



UNIVERSIDAD PERUANA
CAYETANO HEREDIA

**DESARROLLO DE UNA PROTEÍNA
MULTIEPITÓPICA MEDIANTE
ANÁLISIS INMUNOINFORMÁTICO
PARA EL INMUNODIAGNÓSTICO DE
NOROVIRUS GII.4**

TESIS PARA OPTAR EL GRADO DE
MAESTRO EN BIOQUÍMICA Y BIOLOGÍA
MOLECULAR

JUAN ALEJANDRO ALMERCO FRANCO

LIMA - PERÚ

2017

ASESOR

DR. MIRKO JUAN ZIMIC PERALTA

JURADO DE TESIS

PRESIDENTE

DRA. ROSA PATRICIA HERRERA VELIT

VOCAL

DR. HOLGER MAYTA MALPARTIDA

SECRETARIO

MG. CARLA MARIA GALLO LOPEZ-ALIAGA.

DEDICATORIA

A mis padres, por ser mi motivo para seguir adelante.

AGRADECIMIENTOS

Al Consejo Nacional de Ciencia y Tecnología por financiar el presente trabajo de investigación.

Al Dr. Mirko Zimic por su asesoramiento y sugerencias, y por darme la oportunidad de participar en este estudio mediante el desarrollo de la presente tesis. De igual manera a la Dra. Patricia Sheen, por permitirme ser parte del Laboratorio de Bioinformática y Biología Molecular.

A Manuel Ramírez, Hugo Valdivia y David Requena, por sus comentarios y apoyo en el análisis bioinformático de esta investigación.

A José Luis Roldan, por enseñarme y guiarme en mis primeras pruebas experimentales en el laboratorio.

A todos mis compañeros y amigos del Laboratorio de Bioinformática y Biología Molecular, por su tiempo y consejos en el desarrollo de mi trabajo.

Y de manera muy especial a Edith Málaga por su gran apoyo y colaboración en el tramo final de este gran reto.

FUENTES DE FINANCIAMIENTO

La realización del presente trabajo de investigación para optar el grado de Maestro en Bioquímica y Biología Molecular ha sido posible gracias al financiamiento brindado por el Programa de Maestría en Bioquímica y Biología Molecular de la Universidad Peruana Cayetano Heredia, subvencionado por FONDECYT-CONCYTEC, según Convenio de Gestión N°014-2013-FONDECYT.

TABLA DE CONTENIDOS

RESUMEN

ABSTRACT

I. INTRODUCCIÓN	1
II. PLANTEAMIENTO DEL PROBLEMA	4
III. MARCO TEÓRICO	7
3.1 Virología Molecular	7
3.1.1 Genoma.....	7
3.1.2 Estructura de la cápside	8
3.1.3 Clasificación	9
3.1.4 Evolución y variabilidad antigénica	10
3.2 Patogénesis	12
3.2.1 Susceptibilidad Genética	12
3.2.2 Respuesta Inmune.....	13
3.3 Epidemiología	14
3.4 Diagnóstico.....	17
3.4.1 Microscopía Electrónica	17
3.4.2 Real-time PCR.....	17
3.4.3 Inmunoensayo.....	18
3.5 Inmunoinformática	18

3.5.1 Predicción de epítopes para células B	18
3.5.2 Predicción de afinidad a moléculas MHC-II	21
3.5.3 Diseño de proteínas multiepitópicas.....	23
IV. JUSTIFICACIÓN DEL ESTUDIO.....	24
V. HIPÓTESIS	25
VI. OBJETIVOS.....	26
6.1 Objetivo general	26
6.2 Objetivos específicos.....	26
VII. METODOLOGÍA	27
7.1 Análisis Immunoinformático.....	27
7.1.1 Alineamiento de secuencias.....	27
7.1.2 Análisis filogenético	28
7.1.3 Modelamiento por homología	29
7.1.4 Predicción de epítopes para células B	30
7.1.5 Predicción de afinidad a MHC clase II.....	31
7.1.6 Conservación y especificidad	32
7.1.7 Diseño del gen para la proteína multiepitópica MEP-GII.4	32
7.2 Producción de la proteína multiepitópica MEP-NVGII.4.....	33
7.2.1 Sistema de expresión pET28a-MEP-NVGII.4	33
7.2.2 Transformación de células de <i>E. coli</i> BL21(DE3)	34
7.2.3 Caracterización del vector pET28a-MEP-NVGII.4	35

7.2.4 Expresión de la proteína MEP-NVGII.4	35
7.2.5 Cromatografía de afinidad	37
7.2.6 Método de Bradford para cuantificación de proteína	37
7.2.7 Concentración de la proteína MEP-NVGII.4	38
7.3 Desarrollo de un método de ELISA indirecto	38
7.3.1 Muestras de suero humano	38
7.3.2 Desarrollo de un método de ELISA indirecto	39
7.3.3 Detección de anticuerpos específicos IgG e IgM	41
7.3.4 Análisis de datos	41
VIII. RESULTADOS	43
8.1 Análisis de secuencias	43
8.2 Modelamiento estructural de la proteína VP1	48
8.3 Predicción de epítopes B lineales.....	53
8.4 Predicción de afinidad a MHC Clase II	54
8.5 Conservación y especificidad de los epítopes	56
8.6 Evaluación de colonias de <i>E. coli</i> BL21(DE3) transformadas con el vector pET28a-MEP-NVGII.4	57
8.7 Inducción y optimización de la expresión.....	58
8.8 Purificación de la proteína multiepitópica	60
8.9 Detección de anticuerpos IgG e IgM.....	62
8.10 Análisis estadístico	63

IX. DISCUSIÓN	65
9.1 Predicción inmunoinformática	65
9.2 Reactividad de la proteína MEP-NVGII.4 frente a sueros humanos	70
X. CONCLUSIONES.....	74
XI. RECOMENDACIONES.....	75
XII. REFERENCIAS BIBLIOGRÁFICAS.....	76

LISTA DE TABLAS

Tabla 1. Sitios informativos en la secuencia de la proteína VP1	46
Tabla 2. Evaluación de los modelos 3D de la proteína VP1 de NV GII.4 generados por homología.	51
Tabla 3. Epítopes predichos para células B de NV GII.4 con afinidad a moléculas de MHC clase II	54

LISTA DE FIGURAS

Figura 1. Estructura del genoma de norovirus.....	7
Figura 2. Aspectos estructurales de la cápside de norovirus.....	9
Figura 3. Mapa del vector pET28a-MEP-NVGII.4.	34
Figura 4. Análisis filogenético de máxima verosimilitud de 716 secuencias completas de la proteína VP1 de NV GII.4.....	44
Figura 5. Sitios hipervariables del subdominio P de la proteína VP1, identificados en secuencias representativas de NV GII.4 desde el año 1975 hasta el 2015.....	47
Figura 6. Comparación de los valores de energía DOPE del modelo por homología de la proteína VP1 de NV GII.4 Sydney 2012.....	49
Figura 7. Superposición de las estructuras de la proteína VP1 entre variantes de NV GII.4.	50
Figura 8. Mapeo de la variación de sitios informativos en la estructura 3D de la proteína VP1 de NV GII.4 entre el año 1970 y el 2015.....	52
Figura 9. Mapeo de regiones epitópicas en la estructura 3D de la proteína VP1 de NV GII.4 variante Sydney 2012.....	53
Figura 10. Electroforesis en gel de agarosa de la digestión del plásmido pET28a-MEP-NVGII.4, empleando las enzimas de restricción NcoI y XhoI.....	57
Figura 11. SDS-PAGE de la proteína MEP-NVGII.4 después de la inducción de su expresión con 1 mM de IPTG.	59

Figura 12. Fracción soluble e insoluble de la proteína MEP-NVGII.4 expresada luego de 5 horas de inducción con 1 mM de IPTG.....	60
Figura 13. SDS-PAGE de las fracciones de purificación de la proteína MEP-NVGII.4 por cromatografía de afinidad.....	61
Figura 14. Gráfica de puntos de los valores de OD _{450nm} de anticuerpos IgG frente a la proteína MEP-NVGII.4.	63
Figura 15. Gráfica de puntos de los valores de OD _{450nm} de anticuerpos IgM frente a la proteína MEP-NVGII.4.	64

LISTA DE ABREVIATURAS

°C	Grados Celsius
6xHis	6-Histidinas
μL	Microlitro
Å	Ángstrom
ANN	Artificial neural network
AND	Ácido desoxirribonucleico
ARN	Ácido ribonucleico
BLAST	Basic Local Alignment Search Tool
cADN	Ácido desoxirribonucleico complementario
DOPE	Discrete Optimized Protein Energy
<i>E. coli</i>	<i>Escherichia coli</i>
EE.UU.	Estados Unidos de América
ELISA	Enzyme-linked immunosorbent assay
FUT2	α-1,2-fucosiltransferasa
HBGA	Histo-blood group antigens
HLA	Human leukocyte antigen
HRP	Horseradish peroxidase
IC ₅₀	Half-maximal inhibitory concentration
IEDB	Immune Epitope Database
IFN-γ	Interferón γ
IgA	Inmunoglobulina A

IgG	Inmunoglobulina G
IgM	Inmunoglobulina M
IL	Interleucina
IMAC	Immobilized metal affinity chromatography
IME	Inmunomicroscopía electrónica
IPD	Immuno Polimorphism Database
IPTG	isopropil- β -D-1-tiogalactopiranosido
kb	Kilobase
kDa	Kilodalton
LB	Luria-Bertani
ME	Microscopía electrónica
MHC	Major histocompatibility complex
mL	Mililitro
mM	Milimolar
nm	Nanómetro
NCBI	National Center for Biotechnology Information
NV	Norovirus
OD _{450nm}	Densidad óptica a 450 nm
OD _{600nm}	Densidad óptica a 600 nm
ORF	Open reading frame
PBMC	Peripheral blood mononuclear cells
PBS	Phosphate-buffered saline
PDB	Protein Data Bank
Poli-A	Poliadenina

qPCR	quantitative polymerase chain reaction
RIA	Radioinmunoensayo
RMSD	Root-mean-square deviation
RT-PCR	Reverse transcription polymerase chain reaction
SDS-PAGE	Sodium dodecyl sulfate–polyacrylamide gel electrophoresis
SMV	Snow Mountain virus
Th	Linfocito T helper
TMB	3,3',5,5'-tetrametilbencidina
TNF- α	Factor de necrosis tumoral alfa
VLP	Virus-like particles

RESUMEN

Norovirus es la principal causa de gastroenteritis aguda no bacteriana a nivel mundial, siendo el genotipo GII.4 el responsable de la mayoría de casos reportados hasta la actualidad. Debido a una falta de tratamiento farmacológico, un diagnóstico temprano toma un papel importante. Actualmente, los métodos serológicos para la detección de anticuerpos específicos utilizan partículas tipo virus, las cuales representan un límite para los laboratorios en condiciones de bajos recursos. El presente estudio tuvo como objetivo diseñar *in silico* una proteína multiepitópica para la detección de anticuerpos IgG e IgM contra norovirus GII.4 en suero humano. Un total 20 epitopes inmunodominantes de la proteína VP1 de la cápside de norovirus GII.4 fueron predichos mediante análisis inmunoinformático, de los cuales 10 fueron seleccionados para construir el gen sintético *MEP-NVGII.4*. Este gen recombinante fue expresado en *E. coli* BL21, y la proteína recombinante, MEP-NVGII.4, purificada y empleada como antígeno para el desarrollo de un ELISA indirecto para la detección de IgG e IgM en un grupo de 26 sueros de bebés con edades comprendidas entre cero y 2 años. Se observaron valores similares entre los niveles séricos de IgG de los individuos infectados y los controles. Sin embargo, se obtuvo una diferencia estadística significativamente alta ($p < 0.0001$) al comparar los niveles séricos de IgM. MEP-NVGII.4 demostró una potencial utilidad para estudios de seroprevalencia como de diagnóstico de infección aguda y convaleciente por norovirus GII.4.

Palabras clave: Norovirus GII.4, proteína multiepitópica, predicción de epítopes, inmunoinformática, inmunodiagnóstico, ELISA.

ABSTRACT

Norovirus is the leading cause of acute non-bacterial gastroenteritis worldwide, with the GII.4 genotype responsible for the majority of cases reported to date. Due to a lack of drug treatment, an early diagnosis takes an important role. Currently, serological methods for the detection of specific antibodies use virus-like particles, which represent a limit for laboratories in low-resource conditions. The present study aimed to design a multiepitope protein *in silico* for the detection of IgG and IgM antibodies against Norovirus GII.4 in human serum. A total of 20 immunodominant epitopes of the VP1 protein of the norovirus GII.4 capsid were predicted by immunoinformatic analysis, of which 10 were selected to construct the synthetic gene *MEP-NVGII.4*. This recombinant gene was expressed in *E. coli* BL21, and the recombinant protein, MEP-NVGII.4, purified and used as an antigen for the development of an indirect ELISA for the detection of IgG and IgM in a group of 26 sera of babies aged between zero and 2 years. Similar values were observed between serum IgG levels of infected individuals and controls. However, a significantly high statistical difference ($p < 0.0001$) was obtained when comparing serum IgM levels. MEP-NVGII.4 demonstrated a potential utility for seroprevalence studies as a diagnosis of acute and convalescent infection with norovirus GII.4.

Keywords: Norovirus GII.4, multiepitope protein, epitope prediction, immunoinformatics, immunodiagnostics, ELISA.

I. INTRODUCCIÓN

A principios del siglo pasado, en el año 1929, el doctor John Zahorsky realizó una investigación sobre casos esporádicos de vómitos y diarrea aguda en niños, los cuales fueron descritos como la enfermedad emética del invierno (*Hyperemesis hiemis*) (1). La etiología que originaba dicha enfermedad era en aquel entonces desconocida, sin embargo, la contagiosidad de la misma fue evidenciada gracias a la información recopilada de experimentos con personas voluntarias y brotes epidémicos con características similares.

Años más tarde, uno de aquellos brotes aconteció en una escuela, esta vez la de una pequeña localidad llamada Norwalk, en Ohio (EE.UU.), en el cual se reportaron 116 casos de un total de 232 individuos, entre alumnos y trabajadores, durante dos días de octubre de 1968. Además, hubo bastantes casos secundarios entre los familiares, donde se evidenció cuadros de vómitos y diarrea. Este episodio fue diagnosticado como brote de la enfermedad emética del invierno, y el agente causal, denominado en este caso “agente Norwalk”, permaneció aun sin ser identificado (2).

Sin embargo, algunas muestras de origen fecal obtenidas en aquel evento fueron guardadas y entregadas pocos años después a otro grupo de investigadores dirigidos por Albert Kapikian. En esta oportunidad, se estudiaron las heces de voluntarios a quienes se les había administrado el filtrado de heces, libre de bacterias, procedente del brote de 1968. Estas muestras fueron analizadas mediante inmunomicroscopía electrónica (IME), la cual permite observar directamente la presencia de

inmunocomplejos debido a la interacción de anticuerpos específicos presentes en el suero de individuos convalecientes con partículas víricas. Es así como en 1972 logró identificarse unos conglomerados de partículas redondas estructuradas de pequeño tamaño (27 nm) que resultaron ser el agente causal de aquel brote al que se le denominó virus Norwalk (3).

A la par de los grandes avances que se han producido sobre el conocimiento de este agente viral, el cual actualmente es llamado norovirus, su importancia a nivel de salud pública ha alcanzado una escala mundial. Se estima que norovirus, miembro perteneciente a la familia *Caliciviridae*, constituye la principal causa de enfermedad diarreica a nivel mundial, caracterizada por presentar síntomas como la diarrea, vómitos, dolor abdominal, cefalea, etc. (4). La mayor incidencia de infección por norovirus ocurre en niños menores de cinco años, y ocasiona la muerte de 70000 niños cada año, lo que constituiría la causa de muerte por diarrea más común en esta población (5).

Dentro de este género se han clasificado seis genogrupos, desde el GI hasta el GVI, y - basados en la gran variabilidad genética de la proteína de la cápside- cada genogrupo se conforma de distintos genotipos, siendo el genotipo GII.4 el que más afecta a los seres humanos tanto en países desarrollados como en vías de desarrollo.

La posibilidad de generar y amplificar el cADN viral a través de la técnica de RT-PCR la ha conllevado a ser el método de elección para detectar la presencia de norovirus en muestras fecales (6). Posteriormente, la expresión de proteínas de la cápside viral en sistemas de baculovirus y la formación de partículas virales contribuyeron con el conocimiento y uso de antígenos y anticuerpos para el

desarrollo de ensayos inmunológico. Sin embargo, debido al empleo de numerosos equipos de alto costo, el requerimiento de personal adecuadamente entrenado para realizar esa labor, y el continuo cambio a nivel genómico y antigénico propio de carácter viral, han conllevado a la búsqueda de nuevas herramientas para mejorar la eficiencia y rapidez en la detección de norovirus.

Recientemente, el avance de la bioinformática y la inmunoinformática nos permite analizar comparativamente a los epítopes situados en la superficie de la cápside viral y de esta manera producir antígenos sintéticos que predictivamente puedan poseer la capacidad de ser reconocidos por anticuerpos séricos. Es por ello que el objetivo del presente estudio fue evaluar bioinformáticamente las secuencias de la proteína VP1 de norovirus GII.4 reportadas hasta la fecha, identificar y evaluar regiones antigénicas basadas en epítopes inmunodominantes predichos, y diseñar una proteína multiepitópica para su uso en el desarrollo de una nueva y más rápida metodología de detección de anticuerpos séricos IgG e IgM.

II. PLANTEAMIENTO DEL PROBLEMA

En los últimos años se han realizado importantes avances en el conocimiento de las características moleculares y evolutivas del género norovirus en seres humanos. El manejo óptimo de los brotes de norovirus GII.4 así como la necesidad de realizar estudios de prevalencia que aporten a la información epidemiológica en regiones más vulnerables deben incluir un rápido diagnóstico del agente. En el ámbito hospitalario se puede limitar la diseminación de la infección mediante la rápida implementación de medidas de control tales como restricciones de personal, limpieza de superficies con desinfectantes adecuados y mantenimiento de las áreas afectadas. Por otro lado, dado que los brotes se asocian con el contagio a través del agua y los alimentos contaminados, la detección temprana de la presencia de este patógeno en zonas socioeconómicamente afectadas podría prevenir la reaparición de casos de infección en personas anteriormente expuestas. Es por ello que sería deseable disponer de métodos rápidos, sensibles, específicos y no muy costosos, los cuales hagan más asequible la detección de estos virus a la mayoría de los laboratorios.

Hasta la fecha, la RT-PCR ha sido la técnica de referencia en cuanto a la detección de norovirus del genogrupo GI y GII en muestras de heces. Sin embargo, este método molecular suele quedar fuera del alcance de muchos laboratorios y, por otro lado, la variabilidad genética de norovirus, en especial el genogrupo GII.4, hace difícil el desarrollo de una prueba de amplificación simple y genérica.

En cuanto al uso de técnicas de inmunoensayo, el empleo de partículas virales en la gran mayoría de kits comerciales y técnicas experimentales, generadas por expresión mediante baculovirus, simboliza un alto costo en cuanto a su producción y manejo. Sin embargo, dada la naturaleza biológica del género norovirus, su alta tasa de mutación y por ende, su gran variabilidad a través del tiempo, los esfuerzos por generar medidas preventivas tales como vacunas o el diseño de fármacos antivirales capaces de bloquear directamente la replicación han dado pocos o ningún resultado satisfactorio. Por esta razón, el conocimiento del patrón evolutivo de norovirus GII.4, lo cual va de la mano con el origen de las epidemias reportadas en todo el mundo desde el brote reportado en 1972 en Norwalk, Ohio, es de vital importancia porque permite identificar los cambios a nivel funcional y estructural que se dan en este patógeno con el fin de adaptarse y, luego de cada cierta temporada, reaparecer con un fenotipo distinto.

Recientemente, el aporte de herramientas para análisis bioinformático e inmunoinformático ha permitido no solo evaluar un amplio número de datos de secuencias del genoma y proteoma de diversos patógenos, sino también diseñar moléculas que puedan reemplazar un antígeno en el proceso de producción y/o detección de anticuerpos. Es por esta razón que se han publicado varias propuestas sobre el desarrollo de proteínas que contengan diversos epitopes conservados para el inmunodiagnóstico de enfermedades infecciosas. Sin embargo, a la fecha, no hay antecedentes de una metodología similar para la detección de norovirus GII.4.

Por lo tanto, el presente trabajo plantea responder la siguiente pregunta de investigación: ¿Se podrá generar una proteína multiepitópica recombinante, basada en diversos epitopes inmunodominantes de la proteína de la cápside viral de norovirus GII4, con capacidad de detección de anticuerpos IgG e IgM en suero humano?

III. MARCO TEÓRICO

3.1 Virología Molecular

3.1.1 Genoma

A pesar de su temprano descubrimiento, el estudio de la biología molecular del género norovirus (NV) permaneció retrasado por mucho tiempo debido al escaso material genético que se obtenía de personas infectadas. No fue sino hasta el año 1990 en que Jiang et al. lograron clonar y secuenciar el cADN del genoma de NV (7), dando paso a grandes avances en la comprensión de su organización genómica.

El genoma de NV está conformado por un ARN de cadena simple y en sentido positivo, de aproximadamente 7.7 kb de longitud, unido a una cadena de poli-A en el extremo 3', y que codifica 3 marcos abiertos de lectura (ORF) (8) (Figura 1).

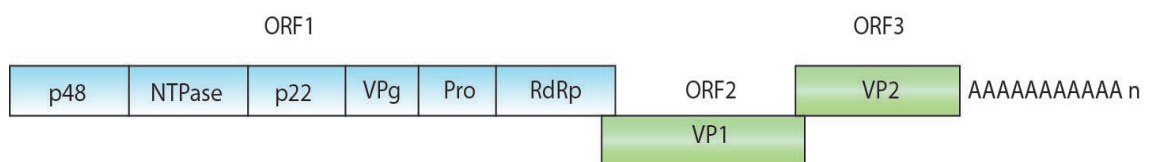


Figura 1. Estructura del genoma de norovirus.

El ORF1 comprende alrededor de los dos tercios del genoma (~5 kb) y codifica a un polipéptido (200 kDa) que es proteolíticamente procesado para dar origen a 6 proteínas no estructurales; la proteína N-terminal p48 (37-48 kDa), la cual promueve el desmontaje del aparato de Golgi e induce cambios en la membrana intracelular (9); NTPasa (40 kDa), una nucleosido trifosfatasa (10); p22 (20 kDa), la cual inhibe la vía secretora de la célula (11); Vpg (16 kDa), una proteína unida al

genoma viral y que participa tanto en la replicación del genoma viral (12) como en su traducción (13); Pro (19 kDa), una proteasa similar a quimotripsina (14); y RdRp (57 kDa), una ARN polimerasa dependiente de ARN (15). El ORF2 (1,8 kb) codifica a la proteína mayoritaria de la cápside de NV (57 kDa), la cual conforma la superficie externa del virión, mientras que el ORF3 (0,6 kb) codifica a la proteína minoritaria estructural (22 KDa) (16).

3.1.2 Estructura de la cápside

A inicios de los años 90 se evidenció que la proteína estructural VP1, de 530 a 555 aminoácidos y con un peso molecular de aproximadamente 58 kDa, tenía la propiedad de autoensamblarse en partículas tipo virus (*virus-like particles*, VLP) cuando era expresada mediante el sistema de baculovirus, manteniendo éstas las propiedades morfológicas del virus nativo (17). La estructura de VP1 se caracteriza por poseer dos dominios principales: los primeros 225 aminoácidos desde el extremo N-terminal componen el dominio S (*shell*), el cual está involucrado en la formación interna de la estructura icosaédrica, mientras que el resto de residuos hasta el extremo C-terminal constituyen el dominio protuberante P (*protruding*). Este dominio a su vez está conformado por dos subdominios, P1, constituido desde el residuo 226 al 278 y del 406 al 530; y P2, constituido del residuo 279 al 405, siendo este una inserción en el subdominio P1 (Figura 2) (18).

