

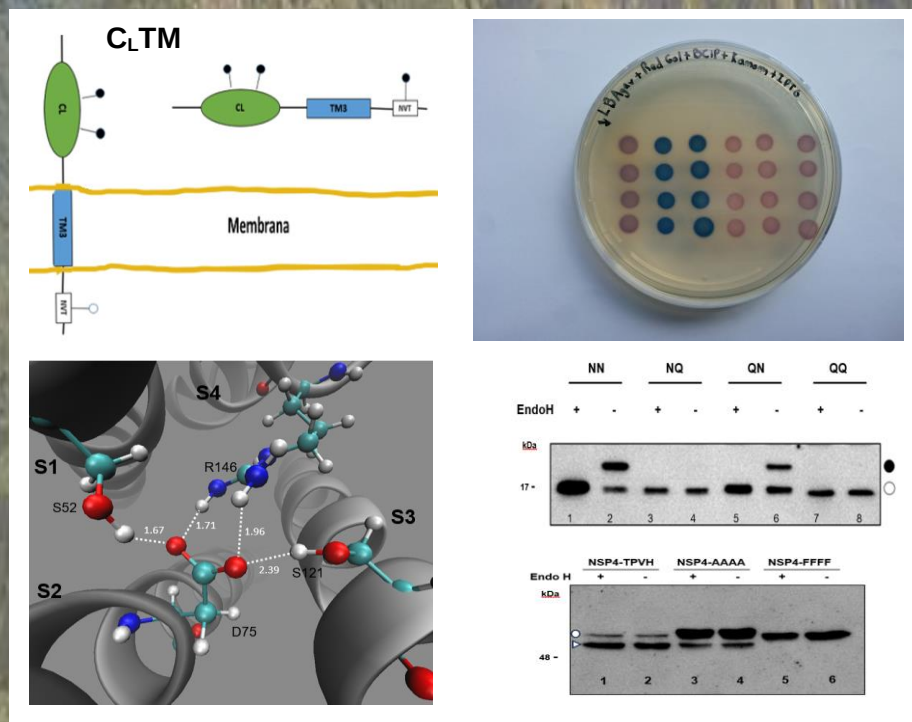


Universidad de Valencia

Biogénesis del Transportador NALCN y de la proteína no estructural 4 (nsp4) del SARS-CoV-2

Programa de Doctorado en Biomedicina y Biotecnología

José Ramiro Miguel Acosta Cáceres



Mayo 2024

Directores: Dr. Ismael Mingarro Muñoz

Dra. María Jesús García Murria



Universitat de València

Departamento de Bioquímica y Biología Molecular

Instituto Universitario de Biotecnología y Biomedicina (BIOTECMED)

José Ramiro Miguel Acosta Cáceres

TESIS DOCTORAL

Marzo 2024

Biogénesis del Transportador NALCN y de la
proteína no estructural 4 (nsp4) del SARS-CoV-2.

Directores: Dr. Ismael Mingarro

Dra. María Jesús García Murria

Memoria presentada para optar al título de Doctor

Programa de Doctorado en Biomedicina y Biotecnología

ISMAEL MINGARRO MUÑOZ, Doctor en Ciencias Biológicas y Catedrático del Departamento de Bioquímica y Biología Molecular de la Universitat de València.

MARÍA JESÚS GARCÍA MURRIA, Doctora en Bioquímica i Profesora Contractada Doctora del Departamento de Bioquímica y Biología Molecular de la Universitat de València.

MANIFIESTAN que, Don JOSÉ RAMIRO MIGUEL ACOSTA CÁCERES, Licenciado en Biología, Microbiología y Parasitología por la Universidad Nacional Mayor de San Marcos, ha realizado bajo nuestra dirección en el Laboratorio de Proteínas de Membrana del Instituto de Biotecnología y Biomedicina (BIOTECMED) y financiado con un Contrato del Programa “Beca Presidente de la República (PRONABEC 2018/2) del Ministerio de Educación de Perú, el trabajo con el título : “*Biogénesis del Transportador NALCN y de la proteína no estructural 4 del SARS-CoV-2*”, que presenta para optar al grado de Doctor por la Universitat de València. Parte de este trabajo se realizó en una estancia en el Instituto de Química Orgánica y Bioquímica de la Academia Checa de Ciencias de la República Checa.

Asimismo, aprovechamos para autorizar la presentación y defensa pública de este trabajo.

Y para que así conste y surta los efectos oportunos, firmamos el presente documento en Burjasot el 4 de marzo de 2024.

ISMAEL JOSE MINGARRO MUÑOZ -
NIF:29017960X
OX

Digitally signed
by ISMAEL JOSE
MINGARRO
MUÑOZ -
NIF:29017960X
Date: 2024.03.04
17:32:12 +01'00'

Dr. Ismael Mingarro

MARIA JESUS|
GARCIA|
MURRIA

Firmado
digitalmente por
MARIA JESUS|
GARCIA|MURRIA
Fecha: 2024.03.04
17:17:00 +01'00'

Dra. M^a Jesús García Murria

Gracias a todas las personas que contribuyeron al desarrollo y finalización de esta tesis, especialmente a mi madre Gloria Cáceres Cadenas, y abuelos José Cáceres y María Cadenas.

RESUMEN

Esta tesis se centra en investigar la topología y las etapas iniciales de la biogénesis del Dominio I (DI) del canal de sodio NALCN humano y la proteína no estructural NSP4 del SARS-CoV-2.

NALCN desempeña un papel clave en el sistema nervioso central, habiéndose vinculado algunas mutaciones a trastornos del neurodesarrollo, como el síndrome CLIFAHDD. Recientemente se ha resuelto la estructura de NALCN confirmando que contiene cuatro dominios estructurales (dominios I-IV), y cada uno de ellos está formado por seis segmentos transmembrana (TM) (S1-S6). Es importante destacar que un tercio de las mutaciones relacionadas con el síndrome CLIFAHDD ocurren en el DI, siendo éstas especialmente abundantes en su hélice TM S6. En esta tesis nos centramos en investigar la topología y las etapas iniciales de la biogénesis del DI de NALCN en membranas biológicas. Descubrimos que, aunque la primera región hidrofóbica direcciona la proteína a la membrana, hasta que no aparece el S3, la proteína no se inserta adecuadamente en ella. Por otro lado, exploramos experimentalmente y por simulaciones de dinámica molecular el impacto de las mutaciones CLIFAHDD descritas. Los resultados muestran que estas mutaciones no afectan a la inserción de S6 en la membrana. Sin embargo, observamos que algunas mutaciones modifican la flexibilidad y las interacciones entre los residuos de una región de S6 responsable de la apertura del canal, mientras que otras mutaciones provocan la reorganización de residuos en la membrana. Todos estos cambios a nivel estructural podrían a su vez generar alteraciones en la funcionalidad de NALCN.

La proteína no estructural 4 (NSP4) del SARS-CoV-2 juega un papel crucial en los reordenamientos de membrana y la formación de vesículas de doble membrana que son esenciales para la síntesis y replicación eficiente del RNA viral. Debido a las inconsistencias encontradas en las predicciones computacionales, en esta tesis hemos abordado la caracterización estructural de la proteína, con el objetivo de determinar su topología, el número de segmentos TM y las posibles modificaciones co- y/o post-traduccionales. Observamos que al expresar individualmente la proteína NSP4 en células humanas, ésta se integra en la membrana co-traduccionally como una proteína politópica y delimitamos, además, la posición de las regiones TM. Por otro lado, hemos descubierto que su primera región hidrofóbica es reconocida como secuencia señal “críptica” y procesada parcialmente por el complejo peptidasa (SPC) en diferentes líneas celulares. Y, por último, demostramos que la diana de glicosilación no canónica de la NSP4 no es modificada en células humanas. Globalmente, nuestro estudio proporciona información valiosa sobre las características estructurales de NSP4 del SARS-CoV-2 y puede abrir nuevas vías para reducir la capacidad infectiva del virus inhibiendo el procesamiento de la proteína.

Contenido

Lista de Figuras	9
Lista de tablas	11
ABREVIATURAS	12
NOMENCLATURA DE AMINOÁCIDOS	13
INTRODUCCIÓN	14
I.1. Membranas biológicas	15
I.2. Proteínas de las membranas biológicas	17
Clasificación de las proteínas integrales de membrana	18
Estructura de las proteínas integrales de membrana	19
Composición y distribución de aminoácidos en las IMP	21
I.3. Biogénesis de las proteínas helicoidales de membrana	22
Complejo Sec61/SecY	24
Complejo EMC	26
Complejo TMC01	28
Complejo PAT	29
Determinantes topológicos y de inserción	30
Puentes salinos	31
Puentes de hidrógeno	32
Interacciones hidrofóbicas	32
Motivos implicados en el empaquetamiento de hélices TM	33
I.4. Modificaciones cotraduccionales de las IMPs	33
I.5. Proteína NALCN (Sodium Leak Channel Non-selective)	35
Estructura de la proteína NALCN	36
Dominio sensor de voltaje (VSD) de NALCN	37
Dominio del poro de NALCN	38
Puerta intracelular de NALCN	39
Enfermedades asociadas a NALCN	41
NALCN y el síndrome de CLIFAHDD	41
I.6. Virus SARS-CoV-2	44
Genoma y proteoma del SARS-CoV-2	44
Ciclo infeccioso del SARS-CoV-2	47
Proteína no estructural 4 (NSP4) del virus SARS-CoV-2	49
OBJETIVOS	52
MATERIALES Y MÉTODOS	54

M.1. Análisis computacional de las secuencias proteicas	55
Predicción de hélices TM y topología.....	55
Simulaciones por dinámica molecular	55
M.2. Manipulación de bacterias.....	56
Cepas bacterianas:	56
Cultivo de cepas de <i>E. coli</i>	56
Transformación de las cepas de <i>E. coli</i> :	56
M.3. Manipulación de líneas celulares humanas	57
Líneas celulares humanas:.....	57
Mantenimiento de las líneas celulares humanas:.....	57
Transfección de líneas celulares humanas:	57
M.4. Manipulación genética.....	58
Vectores de DNA plasmídico:	58
Amplificación de material genético:.....	58
Inserción y clonación de material genético.....	59
Mutagénesis dirigida:	61
Secuenciación de material genético:.....	61
M.5. Ensayo topológico del dominio DI de NALCN y NSP4 en membranas de bacterias.....	61
M.6. Ensayo topológico del dominio DI de NALCN y NSP4 en membranas de células humanas.....	63
Obtención de extracto proteico de células humanas.....	63
Tratamiento con EndoH.....	63
Electroforesis SDS-PAGE y Western Blot	64
M.7. Procesamiento de la diana putativa de NSP4	64
M.8. Inhibición del procesamiento de NSP4 por cavinafungin	65
M.9. Expresión, purificación y cristalización de dominios solubles de la proteína NSP4.....	66
Expresión de dominios solubles	66
Purificación de dominios solubles	66
Solubilización de cuerpos de inclusión del dominio S40-P270 de NSP4.	67
Eliminación de la etiqueta de Histidinas del dominio citosólico NSP4 (T411-Q500)	67
Purificación del dominio citosólico NSP4 (T411-Q500) sin etiqueta	67
Cristalización del dominio citosólico	68
RESULTADOS Y DISCUSIÓN	69
I. Estudio topológico del dominio DI de NALCN	70
R1. Análisis <i>in silico</i> de la topología del dominio DI de NALCN	71

R.2. Inserción en membrana de los segmentos TM aislados	75
Diseño experimental del ensayo	75
Estudio de la inserción de los TMs aislados del dominio DI en membranas.....	76
R.3. Topología del dominio DI de NALCN en membranas procariotas	78
Diseño experimental del ensayo	78
Topología del dominio DI de NALCN en membranas de bacterias.....	79
R.4. Modificación de las dianas de glicosilación N210DT y N216VT	81
R.5. Topología del dominio DI de NALCN en membranas celulares humanas	82
Diseño experimental del ensayo	82
Biogénesis del dominio DI de NALCN en células humanas	83
R.6. Interacciones entre los segmentos del VSD	88
R.7. Posicionamiento en membrana del S6.....	97
R.8. Efecto de mutaciones CLIFAHDD en la inserción de S6 en membrana	100
R.9. Efecto de las mutaciones CLIFAHDD en la estructura del S6.....	101
Variación de contactos intra-hélice en el S6.....	101
Variación de los ángulos de torsión de los residuos W311, L309 y F308	103
Efecto de la mutación I322T en el S6 del DI de NALCN	109
D.1. Discusión del capítulo I	110
II. Estudio topológico de la proteína NSP4	116
R.10. Análisis <i>in silico</i> de la topología de NSP4.	117
R.11. Topología de NSP4 en membranas de bacterias.....	119
Diseño experimental del ensayo	119
Topología de NSP4 en membranas de bacterias	119
R.12. Inserción en membranas de la primera región hidrofóbica de NSP4	121
Diseño experimental del ensayo	121
Estudio de la inserción de la primera región hidrofóbica de NSP4	122
R.13. Estudio del procesamiento co-/post-traducciona l de NSP4.....	124
R13A. Glicosilación de NSP4.....	124
R13B. Caracterización del procesamiento de NSP4	125
Procesamiento de la proteína NSP4 en otras células.....	127
R.14. Topología de NSP4 en membranas celulares humanas	128
Topología de NSP4 en células eucarióticas.....	128
R.15. Purificación de los dominios Solubles de la proteína NSP4 del SARS-CoV-2.....	133
Dominio soluble luminal de la proteína NSP4.....	133
Expresión y purificación del dominio soluble citosólico de la proteína NSP4.....	135

R.16. Cristalización del dominio citosólico de la proteína NSP4	139
D.2. Discusión del capítulo II	140
CONCLUSIONES	145
BIBLIOGRAFÍA	147
ANEXOS	162
ANEXO 1: Oligonucleótidos utilizados en la presente tesis.....	163
ANEXO 2: Construcciones realizadas en la Tesis.....	167
ANEXO 3: Metodología para determinar la topología del DI NALCN y NSP4	185
ANEXO 4: Comparación de secuencias.....	187

Lista de Figuras

Figura 1. Membranas biológicas.....	15
Figura 2. Proteínas de membrana.....	16
Figura 3. Proteínas integrales de las membranas biológicas.....	18
Figura 4. Clasificación de las proteínas integrales de membrana.....	18
Figura 5. Prevalencia de aminoácidos en proteínas integrales de membrana.....	22
Figura 6. Direccionamiento co-traducciona l de proteínas de membrana y de secreción.....	23
Figura 7. Topología y estructura de Sec61 α , componente central del traslocón.....	25
Figura 8. Estructura de EMC y sus homólogos funcionales YidC y GET.....	27
Figura 9. Inserción de proteínas de membrana politópicas mediada por EMC.....	28
Figura 10. Translocón TMCO1.....	29
Figura 11. Complejo PAT.....	30
Figura 12. El complejo señal peptidasa (SPC).....	34
Figura 13. Esquema de la maquinaria de glicosilación (OST) cotraducciona l.....	35
Figura 14. Estructura de la proteína NALCN.....	38
Figura 15. Estructura del filtro selectivo de NALCN.....	39
Figura 16. Estado abierto y cerrado de NALCN y del canal transitorio selectivo de la subfamilia vanilloide (TRPV 6).....	40
Figura 17. Mutaciones en el DI de NALCN estudiadas en la presente tesis.....	42
Figura 18. Estructura del genoma de SARS-CoV-2.....	47
Figura 19. Ciclo infe ctivo del SARS-CoV-2.....	49
Figura 20. Morfología de las vesículas de doble membrana (DMVs).....	51
Figura 21. Predicción del ΔG predictor.....	72
Figura 22. Predicción del TMHMM.....	73
Figura 23. Análisis de las modificaciones cotraducciona les.....	74
Figura 24. Sistema C _L TM-3G.....	76
Figura 25. Inserción en membrana de los 6 segmentos TM del DI.....	77
Figura 26. Variantes fusionadas de truncados del DI /phoA-lacZ.....	78
Figura 27. Topología del dominio DI de NALCN en membranas bacterianas.....	80
Figura 28. Las dianas de glicosilación N210DT y N216VT son modificadas en células eucariotas.....	82
Figura 29. Diseño de las dianas de glicosilación reporteras.....	83
Figura 30. El truncado 65-mer se inserta poco eficientemente en la membrana con topología Nt citoplasmática.....	84
Figura 31. Topología de la proteína truncada 97-mer.....	85
Figura 32. Topología de la proteína truncada 136-mer.....	86
Figura 33. Topología de la proteína truncada 163-mer.....	86
Figura 34. Topología de la proteína truncada 233-mer.....	87
Figura 35. Topología de la proteína truncada 380-mer.....	88
Figura 36. Interacciones entre S1 y S2.....	89
Figura 37. Interacciones entre S1 y S4.....	91
Figura 38. Interacciones entre S2 y S3.....	92
Figura 39. Interacciones entre S2 y S4.....	94
Figura 40. Interacciones entre S3 y S4.....	95
Figura 41. Principales interacciones en el dominio VSD.....	96
Figura 42. Puente salino entre D75 y R146.....	96
Figura 43. Posicionamiento del S6 del DI en las membranas del RE.....	99
Figura 44. Efecto de mutaciones CLIFAHDD en la inserción de S6 en membranas de RE.....	101
Figura 45. Variación de contactos intra-hélice originados por las mutaciones L312V, L312I y V313G.....	102
Figura 46. Efecto de las mutaciones L312V, L312I y V313G en los ángulos de torsión phi (ϕ) y psi (Ψ) del residuo W311.....	104

Figura 47. Efecto de las mutaciones L312V, L312I y V313G en los ángulos phi (ϕ) y psi (Ψ) del residuo L309.....	105
Figura 48. Efecto de las mutaciones L312V, L312I y V313G en los ángulos phi (ϕ) y psi (Ψ) del residuo F308.....	106
Figura 49. Efecto de las mutaciones L312V, L312I y V313G en el acercamiento de los residuos 308-313 y 309-314.....	107
Figura 50. Análisis de las distancias entre residuos 308-313 y 309-314 de la secuencia WT y de las secuencias mutantes (L312V, L312I y V313G).....	108
Figura 51. Cambio de posición del residuo 322 en membrana.....	109
Figura 52. Principales características estructurales de NALCN estudiadas.....	115
Figura 53. Análisis <i>in silico</i> de la proteína NSP4 de SARS-CoV-2.....	118
Figura 54. Variantes fusionadas de los truncados NSP4/pho-lac.....	119
Figura 55. Topología de NSP4 en membranas bacterianas.....	120
Figura 56. Ensayo C _L TM-2G.....	122
Figura 57. Inserción en membrana del segmento TM ₁₀₋₃₂ de NSP4 de SARS-CoV-2.....	123
Figura 58. La diana de glicosilación no canónica NIC no es modificada en células eucariotas.....	125
Figura 59. La primera región hidrofóbica de NSP4 es procesada como péptido señal.....	126
Figura 60. NSP4 es procesada por el complejo SPC.....	127
Figura 61. Procesamiento de la proteína NSP4 en diferentes líneas celulares.....	128
Figura 62. Topología de proteínas NSP4 truncadas 279-mer, 313-mer, 335-mer y 346-mer.....	129
Figura 63. Procesamiento del truncado 346-mer por el SPC.....	131
Figura 64. Topología de los truncados 362-mer, 371-mer, 397-mer y de NSP4 completa (500-mer).....	132
Figura 65. Expresión a pequeña escala del dominio luminal de NSP4 (S40-P270).....	133
Figura 66. Cromatografía de afinidad HisTrapHP del dominio luminal de NSP4 (HisTag+NSP4(S40-P270)).....	134
Figura 67. Solubilización de los cuerpos de inclusión formados por el dominio S40-P270 de NSP4.....	135
Figura 68. Expresión de la proteína HisTag+NSP4(T411-Q500).....	136
Figura 69. Purificación por cromatografía de afinidad del dominio T411-Q500 de NSP4.....	136
Figura 70. Purificación por cromatografía en columna HisTrap ^{HP} del producto de la digestión de HisTag+NSP4(T411-Q500).....	137
Figura 71. Purificación por exclusión molecular del fragmento T411-Q500 de NSP4.....	137
Figura 72. Método alternativo de purificación del fragmento T411-Q500 de NSP4.....	138
Figura 73. Disposición de la proteína NSP4 del SARS-CoV-2 en membrana.....	143
Figura 74. Gráfico resumido del ensayo cualitativo del reportero dual PhoA-LacZ.....	185
Figura 75. Gráfico resumido del ensayo cuantitativo del reportero dual PhoA-LacZ.....	185
Figura 76. Gráfico resumido del ensayo de mapeo de glicosilaciones en líneas celulares humanas.....	186
Figura 77. Alineamiento de las proteínas NALCN, Nav1.2 y Cav1.1.....	188
Figura 78. Alineación de segmentos S6 de NALCN y TRPV6.....	189
Figura 79. Alineamiento de la región N-terminal de las proteínas NSP4 y 2K-NS4B.....	190
Figura 80. Alineamiento de las proteínas nsp4 de cinco coronavirus.....	192

Lista de tablas

Tabla 1. Composición lipídica de membranas celulares.....	17
Tabla 2. Parámetros geométricos de las hélices α y π	20
Tabla 3. Lista de mutaciones en la proteína NALCN que generan un espectro fenotípico CLIFAHDD.....	43
Tabla 4. Programas computacionales de predicción de segmentos TM, topología y péptido señal (SP) disponibles en la web.....	55
Tabla 5. Programas computacionales de diseño, simulación, análisis y procesamiento de trayectorias y datos utilizados en dinámica molecular.....	55
Tabla 6. Lista de Vectores plasmídicos utilizados.....	58
Tabla 7. Predicción del ΔG Predictor de segmentos TM del DI de NALCN con sus respectivas posiciones, longitudes, valor ΔG_{app} y secuencia de aminoácidos.....	72
Tabla 8. Predicción del algoritmo TMHMM de los segmentos TM y la topología del DI de NALCN con sus respectivas posiciones y localizaciones (citosol, lumen o hélice TM).....	73
Tabla 9. Residuos incluidos en las construcciones de los 6 segmentos TM del dominio DI.....	76
Tabla 10. Valores cuantitativos de la actividad relativa normalizada (NAR) de los seis polipéptidos truncados de NALCN fusionados a PhoA-LacZ α	81
Tabla 11. Estadística descriptiva de las distancias de interacción entre S1 y S2.....	90
Tabla 12. Estadística descriptiva de las distancias de interacción entre S1 y S4.....	91
Tabla 13. Estadística descriptiva de las distancias de interacción entre S2 y S3.....	93
Tabla 14. Estadística descriptiva de las distancias de interacción entre S2 y S4.....	94
Tabla 15. Estadística descriptiva de las distancias de interacción entre S3 y S4.....	95
Tabla 16. Secuencias diseñadas para el estudio de posicionamiento de S6 en membranas del RE.....	98
Tabla 17. Estadística descriptiva de las distancias entre los residuos 308-313 y 309-314.....	108
Tabla 18. Valores cuantitativos de la actividad relativa normalizada (NAR) de las diez proteínas truncadas de NSP4 fusionadas a PhoA-LacZ α	121
Tabla 19. Oligonucleótidos usados para la construcción del sistema CLTM en el vector pSVLd.....	163
Tabla 20. Oligonucleótidos utilizados para clonar los 6 TM del DI de NALCN individualmente en el sistema C _L TM-3G.....	163
Tabla 21. Oligonucleótidos usados para clonar el segmento TM ₁₀₋₃₂ de NSP4 en C _L TM-2G.....	163
Tabla 22. Oligonucleótidos utilizados para clonar los polipéptidos truncados del DI de NALCN en el vector de bacterias pKTop.....	164
Tabla 23. Oligonucleótidos utilizados para clonar los polipéptidos truncados de NSP4 en el vector de bacterias pKTop.....	164
Tabla 24. Oligonucleótidos usados para modificar la diana de glicosilación no canónica de NSP4.....	164
Tabla 25. Oligonucleótidos usados para clonar los polipéptidos truncados del dominio DI de NALCN y de NSP4 en el vector de eucariotas pCAGGS.....	165
Tabla 26. Oligonucleótidos usados para modificar el sitio de corte de la peptidasa SPC de NSP4.....	166
Tabla 27. Oligonucleótidos usados para clonar el segmento S6 del DI de NALCN en sistema Lep, diseñar las mutaciones CLIFAHDD en S6 y reemplazar residuos específicos en S4 del DI de NALCN.....	166
Tabla 28. Oligonucleótidos para clonar las regiones solubles de NSP4 en el vector pETNKL 1.1.....	166
Tabla 29. Datos de las proteínas humanas NALCN, Nav1.2 y Cav1.1.....	187
Tabla 30. Segmentos TM de las proteínas NALCN, Nav1.2 y Cav1.1.....	189
Tabla 31. Información de las proteínas NSP4 y 2K-NS4B.....	190
Tabla 32. Datos de la proteína nsp4 de 5 coronavirus.....	190
Tabla 33. Segmentos TM de NSP4 de cinco coronavirus.....	193

ABREVIATURAS

CHOL/ch: colesterol.
CL: cardiolipina.
CLIFAHDD: contracturas congénitas de extremidades y cara, hipotonía y retraso del desarrollo.
C_LTM: sistema transmembrana de la cadena ligera de la inmunoglobulina
Ct: extremo C-terminal.
DGlcDAG: di-glucosil-diacil-glicerol.
ECL: loop o bucle extracelular.
EMD: dominio extramembranoso.
ERG: ergosterol.
FS: filtro selectivo del canal iónico
GPCR: receptor acoplado a proteína G.
Hek293T: células embrionarias de riñón humano 293.
IHPRF: hipotonía infantil con retraso psicomotor y características faciales.
IMP: proteína integral de membrana
lacZα: gen codifica el fragmento α de la β galactosidasa.
LB: Medio Luria-Bertani.
LPG: lisil-fosfatidilglicerol.
MD: Dinámica molecular
ME: membrana externa bacterias.
MI: membrana interna bacterias.
Nt: extremo N-terminal.
NALCN: canal no selectivo de flujo de sodio.
NAR: actividad relativa normalizada.
OMIM: Herencia mendeliana en línea del hombre.
ONPG: orto-nitrofenil-β-D-galactósido.
ORF: marco de lectura abierto.
OST: Oligosacaril transferasa
PC: fosfatidilcolina.
PCR: reacción en cadena de la polimerasa.
PD: dominio del poro del canal iónico
PDB: Banco de datos de proteínas.
PE: fosfatidiletanolamina.
PEI: Polietilenimina.
PG: fosfatidilglicerol.
phoA: gen codifica fosfatasa alcalina.
PI: fosfatidilinositol.
PL: fosfolípidos.
pNPP: p-nitrofenil fosfato.
POPC: 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine.

PP: proteína periférica.
PS: fosfatidilserina.
RE: Retículo endoplásmico.
Red Gal: 6-cloro-3-indolil-β-D-galactósido.
RER: retículo endoplasmático rugoso.
RNAg: ARN genómico.
RNAm: ácido ribonucleico mensajero.
RNAr: ácido ribonucleico ribosomal.
RNAsg: ARN sub-genómico.
RNC: Complejo cadena naciente-ribosoma
SARS-CoV-2: Coronavirus 2 del Síndrome Respiratorio Agudo Severo.
SM: esfingomielina.
SNP: sitios de polimorfismo de un solo nucleótido.
SP: Péptido señal
SR: SRP receptor
SRP: Partícula de reconocimiento de señal.
TM: transmembrana.
TRAM: proteína asociada a la cadena de translocación.
TRAP: Proteína asociada al translocon
DMV: vesículas de doble membrana.
VSD: dominio sensor de voltaje.
X-Pho: 5-bromo-4-cloro-3-indolil-fosfato.

NOMENCLATURA DE AMINOÁCIDOS

Ácido aspártico - Asp - D

Ácido glutámico - Glu - E

Alanina - Ala - A

Arginina - Arg - R

Asparragina - Asn - N

Cisteina - Cys - C

Fenilalanina - Phe – F

Glicina - Gly - G

Glutamina - Gln - Q

Histidina - His - H

Isoleucina - Iso - I

Lisina - Lys – K

Metionina - Met - M

Prolina - Pro - P

Serina - Ser - S

Tirosina - Tyr - Y

Treonina - Thr - T

Triptofano - Trp - W

Valina - Val - V

X: cualquier aminoácido.

INTRODUCCIÓN

I.1. Membranas biológicas

Las membranas biológicas están constituidas principalmente por lípidos y proteínas organizados estructuralmente, de forma simplificada, según el modelo del mosaico fluido (Singer y Nicolson, 1972). Esta estructura lipoproteica rodea a las células de todos los organismos vivos, procariotas, arqueas y eucariotas, representando una barrera física que las protege de moléculas extrañas y cumple diferentes funciones como la regulación del transporte de iones, la endocitosis mediada por receptores, la síntesis de componentes de la membrana, la secreción de sustancias tóxicas, la síntesis de ATP, el transporte de electrones, la transducción de señales, diversas reacciones enzimáticas, comunicación intercelular, señalización celular, entre otras (Jelokhani-Niaraki et al., 2023). El componente lipídico de la membrana está compuesto por dos monocapas de lípidos (bicapa lipídica) en la que pueden diferenciarse dos regiones, un núcleo hidrofóbico y dos interfases polares (Figura 1), cuyas dimensiones fueron calculadas usando un modelo lipídico de dioleoilfosfatidilcolina (DOPC) (Wiener y White, 1992).

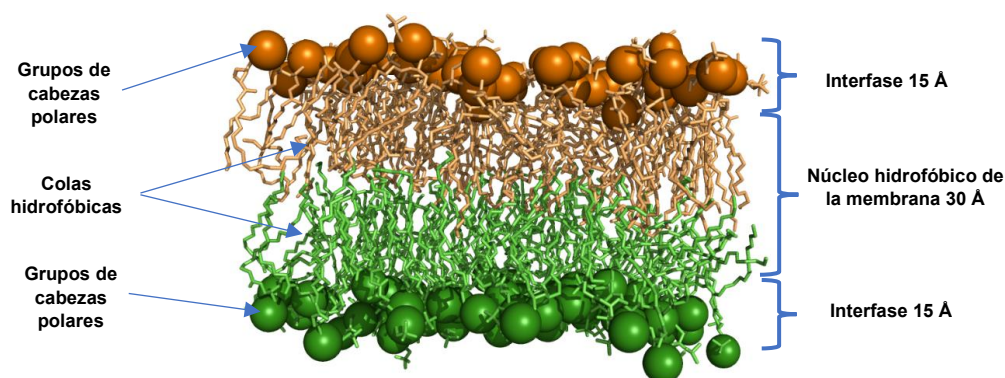


Figura 1. Membranas biológicas. Los fosfolípidos con sus grupos de cabezas polares y colas hidrofóbicas son los componentes más abundantes en la membrana. La longitud de las colas hidrocarbonadas determina el grosor del núcleo hidrofóbico de la membrana que normalmente es de unos 30 Å y el tamaño de los grupos de las cabezas polares a cada lado del núcleo es de alrededor de 15 Å, dándole a la bicapa (anaranjado y verde) un grosor aproximado de 60 Å.

El componente proteico de la membrana está compuesto principalmente por proteínas integrales (atraviesan el núcleo hidrofóbico) y periféricas (no atraviesan el núcleo hidrofóbico) (Figura 2). Estas últimas permanecen en la superficie de la membrana interactuando con los grupos de las cabezas polares de los lípidos o adheridas a otras proteínas de membrana. Por su parte, las proteínas integrales de membrana (IMP), ajustan la longitud de sus dominios transmembrana (TM) con el grosor de la bicapa. Un desajuste entre ambos provoca el desplazamiento lateral de la proteína, la agregación de proteínas, la inclinación y/o el empaquetamiento entre segmentos TM o el ajuste de

la bicapa mediante deformación lipídica (Grau et al., 2017; Harayama y Riezman, 2018). Asimismo, se ha comprobado que la presión lateral de la membrana (por los grupos de sus cabezas polares y de las cadenas hidrofóbicas) ejerce diversas fuerzas sobre las IMP, habiéndose sugerido que afecta a su función, localización y conformación (Lee, 2011).

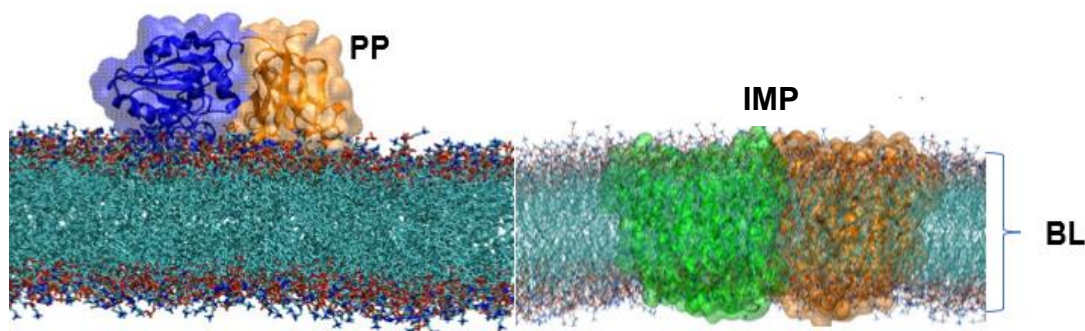


Figura 2. Proteínas de membrana. En la bicapa lipídica (BL) encontramos proteínas periféricas de membrana (PP), las cuales no atraviesan el núcleo hidrofóbico; y proteínas integrales de membrana (IMPs) que atraviesan el núcleo hidrofóbico.

Los fosfolípidos de membrana son moléculas anfipáticas con grupos de cabeza hidrofílica (que contienen fosfatos) y colas hidrocarbonadas hidrofóbicas (ácidos grasos). En un medio acuoso se organizan espontáneamente en una bicapa con los grupos polares orientados hacia el medio acuoso y las colas hidrofóbicas de cada monocapa asociándose entre ellas, evitando el contacto con el agua. Esto se debe a que las cadenas hidrocarbonadas de los ácidos grasos no pueden formar puentes de hidrógeno con el agua; sino que establecen entre ellas interacciones de van der Waals (Casares et al. 2019). La estructura, abundancia y tipo de lípidos de membrana dependen del organismo, del orgánulo e incluso de la monocapa (Sanders et al. 2018), y sus propiedades electrostáticas y mecánicas están determinadas por dos factores. El primero de estos factores es el tamaño y la carga del grupo que constituye la cabeza polar de los lípidos que puede estar cargado y ser catiónico (O-lisil-fosfatidilglicerol) o aniónico (fosfatidilglicerol (PG), cardiolipina (CL), fosfatidilserina (PS), fosfatidilinositol (PI)); o neutro (diglucosildiacilglicerol (DAG)) o zwitteriónico (fosfatidilcolina (PC) y fosfatidiletanolamina (PE)). El segundo factor es, la longitud y grado de insaturación de las cadenas hidrocarbonadas de sus ácidos grasos. En la membrana plasmática de bacterias, tanto Gram negativas como Gram positivas, predominan PE, PG y CL (Raetz y Dowhan 1990), mientras que las membranas plasmáticas de células de mamíferos son bicapas lipídicas que contienen cuatro fosfolípidos mayoritarios: PC, PE, PS y esfingomielina (SM); los cuales se distribuyen asimétricamente entre las dos

monocapas, siendo los lípidos aniónicos más abundantes en la monocapa citoplasmática y generando una carga neta negativa en la cara citosólica. Las células eucariotas presentan un rico sistema de endomembranas en el que destaca el retículo endoplásmico (RE). El RE rugoso tiene sus membranas cubiertas por ribosomas unidos al complejo multiproteico denominado translocón, a través del cual se realiza la inserción o translocación de proteínas integrales de membrana y de secreción, respectivamente. En la membrana del RE predomina principalmente la PC (Tabla 1) (Casares et al., 2019; Van Meer et al., 2008).

Tabla 1. Composición lipídica de membranas celulares. Fosfatidiletanolamina (PE), fosfatidilglicerol (PG), fosfatidilcolina (PC), fosfatidilserina (PS), fosfatidilinositol (PI), cardiolipina (CL), colesterol (Ch), esfingomielina (SM). Porcentajes del total de lípidos en la membrana (Harayama y Riezman, 2018). Cuando la composición de lípidos no sea igual al 100%, el resto de lípidos menos frecuentes se han omitido.

Membrana	PE	PG	PC	PS	PI	CL	Ch	SM
Membrana plasmática de <i>E. coli</i>	70-75	20-25	0	0	0	5-10	0	0
Membrana plasmática de eucariotas	11	0	23	8	0	0	34	17
RE	20	0	54	0	11	0	8	0

I.2. Proteínas de las membranas biológicas

Además de los lípidos encontramos como componente principal de las membranas biológicas a las proteínas de membrana cuya importancia radica en su diversidad de funciones y en su abundancia celular. Se estima que el 20-30% de los genes de los organismos secuenciados codifican proteínas de membrana (Krogh et al., 2001; Wallin y von Heijne, 1998). Además, desde el punto de vista farmacéutico también son muy importantes, ya que más de la mitad de las dianas farmacológicas de nuestro arsenal terapéutico son proteínas de membrana (Hopkins y Groom, 2002; Russell y Eggleston, 2000). Como mencionamos anteriormente, las proteínas de membrana se pueden integrar en la estructura de la bicapa lipídica o simplemente se pueden asociar con un componente lipídico o proteico a su superficie. Por su parte, las proteínas integrales de membrana pueden subdividirse en bitópicas y politópicas, dependiendo del número de veces que atraviesan el núcleo hidrofóbico de la membrana. Se denominan bitópicas aquellas que presentan una sola región transmembrana (TM), y politópicas a aquellas que presentan dos o más segmentos TM (Figura 3), siendo más numerosas estas últimas (Allen et al., 2019; Overduin et al., 2023).

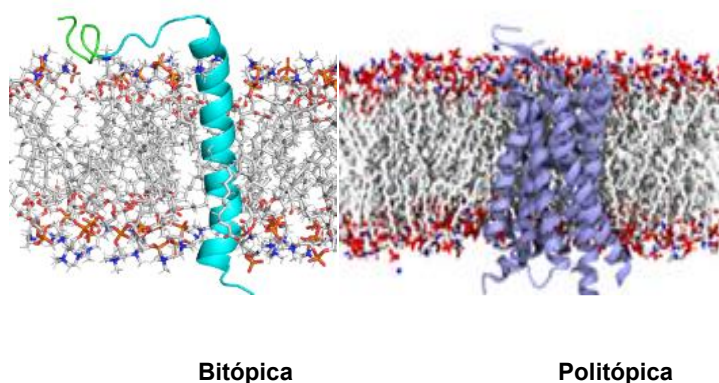


Figura 3. Proteínas integrales de las membranas biológicas. Las IMPs se subdividen en bitópicas (celeste) que atraviesan el núcleo hidrofóbico una sola vez y politópicas (morado) que atraviesan el núcleo hidrofóbico más de una vez.

Clasificación de las proteínas integrales de membrana

Las proteínas integrales de membrana (IMP) tanto bitópicas como politópicas se pueden clasificar en proteínas de tipo I (tienen péptido señal escindible en el extremo N-terminal), tipo II (con N- terminal orientado al citosol) o tipo III (sin péptido señal escindible y con su N-terminal corto extracitoplasmático) (Figura 4). Además, las proteínas bitópicas que tienen su segmento TM próximo al extremo C-terminal se conocen como TA (*tail anchoring*) (C-terminal corto extracitoplasmático) (Figura 4) (Hegde y Keenan, 2022).

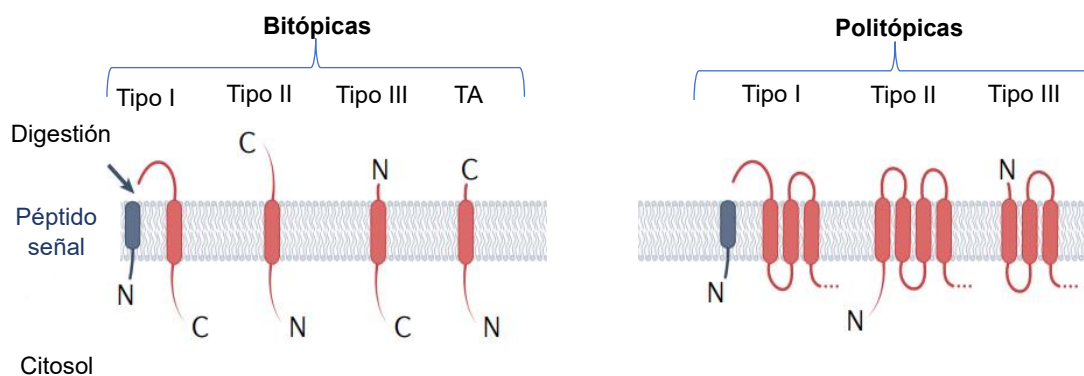


Figura 4. Clasificación de las proteínas integrales de membrana. Aproximadamente el 46% de las proteínas integrales de membrana son bitópicas y el 54% politópicas en humanos. En el esquema se muestra los segmentos TM (rojo) y el péptido señal (azul). N y C indican los extremos N-terminal y C-terminal, respectivamente. Adaptado de Hegde y Keenan, 2022.

Estructura de las proteínas integrales de membrana

Las proteínas están constituidas por aminoácidos unidos por enlaces peptídicos entre sus grupos α -carboxilo y α -amino. Este enlace covalente tiene un carácter parcial de doble enlace que restringe su rotación y le confiere planaridad (Pauling, 1960), permitiéndole únicamente libertad de rotación alrededor de los enlaces N-C α y C α -C β , lo que da lugar a ángulos de torsión o diédricos phi (ϕ) y psi (Ψ) respectivamente, que permiten a la cadena polipeptídica adoptar diferentes conformaciones. Los ángulos de torsión en proteínas (ϕ y Ψ) pueden tener cualquier valor entre -180° y $+180^\circ$. Sin embargo, ciertos valores de ϕ y Ψ están prohibidos porque podrían provocar choques estéricos entre los átomos presentes en el esqueleto polipeptídico, así como entre las cadenas laterales de los aminoácidos (van der Wel, 2021). Por otra parte, el enlace peptídico se caracteriza por adoptar la conformación *trans* en la que los carbonos- α (C α) de los dos aminoácidos que forman el enlace peptídico apuntan en direcciones opuestas (Golden et al., 2017; Zacharias et al., 2013).

Debido a la polaridad del enlace peptídico, el esqueleto polipeptídico de las proteínas de membrana debe adoptar en las regiones embebidas en la membrana conformaciones o estructuras secundarias (locales) estabilizadas por interacciones por puentes de hidrógeno que minimicen su polaridad y que les permitan atravesar de forma termodinámicamente estable el núcleo hidrofóbico de la membrana. De las estructuras secundarias que se han descrito en proteínas de membrana, destacan principalmente las hélices α y la hoja β plegada, si bien también se pueden encontrar hélices π y 3_{10} , aunque con mucha menos frecuencia. Así pues, las proteínas de membrana se pueden clasificar según la estructura en su región TM como helicoidales o como barriles beta, estas segundas sólo se han observado en membrana externa procariótica, mitocondrial y cloroplástica, siendo las proteínas de membrana helicoidales las absolutamente mayoritarias y las que son objeto de estudio en la presente tesis, por lo que se describen brevemente a continuación.

La conformación en hélice α es la más abundante en proteínas en general y también en las proteínas de membrana. Estas hélices tienen 3,6 residuos por vuelta y 13 átomos en una vuelta del esqueleto polipeptídico, por lo que se denominan hélices $3,6_{13}$ (Tabla 2), se estabilizan por puentes de hidrógeno entre los grupos $\text{NH}_{(i+4)}-\text{C}=\text{O}_i$, y los valores de sus ángulos diedros (ϕ , Ψ) son aproximadamente -57° y -47° (Kendrew et al., 1958; Pauling et al., 1951). La segunda estructura helicoidal relevante para la presente tesis es la hélice π o $4,4_{16}$. Este tipo de hélices tienen generalmente una longitud de 5 a 18 residuos, presentan 4,4 residuos y 16 átomos por vuelta del esqueleto peptídico, y están

estabilizadas por puentes de hidrógeno entre los grupos $\text{NH}_{(i+5)}-\text{C}=\text{O}_i$. Además de los enlaces de hidrógeno intrahelicoidales, las hélices π también se estabilizan mediante otras interacciones interhelicoidales y muestran distintas preferencias por residuos específicos (Xie et al., 2020; Kumar y Bansal, 2015). Por otra parte, sus ángulos de torsión presentan valores variables en diferentes proteínas (Tabla 2) (Kumar y Bansal, 2015; Fodje y Al-Karadaghi, 2002). La mayoría de las hélices π se encuentran junto a hélices α en las proteínas de membrana causándoles distorsión (Kumar y Bansal, 2015) y a menudo están implicadas en la formación del centro activo de algunas enzimas (Medlock et al., 2007) o en la apertura de canales TM transportadores de iones (Xie et al., 2020, McGoldrick et al., 2018; Zubcevic et al., 2019), como el estudiado en la presente tesis.

Tabla 2. Parámetros geométricos de las hélices α y π . Los valores de los rotámeros ϕ y ψ son la media y desviación estándar (entre paréntesis) de 659 hélices π y 25,368 hélices α asignados por ASSP (*Assignment of Secondary Structure in Proteins*). Datos obtenidos de Kumar y Bansal, 2015.

Tipo de estructura	Residuos/vuelta	Número de átomos/vuelta	ϕ (°)	ψ (°)
Hélice α (3.6 ₁₃)	3.6	13	-63.3 (20.2)	-44.7 (14.9)
Hélice π (4.4 ₁₆)	4.4	16	-81.2 (20.7)	-37.5 (22.6)

Las hélices TM adaptan su longitud a las dimensiones y limitaciones de las membranas biológicas, mientras que las hélices de proteínas solubles son generalmente más cortas porque no tienen las restricciones del entorno impuestas por la membrana (Baeza-Delgado et al., 2013). Estos dominios TM están en contacto con el interior hidrofóbico de la membrana y no solo sirven de anclaje a la misma, sino que también realizan funciones biológicas específicas (Cheng W. y Li W., 2014; Zhou et al., 2012). Además de los dominios TM en las proteínas integrales de membrana hay regiones expuestas al entorno acuoso denominadas dominios extramembranosos (EMD). Los EMD tienen propiedades estructurales similares a las proteínas solubles y algunos presentan funciones biológicas independientes de los dominios TM, como el sitio de hidrólisis de ATP en la ATPasa (bomba Na/K), o los péptidos de fusión presentes en glicoproteínas de virus con envuelta lipídica como gp41 del virus influenza y de la proteína S del SARS-CoV-2 (Ohta et al., 1986; Zhang et al., 2022).

Composición y distribución de aminoácidos en las IMP

Los segmentos TM de las proteínas de membrana presentan, además de estructura secundaria en su esqueleto polipeptídico, aminoácidos generalmente hidrofóbicos con cadenas laterales compatibles con el núcleo hidrofóbico de la bicapa lipídica. La baja presencia de aminoácidos cargados en los segmentos TM hace que su inserción en la bicapa lipídica sea termodinámicamente favorable. Los aminoácidos con cadena lateral alifática valina (V), isoleucina (I), glicina (G), fenilalanina (F), alanina (A), leucina (L) son abundantes en las hélices TM de las proteínas integrales de membrana, mientras que los aminoácidos polares y cargados a pH fisiológico como aspártico (D), glutámico (E), lisina (K), arginina (R) o histidina (H), son menos frecuentes (Baeza-Delgado et al., 2013; Punta et al., 2013). Asimismo, la figura 5A muestra que en el centro hidrofóbico hay una abundancia relativa de los aminoácidos hidrofóbicos y ligeramente polares valina (V), isoleucina (I), cisteína (C), metionina (M), leucina (L), alanina (A), fenilalanina (F), glicina (G), treonina (T) y serina (S). También se observa la presencia de triptófano (W), tirosina (Y) y prolina (P) en la interfase extracitoplasmática. Los aminoácidos polares glutamina (Q), histidina (H) y asparagina (N) predominan en ambas regiones interfaciales, siendo su presencia dentro del núcleo hidrofóbico de la membrana energéticamente desfavorable. Los aminoácidos cargados aspártico (D), glutámico (E), lisina (K) y arginina (R) son prevalentes en la interfase citoplasmática, siendo los aminoácidos ácidos (D y E) más prevalentes en la interfase extracitoplasmática. Es importante destacar que los aminoácidos básicos K y R son fuertes determinantes topológicos de las proteínas de membrana y están presentes en gran medida en la región flanqueante citoplasmática. Estudios estadísticos de proteínas de membrana con estructura resuelta (Figura 5B) mostraron que, a parte de los residuos hidrófobos, glicina (G) y los aminoácidos hidroxilados serina (S) y treonina (T) están igualmente distribuidos a lo largo del núcleo hidrófobo de la membrana (posición 9,-9). Metionina (M) y cisteína (C) también ocupan diferentes ubicaciones dentro del núcleo hidrofóbico, presentando una cierta prevalencia en la región de los grupos glicerol y fosfato de las cabezas polares de los fosfolípidos. En el caso del triptófano (W), tirosina (Y) y prolina (P) son más abundantes dentro de la región de la interfase (a 8-9 residuos del centro de la membrana), pero con cierta preferencia por la interfase extracitoplasmática. Finalmente, asparagina (N), histidina (H) y glutamina (Q) tienen una tendencia interfacial cercana al final de las regiones TM (Baeza-Delgado et al., 2013).

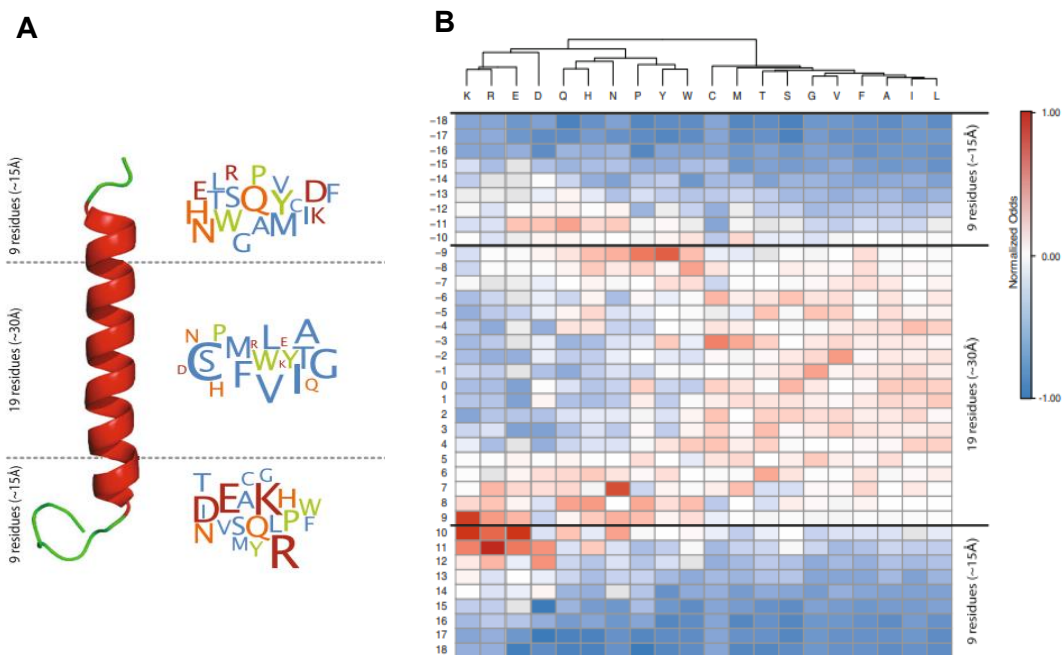


Figura 5. Prevalencia de aminoácidos en proteínas integrales de membrana. (A) Los aminoácidos hidrofóbicos se muestran en azul. W, Y y P en verde. Los aminoácidos polares Q, H y N anaranjados. Los aminoácidos cargados D, E, K y R en rojo. El tamaño de las letras es proporcional a su frecuencia relativa (*Odds* normalizadas). (B) Distribución de los aminoácidos en las hélices TM. En este gráfico se muestra el posicionamiento normalizado según su prevalencia (*Odds*, en escala de rojo a azul se representa respectivamente la abundancia relativa de cada residuo en cada posición a lo largo de la membrana siendo el 0 el centro de la bicapa). Tomado de Baeza-Delgado et al., 2013.

1.3. Biogénesis de las proteínas helicoidales de membrana

Al igual que el resto de proteínas, las proteínas de membrana y de secreción son sintetizadas por ribosomas, si bien, su síntesis la realizan ribosomas asociados a membranas. El direccionamiento de los ribosomas a las membranas puede ser co-traducciona l o post-traducciona l, siendo el primero mucho más común y del que disponemos de más información. Así, la biosíntesis co-traducciona l de las proteínas de membrana puede dividirse en 3 etapas: direccionamiento, inserción y plegamiento. A continuación, pasaremos a describir de forma breve estas etapas, para mayor detalle se recomiendan las revisiones actualizadas de Hegde y Keenan, 2022; Li et al., 2021, O'Keefe et al., 2021 y Duarte et al., 2023.

El *direccionamiento* co-traducciona l de las proteínas de membrana y de secreción está mediado por la presencia de una región hidrofóbica próxima al extremo N-terminal de la

proteína: una secuencia o péptido señal (de 7 a 9 aminoácidos hidrofóbicos) escindible, o en su ausencia, un primer dominio TM (de 15 a 25 aminoácidos hidrofóbicos). Generalmente, esta región hidrofóbica es reconocida por una chaperona citoplasmática, la partícula de reconocimiento de señal (SRP); la unión de la ribonucleoproteína SRP a la cadena nascente genera el complejo de direccionamiento (TC), detiene temporalmente la traducción y dirige el complejo cadena nascente-ribosoma (RNC) hacia la membrana plasmática en el caso de procariotas o hacia la membrana del retículo endoplasmático (RE) en el caso de eucariotas, donde se encuentra el receptor de la SRP asociado al complejo proteico del translocón (Figura 6). Una vez se ha posicionado el ribosoma sobre el translocón, la SRP se separa (mediante la hidrólisis de GTP) del ribosoma reiniciándose la síntesis de la proteína (Li et al., 2021).

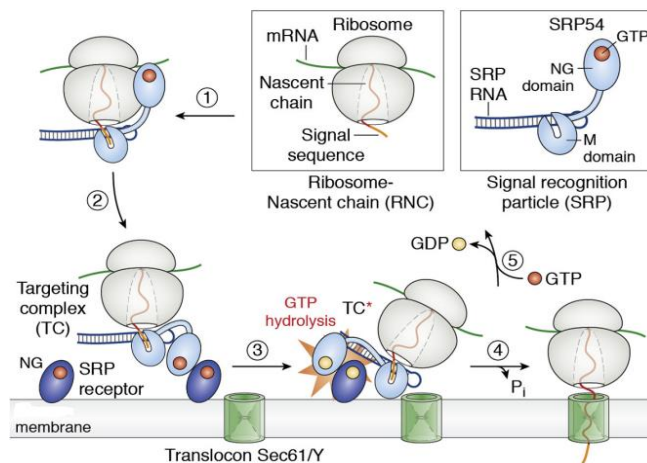


Figura 6. Direccionamiento co-traduccion de proteínas de membrana y de secreción. El direccionamiento es mediado por la secuencia señal (o en su ausencia por el primer segmento TM de la cadena nascente), que es reconocida por la SRP (azul claro). La unión de la SRP detiene temporalmente la traducción y direcciona el complejo (TC) a la membrana donde se encuentra el receptor SRP (azul oscuro) asociado al translocón (verde). La interacción de la SRP con su receptor facilita el posicionamiento del ribosoma sobre el translocón. A continuación, mediante hidrólisis de GTP, la SRP y su receptor se disocian del ribosoma y se reinicia la síntesis de la cadena nascente. Tomado de Li et al., 2021.

En el caso de las proteínas de membrana en esta etapa se produce la *inserción* de la cadena nascente, dado que a medida que la proteína va siendo sintetizada, sus segmentos TM deben ser insertados en la bicapa lipídica a través del translocón Sec61/SecY. Recientemente se han descrito, otros complejos insertasas que se ha propuesto que participan en la inserción de proteínas que se anclan a través de una región hidrofóbica próxima a su extremo C-terminal (*C-tail anchoring*, TA) y en la inserción de proteínas politépicas, como el complejo GET1-GET2, el complejo EMC y el complejo TMCO1 (Hegde y Keenan, 2022; O'Keefe et al., 2021) que pasaremos a describir brevemente.

Complejo Sec61/SecY. El translocón es un complejo multiproteico de membrana conservado y presente en todos los organismos, procariotas, eucariotas y arqueas (Park y Rapoport, 2012). El componente central del translocón es un complejo heterotrimérico, formado por tres subunidades Sec61 α , Sec61 β y Sec61 γ en mamíferos (Görlich et al., 1992; Görlich and Rapoport, 1993), SecY, Sec β y SecE en arqueobacterias y SecY, SecE y SecG en bacterias. El componente central del canal es Sec61 α (o sus homólogas SecY), proteína de 53 kDa que consta de 10 dominios TM organizados en dos dominios de cinco hélices cada uno (TMs 1-5 y TMs 6-10) que dejan un poro hidrofílico en el centro (Voorhees et al., 2014), similar a SecY en arqueobacterias (Van den Berg et al., 2004) y bacterias (Voorhees y Hegde, 2016). El lazo entre las hélices TM 5 y 6 actúa como una bisagra, conectando ambas mitades (Figura 7 A). En el lado opuesto de la bisagra, el TM2 y el TM7 forman una puerta lateral (Martínez-Gil et al., 2011), que permite la difusión de los segmentos TM a la bicapa lipídica y que permanece cerrada en estado de reposo, es decir sin ribosoma unido. En ausencia de ribosoma unido el poro central del translocón está obstruido por una pequeña hélice (solapa) entre el TM1 y TM2. Además, en el centro del poro se disponen residuos aromáticos que forman un anillo hidrofóbico y producen una constricción central, que estrecha el canal del translocón, dando al poro una forma de reloj de arena con un diámetro de sólo 5-8 Å en dicha región (Figura 7 B) (Junne et al., 2010; Li et al., 2021) que impide el paso de moléculas e iones. En mamíferos, Sec61 β y Sec61 γ son proteínas de membrana asociadas a Sec61 α . Sec61 γ es la subunidad más pequeña (7.7 kDa), formada por un dominio TM largo y curvo orientado en diagonal a la membrana, por lo que entra en contacto con ambas mitades de Sec61 α manteniéndolas unidas (Rapoport et al., 2017; van den Berg et al., 2004). Por su parte, Sec61 β es una proteína de 10 kDa, ubicada cerca de las hélices TM1 y TM4 de Sec61 α , y se ha propuesto que participa en el reconocimiento del complejo RNC (Jiang et al., 2008).

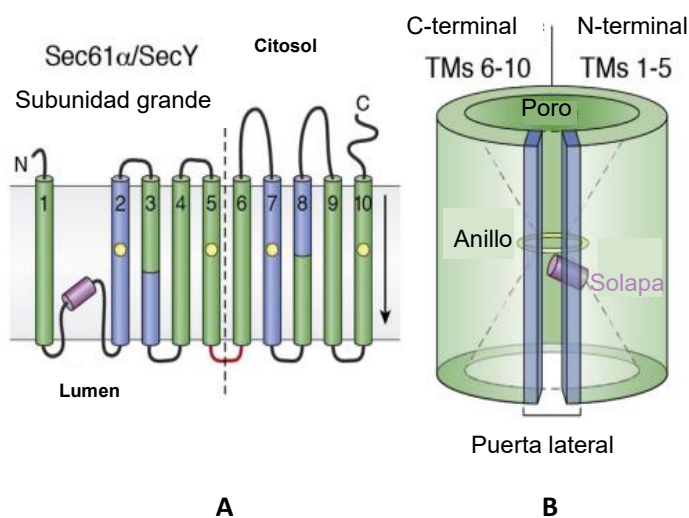


Figura 7. Topología y estructura de Sec61α. (A) Ambos extremos de Sec61α/SecY (N y C terminal) están orientados hacia el citosol. Esta proteína consta de 10 hélices TM. Entre el TM5 y TM6 se forma una bisagra (línea discontinua). El TM2 y TM7 junto a un segmento de los TM3 y TM8 (en azul) forman la puerta lateral. La flecha hacia abajo indica la dirección de entrada en el poro y los residuos del anillo hidrofóbico se indican en amarillo. (B) Esquema del poro formado por Sec61α. Se muestra el anillo hidrofóbico (amarillo), la puerta lateral (azul), y la solapa (morado). Adaptado de Li et al., 2021

El complejo RNC se une al lazo extramembrana que une los TMs 8 y 9, y al C-terminal de Sec61α. Este acoplamiento provoca un cambio conformacional en el poro del translocón que permite la apertura de la puerta lateral hacia la bicapa del RE y a su vez la modificación estructural de su extremo citosólico para facilitar el acoplamiento de las secuencias señales (péptido señal o primer dominio TM) entre los TMs 2 y 7 (Plath et al., 1998; Raden et al., 2000). La interacción ribosoma-Sec61α genera movimientos menores de las hélices TM 6, 7, 8 y 9 de Sec61α; sin embargo, las otras hélices TM del complejo Sec61α experimentan un giro de aproximadamente 22°, creando un espacio entre las hélices 2 y 7 que permite intercalar el péptido señal (Pfeffer et al., 2015; Voorhees y Hegde, 2016). Esta rotación de las hélices 2, 5 y 10 lejos de la hélice 7 desplaza al conjunto de residuos hidrofóbicos que forman el anillo del poro, alterando su configuración y desestabilizando a la solapa; lo que crea un canal sin obstrucciones a través de la membrana que permite el paso de la cadena nascente. La apertura de este canal está directamente relacionada con el reconocimiento de la secuencia señal a través de su intercalación en la puerta lateral de Sec61α (Voorhees y Hegde, 2016). Cuando la secuencia señal (o en su defecto el primer segmento TM) es insertada en la bicapa vía la puerta lateral, el péptido señal se une a Sec61α, con su N-terminal hacia el citosol, dejando su sitio de digestión accesible al complejo del péptido señal (SPC) hacia el lumen. Esta disposición y orientación hace que la secuencia polipeptídica que sigue al péptido señal sea atraída hacia el canal hidrofílico central de Sec61. A partir de este punto, la síntesis de cadenas nascentes correspondientes a proteínas de secreción da como resultado la translocación del polipéptido al otro lado de la membrana, y el

péptido señal hidrofóbico es escindido por el complejo peptidasa del péptido señal (SPC), cuyo sitio activo está orientado hacia el lumen del RE en eucariotas y arqueas o hacia el periplasma en bacterias. Por lo que respecta a las proteínas de membrana, el translocón Sec61/SecY permite la inserción subsiguiente de los distintos segmentos TM. La hidrofobicidad de los segmentos TM favorece su difusión a través de la puerta lateral del canal e inserción en el entorno hidrofóbico de la membrana circundante (Hegde et al., 2022; Whitley y Mingarro, 2014).

Complejo EMC. El complejo de proteínas de membrana del RE (EMC) fue implicado en la biogénesis de proteínas de membrana en experimentos de disrupción génica de sus subunidades, los cuales generaron efectos pleiotrópicos en la expresión de proteínas politópicas (Bai et al., 2020; Gaspar et al., 2021; Miller-Vedam et al., 2020; O'Donnell et al., 2020; Satoh et al., 2015; Tang et al., 2017; Volkmar et al., 2019). EMC es un complejo de 10 subunidades que reside en el RE y presenta dominios luminales y citosólicos voluminosos (Figura 8). Los dominios citosólicos de EMC tienen sitios de unión a los que pueden asociarse segmentos TM poco hidrofóbicos, lo cual podría favorecer la inserción en membrana de proteínas TA y de proteínas politópicas de tipo III con hélices TM de baja/moderada hidrofobicidad (Hegde y Keenan, 2022). La inserción en membrana de ambos tipos de proteínas se ve facilitada por la actividad insertasa y chaperona de este complejo. En la estructura de este complejo, se ha observado que su función insertasa recae fundamentalmente en las subunidades EMC3, EMC4 y EMC6, las cuales forman un vestíbulo hidrofílico conservado situado en el lado citosólico de la bicapa. Cabe destacar que EMC3 es homóloga a YidC (una insertasa de bacterias) (Figura 8) (McDowell et al., 2021), la cual tiene dominios citosólicos que contactan transitoriamente con SecY (ortólogo de Sec61 α) durante la inserción de las proteínas de membrana. Se ha propuesto que EMC3 y YidC podrían capturar selectivamente dominios TM de proteínas tipo III y dirigirlos hacia el sitio de la insertasa de EMC (Chitwood y Hegde, 2019).

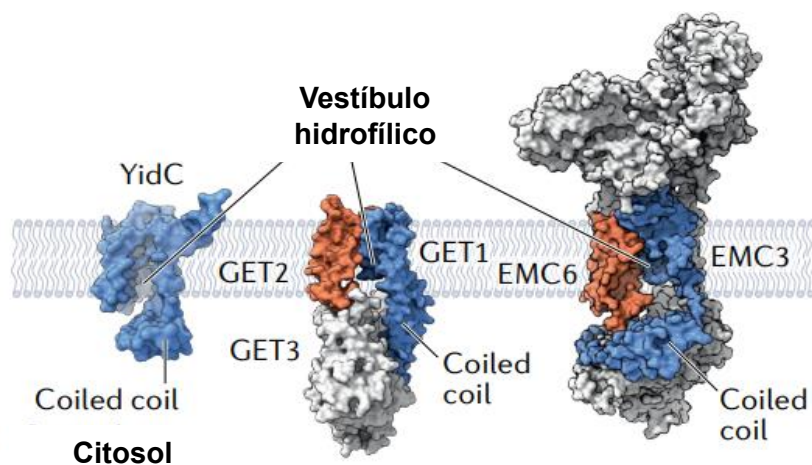


Figura 8. Estructura de EMC y sus homólogos funcionales YidC y GET. Comparación estructural de YidC bacteriana con GET1 y EMC3 de eucariotas. Estas 3 proteínas (en azul) conservan un vestíbulo hidrofílico característico que facilita la translocación regiones hidrofílicas que flanquean los TMD durante su inserción. EMC 3 (azul) y 6 (anaranjado) forman el vestíbulo hidrofílico del complejo EMC en humanos, el cual está orientado hacia el citoplasma. Adaptado de Hegde y Keenan, 2022.

Además, al igual que YidC y SecYEG en bacterias, el complejo EMC se ha propuesto que pueda realizar la inserción co-traducciona de proteínas de membrana tanto sola como actuando con el complejo Sec61; de forma que EMC facilitaría la translocación del extremo N-terminal hacia el lumen, permitiendo posteriormente que el ribosoma se acople a Sec61, donde se sintetizaría e insertaría el resto de la proteína (Figura 9). Cabe destacar que la función de estabilización e inserción de EMC no está estrictamente limitada al primer TMD de las proteínas politópicas tipo III, sino que su función, al menos en algunos casos, se extiende también hacia los siguientes segmentos TM, independientemente de la orientación del primer TM (O'Keefe et al., 2021). El factor decisivo que determina la dependencia de las proteínas politópicas por el complejo EMC parece ser la reducida hidrofobicidad (y/o su aumento de carga o polaridad) de algunos de sus segmentos TM. De hecho, experimentalmente, la introducción de residuos polares/cargados en segmentos TM resulta en una inserción dependiente de EMC (Coelho et al., 2019; Lin et al., 2019; Tian et al., 2019). Asimismo, también se ha descrito que, en ausencia de un complejo EMC funcional, una proteína de membrana con segmentos TM subóptimos en términos de hidrofobicidad es degradada vía de control de calidad, resaltando el papel del complejo EMC como chaperona de proteínas politópicas y diferenciándolo funcionalmente del translocón Sec61 (Hegde y Keenan, 2022; O'Keefe et al., 2021).

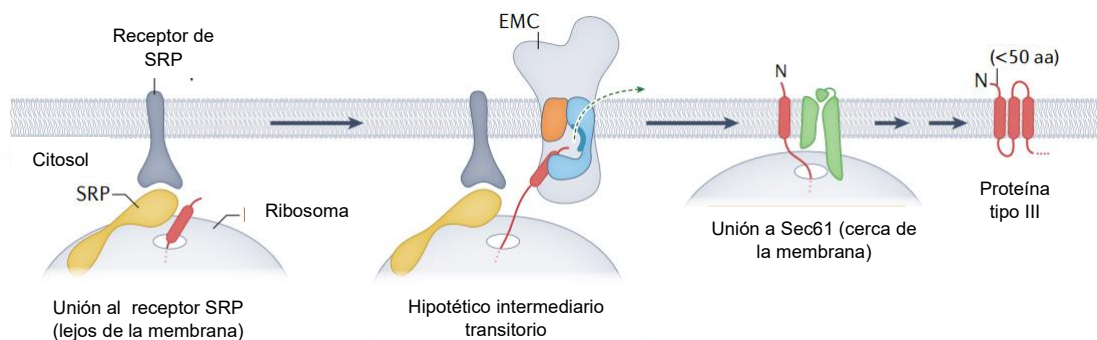


Figura 9. Inserción de proteínas de membrana politópicas mediada por EMC. Se ha propuesto que en algunos casos EMC acomode inicialmente el segmento TM facilitando su inserción. Para ello se ha propuesto que tras ser dirigido a la membrana por la SRP (amarillo) donde se encuentra su receptor (gris oscuro), el ribosoma deposita la cadena nascente en el vestíbulo hidrofílico de EMC3 (azul), facilitando la translocación del extremo N-terminal hacia el lumen, posicionándose a continuación sobre Sec61 (verde) y facilitando la inserción del resto de la proteína. Adaptado de Hegde y Keenan, 2022.

Complejo TMCO1. La proteína TMCO1 es una insertasa que también presenta un vestíbulo hidrofílico. Esta proteína se asocia con la proteína CCDC47 y Nicalin-TMEM147-NOMO para formar el translocón TMCO1 (Figura 10), que se ha vinculado también en la biogénesis de las proteínas integrales politópicas (McGilvray et al., 2020). Además, la similitud estructural entre TMCO1 y EMC, sugiere que TMCO1 podría integrar preferentemente segmentos TM poco hidrofóbicos. Se ha demostrado que TMCO1 puede asociarse con el complejo Sec61 unido al ribosoma y colaborar en la inserción de dominios TM subóptimos, así como en la protección de TMs recién integrados durante el ensamblaje de proteínas de membrana politópicas. Sin embargo, parece ser que la formación del complejo TMCO1 es transitoria, dado que se ensambla y desensambla de acuerdo con las necesidades del complejo Sec61 y RNCs, habiéndose observado que sus subunidades no forman un complejo estable en ausencia de ribosomas unidos (O'Keefe et al., 2021).

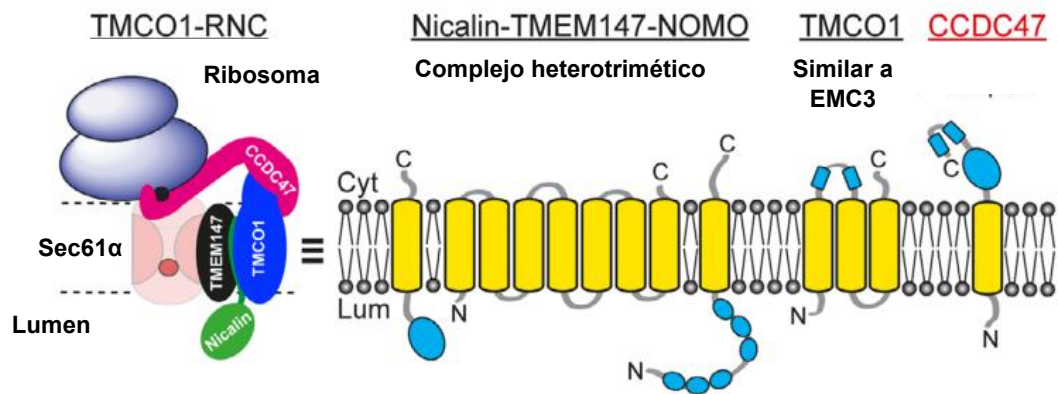


Figura 10. Translocón TMCO1. La asociación de la proteína TMCO1 con el complejo Nicalin-TMEM147-NOMO y CCDC47 forma el holocomplejo TMCO1, el cual tiene actividad insertasa y chaperona. CCDC47 es la responsable del contacto con el ribosoma. Todas las proteínas que forman parte del translocón TMCO1 son proteínas de membrana (cuyos segmentos TM se muestran en amarillo) con dominios y bucles extramembrana (en celeste). Sec61 α se muestra de color rosa. Adaptado de O'Keefe et al., 2021.

Complejo PAT. El complejo PAT es un heterodímero formado por la proteína Astérix (inicialmente denominada PAT10) y CCDC47 que fue descubierto mediante estudios *in vitro* de biogénesis de proteínas politópicas (Meacock et al., 2002). Se ha sugerido que después de que el primer segmento TM de una proteína politópica abandona el translocón Sec61, puede ser reconocido por Astérix y asociarse a ésta, mientras que los siguientes segmentos TM de la misma cadena naciente podrían integrarse a través del complejo Sec61 (Ismail et al., 2006 y 2008). En este contexto, se ha propuesto que la subunidad Astérix del complejo PAT se asociaría a segmentos TM poco hidrofóbicos (que presentan residuos cargados/polares) permaneciendo asociada a estos incluso después de la terminación de la traducción, protegiendo estos segmentos TM “subóptimos” hasta que se empaquetan correctamente para posibilitar la adopción de la conformación nativa de la proteína, como se observa en la figura 11 (Hegde y Keenan, 2022; O'Keefe et al., 2021).

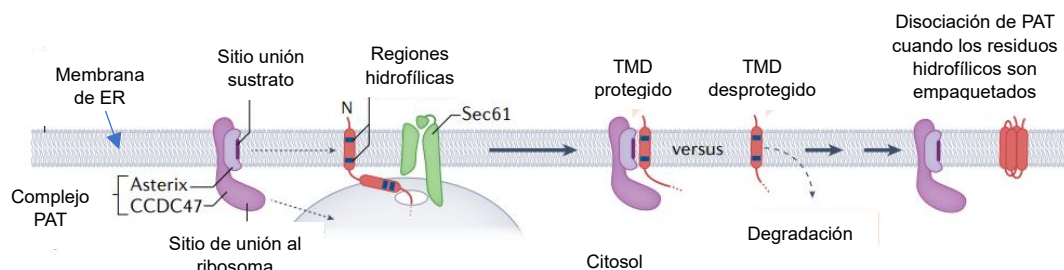


Figura 11. Complejo PAT. Este complejo con actividad chaperona está formado por la asociación de las proteínas de membrana CCDC47 (morado) y Asterix (lila). Los dominios citosólicos de CCDC47 interaccionan con el ribosoma y Asterix se asocia con segmentos TMs que presentan residuos hidrofílicos (en azul) evitando que tales TMs sean direccionados al control de calidad y posteriormente degradados. El complejo PAT se disocia de sus sustratos facilitando la adopción de la estructura nativa de la proteína nascente (rojo). Adaptado de Hegde y Keenan, 2022.

Esta preferencia del complejo PAT por los TMs poco hidrofóbicos junto a su capacidad para proteger TMs independientemente de su orientación caracterizan su actividad chaperona, de alguna forma similar a la actividad de EMC (Chitwood et al., 2020; Meacock et al., 2002), de hecho, se ha demostrado que algunas proteínas politópicas que se insertan a través del complejo EMC también requieren el complejo PAT para adquirir su conformación nativa (Chitwood et al., 2020).

Determinantes topológicos y de inserción.

La longitud de la región hidrofóbica de las hélices TM es un factor que afecta a su inserción. Como se ha mencionado anteriormente, el núcleo hidrofóbico de la bicapa lipídica tiene un espesor aproximado de 30 Å, por lo que son necesarios alrededor de 20 residuos hidrofóbicos consecutivos para que su inserción sea termodinámicamente favorable, dado que en la hélice- α canónica la distancia de un residuo al siguiente es de 1,5 Å. En las estructuras de proteínas resueltas, la longitud promedio de las hélices TM es de 24,0 (+/- 5,6) residuos (Baeza-Delgado et al., 2013). Varios estudios experimentales y computacionales sugieren que se requiere un mínimo de 10-12 residuos altamente hidrofóbicos para una inserción eficiente (Baeza-Delgado et al., 2016; Jaud et al, 2009; Ulmschneider et al, 2011). La elevada hidrofobicidad de las regiones TM favorece su inserción con una topología N-terminal extracitoplasmática (Saurí et al., 2009; Wahlberg y Spiess, 1997). Además de residuos hidrofóbicos, también encontramos en las hélices TM residuos polares y cargados (Bañó-Polo et al., 2012; Duart et al., 2022), que suelen estar muy conservados. Debemos tener en cuenta que en los segmentos TM, no todos los residuos interaccionan con los lípidos, ya que el empaquetamiento de las hélices crea entornos donde las cadenas laterales de los

residuos menos hidrofóbicos están orientadas hacia el interior de la proteína, y por tanto “protegidos” del entorno lipídico. Esto explica el hecho de que las hélices TM de las proteínas politópicas sean, en general, menos hidrofóbicas que las proteínas bitópicas (Hegde y Keenan, 2022). En cualquier caso, hay que tener en cuenta que tanto los residuos hidrofóbicos como los residuos polares de las hélices TM establecen enlaces intra e inter-helicoidales dando estabilidad y facilitando el empaquetamiento y plegamiento de las proteínas de membrana, al tiempo que, en el caso de los residuos polares y cargados, reducen su polaridad (Popot et al., 2000; Ubarretxena-Belandia et al., 2001). Otro factor determinante en la inserción y sobre todo de la topología de las IMPs son los residuos cargados que se encuentran en las regiones flanqueantes de las hélices TM. El análisis de secuencias y estructuras de proteínas integrales de membrana muestra la prevalencia de los residuos cargados positivamente (arginina y lisina) en el extremo de las hélices TM y sus regiones flanqueantes orientadas al citoplasma (Figura 5). Se ha sugerido que estos residuos podrían dirigir la orientación de la hélice al interaccionar con la carga negativa de los fosfolípidos y prevenir la translocación de las regiones de la cadena polipeptídica con carga positiva (von Heijne, 1986). Por otra parte, los residuos cargados negativamente (aspártico y glutámico) parecen ser señales topológicas de translocación cuando están flanqueando una hélice TM; si bien, se requieren varios residuos cargados negativamente para inducir la translocación de los polipéptidos o atenuar el efecto topológico de los residuos cargados positivamente tanto en proteínas de membrana bacterianas como eucariotas (Nilsson et al., 1990).

Las hélices de las proteínas de membrana politópicas una vez insertadas se ensamblan entre ellas para formar haces y a este proceso se le llama *plegamiento*. Se ha demostrado que muchos segmentos TM contienen residuos polares o cargados que son necesarios para su funcionamiento en el entorno de la membrana, siendo poco estables en ausencia de sus vecinos, por lo que requieren el establecimiento de contactos entre ellos para su estabilidad en membrana (Choi et al., 2019; Enquist et al., 2009).

Puentes salinos. Las hélices TM interactúan entre sí formando estructuras funcionales. Para formar estas estructuras, los segmentos TM se asocian a través de interacciones entre los aminoácidos que incluyen la formación de puentes salinos, permitiendo el plegamiento de la proteína y la adopción de estructuras terciarias funcionales. Los aminoácidos que participan en la formación de puentes salinos son lisina (K), arginina (R), histidina (H), aspártico (D) y glutámico (E); que, a pesar de su menor prevalencia, están presentes en muchas proteínas politópicas formando puentes salinos, como por ejemplo en el dominio sensor de voltaje de la proteína transportadora de sodio (Barber

et al., 2012) y en la proteína transportadora de dopamina (Pedersen et al., 2014). La formación de estos puentes salinos reduce la polaridad de las hélices TM que presentan estos residuos cargados, lo que facilita su inserción. De hecho, se ha estimado en 0,5-0,7 kcal/mol la contribución de la formación de un puente salino intra-hélice a la inserción de hélices TM que presentan residuos con cargas opuestas a distancias i , $i+3$ o $i+4$ (Duart et al., 2022).

