

**Centro de Investigación Científica y de Educación
Superior de Ensenada, Baja California**



**Maestría en Ciencias
en Ciencias de la Vida con orientación en Microbiología**

**Estudio *in silico* del genoma del virus de la hepatitis C para
buscar su interrupción viral mediante la identificación de
fármacos candidatos**

Tesis
para cubrir parcialmente los requisitos necesarios para obtener el grado de
Maestro en Ciencias

Presenta:

Víctor Hugo Vargas Bermúdez

Ensenada, Baja California, México
2017

Tesis defendida por
Víctor Hugo Vargas Bermúdez

y aprobada por el siguiente Comité

Dr. Carlos Alberto Brizuela Rodríguez
Co-director de tesis

Dr. Sergio Andrés Águila Puentes
Co-director de tesis

Dra. María Asunción Lago Lestón

Dr. Israel Marck Martínez Pérez



Dra. Clara Elizabeth Galindo Sánchez
Coordinador del Posgrado en Ciencias de la Vida

Dra. Rufina Hernández Martínez
Directora de Estudios de Posgrado

Resumen de la tesis que presenta **Víctor Hugo Vargas Bermúdez** como requisito parcial para la obtención del grado de Maestro en Ciencias en Ciencias de la Vida con orientación en Microbiología.

Estudio *in silico* del genoma del virus de la hepatitis C para buscar su interrupción viral mediante la identificación de fármacos candidatos

Resumen aprobado por:

Dr. Carlos Alberto Brizuela Rodríguez
Co-director de tesis

Dr. Sergio Andrés Águila Puentes
Co-director de tesis

La polimerasa NS5B del virus de la hepatitis C (HCV) se ha convertido en un blanco clave para la búsqueda de fármacos que puedan interferir con su actividad, inhibiendo de esta manera el ciclo de infección de HCV. La presente investigación tuvo por objetivo el desarrollo de un protocolo bioinformático para la identificación y prueba de fármacos que permitieran inhibir *in silico* la polimerasa NS5B de HCV. La metodología desarrollada consistió en la comparación de distintas secuencias proteínicas para buscar aminoácidos conservados en el sitio catalítico, ensayos de acoplamiento de diferentes librerías de ligandos sobre la polimerasa y la simulación molecular de distintos complejos con ligandos candidatos. Los diferentes ensayos permitieron encontrar cuatro ligandos que mostraron interferir con la actividad de la proteína en igual o mayor grado que los ligandos de referencia sofosbuvir y canabidiol, de los que se ha comprobado su efecto inhibitorio sobre ésta. Además de los resultados obtenidos, se establecieron las bases para desarrollar estudios posteriores que consistan en el diseño de otros fármacos y la posibilidad de probar *in vitro* los fármacos definidos durante el estudio.

Palabras clave: virus de la hepatitis C, polimerasa NS5B, acoplamiento, cribado virtual

Abstract of the thesis presented by Víctor Hugo Vargas Bermúdez as a partial requirement to obtain the Master of Science degree in Life Sciences with orientation in Microbiology

Bioinformatic study of hepatitis C virus genome in search of its viral interruption through the identification of candidates drugs

Abstract approved by:

Dr. Carlos Alberto Brizuela Rodríguez

Thesis Co-director

Dr. Sergio Andrés Águila Puentes

Thesis Co-director

Hepatitis C virus (HCV) polymerase NS5B has become the main target for the search of potential drugs that can interfere with its activity and inhibit its infection cycle. The objective of the present work is to develop a bioinformatic pipeline for the identification and screening of different lead compounds that could inhibit the activity of the NS5B polymerase. The experimental design consisted of the following steps: the comparison of different protein sequences in search of conserved amino acids in their catalytic site; the analysis of docking assays with different lead compound libraries against NS5B; and the molecular simulation of distinct complexes with candidate ligands. The execution of the different assays herein mentioned, resulted in the discovery of four ligands that demonstrated to interfere equally or even better with protein activity than both sofosbuvir and cannabidiol compounds; two compounds with proven inhibitory effects on NS5B polymerase. Furthermore, these results pave the way to continue research oriented to drug design and *in vitro* assays with the molecules identified in this study.

Keywords: hepatitis C virus, NS5B polymerase, docking, virtual screening

Dedicatoria

A mis padres, mis hermanas y mi hermano. Muchas gracias por todo.

Las palabras nunca serán suficientes, saben que los adoro.

Agradecimientos

Gracias al CICESE por permitirme pertenecer a la Institución durante estos dos años y al CONACyT (No. de becario 589177) por el financiamiento otorgado.

A mis directores de tesis, el Dr. Sergio A. Águila Puentes y el Dr. Carlos A. Brizuela Rodríguez, les estaré infinitamente agradecido por el apoyo brindado en todo momento, especialmente en las épocas más críticas. Gracias por darme las herramientas necesarias para avanzar a buen paso en la ejecución de este proyecto, por los consejos brindados, por la libertad que me otorgaron en la toma de decisiones durante algunas etapas del trabajo y por el tiempo invertido en mí. Gracias también por los recursos económicos que destinaron para que incrementara mi aprendizaje.

Agradezco a los miembros de mi comité, la Dra. Asunción Lago y el Dr. Israel Pérez, por sus observaciones, por la buena disposición que tuvieron en el avance de todo el trabajo y por su tiempo.

Gracias al Dr. Joel Alderete (Universidad de Concepción), a la Dra. Verónica Jiménez (Universidad Andrés Bello) y a sus grupos de trabajo por todo el apoyo, enseñanzas y las infinitas atenciones que me brindaron. Me hicieron sentir como en casa.

Joel Ricci, fuiste una parte muy importante en el desarrollo de este trabajo, gracias por esas pláticas de retroalimentación y por tu ayuda en la solución de problemas.

Gracias a mis compañeros de Ciencias de la Vida y a los grupos de trabajo de Ciencias de la Computación por los consejos y conocimientos compartidos. Sus aportaciones enriquecieron en todo momento la culminación del proyecto.

Tabla de contenido

Resumen en español	ii
Resumen en inglés	iii
Dedicatoria	iv
Agradecimientos	v
Lista de figuras	ix
Lista de tablas.....	x
Capítulo 1. Introducción	1
1.1 Antecedentes	1
1.1.1 Prevalencia e incidencia del virus de la hepatitis C (HCV)	1
1.1.2 Clasificación de HCV y su genoma	3
1.1.3 Partícula del HCV.....	3
1.1.4 Ciclo de infección de HCV	3
1.1.4.1 Transmisión y tropismo celular.....	3
1.1.4.2 Mecanismo de ingreso del virión al hospedero.....	4
1.1.4.3 Replicación de HCV	6
1.1.5 Proteínas traducidas del genoma de HCV.....	7
1.1.5.1 Procesamiento de la poliproteína.....	7
1.1.5.2 Proteínas estructurales	7
1.1.5.3 Viroporina	8
1.1.5.4 Proteínas no estructurales (NS)	8
1.1.6 Estructuras cristalográficas de HCV	9
1.1.7 Inhibidores de la replicación de HCV	10
1.1.8 Aplicaciones del descubrimiento de fármacos asistido por computadora (CADD)	12
1.2 Justificación.....	13
1.3 Hipótesis.....	14
1.4 Objetivos	14

1.4.1	Objetivo general	14
1.4.2	Objetivos específicos.....	14
Capítulo 2.	Metodología	15
2.1	Selección y preparación de la estructura de la polimerasa viral NS5B.....	15
2.1.1	Obtención de la estructura cristalográfica de la polimerasa viral NS5B.....	15
2.1.2	Modelado homólogo.....	15
2.2	Alineamiento y búsqueda de regiones conservadas	15
2.2.1	Búsqueda de las regiones conservadas a nivel estructural de NS5B	15
2.2.2	Búsqueda de las regiones conservadas a nivel secuencia de NS5B.....	16
2.2.3	Identificación de aminoácidos de interés dentro de las regiones conservadas	16
2.3	Selección y preparación de ligandos para ensayos de inhibición de la polimerasa	17
2.3.1	Generación de librerías de ligandos a partir de base de datos ZINC15	17
2.3.2	Preparación de librería de ligandos	18
2.4	Ensayos de cribado virtual polimerasa-ligandos.....	18
2.4.1	Delimitación de región en la polimerasa para realizar el cribado virtual	18
2.4.2	Cribado virtual.....	18
2.4.3	Selección de ligandos candidatos para ensayar simulaciones moleculares acopladas a la polimerasa	19
2.5	Simulaciones moleculares acopladas del complejo polimerasa-ligando.....	19
2.5.1	Preparación de los ligandos	19
2.5.2	Preparación del complejo	21
2.5.3	Simulaciones moleculares.....	21
2.5.4	Análisis de enlaces de hidrógeno	21
2.6	Ensayos de validación MM-GBSA.....	22
2.6.1	Validación de la simulación acoplada con un modelo de solvente implícito.....	22

Capítulo 3. Resultados	23
3.1 Corrección obtenida por la aplicación del modelado homólogo	23
3.2 Análisis de la conservación de regiones y aminoácidos en NS5B	24
3.3 Selección de los ligandos sometidos al cribado virtual	26
3.4 Evaluación de la estabilidad de la polimerasa con cada ligando acoplado.....	28
3.5 Evaluación de la interacción de aminoácidos con los diferentes ligandos	39
3.6 Verificación de energía libre con MM-GBSA.....	40
Capítulo 4. Discusión	41
Capítulo 5. Conclusiones	44
Literatura citada	45

Lista de figuras

Figura 1. Comparación de las principales causas de muerte a nivel mundial en los años 1990 y 2013.....	2
Figura 2. Factores involucrados durante el ingreso del virión hacia los hepatocitos.....	6
Figura 3. Representación de la polimerasa NS5B de HCV: su estructura asemeja a un puño, típico de las polimerasas virales	10
Figura 4. Ligandos seleccionados para simular acoplados a la polimerasa y su identificador de residuo. .	20
Figura 5. Comparativo del modelado homólogo.	23
Figura 6. Análisis de la secuencia completa de aminoácidos de la polimerasa.	24
Figura 7. Logos generados para la verificación de aminoácidos conservados.	25
Figura 8. Gráfico de ligandos que presentaron una energía superior a los ligandos de referencia.	26
Figura 9. Resultado de la simulación NS5B-SOF que contiene el ligando de referencia SOF	29
Figura 10. Resultado de la simulación del complejo NS5B-ESM.....	30
Figura 11. Resultado de la simulación del complejo NS5B-SSM.....	31
Figura 12. Resultado de la simulación del complejo NS5B-FSM.....	32
Figura 13. Resultado de la simulación del complejo NS5B-ASM.	33
Figura 14. Resultado de la simulación NS5B-CAN que contiene el ligando de referencia CAN.....	35
Figura 15. Resultado de la simulación del complejo NS5B-CAA.	36
Figura 16. Resultado de la simulación del complejo NS5B-CAB..	37
Figura 17. Resultado de la simulación NS5B-CTH.	38

Lista de tablas

Tabla 1. Compuestos utilizados para inhibir HCV sobre diferentes blancos.	12
Tabla 2. Agrupación inicial de librerías de ligandos.....	17
Tabla 3. Ligandos elegidos para ser acoplados a la polimerasa.....	27
Tabla 4. Ligandos y su porcentaje de interacción con diferentes aminoácidos de la polimerasa.....	39
Tabla 5. Ligandos acoplados a la polimerasa con un solvente implícito.	40

Capítulo 1. Introducción

1.1 Antecedentes

1.1.1 Prevalencia e incidencia del virus de la hepatitis C (HCV)

En general, a nivel mundial, la hepatitis viral –HAV, HBV, HCV y HEV– pasó de ser la décima causa de muerte en 1990 a la séptima en el año 2013 (Figura 1). Esto la ha convertido en una de las principales causas de decesos a nivel mundial comparado con la cantidad de muertes ocasionadas por otras enfermedades como tuberculosis, malaria, diabetes y VIH, entre otras. De manera más específica, para cuantificar la epidemiología global de HCV se han utilizado datos disponibles desde el año 2000 al año 2015, de un total de 138 países –cerca del 90% de la población mundial–. La prevalencia total a nivel global del HCV está estimada en 2.5% –177.5 millones de adultos infectados–. Una característica notable de la infección por HCV es su tendencia a volverse crónica: alrededor del 70% de las infecciones agudas se vuelven persistentes y estos casos crónicos conllevan a serios fallos renales (Stanaway *et al.*, 2016; Chisari, 2005; Messina *et al.*, 2015).

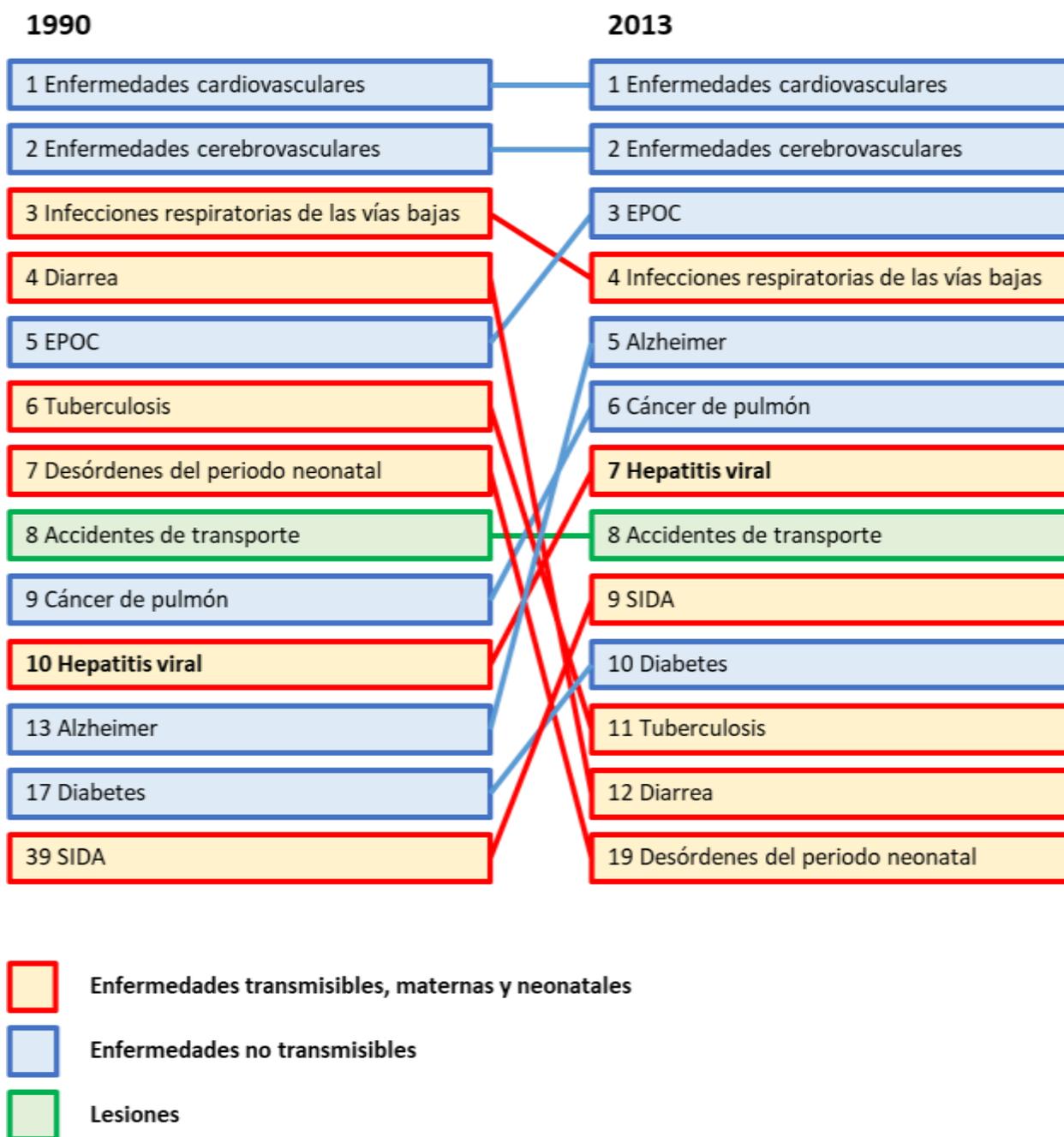


Figura 1. Comparación de las principales causas de muerte a nivel mundial en los años 1990 y 2013 (modificado de Stanaway *et al.*, 2016).

El genotipo 1 de HCV es el más prevalente a nivel mundial con el 49.1% de los casos, seguido del genotipo 3 con 17.9%, el genotipo 4 comprende el 16.8% y el genotipo 2 el 11%, los genotipos 5 y 6 corresponden a menos del 6% de los casos restantes. Mientras los genotipos 1 y 3 son comunes en todo el mundo, la más grande proporción de los genotipos 4 y 5 se encuentra en países de bajos ingresos (Petruzzello *et al.*, 2016). En México, estudios epidemiológicos y dos encuestas nacionales de salud señalan una prevalencia que ha aumentado del 0.7% en 2005 al 1.4% en 2011. No existen estudios directos sobre la incidencia de la enfermedad, sin embargo, se estima que aparecen 19,300 casos nuevos por año (Vera de León *et al.*, 2005; Kershenobich, 2011).

1.1.2 Clasificación de HCV y su genoma

El HCV está clasificado en el género *Hepacivirus* y forma parte de la familia Flaviviridae. Es un virus de RNA de cadena positiva, envuelto; con un genoma de alrededor de 9.6 kb compuesto de una región no codificante (NCR) 5' la cual incluye un sitio de entrada al ribosoma (IRES), un marco de lectura abierto (ORF) que codifica para las proteínas estructurales y no estructurales, y una NCR-3' (Moradpour *et al.*, 2007).

1.1.3 Partícula del HCV

Los componentes virales conocidos del virión de HCV son tres: el RNA genómico, la proteína core y las glicoproteínas de la envoltura (E1, E2). Estas glicoproteínas se cree que forman un heterodímero no-covalente, estabilizado por puentes disulfuro; siendo partícipes en el mecanismo de internalización de HCV hacia los hepatocitos a través de endocitosis mediada por receptor (Paul *et al.*, 2014).

