

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



**Maestría en Ciencias
en Nanociencias**

Determinación del potencial oncogénico de algunas especies de
VPH mediante el análisis *in silico* de la interacción entre las
proteínas E6 y E6AP

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

Presenta:

Lizt Selene Sibila Osorio Pando

Ensenada, Baja California, México
2018

Tesis defendida por

Lizt Selene Sibila Osorio Pando

y aprobada por el siguiente Comité

Dr. Sergio Andrés Águila Puentes

Codirector de tesis

Dr. Abraham Marcelino Vidal Limón

Codirector de tesis

Miembros del comité

Dr. Carlos Alberto Brizuela Rodríguez

Dr. María Guadalupe Moreno Armenta



Dr. Sergio Fuentes Moyado

Coordinador del Posgrado en Nanociencias

Dra. Rufina Hernández Martínez

Directora de Estudios de Posgrado

Lizt Selene Sibila Osorio Pando © 2018

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

Resumen de la tesis que presenta **Lizt Selene Sibila Osorio Pando** como requisito parcial para la obtención del grado de Maestro en Ciencias en Nanociencias

Determinación del potencial oncogénico de algunas especies de VPH mediante el análisis in silico de la interacción entre las proteínas E6 y E6AP

Resumen aprobado por:

Dr. Sergio Andrés Águila Puentes
Codirector de tesis

Dr. Abraham Marcelino Vidal Limón
Codirector de tesis

A la fecha, se han reportado más de 200 tipos de papilomavirus que pueden infectar el tejido epitelial en humanos, estos son llamados, virus de papiloma humano (VPH). A su vez, doce de ellos pueden llevar a la célula infectada hacia un fenotipo cancerígeno, a estos tipos se les conoce como de alto riesgo. Una de las características que determina si un VPH es o no cancerígeno, es la presencia de dos proteínas, la E6 y E7. La E6 es una proteína capaz de interactuar con la proteína celular E6AP, causando un cambio conformacional en E6 que le permite el secuestro y degradación de la proteína p53, como consecuencia, la célula no puede entrar en apoptosis. La mayoría de las investigaciones que buscan dilucidar la función de ésta proteína han sido realizadas de forma *in vitro* e *in vivo*, lo que incrementa, por mucho el costo y el tiempo para la obtención de resultados. Sin embargo, los métodos de química computacional o *in silico*, ofrecen resultados a una fracción del costo y en un menor tiempo. A la fecha, los métodos *in silico* han sido poco aplicados al estudio de la proteína E6, debido principalmente a que no se contaba con la estructura cristalina, no obstante, en los años 2013 y 2016 se reportó dicha estructura. El objetivo de este trabajo, es estudiar por métodos *in silico* la interacción de la proteína E6, tanto de tipos de VPH de alto y bajo riesgo, y la proteína E6AP; con la intención de determinar si existe una correlación entre la energía libre de interacción teórica de cada complejo E6/E6AP, y la capacidad cancerígena del virus. Para ello, se utilizaron modelos estructurales de: la proteína E6, obtenidos por modelado por enhebrado, de 24 tipos de VPH y el motivo proteico de la E6AP implicado en la interacción (LxxLL). El cálculo de la energía libre de interacción se realizó a partir simulaciones por dinámica molecular utilizando los métodos MM/GBSA y MM/PBSA incluidos en el paquete de programas Amber16. Los resultados indicaron que los valores de energía libre de interacción entre la proteína E6 y el motivo LxxLL de los VPH de bajo riesgo son muy parecidos a los valores que se obtuvo para algunos tipos de alto riesgo, éste resultado se obtuvo con ambos métodos de cálculo (MM/GBSA y MM/PBSA). Lo que indica, que la capacidad de interacción entre las proteínas E6 y el motivo LxxLL, no es determinante en el potencial cancerígeno de los VPH, al menos en los tipos correspondientes a la especie Alfa y bajo las condiciones experimentales desarrolladas.

Palabras clave: Proteína E6; Energía Libre de Interacción; Virus de Papiloma Humano; Dinámica molecular; MM/GBSA; MM/PBSA.

Abstract of the thesis presented by **Lizt Selene Sibila Osorio Pando** as a partial requirement to obtain the Master of Science degree in Nanoscience

Oncogenic potential determination of some HPV species through an *in silico* analysis of the interaction between E6 and E6AP proteins.

Abstract approved by:

Dr. Sergio Andrés Águila Puentes
Codirector de tesis

Dr. Abraham Marcelino Vidal Limon
Codirector de tesis

To date, more than 200 species of papillomaviruses have been reported with the ability to infect epithelial tissue in humans, they are called human papillomavirus (HPV). In addition, there are some species, 12 reported so far, with the ability to infect mucous epithelial tissue, e.g. the anogenital areas, and lead the cell towards a carcinogenic phenotype. While the onset of cellular transformation still remains unclear, the presence of two proteins, E6 and E7, is determinant of HPV carcinogenic potential. E6 protein can interact with a cellular protein called E6AP, which leads to a conformational change in E6 that allows the hijacking and degradation of the p53 protein; consequently, the cell cannot enter apoptosis. Most of the investigations on the function of E6 are carried out *in vitro* and *in vivo*, which increases, by far, the cost and time to obtain results. Nowadays, computational simulations or *in silico* methods offer an attractive alternative for faster and cheaper screening of interacting proteins than *in vitro* tests. Recently, a research group reported the crystallographic structure of the E6-E6AP complex, opening a new alternative to study molecular interactions between these regulatory proteins. The aim of this work is to study, by *in silico* methods, the interaction capacity of the E6 protein, from both oncogenic or non-HPV species and the E6AP protein to determine if there is a correlation between the interaction energy of each E6/E6AP complex and the carcinogenic capacity of the virus. To achieve this goal, structural models of E6 proteins were obtained by threading modeling of 24 types of HPV types and the interaction motif of the E6AP (LxxLL). End-point free energy calculations were performed, the calculation of the interaction-free energy was performed based on molecular dynamics simulations and the MM / GBSA and MM / PBSA methods included in the Amber16 program package. The results indicated that overall free energy of interaction between the E6 protein and the motif LxxLL from low risk HPV are very similar to the values obtained for some types of high risk, this result was obtained with both methods of calculation (MM / GBSA and MM / PBSA). Our results indicate that the interaction capacity between E6 proteins and the LxxLL motif is not determinant in the carcinogenic potential of HPV, at least in the types corresponding to the alpha species and under the experimental conditions developed.

Keywords: E6 protein; Free energy calculations; Human Papillomavirus; Molecular dynamics; MM/GBSA; MM/PBSA.

Dedicatoria

A todas aquellas personas que por desinformación o taboo padecen de cáncer a causa de la infección con VPH.

Agradecimientos

Al Centro de Investigación Científica y Educación Superior de Ensenada (CICESE) y al Centro de Nanociencias y Nanotecnología (CNYN), por ser las instituciones que me formaron académicamente durante la maestría.

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. Abraham Limón, por recibirme y darme la oportunidad de formar parte de su grupo de investigación.

A todos los integrantes del grupo de investigación por hacer de éste viaje más cómodo.

A todos los integrantes del LEP, que siguen siendo más que mis amigos mi familia.

A mi familia (papi, mami y las niñas) por el apoyo dado desde la distancia. Ellos saben que todo sale bien al final.

A Dani (el güero) y Aldo por hacer de mi estancia en el CNYN más interesante, amo nuestras pláticas.

A Mayra por el amor de amiga, al mejor estilo venezolano, que recibí de tí.

A Oscar por todo su apoyo durante mi estancia en México.

A todas las personas que conocí en la roca y el oceano, cuya fuerza me guió en el camino.

Y

A Emmanuel... No se que habria sido de mi sin tí.

Tabla de contenido

	Página
Resumen en español.....	ii
Resumen en inglés.....	iii
Dedicatorias.....	iv
Agradecimientos.....	v
Lista de figuras.....	ix
Lista de tablas.....	xiii
Capítulo 1. Introducción.....	1
Capítulo 2. Antecedentes.....	3
2.1. Biología y clasificación de los papilomavirus (PVs).....	3
2.2. Ciclo reproductivo del VPH.....	4
2.3. Proteínas E6 y E7 y su papel en la transformación celular a un fenotipo cancerígeno.....	5
2.4. Estudios de estructura-función de la proteína E6.....	7
2.5. Dinámica molecular (DM) en sistemas biológicos.....	9
Capítulo 3. Hipótesis y Objetivos.....	10
3.1. Hipótesis.....	10
3.2. Objetivos.....	10
3.2.1. Objetivo general	10
3.2.2. Objetivos específicos.....	10
Capítulo 4. Metodología.....	11
4.1. Elección y análisis de las secuencias de la proteína E6 a estudiar.....	11
4.1.1. Elección de las secuencias de E6.....	11
4.1.2. Alineamiento y construcción de fenogramas	11
4.1.3. Análisis de la estructura primaria de las E6.....	11
4.2. Modelado por enhebrado de las E6.....	12

4.2.1. Modelado por enhebrado con I-TASSER.....	12
4.3. Modelado por dinámica molecular del complejo E6-Zn/LxxLL.....	12
4.3.1. Preparación del complejo E6-Zn/LxxLL.....	12
4.3.2. Minimización de la energía.....	13
4.3.3. Recocido simulado y equilibrado.....	13
4.3.4. Evaluación del equilibrio.....	14
4.3.5. Simulación por dinámica molecular.....	15
4.3.6. Análisis de las trayectorias obtenidas durante la DM.....	15
4.3.6.1. RMSD (root-mean-square deviation).....	15
4.3.6.2. RMSF (root-mean-square fluctuation).....	15
4.3.6.3. Enlaces de hidrógeno.....	16
4.3.6.4. Contactos nativos.....	16
4.3.6.5. Distancia entre los centros de masa.....	17
4.3.6.6. Análisis por distribución de Boltzmann.....	17
4.4. Cálculo de la energía libre de interacción entre la proteína E6 y el motivo LxxLL.....	18
4.4.1. Métodos de MM-GBSA y MM-PBSA.....	18
Capítulo 5. Resultados.....	20
5.1. Elección y análisis de las secuencias de la proteína E6 a estudiar.....	20
5.1.1. Elección de las secuencias de E6.....	20
5.1.2. Alineamiento y construcción de fenogramas	20
5.1.3. Análisis de la estructura primaria de las E6.....	24
5.2. Modelado por enhebrado de las E6.....	28
4.2.1. Modelado por enhebrado con I-TASSER.....	28
5.3. Modelado por dinámica molecular del complejo E6-Zn/LxxLL.....	29
5.3.1. Preparación del complejo E6-Zn/LxxLL.....	29
5.3.2. Minimización de la energía.....	29
5.3.3. Recocido simulado y equilibrado.....	30
5.3.4. Evaluación del equilibrio.....	31
5.3.5. Simulación por dinámica molecular.....	34
5.3.6. Análisis de las trayectorias obtenidas durante la DM.....	35
5.3.6.1. RMSD (root-mean-square deviation).....	35
5.3.6.2. RMSF (root-mean-square fluctuation).....	35

5.3.6.3. Enlaces de hidrógeno.....	38
5.3.6.4. Contactos nativos.....	39
5.3.6.5. Distancia entre los centros de masa.....	39
5.3.6.6. Análisis por distribución de Boltzmann.....	40
5.4. Cálculo de la energía libre de interacción entre la proteína E6 y el motivo LxxLL.....	40
Capítulo 6. Discusión.....	44
6.1. Elección y análisis de las secuencias de la proteína E6 a estudiar.....	44
6.2. Modelado por enhebrado de las E6.....	47
6.3. Modelado por dinámica molecular del complejo E6-Zn/LxxLL.....	49
6.3.1. Preparación del complejo E6/LxxLL y DM.....	49
6.3.2. Análisis de las trayectorias obtenidas en la fase de DM.....	49
6.4. Cálculo de la energía libre de interacción entre la proteína E6 y el motivo LxxLL.....	50
Capítulo 7. Conclusiones.....	55
Capítulo 8. Perspectivas.....	56
Literatura citada.....	57
Anexos.....	68