Puentes de hidrógeno. Los enlaces de hidrógeno tienen una mayor contribución en la estabilización de las proteínas de membrana que en las proteínas solubles; ya que, dentro de la región apolar de la bicapa la constante dieléctrica del medio es menor que en un entorno acuoso y los grupos que participan en la formación de un enlace de hidrógeno no tienen la competencia de las moléculas de agua. Por lo tanto, estas interacciones proporcionan una fuerza impulsora en el plegado de segmentos TM (DeGrado et al., 2003), siendo los enlaces de hidrógeno inter-helicoidales particularmente importantes para estabilizar las proteínas de membrana (Adamian y Liang 2001). Esto se demostró en estudios con hélices TM diseñadas en donde un solo residuo polar, como Asn, Asp, Glu, Gln o His, fue suficiente para mediar la homooligomerización de segmentos de polileucinas en membranas biológicas y micelas (Gratkowski et al., 2001, Zhou et al., 2001).

Interacciones hidrofóbicas. El peculiar entorno de la membrana hace que las interacciones hidrofóbicas en las interfases hélice-hélice muestren preferencia por formarse entre residuos con cadenas laterales cortas, como por ejemplo glicinas (G) y alaninas (A), que permitan un mejor empaquetamiento y a su vez estabilizan la oligomerización (Senes et al., 2000; Popot et al., 1990). Por el contrario, los residuos hidrofóbicos con cadenas laterales voluminosas como fenilalanina (F), Leucina (L), Isoleucina (I) y Valina (V) están más frecuentemente ubicados en las superficies de las hélices TM expuestas a las cadenas hidrocarbonadas de los lípidos (Arkin et al., 1996; Ulmschneider et al., 2005). Así, en las proteínas de membrana las interacciones de van der Waals son consideradas la principal fuerza impulsora del empaquetamiento entre hélices. El empaquetamiento de este tipo de proteínas en el núcleo hidrofóbico de la membrana es diferente al de las proteínas solubles (Adamian y Liang, 2001; Hildebrand et al., 2006), incrementándose en las regiones que presentan residuos apolares mediante interacciones de van der Waals, siendo más fuerte la interacción cuando los residuos están más cerca (Oberai et al., 2009). Además, una consecuencia de este mayor empaquetamiento entre hélices TM es una mayor sensibilidad a las mutaciones que causan enfermedades (Oberai et al., 2009).

Motivos implicados en el empaquetamiento de hélices TM. Las proteínas integrales de membrana a menudo expresan su función biológica como oligómeros. El dímero que forma la hélice TM de glicoforina A (GpA) ha sido un modelo de estudio fundamental para empezar a comprender este tipo de interacciones (Mackenzie et al., 2006). Su estructura revela una organización en la que las dos hélices adyacentes no son totalmente paralelas entre sí, pues forman un ángulo y un punto de cruce donde tienen un mayor acercamiento (Mackenzie et al., 1997). Estos puntos de cruce en las hélices TM de los dímeros introducen restricciones en la secuencia de la proteína favoreciendo pequeños residuos en dichas posiciones, siendo la glicina en particular la que estabiliza la interacción entre las hélices (Lemmon et al., 1992; Mingarro et al., 1997). En el caso de GpA, el motivo principal que dirige la dimerización de sus segmentos TM es G⁷⁹XXXG⁸³, en donde la glicina 83 es el residuo más importante para formación del dímero (Trenker et al., 2015). Este y otros motivos de dimerización, ampliamente descritos pueden ir acompañados de residuos pequeños como serina, treonina y alanina (Dawson et al., 2002; DeGrado et al., 2003; Orzáez et al., 2005; Russ y Engelman, 2000; Zhou et al., 2001).

I.4. Modificaciones cotraduccionales de las IMPs

Como hemos mencionado los péptidos señal son *procesados por el complejo peptidasa SPC*. Existen dos parálogos SPC heterotetraméricos funcionales formados por las proteínas accesorias SPCS1, SPCS3 y SPCS2 y una subunidad proteolítica SEC11A en las cuales se encuentra su sitio catalítico formado por una triada serina-histidina-aspartato orientado al lumen de RE (Figura 12) (Liaci et al., 2021). Recientemente se ha identificado, que el complejo SPC además de eliminar los péptidos señal en el RE, actúa como factor de control de calidad de proteínas de membrana. En este sentido, SPC reconoce y corta por unos sitios de corte crípticos que son expuestos cuando las proteínas de membrana no se pliegan o ensamblan correctamente, dirigiéndolas de esta manera a su degradación (Zanotti et al., 2022). La subunidad SPCS1 no catalítica, que recientemente se ha demostrado que está regulada negativamente en el cerebro de los pacientes con enfermedad de Alzheimer, participa de manera crítica en las funciones de control de calidad del SPC (Zanotti et al., 2022).

Además, el complejo SPC es utilizado por algunos virus para la maduración de sus poliproteínas virales, incluidas las pre-proteínas de la mayoría de los flavivirus (por ejemplo, el virus del Zika, el dengue y la hepatitis C), el VIH y el coronavirus del SARS (Estoppey et al., 2017; Oostra et al., 2007; Snapp et al., 2017; Suzuki et al., 2013; Zhang

et al., 2016). Cabe destacar también, que el procesamiento de proteínas virales no estructurales de los flavivirus favorece la evasión de la respuesta inmune innata mediada por interferón, en la célula hospedera (Chen et al., 2017). Todo lo anterior, indica que los virus aprovechan una característica inherente a esta enzima revelando que el rango de sustratos de la SPC es más amplio de lo que se creía (Zanotti et al., 2022).

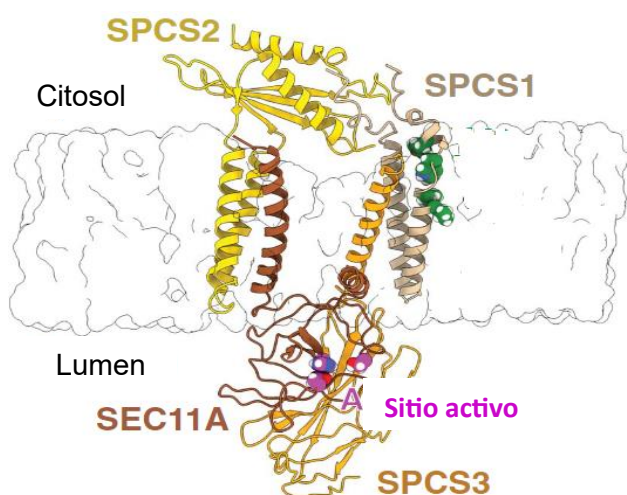


Figura 12. El complejo señal peptidasa (SPC). El esquema muestra el complejo SPC humano, que comprende las proteínas accesorias SPCS1 en gris, SPCS3 en anaranjado y SPCS2 en amarillo y la subunidad proteolítica SEC11A en marrón. La triada catalítica de SPC se muestra en morado (serina-aspartato-histidina). Adaptado de Zanotti et al., 2022.

La translocación de proteínas a través de la membrana es acompañada de algunos mecanismos de modificación de la cadena nascente más allá de la digestión del péptido señal, como por ejemplo la glicosilación de la cadena polipeptídica. En las células eucariotas, *la glicosilación de las proteínas de membrana y de secreción* es un proceso muy común que tiene lugar en el lumen del RE. La adición cotraduccional de azúcares a las cadenas nascentes la realiza el complejo multienzimático *oligosacaril transferasa* (OST) asociado al translocón, formado por 9 subunidades (Ost1p, Ost2p, Ost3p, Ost4p, Ost5p, Ost6p, Stt3p, Swp1p and Wbp1p) e integrado en la membrana del RE con su sitio catalítico orientado hacia el lumen (Figura 13). Esta enzima agrega oligosacáridos al grupo amida de la cadena lateral de asparaginas en la secuencia consenso Asn-X-Thr/Ser que se encuentran a una distancia de al menos a 12-14 residuos de un segmento TM (Orzáez et al., 2004; Nilsson y von Heijne, 1993; Tai y Imperiali, 2001; Welply et al., 1983). Este proceso específico puede ser aprovechado experimentalmente para obtener información topológica. De esta manera se pueden añadir sitios de glicosilación a secuencias de proteínas de membrana desprovistas de ellos para estudiar su topología en sistemas de expresión *in vivo* o *in vitro* (Tamborero et al., 2011; Bañó-Polo et al., 2011). En estos sistemas el estado de glicosilación puede ser detectado mediante técnicas electroforéticas, pues un oligosacárido añadido en la

cadena polipeptídica incrementa su masa molecular en aproximadamente 2,5 kDa (van Geest et al., 2000).

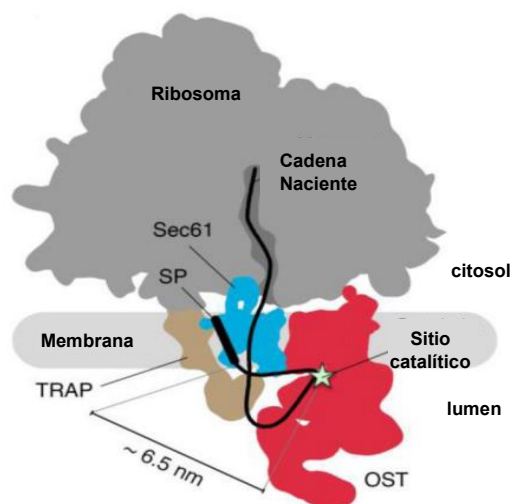


Figura 13. Esquema de la maquinaria de glicosilación (OST) cotraduccional.

En células eucariotas la glicosilación ocurre en el lumen del RE y lo realiza la enzima oligosacaril transferasa (OST) (en rojo), asociada al translocón (azul) que agrega oligosacáridos a la cadena naciente (negro) mientras está siendo sintetizada por el ribosoma (gris). La glicosilación eficiente de Asn por el sitio activo de la OST requiere una distancia de la membrana de 6.5 nm aproximadamente. Adaptado de Braunger et al., 2018.

I.5. Proteína NALCN (Sodium Leak Channel Non-selective)

El gen que la codifica la proteína NALCN humana está ubicado en el cromosoma 13 (13q33.1) y comprende al menos 44 exones. Se expresa principalmente en el sistema nervioso central (SNC), pero también en el corazón, la glándula suprarrenal, la glándula tiroides, los ganglios linfáticos y los islotes de Langerhans (Kutlu et al., 2009; Swayne et al., 2009). En el caso concreto del SNC, NALCN se expresa principalmente en neuronas, y en menor medida en oligodendrocitos y en astrocitos (Cahoy et al., 2008). En invertebrados, los ortólogos de NALCN se concentran en la zona de sinapsis y a lo largo del axón de neuronas en *Drosophila melanogaster* y *Caenorhabditis elegans* (Nash et al., 2002; Yeh et al., 2008). En el gen NALCN (*na*) (también llamado Rb21, VGCNL-1, NA en *D. melanogaster* y NCA-1/2 en *C. elegans*) aparecen mutaciones en la región codificante, ocasionando diferentes fenotipos tanto en mamíferos como en invertebrados. Algunas de estas mutaciones producen descendencia viable en *C. elegans* y *D. melanogaster*, mientras que en ratones homocigóticos son letales (Krishnan y Nash, 1990; Lu et al., 2007; Nakayama et al., 2006; Specca et al., 2010). En diferentes estudios realizados en *D. melanogaster* se observó que la supresión del gen *na* genera la aparición de fenotipos diferentes al silvestre, generando moscas mutantes más pequeñas y con abdómenes más delgados y alargados (Nash et al., 2002; Humphrey et al., 2007). Por otra parte, NALCN es importante en los ritmos circadianos de *D. melanogaster*, ya que mutantes nulos para esta proteína muestran un ritmo

circadiano interrumpido, hasta el punto de que siendo esta especie diurna, los mutantes nulos exhiben la mayor parte de su actividad durante la noche (Nash et al., 2002). Asimismo, también se ha evaluado la participación de NALCN en la agrupación social de animales en *D. melanogaster*, observándose que determinados mutantes muestran un fenotipo social deficiente en comparación con las moscas de tipo silvestre, y dicho fenotipo se corrigió cuando se expresó exógenamente el gen *na* en neuronas colinérgicas y glutamatérgicas (Burg et al., 2013). Por otra parte, estudios funcionales en ratones demostraron que NALCN es importante en la supervivencia de mamíferos, ya que crías de ratones homocigotos *knockout* (NALCN^{-/-}) mueren 12 horas después del nacimiento por la interrupción de su ritmo respiratorio. Esta interrupción se debe a una alteración en la señalización eléctrica del sistema nervioso, ya que no se observaron anomalías en el desarrollo embrionario ni en el movimiento espontáneo de las extremidades (Lu et al., 2007). Otro estudio en ratones demostró que NALCN participa en la osmorregulación sistémica al controlar la concentración sérica de sodio (Sinke et al., 2011), y también se ha vinculado a NALCN en la regulación de la actividad marcapasos intestinal en las células intersticiales de Cajal, es decir las contracciones fásicas de los músculos gastrointestinales (Kim et al., 2012)

Estructura de la proteína NALCN.

NALCN es una proteína politópica de 1.738 aminoácidos que consta de 24 hélices TM organizadas en 4 dominios (DI-DII-DIII-DIV) de 6 segmentos TM (S1-S6) cada uno (Kschonsak, et al 2020). Las regiones extramembranas, de alrededor de 50 Å, sobresalen a ambos lados de las hélices TM formando lazos extracelulares (ECL) e intracelulares (ICL). El análisis de la secuencia de NALCN identifica tres sitios de glicosilación putativos en los residuos N210, N216 y N1064, los dos primeros ubicados en el ECL entre S5-S6 del DI y el tercero en el ECL entre S5-S6 del DIII. Además, también se han identificado cuatro puentes disulfuro repartidos en los ECL de los dominios DI, DIII y IV (Xie et al 2020). La organización estructural de los 4 dominios de NALCN es similar, con los 4 primeros segmentos TM (S1-S4) formando el dominio sensor de voltaje (VSD) y los segmentos S5 y S6 formando el dominio del poro (PD) con un filtro selectivo (FS). La estructura de NALCN en humanos resuelta por microscopía crioelectrónica (cryo-EM) (Kschonsak et al., 2020) es semejante a la de los canales de sodio y calcio dependientes de voltaje. A pesar de estas similitudes estructurales NALCN ha sido clasificado como miembro de una nueva subclase de la familia de canales iónicos. Entre las diferencias con otros canales podemos mencionar que los segmentos TM S4 de los dominios DII y DIV de NALCN presentan menos residuos cargados

positivamente (argininas o lisinas) en comparación con los canales de sodio y calcio, esto posiblemente explique la menor sensibilidad de NALCN al voltaje (Chua et al., 2020). Además, la región ICL entre el S4 y el S5 del DI comprendida entre los residuos 157 y 162 (¹⁵⁷YFRFEL¹⁶²), a diferencia de lo que ocurre en los otros tres dominios de NALCN y en otros canales, forma un lazo flexible en lugar de una hélice α . Adicionalmente, en una de las estructuras resueltas se observa que este mismo lazo ICL entre S4-S5 de DI y el propio segmento S5 forman una cavidad que aloja un lípido que podría contribuir a la regulación de NALCN (Xie et al 2020). Finalmente, NALCN tiene un FS iónico particular, formado por los residuos EEKE en sus 4 dominios, el cual difiere del FS de los canales de sodio (DEKA) y calcio (EEEE/EEDD).

Dominio sensor de voltaje (VSD) de NALCN.

La estructura resuelta de NALCN muestra un dominio VSD con una distribución de cargas positivas “*gating charges*” (GC) particular (Kschonsak, et al 2020). El segmento S4 del VSD del DI tiene 4 residuos cargados positivamente (R143, R146, R152, R155), el S4 del VSD del DII tiene 3 residuos básicos (R481, R484, K487), el S4 del VSD del DIII tiene 4 (R989, R992, R995, K998) y el S4 del VSD del DIV presenta sólo 2 residuos cargados positivamente (R1310, K1317) (Figura 14). Se ha demostrado que neutralizando las cargas del S4 en los dominios VSDI y VSDII, se altera la modulación dependiente del voltaje de NALCN, a diferencia de lo que ocurre cuando se neutralizan las cargas del S4 de VSDIII o VSDIV (Kschonsak, et al 2020; Xie et al 2020). Así pues, estos autores han sugerido que la baja sensibilidad al voltaje de NALCN comparado con otros canales podría ser por 4 motivos. *Primero*, el S2-DIII tiene una metionina en posición 929 en lugar de los residuos de fenilalanina o tirosina conservados en los canales de sodio y calcio, los cuales son cruciales para la transferencia de iones a través de las GC durante la activación del canal. *Segundo*, las hélices S4 de los dominios VSD presentan regiones con conformaciones en hélices 3_{10} , excepto S4-DIV que forma mayormente una hélice α teniendo solo una pequeña región en hélice 3_{10} , donde hay una serina en lugar de un residuo cargado positivamente, lo que podría explicar que VSDIII y VSDIV contribuyan poco a la sensibilidad al voltaje. En *tercer* lugar, las GC del lado extracelular (R143-R481-R992) del S4 de los dominios DI, DII y DIII forman numerosos puentes de hidrógeno con los residuos polares o con carga negativa circundantes en comparación con otros canales de iones. En *cuarto* lugar, el ICL entre S4-S5 del dominio DI tiene una estructura diferente a las observadas en los dominios DII, DIII y DIV de NALCN, lo que puede dificultar la transducción efectiva del

acoplamiento electromecánico, ya que estos ICLs son transductores clave que enlazan los VSDs al dominio del poro conductor de iones (Xie et al 2020).

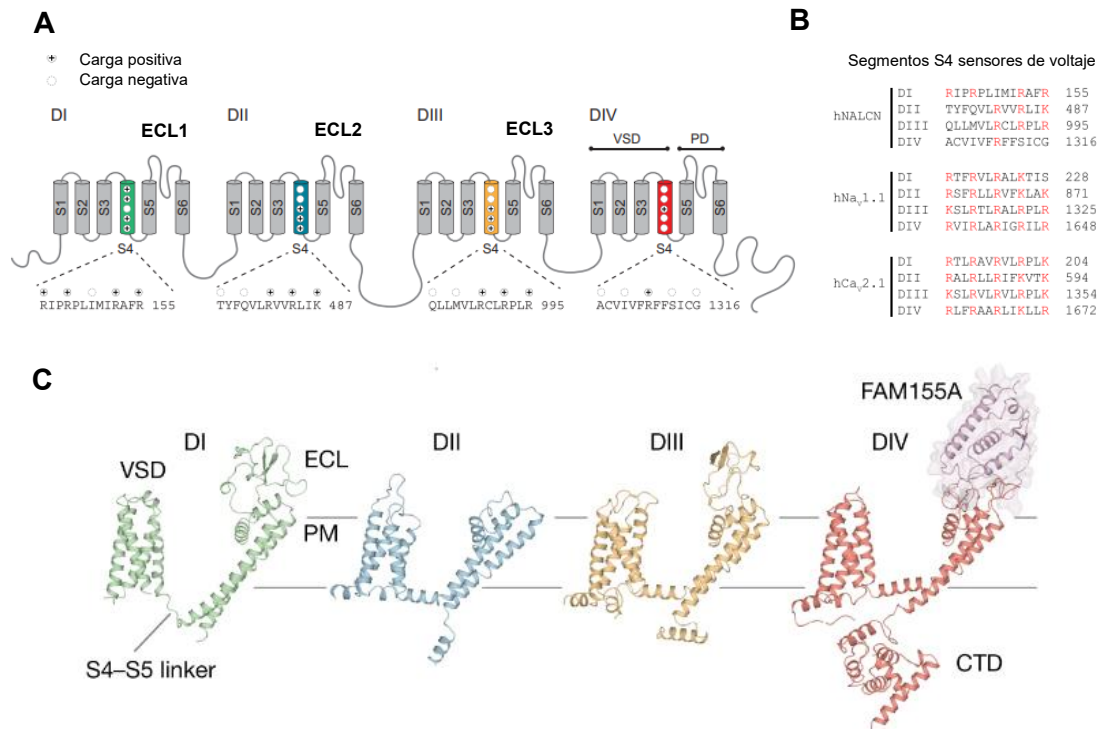


Figura 14. Estructura de la proteína NALCN. (A) Esquema de los 4 dominios (DI, DII, DIII, DIV), representando los segmentos TM como cilindros. Sus extremos N y C terminal están orientados hacia el citoplasma. En el VSD (S1-S4) de cada dominio se destaca de diferente color el S4, en el que se resaltan los residuos básicos (+). Entre los segmentos que forman el dominio del poro (PD) (S5 y S6) hay lazos extracelulares (ECL1,2,3,4), los cuales tienen una región de residuos conservados que forman un filtro selectivo (FS) de iones. (B) Alineamiento de las cargas de activación (en rojo) y su posición en la hélice S4 de cada dominio de NALCN con las de otras proteínas transportadoras de iones Nav 1.1 y Cav 1.1. Adaptado de Chua et al., 2020. (C) Esquema de la estructura de los dominios individualizados siguiendo el código de color utilizado para resaltar la hélice S4 en el panel A. Adaptado de Kschonsak et al., 2020.

Dominio del poro de NALCN

El poro de NALCN tiene un vestíbulo exterior, un filtro de selectividad (FS) de iones (Figura 15), una cavidad central y una puerta de activación (cerrada en reposo). En el vestíbulo exterior la estructura formada por los ECLs permite a los iones acercarse al FS a través de una cavidad extracelular de $9 \times 15 \text{ Å}$, donde un potencial electronegativo concentra cationes en las proximidades del filtro de selectividad. El FS de NALCN es más estrecho y electronegativo que su equivalente en los canales de sodio, pero similar en dimensión al de los canales de calcio. Por tal motivo, los canales de calcio se han usado para establecer similitudes estructurales y funcionales con NALCN (Kschonsak et al., 2020; Xie et al., 2020).

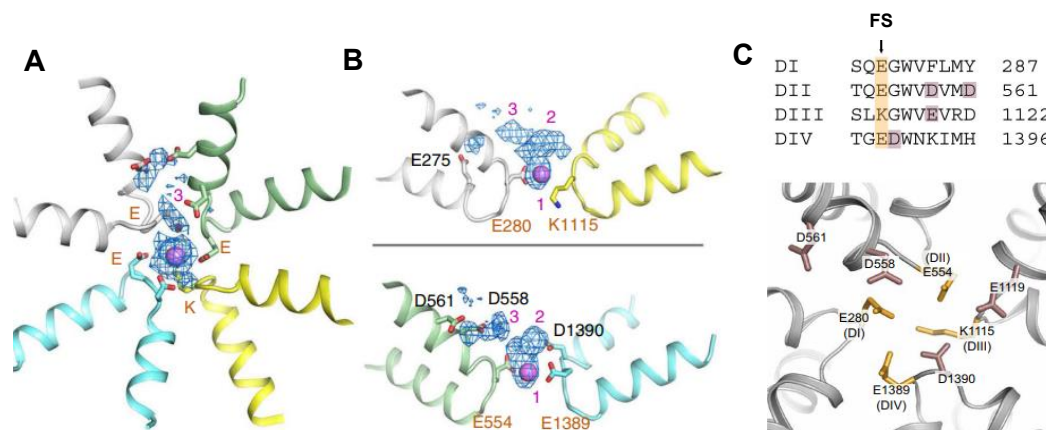


Figura 15. Estructura del filtro selectivo de NALCN. (A) Vista superior. Los principales residuos del FS son E280, E554, K1115, E1389 (en anaranjado) y cada S6 se muestra con su color respectivo (Figura 14C). (B) Vista lateral. La región de FS tiene tres sitios de unión a iones de Na⁺ (esfera morada) con residuos cargados negativamente (E275, D558, D561 y D1390). El Sitio 1 se ubica en el medio de FS y es más estable, mientras el sitio 2 y 3 son transitorios y posiblemente atraen cationes. (C) Residuos del FS de los 4 dominios alineados (anaranjado) y residuos cargados negativamente (en morado). Tomado de Kschonsak et al., 2020; y Xie et al., 2020.

Puerta intracelular de NALCN

La puerta de activación intracelular está cerrada en la estructura resuelta, debido probablemente a que NALCN requiere la presencia de las proteínas UNC80, UNC79 y FAM155A para funcionar como un canal de sodio. El complejo NALCN-FAM155A debe sufrir cambios conformacionales sustanciales para la apertura de esta puerta hacia el citoplasma en presencia de UNC80 y UNC79 (Kschonsak, et al 2020). Los análisis estructurales revelan que los ICLs que unen los cuatro dominios podrían también participar en el cierre y estabilización de la puerta de entrada intracelular mediante enlaces de hidrógeno o interacciones polares entre S6DI y el ICL entre los dominios DII y DIII. La mayoría de los residuos involucrados en esta región de NALCN no están conservados en otros canales de iones, sugiriendo que las interacciones locales entre S6DI, el lazo entre DII-DIII y el lazo entre DIII-DIV son específicos de NALCN y hacen que esta proteína tenga un mecanismo de activación distinto, habiéndose sugerido que los S6 de los cuatro dominios deben sufrir un cambio conformacional que induzca la apertura de esta puerta para que permita el paso de cationes hidratados (Xie et al 2020). La ruta de tránsito de iones debajo del filtro selectivo en NALCN está formada por un grupo tetrahelicoidal de cuatro segmentos S6, una característica estructural presente en canales de iones. Este grupo de hélices deben sufrir un cambio conformacional de hélice

α a π para permitir el paso de iones. Estudios de modelado por homología sugieren que las hélices π se estabilizarían por enlaces de hidrógeno intra-hélice entre W311-N315, L588-S592, M1145-T1149, Y1436-N1440 del segmento S6 de DI, DII, DIII y DIV, respectivamente (Kschonsak et al., 2020). Otro estudio comparativo entre la proteína transportadora de calcio TRPV6 y NALCN describe una región de 7 residuos ubicados en el centro de cada segmento S6 del dominio DI de NALCN (residuos F308-K314) con el mismo cambio conformacional de hélice α a π cuando el poro de NALCN se abre, estabilizado por enlaces intra-hélice en los cuatro segmentos S6 (F308-V313, L309-K314 en DI) e inter-hélice (W311- L1439, L588- K314, M1145- L591 y Y1436- L1148). Además, en este mismo estudio se describe otra región de residuos hidrofóbicos conservados (I318-V595-V1152-V1443 y I322-L599-I1156-V1447) que también participan en la apertura del poro de NALCN (Xie et al., 2020) (Figura 16).

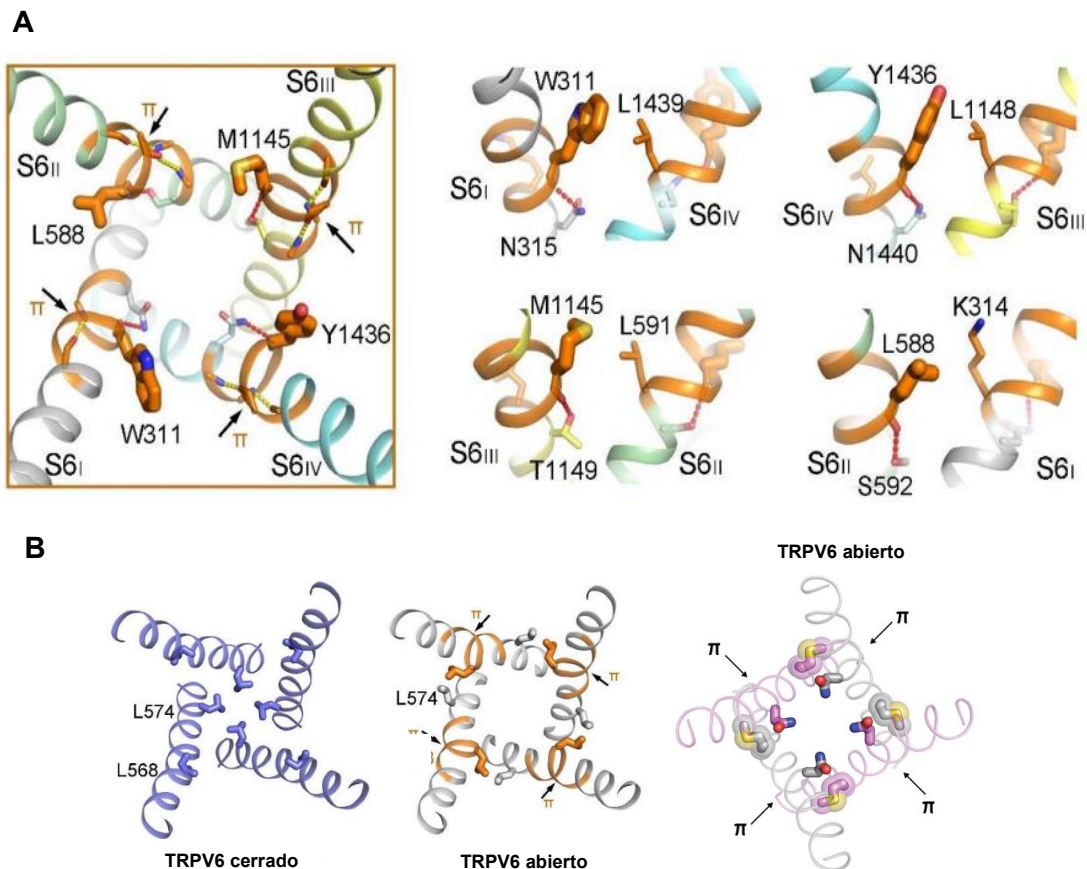


Figura 16. Estado abierto y cerrado de NALCN y del canal transitorio selectivo de la subfamilia vanilloide (TRPV 6). (A) Representación molecular de la probable estructura del canal abierto de NALCN, donde la flecha señala la estructura π (en anaranjado) en los S6_{I,II,III,IV} (B) Los segmentos que conforman el poro de TRPV 6 sufren transiciones de hélice α (en azul para TRPV6) a π (en anaranjado para TRPV6), donde la inducción de la estructura π en los S6 puede abrir los canales tras la unión del ligando. Potencialmente, una transición S6 análoga puede dilatar la puerta S6 en NALCN para iniciar la conductancia iónica. Tomado de Xie et al., 2020 y McGoldrick et al., 2018..

Enfermedades asociadas a NALCN

NALCN es una proteína esencial en mamíferos que regula el potencial de membrana neuronal en reposo, y mutaciones en su gen afectan significativamente a sus funciones provocando diversas enfermedades (Lu et al., 2007; Nakayama et al., 2006). Así, mutaciones en NALCN se han relacionado con distrofia neuroaxonal infantil, en concreto la conversión de C a T en el exón 16 de NALCN (c.1924C > T), lo que crea una señal de terminación de la traducción prematura en el codón 642 (Q642Stop) y por ende un canal NALCN truncado (Koroglu et al., 2013; Seven et al., 2002). En personas con síndrome autosómico recesivo caracterizado por hipotonía severa, deterioro del habla y retraso cognitivo, se ha encontrado una delección ubicada en la región que codifica la hélice S4 del dominio DII, y en otros pacientes una mutación que genera el cambio W1287L en la hélice S4 del dominio DIV (Lu et al., 2007). Además, se ha observado que individuos con epilepsia presentan delecciones en regiones del gen NALCN (Kirchhoff et al., 2009; Lalani et al., 2013), y se han identificado mutaciones en NALCN asociadas con cáncer de pulmón (Lee et al., 2013), y la metástasis del cáncer (Rahrmann et al., 2022).

NALCN y el síndrome de CLIFAHDD.

El síndrome CLIFAHDD (*Congenital contractures of the limbs and face, hypotonia, and developmental delay*), cuyo número MIM (*Mendelian Inheritance in Man*) es 616266 (<https://www.omim.org/entry/616266>), es un trastorno causado por mutaciones heterocigotas en el gen NALCN. Todos los casos notificados con el síndrome de CLIFAHDD han ocurrido *de novo* (Chong et al., 2015), es decir, no son fruto de herencia congénita. Todas estas sustituciones descritas estaban ausentes en las bases de datos del Proyecto Secuenciación del Exoma, Proyecto 1000 Genomas y Consorcio de Agregación del Exoma (Chong et al., 2015). De todas estas mutaciones registradas asociadas al fenotipo CLIFAHDD, en la presente tesis estudiamos en detalle 5 de ellas localizadas en el dominio DI (Figura 17). El resto de las mutaciones asociadas al síndrome de CLIFAHDD se indican en la Tabla 3.

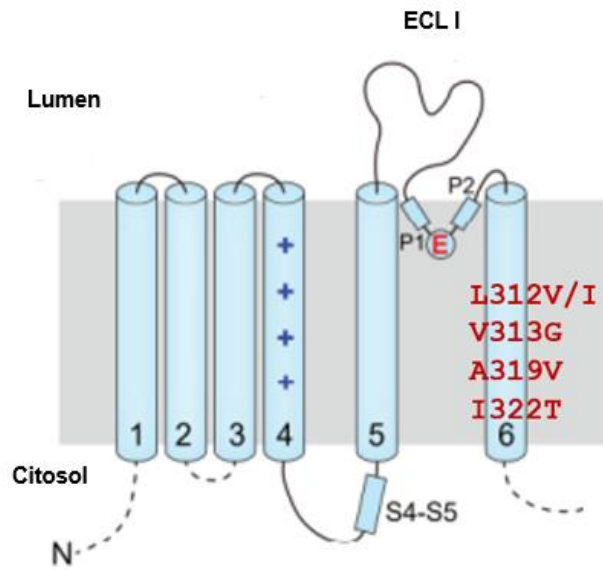


Figura 17. Mutaciones en el DI de NALCN estudiadas en la presente tesis. Las mutaciones estudiadas en esta tesis se muestran en rojo. El esquema muestra también las cuatro cargas (+) del segmento TM S4 y el ECL I del DI donde se muestra un lazo re-entrante (*reentrant loop*) y dos hélices (P1 y P2) flanqueando al residuo de glutámico (E) del filtro selectivo.

Tabla 3. Lista de mutaciones en la proteína NALCN que generan un espectro fenotípico CLIFAHDD.

Mutación	Aminoácido alterado	Ubicación en NALCN	Referencia
c.191A>G	Tyr64Lys	S1-S2 DI	Pergande et al., 2020
c.530A>C	Gln177Pro	S5 DI	Chong et al., 2015
c.934C>A	Leu312Ile	S6 DI	Chong et al., 2015
c.934C>G	Leu312Val	S6 DI	Fukai et al., 2016
c.938T>G	Val313Gly	S6 DI	Chong et al., 2015
c.950T>G	Phe317Cys	S6 DI	Karakaya et al., 2016
c.950T>G	Phe317Lys	S6 DI	Pergande et al., 2020
c.956C>T	Ala319 Val	S6 DI	Lozic et al., 2016
C.965T>C	Ile322Thr	S6 DI	Sivaraman et al., 2016
c.979G>A	Glu327Lys	S6 DI	Chong et al., 2015
c.985A>G	Arg329 Gly	DI-DII	Stark et al., 2016
c.1526T>C	Leu509Ser	S5 DII	Chong et al., 2015
c.1534T>G	Phe512Val	S5 DII	Chong et al., 2015
c.1538C>A	Thr513Asn	S5 DII	Chong et al., 2015
c.1571G>A	Ser524Asn	S5 DII	Angius et al., 2019
c.1640T>C	Met547Thr	S5-S6 DII	Kumaki et al., 2022
c.1733A>G	Tyr578Cys	S6 DII	Vivero et al., 2017
c.1733A>C	Tyr578Ser	S6 DII	Chong et al., 2015
c.1745A>C	Tyr582Ser	S6 DII	Vivero et al., 2017
c.1768C>T	Leu590Phe	S6 DII	Chong et al., 2015; Bend et al., 2016
c.1783G>T	Val595Phe	S6 DII	Karakaya et al., 2016; Pergande et al., 2020
c.1789G>A	Val597Ile	S6 DII	Wang et al., 2016
c.1800C>A	Asp600Glu	S6 DII	Angius et al., 2019
c.1870G>C	Glu603Gln	S6 DII	Winczewska-Wiktor et al., 2022
c.1838A>G	Gln613Arg	DII-DIII	Singh et al. 2022
c.3017T>C	Val1006Ala	S4-S5 DIII	Chong et al., 2015
c.3016G>T	Val1006 Phe	S4-S5 DIII	Li et al., 2020
c.3050T>C	Ile1017Thr	S5 DIII	Chong et al., 2015
c.3058G>T	Val1020Phe	S5 DIII	Fukai et al., 2016
c.3448C>G	Leu1150Val	DIII-DIV	Vivero et al., 2017
c.3493A>C	Thr1165Pro	DIII-DIV	Chong et al., 2015
c.3542G>A	Arg1181 Gln	DIII-DIV	Aoyagi et al., 2015, Chong et al., 2015
c.3553G>A	Ala1185Thr	DIII-DIV	Vissers et al., 2017
c.4205C>T	Leu1324Phe	S4 DIV	Byrne et al., 2023
c.4300A>G	Ile1434Val	S6 DIV	Liao et al., 2022
c.4333A>T	Ile1445Leu	S6 DIV	Bramswig et al., 2018
c.4338T>G	Ile1446Met	S6 DIV	Chong et al., 2015

I.6. Virus SARS-CoV-2

El virus SARS-CoV-2 (**Severe Acute Respiratory Syndrome Coronavirus 2**), es el agente responsable de la enfermedad COVID-19. Se trata de un virus de RNA de cadena simple perteneciente a la familia de los coronavirus que fue identificado por primera vez en la ciudad china de Wuhan a finales de 2019. La Organización Mundial de la Salud (OMS) declaró la propagación del virus como una pandemia en marzo de 2020 (Wu et al., 2020). El SARS-CoV-2 se propaga principalmente a través de aerosoles y afecta principalmente al sistema respiratorio, provocando síntomas que varían en gravedad. La enfermedad registra más de 770 millones de infecciones y más de 7 millones de muertes a principios de 2024 (<https://covid19.who.int/>), además de innumerables hospitalizaciones masivas, pérdidas económicas, despidos laborales, cierre de escuelas y universidades, etc. Las estrategias efectivas para disminuir la incidencia de la enfermedad fueron el distanciamiento social, el uso de mascarillas y el confinamiento en áreas gravemente afectadas y, sin lugar a duda, el desarrollo de vacunas. Sin embargo, la comprensión de la biología del virus es fundamental para controlar su expansión y proteger la salud pública a nivel mundial, así como prepararnos para posibles brotes futuros de virus similares (Chen et al., 2020; Dong et al., 2020; Li et al., 2020; Johns Hopkins Coronavirus Resource Center. COVID-19 dashboard. JHU. 2021. <https://coronavirus.jhu.edu/map.html>).

Genoma y proteoma del SARS-CoV-2. El SARS-CoV-2 es un virus del género *Betacoronavirus* de la familia *Coronaviridae*. Los coronavirus son virus con envuelta lipídica con un genoma de RNA monocatenario de polaridad positiva. El SARS-CoV-2 tiene uno de los genomas más grandes (29,9 kb), comparado con otros virus de RNA (Zhu et al., 2020; Wu et al., 2020). La estructura del genoma viral es similar al de otros coronavirus conocidos, con las dos primeras pautas de lectura abierta en el extremo 5' del genoma, ORF1a y ORF1ab (Figura 18), que codifican para dos poliproteínas (pp1a y pp1ab) que una vez procesadas darán lugar a 16 proteínas no estructurales (nsps) (Khan et al., 2023). Las nsps del SARS-CoV-2 están involucradas en la inhibición de la inmunidad innata y en distintas fases del ciclo vital del virus (Thomas et al., 2021; Kakavandi et al., 2023; Lu et al., 2022), en el que cada nsp desempeña funciones específicas.

NSP1 inhibe la traducción de las proteínas celulares, uniéndose a la subunidad 40S ribosomal y obstruyendo la entrada de mRNAs del huésped (Lapointe et al., 2021; Schubert K. et al., 2020). NSP2 participa en la inhibición de la expresión del mRNA del huésped, reduciendo la respuesta antiviral y facilitando la replicación del SARS-CoV-2.

Además, también participa en la transcripción y traducción viral interactuando con ribosomas y con complejos de replicación-transcripción (CRT) (Xu et al., 2022; Gupta et al., 2021). La proteína NSP3 es una proteasa similar a la papaína (PLpro) que escinde y genera las nsps 1-3 maduras. Se ha comprobado que participa junto con NSP4 en el ensamblaje de vesículas citoplasmáticas de doble membrana (DMV), necesarias para la replicación viral (Ricciardi et al., 2022). NSP3, también interfiere en la inducción de la respuesta inmune innata del interferón al bloquear la fosforilación, dimerización y posterior translocación nuclear del factor regulador de interferón 3 (IRF3) del huésped y previene la señalización NF-kappa-B del huésped (Lei et al., 2020). NSP4 es una proteína de membrana que desempeña un papel relevante en la formación y mantenimiento de los orgánulos de replicación o DMV junto a NSP3 y NSP6 (Ricciardi et al., 2022). La NSP5, o proteasa principal (Mpro), es una proteasa similar a la quimotripsina, escinde la región C-terminal de las poliproteínas pp1a y pp1ab en 11 sitios y origina las nsps 4-16 maduras (Dai et al., 2020; Shin et al., 2020). Además, puede activar en células epiteliales pulmonares la vía del inflammasoma NLRP1 y desencadenar la muerte celular epitelial (Planés et al 2022). NSP6 es una proteína de membrana que desempeña un papel en la formación y mantenimiento de orgánulos de replicación de DMVs (Oostra et al., 2008; Ricciardi et al., 2022). Esta proteína se une a la proteína quinasa serina/treonina TBK1 del huésped ocasionando una disminución en la fosforilación de IRF3, lo que conduce a una reducción de la producción de interferón IFN- β (Xia et al, 2020). NSP7 y NSP8 forman hexadecámeros que podrían participar en la replicación viral. Se ha demostrado también que ambas proteínas se ensamblan con NSP12 para mejorar la actividad de la RNA polimerasa viral, y conjuntamente NSP7 y NSP16 participan en la formación del complejo de replicación y transcripción (RTC) viral para regular la replicación y transcripción viral (Gao et al 2020; Wang et al 2020). Cabe destacar, que la proteína NSP8 interactúa con la SRP y junto con NSP9 suprimen la integración de proteínas en la membrana celular, afectando a las defensas inmunes del huésped (Banerjee et al., 2020). NSP10 es un cofactor de NSP14 y NSP16 (Yan et al., 2021; Jiang et al., 2021) y tiene un papel principal en la metilación de los mRNAs virales (Park et al., 2022). NSP12 es la RNA polimerasa RNA dependiente que cataliza la transcripción de los RNAs genómicos y subgenómicos del virus (Yin et al., 2021). NSP13 tiene actividad helicasa y se une a la proteína TBK1 del huésped e inhibe su fosforilación, lo que genera una reducción del interferón IFN- β (Xia et al, 2020). NSP14 es una exorribonucleasa correctora de errores que reduce la sensibilidad del virus a los mutágenos de RNA durante la replicación, además tiene una función N7-guanina metiltransferasa originando GpppA-RNA (Lin et al., 2021; Park et al., 2022). NSP15 tiene

actividad endorribonucleasa y actúa degradando las 5'-poliuridinas generadas durante la replicación de la región poli(A) de los RNAs genómicos y subgenómicos virales, lo que evita que el RNA poli(U) hibride con las colas poli(A) y active los sensores de RNA de doble cadena (dsRNA) del huésped (Kim et al., 2020; Kim et al., 2021; Pillon et al., 2021). Finalmente, NSP16 tiene actividad 2'-O-metiltransferasa que media en el proceso de protección del mRNA (Rosas-Lemus et al., 2020) a través de la metilación de los mRNA virales, lo cual es importante para evadir el sistema inmunológico (Park et al., 2022). Además, esta proteína interrumpe el *splicing* de los mRNAs del huésped en el núcleo celular al interactuar con los dominios de reconocimiento de pre-mRNA de los RNAs nucleares pequeños (snRNA) U1 y U2 (Banerjee et al., 2020).

En el extremo 3' del genoma, que ocupa aproximadamente un tercio del genoma, se encuentran las pautas de lectura abiertas (ORF) que codifican las proteínas estructurales intercaladas con varias ORFs que codifican proteínas accesorias, incluidas las ORFs 3a, 3b, 6, 7a, 7b, 8, 9b, 9c y 10 (Figura 18). Las proteínas estructurales del SARS-CoV-2 incluyen la glicoproteína de la espícula (S o *spike*) que se une al receptor del huésped y facilita la fusión de membranas (Moreira et al., 2020; Huang et al., 2020); la glicoproteína M (*membrane*) que es la mayoritaria en la envuelta viral y está asociada con la proteína de la nucleocápside y otras proteínas estructurales virales para facilitar el ensamblaje viral, además de participar en el proceso de patogénesis (Fu et al., 2021); la proteína E (*envelope*) que es la proteína de membrana más pequeña (Duart et al., 2020 y 2021) y forma un canal iónico que promueve el ensamblaje y la patogénesis del virus (Nieto-Torres et al., 2014; Schoeman et al., 2019); y la proteína soluble N (*nucleocapsid*) que facilita el empaquetado del genoma viral en complejos ribonucleoproteicos (RNPs) protegidos en el interior del virus y promueve el ensamblaje del virión (Sheikh et al., 2020; Zeng et al., 2020). Profundizar en la comprensión estructural y funcional tanto de estas proteínas estructurales como de las nsps es de particular interés para comprender mejor la patogénesis del SARS-CoV-2 (Gordon et al., 2020; Yoshimoto, 2020).

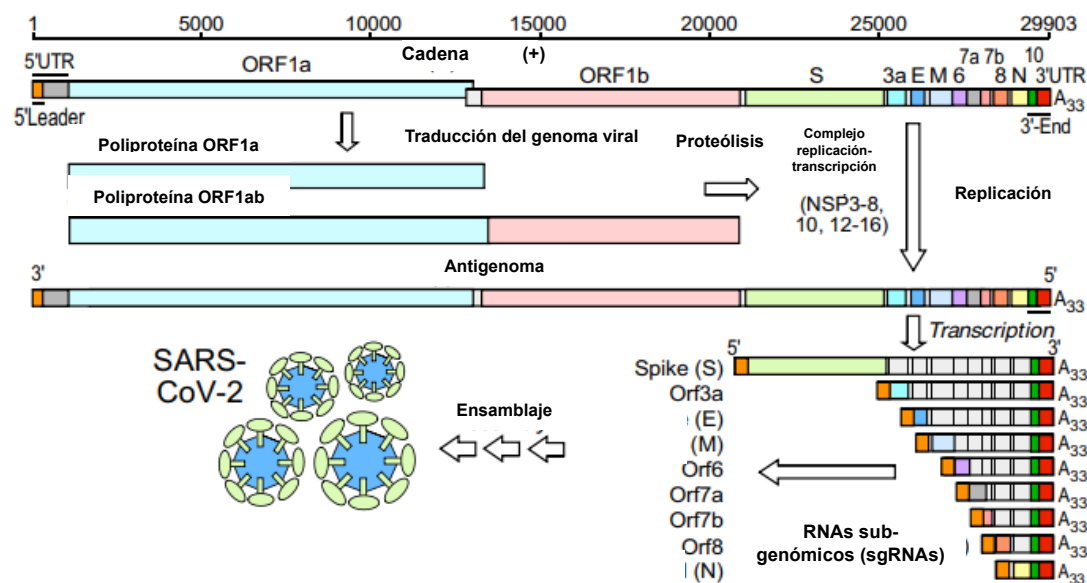


Figura 18. Estructura del genoma y proteoma de SARS-CoV-2. A partir de las dos primeras ORFs 1a y 1b se sintetizan las poliproteínas pp1a y pp1ab. La proteólisis de estas poliproteínas da lugar a las proteínas no estructurales (nsps), que promueven la formación del complejo de replicación-transcripción (CRT) y las vesículas de doble membrana (DMV), donde ocurre la replicación del genoma viral. En estas estructuras se sintetiza un genoma viral de sentido negativo (antigenoma) que servirá como molde para la transcripción de los RNAs subgenómicos (sgRNA) y del RNA viral de sentido positivo. A partir de los sgRNA se traducirán las proteínas estructurales (S, E, M, N) y otras proteínas accesorias. Adaptado de Khan et al., 2023.

Ciclo infeccioso del SARS-CoV-2. Los coronavirus utilizan factores del huésped y estrategias propias para la infección y expresión génica, tanto a nivel transcripcional como traduccional. Los principales factores que participan en la entrada del SARS-CoV-2 a la célula huésped, son la glicoproteína viral *spike* (S) y el receptor de la enzima convertidora de angiotensina 2 (ACE2) de la célula huésped (V'kovski et al., 2021). Cuando la proteína S se une al receptor ACE2 es hidrolizada por la proteasa celular furina dividiéndola en dos fragmentos S1 y S2, posteriormente la serin-proteasa transmembrana 2 (TMPRSS2) escinde el fragmento S2 viral provocando un cambio conformacional de éste que permite la exposición de una región hidrofóbica conocida como péptido de fusión. La exposición del péptido de fusión facilita la interacción con la membrana celular lo que resulta en la fusión de las membranas virales y celulares. Este proceso permite la liberación del genoma viral al interior celular (Ou et al., 2020). El genoma viral del SARS-CoV-2, posee polaridad positiva, es decir, es reconocido por los ribosomas celulares y actúa directamente como RNA mensajero. Así, se traducen las ORFs del extremo 5' generándose las poliproteínas pp1a y pp1ab que tienen 4.405 y 7.096 aminoácidos de longitud, respectivamente. Cabe señalar que la producción de

pp1ab depende de la aparición de un cambio de pauta de lectura del ribosoma, lo que ocurre aproximadamente en el 60% de los casos, que elimina la aparición de un codón de parada y amplía la región codificante. Esto genera que la pp1a codifique 11 nsps (NSP1-NSP11), y la poliproteína pp1ab codifique 16 nsps (NSP 1- NSP 16) (Wu et al., 2020; Whittaker et al., 2021). Como se ha mencionado anteriormente, las poliproteínas son procesadas por las proteasas víricas PLpro (NSP3) y Mpro (NSP5). Algunas de las nsps se ensamblan formando el complejo de replicación-transcripción (CRT) responsable de la replicación y transcripción del genoma viral (Malone et al., 2022). Al mismo tiempo se forman las vesículas de doble membrana (DMVs) mediante la reorganización de membranas del RE por las IMPs NSP3, NSP4 y NSP6, desempeñando un papel clave en la regulación de la inmunidad del huésped y la modificación de los orgánulos celulares para la replicación y salida de los viriones (Lu et al., 2022; Sakai et al., 2017). Las DMVs ofrecen el ambiente favorable para la síntesis de RNA y a su vez lo protegen de la respuesta inmunitaria activada por los dsRNA. De esta manera, a partir del RNA genómico (RNAg) se sintetiza en los CRT una cadena de RNA complementaria llamado antígenoma y un conjunto de RNA-subgenómicos (RNAsg) de polaridad negativa. El antígenoma sirve como molde para producir nuevo RNAg viral, mientras que los RNAsg servirán para la expresión de las cuatro proteínas estructurales necesarias para el ensamblaje y salida del virión, y otras proteínas accesorias implicadas en la modulación de la respuesta inmune innata del huésped. El ensamblaje del SARS-CoV-2 comienza con el recubrimiento de los RNAg con proteínas de la nucleocápside (N) generando estructuras que emergen del ERGIC (RE-compartimento intermedio de Golgi), adquiriendo así una bicapa lipídica que contiene las proteínas S, M y E del virus (Figura 19) (Malone et al., 2022). Finalmente, los viriones son secretados de la célula infectada mediante exocitosis.

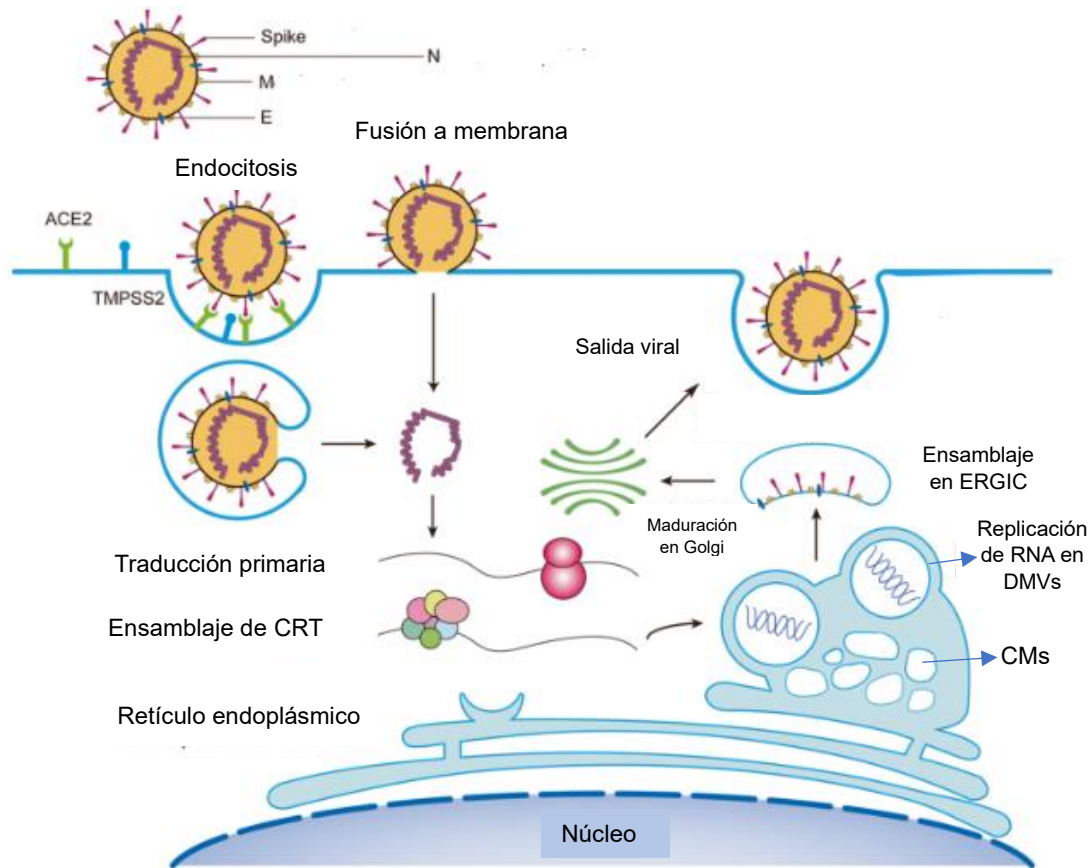


Figura 19. Ciclo infeccioso del SARS-CoV-2. La entrada del virus se realiza tras el reconocimiento de la proteína S por el receptor vía endocitosis o fusión de membranas, posteriormente tiene lugar la traducción de las poliproteínas pp1a y pp1ab que después de ser procesadas por NSP3 y NSP5 producirán 16 proteínas no estructurales (nsp) de las cuales algunas se ensamblan formando el complejo de replicación-transcripción (CRT). La replicación del RNA viral ocurre en DMVs, el ensamblaje de viriones en ERGIC, la maduración del virión en Golgi y finalmente la salida de los viriones se produce por exocitosis. S, proteína de la espícula; M, proteína de membrana; E, proteína de envoltura; N, proteína de nucleocápside; TMPRSS2, serina proteasa transmembrana 2; ERGIC, retículo endoplasmático-compartimento intermedio de Golgi; CMs, membranas contorneadas. Adaptado de Chen et al., 2023.

Proteína no estructural 4 (NSP4) del virus SARS-CoV-2

NSP4 es la proteína no estructural más grande y la segunda proteína integral de membrana de mayor tamaño del SARS-CoV-2 después de la proteína S. Como se ha mencionado, NSP4 es clave para la formación de DMVs (Angelini et al., 2013; Oudshoorn, et al., 2017; Sakai et al., 2017; Wolff et al., 2020), permitiendo su interacción con NSP3 y NSP6 para la formación de estas vesículas (Hagemeijer et al., 2011). La proteína NSP3 se encarga inicialmente de la deformación de la membrana del RE, necesaria para hacer posible la formación de las DMVs. Posteriormente, la interacción con NSP4 genera la reorganización de las membranas del RE, y la subsiguiente interacción entre las proteínas NSP4 y NSP6 permite la formación de las DMVs (Figura

20A). Estas interacciones son previas al corte proteolítico de la poliproteína, lo que permite mantener cerca las 3 proteínas asociadas a la membrana para facilitar su interacción (Angelini et al., 2013; Harcourt et al., 2004; Sawicki et al., 2005). La síntesis del RNA viral se lleva a cabo en las DMVs, que actúan como centros neurálgicos, proporcionando un microambiente favorable y posiblemente protegiendo al RNA viral de los sensores inmunitarios innatos de la célula huésped. El RNA viral recién sintetizado se exporta al citosol a través de un poro que atraviesa las DMVs (Figura 20A). Los componentes de este poro incluyen NSP3, NSP4 y NSP6 (Wolff et al., 2020).

La proteína NSP4 está presente en otros coronavirus y se ha demostrado su importancia en el fenotipo y replicación viral. Así, en el coronavirus de ratón (MHV) se encontró que las sustituciones de aminoácidos (N176A y N258T) en la proteína NSP4 originaron un fenotipo viral sensible a la temperatura, y el análisis de las células infectadas con este clon infectivo reveló una disminución drástica de las DMVs y del título viral, demostrando la importancia de NSP4 en la dirección del ensamblaje de las DMVs y en la replicación viral (Clementz et al., 2008). Por otro lado, en cultivos celulares con los mutantes de NSP4 N176A y N237A del MHV, se encontró que había una disminución en la síntesis de RNA viral, lo cual sugiere que la disminución de RNA viral es la responsable de los defectos de producción de viriones. Además, por microscopía electrónica se observó que las células infectadas con estos mutantes tenían DMVs con morfología alterada o aberrante en comparación con las células infectadas con MHV de tipo silvestre, destacando que el grado de DMV irregulares en las células infectadas con el virus que contenía la NSP4 mutante está directamente relacionado con la alteración en la síntesis de RNA y el desarrollo viral (Gadlage et al., 2010). Ambos estudios fortalecen el papel de NSP4 en la organización y estabilidad de las DMV, que son esenciales para la síntesis eficiente del RNA viral y la replicación óptima del coronavirus MHV. En el caso del coronavirus causante del síndrome respiratorio de oriente medio (MERS), se ha demostrado que la expresión de NSP3 y NSP4 es suficiente para inducir la formación de DMVs (Oudshoorn et al., 2017). Estos resultados fueron corroborados en el virus SARS-CoV, descubriéndose que la interacción de NSP4 con NSP3 es necesaria y suficiente para inducir la formación de DMVs. En este último caso se demostró que los residuos implicados en la interacción con la NSP3 son la H120 y la F121 de NSP4; ya que mutantes H120N y F121L alteraron la interacción de ambas proteínas afectando los reordenamientos de la membrana del RE y, por ende, la replicación del RNA viral (Sakai et al., 2017). Todo lo anterior sugiere que NSP4 juega un papel en la replicación, organización y estabilidad de las DMVs, siendo esta proteína de especial interés para entender la patogénesis del virus. Por tal motivo, resulta importante investigar la

biogénesis de la proteína NSP4 del SARS-CoV-2 con el fin de comprender a nivel molecular su papel en la patología del virus y desarrollar estrategias eficaces para combatir este virus (Ghosh et al., 2020). Hasta el momento no se ha podido determinar la estructura con resolución a nivel atómico de NSP4 de ningún coronavirus. Las primeras caracterizaciones apuntan la presencia de 4 regiones TM tanto en el coronavirus de ratones (MHV) como en SARS-CoV humano, con los extremos N y C-terminal orientados al citoplasma (Oostra et al., 2007). Además, NSP4 de SARS-CoV contiene un sitio de glicosilación no canónico (NIC) ubicado entre la primera y la segunda región hidrofóbica que podría estar modificado. Por otro lado, se ha sugerido que el extremo N-terminal de NSP4 de SARS-CoV, pero no el de MHV, es parcialmente procesado por una peptidasa señal (Oostra et al., 2007). En el caso del SARS-CoV-2 no se ha demostrado experimentalmente la topología de NSP4 ni sus modificaciones co-/post-traduccionales, de hecho, aunque hay distintos modelos topológicos (Thomas, 2021), no hay un consenso entre los distintos algoritmos de predicción. Recientemente, se simuló el impacto de las hélices TM en la interacción entre los dominios luminales de NSP3 y NSP4 concluyendo que algunos modelos podrían ser irreales (Figura 20B), probablemente derivados de la ubicación inexacta de los segmentos TM (Klatte et al., 2022), siendo un ejemplo del impacto negativo que tiene nuestro limitado conocimiento de los detalles moleculares de NSP4 y, por ende, de su función en la propagación viral.

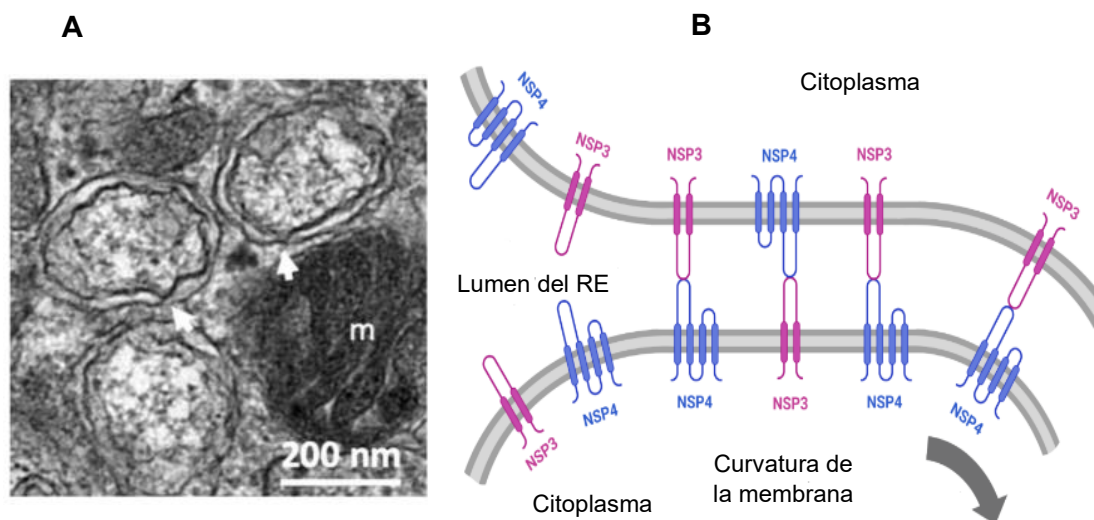


Figura 20. Morfología de las vesículas de doble membrana (DMVs). (A) Imagen de las DMVs obtenida por microscopía electrónica (TEM) 12 h.p.i (10000X) inducidas por la cepa Wuhan del SARS-CoV-2 en células Vero E6. El diámetro de estas vesículas es de 100-300 nm, las flechas indican los puntos de contacto entre las dos membranas, probablemente correspondiente a poros que permiten el intercambio con el citoplasma. M, mitocondria. Tomado de Roingeard et al., 2022. (B) Esquema ilustrativo que muestra la interacción entre las proteínas NSP3 (morado) y NSP4 (azul) para formar las DMVs. Tomado de Klatte et al., 2023).

OBJETIVOS

El objetivo general I: Estudio de la biogénesis del dominio DI del canal de sodio no selectivo humano (NALCN). Para ello se han abordado los siguientes objetivos específicos:

1. Estudio de la inserción individual en membrana de los 6 segmentos TM del dominio I (DI) de la proteína NALCN.
2. Determinación de la topología del DI de NALCN en membranas bacterianas y de células humanas.
3. Estudio computacional de las interacciones entre segmentos TM del DI de NALCN.
4. Posicionamiento del S6 del DI de NALCN en membranas y efecto de las mutaciones asociadas a la enfermedad CLIFAHDD en su integración en membranas.

El objetivo general II: Estudio topológico de la proteína no estructural 4 (NSP4) del SARS-CoV-2. Para ello se han abordado los siguientes objetivos específicos:

1. Determinación de la topología de la proteína NSP4 en membranas de células bacterianas y en membranas de células humanas.
2. Verificar si la primera región hidrofóbica de la proteína NSP4 puede ser sustrato del SPC y de su posible procesamiento en otras líneas celulares.
3. Estudios preliminares para la caracterización estructural de dominios solubles de la proteína NSP4.

MATERIALES Y MÉTODOS

M.1. Análisis computacional de las secuencias proteicas

Predicción de hélices TM y topología. Para la predicción de hélices TM y topología del Dominio I (DI) de NALCN y NSP4, utilizamos los programas computacionales accesibles en la red especificados en la tabla 4 (Duart et al., 2024).

Tabla 4. Programas computacionales de predicción de segmentos TM, topología y péptido señal (SP) disponibles en la web.

Programa	Predicción	Enlace web
ΔG predictor	TM	http://www.cbr.su.se/DGpred/
TMHMM	TM/Topología	http://www.cbs.dtu.dk/services/TMHMM/
HMMTOP	TM/Topología	http://www.enzim.hu/hmmtop/
TOPCONS	TM/Topología/SP	http://topcons.net/
MEMSAT	TM/Topología/SP	http://bioinf.cs.ucl.ac.uk/psipred/
SPOCTOPUS	TM/Topología/SP	http://octopus.cbr.su.se/
Protter	TM/Topología	http://ulo.github.io/Protter/
Uniprot	TM	https://www.uniprot.org/

Simulaciones por dinámica molecular. Para la realización de los cálculos computacionales en las simulaciones de dinámica molecular del dominio DI de NALCN, utilizamos los programas/software de acceso libre en la red indicados en la tabla 5.

Tabla 5. Programas computacionales de diseño, simulación, análisis y procesamiento de trayectorias y datos utilizados en dinámica molecular.

Programa	Enlace web/Referencia
CHARMM-GUI	http://www.charmm-gui.org/
GROMACS 2023.2	http://www.gromacs.org/
Dinámica Molecular Visual (VMD) 1.9.4	https://www.ks.uiuc.edu/Research/vmd/vmd-1.9.4/
Phyton 3.11.0	https://www.python.org/downloads/release/python-3110/
JupyterLab	https://jupyter.org/

En las simulaciones por Dinámica Molecular (DM) usamos una representación a nivel atómico, en la que la posición y orientación del modelo del dominio DI de NALCN se obtuvo del Banco de Datos de Proteínas (PDB) (<https://www.rcsb.org>) con código PDB (7cm3). Usando el algoritmo CHARMM-GUI (<http://www.charmm-gui.org/>), insertamos la estructura proteica resultante en una bicapa de POPC (16:0–18:1 PC) preequilibrada siguiendo un modelo simplificado de membrana de RE que abarca el plano XY del cuadro de simulación periódica. El complejo proteína-membrana se solató con agua

TIP3P que contenía KCl 150 mM (Jo et al., 2008 y 2009). A continuación, procedimos a la simulación de todo el DI de NALCN utilizando GROMACS (versión 2023.2). Primero realizamos una minimización de energía. A continuación, se realizaron seis rondas cortas de equilibración con restricciones de posición en los átomos de la proteína. A partir de la producción se realizaron 3 réplicas. Los sistemas se simularon con un paso de tiempo de 2 fs durante 1 μ s utilizando el campo de fuerza CHARMM36m (Huang et al., 2017). La temperatura y la presión del sistema se mantuvieron a 310 K y 1 atm utilizando el termostato v-rescale y el barostato c-rescale (Bussi et al., 2007; Bernetti et al., 2020). El análisis de las simulaciones fue configurado con Python 3.11.0 en la interfase de usuario Jupyter Lab, y para la visualización de las trayectorias y diseño de imágenes de las estructuras simuladas utilizamos VMD (Humphrey et al., 1996). Para la estadística descriptiva se utilizaron los valores cuantitativos (n=2700) de cada *frame* que fueron analizados en GraphPad Prism 8.0.2.

M.2. Manipulación de bacterias.

Cepas bacterianas: Empleamos la cepa de *E. coli* DH5 α (genotipo: F– ϕ 80lacZ Δ M15 Δ (lacZYA-argF) U169 recA1 endA1 hsdR17(rK–, mK+) phoA supE44 λ -thi-1 gyrA96 relA1) para procedimientos de clonación, transformación y amplificación de DNA plasmídico. Asimismo, esta cepa bacteriana también fue utilizada en los ensayos de topología de las proteínas DI NALCN y NSP4 utilizando el reportero dual PhoA-LacZ α ; ya que permite la α -complementación de LacZ y es deficiente de phoA. También, utilizamos la cepa bacteriana *E. coli* C43(DE3) (genotipo: F – ompT hsdSB (rB- mB-) gal dcm (DE3)) para la expresión de fragmentos solubles de la proteína NSP4. Las células competentes de *E. coli* DH5 α y *E. coli* C43(DE3) se prepararon en el laboratorio siguiendo el protocolo de Green y Sambrook, 2020.

Cultivo de cepas de *E. coli*: Las cepas se cultivaron a 37°C y 200 rpm en medio LB (Luria Bertani) líquido (triptona 1% (p/v) (Pronadisa), extracto de levadura 0.5% (p/v) (Pronadisa), NaCl 1% (p/v)) o sólido, añadiendo 2% (p/v) de agar bacteriológico (Pronadisa). En ambos casos se esterilizaron los medios en autoclave durante 20 minutos a 1 atmósfera de presión y 121°C. Los medios LB líquido o sólido se suplementaron con el antibiótico apropiado para el caso, ampicilina a 50 μ g/ml o kanamicina a 35 μ g/ml.

Transformación de las cepas de *E. coli*: Para la incorporación de DNA plasmídico exógeno dentro de estas cepas utilizamos el método de electroporación. Para ello 50 ó 100 μ L de células electrocompetentes se mezclaron con 5 ng de DNA plasmídico,

después las células fueron sometidas a un choque eléctrico de 1.8 kV/cm² durante 5 milisegundos, para lo cual se emplearon cubetas de electroporación con una distancia entre los electrodos de 0,1 cm. Después del choque eléctrico las células se recuperaron en medio LB incubándolas 1 hora a 37°C en agitación a 350 rpm. Finalmente, las células fueron sembradas en una placa de LB con el antibiótico adecuado.

M.3. Manipulación de líneas celulares humanas

Líneas celulares humanas: En la presente tesis empleamos 2 líneas celulares humanas y una línea celular de mono que crecen en monocapa adheridas a la superficie del frasco de cultivo. La línea HEK-293T (células embrionarias de riñón humano 293) fueron usadas en ensayos de topología de las proteínas DI NALCN y NSP4, inserción individual de segmentos TM en membrana del RE, verificación del residuo inicial del TM del S6 del DI de NALCN, efectos de mutaciones asociadas a enfermedades en DI NALCN y en el procesamiento de la diana putativa de la proteína NSP4. La línea Vero E6 (células epiteliales de riñón extraídas del primate africano cercopiteco verde) y la línea A549 (células epiteliales del carcinoma de pulmón humano) se utilizaron para ensayos de procesamiento de la diana putativa del complejo peptidasa (SPC) de la proteína NSP4.

Mantenimiento de las líneas celulares humanas: Las 3 líneas celulares se cultivaron en monocapa en frascos de 75cm² recubiertos con medio de cultivo que contenía DMEM (Dulbecco's Modified Eagle Media) (Gibco) con 4.8 g/L de glucosa suplementado con 2 mM glutamina y 10% (v/v) de suero bovino fetal (Gibco) y penicilina/estreptomicina (100 U/ml) (Gibco). Cuando las células alcanzaban el 90% de confluencia se tripsinizaban y se pasaba una dilución a otros frascos conteniendo DMEM suplementado fresco con el objetivo de mantener la línea celular.

Transfección de líneas celulares humanas: 24 horas antes de la transfección se sembraban 2 millones de células por placa de 24 pocillos con 500 µL de medio de cultivo en cada pocillo. En general se transfectó 500ng de DNA plasmídico por pocillo. La línea celular HEK-293 se transfectó con una mezcla que contenía 1µL de polietilenimina 1mg/ml y 50µL de DMEM por cada µg de DNA. Las líneas celulares Vero E6 y A549 fueron transfectadas con Lipofectamina 3000 siguiendo el protocolo del fabricante (Invitrogen). Para la preparación, se combinaron 1.5 µL de lipofectamina 3000 con 25 µL de DMEM con una mezcla que contenía 1 µL del reactivo potenciador de la transfección (Invitrogen), 500 ng de DNA plasmídico y 25 µL de DMEM. En ambos casos, la mezcla final se incubó a temperatura ambiente durante 15 minutos y después se

agregó gota a gota al pocillo. Las células se cultivaron a 37°C y 5% CO₂ durante 24-48 horas según sea el caso, antes de proceder a la obtención del extracto proteico.

M.4. Manipulación genética

Vectores de DNA plasmídico: Durante el desarrollo de la tesis utilizamos 6 tipos de vectores (Tabla 6). En el plásmido pGEM-T se clonó la secuencia de DNA que codifica el dominio DI de NALCN y el sistema Lep; el plásmido pLVX-EF1alpha-IRES-Puro contenía la secuencia de DNA que codifica la proteína NSP4; en el plásmido pSVLd construimos el sistema C_LTM con tres y dos dianas de glicosilación (3G y 2G respectivamente); el plásmido pKTop contiene el sistema dual PhoA-LacZ; el plásmido pCAGGS se usó para la expresión de proteínas en células de mamífero; y el plásmido pETNKI se utilizó para expresión de fragmentos solubles de la proteína NSP4. Para obtener una mayor cantidad de copias de los plásmidos y de las construcciones derivadas, estos fueron transformados en *E. coli* DH5α y se realizó purificación del DNA plasmídico mediante lisis alcalina siguiendo las instrucciones de GeneJET Plasmid Miniprep (ThermoFisher Scientific).

Tabla 6. Lista de Vectores plasmídicos utilizados

Vectores DNA plasmídico	Origen	Referencia
pGEM-T	Donado por Dr. von Heijne	Hessa et al. 2005
pLVX-EF1alpha-IRES-Puro	Donado por Dr. Nevan Krogan	Gordon et al. 2020
pSVLd	Donado por el Dr. Feige	Behnke et al., 2016
pKTop	Donado por Dra. Karimova	Karimova et al., 2009
pCAGGS	Life Science Market PVT1020	Fan et al., 2022
pETNKI 1.1	Donado por Dra. Casino	Huesa et al., 2021

Amplificación de material genético: Las reacciones en cadena de la polimerasa (PCR) se realizaron en un termociclador Doppio Ristretto (Hamburgo, Alemania). Las mezclas de reacción contenían 50 ng de DNA molde, 10 μM de cada oligonucleótido, 0.25 mM de dNTPs, tampón de PCR comercial y 1 unidad de la polimerasa Taq (*Thermus aquaticus*) de Promega (Madison, WI, U.S.A.), para un volumen final de reacción de 50 μL. Las condiciones de la reacción fueron una desnaturalización inicial a 95°C, seguido de 32 ciclos de tres segmentos (1 min a 95°C, 1 min. a 55°C, 1 min/kb. a 72°C) y con un tiempo de extensión final de 5 min a 72°C. El DNA amplificado se purificó

siguiendo las instrucciones del GeneJET PCR Purification Kit (ThermoFisher Scientific) y GeneJET Gel Extraction (ThermoFisher Scientific).

Inserción y clonación de material genético: A continuación, describiremos las diferentes construcciones genéticas obtenidas mediante inserción, clonación molecular y mutagénesis para la realización de la presente Tesis. Para ello, el diseño de los oligonucleótidos fue realizado manualmente o utilizando la herramienta “Quik Change II Primer Design tool”, según sea el caso. Los oligonucleótidos utilizados y las construcciones obtenidas se describen en el Anexo 1 y 2, respectivamente.

- Construcción del sistema C_LTM: Primero, al plásmido pSVLd le realizamos algunas modificaciones. En la región codificante de la cadena ligera λ de inmunoglobulina (C_L) diseñamos dos sitios de glicosilación *aat agc agc* y *aat gtc act* que codifican las secuencias de aminoácidos NSS y NVT, respectivamente mediante mutagénesis dirigida; los cuales nos servirán como reporteros de glicosilación. Los sitios de digestión de las enzimas *SpeI* (*actagt*) y *SacI* (*gagctc*) fueron diseñados por inserción en la región que codifica la secuencia de glicinas y serinas repetidas (Figura 24). La etiqueta con el epítipo c-myc (EQKLISEEDL) y los dos codones de parada (*tag taa*), situados al final del sistema C_LTM, fueron añadidos por PCR y el producto amplificado fue clonado en el plásmido pSVLd utilizando las enzimas de restricción *XhoI* y *BamHI* (New England Biolabs NEB). Posteriormente, por PCR amplificamos los 6 segmentos TM del dominio DI de NALCN y el TM₁₀₋₃₂ de NSP4 (flanqueados por los tetrapéptidos GGPG/GPGG y por sitios de digestión); después, tanto los productos de PCR como el vector que contiene el sistema C_LTM fueron digeridos con las enzimas de restricción *SpeI/SacI* (New England Biolabs NEB) y tratados con la T4 DNA *ligase* (NZYTECH) a 16°C durante 16h. Los protocolos de clonación con enzimas de restricción (*XhoI*, *BamHI*, *SpeI* y *SacI*), se realizaron siguiendo las instrucciones del fabricante New England Biolabs (NEB) y NZYTECH. El sistema C_LTM y las construcciones realizadas en éste se muestran en Anexo 2 A.

- Construcciones de proteínas truncadas en el vector pKTop: En las construcciones realizadas en el sistema de reportero dual PhoA-LacZ, las diferentes secuencias que codifican los truncados de longitud creciente del dominio DI de NALCN y NSP4 fueron amplificadas por PCR. Tanto los productos de PCR como el plásmido pKTop fueron digeridos con *XbaI/SacI* siguiendo las instrucciones del fabricante NEB y tratados con T4 DNA *ligase* (NZYTECH) a 16 °C durante 16h. Las construcciones se muestran en el Anexo 2 B.

- **Construcciones de proteínas truncadas en el vector pCAGGS:** Para el mapeo de glicosilaciones de las proteínas truncadas del dominio DI de NALCN y NSP4, se diseñaron cebadores para amplificar por PCR los diferentes truncados. En todas las construcciones, los cebadores incluyen sitios glicosilables (NSTMGM) o no glicosilables (QSTMGM) en N- y/o C-terminal, etiqueta con el epítipo c-myc (EQKLISEEDL) sólo en el cebador C-terminal, y secuencia homóloga de 15pb del plásmido pCAGGS en ambos cebadores directo (*forward*) y reverso (*reverse*). Los productos de PCR se purificaron y se clonaron en el plásmido pCAGGS que fue linealizado por digestión con *KpnI* (NEB) utilizando la tecnología In-Fusion HD (Takara), siguiendo las instrucciones del fabricante. Las construcciones se muestran en el Anexo 2 C.

- **Construcciones en pGEM-T(Lep):** Para los estudios de inserción de las regiones hidrofóbicas seleccionadas se utilizó el ensayo basado en la proteína modelo *leader peptidase* (Lep), ésta fue modificada añadiéndole mediante mutagénesis dirigida sitios de glicosilación (NTS) a distancias de 9, 11, 13 y 15 codones del inicio de la secuencia testada. Al S6 del dominio DI de NALCN se le añadieron por PCR tetrapéptidos GGPG/GPGG en ambos extremos y sitios de restricción *SpeI/KpnI* flanqueantes. Tanto los productos de PCR del S6 como el plásmido pGEM+Lep fueron digeridos con *SpeI/KpnI* siguiendo las instrucciones del fabricante (NEB) y ligados con T4 DNA *ligase* (NZYTECH) a 16 °C durante 16h. Los controles de inserción (la secuencia KVLLVTGVLGLLLLIKWK), translocación (la secuencia YNLYIYAFFASSLVHGFI), S6 silvestre y los mutantes fueron clonados en el sistema pGEM+Lep por la misma metodología. En el caso de las mutaciones asociadas a la enfermedad CLIFAHDD, éstas fueron diseñadas en el S6 mediante mutagénesis dirigida siguiendo el procedimiento que se detalla más adelante. Las construcciones realizadas se describen en el Anexo 2 E.