1.1.4 Ciclo de infección de HCV

1.1.4.1 Transmisión y tropismo celular

La transmisión más eficiente de HCV se da por la vía parenteral: transfusiones de sangre, hemodiálisis, tatuajes, malas prácticas donde se compartan agujas y lesiones con instrumentos punzocortantes. Sin embargo, un gran número de personas –hasta 75%–, quienes no están sometidas a factores obvios de riesgo, son infectados con el virus; en estas personas la transmisión pudo haberse dado

por otras vías: perinatal, intrafamiliar y por transmisión sexual (Ackerman *et al.*, 1998). De este porcentaje, entre el 3-5% de las transmisiones ocurre de manera vertical, es decir, a través de madres infectadas con HCV a los recién nacidos (Nawaz *et al.*, 2015).

El virus de la hepatitis C es hepatotrópico y no-citopático; causante de la hepatitis aguda y crónica, y del carcinoma hepatocelular. El hígado es el órgano blanco primario y el hepatocito su célula blanco primaria. La infección aguda es usualmente asintomática lo que hace difícil el diagnóstico temprano. El HCV puede generar infección crónica en humanos y chimpancés, solamente. Otro tipo de especies, como los roedores, son resistentes a la infección. Debido a que este tropismo es altamente restringido, no ha sido completamente entendido; incluso con varios estudios desarrollados a la fecha, aún no se encuentra bien definido en etapas tempranas de la entrada viral hacia las células. La proteína tetraspanina CD81 y la proteína ocludina (OCLN) constituyen el mínimo grupo de factores específico de humanos necesarios para la asimilación viral en células de ratón o hámster. Sin embargo, en estudios previos, los ratones no han mostrado susceptibilidad incluso con la expresión transgénica de CD81 humanas en ellos. Estas observaciones podrían ser parcialmente explicadas por diferencias, entre humanos y otras especies, del segundo “loop” extracelular, tanto de CD81 como de OCLN, el cual se une al dímero de la glicoproteína viral E2 para permitir la interacción y entrada del HCV (Flint *et al.*, 2006; Michta *et al.*, 2010; Ploss *et al.*, 2009).

A la fecha, la variada susceptibilidad de células no humanas a la entrada de HCV es representada en gran parte por diferencias en residuos críticos en ortólogos de CD81, el receptor removedor clase-B tipo-1 (SCARB1), las proteínas de unión estrecha claudina-1 (CLDN1) y OCLN o por insuficiente expresión de estas proteínas. Aún así, se debe determinar la contribución de otros factores de entrada a los hepatocitos para el tropismo de las especies de HCV (Ding *et al.*, 2014).

1.1.4.2 Mecanismo de ingreso del virión al hospedero

El ingreso del HCV hacia los hepatocitos es un proceso que involucra a las glicoproteínas que forman parte del virión, a las lipoproteínas del hospedero asociadas al virión y a un gran número de factores asociados al hospedero (Figura 2). Muchos de estos factores probablemente contribuyen al hepatotropismo y a la limitada existencia de células hospederas de este virus. Durante el inicio de la infección, las partículas lipovirales del HCV se unen a los glicosaminoglicanos presentes en heparan sulfato proteoglicanos (HSPG), receptor de lipoproteínas de baja densidad (LDLR9), y SCARB1. La posterior interacción con SCARB1 induce cambios conformacionales en E2, llevando a la unión de E2 con CD81. La

unión a CD81 activa rutas de señalización vía receptor factor de crecimiento epidermal (EGFR) tirosina quinasa, Ras, y Rho GTPasa, desencadenando la difusión membranal de HCV. Existen otras proteínas como el receptor transferrina 1 (TfR1), el transportador de colesterol Niemann-Pick 1 tipo-C1 (NPC1L1), CLDN1, y OCLN que están involucradas en el mecanismo post-unión HCV-CD81, pero sus roles exactos no están bien definidos aún. La interacción del complejo E2-CD81/CLDN1 inicia la endocitosis mediada por clatrina. Finalmente, promovido por una interacción con el efector b tipo-DFFA inductor de la muerte celular (CIDEB), la envoltura viral y la membrana endosomal se someten a fusión en un ambiente a bajo pH. Seguido a la fusión, el RNA genómico de HCV es liberado de la nucleocápside viral al citosol, donde es traducido dando inicio a los subsecuentes pasos en el ciclo de vida de HCV (Ding *et al.*, 2014).

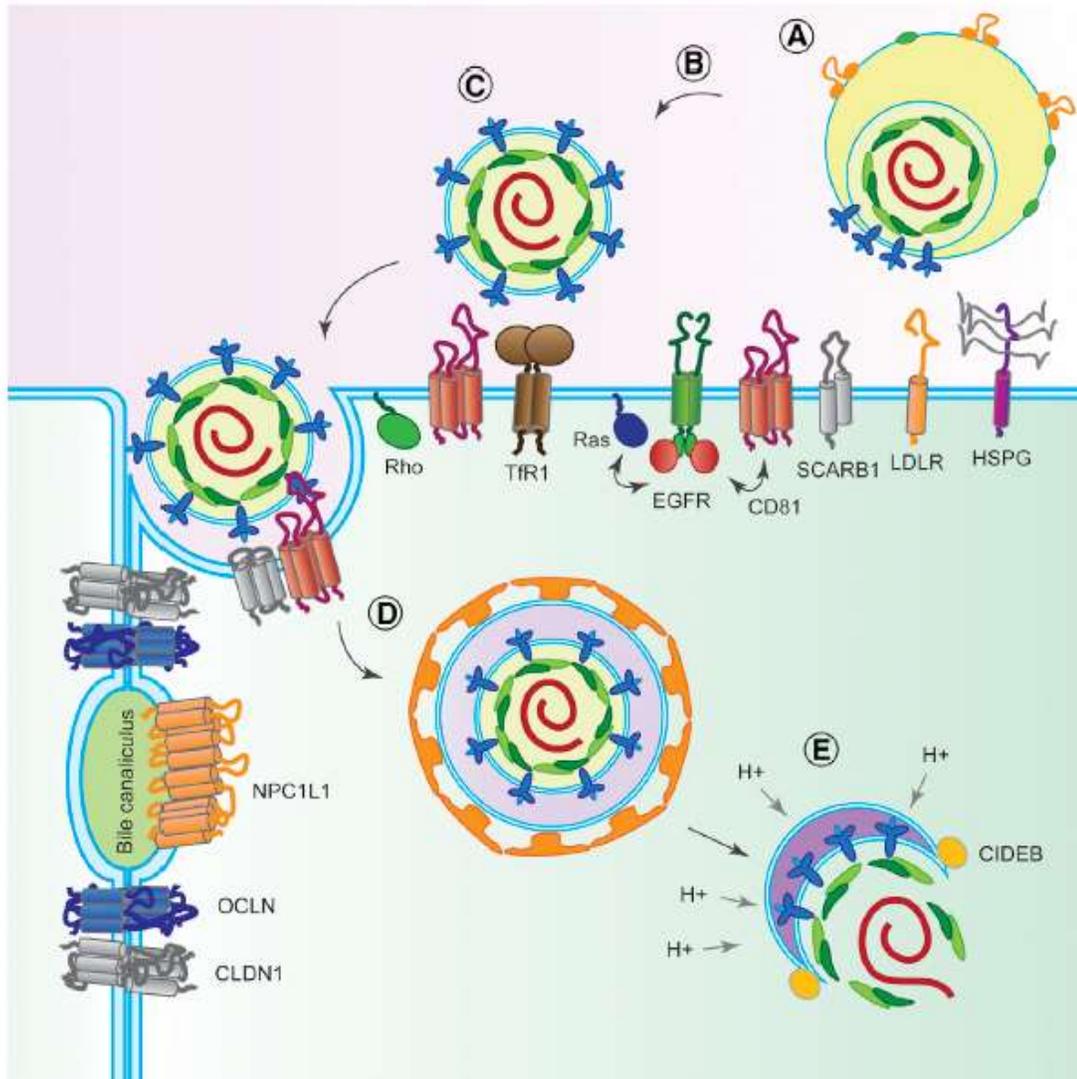


Figura 2. Factores involucrados durante el ingreso del virión hacia los hepatocitos. A) Unión de partículas lipovirales de HCV a HSPG, LDLR y SCARB1; B) la interacción con SCARB1 induce cambios conformacionales en HCV-E2, llevando a la consecuente unión de E2 con CD81; C) la unión a CD81 activa vías de señalización mediadas por EGFR, Ras y Rho GTPasas para iniciar la difusión a través de la membrana del HCV; D) la interacción del complejo E2-CD81/CLDN1 inicia la endocitosis mediada por clatrina; E) promovida por la interacción con CIDEB, la envoltura viral y la membrana endosomal se someten a fusión en un ambiente a bajo pH con la consecuente liberación del genoma del HCV (Ding *et al.*, 2014).

1.1.4.3 Replicación de HCV

Después de que el genoma es liberado, se traduce en una sola poliproteína para posteriormente ser procesado por las proteasas celulares y virales. Para llevar a cabo la replicación de RNA viral se tiene que dar la formación de la red membranosa, la cual es una alteración de la membrana del retículo endoplasmático (ER, por sus siglas en inglés) que es inducida por la proteína NS4A. Sin encontrarse bien definido el mecanismo, estudios han revelado que existe una compleja interacción entre la replicación de

RNA del HCV y el metabolismo de lípidos en la célula, esto puede deberse al tráfico y asociación de proteínas virales y del hospedero con membranas intracelulares sugiriendo que la fluidez de la membrana es importante para el funcionamiento del complejo de replicación (Moradpour *et al.*, 2007; Kapadia y Chisari, 2005). Poco se sabe sobre las etapas finales, el ensamblaje y liberación del virión, del ciclo de vida del HCV.

1.1.5 Proteínas traducidas del genoma de HCV

1.1.5.1 Procesamiento de la poliproteína

La poliproteína de HCV es procesada por proteasas codificadas por el virus y proteasas propias del hospedero dando lugar a diez proteínas maduras: las proteínas estructurales, las cuales incluyen la proteína “core” y las glicoproteínas E1 y E2; la viroporina p7; y las proteínas no estructurales NS2, NS3, NS4A, NS4B, NS5A y NS5B. Una peptidasa péptido-sígnal lleva a cabo la maduración de las proteínas estructurales mediante la escisión entre C/E1, E1/E2 y E2/p7; así como el corte entre la unión de p7/NS2. Con respecto a las proteínas no estructurales, el procesamiento proteolítico es realizado por la acción de dos enzimas virales: la autoproteasa NS2 que escinde la unión de NS2/3, y la serina proteasa NS3/4A que procesa el resto de las proteínas no estructurales (Ding *et al.*, 2014; Lindenbach y Rice, 2005). HCV también codifica una pequeña proteína llamada F –“frame shift”– o ARFP –proteína de lectura de marco alterno–, que puede producir una variación generando un marco de lectura diferente dentro del gen “core” (Xu *et al.*, 2003).

1.1.5.2 Proteínas estructurales

La proteína “core” es la primera proteína estructural codificada por el ORF del HCV y forma la nucleocápside viral. La proteína madura tiene una masa de 21 kDa y consta de 173-179 aminoácidos. Su residuo hidrofílico N-terminal contiene una alta proporción de residuos aminoácidos básicos que participan en su unión al RNA y la homo-oligomerización; y su asociación con adiposomas, mediada por el dominio central relativamente hidrofóbico, podría tener un rol durante la morfogénesis del virión (Boulant *et al.*, 2006). El complejo no covalente que forman las proteínas glicosiladas E1 y E2 es el responsable de la formación de la envoltura viral. La maduración y el plegamiento de estas glicoproteínas es un proceso complejo que involucra a la maquinaria chaperona del ER y depende de la formación de enlaces disulfuro, así como la glicosilación. Sus dominios transmembranales en C-terminal están involucrados en la

heterodimerización y funcionan como una péptido señal interna para la reorientación de la partícula viral (Moradpour *et al.*, 2007).

1.1.5.3 Viroporina

La viroporina p7 es un polipéptido de 67 aminoácidos que se mantiene parcialmente unido a E2, forma oligómeros y parece funcionar como un canal iónico codificado por virus. Se ha reportado que no se requiere para la replicación de RNA *in vitro* pero es necesaria para la infección *in vivo*. Se ha sugerido que al pertenecer a la familia de las viroporinas podría tener un papel importante en la maduración de la partícula viral y su liberación, por lo que es un blanco atractivo para el desarrollo de un antiviral (Lindenbach y Rice, 2005; Moradpour *et al.*, 2007).

1.1.5.4 Proteínas no estructurales (NS)

La autoproteasa NS2-3 posee su actividad proteolítica en C-terminal de NS2 y N-terminal de NS3 y se ha demostrado que los aminoácidos His143, Glu163 y Cys184 son esenciales para su actividad. Como todas las proteínas del HCV, NS2 está asociada con membranas intracelulares y contiene al menos un segmento transmembranal. El complejo NS3-4A es una proteína multifuncional con una RNA helicasa/NTPasa localizada en C-terminal y una serina proteasa localizada en la parte N-terminal. Su región catalítica es formada por His57, Asp81 y Ser139. NS4A funciona como un cofactor para NS3 y su extremo N-terminal es responsable de la asociación membranal del complejo. NS3/4A escinde la proteína precursora en cuatro diferentes posiciones, separándola en varias proteínas no estructurales necesarias para la replicación (Götte y Feld, 2016). La proteína NS4B está poco caracterizada pero una de sus funciones conocidas es inducir la formación de una red membranosa. Esta red es una alteración de la membrana que consiste de vesículas en una matriz membranosa, y sirve como un ambiente para la expresión del complejo de replicación del HCV (Egger *et al.*, 2002). NS5A es una fosfoproteína que puede ser encontrada en formas fosforilada –56kDa– e hiperfosforilada –58kDa–. El estado de fosforilación podría modular la eficiencia de la replicación del RNA viral. Estudios funcionales y bioquímicos sugieren que la isoforma- α de la proteína cinasa CKI está involucrada en este proceso de hiperfosforilación. No se conoce si otras cinasas celulares se encuentran involucradas en generar las diferentes fosfoformas de NS5A. La proteína NS5B es la polimerasa de RNA dependiente de RNA viral (RdRp) y es responsable de la síntesis de la cadena negativa-complementaria de RNA y la subsecuente cadena positiva de RNA. La asociación de esta proteína con la membrana es mediada por un residuo de 21 aminoácidos en el extremo

C-terminal. Esta asociación no es necesaria para su actividad polimerasa *in vitro* pero es indispensable para su replicación en las células (Bressanelli *et al.*, 2002; Moradpour *et al.*, 2007).

1.1.6 Estructuras cristalográficas de HCV

Las estructuras cristalinas formando complejos con compuestos candidatos han facilitado los procesos iterativos de síntesis química, valoración biológica, evaluación estructural y predicciones basadas en modelos que han guiado al diseño de inhibidores que presentan mejoras continuas. La limitada información estructural que se encuentra actualmente disponible está basada en proteínas NS de HCV y la proteína E2. La E2 es considerada una proteína de fusión viral clase II, aunque posee la particularidad de no presentar cambios estructurales significativos al exponerse a bajo pH. Contiene una estructura con dominio globular y compacto que consiste, principalmente, de dos hebras- β arregladas como dos hojas perpendiculares que se mantienen unidas por un amplio núcleo hidrofóbico y enlaces disulfuro con una cola variable con dos pequeñas hélices- α (Khan *et al.*, 2014).

Las estructuras cristalográficas de NS3 con y sin el cofactor NS4A revelan un surco de poca profundidad entre los dominios N-terminal y C-terminal que pueden acoplar el sustrato, además del hecho de poseer un plegado similar a la quimiotripsina (Götte y Feld, 2016). La NS5A consiste en una hélice anfipática N-terminal y tres distintos dominios de los cuales sólo el dominio I ha sido cristalizado en diferentes conformaciones diméricas. La estructura cristalina de NS5B revela un plegamiento terciario típico de polimerasas virales monoméricas; conteniendo una estructura tipo puño (Figura 3) que previene la unión de la cadena de RNA en síntesis, en una doble cadena y con elementos flexibles que pueden existir en conformaciones abiertas y cerradas para acomodar la nueva cadena (Appleby *et al.*, 2015).

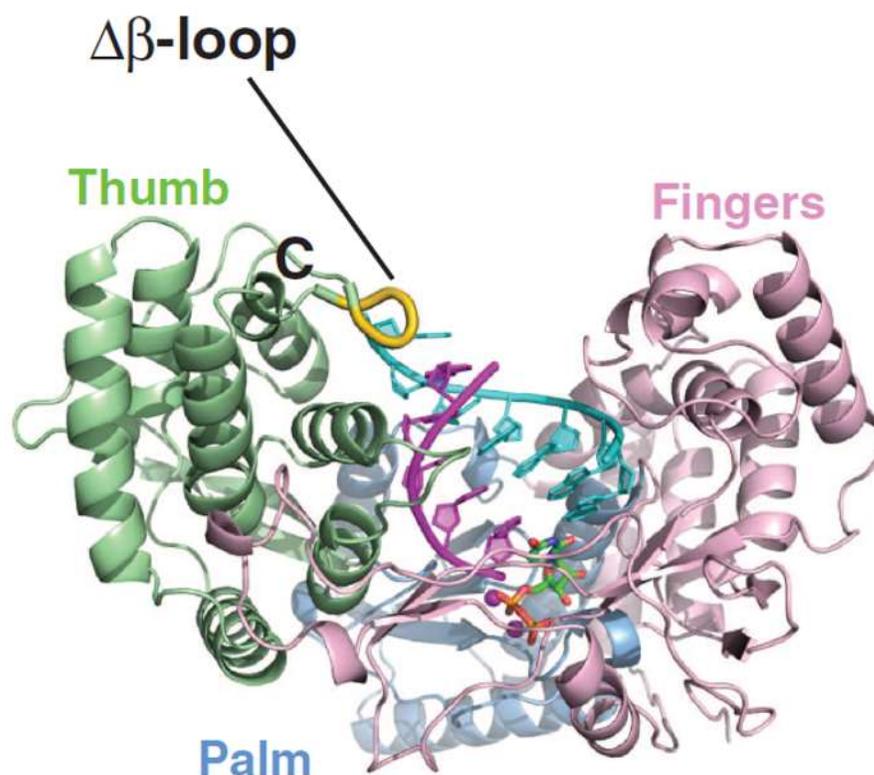


Figura 3. Representación de la polimerasa NS5B de HCV: su estructura asemeja a un puño, típico de las polimerasas virales (Appleby *et al.*, 2015).