Lista de figuras

Figura	Página
1 Organización del genoma de algunos papilomavirus. El genoma de lo Pvs es circular de DNA doble hebra y se divide en tres regiones: early (E: por su sigla en inglés), late (L: por su sigla en inglés) y long codon region (LCR: por sus siglas en inglés).....	3
2 Función de la oncoproteína E6 en un VPH de alto riesgo. Proteínas celulares blanco de la E6 y su actividad asociada.....	7
3 Red de interacciones entre la proteína E6 de VPH16 y el motivo LxxLL. La información fue obtenida a partir de la estructura cristalina del complejo E6/LxxLL. Panel izquierdo: representación de las interacciones polares entre aminoácidos; la estructura en verde se refiere al motivo LxxLL, el resto a la proteína E6, la estructura en amarillo, son los aminoácidos de la región cercana al C-terminal y los violeta al N-terminal. Panel derecho: Representación de todos los contactos entre E6 y LxxLL; las líneas punteadas de color rosa representan interacciones de tipo hidrofóbicas, las líneas negras a los contactos polares, medidas por la cadena lateral (línea continua) y por la cadena principal (línea punteada).....	8
4 Alineamiento de la proteína E6 de diferentes tipos de VPH. Las secuencias están ordenadas en orden descendente según especie y capacidad cancerígena. El porcentaje de identidad se muestra en escala de intensidad de azules, azul intenso (identidad > 80%), azul medianamente intenso (entre 60 y 80 %) azul claro (entre 60 y 40 %) y blanco (< al 40%). La primera secuencia corresponde a la proteína E6 del cristal del complejo E6/LxxLL (codigo 4GIZ en el PDB), en ella se resaltan los aminoácidos de cisteínas indispensables para la formación del complejo E6/LxxLL (verde) y la posterior degradación de la proteína p53, además de las cistenías mutadas por serinas dispensable para la formación del complejo y degradación de p53 (Zanier et al., 2012), los residuos enmarcados en negro corresponde a los aminoácidos implicados en la interacción de E6 y LxxLL (Zanier et al., 2013).....	22
5 Fenograma de la proteína E6 de diferentes tipos de VPH. Haciendo uso del alineamiento con MUSCLE se construyó un fenograma mediante el método Neighbor Joining (NJ) y un Bootstrap de 1000. Las especies resultantes se muestran con diferentes tonalidades de gris. Los integrantes del grupo 1 (categorías IARC) de alto riesgo están escritos con letras amarillas, grupo 2A violeta, grupo 2B aguamarina y grupo 3 de bajo riesgo en azul. El bootstrap está representado con 3 tamaños diferentes de círculo: 85-100 % (círculo grande), 70-84% (círculo mediano), 50-69% (círculo pequeño). La longitud de la barra indica la p-distance.....	24
6 Representación gráfica de la diferencia en composición de aminoácidos de la proteína E6 para cada tipo de VPH. (A) porcentaje de los 20 aminoácidos a lo largo de toda la secuencia de cada E6, (B) porcentaje de los 20 aminoácidos en la región comprendida por los aminoácidos 17 y 146 obtenida a partir del alineamiento (Figura 4) de cada E6. Los integrantes del grupo 1 (categorías IARC) de alto riesgo están representados con líneas amarillas, grupo 2A violeta, grupo 2B aguamarina, grupo 3 de bajo riesgo en azul y el grupo control en naranja.....	26

7	Representación gráfica de la estructura secundaria determinada por el programa Jpred de la proteína E6 para cada tipo de VPH. Las secuencias están ordenadas de forma descendente, en cuanto a la capacidad oncogénica según las categorías IARC (de la misma forma que en la figura 4), las cisteínas implicadas en la unión a átomos de zinc están marcadas con el símbolo (*), los aminoácidos sombreados en blanco y negro corresponden a los aminoácidos implicados en la interacción de E6 y LxxLL (Zanier et al., 2013). Panel inferior: logo de los aminoácidos implicados en la interacción: polares positivos (en morado), polares negativos (naranja), hidrofóbicos alifáticos (azul), hidrofóbicos aromáticos (amarillo) demás aminoácidos (fucsia). Las conformaciones en hélice se denotan como barras de color magenta y las láminas como flechas negras.....	27
8	Representación gráfica de las 24 proteínas E6 superpuestas con el complejo E6/LxxLL descrita en la estructura cristalográfica 4GIZ.....	29
9	Minimización de la energía del sistema E6/LxxLL de VPH16. Descenso de la energía potencial tras el ciclo de minimización.....	30
10	Recocido simulado y equilibrado del sistema E6/LxxLL de VPH16. (A) cambios en la temperatura, (B) cambios en los valores de energía potencial (azul), cinética (verde) y total (gris) del sistema. La fase de recocido fue de 2.5 ns y la de equilibrado de 5.5 ns.....	31
11	Valores de RMSD para cada sistema E6/LxxLL en la fase de equilibrado. Las gráficas representan la frecuencia relativa de los valores de RMSD para cada sistema E6/LxxLL, para los diferentes tipos de VPH estudiados. En color amarillo, corresponden a los VPH de categoría: 1: cancerígeno para humanos, aguamarina 2A: probablemente cancerígeno para humanos, violeta 2B: posiblemente cancerígeno para humanos y azul 3: no clasificable. VPH16 está representado en negro. X= media, S= desviación estándar.....	32
12	Propiedades físicas del sistema E6/LxxLL de VPH16 durante la fase de DM. Fluctuación de la (A) temperatura; (B) presión; (C) volumen, (D) energía potencial (azul), cinética (verde) y total (gris) a lo largo de 100 ns de simulación.	34
13	Valores de RMSD para cada sistema E6/LxxLL durante la DM. Las gráficas representan la frecuencia relativa de los valores de RMSD para cada sistema E6/LxxLL de los diferentes tipos de VPH estudiados. En color amarillo, se representan a los VPH de la categoría: 1: cancerígeno para humanos, aguamarina 2A: probablemente cancerígeno para humanos, violeta 2B: posiblemente cancerígeno para humanos y azul 3: no clasificable; de las categorías IARC. VPH16 está representado en negro y el control negativo en naranja. Los histogramas en color gris corresponden al RMSD de las E6 sin extremos C y N y los coloreados al RMSD de los aminoácidos adyacentes al motivo LxxLL. En el panel inferior están tabulados los valores de (x) promedio, (s) desviación estándar y (r) correlación para los RMSD de toda la E6 y los aminoácidos adyacentes.....	36

- 14 RMSF para cada sistema E6/LxxLL durante la DM. Valores de RMSF de cada uno de los aminoácidos de la proteína E6 de los diferentes tipos de VPH. Los integrantes del grupo 1 (categorías IARC) de alto riesgo están representados con líneas amarillas, a excepción de VPH16 en negro, grupo 2A violeta, grupo 2B aguamarina, grupo 3 de bajo riesgo en azul y el grupo control en naranja. Los aminoácidos sombreados en rosa y gris corresponden a los aminoácidos implicados en la interacción de E6 y LxxLL (Zanier et al., 2013). Panel inferior: logo de los aminoácidos implicados en la interacción: polares positivos (en morado), polares negativos (naranja), hidrofóbicos alifáticos (azul), hidrofóbicos aromáticos (amarillo) demás aminoácidos (fucsia)..... 38
- 15 Promedio de enlaces de hidrógeno para cada sistema E6/LxxLL durante la DM. Las barras con valores < 5 corresponden al promedio de enlaces de hidrógeno que se forman entre cada proteína E6 y el motivo LxxLL y las barras con valores ≥ 15 corresponde al promedio de enlaces de hidrógeno entre el motivo LxxLL y el agua circundante a 5 Å. Los integrantes del grupo 1 (categorías IARC) de alto riesgo están representados con color amarillo, a excepción de VPH16 en gris, grupo 2A violeta, grupo 2B aguamarina, grupo 3 de bajo riesgo en azul y el grupo control en naranja..... 39
- 16 Promedio de la distancia entre los centros de masa de las proteínas E6 y el motivo LxxLL para cada sistema E6/LxxLL durante la DM. Los integrantes del grupo 1 (categorías IARC) de alto riesgo están representados con color amarillo, a excepción de VPH16 en gris, grupo 2A violeta, grupo 2B aguamarina, grupo 3 de bajo riesgo en azul y el grupo control en naranja..... 40
- 17 Valores de ΔG por MM/GBSA entre las proteínas E6 y el motivo LxxLL para cada sistema E6/LxxLL durante la DM. Los integrantes del grupo 1 (categorías IARC) de alto riesgo están representados con color amarillo, a excepción de VPH16 en gris, grupo 2A violeta, grupo 2B aguamarina, grupo 3 de bajo riesgo en azul y el grupo control en naranja..... 41
- 18 Valores de ΔG por MM/PBSA entre las proteínas E6 y el motivo LxxLL para cada sistema E6/LxxLL durante la DM. Los integrantes del grupo 1 (categorías IARC) de alto riesgo están representados con color amarillo, a excepción de VPH16 en gris, grupo 2A violeta, grupo 2B aguamarina, grupo 3 de bajo riesgo en azul y el grupo control en naranja..... 42
- 19 Contribución del componente ΔE_{elect} en la interacción entre las proteínas E6 y el motivo LxxLL para cada sistema E6/LxxLL durante la DM. Los integrantes del grupo 1 (categorías IARC) de alto riesgo están representados con color amarillo, a excepción de VPH16 en gris, grupo 2A violeta, grupo 2B aguamarina, grupo 3 de bajo riesgo en azul y el grupo control en naranja. Los puntos rojos indican los valores de contribución de las argininas a 5Å..... 42
- 20 Contribución del componente ΔE_{vdw} en la interacción entre las proteínas E6 y el motivo LxxLL para cada sistema E6/LxxLL durante la DM. Los integrantes del grupo 1 (categorías IARC) de alto riesgo están representados con color amarillo, a excepción de VPH16 en gris, grupo 2A violeta, grupo 2B aguamarina, grupo 3 de bajo riesgo en azul y el grupo control en naranja. Los puntos rojos indican los valores de contribución de las argininas a 5Å..... 43

- 21 Representación esquemática del protocolo de I-Tasser para el modelado por enhebrado. En primer lugar, la secuencia primaria de la proteína es analizada por la herramienta PSI-BLAST, la cual busca relaciones evolutivas con proteínas alojadas en diversas bases de datos, a continuación, crea una secuencia consenso y construye la estructura secundaria por medio del programa PSIPRED, el ensamble (estructura secundaria-secuencia consenso) es ahora enhebrada (comparado) con estructuras en el PDB, la comparación toma en cuenta secuencia y estructura, para el análisis de scoring utiliza 7 programas (LOMETS) y la selección del mejor empatamiento lo evalúa tomando en cuenta cuántos de los 7 programas llegaron al mismo resultado; los fragmentos que mejor enhebraron son cortados y utilizados en la construcción de una nueva conformación, mientras que los no resueltos son confinadas a un sistema de red; a continuación, los fragmentos resueltos se mantienen rígidos y ocurre un primer paso de optimización, haciendo uso de algoritmos Monte Carlo a diferentes temperaturas y campos de fuerza; por medio del programa SPICKER se agrupan las conformaciones de menor energía; a continuación, en la última etapa del proceso, se eliminan choques estéricos por el método steepest descent, y realiza un nuevo paso de optimización mediante el cálculo de enlaces de hidrógeno y minimizando por el método Monte Carlo.....