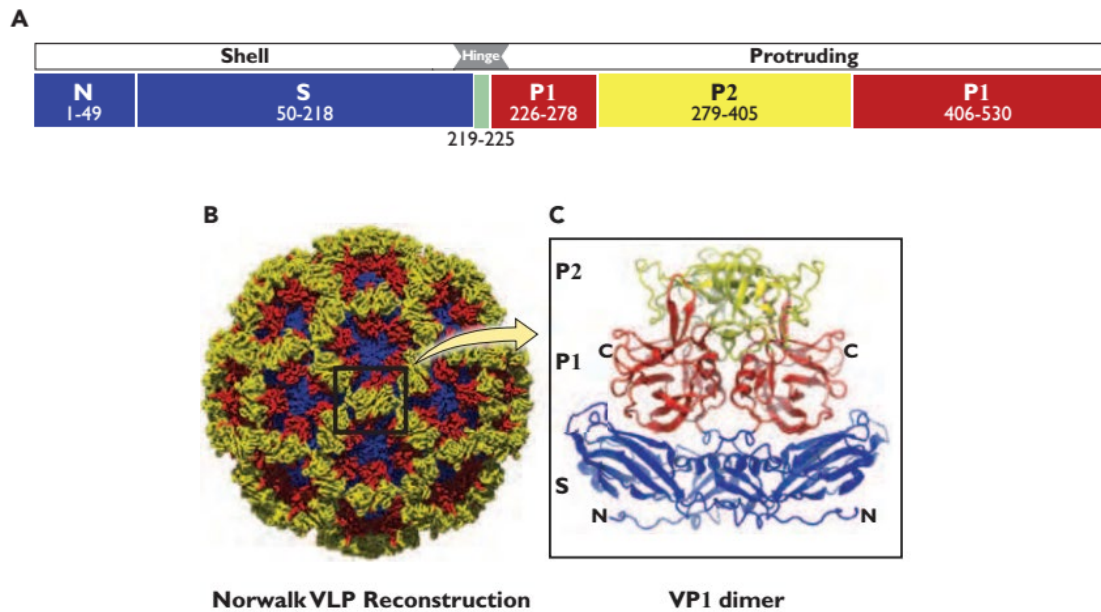


Figura 2. Aspectos estructurales de la cápside de norovirus. **A.** Dominios que conforman la proteína VP1: región brazo N-terminal (N), dominio shell (S), y dominio saliente o protruding (P). **B.** Modelo completo de la estructura T = 3 de la cápside de NV, determinada a 3.4 Å de resolución. **C.** Representación tridimensional en cintas de un dímero de la proteína VP1, mostrando el dominio S, el subdominio P1 y el subdominio P2 de color azul, rojo y amarillo respectivamente. Se indican además las regiones N-terminal (N) y C-terminal (C). (Adaptado de Green, 2013) (19)

3.1.3 Clasificación

Pese al desarrollo reciente de un cultivo de NV humano en enterocitos derivados de células madre (20), la falta de un modelo in vitro de replicación durante años anteriores había impedido determinar un sistema de clasificación mediante serotipificación (21). Es por ello que las técnicas de RT-PCR y secuenciamiento genómico mostraron una importancia notable para dicho propósito. En base a la

secuencia de la proteína VP1, se han establecido genéticamente 6 genogrupos (GI a GVI): GI y GII son los genogrupos responsables de la mayoría de infecciones que afectan al ser humano; GIII afecta al ganado bovino; GIV en especies felinas y caninas, aunque raramente se ha detectado también en seres humanos; GV, engloba a cepas murinas; y finalmente, GVI, en caninos (22). Un genogrupo VII presente en caninos ha sido recientemente propuesto, no obstante, se requieren aún de mayores hallazgos para su sustento. Cada genogrupo se puede subdividir en genotipos, los cuales son asignados al haber un rango de distancia entre par de secuencias (pairwise distance) de 14-43.8 % (23,24). Se han reconocido 9 genotipos en GI y 22 en GII; se ha registrado un mayor número de casos de infección por norovirus cuando la cepa proviene del genogrupo GII (89% de los casos) en comparación con el genogrupo GI.

A pesar de toda la gran variabilidad expuesta, el genotipo GII.4 es el principal causante de brotes infecciosos a nivel global debido a su evolución por épocas que ha dado lugar a la aparición de nuevas variantes genéticas y, por ende, a la persistencia que tiene en la población (25). La cepa denominada Sidney 2012 es la más prevalente en años recientes (22).

3.1.4 Evolución y variabilidad antigénica

Diversos estudios han identificado residuos aminoacídicos en la cápside viral de NV que son importantes tanto para su reconocimiento por anticuerpos específicos como para la interacción con receptores HBGA (*histo-blood group antigens*) del hospedero. Los primeros hallazgos de zonas antigénicas específicas para anticuerpos monoclonales diseñados contra VLPs de NV GII.4 determinaron que el subdominio P2 de la proteína VP1 presentaba epitopes localizados en distintas

posiciones (26), conteniendo uno de ellos un motivo similar a la secuencia RGD altamente conservado entre los residuos 288 y 290 (27). Un análisis de diversidad genética, abarcando distintos brotes desde el año 1987 al 2006 a lo largo de la secuencia de VP1, identificó la presencia de residuos altamente variables (6, 9, 355, 372, 393, 394, 395, 412, 505 y 534) (28) y sugirió que estarían en un proceso de selección en respuesta a la inmunidad del hospedero. Por otro lado, Allen et al. (29,30) determinó la presencia de dos zonas altamente variables, denominadas sitio A (296-298) y sitio B (393-395), los cuales fueron ubicados en la superficie de bucles expuestos; el sitio B cumpliría un rol importante en la especificidad de interacción con los residuos glucosídicos de los receptores HBGA a nivel de genogrupo (31). Otros estudios más recientes reconocieron residuos con capacidad antigénica en posiciones similares (32–35).

Se ha observado que los cambios a nivel del subdominio P2 determinarían la aparición de una nueva epidemia en un modelo de evolución periódica, en el cual existen periodos estacionarios de baja variabilidad seguidos de la rápida aparición de nuevas variantes (31,36). Esto se asoció a un proceso de deriva antigénica, donde se da una acumulación de mutaciones en las secuencias que codifican para los sitios de unión a anticuerpos, derivando en cambios a nivel antigénico y, por lo tanto, evadiendo de esta manera la respuesta del sistema inmune incluso en individuos que hayan sido expuestos anteriormente a la infección con el mismo genotipo de NV (37). Esta variabilidad propia del genotipo GII.4 está reflejada en su una alta tasa de mutación y una fidelidad de replicación baja en comparación con otros genogrupos (38,39). Un caso similar ocurre con el virus de la influenza, el cual

modifica los residuos epitópicos de la proteína hemaglutinina necesaria para la interacción con su hospedero (40).

Sin embargo, Tan et al. 2009 mostró que las secuencias primarias de los sitios de unión a receptores HBGA son altamente conservados entre las distintas cepas dentro de cada genogrupo. De esta manera, planteó que la diversidad tanto genética como antigénica de NV está influenciada tanto por un proceso de evolución convergente, donde cepas con los mismos receptores blancos compartirán una similitud estructural a pesar de los cambios en la composición aminoacídica, como por un proceso de evolución divergente, regida primordialmente por la deriva antigénica producto del escape de la inmunidad humana, y que produce la aparición de nuevos episodios epidémicos y su persistencia a través del tiempo (41).

3.2 Patogénesis

3.2.1 Susceptibilidad Genética

Los primeros estudios de patogénesis y respuesta inmunológica a NV realizados en personas voluntarias demostraron que algunos individuos eran más susceptibles de padecer la infección que otros (42). La búsqueda de factores de susceptibilidad o resistencia del hospedero asociados con la infección por NV fueron apoyados con el uso de distintas líneas celulares para estudios de interacción e internalización de partículas virales, determinando que las células intestinales CaCo2, una línea de carcinoma de colon humano, poseen una mayor capacidad de unión a la cápside (43). La presencia de un receptor específico de entrada para NV fue relacionado con estudios realizados en otros miembros de la familia *Caliciviridae* (44), los cuales mostraron tener afinidad por los antígenos de grupo sanguíneo o HBGA.

Hutson et al. 2002 determinó que el fenotipo ABO de un individuo, el cual se encuentra no solo en glóbulos rojos sino también en células epiteliales, jugaba un rol importante en la susceptibilidad a la infección por NV (45). Algunas cepas de NV infectan individuos quienes llevan el gen que codifica la enzima α -1,2-fucosiltransferasa (FUT2), designándose como tipo secretor para indicar que ellos expresan los antígenos sanguíneos ABH. En cambio, aquellos individuos que son deficientes en el gen FUT2 son denominados no secretores, por lo tanto, no expresan los receptores HBGA apropiados para una interacción adecuada con el patógeno, y generan un grado de resistencia a la infección (46,47).

3.2.2 Respuesta Inmune

Desde 1970 se han llevado a cabo algunos estudios clínicos consistentes en los cuales se administró NV a individuos sanos, lo que permitió evaluar el orden temporal en el cual aparecen los diferentes isotipos de anticuerpos específicos para NV (48–51). Se ha reportado que los niveles séricos de IgG, IgM e IgA pueden ser detectados luego de 8 a 11 días de ocurrida la infección (49,51); IgG alcanza un máximo luego de 3 semanas, a diferencia de IgM e IgA que son detectados generalmente durante la segunda semana (48,50). Los anticuerpos IgG muestran una tendencia a permanecer durante mayor tiempo y a niveles más elevados comparados con los niveles de IgM e IgA, manteniendo un valor máximo constante más allá de los 90 días luego de los cuales, declina (49). En cambio, los niveles IgM retornan a sus valores anteriores a la infección luego de 2 a 4 meses (50), mientras que los niveles de IgA persisten por 1-2 meses y descienden luego de dicho período (48). Por otro lado, se ha observado que la respuesta inducida de anticuerpos contra NV puede presentar una alta reactividad cruzada, siendo mayor la capacidad de

unirse a diversas cepas de NV que pertenezcan a un mismo genogrupo en comparación con aquellas de genogrupos diferentes (52).

En un estudio realizado en voluntarios sanos con evidencia previa de infección por NV GII.4 se reportó que las células mononucleares de sangre periférica (PBMC) expresaron a los marcadores de activación CD80, CD86, CD40, y HLA-DR, luego de ser expuestos a partículas virales de GII.4 (53). Por otro lado, se observó producción *in vitro* de citoquinas por PBMCs recolectadas entre 0, 8 y 21 días de iniciado el desafío, destacando una elevación en los niveles de IFN- γ , IL-2 e IL-5, pero con pocos cambios en TNF- α , IL-4 e IL-10 (54). Todos estos resultados fueron consistentes con un estudio similar, en donde la infección con el virus SMV, miembro del genogrupo G.II, produjo también niveles elevados de IFN- γ e IL-2 luego de 2 días de inoculación (55). En un estudio más reciente, los niveles de citoquinas Th1 y Th2 e IL-8 fueron mayores en el grupo sintomático en comparación con el grupo asintomático ($p < 0.01$) (56). Adicionalmente, se observó que los síntomas no estaban asociados significativamente ni a la carga viral ni al tiempo de eliminación del virus en heces, sugiriendo un papel predominante de la respuesta inmune del hospedero en el curso de la infección por NV.

3.3 Epidemiología

La infección por NV en seres humanos puede ser sintomática o asintomática. De acuerdo a algunos experimentos realizados en humanos y la información proveniente de reportes de brotes infecciosos, 11.7 a 49.2 % de los individuos son asintomáticos (57). En una infección sintomática, el periodo de incubación suele ser de 24 – 48 horas luego de la exposición al virus (95% IC, 1.49 a 1.62 días) (58). La enfermedad se caracteriza por la aparición de diarrea no sanguinolenta, náuseas,

vómitos, dolores abdominales y mialgia. Estos síntomas, a pesar de su posible manifestación severa en algunos individuos, suelen desaparecer en un periodo de 2 a 3 días. La emesis es relativamente más prevalente en niños mayores a un año mientras que en aquellos menores de 1 año, la diarrea suele ser más frecuente. Se sabe que la infección por NV podría ocasionar la muerte en individuos infantes y seniles luego de un periodo prolongado de deshidratación.

El modo de transmisión de NV más común ocurre mediante la ruta fecal-oral, abarcando la ingestión de alimentos o agua contaminados, la exposición a fómites o vómito aerosolizado, y el contacto directo persona-persona, dando lugar a una rápida propagación en establecimientos como hospitales, colegios, hoteles, etc. Entre las características que facilitan esta transmisión se incluyen: una alta capacidad infectiva, siendo necesarias menos de 10 copias genómicas para ocasionar una infección en un individuo sano (59); el tiempo de excreción viral persistente, el cual se puede prolongar alrededor de 3 semanas posterior a la infección, tanto en individuos sintomáticos como asintomáticos; la estabilidad del virus bajo concentraciones altas de desinfectantes y en un rango amplio de temperaturas (desde la congelación hasta 60°C); y, finalmente, la posibilidad de ocurrir repetidas infecciones a lo largo de la vida por falta de una inmunidad a largo plazo.

En todo el mundo, aproximadamente uno de cada cinco casos de gastroenteritis aguda (diarrea y vómitos) es causado por norovirus. Globalmente, se estima que norovirus es la causa más común de gastroenteritis aguda. Es responsable de 685 millones de casos anualmente, entre los cuales, 200 millones ocurren en niños menores de 5 años de edad (4). Las muertes a nivel global por gastroenteritis se han

reducido dramáticamente, de 2.6 millones por año en 1990 a 1.3 millones en 2013, debido a que la sanitización de aguas, las medidas de higiene y el abastecimiento de soluciones rehidratantes han mejorado en diversos países. A pesar de todo ello, la gastroenteritis permanece como la segunda causa más común de mortalidad en niños menores de 5 años (5). Se estima que norovirus es el responsable de aproximadamente 200000 muertes anualmente en todo el mundo, con 70000 o más en la población infantil de países en desarrollo.

Una revisión sistemática de 175 estudios a nivel mundial, los cuales abarcaron 187336 casos de gastroenteritis aguda, mostró que la prevalencia de NV fue del 18% (95% de IC: 17% - 20%), siendo mayores los casos ocurridos en comunidades (24%) y países en vías de desarrollo (19%) (60). En Perú, un estudio de cohorte de nacimiento reveló que, en los dos primeros años de vida, 71% de los individuos experimentaron al menos 1 caso de enfermedad diarreica asociada a NV, siendo mayor la incidencia en el periodo de 6 a 23 meses de edad y detectándose el genotipo GII.4 en el 41% de los casos (61). En cuanto al estudio de la epidemiología molecular de NV, la mayor parte de la información ha provenido históricamente de muestras obtenidas en brotes; sin embargo, aquellas cepas con mayor relevancia actual a nivel mundial, como GII.4 Sidney 2012, han sido detectadas mayormente en muestras de casos esporádicos en comparación con casos de brotes, muchos años antes de tener predominancia mundial. Esto sugiere la importancia de hacer vigilancia a nivel mundial de la enfermedad esporádica, sobre todo en lugares de gran incidencia en países en desarrollo, ya que esto permitiría entender y anticipar cepas emergentes, así como predecir cepas pandémicas potenciales para futuras investigaciones (62). En la actualidad no existen estudios sobre la distribución de

esta cepa en nuestro país y, dado que el genotipo GII.4 tiene una prevalencia alta a nivel mundial, sería de vital importancia evaluar su distribución a nivel poblacional.

3.4 Diagnóstico

3.4.1 Microscopía Electrónica

La cepa original de norovirus, el virus Norwalk (genogrupo GI), fue identificado en 1972 mediante microscopía electrónica (ME). En un inicio, el diagnóstico de este tipo de infecciones se basó en el uso de esta técnica hasta principios de la década de los noventa. Sin embargo, la baja sensibilidad de la ME (requiriéndose $10^6 - 10^7$ partículas virales) hizo que, a partir de la última década del pasado siglo, y con el desarrollo de técnicas moleculares, la ME sea reemplazada por los métodos genómicos de amplificación actuales.

3.4.2 Real-time PCR

Una vez conocida la secuencia nucleotídica del genoma de NV, la metodología de transcripción reversa seguida de la reacción en cadena de la polimerasa (RT-PCR) se ha convertido en la técnica de referencia o *gold standard* para su diagnóstico, permitiendo la detección de NV en un rango de $10^2 - 10^4$ partículas virales por mililitro (63). Se han descrito múltiples primers para detectar distintas regiones del genoma de NV, especialmente en la región de la polimerasa (región A) y en la proteína VP1 de la cápside (Región D) (64). Además, la secuenciación de los amplicones permite identificar el genogrupo, genotipo e incluso, la variante a la cual pertenece la muestra al comparar la secuencia con otras de referencia. Posteriormente esta metodología se mejoró con la aplicación de la técnica de RT-PCR en tiempo real, siendo más sensible y permitiendo detectar la presencia de una carga viral de manera más rápida (65,66). Por ello, el empleo de la técnica de RT-

PCR a tiempo real se ha generalizado últimamente en laboratorios que procesan un elevado número de muestras y ha sido empleada más frecuentemente con fines diagnósticos por laboratorios adecuadamente equipados.

3.4.3 Inmunoensayo

Algunos métodos de inmunodiagnóstico como radioinmunoensayos (RIA) o inmunoensayos enzimáticos (ELISA) empleaban suero de personas en fase aguda y convaleciente de la infección por NV o de individuos voluntarios infectados experimentalmente (67,68), pero sus inconvenientes en cuanto a cantidad disponible de antisuero y su especificidad la hacían muy limitadas. Gracias a la posibilidad de generar partículas virales recombinantes mediante el uso de baculovirus a comienzos de la década de los 90, fue posible la producción de anticuerpos específicos que permitieron el desarrollo de inmunoensayos destinados a la detección de anticuerpos (69). Estas técnicas han mostrado una sensibilidad intermedia entre la ME y la RT-PCR, estimándose que pueden detectar del orden de 10^4 partículas virales por mililitro (63). Actualmente existen algunos kits de ELISA e inmunocromatografía comercialmente disponibles que podrían facilitar la identificación de este tipo de infección (70–73). Sin embargo, presentan el inconveniente de ser específicos para una determinada variante de NV GII.4, por lo que los resultados pueden variar según las cepas circulantes.

3.5 Inmunoinformática

3.5.1 Predicción de epítopes para células B

La inmunoinformática es la aplicación de herramientas bioinformáticas para el estudio de los distintos componentes del sistema inmunológico tales como las inmunoglobulinas, el complejo mayor de histocompatibilidad (MHC), los

receptores para células T y B, antígenos y otras moléculas involucradas en el desarrollo de la respuesta inmune. Una de las aplicaciones más importantes es la del estudio y desarrollo de algoritmos de predicción de epítopes, lo cual favorece la rapidez, pero, sobre todo, el bajo costo del desarrollo de métodos potenciales de diagnóstico y vacunas para enfermedades infecciosas.

Los epítopes para células B, los cuales corresponden a regiones reconocidas por sitios de unión o paratopos de inmunoglobulinas solubles o ancladas a la membrana celular, pueden ser clasificados como continuos o discontinuos. Se puede denominar un epítope continuo a aquella región o secuencia contigua de aminoácidos en una proteína mientras que un epítope discontinuo o conformacional consiste en un grupo de aminoácidos no adyacentes que se agrupan a través del plegamiento de las regiones peptídicas donde están ubicados.

La predicción de epítopes discontinuos en una proteína requiere conocer su estructura tridimensional. Hace más de 3 décadas, existían tan solo 57 estructuras de proteínas en el Protein Data Bank (PDB), razón por la cual los métodos de predicción basados en secuencias buscaban identificar epítopes continuos. Entre estos se pueden mencionar el servidor BCpred, el cual atribuye a cada aminoácido un puntaje de proclividad basado en las propiedades de hidrofiliidad, hidrofobicidad, formación de estructuras secundarias, etc., y el servidor ABCpred basado en una red neuronal artificial la cual ha sido entrenada con 700 epítopes lineales para células B del repositorio Bcipep y 700 péptidos aleatorios provenientes de la base de datos SWISSPROT; un ejemplo de la aplicación de ambos predictores permitió identificar epítopes putativos B de la proteína VP1 de sapovirus, siendo el primer trabajo reportado en este patógeno (74). Hoy en día, y

con el gran número de estructuras cristalinas almacenadas en diversos repositorios, es posible desarrollar algoritmos informáticos que puedan predecir si una proteína contiene epitopes discontinuos o también denominados conformacionales.

DiscoTope 2.0 es un predictor basado en estructura cuya metodología ha sido desarrollada a partir de la información experimental de 75 estructuras cristalinas de complejos anticuerpo-antígeno obtenidos por difracción de rayos X (75). Cada uno de los 20 aminoácidos conocidos recibe un puntaje de acuerdo con un parámetro denominado *log-odds ratio*, el cual representa el logaritmo del cociente entre la frecuencia aminoacídica observada en epitopes y la observada en regiones no epitópicas. En base a lo anterior mencionado, Discotope evalúa dos medidas principales: la primera, el cálculo integrado de los log-odds ratio de aminoácidos vecinos a un determinado residuo en un espacio tridimensional definido, lo cual da como resultado el log-odds ratio para cada residuo en una proteína de estudio, y la segunda, el grado de exposición dado por el parámetro conocido como exposición en media esfera (*half-sphere exposure*), el cual mide el número de átomos Ca ubicados en dos mitades de un espacio esférico alrededor del carbono α de un determinado residuo (76). En este último caso, un puntaje mayor se correlaciona con un número reducido de contactos con el carbono α del residuo evaluado, indicando una localización cercana a la superficie o, como en la mayoría de casos, en regiones protuberantes del antígeno (77). Finalmente, el predictor combina ambas medidas para asignar un puntaje global a cada residuo de la proteína estudiada.

3.5.2 Predicción de afinidad a moléculas MHC-II

A diferencia de las células B que reconocen epitopes en la superficie de un antígeno en estado nativo, las células T reconocen dicho antígeno una vez procesado y presentado como péptido en el contexto de una molécula del complejo mayor de histocompatibilidad (MHC). Por esta razón, el hecho de predecir si, y con qué afinidad, un determinado péptido se unirá a una molécula de MHC específica es un paso importante para identificar epitopes T potenciales. Existen dos clases de moléculas MHC: moléculas MHC clase I (MHC-I), las cuales se caracterizan por unirse a péptidos cortos, usualmente de 9 residuos de aminoácidos, y las moléculas MHC clase II (MHC-II), que se unen a péptidos de diversa longitud; típicamente puede variar de 11 a 30 residuos. Este último hecho se debe a que, a diferencia de las moléculas MHC-I cuya hendidura de unión al péptido es restringida, la hendidura de las moléculas MHC-II es abierta en ambos lados, permitiendo el alojamiento de péptidos mayores a 9 residuos. Sin embargo, se ha reportado que un núcleo o *core* de 9 residuos es esencial para la unión, siendo su ubicación precisa en el péptido de unión a MHC-II desconocida a priori, y su predicción, dificultosa en comparación con la predicción de unión a MHC-I (78).

En seres humanos, el locus de la región MHC-II alberga genes para las subunidades α y β de las proteínas HLA-DR, HLA-DQ y HLA-DP. Estos genes son altamente polimórficos (es decir, existen muchos alelos para cada locus) y, de acuerdo con la última versión de la base de datos IPD (*Immuno Polymorphism Database*), el número de proteínas DR α , DR β , DQ α , DQ β , DP α y DP β registradas es de 2, 1736, 35, 774, 24 y 641, respectivamente (79). Por lo tanto, existe la posibilidad de

generar miles de combinaciones de subunidades α y β de proteínas HLA clase II y, en consecuencia, se complica más el proceso de predicción de afinidad.

Hasta la fecha se ha desarrollado una variedad de métodos computacionales para predecir la unión de péptidos a MHC-II a partir de la secuencia primaria de una determinada proteína. Ejemplos de estos predictores incluyen métodos basados en matrices de puntaje, los cuales asignan un peso a cada residuo en base a un alineamiento de péptidos de unión a MHC-II; y métodos de aprendizaje de máquina (*machine learning*) basados en inteligencia artificial de redes neuronales (ANN). Sin embargo, muchos de estos predictores han sido entrenados empleando un número limitado tanto de datos experimentales de afinidad a MHC-II como de alelos evaluados, limitándose a las moléculas HLA-DR.

El predictor *NetMHC-IIpan 3.0* es el primer predictor común de afinidad de péptidos para las moléculas HLA-DR, HLA-DP y HLA-DQ (80). Su metodología está fundada en una AAN la cual ha sido entrenada con 52062 datos cuantitativos de afinidad peptídica a HLA almacenados en el repositorio IEDB (Immune Epitope Database), definidos según el valor de concentración inhibitoria máxima media o IC_{50} . Primero, se construye una pseudo secuencia representativa lineal de cada alelo a partir de los residuos de su respectiva molécula HLA con mayor potencial de contacto con un péptido anclado; por otro lado, el conjunto de péptidos que tengan afinidad demostrada por un determinado HLA es empleado para generar una matriz de puntaje que identifica el nonúmero con máxima probabilidad de explicar el IC_{50} observado para cada péptido. En base a lo anterior, *NetMHC-IIpan 3.0* se entrena continuamente integrando los datos de todos las pseudo secuencias alélicas con los

valores de afinidad o IC_{50} de los péptidos correspondientes a nanómeros con máximos puntajes.