- **Construcciones en el vector pETNKI 1.1:** Para la purificación de dominios solubles de la proteína NSP4, se realizaron dos construcciones, una que comprende la región entre el primer y segundo TM (S40-P270) y la otra que corresponde al extremo C-terminal de la proteína (T411-Q500). Los fragmentos correspondientes a estas regiones de la proteína fueron amplificados por PCR y clonados en el plásmido pETNKI 1.1 (resistente a kanamicina, con etiqueta Histag en N-terminal seguido de sitio de corte para proteasa PreScission) mediante ligación independiente de ligasa, una técnica donde se crean extremos sobresalientes complementarios entre las secuencias de nucleótidos del plásmido y del fragmento amplificado. Para ello, el plásmido digerido con *KpnI* y los fragmentos amplificados por PCR fueron purificados por GeneJET Gel

Extraction (ThermoFisher Scientific). Luego, para crear los extremos sobresalientes en el plásmido preparamos la siguiente reacción (5 µL de plásmido digerido, 2 µL T4-DNAPol buffer 5X, 1 µL de 25 mM dTTP, 0.5 µL 100 mM DTT, 1 µL BSA (1 mg/ml), 0.5 µL de la DNA polimerasa T4 (NEB) y 2.5 µL H₂O); y para los fragmentos de PCR (5 µL de producto de PCR, 2 µL T4-DNAPol buffer 5X, 1 µL de 25 mM dATP, 0.5 µL 100 mM DTT, 1 µL BSA (1 mg/ml), 0.5 µL de la DNA polimerasa T4 (NEB) y 2.5 µL H₂O). Las reacciones fueron incubadas durante 30 minutos a 22 °C seguido de 20 minutos a 75 °C para inactivar la polimerasa. En la reacción de hibridación mezclamos 1.5 µL de plásmido y 3 µL de inserto tratados con la DNA polimerasa T4, incubamos 10 minutos a temperatura ambiente y transferimos la reacción a 4°C. Finalmente, transformamos 3 µL de cada construcción en 50 µL de células *E. coli* DH5α competentes y sembramos en placas de LB agar con kanamicina 35µg/mL para luego incubarlas a 37°C por toda la noche. Las construcciones se describen en el Anexo 2D.

Mutagénesis dirigida: Para la reacción de mutagénesis seguimos las instrucciones del fabricante (Agilent Technologies). En cada reacción usamos 1 µL de Pfu Turbo DNA polimerasa, 50 ng de DNA molde, 5 µM dNTPs, 5 µL de tampón 10X, 0.05µM de concentración final de cada oligonucleótido que contiene la mutación diseñada y completamos con agua hasta 50 µL de volumen final de reacción. Se realizaron 18 ciclos de amplificación (1 minuto a 95°C, 1 min. a 58°C y 1 min/kb de plásmido a 68°C precedidos por un ciclo de desnaturalización inicial de 2 min. a 95°C). El DNA amplificado se digirió con la enzima *DpnI* para degradar el DNA molde metilado y después se transformaron células competentes DH5α de *E. coli* con 1 µL del producto obtenido. Las células se sembraron en placas de LB seleccionando con el antibiótico apropiado y se dejaron crecer 37°C durante 16-24 horas. Las construcciones obtenidas por mutagénesis en la secuencia del dominio DI de NALCN (D75R, N210Q y N216Q) y en la proteína NSP4 (NIS, QIC, AAAA, FFFF) se muestran en el Anexo 2C.

Secuenciación de material genético: Las inserciones, clonaciones, mutagénesis dirigidas y las demás construcciones realizadas fueron comprobadas mediante secuenciación del DNA por la empresa MacroGen Inc. (Seoul, Korea).

M.5. Ensayo topológico del dominio DI de NALCN y NSP4 en membranas de bacterias

En este ensayo utilizamos el reportero dual *pho-lac* del plásmido pKTop. Las seis construcciones del dominio DI de NALCN, 61-mer (pKTop-NALCN₁₋₆₁-pho-lac), 88-mer (pKTop-NALCN₁₋₈₈-pho-lac), 127-mer (pKTop-NALCN₁₋₁₂₇-pho-lac), 160-mer (pKTop-

NALCN₁₋₁₆₀-pho-lac), 200-mer (pKTop-NALCN₁₋₂₀₀-pho-lac) y 380-mer (pKTop-NALCN₁₋₃₈₀-pho-lac)) y las diez construcciones de la proteína NSP4, 141-mer (pKTop-NSP4₁₋₁₄₁-pho-lac), 164-mer (pKTop-NSP4₁₋₁₆₄-pho-lac), 279-mer (pKTop-NSP4₁₋₂₇₉-pho-lac), 313-mer (pKTop-NSP4₁₋₃₁₃-pho-lac), 335-mer (pKTop-NSP4₁₋₃₃₅-pho-lac), 346-mer (pKTop-NSP4₁₋₃₄₆-pho-lac), 362-mer (pKTop-NSP4₁₋₃₆₂-pho-lac), 371-mer (pKTop-NSP4₁₋₃₇₁-pho-lac), 397-mer (pKTop-NSP4₁₋₃₉₇-pho-lac) y 500-mer (pKTop-NSP4₁₋₅₀₀-pho-lac)) fueron transformadas en *E. coli* DH5α y cultivadas a 37°C en placas de agar LB con 50 µg/mL kanamicina. Para estudiar la topología de todas las construcciones mediante el ensayo cualitativo, una colonia fresca de cada construcción fue crecida hasta un OD_{600nm} de 0.5 en el medio LB con 50 µg/mL de kanamicina y se sembraron por goteo 7 µL en placas con agar LB suplementado con los indicadores 80 µg/mL de disodio 5-bromo-4-cloro-3-indolil fosfato (BCIP) y 100 µg/mL de 6-cloro-3-indolil-β-D-galactósido (Red-Gal), el inductor IPTG (1 mM) y 50 µg/mL de kanamicina. Las placas sembradas se dejaron a 37 °C toda la noche. Para el ensayo cuantitativo, una vez llevado el cultivo a un OD_{600nm} de 0,5 igual que en el apartado anterior, se indujo la expresión de la proteína química (truncados/PhoA-LacZ) con 1 mM de IPTG durante 1 hora a 37 °C. Para medir la actividad β-galactosidasa, se recogieron las células por centrifugación a 7.000 rpm durante 5 minutos y se resuspendieron en medio M63 (100 mM KH₂PO₄, 15 mM (NH₄)₂SO₄, 1,7 mM Fe₂SO₄, 1 mM MgSO₄, pH 7.0). Después las bacterias resuspendidas y permeabilizadas con SDS 0.05%: cloroformo (1:1), fueron mezcladas con tampón PM2 (70 mM Na₂HPO₄, 30 mM NaHPO₄, 1 mM MgSO₄, 0.2 mM MnSO₄, pH 7.0) y 0,15% de Orto-nitrofenil-β-D-galactosido (ONPG) para iniciar la reacción enzimática. Para medir la actividad fosfatasa, las células se recogieron por centrifugación a 7.000 rpm durante 5 minutos, y se resuspendieron en tampón frío PM1 (1 M Tris-HCl, pH 8.0, 0.1 mM ZnCl₂, 1 mM iodoacetamida). A continuación, después de permeabilizar la suspensión bacteriana con SDS0.05%:cloroformo (1:1), la reacción comenzó al añadir una mezcla de la solución sustrato de p-nitrofenil fosfato (pNPP) (0.15% in 1 M Tris-HCl, pH 8.0). Las reacciones β-galactosidasa y fosfatasa se detuvieron cuando se observa la presencia de un color amarillo añadiendo Na₂CO₃ (1 M) y NaOH (2N), respectivamente (Anexo 3, Figura 74 y 75). La actividad enzimática (A) en unidades relativas y la actividad relativa normalizada (NAR) fueron calculados como se describe en Karimova and Ladant, 2017.

$$A = 1000 \times \left(\frac{OD_{405} \text{ muestra} - OD_{405} \text{ control}}{OD_{595} \text{ muestra} - OD_{595} \text{ control}} \right) / t \text{ (min de incubación)}$$

Donde los OD₄₉₅ y OD₅₉₅ del control, corresponden a la mezcla de reacción que no contiene suspensión bacteriana.

$$NAR = \left(\frac{\text{actividad } PhoA \text{ muestra} / \text{máxima actividad } PhoA}{\text{actividad } LacZ \text{ muestra} / \text{máxima actividad } LacZ} \right)$$

Donde la máxima actividad de PhoA o LacZ, corresponde a la máxima actividad medida en cada ensayo.

M.6. Ensayo topológico del dominio DI de NALCN y NSP4 en membranas de células humanas.

La metodología para determinar la topología de ambas proteínas consiste resumidamente en obtener el extracto proteico de las células transfectadas con los correspondientes plásmidos donde se expresaron, realizar un tratamiento con EndoH del extracto proteico a una fracción de la muestra y finalmente, su análisis por electroforesis SDS-PAGE y *Western blot* (Anexo 3, Figura 76).

Obtención de extracto proteico de células humanas

Las construcciones truncadas del dominio DI de NALCN y de la proteína NSP4 fueron clonadas en el plásmido pCAGGS y transfectadas en células HEK-293T en placas multipocillo. Después de 24-48 horas a 37 °C y 5% CO₂, las células se recogieron de los pocillos y se centrifugaron a 250 g durante 10 minutos a 4 °C y se lavaron dos veces con 400 µL de PBS. Finalmente, las células fueron lisadas agregando 100 µL de tampón de lisis (Tris-HCl 20 mM pH 7.5, NaCl 150mM, 1% NP-40) y sonicadas durante 10 minutos en un bioruptor (Diagenode). Después de centrifugar las muestras a 13.000 rpm durante 10 minutos y 4 °C; el sobrenadante se mezcló en proporción 5:1 con tampón de carga (Laemmli *sample buffer*, Tris-HCl 625 mM pH 6.8, glicerol 10% (v/v), SDS 2% (p/v), β-mercaptoetanol 4% (v/v), azul de bromofenol 0.01% (p/v)).

Tratamiento con EndoH

Una fracción de las muestras de proteínas se trataron con endoglicosidasa H (EndoH) que es una enzima capaz de eliminar los N-oligosacáridos de una glicoproteína. El tratamiento de las muestras se realizó siguiendo las indicaciones del proveedor, añadiendo a la muestra en el tampón de carga GlycoBuffer 3 y 2 U de EndoH (New England Biolabs), incubando durante 1 hora a 37 °C.

Electroforesis SDS-PAGE y Western Blot

Después de la incubación con EndoH, las muestras de proteínas fueron calentadas a 96 °C durante 5 minutos para su desnaturalización y analizadas mediante electroforesis en geles de poliacrilamida en presencia de SDS. Las muestras desnaturalizadas fueron cargadas en geles de 12%, 14%, 16% y 20% de poliacrilamida dependiendo del tamaño de las proteínas truncadas y la electroforesis se realizó a un voltaje constante (90V).

Según el procedimiento los geles se tiñeron con azul de Coomassie (0,1 % w/w Coomassie Blue R250, 30 % metanol, 5 % ácido acético) y se decoloraron con solución decolorante (250ml metanol, 70 ml ácido acético y 680ml H₂O destilada), o se transfirieron (90V con límite de 300 mA durante 1 hora) a una membrana de nitrocelulosa Hybond-ECL (Amersham Bioscience) en tampón de transferencia (metanol, solución de transferencia 10x y agua en proporción 1:0,5:3,5, respectivamente). Una vez finalizada la transferencia, la membrana se bloqueó con leche desnatada al 5% (p/v) y para la inmunodetección por transferencia *western blot* se empleó el anticuerpo primario anti-c-myc (Sigma) obtenido en conejo a una dilución 1:10000 y el anticuerpo secundario de cabra contra la IgG (H+L) de conejo conjugado a HRP a 1:8.000. Después de lavados sucesivos con tampón de lavado (10x TBS en agua al 1:10 y 0.05% tween), la detección de la señal quimioluminiscente se obtuvo tras incubar la membrana de nitrocelulosa con el reactivo de revelado ECL de GE Healthcare en el equipo ImageQuant™ LAS 4000mini Biomolecular Imager (GE Healthcare). Esta señal nos permite distinguir la glicosilación de las dianas reporteras en las diferentes construcciones truncadas de las proteínas del DI NALCN y NSP4; ya que una sola glicosilación confiere una diferencia en la movilidad electroforética de ~2.5 kDa permitiéndonos averiguar su localización luminal (diana reportera glicosilada) o citoplásmica (diana reportera no glicosilada).

Para el mapeo de glicosilaciones de las construcciones en el sistema Lep, los 6 segmentos TM del DI de NALCN clonados en C_LTM 3G y el TM₁₀₋₃₂ putativo de NSP4 clonado en C_LTM 2G, se siguió el mismo protocolo descrito arriba para la transfección en células HEK-293T, extracción de proteínas, tratamiento con EndoH, electroforesis e inmunodetección. La diferencia de ~2,5 kDa aproximadamente en la movilidad electroforética nos indicará inserción en membrana del RE o traslocación al lumen de las proteínas evaluadas con respecto a su control.

M.7. Procesamiento de la diana putativa de NSP4

El objetivo de este ensayo fue verificar la existencia de una diana de corte del complejo SPC en la secuencia de NSP4. Para ello, se diseñaron dos mutantes en los que se

reemplazaron la secuencia TPVH (residuos 28-31 en NSP4, 2792-2796 en pp1a) original por cuatro residuos de alanina o fenilalanina. A continuación, sembramos las células HEK-293T en placas de 24 pocillos con las mismas condiciones y medios de crecimiento que en procedimientos anteriores. Cuando las células llegaron a un 70-80% de crecimiento en placa transfectamos el plásmido pCAGGS con la secuencia clonada que codifica la proteína silvestre (pCAGGS-NSP4_{TPVH}) y las secuencias codificantes de ambos mutantes de NSP4 (pCAGGS-NSP4_{AAAA} y pCAGGS-NSP4_{FFFF}); después las células se cultivaron durante 24-42 horas a 37 °C y 5 % de CO₂. La expresión de las construcciones transfectadas se analizó siguiendo la metodología descrita anteriormente en la extracción de proteínas e inmunodetección por *Western blot*.

Con el objetivo de verificar que el procesamiento *in vivo* de la proteína NSP4 ocurre también en otras células humanas y no humanas. Las células Vero y células A549 se cultivaron bajo las mismas condiciones y medios de cultivo empleados anteriormente. Se sembró 2 millones de cada línea celular en placas de 24 pocillos, las cuales fueron transfectadas 24 horas después con el plásmido pCAGGS que contiene la secuencia clonada que codifica la proteína wild type (pCAGGS-NSP4_{TPVH}). Transcurridas 24-42 horas de expresión de la construcción transfectada se continuó con la extracción de proteínas y ensayos de inmunodetección.

M.8. Inhibición del procesamiento de NSP4 por cavinafungin

Sembramos las células HEK-293T en placas de 6 pocillos, siendo las condiciones y medios de crecimiento los mismos que en procedimientos anteriores. Cuando las células llegaron a un 70-80% de crecimiento en placa, transfectamos el plásmido pCAGGS-NSP4_{TPVH}. La cantidad de plásmido transfectado fue de 2 µg, el cual fue mezclado con 250 µL de DMEM sin suero y 5 µL de polietilenimina (PEI) e incubado 5 minutos a temperatura ambiente. Después de tal incubación se agregó a la mezcla 750 µL de DMEM suplementado con FBS al 5% y se incubó durante 10 minutos. Esta mezcla se añadió a las células y se incubaron 4h. Tras la incubación se cambió el medio por DMEM suplementado con 10 % de FBS y con o sin 1µM de cavinafungin (inhibidor del complejo peptidasa, SPC) (Zanotti et al., 2022). Posteriormente, las células se cultivaron durante toda la noche a 37 °C, 5 % de CO₂ y la expresión de las construcciones transfectadas se analizó siguiendo la metodología descrita anteriormente.

M.9. Expresión, purificación y cristalización de dominios solubles de la proteína NSP4

Expresión de dominios solubles

Los dominios solubles de NSP4 S40-P270 y T411-Q500 se clonaron pETNKL que añade cola de His en Nt. Los plásmidos se transformaron en la cepa *E. coli* C43. Para poner a punto las condiciones de expresión de las proteínas se realizó una dilución 1:50 del precultivo de una colonia en medio Hyper Broth™ (AthenaES™) o LB suplementados con Kanamicina en placas de 24 pocillos y dejándose incubar a 37 °C hasta que se alcanzó una OD 0.6-0.8. Se indujo la expresión de la proteína con 0.5mM de IPTG durante 3 horas (Miguel-Romero et al., 2020) a 37 °C ó 20 °C. Antes y después de la inducción se tomaron muestras que se centrifugaron a 6000 rpm durante 5 minutos. El sedimento se resuspendió en tampón de muestras de Laemmli y el extracto proteico se analizó por electroforesis SDS-PAGE en geles de 16% de acrilamida. Los geles se tiñeron con azul de Coomassie y se decoloraron con solución decolorante 2.5:0.7:6.8 de metanol, ácido acético y agua destilada, respectivamente. Para la expresión a gran escala, 25ml de un precultivo incubado a 37°C durante toda la noche se diluyó en 1L de medio Hyper suplementado con glucosa (NSP4 S40-P270) o LB (NSP4 T411-Q500). Los cultivos se incubaron a 37°C en agitación hasta alcanzar a una OD 0.6-0.8, y se indujo con 0.5 mM de IPTG durante 3 horas. Finalmente, las células se centrifugaron a 4200 rpm durante 25 minutos a 4°C. Tras el lavado con LB, el sedimento se guardó a -20°C.

Purificación de dominios solubles

Obtenidas las condiciones óptimas de expresión para cada proteína, calculamos su punto isoeléctrico con la herramienta en línea Compute pI/Mw ExPASy (S40-P270 mer pI:4.93 y T411-Q500 mer pI:6.45). Basados en los puntos isoeléctricos preparamos las siguientes soluciones para su purificación. Para la proteína NSP4(S40-P270) preparamos la solución A1 de homogeneización (50mM Tris-HCl pH 8.0, 500mM NaCl, 1mM de PMSF, 0.5mM de TCEP) y solución B1 o elución (solución A1 con 500mM de Imidazol). Para la proteína NSP4(T411-Q500) se preparó la solución A2 de homogeneización (50mM Tris-HCl pH 8.5, 500mM NaCl, 1mM de PMSF, 0.5mM de TCEP) y solución B2 o elución (solución A2 con 500mM de Imidazol).

Los sedimentos celulares fueron resuspendidos en 20 ml de solución A, y sonicados durante 1seg/pulso; centrifugamos los cultivos sonicados a 12000 rpm durante 30 minutos a 4°C y guardamos el sobrenadante. La purificación de las proteínas del

sobrenadante se realizó por cromatografía de afinidad utilizando columnas de níquel HisTrap™ de 5ml y 1ml (Thermo Fisher Scientific) en el equipo de cromatografía ÄKTA Start de GE Pure™. Tras la carga de la muestra y posterior lavado con solución A, se eluyó con saltos de 7, 40, 70 y 100% de solución B. Todo el procedimiento de purificación de los sobrenadantes fue monitoreado siguiendo la A280 del eluido. Las fracciones se analizaron por electroforesis SDS-PAGE en geles de 16% de acrilamida (Huesa et al., 2021).

Solubilización de cuerpos de inclusión del dominio S40-P270 de NSP4.

En el caso de la proteína NSP4 (S40-P270), que formó cuerpos de inclusión en su proceso de expresión, el sedimento obtenido después de sonicar se resuspendió en solución A1 con diferentes cantidades de urea 2M, 3M, 4M o 6M para disgregar los cuerpos de inclusión (Singh et al., 2015). Las muestras se incubaron a 4°C durante 1 hora y posteriormente se centrifugaron. El sobrenadante y el sedimento de las proteínas tratadas con cada concentración de urea fueron separados y sólo el sobrenadante de las proteínas tratadas con urea 2M se cargó a la columna HisTrap™ de 5ml. Posteriormente, fueron analizados por electroforesis SDS-PAGE junto con los sobrenadantes y sedimentos de los otros tratamientos.

Eliminación de la etiqueta de Histidinas del dominio citosólico NSP4 (T411-Q500)

En el caso de la proteína NSP4(T411-Q500), la fracción eluida con 40% de solución B2 fue colectada y cuantificada por el método de Bradford (Kielkopf et al 2020). Para eliminar la cola de histidinas se trató la muestra con la proteasa GST-Prescission con una proporción proteína:proteasa de 20:1; agregando 1mM de EDTA y 5 mM de β -mercaptoetanol para aumentar la eficiencia de escisión. La mezcla se sometió a diálisis en 1L de solución A2 a 4°C y en agitación durante toda la noche para eliminar el imidazol. Al día siguiente, la muestra dializada se pasó por una columna HisTrap™ de 1ml recogiendo la fracción no retenida (Abdelkader et al., 2021; Huesa et al., 2021).

Purificación del dominio citosólico NSP4 (T411-Q500) sin etiqueta

Para separar la proteína NSP4 (T411-Q500) sin HisTag (10kDa) de la proteasa GST-Prescission de 44 kDa se utilizaron dos estrategias. Por un lado, utilizamos cromatografía de exclusión molecular en columna (Sephadex G10) (Hong et al., 2012) equilibrada en un tampón 50mM Tris pH8.5 y 300 mM NaCl. Y, por otro lado, se usaron columnas GSTrap™ para atrapar la proteasa GST-Prescission (Huesa et al., 2021; Xu et al., 2009; Leong, 1999). Las fracciones se analizaron por electroforesis. Las fracciones en las que apareció la proteína NSP4 (T411-Q500) se concentraron en tubos

concentradores Millipore de 3kDa *cutoff* por centrifugación a 3000 g durante 15 minutos hasta un volumen menor a 2 ml. La cantidad de proteína se cuantificó por el método de Bradford o midiendo la Absorbancia a 280nm (Huesa et al., 2021)

Cristalización del dominio citosólico

Los ensayos de cristalización se realizaron utilizando la técnica de difusión de vapor en gota sentada en placas de 96 pocillos tipo MRC2 que permite testar hasta dos muestras de proteínas purificadas mezclando una gota de la muestra con una gota de solución madre de cristalización presente en cada uno de los 96 reservorios de la placa (Xu et al., 2009; Panjikar et al., 2005). Las placas de cristalización utilizadas son matrices comerciales, que contienen distintos tampones a distinto pH así como diferentes agentes precipitantes, y se denominan JBScreen Classic HTS I y HTS II (Jena Bioscience) y PGA Screen (Molecular Dimensions). Se dispensaron 0.4 µl del fragmento T411-Q500 de NSP4 purificado a 2.7 y 3.96 mg/ml junto con 0.2 µl de solución madre por un equipo de nanodispersión (HoneyBee X8, Genomic Solutions) por el servicio de Cristalogénesis del Instituto de Biomedicina de Valencia (IBV). Las placas se incubaron a 21 °C durante un tiempo indefinido para la aparición de cristales (Xu et al., 2009). Los cristales se congelaron en nitrógeno líquido captándolos con lazos de nylon de diámetros variables entre 0.025 y 0.1 mm (Hampton Research) según su tamaño y pasándolos por solución con condiciones crioprotectoras. Los cristales fueron difractados en sesiones remotas en la línea BL13-XALOC del sincrotrón ALBA (Cerdanyola del Vallés, España).

RESULTADOS Y DISCUSIÓN

I. Estudio topológico del dominio DI de NALCN

La estructura del canal de sodio NALCN se ha resuelto recientemente por microscopía crioeléctrica (cryo-EM) (Kschonsak et al., 2020). Si bien la información estructural resulta valiosísima para entender los detalles moleculares de este canal, no proporciona la información necesaria para explicar todas sus peculiaridades funcionales, ni el efecto que pueden tener las mutaciones observadas asociadas a enfermedades en el correcto funcionamiento de la proteína (Chong et al., 2015).

Como objetivo inicial de esta parte de la tesis nos propusimos estudiar la biogénesis de la proteína en membranas celulares tanto de bacterias como de células eucariotas. Según la estructura resuelta por cryo-EM, esta proteína tiene cuatro dominios estructurales (DI, DII, DIII, DIV), y en cada uno de ellos observamos un dominio sensor de voltaje (VSD, *voltage sensor domain*), formado por 4 segmentos TM (S1, S2, S3, S4), y otra región llamada dominio del poro (PD, *pore domain*), formada por dos segmentos TM (S5 y S6) (Figura 14). En el presente trabajo hemos caracterizado los estadios iniciales de la biogénesis de la proteína, estudiando con detalle molecular la biosíntesis e inserción del primer dominio de este importante canal, el dominio DI. Cabe destacar que en este primer dominio de la proteína se encuentran casi la tercera parte de las mutaciones asociadas al fenotipo de CLIFAHDD, siendo éstas especialmente abundantes en su hélice TM S6 (Tabla 3).

R1. Análisis *in silico* de la topología del dominio DI de NALCN

Para el estudio de la topología del dominio DI realizamos inicialmente un análisis computacional utilizando los algoritmos ΔG Predictor y TMHMM (Sonnhammer et al., 1998; Hessa et al., 2005 y 2007). El dominio DI de NALCN tiene una longitud de 380 aminoácidos y una masa molecular estimada de aproximadamente 44 kDa. Los resultados de los algoritmos mostraron bastante consenso en la predicción de los residuos implicados en la formación de las hélices TM. El ΔG predictor predice que el DI de NALCN tiene 6 segmentos TM (Tabla 7); sin embargo, tres de sus segmentos TM (S1, S2 y S4) tienen valores de energía libre aparente de inserción (ΔG_{app}) positivos, lo cual sugiere una baja probabilidad de que estos segmentos sean reconocidos como verdaderos segmentos TM por el translocón (Tabla 7 y Figura 21).

Tabla 7. Predicción del ΔG Predictor de segmentos TM del DI de NALCN con sus respectivas posiciones, longitudes, valor ΔG_{app} y secuencia de aminoácidos.

TM	Posición	Longitud	ΔG_{app} (kcal/mol)	Secuencia
S1	40-61	22	0,683	SLLRICALISVISVCMNTPMTF
S2	69-88	20	1,828	YVTFTLDTLLMFLYTAEMIA
S3	106-127	22	-0,289	RWCVFDGFMVFCLWVSLVLQVF
S4	139-160	22	1,395	WGMLRIPRPLIMIRAFRIYFRF
S5	178-200	23	-2,556	IWSVSIFLLFLLLYGILGVQMF
S6	295-314	20	-2,028	RWRSYFYFITLIFFLAWLVK

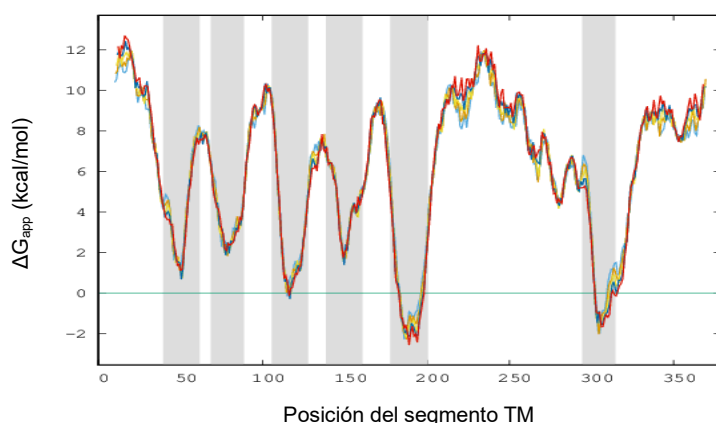


Figura 21. Predicción del ΔG predictor. El gráfico nos muestra la energía libre aparente ΔG_{app} (kcal/mol) del proceso de inserción del dominio DI de NALCN. Los seis segmentos TM predichos se resaltan en gris. Los valores positivos y negativos del ΔG_{app} están delimitados por una línea verde.

El algoritmo TMHMM además de proporcionar información sobre la presencia de segmentos TM predice la topología de la proteína. Los resultados de predicción de segmentos TM son similares a los obtenidos con el ΔG Predictor (Tabla 7 y Figura 21). De los seis segmentos TM predichos por TMHMM, sólo cinco de ellos (S1, S2, S3, S5 y S6) tienen una probabilidad de más del 80% de ser segmentos TM (Tabla 8). TMHMM también predice la topología de los polipéptidos analizados, sugiriendo en este caso que los extremos N- y C-terminal del DI están orientados hacia el citosol y que este dominio presenta entre el S5 y S6 una gran región de casi 100 aminoácidos orientada hacia el lado extracelular (Figura 22).

Tabla 8. Predicción del algoritmo TMHMM de los segmentos TM y la topología del DI de NALCN con sus respectivas posiciones y localizaciones (citósol, lumen o hélice TM).

Posición	Localización	Posición	Localización
1-38	Citósol	139-158	Hélice TM
39-61	Hélice TM	159-177	Citósol
62-65	Lumen	178-200	Hélice TM
66-88	Hélice TM	201-299	Lumen
89-106	Citósol	300-322	Hélice TM
107-129	Hélice TM	323-379	Citósol
130-138	Lumen		

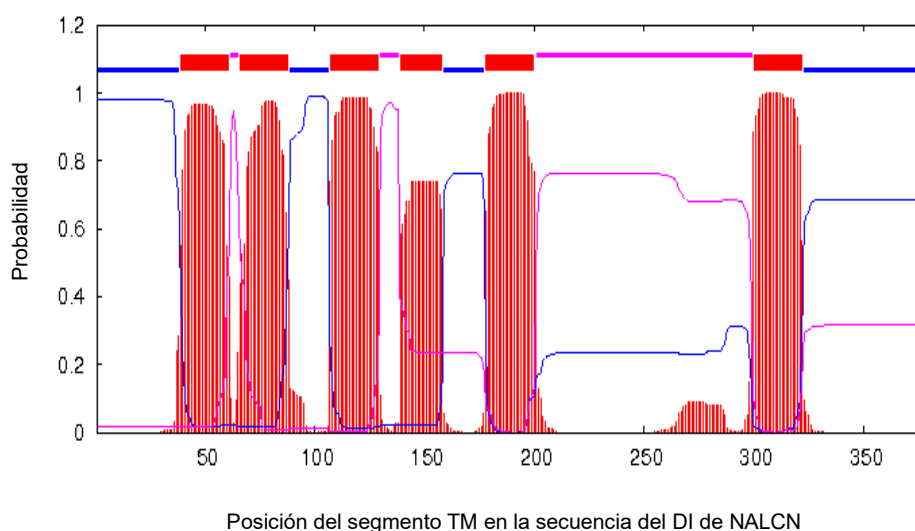


Figura 22. Predicción del TMHMM. Se representa la probabilidad de inserción en la membrana de cada región. En la línea superior del gráfico se indican la posición de los seis segmentos TM en la secuencia del DI de NALCN (en rojo), la línea azul indica orientación citosólica y la línea rosa orientación extracelular o luminal.

A pesar del elevado grado de coincidencia entre ambos algoritmos en la predicción de segmentos TM, se aprecia que hay una cierta discrepancia entre ambos para posicionar el S6; así, ΔG Predictor desplaza unos 5 residuos el S6 hacia el extremo N-terminal (295-314) en comparación con la predicción del TMHMM (300-322). Como se ha mencionado anteriormente, la abundancia de mutaciones CLIFAHDD en este segmento de la proteína sugiere la posibilidad de que estas mutaciones tengan un impacto en la inserción de este segmento TM en la membrana, por lo que resulta interesante determinar experimentalmente cuáles son los residuos de esta región que reconoce el translocón para su inserción en la membrana y la influencia que puedan tener estas

mutaciones en este proceso. Además, ambos algoritmos predicen como TM el segmento S4, a pesar de la elevada presencia de residuos de arginina en esta región de la proteína, para cuya inserción eficiente en la membrana puede requerir la participación las hélices TM adyacentes.

Por lo que respecta a modificaciones cotraduccionales del dominio DI, el análisis computacional de su secuencia de aminoácidos (Código UniProt: Q8IZF0) mediante el algoritmo SignalP 6.0 (<https://services.healthtech.dtu.dk/services/SignalP-6.0/>), diseñado para predecir péptidos señal, sugiere que la proteína NALCN no tiene péptido señal (Figura 23A), y el análisis adicional de la secuencia aminoacídica de NALCN utilizando UniProtKB (<https://www.uniprot.org/uniprotkb/Q8IZF0/entry>) identifica la presencia de dos dianas de glicosilación putativas situadas entre los segmentos S5-S6, en concreto en las posiciones N210 y N216 del dominio DI (Figura 23B).

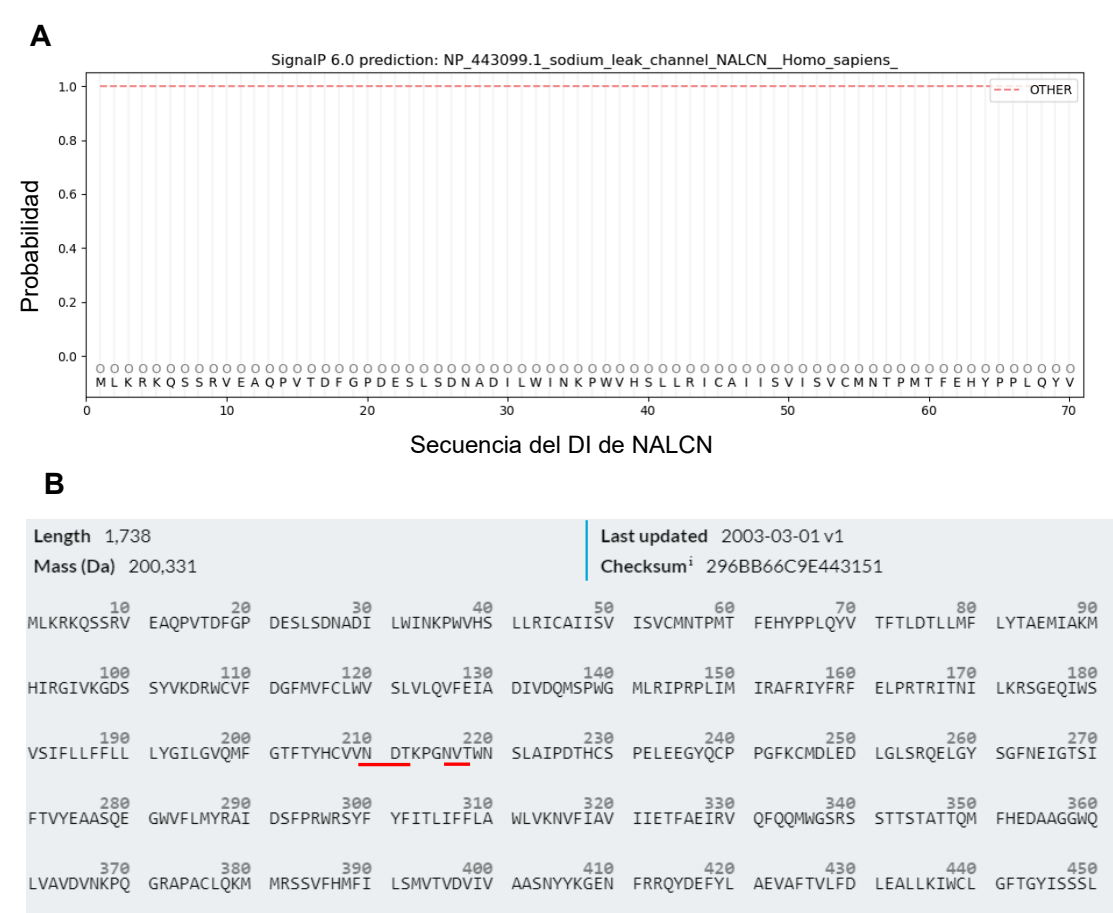


Figura 23. Análisis de las modificaciones cotraduccionales del dominio DI. (A) Identificación de péptido señal con SignalP. (B) La base de datos UniProtKB identifica tres putativos sitios de glicosilación en la secuencia de NALCN, dos de ellos (NDT y NVT) están ubicados en el dominio DI (subrayados con línea roja).

R.2. Inserción en membrana de los segmentos TM aislados.

Diseño experimental del ensayo

Para estudiar la inserción individual en membrana de los seis segmentos TM predichos (S1, S2, S3, S4, S5 y S6) del dominio DI de NALCN, desarrollamos un sistema basado en la glicosilación cotraduccional de proteínas de membrana y de secreción. En este sistema, llamado C_LTM-3G, el plásmido pSVLd incluye un dominio de la cadena ligera λ de una inmunoglobulina (C_L) precedido de su péptido señal que dirige el polipéptido hacia la membrana para su secreción. En la secuencia del dominio C_L diseñamos dos sitios de glicosilación (G1 y G2, Fig. 24A) mediante mutagénesis dirigida. La secuencia de C_L va seguida de una secuencia flexible (rica en residuos de Ser y Gly, *linker*), una región en la que clonamos de forma individual cada uno de los seis segmentos TM del DI (S1, S2, S3, S4, S5 y S6) (Feige et al., 2013; Duarte et al., 2022), una segunda secuencia flexible seguida de una tercera diana de glicosilación (G3) y etiquetada en el extremo C-terminal (c-myc). La transfección de las construcciones obtenidas permite verificar la inserción o translocación de las secuencias estudiadas, utilizando como reporteras las glicosilaciones (G1, G2 y G3). En este ensayo, las dianas G1 y G2 siempre estarán modificadas debido a que C_L está precedida por su péptido señal. En cambio, la diana de glicosilación G3 sólo será modificada si la región evaluada no se inserta en la membrana del RE y queda translocada en el lumen del RE (Figura 24B). Debemos mencionar que este diseño nos permite identificar posibles inversiones de la topología de los polipéptidos quiméricos, dado que la adopción de una topología invertida daría como resultado solo la modificación de la diana G3, quedando la proteína quimérica monoglicosilada, dado que las dianas G1 y G2 quedarían orientadas al citoplasma y por tanto inaccesibles al sitio activo de la OST.

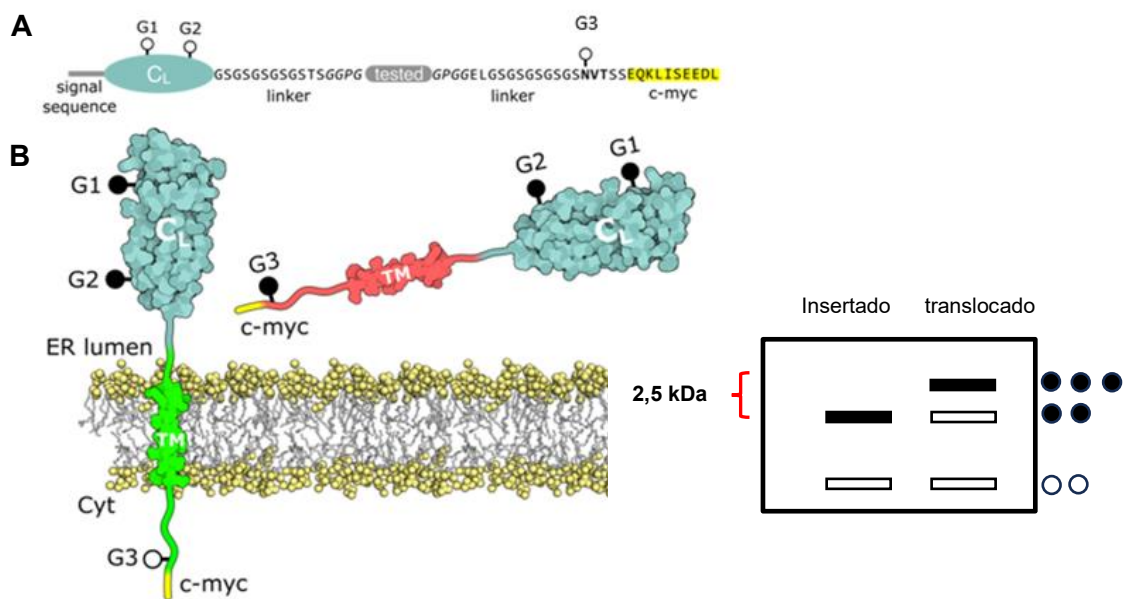


Figura 24. Sistema C_LTM-3G. (A) El esquema de la proteína quimérica muestra el dominio C_L (celeste) precedido de su secuencia señal en N-terminal, las dianas G1, G2 y G3 diseñadas, las secuencias de unión (*linker*) a ambos lados del TM evaluado (*tested*) flanqueado por los tetrapéptidos GGPG y GPGG; en C-terminal se muestra la etiqueta c-myc (amarillo). (B) El esquema de la izquierda muestra la inserción de un segmento TM (en verde), con las dianas G1 y G2 glicosiladas (círculo negro), y la diana G3 no glicosilada (círculo blanco) orientada al citoplasma. La figura de la derecha muestra el ejemplo de un segmento translocado (en rojo) con las 3 dianas (G1, G2, G3) glicosiladas. Cada glicosilación confiere un incremento de ~2,5 kDa en la masa molecular de la proteína.

Basándonos en las predicciones de los algoritmos ΔG predictor y TMHMM, seleccionamos los aminoácidos incluidos en cada uno de los seis segmentos TM que clonamos en el sistema C_LTM-3G y estimamos sus masas moleculares (Tabla 9).

Tabla 9. Residuos incluidos en las construcciones de los 6 segmentos TM del dominio DI. (*)
Peso molecular indica la masa de la construcción TM+ C_LTM-3G,

TM	Peso Molecular (*)	Longitud (aa)	Posición	Secuencia
S1	20.3 kDa	21	39-61	SLLRICAIIISVISVCMNTPMT
S2	20.6 kDa	23	66-88	PLQYVTFITLDTLLMFLYTAEMIA
S3	20.6 kDa	22	106-127	RWCVFDFGMVFCLWWSLVLQVF
S4	20.7 kDa	22	139-160	WGMLRIPRPLIMIRAFRIYFRF
S5	20.6 kDa	23	178-200	IVSVSIFLLFFLLLYGILGVQMF
S6	21.5Da	28	295-322	RWRSYFYFITLIFFLAWLVKNVFAVII

Estudio de la inserción de los TMs aislados del dominio DI en membranas.

Las 6 construcciones que contienen respectivamente los segmentos TM del DI de NALCN fueron analizadas electroforéticamente y visualizadas por *western blot* utilizando un anticuerpo anti-c-myc y quimioluminiscencia. De acuerdo con las

predicciones del ΔG Predictor encontramos que los polipéptidos que contienen las regiones poco hidrofóbicas (S1, S2 y S4) son predominantemente triplemente glicosilados, lo que indica que aislados del contexto de la proteína NALCN estos segmentos no son reconocidos como verdaderos TM por el translocón y no se insertan en la membrana del RE (Figura 25A, carriles 6, 8 y 10). Mientras que las proteínas quiméricas que contienen los segmentos TM más hidrofóbicos (S3, S5 y S6) fueron doblemente glicosiladas, lo que indica la inserción de S3, S5 y S6 en la membrana del RE (Figura 25B, carriles 6,8,10). Para verificar que la variación en la movilidad electroforética de estas proteínas se debe a la presencia de los azúcares se trataron las muestras con EndoH (Figuras 25A y B, carriles impares), una enzima altamente específica que escinde los oligosacáridos unidos a residuos de asparagina. Las células Hek293T se transfectaron en paralelo con construcciones control que contenían una secuencia TM (Figura 25A, carriles 3 y 4; Figura 25B, carriles 1 y 2) o una secuencia soluble (Figura 25A, carriles 1 y 2; Figura 25B, carriles 3 y 4) que se usaron como control de inserción y translocación (Martínez-Gil et al., 2010), respectivamente. Estos resultados sugieren que, aislados del resto de la secuencia de la proteína, sólo los segmentos S3, S5 y S6 son reconocidos por el translocón como verdaderos segmentos TM en células eucariotas.

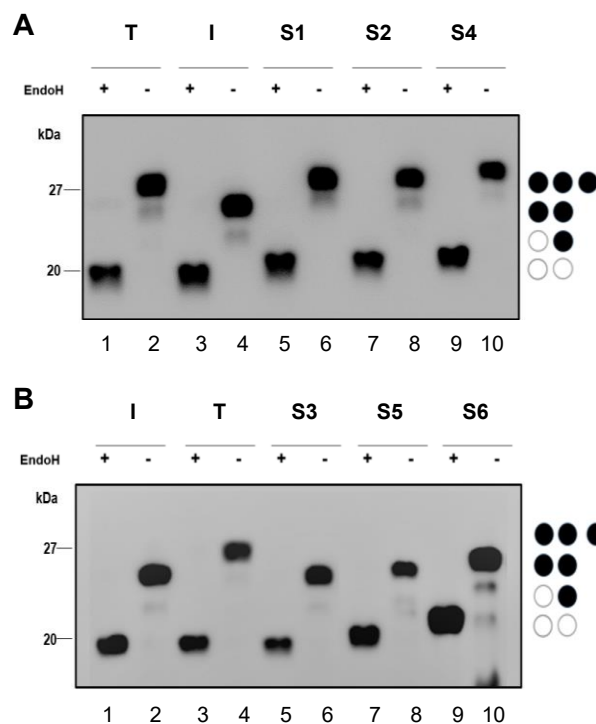


Figura 25. Inserción en membrana de los 6 segmentos TM del DI. Las células Hek293T transfectadas con las construcciones que contienen un control de inserción (I), un control translocación (T) y cada uno de los segmentos TM (S1-S6), fueron lisadas después de un tiempo de incubación. Los lisados celulares que contienen las proteínas expresadas fueron tratados en presencia (+) o ausencia (-) de EndoH, separadas por electroforesis SDS-PAGE en geles de 14% de acrilamida, transferidas a una membrana de nitrocelulosa y tratadas con anti-c-myc. Las proteínas etiquetadas fueron detectadas por quimioluminiscencia. La secuencia del TM utilizado como control de inserción (I) fue KVLLVTGVLGLLLLIKWK y la secuencia de translocación (T) fue YNLYIYAFFASSLVHGFI. Las proteínas no glicosiladas se indican con puntos blancos, las proteínas doble y triplemente glicosiladas se indican con dos y tres puntos negros, respectivamente.

R.3. Topología del dominio DI de NALCN en membranas procariotas

Diseño experimental del ensayo

Para estudiar la biogénesis del dominio DI de NALCN en membranas de bacterias, utilizamos el sistema reportero dual PhoA-LacZ α . En el sistema génico *pho-lac*, la molécula indicadora es una proteína quimérica compuesta por el péptido α de LacZ (residuos 4–60), que se fusiona con el extremo C-terminal de la enzima fosfatasa alcalina (PhoA) madura (residuos 22–472). Esta proteína quimérica PhoA-LacZ α sólo muestra actividad fosfatasa cuando está orientada al periplasma, mientras que cuando se orienta al citoplasma muestra actividad β -galactosidasa porque el fragmento α de LacZ interactúa con una variante inactiva truncada de LacZ (fragmento ω) presente en células *E. coli* DH5 α LacZ Δ M15, restaurando su actividad enzimática debido al fenómeno de α complementación (Karimova et al., 2017). Así, la fusión del reportero *pho-lac* a los polipéptidos objeto de estudio permite determinar *in vivo* la ubicación del extremo C-terminal de las proteínas de membrana evaluadas (Karimova et al., 2017). Para estudiar la inserción de los distintos segmentos TM de DI en su contexto polipeptídico se construyeron fusiones en pauta del indicador *pho-lac* en seis codones seleccionados del dominio DI (61-mer, 88-mer, 127-mer, 160-mer, 200-mer, 380-mer o longitud completa del dominio DI de NALCN) que incluían de forma acumulativa los distintos segmentos TM (Figura 26). A continuación, se transformaron de forma individual cada una de estas fusiones en células de *E. coli* y se realizaron ensayos cualitativos y cuantitativos.

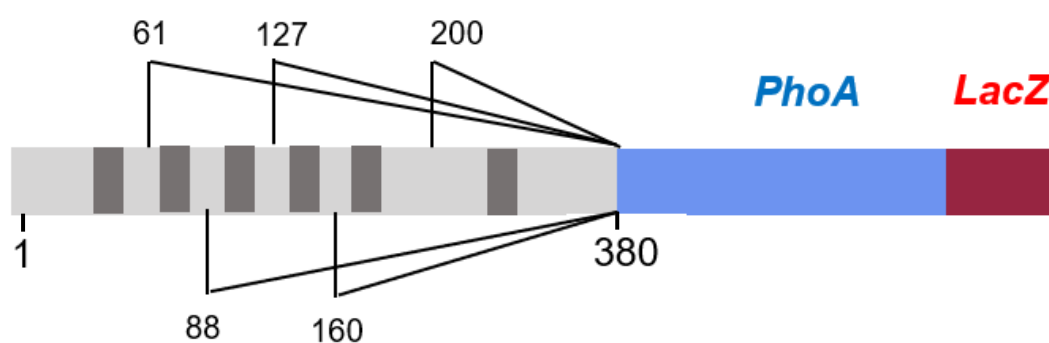


Figura 26. Variantes fusionadas de truncados del DI /phoA-lacZ. Se muestra de forma esquemática el dominio DI (380 aminoácidos, en gris), destacando en gris oscuro la posición de los segmentos TM y el residuo de fusión (61-mer, 88-mer, 127-mer, 160-mer, 200-mer, 380-mer). La fosfatasa alcalina y el péptido α de la β -galactosidasa se muestran en azul y rojo respectivamente.

Topología del dominio DI de NALCN en membranas de bacterias

Los transformantes resultantes se analizaron en medios de indicador dual que contenían un sustrato cromogénico azul para la actividad fosfatasa y un sustrato cromogénico rojo para la actividad β -galactosidasa. Como se muestra en la imagen superior de la Figura 27, las células que expresan los polipéptidos truncados de 61 residuos mostraron un fenotipo azul, lo que indica una ubicación periplásmica del indicador *pho-lac* y, en consecuencia, del residuo correspondiente F61 de NALCN. Esto sugiere que el primer segmento TM (S1) adopta una topología C-terminal periplásmica. La fusión del indicador *pho-lac* después del residuo de A88 (88-mer) produjo colonias rojas-granate en los medios de indicador dual, lo que indica que el residuo A88 está localizado en el citosol (o estrechamente asociado a la membrana). Con ambos resultados cualitativos podemos sugerir la presencia de un segmento TM (S2) entre los residuos 61 y 88 donde se realizaron las fusiones. Cuando el indicador *pho-lac* se fusionó después del residuo F127 (127-mer), la proteína quimérica exhibió una alta actividad fosfatasa (azul intenso), consistente con una ubicación periplásmica del residuo de F127 y lo que sugiere la presencia de un segmento TM (S3) entre los residuos 88 y 127. La fusión del reportero *pho-lac* después del residuo de F160 produjo colonias rojas en el medio de indicador dual, lo que indica que el residuo F160 está localizado en el citosol y sugiere que existe un segmento TM (S4) entre los residuos 127 y 160. Cuando el indicador *pho-lac* se fusionó después del residuo F200 las proteínas quiméricas exhibieron una alta actividad fosfatasa (fenotipo azul), consistente con una ubicación periplásmica del correspondiente residuo y sugiriendo la presencia de un segmento TM (S5) entre los residuos 160 y 200. Finalmente, la fusión del reportero *pho-lac* después del residuo de M380 (último residuo del dominio DI de NALCN) produjo colonias rojas en el medio de indicador dual, lo cual indica que el extremo C-terminal del dominio DI está localizado en el citosol y sugiere que existe un segmento TM (S6) entre los residuos 200 y 380. Para corroborar los resultados cualitativos obtenidos con los distintos truncados del dominio DI en las suspensiones bacterianas transformadas con las proteínas truncadas fusionadas a PhoA-LacZ α , se determinó la actividad relativa normalizada (NAR, (PhoA muestra/PhoA máxima)/(β Gal muestra/ β Gal máxima)) en base a la actividad enzimática fosfatasa alcalina y β -galactosidasa de cada quimera (Figura 27, Tabla 10). Los truncados 127-mer y 200-mer, tienen valores de NAR claramente mayores que 2, indicativos de una actividad fosfatasa alcalina prevalente de estas quimeras y que los respectivos aminoácidos F127 y F200 se ubican en el periplasma. Por otro lado, los truncados 160-mer y 380-mer presentan valores NAR<0.5, lo cual indica una actividad β -galactosidasa prevalente de estas quimeras y que sus aminoácidos F160 y M380

están ubicados en el citosol. Sin embargo, los truncados más pequeños de 61-mer y 88-mer tienen algunos valores NAR comprendidos entre 0.5 y 2 (Tabla 10), sugiriendo que los aminoácidos F61 y A88 podrían estar ubicados en el periplasma y citosol, respectivamente y/o ligeramente asociados a membrana.

Estos resultados indican que la proteína NALCN tiene seis segmentos TM y ambos extremos N- y C-terminal se ubican hacia el citosol en membranas de bacterias, pero que esta topología empieza a estar claramente definida a partir de la biosíntesis e inserción del tercer segmento TM (S3), presentando los truncados más cortos una disposición en la membrana menos definida. Es interesante mencionar al respecto que el segmento S3 es el primero de la secuencia del dominio DI que es reconocido como verdadero segmento TM por el translocón cuando se estudia su inserción aislada del resto de la secuencia de NALCN (Figura 25).

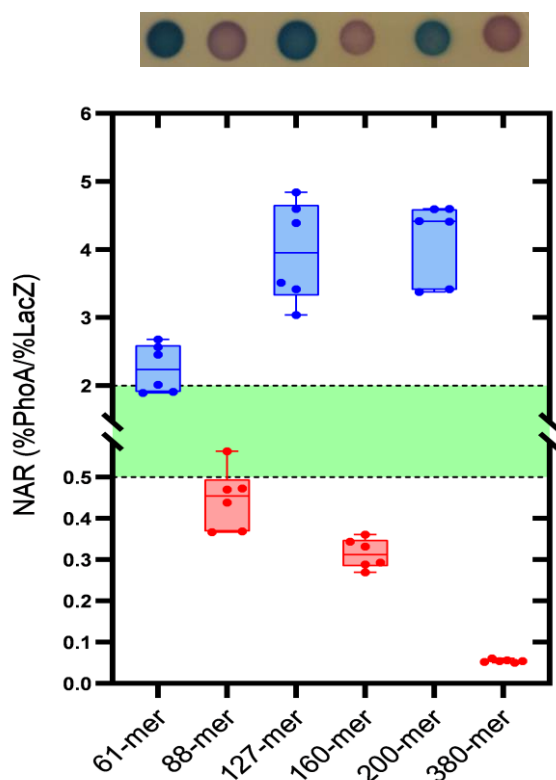


Figura 27. Topología del dominio DI de NALCN en membranas bacterianas. (Arriba) Ensayo cualitativo, se sembraron células de *E. coli* DH5 α que expresaban las diferentes fusiones DI NALCN/Pho-Lac en medio indicador con dos sustratos cromogénicos, Red-Gal (para la actividad de β -galactosidasa) y X-Pho (para la actividad de fosfatasa alcalina). (Abajo) El diagrama de cajas muestra la actividad relativa normalizada (NAR) de los ensayos enzimáticos cuantitativos de cada variante fusionada al reportero *pho-Lac*. Los puntos sólidos (rojos o azules) muestran los valores de experimentos individuales. Los valores de NAR >2 son indicativos de ubicación periplásmica, entre 2 y 0,5 son indicativos de ubicación TM (verde) y los valores <0,5 son indicativos de ubicación citosólica.

Tabla 10. Valores cuantitativos de la actividad relativa normalizada (NAR) de los seis polipéptidos truncados de NALCN fusionados a PhoA-LacZ α . Los valores de cada construcción corresponden a seis réplicas independientes del ensayo (valores >2 en azul y <0,5 en rojo).

61-mer	88-mer	127-mer	160-mer	200-mer	380-mer
2,013	0,562	4,596	0,331	4,409	0,060
1,910	0,368	3,414	0,343	4,593	0,054
1,891	0,469	4,388	0,269	3,377	0,050
2,677	0,438	4,840	0,288	4,417	0,052
2,455	0,366	3,037	0,360	4,596	0,056
2,568	0,472	3,511	0,292	3,414	0,054

R.4. Modificación de las dianas de glicosilación N210DT y N216VT

El análisis de la secuencia de aminoácidos del dominio DI revela la presencia de dos sitios de glicosilación putativos (NDT y NVT) en el ECL que une los segmentos S5 y S6 (Figura 23B) en las asparaginas 210 y 216. Como se ha descrito anteriormente, la glicosilación cotraduccional de cadenas nacientes en células eucariotas la realiza el complejo multienzimático oligosacaril transferasa (OST) (Figura 13). Para verificar si ambas dianas putativas son glicosiladas por la OST, construimos sobre la secuencia silvestre del DI de NALCN dos mutantes, uno simple en el que reemplazamos por mutagénesis dirigida la asparagina por glutamina (N216Q) y otro mutante doble en el que reemplazamos por mutagénesis las dos asparaginas por dos glutaminas, eliminando de esta forma ambas dianas de glicosilación (N210Q-N216Q). A ambas variantes mutantes y a la construcción silvestre del dominio I de NALCN se les añadió una etiqueta c-myc para su inmunodetección y fueron clonadas en un plásmido de expresión en eucariotas y posteriormente las tres construcciones fueron transfectadas en células Hek293T. Tras el crecimiento y lisis celular, las proteínas expresadas fueron analizadas por *western blot* y quimioluminiscencia.

El mutante simple N216Q produce formas monoglicosiladas en (Figura 28, carril 2), lo que nos indica que la N210 de la diana NDT es modificada por la OST mediante la adición de azúcares. Como era de esperar, en el mutante doble N210Q-N216Q no observamos glicosilaciones (Figura 28, carril 6), ya que los residuos de glutamina (Q) no son sustrato de la actividad OST. Por lo que respecta al polipéptido silvestre del dominio DI (380-mer), se observó una banda de proteínas doblemente glicosiladas (Figura 28, carril 4). El tratamiento con EndoH de las muestras verificó la presencia de azúcares en las bandas de menor movilidad electroforética correspondientes a las formas glicosiladas de las proteínas (Figura 28, carriles 1 y 3). Estos resultados

demuestran que el ECL comprendido entre los segmentos TM S5 y S6 está orientado hacia el lumen del RE; y que ambas dianas canónicas (N²¹⁰DT y N²¹⁶VT) son susceptibles de ser modificadas (glicosiladas) por la OST en células humanas.

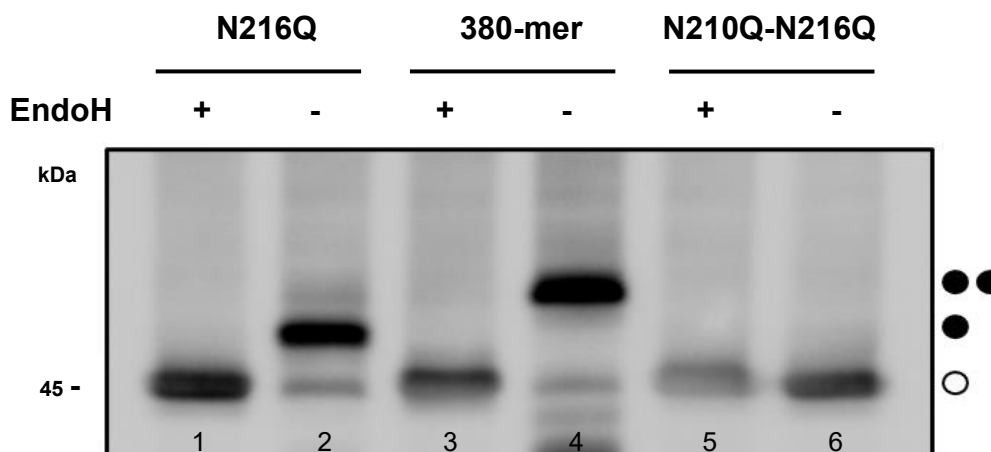


Figura 28. Las dianas de glicosilación N210DT y N216VT son modificadas en células eucariotas. Las células Hek293T transfectadas con las construcciones silvestre 380-mer (carril 3 y 4), N216Q (carril 1 y 2) y N210Q-N216Q (carril 5 y 6) fueron lisadas y tratadas con (+) o sin (-) EndoH. La presencia de proteínas se detectó por inmunotransferencia. Las proteínas no glicosiladas se indican con puntos blancos y las proteínas monoglicosiladas y doblemente glicosiladas con uno y dos puntos negros, respectivamente.

R.5. Topología del dominio DI de NALCN en membranas celulares humanas

Diseño experimental del ensayo

Para estudiar los estadios iniciales de la biogénesis de NALCN y determinar experimentalmente cómo la proteína va adoptando su topología en células eucariotas utilizamos una técnica basada en el mapeo de glicosilaciones del dominio DI (Tamborero et al., 2011; Bañó-Polo et al., 2011). Para ello, se diseñaron una serie de polipéptidos truncados de longitud creciente que contienen etiquetas de glicosilación en N- y C-terminal optimizadas, y como controles, etiquetas no glicosilables en las que se reemplazó el residuo de asparagina aceptor por una glutamina no aceptora. Las predicciones computacionales de los segmentos TM (Figuras 21 y 22) y los resultados obtenidos en células procariotas (apartado R.3) nos ayudaron a seleccionar la longitud de cada una de las construcciones truncadas del dominio DI que evaluamos (65-mer, 97-mer, 136-mer, 163-mer, 233-mer y 380-mer), cada una de ellas incluyendo un segmento TM más respecto al truncado anterior hasta completar la secuencia del dominio DI. Estas secuencias fueron amplificadas por PCR, clonadas en el vector apropiado (pCAGGS) para la transformación de células eucariotas y expresadas en cultivos celulares (Hek293T). Mediante PCR añadimos las etiquetas anteriormente mencionadas en los extremos N- y C-terminales de las construcciones truncadas, de tal

forma que por cada truncado tendremos 4 variantes. La variante NN (presenta dos etiquetas de glicosilación diseñadas en ambos extremos N- y C-terminal), NQ (tiene una etiqueta de glicosilación agregada en N-terminal, mientras que en C-terminal se añadió una secuencia no glicosilable), QN (tiene solo una etiqueta glicosilable en C-terminal) y la variante QQ (con etiquetas no glicosilables en ambos extremos) (Figura 29). Estas dianas de glicosilación añadidas nos permiten averiguar la localización de los extremos N y C-terminal de cada una de las construcciones truncadas, lo que nos proporcionará información sobre cómo van integrándose cada uno de los segmentos TM (S1-S6) en las membranas del RE en los estadios iniciales de la biosíntesis de NALCN.

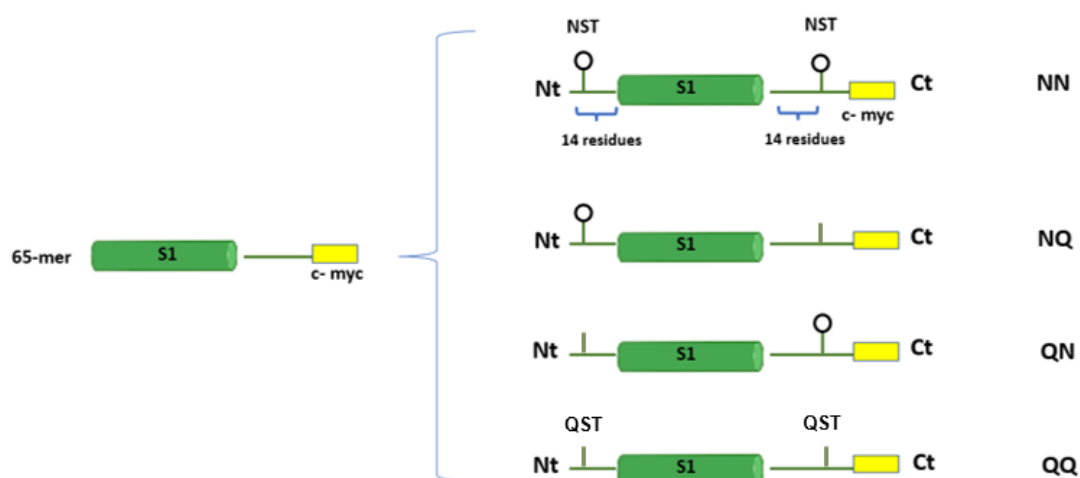


Figura 29. Diseño de las dianas de glicosilación reporteras. Cada una de las 6 construcciones truncadas tiene a su vez 4 variantes. Se muestra como ejemplo el truncado de 65-mer: la proteína 65-mer NN (con diana de glicosilación en N- y C-terminal), 65-mer NQ (con diana de glicosilación en N-terminal), 65-mer QN (con diana de glicosilación en C-terminal) y 65-mer QQ (no glicosilable). Las dianas de glicosilación fueron añadidas a una distancia >12-14 residuos de los extremos del S1 (en verde) para aumentar la eficiencia de glicosilación (Nilsson et al., 1993; Bañó-Polo et al., 2012), y el epítipo c-myc (amarillo) incorporado en C-terminal. Se indica con un círculo la diana de glicosilación.

Biogénesis del dominio DI de NALCN en células humanas

Las proteínas expresadas en células Hek293T fueron analizadas por inmunotransferencia para verificar el estado de glicosilación de sus dianas. En la Figura 30 se muestran los resultados representativos obtenidos con las células transfectadas con el polipéptido truncado de 65 aminoácidos (65-mer) que contiene en su secuencia el extremo N-terminal de NALCN incluyendo el primer segmento TM. La variante 65-mer NN produjo dos bandas de proteínas (carril 2), si bien la banda de menor peso molecular es mayoritaria. La banda de menor movilidad electroforética en los geles corresponde a la forma glicosilada del polipéptido truncado ya que su peso molecular aumentó en ~2.5

kDa y además esta banda de mayor peso molecular desaparece tras el tratamiento con EndoH (comparar carriles 1 y 2). En el resto de variantes de este truncado sólo se observó la presencia de formas glicosiladas en la variante que presenta la diana de glicosilación en C-terminal (QN) (Figura 30, carril 6). Conjuntamente estos resultados sugieren que los truncados de 65 aminoácidos que contienen la primera de las regiones TM de la proteína se inserta poco eficientemente en la membrana del RE, pero la población de moléculas que son insertadas lo hacen con la topología correcta (N-terminal citoplasmático) de la proteína. Por tanto, y en concordancia con los datos obtenidos en los estudios de inserción de los segmentos TM aislados (Fig. 25A), el S1 no es reconocido eficientemente por el translocón como un verdadero segmento TM.

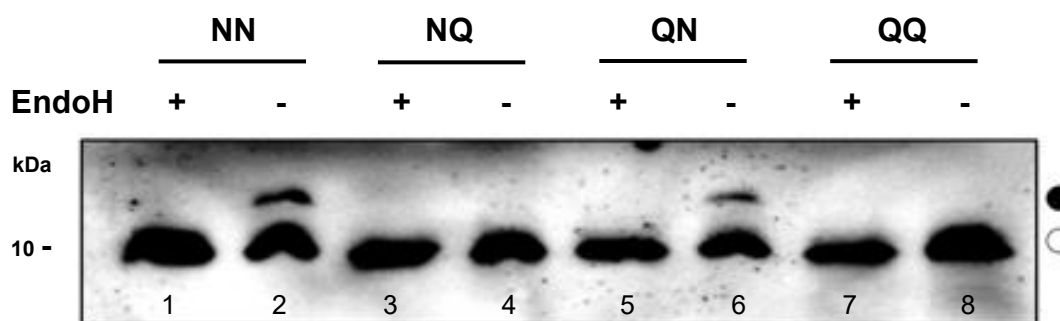


Figura 30. El truncado 65-mer se inserta poco eficientemente en la membrana con topología Nt citoplasmática. Las células Hek293T, transfectadas con las variantes NN, NQ, QN y QQ, fueron lisadas para extraer las diferentes proteínas truncadas expresadas. La inmunodetección de estas proteínas truncadas se realizó por *western blot* y quimioluminiscencia. Se comprobó el estado de glicosilación de las muestras mediante tratamiento con (+) y sin (-) EndoH. Las proteínas no glicosiladas se indican con un punto blanco y las proteínas monoglicosiladas con un punto negro.

El análisis de las variantes truncadas de 97 residuos (97-mer) que contienen los dos primeros segmentos TM del dominio DI (S1 y S2) mostraron formas glicosiladas tanto en el extremo N-terminal como en el C-terminal de la proteína (Figura 31). Así, en las muestras correspondientes a la variante NN expresadas en células, observamos dos bandas de proteínas (Figura 31, carril 2). La banda de proteínas menos intensa y de menor migración electroforética corresponde a la forma doblemente glicosilada, en la que ambas etiquetas en N- y C-terminal son modificadas, ya que la diferencia de peso molecular respecto a la banda no glicosilada es de 5 kDa aproximadamente. Además, en las muestras correspondientes a los dos truncados que presentan una sola diana de glicosilación (NQ y QN) observamos también la presencia de dos bandas de proteínas, pero en este caso las bandas menos intensas y de menor migración electroforética corresponden a las formas monoglicosiladas, ya que el incremento observado en su peso molecular es de ~2.5 kDa (Figura 31, carriles 4 y 6). Estos resultados sugieren en

primer lugar que los polipéptidos truncados que contienen los dos primeros segmentos TM no se insertan eficientemente en la membrana del RE en células eucariotas, de nuevo en concordancia con la baja hidrofobicidad de estas dos regiones TM y con los resultados obtenidos tanto en los estudios de inserción de los TMs aislados (Figura 25A) como en los estudios con los truncados en células procariotas (Figura 27); o bien que se insertan ambos segmentos TM con la topología correcta de la proteína, es decir con sus extremos N- y C-terminal orientados al citoplasma. Las formas glicosiladas que se observan en las tres variantes con dianas de glicosilación en algún extremo (carriles 2, 4 y 6) probablemente corresponden a formas translocadas de estos truncados en las que los segmentos S1 y S2 no son insertados en la membrana, o bien, aunque menos probable, a formas en las que sólo uno de los dos segmentos TM no es correctamente insertado por el translocón, bien con la topología nativa o invertida.

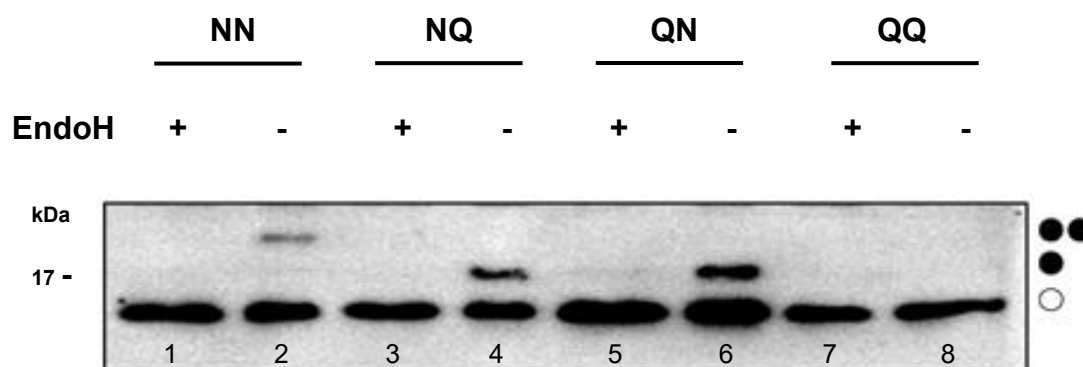


Figura 31. Topología de la proteína truncada 97-mer. Las células Hek293T, transfectadas con las variantes truncadas 97-mer (NN, NQ, QN y QQ), fueron lisadas para extraer las diferentes proteínas expresadas. La inmunodetección de estas proteínas truncadas se realizó con antisuero anti-c-myc, anticuerpo secundario y quimioluminiscencia. Se comprobó el estado de glicosilación de las muestras mediante tratamiento con (+) y sin (-) EndoH. Las proteínas no glicosiladas se indican con un punto blanco, las proteínas mono- y doblemente glicosiladas con uno y dos puntos negros respectivamente.

Los polipéptidos truncados de 136 aminoácidos (136-mer), que incluyen los tres primeros segmentos TM (S1-S3), muestran formas glicosiladas de la proteína sólo en aquellas variantes que contienen una diana de glicosilación en C-terminal (NN y QN), lo que sugiere una topología N-terminal citoplasmática con el C-terminal de estas moléculas orientado al lumen del RE, coincidente con la topología nativa de NALCN (Figura 32). También, se observa por primera vez que la banda menos intensa corresponde a formas no glicosiladas (de mayor migración electroforética), lo cual nos sugiere que la síntesis del segmento S3 es importante para la correcta integración en la membrana de los segmentos TM precedentes (S1 y S2). En este sentido es interesante destacar que S3 es el primer segmento TM del dominio DI suficientemente hidrofóbico como para tener un valor de ΔG_{app} de inserción negativo (Figura 21), indicativo de una

alta probabilidad de inserción que además hemos visto verificada dada su inserción como segmento aislado de su contexto de secuencia en la proteína (Figura 25B).

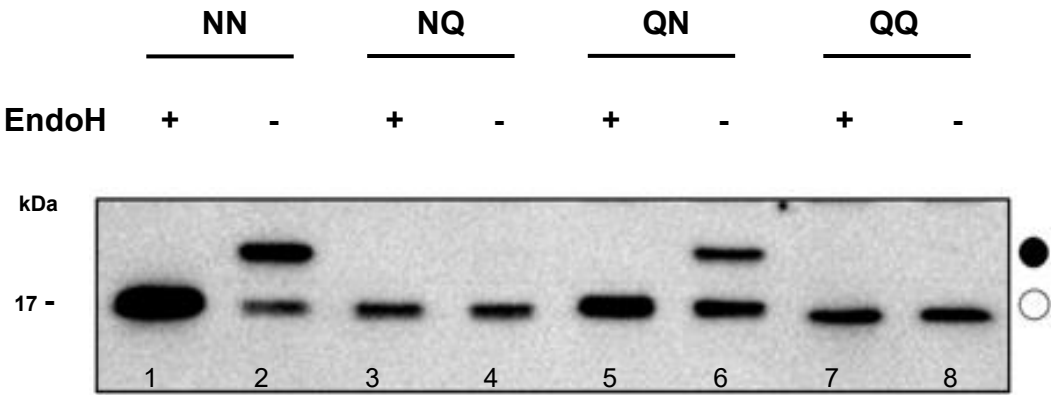


Figura 32. Topología de la proteína truncada 136-mer. Las células Hek293T, transfectadas con las variantes del truncado de 136 aminoácidos, fueron lisadas para extraer las diferentes proteínas expresadas. La inmunodetección de estas proteínas truncadas se realizó con antisuero anti-c-myc, anticuerpo secundario y quimioluminiscencia. Se comprobó el estado de glicosilación de las muestras mediante tratamiento con (+) y sin (-) EndoH. Las proteínas no glicosiladas se indican con un punto blanco y las proteínas monoglicosiladas con un punto negro.

En el caso de las variantes truncadas de 163 aminoácidos que incluyen los cuatro segmentos TM (S1-S4) que forman el dominio sensor de voltaje (VSD), observamos en todos los casos sólo una banda de proteínas correspondiente con la forma no glicosilada del polipéptido (Figura 33). Estos resultados sugieren que la proteína truncada está insertada en la membrana del RE con sus extremos N y C-terminales orientados hacia el citoplasma (coincidiendo con la topología nativa), con el residuo P163 localizado en el citosol y, por ende, la localización del segmento TM S4 entre los residuos 136 y 163.

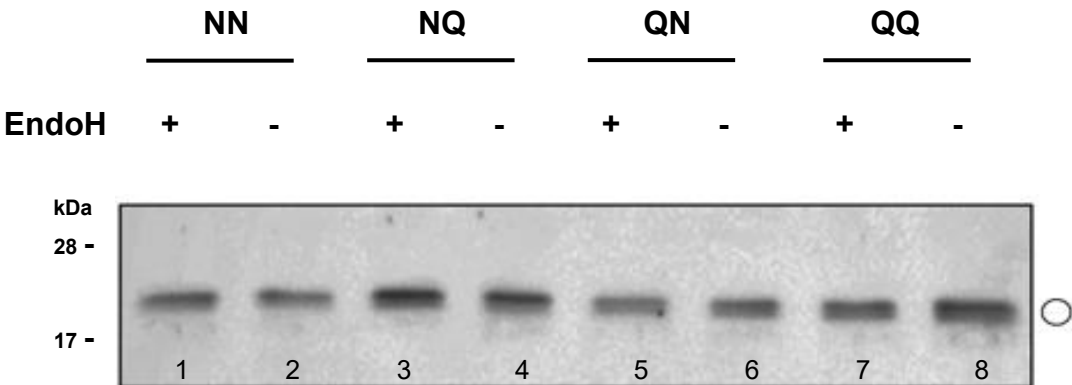


Figura 33. Topología de la proteína truncada 163-mer. Las células Hek293T, transfectadas con los truncados de 163 aminoácidos, fueron lisadas para extraer las diferentes proteínas truncadas expresadas. La inmunodetección de estas proteínas truncadas se realizó con antisuero anti-c-myc, anticuerpo secundario y quimioluminiscencia. Se comprobó el estado de glicosilación de las muestras mediante tratamiento con (+) y sin (-) EndoH. Las proteínas no glicosiladas se indican con un punto blanco.

La expresión de las proteínas truncadas de 233 aminoácidos que incluyen los cinco primeros segmentos TM (S1-S5) y la región ECL que une los segmentos S5 y S6 en la que se encuentran las dos dianas de glicosilación nativas del dominio DI generó formas triplemente glicosiladas en aquellas variantes en las que se incluye una diana de glicosilación en C-terminal (NN y QN) (Figura 34, carriles 2 y 6), mientras que la movilidad electroforética de las variantes sin dianas de glicosilación añadidas o con una diana añadida en el extremo N-terminal (QQ y NQ) corresponde a formas doblemente glicosiladas de la proteína (Figura 34, carriles 4 y 8), originadas por la modificación de los sitios de glicosilación nativos N210 y N216. Globalmente, estos resultados nos indican que el residuo L233 del extremo C-terminal de la proteína truncada está situada en el lumen, ubicando al segmento TM S5 entre los residuos 163 y 233, y sugiriendo una topología global para este polipéptido truncado acorde con la estructura nativa de NALCN.

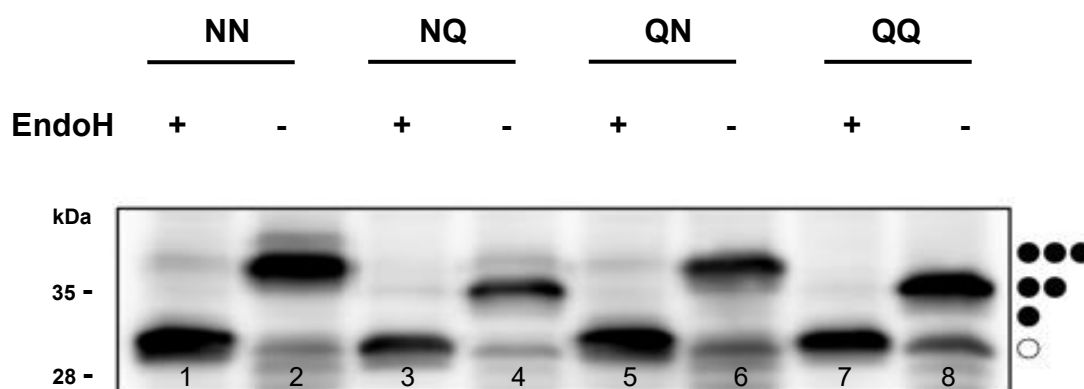


Figura 34. Topología de la proteína truncada 233-mer. Las células Hek293T, transfectadas con las variantes truncadas de 233 aminoácidos, fueron lisadas para extraer las diferentes proteínas expresadas. La inmunodetección de estas proteínas truncadas se realizó mediante *western blot* y quimioluminiscencia. Se comprobó el estado de glicosilación de las muestras mediante tratamiento con (+) y sin (-) EndoH. Las proteínas no glicosiladas se indican con un punto blanco, las proteínas mono-, doble- y triplemente glicosiladas con uno, dos y tres puntos negros, respectivamente.

Finalmente, las células transfectadas con el dominio DI completo de la proteína muestran en todas las variantes formas doblemente glicosiladas (Figura 35), que sugieren una topología N- y C-terminal citoplasmática y la modificación de las dos dianas de glicosilación nativas del dominio DI (N210 y N216).

En conjunto, de los estudios de la biosíntesis del dominio DI de NALCN utilizando truncados de longitud creciente podemos concluir que este primer dominio de la proteína adopta una orientación en la membrana del RE acorde con la estructura resuelta a partir de la aparición del tercer segmento TM (S3), cuya presencia favorece la inserción de los segmentos que le preceden (S1 y S2) y define de forma unívoca la orientación del resto

de segmentos TM del dominio DI (S4-S6), probablemente facilitando la correcta inserción del resto de dominios de NALCN (DII-DIV) y permitiendo la adopción de su estructura nativa.