Lista de tablas

Tabla		Página
1	Tipos de VPH Elegidos.....	20
2	Tabla 2. Número de mutaciones entre la secuencia de la estructura cristalográfica de la proteína E6 (4GIZ) y la E6 de otros tipos de VPH	23
3	RMSD de los modelos de las proteínas E6 a utilizar en el presente trabajo.....	28

Capítulo 1. Introducción

Los papilomavirus son una familia amplia y diversa de virus de DNA con la capacidad de infectar gran variedad de animales como aves, reptiles, marsupiales y mamíferos. Los papilomavirus que infectan específicamente al humano son llamados Virus de Papiloma Humano (VPH) y al igual que el resto del grupo, se caracterizan por ser epiteliotrópicos, es decir, pueden infectar y proliferar en el tejido epitelial (Chouhy et al., 2013).

Las patologías asociadas al VPH varían entre género, especie y tipo. De todos los tipos de VPH reportados, al menos 50 pueden infectar el epitelio mucoso del área anogenital y orofaríngeo, por lo que dicha infección es considerada como una de las enfermedades de transmisión sexual más comunes (Smith, 2013). En la mayoría de los casos, la infección con VPH, en estas áreas, resulta asintomática, pudiendo aparecer verrugas y/o condilomas. Los tipos asociados a esta clase de infecciones son llamados de bajo riesgo. Sin embargo, existen una minoría de tipos, que pueden llevar a la célula infectada hacia un fenotipo tumoral, causando cáncer en zonas como: cuello uterino, vulva, vagina, ano, pene, faringe, entre otros, estos tipos son llamados de alto riesgo (Lavezzo et al., 2016).

Según datos del Instituto Nacional de Salud (*NIH*, por sus siglas en inglés) los agentes infecciosos representan uno de los principales factores de riesgo en el desarrollo de cáncer en el mundo; particularmente, los virus, contribuyen entre el 15-20% de los cánceres. Hasta el momento, se conocen siete tipos de virus cancerígenos para humanos: el virus de la hepatitis B (*HBV*, por sus siglas en inglés), virus de la hepatitis C (*HCV*, por sus siglas en inglés), virus de papiloma humano (*HPV*, por sus siglas en inglés), virus de Epstein Barr (*EBV*, por sus siglas en inglés), virus de herpes humano 8 (*HHV8*, por sus siglas en inglés), virus de células de Merkel (*MCPyV*, por sus siglas en inglés) y Virus linfotrópico de células T humanas (*HTLV-1*, por sus siglas en inglés). Específicamente el VPH, ha sido detectado en prácticamente el 100% de los casos de cáncer de cuello uterino, siendo esta última, el segundo motivo de muerte relacionado a cáncer en mujeres en el mundo, es decir, la infección con éste virus representa un gran problema de salud pública (Bosch et al., 2002).

Para que un tipo de VPH pueda establecer y mantener un fenotipo tumoral, debe expresar las oncoproteínas virales E6 y E7, las cuales actúan de forma cooperativa sobre rutas pro-proliferativas y anti-apoptóticas, tales como adhesión, polaridad y diferenciación celular (Tomaić, 2016). Si se bloquea la expresión de alguna de estas dos proteínas, en la célula infectada, ésta pierde el fenotipo tumoral y entra

a apoptosis o senescencia (Bonetta et al., 2015; Butz et al., 2003; Harden and Munger, 2017). En consecuencia, las proteínas E6 y E7 son consideradas excelentes blancos terapéuticos, por lo que la comprensión de los mecanismos moleculares que subyacen cada una de sus funciones es crítico en el desarrollo de terapias antivirales (Tomaić, 2016).

La mayoría de los estudios, que tratan de dilucidar el mecanismo de acción de las proteínas E6 y E7 son principalmente de tipo *in vitro* e *in vivo* (Chen et al., 1995; Ro et al., 2006; Scott, Nakagawa, and Moscicki, 2001; Vande, Brown, and Turner, 1998; Zanier et al., 2012; Zhao et al., 2013). Sin embargo, debido a los altos costos, la necesidad de infraestructura y la gran cantidad de tiempo que requiere la obtención de resultados a través de estos métodos, son pocos los tipos de VPH que han sido estudiados; hasta el momento, la investigación se ha centrado en los tipos VPH16 y VPH18, puesto que ambos causan más del 70 % de los casos de cáncer cervical (cuello uterino) (World HPV Information, 2017).

La falta de información con respecto a los otros tipos de VPH limita en gran medida la comprensión del fenómeno tumoral de éste virus. Una alternativa rentable, rápida y cada vez más precisa para el análisis a la par de muchos sistemas proteicos, son los estudios *in silico* o computacionales, los cuales buscan simular la estructura tridimensional de éstas, así como dar cuenta de algunas características fisicoquímicas.

Por lo anterior, el objetivo de éste trabajo, es el de analizar, por métodos *in silico*, la interacción entre las proteínas E6 de diferentes tipos de VPH, tanto de alto como bajo riesgo, con la proteína celular E6AP, también llamada E3 ubiquitina ligasa, implicada en el proceso de ubiquitinación. Específicamente, su interacción con un motivo en E6AP, la llamada hélice LxxxLL (Zanier et al., 2013), con el objetivo de determinar la existencia de alguna relación entre la energía libre de interacción, calculada de forma teórica, y la capacidad cancerígena de cada tipo de VPH analizado.

Capítulo 2. Antecedentes

2.1. Biología y clasificación de los papilomavirus (PVs)

Entre las características comunes de los Pvs están: todos poseen un genoma circular doble hebra de DNA entre 6953 pb (un PVs que infecta a *Chelonia mydas*) y 8607 pb (que infecta a *Cotton rabbits*), envuelto en una cápside proteica icosaédrica de 50-60 nm y sin envoltura (Doorbar et al., 2015); que se divide en tres regiones: *early* (E, por su sigla en inglés), *late* (L, por su sigla en inglés) y *LCR* (*long codon region*: por sus siglas en inglés). La región E codifica para proteínas de replicación como: E1, E2 y E4, y las oncoproteínas E5, E6 y E7 (Bocaneti et al., 2014); la región L codifica para las proteínas de la cápside L1 y L2 y la región LCR que no codifica para ninguna proteína, pero contiene algunos elementos esenciales, de regulación de la transcripción y replicación del virus (Figura 1).

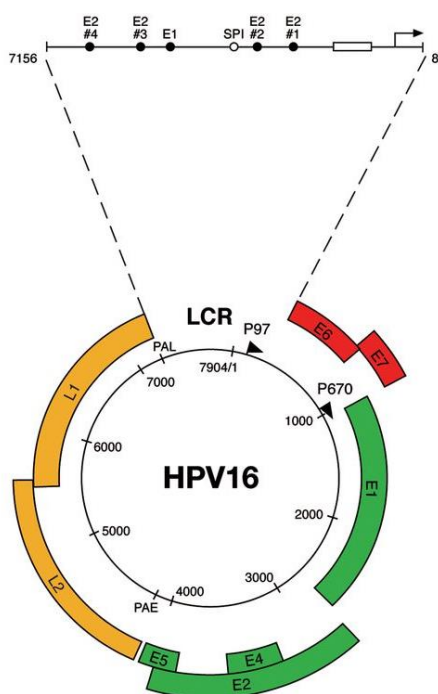


Figura 1. Organización del genoma de algunos papilomavirus. El genoma de los Pvs es circular de DNA doble hebra y se divide en tres regiones: *early* (E: por su sigla en inglés), *late* (L: por su sigla en inglés) y *long codon region* (LCR: por sus siglas en inglés).

De acuerdo con el comité internacional de taxonomía de virus (*ICTV*: por sus siglas en inglés) (Lefkowitz et al., 2018) se han reportado más de 300 tipos de Pvs, de las cuales, al menos 200 infectan únicamente al humano (VPH). Dicha clasificación es de tipo filogenética, basándose en un análisis por homología del marco abierto de lectura de la proteína L1 (L1 es la proteína más conservada de los PVs). Conforme a éste sistema, diferencias mayores al 10% determinan un nuevo tipo de virus, entre el 11-30% una nueva especie y entre el 40% un nuevo género (De Villiers et al., 2004a) .

Según ésta clasificación, los VPH están divididos en cinco grandes géneros: Betapapilomavirus, Gamapapilomavirus, Mupapilomavirus, Nupapilomavirus y Alfapapilomavirus (Van Doorslaer, 2013), éste último género es de gran importancia clínica debido a que en él se encuentran clasificados los VPH de alto riesgo, es decir, aquellos que pueden transformar a la célula infectada en una célula con un fenotipo tumoral. Estos son: VPH16, VPH18, VPH31, VPH33, VPH35, VPH39, VPH45, VPH51, VPH52, VPH56, VPH58, VPH59, VPH66, VPH68, VPH73 y VPH82 (Allison and Maleki, 2016). Cabe acotar, que, aunque los Pvs pueden infectar tejido epitelial tanto mucoso como cutáneo, la gran mayoría de los Alfapapilomavirus tienen mayor tropismo por el tejido epitelial mucoso, que los otros cuatro géneros.

Por otro lado, existe otro método de clasificación de los VPH pero de tipo clínica-epidemiológico, que los agrupa en 5 categorías según la prevalencia del tipo de VPH en un tejido con cáncer. Éstas son las categorías *IARC* (*International Agency for Research on Cancer*, por sus siglas en inglés). Las categorías son: **1**: cancerígeno para humanos, **2A**: probablemente cancerígeno para humanos, **2B**: posiblemente cancerígeno para humanos, **3**: no clasificable y **4**: probablemente no cancerígeno para humanos (World HPV Information, 2017).

2.2. Ciclo reproductivo del VPH

El ciclo viral, de los tipos de VPH que infectan el tejido epitelial mucoso, comienza en las células progenitoras de queratinocitos, específicamente en aquellas localizadas en la membrana basal del epitelio escamoso estratificado. El modo en el que llegan los viriones a las inmediaciones de dichas células aun es tema de debate, sin embargo, la hipótesis más aceptada es que ocurre a través de microabrasiones en el tejido durante el acto sexual, o durante algún otro tipo de contacto físico directo (Kines et al., 2009; Roberts et al., 2007). Una vez en los alrededores de los queratinocitos, la proteína L1 de la cápside viral se une al proteoglicano heparán sulfato que se encuentra en la matriz extracelular, ésto desencadena

cambios conformacionales en L1 y L2, lo que a su vez permite al virión ser transferido a el (los) receptores, aún sin identificar, que llevan a la internalización del virus dentro de la célula (Joyce et al., 1999).

Tras la entrada del virus a la célula, la cápside es eliminada y el DNA viral es transportado al núcleo celular, donde existe como episoma, un elemento genético separado del genoma de la célula huésped. En este momento, el virus emplea enzimas celulares y se replica junto con el genoma celular, por lo que se mantiene en un bajo número de copias (Pyeon et al., 2009). El proceso de replicación comienza con la expresión de las proteínas E1 y E2, las cuales forman un complejo proteico E1-E2 que se une al origen de replicación viral (ori) del genoma del virus (Enemark et al., 2000; García-Vallvé, Alonso, and Bravo, 2005; Wallace and Galloway, 2014), posteriormente son reclutadas algunas proteínas de replicación de la célula huésped tales como la polimerasa α de DNA. Seguidamente, se expresa la proteína E4, muy abundante durante la infección con Pvs, e involucrada tanto en la replicación como en la maduración del virus (Ferraro et al., 2011; Van Doorslaer, 2013). Durante esta fase, el virus también toma control del ciclo celular afectando las rutas de proliferación y muerte celular, expresando los genes tempranos E6 y E7 (Mittal and Banks, 2017). Más adelante se hablará de su papel protagónico en el desarrollo de cáncer.