3.5.3 Diseño de proteínas multiepitópicas

Diversos estudios han mostrado que el uso de proteínas multiepitópicas, cuya estructura contiene regiones antigénicas caracterizadas de manera experimental o mediante herramientas bioinformáticas, tiene un gran potencial como diagnóstico para diversas enfermedades. Una vez identificados o predichos a partir de un determinado antígeno, los epitopes son ensamblados a través de péptidos de unión o espaciadores que otorguen flexibilidad a lo largo de la estructura y, a su vez, eviten la formación accidental de nuevas secuencias antigénicas no correspondientes al patógeno en estudio. Algunos trabajos realizados en el diagnóstico de enfermedades virales incluyen el desarrollo de un método de inmunodiagnóstico para el virus de la hepatitis B, utilizando la proteína recombinante rMEHB (~21kDa) y obteniendo una alta especificidad para detectar anticuerpos IgG mediante ELISA (81); la detección de anticuerpos IgG contra el virus de la hepatitis E, el cual demostró poseer buena sensibilidad y especificidad en comparación con un kit comercial(82); y el diagnóstico del virus del dengue mediante el péptido recombinante r-DME-G (~25kDa) para la detección de anticuerpos IgG e IgM en pacientes humanos(83). Además, se ha utilizado esta aproximación para otro tipo de enfermedades causadas por parásitos tales como toxoplasma(84,85) y leishmania (86), obteniendo antígenos con gran potencial diagnóstico comparado con sus respectivos test comerciales.

IV. JUSTIFICACIÓN DEL ESTUDIO

Los test de diagnóstico basados en proteínas multiepitópicas presentan muchas ventajas sobre los inmunoensayos basados en el uso de una partícula viral completa debido a su seguridad, especificidad y reproducibilidad. Diversos antígenos sintéticos han sido utilizados en metodologías de diagnóstico dirigidas a detectar varias enfermedades, incluyendo infecciones bacterianas, virales, y parasitarias; en estos últimos casos, su aplicación en la detección de un agente infeccioso presenta ventajas en cuanto a la seguridad del personal, ya sea en la manipulación de los antígenos, así como en su producción en forma de proteína no nativa. También, la especificidad de una proteína multiepitópica se ve incrementada considerablemente debido a su composición exclusiva por determinantes antigénicos, elegidos de manera óptima de acuerdo a criterios como frecuencia de distribución, temporalidad, estructura secundaria, etc. El empleo de herramientas inmunoinformáticas para la identificación de dichas regiones cumple aquí un rol importante ya que, además de reducir considerablemente el costo y el tiempo de estudio, contribuye a enfrentar el obstáculo de la gran variabilidad genética que puede poseer un patógeno viral. Finalmente, por tratarse de una proteína recombinante de moderado peso molecular, su obtención representa una especial ventaja en situaciones donde los métodos convencionales para obtener grandes cantidades de antígeno, a través del cultivo de un patógeno o mediante la extracción de moléculas antigénicas, son inaccesibles o no han sido satisfactoriamente desarrollados, generando así un abaratamiento de costos tanto en equipamiento como en el entrenamiento del personal.

V. HIPÓTESIS

Una proteína multiepitópica recombinante, diseñada en base a diversos epitopes inmunodominantes de la proteína de la cápside viral VP1, posee la capacidad de detección de anticuerpos IgG e IgM frente a norovirus GII.4 en suero humano

VI. OBJETIVOS

6.1 Objetivo general

Desarrollar una proteína multiepitópica basada en epitopes inmunodominantes de la proteína VP1 de la cápside de norovirus GII.4 para la detección de anticuerpos IgG e IgM en suero humano.

6.2 Objetivos específicos

- Evaluar bioinformáticamente secuencias completas de aminoácidos de la proteína VP1 de norovirus GII.4 y generar modelos estructurales por homología.
- Predecir epitopes de células B con afinidad a moléculas MHC-II mediante análisis inmunoinformático y ensamblarlos en una secuencia multiepitópica.
- Expresar y purificar la secuencia multiepitópica diseñada en células de *E. coli* BL21.
- Estandarizar una técnica de ELISA para la detección de anticuerpos séricos IgG e IgM en individuos infectados con norovirus GII.4.

VII. METODOLOGÍA

7.1 Análisis Inmunoinformático

7.1.1 Alineamiento de secuencias

Se realizó un alineamiento múltiple de todas las secuencias completas de la proteína VP1 de la cápside de NV GII.4 de la base de datos GenBank, almacenadas hasta el año 2015, mediante el programa CLC MAIN WORKBENCH 7.9. La secuencia de la cápside fue elegida por dos razones: en primer lugar, una gran parte de las secuencias disponibles en la base de datos elegida corresponden a esta región, y, en segundo lugar, el ORF2 codifica a la proteína VP1, la cual contiene la mayoría de epitopes inmunodominantes observados en NV. Además de esto, se consideraron las secuencias más antiguas debido a que no se dispone aún de estudios epidemiológicos en algunos países que confirmen la ausencia de las mismas, y, en cuyo caso, se aumentaría más las probabilidades de poder diagnosticar a NV GII.4 a nivel de dichas regiones. Debido al número elevado de secuencias se empleó el algoritmo de alineamiento progresivo, utilizando los valores de penalidad de 10.0 para la abertura de gaps y de 0.1 para la extensión de los mismos. Con el fin de evitar la sobrerrepresentación de secuencias idénticas contenidas en la data cruda, éstas fueron divididas en grupos que contengan secuencias que compartan un 100% de identidad unas con otras mediante el programa BLASTclust proveído gratuitamente por el Instituto Max Planck (<http://toolkit.tuebingen.mpg.de>), pudiendo de esta manera descartar aquellas que se repiten.

7.1.2 Análisis filogenético

Mediante el programa MEGA 6.0 se infirió un árbol filogenético utilizando el algoritmo de máxima verosimilitud a partir de la información obtenida del alineamiento múltiple de secuencias, con la finalidad de identificar clados correspondientes las diferentes variantes de NV GII.4. El algoritmo inició con la construcción de un árbol basado en distancias genéticas o de *neighbour-joining*, el cual se sometió a una serie de modificaciones tanto en la topología como en la distancia de las ramas, obteniendo así un conjunto de árboles filogenéticos; cada árbol representó una hipótesis evolutiva de la secuencia de la proteína VP1, de tal manera que se eligió aquel con la máxima probabilidad de originar las secuencias observadas.

El modelo de substitución de aminoácidos aplicado fue la matriz de substitución JTT (Jones, et al 1992), la cual estima la tasa de reemplazo de cada aminoácido a partir del alineamiento de diversas proteínas provenientes de bases de datos actualizadas. La significancia estadística de la filogenia analizada fue estimada mediante el método de remuestreo o bootstrap. Esta metodología se basa en la comparación de la topología obtenida por el árbol original y las obtenidas por pseudo réplicas generadas a partir de dicho árbol, obteniendo como resultado un valor bootstrap de 100 si las ramas del árbol original coinciden con el 100% de las réplicas; para el presente estudio se empleó un número de 1000 réplicas y se consideró un valor bootstrap de 70 como adecuado. El árbol obtenido se visualizó a través del programa FIGTREE 1.4. Para facilitar el reconocimiento de clados que correspondan a una epidemia causada por una variante de NV en particular, se ubicó a la secuencia de referencia de cada una de las variantes reportadas hasta la fecha,

de tal manera que las secuencias agrupadas en conjunto con las representativas se consideraron pertenecientes a dicha variante.

Mediante esta aproximación la representatividad de las secuencias en cada clado se debió a la información biológica contenida en la secuencia aminoacídica, descartando la posibilidad de sesgo en cuanto a la diferencia geográfica de los centros donde se obtuvieron dichas secuencias, la cantidad de muestras tomadas, y tiempo en el que fueron recolectadas y/o almacenadas en la base de datos. A partir de cada clado se realizó nuevamente un alineamiento múltiple y se obtuvo una secuencia consenso la cual fue sometida a las evaluaciones posteriores.

7.1.3 Modelamiento por homología

Se generó un modelo estructural 3D de la secuencia consenso de cada clado observado mediante modelamiento por homología empleando el programa MODELLER 9.14. A través de la búsqueda de secuencias homólogas con el servidor SWISS-MODEL, se identificó al cristal de la proteína VP1 de la cepa NV/GII.4/VA387/1998/US (PDB 2OBT) (87) como mejor molde para realizar el modelamiento, sin embargo, este modelo solo contaba con el dominio P razón por la cual se optó por la opción de modelamiento múltiple, añadiendo al cristal de la proteína VP1 del prototipo original NV/GI (PDB 1IHM), cuyo dominio S estaba disponible, con el fin de construir un molde híbrido.

Una vez obtenidos los modelos 3D de cada consenso, éstos fueron sometidos a un proceso de refinamiento estructural mediante minimización de energía y a una evaluación de calidad a través de los siguientes parámetros: la raíz de la desviación cuadrática media (RMSD), la cual mide la semejanza entre las coordenadas del

modelo y la estructura obtenida experimentalmente; el z-score, obtenido mediante el servidor PROSAT-WEB, que mide la desviación de la energía total del modelo propuesto con respecto a una distribución de energía proveniente de conformaciones aleatorias; diagrama de Ramachandran, que permite observar el porcentaje de la distribución de ángulos ϕ/ψ de aminoácidos ubicados en zonas permitidas; y, el análisis de ERRAT que permite identificar que regiones del modelo poseen errores estructurales, a través de la comparación estadística entre las interacciones no covalentes entre pares de átomos en el modelo propuesto y una base de datos de estructuras con alta resolución.

Posteriormente, los aminoácidos que se encontraban bajo selección positiva de acuerdo al análisis de secuencias fueron mapeados en los modelos estructurales obtenidos. Adicionalmente, los modelos correspondientes a las variantes más antigua y reciente fueron superpuestas para comparación. Las estructuras de las proteínas se visualizaron utilizando el programa CHIMERA 1.12.

7.1.4 Predicción de epitopes para células B

Los modelos estructurales de cada secuencia consenso fueron analizados mediante el predictor de epítopes para células B DISCOTOPE 2.0, el cual predice epítopes discontinuos bajo los parámetros de puntaje de proclividad de los residuos y el número de contactos entre ellos; la integración de ambos permite obtener un puntaje final para cada residuo. El valor umbral para dicho puntaje se correlaciona con una combinación determinada de sensibilidad y especificidad en la predicción; a mayores umbrales de puntaje (>1.9) se obtiene una mayor especificidad (0.95), pero se resta sensibilidad (0.17). Por tal motivo, se seleccionó un valor umbral de puntaje final de -3.7, lo cual indica que residuos con valores mayores a esta cantidad fueron

considerados como parte de epitopes B de NV GII.4, con una especificidad de 0.75 y una sensibilidad de 0.47. Debido a que el mayor número de epitopes experimentales para anticuerpos específicos de NV han sido identificados en el subdominio P2 de la proteína VP1, se aplicó un criterio de exclusión para aquellos residuos ubicados fuera de dicho subdominio. Por otro lado, un criterio de inclusión importante fue la localización de los residuos en zonas de mayor exposición, mediante su mapeo en los modelos estructurales predichos. Se contrastaron los resultados obtenidos comparándolos con estudios publicados hasta la fecha.

7.1.5 Predicción de afinidad a MHC clase II

El predictor NetMHCIIpan 3.0 fue escogido para predecir la afinidad de unión de péptidos a moléculas de MHC-II, abarcando los tres isotipos HLA-DR, HLA-DP y HLA-DQ. Para ello, las secuencias consenso son utilizadas como datos de entrada, las cuales fueron virtualmente cortadas con el fin de generar péptidos lineales superpuestos de una longitud de 15 residuos. Una vez evaluados, cada péptido es asignado con un valor de predicción equivalente a la afinidad de unión o valor de IC₅₀, siendo un valor menor a 500 nM correspondiente a un péptido con afinidad intermedia y un valor menor a 50 nM, a un péptido con afinidad fuerte.

Con el fin de evaluar la promiscuidad de los péptidos para los diversos alelos que cubran el mayor número posible de moléculas MHC-II, se seleccionaron los alelos para las subunidades α y β de HLA de mayor frecuencia distribuidos a nivel global reportados en la base de datos *The Allele Frequency* (www.allelefrequencys.net). Los haplotipos más comunes para HLA-DP fueron DPA1*0201/DPB1*0101, DPA1*0103/DPB1*0201, DPA1*01/DPB1*0401, DPA1*0301/DPB1*0402, DPA1*0201/DPB1*0501 y DPA1*0201/DPB1*1401; y para HLA-DQ,

DQA1*0501/DQB1*0201, DQA1*0501/DQB1*0301, DQA1*0301/DQB1*0302, DQA1*0401/DQB1*0402, DQA1*0101/DQB1*0501, DQA1*0102/DQB1*0602. En el caso de HLA-DR, los alelos más comunes fueron DRB1*01:01, DRB1*03:01, DRB1*04:01, DRB1*04:05, DRB1*07:01, DRB1*08:02, DRB1*09:01, DRB1*11:01, DRB1*12:01, DRB1*13:02, DRB1*15:01, DRB3*01:01, DRB3*02:02, DRB4*01:01 y DRB5*01:01

7.1.6 Conservación y especificidad

El criterio para definir que epitopes serían considerados como candidatos fue el grado de conservación evaluado por el servidor BLOCKLOGO, y se definió como la frecuencia con la que un rango específico de residuos contiguos, correspondiente a una región lineal predicha como epítipo para célula B, estaba distribuido en cada clado. La especificidad de los epitopes fue evaluada mediante la herramienta de alineamiento local BLASTp, comparándolos con las secuencias de todas las proteínas almacenadas en el repositorio del NCBI. El criterio de aceptación de similitud ante una determinada proteína y/o especie fue el valor del estadístico $E < 10^{-10}$.

7.1.7 Diseño del gen para la proteína multiepitópica MEP-GII.4

En base a los epitopes candidatos obtenidos del estudio inmunoinformático, se diseñó una proteína multiepitópica denominada MEP-NVGII.4 que consistió en una única cadena polipeptídica construida a partir de 10 epitopes unidos mediante espaciadores flexibles de tetraglicina (Gly-Gly-Gly-Gly), con lo cual se estimó un peso aproximado de 17.7 kDa. Los criterios para seleccionar dichos epitopes fueron un puntaje de predicción alto, afinidad de unión a MHC-II de intermedia a alta, promiscuidad a alelos MHC-II y, finalmente, la temporalidad de las cepas

circulantes correspondientes a la fecha de estudio. Seguidamente, la secuencia aminoacídica MEP-NVGII.4 fue sometida a un proceso de traducción reversa mediante el programa CLC MAIN WORKBENCH 7.9 para diseñar el gen sintético *MEP-NVGII.4*, considerando un uso de codones óptimo para su posterior sobreexpresión en células de *E. coli* BL21(DE3).

7.2 Producción de la proteína multiepitópica MEP-NVGII.4

7.2.1 Sistema de expresión pET28a-MEP-NVGII.4

El gen recombinante *MEP-NVGII.4* fue sintetizado químicamente y clonado en el vector de expresión pET28a (+) por la empresa Genemed Synthesis Inc. (EE.UU.) entre los sitios de corte para las endonucleasas *NcoI* y *XhoI* (Figura 3). Este constructo, denominado pET28a-MEP-NVGII.4, permitió añadir una secuencia codificante 6xHis (cola de histidina) en el extremo carboxilo terminal de la proteína de interés para fines de purificación, precedido por una secuencia de fusión que codifique un sitio de corte para la proteasa TEV (Glu-Asn-Leu-Tyr-Phe-Gln-Gly) proveniente del virus de tabaco, con el fin de realizar ensayos inmunológicos en presencia o ausencia de la cola de histidina. La secuencia de interés se expresó a través del promotor T7, el cual a su vez está fuertemente regulado por el represor *lac*, de tal manera que la expresión fue inducida con isopropil- β -D-1-tiogalactopiranosido (IPTG). Este vector incorporó además un gen codificante para la enzima aminoglucósido-3'-fosfotranserasa (*kan*) la cual confiere resistencia al antibiótico kanamicina para fines de selección de colonias.

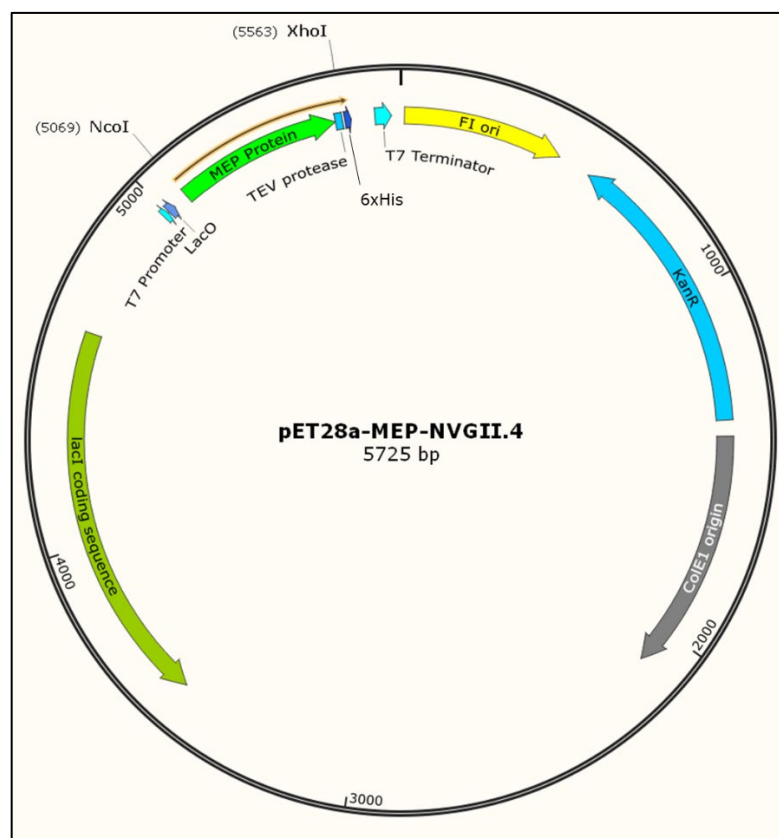


Figura 3. Mapa del vector pET28a-MEP-NVGII.4. MEP Protein: gen que codifica a la proteína multiepitópica recombinante MEP-NVGII.4, flanqueada por los sitios de restricción para las enzimas *NcoI* y *XhoI*.

7.2.2 Transformación de células de *E. coli* BL21(DE3)

Se prepararon células químicamente competentes de *E. coli* BL21(DE3) (ver Anexo 1) para ser transformadas con el vector pET28a-MEP-NVGII.4 mediante el método de shock térmico. Primero, se descongelaron en hielo las células competentes durante 10 minutos y, en seguida, se les añadieron aproximadamente 100 ng del plásmido recombinante, incubándose en hielo durante 30 minutos. Para el control negativo de transformación se añadió un volumen similar de agua Mili Q estéril.

Luego, las células fueron introducidas en baño húmedo a 42°C por 50 segundos y seguidamente incubadas en hielo durante 2 minutos. Se añadió 1 mL de medio LB y se incubó a 37°C con agitación a 250 rpm. Finalmente, se sembraron alícuotas de 100 µL en placas con agar LB con kanamicina (40 µg/mL) y se incubaron a 37°C durante toda la noche.

7.2.3 Caracterización del vector pET28a-MEP-NVGII.4

Se procedió a extraer y purificar el plásmido pET28a-MEP-NVGII.4 a partir de una colonia transformada mediante un protocolo interno de miniprep por lisis alcalina (ver Anexo 2). Así mismo, se llevó a cabo la digestión del plásmido purificado utilizando las enzimas de restricción *XhoI* y *NcoI* (New England Biolabs, EE.UU.) y los reactivos recomendados por el fabricante (ver Anexo 3). Las mezclas de reacción fueron incubadas a 37°C durante 3 horas en baño maría y los productos digeridos fueron posteriormente analizados mediante electroforesis en gel de agarosa al 1,5%. El tamaño del inserto se corroboró por comparación con el marcador de peso molecular 1 Kb Plus (Invitrogen, EE.UU.).

7.2.4 Expresión de la proteína MEP-NVGII.4

Se realizó un estudio de inducción sistematizado para definir las condiciones óptimas de expresión de la proteína multiepitópica. Brevemente, se inocularon 5 mL de caldo LB con kanamicina con una colonia positiva transformada con el plásmido pET28a-MEP-GII4 y se dejó incubar overnight a 37°C, produciendo un cultivo iniciador. El estudio se llevó a cabo inoculando 2mL del cultivo iniciador en un matraz con 50 mL de caldo LB con kanamicina (dilución 1:25) e incubando a 37°C con agitación constante, hasta que el cultivo alcanzara un OD_{600nm} de 0.6 a 0.8.

La expresión de la proteína se indujo mediante la adición de tres volúmenes distintos de IPTG, obteniendo como concentraciones finales 0.125 mM, 0.5 mM y 1mM. Los cultivos fueron incubados a 25°C o 37°C y se recolectaron alícuotas de 1 mL en los intervalos de tiempo T = 0, 1, 2 ,3 ,4 ,5 y 10 horas. Se determinó el OD_{600nm} en cada intervalo y posteriormente las muestras fueron sometidas a centrifugación a 13000 rpm y resuspendidas, en un volumen calculado de buffer fosfato salino (PBS), pH 7.4, según la fórmula:

$$Volumen\ PBS = \frac{OD_{600nm}}{0.5} \times 150$$

Con el fin de determinar en qué fase se encontraba la proteína de interés, se realizó un estudio de solubilidad. Brevemente, 50 mL de cultivo fueron recolectados luego de inducir la expresión con IPTG por T= 5, 7 y 10 horas. Las muestras fueron sometidas a centrifugación refrigerada (4°C) por 20 minutos a 13000 rpm y los pellets fueron resuspendidos con 3 mL de PBS. Las células se colocaron en hielo y seguidamente fueron sonicadas utilizando un equipo de ultrasonido S3000 (Misonix Inc, EE.UU.), empleando 4 pulsos de 60 segundos y un período de enfriamiento de 15 segundos entre cada pulso. Una alícuota de 100 µL de la suspensión celular fueron recolectados para representar a la proteína total (fracción soluble e insoluble). Se centrifugó la suspensión a 13000 rpm por 20 minutos y a 4°C, se separó el sobrenadante (fracción soluble), y el pellet obtenido (fracción insoluble) fue resuspendido en 3mL de buffer de extracción (20mM Tris-HCl pH8, 0,5 M NaCl, 8M urea). Todas las muestras obtenidas se analizaron mediante electroforesis SDS-PAGE discontinua (ver Anexo 4).

7.2.5 Cromatografía de afinidad

La proteína multiepitópica MEP-NVGII.4 fue purificada en un sistema para cromatografía de baja presión BioLogic LP (Bio-Rad, EE.UU.) empleando una columna HisTrap TM HP 1mL (GE Healthcare, EE.UU.) la cual permite la separación de proteínas fusionadas a una cola de histidina mediante cromatografía de afinidad a metales inmovilizados o IMAC (88). Previamente, se realizó la inducción de 1 L de cultivo de *E. coli* BL21(DE3) bajo las condiciones óptimas obtenidas según la metodología 7.2.4. Luego de la liberación de la proteína mediante lisis celular por ultrasonido, se centrifugó a 13000 rpm por 20 minutos y a 4°C, se resuspendió el pellet en 15 mL de *binding buffer* (buffer fosfato 20 mM, NaCl 500 mM, urea 8M, imidazol 20 mM, pH 8) y se filtró a través de una membrana Millipore de 0,45µm. En seguida, la muestra fue inyectada en la columna HisTrap, previamente equilibrada, usando *binding buffer* como fase móvil a un flujo de 1mL/min. La elución de la proteína se llevó a cabo empleando una gradiente de concentraciones de imidazol en la fase móvil de buffer fosfato/NaCl/urea (20, 40, 60, 100, 300 y finalmente 500 mM de imidazol), observando que la señal de absorbancia en el cromatograma sea estable y recolectando alícuotas de 1 mL antes de proseguir con una nueva concentración de imidazol. La pureza de la proteína MEP-NVGII.4 en cada fracción recolectada fue analizada mediante SDS-PAGE discontinua (ver Anexo 4).

