

UNIVERSIDAD DE BUENOS AIRES

Facultad de Ciencias Exactas y Naturales
Departamento de Química Inorgánica,
Analítica y Química Física

Estudios biofísicos de la maquinaria del virus del Dengue: nuevas estrategias antivirales

Tesis presentada para optar al título de Doctora de
la Universidad de Buenos Aires, área Química
Inorgánica, Química Analítica y Química Física

Lic. Carolina Sarto

Directora de Tesis: Dra. Mehrnoosh Arrar

Consejero/a de Estudios: Dr. Damián Scherlis

Lugar de trabajo: Instituto de Cálculo, Facultad de Ciencias Exactas
y Naturales, Universidad de Buenos Aires / CONICET

Fecha de presentación: Buenos Aires, 25 de Julio del 2025

Resumen

Estudios biofísicos de la maquinaria del virus del Dengue: nuevas estrategias antivirales

Se estima que el virus del Dengue es responsable de aproximadamente 390 millones de infecciones a nivel mundial. Actualmente no existe un tratamiento específico disponible y el desarrollo de vacunas se dificulta por la potenciación de virulencia dependiente de anticuerpos, generada por los cuatro serotipos de este virus que infectan a los seres humanos. El objetivo general de esta tesis es comprender mediante herramientas computacionales las relaciones entre estructura, dinámica y función de biomoléculas involucradas en el proceso de replicación del ARN viral, en donde la participación de las proteínas virales no estructurales 3 y 5 (NS3 y NS5, respectivamente) resulta indispensable.

Durante el proceso de replicación, el dominio helicasa de la proteína NS3 (NS3hel) requiere hidrolizar ATP para poder desplazarse unidireccionalmente sobre el ARN viral, dejándolo disponible como cadena simple y facilitando su ingreso al sitio catalítico del dominio polimerasa de la proteína NS5. Esta última enzima sintetiza la cadena complementaria de ARN que servirá como molde para generar nuevas copias del ARN viral. Para poder comenzar la síntesis, la polimerasa debe unirse a un segmento estructurado no codificante del ARN viral, el 5' *Stem Loop A* (SLA), que actúa como promotor específico, activando alostéricamente el sitio catalítico de la polimerasa. En esta tesis se abordan dos fenómenos centrales del contexto de la replicación: el acoplamiento entre las actividades helicasa y ATPasa de la NS3hel y la activación alostérica de la polimerasa NS5.

Diversos trabajos han mostrado que la hidrólisis de ATP es necesaria para que la helicasa pueda desplazarse unidireccionalmente y separar la doble hebra de ARN y, a su vez, la presencia de ARN acelera la hidrólisis de ATP. A partir de simulaciones de dinámica molecular clásica, en un primer paso se exploró el espacio conformacional del complejo NS3hel-ARN y los posibles cambios según la ocupación del sitio catalítico. Como resultado principal, se observó la apertura del *loop* β6β1', el cual genera un canal que favorece la salida del fosfato inorgánico del sitio catalítico y además, se observó que durante el ciclo de hidrólisis de ATP, la afinidad de la NS3hel por el ARN disminuye, permitiendo así el desplazamiento de la helicasa.

En segundo lugar, se caracterizó el mecanismo mediante el cual ocurre la reacción química de hidrólisis de ATP de la NS3hel en presencia y ausencia de ARN. Para esto se empleó una descripción híbrida cuántica/clásica (QM/MM) y una búsqueda de caminos de mínima energía para comparar posibles mecanismos de reacción. Se observó que la presencia de ARN favorece una conformación del sitio catalítico que disminuye la barrera de activación para la reacción. Por último, se identificaron sitios aptos para la unión de pequeñas moléculas que son conservados en los cuatro serotipos y que pueden intervenir en los aspectos mecanísticos estudiados en la tesis.

El otro fenómeno abordado en esta tesis es el reconocimiento molecular entre el SLA y la polimerasa de la NS5. Mediante simulaciones de DM clásica y un análisis basado en la teoría de grafos y redes dinámicas, se identificaron caminos alostéricos entre el sitio de unión del SLA a la NS5 y el sitio catalítico y como mutaciones puntuales en el SLA o la NS5 pueden intervenir en dichos caminos.

Los resultados de esta tesis aportan información relevante para la comprensión de la maquinaria de replicación del virus del Dengue a nivel molecular y permiten diseñar nuevas estrategias terapéuticas.

Abstract

Biophysical Studies of the Dengue Virus Machinery: New Antiviral Strategies

Dengue virus is estimated to be responsible for approximately 390 million infections worldwide. Currently, there is no specific treatment available, and vaccine development is challenging due to a phenomenon known as "antibody-dependent enhancement," which arises from the four serotypes of this virus that infect humans. The general aim of this thesis is to understand, through computational tools, the relationships between structure, dynamics, and function of biomolecules involved in the viral RNA replication process, where the participation of the non-structural viral proteins 3 and 5 (NS3 and NS5, respectively) is essential.

During the replication process, the helicase domain of the NS3 protein (NS3hel) needs to hydrolyze ATP to move unidirectionally along the viral RNA, making it available as a single strand and facilitating its entry into the catalytic site of the polymerase domain of the NS5 protein. The latter enzyme synthesizes the complementary RNA strand that will serve as a template to generate new copies of the viral RNA. To initiate the synthesis, the polymerase must bind to a structured, non-coding segment of the viral RNA called Stem Loop A (SLA), which acts as a specific promoter, allosterically activating the polymerase's catalytic site. Thus, this thesis addresses two central phenomena within the context of replication: the coupling between the helicase and ATPase activities of NS3hel and the allosteric activation of the NS5 polymerase.

Experimentally, several studies have shown that ATP hydrolysis is necessary for the helicase to move unidirectionally and to separate the double-stranded RNA, while the presence of RNA accelerates ATP hydrolysis. Using classical molecular dynamics simulations, the first step involved exploring the conformational space of the NS3hel-RNA complex and the possible changes upon different occupancy of the catalytic site. As a main result, the opening of the β6–β1’ loop was observed, forming a channel that facilitates the release of inorganic phosphate from the catalytic site. Another important result is the decreased affinity of NS3hel for RNA during the ATP hydrolysis cycle, which facilitates helicase translocation.

The second step was to characterize the mechanism of the ATP hydrolysis reaction of NS3hel in the presence and absence of RNA. To this end, the free energy barriers of the rate-limiting step were compared using a hybrid quantum/classical description (QM/MM). The results of this study showed that catalysis occurs through a different mechanism and with lower activation energy in the presence of RNA. Finally, using a variant of classical molecular dynamics that involves the use of mixed solvent, conserved regions were identified in the four serotypes, both in the presence and absence of RNA, that could serve as druggable pockets.

The other phenomenon addressed in this thesis is the promoter effect of SLA on the polymerase activity of NS5. Using classical molecular dynamics, the conformations of NS5 and how they are affected by the interaction with SLA were studied. Through further analysis based on graph theory and dynamic networks, the amino acids responsible for the information flow between the SLA binding site of NS5 and the catalytic site were identified.

The results of this thesis provide relevant information for understanding at a molecular level the replication machinery of the Dengue virus and enable the design of new therapeutic strategies.

Índice

Resumen **2**
Abstract **3**
Índice **4**
Índice de Figuras y Tablas **8**
Glosario **10**
Capítulo 1: Introducción **11**
1.1 Desafíos contra el virus del dengue y posibles estrategias 11
1.2 El proceso de replicación viral 12
1.3 El Complejo de replicación 14
Objetivos de la tesis **16**
Bibliografía 18
Capítulo 2: Métodos **20**
2.1 Introducción 20
2.2 Dinámica molecular 20
2.2.1 Muestreo del espacio configuracional 22
2.2.2 Elección del modelo para representar un sistema 23
2.2.3 Representación clásica del sistema 24
2.2.3.1 Contribuciones de enlace 24
2.2.3.2 Contribuciones de no enlace 26
2.2.3.3 Condiciones periódicas de contorno 26
2.2.3.4 Campos de fuerza y sus parámetros 27
2.2.4 Representación cuántica de un sistema 29
2.2.4.1 Métodos ab initio 29
2.2.4.2 Método DFT y derivados semiempíricos 30
2.2.5 Representación híbrida del sistema 31
2.2.6 Preparación del sistema y etapas de una simulación 32
2.2.6.1 Minimización de la energía 33
2.2.6.2 Calentamiento o termalización 34
2.2.6.3 Equilibración de la presión 34
2.2.6.4 Equilibración final y producción 35
2.3 Análisis de la trayectoria 36
2.3.1 Alineamiento, medidas de desvíos y de fluctuaciones 36
2.3.2 Análisis de componentes principales 38
2.3.3 Bootstrapping 39
2.3.4 Función de distribución radial 39
2.3.5 Cálculo de energía libre de unión a partir de una trayectoria de dinámica molecular 40
2.4 Perfil de energía libre 42
2.4.1 Introducción 42
2.4.2 Muestreo sesgado 43
2.4.2.1 Búsqueda de caminos de mínima energía potencial 44
2.4.2.2 Búsqueda de caminos de mínima energía libre 44

2.4.2.3 Umbrella sampling....45
2.5 Alosterismo: Teoría de redes y grafos....45
2.5.1 Ejemplo sencillo de una red o grafo....46
2.5.2 Correlación lineal versus correlación generalizada....47
2.5.3 Distancias ponderadas....50
2.5.4 Caminos óptimos....50
2.5.5 Intermediación de aristas (edge betweenness)....51
2.5.6 Relación señal/ruido....52
Bibliografía....53
Capítulo 3: Estudio computacional del acoplamiento entre el ciclo de hidrólisis de ATP y el mecanismo de translocación de la helicasa NS3 sobre el ARN....57
3.1 Resumen....57
3.2 Introducción....57
3.3 Métodos....60
3.3.1 Preparación del sistema....60
3.3.2 Parametrización del ion fosfato....61
3.3.3 Simulaciones de dinámica molecular....62
3.3.4 Análisis de las dinámicas....63
3.3.5 Cálculos de energía de unión ARN-proteína....63
3.3.6 Calorimetría de titulación isotérmica....63
3.4 Resultados....64
3.4.1 Cambios en la flexibilidad de la proteína durante el ciclo de hidrólisis de ATP....64
3.4.2 Nueva conformación de loop β6-β1' abierto observada en ausencia de ATP....66
3.4.3 Movimientos relativos de los subdominios a partir del aumento de la flexibilidad conformacional....68
3.4.4 Interacciones proteína-ARN durante el ciclo de hidrólisis de ATP....69
3.4.5 Cambios en el tamaño del sitio de unión al ARN según la ocupación del sitio catalítico....71
3.4.6 Energía libre de unión entre la NS3hel y el ARN....73
3.5 Discusión....74
3.5.1 Alternancia entre estados flexibles y rígidos durante la hidrólisis de ATP....74
3.5.2 La conformación abierta del loop β6-β1' como ruta de salida para el fosfato....74
3.5.3 Tamaño del sitio de unión al ARN y posible sub-paso....75
3.5.4 Es necesario un ARN suficientemente largo para describir la interacción alostérica entre los sitios de unión de ARN y ATP....76
3.6 Conclusiones....77
Bibliografía....78
Capítulo 4: Hidrólisis de ATP estimulada por ARN en la helicasa NS3 del virus del dengue....82
4.1 Resumen....82
4.2 Introducción....82
4.3 Métodos....88
4.3.1 Construcción de los sistemas....89
4.3.1.1 Región cuántica....89
4.3.2 Minimización, calentamiento y equilibración de la presión....90

4.3.3 Camino de mínima energía potencial con el método de la banda elástica.....91
4.3.4 Camino de mínima energía libre con el método de la cuerda.....92
4.3.4.1 Optimización de la cuerda.....92
4.3.4.2 Perfiles de energía libre..... 92
4.3.5 Análisis conformacional de los estados reactivos.....92
4.3.5.1 Simulaciones clásicas de dinámica molecular..... 92
4.3.5.2 Correlación generalizada..... 93
4.3.6 Identificación de bolsillos conservados..... 93
4.3.6.1 Modelos por homología y preparación de sistemas.....93
4.3.6.1 Etapas de las simulaciones de dinámica molecular con solvente mixto... 93
4.3.6.2 Algoritmo de agrupamiento para identificar sitios de etanol.....94
4.3.6.3 Conservación de aminoácidos presentes en los bolsillos conservados....95
4.4 Resultados..... 95
4.4.1 Conformación del sitio activo de NS3hel unido a ARN favorece la hidrólisis de ATP.....95
4.4.1.1 Participación del residuo Glu conservado (motivo DEAH o motivo II) en el mecanismo de reacción..... 95
4.4.1.2 Estabilización del intermediario metafosfato pentavalente..... 98
4.4.2 La presencia y longitud del ARN impactan diferencialmente en la dinámica de los residuos del sitio catalítico y del agua catalítica.....99
4.4.3 La conformación del ARN afecta el posicionamiento del agua catalítica.....101
4.4.4 Implicancias para nuevas estrategias terapéuticas: sitios crípticos pan-serotipo para el descubrimiento de antivirales.....104
4.5 Discusión.....105
4.5.1 Detalles mecanísticos de la hidrólisis de ATP y el rol de E285.....105
4.5.2 Rol del ARN en la modulación de la hidrólisis de ATP.....106
4.5.3 Dinámica del extremo 5′ del ARN unido y del loop β6–β1′..... 107
4.5.4 Bolsillos drogables conservados entre serotipos de la NS3hel..... 107
4.5.5 Implicancias para la actividad NTPasa in vivo..... 108
4.6 Conclusiones.....108
Bibliografía..... 109
Capítulo 5: Activación alostérica de la polimerasa del virus del dengue.....114
5.1 Resumen..... 114
5.2 Introducción.....114
5.3 Métodos.....118
5.3.1 Preparación del sistema..... 118
5.3.2 Simulaciones de dinámica molecular.....120
5.3.3 Análisis de las trayectorias.....121
5.4 Resultados..... 121
5.4.1 Mecanismo de selección conformacional de la NS5 para la unión del SLA.... 121
5.4.2 Interacciones clave entre la NS5-pulgar y el top loop de SLA..... 123
5.4.3 Análisis de correlación generalizada.....126
5.4.4 Vías de comunicación alostérica entre el sitio de unión a SLA y el sitio catalítico de la polimerasa.....127
5.5 Discusión.....130

5.5.1 Requisitos conformacionales entre los subdominios de la NS5 para la actividad polimerasa....130
5.5.2 La importancia de un grupo carbonilo en la base de la posición 32 del SLA...131
5.5.3 Aparición de rutas alternativas y pérdida de eficiencia en la comunicación por la presencia de mutaciones.... 132
5.5.4 Mutaciones propuestas para el desarrollo de virus atenuados como estrategia preventiva contra el DENV....134
5.6 Conclusiones....134
Bibliografía.... 135
Capítulo 6: Conclusiones y Perspectivas....138
6.1 Conclusiones generales....138
6.2 Perspectivas....139
Anexo Capítulo 4....140
Agradecimientos....144
Publicaciones....146

Índice de Figuras y Tablas

Capítulo 1: Introducción..... 11

Figura 1.1. Ciclo de replicación viral..... **13**

Figura 1.2. Esquema de un evento de replicación y los actores principales involucrados..... **15**

Capítulo 2: Métodos..... 20

Figura 2.1. Relación entre el tamaño del sistema y tiempos para procesos característicos..... **23**

Figura 2.2: Componentes de la energía potencial clásica del sistema..... **25**

Figura 2.3: Ciclo termodinámico utilizado para el cálculo de energía de unión entre la proteína y el ARN..... **41**

Figura 2.4. Representación gráfica del cálculo de la información mutua a partir de la entropía de Shannon..... **48**

Capítulo 3: Estudio computacional del acoplamiento entre el ciclo de hidrólisis de ATP y el mecanismo de translocación de la helicasa NS3 sobre el ARN..... 57

Figura 3.1. Estructura de la helicasa NS3 del virus del dengue..... **58**

Tabla 3.1. Detalle de los sistemas estudiados..... **61**

Figura 3.2. Fluctuaciones en el esqueleto de la helicasa NS3 del virus del dengue en diferentes etapas del ciclo de hidrólisis de ATP..... **65**

Figura 3.3. Nueva conformación de *loop* abierto..... **66**

Figura 3.4. Análisis conformacional bidimensional del *loop* β6-β1' y separación de los subdominios 1 y 2 durante la hidrólisis de ATP..... **67**

Figura 3.5. Histograma 2D para d_{ATP} *versus* d_{RNA}..... **69**

Figura 3.6. Interacciones con el esqueleto fosfodiéster del ARN..... **70**

Figura 3.7. Cambios en el tamaño del sitio y la energía de unión al ARN según la ocupación del sitio catalítico..... **72**

Capítulo 4: Hidrólisis de ATP estimulada por ARN en la helicasa NS3 del virus del dengue..... 82

Figura 4.1. Vista general de la proteína no estructural 3 y el sitio catalítico ATPasa..... **84**

Figura 4.2. Múltiples mecanismos posibles para la hidrólisis de ATP..... **86**

Tabla 4.1. Carga total de cada sistema (unidades: *e*)..... **89**

Tabla 4.2. Etapas de la equilibración QM/MM..... **90**

Figura 4.3 Simulaciones cortas QM/MM para equilibrar los estados de reactivos y productos..... **96**

Figura 4.4. Potencial de fuerza media (PMF, kcal/mol) a lo largo de la coordenada de reacción... **97**

Figura 4.5. Interacciones de los residuos catalíticos con el agua catalítica..... **99**

Figura 4.6. Estabilidad de interacciones clave entre NS3hel y ATP-Mg^{2+} en presencia y ausencia de ARN..... **100**

Figura 4.7. Efecto alostérico de la conformación del RNA sobre el sitio catalítico…….. **103**
Figura 4.8. Bolsillos conservados entre serotipos en NS3hel obtenidos con y sin ARN a partir de sitios de etanol.…… **105**

Capítulo 5: Activación alostérica de la polimerasa del virus del dengue.....114

Figura 5.1. Representación del ARN viral del virus del dengue y del inicio de la replicación viral. Estructura del complejo NS5:SLA.…… **116**
Tabla 5.1. Detalles de los sistemas estudiados.…… **119**
Figura 5.2. Conformaciones de la NS5.…….. **123**
Tabla 5.2. Secuencias del *top loop* de los sistemas NS5:SLA estudiados, su presencia o no en la naturaleza y características experimentales.…… **124**
Figura 5.3. Representación esquemática de la interfaz SLA / NS5pol-pulgar.…… **125**
Figura 5.4. Matrices de correlación lineal y generalizada…….. **126**
Figura 5.5. Porcentaje de ocurrencia de aminoácidos que participan en las Vías de comunicación alostérica entre el sitio de unión a SLA y el sitio catalítico de la polimerasa…….. **128**
Tabla 5.3. Valores de relación señal-ruido para los complejos NS5-SLA…….…… **129**
Figura 5.6. Representación esquemática de la activación alostérica del NS5pol posterior a la unión del SLA.…… **133**

Anexo Capítulo 4.....140

Figura 4A.1. Búsqueda de caminos de mínima energía potencial con el método NEB……... **140**
Figura 4A.2. Descomposición del potencial de fuerza media PMF para cada variable colectiva y distancias de las variables colectivas a lo largo de la coordenada de reacción hallada mediante el método de la cuerda ASM para el camino de mínima energía libre…… **141**
Figura 4A.3. Distancias entre el fosfato gamma (Pγ) y los aminoácidos de motivos conservados que interactúan con el ATP o con el agua catalítica. Las series verdes corresponden a la NS3hel unida al ARN y la gris a la NS3hel en ausencia de ARN.…… **141**
Tabla 4A1. Variabilidad de las secuencias utilizadas en las simulaciones de dinámica molecular con solvente mixto y conservación de los aminoácidos de los bolsillos encontrados…… **142**
Figura 4A.4. Matrices de correlación generalizada para los estados considerados……. **143**

Glosario

Virus

DENV: Virus del dengue
HCV: Virus de la Hepatitis C
VIH: Virus de la inmunodeficiencia humana

Proteínas

NS: Proteína no estructural (Non-structural protein)
NS3: Proteína No Estructural 3 (Non-structural protein 3)
NS3pro: Dominio proteasa de la proteína NS3, región N-terminal
NS3hel: Dominio helicasa de la proteína NS3, región C-terminal
NS5: Proteína No Estructural 5 (Non-structural protein 5)
NS5met: Dominio metiltransferasa de la proteína NS5, región N-terminal
NS5pol: Dominio polimerasa de la proteína NS5, región C-terminal
$NS5^{F1mut}$: Proteína NS5 mutante en el motivo F1
E: Proteína de envoltura
M: Proteína de pre-membrana
C: Proteína de Cápside
SF1 y SF2: Superfamilias 1 y 2 de helicasas monoméricas
D1, D2, D3: Subdominios de la proteína NS3hel
NTPasa: Nucleósido trifosfatasa
RTPasa: Trifosfatasa de ARN 5′
DEAH: Motivo conservado (Asp-Glu-Ala-His)

Ácidos nucleicos

ARN (+): Ácido ribonucleico sentido positivo (genoma viral)
ARN (-): Ácido ribonucleico sentido negativo (compuesto por bases complementarias al genoma viral)
ARN-sc: ARN simple cadena
ARN-dc: ARN doble cadena
ARN7: Simple cadena de ARN de 7 nucleótidos
ARN16: Simple cadena de ARN de 16 nucleótidos
SLA: *Stem Loop A.* Región estructurada del extremo 5’ del ARN viral
SLA^{mut}: *Stem Loop A* mutante
SLA^{rev}: *Stem Loop A* revertante de SLA^{mut}
RTP: Ribonucleótido trifosfato
NTP: Nucleótido trifosfato
ATP: Adenosín trifosfato
ADP: Adenosín difosfato
Pi: Fosfato inorgánico (HPO_4^{2-} o $H_2PO_4^-$)
Pγ: Fosfato gamma del ATP
nts: nucleótidos

Metodología

Cryo-EM: Crío-microscopía electrónica
RMN: Resonancia magnética nuclear
PME: *Particle mesh Ewald*
RESP: *Restrained electrostatic potential fit*
TIP3P: *Transferable intermolecular potential with 3 points*
DFT: Teoría del Funcional de la Densidad
HF: Hartree-Fock
DFTB: *Density Functional Tight Binding*
B3LY: Becke, 3-*parameter,* Lee-Yang-Parr
QM/MM: *Quantum Mechanics/Molecular Mechanics*
NVT: Número de partículas, volumen y temperatura constante
NPT: Número de partículas, presión y temperatura constante
RMSD: *Root Mean Square Deviation*
RMSF: *Root Mean Square Fluctuation*
PCA: Análisis de componentes principales
MM-PB/GBSA: *Molecular Mechanics - Poisson Boltzmann / Generalized Born Surface Area*
PMF: Potential de fuerza media
SEP: Superficie de energía potencial
NEB: Método de la banda elástica (*Nudged elastic band)*
ASM: Método de la cuerda (*Adaptive string method)*
CL: Correlación lineal
CG: Correlación generalizada
IM: Información mutua
H: Entropía de Shannon
S/R: Relación Señal/Ruido
PDB: *Protein Data Bank*
VMD: *Visual Molecular Dynamics*
k_{cat}: Constante de velocidad de la catálisis
K_M: Afinidad aparente
SN1 y SN2: Sustitución nucleofílica uni y bi molecular
WT: *Wild-type*

Capítulo 1: Introducción

1.1 Desafíos contra el virus del dengue y posibles estrategias

El virus del dengue (DENV), perteneciente al género de los Flavivirus, es el agente causante de la fiebre del dengue y representa un problema de salud pública a nivel global [1]. Los mosquitos, vectores de este virus, se encuentran en áreas tropicales y subtropicales, lo que expone a la mitad de la población mundial al riesgo de infección [2,3]. Actualmente no existe un tratamiento antiviral específico disponible para la enfermedad causada por el virus del dengue. Si bien se han aprobado antivirales contra otros Flavivirus, como el virus de la Hepatitis C (HCV), no se ha podido comprobar que estos sean efectivos contra el virus del dengue. Más aún, el inconveniente de no contar con estrategias terapéuticas no es solo del dengue, sino que también abarca a muchos otros flavivirus como el virus de Zika o el de fiebre amarilla [4].

El desarrollo de vacunas para la prevención de infecciones causadas por el virus del dengue se dificulta por un efecto conocido como “potenciación dependiente de anticuerpos”. Este efecto se genera a partir de infecciones sucesivas entre al menos dos de los cuatro serotipos (DENV-1, DENV-2, DENV-3 y DENV-4) que afectan a los seres humanos. Se ha observado la aparición de síntomas más graves en pacientes que, después de ser infectados por un serotipo, son infectados por otro. Lo que se postula es que los anticuerpos generados por la primera infección, si bien le confieren inmunidad para este primer serotipo, a su vez generan un aumento en la capacidad infecciosa del segundo, exponiendo a la persona a una mayor probabilidad de desarrollar formas más graves de dengue [5,6]. Por eso, las vacunas aprobadas son “tetravalentes”, ya que deben conferir inmunidad frente a los cuatro serotipos para evitar este efecto. Además, una de ellas solo fue aprobada para personas previamente infectadas, lo que refleja otra limitación asociada a este fenómeno [7].

El modelado molecular y otros métodos computacionales permiten explorar, de manera más económica que los experimentos en el laboratorio, propiedades termodinámicas, paisajes conformacionales y detalles mecanísticos del funcionamiento de biomoléculas [8]. Cuando se aplican a blancos terapéuticos (típicamente proteínas o ácidos nucleicos) estos enfoques facilitan la identificación de posibles sitios de unión a fármacos, ya sea con el objetivo de bloquear funciones conocidas o de revelar bolsillos que pueden no ser visibles en estructuras cristalográficas. De esta manera, y mediante otros enfoques que no se abordan en esta tesis, como el *docking*, el estudio de estos sistemas a nivel atómico-molecular ha contribuido al desarrollo de agentes terapéuticos asistidos por herramientas computacionales [9,10]. Un caso de éxito para estos métodos fue la búsqueda de inhibidores de la enzima integrasa del virus de inmunodeficiencia humana (VIH). La aplicación de técnicas como las simulaciones de dinámica molecular y el *docking* han sido claves en el diseño y la optimización de estos fármacos. Estas herramientas permitieron predecir la unión de compuestos a la enzima y mejorar su eficacia, acelerando el desarrollo de terapias efectivas para personas con VIH [11]. Otro caso es el de Zanamivir, el primer inhibidor de la neuraminidasa desarrollado mediante diseño racional de fármacos. Usando cristalografía de rayos X y herramientas computacionales, se analizaron las estructuras de la neuraminidasa del virus de la influenza para luego optimizar un análogo del ácido siálico, mejorando su afinidad mediante la adición de un grupo guanidinio [12].

Con el objetivo general de profundizar en la comprensión del funcionamiento de biomoléculas indispensables para la replicación del DENV y futuros desarrollos de nuevas estrategias antivirales, en esta tesis se realizaron experimentos computacionales enfocados en las dos enzimas del virus del dengue: la helicasa NS3 y la polimerasa NS5. A partir de los resultados obtenidos se proponen estrategias para contribuir a un futuro desarrollo de antivirales y virus atenuados.

1.2 El proceso de replicación viral

Los virus generan una infección al tomar el control de las células del organismo huésped y aprovechar las funciones de la maquinaria celular para generar nuevas partículas virales o viriones que irán luego a infectar otras células. Cuando una partícula viral del DENV es internalizada en el citoplasma de una célula huésped, libera su genoma, una simple cadena de ARN sentido positivo de aproximadamente 11 kilobases. Por su clasificación como positivo, se entiende que el genoma puede ser directamente traducido a proteínas mediante la maquinaria celular, al igual que el ARN mensajero [13]. En este caso, la traducción da lugar a una única poliproteína que luego es clivada mediante la acción de proteasas celulares, y la proteasa viral, en diez proteínas: tres estructurales (envoltura (E), pre-membrana (M) y cápside (C)) y siete no estructurales (NS, por sus siglas en inglés): NS1, NS2a, NS2b, NS3, NS4a, NS4b, NS5 (Figura 1.1A, izquierda) [14]. Las proteínas NS desempeñan funciones críticas en la evasión de la respuesta del sistema inmune, la modificación de la membrana del retículo endoplasmático para crear “fábricas” de replicación viral y en el ensamblaje del virión [15].

Para la replicación del ARN viral (sentido positivo), primero es necesaria la síntesis de su cadena complementaria (sentido negativo). Luego, a partir de la cadena complementaria, se pueden sintetizar muchas copias del ARN viral y a cada una de ellas se le adiciona un *cap* tipo I metil-guanosín-trifosfato en el extremo 5’ (Figura 1.1A, abajo). La asociación de proteínas no estructurales en el retículo endoplasmático para permitir la síntesis eficiente del ARN viral se conoce como el complejo de replicación [13]. Finalmente, la cadena simple con *cap* de ARN (+) es compactada con proteínas de cápside, formando una estructura denominada nucleocápside y ensamblada en nuevas partículas virales [16] (Figura 1.1A, derecha).

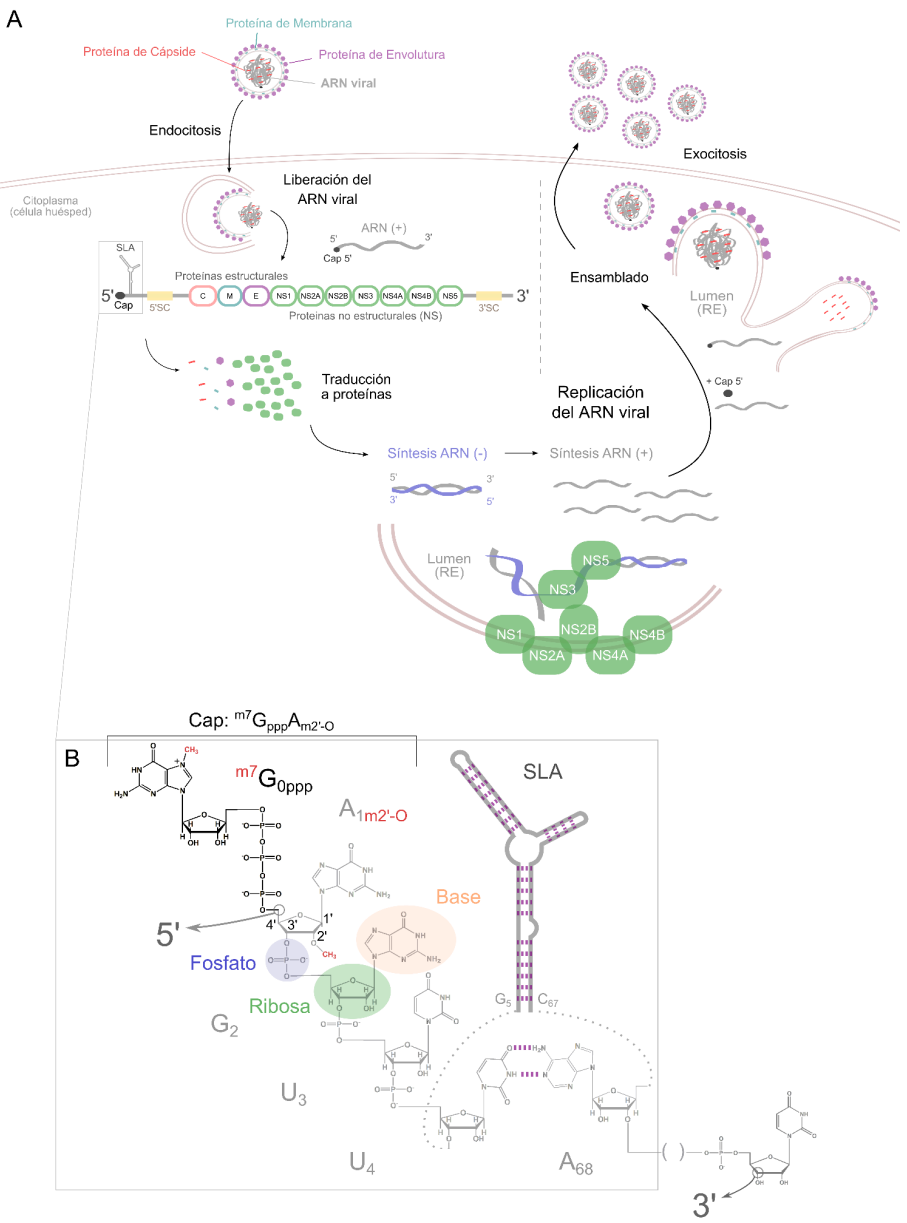


Figura 1.1. Ciclo de replicación viral. **A) De arriba hacia abajo en el lado izquierdo**, siguiendo las flechas que indican el avance del ciclo de replicación: Partícula viral entrando a la célula huésped por endocitosis. Se señala el ARN en gris y las proteínas estructurales (Proteína cápside en rojo, de membrana en celeste y de envoltura en violeta). Una vez dentro de la célula, se libera el ARN (+). Los extremos 5' y 3' están señalados así como el *cap* 5' (esfera negra). Se representa sobre el ARN la región del *Stem Loop A* (SLA), las secuencias complementarias (SC) 5' y 3' en amarillo, las proteínas estructurales coloreadas como se describió antes y las proteínas no estructurales (NS, por sus siglas en inglés) en color verde. **De abajo hacia arriba en el lado derecho,** siguiendo las flechas que indican el avance del ciclo de replicación: Las proteínas virales forman el complejo de replicación en el lumen del retículo endoplásmico. NS1, NS2A, NS2B, NS4A y NS4B son proteínas de membrana, mientras que NS3 y NS5 son solubles. En el complejo de replicación, a partir del ARN viral (+) mostrado en gris se sintetiza la cadena complementaria (-) mostrado en azul y a partir de esta última, se replican muchas cadenas de ARN viral (+). El ARN viral debe incorporar el *cap* 5', compactarse con la proteínas cápside, formando la nucleocápside, y ensamblarse a partir de la membrana del retículo endoplasmático, junto

con las proteínas estructurales para formar nuevas partículas virales que finalmente abandonan la célula por exocitosis. **B)** Detalles de la región 5' del ARN viral y estructura del ARN. Se muestra el detalle atomístico del *cap* 5' en negro, que involucra la adición de un nucleótido difosfato G_0, y en rojo las dos metilaciones realizadas por la NS5met (Ver Panel A). A partir de A_1 empieza el ARN viral, en el cual está señalada la posición 5' en la ribosa que le da nombre a ese extremo (si no estuviera el *cap*, esa posición tendría un grupo hidroxilo en lugar de un enlace como todos los demás). Las partes de un nucleótido están sombreadas en color azul para mostrar el grupo fosfato; en verde, el grupo ribosa; y en naranja, una base (G_2). Desde U_4 hasta A_{68} es la región estructurada no codificante conocida como SLA. La estructura se da por enlaces de hidrógeno entre las bases mayoritariamente, como los que se muestran con líneas discontinuas en violeta. Por fuera del rectángulo, se muestra el detalle del extremo 3', que al no continuar con la cadena tiene un grupo oxhidrilo en esa posición en lugar de un enlace como todos los demás.

1.3 El Complejo de replicación

Como se mencionó en la sección anterior, en el retículo endoplasmático se vinculan las proteínas no estructurales para formar el complejo de replicación. Dentro de este complejo, las proteínas NS3 y NS5, las únicas que contienen dominios catalíticos, llevan a cabo funciones imprescindibles para la replicación del ARN viral. En la Figura 1.2A se muestran en detalle los dominios de cada proteína, los cuales están covalentemente unidos por una región denominada *linker*. Existe un vínculo directo entre estas proteínas ya que el dominio NS5pol, la polimerasa viral que sintetiza ARN es asistida por el dominio NS3hel, la helicasa viral, que permite que el ARN molde esté disponible como simple cadena. Además, para que se inicie la síntesis, es necesario que la NS5pol reconozca primero al ARN viral. Este reconocimiento se da mediante la interacción entre un segmento estructurado de ARN viral no codificante ubicado en el extremo 5' conocido como *Stem Loop A* (SLA, Figura 1.1B) y la NS5.

La proteína NS3, además del dominio helicasa (NS3hel), contiene un dominio proteasa (NS3pro) como se esquematiza en la Figura 1.2A [17]. La NS3pro requiere de la unión a un cofactor, la proteína viral transmembrana NS2B, localizada en la membrana del retículo endoplasmático. De esta manera, al estar unida a la NS2B, la NS3 permanece cerca de la membrana del retículo endoplasmático, donde se organiza el complejo de replicación (Figura 1.1A, abajo). Por otro lado, la NS5 también está formada por dos dominios, como se esquematiza también en la Figura 1.2A: uno con actividad de metiltransferasa (NS5met) y de guanililtransferasa, involucrado en la formación del *cap* del ARN viral (Figura 1.1B), y otro ya mencionado con actividad ARN polimerasa ARN dependiente, NS5pol. Tanto la NS3 [18] como la NS5 [19] se consideran blancos atractivos para el desarrollo de fármacos antivirales, por ser dos proteínas multifuncionales críticas en el ciclo de replicación tanto del virus del dengue como de todos los Flavivirus [20,21].

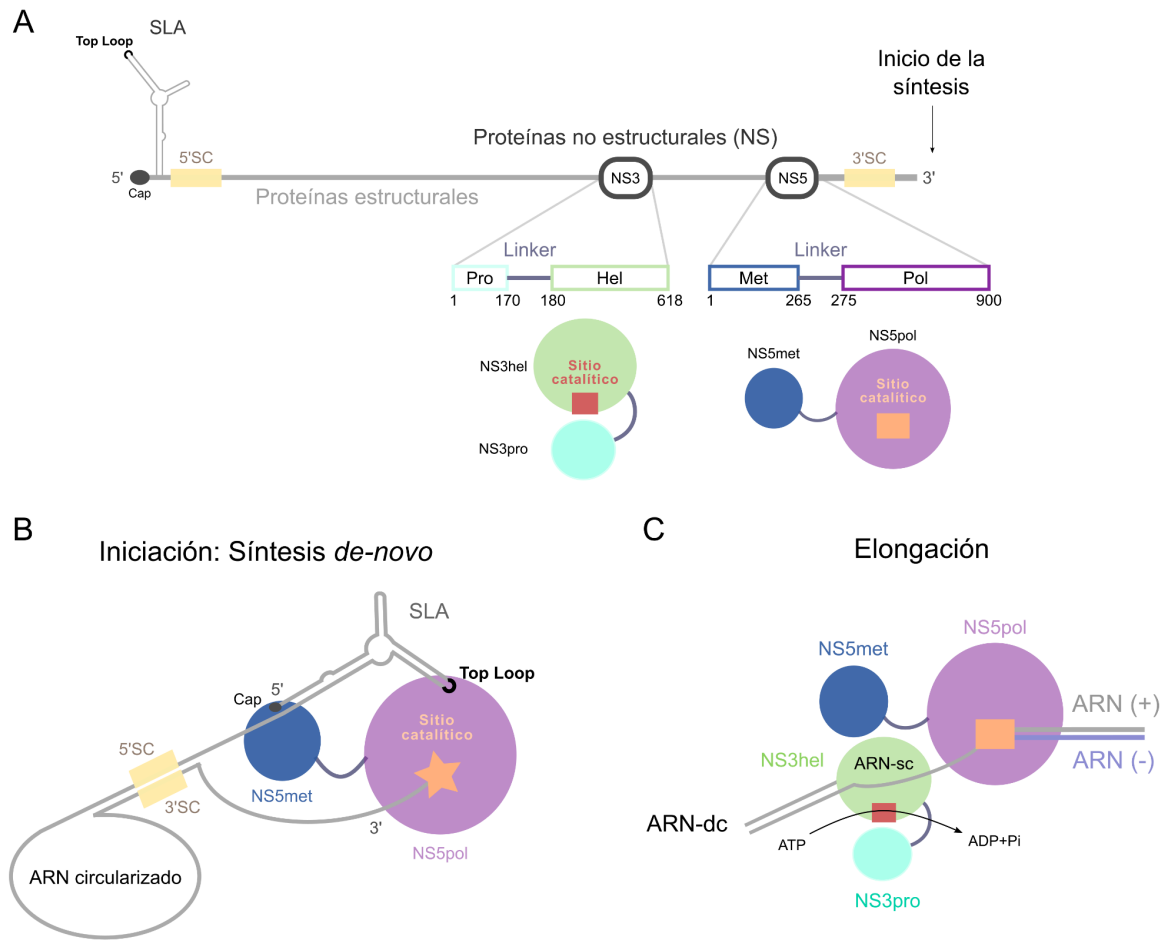


Figura 1.2. Representación de un evento de replicación y los actores principales involucrados. **A)** Se muestra un esquema del ARN viral con los extremos 5' y 3' señalados, el *cap* 5' se representa con una esfera negra. Se representa sobre el ARN la región del *Stem Loop A* (SLA) y resaltado en negro se encuentra el *top loop* y las secuencias complementarias (SC) 5' y 3' están en amarillo. En el extremo 3', una flecha indica que allí comienza la síntesis del ARN complementario a partir del molde. Las proteínas no estructurales NS3 y NS5 se esquematizan según sus dominios unidos por un *linker* (gris) y en los extremos está la numeración de los aminoácidos correspondientes. Para la NS3, se muestra en color cian el dominio proteasa, NS3pro, y en verde el dominio helicasa, NS3hel, en el que además se incluye al sitio catalítico como un rectángulo rojo. Para la NS5, se muestra en color azul el dominio metiltransferasa, NS5met, y en violeta el dominio polimerasa, NS5pol, junto con su sitio catalítico representado como un rectángulo naranja. **B)** Iniciación de la síntesis de ARN o síntesis *de novo.* El ARN se cirulariza apareando las secuencias complementarias mostradas en amarillo. Ubicado en el extremo 5', el SLA se une a través del *top loop* a la NS5pol activando al sitio catalítico (estrella naranja) y permitiendo que inicie la síntesis de ARN desde el extremo 3'. **C)** En el proceso de elongación, la interacción NS5-SLA ya no es un requerimiento para la síntesis. La NS5pol sintetiza el ARN complementario al molde (en este ejemplo se sintetiza ARN (-) mostrado en azul, a partir de ARN (+) mostrado en gris pero valdría lo mismo para el caso contrario). La NS3hel se ubica "adelante" de la NS5pol, en términos del ARN molde y separa estructuras de ARN doble cadena (ARN-dc), mientras cataliza la hidrólisis de ATP representada con una flecha negra, formando los productos ADP y fosfato inorgánico, Pi, en el sitio catalítico representado con un rectángulo rojo.

En el contexto de la replicación del ARN viral, la síntesis de ARN a partir de una simple cadena molde, ocurre en el sitio catalítico de la NS5pol. Muchas polimerasas, como la ADN polimerasa III, que se encarga de replicar el ADN de los seres humanos, requieren de una secuencia corta denominada *primer* que sea complementaria a la secuencia del extremo a partir del cual va a comenzar la síntesis. Por lo tanto, la función de este tipo de polimerasas es la de elongar el *primer* a partir de la secuencia molde. En cambio, la NS5pol tiene la capacidad de sintetizar una cadena de ARN a partir de otro ARN molde sin la necesidad de una secuencia *primer*, lo cual genera que la entrada y apareamiento de la primeras bases para comenzar la síntesis de la cadena complementaria constituya un evento diferente al resto de la síntesis o elongación. El evento de iniciación es conocido como síntesis *de novo.* En el caso de las polimerasas de los Flavivirus, la síntesis *de novo* tiene un requerimiento particular que consiste en la interacción de la NS5pol con el SLA [22], ubicado en el extremo 5' del ARN viral. Sin embargo, la síntesis de la cadena complementaria comienza en el extremo 3'. El genoma viral tiene la capacidad de circularizarse, a partir del apareamiento de secuencias complementarias (SC, Figura 1.2B) ubicadas en sus extremos 5' y 3', para acercar ambos extremos [23]. Por otro lado, al unirse a la NS5 reconociendo al ARN viral, el SLA activa alostéricamente al dominio polimerasa (Figura 1.2B, representado por el cambio de forma en el sitio catalítico). Los mecanismos mediante los cuales sucede la activación no se han comprendido en detalle, aunque hay algunos indicios que señalan una región de la NS5pol, conocida como motivo F1, como clave de este efecto [24].

En el complejo de replicación, la NS3hel se desplaza unidireccionalmente sobre la simple cadena de ARN en el sentido hacia el extremo 5' (mismo sentido que el avance de la síntesis) mientras convierte la energía química de la hidrólisis de NTP en energía mecánica que le permite separar la doble hebra de ARN o desarmar estructuras estables de la misma cadena monocatenaria [25] (como por ejemplo, las estructuras del SLA mostradas en la Figura 1.1B). De esta manera, deja disponible al ARN como cadena simple y facilita su ingreso al sitio catalítico de la NS5pol. Experimentalmente, se ha mostrado que la actividad NTPasa de la NS3 se incrementa en presencia de la NS5 [26]. También se ha verificado que la interacción entre la NS3 y la NS5 ocurre *in vivo* [27], deduciéndose que mientras la NS3 "avanza" hacia el extremo 5' del ARN, la NS5 se mueve "detrás" siguiendo a la NS3 y tomando el extremo 3' liberado como cadena simple para usarlo de molde en la síntesis de la cadena complementaria (Figura 1.2C). Esto ocurre tanto para el paso intermedio de la síntesis de la cadena complementaria, es decir de sentido negativo, como para la síntesis de la cadena positiva o replicación del ARN viral propiamente dicho.

Objetivos de la tesis

El objetivo general de esta tesis es profundizar en la comprensión de los mecanismos moleculares que permiten y promueven la replicación del ARN viral del virus del dengue. En primer lugar, se busca entender el acoplamiento entre el sitio de unión al ARN y el sitio catalítico de hidrólisis de NTP del dominio helicasa de la NS3 (Capítulos 3 y 4). En segundo lugar, se estudia el mecanismo molecular del reconocimiento del ARN viral y la activación del sitio catalítico de la polimerasa NS5 a partir de su interacción con el SLA (Capítulo 5).

Para estudiar los mecanismos que modulan las actividades de estas enzimas virales, se llevaron a cabo experimentos computacionales diseñados en base a las preguntas específicas en las que se basan los estudios de cada capítulo. Las bases teóricas de los

métodos computacionales para el modelado y análisis se describen en el Capítulo 2. En base a la evidencia experimental de que la hidrólisis de ATP es necesaria para la función helicasa de la NS3, en el Capítulo 3 se modelan y se estudian las etapas del ciclo catalítico de hidrólisis de ATP de la NS3hel en presencia de ARN para aportar información sobre cómo está modulada la interacción entre la proteína y el ARN según la ocupación del sitio catalítico. En el Capítulo 4 se retoma el estudio del vínculo entre la interacción con el ARN y la hidrólisis de ATP en la NS3hel desde otro punto de vista. El objetivo específico de este capítulo se centra en describir cómo influye la presencia o ausencia del ARN sobre el sitio catalítico y sobre el mecanismo químico de hidrólisis de ATP. Por último, en el Capítulo 5 se modela el sistema NS5 sin el promotor de la actividad polimerasa, el *stem loop A* (SLA), y se compara con el complejo NS5-SLA. Además, se incluye el estudio de mutantes para ahondar en el mecanismo de activación del sitio catalítico de la NS5pol a partir de la unión del SLA.

Las preguntas formuladas y los resultados computacionales obtenidos intentan apoyarse y ser contrastados permanentemente con datos experimentales. A partir de la información obtenida de los resultados computacionales, se busca no solo aportar nuevos indicios respecto de los mecanismos subyacentes que permiten la replicación del ARN viral del DENV, sino también proveer ideas acerca de posibles estrategias que contribuyan al desarrollo racional de fármacos antivirales y virus atenuados.

Bibliografía

1. Yang X, Quam MBM, Zhang T, Sang S. Global burden for dengue and the evolving pattern in the past 30 years. J Travel Med. 2021;28. doi:10.1093/jtm/taab146

2. Dengue and severe dengue. [cited 31 Mar 2025]. Available: https://www.who.int/news-room/fact-sheets/detail/dengue-and-severe-dengue

3. Bhatt S, Gething PW, Brady OJ, Messina JP, Farlow AW, Moyes CL, et al. The global distribution and burden of dengue. Nature. 2013;496: 504–507. doi:10.1038/nature12060

4. Martins IC, Ricardo RC, Santos NC. Dengue, West Nile, and Zika viruses: Potential novel antiviral biologics drugs currently at discovery and preclinical development stages. Pharmaceutics. 2022;14: 2535. doi:10.3390/pharmaceutics14112535

5. Halstead SB. Neutralization and antibody-dependent enhancement of dengue viruses. Advances in Virus Research. Elsevier; 2003. pp. 421–467. doi:10.1016/s0065-3527(03)60011-4

6. Obi JO, Gutiérrez-Barbosa H, Chua JV, Deredge DJ. Current trends and limitations in dengue antiviral research. Trop Med Infect Dis. 2021;6: 180. doi:10.3390/tropicalmed6040180

7. Cui G, Si L, Wang Y, Zhou J, Yan H, Jiang L. Antibody-dependent enhancement (ADE) of dengue virus: Identification of the key amino acid that is vital in DENV vaccine research. J Gene Med. 2021;23: e3297. doi:10.1002/jgm.3297

8. Leach AR. Molecular modelling. 2nd ed. London, England: Prentice-Hall; 2001.

9. Jones PE, Pérez-Segura C, Bryer AJ, Perilla JR, Hadden-Perilla JA. Molecular dynamics of the viral life cycle: progress and prospects. Curr Opin Virol. 2021;50: 128–138. doi:10.1016/j.coviro.2021.08.003

10. Winkler DA. Computational repurposing of drugs for viral diseases and current and future pandemics. J Math Chem. 2024. doi:10.1007/s10910-023-01568-3

11. Liao C, Nicklaus MC. Computer tools in the discovery of HIV-1 integrase inhibitors. Future Med Chem. 2010;2: 1123–1140. doi:10.4155/fmc.10.193

12. von Itzstein M, Wu WY, Kok GB, Pegg MS, Dyason JC, Jin B, et al. Rational design of potent sialidase-based inhibitors of influenza virus replication. Nature. 1993;363: 418–423. doi:10.1038/363418a0

13. van den Elsen K, Quek JP, Luo D. Molecular insights into the Flavivirus replication complex. Viruses. 2021;13: 956. doi:10.3390/v13060956

14. Kuhn RJ. Structural proteomics of dengue virus. Current Opinion in Microbiology. 2008;11: 369–377. doi:10.1016/j.mib.2008.06.004

15. Bartenschlager R, Miller S. Molecular aspects of Dengue virus replication. Future Microbiol. 2008;3: 155–165. doi:10.2217/17460913.3.2.155

16. Clyde K, Kyle JL, Harris E. Recent advances in deciphering viral and host determinants of dengue virus replication and pathogenesis. J Virol. 2006;80: 11418–11431. doi:10.1128/jvi.01257-06

17. Li H, Clum S, You S, Ebner KE, Padmanabhan R. The serine protease and RNA-stimulated nucleoside triphosphatase and RNA helicase functional domains of dengue virus type 2 NS3 converge within a region of 20 amino acids. J Virol. 1999;73: 3108–3116. doi:10.1128/JVI.73.4.3108-3116.1999

18. Lescar J, Luo D, Xu T, Sampath A, Lim SP, Canard B, et al. Towards the design of antiviral inhibitors against flaviviruses: the case for the multifunctional NS3 protein from Dengue virus as a target. Antiviral Res. 2008;80: 94–101. doi:10.1016/j.antiviral.2008.07.001

19. Lim SP, Noble CG, Shi P-Y. The dengue virus NS5 protein as a target for drug discovery. Antiviral Res. 2015;119: 57–67. doi:10.1016/j.antiviral.2015.04.010

20. Sampath A, Padmanabhan R. Molecular targets for flavivirus drug discovery. Antiviral Res. 2009;81: 6–15. doi:10.1016/j.antiviral.2008.08.004

21. Knyazhanskaya E, Morais MC, Choi KH. Flavivirus enzymes and their inhibitors. Enzymes. 2021;49: 265–303. doi:10.1016/bs.enz.2021.07.006

22. Filomatori CV, Lodeiro MF, Alvarez DE, Samsa MM, Pietrasanta L, Gamarnik AV. A 5' RNA element promotes dengue virus RNA synthesis on a circular genome. Genes Dev. 2006;20: 2238–2249. doi:10.1101/gad.1444206

23. Alvarez DE, Lodeiro MF, Ludueña SJ, Pietrasanta LI, Gamarnik AV. Long-range RNA-RNA interactions circularize the dengue virus genome. J Virol. 2005;79: 6631–6643. doi:10.1128/JVI.79.11.6631-6643.2005

24. Iglesias NG, Filomatori CV, Gamarnik AV. The F1 motif of dengue virus polymerase NS5 is involved in promoter-dependent RNA synthesis. J Virol. 2011;85: 5745–5756. doi:10.1128/JVI.02343-10

25. Jeremías Incicco J, Gebhard LG, González-Lebrero RM, Gamarnik AV, Kaufman SB. Steady-State NTPase Activity of Dengue Virus NS3: Number of Catalytic Sites, Nucleotide Specificity and Activation by ssRNA. PLOS ONE. 2013;8: e58508. doi:10.1371/journal.pone.0058508

26. Cui T, Sugrue RJ, Xu Q, Lee AK, Chan YC, Fu J. Recombinant dengue virus type 1 NS3 protein exhibits specific viral RNA binding and NTPase activity regulated by the NS5 protein. Virology. 1998;246: 409–417. doi:10.1006/viro.1998.9213

27. Kapoor M, Zhang L, Ramachandra M, Kusukawa J, Ebner KE, Padmanabhan R. Association between NS3 and NS5 proteins of dengue virus type 2 in the putative RNA replicase is linked to differential phosphorylation of NS5. J Biol Chem. 1995;270: 19100–19106. doi:10.1074/jbc.270.32.19100

Capítulo 2: Métodos

2.1 Introducción

La función que cumplen los sistemas macromoleculares, como las proteínas y los ácidos nucleicos, en un contexto biológico, está estrechamente vinculada con su estructura y dinámica. Si bien es posible caracterizar muchas propiedades funcionales, estructurales y dinámicas a través de experimentos *in vivo*, en ensayos celulares o *in vitro*, una herramienta complementaria que se viene desarrollando y creciendo hace aproximadamente 50 años para estudiar este tipo de sistemas es el de las simulaciones computacionales de dinámica molecular [1]. La base metodológica de esta tesis se centra en dicha técnica de simulación, ya que permite caracterizar las posibles conformaciones que adoptan los sistemas biomoleculares y su evolución temporal para explorar detalles de sus movimientos internos y vincularlos con funciones conocidas. Más aún, esta técnica tiene la capacidad de predecir comportamientos de las biomoléculas simuladas y proponer hipótesis que motiven nuevos ensayos experimentales.

Las simulaciones de dinámica molecular clásica permiten explorar la superficie de energía potencial y obtener información sobre distintos estados conformacionales de un sistema siempre y cuando la barrera de energía que separa a estos sea menor que la energía térmica disponible. Para estados separados por barreras altas o estados fuera del equilibrio, será necesario recurrir a otras técnicas que se denominan de “muestreo avanzado”. El resultado de realizar simulaciones de dinámica molecular consiste en obtener las posiciones de las partículas del sistema a lo largo del tiempo, lo que se conoce como una trayectoria. Una vez que se conoce una trayectoria del sistema, se aplican diversos métodos de análisis para extraer información de la dinámica como puede ser la caracterización de las conformaciones exploradas, las interacciones presentes, cálculos de energía de unión o cálculos de correlaciones entre distintas regiones del sistema, entre otros. En esta tesis se realizan simulaciones de dinámica molecular con modelos atomísticos y con una representación clásica o híbrida clásica-cuántica.

2.2 Dinámica molecular

La dinámica molecular de un sistema formado por N partículas se basa en obtener su trayectoria, las posiciones (x, y, z) de cada una de las partículas a lo largo del tiempo, a partir de posiciones iniciales conocidas, según lo establecido por la segunda ley de Newton:

$$\frac{d^2 \bar{r}_i}{dt^2} = \frac{\overline{F}_i}{m_i}, \tag{2.1}$$

donde $\bar{r}_i$ representa la posición (x_i, y_i, z_i) de una partícula i de masa m_i y $\overline{F}_i$ representa las componentes de la fuerza (Fx_i, Fy_i, Fz_i) que actúa sobre la partícula.

Aunque existen diversos modelos para biomoléculas, en esta tesis nos centramos en un modelo atomístico, en donde cada partícula representa un átomo. De esta manera, la masa

de la partícula será la masa del átomo en cuestión. Las fuerzas que existen entre las partículas se relacionan con la energía potencial según:

$$\overline{F}_i = - \nabla V(\overline{r}_i) \ . \tag{2.2}$$

Por lo tanto, será necesario conocer la forma funcional de la energía potencial (que se detalla en la Sección 2.2.6), para acceder a las fuerzas que actúan sobre cada partícula.

Dada la gran cantidad de partículas que tiene un sistema y la compleja relación funcional de la energía potencial con las posiciones de las mismas, la Ecuación 2.1 no puede resolverse analíticamente, sino que resulta necesario utilizar métodos numéricos. La forma de resolución numérica empleada, implica aproximar como intervalos de tiempo finitos (Δt) las diferencias infinitesimales de tiempo (dt), de manera tal que se puede aproximar también las fuerzas en esos intervalos como constantes y así calcular las sucesivas posiciones a lo largo del tiempo.

Es necesario conocer las posiciones iniciales del sistema para realizar cálculos de dinámica molecular. Una forma de obtener estos datos es a través de experimentos como pueden ser la cristalografía de rayos X, la crío-microscopía electrónica (Cryo-EM) o la resonancia magnética nuclear (RMN). Si no se conoce la estructura de una macromolécula pero se espera que sea similar y análoga a otra ya conocida, puede usarse el método de “modelado por homología”, que requiere dar una estructura molde y una secuencia, para obtener la estructura de esta última en base a la primera [2]. En los últimos años, además, se ha avanzado en el desarrollo de algoritmos de inteligencia artificial de estructura de proteínas principalmente, pero también de ácidos nucleicos, que se apoyan en las bases de datos existentes para predecir estructuras desconocidas a partir de una secuencia de referencia, sin la necesidad de brindar explícitamente una estructura similar [3]. Por otro lado, si bien es posible obtener posiciones iniciales a partir de cálculos teóricos de estructura electrónica, estos sólo son posibles de realizar para sistemas pequeños.

Además de conocer las posiciones iniciales, es necesario asignar las velocidades iniciales de cada partícula. Para realizar este procedimiento de manera realista, las velocidades iniciales se asignan en base a la distribución de Maxwell-Boltzmann, que vincula la temperatura del sistema con la probabilidad de que una partícula tenga una determinada velocidad. Una vez caracterizada la configuración inicial (posiciones y velocidades), es posible calcular sucesivamente nuevas velocidades y posiciones a cada paso i, para cada partícula, al transcurrir un intervalo de tiempo Δt, según el algoritmo de *Velocity Verlet* [4]:

$$\overline{r}_i = \overline{r}_{i-1} + \overline{v_{i-1}}\,\Delta t + \frac{\overline{F_{i-1}}}{2\,m}\,\Delta t^2 \ , \tag{2.3}$$

$$\overline{v}_i = \overline{v}_{i-1} + \frac{1}{2}\left(\frac{\overline{F}_{i-1} + \overline{F}_i}{m}\right)\Delta t \ . \tag{2.4}$$

A cada paso, una vez obtenidas las nuevas posiciones (Ecuación 2.3), se vuelve a calcular la fuerza y se usa este valor, combinado con la fuerza del paso anterior, para calcular las velocidades actualizadas (Ecuación 2.4). De esta manera, resolviendo sucesivamente las Ecuaciones 2.3 y 2.4, es posible obtener la trayectoria. Existen otros algoritmos para integrar

numéricamente la Ecuación 2.1 como el algoritmo de Verlet [5] o el denominado “leapfrog”, pero se ha mostrado en detalle el algoritmo de *Velocity Verlet* como ejemplo, por ser uno de los más usados [6,7].

El paso de integración (Δt), intervalo de tiempo con el que se actualizan las posiciones y velocidades, en dinámica molecular está limitado por las vibraciones más rápidas del sistema, que suelen ser los enlaces de hidrógeno. Para poder usar un paso de integración más grande y, por lo tanto, simular intervalos de tiempo más largos, es usual aplicar restricciones conocidas como SHAKE [8], que mantienen constantes las longitudes de los enlaces seleccionados, permitiendo así incrementar el paso temporal de 1 fs a 2 fs sin comprometer la estabilidad de la simulación [6].

2.2.1 Muestreo del espacio configuracional

El costo computacional de una simulación de dinámica molecular se verá afectado por el número de partículas que se desee incluir en el análisis, que se define en base a la representación del sistema elegido (Secciones 2.2.3, 2.2.4, 2.2.5), por el intervalo de tiempo que define cada paso para el cual se calcula un nueva configuración (Δt) y la cantidad de pasos realizados.