Finalmente, cuando la célula infectada entra a la última etapa de diferenciación, ocurre un aumento en el número de genomas virales en el núcleo (Bodily and Laimins, 2011). A continuación, se expresan los genes correspondientes a la región L. La proteína L1 que tiene la capacidad de auto ensamblarse y formar estructuras pentaméricas que componen la cápside viral (Lowy and Schiller, 2006) y la proteína L2 que está implicada en la unión del DNA viral durante la encapsulación y consiguiente liberación de las partículas virales (García-Vallvé et al., 2005; Wang and Roden, 2013). Ambas proteínas son transportadas al núcleo donde ocurre la encapsulación. Por último, la maduración ocurre en los queratinocitos más superficiales y senescentes del epitelio, y es a partir de ellos que las nuevas partículas infecciosas son liberadas (Doorbar et al., 2012).

2.3. Proteínas E6 y E7 y su papel en la transformación celular a un fenotipo cancerígeno

El papel específico de las proteínas E6 y E7, durante el ciclo infeccioso, de los tipos de VPH que infectan queratinocitos aún no es completamente claro, ésto se debe, a que ambas proteínas poseen

múltiples motivos con la capacidad de interactuar con una gran variedad de proteínas celulares, las cuales a su vez están implicadas en diversos procesos como: control de ciclo celular, apoptosis y evasión del sistema inmune, afectando varias vías metabólicas o de señalización al mismo tiempo (Araldi et al., 2017; Westrich, Warren, and Pyeon, 2017). En la Figura 2 se listan algunas de las proteínas con las que interacciona E6 y las vías que afecta.

La proteína E6 tiene entre 151 a 158 aminoácidos y no tiene actividad enzimática (Tan et al., 2012), contiene regiones de interacción importantes como un dominio PDZ tipo I en el extremo C-terminal, cuatro dominios conservados de tipo (Cys-X-X-Cys) (Boon et al., 2015) y dos motivos de unión a un átomo de zinc por medio de residuos de cisteína (conocidos como dedos de zinc) (Ristriani et al., 2000). Una de las funciones más conocidas, por su importancia clínica, es la capacidad que tiene de formar un complejo proteico trimérico con las proteínas celulares E6AP ubiquitina ligasa y la proteína supresora de tumores p53. En dicho complejo, E6 se une primero a E6AP, lo que causa un cambio conformacional en E6 que le permite unirse con p53, éste complejo E6AP/E6/p53 es el responsable de la degradación proteolítica de p53 por parte del proteosoma 26S (Cai et al., 2013; Liu et al., 2006; Zanier et al., 2013). La pérdida de p53 evita que la célula pueda realizar apoptosis y contribuye a que haya inestabilidad genómica (Araldi et al., 2015). La proteína E6 puede interactuar con muchas proteínas celulares, se ha demostrado que en los tipos de VPH de alto riesgo, la E6 muestra tendencia a interactuar con la proteína E6AP para la degradación de p53 (Werness, Levine, and Howley, 1990).

Por otro lado, la proteína E7 tiene 127 aminoácidos y es capaz de unir a través de un dominio conservado (L-X-C-X-E) a la proteína celular supresora de tumores pRb, retinoblastoma (Moody and Laimins, 2010; White and Howley, 2013). Dicha unión resulta en la fosforilación de pRb, lo que a su vez conduce al factor de transcripción E2F hacia el núcleo. Una vez dentro, E2F recluta diferentes modificadores de la cromatina como las desacetilasas de histonas, lo que tiene como consecuencia, la expresión constitutiva de los genes regulados por E2F, como los de las ciclina A y E (Moody and Laimins, 2010); ésto lleva a la célula a un incremento de las fases S y G2/M del ciclo celular (Ferraro et al., 2011; Sacco, Siepi, and Crescenzi, 2003) .

Una característica importante, para la transformación y posterior mantenimiento del fenotipo tumoral en células infectadas con tipos de VPH de alto riesgo, es que tanto la proteína E6 como E7 son necesarias, ya que experimentos *in vivo* e *in vitro* han demostrado que si se suprime la producción de alguna de ellas la célula es conducida hacia la apoptosis o arresto del crecimiento celular (Alvarez-Salas et al., 1998; Butz et al., 2003; Butz et al 2000; Steele et al., 1992; Yoshinouchi et al., 2003). Por otro lado, en algunos tipos de VPH de alto riesgo, especialmente VPH16, el DNA del virus se integra de forma aleatoria al genoma del

hospedador, lo que lleva a la pérdida de genes reguladores y la consecuente sobre-expresión de las proteínas E6 y E7 (Androphy, Hubbert, Schiller, and Lowy, 1987; Androphy et al., 1987).



Figura 2. Función de la oncoproteína E6 en un VPH de alto riesgo. Proteínas celulares blanco de la E6 y su actividad asociada.

2.4. Estudios de estructura-función de la proteína E6

Las proteínas son las macromoléculas más versátiles de los sistemas vivos, pueden catalizar reacciones, transportar moléculas, dar soporte mecánico, o brindar protección contra patógenos, entre otras funciones. Es bien sabido, que la función de las proteínas está estrechamente ligada a la estructura, cualquier cambio en la conformación, por más pequeño que sea, puede ocasionar diferencias en el plegamiento total de la proteína y ocasionar la pérdida completa de su función (Nelson, Lehninger, and Cox, 2015).

A la fecha, la mayoría de las investigaciones que buscan dilucidar la función(es) de la proteína E6 son de tipo *in vitro* e *in vivo*, principalmente mediante las técnicas de inmunoprecipitación o el doble híbrido, (Araldi et al., 2017; J. J. Chen et al., 1995; Kaphingst, Persky and Lachance, 2010; Martinez-Zapien et al., 2016a; Nakagawa and Huibregtse, 2000; Zanier et al., 2012, 2013). Sin embargo, dichos estudios presentan limitaciones: la función de la proteína E6 es descrita como un evento de interacción entre la proteína E6 y

alguna proteína celular sin ahondar en cómo ocurre esa interacción a nivel estructural o describir los aminoácidos implicados; además, se centran únicamente en la proteína E6 de los tipos VPH16 y VPH18, olvidándose del resto de los tipos; y hay que recordar, que a pesar de que el resto de los VPH son menos cancerígenos que el 16 y 18, la comprensión de su mecanismo de acción es importante para entender el fenómeno tumoral de éste virus.

Dentro de los pocos estudios estructurales (*in silico*, *in vitro* e *in vivo*) de la proteína E6, resaltan los reportes de (Zanier et al., 2013 y Martínez-Zapien et al., 2016) quienes describen la estructura cristalográfica elucidada mediante difracción de rayos-X del complejo E6 y la proteína E6AP y el complejo trimérico de degradación p53/E6/E6AP, respectivamente. En ambos estudios, se describe la interacción de la proteína E6 del tipo VPH16 con la proteína celular E6AP mediante un bolsillo hidrofóbico cargado, en E6, y un motivo peptídico en forma de hélice, en E6AP, llamado LxxLL por su alto contenido en el aminoácido leucina. De igual forma, detallan los aminoácidos implicados en la interacción (ver Figura 3) y el tipo de enlace que forman; éste último resultado lo obtienen mediante métodos de simulación por dinámica molecular.

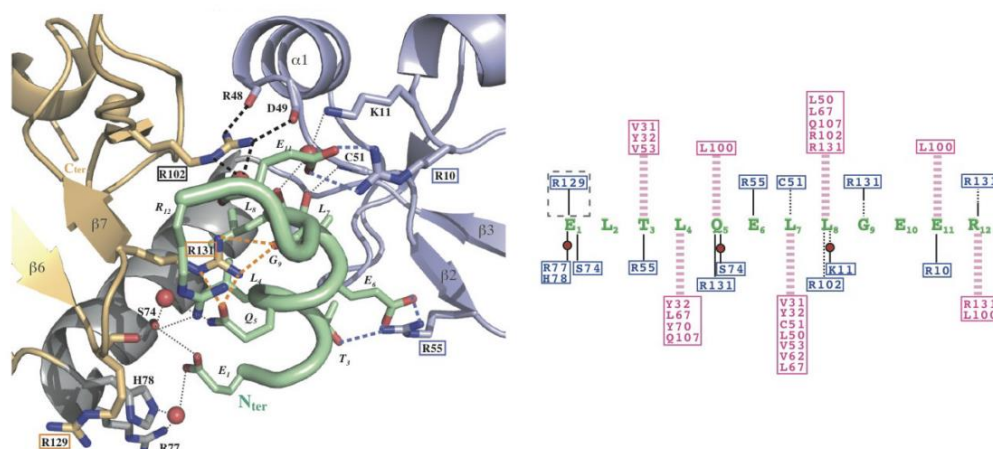


Figura 3. Red de interacciones entre la proteína E6 de VPH16 y el motivo LxxLL. La información fue obtenida a partir de la estructura cristalina del complejo E6/LxxLL. Panel izquierdo: representación de las interacciones polares entre aminoácidos; la estructura en verde se refiere al motivo LxxLL, el resto a la proteína E6, la estructura en amarillo, son los aminoácidos de la región cercana al C-terminal y los violeta al N-terminal. Panel derecho: Representación de todos los contactos entre E6 y LxxLL; las líneas punteadas de color rosa representan interacciones de tipo hidrofóbicas, las líneas negras a los contactos polares, medidas por la cadena lateral (línea continua) y por la cadena principal (línea punteada).

2.5. Dinámica molecular (DM) en sistemas biológicos

Las simulaciones por DM en sistemas biológicos es un conjunto de métodos que busca simular en un espacio computacional, el movimiento de los átomos que componen biomoléculas, con base al potencial de interacción de las mismas y por medio del cálculo de la fuerza de cada átomo con ecuaciones newtonianas de movimiento (Martin Karplus and Petsko, 1990). En consecuencia, es una poderosa herramienta que permite explorar cambios conformacionales en diferentes estados energéticos a lo largo del tiempo y obtener información como flexibilidad y propiedades cinéticas y termodinámicas de la(s) molécula(s) en estudio (Karplus and Kuriyan, 2005).

Por otro lado, el cálculo de la energía libre de interacción basado en datos de mecánica molecular con DM, se ha convertido en una herramienta rutinaria en el estudio de interacciones intermoleculares sobre todo por su aplicación en la industria farmacéutica para la búsqueda y validación de nuevos fármacos dirigidos a proteínas blanco (De Vivo et al., 2016). Entre los métodos de cálculo de energía libre más utilizados, por su rapidez y precisión, están los métodos de punto final MM-GBSA y MM-PBSA (Molecular Mechanics Generalized Born Surface Area y Molecular Mechanics Poisson-Boltzmann Surface Area) (Genheden and Ryde, 2015), los cuales han demostrado gran precisión en el cálculo de afinidad entre complejos proteína-proteína (B. K. and Kollman, 2000; I. M. and Kollman*, 1999; Chong et al., 1999; Gohlke, Kiel, and Case, 2003; Huo et al., 2002; Villacañas and Rubio-Martinez, 2006), proteína-RNA (Reyes and Kollman, 2000) y RNA- metales (V. T. and Case, 2001).

En el caso particular del VPH el único reporte de la energía libre de interacción teórico entre las proteínas E6 y E6AP fue realizado por el método de MM-PBSA (Zanier et al., 2013) para la E6 del tipo VPH16. Sin embargo, en éste no se especifica el valor total de la energía libre de interacción ($\Delta G_{\text{unión}}$), por su parte, detalla el aporte a la energía libre de interacción de cada aminoácido implicado en la interacción.