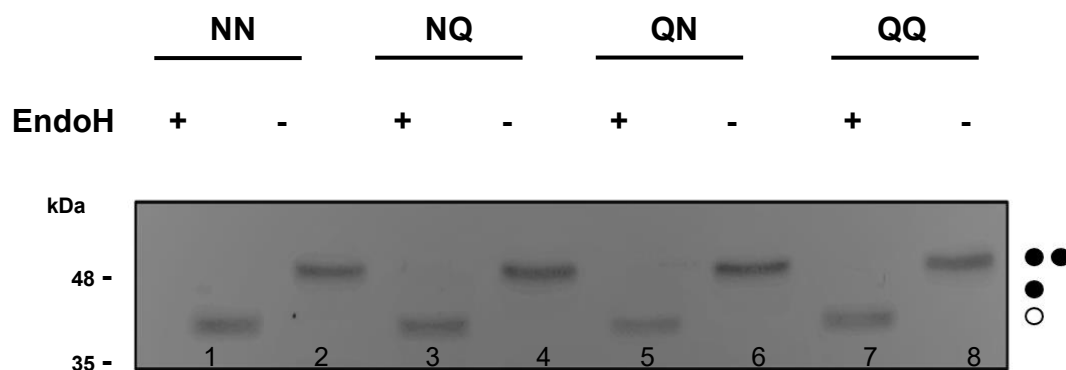


Figura 35. Topología de la proteína truncada 380-mer. Las células Hek293T, transfectadas con las variantes del dominio DI completo fueron lisadas para extraer las diferentes proteínas expresadas. La inmunodetección de estas proteínas truncadas se realizó con antisuero anti-c-myc, anticuerpo secundario y quimioluminiscencia. Se comprobó el estado de glicosilación de las muestras mediante tratamiento con (+) y sin (-) EndoH. Las proteínas no glicosiladas se indican con puntos blancos y las formas mono- y doblemente glicosiladas con uno y dos puntos negros, respectivamente.

R.6. Interacciones entre los segmentos del VSD

En el caso de las proteínas de membrana se ha descrito que sus segmentos TM interactúan entre sí a través de contactos entre los aminoácidos que contribuyen a su empaquetamiento y a la consiguiente formación de estructuras funcionales (Barber et al., 2012; Pedersen et al., 2014). NALCN, proteína que forma un canal selectivo para cationes monovalentes (Chua et al., 2020), tiene en su estructura segmentos TM poco hidrofóbicos cuyos valores de ΔG hacen que sea poco probable que el translocón los reconozca e inserte en membrana (Tabla 7). A pesar de ello, los 24 segmentos TM de NALCN están integrados en la membrana (Kschonsak et al., 2020), donde es muy probable que los contactos o interacciones inter-hélice entre los aminoácidos de sus segmentos TM contribuyan a su estabilidad. Para estudiar con detalle atómico estos contactos entre los 4 primeros segmentos TM (los menos hidrofóbicos) del dominio DI de NALCN, analizamos *in silico* sus trayectorias en membrana mediante simulaciones por dinámica molecular. Buscamos contactos inter-hélice entre aminoácidos a distancias menores de 3,5 Å, distancia considerada como un rango de interacción muy corto (Onofrio et al., 2014; Tam et al., 2022). En casi todos los segmentos encontramos contactos inter-hélice entre sus aminoácidos; sin embargo, sólo seleccionamos aquellos que eran reproducibles a través de las réplicas. Entre los segmentos S1 y S2, seleccionamos los siguientes contactos: M55-T71, S52-L74, S52-D75, S49-L78, L42-

E85, Y64-L67, P65-P66 y P65-L67 (Figura 36, A y B). Los valores de medianas y distancias medias de interacción en estas parejas de aminoácidos están por debajo del rango de interacción indicado (Tabla 11). Las parejas de aminoácidos M55-T71, S52-D75 y P65-P66 registran una menor dispersión, lo que indica una menor variación de sus distancias de interacción (Figura 36C, Tabla 11). De este grupo destacamos la pareja de aminoácidos S52-D75, donde el grupo hidroxilo (OH) de la cadena lateral de la S52 está muy cerca del grupo carboxilo de la cadena lateral del D75, a una distancia media de 1.67 Å, indicándonos una alta probabilidad de presencia de un puente de hidrógeno.

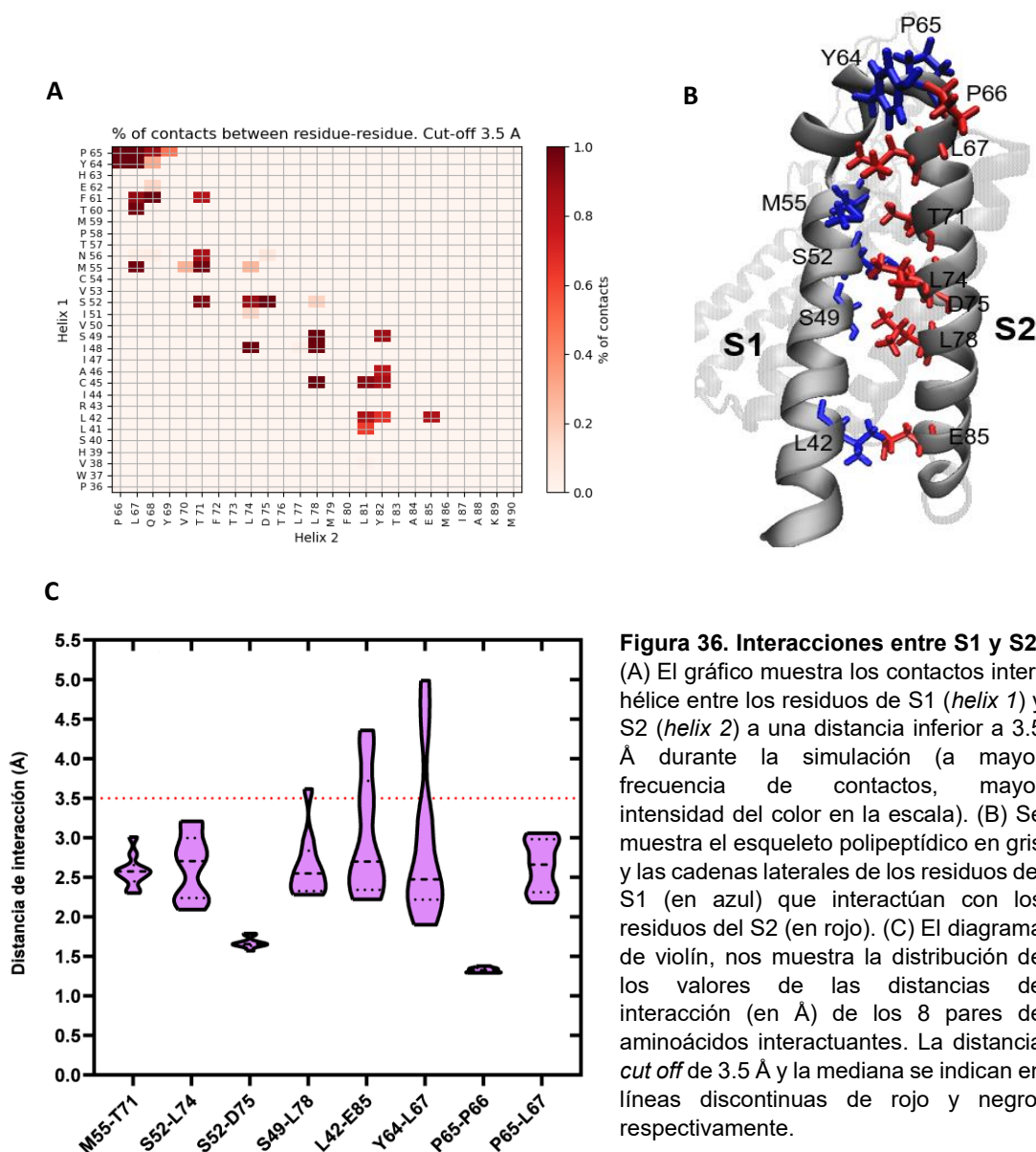


Figura 36. Interacciones entre S1 y S2. (A) El gráfico muestra los contactos inter-hélice entre los residuos de S1 (*helix 1*) y S2 (*helix 2*) a una distancia inferior a 3.5 Å durante la simulación (a mayor frecuencia de contactos, mayor intensidad del color en la escala). (B) Se muestra el esqueleto polipeptídico en gris y las cadenas laterales de los residuos del S1 (en azul) que interactúan con los residuos del S2 (en rojo). (C) El diagrama de violín, nos muestra la distribución de los valores de las distancias de interacción (en Å) de los 8 pares de aminoácidos interactuantes. La distancia *cut off* de 3.5 Å y la mediana se indican en líneas discontinuas de rojo y negro, respectivamente.

Tabla 11. Estadística descriptiva de las distancias de interacción entre S1 y S2

Residuos S1-S2	Mediana (Å)	Media (Å)	SD	Error estándar	Coefficiente Variación
M55-T71	2.57	2.57	0.20	0.06	7.93%
S52-L74	2.70	2.64	0.41	0.13	15.81%
S52-D75	1.65	1.67	0.06	0.02	3.94%
S49-L78	2.55	2.64	0.41	0.12	15.53%
L42-E85	2.70	3.02	0.78	0.24	25.86%
Y64-L67	2.47	2.92	1.03	0.32	35.21%
P65-P66	1.32	1.32	0.03	0.01	2.30%
P65-L67	2.66	2.64	0.32	0.10	12.22%

Las parejas de aminoácidos de las hélices S1 y S4 con distancias medias de interacción menores a 3.5 Å son R43-A153, A46-I149, A46-I153, V50-M150, S52-R146 y V53-R146 (Figura 37, A y B) (Tabla 12). En este grupo no hay una dispersión notable entre las distancias de los residuos interactuantes, excepto entre V53-R146 (Figura 37C, Tabla 12). Esto nos sugiere que las distancias de interacción entre las hélices S1 y S4 son relativamente mayores que las observadas entre las hélices S1 y S2, pero con menor dispersión.

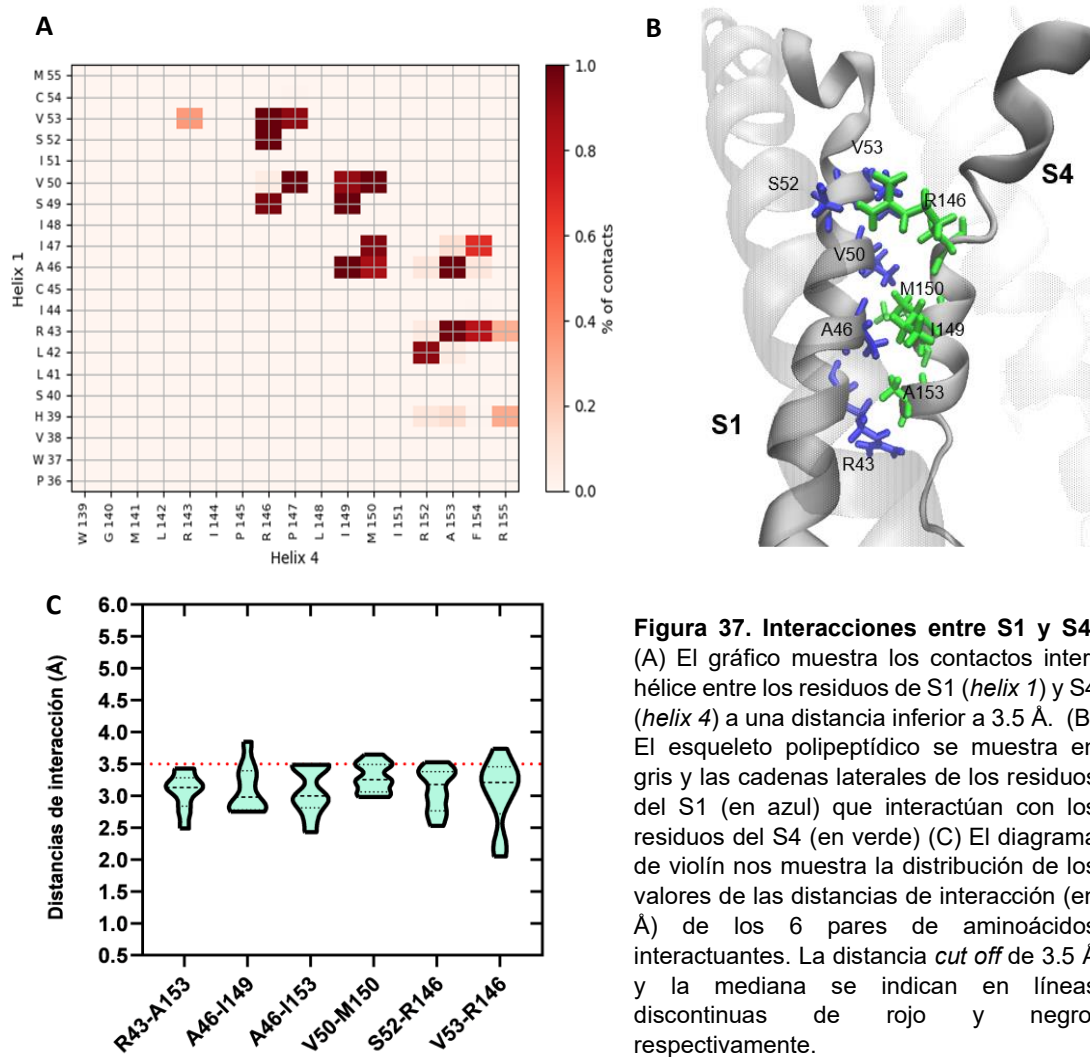


Tabla 12. Estadística descriptiva de las distancias de interacción entre S1 y S4.

Residuos S1-S4	Mediana (Å)	Media (Å)	SD	Error estándar	Coefficiente Variación
R43-A153	3.13	3.07	0.29	0.09	9.68%
A46-I149	2.98	3.11	0.37	0.12	12.05%
A46-I153	3.00	3.07	0.38	0.11	12.28%
V50-M150	3.25	3.29	0.23	0.07	7.06%
S52-R146	3.18	3.10	0.34	0.10	11.09%
V53-R146	3.21	3.06	0.53	0.17	17.35%

De las interacciones entre los segmentos S2 y S3 destacan F72-Q125, D75-S121, D75-Q125, M79-L118, Y82-L118 y K89-D111 (Figura 38 A y B, Tabla 13). Las parejas de aminoácidos D75-Q125 y D75-S121 muestran valores con una mayor dispersión en sus distancias (Tabla 13). El análisis estructural detallado de las trayectorias de la simulación entre los residuos D75-S121, nos permitió determinar que el grupo hidroxilo de la cadena lateral de la S121 está muy cerca al carboxilo de la cadena lateral del D75, indicativo de una alta probabilidad de interacción por puente de hidrógeno.

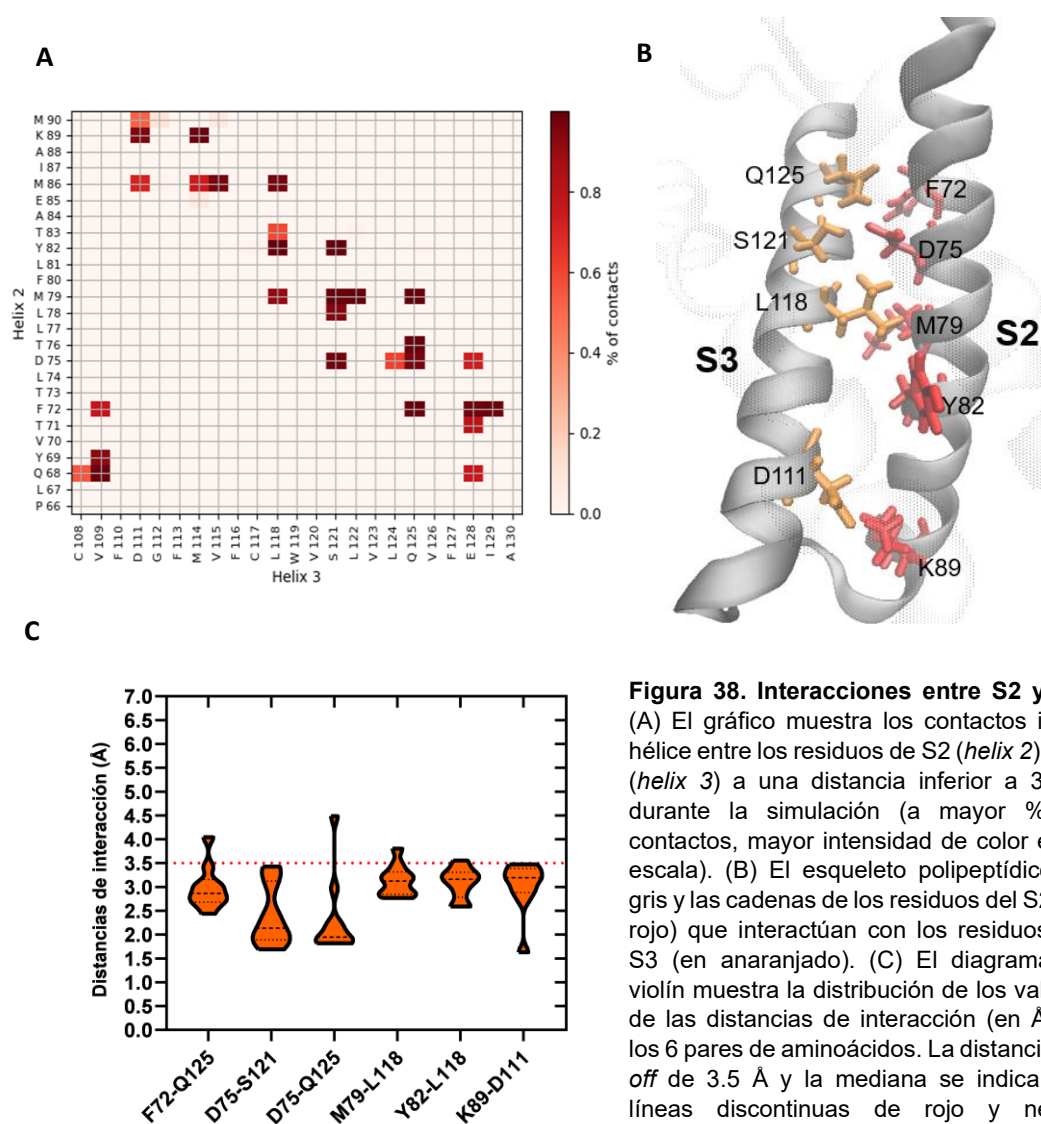


Figura 38. Interacciones entre S2 y S3. (A) El gráfico muestra los contactos inter-hélice entre los residuos de S2 (*helix 2*) y S3 (*helix 3*) a una distancia inferior a 3.5 Å durante la simulación (a mayor % de contactos, mayor intensidad de color en la escala). (B) El esqueleto polipeptídico en gris y las cadenas de los residuos del S2 (en rojo) que interactúan con los residuos del S3 (en anaranjado). (C) El diagrama de violín muestra la distribución de los valores de las distancias de interacción (en Å) de los 6 pares de aminoácidos. La distancia *cut off* de 3.5 Å y la mediana se indican en líneas discontinuas de rojo y negro, respectivamente.

Tabla 13. Estadística descriptiva de las distancias de interacción entre S2 y S3.

Residuos S2-S3	Mediana (Å)	Media (Å)	SD	Error estándar	Coefficiente Variación
F72-Q125	2.86	2.98	0.46	0.14	15.45%
D75-S121	2.14	2.39	0.64	0.20	27.10%
D75-Q125	1.95	2.32	0.83	0.26	36.17%
M79-L118	3.12	3.14	0.32	0.10	10.41%
Y82-L118	3.16	3.08	0.33	0.10	10.58%
K89-D111	3.19	3.01	0.54	0.17	17.94%

Al analizar las interacciones entre los segmentos S2 y S4, encontramos sólo dos parejas de aminoácidos (D75-R146 y Y82-L149) cuyas distancias cercanas entre ellos permanecen constantes a través de las réplicas (Figura 39 A y B). En la pareja de aminoácidos cercanos D75-R146, se aprecia también un bajo coeficiente de variación de sus distancias medias, lo que nos indica una menor dispersión de sus distancias de interacción y una alta probabilidad de que ambos residuos interaccionen a cortas distancias entre ellos (Figura 39C, Tabla 14). Un análisis estructural de las trayectorias de ambos aminoácidos en la simulación, nos permitió calcular sus distancias medias de interacción, siendo éstas de 1,70 y 1,95 Å, correspondientes a las distancias entre los dos oxígenos de los carboxilos de la cadena lateral del D75 y los dos NH del grupo guanidino de la R146, lo que sugiere que probablemente estos grupos formen un puente salino.

Entre los segmentos S2 y S4, se evidencia una segunda interacción entre los aminoácidos Y82-L149, si bien al analizar las trayectorias durante la simulación se detectan algunos valores de distancias mayores a 3.5 Å, lo que incrementa su coeficiente de variación (Figura 39C, Tabla 14).

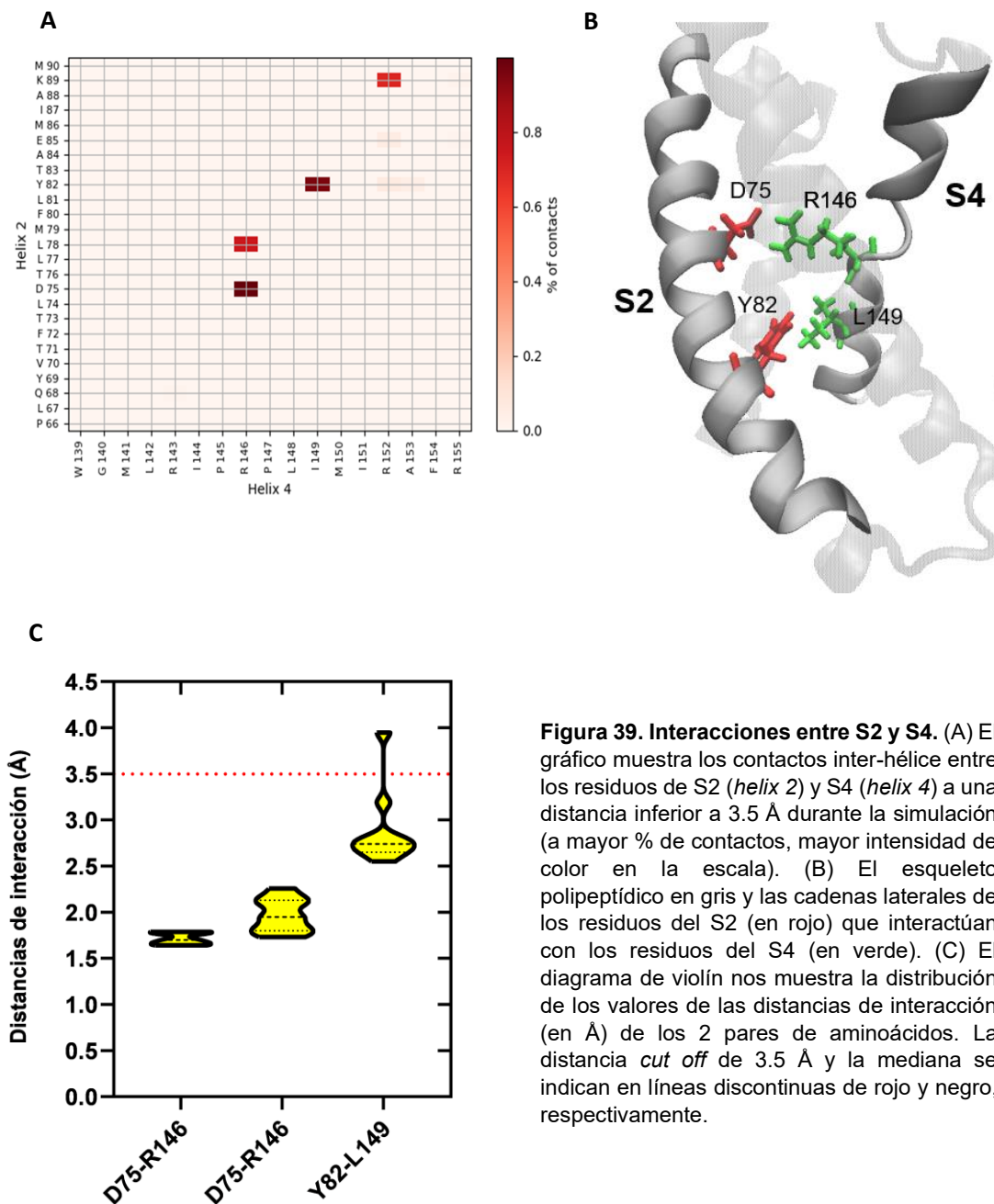


Tabla 14. Estadística descriptiva de las distancias de interacción entre S2 y S4.

Residuos S2-S4	Mediana (Å)	Media (Å)	SD	Error estándar	Coefficiente Variación
D75-R146	1.70	1.71	0.06	0.02	3.66%
D75-R146	1.95	1.96	0.18	0.05	9.24%
Y82-L149	2.74	2.87	0.41	0.13	14.4%

De los contactos observados entre los segmentos S3 y S4, sólo dos interacciones (S121-R146 y L124-R143) se encuentran en todas las réplicas (Figura 40 A y B) y presentan una baja dispersión (Figura 40 C, Tabla 15), lo que sugiere que ambas parejas de aminoácidos tienen una distribución homogénea en las simulaciones y por lo tanto, forman contactos estables.

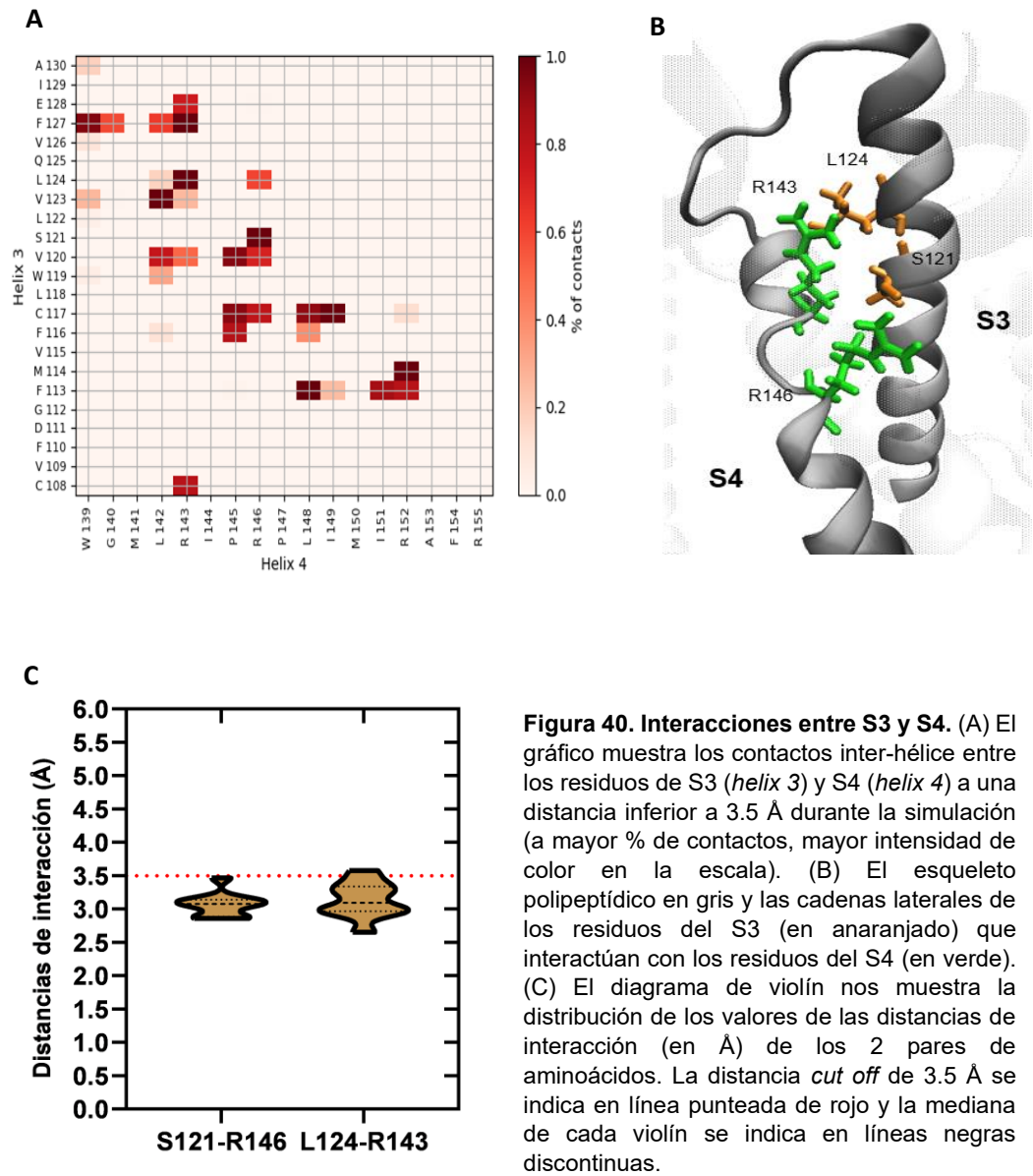


Figura 40. Interacciones entre S3 y S4. (A) El gráfico muestra los contactos inter-hélice entre los residuos de S3 (*helix 3*) y S4 (*helix 4*) a una distancia inferior a 3.5 Å durante la simulación (a mayor % de contactos, mayor intensidad de color en la escala). (B) El esqueleto polipeptídico en gris y las cadenas laterales de los residuos del S3 (en anaranjado) que interactúan con los residuos del S4 (en verde). (C) El diagrama de violín nos muestra la distribución de los valores de las distancias de interacción (en Å) de los 2 pares de aminoácidos. La distancia *cut off* de 3.5 Å se indica en línea punteada de rojo y la mediana de cada violín se indica en líneas negras discontinuas.

Tabla 15. Estadística descriptiva de las distancias de interacción entre S3 y S4.

Residuos S3-S4	Mediana (Å)	Media (Å)	SD	Error estándar	Coefficiente Variación
S121-R146	3.08	3.07	0.18	0.05	5.85%
L124-R143	3.09	3.14	0.27	0.08	8.72%

De todos los contactos analizados en los cuatro segmentos (S1-S4) del dominio sensor de voltaje (VSD), las interacciones entre los aminoácidos S52-D75 y D75-R146 muestran distancias medias de tan solo 1,67, 1,71 y 1,96 Å entre los grupos funcionales de sus cadenas laterales (Figura 41), siendo probablemente el puente salino entre D75 y R146 el principal enlace entre S2 y S4 que contribuye a su estabilidad en membrana (Figura 42). En las simulaciones también se observan los residuos S52 y S121 de los segmentos S1 y S3 respectivamente, a distancias compatibles con los residuos implicados en el puente salino (D75 y R146). Para determinar experimentalmente si la interacción entre D75 de S2 y R146 de S4 altera la inserción en membrana de la proteína truncada 163-mer, mutamos el aspártico 75 por una arginina (D75R). La inserción en membrana del mutante 163-mer (D75R) fue evaluada por mapeo de glicosilaciones y comparada con el polipéptido con la secuencia silvestre 163-mer. Los resultados muestran la presencia de una forma monoglicosilada en el mutante 163-mer (D75R) derivada de la translocación al lumen del RE del extremo C-terminal, lo cual confirma que la presencia del puente salino entre estos dos residuos permite la inserción eficiente del S4 en la membrana del RE (Figura 42).

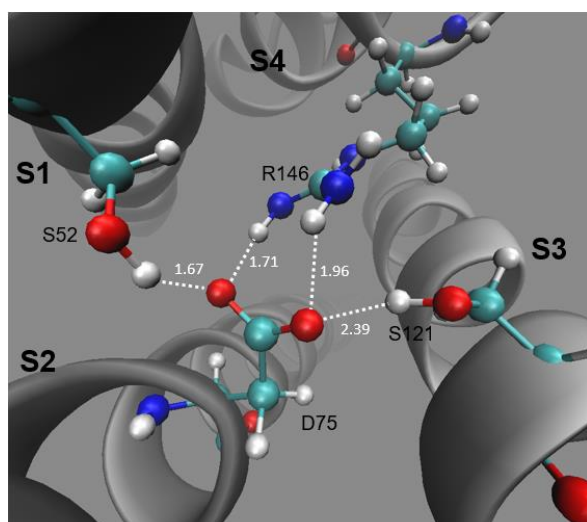


Figura 41. Principales interacciones en el dominio VSD. El gráfico de la simulación del DI de NALCN, nos muestra una vista desde arriba de la estructura de los 4 primeros segmentos TM (S1-S4) en gris. Asimismo, se destacan los 4 principales aminoácidos (S52, D75, S121 y R146) con sus distancias medias de interacción (en blanco) en Å. Las esferas rojas representan los oxígenos, las azules los nitrógenos, las blancas los hidrógenos y las turquesas los átomos de carbono.

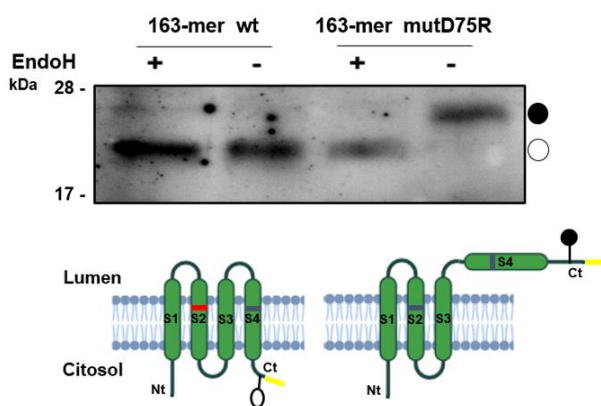


Figura 42. Interacción entre D75 (S2) y R146 (S4). Arriba, *Western blot* del truncado 163-mer silvestre (wt) y mutante (mutD75R), ambas muestras tratadas con EndoH (+). Abajo, esquema ilustrativo del resultado con los segmentos TM (en verde) y c-myc (en amarillo). La marca roja en S2 indica la posición del D75 y la azul la posición del R146 en S4.

R.7. Posicionamiento en membrana del S6.

Debido a las discrepancias observadas entre las predicciones de los residuos que atraviesan el núcleo hidrofóbico de la membrana en el segmento TM S6 (Tablas 7 y 8) y la estructura tridimensional resuelta, decidimos estudiar en detalle el posicionamiento de esta región de la proteína tanto aplicando técnicas experimentales como computacionales. Además, cabe mencionar que un número significativo de las mutaciones que conllevan la aparición de patologías severas ocurren en este segmento de NALCN, por lo que resulta relevante el estudio de su posición en membrana para evaluar el impacto que puedan tener las mutaciones en esta región. Para abordar este objetivo de la tesis utilizamos el mapeo por glicosilación como una 'regla' molecular que nos permita definir con precisión la posición del segmento S6 en las membranas del RE. Como se ha descrito en la introducción, las proteínas de membrana son sintetizadas por ribosomas asociados a las membranas del RE e insertadas mayoritariamente en la bicapa lipídica a través del translocón Sec61. En ese contexto, las cadenas nacientes correspondientes a proteínas de membrana (como NALCN) y de secreción son modificadas cotraduccionalmente mediante la adición de azúcares catalizada por el complejo multiproteico OST (Fig. 13). Debido a la ubicación del sitio catalítico de la OST en el lumen del RE se ha demostrado que la asparagina aceptora de un sitio de glicosilación debe estar a una distancia mínima de 14 residuos del inicio de un segmento TM para ser eficientemente modificada (Nilsson & von Heijne, 1993), lo que permite la verificación experimental del posicionamiento preciso de segmentos TM en el proceso de inserción en membranas biológicas (Kauko et al., 2010). Nuestra aproximación experimental se basa en la utilización como vehículo a la peptidasa del péptido señal (*leader peptidase*, Lep) de *E. coli*. Lep es una proteína de membrana que contiene dos segmentos TM (H1 y H2) conectados por un lazo citoplasmático (P1) y un dominio C-terminal voluminoso (P2) orientado al periplasma en el que reside el centro activo peptidasa (van Geest et al., 2000). Cuando se expresa esta proteína en células eucariotas el dominio H1 dirige la proteína a la membrana del RE, donde se inserta con el extremo N-terminal hacia el lumen (Figura 43A). El diseño experimental utilizado consistió en la clonación del segmento S6 en el dominio globular P2 de Lep y diseñar en esta quimera dianas de glicosilación precediendo la secuencia de S6 en distintas posiciones. Así, las construcciones diseñadas incluyen hasta tres dianas de glicosilación. Una primera diana (G1) ubicada a 136 aminoácidos de distancia del inicio de la secuencia de S6, que por su disposición deberá ser sustrato de la OST. Una segunda diana de glicosilación (G2) que se diseñó en distintas posiciones en N-terminal del inicio del primer residuo de S6, es decir, 9, 11, 13 o 15 residuos antes de la R295

que marca el inicio de la región hidrofóbica objeto de estudio (Tabla 16), y que será modificada solo si se encuentran a una distancia superior a 14 residuos antes del primer residuo TM. Y una tercera glicosilación (G3) diseñada en C-terminal de S6 y que sólo será modificada en el caso de que S6 fuese traslocado al lumen del RE, sirviéndonos de control de inserción (Figura 43A). Estas construcciones fueron transfectadas en células Hek293T y el estado de glicosilación de sus dianas fue analizado por SDS-PAGE y visualizado mediante inmunotransferencia (Figura 43B). En paralelo se transfectaron los pertinentes controles de inserción y translocación que carecen de la diana G2. En la construcción con sólo dos dianas de glicosilación G1 y G3, (Figura 43B, carril 6) se detecta una banda de proteínas de mayor peso molecular correspondientes a formas monoglicosiladas, producto de la modificación de la primera diana de glicosilación (G1), lo que fue corroborado mediante tratamiento con EndoH (+) (Figura 43B, carril 5). Estos resultados nos indican que el segmento S6 está correctamente integrado en la membrana, como ya demostramos utilizando como proteína modelo la cadena ligera de la inmunoglobulina C_L (Figura 25B). De forma análoga, las construcciones en las que el residuo de asparagina aceptor se encontraba a 9 y 11 residuos en N-terminal de la R295 generaron formas también monoglicosiladas producto de la modificación de la diana G1 (Figura 43B, carriles 7-10). Sin embargo, en las muestras procedentes de células transfectadas con la construcción que contiene la diana G2 a 13 residuos antes que la R295 puede observarse, además de la banda correspondiente a formas monoglicosiladas de la proteína, la aparición de una banda tenue de proteína de mayor peso molecular que corresponde a formas doblemente glicosiladas (Figura 43B, carril 12). Estas formas doblemente glicosiladas son mucho más evidentes en las muestras procedentes de las células transfectadas con la construcción en la que la diana G2 está 15 residuos antes de la R295, lo que fue de nuevo corroborado al tratar las muestras con Endo H (Figura 43B, carriles 13 y 14). Estos resultados demuestran que el segmento TM S6 del dominio DI se inserta en la membrana con la arginina de la posición 295 (R295) situada entre la interfase y el núcleo hidrofóbico de la membrana, de acuerdo con las predicciones del algoritmo ΔG Predictor.

Tabla 16. Secuencias diseñadas para el estudio del posicionamiento de S6 en membranas del RE. Constructos Lep-S6 con dianas glicosilables NTS (G2) y NAT (G3) en rojo. La región estudiada de S6 se muestra en verde y está flanqueado por los tetrapéptidos 'aislantes' GGPG y GPGG.

Constructo	Secuencias
Lep+S6	QETKENGIRLSETSGGPG RWRSYFYFITLIFFLAWLVKNVFI AVIIGPGG GVPGQQ NAT WIVPPG
Lep+N9+S6	QETKENGIR NTS ETSGGPG RWRSYFYFITLIFFLAWLVKNVFI AVIIGPGG GVPGQQ NAT WIVPPG
Lep+N11+S6	QETKENGIR NTS ETSGGPG RWRSYFYFITLIFFLAWLVKNVFI AVIIGPGG GVPGQQ NAT WIVPPG
Lep+N13+S6	QETK NTS IRLSETSGGPG RWRSYFYFITLIFFLAWLVKNVFI AVIIGPGG GVPGQQ NAT WIVPPG
Lep+N15+S6	QET NTS GIRLSETSGGPG RWRSYFYFITLIFFLAWLVKNVFI AVIIGPGG GVPGQQ NAT WIVPPG

Para corroborar la ubicación de la R295 realizamos simulaciones de dinámica molecular en las que observamos que este residuo se encuentra a una distancia de 21,41 Å del centro de la membrana compatible con una disposición interfacial, a diferencia del residuo D291 situado a 27,72 Å del centro de la membrana y por lo tanto, expuesto a un entorno acuoso (Figura 43C y D)

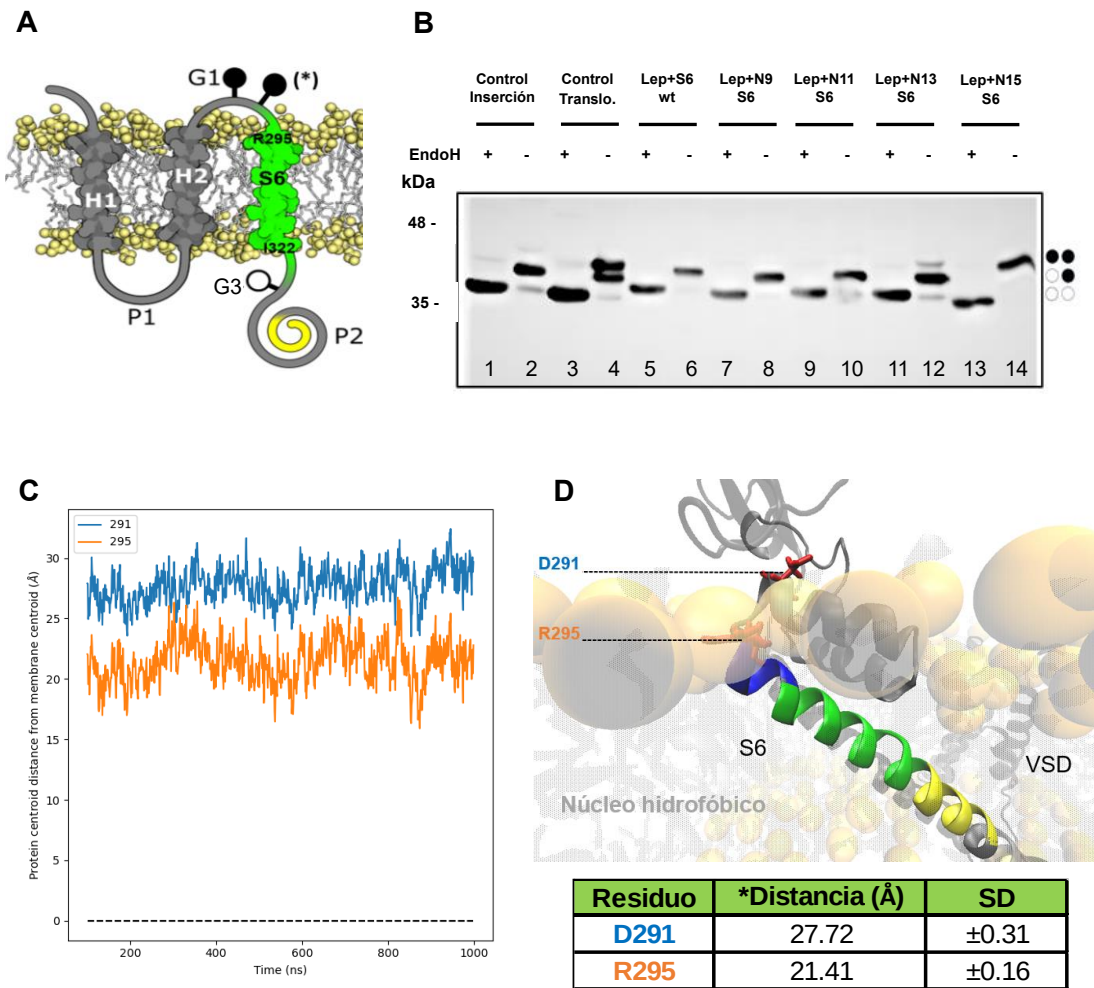
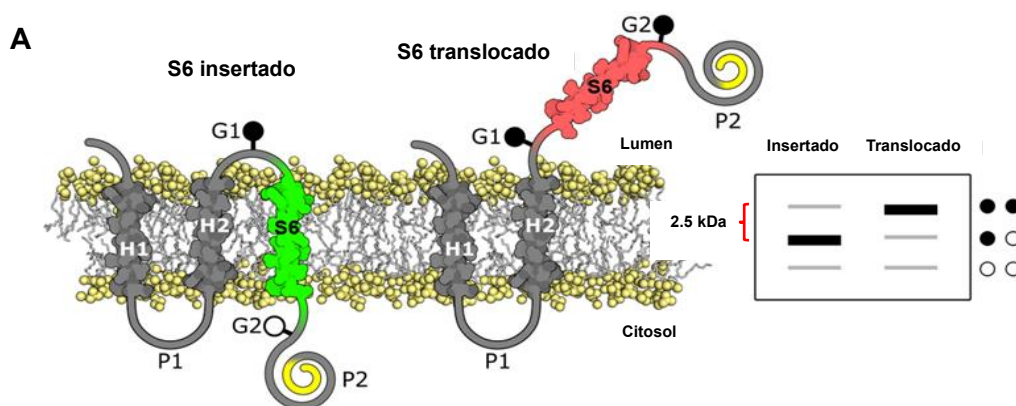


Figura 43. Posicionamiento del S6 del DI en las membranas del RE. (A) Representación esquemática del sistema Lep con sus segmentos TM H1 y H2 (ambos en gris) separados por el lazo P1. El asterisco (*) indica la posición de G2 a 9, 11, 13 y 15 aminoácidos de R295. Las dianas G1, G2 y G3 flanquean el S6 (en verde, residuos R295-I322) y el epítipo c-myc (amarillo) en el extremo Ct. (B) Se analizó por inmunotransferencia el extracto de las células Hek293T, transfectadas con las distintas construcciones. Los controles de inserción (KVLLVTGVLGLLLLIKWK) y de translocación (YNLYIYAFFASSLVHGFI) que contienen únicamente las dianas G1 y G3. Todas las muestras fueron divididas y tratadas con EndoH (carriles impares). Las proteínas no glicosiladas se indican con puntos blancos, las proteínas mono y doblemente glicosiladas con uno y dos puntos negros, respectivamente. (C) Gráfico que muestra la evolución temporal de la distancia de los residuos D291 y R295 hasta el centro geométrico de la membrana (línea punteada). Los valores positivos indican que los residuos estudiados están orientados hacia el lumen. (D) Vista del S6 del DI de NALCN que incluye PD (S5-S6) y VSD (S1-S4) en membrana (tomada de la simulación mediante VMD), la tabla indica los valores de las distancias medias (*) de cada residuo con respecto al centro de la membrana con su respectiva desviación estándar (SD).

R.8. Efecto de mutaciones CLIFAHDD en la inserción de S6 en membrana

Las variantes mutantes de NALCN están fundamentalmente relacionadas con dos síndromes: hipotonía infantil con retraso psicomotor y características faciales (IHPRF) y contracturas congénitas de extremidades y cara, hipotonía y retraso del desarrollo (CLIFAHDD) (Chong et al., 2015). Los efectos funcionales de dichas variantes han sido investigados previamente (Bouasse et al, 2019), sin embargo, el impacto molecular-estructural que ocasionan en NALCN sigue sin ser desvelado. Tanto los pacientes con IHPRF como con CLIFAHDD exhiben rasgos clínicos complejos de gravedad variable y síntomas comunes. Desafortunadamente, no hay un tratamiento disponible por lo que es necesario profundizar en el impacto de estas variantes mutantes a nivel estructural-funcional para desarrollar terapias eficaces. Para ello decidimos evaluar el efecto de algunas variantes mutantes asociadas a la enfermedad CLIFAHDD en la inserción en membrana de S6 utilizando el ensayo basado en la proteína Lep descrito en el apartado anterior. Para ello, en la secuencia de Lep en la que disponemos del segmento S6 flanqueado por dos dianas de glicosilación (Figura 44A) se diseñaron por mutagénesis dirigida los cambios encontrados en distintos pacientes (L312I, L312V, V313G, A319V e I322T). Como muestra la Figura 44B, las mutaciones estudiadas no afectan a la integración del S6 en membranas del RE; ya que tanto en la construcción que lleva la secuencia silvestre como en las distintas variantes que codifican las mutaciones individuales observamos formas monoglicosiladas de la proteína quimérica, lo que nos indica la correcta inserción en membrana de todas las secuencias mutantes. Como en anteriores experimentos, la identidad de la banda de mayor peso molecular fue verificada mediante el tratamiento de las muestras con la enzima EndoH (Fig. 44B, carriles impares).



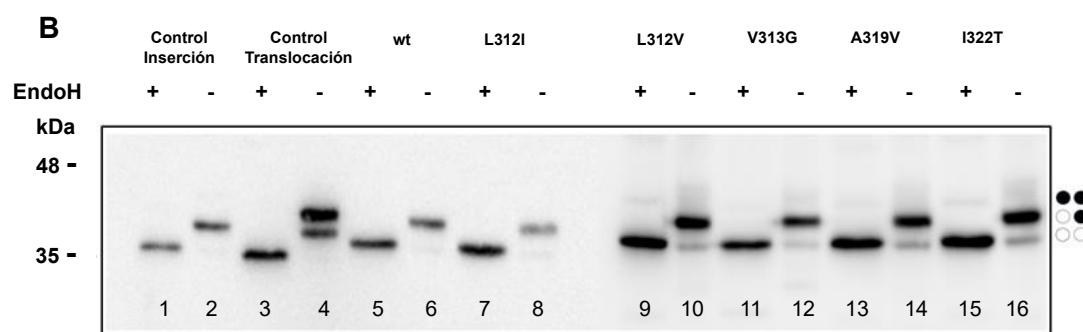


Figura 44. Efecto de mutaciones CLIFAHDD en la inserción de S6 en membranas de RE. (A) Representación esquemática del sistema Lep con sus segmentos TM H1 y H2 (ambos en gris) separados por el lazo P1. Las dianas de glicosilación (G1 y G2) flanquean el S6 (en verde si está insertado y en rojo si es translocado) y el epítipo c-myc (amarillo) se encuentra en el extremo Ct del dominio P2 (B) Las células Hek293T, transfectadas con las variantes indicadas, fueron lisadas para extraer las diferentes proteínas expresadas. La inmunodetección de estas proteínas se realizó con antisuero anti-c-myc, anticuerpo secundario y quimioluminiscencia. En paralelo se transfectaron variantes de Lep con secuencias del control de inserción (KVLLVTGVLGILLIKWK) y de translocación (YNLYIYAFFASSLVHGFI). Todas las muestras fueron divididas y una alícuota tratada con EndoH (+). Las proteínas no glicosiladas se indican con puntos blancos, las proteínas mono y doblemente glicosiladas con uno y dos puntos negros respectivamente.

R.9. Efecto de las mutaciones CLIFAHDD en la estructura del S6

Variación de contactos intra-hélice en el S6

Para estudiar el impacto de las mutaciones asociadas al síndrome de CLIFAHDD ubicadas en el S6 (L312V, L312I, V313G, A319V y I322T) se realizaron simulaciones de dinámica molecular utilizando el algoritmo CHARMM-GUI y el código PDB de la proteína NALCN resuelta por cryoEM (7cm3), diseñando las 5 mutaciones estudiadas. Las estructuras del dominio DI silvestre y de los cinco mutantes fueron insertadas en una bicapa de POPC (palmitoil oleoil fosfatidilcolina) pre-equilibrada. Después, ajustamos las posiciones de las moléculas en el sistema proteína-membrana mediante procesos de minimización y equilibración. Una vez las moléculas están posicionadas correctamente en el sistema, realizamos las simulaciones usando el programa GROMACS. El análisis de las simulaciones muestra que las mutaciones CLIFAHDD L312V, L312I y V313G tienen un efecto diferente, alterando los contactos o interacciones próximas en la región de la hélice que incluye desde la leucina 305 hasta la lisina 314 (L305-K314) (Figura 45). Así, respecto a la estructura de la proteína nativa, la mutación L312V incrementa el porcentaje de contactos en las parejas de residuos I306-F308, W311-F308, W311-L309 y K314-L309, y disminuye los contactos en las parejas W311-I306, V312-A310 y V313-A310. La mutación L312I incrementa los contactos en las

parejas de residuos L309-I306, A310-F307, W311-F308 y W311-L309, y los disminuye en A30-L305. Finalmente, observamos que la mutación V313G incrementa los contactos entre los residuos A310-L305 y W311-L309, disminuyéndolos entre A310-F307, G313-L309, L312-A310 y G313-A310. En general estas mutaciones tienen un mayor impacto en la vuelta de hélice que implica a los residuos W311, L309 y F308 (Figura 45).

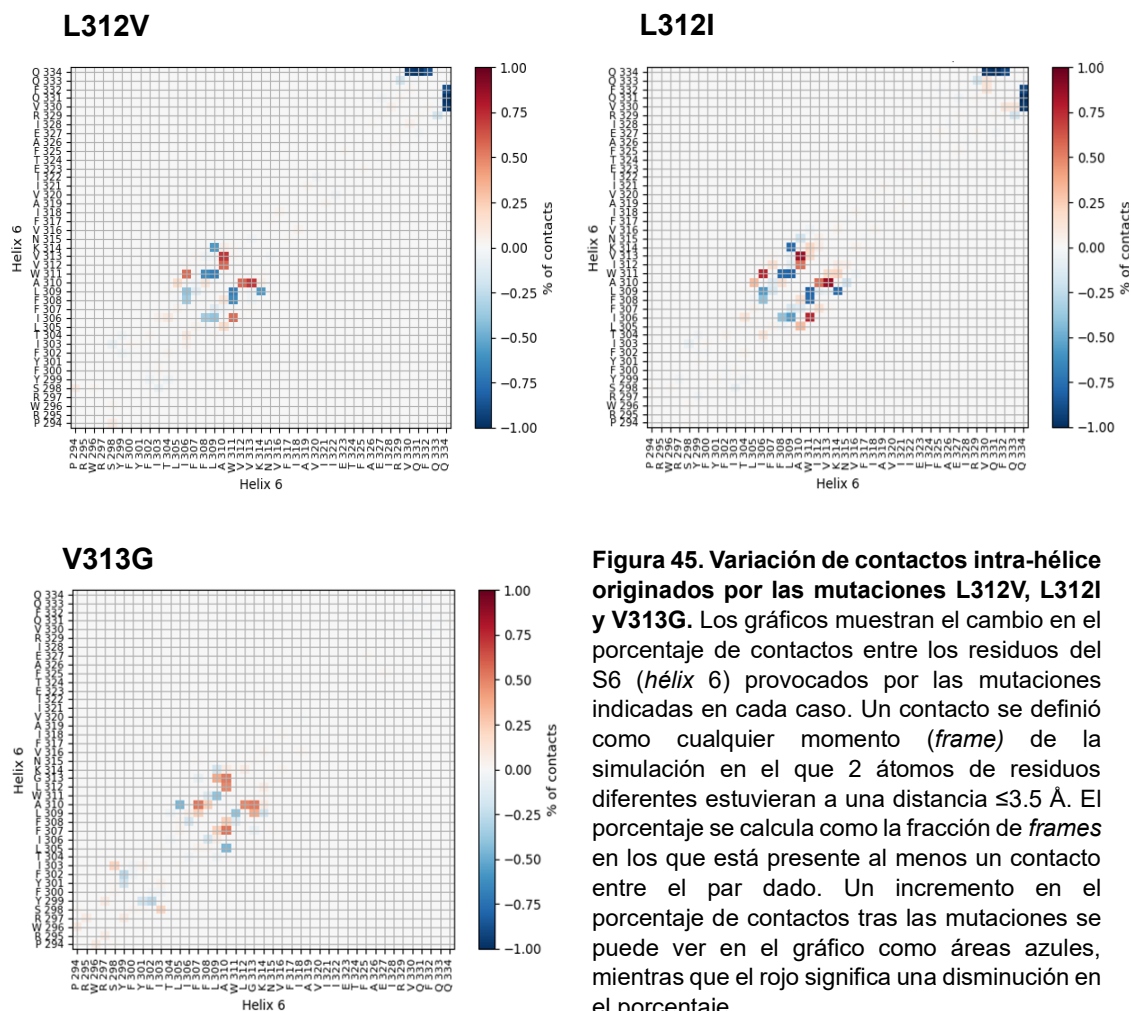


Figura 45. Variación de contactos intra-hélice originados por las mutaciones L312V, L312I y V313G. Los gráficos muestran el cambio en el porcentaje de contactos entre los residuos del S6 (*hélix* 6) provocados por las mutaciones indicadas en cada caso. Un contacto se definió como cualquier momento (*frame*) de la simulación en el que 2 átomos de residuos diferentes estuvieran a una distancia ≤ 3.5 Å. El porcentaje se calcula como la fracción de *frames* en los que está presente al menos un contacto entre el par dado. Un incremento en el porcentaje de contactos tras las mutaciones se puede ver en el gráfico como áreas azules, mientras que el rojo significa una disminución en el porcentaje.

Variación de los ángulos de torsión de los residuos W311, L309 y F308

Un cambio en la variación de contactos entre los aminoácidos de una misma cadena polipeptídica está estrechamente relacionado con un cambio de sus ángulos de torsión psi (Ψ) y phi (ϕ). El cambio de estos ángulos de torsión en muchas ocasiones no sólo genera nuevas interacciones con otros átomos del entorno sino también un movimiento o cambio de posición de las cadenas laterales R de los aminoácidos involucrados (Zacharias et al., 2013; Kumar y Bansal, 2015). Después de analizar y encontrar variación de contactos intra-helicoidales entre la secuencia del S6 wt comparada con sus mutantes (L312V, L312I y V313G), analizamos si esta variación de contactos genera un cambio en los ángulos de torsión psi (Ψ) y phi (ϕ) de los residuos W311, L309 y F308. Para ello, en las trayectorias de las simulaciones calculamos la zona de mayor densidad de ángulos de torsión para cada uno de estos residuos y asociamos la estabilidad de sus ángulos con valores de energía libre de Gibbs. De esta forma, encontramos zonas de mayor densidad y estabilidad de sus ángulos de torsión ϕ y Ψ tanto en la secuencia de S6 wt como en las secuencias mutantes.

Los resultados nos muestran que las mutaciones CLIFAHDD L312V, L312I y V313G generan un cambio considerable en la densidad y estabilidad de los ángulos de torsión de los residuos W311, L309 y F308. Las mutaciones L312V y L312I generaron una mayor modificación en los ángulos de torsión que los residuos W311, L309 y F308. La mutación V313G, aunque no modificó considerablemente la estabilidad de los ángulos de torsión de los residuos W311 y F308, generó una mayor dispersión de estos, lo cual se ve reflejado en sus gráficos de densidades. Estas modificaciones y otras más las podemos observar en los gráficos de densidades y de energía libre de Gibbs de las Figuras 46-48.

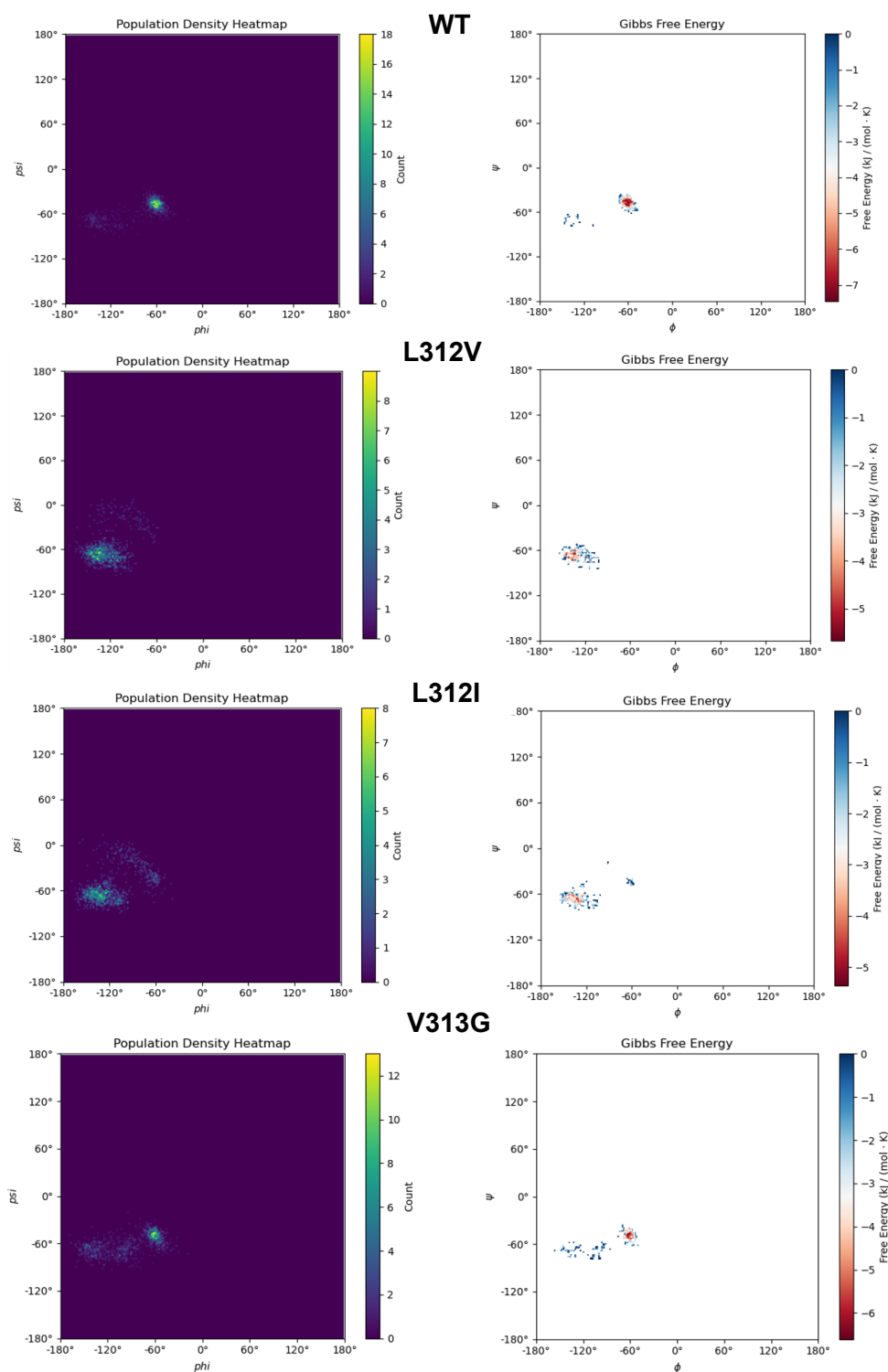


Figura 46. Efecto de las mutaciones L312V, L312I y V313G en los ángulos de torsión phi (ϕ) y psi (Ψ) del residuo W311. El gráfico de densidad de población muestra la distribución predominante (en amarillo) de los ángulos phi (rotación alrededor del enlace N-C α) y psi (rotación alrededor del enlace C α -C β) del residuo 311 del S6 durante la simulación. A la derecha, el gráfico de energía libre de Gibbs muestra la estimación de la energía libre (kJ/mol·K) asociada a los ángulos phi y psi del residuo W311. La energía libre fue calculada con $\Delta G = -kT \ln(p)$; donde k es la constante de Boltzmann, T es la temperatura del sistema de simulación 310K, p es la distribución de densidades. En rojo se muestra la zona de mayor estabilidad de los ángulos de torsión y en azul la de menor estabilidad.

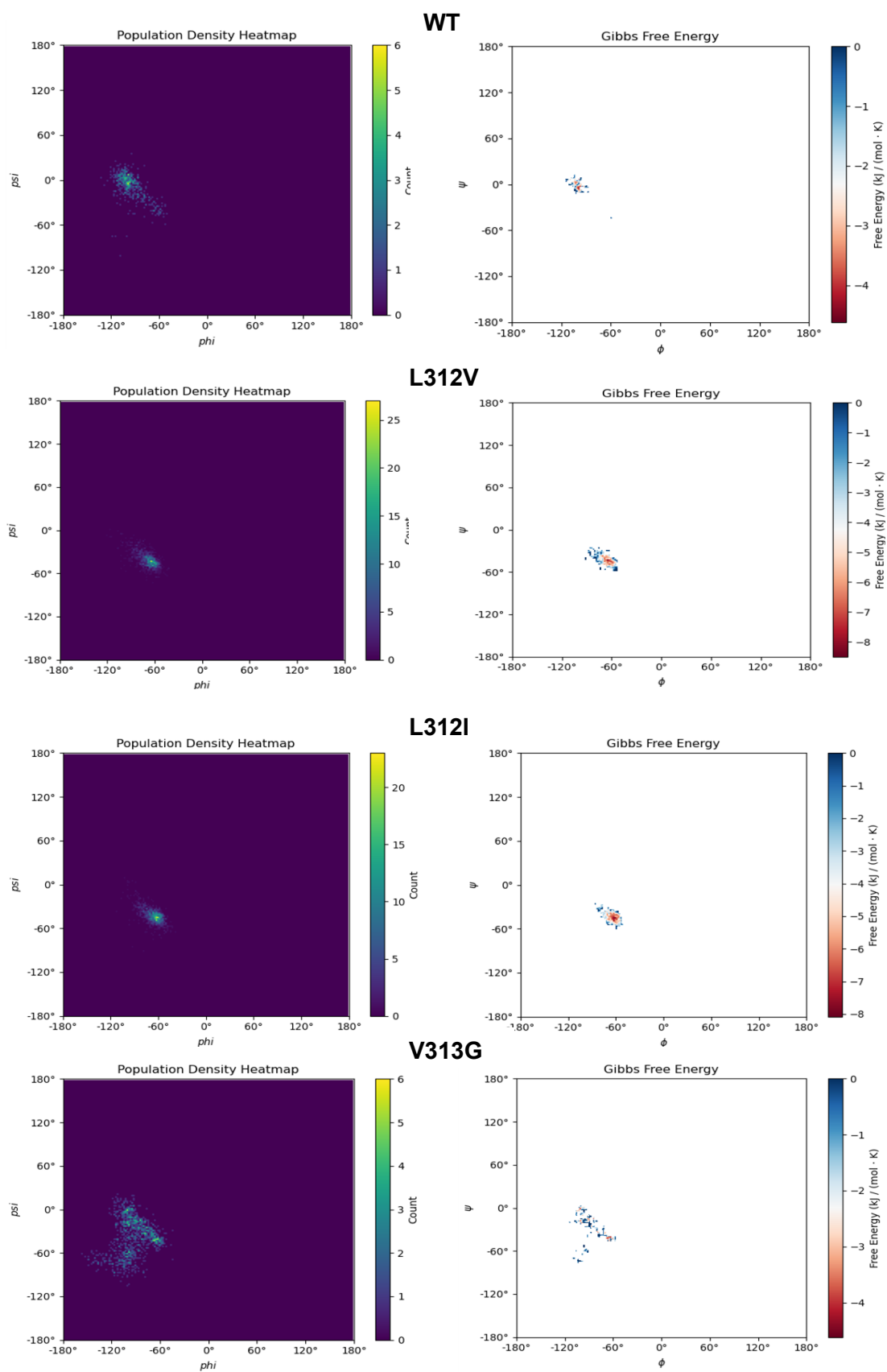


Figura 47. Efecto de las mutaciones L312V, L312I y V313G en los ángulos phi (ϕ) y psi (ψ) del residuo L309

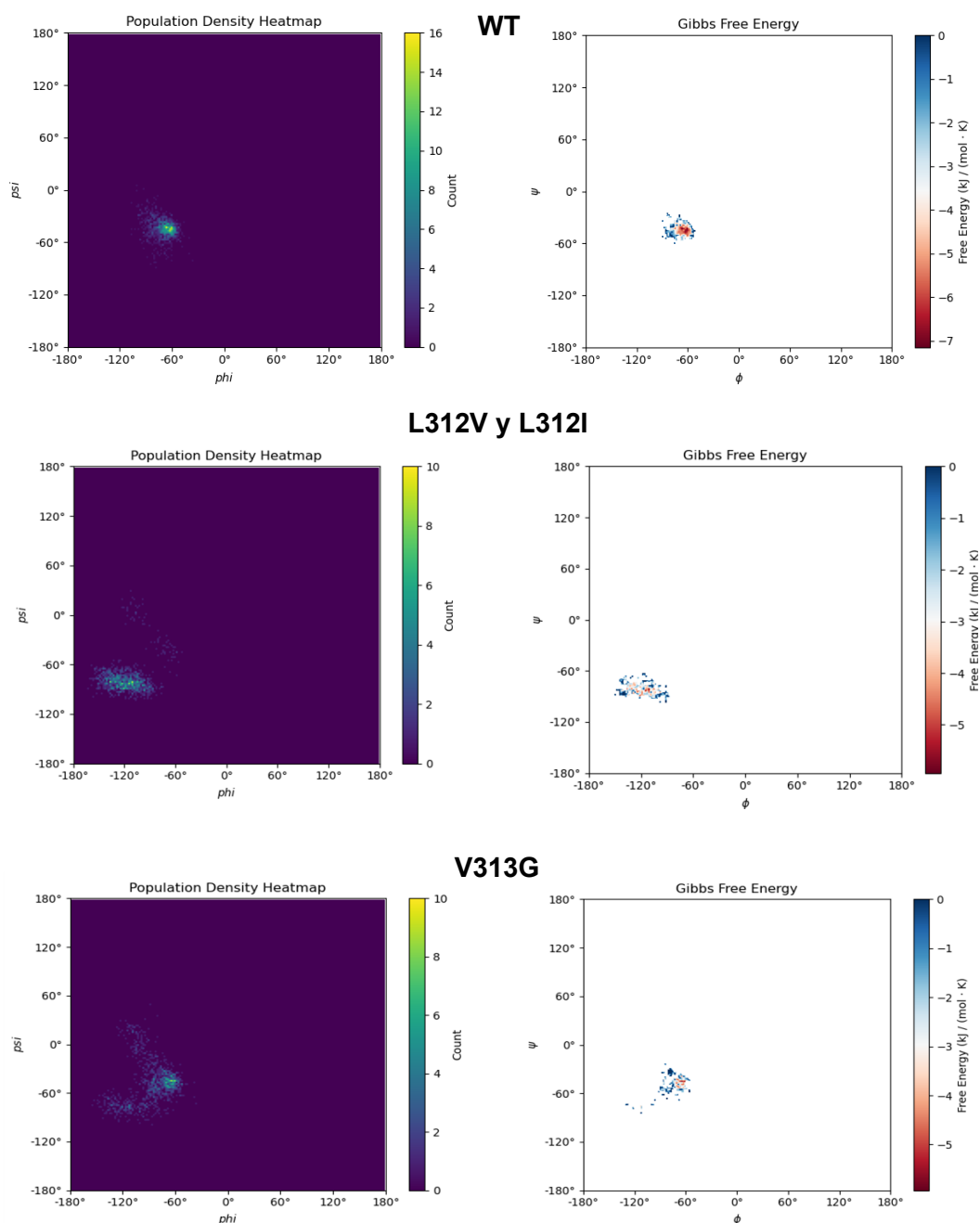


Figura 48. Efecto de las mutaciones L312V, L312I y V313G en los ángulos phi (ϕ) y psi (Ψ) del residuo F308

La variación de contactos y modificación de los ángulos de torsión entre los residuos de una hélice está asociada con un cambio de conformación secundaria de ésta, y dicho cambio está relacionado con la apertura de los canales de las proteínas transportadoras de iones, entre ellas NALCN (Kschonsack et al., 2020; McGoldrick et al., 2018). En NALCN se ha descrito que los residuos W311, L309 y F308 del S6 del DI estabilizan la estructura π del canal abierto mediante la formación de enlaces puentes de hidrógeno intra-hélice ($i, i+5$) entre F308-V313 y L309-K314 e inter-hélice entre W311-L1439 (Xie et al., 2020). Por tal motivo, investigamos si los cambios antes descritos en estos

residuos generan una alteración conformacional en la región L305-K314 del S6 del dominio DI. Nuestros hallazgos principalmente muestran que las mutaciones CLIFAHDD L312V, L312I y V313G generan una alteración de la distancia por vuelta de la hélice α y por ende, un mayor acercamiento entre los residuos F308-V313 y L309-K314 (Figura 49A). Así, en la secuencia nativa S6 (WT) la distancia media de 3,41 Å, entre el oxígeno del grupo carbonilo C=O (C_β) de F308 y el hidrógeno del grupo amino N-H de V313, disminuye a 2,22, 2,52 y 2,63 Å respectivamente, en las secuencias mutantes (L312V, L312I y V313G). De igual forma, la distancia media de 5,58 Å entre el oxígeno del C=O (C_β) de L309 y el hidrógeno del grupo amino N-H de K314 de la secuencia nativa disminuye a 4,09, 3,5 y 4,03 Å respectivamente, en las secuencias mutantes (L312V, L312I y V313G). Con respecto al residuo W311, estas tres mutaciones CLIFAHDD generaron un cambio posicional oscilante de su cadena lateral R.

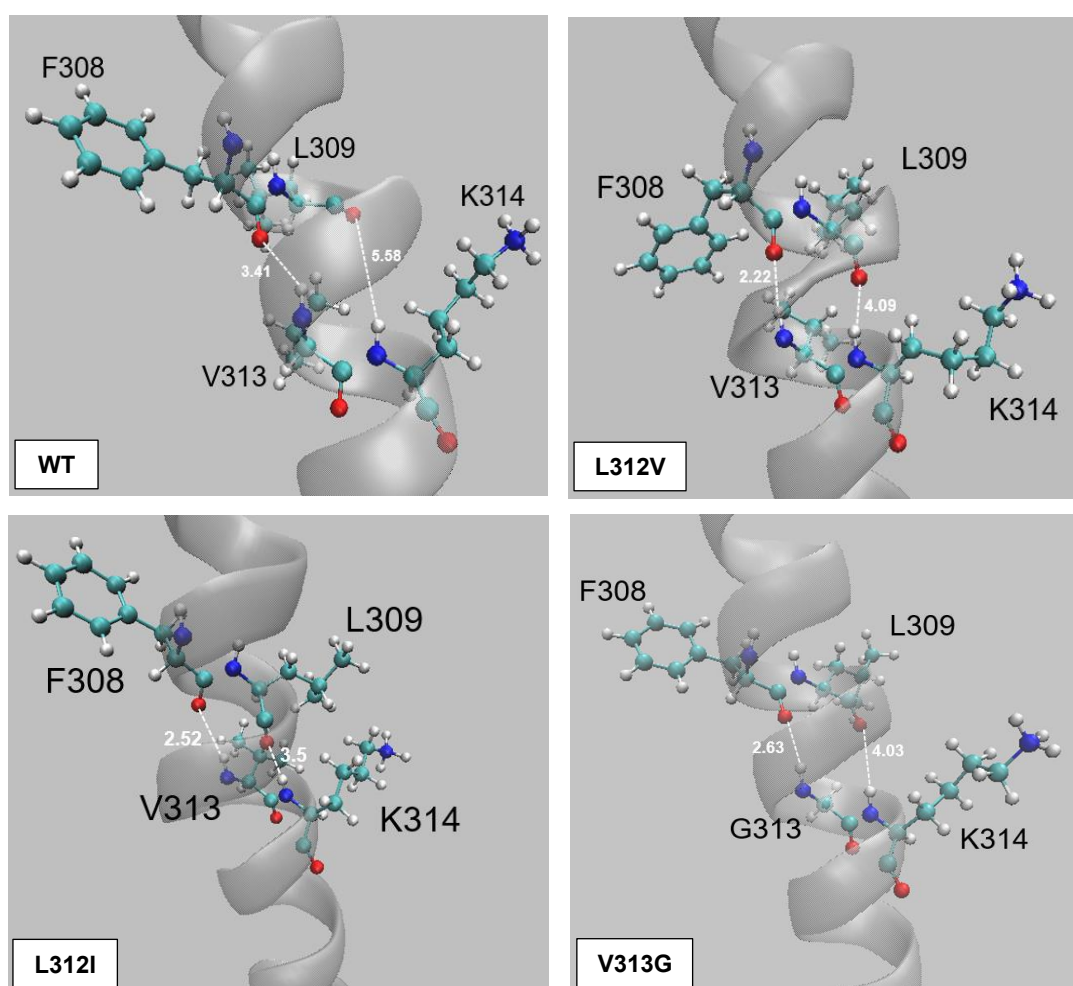


Figura 49. Efecto de las mutaciones L312V, L312I y V313G en el acercamiento de los residuos 308-313 y 309-314. El gráfico nos muestra la posición de los residuos 308, 309, 313 y 314 en la hélice S6 wt y mutante (L312V, L312I y V313G). La línea blanca discontinua nos indica la distancia media entre el oxígeno (rojo) de los residuos 308 (i) y 309 (i) y el hidrógeno (blanco) de los residuos 313 $(i+5)$ y 314 $(i+5)$.

Al analizar todos los valores de distancias de los residuos 308-313 y 309-314 calculados durante la simulación, descubrimos que la mediana, la moda y la distribución de estos son diferentes en la secuencia silvestre (WT) y en las secuencias mutadas (Figura 50). Notándose que las mutaciones CLIFAHDD L312V, L312I y V313G aproximan los grupos carbonilos y aminos de los residuos 308-313 y 309-314, lo cual podría facilitar la formación de puentes de hidrógeno entre ellos.

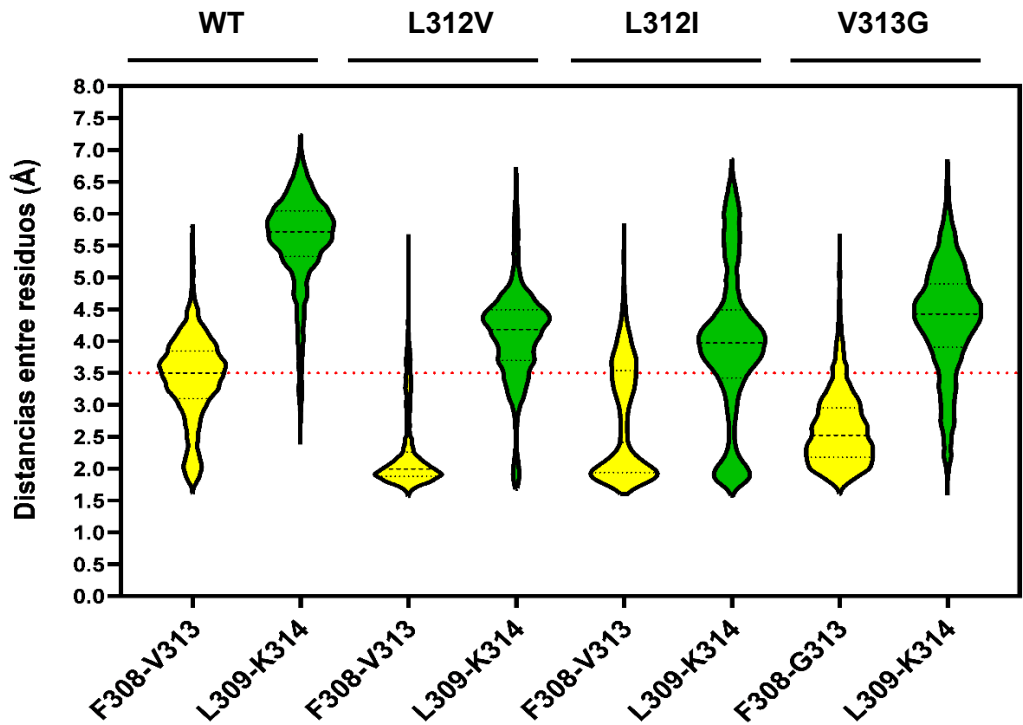


Figura 50. Análisis de las distancias entre residuos 308-313 y 309-314 de la secuencia WT y de las secuencias mutantes (L312V, L312I y V313G). El gráfico muestra la distribución de las distancias entre N-H_(i+5)-C=O_(i) para cada pareja de residuos indicada. La distancia de corte de 3.5 Å se indica en línea punteada de rojo. La mediana de cada diagrama se indica en líneas negras discontinuas. En amarillo se indica la distribución de distancias de la pareja de residuos 308-313 y en verde, la de los residuos 309-314.

