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

Programa de Posgrado en Ciencias de la Vida con orientación en Microbiología

Estudio *in silico* de la proteína E6 del virus del papiloma humano para identificar potenciales fármacos antivirales

Tesis

para cubrir parcialmente los requisitos necesarios para obtener el grado de

Maestro en Ciencias

Presenta:

Joel Ricci López

Ensenada, Baja California, México

2017

Tesis defendida por

Joel Ricci López

y aprobada por el siguiente Comité

Dr. Sergio Andrés Águila Puentes

Codirector del Comité

Dr. Carlos Alberto Brizuela Rodríguez

Codirector del Comité

Dra. María Asunción Lago Lestón

Dra. Rosa Reyna Mouriño Pérez

Dr. Israel Marck Martínez Pérez

Dra. Clara Elizabeth Galindo Sánchez

Coordinador del Programa de Posgrado en Ciencias de la Vida

Dra. Rufina Hernández Martínez

Director de Estudios de Posgrado

Joel Ricci López © 2017

Queda prohibida la reproducción parcial o total de esta obra sin el permiso formal y explícito del autor y director de la tesis.

Resumen de la tesis que presenta Joel Ricci López como requisito parcial para la obtención del grado de Maestro en Ciencias en de la Vida con orientación en Microbiología.

Estudio *in silico* de la proteína E6 del virus del papiloma humano para identificar potenciales fármacos antivirales

Resumen aprobado por:

Dr. Sergio Andrés Águila Puentes
Codirector de Tesis

Dr. Carlos Alberto Brizuela Rodríguez
Codirector de Tesis

La infección por el virus del papiloma humano (VPH) es una de las enfermedades de transmisión sexual más comunes, y algunos VPH han sido identificados como el agente etiológico de diversos tipos de cáncer. Actualmente existen vacunas capaces de prevenir la infección por estos virus. Sin embargo, para el caso de personas ya infectadas, aún sigue siendo necesario el desarrollo de un tratamiento farmacológico eficaz y seguro para combatir la infección y tratar sus efectos oncogénicos. Uno de los blancos terapéuticos más estudiados del VPH ha sido la oncoproteína E6, identificada como uno de los factores clave en la inmortalización celular y la progresión tumoral en células VPH-positivas. E6 es capaz de promover la degradación de la proteína de supresión tumoral p53, a través de la interacción con la proteína celular ubiquitina ligasa E6AP. Por lo anterior, impedir el acoplamiento E6-E6AP es una de las principales estrategias para inhibir la viabilidad y proliferación de las células infectadas.

En el presente trabajo se propuso una metodología *in silico* enfocada en el descubrimiento de moléculas capaces de bloquear esta interacción. Para ello, se partió de la estructura cristalográfica de la proteína E6 VPH-16 acoplada al motivo *LxxLL* de E6AP, y de 34,804 moléculas obtenidas de la base de datos *ZINC15* a partir de su similitud estructural con 26 compuestos cuya actividad anti-VPH ha sido evaluada previamente. Después, se realizó un proceso conocido como *Cribado Virtual* mediante la predicción de la propiedades ADMETox de las moléculas y la *Simulación de Acoplamiento Molecular* con la proteína E6. Con ello, fue posible filtrar compuestos candidato con características farmacocinéticas favorables y con capacidad de interactuar con E6. Por último, cuatro de estos compuestos fueron seleccionados y, mediante *Simulación de Dinámica Molecular*, se corroboró la estabilidad de su acoplamiento con E6, y su efecto en la reducción de la afinidad de E6 hacia E6AP. Por lo anterior, la metodología empleada y las moléculas identificadas en este trabajo representan un nuevo punto de partida en el desarrollo de fármacos anti-VPH.

Palabras Clave: **Virus del Papiloma Humano, Proteína E6, Cribado Virtual Molecular, Acoplamiento Molecular, Dinámica Molecular**

Abstract of the thesis presented by Joel Ricci López as a partial requirement to obtain the Master of Science degree in Life Sciences with orientation in Microbiology.

In silico study of Human Papillomavirus E6 protein to identify antiviral drugs

Abstract approved by:

Dr. Sergio Andrés Águila Puentes
Thesis Co-Director

Dr. Carlos Alberto Brizuela Rodríguez
Thesis Co-Director

Human papillomavirus (HPV) is one of the most common sexually transmitted infections, and some HPV strains have been identified as the etiologic agent of some types of cancer. Currently, there are vaccines capable of preventing infection by these viruses. However, for people that have already been infected, it is still necessary a drug-based treatment againts the infection and its oncogenic effects. The oncoprotein E6 is one of the most studied therapeutic targets of HPV, it has been identified as a key factor in cell immortalization and tumor progression in HPV-positive cells. E6 is able to promote the degradation of p53, a tumor suppression protein, through the interaction with the cellular ubiquitin ligase, E6AP. Therefore, preventing E6-E6AP docking is one of the main strategies to inhibit the viability and proliferation of infected cells.

The present study proposes an *in silico* methodology focused on the discovery of inhibitors of the E6-E6AP interaction. To achive this goal, we started with the crystallographic structure of the E6 HPV-16 protein docked to the LxxLL motif of E6AP, and with 34,804 molecules from the ZINC15 database obtained by their structural similarity with 26 compounds, whose anti-HPV activity has been previously evaluated. Therefore, a process known as Virtual Screening was performed by predicting the ADMETox properties of the molecules and preforming docking simulations with the E6 protein. Then, it was possible to find candidate compounds with favorable pharmacokinetic characteristics and capable to dock with E6. Finally, four of these compounds were selected, and their stability in the E6 docked complex and their effect in the inhibition of the E6-E6AP interaction were corroborated by *Molecular Dynamics Simulation*. Therefore, this methodology and the identified molecules represent a new starting point in the development of anti-HPV drugs.

Keywords: **Human Papillomavirus, E6 protein, Virtual Screening, Molecular Docking, Molecular Dynamics**

Dedicatoria

A mi mamá Liz, que ha sido el pilar de mi vida. Estoy muy orgulloso de ser tu hijo.

Agradecimientos

Al Centro de Investigación Científica y de Educación Superior de Ensenada, sus investigadores y profesores, y a todo su personal que ahí trabaja.

Al Consejo Nacional de Ciencia y Tecnología (CONACyT) por brindarme el apoyo económico para realizar mis estudios de maestría.

A mis directores de tesis, el Dr. Sergio Águila y el Dr. Carlos Brizuela, por recibirme y darme la oportunidad de formar parte de sus grupos de investigación. Muchas gracias por su apoyo y sus asesorías, por brindarme los recursos necesarios para realizar este trabajo, y todo el aprendizaje obtenido. Ha sido un placer colaborar con ustedes.

A los miembros del comité de tesis, la Dra. Asunción Lago, la Dra. Rosa Mouriño y el Dr. Israel Pérez, por sus valiosas observaciones y correcciones a lo largo del desarrollo de este trabajo.

Gracias al Dr. Joel Alderete de la Universidad de Concepción, Chile, y a la Dra. Verónica Jiménez de la Universidad Andrés Bello, Chile, y a sus respectivos equipos de trabajo. Su apoyo, recursos, enseñanzas y colaboración fueron indispensables para el desarrollo de este trabajo, y para mi persona.

Gracias al Dr. Abraham Vidal Limón, y al resto del equipo de trabajo del Dr. Segio Águila. Ha sido una gran retroalimentación conocer y colaborar con compañeros con diferentes formaciones académicas. Lo mismo para los compañeros del grupo de trabajo del Dr. Carlos Brizuela y del laboratorio de algoritmos de Ciencias de la Computación. Agradezco mucho sus charlas, enseñanzas y consejos, y que siempre estuvieran ahí para ayudarme a resolver dudas.

Gracias a todos mis compañeros y amigos de generación, en especial a Víctor Vargas cuya compañía y apoyo hicieron muy llevadero el desarrollo de este trabajo. Fue placer cursar estos dos años, aprendiendo de ellos y junto a ellos. Mucho éxito en todo lo que se propongan.

A mi papá, Joel, por ser un gran ejemplo para mí.

A Jaky, mi novia y compañera de vida. Gracias por tu paciencia y tu infinito apoyo, ha sido un honor compartir una aventura más junto a ti. Gracias también a tu familia por todo el cariño y apoyo.

A mi familia, a mi mamá Liz, mi abuelita Teresa, a mi adorado hermano Oscar, y a toda mi enorme y hermosa familia. Gracias por estar siempre conmigo y confiar en mí. Los amo.

Tabla de contenido

Página

Resumen en español . . . . . **ii**

Resumen en inglés . . . . . **iii**

Dedicatoria . . . . . **iv**

Agradecimientos . . . . . **v**

Lista de figuras . . . . . **ix**

Lista de tablas . . . . . **xi**

1. Introducción **1**
1.1. Antecedentes 2
1.1.1. Clasificación y prevalencia de los virus del papiloma humano 2
1.1.2. Virología y ciclo replicativo del VPH 3
1.1.3. El papel de la oncoproteína E6 en la carcinogénesis 5
1.1.4. Fármacos para el tratamiento del VPH de alto riesgo 6
1.1.5. Descubrimiento *in silico* de fármacos antivirales 7
1.2. Justificación 9
1.3. Hipótesis 9
1.4. Objetivos 10
1.4.1. Objetivo General 10
1.4.2. Objetivos Particulares 10

2. Metodología **11**
2.1. OBJETIVO 1: Análisis secuencia-estructura de la proteína E6 12
2.1.1. Análisis de secuencias de la proteína E6 12
2.1.2. Caracterización del *pocket* de unión a E6AP 12
2.2. OBJETIVO 2: Refinamiento estructural de la proteína E6 12
2.2.1. Modelaje homólogo de la proteína E6 12
2.2.2. Refinamiento por dinámica molecular de la proteína E6 13
2.2.2.1. Preparación de la proteína 13
2.2.2.2. Minimización, *Simulated annealing* y equilibrado 13
2.2.2.3. Simulación de dinámica molecular 14
2.2.3. Análisis de las trayectorias de dinámica molecular 15
2.2.3.1. RMSD: *Root Mean Square Deviation* 15
2.2.3.2. RMSF: *Root Mean Square Fluctuation* 15
2.2.3.3. Análisis de componentes principales y *clustering* Jerárquico 16
2.3. OBJETIVO 3: Cribado virtual de las propiedades ADMETox de los ligandos . . 17
2.3.1. Selección del espacio químico 17
2.3.2. Refinamiento estructural de las moléculas 18
2.3.3. Predicción de propiedades ADME-Tox 18
2.3.4. Cribado virtual utilizando los descriptores ADME-Tox 18
2.4. OBJETIVO 4: Cribado virtual basado en acoplamiento molecular 20
2.4.1. Preparación del receptor y los ligando 20
2.4.2. Acoplamiento molecular con *Atodock Vina* 21
2.4.3. Acoplamiento molecular con *Autodock 4.2* 21
2.5. OBJETIVO 5: Validación de las moléculas candidato 22

2.5.1. Validación del acoplamiento con MM/GBSA y MM/PBSA 22
2.5.2. Dinámica molecular del complejo proteína-ligando 23

3. Resultados 24
3.1. OBJETIVO 1: Análisis secuencia-estructura de la proteína E6 24
3.1.1. Determinación de los tipos de VPH a analizar 24
3.1.2. Obtención y alineamiento múltiple de secuencias 24
3.1.3. Comparación secuencia-estructura de la proteína E6 26
3.1.4. Caracterización del *pocket* de unión a E6AP 27
3.2. OBJETIVO 2: Refinamiento estructural de la proteína E6 29
3.2.1. Modelaje homólogo de la proteína E6 30
3.2.2. Refinamiento por dinámica molecular 30
3.2.2.1. Preparación de los sistemas *E6 sola* y *E6-hx* 31
3.2.2.2. Minimización, *Simulated annealing* y equilibrado 32
3.2.2.3. Producción de dinámica molecular 33
3.2.2.4. Análisis de las trayectorias de dinámica molecular 35
3.2.2.5. Análisis de componentes principales y *clustering* jerárquico 37
3.3. OBJETIVO 3: Cribado virtual de las propiedades ADMETox de los ligandos . . 40
3.3.1. Obtención de las moléculas de referencia 40
3.3.2. Cribado virtual utilizando los descriptores ADME-Tox 43
3.4. OBJETIVO 4: Cribado virtual basado en acoplamiento molecular 44
3.4.1. Acoplamiento Molecular con *Autodock Vina* 44
3.4.2. Acoplamiento molecular con *Autodock 4* 45
3.5. OBJETIVO 5: Validación de las moléculas candidato 47
3.5.1. Validación del acoplamiento con MM/GBSA y MM/PBSA 47
3.5.2. Validación del acoplamiento con dinámica molecular 48

4. Discusión 54
4.1. OBJETIVO 1: Análisis secuencia-estructura de la proteína E6 54
4.2. OBJETIVO 2: Refinamiento de la estructura proteína E6 55
4.3. OBJETIVO 3: Cribado virtual basado en propiedades ADMETox 56
4.4. OBJETIVO 4: Cribado virtual basado en acoplamiento molecular 58
4.4.1. Acoplamiento molecular con *Autodock Vina* y *Autodock 4* 58
4.5. OBJETIVO 5: Validación con simulación de dinámica molecular 59
4.5.1. Reevaluación de los acoplamientos E6+lig con los métodos MM/GBSA y MM/PBSA 59
4.5.2. SDM de los complejos *E6+lig* y *[E6+lig]+hx* 60
4.5.2.1. SDM del acoplamiento proteína E6+luteolina 60
4.5.2.2. SDM del acoplamiento proteína E6+ligando 5 61
4.5.2.3. SDM del acoplamiento proteína E6+ligando 32 62
4.5.2.4. SDM del acoplamiento proteína E6+ligando 70 63
4.5.2.5. SDM del acoplamiento proteína E6+ligando 82 64

5. Conclusión 65

Literatura citada 67

A. Metodología y Resultados 77

A.1. Protocolo de dinámica molecular 77
A.2. Análisis de trayectorias 78
A.3. Alineamiento múltiple de secuencias. 80
A.4. Modelado homólogo y dinámica molecular de E6 VPH-11 81
A.5. Descriptores para el cribado virtual ADMETox 84
A.6. Evaluación de *Autodock 4* y *Autodock Vina* 85
A.7. Dinámica molecular de los complejos E6-hx, E6 VPH-16 y E6 VPH-11 89
A.8. Dinámica molecular de los complejos E6-lig 91

Lista de figuras

Figura Página

1. Metodología empleada para identificar moléculas con potencial de inhibir la interacción E6-E6AP . . . 11
2. Alineamiento múltiple de secuencias de la proteína E6. . . . 25
3. Fenogramas generados con las secuencias de las proteínas L1 y E6 . . . 26
4. Representación esquemática de la estructura cristalográfica de la proteína E6 (modelo 4XR8) . . . 27
5. Estructura de la proteína E6 (VPH-6) unida a la hélice *LxxLL* de E6AP . . . 28
6. Cavidad en la proteína E6 correspondiente al *pocket* de unión a la hélice *LxxLL* . . . 29
7. Logo de la subsecuencia del *pocket* de unión a E6AP con los 21 tipos de VPH analizados . . . 30
8. Preparación de la proteína E6 para la simulación de dinámica molecular . . . 31
9. Minimización, *Simulated annealing* y equilibrado del sistema *E6 sola* . . . 32
10. Propiedades físicas del sistema *E6 sola* evaluadas en la fase de producción de SDM 34
11. RMSD de la proteína E6 en los tres ensayos del sistema *E6 sola*. . . . 35
12. RMSD de la proteína E6 en los tres ensayos del sistema *E6+hx*. . . . 36
13. RMSF de la proteína E6 en los sistemas *E6 sola* y *E6-hx*. . . . 37
14. Análisis de componentes principales de las trayectorias de SDM de *E6 sola* . . . 38
15. Clusters conformacionales seleccionados en cada trayectoria de *E6 sola*. . . . 39
16. Conformaciones representativas de la dinámica de la proteína E6. . . . 40
17. Gráfico de Tanimoto comparando las 26 moléculas de referencia. . . . 41
18. *Pocket* de unión a E6AP de las conformaciones de E6 utilizadas para la SAM. . . . 45
19. Anclaje molecular de moléculas pequeñas a la proteína E6 VPH-16 . . . 46
20. Análisis de la interacción de los 100 ligandos obtenidos con *Autodock 4*. . . . 48
21. Diagramas de interacción con los residuos E6 de las cuatro moléculas candidato. . . 49
22. Acoplamiento y dinámica molecular de los sistemas *E6-lig*. . . . 51
23. Dinámica molecular de los complejos *proteína-hélice-ligando* (50ns) . . . 52
24. Enlaces de hidrógeno entre el *pocket* de la proteína E6 con un ligando acoplado y la hélice *LxxLL* . . . 53

A.1. Alineamiento múltiple con las 21 secuencias de la proteína E6 de distintos VPH. . . 80
A.2. SDM de la proteína E6 VPH-11 (BR) . . . 82
A.3. Dendogramas resultantes del *clustering* Jerárquico de los ensayos de SDM del sistema *E6 sola*. . . . 83

Figura Página

A.4. Evaluación de *Autodock 4* y *Autodock Vina* 86
A.5. Análisis de los mejores 100 ligandos identificados con AD4 87
A.6. Gráfico de Tanimoto mostrando el grado de similitud entre las mejores 30 moléculas 88
A.7. Dinámica molecular de los complejos E6-hx (50 ns) 89
A.8. Enlaces de hidrógeno entre el *pocket* de la proteínas E6 VPH-16 y E6 VPH-11 con la hélice *LxxLL* 90
A.9. Dinámica molecular del complejo *E6-Luteolina* (50 ns) 91
A.10. Dinámica molecular del complejo *E6-Lig5* (50 ns) 92
A.11. Dinámica molecular del complejo *E6-Lig32* (50 ns) 93
A.12. Dinámica molecular del complejo *E6-Lig70* (50 ns) 94
A.13. Dinámica molecular del complejo *E6-Lig82* (50 ns) 95

Lista de tablas

Tabla Página

1. Descriptores de *QikProp* utilizados para el filtrado de moléculas según sus características farmacológicas predichas. 19

2. Moléculas de referencia elegidas a partir de la revisión bibliográfica realizada. 42

2. Moléculas de referencia elegidas a partir de la revisión bibliográfica realizada. 43

3. Número de moléculas filtradas en cada descriptor de *Qikprop* utilizado. . . . 44

4. Resultados principales de las SDM de los sistemas *E6+lig* y *[E6+lig]+hx*. . . 50

Capítulo 1. Introducción

Los papilomavirus son una familia de virus de ADN que infectan células epiteliales, causando lesiones cutáneas y en la mucosa de una gran variedad de vertebrados, principalmente en mamíferos (Baleja *et al.*, 2006). En el caso del ser humano, se han identificado más de 200 especies de papilomavirus capaces de infectarlo, todas éstas conocidas en conjunto como virus del papiloma humano (VPH) (Chouhy *et al.*, 2013). Las especies de VPH poseen un marcado tropismo celular, pero en general la infección por estos virus se presenta en la piel y en las membranas mucosas de los tractos respiratorio, gastrointestinal y urogenital (Zanier *et al.*, 2012).

Alrededor de 50 especies de VPH son capaces de infectar las áreas anogenitales, por lo que la infección por estos virus se considera como una de las enfermedades de transmisión sexual más comunes, pudiéndose transmitir a través de contacto piel a piel (Smith, 2013). En la mayoría de los casos, los VPH resultan asintomáticos o causan síntomas menores en las personas infectadas. Sin embargo, determinadas especies son capaces de causar verrugas o condilomas (WHO, 2014). Y en casos más extremos, algunas han sido señaladas como agentes causales de neoplasias que pueden dar lugar a cáncer cervical, de vulva, vagina, ano, pene, cuello o garganta (Lavezzo *et al.*, 2016).

La prevalencia de estos VPH oncogénicos, identificados como VPH de alto riesgo, implica un problema de salud pública global que afecta principalmente a los países en vías de desarrollo (WHO, 2014). Pese a esto, aún no se ha podido desarrollar y validar clínicamente ningún tratamiento farmacológico efectivo, seguro y económico. Por ello es necesario la identificación de nuevos fármacos que muestren una mejor eficacia en los ensayos clínicos, con la menor cantidad de efectos secundarios (Hampson *et al.*, 2015).

Sin embargo, el descubrimiento y diseño de nuevos medicamentos es un proceso altamente demandante de recursos económicos y de tiempo. Por lo que una alternativa ha sido el empleo de tecnologías y plataformas computacionales que, mediante simulaciones de acoplamiento y dinámica molecular, permiten predecir el comportamiento de moléculas candidato como potenciales antivirales (Cerqueira *et al.*, 2015; Marz *et al.*, 2014; Murgueitio *et al.*, 2012). Estudios previos han demostrado la eficacia de estas metodologías al reducir en tiempo y costo económico el descubrimiento de nuevos fármacos antivirales candidatos a pruebas clínicas, como ha sido el caso del virus de la influenza (Kirchmair *et al.*, 2011; Nguyen *et al.*, 2009), el virus de inmunodeficiencia humana (VIH) (Chang *et al.*, 2010; Curreli *et al.*, 2011; Schames *et al.*, 2004), el virus de Epstein-Barr

(Li *et al.*, 2010) y algunos miembros de la familia Flaviviridae (Mirza *et al.*, 2016; Umamaheswari *et al.*, 2011; Podvinec *et al.*, 2010).

Por lo anterior, y debido a la prevalencia del VPH y la evidente demanda por tratamientos contra el mismo, el presente trabajo tiene como objetivo emplear esta metodología computacional para identificar y proponer fármacos dirigidos a la oncoproteína E6 del VPH, cuya participación en la replicación viral y transformación celular la convierten en un blanco farmacológico ideal. Fármacos que posteriormente puedan ser evaluados mediante estudios *in vitro* y pruebas clínicas en miras del desarrollo de un tratamiento contra el VPH.

1.1. Antecedentes

1.1.1. Clasificación y prevalencia de los virus del papiloma humano

Filogenéticamente los VPH no forman parte de un único grupo monofilético por lo que son clasificados dentro de cinco géneros distintos de la familia Papilomaviridae: *Alfapapilomavirus*, *Betapapilomavirus*, *Gamapapilomavirus*, *Mupapilomavirus* y *Nupapilomavirus*. Esta clasificación taxonómica ha sido establecida con base en el tipo de lesiones que causan, al tropismo celular y, sobretodo, a la similitud de las secuencias genéticas (Doorbar *et al.*, 2012; Villiers *et al.*, 2004). De este modo, dentro de cada género, los genotipos cuyas secuencias del gen L1 (proteína de cápside mayor) difieren entre sí en al menos un 10 % son considerados como tipos de VPH distintos, y su nomenclatura es únicamente numérica. Por otro lado, los subtipos (o linajes) de cada tipo de VPH son aquellos cuyas secuencias del gen L1 comparten como mínimo un 90 % de identidad y, a nivel de todo el genoma, un 99 % de identidad (Burk *et al.*, 2013; Bernard *et al.*, 2010).

De los cinco géneros de VPH, el género Alfapapilomavirus engloba a la mayoría de las especies capaces de causar lesiones epiteliales, verrugas, neoplasias y cáncer en tejidos orales y anogenitales (Doorbar *et al.*, 2012). Subsecuentemente, los tipos de Alfapapilomavirus pueden categorizarse en dos grupos dependiendo del tejido que afectan; los VPH cutáneos y los VPH de la mucosa, grupo al que pertenecen los VPH de mayor importancia clínica. Finalmente, una última clasificación permite distinguir a los VPH de mucosa en VPH de alto riesgo (VPH-AR) y en VPH de bajo riesgo (VPH-BR), según la gravedad de las lesiones que producen (Bernard *et al.*, 2010).

Como su nombre lo indica, los VPH-AR son los tipos de papilomavirus que mayor riesgo representan al poseer un potencial oncogénico en los seres humanos. Actualmente los VPH-AR identificados son los VPH-: 16, 18, 31, 33, 35, 39, 45, 51, 52, 56, 58, 59, 66, 68, 73 y 82 (Allison y

Maleki, 2015). Para los cuales se ha demostrado, experimental y clínicamente, un papel causal en el desarrollo de cáncer cervical, pene, de ano y orofaríngeo (Lavezzo *et al.*, 2016). De estos genotipos, el VPH-16 y el VPH-18 son los más comunes, siendo responsables, respectivamente, del 63 % y el 16 % de los casos de cáncer cervicouterino. Y en conjunto, del 63-80 % de los casos de cáncer de pene, del 85 % de los casos de cáncer de vulva/vagina, del 93 % de los casos de cáncer de ano y del 89-95 % de los casos de cáncer de cuello y cabeza (Bruni *et al.*, 2016; Chaturvedi, 2010).

Desde el punto de vista epidemiológico las infecciones por VPH de alto riesgo representan un problema de salud pública global capaz de afectar a hombres y mujeres en edad reproductiva (WHO, 2014). Sin embargo, estas últimas tienden a ser más susceptibles a complicaciones por infección de VPH debido a la estrecha relación de estos virus con el cáncer cervicouterino (zur Hausen, 2009). Ya que, por un lado, se ha estimado que prácticamente todos los casos de esta enfermedad (99.7 %) son causados por estos agentes infecciosos (Beaudenon y Huibregtse, 2008). Y por el otro, el cáncer cervical es el segundo tipo de cáncer más frecuente en mujeres a nivel mundial; con un estimado global de cerca 500,000 diagnósticos y 270,000 muertes cada año (Bruni *et al.*, 2014).

No obstante, la mayoría de las mujeres que presenta una infección por VPH-AR no desarrolla ningún tipo de cáncer y el virus es, especulativamente, eliminado del cuerpo dentro de los primeros dos años de la infección (WHO, 2014). Pese a esto, factores ajenos al virus, relacionados con la situación médica y socieconómica del paciente, favorecen su persistencia y el desarrollo de carcinomas (Bruni *et al.*, 2014); con lo que la mayoría de los casos de cáncer cervical (>80 %) se presentan en países en vías de desarrollo (WHO, 2014). En el caso de México esto implica un promedio de 13,960 casos nuevos anuales de cáncer cervicouterino (Bruni *et al.*, 2016), con una mayor prevalencia de los tipos VPH-16, VPH-18, VPH-31, VPH-45 y VPH-58 (Flores-Miramontes *et al.*, 2015).

1.1.2. Virología y ciclo replicativo del VPH

La partícula viral de los VPH carece de envoltura y está formada únicamente por una cápside de geometría icosaédrica de aproximadamente 55 nm, la cual está constituida por las proteínas L1 y L2 (Doorbar *et al.*, 2012). En el interior de la cápside se encuentra una única copia del genoma que consiste en una doble hebra de ADN circular de aproximadamente 8 kb, por lo que se agrupan en la clase I de la clasificación de Baltimore. El genoma se estructura en tres regiones principales:

región temprana (E: *early*), región tardía (L: *late*) y región larga de control (LCR: *long control region*) (Zheng y Baker, 2006).

La región temprana (E) ocupa la primera mitad del genoma del VPH en el extremo 5', y posee seis marcos abiertos de lectura (ORF: *open reading frames*) que codifican, respectivamente, para las proteínas tempranas E1, E2, E4, E5, E6 y E7. La región tardía codifica para las proteínas estructurales de la cápside viral L1 (proteína de cápside mayor) y L2 (proteína de cápside menor) (Zheng y Baker, 2006). Finalmente, la región larga de control, también conocida como región reguladora ascendente, consta de 400 a 1,000 pares de bases y posee secuencias de regulación tipo *cis* que actúan como blanco de varios factores de transcripción esenciales para la replicación del ADN viral (Schmitz *et al.*, 2012).

Debido a la diversidad de los tipos de VPH, y a factores extrínsecos al virus como las hormonas y citosinas, no existe un único mecanismo general de infección celular y de proliferación del virus (Doorbar *et al.*, 2012). Sin embargo, para los tipos de VPH que afectan la mucosa, la infección comienza con la entrada de la partícula a través de micro laceraciones en la piel hasta alcanzar la capa basal del epitelio (Allison y Maleki, 2015). Una vez ahí, el virus infecta los queratinocitos del epitelio escamoso estratificado adentrándose en ellos a través de una endocitosis mediada por receptor. Posteriormente, dentro de la célula, los viriones se desensamblan y el ADN vírico, unido a la proteína L2, es internalizado en el núcleo celular (Doorbar *et al.*, 2012).

Dentro del núcleo de la célula, el genoma del VPH se mantiene en forma de episoma e inicia una primer etapa de amplificación a bajo nivel regulada por la actividad de las proteínas E1 y E2 (Maglennon *et al.*, 2011). Durante esta fase, el VPH toma el control del ciclo celular estimulando la proliferación de la célula infectada (en los tipos AR) y con ello la diseminación del genoma viral en las células hijas. Para lograr esto, la expresión de las proteínas E6 y E7 tiene un papel fundamental al promover un estado vegetativo de constante replicación de la célula y al permitirle evadir la senescencia celular (Langsfeld y Laimins, 2016; Zheng y Baker, 2006).

Posteriormente, cuando el queratinocito infectado entra en el estadio final de diferenciación, al pasar de la fase S a la fase G2, el virus incrementa sus niveles de replicación con ayuda de las proteínas E1, E2, E4 y E5; aumentando de dos a tres órdenes de magnitud el número de episomas por célula (Bodily y Laimins, 2011). Tras la replicación, da comienzo la expresión de las proteínas L1 y L2 las cuales son reclutadas dentro del núcleo en donde llevan a cabo la encapsidación del genoma. Por último, la maduración de estas partículas virales ocurre en los queratinocitos

más superficiales y senescentes del epitelio, a partir de los cuales las partículas infecciosas son liberadas (Doorbar *et al.*, 2012; Langsfeld y Laimins, 2016).

1.1.3. El papel de la oncoproteína E6 en la carcinogénesis

Durante el ciclo de vida de los VPH la función de las proteínas E6 y E7 es estimular el ciclo celular y promover la proliferación de las células epiteliales infectadas. Sin embargo, en los tipos de VPH de alto riesgo, estas proteínas muestran una mayor capacidad de interacción con múltiples proteínas celulares (Doorbar *et al.*, 2012). De esta forma, son capaces de inhibir o activar rutas de señalización cuya regulación brinda al virus una mayor capacidad de evasión de las defensas del huésped y de prevención de la apoptosis (Howie *et al.*, 2009).

Los principales blancos de E6 y E7 son las proteínas p53 y pRb, respectivamente (Langsfeld y Laimins, 2016). Por un lado, E7, al promover la degradación de proteínas de supresión tumoral como pRB, p107 y p130, es capaz de inducir la fase S del ciclo celular, pero sin los mecanismos de control adecuados para este proceso (Doorbar *et al.*, 2012). Esto conduce a la inestabilidad genómica de la célula, produciendo aberraciones en la estructura de los cromosomas. En una célula no infectada, este daño a la arquitectura cromosómica llevaría al queratinocito a una fase de arresto o apoptosis celular mediada por la proteína de supresión tumoral p53 (Beaudenon y Huibregtse, 2008). No obstante, es aquí donde la proteína E6 entra en juego al promover la degradación de p53 y la activación de la telomerasa, garantizando así la inmortalización de la célula para continuar con la replicación y proliferación viral (Howie *et al.*, 2009; McLaughlin-Drubin y Münger, 2009).

La expresión de los genes E6 y E7 está fuertemente regulada por E2. Sin embargo, se ha visto que en células cancerígenas el genoma viral es capaz de integrarse en los cromosomas de la célula en proliferación. Esto conduce a la disrupción del control de E2 sobre E6 y E7, lo que acentúa los efectos oncogénicos de estas dos proteínas (Flint *et al.*, 2015). De este modo, los genes E6 y E7 de los VPH-AR actúan como oncogenes cuya expresión e interacción con determinados factores celulares implica un alto riesgo para la formación de neoplasias epiteliales y cáncer. Es por esto que las proteínas E6 y E7 son el blanco ideal para el desarrollo de fármacos antivirales y de tratamientos contra tipos de cáncer cuyo agente etiológico es el VPH (Cherry *et al.*, 2013; Malecka *et al.*, 2014).

1.1.4. Fármacos para el tratamiento del VPH de alto riesgo

Como método preventivo contra el VPH, han sido desarrolladas y empleadas dos vacunas para los genotipos de alto y bajo riesgo más comunes: la vacuna bivalente *Cervarix*, que evita la infección por los tipos oncogénicos VPH-16 y VPH-18; y la tetravalente *Gardasil*, que además de prevenir los dos tipos anteriores también provee protección contra los tipos VPH-6 y VPH-11. Ambas vacunas emplean a las proteínas de cápside L1 y L2 como agentes inmunogénicos (Monie *et al.*, 2008; Shi *et al.*, 2007). Por otro lado, para las personas ya infectadas los únicos tratamientos clínicos disponibles son las prácticas quirúrgicas ablativas o excisionales. Éstas conllevan la destrucción o remoción directa del tejido dañado, implicando riesgos de complicación, una efectividad limitada y elevados costos económicos (Tambunan y Parikesit, 2012). Debido a esto, actualmente existe la necesidad de desarrollar métodos no quirúrgicos, seguros y de bajo costo para combatir esta enfermedad. Por lo que una alternativa ha sido el desarrollo de tratamientos basados en fármacos (Hampson *et al.*, 2015).

Por su papel en la evasión de la apoptosis celular, la proteína E6 ha sido uno de los blancos ideales en el descubrimiento y desarrollo de estos fármacos inhibidores de VPH (Howie *et al.*, 2009; Malecka *et al.*, 2014). Hoy en día se sabe que las proteínas E6 de los VPH de alto riesgo comparten una secuencia de aminoácidos similar y tienen la capacidad de promover la degradación de proteínas celulares (entre ellas p53) mediante la interacción con la proteína asociada a E6 (E6AP); una enzima ubiquitina ligasa E3 (UBE3A) expresada por las células humanas, cuya función es ubiquitinizar proteínas que deben ser degradas por el proteosoma 26S (Howie *et al.*, 2009). E6AP posee un motivo helicoidal *LxxLL*, comúnmente presente en interacciones proteína-proteína en la regulación transcripcional celular. Dicho motivo es reconocido por E6 mediante una cavidad conocida como *pocket* de unión a E6AP. Sin embargo, esta región muestra una topología distinta a los dominios de reconocimiento de *LxxLL* de las proteínas celulares (Zanier *et al.*, 2013).

Por lo anterior, y sobre todo en la última década, se ha promovido el diseño y descubrimiento de diversos fármacos que bloqueen la interacción E6-E6AP; de tal forma que sea posible reestablecer la actividad de p53 y desencadenar la apoptosis de la célula infectada. Anticuerpos, péptidos, ARNs de interferencia y moléculas pequeñas han sido probadas para este fin (Yuan *et al.*, 2016; Hampson *et al.*, 2015; Milrot *et al.*, 2012; Lee *et al.*, 2005). Sin embargo, al día de hoy todas han mostrado una actividad muy modesta. Por ello sigue siendo necesario el empleo de nuevas alternativas en el descubrimiento de mejores fármacos inhibidores de los VPH de alto riesgo (Malecka *et al.*, 2014; Tambunan y Parikesit, 2012).

1.1.5. Descubrimiento *in silico* de fármacos antivirales

El descubrimiento y desarrollo de fármacos, incluidos las vacunas y antivirales, es un proceso altamente demandante de recursos, sobre todo en términos de tiempo y dinero. De acuerdo con Cerqueira *et al.* (2015), se ha estimado que cinco de cada 40, 000 compuestos probados en modelos animales logran alcanzar las pruebas clínicas. Y de estos cinco sólo uno logra ser aprobado. Además, cuando esto sucede, el proceso desde la identificación del compuesto hasta su aprobación llega a durar en promedio 10 años y a costar 2 600 millones de dólares (DiMasi *et al.*, 2016).

Por lo anterior, una de las alternativas para reducir esta elevada demanda de recursos ha sido el empleo de herramientas computacionales, lo que actualmente se conoce como diseño de fármacos asistido por computadora (DFAC) (Drie, 2007). Esta tecnología consiste en la aplicación de algoritmos y modelos computacionales que permiten describir las interacciones entre una macromolécula (proteína o ácido nucleico) y un ligando (péptidos, ARNs de interferencia o moléculas pequeñas) (Murgueitio *et al.*, 2012), evaluando las características fisicoquímicas, farmacológicas y farmacocinéticas de éste último para determinar si puede ser utilizado o no en el área clínica (Jorgensen, 2004).

Uno de los enfoques más utilizados en el DFAC es el cribado virtual (CV), que ha sido aplicado ampliamente en la industria farmacéutica como un complemento, o alternativa, al cribado farmacológico de alto rendimiento (Cerqueira *et al.*, 2015; Schneider, 2010; Clark, 2008). El objetivo del CV es utilizar una base de datos de compuestos químicos sintéticamente accesibles para filtrar e identificar moléculas con propiedades químicas deseables (actividad, metabolismo, inocuidad, etc.) y que sean capaces de interactuar con una macromolécula para inhibir o potencializar su actividad (Cerqueira *et al.*, 2015; Cuanalo, 2011). Esto permite establecer un punto de partida para el descubrimiento de nuevos medicamentos de una forma más rápida y menos costosa (Murgueitio *et al.*, 2012).

La aplicabilidad y efectividad del CV ha sido puesta a prueba en el descubrimiento de nuevos fármacos antivirales. Uno de los ejemplos más representativos fue el trabajo de Schames *et al.* (2004), quienes identificaron, mediante dinámica molecular y análisis de acoplamiento molecular, una región en el sitio catalítico de la proteína integrasa del VIH que posteriormente sería utilizada como blanco del antiviral raltegravir, aprobado en 2007 por la FDA para el tratamiento contra la infección del VIH (Summa *et al.*, 2008). Desde entonces, diversos estudios se han planteado

la identificación y la evaluación de nuevos inhibidores de otros virus, tales como el virus de la influenza, el virus del herpes y algunos integrantes de la familia de los flavivirus (Murgueitio *et al.*, 2012).

Para el caso de los VPH, en especial de los tipos de alto riesgo, la aplicación de esta tecnología ha seguido la tendencia de los estudios experimentales en la búsqueda de la inhibición de la actividad de las proteínas oncogénicas E6 y E7 (Howie *et al.*, 2009). Y en el caso particular de la proteína E6 un aspecto clave de la aplicación del cribado molecular (tanto virtual como *in vitro*) fue el descubrimiento del sitio de unión a E6 de la proteína E6AP: el motivo *LxxLL* (Be *et al.*, 2001; Chen *et al.*, 1998). Esto supuso la presencia de un “bolsillo” (*pocket*) estructural en E6 capaz de reconocer al motivo de E6AP (Be *et al.*, 2001). La existencia de este sitio de interacción ha sido fuertemente corroborada por las estructuras cristalográficas obtenidas por Zanier *et al.* (2013) y Martinez-Zapien *et al.* (2016), y representa un blanco farmacológico muy importante para la inhibición del VPH y sus efectos oncogénicos (Martinez-Zapien *et al.*, 2016; Rietz *et al.*, 2016; Zanier *et al.*, 2014).

Sin embargo, en sus inicios, la aplicación de DFAC para la inhibición del VPH se vio obstaculizada por la falta de la estructura cristalográfica de las oncoproteínas E6 y E7, debido a su alta tendencia a formar agregados (Howie *et al.*, 2009; Zanier *et al.*, 2012). A causa de esto, inicialmente las principales aplicaciones del CV han estado basadas en los ligandos previamente identificados en las investigaciones experimentales. Uno de los primeros aportes en el estudio *in silico* de la interacción E6-E6AP fue llevado a cabo por Baleja *et al.* (2006) valiéndose de la información proporcionada por Nominé *et al.* (2003). Baleja y su equipo lograron establecer un farmacóforo a partir de la información (geometría espacial y características fisicoquímicas) de los grupos funcionales de los aminoácidos en el motivo de E6AP. Con esto establecieron un modelo tridimensional que sirvió como base para realizar CV de farmacóforos basados en ligando. Posteriormente, Cherry *et al.* (2013) utilizaron este farmacóforo como un criterio de búsqueda para un CV apoyado en la estructura cristalográfica de E6 de Zanier *et al.* (2013), con lo cual obtuvieron una serie de compuestos flavonoides capaces de inhibir en ensayos *in vitro* la interacción de E6 (VPH-16) con E6AP.

Otra alternativa en esta área ha sido el modelaje de novo (*ab initio*) y homólogo de la proteína E6, principalmente de VPH-16 y VPH-18 (John y Kotadiya, 2015; Mamgain *et al.*, 2015; Kumar *et al.*, 2014). Esto permitió estudiar, mediante anclaje molecular, la interacción de la proteína E6 con moléculas pequeñas y de origen vegetal que habían sido previamente identificadas en labo-

ratorio como inhibidoras de algunos VPH-AR (Rietz *et al.*, 2016; Malecka *et al.*, 2014). Además, este enfoque ha dado lugar en los últimos años al empleo del CV basado en estructura, ya sea en combinación con un preprocesamiento de los ligandos (Kumar *et al.*, 2015a) o con un refinamiento con farmacóforos (Johnson y Karanicolas, 2016; Malecka *et al.*, 2014; Cherry *et al.*, 2013). El resultado de esto ha sido la identificación y validación de moléculas como la jaceosidina, la luteolína, la curcumina, la whitaferrina A, entre otros. Todas ellas capaces de inhibir, en mayor o menor medida, la actividad de E6-E6AP en la degradación de p53 en estudios tanto *in silico* como *in vitro*.

1.2. Justificación

Pese al éxito del CV, hay pocos compuestos han mostrado en ensayos *in vitro* una interacción eficaz con la proteína E6 y aún menos han alcanzado la etapa clínica (Hampson *et al.*, 2015). Sin embargo, con la reciente publicación de la estructura cristalográfica de la proteína E6 VPH-16 en interacción con E6AP y p53 (Martinez-Zapien *et al.*, 2016), y con la caracterización del *pocket* de unión de E6 a E6AP (Rietz *et al.*, 2016; Zanier *et al.*, 2014), hoy en día es posible refinar y precisar aún más los estudios de CV basado en estructura, en busca de moléculas candidato con mejores propiedades farmacológicas. Además, es necesario ampliar estos análisis a otros tipos de VPH de alto y bajo riesgo que sean importantes por su prevalencia, tanto a nivel mundial como en México (Bruni *et al.*, 2016). Por lo anterior, el presente trabajo tiene como propósito emplear estas herramientas computacionales previamente descritas, en la búsqueda y validación de moléculas capaces de interactuar con el sitio de unión a E6AP de la proteína E6 para combatir la infección por VPH de alto riesgo y sus efectos oncogénicos.

1.3. Hipótesis

El sitio de unión a E6AP de la proteína viral E6 de los VPH de alto riesgo es un blanco farmacológico cuya evaluación *in silico*, mediante Cribado Virtual y Dinámica Molecular, permitirá identificar moléculas con propiedades farmacológicas capaces de inhibir la interacción E6-E6AP.

1.4. Objetivos

1.4.1. Objetivo General

Caracterizar y evaluar *in silico* el sitio de unión a E6AP de la proteína E6 del virus del papiloma humano de alto riesgo para identificar y proponer moléculas con potencial farmacológico en el tratamiento contra el VPH de alto riesgo.

1.4.2. Objetivos Particulares

1. Identificar regiones clave dentro de la estructura de la proteína E6 mediante la comparación de secuencias de aminoácidos de las proteínas E6 de los tipos de VPH de alto y bajo riesgo más importantes.
2. Obtener el modelo estructural de las proteínas E6-HPV de alto y bajo riesgo más representativas.
3. Identificar moléculas pequeñas con propiedades farmacocinéticas favorables.
4. Seleccionar posibles fármacos inhibidores de VPH de alto riesgo capaces de interactuar con el *pocket* de unión a E6AP de la proteína E6.
5. Validar la interacción de las moléculas seleccionadas con la proteína E6 mediante dinámica molecular.

Capítulo 2. Metodología

En este capítulo se presenta la metodología empleada haciendo hincapié en cada uno de los objetivos de la Sección 1.4. La Figura 1 esboza el flujo de trabajo y los análisis realizados en cada paso. Los cálculos de simulación de dinámica molecular se realizaron en GPUs del *cluster* de supercómputo *Aikanaro* de la Universidad Andrés Bello, Chile. Los cálculos de simulación de acoplamiento molecular fueron llevados a cabo en el cluster computacional *OMICAs* del CICESE y en el cluster de supercómputo *Miztli* de la UNAM, México. El resto de los análisis se realizaron en un equipo *Aspire R14 Intel® Celeron dualcore* 1.50GHz.