1.1.7 Inhibidores de la replicación de HCV

A la fecha, la inhibición de la replicación de HCV se ha centrado en tres objetivos principales: la proteasa NS3/4A, la RdRp NS5B y la proteína NS5A. Ciluprevir fue el primer inhibidor de la proteasa (PI) de HCV que fue llevado hasta pruebas clínicas, aunque su desarrollo fue eventualmente detenido debido a los efectos tóxicos que fueron presentados en animales. Telaprevir y boceprevir, otros compuestos inhibidores de NS3, representaron la primera generación de agentes antivirales de acción directa (DAAs) usados con éxito en pruebas clínicas. En contraste con ciluprevir, éstas son estructuras lineales peptidomiméticas que reaccionan reversiblemente con la serina catalítica para formar un enlace covalente; también presentan algunas limitaciones como la reducida especificidad genotípica y baja barrera a la resistencia. Existe una segunda generación de PIs, como grazoprevir, que mantiene su actividad contra la mutación R155K, a diferencia de PIs de primera clase que presentan resistencia contra sustitución de aminoácidos –Arg155, Ala156 y Asp168– afectando su unión a la proteína (Kwong *et al.*, 2011; Hazuda *et al.*, 2013; Götte y Feld, 2016).

Con respecto a la proteína NS5A, el limitado conocimiento de sus propiedades moleculares y su relación con funciones biológicas han obstaculizado el diseño de drogas que pueden ser dirigidas a ésta. Sin embargo, se ha logrado el desarrollo de inhibidores como daclatasvir, ledipasvir y ombitasvir los cuales han sido aprobados para uso clínico. Las mutaciones que confieren resistencia son las que se encuentran en Leu31 y Tyr93 del dominio I, genotipo 1. La búsqueda de inhibidores para NS5A toma importancia debido a que esta proteína se encuentra involucrada en varias etapas del ciclo viral (Feld, 2014).

Los inhibidores de NS5B pueden ser clasificados en inhibidores análogos de no-nucleótidos (NNIs), que interfieren con la dinámica conformacional durante la transición de la iniciación a la elongación; e inhibidores análogos de nucleótidos (NIs), que compiten con la incorporación y unión de nucleósidos trifosfato actuando en distintas etapas de la síntesis de RNA. En contraste a los NNIs, los NIs comúnmente muestran actividad pangénica con gran barrera a la resistencia. Con respecto a los NNIs, los benzimidazoles como el beclabuvir, han demostrado competir con las regiones de “loop” –“thumb” sitio 1–, aunque existen mutaciones que confieren resistencia principalmente en las posiciones Pro495, Pro496 y Val499. Otros inhibidores basados en tiofeno han mostrado actividad en regiones diferentes –thumb sitio 2– contra mutaciones como Met423Thr. Las benzotiadiazinas y el dasabuvir podrían afectar la síntesis de RNA en su etapa inicial, aunque también se ha presentado resistencia contra mutaciones en las posiciones Cys316, Met441 y Ser556. Los NIs son administrados como profármacos¹ que requieren la activación metabólica a la forma trifosfato que eventualmente se posicionará en el sitio de unión del nucleótido de NS5B; el fosforamidato sofosbuvir, es el único de su clase –análogo de uridina– que ha sido aprobado (Götte y Feld, 2016).

Existen ensayos limitados sobre compuestos que tengan efecto sobre otros tipos de proteínas y etapas del ciclo de infección de HCV. Por ejemplo: utilizando derivados de iminoazúcares, se ha reportado una inhibición dependiente de la concentración, sobre la actividad del canal de la proteína p7 (Pavlovich *et al.*, 2003); por otro lado, se ha demostrado que compuestos, como el flavonoide quercetina, poseen propiedades anti-HCV al decrementar la replicación del genoma viral, así como una afectación sobre la integridad del virión (Rojas *et al.*, 2016). Un compuesto natural de la familia butenolida ha demostrado tener efectos positivos sobre las etapas tempranas de infección, pudiendo ser un compuesto valioso como inhibidor de la entrada de HCV (Chung *et al.*, 2016). También se han encontrado moléculas, como miR-122, que sumadas a otros DAAs, logran un efecto sinérgico potenciando de esta manera los efectos antivirales (Liu *et al.*, 2016). Finalmente, se ha publicado de manera reciente que el compuesto canabidiol

¹ Profármacos: sustancias que se administran en forma inactiva y requieren de su activación metabólica para la conversión a una molécula terapéuticamente activa.

ha inhibido la replicación del HCV hasta en un 86.4% (Lowe *et al.*, 2017). Un resumen de los diferentes compuestos que se han utilizado para inhibir el HCV sobre diferentes blancos se presenta en la Tabla 1.

Tabla 1. Compuestos utilizados para inhibir HCV sobre diferentes blancos.

Tipo de proteína que inhibe	Proteína	Clase de inhibidor	Ejemplos
-	Virión	-	Quercetina Butenolida
Estructural	p7	-	Iminoazúcares
No-Estructural	NS3	PI	Ciluprevir Telaprevir Boceprevir Grazoprevir*
			NS5A
	NS5B	NNI	
			NI

PI: inhibidor de la proteasa; NNI: inhibidor análogo de no-nucleótido; NI: inhibidor análogo de nucleótido. *Inhibidor de segunda generación; **Posee un efecto sinergista junto a sofosbuvir; ***Único tratamiento pangotípico desarrollado.

1.1.8 Aplicaciones del descubrimiento de fármacos asistido por computadora (CADD)

La mayoría de los efectos de los medicamentos son basados en la interacción entre compuestos químicos terapéuticos –fármacos– y proteínas –blancos–. Considerando que el genoma humano genera un estimado de 500,000 proteínas –sólo el 2% caracterizadas cristalográficamente– y que existen bases de datos con alrededor de 100 millones de compuestos disponibles comercialmente, el diseño y descubrimiento de fármacos por métodos convencionales resulta un proceso complejo, costoso y tardado, el cual puede llegar a generar un gasto de alrededor de mil millones de dólares en un periodo de 15 años. Por esto, las metodologías computacionales se han vuelto un componente crucial para el descubrimiento de fármacos (Vu *et al.*, 2015; Sterling e Irwin, 2015; Prada-Gracia *et al.*, 2016).

Existen diferentes casos de éxito logrados con la aplicación de estos métodos. El primero de ellos reportado fue el diseño de captopril, utilizado como inhibidor de la enzima convertidora de angiotensina –ACE– para el tratamiento de la hipertensión; la información necesaria para su diseño consistió en el conocimiento de los mecanismos enzimáticos de ACE y del diseño del fármaco basado en ligando y basado en estructura (Cushman *et al.*, 1978). Otro ejemplo consistió en el desarrollo de zanamivir que es un inhibidor de la neuroaminidasa utilizado en el tratamiento y profilaxis de la influenza causada por los virus de la influenza A e influenza B; la resolución de la estructura de la proteína por cristalografía de rayos X permitió elucidar la topología del sitio activo para posteriormente prevenir la infección continua del virus de una célula hospedera a otra (Elliot, 2001). Cuando en el año 2003 ocurrió en China un brote del síndrome respiratorio agudo severo –SARS–, un grupo de investigadores identificó por medio de cribado virtual que la cinanserina, un viejo fármaco antagonista de la serotonina, tenía efectos inhibitorios sobre la proteasa tipo 3C del coronavirus; las subsecuentes pruebas experimentales demostrarían detener la infección y propagación por el virus SARS (Chen *et al.*, 2005).

1.2 Justificación

El virus de la hepatitis C –HCV– es una de las principales causas de muerte en el mundo, las cifras de decesos son equiparables a las registradas por enfermedades como la tuberculosis y diabetes. Alrededor del 2.5% de la población mundial se encuentra infectada por el HCV, que en sus etapas tempranas llega a ser asintomático pero puede alcanzar complicaciones serias como el desarrollo de hepatitis crónica y carcinoma hepatocelular. Los problemas para el desarrollo de un tratamiento efectivo que inhiba el ciclo de infección del HCV son variados; además de la alta tasa de mutación que presentan los virus, los únicos hospederos donde el HCV puede realizar su ciclo de infección completo son los humanos y los chimpancés, lo cual limita los modelos de estudio disponibles para la búsqueda de fármacos que inhiban su proliferación. Aunado a ello, como sucede de manera regular, la inversión monetaria y el tiempo requerido para el desarrollo de un fármaco son muy altos; por lo que es necesaria la búsqueda de alternativas metodológicas para disminuir estos factores. Por esto, el presente trabajo establece un enfoque computacional para la búsqueda de inhibidores que puedan interferir con el ciclo de HCV en alguna de sus etapas de infección; reduciendo de esta manera el tiempo, los costos y las limitaciones experimentales que se presentan durante la búsqueda de un fármaco efectivo para inhibir HCV.

1.3 Hipótesis

La determinación de patrones a nivel de secuencia o de estructura del virus de la hepatitis C, permitirá encontrar fármacos candidatos para inhibir la acción del virus.

1.4 Objetivos

1.4.1 Objetivo general

Diseñar y desarrollar un estudio bioinformático para el análisis del genoma del virus de la hepatitis C y la posterior identificación y prueba de fármacos anti-virales por medio de herramientas bioinformáticas de acoplamiento y dinámica molecular.

1.4.2 Objetivos específicos

- 1 Analizar comparativamente a nivel de estructura los diferentes genotipos reportados.
- 2 Analizar comparativamente a nivel de secuencia los diferentes genotipos reportados.
- 3 Modelar estructuralmente y refinar por dinámica molecular las estructuras conservadas.
- 4 Identificar posibles fármacos que interfieran con la estructura del blanco o se unan a la secuencia conservada basado en cálculos de acoplamiento molecular.
- 5 Simular los complejos proteína-ligando y realizar su validación con solvente implícito.

Capítulo 2. Metodología

2.1 Selección y preparación de la estructura de la polimerasa viral NS5B

2.1.1 Obtención de la estructura cristalográfica de la polimerasa viral NS5B

La estructura cristalográfica de la polimerasa que fue utilizada para todos los ensayos se obtuvo de la base de datos Protein Data Bank (PDB) (Berman *et al.*, 2000), el número de acceso con el que se encuentra depositada es PDB ID: 3gsz. El cristal fue obtenido por difracción de rayos X y se encuentra como una unidad asimétrica conteniendo dos macromoléculas de la proteína: cadena A y cadena B (Rydberg *et al.*, 2009). Los cristales depositados en PDB, en ocasiones, contienen errores propios del proceso de cristalización o moléculas adicionales que facilitan su caracterización cristalográfica, los cuales deben ser revisados con antelación a los ensayos bioinformáticos. Para el análisis del cristal, la polimerasa se trató en primera instancia separando las dos cadenas, se seleccionó la cadena A de la polimerasa y se le retiraron las moléculas de agua. La estructura obtenida de la proteína se utilizó para todos los ensayos posteriores.

2.1.2 Modelado homólogo

Para verificar que el cristal de la macromolécula de la polimerasa, obtenida de la base de datos PDB, contenga todos los aminoácidos que la conforman, se sometió a un proceso de modelado homólogo utilizando el servidor para modelado homólogo de estructuras proteínicas Swiss Model (Kiefer *et al.*, 2009). El cristal de la polimerasa que se modeló contenía inicialmente 558 aminoácidos y se modeló a un pH de 6.8.

2.2 Alineamiento y búsqueda de regiones conservadas

2.2.1 Búsqueda de las regiones conservadas a nivel estructural de NS5B

Se hizo un alineamiento contra secuencias proteínicas depositadas en PDB, utilizando el servidor PDBsum (de Beer *et al.*, 2014). El estudio realizado se generó para la búsqueda de aminoácidos y regiones conservadas a lo largo de toda la secuencia de NS5B. El total de secuencias existentes que contenían información sobre la secuencia de la estructura de la proteína NS5B fueron 229. Para este análisis, la referencia de entrada fue 3gsz_A. El alineamiento de donde se obtuvieron los resultados globales se realizó

solamente contra 28 secuencias de la proteína NS5B, de las cuales 21 son representativas –diferentes entre ellas– y donde cada una de estas secuencias tuvo establecidos los siguientes parámetros: porcentaje de identidad mayor al 85% con respecto al cristal de referencia, E-value menor a 0.0001; y un traslape del 97% de los aminoácidos entre el cristal de referencia y todas las secuencias alineadas, en el menor de los casos.

2.2.2 Búsqueda de las regiones conservadas a nivel secuencia de NS5B

Para obtener una segunda referencia y poder corroborar o desechar los resultados de las regiones conservadas encontradas en el análisis a nivel estructural en la proteína NS5B se realizó un alineamiento a nivel de secuencia. Para llevarlo a cabo se utilizó la herramienta en línea CS-BLAST (Context-Specific Basic Local Alignment Search Tool, por sus siglas en inglés) (Alva *et al.*, 2016), la cual mejora significativamente la calidad de los alineamientos y duplica la sensibilidad, colectando homólogos de SwissProtKB (KnowledgeBase, por sus siglas en inglés) (Pundir *et al.*, 2017) que es una base de datos revisada y con secuencias no redundantes y de alta calidad. El umbral del porcentaje de identidad se disminuyó a un mínimo del 75% para abarcar más secuencias de NS5B. La cantidad final de secuencias representativas alineadas fue de 72, de un total de 85; el E-value se mantuvo muy por debajo de 0.0001 y siguió existiendo un traslape del 97% de los aminoácidos entre el cristal de referencia y todas las secuencias alineadas.

2.2.3 Identificación de aminoácidos de interés dentro de las regiones conservadas

Para la búsqueda y representación de aminoácidos que se localizan en el núcleo catalítico del complejo de replicación viral y que además participan en el proceso de polimerización, se utilizó el software WebLogo (Crooks *et al.*, 2004). Con él se generaron diferentes cadenas de once aminoácidos – longitud suficiente para observar la presencia de regiones conservadas– en formato PNG, que contienen uno o varios aminoácidos conservados y que a su vez forman enlaces de hidrógeno mientras se realiza el proceso de polimerización en sus etapas iniciales.

2.3 Selección y preparación de ligandos para ensayos de inhibición de la polimerasa

2.3.1 Generación de librerías de ligandos a partir de base de datos ZINC15

La pre-selección de ligandos para los análisis de acoplamiento sobre el sitio activo de la polimerasa se obtuvo utilizando la base de datos ZINC15 (Sterling e Irwin, 2015) –versión con más de 120 millones de compuestos “tipo fármaco”, disponibles comercialmente–. El ligando control se definió con previa búsqueda bibliográfica. El ligando de referencia elegido, que ha presentado la mejor barrera al desarrollo de resistencia de HCV y además tiene un efecto pangénotípico, es el compuesto sofosbuvir (Tabla 2). Los filtros utilizados para la búsqueda de grupos de ligandos utilizando una sola molécula de referencia, fueron cuatro: por subestructura, son compuestos que presentan la misma estructura molecular del ligando de referencia más átomos o grupos funcionales extra unidos a ella; el índice “Dice” que arroja resultados basados en la similitud química con respecto al compuesto de referencia más el índice de similitud que va de cero a uno, mientras más alto sea el valor del índice mayor será la similitud, el valor mínimo en las moléculas arrojadas de la búsqueda fue de 0.63; el índice “Tanimoto” que funciona de manera similar al índice Dice, el valor mínimo de este índice correspondió a 0.67; y por grupo funcional que se encuentre contenido en el ligando de referencia, el único grupo funcional considerado fueron las fosforamidas.

Los formatos de descarga de los ligandos fueron en 2D, correspondiente a los SMILES (simplified molecular input line entry specification, por sus siglas en inglés); y en 3D con formato .mol2.

Tabla 2. Agrupación inicial de librerías de ligandos.

NNI (Inhibidores análogos de No Nucleótidos)			
Sitio de inhibición	Control positivo (molécula)	Selección basada en:	Cantidad de moléculas
Pulgar, sitio I	Beclabuvir	Grupo funcional: indoles	484
		Grupo funcional: benzimidazoles	146
Pulgar, sitio II	Radalbuvir	Grupo funcional: tiofeno	332
Palma	Dasabuvir	Grupo funcional: benzotiadiazina	188
	Canabidiol	Índice Dice	283
		Subestructura	51
NI (Inhibidores análogos de Nucleótidos)			
Palma	Sofosbuvir	Subestructura	20
		Grupo funcional: fosforamidas	703
	GS-461213*	Índice Dice	167

*ID de acceso del metabolito en base de datos Drug Bank.

2.3.2 Preparación de librería de ligandos

Para generar estructuras 3D a partir de librerías 2D se utilizó el software LigPrep. Para realizar la corrección de librerías conteniendo pocos ligandos con estructuras 3D se utilizó el Software Maestro; lo que se buscó en esta etapa fue revisar las estructuras y verificar que cada ligando estuviera constituido por el número y tipo de átomos correcto, obtener cada molécula en su estado relajado y adicionar los átomos faltantes en caso de ser necesario, que en algunos casos consistieron solamente en hidrógenos. Ambos softwares forman parte de la paquetería Schrödinger (Schrödinger, 2016).

2.4 Ensayos de cribado virtual polimerasa-ligandos

2.4.1 Delimitación de región en la polimerasa para realizar el cribado virtual

Previo a la ejecución del cribado virtual se delimitó la región de estudio en la polimerasa, la cual fue el blanco para todos los ensayos de acoplamiento. La región se definió basado en los 11 aminoácidos que demostraron formar enlaces de hidrógeno entre el sustrato y la polimerasa cuando ésta se encuentra en sus etapas iniciales de polimerización (Appleby *et al.*, 2015). El enrejado configurado se centró en el sitio activo de la polimerasa (56.177, 2.524, 53.351) y sus medidas tridimensionales estuvieron definidas en 74 puntos por eje con un enrejado de 0.375 Angstroms (Å).