En consecuencia y debido a que los métodos de MM-PBSA y MM-GBSA han demostrado tener precisión en el estudio de sistemas biológicos y que no hay información, teórica, concerniente a la afinidad entre las proteínas E6 y E6AP (la ruta de interacción conspicua entre los tipos de VPH de alto riesgo) para cualquier otro tipo de VPH que no sea el VPH16. El objetivo de éste trabajo es el de calcular los valores de $\Delta G_{\text{unión}}$ entre las proteínas E6 y el motivo de interacción LxxLL en E6AP con ambos métodos, para determinar si existe alguna relación entre la capacidad oncogénica del virus y la energía de interacción entre éstas dos proteínas.

Capítulo 3. Hipótesis y Objetivos

3.1. Hipótesis

La energía libre de interacción, calculada a partir de un análisis in silico de dinámica molecular, entre las proteínas E6 de distintos tipos de VPH de alto y bajo riesgo, y el motivo de interacción LxxLL de la proteína celular E6AP, determinará el potencial oncogénico de los tipos de VPH estudiados.

3.2. Objetivos

3.2.1. Objetivo general

Evaluar la afinidad entre las proteínas E6 y el motivo LxxLL de la proteína celular E6AP, de varios tipos de VPH de alto y bajo riesgo, para determinar si existe correlación entre este valor y la capacidad oncogénica del virus.

3.2.2. Objetivos específicos

- Seleccionar secuencias de la proteína E6 a partir de los tipos de VPH más representativos dentro de los grupos de alto y bajo riesgo.
- Obtener los modelos estructurales, por homología, de cada una de las E6 seleccionadas.
- Obtener los modelos estructurales, mediante dinámica molecular, del complejo E6-Zn/LxxLL.
- Determinar la energía de interacción entre E6 y el motivo LxxLL, a partir de las estructuras
- Obtenidas por dinámica molecular, mediante los algoritmos MMGBSA y MMPBSA.

Capítulo 4. Metodología

4.1. Elección y análisis de las secuencias de la proteína E6 a estudiar

4.1.1. Elección de las secuencias de E6

Se tomaron 24 secuencias, en aminoácidos, de la proteína E6 de diferentes tipos de VPH a partir de la base de datos NCBI (National Center for Biotechnology Information: por sus siglas en inglés) (<http://www.ncbi.nlm.nih.gov>) en formato fasta. La selección de secuencias se hizo, a partir de tipos de VPH que pertenecieran a la misma categoría según el IARC y de la misma especie (Bravo and Félez-Sánchez, 2015; Hsa, 2014). Como grupo control, se tomó la E6 de un tipo de VPH correspondiente a otra especie.

4.1.2. Alineamiento y construcción de fenogramas

Tanto los métodos de alineamiento como los de construcción de fenogramas, son algoritmos bioinformáticos que permiten identificar la posible relación estructural, funcional y/o evolutiva, entre secuencias de nucleótidos o aminoácidos. A través del uso del programa JalView, se realizaron alineamientos de las 24 secuencias de la E6, con los métodos MUSCLE, CLUSTAL-O y T-coffe (Edgar, 2004); con los que se generaron fenogramas por medio del uso del algoritmo Neighbor Joining y Maximum likelihood con un bootstrap de 1000, dichos fenogramas fueron realizados mediante el uso del programa Mega-cc (Kumar, et al., 2012; Pattengale et al., 2010; Saitou and Nei, 1987).

4.1.3. Análisis de la estructura primaria de las E6

A partir de la secuencia primaria de las 24 secuencias de la E6, se realizó un análisis de composición de aminoácidos, por medio del uso de la herramienta en línea protparam de Expsy (<http://web.expasy.org/protparam>) y se determinó la estructura secundaria con la herramienta en línea Jpred (Drozdetskiy et al., 2015).

4.2. Modelado por enhebrado de las E6

4.2.1. Modelado por enhebrado con I-TASSER

El método de modelado por enhebrado, o de reconocimiento de plegamientos, consiste en enhebrar fragmentos de la secuencia consulta (estructura primaria de la proteína) con una conformación 3D conocida, el mejor modelo estructural será aquel que minimice la energía libre de la secuencia consulta (Bowie, Lüthy, and Eisenberg, 1991). Para dicho cálculo, se empleó el programa en línea I-TASSER (<https://zhanglab.ccmb.med.umich.edu/I-TASSER/>) (Yang et al., 2014) con lo que se obtuvo los modelos estructurales para las 24 proteínas de la E6 elegidas. El mejor modelo, se eligió con base al RMSD (root-mean-square deviation of atomic positions) o medida de la distancia promedio entre átomos, calculado por I-TASSER entre el modelo generado y la estructura de referencia seleccionada por el programa. Finalmente, los modelos generados fueron guardados en formato pdb (Protein data bank format). Por otro lado, el modelo estructural del motivo LxxLL se obtuvo a partir del complejo E6/E6AP resuelto por difracción de rayos X (código de identificación: 4GIZ en <https://www.rcsb.org/structure/4giz>) reportado en (Zanier et al., 2013).

4.3. Modelado por dinámica molecular del complejo E6-Zn/LxxLL

4.3.1. Preparación del complejo E6-Zn/LxxLL

Una vez seleccionados los 24 modelos con el mejor RMSD, se realizó el siguiente flujo de trabajo: (I) debido a que los archivos pdb generados por I-TASSER contienen todos los átomos de hidrógeno enlazados a cada átomo de carbono, nitrógeno, oxígeno y azufre de las E6, se realizó un paso de oxidación mediante el programa Reduce del paquete de AmberTool17 (Case et al., 2017), el motivo de esto es que los estados de protonación de los aminoácidos serán calculados por el programa LeaP más adelante, en el paso VI, haciendo uso de campos de fuerza de AMBER (programa a utilizar para la fase de producción); (II) de igual forma, algunos de los pdb generado por I-TASSER, tienen los átomos de azufre de las cisteínas implicadas en los motivos de unión a Zinc en direcciones diferentes a las reportadas en el cristal 4GIZ (Zanier et al., 2012), por lo que se ajustaron algunos ángulos de torsión, de forma manual, haciendo uso del programa de visualización CHIMERA (Pettersen et al., 2004) en función de las coordenadas del cristal;

(III) se adicionó las coordenadas del péptido LxxLL y los dos átomos de zinc a cada archivo pdb de la E6 de tal forma que quedaran ajustadas a las coordenadas de la E6 y el motivo LxxLL en el cristal 4GIZ; (IV) y se adecuó cada nuevo archivo pdb (E6+Zn+LxxLL) con el programa pdb4amber de AmberTool17; (V) debido a que ninguno de los pdb contiene información del estado de protonación de los residuos ionizables, se utilizó la herramienta en línea PDB2PQR para dicho cálculo a pH 7 (http://nbcrc-222.ucsd.edu/pdb2pqr_2.1.1/) (Pettersen et al., 2004); (VI) Finalmente, con el programa LEaP de AmberTool17 se generaron los archivos de topología (.prmtop) y de coordenada (.inpcrd) necesarios para realizar los cálculos de minimización y dinámica molecular; los campos de fuerza utilizados para modelar la E6 fueron, leaprc.gaf (Wang et al., 2004), leaprc.protein.ff14SB y ZAFF.frcmod, éste último corresponde a los campos de fuerza para los dedos de Zinc (Yu et al., 2018); de igual forma, se solató cada complejo con una caja de agua de tipo TIP3P (Price et al., 2004) de 15 Å a partir del último átomo de la proteína en los tres ejes, el sistema se neutralizó con iones Cl⁻ y Na⁺.

4.3.2. Minimización de la energía

El objetivo del proceso de minimización de energía, es encontrar las coordenadas de los átomos que constituyen un sistema, donde la fuerza interatómica neta entre cada átomo sea lo suficientemente bajo (Mackay, Cross, and Hagler, 1989). Con el programa pmemd del paquete AMBER16 se realizó el proceso de minimización para cada complejo E6-Zn/LxxLL junto con la caja de agua; dicho programa relaja la estructura moviendo los átomos de forma iterativa de forma descendente en el gradiente de energías, hasta que se obtiene un gradiente promedio lo suficientemente bajo; los archivos de entrada fueron los .prmtop, .inpcrd y un .in, en éste último se especifican los pasos a seguir durante el proceso. Se utilizaron dos algoritmos de minimización, el Steepest descent para los primeros 5000 ciclos y el Conjugate gradient para los 5000 restantes (Standard, 2015). El archivo de salida corresponde a un .rst donde se especifican las nuevas coordenadas de los átomos del sistema.

4.3.3. Recocido simulado y equilibrado

El recocido simulado, en dinámica molecular, es un algoritmo que permite explorar las posibles conformaciones que puede adoptar una molécula a una temperatura dada, por lo tanto, se pueden

analizar estructuras de alta y baja energía (Tidor, 1993) y así evitar caer en mínimos locales. Como archivo de entrada, se utilizó: el .rst del paso de minimización y un .in donde se especifican las características de la simulación; los pasos de calentamiento fueron los siguientes: (I) calentamiento de 0-300 K en 0.5 ns, luego (II) se mantuvo a 300 oK por 0.5 ns, (III) calentamiento de 300-400 K en 0.5 ns, (IV) se mantuvo a 400 K por 0.5 ns y (v) se enfrió de 400 a 310 K en 0.5 ns. A continuación, se realizó un paso de equilibrado de 5.5 ns a 310 K. Las condiciones de simulación durante el recocido y equilibrado fue de un colectivo NVT con condiciones de frontera periódica. 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). Los archivos de salida, corresponden a un .rst que contiene las coordenadas del complejo y las velocidades después del equilibrio y un .nc con las trayectorias (coordenadas para cada frame o conformación a lo largo de la simulación).

4.3.4. Evaluación del equilibrio

Para evaluar la estabilidad conformacional del sistema luego del recocido simulado se calculó el RMSD (root-mean-square deviation: por sus siglas en ingles) cuyo resultado corresponde a la medida (en Å) del desplazamiento atómico promedio entre dos conformaciones dadas, en nuestro caso, entre la estructura del primer paso de equilibrado y el resto de las obtenidas en los 5.5 ns de la fase de equilibrado. Para dicho cálculo se utilizó la herramienta de análisis de trayectorias cpptraj de AmberTool17 (Roe and Cheatham, 2000). Como archivos de entrada se utilizaron: el correspondiente a las trayectorias (.nc, formato NetCDF) y el de topología (.prmtop); como estructura de referencia se eligió el primer conjunto de coordenadas de la trayectoria, y se evaluó únicamente los carbonos (α) sin los C y N terminales. La ecuación que describe el cálculo del RMSD es:

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

Donde RMSD es la distancia en Å entre una estructura en el instante t y una estructura de referencia en el instante 0. N es el número de átomos (carbonos α) 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.

4.3.5. Simulación por dinámica molecular

Las simulaciones por dinámica molecular buscan calcular el movimiento de átomos individuales en modelos sólidos, líquidos y gaseosos, para describir cambios de posición, velocidad y orientación a lo largo del tiempo (Leach, 2001). En éste paso, al igual que en el recocido simulado, se empleó el programa pmemd del paquete AMBER16. Para la simulación se utilizó el colectivo NPT, donde, el número de partículas (N), la presión (P) y la temperatura del sistema se mantienen constante; la temperatura y la presión se mantuvieron, con pequeñas oscilaciones, alrededor de 310 K y 1 atm, respectivamente. El tiempo total de simulación fue de 100 ns, registradas con un tiempo de integración de 0.02 ns, el cálculo de las interacciones electrostáticas de largo alcance se hizo con el método de Particle Mesh Ewald y un cutoff de 10 Å, todos los enlaces que involucraran átomos de hidrógeno se mantuvieron rígidos por medio del algoritmo SHAKE. Para mantener la temperatura y la presión se utilizó el termostato Langenvin y el barostato descrito en (Berendsen et al., 1984). Como archivo de entrada se utilizó el .rst con los valores de posición y velocidad del último paso de equilibrado y un .in con las especificaciones de la simulación. Como archivos de salida, se obtuvieron un archivo .nc que contiene las coordenadas de las conformaciones adoptadas a lo largo de la simulación (frames) y un archivo con los datos de temperatura, presión, volumen y energía del sistema a lo largo de la simulación.