En un instante dado, para definir la configuración de un sistema macroscópico formado por N partículas, se requiere conocer el valor de cada uno de sus 6N grados de libertad, dados por las posiciones ($x,\ y,\ z$) y velocidades ($v_x,\ v_y,\ v_z$) de cada partícula. El conjunto de estas variables constituyen el “espacio de las fases”. Según la termodinámica estadística, es posible acceder a las propiedades de un sistema macroscópico a partir de conocer las propiedades de los sistemas microscópicos que lo componen o “ensamble”. El valor obtenido al medir un observable O en el sistema macroscópico, equivale al promedio de los valores individuales de cada microestado del ensamble.

En el contexto de una simulación de dinámica molecular, se aproxima el promedio en el ensamble por el promedio temporal de una propiedad, bajo la hipótesis ergódica, que estrictamente es válida al simular un solo microestado durante un tiempo infinito:

$$< O >_{t\to\infty} = \frac{1}{\tau}\int O(t)dt = < O >_{ensamble} . \qquad (2.6)$$

De esta manera, la hipótesis ergódica establece que, si el tiempo simulado τ es suficientemente largo, un sistema explorará todo el espacio de fases con una frecuencia proporcional a su densidad de probabilidad. Sin embargo, debido a que las trayectorias tienden a recorrer regiones de baja energía sobre la superficie de energía potencial que además están sesgadas por las estructuras iniciales, los estados menos probables, o aquellos separados por barreras más grandes que la energía térmica, pueden quedar subrepresentados. Para sistemas como las macromoléculas resulta imposible realizar una exploración completa del espacio de las fases, pero lo que sí se busca es hacer este recorrido lo más exhaustivo posible. Para mejorar la exploración del espacio conformacional, se realizan múltiples réplicas con diferentes condiciones iniciales, o bien se emplean técnicas de

muestreo sesgado. Estas estrategias mejoran la descripción del comportamiento estadístico del sistema y permiten un mayor acercamiento de la simulación teórica a la realidad.

2.2.2 Elección del modelo para representar un sistema

Los métodos de simulación computacional requieren un modelo que represente al sistema. Esta representación se define a partir del nivel de detalle del modelo que depende tanto del sistema en estudio como del tipo de pregunta que se quiera responder. Con respecto al sistema, la variable fundamental va a ser su tamaño, ya que la cantidad de cálculos a realizar escala con la cantidad de partículas que definen al sistema y por lo tanto, escala también su costo computacional. Con respecto al tipo de pregunta, las variables que influirán en el modelado serán el proceso que se quiera estudiar (cambio conformacional, reacción química, proceso difusivo, etc) y el tiempo característico de dicho proceso (Figura 2.1).

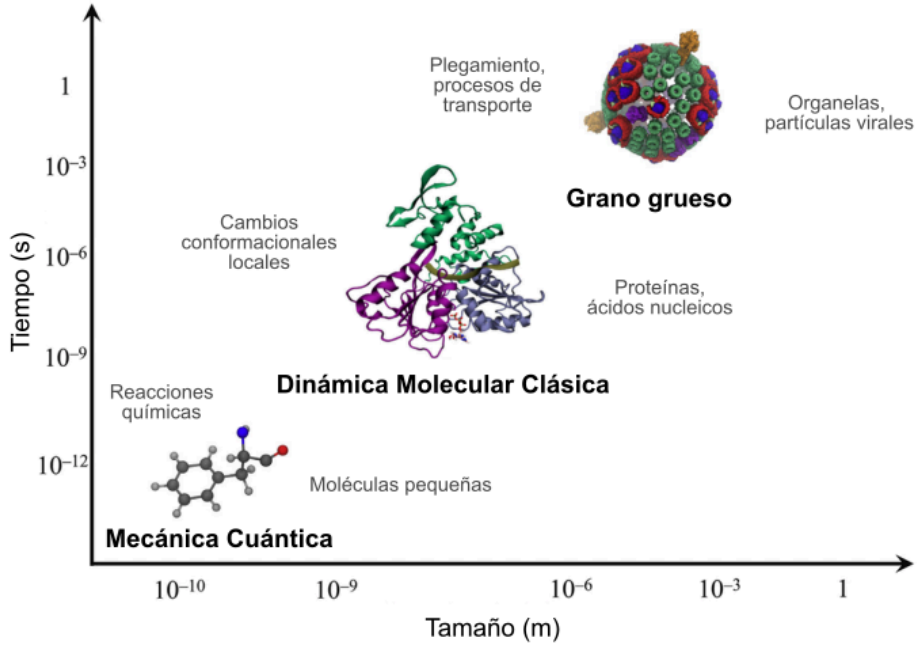


Figura 2.1. Relación entre el tamaño del sistema y tiempos para procesos característicos. De abajo hacia arriba se presentan ejemplos de: una molécula pequeña, una proteína y una partícula viral. Figura adaptada de [9,10].

En el caso de los sistemas biomoleculares estudiados en esta tesis, la principal representación elegida fue la clásica y atomística, es decir, una partícula por átomo, ya que permite explorar el espacio conformacional con costos computacionales razonables para procesos con tiempos característicos en el orden de los nano a microsegundos, como los cambios conformacionales. Esta representación, sin embargo, puede ignorar aspectos relevantes del entorno proteico, como cambios en los estados de protonación de algunos aminoácidos, intercambios entre formas tautoméricas u otros efectos electrónicos sutiles que podrían influir en el comportamiento global del sistema. Por otro lado, como se analizará en el Capítulo 4, el estudio de reacciones químicas requiere una representación cuántica, ya que durante estos procesos se produce una redistribución de la densidad electrónica que no puede capturarse con un modelo clásico, donde dicha densidad permanece fija en cada átomo. Por lo tanto, para este tipo de procesos, será necesario describir a los electrones

explícitamente bajo las leyes de la mecánica cuántica, lo cual genera un aumento significativo en las variables del sistema que hace que sólo sea posible estudiar procesos con tiempos característicos del orden de los femto-picosegundos.

2.2.3 Representación clásica del sistema

Para los sistemas estudiados en esta tesis fue adecuado realizar simulaciones de dinámica molecular clásica con una representación atomística. Como se muestra en las Ecuaciones 2.1 y 2.2, el movimiento de cada uno de los átomos-partículas se debe a la energía potencial (V) que se origina por la interacción con las demás partículas del sistema y depende de su posición relativa. La hipótesis de Born-Oppenheimer es fundamental para poder escribir la energía potencial en función de las coordenadas de los átomos ya que permite desacoplar el movimiento nuclear del electrónico por su diferencia de velocidades, siendo el movimiento nuclear significativamente más lento.

La energía potencial que gobierna el movimiento de las partículas puede escribirse como la suma de dos tipos de contribuciones: de enlace (enlaces, ángulos, diedros) y de no enlace (interacciones de Lennard-Jones y Coulomb), según:

$$
\begin{aligned}
V = & \sum_{\text{enlaces}} k_l (l - l_0)^2 \\
& + \sum_{\text{ángulos}} k_\theta (\theta - \theta_0)^2 \\
& + \sum_{\text{diedros}} \frac{V_n}{2} [1 + \cos(n\phi - \gamma)] \\
& + \sum_{i<j}^{\text{no enlace}} \left[\frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^{6}} + \frac{q_i q_j}{4\pi\varepsilon_0 r_{ij}} \right]
\end{aligned}
\tag{2.7}
$$

A continuación y en la Figura 2.2 se detalla cada término de la Ecuación 2.7 junto con sus variables y parámetros. Se conoce el conjunto de parámetros utilizados en la descripción de la energía potencial como un campo de fuerza.

2.2.3.1 Contribuciones de enlace

Las contribuciones de enlace están representadas en los tres primeros términos de la Ecuación 2.7. El primero corresponde a la energía asociada a cada par de átomos enlazados, modelada mediante un potencial armónico en función de la distancia l respecto a su valor de equilibrio l_0. El segundo término describe la energía angular entre tres átomos conectados, también representada por un potencial armónico centrado en el ángulo de equilibrio θ_0. El tercer término incluye la energía de torsión entre cuatro átomos consecutivos, expresada como una función periódica del ángulo diedro ϕ, con parámetros que controlan la amplitud de la barrera torsional Vn, la periodicidad n, y el desfase γ. Este ángulo diedro se define como el ángulo entre los planos formados por los tres primeros átomos (1, 2 y 3) y los tres últimos (2, 3 y 4), es decir, el ángulo entre los planos 1–2–3 y 2–3–4 (Figura 2.2).

Los parámetros que definen el campo de fuerza en estas tres contribuciones son: las constantes de fuerza k_l y k_θ, las posiciones de equilibrio l_0 y θ_0, y los parámetros torsionales Vn, n y γ, que determinan la forma y altura del potencial de torsión. Adicionalmente, algunos campos de fuerza pueden incluir términos cruzados para capturar efectos de acoplamiento, como por ejemplo, entre dos distancias de enlace adyacentes y el ángulo que estas forman [11].

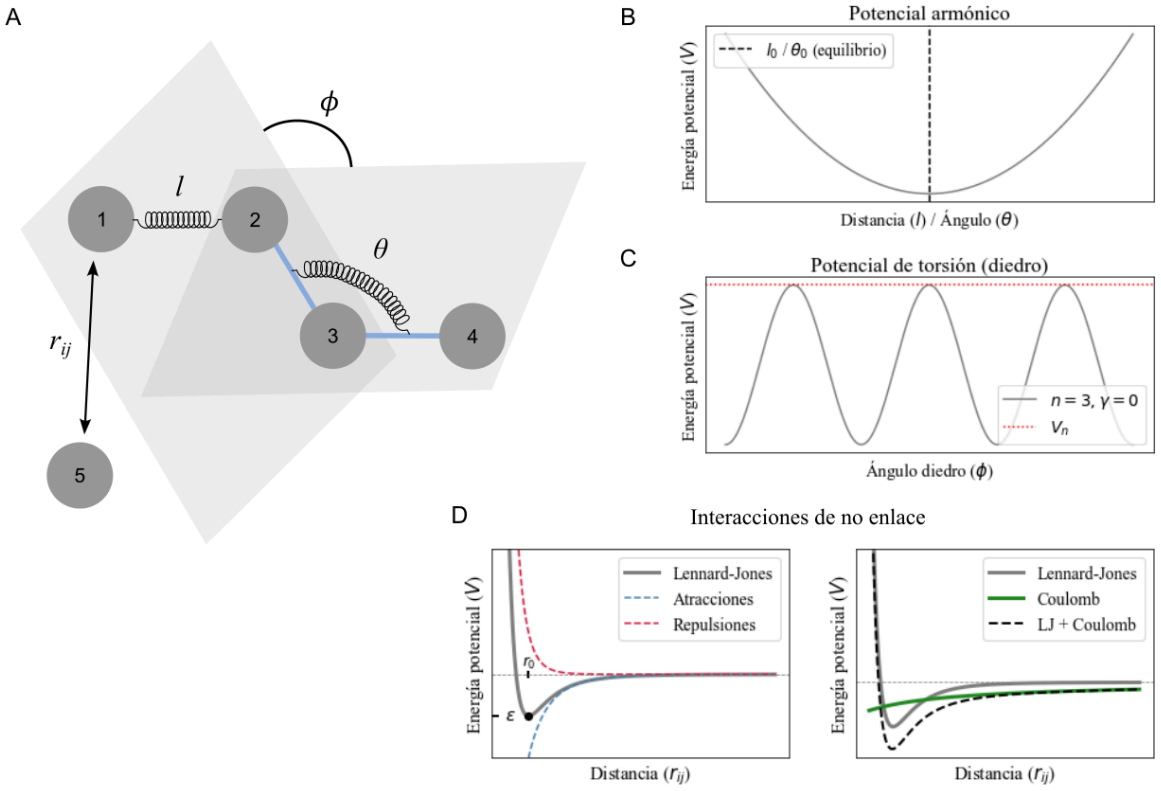


Figura 2.2: Componentes de la energía potencial clásica del sistema, según la Ecuación 2.7. **A**) Representación del enlace (l) entre dos átomos o el ángulo (θ) entre tres átomos como oscilador armónico clásico mediante resortes. Representación del ángulo diedro como ángulo (ϕ) entre los planos 1–2–3 y 2–3–4. Representación de las interacciones de no enlace (Van der Waals y electrostáticas) entre los átomos 1 y 5 con una flecha doble que se encuentran a una distancia r_{ij}. Figura adaptada de [7] **B**) Forma funcional del potencial armónico en función de la distancia/ángulo. Una línea discontinua vertical indica la posición de equilibrio (l_0 ;θ_0) que corresponde al mínimo de energía potencial. **C)** Forma funcional del potencial que modela el comportamiento de los ángulos diedros con una periodicidad $n = 3$ y un desfasaje $\gamma = 0$. Una línea horizontal punteada roja marca el valor máximo del potencial Vn. **D)** El gráfico de la izquierda muestra con líneas discontinuas las componentes atractiva (azul) y repulsiva (rojo) del potencial de Lennard-Jones que se muestra en gris. Sobre el eje x se encuentra señalada la distancia correspondiente al mínimo de energía potencial (r_0) y sobre el eje y, la profundidad del pozo (ε). A la derecha se muestra nuevamente la misma curva gris de la izquierda correspondiente al potencial de Lennard-Jones, superpuesta con el potencial electrostático o de Coulomb (verde) y la suma de ambas funciones (línea discontinua negra).

2.2.3.2 Contribuciones de no enlace

Las contribuciones de no enlace están representadas en el cuarto término de la Ecuación 2.7. Estas incluyen dos componentes fundamentales: el potencial de Lennard-Jones, que modela las interacciones de Van der Waals, y el potencial coulómbico, que representa las interacciones electrostáticas entre pares de partículas separadas por tres o más enlaces covalentes.

En el caso de las interacciones de Van der Waals, el modelo de Lennard-Jones 12–6 es ampliamente utilizado. Esta función incluye un término repulsivo proporcional a r^{-12} y un término atractivo proporcional a r^{-6} siendo r la distancia entre el par de partículas. El término repulsivo describe la repulsión estérica de corto alcance, aunque no tiene una base teórica estricta; se elige por conveniencia matemática, ya que es el cuadrado del exponente 6. Por su parte, el término atractivo, que representa las interacciones de dispersión de largo alcance, sí puede derivarse de la mecánica cuántica a partir de interacciones dipolares inducidas.

Los parámetros necesarios para definir este potencial son: la profundidad del pozo de energía potencial ε y la distancia a la cual la energía se anula r_0. Para un sistema con N tipos de átomos, serían necesarios N(N−1)/2 conjuntos de parámetros distintos para cada interacción entre pares. Para evitar una parametrización exhaustiva, se emplean reglas de combinación, que permiten derivar los parámetros entre diferentes tipos de átomos a partir de los valores definidos para cada uno por separado. Estos parámetros pueden obtenerse ajustando propiedades experimentales, como las energías de vaporización o sublimación, con el objetivo de que el campo de fuerza reproduzca correctamente el comportamiento termodinámico del sistema.

Para las interacciones electrostáticas, los parámetros a definir son las cargas parciales de cada átomo-partícula (q_i). Es importante recordar que la carga parcial no es una propiedad experimentalmente medible, sino que es una discretización de la distribución electrónica continua o potencial electrostático.

2.2.3.3 Condiciones periódicas de contorno

Para sistemas biomoleculares, se colocan las biomoléculas en el centro de una caja solvente, por lo que las dimensiones deben mantenerse lo más reducidas posibles para mantener un costo computacional razonable. Por otro lado, se debe permitir que haya suficientes moléculas de solvente alrededor del sistema biomolecular para garantizar un modelado realista de la “solución acuosa”. Las formas más comunes para definir la caja de solvente son la cúbica y la de octaedro truncado. Esta última resulta más eficiente para sistemas aproximadamente esféricos, como los considerados en esta tesis, y permite reducir la cantidad de moléculas de solvente necesarias, optimizando los recursos computacionales.

El hecho de que la caja sea reducida, implica que una proporción significativa de las partículas se ubique cerca de los bordes, donde el entorno es distinto al que experimentan las partículas del seno del sistema y esta asimetría podría afectar su comportamiento y dar lugar a artefactos en la simulación. Para evitar este problema sin aumentar significativamente el tamaño del sistema, se utilizan condiciones periódicas de contorno, en las que la celda se replica en todas las direcciones espaciales. Así, las partículas interactúan con sus vecinos más cercanos, ya sea porque pertenecen a la caja original o a una imagen adyacente. Las

condiciones periódicas de contorno implican que cuando una partícula abandona la caja, se reemplaza por una imagen que ingresa por el lado opuesto, y sus coordenadas se ajustan sumando o restando múltiplos enteros de los vectores de celda.

Las interacciones de no enlace representan un desafío adicional al considerar condiciones periódicas de contorno ya que la cantidad de este tipo de interacciones se vuelve infinita. Una forma de limitar la cantidad de interacciones contempladas podría ser restringirse a una sola celda adyacente en cada dirección. Otra forma, que es la más utilizada para resolver esto, es establecer un radio de corte (*cutoff*) a partir de las cuales las interacciones se consideran nulas. Esta aproximación es válida para el potencial de Lennard-Jones, que describe fuerzas de corto alcance que decaen rápidamente con la distancia. Sin embargo, las interacciones electrostáticas, no pueden truncarse abruptamente sin generar errores en la energía total y afectar la estabilidad de la simulación. Como este efecto es más relevante cuanto más alcance tenga la interacción considerada, las interacciones electrostáticas que son de mayor alcance, requieren un tratamiento más preciso que simplemente truncar a un cierto valor de distancia.

Para sobrellevar este inconveniente de manera eficiente y precisa se utiliza el método de sumas de Ewald, que transforma el cálculo de las interacciones electrostáticas (suma infinita con convergencia muy lenta y condicionalmente convergente) en dos componentes, como se explica brevemente a continuación. Un componente reside en el espacio real, en donde se consideran solo las interacciones cercanas, suavizando las cargas puntuales mediante distribuciones gaussianas que permiten truncar la suma sin perder tanta precisión. El otro componente consiste en la corrección necesaria para recuperar la interacción original y se calcula en el espacio recíproco, que es el espacio de Fourier, donde las interacciones de largo alcance se evalúan eficientemente mediante la transformada rápida de Fourier (FFT). Esta técnica, conocida como *particle mesh Ewald* (PME) [12], es indispensable para asegurar una evaluación eficiente de las interacciones electrostáticas de largo alcance, manteniendo la conservación de energía y optimizando el costo computacional.

2.2.3.4 Campos de fuerza y sus parámetros

Un campo de fuerza es un conjunto de parámetros que define una función de energía potencial, como la presentada en la Ecuación 2.7, empleada para modelar la energía de un sistema molecular. Se conoce como proceso de parametrización al ajuste de estos parámetros para reproducir propiedades estructurales y dinámicas conocidas de moléculas o sistemas similares. Una vez obtenido el campo de fuerzas, su utilidad radica en la extrapolación de los parámetros a nuevos sistemas, distintos a los que se utilizaron en la parametrización, para estudiar su comportamiento.

Determinar estos parámetros es un proceso complejo incluso para moléculas pequeñas, debido a la diversidad de interacciones involucradas. Entre los parámetros más críticos, y por lo tanto susceptibles a mejoras, se encuentran los que describen las interacciones no enlazantes (la de van der Waals y las electrostáticas) y los términos torsionales (diedros), ya que influyen notablemente en la conformación y dinámica del sistema.

Dado que muchas veces no se dispone de datos experimentales suficientes para validar estos parámetros, se recurre a cálculos de mecánica cuántica de alta precisión. Un enfoque habitual consiste en optimizar la geometría molecular, especificando enlaces, ángulos y torsiones,

resolviendo la ecuación de Schrödinger con la mejor aproximación posible (ver Sección 2.2.4) para obtener la configuración de mínima energía. Los métodos más populares para obtener cargas parciales, como por ejemplo el llamado RESP (*restrained electrostatic potential fit*, por sus siglas en inglés), consisten en minimizar iterativamente la diferencia entre el potencial electrostático "real" calculado a partir de la función de onda del sistema y el generado por los valores asignados a las cargas parciales. En particular, para biomoléculas, al no poder acceder a las propiedades cuánticas del sistema completo, dado su tamaño y el consecuente costo computacional, lo que se hace es calcular las cargas parciales para fragmentos del sistema, como por ejemplo, péptidos de dos o tres aminoácidos. Otra consideración tiene que ver con que la distribución de la carga es que la misma varía con los cambios conformacionales, por lo que el valor final de la carga a utilizar puede provenir de una promedio pesado que contemple las frecuencias poblacionales calculadas según la distribución de Boltzmann [11].

Para asegurar la transferibilidad de los parámetros, no basta con identificar el elemento químico de cada átomo, sino que es necesario definir su "tipo", considerando su hibridación y entorno químico. Por ejemplo, los carbonos sp^2 y sp^3 requieren parámetros distintos debido a sus geometrías y características electrónicas. En el caso de biomoléculas, es fundamental especificar el estado de protonación de residuos ácidos (como aspártico y glutámico) y básicos (como lisina y arginina), así como la presencia de puentes disulfuro, ya que estos enlaces y estados se mantienen fijos durante las simulaciones clásicas. También deben fijarse los estados tautómeros y de protonación de histidinas y otras moléculas con variantes estructurales. Estas simplificaciones, necesarias para garantizar la estabilidad y eficiencia computacional, representan una limitación inherente al modelo clásico, ya que no permiten que ocurran cambios dinámicos en estados de protonación o estados tautoméricos a lo largo de la simulación.

Uno de los modelos más simples para las moléculas de agua es el TIP3P (por sus siglas en inglés, *transferable intermolecular potential with 3 points*) [13], en el cual se utilizan tres partículas en una geometría rígida, con cargas parciales asignadas a cada átomo. Modelos más elaborados, como TIP4P, introducen una cuarta partícula con carga negativa desplazada del átomo de oxígeno, mejorando la descripción electrostática. Sin embargo, el modelo TIP3P es más utilizado principalmente por compatibilidad con campos de fuerza desarrollados originalmente con este modelo, además de su menor costo computacional. La parametrización de estos modelos se orienta a reproducir propiedades estructurales y termodinámicas del agua, como la densidad, la función de distribución radial y la capacidad calorífica. En simulaciones de sistemas bioquímicos en solución acuosa, es fundamental emplear el mismo modelo de agua con el que se parametrizó el sistema.

En esta tesis, para la descripción de los sistemas proteicos, se empleó el campo de fuerza ff14SB de Amber, que representa una evolución significativa respecto a ff99SB, su antecesor, al incorporar mejoras en los términos diedros tanto del esqueleto proteico como de las cadenas laterales. Estas mejoras se basan en cálculos de mecánica cuántica y validaciones mediante simulaciones de dinámica molecular, lo que permite una representación más precisa de la estructura y dinámica proteica [14]. Las correcciones ILDN, que ajustan específicamente las torsiones de isoleucina, leucina, aspartato y asparagina, mejoran la estabilidad y el comportamiento conformacional de estos residuos.

Para ácidos nucleicos, Amber cuenta con campos de fuerza específicos que refinan los parámetros de los grupos fosfato y torsiones nucleotídicas. El modelo más actualizado que ofrece Amber fue desarrollado por Aytenfisu et al. [15], quienes introdujeron correcciones puntuales en las torsiones del ARN, mejorando desviaciones estructurales observadas en simulaciones anteriores, así como la fidelidad dinámica. Es importante destacar que estos campos de fuerza están diseñados para integrarse de forma coherente con los parámetros para proteínas de Amber y con los modelos de agua mencionados, permitiendo simulaciones realistas de sistemas biomoleculares complejos.

Además de los campos de fuerza de Amber, existen otros campos de fuerza ampliamente utilizados como CHARMM (siglas en inglés de *Chemistry at HARvard using Molecular Mechanics*), OPLS (siglas en inglés de *Optimized Potentials for Liquid Simulations*) y GROMOS (siglas en inglés de *GROningen MOlecular Simulation*). Cada uno presenta particularidades en su enfoque de parametrización, en la representación de ciertas interacciones y en su compatibilidad con diferentes modelos de agua o moléculas [16].

2.2.4 Representación cuántica de un sistema

Para estudiar procesos que implican la formación o ruptura de enlaces químicos, los modelos clásicos resultan inadecuados debido a su representación armónica de los enlaces covalentes, que impide describir cambios en la conectividad química. Por eso, es necesario recurrir a métodos cuánticos capaces de describir la densidad electrónica y sus variaciones durante una reacción. Entre los enfoques más relevantes para estudiar este tipo de procesos se destacan los métodos de primeros principios o *ab initio*, la Teoría del Funcional de la Densidad (DFT, por sus siglas en inglés) y métodos semiempíricos, cada uno con diferentes niveles de precisión y costo computacional.

La base teórica fundamental de estos métodos es la ecuación de Schrödinger [17], que describe el comportamiento cuántico de sistemas físicos a nivel atómico y molecular. En su forma independiente del tiempo, esta ecuación se presenta como un problema de autovalores:

$$\widehat{H}\Psi = E\Psi \,, \tag{2.8}$$

donde es $\widehat{H}$ el operador Hamiltoniano que refleja las contribuciones cinética y potencial de la energía, Ψ representa la función de onda que describe el estado cuántico de un sistema, y E es la energía asociada con ese estado. Resolver esta ecuación permite obtener las funciones de onda y las energías de los sistemas, que resultan útiles para comprender propiedades del sistema como su estabilidad o la densidad de probabilidad electrónica.

2.2.4.1 Métodos *ab initio*

Los métodos *ab initio* se caracterizan por resolver la ecuación de Schrödinger utilizando únicamente constantes fundamentales, sin incorporar parámetros empíricos ajustados. Sin embargo, a partir de la ecuación de Schrödinger sólo se obtienen soluciones exactas en sistemas muy simples, como los átomos hidrogenoides, que consisten en un núcleo con una determinada carga positiva y un solo electrón. En estos casos, la ecuación puede resolverse analíticamente, proporcionando orbitales atómicos bien definidos y niveles de energía

discretos que explican la estructura electrónica y propiedades espectroscópicas de estos sistemas.

Al abordar sistemas más complejos, como átomos o moléculas con varios electrones, resolver exactamente la ecuación de Schrödinger deja de ser viable, por lo que deben realizarse aproximaciones. Para el caso de las moléculas, se aplica la teoría de orbitales moleculares, que consiste en construir estos orbitales como combinaciones lineales de orbitales atómicos. Esta teoría constituye la base de métodos computacionales como Hartree-Fock (HF). Este método introduce una aproximación crucial para tratar sistemas con múltiples electrones: considera que cada electrón se mueve en un campo efectivo promedio generado por los demás electrones, lo que simplifica el problema al evitar el tratamiento explícito de la interacción electrónica instantánea, que es lo que impide la resolución exacta de la ecuación de Schrödinger. Este enfoque se basa en un procedimiento autoconsistente, donde los orbitales moleculares se ajustan iterativamente hasta que la energía total del sistema converge a un valor mínimo, garantizando un límite superior para la energía gracias al teorema variacional.

El teorema variacional establece que, para cualquier función de onda “de prueba”, el valor esperado de la energía calculado con el operador Hamiltoniano es siempre mayor o igual que la energía real del estado fundamental del sistema. Esto significa que al variar los parámetros de una función de onda aproximada, la energía obtenida nunca será menor que la energía verdadera, y el objetivo es encontrar la función de onda que minimice este valor esperado, ya que esto permite acercarse lo más posible a la energía real. Este principio proporciona un criterio sólido para optimizar funciones de onda aproximadas y garantiza que el método HF establezca un límite superior para la energía del sistema.

Si bien el método HF mejora la calidad de las funciones de onda calculadas, solo considera la correlación electrónica de forma promedio, sin modelar las fluctuaciones instantáneas en la posición y el movimiento de los electrones, lo que limita la precisión del método. Para superar esta limitación, se desarrollaron métodos post-HF, como la perturbación de Møller–Plesset de segundo orden (MP2) y la Interacción de Configuraciones (CI), que incorporan correcciones para la correlación electrónica mediante técnicas más sofisticadas, aunque a costa de un considerable aumento en el costo computacional. Por esta razón, los métodos *ab initio*, especialmente estos últimos que incluyen correlación electrónica, se aplican principalmente a sistemas pequeños donde el costo computacional es manejable.

2.2.4.2 Método DFT y derivados semiempíricos

La Teoría del Funcional de la Densidad (DFT, Density Functional Theory) reformula el problema cuántico expresando la energía electrónica E en términos de la función densidad electrónica (ρ), lo que reduce considerablemente el costo computacional e incluye efectos de correlación electrónica. La expresión general para la energía electrónica total es:

$$E[\rho] = T[\rho] + V_{ext}[\rho] + E_{Hartree}[\rho] + E_{XC}[\rho] \,, \tag{2.9}$$

donde T es la energía cinética, V_{ext} es la interacción con el potencial externo, E_{Hartee} es la energía de Coulomb, y E_{XC} es la energía de intercambio-correlación. En este sentido, DFT puede entenderse como un procedimiento variacional alternativo a la solución directa de la

ecuación de Schrödinger, en el que se minimiza un funcional de la energía electrónica respecto a la densidad electrónica del sistema en lugar de hacerlo respecto a la función de onda. Según el principio variacional aplicado en DFT, la energía calculada para una densidad electrónica propuesta siempre será mayor o igual que la energía real del estado fundamental, y la densidad correcta es aquella que minimiza este funcional de energía.

La precisión de DFT depende de la elección del funcional de intercambio-correlación $E_{XC}[\rho]$, ya que los demás términos cuentan con expresiones analíticas precisas. La forma más sencilla de aproximar el funcional de intercambio-correlación es la aproximación de densidad local (LDA), que se basa en suponer que la densidad electrónica es constante en todo el espacio, aunque el valor que toma esta constante va cambiando en cada elemento de volumen. Para hacer estos cálculos, se emplean métodos numéricos que dividen el espacio tridimensional en una grilla. Para mejorar esto, se pueden incorporar funcionales no locales que dependen del gradiente de la densidad y ya no solamente de su valor en un punto. De esta manera, los funcionales GGA (Generalized Gradient Approximation) mejoran la aproximación local (LDA, Local Density Approximation) al considerar no solo el valor local de la densidad electrónica, sino también su gradiente espacial, permitiendo describir al sistema molecular con mayor realismo. Ejemplos comunes de funcionales GGA son PBE (Perdew–Burke–Ernzerhof) y BLYP (Becke–Lee–Yang–Parr). Además, existen funcionales híbridos como B3LYP (Becke, 3-parameter, Lee-Yang-Parr), que combinan una fracción del intercambio exacto de Hartree-Fock con términos de intercambio y correlación de funcionales GGA. Esta mezcla incrementa la precisión en el cálculo de propiedades electrónicas y energías de activación, aunque con un costo computacional algo mayor que los GGA puros.

Para reducir aún más el costo computacional, y de esta manera poder acceder al estudio de sistemas más complejos, se desarrolló el método semiempírico DFTB (Density Functional Tight Binding), en el que se parametrizan ciertas integrales, evitando así algunos cálculos costosos computacionalmente. En su versión auto-consistente (SCC-DFTB o DFTB2), permite calcular cargas atómicas y describir interacciones electrostáticas con una precisión cercana a la de DFT, con un menor costo computacional. La eficiencia se logra a partir del uso de una densidad de referencia, que se calcula como una superposición de densidades atómicas neutras, lo cual localiza la densidad electrónica en cada átomo y, a lo sumo, considera a los primeros vecinos. En su última versión, DFTB3 [18], se hace una expansión de la energía de DFT en series de Taylor respecto a la densidad electrónica, y se trunca en tercer orden. Si bien el método general DFTB es especialmente útil para el estudio de sistemas grandes, como las biomoléculas, su precisión depende de la calidad de las parametrizaciones disponibles para los distintos tipos de átomos.

2.2.5 Representación híbrida del sistema

La descripción cuántica de un sistema representa con mayor detalle el comportamiento de las partículas, ya que incorpora propiedades como la distribución electrónica, a diferencia de la descripción clásica, que utiliza modelos mucho más simplificados. Sin embargo, este nivel de detalle incrementa notablemente la cantidad de variables necesarias para los cálculos de dinámica, lo que eleva de forma significativa el costo computacional. Por este motivo, resulta inviable aplicar una descripción cuántica completa a sistemas de gran tamaño, como una proteína entera.

Ante esta limitación, en ciertos sistemas biomoleculares, como las enzimas, se recurre a estrategias que combinan distintas escalas de modelado. En estos casos, las reacciones químicas ocurren en el sitio catalítico, donde es crucial conocer con precisión la densidad electrónica y sus variaciones, mientras que en el resto de la proteína los enlaces pueden considerarse invariables. Para ello, se emplean enfoques híbridos que integran descripciones cuánticas y clásicas: la región catalítica (incluyendo aminoácidos, iones y otras moléculas cercanas) se modela cuánticamente, mientras que el resto del sistema se trata mediante mecánica clásica.

El enfoque QM/MM (*Quantum Mechanics/Molecular Mechanics*, por sus siglas en inglés) surge como una solución eficaz para estudiar sistemas moleculares complejos. Matemáticamente, esto se refleja en un Hamiltoniano total que suma los términos de energía de la región QM (H_{QM}), la región MM (H_{MM}) y un término adicional de acoplamiento ($H_{QM\text{-}MM}$) que describe la interacción entre ambas:

$$H_{QM/MM} = H_{QM} + H_{MM} + H_{QM-MM}\,, \tag{2.10}$$

donde el término de acoplamiento $H_{QM\text{-}MM}$ es crucial porque conecta dos descripciones físicas diferentes. En la práctica, el programa Amber (muy utilizado en general y en particular en los trabajos realizados para esta tesis) resuelve este desafío mediante el esquema de acoplamiento electrostático [19], incluyendo las cargas puntuales de la región MM en el Hamiltoniano cuántico, de modo que la densidad electrónica de la región QM se polariza por el entorno clásico, logrando así una interacción realista y eficiente entre ambas regiones. Este enfoque es el más habitual y recomendado, ya que equilibra precisión y costo computacional.

En el término de acoplamiento entre las regiones QM y MM se tratan de manera diferente las interacciones de enlace que las de no enlace. Las interacciones de no enlace (electrostática y Van der Waals), se tratan empleando los potenciales clásicos habituales, en donde las cargas MM contribuyen al cálculo de la energía y la distribución electrónica de la región QM mediante el acoplamiento electrostático. En cambio, si existen enlaces covalentes que cruzan la frontera QM/MM, se requiere un manejo particular. Cuando un átomo de la región QM está unido covalentemente con otro en la región MM, se introduce un “átomo *link*” (usualmente un hidrógeno virtual) que completa la valencia del átomo QM en el borde, evitando artefactos electrónicos y manteniendo una descripción consistente de la química local. En el cálculo cuántico, el átomo *link* reemplaza al átomo MM, pero en la región MM se conserva la topología original, y se aplican correcciones específicas para evitar dobles conteos o discontinuidades en la energía.

2.2.6 Preparación del sistema y etapas de una simulación

Como las posiciones iniciales han sido obtenidas de manera experimental en la mayoría de los casos, es posible que no todos los átomos hayan podido resolverse. Por ejemplo, las cadenas laterales expuestas al solvente o con alta flexibilidad, no suelen estar bien definidas en las estructuras obtenidas por cristalografía de rayos X. Además, para realizar simulaciones en donde el sistema biomolecular se encuentre en un contexto más realista, se incluyen moléculas de agua y una cierta cantidad de iones alrededor de las biomoléculas para imitar el estado de solución acuosa. Dependiendo del nivel de resolución de las estructuras, puede o no haber moléculas de agua resueltas. En el caso de los experimentos de Cryo-EM, por

ejemplo, no suelen encontrarse por la baja resolución de la técnica. Por esta razón, se agregan geométricamente moléculas de agua alrededor del sistema biomolecular, de forma tal de ocupar el volumen de una caja que puede ser cúbica o en forma de octaedro truncado y la cantidad de iones indicada, repartidos homogéneamente en la caja. Los detalles de este procedimiento para los sistemas estudiados se encuentran en los métodos específicos de los Capítulos 3, 4 y 5.

Antes de comenzar a realizar una simulación computacional de biomoléculas a partir de las posiciones iniciales conocidas y las velocidades iniciales asignadas, es necesario minimizar la energía del sistema, optimizando su geometría, para partir desde un mínimo de la energía potencial. Este paso es importante ya que las posiciones iniciales de los átomos que se agregan mediante un algoritmo como se explicó en el párrafo anterior, no los ubica de manera favorable energéticamente. Además, como las condiciones experimentales de cristalización como la temperatura, la presión, la concentración, entre otras, no son las condiciones de interés para una simulación de biomoléculas en solución acusa, una vez que el sistema se relajó con la minimización de energía, se requerirá llevarlo a condiciones realistas de presión y temperatura, en los casos estudiados, 300 K y 1 bar.

Tanto en la etapa de minimización como en las etapas de calentamiento y equilibración, es posible incluir restricciones sobre parte del sistema para hacer que los cambios sobre esas regiones sean menos bruscos. Por ejemplo, puede minimizarse la energía del sistema permitiendo total movimiento de las moléculas de agua y los iones, dejando restringidos los demás átomos o puede aumentarse la temperatura permitiendo que todos los átomos se muevan pero restringir los átomos de la proteína para que no puedan alejarse demasiado de su posición inicial. Las restricciones se logran términos a la energía potencial que incluyen potenciales armónicos con una constante de fuerza que varía según la etapa o paso, mediante la cual se establece la fuerza de la restricción: a mayor constante, mayor restricción y por lo tanto, menor movimiento; y también involucran una “posición de referencia” sobre la cual restringir el movimiento o estructura de referencia que para esta tesis siempre fue la estructura inicial que se usa para comenzar la etapa o paso.

2.2.6.1 Minimización de la energía

La energía potencial es función de las posiciones de todos los átomos del sistema, por lo tanto, es función de 3N variables. La minimización de la energía del sistema requiere entonces un algoritmo que evalúe hacia donde deben moverse los átomos para lograr que ese movimiento resulte en una disminución de la energía. En el programa Sander del paquete Amber, ampliamente empleado en esta tesis, están implementados varios algoritmos de minimización, entre ellos los más usados para simulaciones de dinámica molecular clásica son *steepest descent* y *conjugate gradient*. Para otro tipo de simulaciones como las QM/MM, se recurre a otros algoritmos (Ver Capítulo 4, Sección 4.3). Por las características de estos algoritmos, en primer lugar se realizan una cierta cantidad de pasos con el algoritmo *steepest descent* que permite guiar al sistema rápidamente hacia energías menores; y en segundo lugar, una vez más cerca del mínimo, se cambia el algoritmo por *conjugate gradient* que, a pesar de ser más costoso computacionalmente, es más eficiente en el refinamiento de la estructura para alcanzar menores energías.

Al dar pasos en la búsqueda de menores energías, deben definirse dos valores: el tamaño del paso, en términos de distancia que se moverán los átomos y que no debe ser demasiado grande para evitar cambios bruscos y estructuras poco realistas ni tampoco muy pequeño porque esto aumentaría demasiado el costo computacional (en el caso del programa Sander, el valor por defecto es 0.01 Å [20]); y el criterio de convergencia, mediante el cual se compara la energía obtenida en el paso anterior con el actual y se evalúa si la diferencia es menor al valor elegido en el criterio (en el caso del programa Sander, el valor por defecto es 10^{-4} kcal·mol^{-1}·Å^{-1} [20]).

2.2.6.2 Calentamiento o termalización

Como el calentamiento implica un aumento en la energía, es conveniente realizar este proceso de manera progresiva para evitar estructuras inestables y poco probables. Para lograr esto, se implementan incrementos de temperatura en pasos sucesivos que en las primeras etapas, en general incluyen restricciones sobre los átomos de las biomoléculas y ligandos del sistema, mientras que se permite el movimiento completo de las moléculas de agua y los iones. A medida que el sistema se adapta, estas restricciones se eliminan de manera gradual, hasta que finalmente se deja al sistema evolucionar libremente. En general, la termalización suele hacerse a volumen constante.

Dado que la temperatura de un sistema está directamente relacionada con la energía cinética promedio ($< E_c >$) de sus N partículas, según el teorema de equipartición de la energía:

$$< E_c >= \sum_{i=1}^{N} \frac{1}{2} m_i v_i^2 = \frac{3}{2} k_B T \ , \tag{2.11}$$

donde m_i es la masa de la partícula i y k_B es la constante de Boltzmann, modificar la temperatura (T) implica ajustar las velocidades de las partículas (v_i).

Para mantener la temperatura constante durante la simulación, uno de los termostatos más usados es el de *Langevin* [21], un modelo estocástico que simula el efecto de un baño térmico permitiendo el intercambio eficiente de energía con el sistema, lo que facilita alcanzar y mantener la temperatura deseada durante la simulación. Para aplicar este efecto, se modifica la Ecuación 2.1, agregando un término de fricción y un término de fluctuaciones aleatorias por choques estocásticos:

$$\frac{d^2 \bar{r}_i}{dt^2} = \frac{\bar{F}_i}{m_i} - \gamma m_i \frac{d \bar{r}_i}{dt} + \bar{R}_i(t) \ , \tag{2.12}$$

donde γ es la frecuencia de colisiones, que regula la intensidad del acoplamiento térmico y $\bar{R}_i(t)$ es la fuerza aleatoria que obedece al teorema fluctuación-disipación. Los valores típicos utilizados para la frecuencia de colisiones en este tipo de análisis oscilan entre 1 y 5 ps^{-1}.

2.2.6.3 Equilibración de la presión

Luego de alcanzar la temperatura final, que para esta tesis fue de 300 K, se procede a realizar la equilibración de la presión. Manteniendo la temperatura constante, se deja que el

volumen de la caja se ajuste a una presión constante, en el ensamble NPT (número de partículas, la presión y la temperatura constantes), controlando que la densidad llegue a un valor cercano a 1 g/cm^3 sin ningún cambio brusco en la energía y las coordenadas del sistema.

Para lograr este control de presión en las simulaciones, se emplean algoritmos denominados barostatos, que ajustan el volumen de la caja simulada con el objetivo de alcanzar y mantener la presión deseada. Un barostato ampliamente utilizado es el de Berendsen [22], que acopla el sistema a un baño de presión externa, según:

$$\frac{dP}{dt} = \frac{\left(P_0 - P\right)}{\tau_P}, \tag{2.13}$$

donde P_0 es la presión de referencia a la cual se quiere llevar el sistema y τ_P es la constante de acoplamiento de la presión, que típicamente toma valores entre 0,5 y 5 ps. Cuanto más grande sea la constante, más suave será el acoplamiento, ya que funciona como un tiempo característico del ajuste. A cada paso de integración, el volumen se escala de manera isotrópica, es decir, aplicando el mismo factor en las tres direcciones espaciales, lo que implica también un escalado proporcional de las coordenadas atómicas.

En simulaciones de dinámica molecular de sistemas en fase condensada, como biomoléculas en solución acuosa, la diferencia entre realizar la simulación a volumen constante (NVT) o a presión constante (NPT) suele ser mínima, ya que las fluctuaciones de volumen a presión constante son muy pequeñas y, en la práctica, ambas condiciones producen resultados muy similares para la mayoría de las propiedades estructurales y dinámicas. Por eso, es común utilizar el ensamble NPT solo durante la etapa de equilibración, permitiendo que el sistema equilibre su densidad, aproximadamente 1 g/cm^3, y volumen de equilibrio para 1 bar de presión. Una vez alcanzado ese equilibrio, se continúa la simulación en el ensamble NVT, donde el volumen permanece fijo.

2.2.6.4 Equilibración final y producción

Para las simulaciones realizadas para esta tesis, la última etapa de equilibración se lleva a cabo en las mismas condiciones NVT que la etapa de producción. Por lo tanto, la equilibración final consiste en descartar la porción inicial de la trayectoria (típicamente entre el 10 y el 20%) en donde el sistema evoluciona hasta alcanzar el equilibrio termodinámico, que suelen determinarse de manera arbitraria, monitoreando propiedades como el RMSD (ver Sección 2.3). Una vez alcanzado el equilibrio, la secuencia de configuraciones o imágenes obtenidas se conoce como la trayectoria asociada a la producción, que es la base de datos que servirá para el análisis posterior.

Cuando se simula durante un tiempo suficientemente largo, se dice que una simulación de dinámica molecular ha convergido si las propiedades físicas o estructurales de interés alcanzan un comportamiento estable y representativo del equilibrio del sistema. En otras palabras, esto indica que se ha muestreado de forma adecuada el espacio conformacional relevante y que magnitudes como la energía, el RMSD o distintos parámetros estructurales fluctúan alrededor de valores constantes. Un estudio reciente muestra que muchas propiedades estructurales y dinámicas de interés biológico tienden a estabilizarse en escalas

de tiempo del orden de los microsegundos. En cambio, otras propiedades, como las tasas de transición hacia conformaciones poco frecuentes, pueden requerir trayectorias mucho más largas para alcanzar la convergencia. Esta diferencia se debe a que distintos mecanismos moleculares presentan escalas de tiempo muy variadas [23].

En sistemas grandes o muy flexibles, alcanzar la convergencia puede ser particularmente desafiante debido a la gran cantidad de configuraciones accesibles y la complejidad del paisaje energético. Por eso, aunque un sistema puede no estar en equilibrio físico completo, desde un punto de vista funcional se considera que ha alcanzado el equilibrio cuando los parámetros de análisis relevantes para la función biológica que se quiere estudiar han convergido. Esta perspectiva se enmarca dentro del enfoque que vincula estructura, dinámica y función en el estudio de biomoléculas [23]. Determinar el tiempo total de simulación y cuánto descartar como equilibración es complejo y no tiene un criterio único, ya que depende del sistema y de las propiedades analizadas, por lo que se requiere un análisis cuidadoso para asegurar resultados confiables.

2.3 Análisis de la trayectoria

2.3.1 Alineamiento, medidas de desvíos y de fluctuaciones

La obtención de la trayectoria asociada a la producción, significa contar con los datos de las posiciones de todas las partículas en función del tiempo, a partir de las cuales se busca extraer información relevante. En cambio, si se visualiza la trayectoria con un software apropiado como puede ser VMD (*Visual Molecular Dynamics*, por sus siglas en inglés) [24], es posible observar el movimiento colectivo del sistema a lo largo del tiempo.

Una manera habitual de comparar cuantitativamente diferentes configuraciones de la trayectoria consiste en calcular el desvío cuadrático medio o RMSD (del inglés, *Root Mean Square Deviation*), que mide la desviación cuadrática media de las posiciones atómicas respecto de una conformación de referencia:

$$RMSD(t) = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left|\overline{r_i}(t) - \overline{r_i^{ref}}\right|^2}\ , \qquad (2.14)$$

donde t se refiere a la configuración de la trayectoria en un tiempo determinado, N es la cantidad de átomos incluidos en el cálculo (típicamente se usan los átomos del esqueleto de la proteína), $\overline{r_i}(t)$ son las coordenadas (x, y, z) del átomo i en el tiempo t considerada y $\overline{r_i^{ref}}$ son las coordenadas (x, y, z) del mismo átomo en la configuración de referencia.

Durante la trayectoria de dinámica molecular, el sistema experimenta movimientos globales de traslación y rotación dentro de una caja, por un lado y a través de ella debido a las condiciones periódicas de contorno, por otro lado. En este último caso, dado que las moléculas pueden cruzar las fronteras de la caja y reaparecer en el lado opuesto, es deseable que las posiciones atómicas sean ajustadas previo al análisis, y devueltas a la caja principal. Muchos de los cambios estructurales internos relevantes, requieren eliminar los movimientos globales mencionados. Esto se logra mediante la alineación de cada configuración de la

trayectoria con respecto a una configuración de referencia (usualmente el inicial), aplicando una rotación y traslación óptimas que minimizan el RMSD global entre las configuraciones. El algoritmo de alineación típicamente usado es el de Kabsch [25]. Esta minimización del RMSD permite independizar el análisis de las fluctuaciones internas de la molécula de los movimientos globales.

Es importante aclarar que el resultado del RMSD depende del conjunto de átomos utilizado para alinear la trayectoria. En un sistema proteico, si bien una opción común es alinear considerando todos los aminoácidos, también puede hacerse con un subconjunto que excluya regiones móviles (como *loops* flexibles o extremos terminales) que podrían sesgar el análisis.

Otra métrica útil para hacer un análisis cuantitativo y caracterizar la trayectoria es el cálculo de la fluctuación por átomo o RMSF (del inglés, *Root Mean Square Fluctuation*), que mide la fluctuación promedio de la posición de cada átomo respecto a su posición promedio a lo largo de la trayectoria de simulación:

$$RMSF(i) = \sqrt{\frac{1}{T}\sum_{t=1}^{T}\left|\overline{r}_i(t) - <\overline{r}_i>\right|^2}, \qquad (2.15)$$

donde i se refiere al átomo considerado, T es la cantidad de imágenes o configuraciones de la trayectoria, $\overline{r}_i(t)$ son las coordenadas (x, y, z) del átomo i en la imagen t y $<\overline{r}_i>$ son las coordenadas (x, y, z) que corresponden a la posición promedio del mismo átomo a lo largo de toda la trayectoria.

Este valor cuantifica la movilidad local de átomos individuales, pero es posible combinar los cálculos de los átomos que componen cada aminoácido o nucleótido para obtener el RMSF por residuo, lo que permite identificar fácilmente regiones flexibles o rígidas dentro de una biomolécula. Para obtener el RMSF por residuo, se promedian los RMSF de los átomos que conforman ese residuo (o los especificados, como pueden ser los átomos del esqueleto solamente), ponderados por su masa.

Todos los cálculos mencionados se pueden realizar con el módulo CPPTRAJ del paquete Amber [26]. Además, este programa permite calcular parámetros que resultan muy informativos como lo son las distancias, ángulos y diedros entre dos, tres y cuatro centros de masa, respectivamente; estimar la superficie accesible al solvente para un conjunto de átomos, conocer la estructura secundaria de una proteína, implementar algoritmos de clusterización para las conformaciones, entre otros. Es posible también extraer de la trayectoria la información detallada de las interacciones, como la caracterización de los puentes de hidrógeno tanto intra como intermoleculares, e incluso se puede obtener el detalle de este tipo de interacciones con el solvente. Más aún, se puede identificar qué átomos son donores o aceptores y qué porcentaje de la trayectoria esa interacción está formada. Tanto el programa CPPTRAJ como VMD y otros, suelen determinar las interacciones de puente hidrógeno contemplando la identidad de los átomos y a partir de criterios geométricos, buscando distancias menores a 3,5 Å y ángulos cercanos a 180°.

2.3.2 Análisis de componentes principales

Los sistemas que suelen estudiarse con dinámica molecular están atravesados por la problemática de la alta dimensionalidad. Los cambios conformacionales se dan por cambios en las coordenadas de los átomos que componen una proteína y gracias a los programas de visualización como VMD, muchas veces es posible observar representaciones que son útiles para interpretar esos cambios. Si se observa un cambio en la dinámica, a simple vista no siempre resulta obvia una definición de coordenadas que muestre el cambio observado. Otra posibilidad es que a simple vista no se observe un cambio conformacional, más allá de los movimientos de "respiración" que puede tener una proteína. Para ambos casos, podría realizarse un análisis de componentes principales que tiene como objetivo encontrar las direcciones que maximizan la variabilidad de las coordenadas y de esta manera, reducir la dimensionalidad. En este contexto, puede resultar útil realizar un análisis de componentes principales (PCA, por sus siglas en inglés), cuyo objetivo es encontrar los vectores que describen las fluctuaciones más amplias de la biomolécula, bajo la hipótesis de que estos movimientos tienen alta probabilidad de ser funcionales o estar relacionados con la actividad biológica.

Para realizar un análisis de PCA, se parte de una trayectoria de dinámica molecular que describe las posiciones de los átomos que se quiere incluir (se suele hacer este análisis con los átomos del esqueleto de la proteína). Cada configuración o imagen de la trayectoria se representa como un vector $\bar{r}$ de dimensión $3N$ (con N es el número de átomos considerados, 3 son las coordenadas x, y, z). Para evitar que las diferencias globales de traslación y rotación afecten el análisis, todas las estructuras se superponen a una referencia común mediante un ajuste como la minimizando de RMSD que se describe en la sección anterior. Luego, se calcula la posición promedio global del sistema a lo largo de la trayectoria, según:

$$< \bar{r} >= \frac{1}{T} \sum_{t=1}^{T} \bar{r}(t) \, , \tag{2.16}$$

donde T es la cantidad de imágenes o configuraciones de la trayectoria y $\bar{r}(t)$ son las coordenadas (x, y, z) de todos los átomos en la imagen t. A continuación, para garantizar que en el PCA se analicen las fluctuaciones alrededor de la conformación promedio, se centran los datos restando esta media:

$$\overline{r}_c(t) = \bar{r}(t) - < \bar{r} > . \tag{2.17}$$

Una vez obtenidos los valores centrados, se procede a construir la matriz de covarianza **C** de las fluctuaciones, en donde cada elemento C_{ij} correspondiente a la covarianza entre las posiciones i y j del vector $\overline{r}_c(t)$ se calcula según:

$$C_{ij} = \frac{1}{T-1} \sum_{t=1}^{T} \overline{r}_c(t)_i \, \overline{r}_c(t)_j \, . \tag{2.18}$$

Esta matriz es de tamaño $3N \; x \; 3N$ y contiene la correlación entre los movimientos de todos los átomos seleccionados. Para obtener una estimación estadísticamente correcta de la covarianza, se usa T - 1 en lugar de T, que está relacionado con haber restado la media, es decir que se ha perdido un grado de libertad.

El paso siguiente consiste en diagonalizar la matriz de covarianza para obtener sus autovectores y autovalores. La matriz de covarianza es simétrica y positiva semidefinida (cuyos autovalores son todos mayores o iguales a cero), por lo que siempre es posible encontrar una base ortonormal de autovectores que la diagonalizan. Los autovalores se ordenan de manera decreciente porque los mayores, que son lo de interés, indican las direcciones de mayor variabilidad en el sistema. En otras palabras, los autovectores asociados a estos autovalores describen los movimientos colectivos más importantes de la proteína.

Para analizar la dinámica, se proyectan las coordenadas centradas de cada imagen de la trayectoria sobre estos autovectores, obteniendo así los coeficientes de proyección. Estos coeficientes son valores escalares que indican cuánto contribuye cada autovector o componente principal a la conformación particular en ese instante de la trayectoria. Al graficar estos valores a lo largo del tiempo, en forma de histograma o en espacios de dos o tres componentes principales, se puede visualizar y caracterizar la exploración del espacio conformacional en términos de los componentes elegidos.

2.3.3 *Bootstrapping*

El método *bootstrap* se usa para estimar la incertidumbre de una característica o valor calculado a partir de una muestra sin necesidad de suposiciones fuertes sobre la distribución poblacional. En particular, en esta tesis se utilizó este método para estimar la incertidumbre asociada al cálculo de RMSF (Capítulo 3, Sección 3.4.1). Este método se basa en un proceso de remuestreo con reemplazo de los datos originales. En lugar de asumir una distribución teórica para estimar la incertidumbre, el *bootstrap* genera múltiples muestras nuevas (llamadas muestras *bootstrap*) tomando aleatoriamente una cierta cantidad de datos de la muestra original, permitiendo que un mismo dato pueda repetirse en cada muestra. Para cada una de estas muestras se calcula el promedio, y a partir de la distribución de estos promedios se estima la variabilidad o error del promedio original. De esta forma, el bootstrap resulta una manera sencilla y robusta para cuantificar la incertidumbre sin necesidad de supuestos fuertes sobre la distribución de los datos, aprovechando la gran cantidad de datos obtenidos de la simulación computacional.

2.3.4 Función de distribución radial

La función de distribución radial $g(r)$ describe cómo varía la densidad promedio de partículas a una distancia r desde una partícula de referencia. Esta función es útil para entender el ordenamiento y la organización espacial de distintos componentes de un sistema, mostrando picos donde hay capas preferenciales de vecinos. Se calcula la función de distribución radial $g(r)$ a partir del histograma del número de átomos considerados (por ejemplo, átomos de oxígeno de las moléculas de agua) alrededor de un átomo en particular (por ejemplo, un átomo de fósforo de un ion fosfato, ver Capítulo 3, Sección 3.4.2). Por último, se normalizan los valores obtenidos a partir de la densidad calculada con las propiedades del sistema. Esto

significa que, en lugar de usar un valor fijo o asumido, la densidad ρ se calcula a partir del volumen V de la caja de solvente y el número N de moléculas de agua, según:

$$\rho = \frac{N}{V}. \tag{3.1}$$

La función de distribución radial para una dada configuración o imagen de la trayectoria se calcula entonces como:

$$g(r) = \frac{n(r)}{4\pi r^2 \Delta r \rho}, \tag{3.2}$$

donde $n(r)$ es el número de partículas encontradas en un cascarón esférico de radio r y espesor Δr (en este caso se tomó 0,5 Å) alrededor de la partícula de referencia, $4\pi r2\Delta r$ es el volumen del cascarón y ρ es la densidad.

2.3.5 Cálculo de energía libre de unión a partir de una trayectoria de dinámica molecular

La unión de un ligando a una proteína es un proceso central en la biología molecular, ya que constituye la base de numerosos mecanismos de regulación y señalización. Comprender cómo se produce este reconocimiento molecular y cómo está modulada la afinidad de la unión permite profundizar en la comprensión del funcionamiento de sistemas biológicos. De esta manera, el estudio detallado de la energía libre de unión permite identificar los determinantes estructurales y dinámicos que favorecen o dificultan la formación del complejo, y aporta información clave sobre los residuos y regiones involucrados en la interacción. Más aún, este tipo de análisis resulta especialmente valioso para investigar fenómenos alostéricos, en los cuales la unión de un ligando en un sitio específico induce cambios conformacionales que afectan la función o la afinidad de sitios distantes en la proteína.

El proceso de formación del complejo proteína-ligando (PL) a partir de la unión de una proteína P con un ligando L se puede representar de la siguiente manera:

$$P + L \rightarrow PL\ . \tag{2.19}$$

Para caracterizar cuantitativamente la afinidad de la proteína por el ligando, se recurre a la constante de unión o afinidad (K_a), que se define en solución acuosa y en equilibrio termodinámico a una dada presión y temperatura, según:

$$\frac{[PL]}{[P][L]} = K_a = \frac{1}{K_d}, \tag{2.20}$$

donde [ ] representa la concentración en equilibrio. Estos valores de concentración se encuentran divididos por concentraciones estándar 1 M por lo que la constante no tiene unidades. La constante de disociación, K_d, correspondiente a la inversa de la constante de unión, muchas veces aparece reportada con unidades, representando la concentración de ligando necesaria para que la mitad de la concentración de proteína esté formando parte del complejo unido en equilibrio (si $K_d = [L]$ cuando $[P] = [PL]$).

A partir de K_a, es posible definir la energía libre de unión (ΔGu) proteína-ligando a temperatura T y presión constantes, según:

$$\Delta Gu = - RT\, ln\left(K_a\right), \tag{2.21}$$

donde R es la constante de los gases.

Una gran parte de la bibliografía disponible propone métodos para estimar esta afinidad en el contexto del desarrollo de fármacos, donde el ligando suele ser una molécula pequeña y la proteína, el blanco. Sin embargo, para esta tesis, los cálculos de energía libre de unión fueron realizados para estimar la afinidad entre una proteína y una cadena simple de ARN, mediante un método denominado MM-PB/GBSA (por sus siglas en inglés provenientes de *Molecular Mechanics - Poisson Boltzmann / Generalized Born Surface Area*), que tiene la ventaja de ser de bajo costo computacional. A diferencia de los ligandos pequeños, el ARN presenta una alta densidad de carga negativa y una notable flexibilidad estructural, lo que plantea desafíos adicionales para la caracterización precisa de su afinidad con proteínas.

El método MM-PB/GBSA estima la energía libre de unión a partir de trayectorias de dinámica molecular, combinando la energía mecánica molecular con modelos de solvatación implícita. Este enfoque se basa en el cálculo de la energía libre de unión (ΔGu) en base al ciclo termodinámico planteado en la Figura 2.3: la energía libre de unión en ausencia de solvente (en vacío o fase gaseosa, $\Delta Gu^{vacío}$) y la energía libre de solvatación (ΔGu^{sv}). La energía libre de solvatación se divide en una parte polar (calculada mediante el modelo PB o GB) y una parte no polar (proporcional a la superficie accesible al solvente), mientras que la energía libre de unión en vacío se compone de una contribución entálpica, que incluye las interacciones electrostáticas y de van der Waals (calculadas de la manera explicada en la Sección 2.2.3.2), que corresponden a las contribuciones de no enlace entre la proteína y el ARN, y una contribución entrópica que requiere el mayor costo computacional. En muchas ocasiones, se suele calcular el $\Delta\Delta Gu$ para comparar sistemas similares, por lo que es razonable suponer que la contribución entrópica es aproximadamente la misma y de esta manera es posible no considerar esta contribución.