7.2.6 Método de Bradford para cuantificación de proteína

Este método colorimétrico se basa en la unión proporcional del colorante azul de Coomassie (Sigma Aldrich, EE.UU.) a las proteínas. Una solución de albúmina de suero bovino fue preparada para un stock de 10 mg/mL y diluida 1/10 para la

preparación de los estándares. 20 µL de estándar (0.25, 0.5, 0.75, 1 y 1.4 mg/mL) y de las fracciones diluidas fueron colocados en una placa de poliestireno de 96 pocillos con fondo plano, junto con 200 µL del colorante diluido 1/5. Luego de un reposo a temperatura ambiente por 5 minutos, la placa fue leída a 595 nm en el lector de microplacas SpectraMax® 190 (Molecular Devices, EE.UU.).

7.2.7 Concentración de la proteína MEP-NVGII.4

Aquellas fracciones con mayor cantidad de proteína fueron sometidas a un proceso de concentración al vacío, utilizando el sistema AMICON® Stirred Cell (Merck, EE.UU.) (ver Anexo 5). La solución concentrada de proteína recombinante fue nuevamente cuantificada según la metodología 7.2.6.

7.3 Desarrollo de un método de ELISA indirecto

7.3.1 Muestras de suero humano

El material clínico de estudio estuvo constituido por sueros de bebés con edades comprendidas entre cero y 2 años, provenientes de un estudio de cohorte realizado entre los años 2007 y 2011 (61) el cual fue aprobado por el Comité Institucional de Ética de la Universidad Peruana Cayetano Heredia. Los individuos fueron divididos en dos paneles: el panel A, conformado por 20 sueros de bebés (B1-B20) con confirmación de infección por NV GII.4 mediante qPCR en el Laboratorio de Investigación y Desarrollo (LID) de la Universidad Peruana Cayetano Heredia, los cuales fueron considerados como controles positivos, y el panel B, conformado por 6 sueros de bebés recién nacidos (R1-R6), los cuales fueron considerados como controles negativos. Todos los procedimientos fueron realizados siguiendo los lineamientos establecidos en la Declaración de Helsinki para investigación en humanos.

7.3.2 Desarrollo de un método de ELISA indirecto

La prueba inmunológica que se desarrolló fue un ELISA indirecto para la cual, la concentración óptima de la proteína MEP-NVGII.4 como antígeno, las diluciones de trabajo del conjugado y la dilución de los sueros fueron determinadas previamente mediante titulaciones del tipo *checkerboard* o tablero de ajedrez modificado.

En una primera etapa, se tomaron dos placas NUNC Maxisorp de 96 pozos; en los pares de filas A y E, B y D, y C y F se colocaron respectivamente 100 µL del antígeno a 0.5, 1 y 2 µg/mL, y se incubó toda la noche a temperatura ambiente con rotación constante. En seguida, la placa se lavó 5 veces con tampón fosfato salino 0.075 M Na₂HPO₄, 0.025 M NaH₂PO₄, 0.014 M NaCl, pH 7.2, que contenía 0.5% Tween-20 (PBS/Tween) y se añadió 200 µL de una solución de bloqueo distinto en cada placa, una para evaluar el bloqueo con leche al 5% en PBS/Tween y otra, con leche al 5% y seroalbúmina bovina (BSA) al 2.5% en PBS/Tween. El bloqueo se realizó durante una hora a 37°C y se lavó 5 veces. Posteriormente en cada placa, se adicionaron 100 µL de suero (3 controles negativos y 3 controles positivos) diluidos 1:100 en las filas comprendidas entre la A y la C, y 100 µL de suero diluidos 1:200 en las filas comprendidas entre la D y la F. Los sueros fueron diluidos en solución de leche al 1% en PBS/Tween. Se incubó por hora y media a 37°C y se lavó 5 veces. Luego, se agregaron en ambas placas 100 µL del conjugado anti IgG-peroxidasa HRP (Kirkegaard & Perry Laboratories, EE.UU.) en la dilución de 1:5000 para las columnas 1 a la 6, y en la dilución 1:10000 para las columnas 7 a la 12, y se incubó por una hora a 37°C. Se lavó 5 veces y se adicionaron 50 µL del sustrato 3,3',5,5'-tetrametilbencidina (TMB), midiéndose la absorbancia a 450 nm luego de 5

minutos en el lector de microplacas. Las filas G y H en cada placa se emplearon para evaluar los controles experimentales de antígeno (suero + conjugado), de conjugado (antígeno + suero) y de suero (antígeno + conjugado).

En una segunda etapa, se tomó una placa NUNC Maxisorp de 96 pozos; en los grupos de filas A, B, E, F y C, D, G, H se colocaron respectivamente 100 μ L del antígeno a 0.5 μ g/mL y 1 μ g/mL y se incubó toda la noche a temperatura ambiente con rotación constante. Seguidamente, la placa se lavó 5 veces y fue bloqueada con 200 μ L de leche descremada al 5% en PBS/Tween, incubándose durante una hora a 37 °C. Se lavó 5 veces con PBS/Tween e inmediatamente, en los pares de columnas 1:7, 2:8, 3:9, 4:10, 5:11, y 6:12 se añadieron respectivamente 100 μ L de dilución 1:100 de los sueros B1, B2, B3, B4, B5 y B6. Los sueros fueron diluidos en solución de leche al 1% en PBS/Tween. Se incubó por hora y media a 37°C y se lavó 5 veces. Luego, se agregaron 100 μ L del conjugado anti IgG-peroxidasa HRP en las columnas 1-6, en la dilución de 1:5000 para las filas A-D y de 1:10000 para las filas E-H; de modo similar, se agregaron 100 μ L del conjugado anti IgM-peroxidasa HRP en las columnas 7-12, en la dilución de 1:4000 para las filas A-D y de 1:8000 para las filas E-H. La placa fue incubada por una hora a 37°C, se lavó 5 veces y se añadieron 50 μ L de TMB, midiéndose la absorbancia a 450 nm luego de 5 minutos.

Finalmente, se seleccionaron las condiciones que permitieron discriminar de manera óptima la absorbancia generada entre controles positivos y negativos, a la vez de obtener una señal de fondo por debajo del valor de densidad óptica (OD) de 0.05.

7.3.3 Detección de anticuerpos específicos IgG e IgM

Brevemente, se tomaron dos placas NUNC Maxisorp de 96 pozos y se inmovilizaron 100 μ L de antígeno a 1 μ g/mL durante toda la noche a temperatura ambiente y con rotación constante. En seguida, las placas se lavaron 5 veces con PBS/Tween y fueron bloqueadas con 200 μ L de leche al 5% en tampón PBS/Tween, incubándose durante una hora a 37°C. Se lavó 5 veces, se agregaron 100 μ L tanto de controles positivos como negativos por duplicado, diluidos 1:100 con leche descremada al 1% en PBS/Tween, y se incubaron las placas por hora y media a 37°C. Luego se lavó 5 veces y se agregaron 100 μ L del conjugado anti IgG-peroxidasa HRP diluido 1:10000 en una de las placas, mientras que en la otra se agregaron 100 μ L del conjugado anti IgM-peroxidasa HRP diluido 1:5000; ambos conjugados se diluyeron en leche descremada al 1% en PBS/Tween. Se incubaron las placas por una hora a 37°C y se lavaron 5 veces. Después, se añadieron 50 μ L de TMB, leyéndose la absorbancia generada a 450 nm luego de 5 minutos en el lector de microplacas.

7.3.4 Análisis de datos

La presencia o ausencia de anticuerpos específicos de NV GII.4 se determinó comparando la absorbancia de las muestras respecto al valor de corte o *cut-off*, el cual fue calculado a través de la media aritmética de las lecturas de OD de los 6 controles negativos más 3 veces la desviación estándar. La reactividad de los sueros individuales frente a la proteína MEP-NVGII.4 fue graficada empleando el programa GRAPHPAD PRISM 6 (GraphPad Software, EE.UU.). La diferencia estadística entre los controles positivos y los controles negativos fue calculada

mediante el test no paramétrico U de Mann-Whitney, considerándose un valor $p < 0,001$ como estadísticamente significativo.

VIII. RESULTADOS

8.1 Análisis de secuencias

Se descargaron todas las secuencias completas de la proteína VP1 de la cápside de NV GII.4 entre los años 1975 y 2015 ($n = 925$), de las cuales sólo 716 fueron seleccionadas luego del proceso de curado (ver Anexo 6), descartándose aquellas que contenían residuos no definidos o algún error de secuenciamiento, y que además fueran redundantes. A partir de dichas secuencias se generó un alineamiento múltiple, el cual fue exportado en formato FASTA para luego construir un árbol filogenético utilizando el algoritmo de máxima verosimilitud.

En la figura 4 se observa que las secuencias se segregaron en seis ramas principales, con múltiples cepas conformando cada rama, seguido de ramas más pequeñas o clados, los cuales representaron brotes epidémicos específicos alrededor del mundo. Según orden cronológico, los clados estuvieron asociados a las variantes CHDC 1970s (1974-1977), 95/96-US (1995-2002), Henry (2000-2004), Farmington Hills (2002-2004), Hunter (2004-2006), Asia (2003-2006), 2006a (2006-2008), 2006b (2006-2012), Osaka (2007-2008), Apeldoorn 2008 (2007-2011), New Orleans (2009 al presente) y la más reciente, Sydney (2012 al presente). Algunas de estas variantes estuvieron asociadas a pandemias de gastroenteritis aguda y han mostrado distribución global, tales como las variantes 95/96-US, Farmington Hills 2002, Hunter 2004, 2006b, New Orleans 2009 y Sydney 2012.

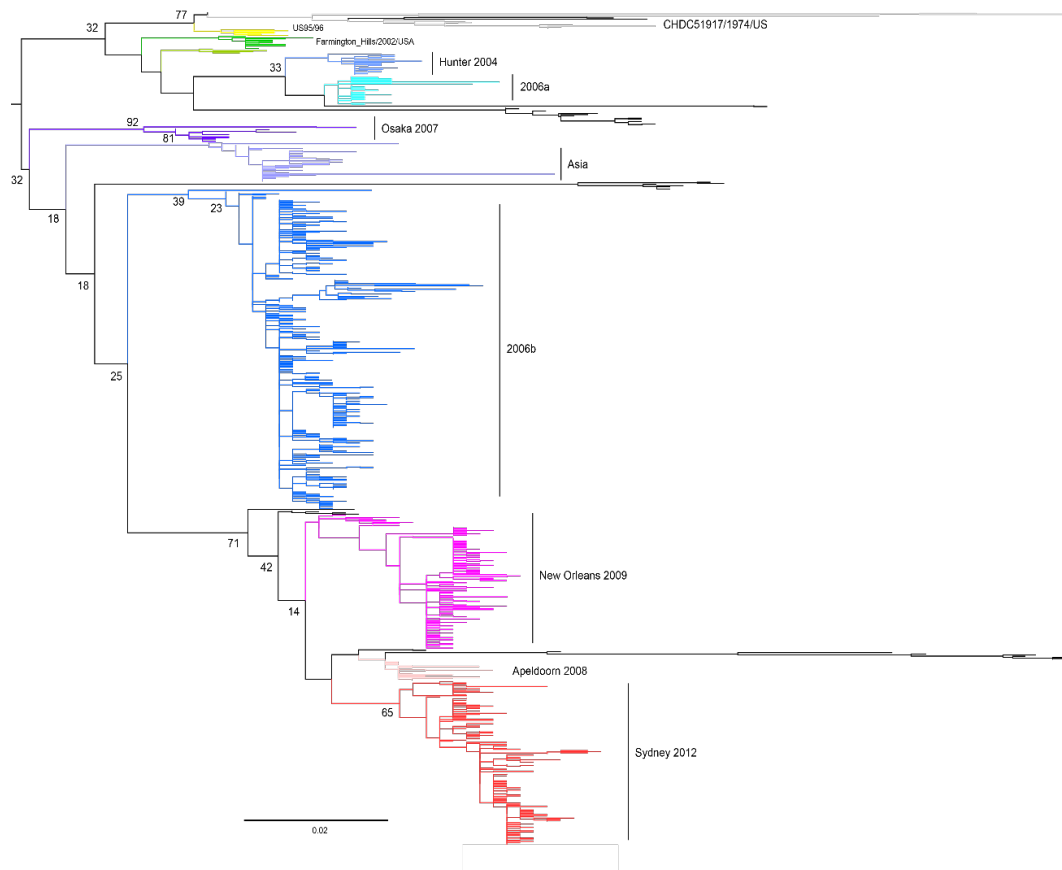


Figura 4. Análisis filogenético de máxima verosimilitud de 716 secuencias completas de la proteína VP1 de NV GII.4. El árbol muestra que distintos clados emergieron en el tiempo, los cuales están representados por un color diferente. La longitud de las ramas es proporcional al número de sustituciones por sitio. Los valores de bootstrap, expresados como el porcentaje de veces que se observa un clado en un total de 1000 réplicas, están indicados en los principales nodos del árbol filogenético.

Por otro lado, se observaron 4 ramas que no estuvieron asociadas a ningún grupo anteriormente mencionado (denominadas *outlayers*) y que, según la fecha en la que dichas variantes fueron secuenciadas, debieron pertenecer a alguno de los clados más recientes desde el 2012.

Los valores de bootstrap obtenidos fueron menores de los esperados tras el análisis filogenético, siendo el valor de 92 en el nodo que define a la variante Osaka 2007 el mayor de los observados. A pesar de este hecho, la topología obtenida que nos aproxima a la historia evolutiva de NV GII.4 resultó ser consistente con estudios previamente reportados.

De acuerdo a algunos estudios, se ha sugerido que los virus pertenecientes al genotipo GII.4 evolucionaron mediante un modelo de evolución por época, similar al virus de la influenza, lo cual significa un proceso en donde existen períodos estacionarios, sin diferenciación considerable en el fenotipo (épocas), seguidos de rápidas tasas de evolución, en los cuales nuevos fenotipos emergen. En el caso de NV GII.4, algunos clados han persistido por varios años, siendo algunos ejemplares identificados incluso en estudios muy recientes. Cabe señalar que, de manera interesante, se puede observar que los clados más recientes han emergido de manera más rápida, dando lugar a la aparición de nuevas variantes aproximadamente cada 2 a 3 años.

Más adelante, una secuencia representativa de cada clado fue seleccionada y nombrada según la fecha en la que fue aislada. El criterio de representatividad de cada secuencia se basó en la condición de haber sido la primera en ser reportada desde su correspondiente brote epidémico. Para el clado 1970, se seleccionó la

secuencia con código de acceso al GenBank ACT76139.1; para US95/96, ABY27560.1; para Henry 2001, ACL27298.1; para Farmington Hills 2002, AFJ04708.1; para 2006a, ABQ63283.2; para Hunter 2004, BAF95509.1; para 2006b, ABV55634.1, para Asia, ABP88833.1; para Osaka 2007, BAJ13894.1, para Apeldoorn 2008, ADM52743.1; para New Orleans 2010, ACX31889.1; y para Sydney 2012, AFV08795.1. Estas secuencias fueron exportadas, alineadas y ordenadas por fecha en una tabla de Excel para la evaluación de sitios informativos (Ver anexo 7). Se consideraron sitios informativos a aquellas posiciones en donde ocurrió una o más sustituciones aminoacídicas entre las variantes. En base a ello, se identificaron 87 sitios (16.1 % de 540 aminoácidos) como informativos (Tabla 1); 41 de estos sitios estuvieron localizados en el dominio P2 (32.3 % de aminoácidos en este dominio), siendo su número significativamente mayor (test de chi cuadrado, $p < 0.001$) en comparación con el dominio N-terminal (10%), el dominio S (7.4%), y el dominio P1 (14.9%).

Tabla 1. Sitios informativos en la secuencia de la proteína VP1

Dominio	Posición	Longitud total (aa)*	Nº de sitios informativos (%)
N-terminal	1 a 50	50	5 (10)
S	51-225	175	13 (7.4)
P1	226-278, 406-540	188	28 (14.9)
P2	279-405	127	41 (32.3)
Total		540	87 (16.1)

*aa = aminoácidos

Además, se pudieron identificar sitios hipervariables, es decir, posiciones en las cuales se observaron 3 o más residuos diferentes a lo largo del periodo de estudio,

ubicados predominantemente en el dominio protuberante P2 (Figura 5), en los residuos 294, 297, 298, 310, 333, 340, 352, 357, 359, 368, 372, 376, 377, 378, 389, 393, 394 y 395.

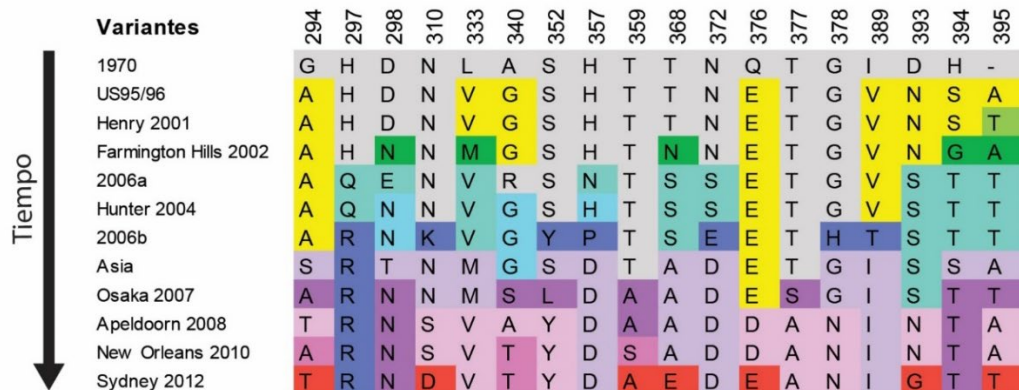


Figura 5. Sitios hipervariables del subdominio P de la proteína VP1, identificados en secuencias representativas de NV GII.4 desde el año 1975 hasta el 2015. La posición de los aminoácidos está indicada en la parte superior mientras que el nombre y año de origen de las variantes se muestran en el lado izquierdo. De arriba hacia abajo, cada color está asociado a la variante con la cual aparecieron los cambios de residuos observados. Variante ancestral 1970, gris; US95/96, amarillo; Henry 2001, verde claro; Farmington Hills 2002, verde oscuro; 2006a, turquesa; Hunter 2004, celeste; 2006b, azul; Asia, púrpura; Osaka 2007, morado; Apeldoorn 2008, rosado; New Orleans 2010, lila; y Sydney 2012, rojo.

Como se puede observar en la figura 5, los cambios en los sitios informativos del dominio P2 ocurrieron en un modo por etapas, mostrando una correlación entre dichos cambios y la aparición de una nueva cepa circulante.

8.2 Modelamiento estructural de la proteína VP1

Los modelos por homología de la proteína de la cápside de NV GII.4, correspondientes a la secuencia consenso de cada clado, fueron generados en base al uso de un molde híbrido de las estructuras cristalinas del dominio P de la proteína VP1 de la cepa VA387 de NV GII.4 (PDB 2OBT) y de la proteína VP1 completa de NV GI (PDB 1IHM). Los valores de la raíz de la desviación cuadrática media (RMSD) obtenidos para las 12 estructuras oscilan entre 0.92 y 1.05 Å, lo cual indica una similitud alta a nivel de los esqueletos carbonados del modelo y las secuencias consenso.

El empleo de un molde híbrido fue justificado al observar que los valores de energía potencial generados (*Discrete Optimized Protein Energy*, DOPE) fueron menores en comparación con el uso de un único modelo (Figura 6), mostrando que ciertas regiones, en especial los residuos ubicados en el subdominio P2 (280 – 405), redujeron sus valores DOPE. Esto se debe a un mejor alineamiento del modelo 2OBT ya que pertenece a una cepa del genotipo GII.4 más cercano en comparación al dominio P del modelo 1IHM de NV GI.

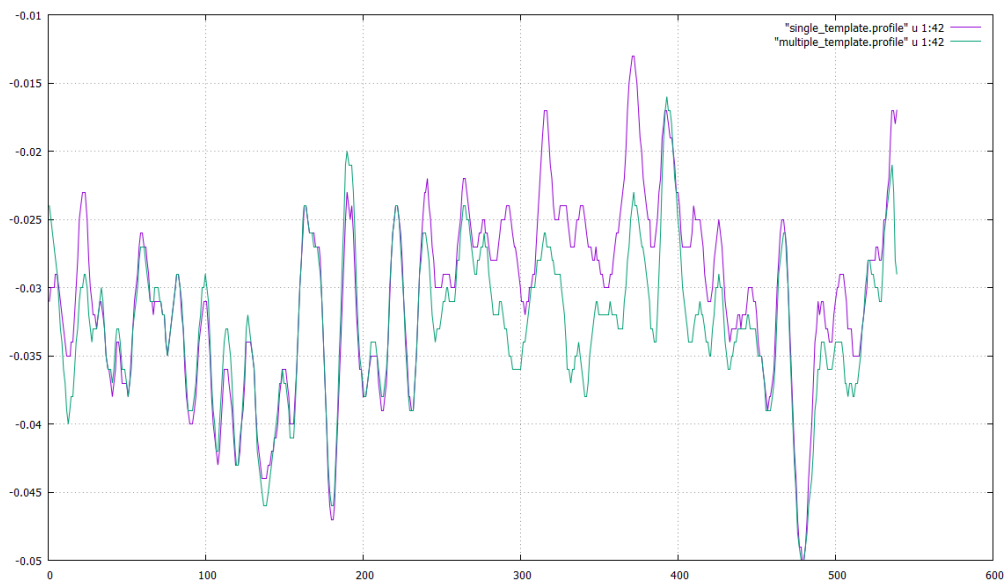


Figura 6. Comparación de los valores de energía DOPE del modelo por homología de la proteína VP1 de NV GII.4 Sydney 2012. El cálculo fue realizado con un molde simple (color lila) y un molde híbrido (color verde). El eje “Y” representa el valor DOPE y el eje “X”, el número de residuo aminoacídico en la secuencia de la proteína.

Con el propósito de examinar si la estructura de VP1 fue sometida a cambios estructurales en los últimos cuarenta años, se realizó la superposición de los modelos correspondientes a las variantes 1970s y Sydney 2012 (Figura 7). Se observó que 526 residuos poseen similitud en cuanto a su ubicación espacial, dando como resultado un valor de RMSD de 0.53 Å e indicando una conservación estructural mayoritaria a nivel estructural durante el periodo mencionado. Sin embargo, al considerar el RMSD entre las secuencias completas de ambas variantes se obtuvo un valor de 2.927; esta diferencia significativa se atribuye principalmente a la orientación de tres bucles expuestos y localizados en el dominio P2 de la

proteína. La variabilidad estructural observada en estas regiones podría sugerir una modulación por parte del virus en cuanto a la especificidad por su respectivo receptor dentro del hospedero.

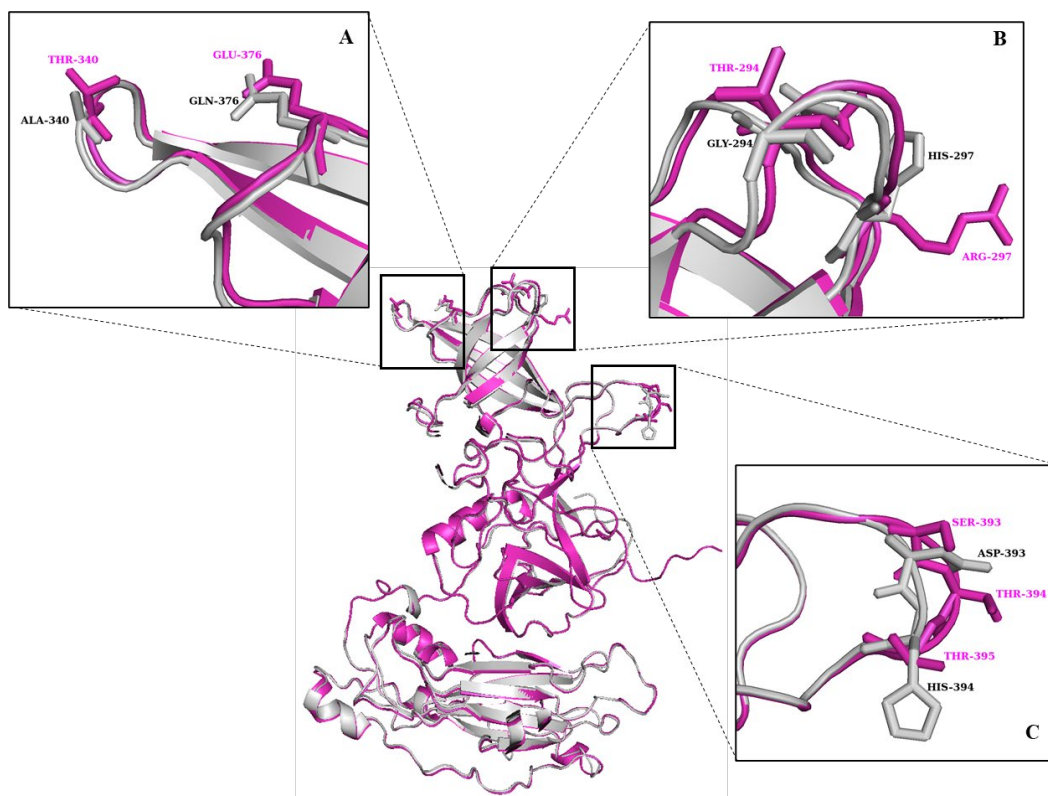


Figura 7. Superposición de las estructuras de la proteína VP1 entre variantes de NV GII.4. Variante 1970s en color gris y variante Sydney 2012 en color magenta. Las cajas negras indican las zonas de mayor variabilidad en el subdominio P2, resaltando los residuos 340 y 376 (A); 294 y 297 (B); 393, 394 y 395 (C).

