reevaluar la viabilidad de las vacunas atenuadas que se están desarrollando para el SARS-CoV-2 y evitar sucesos inesperados.

#### **V.VIII.VI Anticuerpos**

Se han analizado y comprobado la efectividad de anticuerpos generados a partir del método de despliegue de fago usando bibliotecas de genes de anticuerpos de personas que no han tenido contacto con agentes infecciosos como los virus de la hepatitis C, de la influenza y el VIH entre otros (256). A partir de esta metodología se ha conseguido obtener anticuerpos anti-SARS-CoV-2 a partir de una biblioteca generada en pre-pandemia proveniente de un donador sano, estos anticuerpos son capaces de unirse a la región S1, RBD y poseen actividad neutralizante, lo que demuestra que las bibliotecas universales de donantes humanos sanos ofrecen una gran alternativa para generar anticuerpos terapéuticos de forma rápida e independiente de la disponibilidad de material biológico de pacientes y puede ser una solución rápida en un caso de pandemia o de la aparición de una nueva clase de microorganismo que represente un riesgo sanitario para la población (257). La posibilidad de contar con una librería de anticuerpos de cerdo empleando la misma metodología o similar ofrecería una herramienta para enfrentar nuevas variantes del vDEP o nuevos virus que afecten a los cerdos o humanos.

#### **V.VIII.VII Conclusiones de la revisión bibliográfica**

El vDEP y el SARS-CoV-2 comparten muchas de sus características en la infección viral, así como en la respuesta celular inducida. Gracias al trabajo previo realizado en las investigaciones con el vDEP y otros coronavirus se ha logrado avanzar significativamente en el entendimiento del SARS-CoV-2. Si bien, algunos descubrimientos del SARS-CoV-2 ya habían sido descritos previamente en otros coronavirus, la magnitud de la pandemia ocasionada por este virus ha permitido que más y diferentes áreas científicas se involucren en el estudio de éste, abriendo nuevas posibilidades de análisis o conocimiento para otros coronavirus incluyendo al vDEP.

La propuesta de la formación en el país de una “red nacional de investigadores de coronavirus” podría impulsar el fortalecimiento científico y económico del país, ya que el vDEP afecta a la industria porcina de manera importante y el SARS-CoV-2 ha afectado a diferentes áreas productivas en todo el país, esta red ayudaría a afrontar nuevos retos que provengan de estos virus, ya sea por cepas mutadas que originen una reemergencia del virus, o incluso por eventos de recombinación que originen virus alternativos y nuevos. SARS-CoV-2, es una muestra del riesgo constante de zoonosis que existe en el planeta y representa un futuro reto para el país en caso de existir un nuevo coronavirus que infecte a humanos, u otro coronavirus que infecte al ganado e impacte económicamente al país.

### **VI. Discusión**

La industria porcina se ve afectada cada año por la pérdida de cerdos por infecciones virales, la diarrea epidémica porcina una de las enfermedades más preocupantes para los productores. La enfermedad, que es muy contagiosa, es causada por vDEP. El vDEP es un agente etiológico asociado con brotes de infección en todo el mundo, incluido México. Si bien esta enfermedad se aborda mediante programas de bioseguridad, prácticas de manejo zootécnico adecuado y vacunas, la ocurrencia de mutaciones en el genoma viral motiva la búsqueda continua de nuevas estrategias de control. Esto requiere conocimiento

molecular de los posibles desencadenantes virales en la respuesta inmune del hospedero. La proteína M cumple varias funciones virales importantes en diferentes regiones de la estructura de la partícula viral. El conocimiento de estas funciones podría ayudar a desarrollar nuevas estrategias para controlar la enfermedad. Sin embargo, se desconocía la estructura 3D de la proteína M de vDEP. Por lo tanto, el presente trabajo generó un modelo 3D de la proteína M y lo comparó con modelos validados de SARS-CoV-2. Las regiones inductoras potenciales de la respuesta inmune se evaluaron en el modelo 3D; Se encontraron 4 sitios prometedores que deberían ser objeto de más estudios. El modelo 3D se construyó a partir de la secuencia de aminoácidos de la proteína M de CV777 y de una cepa mexicana aislada en Michoacán, México, que muestra cambios de aminoácidos en comparación con el prototipo. Esto podría ayudar a determinar si la estructura 3D de la proteína se ve afectada significativamente cuando ocurren cambios en la secuencia de aminoácidos, especialmente en la región globular de la proteína donde se encuentran los residuos asociados con las funciones de la proteína. Para determinar si las secuencias propuestas en la generación del modelo 3D arrojarían información útil sobre las proteínas M de otras cepas y virus de diferentes partes del mundo, se realizó una comparación. A pesar de la presencia de 8 cambios de aminoácidos, la secuencia de proteínas de 2013MMV tiene 97.35-99.56% de identidad entre cepas de México, Estados Unidos, Europa y Asia. Esto muestra un alto grado de conservación de la secuencia de la proteína M, lo que permite generar un modelo útil para el estudio profundo de la proteína M a partir de la secuencia de la cepa CV777 y la secuencia mexicana.

El presente trabajo propone por primera vez un modelo 3D predicho de la proteína M de vDEP. El modelo fue generado por Robetta mediante un modelo comparativo y con la proteína ORF3 de SARS-CoV-2 como una de las principales plantillas de modelado. La estructura de ORF3 se resolvió este año mediante Cryo-EM y tiene la misma topología en el endodominio globular M que la proteína vDEP M [17]. El modelo 3D de la proteína M que se muestra en la Ilustración 2 muestra una topología consistente con la arquitectura informada para la proteína M del coronavirus. Está organizado en dos dominios, uno transmembranal y otro globular, con 3 hélices  $\alpha$  que atravesarían la membrana, con regiones de bucle dentro y fuera de la envoltura viral y un dominio globular formado principalmente por cadenas  $\beta$  antiparalelas. Curiosamente, el modelo 3D conserva la topología reportada en Uniprot para la proteína CV777 M: superficie del virión (posiciones 1-11), transmembranal (12-32, 21 residuos de longitud), intravirión (33-41, 9 residuos), transmembranal (42-62, 21 residuos), superficie del virión (63-75, 13 residuos), transmembranal (76-96, 21 residuos), intravirión (97-226, 130 residuos) (<https://www.uniprot.org/uniprot/P59771>). La proteína M es la proteína estructural más abundante en la envoltura viral del vDEP y otros coronavirus. Se sabe que juega un papel central en el ensamblaje del virus. Se ha informado de la homodimerización de la proteína M a través de la región transmembrana [18]. En el virus de la hepatitis del ratón (MHV), la interacción de la proteína M con la proteína E es el requisito mínimo para el ensamblaje de la partícula viral [18]. Kuo y Masters informaron de una interacción entre las proteínas M y N del MHV a través del extremo C-terminal de la proteína M [19]. Curiosamente, se demostró que, en ausencia de la proteína N, la proteína M del MHV puede interactuar con los ARN sintéticos que contienen una señal de empaquetamiento viral que favorece el empaquetamiento específico de los ARN en partículas similares al coronavirus [20]. Un estudio reciente mostró que la proteína vDEP M interactúa con las proteínas virales S y E, con ORF3 y con 218 proteínas del hospedero [3]. Algunos de ellos están asociados con 131 vías de señalización y 10 procesos biológicos. La proteína M interactúa con el factor de iniciación de la traducción 3 (eIF3L) cuya regulación a la baja aumenta

significativamente la producción viral. El factor de iniciación de la traducción 3 podría actuar como un regulador negativo de la replicación del vDEP [3]. Se ha demostrado que la proteína M de tres coronavirus, SARS-Cov, MHV y el coronavirus felino (FCoV), adopta 2 conformaciones, llamadas largas y compactas, que influyen en la curvatura de la envoltura viral. La conformación larga (M-long) proporciona rigidez y uniformidad a la envoltura viral [21]. También se ha demostrado que la proteína M forma homodímeros por contacto con la región correspondiente al endodominio, y que la conformación M-long interactúa principalmente con las proteínas N y S en una estequiometría del tipo 8M2: 4N: 1S3. La multimerización de la proteína M es un proceso conservado entre los coronavirus [21].

Recientemente, Lim Heo y Michael Feig generaron modelos de homodímeros de la proteína M, incluidos 4 homodímeros de la proteína M que interactúan con un trímero en la región transmembranal de la proteína S del SARS-CoV-2. Los modelos están depositados en GitHub (<https://github.com/Feiglab/sars-cov-2-proteins>) [16]. Conocer las interacciones de la proteína M, así como conocer su estructura 3D, o incluso tener un buen modelo 3D de la misma, permitiría hacer un análisis profundo de los procesos intracelulares que lleva a cabo esta proteína, lo que a su vez permitiría diseñar fármacos. y otras alternativas que podrían impedir que la proteína M realice sus funciones como método para controlar infecciones causadas por vDEP y otros coronavirus. La falta de estrategias de control eficientes para reducir la propagación del vDEP y, por tanto, la muerte de los lechones, mantiene vivo el interés por el desarrollo de nuevos métodos de control. Sin embargo, el genoma viral en constante cambio dificulta la consecución de este objetivo. Diferentes estudios han informado de la existencia de diferentes eventos de recombinación entre cepas de vDEP [22-25]. En Rusia e Italia se informó de un evento de recombinación entre una cepa de vDEP y uno del virus de la gastroenteritis transmisible (TGEV) [26-27]. También se ha informado de un evento de recombinación entre una cepa de vDEP muy virulenta y una cepa de vDEP atenuada; este evento generó una nueva cepa virulenta [28]. El desarrollo de una estrategia de control eficaz se ha visto obstaculizado por la aparición de mutaciones y recombinaciones entre cepas de vDEP en la región codificante de la proteína S. Por tanto, es importante estudiar el potencial inmunológico y funcional de otras proteínas estructurales del virus. La proteína M es un buen candidato debido a su baja variabilidad genética, su estructura y las funciones que podría realizar. Es particularmente interesante analizar el potencial de la proteína M para inducir la activación del sistema inmunológico. Se ha demostrado que, en otros coronavirus, como el MHV, el sistema inmunológico produce anticuerpos monoclonales dirigidos a la proteína M; estos anticuerpos protegen a los ratones de la infección viral [29]. La inmunización de 8 cachorros gato con la proteína M recombinante del virus de la peritonitis infecciosa felina (FIPV) permitió que 3 de ellos sobrevivieran a la infección viral [30]. Un estudio experimental identificó un epítipo de células B en la proteína M de vDEP que corresponde al motivo 195WAFYVR200 [6]. También se han encontrado epítopos de células B, como en la proteína M del SARS-CoV y el virus de la bronquitis infecciosa aviar [31] [32]. Se ha demostrado experimentalmente que algunos péptidos de la proteína M del SARS-CoV son altamente reactivos con sueros de la fase convaleciente de pacientes con SARS. También se obtuvieron títulos altos de anticuerpos contra la proteína M en ratones y conejos inmunizados, lo que sugiere que la proteína M es un componente altamente inmunogénico en las preparaciones de vacunas inactivadas contra el SARS-CoV [33]. El epítipo encontrado por Zhang [6] no coincide exactamente con los epítopos predichos para las células B obtenidos en el presente estudio, como se muestra en la última ilustración del artículo. Algunos epítopos de la región globular de la proteína M se localizaron antes o después de la secuencia WAFYVRS [6]. La investigación adicional del potencial

inmunogénico de la proteína M podría implicar la expresión de la proteína M de los coronavirus en sistemas eucariotas y procariotas con el fin de analizar posibles cambios en la conformación de la proteína. También sería necesario analizar las regiones de contacto de la proteína M con los anticuerpos producidos por los cerdos inmunizados con la proteína vDEP M [6].