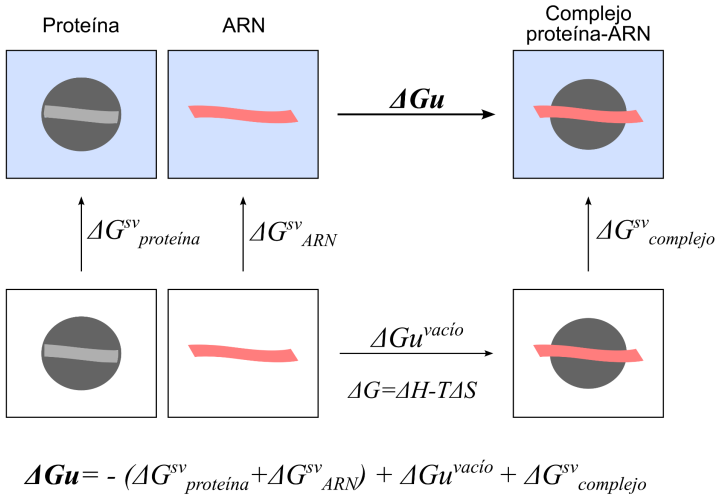


Figura 2.3: Ciclo termodinámico utilizado para el cálculo de energía de unión entre la proteína y el ARN.

Para aplicar MM-PB/GBSA, se extraen múltiples conformaciones o imágenes de una trayectoria obtenida de una simulación de dinámica molecular del complejo en solución acuosa. Sobre cada una de estas configuraciones, se calcula su energía de unión considerando tanto el complejo como sus componentes individuales (las conformaciones de la proteína y el ARN por separado en esta tesis se extraen de la simulación del complejo, pero podrían proporcionarse datos de simulaciones realizadas por separado). Finalmente, la energía libre de unión del ensamble se obtiene promediando los resultados sobre las diferentes imágenes o configuraciones.

Tanto MM/PBSA como MM/GBSA son ampliamente utilizados para estimar la energía libre de unión en complejos proteína-ARN. El modelo PB es más preciso porque resuelve la ecuación de Poisson-Boltzmann, representando mejor la frontera entre el soluto y el solvente pero resulta computacionalmente más costoso, mientras que el modelo GB es una aproximación eficiente que ha sido mejorada con el correr del tiempo para poder aplicarse a sistemas biomoleculares [27]. Debido al desarrollo existente y al conveniente costo computacional, para los cálculos realizados en esta tesis se utilizó un modelo GB avanzado [28] para la estimación de la parte polar de ΔGu^{sol}, que está implementado en el paquete Amber [20], y que incorpora mejoras específicas para sistemas compuestos por proteínas y ácidos nucleicos.

El volumen de cada partícula juega un rol fundamental en la aproximación GB, lo que genera que la descripción del radio sea un parámetro susceptible a análisis y perfeccionamiento. En este sentido, el método utilizado redefine parámetros individuales, optimizados de manera separada para cada tipo biomolecular (proteínas o ácidos nucleicos), lo que resulta especialmente importante para representar con mayor precisión las características electrostáticas y de solvatación de complejos proteína-ARN [28–30].

2.4 Perfil de energía libre

2.4.1 Introducción

La energía libre es una función termodinámica fundamental que describe el estado de equilibrio de un sistema y su capacidad para realizar trabajo útil bajo ciertas condiciones. Existen dos formas de energía libre: la energía libre de Helmholtz y la energía libre de Gibbs. La energía libre de Helmholtz es apropiada para sistemas a volumen y temperatura constantes, mientras que la energía libre de Gibbs es la función relevante cuando el sistema está a presión y temperatura constantes. En la práctica, para sistemas en fases condensadas, ambas funciones suelen ser numéricamente muy similares. Estas funciones dependen de variables macroscópicas como el número de partículas, el volumen (o presión) y la temperatura. Su valor no puede asignarse a una configuración microscópica particular, sino que representa una propiedad macroscópica del sistema, resultado de considerar todas las posibles configuraciones de las partículas que lo componen.

En química y biofísica computacional es habitual estudiar procesos que involucran cambios conformacionales o reacciones químicas, en los que los estados inicial y final difieren en características microscópicas, como la geometría molecular. Para analizar estos procesos, se introduce el concepto de coordenada de reacción, que es una variable (o conjunto de variables) que describe el progreso del sistema entre dos estados. Esta coordenada puede ser, por ejemplo, la distancia entre dos átomos, un ángulo, o una combinación de ellas.

El potencial de fuerza media (PMF, por sus siglas en inglés) describe cómo varía la energía libre de un sistema a lo largo de una coordenada de reacción, es decir, una variable que describe el progreso de un dado proceso (como una reacción química o un cambio conformacional). Los gráficos de PMF en función de la coordenada de reacción, o perfiles de energía libre, permiten visualizar los estados estables y las barreras energéticas que el sistema debe superar durante el proceso en estudio [11].

Para construir el PMF, primero se define la coordenada de reacción, ξ, y se estima la probabilidad de que el sistema se encuentre en un cierto valor de esa coordenada, P(ξ). Esta probabilidad se puede obtener, bajo el supuesto de ergodicidad (es decir, que el sistema explora todas las configuraciones accesibles a lo largo del tiempo), a partir de simulaciones. Según la termodinámica estadística, la relación entre la probabilidad y la energía libre como función de la coordenada de reacción es:

$$A(\xi) = -\, k_B T\, ln(P(\xi))\ , \tag{2.22}$$

Donde $A(\xi)$ representa el perfil de energía libre (de Helmholtz en este caso, pero podría ser de Gibbs, dependiendo las condiciones), k_B es la constante de Boltzmann y T, la temperatura. La relación logarítmica implica que los valores de ξ donde la probabilidad P(ξ) es alta (es decir, donde el sistema pasa más tiempo) corresponden a “valles” en el perfil de energía libre (estados estables), mientras que los valores donde P(ξ) es baja corresponden a “picos” o barreras (estados de transición).

El punto más alto del perfil de energía libre entre reactivos y productos se denomina estado de transición, presumiendo la identificación de la verdadera coordenada de reacción. Este estado representa el “cuello de botella” energético que el sistema debe atravesar para que la reacción ocurra. La teoría del estado de transición (TST) [31], establece que la velocidad de una reacción está determinada por la altura de esta barrera de energía libre. La constante de velocidad k se relaciona con la barrera de energía libre de activación mediante la ecuación de Eyring:

$$k = \frac{k_B T}{h} exp\left(-\frac{\Delta A^{\ddagger}}{k_B T}\right), \tag{2.23}$$

donde h es la constante de Planck. Esta ecuación permite vincular el parámetro termodinámico de la barrera de energía libre de un proceso con la constante de velocidad que caracteriza la cinética del mismo.

2.4.2 Muestreo sesgado

Como se mencionó anteriormente, a partir de simulaciones de dinámica molecular se explora una pequeña porción de la superficie de energía potencial (SEP), delimitada por la estructura inicial y las conformaciones accesibles mediante la superación de barreras del orden de k_BT. Esto significa que una gran cantidad de estados o configuraciones no serán visitados. Al querer estudiar un proceso en particular que lleva al sistema desde un estado inicial hacia uno final, es posible plantear y proponer configuraciones intermedias y/o imponer restricciones en las simulaciones o sesgos que “obliguen” al sistema a adoptar esas configuraciones. Algunos métodos conocidos y ampliamente utilizados con este fin son *Umbrella Sampling [32]*,

Dinámica Molecular Guiada [33], *Replica Exchange [34]* y Metadinámica [35]. Por otro lado, existen algoritmos que permiten encontrar caminos de mínima energía, mediante minimizaciones sucesivas con ciertas restricciones, como el método de la banda elástica (NEB) [36] y el método de la cuerda (ASM) [37,38].

Este tipo de muestreo nos permite calcular propiedades sobre configuraciones con baja probabilidad que al ser contrastadas con resultados experimentales pueden aportar información valiosa a nivel atómico o descartar ciertas hipótesis. A continuación se encuentra el marco teórico de las distintas maneras de sesgar el muestreo, dado un proceso de interés, que fueron empleadas en esta tesis.

2.4.2.1 Búsqueda de caminos de mínima energía potencial

Cuando se quiere estudiar el proceso mediante el cual un sistema cambia entre dos estados estables (para el caso de estudio de esta tesis se estudió una reacción química pero podría aplicarse a un cambio conformacional u otro tipo de procesos), estos estados se caracterizan por estar ubicados en mínimos local dentro de la superficie de energía potencial. La transición de un estado a otro implica entonces pasar por regiones que aumenten la energía. Si bien el sistema podría llegar de un estado a otro por cualquier camino, suponiendo que posee la energía suficiente para hacerlo, la transición ocurrirá con mayor probabilidad cuando esta sea a través de la barrera de mínima energía. El punto máximo en el camino de mínima energía que conecta dos estados tiene la particularidad de ser un “punto silla” de la SEP ya que corresponde a un mínimo en la dirección perpendicular al camino y un máximo en la dirección paralela al camino.

El método de la banda elástica NEB (del inglés, Nudged Elastic Band) se utiliza para encontrar caminos de mínima energía potencial entre dos estados definidos *a priori* de un sistema. Este enfoque consiste en interpolar varias configuraciones entre los estados inicial y final para formar una "banda elástica" que conecte ambos estados. Estos puntos intermedios, llamados “imágenes” o nodos, se optimizan para minimizar la energía potencial mientras se mantienen distribuidos de manera uniforme a lo largo del camino. Para lograrlo, se aplican fuerzas armónicas que restringen el movimiento de las imágenes en la dirección tangente al camino, evitando así que se acumulen los puntos en los mínimos que corresponden a los estados de partida y llegada, permitiendo únicamente el movimiento perpendicular a él. Esto permite identificar el estado de transición y su barrera de energía potencial asociada, proporcionando información útil para entender los mecanismos de reacción o los cambios estructurales en biomoléculas [39].

2.4.2.2 Búsqueda de caminos de mínima energía libre

En esta sección se describe la primera etapa del método de la cuerda ASM (del inglés, Adaptive String Method) [38] que se utiliza para encontrar caminos de mínima energía libre en procesos de transición entre dos estados, similar al método NEB, pero con mayor flexibilidad ya que su algoritmo no tiene el requisito de mantener los nodos equidistantes en la dirección del camino. El hecho de que cada nodo puede muestrear sus alrededores hace que se pueda obtener el camino de mínima energía *libre*, a diferencia del método NEB descrito con el que busca el camino de mínima energía *potencial*. La energía potencial en función de las coordenadas del sistema está definida según la representación del modelo elegido, como fue explicado en las Secciones 2.2.3, 2.2.4, 2.2.5. En lugar de mantener puntos fijos o nodos

equidistantes, la cuerda (análoga a la banda de NEB) que representa el camino se adapta dinámicamente durante la optimización para seguir las regiones de menor energía en la superficie de energía potencial. Esto significa que la cuerda se ajusta constantemente para pasar por zonas donde la energía es más baja, evitando áreas energéticamente desfavorables. Para lograrlo, los puntos que forman la cuerda se redistribuyen a partir de un algoritmo que evalúa la curvatura local: en zonas de alta curvatura o cambios abruptos en la energía, los puntos se agrupan más, mientras que en regiones más planas se distribuyen de manera más uniforme. Esto permite representar con mayor precisión los puntos críticos, como las barreras de energía o los estados de transición, logrando un camino continuo y eficiente que refleja el cambio estructural del sistema con la menor barrera energética.

Para hacer la exploración de la SEP eficiente, el método ASM tiene implementado simulaciones de dinámica molecular con intercambio de réplicas (*Replica Exchange*), Se intentan intercambios periódicos entre las configuraciones de nodos adyacentes, los cuales son aceptados o rechazados según el criterio de Metropolis.

Para evaluar el avance de la búsqueda, se calcula el RMSD en promedio de la cuerda respecto a la cuerda anterior y al observar que este no cambia durante cierto tiempo de simulación, puede considerarse que el camino de mínima energía libre ha convergido.

2.4.2.3 *Umbrella sampling*

Umbrella Sampling es una técnica de simulación utilizada para estudiar transiciones no observadas en una dinámica molecular convencional. Se utiliza para "forzar" al sistema a explorar ciertas regiones del espacio de fases, que de otro modo serían visitadas sólo con una probabilidad muy baja. Este método se aplica comúnmente en estudios de procesos como cambios conformacionales de proteínas, interacciones proteína-ligando y reacciones químicas [32].

La técnica de umbrella sampling implica primero definir una coordenada de reacción (ξ) que permite describir el proceso de interés y dividir el proceso en valores discretos de ξ. Para cada valor, se realiza una simulación de DM sesgada con un potencial armónico que restringe el valor de la coordenada de reacción. A partir de las probabilidades sesgadas de ξ de cada ventana se puede reconstruir el perfil de energía libre sin sesgo, mediante diversas técnicas como el método de histogramas pesados (*Weighted Histogram Analysis Method* o WHAM [40]) o integración termodinámica [32].

2.5 Alosterismo: Teoría de redes y grafos

Una red o grafo permite representar las relaciones entre componentes dentro de un sistema complejo. Este tipo de representaciones se compone de nodos (o vértices) y aristas (o enlaces) que indican interacciones directas y puede analizarse mediante la teoría de grafos, una rama de las matemáticas orientada a identificar propiedades globales y patrones locales en distintos tipos de sistemas. Estos métodos han resultado particularmente útiles en el estudio de redes sociales, circuitos eléctricos, sistemas de transporte y comunicación, así como para analizar sistemas químicos y biomoleculares. Para el caso de las simulaciones de dinámica molecular el análisis basado en grafos resulta beneficioso para interpretar grandes volúmenes de datos y estudiar fenómenos como la regulación alostérica, un mecanismo clave

ejemplificado en paradigma clásico de que la unión de un ligando en un sitio alostérico (lejano al sitio catalítico) induce cambios conformacionales sustanciales en el sitio catalítico. Esta forma de comunicación estructural que ocurre en las biomoléculas puede modelarse mediante redes dinámicas derivadas de simulaciones computacionales [41].

Los sistemas biológicos modelados atómicamente para realizar simulaciones de dinámica molecular pueden representarse como grafos, donde los átomos o grupos de átomos son nodos y sus interacciones, aristas. Este enfoque permite caracterizar propiedades como la conectividad, con qué otros nodos interactúa, y centralidad, qué relevancia tiene un nodo en función de su rol en la transmisión de información. Esto es particularmente útil para comprender la regulación alostérica en complejos biomoleculares, identificando residuos o regiones clave en la propagación de señales. La aplicación de la teoría de grafos en simulaciones de dinámica molecular ha llevado a importantes avances en la comprensión de mecanismos alostéricos en complejos ARN-proteína de gran tamaño, como el sistema de edición génica CRISPR-Cas9, donde el uso combinado de estas herramientas permitió identificar residuos clave y cambios conformacionales que regulan la catálisis y la selectividad [42].

Si bien estas herramientas permiten el estudio completo de complejos ARN-proteína, en esta tesis el enfoque está puesto en los cambios alostéricos dentro del sistema proteico y aunque el ARN está presente en las simulaciones de dinámica molecular, el análisis solamente fue aplicado a los aminoácidos. Además, en los casos estudiados se eligió posicionar los nodos en los átomos Cα de los aminoácidos, de las muchas formas posibles de definir qué es un nodo y donde posicionarlo. Esto se debe a la implementación de los algoritmos empleados, que a su vez refieren a la bibliografía en la que se apoyan los análisis realizados. Es por esto que de acá en adelante los grafos o redes mencionados siempre se referirán (explícita o implícitamente) al sistema proteico formado por M aminoácidos representados por la posición de su átomo Cα en el contexto de un estudio de simulación computacional de dinámica molecular en donde se obtuvieron N imágenes de la trayectoria. Cabe destacar que podría haberse definido la posición de cada nodo como el centro de masa del aminoácido o incluso podría describirse el sistema con 2M nodos en donde la mitad de los nodos representan el esqueleto de cada aminoácido y la otra mitad, su cadena lateral, entre muchas otras maneras.

2.5.1 Ejemplo sencillo de una red o grafo

La red o grafo que modela un sistema proteico dinámico puede pensarse como una matriz de MxM (M es la cantidad de aminoácidos) en el que cada elemento de matriz contiene alguna característica de los nodos i y j. De esta manera, quedan caracterizadas las aristas con un valor, como podría ser por ejemplo, la distancia promedio ($<r(A_i, A_j)>$) entre los átomos Cα de cada par de aminoácidos a lo largo de una trayectoria con N imágenes, obteniéndose un mapa de distancias. Esto se conoce como grafo completo ya que todos los nodos están vinculados entre sí mediante aristas. La distancia promedio se calcula a partir de la siguiente ecuación, siendo $r(A_i, A_j)$ la distancia entre los átomos Cα, A_i y A_j, de cada par de aminoácidos.

$$< r(A_i, A_j) > \; = \frac{\sum^{N} \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2 + (z_i - z_j)^2}}{N} \; . \qquad (2.24)$$

Si bien es sencillo calcular un mapa de distancias completo o filtrado, la información sobre la dinámica que aportaría este tipo de análisis no parece ser mucho más reveladora que otro tipo de análisis como una inspección visual de la trayectoria, pero permite comprender mediante un ejemplo simple y utilizando una métrica fácil de racionalizar (la distancia) el concepto de grafo o red en el contexto de dinámica molecular. A continuación se encuentran una serie de caracterizaciones de redes dinámicas y grafos que permiten extraer información de la dinámica "encriptadas" dentro del movimiento global de sistema.

2.5.2 Correlación lineal *versus* correlación generalizada

Obtener experimentalmente información sobre los movimientos atómicos y su comportamiento colectivo para racionalizar el efecto alostérico en proteínas no es sencillo. Por esta razón, el análisis de correlación en simulaciones de dinámica molecular representa una oportunidad para modelar estos comportamientos. El método más conocido y simple para calcular la correlación entre dos variables es calcular el coeficiente de correlación de Pearson, que evalúa la magnitud y la dirección de la relación lineal entre dos variables. Es posible, entonces, calcular este coeficiente y obtener la correlación lineal (CL) a partir de las posiciones de dos átomos (como por ejemplo, los átomos Cα de dos aminoácidos) a lo largo de una trayectoria de dinámica molecular mediante la siguiente ecuación:

$$CL_{ij} = \frac{<(r_i - \overline{r_i}) \cdot (r_j - \overline{r_j})>}{\sqrt{<(r_i - \overline{r_i})^2><(r_j - \overline{r_j})^2>}} \quad , \tag{2.25}$$

donde r_i, r_j corresponden a las posiciones en cada una de las N imágenes de la trayectoria dadas por las coordenadas (x_i, y_i, z_i) y (x_j, y_j, z_j) de los átomos i y j respectivamente; $\overline{r_i}$, $\overline{r_j}$ corresponden a las posiciones promedio de cada átomo y < > indica el promedio sobre las N imágenes de la trayectoria.

Análogamente al ejemplo mencionado del mapa de distancias, el resultado de calcular la correlación entre todos los átomos de una proteína a partir de una dinámica molecular puede pensarse como una matriz de MxM (M es la cantidad de aminoácidos), en donde cada elemento de matriz CL_{ij} contiene el valor correspondiente a la correlación entre el átomo i y el j que puede variar entre -1 y 1, siendo estos valores extremos indicativos de correlación ("se mueven en la misma dirección y sentido) y anti-correlación (se mueven en la misma dirección pero sentidos opuestos), respectivamente, y el 0, ausencia de correlación. Los elementos diagonales de esta matriz serán todos 1, ya que cada elemento correlaciona linealmente consigo mismo. Un tipo de correlación totalmente ignorada por este cálculo sería el caso de que dos átomos se movieran correlacionadamente formando un ángulo de 90º entre ellos, lo cual se evidencia en la Ecuación 2.25 por la presencia del producto escalar. Cuanto más se aleje el ángulo entre los átomos de 180º, o si ellos se mueven en la misma dirección pero la variación es no-lineal (podría ser cuadrática, exponencial, logarítmica, etc) o si la correlación está desfasada en el tiempo, este coeficiente tampoco capturará correctamente la existencia del vínculo entre los movimientos.

En la literatura se han propuesto muchas formas de considerar correlaciones de orden superior, es decir, tanto lineales como no lineales, proporcionando un análisis más robusto de

las dependencias dentro del sistema. Una de ellas consiste en calcular la correlación generalizada (CG), un coeficiente que se basa en técnicas de estadística no paramétrica, es decir que no se asume una forma funcional predeterminada para la distribución de los datos, lo cual resulta especialmente útil para el estudio de sistemas complejos como las biomoléculas. Es posible calcular la CG a partir de la información mutua (IM), que evalúa la cantidad de información compartida entre dos variables y se obtiene al comparar las distribuciones marginales (distribuciones individuales que consideran de cada variable por separado), con la distribución conjunta. En el contexto de dinámica molecular, cada distribución es el conjunto de posiciones (x, y, z) que adopta cada átomo a lo largo de la trayectoria y se calcula la CG a partir de la información mutua comparando las distribuciones de a dos átomos A_i y A_j . La información mutua puede tomar valores entre 0 y +∞ y está íntimamente relacionada con la entropía de Shannon (H) a través de la siguiente ecuación:

$$IM(A_i, A_j) = H(A_i) + H(A_j) - H(A_i, A_j)\,, \tag{2.26}$$

que indica que la información mutua se puede obtener al sumar las entropías individuales de cada conjunto de datos y restar la entropía de la distribución conjunta. La entropía, en este contexto, se puede pensar como la “sorpresa” al conocer un valor extra, habiendo determinado los demás datos de la distribución. Esto significa que si la mayoría de los datos toman valores cercanos entre sí, habrá baja “sorpresa” al conocer uno nuevo. En el límite se obtiene entropía igual a cero cuando todos los valores de una distribución son iguales. Por el contrario, cuanto más dispersos estén los datos, mayor será su entropía. En este sentido, se puede pensar a la entropía como relacionada con la “incertidumbre”, ya que antes de conocer un valor nuevo de una distribución con entropía alta, será difícil tener información sobre el resultado que se va a obtener.

Cuando dos distribuciones son completamente independientes, no comparten información mutua (IM = 0), por lo tanto, la suma de sus entropías será igual a la entropía de la distribución conjunta. En cambio, si entre las distribuciones existe algún tipo de relación, se espera que la entropía de la distribución conjunta sea menor que la suma de las entropías de las distribuciones individuales, ya que al integrar los datos, habrá menos “sorpresa” al conocer aquellos datos de una distribución que estén relacionados con la otra (Figura 2.4).

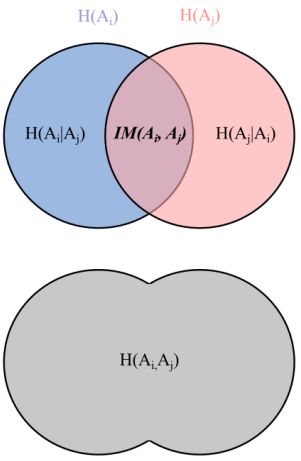


Figura 2.4. Representación gráfica del cálculo de la información mutua a partir de la entropía de Shannon.

Uno de los estimadores reportados para calcular la información mutua involucra el concepto de k vecinos más cercanos [43,44]. Según esta propuesta, para estimar la información mutua, se evalúan las variaciones en la posición de los dos átomos A_i y A_j considerados evaluando todas las imágenes de la trayectoria en función de una que funciona de referencia y que se va cambiando iterativamente. Para cada referencia, se seleccionan las k imágenes vecinas con menor variación teniendo en cuenta simultáneamente ambos átomos A_i y A_j (el número k es fijado en 6, como fue propuesto por Lange et al. [44]). Una vez seleccionadas estas imágenes, se reduce la búsqueda entre ellas para seleccionar la mayor variación y establecer un límite de distancia (ε) pero esta vez, separando el análisis para cada átomo (ε_i para átomo A_i y ε_j para átomo A_j). Esta distancia límite va cambiando a medida que cambia la imagen de referencia. Una vez establecido este límite, se contabilizan las imágenes de la trayectoria (n_i y n_j) para las cuales cada átomo está a una distancia menor que el límite establecido. Por último, se calcula la información mutua a partir de la siguiente ecuación (el desarrollo teórico de este estimador se encuentra detallado en [43]):

$$IM \simeq \psi(k) - 1/k - \langle \psi(n_i) + \psi(n_j) \rangle + \psi(N), \qquad (2.27)$$

en la cual < > representa el promedio sobre las imágenes de la trayectoria, ya que cada una de ellas conlleva una referencia distinta y por lo tanto genera límites distintos de distancias para el conteo. En el desarrollo teórico de este estimador para la información mutua, surge la función digamma (ψ) aplicada a n_i, n_j, k y N, que para valores reales positivos mayores a 2 (como los valores que toman las variables en los casos estudiados) esta función es creciente y positiva, con una forma funcional muy similar a la del logaritmo natural.

Cuanto más independientes son las distribuciones, se obtienen valores más bajos de información mutua, dados por valores más altos de n_i y n_j, aunque estos últimos tienen un límite dado por la cantidad de imágenes de la trayectoria y la cantidad de vecinos considerados, impidiendo que la información mutua sea negativa. En el límite de dos distribuciones completamente independientes, el estimador tiende a cero cuando N tiende a infinito [43]. Por el contrario, cuanta más información comparten las distribuciones, n_i y n_j son más bajos y por lo tanto, la información mutua aumenta.

Si bien la información mutua permite comparar dos distribuciones, el límite superior tendiendo a +∞ dificulta la comparación con otros coeficientes de correlación como el de Pearson. A partir de esta observación, se ha propuesto la siguiente normalización [44], mediante la cual la correlación generalizada toma valores entre 0 (variables independientes) y 1 (correlación total).

$$CG_{ij} = \left\{1 - e^{\frac{2IM(A_i, A_j)}{d}}\right\}^{-1/2}, \qquad (2.28)$$

donde d es dimensión, que en los casos considerados es igual a 3. A diferencia del coeficiente de Pearson, esta medida permite detectar relaciones entre movimientos atómicos que no ocurren en la misma dirección ni al mismo tiempo, es decir, que pueden ser no colineales o estar desfasados temporalmente. Esto es particularmente relevante en dinámica

molecular, donde las fluctuaciones funcionales pueden involucrar movimientos asincrónicos o indirectos entre regiones estructuralmente distantes.

El estimador de la información mutua ha mostrado ser eficiente [43] y al ser aplicado a simulaciones de dinámica molecular, el cálculo de correlación generalizada ha permitido ampliar el conocimiento de las redes de comunicación dentro de una proteína, evidenciando relaciones no capturadas por la correlación cruzada [41,42,44–46]

2.5.3 Distancias ponderadas

Bajo la premisa de que dos regiones pueden presentar alta correlación entre sus nodos aun estando alejadas espacialmente, es posible asignar a cada arista un valor que represente su "longitud" en el espacio de las correlaciones, definido por el grado de correlación entre nodos. Así, nodos altamente correlacionados se consideran más "próximos". Esta representación permite identificar correlaciones dinámicas entre sitios distantes, diferenciándose del concepto tradicional de distancia euclidiana. De este modo, se puede analizar cómo se transmite información dentro de la proteína, por ejemplo, cuando un evento como la unión de un ligando en una región induce movimientos correlacionados en otra, como el sitio catalítico.

Como se mencionó anteriormente, es útil caracterizar las aristas de una red mediante métricas como la distancia promedio o la correlación (lineal o generalizada). Otra opción es la distancia ponderada, que representa la "longitud" en el espacio de las correlaciones. En este caso, se define a partir de la correlación generalizada entre los átomos A_i y A_j, se calcula a través de la siguiente ecuación:

$$d_{CG_{ij}} = - \, log[CG_{ij}] \, , \tag{2.29}$$

Este cambio funcional hace que una mayor correlación tenga una menor distancia ponderada, dando la idea de cercanía, aunque ya no en el espacio euclidiano, sino en un nuevo espacio de correlaciones, lo que permite una posterior búsqueda de caminos óptimos para la transmisión de información.

2.5.4 Caminos óptimos

Aunque dos nodos estén alejados en el espacio euclidiano, pueden estar comunicados, lo que podría verse reflejado directamente en una distancia ponderada corta, que proviene de una alta correlación. Una manera de racionalizar la transmisión de información que permite que este tipo de comunicación se establezca, supone que la señal debe transmitirse a través de nodos adyacentes en la red. En este contexto, los caminos óptimos que conectan dos nodos se pueden definir como las secuencias de aristas que minimizan la suma total de las distancias ponderadas, es decir, aquellas rutas que maximizan la correlación a lo largo del trayecto. Cabe aclarar que entre dos nodos puede haber uno o más caminos óptimos con el mismo valor de distancia total ponderada.

Antes de realizar la búsqueda de caminos óptimos a partir de la matriz de distancias ponderadas, se requiere eliminar las aristas correspondientes a átomos no adyacentes, ya que a través de ellas no se espera que la información sea transmitida directamente. Para

hacer esto, la matriz se filtra utilizando dos criterios *a priori* modulables, uno de distancia y otro de permanencia de la interacción. Si bien estos criterios son modulables, es usual encontrar este tipo de análisis bajo la condición de que una arista es considerada una interacción relevante cuando la distancia euclidiana entre los átomos pesados de los aminoácidos (representados en los nodos que componen la arista) es menor a 4.5 Å durante al menos el 75% de la simulación. Este filtro genera un subgrafo de distancias ponderadas donde solo se conservan las aristas que conectan nodos físicamente cercanos.

Al imaginar una búsqueda de caminos óptimos entre dos nodos, se suele definir un nodo de partida y un nodo de llegada. Para el tipo de análisis que se realizan en esta tesis, es indistinta la definición del nodo de partida y llegada, ya que el grafo o red que representa el sistema biomolecular es lo que se conoce como “no dirigido” a diferencia de los grafos “dirigidos” en donde esta definición sí afectaría.

En este contexto, los algoritmos más usados para el cálculo de caminos óptimos son el algoritmo de Dijkstra [47] y el de Floyd-Warshall [48]. La principal diferencia radica en que el primero fue diseñado para encontrar el camino óptimo partiendo de un nodo específico, mientras que el segundo fue diseñado para caracterizar todos los pares de nodos con la distancia correspondiente al camino óptimo. Para caracterizar todos los caminos óptimos que conectan todos los nodos entre sí, es preferible iterar sobre el algoritmo de Dijkstra, ya que además de dar información sobre la distancia óptima, también muestra los nodos involucrados en ese camino. Sin embargo, es posible modificar el algoritmo de Floyd-Warshall para incluir la descripción de los nodos involucrados en los caminos óptimos [49].

La propagación de señales en sistemas biomoleculares no necesariamente sigue una única trayectoria óptima, sino que puede involucrar múltiples rutas alternativas que, si bien no son las más cortas, presentan longitudes comparables. Por ello, el análisis de caminos mínimos suele ir un poco más allá y contemplar también principales rutas subóptimas. Tanto los caminos óptimos como los subóptimos se determinan en función de aristas ponderadas y su identificación resulta clave para comprender los posibles mecanismos de transducción de señales dentro del sistema.

2.5.5 Intermediación de aristas (*edge betweenness*)

La intermediación de aristas (EB, por sus siglas en inglés *edge betweenness*) es otra medida utilizada en teoría de grafos para evaluar la importancia de una arista o conexión dentro de una red, basada en cuántos caminos óptimos pasan a través de ella. En otras palabras, mide el número de caminos óptimos que conectan pares de nodos y que atraviesan una arista específica. Esta métrica es útil para identificar las aristas que actúan como intersecciones clave en la conectividad de la red. Un valor alto de intermediación de aristas indica que la arista juega un papel crucial en la comunicación entre diferentes partes de la red, lo que podría hacerla un punto crítico en términos de flujo de información. La intermediación de aristas se calcula sumando, para cada par de nodos, la fracción de caminos más cortos que pasan por la arista en cuestión.

$$EB(e_{ij}) = \sum_{s,t \in V \neq i,j} \frac{\sigma(s,t|e_{ij})}{\sigma(s,t)} , \tag{2.30}$$

donde V es el conjunto de todos los nodos sin tener en cuenta los nodos i y j que son los que participan en la arista e_{ij} a caracterizar, $\sigma(s,t)$ es la cantidad de caminos óptimos entre los nodos s y t; y $\sigma(s,t|e_{ij})$ es la cantidad de los caminos óptimos entre s y t que pasan por la arista e_{ij} [49]. Este parámetro cuantifica el "tráfico" que pasa a través de las aristas, midiendo con qué frecuencia una arista funciona como un puente en el flujo óptimo de información entre nodos [45].

2.5.6 Relación señal/ruido

Un cálculo de Relación Señal/Ruido (S/R) evalúa la efectividad de las señales dentro de un sistema frente al ruido o detecciones irrelevantes. En este enfoque, se consideran las correlaciones significativas (señales) dentro de la proteína, y se compara su intensidad con la del ruido de fondo, que puede ser causado por fluctuaciones o dinámicas no relevantes. Volviendo al ejemplo del efecto alostérico de la unión de un ligando sobre el sitio catalítico de una enzima, se espera que la señal dada por la unión del ligando se transmita hacia el sitio catalítico y que la transmisión esté amplificada con respecto al ruido estructural para que la comunicación sea efectiva.

Al querer estudiar dos regiones comunicadas alostéricamente , en primer lugar, es conveniente considerar varios nodos (aminoácidos) en cada una de las regiones y, para cada combinación, caracterizar los caminos óptimos y subóptimos a través de su longitud (cantidad de nodos en el camino). Se obtiene, de esta manera, una distribución dada por los largos de los caminos calculados que permite descomponer el posterior análisis de S/N, dividiendo los caminos en tres categorías: caminos cortos, intermedios y largos, cuyas longitudes se establecen al inspeccionar visualmente la distribución. En segundo lugar, para cada camino se calcula la suma de las intermediaciones de aristas y, separando por categoría de largo, se agrupan generando tres distribuciones que representan tres tipos de "señal": corta, intermedia y larga. Por otro lado, para cada categoría de largo, se busca en la red todos los caminos entre dos nodos cualesquiera que tengan ese largo y se calcula la suma de intermediaciones de aristas para todos esos caminos, constituyendo así tres distribuciones que representan tres tipos de "ruido": corto, intermedio y largo. Finalmente, las distribuciones se aproximan por una gaussiana y se extraen los valores de la esperanza y la varianza de cada una para calcular S/R de la siguiente manera:

$$S/R = \frac{E[S]/Var(S)}{E[R]/Var(R)}\ , \tag{2.31}$$

donde E es la esperanza, Var es la varianza y ambas se estiman para la señal S y para el ruido R [45].

Bibliografía

1. Karplus M, McCammon JA. Molecular dynamics simulations of biomolecules. Nat Struct Biol. 2002;9: 646–652. doi:10.1038/nsb0902-646

2. Waterhouse A, Bertoni M, Bienert S, Studer G, Tauriello G, Gumienny R, et al. SWISS-MODEL: homology modelling of protein structures and complexes. Nucleic Acids Res. 2018;46: W296–W303. doi:10.1093/nar/gky427

3. Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, et al. Highly accurate protein structure prediction with AlphaFold. Nature. 2021;596: 583–589. doi:10.1038/s41586-021-03819-2

4. Swope WC, Andersen HC, Berens PH, Wilson KR. A computer simulation method for the calculation of equilibrium constants for the formation of physical clusters of molecules: Application to small water clusters. J Chem Phys. 1982;76: 637–649. doi:10.1063/1.442716

5. Verlet L. Computer “experiments” on classical fluids. I. thermodynamical properties of Lennard-Jones molecules. Phys Rev. 1967;159: 98–103. doi:10.1103/physrev.159.98

6. Braun E, Gilmer J, Mayes HB, Mobley DL, Monroe JI, Prasad S, et al. Best practices for foundations in molecular simulations [article v1.0]. Living J Comput Mol Sci. 2019;1. doi:10.33011/livecoms.1.1.5957

7. Becker/Mackerel. Computational biochemistry and biophysics. New York, NY: Marcel Dekker; 2001.

8. Ryckaert J-P, Ciccotti G, Berendsen HJC. Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. J Comput Phys. 1977;23: 327–341. doi:10.1016/0021-9991(77)90098-5

9. Perilla JR, Goh BC, Cassidy CK, Liu B, Bernardi RC, Rudack T, et al. Molecular dynamics simulations of large macromolecular complexes. Curr Opin Struct Biol. 2015;31: 64–74. doi:10.1016/j.sbi.2015.03.007

10. Lin K, Wang Z. Multiscale mechanics and molecular dynamics simulations of the durability of fiber-reinforced polymer composites. Commun Mater. 2023;4: 1–16. doi:10.1038/s43246-023-00391-2

11. Leach AR. Molecular modelling. 2nd ed. London, England: Prentice-Hall; 2001.

12. Darden T, York D, Pedersen L. Particle mesh Ewald: An $N \cdot \log(N)$ method for Ewald sums in large systems. J Chem Phys. 1993;98: 10089–10092. doi:10.1063/1.464397

13. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML. Comparison of simple potential functions for simulating liquid water. J Chem Phys. 1983;79: 926–935. doi:10.1063/1.445869

14. Maier JA, Martinez C, Kasavajhala K, Wickstrom L, Hauser KE, Simmerling C. Ff14SB: Improving the accuracy of protein side chain and backbone parameters from ff99SB. J Chem Theory Comput. 2015;11: 3696–3713. doi:10.1021/acs.jctc.5b00255

15. Aytenfisu AH, Spasic A, Grossfield A, Stern HA, Mathews DH. Revised RNA dihedral parameters for the Amber force field improve RNA molecular dynamics. J Chem Theory Comput. 2017;13: 900–915. doi:10.1021/acs.jctc.6b00870

16. Ponder JW, Case DA. Force Fields for Protein Simulations. Protein Simulations. Elsevier; 2003. pp. 27–85. doi:10.1016/s0065-3233(03)66002-x

17. Schrödinger E. An undulatory theory of the mechanics of atoms and molecules. Phys Rev. 1926;28: 1049–1070. doi:10.1103/physrev.28.1049

18. Gaus M, Cui Q, Elstner M. DFTB3: Extension of the self-consistent-charge density-functional tight-binding method (SCC-DFTB). J Chem Theory Comput. 2012;7: 931–948. doi:10.1021/ct100684s

19. [cited 21 Jun 2025]. Available: https://ambermd.org/doc12/Amber25.pdf

20. Case DA, Aktulga HM, Belfon K, Cerutti DS, Cisneros GA, Cruzeiro VWD, et al. AmberTools. J Chem Inf Model. 2023;63: 6183–6191. doi:10.1021/acs.jcim.3c01153

21. Pastor RW, Brooks BR, Szabo A. An analysis of the accuracy of Langevin and molecular dynamics algorithms. Mol Phys. 1988;65: 1409–1419. doi:10.1080/00268978800101881

22. Berendsen HJC, Postma JPM, van Gunsteren WF, DiNola A, Haak JR. Molecular dynamics with coupling to an external bath. J Chem Phys. 1984;81: 3684–3690. doi:10.1063/1.448118

23. Ormeño F, General IJ. Convergence and equilibrium in molecular dynamics simulations. Commun Chem. 2024;7: 26. doi:10.1038/s42004-024-01114-5

24. Humphrey W, Dalke A, Schulten K. VMD: visual molecular dynamics. J Mol Graph. 1996;14: 33–8, 27–8. doi:10.1016/0263-7855(96)00018-5

25. Kabsch W. A solution for the best rotation to relate two sets of vectors. Acta Crystallogr A. 1976;32: 922–923. doi:10.1107/s0567739476001873

26. Roe DR, Cheatham TE 3rd. PTRAJ and CPPTRAJ: Software for processing and analysis of molecular dynamics trajectory data. J Chem Theory Comput. 2013;9: 3084–3095. doi:10.1021/ct400341p

27. Hou T, Wang J, Li Y, Wang W. Assessing the performance of the MM/PBSA and MM/GBSA methods. 1. The accuracy of binding free energy calculations based on molecular dynamics simulations. J Chem Inf Model. 2011;51: 69–82. doi:10.1021/ci100275a

28. Mongan J, Simmerling C, McCammon JA, Case DA, Onufriev A. Generalized Born model with a simple, robust molecular volume correction. J Chem Theory Comput. 2007;3: 156–169. doi:10.1021/ct600085e

29. Nguyen H, Roe DR, Simmerling C. Improved Generalized Born Solvent Model Parameters for Protein Simulations. J Chem Theory Comput. 2013;9: 2020–2034. doi:10.1021/ct3010485

30. Nguyen H, Pérez A, Bermeo S, Simmerling C. Refinement of Generalized Born Implicit Solvation Parameters for Nucleic Acids and Their Complexes with Proteins. J Chem Theory Comput. 2015;11: 3714–3728. doi:10.1021/acs.jctc.5b00271

31. Eyring H. The activated complex in chemical reactions. J Chem Phys. 1935;3: 107–115. doi:10.1063/1.1749604

32. Kästner J. Umbrella sampling: Umbrella sampling. Wiley Interdiscip Rev Comput Mol Sci. 2011;1: 932–942. doi:10.1002/wcms.66

33. Izrailev S, Stepaniants S, Isralewitz B, Kosztin D, Lu H, Molnar F, et al. Steered Molecular Dynamics. Computational Molecular Dynamics: Challenges, Methods, Ideas. Berlin, Heidelberg: Springer Berlin Heidelberg; 1999. pp. 39–65. doi:10.1007/978-3-642-58360-5_2

34. Sugita Y, Okamoto Y. Replica-exchange molecular dynamics method for protein folding. Chem Phys Lett. 1999;314: 141–151. doi:10.1016/s0009-2614(99)01123-9

35. Laio A, Parrinello M. Escaping free-energy minima. arXiv [cond-mat.stat-mech]. 2002. Available: http://arxiv.org/abs/cond-mat/0208352

36. Jónsson H, Mills G, Jacobsen KW. Nudged elastic band method for finding minimum energy paths of transitions. Classical and Quantum Dynamics in Condensed Phase Simulations. WORLD SCIENTIFIC; 1998. doi:10.1142/9789812839664_0016

37. Peters B, Heyden A, Bell AT, Chakraborty A. A growing string method for determining transition states: comparison to the nudged elastic band and string methods. J Chem Phys. 2004;120: 7877–7886. doi:10.1063/1.1691018

38. Zinovjev K, Tuñón I. Adaptive finite temperature string method in collective variables. J Phys Chem A. 2017;121: 9764–9772. doi:10.1021/acs.jpca.7b10842

39. Ásgeirsson V, Birgisson BO, Bjornsson R, Becker U, Neese F, Riplinger C, et al. Nudged elastic band method for molecular reactions using energy-weighted springs combined with eigenvector following. J Chem Theory Comput. 2021;17: 4929–4945. doi:10.1021/acs.jctc.1c00462

40. Kumar S, Rosenberg JM, Bouzida D, Swendsen RH, Kollman PA. THE weighted histogram analysis method for free-energy calculations on biomolecules. I. The method. J Comput Chem. 1992;13: 1011–1021. doi:10.1002/jcc.540130812

41. Patel AC, Sinha S, Palermo G. Graph theory approaches for molecular dynamics simulations. Q Rev Biophys. 2024;57: e15. doi:10.1017/S0033583524000143

42. Saha A, Arantes PR, Palermo G. Dynamics and mechanisms of CRISPR-Cas9 through the lens of computational methods. Curr Opin Struct Biol. 2022;75: 102400. doi:10.1016/j.sbi.2022.102400

43. Kraskov A, Stögbauer H, Grassberger P. Estimating mutual information. Phys Rev E Stat Nonlin Soft Matter Phys. 2004;69: 066138. doi:10.1103/PhysRevE.69.066138

44. Lange OF, Grubmüller H. Generalized correlation for biomolecular dynamics. Proteins. 2006;62: 1053–1061. doi:10.1002/prot.20784

45. Sinha S, Molina Vargas AM, Arantes PR, Patel A, O'Connell MR, Palermo G. Unveiling the RNA-mediated allosteric activation discloses functional hotspots in CRISPR-Cas13a. Nucleic Acids Res. 2024;52: 906–920. doi:10.1093/nar/gkad1127

46. Melo MCR, Bernardi RC, de la Fuente-Nunez C, Luthey-Schulten Z. Generalized correlation-based dynamical network analysis: a new high-performance approach for identifying allosteric communications in molecular dynamics trajectories. bioRxiv. bioRxiv; 2020. doi:10.1101/2020.06.18.160572

47. Dijkstra EW. A note on two problems in connexion with graphs. Numer Math (Heidelb). 1959;1: 269–271. doi:10.1007/bf01386390

48. Floyd RW. Algorithm 97: Shortest path. Commun ACM. 1962;5: 345–345.

doi:10.1145/367766.368168

49. Hagberg AA, Schult DA, Swart PJ. Exploring network structure, dynamics, and function using NetworkX. Proceedings of the Python in Science Conference. SciPy; 2008. pp. 11–15. doi:10.25080/tcwv9851

Capítulo 3: Estudio computacional del acoplamiento entre el ciclo de hidrólisis de ATP y el mecanismo de translocación de la helicasa NS3 sobre el ARN

3.1 Resumen

En este capítulo se estudia el efecto alostérico generado por la ocupación del sitio catalítico en la interfaz proteína-ARN de la helicasa NS3 (NS3hel). En base a las simulaciones computacionales y cálculos de energía, se evalúan varios modelos de translocación propuestos para helicasas estructuralmente similares. Además, se destaca la observación de un *loop* cercano al sitio catalítico que podría funcionar como canal de salida del ion fosfato, producto de la hidrólisis. Por último, se extiende el análisis a partir del uso de un modelo más realista del sistema NS3hel:ARN que consiste en emplear una simple cadena de ARN más larga que la resuelta en la estructura cristalográfica, ya que se ha visto que reproduce con mayor fidelidad los hallazgos experimentales. Los resultados de este capítulo se enfocan en las contribuciones que realicé para dos trabajos publicados con revisión de pares. En uno de ellos me encuentro como primera autora [1] y en el otro, compartiendo la primera autoría con Fernando Amrein [2]. Entre ambos trabajos transcurrieron dos años, durante los cuales hubo una profundización en el conocimiento del sistema, en particular por mi participación en la escritura de un *review* sobre la NS3hel [3] y en otro trabajo a partir del cual se determinó el orden de liberación de productos de la hidrólisis de ATP [4]. Esta evolución puede dar lugar a ciertas diferencias en el enfoque o en la representación de resultados entre los trabajos, que se busca integrar y contextualizar a lo largo del capítulo.

3.2 Introducción

La proteína no-estructural 3 (NS3) del virus del dengue es considerada un blanco atractivo para la búsqueda de antivirales ya que sus funciones son imprescindibles para la replicación viral [5,6]. Esta proteína está compuesta por dos dominios (Figura 3.1A): la región N-terminal se clasifica como una serina-proteasa (NS3pro, residuos 1 a 170), mientras que la región C-terminal (NS3hel, residuos 180 a 618) actúa como una helicasa separando la doble hebra de ácidos ribonucleicos (ARN), hidroliza nucleósidos trifosfato (NTP) y ribonucleósidos trifosfato (RTP) [7,8]. Estudios *in vitro* han mostrado que la mayoría de las actividades que realiza la NS3hel se conservan aún en ausencia del dominio proteasa [7,9–11].

En la NS3hel se distinguen tres subdominios (Figura 3.1 A y B). Los subdominios 1 y 2 (D1 y D2) son dominios que contienen motivos conservados (I, IA, II, III, IV, IVA, V, VI) esenciales para unir ARN simple cadena (ARN-sc) en un lado y, en el lado opuesto, formar el sitio de unión para el NTP [12]. Esta arquitectura de helicasa es, en gran medida, conservada dentro de las superfamilias SF1 y SF2 de helicasas monoméricas. A su vez, las NS3hel de los flavivirus cuentan con el subdominio 3 (D3) que encierra al ARN generando un “techo” en su sitio de unión. Este subdominio, al no estar presente en las helicasas monoméricas de los seres humanos, representa un blanco molecular atractivo para el diseño racional de fármacos contra los flavivirus. Dos hélices alfa, una en D2 y otra en D3 forman una estructura de “puerta” (*α-helical gate*) por donde ingresa la simple cadena de ARN al sitio catalítico [13].

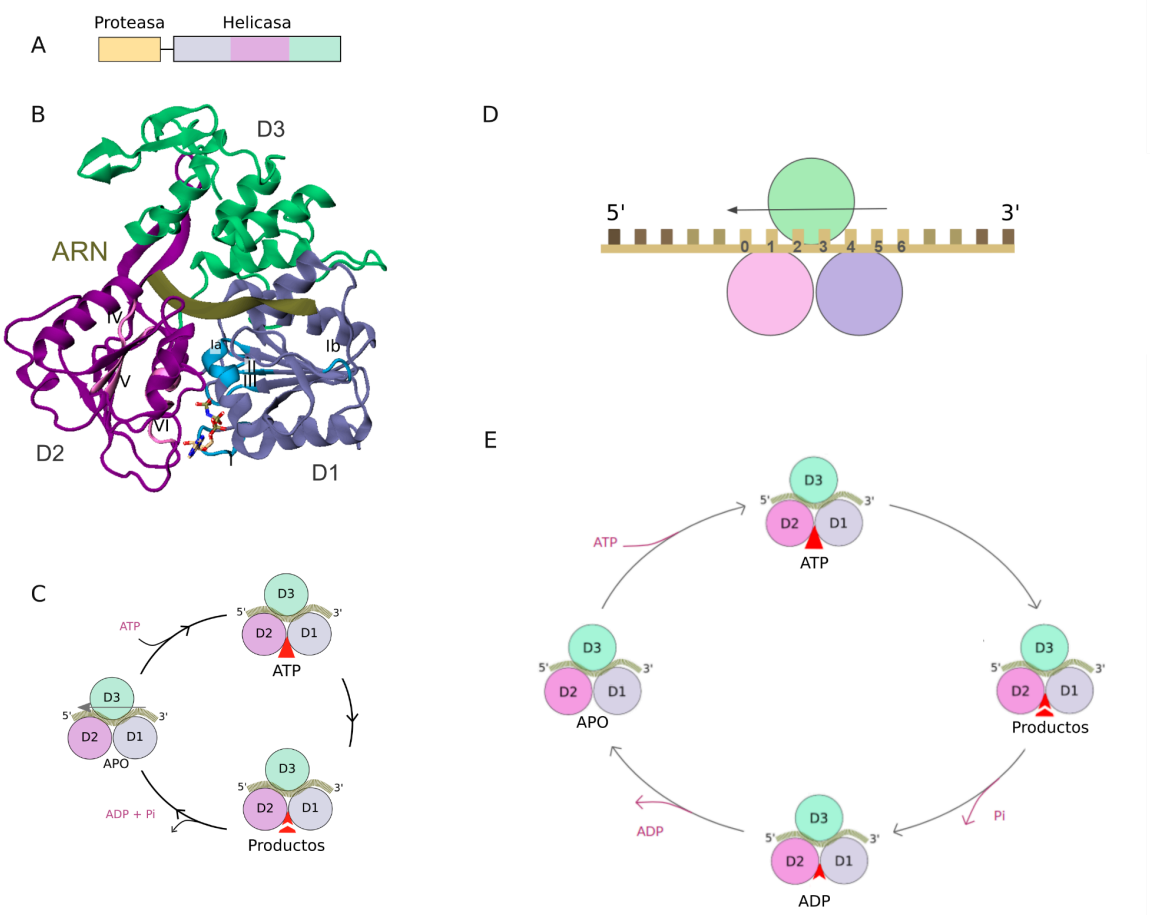


Figura 3.1. Estructura de la helicasa NS3 del virus del dengue. **A)** Organización de los dominios proteasa y helicasa de NS3 del dengue. Este trabajo se enfoca únicamente en el dominio helicasa. **B)** Subdominios de NS3hel codificados por colores: los subdominios motores tipo RecA D1 y D2 en azul y violeta, respectivamente, y D3 en verde. Un segmento de ARN de simple cadena (ARN-sc) está representado como una cinta en color verde amarronado, y la molécula de ATP se muestra representada con bastones coloreados según la identidad de los átomos que lo componen. Los motivos conservados de la helicasa en los dominios D1 y D2 están etiquetados y destacados en azul y rosa, respectivamente. **C)** Representación esquemática del ciclo de hidrólisis de ATP en NS3hel, incluyendo los estados apo, unido a ATP y unido a los productos de la hidrólisis. En rojo se representa el ATP o ADP y Pi, según corresponda. La flecha gris indica el sentido de translocación 3'-a-5' a lo largo del ARN-sc. **D)** Representación esquemática de la numeración de las bases del ARN. Los subdominios de NS3hel están codificados por colores como en el panel A. Los nucleótidos están numerados en dirección 5′ a 3′ según su base correspondiente, y los números del 0 al 6 están reservados para los nucleótidos resueltos en la estructura cristalográfica. La flecha gris indica la dirección de translocación de NS3hel, de 3′ a 5′. **E)** Ciclo de hidrólisis de ATP, teniendo en cuenta el orden de liberación de los productos ADP y Pi [4].

La translocación unidireccional de 3' a 5' de NS3hel a lo largo del ARN simple cadena está acoplada a la hidrólisis de NTP [7,10]. A pesar de la arquitectura conservada de las helicasas SF1 y SF2 [14,15], se han reportado mecanismos diferentes para esta translocación unidireccional [16–18]. Estas diferencias pueden atribuirse a divergencias estructurales sutiles entre helicasas y, en cierta medida, al uso de técnicas de medición diferentes. Estructuras recientes de helicasas virales SF2, han sido fundamentales para la propuesta de los posibles mecanismos de translocación [19–21].

Si bien la NS3hel puede unir e hidrolizar cualquier NTP, en esa tesis se ejemplifica la función catalítica con el adenosín trifosfato (ATP). El ciclo de hidrólisis de ATP en las helicasas virales es un proceso dinámico que involucra la unión de ATP, la hidrólisis de ATP a ADP y fosfato inorgánico (Pi), y la liberación del producto de la hidrólisis. Durante este ciclo, se especula que las helicasas deberían experimentar cambios conformacionales que le permitan avanzar unidireccionalmente desde el 3' hasta el 5' [7]. Esta especulación surge por el hecho de que la translocación en las helicasas virales es un proceso acoplado a la hidrólisis de ATP y permite el movimiento de la helicasa sobre la cadena de ARN, facilitando el desdoblamiento de la doble hélice de ARN o posibles estructuras que pudieran formarse debidas a interacciones de una misma cadena, como los bucles o *loops* [17]. Existen al menos dos modelos para clasificar mecanismos de translocación. Uno de ellos se conoce como el modelo de motor browniano, donde las fluctuaciones térmicas permiten que la helicasa avance sobre el ARN, mientras el ATP es hidrolizado, al experimentar transiciones entre estados flexibles y rígidos [22]. Este mecanismo puede entenderse también como pasivo. Otra hipótesis es que la helicasa se comporte según el modelo de gusano, en donde se va acumulando tensión en el sistema que es liberada cuando la helicasa avanza sobre el ARN, permitiéndole dar "pasos hacia adelante". Otra manera de entender este mecanismo es como activo [18].

A diferencia de la evidencia estructural reportada para la helicasa de otro Flavivirus, el virus de la hepatitis C (HCV), al comparar las estructuras cristalográficas de NS3hel del dengue no se observan diferencias significativas en la proteína por estar unida a nucleótidos correspondientes con diferentes etapas del ciclo catalítico de hidrólisis de ATP [23]. Por lo tanto, estas similitudes entre las estructuras no permiten explicar cómo se acopla la hidrólisis de ATP al movimiento unidireccional de la NS3hel del dengue sobre el ARN, ni por qué la hidrólisis de ATP es imprescindible para su avance. Por el contrario, la NS3hel del HCV que ha sido extensamente estudiada, se ha capturado en distintos estados conformacionales (conformaciones abierta y cerrada) durante la hidrólisis de ATP [19], lo que indica diferencias claras entre los mecanismos de translocación de las helicasas NS3 de Flavivirus y subraya la necesidad de estudios más detallados sobre NS3hel del dengue.

Estudios recientes de la NS3hel del dengue han proporcionado algunos detalles sobre la comunicación entre el sitio de unión del ARN y el sitio catalítico de hidrólisis de ATP, proponiendo el motivo V como principal regulador alostérico [24]. Además, reportaron algunos cambios dependientes de nucleótidos en las interacciones ARN-proteína [25]. Sin embargo, estos estudios no reportan diferencias conformacionales significativas, al igual que las estructuras cristalográficas. A partir de estos resultados reportados y con el objetivo de estudiar los mecanismos moleculares que modulan la interacción entre la NS3hel y el ARN, en este capítulo se explora el espacio conformacional de la NS3hel por medio de simulaciones de dinámica molecular en diferentes estados del ciclo de hidrólisis de ATP: apo, unido a ATP y unido a los productos de la hidrólisis de ATP, ADP y fosfato inorgánico (Pi, HPO_4^{2-}). En los tres sistemas, se considera el estado unido a un ARN-sc de 7 nucleótidos (NS3hel:ARN7).

Aunque las estructuras cristalográficas disponibles de NS3 fueron obtenidas unidas a un ARN-sc de 12 nucleótidos, solo se ha reportado la estructura de 7 nucleótidos [23], de los cuales únicamente 4 parecen formar interacciones estables con la proteína. Curiosamente, estudios termodinámicos sobre la unión de la helicasa a oligonucleótidos de diferentes longitudes han indicado que el tamaño mínimo del sitio de unión a ARN-sc (es decir, el

número de nucleótidos que forman la interfaz NS3hel–ARN) en NS3hel es de 10-11 nucleótidos [26]. Por otro lado, luego del primer estudio realizado y publicado en donde se realizaron modelos de la NS3hel en tres estados del ciclo catalítico (*apo,* unida a ATP y unida a ADP y Pi, Figura 3.1C) [1], se demostró que la salida de fosfato ocurre antes que la salida de ADP [4]. En base a estas dos observaciones, se repitió el estudio de los mecanismos moleculares que modulan la interacción entre la NS3hel y el ARN, empleando modelos de la NS3hel unida a una cadena de ARN de 16 nucleótidos (NS3hel:ARN16, Figura 3.1D) y se incorporó el estado unido a ADP (Figura 3.1E) [2].

Cabe mencionar que a partir de simulaciones de la NS3hel unida a cadenas de 11, 16, 26 nucleótidos, cuyos resultados no están incluidos en esta tesis, se validaron los modelos con ARN de mayores longitudes que la resuelta (7 nucleótidos), demostrando que estos modelos sí son compatibles con el resultado experimental del tamaño del sitio de unión al ARN. Más aún, la unión de un ARN-sc lo suficientemente largo mostró ser necesaria para evidenciar la comunicación alostérica entre los sitios de unión a ATP y ARN de la NS3hel del DENV. Las simulaciones de dinámica molecular de estos sistemas y los detalles del análisis del tamaño del sitio de unión de la NS3hel al ARN para la comparación con los experimentos, así como la afinidad relativa de unión al ARN en presencia y ausencia de ATP con los diferentes largos de ARN forman parte de la tesis doctoral del estudiante de doctorado Fernando Amrein y están publicados [2].

3.3 Métodos

3.3.1 Preparación del sistema

La helicasa NS3 (NS3hel) del virus del dengue fue cristalizada unida a un fragmento de ARN de simple cadena (ARN-sc) de 7 bases en los distintos estados del ciclo de hidrólisis de ATP: sitio activo vacío (apo), ocupado por un análogo no hidrolizable de ATP (AMP-PNP), unido a ADP y unido a ADP e ion fosfato inorgánico (PO_4^{3-}) (PDB IDs: 2JLU, 2JLV, 2JLZ y 2JLY, respectivamente [23]).

Estas estructuras cristalográficas fueron utilizadas como configuraciones iniciales para cada estado de NS3hel unida a ARN-sc según la Tabla 3.1. Se revirtieron las mutaciones reportadas en cada entrada de la base de datos PDB (E250D, S292C, T321S, K381R y R480K en los tres sistemas; R480K en 2JLU y 2JLV; K213L en 2JLY). Para los sistemas unidos a una simple cadena de ARN de 16, los nucleótidos adicionales fueron agregados utilizando el programa *Visual Molecular Dynamics* (VMD) [27], moviendo copias de nucleótidos existentes a los extremos correspondientes del ARN y guardando las coordenadas. De esta manera, se agregaron 5 y 4 adenosinas en los extremos 5’ y 3’, respectivamente.