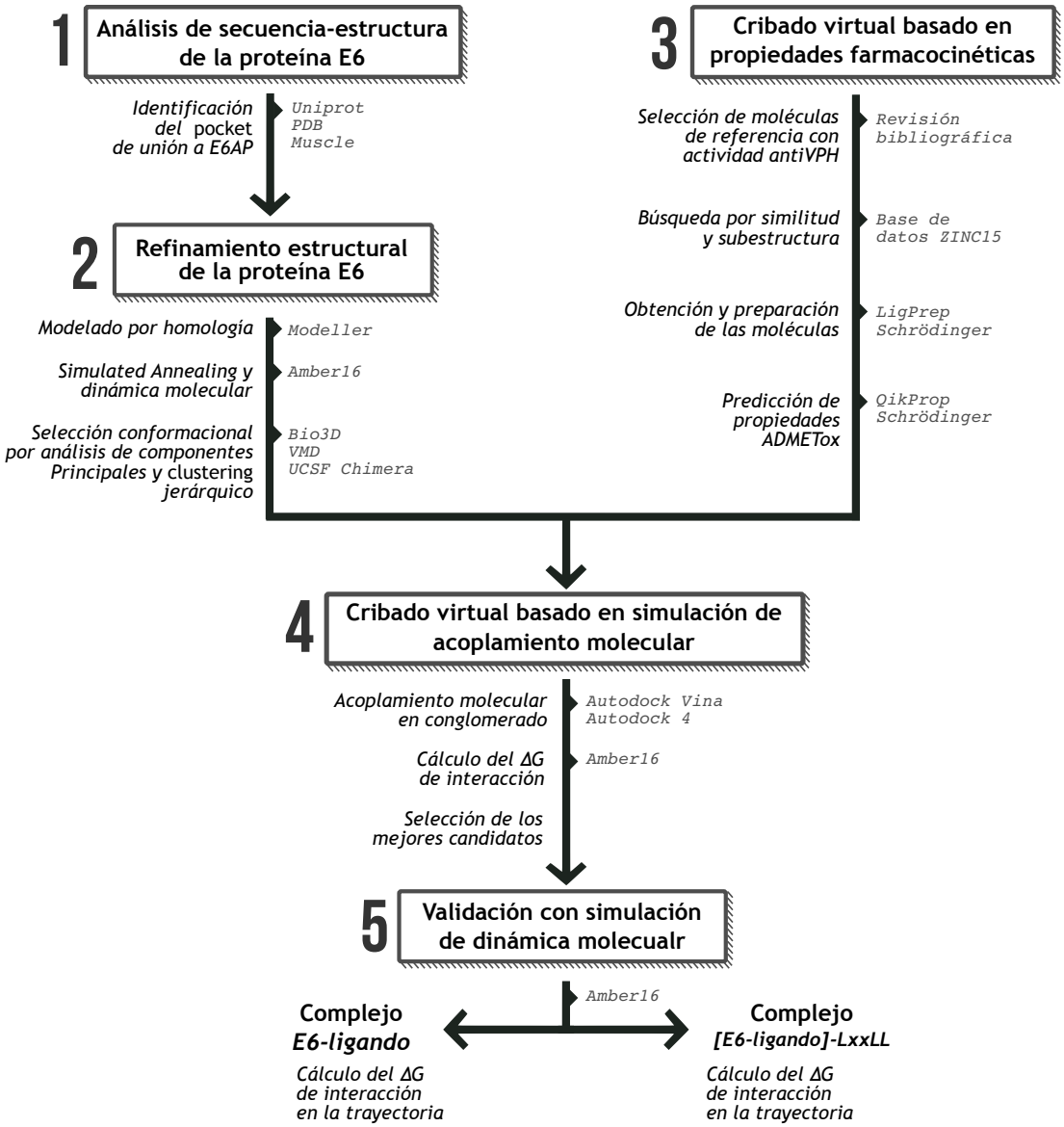


Figura 1: Metodología empleada para identificar moléculas con potencial de inhibir la interacción E6-E6AP. Los métodos realizados están desglosados de acuerdo con los cinco objetivos. Además se hace referencia a las principales herramientas utilizadas en cada paso.

2.1. OBJETIVO 1: Análisis secuencia-estructura de la proteína E6

2.1.1. Análisis de secuencias de la proteína E6

Se utilizó la base de datos *HPV Centre* (Bruni *et al.*, 2014) para seleccionar los tipos de VPH de alto riesgo (AR) y bajo riesgo (BR) más representativos, tomando en cuenta su prevalencia y el riesgo asociado a la formación de carcinomas. Posteriormente, a partir de la base de datos Uniprot (The UniProt Consortium, 2014), se obtuvieron las secuencias de aminoácidos de la proteínas E6 y L1 correspondientes a los tipos de VPH seleccionados. El alineamiento múltiple de estas secuencias se llevó acabo a través del software *SeaView* (Gouy *et al.*, 2010), utilizando los algoritmos *MUSCLE* y *Clustal Omega*. Finalmente, con los alineamientos se generaron fenogramas empleando el método de *BIO Neighbor Joining* (Gascuel, 1997) y un *boostrap* de 1000 réplicas. La visualización se llevó a cabo con el software *FigTree v1.4.2* (Rambaut, 2012).

2.1.2. Caracterización del *pocket* de unión a E6AP

El modelo estructural de la proteína E6 VPH-16 (Martinez-Zapien *et al.*, 2016), fue utilizado como referencia para identificar los residuos de E6 correspondientes al *pocket* de unión al motivo *LxxLL* de E6AP (*pocket* de unión). Este proceso se llevó a cabo empleando la herramienta en línea *CASTp* (*Computed Atlas of Surface Topography of proteins*) (Dundas *et al.*, 2006) y el software de visualización *UCSF Chimera* (Pettersen *et al.*, 2004). Posteriormente, con la herramienta *WebLogo* (Crooks *et al.*, 2004) se generaron los logos de la subsecuencia de los *pockets* de unión a E6AP de las proteínas E6 de AR y de BR.

2.2. OBJETIVO 2: Refinamiento estructural de la proteína E6

2.2.1. Modelaje homólogo de la proteína E6

Una vez seleccionadas las secuencias de los tipos representativos de AR y de BR se procedió a obtener los modelos estructurales de la proteína E6 mediante modelado por homología. Este proceso fue llevado a cabo empleando el software *Modeller* (Šali, 2013) a través de la interfaz de *UCSF Chimera*. Como primer paso, se empleó el programa BLASTp (*protein-protein BLAST*) utilizando como *query* las secuencias de aminoácidos de la proteína E6, de los tipos de AR y de BR, obtenidas de *Uniprot*. La base de datos utilizada para la búsqueda de la estructura molde fue *The Protein Data Bank* (Berman *et al.*, 2000), y la matriz de sustitución empleada fue BLOSUM62.

El mejor modelo fue elegido con base en los métodos de *scoring* DOPE (*Discrete Optimized Protein Energy*) y GA341, ambos implementados por *Modeller*. Finalmente, los modelos generados fueron guardados en formato *"pdb"*.

2.2.2. Refinamiento por dinámica molecular de la proteína E6

Diversos ensayos de simulación de dinámica molecular (SDM) fueron llevados a cabo tanto para refinar la estructura de la proteína E6 previo a los ensayos de acoplamiento molecular; para evaluar la interacción de E6 con el motivo LxxLL de E6AP; o para validar los resultados del acoplamiento de E6 con moléculas pequeñas. Para cada caso los ensayos de SDM fueron realizados con el software *The Amber Molecular Dynamics Package* (Amber16) (Case *et al.*, 2016), siguiendo el protocolo que se describe a continuación.

2.2.2.1. Preparación de la proteína

Como primer paso se utilizó la herramienta en línea *PROPKA* (Olsson *et al.*, 2011) para calcular los valores de pKa de los aminoácidos ionizables de la proteína, y sus estados de protonación a un pH de 7.0. Posteriormente, se utilizó la herramienta LEaP de *AmberTools* para solvatar la proteína (archivo *"pdb"*) en una caja de agua (modelo TIP3P), cuyas dimensiones fueron de 15 Å más allá del último átomo de la proteína en los tres ejes. Además, el sistema se neutralizó añadiendo átomos de Cl^- o Na^+, según el caso. Como resultado se obtuvo un archivo de coordenadas (*"inpcrd"*) y un archivo de parámetros y topología (*"prmtop"*), los cuales fueron creados utilizando los campos de fuerza *Amber-FF14SB* (Maier *et al.*, 2015) y *Zinc AMBER force field* (ZAFF) (Peters *et al.*, 2010). Este último fue usado para modelar los dedos Cys_4Zinc presentes en la proteína E6.

2.2.2.2. Minimización, *Simulated annealing* y equilibrado

La minimización de la energía potencial del sistema fue llevada a cabo mediante un proceso de 10 000 ciclos de refinamiento utilizando el algoritmo del "descenso más pronunciado"(*Steepest descent*) para los primeros 5 000 ciclos, y el algoritmo del "gradiente conjugado"(*Conjugate gradient*) para los 5 000 ciclos restantes. Posteriormente, se empleó el método de *Simulated Annealing* (SA) para relajar el sistema y favorecer su equilibrio termodinámico a 310 K. El protocolo de SA fue el que indica a continuación: 1) Calentamiento de 0 K a 300 K en un lapso de 0.5 ns; 2) mantenimiento a 300 K por 0.5 ns; 3) calentamiento de 300 K a 400 K en 0.5 ns; 4) mantenimiento

a 400 K por 0.5 ns; y 5) enfriamiento de 400 K a 310 en 0.5 ns. Posteriormente, se llevó a cabo una fase de equilibrado de 5.5 ns a 310 K.

El tipo de simulación utilizado durante el SA y el equilibrado fue de un colectivo NVT con condiciones de frontera periódicas. Se empleó el método *Langenvin dynamics* como termostato, y el tiempo de integración utilizado fue de 1 fs por paso. Para el cálculo de las interacciones electrostáticas se utilizó el método de *Particle Mesh Ewald* con un *cutoff* de 10 Å. Por último, los enlaces que implicaban una interacción covalente con hidrógeno se mantuvieron rígidos utilizando el algoritmo *SHAKE* (Case *et al.*, 2016). Como resultado se obtuvo un archivo *"restrt"* con las coordenadas finales de los átomos y sus velocidades.

Finalmente, para asegurar que el sistema se encontraba en equilibrio termodinámico a 310 K al final de la etapa de equilibrado, se corroboró que las energías cinéticas (E_k) de los átomos se ajustaran a la distribución de Maxwell-Boltzmann para energías cinéticas (Ecuación 1). Para ello se obtuvo la masa de cada átomo del sistema (archivo *"prmtop"*) y su velocidad correspondiente al último paso de equilibrado (archivo *"restrt"*), con ello se obtuvo un histograma con la frecuencia relativa de las E_k observadas. Después, para evaluar el ajuste con la probabilidad distribución E_k a una T determinada, se calculó el error total, el cual corresponde a la suma del cuadrado de las diferencias entre los rangos numéricos (bins) de las frecuencias esperadas menos las observadas.

$$f(E_k) = \ 2\sqrt{\frac{E_k}{\pi}} \cdot \left(\frac{1}{k_B T}\right)^{3/2} \cdot \exp\left(\frac{-E_k}{k_B T}\right) \tag{1}$$

donde E_k es la energía cinética de cada átomo, T la temperatura del sistema, k_B la constante de Boltzmann equivalente a $0.00198657\ kcal/mol$, y $f(E_k)$ la densidad de probabilidad de las energías cinéticas a una temperatura T (Satoh, 2011; Phillips *et al.*, 2005).

2.2.2.3. Simulación de dinámica molecular

Una vez concluido el equilibrado se procedió a realizar la fase de producción de la SDM. Para ello se utilizó el archivo *"prmtop"* y el archivo *"restrt"*, que posee las coordenadas de los átomos y sus velocidades correspondientes al último paso del equilibrado. El intervalo de integración utilizado fue de 2 fs, y se utilizó un colectivo NPT, manteniendo la temperatura del sistema a $T \approx$ 310 K, y la presión a $P \approx$ 1 atm. El tiempo total de simulación fue de 100 ns, y se registró la información del sistema cada 0.02 ns. Como resultado se obtuvo un archivo de trayectoria *"mdcrd"* con las

coordenadas de los átomos de cada una de las conformaciones (*frames*) correspondientes a la dinámica del sistema. También se generó un archivo *"mdout"* que registró cada 0.02 ns los valores de la temperatura, presión, volumen y energía del sistema.

2.2.3. Análisis de las trayectorias de dinámica molecular

El análisis de las trayectorias de las SDM fue llevado a cabo utilizando el programa de visualización molecular VMD 1.9.3 (*Visual molecular dynamics*) (Humphrey *et al.*, 1996) y el paquete Bio3D v2.3 (Grant *et al.*, 2006) implementado en lenguaje de programación R.

2.2.3.1. RMSD: *Root Mean Square Deviation*

Para el análisis conformacional se utilizó la medida RMSD, la cual mide la distancia media entre los átomos de dos estructuras moleculares previamente superpuestas mediante el algoritmo de Kabsch (Grant *et al.*, 2006). En este caso, se consideraron únicamente los carbonos alfa ($C\alpha$) de la proteína E6, y se evaluó cada *frame* t a lo largo del tiempo de la trayectoria frente al *frame* de referencia $t = 0$, el cual corresponde a la estructura cristalográfica de la proteína. La ecuación es la siguiente:

$$RMSD_{t,0} = \sqrt{\frac{1}{N}\sum_{i=1}^{N} || \, r_i(t) - r_i(0) \, ||^2} \tag{2}$$

donde $RMSD_{t,0}$ es el valor de RMSD en Å entre una estructura en el instante t y una estructura de referencia en el instante 0. N es el número de átomos ($C\alpha$) en la proteína; $r_i(t)$ y $r_i(0)$ son los vectores de coordenadas del átomo i en los *frames* t y 0, respectivamente.

2.2.3.2. RMSF: *Root Mean Square Fluctuation*

Posteriormente, para evaluar la flexibilidad estructural por residuo de la proteína se utilizó la medida RMSF, la cual sigue un principio similar al del RMSD con la diferencia de que se evalúa de forma independiente para cada $C\alpha$ de cada residuo, y de que la estructura de referencia es la media de todas las coordenadas (x, y, z) que dicho átomo tiene a largo de una trayectoria. La ecuación es la siguiente:

$$RMSF_j = \sqrt{\frac{1}{T}\sum_{i=1}^{T} || \, \alpha_i j - \overline{\alpha} \, ||^2} \, , \qquad \text{con} \qquad \overline{\alpha} = \frac{1}{T}\sum_{i=1}^{T} \alpha_i \tag{3}$$

donde $RMSF_j$ es el valor de RMSF en Å del $C\alpha$ de un residuo j de la proteína. T es el número *frames* en la trayectoria; $\alpha_i j$ es el vector de coordenadas de α en el *frame* i del residuo j; y $\overline{\alpha_j}$ es el vector de la media de las coordenadas de α en el residuo j tomando en cuenta todos los *frames*.

2.2.3.3. Análisis de componentes principales y *clustering* Jerárquico

El siguiente paso en el análisis de trayectorias fue seleccionar un grupo de conformaciones representativas de la dinámica de la proteína, y en particular del *pocket* de unión a E6AP. Como primer paso se realizó un análisis de componentes principales (ACP), el cual permite analizar la "dinámica esencial" de la proteína al describir los cambios conformacionales asociados a las fluctuaciones estructurales más relevantes que ocurren a lo largo de la trayectoria (Skjærven *et al.*, 2014; Stein *et al.*, 2006).

Para llevar a cabo el ACP se empleó la función `pca.xyz()` del paquete *Bio3D* (Grant *et al.*, 2006), la cual utiliza la información de la trayectoria, representada en una matriz donde las columnas corresponden a las coordenadas cartesianas de los átomos y las filas a los *frames* de la etapa de producción de la SDM. Para el caso del presente trabajo se utilizaron únicamente los $C\alpha$ de los residuos pertenecientes al *pocket* de unión a E6AP.

Posteriormente, se llevó a cabo un *clusering* jerárquico con el objetivo de identificar los macroestados conformacionales de la proteína, tomando en cuenta su geometría a lo largo de la trayectoria (*clustering* geométrico) (Skjærven *et al.*, 2014). Primero se generó la matriz de distancia a partir de los *scores* de los Componentes Principales. Para ello se empleó la función `dist()` con el método de Ward. Después se usó la función `hclust()` seguida de la función `cutree()` para generar un dendograma y seleccionar el número de grupos deseado (R Core Team, 2017). Finalmente, para cada *cluster* de conformaciones se obtuvo una estructura a partir de la media de las coordenadas de los $C\alpha$ de los residuos de la proteína. Y se seleccionó la conformación cuyo valor de RMSD del *pocket* fuera el mínimo con respecto al *pocket* de la estructura media (Grant *et al.*, 2006).

2.3. OBJETIVO 3: Cribado virtual de las propiedades ADMETox de los ligandos

2.3.1. Selección del espacio químico

Se llevó a cabo una revisión bibliográfica para identificar moléculas orgánicas pequeñas (peso molecular < 500 Da) que en estudios previos, tanto experimentales y/o computacionales, hayan mostrado algún tipo de actividad anti-VPH. El objetivo fue que estas moléculas seleccionadas fungieran como referencia estructural en la búsqueda y obtención de nuevas moléculas, cuya actividad frente al VPH no hubiese sido evaluada anteriormente. Para esto, cada una de las moléculas de referencia identificadas fue utilizada como criterio de consulta (*query*) en la base de datos *Zinc15* (Sterling y Irwin, 2015), la cual contiene información de más 120 millones de moléculas comercialmente disponibles.

Para la selección de nuevas moléculas en *Zinc15* se utilizaron tres métodos de búsqueda: *subestructura*, coeficiente de Tanimoto y coeficiente de Dice. Los coeficientes de Tanimoto y Dice son usados para medir la similitud entre dos moléculas (con valores que van de 0 a 1, siendo este último el valor máximo de similitud). Ambos coeficientes emplean el criterio *Extended Connectivity Fingerprint* (512 bit ECFP4), el cual utiliza huellas topológicas que a través de variables dicotómicas (*bits*) determinan la presencia (1) o la ausencia (0) de subestructuras en ambas moléculas. Sin embargo, como se observa en la Ecuación 4, Tanimoto y Dice difieren en la manera de evaluar la relación presencia/ausencia de dichas huellas. Finalmente, se seleccionaron moléculas con un coeficiente de similitud igual o mayor a 0.6, y con la búsqueda por *subestructura* se obtuvieron las moléculas que tuviesen como estructura base a la estructura de la molécula consulta.

$$\begin{aligned} \text{Coeficiente de Tanimoto:} \quad & S_{A,B} = 2c/(a+b). \\ \text{Coeficiente de Dice:} \quad & S_{A,B} = c/(a+b-c). \end{aligned} \tag{4}$$

Donde $S_{A,B}$ es el valor de similitud entre una molécula A y una molécula B, a es el número de *bits* activos en A (molécula consulta), b es el número de *bits* activos en B (cada una de las moléculas en la base de datos *Zinc15*), y c es el número de *bits* activos compartidos por ambas moléculas (Bajusz *et al.*, 2015).

2.3.2. Refinamiento estructural de las moléculas

Tras la búsqueda realizada en *Zinc15*, las moléculas seleccionadas fueron descargadas en formato SMILES (*Simplified Molecular Input Line Entry Specification*) y se utilizó el módulo *LigPrep* (Schrödinger, 2017b) de la suite *Schrödinger* para construir los modelos tridimencionales de las mismas. Esto se llevó a cabo mediante un proceso de minimización de la energía potencial de la molécula empleando el campo de fuerza OPLS3 (Harder *et al.*, 2016), y un rango de pH de 7.0 $\pm$ 0.2.

2.3.3. Predicción de propiedades ADME-Tox

Mediante el empleo del software *QikProp* (Schrödinger, 2017b) se llevó a cabo la predicción *in silico* de 51 descriptores relacionados con la farmacocinética de cada una de las moléculas previamente refinadas con *LigPrep*. Estos descriptores, calculados a partir de la estructura tridimencional de las moléculas, permiten establecer un perfil ADME-Tox (*Absortion, Distribution, Metabolism, Excretion and Toxcicity*) que describa la disposición de un compuesto farmacéutico en el organismo.

2.3.4. Cribado virtual utilizando los descriptores ADME-Tox

Con la información generada con *Qikprop* se llevó a cabo un filtrado de moléculas. Para esto, se tomaron en cuenta únicamente cinco descriptores que, sin embargo, englobaron un total de 25 propiedades químicas y farmacocinéticas; las cuales son desglosadas en la Tabla 1. Los descriptores, y los criterios de filtrado usados, se seleccionaron tomando como referencia el trabajo publicado por Ntie-Kang (2013) y la documentación de *Qikprop* (Schrödinger, 2017b).

Adicionalmente, 12 fármacos fueron elegidos y también evaluados en esta fase del cribado; sirviendo como controles positivos. Dichos fármacos están actualmente aprobados por la FDA (*Food and Drug Administration*), se catalogan como moléculas pequeñas, actúan sobre proteínas y su método de administración es oral (Law *et al.*, 2014). Los fármacos elegidos fueron los siguientes: *amoxicilina, ácido acetil salicílico, cafeína, ibuprofeno, metoxalen, omeoprazol, paracetamol, atazanavir, nelfinavir, oseltamivir, raltegravir y sofosvuvir*.

Tabla 1: Descriptores de *QikProp* utilizados para el filtrado de moléculas según sus características farmacológicas predichas.

	Descriptor	Criterio utilizado para el filtrado
1)	**Regla de cinco de Lipinsky**	$\leq$ 1 (máximo una violación a la regla)
	Implica el número de violaciones a la regla de cinco de Lipinski (Lipinski *et al.*, 2001). Incluye cuatro reglas: a) Peso molecular menor a 500 Da (*mol_MW* < 500). b) Cinco o menos donadores de enlaces de hidrógeno (*donorHB* $\leq$ 5). c) Menos de 10 aceptadores de enlaces de hidrógeno (*accptHB* < 10). d) Logaritmo del coeficiente de partición octanol/agua menor a 5.0 (*QPlogPo/w* < 5).	
2)	**Regla de tres de Jorgensen**	$\leq$ 1 (máximo una violación a la regla)
	Número de violaciones a la regla de tres de Jorgensen (Jorgensen, 2004). Las tres reglas son: a) Logaritmo de la solubilidad acuosa (*logS*); concentración del soluto en una disolución acuosa saturada (mol/L) (-20 $\leq$ *logS* $\leq$ 5.7). b) Permeabilidad celular en células Caco-2 mayor a 20 nm/s (*QPPCaco* > 20); las células Caco-2 son células modelo de la barrera hemato-intestinal. c) Menos de siete reacciones metabólicas probables a partir de los grupos funcionales de la molécula (*#metab* < 7).	
3)	**Absorción Oral Humana**	Valor de 2 o 3
	Valor cualitativo de absorción oral humana: 1 = baja, 2 = media, 3 = alta. Además de *QPPCaco* y *logS*, toma también en cuenta los siguientes descriptores: a) Número de rotámeros menor a 15 (*#rotor* < 15). b) Logaritmo de la permeabilidad celular .	
4)	**Número de *stars* de *Qikprop***	$\leq$ 1 (máximo un descriptor fuera de intervalo)
	Implica el número de descriptores cuyo valor cae fuera del rango del 95 % de los fármacos conocidos. Los descriptores tomados en cuenta son: a) *SASA*: Área superficial accesible al solvente en Å^2 (intervalo 300.0 a 1000.0). b) *FOSA*: Componente hidrofóbico de SASA en Å^2 (intervalo 0.0 a 750.0). c) *FISA*: Componente hidrofílico de SASA en Å^2 (intervalo 0.0 a 330.0). d) *PISA*: Componente pi de SASA en Å^2 (intervalo 0.0 a 400.0). e) *WPSA*: Componente polar débil de SASA; halógenos P y S (intervalo 0.0 a 175.0). f) *PSA*: Área superficial vdW de Oxígenos y Nitrógenos polares (intervalo 7.0 a 200.0). g) Volumen: Volumen total en Å^3 (intervalo 500.0 a 2000.0). h) Globularidad: grado de esfericidad de la molécula; 1 = esfera. Se calcula a partir del radio de una esfera con un volumen igual al de la molécula (intervalo 0.75 a 0.95). i) *QPpolrz*: Logaritmo de polarizabilidad en Å^3 (13.0 a 70.0). j) *QPlogPC16*: Logaritmo del coeficiente de partición hexadecano/gas (4.0 a 18.0). k) *QPlogPoct*: Logaritmo del coeficiente de partición octanol/gas (8.0 a.2.0) l) *IP*: Potencial de ionización calculado PM3; HOMO (7.9 a 10.5). m) *EA*: Afinidad electrónica calculada PM3; LUMO (-0.9 a 1.7) n) *QPlogKhsa*: Predicción de unión a albumina humana (-1.5 a 1.5). o) *QPlogBB*: Coeficiente de partición para la barrera hemato-encefálica (-3.0 a 1.2).	
5)	**Penetración dérmica**	Valores entre -8.0 y 1.0
	QPlogKp: Logaritmo de permeabilidad dérmica predicha Kp (cm/s). Se calcula mediante un algoritmo basado en *QSAR* que evalúa la relación entre el peso molecular, el volumen molecular y el coeficiente de partición octanol/agua.	

2.4. OBJETIVO 4: Cribado virtual basado en acoplamiento molecular

Una vez que las moléculas fueron filtradas según su perfil ADME-Tox, se procedió a realizar la etapa de CV basado en receptor. Para ello, se llevaron a cabo simulaciones de acoplamiento molecular (SAM) cuya finalidad fue estimar la conformación de acoplamiento energéticamente más favorable entre una molécula ligando con flexibilidad rotacional y una molécula receptora rígida. Pero además, al calcular un valor de afinidad receptor-ligando, la SAM posibilitó el ordenamiento del conjunto de ligandos de acuerdo con su valor de interacción con un receptor en común (Cerqueira *et al.*, 2015). Por tanto, en el caso del presente trabajo, esta fase tuvo como objetivo identificar qué moléculas (llamadas desde ahora ”ligandos”) poseían una mejor afinidad hacia la proteína E6.

Para lograr lo anterior se realizaron dos fases de SAM consecutivas empleando dos programas de *docking* distintos: *Autodock Vina* (Trott y Olson, 2010a) y *Autodock 4.2* (*AD4*) (Morris y Huey, 2009). Ambos programas llevan a cabo el acoplamiento en un proceso iterativo que consiste en 1) predecir la *pose* del ligando, es decir, la conformación, la posición y la orientación del ligando frente al receptor; y 2) calcular, mediante una función de *scoring* (puntuación), la energía libre de interacción (ΔG en $kcal/mol$) entre el receptor y dicha *pose* (Meng *et al.*, 2011). Sin embargo, *AD4* y *Vina* utilizan funciones de *scoring* y algoritmos de búsqueda diferentes. *AD4* emplea un campo fuerza híbrido, que combina la aplicación del campo de fuerza Amber y de términos empíricos basados en constantes de interacción determinadas experimentalmente (Morris y Huey, 2009). Por su parte, *Vina* usa una función de *scoring* empírica inspirada en la función *X-Score* (Wang *et al.*, 2002).

2.4.1. Preparación del receptor y los ligando

Tanto *AD4* como *Vina* requieren como entrada el archivo del receptor y el ligando, ambos en formato *”pdbqt"*. Este formato, además de contener las coordenadas (x, y, z) de los átomos, posee también sus cargas parciales y la nomenclatura del tipo de átomo correspondiente al campo de fuerza usado por *Autodock*. Para generar estos archivos se utilizaron los *scripts* de *python prepare_ligand4.py* y *prepare_receptor4.py*, proporcionados por el programa *Autodock Tools* (Morris y Huey, 2009).

El proceso de generación para cada molécula consistió en los siguientes pasos: 1) Asignación de las cargas parciales empíricas de cada átomo aplicando el método de Gasteiger. 2) Remoción

de los hidrógenos apolares y reasignación de sus cargas parciales a los carbonos con los que se enlazan (*united atom representation*). 3) Adición del nombre del tipo atómico según correspondiente a la nomenclatura del campo de fuerza utilizado *Autodock*.

2.4.2. Acoplamiento molecular con *Atodock Vina*

En una primera etapa de SAM se empleó una metodología conocida como *Conformation Ensemble Docking* (ancoplamiento molecular en colectivo conformacional). Ésta se basa en la aplicación de SDM a una muestra de confórmeros de la proteína receptora, que en conjunto buscan representar la flexibilidad conformacional de la proteína (Korb *et al.*, 2012). Por lo tanto, para efectuar este proceso se utilizó como receptor a cada una de las conformaciones de la proteína E6 (obtenidas en el objetivo 2; Sección 2.2.2). Y las moléculas usadas como ligandos fueron todas aquellas filtradas al final del objetivo 3 (Sección 2.3.4).

Para efectuar este proceso se empleó el programa *Vina*, que puede llegar a ser varios ordenes más rápido que *AD4* (Trott y Olson, 2010a), lo que posibilitó evaluar el elevado número de ligandos iniciales tras el filtrado ADME-Tox. Los parámetros utilizados para el *docking* con *Vina* fueron los siguientes: una exhaustividad de 64 y una caja de búsqueda centrada en el *pocket* de unión a E6AP, con dimensiones de 21x21x21 Å, y con un tamaño de rejilla de 1 $Å^2$. Para el análisis de resultados se tomó en cuenta el valor de la energía libre de interacción ($kcal/mol$) de cada molécula con respecto a E6.

2.4.3. Acoplamiento molecular con *Autodock 4.2*

La segunda etapa de SAM se llevó a cabo con *AD4*. Para esta fase sólo se evaluaron ligandos que con *Vina* tuvieron una mejor energía de interacción con cada una de las conformaciones usadas de E6. Los parámetros utilizados fueron los siguientes: se empleó el algoritmo genético Lamarckiano de *AD4*, con un tamaño de población de 150, 2 700 generaciones, y una tasa de mutación de 0.02. Se realizaron 50 000 000 evaluaciones de energía por corrida (ensayo de *docking*), y un número total de 50 corridas por molécula. Además, el tamaño de la caja de búsqueda fue de 21x21x21 Å, con un tamaño de rejilla de 0.375 Å por 0.375 Å. Finalmente, las moléculas con valores de interacción más bajos fueron seleccionadas y se obtuvo la estructura del complejo de acoplamiento en formato *"pdb"*, que sería finalmente utilizada para los ensayos de validación por dinámica molecular.

2.5. OBJETIVO 5: Validación de las moléculas candidato

2.5.1. Validación del acoplamiento con MM/GBSA y MM/PBSA

Para corroborar los resultados que mostraron los mejores ligandos con *AD4*, se llevó a cabo la aplicación de los métodos MM/GBSA (*Molecular Mechanics with Generalized Born and Surface Area solvation*) y MM/PBSA (*Molecular Mechanics with Poisson–Boltzmann and Surface Area solvation*). En estos métodos la energía libre de interacción receptor-ligando (ΔG_{bind}) es estimada a partir de la diferencia entre la energía de tres estados independientes: *complejo receptor-ligando* (ΔG_{comp}), *receptor* (ΔG_{rec}) y *ligando* (ΔG_{lig}):

$$\Delta G_{bind} = \Delta G_{compl} - \Delta G_{rec} - \Delta G_{lig}. \tag{5}$$

La energía libre de cada estado en la Ecuación 5 es calculada a partir de los términos de energía de mecánica molecular (ΔE_{MM}) más la contribución de la energía del solvente (ΔG_{solv}), menos el producto del valor de la entropía (ΔS) por la temperatura (T) del sistema (Ecuación 6). El valor de ΔG_{solv} es la suma del componente apolar del solvente (ΔG_{SA}), calculado como una función lineal del área accesible al solvente (*ASA: Accessible Surface Area*); y del componente polar o electrostático ($\Delta G_{PB/GB}$), el cual se calcula utilizando la ecuación de Poisson-Boltzmann (PB), o el modelo de aproximación Generalizada de Borh (GB) (Kantabutra, 2015).

$$\begin{aligned} \Delta G_{comp/rec/lig} &= \Delta E_{MM} + \Delta G_{solv} - T\Delta S, \\ \Delta E_{MM} &= \Delta E_{bond} + \Delta E_{elect} + \Delta E_{vdW}, \\ \Delta G_{solv} &= \Delta G_{ASA} + \Delta E_{PB/GB}. \end{aligned} \tag{6}$$

La aplicación de ambos métodos se llevó a cabo utilizando el software *Amber16* mediante la ejecución del *script* de *python* MMPBSA.py, proporcionado por *AmberTools16*. Para ello, primero se efectuó el protocolo de preparación del sistema solvatado (Sección 2.2.2.1) y la minimización de la energía del mismo (Sección 2.2.2.2). Sin embargo, antes fue necesario obtener los parámetros de campo de fuerza requeridos para simular la dinámica de los ligandos. Para esto se utilizó la herramienta *Antechamber* (Wang *et al.*, 2006) y el campo de fuerza *Amber-GAFF* (Wang *et al.*, 2004).

2.5.2. Dinámica molecular del complejo proteína-ligando

Los complejos proteína-ligando con los mejores valores de energía de interacción (ΔG_{bind} más bajos) fueron seleccionados para llevar a cabo SDM siguiendo el protocolo y análisis descrito en la Sección 2.2. Adicionalmente, para evaluar la afinidad y estabilidad del complejo se aplicaron los métodos de MM/GBSA y MM/PBSA a lo largo de los *frames* correspondientes la trayectoria. Finalmente, se llevaron a cabo simulaciones de dinámica molecular de un complejo formado por la proteína E6, el ligando y la hélice de E6AP. El objetivo de esta fase fue calcular la energía de interacción entre la hélice de E6AP hacia el complejo del ligando unido a E6, y así intuir el posible efecto que tendría la presencia del ligando sobre la unión de E6AP al *pocket* de la proteína E6.

Capítulo 3. Resultados

3.1. OBJETIVO 1: Análisis secuencia-estructura de la proteína E6

3.1.1. Determinación de los tipos de VPH a analizar

A partir de la base de datos HPV Centre (Bruni *et al.*, 2014) se identificaron los tipos de VPH de alto y bajo riesgo a tomar en cuenta en el análisis de secuencias. Para esto, se realizó una búsqueda de la prevalencia de los tipos de VPH en muestras de lesiones cancerígenas de cérvix a nivel mundial. La secuencia de búsqueda empleada fue la siguiente: $HPV\ prevalence\ estimates > Invasive\ cervical\ cancer > HPV\ type\ distribution > World$.

Para seleccionar los tipos de VPH analizados se consideró la prevalencia total del tipo (criterio de búsqueda: $Any\ Histology$); la prevalencia parcial asociada únicamente a carcinomas de células escamosas y a adenocarcinomas (criterios de búsqueda: $Squamous\ Cell\ Carcinoma$ y $Adenocarcinoma$); la clasificación del tipo de alto riesgo o bajo riesgo; y la disponibilidad de la secuencia de aminoácidos de la proteína E6 de cada tipo en la base de datos *Uniprot*.

Tomando en cuenta los datos anteriores, se eligieron 21 tipos de VPH, todos pertenecientes al género *Alphapapilomavirus* (A), y correspondientes a los subgrupos A5, A7, A8, A9, A10 y A13, de acuerdo con la filogenia propuesta por Villiers *et al.* (2004). Posteriormente los 21 tipos fueron clasificados en dos grupos principales: alto riesgo (AR) y bajo riesgo (BR). Además, según la prevalencia en los casos $Carcinoma$ y $Adenocarcinoma$, cada grupo se dividió en dos subgrupos; siendo los subgrupos AR1 y BR1 los tipos de alto y bajo riesgo más frecuentes, y los AR2 y BR2 los menos frecuentes. Los tipos seleccionados se listan a continuación:

- VPH AR1: 16_{A9}, 31_{A9}, 33_{A9}, 52_{A9}, 58_{A9}, 18_{A7}, 45_{A7}
- VPH AR2: 35_{A9}, 67_{A9}, 39_{A7}, 59_{A7}, 68_{A7}, 70_{A7}, 26_{A5}
- VPH BR1: $6A_{A10}$, 11_{A10}, 44_{A10}, 54_{A10}, 74_{A10}
- VPH BR2: 40_{A8}, 43_{A8}

3.1.2. Obtención y alineamiento múltiple de secuencias

Para cada tipo de VPH seleccionado se obtuvo la secuencia de aminoácidos en formato FASTA de las proteínas E6 y L1. La proteína L1 fue utilizada como referencia, debido a que la filogenia

de los VPH está basada principalmente en la secuencia de esta proteína (Bernard *et al.*, 2010; Villiers *et al.*, 2004). Posteriormente, dichas secuencias fueron alineadas empleando los algoritmos *MUSCLE* y *Clustal Omega*. Tras un análisis con ambos, se eligió proceder utilizando *MUSCLE*, pues mostró un mejor alineamiento en las regiones terminales de la secuencia. La Figura 2 muestra parte del alineamiento haciendo hincapié en las secuencias de los tipos AR y BR con mayor prevalencia.

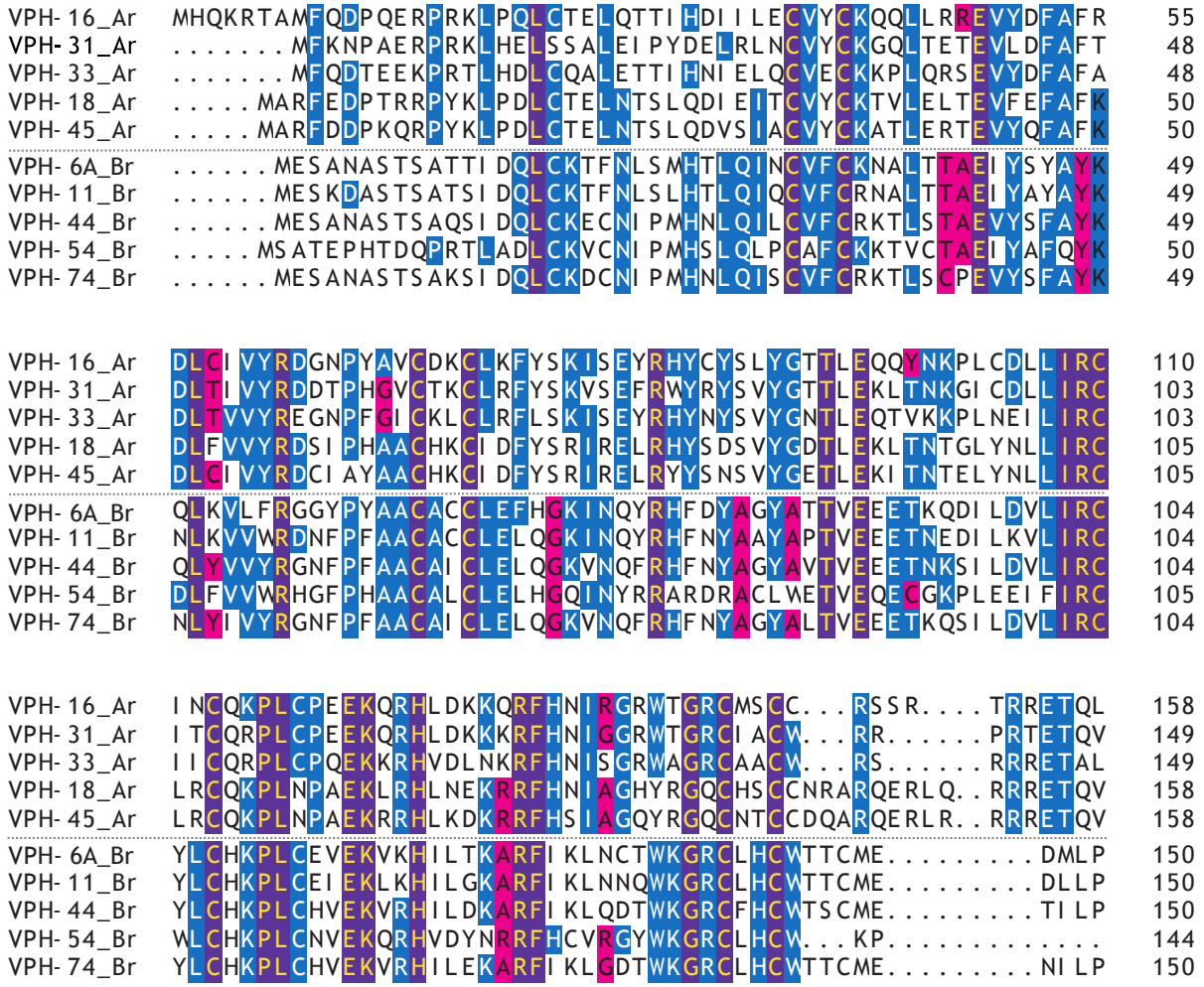


Figura 2: Alineamiento múltiple de secuencias de la proteína E6. Se muestran únicamente las secuencias de aminoácidos de la proteína E6 de los cinco tipos de AR y de BR con mayor incidencia (AR1 y BR1). El color morado señala los sitios con una conservación igual al 100 %, el azul los sitios con una conservación $\geq$ 50 %, y el magenta sitios con una similitud $\geq$ 50 %, es decir, residuos con propiedades fisicoquímicas similares. Figura obtenida con el paquete TeXshade (Beitz, 2000)

Tras obtener los alineamientos, se generaron los fenogramas de la proteína E6 y de la proteína L1. El objetivo fue verificar si los *clusters* de VPH formados con las secuencias de E6 coinciden con los de la proteína L1. Para construir cada fenograma se usó *SeaView* con el método de *BIO Neighbor Joining* (BIONJ) y un *boostrap* de 1000 réplicas. La visualización se llevó a cabo con el software *FigTree v1.4.2*.

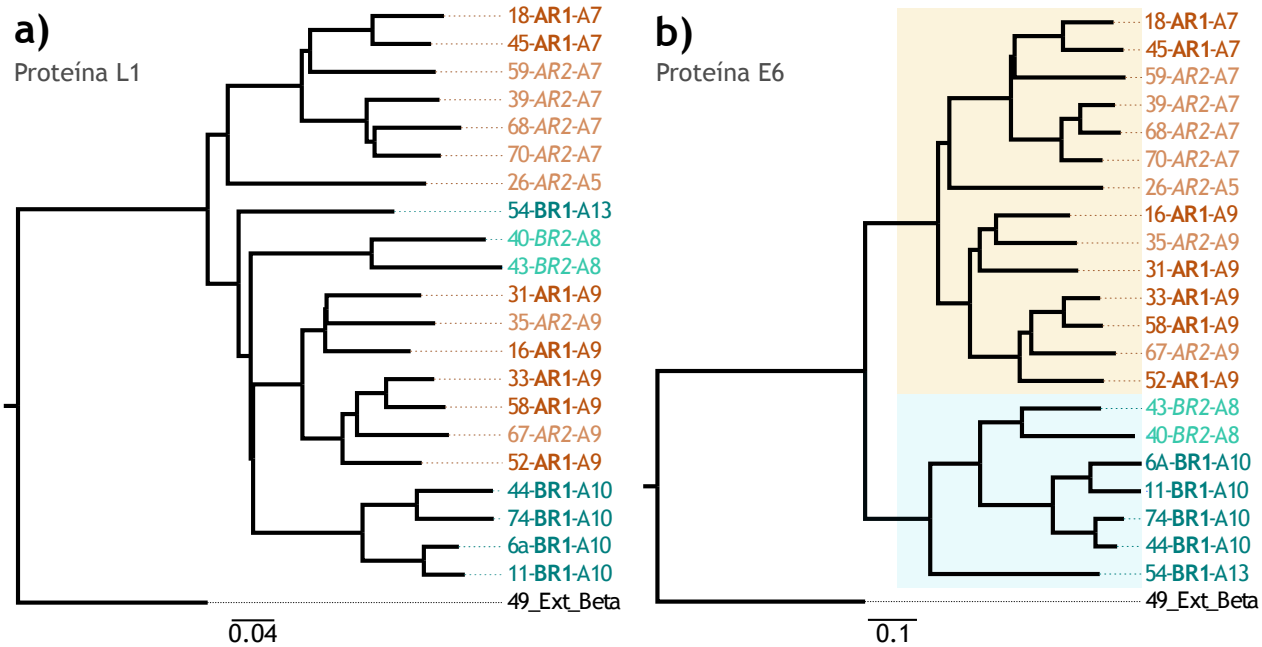


Figura 3: Fenogramas generados con las secuencias de las proteínas L1 y E6. ***a):*** Fenograma construido con las las 21 secuencias de aminoácidos de la proteína L1. La distribución de los grupos es similar al reportado por Villiers *et al.* (2004). ***b):*** Fenograma de la proteína E6. Se observan dos grupos definidos correspondientes a los tipos de alto riesgo (AR) y a los de bajo riesgo (BR), en café y azul, respectivamente. El *Betapapillomavirus* VPH-49 fue utilizado como grupo externo.

3.1.3. Comparación secuencia-estructura de la proteína E6

La identificación de la subsecuencia correspondiente al *pocket*[1] de unión a la proteína E6AP fue llevada a cabo utilizando el modelo cristalográfico publicado por Martinez-Zapien *et al.* (2016), con el número de acceso 4XR8. Dicho modelo posee la estructura de la proteína E6 VPH-16, unida al dominio central (*central core*) de p53 y a una proteína quimérica formada por la hélice *LxxLL* de la proteína E6AP humana (AAAELT**L**QE**LL**GEE) y la proteína periplásmica de unión a maltosa (MBP) de *E. coli* (Figura 4a).

Sin embargo, la proteína E6 de este modelo es en realidad una proteína mutada (E6 4C/4S, Martinez-Zapien *et al.* (2016); Zanier *et al.* (2013)) que posee una deleción de los primeros siete residuos (MHQKRTA), y con cuatro sustituciones C/S en los sitios 87, 104, 117 y 147 (Figura 4b). Por este motivo se utilizó como referencia la secuencia anotada P03126 (*Unirprot*), la cual es resultado de la traducción de la secuencia de nucleótidos del gen *e6* del VPH-16, obtenida mediante secuenciación (Seedorf *et al.*, 1985).