Los resultados del presente estudio confirman que la proteína M es un candidato potencial para la activación de la respuesta inmune celular. El análisis de acoplamiento confirmó que los epítomos encontrados pueden interactuar con el alelo SLA-1: 0401, lo que significa que la respuesta inmune celular podría complementarse con la respuesta humoral. Sin embargo, existen limitaciones en los programas para la predicción de epítomos T, al arrojar péptidos con tamaños variables, ocasionando múltiples especificidades hacia las mismas regiones de epítomo al permitir estas variantes de longitudes, la poca caracterización de motivos de HLA de clase I poco estudiados (146). También se demostró un alto grado de conservación de cada uno de los epítomos de células B y T pronosticados en este estudio, en comparación con diferentes cepas de vDEP que circulan en México y otras partes del mundo.

## VII. Conclusión

Se realizó la construcción exitosa de los vectores pPICZαB-M2-226 y pPICZαB-M102-212 y se generaron las cepas productoras de *Pichia pastoris* incorporando los vectores linearizados al cromosoma en el gen AOX. Se obtuvieron modelos 3D predichos de las proteínas CV777 M y 2013MMV M. Se identificaron epítomos de células B continuos y discontinuos y epítomos de células T en la proteína M. La comparación entre los modelos 3D de proteínas vDEP M obtenidos en este estudio y los modelos de proteína M del SARS-CoV-2 obtenidos por el grupo AlphaFold y el laboratorio Feig mostró una alta conservación de la estructura 3D en ambos virus, así como la presencia de epítomos discontinuos en las mismas regiones en ambas proteínas. El modelo 3D de la proteína M de vDEP generado en el presente estudio demostró que sería una herramienta útil para realizar más investigaciones sobre el desarrollo de vacunas o fármacos para controlar la PED.

## VIII. Perspectivas del Trabajo

Como perspectiva inmediata, debido a los acontecimientos de la pandemia por SARS-CoV-2, diferentes experimentos no se pudieron llevar a cabo como consecuencia de las medidas de aislamiento social. Dentro de los experimentos no realizados se encuentran; la caracterización del fenotipo de las células productoras de *Pichia pastoris*, la inducción de la expresión de la proteína M en *Pichia pastoris*, su posterior purificación mediante columnas de afinidad y su evaluación en contra de sueros de cerdos infectados por el vDEP para confirmar el potencial antigénico de la proteína M.

## IX. Bibliografía

1. Su S, Wong G, Shi W, Liu J, Lai ACK, Zhou J, et al. Epidemiology, Genetic Recombination, and Pathogenesis of Coronaviruses. *Trends Microbiol.* 2016 Jun;24(6):490–502.
2. Vijaykrishna D, Smith GJD, Zhang JX, Peiris JSM, Chen H, Guan Y. Evolutionary insights into the ecology of coronaviruses. *J Virol* [Internet]. 2007/01/31. 2007 Apr;81(8):4012–20. Available from: <https://pubmed.ncbi.nlm.nih.gov/17267506>
3. Miranda C, Silva V, Igrejas G, Poeta P. Genomic evolution of the human and animal coronavirus diseases. *Mol Biol Rep.* 2021 Sep;48(9):6645–53.
4. Woo PCY, Lau SKP, Huang Y, Yuen K-Y. Coronavirus Diversity, Phylogeny and Interspecies Jumping. *Exp Biol Med* [Internet]. 2009 Oct 1;234(10):1117–27. Available from: <https://doi.org/10.3181/0903-MR-94>
5. Ye Z-W, Yuan S, Yuen K-S, Fung S-Y, Chan C-P, Jin D-Y. Zoonotic origins of human coronaviruses. *Int J Biol Sci.* 2020;16(10):1686–97.
6. Brown J, Rademacher C, Baker S, Poonsuk K, Cheng TY, Skoland K, et al. Efficacy of a commercial porcine epidemic diarrhea virus vaccine at reducing duration of viral shedding in gilts. *J Swine Heal Prod.* 2019;27(5):256–64.
7. FIRA. Carne de cerdo 2017. Panor Agroaliment [Internet]. 2017;28. Available from: [http://www.ugrpg.org.mx/pdfs/Panorama Agroalimentario Carne de cerdo 2017.pdf](http://www.ugrpg.org.mx/pdfs/Panorama%20Agroalimentario%20Carne%20de%20cerdo%202017.pdf)
8. ALARCÓN CGG, RONQUILLO JCC, SÁNCHEZ JG. Produccion de Cerdos. Col postgraduados. 2005;1–82.
9. OECD. Market examinations in Mexico CASE STUDY OF THE PORK MEAT MARKET. 2019; Available from: <https://www.oecd.org/daf/competition/market-examinations-mexico-pork-meat-market-web-eng.pdf>
10. Vanderwaal K, Deen J. Global trends in infectious diseases of swine. *PNAS.* 2018;115(45):11495–500.
11. Holtkamp DJ, Kliebenstein JB, Neumann EJ, Zimmerman JJ, Rotto HF, Yoder TK, et al. Assessment of the economic impact of porcine reproductive and respiratory syndrome virus on United States pork producers. *J Swine Heal Prod.* 2013;21(2):72–84.
12. Lunney JK, Fang Y, Ladinig A, Chen N, Li Y, Rowland B, et al. Porcine Reproductive and Respiratory Syndrome Virus ( PRRSV ): Pathogenesis and Interaction with the Immune System. *Annu Rev.* 2016;4.
13. Dold C, Holland C V. Ascaris and ascariasis. *Microbes Infect* [Internet]. 2011;13(7):632–7. Available from: <http://dx.doi.org/10.1016/j.micinf.2010.09.012>
14. Galindo I, Alonso C. African Swine Fever Virus : A Review. *Viruses.* 2017;9(103):1–10.
15. Goyette-desjardins G, Auger J, Xu J, Segura M, Gottschalk M. Streptococcus suis , an important pig pathogen and emerging zoonotic agent — an update on the worldwide distribution based on serotyping and sequence typing. *Emerg microbes Infect.* 2014;3(April):1–20.
16. Segalés J, Barcellos D, Alfieri A, Burrough E, Marthaler D. Senecavirus A : An Emerging Pathogen Causing Vesicular Disease and Mortality in Pigs ? *Vet Pathol.* 2017;54(1):11–21.
17. Qi W, Zhu R, Bao C, Xiao J, Liu B, Sun M, et al. Porcine circovirus type 2 promotes Actinobacillus pleuropneumoniae survival during coinfection of porcine alveolar macrophages by inhibiting ROS production. *Vet Microbiol* [Internet]. 2019;233(April):93–101. Available from: <https://doi.org/10.1016/j.vetmic.2019.04.028>
18. Opriessnig T, Giménez-Lirola LG, Halbur PG. Polymicrobial respiratory disease in pigs. *Anim Heal Res Rev.* 2011;12(2):133–48.
19. Franzoni G, Graham SP, Giudici SD, Oggiano A. Porcine Dendritic Cells and Viruses : An Update. *Viruses.* 2019;11(445):1–17.
20. Zhang M, Li W, Zhou P, Liu D, Luo R. Genetic manipulation of porcine deltacoronavirus reveals insights into NS6 and NS7 functions : a novel strategy for vaccine design. *Emerg Microbes Infect.* 2020;9(1):20–31.
21. Chen F, Knutson TP, Rossow S, Saif LJ, Marthaler DG. Decline of transmissible gastroenteritis virus and its complex evolutionary relationship with porcine respiratory coronavirus in the United States. *Sci Rep* [Internet]. 2019;9(3953):1–11. Available from: <http://dx.doi.org/10.1038/s41598-019-40564-z>
22. Schulz LL, Tonsor GT. Assessment of the economic impacts of porcine epidemic diarrhea virus in the United States. *J Anim Sci.* 2015;93(11):5111–8.
23. Paarlberg P. Updated Estimated Economic Welfare Impacts of Porcine Epidemic Diarrhea Virus (Pedv). 2014;4(14):1–38.