2.4.2 Cribado virtual

Los grupos de ligandos sometidos al cribado virtual tuvieron como control positivo a moléculas que han demostrado tener actividad sobre el núcleo catalítico de la polimerasa, como se muestra en la Tabla 2. Se incluyó al canabidiol como inhibidor análogo de no-nucleótido, y a dos metabolitos análogos de nucleótido; las librerías generadas se hicieron con una búsqueda por subestructura, índice Dice y por grupo funcional. El cribado virtual final que se realizó incluyó 1224 ligandos. Para ejecutar el cribado virtual se utilizó la paquetería de AutoDockTools (ADT) (Morris *et al.*, 1998). Para la generación de los mapas y los archivos de parámetros para la librería se utilizaron AutoGrid y AutoDock, respectivamente. Todo se ejecutó con línea de comandos en UNIX. El procesamiento de los datos se realizó en el clúster OMICA, en 4 nodos, con 2 procesadores Intel Xeon (E5-2670) cada nodo y 12 núcleos cada procesador. Para todos los ensayos, el número de evaluaciones de acoplamiento polimerasa-ligando que se realizó con cada complejo

inició en mil ($1.00E+03$) evaluaciones, a partir de este número se hizo un incremento en un orden de magnitud hasta llegar a los 10 millones ($1.00E+07$) de evaluaciones.

2.4.3 Selección de ligandos candidatos para ensayar simulaciones moleculares acopladas a la polimerasa

Seguido al cribado virtual, se enlistaron todos los ligandos que mostraron una energía libre de acoplamiento mejor que -6.5 kcal/mol. Este criterio se estableció basado en que fueron valores de energía superiores al mostrado por uno de los ligandos control, sofosbuvir. A partir de este punto, se seleccionaron algunos de estos ligandos para acoplarlos a la polimerasa y someterlos a simulaciones moleculares.

2.5 Simulaciones moleculares acopladas del complejo polimerasa-ligando

2.5.1 Preparación de los ligandos

Cada uno de los ligandos seleccionados para las simulaciones fueron verificados inicialmente con ADT (Morris *et al.*, 1998), con el cual se analizaron las diez mejores conformaciones de cada ligando que mejor se acoplaron a la polimerasa. De éstas, se eligió la conformación que mostró la clasificación más alta basado en su energía, para cada uno de los ligandos. Además, se asignó un nombre de residuo para cada ligando a simular. La Figura 4 muestra la estructura de los ligandos elegidos y el identificador de residuo con el que se distinguió cada uno para su acoplamiento.

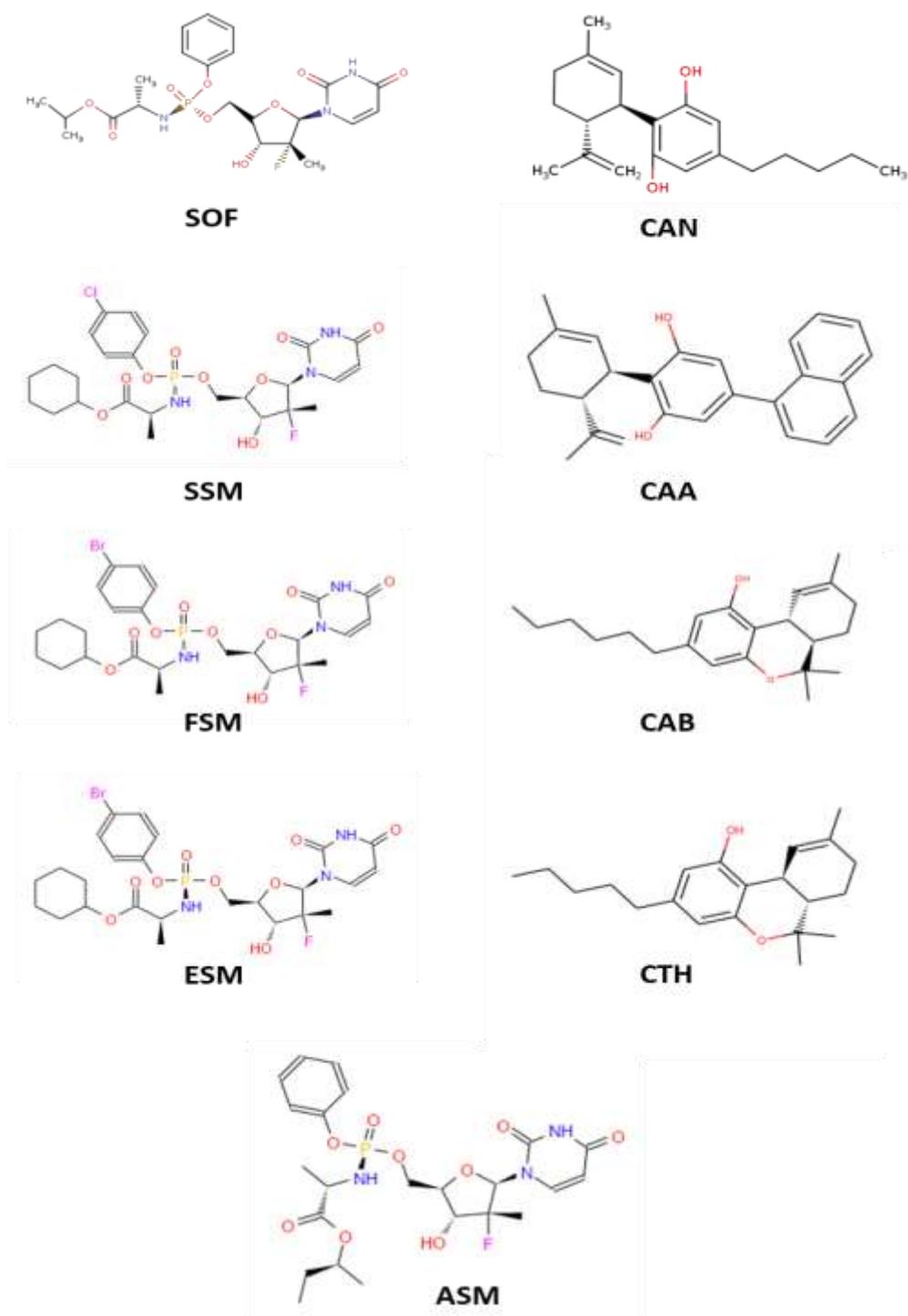


Figura 4. Ligandos seleccionados para simular acoplados a la polimerasa y su identificador de residuo.

2.5.2 Preparación del complejo

Para simular los diferentes complejos se definieron tres modelos de campos de fuerza: “ff14SB”, para la polimerasa viral; “gaff”, como campo de fuerza general para cada uno de los diferentes ligandos; y “tip3p” para el agua e iones atómicos. Cada uno de los campos de fuerza se encuentran contenidos en la paquetería AmberTools y están diseñados para trabajar con Amber. Aunado a este proceso se prepararon los diferentes complejos ligando-polimerasa; se neutralizó el sistema, se generó su solvatación en una caja de 12x12x12 Å; y se crearon los diferentes parámetros y topologías para simular en Amber (Case *et al.*, 2016).

2.5.3 Simulaciones moleculares

Todas las simulaciones moleculares acopladas se desarrollaron en Amber, implementado en GPUs. La trayectoria completa de la simulación constó de cuatro etapas: minimización, calentamiento, densidad y producción. Cada proceso de simulación estuvo constituido, de manera general, de 2500 “frames”; la duración total de cada una de las simulaciones fue de 50 nanosegundos (ns). Los tiempos de simulación se reportaron en porcentajes para facilitar su comprensión, por lo que 1% del tiempo de simulación tuvo un equivalente a 0.5 ns y comprendió 25 frames. Para cada simulación se obtuvieron los RMSD (Root Mean Square Deviation, por sus siglas en inglés) de las posiciones atómicas del esqueleto de la proteína –eje de los carbonos alfa–, esto se realizó para observar la variación conformacional de la proteína NS5B con cada uno de los ligandos a los que se acopló para formar los diferentes complejos.

2.5.4 Análisis de enlaces de hidrógeno

Para definir qué aminoácidos de la polimerasa interaccionaron durante la simulación con los distintos átomos de los ligandos, se analizaron los enlaces de hidrógeno formados entre ellos. El porcentaje de interacción establecido como válido para cada una de las trayectorias fue del 20%, el equivalente en tiempo de simulación a este porcentaje correspondió a 10ns.

2.6 Ensayos de validación MM-GBSA

2.6.1 Validación de la simulación acoplada con un modelo de solvente implícito

El método MM-GBSA (Molecular Mechanics, Generalized Born Surface Area, por sus siglas en inglés), con la variante de ser un modelo de solvente implícito, fue utilizado para validar los resultados arrojados por la trayectoria de la simulación de los complejos polimerasa-ligando (Miller *et al.*, 2012). Se estimó el cambio en la energía libre [$\Delta (\Delta G)$] de la unión de cada uno de los ligandos a la polimerasa. Posterior a esto se obtuvieron los valores de $\Delta (\Delta G)$ para el ligando, la polimerasa y el complejo, y se calculó la energía de cada reacción. Finalmente, para identificar los ligandos que tuvieron una mejor actividad sobre la polimerasa, se normalizaron los datos con respecto a la energía de reacción del ligando de referencia SOF acoplado a la proteína. Esto se realizó dividiendo la energía de reacción de cada ligando entre la energía de reacción del ligando de referencia.

Capítulo 3. Resultados

3.1 Corrección obtenida por la aplicación del modelado homólogo

El cristal de referencia obtenido del PDB contenía 558 aminoácidos y la estructura obtenida del modelado fue de 563 aminoácidos. La región de 5 aminoácidos faltante corresponde a la secuencia continua de P149-A150-K151-G152-G153. En el comparativo con la constitución estructural de otras proteínas, esta secuencia faltante no se encuentra formando estructuras secundarias correspondientes a hebras beta, hojas beta o hélices; sin embargo, forman parte de dos vueltas beta que se observan en otras polimerasas NS5B representadas en la Figura 5.

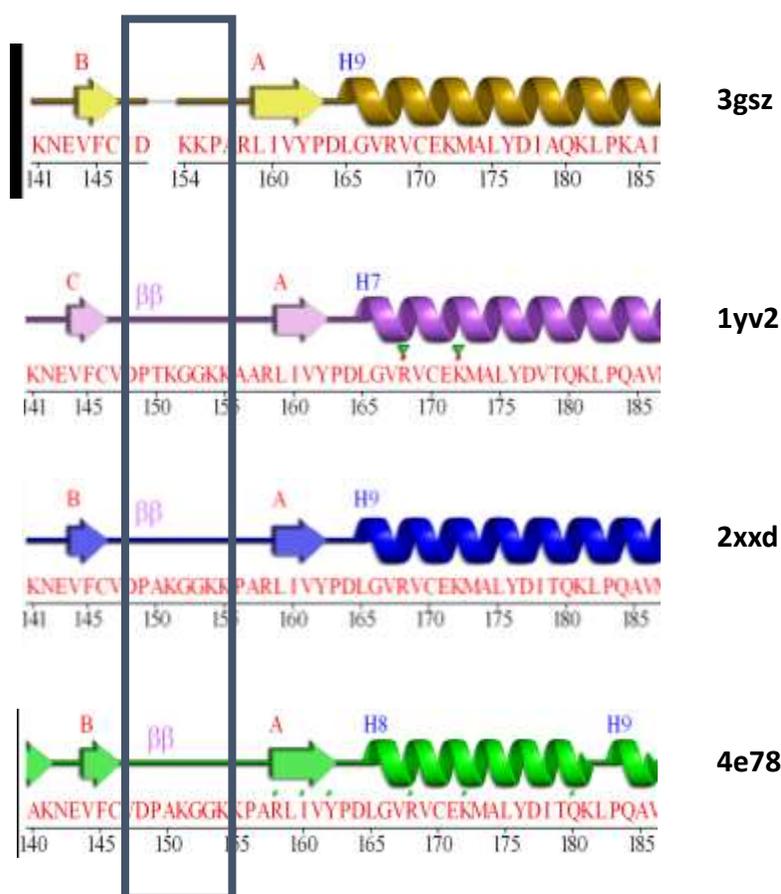


Figura 5. Comparativo del modelado homólogo. La región de cinco aminoácidos faltantes forma dos vueltas beta en la estructura consenso. Izquierda: sección de región que contiene la secuencia faltante en el cristal referencia –3gsz– y tres secuencias distintas; derecha: PDB ID de cada secuencia.

3.2 Análisis de la conservación de regiones y aminoácidos en NS5B

La polimerasa NS5B presentó varias regiones conservadas a lo largo de su secuencia y se encuentran representadas en color rojo en la Figura 6. Algunas de estas regiones se encuentran comprendidas entre los aminoácidos D220-E230, S282-T292 y G317-D319, las cuales contienen aminoácidos localizados en el sitio activo de la proteína y son partícipes del proceso de polimerización.

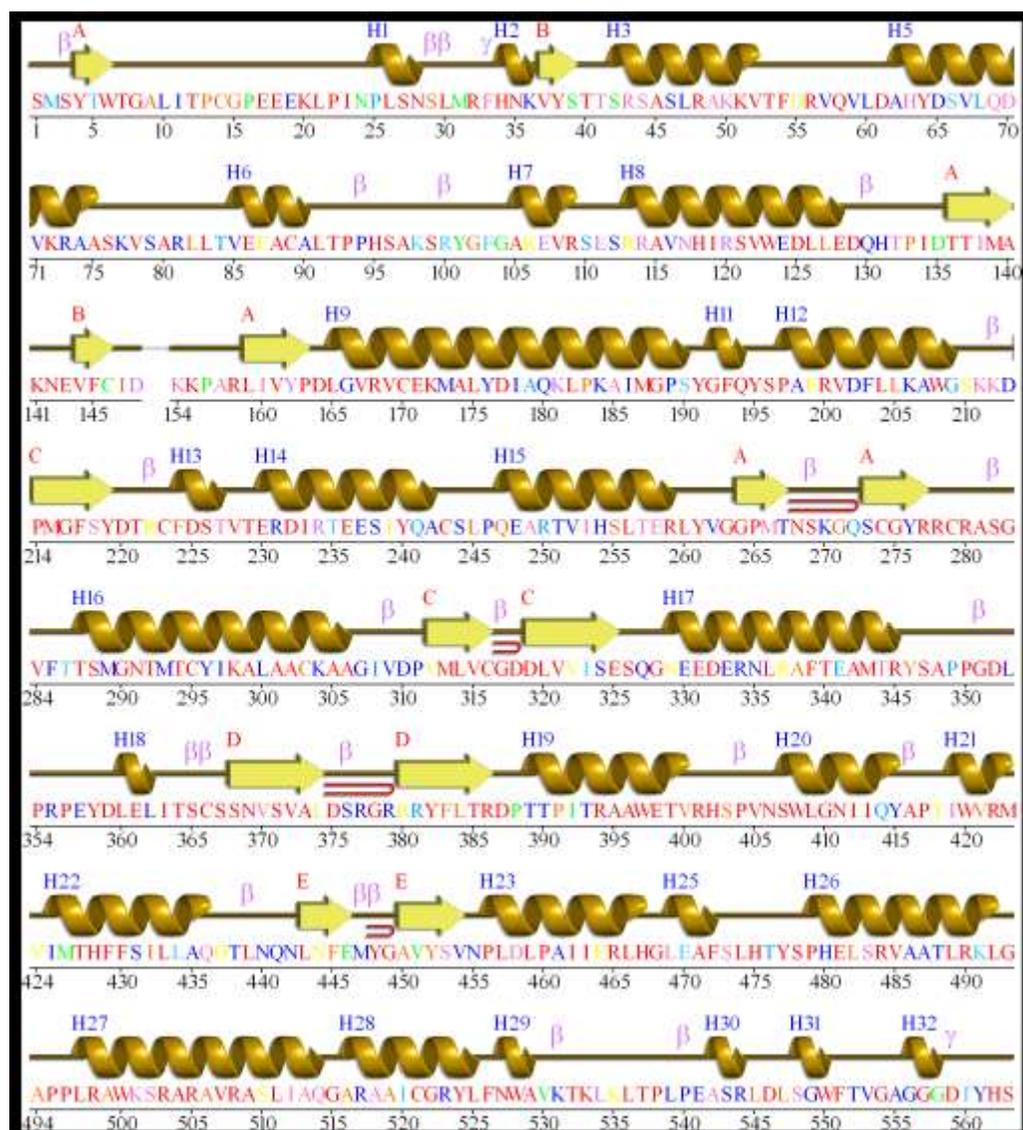


Figura 6. Análisis de la secuencia completa de aminoácidos de la cadena A de la polimerasa. Los aminoácidos y regiones conservadas en su totalidad se encuentran en color rojo; los aminoácidos con el menor grado de conservación están en color azul; la restante gama de colores representa variedad en el grado de conservación.

En la Figura 7 se presentan los códigos de secuencias conteniendo aminoácidos que participan en el proceso de polimerización en sus etapas iniciales formando enlaces con diferentes átomos. Los aminoácidos que reconocen de manera natural al sustrato son: D225, S282, N291; aminoácidos que interaccionan con Mg^{+2} catalítico: D220, D318, D319; y aminoácidos que forman enlaces de hidrógeno con los nucleótidos de la cadena molde o con los nucleótidos que ingresan al proceso de polimerización en sus etapas iniciales: K141, R158, S367, R386, R394.