La calidad de las estructuras obtenidas fue evaluada mediante varios métodos, incluyendo el Z-score, evaluación ERRAT y análisis de Ramachandran (Tabla 2). El Z-score de los modelos varía de -6.71 a -6.23, ubicándose dentro de los valores

aceptables, mientras que el resultado del análisis de Ramachandran mostró que más del 95% de residuos se ubican en la región favorable.

Tabla 2. Evaluación de los modelos 3D de la proteína VP1 de NV GII.4 generados por homología.

Clado	RMSD* (Å)	Z score	Ramachandran			Calidad de modelo ERRAT (%)
			Residuos favorecidos (%)	Residuos permitidos (%)	Residuos atípicos (%)	
1970s	1.03	-6.23	95.5	3.6	0.9	67.297
US 95/96	0.97	-6.44	96.3	3	0.7	70.209
Henry	1.04	-6.27	95.9	3	1.1	71.591
Farmington Hills	0.99	-6.49	96.5	2.4	1.1	68.75
2006a	1	-6.37	95.4	3.3	1.3	72.075
Hunter	0.92	-6.23	95.7	3.5	0.7	74.528
2006b	1.05	-6.42	97.2	2	0.7	72.814
Asia	1	-6.32	96.3	2.8	0.9	74.053
Osaka	1.02	-6.65	95.9	3	1.1	74.858
Apeldoorn	1.02	-6.71	95.7	3	1.3	69.905
New Orleans	1	-6.44	96.1	2.8	1.1	68.053
Sydney	0.97	-6.44	96.3	2.4	1.3	73.423

*RMSD = raíz de la desviación cuadrática media

La evaluación del servidor ERRAT mostró resultados comprendidos entre 67.297% y 74.528% en calidad de modelo, sin embargo, las regiones con mayor error se encontraron en el dominio S y en bucles pertenecientes al dominio P, siendo en este último caso justificable por el grado de libertad que poseen estas estructuras secundarias.

El mapeo de los residuos hipervariables en las estructuras de las diversas variantes, los cuales fueron identificados de acuerdo al color asignado por clado según la figura 5, indicó que la gran mayoría de sustituciones ocurrieron en la superficie del subdominio P2, incluyendo los dominios de unión al receptor HBGA (Figura 8).

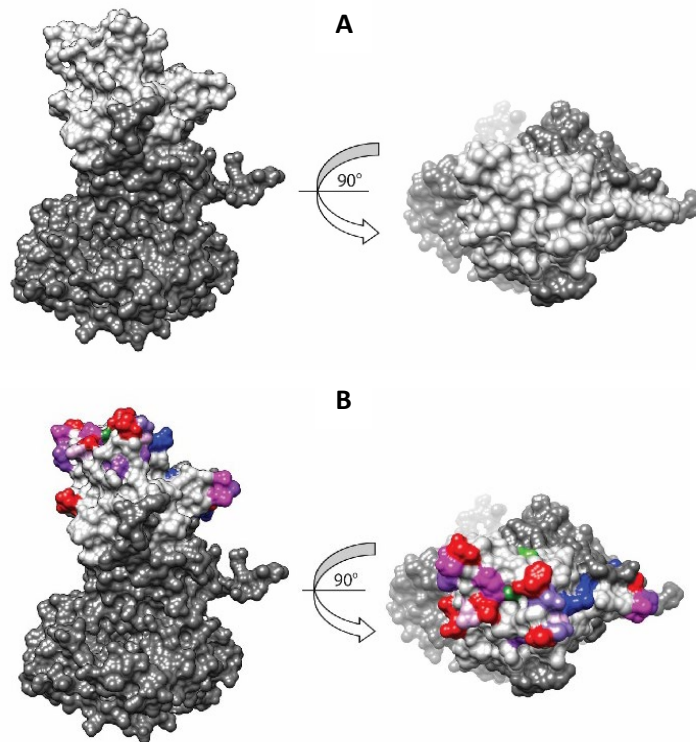


Figura 8. Mapeo de la variación de sitios informativos en la estructura 3D de la proteína VP1 de NV GII.4 entre el año 1970 y el 2015. Los modelos por homología se muestran en color gris oscuro, y el subdominio P2 (parte superior) está de color gris claro. A) Modelo por homología de la variante ancestral GII.4 1970. B) Modelo por homología de la variante GII.4 Sydney 2012 que muestra los distintos cambios acumulados en los sitios informativos hasta el 2015. Cada color representa los residuos asociados a las variantes de NV GII.4: US95/96, amarillo; Henry 2001, verde claro; Farmington Hills 2002, verde oscuro; 2006a, turquesa; Hunter 2004, celeste; 2006b, azul; Asia, púrpura; Osaka 2007, morado; Apeldoorn 2008, rosado; New Orleans 2010, lila; y Sydney 2012, rojo.

8.3 Predicción de epítopes B lineales

Empleando el servidor DISCOTOPE 2.0 se predijeron un total de 65 péptidos como epítopes para células B putativos, de los cuales la gran mayoría se ubicaron en el subdominio P2, considerando a estos últimos para las posteriores predicciones. Se reconocieron cinco regiones potencialmente antigénicas: 1) 287-298, 2) 302-315, 3) 302-316, 4) 370-383, y 5) 389-398, estando todas estas regiones asociadas a conformaciones aleatorias o bucles expuestos en la superficie (Figura 9) e incluyendo los sitios hipervariables descritos en la sección 8.1. Por tal motivo, se optó por considerar las regiones ubicadas exclusivamente en dichas estructuras, asignando así la denominación de epítopes B lineales.

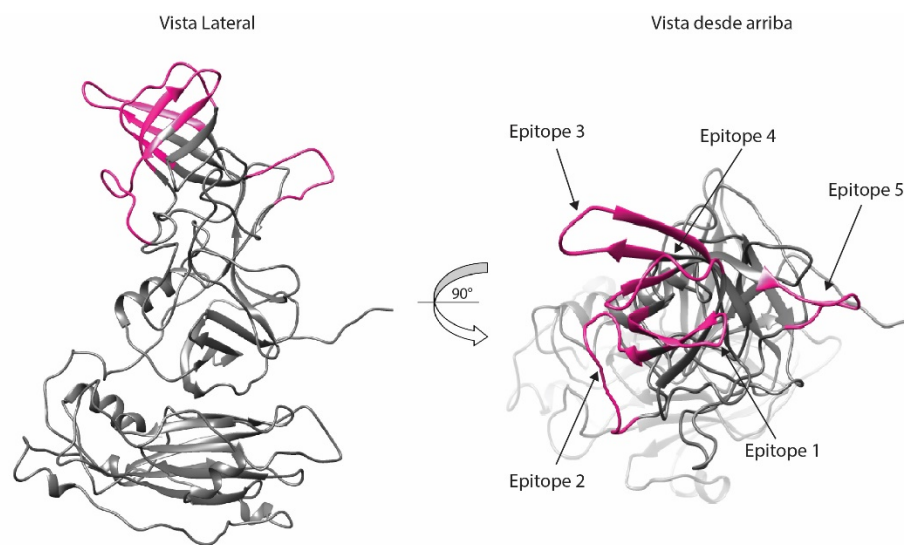


Figura 9. Mapeo de regiones epitópicas en la estructura 3D de la proteína VP1 de NV GII.4 variante Sydney 2012. El modelo estructural se encuentra de color gris mientras que las zonas predichas como potenciales epítopes están de color lila. Las

regiones con mayor puntaje de predicción fueron: epítotope 1 (287-298), epítotope 2 (302-315), epítotope 3 (302-316), epítotope 4 (370-383), y epítotope 5 (389-398).

8.4 Predicción de afinidad a MHC Clase II

El servidor NetMHCIIpan 3.1 predijo la afinidad de péptidos de 15 residuos de longitud, provenientes de las secuencias representativas de los clados de NV GII.4, a moléculas de MHC clase II humanos, teniendo en cuenta los 3 isotipos HLA-DR, HLA-DP y HLA-DQ y los alelos asociados a una cobertura en la población mundial mayor a 99%. Los valores de predicción fueron obtenidos en términos de concentración inhibitoria 50 (IC₅₀), siendo el valor límite de 500 nM usado por defecto para definir afinidad intermedia entre los péptidos y las moléculas MHC clase II y el valor de 50 nM para una afinidad fuerte. De todos los isotipos analizados, sólo se obtuvieron resultados de afinidad significativa para los alelos DQA1_0501-DQB1_0301, DRB1_0101, DRB1_0401, DRB1_0701, DRB1_0901, DRB1_1101, DRB1_1201, y DRB5_0101, dando un total de 23 epítotos candidatos (Tabla 3).

Tabla 3. Epítotos predichos para células B de NV GII.4 con afinidad a moléculas de MHC clase II

Epítotos B	Posición	Alelos MHC clase II	Variante
RGDVTHIAGSHDYT	287-300	DQA1*05:01/DQB1*03:01	US95/96, Farmington Hills 2002, 2006a, Hunter 2006b
RGDVTHIAGT	287-296	DQA1*05:01/DQB1*03:01	
GDVTHIAGSRN	288-298	DQA1*05:01/DQB1*03:01, DRB1*01:01, DRB1*07:01, DRB1*12:01, DRB5*01:01	

RGDVTHIPGTRT	287-298	DQA1*05:01/DQB1*03:01, DRB1*01:01, DRB1*07:01, DRB1*11:01, DRB1*12:01, DRB5*01:01	Asia
RGDVTHIAGS	287-296	DQA1*05:01/DQB1*03:01	Osaka 2007
RGDVTHIPGSRNYT	287-300	DQA1*05:01/DQB1*03:01, DRB1*01:01, DRB1*12:01, DRB5*01:01	New Orleans 2009
RGDVTHITGSRNYT	287-300	DRB1*01:01, DRB1*07:01, DRB1*11:01, DRB1*12:01, DRB5*01:01	Sydney 2012
NLASQNWNNYDPTEE	302-316	DRB1*12:01	US95/96, Farmington Hills 2002, 2006a, Hunter 2004, Asia
NLASQNWNSYDPTE	302-315	DRB1*12:01	New Orleans 2009
NLASQNWNSYDPTE	302-315	DRB1*12:01	Sydney 2012
TQTTRADGSTRA	335-346	DRB1*12:01	1970
TQTTREDGSTRAHK	335-348	DQA1*05:01/DQB1*03:01	US95/96
TQTTRRDGSTRG	335-346	DRB1*12:01	2006a, Hunter 2004
TQTTRADGSTRG	335-346	DRB1*12:01	Apeldoorn 2008
TQTTRTNGSTRGHK	335-348	DQA1*05:01/DQB1*03:01	New Orleans 2009
DTNNDFQTGQNTK	370-382	DRB1*01:01, DRB1*07:01, DRB1*09:01	US95/96
DTNNDFETGQNTKF	370-383	DRB1*01:01	Farmington Hills 2002
TSNDFETGQNTR	371-382	DRB1*01:01	2006a, Hunter 2004
DTDNDFETGQNTR	370-382	DRB1*01:01	Asia
VGDGSTTHQN	389-398	DQA1*05:01/DQB1*03:01	2006a, Hunter 2004

IQDGGTTHRN	389-398	DQA1*05:01/DQB1*03:01	2006b
DGSSAHRN	391-398	DQA1*05:01/DQB1*03:01	Asia
IQDGSTTHRN	389-398	DQA1*05:01/DQB1*03:01	Sydney 2012

El péptido VTHIPGTRT, positivo para DRB1_0101, mostró tener la más alta afinidad ($IC_{50} = 58$ nM) mientras que para el alelo DRB1_1201, el péptido LTQTTRADG mostró el valor IC_{50} más alto (497.23 nM). El epítipo para células T VTHIPGTRT ($IC_{50} = 58$ nM) se ubicó dentro de la secuencia del epítipo B 287 RGDVTHIPGTRT y fue el que mostró mayor promiscuidad entre alelos MHC clase II, teniendo afinidad para los alelos DRB1_0101, DRB1_0701, DRB1_1101, DRB1_1201, DRB5_0101 y DQA1_0501-DQB1_0301. Todos los péptidos con menor valor IC_{50} fueron mapeados en las estructuras 3D predichas para VP1, corroborándose su ubicación en las zonas expuestas de la proteína VP1.

8.5 Conservación y especificidad de los epítopes

En el estudio de conservación y especificidad de epítopes predichos, se obtuvo que para la región 1, el péptido con mayor frecuencia dentro del clado 2006b fue GDVTHIAGSRN (100%), mientras que en el clado Sydney 2012, fue el péptido RGDVTHITGSRNYT (79.43%). En la región 2, sólo dos péptidos pertenecientes a los clados 2006b y 2012 Sidney fueron analizados, teniendo una frecuencia 100% y 48.94% respectivamente. En la región 3, el péptido TQTTRADGSTRG mostró una frecuencia de 69.23% siendo la más alta en este bloque, seguidos de los péptidos TQTTRRDGSTRG (60%) y TQTTRTNGSTRGHK (42.61%). En la región 4, el péptido TSNDFETQGNTR fue el más frecuente y se encontraba distribuido en los clados Hunter 2004 (94.74%) y 2006a (88%) respectivamente, seguido del péptido DTDNDFETGQNTR (85.29%). Finalmente, en la región 5 se

encontraron los péptidos IQDGSTTHRN (51.06%) e IQDGGTTHRN (49.64%) pertenecientes a los clados Sydney 2012 y 2006b respectivamente. En el análisis de BLASTp no se observó similitud significativa con ninguna otra especie, lo que demuestra que los péptidos candidatos son altamente específicos para el genotipo NV GII.4.

8.6 Evaluación de colonias de *E. coli* BL21(DE3) transformadas con el vector pET28a-MEP-NVGII.4

Los resultados de la digestión del plásmido pET28a-MEP-NVGII.4 obtenido de células de *E. coli* BL21 transformadas con dicho vector se ilustran en la Figura 10.

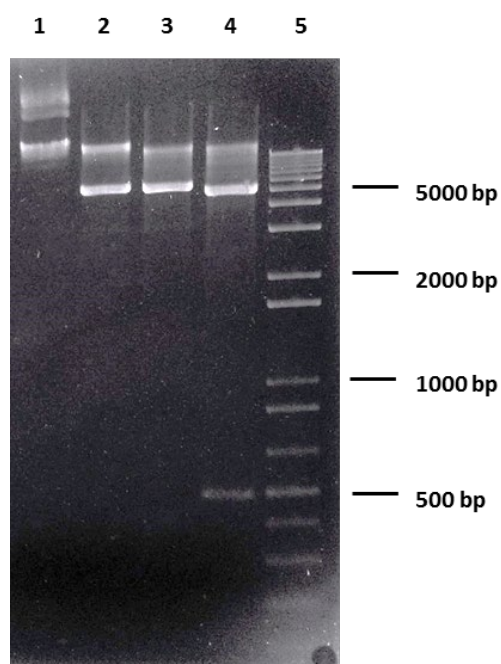


Figura 10. Electroforesis en gel de agarosa de la digestión del plásmido pET28a-MEP-NVGII.4, empleando las enzimas de restricción *Nco*I y *Xho*I. Carril 1: pET28a-MEP-NVGII.4 no digerido, carril 2: pET28a-MEP-NVGII.4 digerido con *Nco*I; carril3: pET28a-MEP-NVGII.4 digerido con *Xho*I; carril 4: Plásmido

pET28a-MEP-NVGII.4 digerido con *NcoI* y *XhoI*; carril 5: Marcador de tamaño de ADN 1 kb Plus.

En el carril 1, se puede observar que el tamaño del plásmido es visible alrededor de la banda 5000 kb, correspondiendo a la forma no enrollada, mientras que en la parte superior se observa una banda de mayor peso correspondiente a la forma enrollada del plásmido. En los carriles 2 y 3 se pueden observar la presencia de la forma linear del plásmido (aproximadamente 5000 kb) en los productos de digestión con las enzimas *NcoI* y *XhoI* respectivamente. En el carril 4, se confirmó el producto de doble digestión *NcoI/XhoI* correspondiente al gen que codifica la proteína MEP-NVGII.4 de 487 kb (~500 kb).

8.7 Inducción y optimización de la expresión

En el presente estudio, la proteína recombinante MEP-NVGII.4 fue expresada en células de *E. coli* BL21 y mediante inducción con IPTG. Los resultados del SDS-PAGE confirmaron la expresión de la proteína recombinante en diferentes intervalos de tiempo de inducción, siendo el peso de la proteína de ~ 18 kDa (Figura 11). A partir de la comparación de bandas en los geles de los diferentes ensayos y considerando que el número de células analizadas luego de los respectivos tiempos de inducción fue normalizado según el OD_{600nm} obtenido, se evidenció que la concentración de IPTG tiene poco efecto en el incremento de la expresión de la proteína recombinante, y, además, que ésta comienza a detectarse de manera progresiva a partir de T = 1 h hasta alcanzar un máximo a T = 2 h, con menores cambios en horas posteriores.

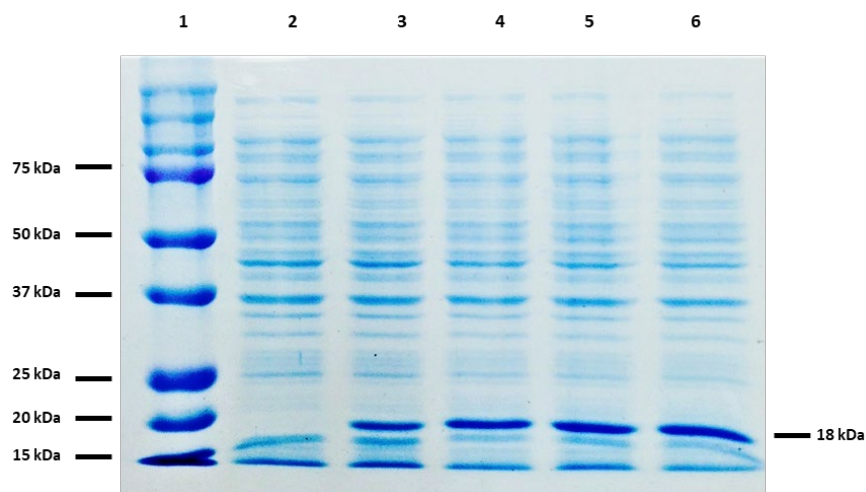


Figura 11. SDS-PAGE de la proteína MEP-NVGII.4 después de la inducción de su expresión con 1 mM de IPTG. Carril 1: Marcador de peso molecular (10 kDa –250 kDa); Carriles 2 – 6: expresión de la proteína multiepitópica después de la inducción con IPTG a T = 0, 1, 2, 3 y 5 h respectivamente.

El estudio de solubilidad en diferentes intervalos de tiempo, mostro que la proteína recombinante estaba presente predominantemente en la fracción insoluble a T = 5 h mientras que los niveles de expresión en la fracción soluble fueron indetectables, razón por la cual se procedió a su purificación en condiciones desnaturalizantes (Figura 12).

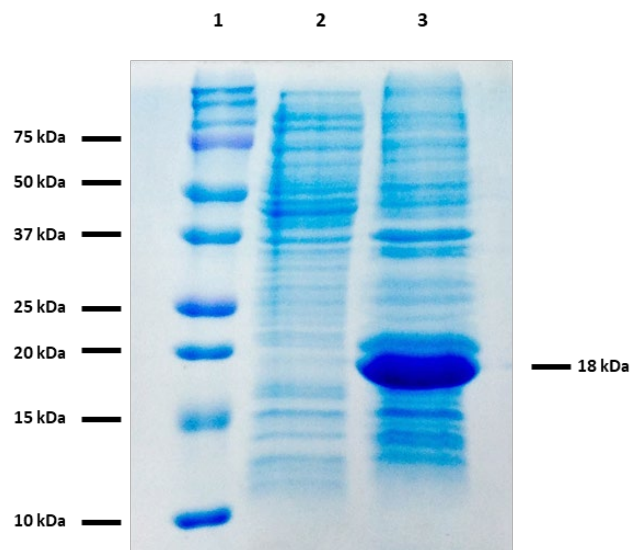


Figura 12. Fracción soluble e insoluble de la proteína MEP-NVGII.4 expresada luego de 5 horas de inducción con 1 mM de IPTG. Carril 1: Marcador de peso molecular (10 kDa – 250 kDa); carril 2: fracción soluble de la proteína multiepitópica; carril 3: fracción insoluble de la proteína multiepitópica.

8.8 Purificación de la proteína multiepitópica

La proteína MEP-NVGII.4 fue purificada a partir de un cultivo de *E. coli* BL21(DE3) con un tiempo de inducción de la expresión de 5 horas, mediante cromatografía de afinidad, en la cual los residuos de histidina de la proteína recombinante se unen a los átomos de níquel inmovilizados en la columna cromatográfica. La presencia de la proteína (~18 kDa) fue confirmada mediante SDS-PAGE tal como se muestra en la Figura 13.

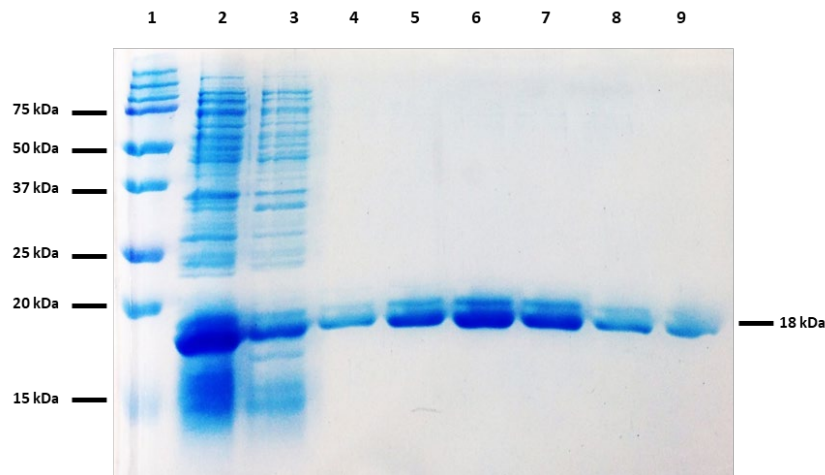


Figura 13. SDS-PAGE de las fracciones de purificación de la proteína MEP-NVGII.4 por cromatografía de afinidad. La proteína multiepitópica fue expresada en células de *E. coli* BL21 a 5 horas después de la inducción. Carril 1: marcador de peso molecular (10 kDa – 250 kDa); carril 2 y 3: elución de carga o *flow-trough*, con tampón de imidazol 20 mM; carriles 4 y 5: elución con tampón de imidazol 40 mM; carriles 6 y 7: elución con tampón de imidazol 60 mM, donde se aprecia una mayor cantidad de proteína eluída; carriles 8 y 9: elución con tampón de imidazol 100 mM.

La concentración de proteína se determinó a partir de la curva de calibración obtenida con el método de Bradford (ver Anexo 8). El ensayo arrojó una concentración de 0.1611 mg/mL para las bandas más intensas (Figura 13, carriles 6 y 7), que luego del proceso de concentración al vacío alcanzó el valor de 1.074 mg/mL.

8.9 Detección de anticuerpos IgG e IgM

Para determinar la reactividad de los anticuerpos IgG e IgM frente a la proteína MEP-NVGII.4, se evaluaron diversos parámetros experimentales de la técnica de ELISA indirecto, entre ellos el tipo de solución bloqueante, y las diluciones adecuadas de los reactivos usados. El empleo de la solución bloqueante de leche descremada al 5% en PBS/Tween 0.5% permitió obtener un menor ruido de fondo según las absorbancias de los controles experimentales (control de antígeno usados (rango de O.D. = 0.043 - 0.1271). La dilución óptima de antígeno y suero fueron de 1µg/mL y 1:100 respectivamente, mientras que las soluciones de conjugado anti-IgG y conjugado anti-IgG e IgM en las que se obtuvo una mejor proporción entre absorbancias de muestras positivas y negativas fue establecida en 1:10000 y 1:5000 respectivamente.