La estadística descriptiva nos muestra que los valores de distancias medias y mediana de los datos de 308-313 y 309-314 en las secuencias mutantes son menores que las encontradas en la secuencia WT. Asimismo, se aprecia que la variación de las distancias entre los residuos es mayor en la secuencia mutada que en la secuencia WT, lo cual se observa en el coeficiente de variación (Tabla 17).

Tabla 17. Estadística descriptiva de las distancias entre los residuos 308-313 y 309-314.

	WT		L312V		L312I		V313G	
Parámetro	F308-V313	L309-K314	F308-V313	L309-K314	F308-V313	L309-K314	F308-V313	L309-K314
Mediana	3,49	5,72	1,99	4,18	2,41	3,97	2,52	4,42
Media	3,41	5,58	2,22	4,09	2,52	3,5	2,63	4,03
Desv. Estándar	0,67	0,71	0,59	0,73	0,87	1,11	0,57	0,85
Coef. Variación	19,64%	12,62%	26,53%	17,97%	31,92%	29,69%	21,87%	19,59%

Efecto de la mutación I322T en el S6 del DI de NALCN

Mediante simulación molecular, observamos cómo cambia la posición del residuo mutante T322 en comparación con el residuo original I322. Encontramos que la mutación I322T altera posición de este residuo con respecto al centro de la membrana. El residuo I322 se encuentra más cerca del centro de la membrana, a unos $-6.1 \text{ \AA}$ en promedio, mientras que el residuo mutante T322 está más lejos, a $-10.74 \text{ \AA}$ en promedio. El análisis estadístico nos muestra diferencias significativas entre las distancias medias de ambos residuos con respecto al centro de la membrana. Asimismo, la observación de estos residuos durante las simulaciones nos muestra que el residuo T322 tiende a desplazarse hacia el interior de la célula, hacia el citoplasma (Figura 51).

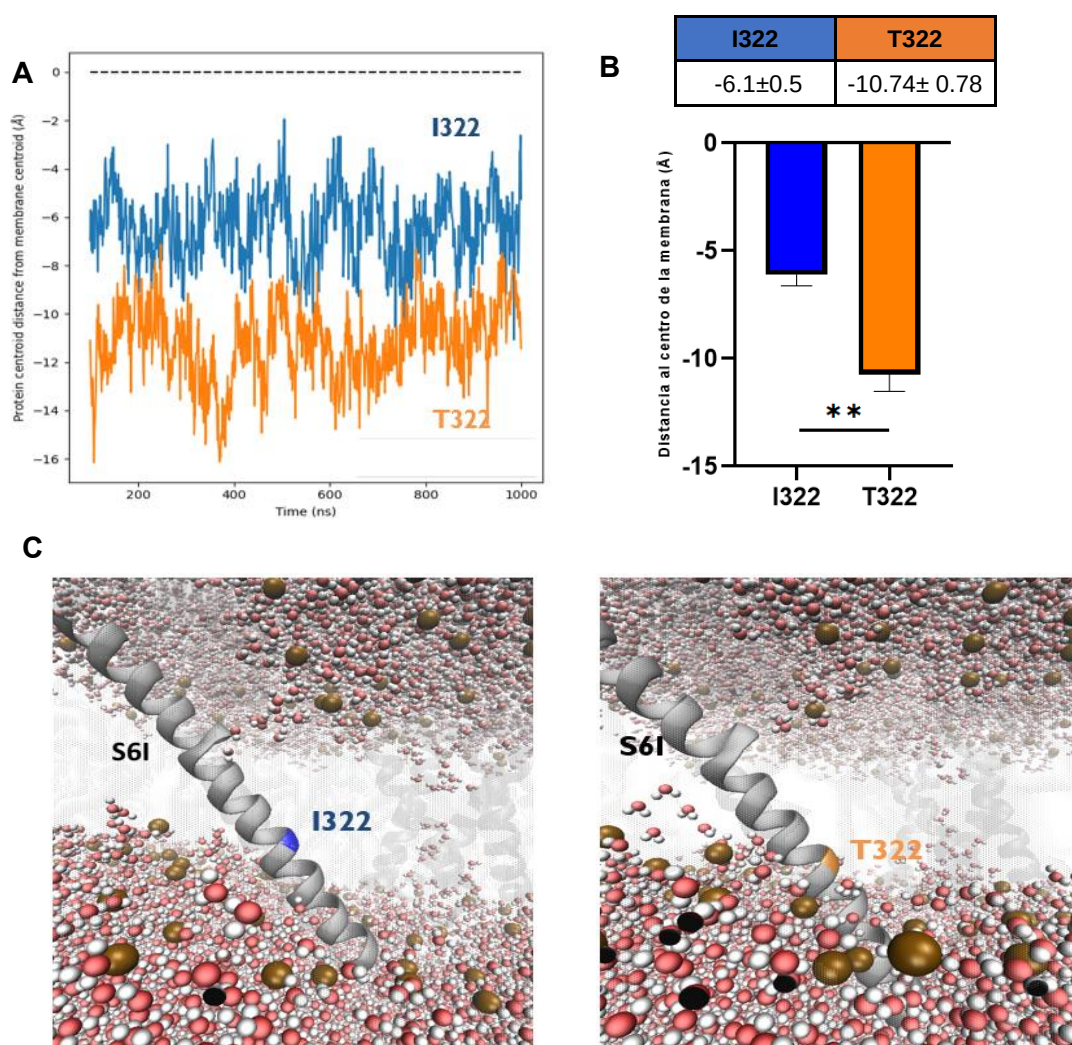


Figura 51. Cambio de posición del residuo 322 en membrana. (A) Gráfico que muestra la evolución temporal de la distancia desde el centro geométrico del residuo 322 hasta el centro geométrico de la membrana. Los valores negativos indican que el centro geométrico del residuo está por debajo del centro de la membrana, hacia el compartimento citosólico. (B) Comparación de las distancias medias de los residuos I322 y T322 con respecto al centro de la membrana. Los asteriscos indican diferencias significativas ($p\text{-value} < 0.05$). (C) Visualización en VMD de las trayectorias de los residuos I322 y T322 en membrana durante las simulaciones. El I322 se representa en azul y el T322 en anaranjado.

D.1. Discusión del capítulo I

NALCN es una proteína con 24 segmentos TM agrupados en cuatro dominios de estructura similar (Kschonsak et al., 2020; Xie et al., 2020), de los cuales el dominio DI de NALCN acumula aproximadamente la tercera parte de las mutaciones asociadas al síndrome de CLIFAHDD (Montiel et al., 2024), siendo su estudio especialmente interesante.

Como punto de partida se emplearon programas de predicción de segmentos TM y topología de proteínas de membrana, identificándose seis segmentos TM. El programa TMHMM, que además de identificar los fragmentos TM es capaz de predecir topología, localiza el extremo amino terminal y el carboxilo terminal del dominio DI en el citosol. Estas predicciones de estos programas coinciden con la estructura de NALCN recientemente publicada (Kschonsak et al., 2020). Algunos de los segmentos TM tanto del dominio DI de NALCN como de los dominios equivalentes de otros canales de sodio y calcio tienen valores positivos de energía libre para su inserción, lo cual sugiere que los segmentos TM menos hidrofóbicos de NALCN, Nav1.4 y TRPV6 podrían no ser reconocidos y eficientemente insertados por el translocón en la membrana del RE. Sin embargo, estos segmentos tienen una conformación helicoidal TM en las estructuras resueltas de estas tres proteínas de membrana (Chua et al., 2020; Pan et al., 2018, McGoldrick et al., 2018). Por ello, nos propusimos estudiar en los estadios iniciales de la síntesis de la proteína cómo se realiza la inserción individual en membrana de cada uno de los seis segmentos TM del dominio DI de NALCN. Así, tras la traducción *in vivo* observamos que los segmentos S1, S2 y S4 aislados del contexto del resto de la secuencia del polipéptido no se integraron en la membrana del RE, mientras los segmentos S3, S5 y S6 fueron eficientemente insertados. Estos resultados coinciden con las predicciones de los algoritmos utilizados; ya que los TM poco hidrofóbicos no se integran de manera estable en membrana.

Sabiendo que la mitad de los segmentos TM del dominio DI de NALCN no se integran individualmente en la membrana del RE, nos planteamos como hipótesis de trabajo la posibilidad de que, como se ha demostrado en otras proteínas de membrana (Bañó-Polo et al., 2017; Hedin et al., 2010; Tamborero et al., 2011), la inserción de estos segmentos marginalmente hidrofóbicos dependa de su contexto de secuencia. Para abordar esta hipótesis experimentalmente evaluamos la biogénesis de diferentes polipéptidos truncados del dominio DI de NALCN tanto en bacterias como en células humanas. La síntesis de estos truncados nos permite tener información de la biogénesis de la proteína paso a paso. Los resultados del sistema del indicador dual en bacterias y

del mapeo de glicosilaciones en células eucariotas son coincidentes y nos muestran que los polipéptidos truncados más pequeños, que incluyen los segmentos S1 y S2, no se insertan correctamente en la membrana a pesar de que, como hemos mencionado antes, tienen una disposición transmembranal en las estructuras resueltas (Kschonsak et al., 2020; Xie et al 2020). Sin embargo, en los polipéptidos truncados más largos, concomitante con la presencia de la tercera hélice TM (S3), las formas truncadas de la proteína insertan los sucesivos segmentos TM de forma secuencial y con la topología nativa. Estos resultados sugieren un papel destacado del segmento S3 en los estadios iniciales de la biosíntesis de NALCN. Los bajos niveles de glicosilación (Figura 30) observados en nuestros estudios del truncado 65-mer (que contiene solo la secuencia del primer segmento TM, S1) probablemente se deben a que la secuencia S1 no es correctamente detectada por la SRP debido su baja hidrofobicidad impidiendo/dificultando el adecuado direccionamiento a la membrana del complejo ribosoma/cadena nascente (RNC). El mismo efecto podría tener lugar en el caso del truncado 97-mer, como se deduce de su baja eficiencia de glicosilación en las variantes diseñadas (Figura 31). A diferencia de lo que ocurre en los dos casos anteriores, el truncado 136-mer presenta una mayor eficiencia de glicosilación (Figura 32) lo que sugiere un mejor direccionamiento del RNC hacia la membrana del RE y demuestra una orientación de su extremo C-terminal hacia el lumen, acorde con las predicciones y con la estructura resuelta de NALCN (Kschonsak et al., 2020). Para la inserción de este truncado (136-mer) el translocón Sec61 reconoce a S3 y lo integra en la membrana del RE, mientras que los segmentos S1 y S2 del truncado podrían ser reconocidos por los dominios citosólicos del complejo EMC quedando retenidos hasta la inserción cotraduccional en membrana del S3, como se ha sugerido recientemente para la inserción de otras proteínas de membrana (McDowell et al., 2021), en las que se ha observado la participación en conjunto de ambos complejos insertasa (EMC y Sec61) en la inserción en membrana de proteínas politópicas con segmentos TM de baja hidrofobicidad (Hegde y Keenan, 2022; O'Keefe et al., 2021).

La participación de varios complejos insertasa también podría explicar la inserción del truncado 163-mer, en el que tres de sus cuatro segmentos (S1, S2 y S4) son poco hidrofóbicos. Recientemente, se han publicado revisiones sobre los complejos EMC, TMC y PAT y su papel en la biogénesis de las proteínas politópicas. Cada uno de estos complejos puede asociarse con el translocón Sec61 para facilitar la integración en membrana de regiones TM poco hidrofóbicas (O'Keefe et al., 2021; Hegde y Keenan, 2022). Así que es muy posible que los segmentos menos hidrofóbicos o con residuos cargados (S4) del dominio DI de NALCN permanezcan asociados a estos complejos

hasta que los segmentos más hidrofóbicos (S3, S5 y S6) emerjan del ribosoma. Es interesante mencionar que la inserción y/o el ensamblaje del dominio sensor de voltaje (VSD) que incluye los cuatro primeros segmentos TM (S1-S4) ocurra de forma independiente de la inserción del dominio del poro (PD) que incluye la horquilla formada por los segmentos S5 y S6, como se ha demostrado anteriormente en proteínas canal (Doñate-Macian et al., 2015). Para que esto ocurra es probable que las interacciones inter-hélice entre los cuatro primeros segmentos TM (S1-S4) tengan un papel relevante.

Para investigar las interacciones a nivel atómico entre estos cuatro segmentos TM, realizamos simulaciones de dinámica molecular evaluando las interacciones entre S1-S2, S1-S4, S2-S3, S2-S4 y S3-S4. Esto nos permitió identificar contactos o interacciones entre los segmentos más próximos en las estructuras resueltas. Destacamos principalmente la presencia de enlaces de hidrógeno y un puente salino entre las cadenas laterales de los residuos S52, D75, S121 y R146. Cabe mencionar que nuestros resultados describen por primera vez la presencia del puente salino entre D75-R146 y su importancia en el plegamiento y estabilidad en membrana del VSD del dominio DI de NALCN, tal y como ha sido demostrado en interacciones electrostáticas de otras proteínas (Barber et al., 2012; Groome et al., 2020). Asimismo, nuestros resultados *in silico* muestran la corta distancia de interacción entre las cargas de activación de S4 (R143 y R146) y los residuos del centro de transferencia de carga de los segmentos S1 (S49 y S52), S2 (D75, Y82 y E85) y S3 (D111) y comprueban la formación de interacciones entre ellos; hecho que fue sugerido al analizar la estructura resuelta de NALCN por la proximidad observada entre estos residuos observados (Kschonsak, et al 2020; Xie et al., 2020).

La proximidad observada entre estos y otros aminoácidos hace probable que estos contactos favorezcan el empaquetamiento de los cuatro primeros segmentos TM (S1-S4) y contribuyan a la inserción, plegamiento y estabilidad en membrana del VSD, de forma similar a lo observado en las interacciones entre segmentos TM de otras proteínas de membrana descritas (Barber et al., 2012; DeGrado et al., 2003; Adamian y Liang 2001). Además, se ha descrito que estas interacciones desempeñan un papel importante en el estado de reposo y apertura de canales de sodio, calcio y potasio para mantener funciones claves como la transmisión sináptica en las neuronas y la contracción del corazón y los músculos esqueléticos (Fernández-Quintero et al., 2021; Groome et al., 2020). Por otro lado, los alineamientos de secuencias de las proteínas NALCN, Nav 1.2 y Cav 1.1 muestran que los residuos D75, S121 y R146 se encuentran conservados, lo que sugiere que desempeñan un papel relevante en las estructuras de

dichas proteínas (ver Anexo 4A, Figura 77). En el caso específico del residuo R146, su mutación se ha encontrado en pacientes con cáncer en los que se ve alterada la sensibilidad al voltaje de la proteína NALCN (Chua et al., 2020). Según nuestros resultados, esta alteración podría atribuirse a la pérdida de interacciones del residuo mutante R146Q con los residuos D75 y S121 (Figura 41). Por lo que respecta a las mutaciones asociadas al síndrome de CLIFAHDD en la integración a membrana del S6, nos planteamos estudiar el posicionamiento del S6 en membrana dado que en esta hélice recaen varias de estas mutaciones. Los algoritmos de predicción utilizados en este trabajo no llegan a un consenso sobre la posición de éste en la membrana (Tablas 7 y 8). Por ello decidimos investigar si la región embebida en la membrana comienza en el residuo R295 como predice el ΔG Predictor (Tabla 7) o en la F300 como predice TMHMM (Tabla 8). La técnica de mapeo por glicosilación permite averiguar con precisión a nivel de residuo la distancia que separa el residuo de asparagina receptor de un sitio de glicosilación de la interfase de la membrana (Orzáez et al., 2004; Nilsson et al., 1993). Nuestros resultados *in vivo*, muestran una eficiente glicosilación de la asparagina situada en la posición equivalente al residuo E280 en la secuencia de NALCN, es decir a 15 residuos de distancia del inicio de la región TM predicha por el algoritmo ΔG Predictor (R295). Para contrastar estos resultados con la estructura de NALCN depositada en las bases de datos, simulamos por dinámica molecular esta región del dominio DI y observamos que S6 presenta una región hidrofóbica inusualmente larga que se encuentra embebida en la membrana desde el residuo R295 hasta el residuo R329, lo que obliga a esta hélice disponerse inclinada en el núcleo hidrofóbico de la membrana con cierto ángulo de inclinación, excediendo en longitud las predicciones de los algoritmos, lo que pone de relieve la importancia de verificar experimentalmente las predicciones computacionales.

En el análisis *in silico* de la estructura de la proteína NALCN se puede apreciar que sus cuatro segmentos S6_{I-II-III-IV} presentan una región central de 7 aminoácidos que participan en la transición de hélice α a hélice π durante la apertura del dominio del poro central (Xie et al., 2020). En el S6 del dominio DI tal secuencia implica los residuos ³⁰⁸FLAWLVK³¹⁴, siendo esta secuencia similar a las reportadas en otros S6 de canales transportadores de iones (Pan et al., 2018, McGoldrick et al., 2018) (Figura 77; Anexo 4B, Figura 78), además la frecuencia y posición de estos mismos residuos se han descrito en otras hélices π (Fodje et al., 2002; Kumar et al 2015). La presencia de estos residuos y su implicación en el cambio conformacional sugiere que esta transición podría regular la apertura del poro del canal NALCN y con ello el paso de los iones Na⁺. Además, en este segmento TM S6 se acumulan mutaciones asociadas al síndrome de

CLIFAHDD (Tabla 3) y a pesar de que nuestros resultados experimentales muestran que las mutaciones L312V, L312I y V313G no alteran la integración de S6 en la membrana, nuestros cálculos de dinámica molecular muestran un nuevo perfil de interacciones y un cambio en los ángulos de torsión en los residuos de esta misma región (³⁰⁸FLAWLVK³¹⁴). Los cálculos computacionales nos permiten observar que las mutaciones L312V, L312I y V313G ocasionan un acercamiento entre los residuos de esta región, principalmente entre los residuos 308-313 y 309-314, generando la formación de un nuevo perfil de interacciones en la región mutada. No sabemos qué efecto tiene esto en la sensibilización y/o apertura del canal de NALCN; pero se ha descrito que los residuos F308, L309 y W311 estabilizan la estructura π del canal abierto mediante interacciones intra-hélice e inter-hélice (Xie et al., 2020). Además, si observamos en las secuencias mutantes la distribución de los ángulos de torsión ϕ y Ψ de los residuos W311, L309 y F308, notaremos que tienen valores semejantes a los ángulos ϕ y Ψ de residuos que conforman una hélice π y que han sido reportadas en otras proteínas (Cooley et al., 2010; Fodje et al., 2002; Kumar et al., 2015). Todo parece indicar que las mutaciones L312V, L312I y V313G propician un nuevo orden de interacciones y ángulos de torsión que estabilizan la región ³⁰⁸FLAWLVK³¹⁴ en una conformación que podría favorecer más fácilmente su transición a hélice π y/o la apertura del canal. Tal afirmación se sustenta con los descubrimientos publicados en otras proteínas en donde algunas hélices π son generadas por la inserción o mutación de un aminoácido en una región conservada de una hélice α y este cambio conformacional también altera la funcionalidad de la proteína (Cooley et al., 2010; Duneau et al., 1996; Weaver et al., 2000). Hasta ahora, en NALCN no se ha descrito que el cambio aminoacídico genere una estructura π ; sin embargo, las mutaciones CLIFAHDD en esta región causan hiperactividad neuronal a través de una ganancia de función en NALCN; ya que se altera su sensibilidad al voltaje (Aoyagi et al., 2015; Bend et al., 2016; Kschonsak et al., 2020; Monteil et al., 2024). Tal alteración funcional debido al cambio conformacional ($\alpha \rightarrow \pi$) generado por mutaciones se ha descrito también en otras proteínas de membrana (Zubcevic et al., 2019; Kim et al., 2007; Segatto et al., 1988).

No se observaron efectos estructurales de la mutación A319V, tal vez porque el residuo A319 está alejado de los residuos centrales (³⁰⁸FLAWLVK³¹⁴) implicados en la estabilización de la estructura π del S6 del dominio DI de NALCN. Sin embargo, al comparar las secuencias de las proteínas del DI de NALCN, Nav 1.2 y Cav 1.1 identificamos que la alanina en posición 319 de NALCN al igual que la asparagina 315 están conservadas en los segmentos S6 de estas proteínas y TRPV6 (ver Anexo 4 A y B, Figura 77 y 78). Revisando las trayectorias de los átomos en la región ³¹⁵NVFIIV³²²

de la secuencia nativa de S6 encontramos que en los residuos de dicha región existe una red de contactos cercanos mediante la formación de enlaces; y que el residuo mutante V319 registra distancias de rango corto de interacción con el residuo V316 (resultados no mostrados). Este cambio también podría estar asociado con la alteración de la sensibilidad al voltaje y la funcionalidad de la proteína NALCN como se ha descrito para esta mutación (Kschonsak et al., 2020; Lozic et al., 2016); pero tal descubrimiento requiere una mayor revisión. Por otro lado, la mutación I322T genera un cambio de posición significativo de este residuo con respecto al centro de la membrana, adoptando una ubicación más cerca al citosol. Tal cambio de posición podría ocasionar una desestabilización de interacciones inter-hélice entre residuos que conforman dicha región llamada puerta de entrada intracelular del canal de NALCN; ya que tales residuos hidrofóbicos (I318, V595, V1152, V1443, I322, L599, I1156 y V1447) están muy conservados en las proteínas transportadoras de iones (Xie et al., 2020) y también en el segmento S6 del dominio DI y en la proteína TRPV6 (ver Anexo 4 A y B, Figura 77 y 78). Finalmente, esperamos que todos estos resultados contribuyan a conocer un poco más sobre el efecto de las mutaciones CLIFAHHD en la proteína NALCN y también, creemos necesario continuar investigando este tipo de mutaciones y otras más asociadas a enfermedades mediante el uso de una línea multidisciplinar complementaria.

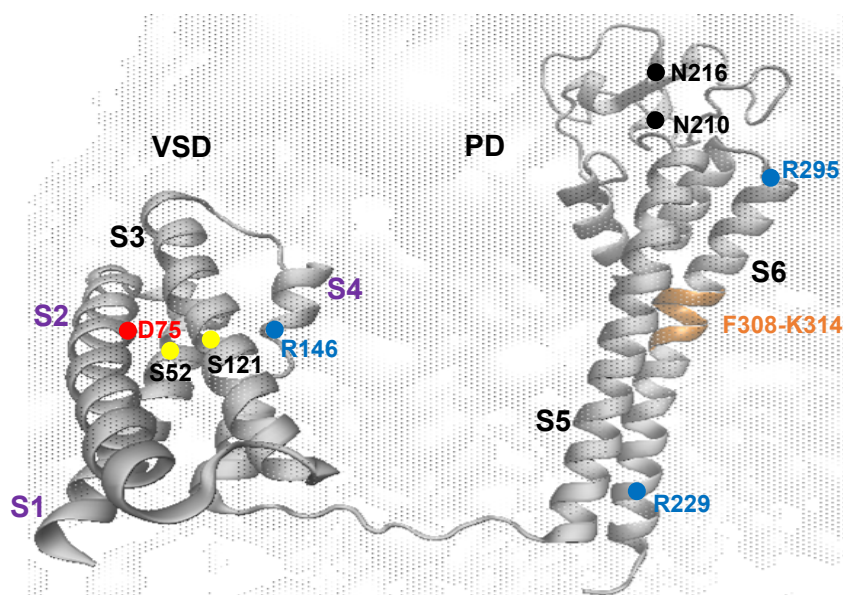


Figura 52. Principales características estructurales de NALCN estudiadas. Los segmentos menos hidrofóbicos S1, S2 y S4 se indican en letras violetas y los más hidrofóbicos S3, S5 y S6 se indican en letras negras. Los puntos azules indican los residuos de argininas R146, R295 y R229, el rojo el aspártico D75 y los amarillos los residuos serinas S52 y S121. Los círculos negros indican las asparaginas N210 y N216 de las dianas naturales de glicosilación. La secuencia de siete residuos F308-K314 implicada en la transición de $\alpha \rightarrow \pi$ se indica en anaranjado. VSD, el dominio sensor de voltaje (S1-S4) y el PD, el dominio del poro (S5 y S6).

II. Estudio topológico de la proteína NSP4

La proteína NSP4 de SARS-CoV-2 desempeña un papel crucial en la replicación del virus al participar en la organización y estabilización de las DMVs. La caracterización detallada de esta proteína es de especial interés para comprender la patogénesis del virus y para el potencial desarrollo de estrategias terapéuticas efectivas. Hasta el momento, la estructura de NSP4 no ha sido resuelta experimentalmente, y las predicciones *in silico* sobre su topología muestran discrepancias. Por lo tanto, en este capítulo llevamos a cabo la caracterización molecular de NSP4 del SARS-CoV-2 con el objetivo de determinar su topología, el número de segmentos TM y las posibles modificaciones co- y/o post-traduccionales de la proteína.

R.10. Análisis *in silico* de la topología de NSP4.

Para estudiar la topología de la proteína NSP4 realizamos inicialmente un análisis computacional utilizando los algoritmos y bases de datos más avanzados disponibles en internet (Δ G predictor, TMHMM, HMMTOP, TOPCONS, MEMSAT, SPOCTOPUS, Protter y Uniprot). La proteína NSP4 del SARS-CoV-2 tiene una longitud de 500 aminoácidos (residuos 2764-3263 de las poliproteínas pp1a y pp1ab) con una masa molecular estimada de aproximadamente 56 kDa. Las predicciones de segmentos TM y topología de la proteína NSP4 mostraron variabilidad según el método utilizado (Figura 53A). La entrada de UniProt (P0DTC1) predice seis segmentos TM en la región de NSP4, mientras que el resultado de los otros siete algoritmos de predicción varía de 4 a 7 segmentos TM. Por lo que respecta a la predicción de la topología (Figura 53B) se observó un notable grado de coincidencia entre los diferentes algoritmos, ya que en su mayoría predijeron que los extremos N- y C-terminal están orientados hacia el citosol; excepto los algoritmos Protter y HMMTOP, que predijeron una orientación N-terminal o C-terminal extracelular, respectivamente (Figura 53B).

El análisis computacional de las modificaciones co- y post-traduccionales, muestra que los algoritmos como MEMSAT y SPOCTOPUS predicen una secuencia señal escindible (Figura 53B) coincidiendo con lo sugerido previamente para la proteína NSP4 de SARS-CoV (Oostra et al., 2007). Además, el análisis de la secuencia de NSP4 con el servidor UniProt (P0DTD1) permitió identificar una diana de glicosilación no canónica (Asn-Ile-Cys) en la posición N131. Cabe mencionar que en eucariotas la adición de azúcares a proteínas de membrana y de secreción la realizan complejos multiproteicos (OST) que transfieren los azúcares desde el dolicol fosfato donador de los azúcares a residuos de asparagina en secuencias consenso Asn-X-Ser/Thr de polipéptidos sustrato (Ramírez et al., 2023), siendo X cualquier aminoácido excepto prolina (Welply et al., 1983; Bañó-

Polo et al., 2011). La potencial diana de glicosilación no canónica (NIC) encontrada en la proteína NSP4 del SARS-CoV-2 es similar a la de su homóloga en el SARS-CoV, la cual se ha sugerido estar parcialmente glicosilada (Oostra et al., 2007).

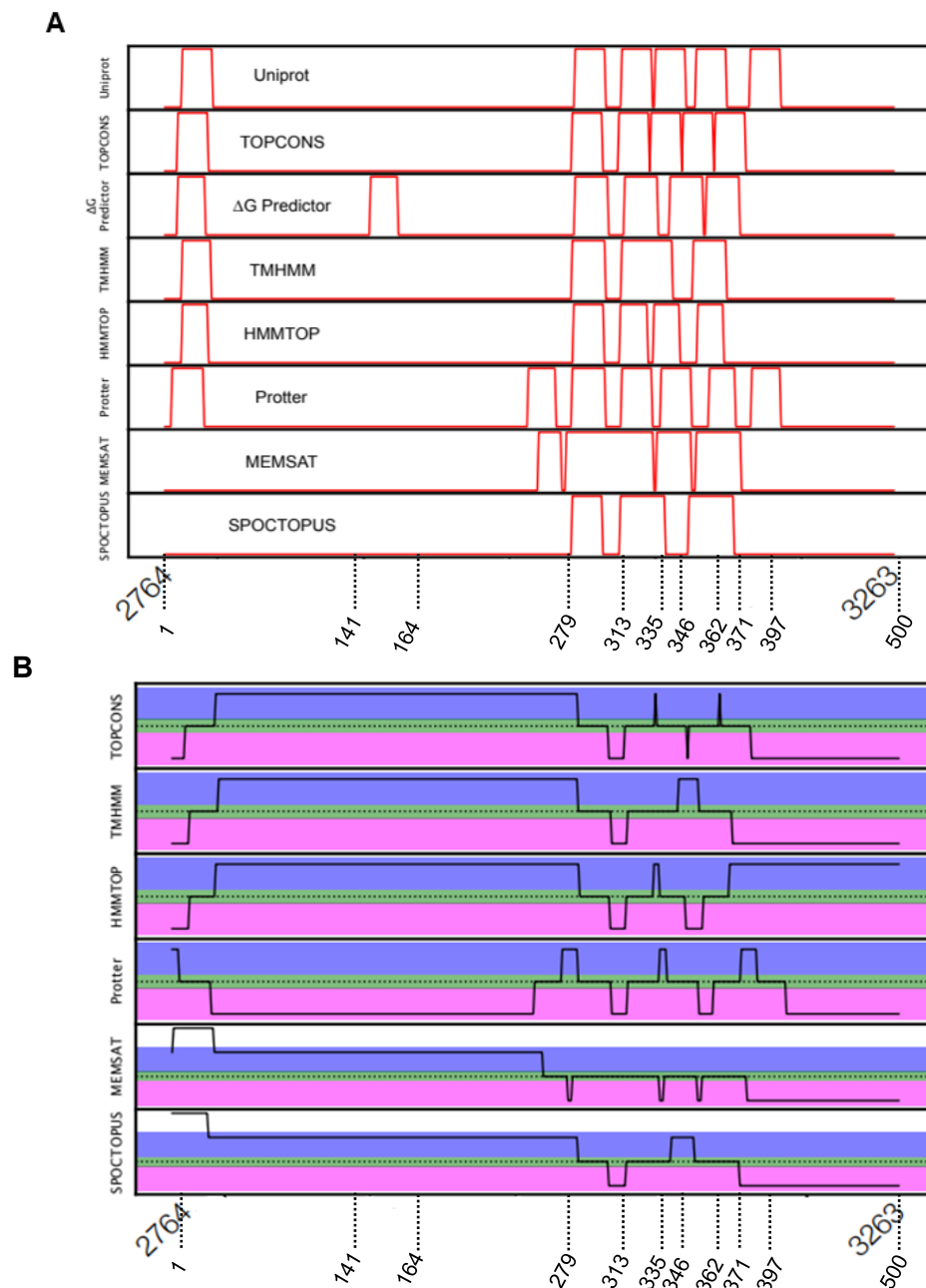


Figura 53. Análisis *in silico* de la proteína NSP4 de SARS-CoV-2 (A) Predicción de segmentos TM. La secuencia de 500 aminoácidos (residuos 2764-3263 en pp1a) de NSP4 fue analizada por 8 algoritmos. La línea roja representa la proteína NSP4 y las regiones TM predichas por cada uno de los 8 algoritmos. **(B) Predicción de la topología de NSP4.** La secuencia de 500 aminoácidos (residuos 2764-3263 en pp1a) de NSP4 fue analizada en 6 algoritmos de predicción de topología. Se muestra la secuencia de la proteína NSP4 (línea negra) y sus diferentes regiones orientadas hacia zonas extracelulares (morado), intracelulares (rosa) y de membrana (verde). Además, tanto SPOCTOPUS y MEMSAT predicen una secuencia señal en N-terminal (blanco).

R.11. Topología de NSP4 en membranas de bacterias.

Diseño experimental del ensayo.

Después del análisis computacional la topología en membranas de bacterias de la proteína NSP4 de SARS-CoV-2 se analizó experimentalmente utilizando el sistema de reportero dual PhoA-LacZα (Karimova *et al.*, 2017) que se describió en el capítulo anterior. De acuerdo con las predicciones computacionales, el indicador dual se fusionó al extremo C-terminal de diferentes variantes truncadas de NSP4 (141-mer, 164-mer, 279-mer, 313-mer, 335-mer, 346-mer, 362-mer, 371-mer, 397-mer y 500-mer o longitud completa de NSP4) que incluían de forma acumulativa los distintos segmentos TM (Figura 54). Las construcciones se transformaron en células de *E. coli* y se realizaron ensayos cualitativos y cuantitativos.

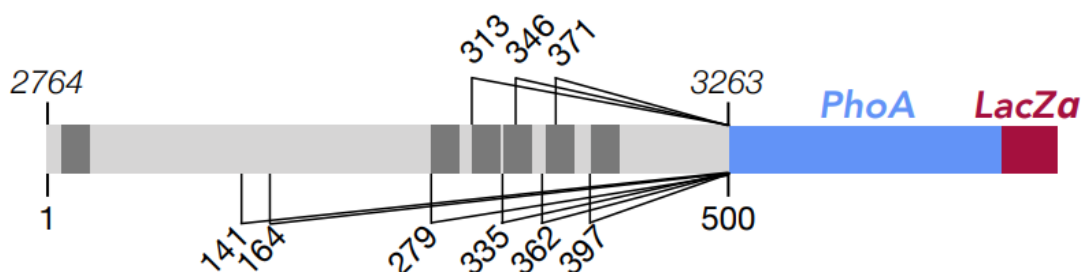


Figura 54. Variantes fusionadas de los truncados NSP4/pho-lac. Se muestra un esquema de la secuencia de 500 aminoácidos de NSP4 (en gris), destacando en gris oscuro los segmentos TM predichos por los algoritmos computacionales y la posición del último codón seleccionado en cada polipéptido truncado (141-mer, 164-mer, 279-mer, 313-mer, 335-mer, 346-mer, 362-mer, 371-mer, 397-mer y 500-mer). Cada codón fue fusionado con las secuencias de pho-lac (destacadas en azul y rojo, respectivamente). La posición de cada codón se eligió en base a los TM predichos en el análisis computacional.

Topología de NSP4 en membranas de bacterias

Los transformantes resultantes se analizaron en medios sólidos de indicador dual que contenían dos sustratos cromogénicos distintos. Uno de los sustratos se vuelve azul en presencia de la actividad fosfatasa, mientras que el otro adquiere un tono rojo cuando β-galactosidasa está activa. Como se muestra en la parte superior de la Figura 55, las células que expresan las fusiones a los polipéptidos truncados de 141, 164 y 279 residuos mostraron un fenotipo azul, lo que indica una ubicación periplásmica del indicador *pho-lac* y, en consecuencia, de los residuos correspondientes E141, G164 y D279 de NSP4. Esta ubicación sugiere que si la primera región hidrofóbica (residuos 10-32) se inserta en la membrana interna de las bacterias debe hacerlo con una orientación N-terminal citoplásmica y C-terminal periplásmica. La fusión del indicador *pho-lac* después del residuo H313 (313-mer) produjo colonias rojas en los medios de

indicador dual, lo que indica una ubicación intracelular para el residuo H313 y sugiere, por tanto, la presencia de un segmento TM entre los residuos 279 y 313. Cuando el indicador *pho-lac* se fusionó después del residuo Y335 (335-mer), la proteína quimérica exhibió una alta actividad fosfatasa (azul intenso), consistente con una ubicación periplásmica del residuo Y335, lo que sugiere la presencia de un segmento TM entre los residuos 313 y 335. La fusión del reportero *pho-lac* después del residuo Y346 de NSP4 produjo colonias azul-púrpura en el medio de cultivo con indicador dual, lo que sugiere que el residuo Y346 está localizado en (o estrechamente asociado a) la membrana. Cuando el indicador *pho-lac* se fusionó después de los residuos H362, P371, Y397 o Q500 (el último de la secuencia de NSP4), las proteínas quiméricas exhibieron una alta actividad de β -galactosidasa (fenotipo rojo), consistente con una ubicación citoplásmica de los correspondientes residuos de NSP4 (Figura 55, arriba) lo que sugiere la presencia de un segmento TM entre los residuos 335 y 362.

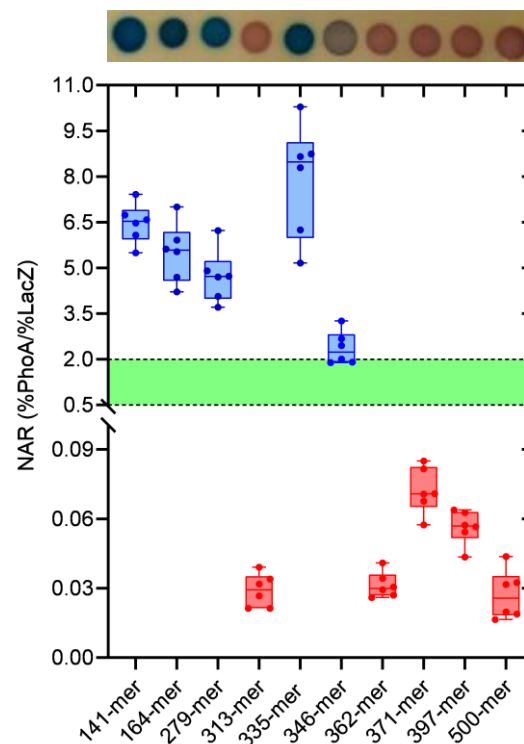


Figura 55. Topología de NSP4 en membranas bacterianas. (Arriba) Ensayo cualitativo, se sembraron células de *E. coli* DH5 α que expresaban las diferentes fusiones NSP4/Pho-Lac en medio indicador con dos sustratos cromogénicos, Red-Gal (para la actividad de β -galactosidasa) y X-Pho (para la actividad de fosfatasa alcalina). La coloración azul de las colonias (alta actividad fosfatasa) indica una ubicación periplásmica del residuo de fusión. La coloración roja de las colonias (alta actividad de β -galactosidasa) indica la ubicación citosólica del residuo de fusión. (Abajo) En el diagrama de cajas y bigotes se muestra la actividad relativa normalizada (NAR) de los ensayos enzimáticos cuantitativos de cada variante fusionada a Pho-Lac. Los puntos sólidos representan los valores de 6 réplicas independientes. Los valores de NAR >2 son indicativos de ubicación periplásmica, entre 2 y 0,5 son indicativos de ubicación en la membrana y los valores <0,5 son indicativos de ubicación citosólica.

Para corroborar los resultados cualitativos obtenidos para cada truncado se realizó un ensayo cuantitativo. Para ello, se determinó la actividad relativa normalizada (NAR: $(\text{PhoA}_{\text{muestra}}/\text{PhoA}_{\text{máxima}})/(\beta\text{Gal}_{\text{muestra}}/\beta\text{Gal}_{\text{máxima}})$) en base a la actividad enzimática fosfatasa alcalina y β -galactosidasa de cada quimera de las suspensiones bacterianas transformadas con los diferentes truncados fusionados al reportero dual de cada quimera (Tabla 18 y Figura 55). Los truncados 141-mer, 164-mer, 279-mer y 335-mer tienen valores de $\text{NAR} > 2$, lo cual nos indica una actividad fosfatasa alcalina prevalente de estas quimeras y que los aminoácidos E141, G164, D279 y Y335 se ubican en el periplasma. Por otro lado, los truncados 313-mer, 362-mer, 371-mer, 397-mer y 500-mer tienen valores $\text{NAR} < 0,5$, lo cual indica una actividad β -galactosidasa prevalente de estas quimeras y que los aminoácidos H313, H362, P371 y Q500 de NSP4 están ubicados en el citosol. Sin embargo, el polipéptido 346-mer tiene valores de $\text{NAR} \sim 2$, lo que sugiere que el aminoácido Y346 podría estar ubicado en el periplasma y/o ligeramente asociado a la membrana. De esta manera, los ensayos en membranas de bacterias mediante el uso de este reportero dual nos indican como resultados preliminares que la proteína NSP4 tiene cuatro segmentos TM y ambos extremos N- y C-terminal se ubican hacia el citosol. Sin embargo, corresponde verificar dicha topología en membranas eucariotas.

Tabla 18. Valores cuantitativos de la actividad relativa normalizada (NAR) de las diez proteínas truncadas de NSP4 fusionadas a PhoA-LacZ α . Los valores de cada construcción corresponden a las seis réplicas del ensayo.

141-mer	164-mer	279-mer	313-mer	335-mer	346-mer	362-mer	371-mer	397-mer	500-mer
7,416	7,005	6,225	0,032	10,291	3,261	0,034	0,071	0,056	0,031
6,582	5,916	4,071	0,021	8,661	2,013	0,031	0,071	0,054	0,019
6,741	4,696	4,905	0,034	6,253	1,911	0,027	0,057	0,063	0,016
6,081	5,541	4,706	0,039	5,166	1,891	0,029	0,067	0,062	0,018
6,472	4,218	3,708	0,021	8,291	2,677	0,041	0,081	0,043	0,043
5,501	5,621	4,731	0,027	8,742	2,455	0,026	0,085	0,057	0,032

R.12. Inserción en membranas de la primera región hidrofóbica de NSP4

Diseño experimental del ensayo

Dado que NSP4 forma parte de la poliproteína 1a, para investigar si la primera región hidrofóbica de NSP4 es reconocida por el translocón del RE en células eucariotas, utilizamos un sistema reportero (C_L TM) basado en la glicosilación cotraduccional de proteínas de membrana y de secreción similar al utilizado en el capítulo anterior, pero en este caso con sólo 2 dianas de glicosilación reporteras. De esta manera, la primera región hidrofóbica (del residuo 10 al 32) de NSP4 se fusionó en pauta tras el extremo C-terminal de la cadena ligera λ (C_L) de la inmunoglobulina, que incluye una secuencia

señal y sólo un sitio de glicosilación (G1) (Feige et al., 2013; Duart et al., 2022), seguida de dos secuencias flexibles (*linkers*) que flanquean la secuencia objeto de estudio (*tested*). En el extremo C-terminal se añadió otro sitio de glicosilación (G2) y un epítipo c-myc (Figura 56A). Tras la transfección de células eucariotas con la construcción obtenida (C_L-2G+TM₁₀₋₃₂), se analizan las glicosilaciones de los reporteros (G1 y G2) una vez que la proteína quimérica ha sido expresada. La diana de glicosilación G1 siempre estará modificada debido a la translocación nativa del dominio C_L de la inmunoglobulina. En cambio, la diana de glicosilación G2 sólo se modificará si la región hidrofóbica estudiada (residuos 10-32) no se inserta en la membrana del RE y acaba siendo translocada al lumen del RE (Figura 56B).

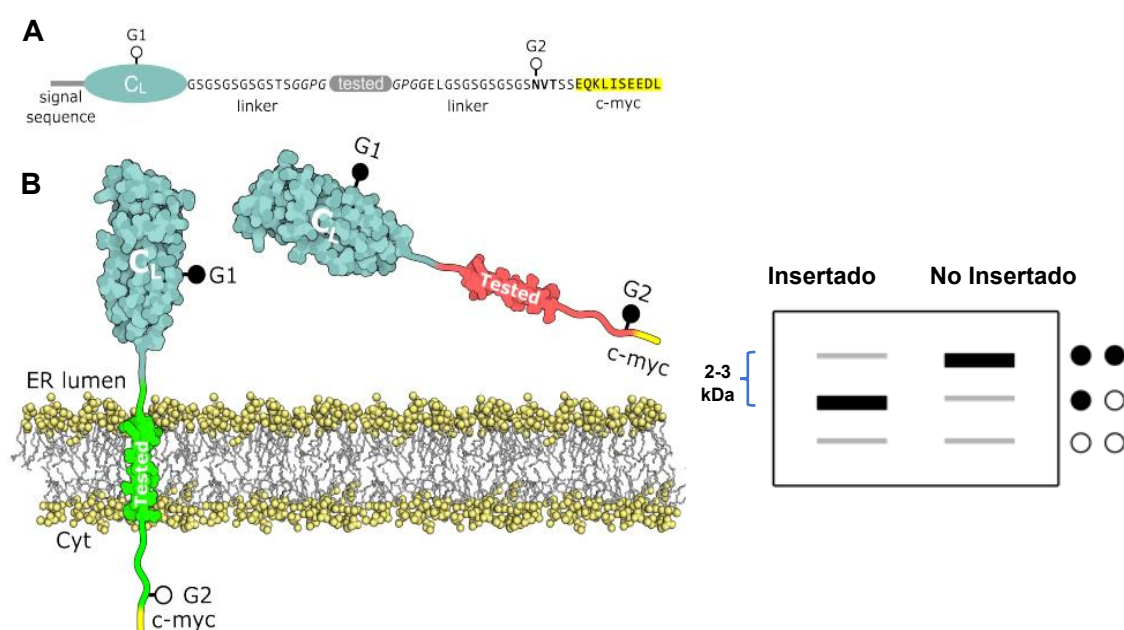


Figura 56. Ensayo C_LTM-2G. (A) Esquema de la construcción diseñada. El sistema C_LTM-2G incluye un dominio de la cadena ligera λ (C_L) de la inmunoglobulina en el que diseñamos una diana de glicosilación (G1), una región repetitiva de glicinas y serinas (*linker*) a ambos lados del TM evaluado flanqueado por los tetrapéptidos GGPG y GPGG, un sitio de glicosilación (G2), y la etiqueta c-myc (amarilla). (B) Representación esquemática del sistema C_LTM-2G. Los sitios de glicosilación diseñados permiten distinguir entre la inserción (una glicosilación, G1) y translocación (doble glicosilación, G1 y G2) de la secuencia evaluada. Las diferencias en el patrón de glicosilación se pueden detectar en la inmunotransferencia dado que cada glicosilación incrementa el peso molecular de la proteína en ~2.5 kDa.

Estudio de la inserción de la primera región hidrofóbica de NSP4.

La construcción C_LTM-2G+TM₁₀₋₃₂, que contiene los residuos de la primera región hidrofóbica de NSP4, se transfectó en células Hek293T. Las proteínas expresadas se extrajeron y analizaron por electroforesis en presencia de SDS (SDS-PAGE), y se visualizaron por *western blot* utilizando un anticuerpo anti-c-myc y quimioluminiscencia. En estos experimentos, se observó predominantemente la forma monoglicosilada de la proteína, lo que indica que la región TM₁₀₋₃₂ es reconocida e insertada por el translocón

en la membrana del RE (Figura 57, carril 6). Se confirmó que la variación en la movilidad electroforética se debe a la glicosilación mediante el tratamiento de las muestras con EndoH (Figura 57, carril 5). En paralelo, las células Hek293T se transfectaron con construcciones control que contenían una secuencia TM (Figura 57, carriles 1 y 2) o una secuencia soluble (Figura 57, carriles 3 y 4) que se usaron como control de inserción y translocación (Martínez-Gil *et al.*, 2010), respectivamente. De estos resultados se puede deducir que la primera región hidrofóbica de NSP4 es reconocida por el translocón del RE como un segmento TM en células eucariotas.

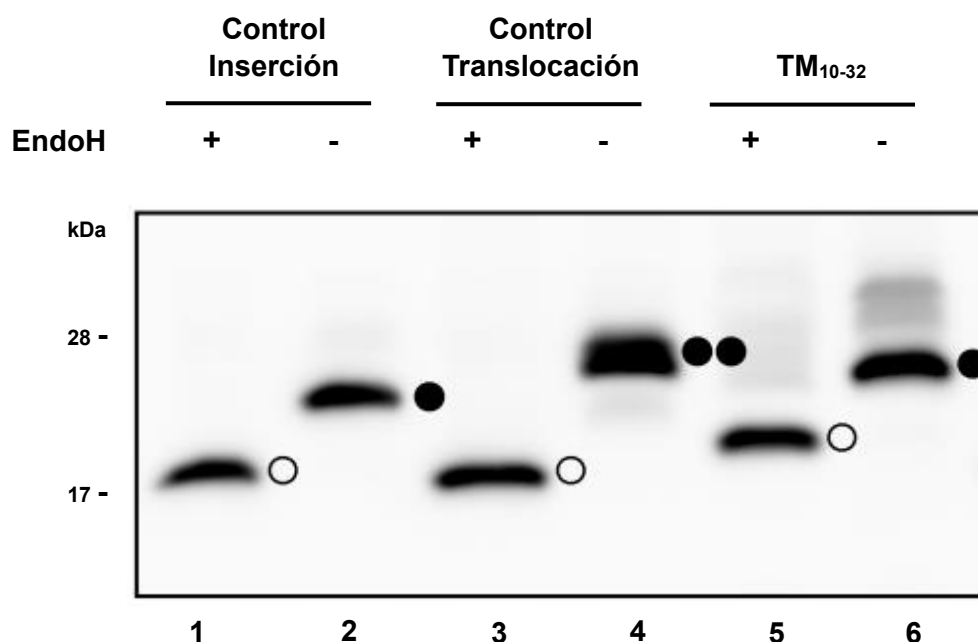


Figura 57. Inserción en membrana del segmento TM₁₀₋₃₂ de NSP4 de SARS-CoV-2. Las células Hek293T transfectadas con las construcciones indicadas (C_LTM-2G+control inserción/translocación o C_LTM-2G+TM₁₀₋₃₂) fueron lisadas después de 24h de la transfección. Los extractos celulares en presencia (+) o ausencia (-) de EndoH, se analizaron por inmunotransferencia para detectar las proteínas etiquetadas con c-myc. Las secuencias control fueron KVLLVTGVLGLLLLIKWK (carriles 1 y 2) y YNLYIYAFFASSLVHGFI (carriles 3 y 4). Las proteínas no glicosiladas se indican con círculos blancos, las proteínas mono y doblemente glicosiladas se indican con uno o dos círculos negros, respectivamente.

R.13. Estudio del procesamiento co-/post-traduccion de NSP4

R13A. Glicosilación de NSP4

La proteína NSP4 del SARS-CoV-2 y de otros coronavirus tiene un gran lazo hidrofílico comprendido entre las dos primeras regiones hidrofóbicas expuesto al lumen del RE durante su síntesis en células eucarióticas. En esta región se ha descrito la presencia de sitios glicosilables en otros coronavirus que podrían cumplir funciones relevantes. Así, se ha demostrado que la sustitución de las asparaginas N237A y N176A del sitio de glicosilación de la proteína NSP4 del virus de la hepatitis murina MHV provoca problemas en el crecimiento del virus y la síntesis del RNA viral (Clementz et al., 2008; Gadlage et al., 2010). En la misma línea se ha detectado que una proporción considerable de la proteína NSP4 de SARS-CoV, se encuentra modificada parcialmente en una diana de glicosilación no canónica (NIC). Dada la elevada similitud de secuencia entre las proteínas NSP4 de los dos SARS (Anexo 4D, Figura 81, Tabla 32), procedimos a comprobar si la diana no canónica (NIC, residuos 131-133) del SARS-CoV-2 es modificada en células eucariotas.

El análisis comparativo de la secuencia de NSP4 del SARS-CoV-2 y los otros seis coronavirus humanos conocidos muestran que el motivo de glicosilación no canónico NIC sólo está presente en los coronavirus del SARS (Figura 58B). Para investigar su posible glicosilación, diseñamos dos mutantes, uno para impedir la N-glicosilación reemplazando la asparagina aceptora por una glutamina no aceptora (N131Q), y un segundo mutante (C133S) diseñado para convertir el sitio de glicosilación no canónico (NXC) en canónico (NXS/T). Como muestra la Figura 58C las proteínas NSP4 silvestre y mutante (N131Q) (carriles 1 al 4) presentan una doble banda, si bien el tratamiento con EndoH de estas muestras indicó que ninguna de las dos bandas corresponde a formas N-glicosiladas de NSP4. Por otra parte, sólo la proteína mutante C133S que contiene una diana canónica muestra una movilidad electroforética compatible con la glicosilación de la proteína (carril 5). La modificación por glicosilación de esta variante fue verificada por tratamiento con EndoH (carril 6). Estos resultados demuestran que en la proteína NSP4 del SARS-CoV-2 el sitio de glicosilación no canónico no es modificado en células eucariotas. Además, la glicosilación de la NSP4 mutante C133S corrobora que esta región ubicada entre los dos primeros segmentos TM de la proteína está orientada al lumen del RE.

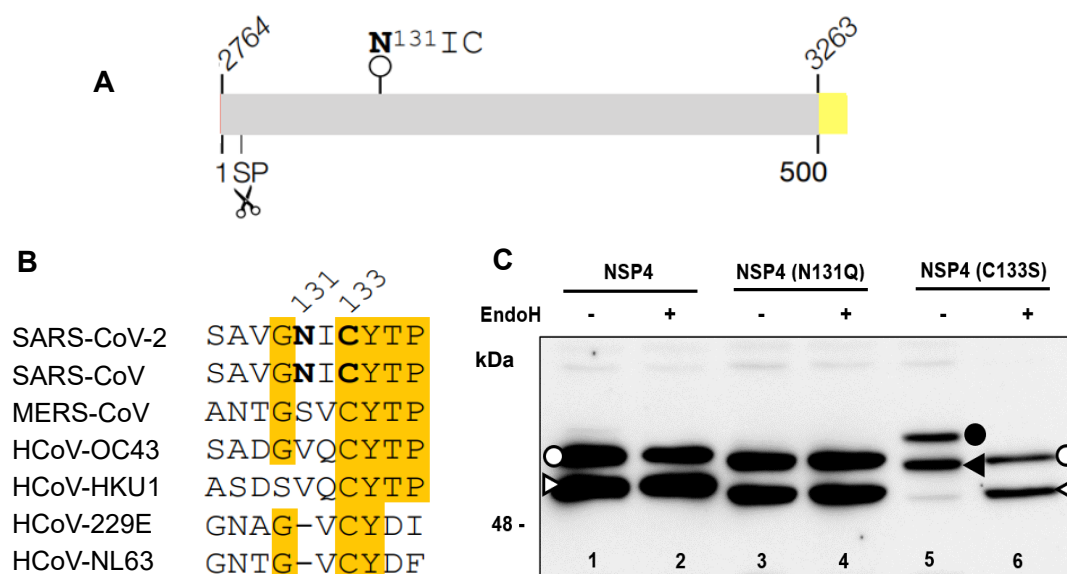


Figura 58. La diana de glicosilación no canónica NIC no es modificada en células eucariotas. (A) Esquema de la secuencia de NSP4 (la numeración de residuos en la secuencia de poliproteína pp1a se muestra en cursiva). Se indica el sitio de glicosilación no canónico (N₁₃₁IC) (círculo blanco). El epítipo c-myc se muestra en amarillo. **(B)** El sitio de glicosilación no canónico (N131IC) de SARS-CoV-2 sólo se conserva en el SARS-CoV entre los siete coronavirus humanos SARS-CoV-2, SARS-CoV (UniProt P59637), MERS-CoV (UniProt K9N5R3); HCoV-OC43 (UniProt Q4VID3), HCoV-HKU1 (UniProt Q0ZJ83), HCoV-229E (UniProt P19741) y HCoV-NL63 (UniProt Q5SBN7). Los residuos con más del 50% de conservación entre los coronavirus humanos se resaltan en anaranjado. **(C)** Las células Hek293T transfectadas con las construcciones NSP4, NSP4 (N131Q) y NSP4 (C133S) fueron lisadas y tratadas con (+) o sin (-) EndoH. La presencia de las proteínas etiquetadas se detectó por inmunotransferencia. Las proteínas no glicosiladas y no procesadas se indican con círculos blancos y las glicosiladas con círculos negros. Los triángulos señalan formas de las proteínas de menor tamaño molecular.

R13B. Caracterización del procesamiento de NSP4

La aparición del doble bandeo aún en presencia de EndoH en el apartado anterior sugiere el procesamiento de la proteína. Dado que las predicciones de algunos de los algoritmos utilizados (Figura 53B) sugieren la presencia de un sitio de corte tras la primera región hidrofóbica que se comportaría como péptido señal, decidimos estudiar el posible procesamiento de la NSP4 por el Complejo de la Peptidasa Señal (SPC). Para ello se mutó la región de corte predicha (T28-H31) reemplazando los residuos naturales (28TPHV₃₁) por 4 alaninas NSP4-AAAA (TPHV₂₈₋₃₁ → AAAA) o por 4 fenilalaninas NSP4-FFFF (TPHV₂₈₋₃₁ → FFFF). Cuando las células se transfectaron con estas variantes mutadas de NSP4 la banda de mayor movilidad electroforética desapareció parcialmente en el mutante NSP4-AAAA (Figura 59, carril 3 y 4) y completamente en la variante NSP4-FFFF (Figura 59, carril 5 y 6). En ningún caso las muestras variaron su movilidad electroforética tras el tratamiento con EndoH (Figura 59, carriles 2, 4 y 6),

descartando la posible glicosilación de las proteínas. Estos resultados demuestran que en este tipo celular la primera región hidrofóbica de la proteína es escindida de manera mayoritaria y además la proteína no es glicosilada en células eucarióticas a diferencia de lo observado anteriormente para la NSP4 de SARS-CoV (Oostra et al., 2007).

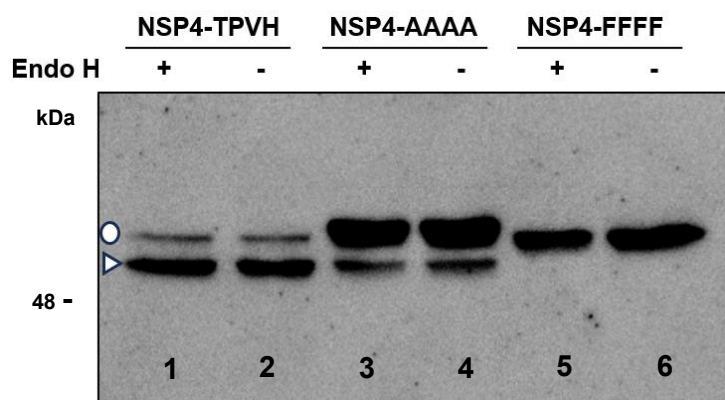


Figura 59. La primera región hidrofóbica de NSP4 es procesada como péptido señal. Los extractos celulares en presencia (+) o ausencia (–) de EndoH, se analizaron por inmunotransferencia para detectar las proteínas etiquetadas con c-myc. El círculo blanco indica proteínas no procesadas y no glicosiladas, el triángulo blanco indica proteínas digeridas y no glicosiladas.

A continuación, decidimos identificar la proteasa celular responsable del procesamiento de NSP4. Estudios previos han demostrado, mediante la utilización de un inhibidor específico de SPC (cavinafungin), la participación de este complejo en la maduración de poliproteínas virales (Snapp et al., 2017; Zhang et al., 2016). Cavinafungin, un lipopéptido (Figura 60 A) de origen fúngico extraído de *Colispora cavicola* (Ortíz-López et al., 2015), ha demostrado ser un compuesto potente y selectivamente activo contra el virus Zika y el virus del dengue (Estoppey et al., 2017). Su acción consiste en inhibir el complejo SPC, el cual desempeña un papel crucial en la maduración de las pre-proteínas y en el ciclo replicativo de estos flavivirus (Estoppey et al., 2017).

Cuando las células transfectadas con la secuencia que codifica la proteína NSP4 se trataron con cavinafungin, desapareció la banda de menor tamaño molecular (mayor movilidad electroforética) (Figura 60 B, carriles 5 y 6), demostrando que la proteína NSP4 del SARS-CoV-2 es sustrato de la SPC, siendo parcialmente digerida en las células Hek293T. Una vez más, el tratamiento de las muestras con EndoH descarta la posible glicosilación de la proteína.

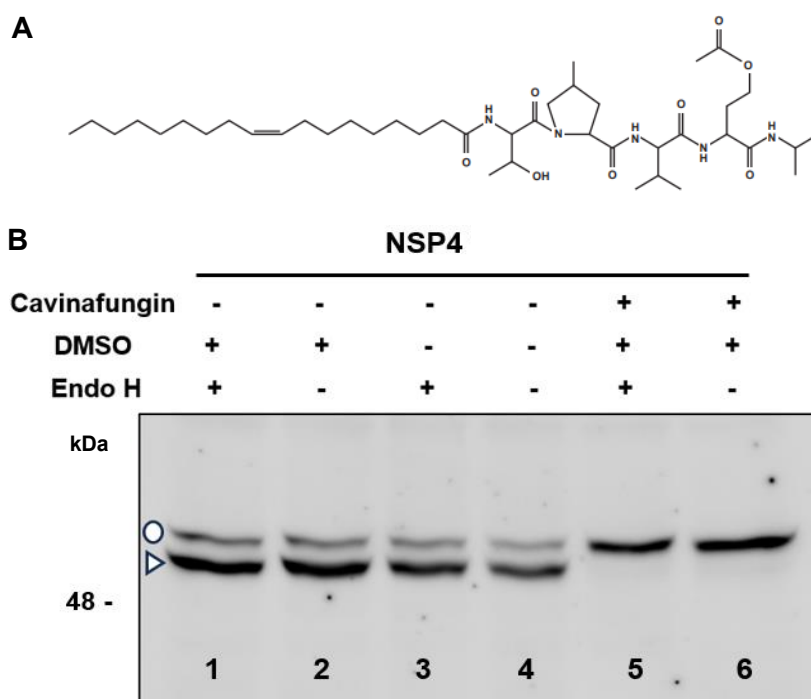


Figura 60. NSP4 es procesada por el complejo SPC. (A) Estructura química de Cavinafungin. (B) La proteína silvestre NSP4 fue transfectada en células Hek293T y los cultivos inoculados con cavinafungin y DMSO (+), medio de dilución del cavinafungin. El extracto celular fue tratado con EndoH (+) y las muestras proteicas se evaluaron por inmunotransferencia. El círculo blanco indica proteínas no procesadas y no glicosiladas, el triángulo blanco indica proteínas procesadas y no glicosiladas.

Procesamiento de la proteína NSP4 en otras células

Para investigar si el procesamiento de NSP4 por el complejo SPC observado en células Hek293T es generalizable y ocurre en otras líneas celulares, se transfectaron otros dos tipos celulares (Vero y A549) con el plásmido de expresión en eucariotas que contiene la secuencia de NSP4 completa y una etiqueta c-myc (Figura 61). El extracto celular analizado tras 36 horas de la transfección mostró la presencia de dos poblaciones de proteínas en los tres tipos de células. El tratamiento con EndoH de estas muestras confirmó de nuevo que ninguna de las dos bandas de proteínas corresponde a formas N-glicosiladas de NSP4 y, por lo tanto, que la proteína NSP4 es procesada parcialmente por los complejos SPC del RE en los tres tipos de células, incluyendo las células A549 que son derivadas del epitelio basal alveolar, más directamente relacionadas con el ciclo vital del SARS-CoV-2.

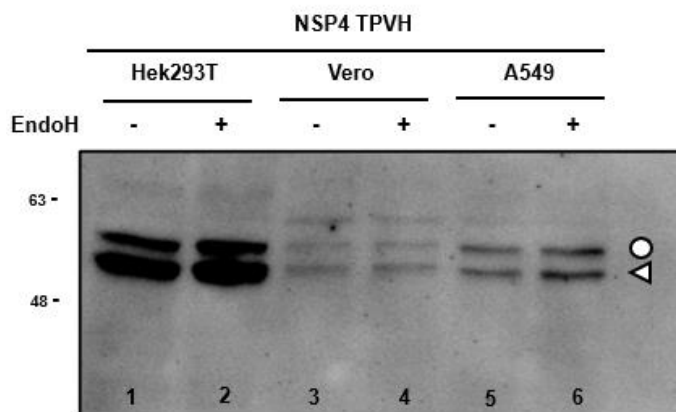


Figura 61. Procesamiento de la proteína NSP4 en diferentes líneas celulares. Las células Hek293T, Vero y A549 transfectadas con NSP4 fueron lisadas después de 36 horas para obtener el extracto proteico y detectar las proteínas expresadas por inmunotransferencia. Se comprobó el estado de glicosilación de las muestras mediante tratamiento con EndoH (+). Las proteínas no procesadas y no glicosiladas se indican con puntos blancos y las proteínas procesadas no glicosiladas con triángulos blancos.

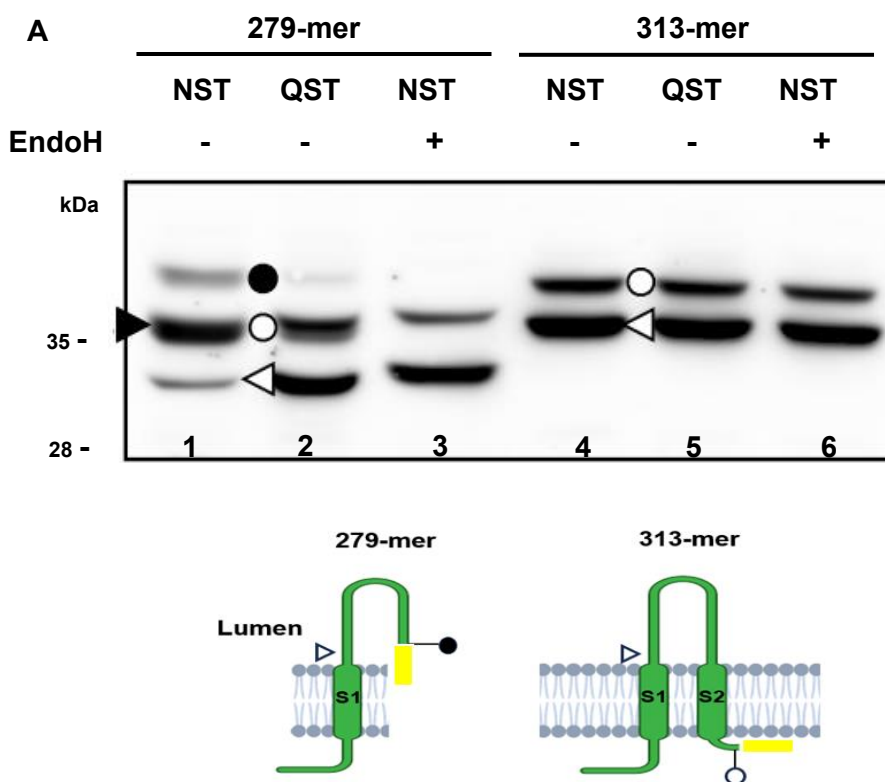
R.14. Topología de NSP4 en membranas celulares humanas

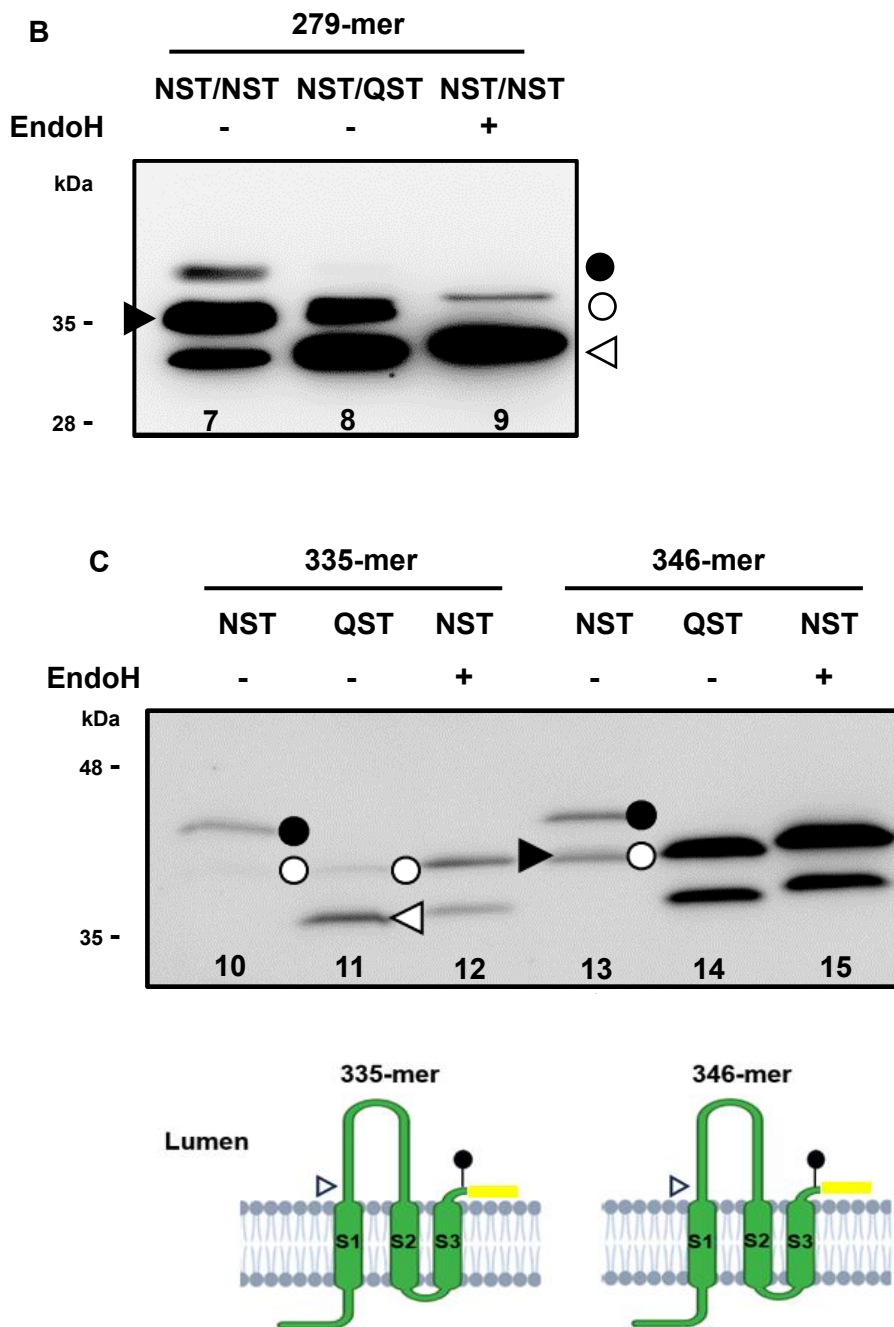
Topología de NSP4 en células eucarióticas

Para determinar experimentalmente la topología de NSP4 de SARS-CoV-2 en células eucariotas utilizamos la aproximación molecular basada en mapeo por glicosilación anteriormente descrita. Para ello, basándonos en las predicciones (Figura 53) y en los resultados obtenidos en células procariotas (Figura 55) diseñamos una serie de polipéptidos truncados de longitud creciente que contienen una etiqueta optimizada de glicosilación en C-terminal (NSTMGM) (Bañó-Polo et al., 2011; Tamborero et al., 2011), o, como control, una etiqueta equivalente pero no glicosilable (QSTMGM). Las células transfectadas con una construcción que contiene aproximadamente la mitad N-terminal de NSP4 (279 aminoácidos) con la etiqueta de glicosilación en C-terminal dio lugar a tres bandas en los análisis electroforéticos correspondientes a tres formas distintas de la proteína (Figura 62 A, carriles 1-3). Como se ha demostrado en apartado anterior, la proteína se procesa parcialmente por el SPC, por lo que la banda de mayor migración electroforética corresponderá a la proteína procesada; la banda de menor migración corresponderá al polipéptido sin procesar y con la diana reportera en C-terminal modificada, mientras que la banda intermedia procederá de una mezcla de proteína sin procesar y proteína procesada y glicosilada, dado que el incremento de peso molecular en los polipéptidos debido a la glicosilación (~2,5 kDa) coincide en buena medida con la pérdida estimada por la digestión de la proteína debida al procesamiento del complejo SPC. El tratamiento con EndoH elimina el resto glicano de la especie proteica de mayor peso molecular (sin procesar y glicosilada en la etiqueta C-terminal), y las moléculas de proteína de la banda intermedia que corresponde a la proteína procesada y glicosilada dando lugar a la aparición de las bandas de menor tamaño respectivamente. Estos resultados indican que el residuo D279 está ubicado en la zona luminal del RE. Para

demostrar que el resultado obtenido es debido a que la primera región hidrofóbica se inserta en la membrana y no es traslocada, se realizó el experimento con el truncado 279-mer con dos dianas de glicosilación diseñadas en N- y C-terminal (Figura 62B). Los resultados muestran que sólo se modificó la diana situada en el extremo C-terminal indicando que la primera región hidrofóbica direcciona la proteína NSP4 hacia la membrana del RE, es insertada en membrana y procesada por SPC en algunos casos. Cuando las células se transfectaron con el truncado que contiene 313 aminoácidos (313-mer), dio lugar a especies proteicas que no eran glicosiladas (Figura 62 A, carriles 4-6), lo que sugiere la presencia de un segmento TM entre los residuos 279 y 313, en concordancia con la topología de NSP4 observada en células bacterianas (Figura 55). Por el contrario, los polipéptidos truncados de 335 y 346 aminoácidos (335-mer y 346-mer, respectivamente) fueron glicosilados eficientemente (Figura 62 C, carriles 10-12 y 13-15, respectivamente), lo que demuestra la disposición luminal de residuos Y335 e Y346, y sugiere la presencia de un dominio TM entre los residuos 313 y 335.

Figura 62. Topología de proteínas NSP4 truncadas 279-mer, 313-mer, 335-mer y 346-mer. Las células Hek293T transfectadas con las proteínas truncadas 279-mer y 313-mer (A) y 335-mer y 346-mer (C) con dianas indicadas en C-terminal y 279-mer (B) con dianas en N-y C-terminal fueron lisadas y tratadas con (+) o sin (-) EndoH. La presencia de proteínas con etiqueta se detectó por inmunotransferencia. Las proteínas no glicosiladas y no procesadas se indican con círculos blancos y las glicosiladas con círculos negros. Los triángulos señalan proteínas de menor tamaño molecular procedentes del procesamiento del primer TM. Abajo de cada inmunodetección, se muestra un esquema de la disposición de los segmentos TM de cada truncado en la membrana del RE de acuerdo con la interpretación de los resultados obtenidos, el triángulo blanco indica el sitio de escisión de la SPC y en C-terminal la etiqueta c-myc (amarillo).





El procesamiento de NSP4 por la SPC también se comprobó en el truncado de 346 aminoácidos. Para ello se transfectaron células Hek293T con un plásmido que contenía la secuencia de NSP4 truncada silvestre (346-mer_{28TPVH₃₁}) y la variante mutante en la que estos residuos fueron reemplazados por fenilalaninas (346-mer TPVH₂₈₋₃₁→FFFF) (Figura 63 B). La Figura 63 muestra como la variante 346-mer FFFF en la que se altera el sitio de digestión de la SPC (TPVH₂₈₋₃₁→FFFF) aparece una única banda (carril 4), a diferencia de lo que ocurre con el polipéptido silvestre en el que se volvió a evidenciar

la aparición de dos bandas. En conjunto, estos resultados indican que tanto la proteína NSP4 completa del SARS-CoV-2 como todos los truncados que de ella se derivan son parcialmente procesados por el SPC. De manera, que podemos concluir que NSP4 es sustrato del complejo SPC del RE, que actúa reconociendo la primera región hidrofóbica de la proteína como secuencia señal y procediendo parcialmente a su digestión.

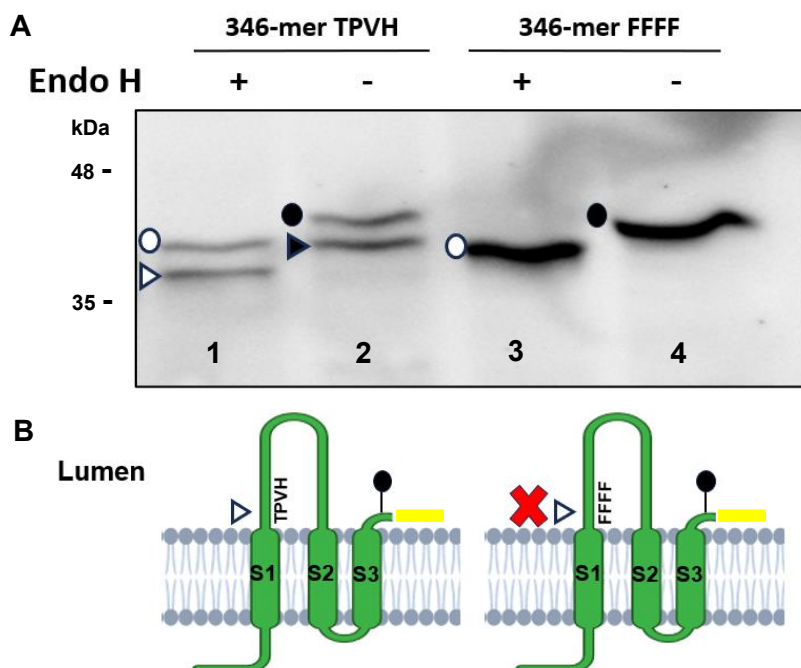


Figura 63. Procesamiento del truncado 346-mer por el SPC. (A) Las células Hek293T transfectadas con las proteínas truncadas 346-mer fueron lisadas y tratadas con (+) o sin (-) EndoH. La presencia de proteínas se detectó por inmunotransferencia. Las proteínas no glicosiladas y no procesadas se indican con círculos blancos y las glicosiladas con círculos negros. Los triángulos señalan proteínas de menor tamaño molecular procedentes del procesamiento del primer TM. (B) Esquema que ilustra los segmentos TM de cada truncado en la membrana del RE de acuerdo con la interpretación de los resultados obtenidos, el triángulo blanco indica sitio de escisión de la SPC y en amarillo se muestra la etiqueta c-myc en el C-terminal.

A continuación, se analizaron polipéptidos truncados de 362 aminoácidos (362-mer) que produjeron formas no glicosiladas (Figura 64 A, carriles 1-3), y por ende el C-terminal de este truncado se ubica hacia el citosol, sugiriendo la presencia de una región TM entre los residuos 335 y 362. Finalmente, los truncados de mayor tamaño (371-mer y 397-mer) y la secuencia de NSP4 completa (500-mer) produjeron formas no glicosiladas (Figura 64, carriles 4-9 y Figura 64 B, carriles 1-3, respectivamente), ya que en todas ellas la presencia de dos bandas se debe al procesamiento parcial por peptidasas señal (SPC) del RE, como corroboran los tratamientos con EndoH. Estos últimos resultados sugieren que el extremo C-terminal de NSP4 es una región grande (alrededor de 140 residuos) orientada al citoplasma. En base a los resultados obtenidos hasta ahora, se infiere que la proteína NSP4 consta de cuatro segmentos TM, el primero de ellos es

reconocido e integrado en la membrana del RE como segmento TM por el translocón con su extremo N-terminal hacia el citosol. A medida que se sintetiza el resto de la cadena polipeptídica, el translocón permite la inserción de los otros tres segmentos TM a través de su puerta lateral. En consecuencia, la proteína NSP4 adopta una topología con sus extremos N y C-terminales orientados hacia el citosol (proteína tipo II), presentando un gran lazo entre las dos primeras regiones hidrofóbicas (residuos 33-279) traslocado al lumen del RE y una región extensa (entre los residuos 362-500) ubicada en el citosol. El tratamiento con el inhibidor de la SPC (cavinafugin) y el resultado de los mutantes del sitio putativo de corte del SPC sugieren que el primer TM de NSP4 es procesado parcialmente dando lugar a dos poblaciones de la proteína; una con tres segmentos TM y su extremo N-terminal orientado hacia el lumen del RE (proteína tipo I), y otra con cuatro segmentos TM y ambos extremos N- y C-terminales orientados hacia el citosol (proteína tipo II), como mencionamos anteriormente.