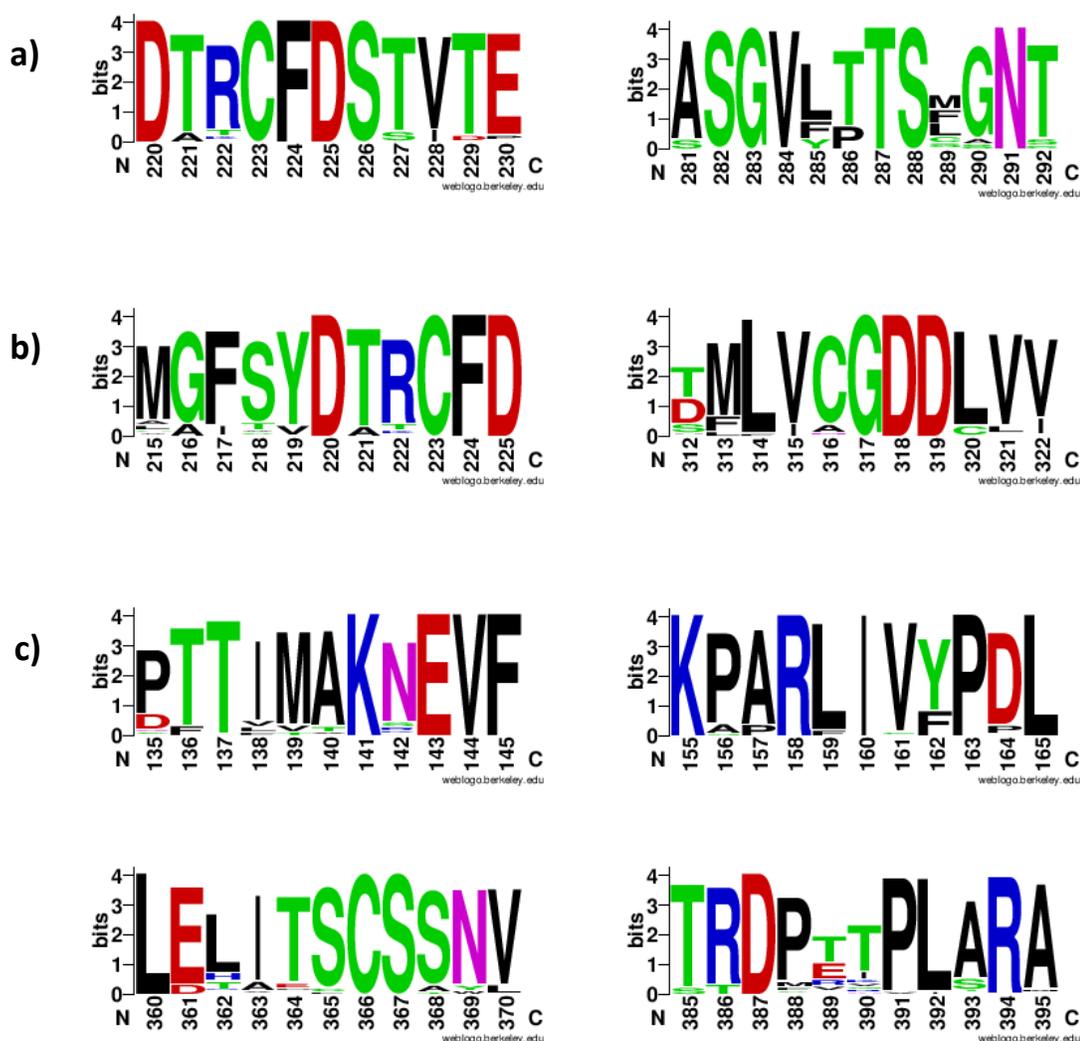


Figura 7. Logos generados para la verificación de aminoácidos conservados. a) D225, S282 y N291 son los primeros aminoácidos en participar en el reconocimiento del sustrato entrante; b) D220, D318 y D319 interaccionan con Mg para estabilizar el nucleótido entrante; c) K141, R158, S367, R386 y R394 forman enlaces de hidrógeno con el molde y los nucleótidos que ingresan para su polimerización.

3.3 Selección de los ligandos sometidos al cribado virtual

De los 1224 ligandos evaluados durante el cribado virtual se eligieron 34 candidatos, como se muestra en la Figura 8. Sofosbuvir (SOF), canabidiol (CAN) y GS-461213² (ACT) son tres ligandos de referencia que presentaron valores de energía libre de -6.17 kcal/mol, -5.97 kcal/mol y -5.67 kcal/mol, respectivamente. El resto de los 31 ligandos arrojaron valores de energía libre mejores a éstos y hasta un valor de -8.78 kcal/mol.

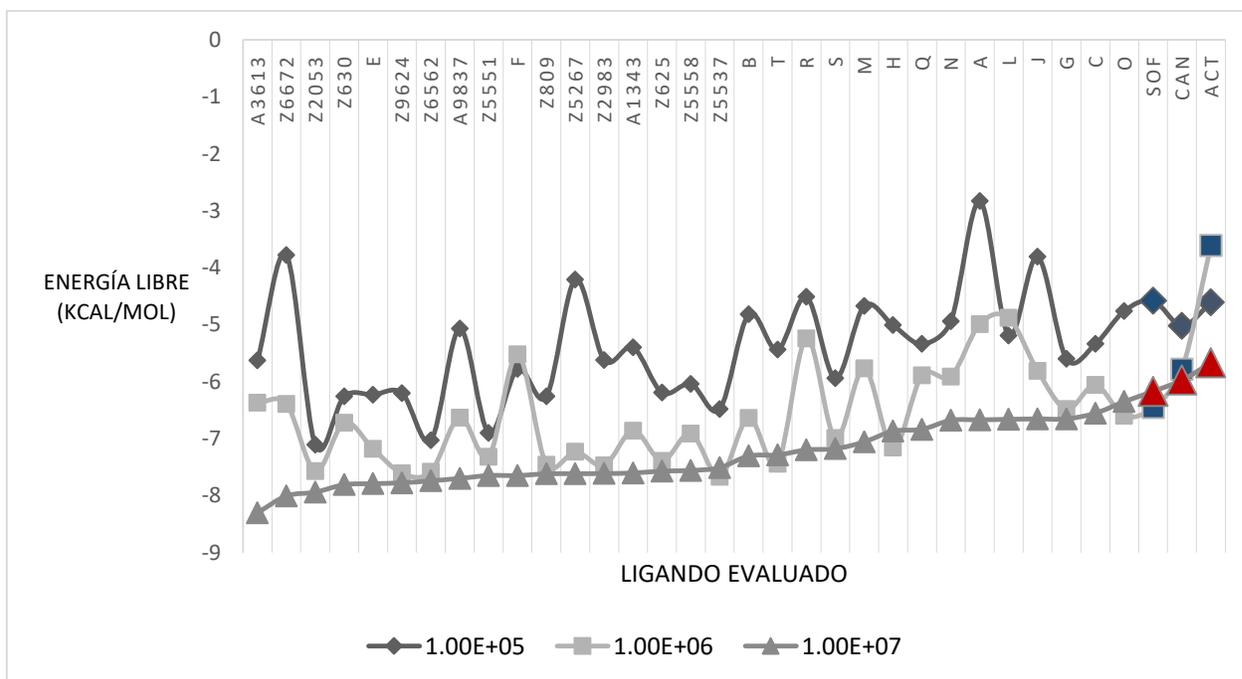


Figura 8. Gráfico de ligandos que presentaron una energía superior a los ligandos de referencia. 1.00E+05, 1.00E+06 y 1.00E+07 corresponden a 100 mil, 1 millón y 10 millones de evaluaciones, respectivamente, de cada ligando con el sitio activo de la polimerasa. SOF, CAN y ACT corresponden a ligandos utilizados como referencia. Los datos se encuentran ordenados del valor de energía más negativo hacia el más positivo con respecto a la línea obtenida con 1.00E+07 evaluaciones.

Posterior a este filtro se eligieron 9 ligandos y se les asignó un nombre de residuo a cada uno de ellos con el cual se identificaron a partir de este punto. En la Tabla 3 se establecen los ligandos elegidos, el identificador inicial del ligando y la relación con su nuevo identificador por residuo, de esta manera se comenzaron a referenciar cada uno de los ligandos; también se alcanza a observar que se trata de moléculas diferentes entre ellas con respecto a su número de átomos y de planos de rotación. Además de SOF, que es el fármaco pangenotípico que ha mostrado mejor inhibición de la polimerasa NS5B, y de CAN, que es una molécula de origen natural de la que también se ha reportado efecto inhibitorio, se

² ID del metabolito en base de datos DrugBank. <https://www.drugbank.ca/>

seleccionaron cuatro ligandos (ESM, SSM, FSM y ASM) similares a SOF por subestructura, y tres (CAA, CAB y CTH) relacionados a CAN por índice de similitud química.

Tabla 3. Ligandos elegidos para ser acoplados a la polimerasa. Se eligieron cuatro ligandos (SSM, FSM, ESM y ASM) relacionados con SOF y tres (CAA, CAB y CTH) similares a CAN, todos ellos mejores en sus valores de energía.

ID Ligando	Residuo	# Átomos	# Enlaces de rotación	Energía con 1.00E+07 evaluaciones
S	SSM	43	12	-8.78
F	FSM	43	12	-8.31
z2053	CAA	30	5	-7.94
E	ESM	43	12	-7.83
z630	CAB	25	6	-7.8075
z9624	CTH	24	5	-7.775
A	ASM	40	13	-7.53
SOF	SOF*	39	12	-6.17
CAN	CAN*	25	8	-5.9725

La columna con el encabezado "Residuo" corresponde al nuevo identificador con el que se distinguirá cada ligando.

*Ligandos de referencia.

3.4 Evaluación de la estabilidad de la polimerasa con cada ligando acoplado

Los resultados de las simulaciones acopladas se pueden agrupar en dos bloques: uno para las simulaciones con el ligando referencia SOF y sus cuatro ligandos similares por subestructura, y otro para el ligando referencia CAN y sus tres ligandos relacionados por coeficiente de similitud química. De manera global, estos dos grupos mostraron diferentes patrones en el número de enlaces de hidrógeno formados y en los aminoácidos con los que los forman; y comportamientos muy similares en la estabilidad observada en los gráficos de RMSD para la mayoría de todos los ligandos.

A modo general, para las simulaciones de la polimerasa acoplada a SOF, ESM, SSM, FSM y ASM; la mayor parte del tiempo de simulación se observó estabilidad en la polimerasa, la cual es mostrada en los gráficos de RMSD durante todo el tiempo de simulación. El RMSD promedio para los complejos de NS5B acoplada con el ligando de referencia SOF y sus cuatro ligandos (ESM, SSM, FSM y ASM) similares por subestructura fueron: NS5B-SOF: 1.157 Å; NS5B-ESM: 1.222 Å; NS5B-SSM: 1.170 Å; NS5B-FSM: 1.060 Å y NS5B-ASM: 1.132 Å.

El número de enlaces de hidrógeno formados la mayor parte del tiempo de cada una de las dinámicas fue de dos o tres enlaces, lo que correspondió a alrededor del 25% del tiempo de simulación para dos enlaces y 25% para tres enlaces. De manera más específica, el número de enlaces de hidrógeno que se formó durante la mayor parte del tiempo de simulación para cada ligando fueron: SOF: dos enlaces durante el 26.8% del tiempo; ESM: tres enlaces durante el 29% del tiempo; SSM: tres enlaces durante el 25.4% del tiempo; FSM: tres enlaces durante el 25.4% del tiempo y ASM: tres enlaces durante el 21% del tiempo. En los diagramas de interacciones, SOF mostró formar enlaces de hidrógeno con los aminoácidos D318, L159 y R158; ESM con D318, Q414 y R200; SSM con D318, S288, T287 y R200; FSM con N411, E446, R386 y Y448; por último, ASM formó enlaces de hidrógeno con G557 y Q414. En las figuras 9-13 se muestran los RMSD para cada simulación acoplada con cada uno de estos cinco ligandos, así como el número de enlaces formados y su porcentaje de formación con respecto a la simulación completa, también se muestra el diagrama de interacción ligando-aminoácido.

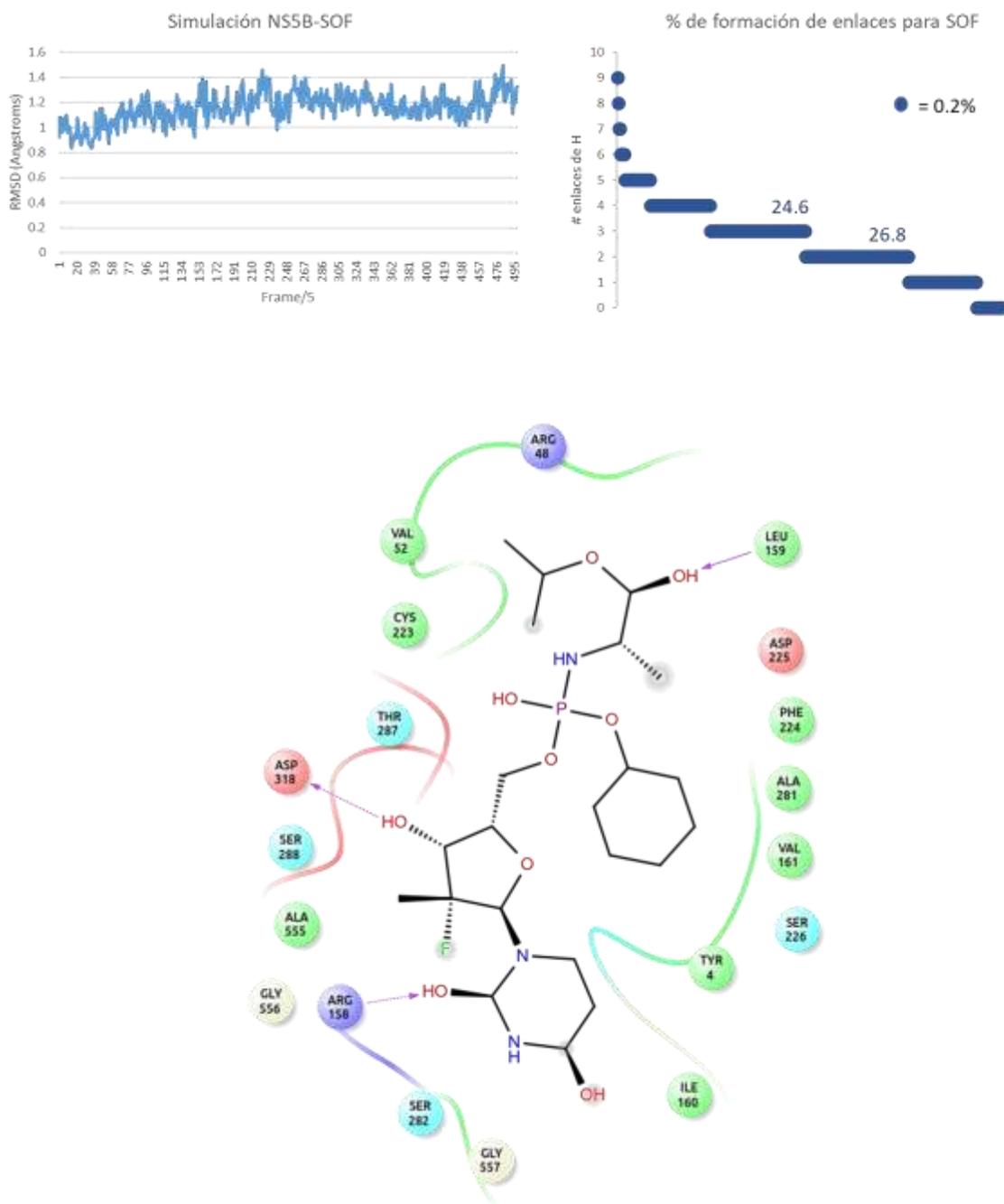


Figura 9. Resultado de la simulación del complejo NS5B-SOF que contiene el ligando de referencia SOF. Superior izquierda: RMSD para la simulación acoplada NS5B-SOF; RMSD promedio: 1.157 Å, mínimo: 0.837 Å, máximo: 1.492 Å, sd: 0.118; superior derecha: porcentaje de formación para los diferentes números de enlaces de hidrógeno; el 26.8% del tiempo total se formaron dos enlaces y el 24.6% tres enlaces entre NS5B-SOF; inferior: esquema de interacción, SOF mostró enlazarse con D318, L159 y R158.

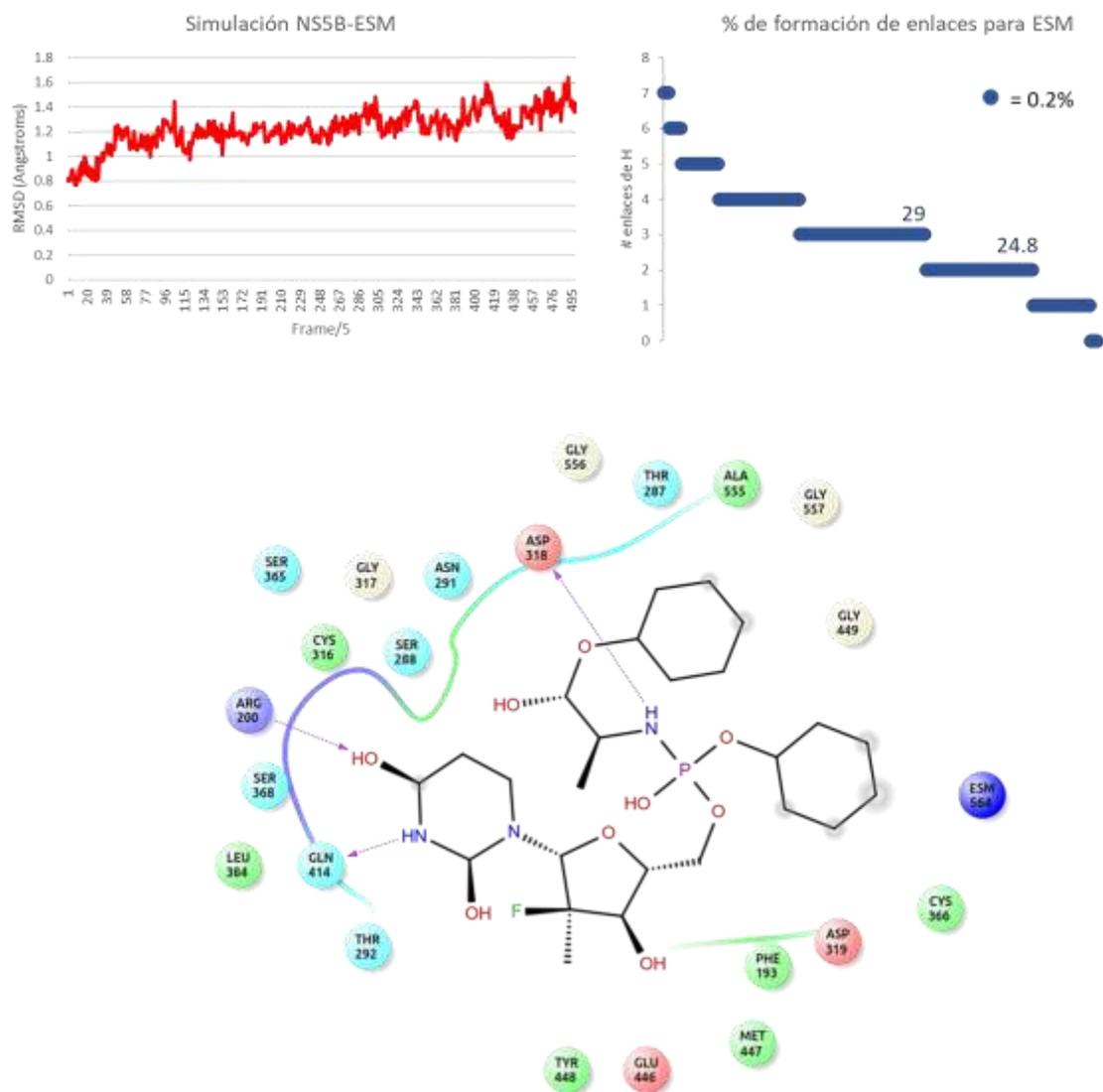


Figura 10. Resultado de la simulación del complejo NS5B-ESM. Superior izquierda: RMSD para la simulación acoplada NS5B-ESM; RMSD promedio: 1.222 Å, mínimo: 0.769 Å, máximo: 1.644 Å, sd: 0.148; superior derecha: porcentaje de formación para los diferentes números de enlaces de hidrógeno; el 29% del tiempo total se formaron tres enlaces y el 24.8% dos enlaces entre NS5B-ESM.; inferior: esquema de interacción, ESM mostró enlazarse con D318, Q414 y R200.