Se emplearon un total de 26 sueros de bebés, de los cuales 20 eran confirmados como positivos y los 6 restantes provinieron de recién nacidos (controles negativos), y los análisis fueron llevados a cabo por duplicado. Todos los sueros mostraron una alta absorbancia para anticuerpos IgG frente a la proteína MEP-NVGII.4 (ver Anexo 9), hallando valores similares entre los bebés infectados (rango de O.D. = 0.2712 – 2.8708) y los bebés recién nacidos (rango de O.D. = 0.5352 – 2.8689). Sin embargo, los valores de absorbancia para anticuerpos IgM en muestras positivas (rango de O.D. = 0.2097 – 2.5662) fueron considerablemente mayores en contraste con las muestras negativas (rango de O.D. = 0.0927 – 0.1890).

8.10 Análisis estadístico

La comparación de medianas de los valores de IgG detectados por la proteína multiepitópica entre los bebés infectados y los bebés recién nacidos no mostró diferencia estadística significativa ($p = 0.4938$) de acuerdo a la prueba de U de Mann-Whitney (Figura 14).

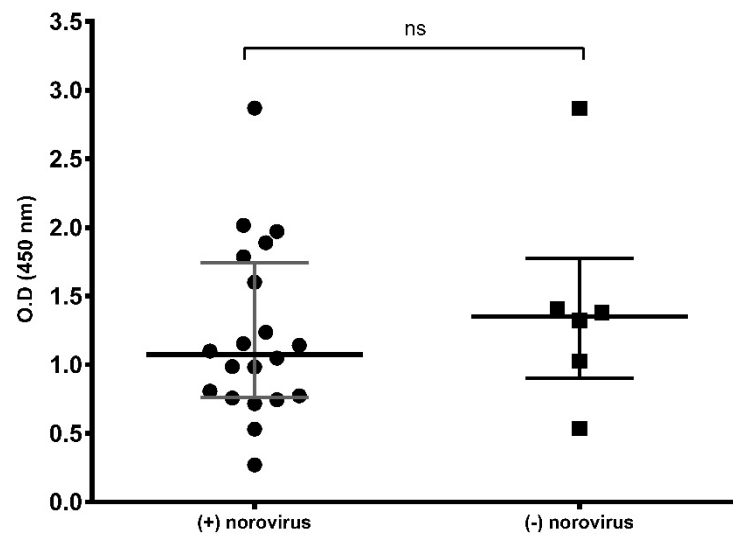


Figura 14. Gráfica de puntos de los valores de OD_{450nm} de anticuerpos IgG frente a la proteína MEP-NVGII.4. La prueba inmunológica se realizó entre dos grupos: (+) norovirus, bebés infectados con NV; y (-) norovirus, bebés recién nacidos. El valor de la mediana de cada grupo está indicado por una línea sólida horizontal. El símbolo ns representa un valor $p > 0.05$, indicando ausencia de diferencia significativa.

Por otro lado, la comparación de medianas de los valores de IgM detectados por la proteína multiepitópica entre los bebés infectados y los bebés recién nacidos mostró diferencia estadística muy significativa ($p < 0.0001$) (Figura 15).

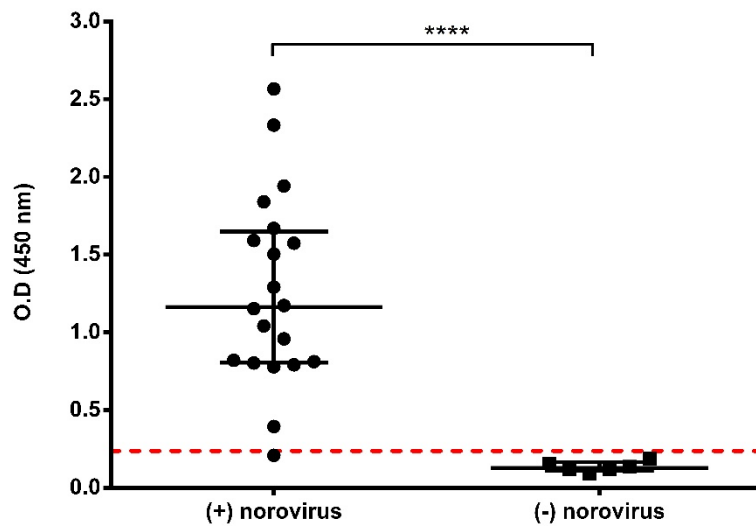


Figura 15. Gráfica de puntos de los valores de OD_{450nm} de anticuerpos IgM frente a la proteína MEP-NVGII.4. La prueba inmunológica se realizó entre dos grupos: (+) norovirus, bebés infectados con NV; y (-) norovirus, bebés recién nacidos. El valor de la mediana de cada grupo está indicado por una línea sólida horizontal mientras que la línea punteada (color rojo) representa el valor de corte según promedio + 3SD, el cual fue de 0.2375. El símbolo **** representa un valor $p < 0.0001$.

Los valores promedio de O.D usando los controles negativos fueron de 0.09265, 0.1169, 0.1194, 0.1352, 0.1581 y 0.189 respectivamente. El promedio fue calculado como 0.1352 y la desviación estándar fue determinada como 0.0341, siendo el valor de corte (promedio + 3SD) de 0.2375.

IX. DISCUSIÓN

9.1 Predicción inmunoinformática

En el presente trabajo se realizó el análisis filogenético, modelamiento por homología, y predicción de epítopes B lineales así como de su afinidad por moléculas de MHC clase II comparando secuencias completas de la proteína VP1 de la cápside de NV GII.4. A través del análisis de secuencias, se confirmaron hallazgos previos de la aparición de brotes de NV GII.4 a lo largo del tiempo a través de un modelo de selección impuesta por la inmunidad colectiva de la población (89), dando lugar a la formación de 12 clados temporales desde el año 1975 hasta la fecha.

Cada nuevo clado que ha emergido estuvo acompañado de mutaciones acumuladas en determinados residuos, denominados sitios informativos, de la proteína VP1. La mayor densidad de estos sitios se ubicó en el subdominio protuberante P2 en comparación con el resto de la secuencia de la proteína de la cápside, obteniendo así evidencia de que existió una fuerza de selección que dispuso a cada nueva variante de NV GII.4 de cierta ventaja evolutiva con respecto a previas cepas circulantes. Algunos de estos residuos han sido señalados como partícipes importantes en la interacción entre la cápside viral y los receptores asociados de tipo HBGA o de grupo sanguíneo presentes en las células epiteliales del intestino, principal puerta de entrada para este agente infeccioso (90). Dentro del subdominio P2, el alineamiento de secuencias mostró que los residuos S₃₄₃, T₃₄₄, R₃₄₅, G₃₄₆, H₃₄₇, y D₃₇₄ presentan un alto grado de conservación a lo largo de todo el genotipo GII.4, siendo éstos los que conforman el pocket de unión al receptor, mientras que

los residuos vecinos presentaron gran variabilidad. Por otro lado, los bloques antigénicos designados en otro estudio como A y B (29) contenían residuos informativos (A: 296-298 y B 393-395) cuya variabilidad estaba muy asociada a la generación de una nueva cepa. Como ejemplo, se observa que el conjunto de residuos T₂₉₆, H₂₉₇, D₂₉₈, N₃₉₃, S₃₉₄, y A₃₉₅ en la variante US95/96 fue reemplazado por T₂₉₆, Q₂₉₇, N₂₉₈, S₃₉₃, T₃₉₄, y T₃₉₅ con la aparición de la variante Hunter 2004 y, posteriormente, a S₂₉₆, R₂₉₇, N₂₉₈, G₃₉₃, T₃₉₄, y T₃₉₅ con el surgimiento de la variante Sidney 2012.

En el análisis estructural, se observó un patrón de plegamiento muy similar entre todas secuencias representativas con respecto a los moldes 1IHM y 2OBT, (RMSD ~ 1%) debido al gran porcentaje de similitud que poseían entre sí. El empleo del cristal 2OBT (NV GII.4/VA387) ha sido ampliamente referido en muchas investigaciones por abarcar el dominio P de la proteína VP1 y por pertenecer específicamente al genotipo GII.4. Por otro lado, no se descartó el uso simultáneo del cristal de la proteína completa de NV GI ya que éste incluye el dominio S, con lo que se tuvo en cuenta en el mapeo de posibles epitopes correspondientes a esta región. Además, la comparación entre las secuencias de NV GI y NV GII señala que este último posee 3 inserciones de 6, 3 y 7 aminoácidos ubicados dentro del subdominio P2 (36), por lo que el modelamiento de esta región fue realizado considerando predominantemente el molde 2OBT. La evaluación de los modelos generados mostró que son válidos para los posteriores análisis de predicción antigénica.

Las secuencias asignadas como potenciales epítopes fueron seleccionadas en base a los resultados de predicción para epítopes de células B, predicción de afinidad a

moléculas de MHC clase II, mapeo en las estructuras generadas por modelamiento tanto en la proteína individual como en la cápside completa, y accesibilidad del solvente. Los epítopes para células B son regiones antigénicas situadas en la superficie de la proteína, los cuales son preferentemente reconocidos por anticuerpos, y, de acuerdo a la continuidad en su secuencia, pueden clasificarse como lineales o conformacionales. Si bien existe la posibilidad de que un segmento corto y continuo de una proteína pueda tener la capacidad de unirse a un anticuerpo específico contra dicha proteína, los estudios cristalográficos de complejos proteína anticuerpo revelan que la mayoría de epítopes son discontinuos, es decir, están conformados de átomos provenientes de residuos distantes agrupados en la superficie. Sin embargo, un péptido sintético puede ser diseñado de tal forma que simule la estructura de un epítipo discontinuo de una proteína. Por lo tanto, una aproximación de estudio es identificar aquellos epítopes que muestren afinidad por moléculas HLA pero que, a su vez, estén ubicadas en las regiones continuas de un epítipo para células B.

Para el presente trabajo, se realizaron predicciones en base a los modelos 3D de cada clado, teniendo en cuenta que los epítopes predichos estén ubicados preferentemente en regiones aleatorias o bucles expuestos en la superficie. En total se reconocieron 5 regiones potencialmente antigénicas, en el rango de 8 a 15 aminoácidos, y todas ubicadas en el subdominio P2.

Diversos anticuerpos monoclonales han sido reportados por tener la propiedad de detectar diferentes partículas virales asociadas a variantes específicas de NV GII.4, y la gran mayoría de los epítopes reconocidos por dichos anticuerpos se localizaron en el dominio P de la proteína VP1. Tres anticuerpos anti-GII-4v0 mostraron

reducir su capacidad de reconocimiento del antígeno en ~80% cuando éste es sometido a mutagénesis dirigida hacia las regiones 294-296 y 393-395, mientras que la unión de otros dos anticuerpos anti-GII-4v2 fue dependiente solo de la región 294-296 (30). Estos dos epitopes reconocidos experimentalmente coinciden con las predicciones llevadas a cabo en el presente estudio, ubicándose dentro de las regiones 287-300 y 389-398 presentes en la mayoría de clados de NV GII.4. En otro estudio, los aminoácidos 294-298 y 368-372 (Epitope A) fueron reconocidos por los anticuerpos monoclonales NVB 114, NVB 111 y NVB 43.9 generados contra un panel de 6 partículas virales representativas de variantes entre los años 1987 y 2009, a la vez que los residuos 393-395 (Epitope D) fueron esenciales para la unión con el anticuerpo NVB 97 (34). A pesar que todos los residuos que conforman el epítape D están contenidos en la región 389-398, sólo dos residuos del epítape A (368 y 369) no fueron considerados como parte de la región 3 (370-381) debido a que el residuo 369 sobrepasó el umbral de puntaje establecido en el servidor de predicción, dando como resultado un epítape no lineal. Otro estudio similar de predicción de epitopes conformacionales tanto para genogrupo I y II identificó, a través del servidor SEPPA, una región antigénica exclusiva para el genogrupo II situada en los residuos 395-406 (35), estando parcialmente superpuesta con la región 5 (389-398) descrita en este estudio.

En contraste con las predicciones de epitopes para células B, los epitopes para células T son péptidos cortos con una longitud de 8-9 aminoácidos, y, a diferencia de los primeros, pueden ser reconocidos en cualquier región de la proteína viral (tanto en la superficie como en la parte interna). La presencia de regiones con afinidad a moléculas MHC de clase II asociadas con epitopes B lineales podría

asegurar que dichos péptidos pudieron haber sido capaces de activar el sistema inmune y, por ende, haber generado la producción de anticuerpos. De esta evaluación se pudieron escoger 19 péptidos *core* con la mejor afinidad a moléculas MHC-II ($IC_{50} < 500$ nM) distribuidos en 23 de los 65 péptidos predichos como epítopes B lineales. Algunos de ellos mostraron ser promiscuos tanto para el isotipo HLA-DQ como para el HLA-DR, mientras que ninguno mostró afinidad significativa por el isotipo HLA-DP (Tabla 3). Los péptidos *core* 290VTHIPGTRT298, 290VTHITGSRN298, 290VTHIAGSRN y 290VTHIPGSRN298 fueron los que tuvieron afinidad a un mayor número de alelos, siendo comunes a la región 1 predicha como epítope de células B, la cual corresponde a sitios antigénicos reconocidos de manera experimental en los estudios mencionados. El estudio de conservación de epítopes determinó que la región 2 proveniente de las secuencias predichas (302NLASQNWNNYDPTEE316) se superponían en un 100% en casi todas las variantes, siendo considerada en la construcción de la proteína multiepitópica. A su vez, para el resto de regiones predichas se escogieron aquellas secuencias que cumplan el criterio de temporalidad deseado y hayan mostrado un cierto grado de conservación en cada clado, haciendo un total de 10 epítopes elegidos. La búsqueda BLAST de estas secuencias frente a todo el repositorio GenBank confirmó que no presentan ninguna homología con proteínas humanas, minimizando la posibilidad de una reacción inmunológica indeseable.

Existen una utilidad potencial de estos péptidos para que puedan ser reconocidos por anticuerpos para NV GII.4 pertenecientes a distintas variantes circulantes, entre ellas, algunas más recientes y otras de clados con más antigüedad. Por lo tanto, la

construcción de una proteína multiepitópica en base a los péptidos mencionados representó una alternativa como antígeno de diagnóstico, con alta probabilidad de reaccionar ante la mayoría de anticuerpos presentes en la actualidad.

9.2 Reactividad de la proteína MEP-NVGII.4 frente a sueros humanos

Los estudios serológicos han representado una herramienta muy valiosa para entender el comportamiento del sistema inmune frente a las infecciones por NV. Por medio de la producción de partículas virales representativas de los brotes donde fueron diseñadas, su utilidad como antígenos para pruebas de diagnóstico, así como para estudios epidemiológicos ha sido reconocida ampliamente en las décadas anteriores. Sin embargo, hoy en día los requerimientos tanto a nivel de personal adecuado como infraestructura en el laboratorio con el fin de desarrollar este tipo de material biológico impiden en nuestro entorno local considerar a las técnicas de inmunoensayo como pruebas rápidas de diagnóstico, especialmente en trabajos de campo. Además, hasta la fecha, todos los intentos por generar partículas virales a través del sistema de baculovirus no han sido reproducibles en comparación con trabajos de rutina de otros países (Datos no publicados).

Las proteínas multiepitópicas han simbolizado no sólo una alternativa menos costosa a los antígenos convencionales, sino también, y de acuerdo a las herramientas y procedimientos bioinformáticos empleados para su diseño, pueden ser altamente específicos para un determinado patógeno. Actualmente, no hay evidencia del uso de esta aproximación para el inmunodiagnóstico de NV por lo que los resultados obtenidos en el presente estudio son los primeros en base a combinar técnicas computacionales y experimentales.

En cuanto a los detalles técnicos observados durante el *checkerboard* para optimizar las condiciones de los inmunoensayos, se pudo destacar el hecho de que no hubo la necesidad de realizar el replegamiento de la proteína a partir de la solución obtenida tras la purificación. Siendo una proteína conformada íntegramente por secuencias predichas de poseer conformaciones aleatorias, resultó bastante ventajoso mantenerla bajo condiciones desnaturalizantes ya que permitió emplearla directamente en la etapa del inmunoensayo y, además, otorgar cierta flexibilidad a la proteína al momento de su fijación en las placas de ELISA, de tal modo que los epítopes individuales puedan mantenerse expuestos para su reconocimiento.

En primer lugar, se diseñó y optimizó una metodología de ELISA indirecto para detectar anticuerpos IgG en suero humano. La proteína MEP-NVGII.4 obtenida por predicción inmunoinformática mostró reactividad notable frente a anticuerpos IgG en el 96% de los casos (Figura 14), observando resultados similares tanto en las muestras de bebés confirmados con infección a NV GII.4 como en las muestras de bebés recién nacidos, considerados como controles negativos en este estudio. Este hecho se explica debido al rol que cumplen los anticuerpos IgG de la madre en la protección del bebé, los cuales son transferidos durante el desarrollo del feto por vía transplacentaria. Estos anticuerpos pueden permanecer en el individuo hasta mostrar una reducción considerable luego de un periodo de entre 3 y 6 meses (49); sin embargo, dichos niveles séricos pueden ser detectables hasta un periodo de 9 meses de edad. Además, cabe señalar que el valor de corte para esta metodología no pudo ser obtenido adecuadamente por la falta de un número requerido de muestras con reactividad negativa. Por todo lo anterior dicho, la utilidad de emplear

a los anticuerpos IgG como marcador de infección por NV GII.4 resultó ser muy limitada.

En segundo lugar, se realizó una metodología similar para la detección de anticuerpos IgM específicos contra NV GII.4. Los resultados mostraron una reactividad muy alta en los individuos con infección confirmada en comparación con los controles negativos (Figura 15), razón por la cual se obtuvo una diferencia estadística significativamente alta ($p < 0.0001$). La capacidad de un antígeno para diagnosticar una infección por NV a través de la seroconversión de anticuerpos ha sido indicada en algunos estudios. Sin embargo, se requiere al menos un levantamiento de 4 veces en los niveles séricos comparando una muestra en fase aguda temprana con otra en fase convaleciente; en los resultados anteriormente mostrados, y considerando los valores máximos y mínimos de respuesta frente a IgM, la proteína multiepitópica pudo ser capaz de producir un levantamiento de 27 veces en los niveles séricos. Se sabe que la respuesta inmunológica a través de la producción de anticuerpos IgM es detectable aproximadamente luego de dos semanas de iniciada la infección y que mantiene valores séricos altos hasta por 100 días, según estudios clínicos (50). Tomando en cuenta que las muestras de suero de bebés infectados fueron obtenidas en un intervalo de 1 a 2 meses de haber confirmado una infección reciente con NV GII.4, se puede dar por demostrado que la metodología establecida es adecuada para evaluar la presencia de una infección por NV, ya sea aguda o convaleciente.

Adicionalmente, en los resultados mostrados sólo hubo un caso de bebé infectado cuya respuesta inmunológica tanto para IgG como para IgM fue muy baja. La hipótesis de una falta de especificidad esperada para otros genogrupos de NV, como

es el caso del genogrupo GI, fue descartada al confirmar que la cepa obtenida de la muestra pertenecía al genotipo GII.4. Otra posible explicación se relaciona con la propiedad de algunos individuos de ser genéticamente resistentes a la infección por NV, (fenotipo no secretor) lo que obstaculizaría el desarrollo de la replicación viral y, por ende, la formación de anticuerpos específicos. Sin embargo, se necesitan pruebas específicas para confirmar esta hipótesis.

Es importante señalar también que el tiempo en el cual se obtuvieron las muestras de sangre provienen de una época en la cual la cepa circulante actual, GII.4 Sydney 2012, aun no existía, o por lo menos, no en nuestro entorno. Esta información es relevante y apoya la utilidad de la proteína multiepitópica en la detección de anticuerpos circulantes correspondientes a variantes de años anteriores, ya que en el diseño de la misma se abarcaron las variantes GII.4 2006, siendo estas las más prevalentes de dicho estudio de cohorte. De esto último, se deduce la necesidad de evaluar y renovar la estructura de la proteína multiepitópica cada cierto periodo de tiempo, resaltando de este modo el papel que cumple el criterio de temporalidad al permitirnos mantener la especificidad deseada con la inclusión de epítopes provenientes de las nuevas variantes emergentes.

Finalmente, es importante señalar que a pesar de haber obtenido resultados que van acorde con la información expuesta en la literatura, la obtención de muestras tuvo muchas limitaciones debido a la población en estudio, razón por la cual los resultados mostrados conforman una base potencial para futuros ensayos en el estudio la respuesta inmune frente a NV en nuestro entorno.

X. CONCLUSIONES

1. El análisis inmunoinformático de modelos estructurales consenso para cada variante de norovirus GII.4 logró reconocer epítopes B lineales ubicados preferentemente en bucles expuestos.
2. Los epítopes predichos obtuvieron puntajes de afinidad por MHC-II relativamente altos y en diferentes alelos de mayor distribución, lo que permitió asegurar la capacidad *in silico* de reconocimiento por anticuerpos específicos.
3. La proteína multiepitópica posee la capacidad potencial de diagnosticar una infección aguda y/o convaleciente a través de la detección de niveles séricos de IgM.

XI. RECOMENDACIONES

1. En base a la capacidad de reconocimiento de la proteína multiepitópica por parte de los anticuerpos séricos IgG e IgM, se debería establecer una metodología de inmunoensayo similar para la detección de anticuerpos específicos IgA en muestras de saliva y heces.
2. Se podría llevar a cabo un estudio de cohorte de madres que posean reactividad IgG negativa ante la proteína multiepitópica, con el objetivo de obtener y evaluar adecuadamente controles negativos derivados de sueros de los bebés recién nacidos.
3. Evaluar la reactividad inmunológica de la proteína multiepitópica en ausencia de la cola de histidina.
4. Sintetizar y evaluar la capacidad de detección de IgG e IgM de los epítopes individuales predichos en el estudio mediante técnicas de *cross-linking* tanto en nanopartículas como en dendrímeros.
5. Realizar la búsqueda y evaluar a los epítopes B lineales ubicados preferentemente en la zona más expuesta del dominio P1 de la proteína VP1 de NV GII.4.

XII. REFERENCIAS BIBLIOGRÁFICAS

1. Zahorsky J. Hyperemesis Hiemis or the Winter Vomiting Disease. Archives de Pédiatrie. 1929; 46:391.
2. Adler JL, Zickl R. Winter vomiting disease. J Infect Dis. 1969 Jun;119(6):668–73.
3. Kapikian AZ, Wyatt RG, Dolin R, Thornhill TS, Kalica AR, Chanock RM. Visualization by Immune Electron Microscopy of a 27-nm Particle Associated with Acute Infectious Nonbacterial Gastroenteritis. J Virol. 1972;10(5):1075–81.
4. Kirk MD, Pires SM, Black RE, Caipo M, Crump JA, Devleesschauwer B, et al. World Health Organization Estimates of the Global and Regional Disease Burden of 22 Foodborne Bacterial, Protozoal, and Viral Diseases, 2010: A Data Synthesis. PLoS Med. 2015;12(12):1–21.
5. Lopman B, Atmar R, Baric R, Estes M, Green K, Glass R, et al. Global burden of norovirus and prospects for vaccine development. Centers for Disease Control and Prevention. 2015;1–46.
6. Jiang XI, Wang J, Graham DY, Estes MK. Detection of Norwalk Virus in Stool by Polymerase Chain Reaction. J Clin Microbiol. 1992; 30(10).
7. Jiang X, Graham DY, Wang K, Estes MK. Norwalk virus genome cloning and characterization. Science (1979). 1990;250(4987):1580–3.