[1] Para este trabajo se define como "*pocket*.ª la cavidad en la proteína E6 que interactúa con el motivo *LxxLL* de E6AP y de otras proteínas celulares.

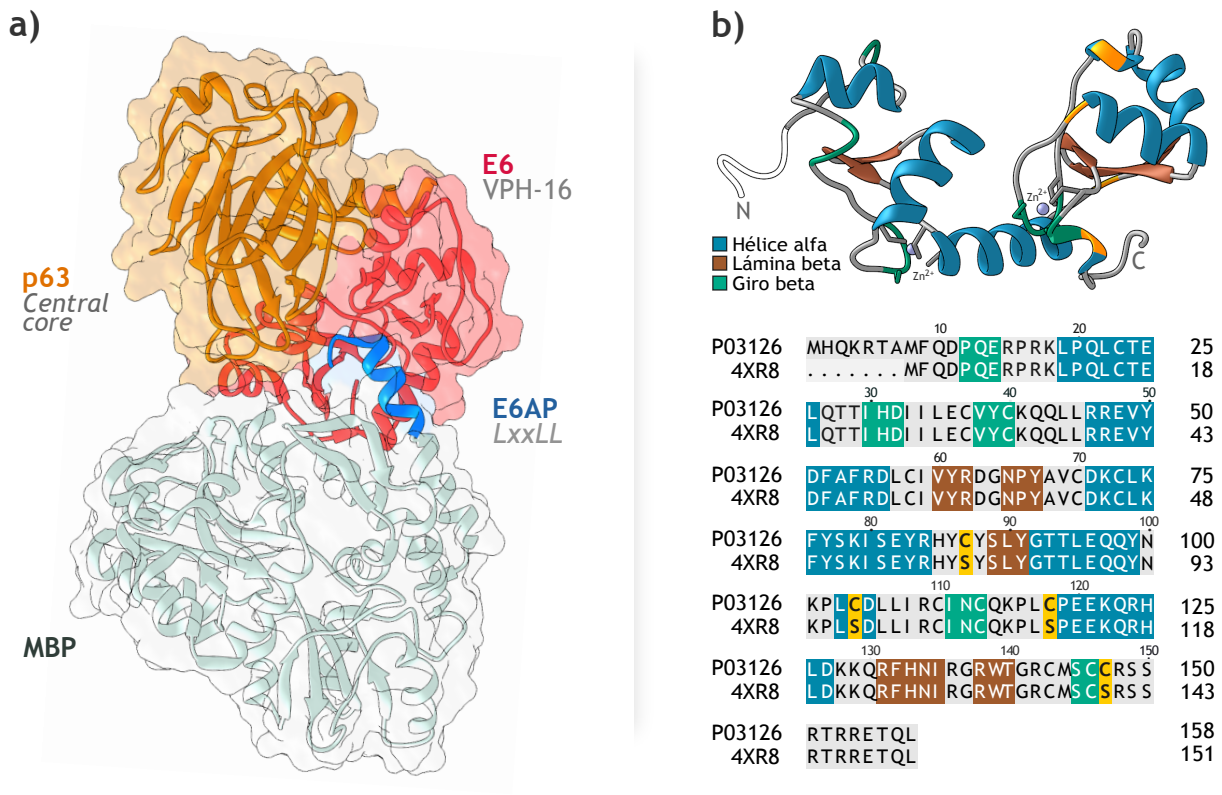


Figura 4: Representación esquemática de la estructura cristalográfica de la proteína E6 (modelo 4XR8, difracción de rayos X a 2.25 Å). a) Heterodimero 1 del complejo E6/E6AP/p53 (Martinez-Zapien *et al.*, 2016) requerido para la degradación de p53. **b)** Estructura y secuencia de aminoácidos de la proteína E6 (4XR8) en comparación con la secuencia P03126 (*Uniprot*). Se muestran las estructuras secundarias, sus subsecuencias, y los sitios mutados C/S.

3.1.4. Caracterización del *pocket* de unión a E6AP

Los aminoácidos que podrían estar interactuando con la hélice *LxxLL* fueron identificados al seleccionar los residuos de E6 que tuviesen al menos un átomo a una distancia $\leq$ 5 Å de cualquier átomo de la hélice *LxxLL*. Un total de 25 residuos fueron seleccionados (Figura 5), conformando una subsecuencia no continua de la cadena de aminoácidos de E6. Las posiciones correspondientes a estos residuos dentro de la secuencia P03126 fueron las siguientes: 17, 18, 38, 39, 52, 57, 58, 59, 60, 61, 62, 68, 69, 74, 77, 78, 81, 82, 85, 107, 109, 114, 136, 138, 139.

Por otro lado el análisis de topología con el servidor *CASTp* identificó esta misma región como la cavidad más grande de la proteína, con una superficie de 507.6 Å^2 y un área de apertura de 316.9 Å^2 (Figura 6a). Los residuos superficiales marcados por *CASTp* son los mismos identificados en el paso anterior, salvo por los residuos I59 y Y61, que no fueron señaladas como parte de la cavidad debido a que sus cadenas laterales se hallan en dirección opuesta a la zona de interacción con *LxxLL*, pero su cadena principal está a menos de 5 Å de la hélice (Fig. 5b superior). Cabe señalar, en la Figura 6b, que al visualizar la superficie del *pocket* es notoria la región

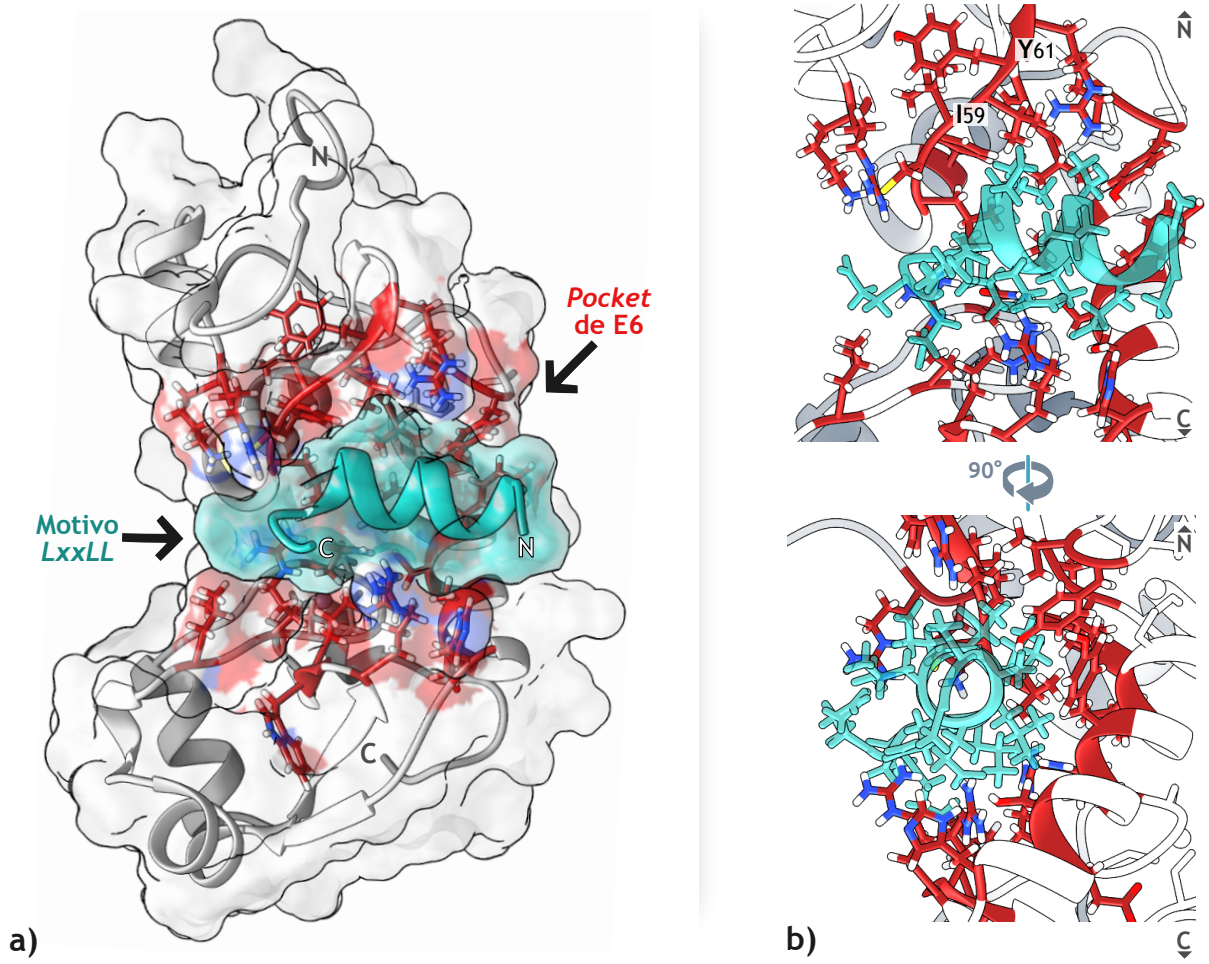


Figura 5: Estructura de la proteína E6 (VPH-6) unida a la hélice *LxxLL* de E6AP. a) Superficie y cadena principal de la proteína E6 (gris) en interacción con la hélice *LxxLL* (verde). Se muestran color rojo las cadenas laterales de los residuos de E6 a menos de 5 Å de la hélice. **b)** Acercamiento al *pocket* de unión de E6 a E6AP. Se muestra la estructura de las cadenas principales y laterales de los residuos que conforman el *pocket* de E6.

hidrofóbica, marcada en color naranja, que interactúa con las tres leucinas del motivo *LxxLL* (AAAELT**L**QE**LL**GEE). Por otro lado la zona azul del *pocket* es altamente hidrofílica e interactúa con los glutamatos 4 y 13, y la glutamina 8 de la hélice *LxxLL* (AAA**E**LTL**Q**ELLG**E**E).

Posteriormente, con las 25 posiciones del *pocket* obtenidas en el paso anterior, y tomando como referencia el alineamiento de la Sección 3.1.2, se identificaron los residuos pertenecientes al *pocket* de los 21 tipos de VPH analizados. Con esta fracción del alineamiento se construyeron dos logos de secuencia utilizando la herramienta *WebLogo*. Como se observa en la Figura 7, cada logo corresponde a los tipos BR y AR, respectivamente. El objetivo fue poner de manifiesto el grado de conservación de los residuos en cada posición, lo que más adelante fue usado como pauta para seleccionar ligandos capaces de interactuar con sitios conservados en AR, o en AR y BR juntos.

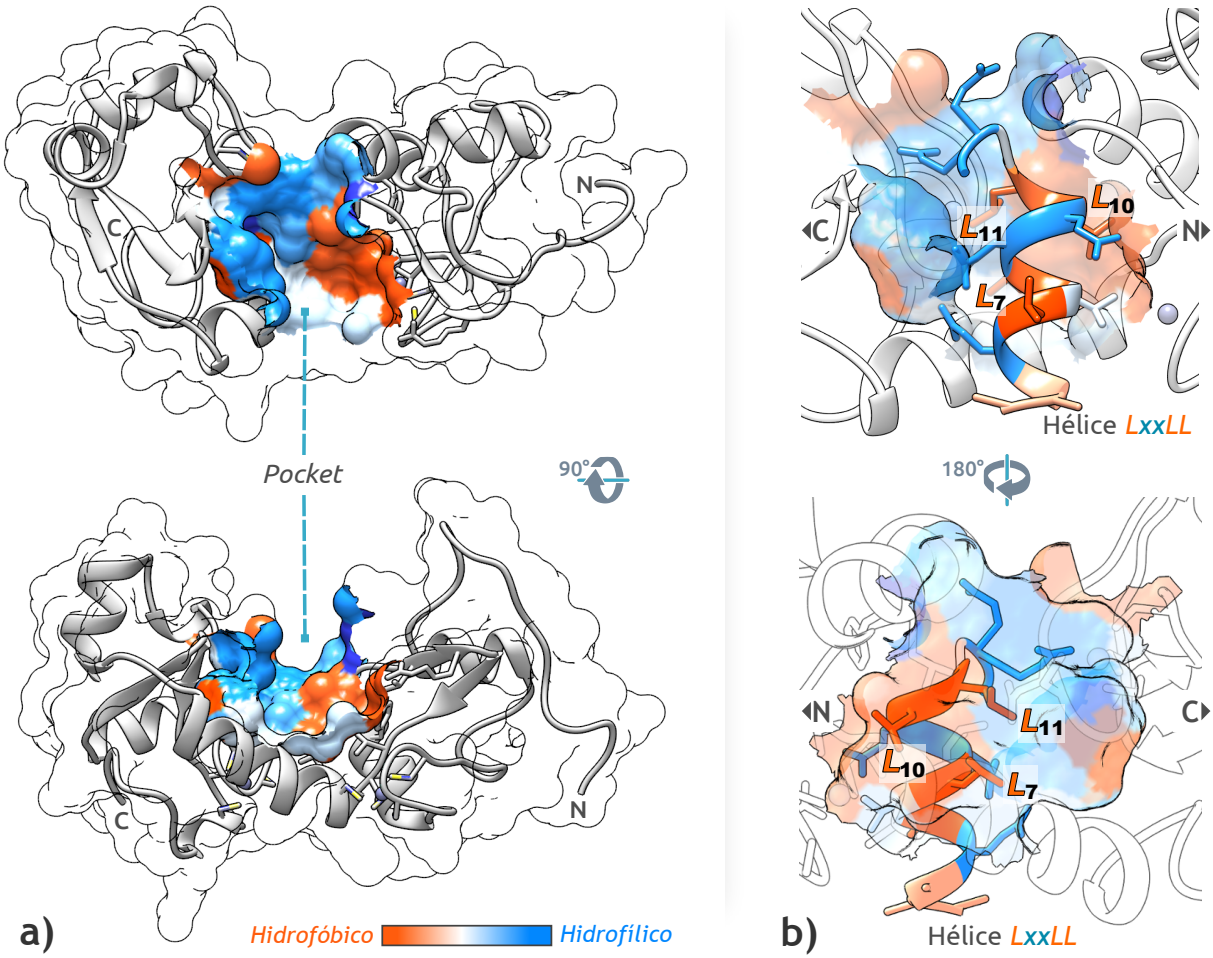


Figura 6: Cavidad en la proteína E6 correspondiente al pocket de unión a la hélice *LxxLL*. a) Superficie de la cavidad mayor identificada por *CASTp* en la estructura de E6. Se muestra la vista frontal (arriba) y la vista lateral (abajo) del *pocket* **b)** Acercamiento al *pocket* en interacción con la hélice, destacando la posición de los residuos de leucina del motivo *LxxLL*. Se muestra la vista frontal (arriba) y posterior (abajo). El color naranja corresponde a zonas con residuos con cadenas laterales altamente hidrofóbicas. En contraste, el color azul representa regiones con residuos altamente hidrofílicos.

Por último, se eligió modelar y evaluar la interacción con la hélice *LxxLL* de una proteína E6 de BR con el objetivo de compararla con los resultados de interacción proteína-hélice y proteína-ligando-hélice de la proteína E6 VPH-16 (ver Sección 3.5). Para seleccionar la proteína E6 de BR se consideró el valor de disimilitud entre las subsecuencias de los *pockets* y la prevalencia del tipo. Como resultado se eligió proceder con la proteína E6 del VPH tipo 11 como la representativa de VPH-BR. Los resultados de este análisis pueden verse en Apéndice A.4.

3.2. OBJETIVO 2: Refinamiento estructural de la proteína E6

El objetivo 2 está enfocado en la obtención de estructuras de la proteína E6 (AR) que representen el espacio conformacional de ésta en condiciones fisiológicas y cuando aún no se encuentra

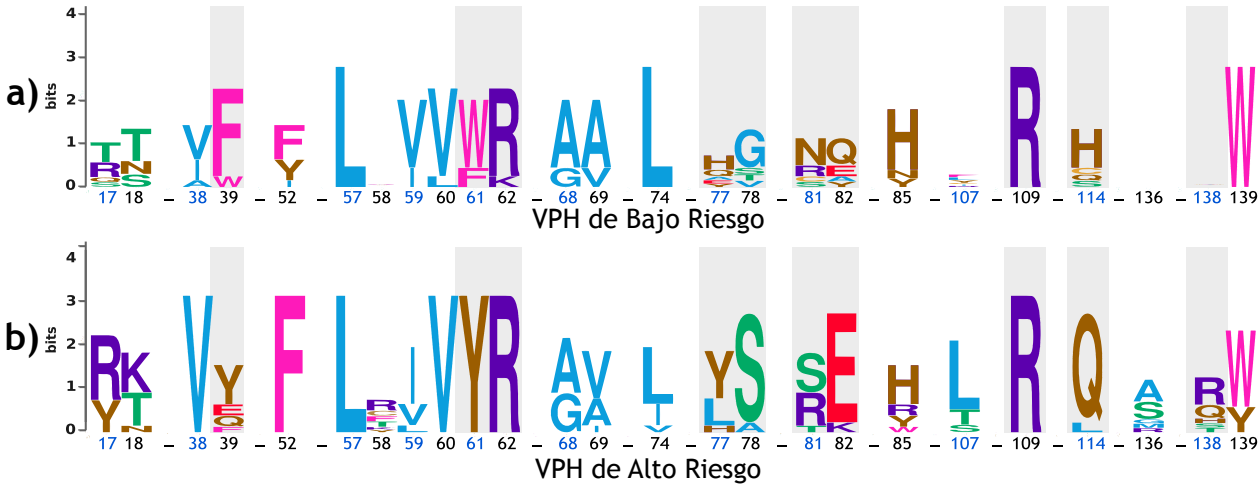


Figura 7: Logo de la subsecuencia del *pocket* de unión a E6AP con los 21 tipos de VPH analizados. a) Logo del *pocket* obtenido a partir del alineamiento de las siete secuencias de la proteína E6 de VPHs de bajo riesgo. **b)** Logo del *pocket* correspondiente al alineamiento de las catorce secuencias de la proteína E6 de VPHs de alto riesgo. Código de colores: morado = carga positiva, rojo = carga negativa, rosa = aromáticos-no polares, azul = alifáticos-no polares, café = polares-neutros, verde = polares-neutros-pequeños.

unida a ninguna proteína celular. Por este motivo, la presente sección destaca los resultados de la Simulación de Dinámica Molecular (SDM) de la proteína E6 del VPH-16 de AR (*E6 sola*). No obstante, con el objetivo de contrastar los resultados se muestran además algunos de los datos obtenidos en la SDM de E6 VPH-16 en interacción con la hélice *LxxLL* de E6AP.

3.2.1. Modelaje homólogo de la proteína E6

El *BLASTp* de la secuencia P03126 identificó como mejor molde a la cadena F del modelo 4XR8; con un *E-value* de 2.46×10^{110}, un *score* de 324, una identidad del 97.4 %, una similitud de 93.0 % y una cobertura del 96.0 %. Esta cadena se utilizó como molde para generar cien modelos homólogos utilizando el programa *Modeller*. Todos los modelos tuvieron un GA341 (*Model Score*) de 1.00 (valores arriba de 0.7 indican un modelo válido con un 95.0 % de tener un plegamiento correcto (John y Sali, 2003) y un valor de zDOPE de -0.5 (valores menores a 0 indican buenos resultados). Finalmente se seleccionó el mejor modelo (modelo P03126), el cual registró un zDOPE de -0.55, y un RMSD estimado de 0.8 Å. Los resultados del modelado homólogo de la proteína E6 VPH-11 se muestran en el Apéndice A.4.

3.2.2. Refinamiento por dinámica molecular

La SDM de la proteína E6 VPH-16 (modelo P03126) fue realizada con el paquete *Amber16* en dos experimentos independientes. El primer experimento consistió en la simulación de la proteína E6 sola (*E6 sola*). El segundo consistió en la simulación de E6 unida a la hélice *LxxLL* (*E6-hx*).

Para cada experimento se realizaron tres ensayos de SDM utilizando los sistemas correspondientes (*E6 sola* y *E6+hx*) y los mismos parámetros, con excepción de las velocidades iniciales de los átomos al comenzar la fase de SA, las cuales son asignadas aleatoriamente por *Amber16*. Dos experimentos más, aunque de un solo ensayo, fueron realizados con el modelo homólogo de E6 VPH-11 (P04019) y los resultados no mostrados en esta sección se muestran en el Apéndice A.4.

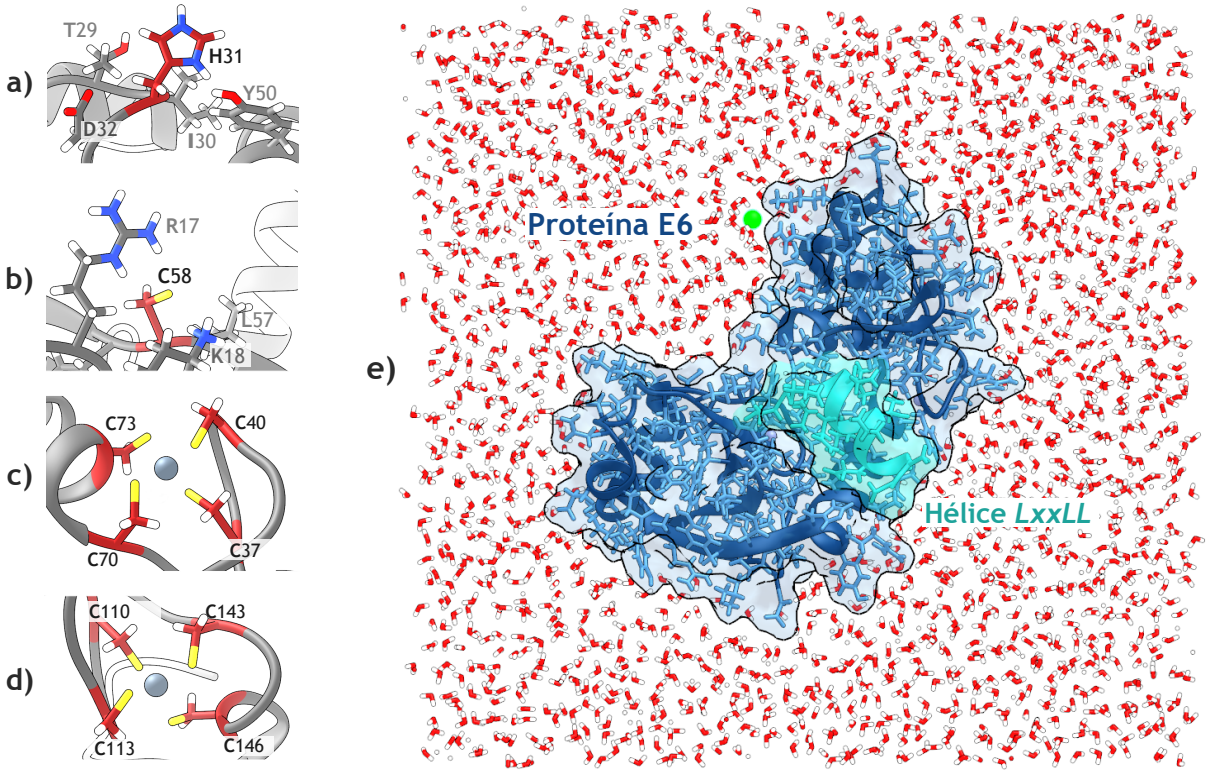


Figura 8: Preparación de la proteína E6 para la simulación de dinámica molecular. a) Histidina 31 de E6, muestra un estado biprotonado en el anillo imidazol (estado HIP) **b)** Cisteína 58 en su forma aniónica (tiolato), debido a su cercanía con ARG-17 y LYS-18. **c)** y **d)** Dedos de Cisteína-Zinc característicos de la proteína E6, altamente conservados entre los tipos de VPH (ver Figura 2). **e)** Caja de agua con la proteína E6 en interacción a la hélice *LxxLL* (*E6-hx*) que en conjunto conforman un sistema de aproximadamente 45 000 átomos. Para efectos de visualización se han omitido las moléculas de agua en los planos anterior y posterior de la proteína.

3.2.2.1. Preparación de los sistemas *E6 sola* y *E6-hx*

Previo a la SDM, tanto E6 como la hélice fueron evaluadas utilizando la herramienta *PROPKA* para determinar los estados de protonación de los residuos a un pH de 7.0. Cuatro de las cinco histidinas de E6 tuvieron una forma imidazol con el protón en el nitrógeno delta. No obstante, la histidina 31 mostró el estado catiónico (+1) con los nitrógenos delta y épsilon protonados (Figura 8a). La cisteína 58 tuvo una carga de -1 al estar desprotonada (Figura 8b). Por otro lado, los dos dedos de Cys_4Zn ($C_{37}C_{40}C_{70}C_{73}$ y $C_{110}C_{113}C_{143}C_{146}$) fueron construidos utilizando la topología y el campo de fuerza *ZAFF* (figuras 8c y 8d). La carga neta final de la proteína fue de +8, y de la

hélice -4. Finalmente, tanto *E6 sola* como *E6+hx* fueron solvatadas en una caja de agua y cada sistema se neutralizó agregando átomos de Cl^- (Figura 8e).

3.2.2.2. Minimización, *Simulated annealing* y equilibrado

Los resultados de las fases de minimización, SA y equilibrado fueron evaluados para cada ensayo de cada experimento. En la Figura 9 pueden observarse dichos resultados tomando como ejemplo uno de los ensayos del sistema *E6 sola*. En el inciso *a* se muestra que la energía potencial inicial del sistema llegó a descender hasta -1.7x10^5 kcal/mol tras los 10,000 ciclos de minimización.

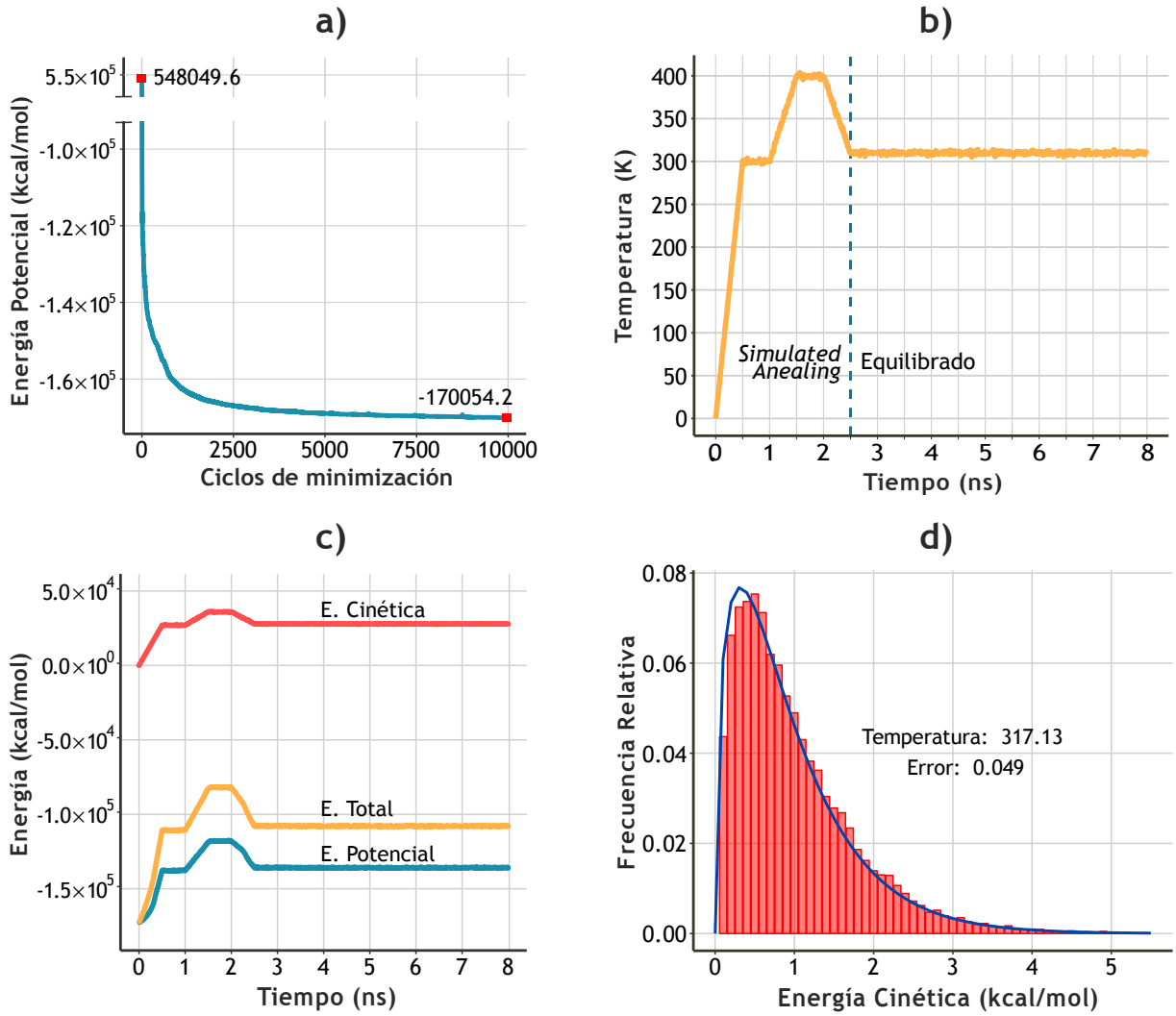


Figura 9: Minimización, *Simulated annealing* y equilibrado del sistema *E6 sola*. a) Descenso en la energía potencial del sistema *E6 sola* tras los 10 000 ciclos de minimización, partiendo de la estructura solvatada del modelo homólogo P03126. **b)** Temperatura del sistema durante el proceso de *SA* y Equilibrado. **c)** Valores de energía Cinética (rojo), Potencial (azul) y Total (amarillo) del sistema durante el *SA* y Equilibrado. **d)** Histograma de las energías cinéticas observadas para cada átomo del sistema (rojo) y densidad de probabilidad de Maxwell-Boltzmann (línea azul) correspondiente a las energías cinéticas esperadas para un sistema a T = 317.13 K.

Las figuras 9b y 9c muestran la variación en la temperatura y las energías del sistema a lo largo de los 8 ns de simulación correspondientes a la fase de SA (0 a 2.5 ns) y equilibrado (2.5 a 8 ns). Los últimos 5.5 ns en estas gráficas muestran la estabilidad del sistema en las energías y en la temperatura a 310 K. Finalmente, la gráfica en la Figura 9d corrobora que el sistema se encontró en equilibro termodinámico al finalizar la fase de equilibrado, ya que la distribución observada de las energías cinéticas de los átomos del sistema se ajustan (con un error < 0.05) a la distribución de Maxwell-Boltzmann de energías cinéticas a una temperatura de 317.13 K (Ecuación 1).

3.2.2.3. Producción de dinámica molecular

En la fase de producción de SDM, las propiedades físicas del sistema también fueron evaluadas en cada ensayo. La medición de las energías, la temperatura, el volumen y la presión fueron registradas cada 0.02 ns a lo largo de los 100 ns de la producción. La Figura 10 muestra, a modo de ejemplo, los resultados de los ensayos del sistema *E6 sola*. Como puede verse, las propiedades evaluadas se mantuvieron cercanas a los esperados para el sistema: la temperatura se mantuvo estable y próxima a los 310 K (Figura 10b), y el volumen tuvo una media de 482.5 nm^3 con una variación de $\pm$ 1 nm^3 (Figura 10c). Por su parte, la densidad del sistema se mantuvo a 0.995 g/cm3, un valor cercano a la densidad del agua a T = 310 K (36.85 ℃) y a P = 1 bar (0.99 atm).

Por otro lado, la presión mostró una amplia variación y la presencia de valores negativos (Figura 10d). Sin embargo, este comportamiento es esperable, aún para la simulación de un sistema NTP. Esto se debe a que el valor de la compresibilidad isotérmica (β_T) del agua en estado líquido es muy bajo, lo que significa que pequeños cambios en el volumen del sistema conllevan a grandes cambios en los valores de presión.

Por tanto, de acuerdo con la documentación de *Amber16* (Case *et al.*, 2016), para mantener la media de la presión cercana al valor deseado en un sistema NTP (en este caso P_0 = 1 bar) el volumen del sistema es regulado en relación con el valor de P en un tiempo t; los valores negativos de P_t corresponden a la aplicación de una fuerza que actúa disminuyendo el tamaño de la caja, y los positivos a una fuerza actuando para aumentar el tamaño de ésta. Por lo anterior, fluctuaciones mayores a 100 bar en P_t son esperables, siempre que la media de P se mantenga cercana al valor esperado (Case *et al.*, 2016), en este caso 1.8 bar.

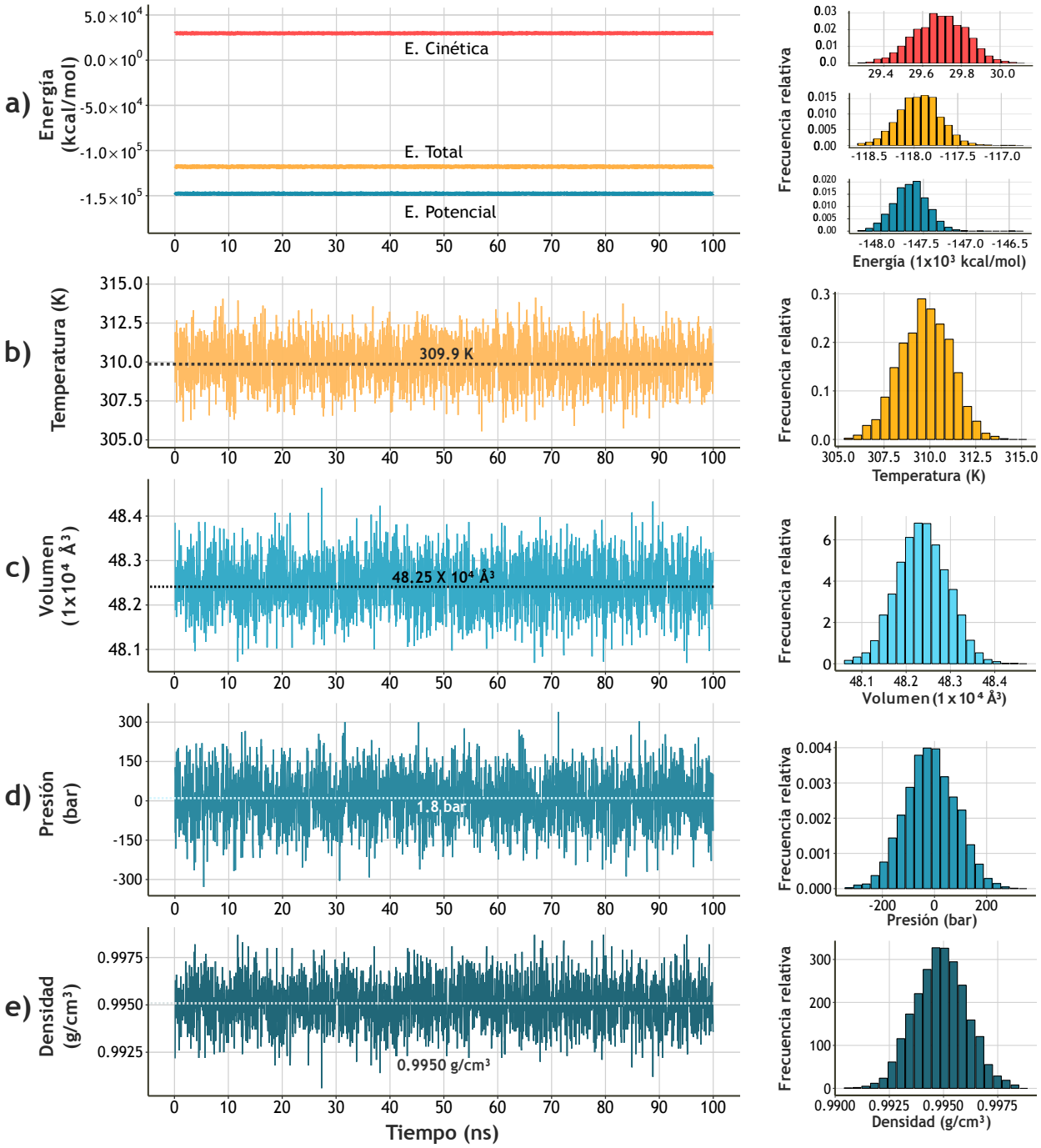


Figura 10: Propiedades físicas del sistema *E6 sola* evaluadas en fase de producción de SDM. Se muestra la fluctuación de los valores de cada propiedad a lo largo de los 100 ns de la simulación (derecha), y los histogramas con la frecuencia relativa de los valores de la propiedad correspondiente (izquierda). **a)** Valores de energía Cinética (rojo), Potencial (azul) y Total (amarillo) del sistema. **b)** Temperatura del sistema, con un valor promedio de 309.9 K. **c)** Volumen del sistema, con un valor promedio de 482.5 nm^3. **d)** Presión del sistema, con un valor promedio de 1.8 bar. **e)** Densidad del sistema, con un valor promedio de 0.995 g/cm^3. Cabe mencionar que el 93 % de la masa total del sistema corresponde al agua.

3.2.2.4. Análisis de las trayectorias de dinámica molecular

Para este análisis cada trayectoria se representó en una matriz de 5000 filas por 3x158 columnas. Donde las columnas corresponden a las coordenadas cartesianas del Cα de cada residuo de la proteína, y cada fila a un conformero (*frame*) de la trayectoria. La Figura 11 muestra los resultados del cálculo RMSD (Ecuación 2) de los tres ensayos de *E6 sola*. La Figura 11a corresponde los valores de RMSD de toda la cadena principal de la proteína, a excepción de los primeros cinco residuos de los extremos N-terminal y C-terminal. Por su parte, la Figura 11b muestra los valores de RMSD de los tres ensayos considerando únicamente los residuos correspondientes al *pocket* (Sección 3.1.4). Por último, la Figura 11c permite la apreciación de la distribución de los valores de RMSD en cada ensayo, tanto para el *pocket* (rosa, amarillo o rojo) como para la proteína total (gris).

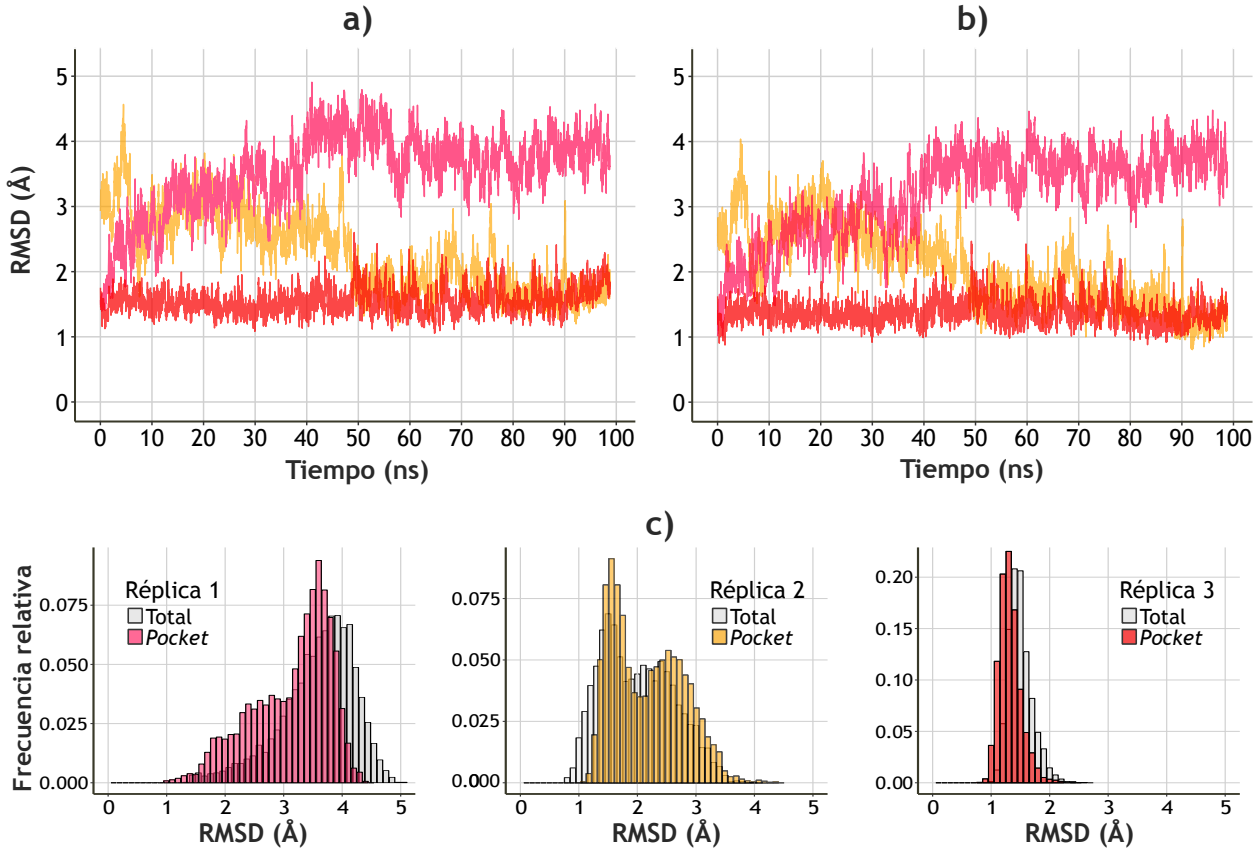


Figura 11: RMSD de la proteína E6 en los tres ensayos del sistema *E6 sola*. a) RMSD de la cadena principal (Cα) de la proteína E6 en cada ensayo. **b)** RMSD considerando únicamente los Cα de los residuos correspondientes al *pocket* de E6. **c)** Histogramas de RMSD de la cadena principal Total y del *pocket* para cada ensayo. Aunque se observan pequeñas diferencias, la variación conformacional del *pocket* es similar a la que presenta toda la cadena principal de E6 a lo largo de los 100 ns de la fase de producción.

Un análisis equivalente fue realizado para el sistema *E6+hx*. Los resultados de los tres ensayos realizados, mostrados en la Figura 12, resaltan la similitud entre las conformaciones de E6 a lo largo de la trayectoria. Evidenciando una mayor rigidez conferida a E6 cuando se encuentra unida a la hélice de E6AP a través del *pocket* de unión.

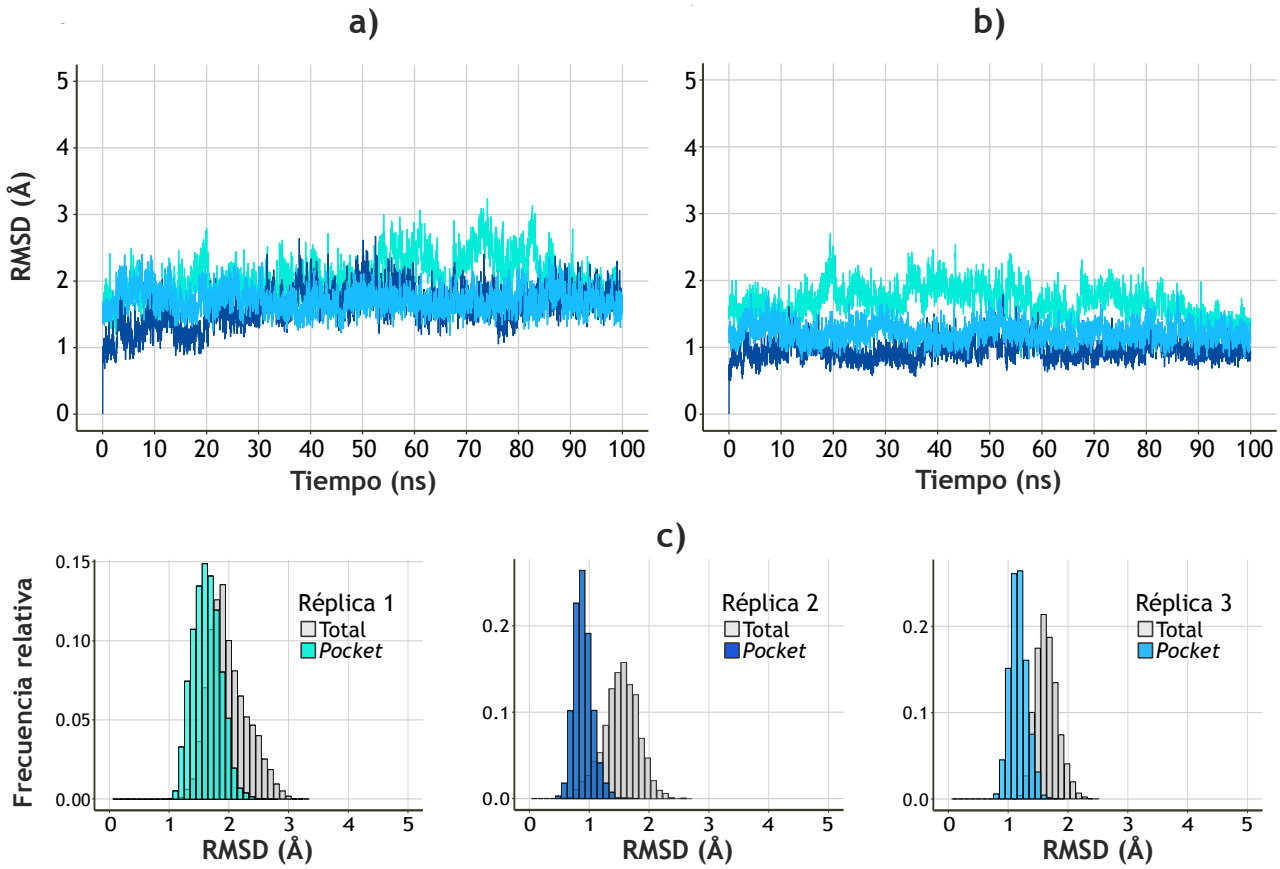


Figura 12: RMSD de la proteína E6 en los tres ensayos del sistema *E6+hx*. a) RMSD de la cadena principal (Cα) de la proteína E6 en cada ensayo. **b)** RMSD considerando únicamente los Cα de los residuos correspondientes al *pocket* de E6. **c)** Histogramas de RMSD de la cadena principal Total y del *pocket* para cada ensayo.