Los estados de protonación de los residuos ionizables se determinaron en función de los valores de pKa estimados para las cadenas laterales mediante el programa Propka 3.0 [28], un predictor empírico de pKa para residuos superficiales y enterrados en proteínas. Para el caso particular de los residuos histidina, además de estimar el pKa es importante definir el estado tautomérico, es decir, en qué átomo de nitrógeno va unido el protón, ya que admite dos posibilidades: Nδ o Nε. La elección del estado tautomérico de cada residuo de histidina fue realizada mediante una inspección visual del entorno del residuo y en todos los casos fue

adecuado la protonación en el Nε. Para el estado unido a ATP, la molécula de adenosín imido difosfato (AMP-PNP, análogo de ATP no hidrolizable) fue transformada en ATP cambiando el átomo de N por O, y los parámetros de ATP y ADP fueron tomados de la literatura [29]. La parametrización de HPO_4^{2-} se describe en la Sección siguiente, utilizando ajustes recomendados para los parámetros de Lennard-Jones reportados en la literatura [30]. La proteína y el ARN fueron modelados utilizando el campo de fuerza Amber ff14SB [31] con la corrección ILDN para residuos proteicos y parámetros diedros recientemente revisados para ARN [32].

Todos los sistemas fueron solvatados con el modelo de agua TIP3P [33] en cajas octaédricas con una distancia de 8 Å desde la superficie de la proteína. Esto resultó en la adición de aproximadamente 10,000 moléculas de agua TIP3P. Se empleó una concentración salina fisiológica de 0.15 M mediante la adición de aproximadamente 25 iones Na^+ y 25 Cl^-, después de neutralizar la carga total del sistema en cada caso (ver Tabla 3.1). Las aguas TIP3P, los hidrógenos y otros átomos no resueltos en las estructuras cristalográficas fueron añadidos utilizando el módulo tLeAP del paquete Amber [34].

Tabla 3.1. Detalle de los sistemas estudiados.

Sistema	Detalle	PDB ID	Carga (*e*)	Número de Réplicas (Tiempo de simulación por réplica, μs)	Tiempo total de simulación por sistema (μs)
apo	NS3hel:ARN7	2JLU	-1	6 (0,2)	1,2
	NS3hel:ARN16		-16	3 (1,0)	3,0
ATP	NS3hel:ARN7:ATP	2JLV	-3	6 (0,2)	1,2
	NS3hel:ARN16:ATP		-18	3 (1,0)	3,0
ADP-Pi	NS3hel:ARN7:ADP:HPO_4^{2-}	2JLY	-4	6 (0,2)	1,2
ADP-Pi	NS3hel:ARN16:ADP:HPO_4^{2-}		-19	3 (1,0)	3,0
ADP	NS3hel:ARN16:ADP		-17	3 (1,0)	3,0

3.3.2 Parametrización del ion fosfato

Para simular el sistema de productos que se obtendría luego de la hidrólisis de ATP, se generó un modelo a partir de la estructura cristalográfica de NS3hel con ARN simple cadena, ADP y fosfato inorgánico unidos (PDB ID 2JLY [23]). El fosfato inorgánico fue modelado en su estado dianiónico (Pi, HPO_4^{2-}). El modelo de cargas se obtuvo de manera consistente con los parámetros de Amber utilizados para modelar biomoléculas y ácidos nucleicos. Resumidamente, la geometría del ion Pi se optimizó con Gaussian 03 [35], utilizando la teoría de Hartree-Fock y el conjunto de bases 6-31G*, y se realizó un cálculo de *single-point energy*, cuyo procedimiento se basa en determinar la energía total del sistema para una geometría fija, sin modificar la disposición de los átomos durante el cálculo, para obtener el potencial electrostático. Posteriormente, se utilizó el procedimiento de Potencial Electroestático Restringido (RESP, por sus siglas en inglés) para derivar las cargas atómicas, según la

implementación del módulo Antechamber de AmberTools [34]. Además, se emplearon nuevos tipos atómicos, ‘OP’ y ‘OQ’, con radios de van der Waals mayores que los correspondientes tipos atómicos ‘O2’ y ‘OH’ de Amber, y con constantes de fuerza aumentadas para los ángulos OP-P-OQ y P-OQ-HO, tal como fue propuesto por Steinbrecher et al. [30]. Cabe destacar que, como se menciona en esta evaluación reciente de parámetros para fosfatos en simulaciones, el ácido fosfórico es una molécula difícil de modelar con métodos clásicos; los radios aumentados que se introdujeron aquí para el estado dianiónico se tomaron por analogía con los oxígenos fosfato en residuos de SER/THR fosforilados en su forma dianiónica. En conjunto con las cargas derivadas y las constantes de fuerza ajustadas, estos parámetros resultaron en simulaciones estables del sistema con los productos unidos.

3.3.3 Simulaciones de dinámica molecular

Para cada sistema, se realizaron simulaciones por triplicado con el módulo PMEMD cuda del paquete del programa Amber [36]. Cada réplica incluyó las tres etapas previas a la dinámica de producción (minimización, calentamiento y equilibración, ver Capítulo 2) que se describen a continuación.

Primero se realizaron 500 pasos de minimización con el algoritmo *steepest descent* y luego 1500 pasos más con el algoritmo *conjugate gradient*. Una vez que el sistema se relajó con la minimización de energía, el objetivo siguiente fue llevar al sistema a condiciones realistas de presión y temperatura, en este caso, 300 K y 1 bar. En un primer paso, se aumentó la temperatura de 0 a 200 K imponiendo fuertes restricciones en forma de potencial armónico con una constante de fuerza de 100 kcal/mol/Å^2 sobre los átomos de la proteína. De esta manera, se permitió el movimiento y se relaja únicamente la caja de solvente durante 50 ps. Análogamente, en un segundo paso se redujeron las restricciones sobre la proteína a 50 kcal/mol/Å², para aumentar la temperatura de 200 K a 300 K, nuevamente durante 50 ps. Ambos pasos de calentamiento se llevaron a cabo a volumen constante. Finalmente, se realizó una equilibración en dos pasos y sin restricciones. En primer lugar, una vez que el sistema llegó a 300 K, se equilibró el sistema durante 50 ps a temperatura y volumen constante. En segundo lugar, se equilibró el sistema durante 100 ps a 1 bar de presión y temperatura constante. En todos los pasos de calentamiento y equilibración, el paso de tiempo utilizado fue de 1 fs.

Tanto en las etapas de calentamiento y equilibración como para la producción de la trayectoria, se controló la temperatura del sistema mediante con el termostato de Langevin [37] usando una frecuencia de colisión de 1.0 ps^{-1}. Además, el segundo paso de equilibración a presión constante, se llevó a cabo con el barostato de Berendsen [38] y un tiempo de relajación de 2.0 ps. Para disminuir el costo computacional, se empleó el algoritmo SHAKE [39] que restringe los enlaces con hidrógenos no polares y permite usar un paso de tiempo mayor (1-2 fs) y como límite en la consideración de las interacciones de no unión, se estableció una distancia de corte de 8 Å, lo que significa que no se contarán las interacciones entre dos partículas que superen esa distancia.

Para los sistemas de NS3hel:ARN7, se simularon seis réplicas independientes de 0.200 μs para cada estado, y en tres de las seis réplicas de cada sistema se conservaron las aguas cristalográficas originales. La decisión de simular los mismos sistemas partiendo de configuraciones sin las aguas cristalográficas se basa la búsqueda de aumentar la

exploración de la superficie de energía potencial, ya que este cambio genera un cambio en el punto de partida de las simulaciones más grande que el cambio de modificar la asignación aleatoria de las velocidades iniciales, como ocurre entre las tres réplicas para las cuales se conservan las aguas cristalográficas.

Para los de NS3hel:ARN16, se realizaron tres simulaciones independientes para cada sistema, conservando las aguas cristalográficas, de 1 μs cada una. Toda la producción se obtuvo bajo las mismas condiciones de NVT mencionadas anteriormente, utilizando un paso de tiempo de 2 fs.

3.3.4 Análisis de las dinámicas

En el Capítulo 2 se describen en detalle los métodos de análisis aplicados a las simulaciones de dinámica molecular de este capítulo. Con el programa CPPTRAJ del paquete Amber [40], se calculó la función de distribución radial $g(r)$ a partir del histograma del número de átomos de oxígeno de las moléculas de agua alrededor del átomo de fósforo del fosfato, Pi. Por otro lado, con el mismo programa, se calculó el RMSF para los aminoácidos y luego se empleó el método *bootstrap* para estimar la incertidumbre.

3.3.5 Cálculos de energía de unión ARN-proteína

Se realizó el cálculo de la energía libre de unión para los sistemas compuestos por ARN de 7 y 16 nucleótidos. Para el primero, se tomaron los últimos 100 ns de un total de 200 ns por cada réplica, analizado un frame cada 0,2 ns. Para el segundo, se tomaron los últimos 800 ns de un total de 1000 ns por cada réplica, analizado un frame cada 10 ns. Para ambos largos de ARN se analizaron los tres o cuatro estados de ocupación del sitio catalítico simulados, respectivamente (Tabla 3.1).

Para estimar la energía libre de unión (ΔGu) entre la proteína y el ARN para cada sistema se empleó el método MM-GBSA (por sus siglas en inglés provenientes de *Molecular Mechanics - Generalized Born Surface Area;* ver Capítulo 2, Sección 2.3.3). A modo de resumen, se calculó la energía libre de unión, con el programa Amber [36], que emplea un modelo GB mejorado [41] para aplicar específicamente a sistemas compuestos por proteínas y ácidos nucleicos, además de contar con la ventaja de ser de bajo costo computacional. Este método involucra, por un lado, el cálculo por separado la energía libre de solvatación de la proteína (en este caso se incluye el ATP y Mn(II) o el ADP, P y Mn(II), según corresponda), del ligando (en este caso, del ARN) y del complejo proteína-ligando. El medio acuoso se modela con solvente implícito y una concentración de sal de 0,15 M. Por otro lado, se calcula la entalpía de unión en fase gaseosa o vacío a partir de las contribuciones de no enlace a la energía potencial. Al comprar la energía libre de unión entre varios sistemas similares se está evaluando el $\Delta\Delta Gu$, en donde se modifica la ocupación del sitio catalítico pero se mantiene la identidad de la proteína y el ARN. Por lo tanto, es razonable suponer que la contribución entrópica será similar y se cancelará, lo cual resulta conveniente por el gran costo computacional que conlleva su estimación. Habiendo dicho esto, si existiera un efecto entrópico diferencial que afecta la unión entre la proteína y el ARN por la ocupación del sitio catalítico, se estaría perdiendo por esta aproximación.

3.3.6 Calorimetría de titulación isotérmica

En el laboratorio del Dr. Sergio Kaufman, cuyo grupo realizó los experimentos que se detallan a continuación, se obtuvo un constructo que codifica el dominio helicasa de NS3 (residuos 171–618) a partir del cDNA de un clon infeccioso del serotipo 2 de DENV (número de acceso en GenBank: U87411) [42]. Se realizaron titulaciones calorimétricas entre NS3hel y el ARN-sc p-5′-AGUUGAGUUG-3′ (p-R10) en ausencia y presencia de ADP (0,18 M) para obtener las constantes de equilibrio de unión que caracterizan la interacción NS3hel–ARN-sc en estos dos estados. Los experimentos de ITC y espectroscopía se llevaron a cabo a 25 ºC y pH 6.5 en presencia de 0,11 M de cationes monovalentes.

3.4 Resultados

3.4.1 Cambios en la flexibilidad de la proteína durante el ciclo de hidrólisis de ATP

El primer objetivo fue evaluar los cambios en la flexibilidad de la proteína debidos a la presencia de ATP y a los productos tras la hidrólisis. Para esto se analizaron las simulaciones clásicas de dinámica molecular realizadas para del dominio helicasa de NS3 unido a un fragmento de 7 nucleótidos de ARN (NS3hel:ARN7) en los estados apo, unido al ATP y unido a productos (Figura 3.1C), partiendo de las estructuras cristalográficas disponibles para cada estado [23]. También se simuló la misma proteína unida a un fragmento de 16 nucleótidos de ARN-sc (NS3hel:ARN16) en los mismos estados y en el estado unido a ADP (Figura 3.1 E).

Utilizando un análisis de *bootstrapping* sobre el total de 1.2 μs simulados para cada estado NS3hel:ARN7 y 3 μs simulados para cada estado NS3hel:ARN16, se calcularon las fluctuaciones cuadráticas medias por residuo (RMSF, por sus siglas en inglés. Ver Capítulo 2, Sección 2.3.1) con respecto a la estructura promedio correspondiente. Posteriormente, se analizaron las diferencias de RMSF (ΔRMSF) entre estados sucesivos del ciclo catalítico, para comparar el efecto de la unión del ATP (ATP - apo), la hidrólisis (ADP:Pi - ATP), y en el caso del ARN16, la liberación de fosfato (ADP:Pi - ADP).

Para ambos largos de ARN, tras la unión de ATP al complejo NS3hel:ARN, se observa una pérdida en la flexibilidad de la proteína, reflejada por los picos negativos en la Figura 3.2B y 3.2C (serie negra), y esta diferencia es más marcada en presencia del ARN largo. En particular, esta estabilización ocurre en regiones próximas al sitio activo, como los motivos conservados I (perteneciente a D1) y VI (perteneciente a D2), el *loop* de unión al ARN (perteneciente a D1) y el *loop* β6-β1’ (resaltados en la Figura 3.2A) que conecta los subdominios D1 y D2. Además, en los sistemas NS3hel:ARN16 se observó la desestabilización de otras dos regiones a partir de la unión de ATP: el *loop* de unión al ARN y una zona del subdominio D3 que consiste en un *loop* no estructurado junto con el α-hélice 2’’ (resaltados en la Figura 3.2A). Este último efecto resulta particularmente interesante ya que el subdominio D3 se encuentra alejado del sitio catalítico.

Al comparar el estado unido a productos ADP:Pi con el estado unido al ATP, se observó el efecto opuesto: el mantenimiento o la recuperación de flexibilidad en las regiones previamente estabilizadas por la presencia de ATP, incluyendo los motivos conservados I y VI y el *loop* β6-β1’ (serie gris de las Figuras 3.2 B y C). Análogamente, las dos regiones que se habían flexibilizado en presencia de ATP para los sistemas unidos al ARN16, vuelven a estabilizarse en presencia de productos de hidrólisis. Al liberarse el fosfato, generando así el sistema

NS3hel:ARN16 unido a ADP, se observó nuevamente la pérdida de flexibilidad para la zona del motivo I y el *loop* β6-β1', similar al efecto de unión de ATP. En cambio, para el motivo VI y la región señalada del subdominio D3 no se observaron diferencias significativas con el estado unido a productos, mientras que, en esta misma comparación, el *loop* de unión al ARN presentó una pérdida de flexibilidad.

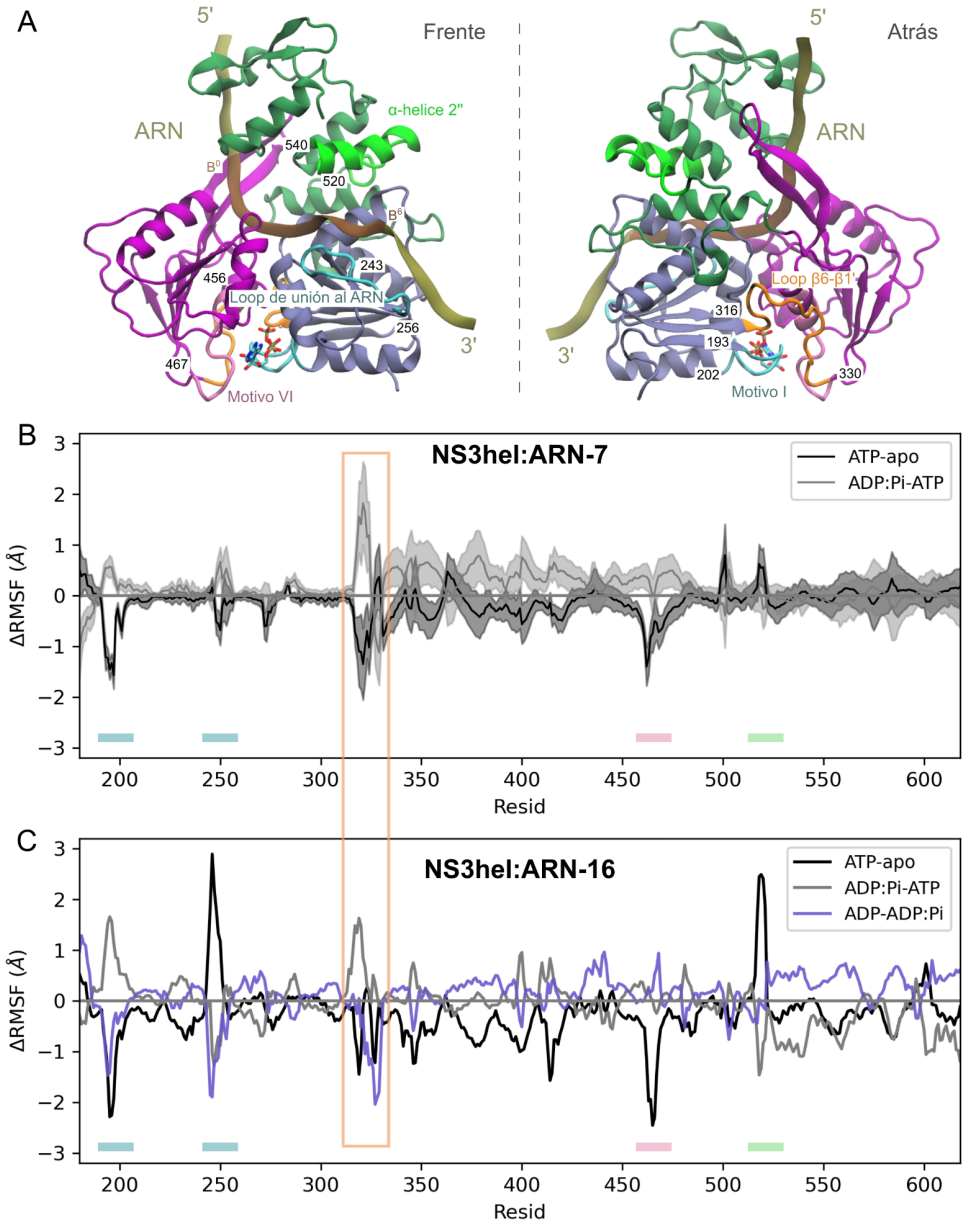


Figura 3.2. Fluctuaciones en el esqueleto de la helicasa NS3 del virus del dengue en diferentes etapas del ciclo de hidrólisis de ATP. **A)** Estructura de la proteína NS3hel coloreada según los subdominios (D1 en azul, D2 en violeta y D3 en verde. La estructura de la derecha representa una rotación sobre el eje vertical de 180°. **B)** ΔRMSF (Å) para el estado de la NS3hel:ARN7 unido al ATP con respecto al estado *apo* (ATP menos *apo*; serie en negro) y para el estado unido a productos con respecto al estado ATP (productos menos ATP; serie gris). Las barras de error para ambas series fueron calculadas mediante un análisis de *bootstrapping*, como se detalla en los métodos. **C)** ΔRMSF (Å) para el estado de la NS3hel:ARN16 unido al ATP con respecto al estado *apo* (ATP menos *apo*; serie en negro) y para el estado unido a productos con respecto al estado ATP (productos menos ATP; serie gris) y para el estado unido a ADP con respecto al estado unido a ADP:Pi (ADP-productos; serie violeta azulada).

3.4.2 Nueva conformación de *loop* β6-β1’ abierto observada en ausencia de ATP

La mayor flexibilidad del *loop* β6-β1’ en los estados apo y unido a productos merece atención adicional, ya que este cambio conformacional (Figura 3.3A) da lugar a la apertura de un canal estrecho que resulta en el aumento de la solvatación del Pi, como se observa en la función de distribución radial (RDF, por sus siglas en inglés) de Pi y la cantidad de átomos de oxígeno de las moléculas de agua para las conformaciones de *loop* abierto y cerrado (Figura 3.3B), tomando dos configuraciones representativas del sistema NS3hel:ARN7:ADP:Pi, una para cada caso. Visualmente, este loop parecería actuar como una "puerta trasera" del sitio activo (Figura 3.3C).

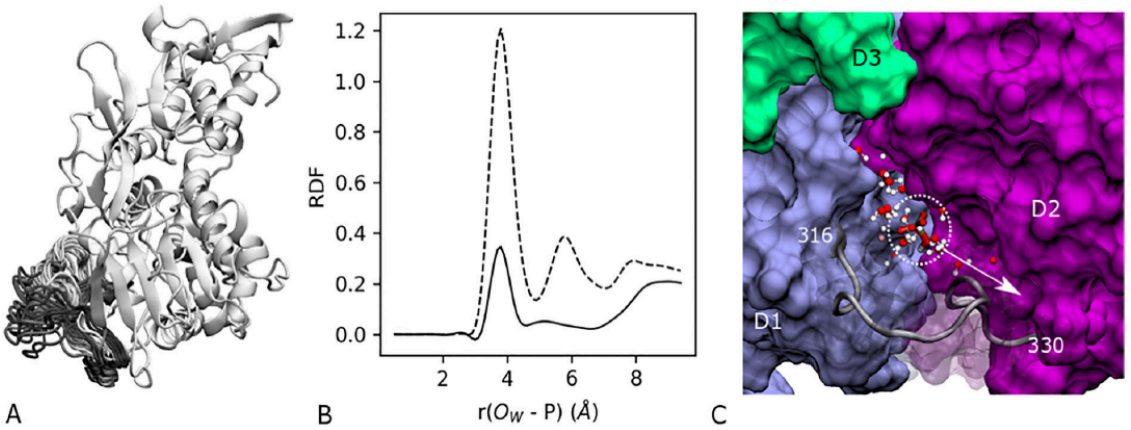


Figura 3.3. Nueva conformación de *loop* abierto. **A)** Conformaciones superpuestas del *loop* β6-β1’ durante un evento de apertura, coloreadas temporalmente de gris a negro. NS3hel está rotada 90 grados sobre su eje vertical respecto a la Figura 3.1, las imágenes fueron tomadas de una simulación del estado unido a productos. **B)** Función de distribución radial (RDF, por sus siglas en inglés) de Pi y átomos de oxígeno de moléculas de agua para las conformaciones abierta (serie punteada) y cerrada (serie sólida). **C)** El detalle de la conformación del *loop* β6-β1’ abierto revela un canal o “puerta trasera” que conecta el sitio activo con el solvente. La proteína está representada con su superficie accesible al solvente, con el mismo esquema de colores que en la Figura 3.1, rotada 180 grados sobre el eje vertical; el *loop* β6-β1’ se muestra como una cinta gris, con los números de sus residuos terminales identificados. Se muestran las moléculas de agua (esferas rojas y blancas), el fosfato (bastones color marrón y rojo, rodeadas por círculo punteado) y el Mn (esfera rosa). La flecha blanca indica una posible ruta de salida para el Pi.

Además de una nueva posible ruta de salida para el ion fosfato post-hidrólisis, se evaluó si la dinámica del *loop* β6-β1’ podría influir en los sitios de unión del ARN o ATP, los cuales se forman a través de los subdominios D1 y D2. Para obtener primero una métrica cuantitativa para la clasificación de la conformación del *loop* β6-β1’, se realizó un análisis de componentes principales (PCA, por sus siglas en inglés) de los residuos del esqueleto del loop, concatenando las trayectorias obtenidas de los tres sistemas de estudio de la NS3hel unida al ARN7. Los detalles de esta metodología se encuentran en el Capítulo 2, Sección 2.3.2. El primer componente principal o PC, por sus siglas en inglés (PC1 del loop β6-β1’, porcentaje de varianza de 41,3%) resultó ser un adecuado descriptor para la apertura y cierre del loop.

La proyección de las estructuras de *loop* cerrado sobre PC1 da como resultado valores negativos, mientras que la proyección de las estructuras semiabiertas y abiertas del loop da como resultado valores positivos cada vez mayores, lo que convierte al PC1 del *loop* β6-β1' en una métrica útil para evaluar su conformación.

Para medir los cambios en el modo de unión del ARN, se calculó a la distancia entre el centro de masa de la Arg 225 y la Thr 244, ambos pertenecientes a D1, y el centro de masa de la Arg 387 y la Thr 408, ambos pertenecientes a D2. Esta distancia inter-subdominio entre los dos pares de "anclas" conservadas de ARN se denominó d_{ARN}. De manera similar, para monitorear los cambios en el sitio activo durante la hidrólisis de ATP, se recurrió a una segunda distancia inter-subdominio, d_{ATP}, calculada como la distancia entre los centros de masa de los motivos conservados I y III (pertenecientes a D1) y VI (perteneciente a D2) detallados a continuación. Del motivo I, se seleccionaron los residuos 195, 196, 197, mientras que del motivo III, 316 y 317 y del motivo VI, los residuos desde 458 a 463. Estos motivos conservados de la helicasa son conocidos por estar implicados en la unión e hidrólisis de NTP y se ubican a ambos lados del sitio activo.

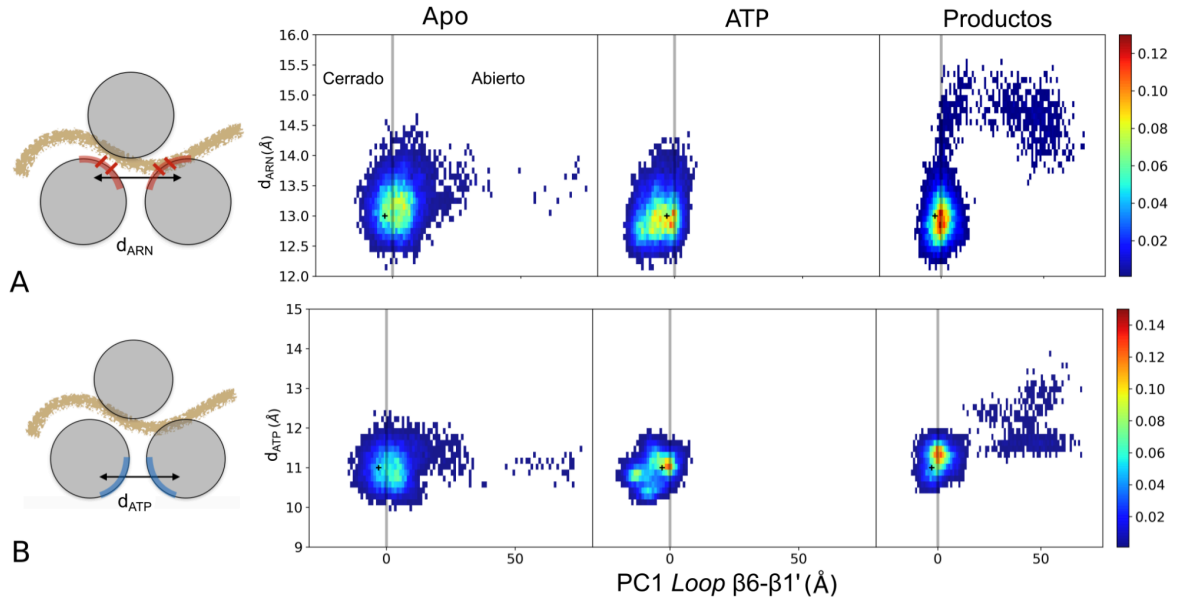


Figura 3.4. Análisis conformacional bidimensional del *loop* β6-β1' y separación de los subdominios 1 y 2 durante la hidrólisis de ATP**.** Las trayectorias para los estados apo, ATP y productos se caracterizaron de acuerdo con el primer componente principal de la dinámica del esqueleto del *loop* β6-β1' y **(A)** la separación de los residuos de unión al ARN en los subdominios 1 y 2 (d_{ARN}) y **(B)** la separación de los residuos de unión al ATP en los subdominios 1 y 2 (d_{ATP}). Las rayas rojas representan las posiciones de Arg 225 y Thr 244 en D1, y Arg 387 y Thr 408 en D2, residuos que interactúan directamente con el ARN. Las regiones involucradas en la unión directa con ARN o ATP se indican esquemáticamente con arcos rojos y azules, respectivamente.

Utilizando la métrica del PC1 del *loop* β6-β1' y las dos distancias inter-subdominio definidas anteriormente, se construyeron las densidades de probabilidad en 2D para los estados *apo*, ATP y productos. En la Figura 3.4A se muestra la dinámica del *loop* β6-β1' en función de la distancia inter-subdominio de unión al ARN (d_{ARN}), mientras que en la Figura 3.4B, la misma

dinámica del *loop* se muestra en función de la distancia inter-subdominio de unión al ATP (d_{ATP}).

De ambos paneles de la Figura 3.4, se observa claramente que la unión de ATP favorece fuertemente una conformación de *loop* cerrado, mientras que tanto los estados apo como los estados unidos a productos exploran conformaciones de *loop* abierto. Además, estos mismos estados exploran un rango ligeramente más amplio de distancias d_{ARN} y d_{ATP} que en el estado NS3hel:ARN7:ATP, sin una correlación específica entre la distancia inter-subdominio y la dinámica del *loop*. Por otro lado, el análisis separado de estas variables en la Figura 4A y 4B en presencia de los productos de la hidrólisis sugiere una secuencia de cambios conformacionales: las fluctuaciones aumentadas en la distancia del sitio de unión al ARN (d_{ARN}) anteceden a la apertura del *loop* β6-β1' (Figura 3.4A), y el sitio de unión al ATP se ensancha en respuesta a la apertura del *loop* en el estado de productos (Figura 3.4B). En otras palabras, no es posible observar la apertura del *loop* β6-β1' si los subdominios D1 y D2 no alejan sus anclas entre sí, modificando la configuración inter-subdominio en la región de unión al ARN.

3.4.3 Movimientos relativos de los subdominios a partir del aumento de la flexibilidad conformacional

La apertura del *loop* β6-β1' fue mucho más evidente para el sistema de productos ADP:Pi del complejo NS3hel:ARN7, y más específicamente para una de las réplicas cuyas configuraciones iniciales se realizaron en ausencia de aguas cristalográficas, alcanzando el objetivo buscado del aumento de la exploración conformacional sobre la superficie de energía potencial. En este sentido, como las simulaciones del complejo NS3hel:ARN16 se hicieron a partir de configuraciones iniciales en presencia de las aguas cristalográfica y sólo se observaron diferencias sutiles (esto se evidencia en la menor diferencia de fluctuaciones de la Figura 3.2), no se repitió el análisis de PCA para estos sistemas. En su lugar, se evaluaron los histogramas 2D entre d_{ARN} y d_{ATP} para cada uno de los cuatro estados del ciclo catalítico de hidrólisis de ATP del complejo NS3hel:ARN16 (Figura 3.5).

En el estado apo, se observó que la proteína explora un amplio rango de valores de d_{ARN} (de 12 a 16 Å), así como valores de d_{ATP} (de 8 a 15 Å), siendo más favorables las distancias mayores de d_{ATP}. Al unirse el ATP, mientras los valores de d_{ATP} se restringen sustancialmente (agrupados alrededor de 9 Å), se observan dos poblaciones en la dimensión de d_{ARN}, siendo la más probable, la que posiciona a las "anclas" del ARN a una distancia de alrededor de 13 Å, menor que esa distancia en la configuración alternativa. Después de la hidrólisis, se observó que estas dos poblaciones persisten, pero la probabilidad de explorar valores más grandes de d_{ARN} aumenta, y además, se recupera la flexibilidad en la dimensión de d_{ATP}.

Consistentemente con la observación de mayores fluctuaciones del *loop* β6-β1' tanto para los complejos unidos al ARN de 7 nucleótidos como los de 16, es en esta etapa del ciclo catalítico (en presencia de ambos productos de hidrólisis), cuando el sitio de unión al ATP vuelve a explorar una conformación más abierta, así como también lo hacía en el estado apo. En esta etapa de productos, mientras ocurre esta apertura del sitio catalítico, el sitio de unión al ARN se muestra siempre en conformación cerrada (menores distancias d_{ARN}) y viceversa, evidenciando la correlación entre estos movimientos por la falta de población de altas fluctuaciones para d_{ARN} y d_{ATP} a la vez. En contraste, después de la liberación del Pi, se pierde

algo de flexibilidad nuevamente y aparece un nuevo grupo conformacional con valores intermedios de ambas distancias al mismo tiempo.

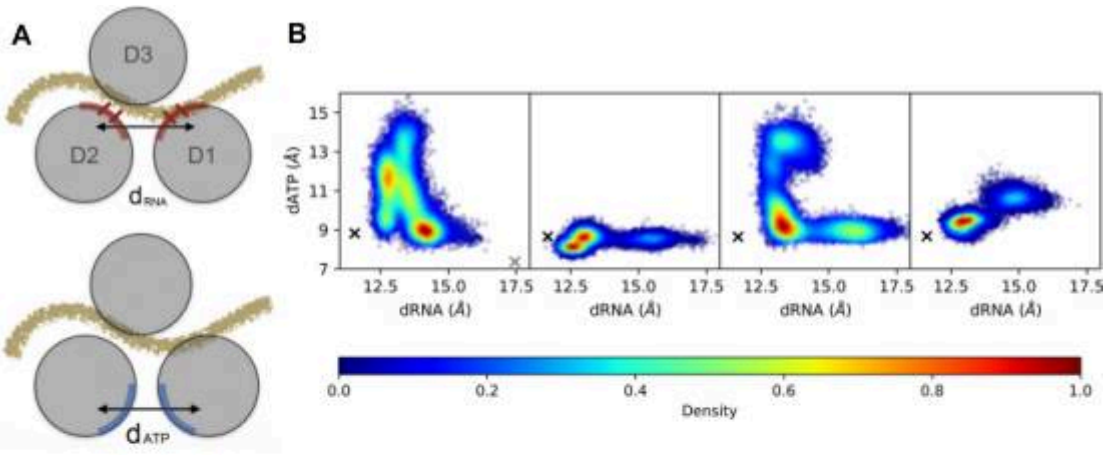


Figura 3.5. A) Representación esquemática de la proteína que muestra las distancias d_{RNA} y d_{ATP}. Se indican los subdominios D1, D2 y D3. **B)** Histograma 2D para d_{ATP} *versus* d_{RNA}. Los complejos NS3hel:ARN16, de izquierda a derecha, son: apo, unido a ATP, unido a ADP-Pi y unido a ADP. Las cruces negras muestran las distancias correspondientes en cada estructura cristalográfica, y la cruz gris muestra también esas distancias para la proteína en ausencia de ARN.

Como se mencionó anteriormente, los cristales unidos a ARN no muestran casi ninguna diferencia entre las diferentes etapas del ciclo de hidrólisis (Figura 3.5b, cruces negras), y los valores para d_{ATP} corresponden a la conformación que se observa frecuentemente a ~9 Å. Sin embargo, no ocurre lo mismo en todos los casos para los valores de d_{ARN}, en donde estas distancias se alejan notablemente del valor cristalográfico para los sistemas apo y en presencia de ambos productos de hidrólisis. En el cristal apo en ausencia de ARN, se observa una longitud de d_{ARN} cercana a 17 Å (Figura 3.5b, cruz gris) cuyo valor resulta cercano a la conformación observada en estados más flexibles.

3.4.4 Interacciones proteína-ARN durante el ciclo de hidrólisis de ATP

Dado que los mecanismos propuestos de translocación para la helicasa NS3 de HCV, estrechamente relacionada, involucran cambios específicos en las interacciones entre la proteína y los enlaces fosfodiéster del ácido nucleico, a continuación se analizan las interacciones proteína-ARN (Figura 3.6A) en cada uno de los estados simulados. Para los tres sistemas NS3hel:ARN7, el análisis se limitó a las interacciones de puentes salinos o enlaces de hidrógeno presentes en al menos el 50% de las trayectorias obtenidas en cada estado de los complejos (Figura 3.6B), mientras que para los cuatro sistemas NS3hel:ARN16, el análisis se limitó a las interacciones con los residuos definidos en d_{ARN} (Arg 225 y Thr 244 en D1, y Arg 387 y Thr 408 en D2, Figura 3.6C).

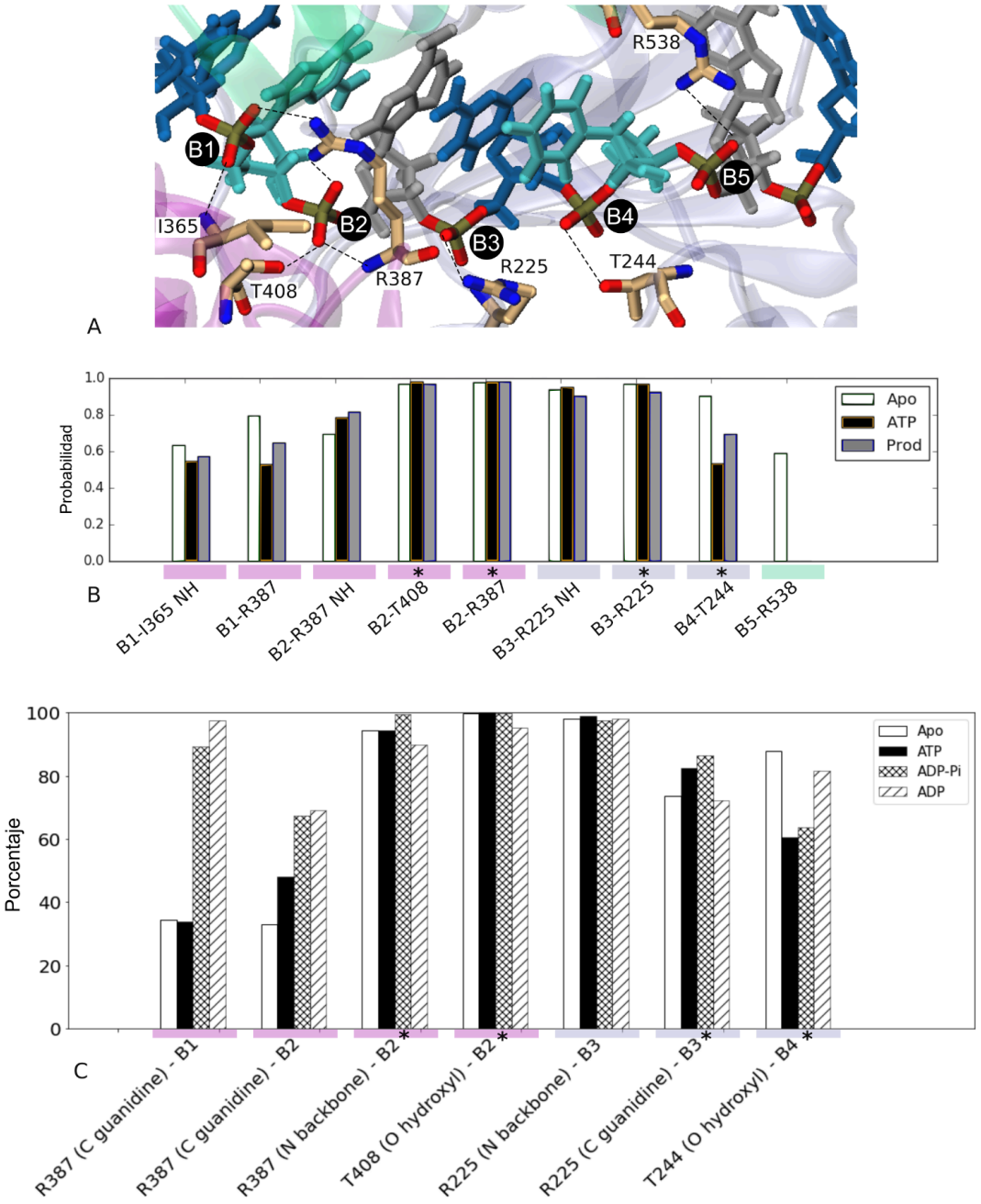


Figura 3.6. Interacciones con el esqueleto fosfodiéster del ARN. **A)** Representación esquemática del sitio de unión al ARN. En una representación de cintas, la proteína está coloreada con el mismo esquema de colores que en la Figura 3.1. Los residuos relevantes de la proteína están etiquetados y coloreados según el tipo de átomo. Los nucleótidos involucrados en interacciones con la proteína están numerados según la base (B1-B5). Las líneas discontinuas representan las interacciones proteína-ARN. La imagen fue tomada de una simulación en el estado apo. **B)** Fracción de la simulación combinada de 1.2 μs en la que se observó cada interacción para los estados *apo*, ATP y productos en los complejos NS3hel:ARN7. Las interacciones están coloreadas según el subdominio. Las interacciones señaladas con asteriscos corresponden a la unión más conservada de cada cadena lateral de los residuos de unión al ARN considerados en dARN. **C)** Porcentaje de cada interacción de los residuos residuos de unión al ARN considerados en dARN. Las interacciones señaladas con asteriscos corresponden a la unión más conservada de cada cadena lateral. Los fosfatos están numerados según la base (B1-B4). Análisis realizado para los estados *apo*, ATP y productos y ADP en los complejos NS3hel:ARN16.

Para los tres estados catalìticos de NS3hel:ARN7, casi todas las interacciones de enlaces de hidrógeno/puentes salinos proteína-ARN se formaron con los subdominios D1 o D2. De hecho, la única interacción entre el esqueleto fosfodiéster y D3 que cumplió con nuestros criterios se observó exclusivamente en el estado apo del complejo NS3hel:ARN7 (Figura 3.6 A y B). Sin contar esta última interacción formada con el subdominio D3 en el estado apo, todas las interacciones fueron observadas en los tres estados. Estas interacciones involucran principalmente los dos pares conservados de Thr/Arg en D1 y D2, mencionados previamente en la definición de d_{ARN}, que interactúan con B0 y B1 en el subdominio D2 y B2 y B3 en el subdominio D1. Además de las interacciones con estas cuatro “anclas”, se observó otra interacción conservada entre el NH del esqueleto de Ile 365 y B0 en los tres estados. Para B0 y B3, hay una clara tendencia a la disminución de las interacciones específicamente en presencia de ATP, que se recupera ligeramente tras la hidrólisis. A pesar de la ausencia de cambios claros en las interacciones con el esqueleto fosfodiéster debido a la unión de ATP y la hidrólisis, los resultados para el estudio de los complejos NS3hel:ARN7 indican que las interacciones proteína-fosfodiéster están más desestabilizadas hacia el extremo 3′ del segmento de ARN (notar que los nucleótidos B4, B5, B6, a excepción de la interacción con B4 para el sistema apo, no tienen interacciones estables con la proteína) que hacia el extremo 5′.

Es relevante retomar en este punto la evidencia existente sobre estudios termodinámicos de la unión entre la helicasa y oligonucleótidos de diferentes longitudes, que muestran que el tamaño mínimo del sitio de unión al ARN-sc en NS3hel del DENV es de 10-11 nucleótidos [26]. Aunque las estructuras cristalográficas disponibles de NS3hel fueron obtenidas en complejo con un ARN de 12 nucleótidos, solo se resolvieron 7 nucleótidos [23], de los cuales únicamente cuatro (B1-B4) parecen formar interacciones estables con la proteína. Al evaluar computacionalmente el tamaño del sitio de unión a ARN (es decir, el número de nucleótidos que forman la interfaz NS3hel-ARN), empleando cadenas más largas de oligonucleótidos (específicamente, ARN-sc de 11, 16 y 26 nucleótidos), se encontró que este es efectivamente mayor que los 7 nucleótidos resueltos [2]. Los cambios en el tamaño de sitio de unión al ARN se analizan en la Sección 3.4.5.

Los resultados obtenidos al alargar la cadena simple de ARN, empleando 16 nucleótidos, muestran cambios más evidentes, en comparación al ARN de 7 nucleótidos, en los porcentajes de las interacciones entre el residuo básico Arg 387 y las bases B1 y B2 en el extremo 5’. Tanto en el estado apo como unido a ATP de la NS3hel:ARN16, estas interacciones se ven desestabilizadas, mientras que en ambos estados de productos de hidrólisis, se estabilizan. Por el contrario, y consistentemente con lo observado para NS3hel:ARN7, el estado apo estabiliza las interacciones en el extremo 3’ del ARN, evidenciado en el mayor porcentaje obtenido para la interacción entre la Thr 244 y B4.

3.4.5 Cambios en el tamaño del sitio de unión al ARN según la ocupación del sitio catalítico

El análisis exhaustivo y la metodología empleada para la determinación computacional del sitio de unión del ARN en la NS3hel no está incluido en esta tesis, ya que fueron desarrollados por el estudiante de doctorado Fernando Amrein. Estos se basan en el cálculo de la superficie accesible al solvente de los nucleótidos, habiendo establecido un valor límite o umbral a partir del cual un nucleótido se considera unido a la proteína y de esta manera, forma parte del sitio de unión. Estos detalles se encuentran publicados [2]. Al haber

determinado computacionalmente que es necesaria una cadena de ARN más larga para modelar correctamente el complejo NS3hel:ARN, el análisis de los cambios en el sitio de unión del ARN en la NS3hel en todas las etapas del ciclo de hidrólisis de ATP solamente se realizó para los complejos con ARN de 16 nucleótidos (Figura 3.7A). Sin embargo, los sistemas apo y en presencia de ATP cuentan con determinaciones análogas para ambos largos de ARN, por lo que se mencionan en la siguiente descripción a modo de comparación.

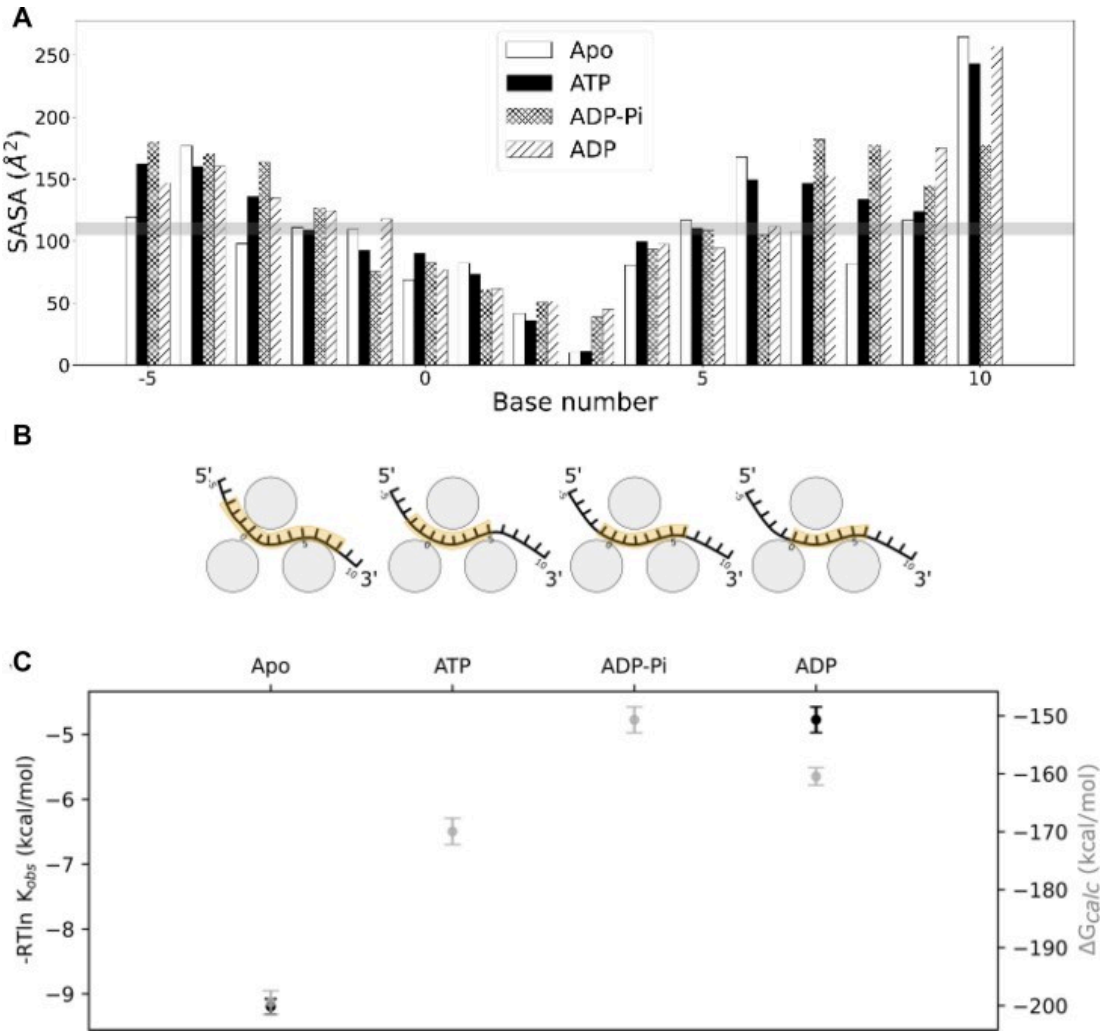


Figura 3.7. A) Gráficos de barras de superficie accesible al solvente para las distintas cuatro etapas del ciclo catalítico estudiadas de los complejos NS3hel:ARN16. Una línea horizontal gris indica el umbral para determinar las bases consideradas como parte del sitio de unión al ARN. **B)** Representación esquemática de NS3hel unido al oligonucleótido de 16 nucleótidos, indicando en amarillo el tamaño del sitio de unión al ARN para los estados apo, unido a ATP, unido a productos, ADP y Pi, y unido a ADP (de izquierda a derecha). Se indican los extremos 5′ y 3′ del ssRNA. **C)** Energías de unión para el complejo NS3hel:ARN16 en los mismos cuatro estados que los analizados en el panel A y B mostradas con círculos grises, junto con las energías experimentales calculadas a partir de constantes de equilibrio Kobs, indicadas con círculos negros, con sus respectivas barras de error.

En el complejo NS3hel:ARN16 apo, el tamaño del sitio de unión al ARN es de aproximadamente 12 nucleótidos (nts) (B-3 a B8), y la presencia de ATP en el sitio catalítico reduce este tamaño hasta 8 nts (B-2 a B5) (Figura 3.7 A y B). Usando la misma metodología,

el tamaño de sitio obtenido para los complejos NS3hel:ARN7 apo y unido a ATP fue de 5 nucleótidos [2], similar a los 5 y 4 nucleótidos señalados en el análisis de interacciones, respectivamente. Volviendo a los sistemas NS3hel:ARN16, aunque la presencia de ADP-Pi no causa un cambio en el tamaño del sitio de unión al ARN, el sitio se desplaza en una base hacia el extremo 3′ (B-1 a B6) (Figura 3.7 A y B). La unión de ATP viene acompañada de un aumento en la flexibilidad del extremo 3′ que se conserva tanto en los complejos unidos a ADP-Pi como en los unidos a ADP, mientras que el extremo 5′ parece estabilizarse después de la hidrólisis y luego desestabilizarse nuevamente con la liberación del primer producto de la hidrólisis de ATP, Pi. Con respecto al extremo 3' del ARN, la mayor flexibilidad del *loop* de unión al ARN en presencia de ATP se conserva en presencia de los productos de la hidrólisis, pero después de la liberación del fosfato, este *loop* se estabiliza nuevamente (Figura 3.2C).

3.4.6 Energía libre de unión entre la NS3hel y el ARN

Usando experimentos de calorimetría de titulación isotérmica (ITC, por sus siglas en inglés) y espectroscopía, el grupo del Dr. Kaufman obtuvo las constantes de equilibrio para la unión de la NS3hel a un ARN de 10 nucleótidos en ausencia ($5 \pm 1 \times 10^{6}\ M^{-1}$) y en presencia ($3 \pm 1 \times 10^{3}\ M^{-1}$) de ADP (Figura 3.6C). Estos resultados concuerdan con la existencia de la comunicación alostérica entre los sitios de unión al NTP y ARN, ya que la presencia de ADP resulta en una reducción de la constante de asociación de equilibrio para la unión NS3hel–ARN.

En la Figura 3.7C, se muestra la energía libre de unión NS3hel–ARN estimada para los cuatro estados (apo, unido a ATP unido a productos, ADP-Pi, y unido a ADP) de los complejos NS3hel:ARN16, utilizando el método MM-GBSA (Sección 3.3.5 y Capítulo 2, Sección 2.3.3). Se observa una reducción en la afinidad por el ARN-sc en presencia de ATP, en comparación con el complejo apo, y energías de unión similares para los complejos unidos a ADP, ADP-Pi y ATP. Estos resultados indican al menos dos estados energéticos del complejo NS3hel–ARN: un estado fuertemente unido con un sitio catalítico de NTP libre, en el que el tamaño del sitio de unión al ARN de la NS3hel es de aproximadamente 10 nts; y un estado débilmente unido, en el que este tamaño es de 8 nucleótidos, observado al unirse ATP y después de la hidrólisis. Las energías libres de unión calculadas para los cuatro complejos modelados (apo, unido a ATP, unido a productos, ADP y P, y unido a ADP) indican que el complejo NS3hel–ARN apo está más fuertemente unido que el complejo con el sitio catalítico ocupado, lo que también es consistente con la diferencia observada en las constantes de asociación de equilibrio medidas experimentalmente para los complejos apo y ADP-unido, representando al complejo proteína-ARN con el sitio catalítico ocupado.

Cabe destacar que lo interesante de este análisis está en la tendencia observada del aumento de la energía libre de unión al estar ocupado el sitio catalítico que es consistente en ambos enfoques, teórico y experimental, lo cual significa estar analizando implícitamente el valor del $\Delta\Delta G$ entre los sistemas (ver Sección 3.3.5). El mismo cálculo realizado para comparar la energía libre de unión entre los estados apo y unido a ATP del complejo NS3hel:ARN7 mostró que no hay diferencias significativas [2], reforzando el resultado obtenido a partir del cálculo del tamaño del sitio de unión del ARN que sugería que este modelo no es adecuado para representar el complejo proteína ARN de la helicasa NS3 del DENV.

3.5 Discusión

3.5.1 Alternancia entre estados flexibles y rígidos durante la hidrólisis de ATP

Las conformaciones obtenidas en las simulaciones de dinámica molecular para los distintos estados de la NS3hel, unida tanto a una simple cadena de ARN7 como de ARN16 durante el ciclo de hidrólisis de ATP, indican que la helicasa NS3 del DENV alterna entre conformaciones flexibles y rígidas durante la hidrólisis de ATP (Figura 3.2). La flexibilidad observada en ausencia de ATP (es decir, en los estados apo y unido a productos) habilita, por un lado, la apertura del *loop* β6-β1' (Figuras 3.3 y 3.4), y por otro, interacciones más estables en el extremo 3' (Figura 3.6).

Para el análisis de los complejos NS3hel:ARN16, cuando se tiene en cuenta la distancia entre residuos clave de unión al ARN (d_{ARN}) o residuos clave de unión al ATP (d_{ATP}) en los dos subdominios D1 y D2, se observa un claro cambio de población en términos de separaciones durante el ciclo de hidrólisis de ATP (Figura 3.5). En términos de la separación de los residuos de unión al ARN, se observan aproximadamente dos poblaciones en ~14.5 y 12.5 Å, 'relajado' y 'tenso', respectivamente. Curiosamente, las variaciones en la distancia d_{ARN} son características de la diferencia entre las estructuras cristalográficas de la DENV NS3hel obtenidas en presencia y ausencia de ARN-sc [23], observándose valores más grandes en ausencia de ARN-sc (ver cruces en la Figura 3.5B). En el estado apo, el modo relajado de unión al ARN es más favorable que el modo tenso, mientras que lo contrario se observa para los estados unidos a ATP y ADP-Pi, y la brecha física entre los dos estados se amplía ligeramente en estos casos. De forma similar, se puede identificar una conformación relajada y tensa del sitio de unión al ATP en el estado apo, siendo más favorable el estado relajado. Al unirse el ATP, estos residuos están exclusivamente en el estado tenso, caracterizado por distancias más pequeñas d_{ATP}, y tras la hidrólisis de ATP, se explora nuevamente el estado relajado. Es interesante notar que, aunque el modo tenso de unión al ARN se favorece en presencia de ATP, el modo relajado caracterizado por distancias mayores d_{ARN} todavía se explora. Mecanísticamente, esto podría ser interpretado como que la proteína "suelta" al ARN al querer "agarrar" al ATP, lo que permite la exploración de múltiples modos de unión a lo largo del ácido nucleico.

Debido a las alternancias en la flexibilidad de la proteína (Figuras 3.2 y 3.5) y al cambio asimétrico en las interacciones proteína-ARN (Figuras 3.6 y 3.7), los resultados son consistentes con el modelo de motor browniano para la translocación direccional. Este modelo describe correctamente los cambios observados según las distintas ocupaciones del sitio catalítico de la NS3hel del DENV unida a ARN-sc a partir de todas las simulaciones realizadas, independientemente del largo de ARN considerado.

3.5.2 La conformación abierta del *loop* β6-β1' como ruta de salida para el fosfato

La apertura observada del *loop* β6-β1’ ofrece pruebas de una nueva posible ruta de salida para el fosfato, Pi, después de la hidrólisis. Según las estructuras cristalográficas, se ha propuesto un pequeño túnel formado por Pro 195, Ala 316, Thr 317, Pro 326 y Gln 456 como una posible ruta de salida [23]. Al abrirse el loop β6-β1’, esta región se hidrata y resulta en una ruta viable para la salida del Pi. Estos resultados indican que esta conformación de loop abierto no solo abre una clara "puerta trasera" al sitio activo, sino que también provoca la expansión de el sitio de unión a NTP (d_{ATP}), lo que sugiere aún más que este cambio conformacional favorece la liberación de los productos de la hidrólisis. En un trabajo posterior a estos resultados, exploramos la ruta de salida para el Pi a través del *loop* β6-β1’ y se demostró que efectivamente este proceso resulta exergónico, mientras la salida de Pi por la ruta canónica es un proceso endergónico [4]. Más aún, las barreras energéticas para las salidas tanto de del Pi, primer producto en abandonar el sitio catalítico, como del ADP, en un paso posterior, solo resultaron consistentes con las constantes cinéticas obtenidas experimentalmente, al tener en cuenta la conformación del *loop* abierto.

3.5.3 Tamaño del sitio de unión al ARN y posible sub-paso

Las estructuras cristalográficas disponibles para la NS3hel de DENV unida a ARN fueron obtenidas utilizando un constructo de ARN-sc de 12 nucleótidos. Sin embargo, en estas estructuras solo se resolvieron 7 nucleótidos unidos a la NS3hel, lo que probablemente se deba a un mayor grado de flexibilidad en los extremos 3′ y 5′ no resueltos. En las simulaciones dinámica molecular de esta proteína unida al ARN-sc de 7 nucleótidos (ARN7), se observa además una gran flexibilidad en el extremo 3′ del ARN7. Es importante destacar que ni las estructuras cristalográficas ni estas simulaciones pudieron capturar cambios conformacionales que serían esperables debido al efecto alostérico del cambio en la ocupación del sitio catalítico NTPasa, con excepción de la alternancia flexible-rígido-flexible (Sección 3.5.1).

Uno de los objetivos de este capítulo fue comparar el efecto de un ARN-sc más largo que 7 nucleótidos, durante el ciclo de hidrólisis de ATP que tiene lugar en el sitio catalítico de la NS3hel. Esta propuesta surge de los resultados reportados a partir de experimentos de titulación fluorescente de la NS3hel con oligonucleótidos de diferentes longitudes, en los que se informó que el tamaño mínimo del sitio de unión al ARN era de 10-11 nucleótidos. Computacionalmente, se observó un tamaño de sitio de unión al ARN de aproximadamente 10 nucleótidos en el estado apo, consistente con los experimentos, únicamente en los complejos unidos a una simple cadena de ARN de 16 y 26 nucleótidos (metodología y resultados publicados [2] y detallados en la tesis doctoral de Fernando Amrein). Teniendo en cuenta el aumento de seis veces en el tiempo computacional necesario para simular el sistema de 26 nucleótidos en comparación con el de 16, como ya se mencionó anteriormente, este último fue elegido como un modelo representativo con ARN-sc suficientemente largo. Es importante notar que estos modelos se diseñaron para representar el complejo relevante en la translocación de la NS3hel a lo largo del ARN-sc. En futuros estudios se podría modelar a la NS3hel enfrentándose a una cadena de ARN estructurado en el extremo 5′ o más aún, a una doble cadena de ARN, para profundizar en el alosterismo de esta proteína.

Para los complejos NS3hel:ARN16, que representan un modelo mejorado con respecto a uno con un ARN-sc de 7 nucleótidos, en los cuatro estados del ciclo catalítico (apo, unido a ATP unido a productos, ADP-Pi, y unido a ADP), se estimaron computacionalmente y compararon

los tamaños de sitio de unión al ARN en la NS3hel del DENV. El tamaño del sitio de unión al ARN se redujo en aproximadamente 3 nts en presencia de ATP con respecto al estado apo, y esta reducción se mantuvo después de la hidrólisis antes de la liberación de productos. Aunque el tamaño reducido del sitio de unión al ARN no cambia después de la hidrólisis, se desplaza un nucleótido hacia el extremo 3′. Para el estado apo del complejo NS3hel:ARN16, la primera base del sitio es B^{-3}, pero al unirse el ATP, se libera un nucleótido y el sitio comienza en B^{-2}. Además, la tendencia continúa después de la hidrólisis (la primera base es B^{-1}) y después de la liberación del primer producto, Pi (B^{0}). Resumidamente, a lo largo del ciclo de hidrólisis de ATP, la proteína va "soltando" las bases en el extremo 5' del ARN.