8. Jiang X, Wang M, Wang K, Estes MK. Sequence and Genomic Organization of Norwalk Virus. *Virology*. 1993; 195:51–61.
9. Fernandez-Vega V, Sosnovtsev S V, King AD, Mitra T, Gorbalenya A, Green KY. Norwalk Virus N-Terminal Nonstructural Protein Is Associated with Disassembly of the Golgi Complex in Transfected Cells. *J Virol*. 2004;78(9):4827–37.
10. Pfister T, Wimmer E. Polypeptide p41 of a Norwalk-Like Virus Is a Nucleic Acid-Independent Nucleoside Triphosphatase. *J Virol*. 2001;75(4):1611–9.
11. Sharp TM, Crawford SE, Ajami NJ, Neill F, Atmar RL, Katayama K, et al. Secretory pathway antagonism by calicivirus homologues of Norwalk virus nonstructural protein p22 is restricted to noroviruses. *Virol J*. 2012; 9:181.
12. Rohayem J, Robel I, Jager K, Scheffler U, Rudolph W. Protein-primed and de novo initiation of RNA synthesis by norovirus 3Dpol. *J Virol*. 2006;80(14):7060–9.
13. Chung L, Bailey D, Leen EN, Emmott EP, Chaudhry Y, Roberts LO, et al. Norovirus Translation Requires an Interaction between the C Terminus of the Genome-linked Viral Protein VPg and Eukaryotic Translation Initiation Factor 4G *. *J Biol Chem*. 2014;289(31):21738–50.
14. Belliot G, Sosnovtsev S V, Mitra T, Hammer C, Garfield M, Green KY. In Vitro Proteolytic Processing of the MD145 Norovirus ORF1 Nonstructural Polyprotein Yields Stable Precursors and Products Similar to Those Detected in Calicivirus-Infected Cells. *J Virol*. 2003;77(20):10957–74.

15. Rohayem J, Jäger K, Robel I, Scheffler U, Temme A, Rudolph W. Characterization of norovirus 3Dpol RNA-dependent RNA polymerase activity and initiation of RNA synthesis. *Journal of General Virology*. 2006;87(9):2621–30.
16. Glass PJ, White LJ, Ball JM, Leparac-Goffart I, Hardy ME, Estes MK. Norwalk virus open reading frame 3 encodes a minor structural protein. *J Virol*. 2000;74(14):6581–91.
17. Jiang XI, Wang MIN, Graham DY, Estes MK. Expression, Self-Assembly, and Antigenicity of the Norwalk Virus Capsid Protein. *J Virol*. 1992;66(11):6527–32.
18. Prasad BVV, Hardy ME, Dokland T, Bella J, Rossmann MG, Estes MK. X-ray Crystallographic Structure of the Norwalk Virus Capsid. *Science* (1979). 1999;286(October):287–90.
19. Green KY. Caliciviridae: The Noroviruses. In: Knipe DM, Howley PM, editors. *Fields Virology*. 6th ed. Philadelphia: Lippincott Williams & Wilkins; 2013. p. 582–608.
20. Ettayebi K, Crawford SE, Murakami K, Broughman JR, Karandikar U, Tenge VR, et al. Replication of human noroviruses in stem cell-derived human enteroids. *Science* (1979). 2016;353(6306):1387–93.
21. Duizer E, Schwab KJ, Neill FH, Atmar RL, Koopmans MPG, Estes MK. Laboratory efforts to cultivate noroviruses. *Journal of General Virology*. 2004;85(1):79–87.

22. Vinjé J. Advances in laboratory methods for detection and typing of norovirus. *J Clin Microbiol.* 2015;53(2):373–81.
23. Zheng DP, Ando T, Fankhauser RL, Beard RS, Glass RI, Monroe SS. Norovirus classification and proposed strain nomenclature. *Virology.* 2006;346(2):312–23.
24. Kroneman A, Vega E, Vennema H, Vinjé J, White PA, Hansman G, et al. Proposal for a unified norovirus nomenclature and genotyping. *Arch Virol.* 2013;158(10):2059–68.
25. Hoa Tran TN, Trainor E, Nakagomi T, Cunliffe NA, Nakagomi O. Molecular epidemiology of noroviruses associated with acute sporadic gastroenteritis in children: Global distribution of genogroups, genotypes and GII.4 variants. *Journal of Clinical Virology.* 2013;56(3):185–93.
26. Lochridge VP, Jutila KL, Graff JW, Hardy ME. Epitopes in the P2 domain of norovirus VP1 recognized by monoclonal antibodies that block cell interactions. *J Gen Virol.* 2005 Oct;86(Pt 10):2799–806.
27. Tan M, Huang P, Meller J, Zhong W, Farkas T, Jiang X. Mutations within the P2 Domain of Norovirus Capsid Affect Binding to Human Histo-Blood Group Antigens: Evidence for a Binding Pocket. *J Virol.* 2003 Nov;77(23):12562–71.
28. Lindesmith LC, Donaldson EF, Lobue AD, Cannon JL, Zheng DP, Vinje J, et al. Mechanisms of GII.4 norovirus persistence in human populations. *PLoS Med.* 2008 Feb;5(2): e31.

29. Allen DJ, Gray JJ, Gallimore CI, Xerry J, Iturriza-go M, Infections CF, et al. Analysis of Amino Acid Variation in the P2 Domain of the GII-4 Norovirus VP1 Protein Reveals Putative Variant- Specific Epitopes. PLoS One. 2008;(1):1–9.
30. Allen DJ, Noad R, Samuel D, Gray JJ, Roy P, Iturriza-Gómara M. Characterisation of a GII-4 norovirus variant-specific surface-exposed site involved in antibody binding. Virol J. 2009 Jan; 6:150.
31. Lindesmith LC, Donaldson EF, Lobue AD, Cannon JL, Zheng DP, Vinje J, et al. Mechanisms of GII.4 norovirus persistence in human populations. PLoS Med. 2008 Feb;5(2): e31.
32. Debbink K, Donaldson EF, Lindesmith LC, Baric RS. Genetic mapping of a highly variable norovirus GII.4 blockade epitope: potential role in escape from human herd immunity. J Virol. 2012 Jan;86(2):1214–26.
33. Zakikhany K, Allen DJ, Brown D, Iturriza-go M. Molecular Evolution of GII-4 Norovirus Strains. PLoS One. 2012;7(7).
34. Lindesmith LC, Beltramello M, Donaldson EF, Corti D, Swanstrom J, Debbink K, et al. Immunogenetic mechanisms driving norovirus GII.4 antigenic variation. PLoS Pathog. 2012 Jan;8(5): e1002705.
35. Chen L, Wu D, Ji L, Wu X, Xu D, Cao Z, et al. Bioinformatics analysis of the epitope regions for norovirus capsid protein. BMC Bioinformatics. 2013 Jan;14 Suppl 4(Suppl 4): S5.

36. Siebenga JJ, Vennema H, Renckens B, de Bruin E, van der Veer B, Siezen RJ, et al. Epochal evolution of GGII.4 norovirus capsid proteins from 1995 to 2006. *J Virol.* 2007 Sep;81(18):9932–41.
37. Debbink K, Lindesmith LC, Donaldson EF, Baric RS. Norovirus immunity and the great escape. *PLoS Pathog.* 2012 Jan;8(10):e1002921.
38. Bull R a, Eden JS, Rawlinson WD, White P a. Rapid evolution of pandemic noroviruses of the GII.4 lineage. *PLoS Pathog.* 2010 Mar;6(3):e1000831.
39. Bok K, Abente EJ, Realpe-Quintero M, Mitra T, Sosnovtsev S V., Kapikian a. Z, et al. Evolutionary Dynamics of GII.4 Noroviruses over a 34-Year Period. *J Virol.* 2009 Sep;83(22):11890–901.
40. Hensley SE, Das SR, Bailey AL, Schmidt LM, Hickman HD, Jayaraman A, et al. Hemagglutinin receptor binding avidity drives influenza A virus antigenic drift. *Science (1979).* 2009 Oct;326(5953):734–6.
41. Tan M, Xia M, Chen Y, Bu W, Hegde RS, Meller J, et al. Conservation of carbohydrate binding interfaces: evidence of human HBGA selection in norovirus evolution. *PLoS One.* 2009 Jan;4(4): e5058.
42. Gary GW, Anderson LJ, Keswick BH, Johnson PC, Dupont HL, Stine SE, et al. Norwalk Virus Antigen and Antibody Response in an Adult Volunteer Study. *J Clin Microbiol.* 2003;25(10):2001–3.
43. White LJ, Ball JM, Hardy ME, Tanaka TN, Kitamoto N, Estes MK, et al. Attachment and Entry of Recombinant Norwalk Virus Capsids to Cultured Human and Animal Cell Lines. *J Virol.* 1996;70(10):6589–97.

44. Ruvoën-clouet N, Ganière JP, André-fontaine G, Blanchard D, Pendu JLE. Binding of Rabbit Hemorrhagic Disease Virus to Antigens of the ABH Histo-Blood Group Family. *J Virol.* 2000;74(24):11950–4.
45. Hutson AM, Atmar RL, Graham DY, Estes MK. Norwalk Virus Infection and Disease Is Associated with ABO Histo – Blood Group Type. *Journal of Infectious Diseases.* 2002; 185:1335–7.
46. Huang P, Farkas T, Zhong W, Ruvoe N, Morrow AL, Altaye M, et al. Noroviruses Bind to Human ABO, Lewis, and Secretor Histo–Blood Group Antigens: Identification of 4 Distinct Strain-Specific Patterns. *J Infect Dis.* 2003; 3039:19–31.
47. Lindesmith L, Moe C, Marionneau S, Ruvoen N, Iang XII, Indblad LAL, et al. Human susceptibility and resistance to Norwalk virus infection. *Nat Med.* 2003;9(5).
48. Erdman DD, Gary GW, Anderson LJ. Serum Immunoglobulin A Response to Norwalk Virus Infection. *J Clin Microbiol.* 1989;27(6):1417–8.
49. Gray JJ, Cunliffe C, Ball J, Graham DY, Desselberger U, Estes MK. Detection of Immunoglobulin M (IgM), IgA, and IgG Norwalk Virus-Specific Antibodies by Indirect Enzyme-Linked Immunosorbent Assay with Baculovirus-Expressed Norwalk Virus Capsid Antigen in Adult Volunteers Challenged with Norwalk Virus. *J Clin Microbiol.* 1994;32(12):3059–63.
50. Cukor G, Nowak NA, Blacklow NR. Immunoglobulin M Responses to the Norwalk Virus of Gastroenteritis. *Infect Immun.* 1982;37(2):463–8.

51. Brinker JP, Blacklow NR, Estes MK, Moe CL, Schwab KJ, Herrmann JE. Detection of Norwalk Virus and Other Genogroup 1 Human Caliciviruses by a Monoclonal Antibody, Recombinant- Antigen-Based Immunoglobulin M Capture Enzyme Immunoassay. *J Clin Microbiol.* 1998;36(4):1064–9.
52. Noel JS, Ando T, Leite JP, Green KY, Dingle KE, Estes MK, et al. Correlation of patient immune responses with genetically characterized small round-structured viruses involved in outbreaks of nonbacterial acute gastroenteritis in the United States, 1990 to 1995. *J Med Virol.* 1997;53(4):372–83.
53. Ponterio E, Petrizzo A, Di Bartolo I, Buonaguro FM, Buonaguro L, Ruggeri FM. Pattern of activation of human antigen presenting cells by genotype GII.4 norovirus virus-like particles. *J Transl Med.* 2013;11(1):127.
54. Lindesmith LC, Donaldson E, Leon J, Moe CL, Frelinger JA, Johnston RE, et al. Heterotypic Humoral and Cellular Immune Responses following Norwalk Virus Infection. *J Virol.* 2010;84(4):1800–15.
55. Lindesmith L, Moe C, Lependu J, Frelinger a, Treanor J, Baric RS. Cellular and Humoral Immunity following Snow Mountain Virus Challenge. *J Virol.* 2005;79(5):2900.
56. Newman KL, Moe CL, Kirby AE, Flanders WD, Parkos CA, Leon JS. Norovirus in symptomatic and asymptomatic individuals: Cytokines and viral shedding. *Clin Exp Immunol.* 2016;184(3):347–57.
57. Robilotti E, Deresinski S, Pinsky BA. Norovirus. *Clin Microbiol Rev.* 2015;28(1):134–64.

58. Lee RM, Lessler J, Lee RA, Rudolph KE, Reich NG, Perl TM, et al. Incubation periods of viral gastroenteritis: a systematic review. *BMC Infect Dis.* 2013;13(1):446.
59. Teunis PFM, Moe CL, Liu P, Miller SE, Lindesmith L, Baric RS, et al. Norwalk Virus: How Infectious is It? *J Med Virol.* 2008; 80:1468–76.
60. Ahmed SM, Hall AJ, Robinson AE, Verhoef L, Premkumar P, Parashar UD, et al. Global prevalence of norovirus in cases of gastroenteritis: A systematic review and meta-analysis. *Lancet Infect Dis.* 2014;14(8):725–30.
61. Saito M, Goel-Apaza S, Espetia S, Velasquez D, Cabrera L, Loli S, et al. Multiple norovirus infections in a birth cohort in a peruvian periurban community. *Clinical Infectious Diseases.* 2014;58(4):483–91.
62. Allen DJ, Trainor E, Callaghan A, O’Brien SJ, Cunliffe NA, Iturriza-Gómara M. Early detection of epidemic GII-4 norovirus strains in UK and Malawi: Role of surveillance of sporadic acute gastroenteritis in anticipating global epidemics. *PLoS One.* 2016 Apr 1;11(4).
63. Parashar U, Quiroz ES, Mounts AW, Monroe SS, Fankhauser RL, Ando T, et al. “Norwalk-like viruses”. Public health consequences and outbreak management. *MMWR Recommendations and reports: Morbidity and mortality weekly report Recommendations and reports / Centers for Disease Control.* 2001 Jun;50(RR-9):1–17.

64. Vinjé J, Hamidjaja R a., Sobsey MD. Development and application of a capsid VP1 (region D) based reverse transcription PCR assay for genotyping of genogroup I and II noroviruses. *J Virol Methods*. 2004 Mar;116(2):109–17.
65. Wildermann N, Porto-espinoza L, Moronta R, Bracho M, Costa L, Callejas D. Detección molecular mediante RT-PCR de calicivirus y enterovirus en niños menores de 6 años con síndrome diarreico. *Revista de la Sociedad Venezolana de Microbiología*. 2010; 30:145–50.
66. Yan Y, Wang H hui, Gao L, Ji J mei, Ge Z jie, Zhu X qiang, et al. A one-step multiplex real-time RT-PCR assay for rapid and simultaneous detection of human norovirus genogroup I, II and IV. *J Virol Methods*. 2013 May;189(2):277–82.
67. Madore HP, Treanor JJ, Pray KA, Dolin R. Enzyme-Linked Immunosorbent Assays for Snow Mountain and Norwalk Agents of Viral Gastroenteritis. *J Clin Microbiol*. 1986;24(3):456–9.
68. Greenberg HB, Wyatt RG, Valdesuso J, Kalica AR, London WT, Chanock RM, et al. Solid-phase microtiter radioimmunoassay for detection of the Norwalk strain of acute nonbacterial, epidemic gastroenteritis virus and its antibodies. *J Med Virol*. 1978;2(2):97–108.
69. Jiang X, Wilton N, Zhong WM, Farkas T, Huang PW, Barrett E, et al. Diagnosis of human caliciviruses by use of enzyme immunoassays. *J Infect Dis*. 2000 May;181 Suppl: S349-59.

70. Burton-MacLeod J a, Kane EM, Beard RS, Hadley L a, Glass RI, Ando T. Evaluation and comparison of two commercial enzyme-linked immunosorbent assay kits for detection of antigenically diverse human noroviruses in stool samples. *J Clin Microbiol*. 2004 Jun;42(6):2587–95.
71. Takanashi S, Okame M, Shiota T, Takagi M, Yagyu F, Tung PG, et al. Development of a rapid immunochromatographic test for noroviruses genogroups I and II. *J Virol Methods*. 2008 Mar;148(1–2):1–8.
72. Morillo SG, Luchs A, Cilli A, Ribeiro CD, Calux SJ, Carmona RDCC, et al. Norovirus 3rd Generation kit: an improvement for rapid diagnosis of sporadic gastroenteritis cases and valuable for outbreak detection. *J Virol Methods*. 2011 Apr;173(1):13–6.
73. Bruggink LD, Witlox KJ, Sameer R, Catton MG, Marshall J a. Evaluation of the RIDA(®)QUICK immunochromatographic norovirus detection assay using specimens from Australian gastroenteritis incidents. *J Virol Methods*. 2011 Apr;173(1):121–6.
74. Amin MR, Siddiqui MS, Ahmed D, Ahmed F, Hossain A. B- and T-cell epitope mapping of human sapovirus capsid protein: an immunomics approach. *Int J Bioinform Res Appl*. 2011;7(3):287–98.
75. Kringelum JV, Lundegaard C, Lund O, Nielsen M. Reliable B Cell Epitope Predictions: Impacts of Method Development and Improved Benchmarking. *PLoS Comput Biol*. 2012;8(12).

76. Hamelryck T. An amino acid has two sides: A new 2D measure provides a different view of solvent exposure. *Proteins: Structure, Function and Genetics*. 2005;59(1):38–48.
77. Thornton JM, Edwards MS, Taylor WR, Barlow DJ. Location of “continuous” antigenic determinants in the protruding regions of proteins. *EMBO J*. 1986;5(2):409–13.
78. Nielsen M, Lund O, Buus S, Lundegaard C. MHC Class II epitope predictive algorithms. *Immunology*. 2010;130(3):319–28.
79. Robinson J, Halliwell JA, Hayhurst JD, Flicek P, Parham P, Marsh SGE. The IPD and IMGT/HLA database: Allele variant databases. *Nucleic Acids Res*. 2015;43(D1): D423–31.
80. Karosiene E, Rasmussen M, Blicher T, Lund O, Buus S, Nielsen M. NetMHCIIpan-3.0, a common pan-specific MHC class II prediction method including all three human MHC class II isotypes, HLA-DR, HLA-DP and HLA-DQ. *Immunogenetics*. 2013;65(10):711–24.
81. Souza MQ De, Galdino AS, Carlos J, Soares MV, Nóbrega YC De, Morales C, et al. A Recombinant Multiepitope Protein for Hepatitis B Diagnosis. 2013;2013.
82. Shenyang G, Dandan L, Chen F, Shuliang W, Tiezhong Z. Design and evaluation of a recombinant multi-epitope-based ELISA for the serological surveillance of HEV infection in northern China. *Arch Virol*. 2011; 156:1621–6.

83. AnandaRao R, Swaminathan S, Fernando S, Jana AM, Khanna N. A custom-designed recombinant multiepitope protein as a dengue diagnostic reagent. *Protein Expr Purif.* 2005; 41:136–47.
84. Dai J, Jiang M, Wang Y, Qu L, Gong R, Si J. Evaluation of a recombinant multiepitope peptide for serodiagnosis of *Toxoplasma gondii* infection. *Clinical and Vaccine Immunology.* 2012; 19:338–42.
85. Dai JF, Jiang M, Qu LL, Sun L, Wang YY, Gong LL, et al. *Toxoplasma gondii*: Enzyme-linked immunosorbent assay based on a recombinant multi-epitope peptide for distinguishing recent from past infection in human sera. *Exp Parasitol.* 2013; 133:95–100.
86. Faria R, Veloso LDC, Coura-vital W, Reis AB. Novel Recombinant Multiepitope Proteins for the Diagnosis of Asymptomatic *Leishmania infantum*- Infected Dogs. 2015;9(1):13–6.
87. Cao S, Lou Z, Tan M, Chen Y, Liu Y, Zhang Z, et al. Structural basis for the recognition of blood group trisaccharides by norovirus. *J Virol.* 2007 Jun;81(11):5949–57.
88. Crowe J, Masone BS, Ribbe J. One-step purification of recombinant proteins with the 6xHis tag and Ni-NTA resin. *Mol Biotechnol* [Internet]. 1995;4(3):247–58. Available from: <https://doi.org/10.1007/BF02779018>
89. Lindesmith LC, Donaldson EF, Baric RS. Norovirus GII.4 Strain Antigenic Variation. *J Virol.* 2011 Jan;85(1):231–42.

90. Tan M, Jiang X. Norovirus gastroenteritis, carbohydrate receptors, and animal models. *PLoS Pathog.* 2010 Aug;6(8):3–4.

ANEXOS

Anexo 1: Preparación de células competentes de *E. coli*

- Sembrar por agotamiento 1 colonia de la cepa deseada en una placa con LB-agar sin antibiótico.
- Transferir una colonia a 2 mL de medio LB sin antibiótico e incubar a 37°C con agitación durante toda la noche.
- Tomar 1 mL del cultivo e inocular 50 mL de medio LB sin antibiótico en un matraz de 250 mL e incubar a 37°C con agitación hasta un OD₆₀₀ de 0.35 a 0.45 (aproximadamente en 2 horas de incubación).
- Transferir el contenido a tubos de 50 mL enfriados previamente e incubar en hielo por 10 min. Centrifugar el cultivo a 4000 rpm y a 4°C por 10 min.
- Resuspender el pellet con 1mL de 0.1 M CaCl₂, 15% glicerol frío y homogenizar.
- Preparar alícuotas de fracciones de 50 µL en criotubos colocados en una gradilla con isopropanol previamente congelados a -80°C. Almacenar a -80°C.

Anexo 2. Método de extracción del ADN plasmídico por lisis alcalina

- Seleccionar una de las colonias obtenidas en las placas de agar LB-kanamicina, se transfirió a 5mL de caldo LB con kanamicina y se incubó durante toda la noche a 37°C y con agitación constante a 200 rpm.
- Después, transferir 1.5 mL del cultivo incubado en un tubo eppendorf, centrifugar la muestra a 10000 rpm durante 5 min y eliminar el sobrenadante.
- Añadir 150 µL de Solución I fría, colocar en un agitador vórtex de manera cuidadosa, evitando formar grumos.
- Añadir 300 µL de Solución II (preparada recientemente) y mezclar 5 veces por inversión y dejar incubar a temperatura ambiente por 5 minutos.
- Añadir 225 µL de Solución III fría, mezclar 5 veces por inversión y dejar incubar a 4°C durante 5 minutos. Luego, centrifugar a 13000 rpm por 5 minutos y transferir el sobrenadante en otro tubo eppendorf.
- Añadir 675 µL de solución de fenol/cloroformo/alcohol isoamílico 25:24:1, mezclando la solución con un agitador vórtex cuidadosamente y centrifugar a 12000 rpm durante 2 minutos.
- Transferir la fase acuosa a otro tubo y repetir el punto 6 una vez más.
- Añadir 2 volúmenes de etanol absoluto al sobrenadante, mezclar por inversión y dejar reposar por 20 minutos a 4°C.
- Centrifugar a 12000 rpm durante 5 minutos, remover el sobrenadante y colocar el tubo invertido sobre papel toalla hasta que el pellet se seque.
- Añadir 1mL de etanol 70% a 4°C. Repetir el paso 9.

- Resuspender el pellet en 50 μ L buffer EB conteniendo RNAsa y evaluar la cantidad y pureza del ADN extraído mediante espectrofotometría UV a 260 nm y 280 nm.

Anexo 3. Componentes de la reacción de digestión de ADN con las enzimas de restricción *NcoI* y *XhoI*

Reactivo	Control	Rxn <i>XhoI</i>	Rxn <i>NcoI</i>	Rxn Doble
H ₂ O (libre de nucleasas)	6 µL	5 µL	5 µL	5 µL
Buffer NE 3.1 10X	2 µL	2 µL	2 µL	2 µL
<i>XhoI</i> (20U/µL)	-	1 µL	-	0.5 µL
<i>NcoI</i> (10U/µL)	-	-	1 µL	0.5 µL
ADN plasmídico	12 µL	12 µL	12 µL	12 µL
Total	20 µL	20 µL	20 µL	20 µL

Anexo 4. Electroforesis SDS-PAGE

Para el análisis de las muestras de proteína, se empleó la metodología de SDS-PAGE descrita por Laemmli en 1970 con algunas modificaciones. Brevemente, 10 μ L de cada muestra fueron mezclados con un volumen igual de buffer de carga (Buffer Tris-HCl 0,3 M pH 6,8, SDS 10%, Azul de Bromofenol 0,05%, glicerol 50%, DTT 30%) y calentados por 5 minutos a 95°C. Se colocaron 10 μ L de muestra en los pocillos correspondientes y el tamaño de las bandas de proteínas fue estimado mediante el marcador de peso molecular (10 kDa – 250 kDa) Precision Plus Protein Dual Color (Bio-Rad, USA). La electroforesis se llevó a cabo en una cámara vertical Mini-PROTEAN® (Bio-Rad, EE.UU.) llena con buffer 1X Tris-glicina-SDS (Tris 25mM, glicina 192 mM y SDS 0,1%, pH 8,3) y empleando una corriente constante de 15 mA para el gel de apilamiento y 30 mA para el gel separador. Finalmente, los geles se tiñeron con azul de Coomassie 0,25% (0,25g de azul de Coomassie R-250 en H₂O/MeOH/acético glacial 45:45:10) *overnight* y posteriormente desteñidos con solución decolorante (H₂O/MeOH/acético glacial 45:45:10).