4.3.6. Análisis de las trayectorias obtenidas durante la DM

4.3.6.1. RMSD (root-mean-square deviation)

Se empleó el cálculo de RMSD, descrito en la ec.1, tanto para los carbonos α de cada estructura de la E6, como para la región de interacción entre la E6 con la hélice LxxLL únicamente.

4.3.6.2. RMSF (root-mean-square fluctuation)

Se calculó la flexibilidad estructural de la proteína por medio del valor del RMSF, el principio matemático es el mismo que el del RMSD, con la diferencia, de que se hace por residuo de aminoácido y toma como estructura de referencia la media para cada coordenada (x, y, z) de cada átomo (Bornot, Etchebest, and De Brevern, 2011). Se utilizó el programa de análisis de trayectorias cpptraj de AmberTool17. Como archivos de entrada se utilizaron: el correspondiente a las coordenadas a lo largo de la trayectoria .nc y el de topología (.prmtop) y se evaluó únicamente los carbonos α de cada aminoácido. La ecuación que describe el cálculo del RMSF es:

$$\text{RMSF}_i = \sqrt{\frac{1}{T} \sum_{t=0}^T (\vec{R}_t^i - \vec{R}_{\text{ave}}^i)^2} \quad (2)$$

Donde el RMSF es la distancia en Å entre las coordenadas del residuo i en un tiempo de referencia t y las coordenadas del mismo residuo en una estructura promedio. T corresponde al tiempo en número de frames a lo largo de la trayectoria, R_t son las coordenadas del carbono α del aminoácido i en un tiempo t y R_{ave} son las coordenadas promedio del carbono α del aminoácido i durante todo el tiempo de simulación.

4.3.6.3. Enlaces de hidrógeno

Por medio del programa de visualización de dinámicas moleculares (VMD: por sus siglas en inglés) (Stone, 2011), se cuantificó el número de enlaces de hidrógeno que se formaron a lo largo de la dinámica:

entre la hélice LxxLL y cada una de las E6 estudiadas, y entre las moléculas de agua a 5 Å de la hélice y la hélice.

4.3.6.4. Contactos nativos

Un contacto nativo, corresponde al contacto que ocurre entre las cadenas laterales de dos de aminoácidos, que no se están continuos en la secuencia de la proteína, pero si se encuentran adyacentes en el espacio (Turjanski et al., 2008) . En aras de determinar la unión entre la E6 y la hélice LxxLL, se calculó el número de los contactos nativos entre éstas a lo largo de toda la simulación; para ésto, se utilizó la herramienta cpptraj de AmberTool17. Como archivos de entrada se utilizaron: el correspondiente a las coordenadas a lo largo de la trayectoria .nc y el de topología (.prmtop) evaluando todo lo que estuviera en un radio de 5Å entre el motivo LxLL y las E6.

4.3.6.5. Distancia entre los centros de masa

El centro de masa de una proteína es un punto geométrico artificial (x, y, z) que define el centro de la molécula con base en valores de coordenadas y masa (Costantini, Paladino, and Facchiano, 2008). Para identificar si la distancia entre las proteínas E6 y la hélice LxxLL es diferente entre los distintos tipos de VPH, se calculó la distancia entre el centro de masas de cada E6 y de la hélice a lo largo de la trayectoria. Para ésto, se utilizó la herramienta cpptraj de AmberTool17. Como archivos de entrada se utilizaron: el correspondiente a las coordenadas a lo largo de la trayectoria .nc y el de topología (.prmtop). La ecuación que define el cálculo del centro de masa es:

$$a = \frac{\sum_{i=1}^N m_i a_i}{\sum_{i=1}^N m_i} \quad (3)$$

Donde **a** es la coordenada (x, y, z) del centro de masas, **N** es el número de átomos total, **m_i** la masa de cada átomo y **a_i** una de las tres coordenadas (x, y, z).

4.3.6.6. Análisis por distribución de Boltzmann

En mecánica estadística, la distribución de Boltzmann corresponde a la función matemática que describe el número de moléculas, con una energía asociada, dentro de un intervalo de energías (Kovacs and Mcharris, 2000). Partiendo de la premisa, de que la conformación con la energía más frecuente corresponde a la estructura que con mayor probabilidad se podría encontrar en la naturaleza, por ser la más favorable desde el punto de vista energético; se utilizó éste método para seleccionar a lo largo de cada sistema E6/LxxLL simulado dicha conformación. La ecuación de la distribución de Boltzmann es:

$$f = \frac{e^{-\Delta E / K_b T}}{\sum e^{-\Delta E / K_b T}} \quad (4)$$

Donde **f** es la función probabilística, **E** la energía total de cada conformación, **K_b** la constante de Boltzmann (1.987*10⁻³ Kcal/mol), **T** temperatura absoluta en grados kelvin, **-ΔE** la diferencia en energía entre cada conformación y la de mínima energía.

4.4. Cálculo de la energía libre de interacción entre la proteína E6 y el motivo LxxLL

4.4.1. Métodos de MM-GBSA y MM-PBSA

Para calcular la energía libre de interacción (teórica) entre las proteínas E6 de cada tipo de VPH y la hélice LxxLL, se utilizaron los métodos MMPBSA (Molecular Mechanics Poisson-Boltzman Surface Area) (Kollman et al., 2000) y MMGBSA (Molecular Mechanics-Generalized Born Surface Area) (Clark et Still, Tempczyk, and Hawley, 1990). En ambos algoritmos, la energía libre de interacción ($\Delta G_{\text{unión, solvente}}$) es calculada por medio de la ecuación:

$$\Delta G_{\text{unión, solv}} = \Delta G_{\text{solv, complejo}} - (\Delta G_{\text{solv, receptor}} + \Delta G_{\text{solv, ligando}}) \quad (5)$$

Donde el cambio de energía libre (ΔG) de cada estado: complejo, receptor y ligando, es calculada a través de la ecuación:

$$\Delta G_{solv;com,rec,lig} = \Delta E_{com,rec,lig} + \Delta G_{solv} - T\Delta S \quad (6)$$

A su vez el ΔE o cambio en la energía, es calculado en términos de mecánica molecular, por medio de la ecuación:

$$\Delta E_{com,rec,lig} = \Delta E_{enlaces} + \Delta E_{ángulos} + \Delta E_{torciones} + \Delta E_{vdw} + \Delta E_{elec} \quad (7)$$

y el término de contribución de la energía libre de solvatación (ΔG_{solv}) mediante la ecuación:

$$\Delta G_{solv} = \Delta G_{pb/gb} + \Delta G_{asa} \quad (8)$$

El ΔG_{asa} corresponde a la contribución del componente apolar del solvente y $\Delta G_{pb/gb}$ al componente polar o electrostático del solvente. Si se utilizan las ecuaciones de Poisson-Boltzman para resolver éste último término, se estará empleando el modelo de MM-PBSA y si son las ecuaciones del modelo generalizado de Born se estará empleando el método de MM-GBSA.

La aplicación de ambos métodos se realizó mediante la herramienta MMPBSA.py de AmberTool17. Como archivos de entrada se utilizaron: el correspondiente a las coordenadas a lo largo de la trayectoria .nc y el de topología (.prmtop), con los que se crearon nuevos archivos de topología del complejo, receptor y ligando libres del solvente de la DM. Para el cálculo se utilizaron 250 frames de los 5000 frames de cada trayectoria.

Capítulo 5. Resultados

5.1. Elección y análisis de las proteínas E6 a estudiar

5.1.1. Elección de las secuencias de E6

Se eligieron las secuencias de la proteína E6 a partir de tipos de VPH que cumplieran con dos criterios: que pertenecieran al género Alfacapilomavirus según la clasificación hecha con el ORF de la proteína L1, esto, debido a que dentro del género Alfa están clasificados todos los tipos de VPH de alto riesgo (cancerígenos) y de bajo riesgo (no cancerígenos) que infectan el mismo tipo de tejido epitelial, el mucoso (Bravo and Félez-Sánchez, 2015; De Villiers et al., 2004b) y que además estuvieran clasificados dentro de las categorías del IARC (International Agency for Research on Cancer): **1**: cancerígenos, **2A**: probablemente cancerígenos, **2B**: posiblemente cancerígenos y **3**: no clasificable (Hsa, 2014). Como control, se eligió un tipo de VPH del género Mupapilomavirus, el cual infecta tejido epitelial cutáneo y produce verrugas en manos y plantas de pies (no cancerígenos) (Bravo, de Sanjos, and Gottschling, 2010), En total se eligieron 24 secuencias de la proteína E6.

Tabla 1. Tipos de VPH Elegidos

		Género		Alfa				Mu	
		Especie		5	6	7	9	10	1
Categorías IARC	1			18, 39, 45, 59	16, 31, 33, 35, 52, 58				
	2A			68					
	2B	69, 82	30, 53, 66	70	67				
	3						6, 11, 13, 44, 55, 74	1	

Las secuencias de las proteínas E6 se extrajeron de la base de datos PaVe (Papillomavirus Epitome: por sus siglas en inglés) (<https://pave.niaid.nih.gov/>). En la Tabla 1 se muestran los 24 tipos de VPH seleccionados.

5.1.2. Alineamiento y construcción de fenogramas

Las secuencias de E6 elegidas en formato fasta (ver Tabla 2) fueron utilizadas para realizar un alineamiento múltiple mediante los algoritmos MUSCLE, CLUSTAL-O y T-coffe. Los resultados con los algoritmos T-coffe y CLUSTAL-O mostraron GAPS (aperturas entre las secuencias) evolutivamente improbables (resultados no mostrados), por lo que se decidió trabajar con el alineamiento construido con el programa MUSCLE utilizando los parámetros predeterminados del programa MEGA. Tal como se muestra en la Figura 4. De forma adicional, se comparó el alineamiento obtenido con la clasificación propuesta a partir del análisis comparativo del ORF de la proteína L1, ver Figura 1 del artículo (De Villiers et al., 2004a). Como se puede observar, las secuencias que fueron alineadas formaron grupos definidos, que al compararlos con las especies reportadas para la proteína L1, el alineamiento parece obedecer la misma clasificación en cuanto a especie. Al suponer ésto, los tipos del grupo 1 (categorías IARC) de alto riesgo, quedan clasificados en las especies Alfa 7 y 9 (sombreado amarillo), los del grupo 2A en Alfa 7, los del grupo 2B en Alfa 5, 6, 7 y 9 (sombreado violeta) y los de grupo 3, bajo riesgo, en Alfa 10 (sombreado azul) y el tipo HPV1 (control negativo) no se relaciona con ninguno. Por otro lado, se observó que la proteína E6 entre las diferentes especies de VPH es poco conservada, sin embargo, algunos aminoácidos importantes desde el punto de vista estructural y de interacción con la proteína E6AP están conservados (Scheffner et al., 1990; Zanier et al., 2013).

Se realizaron fenogramas por medio de dos modelos, Neighbor Joining (NJ) y Maximum likelihood, ambos con un Bootstrap de 1000 (número de veces que se reconstruye el fenograma a lo largo del cálculo) y el método de sustitución p-distance (proporción de sitios de aminoácidos diferentes entre las secuencias); la topología de ambos fenogramas fue parecida (datos no mostrados). En la Figura 5 se muestra el fenograma construido por NJ. Como se puede observar, los tipos que constituyen cada especie se agrupan igual que la reportada para la ORF de la proteína L1 (tal como se observó para la Figura 4), sin embargo, la relación entre las diferentes especies no es la misma. Según nuestros resultados, las especies alfa 6, 7 y 10 son más parecidas entre sí al compararlas con las especies alfa 5 y 9, y el grupo con mayor diferencia corresponde a los Alfa 10; caso diferente al reportado, donde las especies 5, 6 y 7 son las más parecidas, y en conjunto, estos se relacionan más con los alfa 9 que con los Alfa 10, es decir, los Alfa 10 (de bajo riesgo) siguen siendo el grupo que más diverge (De Villiers et al., 2004a).