En el caso del RMSF (Ecuación 3), la Figura 13 condensa los resultados obtenidos para el Cα de cada aminoácido de E6 en los sistemas *E6 sola* (rojo) y *E6+hx* (azul). La gráfica muestra la media y la desviación estándar del RMSF de cada residuo en los tres ensayos de cada experimento. Como puede observarse, la fluctuación en Å de los residuos de E6 tiende a ser menor, en términos generales, cuando se encuentra unida a la hélice. Esta rigidez de *E6-hx* se ve representada a nivel de estructura en la Figura 13c que posee valores de fluctuación (en Å) más cercanos a cero, sobre todo en la región correspondiente al *pocket* de unión a E6AP. Por su parte, en la Figura 13b, E*6 sola* mostró una mayor flexibilidad, sobre todo en el extremo N-terminal.

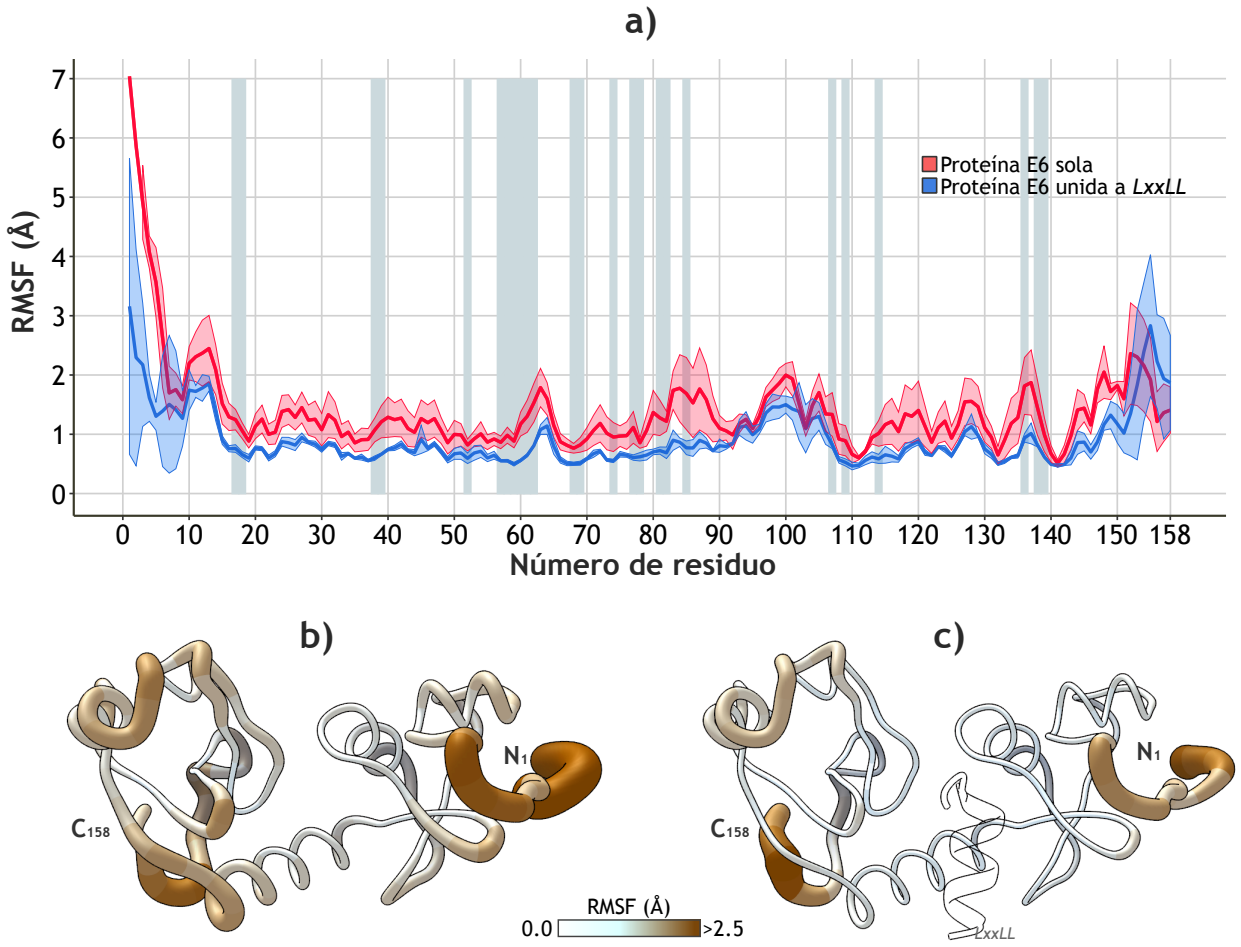


Figura 13: RMSF de la proteína E6 en los sistemas *E6 sola* y *E6-hx*. a) Gráfica con los valores de RMSF de cada residuo de la proteína E6 en los experimentos *E6 sola* y *E6-hx*. Las líneas opacas corresponden a la media de los valores de RMSF de los tres ensayos de cada experimento y las bandas semitransparentes a su desviación estándar. *E6 sola* = rojo, y *E6-hx* = azul. Las columnas grises del fondo indican las posiciones de los residuos correspondientes al *pocket*. **b)** y **c)** Representación estructural de la cadena principal de E6 que muestra, en un rango de color (blanco a café) y d grosor, las regiones más flexibles de la proteína de acuerdo con el valor de RMSF de cada residuo en cada experimento: *E6 sola* (b) y *E6-hx* (c).

3.2.2.5. Análisis de componentes principales y *clustering* jerárquico

Para analizar la dinámica y obtener conformeros representativos de la proteína E6, se realizaron Análisis de Componentes Principales utilizando las trayectorias de los tres ensayos de *E6 sola*. La Figura 14 muestra los resultados del ACP para cada ensayo. En el inciso *a* aparecen las gráficas de sedimentación (*scree plots*) con el porcentaje de varianza correspondiente a cada eigenvector. Para los ensayos 1 y 2, los tres primeros CP explican más del 75 % de la varianza conformacional. Sin embargo, esto no ocurre en el ensayo 3, donde apenas el 60 % de la varianza es descrito por los primeros tres CP.

En la Figura 14b, se muestran los diagramas de dispersión de los conformeros en el subespa-

cio correspondiente a los dos primeros CP. Finalmente, la Figura 14c la superposición de estructuras obtenidas a partir de la interpolación de la trayectoria en términos del primer componente principal, en cada ensayo. La apertura de cada estructura en determinadas regiones representa las contribuciones de cada residuo a la varianza en el primer CP. Esto a su vez corresponde al grado de desplazamiento de cada residuo en dicha región a lo largo de la trayectoria (en términos del primer eigenvector). Lo que, en última instancia, permitió evidenciar los movimientos dominantes de la proteína (Grant *et al.*, 2006).

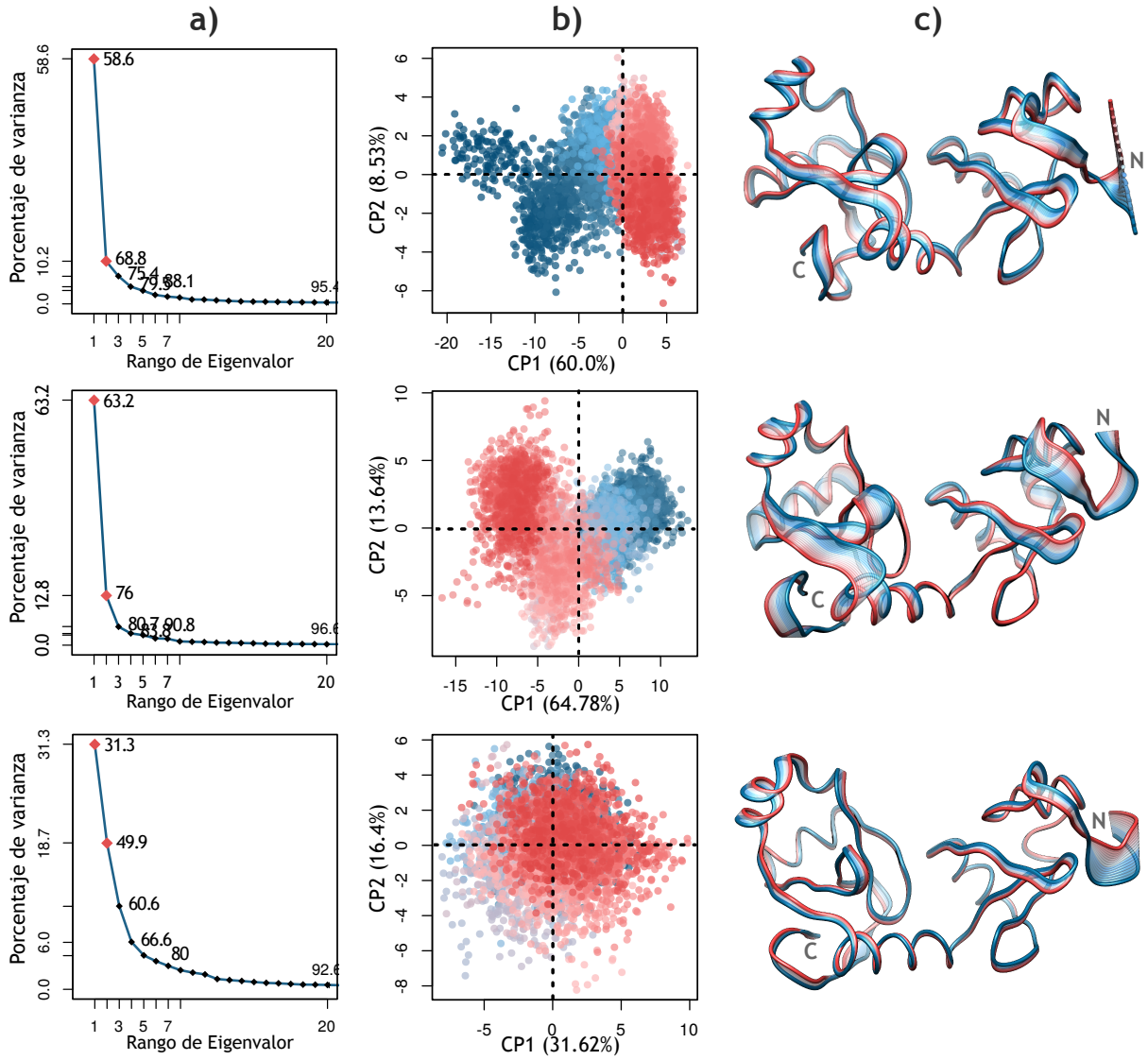


Figura 14: Análisis de Componentes Principales de las trayectorias de SDM de *E6 sola*. a) Gráficas de sedimentación de los ACP mostrando la varianza correspondiente a los primeros 20 CP de cada ensayo. **b)** Diagramas de dispersión de los conformeros de la trayectoria en el subespacio del los dos primeros CP de cada ensayo. La escala de color continua indica la posición de cada conformero en la trayectoria (rojo = 0 ns a azul = 100 ns). **c)** Estructuras superpuestas que representan una ”trayectoria” obtenida a partir de las conformaciones más disimiles (de rojo a azul) en el subespacio del CP1 de cada ensayo. La apertura de las cadenas muestra las zonas con mayor movimiento en la proteína según lo descrito por el primer CP.

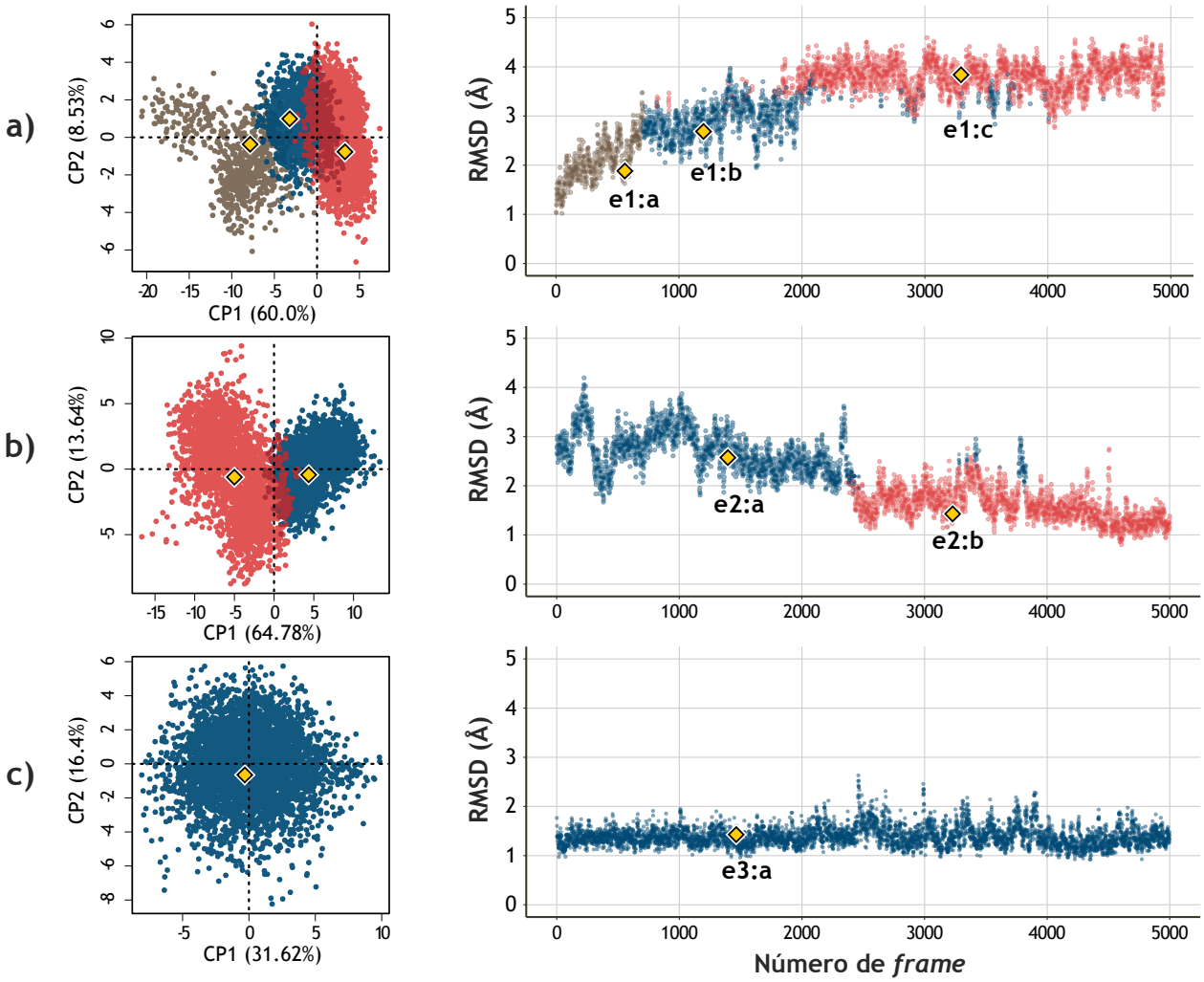


Figura 15: Clusters conformacionales seleccionados en cada trayectoria de *E6 sola*. A la izquierda, diagramas de dispersión de los conformeros en los primeros dos eigenvectores del ACP. A la derecha, gráficas de RMSD de cada ensayo (Figura 11) en la que cada punto corresponde a un conformero de E6. Cada color (rojo, azul o café) representa un grupo de estructuras. El rombo amarillo represnta la conformación representativa seleccionada para cada conjunto de conformeros. **a)** Réplica 1 con tres grupos seleccionados. **b)** Réplica 2 con dos grupos seleccionados. **c)** Réplica 3 con un único grupo seleccionado.

Posteriormente con el *clustering* jerárquico se identificaron los principales estados conformacionales de la proteína *E6 sola* en cada trayectoria. El *clustering* fue efectuado sobre los primeros CP de cada ensayo, y la decisión del número de grupos (k) y de CP a utilizar fue tomada con base en la inspección visual de las gráficas de RMSD, los diagramas de dispersión del ACP y los dendogramas generados por el propio *clustering* jerárquico (Figura A.3). Para el caso del ensayo 1 se tomaron en cuenta los tres primeros CP (75.4 % de varianza) y las conformaciones se agruparon en tres grupos distintos (Figura 15a). Para el ensayo 2, se utilizaron los dos primeros CP (76 % de varianza) y un k = 2 (15b). Finalmente, para el ensayo 3 se seleccionaron los cuatro primeros CP (66.6 % de varianza), aunque esto no tuvo un efecto relevante, pues se decidió que todas las conformaciones de la trayectoria corresponderían a un sólo *cluster* (15b).

Para la selección de las conformaciones representativas de E6, primero se obtuvo la media aritmética de las coordenadas de los Cα de todos los *frames* pertenecientes a cada grupo. Después, se identificó el conformero que tuviera el valor mínimo de RMSD frente a la estructura media calculada, pero sólo los residuos del *pocket* fueron considerados. Esto dio como resultado la identificación de seis conformeros de la proteína *E6 sola* (Figura 15). Por último, la Figura 16a muestra la superposición de las cadenas principales de los conformeros elegidos, y la gráfica adyacente (16b) contempla los valores del RMSD calculados para cada par de estructuras.

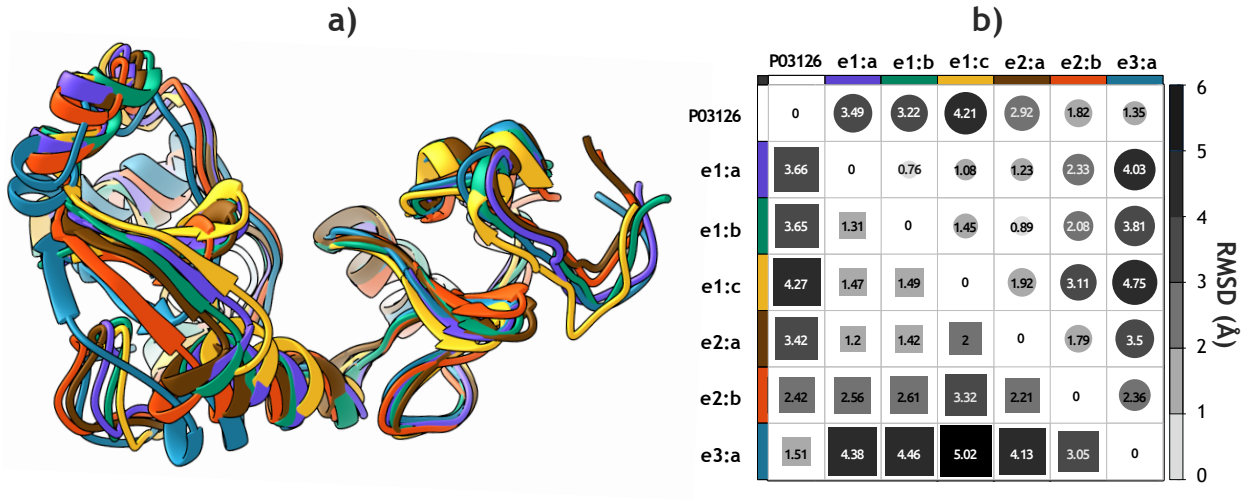


Figura 16: Conformaciones representativas de la dinámica de la proteína E6. a) Estructuras superpuestas de los seis conformeros elegidos en la Figura 15. **b)** Valores de RMSD de cada par de conformeros, incluyendo al modelo P03126. Los cuadrados representan el valor de RMSD de toda la cadena principal. Los círculos corresponden al RMSD tomando en cuenta sólo los residuos del *pocket*. La superposición fue realizada con la herramienta *MatchMaker* de *UCSF Chimera* (Pettersen *et al.*, 2004)**.**

3.3. OBJETIVO 3: Cribado virtual de las propiedades ADMETox de los ligandos

3.3.1. Obtención de las moléculas de referencia

Un total de 26 moléculas de referencia fueron elegidas a partir de la revisión bibliográfica. La Tabla 3.3.1 muestra la estructura y el nombre de cada molécula, su código de registro en la base de datos *ZINC15* y la fuente bibliográfica a partir de la que fue identificada. Adicionalmente, el gráfico de Tanimoto de la Figura 17 muestra la relación estructural entre las 26 moléculas elegidas. Cabe mencionar que la actividad anti-VPH de las moléculas elegidas ha sido previamente evaluada en ensayos *in vitro*, *in silico*, o en ambos. Sin embargo, en algunos casos no necesariamente ha sido demostrado que estas moléculas actúen directamente sobre la proteína E6, y tampoco ha sido evidenciada su eficacia en experimentos *in vivo*.

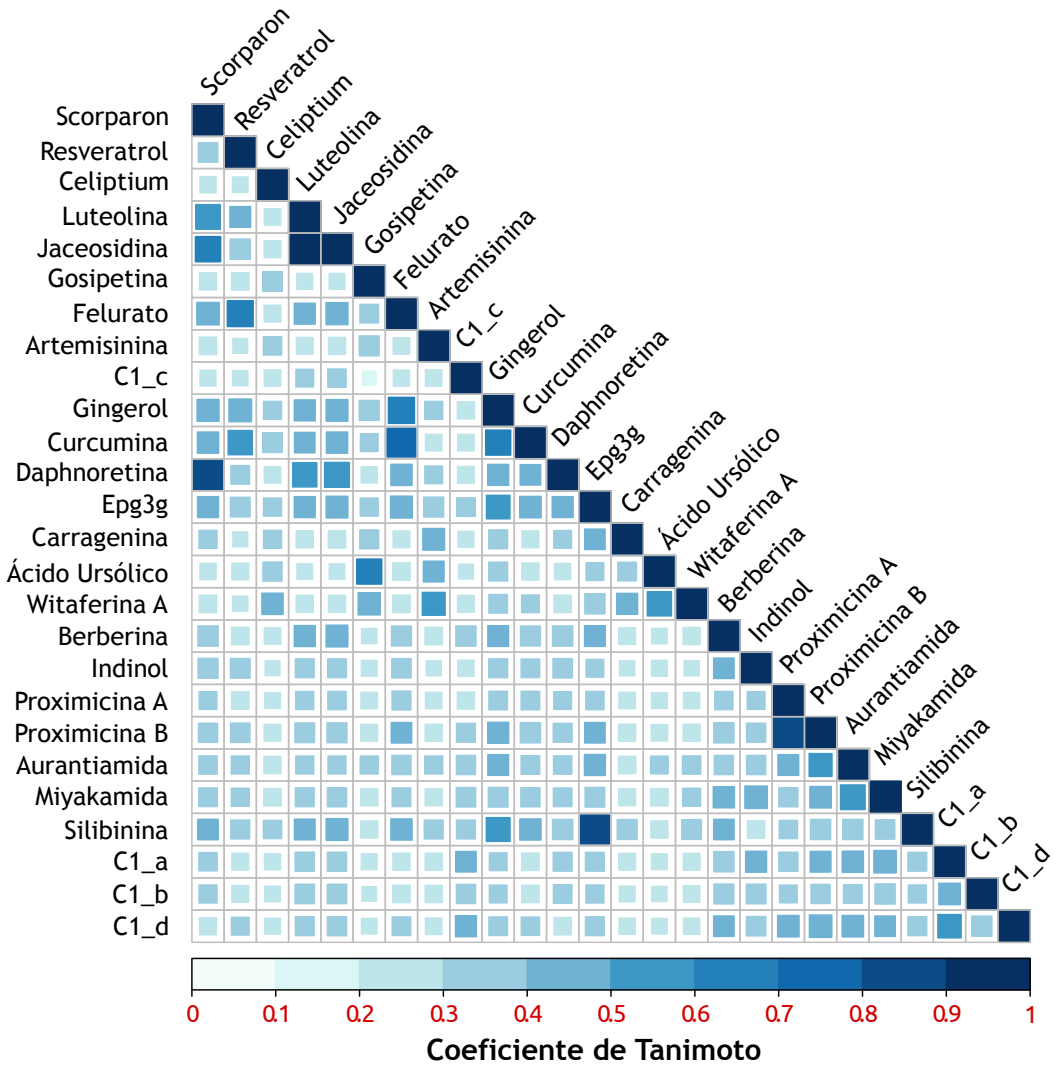


Figura 17: Gráfico de Tanimoto comparando las 26 moléculas de referencia obtenidas de la bibliografía. Valores del coeficiente de Tanimoto (Ecuación 4) calculados para cada par de moléculas. Valores cercanos a 1 (color azul oscuro) implican una similitud química total entre las moléculas comparadas.

Posteriormente, cada molécula sirvió como consulta en la base de datos *ZINC15* para buscar más moléculas con características estructurales relacionadas. Los métodos de búsqueda utilizados fueron subestructura, coeficiente de Tanimoto y coeficiente de Dice. Como resultado se obtuvo un total de 34 804 moléculas, las cuales fueron descargadas en formato SMILES de la base de datos *ZINC15*. Después, para obtener las estructuras tridimensionales de las moléculas, se utilizó *LigPrep* (Schrödinger, 2017a) y el campo de fuerza OPLS3, generando un total de 35,319 moléculas. Un número mayor al inicial debido a que se consideraron los estados de protonación que podrían tener algunas de las moléculas a un pH de 7 ± 0.2.

Tabla 2: Moléculas de referencia elegidas a partir de la revisión bibliográfica realizada.

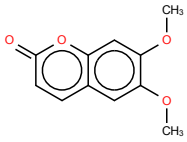
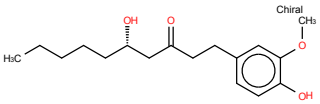
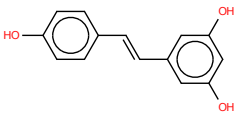
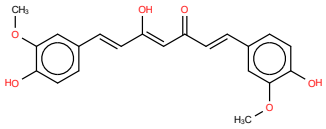
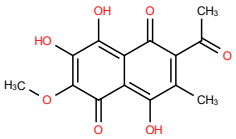
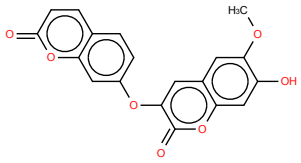
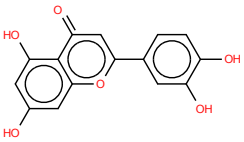
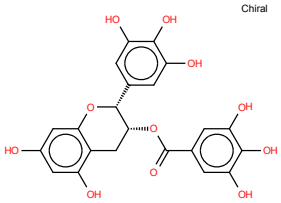
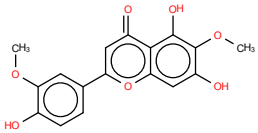
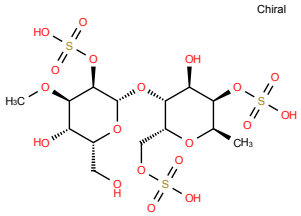
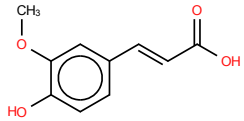
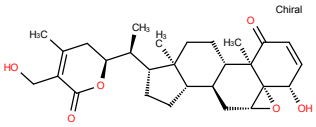
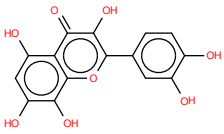
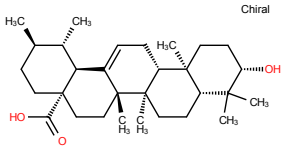
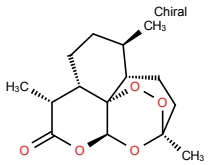
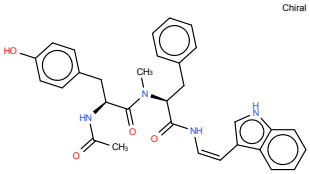
Estructura	Datos	Estructura	Datos
	Escoparona ZINC000000002067 (Rietz *et al.*, 2016)		Gingerol ZINC000001531846 (Kapoor *et al.*, 2016; Kumar *et al.*, 2016)
	Resveratrol ZINC000000006787 (García-Zepeda *et al.*, 2013; Kumar *et al.*, 2015a)		Curcumina ZINC000100067274 (Divya y Pillai, 2006; Maher *et al.*, 2011; Basu y et al, 2013; Singh y Misra, 2013; Mamgain *et al.*, 2015)
	Celiptium ZINC000100089149 (Mamgain *et al.*, 2015; Kumar *et al.*, 2015b)		Daphnoretina ZINC000000689683 (Mamgain *et al.*, 2015)
	Luteolina ZINC000018185774 (Malecka *et al.*, 2014; Ham *et al.*, 2014; Shah *et al.*, 2015; Cherry *et al.*, 2013)		Epg-3-g ZINC000003870412 (García-Zepeda *et al.*, 2013; He *et al.*, 2013; Kumar *et al.*, 2016)
	Jaceosidina ZINC000014779854 (Lee *et al.*, 2005; Kumar *et al.*, 2016; Cherry *et al.*, 2013)		Carragenano ZINC000064373300 (Buck *et al.*, 2006)
	Felurato ZINC000000058258 (Palasap *et al.*, 2014; Kumar *et al.*, 2015b)		Witaferina-A ZINC000064373300 (Park *et al.*, 2015; Munagala *et al.*, 2011; Kumar *et al.*, 2015a)
	Gosipetina ZINC000006525297 (Malecka *et al.*, 2014)		Ácido-Ursólico ZINC000003978827 (Kassi *et al.*, 2007; Yim *et al.*, 2006; Kumar *et al.*, 2015a)
	Artemisinina ZINC000008681831 (Jansen *et al.*, 2011; Kumar *et al.*, 2016)		Miyakamida ZINC000014594537 (John y Kotadiya, 2015)

Tabla 2: Moléculas de referencia elegidas a partir de la revisión bibliográfica realizada.

Estructura	Datos	Estructura	Datos
	Indol-3-carbinol ZINC000000158743 (Bell *et al.*, 2000b)		Proximicina-A ZINC000043073751 (John y Kotadiya, 2015)
	Berberina ZINC000003779067 (Saha y Khuda-Bukhsh, 2015)		Proximicina-B ZINC000043078506 (John y Kotadiya, 2015)
Chiral	Acetato de Aurantiamida ZINC000013374322 (John y Kotadiya, 2015)	Chiral	Silibinina ZINC000002033588 (Pirouzpanah *et al.*, 2015; Huang *et al.*, 2005; Yu *et al.*, 2012)
	C1-a ZINC000000100126 (Kumar *et al.*, 2015a)		C1-b ZINC000009679272 (Kumar *et al.*, 2015a)
Chiral	C1-c ZINC000000100126 (Kumar *et al.*, 2015a)		C1-d ZINC000009679272 (Kumar *et al.*, 2015a)

3.3.2. Cribado virtual utilizando los descriptores ADME-Tox

El siguiente paso fue la obtención de los descriptores moleculares utilizando el módulo *QikProp* (Schrödinger, 2017b). De los 51 descriptores generados sólo cinco fueron elegidos para llevar a cabo el filtro ADMETox del conjunto de compuestos obtenido en el paso anterior. Los filtros también fueron aplicados a las 12 moléculas aprobadas por la FDA, mencionadas en la Sección 2.3.4. Este proceso fue llevado a cabo utilizando un *script* de R para descartar las moléculas que no cumplieran con las condiciones establecidas (Tabla 1). Los resultados se muestran en la Tabla 3, destacando el número de compuestos filtrados en cada paso. También se hace énfasis en qué moléculas aprobadas por la FDA fueron descartadas en cada etapa de selección. Al final, un total de 19,119 moléculas cumplieron con los criterios definidos en el cribado, con lo cual fueron elegidas para pasar al siguiente análisis.

Tabla 3: Número de moléculas filtradas en cada descriptor de *Qikprop* utilizado.

	Descriptor	Criterio utilizado para el filtrado	Número de moléculas filtradas	Moléculas FDA descartadas
1)	**Regla de 5 de Lipinsky**	$\leq$ 1 (máximo una violación a la regla)	34 804 a 27 674	Atazanavir y sofosvuvir
2)	**Regla de tres de Jorgensen**	$\leq$ 1 (máximo una violación a la regla)	27 674 a 24 682	Ninguna
3)	**Absorción Oral Humana**	Valor de 2 o 3	24 682 a 20 856	Amoxicilina
4)	**Número de *stars* de *Qikprop***	$\leq$ 1 (máximo un descriptor fuera de rango)	20 856 a 20 578	Ninguna
5)	**Penetración dérmica**	$\leq$ Valores entre -8.0 y 1.0	20 578 a 19 119	Ninguna

3.4. OBJETIVO 4: Cribado virtual basado en acoplamiento molecular

3.4.1. Acoplamiento Molecular con *Autodock Vina*

La primera fase del proceso de Simulación de Acoplamiento Molecular (SAM) fue llevada a cabo con el programa *Autodock Vina*, evaluando el acoplamiento de las 19,119 moléculas (Sección 3.3.2) con cada una de las estructuras representativas de la proteína *E6 sola* (Sección 3.2.2.5). Sin embargo, debido al elevado número de ligandos a evaluar, de los seis conformeros disponibles sólo cuatro fueron elegidos: *R1:c*, *R2:a*, *R2:b* y *R3:a*. Éstos pertenecen a los grupos más grandes obtenidos en la Sección 2.2.3.3 (Figura 15), y pueden observarse en la Figura 18, la cual recalca cada residuo del *pocket* de unión a E6AP. Además, muestra los valores del área superficial accesible al solvente, volumen y apertura de la cavidad correspondiente; todos calculados utilizando el servidor *CASTp*.

Al finalizar la SAM con *Vina*, cada acoplamiento dio como resultado una pose y un valor de energía de interacción (ΔG) para cada ligando en cada conformero. Estos valores fueron usados para elegir a los mejores ligandos, es decir, aquellos capaces de interactuar eficientemente con E6. Para ello se seleccionaron a las 3,000 moléculas con los valores más bajos de ΔG en cada conformero, se obtuvieron cuatro conjuntos de 3,000 compuestos cada uno. Después se efectuó la intersección de los cuatro conjuntos, con lo que se identificaron 834 ligandos los cuales habían obtenido valores de interacción favorables con las cuatro conformaciones de E6. La Figura 19a facilita la visualización de este proceso al mostrar los valores de interacción de cada ligando con tres

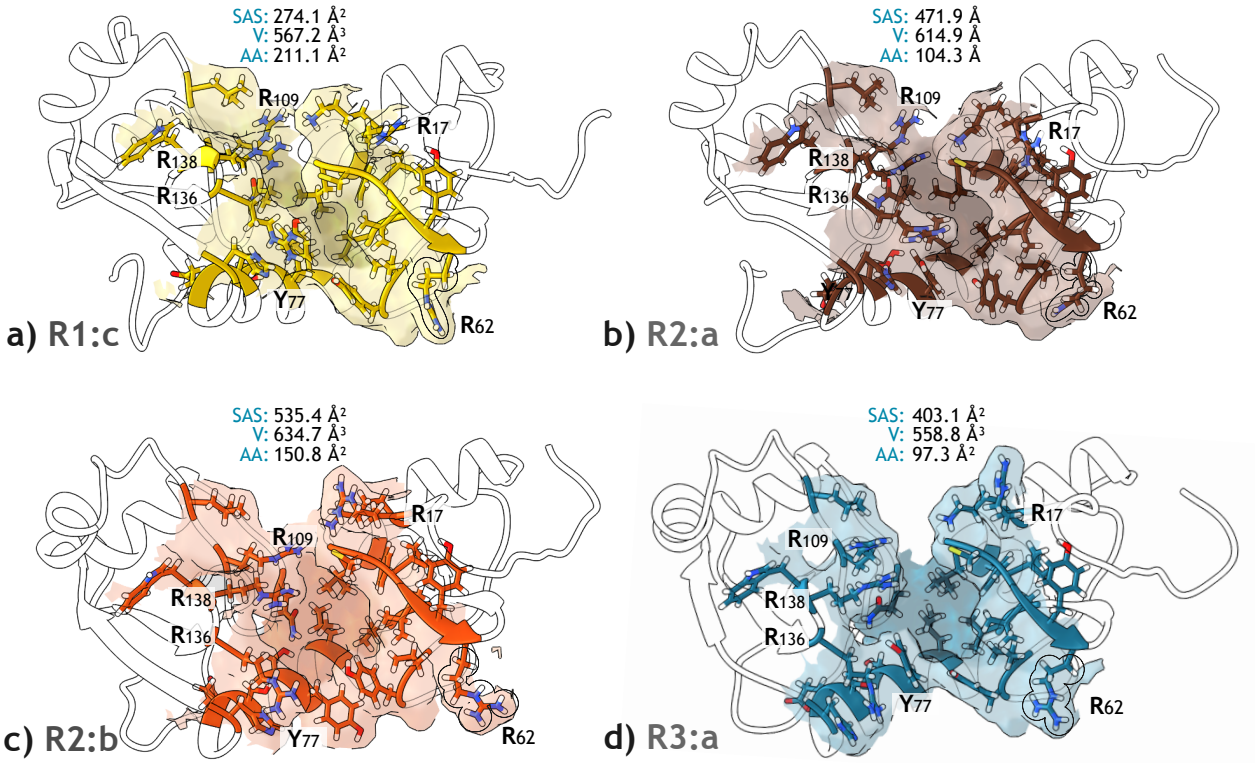


Figura 18: *Pocket* de unión a E6AP de las conformaciones de E6 utilizadas para la SAM. Se muestran los aminoácidos correspondientes al *pocket* de cada conformero y la representación de la superficie de las cadenas laterales. Además, se muestran las dimensiones del área superficial accesible al solvente (ASAS), el volumen y el área de apertura (AA) de cada cavidad. **a)** Réplica 1, confórmero *c*. **b)** Réplica 2, confórmero *a*. **c)** Réplica 2, confórmero *b*. **d)** Réplica 3, confórmero *a*.

de las conformaciones evaluadas. Como puede verse, los mejores ligandos se concentran en la esquina inferior izquierda, es decir, en los valores más bajos de energía para cada coordenada. A modo de referencia se destaca la posición de los 26 ligandos obtenidos de la revisión bibliográfica y la de las moléculas aprobadas por la FDA.

3.4.2. Acoplamiento molecular con *Autodock 4*

Tras la selección de las 834 moléculas elegidas con *Vina*, el siguiente paso fue la aplicación de la SAM utilizando *Autodock 4*. La Figura 19b muestra los resultados de los acoplamientos los 834 ligandos evaluados. Al igual que en la Figura 19a, se resalta la posición de los ligandos de referencia y los aprobados por la FDA, que aunque no estuviesen incluidos dentro de las 834 moléculas también fueron incluidos en el análisis. Finalmente, los 100 ligandos con valores más bajos fueron elegidos para pasar al siguiente filtro. En la Figura A.5 (Apéndice A), se muestra el gráfico de similitud de Tanimoto entre éstos y las 26 moléculas iniciales de referencia. Además, la Figura A.5a indica qué residuos del *pocket* de E6 han contribuido más en la formación de enlaces de hidrógeno al interaccionar con estos 100 ligandos.

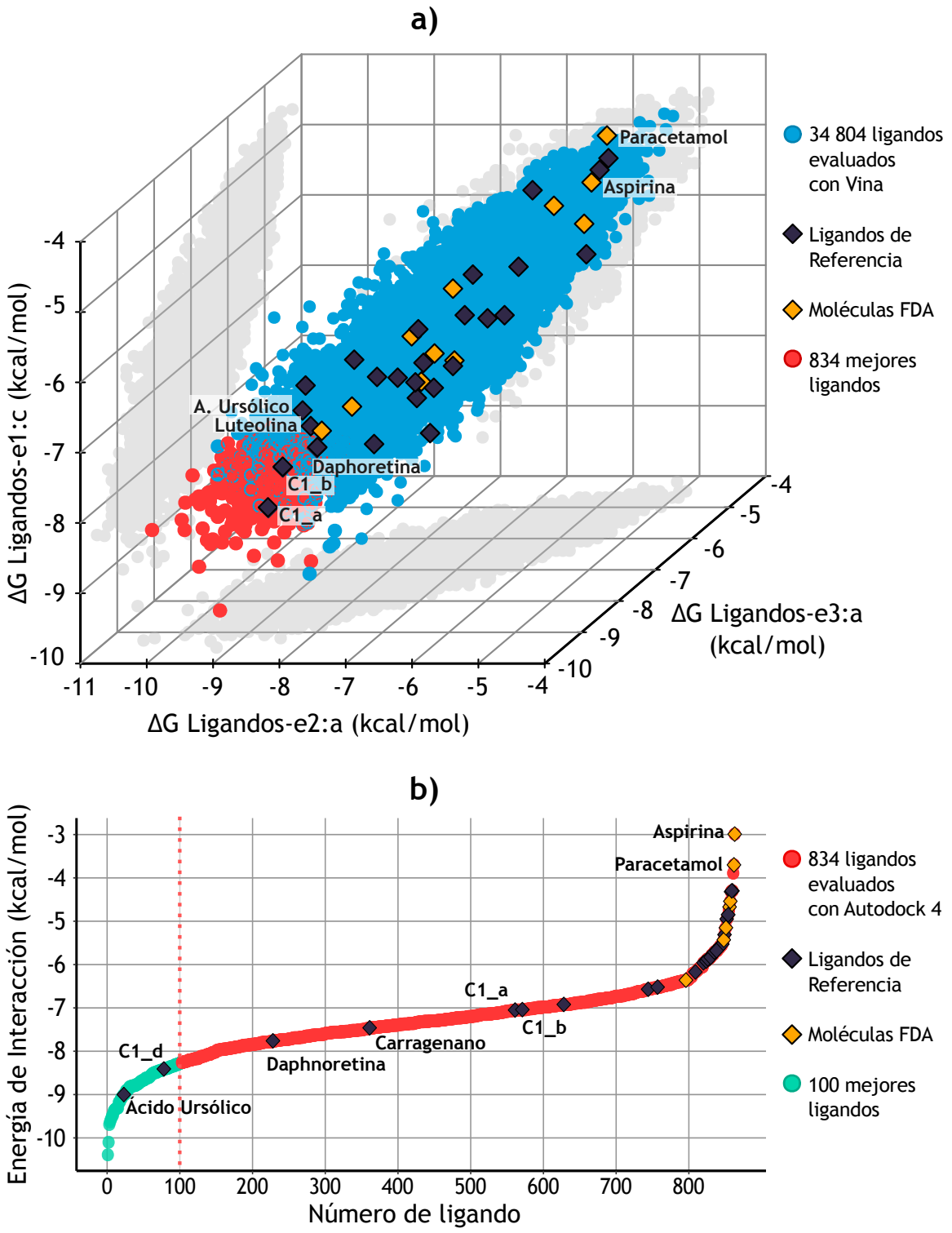


Figura 19: Anclaje molecular de moléculas pequeñas a la proteína E6 VPH-16. a) Resultado del anclaje molecular llevado a cabo con *Autodock Vina*. 19 119 moléculas con tres de las conformaciones elegidas de E6. Las 834 moléculas con valores de interacción más favorables (rojo) fueron seleccionadas para evaluarlas con *Autodock 4*. Para efectos de visualización sólo se muestran los resultados de tres de las conformaciones. **b)** Resultado de la evaluación de las 834 moléculas con *Autodock 4*. Las 100 moléculas con valores de energía de interacción más bajos fueron seleccionadas. Tanto para *Vina* como para *AD4* las moléculas de referencia y las aprobadas por la FDA fueron también analizadas.

3.5. OBJETIVO 5: Validación de las moléculas candidato

3.5.1. Validación del acoplamiento con MM/GBSA y MM/PBSA

Empleando los métodos MM/GBSA y MM/PBSA se llevó a cabo la evaluación de los acoplamientos entre la proteína E6 y cada uno de los 100 ligandos seleccionados. Primero fue necesario preparar la proteína y el ligando acoplado utilizando las herramientas LEaP y *Antechamber*, tal como se menciona en las sección 2.5. Después, a cada sistema *E6-lig* solvatado se le realizó un proceso de minimización de la energía potencial (Sección 2.2.2.2) y finalmente se evaluó la energía de interacción (ΔG) proteína-ligando con los métodos MM/GBSA y MM/PBSA. Cabe aclarar que, si bien ambos métodos consideran la contribución entrópica ($-T\Delta S$) para el cálculo (Ecuación 6), en el presente trabajo se asumió un valor de $\Delta S = 0$, debido al costo computacional de su cálculo y a la baja precisión que ha mostrado en estudios previos (Genheden y Ryde, 2015). Por lo anterior, los valores calculados de ΔG son relativos y no representan el valor experimental de interacción, pues sólo fungen como un punto de comparación para evaluar la afinidad e cada ligando a la proteína E6. Por lo tanto, en adelante se hará referencia a energías de interacción ΔG^{GB} y ΔG^{PB}.