24. Stadler J, Moser L, Numberger J, Rieger A, Strutzberg-minder K, Stellberger T, et al. Investigation of three outbreaks of Porcine Epidemic Diarrhea in Germany in 2016 demonstrates age dependent differences in the development of humoral immune response. *Prev Vet Med* [Internet]. 2018;150(March 2017):93–100. Available from: <https://doi.org/10.1016/j.prevetmed.2017.12.012>
25. Wood E. an apparently new syndrome of porcine epidemic diarrhoea. *vet rec.* 1977;100(12):243–4.
26. Pensaert MB, de Bouck P. A New Coronavirus-Like Partiele Associeated With Diarrhea in Swine. *Arch Virol.* 1978;58.
27. Pijpers A, van Nieuwstadt AP, Terpstra C, Verheijden JH. Porcine epidemic diarrhoea virus as a cause of persistent diarrhoea in a herd of breeding and finishing pigs. *Vet Rec.* 1993;132(6):129–31.
28. Nagy B, Nagy G, Meder M ME. Enterotoxigenic Escherichia coli, rotavirus, porcine epidemic diarrhoea virus, adenovirus and calici-like virus in porcine postweaning diarrhoea in Hungary. *acta vet hung.* 1996;44(1):9–19.
29. Pritchard GC, Paton DJ, Wibberley G, Iyata G. Transmissible gastroenteritis and porcine epidemic diarrhoea in Britain. *Vet Rec.* 1999;144(22):616–8.
30. Martelli P, Lavazza A, Nigrelli AD, Meriardi G, Alborali LG, Pensaert MB. Epidemic of diarrhoea caused by porcine epidemic diarrhoea virus in Italy. *Vet Rec.* 2008;162(10):307–10.
31. Li W, Li H, Liu Y, Pan Y, Deng F, Song Y, et al. New Variants of Porcine Epidemic Diarrhea Virus, China, 2011. *Emerg Infect Dis CDC.* 2012;18(8).
32. Puranaveja S, Poolperm P, Lertwatcharasarakul P, Kesdaengsakonwut S, Boonsoongnern A, Urairong K, et al. Chinese-like Strain of Porcine Epidemic diarrhea virus, thailand. *Emerg Infect Dis CDC.* 2009;15(7).
33. Takahashi K, Okada K, Ohshima K. an outbreak of swine diarrhea of a new type associated with coronavirus like particles in japan. *jpn J vet sci.* 1983;45(6):829–32.
34. Chung H, Nguyen VG, Moon H, Lee J, Park S, Lee G, et al. Isolation of Porcine Epidemic Diarrhea Virus during outbreaks in South Korea, 2013–2014. *Emerg Infect Dis CDC.* 2015;21(12).
35. Stevenson GW, Hoang H, Schwartz KJ, Burrough ER, Sun D, Madson D, et al. Emergence of Porcine epidemic diarrhea virus in the United States: Clinical signs, lesions, and viral genomic sequences. *J Vet Diagnostic Investig.* 2013;25(5):649–54.
36. Trujillo-ortega ME, Beltrán-figueroa R, Juárez-ramírez M, Sotomayor-gonzález A, Hernández-villegas EN, Becerra-hernández JF, et al. Isolation and characterization of porcine epidemic diarrhea virus associated with the 2014 disease outbreak in Mexico : case report. *BMC Vet Res* [Internet]. 2016;12(132):1–8. Available from: <http://dx.doi.org/10.1186/s12917-016-0763-z>
37. Vlasova AN, Marthaler D, Wang Q, Culhane MR, Rossow KD, Rovira A, et al. Distinct characteristics and complex evolution of pedv strains, North america, May 2013-february 2014. *Emerg Infect Dis.* 2014;20(10):1620–8.
38. Reveles-Félix S, Carreón-Nápoles R, Mendoza-Elvira S, Quintero-Ramírez V, García-Sánchez J, Martínez-Bautista R, et al. Emerging strains of porcine epidemic diarrhoea virus ( PEDV ) in Mexico. *Transbound Emerg Dis.* 2020;67(July 2019):1035–41.
39. Romero RL, Gómez L, José N, Cerriteño L, Laura S, Valdelamar M, et al. Molecular characterization of the spike gene of the porcine epidemic diarrhea virus in Mexico , 2013 – 2016. *Virus Genes* [Internet]. 2018;54(2):215–24. Available from: <https://doi.org/10.1007/s11262-017-1528-x>
40. Hofmann M, Wyler R. Quantitation, biological and physicochemical properties of cell culture-adapted porcine epidemic diarrhea coronavirus (PEDV). *Vet Microbiol.* 1989;20(2):131–42.
41. Pospischil A, Stuedli A, Kiupel M. Update on porcine epidemic diarrhea. *J Swine Heal Prod.* 2002;10(2):81–5.
42. Shirato K, Matsuyama S, Ujike M, Taguchi F. Role of Proteases in the Release of Porcine Epidemic Diarrhea Virus from Infected Cells. *J Virol.* 2011;85(15):7872–80.
43. Kocherhans R, Bridgen A, Ackermann M, Tobler K. Completion of the Porcine Epidemic Diarrhoea Coronavirus ( PEDV ) Genome Sequence. *Virus Genes.* 2001;23(2):137–44.
44. Luo Y, Zhang J, Deng X, Ye Y, Liao M, Fan H. Complete Genome Sequence of a Highly Prevalent Isolate of Porcine Epidemic Diarrhea Virus in South China. *J Virol.* 2012;86(17):9551–9551.
45. Wei Z -y., Lu W -h., Li Z -l., Mo J -y., Zeng X -d., Zeng Z -l., et al. Complete Genome Sequence of Novel Porcine Epidemic Diarrhea Virus Strain GD-1 in China. *J Virol.* 2012;86(24):13824–5.
46. Park S-J, Kim H-K, Song D-S, An D-J, Park B-K. Complete Genome Sequences of a Korean Virulent Porcine Epidemic Diarrhea Virus and Its Attenuated Counterpart. *J Virol.* 2012;86(10):5964–5964.
47. Tang XC, Zhang JX, Zhang SY, Wang P, Fan XH, Li LF, et al. Prevalence and Genetic Diversity of Coronaviruses in Bats from China. *J Virol.* 2006;80(15):7481–90.

48. Alonso C, Goede DP, Morrison RB, Davies PR, Rovira A, Marthaler DG, et al. Evidence of infectivity of airborne porcine epidemic diarrhea virus and detection of airborne viral RNA at long distances from infected herds. *Vet Res.* 2014;45(1):1–5.
49. Lowe J, Gauger P, Harmon K, Zhang J, Connor J, Yeske P, et al. Role of transportation in spread of porcine epidemic diarrhea virus infection, United States. *Emerg Infect Dis.* 2014;20(5):872–4.
50. Dee S, Clement T, Schelkopf A, Nerem J, Knudsen D, Christopher-Hennings J, et al. An evaluation of contaminated complete feed as a vehicle for porcine epidemic diarrhea virus infection of naïve pigs following consumption via natural feeding behavior: Proof of concept. *BMC Vet Res.* 2014;10(1):1–9.
51. Jung K, Wang Q, Scheuer KA, Lu Z, Zhang Y, Saif LJ. Pathology of US Porcine Epidemic Diarrhea Virus Strain PC21A in. *Emerg Infect Dis.* 2014;20(4):662–5.
52. Park J, Shin H. Porcine epidemic diarrhea virus infects and replicates in porcine alveolar macrophages. *Virus Res [Internet].* 2014;191:143–52. Available from: <http://dx.doi.org/10.1016/j.virusres.2014.07.038>
53. MacLachlan NJ, Dubovi EJ. Coronaviridae. *Fenner's Vet Virol.* 2017;435–61.
54. Burrell CJ, Howard CR, Murphy FA. Chapter 13 - Coronaviruses. *Fenner White's Med Virol.* 2017;437–46.
55. Khamis Z, Menassa R. Porcine Epidemic Diarrhea Virus [Internet]. *Prospects of Plant-Based Vaccines in Veterinary Medicine.* Springer International Publishing; 2018. 1–367 p. Available from: [http://dx.doi.org/10.1007/978-3-319-90137-4\\_12](http://dx.doi.org/10.1007/978-3-319-90137-4_12)
56. Wang K, Lu W, Chen J, Xie S, Shi H, Hsu H, et al. PEDV ORF3 encodes an ion channel protein and regulates virus production. *FEBS Lett.* 2012;586(4):384–91.
57. Wicht O, Li W, Willems L, Meuleman TJ, Wubbolts RW, van Kuppeveld FJM, et al. Proteolytic Activation of the Porcine Epidemic Diarrhea Coronavirus Spike Fusion Protein by Trypsin in Cell Culture. *J Virol.* 2014;88(14):7952–61.
58. Burrell CJ, Howard CR, Murphy FA. Virus Replication. *Fenner White's Med Virol.* 2017;39–55.
59. Baudoux P, Carrat C, Besnardeau L, Charley B, Laude H. Coronavirus Pseudoparticles Formed with Recombinant M and E Proteins Induce Alpha Interferon Synthesis by Leukocytes. *J Virol.* 1998;72(11):8636–43.
60. Vennema H, Godeke G, Rossen JWA, Voorhout WF, Horzinek MC, Opstelten DE, et al. Nucleocapsid-independent assembly of coronavirus-like particles by co-expression of viral envelope protein genes. *EMBO J.* 1996;15(8):2020–8.
61. Huang Y, Yang Z, Kong W, Nabel GJ. Generation of Synthetic Severe Acute Respiratory Syndrome Coronavirus Pseudoparticles : Implications for Assembly and Vaccine Production. *J Virol.* 2004;78(22):12557–65.
62. Siu YL, Teoh KT, Lo J, Chan CM, Kien F, Escriu N, et al. The M , E , and N Structural Proteins of the Severe Acute Respiratory Syndrome Coronavirus Are Required for Efficient Assembly , Trafficking , and Release of Virus-Like Particles . *J Virol.* 2008;82(22):11318–30.
63. Huan C chao, Wang Y, Ni B, Wang R, Huang L, Ren X feng, et al. Porcine epidemic diarrhea virus uses cell-surface heparan sulfate as an attachment factor. *Arch Virol.* 2015;160(7):1621–8.
64. Milewska A, Zarebski M, Nowak P, Stozek K, Potempa J, Pyrc K. Human Coronavirus NL63 Utilizes Heparan Sulfate Proteoglycans for Attachment to Target Cells. *J Virol.* 2014;88(22):13221–30.
65. Watanabe R, Sawicki SG, Taguchi F. Heparan sulfate is a binding molecule but not a receptor for CEACAM1-independent infection of murine coronavirus. *Virology.* 2007;366(1):16–22.
66. Lang J, Yang N, Deng J, Liu K, Yang P, Zhang G, et al. Inhibition of SARS pseudovirus cell entry by lactoferrin binding to heparan sulfate proteoglycans. *PLoS One.* 2011;6(8).
67. Naskalska A, Dabrowska A, Szczepanski A, Milewska A, Jasik KP, Pyrc K. Membrane Protein of Human Coronavirus NL63 Is Responsible for Interaction with the Adhesion Receptor. *J Virol.* 2019;93(19):1–14.
68. Voß D, Pfefferle S, Drosten C, Stevermann L, Traggiai E, Lanzavecchia A, et al. Studies on membrane topology, N-glycosylation and functionality of SARS-CoV membrane protein. *Viol J.* 2009;6(M):1–13.
69. Armstrong J, Niemann H, Smeekens S, Rottier P, Warren G. Sequence and topology of a model intracellular membrane protein, E1 glycoprotein, from a coronavirus. *Nature.* 1984;308(5961):751–2.
70. Rottier PJM, Niesters HGM, Lenstra JA, Van der Zeijst BAM, Welling GW, Welling-Wester S. Predicted Membrane Topology of the Coronavirus Protein E1. *Biochemistry.* 1986;25(6):1335–9.
71. Locker JK, Rose JK, Horzinek MC, Rottier PJM. Membrane assembly of the triple-spanning coronavirus M protein. Individual transmembrane domains show preferred orientation. *J Biol Chem.* 1992;267(30):21911–8.
72. Chan C, Kien F, Siu L, Tse J, Chu K, Kam J, et al. Differential maturation and subcellular localization of severe