Teniendo en cuenta que la dirección de la NS3hel a lo largo del ARN es hacia el extremo 5′ y que se espera que se mueva un nucleótido por ciclo de hidrólisis de ATP, se esperaría entonces que después de la liberación del ADP, la primera base en el sitio sea B^{-4}, y así sucesivamente. Con respecto a las interacciones entre las "anclas" de la NS3hel (Arg 225, Thr 244, Arg 387 y Thr 408) y el ARN-sc, se mantienen mayormente conservadas en las cuatro etapas del ciclo catalítico, excepto por Arg 387 (motivo conservado IVa), que no establece interacciones estables con ningún nucleótido en los complejos apo y unidos a ATP (porcentaje de interacción menor al 50%), pero juega un papel importante en ambas etapas unidas a productos. Curiosamente, las simulaciones muestran que la cadena lateral de Arg 387 tiende a mantener su interacción con B^{1} durante más tiempo que con B^{2} post-hidrólisis (Figura 3.6C). Por lo tanto, parecería que la probabilidad de un paso en la dirección 3′-a-5′ sería más alta que en la dirección opuesta. Los cambios conformacionales en la cadena lateral de este aminoácido también se han reportado como relacionados con un paso de desplazamiento de la NS3hel a lo largo del ARN-sc en otro estudio computacional de la NS3hel de DENV [25].

3.5.4 Es necesario un ARN suficientemente largo para describir la interacción alostérica entre los sitios de unión de ARN y ATP

La disminución en el tamaño del sitio de unión al ARN de la NS3hel debido a la ocupación del sitio catalítico es consistente con la disminución observada en la afinidad de unión al ARN debido a la presencia de ATP, ADP-Pi o ADP. Se ha reportado también una disminución en la afinidad de unión al ARN debido a la presencia de ATP para la NS3hel de HCV [43,44].

Los resultados presentados indican que el alosterismo entre los sitios de unión al ARN y ATP se ve influenciado por la longitud del oligonucleótido: sin un oligonucleótido suficientemente largo, no se observa dicha comunicación. Esto indica que las interacciones transitorias proteína-ARN juegan un papel crucial en el alosterismo entre los sitios de unión al ARN y ATP, a pesar de no ser lo suficientemente estables para permitir su resolución cristalográfica.

El genoma de DENV (~11 kb) es mucho más largo que las longitudes de ARN-sc estudiadas aquí; sin embargo, este ARN genómico está en gran parte estructurado sobre sí mismo y en complejo con proteínas de unión a ARN virales y del huésped, dejando probablemente pequeños tramos de ARN-sc expuestos durante el ciclo viral [45]. Además, fragmentos pequeños de ARN de 15 a 30 nts son conocidos por estar presentes en células hospedadoras mamíferas y se ha demostrado que la distribución de estas longitudes se ve afectada por la infección viral [46]. Por lo tanto, se deduce que entender la dependencia de longitud de la interacción entre NS3hel y ARN-sc es un aspecto crucial en futuros esfuerzos de descubrimiento de fármacos, especialmente en el diseño de inhibidores alostéricos que

puedan aprovechar nuevos bolsillos y cambios conformacionales en la estructura de la proteína.

3.6 Conclusiones

En este capítulo, se analizan los resultados de las simulaciones de dinámica molecular de la helicasa NS3 del virus del dengue unida a un segmento de ARN-sc de 7 nucleótidos (NS3hel:ARN7) en tres estados del ciclo de hidrólisis de ATP (apo, unida a ATP y unida a productos de la hidrólisis, ADP y Pi), por un lado, y unida a un segmento de ARN-sc de 16 nucleótidos (NS3hel:ARN16) en cuatro estados del ciclo de hidrólisis de ATP (apo, unida a ATP y unida a productos de la hidrólisis, ADP y Pi, y unida a ADP), por otro. Al unirse el ATP, observamos una transición de flexible a rígida, que se invierte tras la hidrólisis, lo que respalda un modelo de motor browniano para la translocación. La novedosa conformación abierta de *loop* β6-β1', observada en los estados unidos a productos y apo, proporciona una nueva visión sobre la base estructural de las actividades acopladas de hidrólisis y translocación.

El uso de modelos atomísticos y simulaciones de dinámica molecular (DM) se ha convertido en un componente integral en los procesos de descubrimiento de fármacos. Tanto en nuestro grupo de investigación como otros hemos simulado el complejo NS3hel–ARN-sc tal como se ha cristalizado, con un fragmento de ARN-sc de 7 nucleótidos unido, para dilucidar las redes alostéricas que conectan el sitio activo de la NTPasa con el sitio de unión al ARN. Contrario a las expectativas, las simulaciones no presentaron cambios sustanciales en la interacción NS3hel–ARN debido a la presencia de ATP o los productos de la hidrólisis. En este capítulo se aporta evidencia que muestra que la longitud del fragmento de ARN-sc unido es crucial para capturar adecuadamente esta comunicación alostérica, en donde la ocupación del sitio catalítico resulta en un complejo NS3hel:ARN más débilmente unido que en el estado apo. Estos resultados tienen implicaciones directas en la interpretación de estudios experimentales de la NS3hel, así como en los esfuerzos futuros de descubrimiento computacional de inhibidores de esta proteína contra el DENV.

Bibliografía

1. Sarto C, Kaufman SB, Estrin DA, Arrar M. Nucleotide-dependent dynamics of the Dengue NS3 helicase. Biochim Biophys Acta Proteins Proteom. 2020;1868: 140441. doi:10.1016/j.bbapap.2020.140441

2. Amrein F, Sarto C, Cababie LA, Gonzalez Flecha FL, Kaufman SB, Arrar M. Impact of bound ssRNA length on allostery in the Dengue Virus NS3 helicase. Nucleic Acids Res. 2023;51: 11213–11224. doi:10.1093/nar/gkad808

3. Incicco JJ, Cababie LA, Sarto C, Adler NS, Amrein F, Mikkelsen E, et al. Thermodynamic and mechanistic analysis of the functional properties of dengue virus NS3 helicase. Biophys Rev. 2023;15: 591–600. doi:10.1007/s12551-023-01101-5

4. Adler NS, Cababie LA, Sarto C, Cavasotto CN, Gebhard LG, Estrin DA, et al. Insights into the product release mechanism of dengue virus NS3 helicase. Nucleic Acids Res. 2022;50: 6968–6979. doi:10.1093/nar/gkac473

5. Matusan AE, Pryor MJ, Davidson AD, Wright PJ. Mutagenesis of the Dengue virus type 2 NS3 protein within and outside helicase motifs: effects on enzyme activity and virus replication. J Virol. 2001;75: 9633–9643. doi:10.1128/JVI.75.20.9633-9643.2001

6. Lescar J, Luo D, Xu T, Sampath A, Lim SP, Canard B, et al. Towards the design of antiviral inhibitors against flaviviruses: the case for the multifunctional NS3 protein from Dengue virus as a target. Antiviral Res. 2008;80: 94–101. doi:10.1016/j.antiviral.2008.07.001

7. Li H, Clum S, You S, Ebner KE, Padmanabhan R. The serine protease and RNA-stimulated nucleoside triphosphatase and RNA helicase functional domains of dengue virus type 2 NS3 converge within a region of 20 amino acids. J Virol. 1999;73: 3108–3116. doi:10.1128/JVI.73.4.3108-3116.1999

8. The RNA helicase, nucleotide 5′-triphosphatase, and RNA 5′-triphosphatase activities of Dengue virus protein NS3 are Mg2+-dependent and require a functional Walker B motif in the helicase catalytic core. Virology. 2004;328: 208–218. doi:10.1016/j.virol.2004.07.004

9. Gebhard LG, Kaufman SB, Gamarnik AV. Novel ATP-Independent RNA Annealing Activity of the Dengue Virus NS3 Helicase. PLOS ONE. 2012;7: e36244. doi:10.1371/journal.pone.0036244

10. Jeremías Incicco J, Gebhard LG, González-Lebrero RM, Gamarnik AV, Kaufman SB. Steady-State NTPase Activity of Dengue Virus NS3: Number of Catalytic Sites, Nucleotide Specificity and Activation by ssRNA. PLOS ONE. 2013;8: e58508. doi:10.1371/journal.pone.0058508

11. Structure of the Dengue virus helicase/nucleoside triphosphatase catalytic domain at a resolution of 2.4 Å. J Virol. 2005.

12. Du Pont KE, McCullagh M, Geiss BJ. Conserved motifs in the flavivirus NS3 RNA helicase enzyme. Wiley Interdiscip Rev RNA. 2022;13: e1688. doi:10.1002/wrna.1688

13. Mastrangelo E, Bolognesi M, Milani M. Flaviviral helicase: insights into the mechanism of action of a motor protein. Biochem Biophys Res Commun. 2012;417: 84–87. doi:10.1016/j.bbrc.2011.11.060

14. SF1 and SF2 helicases: family matters. Current Opinion in Structural Biology. 2010;20:

313–324. doi:10.1016/j.sbi.2010.03.011

15. Dumont S, Cheng W, Serebrov V, Beran RK, Tinoco I Jr, Pyle AM, et al. RNA translocation and unwinding mechanism of HCV NS3 helicase and its coordination by ATP. Nature. 2006;439: 105–108. doi:10.1038/nature04331

16. Levin MK, Gurjar M, Patel SS. A Brownian motor mechanism of translocation and strand separation by hepatitis C virus helicase. Nat Struct Mol Biol. 2005;12: 429–435. doi:10.1038/nsmb920

17. Jankowsky E. RNA helicases at work: binding and rearranging. Trends Biochem Sci. 2011;36: 19–29. doi:10.1016/j.tibs.2010.07.008

18. Betterton MD, Jülicher F. Opening of nucleic-acid double strands by helicases: active versus passive opening. Phys Rev E Stat Nonlin Soft Matter Phys. 2005;71: 011904. doi:10.1103/PhysRevE.71.011904

19. Büttner K, Nehring S, Hopfner K-P. Structural basis for DNA duplex separation by a superfamily-2 helicase. Nat Struct Mol Biol. 2007;14: 647–652. doi:10.1038/nsmb1246

20. Gu M, Rice CM. Three conformational snapshots of the hepatitis C virus NS3 helicase reveal a ratchet translocation mechanism. Proc Natl Acad Sci U S A. 2010;107: 521–528. doi:10.1073/pnas.0913380107

21. Kim JL, Morgenstern KA, Griffith JP, Dwyer MD, Thomson JA, Murcko MA, et al. Hepatitis C virus NS3 RNA helicase domain with a bound oligonucleotide: the crystal structure provides insights into the mode of unwinding. Structure. 1998;6: 89–100. doi:10.1016/s0969-2126(98)00010-0

22. Chowdhury D, (Amsterdam). E. Stochastic Mechano-chemical Kinetics of Molecular Motors: A Multidisciplinary Enterprise from a Physicist's Perspective. 2013. Available: https://books.google.com/books/about/Stochastic_Mechano_chemical_Kinetics_of.html?hl=&id=j3EIngEACAAJ

23. Luo D, Xu T, Watson RP, Scherer-Becker D, Sampath A, Jahnke W, et al. Insights into RNA unwinding and ATP hydrolysis by the flavivirus NS3 protein. EMBO J. 2008;27: 3209–3219. doi:10.1038/emboj.2008.232

24. Du Pont KE, Davidson RB, McCullagh M, Geiss BJ. Motif V regulates energy transduction between the flavivirus NS3 ATPase and RNA-binding cleft. J Biol Chem. 2020;295: 1551–1564. doi:10.1074/jbc.RA119.011922

25. Davidson RB, Hendrix J, Geiss BJ, McCullagh M. Allostery in the dengue virus NS3 helicase: Insights into the NTPase cycle from molecular simulations. PLoS Comput Biol. 2018;14: e1006103. doi:10.1371/journal.pcbi.1006103

26. Gebhard LG, Incicco JJ, Smal C, Gallo M, Gamarnik AV, Kaufman SB. Monomeric nature of dengue virus NS3 helicase and thermodynamic analysis of the interaction with single-stranded RNA. Nucleic Acids Res. 2014;42: 11668–11686. doi:10.1093/nar/gku812

27. Humphrey W, Dalke A, Schulten K. VMD: visual molecular dynamics. J Mol Graph. 1996;14: 33–8, 27–8. doi:10.1016/0263-7855(96)00018-5

28. Olsson MHM, Søndergaard CR, Rostkowski M, Jensen JH. PROPKA3: Consistent treatment of internal and surface residues in empirical pKa predictions. J Chem Theory Comput. 2011;7: 525–537. doi:10.1021/ct100578z

29. Meagher KL, Redman LT, Carlson HA. Development of polyphosphate parameters for use with the AMBER force field. J Comput Chem. 2003;24: 1016–1025. doi:10.1002/jcc.10262

30. Steinbrecher T, Latzer J, Case DA. Revised AMBER parameters for bioorganic phosphates. J Chem Theory Comput. 2012;8: 4405–4412. doi:10.1021/ct300613v

31. Maier JA, Martinez C, Kasavajhala K, Wickstrom L, Hauser KE, Simmerling C. Ff14SB: Improving the accuracy of protein side chain and backbone parameters from ff99SB. J Chem Theory Comput. 2015;11: 3696–3713. doi:10.1021/acs.jctc.5b00255

32. Aytenfisu AH, Spasic A, Grossfield A, Stern HA, Mathews DH. Revised RNA dihedral parameters for the Amber force field improve RNA molecular dynamics. J Chem Theory Comput. 2017;13: 900–915. doi:10.1021/acs.jctc.6b00870

33. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML. Comparison of simple potential functions for simulating liquid water. J Chem Phys. 1983;79: 926–935. doi:10.1063/1.445869

34. Case DA, Aktulga HM, Belfon K, Cerutti DS, Cisneros GA, Cruzeiro VWD, et al. AmberTools. J Chem Inf Model. 2023;63: 6183–6191. doi:10.1021/acs.jcim.3c01153

35. People O. Gaussian.com. [cited 30 Jun 2025]. Available: https://gaussian.com/citation/

36. Case DA, Aktulga HM, Belfon K, Ben-Shalom IY, Berryman JT, Brozell SR, et al. Amber 2023. 2023. Available: https://books.google.com/books?hl=en&lr=&id=F1rGEAAAQBAJ&oi=fnd&pg=PA5&ots=4M6PfoP7wq&sig=r2fOLDcHA6Gdg_fzwB8W0NKP6mM

37. Sindhikara DJ, Kim S, Voter AF, Roitberg AE. Bad seeds sprout perilous dynamics: Stochastic thermostat induced trajectory synchronization in biomolecules. J Chem Theory Comput. 2009;5: 1624–1631. doi:10.1021/ct800573m

38. Berendsen HJC, Postma JPM, van Gunsteren WF, DiNola A, Haak JR. Molecular dynamics with coupling to an external bath. J Chem Phys. 1984;81: 3684–3690. doi:10.1063/1.448118

39. Ryckaert J-P, Ciccotti G, Berendsen HJC. Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. J Comput Phys. 1977;23: 327–341. doi:10.1016/0021-9991(77)90098-5

40. Roe DR, Cheatham TE 3rd. PTRAJ and CPPTRAJ: Software for processing and analysis of molecular dynamics trajectory data. J Chem Theory Comput. 2013;9: 3084–3095. doi:10.1021/ct400341p

41. Onufriev A, Bashford D, Case DA. Exploring protein native states and large-scale conformational changes with a modified generalized born model. Proteins. 2004;55: 383–394. doi:10.1002/prot.20033

42. Kinney RM, Butrapet S, Chang GJ, Tsuchiya KR, Roehrig JT, Bhamarapravati N, et al. Construction of infectious cDNA clones for dengue 2 virus: strain 16681 and its attenuated vaccine derivative, strain PDK-53. Virology. 1997;230: 300–308. doi:10.1006/viro.1997.8500

43. Porter DJ, Preugschat F. Strand-separating activity of hepatitis C virus helicase in the absence of ATP. Biochemistry. 2000;39: 5166–5173. doi:10.1021/bi992384f

44. Levin MK, Gurjar MM, Patel SS. ATP binding modulates the nucleic acid affinity of

hepatitis C virus helicase. J Biol Chem. 2003;278: 23311–23316. doi:10.1074/jbc.M301283200

45. Chambers TJ, Hahn CS, Galler R, Rice CM. Flavivirus genome organization, expression, and replication. Annu Rev Microbiol. 1990;44: 649–688. doi:10.1146/annurev.mi.44.100190.003245

46. Jackowiak P, Hojka-Osinska A, Philips A, Zmienko A, Budzko L, Maillard P, et al. Small RNA fragments derived from multiple RNA classes – the missing element of multi-omics characteristics of the hepatitis C virus cell culture model. BMC Genomics. 2017;18. doi:10.1186/s12864-017-3891-3

Capítulo 4: Hidrólisis de ATP estimulada por ARN en la helicasa NS3 del virus del dengue

4.1 Resumen

La helicasa NS3 del virus del dengue (NS3hel), una NTPasa del tipo DEAH-box esencial para la replicación viral. Su actividad NTPasa es altamente sensible al entorno, una característica que podría aprovecharse en el desarrollo de tratamientos antivirales, aunque los mecanismos subyacentes de esta sensibilidad aún no están del todo claros. A partir de conformaciones resueltas en estructuras cristalográficas de NS3hel en ausencia y en presencia de ARN, se ha propuesto que la posición de una base conservada genera un aumento en la actividad NTPasa en presencia de ARN. La activación se da por la posición del residuo E285, ampliamente reconocido como la base catalítica de la hidrólisis, que se encuentra disponible en el estado unido al ARN, pero está secuestrado en la esfera de coordinación del centro metálico divalente en ausencia de ARN. En este capítulo, se comparó sistemáticamente la viabilidad energética de los mecanismos asistidos por la base o por el sustrato para la hidrólisis de ATP catalizada por NS3hel, tanto en ausencia como en presencia de ARN. Se encontró que E285 no necesariamente actúa como una base general durante la hidrólisis, incluso en el estado unido al ARN. En ambos casos, la transferencia de protón desde la molécula de agua nucleófila ocurre después de la ruptura del enlace fosfoanhídrido y no representa el paso limitante de la reacción. A partir de estos resultados, se identificaron un subconjunto de residuos del sitio catalítico que proponemos como claves en la estimulación de la hidrólisis de ATP por ARN, específicamente por su papel en la estabilización del intermedio metafosfato y en la orientación de la molécula de agua catalítica. La configuración de estos residuos es sensible a la presencia, longitud y conformación del ARN. Finalmente, identificamos una región hélice-giro-hélice (residuos 528-555) en el subdominio 3 que se correlaciona con este subconjunto de forma dependiente del ARN, y que se encuentra adyacente a bolsillos estratégicos con potencial para el desarrollo de fármacos pan-serotipo contra el dengue. Los resultados de este trabajo están siendo incorporados en un manuscrito del que soy primera autora y será enviado a una revista con revisión por pares para su publicación.

4.2 Introducción

La proteína NS3 es una enzima multifuncional de dos dominios, esencial para la replicación del virus del dengue. Su dominio N-terminal corresponde a una serina proteasa, mientras que el dominio C-terminal actúa como una helicasa de ARN (Figura 4.1A), con actividades de nucleósido trifosfatasa (NTPasa) y trifosfatasa de ARN 5′ (RTPasa) [1]. En este capítulo, al igual que en el anterior, el foco estará puesto en el dominio helicasa de NS3 (NS3hel), que pertenece a la familia DEAH-box dentro de la superfamilia 2 de helicasas [2], un grupo de proteínas que separan *loops* y dobles cadenas de ARN con el requerimiento de la hidrólisis de ATP y que se caracterizan por presentar motivos conservados. Cuatro de estos motivos están asociados a la actividad ATPasa, incluyendo el Motivo I (también conocido como motivo Walker A o *P-loop*), II, III y VI, mientras que otros cuatro (IA, IV, IVA y V) están relacionados con la unión al ARN y con la translocación de la proteína a lo largo de la hebra de ácido nucleico [3] (Figura 4.1B).

La proteína NS3hel está compuesta por tres subdominios: D1, D2 y D3. El sitio de unión de ATP se encuentra entre D1 y D2, mientras que el sitio de unión del ARN se localiza entre la interfaz D1-D2 y D3 (Figura 4.1B). En el capítulo anterior (y en la publicación relacionada con estos resultados), mostré que la NS3hel alterna entre conformaciones rígidas y flexibles durante el ciclo de hidrólisis de ATP y también que mostré una nueva conformación abierta del *loop* β6–β1′ [4], que conecta D1 y D2 y que favorece la liberación de fosfato [5]. Además,se ha establecido que la comunicación alostérica entre los sitios de unión de ATP y de ARN en NS3hel del DENV depende de la longitud del ARN simple cadena (ARN-sc) unido [6].

La actividad ATPasa de NS3hel del DENV está modulada por la presencia de ARN, con un aumento reportado de 10 veces en la constante de velocidad (k_{cat}) y una disminución de 100 veces en la afinidad aparente por ATP (K_M) en comparación con la enzima en ausencia de ARN [7,8]. Las estructuras cristalográficas de NS3hel del virus del dengue han aportado información sobre cómo el ARN puede modular su actividad ATPasa: en estructuras sin ARN, el residuo E285 (perteneciente al motivo DEAH) está coordinado al catión divalente, mientras que en estructuras con ARN, E285 es desplazado por una molécula de agua [9]. Esta observación dio lugar a la siguiente hipótesis: que la estimulación de la actividad ATPasa por el ARN podría deberse a la capacidad de E285 para actuar como base general y aumentar la nucleofilicidad del agua catalítica (Figura 4.1C).

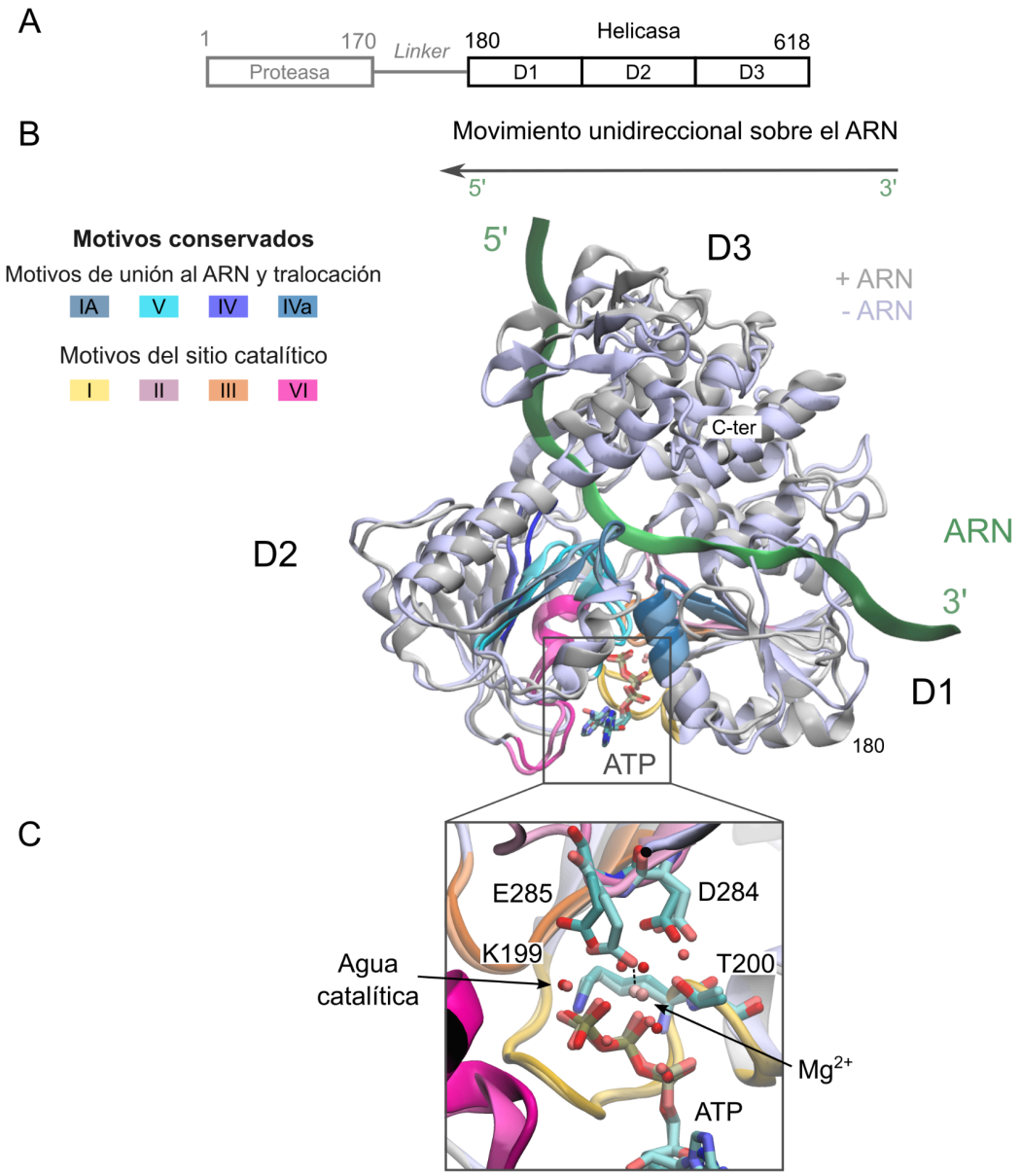
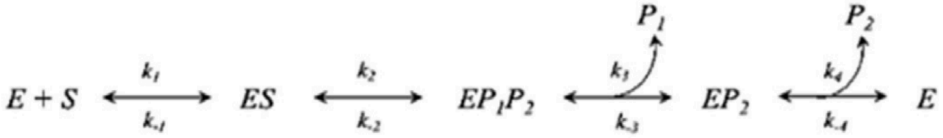


Figura 4.1. Vista general de la proteína no estructural 3 y el sitio catalítico ATPasa. **A)** Representación esquemática de los dominios proteasa y helicasa de la NS3 conectados por una región *linker*, señalando los tres subdominios de la helicasa (D1, D2 y D3). **B)** Estructuras cristalográficas superpuestas de la helicasa NS3hel con ARN (gris, PDB ID: 2JLV) y sin ARN (celeste claro, PDB ID: 2JLR). El ARN simple cadena se muestra como una cinta verde; la región verde clara corresponde a los nucleótidos resueltos en la estructura cristalográfica 2JLV. La región verde oscura representa un modelo de un ARN-sc de 16 nucleótidos. Una flecha horizontal indica el movimiento unidireccional de la proteína desde el extremo 3' hacia el 5' a lo largo del ARN-sc. Los motivos conservados de la familia de helicasas DEAH-box están coloreados según la leyenda. Los subdominios D1, D2 y D3 están etiquetados. Se indica el residuo N-terminal de NS3hel (residuo 180) y también la posición del C-terminal. **C)** Vista en detalle del sitio catalítico ATPasa en presencia (tonos brillantes) y ausencia (tonos opacos) de ARN, mostrando ATP, Mg^{2+}, residuos clave incluyendo D284 y E285 del motivo DEAH, y la molécula de agua catalítica. En ausencia de ARN, E285 está coordinado al Mg^{2+}, mientras que en presencia de ARN es desplazado por una molécula de agua.

En trabajos publicados por nuestro grupo, demostramos que el ciclo catalítico de la NS3hel del virus del dengue se describe adecuadamente mediante un mecanismo secuencial ordenado (Esquema 4.1, donde P1 = Pi y P2 = ADP). Es importante señalar que los valores reportados de k_{cat} para la hidrólisis de ATP reflejan la secuencia completa de procesos microscópicos, desde la unión del ATP hasta la liberación de Pi y ADP, y no únicamente la reacción química; no obstante, estos valores proporcionan límites inferiores útiles para las constantes de velocidad de cada uno de estos procesos (y límites superiores para las barreras libres de energía asociadas).



Esquema 4.1

La reacción de hidrólisis en sí misma (con constantes de velocidad directa e inversa k_2 y k_{-2} en el Esquema 4.1) puede describirse a su vez mediante dos subpasos, considerando que inicialmente ambos productos se encuentran coordinados al centro metálico divalente (M^{2+}), y que la disociación de Pi del catión divalente debe ocurrir antes de su liberación del complejo ternario (Esquema 4.2). Este proceso ha demostrado aportar una contribución significativa al cambio de energía libre total asociado a la hidrólisis de ATP [10–13].

$$ATP \cdot M^{2+} + H_2O \rightarrow ADP \cdot P_i \cdot M^{2+} \rightarrow ADP \cdot M^{2+} + P_i$$

Esquema 4.2

$$ES \qquad EP_1P_2 \qquad EP_1P_2$$

En este trabajo el enfoque estará en el primero de los dos subpasos: la conversión de ATP en ADP y Pi, todos coordinados al Mg^{2+} en el sitio catalítico de la NS3hel. Esta reacción corresponde a una sustitución nucleofílica en la que una agua catalítica (Wc, por sus siglas en inglés) ataca al fosfato gamma (Pγ) del ATP, lo que resulta en la ruptura del enlace fosfoanhídrido y la formación de ADP y fosfato inorgánico (Pi), acompañada por una transferencia de protón desde el agua catalítica. El estudio detallado de esta reacción es más complejo aún por el hecho de que Pi puede referirse tanto a $H_2PO_4^-$ como a HPO_4^{2-} y un protón separado, existiendo múltiples rutas posibles que conducen a cada uno de estos estados productos. No obstante, es posible distinguir entre mecanismos asistidos por el sustrato o por el solvente (Figura 4.2A) y mecanismos asistidos por una base (Figura 4.2B), evaluando si se forma inicialmente fosfato diácido o fosfato monoácido (acompañado por una base protonada).

A. Mecanismo asistido por el sustrato o por el solvente



B. Mecanismo asistido por una base

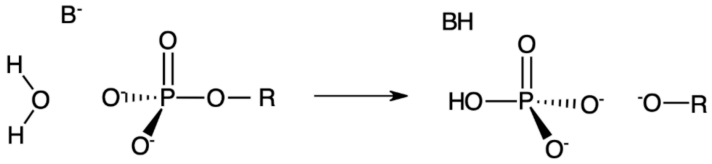


Figura 4.2. Múltiples mecanismos posibles para la hidrólisis de ATP. **A)** Mecanismo asistido por el sustrato o por el solvente, en los cuales el propio grupo fosfato actúa como base, ya sea de forma directa o a través de moléculas de agua intermedias, dando lugar a fosfato diácido como producto inicial de la hidrólisis. **B)** Mecanismo asistido por una base, en el que un protón es transferido primero a una base general, formándose fosfato monoácido.

En solución, el ATP es extremadamente estable, con una constante de velocidad del orden de 10^{-5} s^{-1} [14]. Las enzimas ATPasas aumentan esta velocidad en un factor de 10^5 [15]. Existe una extensa bibliografía sobre las rutas enzimáticas y no enzimáticas de esta reacción, abordada tanto mediante enfoques experimentales como computacionales, que indican que se trata de un mecanismo disociativo, en el que la ruptura del enlace precede a la formación del nuevo enlace y a la transferencia de protón desde el agua catalítica [12,15–20]. Sin embargo, los numerosos trabajos que describen la reacción como un mecanismo concertado de tipo SN2 [21] evidencian que esta reacción central en biología sigue representando un desafío tanto para la química física experimental como para la química computacional [22,23]. La aceleración enzimática ha sido atribuida a diversos factores, incluyendo la desestabilización de los reactivos, la estabilización del intermediario (o estado de transición) tipo metafosfato, y el posicionamiento y aumento de la nucleofilicidad de la molécula de agua catalítica [24].

La creciente observación de mecanismos similares entre distintos motores moleculares que hidrolizan ATP, incluyendo miosina, kinesina, F1-ATPasa, así como otras helicasas de la familia DEAH-box, no resulta sorprendente si se considera la presencia de un motivo P-loop o Walker A universalmente conservado (aquí referido como motivo I), así como de los altamente conservados dedos de arginina (motivo VI) y el residuo de glutamato (motivo II) en los sitios activos de estas enzimas [25]. Estudios computacionales recientes sobre la NS3hel del virus del dengue y del Zika han destacado el rol del motivo V en la transducción de energía desde el sitio activo de ATP hacia la hendidura de unión al ARN [12,26,27]. Se ha reportado también que la presencia de ARN afecta el posicionamiento tanto de la molécula de agua catalítica como de la base general en Prp2 y en la NS3hel del dengue, observándose menores barreras de activación en los estados unidos al ARN [27,28]. Hasta donde se sabe, no se ha realizado una comparación sistemática de los mecanismos asistidos por el sustrato y por una base general en los estados con y sin ARN. Una comprensión más clara del papel del residuo conservado E285 en helicasas DEAH-box resulta aún más relevante a la luz del mecanismo

asistido por el solvente recientemente propuesto para Ras y GTPasas relacionadas, que carecen de una base general en sus sitios activos [29].

A continuación agrego algunos detalles extraídos de los aporte de estudios computacionales sobre los mecanismos de hidrólisis de ATP en diversas helicasas dependientes de ATPasa. Utilizando el método adaptativo de *string*, se reportó una barrera energética de 15,1 kcal·mol^{-1} para la hidrólisis de ATP en la NS3hel del virus del Zika unida al ARN. Curiosamente, los autores sugirieron que la ausencia de ARN no modificaría la barrera, basándose en la conformación similar del sitio catalítico observada en estructuras cristalográficas. Sus resultados destacaron el rol de la conformación del motivo V en la modulación de la barrera energética de la hidrólisis, así como un paso posterior a la hidrólisis que involucra la liberación del fosfato desde la esfera de coordinación del Mg^{2+} y la posterior entrada de agua [12]. Mecanismos similares han sido observados en miosina, donde se propuso que la hidratación del magnesio es clave para la liberación del fosfato, lo que sugiere que esta característica podría estar conservada entre las NTPasas [10,11]. Para la helicasa DEAH-box Prp2, simulaciones con metadinámica revelaron una barrera de energía libre de 16,7 ± 0,8 kcal·mol^{-1} en presencia de ARN, que aumentó a 23,5 ± 3,2 kcal·mol^{-1} en ausencia del mismo [28]. Sin embargo, para la helicasa NS3 del virus del dengue, un estudio teórico reportó barreras energéticas comparables en los estados con y sin ARN (27,5 y 30,5 kcal·mol^{-1}, respectivamente), a pesar de la diferencia en la conformación del sitio catalítico [27].

La helicasa NS3hel es un blanco antiviral clave debido al papel que juega en la replicación no solo del DENV, sino de todos los flavivirus. Si bien los sitios de unión al ARN y al ATP son bolsillos drogables canónicos, su alta conservación resulta polémica respecto a posibles efectos *off-target* sobre helicasas y ATPasas del huésped [30,31]. Para superar esto, estudios recientes han explorado bolsillos crípticos, que son sitios transitorios revelados mediante enfoques basados en fragmentos de moléculas orgánicas [32] y simulaciones [33], y que podrían servir como blancos alternativos para generar una inhibición alostérica. En el virus del dengue, este enfoque resulta particularmente prometedor dado el desafío que representa el desarrollo de vacunas para sus cuatro serotipos debido al fenómeno de potenciación dependiente de anticuerpos. Un inhibidor pan-serotipo de la NS3hel del DENV representa una estrategia antiviral atractiva, con potencial aplicabilidad a otros flavivirus. Sin embargo, aún no está claro cómo la modulación de estos bolsillos alostéricos afectaría la función enzimática, lo que añade incertidumbre a su uso en comparación con los sitios canónicos bien caracterizados.

En este capítulo, se exploraron posibles mecanismos de hidrólisis de ATP en la NS3hel del DENV en los estados con y sin ARN. Utilizando el Método Adaptativo de *String* con una descripción QM/MM, se buscó el camino de mínima energía libre (MFEP, por sus siglas en inglés) para cada estado, considerando productos derivados de mecanismos asistidos por el sustrato o por una base (Figura 4.2). Encontramos que la formación inicial del intermediario pentavalente metafosfato es sustancialmente más favorable en el estado unido al ARN que en el estado sin ARN. Aunque la presencia de ARN favorece la protonación de E285, nuestros resultados muestran que esta transferencia de protón ocurre después del ataque nucleofílico del agua catalítica sobre el intermediario metafosfato, lo que modifica nuestra comprensión del papel de E285 como base general. Además, observamos que no solo la presencia, sino también la longitud y la conformación del ARN inducen cambios en un subconjunto de residuos del sitio catalítico junto con el posicionamiento del agua catalítica. Finalmente,

identificamos bolsillos drogables pan-serotípicos que se encuentran alostéricamente ligados a este subconjunto de residuos del sitio catalítico. Se espera que estos resultados sean útiles para el desarrollo de antivirales dirigidos a la actividad NTPasa del dominio helicasa NS3, así como para el diseño de virus atenuados.

4.3 Métodos

Para comprender en detalle los aspectos mecanísticos relevantes en la hidrólisis de ATP estimulada por ARN en DENV2, así como sus implicancias para futuras intervenciones terapéuticas, este capítulo combina resultados obtenidos a partir de tres enfoques metodológicos complementarios:

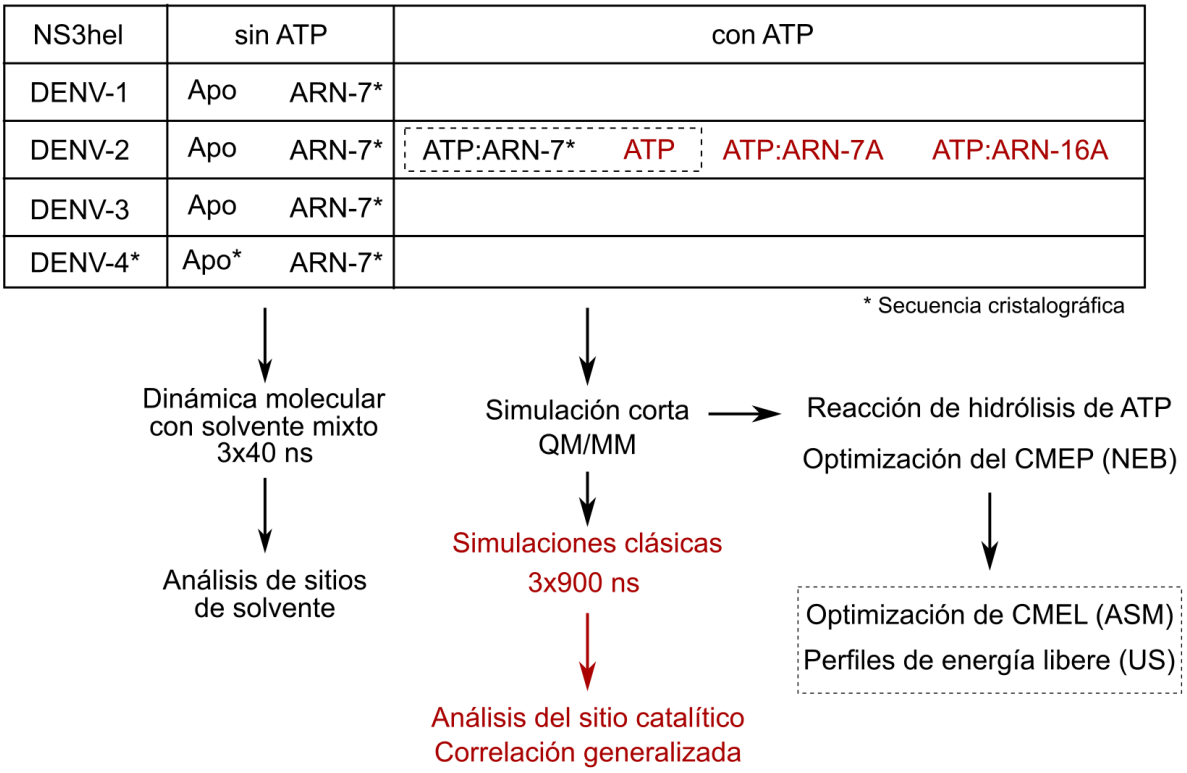
1. Exploración de trayectorias de mínima energía libre para la hidrólisis de ATP catalizada por la NS3hel del DENV2 en estados con y sin ARN, utilizando una descripción QM/MM.

2. Comparación de los conjuntos conformacionales del complejo de Michaelis (complejo NS3hel:ATP) del DENV2 en sus estados con y sin ARN mediante simulaciones clásicas de dinámica molecular.

3. Identificación de bolsillos drogables crípticos en la NS3hel, conservados en los cuatro serotipos del DENV, tanto en presencia como en ausencia de ARN.

Los protocolos utilizados para cada uno de estos enfoques se describen a continuación. Un resumen más detallado de los sistemas y simulaciones se incluye en el Esquema 4.3.

Esquema 4.3 Resumen de todos los sistemas estudiados y técnicas aplicadas



4.3.1 Construcción de los sistemas

Los sistemas NS3hel:ATP y NS3hel:ATP:ARN7 (residuos 180–618 de la NS3hel) se obtuvieron mediante el programa de modelado por homología Swiss-Model [34], utilizando una secuencia de DENV2 (la misma utilizada en los datos experimentales [7], número de acceso GenBank U87411) y las estructuras cristalográficas con PDB ID 2JLR y 2JLV [9], respectivamente. Los complejos ternarios NS3hel:ATP:ARN-7A y NS3hel:ATP:ARN-16A se construyeron siguiendo el mismo protocolo utilizado para NS3hel:ATP:ARN7.

Para el complejo NS3hel:ATP:ARN-16A, la estructura proteica se superpuso a la primera configuración de una simulación previamente equilibrada del sistema con ARN de 16 nucleótidos, proveniente de un trabajo anterior [6]. Todas las bases del ARN fueron modificadas a adenina en los sistemas NS3hel:ATP:ARN-7A y NS3hel:ATP:ARN-16A. Los análogos de ATP (AMP-PNP) fueron reemplazados por ATP, y los iones Mn^{2+} por Mg^{2+}. Las moléculas de agua presentes en las estructuras cristalográficas se conservaron para preservar la conformación adecuada del sitio catalítico. Se añadió una concentración fisiológica de 0,15 M de cloruro de sodio, junto con iones Na^{+} adicionales para neutralizar el sistema. Las cargas totales de cada sistema se detallan en la Tabla 4.1.

Tabla 4.1. Carga total de cada sistema (unidades: *e*)

	Análisis del mecanismo de reacción y de las conformaciones de NS3hel:ARN				Búsqueda de bolsillos conservados	
	ATP	ATP:ARN7	ATP:ARN7A	ATP:ARN16A	Apo	ARN7
DV1	-	-	-	-	+5	-1
DV2	0	-6	-6	-15	+2	-4
DV3	-	-	-	-	+6	0
DV4	-	-	-	-	+5	-1

Los estados de protonación se asignaron tras calcular los valores de pKa utilizando Propka [35]; no se sugirieron cambios respecto a los estados por defecto para pH neutro. Todas las histidinas fueron consideradas neutras y protonadas en la posición Nε. Los átomos de hidrógeno de las moléculas de agua cristalográficas y los iones de cloruro de sodio fueron agregados utilizando el módulo tLeAP [36] del paquete de simulación Amber. Las proteínas y los ARN se modelaron utilizando el campo de fuerza Amber ff14SB, con la corrección ILDN para residuos proteicos y los parámetros de las torsiones recientemente revisados para ARN [37,38]. Todos los sistemas se solvataron con el modelo de agua TIP3P [39] mediante el mismo módulo tLeAP, utilizando cajas octaédricas extendidas 8 Å desde la superficie del soluto, con aproximadamente 10.000 moléculas de agua TIP3P para cada sistema.

4.3.1.1 Región cuántica

Cada sistema incluyó una región cuántica (QM) descrita utilizando el funcional DFT-B3LYP y el conjunto base 6-31G* [40,41] que se empleó inicialmente en la exploración de los mecanismos de reacción mediante el método de la banda elástica NEB (de las siglas en inglés, *Nudged Elastic Band*) y posteriormente se empleó el funcional semiempírico DFTB3

[42–44] en el refinamiento de dichos caminos dentro del marco del método de la cuerda ASM (de las siglas en inglés, *Adaptive String Method)* y los cálculos de energía libre (ver sección correspondiente más abajo).

La región QM comprendió 49 átomos para el estado sin ARN y 52 átomos para el estado con ARN. Específicamente, incluyó las cadenas laterales de Thr 200 y Glu 285, la molécula de agua catalítica y una molécula de agua auxiliar, dos (estado sin ARN) o tres (estado con ARN) moléculas de agua pertenecientes a la esfera de coordinación del Mg^{2+}, y el grupo trifosfato del ATP (Figura 4.1). Los átomos de enlace (átomos *link*) se ubicaron en los carbonos β de Thr 200 y Glu 285, y en el carbono 5' del ATP.

4.3.2 Minimización, calentamiento y equilibración de la presión

Inicialmente, se realizaron 3000 pasos de minimización de energía con la región QM fijada, aplicando el método de *steepest descent* durante los primeros 1000 pasos y luego cambiando al método de *conjugate gradient* para los pasos restantes. A continuación, se llevó a cabo una minimización QM/MM aplicando fuertes restricciones a la región MM (constante de restricción: 200 kcal/mol·Å²), utilizando el algoritmo LBFGS (*Limited-memory Broyden-Fletcher-Goldfarb-Shanno*) [45], hasta que el valor cuadrático medio de las componentes cartesianas del gradiente fuera menor a 10^{-4} kcal·mol^{-1}·Å^{-1}, lo cual tomó aproximadamente 200 pasos.

Luego, todos los sistemas fueron calentados gradualmente utilizando un termostato de Langevin con una frecuencia de colisión de 5,0 ps^{-1} [46], alternando entre simulaciones clásicas (MM) con la región QM fija (paso de integración: 1,0 fs) y simulaciones QM/MM (paso de integración: 0,5 fs). La temperatura se incrementó por etapas: de 0 K a 100 K, de 100 K a 200 K y finalmente de 200 K a 300 K. Se realizó un paso de equilibración a temperatura constante (300 K) mediante etapas secuenciales: primero, simulaciones MM con restricciones sobre la región QM (k_{rest} = 200 kcal/mol·Å^2), seguidas por simulaciones QM/MM. Finalmente, se llevó a cabo la equilibración a presión constante (1 bar), utilizando un barostato de Berendsen [47] con un tiempo de relajación de presión de 1,0 ps. Esta etapa también consistió en dos fases: una simulación MM con la región QM restringida (k_{rest} = 200 kcal/mol·Å^2), seguida por una simulación QM/MM.

Todas las simulaciones, excepto los pasos de equilibración MM (realizados con PMEMD-CUDA), se llevaron a cabo con el programa Sander [36]. Para las interacciones de no enlace, se utilizó un radio de corte de 10,0 Å en todas las etapas descritas anteriormente. Un resumen de estas etapas se presenta en la Tabla 4.2.

Tabla 4.2. Etapas de la equilibración QM/MM

Etapas	Condiciones	Tiempo (ps)
Calentamiento 0-100 K	MM (QM fijo)	5
	QM/MM	2
Calentamiento 100-200 K	MM (QM fijo)	5
	QM/MM	2

Calentamiento 200-300 K	MM (QM fijo)	5
	QM/MM	2
Equilibración T constante	MM (QM restringido)	50
	QM/MM	5
Equilibración p constante	MM (QM restringido)	50
	QM/MM	5

4.3.3 Camino de mínima energía potencial con el método de la banda elástica

Para identificar el camino de energía mínima (CME) para la hidrólisis de ATP por la NS3hel en presencia y ausencia de ARN, se utilizó el método *climbing image nudged elastic band* (NEB) implementado en ORCA [48,49]. Se empleó la descripción QM/MM descrita anteriormente, con el funcional DFT-B3LYP y el conjunto base 6-31G* [40,41].

La región activa involucrada en la reacción comprendió 93 átomos (estado unido a ARN) o 90 átomos (estado sin ARN), que incluyó la región QM junto con Lys 199 (CH_2NH_3), Gly 414, Gln 456 (CH_2CONH_2), Arg 460 y Arg 463 ($CH_2NHCNH_2NH_2$). Para cada sistema, la estructura de reactivo (R) se tomó de la fase de equilibración previamente descrita, mientras que la estructura de producto (P) fue construida manualmente utilizando el programa Visual Molecular Dynamics (VMD) [50], alargando el enlace Pγ–Oβγ, acercando el átomo Pγ al oxígeno del agua catalítica y, dependiendo del mecanismo de reacción bajo estudio, posicionando el hidrógeno del agua catalítica cerca de Glu 285 o más próximo al grupo fosfato.

Resumidamente, el cálculo se desarrolló de la siguiente manera: primero se minimizaron energéticamente las estructuras R y P; luego se generó una interpolación geométrica entre ellas que produjo una serie de estructuras intermedias o nodos a lo largo de la trayectoria; se calculó la energía de cada estructura intermedia; y finalmente, la trayectoria se refinó iterativamente permitiendo que las estructuras intermedias se relajaran perpendicularmente al camino mientras se mantenía una distancia uniforme entre ellas. Este proceso continuó hasta alcanzar la convergencia predeterminada hacia la CME [51]. Este método también se encuentra descrito en el Capítulo 2, Sección 2.4.2.1.

Para asegurar la identificación de un verdadero estado de transición, el nodo con mayor energía a lo largo de la CME se optimizó adicionalmente en la dirección del camino y sentido ascendente para localizar un punto silla en la superficie energética. En total, se emplearon 13 nodos intermedios para cada cálculo NEB.

4.3.4 Camino de mínima energía libre con el método de la cuerda

4.3.4.1 Optimización de la cuerda

Los caminos obtenidos mediante el método NEB se utilizaron como sugerencia inicial para calcular los caminos de mínima energía libre (CMEL) empleando el método de la cuerda ASM (Capítulo 2, Sección 2.4.2.2) [52]. Este método tiene en cuenta las contribuciones entrópicas mediante la incorporación de simulaciones de dinámica molecular entre los pasos de optimización de la cuerda (análoga a la banda elástica en NEB).

En lugar de definir una región activa, ASM se basa en la definición de variables colectivas (CVs, por sus siglas en inglés). Los resultados de NEB también se utilizaron para seleccionar cuatro distancias como CVs para NS3hel unido al ARN: Pγ–Oβγ, OWc–HWc, Pγ–OWc y OE285–HWc. Para el sistema sin ARN se usaron las mismas cuatro CVs, más Oγ–HWc. La cuerda para cada sistema consistió en 48 nodos. Las simulaciones se realizaron con *multisander* del paquete Amber [36]. Antes de iniciar los cálculos ASM, se realizó una minimización y una trayectoria QM/MM en el ensamble NVT de 50 ps para los estados reactivo y producto, utilizando Sander con la descripción DFTB3 para la región QM, a fin de compensar el mayor costo computacional de los 48 nodos y el marco ASM. Estas simulaciones NVT se llevaron a cabo como se describió anteriormente, con la excepción de un mayor paso de tiempo de 1.0 fs y una sustitución de los átomos de hidrógeno en la región QM por deuterio; se aplicó el algoritmo SHAKE para restringir los enlaces a hidrógenos no polares en la región MM [53]. Durante la optimización de la cuerda se realizó intercambio de réplicas, intentando un intercambio cada 0.1 ps entre nodos adyacentes. La optimización de la cuerda se consideró convergente cuando la evolución temporal del desplazamiento cuadrático medio (RMSD) del estado del string tendió a valores por debajo de 0.05 $amu^{1/2}$ Å durante 2 ps y 0.15 $amu^{1/2}$ Å durante 4 ps para el complejo con y sin ARN, respectivamente.

4.3.4.2 Perfiles de energía libre

Una vez convergida la CMEL, se aplicó el método de *umbrella sampling* (Capítulo 2, Sección 2.4.2.3), restringiendo las variables colectivas mediante un sesgo armónico centrado en los valores optimizados de CV en cada nodo. La constante del sesgo fue establecida para cada nodo en base a los resultados de la etapa anterior. En esta etapa también se aplicó el enfoque de intercambio de réplicas, intentando un intercambio cada 0.1 ps entre nodos adyacentes. Para cada nodo, las simulaciones sesgadas se extendieron entre 90 y 100 ps. El método de análisis de histogramas ponderados (WHAM) se utilizó para construir los perfiles de energía libre [54].

4.3.5 Análisis conformacional de los estados reactivos

4.3.5.1 Simulaciones clásicas de dinámica molecular

Para los complejos NS3hel:ATP, NS3hel:ATP:ARN-7A y NS3hel:ATP:ARN-16A, se obtuvieron tres réplicas de simulaciones clásicas de dinámica molecular en el ensamble NVT de 1 μs utilizando PMEMD-CUDA, con un paso de integración de 2.0 fs. Las estructuras iniciales se tomaron del último paso de las simulaciones QM/MM descritas anteriormente. Se aplicó el termostato de Langevin con una frecuencia de colisión de 1.0 ps^{-1} [46] para mantener una

temperatura constante de 300 K. El algoritmo SHAKE se utilizó para restringir los enlaces a hidrógenos no polares [53]. Se empleó un radio de corte de 10.0 Å para las interacciones de no enlace.

Para el análisis, se descartaron los primeros 100 ns como una etapa adicional de equilibración. Las trayectorias se procesaron y analizaron con el programa CPPTRAJ [55] (cálculos de RMSD, distancias y ángulos). Para el complejo con ARN de 16 nucleótidos, se aplicó un algoritmo jerárquico aglomerativo (de abajo hacia arriba) de *clustering*, alineando los átomos P del ARN y dos átomos de carbono de cada base (C4 y C5). Para este análisis, se tomaron 1 de cada 10 imágenes de la trayectoria, la primera de las cuales se eligió aleatoriamente, y el algoritmo se configuró para finalizar cuando la distancia mínima entre clústeres fuera mayor a 5 Å. Se utilizó un enfoque de enlace promedio (*average-linkage*) para comparar la distancia entre miembros de dos clústeres. Tras finalizar el *clustering*, se reasignaron todas las demás imágenes a los clústeres obtenidos. Finalmente, se examinaron visualmente los cuatro clústeres identificados para corroborar la separación, y se agruparon en dos conjuntos que cumplían con el criterio de que el extremo 5' del ARN se encontrara más próximo al subdominio D2 o D3.

4.3.5.2 Correlación generalizada

La correlación generalizada (CG) evalúa la correlación entre residuos a partir de la información mutua (IM), una medida que captura dependencias tanto lineales como no lineales [56]. El coeficiente de CG normalizado varía entre 0 (sin correlación) y 1 (correlación total). Este cálculo se basa en la construcción de modelos de grafos que permiten revelar dinámicas estructurales y funcionales en biomoléculas, y resulta útil para detectar acoplamientos a larga distancia implicados en la regulación alostérica [57]. Para los cálculos de CG, se concatenaron las tres trayectorias de 900 ns (1 µs descartando los primeros 100 ns como equilibración adicional), tomando una imagen cada 2 ns. Solo se analizaron los átomos Cα de la proteína [58].

4.3.6 Identificación de bolsillos conservados

4.3.6.1 Modelos por homología y preparación de sistemas

Los modelos por homología de la NS3hel de los serotipos DENV1, DENV2 y DENV3 se obtuvieron de manera análoga a lo descrito anteriormente, utilizando como molde las estructuras cristalográficas del serotipo DENV4 (PDB IDs: 2JLQ y 2JLU). Los números de acceso de GenBank correspondientes a DENV1, DENV2 y DENV3 fueron: U88535, U87411 y AY099336, respectivamente. Con el programa tLeAP, se añadió una caja de solvente mixto que contenía un 20% de etanol en agua (aproximadamente 10.000 moléculas de agua y 800 de etanol). Los parámetros del solvente mixto se tomaron de la bibliografía [59].

4.3.6.1 Etapas de las simulaciones de dinámica molecular con solvente mixto

Se prepararon ocho sistemas, con y sin ARN para cada serotipo (DENV1-4), y se llevaron a cabo simulaciones de dinámica molecular siguiendo el siguiente protocolo. Inicialmente, se realizaron 2000 pasos de minimización de energía utilizando el método *steepest descent*

durante los primeros 500 pasos, seguido del método *conjugate gradient*. Se empleó un radio de corte de 10.0 Å para las interacciones de no enlace.

La estructura minimizada de cada sistema se utilizó para generar tres réplicas independientes que fueron equilibradas en distintas etapas: calentamiento de 0 a 200 K durante 50 ps con un paso de integración de 1 fs y fuertes restricciones sobre la proteína (k_{rest} de 100 kcal/mol·Å^2); luego, calentamiento de 200 a 300 K bajo las mismas condiciones pero con restricciones más débiles (k_{rest} de 50 kcal/mol·Å^2); posteriormente, una etapa de equilibración de 50 ps a temperatura y volumen constantes sin restricciones; y finalmente, una etapa a temperatura y presión constantes sin restricciones durante 100 ps adicionales. Para las simulaciones a presión constante se realizó un ajuste isotrópico de las posiciones con el barostato de Berendsen y un tiempo de relajación de presión de 2.0 ps. Se aplicó el termostato de Langevin con una frecuencia de colisión de 1.0 ps^{-1}. El algoritmo SHAKE se utilizó para restringir los enlaces a hidrógenos no polares. En estas etapas, se aplicó un radio de corte de 8.0 Å para las interacciones de no enlace. Para cada una de las tres réplicas, se realizó una simulación de producción de 50 ns en condiciones NVT, como se describió previamente, con un paso de integración de 2.0 fs. A los sistemas cargados se les aplicó el tratamiento por defecto de plasma uniformemente neutralizante de PMEMD (ver Tabla 4.1). Para el análisis, se descartaron los primeros 10 ns de cada réplica como equilibración adicional.

4.3.6.2 Algoritmo de agrupamiento para identificar sitios de etanol

Se utilizó el algoritmo WatClust [60] (diseñado para identificar “sitios de agua” a partir de una trayectoria) para detectar los “sitios de solvente” [61], es decir, regiones adyacentes a la superficie de la proteína donde la probabilidad de encontrar un átomo de referencia es mayor que en el seno del solvente. Se estableció el parámetro ajustable de distancia “dist” en 0,28 y el parámetro “watnumbermin” (número mínimo de moléculas de agua que un clúster debe contener para ser considerado un sitio de solvente) en 120 (10% de los 1200 imágenes obtenidas al concatenar las tres réplicas de simulación, con 10 imágenes por nanosegundo). Las trayectorias fueron alineadas por separado para cada sistema considerando todos los residuos de la proteína.

Si bien los sitios identificados por el algoritmo derivan de dos tipos de referencias (el átomo de carbono del grupo metilo y el átomo de oxígeno de las moléculas de etanol), el análisis presentado a continuación, orientado a encontrar bolsillos conservados que puedan alojar uno o ambos átomos de referencia, no hace distinción entre ellas y se refiere a estos como “sitios de etanol”.

El número de sitios de etanol (ES, por sus siglas en inglés) considerado para una comparación posterior entre serotipos fue reducido mediante el siguiente filtro: si dos o más ES estaban separados por menos de 2 Å, se consideró que pertenecían al mismo bolsillo y se representaron mediante un único ES ubicado en la posición promedio. Luego, se realizó una alineación de los esqueletos proteicos de los cuatro serotipos, y se conservaron los bolsillos filtrados si se encontraban a menos de 4 Å entre sí, analizando por separado los sistemas con y sin ARN. Finalmente, los residuos adyacentes a los bolsillos de NS3hel presentes en ambos estados se reportan en este capítulo a partir de una inspección visual realizada con el programa VMD.

4.3.6.3 Conservación de aminoácidos presentes en los bolsillos conservados

Para cada serotipo, se descargaron secuencias completas del genoma de DENV desde GenBank, y se extrajeron los últimos 2500 aminoácidos (correspondientes aproximadamente a las regiones NS3-4-5). Luego, se seleccionó el segmento comprendido entre los residuos 180 y 618 de NS3 utilizando el programa Jalview [62]. Se eliminaron las secuencias que contenían cinco o más aminoácidos consecutivos desconocidos y, finalmente, se realizó un alineamiento con ClustalW2 [63].

4.4 Resultados

4.4.1 Conformación del sitio activo de NS3hel unido a ARN favorece la hidrólisis de ATP

4.4.1.1 Participación del residuo Glu conservado (motivo DEAH o motivo II) en el mecanismo de reacción

Las estructuras cristalográficas correspondientes al complejo de Michaelis de NS3hel del DENV en presencia y ausencia de ARN sugieren una posible base estructural para la hidrólisis de ATP estimulada por ARN. Se ha sugerido que el desplazamiento del residuo Glu 285 desde la esfera de coordinación del catión divalente en el complejo unido a ARN favorece su papel como base general en la reacción. Para evaluar esta hipótesis, se compararon los mecanismos asistidos por base y por sustrato para la hidrólisis de ATP en presencia y ausencia de ARN (Figura 4.2). En el mecanismo asistido por base, el ataque nucleofílico del fosfato gamma por una molécula de agua catalítica es facilitado por Glu 285 actuando como base general, mientras que en el mecanismo asistido por sustrato, el grupo fosfato gamma actúa como la base. En lugar de sesgar el mecanismo exacto, se usaron los métodos nudged elastic band (NEB) y adaptive string method (ASM) para comparar caminos de mínima energía entre los estados de reactivos y producto en los complejos de Michaelis libres y unidos a ARN.