La figuras 20a y 20b muestran los valores de ΔG^{GB} y ΔG^{PB} calculados. De manera adicional, en el inciso *c* se muestra el número de enlaces de hidrógeno que cada ligando forma con los residuos del *pocket* de E6. A modo de referencia, las moléculas identificadas de la bibliografía: ácido ursólico (Kassi *et al.*, 2007; Yim *et al.*, 2006), Cd_1 (Kumar y Zhang, 2015), daphnoretina (García-Zepeda *et al.*, 2013; He *et al.*, 2013), luteolina (Cherry *et al.*, 2013; Malecka *et al.*, 2014) y gosipetina (Malecka *et al.*, 2014) fueron incluidas en este análisis. Además del ácido acetil salicílico. Posteriormente, estos resultados fueron tomados en cuenta para elegir los mejores ligandos en esta fase al considerar los ΔG^{GB} y ΔG^{PB} más bajos, además de tomar en cuenta el número de enlaces de hidrógeno que puede llegar a formar y la similitud estructural calculada utilizando el coeficiente de Tanimoto, mostrado en la Figura A.6.

Cuatro moléculas fueron elegidas para la etapa final de validación con dinámica molecular y corresponden a los ligandos 5, 32, 70 y 82 de la Figura 20. Los acoplamientos de estas moléculas con E6 mostraron los valores más bajos tanto de ΔG$^{MM/GBSA}$ y ΔG$^{MM/PBSA}$ y como mínimo dos enlaces de hidrógeno con los residuos de E6. Por otro lado, el ligando 5 mostró buenos resultados de interacción pero fue seleccionado principalmente por su diferencia estructural en comparación con los otros tres ligandos (ver Figura A.6). La Figura 21 muestra los diagramas de interacción

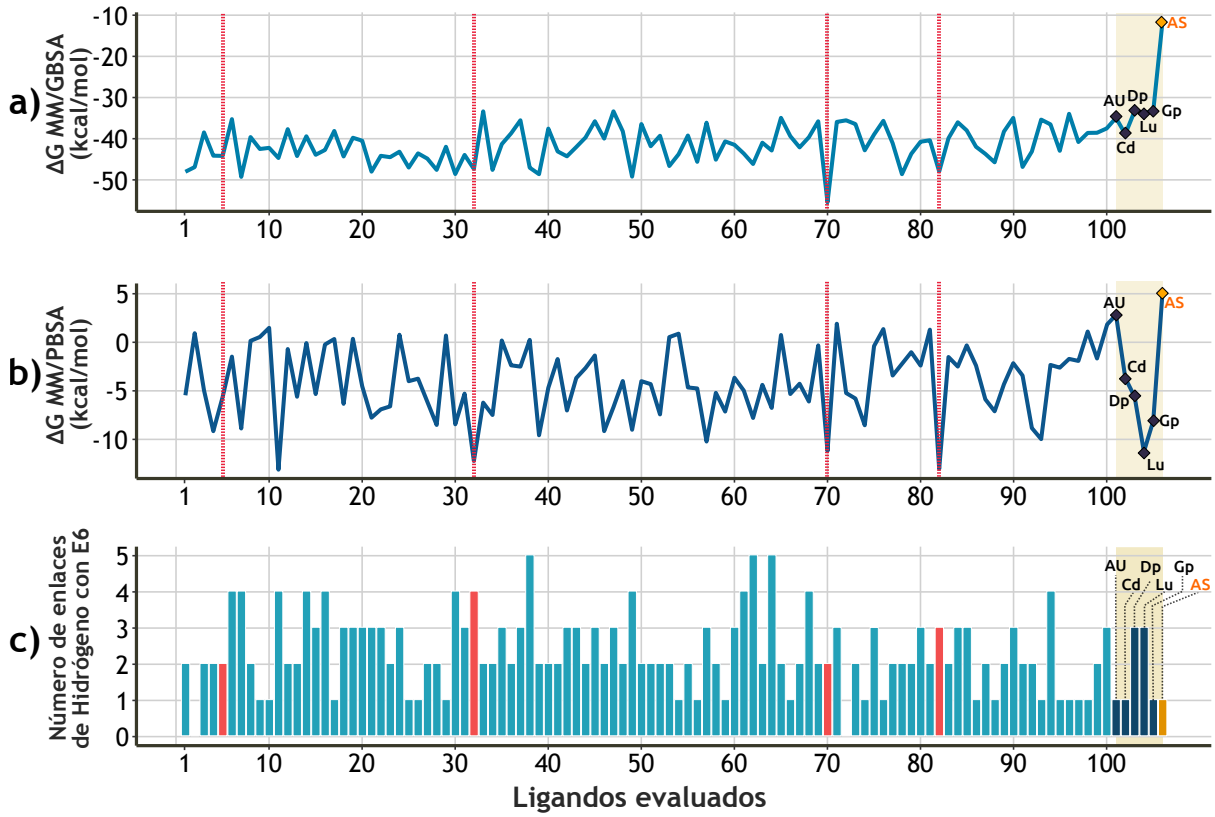


Figura 20: Análisis de la interacción de los 100 ligandos obtenidos con *Autodock 4*. a) y **b)** Valores de ΔG^{GB} y ΔG^{PB} calculados para cada ligando. **c)** Número de enlaces de hidrógeno que cada ligando forma con la proteína E6. Los cinco ligandos elegidos para validación por dinámica molecular se muestran en color rojo.

de cada molécula candidato con los residuos del *pocket* de E6 de acuerdo a los acoplamientos obtenidos con *Autodock 4*.

3.5.2. Validación del acoplamiento con dinámica molecular

La validación *in silico* de los cuatro ligandos candidato se llevó a cabo en dos fases de SDM. La primera consistió en la SDM del complejo *proteína-ligando* utilizando los acoplamientos de E6 con cada una de las cuatro moléculas elegidas y también con la molécula de referencia luteolina. Elegida como control positivo debido a que su actividad como inhibidor de la interacción E6-E6AP ha sido evaluada tanto *in silico* como *in vitro*, mostrando resultados favorables (Shah *et al.*, 2015; Malecka *et al.*, 2014; Ham *et al.*, 2014; Cherry *et al.*, 2013). Un análisis extra fue llevado a cabo utilizando un complejo *proteína-hélice*, evaluando la interacción de la proteína E6 y la hélice *LxxLL*, para lo cual se simularon dos sistemas: proteína *E6 VPH-16 + hx* y proteína *E6 VPH-11 + hx*. Los resultados de este análisis se muestran en las figuras A.7 y A.8.

En la segunda fase se simuló la dinámica del complejo *proteína-ligando-hélice*. Para ello, a cada estructura de acoplamiento *E6-lig* (Figura 21) se le colocó la hélice *LxxLL* en una posición

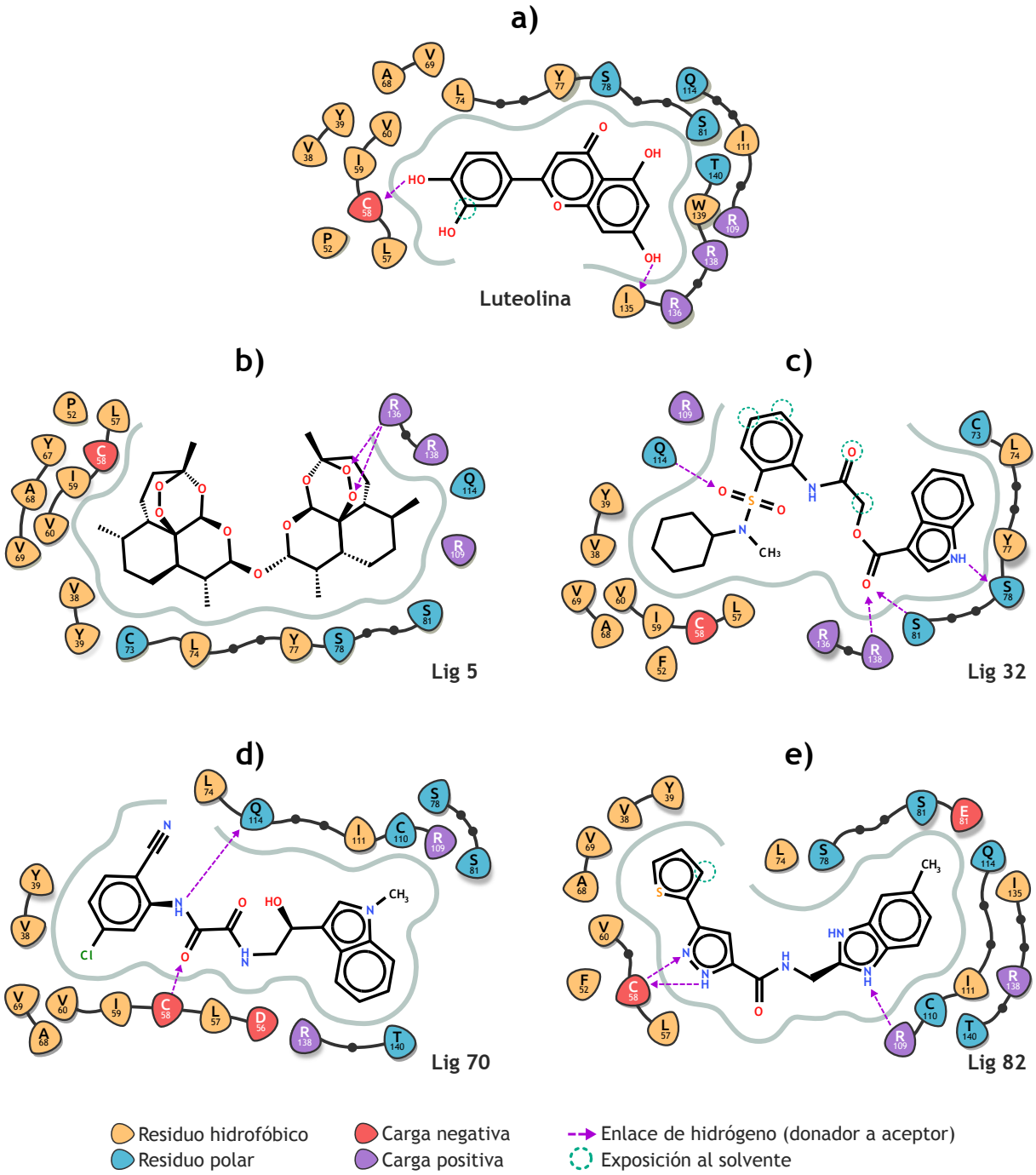


Figura 21: Diagramas de interacción con los residuos E6 de las cuatro moléculas candidato. Se muestra la estructura de los cuatro ligandos candidato y los residuos del *pocket* de E6 con los que interacciona. **a) Luteolina:** ΔG^{GB}= -44.2 kcal/mol y ΔG^{PB}= -5.6 kcal/mol. Establece enlaces de Hidrógeno (H) con la cadena principal de los residuos C58 e I135. **b) Ligando 5:** ΔG^{GB}= -44.2 kcal/mol y ΔG^{PB}= -5.6 kcal/mol. Dos de sus Oxígenos (O) son capaces de establecer un enlace de H con la cadena lateral del residuo R136. **c) Ligando 32:** ΔG^{GB}= -47.6 kcal/mol y ΔG^{PB}= -12.2 kcal/mol. Establece enlaces de H con los residuos Q114, S78 y S81/R136. **d) Ligando 70:** ΔG^{GB}= -55.4 kcal/mol y ΔG^{PB}= -11.2 kcal/mol. Forma enlaces de H con Q114 y con el H del amino de la cadena principal de C58. **e) Ligando 82:** ΔG^{GB}= -47.9 kcal/mol y ΔG^{PB}= -13.0 kcal/mol. Establece un enlace de H con la cadena lateral del residuo R109, y dos con la cadena principal de C58; uno con el H del grupo amino y otro con el O del grupo carboxilo. Diagramas generados utilizando el programa *Maestro* (Schrödinger, 2017a)

Tabla 4: Resultados principales de las SDM de los sistemas *E6+lig* y *[E6+lig]+hx*.

Nombre, fórmula molecular e ID en *ZINC15*	**Sistema *E6+lig***			**Sistema *[E6+lig]+hx***		
	RMSD lig (Å)	$\Delta\overline{G}^{GB}$ (kcal/mol)	$\Delta\overline{G}^{GB}$ (kcal/mol)	Dist.(Å) E6-hx	$\Delta\overline{G}^{GB}$ (kcal/mol)	$\Delta\overline{G}^{PB}$ (kcal/mol)
hx: Hélice *LxxLL*	*NA*	*NA*	*NA*	9.5±0.6	-69.3±9.5	-15.5±13.7
Luteolina ($C_{15}H_{10}O_6$) ZINC000018185774	0.6±0.2	-29.45±4.08	-0.4±4.8	14.1±0.6	-48.7±8.1	-25.11±9.5
lig5 ($C_30H_46O_9$) ZINC000049833587	1.4±0.6	**-45.07±7.4***	5.62±6.4	14.3±0.5	-46.2±6.4	-13.3±7.1
lig32 ($C_{24}H_{27}N_3O_5S$) ZINC000017118054	2.3±0.3	-33.97±5.9	2.19±4.4	12.9±1.0	-51.1±8.2	-17.8±10.9
lig70 ($C_{20}H_{17}ClN_4O_3$) ZINC000096096545	0.8±0.2	-37.65±3.77	0.62±4.5	13.5±0.8	-52.0±5.9	-12.7±8.3
lig82 ($C_{19}H_{19}N_5OS$) ZINC000362643639	0.9±0.2	-34.8±3.32	**-0.77±4.7***	13.2±0.4	**-38.5±7.3***	**-8.6±9.5***

* Mejores valores de energía libre de interacción en las SDM según el objetivo evaluado; para sistemas *E6+lig* se señalan los $\Delta\overline{G}$ más bajos. Para *[E6+lig]+hx*, los más altos.

cercana a la mostrada en el cristal 4XR8 (a 9.8 Å de distancia entre los centros de masa de E6 y la hélice), aunque evitando el solapamiento de los átomos entre ésta y los del ligando acoplado a E6. Con ello, la posición inicial de la hélice se estableció a 12.3 Å de distancia al centro de masa de E6, siendo la misma para cada sistema de cada ligando (*E6+lig+hx*). Así, el ligando quedó en medio del *pocket* de E6 y de la hélice de E6AP, permitiendo evaluar el efecto de la molécula pequeña sobre la interacción de E6 y el motivo *LxxLL*. Además se realizó una simulación extra con la hélice en esta misma posición, pero sin un ligando de por medio (sistema *E6+hx*); sirviendo como control negativo.

En ambas fases, todas las SDM tuvieron una etapa de producción de 50 ns y se calcularon los valores de $\Delta G^{GB/PB}$ a lo largo de toda la trayectoria. Además, según el caso, se obtuvieron los valores de RMSD de la proteína E6, del ligando y de la hélice, tomando la conformación del acoplamiento *E6-lig* como estructura inicial (Fig 21). También se obtuvo la distancia entre los centros de masa de la proteína E6 y la hélice o el ligando. Los valores promedio de la energía libre de interacción en los sistemas de ambos análisis se sintetizan en la Tabla 4, y las figuras A.9 a la A.13 detallan, respectivamente, los resultados obtenidos para las SDM de los complejos proteína-ligando con las moléculas luteolina, lig. 5, lig.32, lig.70 y lig. 82. Por último, las figuras 23 y 24 condensan los resultados de las SDM en los complejos *E6+lig+hx* de cada ligando, lo que permite visualizar el comportamiento de E6 y de la hélice ante la presencia de cada molécula.

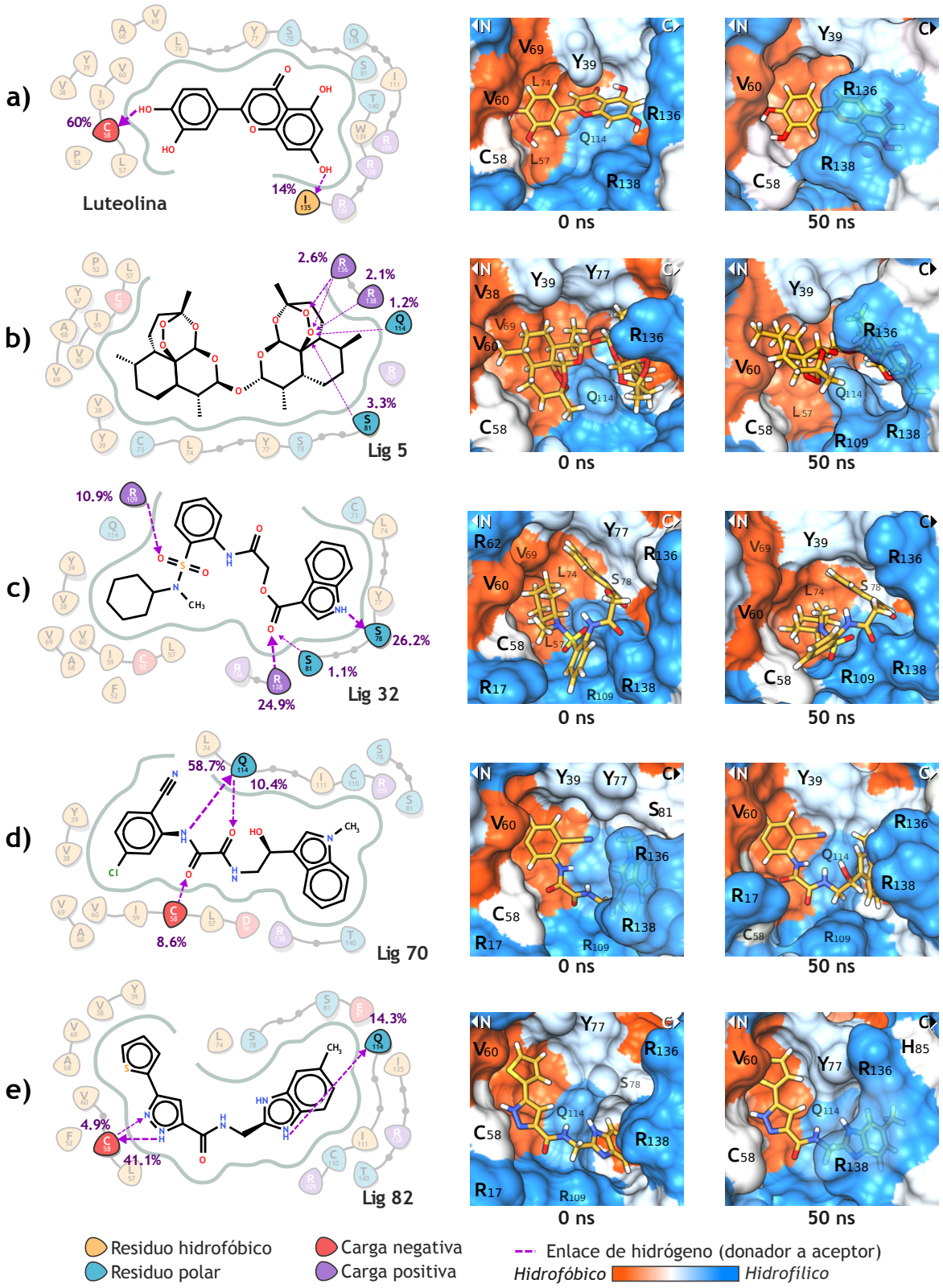


Figura 22: Acoplamiento y dinámica molecular de los sistemas *E6-lig*. En cada inciso se muestra un esquema de la interacción del ligando con los residuos del pocket de E6, haciendo hincapié en los enlaces de hidrógeno proteína-ligando establecidos durante la dinámica y mostrando el porcentaje de ocupancia del tiempo de interacción de dicho enlace. También se muestran esquemas de los acoplamientos de cada molécula con la superficie del *pocket* de E6 correspondientes al los tiempos 0 ns y 50 ns. **a)** Luteolina**. b)** *Lig5*. **c)** *Lig32*. **d)** *Lig70*. **e)** *Lig82*.

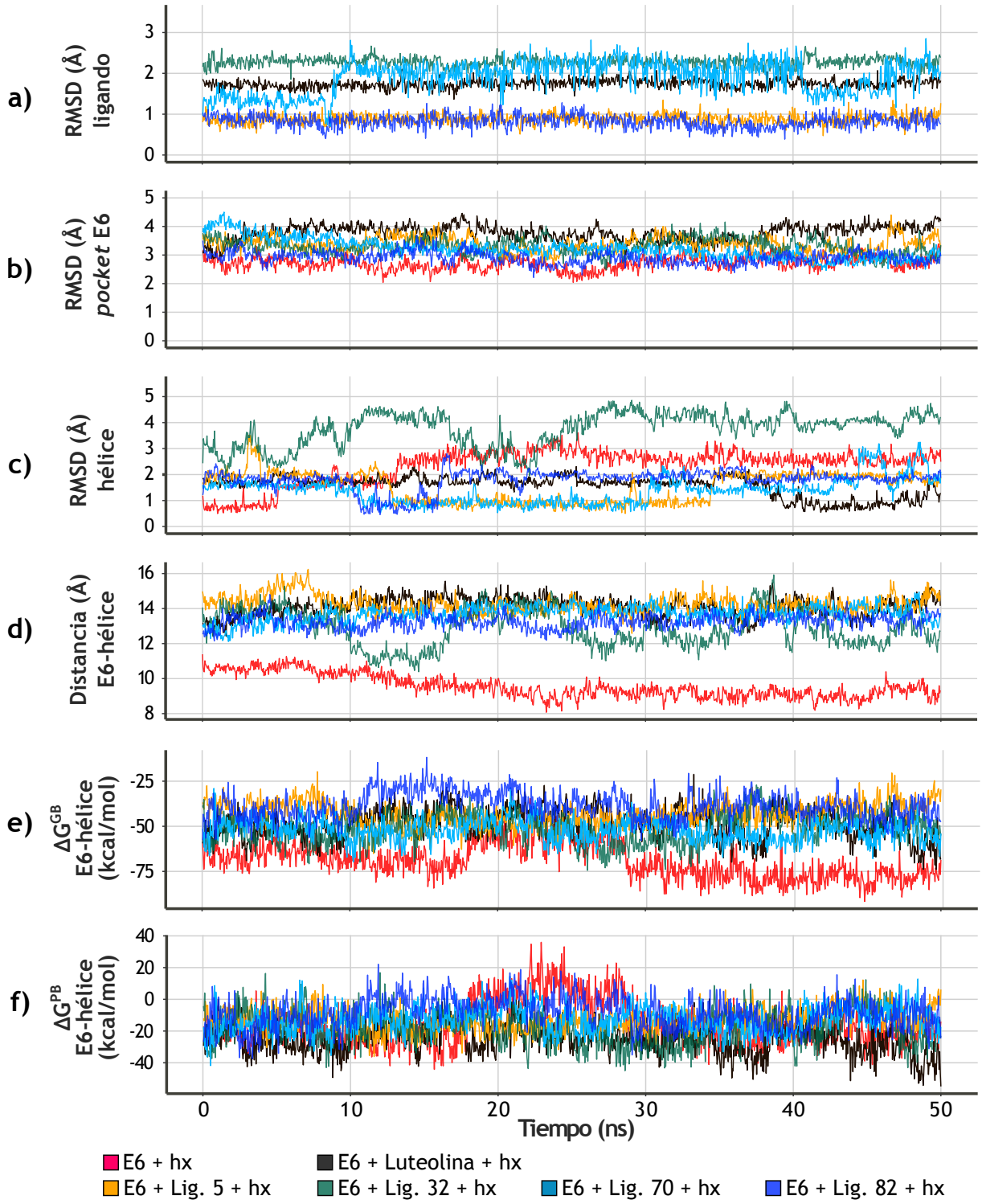


Figura 23: Dinámica molecular de los complejos *proteína-hélice-ligando* (50ns). a) Valores de RMSD de cada uno de los cinco ligandos en comparación con su posición de anclaje inicial. **b)** Valores de RMSD del *pocket* de la proteína E6 mostrados en cada ensayo con cada ligando y sin ligando (rojo). **c)** Valores de RMSD de la hélice *LxxLL* mostrados en cada ensayo con cada ligando y sin ligando (rojo). **d)** Distancia entre el centro de masa de E6 y el centro de masa de la hélice *LxxLL*. **e)** Valores de ΔG^{GB} de la interacción *[proteína-ligando]-hélice*. **f)** Valores de ΔG^{PB} de la interacción *[proteína-ligando]-hélice*.

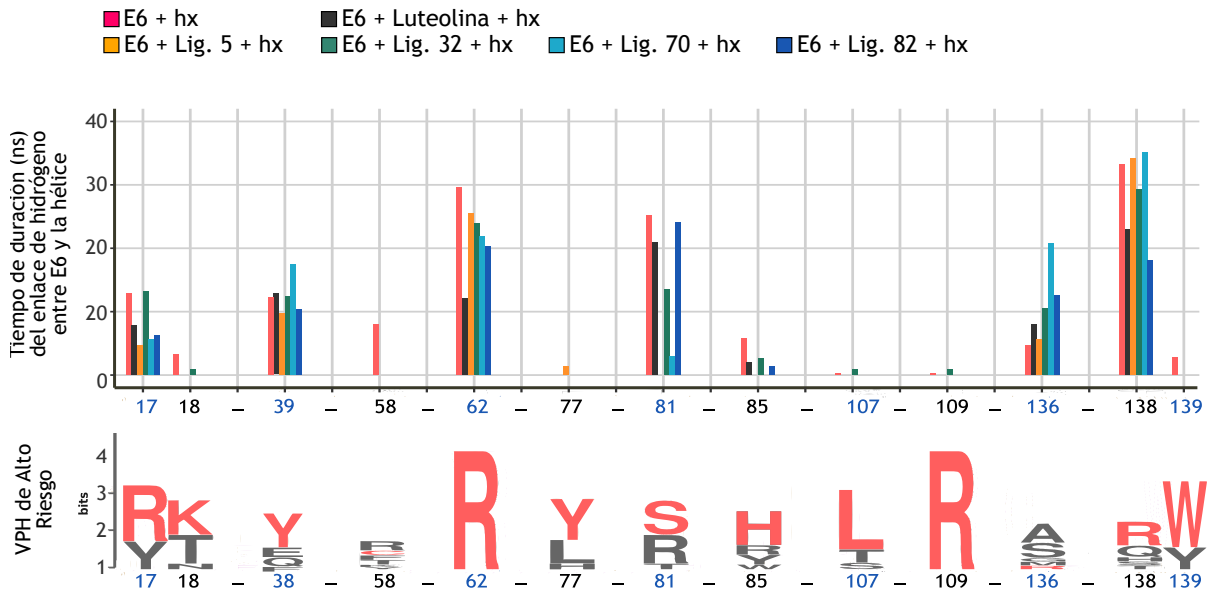


Figura 24: Enlaces de hidrógeno entre el *pocket* de la proteína E6 con un ligando acoplado y la hélice *LxxLL*. Se muestran los residuos de E6 VPH-16 (en rojo) utilizando parte de la información del logo de secuencia de la Fig. 7. Cada barra representa el tiempo de interacción por enlace de hidrógeno de cada residuo del *pocket* en cada ensayo de los complejos *[E6+lig+hx]* y *E6+hx* (rojo). El objetivo fue identificar el efecto de la presencia de los ligandos sobre estas interacciones. Estos datos son comparables con los mostrados en la Fig. A.8.

Capítulo 4. Discusión

4.1. OBJETIVO 1: Análisis secuencia-estructura de la proteína E6

Actualmente, en la base de datos *Uniprot* existen más de 600 secuencias de aminoácidos correspondientes a la proteína E6 VPH-16. Sin embargo, en el presente trabajo se seleccionó como referencia a la secuencia P03126, que equivale al consenso del resto de las secuencias de E6 VPH-16 depositadas en esta base de datos. P03126 fue comparada con otras secuencias de E6 pertenecientes a otros tipos de VPH de alto riesgo (AR) y bajo riego (BR) (Bruni *et al.*, 2014). Los tipos AR elegidos están clasificados como carcinogénicos por la Agencia Internacional de Investigación sobre el Cáncer. Por otro lado, los tipos BR pertenecen a los grupos Alfa 8 y 10, caracterizados por su tropismo hacia células de la mucosa (Doorbar *et al.*, 2012).

El resultado de este análisis indicó que, a pesar de no pertenecer a un grupo monofilético, las secuencias completas de E6 de los VPH de AR son más similares entre sí al compararlas con los tipos de BR (Figura 3). Específicamente, al medir la disimilitud de las subsecuencias de los bolsillos de unión a E6AP, se observó que los *pockets* de los VPH de BR son los más disimiles con respecto al *pocket* de VPH-16 (ver Figura A.2a). Es decir, que independientemente al subgénero al que pertenezca cada tipo, las secuencias del *pocket* de los tipos AR poseen residuos conservados no compartidos con los tipos BR.

Posteriormente, en los ensayos de Simulación de Dinámica Molecular (SDM) de los complejos *E6 VPH-16+hx* (AR) y *E6 VPH-11+hx* (BR), se observó que la hélice *LxxLL* tuvo una mayor afinidad hacia la proteína E6 VPH-16 (ver Figura A.7), una menor distancia entre los centros de masa y valores más bajos de RMSD (ver figuras A.7b y A.7c). Por último, como se muestra en la Figura A.8, la proteína E6 VPH-16 mantuvo un mayor número de enlaces de hidrógeno (H) con la hélice que la proteína E6 VPH-11, particularmente en los residuos R17, Y39, C58, R62, S81 y R138, de los cuales cuatro no están presentes en la secuencia de E6 VPH-11, y todos, a excepción de C58, están muy o medianamente conservados entre las proteínas E6 de los tipos de AR (Figura 7). Por otro lado, Zanier *et al.* (2013) llevaron a cabo un experimento *in vitro* mutando los residuos L57, R109 y R138 de E6 VPH-16, indentificándolos como primordiales en la interacción con E6AP. Sin embargo, L57 y R109 están conservados tanto en los tipos de AR como en los de BR, y sólo el sitio 138 de VPH-11 posee una glutamina en lugar de la arginina de la E6 VPH-16.

Todo lo anterior parece implicar una mayor estabilidad en la interacción entre la hélice y la proteína E6 VPH-16, lo que podría extrapolarse a otros tipos de AR, sobre todo si comparten

una marcada similitud con la subsecuencia del *pocket* de E6 VPH-16. Y es que esta distinción a nivel de secuencia parece estar relacionada con las diferencias funcionales entre las E6 de AR y de BR reportadas por Doorbar *et al.* (2012), entre las que destaca la aparente capacidad de las E6 de BR de interaccionar con E6AP pero sin ubiquitinización de p53. Además, trabajos previos han sugerido que las proteínas E6 y E7 de AR son suficientes para inducir la inmortalización de queratinocitos humanos primarios (Hawley-Nelson *et al.*, 1989), y que las E6 de AR son las principales responsables de inducir la progresión tumoral (Song *et al.*, 2000).

Esto indica que las características intrínsecas de la secuencia-estructura de las E6 de AR tienen por sí solas un peso muy importante en la oncogenia del virus. Un modelo computacional que refuerza esta idea fue desarrollado por Wang *et al.* (2015), quienes mostraron que a partir de la información contenida en la secuencia de una proteína E6 de determinado tipo, es posible predecir si éste pertenece o no a un VPH de AR. Sin embargo, es necesario llevar a cabo más estudios, tanto *in silico* como *in vitro*, para validar mejor esta hipótesis.

4.2. OBJETIVO 2: Refinamiento de la estructura proteína E6

El primer paso del refinamiento de la estructura de la proteína E6 se llevó a cabo mediante modelado por homología de la secuencia P03126, usando el modelo 4XR8. Estudios *in silico* como los de Kumar *et al.* (2015a, 2014) han seguido esta metodología previo a una etapa de Simulación de Acoplamiento Molecular (SAM). Sin embargo, en el presente trabajo se propuso el empleo de Simulación de Dinámica Molecular (SDM) para enriquecer aun más el refinamiento de E6, y así obtener conformaciones que representen mejor su variabilidad estructural en condiciones fisiológicas.

Para lograr lo anterior, en la primer fase de SDM se aplicó el método *Simulated Annealing* (SA) (Sección 2.2.2.2), diseñado para favorecer la búsqueda en el espacio conformacional de una biomolécula al aumentar, y posteriormente disminuir gradualmente, la temperatura del sistema en simulación. Esto permitió sortear barreras energéticas elevadas que de otro modo ”atraparían” la dinámica conformacional de la proteína en mínimos energéticos locales, dificultando el muestreo de otros estados conformacionales que pudieran ser relevantes (Bernardi *et al.*, 2015). De este modo, se obtuvieron diferentes macroestados conformacionales de la proteína E6 (Figura 11a), y se observó una considerable flexibilidad en su estructura (Figura 13), sobre todo en la región N-terminal, algo reportado anteriormente en los estudios experimentales de Zanier *et al.* (2012) y Nominé *et al.* (2003). Por su parte, en los análisis *in silico* de Rietz *et al.* (2016) y Shah *et al.*

(2015), también se ha puesto de manifiesto la flexibilidad de esta proteína. Sin embargo, es necesario aclarar que estos autores utilizaron como estructura inicial el modelo cristalográfico de E6 4C/4S (4GIZ, Zanier *et al.* (2013)), y no un modelo homólogo de la secuencia completa de E6 VPH-16.

Por otro lado, los resultados del RMSD del *pocket* de la proteína *E6 sola* (Fig. 11b) no mostraron diferencias frente a los cambios conformacionales de toda la proteína (ver Figura 11a). Esta idea se ve reforzada con los resultados del ACP, particularmente al considerar las trayectorias correspondientes al subespacio del CP1 de cada ensayo, mostradas en la Fig. 14c. Estas representaciones estructurales recalcan la movilidad relativa de los dominios N-terminal y C-terminal de la proteína, lo que se traduce como la capacidad de apertura y cierre del bolsillo de unión a E6AP, pues el *pocket* está estructurado por residuos de ambos dominios (Figura 5).

Respecto a lo anterior, Rietz *et al.* (2016) observaron un estado ”abierto” de la proteína, haciendo hincapié en la flexibilidad del *pocket*; particularmente en el movimiento de la cadena lateral de los residuos de arginina (ver Figura 5 de Rietz *et al.* (2016)). En el presente trabajo se observó un comportamiento similar, de hecho, al observar las conformaciones de E6 en la Figura 16 es posible notar que gran parte de las diferencias en la configuración superficial de esta cavidad se debe a la posición de las cadenas laterales de arginina (R17, R62, R109, R136 y R138), además del residuo Y77. Por ello, se puede concluir que tanto la cadena principal como las cadenas laterales de algunos residuos contribuyen a la plasticidad de esta cavidad, la cual se vio representada en las seis conformaciones obtenidas y mostradas en la Figura 16a y en los estados de ”apertura” y ”cierre” del *pocket* descritos en la Figura 14c.

4.3. OBJETIVO 3: Cribado virtual basado en propiedades ADMETox

La farmacobilidad de E6 con moléculas pequeñas fue una idea propuesta inicialmente por Baleja *et al.* (2006), que ha sido reforzada por los estudios de Zanier *et al.* (2013, 2014) sobre la estructura de E6 VPH-16. Particularmente, el *pocket* de E6 ha sido descrito como una cavidad única y relativamente profunda capaz de interactuar con el motivo *LxxLL* de algunas proteínas celulares, incluida E6AP. En contraste, las proteínas celulares que reconocen al motivo *LxxLL* muestran dominios helicoidales poco profundos y distintos a la topología del *pocket* de E6. Por lo anterior, esta región ha sido señalada como un blanco farmacológico ideal, debido a su probable especificidad, en el desarrollo de fármacos anti-VPH (Zanier *et al.*, 2013). En lo que respecta a moléculas pequeñas, algunos estudios han mostrado buenos resultados experimentales en lí-

neas celulares con genomas de VPH integrados. E incluso compuestos como la Artemisinina, la Epigallocatequina-3.gallato (Epg-3-g), el Indol-3-carbinol o la Curcumina han alcanzado etapas clínicas. Sin embargo, hasta ahora ninguna ha reportado resultados posteriores favorables (Hampson *et al.*, 2015; Bell *et al.*, 2000a).

Tomando lo anterior como punto de partida, se llevó a cabo la selección de las 26 moléculas mostradas en la Tabla 3.3.1. Estas moléculas mostraron una marcada diversidad estructural (Figura 17) impactando en el número de moléculas obtenido tras la búsqueda por similitud en la base de datos *ZINC15*; un total 34,804 compuestos, que correponderían al conjunto inicial de moléculas a evaluar (conjunto de evaluación). A partir de este conjunto de evaluación se planteó la identificación de nuevas moléculas que además de compartir similitudes estructurales con las 26 iniciales, pudieran tener una mejor actividad anti-VPH y propiedades farmacológicas favorables.

Para logralo, se realizó la predicción *in silico* de las propiedades farmacocinéticas (ADMETox) de todo el conjunto de evaluación. La metodología empleada se basó en descriptores obtenidos de la relación cuantitativa estructura-actividad/propiedad de las moléculas (en inglés, *Quantitative structure-activity relationship*, QSAR, o, *quantitative structure-property relationship*, QSPR), con lo cual las propiedades fisicoquímicas y/o estructurales son vinculadas a una actividad en particular (Schrödinger, 2017b; Gola *et al.*, 2006). Para este trabajo, la selección de los descriptores utilizados (Tabla 1) se basó en la documentación de *Qikprop* (Schrödinger, 2017b), y corresponde a un esquema relajado en el que se permitió como máximo una violación a los parámetros de cada descriptor. Hay que resaltar que, al aplicar de forma jerárquica cada filtro, algunas combinaciones de propiedades tuvieron más peso que otras al momento de seleccionar las moléculas.

En el Apéndice A.5 se describe el resultado de cada filtro y su efecto sobre la reducción del tamaño del conjunto de evaluación de moléculas. Como resultado, se descartaron 15,685 moléculas del conjunto de evaluación inicial, y las otras 19,119 restantes pasaron a la fase de SAM. De ellas se puede concluir que todas cumplieron, como mínimo, con 21 de las 25 propiedades fisicoquímicas evaluadas, al igual que lo hicieron nueve de los 12 fármacos aprobados por la FDA. Hay que recordar también que, de acuerdo a la base de datos *ZINC15*, todas estas moléculas están comercialmente disponibles.

4.4. OBJETIVO 4: Cribado virtual basado en acoplamiento molecular

La Simulación de Acoplamiento Molecular (SAM) es una de las alternativas principales utilizadas en el descubrimiento de fármacos asistido por computadora (Cerqueira *et al.*, 2015), y es parte integral en los protocolos de la industria farmacéutica (Lape *et al.*, 2010). En el presente trabajo se usaron los programas *Autodock 4* (AD4) y *Autodock Vina*, ambos ampliamente utilizados y evaluados en investigaciones previas (Jaghoori *et al.*, 2016), *AD4*: (Kumar *et al.*, 2016, 2015a, 2014; Tambunan y Parikesit, 2012) y *Vina*: (Kumar *et al.*, 2015b; Mamgain *et al.*, 2015)).Ambos programas han sido desarrollados por el mismo equipo de trabajo, no obstante, poseen características distintas (ver Sección 2.4). Y los propios autores han señalado que, para un sistema receptor-ligando determinado, cualquiera de los dos programas puede dar mejores resultados que el otro (Trott y Olson, 2010b). Por este motivo, previo a la evaluación de los 19,119 ligandos obtenidos en el objetivo 3, se llevó a cabo un análisis preliminar con ambos programas, y los resultados se muestran en el Apéndice A.6.

4.4.1. Acoplamiento molecular con *Autodock Vina* y *Autodock 4*

En su investigación, Rietz *et al.* (2016) concluyen que la flexibilidad observada del *pocket* de E6 debe ser tomada en cuenta en los estudios de descubrimiento de fármacos basado en la estructura, y proponen que la apertura de esta cavidad puede ser la causal de la baja efectividad de algunos candidatos anti-VPH propuestos a través de métodologías *in silico* previas. Así que, para tomar esto en consideración, en la presente investigación se realizó un procedimiento conocido como Acoplamiento Molecular en Conglomerado (Korb *et al.*, 2012; Amaro *et al.*, 2008) utilizando *Vina* y las conformaciones obtenidas en la Sección 2.2.3.3.

Dada la disimilitud del *pocket* en las cuatro estructuras de las conformaciones de E6 usadas (ver figuras 16b y 18), era de esperarse que la distribución de las energías de interacción de los 19,119 ligandos no fuese la misma para cada conformación. Es decir, algunos ligandos que presentaron valores de ΔG muy favorables con un confórmero no necesariamente tuvieron este comportamiento con los otros. Sin embargo, en la Figura 19a fue posible observar una tendencia en el ΔG de los ligandos, en la que las interacciones menos favorables se ubican en la esquina superior derecha. Por otro lado, el método de selección de los mejores ligandos procuró identificar a las moléculas que mostraron las energías de interacción más bajas en los cuatro confórmeros.

Posteriormente, estos ligandos fueron reacoplados con *AD4*, pero utilizando solamente una

conformación de E6 debido a la demanda en tiempo computacional que exige *AD4*, sobre todo por los parámetros empleados (sección 2.4.3). La estructura de E6 elegida fue la de la réplica *e3a* (Figura 18d), pues el *cluster* conformacional que representa fue el más grande (Figura 15c). Al final los 100 mejores ligandos en este proceso mostraron valores entre -8.5 y -10.5 kcal/mol en el ΔG de unión a E6 (Figura 19b).

4.5. OBJETIVO 5: Validación con simulación de dinámica molecular

4.5.1. Reevaluación de los acoplamientos E6+lig con los métodos MM/GBSA y MM/PBSA

Comúnmente los ΔG de unión receptor-ligando calculados por programas de SAM difieren de los valores de interacción obtenidos experimentalmente. Esto se debe a que las funciones de *scoring* de los programas de SAM tienden a sacrificar precisión a cambio de rapidez para poder analizar un elevado conjunto de ligandos en un tiempo considerable. De hecho, se ha reportado que generalmente la SAM no permite discriminar entre moléculas con menos de 1.5 kcal/mol en su energía libre de interacción con el receptor (Genheden y Ryde, 2015). Por lo anterior, tras la SAM es recomendable una etapa de *rescoring* con métodos más robustos que permitan refinar los resultados (Zhang *et al.*, 2014). En el caso de esta investigación se utilizaron los métodos MM/GBSA (GB) y MM/PBSA (PB) para reevaluar los acoplamientos obtenidos con *AD4*.

Al comparar las figuras 19b y 20 puede observarse que no necesariamente las moléculas con los valores de ΔG más bajos en *AD4* tuvieron los valores más bajos de $\Delta G^{GB/PB}$. Esto puede deberse a lo mencionado anteriormente acerca de la baja precisión de la SAM si se compara con la de los métodos GB/PB. Sin embargo, incluso entre GB y PB no necesariamente existe una correlación entre los valores de cada ligando. Un ejemplo claro es la luteolina, cuyo ΔG^{PB} la coloca dentro de las cinco moléculas con interacciones más favorables hacia E6. Pero esto no ocurre si se considera el ΔG^{GB}, que no destaca frente a los obtenidos por las otras moléculas.

Análisis previos han puesto de manifiesto esta discrepancia entre GB y PB al intentar determinar cuál es más preciso. No obstante, no hay un claro consenso al respecto, pues el resultado de cada método depende del sistema evaluado (Genheden y Ryde, 2015). Para la proteína E6 no se han reportado valores de interacción experimentales con una molécula pequeña, lo que impide contar con una referencia para determinar qué método funciona mejor para este sistema. Por lo anterior, los mejores cuatro ligandos fueron elegidos tomando en cuenta ambos valores, ΔG^{GB} y ΔG^{PB} (Fig. 20).

Además, el número de enlaces de H fue también considerado, ya que si bien el *pocket* de E6 posee una región altamente hidrofóbica, también es cierto que la apertura de la cavidad y su superficie accesible al solvente son bastante amplias. Por lo que un acoplamiento basado principalmente en interacciones hidrofóbicas podría no ser suficiente, sobre todo si se considera la flexibilidad observada en esta región. Como resultado cuatro moléculas fueron elegidas (Fig. 21) para la última fase de este trabajo y sus resultados se discuten la siguiente sección.

Sin embargo, es necesario hacer hincapié en que el resto de los 100 ligandos obtenidos hasta esta fase, y sobre todo los 30 mostrados en la Figura A.6, no dejan de ser buenos candidatos para ser evaluados en experimentos *in vitro*. Ya que hay que recordar que un protocolo de cribado virtual molecular para el descubrimiento de fármacos tiene como objetivo evaluar un gran número de moléculas para descartar aquellas con propiedades no deseadas o desfavorables y conservar un conjunto de candidatos potencialmente útiles, que después deban ser validados experimentalmente (Cerqueira *et al.*, 2015). Es por ello que los resultados obtenidos hasta este punto cumplen con ese criterio y con el objetivo principal de esta tesis, ofreciendo un nuevo punto de partida para el desarrollo de fármacos anti-VPH.