Anexo 5. Concentración de la proteína recombinante

Para concentrar la proteína purificada por cromatografía de afinidad en condiciones desnaturalizantes, se usó el sistema AMICON® Stirred Cell (Merck, EE.UU.). Este sistema fue conectado a gas nitrógeno, el cual ejerció la presión necesaria para lograr que la solución de proteína pase a través de una membrana con 10 kDa de tamaño de poro, lo cual evitó filtrar la proteína de 18 kDa. Este ensayo se realizó a temperatura de 4°C y en agitación. Inicialmente, el sistema AMICON® se desarmó y se lavó con agua hervida, la membrana de concentración también fue tratada de la misma manera. Seguidamente, se rearmó el sistema, y se colocaron las fracciones que contenían a la proteína purificada hasta alcanzar un volumen igual a $\frac{3}{4}$ de la capacidad total del sistema. Luego, se agregó 10 mL de Buffer A (buffer fosfato 20 mM, NaCl 500 mM, urea 8M, pH 8) como tampón de lavado, y se abrió el flujo de presión para iniciar la concentración. Cuando el volumen de solución en la parte superior de la membrana llegó a 10 mL, se agregaron más fracciones de proteína purificada hasta que el sistema nuevamente alcanzó un volumen de solución en la parte superior de la membrana equivalente a 10 mL. A continuación, se agregaron 35 mL de Buffer A al sistema en agitación hasta alcanzar un volumen de 4 mL dentro del sistema. Finalmente, estos 4 mL fueron recolectados en tubos de 1.5 mL y centrifugados por 1 minuto a 10 000 rpm para precipitar cualquier contaminante.

Anexo 6. Números de acceso al GenBank de las 716 secuencias de la proteína

VP1 de NV GII.4 empleadas en este estudio

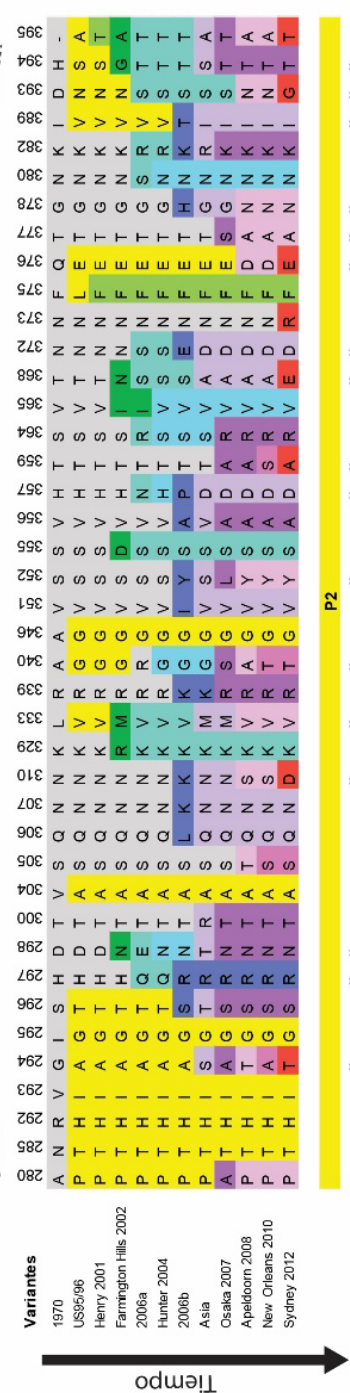
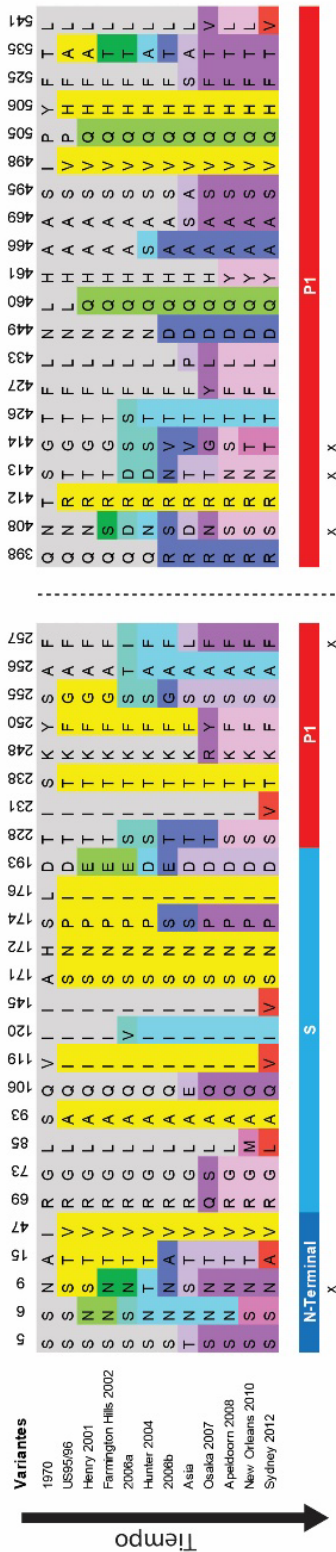
ABG49509.1, BAE98191.1, BAE98194.1, BAE98197.1, ABP88833.1,
ABV55634.1, ABW74532.1, ABW74533.1, ABW74534.1, ABW74535.1,
ABW74536.1, ABW74537.1, ABW74538.1, ABW74539.1, ABW74540.1,
ABW74541.1, ABW74542.1, ABW74543.1, ABW74544.1, ABW74545.1,
ABW74546.1, BAF95499.1, BAF95501.1, BAF95503.1, BAF95505.1,
BAF95509.1, BAF95511.1, BAF95513.1, BAF95515.1, BAF95517.1,
BAF95519.1, BAF95521.1, BAF95527.1, BAF95529.1, ABY27560.1,
ABY65350.1, P54635.1, BAG30923.1, ACF41177.1, ACF41180.1, ACF41181.1,
ACF41182.1, ACF41183.1, ACF41184.1, ACF41185.1, ACF41186.1,
ACF41187.1, ACF41188.1, ACF41189.1, ACF41190.1, ACF41191.1,
BAG70434.1, BAG70437.1, BAG70440.1, BAG70449.1, BAG70452.1,
BAG70455.1, BAG70458.1, BAG70461.1, BAG70476.1, BAG70479.1,
BAG70491.1, BAG70494.1, BAG70500.1, BAG70503.1, BAG70509.1,
BAG70512.1, BAG70521.1, BAG70524.1, BAG70527.1, ACJ66263.1,
ACJ66265.1, ACJ66267.1, ACJ66269.1, ACJ66271.1, ACJ66273.1, ACJ66275.1,
ACJ66279.1, ACJ66281.1, ACJ66283.1, ACJ66285.1, ACL27297.1, ACL27298.1,
ACL27299.1, ACL27300.1, ACL36375.1, BAH29786.1, BAD88558.2,
ACO55068.1, ACT76139.1, ACT76142.1, ACT76145.1, ACT76148.1,
ACT76151.1, ACV41093.1, ACW19927.1, ACX31821.1, ACX31885.1,
ACX31887.1, ACX31889.1, ACX31891.1, ACX85806.1, BAI49904.1,
BAI49905.1, BAI49906.1, BAI49907.1, BAI49908.1, BAI49909.1, BAI49910.1,
BAI49911.1, BAI49912.1, BAI49916.1, BAI49917.1, BAI49918.1, BAI49919.1,
BAI49920.1, BAI49921.1, BAI49924.1, BAI49925.1, BAI49926.1, BAI49927.1,
BAI49928.1, ADB27027.1, ADB27914.1, ADD10375.1, ADF47125.1,
ADI45808.1, ADI45809.1, ADI45810.1, ADI45811.1, ADI45813.1, ADI45815.1,
ADI45816.1, ADI45817.1, ADI45818.1, ACX31841.2, BAJ13508.1, BAJ13513.1,
BAJ13520.1, BAJ13534.1, BAJ13537.1, BAJ13540.1, BAJ13546.1, BAJ13555.1,
BAJ13564.1, BAJ13567.1, BAJ13570.1, BAJ13576.1, BAJ13588.1, BAJ13591.1,
BAJ13600.1, BAJ13603.1, BAJ13606.1, BAJ13609.1, BAJ13612.1, BAJ13615.1,
BAJ13618.1, BAJ13624.1, BAJ13627.1, BAJ13630.1, BAJ13633.1, BAJ13639.1,
BAJ13645.1, BAJ13651.1, BAJ13654.1, BAJ13657.1, BAJ13660.1, BAJ13663.1,
BAJ13666.1, BAJ13672.1, BAJ13675.1, BAJ13678.1, BAJ13681.1, BAJ13684.1,
BAJ13687.1, BAJ13690.1, BAJ13693.1, BAJ13696.1, BAJ13699.1, BAJ13702.1,
BAJ13708.1, BAJ13711.1, BAJ13717.1, BAJ13720.1, BAJ13723.1, BAJ13726.1,
BAJ13735.1, BAJ13741.1, BAJ13756.1, BAJ13759.1, BAJ13771.1, BAJ13777.1,
BAJ13789.1, BAJ13795.1, BAJ13801.1, BAJ13813.1, BAJ13825.1, BAJ13828.1,
BAJ13840.1, BAJ13852.1, BAJ13855.1, BAJ13861.1, BAJ13867.1, BAJ13873.1,
BAJ13879.1, BAJ13891.1, BAJ13894.1, BAJ13909.1, BAJ13915.1, BAJ13921.1,
BAJ13927.1, BAJ13930.1, BAJ13942.1, BAJ13948.1, BAJ13954.1, BAJ13957.1,
BAJ13972.1, BAJ13975.1, BAJ13978.1, BAJ13990.1, BAJ14005.1, BAJ14008.1,
BAJ14011.1, BAJ14017.1, ACL31316.2, ADM52743.1, ADQ43779.1,
ADQ43781.1, ADR78872.1, ADR78887.1, ADR78899.1, ADR78911.1,

ADR78917.1, ADR78920.1, ADR78923.1, ADR78929.1, ADR78932.1,
ADR78935.1, ADR78944.1, ADR78950.1, AEB71774.1, AEG79284.1,
BAK43272.1, BAK43278.1, BAK43281.1, AEJ82646.1, AEK05311.1,
AER11310.1, AER11312.1, AER11320.1, AER11326.1, AER11330.1,
AER11336.1, AER11338.1, AER11344.1, AER11348.1, AER11350.1,
AER11352.1, AER11354.1, AER11360.1, AEX91910.1, AEY77037.1,
ACX31831.2, ACX31836.2, ACX31851.2, ACX31855.2, ACX31856.2,
ACX31860.2, ACX31864.2, ABQ63283.2, AFI08234.1, AFI08237.1,
AFI08243.1, AFJ04708.1, AFJ04709.1, AFJ04707.1, AFJ21385.1, AFJ21387.1,
AFJ21390.1, AFJ21391.1, AFJ21392.1, AFJ21397.1, AFJ21399.1, AFJ21401.1,
AFJ21405.1, AFJ21408.1, AFJ21412.1, AFJ21418.1, AFJ21421.1, AFJ21423.1,
AFJ21425.1, AFJ21427.1, AFJ21428.1, AFJ21432.1, AFJ21433.1, AFJ21435.1,
AFJ21438.1, AFJ21440.1, AFJ21441.1, AFJ21442.1, AFJ21443.1, AFJ21444.1,
AFJ21446.1, AFJ21451.1, AFJ21454.1, AFJ24945.1, AFL70023.1, AFN09918.1,
AFN09924.1, AFN09930.1, AFN09939.1, AFN09945.1, AFN09954.1,
AFN09957.1, AFN09963.1, AFN09969.1, AFN09972.1, AFN09978.1,
AFN09981.1, AFN09984.1, AFN09993.1, AFP89593.1, AFQ00511.1,
AFQ00513.1, AFQ00514.1, AFQ00516.1, AFQ00519.1, AFQ00520.1,
AFQ00521.1, AFQ00526.1, AFQ00528.1, AFS33552.1, AFU55686.1,
AFU55689.1, AFU55692.1, AFU55698.1, AFU55707.1, AFU55713.1,
AFU55716.1, AFU55725.1, AFU55728.1, AFU55731.1, AFU92659.1,
AFU92660.1, AFU92661.1, AFU92662.1, AFU92663.1, AFU92664.1,
AFU92666.1, AFU92667.1, AFU92668.1, AFU92670.1, AFU92671.1,
AFU92672.1, AFU92673.1, AFU92674.1, AFU92675.1, AFU92676.1,
AFU92678.1, AFU92679.1, AFU92685.1, AFU92686.1, AFU92687.1,
AFU92688.1, AFU92693.1, AFU92699.1, AFU92700.1, AFU92703.1,
AFU92704.1, AFU92707.1, AFU92708.1, AFU92712.1, AFU92713.1,
AFU92715.1, AFU92716.1, AFU92717.1, AFU92718.1, AFU92721.1,
AFU92722.1, AFU92723.1, AFV08771.1, AFV08774.1, AFV08777.1,
AFV08780.1, AFV08783.1, AFV08789.1, AFV08792.1, AFV08795.1,
AFV99155.1, AFW15938.1, AFW15939.1, AFW15943.1, AFW15945.1,
AFW15948.1, AFW15950.1, AFW15951.1, AFX71668.1, AFX81124.1,
AFX81133.1, AFX81151.1, AFX81160.1, AFX81166.1, AFX81175.1,
AFX81184.1, AFX81187.1, AFX81217.1, AFX81220.1, AFX81223.1,
AFX81229.1, AFX81247.1, AFX81250.1, AFX81259.1, AFX81274.1,
AFX81280.1, AFX81286.1, AFX81301.1, AFX81310.1, AFZ84663.1,
AGC66783.1, AGC66784.1, AGC66787.1, AGC66788.1, AGC66789.1,
AGE89442.1, AGE89445.1, AGE89469.1, AGE89484.1, AGE89487.1,
AGE89508.1, AGE89538.1, AGE89562.1, AGE89565.1, AGE89586.1,
AGE89598.1, AGE89616.1, AGE89619.1, AGE89625.1, AGE89628.1,
AGE89643.1, AGE89652.1, AGE89670.1, AGE99596.1, AGE99612.1,
AGG37871.1, AGH24750.1, AGI96383.1, AGI99552.1, AGI99553.1,
AGI99554.1, AGI99555.1, AGJ52172.1, AGJ52173.1, AGJ52174.1,
AGK25882.1, AGK25884.1, AGK25887.1, AGK25890.1, AGK25895.1,
AGK25898.1, AGK25900.1, AGK25902.1, AGK25905.1, AGK25908.1,
AGK25910.1, AGK25912.1, AGK25915.1, AGK25917.1, AGK25919.1,
AGK25921.1, AGK25924.1, AGK25927.1, AGK25930.1, AGK36311.1,
AGL98413.1, AGL98416.1, AGN11943.1, CCX28607.1, CCX28610.1,

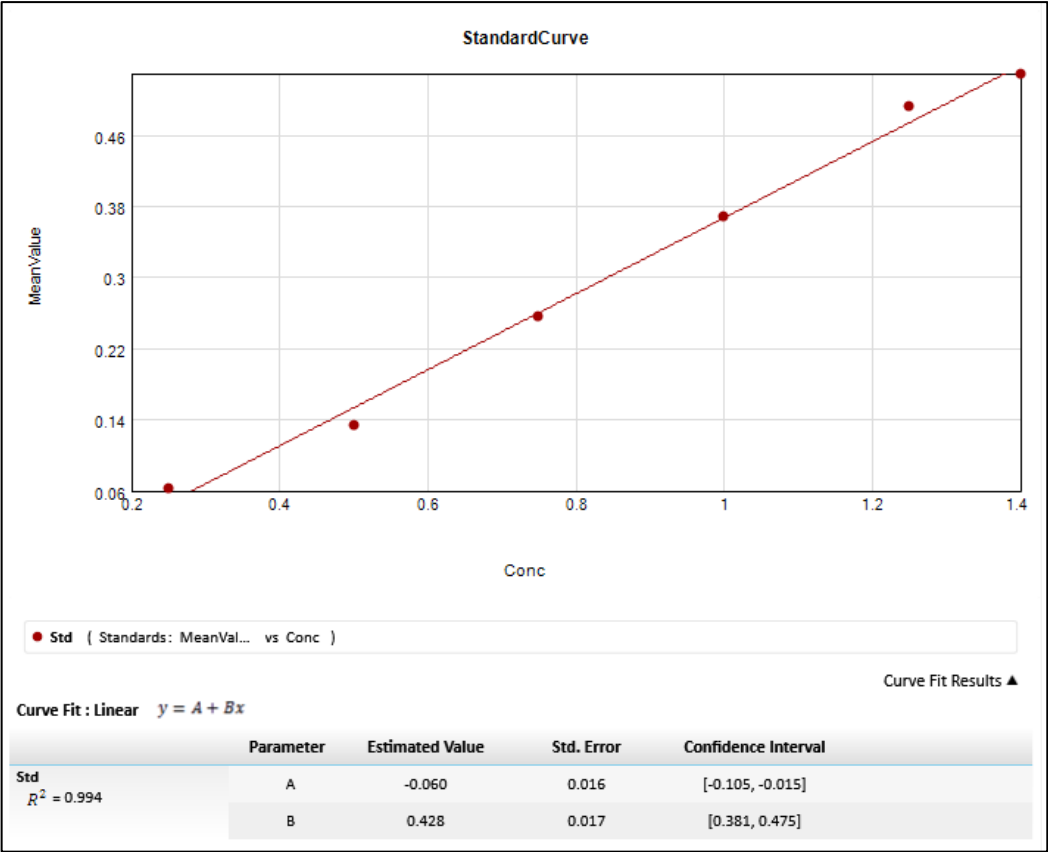
CCX28613.1, CCX28616.1, AGN90584.1, AGO20263.1, AGS08025.1,
AGS08027.1, AGS08028.1, AGS08034.1, AGS08035.1, AGS08038.1,
AGS08039.1, AGS08043.1, AGS08044.1, AGS08045.1, AGS08046.1,
AGS08047.1, AGS08049.1, AGS08051.1, AGS08052.1, AGS08053.1,
AGS08054.1, AGS08055.1, AGS08057.1, AGS08059.1, AGS08068.1,
AGS08070.1, AGS08075.1, AGS08077.1, AGS08078.1, AGS08081.1,
AGS08084.1, AGS08085.1, AGS08087.1, AGS08091.1, AGS08098.1,
AGS08100.1, AGS08101.1, AGS08108.1, AGS08109.1, AGS08110.1,
AGS08113.1, AGS08117.1, AGS08118.1, AGS08122.1, AGS08123.1,
AGS08124.1, AGS08125.1, AGS08130.1, AGS08131.1, AGS08132.1,
AGS08133.1, AGS08134.1, AGS08135.1, AGS08137.1, AGS08140.1,
AGS08141.1, AGS08142.1, AGS08144.1, AGS08149.1, AGS08151.1,
AGS08153.1, AGS08154.1, AGS08155.1, AGS08156.1, AGS08158.1,
AGS08159.1, AGS08167.1, AGS08169.1, AGS08171.1, AGS41450.1,
AGT17782.1, AGT17788.1, AGT17795.1, AGT17798.1, AGT17813.1,
AGT17816.1, AGT17819.1, AGT17823.1, AGT17830.1, AGT17833.1,
AGT17839.1, AGT17842.1, AGT95926.1, AGT95927.1, AGT95928.1,
AGT95929.1, AGT95930.1, AGT95931.1, AGU68328.1, AGV40221.1,
AGV76553.1, AGX01098.1, AGX85889.1, AGX85892.1, AGX85895.1,
AGX85901.1, AGX85904.1, AGX85910.1, AGX85934.1, AHA44847.1,
AHB62404.1, AHF45991.1, AHF45995.1, AHH44876.1, AHH44877.1,
AHH44878.1, AHH44879.1, AHH44880.1, AHH44881.1, AHH44884.1,
AHH44886.1, AHH44887.1, AHH44889.1, AHH44894.1, AHH44895.1,
AHH44900.1, AHH44904.1, AHH44906.1, AHH44907.1, AHH44908.1,
AHH44909.1, AHH44910.1, AHH44911.1, AHH44913.1, AHH44914.1,
AHH44915.1, AHI59157.1, AHN52598.1, AHX22018.1, AHX22021.1,
AHX22025.1, AHX22028.1, AHX22032.1, AHX22038.1, AHX22041.1,
AHX22044.1, AHX22047.1, AID51515.1, AID51519.1, AID51522.1,
AID68576.1, AID68577.1, AID68578.1, AID68579.1, AIG59994.1, AIG59996.1,
AIG60000.1, AIG60002.1, AIG60020.1, AII02584.1, AII02587.1, AII02680.1,
AII02725.1, AII02761.1, AII02794.1, AII02815.1, AII02821.1, AII02833.1,
AII02857.1, AII73750.1, AIS40019.1, AIV43150.1, AIV43151.1, AIV43156.1,
AIV43157.1, AIV43158.1, AIV43165.1, AIV43166.1, AIV43167.1, AIV43168.1,
AIV43169.1, AIV43170.1, AIV43173.1, AIV43179.1, AIY25491.1, AIY27747.1,
AIZ46843.1, AIZ46845.1, AIZ46846.1, AIZ46849.1, AIZ46850.1, AIZ46852.1,
AIZ46855.1, AIZ46858.1, AIZ46860.1, AIZ46863.1, AJA71388.1, AJA71390.1,
AJA71391.1, AJA71394.1, AJA71397.1, AJA71398.1, AJA71399.1, AJA71401.1,
AJA71405.1, AJA71407.1, AJA71412.1, AJA71414.1, AJA71415.1, AJA71417.1,
AJA71423.1, AJA71424.1, AJA71425.1, AJA71426.1, AJA71427.1, AJA71430.1,
AJA71456.1, AJA71457.1, AJA71458.1, AJA71459.1, AJA71466.1, AJA71467.1,
AJA71468.1, AJA71469.1, AJA71471.1, AJA71473.1, AJA71474.1, AJF21974.1,
AJF21975.1, AJF21977.1, AJF21978.1, AJF21979.1, AJF21980.1, AJF21983.1,
AJF21984.1, AJF39151.1, AJF39153.1, AJF39155.1, BAQ55211.1, BAQ55213.1,
BAQ55214.1, BAQ55215.1, BAQ55216.1, BAQ55217.1, BAQ55218.1,
ABC96332.1, ABD77588.1, ABD77589.1.

Anexo 7: Evaluación de sitios informativos en las secuencias de VP1 representativas de cada variante de NV GII.4

A continuación, se muestra el análisis evolutivo de las secuencias de la proteína VP1 de NV GII.4 representativas de las variantes que surgieron entre los años 1975 hasta 2015. Se observaron 12 patrones evolutivos en base a las sustituciones de residuos asociados con cada nueva variante. Gris, aminoácidos presentes en la variante 1970; amarillo, cambios que ocurrieron para generar la variante US95/96; verde claro, cambios específicos para la variante Henry 2001; verde oscuro, cambios asociados a la variante Farmington Hills 2002; turquesa, cambios asociados a la variante 2006a; celeste, cambios asociados a la variante Hunter 2004; azul, cambios asociados a la variante 2006b; púrpura, cambios asociados a la variante Asia; morado, cambios asociados a la variante Osaka 2007; rosado, cambios asociados a la variante Apeldoorn 2008; lila, cambios asociados a la variante New Orleans 2010; y rojo, cambios asociados a la variante Sidney 2012. El subdominio P2 se encuentra resaltado en color amarillo, cuyos flancos N-terminal y C-terminal están resaltados en color azul para el extremo N-terminal de la proteína total, color celeste para el dominio S, y color rojo para el subdominio P1. Los aminoácidos identificados bajo presión selectiva (denominados hipervariables) están marcados debajo de su columna por un aspa.



Anexo 8. Curva de calibración obtenida con el método de Bradford para cuantificación de proteínas



Anexo 9. Reactividad de sueros de bebés frente a la proteína multiepitópica MEP-NVGII.4 mediante ELISA indirecto

A continuación, se muestra la gráfica en barras de las lecturas de ELISA indirecto para la detección de anticuerpos IgG e IgM, mostrando los valores promedio de OD_{450nm} en cada muestra, así como la desviación estándar. El grupo R1-R6 representa aquellos sueros provenientes de bebés recién nacidos (controles negativos), y el grupo B1-B20, sueros obtenidos de bebés infectados con NV GII.4.