Inicialmente se exploraron las superficies de energía potencial utilizando el método NEB y el funcional híbrido DFT B3LYP con el conjunto base 6-31G* [48]. En el sistema libre de ARN, ambos mecanismos presentaron barreras de activación altas, del orden de ~50 kcal/mol (Anexo Figura 4A.1). En contraste, en el estado unido a ARN, la reacción procede exclusivamente a través del mecanismo asistido por base, con una barrera significativamente menor, aproximadamente 20 kcal/mol (Anexo Figura 4A.1); además, se observó protonación espontánea de Glu 285 (E-H) al intentar muestrear el mecanismo asistido por sustrato.

Se utilizaron estas simulaciones NEB para definir un espacio de variables colectivas (CV) y se continuó con el método ASM [52] para explorar caminos de mínima energía libre que conectan los estados de reactivos y productos, utilizando cuerdas de 48 nodos, en comparación con los 13 nodos empleadas en la optimización con NEB. Para compensar el mayor costo computacional, el nivel teórico utilizado para la región QM se cambió a DFTB3. Debido a este cambio en la descripción QM, primero se realizaron simulaciones QM/MM de 50 ps usando estructuras optimizadas obtenidas con NEB. Sorprendentemente, estas trayectorias revelaron reordenamientos espontáneos en la región QM. En la trayectoria del estado de reactivos del complejo unido a ARN, el enlace Pγ–Oβγ de la molécula de ATP se

alarga, conduciendo a la formación de un intermedio pentavalente de metafosfato, que no se observó en las optimizaciones iniciales con NEB (Figura 4.3A). En ausencia de ARN, el enlace Pγ–O permaneció estable. En la trayectoria del estado de productos del complejo libre de ARN, se observó un intercambio de protones, explorando tanto las configuraciones E-H/HPO_4^{2-} como E/$H_2PO_4^-$ en la simulación corta, mientras que la configuración E-H/HPO_4^- fue estable en el complejo unido a ARN (Figura 4.3B).

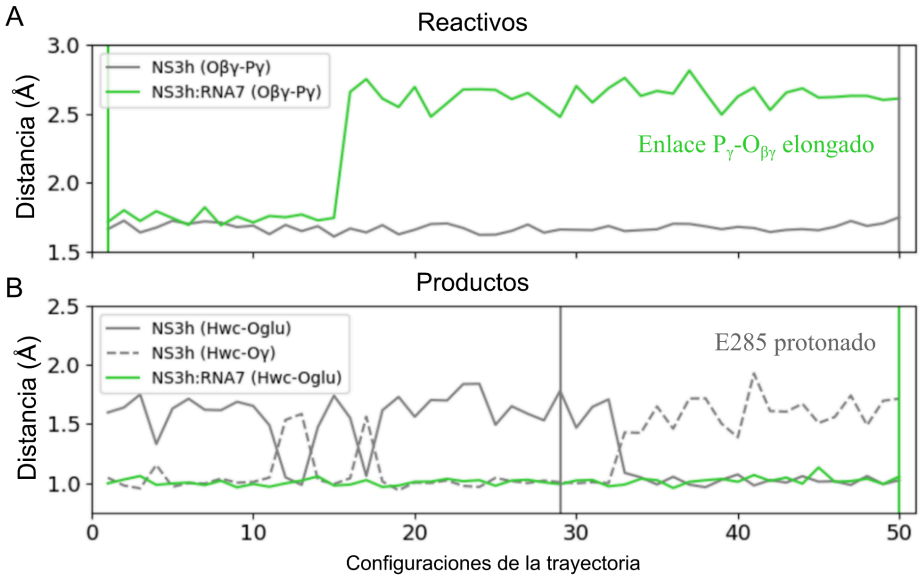


Figura 4.3 Simulaciones cortas QM/MM para equilibrar los estados de reactivos **(A)** y productos **(B)**. Las distancias mostradas fueron seleccionadas para evidenciar la estabilidad de cada estado. Las líneas verticales representan las configuraciones empleadas para la exploración de caminos de mínima energía libre mediante el método de la cuerda ASM.

En base a esta exploración inicial, se seleccionó una configuración de estado de reactivos similar a la estructura cristalográfica, es decir, antes del estiramiento del enlace Pγ–Oβγ en presencia de ARN. Se eligió el estado de productos E-H/HPO_4^{2-} para el complejo unido a ARN (consistente con un mecanismo asistido por base) y la configuración E/$H_2PO_4^-$ para el complejo libre de ARN (consistente con un mecanismo asistido por sustrato). Las variables colectivas usadas para los cálculos ASM fueron las siguientes distancias para ambos complejos: Pγ–Oβγ, Pγ–OWc, OWc–HWc, OE285–HWc. Para el complejo libre de ARN se agregó una distancia adicional, Oγ–HW (Figura 4.3).

Los perfiles de energía libre se muestran en la Figura 4.4 para NS3hel en los complejos libres y unidos a ARN, junto con las configuraciones de los reactivos (R), primer estado de transición (TS1, por sus siglas en inglés), intermediario (I), segundo estado de transición (TS2, por sus siglas en inglés), Glu 285 protonado (E-H) y productos (P). La diferencia clave entre los dos perfiles energéticos radica en la formación del estado intermediario: en el sistema libre de ARN, este paso es endergónico con una barrera de 13 ± 2 kcal/mol, mientras que en el sistema unido a ARN, es exergónico con una barrera menor, de aproximadamente 5 kcal/mol. En ambos casos, las barreras corresponden a la formación del enlace Pγ–OWc. La posterior transferencia de protón al átomo de oxígeno del HPO_4^{2-}, junto con la elongación del enlace H–O de la molécula de agua catalítica tras el ataque, contribuyen a la disminución global de energía libre. La contribución de cada variable colectiva al perfil de energía libre, así como la evolución de las distancias a lo largo de la coordenada de reacción obtenida del

camino de mínima energía libre, se muestran en la Figura 4A.2 del Anexo. Aunque el Glu 285 protonado (E-H) participa transitoriamente en la reacción en ausencia de ARN, la alta energía libre de esta configuración impide considerarla un intermediario estable, a diferencia del estado unido a ARN. Cabe destacar que en el sistema libre de ARN, Glu 285 se disocia transitoriamente del ion Mg^{2+} en el estado de transición TS2.

Aunque la reacción explorada representa solo la primera parte del proceso de hidrólisis, sin considerar eventos subsiguientes de tautomerización o la disociación de la molécula de Pi del centro de Mg^{2+}, es interesante observar que esta conversión inicial de ATP a ADP y Pi es endergónica en el estado libre de ARN y exergónica en el estado unido a ARN. Además, a pesar de la selección inicial del estado de productos para reflejar mecanismos asistidos por base o por sustrato, se concluye que la transferencia de protón desde la molécula de agua catalítica no limita la velocidad de la reacción, independientemente de la presencia de ARN y de la diferente coordinación de Glu 285 al centro Mg^{2+}.

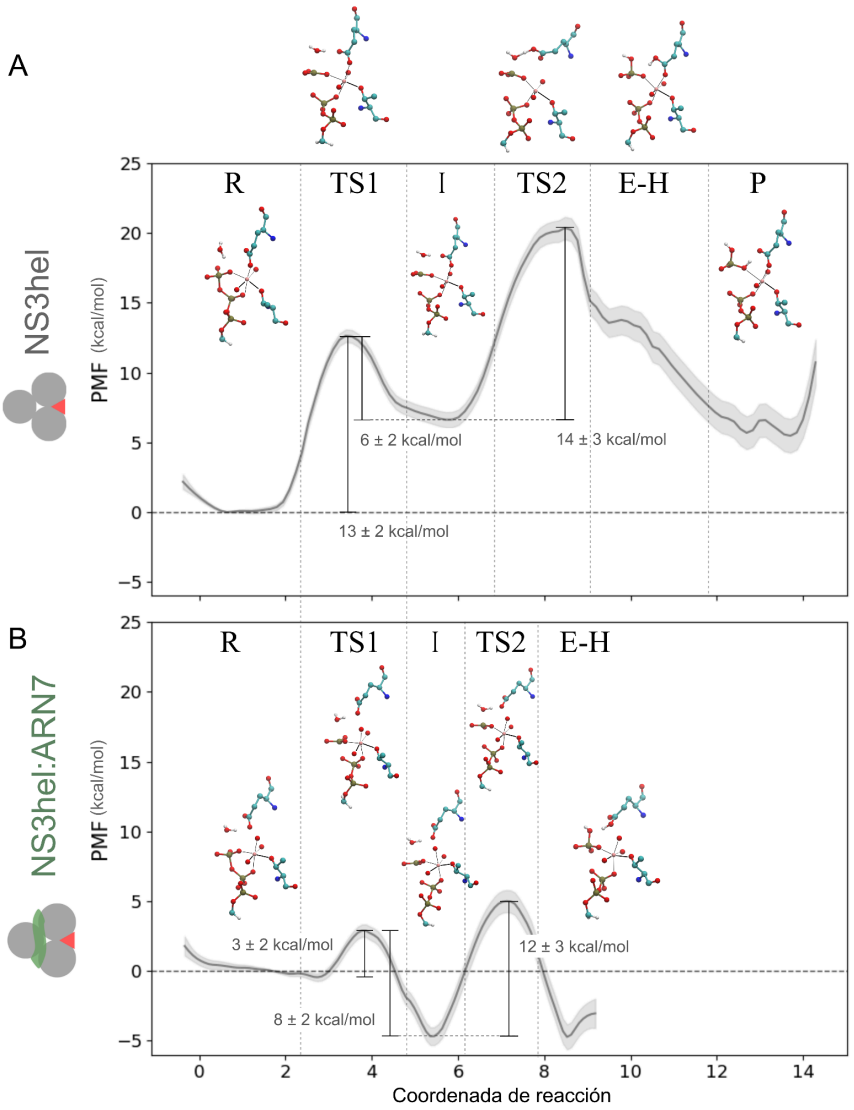


Figura 4.4. Potencial de fuerza media (PMF, kcal/mol) a lo largo de la coordenada de reacción para la hidrólisis de ATP de la NS3hel sin **(A)** y con **(B)** ARN. Las líneas discontinuas verticales dividen las etapas significativas de la reacción: estado de reactivos (R), estados de transición (TS), estado

intermediario (I) y estado de productos (P). El Glu 285 protonado está etiquetado como E-H. En el sistema sin ARN, el progreso de la reacción llega hasta la formación de fosfato diácido y ADP (ver mecanismo asistido por sustrato en la Figura 4.2), mientras que en el sistema con ARN, la reacción progresa hasta la formación de fosfato monoácido y ADP, con el Glu 285 protonado (ver mecanismo asistido por base en la Figura 4.2).

4.4.1.2 Estabilización del intermediario metafosfato pentavalente

A pesar de la observación de una nueva conformación durante la etapa de equilibración, se seleccionó una configuración del estado de reactivos en la que el ATP mantenía distancias más cercanas a las observadas en la estructura cristalográfica (Figura 4.3). En consecuencia, el complejo intermediario observado durante la equilibración inicial también surgió durante la optimización con ASM, luego de una barrera de energía baja, como era de esperar. Este intermediario metafosfato pentavalente también se observó en el sistema sin ARN, aunque con una barrera energética más alta (13 frente a 5 kcal/mol). La variable colectiva que más contribuyó al aumento de energía en el estado de transición TS1 correspondió a la ruptura del enlace Pγ–Oβγ del ATP (Anexo Figura 4A.2).

A lo largo de la reacción, el sistema con ARN no solo genera un cambio en la contribución de Glu 285, sino también un enlace de puente hidrógeno adicional con el agua catalítica, la cual interactúa simultáneamente con Gly 414 y Gln 456 en el sistema con ARN. En el complejo sin ARN, solo una de estas interacciones es posible: mientras que la interacción de la Wc con Gly 414 (motivo V) es estable solo en los estados de reactivos y TS1, la interacción de Gln 456 (motivo VI) con Wc ocurre en TS2 y en el estado de productos, pero no logra estabilizar los estados de reactivos, TS1 o intermediario (Figura 4.5). El ángulo OWc–Pγ–Oβγ de aproximadamente 170° se mantiene durante toda la reacción de hidrólisis en el sistema con ARN, contribuyendo a la estabilidad del intermediario metafosfato, mientras que el valor para el sistema sin ARN aumenta de 140 a ~170° recién en el segundo estado de transición y en el estado de productos (el análisis detallado del ángulo OWc–Pγ–Oβγ y las interacciones durante la hidrólisis se muestra en las Figura 4.5 y Anexo Figura 4A.3).

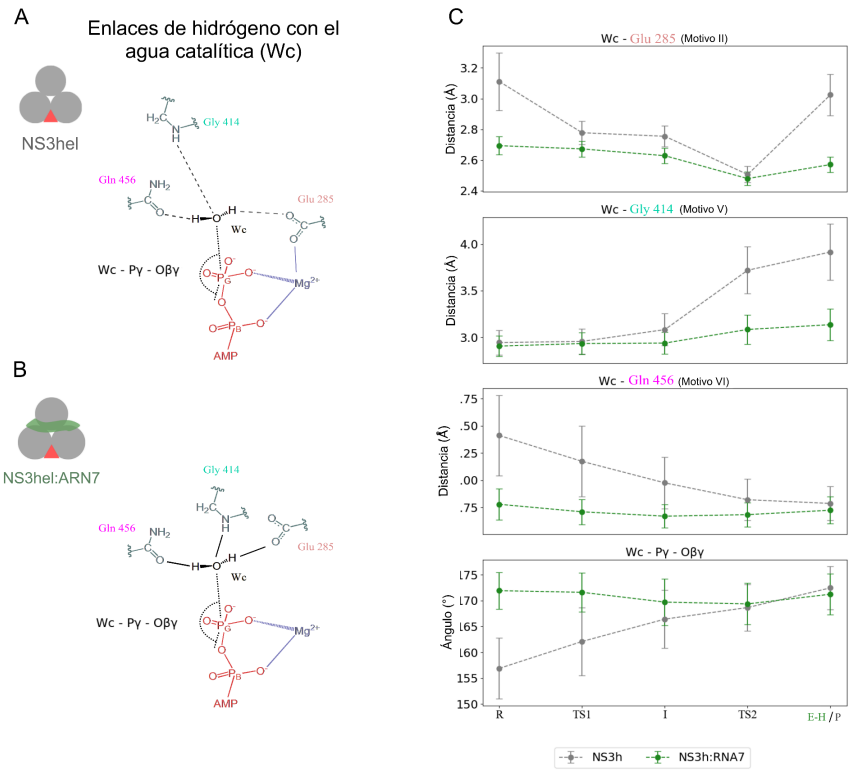


Figura 4.5. Interacciones de los residuos catalíticos con el agua catalítica para **(A)** NS3hel:ATP y **(B)** NS3hel:ARN7:ATP. Las etiquetas están coloreadas según el código de colores de los motivos conservados de la Figura 4.1. Se muestran las interacciones con líneas sólidas (siempre presente) y discontinuas (parcialmente presente) entre los residuos catalíticos y el agua catalítica que forman enlaces de hidrógeno. Además se muestra el ángulo formado por OWc–Pγ–Oβγ. **C)** Las distancias relacionadas con las interacciones del Panel B y el ángulo señalado se grafican a lo largo de las etapas definidas de la Figura 4.4 para la reacción de hidrólisis de ATP. Las series en gris corresponden a NS3hel:ATP y las verdes a NS3hel:ARN7:ATP.

En conjunto, estas observaciones destacan que la unión del ARN no solo disminuye la barrera de activación para la formación del intermediario metafosfato pentavalente, sino que también reorganiza la red de enlaces de hidrógeno para facilitar el primer paso de la hidrólisis de ATP.

4.4.2 La presencia y longitud del ARN impactan diferencialmente en la dinámica de los residuos del sitio catalítico y del agua catalítica

Para investigar el impacto del ARN sobre la conformación del sitio catalítico, se realizaron simulaciones clásicas de dinámica molecular del complejo de Michaelis tanto en ausencia como en presencia de ARN. En los sistemas con ARN se consideraron dos longitudes diferentes, de 7 y 16 nucleótidos, basándonos en trabajos previos en los que se determinó que una mayor longitud de ARN es crítica para representar con precisión la comunicación alostérica entre los sitios de unión de ARN y NTP [6]. A lo largo del texto, se hará referencia a estos tres estados como NS3hel:ATP, NS3hel:ATP:ARN7A y NS3hel:ATP:ARN16A, donde la 'A' indica que la secuencia de ARN fue cambiada por una secuencia de adeninas (poliA).

Se analizaron las distancias entre Glu 285 (motivo II) y Mg^{2+} para evaluar la coordinación del catión divalente, así como las distancias entre el fosfato gamma (Pγ) y residuos conservados

que interactúan con el agua catalítica (Gly 414, motivo V) o con el ATP (Lys 199, motivo I; Arg 460 y Arg 463, motivo VI). Si bien la mayoría de las distribuciones de distancia estuvieron centradas cerca de los valores cristalográficos, las correspondientes al sistema con ARN de 7 nucleótidos, en ocasiones mostraron distribuciones más amplias que las del sistema con ARN de 16 nucleótidos (Figura 4.6, curvas en verde claro y oscuro, respectivamente). Por su parte, el complejo sin ARN presentó las mayores desviaciones (Figura 4.6, curva gris) en algunas distribuciones, como se detalla a continuación.

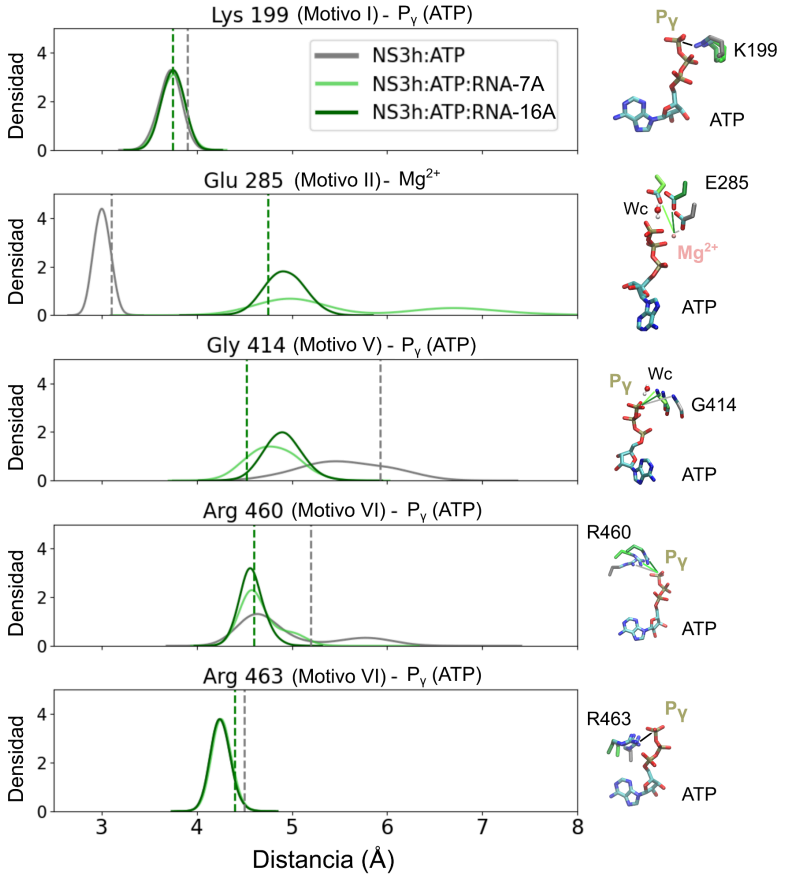


Figura 4.6. Estabilidad de interacciones clave entre NS3hel y ATP-Mg^{2+} en presencia y ausencia de ARN. Cada curva representa la distribución de las distancias entre un residuo catalítico y el fosfato gamma (Pγ) o el Mg^{2+}. El eje x indica la distancia, mientras que el eje y corresponde a la densidad estimada. Las líneas grises representan NS3hel sin ARN, las verdes representan NS3hel con ARN (verde oscuro para ARN largo y verde claro para ARN corto). Las líneas discontinuas verticales indican el valor de cada distancia en las estructuras cristalográficas, con el mismo código de colores. En cada panel, el ATP y los residuos superpuestos de NS3hel sin ARN, con ARN-7 y con ARN-16 se muestran con una representación tipo bastones. Si la interacción involucra al agua catalítica, se incluye una molécula de agua representada con esferas y cilindros.

Se observó que el catión divalente conservó un patrón de coordinación distinto según la presencia de ARN. En ausencia de ARN, el Glu 285 (Motivo II) permaneció coordinado al Mg^{2+}, mientras que en los complejos con ARN esta posición de coordinación se mantuvo ocupada por una molécula de agua. En el complejo NS3hel:ATP:ARN-16A, la distancia entre el Glu 285 y el fosfato γ (Pγ) permaneció cercana al valor cristalográfico a lo largo de toda la trayectoria. En cambio, con un ARN más corto, el Glu 285 se alejó parcialmente del Pγ.

Utilizando el mismo protocolo NEB descrito anteriormente, se verificó que un estado de reactivos con Glu 285 en esta conformación, explorado con ARN corto, resultó en una barrera más alta para el primer paso de la reacción, similar a la observada en ausencia de ARN (~45 kcal/mol).

El residuo Gly 414 (Motivo V), que estabiliza el agua catalítica, permaneció a una distancia óptima del Pγ en ambos complejos con ARN, aunque levemente desplazado hacia valores mayores en comparación con la estructura cristalográfica. Estas configuraciones permiten el establecimiento de puentes de hidrógeno con una molécula de agua posicionada para el ataque nucleofílico, los cuales, al iniciarse la reacción, estabilizarán el estado intermediario y los estados de transición durante la hidrólisis del ATP. En el complejo sin ARN, la distribución de distancias fue más amplia y Gly 414 sólo ocasionalmente se acercó al Pγ. Para evaluar el impacto de esta posición, se seleccionó una imagen representativa de la trayectoria sin ARN con una distancia reducida Gly 414–Pγ de 4.7 Å. A modo de comparación, las distancias cristalográficas son de 5.8 Å en la estructura sin ARN y de 4.4 Å en la estructura con ARN, lo que indica que esta configuración se asemeja a la geometría del estado con ARN. Usando el mismo protocolo NEB, se estimaron las barreras energéticas para la hidrólisis del ATP y se encontró que, si bien ambos mecanismos eran accesibles con una barrera menor (~40 kcal/mol), los valores seguían siendo significativamente más altos que en presencia de ARN, lo que indica que esta variación en la conformación del Motivo V no es crítica para la diferencia mecanística observada.

Las interacciones electrostáticas entre el ATP y tanto Arg 463 (Motivo VI, “dedo de arginina”) como Lys 199 (Motivo I) fueron estables en todos los sistemas, mostrando distribuciones similares a pesar de leves variaciones respecto a los valores cristalográficos. En cambio, el puente salino entre Arg 460 (Motivo VI, “dedo de arginina”) y el fosfato γ fue menos estable en el complejo sin ARN. El ARN de 16 nucleótidos estabilizó esta interacción, acercándose al valor cristalográfico, mientras que la configuración con ARN corto mostró una leve disminución en la estabilidad, con desviaciones hacia distancias intermedias, aunque no tan largas como las observadas en el estado sin ARN.

En conjunto, se observó que la dinámica de los residuos del sitio catalítico pertenecientes a motivos conservados varió no sólo con la presencia de ARN sino también con su longitud. En particular, la simple cadena de ARN más larga contribuyó a estabilizar la conformación de los residuos catalíticos que establecen diferencias clave en las rutas de reacción, lo que presumiblemente favorece la hidrólisis del ATP.

4.4.3 La conformación del ARN afecta el posicionamiento del agua catalítica

Las simulaciones de dinámica molecular del complejo NS3hel:ATP:ARN-16A mostraron grandes fluctuaciones conformacionales en el extremo 5’ del ARN, alternando entre dos grupos principales con interacciones debidas al posicionamiento diferencial cercano a los subdominios D2 o D3, respectivamente (Figura 4.7). Estas conformaciones se denominarán “5’-D2” y “5’-D3” (Figura 4.7A). Se clasificaron las imágenes de la trayectoria a partir del algoritmo de *clustering* detallado en la Sección de 4.3.5.1 de los métodos de este capítulo, separando la trayectoria en dichas conformaciones o grupos (*clusters*). En un trabajo previo, se identificó una conformación abierta novedosa del *loop* β6–β1′ (residuos 316–330) [4], que

abre un canal adyacente al sitio catalítico para la liberación de productos. La inspección visual de las conformaciones del *loop* β6–β1′ en las trayectorias separadas en los dos grupos reveló que la conformación 5’-D2 mantuvo consistentemente el *loop* cerrado, mientras que 5’-D3 mostró mayor flexibilidad, explorando conformaciones “semiabiertas” (Figura 4.7B).

En el sistema sin ARN, la conformación cerrada del *loop* se estabiliza mediante una interacción entre Gln 456 (motivo VI) y Thr 317 (motivo III), un residuo del *loop* β6–β1′ (Figura 4.7C, curva gris). En contraste, el complejo unido a ARN-7A mostró una conformación semiabierta, reflejada en un corrimiento hacia mayores distancias Gln–Thr (Figura 4.7C, verde). El grupo correspondiente a la conformación 5’-D2 se asemeja a la distribución sin ARN, aunque con distancias levemente mayores (Figura 4.7C, violeta), mientras que estos valores para 5’-D3 se desplazaron completamente hacia la conformación semiabierta (Figura 4.7C, celeste). Para comparar más detalladamente las conformaciones del *loop*, se calculó el RMSD de los átomos del esqueleto del *loop* β6–β1′ respecto a la estructura cristalográfica. Se utilizó como referencia del estado abierto una imágen con el *loop* abierto tomado del trabajo anterior [4] (Figura 4.7D). El rol de Gln 456 (motivo VI) en la hidrólisis de ATP fue discutido en la subsección A. Se mostró que, en presencia de ARN, se forma un enlace puente hidrógeno estable entre Gln 456 y el agua catalítica durante toda la reacción, mientras que en ausencia de ARN esta interacción aparece recién a partir del segundo estado de transición. En todos los análisis de hidrólisis de ATP descritos previamente, el *loop* β6–β1′ permaneció en una conformación cerrada, consistente con la estructura cristalográfica.

A continuación, se evaluó si estas diferencias en la conformación del *loop* β6–β1′ están relacionadas con la probabilidad de encontrar una molécula de agua catalítica en el sitio activo. Se definió como agua catalítica aquella con un átomo de oxígeno a menos de 2.7 Å del átomo Pγ del ATP y con un ángulo Oβγ–Pγ–OWc entre 160° y 180°. Todos los sistemas mostraron al menos una molécula de agua cerca del Pγ en casi todas las imágenes de la trayectoria sin aplicar restricciones angulares. Sin embargo, al restringir a geometrías catalíticamente competentes, la probabilidad disminuyó notablemente en el sistema sin ARN (33 %, Figura 4.7E, barra gris), mientras que el sistema unido a ARN-7A retuvo aproximadamente el 60 %. Sorprendentemente, el sistema NS3hel:ATP:ARN-16A mostró una probabilidad general más baja, similar a la observada en ausencia de ARN. No obstante, al separar las trayectorias de ARN-16A en las conformaciones 5’-D2 y 5’-D3, se encontró que la probabilidad de presencia de un agua catalítica para el grupo 5’-D3 (61 %, barra celeste en Figura 4.7E) fue similar a la del complejo con ARN-7, mientras que para el grupo 5’-D2 (34 %, barra violeta) lo fue al sistema sin ARN.

Estos resultados refuerzan el papel de la conformación del ARN en la modulación de la posición de residuos conservados relacionados con la actividad catalítica de NS3hel. En particular, la disrupción de la interacción Gln 456–Thr 317, observada en la conformación 5’-D3 del ARN largo, coincide con un estado semiabierto del lazo β6–β1′ que facilita el ingreso y posicionamiento del agua catalítica en el sitio activo. En cambio, la conformación cerrada del lazo, con una interacción Gln–Thr estable, mantenida en el sistema sin ARN y en el agrupamiento 5’-D2 del ARN largo, parece restringir el acceso de moléculas de agua (Figura 4.7F). Esto indica que no solo la presencia, sino también la conformación del ARN unido, gobierna la accesibilidad y organización del sitio catalítico, modulando la probabilidad de hidrólisis de ATP mediante efectos estructurales y dinámicos. En conjunto, estos hallazgos

revelan un nuevo rol del lazo β6–β1′ en la regulación del acceso del agua catalítica, vinculando cambios conformacionales inducidos por el ARN con la eficiencia hidrolítica.

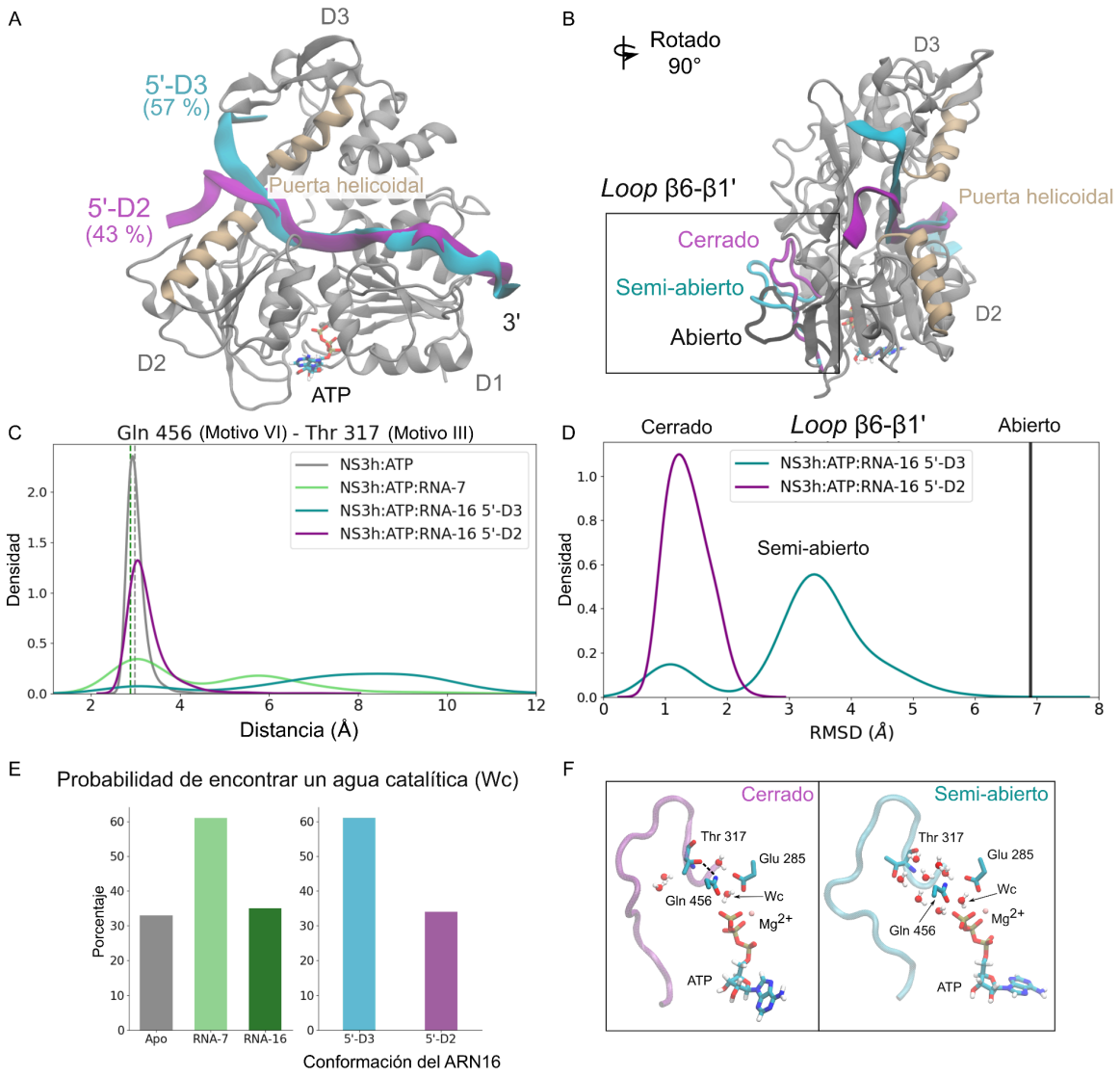


Figura 4.7. Efecto alostérico de la conformación del RNA sobre el sitio catalítico. **A)** Se muestra el esqueleto de la helicasa NS3hel en forma de caricatura con cintas en gris unida a un RNA de 16 nucleótidos representado como una cinta. Se superponen las conformaciones del RNA: una imágen del grupo de conformaciones 5'-D3 está en cian y de conformación 5'-D2, en violeta. Se indica el porcentaje de cada conformación obtenido luego del *clustering*. Se incluyen etiquetas para los subdominios D1, D2 y D3, así como la molécula de ATP, mostrada con representación en bastones coloreados según el átomo. La puerta helicoidal (*helical gate,* α-hélices 2' y 6'') está resaltada en dorado, a modo de referencia. **B)** El complejo NS3hel:ARN16 está rotado 90° respecto al eje vertical con respecto a la representación superior izquierda. El *loop* β6–β1' está resaltado con colores según la conformación del RNA, que corresponde a las conformaciones cerrada y semiabierta, con el código del panel A. La conformación abierta mostrada en negro fue reportada en un trabajo previo [4]. **C)** Estimación de densidad *kernel* de la distancia entre Gln 456 y Thr 317 usando el mismo código de colores del panel A. **D)** Estimación de densidad *kernel* del RMSD calculado para los átomos del esqueleto del *loop* β6–β1' (residuos 316-330). El RMSD se refiere a la estructura cristalina (conformación cerrada) y el valor para la conformación abierta se calculó usando la imagen del panel B con la misma referencia. **E)** Probabilidad de encontrar una molécula de agua en posición catalítica para

NS3hel en ausencia de RNA (gris), en presencia de RNA (verde claro para 7-mer y verde oscuro para 16-mer) y luego de agrupar el RNA en conformaciones 5'-D3/D2 (cian y púrpura, respectivamente). **F)** Imágenes representativas de los estados cerrado (conformación 5'-D2 de RNA 16-mer) y semiabierto (conformación 5'-D3 de RNA 16-mer) del *loop* β6–β1'. La línea discontinua en el panel izquierdo indica la interacción estable entre Gln 456 y Thr 317. ATP, Glu 285 y Mg^{2+} se muestran como referencia dentro del sitio catalítico, así como las moléculas de agua del entorno local, incluida el agua catalítica (Wc).

4.4.4 Implicancias para nuevas estrategias terapéuticas: sitios crípticos pan-serotipo para el descubrimiento de antivirales

Los resultados hasta ahora demuestran cómo cambios sutiles en la configuración de los motivos conservados involucrados en la hidrólisis de ATP favorecen la formación del intermediario metafosfato pentavalente, y cómo tanto la presencia del ARN como su conformación pueden promover la correcta posición del agua catalítica a través de la dinámica del *loop* β6–β1'. La hidrólisis de ATP estimulada por ARN, de esta manera, se muestra como un mecanismo adecuado para la intervención alostérica mediante pequeñas moléculas.

Para evaluar esta hipótesis, se identificaron bolsillos (sitios drogables o sitios crípticos) conservados en los cuatro serotipos del DENV para un desarrollo antiviral futuro. Se realizaron simulaciones de dinámica molecular con solvente mixto usando una mezcla 20% etanol-agua [61] tanto en la proteína NS3hel apo como en el complejo NS3hel:ARN7 para los cuatro serotipos del DENV. Los serotipos sin estructura cristalográfica resuelta fueron modelados por homología [34].

A través de las simulaciones con solvente mixto se identificaron cuatro sitios drogables o bolsillos (Figura 4.8), caracterizados por residuos altamente conservados entre los serotipos del DENV (Anexo Tabla 4A.1). Dos de ellos, incluyen motivos conservados que han demostrado ser clave en la hidrólisis de ATP estimulada por ARN. El primero se denomina "bolsillo puerta helicoidal" (Figura 4.8A), rodeado por los motivos IV, IVa y V, y el segundo es el "bolsillo *loop* β6–β1'," que incluye residuos de los motivos II y III. Los otros dos bolsillos crípticos conservados no incluyen ninguno de los motivos conservados y se denominan "bolsillo D3" y "bolsillo RNA-D3," ambos adyacentes a una región hélice-giro-hélice (residuos 528 a 555) del subdominio D3.

En base a un análisis de correlación generalizada de trayectorias concatenadas de 2.7 μs del serotipo DENV2 con y sin ARN (simulaciones analizadas en las Secciones 4.4.2 y 4.4.3, Anexo Figura 4A.4), se encontró que esta región está comunicada alostéricamente con los siguientes motivos conservados: I, IA, II, III y IV. Para los sistemas con ARN, se realizaron análisis separados para los ARN de 7 y 16 nucleótidos. En el caso del ARN más largo, las trayectorias se dividieron además en los estados conformacionales 5'-D2 y 5'-D3. Sin embargo, no se observaron diferencias sustanciales en los patrones de conectividad alostérica bajo estas condiciones.

Cabe destacar que el motivo II incluye a Glu 285 y el motivo III es adyacente al *loop* β6–β1', ambos elementos identificados como clave en la hidrólisis de ATP estimulada por ARN (Figura 4.8B). Por lo tanto, estos bolsillos sensibles al ARN (D3 y RNA-D3) podrían ser la

base para el desarrollo de inhibidores alostéricos e inspirar el diseño de virus atenuados específicamente con dominios helicasa NS3 modificados.

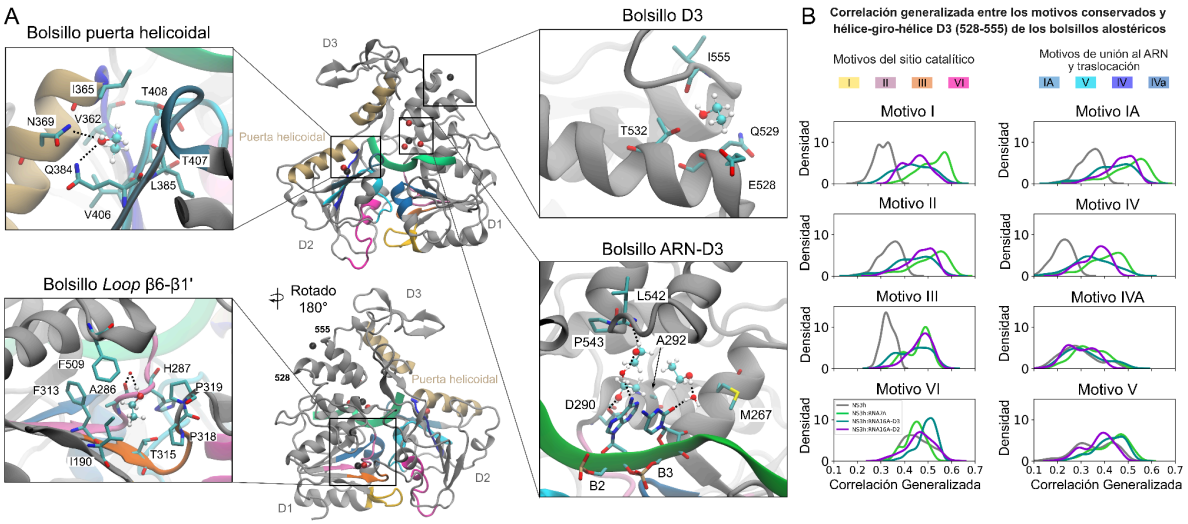


Figura 4.8. Bolsillos conservados entre serotipos en NS3hel obtenidos con y sin ARN a partir de sitios de etanol. **A)** Representación en forma de caricatura con cintas en gris de NS3hel, ARN con una cinta verde, la puerta helicoidal (*helical gate,* α-hélices 2' y 6'') marcada en dorado y motivos conservados coloreados según el código del panel B. Se indican los subdominios D1, D2 y D3. Las esferas negras y rojas representan los sitios de etanol correspondientes a los grupos metilo e hidroxilo, respectivamente. Las regiones ampliadas muestran moléculas de etanol en representación de esferas y cilindros, y residuos del entorno local en bastones, con etiquetas. Las líneas discontinuas indican interacciones entre las moléculas de etanol y los residuos del bolsillo. Los residuos extremos de la región hélice-giro-hélice del subdominio D3 están etiquetados en el complejo NS3hel–ARN rotado 180° respecto del eje vertical con relación a la vista superior. **B)** Estimación de densidad *kernel* de la correlación generalizada entre la región hélice-giro-hélice de D3 y los motivos conservados. Las curvas grises corresponden al sistema sin ARN, las verdes al complejo con ARN de 7 nucleótidos, y las curvas cian y violeta-rosa, al complejo con ARN de 16 nucleótidos.

4.5 Discusión

4.5.1 Detalles mecanísticos de la hidrólisis de ATP y el rol de E285

El valor de la constante de velocidad (k_{cat}) de la actividad ATPasa de NS3hel de DENV2 es de aproximadamente 3 s^{-1} en ausencia de ARN y de casi 28 s^{-1} en presencia de ARN [7]. Estos valores de k_{cat} dependen de múltiples etapas del ciclo catalítico, que incluyen la unión del ATP, la energía de activación del paso catalítico y la liberación de productos. Sin embargo, resultan útiles como límites inferiores para las constantes de velocidad —o límites superiores para las energías de activación— de cada uno de estos procesos. Usando la ecuación de Eyring, estos límites superiores se estiman en aproximadamente 17 kcal/mol sin ARN y 16 kcal/mol con ARN. Las barreras reportadas son consistentes con estos límites experimentales superiores y aportan información mecanística sobre la modulación de la actividad ATPasa de NS3hel por el ARN.

En particular, se observa que la formación del intermedio metafosfato, junto con el posterior ataque del agua catalítica, ocurre antes que la transferencia de protón desde esta molécula

de agua. En este sentido, los resultados sugieren que, aunque E285 acepta este protón más fácilmente en el estado con ARN, y puede o no hacerlo en el estado sin ARN, ninguna de los caminos de reacción de mínima energía identificados refleja un mecanismo catalizado por una base general. Esta observación, es decir, que la transferencia del protón desde la molécula de agua catalítica ocurre después del ataque nucleofílico y no es el paso limitante de la reacción, es coherente con lo observado en reacciones de hidrólisis estudiadas en otras NTPasas [12].

A pesar del consenso emergente sobre la necesidad (o la prescindencia) de una base general en las NTPasas de tipo DEAH-box, hay que reconocer que la exploración exhaustiva de todos los caminos de reacción posibles —algunos de los cuales pueden ser isoenergéticos— sigue siendo un desafío. Una muestra de esta dificultad surge al comparar más detalladamente los mecanismos de reacción. Usando el método de la banda elástica (NEB), los caminos de energía mínima (CME) reflejan un mecanismo concertado (tipo SN2) tanto para el estado sin ARN como para el estado con ARN (ver Anexo Figura 4A.1, similar al mecanismo propuesto por [27]), mientras que los caminos obtenidos mediante el método de la cuerda (ASM) corresponden a un mecanismo disociativo, tipo SN1. Una fuente probable de esta diferencia es el cambio en la resolución a lo largo de la banda/cuerda, con 13 nodos utilizados en los cálculos NEB y 48 en los cálculos ASM. La estabilidad relativa del intermedio metafosfato también se ve claramente afectada por el nivel de teoría elegido para describir la región QM [12].

Si bien este estudio proporciona un panorama integral sobre la influencia del ARN en la hidrólisis de ATP catalizada por NS3hel, quedan varios aspectos por resolver. Entre ellos se incluyen el rol y el número de moléculas de agua involucradas en la transferencia de protón, vías alternativas de aceptación de protón hacia el grupo fosfato o el ADP, así como la posterior disociación del centro metálico divalente, variables que sería relevante considerar en estudios futuros.

4.5.2 Rol del ARN en la modulación de la hidrólisis de ATP

Los resultados obtenidos destacan cómo la unión del ARN influye en la conformación del sitio catalítico de NS3hel, favoreciendo el perfil energético del mecanismo de hidrólisis de ATP. Esta modulación surge no sólo de cambios en la coordinación del catión divalente, sino también de la contribución de residuos conservados que forman enlaces de puente hidrógeno con la molécula de agua catalítica. La importancia de estos enlaces de puente [26] hidrógeno, en particular mediante el motivo V, ha sido resaltada en otros estudios de la hidrólisis de ATP por la NS3hel del DENV [26] y por la NS3hel del virus del Zika (ZIKV) [12]. De manera similar, la posición dependiente del ARN de la molécula de agua catalítica ha sido observada en otros estudios sobre la NS3hel del DENV y sobre Prp2, otra helicasa del tipo DEAH-box [27,28].

El estudio de la hidrólisis de ATP en la NS3hel del ZIKV con Glu 286 desplazado de la esfera de coordinación del metal divalente incluso en ausencia de ARN [12] podría poner en duda si la conformación catalíticamente favorable es necesariamente consecuencia de la unión del ARN. Sin embargo, cabe señalar que la estructura sin ARN utilizada por García Martínez et al. fue obtenida con un análogo del estado de transición en el sitio catalítico (PDB ID 6S0J), mientras que otras estructuras de NS3hel del ZIKV sin ARN proporcionan evidencia de que Glu 286 se encuentra coordinado al catión metálico divalente (PDB IDs 6ADY, 5K8U, 6RWZ, 5GJC). En conjunto, estas observaciones respaldan la hipótesis de que el rol mecanístico del

ARN en la modulación de la hidrólisis de NTPs probablemente se conserve entre las helicasas del tipo DEAH-box.

4.5.3 Dinámica del extremo 5′ del ARN unido y del *loop* β6–β1′

Los resultados obtenidos destacan el rol fundamental de la longitud del ARN en el modelado del complejo NS3hel:ARN, ya que la dinámica de residuos clave del sitio catalítico en la hidrólisis estimulada por ARN —como Glu 285 (motivo II), Gly 414 (motivo V) y Arg 460 (motivo VI)— depende de la longitud del oligonucleótido (Figura 4.6), siendo el ARN simple cadena de 16 nucleótidos más efectivo que el de 7 nucleótidos para estabilizar la conformación catalíticamente óptima de estos residuos.

Además de su impacto sobre los motivos del sitio catalítico, se encontró que diferentes conformaciones del ARN en el extremo 5′, evidenciadas en modelos con oligonucleótidos más largos, influye en la dinámica del *loop* β6–β1′ y en la localización de moléculas de agua en el sitio catalítico. En trabajos previos, se identificó una conformación abierta de este loop que crea un canal alternativo para la liberación de productos, mientras que en el presente trabajo se determinó que su dinámica depende de la longitud del ARN [4–6]. En particular, cuando el ARN adopta la conformación “5′-D3” (es decir, con el extremo 5′ más cerca del subdominio D3), se favorece el muestreo de un estado semiabierto del *loop*, rompiéndose la interacción entre Gln 456 y Thr 317, y facilitando el acceso de moléculas de agua al sitio catalítico. Más aún, la conformación 5′-D3 favorece la interacción de los nucleótidos con la horquilla hidrofóbica β5A′–β5B′ (residuos 426 a 448), la cual ha sido propuesta como un elemento que facilita la separación de la doble cadena [6,64,65]. En cambio, la conformación “5′-D2” no sólo estabiliza el *loop* en un estado cerrado, similar al observado en ausencia de ARN, sino que también impide que la hebra de ARN interactúe con la horquilla hidrofóbica. Estos resultados refuerzan la importancia de considerar la longitud del ARN simple cadena para abordar la dinámica del complejo NS3hel:ARN y su comunicación alostérica con el sitio catalítico [6].

4.5.4 Bolsillos drogables conservados entre serotipos de la NS3hel

Los bolsillos crípticos identificados en simulaciones de NS3hel con solvente mixto de los cuatro serotipos del DENV, tanto en ausencia como en presencia de ARN, representan sitios prometedores para futuras estrategias terapéuticas. Estos sitios (Figura 4.8) se ubican cerca de motivos conservados implicados en el reconocimiento de ARN y la catálisis, o bien se encuentran conectados alostéricamente con dichos motivos.

Dos de los sitios identificados en proximidad a motivos conservados son los denominados bolsillos del “puerta helicoidal” y del “*loop* β6–β1′”. El primero se encuentra adyacente al motivo V y próximo a los residuos Ile 365 y Thr 408, ambos en contacto directo con el ARN y pertenecientes a los motivos conservados IV y IVa, respectivamente, cuya relevancia funcional ha sido demostrada en NS3hel del DENV y en otras helicasas de tipo DEAH-box [66,67]. El bolsillo del *loop* β6–β1′ representa un sitio estratégico para modular la dinámica de dicho loop, cuya importancia funcional se evidencia tanto en su rol para regular el acceso de agua al sitio catalítico (Figura 4.7) como en la liberación de productos tras la hidrólisis de ATP [4,5], además de estudios de mutagénesis dirigidos al motivo III [68]. Notablemente, este bolsillo aloja de manera recurrente una molécula de glicerol resuelta en varias estructuras cristalográficas (PDB IDs: 2WZQ, 5XC7, 2JLQ, 2JLS [65,69,70]), lo que refuerza su potencial

como sitio de unión de ligandos para el diseño racional de fármacos. En relación con esta observación, ensayos de *screening* de fragmentos por cristalografía de rayos X en la NS3hel del virus Zika (ZIKV) también han identificado esta región como blanco de pequeñas moléculas [32].

Los dos bolsillos restantes —denominados “D3” y “RNA-D3”— se encuentran comunicados alostéricamente con los residuos clave del sitio catalítico que participan en la hidrólisis de ATP estimulada por ARN. Más aún, el bolsillo D3 coincide con una de las interfaces identificadas entre NS3 y NS5 (PDB ID: 8GZR, [71]). Por su parte, el bolsillo RNA-D3, localizado cerca del surco de unión al ARN, ha sido previamente identificado en la NS3hel del virus de la hepatitis C (HCV), donde se resolvió una molécula pequeña unida a este sitio (PDB ID: 4WXR, [72]), lo que refuerza su potencial para el desarrollo de antivirales.

En conjunto, estos bolsillos conservados entre serotipos y en presencia y ausencia de ARN, constituyen una base para el diseño de inhibidores por técnicas de modelado estructural, orientados a interferir con mecanismos regulatorios clave de la función helicasa.

4.5.5 Implicancias para la actividad NTPasa *in vivo*

Estructuras recientes obtenidas por cryo-EM aportan información adicional sobre el rol de la NS3hel en distintas etapas de la replicación viral [71]. En ausencia de ARN, la estructura del complejo NS3:NS5 muestra que la NS5 bloquearía la posible apertura del *loop* β6–β1’ de NS3hel (PDB ID: 8GZQ), lo cual, según los resultados presentados, impediría tanto la hidratación del sitio catalítico como la liberación de Pi y ADP. Sin embargo, en las estructuras del complejo NS3:NS5 unido a ARN durante la etapa de elongación de la síntesis de ARN, el *loop* β6–β1’ ya no se encuentra bloqueado (PDB ID: 8GZR). De modo similar, las estructuras de NS3 de longitud completa (con ambos dominios presentes, ambas en ausencia de ARN) revelan orientaciones interdominio distintas entre la proteasa y la helicasa, en una de las cuales el dominio proteasa puede obstruir el sitio de unión al ATP de la NTPasa (PDB IDs: 2WHX, 2VBC).

Considerando que la unión al ARN potencia, pero no es un requisito para la actividad NTPasa de NS3hel, estas observaciones respaldan la hipótesis de que los reordenamientos en las orientaciones interdominio, así como en las interacciones biomoleculares de NS3, regulan su actividad NTPasa en condiciones fisiológicas.

4.6 Conclusiones

La estabilidad del sitio activo y la eficiencia de la hidrólisis de ATP en la helicasa NS3 están determinadas en gran medida por la presencia y la longitud del ARN, que modulan diferencialmente estos procesos. Un aspecto clave del mecanismo catalítico es la coordinación de Glu con Mg^{2+}, lo que resulta fundamental para garantizar su participación en la reacción. Los resultados obtenidos señalan la existencia de bolsillos crípticos y sitios alostéricos que representan objetivos prometedores para el desarrollo de nuevas estrategias terapéuticas dirigidas contra el virus del dengue y los flavivirus. El uso combinado de simulaciones moleculares y análisis estructural permite comprender en mayor profundidad los determinantes moleculares que gobiernan la actividad enzimática, proporcionando información valiosa para el diseño racional de inhibidores.