Se cuantificó las mutaciones entre la secuencia de la proteína E6 de VPH16 reportada en la estructura cristalográfica 4GIZ y el resto de las secuencias, tanto a lo largo de las secuencias como en la región de interacción con el motivo LxxLL. Tal como se observa en la Tabla 2, los tipos de VPH en la especie Alfa 9

difieren en promedio en 53 aminoácidos (aa) con respecto al cristal, en 72 aa para la especie Alfa 7, en 75 para Alfa 6, en 68 para Alfa 5, en 98 para Alfa 10 y 102 con respecto a la especie Mu. Por otro lado, en cuanto a la región de interacción, se utilizaron como referencia los aa determinados para la E6 de VPH16 (Zanier et al., 2013), en la especie Alfa 9 hay diferencia en promedio de 5 mutaciones, de 8 para Alfa 7, 7 para Alfa 6, 7 para Alfa 5, 10 para Alfa 10 y 15 con respecto a la especie Mu. Según éste último resultado, las especies alfa 9, 6 y 5 se parecen más entre si que con respecto a alfa 7 y 10, lo cual concuerda con lo obtenido en el fenograma.

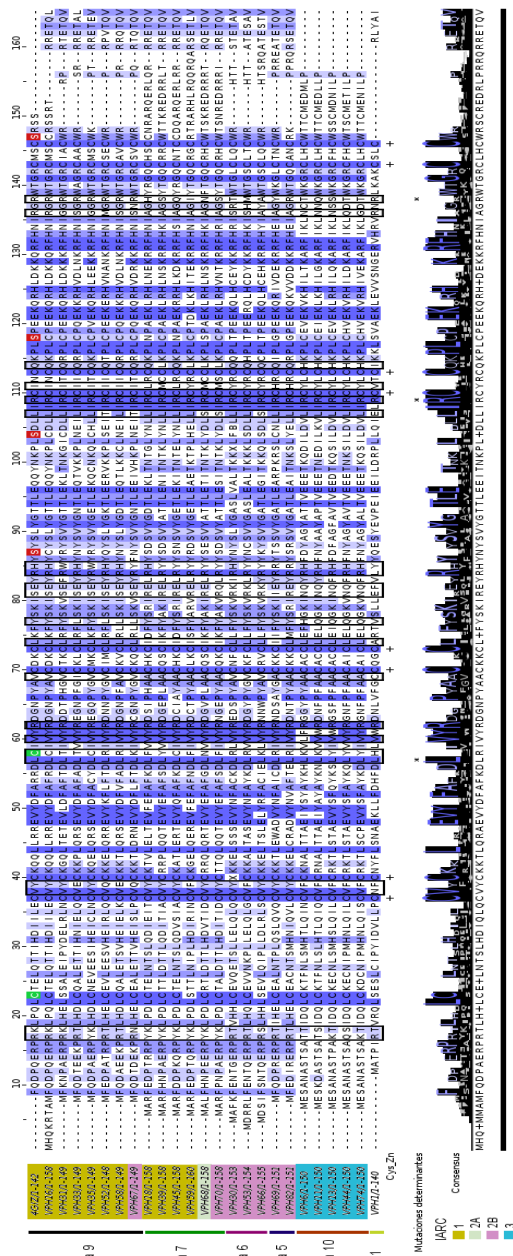


Figura 4. Alineamiento de la proteína E6 de diferentes tipos de VPH. Las secuencias están ordenadas en orden descendente según especie y capacidad cancerígena. El porcentaje de identidad se muestra en escala de intensidad de azules, azul intenso (identidad > 80%), azul medianamente intenso (entre 60 y 80 %) azul claro (entre 40 y 60 %) y blanco (< 40%). La primera secuencia corresponde a la proteína E6 del cristal del complejo E6/LxxLL (codigo 4GIZ en el PDB), en ella se resaltan los aminoácidos de cisteínas indispensables para la formación del complejo E6/LxxLL (verde) y la posterior degradación de la proteína p53, además de las cisteínas mutadas por serinas dispensable para la formación del complejo y degradación de p53 (Zanier et al., 2012), los residuos enmarcados en negro corresponde a los aminoácidos implicados en la interacción de E6 y LxxLL (Zanier et al., 2013).

Tabla 2. Número de mutaciones entre la secuencia de la estructura cristalográfica de la proteína E6 (4GIZ) y la E6 de otros tipos de VPH

Tipo	GenBank ID	Especie	IARC	No. Mutaciones (toda la secuencia)	No. Mutaciones (aminoácidos que interaccionan con la hélice LxxLL)
VPH16	NP_041325.1	Alfa 9	1	7	0
VPH31	AAA46950.1		1	53	3
VPH33	AAA46958.1		1	54	6
VPH35	CAA52561.1		1	42	3
VPH52	CAA52585.1		1	57	6
VPH58	BAA31845.1		1	55	5
VPH67	BAA66109.1		2B	59	7
VPH18	NP_040310.1	Alfa 7	1	70	6
VPH39	AAA47050.1		1	70	7
VPH45	CAA52573.1		1	70	9
VPH59	CAA54849.1		1	70	7
VPH68	ACX32377.1		2A	81	10
VPH70	AGU90872.1		2B	71	10
VPH30	ALT54643.1	Alfa 6	2B	70	10
VPH53	ACX32363.1		2B	79	5
VPH66	AAY69378.1		2B	77	7
VPH69	BAA90727.1	Alfa 5	2B	70	8
VPH82	BAA90735.1		2B	65	6
VPH6	NP_040296.1	Alfa 10	3	98	11
VPH11	AFC88060.1		3	100	10
VPH13	ABC79057.1		3	101	10
VPH44	ALT54746.1		3	95	10
VPH74	APX52883.1		3	98	10
VPH1	NP_040305.1	Mu 1	-	102	15

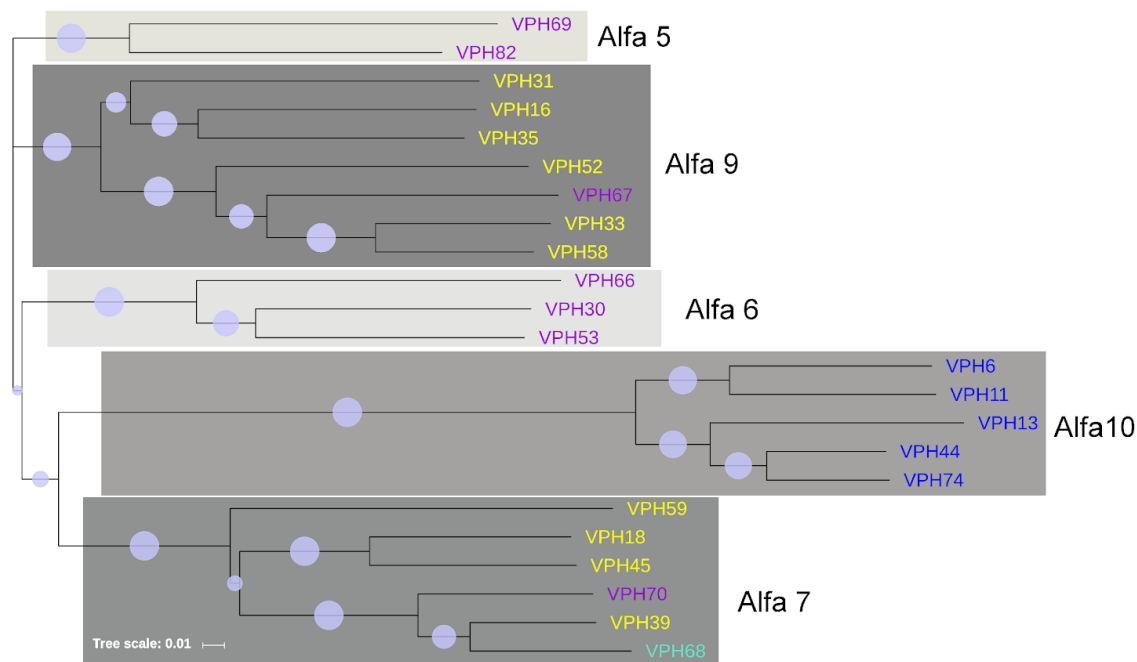


Figura 5. Fenograma de la proteína E6 de diferentes tipos de VPH. Haciendo uso del alineamiento con MUSCLE se construyó un fenograma mediante el método Neighbor Joining (NJ) y un Bootstrap de 1000. Las especies resultantes se muestran con diferentes tonalidades de gris. Los integrantes del grupo 1 (categorías IARC) de alto riesgo están escritos con letras amarillas, grupo 2A violeta, grupo 2B aguamarina y grupo 3 de bajo riesgo en azul. El bootstrap está representado con 3 tamaños diferentes de círculo: 85-100 % (círculo grande), 70-84% (círculo mediano), 50-69% (círculo pequeño). La longitud de la barra indica la p-distance.

5.1.3. Análisis de la estructura primaria de las E6

Por medio del uso de la herramienta en línea protparam de ExPASy se cuantificó la cantidad de cada uno de los 20 aminoácidos para cada una de las proteínas E6 estudiadas. Ésto se hizo a lo largo de toda la secuencia de las E6 (Figura 6A) y en la región comprendida entre los aminoácidos 17 y 146 a partir del alineamiento (Figura 4), dentro de esta región se encuentran los aminoácidos implicados en la interacción de la proteína E6 y el motivo LxxLL en VPH16 (Figura 6B). En la Figura 6A, la cantidad de cada aminoácido es muy parecida entre las proteínas E6, con algunas diferencias en los aminoácidos arginina, fenilalanina, leucina e isoleucina. Por un lado, los VPH de bajo riesgo contienen un porcentaje menor de argininas y mayor de fenilalaninas que el resto de las E6, mientras que la E6 de los tipos probable y posiblemente cancerígenos tienen mayor cantidad de leucinas y menor de isoleucina que el resto. Este

mismo patrón se observó al cuantificar los aminoácidos en la región libre de los C y N terminales (Figura 6B), con la adición del aminoácido serina, el cual es menor en la E6 de los VPH de bajo riesgo.

De igual forma, se determinó la estructura secundaria de cada proteína E6 por medio de la herramienta en línea Jpred. Tal como se muestra en la Figura 7. Como se puede ver, la estructura secundaria de las E6 está muy conservada entre los VPH, incluso y a pesar del gran número de mutaciones que hay entre ellas (Tabla 2). Por otro lado, se vió que gran parte de los aminoácidos implicados en la interacción entre la proteína E6 y el motivo LxxLL de VPH16 están comprometidos en la formación de las estructuras en hélice y láminas de las proteínas (aminoácidos sombreados). Un análisis adicional, del tipo de aminoácido implicado en la interacción E6/LxxLL de VPH16, y, cómo éste varía entre los distintos tipos de VPH arrojó (ver panel inferior de la Figura 7), que gran parte de los aminoácidos implicados son argininas, 6 de 21, y que 3 de ellos se conservan entre todos los tipos, en aquellos casos donde no se conserva, aparecen los aminoácidos ionizables: tirosina, serina, glutamina y asparagina y los aminoácidos hidrofóbicos pequeños: glicina, treonina y alanina, esto concuerda con la pérdida de argininas observada en el inciso anterior (Figura 6B); en general, se observa que el resto de los aminoácidos implicados son de carácter hidrofóbico, tales como valina, leucina, tirosina, fenilalanina e isoleucina.