4.5.2. SDM de los complejos *E6+lig* y *[E6+lig]+hx*

4.5.2.1. SDM del acoplamiento proteína E6+luteolina

La luteolina es un compuesto flavonoide de 286.2 Da. Fue identificada como inhibidor de la interacción E6-E6AP por Cherry *et al.* (2013) a través de un cribado *in silico* basado en farmacóforo (Baleja *et al.*, 2006). Posteriormente, se demostró su actividad anti-VPH en las líneas celulares Caski y HeLa, derivadas de carcinomas cervicales infectados por VPH-16 y VPH-18, respectivamente. Estos resultados mostraron un incremento de p53, y un decremento en la viabilidad de las células. Además, se corroboró la interacción directa entre la luteolina con la proteína E6 mediante resonancia de plasmón de superficie. Y finalmente, utilizando el programa de acoplamiento molecular *GLIDE*, Cherry *et al.* (2013) descubrieron que la luteolina era capaz de interactuar con E6 a través de su *pocket* de unión a E6AP.

En el presente trabajo, la luteolina mostró valores de interacción moderados con *Vina* y sobre todo con *AD4* en comparación con otros ligandos. Pese a esto, la posición de acoplamiento a E6 obtenida con *AD4* (Fig. 22a) coincidió con la publicada por Cherry *et al.* (2013) (ver Figura 11A de (Cherry *et al.*, 2013)), y su evaluación con el método PB indicó valores muy favorables de

ΔG de interacción (Figura 20). Durante la SDM, esta pose se mantuvo constante, con una baja fluctuación en sus valores de RMSD (Figura A.9a). Además, llegó a establecer hasta dos enlaces de H con los residuos de E6 (Figura 22a), uno de los cuales estuvo presente en más del 60 % del tiempo de duración de la dinámica. Con respecto a la energía libre de interacción con E6, los valores promedio obtenidos fueron los siguientes: ΔG^{GB} = -29.4 kcal/mol y ΔG^{PB} = -0.4 kcal/mol.

Por otro lado, en el sistema *[E6+luteolina]+hx*, la interacción de la hélice con E6 se vio afectada pues la presencia de la luteolina bloqueó la formación de enlaces de H de la hélice y los residuos C58 y R62 del *pocket* (Fig. 24). Así mismo, los valores ΔG calculados para este sistema fueron los siguientes: ΔG^{GB} = -48.7 kcal/mol y ΔG^{PB} = -25.11 kcal/mol. De acuerdo con esto, luteolina aumentó el valor de ΔG^{GB} en comparación con los del sistema *E6+hx* (Fig. 23 e y f (rojo)), cuyo ΔG^{GB} fue de -69.3 kcal/mol. Sin embargo, con el método PB ocurrió lo contrario pues el valor promedio de la interacción proteína-hélice en el sistema *E6+hx* fue de -15.5 kcal/mol, curiosamente, menos favorable que el del sistema *[E6-luteolina]+hx*.

4.5.2.2. SDM del acoplamiento proteína E6+ligando 5

El ligando 5 (*lig5*) es una molécula de 550.7 Da. Posee 2 rotámeros, 0 donadores y 9 aceptores de enlaces de H. Estructuralmente se le reconoce con un dímero de la Artemisinina (Fig. A.5), un fármaco comúnmente empleado contra la malaria (Galal *et al.*, 2009). Y al igual que ésta, *lig5* es una lactona sesquiterpénica en la que cada monómero posee un puente peróxido muy característico. De acuerdo a la base de datos *PubChem*, *lig5* ha sido previamente evaluado contra especies del género *Plasmodium*, pero también ha mostrado resultados favorables en evaluaciones de citotoxicidad en líneas celulares tumorales (Galal *et al.*, 2009), incluyendo células HeLa (Beekman *et al.*, 1998).

En el presente trabajo esta molécula mostró buenos resultados de acoplamiento con E6 durante la SDM, comparándola con los obtenidos con luteolina (Tabla 4). Durante los 20 primeros ns de la producción de la SDM, el valor de ΔG^{GB} de *E6+lig5* alcanzó las -60 kcal/mol (Fig. A.10c), los valores más bajos que llegó a obtener un ligando y que curiosamente fueron cercanos al ΔG^{GB} del sistema *E6 VPH-16+hx* (Fig. A.7e, rojo) y más favorables que los del $\Delta G^{GB/PB}$ de *E6 VPH-11+hx* (Fig. A.7e y A.7f, azul). Sin embargo, cabe aclarar desde ahora que esta comparación entre los $\Delta G^{GB/PB}$ de los sistemas *E6+lig* y de los sistemas *E6+hx* debe ser tomada con reserva, pues este valor de interacción calculado es sensible al tamaño del sistema *receptor-ligando*, y en este caso la hélice *LxxLL* su peso molecular, de 1,379 Da, es más del doble que el de cualquier ligando

evaluado.

Por otro lado, durante SDM, uno de los monómeros de Artemisinina de *lig5* interaccionó con la región hidrofílica del *pocket*, mientras que el otro lo hizo con la región hidrofóbica. Por su parte, la formación de enlaces de H con E6 fue bastante moderada (Fig. 22b). *lig5* mostró una conformación rígida los primeros 30 ns, coincidiendo con los valores más bajos de $\Delta G^{GB/PB}$ mencionados anteriormente. Sin embargo, a partir del ns 30 el monómero de la región hidrofóbica se desplazó en dirección opuesta a la superficie del *pocket*, aumentando la distancia entre los centros de masa *E6-lig5* e incrementando el ángulo de rotación de los enlaces del grupo acetal de la molécula, haciéndola más flexible (Fig. A.10a). Esto repercutió en la energía de interacción final, que aun así, fue prácticamente menor en todo momento que la de la luteolina en el caso del ΔG^{PB}, pero curiosamente, mayor para el ΔG^{PB} (Tabla 4). Sin embargo, dado que el monómero de la región hidrofílica mostró poca variación en su posición, es probable que *lig5* tenga la posibilidad de volver a la conformación de acoplamiento más favorable.

Por último, en los resultados de la SDM del sistema *[E6+lig5]+hx*, *lig5* mostró mayor rigidez, y su presencia en el *pocket* aumentó la distancia entre el centro de masa de E6 y la hélice del valor inicial 12 Å a hasta cerca de 15 Å (Fig. 23, amarillo), 5 Å más que del valor final obtenido en el sistema *E6+hx* (Fig. 23, rojo). También, *lig5* evitó el establecimiento de enlaces de H entre la hélice y los residuos C58, S81 y H85 del pocket de E6 (Fig. 23, amarillo). Finalmente, en los resultados de la energía de interacción *[E6+lig5]+hx*, la afinidad media de la hélice hacia E6 mostró un mejor efecto inhibitorio de la afinidad *E6+hx* que el del sistema *[E6-luteolina]+hx*.

4.5.2.3. SDM del acoplamiento proteína E6+ligando 32

El ligando 32 (*lig32*) tiene un peso molecular de 469.5 Da, y posee 8 rotámeros, 2 donadores y 9 aceptores de enlaces de H. Estructuralmente se caracteriza por la presencia de un anillo ciclohexilo, un grupo sulfonilo y un grupo indólico. Fue obtenida en la sección 3.3.1 por poseer como subestructura al indinol (Fig. A.5). De acuerdo a las bases de datos *ZINC15* y *PubChem*, no hay una actividad conocida por este compuesto.

En la SDM del sistema *E6+lig32*, *lig32* mostró poca variación en sus valores de RMSD, manteniendo la mayor parte del tiempo el contacto del anillo ciclohexilo con la superficie hidrofóbica del *pocket*, y el grupo indólico interaccionando con la región hidrofílica (Fig. 22b). Por su parte, la región intermedia del ligando, con el anillo aromático, fue la más fluctuante al quedar expuesta al

solvente (Fig. 21b). En cuanto a su interacción con E6 (Tabla 4), el valor promedio obtenido de ΔG^{GB} fue 4.5 kcal/mol menor al de la luteolina, pero 1.8 kcal/mol que el ΔG^{PB} de ésta.

En los resultados de la dinámica del sistema *[E6+lig32]+hx* (Fig. 23, verde), la conformación de *lig32* fue más uniforme. En cambio, la hélice *LxxLL* mostró los valores de fluctuación más altos y variaciones drásticas en la distancia a E6. Esto se debió a la flexibilidad en región intermedia del ligando, que en determinado momento permitió un mayor acercamiento de la hélice hacia E6 (Fig. 23d, verde). Por su parte, *lig32* mermó la formación de enlaces de H entre la hélice y los residuos C58 y S81 del *pocket* (Fig. 24). Por último, en los resultados de la energía libre de interacción *[E6+lig32]+hx*, la presencia de *lig32* no mostró buenos resultados en la reducción de la afinidad media de la hélice hacia E6, tanto para el ΔG^{GB} y el ΔG^{PB} calculados. Aunque en este último fue +7.4 kcal/mol mayor que el valor de (ΔG^{PB}) del sistema *[E6-luteolina]+hx* (Tabla 4).

4.5.2.4. SDM del acoplamiento proteína E6+ligando 70

El ligando 70 (*lig70*) tiene un peso molecular es de 469.5 Da, y posee 6 rotámeros, 3 donadores y 7 aceptores de enlaces de H. Estructuralmente se caracteriza por la presencia de un átomo de cloro unido a un anillo cianofenil y un grupo indólico en el extremo opuesto. Fue obtenido en la sección 3.3.1 a partir de su similitud con el compuesto Myakamida (Fig. A.5). De acuerdo a las bases de datos *ZINC15* y *PubChem*, no hay una actividad conocida por este compuesto.

En la SDM del sistema *E6+lig70*, *lig70* conservó la posición de acoplamiento con E6 con el anillo cianofenil en zona hidrofóbica del pocket de E6 (Fig. 22f), y estableciendo enlaces de H con los residuos C58 (cadena principal) y Q114 (cadena lateral). Este último por un lapso mayor al 69 % del tiempo total de la trayectoria (Fig. 22f). Por su parte, en los resultados de interacción con E6, mostró un buen valor de ΔG^{GB} en comparación con el sistema *E6+luteolina*; 8.5 kcal/mol menos.

Por su parte, en el sistema *[E6+lig70]+hx* (Fig. 23, cían), *lig70* fue el único ligando que tuvo un cambio conformacional ante la presencia de la hélice (figuras 23a y 23c, cían, ns 10 al 30). La inspección visual de la trayectoria permitió observar que este cambio en la conformación de *lig70* ocurrió por el reacomodo del grupo intermedio etanodiamida, pero que los extremos de la molécula se mantuvieron estables en todo momento. Finalmente, el ΔG^{GB} promedio de este sistema fue de -52.0 kcal/mol, 17.3 kcal/mol más que el ΔG^{GB} del sistema *E6+hx*, por lo que para el caso del método GB, *lig70* mostró el peor efecto inhibitorio de la afinidad de la hélice hacia E6 (Tabla 4).

4.5.2.5. SDM del acoplamiento proteína E6+ligando 82

El ligando 82 (*lig82*) tiene un peso molecular es de 396.834 Da, posee 4 rotámeros, 2 donadores y 4 aceptores de enlaces de H. Estructuralmente posee un grupo metil-benzimidazol en uno de sus extremos, un tiofeno en el extremo opuesto y un anillo pirazol intermedio. Fue obtenido en la sección 3.3.1 a partir de su similitud con el compuesto C1_a (Fig. A.5). De acuerdo a las bases de datos *ZINC15* y *PubChem*, no hay una actividad conocida por este compuesto.

En la SDM del sistema *E6+lig82*, *lig82* mantuvo la pose de acoplamiento en todo momento (Fig. 22e), con el grupo tiofeno en la región hidrofóbica del *pocket* y el grupo metil-benzimidazol en la región hidrofílica. Cabe mencionar que *lig32* y *lig70* también mostraron una subestructura indano en esta región de E6. Por otro lado, la distancia del centro de masa de la molécula al de la proteína E6 fue la menor registrada. En cuanto al ΔG *E6+lig82*, los valores promedio de ΔG^{GB} y ΔG^{PB} fueron ligeramente mejores que los del sistema *E6+luteolina* (Tabla 4).

Finalmente, en la SDM de *[E6+lig82]-hx* (Fig. 23, azul), *lig82* mantuvo una única conformación muy cercana a la inicial (figuras 23a y 22e) mientras que su efecto sobre la distancia de la hélice a E6 y sus cambios conformacionales fue similar al observado con la luteolina, *lig5* y *lig70*, afectando la formación de enlaces de H entre la hélice y los residuos C58, S81 y 138 del *pocket* (Fig. 24). Por último, en los resultados de la energía de interacción *[E6+lig82]+hx*, la presencia de *lig82* mostró los mejores efectos contra la interacción *E6+hx*. Ya que los valores de afinidad calculados, ΔG^{GB} = -38.5 kcal/mol y ΔG^{PB} = -8.5 kcal/mol, mostraron una diferencia de +30.7 kcal/mol (ΔG^{GB}) y +7 kcal/mol (ΔG^{PB}) en comparación con el sistema *E6+hx*. También se observó una diferencia de +10.2 kcal/mol (ΔG^{GB}) y de +16.6 kcal/mol (ΔG^{PB}) frente al sistema *[E6-luteolina]+hx*.

Capítulo 5. Conclusión

El *pocket* de unión al motivo *LxxLL* de la proteína E6 ha sido señalado como un blanco farmacológico ideal para el desarrollo de medicamentos contra la infección y los efectos oncogénicos de los VPH de alto riesgo (Zanier *et al.*, 2014; Cherry *et al.*, 2013). No obstante, el descubrimiento y desarrollo de fármacos es un proceso sumamente demandante de tiempo y de recursos económicos. Por lo que en las últimas décadas el empleo de métodos *in silico* ha emergido como un complemento para simplificar y facilitar este proceso, sobre todo en sus etapas inicíales (Cerqueira *et al.*, 2015).

En el presente trabajo se llevó a cabo un Cribado Virtual (CV) de moléculas pequeñas combinando métodos basados en el análisis de la proteína E6 y en la estructura y predicción de las propiedades fisicoquímicas de los compuestos evaluados. Como primer paso, se analizaron las secuencias de las proteínas E6 de los tipos de AR de mayor incidencia y se obtuvo un modelo homólogo de la proteína E6 VPH-16. Después, la Simulación de Dinámica Molecular (SDM) de E6 en condiciones fisiológicas, y el empleo del método *Simulated Annealing*, permitieron estimar la variación conformacional de esta proteína, particularmente en el *pocket* de unión a E6AP. Esto brindó una mejor perspectiva acerca de la flexibilidad de esta región, algo que no había sido tomado en cuenta en trabajos previos (Kumar *et al.*, 2016; John y Kotadiya, 2015; Cherry *et al.*, 2013).

Por otro lado, 34,804 moléculas fueron obtenidas de la base de datos *ZINC15* por su similitud estructural con 26 moléculas que en trabajos anteriores han mostrado, en experimentos *in vitro* y/o *in silico*, algún grado de actividad anti-VPH. Después, algunas propiedades fisicoquímicas de las 34,804 moléculas fueron predichas a partir de su estructura tridimensional. Y tras una serie de filtros se obtuvieron 19,119 moléculas con características farmacocinéticas deseables (ATMETox).

La Simulación de Acoplamiento Molecular (SAM) permitió identificar los compuestos con mayor afinidad hacia E6 y brindó más evidencia a la hipótesis de la farmacobilidad del *pocket* con moléculas pequeñas (Baleja *et al.*, 2006), al mostrar que los acoplamientos *E6-lig* más favorables ocurrieron en esta cavidad. Aunado a esto, se observó que la variabilidad en la estructura del *pocket* parece haber afectado considerablemente la afinidad de algunas moléculas hacia E6. Por lo tanto, resultó necesario tomar en cuenta la capacidad de apertura y cierre de esta región, enfocando la búsqueda en ligandos capaces de adaptarse a los cambios conformacionales del *pocket*.

Finalmente, las cuatro moléculas identificadas y validadas por SDM corresponden a las de mejor potencial anti-VPH de las 34,804 moléculas evaluadas. Pero esto únicamente bajo los criterios del diseño experimental utilizado, y de las herramientas y aproximaciones usadas. Por lo tanto, el siguiente paso de este trabajo es la validación de estos compuestos en ensayos *in vitro* e *in vivo*. Pues sólo así se podrá corroborar su posible efecto terapéutico contra los VPHs de AR. Sin embargo, es necesario mencionar que el resto de los compuestos listados en la Figura A.6 puede ser utilizado como punto de partida para futuros trabajos computacionales, o incluso para su evaluación inmediata en el laboratorio.

En cuanto las perspectivas para el análisis *in silico*, la presente investigación establece una metodología capaz de adaptarse al estudio de otros blancos terapéuticos. Sin embargo, aún le queda mucho por refinar y enriquecer. Principalmente, es importante extender el análisis a un mayor número de moléculas, e inclusive a nuevas bases de datos de compuestos químicos. En cuanto las fases de SAM y SDM, aumentar el número de ensayos y evaluaciones, o refinar el método de muestreo conformacional de la proteína, debería favorecer considerablemente la precisión de los resultados. Además, la incorporación de procedimientos más robustos para la evaluación de la energía libre de interacción; como el método de Perturbación de la Energía Libre (FEP) o el método de Integración Termodinámica (TI), podrían facilitar la selección final entre los mejores candidatos.

Sin embargo, no debe perderse de vista que el CV tiene como objetivo reducir el espacio químico inicial a un número de moléculas ”manejable”, minimizando en lo posible la inversión en tiempo y dinero dedicada a los análisis en laboratorio. Es por esto, que para mantenerse fiel a esta idea, algunas de las contribuciones y correcciones que se le puedan ofrecer a la metodología presentada en este trabajo dependerán en gran medida del tiempo y de los recursos computacionales con los que se disponga.

Literatura Citada

Allison, D. B. y Maleki, Z. (2015). HPV-related Head and Neck Squamous Cell Carcinoma: An Update and Review Article. *Journal of the American Society of Cytopathology*, **5**(4): 1–13.

Amaro, R. E., Baron, R., y McCammon, J. A. (2008). An improved relaxed complex scheme for receptor flexibility in computer-aided drug design. *Journal of Computer-Aided Molecular Design*, **22**(9): 693–705.

Bajusz, D., Rácz, A., y Héberger, K. (2015). Why is Tanimoto index an appropriate choice for fingerprint-based similarity calculations? *Journal of Cheminformatics*, **7**(1): 1–13.

Baleja, J. D., Cherry, J. J., Liu, Z., Gao, H., Nicklaus, M. C., Voigt, J. H., Chen, J. J., y Androphy, E. J. (2006). Identification of inhibitors to papillomavirus type 16 E6 protein based on three-dimensional structures of interacting proteins. *Antiviral Research*, **72**(1): 49–59.

Basu, P. y et al (2013). Clearance of Cervical Human Papillomavirus Infection by Topical Application of Curcumin and Curcumin. *Asian Pacific Journal of Cancer Prevention*, **14**(10): 5753–5759.

Be, X., Hong, Y., Wei, J., Androphy, E. J., Chen, J. J., y Baleja, J. D. (2001). Solution structure determination and mutational analysis of the papillomavirus E6 interacting peptide of E6AP. *Biochemistry*, **40**(5): 1293–1299.

Beaudenon, S. y Huibregtse, J. M. (2008). HPV E6, E6AP and cervical cancer. *BMC biochemistry*, **9 Suppl 1**(Suppl 1): S4.

Beekman, A., Wierenga, P., Woerdenbag, H., Uden, W., Pras, N., Konings, A., El-Feraly, F., Galal, A., y Wikström, H. (1998). Artemisinin-Derived Sesquiterpene Lactones as Potential Antitumour Compounds: Cytotoxic Action against Bone Marrow and Tumour Cells. *Planta Medica*, **64**(07): 615–619.

Beitz, E. (2000). TeXshade: shading and labeling of multiple sequence alignments using LaTeX2e. *Bioinformatics*, **16**(2): 135–139.

Bell, M. C., Crowley-Nowick, P., Bradlow, H., Sepkovic, D. W., Schmidt-Grimminger, D., Howell, P., Mayeaux, E., Tucker, A., Turbat-Herrera, E. A., y Mathis, J. (2000a). Placebo-Controlled Trial of Indole-3-Carbinol in the Treatment of CIN. *Gynecologic Oncology*, **78**(2): 123–129.

Bell, M. C., Crowley-Nowick, P., Bradlow, H. L., Sepkovic, D. W., Schmidt-Grimminger, D., Howell, P., Mayeaux, E. J., Tucker, a., Turbat-Herrera, E. a., y Mathis, J. M. (2000b). Placebo-controlled trial of indole-3-carbinol in the treatment of CIN. *Gynecologic oncology*, **78**(2): 123–9.

Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N., y Bourne, P. E. (2000). The protein data bank. *Nucleic acids research*, **28**(1): 235–242.

Bernard, H. U., Burk, R. D., Chen, Z., van Doorslaer, K., Hausen, H. z., y de Villiers, E. M. (2010). Classification of papillomaviruses (PVs) based on 189 PV types and proposal of taxonomic amendments. *Virology*, **401**(1): 70–79.

Bernardi, R. C., Melo, M. C. R., y Schulten, K. (2015). Enhanced sampling techniques in molecular dynamics simulations of biological systems. *Biochimica et biophysica acta*, **1850**(5): 872–7.

Bodily, J. y Laimins, L. A. (2011). Persistence of human papillomavirus infection: Keys to malignant progression.

Bruni, L., Barrionuevo-Rosas, L., Albero, Aldea, Serrano, Valencia, Brotons, Mena, Cosano, Muñoz, Bosch, De-Sanjosé, y Castellsagué (2014). Human Papillomavirus and Related Diseases Report. *HPV Centre*, (Afghanistan): iii.

Bruni, L., Barrionuevo-Rosas, L., Albero, Aldea, Serrano, Valencia, Brotons, Mena, Cosano, Muñoz, Bosch, De-Sanjosé, y Castellsagué (2016). Human Papillomavirus and Related Diseases Report: México. *HPV Centre*, (Afghanistan): iii.

Buck, C. B., Thompson, C. D., Roberts, J. N., Müller, M., Lowy, D. R., y Schiller, J. T. (2006). Carrageenan is a potent inhibitor of papillomavirus infection. *PLoS Pathogens*, **2**(7): 0671–0680.

Burk, R. D., Harari, A., y Chen, Z. (2013). Human papillomavirus genome variants. *Virology*, **445**(1-2): 232–243.

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, S., Li, P., Lin, C., Luchko, T., Luo, R., Madej, B., Mermelstein, D., Merz, K., Monard, G., Nguyen, H., Nguyen, H., 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.

Cerqueira, N. M. F. S. A., Gesto, D., Oliveira, E. F., Santos-Martins, D., Br??s, N. F., Sousa, S. F., Fernandes, P. A., y Ramos, M. J. (2015). Receptor-based virtual screening protocol for drug discovery. *Archives of Biochemistry and Biophysics*, **582**(June): 56–67.

Chang, M. W., Ayeni, C., Breuer, S., y Torbett, B. E. (2010). Virtual screening for HIV protease inhibitors: A comparison of AutoDock 4 and Vina. *PLoS ONE*, **5**(8).

Chaturvedi, A. K. (2010). Beyond Cervical Cancer: Burden of Other HPV-Related Cancers Among Men and Women.

Chen, J. J., Hong, Y., Rustamzadeh, E., Baleja, J. D., y Androphy, E. J. (1998). Identification of an Helical Motif Sufficient for Association with Papillomavirus E6. *Journal of Biological Chemistry*, **273**(22): 13537–13544.

Cherry, J. J., Rietz, A., Malinkevich, A., Liu, Y., Xie, M., Bartolowits, M., Davisson, V. J., Baleja, J. D., y Androphy, E. J. (2013). Structure based identification and characterization of flavonoids that disrupt human papillomavirus-16 E6 function. *PLoS ONE*, **8**(12).

Chouhy, D., Bolatti, E. M., Pérez, G. R., y Giri, A. A. (2013). Analysis of the genetic diversity and phylogenetic relationships of putative human papillomavirus types. *The Journal of general virology*, **94**(Pt 11): 2480–8.

Clark, D. E. (2008). What has virtual screening ever done for drug discovery? *Expert Opinion on Drug Discovery*, **3**(8): 841–851.

Crooks, G. E., Hon, G., Chandonia, J. M., y Brenner, S. E. (2004). WebLogo: A sequence logo generator. *Genome Research*, **14**(6): 1188–1190.

Cuanalo, Luz. Merino, S. S. T. (2011). ¿Qué sabe Ud. acerca de... Cribado Virtual? *Revista mexicana de ciencias farmacéuticas*, **42**(2): 54–55.

Curreli, F., Zhang, H., Zhang, X., Pyatkin, I., Victor, Z., Altieri, A., y Debnath, A. K. (2011). Virtual screening based identification of novel small-molecule inhibitors targeted to the HIV-1 capsid. *Bioorganic & Medicinal Chemistry*, **19**(1): 77–90.

DiMasi, J. A., Grabowski, H. G., y Hansen, R. W. (2016). Innovation in the pharmaceutical industry: New estimates of R&D costs. *Journal of Health Economics*, **47**: 20–33.

Divya, C. S. y Pillai, M. R. (2006). Antitumor action of curcumin in human papillomavirus associated cells involves downregulation of viral oncogenes, prevention of NFkB and AP-1 translocation, and modulation of apoptosis. *Molecular Carcinogenesis*, **45**(5): 320–332.

Doorbar, J., Quint, W., Banks, L., Bravo, I. G., Stoler, M., Broker, T. R., y Stanley, M. A. (2012). The biology and life-cycle of human papillomaviruses. *Vaccine*, **30**(SUPPL.5): F55–F70.

Drie, J. H. (2007). Computer-aided drug design: The next 20 years. *Journal of Computer-Aided Molecular Design*, **21**(10-11): 591–601.

DrugBank (2015). DrugBank v4.2.

Dundas, J., Ouyang, Z., Tseng, J., Binkowski, A., Turpaz, Y., y Liang, J. (2006). CASTp: Computed atlas of surface topography of proteins with structural and topographical mapping of functionally annotated residues. *Nucleic Acids Research*, **34**(WEB. SERV. ISS.).

Flint, J., Rall, G. F., Racaniello, V. R., y Skalka, A. M. (2015). *Principles of Virology, Bundle*, Vol. 70. pp. 156–7.

Flores-Miramontes, M. G., Torres-Reyes, L. A., Aguilar-Lemarroy, A., Vallejo-Ruíz, V., Piña-Sánchez, P., Cortés-Gutiérrez, E., Reyes-Leyva, J., y Jave-Suárez, L. F. (2015). Prevalencia de genotipos de VPH en México y en el mundo detectados mediante Linear Array. *HPV genotypes prevalence in México and worldwide detected by Linear Array.*, **53**(33): S122–S130.

Galal, A. M., Gul, W., Slade, D., Ross, S. A., Feng, S., Hollingshead, M. G., Alley, M. C., Kaur, G., y ElSohly, M. A. (2009). Synthesis and evaluation of dihydroartemisinin and dihydroartemisitene acetal dimers showing anticancer and antiprotozoal activity. *Bioorganic & Medicinal Chemistry*, **17**(2): 741–751.

García-Zepeda, S. P., García-Villa, E., Díaz-Chávez, J., Hernández-Pando, R., y Gariglio, P. (2013). Resveratrol induces cell death in cervical cancer cells through apoptosis and autophagy. *European journal of cancer prevention : the official journal of the European Cancer Prevention Organisation (ECP)*, **22**(6): 577–84.

Gascuel, O. (1997). BIONJ: an improved version of the NJ algorithm based on a simple model of sequence data. *Mol Biol Evol*, **14**(7): 685–95.

Ge, Z.-Q., Du, X.-Y., Huang, X.-N., y Qiao, B. (2015). Enhanced oral bioavailability of ursolic acid nanoparticles via antisolvent precipitation with TPGS1000 as a stabilizer. *Journal of Drug Delivery Science and Technology*, **29**: 210–217.

Genheden, S. y Ryde, U. (2015). The MM/PBSA and MM/GBSA methods to estimate ligand-binding affinities. *Expert opinion on drug discovery*, **10**(5): 449–61.

Gola, J., Obrezanova, O., Champness, E., y Segall, M. (2006). ADMET Property Prediction: The State of the Art and Current Challenges. *QSAR & Combinatorial Science*, **25**(12): 1172–1180.

Gouy, M., Guindon, S., y Gascuel, O. (2010). SeaView Version 4: A Multiplatform Graphical User Interface for Sequence Alignment and Phylogenetic Tree Building. *Molecular Biology and Evolution*, **27**(2): 221–224.

Grant, B. J., Rodrigues, A. P. C., ElSawy, K. M., McCammon, J. A., y Caves, L. S. D. (2006). Bio3d: An R package for the comparative analysis of protein structures. *Bioinformatics*, **22**(21): 2695–2696.

Ham, S., Kim, K. H., Kwon, T. H., Bak, Y., Lee, D. H., Song, Y. S., Park, S. H., Park, Y. S., Kim, M. S., Kang, J. W., Hong, J. T., y Yoon, D. Y. (2014). Luteolin induces intrinsic apoptosis via inhibition of E6/E7 oncogenes and activation of extrinsic and intrinsic signaling pathways in HPV-18-associated cells. *Oncology Reports*, **31**(6): 2683–2691.

Hampson, L., Martin-Hirsch, P., y Hampson, I. N. (2015). An overview of early investigational drugs for the treatment of human papilloma virus infection and associated dysplasia. *Expert Opin Investig Drugs*, **24**(12): 1529–1537.

Harder, E., Damm, W., Maple, J., Wu, C., Reboul, M., Xiang, J. Y., Wang, L., Lupyan, D., Dahlgren, M. K., Knight, J. L., Kaus, J. W., Cerutti, D. S., Krilov, G., Jorgensen, W. L., Abel, R., y Friesner, R. A. (2016). OPLS3: A Force Field Providing Broad Coverage of Drug-like Small Molecules and Proteins. *Journal of Chemical Theory and Computation*, **12**(1): 281–296.

Hawley-Nelson, P., Vousden, K. H., Hubbert, N. L., Lowy, D. R., y Schiller, J. T. (1989). HPV16 E6 and E7 proteins cooperate to immortalize human foreskin keratinocytes. *The EMBO journal*, **8**(12): 3905–10.

He, L., Zhang, E., Shi, J., Li, X., Zhou, K., Zhang, Q., Le, A. D., y Tang, X. (2013). (-)-Epigallocatechin-3-gallate inhibits human papillomavirus (HPV)-16 oncoprotein-induced angiogenesis in non-small cell lung cancer cells by targeting HIF-1alpha. *Cancer Chemother Pharmacol*, **71**(3): 713–725.

Howie, H. L., Katzenellenbogen, R. A., y Galloway, D. A. (2009). Papillomavirus E6 proteins. *Virology*, **384**(2): 324–334.

Huang, Q., Wu, L. J., Tashiro, S., Onodera, S., Li, L. H., y Ikejima, T. (2005). Silymarin augments human cervical cancer HeLa cell apoptosis via P38/JNK MAPK pathways in serum-free medium. *J Asian Nat Prod Res*, **7**(5): 701–709.

Humphrey, W., Dalke, A., y Schulten, K. (1996). VMD: Visual molecular dynamics. *Journal of Molecular Graphics*, **14**(1): 33–38.

Jaghoori, M. M., Bleijlevens, B., y Olabarriaga, S. D. (2016). 1001 Ways to run AutoDock Vina for virtual screening. *Journal of Computer-Aided Molecular Design*, **30**(3): 237–249.

Jansen, F. H., Adoubi, I., J, C. K., T, D. E. C., Jansen, N., Tschulakow, A., y Efferth, T. (2011). First study of oral Artenimol-R in advanced cervical cancer: clinical benefit, tolerability and tumor markers. *Anticancer research*, **31**(12): 4417–4422.

John, B. y Sali, A. (2003). Comparative protein structure modeling by iterative alignment, model building and model assessment. *Nucleic Acids Research*, **31**(14): 3982–3992.

John, G. y Kotadiya, R. (2015). In silico approach to identify putative drugs from natural products for human papillomavirus (HPV) which cause cervical cancer. (August 2016).

Johnson, D. K. y Karanicolas, J. (2016). Ultra-High-Throughput Structure-Based Virtual Screening for Small-Molecule Inhibitors of Protein–Protein Interactions. *Journal of Chemical Information and Modeling*, **56**(2): 399–411.

Jorgensen, W. L. (2004). The many roles of computation in drug discovery. *Science (New York, N.Y.)*, **303**(5665): 1813–8.

Kantabutra, V. (2015). Intentionally Linked Entities. En: *Emerging Trends in Computational Biology, Bioinformatics, and Systems Biology*. Elsevier, pp. 275–294.

Kapoor, V., Aggarwal, S., y Das, S. N. (2016). 6-Gingerol Mediates its Anti Tumor Activities in Human Oral and Cervical Cancer Cell Lines through Apoptosis and Cell Cycle Arrest. *Phytotherapy Research*, **30**(4): 588–595.

Kassi, E., Papoutsi, Z., Pratsinis, H., Aligiannis, N., Manoussakis, M., y Moutsatsou, P. (2007). Ursolic acid, a naturally occurring triterpenoid, demonstrates anticancer activity on human prostate cancer cells. *Journal of Cancer Research and Clinical Oncology*, **133**(7): 493–500.

Kirchmair, J., Rollinger, J. M., Liedl, K. R., Seidel, N., Krumbholz, A., y Schmidtke, M. (2011). Novel neuraminidase inhibitors: identification, biological evaluation and investigations of the binding mode. *Future Medicinal Chemistry*, **3**(4): 437–450.

Korb, O., Olsson, T. S. G., Bowden, S. J., Hall, R. J., Verdonk, M. L., Liebeschuetz, J. W., y Cole, J. C. (2012). Potential and Limitations of Ensemble Docking. *Journal of Chemical Information and Modeling*, **52**(5): 1262–1274.

Kumar, A. y Zhang, K. Y. J. (2015). Hierarchical virtual screening approaches in small molecule drug discovery. *Methods*, **71**(C): 26–37.

Kumar, S., Jena, L., Galande, S., Daf, S., Mohod, K., y Varma, A. K. (2014). Elucidating Molecular Interactions of Natural Inhibitors with HPV-16 E6 Oncoprotein through Docking Analysis. *Genomics & informatics*, **12**(2): 64–70.

Kumar, S., Jena, L., Mohod, K., Daf, S., y Varma, A. K. (2015a). Virtual Screening for Potential Inhibitors of High-Risk Human Papillomavirus 16 E6 Protein. *Interdisciplinary Sciences: Computational Life Sciences*, **7**(2): 136–142.

Kumar, S., Jena, L., Sahoo, M., Kakde, M., Daf, S., y Varma, A. K. (2015b). In Silico Docking to Explicate Interface between Plant-Originated Inhibitors and E6 Oncogenic Protein of Highly Threatening Human Papillomavirus 18. *Genomics & Informatics*, **13**(2): 60.

Kumar, S., Jena, L., Sahoo, M., Nayak, T., Mohod, K., Daf, S., y Varma, A. K. (2016). The in Silico Approach to Identifying a Unique Plant-Derived Inhibitor Against E6 and E7 Oncogenic Proteins of High-Risk Human Papillomavirus 16 and 18. *Avicenna Journal of Medical Biochemistry*, **In Press**(In Press).

Langsfeld, E. y Laimins, L. A. (2016). Human Papillomaviruses: Research Priorities for the Next Decade. *Trends in Cancer*, **2**(5): 234–240.

Lape, M., Elam, C., y Paula, S. (2010). Comparison of current docking tools for the simulation of inhibitor binding by the transmembrane domain of the sarco/endoplasmic reticulum calcium ATPase. *Biophysical chemistry*, **150**(1-3): 88–97.

Lavezzo, E., Masi, G., Toppo, S., Franchin, E., Gazzola, V., Sinigaglia, A., Masiero, S., Trevisan, M., Pagni, S., Palù, G., y Barzon, L. (2016). Characterization of Intra-Type Variants of Oncogenic Human Papillomaviruses by Next-Generation Deep Sequencing of the E6/E7 Region. *Viruses*, **8**(3): 79.

Law, V., Knox, C., Djoumbou, Y., Jewison, T., Guo, A. C., Liu, Y., MacIejewski, A., Arndt, D., Wilson, M., Neveu, V., Tang, A., Gabriel, G., Ly, C., Adamjee, S., Dame, Z. T., Han, B., Zhou, Y., y Wishart, D. S. (2014). DrugBank 4.0: Shedding new light on drug metabolism. *Nucleic Acids Research*, **42**(D1).

Lee, H. G., Yu, K. A., Oh, W. K., Baeg, T. W., Oh, H. C., Ahn, J. S., Jang, W. C., Kim, J. W., Lim, J. S., Choe, Y. K., y Yoon, D. Y. (2005). Inhibitory effect of jaceosidin isolated from Artemisia argyi on the function of E6 and E7 oncoproteins of HPV 16. *Journal of Ethnopharmacology*, **98**(3): 339–343.

Li, N., Thompson, S., Schultz, D. C., Zhu, W., Jiang, H., Luo, C., y Lieberman, P. M. (2010). Discovery of selective inhibitors against ebna1 via high throughput in silico virtual screening. *PLoS ONE*, **5**(4).

Lipinski, C. A., Lombardo, F., Dominy, B. W., y Feeney, P. J. (2001). Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings1PII of original article: S0169-409X(96)00423-1. The article was originally published in Advanced Drug Delivery Reviews 23 (1997) 3. *Advanced Drug Delivery Reviews*, **46**(1-3): 3–26.

Loguercio, C. y Festi, D. (2011). Silybin and the liver: from basic research to clinical practice. *World journal of gastroenterology*, **17**(18): 2288–301.

Maglennon, G. A., McIntosh, P., y Doorbar, J. (2011). Persistence of viral DNA in the epithelial basal layer suggests a model for papillomavirus latency following immune regression. *Virology*, **414**(2): 153–163.

Maher, D. M., Bell, M. C., O'Donnell, E. A., Gupta, B. K., Jaggi, M., y Chauhan, S. C. (2011). Curcumin suppresses human papillomavirus oncoproteins, restores p53, rb, and ptpn13 proteins and inhibits benzo[a]pyrene-induced upregulation of HPV E7. *Molecular Carcinogenesis*, **50**(1): 47–57.

Maier, J. A., Martinez, C., Kasavajhala, K., Wickstrom, L., Hauser, K. E., y Simmerling, C. (2015). ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. *Journal of Chemical Theory and Computation*, **11**(8): 3696–3713.

Malecka, K. A., Fera, D., Schultz, D. C., Hodawadekar, S., Reichman, M., Donover, P. S., Murphy, M. E., y Marmorstein, R. (2014). Identification and characterization of small molecule human papillomavirus E6 inhibitors. *ACS Chemical Biology*, **9**(7): 1603–1612.

Mamgain, S., Sharma, P., Pathak, R. K., y Baunthiyal, M. (2015). Computer aided screening of natural compounds targeting the E6 protein of HPV using molecular docking. *Bioinformation*, **11**(5): 236–242.

Martinez-Zapien, D., Ruiz, F. X., Poirson, J., Mitschler, A., Ramirez, J., Forster, A., Cousido-Siah, A., Masson, M., Vande Pol, S., Podjarny, A., Travé, G., y Zanier, K. (2016). Structure of the E6/E6AP/p53 complex required for HPV-mediated degradation of p53. *Nature*, **529**(7587): 541–5.

Marz, M., Beerenwinkel, N., Drosten, C., Fricke, M., Frishman, D., Hofacker, I. L., Hoffmann, D., Middendorf, M., Rattei, T., Stadler, P. F., y T??pfer, A. (2014). Challenges in RNA virus bioinformatics. *Bioinformatics*, **30**(13): 1793–1799.

McLaughlin-Drubin, M. E. y Münger, K. (2009). The human papillomavirus E7 oncoprotein. *Virology*, **384**(2): 335–344.

Meng, X.-Y., Zhang, H.-X., Mezei, M., y Cui, M. (2011). Molecular Docking: A Powerful Approach for Structure-Based Drug Discovery. *Current Computer Aided-Drug Design*, **7**(2): 146–157.

Milrot, E., Jackman, A., Kniazhanski, T., Gonen, P., Flescher, E., y Sherman, L. (2012). Methyl jasmonate reduces the survival of cervical cancer cells and downregulates HPV E6 and E7, and survivin. *Cancer Letters*, **319**(1): 31–38.

Mirza, S. B., Salmas, R. E., Fatmi, M., y Durdagi, S. (2016). Virtual Screening of Eighteen Million Compounds against Dengue Virus: Combined Molecular Docking and Molecular Dynamics Simulations Study. *Journal of Molecular Graphics and Modelling*, **66**: 99–107.

Monie, A., Hung, C.-F., Roden, R., y Wu, T.-C. (2008). Cervarix: a vaccine for the prevention of HPV 16, 18-associated cervical cancer. *Biologics : targets & therapy*, **2**(1): 97–105.

Morris, G. y Huey, R. (2009). AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. *Journal of . . .* , **30**(16): 2785–2791.

Munagala, R., Kausar, H., Munjal, C., y Gupta, R. C. (2011). Withaferin a induces p53-dependent apoptosis by repression of HPV oncogenes and upregulation of tumor suppressor proteins in human cervical cancer cells. *Carcinogenesis*, **32**(11): 1697–1705.

Murgueitio, M. S., Bermudez, M., Mortier, J., y Wolber, G. (2012). In silico virtual screening approaches for anti-viral drug discovery. *Drug Discovery Today: Technologies*, **9**(3): 219–225.

Nguyen, H. T., Le, L., y Truong, T. N. (2009). Top-hits for A/H1N1 Identified by Virtual Screening Using Ensemble-based Docking. *PLoS Currents*, **1**: RRN1030.

Nominé, Y., Charbonnier, S., Ristriani, T., Stier, G., Masson, M., Cavusoglu, N., Van Dorsselaer, A., Weiss, E., Kieffer, B., y Travé, G. (2003). Domain substructure of HPV E6 oncoprotein: Biophysical characterization of the E6 C-terminal DNA-binding domain. *Biochemistry*, **42**(17): 4909–4917.

Ntie-Kang, F. (2013). An in silico evaluation of the ADMET profile of the StreptomeDB database. *SpringerPlus*, **2**(Hodgson 2001): 353.

Olsson, M. H. M., SØndergaard, C. R., Rostkowski, M., y Jensen, J. H. (2011). PROPKA3: Consistent treatment of internal and surface residues in empirical p K a predictions. *Journal of Chemical Theory and Computation*, **7**(2): 525–537.

Palasap, A., Limpaiboon, T., Boonsiri, P., Thapphasaraphong, S., Daduang, S., Suwannalert, P., y Daduang, J. (2014). Cytotoxic effects of phytophenolics from Caesalpinia mimosoides Lamk on cervical carcinoma cell lines through an apoptotic pathway. *Asian Pacific Journal of Cancer Prevention*, **15**(1): 449–454.

Park, J. W., Min, K.-J., Kim, D. E., y Kwon, T. K. (2015). Withaferin A induces apoptosis through the generation of thiol oxidation in human head and neck cancer cells. *International Journal of Molecular Medicine*, **35**(1): 247–252.

Peters, M. B., Yang, Y., Wang, B., Füsti-Molnár, L., Weaver, M. N., y Merz, K. M. (2010). Structural survey of zinc-containing proteins and development of the zinc AMBER force field (ZAFF). *Journal of Chemical Theory and Computation*, **6**(9): 2935–2947.

Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C., y Ferrin, T. E. (2004). UCSF Chimera - A visualization system for exploratory research and analysis. *Journal of Computational Chemistry*, **25**(13): 1605–1612.

Phillips, J. C., Braun, R., Wang, W., Gumbart, J., Tajkhorshid, E., Villa, E., Chipot, C., Skeel, R. D., Kalé, L., y Schulten, K. (2005). Scalable molecular dynamics with NAMD.

Pirouzpanah, M. B., Sabzichi, M., Pirouzpanah, S., Chavoshi, H., y Samadi, N. (2015). Silibilin-induces apoptosis in breast cancer cells by modulating p53, p21, bak and bcl-xl pathways. *Asian Pacific Journal of Cancer Prevention*, **16**(5): 2087–2092.

Podvinec, M., Lim, S. P., Schmidt, T., Scarsi, M., Wen, D., Sonntag, L. S., Sanschagrin, P., Shenkin, P. S., y Schwede, T. (2010). Novel inhibitors of dengue virus methyltransferase: Discovery by in vitro-driven virtual screening on a desktop computer grid. *Journal of Medicinal Chemistry*, **53**(4): 1483–1495.

R Core Team (2017). R: A Language and Environment for Statistical Computing.

Rambaut, A. (2012). FigTree v1. 4.0. A graphical viewer of phylogenetic trees. *Institute of Evolutionary Biology University of Edinburgh*.

Reed, G. A., Arneson, D. W., Putnam, W. C., Smith, H. J., Gray, J. C., Sullivan, D. K., Mayo, M. S., Crowell, J. A., y Hurwitz, A. (2006). Single-Dose and Multiple-Dose Administration of Indole-3-Carbinol to Women: Pharmacokinetics Based on 3,3'-Diindolylmethane. *Cancer Epidemiology Biomarkers & Prevention*, **15**(12): 2477–2481.