- acute respiratory syndrome coronavirus surface proteins S , M and E Printed in Great Britain. *J Gen Virol.* 2005;86:1423–34.
73. Klumperman J, Locker JK, Meijer A, Horzinek MC, Geuze HJ, Rottier PJ. Coronavirus M proteins accumulate in the Golgi complex beyond the site of virion budding. *J Virol.* 1994;68(10):6523–34.
  74. INIFAP. 1ª Reunión Ordinaria 2018. 2018.
  75. SAGARPA, SENASICA. GUÍA RÁPIDA PARA LA VIGILANCIA E INVESTIGACIÓN EPIDEMIOLÓGICA DE DIARREA EPIDÉMICA PORCINA (DEP). 2015;1–29.
  76. Deejai N, Roshorm YM, Kubera A. Antiviral Compounds Against Nucleocapsid Protein of Porcine Epidemic Diarrhea Virus. *Anim Biotechnol* [Internet]. 2017;28(2):120–30. Available from: <http://dx.doi.org/10.1080/10495398.2016.1232268>
  77. Li H, Gao D, Li Y, Wang Y, Liu H, Zhao J. Research in Veterinary Science Antiviral effect of lithium chloride on porcine epidemic diarrhea virus in vitro. *Res Vet Sci* [Internet]. 2018;118(September 2017):288–94. Available from: <https://doi.org/10.1016/j.rvsc.2018.03.002>
  78. Li L, Yu X, Zhang H, Cheng H, Hou L, Zheng Q, et al. In vitro antiviral activity of Griffithsin against porcine epidemic diarrhea virus. *Virus Genes* [Internet]. 2019;55(2):174–81. Available from: <https://doi.org/10.1007/s11262-019-01633-7>
  79. Choi H, Kim J, Lee C, Ahn Y, Song J. Antiviral activity of quercetin 7-rhamnoside against porcine epidemic diarrhea virus. *Antiviral Res.* 2009;81:77–81.
  80. Peng JY, Horng YB, Wu CH, Chang CY, Chang YC, Tsai PS. Evaluation of antiviral activity of *Bacillus licheniformis* - fermented products against porcine epidemic diarrhea virus. *AMB Express* [Internet]. 2019;9(191):1–11. Available from: <https://doi.org/10.1186/s13568-019-0916-0>
  81. Mogler MA, Gander JR, Ray DD, Harris DLH. Vaccination of PEDV-naive dams with replicon RNA Particle vaccine protects suckling piglets from challenge. *North Am PRRS Symp.* 2014;(abstract 75).
  82. Makadiya N, Brownlie R, Hurk J Van Den, Berube N, Allan B, Gerdt V, et al. S1 domain of the porcine epidemic diarrhea virus spike protein as a vaccine antigen. *Virol J* [Internet]. 2016;13(57):1–10. Available from: <http://dx.doi.org/10.1186/s12985-016-0512-8>
  83. Kim S, Cho B, Lee K, Jang Y. N-terminal Domain of the Spike Protein of Porcine Epidemic Diarrhea Virus as a New Candidate Molecule for a Mucosal Vaccine. *immune Netw.* 2018;18(3):1–9.
  84. Singh G, Singh P, Pillatzki A, Nelson E, Webb B, Dillberger-lawson S, et al. A Minimally Replicative Vaccine Protects Vaccinated Piglets Against Challenge With the Porcine Epidemic Diarrhea Virus. *Front Vet Sci.* 2019;6(October):1–11.
  85. Song DS, Oh JS, Kang BK, Yang JS, Moon HJ, Yoo HS, et al. Oral efficacy of Vero cell attenuated porcine epidemic diarrhea virus DR13 strain. *Res Vet Sci.* 2007;82(January):134–40.
  86. Kweon C, Kwon B, Lee J, Kwon G. Derivation of attenuated porcine epidemic diarrhea virus ( PEDV ) as vaccine candidate. *Vaccine.* 1999;17:2546–53.
  87. Hou Y, Ke H, Kim J, Yoo D, Su Y, Boley P, et al. crossm Engineering a Live Attenuated Porcine Epidemic Diarrhea Virus Vaccine Candidate via Inactivation of the Viral 2' - O - Methyltransferase and the Endocytosis Signal of the Spike Protein. *J Virol.* 2019;93(15):1–19.
  88. Hou X, Jiang X, Jiang Y, Tang L, Xu Y, Qiao X. Oral Immunization against PEDV with Recombinant *Lactobacillus casei* Expressing Dendritic Cell-Targeting Peptide Fusing COE Protein of PEDV in Piglets. *Viruses.* 2018;10(106):1–14.
  89. Chang C-Y, Hsu W-T, Chao Y-C, Chang H-W. Display of Porcine Epidemic Diarrhea Virus Spike Protein on Baculovirus to Improve Immunogenicity and Protective Efficacy. *Viruses.* 2018;10(346):1–16.
  90. Liu X, Zhao D, Zhou P, Zhang Y, Wang Y. Evaluation of the Efficacy of a Recombinant Adenovirus Expressing the Spike Protein of Porcine Epidemic Diarrhea Virus in Pigs. *Biomed Res Int.* 2019;2019:1–8.
  91. Zhang Q, Shi K, Yoo D. Suppression of type I interferon production by porcine epidemic diarrhea virus and degradation of CREB-binding protein by nsp1. *Virology* [Internet]. 2016;489:252–68. Available from: <http://dx.doi.org/10.1016/j.virol.2015.12.010>
  92. Yang S, Li L, Yin S, Shang Y, Umar MKZ, He X, et al. Single-domain antibodies as promising experimental tools in imaging and isolation of porcine epidemic diarrhea virus. *Appl Microbiol Biotechnol.* 2018;102:8931–42.
  93. Zhang Z, Chen J, Shi H, Chen X, Shi D, Feng L, et al. Identification of a conserved linear B-cell epitope in the M protein of porcine epidemic diarrhea virus. *Virol J* [Internet]. 2012;9(225):1–9. Available from: *Virology Journal*
  94. Shasha Li, Yang J, Zhu Z, Zheng H. Porcine Epidemic Diarrhea Virus and the Host Innate. *Pathogens.*

- 2020;9(367):1–27.
95. Sato T, Takeyama N, Katsumata A, Tuchiya K, Kodama T, Kusanagi KI. Mutations in the spike gene of porcine epidemic diarrhea virus associated with growth adaptation in vitro and attenuation of virulence in vivo. *Virus Genes*. 2011;43(1):72–8.
  96. Wang R, Yu R, Chen B, Si F, Wang J, Xie C, et al. Identification of host cell proteins that interact with the M protein of porcine epidemic diarrhea virus. *Vet Microbiol* [Internet]. 2020;246:108729. Available from: <https://doi.org/10.1016/j.vetmic.2020.108729>
  97. Wrapp D, McLellan JS. The 3.1-Angstrom Cryo-electron Microscopy Structure of the Porcine Epidemic Diarrhea Virus Spike Protein in the Prefusion Conformation. *J Virol*. 2019;93(23):1–13.
  98. Ye G, Wang X, Tong X, Shi Y, Fu Z, Peng G. Structural Basis for Inhibiting Porcine Epidemic Diarrhea Virus Replication with the 3C-Like Protease Inhibitor GC376. *Viruses*. 2020;12(240):1–15.
  99. Zeng Z, Deng F, Shi K, Ye G, Wang G, Fang L, et al. Dimerization of Coronavirus nsp9 with Diverse Modes Enhances Its Nucleic Acid Binding Affinity. *J Virol*. 2018;92(17):1–15.
  100. Ahmad M, Hirz M, Pichler H. Protein expression in *Pichia pastoris* : recent achievements and perspectives for heterologous protein production. *applied microbiology and biotechnology*. 2014;98:5301–17.
  101. Tran AM, Nguyen TT, Nguyen CT, Mai X, Thi H, Nguyen CT. *Pichia pastoris* versus *Saccharomyces cerevisiae* : a case study on the recombinant production of human granulocyte - macrophage colony - stimulating factor. *BMC Res Notes*. 2017;10(148):6–13.
  102. Yang Z, Zhang Z. Engineering strategies for enhanced production of protein and bio-products in *Pichia pastoris* : A review. *Biotechnol Adv* [Internet]. 2018;36(1):182–95. Available from: <https://doi.org/10.1016/j.biotechadv.2017.11.002>
  103. Zhu T, Sun H, Wang M, Li Y. *Pichia pastoris* as a Versatile Cell Factory for the Production of Industrial Enzymes and Chemicals : Current Status and Future Perspectives. *Biotechnol J* [Internet]. 2019;14(6):1–13. Available from: <http://dx.doi.org/10.1002/biot.201800694>
  104. Looser V, Bruhlmann B, Bumbak F, Stenger C, Costa M, Camattari A, et al. Cultivation strategies to enhance productivity of *Pichia pastoris* : A review. *Biotechnol Adv* [Internet]. 2015;33(6):1177–93. Available from: <http://dx.doi.org/10.1016/j.biotechadv.2015.05.008>
  105. Hanke D, Pohlmann A, Sauter-louis C, Höper D, Stadler J, Ritzmann M, et al. Porcine Epidemic Diarrhea in Europe : In-Detail Molecular Epidemiology. *Viruses*. 2017;9(177):1–13.
  106. Song D, Huang D, Peng Q, Huang T, Chen Y, Zhang T, et al. Molecular Characterization and Phylogenetic Analysis of Porcine Epidemic Diarrhea Viruses Associated with Outbreaks of Severe Diarrhea in Piglets in Jiangxi , China 2013. *PLoS One*. 2015;10(3):1–16.
  107. Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res*. 1997;25(17):3389–402.
  108. Waterhouse A, Bertoni M, Bienert S, Studer G, Tauriello G, Gumienny R, et al. SWISS-MODEL: Homology modelling of protein structures and complexes. *Nucleic Acids Res*. 2018;46:W296–303.
  109. Yang Y, Faraggi E, Zhao H, Zhou Y. Improving protein fold recognition and template-based modeling by employing probabilistic-based matching between predicted one-dimensional structural properties of query and corresponding native properties of templates. *Bioinformatics*. 2011;27(15):2076–82.
  110. McGuffin LJ, Bryson K, Jones DT. The PSIPRED protein structure prediction server. *Bioinforma Appl note*. 2000;16(4):404–5.
  111. Yoon B. Hidden Markov Models and their Applications in Biological Sequence Analysis. *Curr Genomics*. 2009;10(6):402–15.
  112. Kelley LA, Mezulis S, Yates CM, Wass MN, Sternberg MJE. The Phyre2 web portal for protein modeling , prediction and analysis. *Nat Protoc*. 2015;10(6):845–58.
  113. Källberg M, Wang H, Wang S, Peng J, Wang Z, Lu H, et al. Template-based protein structure modeling using the RaptorX web server. *Nat P*. 2016;7(8):1511–22.
  114. Roy A, Kucukural A, Zhang Y. I-TASSER: a unified platform for automated protein structure and function prediction. *Nat Protoc* [Internet]. 2010;5(4):725–38. Available from: [http://www.proteopedia.org/w/Ramachandran\\_Plot&oldid=1772128%0Ahttp://www.jove.com/details.php?id=3259](http://www.proteopedia.org/w/Ramachandran_Plot&oldid=1772128%0Ahttp://www.jove.com/details.php?id=3259)
  115. Meier A, Söding J. Automatic Prediction of Protein 3D Structures by Probabilistic Multi-template Homology Modeling. *PLoS Comput Biol*. 2015;11(10):1–20.