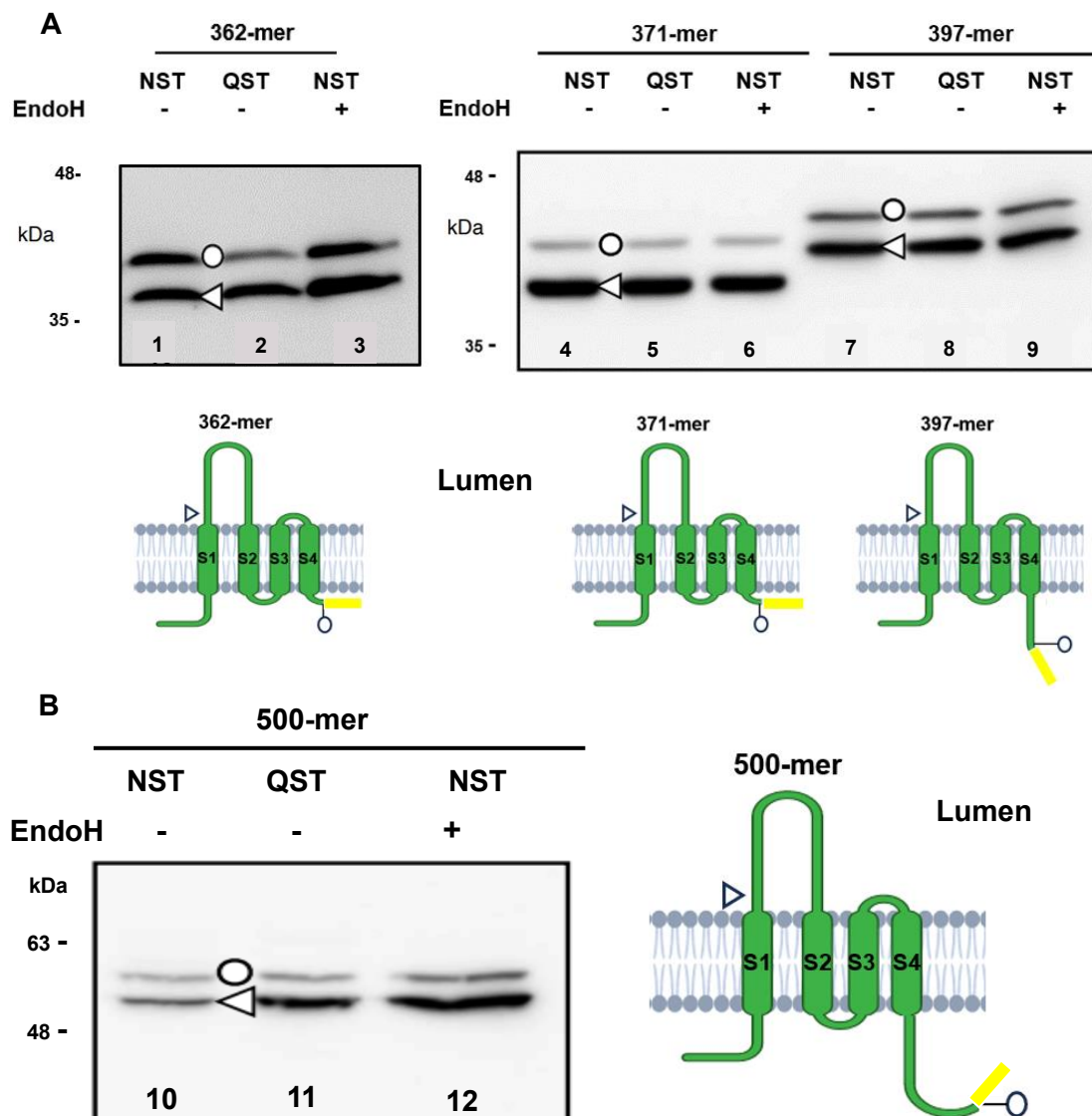


Figura 64. Topología de los truncados 362-mer, 371-mer, 397-mer y de NSP4 completa (500-mer). Las células Hek293T transfectadas con las proteínas truncadas 362-mer, 371-mer (A) 397-mer y 500-mer (B) fueron lisadas y tratadas con (+) o sin (-) EndoH. La presencia de las proteínas con etiqueta c-myc se detectó por inmunotransferencia. Las proteínas no glicosiladas y no procesadas se indican con círculos blancos y los triángulos señalan proteínas de menor tamaño molecular procedentes del procesamiento del primer TM. Adicionalmente, se muestra un esquema de la disposición de los segmentos TM de cada truncado en la membrana del RE de acuerdo con la interpretación de los resultados obtenidos, el triángulo blanco indica sitio de escisión de la SPC y en C-terminal la etiqueta c-myc (amarillo).

R.15. Purificación de los dominios Solubles de la proteína NSP4 del SARS-CoV-2

Dominio soluble luminal de la proteína NSP4

La región que comprende el dominio soluble luminal abarca desde el residuo S40 al P270 y está orientado hacia el lumen del RE según los resultados experimentales mostrados en esta tesis. Asimismo, regiones homólogas a este dominio luminal están presentes en otros coronavirus y están vinculadas con la interacción con la proteína NSP3 (Sakai *et al.*, 2017), con la formación de DMVs (Sakai *et al.*, 2017; Gadlage *et al.*, 2010; Clementz *et al.*, 2008) y con la replicación viral (Gadlage *et al.*, 2010; Clementz *et al.*, 2008).

Para la expresión de este dominio de 231 residuos (~25 kDa de peso molecular) se transformó la construcción 6xHis-NSP4(S40-P270) clonada en el plásmido pETNKL en la cepa de *E. coli* C43 (DE3). Los transformantes fueron sembrados en dos medios de cultivo (Hyper y LB) y la inducción se testó a dos temperaturas (37°C y 20°C) (Miguel-Romero *et al.*, 2020; Broedel *et al.*, 2001). Como muestra la Figura 65, en medio Hyper, con una inducción a 37°C durante 3h se obtuvo la mayor inducción del fragmento 6xHis-NSP4(S40-P270), por lo que estas fueron las condiciones elegidas para la expresión a gran escala.

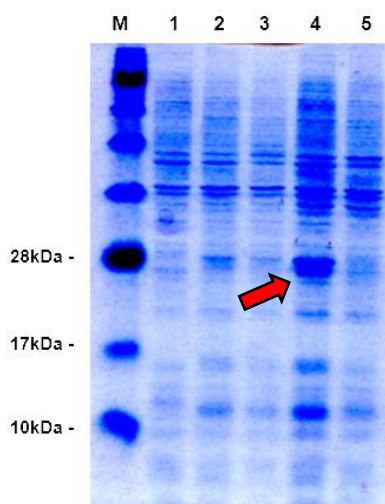


Figura 65. Expresión a pequeña escala del dominio luminal de NSP4 (S40-P270). Células de *E. coli* fueron transformadas con la construcción pETNKL+HisTag+NSP4(S40-P270), crecidas independientemente en dos medios de cultivo e inducidas a dos temperaturas de incubación. Se muestra el análisis de los extractos crudos por SDS-PAGE al 16% de acrilamida. La flecha indica la migración esperada del dominio luminal NSP4(S40-P270). M (marcador de peso *Prestained Protein Ladder - Mid-range molecular weight* (10 - 180 kDa). (1) Cultivo sin inducir, (2) Inducción en LB a 37°C 3h (3) Inducción en LB 20°C toda la noche, (4) Inducción en medio Hyper a 37°C 3h (5) Inducción en Hyper 20°C toda la noche.

Tras la expresión a gran escala, el dominio luminal de NSP4 se intentó purificar por cromatografía de afinidad en columnas HisTrap™, ya que la quimera contenía una etiqueta de 6 histidinas en N-terminal.

Después del análisis de las muestras por electroforesis, encontramos que el fragmento del dominio luminal 6xHis-NSP4(S40-P270) no apareció en ninguna fracción de la cromatografía, es decir, ni en la fracción no retenida (Figura 66, carril 3) y ni en las fracciones eluidas con 7, 40, 70 ó 100% de solución B1 que contenía imidazol (500mM) (Figura 66, carriles 4, 5, 6 y 7 respectivamente). Sin embargo, se detectó en gran cantidad en el sedimento obtenido después de la sonicación (Figura 66, carril 2). Estos resultados indican que la sobreexpresión el dominio NSP4(S40-P270) genera agregados (posiblemente cuerpos de inclusión) que imposibilitaron su purificación.

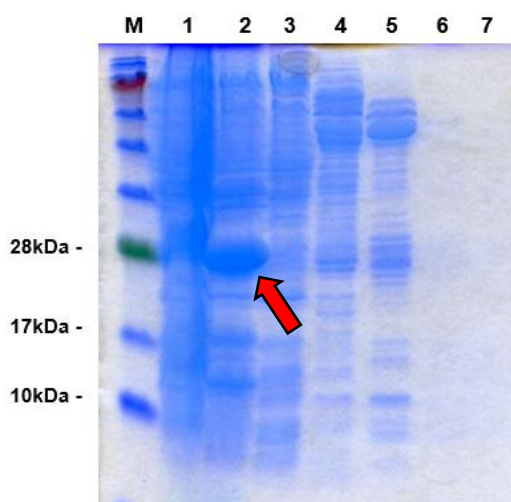


Figura 66. Cromatografía de afinidad HisTrapHP del dominio luminal de NSP4 (HisTag+NSP4(S40-P270)). Análisis de las muestras por electroforesis SDS-PAGE al 16% de acrilamida. La flecha indica la migración de la proteína sobreexpresada en la fracción no soluble. M (marcador de peso Prestained Protein Ladder - Mid-range molecular weight (10 - 180 kDa). (1) Muestra antes de sonicar, (2) Fracción no soluble después de sonicar, (3) Fracción no retenida en columna His Trap. Elución con 7% (4); 40% (5); 70% (6) y 100% (7) de solución B1 con 500mM de imidazol.

Para la solubilización de estos agregados el sedimento se incubó con diferentes concentraciones de urea (2, 3, 4 y 6M) (Singh et al., 2015). Después de la incubación con urea, se volvió a centrifugar y se tomaron muestras de la fracción soluble y sedimento. Al analizar los sobrenadantes de todas las muestras tratadas con urea mediante electroforesis en SDS-PAGE, observamos que en todos los casos se consiguió solubilizar aproximadamente un 50% de la proteína 6xHis-NSP4(S40-P270) (Figura 67, carriles 1, 5, 7 y 9). Por ello, la fracción soluble de las muestras tratadas con urea 2M se cargó en una columna HisTrap, pero la mayor parte de la muestra salió en la fracción no retenida (Figura 67, carril 3). Ante la imposibilidad de poder solubilizar y purificar el fragmento soluble NSP4(S40-P270) se decidió abandonar su caracterización estructural.

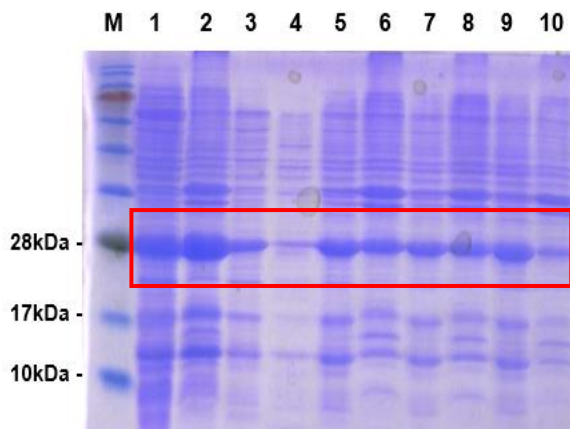


Figura 67. Solubilización de los cuerpos de inclusión formados por el dominio S40-P270 de NSP4. Se muestra el análisis por SDS-PAGE al 16% para evaluar el resultado tras el tratamiento de los cuerpos de inclusión formados por la proteína HisTag+NSP4(S40-P270) con diferentes concentraciones de urea. M (marcador de peso Prestained Protein Ladder - Mid-range molecular weight (10 - 180 kDa). El recuadro rojo indica la proteína HisTag+NSP4(S40-P270). (1) Sobrenadante 2M urea, (2) Fracción no soluble tras tratamiento con 2M urea, (3) Fracción no retenida (2M urea), (4) Elución con 40% de solución B1 (2M urea), (5) Sobrenadante 3M urea, (6) Fracción no soluble tras tratamiento con 3M urea, (7) Sobrenadante 4M urea, (8) Fracción no soluble tras tratamiento con 4M urea, (9) Sobrenadante 6M urea, (10) Fracción no soluble tras tratamiento con 6M urea.

Expresión y purificación del dominio soluble citosólico de la proteína NSP4

El dominio C-terminal de la proteína NSP4 del SARS-CoV-2, altamente conservado entre todas las especies de coronavirus desempeña un papel crucial en el ensamblaje de las DMVs. De hecho, se ha descrito que este dominio citosólico de NSP4 es esencial para la formación eficiente del complejo de replicación del coronavirus SARS-CoV (van Hemert et al., 2008).

En base a estudios previos de cristalización del dominio C-terminal de NSP4 de SARS-CoV (Xu et al., 2009) se eligió expresar y purificar el fragmento C-terminal desde el residuo T411 hasta el final de la proteína (residuo Q500), de unos ~10 kDa de peso molecular. Este dominio 6xHis-NSP4(T411-Q500) fue expresado en la cepa de *E. coli* C43(DE3) utilizando de nuevo dos medios de cultivo (Hyper y LB) y dos temperaturas de inducción (37°C y 20°C) (Miguel-Romero et al., 2020; Broedel et al., 2001). El fragmento se sobreexpresó en mayor cantidad a 37°C, tanto en medio LB como en medio Hyper (Figura 68, carriles 2, 4), por lo que se eligió el LB e inducción durante 3h a 37°C para su expresión a gran escala.

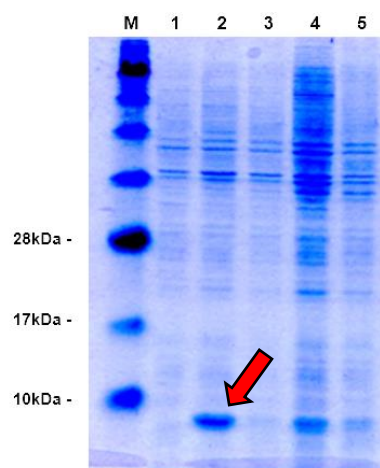


Figura 68. Expresión de la proteína HisTag+NSP4(T411-Q500). Células de *E. coli* fueron transformadas con la construcción pETNKI+HisTag+NSP4(T411-Q500) y crecidas independientemente en dos medios de cultivo (Hyper y LB) e inducidas a dos temperaturas diferentes (20°C y 37°C). Se muestra el análisis de los extractos crudos por SDS-PAGE al 16% de acrilamida. La flecha indica la migración esperada del dominio luminal NSP4(T411-Q500). M (marcador de peso Prestained Protein Ladder - Mid-range molecular weight (10 - 180 kDa). (1) Cultivo sin inducir, (2) Inducción en LB a 37°C 3h (3) Inducción en LB 20°C toda la noche, (4) Inducción en Hyper a 37°C 3h, (5) Inducción en Hyper 20°C toda la noche.

El extracto proteico obtenido después de la expresión a gran escala del dominio citosólico de NSP4 (T411-Q500) se cargó y purificó en una columna HisTrap™, eluyendo con concentraciones crecientes de imidazol (tampón B2). El análisis electroforético de las fracciones permitió comprobar que el fragmento proteico 6xHis-NSP4(T411-Q500) eluyó principalmente a 40% de solución B2 junto con otras proteínas contaminantes (Figura 69, carril 5).

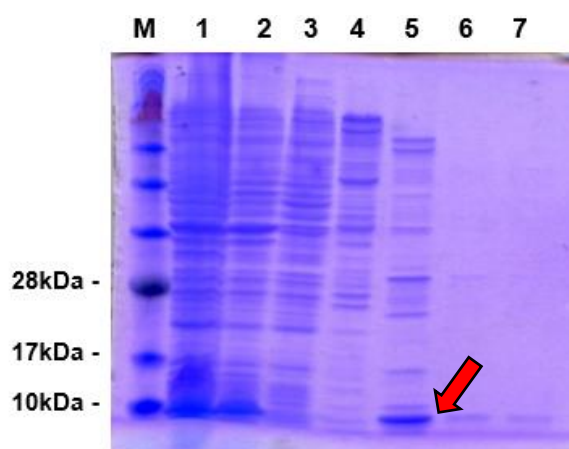


Figura 69. Purificación por cromatografía de afinidad del dominio T411-Q500 de NSP4. Análisis de las muestras por electroforesis SDS-PAGE al 16% de acrilamida. La flecha indica la proteína HisTag+NSP4(T411-Q500). M (marcador de peso Prestained Protein Ladder - Mid-range molecular weight (10 - 180 kDa). (1) Muestra antes de sonicar, (2) Fracción soluble después de sonicar, (3) Fracción no retenida en columna His Trap. Elución con 7% (4); 40%(5); 70%(6) 100% (7) de solución B2 con 500 mM de imidazol.

Con el objetivo de eliminar las proteínas contaminantes e incrementar la probabilidad de cristalización, se eliminó la cola de Histidinas de la construcción 6xHis-NSP4(T411-Q500) (Miguel-Romero et al., 2020). Para ello, la fracción eluida al 40% de B2 se dializó y se digirió con la proteasa GST-PreScission que corta en la secuencia LEVLFLQ/GP ubicada entre la cola de histidinas y el fragmento NSP4(T411-Q500). El resultado de la digestión se pasó de nuevo por una columna HisTrap^{HP} en las mismas condiciones que antes. El fragmento NSP4(T411-Q500) de ~10 kDa eluyó en la fracción no retenida junto

la GST-Prescission de 44kDa (Figura 70, carril 2). El resto de las proteínas contaminantes eluyeron al 40% de solución B2 (Figura 70, carril 4).

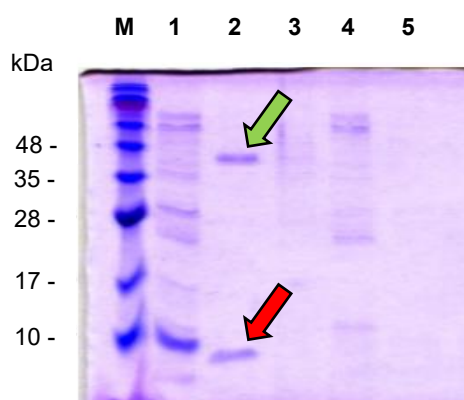


Figura 70. Purificación por cromatografía en columna HisTrap^{HP} del producto de la digestión de HisTag+NSP4(T411-Q500). Análisis en SDS-PAGE al 16% de las muestras y las fracciones colectadas tras la carga de la construcción HisTag+NSP4(T411-Q500) digerida con la proteasa GST-Prescission. M (marcador de peso Prestained Protein Ladder - Mid-range molecular weight (10 - 180 kDa)). La flecha verde indica la proteasa GST-Prescission y la flecha roja la proteína NSP4(T411-Q500) sin la cola de histidinas. (1) Muestra no digerida, (2) Fracción no retenida en columna HisTrap de la muestra digerida. Elución con 7% (3), 40% (4) y 100% (5) de solución B2 con 500 mM de imidazol.

La separación del dominio citosólico NSP4(T411-Q500) de la GST-Prescission se realizó mediante cromatografía de exclusión molecular en una columna de Proteosec 16/60 6-600 de Neo Biotech (Hong et al., 2012). Debido a su mayor tamaño (44 kDa) la proteasa no quedó retenida en la columna y eluyó primero (Figura 71, carriles 5 y 6). A diferencia del fragmento proteico NSP4(T411-Q500) (sin HisTag) que por su menor tamaño (~10 kDa) quedó retenido en la columna y eluyó en un volumen de elución entre 100-130mL (Figura 71, carriles 7-11). A continuación, concentramos las fracciones en las que eluyó el dominio NSP4(T411-Q500) hasta una concentración final de 2,7 mg/mL.

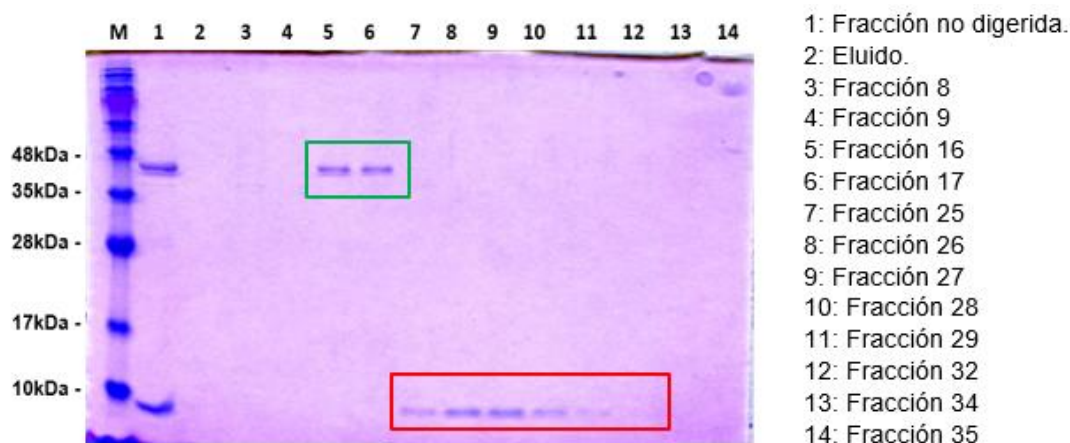


Figura 71. Purificación por exclusión molecular del fragmento T411-Q500 de NSP4. Análisis por SDS-PAGE al 16% de las muestras y las fracciones colectadas en la purificación de la proteína recombinante NSP4(T411-Q500). Las fracciones del fragmento NSP4(T411-Q500) se muestra en rojo (carriles del 7-12) y la proteasa GST-Prescission usada en la digestión se muestra en verde (carriles 5 y 6). M (marcador de peso Prestained Protein Ladder - Mid-range molecular weight (10 - 180 kDa)).

Para intentar conseguir mayor cantidad de la proteína se diseñó un método alternativo de purificación del fragmento NSP4(T411-Q500). Después de eluir la muestra con 200mM de imidazol (Figura 72, carril 1), se procedió a dializarla y digerirla utilizando GST-PresCission. Posteriormente, se pasó nuevamente a través de una columna HisTrap™, recolectando la fracción no retenida como se hizo en el método anterior. En esta ocasión, la proteasa se separó de NSP4(T411-Q500) al pasar la mezcla por una columna GSTrap™. El fragmento NSP4(T411-Q500) se recuperó en la fracción no retenida (Figura 72 carril 2) mientras que la proteasa eluyó con 20mM de GST (Figura 72 carril 4). La fracción que contenía el fragmento NSP4(T411-Q500) se concentró hasta alcanzar una concentración de 3,96 mg/mL.

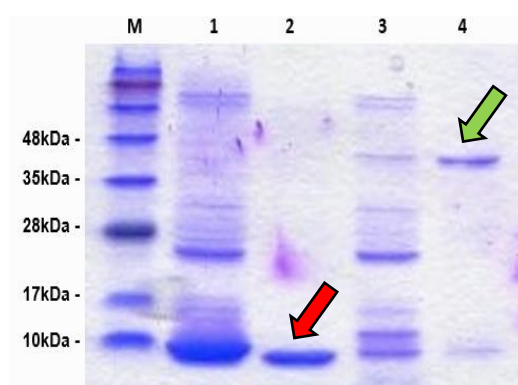


Figura 72. Método alternativo de purificación del fragmento T411-Q500 de NSP4. Análisis por SDS-PAGE al 16% de las muestras y las fracciones obtenidas en la purificación de la proteína recombinante NSP4(T411-Q500) digerida con GST-PresCission. Tras la digestión se analizó: (1) Fracción eluida con 40% solución B2 en columna HisTrap, (2) Fracción no retenida de GSTrap, (3) Fracción eluida de HisTrap con 40% de solución B2, (4) fracción eluida con 20mM de GSH tras columna GSTrap. La flecha roja muestra el fragmento NSP4(T411-Q500) y la flecha verde a la proteasa GST-PresCission. M (marcador de peso Prestained Protein Ladder - Mid-range molecular weight (10 - 180 kDa).

R.16. Cristalización del dominio citosólico de la proteína NSP4

Con el objeto de obtener la estructura por difracción de rayos X, las muestras de NSP4(T411-Q500) se intentaron cristalizar. Para ello, se realizaron una serie de cribados JBScreen Classic HTS I y HTS II (Jena Bioscience) y PGA Screen (Molecular Dimensions) en los que se probaron distintas condiciones de cristalización, así como diferentes concentraciones de la proteína (2,7 y 3,96 mg/mL).

La formación de cristales se monitoreó diariamente con una lupa. En la muestra concentrada a 2,7 mg/mL no se observó formación de cristales. Sin embargo, en la muestra concentrada a 3,6 mg/mL, con el tiempo se observó la formación de cristales putativos, aunque desgraciadamente comprobamos que eran cristales de sal provenientes de las condiciones de cristalización.

D.2. Discusión del capítulo II

La proteína NSP4 del SARS-CoV-2 es una proteína de membrana que se sintetiza como parte de una poliproteína que es procesada por proteasas virales. Su función ha sido principalmente asociada a la formación de DMVs (Roingeard et al., 2022) mediante interacciones con NSP3 (Sakai et al., 2017) y con otras proteínas del retículo 3 y 4 del RE (RTN3 y RTN4) (Williams et al., 2023). También se la ha relacionado con la regulación de la respuesta inmune innata del huésped (Kakavandi et al., 2023; Lu et al., 2022). En este contexto, la disposición en la membrana puede ayudar a comprender y explicar a nivel molecular las interacciones de NSP4 con el resto de componentes virales y celulares con los que interactúa.

Las predicciones *in silico* del número de segmentos TM y de la topología de NSP4 mostraron mucha variabilidad, en línea con hallazgos previos (Thomas et al., 2021). Por ejemplo, mientras UniProt (P0DTC1) predice seis segmentos TM, otros algoritmos sugieren la presencia de 4 a 7 segmentos TM para (Figura 53A). Asimismo, también encontramos discrepancias en la predicción topológica: algunos algoritmos predicen una orientación intracelular de ambos extremos de la proteína NSP4, mientras que Protter y HMMTOP sugieren una orientación extracelular de los extremos N- o C-terminal, respectivamente (Figura 53B). Aunque el sistema de inteligencia artificial AlphaFold (Jumper et al., 2020) no predice topología ni segmentos TM, la estructura 3D de NSP4 predicha por este algoritmo muestra una hélice α en el extremo N-terminal (que abarca desde la isoleucina en posición 2 hasta la isoleucina en posición 27) y, por otro lado, muestra también una región de hélices α (desde la alanina 277 hasta la leucina 398) agrupadas entre sí y que corresponderían a segmentos TM.

Nuestro objetivo inicial fue determinar experimentalmente la topología de la proteína NSP4, teniendo en cuenta que estamos estudiándola aislada de contexto de la poliproteína. En primer lugar, analizamos si la secuencia de NSP4 aislada del contexto de la poliproteína es correctamente dirigida a las membranas del RE. Nuestros resultados utilizando el ensayo C_LTM demuestran que la primera región hidrofóbica se inserta en membrana (Figura 57), lo cual coincide con el análisis *in silico* de los algoritmos TOPCONS, ΔG predictor, TMHMM, HMMTOP y Protter (Figura 53). Es más, los experimentos realizados con el truncado 279-mer de la proteína, con dos dianas de glicosilación diseñadas en N- y C-terminal, demuestran no sólo que la primera región hidrofóbica es, al menos inicialmente, un segmento TM sino también que esta región hidrofóbica direcciona la proteína NSP4 hacia la membrana del RE (Figura 62B). Como en el caso de otras proteínas de membrana virales (Saurí et al., 2005; Martínez-Gil et

al., 2010). Esta capacidad de direccionamiento a membranas de la primera región hidrofóbica de NSP4 está en la línea del análisis *in silico* de los algoritmos de predicción MEMSAT y SPOCTOPUS, que predicen esta región como secuencia señal escindible (Figura 53B), lo que también se ha sugerido previamente, si bien no demostrado para la NSP4 del SARS-CoV (Oostra et al., 2007).

Es interesante mencionar que la primera región hidrofóbica de la proteína NS4B funciona como péptido señal en flavivirus (virus Zika, Dengue serotipo 4, Fiebre amarilla y virus del Nilo occidental) (Estoppey et al., 2017), y presenta una secuencia de aminoácidos conservados similares a la primera región hidrofóbica de la proteína NSP4 (ver Anexo 4 C, Figura 79 y Tabla 31). Por otro lado, un análisis proteómico identificó sitios de escisión para SPC ubicados después de las hélices TM no iniciales en proteínas de membrana humanas (Zanotti et al., 2022). Ambas investigaciones citadas, nos permitieron identificar un posible sitio de corte en NSP4 (²⁸TPVH³¹) y caracterizar su procesamiento por SPC. Este sitio de corte de NSP4 de SARS-CoV-2 está conservado en la proteína homóloga de SARS-CoV, si bien no está conservado en las proteínas homólogas de otros cinco coronavirus humanos (ver Anexo 4 D, Figura 80 y Tabla 32). Nuestros resultados demuestran que la proteína NSP4 del SARS-CoV-2 (y posiblemente también la del SARS-CoV) es sustrato del complejo SPC, por lo que probablemente esta peptidasa participa en la maduración de sus poliproteínas pp1a y pp1ab, como se ha comprobado en los casos de otros virus (Estoppey et al., 2017; Snapp et al., 2017; Suzuki et al., 2013; Zhang et al., 2016). Esto último, podría también asignarle a la proteína NSP4 un nuevo papel patológico en la evasión de la respuesta inmune innata descrita anteriormente para el procesamiento de la proteína NS4B de flavivirus (Chen et al., 2017), o que tal escisión podría estabilizar la interacción con NSP3 para la formación de DMVs (Klatte et al., 2022).

Otra modificación de la proteína NSP4 que investigamos en la presente tesis fue la glicosilación de la diana no canónica NIC presente en su secuencia, la cual se conserva en las proteínas NSP4 de SARS-CoV y SARS-CoV-2, pero no en el resto de coronavirus humanos (Figura 58B). La glicosilación parcial de esta diana no canónica fue reportada previamente en la proteína NSP4 de SARS-CoV (Oostra et al., 2007); sin embargo, en nuestros ensayos no observamos la glicosilación de esta diana en ninguna de las líneas celulares que evaluamos (Figura 61). Por otro lado, se ha observado la glicosilación de ciertos motivos no canónicos (NVC y NPC) en otros canales de calcio (Ficelova et al., 2020). En cambio, en el caso de NSP4 los mutantes de la diana no canónica N131Q y

C133S confirman su no glicosilación y la localización luminal de esta región de la proteína.

El estudio de la topología de la proteína NSP4 en membranas de células humanas mostró un doble bandeo en todas las proteínas truncadas evaluadas producto de la digestión por el complejo SPC. Este doble bandeo también se ha observado en otros estudios con NSP4, en los que los autores sugirieron que se debía a glicosilaciones (Davies et al., 2024 y 2020; Pahmeier et al., 2023). Los resultados obtenidos en la presente tesis demuestran que la banda de menor tamaño molecular se origina por el procesamiento parcial de la proteína NSP4 por el complejo SPC del RE (Figura 60). Por lo que respecta a la orientación del resto de la proteína, los resultados *in silico* coinciden con nuestros resultados experimentales, mostrando una ubicación lumen/periplásmica de los residuos E141, G164 y D279 y la ubicación citosólica de los residuos H313, Y397 y Q500 (Figura 53). Sin embargo, la principal discrepancia en cuanto a número de segmentos TM y topología ocurre en la región comprendida entre los residuos Y335 y P371 de NSP4 (Figura 53 y 80). Las predicciones *in silico* para el residuo Y335 muestran discrepancias ubicándolo asociado a membrana (TMHMM, SPOCTOPUS), en la región extracelular (HMMTOP, TOPCONS) o en la región intracelular (MEMSAT). Por el contrario, los hallazgos *in vivo* son unánimes y ubican al residuo Y335 en el lumen/periplasma; lo cual coincide sólo con las predicciones de HMMTOP y TOPCONS. Para el residuo Y346 los resultados *in vivo* en membranas de células humanas sólo muestran concordancia con las predicciones de SPOCTOPUS y TMHMM, situándolo en el lumen y en la zona extracelular para cada tipo de ensayo respectivamente. Además, tanto los resultados cualitativos como cuantitativos en membranas de bacterias indican que este residuo está orientado hacia el periplasma, pero estrechamente asociado a la membrana. En el caso del residuo H362, la ausencia de formas glicosiladas en el truncado 362-mer (Figura 62) sugiere que dicho residuo está orientado hacia el citosol o asociado a membrana, en concordancia con la mayoría de las predicciones (Figura 50). Finalmente, los ensayos de topología *in vivo* muestran que el residuo P371 se orienta hacia el citosol, lo cual está en línea con las predicciones de los algoritmos SPOCTOPUS y TMHMM. Globalmente los resultados experimentales obtenidos en membranas procarióticas y eucarióticas coinciden con la predicción del algoritmo TMHMM. El procesamiento parcial de NSP4 del SARS-CoV-2 genera dos formas de la proteína sin procesar, con 4 segmentos TM y extremos N- y C-terminal orientados hacia el citosol (politópica tipo II) y procesada con 3 segmentos TM, un dominio N-terminal voluminoso orientado al lumen del RE y su extremo C-terminal orientado hacia el citosol (politópica tipo I) (Figura 73). Los resultados obtenidos en el estudio de las

modificaciones co-/post-traduccionales de la proteína NSP4 pueden abrir nuevas vías tanto para reducir la capacidad infectiva del virus inhibiendo el procesamiento de la proteína como en la dilucidación de la red de interacciones de NSP4 con otras proteínas virales o del huésped, si bien debemos tener presente que las modificaciones observadas con la proteína asilada podrían ser diferentes a las que tengan lugar en la proteína en el contexto de las poliproteínas pp1a y pp1ab (Pahmeier et al., 2023).

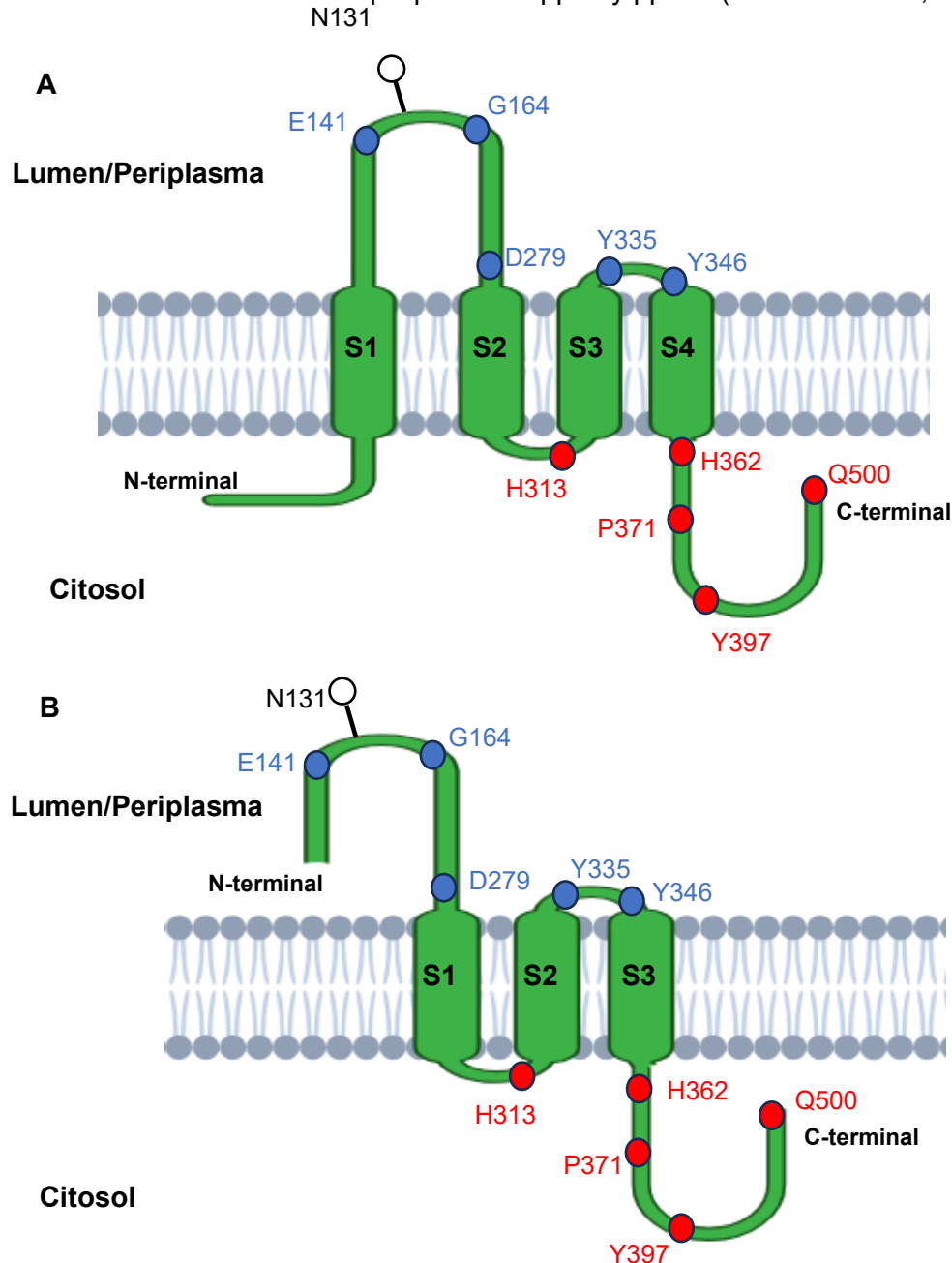


Figura 73. Disposición de la proteína NSP4 del SARS-CoV-2 en membrana. Se muestra gráficamente los resultados de topología obtenidos en membranas de bacterias y del RE. Los segmentos verdes indican los segmentos TM de NSP4 politópica tipo II (A) y tipo I (B). Los puntos azules indican los residuos ubicados hacia el lumen/periplasma y los rojos indican los residuos ubicados hacia el citosol. El círculo blanco indica la diana NIC en posición 131 no glicosilada.

Una vez definida experimentalmente la topología de NSP4 procedimos a caracterizar estructuralmente los dos dominios solubles de mayor tamaño de esta proteína. Uno de estos dominios es el bucle periplasmático/luminal (desde el residuo S40 hasta P270) y el otro es el dominio citosólico C-terminal de la proteína (desde el residuo T411 hasta Q500). El fragmento S40-P270 no se pudo purificar, pues formó cuerpos de inclusión en el proceso. En el caso del fragmento soluble T411-Q500 no formó cuerpos de inclusión y pudimos purificarlo, si bien no obtuvimos cristales. La imposibilidad de purificar y cristalizar los fragmentos S40-P270 y T411-Q500 podría deberse a muchos factores (temperatura, pH, concentración de sales) que les impidan formar las interacciones adecuadas para la formación de cristales (McPherson et al., 2014), dado que las proteínas necesitan encontrar un equilibrio y minimizar la exposición de sus regiones hidrofóbicas al agua, para formar contactos estables que conduzcan a la formación de cristales (Price et al., 2009). La presencia de regiones flexibles y dinámicas que permitan los reordenamientos necesarios para la formación de las DMVs también pueden afectar su cristalización, ya que estas regiones pueden interferir con la formación de contactos regulares y repetidos entre moléculas de proteínas que son necesarios para la formación de cristales (Price et al., 2009). En este sentido, es interesante destacar que ambos dominios de NSP4 participan en la formación de DMVs mediante interacciones con otras proteínas (Sakai et al., 2017; van Hemert et al., 2008), cuya presencia puede ser necesaria para la cristalización de la proteína.

CONCLUSIONES

1. La proteína NALCN humana se integra en membranas co-traduccionamente gracias al direccionamiento de la primera región hidrofóbica.
2. Las interacciones inter-hélice y la hidrofobicidad de cada uno de los seis segmentos transmembrana tiene un efecto sobre la estabilidad y la topología correcta del Dominio I de NALCN.
3. Las mutaciones CLIFAHDD reportadas en el segmento S6 del dominio DI no afectan su inserción, pero causan modificaciones estructurales que podrían generar alteraciones en la funcionalidad de la proteína NALCN.
4. La proteína NSP4 del SARS-CoV-2 se integra en membrana co-traduccionamente como una proteína politópica.
5. Al expresar individualmente la proteína NSP4 en células humanas, su primera región hidrofóbica es reconocida como secuencia señal “críptica” y procesada parcialmente por el complejo peptidasa SPC en diferentes líneas celulares.
6. La diana de glicosilación no canónica de la NSP4 del SARS-CoV-2 no es modificada en células humanas.

BIBLIOGRAFÍA

- Abdelkader EH, Otting G., 2021. NT*-HRV3CP: An optimized construct of human rhinovirus 14 3C protease for high-yield expression and fast affinity-tag cleavage. *J Biotechnol.* Jan 10;325:145-151. doi: 10.1016/j.jbiotec.2020.11.005.
- Adamian L, Liang J., 2001. Helix-helix packing and interfacial pairwise interactions of residues in membrane proteins. *J Mol Biol.* Aug 24;311(4):891-907. doi: 10.1006/jmbi.2001.4908.
- Alexeyev, M. F. & Winkler, H., 1999. Membrane topology of the Rickettsia prowazekii ATP/ADP translocase revealed by novel dual pho-lac reporters. *J Mol Biol* 285, 1503–1513.
- Allen KN, Entova S, Ray LC, Imperiali B., 2019. Monotopic Membrane Proteins Join the Fold. *Trends Biochem Sci.* Jan;44(1):7-20. doi: 10.1016/j.tibs.2018.09.013.
- Angelini MM, Akhlaghpour M, Neuman BW, Buchmeier MJ., 2013. Severe acute respiratory syndrome coronavirus nonstructural proteins 3, 4, and 6 induce double-membrane vesicles. *mBio.* Aug 13;4(4):e00524-13. doi: 10.1128/mBio.00524-13.
- Angius A. et al., 2019. Exome sequencing in Crisponi/cold-induced sweating syndrome-like individuals reveals unpredicted alternative diagnoses. *Clin Genet.* May;95(5):607-614. doi: 10.1111/cge.13532.
- Aoyagi K. et al., 2015. A Gain-of-Function Mutation in NALCN in a Child with Intellectual Disability, Ataxia, and Arthrogryposis. *Hum Mutat.* Aug;36(8):753-7. doi: 10.1002/humu.22797.
- Arkin IT, MacKenzie KR, Fisher L, Aimoto S, Engelman DM, Smith SO., 1996. Mapping the lipid-exposed surfaces of membrane proteins. *Nat Struct Biol.* Mar;3(3):240-3. doi: 10.1038/nsb0396-240.
- Baeza-Delgado C, Marti-Renom MA, Mingarro I., 2013. Structure-based statistical analysis of transmembrane helices. *Eur Biophys J.* Mar;42(2-3):199-207. doi: 10.1007/s00249-012-0813-9.
- Baeza-Delgado C, von Heijne G, Marti-Renom MA, Mingarro I., 2016. Biological insertion of computationally designed short transmembrane segments. *Sci Rep.* Mar 18;6:23397. doi: 10.1038/srep23397.
- Bai L, You Q, Feng X, Kovach A, Li H., 2020. Structure of the ER membrane complex, a transmembrane-domain insertase. *Nature.* Aug;584(7821):475-478. doi: 10.1038/s41586-020-2389-3.
- Banerjee AK. et al., 2020. SARS-CoV-2 Disrupts Splicing, Translation, and Protein Trafficking to Suppress Host Defenses. *Cell.* Nov 25;183(5):1325-1339.e21. doi: 10.1016/j.cell.2020.10.004.
- Bañó-Polo M, Baeza-Delgado C, Orzáez M, Marti-Renom MA, Abad C, Mingarro I., 2012. Polar/Ionizable residues in transmembrane segments: effects on helix-helix packing. *PLoS One.* 7(9):e44263. doi: 10.1371/journal.pone.0044263.
- Bañó-Polo M, Baldin F, Tamborero S, Marti-Renom MA, Mingarro I., 2011. N-glycosylation efficiency is determined by the distance to the C-terminus and the amino acid preceding an Asn-Ser-Thr sequon. *Protein Sci.* Jan;20(1):179-86. doi: 10.1002/pro.551.
- Bañó-Polo M. et al., 2017. Membrane insertion and topology of the translocon-associated protein (TRAP) gamma subunit. *Biochim Biophys Acta Biomembr.* May;1859(5):903-909. doi: 10.1016/j.bbamem.2017.01.027.
- Barber AF, Carnevale V, Raju SG, Amaral C, Treptow W, Klein ML., 2012. Hinge-bending motions in the pore domain of a bacterial voltage-gated sodium channel. *Biochim Biophys Acta.* Sep;1818(9):2120-5. doi: 10.1016/j.bbamem.2012.05.002.
- Behnke J, Mann MJ, Scruggs FL, Feige MJ, Hendershot LM., 2016. Members of the Hsp70 Family Recognize Distinct Types of Sequences to Execute ER Quality Control. *Mol Cell.* Sep 1;63(5):739-52. doi: 10.1016/j.molcel.2016.07.012.
- Bend EG, Si Y, Stevenson DA, Bayrak-Toydemir P, Newcomb TM, Jorgensen EM, Swoboda KJ., 2016. NALCN channelopathies: Distinguishing gain-of-function and loss-of-function mutations. *Neurology.* Sep 13;87(11):1131-9. doi: 10.1212/WNL.0000000000003095.
- Bend EG. et al., 2016. NALCN channelopathies: Distinguishing gain-of-function and loss-of-function mutations. *Neurology.* Sep 13;87(11):1131-9. doi: 10.1212/WNL.0000000000003095.
- Bernetti M, Bussi G., 2020. Pressure control using stochastic cell rescaling. *J Chem Phys.* Sep 21;153(11):114107. doi: 10.1063/5.0020514.

- Bouasse, M. et al., 2019. Functional expression of CLIFAHDD and IHPRF pathogenic variants of the NALCN channel in neuronal cells reveals both gain- and loss-of-function properties. *Sci Rep* 9, 11791 (2019). doi: <https://doi.org/10.1038/s41598-019-48071-x>.
- Bramswig NC. et al., 2018. Genetic variants in components of the NALCN-UNC80-UNC79 ion channel complex cause a broad clinical phenotype (NALCN channelopathies). *Hum Genet.* Sep;137(9):753-768. doi: 10.1007/s00439-018-1929-5.
- Braunger K. et al., 2018. Structural basis for coupling protein transport and N-glycosylation at the mammalian endoplasmic reticulum. *Science.* Apr 13;360(6385):215-219. doi: 10.1126/science.aar7899.
- Burg, E. D., Langan, S. T., and Nash, H. A., 2013. Drosophila social clustering is disrupted by anesthetics and in narrow abdomen ion channel mutants. *Genes Brain Behav.* 12, 338–347. doi: 10.1111/gbb.12025.
- Bussi G, Donadio D, Parrinello M., 2007. Canonical sampling through velocity rescaling. *J Chem Phys.* Jan 7;126(1):014101. doi: 10.1063/1.2408420.
- Byrne AB. et al., 2023. Genomic autopsy to identify underlying causes of pregnancy loss and perinatal death. *Nat Med.* Jan;29(1):180-189. doi: 10.1038/s41591-022-02142-1.
- Cahoy, J. D. et al., 2008. A transcriptome database for astrocytes, neurons, and oligodendrocytes: a new resource for understanding brain development and function. *J. Neurosci.* 28, 264–278. doi: 10.1523/JNEUROSCI.4178-07.2008.
- Casares D, Escribá PV, Rosselló CA. et al., 2019. Membrane Lipid Composition: Effect on Membrane and Organelle Structure, Function and Compartmentalization and Therapeutic Avenues. *Int J Mol Sci.* May 1;20(9):2167. doi: 10.3390/ijms20092167.
- Chen N, Zhou M, Dong X, Qu J, Gong F, Han Y, Qiu Y, Wang J, Liu Y, Wei Y, Xia J, Yu T, Zhang X, Zhang L., 2020. Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in Wuhan, China: a descriptive study. *Lancet.* Feb 15;395(10223):507-513. doi: 10.1016/S0140-6736(20)30211-7.
- Chen P, Wu M, He Y, Jiang B, He ML., 2023. Metabolic alterations upon SARS-CoV-2 infection and potential therapeutic targets against coronavirus infection. *Signal Transduct Target Ther.* Jun 7;8(1):237. doi: 10.1038/s41392-023-01510-8.
- Chen S, Wu Z, Wang M, Cheng A., 2017. Innate Immune Evasion Mediated by Flaviviridae Non-Structural Proteins. *Viruses.* Oct 7;9(10):291. doi: 10.3390/v9100291.
- Cheng W, Li W., 2014. Structural insights into ubiquinone biosynthesis in membranes. *Science.* Feb 21;343(6173):878-81. doi: 10.1126/science.1246774.
- Chitwood PJ, Hegde RS., 2019. The Role of EMC during Membrane Protein Biogenesis. *Trends Cell Biol.* May;29(5):371-384. doi: 10.1016/j.tcb.2019.01.007.
- Chitwood PJ, Hegde RS., 2020. An intramembrane chaperone complex facilitates membrane protein biogenesis. *Nature.* Aug;584(7822):630-634. doi: 10.1038/s41586-020-2624-y.
- Choi HK. et al., 2019. Watching helical membrane proteins fold reveals a common N-to-C-terminal folding pathway. *Science.* Nov 29;366(6469):1150-1156. doi: 10.1126/science.aaw8208.
- Chong JX. et al., 2015. De novo mutations in NALCN cause a syndrome characterized by congenital contractures of the limbs and face, hypotonia, and developmental delay. *Am J Hum Genet.* Mar 5;96(3):462-73. doi: 10.1016/j.ajhg.2015.01.003.
- Chua HC, Wulf M, Weidling C, Rasmussen LP, Pless SA., 2020. The NALCN channel complex is voltage sensitive and directly modulated by extracellular calcium. *Sci Adv.* Apr 24;6(17):eaaz3154. doi: 10.1126/sciadv.aaz3154.
- Clementz MA, Kanjanahaluethai A, O'Brien TE, Baker SC., 2008. Mutation in murine coronavirus replication protein nsp4 alters assembly of double membrane vesicles. *Virology.* May 25;375(1):118-29. doi: 10.1016/j.virol.2008.01.018.
- Coelho JPL. et al., 2019. A network of chaperones prevents and detects failures in membrane protein lipid bilayer integration. *Nat Commun.* Feb 8;10(1):672. doi: 10.1038/s41467-019-08632-0.
- Cooley RB, Arp DJ, Karplus PA., 2010. Evolutionary origin of a secondary structure: π -helices as cryptic but widespread insertional variations of α -helices that enhance protein functionality. *J Mol Biol.* 404: 232–246.

- Dai W. et al., 2020. Structure-based design of antiviral drug candidates targeting the SARS-CoV-2 main protease. *Science*. Jun 19;368(6497):1331-1335. doi: 10.1126/science.abb4489.
- Davies J. et al., 2020. Comparative Multiplexed Interactomics of SARS-CoV-2 and Homologous Coronavirus Nonstructural Proteins Identifies Unique and Shared Host-Cell Dependencies. *ACS Infect Dis*. Dec 11;6(12):3174-3189. doi: 10.1021/acsinfecdis.0c00500.
- Davies J. et al., 2024. Expression of SARS-CoV-2 Nonstructural Proteins 3 and 4 Can Tune the Unfolded Protein Response in Cell Culture. *Journal of Proteome Research*. 23 (1), 356-367. doi: 10.1021/acs.jproteome.3c00600.
- Dawson JP, Weinger JS, Engelman DM., 2002. Motifs of serine and threonine can drive association of transmembrane helices. *J Mol Biol*. Feb 22;316(3):799-805. doi: 10.1006/jmbi.2001.5353.
- DeGrado WF, Gratkowski H, Lear JD., 2003. How do helix-helix interactions help determine the folds of membrane proteins? Perspectives from the study of homo-oligomeric helical bundles. *Protein Sci*. Apr;12(4):647-65. doi: 10.1110/ps.0236503.
- Dong E, Du H, Gardner L., 2020. An interactive web-based dashboard to track COVID-19 in real time. *Lancet Infect Dis*. May;20(5):533-534. doi: 10.1016/S1473-3099(20)30120-1.
- Doñate-Macian P, Bañó-Polo M, Vazquez-Ibar JL, Mingarro I, Perálvarez-Marín A., 2015. Molecular and topological membrane folding determinants of transient receptor potential vanilloid 2 channel. *Biochem Biophys Res Commun*. Jul 3;462(3):221-6. doi: 10.1016/j.bbrc.2015.04.120.
- Duart G. et al., 2020. SARS-CoV-2 envelope protein topology in eukaryotic membranes. *Open Biol*. Sep;10(9):200209. doi: 10.1098/rsob.200209.
- Duart G, García-Murria MJ, Mingarro I., 2021. The SARS-CoV-2 envelope (E) protein has evolved towards membrane topology robustness. *Biochim Biophys Acta Biomembr*. Jul 1;1863(7):183608. doi: 10.1016/j.bbamem.2021.183608.
- Duart G, Lamb J, Ortiz-Mateu J, Elofsson A, Mingarro I., 2022. Intra-Helical Salt Bridge Contribution to Membrane Protein Insertion. *J Mol Biol*. Mar 15;434(5):167467. doi: 10.1016/j.jmb.2022.167467.
- Duart G, Marínez-Gil L, Mingarro I, 2023. Nadando en un mar de lípidos. *Revista SEBBM Interacciones lípido-proteína*. Marzo.
- Duart G. et al., 2024. Experimental and computational approaches for membrane protein insertion and topology determination. *Methods*. Apr 9:S1046-2023(24)00088-4. doi: 10.1016/j.ymeth.2024.03.012.
- Duneau JP, Genest D, Genest M., 1996. Detailed description of an α helix $\rightarrow\pi$ bulge transition detected by molecular dynamics simulations of the p185c-erbB2 V659G transmembrane domain. *J Biomol Struct Dyn*. Apr;13(5):753-69. doi: 10.1080/07391102.1996.10508889.
- Enquist K. et al., 2009. Membrane-integration characteristics of two ABC transporters, CFTR and P-glycoprotein. *J Mol Biol*. Apr 17;387(5):1153-64. doi: 10.1016/j.jmb.2009.02.035.
- Estoppey D. et al., 2017. The Natural Product Cavinafungin Selectively Interferes with Zika and Dengue Virus Replication by Inhibition of the Host Signal Peptidase. *Cell Rep*. Apr 18;19(3):451-460. doi: 10.1016/j.celrep.2017.03.071.
- Fan W. et al., 2022. Estrogen receptor β activation inhibits colitis by promoting NLRP6-mediated autophagy. *Cell Rep*. Oct 11;41(2):111454. doi: 10.1016/j.celrep.2022.111454.
- Feige MJ, Hendershot LM., 2013. Quality control of integral membrane proteins by assembly-dependent membrane integration. *Mol Cell*. Aug 8;51(3):297-309. doi: 10.1016/j.molcel.2013.07.013.
- Fernández-Quintero ML, El Ghaleb Y, Tuluc P, Campiglio M, Liedl KR, Flucher BE., 2021. Structural determinants of voltage-gating properties in calcium channels. *Elife*. Mar 30;10:e64087. doi: 10.7554/eLife.64087.
- Ficelova V, Souza IA, Cmarko L, Gandini MA, Stringer RN, Zamponi GW, Weiss N., 2020. Functional identification of potential non-canonical N-glycosylation sites within Cav3.2 T-type calcium channels. *Mol Brain*. Nov 11;13(1):149. doi: 10.1186/s13041-020-00697-z.
- Fodje MN, Al-Karadaghi S. et al., 2002. Occurrence, conformational features and amino acid propensities for the π -helix. *Protein Eng*. May;15(5):353-8. doi: 10.1093/protein/15.5.353.
- Fodje MN, Al-Karadaghi S., 2002. Occurrence, conformational features and amino acid propensities for the π -helix. *Protein Eng*. May;15(5):353-8. doi: 10.1093/protein/15.5.353.

- Fu YZ, Wang SY, Zheng ZQ, Yi Huang, Li WW, Xu ZS, Wang YY., 2021. SARS-CoV-2 membrane glycoprotein M antagonizes the MAVS-mediated innate antiviral response. *Cell Mol Immunol.* Mar;18(3):613-620. doi: 10.1038/s41423-020-00571-x.
- Fukai R. et al., 2016. De novo missense mutations in NALCN cause developmental and intellectual impairment with hypotonia. *J Hum Genet.* May;61(5):451-5. doi: 10.1038/jhg.2015.163.
- Fung TS, Liu DX., 2018. Post-translational modifications of coronavirus proteins: roles and function. *Future Virol.* Jun;13(6):405-430. doi: 10.2217/fvl-2018-0008.
- Gadlage MJ. et al., 2010. Murine hepatitis virus nonstructural protein 4 regulates virus-induced membrane modifications and replication complex function. *J Virol.* Jan;84(1):280-90. doi: 10.1128/JVI.01772-09.
- Gao Y. et al., 2020. Structure of the RNA-dependent RNA polymerase from COVID-19 virus. *Science.* May 15;368(6492):779-782. doi: 10.1126/science.abb7498.
- Gaspar CJ. et al., 2022. EMC is required for biogenesis of Xport-A, an essential chaperone of Rhodopsin-1 and the TRP channel. *EMBO Rep.* Jan 5;23(1):e53210. doi: 10.15252/embr.202153210.
- Ghosh S. et al., 2020. β -Coronaviruses Use Lysosomes for Egress Instead of the Biosynthetic Secretory Pathway. *Cell.* Dec 10;183(6):1520-1535.e14. doi: 10.1016/j.cell.2020.10.039.
- Golden M. et al., 2017. A Generative Angular Model of Protein Structure Evolution. *Mol Biol Evol.* Nov 1;34(11):3040. doi: 10.1093/molbev/msx214.
- Gordon DE. et al., 2020. A SARS-CoV-2 protein interaction map reveals targets for drug repurposing. *Nature.* Jul;583(7816):459-468. doi: 10.1038/s41586-020-2286-9.
- Görlich D, Prehn S, Hartmann E, Kalies KU, Rapoport TA., 1992. A mammalian homolog of SEC61p and SECYp is associated with ribosomes and nascent polypeptides during translocation. *Cell.* Oct 30;71(3):489-503. doi: 10.1016/0092-8674(92)90517-g.
- Görlich D, Rapoport TA. 1993. Protein translocation into proteoliposomes reconstituted from purified components of the endoplasmic reticulum membrane. *Cell.* Nov 19;75(4):615-30. doi: 10.1016/0092-8674(93)90483-7.
- Gratkowski H, Lear JD, DeGrado WF., 2001. Polar side chains drive the association of model transmembrane peptides. *Proc Natl Acad Sci U S A.* Jan 30;98(3):880-5. doi: 10.1073/pnas.98.3.880.
- Grau B, Javanainen M, García-Murria MJ, Kulig W, Vattulainen I, Mingarro I, Martínez-Gil L., 2017. The role of hydrophobic matching on transmembrane helix packing in cells. *Cell Stress.* Nov 2;1(2):90-106. doi: 10.15698/cst2017.11.111.
- Green MR, Sambrook J., 2020. Transformation of *Escherichia coli* by Electroporation. *Cold Spring Harb Protoc.* Jun 1;2020(6):101220. doi: 10.1101/pdb.prot101220.
- Groome, J. R., & Bayless-Edwards, L., 2020. Roles for Countercharge in the Voltage Sensor Domain of Ion Channels. *Frontiers in Pharmacology*, 11. <https://doi.org/10.3389/fphar.2020.00160>.
- Gupta M. et al., 2021. CryoEM and AI reveal a structure of SARS-CoV-2 Nsp2, a multifunctional protein involved in key host processes. doi: 10.1101/2021.05.10.443524.
- Hagemeijer MC, Ulasli M, Vonk AM, Reggiori F, Rottier PJ, de Haan CA., 2011. Mobility and interactions of coronavirus nonstructural protein 4. *J Virol.* May;85(9):4572-7. doi: 10.1128/JVI.00042-11.
- Harayama T, Riezman H., 2018. Understanding the diversity of membrane lipid composition. *Nat Rev Mol Cell Biol.* May;19(5):281-296. doi: 10.1038/nrm.2017.138.
- Harcourt BH, Jukneliene D, Kanjanahaluethai A, Bechill J, Severson KM, Smith CM, Rota PA, Baker SC., 2004. Identification of severe acute respiratory syndrome coronavirus replicase products and characterization of papain-like protease activity. *J Virol.* Dec;78(24):13600-12. doi: 10.1128/JVI.78.24.13600-13612.2004.
- Hedin LE. et al., 2010. Membrane insertion of marginally hydrophobic transmembrane helices depends on sequence context. *Journal of Molecular Biology.* Feb;396(1):221-229. doi: 10.1016/j.jmb.2009.11.036.
- Hegde RS, Keenan RJ., 2022. The mechanisms of integral membrane protein biogenesis. *Nat Rev Mol Cell Biol.* Feb;23(2):107-124. doi: 10.1038/s41580-021-00413-2.

- Hessa T, Kim H, Bihlmaier K, Lundin C, Boekel J, Andersson H, Nilsson I, White SH, von Heijne G., 2005. Recognition of transmembrane helices by the endoplasmic reticulum translocon. *Nature*. Jan 27;433(7024):377-81. doi: 10.1038/nature03216.
- Hessa T, Meindl-Beinker NM, Bernsel A, Kim H, Sato Y, Lerch-Bader M, Nilsson I, White SH, von Heijne G., 2007. Molecular code for transmembrane-helix recognition by the Sec61 translocon. *Nature*. Dec 13;450(7172):1026-30. doi: 10.1038/nature06387.
- Hildebrand PW, Lorenzen S, Goede A, Preissner R., 2006. Analysis and prediction of helix-helix interactions in membrane channels and transporters. *Proteins*. Jul 1;64(1):253-62. dyehoi: 10.1002/prot.20959.
- Hong P, Koza S, Bouvier ES., 2012. Size-Exclusion Chromatography for the Analysis of Protein Biotherapeutics and their Aggregates. *J Liq Chromatogr Relat Technol*. Nov;35(20):2923-2950. doi: 10.1080/10826076.2012.743724.
- Hopkins AL, Groom CR., 2002. The druggable genome. *Nat Rev Drug Discov*. Sep;1(9):727-30. doi: 10.1038/nrd892.
- Huang J. et al., 2017. CHARMM36m: an improved force field for folded and intrinsically disordered proteins. *Nat Methods*. Jan;14(1):71-73. doi: 10.1038/nmeth.4067.
- Huang Y, Yang C, Xu XF, Xu W, Liu SW., 2020. Structural and functional properties of SARS-CoV-2 spike protein: potential antiviral drug development for COVID-19. *Acta Pharmacol Sin*. Sep;41(9):1141-1149. doi: 10.1038/s41401-020-0485-4.
- Huesa J, Giner-Lamia J, Pucciarelli MG, Paredes-Martínez F, García-del Portillo F, Marina A, Casino P., 2021. Structure-based analyses of Salmonella RcsB variants unravel new features of the Rcs regulon. *Nucleic Acids Res*. Feb 26;49(4):2357-2374. doi: 10.1093/nar/gkab060.
- Humphrey W, Dalke A, Schulten K., 1996. VMD: visual molecular dynamics. *J Mol Graph*. Feb;14(1):33-8, 27-8. doi: 10.1016/0263-7855(96)00018-5.
- Humphrey, J. A. et al., 2007. A putative cation channel and its novel regulator: cross-species conservation of effects on general anesthesia. *Curr. Biol*. 17, 624–629. doi: 10.1016/j.cub.2007.02.037.
- Ismail N, Crawshaw SG, Cross BC, Haagsma AC, High S., 2008. Specific transmembrane segments are selectively delayed at the ER translocon during opsin biogenesis. *Biochem J*. May 1;411(3):495-506. doi: 10.1042/BJ20071597.
- Ismail N, Crawshaw SG, High S., 2006. Active and passive displacement of transmembrane domains both occur during opsin biogenesis at the Sec61 translocon. *J Cell Sci*. Jul 1;119(Pt 13):2826-36. doi: 10.1242/jcs.03018.
- Jaud S. et al., 2009. Insertion of short transmembrane helices by the Sec61 translocon. *Proc Natl Acad Sci U S A*. Jul 14;106(28):11588-93. doi: 10.1073/pnas.0900638106. Epub 2009 Jul 6. PMID: 19581593; PMCID: PMC2710650.
- Jelokhani-Niaraki M., 2022. Membrane Proteins: Structure, Function and Motion. *Int J Mol Sci*. Dec 27;24(1):468. doi: 10.3390/ijms24010468.
- Jiang Y, Cheng Z, Mandon EC, Gilmore R., 2008. An interaction between the SRP receptor and the translocon is critical during cotranslational protein translocation. *J Cell Biol*. Mar 24;180(6):1149-61. doi: 10.1083/jcb.200707196.
- Jiang Y, Yin W, Xu HE., 2021. RNA-dependent RNA polymerase: Structure, mechanism, and drug discovery for COVID-19. *Biochem Biophys Res Commun*. Jan 29;538:47-53. doi: 10.1016/j.bbrc.2020.08.116.
- Jo S, Kim T, Iyer VG, Im W., 2008. CHARMM-GUI: a web-based graphical user interface for CHARMM. *J Comput Chem*. Aug;29(11):1859-65. doi: 10.1002/jcc.20945.
- Jo S, Lim JB, Klauda JB, Im W., 2009. CHARMM-GUI Membrane Builder for mixed bilayers and its application to yeast membranes. *Biophys J*. Jul 8;97(1):50-8. doi: 10.1016/j.bpj.2009.04.013.
- Jumper J. et al., 2020. "Computational predictions of protein structures associated with COVID-19", Version 3, DeepMind website, 4 August 2020.
- Junne T, Kocik L, Spiess M., 2010. The hydrophobic core of the Sec61 translocon defines the hydrophobicity threshold for membrane integration. *Mol Biol Cell*. May 15;21(10):1662-70. doi: 10.1091/mbc.e10-01-0060.

- Kakavandi S. et al., 2023. Structural and non-structural proteins in SARS-CoV-2: potential aspects to COVID-19 treatment or prevention of progression of related diseases. *Cell Commun Signal*. May 15;21(1):110. doi: 10.1186/s12964-023-01104-5.
- Kanki T, Sakaguchi M, Kitamura A, Sato T, Mihara K, Hamasaki N., 2002. The tenth membrane region of band 3 is initially exposed to the luminal side of the endoplasmic reticulum and then integrated into a partially folded band 3 intermediate. *Biochemistry*. Nov 26;41(47):13973-81. doi: 10.1021/bi026619q.
- Karakaya M, Heller R, Kunde V, Zimmer KP, Chao CM, Nürnberg P, Cirak S., 2016. Novel Mutations in the Nonselective Sodium Leak Channel (NALCN) Lead to Distal Arthrogryposis with Increased Muscle Tone. *Neuropediatrics*. Aug;47(4):273-7. doi: 10.1055/s-0036-1584084.
- Karimova G, Ladant D., 2017. Defining Membrane Protein Topology Using pho-lac Reporter Fusions. *Methods Mol Biol*. 2017;1615:129-142. doi: 10.1007/978-1-4939-7033-9_10.
- Karimova G, Robichon C, Ladant D., 2009. Characterization of YmgF, a 72-residue inner membrane protein that associates with the Escherichia coli cell division machinery. *J Bacteriol*. Jan;191(1):333-46. doi: 10.1128/JB.00331-08.
- Kauko A, Hedin LE, Thebaud E, Cristobal S, Elofsson A, von Heijne G., 2010. Repositioning of transmembrane alpha-helices during membrane protein folding. *J Mol Biol*. Mar 19;397(1):190-201. doi: 10.1016/j.jmb.2010.01.042.
- Kendrew JC, Bodo G, Dintzis HM, Parrish RG, Wyckoff H, Phillips DC., 1958. A three-dimensional model of the myoglobin molecule obtained by x-ray analysis. *Nature*. Mar 8;181(4610):662-6. doi: 10.1038/181662a0.
- Khan, D. et al., 2023. A viral pan-end RNA element and host complex define a SARS-CoV-2 regulon. *Nat Commun* 14, 3385 (2023). <https://doi.org/10.1038/s41467-023-39091-3>
- Kielkopf CL, Bauer W, Urbatsch IL., 2020. Bradford Assay for Determining Protein Concentration. *Cold Spring Harb Protoc*. Apr 1;2020(4):102269. doi: 10.1101/pdb.prot102269.
- Kim HJ. et al., 2007. Gain-of-function mutation in TRPML3 causes the mouse Varitint-Waddler phenotype. *J Biol Chem*. 282, 36138–36142, doi: 10.1074/jbc.C700190200.
- Kim Y, Jedrzejczak R, Maltseva NI, Wilamowski M, Endres M, Godzik A, Michalska K, Joachimiak A., 2020. Crystal structure of Nsp15 endoribonuclease NendoU from SARS-CoV-2. *Protein Sci*. Jul;29(7):1596-1605. doi: 10.1002/pro.3873.
- Kim Y, Wower J, Maltseva N, Chang C, Jedrzejczak R, Wilamowski M, Kang S, Nicolaescu V, Randall G, Michalska K, Joachimiak A., 2021. Tipiracil binds to uridine site and inhibits Nsp15 endoribonuclease NendoU from SARS-CoV-2. *Commun Biol*. Feb 9;4(1):193. doi: 10.1038/s42003-021-01735-9.
- Kim, B. J. et al., 2012. Involvement of Na(+)-leak channel in substance P-induced depolarization of pacemaking activity in interstitial cells of Cajal. *Cell. Physiol. Biochem*. 29, 501–510. doi: 10.1159/000338504.
- Kirchhoff, M. et al., 2009. Phenotype and 244k array-CGH characterization of chromosome 13q deletions: an update of the phenotypic map of 13q21.1-qter. *Am. J. Med. Genet. A* 149A, 894–905. doi: 10.1002/ajmg.a.32814.
- Klatte N, Shields DC, Agoni C., 2022. Modelling the Transitioning of SARS-CoV-2 nsp3 and nsp4 Luminal Regions towards a More Stable State on Complex Formation. *Int J Mol Sci*. Dec 31;24(1):720. doi: 10.3390/ijms24010720.
- Koroglu, C., Seven, M., and Tolun, A., 2013. Recessive truncating NALCN mutation in infantile neuroaxonal dystrophy with facial dysmorphism. *J. Med. Genet*. 50, 515–520. doi: 10.1136/jmedgenet-2013-101634.
- Krishnan, K. S., and Nash, H. A., 1990. A genetic study of the anesthetic response: mutants of *Drosophila melanogaster* altered in sensitivity to halothane. *Proc. Natl. Acad. Sci. U.S.A.* 87, 8632–8636. doi: 10.1073/pnas.87.21.8632.
- Krogh A, Larsson B, von Heijne G, Sonnhammer EL., 2001. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J Mol Biol*. Jan 19;305(3):567-80. doi: 10.1006/jmbi.2000.4315.
- Kschonsak M. et al., 2020. Structure of the human sodium leak channel NALCN. *Nature*. Nov;587(7833):313-318. doi: 10.1038/s41586-020-2570-8.

- Kumaki T, Enomoto Y, Aida N, Goto T, Kurosawa K., 2022. Progression of cerebral and cerebellar atrophy in congenital contractures of limbs and face, hypotonia, and developmental delay. *Pediatr Int. Jan*;64(1):e14734. doi: 10.1111/ped.14734.
- Kumar P, Bansal M., 2015. Dissecting π -helices: sequence, structure and function. *FEBS J. Nov*;282(22):4415-32. doi: 10.1111/febs.13507.
- Kutlu B. et al., 2009. Detailed transcriptome atlas of the pancreatic beta cell. *BMC Med Genomics. Jan* 15;2:3. doi: 10.1186/1755-8794-2-3.
- Lalani, S. R. et al., 2013. Rare DNA copy number variants in cardiovascular malformations with extracardiac abnormalities. *Eur. J. Hum. Genet.* 21, 173–181. doi: 10.1038/ejhg.2012.155.
- Landolt-Marticorena C, Williams KA, Deber CM, Reithmeier RA., 1993. Non-random distribution of amino acids in the transmembrane segments of human type I single span membrane proteins. *J Mol Biol.* Feb 5;229(3):602-8. doi: 10.1006/jmbi.1993.1066.
- Lapointe CP, Grosely R, Johnson AG, Wang J, Fernández IS, Puglisi JD., 2021. Dynamic competition between SARS-CoV-2 NSP1 and mRNA on the human ribosome inhibits translation initiation. *Proc Natl Acad Sci U S A.* Feb 9;118(6):e2017715118. doi: 10.1073/pnas.2017715118.
- Lee AG., 2011. Biological membranes: the importance of molecular detail. *Trends Biochem Sci.* Sep;36(9):493-500. doi: 10.1016/j.tibs.2011.06.007.
- Lee, Y. et al., 2013. Prognostic implications of genetic variants in advanced non-small cell lung cancer: a genomewide association study. *Carcinogenesis* 34, 307–313. doi: 10.1093/carcin/bgs356.
- Lei X. et al., 2020. Activation and evasion of type I interferon responses by SARS-CoV-2. *Nat Commun.* Jul 30;11(1):3810. doi: 10.1038/s41467-020-17665-9.
- Lemmon MA, Flanagan JM, Hunt JF, Adair BD, Bormann BJ, Dempsey CE, Engelman DM., 1992. Glycophorin A dimerization is driven by specific interactions between transmembrane alpha-helices. *J Biol Chem.* Apr 15;267(11):7683-9.
- Leong LE., 1999. The use of recombinant fusion proteases in the affinity purification of recombinant proteins. *Mol Biotechnol.* Oct;12(3):269-74. doi: 10.1385/MB:12:3:269.
- Li F. et al., 2021. Highlighting membrane protein structure and function: A celebration of the Protein Data Bank. *J Biol Chem.* Jan-Jun;296:100557. doi: 10.1016/j.jbc.2021.100557.
- Li Q. et al., 2020. Early transmission dynamics in Wuhan, China, of novel Coronavirus-infected pneumonia. *N Engl J Med.* 2020; 382:1199–207. 4.
- Li, Y. et al., 2020. A Novel Missense Mutation in NALCN Cause CLIFAHDD Syndrome and Prenatal Diagnosis in China. 10.21203/rs.3.rs-136131/v1.
- Liacci AM. et al., 2021. Structure of the human signal peptidase complex reveals the determinants for signal peptide cleavage. *Mol Cell.* Oct 7;81(19):3934-3948.e11. doi: 10.1016/j.molcel.2021.07.031.
- Liao Z, Liu Y, Wang Y, Lu Q, Peng Y, Liu Q., 2022. Case Report: A *de novo* Variant in NALCN Associated with CLIFAHDD Syndrome in a Chinese Infant. *Front Pediatr.* Jul 13;10:927392. doi: 10.3389/fped.2022.927392.
- Lin DL, Inoue T, Chen YJ, Chang A, Tsai B, Tai AW., 2019. The ER Membrane Protein Complex Promotes Biogenesis of Dengue and Zika Virus Non-structural Multi-pass Transmembrane Proteins to Support Infection. *Cell Rep.* May 7;27(6):1666-1674.e4. doi: 10.1016/j.celrep.2019.04.051.
- Lin S. et al., 2021. Crystal structure of SARS-CoV-2 nsp10 bound to nsp14-ExoN domain reveals an exoribonuclease with both structural and functional integrity. *Nucleic Acids Res.* May 21;49(9):5382-5392. doi: 10.1093/nar/gkab320.
- Lozic B, Johansson S, Lovric Kojundzic S, Markic J, Knappskog PM, Hahn AF, Boman H., 2016. Novel NALCN variant: altered respiratory and circadian rhythm, anesthetic sensitivity. *Ann Clin Transl Neurol.* Oct 11;3(11):876-883. doi: 10.1002/acn3.362.
- Lu Y, Michel HA, Wang PH, Smith GL., 2022. Manipulation of innate immune signaling pathways by SARS-CoV-2 non-structural proteins. *Front Microbiol.* Nov 21;13:1027015. doi: 10.3389/fmicb.2022.1027015.

- Lu Y, Turnbull IR, Bragin A, Carveth K, Verkman AS, Skach WR., 2000. Reorientation of aquaporin-1 topology during maturation in the endoplasmic reticulum. *Mol Biol Cell*. Sep;11(9):2973-85. doi: 10.1091/mbc.11.9.2973.
- Lu, B. et al., 2007. The neuronal channel NALCN contributes resting sodium permeability and is required for normal respiratory rhythm. *Cell* 129, 371–383. doi: 10.1016/j.cell.2007.02.041.
- MacKenzie KR, Prestegard JH, Engelman DM., 1997. A transmembrane helix dimer: structure and implications. *Science*. Apr 4;276(5309):131-3. doi:10.1126/science.276.5309.131.
- Mackenzie KR., 2006. Folding and stability of alpha-helical integral membrane proteins. *Chem Rev*. May;106(5):1931-77. doi: 10.1021/cr0404388.
- Malone B, Urakova N, Snijder EJ, Campbell EA., 2022. Structures and functions of coronavirus replication-transcription complexes and their relevance for SARS-CoV-2 drug design. *Nat Rev Mol Cell Biol*. Jan;23(1):21-39. doi: 10.1038/s41580-021-00432-z.
- Martínez-Gil L, Johnson AE, Mingarro I., 2010. Membrane insertion and biogenesis of the Turnip crinkle virus p9 movement protein. *J Virol*. Jun;84(11):5520-7. doi: 10.1128/JVI.00125-10.
- Martínez-Gil L. et al., 2011. Membrane protein integration into the endoplasmic reticulum. *FEBS J*. Oct;278(20):3846-58. doi: 10.1111/j.1742-4658.2011.08185.x.
- McDowell MA, Heimes M, Sinning I., 2021. Structural and molecular mechanisms for membrane protein biogenesis by the Oxa1 superfamily. *Nat Struct Mol Biol*. Mar;28(3):234-239. doi: 10.1038/s41594-021-00567-9.
- McGilvray PT., 2020. An ER translocon for multi-pass membrane protein biogenesis. *Elife*. Aug 21;9:e56889. doi: 10.7554/eLife.56889.
- McGoldrick LL. et al., 2018. Opening of the human epithelial calcium channel TRPV6. *Nature*. Jan 11;553(7687):233-237. doi: 10.1038/nature25182.
- McPherson A, Gavira JA., 2014. Introduction to protein crystallization. *Acta Crystallogr F Struct Biol Commun*. Jan;70(Pt 1):2-20. doi: 10.1107/S2053230X13033141.
- Meacock SL, Lecomte FJ, Crawshaw SG, High S., 2002. Different transmembrane domains associate with distinct endoplasmic reticulum components during membrane integration of a polytopic protein. *Mol Biol Cell*. Dec;13(12):4114-29. doi: 10.1091/mbc.e02-04-0198.
- Medlock AE, Dailey TA, Ross TA, Dailey HA, Lanzilotta WN., 2007. A pi-helix switch selective for porphyrin deprotonation and product release in human ferrochelatase. *J Mol Biol*. Nov 2;373(4):1006-16. doi: 10.1016/j.jmb.2007.08.040.
- Miguel-Romero L. et al., 2020. Structural and Functional Characterization of Autophosphorylation in Bacterial Histidine Kinases. *Methods Mol Biol*. 2020; 2077:121-140. doi: 10.1007/978-1-4939-9884-5_9.
- Miller-Vedam LE., 2020. Structural and mechanistic basis of the EMC-dependent biogenesis of distinct transmembrane clients. *Elife*. Nov 25;9: e62611. doi: 10.7554/eLife.62611.
- Mingarro I, Elofsson A, von Heijne G., 1997. Helix-helix packing in a membrane-like environment. *J Mol Biol*. Oct 3;272(4):633-41. doi: 10.1006/jmbi.1997.1276.
- Monteil A, Guérineau NC, Gil-Nagel A, Parra-Díaz P, Lory P, Senatore A., 2024. New insights into the physiology and pathophysiology of the atypical sodium leak channel NALCN. *Physiol Rev*. Jan 1;104(1):399-472. doi: 10.1152/physrev.00014.2022.
- Moreira RA, Guzman HV, Boopathi S, Baker JL, Poma AB., 2020. Characterization of Structural and Energetic Differences between Conformations of the SARS-CoV-2 Spike Protein. *Materials (Basel)*. Nov 26;13(23):5362. doi: 10.3390/ma13235362.
- Nakayama, M., Iida, M., Koseki, H., and Ohara, O., 2006. A gene-targeting approach for functional characterization of KIAA genes encoding extremely large proteins. *FASEB J*. 20, 1718–1720. doi: 10.1096/fj.06-5952fje.
- Nash, H. A., Scott, R. L., Lear, B. C., and Allada, R., 2002. An unusual cation channel mediates photic control of locomotion in *Drosophila*. *Curr. Biol*. 12, 2152–2158. doi: 10.1016/S0960-9822(02)01358-1.
- Nieto-Torres JL. et al., 2014. Severe acute respiratory syndrome coronavirus envelope protein ion channel activity promotes virus fitness and pathogenesis. *PLoS Pathog*. May 1;10(5): e1004077. doi: 10.1371/journal.ppat.1004077.