Rietz, A., Petrov, D. P., Bartolowits, M., DeSmet, M., Davisson, V. J., y Androphy, E. J. (2016). Molecular Probing of the HPV-16 E6 Protein Alpha Helix Binding Groove with Small Molecule Inhibitors. *Plos One*, **11**(2): e0149845.

Saha, S. K. y Khuda-Bukhsh, A. R. (2015). Berberine alters epigenetic modifications, disrupts microtubule network, and modulates HPV-18 E6-E7 oncoproteins by targeting p53 in cervical cancer cell HeLa: A mechanistic study including molecular docking. *European Journal of Pharmacology*, **744**: 132–146.

Šali, a. (2013). MODELLER: A Program for Protein Structure Modeling Release 9.12, r9480. *Rockefeller University*, pp. 779–815.

Satoh, A. (2011). *Introduction to Practice of Molecular Simulation*. Elsevier.

Schames, J. R., Henchman, R. H., Siegel, J. S., Sotriffer, C. A., Ni, H., y McCammon, J. A. (2004). Discovery of a Novel Binding Trench in HIV Integrase. *Journal of Medicinal Chemistry*, **47**(8): 1879–1881.

Schmitz, M., Driesch, C., Jansen, L., Runnebaum, I. B., y Dürst, M. (2012). Non-random integration of the HPV genome in cervical cancer. *PLoS ONE*, **7**(6).

Schneider, G. (2010). Virtual screening: an endless staircase? *Nat Rev Drug Discov*, **9**(4): 273–276.

Schrödinger (2017a). Schrödinger Release 2017-1: QikProp.

Schrödinger (2017b). Schrödinger Release 2017-1: LigPrep.

Seedorf, K., Durst, M., Dürst, M., Suhai, S., y Röwekamp, W. (1985). Human papillomavirus type 16 DNA sequence. *Virology*, **185**: 181–185.

Shah, M., Anwar, M. A., Park, S., Jafri, S. S., y Choi, S. (2015). In silico mechanistic analysis of IRF3 inactivation and high-risk HPV E6 species-dependent drug response. *Scientific reports*, **5**(August 2014): 13446.

Shi, L., Sings, H. L., Bryan, J. T., Wang, B., Wang, Y., Mach, H., Kosinski, M., Washabaugh, M. W., Sitrin, R., y Barr, E. (2007). GARDASIL: prophylactic human papillomavirus vaccine development–from bench top to bed-side. *Clinical pharmacology and therapeutics*, **81**(2): 259–264.

Singh, A. K. y Misra, K. (2013). Human papilloma virus 16 e6 protein as a target for curcuminoids, curcumin conjugates and congeners for chemoprevention of oral and cervical cancers. *Interdisciplinary Sciences: Computational Life Sciences*, **5**(2): 112–118.

Skjærven, L., Yao, X.-Q., Scarabelli, G., y Grant, B. J. (2014). Integrating protein structural dynamics and evolutionary analysis with Bio3D. *BMC bioinformatics*, **15**(1): 399.

Smith, H. B. (2013). *Handbook on human papillomavirus prevalence, detection and management*. Nova Biomedical. p. 548.

Song, S., Liem, A., Miller, J. A., y Lambert, P. F. (2000). Human papillomavirus types 16 E6 and E7 contribute differently to carcinogenesis. *Virology*, **267**(2): 141–150.

Stein, S. A. M., Loccisano, A. E., Firestine, S. M., y Evanseck, J. D. (2006). Principal Components Analysis: A Review of its application on molecular dynamics data. *Annual Reports in Computational Chemistry*, **2**(06): 233–248.

Sterling, T. y Irwin, J. J. (2015). ZINC15–Ligand Discovery for Everyone. *Journal of Chemical Information and Modeling*, **55**(11): 2324–2337.

Summa, V., Petrocchi, A., Bonelli, F., Crescenzi, B., Donghi, M., Ferrara, M., Fiore, F., Gardelli, C., Gonzalez Paz, O., Hazuda, D. J., Jones, P., Kinzel, O., Laufer, R., Monteagudo, E., Muraglia, E., Nizi, E., Orvieto, F., Pace, P., Pescatore, G., Scarpelli, R., Stillmock, K., Witmer, M. V., y Rowley, M. (2008). Discovery of Raltegravir, a Potent, Selective Orally Bioavailable HIV-Integrase Inhibitor for the Treatment of HIV-AIDS Infection. *Journal of Medicinal Chemistry*, **51**(18): 5843–5855.

Tambunan, U. S. F. y Parikesit, A. A. (2012). HPV Bioinformatics: In Silico Detection, Drug Design and Prevention Agent Development. *Topics on Cervical Cancer with an Advocacy for Prevention*, pp. 237–252.

The UniProt Consortium (2014). UniProt: a hub for protein information. *Nucleic Acids Research*, **43**(Database issue): 204–12.

Trott, O. y Olson, A. J. (2010a). AutoDock Vina: Improving the Speed and Accuracy of Docking with a New Scoring Function, EfficientOptimization, and Multithreading. *Journal of Computational Chemistry*, **31**(2): 455–61.

Trott, O. y Olson, A. J. (2010b). AutoDock Vina - molecular docking and virtual screening program.

Umamaheswari, A., Kumar, M. M., Pradhan, D., y Marisetty, H. (2011). Docking studies towards exploring antiviral compounds against envelope protein of yellow fever virus. *Interdisciplinary Sciences: Computational Life Sciences*, **3**(1): 64–77.

Villiers, E.-m. D., Fauquet, C., Broker, T. R., y Bernard, H.-u. (2004). Classification of papillomaviruses. **324**: 17–27.

Wang, C., Hai, Y., Liu, X., Liu, N., Yao, Y., He, P., y Dai, Q. (2015). Prediction of high-risk types of human papillomaviruses using statistical model of protein "sequence space". *Computational and mathematical methods in medicine*, **2015**: 756345.

Wang, J., Wolf, R. M., Caldwell, J. W., Kollman, P. A., y Case, D. A. (2004). Development and testing of a general Amber force field. *Journal of Computational Chemistry*, **25**(9): 1157–1174.

Wang, J., Wang, W., Kollman, P. A., y Case, D. A. (2006). Automatic atom type and bond type perception in molecular mechanical calculations. *Journal of Molecular Graphics and Modelling*, **25**(2): 247–260.

Wang, R., Lai, L., y Wang, S. (2002). Further development and validation of empirical scoring functions for structure-based binding affinity prediction. *Journal of Computer-Aided Molecular Design*, **16**(1): 11–26.

WHO (2014). Human papillomavirus vaccines: WHO position paper, October 2014. *World Health Organisation Weekly epidemiological record*, **89**(43): 465–492.

Yim, E. K., Lee, M. J., Lee, K. H., Um, S. J., y Park, J. S. (2006). Antiproliferative and antiviral mechanisms of ursolic acid and dexamethasone in cervical carcinoma cell lines. *International Journal of Gynecological Cancer*, **16**(6): 2023–2031.

Yu, H. C., Chen, L. J., Cheng, K. C., Li, Y. X., Yeh, C. H., y Cheng, J. T. (2012). Silymarin inhibits cervical cancer cell through an increase of phosphatase and tensin homolog. *Phytotherapy Research*, **26**(5): 709–715.

Yuan, C.-H., Filippova, M., Krstenansky, J. L., y Duerksen-Hughes, P. J. (2016). Flavonol and imidazole derivatives block HPV16 E6 activities and reactivate apoptotic pathways in HPV+ cells. *Cell Death and Disease*, **7**(1): 2060.

Zanier, K., M'Hamed Ould Sidi, A. O., Boulade-Ladame, C., Rybin, V., Chappelle, A., Atkinson, A., Kieffer, B., y Travé, G. (2012). Solution structure analysis of the HPV16 E6 oncoprotein reveals a self-association mechanism required for E6-mediated degradation of p53. *Structure*, **20**(4): 604–617.

Zanier, K., Charbonnier, S., Ould, A., Ould, M., Mcewen, A. G., Ferrario, M. G., Poussin-courmontagne, P., Cura, V., Brimer, N., Babah, K. O., Ansari, T., Muller, I., Stote, R. H., Cavarelli, J., Pol, S. V., y Travé, G. (2013). Structural Basis for Hijacking of Cellular LxxLL Motifs by. **339**(February).

Zanier, K., Stutz, C., Kintscher, S., Reinz, E., Sehr, P., Bulkescher, J., Hoppe-Seyler, K., Travé, G., y Hoppe-Seyler, F. (2014). The E6AP binding pocket of the HPV16 E6 oncoprotein provides a docking site for a small inhibitory peptide unrelated to E6AP, indicating druggability of E6. *PLoS ONE*, **9**(11).

Zhang, Y., Hou, F., Liu, X., Ma, D., Zhang, Y., Kong, B., y Cui, B. (2014). Tc17 cells in patients with uterine cervical cancer. *PLoS ONE*, **9**(2): 1–8.

Zheng, Z.-M. y Baker, C. C. (2006). Papillomavirus genome structure, expression, and post-transcriptional regulation. *Frontiers in bioscience : a journal and virtual library*, **11**: 2286–302.

zur Hausen, H. (2009). Papillomaviruses in the causation of human cancers - a brief historical account. *Virology*, **384**(2): 260–265.

Apéndice A. Metodología y Resultados

A.1. Protocolo de dinámica molecular

Script con los parámetros utilizados para llevar a cabo la minimización de energía potencial utilizando el programa Amber16.

```
1    Minimización de la energía del sistema:
2 &cntrl
3    imin=1,        ! Cuando = 1, indica llevar a cabo una minimización.
4    ntmin=1,       ! Método de minimización, cuando = 1 implica cambiar de Steepest
      decendant (SD) a Conjugate Gradient (CG).
5    drms=0.1,      ! Criterio de convergencia para el gradiente de energía.
6    maxcyc=10000,! Número máximo de ciclos para la minimización.
7    ncyc=5000,     ! Cuando ntmin =  1, indica el paso para cambiar de SD a CG.
8    ntx=1,         ! Si = 1, lee las coordenadas iniciales del archivo inpcrd.
9    irest=0,       ! Flag para reiniciar una simulación; 0 = no reiniciar.
10   ntpr=100,      ! Intervalo de pasos para guardar información del sistema.
11   ntwr=100,      ! Intervalo de pasos para guardar un archivo restart.
12   iwrap=0,       ! Cuando = 1, realiza un wrap en favor de la visualización de los á
      tomos dentro de la caja. No tiene ningún efecto sobre la dinámica.
13   ntf=1,         ! Determina si evaluar las fuerzas enlazantes con H, cuando = 1 se
      evalúan dichas fuerzas.
14   ntb=1,         ! Controla el tipo de simulación y el PME; cuando = 1 se mantiene V
      constante (ensamble NVT)
15   cut=10.0,      ! Distancia máxima en A para la evaluación de las interacciones
      electrostáticas
16   nsnb=20,       ! Determina la frecuencia en pasos de integración de la evaluación de
       las fuerzas no enlazantes
17 &end
18 /
```

Script con los parámetros utilizados para llevar a cabo el proceso de *Simulated annealing* y equilibrado utilizando el programa Amber16.

```
1    Simulated Annealing y Equilibrado:
2 &cntrl
3     iwrap=1, ntx=1,
4     dt=0.001,     ! Tiempo de integración en ps.
5     nstlim=8000000, ! Número de pasos de simulación a realizar: 8 ns.
6     ntf=2,        ! Si = 2, omite evaluar las interacciones enlazantes de los H.
7     ntc=2,        ! Si = 2, utiliza SHAKE para restringir los enlaces de H.
8     tempi=0.0,    ! Temperatura inicial.
9     temp0=300.0,! Temperatura final o de referencia.
10    ntpr=5000,    ! Intervalo de pasos para guardar información del sistema.
11    ntwx=5000,    ! Intervalo para guardar las coordenadas de los átomos del sistema en
       un archivo "mdrc".
12    cut=10.0, ntb=1,
13    ntt=3,        ! Si = 3, se utiliza el termostato de Langevin para regular T.
14    gamma_ln=2.0, ! Frecuencia de colisión atómica para evaluar T.
15    ig=-1,        ! semilla para generar los números aleatorios, cuando = -1 se utiliza
       la fecha del sistema operativo.
16 /
17 ! Para calentar durante 0.5 ns y equilibrar durante 5 ns.
18 ! 1) Calienta del paso 1 al paso 500 000 (0.5 ns 0K a 300K)
19 &wt type='TEMP0', istep1=0, istep2=500000, value1=0.0, value2=300 /
20 ! 2) Equilibra a 300K por 0.5ns
```

```
21 &wt type='TEMP0', istep1=500001, istep2=1000000, value1=300, value2=300 /
22 ! 3) Calienta a 400K por 0.5ns (0.5 ns, 300K a 400K)
23 &wt type='TEMP0', istep1=1000001, istep2=1500000, value1=300.0, value2=400 /
24 ! 4) Equilibra a 400K por 0.5ns (0.5 ns, 400K)
25 &wt type='TEMP0', istep1=1500001, istep2=2000000, value1=400, value2=400 /
26 ! 5) Enfría a 310K por 0.5ns (0.5 ns, 400K a 310K)
27 &wt type='TEMP0', istep1=2000001, istep2=2500000, value1=400.0, value2=310 /
28 ! 6) Equilibra a 310K por 5.5ns (5.5 ns, 310K)
29 &wt type='TEMP0', istep1=2500001, istep2=8000000, value1=310.0, value2=310 /
30
31 &wt type='END' /
32 &end
33 /
```

Script con los parámetros utilizados para llevar a cabo la fase de producción de dinámica molecular utilizando el programa Amber16.

```
1     Producción de Dinámica Molecular:
2 &cntrl
3      imin=0, iwrap=1,
4      ntx=5,        ! Si = 5, lee las coordenadas y velocidades iniciales del archivo
       restrt.
5      irest=1,      ! Si = 1 reinicia la simulación a partir del archivo restrt.
6      nstlim=50000000, ! Número de pasos de simulación a realizar: 100 ns.
7      dt=0.002,     ! Tiempo de integración en ps.
8      temp0=310.0,! Temperatura final o de referencia.
9      ntf=2, ntc=2, ntpr=10000, ntwx=10000, cut=10.0,
10     ntb=2,        ! Cuando = 2 se mantiene P constante (ensamble NPT).
11     ntp=1,        ! Activa el barostato para regular la presión del sistema.
12     ntt=3, gamma_ln=2.0,  ig=-1,
13 &end
14 /
```

A.2. Análisis de trayectorias

Ejemplo del *script* para el análisis de trayectorias de SDM, utilizando R y el paquete Bio3D.

```
1 # *** Ejemplo del aálisis de trayectorias llevado a cabo, tomando como
      referencia la documentación de Bio3D y el protocolo mostrado en la
      siguiente página: https://bitbucket.org/Grantlab/bio3d/issues/326/give
      -up-trying-to-find-out-how-to-extract ***
2
3 # Se carga el paquete Bio3D
4 library(bio3d)
5
6 # Lectura del archivo pdb de la proteína
7 pdb.e6 <- read.pdb("/ruta/al/archivo/proteina_e6.pdb")
8 # Selección de los carbonos alfa de los residuos de la proteína
9 ca.inds <- atom.select(pdb.e6, elety = "CA")
10 # Lectura del archivo de trayectoria, ncdf.
11 ncdf.e6 <- read.ncdf("/ruta/al/archivo/proteina_e6_sdm_produccion.ncdf")
12 # Superposición de los C-alfa de las conformaciones con el algoritmo de
      Kabsch.
13 xyz.ca.e6 <- fit.xyz(fixed = pdb.e6$xyz, mobile = ncdf.e6,
14            fixed.inds = ca.inds$xyz, mobile.inds = ca.inds$xyz) # Se crea
      una matriz de 3N columnas y T filas. N = # de carbonos alfa, T = #
      frames
15
16 ####### RMSD #######
```

```
rmd.e6 <- rmsd(pdb.e6, xyz.ca.e6[,]) # Devuelve un vector con T valores de rmsd

####### RMSF #######
rmf.e6 <- rmsf(xyz.ca.e6[,]) # Devuelve un vector con N valores de rmsf

####### ACP ########
acp.e6 <- pca.xyz(xyz.ca.e6[,]) # Devuelve una lista de objetos con: los eigenvalores (L), los eigenvectores (U), los scores (z), y los loadings (au)

####### CLUSTERING #####
# Se crea la matriz de distancia
dst.matx <- dist(acp.e6$z[,1:3]) # En este ejemplo se toman en cuenta los CP 1, 2 y 3
# Se crea el dendograma
cj.e6 <- hclust(dst.matx, method="ward.D") # Método de clustering
# Se eligen los clusters en función del k deseado (k = 2 para este ejemplo)
clus.conf.e6 <- cutree(cj.e6, k=2) # k equivale al número de grupos elegido

##### SELECCIÓN DE LA CONFORMACIÓN POR CLUSTER #####
# Se crea una conformación por cluster (a y b para este ejemplo), con la media de las posiciones de los C-alfa en los frames correspondientes a cada cluster.
conf.med.clus.a <- colMeans(xyz.ca.e6[clus.conf.e6==1,])
conf.med.clus.b <- colMeans(xyz.ca.e6[clus.conf.e6==2,])

# *** Se elige la conformación, dentro de la trayectoria original, que poseea el valor mínimo de RMSD frente a la conformación media. Se ejempliica sólo para el cluster a. ***

# RMSD de la trayectoria con la conformación media como referencia
rmd.med.conf.a <- rmsd(conf.med.clus.a, xyz.ca.e6[,])
# Se elige el frame (conformación) más parecida a la media, en función de su valor de RMSD
conf.clus.a <- which.min(rmd.med.conf.a); conf.clus.a
# El resultado es el número de frame utilizado como estructura representativa del cluster
```

A.3. Alineamiento múltiple de secuencias.

```
VPH-16_AR1-A9     MHQKRTAMFQDPQERPRKLPQLCTELQTTIHDIILECVYCKQQLBRREVYDFAFRDLCIVYRDGNPYAVCDKCLK     75
VPH-31_AR1-A9     .......MFKNPAERPRKLHELSSALEIPYDEBRLNCVYCKGQLTETEVLDFAFTDLTIVYRDDTPHGVCTKCBR     68
VPH-33_AR1-A9     .......MFQDTEEKPRTLHDLCQALETTIHNIELQCVECKKPLQRSEVYDFAFADLTVVYREGNPFGICKLCBR     68
VPH-52_AR1-A9     .......MFEDPATRPRTLHELCEVLEESVHEIRLQCVQCKKELQRREVYKFLFTDBRIVYRDNNPYGVCIMCBR     68
VPH-58_AR1-A9     .......MFQDAEEKPRTLHDLCQALETSVHEIELKCVECKKTLQRSEVYDFVFADBRIVYRDGNPFAVCKVCBR     68
VPH-18_AR1-A7     .....MARFEDPTRRPYKLPDLCTELNTSLQDIEITCVYCKTVLELTEVFEFAFKDLFVVYRDSIPHAACHKCID     70
VPH-45_AR1-A7     .....MARFDDPKQRPYKLPDLCTELNTSLQDVSIACVYCKATLERTEVYQFAFKDLCIVYRDCIAYAACHKCID     70
VPH-35_AR2-A9     .......MFQDPAERPYKLHDLCNEVEESIHEICLNCVYCKQELQRSEVYDFACYDLCIVYREGQPYGVCMKCLK     68
VPH-67_AR2-A9     .......MFQDTDEKPRNLHELCEALETTVHEISLPCVQCKKTLDTNEVYDFLFTDLKIVYRCGNPYGVCKQCBR     68
VPH-39_AR2-A7     .....MARFHNPAERPYKLPDLCTTLDTTLQDITIACVYCRRPLQQTEVYEFAFSDLYVVYRDGEPLAACQSCIK     70
VPH-59_AR2-A7     .....MARFEDPTQRPYKLPDLSTTLNIPLHDIRINCVFCKGELQEREVFEFAFNDLFIVYRDCTPYAACLKCIS     70
VPH-68_AR2-A7     .....MALFHNPEERPYKLPDLCRTLDTTLHDVTIDCVYCRRQLQRTEVYEFAFSDLCVVYRDGVPFAACQSCIK     70
VPH-70_AR2-A7     .....MARFPNPAERPYKLPDLCTALDTTLHDITIDCVYCKTQLQQTEVYEFAFSDLFIVYRNGEPYAACQKCIK     70
VPH-26_AR2-A5     .......MFEDPRERPRTLHELCESLNTTLQNLQVQCVYCKETLQWADVYNFAICDBRVVYRDRSPYAACKRCVI     68
VPH-6A_BR1-A10    ......MESANASTSATTIDQLCKTFNLSMHTLQINCVFCKNALTTAEIYSYAYKQLKVLFRGGYPYAACACCLE     69
VPH-11_BR1-A10    ......MESKDASTSATSIDQLCKTFNLSLHTLQIQCVFCRNALTTAEIYAYAYKNLKVVWRDNFPFAACACCLE     69
VPH-44_BR1-A10    ......MESANASTSAQSIDQLCKECNIPMHNLQILCVFCRKTLSTAEVYSFAYKQLYVVYRGNFPFAACAICLE     69
VPH-74_BR1-A10    ......MESANASTSAKSIDQLCKDCNIPMHNLQISCVFCRKTLSCPEVYSFAYKNLYIVYRGNFPFAACAICLE     69
VPH-54_BR1-A13    .....MSATEPHTDQPRTLADLCKVCNIPMHSLQLPCAFCKKTVCTAEIYAFQYKDLFVVWRHGFPHAACALCLE     70
VPH-40_BR2-A8     .......MSARCGSQARTLYELCDQCNITLPTLQIDCVFCKTVLKTAEVLAFAFRELYVVWRDDFPHAACPRCLD     68
VPH-43_BR2-A8     ......MSARSCSQNARTIFELCDECNITLPTLQIGCIFCKKWLLTTEVLSFAFRDBRVVWRDGYPFAACLACLQ     69
VPH-49_Ext_Beta   MARPVKVCELAHHLNIPIWEVLLPCNFCTGFLTYQELLEFDYKDFNLLWKDGFVFGCCAACAYRSAYHEFTNYHQ     75

VPH-16_AR1-A9     FYSKISEYRHYCYSLYGTTLEQQYNKPLCDLLIRCINCQKPLCPEEKQRHLDKKQRFHNIRGRWTGRCMSCC...    147
VPH-31_AR1-A9     FYSKVSEFRWYRYSVYGTTLEKLTNKGICDLLIRCITCQRPLCPEEKQRHLDKKKRFHNIGGRWTGRCIACW...    140
VPH-33_AR1-A9     FLSKISEYRHYNYSVYGNTLEQTVKKPLNEILIRCIICQRPLCPQEKKRHVDLNKRFHNISGRWAGRCAACW...    140
VPH-52_AR1-A9     FLSKISEYRHYQYSLYGKTLEERVKKPLSEITIRCIICQTPLCPEEKERHVNANKRFHNIMGRWTGRCSECW...    140
VPH-58_AR1-A9     LLSKISEYRHYNYSLYGDTLEQTLKKCLNEILIRCIICQRPLCPQEKKRHVDLNKRFHNISGRWTGRCAVCW...    140
VPH-18_AR1-A7     FYSRIREBRHYSDSVYGDTLEKLTNTGLYNLLIRCBRCQKPLNPAEKBRHLNEKRRFHNIAGHYRGQCHSCCNRA    145
VPH-45_AR1-A7     FYSRIREBRYYSNSVYGETLEKITNTELYNLLIRCBRCQKPLNPAEKRRHLKDKRRFHSIAGQYRGQCNTCCDQA    145
VPH-35_AR2-A9     FYSKISEYRWYRYSVYGETLEKQCNKQLCHLLIRCITCQKPLCPVEKQRHLEEKKRFHNIGGRWTGRCMSCW...    140
VPH-67_AR2-A9     LLSKVSEYRYFNYSVYGNTLEDIVHKPLNEITIRCITCQRPLCPQEKQRHVDRKKRFHNISNRWTGRCSVCW...    140
VPH-39_AR2-A7     FYAKIREBRYYSDSVYATTLENITNTKLYNLLIRCMCCLKPLCPAEKBRHLNSKRRFHKIAGSYTGQCRRCWTTK    145
VPH-59_AR2-A7     FYARVREBRYYRDSVYGETLEAETKTPLHELLIRCYRCLKPLCPTDKLKHITEKRRFHNIAGIYTGQCRGCRTRA    145
VPH-68_AR2-A7     FYAKIREBRYYSESVYATTLETITNTKLYNLLIRCMSCLKPLCPAEKBRHLTTKRRLHKIAGNFTGQCRHCWTSK    145
VPH-70_AR2-A7     FHAKVREBRHYSDSVYATTLESITNTKLYDLSIRCMSCLKPLCPAEKBRHVNTKRRFARIAGSYTGQCRHCWTSN    145
VPH-26_AR2-A5     FYSKITEYRRYTCSVYGATLEALTKKSLCNLLIRCARCQMPLGPEEKQRIVDEKRRFHEIAGQWKGLCTNCW...    140
VPH-6A_BR1-A10    FHGKINQYRHFDYAGYATTVEEETKQDILDVLIRCYLCHKPLCEVEKVKHILTKARFIKLNCTWKGRCLHCWTTC    144
VPH-11_BR1-A10    LQGKINQYRHFNYAAYAPTVEEETNEDILKVLIRCYLCHKPLCEIEKLKHILGKARFIKLNNQWKGRCLHCWTTC    144
VPH-44_BR1-A10    LQGKVNQFRHFNYAGYAVTVEEETNKSILDVLIRCYLCHKPLCHVEKVRHILDKARFIKLQDTWKGRCFHCWTSC    144
VPH-74_BR1-A10    LQGKVNQFRHFNYAGYALTVEEETKQSILDVLIRCYLCHKPLCHVEKVRHILEKARFIKLGDTWKGRCLHCWTTC    144
VPH-54_BR1-A13    LHGQINYRRARDRACLWETVEQECGKPLEEIFIRCWLCHKPLCNVEKQRHVDYNRRFHCVRGYWKGRCLHCW...    142
VPH-40_BR2-A8     LHGKVNQYRNFRYAAYAPTVEEETGLTILQVRIRCCKCHKPLSPVEKTNHIVKKTQFFKLKDSWTGYCLHCWKKC    143
VPH-43_BR2-A8     FHGKISQYRHFDYAAYADTVEEETKQTVFDLCIRCCKCHKPLSPVEKVQHIVQKAQFFKIHSVWKGYCLHCWKSC    144
VPH-49_Ext_Beta   EIVVGIEIEGRAAANIAEIVVRCLICLKRLDLLEKLDICAQAREFARVRNRWKGVCRHCRVIE............    138

VPH-16_AR1-A9     RSSR....TRRETQL    158
VPH-31_AR1-A9     RR......PRTETQV    149
VPH-33_AR1-A9     RS......RRRETAL    149
VPH-52_AR1-A9     RP.......RPVTQV    148
VPH-58_AR1-A9     RP......RRRQTQV    149
VPH-18_AR1-A7     RQERLQ..RRRETQV    158
VPH-45_AR1-A7     RQERBR..RRRETQV    158
VPH-35_AR2-A9     KP......TRRETEV    149
VPH-67_AR2-A9     RP......QRTQTQV    149
VPH-39_AR2-A7     REDRRL..TRRETQV    158
VPH-59_AR2-A7     RHBRQQRQARSETLV    160
VPH-68_AR2-A7     REDRRR..IRQETQV    158
VPH-70_AR2-A7     REDRRR..IRRETQV    158
VPH-26_AR2-A5     RP.R....RQTETQV    150
VPH-6A_BR1-A10    ME.........DMLP    150
VPH-11_BR1-A10    ME........DLLP     150
VPH-44_BR1-A10    ME.........TILP    150
VPH-74_BR1-A10    ME.........NILP    150
VPH-54_BR1-A13    KP.............    144
VPH-40_BR2-A8     MEKG....QRSETLC    154
VPH-43_BR2-A8     MEKR....RRSETMC    155
VPH-49_Ext_Beta   ...............    138
```

Figura A.1: Alineamiento Múltiple con las 21 secuencias de la proteína E6 de distintos VPH. El color negro señala los sitios con una conservación igual al 100 % mientras, y el gris los sitios con una conservación $\geq$ 50 %. Figura obtenida con el paquete *TeXshade* (Beitz, 2000)

A.4. Modelado homólogo y dinámica molecular de E6 VPH-11

z Para el modelado homólogo de la proteína E6 VPH-11 se utilizó la secuencia P04019 y como moldes a las estructuras 4XR8.F (identidad: 33.78 %, similitud: 40.0 % y cobertura: 99 %) y 4GIZ (identidad: 41.0 %, similitud: 36.0 % y cobertura: 93.0 %). El mejor modelo seleccionado obtuvo un GA341 de 1.0, un zDOPE de 0.62 (ningún modelo obtuvo un valor menor a 0), un RMSD de 2.7 Å. La dinámica molecular de la proteína siguió el mismo protocolo mencionado para la proteína E6 VPH-16. Se realizó un único ensayo de 100 ns con la proteína E6 VPH-11 sola y otro ensayo con la proteína en interacción con la hélice *LxxLL* del modelo 4XR8. La posición de la hélice con respecto a E6 fue la equivalente a la posición utilizada con E6 VPH-16. La Figura A.2 muestra los resultados de estos análisis.

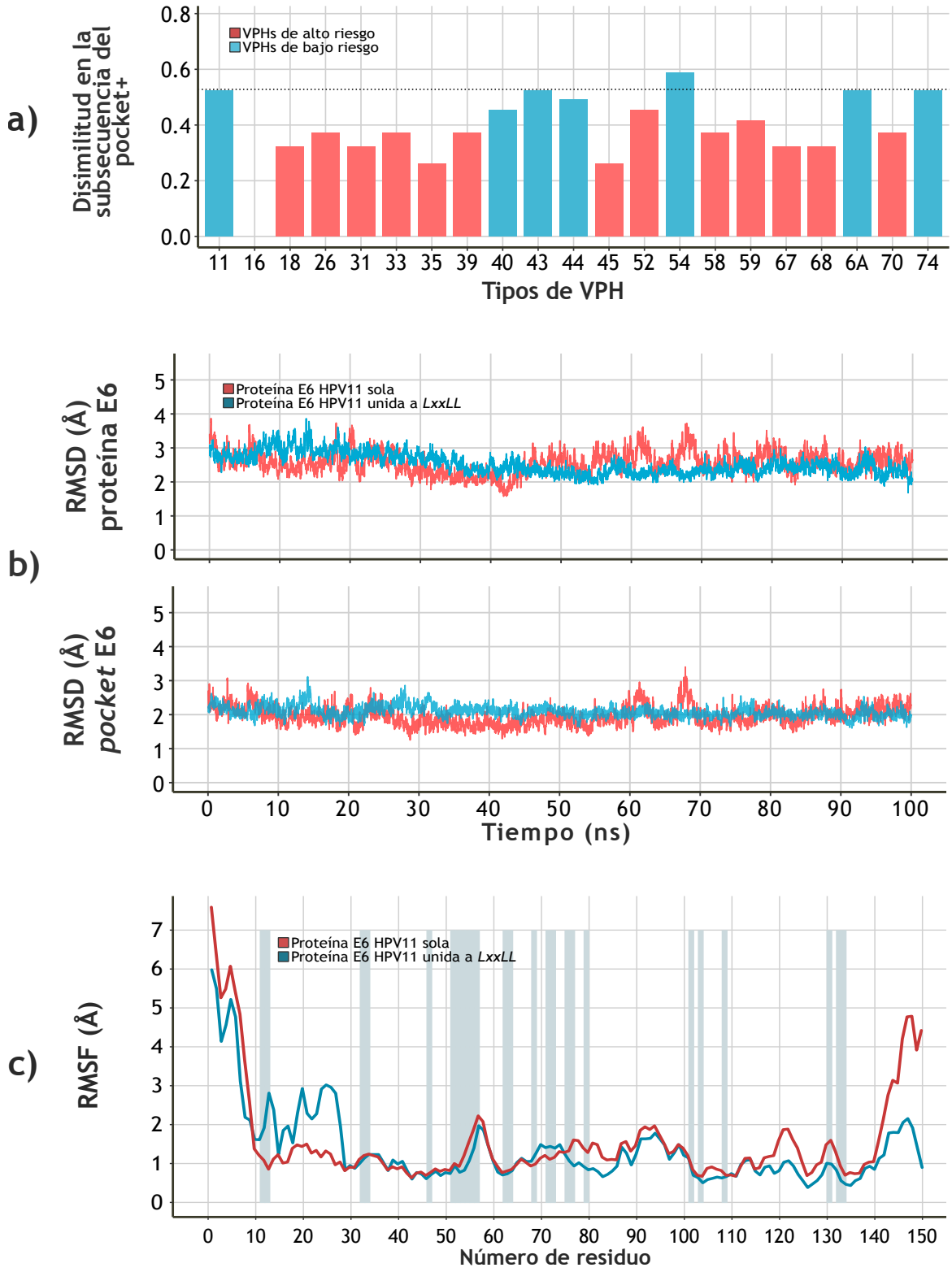


Figura A.2: SDM de la proteína E6 VPH-11 (BR). a) Evaluación de la disimilitud entre las subsecuencias de los *pockets* de proteínas E6 de distintos de VPH de alto y bajo riesgo en comparación con el *pocket* de la proteína E6 VPH-16. La proteína E6 del VPH tipo 11 fue seleccionada por encima de la del tipo 54 tanto por su valor de disimilitud con el tipo 16, como por la importancia clínica del tipo. **b)** RMSD de la secuencia total (arriba) y del *pocket* (abajo) de la proteína E6 VPH-11 sola (rojo) y de E6 VPH-11 en interacción con la hélice (azul). **c)** RMSF de los 150 residuos de E6 VPH-11 sola (rojo) y de E6 VPH-11 en interacción con la hélice (azul).

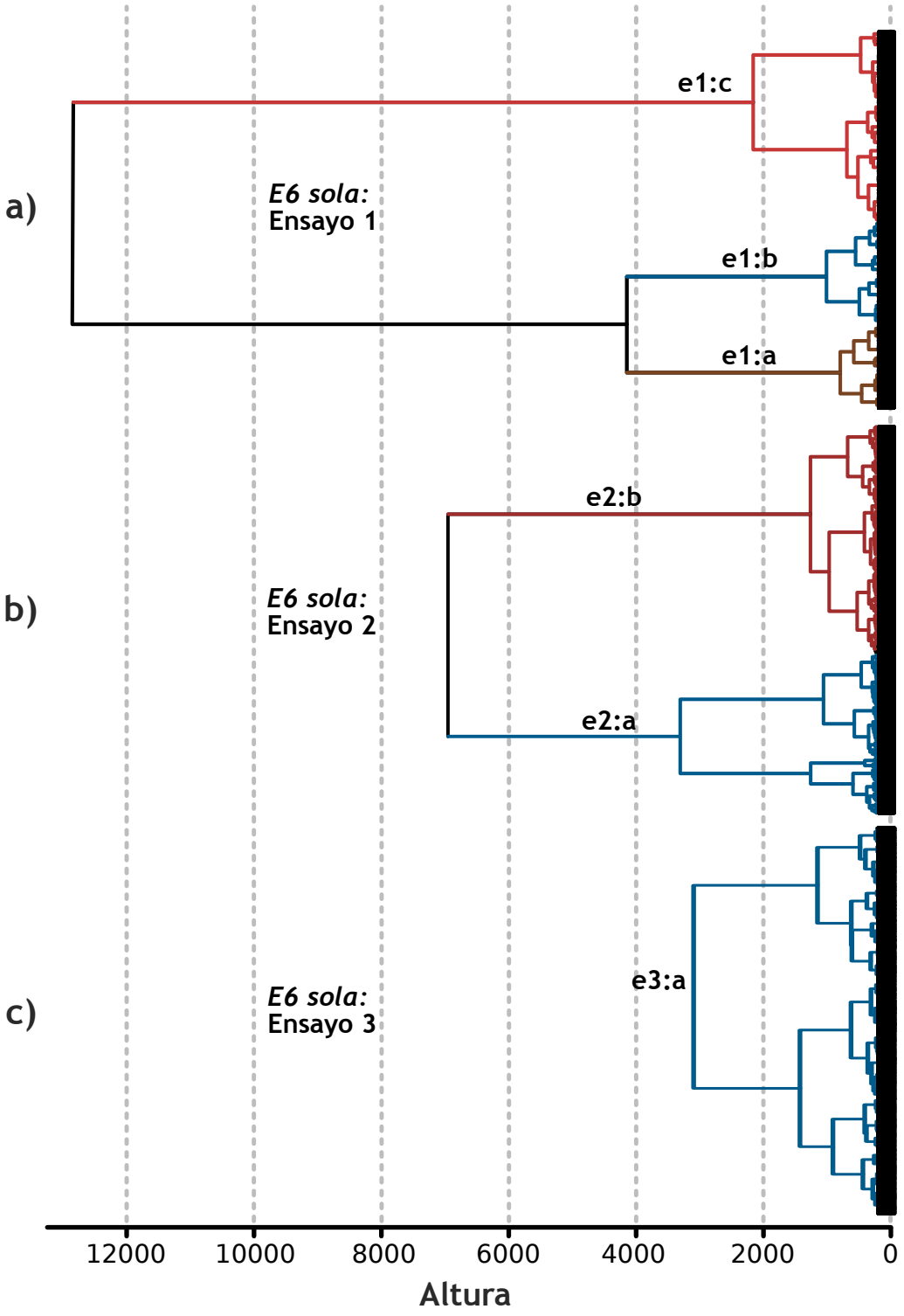


Figura A.3: Dendogramas resultantes del *clustering* Jerárquico de los ensayos de SDM del sistema *E6 sola*. Cada dendograma fue generado empleando la función *hclust()* con el método Ward y la métrica euclidiana implementadas en R, tomando la información de los primeros componentes principales obtenidos en cada ensayo (Sección 2.2.3.3). **a)** ***Ensayo 1. Tres grupos elegidos.*** **b)** ***Ensayo 2. Dos grupos elegidos.*** **c)** ***Ensayo 3. Un grupo elegido.***

A.5. Descriptores para el cribado virtual ADMETox

Regla de cinco de Lipinksi (Lipinski *et al.*, 2001): Este criterio es comúnmente empleado para predecir la permeabilidad celular y la absorción de un compuesto de manera sistémica contemplando factores como la lipofilicidad o el tamaño de la molécula. Al aplicar este filtro se descartaron 7,130 moléculas del conjunto de evaluación. En cuanto a los fármacos control, los antivirales Atazanavir y Sofosvuvir (un profarmaco) fueron retirados al tener un peso molecular mayor a 500 Da y un número de aceptores de enlaces de H mayor a 10. También las moléculas de referencia Epg-3-g y el Carragenano fueron excluidas al violar más de una regla.

Regla de tres de Jorgensen (Jorgensen, 2004): Esta regla está relacionada con la absorción oral, el metabolismo y la biodisponibilidad del compuesto tras su primer paso hepático. Tras su aplicación, 2,992 moléculas fueron excluidas del conjunto de evaluación, incluida la molécula de referencia Silibilina, para la cual estudios experimentales han reportado justamente una baja absorción intestinal y baja biodisponibilidad (Loguercio y Festi, 2011).

Absorción Oral Humana (Schrödinger, 2017b): Este descriptor es representado en valores cualitativos de absorción: bajo, medio y alto, a partir del porcentaje predicho de absorción. Con este filtro 3,826 moléculas fueron descartadas del conjunto de evaluación, incluida la molécula de referencia Ácido Ursólico, un compuesto descrito como altamente hidrofóbico y con baja biodisponibilidad oral (Ge *et al.*, 2015). Y el Indol-3-carbinol, del cual se ha reportado inestabilidad a pHs bajos y una tendencia a oligomerizarse en condiciones de alta acidez (Reed *et al.*, 2006). Por otro lado, el fármaco control Amoxicilina tampoco cumplió con el criterio de esta etapa pues se le predijo una absorción oral baja. No obstante, esto no coincide con los valores experimentales reales, pues la Amoxicilina es reconocida por tener una buena absorción oral, al menos en comparación con otras penicilinas (DrugBank, 2015).

Número de *stars* de *Qikprop* (Schrödinger, 2017b): Este descriptor fue establecido por *Qikprop* para predecir el "perfil tipo fármaco"(*drug-like*) de una molécula a partir de varias propiedades fisicoquímicas. Estas propiedades están relacionadas con la absorción, distribución y toxicidad de los compuestos. Tras la aplicación de este filtro se descartaron 278 moléculas, incluida la de referencia Miyakamida, un antibiótico de origen fúngico reportado por John y Kotadiya (2015) como posible inhibidor de la interacción E6-E6AP.

Penetración dérmica (Schrödinger, 2017b): Este descriptor corresponde al logaritmo de permeabilidad dérmica predicha (*QPlogK*), y con él se descartaron 1,459 compuestos del conjunto de evaluación. Entre estos compuestos se encontró el Acetato de Aurantiamida, también propuesto en los resultados del análisis *in silico* de John y Kotadiya (2015) como posible inhibidor de la interacción E6-E6AP.

A.6. Evaluación de *Autodock 4* y *Autodock Vina*

Para evaluar el desempeño de *Autodock 4* (AD4) y *Autodock Vina* (Vina), se realizó una primer fase de Simulación de Acoplamiento Molecular (SAM) con ambos programas, y utilizando como espacio de búsqueda toda la superficie de la proteína E6 HPV-16; un análisis conocido como anclaje molecular ”ciego”. El proceso consistió en utilizar una únicamente la molécula curcumina como ligando. Para *AD4* se hicieron cinco análisis con diferente número de evaluaciones de la energía de interacción proteína-ligando: 2 000 000, 4 000 000, 8 000 000, 16 000 000 y 32 000 000. Cada análisis constó de 50 corridas (*runs*), 2700 generaciones y una caja 40*40*40 Å con un tamaño de rejilla de 0.375 Å (suficiente para cubrir la proteína E6). Se realizaron 20 SAM independientes con la curcumina, pero partiendo de diferentes posiciones iniciales de ésta en relación a la proteína E6 (Fig. A.4a). El mismo proceso fue llevado a cabo con *Vina*; cinco análisis con diferentes valores de exhaustividad: 8, 16, 32, 64 y 128, y con una caja de 40*40*40 Å .

Los resultados de este análisis se presentan en la Fig. A.4b, en donde se muestran, para cada evaluación, la media y la desviación estándar de los valores de energía libre de interacción calculados entre la proteína E6 y la curcumina en sus 20 posiciones iniciales. Idealmente, se esperaría que a mayor exhuastividad o número de evaluaciones aumente la probabilidad de encontrar energías de interacción más favorables. Es decir, mínimos locales de energía más bajos y menos dispersos entre los ensayos. No obstante, para *AD4* sólo se logró observar una disminución en la media de valores de energía de interacción pero no en su dispersión. Por su parte, *Vina* sí mostró ambos comportamientos, teniendo valores de energía más bajos y menos dispersos a mayor exhaustividad. La diferencia entre los valores de energía de ex8 y ex128 no es tan considerable, aunque éste último es 16 veces más costoso en tiempo computacional.

También se realizó una SAM utilizando las 26 moléculas obtendias de la bibliografía. Para ello se utilizó un tamaño de caja 40*40*40 Å y los valores de evaluaciones y exhaustividad mostrados en la Fig. A.4c. En estos resultados es posible observar un comportamiento similar al visto en el análisis con curcumina. Sin embargo, aquí es más evidente que *AD4* es menos eficiente a mayor número de rotámeros, para los cuales requiere valores más altos de evaluaciones para encontrar mínimos de energía más bajos. Esto es señalado por la propia documentación de *AD4* (Morris y Huey, 2009). Por su parte *Vina* muestra un comportamiento muy similar en todas las evaluaciones de exhaustividad. Es interesante notar también que, al menos en los ligandos más pequeños, los valores de energía entre *AD4* y *Vina* son similares. El comportamiento mostrado por *AD4* y *Vina* ha sido previamente señalado por trabajos más robustos como el de (Jaghoori *et al.*, 2016; Chang *et al.*, 2010). Sin embargo, este análisis ha permitido determinar mejor el número de evaluaciones o de exhaustividad a utilizar para *Vina* y *AD4* en la fase de CV basado en SAM. Además, se ha elegido comenzar esta fase con *Vina*, pues resulta ser eficiente aún con valores bajos de exhaustividad, permitiendo evaluar un gran número de ligandos. Sin embargo, en algunos ligandos *AD4* fue capaz de encontrar acoplamientos más favorables, por lo que se decidió implementarlo como filtro adicional al CV con *Vina*, posibilitando el hallazgo de conformaciones de acoplamiento aún más favorables.