116. Song Y, Dimaio F, Wang RYR, Kim D, Miles C, Brunette T, et al. High-resolution comparative modeling with RosettaCM. *Structure* [Internet]. 2013;21(10):1735–42. Available from: <http://dx.doi.org/10.1016/j.str.2013.08.005>
117. Studer G, Biasini M, Schwede T. Assessing the local structural quality of transmembrane protein models using statistical potentials ( QMEANBrane ). *Bioinformatics*. 2014;30:505–11.
118. Laskowski RA, MacArthur MW, Moss DS, Thornton JM. PROCHECK: a program to check the stereochemical quality of protein structures. *J Appl Cryst*. 1993;26:283–91.
119. Maghrabi AHA, McGuffin LJ. ModFOLD6 : an accurate web server for the global and local quality estimation of 3D protein models. *Nucleic Acids Res*. 2017;45(April):416–21.
120. Sippl MJ. Recognition of Errors in Three-Dimensional Structures of Proteins. *Proteins*. 1993;17:355–62.
121. Wiederstein M, Sippl MJ. ProSA-web : interactive web service for the recognition of errors in three-dimensional structures of proteins. *Nucleic Acids Res*. 2007;35:407–10.
122. Williams CJ, Headd JJ, Moriarty NW, Prisant MG, Videau LL, Deis LN, et al. MolProbity : More and better reference data for improved all-atom structure validation. *protein Soc*. 2017;27:293–315.
123. Senior AW, Evans R, Jumper J, Kirkpatrick J, Sifre L, Green T, et al. Improved protein structure prediction using potentials from deep learning. *Nature*. 2020;577:706–10.
124. Heo L, Feig M. Modeling of Severe Acute Respiratory Syndrome Coronavirus 2 ( SARS-CoV-2 ) Proteins by Machine Learning and Physics-Based Refinement. *bioRxiv Prepr Serv Biol*. 2020;2020.03.25(008904).
125. Jespersen MC, Peters B, Nielsen M, Marcatili P. BepiPred-2.0: improving sequence-based B-cell epitope prediction using conformational epitopes. *Nucleic Acids Res*. 2017;45(May):24–9.
126. Larsen JEP, Lund O, Nielsen M. Improved method for predicting linear B-cell epitopes. *immunome Res*. 2006;2(2):1–7.
127. Reynisson B, Alvarez B, Nielsen M, Paul S, Peters B. NetMHCpan-4.1 and NetMHCIIpan-4.0: improved predictions of MHC antigen presentation by concurrent motif deconvolution and integration of MS MHC eluted ligand data. *Nucleic Acids Res*. 2020;48(May):449–54.
128. Sørensen MR, Ilsøe M, Strube ML, Bishop R, Erbs G, Hartmann SB, et al. Sequence-Based Genotyping of Expressed Swine Leukocyte Antigen Class I Alleles by Next-Generation Sequencing Reveal Novel Swine Leukocyte Antigen Class I Haplotypes and Alleles in Belgian, Danish, and Kenyan Fattening Pigs and Göttingen Minipigs. *Front Immunol*. 2017;8(701):1–11.
129. Yin HS, Wen X, Paterson RG, Lamb RA, Jardetzky TS. Structure of the parainfluenza virus 5 F protein in its metastable, prefusion conformation. *Nature*. 2006;439(7072):38–44.
130. Trott O., Olson AJ. Autodock vina: improving the speed and accuracy of docking. *J Comput Chem*. 2010;31(2):455–61.
131. Saitou N, Nei M. The Neighbor-joining Method : A New Method for Reconstructing Phylogenetic Trees '. *Mol Biol Evol*. 1987;4:406–25.
132. Felsenstein J. Confidence Limits on Phylogenies : An Approach Using the Bootstrap. *Evolution (N Y)*. 2009;39(4):783–91.
133. Zuckerkandl E, Pauling L. Evolutionary Divergence and Convergence in Proteins. *Evol Genes Protein*. 1965;97–166.
134. Kumar S, Stecher G, Li M, Knyaz C, Tamura K. MEGA X : Molecular Evolutionary Genetics Analysis across Computing Platforms. *Mol Biol Evol*. 2018;35(6):1547–9.
135. Benkert P, Ku M, Schwede T. QMEAN server for protein model quality estimation. *Nucleic Acids Res*. 2009;37(May):510–4.
136. Jumper J, Tunyasuvunakool K, Kohli P, Hassabis D, Team A. Computational predictions of protein structures associated with COVID-19. *Deep website*. 2020;Versión 3(4 August).
137. Jung K, Saif LJ, Wang Q. Porcine epidemic diarrhea virus (PEDV): An update on etiology, transmission, pathogenesis, and prevention and control. *Virus Res*. 2020;286:1–13.
138. Dawei Wang, Hu B, Hu C, Zhu F, Liu X, Zhang J, et al. Clinical Characteristics of 138 Hospitalized Patients With 2019 Novel Coronavirus–Infected Pneumonia in Wuhan, China. *JAMA*. 2020;323(11):1061–9.
139. Beam A, Goede D, Fox A, Mccool MJ, Wall G, Haley C, et al. A Porcine Epidemic Diarrhea Virus Outbreak in One Geographic Region of the United States : Descriptive Epidemiology and Investigation of the Possibility of Airborne Virus Spread. *PLoS One*. 2015;10(12):1–12.
140. Alonso C, Raynor PC, Davies PR. Concentration , Size Distribution , and Infectivity of Airborne Particles Carrying

- Swine Viruses. *PLoS One*. 2015;10(8):1–12.
141. Niederwerder MC, Nietfeld JC, Bai J, Peddireddi L, Breazeale B, Anderson J, et al. Tissue localization , shedding , virus carriage , antibody response , and aerosol transmission of Porcine epidemic diarrhea virus following inoculation of 4-week-old feeder pigs. *J Vet Diagnostic Investig*. 2016;28(6):671–8.
  142. Yuan C, Jin Y, Li Y, Zhang E, Zhang P, Yang Q. PEDV infection in neonatal piglets through the nasal cavity is mediated by subepithelial CD3(+) T cells. *Vet Res*. 2021 Feb;52(1):26.
  143. Morawska L, Cao J. Airborne transmission of SARS-CoV-2: The world should face the reality. *Environ Int*. 2020;139:105730.
  144. Van-Doremalen N, Bushmaker T, Morris DH, Holbrook MG, Gamble A, Williamson BN, et al. Aerosol and Surface Stability of SARS-CoV-2 as Compared with SARS-CoV-1. *New Engl J Med*. 2020;382.
  145. Arslan M, Xu B, Gamal El-Din M. Transmission of SARS-CoV-2 via fecal-oral and aerosols-borne routes: Environmental dynamics and implications for wastewater management in underprivileged societies. *Sci Total Environ* [Internet]. 2020;743:140709. Available from: <https://doi.org/10.1016/j.scitotenv.2020.140709>
  146. Zhang W, Dub R-H, Li B, Zheng X-S, Yang X-L, Hu B, et al. Molecular and serological investigation of 2019-nCoV infected patients: implication of multiple shedding routes. *Emerg Microbes Infect*. 2020;9:386–9.
  147. Chen Y, Chen L, Deng Q, Zhang G, Wu K, Ni L, et al. The presence of SARS-CoV-2 RNA in the feces of COVID-19 patients. *J Med Virol*. 2020;92(7):833–40.
  148. Xiao F, Tang M, Zheng X, Liu Y, Li X, Shan H. Evidence for Gastrointestinal Infection of SARS-CoV-2. *Gastroenterology* [Internet]. 2020;158(6):1831-1833.e3. Available from: <https://doi.org/10.1053/j.gastro.2020.02.055>
  149. Young BE, Ong SWX, Kalimuddin S, Low JG, Tan SY, Loh J, et al. Epidemiologic Features and Clinical Course of Patients Infected With SARS-CoV-2 in Singapore. *JAMA*. 2020;323(15):1488–94.
  150. Xu Y, Li X, Zhu B, Liang H, Fang C, Gong Y, et al. Characteristics of pediatric SARS-CoV-2 infection and potential evidence for persistent fecal viral shedding. *Nat Med*. 2020;26(April):502–5.
  151. Xing Y, Ni W, Wu Q, Li W, Li G, Wang W, et al. Prolonged viral shedding in feces of pediatric patients with coronavirus disease 2019. *J Microbiol Immunol Infect* [Internet]. 2020;53(3):473–80. Available from: <https://doi.org/10.1016/j.jmii.2020.03.021>
  152. Lin L, Jiang X, Zhang Z, Huang S, Zhang Z, Fang Z, et al. Gastrointestinal symptoms of 95 cases with SARS- - CoV-2 infection. *Gut*. 2020;69:997–1001.
  153. Chen N, Zhou M, Dong X, Qu J, Gong F, Han Y, et al. Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in Wuhan , China : a descriptive study. *Lancet* [Internet]. 2020;395:507–13. Available from: [http://dx.doi.org/10.1016/S0140-6736\(20\)30211-7](http://dx.doi.org/10.1016/S0140-6736(20)30211-7)
  154. Pan L, Mu M, Yang P, Sun Y, Wang R, Yan J, et al. Clinical Characteristics of COVID-19 Patients With Digestive Symptoms in Hubei , China : A Descriptive , Cross-Sectional, Multicenter Study. *Am J Gastroenterol*. 2020;115:766–73.
  155. Cholanteril G, Podboy A, Aivaliotis VI, Tarlow B, Pham EA, Spencer SP, et al. High Prevalence of Concurrent Gastrointestinal Manifestations in Patients With Severe Acute Respiratory Syndrome Coronavirus 2: Early Experience From California. *Gastroenterology* [Internet]. 2020;159(2):775–7. Available from: <https://doi.org/10.1053/j.gastro.2020.04.008>
  156. D’Amico F, Baumgart DC, Danese S, Peyrin-Biroulet L. Diarrhea During COVID-19 Infection : Pathogenesis, Epidemiology, Prevention, and Management. *Clin Gastroenterol Hepatol*. 2020;18:1663–72.
  157. Lamers MM, Beumer J, Vaart J van der, Knoops K, Puschhof J, Breugem TI, et al. SARS-CoV-2 productively infects human gut enterocytes. *Science* (80- ). 2020;1669(May):1–10.
  158. Stanifer ML, Kee C, Cortese M, Alexandrov T, Bartenschlager R, Boulant S, et al. Report Critical Role of Type III Interferon in Controlling SARS-CoV-2 Infection in Human Intestinal Epithelial II II Critical Role of Type III Interferon in Controlling SARS-CoV-2 Infection in Human Intestinal Epithelial Cells. *CellReports* [Internet]. 2020;32(1):107863. Available from: <https://doi.org/10.1016/j.celrep.2020.107863>
  159. Zang R, Immunol S, Zang R, Florencia M, Castro G, Mccune BT, et al. TMPRSS2 and TMPRSS4 promote SARS-CoV-2 infection of human small intestinal enterocytes. *Sci Immunol*. 2020;3582(May):1–14.
  160. Jung K, Miyazaki A, Saif LJ. Immunohistochemical detection of the vomiting-inducing monoamine neurotransmitter serotonin and enterochromaffin cells in the intestines of conventional or gnotobiotic ( Gn ) pigs infected with porcine epidemic diarrhea virus ( PEDV ) and serum cytokine. *Res Vet Sci*. 2018;119:99–108.