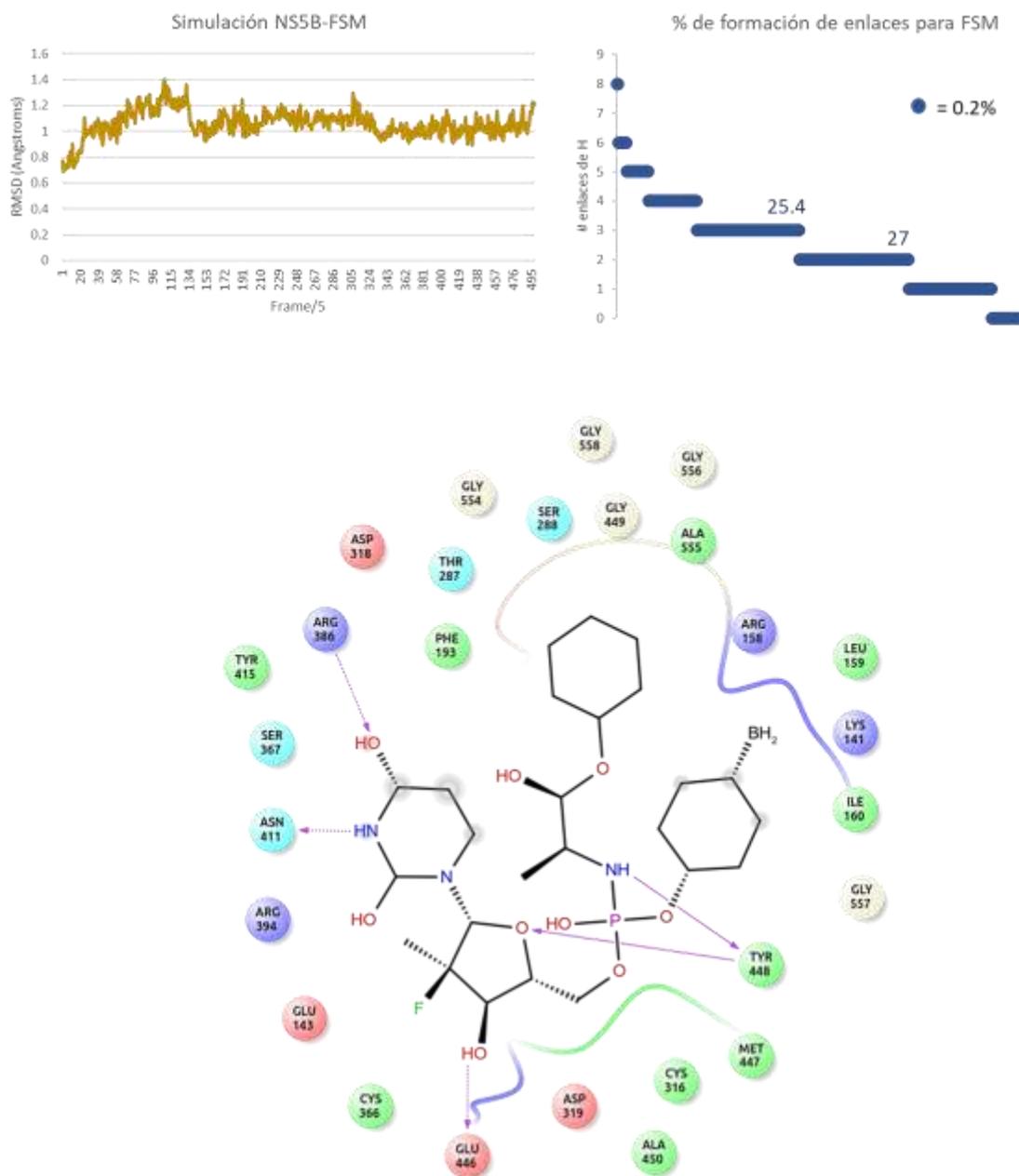


Figura 12. Resultado de la simulación del complejo NS5B-FSM. Superior izquierda: RMSD para la simulación acoplada NS5B-FSM; RMSD promedio: 1.060 Å, mínimo: 0.691 Å, máximo: 1.402 Å, sd: 0.107; superior derecha: porcentaje de formación para los diferentes números de enlaces de hidrógeno; el 25.4% del tiempo total se formaron tres enlaces y el 27% dos enlaces entre NS5B-FSM; inferior: esquema de interacción, FSM mostró enlazarse con N411, E446, R386 y Y448.

Con respecto a las simulaciones acopladas a CAN, CAA, CAB y CTH; se observa estabilidad en los RMSD durante todo el tiempo de simulación. El RMSD promedio para los complejos de NS5B acoplada con el ligando de referencia CAN y sus tres ligandos (CAA, CAB y CTH) relacionados por su coeficiente de similitud química fueron: NS5B-CAN: 1.106 Å; NS5B-CAA: 1.114 Å; NS5B-CAB: 1.200 Å y NS5B-CTH: 1.057 Å.

Durante la mayor parte del tiempo de simulación para este grupo de ligandos acoplados a la polimerasa se estuvieron formando dos enlaces de hidrógeno: CAN formó dos enlaces de hidrógeno durante el 38% del tiempo; CAA dos enlaces durante el 47.6% del tiempo de simulación; CAB no mostró formar enlaces durante el mayor tiempo de simulación correspondiente al 89.2% del tiempo y CTH se encontró formando dos enlaces el 42.6% del tiempo con la polimerasa. CAN mostró interacción con V405, CAA con D319 y CTH formó enlace con C366. Los gráficos con el RMSD, el porcentaje para los diferentes números de enlaces formados y el diagrama de su interacción se representan en las figuras 14-17.

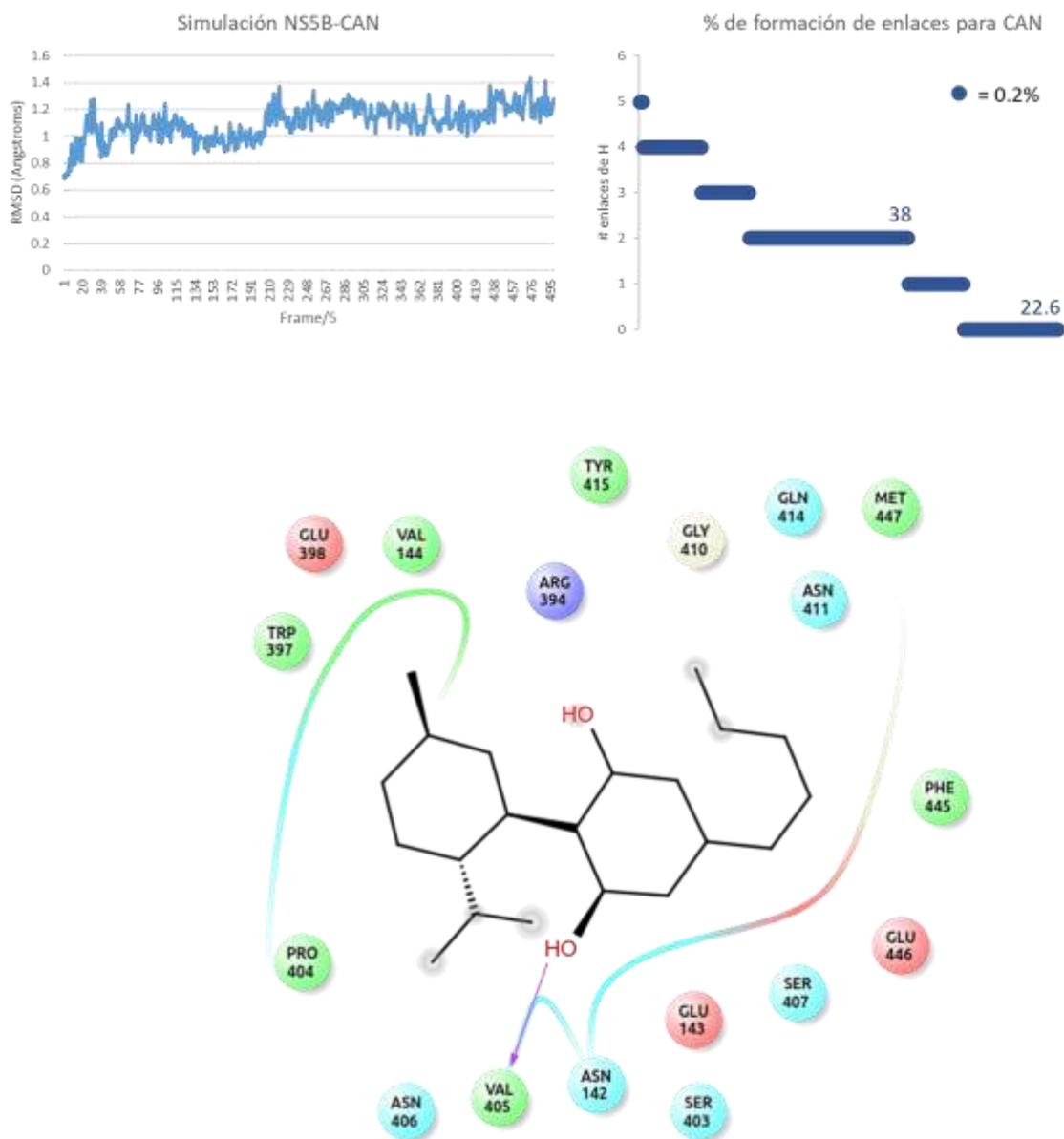


Figura 14. Resultado de la simulación del complejo NS5B-CAN que contiene el ligando de referencia CAN. Superior izquierda: RMSD para la simulación acoplada NS5B-CAN; RMSD promedio: 1.106 Å, mínimo: 0.681 Å, máximo: 1.433 Å, sd: 0.126; superior derecha: porcentaje de formación para los diferentes números de enlaces de hidrógeno; el 38% del tiempo total se formaron dos enlaces y el 22.6% no se presentaron interacciones entre NS5B-CAN; inferior: esquema de interacción, CAN mostró enlazarse con V405.

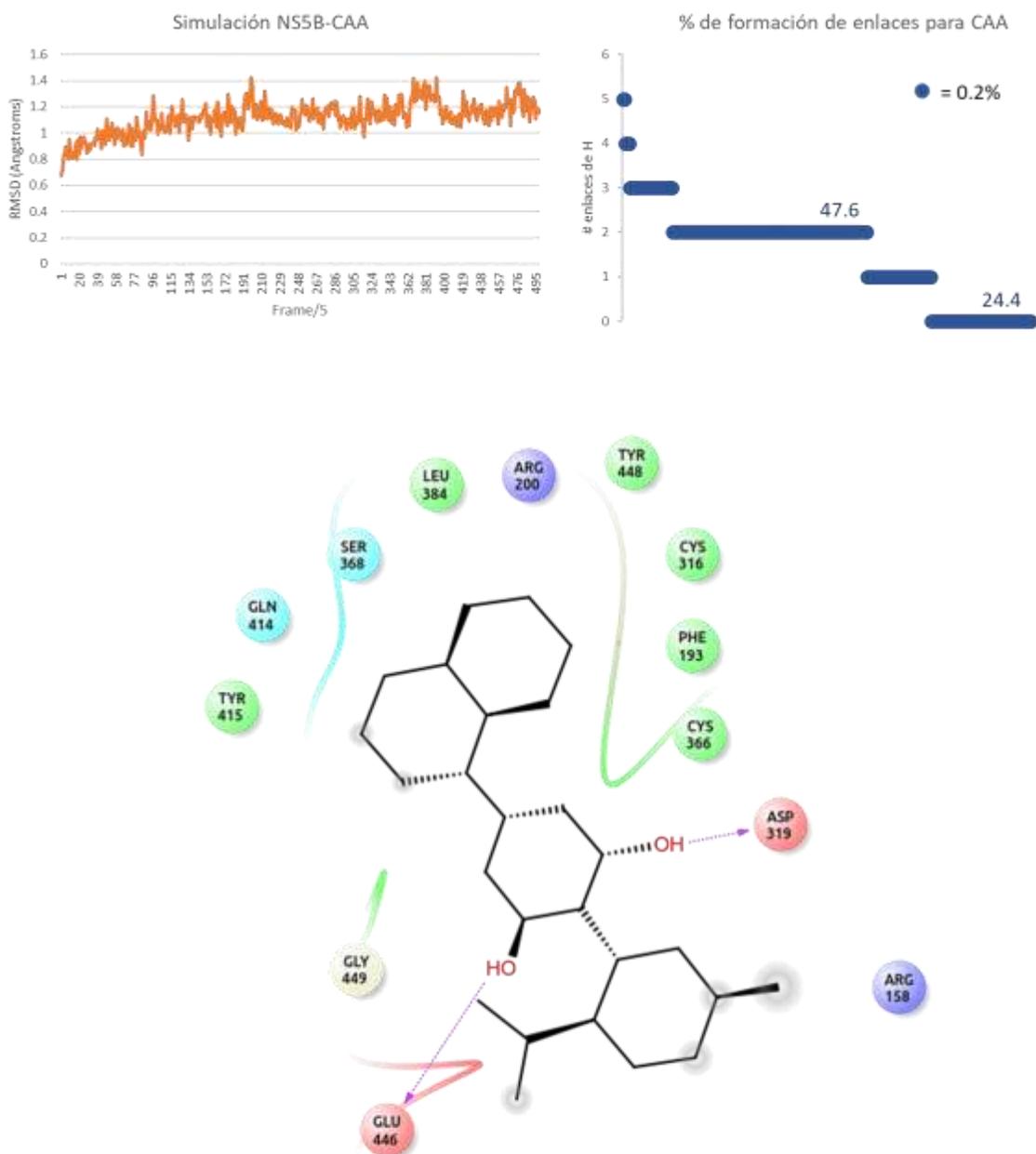


Figura 15. Resultado de la simulación del complejo NS5B-CAA. Superior izquierda: RMSD para la simulación acoplada NS5B-CAA; RMSD promedio: 1.114 Å, mínimo: 0.674 Å, máximo: 1.424 Å, sd: 0.119; superior derecha: porcentaje de formación para los diferentes números de enlaces de hidrógeno; el 47.6% del tiempo total se formaron dos enlaces y el 24.4% no se presentaron interacciones entre NS5B-CAA; inferior: esquema de interacción, CAA mostró enlazarse con D319 y E446.