- Nilsson I, von Heijne G., 1990. Fine-tuning the topology of a polytopic membrane protein: role of positively and negatively charged amino acids. *Cell*. Sep 21;62(6):1135-41. doi: 10.1016/0092-8674(90)90390-z.
- Nilsson IM, von Heijne G., 1993. Determination of the distance between the oligosaccharyltransferase active site and the endoplasmic reticulum membrane. *J Biol Chem*. Mar 15;268(8):5798-801.
- Oberai A, Joh NH, Pettit FK, Bowie JU., 2009. Structural imperatives impose diverse evolutionary constraints on helical membrane proteins. *Proc Natl Acad Sci U S A*. Oct 20;106(42):17747-50. doi: 10.1073/pnas.0906390106.
- O'Donnell JP. et al., 2020. The architecture of EMC reveals a path for membrane protein insertion. *Elife*. May 27;9: e57887. doi: 10.7554/eLife.57887.
- Ohta T, Nagano K, Yoshida M., 1986. The active site structure of Na⁺/K⁺-transporting ATPase: location of the 5'-(p-fluorosulfonyl) benzoyl adenosine binding site and soluble peptides released by trypsin. *Proc Natl Acad Sci U S A*. Apr;83(7):2071-5. doi: 10.1073/pnas.83.7.2071.
- O'Keefe S, Pool MR, High S., 2021. Membrane protein biogenesis at the ER: the highways and byways. *FEBS J*. Nov;289(22):6835-6862. doi: 10.1111/febs.15905.
- Onofrio, A. et al., 2014. Distance-depending hydrophobic-hydrophobic contacts in protein folding simulations. *Phys. Chem. Chem. Phys.* Jul (16): 18907-18917 doi:16. 10.1039/C4CP01131G.
- Oostra M. et al., 2007. Localization and membrane topology of coronavirus nonstructural protein 4: involvement of the early secretory pathway in replication. *J Virol*. Nov; 81(22):12323-36. doi: 10.1128/JVI.01506-07.
- Oostra M. et al., 2008. Topology and membrane anchoring of the coronavirus replication complex: not all hydrophobic domains of nsp3 and nsp6 are membrane spanning. *J Virol*. 2008;82: 12392–405.
- Ortíz-López FJ. et al., 2015. Cyclic colisporifungin and linear cavinafungins, antifungal lipopeptides isolated from *Colispora cavincola*. *J Nat Prod*. Mar 27;78(3):468-75. doi: 10.1021/np500854j.
- Orzáez M, Salgado J, Giménez-Giner A, Pérez-Payá E, Mingarro I. Influence of proline residues in transmembrane helix packing. *J Mol Biol*. 2004 Jan 9;335(2):631-40. doi: 10.1016/j.jmb.2003.10.062.
- Orzáez M. et al., 2005. Influence of hydrophobic matching on association of model transmembrane fragments containing a minimised glycophorin A dimerisation motif. *FEBS Lett*. Mar 14;579(7):1633-8. doi: 10.1016/j.febslet.2005.01.078.
- Ou X. et al., 2020. Characterization of spike glycoprotein of SARS-CoV-2 on virus entry and its immune cross-reactivity with SARS-CoV. *Nat Commun*. Mar 27;11(1):1620. doi: 10.1038/s41467-020-15562-9.
- Oudshoorn D. et al., 2017. Expression and Cleavage of Middle East Respiratory Syndrome Coronavirus nsp3-4 Polyprotein Induce the Formation of Double-Membrane Vesicles That Mimic Those Associated with Coronaviral RNA Replication. *mBio*. Nov 21;8(6): e01658-17. doi: 10.1128/mBio.01658-17.
- Overduin M, Kervin TA, Klarenbach Z, Adra TRC, Bhat RK., 2023. Comprehensive classification of proteins based on structures that engage lipids by COMPOSEL. *Biophys Chem*. Apr;295:106971. doi: 10.1016/j.bpc.2023.106971.
- Pahmeier F. et al., 2023. Identification of host dependency factors involved in SARS-CoV-2 replication organelle formation through proteomics and ultrastructural analysis. *J Virol*. Nov 30;97(11):e0087823. doi: 10.1128/jvi.00878-23.
- Pan X. et al., 2018. Structure of the human voltage-gated sodium channel Nav1.4 in complex with β 1. *Science*. Oct 19;362(6412):eaau2486. doi: 10.1126/science.aau2486.
- Panjikar S, Parthasarathy V, Lamzin VS, Weiss MS, Tucker PA., 2005. Auto-rickshaw: an automated crystal structure determination platform as an efficient tool for the validation of an X-ray diffraction experiment. *Acta Crystallogr D Biol Crystallogr*. Apr;61(Pt 4):449-57. doi: 10.1107/S0907444905001307.
- Park E, Rapoport TA., 2012. Mechanisms of Sec61/SecY-mediated protein translocation across membranes. *Annu Rev Biophys*. 2012;41: 21-40. doi: 10.1146/annurev-biophys-050511-102312.

- Park GJ. et al., 2022. The mechanism of RNA capping by SARS-CoV-2. *Nature*. Sep;609(7928):793-800. doi: 10.1038/s41586-022-05185-z.
- Parks GD, Lamb RA., 1991. Topology of eukaryotic type II membrane proteins: importance of N-terminal positively charged residues flanking the hydrophobic domain. *Cell*. Feb 22;64(4):777-87. doi: 10.1016/0092-8674(91)90507-u.
- Pauling L, Corey RB, Branson HR., 1951. The structure of proteins; two hydrogen-bonded helical configurations of the polypeptide chain. *Proc Natl Acad Sci U S A*. Apr;37(4):205-11. doi: 10.1073/pnas.37.4.205.
- Pauling, L., 1960. The Nature of the Chemical Bond. *Ithaca*, NY: Cornell University Press.
- Pedersen AV, Andreassen TF, Loland CJ., 2014. A conserved salt bridge between transmembrane segments 1 and 10 constitutes an extracellular gate in the dopamine transporter. *J Biol Chem*. Dec 12;289(50):35003-14. doi: 10.1074/jbc.M114.586982.
- Pergande M. et al., 2020. The genomic and clinical landscape of fetal akinesia. *Genet Med*. Mar;22(3):511-523. doi: 10.1038/s41436-019-0680-1.
- Pfeffer S. et al., 2015. Structure of the native Sec61 protein-conducting channel. *Nat Commun*. Sep 28;6: 8403. doi: 10.1038/ncomms9403.
- Pillon MC. et al., 2021. Cryo-EM structures of the SARS-CoV-2 endoribonuclease Nsp15 reveal insight into nuclease specificity and dynamics. *Nat Commun*. Jan 27;12(1):636. doi: 10.1038/s41467-020-20608-z.
- Planès R. et al., 2022. Human NLRP1 is a sensor of pathogenic coronavirus 3CL proteases in lung epithelial cells. *Mol Cell*. Jul 7;82(13):2385-2400.e9. doi: 10.1016/j.molcel.2022.04.033.
- Plath K, Mothes W, Wilkinson BM, Stirling CJ, Rapoport TA., 1998. Signal sequence recognition in posttranslational protein transport across the yeast ER membrane. *Cell*. Sep 18;94(6):795-807. doi: 10.1016/s0092-8674(00)81738-9.
- Popot JL, de Vitry C., 1990. On the microassembly of integral membrane proteins. *Annu Rev Biophys Chem*. 1990;19:369-403. doi: 10.1146/annurev.bb.19.060190.002101.
- Popot JL, Engelman DM., 2000. Helical membrane protein folding, stability, and evolution. *Annu Rev Biochem*. 2000;69:881-922. doi: 10.1146/annurev.biochem.69.1.881.
- Price WN. et al., 2009. Understanding the physical properties that control protein crystallization by analysis of large-scale experimental data. *Nat Biotechnol*. Jan;27(1):51-7. doi: 10.1038/nbt.1514.
- Punta, M., Kloppe, E., 2013. Membrane Protein Structure. In: Roberts, G.C.K. (eds) *Encyclopedia of Biophysics*. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-16712-6_422.
- Raden D, Song W, Gilmore R., 2000. Role of the cytoplasmic segments of Sec61alpha in the ribosome-binding and translocation-promoting activities of the Sec61 complex. *J Cell Biol*. Jul 10;150(1):53-64. doi: 10.1083/jcb.150.1.53.
- Raetz CR, Dowhan W., 1990. Biosynthesis and function of phospholipids in Escherichia coli. *J Biol Chem*. Jan 25;265(3):1235-8.
- Rahrmann EP. et al., 2022. The NALCN channel regulates metastasis and nonmalignant cell dissemination. *Nat Genet*. Dec;54(12):1827-1838. doi: 10.1038/s41588-022-01182-0.
- Ramírez AS, Locher KP., 2023 Structural and mechanistic studies of the N-glycosylation machinery: from lipid-linked oligosaccharide biosynthesis to glycan transfer. *Glycobiology*. Dec 25;33(11):861-872. doi: 10.1093/glycob/cwad053.
- Rapoport TA, Li L, Park E., 2017. Structural and Mechanistic Insights into Protein Translocation. *Annu Rev Cell Dev Biol*. Oct 6;33: 369-390. doi: 10.1146/annurev-cellbio-100616-060439.
- Ricciardi S. et al., 2022. The role of NSP6 in the biogenesis of the SARS-CoV-2 replication organelle. *Nature*. Jun; 606(7915):761-768. doi: 10.1038/s41586-022-04835-6.
- Roingard P. et al., 2022. The double-membrane vesicle (DMV): a virus-induced organelle dedicated to the replication of SARS-CoV-2 and other positive-sense single-stranded RNA viruses. *Cell Mol Life Sci*. 2022 Jul 16;79(8):425. doi: 10.1007/s00018-022-04469-x.
- Rosas-Lemus M. et al., 2020. High-resolution structures of the SARS-CoV-2 2'-O-methyltransferase reveal strategies for structure-based inhibitor design. *Sci Signal*. Sep 29;13(651): eabe1202. doi: 10.1126/scisignal.abe1202.

- Russ WP, Engelman DM., 2000. The GxxxG motif: a framework for transmembrane helix-helix association. *J Mol Biol.* Feb 25;296(3):911-9. doi: 10.1006/jmbi.1999.3489.
- Russell RB, Eggleston DS., 2000. New roles for structure in biology and drug discovery. *Nat Struct Biol.* Nov;7 Suppl:928-30. doi: 10.1038/80691.
- Sakai Y, Kawachi K, Terada Y, Omori H, Matsuura Y, Kamitani W., 2017. Two-amino acids change in the nsp4 of SARS coronavirus abolishes viral replication. *Virology.* Oct;510:165-174. doi: 10.1016/j.virol.2017.07.019.
- Sanders MR, Findlay HE, Booth PJ., 2018. Lipid bilayer composition modulates the unfolding free energy of a knotted α -helical membrane protein. *Proc Natl Acad Sci U S A.* Feb 20;115(8): E1799-E1808. doi: 10.1073/pnas.1714668115.
- Satoh T, Ohba A, Liu Z, Inagaki T, Satoh AK., 2015. dPob/EMC is essential for biosynthesis of rhodopsin and other multi-pass membrane proteins in *Drosophila* photoreceptors. *Elife.* Feb 26;4: e06306. doi: 10.7554/eLife.06306.
- Saurí A. et al., 2005. Double-spanning plant viral movement protein integration into the endoplasmic reticulum membrane is signal recognition particle-dependent, translocon-mediated, and concerted. *J Biol Chem.* Jul 8;280(27):25907-12. doi: 10.1074/jbc.M412476200.
- Saurí A, Tamborero S, Martínez-Gil L, Johnson AE, Mingarro I., 2009. Viral membrane protein topology is dictated by multiple determinants in its sequence. *J Mol Biol.* Mar 20;387(1):113-28. doi: 10.1016/j.jmb.2009.01.063.
- Sawicki SG. et al., 2005. Functional and genetic analysis of coronavirus replicase-transcriptase proteins. *PLoS Pathog.* Dec;1(4): e39. doi: 10.1371/journal.ppat.0010039.
- Schmidt V, Sturgis JN. et al., 2017. Making Monomeric Aquaporin Z by Disrupting the Hydrophobic Tetramer Interface. *ACS Omega.* Jun 28;2(6):3017-3027. doi: 10.1021/acsomega.7b00261.
- Schoeman D, Fielding BC., 2019. Coronavirus envelope protein: current knowledge. *Viol J.* May 27;16(1):69. doi: 10.1186/s12985-019-1182-0.
- Schubert K. et al., 2020. SARS-CoV-2 Nsp1 binds the ribosomal mRNA channel to inhibit translation. *Nat Struct Mol Biol.* Oct;27(10):959-966. doi: 10.1038/s41594-020-0511-8.
- Segatto O, King CR, Pierce JH, Di Fiore PP, Aaronson SA., 1988. Different structural alterations upregulate in vitro tyrosine kinase activity and transforming potency of the erbB-2 gene. *Mol Cell Biol.* Dec;8(12):5570-4. doi: 10.1128/mcb.8.12.5570-5574.1988.
- Senes A, Gerstein M, Engelman DM., 2000. Statistical analysis of amino acid patterns in transmembrane helices: the GxxxG motif occurs frequently and in association with beta-branched residues at neighboring positions. *J Mol Biol.* Feb 25;296(3):921-36. doi: 10.1006/jmbi.1999.3488.
- Seven, M., Ozkilic, A., and Yuksel, A., 2002. Dysmorphic face in two siblings with infantile neuroaxonal dystrophy. *Genet. Couns.* 13, 465–473.
- Sheikh A, Al-Taher A, Al-Nazawi M, Al-Mubarak AI, Kandeel M., 2020. Analysis of preferred codon usage in the coronavirus N genes and their implications for genome evolution and vaccine design. *J Virol Methods.* Mar;277:113806. doi: 10.1016/j.jviromet.2019.113806.
- Shin D. et al., 2020. Papain-like protease regulates SARS-CoV-2 viral spread and innate immunity. *Nature.* 2020;587:657–62.
- Singer SJ, Nicolson GL., 1972. The fluid mosaic model of the structure of cell membranes. *Science.* Feb 18;175(4023):720-31. doi: 10.1126/science.175.4023.720.
- Singh A, Upadhyay V, Upadhyay AK, Singh SM, Panda AK., 2015. Protein recovery from inclusion bodies of *Escherichia coli* using mild solubilization process. *Microb Cell Fact.* Mar 25;14:41. doi: 10.1186/s12934-015-0222-8.
- Singh P, Agrawal N, Maurya RK, Moirangthem A., 2022. Novel variant c.1838A>G, p.(Gln613Arg) in NALCN causes camptodactyly and cognitive delay. *Clin Dysmorphol.* Oct 1;31(4):206-210. doi: 10.1097/MCD.0000000000000431.
- Sinke, A. P. et al., 2011. Genetic analysis of mouse strains with variable serum sodium concentrations identifies the Nalcn sodium channel as a novel player in osmoregulation. *Physiol. Genomics* 43, 265–270. doi: 10.1152/physiolgenomics.00188.2010.
- Sivaraman I, Friedman NR, Prayson RA., 2016. Muscle biopsy findings in a child with NALCN gene mutation. *J Clin Neurosci.* Dec;34: 222-223. doi: 10.1016/j.jocn.2016.06.018.

- Skach WR. et al., 1994. Biogenesis and transmembrane topology of the CHIP28 water channel at the endoplasmic reticulum. *J Cell Biol.* May;125(4):803-15. doi: 10.1083/jcb.125.4.803.
- Snapp EL. et al., 2017. Structure and topology around the cleavage site regulate post-translational cleavage of the HIV-1 gp160 signal peptide. *Elife.* Jul 28;6: e26067. doi: 10.7554/eLife.26067.
- Sonnhammer EL, von Heijne G, Krogh A., 1998. A hidden Markov model for predicting transmembrane helices in protein sequences. *Proc Int Conf Intell Syst Mol Biol.* 1998;6: 175-82.
- Specia, D. J. et al., 2010. Conserved role of unc-79 in ethanol responses in lightweight mutant mice. *PLoS Genet.* 6:1057. doi: 10.1371/journal.pgen.1001057.
- Stark Z. et al., 2016. A prospective evaluation of whole-exome sequencing as a first-tier molecular test in infants with suspected monogenic disorders. *Genet Med.* Nov;18(11):1090-1096. doi: 10.1038/gim.2016.1.
- Suzuki R. et al., 2013. Signal peptidase complex subunit 1 participates in the assembly of hepatitis C virus through an interaction with E2 and NS2. *PLoS Pathog.* 2013;9(8): e1003589. doi: 10.1371/journal.ppat.1003589.
- Swayne LA. et al., 2009. The NALCN ion channel is activated by M3 muscarinic receptors in a pancreatic beta-cell line. *EMBO Rep.* Aug;10(8):873-80. doi: 10.1038/embor.2009.125.
- Tai VW, Imperiali B., 2001. Substrate specificity of the glycosyl donor for oligosaccharyl transferase. *J Org Chem.* Sep 21;66(19):6217-28. doi: 10.1021/jo0100345.
- Tam JZ, Palumbo T, Miwa JM, Chen BY., 2022. Analysis of Protein-Protein Interactions for Intermolecular Bond Prediction. *Molecules.* Sep 21;27(19):6178. doi: 10.3390/molecules27196178.
- Tamborero S, Vilar M, Martínez-Gil L, Johnson AE, Mingarro I., 2011. Membrane insertion and topology of the translocating chain-associating membrane protein (TRAM). *J Mol Biol.* Mar 4;406(4):571-82. doi: 10.1016/j.jmb.2011.01.009.
- Tang X. et al., 2017. EMC3 coordinates surfactant protein and lipid homeostasis required for respiration. *J Clin Invest.* Dec 1;127(12):4314-4325. doi: 10.1172/JCI94152.
- Thomas S., 2021. Mapping the Nonstructural Transmembrane Proteins of Severe Acute Respiratory Syndrome Coronavirus 2. *J Comput Biol.* Sep;28(9):909-921. doi: 10.1089/cmb.2020.0627.
- Tian S. et al., 2019. Proteomic Analysis Identifies Membrane Proteins Dependent on the ER Membrane Protein Complex. *Cell Rep.* Sep 3;28(10):2517-2526.e5. doi: 10.1016/j.celrep.2019.08.006.
- Trenker R, Call ME, Call MJ., 2015. Crystal Structure of the Glycophorin A Transmembrane Dimer in Lipidic Cubic Phase. *J Am Chem Soc.* Dec 23;137(50):15676-9. doi: 10.1021/jacs.5b11354.
- Ubarretxena-Belandia I, Engelman DM., 2001. Helical membrane proteins: diversity of functions in the context of simple architecture. *Curr Opin Struct Biol.* Jun;11(3):370-6. doi: 10.1016/s0959-440x(00)00217-7.
- Ulmschneider JP, Andersson M, Ulmschneider MB., 2011. Determining peptide partitioning properties via computer simulation. *J Membr Biol.* Jan;239(1-2):15-26. doi: 10.1007/s00232-010-9324-8.
- Ulmschneider MB, Sansom MS, Di Nola A., 2005. Properties of integral membrane protein structures: derivation of an implicit membrane potential. *Proteins.* May 1;59(2):252-65. doi: 10.1002/prot.20334.
- V'kovski, P. et al., 2021. Coronavirus biology and replication: implications for SARS-CoV-2. *Nat Rev Microbiol* 19, 155–170. doi: <https://doi.org/10.1038/s41579-020-00468-6>.
- Van den Berg B. et al., 2004. X-ray structure of a protein-conducting channel. *Nature.* Jan 1;427(6969):36-44. doi: 10.1038/nature02218.
- Van der Wel P. et al., 2021. Dihedral Angle Measurements for Structure Determination by Biomolecular Solid-State NMR Spectroscopy. *Front Mol Biosci.* Dec 6;8:791090. doi: 10.3389/fmolb.2021.791090.
- Van Geest M, Lolkema JS., 2000. Membrane topology and insertion of membrane proteins: search for topogenic signals. *Microbiol Mol Biol Rev.* Mar;64(1):13-33. doi: 10.1128/MMBR.64.1.13-33.2000.

- Van Hemert MJ. et al., 2008. SARS-coronavirus replication/transcription complexes are membrane-protected and need a host factor for activity in vitro. *PLoS Pathog.* May 2;4(5): e1000054. doi: 10.1371/journal.ppat.1000054.
- Van Meer G, Voelker DR, Feigenson GW., 2008. Membrane lipids: where they are and how they behave. *Nat Rev Mol Cell Biol.* Feb;9(2):112-24. doi: 10.1038/nrm2330.
- Visser LELM. et al., 2017. A clinical utility study of exome sequencing versus conventional genetic testing in pediatric neurology. *Genet Med.* Sep;19(9):1055-1063. doi: 10.1038/gim.2017.1.
- Vivero M. et al., 2017. Additional de novo missense genetic variants in NALCN associated with CLIFAHDD syndrome. *Clin Genet.* Jun;91(6):929-931. doi: 10.1111/cge.12899.
- Volkmar N. et al., 2019. The ER membrane protein complex promotes biogenesis of sterol-related enzymes maintaining cholesterol homeostasis. *J Cell Sci.* Jan 16;132(2): jcs223453. doi: 10.1242/jcs.223453.
- von Heijne G., 1986. Towards a comparative anatomy of N-terminal topogenic protein sequences. *J Mol Biol.* May 5;189(1):239-42. doi: 10.1016/0022-2836(86)90394-3.
- Von Heijne G., 1989. Control of topology and mode of assembly of a polytopic membrane protein by positively charged residues. *Nature.* Oct 5;341(6241):456-8. doi: 10.1038/341456a0.
- Von Heijne G., 1986. The distribution of positively charged residues in bacterial inner membrane proteins correlates with the trans-membrane topology. *EMBO J.* Nov;5(11):3021-7. doi: 10.1002/j.1460-2075.1986.tb04601.x.
- Voorhees RM, Fernández IS, Scheres SH, Hegde RS., 2014. Structure of the mammalian ribosome-Sec61 complex to 3.4 Å resolution. *Cell.* Jun 19;157(7):1632-43. doi: 10.1016/j.cell.2014.05.024.
- Voorhees RM, Hegde RS., 2016. Structure of the Sec61 channel opened by a signal sequence. *Science.* Jan 1;351(6268):88-91. doi: 10.1126/science.aad4992.
- Wahlberg JM, Spiess M., 1997. Multiple determinants direct the orientation of signal-anchor proteins: the topogenic role of the hydrophobic signal domain. *J Cell Biol.* May 5;137(3):555-62. doi: 10.1083/jcb.137.3.555.
- Wallin E, von Heijne G., 1998. Genome-wide analysis of integral membrane proteins from eubacterial, archaean, and eukaryotic organisms. *Protein Sci.* Apr;7(4):1029-38. doi: 10.1002/pro.5560070420.
- Wang Q. et al., 2020. Structural Basis for RNA Replication by the SARS-CoV-2 Polymerase. *Cell.* Jul 23;182(2):417-428.e13. doi: 10.1016/j.cell.2020.05.034.
- Wang Y, Koh K, Ichinose Y, Yasumura M, Ohtsuka T, Takiyama Y., 2016. A *de novo* mutation in the NALCN gene in an adult patient with cerebellar ataxia associated with intellectual disability and arthrogryposis. *Clin Genet.* Dec;90(6):556-557. doi: 10.1111/cge.12851.
- Wang Y. et al., 2021. Impact of Expressing Cells on Glycosylation and Glycan of the SARS-CoV-2 Spike Glycoprotein. *ACS Omega.* Jun 11;6(24):15988-15999. doi: 10.1021/acsomega.1c01785.
- Weaver TM. 2000. The π -helix translates structure into function. *Protein Sci* 9:201–206. 22.
- Welply JK, Shenbagamurthi P, Lennarz WJ, Naider F., 1983. Substrate recognition by oligosaccharyltransferase. Studies on glycosylation of modified Asn-X-Thr/Ser tripeptides. *J Biol Chem.* Oct 10;258(19):11856-63.
- Whitley P, Mingarro I., 2014. Stitching proteins into membranes, not sew simple. *Biol Chem.* Dec;395(12):1417-24. doi: 10.1515/hsz-2014-0205.
- Whittaker GR, Daniel S, Millet JK. Coronavirus entry: how we arrived at SARS-CoV-2. 2021. *Curr Opin Virol.* Apr;47:113-120. doi: 10.1016/j.coviro.2021.02.006.
- Wiener MC, White SH., 1992. Structure of a fluid dioleoylphosphatidylcholine bilayer determined by joint refinement of x-ray and neutron diffraction data. III. Complete structure. *Biophys J.* Feb;61(2):434-47. doi: 10.1016/S0006-3495(92)81849-0.
- Williams JM. et al., 2023. Reticulons promote formation of ER-derived double-membrane vesicles that facilitate SARS-CoV-2 replication. *J Cell Biol.* Jul 3;222(7):e202203060. doi: 10.1083/jcb.202203060.
- Winczewska-Wiktor A. et al., 2022. Central Apneas Due to the CLIFAHDD Syndrome Successfully Treated with Pyridostigmine. *Int J Environ Res Public Health.* Jan 11;19(2):775. doi: 10.3390/ijerph19020775.

- Wolff G. et al., 2020. A molecular pore spans the double membrane of the coronavirus replication organelle. *Science*. Sep 11;369(6509):1395-1398. doi: 10.1126/science.abd3629.
- Wu F. et al., 2020. A new coronavirus associated with human respiratory disease in China. *Nature*. Mar;579(7798):265-269. doi: 10.1038/s41586-020-2008-3.
- Xia H. et al., 2020. Evasion of Type I Interferon by SARS-CoV-2. *Cell Rep*. Oct 6;33(1):108234. doi: 10.1016/j.celrep.2020.108234.
- Xie J. et al., 2020. Structure of the human sodium leak channel NALCN in complex with FAM155A. *Nat Commun*. Nov 17;11(1):5831. doi: 10.1038/s41467-020-19667-z.
- Xu X. et al., 2009. Crystal structure of the C-terminal cytoplasmic domain of non-structural protein 4 from mouse hepatitis virus A59. *PLoS One*. Jul 10;4(7): e6217. doi: 10.1371/journal.pone.0006217.
- Xu Z. et al., 2022. SARS-CoV-2 impairs interferon production via NSP2-induced repression of mRNA translation. *Proc Natl Acad Sci U S A*. Aug 9;119(32): e2204539119. doi: 10.1073/pnas.2204539119.
- Yan L. et al., 2021. Cryo-EM Structure of an Extended SARS-CoV-2 Replication and Transcription Complex Reveals an Intermediate State in Cap Synthesis. *Cell*. Jan 7;184(1):184-193.e10. doi: 10.1016/j.cell.2020.11.016.
- Yeh, E. et al., 2008. A putative cation channel, NCA-1, and a novel protein, UNC-80, transmit neuronal activity in *C. elegans*. *PLoS Biol*. 6:e55. doi: 10.1371/journal.pbio. 0060055.
- Yin W. et al., 2021. Structural basis for inhibition of the SARS-CoV-2 RNA polymerase by suramin. *Nat Struct Mol Biol*. Mar;28(3):319-325. doi: 10.1038/s41594-021-00570-0.
- Yoshimoto FK., 2020. The Proteins of Severe Acute Respiratory Syndrome Coronavirus-2 (SARS CoV-2 or n-COV19), the Cause of COVID-19. *Protein J*. Jun;39(3):198-216. doi: 10.1007/s10930-020-09901-4.
- Zacharias J, Knapp EW., 2013. Geometry motivated alternative view on local protein backbone structures. *Protein Sci*. Nov; 22(11):1669-74. doi: 10.1002/pro.2364.
- Zanotti A. et al., 2022. The human signal peptidase complex acts as a quality control enzyme for membrane proteins. *Science*. Dec 2;378(6623):996-1000. doi: 10.1126/science.abo5672.
- Zeng W. et al., 2020. Biochemical characterization of SARS-CoV-2 nucleocapsid protein. *Biochem Biophys Res Commun*. Jun 30;527(3):618-623. doi: 10.1016/j.bbrc.2020.04.136.
- Zhang C, Meng X, Zhao H., 2022. Comparison of Cell Fusions Induced by Influenza Virus and SARS-CoV-2. *Int J Mol Sci*. Jul 1;23(13):7365. doi: 10.3390/ijms23137365.
- Zhang R. et al., 2016. A CRISPR screen defines a signal peptide processing pathway required by flaviviruses. *Nature*. Jul 7;535(7610):164-8. doi: 10.1038/nature18625.
- Zhou FX, Merianos HJ, Brunger AT, Engelman DM., 2001. Polar residues drive association of polyleucine transmembrane helices. *Proc Natl Acad Sci U S A*. Feb 27;98(5):2250-5. doi: 10.1073/pnas.041593698.
- Zhou XE, Melcher K, Xu HE., 2012. Structure and activation of rhodopsin. *Acta Pharmacol Sin*. Mar;33(3):291-9. doi: 10.1038/aps.2011.171.
- Zhou Y, Lau FW, Nauli S, Yang D, Bowie JU., 2001. Inactivation mechanism of the membrane protein diacylglycerol kinase in detergent solution. *Protein Sci*. Feb;10(2):378-83. doi: 10.1110/ps.34201.
- Zhu N. et al., 2020. A Novel Coronavirus from Patients with Pneumonia in China, 2019. *N Engl J Med*. Feb 20;382(8):727-733. doi: 10.1056/NEJMoa2001017.
- Zubcevic L, Lee SY., 2019. The role of π -helices in TRP channel gating. *Curr Opin Struct Biol*. Oct;58:314-323. doi: 10.1016/j.sbi.2019.06.011.

ANEXOS

ANEXO 1: Oligonucleótidos utilizados en la presente tesis

Tabla 19. Oligonucleótidos usados para la construcción del sistema C_LTM en el vector pSVLd

Oligo	Secuencia (5' –3')
Mutagenesis (ASS--->NSS)	F-CAGGGTCAGGTAGCTGCTATTCATGTACTTGTGTTGCTC R-GAGCAACAACAAGTACATGAATAGCAGCTACCTGACCCTG
Mutagenesis (QVT--->NVT)	F-TGTGACCTTCATGAGTGACATTGCAGCTGTAAGTCTATGC R-GCATAGCAGTTACAGCTGCAATGTCACTCATGAAGGTCACA
Inserción del sitio SpeI	F- GGCTCTGGGTCCACTAGTGGTACCGGCAGCGGCAGC R- GCTGCCGCTGCCGGTACCACTAGTGGACCCAGAGCC
Inserción del sitio SacI	F- CACTAGTGGTACCTCTGAGCTCGGCAGCGGCAGCG R- CGCTGCCGCTGCCGAGCTCAGAGGTACCACTAGTG
Inserción de etiqueta c-myc	F- CGATCCCTCGAGATGGCCTGG R-GCGGGATCCTTACTACAGGTCTCTCTGAGATCAGCTTCTGCTCGCTGCTGGTCACGTT

Tabla 20. Oligonucleótidos utilizados para clonar los 6 TM del DI de NALCN individualmente en el sistema C_LTM-3G.

Oligo	Secuencia (5' –3')
S1-DINALCN	F- CGGACTAGTGGTGGTCCTGGAAGCCTGCTGAGAAT R- CGGGAGCTCCCCACCAGGTCCGAAGGTCATAGGGG
S2-DINALCN	F- CGGACTAGTGGTGGTCCTGGACCACTGCAGTACGT R- CGGGAGCTCCCCACCAGGTCCGGCGATCATCTCGG
S3-DINALCN	F- CGGACTAGTGGTGGTCCTGGACGTTGGTGCGTGTT R- CGGGAGCTCCCCACCAGGTCCGAACACCTGGAGCA
S4-DINALCN	F-CGGACTAGTGGTGGTCCTGGATGGGGCATGCTGAG R-CGGGAGCTCCCCACCAGGTCCGAATCTGAAGTAGA
S5-DINALCN	F- CGGACTAGTGGTGGTCCTGGAATTTGGAGCGTGTC R- CGGGAGCTCCCCACCAGGTCCGAACATCTGCACGC
S6-DINALCN	F- CGGACTAGTGGTGGTCCTGGAAGATGGCGGAGCTA R- CGGGAGCTCCCCACCAGGTCCGATGATCACGGCAA
Oligonucleótidos de controles (Inserción y translocación)	F- GAGACTAGTGGTGGTCCTGGA R- CGGGAGCTCCCCACCAGGTCC

Tabla 21. Oligonucleótidos usados para clonar el segmento TM₁₀₋₃₂ de NSP4 en C_LTM-2G

Oligo	Secuencia (5' –3')
TM ₁₀₋₃₂ NSP4	F-CGGACTAGTGGTGGTCCTGGACTTATTAAAGTAACCCTG R- CGGGAGCTCCCCACCAGGTCCCACGTGGACCGGTGTTAT

Tabla 22. Oligonucleótidos utilizados para clonar los polipéptidos truncados del DI de NALCN en el vector de bacterias pKTop

Oligo	Secuencia (5' –3')
FORWARD	F-CGGTCTAGAGATGCTGAAGAGAAA
REVERSE	
F61 (PKTOP-DINALCN ₁₋₆₁)	R-CGGGAGCTCGGGAAGGTCATAGGGG
A88 (PKTOP-DINALCN ₁₋₈₈)	R-CGGGAGCTCGGGGCGATCATCTCGG
F127 (PKTOP-DINALCN ₁₋₁₂₇)	R-CGGGAGCTCGGGAACACCTGGAGCA
F160 (PKTOP-DINALCN ₁₋₁₆₀)	R-CGGGAGCTCGGGAATCTGAAGTAGA
F200 (PKTOP-DINALCN ₁₋₂₀₀)	R-CGGGAGCTCGGGAACATCTGCACGC
M380 (PKTOP-DINALCN ₁₋₃₈₀)	R-CGGGAGCTCGGCATCTTCTGCAGGCA

Tabla 23. Oligonucleótidos utilizados para clonar los polipéptidos truncados de NSP4 en el vector de bacterias pKTop.

Oligonucleótidos	Secuencia (5' –3')
FORWARD	F-CGGTCTAGAGATGAAGATTGTAAACAAT
REVERSE	
E141 (PKTOP-NSP4 ₁₋₁₄₁)	R-CGGGAGCTCGGTTCAATCAGCTTGCTTGG
G164 (PKTOP-NSP4 ₁₋₁₆₄)	R-CGGGAGCTCGGTCCGCTCGCATCCTTAAA
D279 (PKTOP-NSP4 ₁₋₂₇₉)	R-CGGGAGCTCGGATCCAGTGCGCCAATGGG
H313 (PKTOP-NSP4 ₁₋₃₁₃)	R-CGGGAGCTCGGGTGGGAATATTCCCCGAA
Y335 (PKTOP-NSP4 ₁₋₃₃₅)	R-CGGGAGCTCGGGTAGACAGGAGTCAAACA
Y346 (PKTOP-NSP4 ₁₋₃₄₆)	R-CGGGAGCTCGGGTAAATAACTGAGTATAC
H362 (PKTOP-NSP4 ₁₋₃₆₂)	R-CGGGAGCTCGGGTAAATAACTGAGTATAC
P371 (PKTOP-NSP4 ₁₋₃₇₁)	R-CGGGAGCTCGGTGGCGTAAACATCACCAT
Y397 (PKTOP-NSP4 ₁₋₃₉₇)	R-CGGGAGCTCGGGTAGTTACTAAAGAACCA
Q 500 (PKTOP-NSP4 ₁₋₅₀₀)	R-CGGGAGCTCGGCTGGAGTACGGCTGAGGT

Tabla 24. Oligonucleótidos usados para modificar la diana de glicosilación no canónica de NSP4

Oligo	Secuencia (5' –3')
Mutagenesis NIC--->QIC	F-GTTTTTAGCGCGGTAGGCCAGATCTGCTATACGCCAAGC R-GCTTGGCGTATAGCAGATCTGGCCTACCGCGCTAAAAAC
Mutagenesis NIC--->NIS	F-CGCGGTAGGCAATATCTCTATACGCCAAGCA R-TGCTTGGCGTATAGGAGATATTGCCTACCGCG

Tabla 25. Oligonucleótidos usados para clonar los polipéptidos truncados del dominio DI de NALCN y de NSP4 en el vector de eucariotas pCAGGS

Oligo	Secuencia (5' –3')
FORWARD	
NST _{Nt} DINALCN	gctcatcgatgcatgccaccatgggcagcggcaacagcacctgaagagaaagCAGAGC
QST _{Nt} DINALCN	gctcatcgatgcatgccaccatgggcagcggcCAAagcacctgaagagaaagCAGAGC
REVERSE	
T1DINALCN-cmyc-NST _{Ct}	ctcgagcatgcccggttatcacatgcccattggtgctgttcaggctcctcctcgctgatcagcttctgctcagggtagtgctcgaaggt
T1DINALCN-cmyc-QST _{Ct}	ctcgagcatgcccggttatcacatgcccattggtgctgttcaggctcctcctcgctgatcagcttctgctcagggtagtgctcgaaggt
T2DINALCN-cmyc-NST _{Ct}	ctcgagcatgcccggttatcacatgcccattggtgctgttcaggctcctcctcgctgatcagcttctgctccttcacgattccccggat
T2DINALCN-cmyc-QST _{Ct}	ctcgagcatgcccggttatcacatgcccattggtgctgttcaggctcctcctcgctgatcagcttctgctccttcacgattccccggat
T3DINALCN-cmyc-NST _{Ct}	ctcgagcatgcccggttatcacatgcccattggtgctgttcaggctcctcctcgctgatcagcttctgctccatctggtccacgatgtc
T3DINALCN-cmyc-QST _{Ct}	ctcgagcatgcccggttatcacatgcccattggtgctgttcaggctcctcctcgctgatcagcttctgctccatctggtccacgatgtc
T4DINALCN-cmyc-NST _{Ct}	ctcgagcatgcccggttatcacatgcccattggtgctgttcaggctcctcctcgctgatcagcttctgctcgggcagctcgaatctgaa
T4DINALCN-cmyc-QST _{Ct}	ctcgagcatgcccggttatcacatgcccattggtgctgttcaggctcctcctcgctgatcagcttctgctcgggcagctcgaatctgaa
T5DINALCN-cmyc-NST _{Ct}	ctcgagcatgcccgggtcacaggctcctcctcgctgatcagcttctgctccatgcccattggtgctgttcagttcagggtgca
T5DINALCN-cmyc-QST _{Ct}	ctcgagcatgcccgggtcacaggctcctcctcgctgatcagcttctgctccatgcccattggtgctGTGcagttcagggtgca
T6DINALCN-cmyc-NST _{Ct}	ctcgagcatgcccggttatcacatgcccattggtgctgttcaggctcctcctcgctgatcagcttctgctccatcatcttctgcaggcaggctgg
T6DINALCN-cmyc-QST _{Ct}	ctcgagcatgcccggttatcacatgcccattggtgctgttcaggctcctcctcgctgatcagcttctgctccatcatcttctgcaggcaggctgg

Oligo	Secuencia (5' –3')
FORWARD	
NSP4-N _t	GCTCATCGATGCATGCCACCATGGGCAGCGGCAAGATTGTAAACAATTGG
REVERSE	
NSP4-cmyc-NST-C _t	CTCGAGCATGCCCGTTATCACATGCCCATGGTGCTGTTTCAAGTCTCTCGTGATCAGCTTCTGCTCTGTCCGGGCCGAGCTCGG
NSP4-cmyc-QST-C _t	CTCGAGCATGCCCGTTATCACATGCCCATGGTGCTCTGCAGGTCTCTCGTGATCAGCTTCTGCTCTGTCCGGGCCGAGCTCGG

Tabla 26. Oligonucleótidos usados para modificar el sitio de corte de la peptidasa SPC de NSP4

Oligo	Secuencia (5' –3')
Mutagenesis TPVH--->AAAA	F-TCTTGTTTGTGCGCCGAATCTTTTACTTGATAGCAGCCGCCGCGTGATGTCTAAACACACAG R-CTGTGTGTTTAGACATCACGGCGGGCGGCTGCTATCAAGTAAAAGATTGCGGCGACAAACAAGA
Mutagenesis TPVH--->FFFF	F-TCTTGTTTGTGCGCCGAATCTTTTACTTGATATTCTTCTTCTTCGTGATGTCTAAACACACAG R-CTGTGTGTTTAGACATCACGAAGAAGAAGAATATCAAGTAAAAGATTGCGGCGACAAACAAGA

Tabla 27. Oligonucleótidos usados para clonar el segmento S6 del DI de NALCN en sistema Lep, diseñar las mutaciones CLIFAHDD en S6 y reemplazar residuos específicos en S4 del DI de NALCN

Oligo	Secuencia (5' –3')
S6 DINALCN EN LEP	F- CGGACTAGTGGTGGTCCTGGACGTTGGCGTTCCTAC R- CGGGGTACCCCCACCAGGTCCAATGATAACAGCAAT
MUTAGÉNESIS L312I EN S6	F-GATCTTCTTCCTGGCCTGGATCGTCAAGAACGTGTTCATTG R-CAATGAACACGTTCTTGACGATCCAGGCCAGGAAGAAGATC
MUTAGÉNESIS L312V EN S6	F-TCTTCTGCGCTGGGTGGTCAAGAACGTG R-CACGTTCTTGACCACCCAGGCCAGGAAGA
MUTAGÉNESIS V313G EN S6	F-CCTGGCCTGGCTGGGCAAGAACGTGTTCA R-TGAACACGTTCTTGCCCAGCCAGGCCAGG
MUTAGÉNESIS A319V EN S6	F-CAAGAACGTGTTTCATTGTCGTGATCATCGGACCTG R-CAGGTCCGATGATCACGACAATGAACACGTTCTTG
MUTAGÉNESIS I322T EN S6	F-ACGTGTTTCATTGCCGTGATCACCGGACCTGGTG R-CACCAGGTCCGGTGATCACGGCAATGAACACGT
MUTAGÉNESIS D75R EN S4	F-GTACGTGACCTTCACACTGCGGACCCTGCTGATGTTCCCTGT R-GTACGTGACCTTCACACTGCGGACCCTGCTGATGTTCCCTGT
MUTAGÉNESIS N216Q EN WT	F-GATACCAAGCCTGGCCAGGTGACCTGGAACAGC R-GCTGTTCCAGGTACCTGGCCAGGCTTGGTATC
MUTACIÓN N210 EN WT	F- CTACCACTGCGTGGTCCAAGATACCAAGCCTGGCC R- GGCCAGGCTTGGTATCTTGGACCACGCAGTGGTAG

Tabla 28. Oligonucleótidos para clonar las regiones solubles de NSP4 en el vector pETNKI 1.1

Oligo	Secuencia (5' –3')
NSP4 (S40-P270)	F-CAGGGACCCGGTATGAGTAGTGAGATAATCGGATAT R-CGAGGAGAAGCCCGGTTAAGGCGTAAACATATTCGTCAA
NSP4 (T411-Q500)	F-CAGGGACCCGGTATGACGTTTGAAGAGGCAGCACTG R-CGAGGAGAAGCCCGGTTACTGGAGTACGGCTGAGGTTAT

ANEXO 2: Construcciones realizadas en la Tesis

A. Construcciones en el sistema C_LTM: La cadena ligera (C_L) de C_LTM (subrayado), sitios de restricción *Xho*I y *Bam*HI (en gris), el sitio de corte del péptido señal de C_LTM (en fucsia), las dianas de glicosilación (en rojo), el sitio de clonación de segmentos TM Spel y SacI (en verde azulado), los tetrapéptidos flanqueantes de segmentos TM (en turquesa), los segmentos TM (en amarillo) y la etiqueta c-myc (en verde)

Sistema C_LTM-3G

PDRSLEMAWISLILSLLALSSGAI SQA/GQ PKSSPSVTLFPPSSEELETNKATLVCTITDFYPG
VVTVDWKVDGTPVTQGMETTQPSKQSNKYM NSSYLTLTARAWERHSSYS NVT HEGHTVEKSL
SRADSRGSGSGSGSGS TSGTSEI GSGSGSGSGS NVTSS EQKLISEEDL GSRHDK

Sistema C_LTM-2G

PDRSLEMAWISLILSLLALSSGAI SQA/GQ PKSSPSVTLFPPSSEELETNKATLVCTITDFYPG
VVTVDWKVDGTPVTQGMETTQPSKQSNKYM ASSYLTLTARAWERHSSYS NVT HEGHTVEKSL
SRADSRGSGSGSGSGS TSGTSEI GSGSGSGSGS NVTSS EQKLISEEDL GSRHDK

Segmentos TM de controles clonados en el sistema C_LTM-2G

C_LTM-2G+1022 (Control de inserción en membrana)

METTQPSKQSNKYM ASSYLTLTARAWERHSSYS NVT HEGHTVEKSLSRADSRGSGSGSGSGS
TS GGPGKVLIVTGVLG LLLIKWK GPGG ELGSGSGSGSGS NVTSS EQKLISEEDL GSRHDK

C_LTM-2G+753 (Control de translocación (no inserción) en membrana)

METTQPSKQSNKYM ASSYLTLTARAWERHSSYS NVT HEGHTVEKSLSRADSRGSGSGSGSGS
TS GGPGYNLYIYAFFASSLVHGF GPGG ELGSGSGSGSGS NVTSS EQKLISEEDL GSRHDK

C_LTM-2G+TM₁₀₋₃₂ de NSP4

METTQPSKQSNKYM ASSYLTLTARAWERHSSYS NVT HEGHTVEKSLSRADSRGSGSGSGSGS
TS GGPGLIKVTLVFLFVA AIFYLITPVHV GPGG ELGSGSGSGSGS NVTSS EQKLISEEDL GSR

Segmentos TM del dominio DI de NALCN clonados en CLTM-3G

S1 de DI de NALCN clonado en CLTM-3G

PDRSLEMAWISLILSLLALSSGAI SQA/GQ PKSSPSVTLFPPSSEELETNKATLVCTITDFYPG
VVTVDWKVDGTPVTQGMETTQPSKQSNKYM NSSYLTLTARAWERHSSYS NVT HEGHTVEKSL
SRADSRGSGSGSGSGS TSGGPG SLLRICAIISVISVCMNTPMT GPGG ELGSGSGSGSGS NVTSS
EQKLISEEDL GSRHDK

S2 de DI de NALCN clonado en C_LTM-3G

PDRSLEMAWISLILSLLALSSGAI SQA/GQ PKSSPSVTLFPPSSEELETNKATLVCTITDFYPG
VVTVDWKVDGTPVTQGMETTQPSKQSNKYM NSSYLTLTARAWERHSSYSC NVT HEGHTVEKSL
SRADSRGSGSGSGSGS TSGGPG PLQYVFTFLDTLLMFLYTAEMIA GPGGEL GSGSGSGSGS NVT
SS EQKLISEEDL GSRHDK

S3 de DI de NALCN clonado en C_LTM-3G

PDRSLEMAWISLILSLLALSSGAI SQA/GQ PKSSPSVTLFPPSSEELETNKATLVCTITDFYPG
VVTVDWKVDGTPVTQGMETTQPSKQSNKYM NSSYLTLTARAWERHSSYSC NVT HEGHTVEKSL
SRADSRGSGSGSGSGS TSGGPG RWCVFDFGMVFC LWVSLVLQVF GPGGEL GSGSGSGSGS NVT S
S EQKLISEEDL GSRHDK

S4 de DI de NALCN clonado en C_LTM-3G

PDRSLEMAWISLILSLLALSSGAI SQA/GQ PKSSPSVTLFPPSSEELETNKATLVCTITDFYPG
VVTVDWKVDGTPVTQGMETTQPSKQSNKYM NSSYLTLTARAWERHSSYSC NVT HEGHTVEKSL
SRADSRGSGSGSGSGS TSGGPG WGMLRIPRPLIMIRAFRIYFRF GPGGEL GSGSGSGSGS NVT S
S EQKLISEEDL GSRHDK

S5 de DI de NALCN clonado en C_LTM-3G

PDRSLEMAWISLILSLLALSSGAI SQA/GQ PKSSPSVTLFPPSSEELETNKATLVCTITDFYPG
VVTVDWKVDGTPVTQGMETTQPSKQSNKYM NSSYLTLTARAWERHSSYSC NVT HEGHTVEKSL
SRADSRGSGSGSGSGS TSGGPG IWSVSIFLLFLLLYGILGVQMF GPGGEL GSGSGSGSGS NVT
SS EQKLISEEDL GSRHDK

S6 de DI de NALCN clonado en C_LTM-3G

PDRSLEMAWISLILSLLALSSGAI SQA/GQ PKSSPSVTLFPPSSEELETNKATLVCTITDFYPG
VVTVDWKVDGTPVTQGMETTQPSKQSNKYM NSSYLTLTARAWERHSSYSC NVT HEGHTVEKSL
SRADSRGSGSGSGSGS TSGGPG RWRSYFYFITLIFFLAWLVKNVFI AVII GPGGEL GSGSGSGS
GS NVT SS EQKLISEEDL GSRHDK

B. Proteínas clonadas en el vector pKTop: Segmento del vector pKTop (en gris), dianas naturales de glicosilación (en rojo), proteína PhoA (letras azules) y subunidad α de LacZ (en letras rojas), primer y último aminoácido de la proteína (en oliva).

Polipéptidos truncados del DI de NALCN clonados en el vector pKTop

61-mer (1-61)

MTMITPSLHACRSTLE MLKRKQSSRV EAQPVTDFGP DESLSDNADI LWINKPWWH_S
LLRICAIIISVISVCMNTPMT FPSSARTPEMPVLENRAAQGDITAPGGARRLTGDQTAALRDSLS
DKPAKNIILLIGDGMGDSEITAARNYAEGAGGFFKGIDALPLTGQYTHYALNKKTKGPDYVTD
AASATAWSTGVKTYNGALGVDIHEKDHPTILEMAKAAGLATGNVSTAELQDATPAALVAHVTSR
KCYGPSATSEKCPGNALEKGGKGSITEQLLNARADVTLGGAKTFAETATAGEWQKTLREQAQ
ARGYQLVSDAASLSNVTEANQQKPLLGLFADGNMPVRWLGPATYHGNI DKPAVTCTPNPQRND
SVPTLAQMTDKAIELLSKNEKGFFLQVEGASIDKQDHAANPCGQIGETVDLDEAVQRALEFAKK
EGNTLVIVTADHAHASQIVAPDTKAPGLTQALNTKDGA VMMSYGNSEEDSQEHTGSQRLRIAY

GPHAANVVGLTDQTDLFYTMKAALGLKTD~~SLAVVLQRRD~~WENPGVTQLNRLAAHPPFASWRNSE
EARTDRPSQQLRSLNGEWR-

88-mer (1-88)

MTMITPSLHACRSTLE MLKRKQSSRV EAQPVTDGFP DESLSDNADI LWINKPWVH~~S~~
LLRICAIIISVISVCMNTPMTFEHYPLQYVTF~~TLDTLLMFLY~~TAEMI~~A~~PSSARTPEMPVLENRA
AQGDITAPGGARRLTGDQTAALRDSLS~~DKPAKNIILLIGDGMGDSEITAARNYAEGAGGFFKGI~~
DALPLTGQYTHYALNKKTGKPDYVTD~~SAASATAWSTGVKTYNGALGVDIHEKDHPTILEMAKAA~~
GLATGNVSTAELQDATPAALVAHVTSRKCYGPSATSEKCPGNALEKGGKGSITEQLLNARADVT
LGGGAKTFAETATAGEWQGKTLREQAQARGYQLVSDAASLNSVTEANQQKPLLGLFADGNMPVR
WLGPKATYHGNIDKPAVTCPTNPQRNDSVPTLAQMTDKAIELLSKNEKGFFLQVEGASIDKQDH
AANPCGQIGETVDLDEAVQRALEFAKKEGNTLVIVTADHAHASQIVAPDTKAPGLTQALNTKDG
AVMMSYGNSEEDSQEHTGSQ~~LRIAAYGPHAANVVGLTDQTDLFYTMKAALGLKTD~~SLAVVLQ~~R~~
RDWENPGVTQLNRLAAHPPFASWRNSEEARTDRPSQQLRSLNGEWR-

127-mer (1-127)

MTMITPSLHACRSTLE MLKRKQSSRV EAQPVTDGFP DESLSDNADI LWINKPWVH~~S~~
LLRICAIIISVISVCMNTPMTFEHYPLQYVTF~~TLDTLLMF~~LYTAEMIAKM HIRGIVKGDS
SYVKDRWC~~VDFGFMVFCLWVSLVLQV~~IPSSARTPEMPVLENRAAQGDITAPGGARRLTGDQTA
ALRDSLS~~DKPAKNIILLIGDGMGDSEITAARNYAEGAGGFFKGI~~DALPLTGQYTHYALNKKTGK
PDYVTD~~SAASATAWSTGVKTYNGALGVDIHEKDHPTILEMAKAA~~GLATGNVSTAELQDATPAALV
AHVTSRKCYGPSATSEKCPGNALEKGGKGSITEQLLNARADVTLGGGAKTFAETATAGEWQGK
TLREQAQARGYQLVSDAASLNSVTEANQQKPLLGLFADGNMPVRWLGPKATYHGNIDKPAVTCPT
NPQRNDSVPTLAQMTDKAIELLSKNEKGFFLQVEGASIDKQDHAANPCGQIGETVDLDEAVQRA
LEFAKKEGNTLVIVTADHAHASQIVAPDTKAPGLTQALNTKDGAVMMSYGNSEEDSQEHTGSQ
LRIAAYGPHAANVVGLTDQTDLFYTMKAALGLKTD~~SLAVVLQRRD~~WENPGVTQLNRLAAHPPFA
SWRNSEEARTDRPSQQLRSLNGEWR-

160-mer (1-160)

MTMITPSLHACRSTLE MLKRKQSSRV EAQPVTDGFP DESLSDNADI LWINKPWVH~~S~~
LLRICAIIISVISVCMNTPMTFEHYPLQYVTF~~TLDTLLMF~~LYTAEMIAKM HIRGIVKGDS
SYVKDRWC~~VF~~DGFMVFCLWVSLVLQVFEIA DIVDQMSPWG MLRIPRPLIM IRAFRIYFR~~E~~
PSSARTPEMPVLENRAAQGDITAPGGARRLTGDQTAALRDSLS~~DKPAKNIILLIGDGMGDSEIT~~
AARNYAEGAGGFFKGI~~DALPLTGQYTHYALNKKTGKPDYVTD~~SAASATAWSTGVKTYNGALGVD
IHEKDHPTILEMAKAA~~GLATGNVSTAELQDATPAALVAHVTSRKCYGPSATSEKCPGNALEKGG~~
KGSITEQLLNARADVTLGGGAKTFAETATAGEWQGKTLREQAQARGYQLVSDAASLNSVTEANQ
QKPLLGLFADGNMPVRWLGPKATYHGNIDKPAVTCPTNPQRNDSVPTLAQMTDKAIELLSKNEK
GFFLQVEGASIDKQDHAANPCGQIGETVDLDEAVQRALEFAKKEGNTLVIVTADHAHASQIVAP
DTKAPGLTQALNTKDGAVMMSYGNSEEDSQEHTGSQ~~LRIAAYGPHAANVVGLTDQTDLFYTMK~~
AALGLKTD~~SLAVVLQRRD~~WENPGVTQLNRLAAHPPFASWRNSEEARTDRPSQQLRSLNGEWR-

200-mer (1-200)

MTMITPSLHACRSTLE MLKRKQSSRV EAQPVTDGFP DESLSDNADI LWINKPWVH~~S~~
LLRICAIIISVISVCMNTPMTFEHYPLQYVTF~~TLDTLLMF~~LYTAEMIAKM HIRGIVKGDS
SYVKDRWC~~VF~~DGFMVFCLWVSLVLQVFEIA DIVDQMSPWG MLRIPRPLIM IRAFRIYFRF
ELPRTRITNILKRSGEQIWSVSIFLLFFLLLYGILGVQMIPSSARTPEMPVLENRAAQGDITAP
GGARRLTGDQTAALRDSLS~~DKPAKNIILLIGDGMGDSEITAARNYAEGAGGFFKGI~~DALPLTGQ
YTHYALNKKTGKPDYVTD~~SAASATAWSTGVKTYNGALGVDIHEKDHPTILEMAKAA~~GLATGNVS
TAE~~LQDATPAALVAHVTSRKCYGPSATSEKCPGNALEKGGKGSITEQLLNARADVT~~LGGGAKTF
AETATAGEWQGKTLREQAQARGYQLVSDAASLNSVTEANQQKPLLGLFADGNMPVRWLGPKATY

HGNIDKPAVTCTPNPQRNDSVPTLAQMTDKAIELLSKNEKGFFLQVEGASIDKQDHAANPCGQI
GETVDLDEAVQRALEFAKKEGNTLVIVTADHAHASQIVAPDTKAPGLTQALNTKDGAVMVMSYG
NSEEDSQEHTGSQLRIAAYGPHAANVVGLTDQTDLFYTMKAALGLK**TDSLAVVLQRRDWENPGV**
TQLNRLAAHPPFASWRNSEEARTDRPSQQLRSLNGEWR–

Dominio DI completo 380-mer (1-380)

MTMITPSLHACRSTLE **M**LKRKQSSRV EAQPVTDGFP DESLSDNADI LWINKPWVH_S
LLRICAIIISV ISVCMNTPMT FEHYPPLOQYV TFTLDTLLMF LYTAEMIAKM HIRGIVKGDS
SYVKDRWCVF DGFMVFLWV SLVLQVFEIA DIVDQMSPWG MLRIPRPLIM IRAFRIYFRF
ELPRTRITNI LKRSGEQIWS VSIFLLFFLL LYGILGVQMF GTFTYHCVVN DTKPGNVTWN
SLAIPDTHCS PELEEGYQCP PGFKCMDLED LGLSRQELGY SGFNEIGTSI FTVYEAASQE
GWVFLMYRAI DSFPRWRSYF YFITLIFFLA WLVKNVFIIV IETFAEIRV QFQQMWGSR
STTSTATTQMFHEDAAGGWQLVAVDVNKPQGRAPACLOK **P**SSARTPE**MPVLENRAAQGDITAP**
GGARRLTGDQTAALRDSLSDKPAKNIILLIGDGMGDSEITAARNYAEGAGGFFKGIDALPLTGQ
YTHYALNKKTGKPDYVTDSAASATAWSTGVKTYNGALGVDIHEKDHPTILEMAKAAGLATGNVS
TAELODATPAALVAHVTSRKCYGPSATSEKCPGNALEKGGKGSITEQLLNARADVTLGGGAKTF
AETATAGEWQGTKTLREQAQARGYQLVSDAASLNSVTEANQQKPLLGLFADGNMPVRWLGP
KATYHGNIDKPAVTCTPNPQRNDSVPTLAQMTDKAIELLSKNEKGFFLQVEGASIDKQDHAANPCGQI
GETVDLDEAVQRALEFAKKEGNTLVIVTADHAHASQIVAPDTKAPGLTQALNTKDGAVMVMSYG
NSEEDSQEHTGSQLRIAAYGPHAANVVGLTDQTDLFYTMKAALGLK**TDSLAVVLQRRDWENPGV**
TQLNRLAAHPPFASWRNSEEARTDRPSQQLRSLNGEWR–

Polipéptidos truncados de NSP4 clonados en el vector pKTop

141-mer (1-141)

MTMITPSLHACRSTLEM **M**IVNNWLKQLIKVTLVFLFVAAIFYLITPVHVMKHTDFSSEIIGYK
AIDGGVTRDIASTDTCFANKRADFDTWFSQRGGSYTNDKACPLIAAVITREVGFFVPGLP
GTILRTTNGDFLHFLPRVFSAVGNICYTPSKLIE **P**SSARTPE**MPVLENRAAQGDITAPGGARRLTGDQ**
TAALRDSLSDKPAKNIILLIGDGMGDSEITAARNYAEGAGGFFKGIDALPLTGQYTHYALNKK
TGKPDYVTDSAASATAWSTGVKTYNGALGVDIHEKDHPTILEMAKAAGLATGNVSTAELQ
DATPAALVAHVTSRKCYGPSATSEKCPGNALEKGGKGSITEQLLNARADVTLGGGAKTF
AETATAGEWQGTKTLREQAQARGYQLVSDAASLNSVTEANQQKPLLGLFADGNMPVRWL
GPKATYHGNIDKPAVTCTPNPQRNDSVPTLAQMTDKAIELLSKNEKGFFLQVEGASIDKQDHAANPCGQI
GETVDLDEAVQRALEFAKKEGNTLVIVTADHAHASQIVAPDTKAPGLTQALNTKDGAVMVMSYGN
SEEDSQEHTGSQLRIAAYGPHAANVVGLTDQTDLFYTMKAALGLK**TDSLAVVLQRRDWENPGVTQLNRLAAH**
PPASWRNSEEARTDRPSQQLRSLNGEWR–

164-mer (1-164)

MTMITPSLHACRSTLEM **M**IVNNWLKQLIKVTLVFLFVAAIFYLITPVHVMKHTDFSSEIIGYK
AIDGGVTRDIASTDTCFANKRADFDTWFSQRGGSYTNDKACPLIAAVITREVGFFVPGLP
GTILRTTNGDFLHFLPRVFSAVGNICYTPSKLIEYTDFAVSACVLAAECTIFKDS **G**SSARTPE**MPV**
LENRAAQGDITAPGGARRLTGDQTAALRDSLSDKPAKNIILLIGDGMGDSEITAARNYAEGAGG
FFKGIDALPLTGQYTHYALNKKTGKPDYVTDSAASATAWSTGVKTYNGALGVDIHEKDHPTILE
MAKAAGLATGNVSTAELQDATPAALVAHVTSRKCYGPSATSEKCPGNALEKGGKGSITEQLLN
ARADVTLGGGAKTF AETATAGEWQGTKTLREQAQARGYQLVSDAASLNSVTEANQQKPLLGLFADG
NMPVRWLGPATYHGNIDKPAVTCTPNPQRNDSVPTLAQMTDKAIELLSKNEKGFFLQVEGASI
DKQDHAANPCGQIGETVDLDEAVQRALEFAKKEGNTLVIVTADHAHASQIVAPDTKAPGLTQAL
NTKDGAVMVMSYGNSEEDSQEHTGSQLRIAAYGPHAANVVGLTDQTDLFYTMKAALGLK**TDSL**
VVLQRRDWENPGVTQLNRLAAHPPFASWRNSEEARTDRPSQQLRSLNGEWR–

279-mer (1-279)

MTMITPSLHACRSTLEMKIVNNWLKQLIKVTLVFLFVAAIFYLITPVHVMKHTDFSSEIIGYK
AIDGGVTRDIASTDTCFANKRADFDTWFSQRGGSYTNDKACPLIAAVITREVGFFVPLPGTIL
RTTNGDFLHFLPRVFSAVGNICYTPSKLIEYTDFAVSACVLAEECTIFKDASGKPVPCYDTNV
LEGSVAYESLRPDTRYVLMDSIIQFPNTYLEGSVRVVTTFDSEYCRHGTCESEAGVCVSTSG
RWVLNNDYYRNLPGVFCGVDAVNLLTNMFTPLIQPIGALDPSSARTPEMPVLENRAAQGDITAP
GGARRLTGDQTAALRDSLSKPAKNIILLIGDGMGDSEITAARNYAEGAGGFFKGIDALPLTGQ
YTHYALNKKTKGPDYVTDASAATAWSTGVKTYNGALGVDIHEKDHPTILEMAKAAGLATGNVS
TAEHQDATPAALVAHVTSRKCYGPSATSEKCPGNALEKGGKGSITEQLLNARADVTLGGGAKTF
AETATAGEWQGKTLREQAARGYQLVSDAASLNSVTEANQQKPLLGLFADGNMPVRWLGPATY
HGNIDKPAVTCTPNPQRNDSVPTLAQMTDKAIELLSKNEKGFFLQVEGASIDKQDHAANPCGQI
GETVDLDEAVQRALEFAKKEGNTLVIVTADHAHASQIVAPDTKAPGLTQALNTKDGAVMVMSYG
NSEEDSQEHTGSQRLIAAYGPHAANVVGLTDQTDLFYTMKAALGLKTDSLAVVLQRRDWENPGV
TQLNRLAAHPPFASWRNSEEARTDRPSQQLRSLNGEWR-

313-mer (1-313)

MTMITPSLHACRSTLEMKIVNNWLKQLIKVTLVFLFVAAIFYLITPVHVMKHTDFSSEIIGYK
AIDGGVTRDIASTDTCFANKHADFDTWFSQRGGSYTNDKACPLIAAVITREVGFFVPLPGTIL
RTTNGDFLHFLPRVFSAVGNICYTPSKLIEYTDFAVSACVLAEECTIFKDASGKPVPCYDTNV
LEGSVAYESLRPDTRYVLMDSIIQFPNTYLEGSVRVVTTFDSEYCRHGTCESEAGVCVSTSG
RWVLNNDYYRSLPGVFCGVDAVNLLTNMFTPLIQPIGALDISASIVAGGIVAIIVVTCLAYYFMR
FRRAFGEYSHPSSARTPEMPVLENRAAQGDITAPGGARRLTGDQTAALRDSLSKPAKNIILLI
GDGMGDSEITAARNYAEGAGGFFKGIDALPLTGQYTHYALNKKTKGPDYVTDASAATAWSTGV
KTYNGALGVDIHEKDHPTILEMAKAAGLATGNVSTAELQDATPAALVAHVTSRKCYGPSATSEK
CPGNALEKGGKGSITEQLLNARADVTLGGGAKTFAETATAGEWQGKTLREQAARGYQLVSDAA
SLNSVTEANQQKPLLGLFADGNMPVRWLGPATYHGNIDKPAVTCTPNPQRNDSVPTLAQMTDK
AIELLSKNEKGFFLQVEGASIDKQDHAANPCGQIGETVDLDEAVQRALEFAKKEGNTLVIVTAD
HAHASQIVAPDTKAPGLTQALNTKDGAVMVMSYGNSEEDSQEHTGSQRLIAAYGPHAANVVGLT
DQTDLFYTMKAALGLKTDSLAVVLQRRDWENPGVTQLNRLAAHPPFASWRNSEEARTDRPSQQL
RSLNGEWR-

335-mer (1-335)

MTMITPSLHACRSTLEMKIVNNWLKQLIKVTLVFLFVAAIFYLITPVHVMKHTDFSSEIIGYK
AIDGGVTRDIASTDTCFANKHADFDTWFSQRGGSYTNDKACPLIAAVITREVGFFVPLPGTIL
RTTNGDFLHFLPRVFSAVGNICYTPSKLIEYTDFAVSACVLAEECTIFKDAGGKPVPCYDTNV
LEGSVAYESLRPDTRYVLMDSIIQFPNTYLEGSVRVVTTFDSEYCRHGTCESEAGVCVSTSG
RWVLNNDYYRSLPGVFCGVDAVNLLTNMFTPLIQPIGALDISASIVAGGIVAIIVVTCLAYYFMR
FRRAFGEYSHVAFNTLLFLMSFTVLCLTPVHPSSARTPEMPVLENRAAQGDITAPGGARRLTG
DQTAALRDSLSKPAKNIILLIGDGMGDSEITAARNYAEGAGGFFKGIDALPLTGQYTHYALNK
KTGKPDYVTDASAATAWSTGVKTYNGALGVDIHEKDHPTILEMAKAAGLATGNVSTAELQDAT
PAALVAHVTSRKCYGPSATSEKCPGNALEKGGKGSITEQLLNARADVTLGGGAKTFAETATAGE
WQGKTLREQAARGYQLVSDAASLNSVTEANQQKPLLGLFADGNMPVRWLGPATYHGNIDKPA
VTCTPNPQRNDSVPTLAQMTDKAIELLSKNEKGFFLQVEGASIDKQDHAANPCGQIGETVDLDE
AVQRALEFAKKEGNTLVIVTADHAHASQIVAPDTKAPGLTQALNTKDGAVMVMSYGNSEEDSQE
HTGSQRLIAAYGPHAANVVGLTDQTDLFYTMKAALGLKTDSLAVVLQRRDWENPGVTQLNRLAA
HPPFASWRNSEEARTDRPSQQLRSLNGEWR-

346-mer (1-346)

MTMITPSLHACRSTLEMIVNNWLKQLIKVTLVFLFVAAIFYLITPVHVMSKHTDFSSEIIGYR
AIDGGVTRDIASTDTCFANKHADFDTWFSQRGGSYTNGKACPLIAAVITREVG FVVPGLPGTIL
RTTNGDFLHFLPRVFSAVGNICYTPSKLIEYTD FATSACVLAAECTIFKDASGKPVPCYDTNV
LEGSVAYESLRPDTRYVLM DGSIIQFPNTYLEGSVRVVTTFDSEYCRHGT CERSEAGVCVSTSG
RWVLNNDYYRSLPGVFCGVDAVNLLTNMFTPLIQPIGALDISASIVAGGIVAI VVTCLAYYFMR
FRRAFGEYSHV VAFNTLLFLMSFTVLCLTPVYSFLPGVYSVIY P SSARTPEMPVLENRAAQGDI
TAPGGARRLTGDQTAALRDSLSDKPAKNI ILLIGDGMGDSEITAARNYAEGAGGFFKGIDALPL
TGQYTHYALNKKTGKPDYVTDSAASATAWSTGVKTYNGALGVDIHEKDHPTILEMAKAAGLATG
NVSTAELQDATPAALVAHVTSRKCYGPSATSEKCPGNALEKGGKGSITEQLLNARADVTLGGGA
KTFAETATAGEWQGKTLREQAQARGYQLVSDAASLNSVTEANQQKPLLGLFADGNMPVRWLGP
KATYHGNIDKPAVTCTPNPQRNDSVPTLAQMTDKAIELLSKNEKGFFLQVEGASIDKQDHAANPC
GQIGETVDLDEAVQRALEFAKKEGNTLVIVTADHAHASQIVAPDTKAPGLTQALNTKDGAVMVM
SYGNSEEDSQEHTGSQLRIAAYGPHAANVVGLTDQTDLFYTMKAALGLKTDSLAVVLQRRDWEN
PGVTQLNRLAAHPPFASWRNSEEARTDRPSQQLRSLNGEWR-

362-mer (1-362)

MTMITPSLHACRSTLEMIVNNWLKQLIKVTLVFLFVAAIFYLITPVHVMSKHTDFSSEIIGYK
AIDGGVTRDIASTDTCFANKRADFDTWFSQRGGSYTNDKACPLIAAVITREVG FVVPGLPGTIL
RTTNGDFLHFLPRVFSAVGNICYTPSKLIEYTD FATSACVLAAECTIFKDASGKPVPCYDTNV
LEGSVAYESLRPDTRYVLM DGSIIQFPNTYLEGSVRVVTTFDSEYCRHGT CERSEAGVCVSTSG
RWVLNNDYYRNLPGVFCGVDAVNLLTNMFTPLIQPIGALDISASIVAGGIVAI VVTCLAHYFMR
FRRAFGEYSHV VAFNTLLFLMSFTVLCLTPVYSFLPGVYSVIYLYLTFYLTNDVSFLAHPSSAR
TPEMPVLENRAAQGDI TAPGGARRLTGDQTAALRDSLSDKPAKNI ILLIGDGMGDSEITAARNY
AEGAGGFFFKGIDALPLTGQYTHYALNKKTGKPDYVTDSAASATAWSTGVKTYNGALGVDIHEKD
HPTILEMAKAAGLATGNVSTAELQDATPAALVAHVTSRKCYGPSATSEKCPGNALEKGGKGSIT
EQLLNARADVTLGGGAKTFAETATAGEWQGKTLREQAQARGYQLVSDAASLNSVTEANQQKPLL
GLFADGNMPVRWLGP KATYHGNIDKPAVTCTPNPQRNDSVPTLAQMTDKAIELLSKNEKGFFLQ
VEGASIDKQDHAANPCGQIGETVDLDEAVQRALEFAKKEGNTLVIVTADHAHASQIVAPDTKAP
GLTQALNTKDGAVMVM SYGNSEEDSQEHTGSQLRIAAYGPHAANVVGLTDQTDLFYTMKAALGL
KTDSLAVVLQRRDWENPGVTQLNRLAAHPPFASWRNSEEARTDRPSQQLRSLNGEWR-

371-mer (1-371)

MTMITPSLHACRSTLEMIVNNWLKQLIKVTLVFLFVAAIFYLITPVHVMSKHTDFSSEIIGYK
AIDGGVTRDIASTDTCFANKRADFDTWFSQRGGSYTNDKACPLIAAVITREVG FVVPGLPGTIL
RTTNGDFLHFLPRVFSAVGNICYTPSKLIEYTD FATSACVLAAECTIFKDASGKPVPCYDTNV
LEGSVAYESLRPDTRYVLM DGSIIQFPNTYLEGSVRVVTTFDSEYCRHGT CERSEAGVCVSTSG
RWVLNNDYYRNLPGVFCGVDAVNLLTNMFTPLIQPIGALDISASIVAGGIVAI VVTCLAHYFMR
FRRAFGEYSHV VAFNTLLFLMSFTVLCLTPVYSFLPGVYSVIYLYLTFYLTNDVSFLAHIQWMV
MFT P SSARTPEMPVLENRAAQGDI TAPGGARRLTGDQTAALRDSLSDKPAKNI ILLIGDGMGD
SEITAARNYAEGAGGFFFKGIDALPLTGQYTHYALNKKTGKPDYVTDSAASATAWSTGVKTYNGA
LGVDIHEKDHPTILEMAKAAGLATGNVSTAELQDATPAALVAHVTSRKCYGPSATSEKCPGNAL
EKGGKGSITEQLLNARADVTLGGGAKTFAETATAGEWQGKTLREQAQARGYQLVSDAASLNSVT
EANQQKPLLGLFADGNMPVRWLGP KATYHGNIDKPAVTCTPNPQRNDSVPTLAQMTDKAIELLS
KNEKGFFLQVEGASIDKQDHAANPCGQIGETVDLDEAVQRALEFAKKEGNTLVIVTADHAHASQ
IVAPDTKAPGLTQALNTKDGAVMVM SYGNSEEDSQEHTGSQLRIAAYGPHAANVVGLTDQTDLF
YTMKAALGLKTDSLAVVLQRRDWENPGVTQLNRLAAHPPFASWRNSEEARTDRPSQQLRSLNGE
WR-

397-mer (1-397)

MTMITPSLHACRSTLEMKIVNNWLKQLIKVTLVFLFVAAIFYLITPVHVMSKHTDFSSEIIGYK
AIDGGVTRDIASTDTCFANKRADFDTWFSQRGGSYTNDKACPLIAAVITREVGFFVPGLPGTIL
RTTNGDFLHFLPRVFSAVGNICYTPSKLIEYTDFAVSACVLAEECTIFKDASGKPVPCYDTNV
LEGSVAYESLRPDTRYVLMDSIIQFPNTYLEGSRVVTTFDSEYCRHGTCESEAGVCVSTSG
RWVLNNDYYRNLPGVFCGVDAVNLLTNMFTPLIQPIGALDISASIVAGGIVAIIVVTCLAHYFMR
FRRAFGEYSHVAFNTLLFLMSFTVLCLTPVYSFLPGVYSVIYLYLTFYLTNDVSFLAHIQWMV
MFTPLVPFWITIAIYIICISTKHFYWFFSNYPSSARTPEMPVLENRAAQGDITAPGGARRLTGDQ
TAALRDSLSDKPAKNIILLIGDGMGDESEITAARNYAEGAGGFFKGIDALPLTGQYTHYALNKKT
GKPDYVTDASAATAWSTGVKTYNGALGVDIHEKDHPTILEMAKAAGLATGNVSTAELQDATPA
ALVAHVTSRKCYGPSATSEKCPGNALEKGGKGSITEQLLNARADVTLGGGAKTFAETATAGEWQ
GKTLREQAQARGYQLVSDAASLNSVTEANQQKPLLGLFADGNMPVRWLGPATYHGNIIDKPAVT
CTPNPQRNDSVPTLAQMTDKAIELLSKNEKGFFLQVEGASIDKQDHAANPCGQIGETVDLDEAV
QRALEFAKKEGNTLVIVTADHAHASQIVAPDTKAPGLTQALNTKDGAVMVMSYGNSEEDSQEHT
GSQRLIAAYGPHAANVVGLTDQTDLFYTMKAALGLKTDSLAVVLQRRDWENPGVTQLNRLAAHP
PFASWRNSEEARTDRPSQQLRSLNGEWR-

500-mer (1-500)

MTMITPSLHACRSTLEMKIVNNWLKQLIKVTLVFLFVAAIFYLITPVHVMSKHTDFSSEIIGYK
AIDGGVTRDIASTDTCFANKRADFDTWFSQRGGSYTNDKACPLIAAVITREVGFFVPGLPGTIL
RTTNGDFLHFLPRVFSAVGNICYTPSKLIEYTDFAVSACVLAEECTIFKDASGKPVPCYDTNV
LEGSVAYESLRPDTRYVLMDSIIQFPNTYLEGSRVVTTFDSEYCRHGTCESEAGVCVSTSG
RWVLNNDYYRNLPGVFCGVDAVNLLTNMFTPLIQPIGALDISASIVAGGIVAIIVVTCLAHYFMR
FRRAFGEYSHVAFNTLLFLMSFTVLCLTPVYSFLPGVYSVIYLYLTFYLTNDVSFLAHIQWMV
MFTPLVPFWITIAIYIICISTKHFYWFFSNYLKRRVVFNGVSFSTFEEAALCTFLLNKEMYLKL
SDVLLPLTQYNRYLALYNKYKYFSGAMDTTSYREAACCHLAKALNDFSNSGSDVLYQPPQTSIT
SAVLQPSARTPEMPVLENRAAQGDITAPGGARRLTGDQTAALRDSLSDKPAKNIILLIGDGMG
DSEITAARNYAEGAGGFFKGIDALPLTGQYTHYALNKKTGKPDYVTDASAATAWSTGVKTYNG
ALGVDIHEKDHPTILEMAKAAGLATGNVSTAELQDATPAALVAHVTSRKCYGPSATSEKCPGNA
LEKGGKGSITEQLLNARADVTLGGGAKTFAETATAGEWQKTLREQAQARGYQLVSDAASLNSV
TEANQQKPLLGLFADGNMPVRWLGPATYHGNIIDKPAVTCTPNPQRNDSVPTLAQMTDKAIELL
SKNEKGFFLQVEGASIDKQDHAANPCGQIGETVDLDEAVQRALEFAKKEGNTLVIVTADHAHAS
QIVAPDTKAPGLTQALNTKDGAVMVMSYGNSEEDSQEHTGSQRLIAAYGPHAANVVGLTDQTDL
FYTMKAALGLKTDSLAVVLQRRDWENPGVTQLNRLAAHPPFASWRNSEEARTDRPSQQLRSLN
EWR-

C. Proteínas clonadas en el vector pCAGGS: Metionina añadida (en fucsia), dianas naturales de glicosilación (en rojo), dianas no glicosilables (en gris), primer y último aminoácido de la proteína truncada clonada (en oliva), etiqueta c-myc (en verde) y las mutaciones D75R, N216Q, N210Q-N216Q, NIS, QIC, AAAA y FFFF (en turquesa).