161. Jung K, Annamalai T, Lu Z, Saif LJ. Comparative pathogenesis of US porcine epidemic diarrhea virus ( PEDV ) strain PC21A in conventional 9-day-old nursing piglets vs . 26-day-old weaned pigs. *Vet Microbiol.* 2015;178:31–40.
162. Park J, Shin H. Porcine epidemic diarrhea virus infects and replicates in porcine alveolar macrophages. *Virus Res.* 2014;191:143–52.
163. Wang X, Ohnstad M, Nelsen A, Nelson E. Porcine epidemic diarrhea virus does not replicate in porcine monocyte- derived dendritic cells , but activates the transcription of type I interferon and chemokine. *Vet Microbiol [Internet].* 2017;208(May):77–81. Available from: <http://dx.doi.org/10.1016/j.vetmic.2017.07.014>
164. Li Y, Wu Q, Huang L, Yuan C, Wang J, Yang Q. An alternative pathway of enteric PEDV dissemination from nasal cavity to intestinal mucosa in swine Yuchen. *Nat Commun.* 2018;9(811):1–14.
165. Schönrich G, Raftery MJ. Dendritic cells as Achilles’ heel and Trojan horse during varicella zoster virus infection. *Front Microbiol.* 2015;6(417):1–6.
166. Gallien S, Moro A, Lediguerher G, Catinot V, Paboef F, Bigault L, et al. Evidence of porcine epidemic diarrhea virus (PEDV) shedding in semen from infected specific pathogen-free boars. *Vet Res [Internet].* 2018;49(7):1–9. Available from: <https://doi.org/10.1186/s13567-018-0505-2>
167. Pensaert MB, de Bouck P. A New Coronavirus-Like Partiele Assoeiated With Diarrhea in Swine. *Arch Virol.* 1978;58:243–7.
168. Walls AC, Park Y-J, Tortorici MA, Wall A, McGuire AT, Veessler D. Structure , Function , and Antigenicity of the SARS-CoV-2 Spike Glycoprotein. *Cell.* 2020;180:281–92.
169. Li BX, Ge JW, Li YJ. Porcine aminopeptidase N is a functional receptor for the PEDV coronavirus. *Virology.* 2007;365(1):166–72.
170. Hashida H, Takabayashi A, Kanai M, Adachi M, Kondo K, Kohno N, et al. Aminopeptidase N Is Involved in Cell Motility and Angiogenesis: Its Clinical Significance in Human Colon Cancer. *Gastroenterology.* 2002;122:376–86.
171. Zhou P, Yang X Lou, Wang XG, Hu B, Zhang L, Zhang W, et al. A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature [Internet].* 2020;579(7798):270–3. Available from: <http://dx.doi.org/10.1038/s41586-020-2012-7>
172. Lan J, Ge J, Yu J, Shan S, Zhou H, Fan S, et al. Structure of the SARS-CoV-2 spike receptor-binding domain bound to the ACE2 receptor. *Nature.* 2020;581(7807):215–20.
173. Li M, Li L, Zhang Y, Wang X. Expression of the SARS-CoV-2 cell receptor gene ACE2 in a wide variety of human tissues. *Infect Dis Poverty.* 2020;9(45):1–7.
174. Hikmet F, Méar L, Edvinsson Å, Micke P, Uhlén M, Lindskog C. The protein expression profile of ACE 2 in human tissues. *Mol Syst Biol.* 2020;16(e9610):1–16.
175. Prasad A, Prasad M. Single Virus Targeting Multiple Organs : What We Know and Where We Are Heading ? *Front Med.* 2020;7(370):1–6.
176. Mokhtari T, Hassani F, Ghaffari N, Ebrahimi B, Yarahmadi A, Hassanzadeh G. COVID-19 and multiorgan failure : A narrative review on potential mechanisms. *J Mol Histol [Internet].* 2020;51:613–28. Available from: <https://doi.org/10.1007/s10735-020-09915-3>
177. Devaux CA, Rolain J, Raoult D. ACE2 receptor polymorphism : Susceptibility to SARS-CoV-2 , hypertension , multi-organ failure , and COVID-19 disease outcome. *J Microbiol Immunol Infect [Internet].* 2020;53:425–35. Available from: <https://doi.org/10.1016/j.jmii.2020.04.015>
178. Renu K, Prasanna PL, Gopalakrishnan AV. Coronaviruses pathogenesis , comorbidities and multi-organ damage – A review. *Life Sci [Internet].* 2020;255(May):117839. Available from: <https://doi.org/10.1016/j.lfs.2020.117839>
179. Lukassen S, Chua RL, Trefzer T, Kahn NC, Schneider MA, Muley T, et al. SARS-CoV- 2 receptor ACE 2 and TMPRSS 2 are primarily expressed in bronchial transient secretory cells. *EMBO J.* 2020;39(e105114):1–15.
180. Bertram S, Heurich A, Lavender H, Gierer S, Danisch S, Soilleux EJ, et al. Influenza and SARS-Coronavirus Activating Proteases TMPRSS2 and HAT Are Expressed at Multiple Sites in Human Respiratory and Gastrointestinal Tracts. *PLoS One.* 2012;7(4):1–8.
181. Sungnak W, Huang N, Bécavin C, Berg M, Queen R, Litvinukova M, et al. SARS-CoV-2 entry factors are highly expressed in nasal epithelial cells together with innate immune genes. *Nat Med.* 2020;26(May):681–7.
182. Zhao Y, Zhao Z, Wang Y, Zhou Y, Ma Y, Zuo W. Single-Cell RNA Expression Profiling of ACE2, the Receptor of SARS-CoV-2. *Am J Respir Crit Care Med.* 2020;202(5):756–9.

183. Zou X, Chen K, Zou J, Han P, Hao J, Han Z. Single-cell RNA-seq data analysis on the receptor ACE2 expression reveals the potential risk of different human organs vulnerable to 2019-nCoV infection. *Fron Med*. 2020;
184. Qi F, Qian S, Zhang S, Zhang Z. Single cell RNA sequencing of 13 human tissues identify cell types and receptors of human coronaviruses. *Biochem Biophys Res Commun* [Internet]. 2020;526(1):135–40. Available from: <https://doi.org/10.1016/j.bbrc.2020.03.044>
185. Li F. Receptor Recognition Mechanisms of Coronaviruses: a Decade of Structural Studies. *J Virol*. 2015;89(4):1954–64.
186. Deng F, Ye G, Liu Q, Navid MT, Zhong X, Li Y, et al. Identification and comparison of receptor binding characteristics of the spike protein of two porcine epidemic diarrhea virus strains. *Viruses*. 2016;8(55):1–15.
187. Ji CM, Wang B, Zhou J, Huang YW. Aminopeptidase-N-independent entry of porcine epidemic diarrhea virus into Vero or porcine small intestine epithelial cells. *Virology* [Internet]. 2018;517(October 2017):16–23. Available from: <https://doi.org/10.1016/j.virol.2018.02.019>
188. Li W, Luo R, He Q, Kuppeveld FJM van, Rottier PJM, Bosch B-J. Aminopeptidase N is not required for porcine epidemic diarrhea virus cell entry. *Virus Res*. 2017;235:6–13.
189. Shirato K, Maejima M, Islam MT, Miyazaki A, Kawase M, Matsuyama S, et al. Porcine aminopeptidase N is not a cellular receptor of porcine epidemic diarrhea virus, but promotes its infectivity via aminopeptidase activity. *J Gen Virol*. 2016;97(10):2528–39.
190. Whitworth KM, Rowland RRR, Petrovan V, Sheahan M, Cino-Ozuna AG, Fang Y, et al. Resistance to coronavirus infection in amino peptidase N-deficient pigs. *Transgenic Res* [Internet]. 2019;28(1):21–32. Available from: <https://doi.org/10.1007/s11248-018-0100-3>
191. Zhang J, Wu Z, Yang H. Aminopeptidase N Knockout Pigs Are Not Resistant to Porcine Epidemic Diarrhea Virus Infection. *Virol Sin* [Internet]. 2019;34(5):592–5. Available from: <https://doi.org/10.1007/s12250-019-00127-y>
192. Cui T, Theuns S, Xie J, van den Broeck W, Nauwynck HJ. Role of porcine aminopeptidase N and sialic acids in porcine coronavirus infections in primary porcine enterocytes. *Viruses*. 2020;12(4):1–18.
193. Liu C, Tang J, Liang X, Yang Y, Peng G, Qi Q, et al. Receptor Usage and Cell Entry of Porcine Epidemic Diarrhea. *J Virol*. 2015;89(11):6121–5.
194. Zhang H, Kang Z, Gong H, Xu D, Wang J, Li Z, et al. The digestive system is a potential route of 2019-nCoV infection: a bioinformatics analysis based on single-cell transcriptomes. *bioRxiv*. 2020;01(30):927806.
195. de Haan CAM, Haijema BJ, Schellen P, Schreur PW, te Lintelo E, Vennema H, et al. Cleavage of Group 1 Coronavirus Spike Proteins: How Furin Cleavage Is Traded Off against Heparan Sulfate Binding upon Cell Culture Adaptation. *J Virol*. 2008;82(12):6078–83.
196. Coutard B, Valle C, de Lamballerie X, Canard B, Seidah NG, Decroly E. The spike glycoprotein of the new coronavirus 2019-nCoV contains a furin-like cleavage site absent in CoV of the same clade. *Antiviral Res* [Internet]. 2020;176:104742. Available from: <https://doi.org/10.1016/j.antiviral.2020.104742>
197. Hoffmann M, Kleine-Weber H, Pöhlmann S. A Multibasic Cleavage Site in the Spike Protein of SARS-CoV-2 Is Essential for Infection of Human Lung Cells. *Mol Cell*. 2020;78(4):779–784.e5.
198. Hoffmann M, Kleine-Weber H, Schroeder S, Kruger N, Herrler T, Erichsen S, et al. SARS-CoV-2 Cell Entry Depends on ACE2 and TMPRSS2 and Is Blocked by a Clinically Proven Protease Inhibitor. *Cell*. 2020;181:271–80.
199. Jeon JH, Lee C. Cellular cholesterol is required for porcine nidovirus infection. *Arch Virol*. 2017;162(12):3753–67.
200. Luo X, Guo L, Zhang J, Xu Y, Gu W, Feng L, et al. Tight Junction Protein Occludin Is a Porcine Epidemic Diarrhea Virus Entry Factor. *J Virol*. 2017;91(10):1–14.
201. Li C, Su M, Yin B, Guo D, Wei S, Kong F, et al. Integrin  $\alpha\beta 3$  enhances replication of porcine epidemic diarrhea virus on Vero E6 and porcine intestinal epithelial cells. *Vet Microbiol* [Internet]. 2019;237(February):108400. Available from: <https://doi.org/10.1016/j.vetmic.2019.108400>
202. Zhang S, Cao Y, Yang Q. Transferrin receptor 1 levels at the cell surface influence the susceptibility of newborn piglets to PEDV infection. *PLoS Pathog* [Internet]. 2020;16(7):1–26. Available from: <http://dx.doi.org/10.1371/journal.ppat.1008682>
203. Huang Y, Dickerman AW, Piñeyro P, Li L, Fang L, Kiehne R, et al. Origin, Evolution, and Genotyping of Emergent Porcine Epidemic Diarrhea Virus Strains in the United States. *MBio*. 2013;4(5):1–8.
204. Fan B, Jiao D, Zhao X, Pang F, Xiao Q, Yu Z, et al. Characterization of Chinese Porcine Epidemic Diarrhea Virus