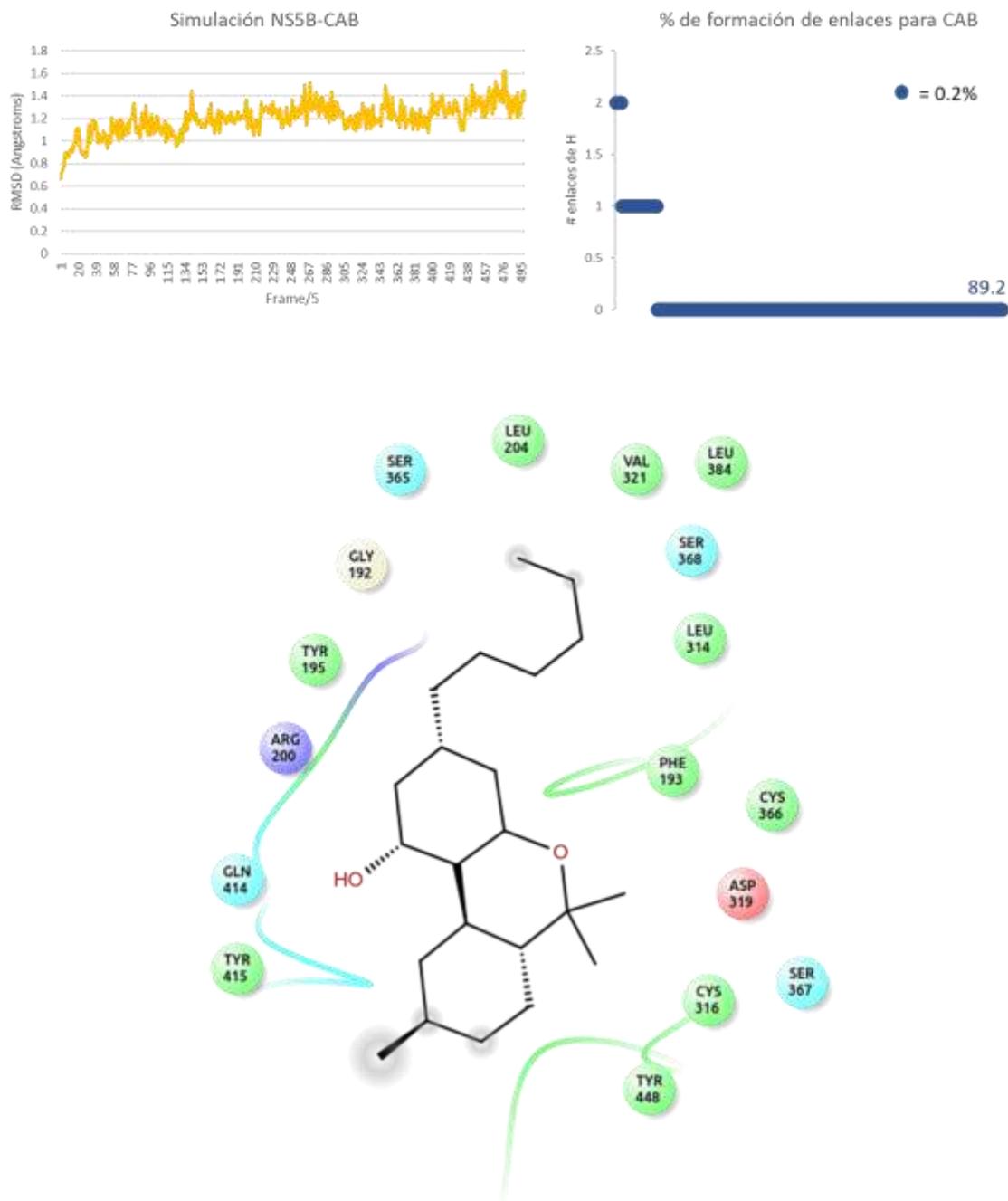


Figura 16. Resultado de la simulación del complejo NS5B-CAB. Superior izquierda: RMSD para la simulación acoplada NS5B-CAB; RMSD promedio: 1.200 Å, mínimo: 0.667 Å, máximo: 1.624 Å, sd: 0.135; superior derecha: porcentaje de formación para los diferentes números de enlaces de hidrógeno; el 89.2% del tiempo de simulación no se presentaron interacciones entre NS5B-CAB; inferior: esquema de interacción, CAB no mostró interacciones representativas.

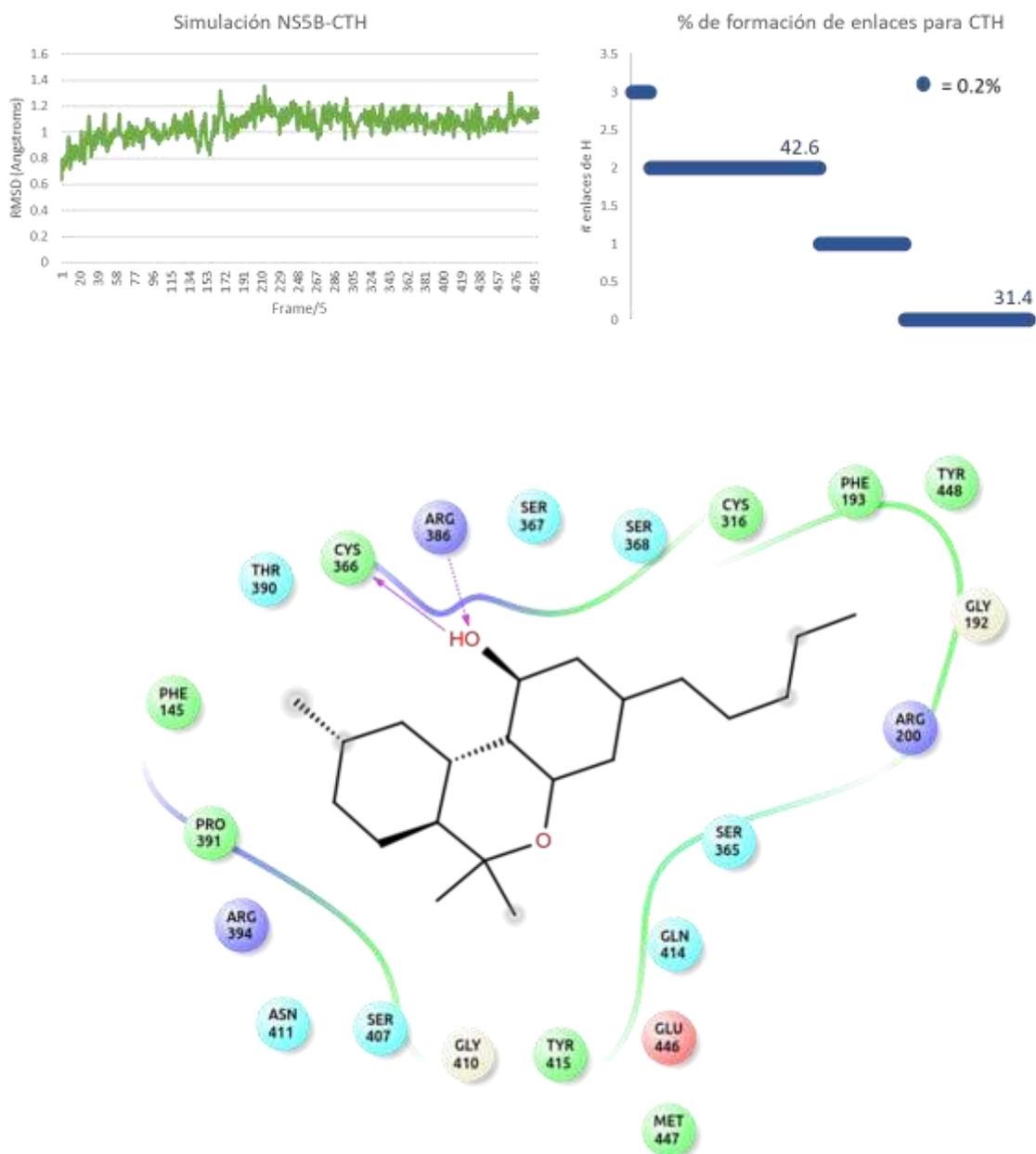


Figura 17. Resultado de la simulación del complejo NS5B-CTH. Superior izquierda: RMSD para la simulación acoplada NS5B-CTH; RMSD promedio: 1.057 Å, mínimo: 0.640 Å, máximo: 1.354 Å, sd: 0.101; superior derecha: porcentaje de formación para los diferentes números de enlaces de hidrógeno; el 42.6% del tiempo total se formaron dos enlaces y el 31.4% no se presentaron interacciones entre NS5B-CTH; inferior: esquema de interacción, CAA mostró enlazarse con C366 y R386.

3.5 Evaluación de la interacción de aminoácidos con los diferentes ligandos

Cada uno de los ligandos presentó diferente porcentaje del tiempo de interacción de algunos de sus átomos con diferentes aminoácidos de la polimerasa. Sin embargo, en algunos de ellos se alcanzó a detectar que, en las diferentes simulaciones acopladas, diferentes ligandos interaccionan con el mismo aminoácido. En la Tabla 4 se encuentran representados cada uno de los ligandos y el porcentaje de tiempo de simulación que interaccionaron con distintos aminoácidos. El grupo de cinco ligandos asociados a SOF por subestructura mostraron interaccionar con un mayor número de aminoácidos que el grupo de cuatro ligandos asociado a CAN por similitud química. Con referencia al grupo de ligandos asociados a SOF, los aminoácidos que presentaron un tiempo de interacción mayor al equivalente a 20% del tiempo de simulación (10 nanosegundos), correspondieron a D318, L159, N291 y S288; para el grupo de ligandos similares a CAN, los aminoácidos más representativos fueron D319, C366 y V405. De todos estos, los aminoácidos más interesantes fueron D318 al interaccionar con SOF, ESM y SSM en 49.8%, 94.2% y 52.4% respectivamente; D319 y C366 que presentaron una interacción el 100% del tiempo con CAA y CTH, respectivamente; así como V405 que interaccionó el 83.8% del tiempo con CAN. Los ligandos ASM y CAB fueron los más débiles al presentar una reducida interacción con la polimerasa.

Tabla 4. Ligandos y su porcentaje de interacción con diferentes aminoácidos de la polimerasa.

% de interacción durante la simulación molecular													
Ligando	Aminoácido												
	D318*	R158*	S282*	Q414	L159	N291	S288	D319*	R386*	S367*	C366	V405	Y415
SOF	49.8	1.4	7.4	-	27.6	-	1	-	-	-	-	-	-
ESM	94.2	8	-	14.4	-	-	-	-	-	-	-	-	-
FSM	-	0.6	-	-	-	32.4	-	-	10.4	-	-	-	10.6
SSM	52.4	-	-	12	-	-	20.2	-	-	-	-	-	-
ASM	-	0.6	-	8.4	-	-	-	-	-	-	-	-	-
CAN	-	-	-	-	-	-	-	-	-	-	-	83.8	-
CAA	-	-	-	-	-	-	-	100	-	-	-	-	-
CTH	-	-	-	-	-	-	-	-	2.8	1	100	-	-
CAB	-	-	-	0.6	-	-	-	-	-	-	2.4	-	6.2

*Aminoácidos de los que se ha reportado participación durante el proceso de polimerización.

3.6 Verificación de energía libre con MM-GBSA

Los resultados de los ensayos para validar la afinidad de los diferentes ligandos por la polimerasa coincidieron, en su mayoría, con lo mostrado en las dinámicas acopladas. Los datos normalizados de todos los ligandos, con respecto al ligando referencia SOF en su valor de energía de reacción –E de reacción, que corresponde a la diferencia de complejo - ligando - receptor–, mostraron a ESM, FSM, ASM y CTH con valores similares e incluso superiores.

Tabla 5. Ligandos acoplados a la polimerasa con un solvente implícito. La energía de reacción (E de reacción) corresponde a la diferencia entre el complejo y su ligando y receptor. Los datos se encuentran normalizados con respecto a SOF.

Ligando acoplado	$\Delta (\Delta G)$ en kcal/mol					
	Receptor	Ligando	Complejo	E de reacción	Normalizados*	Desv. Estándar
SOF	-48024.564	-116.788	-48180.937	-39.585	1.000	0.133
ESM	-48044.722	-132.865	-48224.342	-46.754	1.181	0.130
FSM	-48039.379	-151.452	-48244.428	-53.597	1.354	0.149
SSM	-48129.858	-127.874	-48304.668	-46.936	1.186	0.099
ASM	-48099.195	-120.835	-48241.145	-21.114	0.533	0.126
CAN	-48086.939	39.188	-48079.867	-32.116	0.811	0.070
CAA	-48023.553	28.144	-48019.841	-24.432	0.617	0.100
CTH	-47985.668	38.110	-47986.615	-39.057	0.987	0.061
CAB	-48024.620	36.943	-48019.686	-32.008	0.809	0.067

*Para realizar este cálculo de normalización, se tomó como valor de referencia la E de reacción de -39.585 kcal/mol, correspondiente a SOF. Cada uno de los valores de los demás ligandos se dividió entre el valor de referencia mencionado.

Capítulo 4. Discusión

El primer análisis para la búsqueda de regiones conservadas a nivel estructural comparando 21 secuencias no-redundantes y el segundo análisis para la búsqueda a nivel secuencia comparando 72 secuencias no redundantes demostraron, dentro de las regiones conservadas encontradas, que todos los aminoácidos localizados en el sitio catalítico y que participan en el proceso de polimerización se encuentran altamente conservados. En el año 2012, Waheed *et al.*, realizaron un estudio donde alinearon 236 secuencias de NS5B elegidas de manera aleatoria utilizando el software “CLC Workbench” y encontraron aminoácidos conservados que coinciden con los encontrados en este trabajo: D225, N291, D220, D318, D319, R158, S367, R386 y R394 son los aminoácidos que presentaron coincidencia en ambos análisis lo que da sustento a considerarlos como aminoácidos muy importantes para HCV en diferentes etapas del proceso de replicación del genoma. El único aminoácido que se encontró con un menor grado de conservación en este estudio y se reportó altamente conservado por Waheed *et al.* (2012), fue R386. La aportación de esta etapa del trabajo radica en haber obtenido resultados similares a estudios previos utilizando datos de mejor calidad –menos “gaps” en las secuencias elegidas y mayor cantidad de aminoácidos traslapados (97%)– que se refleja en una mayor sensibilidad y, por tanto, mayor confiabilidad en los datos obtenidos comparado con el estudio del año 2012 donde reportan una mayor cantidad de aminoácidos conservados que necesitan un mejor sustento debido al método aleatorio de muestreo que reportan.

El cristal elegido de la proteína NS5B para realizar todos los ensayos demostró ser un buen candidato en su análisis de modelado por homología ya que la región de los cinco aminoácidos faltantes, correspondientes a la secuencia continua P149-A150-K151-G152-G153, no formó parte de ninguno de los cuatro motivos definidos en estudios previos por Ranjith-Kumar y Kao (2006) como motivos conservados localizados en la palma (sitio catalítico) de NS5B. Sin embargo, sí forman parte de una región que constituye un dominio importante de los dedos de NS5B, la región se denomina F2, está constituida por los aminoácidos K151-G152-G153-R154K y debe ser considerada si se desean hacer estudios de acoplamiento sobre esta región.

Al realizar el cribado virtual para la identificación de los posibles fármacos que interfirieran con el blanco (sitio activo de la polimerasa), se encontraron varios candidatos mejores que los ligandos de referencia SOF y CAN reportados como efectivos por Sofia *et al.* (2010) y Lowe *et al.* (2017), respectivamente. Estos estudios previos no poseen datos de energía libre para comparar con los valores obtenidos en este estudio, sin embargo, se tomaron como referencia los valores de -6.17 kcal/mol para SOF y -5.97 kcal/mol para CAN, que resultaron de los diferentes ensayos de acoplamiento. Se observó que

la mayoría de los ligandos probados mostraron una mejora en el valor de energía libre conforme aumentó el número de evaluaciones de acoplamiento molecular; los valores convergieron a partir de diez millones de evaluaciones. Con esto, se generó una galería de moléculas para ser probadas *in silico*, la cual consistió en 34 ligandos que mostraron valores de energía libre entre 1.0 kcal/mol y 2.5 kcal/mol mejores que los controles. Aunado a esto, se encontró un valor adecuado para el número de evaluaciones de acoplamiento para consecuentes investigaciones. No existen datos en otros estudios que relacionen el uso de AutoDockTools para los acoplamientos, el tipo de moléculas, ni la región de acoplamiento de NS5B que se eligieron en la presente investigación.

Durante las simulaciones moleculares acopladas, la mayoría de los ligandos (SOF, ESM, SSM, FSM, ASM, CAN, CAA y CTH) probados permanecieron formando enlaces de hidrógeno la mayor parte del tiempo con la polimerasa en su sitio activo. Sin embargo, el número de enlaces de hidrógeno formados fue diferente entre los ligandos de alrededor de 40 átomos (SOF, ESM, SSM, FSM y ASM) y el grupo de ligandos con menos de 30 átomos (CAN, CAA, CAB y CTH); para el primer grupo existió la formación de dos o tres enlaces de hidrógeno durante el 50% del tiempo de la dinámica, y para el segundo grupo se formaron dos enlaces durante el 70% del tiempo y ningún enlace para el caso del ligando CAB durante el 89% del tiempo de simulación. No existen datos equiparables del tiempo que se mantienen interaccionando estos ligandos acoplados con la polimerasa NS5B pero la importancia de los resultados obtenidos radica en el hecho de que estos ligandos se mantienen unidos al sitio activo de la polimerasa durante más de la mitad del tiempo de simulación y, por tanto, podrían competir con la adición de nucleótidos que se realiza durante el proceso de polimerización del HCV.

En el análisis posterior se evaluaron los porcentajes de ocupancia con aminoácidos de manera individual y se detectó que, durante las diferentes simulaciones de los complejos proteína-ligando, siete aminoácidos (R158, S282, N291, D318, D319, R386 y S367) coincidieron en su participación durante la polimerización, esto concuerda parcialmente con estudios experimentales previos realizados por Appleby *et al.* (2015), donde detallan la replicación del genoma de HCV determinando la estructura cristalográfica de diferentes complejos de la polimerasa con enzimas, moldes, iniciadores e iones metálicos catalíticos; para definir posibles aminoácidos que participan durante diferentes etapas en el proceso de polimerización. De éstos, D318 y D319 son los que tomaron mayor importancia, el primero por aparecer interaccionando más de la mitad del tiempo con tres ligandos diferentes y D319 porque interacciona todo el tiempo con el ligando CAA; S367 se convirtió en el aminoácido menos importante al interaccionar sólo el 1% del tiempo de simulación. Por otro lado, en estudios computacionales desarrollados por Li *et al.* (2010), donde utilizaron un solvente implícito para las simulaciones moleculares, encontraron varios aminoácidos que interaccionan alrededor del 60% del tiempo de simulación (tiempo de simulación total =

8 ns) y coinciden con algunos aminoácidos que también mostraron interacción durante las simulaciones acopladas realizadas en este estudio (tiempo de simulación total = 50 ns); estos cuatro aminoácidos son: R386, S367, C366 y Y415; de los cuales C366 es el que tomó mayor importancia ya que en ambos estudios demostró formar enlaces de hidrógeno durante gran parte de la simulación (60% del tiempo en estudio previo vs 100% del tiempo en el presente estudio).

Además de lo mencionado, se identificaron otros tres aminoácidos, L159, S288 y V405, que interaccionaron más de la quinta parte del tiempo con uno o dos ligandos diferentes; éstos tomaron gran importancia debido a que se encuentran en el sitio catalítico. Sin embargo, no existen reportes previos de su participación en la polimerización; sumado a esto, los tres aminoácidos se encuentran muy conservados de acuerdo a los resultados obtenidos en este trabajo en el análisis a nivel de estructura y de secuencia.