Polipéptidos truncados del DI de NALCN clonados en el vector pCAGGS

65-mer (2-65) NN

MGSGNSTLKRKQSSRV EAQPVTDFGP DESLSDNADI LWINKPWVHS LLRICAIIISV
ISVRMNTPTM FEHYP EQKLISEEDLNSTMG-

65-mer (2-65) NQ

M G S G **N** S T L K R K Q S S R V E A Q P V T D F G P D E S L S D N A D I L W I N K P W V H S L L R I C A I I S V
I S V R M N T P M T F E H Y **P** **E Q K L I S E E D L** Q S T M G M —

65-mer (2-65) QN

M G S G Q S T L K R K Q S S R V E A Q P V T D F G P D E S L S D N A D I L W I N K P W V H S L L R I C A I I S V
I S V R M N T P M T F E H Y **P** **E Q K L I S E E D L** **N** S T M G M —

65-mer (2-65) QQ

M G S G Q S T L K R K Q S S R V E A Q P V T D F G P D E S L S D N A D I L W I N K P W V H S L L R I C A I I S V
I S V R M N T P M T F E H Y **P** **E Q K L I S E E D L** Q S T M G M —

97-mer (2-97) NN

M G S G **N** S T L K R K Q S S R V E A Q P V T D F G P D E S L S D N A D I L W I N K P W V H S L L R I C A I I S V
I S V R M N T P M T F E H Y P P L Q Y V T F T L D T L L M F L Y T A E M I A K M H I R G I V **K**
E Q K L I S E E D L **N** S T M G M —

97-mer (2-97) NQ

M G S G **N** S T L K R K Q S S R V E A Q P V T D F G P D E S L S D N A D I L W I N K P W V H S L L R I C A I I S V
I S V R M N T P M T F E H Y P P L Q Y V T F T L D T L L M F L Y T A E M I A K M H I R G I V **K**
E Q K L I S E E D L Q S T M G M —

97-mer (2-97) QN

M G S G Q S T L K R K Q S S R V E A Q P V T D F G P D E S L S D N A D I L W I N K P W V H S L L R I C A I I S V
I S V R M N T P M T F E H Y P P L Q Y V T F T L D T L L M F L Y T A E M I A K M H I R G I V **K**
E Q K L I S E E D L **N** S T M G M —

97-mer (2-97) QQ

M G S G Q S T L K R K Q S S R V E A Q P V T D F G P D E S L S D N A D I L W I N K P W V H S L L R I C A I I S V
I S V R M N T P M T F E H Y P P L Q Y V T F T L D T L L M F L Y T A E M I A K M H I R G I V **K**
E Q K L I S E E D L Q S T M G M —

136-mer (2-136) NN

M G S G **N** S T L K R K Q S S R V E A Q P V T D F G P D E S L S D N A D I L W I N K P W V H S L L R I C A I I S V
I S V R M N T P M T F E H Y P P L Q Y V T F T L D T L L M F L Y T A E M I A K M H I R G I V K G D S S Y V K D R W C V F
D G F M V F C L W V S L V L Q V F E I A D I V D Q **M** **E Q K L I S E E D L** **N** S T M G M —

136-mer (2-136) NQ

M G S G **N** S T L K R K Q S S R V E A Q P V T D F G P D E S L S D N A D I L W I N K P W V H S L L R I C A I I S V
I S V R M N T P M T F E H Y P P L Q Y V T F T L D T L L M F L Y T A E M I A K M H I R G I V K G D S S Y V K D R W C V F
D G F M V F C L W V S L V L Q V F E I A D I V D Q **M** **E Q K L I S E E D L** Q S T M G M —

136-mer (2-136) QN

M G S G Q S T L K R K Q S S R V E A Q P V T D F G P D E S L S D N A D I L W I N K P W V H S L L R I C A I I S V
I S V R M N T P M T F E H Y P P L Q Y V T F T L D T L L M F L Y T A E M I A K M H I R G I V K G D S S Y V K D R W C V F
D G F M V F C L W V S L V L Q V F E I A D I V D Q **M** **E Q K L I S E E D L** **N** S T M G M —

136-mer (2-136) QQ

M G S G Q S T L K R K Q S S R V E A Q P V T D F G P D E S L S D N A D I L W I N K P W V H S L L R I C A I I S V
I S V R M N T P M T F E H Y P P L Q Y V T F T L D T L L M F L Y T A E M I A K M H I R G I V K G D S S Y V K D R W C V F
D G F M V F C L W V S L V L Q V F E I A D I V D Q M E Q K L I S E E D I Q S T M G M --

163-mer (2-163) NN

M G S G N S T L K R K Q S S R V E A Q P V T D F G P D E S L S D N A D I L W I N K P W V H S L L R I C A I I S V
I S V R M N T P M T F E H Y P P L Q Y V T F T L D T L L M F L Y T A E M I A K M H I R G I V K G D S S Y V K D R W C V F
D G F M V F C L W V S L V L Q V F E I A D I V D Q M S P W G M L R I P R P L I M I R A F R I Y F R F E L P
E Q K L I S E E D I N S T M G M --

163-mer (2-163) NQ

M G S G N S T L K R K Q S S R V E A Q P V T D F G P D E S L S D N A D I L W I N K P W V H S L L R I C A I I S V
I S V R M N T P M T F E H Y P P L Q Y V T F T L D T L L M F L Y T A E M I A K M H I R G I V K G D S S Y V K D R W C V F
D G F M V F C L W V S L V L Q V F E I A D I V D Q M S P W G M L R I P R P L I M I R A F R I Y F R F E L P
E Q K L I S E E D I Q S T M G M --

163-mer (2-163) QN

M G S G Q S T L K R K Q S S R V E A Q P V T D F G P D E S L S D N A D I L W I N K P W V H S L L R I C A I I S V
I S V R M N T P M T F E H Y P P L Q Y V T F T L D T L L M F L Y T A E M I A K M H I R G I V K G D S S Y V K D R W C V F
D G F M V F C L W V S L V L Q V F E I A D I V D Q M S P W G M L R I P R P L I M I R A F R I Y F R F E L P
E Q K L I S E E D I N S T M G M --

163-mer (2-163) QQ

M G S G Q S T L K R K Q S S R V E A Q P V T D F G P D E S L S D N A D I L W I N K P W V H S L L R I C A I I S V
I S V R M N T P M T F E H Y P P L Q Y V T F T L D T L L M F L Y T A E M I A K M H I R G I V K G D S S Y V K D R W C V F
D G F M V F C L W V S L V L Q V F E I A D I V D Q M S P W G M L R I P R P L I M I R A F R I Y F R F E L P
E Q K L I S E E D I Q S T M G M --

233-mer (2-233) NN

M G S G N S T L K R K Q S S R V E A Q P V T D F G P D E S L S D N A D I L W I N K P W V H S L L R I C A I I S V
I S V R M N T P M T F E H Y P P L Q Y V T F T L D T L L M F L Y T A E M I A K M H I R G I V K G D S S Y V K D R W C V F
D G F M V F C L W V S L V L Q V F E I A D I V D Q M S P W G M L R I P R P L I M I R A F R I Y F R F E L P R T R I T N I
L K R S G E Q I W S V S I F L L F F L L Y G I L G V Q M F G T F T Y H C V V N D T K P G N V T W N S L A I P D T H C S
P E L E Q K L I S E E D I N S T M G M --

233-mer (2-233) NQ

M G S G N S T L K R K Q S S R V E A Q P V T D F G P D E S L S D N A D I L W I N K P W V H S L L R I C A I I S V
I S V R M N T P M T F E H Y P P L Q Y V T F T L D T L L M F L Y T A E M I A K M H I R G I V K G D S S Y V K D R W C V F
D G F M V F C L W V S L V L Q V F E I A D I V D Q M S P W G M L R I P R P L I M I R A F R I Y F R F E L P R T R I T N I
L K R S G E Q I W S V S I F L L F F L L Y G I L G V Q M F G T F T Y H C V V N D T K P G N V T W N S L A I P D T H C S
P E L E Q K L I S E E D I Q S T M G M --

233-mer (2-233) QN

M G S G Q S T L K R K Q S S R V E A Q P V T D F G P D E S L S D N A D I L W I N K P W V H S L L R I C A I I S V
I S V R M N T P M T F E H Y P P L Q Y V T F T L D T L L M F L Y T A E M I A K M H I R G I V K G D S S Y V K D R W C V F
D G F M V F C L W V S L V L Q V F E I A D I V D Q M S P W G M L R I P R P L I M I R A F R I Y F R F E L P R T R I T N I
L K R S G E Q I W S V S I F L L F F L L Y G I L G V Q M F G T F T Y H C V V N D T K P G N V T W N S L A I P D T H C S
P E L E Q K L I S E E D I N S T M G M --

233-mer (2-233) QQ

MSGGQSTLKRKQSSRV EAQPVTDFGP DESLSDNADI LWINKPWVHS LLRICAIIISV
ISVRMNTPTMT FEHYPLQYV TFTLDTLIMF LYTAEMIAKM HIRGIVKGDS SYVKDRWCVF
DGFVMVFLWV SLVLQVFEIA DIVDQMSPWG MLRIPRPLIM IRAFRIYFRF ELPRTITNI
LKRSGEQIWS VSIFLLFFLL YGILGVQMF GTFTYHCVVN DTKPGNVTWN SLAIPDTHCS
PEL EQKLISEEDLQSTMGM -

Dominio DI completo 380-mer (2-380) NN

MSGGNSTLKRKQSSRV EAQPVTDFGP DESLSDNADI LWINKPWVHS LLRICAIIISV
ISVRMNTPTMT FEHYPLQYV TFTLDTLIMF LYTAEMIAKM HIRGIVKGDS SYVKDRWCVF
DGFVMVFLWV SLVLQVFEIA DIVDQMSPWG MLRIPRPLIM IRAFRIYFRF ELPRTITNI
LKRSGEQIWS VSIFLLFFLL YGILGVQMF GTFTYHCVVN DTKPGNVTWN SLAIPDTHCS
PELEEGYQCP PGFKCMDLED LGLSRQELGY SGFNEIGTSI FTVYEAASQE GWVFLMYRAI
DSFPRWRSYF YFITLIFFLA WLKKNVFIIV IETFAEIRV QFQQMWGSRS STTSTATTQM
FHEDAAGGWQ LVAVDVNKPQ GRAPACLQKMEQKLISEEDLNSTMGM--

DI 380-mer (2-380) NQ

MSGGNSTLKRKQSSRV EAQPVTDFGP DESLSDNADI LWINKPWVHS LLRICAIIISV
ISVRMNTPTMT FEHYPLQYV TFTLDTLIMF LYTAEMIAKM HIRGIVKGDS SYVKDRWCVF
DGFVMVFLWV SLVLQVFEIA DIVDQMSPWG MLRIPRPLIM IRAFRIYFRF ELPRTITNI
LKRSGEQIWS VSIFLLFFLL YGILGVQMF GTFTYHCVVN DTKPGNVTWN SLAIPDTHCS
PELEEGYQCP PGFKCMDLED LGLSRQELGY SGFNEIGTSI FTVYEAASQE GWVFLMYRAI
DSFPRWRSYF YFITLIFFLA WLKKNVFIIV IETFAEIRV QFQQMWGSRS STTSTATTQM
FHEDAAGGWQ LVAVDVNKPQ GRAPACLQKMEQKLISEEDLQSTMGM--

DI 380-mer (2-380) QN

MSGGQSTLKRKQSSRV EAQPVTDFGP DESLSDNADI LWINKPWVHS LLRICAIIISV
ISVRMNTPTMT FEHYPLQYV TFTLDTLIMF LYTAEMIAKM HIRGIVKGDS SYVKDRWCVF
DGFVMVFLWV SLVLQVFEIA DIVDQMSPWG MLRIPRPLIM IRAFRIYFRF ELPRTITNI
LKRSGEQIWS VSIFLLFFLL YGILGVQMF GTFTYHCVVN DTKPGNVTWN SLAIPDTHCS
PELEEGYQCP PGFKCMDLED LGLSRQELGY SGFNEIGTSI FTVYEAASQE GWVFLMYRAI
DSFPRWRSYF YFITLIFFLA WLKKNVFIIV IETFAEIRV QFQQMWGSRS STTSTATTQM
FHEDAAGGWQ LVAVDVNKPQ GRAPACLQKMEQKLISEEDLNSTMGM--

DI 380-mer (2-380) QQ

MSGGQSTLKRKQSSRV EAQPVTDFGP DESLSDNADI LWINKPWVHS LLRICAIIISV
ISVRMNTPTMT FEHYPLQYV TFTLDTLIMF LYTAEMIAKM HIRGIVKGDS SYVKDRWCVF
DGFVMVFLWV SLVLQVFEIA DIVDQMSPWG MLRIPRPLIM IRAFRIYFRF ELPRTITNI
LKRSGEQIWS VSIFLLFFLL YGILGVQMF GTFTYHCVVN DTKPGNVTWN SLAIPDTHCS
PELEEGYQCP PGFKCMDLED LGLSRQELGY SGFNEIGTSI FTVYEAASQE GWVFLMYRAI
DSFPRWRSYF YFITLIFFLA WLKKNVFIIV IETFAEIRV QFQQMWGSRS STTSTATTQM
FHEDAAGGWQ LVAVDVNKPQ GRAPACLQKMEQKLISEEDLQSTMGM--

Polipéptido truncado 163-mer modificado (D75R)

163-mer (2-163) NN

M G S G **N** S T I L K R K Q S S R V E A Q P V T D F G P D E S L S D N A D I L W I N K P W V H S L L R I C A I I S V
I S V R M N T P M T F E H Y P P L Q Y V T F T L **R** T L L M F L Y T A E M I A K M H I R G I V K G D S S Y V K D R W C V F
D G F M V F C L W V S L V L Q V F E I A D I V D Q M S P W G M L R I P R P L I M I R A F R I Y F R F E L P
E Q K L I S E E D L **N** S T M G M —

163-mer (2-163) QN

M G S G **Q** S T I L K R K Q S S R V E A Q P V T D F G P D E S L S D N A D I L W I N K P W V H S L L R I C A I I S V
I S V R M N T P M T F E H Y P P L Q Y V T F T L **R** T L L M F L Y T A E M I A K M H I R G I V K G D S S Y V K D R W C V F
D G F M V F C L W V S L V L Q V F E I A D I V D Q M S P W G M L R I P R P L I M I R A F R I Y F R F E L P
E Q K L I S E E D L **N** S T M G M —

DI 380-mer (2-380) modificada (N216Q)

M G S G **Q** S T I L K R K Q S S R V E A Q P V T D F G P D E S L S D N A D I L W I N K P W V H S L L R I C A I I S V
I S V R M N T P M T F E H Y P P L Q Y V T F T L D T L L M F L Y T A E M I A K M H I R G I V K G D S S Y V K D R W C V F
D G F M V F C L W V S L V L Q V F E I A D I V D Q M S P W G M L R I P R P L I M I R A F R I Y F R F E L P R T R I T N I
L K R S G E Q I W S V S I F L L F F L L L Y G I L G V Q M F G T F T Y H C V V N D T K P G **Q** V T W N S L A I P D T H C S
P E L E E G Y Q C P P G F K C M D L E D L G L S R Q E L G Y S G F N E I G T S I F T V Y E A A S Q E G W V F L M Y R A I
D S F P R W R S Y F Y F I T L I F F L A W L V K N V F I A V I I E T F A E I R V Q F Q Q M W G S R S S T T S T A T T Q M
F H E D A A G G W Q L V A V D V N K P Q G R A P A C L Q K **E Q K L I S E E D L** **Q** S T M G M —

DI 380-mer modificada (N210Q, N216Q)

M G S G **Q** S T I L K R K Q S S R V E A Q P V T D F G P D E S L S D N A D I L W I N K P W V H S L L R I C A I I S V
I S V R M N T P M T F E H Y P P L Q Y V T F T L D T L L M F L Y T A E M I A K M H I R G I V K G D S S Y V K D R W C V F
D G F M V F C L W V S L V L Q V F E I A D I V D Q M S P W G M L R I P R P L I M I R A F R I Y F R F E L P R T R I T N I
L K R S G E Q I W S V S I F L L F F L L L Y G I L G V Q M F G T F T Y H C V V **Q** D T K P G **Q** V T W N S L A I P D T H C S
P E L E E G Y Q C P P G F K C M D L E D L G L S R Q E L G Y S G F N E I G T S I F T V Y E A A S Q E G W V F L M Y R A I
D S F P R W R S Y F Y F I T L I F F L A W L V K N V F I A V I I E T F A E I R V Q F Q Q M W G S R S S T T S T A T T Q M
F H E D A A G G W Q L V A V D V N K P Q G R A P A C L Q K **E Q K L I S E E D L** **Q** S T M G M —

Polipéptidos truncados de NSP4 clonadas en el vector pCAGGS

279-mer (1-279) NST

M G S G **K** I V N N W L K Q L I K V T L V F L F V A A I F Y L I T P V H V M S K H T D F S S E I I G Y K A I D G G V T R D I A S T
D T C F A N K H A D F D T W F S Q R G G S Y T N D K A C P L I A A V I T R E V G F V V P G L P G T I L R T T N G D F L H F L P R
V F S A V G N I C Y T P S K L I E Y T D F A T S A C V L A A E C T I F K D A S G K P V P Y C Y D T N V L E G S V A Y E S L R P D
T R Y V L M D G S I I Q F P N T Y L E G S V R V I T F D S E Y C R H G T C E R S E A G V C V S T S G R W V L N N D Y Y R S L P
G V F C G V D A V N L L T N M F T P L I Q P I G A L **D** P S S A R T **E Q K L I S E E D L** **N** S T M G M —

279-mer (1-279) QST

M G S G **K** I V N N W L K Q L I K V T L V F L F V A A I F Y L I T P V H V M S K H T D F S S E I I G Y K A I D G G V T R D I A S T
D T C F A N K H A D F D T W F S Q R G G S Y T N D K A C P L I A A V I T R E V G F V V P G L P G T I L R T T N G D F L H F L P R
V F S A V G N I C Y T P S K L I E Y T D F A T S A C V L A A E C T I F K D A S G K P V P Y C Y D T N V L E G S V A Y E S L R P D
T R Y V L M D G S I I Q F P N T Y L E G S V R V I T F D S E Y C R H G T C E R S E A G V C V S T S G R W V L N N D Y Y R S L P
G V F C G V D A V N L L T N M F T P L I Q P I G A L **D** P S S A R T **E Q K L I S E E D L** **Q** S T M G M —

313-mer (1-313) NST

MSGKIVNNWLKQLIKVTLVFLFVAAlFYLI TPVHVMSKHTDFSSEIIGYKAIDGGVTRDIAS
DTCFANKHADFDTWFSQRGGSYTNDKACPLIAAVITREVG FVVPGLPGTILRTTNGDFLHFLPR
VFSAVGNICYTPSKLIEYTD FATSACVLAAECTIFKDASGKPV PYPICYDTNVLEG SVAYESLRPD
TRYVLMDGSI IQFPNTYLEG SVRVVTTFDSEYCRHGTCERSEAGVCVSTSGRWVLNNDY YRSLP
GVFCGVDAVNLLTNMFTPLIQPIGALDISASIVAGGIVAIVVTCLAYYFMRFRRAFGEYSHPSS
ARTEQKLISEEDLNSTMGM--

313-mer (1-313) QST

MSGKIVNNWLKQLIKVTLVFLFVAAlFYLI TPVHVMSKHTDFSSEIIGYKAIDGGVTRDIAS
DTCFANKHADFDTWFSQRGGSYTNDKACPLIAAVITREVG FVVPGLPGTILRTTNGDFLHFLPR
VFSAVGNICYTPSKLIEYTD FATSACVLAAECTIFKDASGKPV PYPICYDTNVLEG SVAYESLRPD
TRYVLMDGSI IQFPNTYLEG SVRVVTTFDSEYCRHGTCERSEAGVCVSTSGRWVLNNDY YRSLP
GVFCGVDAVNLLTNMFTPLIQPIGALDISASIVAGGIVAIVVTCLAYYFMRFRRAFGEYSHPSS
ARTEQKLISEEDLQSTMGM--

335-mer (1-335) NST

MSGKIVNNWLKQLIKVTLVFLFVAAlFYLI TPVHVMSKHTDFSSEIIGYKAIDGGVTRDIAS
DTCFANKHADFDTWFSQRGGSYTNDKACPLIAAVITREVG FVVPGLPGTILRTTNGDFLHFLPR
VFSAVGNICYTPSKLIEYTD FATSACVLAAECTIFKDAGGKPV PYPICYDTNVLEG SVAYESLRPD
TRYVLMDGSI IQFPNTYLEG SVRVVTTFDSEYCRHGTCERSEAGVCVSTSGRWVLNNDY YRSLP
GVFCGVDAVNLLTNMFTPLIQPIGALDISASIVAGGIVAIVVTCLAYYFMRFRRAFGEYSHVVA
FNTLLFLMSFTVLCLTPVYPSSARTEQKLISEEDLNSTMGM-

335-mer (1-335) QST

MSGKIVNNWLKQLIKVTLVFLFVAAlFYLI TPVHVMSKHTDFSSEIIGYKAIDGGVTRDIAS
DTCFANKHADFDTWFSQRGGSYTNDKACPLIAAVITREVG FVVPGLPGTILRTTNGDFLHFLPR
VFSAVGNICYTPSKLIEYTD FATSACVLAAECTIFKDAGGKPV PYPICYDTNVLEG SVAYESLRPD
TRYVLMDGSI IQFPNTYLEG SVRVVTTFDSEYCRHGTCERSEAGVCVSTSGRWVLNNDY YRSLP
GVFCGVDAVNLLTNMFTPLIQPIGALDISASIVAGGIVAIVVTCLAYYFMRFRRAFGEYSHVVA
FNTLLFLMSFTVLCLTPVYPSSARTEQKLISEEDLQSTMGM-

346-mer (1-346) NST

MSGKIVNNWLKQLIKVTLVFLFVAAlFYLI TPVHVMSKHTDFSSEIIGYRAIDGGVTRDIAS
DTCFANKHADFDTWFSQRGGSYTNGKACPLIAAVITREVG FVVPGLPGTILRTTNGDFLHFLPR
VFSAVGNICYTPSKLIEYTD FATSACVLAAECTIFKDASGKPV PYPICYDTNVLEG SVAYESLRPD
TRYVLMDGSI IQFPNTYLEG SVRVVTTFDSEYCRHGTCERSEAGVCVSTSGRWVLNNDY YRSLP
GVFCGVDAVNLLTNMFTPLIQPIGALDISASIVAGGIVAIVVTCLAYYFMRFRRAFGEYSHVVA
FNTLLFLMSFTVLCLTPVYSFLPGVYSVIYPSSARTEQKLISEEDLNSTMGM-

346-mer (1-346) QST

MSGKIVNNWLKQLIKVTLVFLFVAAlFYLI TPVHVMSKHTDFSSEIIGYRAIDGGVTRDIAS
DTCFANKHADFDTWFSQRGGSYTNGKACPLIAAVITREVG FVVPGLPGTILRTTNGDFLHFLPR
VFSAVGNICYTPSKLIEYTD FATSACVLAAECTIFKDASGKPV PYPICYDTNVLEG SVAYESLRPD
TRYVLMDGSI IQFPNTYLEG SVRVVTTFDSEYCRHGTCERSEAGVCVSTSGRWVLNNDY YRSLP
GVFCGVDAVNLLTNMFTPLIQPIGALDISASIVAGGIVAIVVTCLAYYFMRFRRAFGEYSHVVA
FNTLLFLMSFTVLCLTPVYSFLPGVYSVIYPSSARTEQKLISEEDLQSTMGM-

362–mer (1-362) NST

MSGKIVNNWLKQLIKVTLVFLFVAAlFYLI TPVHVMSKHTDFSSEIIGYKAIDGGVTRDIAS
DTCFANKRADFDTWFSQRGGSYTNDKACPLIAAVITREVG FVVPGLPGTILRTTNGDFLHFLPR
VFSAVGNICYTPSKLIEYTD FATSACVLAAECTIFKDASGKPVPCYDTNVLEG SVAYESLRPD
TRYVLMDGSI IQFPNTYLEG SVRVVTTFDSEYCRHGTCERSEAGVCVSTSGRWVLNNDYYRNLP
GVFCGVDAVNLLTNMFTPLIQPIGALDISASIVAGGIVAI VVTCLAHYFMRFRRAFGEYSHVVA
FNTLLFLMSFTVLCLTPVYSFLPGVYSVIYLYLTFYLTNDVSFLAHIPSSART **EQKLISEEDLNS**
TMGM–

362–mer (1-362) QST

MSGKIVNNWLKQLIKVTLVFLFVAAlFYLI TPVHVMSKHTDFSSEIIGYKAIDGGVTRDIAS
DTCFANKRADFDTWFSQRGGSYTNDKACPLIAAVITREVG FVVPGLPGTILRTTNGDFLHFLPR
VFSAVGNICYTPSKLIEYTD FATSACVLAAECTIFKDASGKPVPCYDTNVLEG SVAYESLRPD
TRYVLMDGSI IQFPNTYLEG SVRVVTTFDSEYCRHGTCERSEAGVCVSTSGRWVLNNDYYRNLP
GVFCGVDAVNLLTNMFTPLIQPIGALDISASIVAGGIVAI VVTCLAHYFMRFRRAFGEYSHVVA
FNTLLFLMSFTVLCLTPVYSFLPGVYSVIYLYLTFYLTNDVSFLAHIPSSART **EQKLISEEDLQS**
TMGM–

371–mer (1-371) NST

MSGKIVNNWLKQLIKVTLVFLFVAAlFYLI TPVHVMSKHTDFSSEIIGYKAIDGGVTRDIAS
DTCFANKRADFDTWFSQRGGSYTNDKACPLIAAVITREVG FVVPGLPGTILRTTNGDFLHFLPR
VFSAVGNICYTPSKLIEYTD FATSACVLAAECTIFKDASGKPVPCYDTNVLEG SVAYESLRPD
TRYVLMDGSI IQFPNTYLEG SVRVVTTFDSEYCRHGTCERSEAGVCVSTSGRWVLNNDYYRNLP
GVFCGVDAVNLLTNMFTPLIQPIGALDISASIVAGGIVAI VVTCLAHYFMRFRRAFGEYSHVVA
FNTLLFLMSFTVLCLTPVYSFLPGVYSVIYLYLTFYLTNDVSFLAHIQWMVMFTIPSSART **EQK**
LISEEDLNSTTMGM–

371–mer (1-371) QST

MSGKIVNNWLKQLIKVTLVFLFVAAlFYLI TPVHVMSKHTDFSSEIIGYKAIDGGVTRDIAS
DTCFANKRADFDTWFSQRGGSYTNDKACPLIAAVITREVG FVVPGLPGTILRTTNGDFLHFLPR
VFSAVGNICYTPSKLIEYTD FATSACVLAAECTIFKDASGKPVPCYDTNVLEG SVAYESLRPD
TRYVLMDGSI IQFPNTYLEG SVRVVTTFDSEYCRHGTCERSEAGVCVSTSGRWVLNNDYYRNLP
GVFCGVDAVNLLTNMFTPLIQPIGALDISASIVAGGIVAI VVTCLAHYFMRFRRAFGEYSHVVA
FNTLLFLMSFTVLCLTPVYSFLPGVYSVIYLYLTFYLTNDVSFLAHIQWMVMFTIPSSART **EQK**
LISEEDLQSTTMGM–

397–mer (1-397) NST

MSGKIVNNWLKQLIKVTLVFLFVAAlFYLI TPVHVMSKHTDFSSEIIGYKAIDGGVTRDIAS
DTCFANKRADFDTWFSQRGGSYTNDKACPLIAAVITREVG FVVPGLPGTILRTTNGDFLHFLPR
VFSAVGNICYTPSKLIEYTD FATSACVLAAECTIFKDASGKPVPCYDTNVLEG SVAYESLRPD
TRYVLMDGSI IQFPNTYLEG SVRVVTTFDSEYCRHGTCERSEAGVCVSTSGRWVLNNDYYRNLP
GVFCGVDAVNLLTNMFTPLIQPIGALDISASIVAGGIVAI VVTCLAHYFMRFRRAFGEYSHVVA
FNTLLFLMSFTVLCLTPVYSFLPGVYSVIYLYLTFYLTNDVSFLAHIQWMVMFTPLVPFWITIA
YIICISTKHfYWFFSNIPSSART **EQKLISEEDLNST**TMGM–

397-mer (1-397) QST

MSGKIVNNWLKQLIKVTLVFLFVAAlFYLI TPVHVMSKHTDFSSEIIGYKAIDGGVTRDIAS
DTCFANKRADFDTWFSQRGGSYTNDKACPLIAAVITREVG FVVPGLPGTILRTTNGDFLHFLPR
VFSAVGNICYTPSKLIEYTD FATSACVLAAECTIFKDASGKPVPCYDTNVLEGSVAYESLRPD
TRYVLMDGSI IQFPNTYLEG SVRVVTTFDSEYCRHGTCERSEAGVCVSTSGRWVLNNDYYRNL
GVFCGVDAVNLLTNMFTPLIQPIGALDISASIVAGGIVAI VVTCLAHYFMRFRRAFGEYSHVVA
FNTLLFLMSFTVLCLTPVYSFLPGVYSVIYLYLT FYLTNDVSFLAHIQWMVMFTPLVPFWITIA
YIICISTKH FYWFFSNYPSSARTEQKLISEEDLQSTMGM-

500-mer (1-500) NST

MSGKIVNNWLKQLIKVTLVFLFVAAlFYLI TPVHVMSKHTDFSSEIIGYKAIDGGVTRDIAS
DTCFANKRADFDTWFSQRGGSYTNDKACPLIAAVITREVG FVVPGLPGTILRTTNGDFLHFLPR
VFSAVGNICYTPSKLIEYTD FATSACVLAAECTIFKDASGKPVPCYDTNVLEGSVAYESLRPD
TRYVLMDGSI IQFPNTYLEG SVRVVTTFDSEYCRHGTCERSEAGVCVSTSGRWVLNNDYYRNL
GVFCGVDAVNLLTNMFTPLIQPIGALDISASIVAGGIVAI VVTCLAHYFMRFRRAFGEYSHVVA
FNTLLFLMSFTVLCLTPVYSFLPGVYSVIYLYLT FYLTNDVSFLAHIQWMVMFTPLVPFWITIA
YIICISTKH FYWFFSNYLKRRVVFNGVSFSTFEEAALCTFLLNKEMYLKLRSDVLLPLTQYNRY
LALYNKYKYFSGAMDTTSYREAA CCHLAKALNDFSNSGSDVLYQPPQTSITS AVLQPSSARTEQ
KLISEEDLNSTMGM-

500-mer (1-500) QST

MSGKIVNNWLKQLIKVTLVFLFVAAlFYLI TPVHVMSKHTDFSSEIIGYKAIDGGVTRDIAS
DTCFANKRADFDTWFSQRGGSYTNDKACPLIAAVITREVG FVVPGLPGTILRTTNGDFLHFLPR
VFSAVGNICYTPSKLIEYTD FATSACVLAAECTIFKDASGKPVPCYDTNVLEGSVAYESLRPD
TRYVLMDGSI IQFPNTYLEG SVRVVTTFDSEYCRHGTCERSEAGVCVSTSGRWVLNNDYYRNL
GVFCGVDAVNLLTNMFTPLIQPIGALDISASIVAGGIVAI VVTCLAHYFMRFRRAFGEYSHVVA
FNTLLFLMSFTVLCLTPVYSFLPGVYSVIYLYLT FYLTNDVSFLAHIQWMVMFTPLVPFWITIA
YIICISTKH FYWFFSNYLKRRVVFNGVSFSTFEEAALCTFLLNKEMYLKLRSDVLLPLTQYNRY
LALYNKYKYFSGAMDTTSYREAA CCHLAKALNDFSNSGSDVLYQPPQTSITS AVLQPSSARTEQ
KLISEEDLQSTMGM-

500-mer (1-500) NST - NST

MSGKNSTKIVNNWLKQLIKVTLVFLFVAAlFYLI TPVHVMSKHTDFSSEIIGYKAIDGGVTRDI
ASTDTCFANKRADFDTWFSQRGGSYTNDKACPLIAAVITREVG FVVPGLPGTILRTTNGDFLHFL
LPRVFSAVGNISYTPSKLIEYTD FATSACVLAAECTIFKDASGKPVPCYDTNVLEGSVAYESL
RPDTRYVLMDGSI IQFPNTYLEG SVRVVTTFDSEYCRHGTCERSEAGVCVSTSGRWVLNNDYYR
NLPGVFCGVDAVNLLTNMFTPLIQPIGALDISASIVAGGIVAI VVTCLAHYFMRFRRAFGEYSH
VVAFNTLLFLMSFTVLCLTPVYSFLPGVYSVIYLYLT FYLTNDVSFLAHIQWMVMFTPLVPFWI
TIAYIICISTKH FYWFFSNYLKRRVVFNGVSFSTFEEAALCTFLLNKEMYLKLRSDVLLPLTQY
NRYLALYNKYKYFSGAMDTTSYREAA CCHLAKALNDFSNSGSDVLYQPPQTSITS AVLQPSSAR
TEQKLISEEDLNSTMGM-

Proteína NSP4 con su diana de glicosilación no canónica modificada (NIC→NIS)

MGSGNSTKIVNNWLKQLIKVTLVFLFVAAIFYLITPVHVMKHTDFSSEIIGYKAIDGGVTRDI
ASTDTCFANKRADFDTWFSQRGGSYTNDKACPLIAAVITREVG FVVPGLPGTILRTTNGDFLHF
LPRVFSAVGNIYTPSKLIEYTD FATSACVLAAECTIFKDASGKVPVPCYDTNVLEGSVAYESL
RPDTRYVLM DGSIIQFPNTYLEG SVRVVTTFDSEYCRHGT CERSEAGVCVSTSGRWVLNNDYYR
NLPGVFCGVDAVNLLTNMFTPLIQPIGALDISASIVAGGIVAIVVTC LAHYFMRFRRAFGEYSH
VVAFN TLLFLMSFTVLCLTPVYSFLPGVYSVIYLYLTFYLTNDVSFLAHIQWMVMFTPLVPFWI
TIAYIIICISTKH FYWFFSNYLKRRVVFNGVSFSTFEEAALCTFLLNKEMYLKLRSDVLLPLTQY
NRYLALYNKYKYFSGAMD TTSYREAA CCHLAKALNDFSNSGSDVLYQPPQTSITS AVLQPSSAR
TEQKLI SEEDINSTMGM—

Proteína NSP4 con su diana de glicosilación no canónica modificada (NIC→QIC)

MGSGNSTKIVNNWLKQLIKVTLVFLFVAAIFYLITPVHVMKHTDFSSEIIGYKAIDGGVTRDI
ASTDTCFANKRADFDTWFSQRGGSYTNDKACPLIAAVITREVG FVVPGLPGTILRTTNGDFLHF
LPRVFSAVGQICYTPSKLIEYTD FATSACVLAAECTIFKDASGKVPVPCYDTNVLEGSVAYESL
RPDTRYVLM DGSIIQFPNTYLEG SVRVVTTFDSEYCRHGT CERSEAGVCVSTSGRWVLNNDYYR
NLPGVFCGVDAVNLLTNMFTPLIQPIGALDISASIVAGGIVAIVVTC LAHYFMRFRRAFGEYSH
VVAFN TLLFLMSFTVLCLTPVYSFLPGVYSVIYLYLTFYLTNDVSFLAHIQWMVMFTPLVPFWI
TIAYIIICISTKH FYWFFSNYLKRRVVFNGVSFSTFEEAALCTFLLNKEMYLKLRSDVLLPLTQY
NRYLALYNKYKYFSGAMD TTSYREAA CCHLAKALNDFSNSGSDVLYQPPQTSITS AVLQPSSAR
TEQKLI SEEDINSTMGM—

Proteína NSP4 con su diana de procesamiento modificada (TPVH→AAAA)

MGSGNSTKIVNNWLKQLIKVTLVFLFVAAIFYLI AAAAVMSKHTDFSSEIIGYKAIDGGVTRDI
ASTDTCFANKRADFDTWFSQRGGSYTNDKACPLIAAVITREVG FVVPGLPGTILRTTNGDFLHF
LPRVFSAVGNICYTPSKLIEYTD FATSACVLAAECTIFKDASGKVPVPCYDTNVLEGSVAYESL
RPDTRYVLM DGSIIQFPNTYLEG SVRVVTTFDSEYCRHGT CERSEAGVCVSTSGRWVLNNDYYR
NLPGVFCGVDAVNLLTNMFTPLIQPIGALDISASIVAGGIVAIVVTC LAHYFMRFRRAFGEYSH
VVAFN TLLFLMSFTVLCLTPVYSFLPGVYSVIYLYLTFYLTNDVSFLAHIQWMVMFTPLVPFWI
TIAYIIICISTKH FYWFFSNYLKRRVVFNGVSFSTFEEAALCTFLLNKEMYLKLRSDVLLPLTQY
NRYLALYNKYKYFSGAMD TTSYREAA CCHLAKALNDFSNSGSDVLYQPPQTSITS AVLQPSSAR
TEQKLI SEEDINSTMGM—

Proteína NSP4 con su diana de procesamiento modificada (TPVH→FFFF)

MGSGNSTKIVNNWLKQLIKVTLVFLFVAAIFYLI FFFFFMSKHTDFSSEIIGYKAIDGGVTRDI
ASTDTCFANKRADFDTWFSQRGGSYTNDKACPLIAAVITREVG FVVPGLPGTILRTTNGDFLHF
LPRVFSAVGNICYTPSKLIEYTD FATSACVLAAECTIFKDASGKVPVPCYDTNVLEGSVAYESL
RPDTRYVLM DGSIIQFPNTYLEG SVRVVTTFDSEYCRHGT CERSEAGVCVSTSGRWVLNNDYYR
NLPGVFCGVDAVNLLTNMFTPLIQPIGALDISASIVAGGIVAIVVTC LAHYFMRFRRAFGEYSH
VVAFN TLLFLMSFTVLCLTPVYSFLPGVYSVIYLYLTFYLTNDVSFLAHIQWMVMFTPLVPFWI
TIAYIIICISTKH FYWFFSNYLKRRVVFNGVSFSTFEEAALCTFLLNKEMYLKLRSDVLLPLTQY
NRYLALYNKYKYFSGAMD TTSYREAA CCHLAKALNDFSNSGSDVLYQPPQTSITS AVLQPSSAR
TEQKLI SEEDINSTMGM—

D. Fragmentos solubles de la proteína NSP4 clonados en vector pETNKI 1.1:

Fragmentos solubles de la proteína NSP4 (en turquesa), etiqueta HisTag (en letras rojas) y diana de escisión de la proteasa PreScission (en subrayado).

Fragmento soluble NSP4 (S40-P270)

FCLTLRRRYTMAHHHHHSAALEVLFQGPGMSSEIIIGYKAIDGGVTRDIASDTCTCFANKHADFD
TWFSQRGGSYTNDKACPLIAAVITREVGFFVPGLPGTILRTTNGDFLHFLPRVFSAVGNICYTP
SKLIEYTDFAVSACVLAAECTIFKDASGKPVPCYDNTNVLEGSVAYESLRPDTRYVLMGDSIIQ
FPNTYLEGSVRVVTTFDSEYCRHGTCESEAGVCVSTSGRWVLNNDYYRSLPGVFCGVDAVNLL
TNMFTP-PGFSSSTTTTTTEIRLLTKPERKLSWLLPPLSNN

Fragmento soluble NSP4 (T411-Q500)

FCLTLRRRYTMAHHHHHSAALEVLFQGPGMTFEEAALCTFLLNKEMYLKLRSDVLLPLTQYNR
YLALYNKYKYFSGAMDTTSYREAAACCHLAKALNDFSNSGSDVLYQPPQTSITS AVLQ-
PGFSSSTTTTTTEIRLLTKPERKLSWLLPPLSNN

E. El S6 del dominio DI de NALCN clonado en sistema Lep con dianas N9, N11, N13 y N15: Segmentos TM H1 y H2 del sistema Lep (en gris), dianas de glicosilación del sistema Lep (en rojo), dianas añadidas en posiciones 9, 11, 13 y 15 de S6 (en fucsia), S6 (en amarillo), tetrapéptido flanqueante de S6 (en turquesa) y mutaciones CLIFAHDD (en verde).

Control de inserción en membrana (#1022) clonado en el sistema Lep

PYVSYTYDLGDTIEYKLGLQVDSRATMANMFALILVIATLVTGILWCVDKFFFAPKRRERQAAA
QAAAGDSLDKATLKKVAPKPGWLETGASVFPVLAIVLIVRSFIYEPPQIPSGSMMP TLNSTDFI
LVEKFAYGIKDPIYQKTLIETGHPKRGDIVVKYPEDPKLDYIKRAVGLPGDKVTYDPVSKELT
IQPGCSSGQACENALPVTYSNVEPSDFVQTFSTRNGGEATSGFFEVPKQETKENGIRLSETS GG
PGKVLLVTGVLGLLLLIKWKGPGGGVPPQQNATWIVPPGQYFMMGDNRDNSADSRWGFVPEAN
LVGRATAIWMSFDKQEGNGRLVCA-VALAASIN

Control de translocación o no inserción en membrana (#753) clonado en el sistema Lep

PYVSYTYDLGDTIEYKLGLQVDSRATMANMFALILVIATLVTGILWCVDKFFFAPKRRERQAAA
QAAAGDSLDKATLKKVAPKPGWLETGASVFPVLAIVLIVRSFIYEPPQIPSGSMMP TLNSTDFI
LVEKFAYGIKDPIYQKTLIETGHPKRGDIVVKYPEDPKLDYIKRAVGLPGDKVTYDPVSKELT
IQPGCSSGQACENALPVTYSNVEPSDFVQTFSTRNGGEATSGFFEVPKQETKENGIRLSETS GG
PGYNLYIYAFFASSLVHGFIPGGGVPPQQNATWIVPPGQYFMMGDNRDNSADSRWGFVPEAN
LVGRATAIWMSFDKQEGNGRLVCA-VALAASIN

S6 clonado en el sistema Lep

PYVSYTYDLGDTIEYKLGLQVDSRATMANMFALILVIATLVTGILWCVDKFFFAPKRRERQAAA
QAAAGDSLDKATLKKVAPKPGWLETGASVFPVLAIVLIVRSFIYEPPQIPSGSMMP TLNSTDFI
LVEKFAYGIKDPIYQKTLIETGHPKRGDIVVKYPEDPKLDYIKRAVGLPGDKVTYDPVSKELT
IQPGCSSGQACENALPVTYSNVEPSDFVQTFSTRNGGEATSGFFEVPKQETKENGIRLSETS GG

PGRWRSYFYFITLIFFLAWLVKNVFIAVII GPGGGVPGQQ NATWIVPPGQYFMMGDNRDNSADS
RYWGFVPEANLVGRATAIWMSFDKQEGNGRLVCA-VALAASIN

S6 clonado en el sistema Lep con sitio de glicosilación a 9 aminoácidos (N9) de su Nt

PYVSYTYDLGDTIEYKLGQLQVDSRATMANMFALILVIATLVTGILWCV DKFFFAPKRRERQAAA
QAAAGDSLDKATLKKVAPKPGWLETGASVFPVLAIVLIVRSFIYEPPQIPSGSMMPTL NSTDFI
LVEKFAYGIKDPIYQKTLIETGHPKRGDIVVFKYPEDPKLDYIKRAVGLPGDKVTYDPVSKELT
IQPGCSSGQACENALPVTYSNVEPSDFVQTFSTRNGGEATSGFFEVPKQETKENGIRNTS TS GG
PGRWRSYFYFITLIFFLAWLVKNVFIAVII GPGGGVPGQQ NATWIVPPGQYFMMGDNRDNSADS
RYWGFVPEANLVGRATAIWMSFDKQEGNGRLVCA-VALAASIN

S6 clonado en el sistema Lep con sitio de glicosilación a 11 aminoácidos (N11) de su Nt

PYVSYTYDLGDTIEYKLGQLQVDSRATMANMFALILVIATLVTGILWCV DKFFFAPKRRERQAAA
QAAAGDSLDKATLKKVAPKPGWLETGASVFPVLAIVLIVRSFIYEPPQIPSGSMMPTL NSTDFI
LVEKFAYGIKDPIYQKTLIETGHPKRGDIVVFKYPEDPKLDYIKRAVGLPGDKVTYDPVSKELT
IQPGCSSGQACENALPVTYSNVEPSDFVQTFSTRNGGEATSGFFEVPKQETKENG NTS SETS GG
PGRWRSYFYFITLIFFLAWLVKNVFIAVII GPGGGVPGQQ NATWIVPPGQYFMMGDNRDNSADS
RYWGFVPEANLVGRATAIWMSFDKQEGNGRLVCA-VALAASIN

S6 clonado en el sistema Lep con sitio de glicosilación a 13 aminoácidos (N13) de su Nt

PYVSYTYDLGDTIEYKLGQLQVDSRATMANMFALILVIATLVTGILWCV DKFFFAPKRRERQAAA
QAAAGDSLDKATLKKVAPKPGWLETGASVFPVLAIVLIVRSFIYEPPQIPSGSMMPTL NSTDFI
LVEKFAYGIKDPIYQKTLIETGHPKRGDIVVFKYPEDPKLDYIKRAVGLPGDKVTYDPVSKELT
IQPGCSSGQACENALPVTYSNVEPSDFVQTFSTRNGGEATSGFFEVPKQETKENTSR LSETS GG
PGRWRSYFYFITLIFFLAWLVKNVFIAVII GPGGGVPGQQ NATWIVPPGQYFMMGDNRDNSADS
RYWGFVPEANLVGRATAIWMSFDKQEGNGRLVCA-VALAASIN

S6 clonado en el sistema Lep con sitio de glicosilación a 15 aminoácidos (N15) de su Nt

PYVSYTYDLGDTIEYKLGQLQVDSRATMANMFALILVIATLVTGILWCV DKFFFAPKRRERQAAA
QAAAGDSLDKATLKKVAPKPGWLETGASVFPVLAIVLIVRSFIYEPPQIPSGSMMPTL NSTDFI
LVEKFAYGIKDPIYQKTLIETGHPKRGDIVVFKYPEDPKLDYIKRAVGLPGDKVTYDPVSKELT
IQPGCSSGQACENALPVTYSNVEPSDFVQTFSTRNGGEATSGFFEVPKQETNTS GIRLSETS GG
PGRWRSYFYFITLIFFLAWLVKNVFIAVII GPGGGVPGQQ NATWIVPPGQYFMMGDNRDNSADS
RYWGFVPEANLVGRATAIWMSFDKQEGNGRLVCA-VALAASIN

S6 del DI de NALCN con mutaciones CLIFAHDD clonado en sistema Lep

S6 con mutación L312I

PYVSYTYDLGDTIEYKLGQLQVDSRATMANMFALILVIATLVTGILWCV DKFFFAPKRRERQAAA
QAAAGDSLDKATLKKVAPKPGWLETGASVFPVLAIVLIVRSFIYEPPQIPSGSMMPTL NSTDFI
LVEKFAYGIKDPIYQKTLIETGHPKRGDIVVFKYPEDPKLDYIKRAVGLPGDKVTYDPVSKELT
IQPGCSSGQACENALPVTYSNVEPSDFVQTFSTRNGGEATSGFFEVPKQETKENGIRLSETS GG

PGRWRSYFYFITLIFFLAWLVKNVFIAVIIGPGGGVPGQQNATWIVPPGQYFMMGDNRDNSADS
RYWGFVPEANLVGRATAIWMSFDKQEGNGRLVCA-VALAASIN

S6 con mutación L312V

PYVSYTYDLGDTIEYKLGGLQVDSRATMANMFALILVIATLVTGILWCVDKFFFAPKRRERQAAA
QAAAGDSLDKATLKKVAPKPGWLETGASVFPVLAIVLIVRSFIYEPFQIPSGSMMPTLNSTDFI
LVEKFAYGIKDPIYQKTLIETGHPKRGDIVVFYKYPEDPKLDYIKRAVGLPGDKVTYDPVSKELT
IQPGCSSGQACENALPVTYSNVEPSDFVQTFSTRNGGEATSGFFEVPKQETKENGIRLSETS
GG
PGRWRSYFYFITLIFFLAWLVKNVFIAVIIGPGGGVPGQQNATWIVPPGQYFMMGDNRDNSADS
RYWGFVPEANLVGRATAIWMSFDKQEGNGRLVCA-VALAASIN

S6 con mutación V313G

PYVSYTYDLGDTIEYKLGGLQVDSRATMANMFALILVIATLVTGILWCVDKFFFAPKRRERQAAA
QAAAGDSLDKATLKKVAPKPGWLETGASVFPVLAIVLIVRSFIYEPFQIPSGSMMPTLNSTDFI
LVEKFAYGIKDPIYQKTLIETGHPKRGDIVVFYKYPEDPKLDYIKRAVGLPGDKVTYDPVSKELT
IQPGCSSGQACENALPVTYSNVEPSDFVQTFSTRNGGEATSGFFEVPKQETKENGIRLSETS
GG
PGRWRSYFYFITLIFFLAWLVKNVFIAVIIGPGGGVPGQQNATWIVPPGQYFMMGDNRDNSADS
RYWGFVPEANLVGRATAIWMSFDKQEGNGRLVCA-VALAASIN

S6 con mutación A319V

PYVSYTYDLGDTIEYKLGGLQVDSRATMANMFALILVIATLVTGILWCVDKFFFAPKRRERQAAA
QAAAGDSLDKATLKKVAPKPGWLETGASVFPVLAIVLIVRSFIYEPFQIPSGSMMPTLNSTDFI
LVEKFAYGIKDPIYQKTLIETGHPKRGDIVVFYKYPEDPKLDYIKRAVGLPGDKVTYDPVSKELT
IQPGCSSGQACENALPVTYSNVEPSDFVQTFSTRNGGEATSGFFEVPKQETKENGIRLSETS
GG
PGRWRSYFYFITLIFFLAWLVKNVFIIVIIIGPGGGVPGQQNATWIVPPGQYFMMGDNRDNSADS
RYWGFVPEANLVGRATAIWMSFDKQEGNGRLVCA-VALAASIN

S6 con mutación I322T

PYVSYTYDLGDTIEYKLGGLQVDSRATMANMFALILVIATLVTGILWCVDKFFFAPKRRERQAAA
QAAAGDSLDKATLKKVAPKPGWLETGASVFPVLAIVLIVRSFIYEPFQIPSGSMMPTLNSTDFI
LVEKFAYGIKDPIYQKTLIETGHPKRGDIVVFYKYPEDPKLDYIKRAVGLPGDKVTYDPVSKELT
IQPGCSSGQACENALPVTYSNVEPSDFVQTFSTRNGGEATSGFFEVPKQETKENGIRLSETS
GG
PGRWRSYFYFITLIFFLAWLVKNVFIAVIIGPGGGVPGQQNATWIVPPGQYFMMGDNRDNSADS
RYWGFVPEANLVGRATAIWMSFDKQEGNGRLVCA-VALAASIN

ANEXO 3: Metodología para determinar la topología del DI NALCN y NSP4

Figura 74. Gráfico resumido del ensayo cualitativo del reportero dual PhoA-LacZ

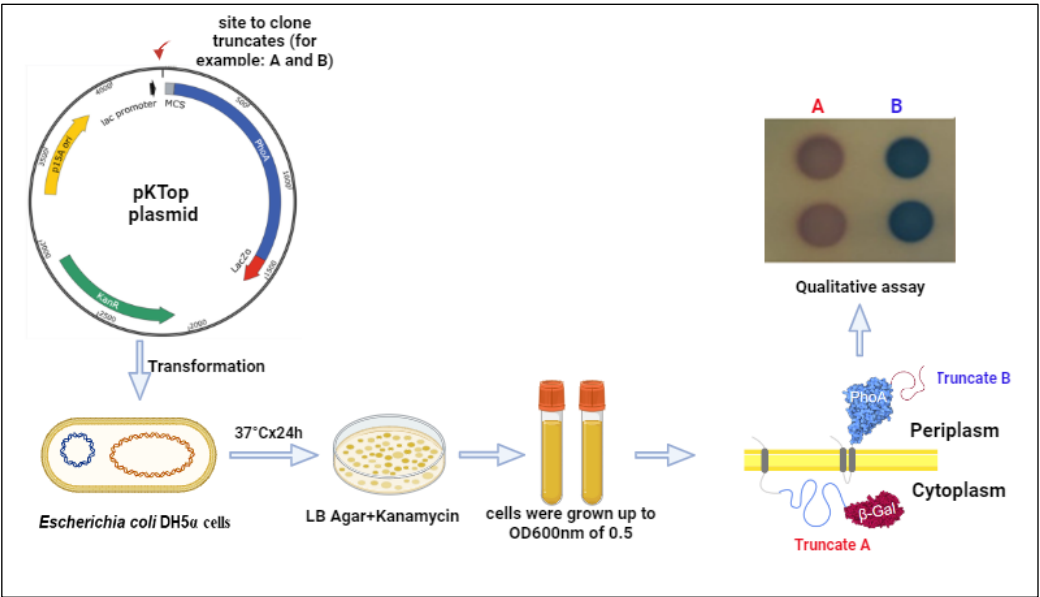
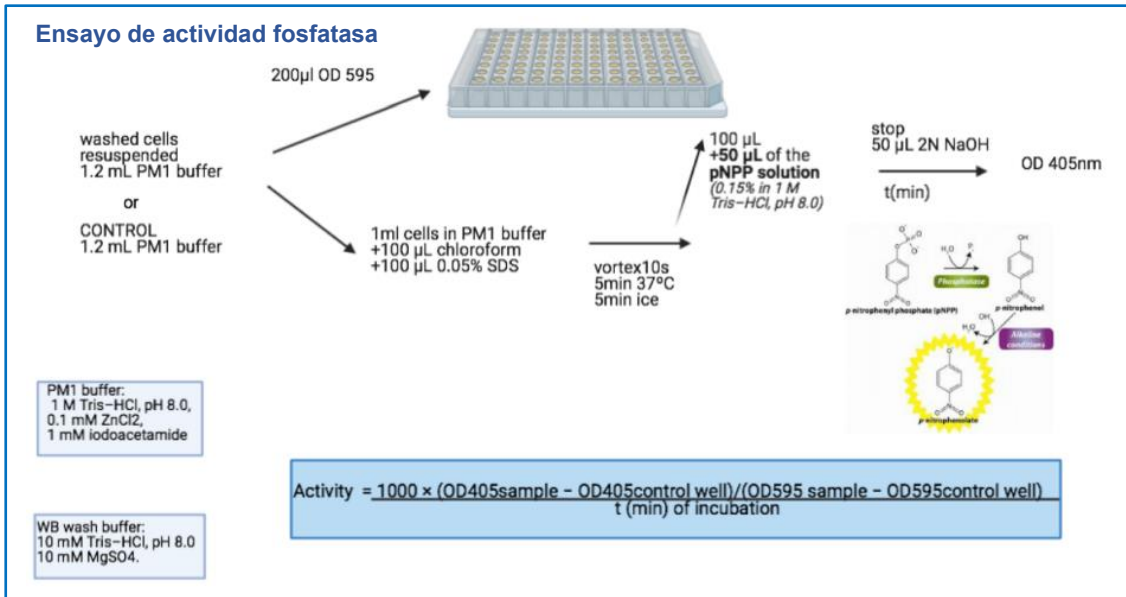
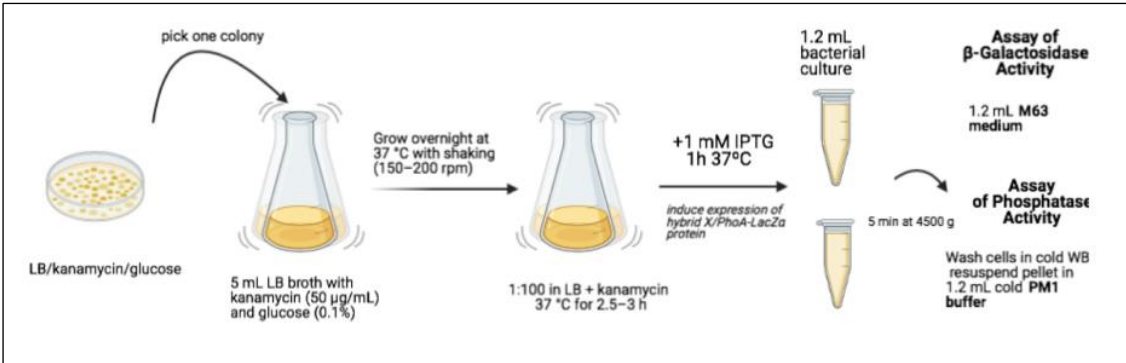


Figura 75. Gráfico resumido del ensayo cuantitativo del reportero dual PhoA-LacZ



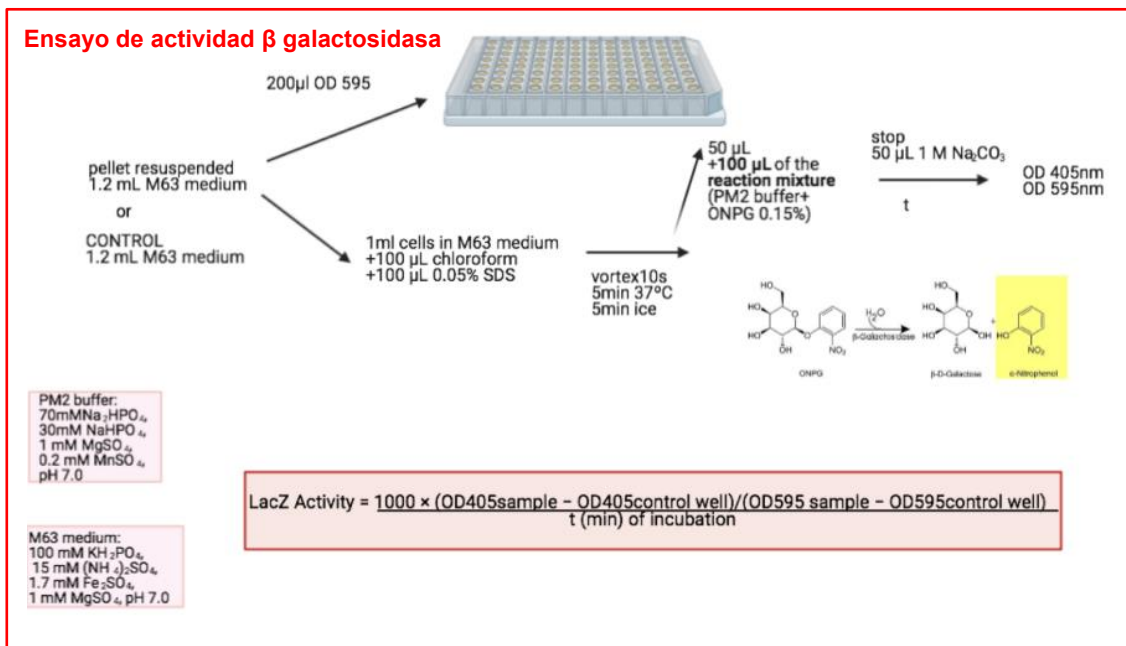
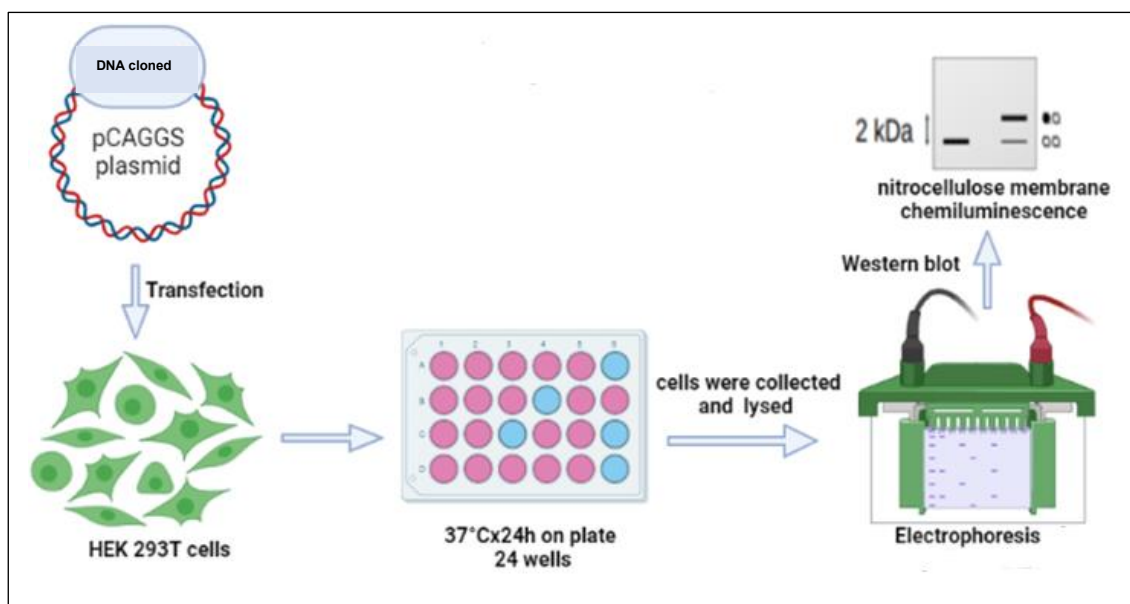


Figura 76. Gráfico resumido del ensayo de mapeo de glicosilaciones en líneas celulares humanas



A. Alineamiento de las secuencias del dominio DI la proteína NALCN humana y los dominios DI de los canales de sodio Nav1.2 y calcio Cav1.1 humanos

Proteína	Código UniProt	Nº de residuos Dominio I	(%) Identidad	Nº TM
hNALCN	Q8IZF0	1-380	100	6
hNav1.2	Q99250	1-451	20.92	6
hCav1.1	Q13698	1-380	20.61	6



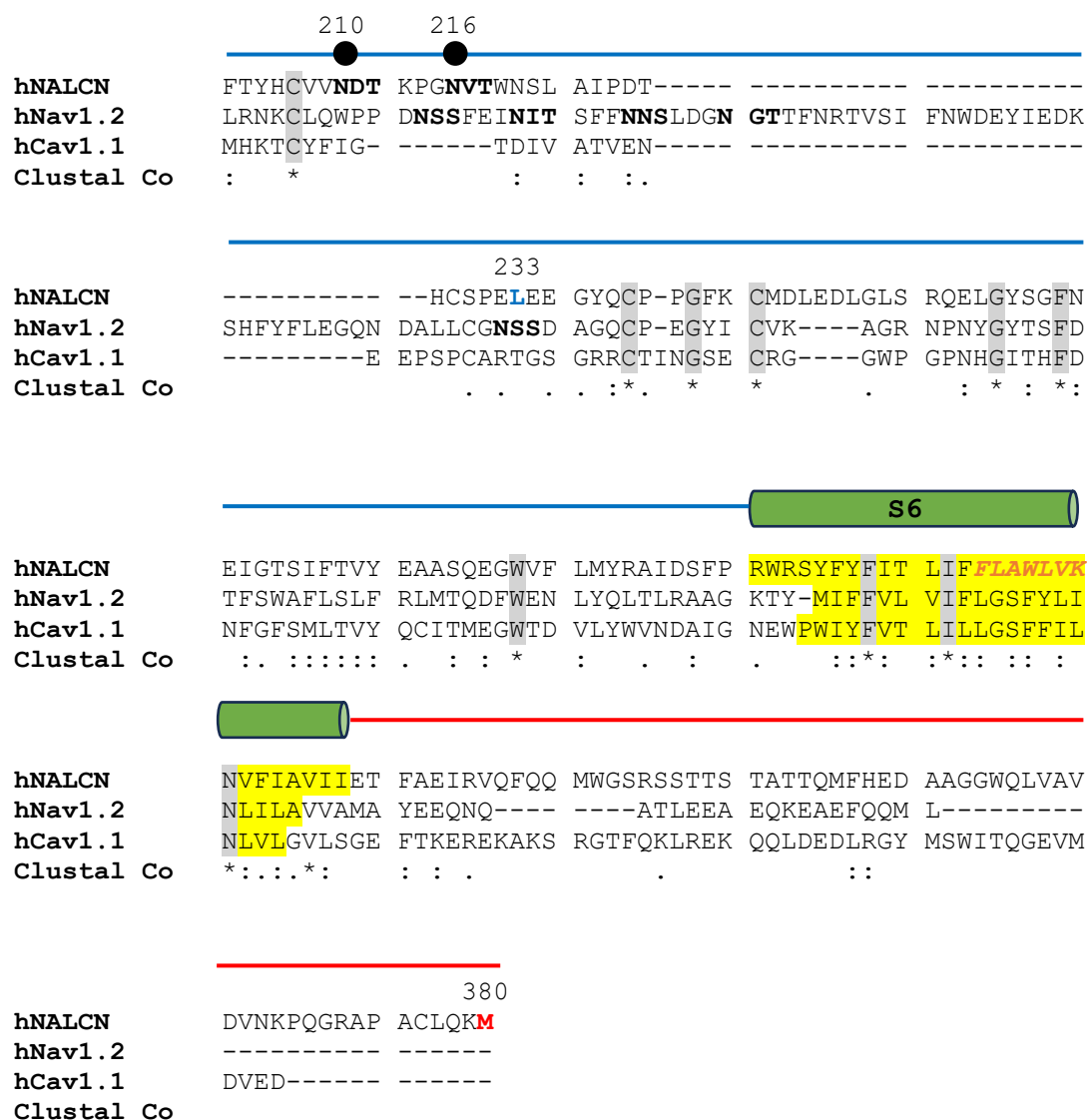


Figura 77. Alineamiento de las proteínas NALCN, Nav1.2 y Cav1.1. Arriba se indica los segmentos TM del DI de NALCN (cilindros verdes), la región citoplásmica (línea roja), la región extracelular (línea azul) y las dianas glicosilables del DI de NALCN en 210 y 216 (esferas negras). Abajo, se muestra el alineamiento mediante ClustalW de las proteínas NALCN, Nav1.2 y Cav1.1. En las secuencias alineadas observamos las regiones TM predichas (amarillo), los residuos de NALCN P65, M136 y L233 en azul (por su ubicación periplásmica/luminal) y K97, P163 y M380 en rojo (por su ubicación citosólica). Las dianas glicosilables del DI de NALCN N210 y N216 (en negrita). Los principales residuos del DI de NALCN implicados en la interacción de los segmentos S1-S4 (S52, D75, R146 y S121) en turquesa, la región de NALCN involucrada en el cambio conformacional de α a π hélice (FLAWLVK). (*) coincidencia en gris, (:) conservación fuerte, (.) conservación débil.

Tabla 30. Segmentos TM de las proteínas NALCN, Nav1.2 y Cav1.1. El número, la posición y la secuencia de los segmentos TM de la tabla se obtuvieron de UniProtKB. Los valores ΔG_{app} de cada segmento TM fueron calculados utilizando ΔG predictor.

Proteína	TM	Posición	Secuencia del TM	ΔG_{app}
hNALCN	S1	37-57	WVHSLLRICAIISVISVCMNT	1.488
	S2	66-90	PLQYVTFTLDTLLMFLYTAEMIAKM	2.382
	S3	106-127	RWCVFDFGMVFCLWVSLVLQVF	-0.289
	S4	138-158	PWGMLRIPRPLIMIRAFRIYF	2.329
	S5	174-199	SGEQIWSVSIFLLFFLLLYGILGVQM	-1,093
	S6	295-322	RWRSYFYFITLIFFLAWLVKNVFIIVII	-2,234
hNav1.2	S1	130-148	LFNMLIMCTILTNCVFMSTM	0.845
	S2	156-176	KNVEYTFTGIYTFESLIKILA	5.135
	S3	191-208	WNWLDFTVITFAYVTEFV	2.900
	S4	215-231	ALRTFRVLRALKTISVI	4.74
	S5	251-270	VMILTVFCLSVFALIGLQLF	-1,539
	S6	402-422	MIFFVLVIFLGSFYLINLILA	-1,938
hCav1.1	S1	52-70	PFETIILLTIFANCVALAV	0.933
	S2	86-106	LEKLEYFFLIVFSIEAAMKII	1.951
	S3	116-136	AYLRSGWNVLDFTIVFLGVFT	2.598
	S4	161-179	VKALRAFRVLRPLRLVSGV	5.024
	S5	197-218	FHIALLVLFMVIIYAIIGLELF	-2.711
	S6	310-330	PWIFYVTLILLGSFFILNLVL	-2.204

B. Alineamiento de los segmentos S6I-IV de NALCN con S6 de la proteína TRPV6.

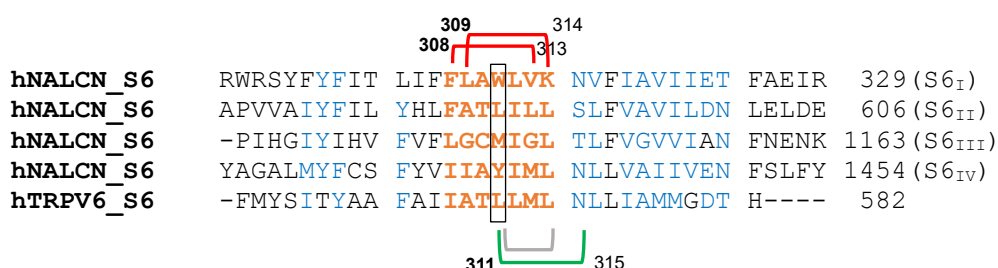


Figura 78. Alineación de segmentos S6 de NALCN y TRPV6. Los residuos conservados están coloreados de azul. Los siete residuos en cada S6 que forman la hélice π están sombreados en amarillo. Dentro de la hélice π , los enlaces de hidrógeno formados por el grupo carbonilo de los residuos (i) y el grupo amino de los residuos (i+5) se indican con líneas rojas. La línea verde indica el enlace de hidrógeno formado por el grupo carbonilo de los residuos (i) delineados en un cuadro negro con los residuos (i+4). La línea gris indica interacciones inter-hélice entre los residuos delineados en el cuadro negro con residuos de otro S6 adyacente. Los residuos estabilizadores de la hélice π del S6 del DI de NALCN son F308, L309 y W311.

C. Alineamiento de la secuencia N_t de NSP4 de SARS-CoV-2 y 2k-NS4B de flavivirus

Tabla 31. Información de las proteínas NSP4 y 2K-NS4B. Los datos de la tabla se obtuvieron de UniProtKB con los respectivos códigos de entrada, la posición de la proteína NSP4 y 2k-ns4B se indica con respecto a su primer ORF; y los valores Δ Gapp de cada primer TM fueron calculados utilizando Δ G Predictor.

Virus	Código UniProt	Posición de la proteína (aa)	Posición del primer TM	Δ Gapp
nsp4 SARS-CoV-2	P0DTD1	2764-3263	2772-2790	-1.409
2K-ns4B ZIKV	Q32ZE1	2243-2516	2248-2265	-0.414
2K-ns4B DENV-4	P09866	2220-2487	2225-2242	-0.745
2K-ns4B YFV	P03314	2234-2506	2239-2256	0.508
2K-ns4B WNV	P06935	2247-2525	2252-2269	-0.233

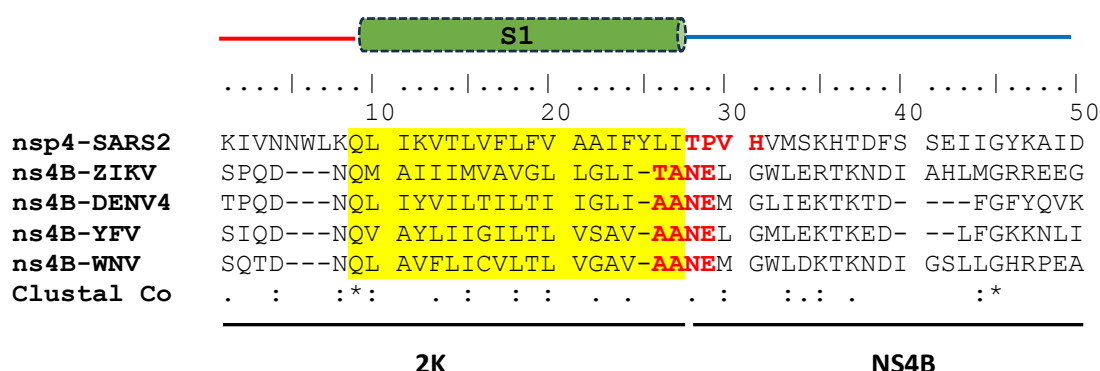


Figura 79. Alineamiento de la región N-terminal de las proteínas NSP4 y 2K-NS4B. El esquema de arriba nos indica el segmento TM S1 (cilindro verde) de ambas proteínas, la región citoplásmica de la proteína (línea roja) y la región orientada hacia el lumen (línea azul). Abajo, se muestra el alineamiento utilizando ClustalW de las proteínas nsp4 y NS4B. La región TM predicha manualmente usando Δ G predictor (amarillo) y el sitio de procesamiento (residuos en rojo). (*) coincidencia, (:) conservación fuerte de residuos, (.) conservación débil. 2K, es la región escindida que funciona como péptido señal de la proteína NS4B del virus Zika (ZIKV), Dengue serotipo 4 (DENV4), Fiebre amarilla (YFV) y virus del Nilo occidental (WNV).

D. Alineamiento de las secuencias de la proteína NSP4 de 5 coronavirus humanos

Tabla 32. Datos de la proteína NSP4 de 5 coronavirus. Los datos de la tabla se obtuvieron de UniProtKB, la ubicación de la proteína nsp4 está indicada según su posición aminoacídica con respecto al ORF1a. El (%) de identidad de las secuencias de la proteína nsp4 de cada coronavirus fue calculado con Blastp del NCBI.

Virus	Código UniProt	Ubicación de la proteína nsp4	Nº de residuos	(%) Identidad
SARS-CoV-2	P0DTD1	2764-3263	500	100
SARS-CoV	P0C6U8	2741-3240	500	80
MERS-CoV	K9N638	2741-3247	507	39.96
HCoV-OC43	P0C6U7	2751-3246	496	42.83
HCoV-HKU1	P0C6X4	2789-3284	496	43.7

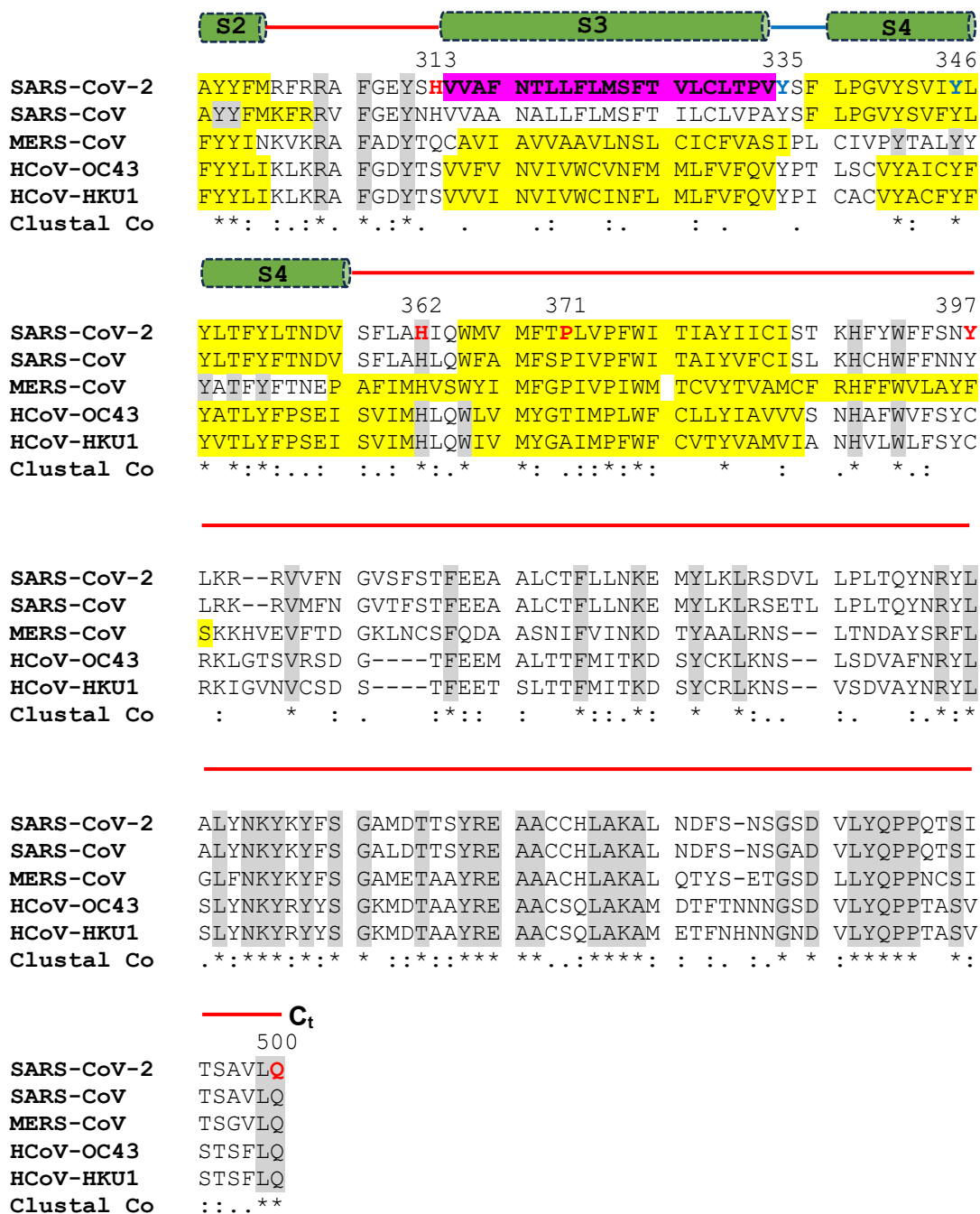


Figura 80. Alineamiento de las proteínas nsp4 de cinco coronavirus. El esquema de arriba nos muestra los segmentos TM (cilindros verdes), la región citoplásmica de la proteína NSP4 (línea roja) y la región luminal (línea azul). Abajo, se muestra el alineamiento mediante ClustalW de las proteínas NSP4 de 5 coronavirus (SARS-CoV-2, SARS-CoV, MERS-CoV, HCoV-OC43 y HCoV-HKU1). En las secuencias observamos las regiones TM predichas (amarillo), el segmento TM₃₁₃₋₃₃₅ (fucsia) el sitio de procesamiento de NSP4 del SARS-CoV-2 (residuos en negrita-cursiva) y sus residuos V32, E141, G164, D279, Y335 en azul (por su ubicación periplásmica/luminal), H313, P371, Y397 y Q500 en rojo (por su ubicación citosólica) y diana no canónica NIC (en turquesa). El residuo Y362 tiene ubicación TM. (*) coincidencia en gris, (:) conservación fuerte de residuos, (.) conservación débil.

Tabla 33. Segmentos TM de NSP4 de cinco coronavirus. El número, la posición y la secuencia de los segmentos TM de la tabla se obtuvieron de UniProtKB. Los valores ΔG_{app} de cada segmento TM fueron calculados utilizando ΔG predictor.

Virus	TMs	Posición del TM	Secuencia del TM	ΔG_{app}
nsp4 SARS-CoV-2	TM1	2776-2796	VTLVFLFV AAIFYLITPV HVM	-0,308
	TM2	3045-3065	ASIVAGGIVAIVVTCLAYYFM	-0,768
	TM3	3100-3120	FLPGVYSVIYLYLTFTYLTNDV	2,21
	TM4	3128-3148	WMVMFTPLVPFWITIAYIICI	0,526
nsp4 SARS-CoV	TM1	2755-2775	LLCVLA ALVCYIVMPV HTLSI	0,313
	TM2	3022-3046	ASVVAGGIIAILVTCAAYYFMKFR	-0,309
	TM3	3077-3097	FLPGVYSVFYLYLTFTYFTNDV	2,464
	TM4	3105-3125	WFAMFSPIVPFWITAIYVFCI	0,976
nsp4 MERS-CoV	TM1	2757-2777	VLATIIVFLC AVLMYLCLPT F	-2,252
	TM2	3028-3048	TTSLVLGIGLCAFLTLLFYI	-2,458
	TM3	3062-3082	AVIAVVAVLNSLCICFVASI	0,678
	TM4	3104-3124	PAFIMHVSWMYIMFGPIVPIWM	2,583
	TM5	3125-3145	TCVYTVAMCFRHHFFWVLAYFS	1,677
nsp4 HCoV-OC43	TM1	2752-2772	VFSYFVYVCFVLSLVCFGLW	-2,111
	TM2	2824-2844	TFGLSYYSNSMACPIVVAVI	4,22
	TM3	3009-3029	VFDLIYQLFKGLAQPVDFLAL	5,747
	TM4	3031-3051	ASSIAGAILAVIVVLVFYYLI	-2,046
	TM5	3063-3083	VVFNVIWCVNFMMLFVFQV	0,967
	TM6	3090-3110	VYAICYFYATLYFPSEISVIM	3,1
	TM7	3115-3135	LVMYGTIMPLWFCLLYIAVVV	-0,557
nsp4 HCoV-HKU1	TM1	2794-2814	LLYILFFISLICFILLWALLP	-5,237
	TM2	3069-3089	ASSIFGAILAIVVVLVFYYLI	-2,344
	TM3	3101-3121	VVINIVWCINFLMLFVFQV	0,5
	TM4	3128-3148	VYACFYFYVTLYFPSEISVIM	2,799
	TM5	3153-3173	IVMYGAIMPFWFVCTYVAMVI	0,751