- with Novel Insertions and Deletions in Genome. *Nat Publ Gr*. 2017;7(44209):1–11.
205. Chen Q, Li G, Stasko J, Thomas JT, Stensland WR, Pillatzki AE, et al. Isolation and characterization of porcine epidemic diarrhea viruses associated with the 2013 disease outbreak among swine in the united states. *J Clin Microbiol*. 2014;52(1):234–43.
  206. Diep N Van, Sueyoshi M, Norimine J, Hirai T, Myint O, Teh APP, et al. Molecular characterization of US-like and Asian non-S INDEL strains of porcine epidemic diarrhea virus ( PEDV ) that circulated in Japan during 2013 – 2016 and PEDVs collected from recurrent outbreaks. *BMC Vet Res*. 2018;14(96):1–12.
  207. Wang L, Byrum B, Zhang Y. New Variant of Porcine Epidemic Diarrhea Virus , United States , 2014 To the Editor : Porcine epidemic. *Emerg Infect Dis CDC*. 2014;20(5):917–9.
  208. Chen Q, Gauger PC, Stafne MR, Thomas JT, Madson DM, Huang H, et al. Pathogenesis comparison between the United States porcine epidemic diarrhoea virus prototype and S-INDEL-variant strains in conventional neonatal piglets. *J Gen Virol*. 2016;97:1107–21.
  209. Marthaler D, Bruner L, Collins J, Rossow K. Third strain of porcine epidemic diarrhea virus, United states. *Emerg Infect Dis*. 2014;20(12):2162–3.
  210. Jarvis MC, Ching H, Zhang Y, Wang L, Hesse RA, Hause BM, et al. Genomic and evolutionary inferences between American and global strains of porcine epidemic diarrhea virus. *Prev Vet Med*. 2016;123:175–84.
  211. Sun M, Ma J, Wang Y, Wang M, Song W, Zhang W, et al. Genomic and Epidemiological Characteristics Provide New Insights into the Phylogeographical and Spatiotemporal Spread of Porcine Epidemic Diarrhea Virus in Asia. *J Clin Microbiol*. 2015;53(5):1484–92.
  212. Wang S, Cheng X, Chen S, Lin F, Jiang B, Zhu X, et al. Classification of emergent U.S. strains of porcine epidemic diarrhea virus by phylogenetic analysis of nucleocapsid and ORF3 genes. *J Clin Microbiol*. 2014;52(9):3509–10.
  213. Kim S, Nguyen V, Huynh T, Park Y, Park B-K, Chung H-C. Molecular Characterization of Porcine Epidemic Diarrhea Virus and Its New Genetic Classification Based on the Nucleocapsid Gene. *Viruses*. 2020;12(790):1–12.
  214. Toyoshima Y, Nemoto K, Matsumoto S, Nakamura Y, Kiyotani K. SARS-CoV-2 genomic variations associated with mortality rate of COVID-19. *J Hum Genet [Internet]*. 2020;65:1075–82. Available from: <http://dx.doi.org/10.1038/s10038-020-0808-9>
  215. Koyama T, Weeraratne D, Snowdon JL, Parida L. Emergence of Drift Variants That May A ff ect COVID-19 Vaccine Development and Antibody Treatment. *Pathogens*. 2020;9(324):1–7.
  216. Cardozo T. SARS-CoV-2 viral spike G614 mutation exhibits higher case fatality rate. *Int J Clin Pr*. 2020;00(e13525).
  217. Eaaswarkhanth M, Madhoun A Al, Al-mulla F. Could the D614G substitution in the SARS-CoV-2 spike ( S ) protein be associated with higher COVID-19 mortality ? *Int J Infect Dis*. 2020;96:459–60.
  218. Gonzalez-Reiche AS, Hernandez MM, Sullivan MJ, Ciferri B, Alshammary H, Obla A, et al. Introductions and early spread of SARS-CoV-2 in the New York City area. *Science (80- )*. 2020;369:297–301.
  219. Deng X, Deng X, Gu W, Federman S, Plessis L, Pybus OG, et al. Genomic surveillance reveals multiple introductions of SARS-CoV-2 into Northern California. *Science (80- )*. 2020;1–10.
  220. Forster P, Forster L, Renfrew C, Forster M. Phylogenetic network analysis of SARS-CoV-2 genomes. *PNAS*. 2020;117(17):9241–3.
  221. Rambaut A, Holmes EC, Toole ÁO, Hill V, Mccrone JT, Ruis C, et al. A dynamic nomenclature proposal for SARS-CoV-2 lineages to assist genomic epidemiology. *Nat Microbiol [Internet]*. 2020;5(November):1403–7. Available from: <http://dx.doi.org/10.1038/s41564-020-0770-5>
  222. Nextstrain. Genomic epidemiology of novel coronavirus - Global subsampling. 2020.
  223. Hadfield J, Megill C, Bell SM, Huddleston J, Potter B, Callender C, et al. Nextstrain: real-time tracking of pathogen evolution. *Bioinformatics [Internet]*. 2018 Dec 1;34(23):4121–3. Available from: <https://doi.org/10.1093/bioinformatics/bty407>
  224. Alm E, Broberg EK, Connor T, Hodcroft EB, Komissarov AB, Maurer-stroh S, et al. Geographical and temporal distribution of SARS-CoV-2 clades in the WHO European Region , January to June 2020. *Euro Surveill*. 2020;25(32).
  225. Tang X, Wu C, Li X, Song Y, Yao X, Wu X, et al. On the origin and continuing evolution of SARS-CoV-2. *Natl Sci Rev*. 2020;7:1012–23.
  226. Hsueh FC, Lin CN, Ying H, Min C, Chia Y, Tang M, et al. Updated phylogenetic analysis of the spike gene and identification of a novel recombinant porcine epidemic diarrhoea virus strain in Taiwan. *Transbound Emerg*

- Dis. 2019;00:1–14.
227. Li K, Song D, Zhang F, Gong W, Guo N, Li A, et al. Complete Genome Sequence of a Recombinant Porcine Epidemic Diarrhea Virus Strain, CH/JXJA/2017, Isolated in Jiangxi, China, in 2017. *Genome Announc.* 2018;6(6):e01590-17.
  228. Chen N, Li S, Zhou R, Zhu M, He S, Ye M, et al. Two novel porcine epidemic diarrhea virus ( PEDV ) recombinants from a natural recombinant and distinct subtypes of PEDV variants. *Virus Res [Internet]*. 2017;242(September):90–5. Available from: <http://dx.doi.org/10.1016/j.virusres.2017.09.013>
  229. Tian P-F, Jin Y-L, Xing G, Qv L-L, Huang Y-W, Zhou J-Y. Evidence of Recombinant Strains of Porcine Epidemic Diarrhea Virus, United States, 2013. *Emerg Infect Dis.* 2014;20(10):1735–8.
  230. Chung H-C, Lee J-H, Nguyen VG, Huynh TM Le, Leea G-E, Moon H-J, et al. New emergence pattern with variant porcine epidemic diarrhea viruses , South Korea , 2012 – 2015. *Virus Res.* 2020;226:14–9.
  231. Guo J, Fang L, Ye X, Chen J, Xu S, Zhu X, et al. Evolutionary and genotypic analyses of global porcine epidemic diarrhea virus strains. *Transbound Emerg Dis.* 2019;66:111–8.
  232. Qin S, Hu C, Yang D, Wu J, Yue H, Tang C, et al. Emergence of porcine epidemic diarrhea viruses with the novel S genes in Tibetan pigs in the Qinghai-Tibetan plateau in China. *Virus Res [Internet]*. 2019;270(June):197652. Available from: <https://doi.org/10.1016/j.virusres.2019.197652>
  233. Nefedeva M, Titov I, Malogolovkin A. Molecular characteristics of a novel recombinant of porcine epidemic diarrhea virus. *Arch Virol [Internet]*. 2019;164(4):1199–204. Available from: <https://doi.org/10.1007/s00705-019-04166-4>
  234. Boniotti MB, Papetti A, Lavazza A, Alborali G, Sozzi E, Chiapponi C, et al. Porcine Epidemic Diarrhea Virus and Discovery of a Recombinant Swine Enteric Coronavirus, Italy M. *Emerg Infect Dis CDC.* 2016;22(1):83–7.
  235. Wang H, Zhang L, Shang Y, Tan R, Ji M, Yue X, et al. Emergence and evolution of highly pathogenic porcine epidemic diarrhea virus by natural recombination of a low pathogenic vaccine isolate and a highly pathogenic strain in the spike gene. *Virus Evol.* 2020;6(2):1–10.
  236. Wang E, Guo D, Li C, Wei S, Wang Z, Liu Q, et al. Molecular Characterization of the ORF3 and S1 Genes of Porcine Epidemic Diarrhea Virus Non S-INDEL Strains in Seven Regions of. *PLoS One.* 2016;11(8):1–15.
  237. Mulligan MJ, Lyke KE, Kitchin N, Absalon J, Gurtman A, Lockhart S, et al. Phase 1/2 Study to Describe the Safety and Immunogenicity of a COVID-19 RNA Vaccine Candidate (BNT162b1) in Adults 18 to 55 Years of Age: Interim Report. *medRxiv [Internet]*. 2020 Jan 1;2020.06.30.20142570. Available from: <http://medrxiv.org/content/early/2020/07/01/2020.06.30.20142570.abstract>
  238. Jackson LA, Anderson EJ, Roupael NG, Roberts PC, Makhene M, Coler RN, et al. An mRNA Vaccine against SARS-CoV-2 — Preliminary Report. *N Engl J Med.* 2020;383(20):1920–31.
  239. Smith TRF, Patel A, Ramos S, Elwood D, Zhu X, Yan J, et al. Immunogenicity of a DNA vaccine candidate for COVID-19. *Nat Commun [Internet]*. 2020;11(1):2601. Available from: <https://doi.org/10.1038/s41467-020-16505-0>
  240. Zhu F-C, Li Y-H, Guan X-H, Hou L-H, Wang W-J, Li J-X, et al. Safety, tolerability, and immunogenicity of a recombinant adenovirus type-5 vectored COVID-19 vaccine: a dose-escalation, open-label, non-randomised, first-in-human trial. *Lancet [Internet]*. 2020 Jun 13;395(10240):1845–54. Available from: [https://doi.org/10.1016/S0140-6736\(20\)31208-3](https://doi.org/10.1016/S0140-6736(20)31208-3)
  241. Mercado NB, Zahn R, Wegmann F, Loos C, Chandrashekar A, Yu J, et al. Single-shot Ad26 vaccine protects against SARS-CoV-2 in rhesus macaques. *Nature [Internet]*. 2020;586(7830):583–8. Available from: <https://doi.org/10.1038/s41586-020-2607-z>
  242. van Doremalen N, Lambe T, Spencer A, Belij-Rammerstorfer S, Purushotham JN, Port JR, et al. ChAdOx1 nCoV-19 vaccine prevents SARS-CoV-2 pneumonia in rhesus macaques. *Nature [Internet]*. 2020;586(7830):578–82. Available from: <https://doi.org/10.1038/s41586-020-2608-y>
  243. Logunov DY, Dolzhikova I V, Zubkova O V, Tukhvatulin AI, Shcheblyakov D V, Dzharullaeva AS, et al. Safety and immunogenicity of an rAd26 and rAd5 vector-based heterologous prime-boost COVID-19 vaccine in two formulations: two open, non-randomised phase 1/2 studies from Russia. *Lancet [Internet]*. 2020 Sep 26;396(10255):887–97. Available from: [https://doi.org/10.1016/S0140-6736\(20\)31866-3](https://doi.org/10.1016/S0140-6736(20)31866-3)
  244. Keech C, Albert G, Cho I, Robertson A, Reed P, Neal S, et al. Phase 1–2 Trial of a SARS-CoV-2 Recombinant Spike Protein Nanoparticle Vaccine. *N Engl J Med.* 2020;383(24):2320–32.
  245. Xia S, Duan K, Zhang Y, Zhao D, Zhang H, Xie Z, et al. Effect of an Inactivated Vaccine Against SARS-CoV-2 on Safety and Immunogenicity Outcomes: Interim Analysis of 2 Randomized Clinical Trials. *JAMA [Internet]*. 2020