Finalmente, durante la validación de la energía libre de unión, estimada con el uso de un solvente implícito utilizando el método MM/GBSA, los datos normalizados indicaron descartar a los ligandos ASM y CAA, por sus valores bajos de 0.553 y 0.617, respectivamente. Al no existir datos con los cuales comparar la efectividad de los ligandos candidato seleccionados, estos resultados de energía libre de acoplamiento con solvente implícito se cotejaron con los porcentajes de interacción y los resultados de la simulación con solvente explícito previos de este estudio y así se definieron a los ligandos ESM, FSM, SSM y CTH como potenciales candidatos para ser probados experimentalmente, incluso para ser usados como punto de partida para el diseño de un fármaco que interfiera con el proceso de polimerización del genoma viral en mayor grado que los fármacos existentes a la fecha.

Capítulo 5. Conclusiones

El presente trabajo permitió, en primera instancia, encontrar patrones sobre la RNA polimerasa dependiente de RNA viral (RdRp) NS5B, los cuales consistieron en aminoácidos altamente conservados localizados en el sitio activo de la proteína, observados como resultado del alineamiento de las distintas secuencias de la polimerasa. Con esto se confirmó que el sitio seleccionado es un buen candidato como blanco de diferentes fármacos.

La existencia de fármacos que han mostrado tener efecto inhibitorio sobre la polimerasa permitió tener una referencia para realizar la búsqueda de otros ligandos que pudieran presentar iguales o mejores efectos inhibitorios *in silico*. El acoplamiento de alrededor de 1200 ligandos al sitio catalítico de la polimerasa permitió obtener una amplia galería de candidatos para ser simulados.

Las diferentes simulaciones de los complejos, aunado a métodos de validación, mostraron la existencia de cuatro fármacos (ESM, FSM, ASM y CTH) que interfirieron con el blanco de la polimerasa durante un tiempo igual o mayor que los ligandos de referencia. Tres de estos fármacos (ESM, FSM y CTH) tuvieron como ligando de referencia a sofosbuvir (SOF) y uno de ellos (CTH) tuvo como ligando de referencia a canabidiol (CAN). Además, se encontraron aminoácidos interaccionando en el sitio catalítico de los que no se tenía registro previo de su participación formando enlaces de hidrógeno.

Finalmente, esta investigación estableció las bases para dar seguimiento a futuros estudios *in silico* para el diseño de un fármaco basado en la estructura y los patrones de enlace de los ligandos encontrados. Además de mostrar potenciales candidatos para ser probados en estudios *in vitro* para inhibir el proceso de polimerización del virus de la hepatitis C y, por consiguiente, su ciclo de infección.

Literatura citada

- Ackerman, Z., Paltiel, O., Glikberg, F., y Ackerman, E. (1998). Hepatitis C virus in various human body fluids: a systematic review. *Hepatology Research*, 11: 26–40.
- Alva, V, Nam, S., Söding, J. y Lupas, A. (2016) The MPI bioinformatics Toolkit as an integrative platform for advanced protein sequence and structure analysis. *Nucleic Acids Research* 44(1): 410-5.
- Appleby, T., Perry, J., Murakami, E., Barauskas, O., Feng, J., Cho, A., Fox, D., Wetmore, D., McGrath, M., Ray, A., Sofia, M., Swaminathan, S., y Edwards, T. (2015). Structural basis for rna replication by the hepatitis c virus polymerase. *Scientific american*, (347):771–775.
- Berman, H., Westbrook, J.,Feng, Z., Gilliland, G., Bhat, T., Weissig, H., Shindyalov, I. y Bourne. P. (2000) The Protein Data Bank *Nucleic Acids Research*, 28: 235-242.
- Boulant, S., Montserret, R., Hope, R. G., Ratinier, M., Targett-Adams, P., Lavergne, J. P., Penin, F., y McLauchlan, J. (2006). Structural determinants that target the hepatitis C virus core protein to lipid droplets. *Journal of Biological Chemistry*, 281(31): 22236–22247.
- Bressanelli, S., Tomei, L., Rey, F. A., y De Francesco, R. (2002). Structural analysis of the hepatitis C virus RNA polymerase in complex with ribonucleotides. *Journal of virology*, 76(7): 3482–92.
- Case, D., Betz, R., Cerutti, D., Cheatham, T., Darden, T., Duke, R., Giese, T., Gohlke, H., Goetz, A., Homeyer, N., Izadi, S., Janowski, P., Kaus, J., Kovalenko, A., Lee, T., LeGrand, ., Li, P., Lin, C., Luchko, T., Luo, R., Madej, B., Mermelstein, D., Merz, K., Monard, G., Nguyen, H., Nguyen, H.T., Omelyan, I., Onufriev, A., Roe, D., Roitberg, A., Sagui, C., Simmerling, C., Botello-Smith, W., Swails, J., Walker, R., Wang, J., Wolf, R., Wu, X., Xiao L y Kollman, p. (2016). AMBER 2016 - Reference Manual. University of California, San Francisco.
- Chen, L., Gui, C., Luo, X., Yang, Q., Günther, S., Scandella, E., Drosten, C., Bai, D., He, X., Ludewig, B., Chen, J., Luo, H., Yang, Y., Yang, Y., Zou, J., Chen, K., Shen, J., Shen, X., Gu, S., y Thiel, V. (2005). Cinanserin Is an Inhibitor of the 3C-Like Proteinase of Severe Acute Respiratory Syndrome Coronavirus and Strongly Reduces Virus Replication In Vitro Cinanserin Is an Inhibitor of the 3C-Like Proteinase of Severe Acute Respiratory Syndrome Coronavirus and. *Society*, 79(11): 7095–7103.
- Chisari, F. V. (2005). Unscrambling hepatitis C virus–host interactions. *Nature*. 436: 930-932
- Chung, C.-Y., Liu, C.-H., Wang, G.-H., Jassey, A., Li, C.-L., Chen, L., Yen, M.-H., Lin, C.- C., y Lin, L.-T. (2016). (4R,6S)-2-Dihydromenisdaurilide is a Butenolide that Efficiently Inhibits Hepatitis C Virus Entry. *Nature Publishing Group*. 6: 29969.
- Crooks, G., Hon, G., Chandonia, J., y Brenner, S. (2004). NCBI GenBank FTP Site\nWebLogo: a sequence logo generator. *Genome Res*, 14: 1188–1190.
- Cushman, D. W., Cheung, H. S., y Sabo, E. F. (1978). Design of New Antihypertensive Drugs: Inhibitors of Angiotensin-Converting. *Progress in Cardiovascular Diseases*, XXI(3): 176–182.
- de Beer, T., Berka, K., Thornton, J. y Laskowski, R. (2014). PDBsum additions. *Nucleic Acids Res.*, 42, D292-D296

- Ding, Q., vonSchaeuwen, M., y Ploss, A. (2014). The Impact of Hepatitis C Virus Entry on Viral Tropism. *Cell Host & Microbe*, 16(5): 562–568.
- Egger, D., Wölk, B., Gosert, R., Blum, H. E., Moradpour, D., Bienz, K., y Bianchi, L. (2002). Expression of Hepatitis C Virus Proteins Induces Distinct Membrane Alterations Including a Candidate Viral Replication Complex Expression of Hepatitis C Virus Proteins Induces Distinct Membrane Alterations Including a Candidate Viral Replication Complex. *Journal of virology*, 76(12): 5974–5984.
- Elliott, M. (2001). Zanamivir: from drug design to the clinic. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 356(1416): 1885–1893.
- Feld, J. J. (2014). Interferon-Free Strategies with a Nucleoside/ Nucleotide Analogue. *Semin Liver Dis*, 34: 37–46.
- Flint, M., von Hahn, T., Zhang, J., Farquhar, M., Jones, C. T., Balfe, P., Rice, C. M., y McKeating, J. A. (2006). Diverse CD81 Proteins Support Hepatitis C Virus Infection. *Journal of Virology*, 80(22): 11331–11342.
- Götte, M. y Feld, J. J. (2016). Direct-acting antiviral agents for hepatitis C: structural and mechanistic insights. *Nature Reviews Gastroenterology & Hepatology*. 13: 338-351.
- Hazuda, D. J., Burroughs, M., Howe, A. Y. M., Wahl, J., y Venkatraman, S. (2013). Development of boceprevir: A first-in-class direct antiviral treatment for chronic hepatitis C infection. *Annals of the New York Academy of Sciences*, 1291(1): 69–76.
- Kapadia, S. B. y Chisari, F. V. (2005). Hepatitis C virus RNA replication is regulated by host geranylgeranylation and fatty acids. *Proceedings of the National Academy of Sciences of the United States of America*, 102(7): 2561–6.
- Kershenobich Stalnikowitz, D. E. A. (2011). Hepatitis C as a public health problem in Mexico. *Mexico Public Health*, 53(1): S61-S67.
- Khan, A. G., Whidby, J., Miller, M. T., Scarborough, H., Zatorski, A. V., Cygan, A., Price, A. A., Yost, S. A., Bohannon, C. D., Jacob, J., Grakoui, A., y Marcotrigiano, J. (2014). Structure of the core ectodomain of the hepatitis C virus envelope glycoprotein 2. *Nature*. 509: 381-384.
- Kiefer, F., Arnold, K., Künzli, M., Bordoli, L., y Schwede, T. (2009). The SWISS-MODEL Repository and associated resources. *Nucleic Acids Research*, 37(SUPPL. 1): 387– 392.
- Kwong, A. D., Kauffman, R. S., Hurter, P., y Mueller, P. (2011). Discovery and development of telaprevir: an NS3-4A protease inhibitor for treating genotype 1 chronic hepatitis C virus. *Nature Biotechnology*, 29(11): 993–1003.
- Li, T., Froeyen, M. y Herdewijn, P. (2010) Insight into ligand selectivity in HCV NS5B polymerase: molecular dynamics simulations, free energy decomposition and docking. *J. Mol. Model*. 16: 49-59.
- Lindenbach, B. D. y Rice, C. M. (2005). Unravelling hepatitis C virus replication from genome to function. *Nature*. 436(7053): 933-938.
- Liu, F., Shimakami, T., Murai, K., Shirasaki, T., Funaki, M., Honda, M., Murakami, S., Yi, M., Tang, H., y Kaneko, S. (2016). Efficient Suppression of Hepatitis C Virus Replication by Combination Treatment with miR-122 Antagonism and Direct-acting Antivirals in Cell Culture Systems. *Scientific Reports*. 6:30939.

- Lowe, H. I., Toyang, N., y McLaughlin, W. (2017). Potential of cannabidiol for the treatment of viral hepatitis. *Pharmacognosy research*, 9(1): 116–118. Consultado el 10 de febrero de 2016, de <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5330095/?report=reader>
- Messina, J. P., Humphreys, I., Flaxman, A., Brown, A., Cooke, G. S., Pybus, O. G., y Barnes, E. (2015). Global distribution and prevalence of hepatitis C virus genotypes. *Hepatology*. 61: 1-3.
- Michta, M. L., Hopcraft, S. E., Narbus, C. M., Kratovac, Z., Israelow, B., Sourisseau, M., y Evans, M. J. (2010). Species-Specific Regions of Occludin Required by Hepatitis C Virus for Cell Entry. *Journal of Virology*, 84(22): 11696–11708.
- Miller, B., McGee, T., Swails, J., Homeyer, N., Gohlke, H., y Roitberg, A. (2012). MMPBSA.py: an efficient program for end-state free energy calculations. *Chem. Theory Comput.*, 8(9): 3314–3321.
- Moradpour, D., Penin, F., y Rice, C. M. (2007). Replication of hepatitis C virus. *Nature Reviews Microbiology*, 5(6): 453–463.
- Morris, G. M., Goodsell, D. S., Halliday, R. S., Huey, R., Hart, W. E., Belew, R. K., y Olson, A. J. (1998). Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function. *Journal of Computational Chemistry*, 19(14): 1639–1662.
- Nawaz, A., Zaidi, S. F., Usmanghani, K., y Ahmad, I. (2015). Concise review on the insight of hepatitis C. *Journal of Taibah University Medical Sciences*, 10(2): 132–139. 25
- Paul, D., Madan, V., y Bartenschlager, R. (2014). Hepatitis C virus RNA replication and assembly: Living on the fat of the land. *Cell Host and Microbe*, 16(5): 569–579.
- Pavlovic, D., Neville, D. C., Argaud, O., Blumberg, B., Dwek, R. A., Fischer, W. B., y Zitzmann, N. (2003). The hepatitis C virus p7 protein forms an ion channel that is inhibited by long-alkyl-chain iminosugar derivatives. *Proc Natl Acad Sci U S A*, 100(10): 6104–6108.
- Petruzzello, A., Marigliano, S., Loquercio, G., Cozzolino, A., y Cacciapuoti, C. (2016). Global epidemiology of hepatitis C virus infection: An up-date of the distribution and circulation of hepatitis C virus genotypes. *World Journal of Gastroenterology*, 22(34): 7824–7840.
- Ploss, A., Evans, M. J., Gaysinskaya, V. A., Panis, M., You, H., de Jong, Y. P., y Rice, C. M. (2009). Human occludin is a hepatitis C virus entry factor required for infection of mouse cells. *Nature*, 457(7231): 882–886.
- Prada-Gracia, D., Huerta-Yépez, S., y Moreno-Vargas, L. M. (2016). Application of computational methods for anticancer drug discovery, design, and optimization. *Boletín Médico del Hospital Infantil de México*, 73(6): 411–423.
- Pundir, S., Martin, M., O'Donovan, C. (2017) UniProt Protein Knowledgebase. En: Wu, C. (Ed) *Protein Bioinformatics. Methods in Molecular Biology*. Humana Press, New York, NY Vol. 1558
- Ranjith-Kumar, C., Kao, C. 2006. Biochemical Activities of the HCV NS5B RNA-Dependent RNA Polymerase. En: Tan, S. (Ed), *Hepatitis C Viruses: Genomes and Molecular Biology*. Horizon Bioscience. Norfolk (UK). pp 293-310.

- Rojas, A., Del Campo, J. A., Clement, S., Lemasson, M., García-Valdecasas, M., Gil- Gómez, A., Ranchal, I., Bartosch, B., Bautista, J. D., Rosenberg, A., Negro, F., y Romero-Gómez, M. (2016). Effect of Quercetin on Hepatitis C Virus Life Cycle: From Viral to Host Targets. *Scientific Reports*. 6: 31777.
- Rydberg, E. H., Cellucci, A., Bartholomew, L., Mattu, M., Barbato, G., Ludmerer, S. W., Graham, D. J., Altamura, S., Paonessa, G., De Francesco, R., Migliaccio, G., y Carfí, A. (2009). Structural Basis for Resistance of the Genotype 2b Hepatitis C Virus NS5B Polymerase to Site A Non-Nucleoside Inhibitors. *Journal of Molecular Biology*, 390(5): 1048–1059.
- Schrödinger. 2016. Consultado en octubre de 2016, de <https://www.schrodinger.com/citations>
- Sofia, M., Bao, D., Chang, W., Du, J., Nagarathnam, D., Rachakonda, S., Reddy, P., Ross, B., Wang, P., Zhang, H., Bansal, S., Espiritu, C. y Keilman, M. (2010). Discovery of a β -D-20-Deoxy-20-r-fluoro-20- β -C-methyluridine Nucleotide Prodrug (PSI-7977) for the Treatment of Hepatitis C Virus. *J. Med. Chem.* 53: 7202–7218
- Angela M. Lam, Holly M. Micolochick Steuer, Congrong Niu, Michael J. Otto, and Phillip A. FurmanStanaway, J. D., Flaxman, A. D., Naghavi, M., Fitzmaurice, C., Vos, T., Abubakar, I., Abu-Raddad, L. J., Assadi, R., Bhalal, N., Cowie, B., Forouzanfar, M. H., Groeger, J., Hanafiah, K. M., Jacobsen, K. H., James, S. L., MacLachlan, J., Malekzadeh, R., Martin, N. K., Mokdad, A. A., Mokdad, A. H., Murray, C. J. L., Plass, D., Rana, S., Rein, D. B., Richardus, J. H., Sanabria, J., Saylan, M., Shahraz, S., So, S., Vlassov, V. V., Weiderpass, E., Wiersma, S. T., Younis, M., Yu, C., El Sayed Zaki, M., y Cooke, G. S. (2016). The global burden of viral hepatitis from 1990 to 2013: findings from the Global Burden of Disease Study 2013. *The Lancet*. 6736(16): 30579-7
- Sterling, T. e Irwin, J. J. (2015). ZINC 15 - Ligand Discovery for Everyone. *Journal of Chemical Information and Modeling*, 55(11): 2324–2337.
- Vera de León, L., Juárez Navarro, J. A., Díaz Gómez, M., Méndez Navarro, J., Chirino Sprung, R. A., Dehesa Violante, M., Casillas Davila, L., Rizo Robles, M. T., Torres Ibarra, R., Cano Domínguez, C., Nava Estrada, J., Ramos Gómez, M., Hernández, M. E., Wong González, S., Félix Tamayo, M., Corona Lechuga, M. D. C., Zárate Negrete, A. R., Rangel Jiménez, M., Rodríguez Hernández, H., González Ortiz, V., Tirado Estrella, M. P., Villanueva Carreto, M. L. d. A., Orta Flores, R., Manteca Argumedo, J. L., Vázquez Avila, I., González Macias, J., Razcón Hernández, O., Torres López, P., y Bazán Pérez, C. (2005). Panorama epidemiológico y situacional de la hepatitis C en México. *Revista de gastroenterología de Mexico*, 70(1): 25–32.
- Vu, L. A., Thi, P., Quyen, C., y Huong, N. T. (2015). In silico Drug Design : Prospective for Drug Lead Discovery. 4(10): 60–70.
- Waheed, Y., Saeed, U., Anjum, S., Afzal, M. y Ashraf, M. (2012). Development of Global Consensus Sequence and Analysis of Highly Conserved Domains of the HCV NS5B Protein. *Hepatitis Monthly*. 12(9): e6142.
- Xu, Z., Choi, J., Lu, W., y Ou, J.-h. (2003). Hepatitis C Virus F Protein Is a Short-Lived Protein Associated with the Endoplasmic Reticulum. *Journal of Virology*. 77(2): 1578-1583.