Bibliografía

1. Benarroch D, Selisko B, Locatelli GA, Maga G, Romette J-L, Canard B. The RNA helicase, nucleotide 5’-triphosphatase, and RNA 5'-triphosphatase activities of Dengue virus protein NS3 are Mg2+-dependent and require a functional Walker B motif in the helicase catalytic core. Virology. 2004;328: 208–218. doi:10.1016/j.virol.2004.07.004

2. Xu T, Sampath A, Chao A, Wen D, Nanao M, Chene P, et al. Structure of the Dengue virus helicase/nucleoside triphosphatase catalytic domain at a resolution of 2.4 A. J Virol. 2005;79: 10278–10288. doi:10.1128/JVI.79.16.10278-10288.2005

3. Du Pont KE, McCullagh M, Geiss BJ. Conserved motifs in the flavivirus NS3 RNA helicase enzyme. Wiley Interdiscip Rev RNA. 2022;13: e1688. doi:10.1002/wrna.1688

4. Sarto C, Kaufman SB, Estrin DA, Arrar M. Nucleotide-dependent dynamics of the Dengue NS3 helicase. Biochim Biophys Acta Proteins Proteom. 2020;1868: 140441. doi:10.1016/j.bbapap.2020.140441

5. Adler NS, Cababie LA, Sarto C, Cavasotto CN, Gebhard LG, Estrin DA, et al. Insights into the product release mechanism of dengue virus NS3 helicase. Nucleic Acids Res. 2022;50: 6968–6979. doi:10.1093/nar/gkac473

6. Amrein F, Sarto C, Cababie LA, Gonzalez Flecha FL, Kaufman SB, Arrar M. Impact of bound ssRNA length on allostery in the Dengue Virus NS3 helicase. Nucleic Acids Res. 2023;51: 11213–11224. doi:10.1093/nar/gkad808

7. Incicco JJ, Gebhard LG, González-Lebrero RM, Gamarnik AV, Kaufman SB. Steady-state NTPase activity of Dengue virus NS3: number of catalytic sites, nucleotide specificity and activation by ssRNA. PLoS One. 2013;8: e58508. doi:10.1371/journal.pone.0058508

8. Gebhard LG, Kaufman SB, Gamarnik AV. Novel ATP-independent RNA annealing activity of the dengue virus NS3 helicase. PLoS One. 2012;7: e36244. doi:10.1371/journal.pone.0036244

9. Luo D, Xu T, Watson RP, Scherer-Becker D, Sampath A, Jahnke W, et al. Insights into RNA unwinding and ATP hydrolysis by the flavivirus NS3 protein. EMBO J. 2008;27: 3209–3219. doi:10.1038/emboj.2008.232

10. Mugnai ML, Thirumalai D. Step-wise hydration of magnesium by four water molecules precedes phosphate release in a myosin motor. J Phys Chem B. 2021;125: 1107–1117. doi:10.1021/acs.jpcb.0c10004

11. Mugnai ML, Thirumalai D. Hydration of magnesium is required for myosin VI phosphate release. Biophys J. 2018;114: 318a. doi:10.1016/j.bpj.2017.11.1788

12. García-Martínez A, Zinovjev K, Ruiz-Pernía JJ, Tuñón I. Conformational changes and ATP hydrolysis in Zika helicase. The molecular basis of a biomolecular motor unveiled by multiscale simulations. ChemRxiv. 2023. doi:10.26434/chemrxiv-2023-5krqd

13. Kamerlin SCL, Warshel A. On the energetics of ATP hydrolysis in solution. J Phys Chem B. 2009;113: 15692–15698. doi:10.1021/jp907223t

14. Trentham DR, Eccleston JF, Bagshaw CR. Kinetic analysis of ATPase mechanisms. Q Rev Biophys. 1976;9: 217–281. doi:10.1017/s0033583500002419

15. Kiani FA, Fischer S. Stabilization of the ADP/metaphosphate intermediate during ATP

hydrolysis in pre-power stroke myosin: quantitative anatomy of an enzyme: Quantitative anatomy of an enzyme. J Biol Chem. 2013;288: 35569–35580. doi:10.1074/jbc.M113.500298

16. Akola J, Jones RO. ATP hydrolysis in water − A density functional study. J Phys Chem B. 2003;107: 11774–11783. doi:10.1021/jp035538g

17. Wang C, Huang W, Liao J-L. QM/MM investigation of ATP hydrolysis in aqueous solution. J Phys Chem B. 2015;119: 3720–3726. doi:10.1021/jp512960e

18. Duarte F, Barrozo A, Åqvist J, Williams NH, Kamerlin SCL. The competing mechanisms of phosphate monoester dianion hydrolysis. J Am Chem Soc. 2016;138: 10664–10673. doi:10.1021/jacs.6b06277

19. Carvalho ATP, Szeler K, Vavitsas K, Åqvist J, Kamerlin SCL. Modeling the mechanisms of biological GTP hydrolysis. Arch Biochem Biophys. 2015;582: 80–90. doi:10.1016/j.abb.2015.02.027

20. Movilla S, Roca M, Moliner V, Magistrato A. Molecular basis of RNA-driven ATP hydrolysis in DExH-box helicases. J Am Chem Soc. 2023;145: 6691–6701. doi:10.1021/jacs.2c11980

21. Spaine RF, Wang F, Foreman KW, Solomon LA. Ab initio exploration of the phosphoryl transfer reaction provides insights into interpreting models of SN2 mechanisms. J Phys Chem C Nanomater Interfaces. 2025;129: 9686–9698. doi:10.1021/acs.jpcc.5c00499

22. Prasad BR, Plotnikov NV, Warshel A. Addressing open questions about phosphate hydrolysis pathways by careful free energy mapping. J Phys Chem B. 2013;117: 153–163. doi:10.1021/jp309778n

23. Petrović D, Szeler K, Kamerlin SCL. Challenges and advances in the computational modeling of biological phosphate hydrolysis. Chem Commun (Camb). 2018;54: 3077–3089. doi:10.1039/c7cc09504j

24. Biological Phosphoryl transfer reactions: understanding mechanisms and catalysis.

25. Kiani FA, Fischer S. Comparing the catalytic strategy of ATP hydrolysis in biomolecular motors. Phys Chem Chem Phys. 2016;18: 20219–20233. doi:10.1039/c6cp01364c

26. Du Pont KE, Davidson RB, McCullagh M, Geiss BJ. Motif V regulates energy transduction between the flavivirus NS3 ATPase and RNA-binding cleft. J Biol Chem. 2020;295: 1551–1564. doi:10.1074/jbc.RA119.011922

27. Davidson RB, Hendrix J, Geiss BJ, McCullagh M. Allostery in the dengue virus NS3 helicase: Insights into the NTPase cycle from molecular simulations. PLoS Comput Biol. 2018;14: e1006103. doi:10.1371/journal.pcbi.1006103

28. Movilla S, Roca M, Moliner V, Magistrato A. Molecular basis of RNA-driven ATP hydrolysis in DExH-box helicases. J Am Chem Soc. 2023;145: 6691–6701. doi:10.1021/jacs.2c11980

29. Calixto AR, Moreira C, Pabis A, Kötting C, Gerwert K, Rudack T, et al. GTP hydrolysis without an active site base: A unifying mechanism for Ras and related GTPases. J Am Chem Soc. 2019;141: 10684–10701. doi:10.1021/jacs.9b03193

30. Knyazhanskaya E, Morais MC, Choi KH. Flavivirus enzymes and their inhibitors. Enzymes. 2021;49: 265–303. doi:10.1016/bs.enz.2021.07.006

31. Sampath A, Padmanabhan R. Molecular targets for flavivirus drug discovery. Antiviral Res. 2009;81: 6–15. doi:10.1016/j.antiviral.2008.08.004

32. Munawar A, Beelen S, Munawar A, Lescrinier E, Strelkov SV. Discovery of novel druggable sites on Zika virus NS3 helicase using X-ray crystallography-based fragment screening. Int J Mol Sci. 2018;19: 3664. doi:10.3390/ijms19113664

33. Cruz-Arreola O, Orduña-Diaz A, Domínguez F, Reyes-Leyva J, Vallejo-Ruiz V, Domínguez-Ramírez L, et al. In silico testing of flavonoids as potential inhibitors of protease and helicase domains of dengue and Zika viruses. PeerJ. 2022;10: e13650. doi:10.7717/peerj.13650

34. Waterhouse A, Bertoni M, Bienert S, Studer G, Tauriello G, Gumienny R, et al. SWISS-MODEL: homology modelling of protein structures and complexes. Nucleic Acids Res. 2018;46: W296–W303. doi:10.1093/nar/gky427

35. Olsson MHM, Søndergaard CR, Rostkowski M, Jensen JH. PROPKA3: Consistent treatment of internal and surface residues in empirical pKa predictions. J Chem Theory Comput. 2011;7: 525–537. doi:10.1021/ct100578z

36. Case DA, Aktulga HM, Belfon K, Cerutti DS, Cisneros GA, Cruzeiro VWD, et al. AmberTools. J Chem Inf Model. 2023;63: 6183–6191. doi:10.1021/acs.jcim.3c01153

37. Maier JA, Martinez C, Kasavajhala K, Wickstrom L, Hauser KE, Simmerling C. Ff14SB: Improving the accuracy of protein side chain and backbone parameters from ff99SB. J Chem Theory Comput. 2015;11: 3696–3713. doi:10.1021/acs.jctc.5b00255

38. Aytenfisu AH, Spasic A, Grossfield A, Stern HA, Mathews DH. Revised RNA dihedral parameters for the Amber force field improve RNA molecular dynamics. J Chem Theory Comput. 2017;13: 900–915. doi:10.1021/acs.jctc.6b00870

39. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML. Comparison of simple potential functions for simulating liquid water. J Chem Phys. 1983;79: 926–935. doi:10.1063/1.445869

40. Lee C, Yang W, Parr RG. Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density. Phys Rev B Condens Matter. 1988;37: 785–789. doi:10.1103/physrevb.37.785

41. Becke AD. Density-functional thermochemistry. III. The role of exact exchange. J Chem Phys. 1993;98: 5648–5652. doi:10.1063/1.464913

42. Gaus M, Lu X, Elstner M, Cui Q. Parameterization of DFTB3/3OB for sulfur and phosphorus for chemical and biological applications. J Chem Theory Comput. 2014;10: 1518–1537. doi:10.1021/ct401002w

43. Lu X, Gaus M, Elstner M, Cui Q. Parametrization of DFTB3/3OB for magnesium and zinc for chemical and biological applications. J Phys Chem B. 2015;119: 1062–1082. doi:10.1021/jp506557r

44. Gaus M, Goez A, Elstner M. Parametrization and benchmark of DFTB3 for organic molecules. J Chem Theory Comput. 2013;9: 338–354. doi:10.1021/ct300849w

45. Liu DC, Nocedal J. On the limited memory BFGS method for large scale optimization. Math Program. 1989;45: 503–528. doi:10.1007/bf01589116

46. Liu J, Li D, Liu X. A simple and accurate algorithm for path integral molecular dynamics

with the Langevin thermostat. J Chem Phys. 2016;145: 024103. doi:10.1063/1.4954990

47. Berendsen HJC, Postma JPM, van Gunsteren WF, DiNola A, Haak JR. Molecular dynamics with coupling to an external bath. J Chem Phys. 1984;81: 3684–3690. doi:10.1063/1.448118

48. Ásgeirsson V, Birgisson BO, Bjornsson R, Becker U, Neese F, Riplinger C, et al. Nudged elastic band method for molecular reactions using energy-weighted springs combined with eigenvector following. J Chem Theory Comput. 2021;17: 4929–4945. doi:10.1021/acs.jctc.1c00462

49. Neese F. Software update: The ORCA program system—Version 5.0. Wiley Interdiscip Rev Comput Mol Sci. 2022;12. doi:10.1002/wcms.1606

50. Humphrey W, Dalke A, Schulten K. VMD: visual molecular dynamics. J Mol Graph. 1996;14: 33–8, 27–8. doi:10.1016/0263-7855(96)00018-5

51. Nudged Elastic Band Method for Molecular Reactions Using EnergyWeighted Springs Combined with Eigenvector Following.

52. Zinovjev K, Tuñón I. Adaptive finite temperature string method in collective variables. J Phys Chem A. 2017;121: 9764–9772. doi:10.1021/acs.jpca.7b10842

53. Ryckaert J-P, Ciccotti G, Berendsen HJC. Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. J Comput Phys. 1977;23: 327–341. doi:10.1016/0021-9991(77)90098-5

54. Kumar S, Rosenberg JM, Bouzida D, Swendsen RH, Kollman PA. THE weighted histogram analysis method for free-energy calculations on biomolecules. I. The method. J Comput Chem. 1992;13: 1011–1021. doi:10.1002/jcc.540130812

55. Roe DR, Cheatham TE 3rd. PTRAJ and CPPTRAJ: Software for processing and analysis of molecular dynamics trajectory data. J Chem Theory Comput. 2013;9: 3084–3095. doi:10.1021/ct400341p

56. Lange OF, Grubmüller H. Generalized correlation for biomolecular dynamics. Proteins. 2006;62: 1053–1061. doi:10.1002/prot.20784

57. Patel AC, Sinha S, Palermo G. Graph theory approaches for molecular dynamics simulations. Q Rev Biophys. 2024;57: e15. doi:10.1017/S0033583524000143

58. Melo MCR, Bernardi RC, de la Fuente-Nunez C, Luthey-Schulten Z. Generalized correlation-based dynamical network analysis: a new high-performance approach for identifying allosteric communications in molecular dynamics trajectories. bioRxiv. bioRxiv; 2020. doi:10.1101/2020.06.18.160572

59. Arcon JP, Defelipe LA, Modenutti CP, López ED, Alvarez-Garcia D, Barril X, et al. Molecular dynamics in mixed solvents reveals protein-ligand interactions, improves docking, and allows accurate binding free energy predictions. J Chem Inf Model. 2017;57: 846–863. doi:10.1021/acs.jcim.6b00678

60. López ED, Arcon JP, Gauto DF, Petruk AA, Modenutti CP, Dumas VG, et al. WATCLUST: a tool for improving the design of drugs based on protein-water interactions. Bioinformatics. 2015;31: 3697–3699. doi:10.1093/bioinformatics/btv411

61. Burastero O, Defelipe LA, Gola G, Tateosian NL, Lopez ED, Martinena CB, et al. Cosolvent sites-based discovery of Mycobacterium tuberculosis protein kinase G

inhibitors. J Med Chem. 2022;65: 9691–9705. doi:10.1021/acs.jmedchem.1c02012

62. Waterhouse AM, Procter JB, Martin DMA, Clamp M, Barton GJ. Jalview Version 2--a multiple sequence alignment editor and analysis workbench. Bioinformatics. 2009;25: 1189–1191. doi:10.1093/bioinformatics/btp033

63. Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, et al. Clustal W and Clustal X version 2.0. Bioinformatics. 2007;23: 2947–2948. doi:10.1093/bioinformatics/btm404

64. He Y, Staley JP, Andersen GR, Nielsen KH. Structure of the DEAH/RHA ATPase Prp43p bound to RNA implicates a pair of hairpins and motif Va in translocation along RNA. RNA. 2017;23: 1110–1124. doi:10.1261/rna.060954.117

65. Luo D, Xu T, Watson RP, Scherer-Becker D, Sampath A, Jahnke W, et al. Insights into RNA unwinding and ATP hydrolysis by the flavivirus NS3 protein. EMBO J. 2008;27: 3209–3219. doi:10.1038/emboj.2008.232

66. Schneider S, Campodonico E, Schwer B. Motifs IV and V in the DEAH box splicing factor Prp22 are important for RNA unwinding, and helicase-defective Prp22 mutants are suppressed by Prp8. J Biol Chem. 2004;279: 8617–8626. doi:10.1074/jbc.M312715200

67. Sampath A, Xu T, Chao A, Luo D, Lescar J, Vasudevan SG. Structure-based mutational analysis of the NS3 helicase from dengue virus. J Virol. 2006;80: 6686–6690. doi:10.1128/jvi.02215-05

68. Dillingham MS, Soultanas P, Wigley DB. Site-directed mutagenesis of motif III in PcrA helicase reveals a role in coupling ATP hydrolysis to strand separation. Nucleic acids research. 1999;27. doi:10.1093/nar/27.16.3310

69. Swarbrick CMD, Basavannacharya C, Chan KWK, Chan S-A, Singh D, Wei N, et al. NS3 helicase from dengue virus specifically recognizes viral RNA sequence to ensure optimal replication. Nucleic Acids Res. 2017;45: 12904–12920. doi:10.1093/nar/gkx1127

70. Flexibility between the Protease and Helicase Domains of the Dengue Virus NS3 Protein Conferred by the Linker Region and Its Functional Implications. Journal of Biological Chemistry. 2010;285: 18817–18827. doi:10.1074/jbc.M109.090936

71. Osawa T, Aoki M, Ehara H, Sekine S-I. Structures of dengue virus RNA replicase complexes. Mol Cell. 2023;83: 2781–2791.e4. doi:10.1016/j.molcel.2023.06.023

72. wwPDB: 4WXR. [cited 13 May 2025]. Available: https://doi.org/10.2210/pdb4WXR/pdb

Capítulo 5: Activación alostérica de la polimerasa del virus del dengue

5.1 Resumen

La proteína no estructural 5 (NS5) es la proteína más grande y altamente conservada del género *flavivirus*, y desempeña un papel central en la replicación viral. La NS5 es una enzima formada por dos dominios, uno con actividad metiltransferasa (NS5met) en el extremo N-terminal, unido covalentemente a un dominio polimerasa de ARN dependiente de ARN (NS5pol) en el extremo C-terminal. Una función clave del dominio NS5pol es la iniciación *de novo* de la síntesis de ARN, la cual requiere el reconocimiento del promotor de ARN *Stem Loop A* (SLA), un elemento estructural conservado en la región no codificante del genoma viral. En este capítulo, se combinan simulaciones de dinámica molecular y análisis de redes para proporcionar una visión a nivel atómico del reconocimiento de SLA y de la posterior activación alostérica del dominio NS5pol. Nuestros resultados revelan una vía de comunicación que conecta la interfaz de unión al SLA con el sitio activo de NS5pol a través del *priming loop*, una región conservada e indispensable para la actividad polimerasa. Mostramos cómo esta vía es modulada adicionalmente por la secuencia específica del *top loop* del SLA y los efectos de la mutación reportada en el motivo F1 de NS5. Estos resultados ofrecen nuevas perspectivas mecanísticas sobre la replicación de los flavivirus y podrían contribuir al desarrollo de vacunas que se basan en el uso de virus atenuados o estrategias antivirales dirigidas contra NS5.

Este trabajo se realizó en su mayor parte durante una pasantía doctoral que tuvo lugar en el grupo de la Dra. Giulia Palermo (University of California Riverside, EEUU), donde se desarrolló la metodología de análisis de redes empleada. El capítulo se enmarca en una colaboración con el grupo de virología molecular de la Dra. Andrea Gamarnik (Fundación Instituto Leloir, Buenos Aires), donde se están llevando a cabo actualmente experimentos para evaluar las hipótesis que surgen del análisis computacional. Los resultados de este trabajo están siendo incorporados en un manuscrito del que soy primera autora y será enviado a una revista con revisión por pares para su publicación.

5.2 Introducción

La proteína no-estructural 5 (NS5, por sus siglas en inglés) del virus del dengue (DENV) es considerada un blanco atractivo para la búsqueda de antivirales, al igual que la proteína NS3 estudiada en los Capítulos 3 y 4. La NS5 es la proteína viral más grande, formada por aproximadamente 900 aminoácidos, y contiene dos dominios con actividad catalítica: uno ubicado en el extremo N-terminal con actividad de metiltransferasa (NS5met), crucial para la formación del *cap* tipo I ($^{m7}G_{ppp}A_{m2'\text{-}O}$) en el extremo 5' del ARN viral y el otro, responsable de la actividad ARN polimerasa dependiente de ARN (NS5pol) en el extremo C-terminal (Figura 5.1) [1]. Dada la relevancia de sus funciones, esta proteína representa un componente clave en la replicación viral y un blanco terapéutico atractivo [2]. Aunque los dos dominios de la NS5 realizan funciones *a priori* no vinculadas, están conectados covalentemente por un segmento flexible o *linker* de aproximadamente 10 aminoácidos que mantiene su proximidad espacial. Se han reportado estructuras de la NS5 que evidencian la existencia de al menos dos

conformaciones dadas por la distancia entre los dominios, una cerrada y otra abierta [3–6]. Además, se ha reportado la capacidad de esta proteína para formar dímeros [7]. Si bien la presencia de uno de estos dominios afecta alostéricamente la actividad del otro, activándolo [8,9], las actividades están conservadas aún cuando se estudia *in vitro* cada dominio por separado.

El sitio catalítico de la NS5met está compuesto por cuatro aminoácidos conservados (K61, D146, K180 y E216) que son capaces tanto de generar metilaciones como de transferir el nucleótido G necesario para la formación del *cap* tipo I en el extremo 5' del ARN viral. Por su parte, la NS5pol contiene tres subdominios conservados en polimerasas de ARN que se denominan haciendo alusión a la distribución de una mano derecha. Estos subdominios son: NS5pol-dedos, NS5pol-palma y NS5pol-pulgar. El primero se conecta con NS5met a través del *linker* y contiene los motivos conservados F y G, dos regiones flexibles que interactúan con el ARN molde, guiándolo hacia el sitio catalítico; el segundo, contiene los motivos conservados A, B, C y D que forman el sitio catalítico; y el tercero contiene los motivos E, indispensable para la unión de dos cationes Zn^{2+}, además del *priming loop [1]*. Estos cationes no se han reportado como relacionados con la actividad catalítica de la NS5pol, sino que su función es aportar estabilidad estructural. La NS5pol es capaz de comenzar la síntesis ARN sin la necesidad de una secuencia corta complementaria al extremo 3' (conocida como secuencia *primer*). Esto se conoce como síntesis *de novo*, en contraposición a los demás pasos de síntesis de ARN que consisten en elongar la cadena naciente de oligonucleótidos según la cadena molde. Para poder llevar a cabo la síntesis *de novo*, la NS5pol contiene una región conocida como *priming loop* localizado en el NS5pol-pulgar que tiene como función estabilizar la cadena molde al entrar al sitio catalítico durante el comienzo de la síntesis de ARN, imitando la función de un *primer*.

En el extremo 5' no codificante del ARN viral, se encuentra una secuencia de 70 nucleótidos de ARN estructurado conocido como *Stem Loop A* (SLA), cuya función es permitir que la NS5 reconozca al ARN viral mediante su interacción, discriminando otros ARN celulares, y actúa además como "promotor" del inicio de la síntesis de ARN [10–12].

El SLA tiene estructura de Y, en donde se distinguen tres regiones: los extremos 3' y 5' se aparean formando el tallo inferior o *bottom stem*, con la excepción de un nucleótido U que se destaca por quedar desapareado (U *bulge*); el tallo superior o *top stem* que termina en una secuencia de entre 4 y 5 nucleótidos conocida como *top loop* (DENV1,3,4: C30, A31, G32, U33; DENV2: C30, A31, G32, A33, U34); y el tallo-*loop* lateral o *side stem-loop*. La región central en la cual se encuentran estas tres estructuras se conoce como "junta triple" o *three-way junction*. Si bien el SLA se encuentra en el extremo 5' del ARN viral, la NS5pol comienza la síntesis en el extremo 3' y avanza en dirección hacia el 5'. En regiones del ARN viral no codificantes, existen secuencias específicas y complementarias en ambos extremos, cuyo apareamiento ha mostrado ser indispensable para la replicación viral [13,14], y que al generar un ARN circular permiten el acercamiento del extremo 3' hacia la NS5pol para el comienzo de la síntesis.

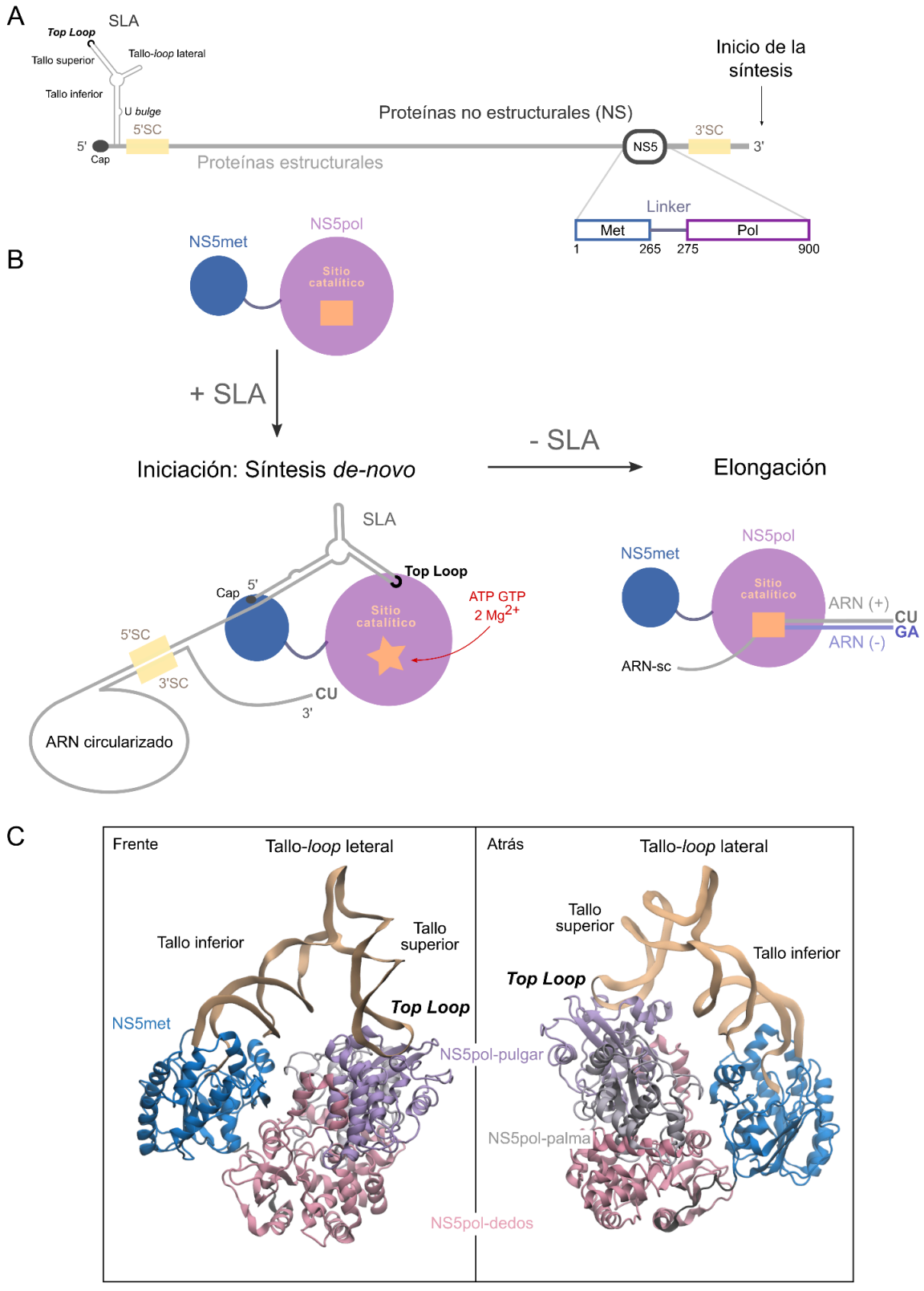


Figura 5.1. A) Representación del ARN viral del virus del dengue. En el extremo 5' se muestra el cap tipo I con un círculo negro y a continuación, la estructura secundaria del *Stem Loop A* (SLA) con las regiones etiquetadas, tallo inferior y superior, tallo-*loop* lateral y el *top loop* resaltado en negro. Las secuencias de circularización (CS) de los extremos 3' y 5' se muestran en amarillo. Región codificante que representa los nucleótidos que al ser traducidos formarán las proteínas estructurales y no

estructurales. Detalle de las regiones del proteína no estructural 5 (NS5), mostrando el dominio N-terminal con actividad de metiltransferasa (NS5met, azul), el dominio C-terminal con actividad polimerasa (NS5pol, violeta) y el linker (gris) que une ambos dominios. Se detallan además los residuos de cada región. **B)** Representación del inicio de la replicación viral. Se representa la NS5 según los colores del panel A y se muestra el sitio catalítico con un rectángulo naranja. Para iniciar la síntesis, el ARN viral se circulariza por las CS complementarias que permiten acercar el extremo 3' finalizado en las bases CU (comienzo de la síntesis) a la NS5 mientras se encuentra unida al SLA (NS5met se une en el extremo 5' y el tallo inferior del SLA y NS5pol, a través del top loop destacado en negro), lo que a su vez permite que el sitio catalítico se active (estrella naranja). Una flecha roja indica que también es necesaria la presencia de Mg^{2+}, ATP y GTP para la iniciación o síntesis *de novo.* Una vez iniciada la síntesis, esta puede continuar (elongación) en ausencia de SLA a partir de una cadena molde de ARN simple cadena (ARN-sc) **C)** Estructura resuelta por crío-microscopía electrónica del complejo NS5:SLA (PDB ID 8GZP). El SLA se muestra en naranja claro, el dominio NS5met está mostrado en azul y se distinguen los subdominios de la NS5pol: NS5pol-dedos en rosa, -palma en gris y -pulgar en violeta. Se señalan también las regiones de estructura secundaria del SLA. La estructura de la derecha se encuentra rotada 180° sobre el eje vertical con respecto a la de la izquierda.

En la literatura se han reportado varias mutaciones en el sistema SLA:NS5 que permiten conocer algunos detalles de las regiones específicas involucradas en la modulación que ejerce el SLA sobre la actividad polimerasa de la NS5pol. De estas, se destacan dos tipos de mutaciones: sobre el *top loop* del ARN viral y sobre la NS5pol-dedos. Por un lado, se reportó que al mutar dos bases del *top loop* en DENV2 (A31G, G32A) del SLA se anula completamente la actividad polimerasa y la replicación viral. Luego, es posible experimentalmente llevar a cabo ensayos para permitir que las mutantes encuentren variantes, a través de mutaciones espontáneas, que recuperen la replicación, para finalmente secuenciar los genomas y obtener lo que se conoce como secuencias revertantes. En este caso, la revertante encontrada presentó un cambio de base (revirtiendo la posición 32 y quedando únicamente mutado en A31G) mediante la cual se recupera el rol del SLA en la replicación [11]. Por otro lado, en un estudio de mutaciones sobre la proteína NS5, la mutación de dos aminoácidos sobre la región denominada F1 (K456A y R457A) adyacente al motivo conservado F de NS5pol) anula la síntesis *de novo* dependiente de SLA pero conserva la actividad polimerasa para elongar un primer unido al ARN molde [15]. Sorprendentemente, ambas mutantes comparten el hecho de conservar la afinidad del SLA por la NS5, a la vez que afectan las capacidades catalíticas de la NS5pol mediada por SLA [15,16]. Estas observaciones muestran que la unión del SLA a la NS5 no alcanza para activar la actividad polimerasa y por lo tanto, resaltan la necesidad de investigar la comunicación alostérica entre el sitio de unión al SLA y el sitio catalítico para entender cómo ocurre la activación de la polimerasa mediada por el SLA.

Recientemente, se han reportado dos estructuras del complejo NS5:SLA, una de NS5-DENV3:SLA-DENV4 y otra de (NS5:SLA)-DENV2, obtenidas por crío-microscopía electrónica [17,18]. Si bien ya se sospechaba cuales eran las regiones de contacto del SLA con la NS5, se verificó con estas estructuras que el extremo 5' y el comienzo del tallo inferior interaccionan con el dominio NS5met y el *top loop* interacciona con NS5pol, específicamente a través de tres nucleótidos: C30, A31, G32. Como era esperable, la unión del SLA a la NS5pol se da en la superficie de la proteína, específicamente en el subdominio NS5pol-pulgar, y el sitio catalítico se encuentra enterrado y en el subdominio NS5pol-palma, por lo tanto y como se ha mencionado anteriormente, la activación de la NS5pol por la unión de SLA es un fenómeno alostérico. Esto significa que la unión del SLA a la NS5 debe generar

cambios estructurales y dinámicos en el sitio de unión que son propagados al sitio catalítico, de manera tal que permitan el inicio de la síntesis de ARN.

En este capítulo, se estudian los sistemas NS5 y NS5:SLA *wild-type* (WT) y las mutantes mencionadas ($NS5^{F1mut}$, $NS5^{F1mut}$:SLA, NS5:SLA^{mut}, NS5:SLA^{rev}) mediante simulaciones de dinámica molecular con el objetivo de identificar las vías de comunicación dentro del dominio NS5pol que conectan el sitio de unión de SLA con el sitio catalítico. Para esto se utilizaron herramientas de análisis derivadas de la teoría de grafos y análisis de redes dinámicas. No solo se encontraron diferencias en cómo se pueblan las conformaciones abierta y cerrada de la NS5 en ausencia de SLA como resultado de las mutaciones sobre la proteína, sino que también estas poblaciones parecen ser afectadas por por la presencia de un *top loop* mutado e incapaz de replicar. Se observó también que para los sistemas *wild-type,* las correlaciones generalizadas de la NS5pol disminuyen al unirse el SLA, mientras que para las mutantes sobre la proteína ocurre lo contrario. Además se caracterizaron las interacciones entre el *top loop* del SLA y el subdominio NS5pol-pulgar, destacándose la posición 32 como principal pérdida en las interacciones, racionalizando la recuperación a través de la reverente y destacándose los aminoácidos R770 y R888 como contrapartes en la interacción sobre la proteína. Por último, un análisis de caminos óptimos basado en las correlaciones generalizadas mostró la importancia de que la comunicación entre el sitio de unión del SLA en la NS5pol y el sitio catalítico involucre la participación del *priming loop.* Se concluye entonces que la aparición de caminos alternativos impedirían la activación de la polimerasa por la presencia del SLA y se proponen aminoácidos que de ser mutados, impedirían esta comunicación y consecuente activación, proporcionando así información valiosa para el diseño racional de estrategias antivirales.

5.3 Métodos

5.3.1 Preparación del sistema

Todos los sistemas analizados fueron obtenidos a partir de la estructura de CryoEM con PDB ID 8GZP [17], que corresponde al complejo NS5-DENV3:SLA-DENV4. Para los sistemas en ausencia de SLA, se borraron los átomos correspondientes a estos nucleótidos. La proteína fue resuelta entre las posiciones 7 y 890 (de un total de 896 aminoácidos) y fueron considerados para este análisis los aminoácidos contenidos entre estos extremos. Las regiones internas que no fueron resueltas debidas a su alta flexibilidad se modelaron por homología usando el programa SWISS-MODEL [19] a partir de la secuencia de NS5-DENV3 reportada en la estructura. De manera similar, en el SLA (nucleótidos 1-69), dos nucleótidos del tallo-*loop* lateral no fueron resueltos y su estructura se completó utilizando otro programa de modelado por homología, específico para ARN, ModeRNA [20]. Si bien la estructura fue resuelta en presencia del *cap* 5', la guanosina difosfato junto con un catión Mg^{2+} fueron eliminados, dada la evidencia experimental de que su presencia no afecta la interacción con la NS5 [15]. Los dos cationes Zn^{2+} resueltos también fueron incluidos en la estructura.

La proteína y el ARN fueron modelados utilizando el campo de fuerza Amber ff14SB [21] con la corrección ILDN para residuos proteicos y parámetros diedros recientemente revisados para ARN [22]. Los parámetros para los cuatro aminoácidos (C778, C847, H712 y H714) de uno de los entornos tetraédricos de los cationes Zn^{2+} fueron obtenidos directamente a partir del Zinc AMBER force field (ZAFF) [23]. Los parámetros de los otros cuatro aminoácidos del

otro Zn^{2+} (C446, C449, H441 y E437), fueron obtenidos a partir del método *Metal Center Parameter Builder* escrito en lenguaje de programación Python: MCPB.py [24,25]. Los parámetros de este tipo de campo de fuerzas funcionan como enlaces similares a los covalentes entre el metal y los 4 aminoácidos del entorno. Para obtener estos parámetros en primer lugar se calculan las distancias o ángulos de equilibrio entre 2 y 3 átomos, respectivamente, para el sistema en fase gaseosa mediante métodos *ab initio*; en segundo lugar, se obtienen las constantes de fuerza, a partir de la minimización del sistema total aproximado sin el metal (por ejemplo, tiol (SH) por iolato (SO^-) si hay cisteínas que coordinan al metal) mediante usando DFT B3LYP/6-31G* (Ver Capítulo 2: Métodos); finalmente, las cargas puntuales son asignadas como cargas parciales centradas en los átomos mediante el método RESP (Ver Capítulo 2: Métodos) implementado en el programa Gaussian 16 [26].

Los estados de protonación de los residuos ionizables se determinaron en función de los valores de pKa estimados para las cadenas laterales mediante propka 3.0 [27], un predictor empírico de pKa para residuos superficiales y enterrados en proteínas. Luego de mirar el entorno de cada histidina, se establecieron los tautómeros, resultando como HID las posiciones H215, H263, H441, H512, H712, H714, H768 y el resto como HIE. Para generar las mutantes, se modificaron los aminoácidos o nucleótidos, según cada caso, evitando solapamientos y se corroboró que las estructuras resultantes relajaron su energía luego de la minimización sin presentar conformaciones anómalas (ver minimización y equilibración). La mayoría de los sistemas fueron solvatados con el modelo de agua TIP3P [28] en cajas octaédricas con una distancia de 8 Å desde la superficie de la proteína. Como algunos sistemas presentaron mayor flexibilidad y se corría el riesgo de generar efectos de borde, para estos, se amplió la caja a 15 Å. Se empleó una concentración salina fisiológica de 0,15 M. El detalle de cada sistema junto con su carga, la cantidad de aguas TIP3P e iones se encuentra en la Tabla 5.1. Las aguas TIP3P, los hidrógenos y otros átomos no resueltos en las estructuras cristalográficas fueron añadidos utilizando el módulo tLeAP del paquete Amber [29].

Tabla 5.1. Detalles de los sistemas estudiados

Sistemas	Mutaciones	Carga (*e*)	Tamaño de la caja a partir de la superficie de la proteína o ARN	Cantidad aproximada de aguas	Cantidad de iones
NS5	*wild-type*	+16	8 Å	24.500	67 Na+ 83 Cl-
$NS5^{F1mut}$	K456A y R457A	+14	8 Å	24.500	67 Na+ 81 Cl-
NS5:SLA	*wild-type*	-53	15 Å	45.000	176 Na+ 123 Cl-
$NS5:SLA^{mut}$	A31G, G32A	-53	15 Å	45.000	176 Na+ 123 Cl-
$NS5:SLA^{rev}$	A31G	-53	8 Å	31.500	139 Na+ 86 Cl-

$NS5^{F1mut}$:SLA	K456A y R457A	-55	8 Å	31.500	137 Na+ 86 Cl-

5.3.2 Simulaciones de dinámica molecular

Para cada sistema, se realizaron simulaciones por triplicado con el módulo PMEMD Cuda del paquete del programa Amber [30]. Cada réplica incluyó las tres etapas previas a la dinámica de producción (minimización, calentamiento y equilibración, ver Capítulo 2) que se describen a continuación.

Primero se minimizaron las moléculas de agua de los iones, aplicando restricciones (constante de fuerza de 300 kcal/mol/$Å^2$) sobre los átomos de las biomoléculas (proteína y en caso de que hubiera, ARN). Para esto, se realizaron 500 pasos de minimización con el algoritmo *steepest descent* y luego 500 pasos más con el algoritmo *conjugate gradient*. El proceso de minimización continuó eliminando las restricciones y realizando 500 pasos más de *steepest descent* y por último, 2500 pasos más con el algoritmo *conjugate gradient*. Una vez que el sistema se relajó con la minimización de energía, el objetivo siguiente fue llevar al sistema a 300 K y 1 bar. En un primer paso de calentamiento, se aumentó la temperatura de 0 a 100 K imponiendo fuertes restricciones en forma de potencial armónico con una constante de fuerza de 100 kcal/mol/$Å^2$ sobre los átomos de las biomoléculas. De esta manera, se permite el movimiento y se relaja únicamente la caja de solvente junto con los iones durante 5 ps. Análogamente, en un segundo paso se redujo las restricciones sobre la proteína a 25 kcal/mol/$Å^2$, para aumentar la temperatura de 100 K a 200 K, esta vez durante 10 ps. Por último, se aumentó la temperatura de 200 K a 300 K reduciendo nuevamente la constante de fuerza de las restricciones a un valor de 10 kcal/mol/$Å^2$ para los mismos átomos durante 100 ps. Estos tres pasos de calentamiento se llevaron a cabo a volumen constante. Finalmente, se realizó una equilibración en dos pasos, un primer paso con restricciones (constante de fuerza de 10 kcal/mol/$Å^2$) y un segundo paso libre de movimiento. En primer lugar, una vez que el sistema fue llevado a 300 K, se equilibró el sistema durante 100 ps a temperatura y presión constante de 1 bar. En segundo lugar, el sistema se equilibró durante 10.000 ps a 1 bar y 300 K. En todos los pasos de calentamiento, el paso de tiempo utilizado fue de 1 fs y en el de equilibración fue de 2 fs.

Tanto en las etapas de calentamiento y equilibración como para la producción de la trayectoria, se controló la temperatura del sistema mediante con el termostato de Langevin usando una frecuencia de colisión de 1 ps^{-1} [31]. Además, el segundo paso de equilibración a presión constante, se llevó a cabo con el barostato de Berendsen y un tiempo de relajación de 2 ps [32]. Para disminuir el costo computacional, se empleó el algoritmo SHAKE que restringe los enlaces con hidrógenos no polares y permite usar un paso de tiempo mayor (1-2 fs) [33] y como límite en la consideración de las interacciones de no unión, se estableció una distancia de corte de 10 Å, lo que significa que no se contarán las interacciones entre dos partículas que superen esa distancia.

Para cada sistema, se realizaron tres réplicas independientes de 2 µs de dinámica molecular partiendo de la estructura obtenida luego de la equilibración y en las mismas condiciones de NVT mencionadas anteriormente, utilizando un paso de tiempo de 2 fs. De esta producción,

se eliminaron los primeros 0,5 μs, ya que fueron tomados como un paso extra de equilibración.

5.3.3 Análisis de las trayectorias

A partir de las imágenes de las trayectorias de cada sistema estudiado se aplicaron las técnicas de análisis, muchas de ellas basadas en teoría de grafos y análisis de redes que se mencionan a continuación y cuyos detalles pueden ser encontrados en el Capítulo 2.

Para todos los análisis, cada trayectoria fue alineada con respecto a los aminoácidos que del dominio NS5pol (posiciones 267-890 de la NS5). Se calcularon las distancias entre los sitios catalíticos de la NS5met (K61, D146, K180 y E216) y de la NS5pol (G656, D657 y D658) a partir de la distancia entre los centros de masa de los aminoácidos que los componen mediante el programa CPPTRAJ [34].

Dado el interés en estudiar el efecto de la unión del SLA sobre el sitio catalítico de la NS5pol, se representó únicamente a los aminoácidos de la NS5pol como un grafo o red dinámica, estableciendo la posición de cada nodo sobre el átomo Cα de cada aminoácido. Luego se calculó la matriz de correlaciones generalizadas [35,36] y se contrastó con los valores de correlaciones lineales convencionales. También se estudiaron las interacciones entre el SLA y la NS5pol, específicamente en la región de NS5pol-pulgar para determinar los aminoácidos involucrados en la interacción en cada uno de los estados simulados mediante el uso del módulo “hbond” implementado en el programa CPPTRAJ y el cálculo de distancias [34].

En base a resultados de las interacciones durante la dinámica molecular y luego de calcular la matriz de distancias ponderadas filtrada por contactos (considerando como adyacentes a los aminoácidos que tienen sus átomos pesados a menos de 4.5 Å durante al menos el 75% de la simulación), se identificaron los caminos óptimos y los 50 subóptimos que comunican el sitio de unión a SLA con el sitio catalítico (G656, D657, D658 y D541) de NS5pol. Se extrajo, además, el porcentaje de ocurrencia de los aminoácidos contenidos dichos caminos. A continuación, se dividió el análisis en tres tipos de caminos: cortos, intermedios y largos; y, junto con la medida de intermediación de aristas, se estimó la relación señal/ruido (S/R) de la suma de la intermediación de aristas a lo largo de estos caminos. A partir de un ajuste gaussiano de las distribuciones de sumas de intermediación de aristas, se obtiene un valor para la esperanza y la varianza de la “señal” y para el “ruido”, a partir de los cuales se puede calcular el cociente S/R [37].

5.4 Resultados

5.4.1 Mecanismo de selección conformacional de la NS5 para la unión del SLA

Al inspeccionar visualmente las trayectorias de los dos sistemas *apo* (NS5 y $NS5^{F1mut}$), se observó un gran cambio conformacional. Los dos dominios de la NS5 se encuentran unidos covalentemente a través de un linker flexible que les permite adoptar diferentes conformaciones evidenciadas en cristales reportados y etiquetadas como “abierta” y “cerrada” (Figura 5.2A). Para este estudio, la estructura inicial en las simulaciones de la NS5 fue tomada del complejo NS5:SLA, por lo que haber observado este tipo de comportamientos

sugiere un amplio muestreo del espacio conformacional. En contraposición, este cambio conformacional no fue observado a partir de las dinámicas de los complejos de la proteína NS5 unida al SLA (NS5:SLA y $NS5^{F1mut}$:SLA, NS5:SLA^{mut}, NS5:SLA^{rev}), sugiriendo que la proteína permanece en la conformación cerrada que es la conformación cristalográfica para todos los casos analizados, independientemente de si tiene mutaciones o no.

En base a la observación anterior, se calculó la distancia entre los sitios catalíticos de la NS5met y la NS5pol, a partir de los centros de masa de aminoácidos conservados (NS5met: K61, D146, K180, E216 y NS5pol: G656, D657 y D658) para cada uno de los sistemas estudiados y se incluyó el mismo cálculo para las siguientes estructuras reportadas: NS5-DV2 en conformación abierta (PDB ID 5ZQK), NS5-DV3 en conformación cerrada (PDB ID 5DTO), el complejo de elongación, NS5-DV3:NS3:ARN (PDB ID 8GZR), y la estructura inicial de este trabajo que corresponde al complejo de iniciación NS5-DV3:SLA-DV4 (PDB ID 8GZP) (Fig 2B).

Además de la estabilización en la conformación cerrada de la NS5 en presencia del SLA, los resultados de este análisis evidencian un cambio en las poblaciones de las dos conformaciones posibles dadas por las mutaciones estudiadas. En el sistema *apo*, la mutante F1 permanece más tiempo en conformaciones intermedias y menos tiempo en la conformación cerrada que la proteína WT, que parece poblar claramente una u otra conformación con preferencia de la conformación cerrada. Notablemente, para la proteína WT y para ambas conformaciones, las distribuciones exploran intervalos de distancia más amplios, no quedando centradas en los valores cristalográficos.

Si bien la presencia del SLA restringe las conformaciones de la NS5 hacia valores consistentes con la conformación cerrada, también se observan diferencias debidas a las mutaciones, a excepción de la mutante F1 que no parece tener un efecto significativo en comparación con el sistema WT. Dentro de las mutaciones introducidas en el SLA, la variante no replicativa NS5:SLA^{mut} se aleja contundentemente hacia valores más pequeños, acercando los sitios catalíticos en una conformación que si bien puede considerarse como cerrada, sugiere ser diferente a la WT. En cambio, al observar el comportamiento de la distribución de distancias para la variante replicativa NS5:SLA^{rev}, esta recupera la similitud con el sistema WT con un máximo centrado en el valor de referencia calculado a partir de la estructura inicial.

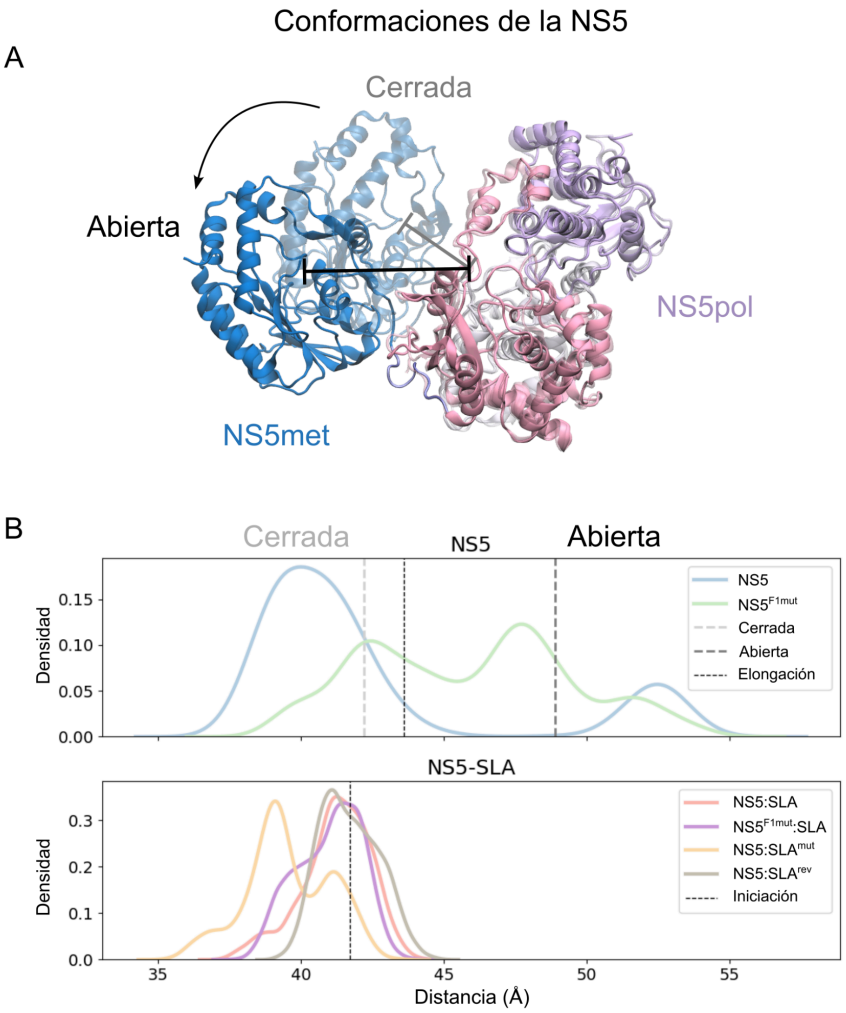


Figura 5.2. A) Imágenes superpuestas de la conformación abierta y cerrada de la NS5 tomadas de la dinámica del sistema NS5 WT. Un segmento negro simboliza la distancia entre los dos sitios cataíticos de la NS5 en la conformación abierta y uno gris para la misma distancia en la conformación cerrada. La fecha curva indica el cambio conformacional de cerrada a abierta observado. La proteína está coloreada según la Figura 5.1C. **B)** Gráficos de estimación de la densidad suavizada *(kernel density estimation)* de los valores de distancia distancia entre los sitios catalíticos de la metiltransferasa (MTase) y la polimerasa (distancia entre el centro de masa de los residuos K61, D146, K180, E216 y los residuos G662, D663, D664) obtenidos a partir de las trayectorias de dinámica molecular. Se divide el análisis para los sistemas en ausencia (arriba) y presencia (abajo) de SLA. Las líneas verticales corresponden al valor de la misma distancia entre sitios catalíticos medidos para la estructura cristalográfica (8GZP, iniciación) y para las conformaciones reportadas como cerrada (PDB ID 5DTO, DV3) y abierta (PDB ID 5ZQK, DV2), así como al complejo de elongación (PDB ID 8GZR).

5.4.2 Interacciones clave entre la NS5-pulgar y el *top loop* de SLA

A partir de un análisis global de las interacciones de los aminoácidos de NS5pol y el SLA, se observaron tres nucleótidos (posiciones 30, 31 y 32 del *top loop*) que mantienen contactos estables en todos los sistemas (Figura 5.3). El hecho de haber encontrado interacciones similares en todos los sistemas resulta coherente con la evidencia de que las mediciones experimentales de la unión del SLA a la NS5pol para los sistemas estudiados no han

reportado diferencias significativas [15,16]. En la Tabla 5.2 se resume la evidencia experimental de los sistemas NS5:SLA estudiados.

Tabla 5.2. Secuencias del *top loop* de los sistemas NS5:SLA estudiados, su presencia o no en la naturaleza y características experimentales (unión proteína-ARN y actividad polimerasa *de novo*).

Sistemas	***Top loop***	**Natural**	**Unión**	**Actividad *de novo***
NS5:SLA	CAG	Sí	+	+
NS5:SLA^{mut}	C**GA**	No	+	-
NS5:SLA^{rev}	CG**G**	Sí	+	+
NS5^{F1mut}:SLA	CAG	Sí	+	-

El nucleótido C30 del SLA, cuya base no ha sido mutada en este estudio, presentó interacciones estables y similares para los sistemas NS5:SLA, NS5^{F1mut}:SLA y NS5:SLA^{rev}. Sin embargo, para la mutante NS5:SLA^{mut}, se observó una desestabilización de la interacción del fosfato con la R888 y un acercamiento de la base hacia la R770. En cambio, la posición 31 (que tiene una Adenina como base para el SLA WT pero en ambas variantes SLA^{mut} y SLA^{rev}, fue reemplazada por una Guanina), no muestra cambios significativos en las interacciones ARN-proteína para los sistemas estudiados, a excepción de un sutil acercamiento de la R888 hacia el fosfato acompañado de una mayor variabilidad en la distancia que caracteriza esta interacción. Este aumento en el desvío estándar de las distancias que caracterizan los enlaces está presente en muchas de las interacciones de los fosfatos de las posiciones 30, 31 y 32 con la proteína en el sistema NS5:SLA^{mut}. Por último, se analizó la posición 32, que resulta la más interesante ya que experimentalmente, se encontró que la secuencia no replicativa de un virus mutado según SLA^{mut} (A31G, G32A), revierte volviendo al virus replicativo a partir de un único cambio en esta posición, recuperando la posición 32 y convirtiéndo su secuencia en SLA^{rev} (A31G). En este caso, si bien no se observaron diferencias significativas en las interacciones del fosfato con la NS5, se evidenció la pérdida de la interacción entre la base A32 y la R770 en presencia de SLA^{mut} y una recuperación parcial del enlace G32-R770 en presencia de SLA^{rev}, enfatizando no sólo la importancia de la posición 32 sobre el SLA, sino también la relevancia de este aminoácido.

Las conformaciones del *top loop*, y específicamente de las posiciones 30, 31 y 32, mantienen interacciones estables en presencia del SLA WT. En ambos sistemas (NS5 y NS5^{F1mut}), la base C30 junto con la ribosa de este nucleótido forman uniones de puente hidrógeno con la base G32 que se mantienen estables a lo largo de toda la dinámica. Específicamente, estas interacciones están dadas entre el carbonilo (aceptor de puente hidrógeno) de la base C30 y uno de los N-H (donor de puente hidrógeno) de la base G32, por un lado, y un hidroxilo (donor de puente hidrógeno) de la ribosa de C30 y el carbonilo (aceptor de puente hidrógeno) de la base G32, por otro lado. Al introducir la mutación G32A en SLA^{mut}, el grupo carbonilo de la base es reemplazado grupo amino, que le impide funcionar como aceptor de puente hidrógeno y, a su vez, un N aromático reemplaza al grupo N-H, lo cual explica que ambas

interacciones se pierdan. Contrariamente a lo esperado, estas interacciones no se recuperan en la revertante, a pesar de haber recuperado la presencia de las dos bases involucradas en las interacciones.

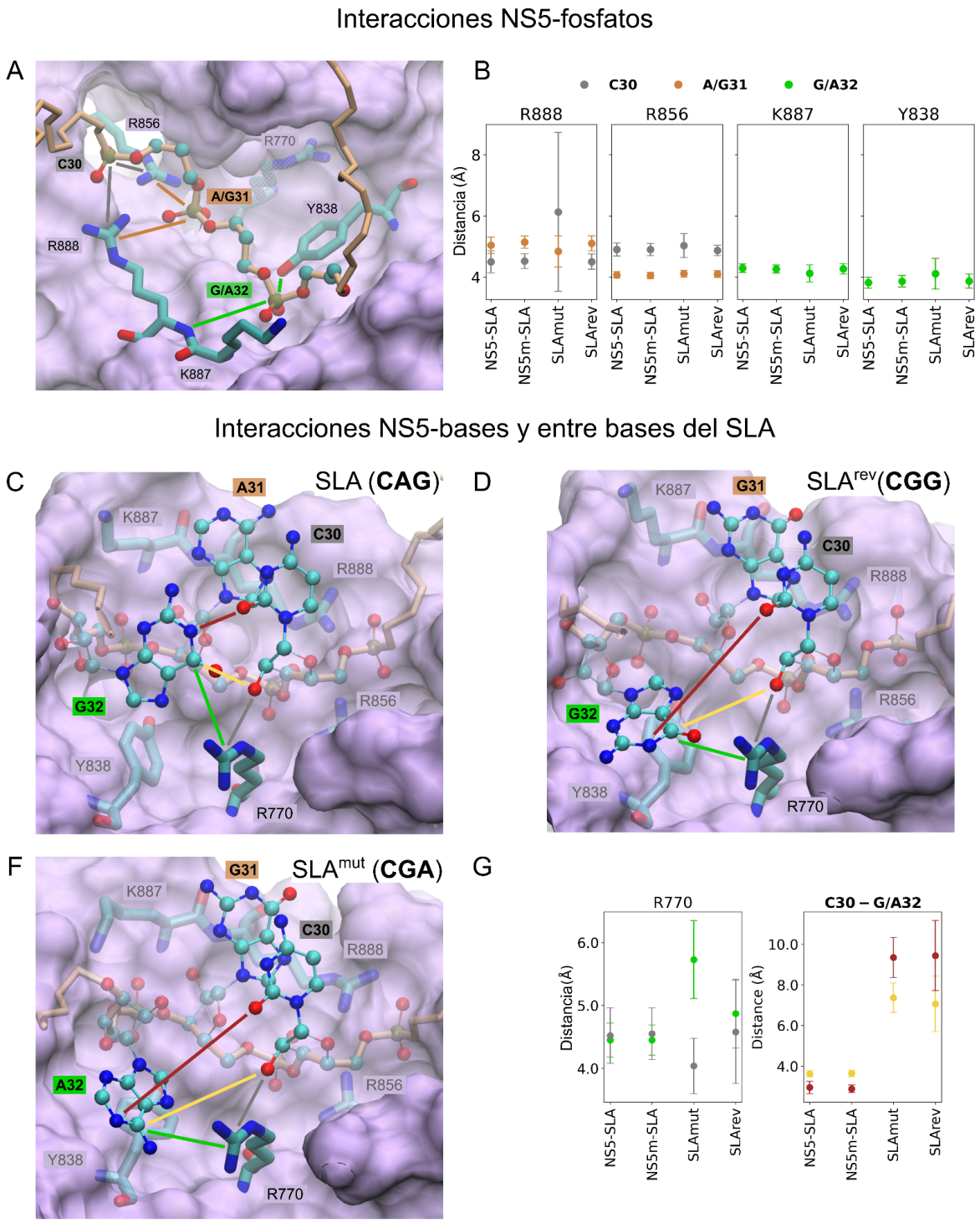


Figura 5.3. Representación esquemática de la interfaz SLA / NS5pol-pulgar. **A)** Se muestran el esqueleto de fosfatos del SLA como bastones dorados y con bastones finos y bolitas coloreados por átomo se destacan los fosfatos de los nucleótidos 30 (etiqueta y líneas representando distancias analizadas en gris), 31 (ídem 30 en naranja) y 32 (ídem 30 en verde). Los aminoácidos involucrados en los enlaces se muestran en bastones coloreados según el color del átomo. **B)** Se grafica el valor

promedio con puntos y las barras que indican ± el desvío estándar de las distancias calculadas según las interacciones del panel A. Los colores representan el nucleótido involucrado y se separa el análisis en paneles según el aminoácido. **C, D, E)** La vista de los tres paneles corresponde a girar 180° el panel A sobre el eje vertical para los sistemas unidos a SLA, SLA^{mut} y SLA^{rev}, respectivamente, con la incorporación de las bases y átomos de la ribosa C30 como bastones finos y bolitas y en el vértice superior derecho se encuentran las bases correspondientes a las posiciones 30, 31 y 32. Las líneas grises y verdes representan las interacciones con la base de la posición 32 y la ribosa de la posición 30, respectivamente. Las líneas roja y amarilla representan las interacciones entre los nucleótidos de las posiciones 30 y 32. **F)** Gráfico de promedio y desvío estándar análogo al panel B pero para las distancias indicadas en los paneles C, D y E.

5.4.3 Análisis de correlación generalizada

Para el siguiente análisis, se tomaron los átomos Cα de la NS5pol y se calcularon y graficaron las matrices de correlación lineal (CL) y correlación generalizada (CG) para todos los sistemas. En la Figura 5.4 se marcan, además, dos regiones correspondientes al vínculo entre el sitio catalítico (señalado con un asterisco y ubicado en ls NS5pol-palma) y por un lado, el sudominio NS5pol-pulgar, en el cual se encuentra el sitio de unión al SLA, y por otro lado, el subdominio NS5pol-dedos, en el cual se encuentra la mutación F1. En la siguiente sección se analiza de manera cuantitativa estos resultados mediante el uso de la cociente señal/ruido.

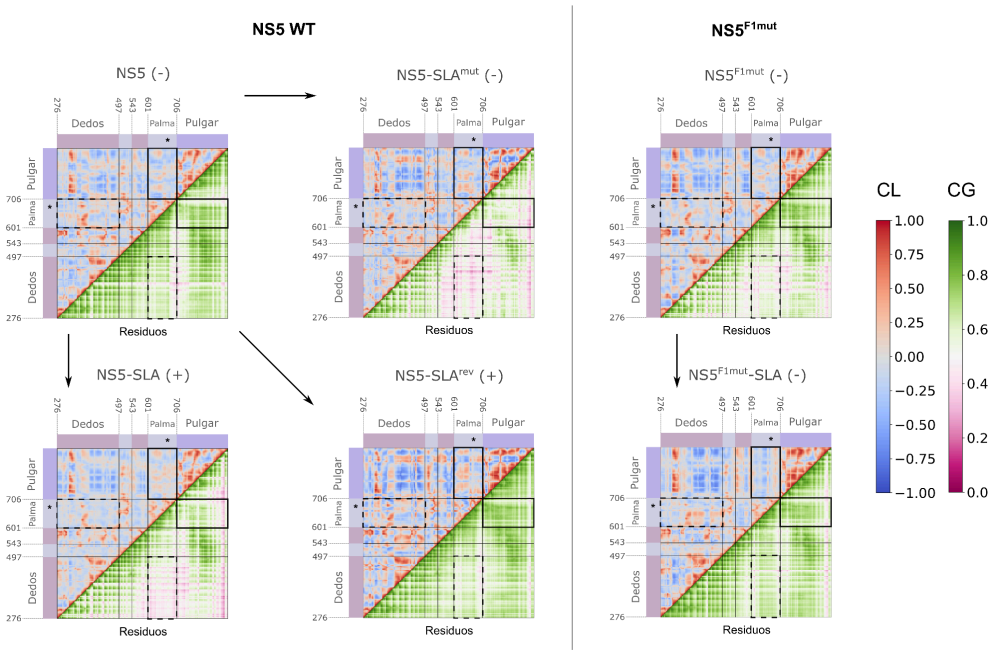


Figura 5.4. Matrices de correlación lineal y generalizada. Se muestran las matrices de correlación lineal (CL) y de correlación generalizada (CG) para los sistemas apo y holo estudiados. El signo más o menos en el título de cada subgráfico indica la actividad reportada de síntesis de ARN viral *de novo*. Los valores de CL varían entre -1 y 1, y se representan con una escala de color de azul a rojo. Los valores de CG varían entre 0 y 1, y se representan con una escala de rosa a verde. En cada gráfico se indican los subdominios de NS5pol-dedos, -pulgar y -palma de la RdRp. El asterisco corresponde a la región donde se encuentran los residuos catalíticos G662, D663 y D664. Las flechas representan la unión a SLA. Los rectángulos negros resaltan la correlación en las regiones entre el NS5pol-pulgar y -palma, así como entre -dedos (específicamente la región donde se encuentra la mutación F1) y -palma.

5.4.4 Vías de comunicación alostérica entre el sitio de unión a SLA y el sitio catalítico de la polimerasa

Con el objetivo de estudiar los caminos óptimos mediante de los cuales se transmite la información desde el sitio de unión al SLA hacia el sitio catalítico de la NS5pol, se calcularon las distancias ponderadas por las correlaciones generalizadas y se redujo el análisis considerando aminoácidos adyacentes. La búsqueda de caminos óptimos y subóptimos requiere definir aminoácidos de partida y de llegada. Por lo tanto, una vez caracterizadas las interacciones entre la NS5pol y el SLA, se seleccionaron los aminoácidos R770, Y838, R856, K887 y R888 como representantes del sitio de unión al SLA. Como se ha mencionado en la Sección 5.4.1, el sitio catalítico de la NS5pol contiene a los aminoácidos G662, D663 y D664. Además, se incorporó a este análisis el aminoácido D533 por participar también de la coordinación de los dos cationes divalentes presentes en el sitio catalítico ocupado que fue resuelto para complejo de elongación (PDB ID 8GZR [17]).

Luego de identificar los caminos óptimos y los primeros 50 caminos subóptimos entre el sitio de unión al SLA hacia el sitio catalítico de la NS5pol, se evaluó el porcentaje de aparición u ocurrencia para cada uno de los aminoácidos que participan en estos caminos (Figura 5.5). Como resultado de este análisis, se observó que para los sistemas WT, los nodos importantes para la comunicación tanto en presencia como en ausencia de SLA pertenecen a los subdominios NS5pol-palma y -pulgar, y dentro de este último, se resaltan en la Figura 5.5 con azul, los aminoácidos pertenecientes al *priming loop*, un segmento clave para la función de la polimerasa. Los primeros dos nodos con mayor ocurrencia corresponden a A799 y H800 con valores de 70-80% aproximadamente. Además, H798 aparece con más del 30% en ambos sistemas y en ausencia de SLA, también se destaca N802. Otros nodos compartidos en ambos sistemas corresponden a E834, T811, M804, M886 y H768. Por último, Y766 se destaca por tener más del 30% de ocurrencia solamente en presencia de SLA.

Para los sistemas con mutaciones, aparecen rutas alternativas que involucran aminoácidos pertenecientes al Motivo E y Pre-Motivo E (ubicados justo antes del Motivo E), señalados en rojo y naranja en la Figura 5.5, respectivamente, que no están presentes en los sistemas WT. Para los aminoácidos con >30% de ocurrencia, las mutantes en el motivo F1 se comportan de manera similar a los sistemas WT, pero con valores más bajos de ocurrencia en los nodos del *priming loop* mencionados. En cambio, al evaluar la mutante no replicativa del *Top Loop,* NS5:SLA^{mut}, los nodos con mayor ocurrencia dejan de ser los del *priming loop* y en su lugar se encuentran aminoácidos del Motivo conservado E y otros cercanos, que si bien cumple una función estructural para la NS5, no tiene una implicancia directa en la actividad polimerasa. Llamativamente, la variante revertante NS5:SLA^{rev}, recupera los nodos del *priming loop* con los mayores porcentajes de ocurrencia para la comunicación entre el sitio de unión del SLA y el sitio catalítico de la NS5pol, casi a los valores del WT.