- Sep 8;324(10):951–60. Available from: <https://doi.org/10.1001/jama.2020.15543>
246. Wang H, Zhang Y, Huang B, Deng W, Quan Y, Wang W, et al. Development of an Inactivated Vaccine Candidate, BBIBP-CorV, with Potent Protection against SARS-CoV-2. *Cell* [Internet]. 2020 Aug 6;182(3):713-721.e9. Available from: <https://doi.org/10.1016/j.cell.2020.06.008>
  247. Gao Q, Bao L, Mao H, Wang L, Xu K, Yang M, et al. Development of an inactivated vaccine candidate for SARS-CoV-2. *Science*. 2020;369(6499):77–81.
  248. Seo SH, Jang Y. Cold-adapted live attenuated SARS-CoV-2 vaccine completely protects human ACE2 transgenic mice from SARS-CoV-2 infection. *bioRxiv* [Internet]. 2020;2:2020.08.04.235689. Available from: <https://doi.org/10.1101/2020.08.04.235689>
  249. Wang D, Fang L, Xiao S. Porcine epidemic diarrhea in China. *Virus Res*. 2016;226:7–13.
  250. Gerdt V, Zakhartchouk A. Vaccines for porcine epidemic diarrhea virus and other swine coronaviruses. *Vet Microbiol* [Internet]. 2017;206:45–51. Available from: <http://dx.doi.org/10.1016/j.vetmic.2016.11.029>
  251. Opriessnig T, Gerber PF, Shen H, De Castro AMMG, Zhang J, Chen Q, et al. Evaluation of the efficacy of a commercial inactivated genogroup 2b-based porcine epidemic diarrhea virus (PEDV) vaccine and experimental live genogroup 1b exposure against 2b challenge. *Vet Res*. 2017;48(1):1–10.
  252. Hou XL, Yu LY, Liu J, Wang GH. Surface-displayed porcine epidemic diarrhea viral (PEDV) antigens on lactic acid bacteria. *Vaccine*. 2007;26(1):24–31.
  253. Di-Qiu L, Jun-Wei G, Xin-Yuan Q, Yan-Ping J, Song-Mei L, Yi-Jing L. High-level mucosal and systemic immune responses induced by oral administration with *Lactobacillus*-expressed porcine epidemic diarrhea virus (PEDV) S1 region combined with *Lactobacillus*-expressed N protein. *Appl Microbiol Biotechnol*. 2012;93(6):2437–46.
  254. Meng F, Ren Y, Suo S, Sun X, Li X, Li P, et al. Evaluation on the Efficacy and Immunogenicity of Recombinant DNA Plasmids Expressing Spike Genes from Porcine Transmissible Gastroenteritis Virus and Porcine Epidemic Diarrhea Virus. *PLoS One*. 2013;8(3).
  255. Zhang Y, Zhang X, Liao X, Huang X, Cao S, Wen X, et al. Construction of a bivalent DNA vaccine co-expressing S genes of transmissible gastroenteritis virus and porcine epidemic diarrhea virus delivered by attenuated *Salmonella typhimurium*. *Virus Genes*. 2016;52(3):354–64.
  256. Huang JX, Bishop-Hurley SL, Cooper MA. Development of anti-infectives using phage display: biological agents against bacteria, viruses, and parasites. *Antimicrob Agents Chemother* [Internet]. 2012/06/04. 2012 Sep;56(9):4569–82. Available from: <https://pubmed.ncbi.nlm.nih.gov/22664969>
  257. Bertoglio F, Meier D, Langreder N, Steinke S, Rand U, Simonelli L, et al. SARS-CoV-2 neutralizing human recombinant antibodies selected from pre-pandemic healthy donors binding at RBD-ACE2 interface. *Nat Commun* [Internet]. 2021;12(1):1577. Available from: <https://doi.org/10.1038/s41467-021-21609-2>

## X. ANEXO

### X.I Publicación

Predicted 3D model of the M protein of Porcine Epidemic Diarrhea Virus and analysis of its immunogenic potential.

Rodríguez-Enríquez A, Herrera-Camacho I, Millán-Pérez-Peña L, Reyes-Leyva J, Santos-López G, Rivera-Benítez JF, Rosas-Murrieta NH.

*PLoS One*. 2022 Feb 9;17(2):e0263582. doi: 10.1371/journal.pone.0263582. eCollection 2022.

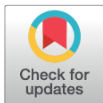
## RESEARCH ARTICLE

# Predicted 3D model of the M protein of Porcine Epidemic Diarrhea Virus and analysis of its immunogenic potential

Alan Rodríguez-Enríquez<sup>1,2\*</sup>, Irma Herrera-Camacho<sup>1\*</sup>, Lourdes Millán-Pérez-Peña<sup>1‡</sup>, Julio Reyes-Leyva<sup>3‡</sup>, Gerardo Santos-López<sup>3‡</sup>, José Francisco Rivera-Benítez<sup>4\*</sup>, Nora Hilda Rosas-Murrieta<sup>1\*</sup>

**1** Laboratory of Biochemistry and Molecular Biology, Chemistry Center, Institute of Science., Benemérita Universidad Autónoma de Puebla, Puebla, México, **2** Postgraduate in Chemical Sciences, Benemérita Universidad Autónoma de Puebla, Puebla, México, **3** Molecular Biology and Virology Laboratory, Centro de Investigación Biomédica de Oriente, Instituto Mexicano del Seguro Social (IMSS), Metepec, México, México, **4** Centro Nacional de Investigación Disciplinaria en Salud Animal e Inocuidad, Instituto Nacional de Investigaciones Forestales, Agrícolas y Pecuarias, Ciudad de México, México

\* These authors contributed equally to this work.  
‡ LMPP, JRL and GSL also contributed equally to this work.  
\* [nora.rosas@correo.buap.mx](mailto:nora.rosas@correo.buap.mx)

**OPEN ACCESS**

**Citation:** Rodríguez-Enríquez A, Herrera-Camacho I, Millán-Pérez-Peña L, Reyes-Leyva J, Santos-López G, Rivera-Benítez JF, et al. (2022) Predicted 3D model of the M protein of Porcine Epidemic Diarrhea Virus and analysis of its immunogenic potential. PLoS ONE 17(2): e0263582. <https://doi.org/10.1371/journal.pone.0263582>

**Editor:** Claude Krummenacher, Rowan University, UNITED STATES

**Received:** May 19, 2021

**Accepted:** January 23, 2022

**Published:** February 9, 2022

**Peer Review History:** PLOS recognizes the benefits of transparency in the peer review process; therefore, we enable the publication of all of the content of peer review and author responses alongside final, published articles. The editorial history of this article is available here: <https://doi.org/10.1371/journal.pone.0263582>

**Copyright:** © 2022 Rodríguez-Enríquez et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** All relevant data are within the paper and its [Supporting Information](#) files.

## Abstract

The membrane protein M of the Porcine Epidemic Diarrhea Virus (PEDV) is the most abundant component of the viral envelope. The M protein plays a central role in the morphogenesis and assembly of the virus through protein interactions of the M-M, M-Spike (S) and M-nucleocapsid (N) type. The M protein is known to induce protective antibodies in pigs and to participate in the antagonistic response of the cellular antiviral system coordinated by the type I and type III interferon pathways. The 3D structure of the PEDV M protein is still unknown. The present work exposes a predicted 3D model of the M protein generated using the Robetta protocol. The M protein model is organized into a transmembrane and a globular region. The obtained 3D model of the PEDV M protein was compared with 3D models of the SARS-CoV-2 M protein created using neural networks and with initial machine learning-based models created using trRosetta. The 3D model of the present study predicted four linear B-cell epitopes (RSVNASSGTG and KHGDYSAVSNPSALT peptides are noteworthy), six discontinuous B-cell epitopes, forty weak binding and fourteen strong binding T-cell epitopes in the CV777 M protein. A high degree of conservation of the epitopes predicted in the PEDV M protein was observed among different PEDV strains isolated in different countries. The data suggest that the M protein could be a potential candidate for the development of new treatments or strategies that activate protective cellular mechanisms against viral diseases.

**Funding:** Grant number: JFR-B and IH-C (FONSEC SADER-CONACYT 2017-06-292826, Consejo Nacional de Ciencia y Tecnología). NHR-M and IH-C (100429633-VIEP2021, Vice-Rectoría for Research and Postgraduate Studies [VIEP]). Alan Rodríguez Enríquez (BUAP: 219470378) was supported by CONACYT (2019-000037-02NACF). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing interests:** The authors have declared that no competing interests exist.

## Introduction

The Porcine Epidemic Diarrhea Virus (PEDV) is the causative agent of porcine epidemic diarrhea, a highly contagious disease that affects farm animals. The disease is diagnosed by clinical symptomatology such as vomiting, diarrhea, dehydration, and anorexia. The infection affects pigs of all ages, with piglets less than one week old being the most affected (100% morbidity and mortality). Although the virus does not affect human health, it causes significant economic losses to the pig breeding industry [1]. Current treatment is primarily aimed at diarrhea controlling. Control strategies have been oriented to the establishment in farms of rigorous biosecurity programs, adequate zootechnical management and vaccination [2]. Although these strategies have been effective in controlling disease, specific treatments are needed, which would require molecular knowledge of each stage of the viral replication cycle, as well as information on the biochemical characteristics of all the viral proteins. PEDV belongs to the order Nidovirales, family *Coronaviridae*, subfamily *Coronavirinae*, genus *Alphacoronavirus*. Its genome is a single-stranded, positive-sense RNA molecule that encodes 16 non-structural proteins (nsp 1–16) and 5 structural proteins, S, M, N, E and ORF3 [2]. Coronavirus M proteins share the same structural characteristics: three TM domains with N-exo-C-endo orientation flanked by a short glycosylated NH<sub>2</sub>-terminal domain on the virion surface and a long carboxy-terminal globular domain within the virion [3, 4]. The 3D structure of the PEDV M protein is still unknown. The functions described for the PEDV M protein include assembly and budding of virions, inhibition of innate immune response by blocking type I and III interferon pathways in infected cells and induction of protective antibodies in pigs [5–9]. The M protein performs its functions by interacting with other viral proteins. It is known that M proteins interact with each other and with spike and nucleocapsid proteins during virus assembly. Interestingly, M-M interactions form the overall scaffold for the viral envelope [3]. The indispensable role played by the M protein in the viral replication cycle can be explained in part by the lower genetic variability of the M gene compared to the S gene [10]. Therefore, the PEDV M protein could be a good candidate for developing specific antiviral drugs and as a target for vaccine design. The three-dimensional structure of the PEDV M protein has not been reported yet. The structure of some crystallized proteins of PEDV can be found in the PDB database, including the S protein [11], 3CLpro protease [12], NSP9 [13], papain-like protease (PDB: 6NOZ) and NSP1 (PDB 5XBC). Due to the importance of the M protein in the viral infection cycle, the present work proposes a 3D model of the M protein generated using the Robetta protocol. An immunoinformatics analysis was also carried out to determine the potential antigenic regions of the M protein.

## Material and methods

### Sequences

For the generation of the 3D model of the M protein, the sequence of the M protein of the PEDV CV777 virulent strain (Uniprot ID: P59771) (226 residues) and the M protein from strain PEDV/MEX/MICH/01/2013 (GenBank number: MH006957.1) isolated in Michoacán, Mexico, called 2013MMV were used. Different sequences of the M protein were retrieved from the NCBI Taxonomy Browser page (<https://www.ncbi.nlm.nih.gov/Taxonomy/Browser/wwwtax.cgi>). The sequence of the SARS-CoV-2 M protein of the AlphaFold and Feig lab models corresponds to Uniprot ID: P0DTC5.

### Cloning sequence of the M gen and phylogenetic analyses

The nucleotide sequence of the 2013MMV was subcloned in the vector pPICZαB from the vector pETSumo and it was called 2013MMV. The recombinant DNA was sequenced, 225 amino