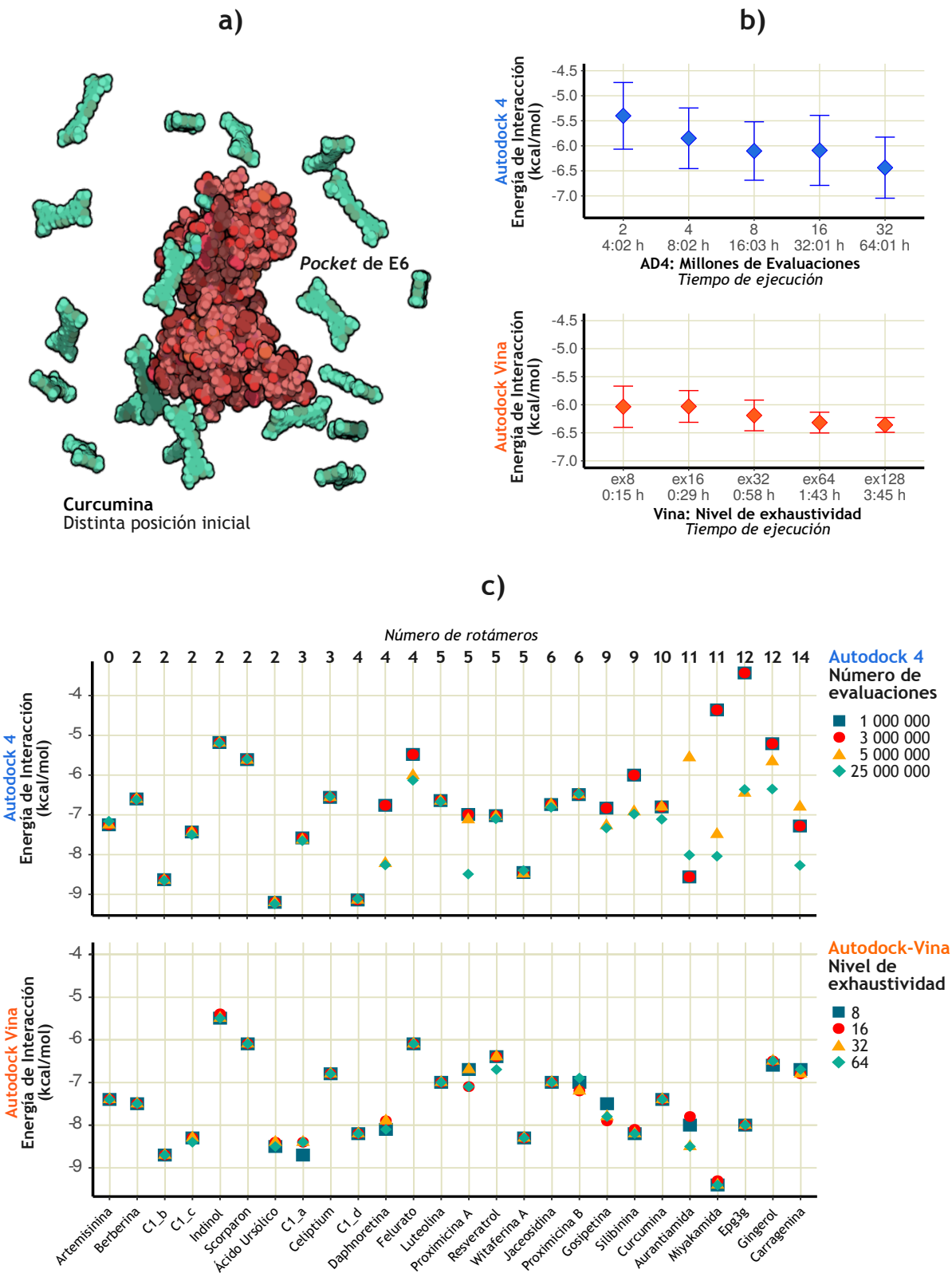


Figura A.4: Evaluación de *Autodock 4* y *Autodock Vina*. a) Esquema que muestra las posiciones iniciales de la molécula curcumina (verde) con respecto a la proteína E6 (rojo). **b)** Resultados del análisis de SAM con *AD4* (arriba) y *Vina* (abajo) con la molécula curcumina y la proteína E6. Estas posiciones iniciales del ligando fueron elegidas de forma aleatoria. **c)** Resultados del análisis de SAM con *AD4* (arriba) y *Vina* (abajo) con las 26 moléculas de referencia y utilizando diferentes números de evaluaciones o niveles de exhaustividad. Las moléculas en el eje *x* están ordenadas de izquierda a derecha por número de rotámeros, y este valor se muestra en la parte superior del gráfico.

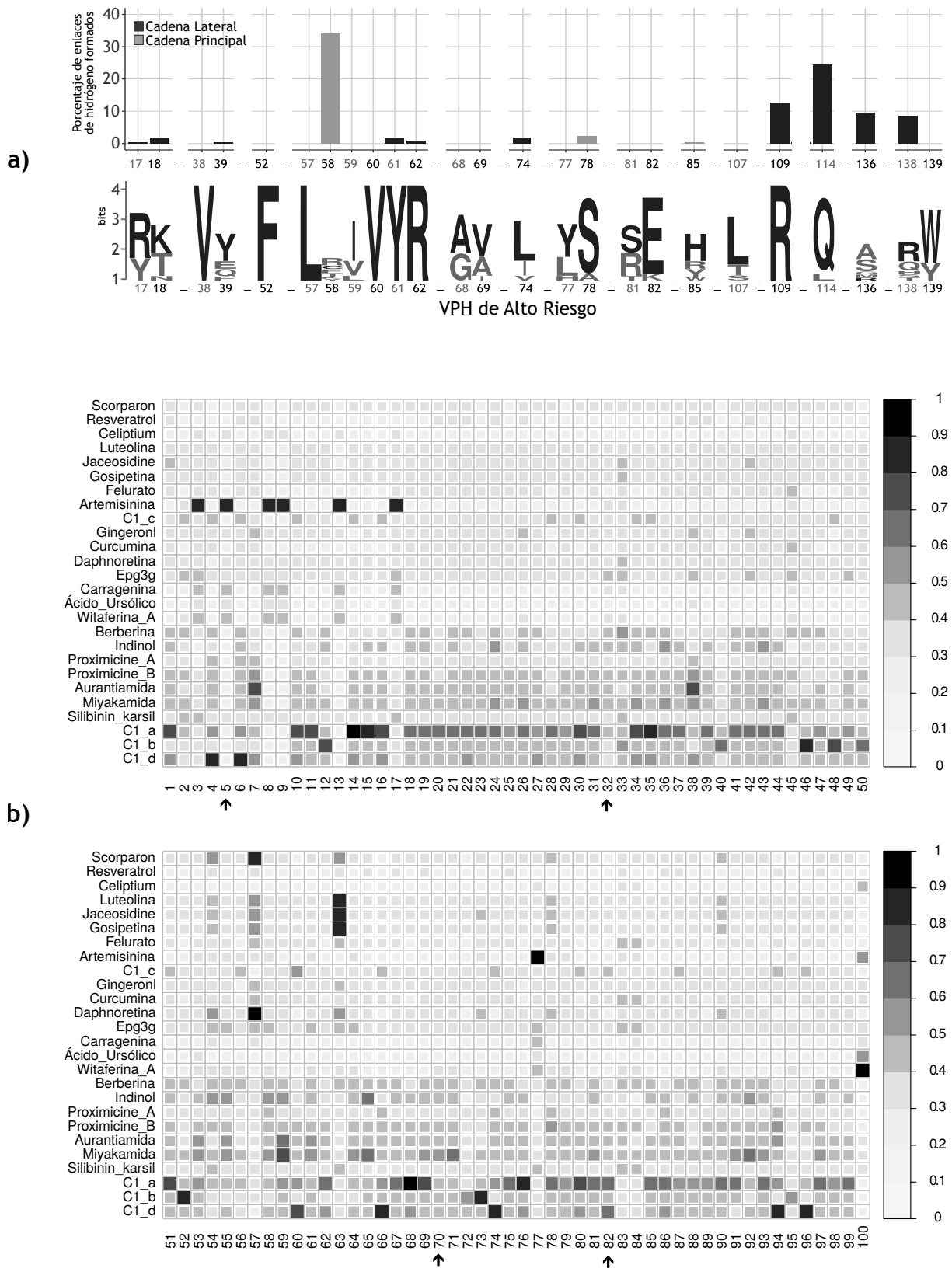


Figura A.5: Análisis de los mejores 100 ligandos identificados con AD4. a) Número de puentes de hidrógeno en el que cada residuo del *pocket* de E6 participa al considerar los acoplamientos de E6 con los 100 ligandos. **b)** Coeficiente de Tanimoto calculado entre los 100 ligandos y las 26 moléculas iniciales. Los valores más altos se representan por casillas más oscuras, las cuales indican el parecido estructural entre los mejores 100 candidatos y los compuestos iniciales. Los cinco ligandos elegidos para validación por dinámica molecular son señalados por las flechas inferiores.

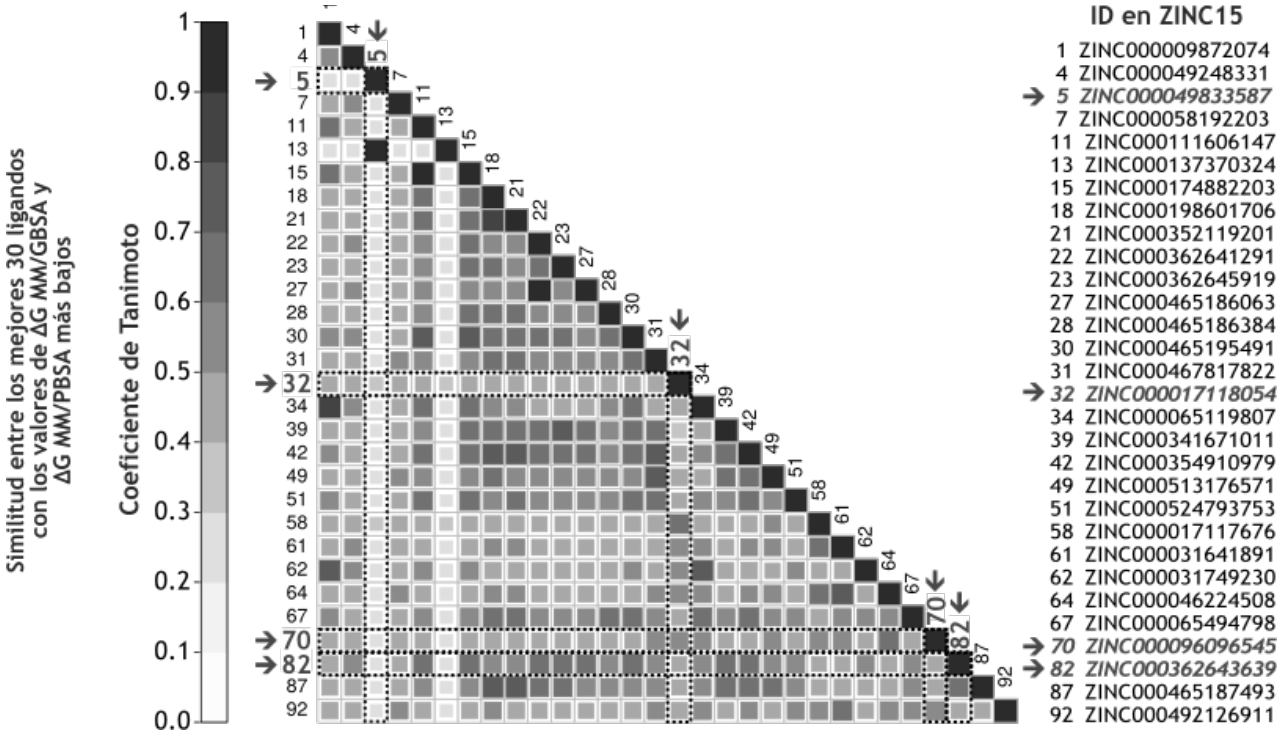


Figura A.6: Gráfico de Tanimoto mostrando el grado de similitud entre las mejores 30 moléculas. Gráfico de Tanimoto mostrando el grado de similitud entre los ligandos. Para efectos de visualización sólo se muestran los 30 mejores ligandos en este análisis, los cuales corresponden a aquellos cuyos valores de energía de interacción en cada método son mayores a la media correspondiente: $\Delta\overline{\mathrm{G}}^{GB}$ = -41.5 kcal/mol y $\Delta\overline{\mathrm{G}}^{PB}$ = 4.0 kcal/mol. Los cinco ligandos elegidos para validación por dinámica molecular se muestran en letra itálica. El ligando 5 mostró una baja similitud con el resto de los ligandos y buenos valores de $\Delta\mathrm{G}^{GB}$ y $\Delta\mathrm{G}^{PB}$, motivo por el cual se decidió también analizarlo.

A.7. Dinámica molecular de los complejos E6-hx, E6 VPH-16 y E6 VPH-11

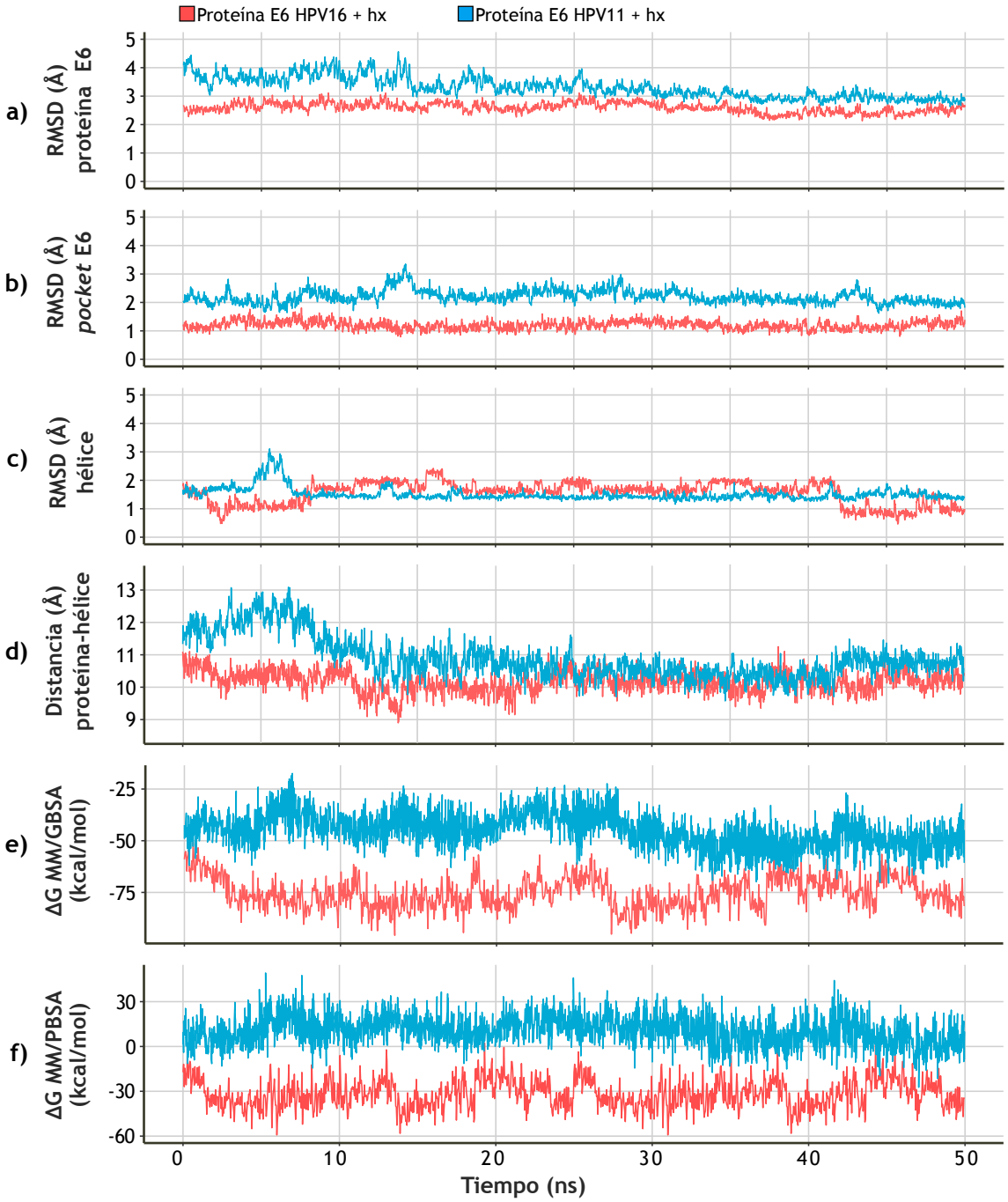


Figura A.7: Dinámica molecular de los complejos E6-hx. a) Valores de RMSD de la cadena principal de la secuencia total de la proteína E6 HPV16 (rojo) y de la proteína E6 HPV11 (azul). **b)** RMSD del *pocket* de E6 HPV16 y de E6 HPV11. **c)** RMSD de la hélice *LxxLL* en interacción con E6 HPV16 y con E6 HPV11. **d)** Distancia entre el centro de masa de la hélice y la proteína E6. **e)** Valores de ΔG^{GB} de la interacción proteína-hélice. Para E6 HPV16 el ΔG^{GB} promedio fue de -75.31 kcal/mol con una desviación estándar (σ) de 7.63 kcal/mol. Para E6 HPV11 el promedio fue de -44.92 kcal/mol con σ= 8.69 kcal/mol. **f)** Valores de ΔG^{PB} de la interacción proteína-hélice. Para E6 HPV16 el ΔG^{PB} promedio fue de -31.76 kcal/mol σ= 9.95 kcal/mol. Para E6 HPV11 tuvo un promedio fue de 10.47 kcal/mol con σ= 10.15 kcal/mol.

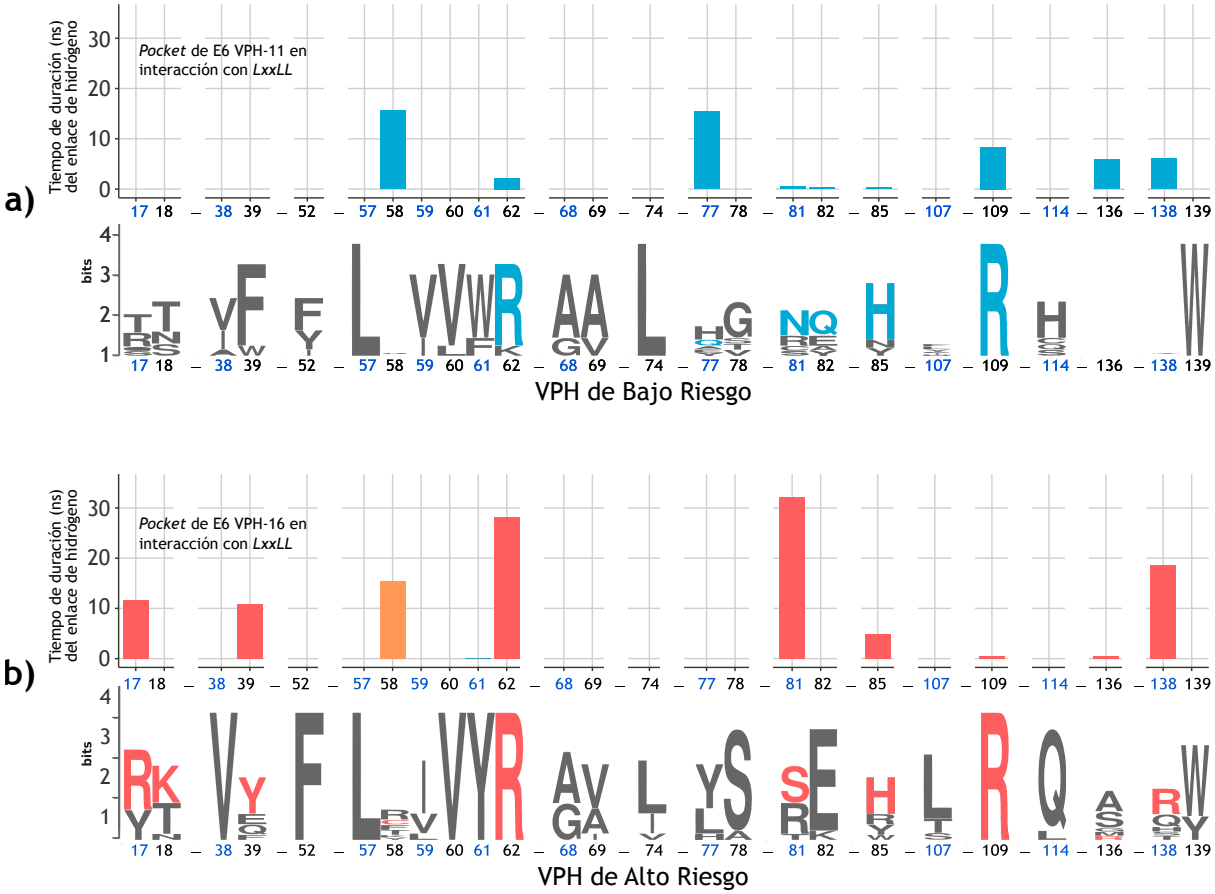


Figura A.8: Enlaces de hidrógeno entre el *pocket* de la proteínas E6 VPH-16 y E6 VPH-11 con la hélice *LxxLL*. Se muestran los residuos de E6 VPH-16 (en rojo) y de E6 VPH-11 (en azul) utilizando parte de la información del logo de secuencia de la Fig. 7. Lo que permite identificar los sitios conservados en los tipos AR y BR que participan en la interacción con la hélice de E6. Cada barra representa el tiempo de interacción por enlace de hidrógeno de cada residuo del *pocket* en cada ensayo de los complejos *E6 VPH-16+hx* y *E6 VPH-11+hx* a lo largo de 50 ns. Resulta interesante observar que la proteína E6 VPH-16 posee más enlaces de hidrógeno con la hélice que la proteína E6 VPH-11.

A.8. Dinámica molecular de los complejos E6-lig

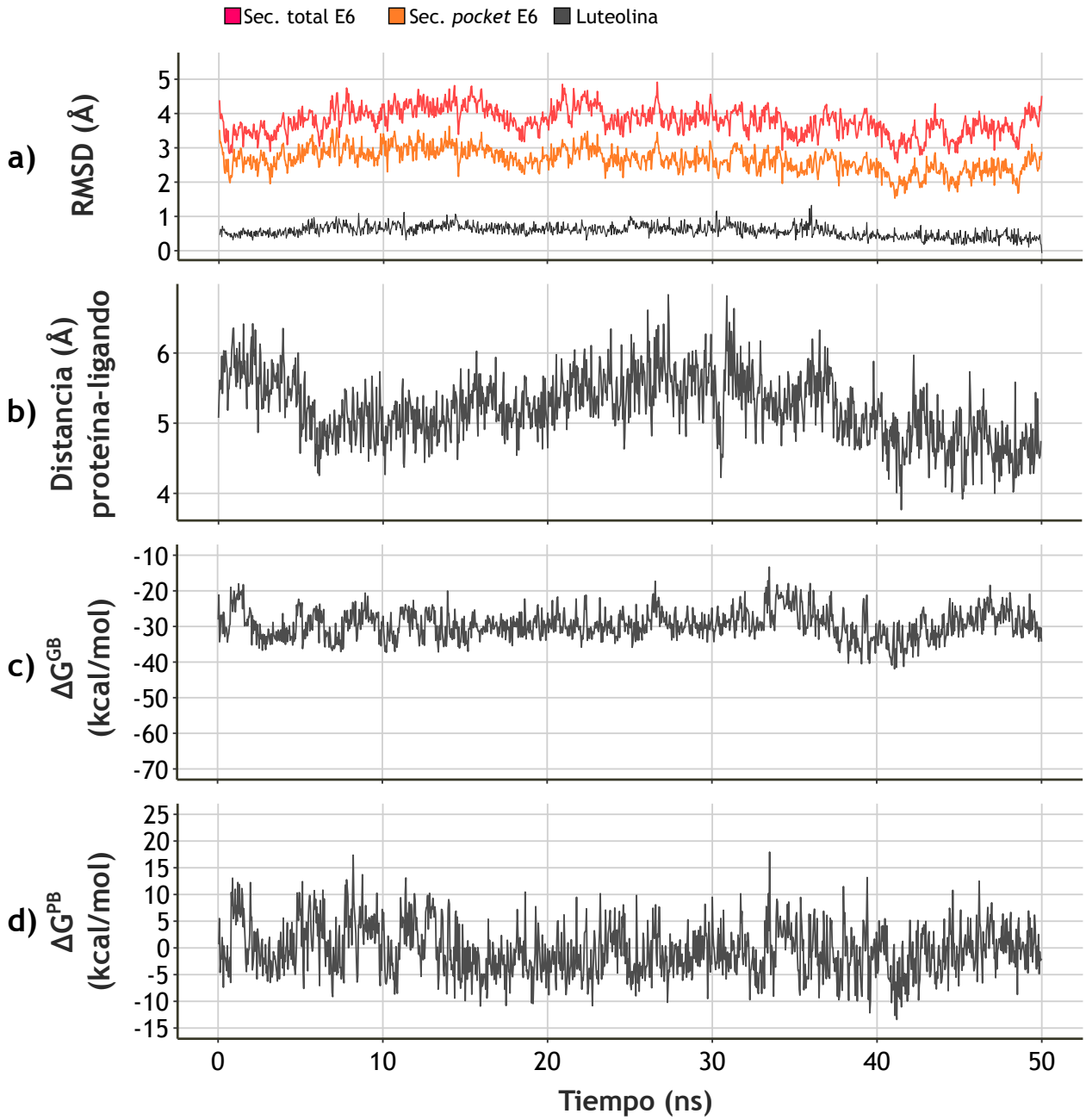


Figura A.9: Dinámica molecular del complejo *E6-Luteolina* (50 ns). a) Valores de RMSD de la cadena principal de la secuencia total de la proteína E6 (rojo), la subsecuencia del *pocket* (naranja) y la molécula Luteolina (gris). **b)** Distancia entre el centro de masa de E6 y el centro de masa de la Luteolina, con un valor promedio de 5.22 Å y una desviación estándar de 0.47 Å **c)** Valores de ΔG^{GB} de la interacción proteína-ligando calculados a lo largo de la trayectoria, con un valor mínimo de -41.90 kcal/mol, un máximo de -13.31 kcal/mol, un promedio de -29.45 kcal/mol y una desviación estándar de 4.08 kcal/mol. **d)** Valores de ΔG^{PB} de la interacción proteína-ligando, con un valor mínimo de -13.40 kcal/mol, un máximo de 17.93 kcal/mol, un promedio de -0.39 kcal/mol y una desviación estándar de 4.8 kcal/mol.

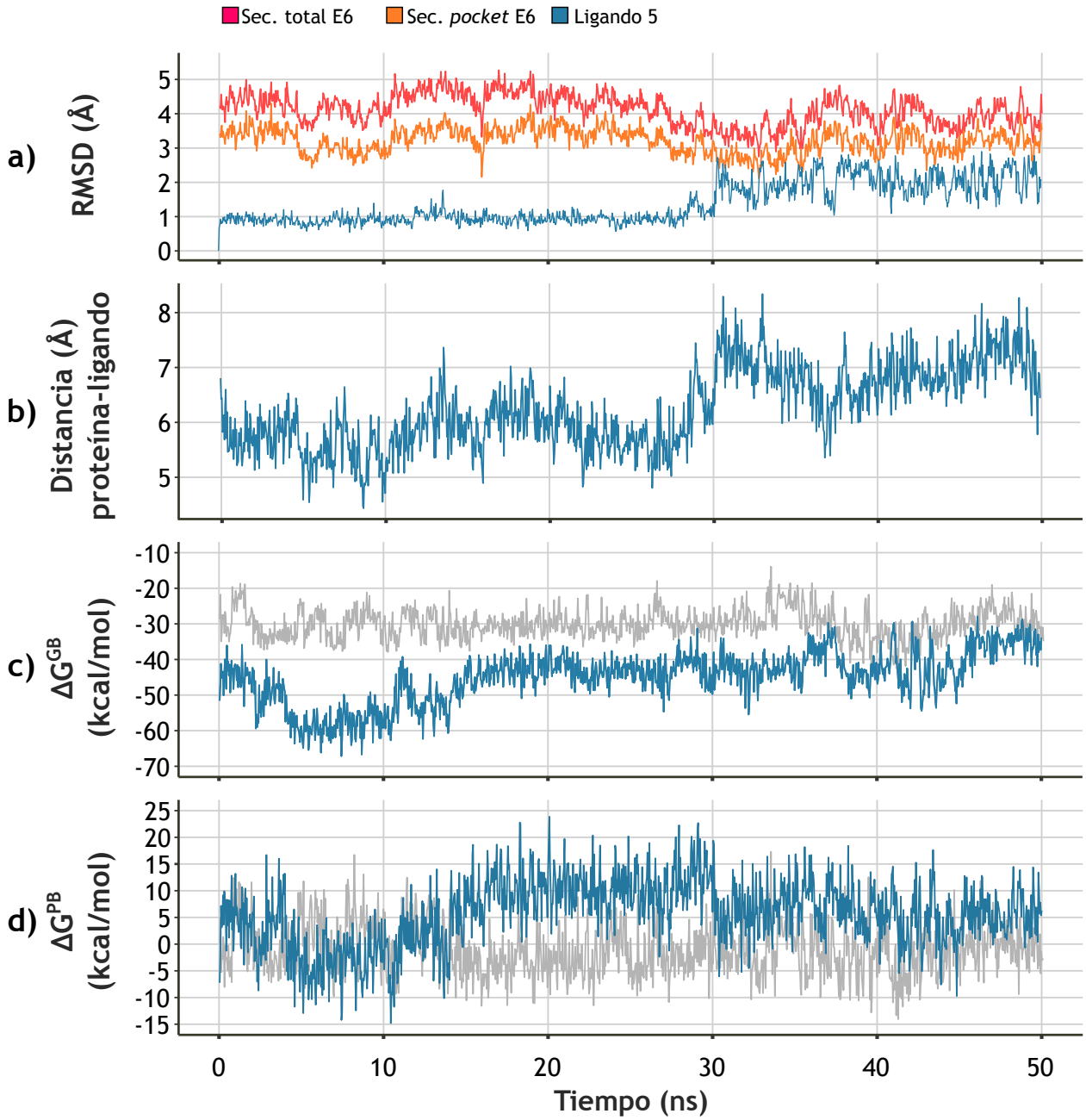


Figura A.10: Dinámica molecular del complejo *E6-Lig5* (50 ns). a) Valores de RMSD de la cadena principal de la secuencia total de la proteína E6 (rojo), la subsecuencia del *pocket* (naranja) y el ligando 5 (azul). **b)** Distancia entre el centro de masa de E6 y el centro de masa del ligando 5, con un valor promedio de 6.24 Å y una desviación estándar de 0.72 Å **c)** Valores de ΔG^{GB} de la interacción proteína-ligando calculados a lo largo de la trayectoria, con un valor mínimo de -67.14 kcal/mol, un máximo de -27.89 kcal/mol, un promedio de -45.07 kcal/mol y una desviación estándar de 7.38 kcal/mol. **d)** Valores de ΔG^{PB} de la interacción proteína-ligando, con un valor mínimo de -14.79 kcal/mol, un máximo de 23.86 kcal/mol, un promedio de 5.62 kcal/mol y una desviación estándar de 6.38 kcal/mol. A modo de referencia se muestra en color gris el resultado obtenido con la molécula control Luteolina (Fig. A.9).

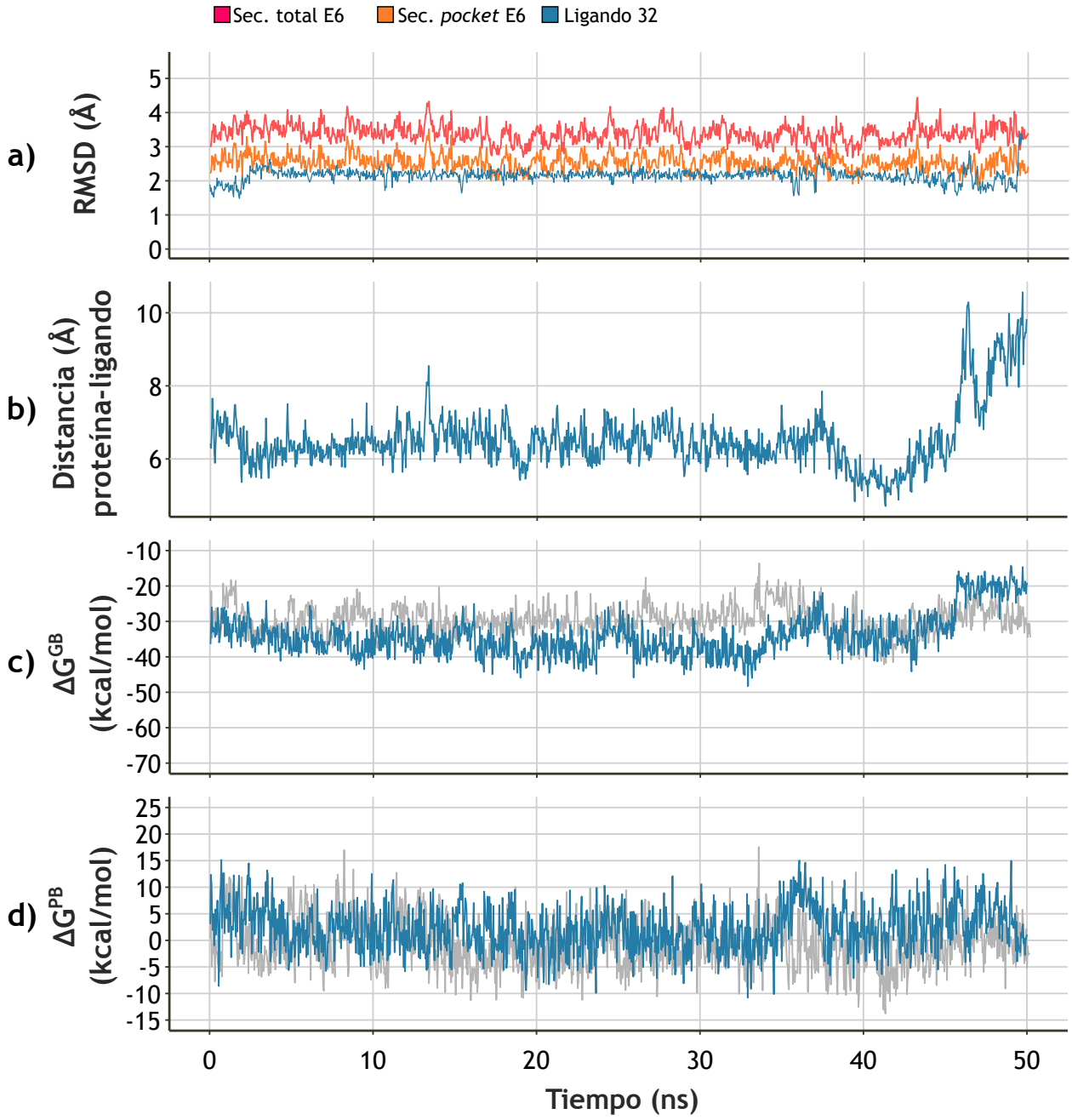


Figura A.11: Dinámica molecular del complejo *E6-Lig32* (50 ns). a) Valores de RMSD de la cadena principal de la secuencia total de la proteína E6 (rojo), la subsecuencia del *pocket* (naranja) y el ligando 32 (azul). **b)** Distancia entre el centro de masa de E6 y el centro de masa del ligando 32, con un valor promedio de 6.54 Å y una desviación estándar de 0.84 Å **c)** Valores de ΔG^{GB} de la interacción proteína-ligando calculados a lo largo de la trayectoria, con un valor mínimo de -48.30 kcal/mol, un máximo de -14.23 kcal/mol, un promedio de -33.97 kcal/mol y una desviación estándar de 5.92 kcal/mol. **d)** Valores de ΔG^{PB} de la interacción proteína-ligando, con un valor mínimo de -10.80 kcal/mol, un máximo de 15.14 kcal/mol, un promedio de 2.19 kcal/mol y una desviación estándar de 4.38 kcal/mol. A modo de referencia se muestra en color gris el resultado obtenido con la molécula control Luteolina (Fig. A.9).

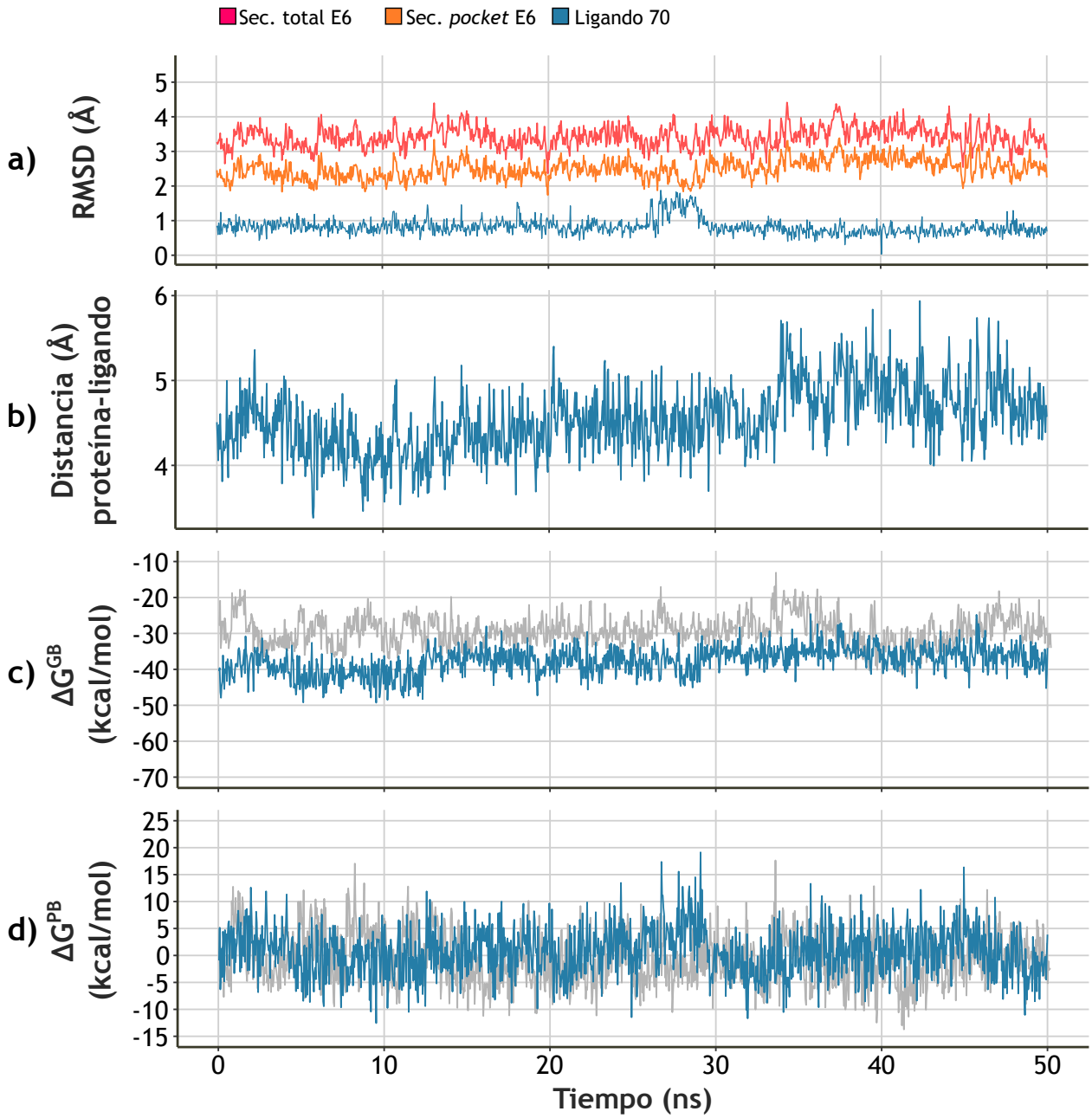


Figura A.12: Dinámica molecular del complejo *E6-Lig70* (50 ns). a) Valores de RMSD de la cadena principal de la secuencia total de la proteína E6 (rojo), la subsecuencia del *pocket* (naranja) y el ligando 70 (azul). **b)** Distancia entre el centro de masa de E6 y el centro de masa del ligando 70, con un valor promedio de 4.54 Å y una desviación estándar de 0.39 Å **c)** Valores de ΔG^{GB} de la interacción proteína-ligando calculados a lo largo de la trayectoria, con un valor mínimo de -49.28 kcal/mol, un máximo de -24.59 kcal/mol, un promedio de -37.65 kcal/mol y una desviación estándar de 3.77 kcal/mol. **d)** Valores de ΔG^{PB} de la interacción proteína-ligando, con un valor mínimo de -12.55 kcal/mol, un máximo de 19.13 kcal/mol, un promedio de 0.62 kcal/mol y una desviación estándar de 4.53 kcal/mol. A modo de referencia se muestra en color gris el resultado obtenido con la molécula control Luteolina (Fig. A.9).

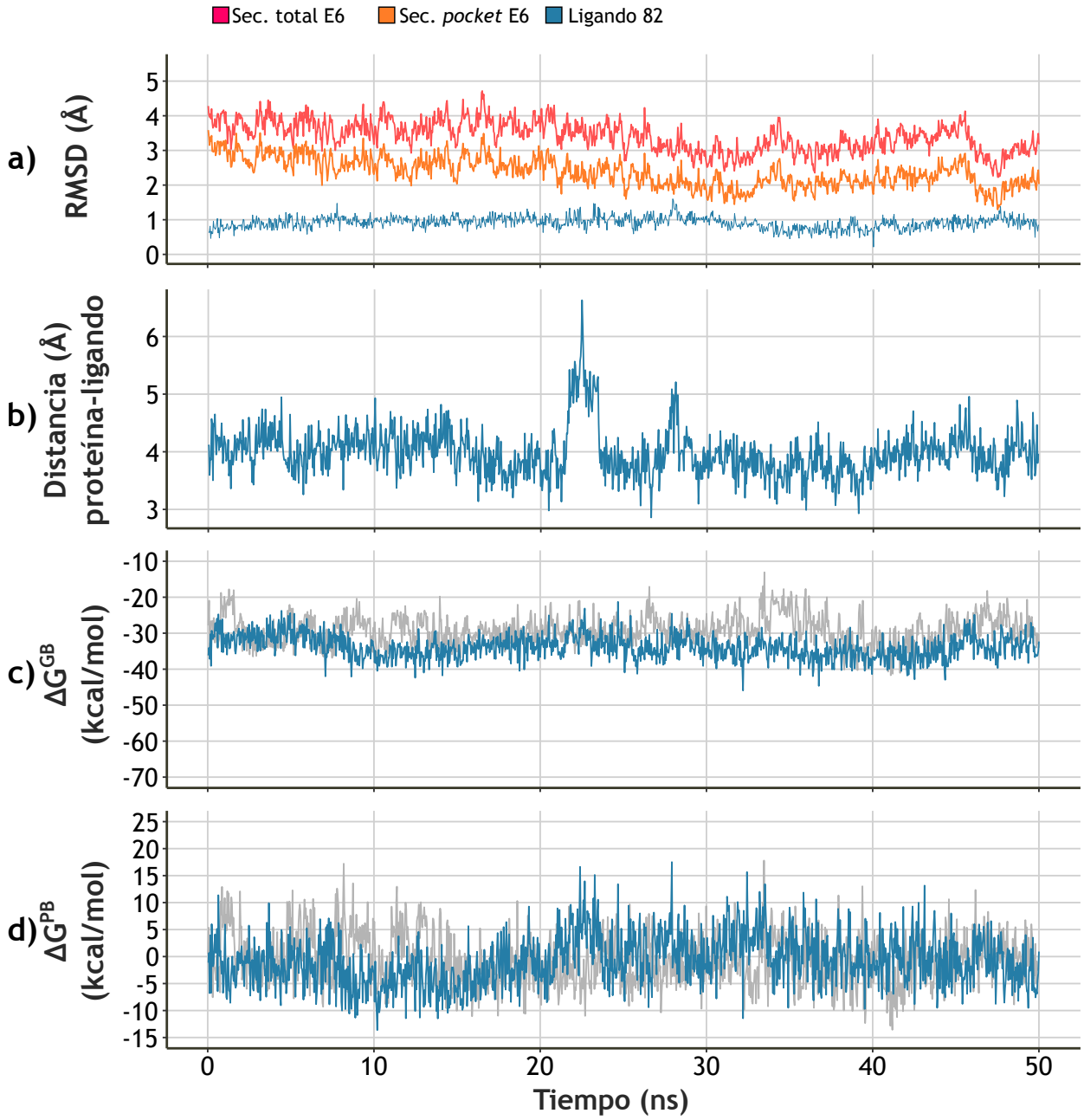


Figura A.13: Dinámica molecular del complejo *E6-Lig82* (50 ns). a) Valores de RMSD de la cadena principal de la secuencia total de la proteína E6 (rojo), la subsecuencia del *pocket* (naranja) y el ligando 82 (azul). **b)** Distancia entre el centro de masa de E6 y el centro de masa del ligando 82, con un valor promedio de 3.99 Å y una desviación estándar de 0.41 Å **c)** Valores de ΔG^{GB} de la interacción proteína-ligando calculados a lo largo de la trayectoria, con un valor mínimo de -42.92 kcal/mol, un máximo de -21.29 kcal/mol, un promedio de -34.84 kcal/mol y una desviación estándar de 3.32 kcal/mol. **d)** Valores de ΔG^{PB} de la interacción proteína-ligando, con un valor mínimo de -13.62 kcal/mol, un máximo de 17.50 kcal/mol, un promedio de -0.77 kcal/mol y una desviación estándar de 4.74 kcal/mol. A modo de referencia se muestra en color gris el resultado obtenido con la molécula control Luteolina (Fig. A.9).