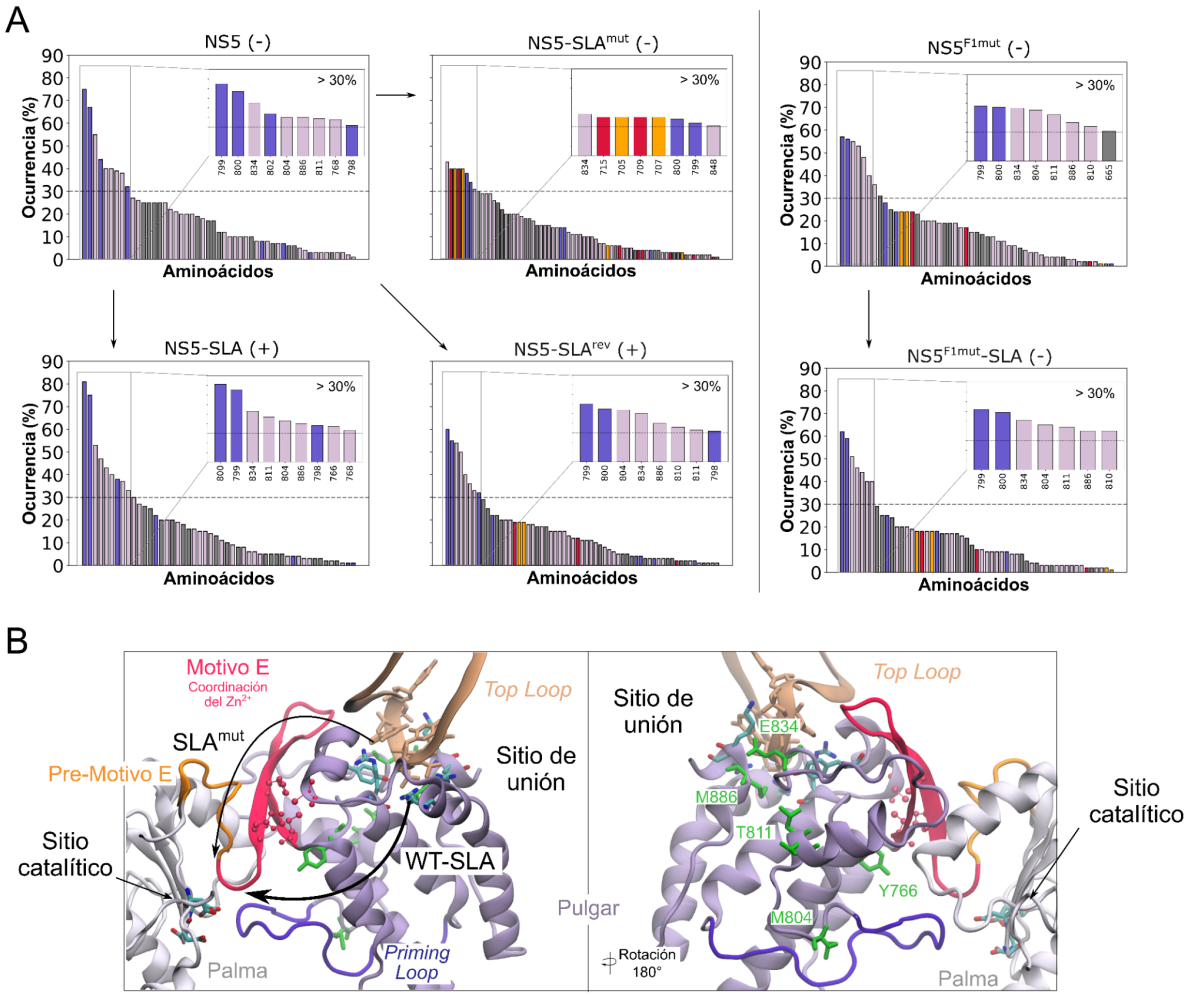


Figura 5.5. A) Porcentajes de ocurrencia mayores al 1% para los aminoácidos que participan en los caminos óptimos y los 50 primeros caminos subóptimos que conectan el sitio de unión (R770, Y838, R856, K887 y R888) al SLA y el sitio catalítico (G662, D663, D664 y D533) de la NS5pol colorados según regiones relevantes y subdominios (*priming loop* en azul, motivo E en rojo, pre-motivo E en naranja, NS5pol-pulgar en violeta, -palma en gris y -dedos en rosa, aunque este último aparece porque no se ha encontrado en este estudio). Para los aminoácidos con más del 30% de ocurrencia en los caminos se muestra la numeración en el gráfico aumentado arriba a la derecha de cada subpanel. **B)** Zoom de la interfaz de interacción SLA-NS5pol mostrando los residuos de la NS5 del sitio de unión y del sitio catalítico con bastones coloreados según el átomo y los del SLA con bastones en naranja claro. Los aminoácidos más ocurrentes involucrados en los caminos óptimos que no pertenecen al *priming loop* se muestran en verde. Los colores representan las regiones relevantes y subdominios con el mismo código de colores usado en el panel A. La imagen de la izquierda se encuentra rotada 180° sobre el eje vertical con respecto a la de la izquierda.

Los aminoácidos que participan en la comunicación entre el sitio de unión y el sitio catalítico pero que no forman parte de motivos conservados resultan interesantes para la propuesta de mutaciones que puedan afectar esa comunicación y generar virus atenuados, así como para la propuesta de regiones y posterior búsqueda de sitios drogables que puedan interferir con esta comunicación. En particular los residuos propuesto son: M804, T811, M886, E834. El residuo Y766 ha sido estudiado por pertenecer a un bolsillo que aloja una molécula inhibidora

del virus del dengue y su mutación por Alanina impide la replicación [38], lo que aporta evidencia a favor de que este método está seleccionando nodos relevantes.

Por último, se caracterizó la efectividad en la comunicación de los caminos obtenidos entre el sitio de unión y el sitio catalítico de la NS5pol, realizando un cálculo de relación señal-ruido (S/R). En primer lugar, se calculó la longitud de cada camino, medida como la cantidad de aminoácidos o nodos que lo componen. En segundo lugar, y mirando las distribuciones de todas las longitudes de los caminos de todos los sistemas, se fragmentó el análisis en tres partes: Los caminos de entre 4 y 8 aminoácidos fueron considerados como “cortos”, los de entre 9 y 13 como “intermedios” y los de entre 14 y 18, como “largos”. Los resultados se muestran en la Tabla 5.3.

En todos los casos, los caminos largos resultan en relaciones S/R cercanas a 1, lo que significa que la transmisión de información a través de estas vías no es eficiente, ya que no se destaca con respecto al ruido. Notablemente, tanto para los caminos cortos como para los intermedios, la relación S/R presentó el valor más alto en el caso de NS5-SLA *wild-type*, sugiriendo la comunicación más efectiva posible en los sistemas estudiados. En ausencia de SLA, el sitio de unión de SLA y el sitio catalítico no presentan caminos óptimos o subóptimos de longitudes cortas, pero sí intermedias, siendo estas levemente más altas que para el caso de la mutantes F1. Esta mutante, sí emplea caminos cortos de transmisión de información, que aumentan levemente su relación S/R en presencia de SLA y por otro lado, presenta el mismo valor para la relación S/R de los caminos intermedios en presencia y ausencia de SLA. Para las variantes de SLA, la relación S/R de la mutante no replicativa SLA^{mut} tanto en los caminos cortos como los intermedios fue mayor que para la replicativa obtenida como revertante SLA^{rev}, pero ambos valores fueron mucho menores que en presencia de SLA WT.

Tabla 5.3. Valores de relación señal-ruido para los complejos NS5-SLA, considerando por separado caminos cortos, intermedios o largos. La actividad reportada para el dominio polimerasa en cada sistema se refleja con un signo -/+.

		Largo del camino (aminoácidos)		
Sistema	**NS5pol**.	**4-8**	**9-13**	**14-18**
NS5	-	ND	1.9	1.2
$NS5^{F1mut}$	-	3.7	1.5	1.0
$NS5^{F1mut}$-SLA	-	4.3	1.5	0.5
NS5-SLA	+	7.4	2.2	1.0
NS5-SLA^{mut}	-	4.9	1.5	0.7
NS5-SLA^{rev}	+	3.9	1.1	0.9

5.5 Discusión

5.5.1 Requisitos conformacionales entre los subdominios de la NS5 para la actividad polimerasa

Para profundizar en la comprensión de la activación alostérica del sitio catalítico del dominio enzimático NS5pol a partir de la unión del segmento de ARN no codificante y estructurado SLA, se realizaron simulaciones de dinámica molecular. Los sistemas modelados consistieron en la proteína *apo* y *holo* para los sistemas WT y para la mutante F1 de la proteína, un sistema que conserva la unión entre SLA y NS5 pero ha perdido la capacidad de activar el sitio catalítico de la NS5pol en presencia de SLA. Además se incluye el estudio de dos variantes de SLA unidas a la NS5 WT: una de ellas no es capaz replicar el virus ($NS5^{mut}$) y la otra, que se diferencia de la anterior por el cambio de una sola base, sí lo es ($NS5^{rev}$). Los resultados obtenidos se interpretan a continuación en el contexto de la bibliografía experimental que incluye ensayos de replicación, actividad polimerasa y afinidad entre la NS5 y el SLA.

La $NS5^{F1mut}$ es capaz de sintetizar ARN a partir de una cadena repetitiva de bases de citosina, poli(rC), pero no replica el ARN viral en ensayos celulares ni es incapaz de sintetizar la cadena complementaria del ARN viral, salvo en condiciones de no-especificidad, que se logran al incorporar Mn^{2+}. Para la NS5 WT ocurre lo contrario, es decir que es incapaz de usar ARN poli(rC) como molde, salvo en condiciones de no-especificidad, pero sintetiza la cadena complementaria al ARN viral, con el requerimiento de la presencia del SLA. Por otro lado, ambas proteínas comparten dos características: tienen la misma afinidad por el SLA y son capaces de elongar una secuencia *primer* unida al ARN molde. Las conclusiones de los experimentos proponen que la región F1 es importante para la activación mediada por el SLA de la polimerasa. Posterior a la publicación de los resultados anteriores, la estructura resuelta para el complejo de elongación muestra que K456 interacciona con una de las bases del ARN molde, mientras que en la estructura de iniciación, que fue la utilizada para este trabajo, esta región no pudo ser resuelta, lo que indica que experimenta una gran flexibilidad.

Parte de los requerimientos para la activación de la actividad polimerasa por la presencia de SLA podrían explicarse a partir de las conformaciones abierta y cerrada que adopta la NS5. Medida como la distancia entre los sitios catalíticos, la estabilidad de la conformación cerrada podría jugar un rol fundamental en la modulación de la actividad catalítica de la NS5pol. En ausencia de SLA, la mutante $NS5^{F1mut}$ adquiere conformaciones intermedias entre abierta y cerrada con alta frecuencia, a diferencia de la proteína WT que se mantiene preferentemente en la conformación cerrada, desplazándose a la conformación abierta con menor frecuencia pero sin adoptar conformaciones intermedias de manera estable. La flexibilidad observada en la $NS5^{F1mut}$ podría conferirle la capacidad de sintetizar ARN sin el requerimiento de la unión al SLA.

Ensayos celulares han mostrado que la mutación de sólo dos bases de todo el ARN viral, localizadas en el *top loop* del SLA impide la replicación del DENV. Al permitir que ocurran mutaciones espontáneas en el ARN viral, se han encontrado secuencias revertantes que sí replican (SLA^{mut} y SLA^{rev}, respectivamente). Para estas variantes de SLA se conocen resultados experimentales de afinidad, mostrando similares valores de las variantes con

respecto al SLA WT, y de actividad polimerasa de la NS5pol, que evidentemente implica la iniciación de la síntesis de ARN; pero no se conocen los resultados análogos para la NS5 completa, ni su posible capacidad de elongación, como los mencionados para la mutante F1.

Para los resultados conformacionales de los complejos proteína:ARN, además de verificar que la NS5 se encuentra en todos los casos en la conformación cerrada, se observó que los dos sistemas no replicativos ($NS5^{F1mut}$:SLA y NS5:SLA^{mut}) desplazan sus distribuciones hacia la izquierda (menores distancias que separan los sitios catalíticos). En cambio, los dos sistemas replicativos (NS5:SLA y NS5:SLA^{rev}) se mantienen alrededor del valor inicial. Por lo tanto, la estabilidad conformacional de los subdominios de la NS5 parecería ser indispensable para asegurar una correcta iniciación de la síntesis de ARN.

5.5.2 La importancia de un grupo carbonilo en la base de la posición 32 del SLA

Se había propuesto que una serie de cuatro (cinco para DENV2) nucleótidos, C30, A31, G32 y A33 (U33 y A34 para DENV2), pertenecientes al *top loop* del SLA WT era la región que intervenía en la unión a la NS5pol, previo a la publicación de la estructura de Cryo-EM. Además, los ensayos de mutaciones del *top loop* para los cuales se obtuvieron secuencias revertantes, mostraron que es indispensable la presencia del par C30-G32 y las estructuras resueltas del SLA unido a la NS5 explican este fenómeno ya que estos nucleótidos interaccionan.

Las interacciones entre la NS5 WT y la $NS5^{F1mut}$ no presentaron diferencias significativas, a pesar de las diferencias conformacionales de la NS5. A partir del estudio de las interacciones, derivado de las simulaciones de dinámica molecular, fue posible corroborar que las dos uniones de puente hidrógeno entre C30 y G32 observadas en la estructura resuelta se mantienen estables. Sin embargo, no ocurrió lo mismo con la variante revertante, la cual se sabe que tiene actividad polimerasa y replica el ARN viral mutado en ensayos celulares. Por otro lado, sería esperable que existan diferencias entre el SLA^{rev} y el SLA WT, ya que la secuencia del *top loop* CAG está conservada para los cuatro serotipos de DENV. La secuencia del *top loop* de SLA^{rev}, CGG, por su parte, difiere en la posición 31 con la de DENV, y se encuentra en el *top loop* de otros flavivirus (por ejemplo el virus de la encefalitis por garrapata , TBEV). Al estudiar la variante mutada y no replicativa SLA^{mut}, las interacciones del par C30-A32 no tienen la posibilidad de establecerse, ya que el cambio de la base G32 por A, impide la formación de los puentes de hidrógeno establecidos en el sistema WT. Curiosamente, la secuencia de SLA^{mut}, CGA, no se encuentra en ningún otro flavivirus. Otra mutante reportada (no estudiada en este trabajo) consistió en realizar los siguientes cambios: C30A y A33C. A variante no replicativa (AAGC) también se le permitió encontrar una secuencia revertante y en este caso, con la recuperación de la posición C30 alcanzó para volver a replicar, enfatizando la importancia de este par.

En la estructura WT, el fosfato del nucleótido C30 se encuentra fuertemente atraído a la R888, manteniendo una interacción estable a lo largo de toda la trayectoria. Por su parte, la ribosa de este mismo nucleótido se encuentra cerca de la R770, así como la base de G32. La pérdida de las interacciones entre C30-G32 en presencia de la $NS5^{rev}$ no modifica apreciablemente la interacción C30-R888, pero sí parcialmente las interacciones mencionadas con R770. Este resultado sugiere que si bien la presencia de la base G31 es

aceptable para la capacidad replicativa del virus, no es óptima para permitir las interacciones que se observan con el SLA WT, como lo es A31, lo cual explica la alta conservación en los serotipos de DENV. Con la presencia de $NS5^{mut}$, la interacción A32-R770 se pierde, ya que el cambio de base reemplaza un oxígeno carbonilo por un grupo amino y en su lugar R770 se posiciona más cerca de la ribosa de C30, en comparación con el WT y la revertante. Este cambio en las interacciones también repercute debilitando notablemente la interacción del fosfato del nucleótido C30 con R888. En el caso mencionado anteriormente de la revertante, si bien la interacción C30-G32 no es estable, la presencia de G32 que contiene el grupo carbonilo, permite conservar el acercamiento con R770 y no debilita el enlace con R888. La importancia de esta última arginina es ampliamente reconocida. Un ejemplo de esto es que se ha propuesto que la interacción entre la región estructurada del extremo 3' del ARN viral y la R888 puede activar la síntesis de novo de manera independiente del SLA [39]. Ambas argininas además, al ser mutadas por alanina, disminuyen la afinidad de la NS5 por el SLA [17,40].

Las secuencias de todos los *Top Loops* de los flavivirus (Figura 5.6), sugieren una conservación evolutiva de esta interacción en diferentes flavivirus dada por el nucleótido G mayoritariamente y en menor medida U (que comparte el carbonilo implicado en la interacción con R770 descrita), pero no se observa la presencia de A o C.

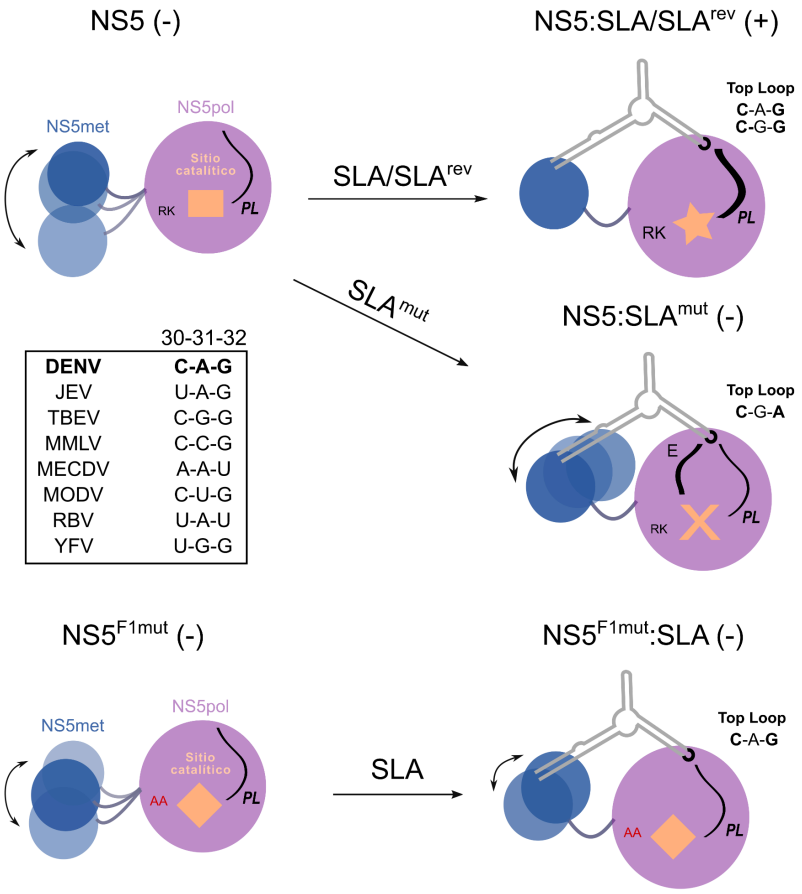
5.5.3 Aparición de rutas alternativas y pérdida de eficiencia en la comunicación por la presencia de mutaciones

Bajo el paradigma del alosterismo, la unión de un ligando induce cambios que se propagan a través de los aminoácidos hasta el sitio catalítico. Para evaluar la comunicación entre el sitio de unión al SLA y el sitio catalítico de la NS5pol, se caracterizaron los caminos óptimos y subóptimos que comunican ambos sitios en términos de identidad de los aminoácidos involucrados y de relación señal/ruido.

La comunicación se establece en todos los casos estudiados a través de aminoácidos localizados en los subdominios NS5pol-pulgar y -palma. En ningún caso se observaron aminoácidos del subdominio NS5pol-dedos con más del 1% de ocurrencia, descartando la posibilidad de que los residuos del motivo F1 estuvieran involucrados directamente. Habiendo encontrado diferencias significativas para las interacciones del SLA^{mut}, era esperable que esto se transfiera a cambios en los canales de comunicación entre el sitio de unión y el sitio catalítico. A diferencia del sistema WT, en el cual se destacaron los residuos con mayor porcentaje de ocurrencia por pertenecer a la región del *priming loop*, en esta mutante se destacan los residuos del motivo conservado E y la región adyacente, denominada pre-motivo E, disminuyendo la capacidad de transmitir información a través del *priming loop*. El análisis de mayores porcentajes de ocurrencia (>30%) enfatiza además el aumento en la participación de aminoácidos del *priming loop* en los sistemas WT con respecto a las mutantes F1, en cantidad de aminoácidos y en porcentaje. Este efecto, aunque más sútil, también se observa al comparar estas mutantes con el sistema $NS5:SLA^{rev}$, que se distingue por ser replicativo en ensayos celulares.

Al comprar la participación de todos los aminoácidos (con > 1% de ocurrencia) en los caminos óptimos y subóptimos estudiados, se observa que los sistemas WT nunca exploran caminos que atraviesen el motivo estructural y pre-motivo E, concentrando así las vías de

comunicación por otros aminoácidos. En cambio, todos los sistemas con mutaciones, incluida la revertante SLA^{rev} que al igual que el sistema WT es capaz de replicar, incorporan los aminoácidos pertenecientes a estos motivos, lo que se interpreta como un desvío del comportamiento óptimo. Más aún, los valores de la relación señal/ruido muestran que el sistema NS5:SLA se destaca por tener la comunicación más eficiente. Por lo tanto, estos resultados aportan evidencia de que el hecho de que las mutantes pierden la capacidad de activar la función polimerasa dependiente del promotor SLA sin perder la afinidad, se debe a cambios en la comunicación entre el el sitio de unión del SLA y el sitio catalítico. En la Figura 5.6 se muestra un resumen de los efectos de las mutantes en las conformaciones relativas de los dominios de la NS5 y los caminos alostéricos que conectan el sitio de unión de NS5pol y el sitio catalítico.



30-31-32

DENV	C-A-G
JEV	U-A-G
TBEV	C-G-G
MMLV	C-C-G
MECDV	A-A-U
MODV	C-U-G
RBV	U-A-U
YFV	U-G-G

Figura 5.6. Representación esquemática de la activación alostérica del NS5pol posterior a la unión del SLA, resumiendo el efecto de mutaciones puntuales en cada biomolécula. La tabla muestra la secuencia de los tres nucleótidos del SLA involucrados en el reconocimiento del NS5pol al genoma viral para diversos flavivirus.

La comparación entre los sistemas NS5:SLA y NS5:SLA^{mut} permite evidenciar claramente a través del análisis realizado las razones por las cuales el segundo no es funcionalmente activo: La conformación de la NS5 cambia sustancialmente en presencia del SLA^{mut}, las interacciones proteína-ARN se modifican, las vías de comunicación disminuyen la participación de una región funcional clave en la actividad polimerasa, el *priming loop,* y en su lugar desvían los canales hacia motivos estructurales, motivo y pre-motivo E, y por último,

disminuye la relación señal/ruido que da cuenta de la eficiencia en esa comunicación. En contraposición, la comparación entre los resultados de NS5:SLA WT con NS5:SLA^{rev} y NS5^{F1mut}:SLA no parece ser tan sencilla de conceptualizar, teniendo en cuenta que de este par de variantes, la primera es funcional y la segunda no, y que a partir de las simulaciones sólo se observan diferencias sutiles entre ellos: La conformación de la NS5 se aleja hacia valores menores en la segunda y a pesar de conservar las interacciones proteína-ARN, la variante NS5^{F1mut}:SLA disminuye el porcentaje de ocurrencia del aminoácido H798, presente con >30% tanto en los sistemas WT como en la revertante.

5.5.4 Mutaciones propuestas para el desarrollo de virus atenuados como estrategia preventiva contra el DENV

Nuestros resultados destacan residuos específicos (M804, T811, M886, Y766 y E834) como nodos críticos en la comunicación alostérica, lo que sugiere que mutaciones en estos sitios podrían interferir con la activación de la replicación viral, abriendo posibilidades para el diseño de virus atenuados o nuevos inhibidores. Se han identificado inhibidores alostéricos para la polimerasa NS5 mediante *screening* basado en fragmentos y cristalografía de rayos X. El bolsillo denominado "*N pocket*" contiene el aminoácido Y766 [38], uno de los nodos destacados para la comunicación entre el sitio de unión al SLA y el sitio catalítico de la NS5pol. Más aún, en el mismo trabajo, Lim *et al* reportan que la mutación Y766A afecta específicamente la síntesis *de novo* de la NS5pol.

5.6 Conclusiones

Estos resultados muestran que la presencia de SLA selecciona una conformación cerrada de la NS5, lo que sugiere que esta estabilización estructural es crucial para la actividad polimerasa. Por otro lado, se deriva de las simulaciones realizadas que la interacción del *top loop* con la R770 a través del carbonilo en la base 32 es un elemento necesario para la activación de la polimerasa NS5 del DENV. La identificación de residuos clave involucrados en la comunicación alostérica entre el sitio de unión al SLA y el sitio catalítico de NS5pol permitió encontrar nodos crítico que ofrecen nuevas perspectivas para el diseño de estrategias antivirales, ya sea mediante la creación de virus atenuados o el desarrollo de inhibidores específicos que bloqueen estos caminos.

Bibliografía

1. El Sahili A, Lescar J. Dengue virus non-structural protein 5. Viruses. 2017;9: 91. doi:10.3390/v9040091

2. Lim SP, Noble CG, Shi P-Y. The dengue virus NS5 protein as a target for drug discovery. Antiviral Res. 2015;119: 57–67. doi:10.1016/j.antiviral.2015.04.010

3. El Sahili A, Soh TS, Schiltz J, Gharbi-Ayachi A, Seh CC, Shi P-Y, et al. NS5 from dengue virus serotype 2 can adopt a conformation analogous to that of its Zika virus and Japanese encephalitis virus homologues. J Virol. 2019;94. doi:10.1128/JVI.01294-19

4. Wu J, Ye H-Q, Zhang Q-Y, Lu G, Zhang B, Gong P. A conformation-based intra-molecular initiation factor identified in the flavivirus RNA-dependent RNA polymerase. PLoS Pathog. 2020;16: e1008484. doi:10.1371/journal.ppat.1008484

5. Zhao Y, Soh TS, Zheng J, Chan KWK, Phoo WW, Lee CC, et al. A crystal structure of the Dengue virus NS5 protein reveals a novel inter-domain interface essential for protein flexibility and virus replication. PLoS Pathog. 2015;11: e1004682. doi:10.1371/journal.ppat.1004682

6. Zhao Y, Soh TS, Lim SP, Chung KY, Swaminathan K, Vasudevan SG, et al. Molecular basis for specific viral RNA recognition and 2’-O-ribose methylation by the dengue virus nonstructural protein 5 (NS5). Proc Natl Acad Sci U S A. 2015;112: 14834–14839. doi:10.1073/pnas.1514978112

7. Klema VJ, Ye M, Hindupur A, Teramoto T, Gottipati K, Padmanabhan R, et al. Dengue virus nonstructural protein 5 (NS5) assembles into a dimer with a unique methyltransferase and polymerase interface. PLoS Pathog. 2016;12: e1005451. doi:10.1371/journal.ppat.1005451

8. Wu J, Lu G, Zhang B, Gong P. Perturbation in the conserved methyltransferase-polymerase interface of flavivirus NS5 differentially affects polymerase initiation and elongation. J Virol. 2015;89: 249–261. doi:10.1128/JVI.02085-14

9. Klema VJ, Padmanabhan R, Choi KH. Flaviviral replication complex: Coordination between RNA synthesis and 5’-RNA capping. Viruses. 2015;7: 4640–4656. doi:10.3390/v7082837

10. Choi KH. The Role of the Stem-Loop A RNA Promoter in Flavivirus Replication. Viruses. 2021;13: 1107. doi:10.3390/v13061107

11. Filomatori CV, Lodeiro MF, Alvarez DE, Samsa MM, Pietrasanta L, Gamarnik AV. A 5′ RNA element promotes dengue virus RNA synthesis on a circular genome. Genes Dev. 2006;20: 2238–2249. doi:10.1101/gad.1444206

12. Lodeiro MF, Filomatori CV, Gamarnik AV. Structural and functional studies of the promoter element for dengue virus RNA replication. J Virol. 2009;83: 993–1008. doi:10.1128/JVI.01647-08

13. Alvarez DE, De Lella Ezcurra AL, Fucito S, Gamarnik AV. Role of RNA structures present at the 3’UTR of dengue virus on translation, RNA synthesis, and viral replication. Virology. 2005;339: 200–212. doi:10.1016/j.virol.2005.06.009

14. You S, Padmanabhan R. A novel in vitro replication system for Dengue virus. Initiation of RNA synthesis at the 3’-end of exogenous viral RNA templates requires 5'- and 3'-terminal

complementary sequence motifs of the viral RNA. J Biol Chem. 1999;274: 33714–33722. doi:10.1074/jbc.274.47.33714

15. Iglesias NG, Filomatori CV, Gamarnik AV. The F1 motif of dengue virus polymerase NS5 is involved in promoter-dependent RNA synthesis. J Virol. 2011;85: 5745–5756. doi:10.1128/JVI.02343-10

16. Filomatori CV, Iglesias NG, Villordo SM, Alvarez DE, Gamarnik AV. RNA sequences and structures required for the recruitment and activity of the dengue virus polymerase. J Biol Chem. 2011;286: 6929–6939. doi:10.1074/jbc.M110.162289

17. Osawa T, Aoki M, Ehara H, Sekine S-I. Structures of dengue virus RNA replicase complexes. Mol Cell. 2023;83: 2781–2791.e4. doi:10.1016/j.molcel.2023.06.023

18. Obi JO, Kihn KC, McQueen L, Fields JK, Snyder GA, Deredge DJ. Structural dynamics of the dengue virus non-structural 5 (NS5) interactions with promoter stem-loop A (SLA). Npj Viruses. 2025;3: 30. doi:10.1038/s44298-025-00112-1

19. Waterhouse A, Bertoni M, Bienert S, Studer G, Tauriello G, Gumienny R, et al. SWISS-MODEL: homology modelling of protein structures and complexes. Nucleic Acids Res. 2018;46: W296–W303. doi:10.1093/nar/gky427

20. Rother M, Rother K, Puton T, Bujnicki JM. ModeRNA: a tool for comparative modeling of RNA 3D structure. Nucleic Acids Res. 2011;39: 4007–4022. doi:10.1093/nar/gkq1320

21. Maier JA, Martinez C, Kasavajhala K, Wickstrom L, Hauser KE, Simmerling C. Ff14SB: Improving the accuracy of protein side chain and backbone parameters from ff99SB. J Chem Theory Comput. 2015;11: 3696–3713. doi:10.1021/acs.jctc.5b00255

22. Aytenfisu AH, Spasic A, Grossfield A, Stern HA, Mathews DH. Revised RNA dihedral parameters for the Amber force field improve RNA molecular dynamics. J Chem Theory Comput. 2017;13: 900–915. doi:10.1021/acs.jctc.6b00870

23. Peters MB, Yang Y, Wang B, Füsti-Molnár L, Weaver MN, Merz KM Jr. Structural Survey of Zinc Containing Proteins and the Development of the Zinc AMBER Force Field (ZAFF). J Chem Theory Comput. 2010;6: 2935–2947. doi:10.1021/ct1002626

24. Li P, Merz KM Jr. MCPB.py: A Python Based Metal Center Parameter Builder. J Chem Inf Model. 2016;56: 599–604. doi:10.1021/acs.jcim.5b00674

25. Li P, Merz KM Jr. Parameterization of a Dioxygen Binding Metal Site Using the MCPB.py Program. Methods Mol Biol. 2021;2199: 257–275. doi:10.1007/978-1-0716-0892-0_15

26. People O. Citation. [cited 27 Jun 2025]. Available: https://gaussian.com/citation/

27. Olsson MHM, Søndergaard CR, Rostkowski M, Jensen JH. PROPKA3: Consistent treatment of internal and surface residues in empirical pKa predictions. J Chem Theory Comput. 2011;7: 525–537. doi:10.1021/ct100578z

28. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML. Comparison of simple potential functions for simulating liquid water. J Chem Phys. 1983;79: 926–935. doi:10.1063/1.445869

29. Case DA, Aktulga HM, Belfon K, Cerutti DS, Cisneros GA, Cruzeiro VWD, et al. AmberTools. J Chem Inf Model. 2023;63: 6183–6191. doi:10.1021/acs.jcim.3c01153

30. Case DA, Aktulga HM, Belfon K, Ben-Shalom IY, Berryman JT, Brozell SR, et al. Amber

2023. 2023. Available: https://books.google.com/books?hl=en&lr=&id=F1rGEAAAQBAJ&oi=fnd&pg=PA5&ots=4M6PfoP7wq&sig=r2fOLDcHA6Gdg_fzwB8W0NKP6mM

31. Liu J, Li D, Liu X. A simple and accurate algorithm for path integral molecular dynamics with the Langevin thermostat. J Chem Phys. 2016;145: 024103. doi:10.1063/1.4954990

32. Berendsen HJC, Postma JPM, van Gunsteren WF, DiNola A, Haak JR. Molecular dynamics with coupling to an external bath. J Chem Phys. 1984;81: 3684–3690. doi:10.1063/1.448118

33. Ryckaert J-P, Ciccotti G, Berendsen HJC. Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. J Comput Phys. 1977;23: 327–341. doi:10.1016/0021-9991(77)90098-5

34. Roe DR, Cheatham TE 3rd. PTRAJ and CPPTRAJ: Software for processing and analysis of molecular dynamics trajectory data. J Chem Theory Comput. 2013;9: 3084–3095. doi:10.1021/ct400341p

35. Melo MCR, Bernardi RC, de la Fuente-Nunez C, Luthey-Schulten Z. Generalized correlation-based dynamical network analysis: a new high-performance approach for identifying allosteric communications in molecular dynamics trajectories. bioRxiv. bioRxiv; 2020. doi:10.1101/2020.06.18.160572

36. Patel AC, Sinha S, Palermo G. Graph theory approaches for molecular dynamics simulations. Q Rev Biophys. 2024;57: e15. doi:10.1017/S0033583524000143

37. Sinha S, Molina Vargas AM, Arantes PR, Patel A, O'Connell MR, Palermo G. Unveiling the RNA-mediated allosteric activation discloses functional hotspots in CRISPR-Cas13a. Nucleic Acids Res. 2024;52: 906–920. doi:10.1093/nar/gkad1127

38. Lim SP, Noble CG, Seh CC, Soh TS, El Sahili A, Chan GKY, et al. Potent allosteric dengue virus NS5 polymerase inhibitors: Mechanism of action and resistance profiling. PLoS Pathog. 2016;12: e1005737. doi:10.1371/journal.ppat.1005737

39. Wang S, Chan KWK, Tan MJA, Flory C, Luo D, Lescar J, et al. A conserved arginine in NS5 binds genomic 3' stem-loop RNA for primer-independent initiation of flavivirus RNA replication. RNA. 2022;28: 177–193. doi:10.1261/rna.078949.121

40. Hodge K, Tunghirun C, Kamkaew M, Limjindaporn T, Yenchitsomanus P-T, Chimnaronk S. Identification of a conserved RNA-dependent RNA polymerase (RdRp)-RNA interface required for flaviviral replication. J Biol Chem. 2016;291: 17437–17449. doi:10.1074/jbc.M116.724013

Capítulo 6: Conclusiones y Perspectivas

6.1 Conclusiones generales

El abordaje computacional de fenómenos fisicoquímicos de las enzimas NS3hel y NS5pol del virus del dengue presentado en esta tesis permiten profundizar en la comprensión de la modulación de sus actividades, específicamente en el contexto de la replicación del ARN viral. A continuación se resumen las conclusiones generales de la tesis.

La proteína NS3hel, en las escalas de tiempo estudiadas (ns-μs), se comporta como un motor molecular browniano que alterna entre estados rígidos con estados de mayor flexibilidad a lo largo del ciclo catalítico de hidrólisis de ATP. Los estados flexibles poseen mayor afinidad para el ARN, mientras los estados de mayor flexibilidad están más débilmente unidos al ARN, así permitiendo el desplazamiento sobre la hebra de ARN. Los resultados obtenidos no son compatibles con el modelo de gusano, que ha sido propuesto para otras helicasas de la misma familia.

Se reporta un cambio conformacional obtenido a partir de la observación de la apertura del *loop* β6-β1', adyacente al sitio catalítico, que acumula evidencia de tener roles fundamentales. Por un lado, se propone que la conformación abierta del *loop* funciona como canal de salida del fosfato, que es el primer producto que abandona el sitio catalítico luego de la hidrólisis de ATP y cuya salida es difícil de racionalizar sin considerar esta conformación. Por otro lado, la conformación semiabierta del *loop* permite el ingreso de moléculas de agua, necesarias para la hidrólisis, al sitio catalítico cuando la NS3hel se encuentra unida a ATP. Este último efecto se ve influenciado por la presencia, el largo y la conformación del ARN.

El complejo NS3hel:ARN requiere que el modelo computacional de ARN contenga una longitud mayor a los 7 nucleótidos resueltos en las estructuras cristalográficas. El modelado de la NS3hel unida a un oligonucleótido de 16 bases, a diferencia del de 7, mostró reproducir exitosamente las tendencias experimentales al calcular diferencias de energía de unión entre el ARN y la NS3hel con cambios en la ocupación del sitio catalítico; también evidenció un aumento en la estabilidad de residuos del sitio catalítico, cuya conformación es indispensable para la hidrólisis de ATP; y por último, se observó la capacidad del ARN en este modelo para modular la probabilidad de acomodar un agua catalítica para la reacción de hidrólisis de ATP según la conformación del extremo 5' del ARN.

La presencia de ARN facilita la reacción de hidrólisis de ATP de una manera diferente a la hipotetizada en base a la evidencia disponible, que resalta la nucleofilicidad diferencial de un aminoácido clave en la reacción, E285, ya que puede funcionar como base o nucleófilo, facilitando el ataque del agua catalítica. La NS3hel unida a ATP, en presencia de ARN, genera la ruptura espontánea del ATP sin la participación activa de la base mencionada y es esta facilidad lo que se destaca al comparar el mismo proceso en ausencia de ARN, en donde la ruptura no es espontánea y ocurre al atravesar una barrera energética cuatro veces más alta. Una vez que se forma el intermediario ADP-PO_3, el ataque del agua catalítica para formar productos, en el que sí participa la base E285, ocurre de manera energéticamente similar.

El elemento hélice-giro-hélice presente en el subdominio D3 la NS3hel del virus del dengue se propone como clave para el desarrollo de antivirales debido a que por un lado, aloja moléculas de etanol en simulaciones de dinámica molecular con solvente mixto en los cuatro serotipos de

DENV y en presencia y ausencia de ARN, y, por otro lado, sus aminoácidos presentan diferencias sustanciales, según la presencia/ausencia de ARN, en las correlaciones con los aminoácidos de los motivos conservados implicados en la hidrólisis de ATP. La modulación alostérica sobre este elemento también fue observada en presencia de ARN con el cambio en la ocupación del sitio catalítico de la NS3hel, dado por la alternancia entre mayor flexibilidad de una de las hélices en presencia de ATP y menor flexibilidad en otro caso. Esta región, además, está ubicada en la interfaz de interacción NS3-NS5.

Cuando la polimerasa NS5pol se une al segmento estructurado de ARN viral no codificante denominado SLA, los aminoácidos del sitio de unión y los del sitio catalítico de la NS5pol establecen canales de comunicación eficientes que involucran al *priming loop*, un elemento conservado y esencial para su actividad. Esta comunicación evidenciada en la alta correlación entre el sitio de unión, el sitio catalítico y el *priming loop* permite racionalizar el efecto promotor del SLA sobre la NS5pol al comparar estos resultados con sistemas inactivados como lo son las mutantes del SLA y de la proteína, y la NS5pol en ausencia de SLA.

Los aminoácidos de la NS5pol que participan en la comunicación entre el sitio de unión al SLA y el sitio catalítico y que no forman parte del *priming loop* constituyen posiciones susceptibles a ser mutadas para generar virus atenuados en el contexto del desarrollo de métodos preventivos del virus del dengue.

6.2 Perspectivas

Los resultados de esta tesis destacan la sensibilidad de las dos enzimas del virus del dengue a modificaciones de su entorno. Futuros trabajos podrían explorar si la unión covalente de dos dominios en cada uno de las proteínas virales NS3 y NS5 efectivamente es un mecanismo regulatorio de sus múltiples funciones.

Si bien muchas de las funciones de la NS3hel son independientes del dominio proteasa, NS3pro, otras han mostrado tener influencia e incluso hay bibliografía contrapuesta al respecto para algunas actividades de la NS3hel. El estudio de la NS3 de longitud completa, incluyendo ambos dominios, junto con el efecto de las conformaciones dadas por el linker que los une, ha quedado por fuera de los alcances de esta tesis. Sin embargo, una estudiante de Licenciatura en Química, Jimena Zarraga, a quien acompañé durante sus primeros pasos en las simulaciones computacionales, ha comenzado a modelar estos sistemas.

Dadas las recientes estructuras resueltas, podrían realizarse modelos atomísticos de los complejos NS3:NS5 en presencia y ausencia de ARN. Sería interesante conocer el efecto de la NS5 sobre las funciones de la NS3 y viceversa.

Algunos de los resultados presentados en la tesis aguardan aún validación experimental. Sería útil, por ejemplo, obtener evidencia experimental directa, mediante experimentos de RMN u otros, de las conformaciones del *loop* β6-β1’ observadas en las simulaciones de esta tesis. Se están planificando actualmente ensayos de replicación viral en vivo para evaluar el efecto de mutaciones en los residuos mencionados como el camino alostérico en la NS5 para profundizar en la activación del dominio NS5pol.

Finalmente, es de interés continuar evaluando la conservación de los mecanismos descritos en otros flavivirus.

Anexo Capítulo 4

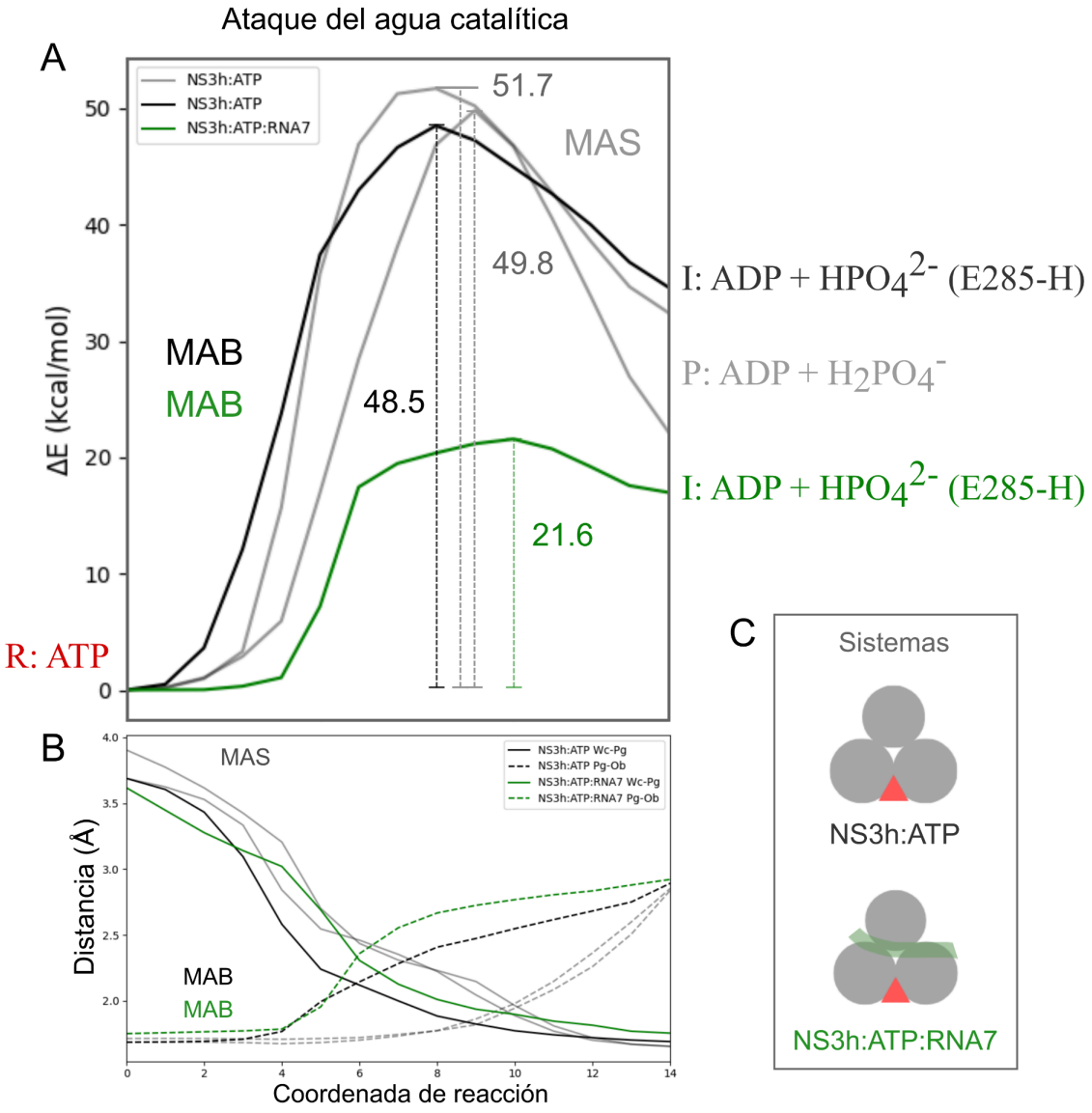


Figura 4A.1. A) Resultados de la búsqueda de caminos de mínima energía potencial con el método NEB para los mecanismos asistido por el sustrato (MAS, serie gris) y asistido por la base (MAB, serie negra) en ausencia de ARN, y para el mecanismo asistido por la base en presencia de ARN (MAB, serie verde). Los intentos de explorar el mecanismo asistido por el sustrato en los complejos NS3hel:ARN resultaron en una protonación espontánea del Glu 285. **B)** Evolución de las distancias durante la reacción para los enlaces Pγ–OWc y Pγ–Oβγ. **C)** Representación de los sistemas incluidos en este análisis.

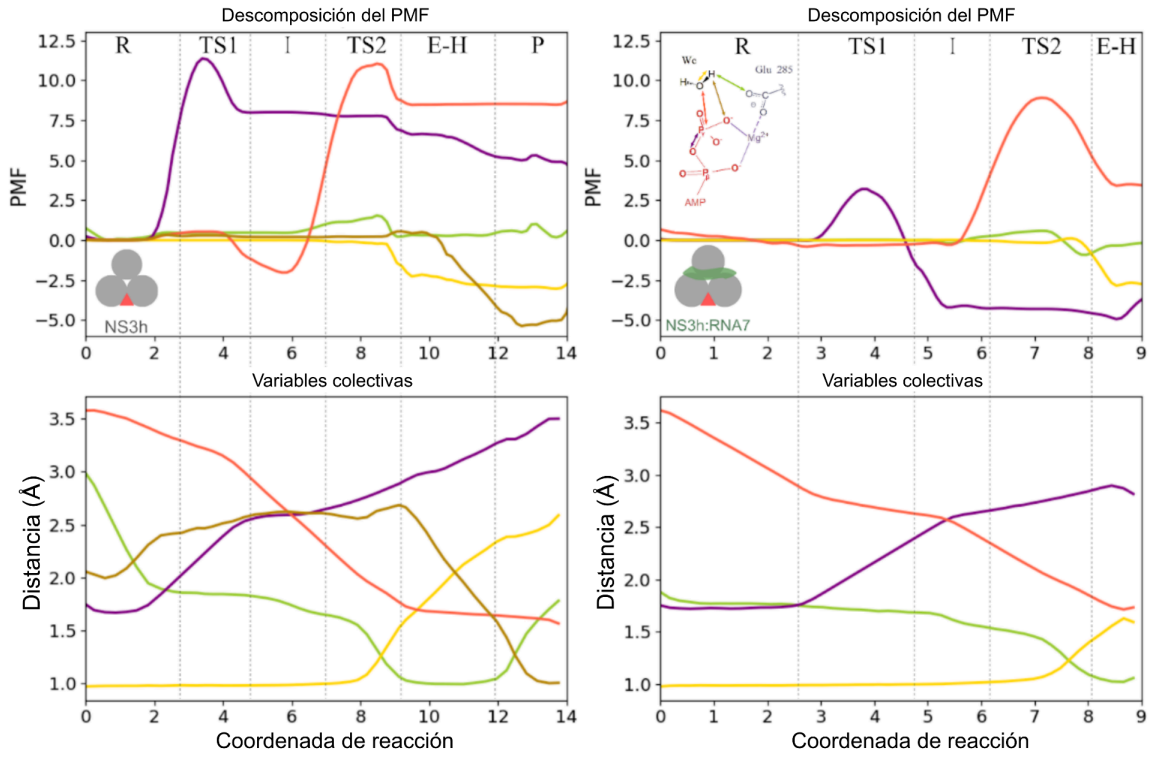


Figura 4A.2. Descomposición del potencial de fuerza media PMF para cada variable colectiva (arriba) y distancias de las variables colectivas a lo largo de la coordenada de reacción hallada mediante el método de la cuerda ASM para el camino de mínima energía libre (abajo). Los gráficos de la izquierda corresponden al sistema NS3hel y los de la derecha a NS3hel:ARN7.

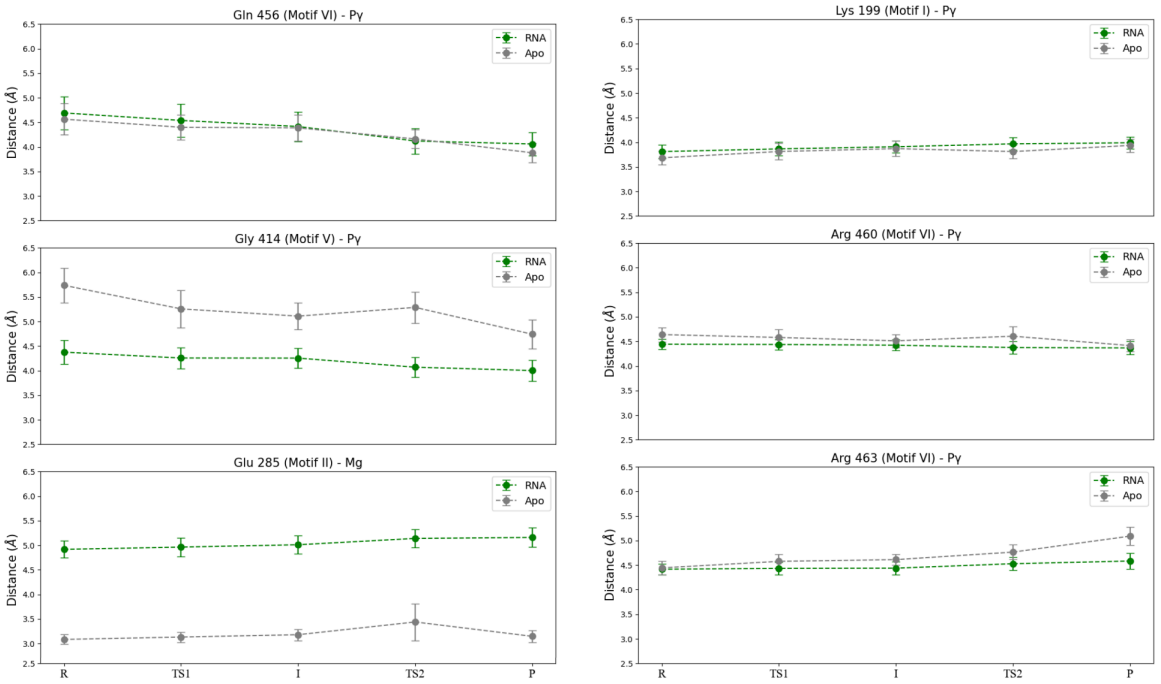


Figura 4A.3. Distancias entre el fosfato gamma (Pγ) y los aminoácidos de motivos conservados que interactúan con el ATP o con el agua catalítica. Las series verdes corresponden a la NS3hel unida al ARN y la gris a la NS3hel en ausencia de ARN.

Tabla 4A1. Variabilidad de las secuencias utilizadas en las simulaciones de dinámica molecular con solvente mixto y conservación de los aminoácidos de los bolsillos encontrados.

Modelo por homología de DENV1				82.7%	88.8%	80.4%
Modelo por homología de DENV2					85.2%	81.8%
Modelo por homología de DENV3						83.4%
Bolsillo	**ResID**		**Conservación DENV1**	**Conservación DENV2**	**Conservación DENV3**	**Conservación DENV4**
Bolsillo D3-ARN	267	M	M: 99.92%, T: 0.08%	M: 99.96%, T: 0.04%	M: 99.69%, T: 0.20%, I: 0.12%	T: 99.14%, M: 0.29%, A: 0.38%, S: 0.19%
	290	D	D: 99.96%, V: 0.02%, A: 0.02%	100%	100%	100%
	292	A	A: 34.34%, T: 0.02%, S: 65.43%, F: 0.17%, C: 0.02%, Y: 0.02%	A: 99.91%, S: 0.09%	100%	S: 82.98%, C: 16.63%, T: 0.19%, A: 0.19%
	542	L	100%	L: 99.96%, I: 0.04%	100%	100%
	543	P	100%	100%	99.88%, T: 0.08%, Q: 0.04%	100%
Bolsillo puerta helical	362	V	V: 99.98%, G: 0.02%	100%	100%	100%
	365	I	100%	I: 99.96%, L: 0.02%, T: 0.02%	100%	I: 99.90%, M: 0.10%
	369	N	N: 99.94%, S: 0.04%, H: 0.02%	N: 99.98%, K: 0.02%	100%	100%
	384	Q	100%	Q: 99.98%, P: 0.02%	100%	100%
	385	L	100%	L: 99.98%, Y: 0.02%	100%	L: 99.90%, M: 0.10%
	406	V	V: 99.90%, I: 0.08%, G: 0.02%	V: 99.93%, D: 0.02%, G: 0.02%, I: 0.02%	V: 99.96%, G: 0.04%	100%
	407	T	100%	T: 99.98%, P: 0.02%	100%	100%
	408	T	100%	100%	100%	100%
Bolsillo D3	570	N	100%	N: 99.96%, S: 0.04%	100%	100%

	571	Q	100%	100%	100%	100%
	572	I	V: 99.23%, I: 0.77%	I: 99.93%, V: 0.04%, F: 0.02%	I: 99.84%, V: 0.16%	100%
	593*	P	P: 99.94%, S: 0.02%, T: 0.02%, H: 0.02%	100%	P: 99.92%, T: 0.04%, Q: 0.04%	100%
	596*	L	L: 99.98%, P: 0.02%	100%	L: 99.96%, R: 0.04%	100%
	601*	Y	100%	Y: 99.98%, C: 0.02%	Y: 99.96%, K: 0.04%	100%
Bolsillo *loop* β6-β1’	190	I	I: 99.96%, V: 0.02%, T: 0.02%	I: 99.91%, V: 0.09%	I: 99.72%, M: 0.08%, V: 0.16%, T: 0.04%	I: 97.43%, V: 2.57%
	286	A	100%	A: 99.96%, F: 0.02%, H: 0.02%	100%	100%
	287	H	100%	H: 99.89%, P: 0.09%, D: 0.02%	100%	100%
	313	F	100%	100%	F: 99.96%, S: 0.04%	100%
	315	T	T: 99.98%, S: 0.02%	100%	T: 99.84%, S: 0.16%	100%
	318	P	100%	P: 99.98%, S: 0.02%	100%	100%
	319	P	P: 99.96%, Q: 0.02%, T: 0.02%	100%	100%	100%
	509	F	F: 99.88%, Y: 0.12%	F: 99.98%, Y: 0.02%	100%	F: 88.59%, Y: 11.41%

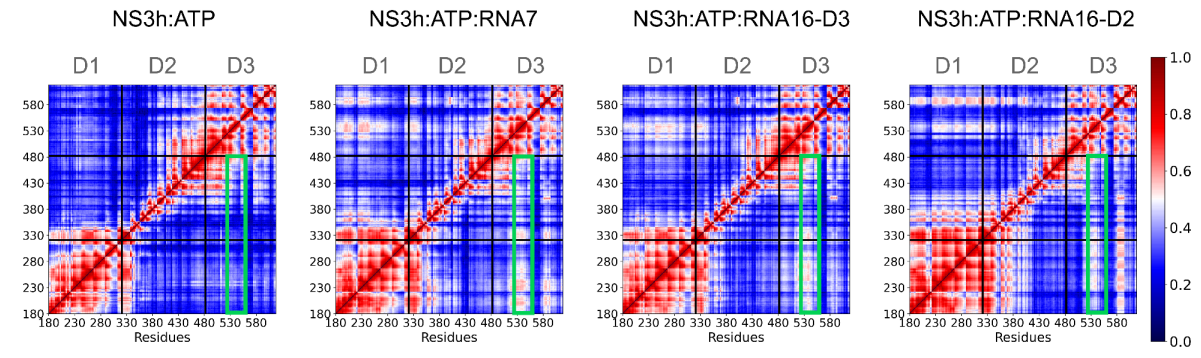


Figura 4A.4. Matrices de correlación generalizada para los estados considerados. El rectángulo verde indica la región hélice-giro-hélice de D3 para la cual se realizó el gráfico de densidad densidad kernel en la Figura 4.8.B.

Agradecimientos

Comienzo agradeciendo al Departamento de Química Inorgánica, Analítica y Química Física por darme un lugar en el departamento como estudiante de doctorado y como docente; a la Facultad de Ciencias Exactas y Naturales (FCEN) y a la Universidad de Buenos Aires (UBA), por ser pública, gratuita y de la más alta calidad; y al CONICET, por haberme permitido dedicarme principalmente a la investigación durante el doctorado.

Luego, le agradezco enormemente a Noosh, por ser una excelente directora, a quien admiro y considero una referente en la ciencia y en la vida por su capacidad intelectual y emocional. Es un verdadero privilegio transitar esta etapa con ella. Agrego acá mis agradecimiento al grupo de trabajo que lidera Noosh, a Nati, Fer y Jime, por compartir las ganas de aprender y entender a la NS3 del virus del dengue.

A continuación, un breve apartado para agradecerle a Santi Di Lella, por permitirme acceder a una beca doctoral, pero más aún por su calidad humana y su apoyo incondicional. Participé en un proyecto liderado por Santi que se centraba en el estudio de la proteína spike del virus SARS-CoV-2. Este trabajo fue realizado en colaboración con dos investigadores alemanes, Daniel Lauster y Christian Hackenberger. Quisiera destacar y agradecer además la participación de Sebastián Florez-Rueda, quien fue el encargado de realizar la mayoría de los experimentos. No solo publicamos los resultados de este trabajo del cual soy primera autora [1], sino que tuve la oportunidad de hacer una pasantía de tres meses en la Freie Universität Berlin, en Alemania, gracias a una beca del Servicio alemán de intercambio académico (DAAD, por sus siglas en alemán). Esta etapa contribuyó indiscutiblemente a mi formación científica, por lo que le agradezco a los colaboradores pero especialmente a Santi, por permitirme formar parte de este proyecto.

Siguiendo con agradecimiento dentro de la FCEN, le doy las gracias al Instituto del Cálculo por darme un espacio para trabajar y por la buena onda que irradia cada uno de sus miembros; en particular, a mis compañeras y compañeros de la oficina, con quienes convivimos tantas horas. Por otro lado, quisiera agradecerle a todo el E1 y en particular a Dario, Lula y Joni, por la ayuda incondicional, los seminarios compartidos y las discusiones científicas que contribuyeron enormemente a mi formación.

Por fuera de la FCEN, quisiera agradecerle a Sergio Kaufman, a Andrea Gamarnik y todo su grupo, especialmente a Santi Oviedo, por las colaboraciones establecidas en donde el intercambio estuvo guiado por el entusiasmo y la pasión por la investigación. Especialmente a Sergio, por haberme acompañado con tanta calidez desde mis primeros acercamientos a la investigación de la NS3, que iniciaron incluso antes del doctorado. Por otro lado, le agradezco a Giulia Palermo por permitirme visitar su laboratorio como estudiante de intercambio y participar en el congreso anual de la Biophysical Society. También le agradezco a los miembros de su grupo de investigación por su amabilidad y por colaborar con nuestro trabajo. Esta segunda pasantía la pude realizar gracias a la ayuda económica brindada por The Company of Biologists y la Universidad de Buenos Aires. Tuve la oportunidad también de participar en otro congreso internacional organizado por Chemical

[1] Sarto C, Florez-Rueda S, Arrar M, Hackenberger CPR, Lauster D, Di Lella S. Atomistic insight into the essential binding event of ACE2-derived peptides to the SARS-CoV-2 spike protein. Biol Chem. 2022;403: 615–624. doi:10.1515/hsz-2021-0426

Computing Group, gracias a su ayuda económica y la brindada por la IUPAB. Volviendo al terreno nacional, a los investigadores e investigadoras que forman parte de la Sociedad Argentina de Biofísica (SAB), les agradezco por generar el sentimiento de comunidad a lo largo y ancho del país y en particular al grupo de jóvenes biofísicos, por permitirme formar parte de la organización de las jornadas YIB. Estoy realmente agradecida con toda la comunidad científica nacional e internacional.

Quisiera incluir también mi agradecimiento a los jurados, Pablo, Mónica y María Ana, por haber aceptado formar parte de la evaluación de este trabajo.

Por último y no menos importante, le agradezco a toda mi familia, amigas y amigos por acompañarme durante esta etapa de tanto aprendizaje.

Publicaciones

- C Sarto, SB Kaufman, DA Estrin, M Arrar. Nucleotide-dependent dynamics of the Dengue NS3 helicase. *Biochimica et Biophysica Acta (BBA)-Proteins and Proteomics.* **1868** (8), 140441 (2020)

- NS Adler†, LA Cababie†, C Sarto, CN Cavasotto, LG Gebhard, DA Estrin AV Gamarnik, M Arrar, SB Kaufman. Insights into the product release mechanism of dengue virus NS3 helicase. *Nucleic Acids Research* **50** (12), 6968-6979 (2022)

- C Sarto, S Florez-Rueda, M Arrar, CPR Hackenberger, D Lauster, S Di Lella. Atomistic insight into the essential binding event of ACE2-derived peptides to the SARS-CoV-2 spike protein *Biological Chemistry* **403** (5-6), 615-624 (2022)

- JJ Incicco, LA Cababie, C Sarto, NS Adler, F Amrein, E Mikkelsen, M Arrar, SB Kaufman.Thermodynamic and mechanistic analysis of the functional properties of dengue virus NS3 helicase *Biophysical Reviews* **15** (4), 591-600 (2023)

- F Amrein†, C Sarto†, LA Cababie, FL Gonzalez Flecha, SB Kaufman, M Arrar. Impact of bound ssRNA length on allostery in the Dengue Virus NS3 helicase *Nucleic Acids Research* **51** (20), 11213-11224 (2023)