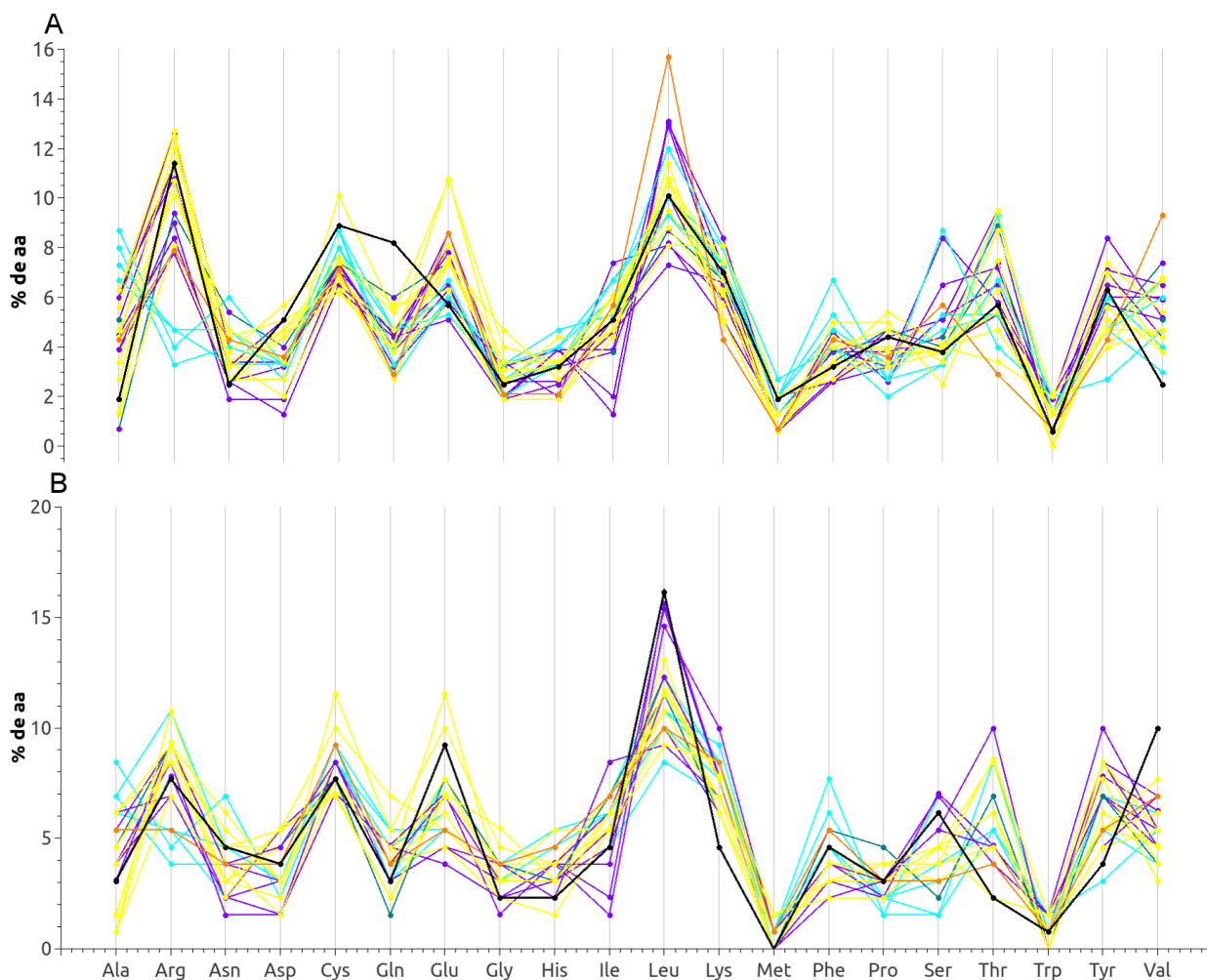


Figura 6. Representación gráfica de la diferencia en composición de aminoácidos de la proteína E6 para cada tipo de VPH. (A) porcentaje de los 20 aminoácidos a lo largo de toda la secuencia de cada E6, (B) porcentaje de los 20 aminoácidos en la región comprendida por los aminoácidos 17 y 146 obtenida a partir del alineamiento (Figura 4) de cada E6. Los integrantes del grupo 1 (categorías IARC) de alto riesgo están representados con líneas amarillas, grupo 2A violeta, grupo 2B aguamarina, grupo 3 de bajo riesgo en azul y el grupo control en naranja.

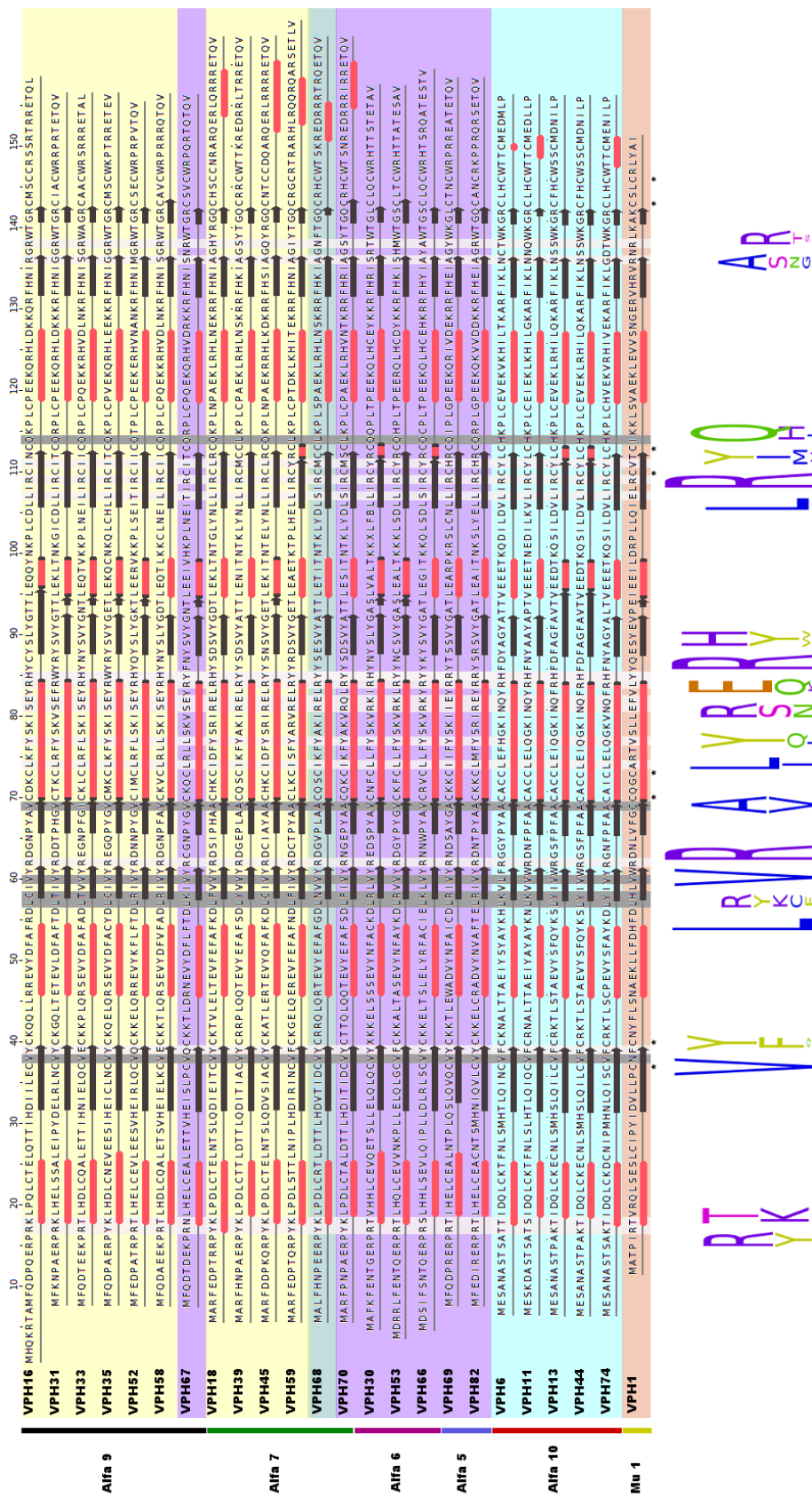


Figura 7. Representación gráfica de la estructura secundaria determinada por el programa Jpred de la proteína E6 para cada tipo de VPH. Las secuencias están ordenadas de forma descendente, en cuanto a la capacidad oncogénica según las categorías IARC (de la misma forma que en la figura 4), las cisteínas implicadas en la unión a átomos de zinc están marcadas con el símbolo (*), los aminoácidos sombreados en blanco y negro corresponden a los aminoácidos implicados en la interacción de E6 y LxxLL (Zanier et al., 2013). Panel inferior: logo de los aminoácidos implicados en la interacción: polares positivos (en morado), polares negativos (naranja), hidrofóbicos alifáticos (azul), hidrofóbicos aromáticos (amarillo) demás aminoácidos (fucsia). Las conformaciones en hélice se denotan como barras de color majenta y las láminas como flechas negras.

5.2. Modelado por enhebrado de las E6

5.2.1 Modelado por enhebrado con I-TASSER

La herramienta de modelado en línea I-Tasser identificó como mejor modelo para el enhebrado a la estructura cristalina de E6/LxxLL con el código de identificación 4GIZ del Protein Data Bank. De los modelos propuestos por I-Tasser (generalmente 5) se eligieron aquellos que tuvieron el menor valor de RMSD al comparar la estructura consulta y el cristal 4GIZ. En la Tabla 3 se muestran los valores del RMSD de las estructuras elegidas y a utilizar para los estudios posteriores en éste trabajo.

Tabla 3. RMSD de los modelos de las proteínas E6 a utilizar en el presente trabajo

Tipo	Especie	RMSD (Å)
VPH16	Alfa 9	-
VPH31		-
VPH33		2.4±1.8
VPH35		2.2±1.7
VPH52		2.3±1.8
VPH58		2.3±1.8
VPH67		2.3±1.8
VPH18	Alfa 7	3.6±2.5
VPH39		3.9±2.7
VPH45		3.4±2.3
VPH59		3.8±2.6
VPH68		3.9±2.7
VPH70	3.4±2.4	
VPH30	Alfa 6	2.6±1.9
VPH53		2.9±2.1
VPH66		3.2±2.3
VPH69	Alfa 5	3.0±2.1
VPH82		3.0±2.1
VPH6	Alfa 10	2.1±1.7
VPH11		2.6±1.9
VPH13		2.1±1.7
VPH44		2.3±1.7
VPH74		2.3±1.8
VPH1	Mu 1	1.3±1.3

5.3. Modelado por dinámica molecular (DM) del complejo E6/LxxLL

5.3.1. Preparación del complejo E6/LxxLL

Previo a la dinámica molecular, cada modelo de E6 en formato .pdb fue sometido a varios programas que lo iban adecuando, en estructura y lenguaje, al programa AMBER16 (programa a utilizar para la simulación). El flujo de trabajo fue el siguiente: desprotonación de la proteína - ajuste de los ángulos de torsión de las cisteínas de los dedos de zinc - adición de los dos átomos de Zn y la hélice LxxLL (Figura 9) - cálculo del estado de protonación de todos residuos de cada proteína - modificación de los nombres de las histidinas, cisteínas y los dos átomos de zinc - producción de los archivos de topología .prmtop y de coordenadas .inpcrd del complejo E6/LxxLL. En todos los casos, los complejos E6/LxxLL fueron solvatados en una caja de agua de 15 Å y cada sistema se neutralizó con iones Cl⁻ o Na⁺. Para ver más detalles de los nombres de los programas ver la sección 3.1 de la metodología.

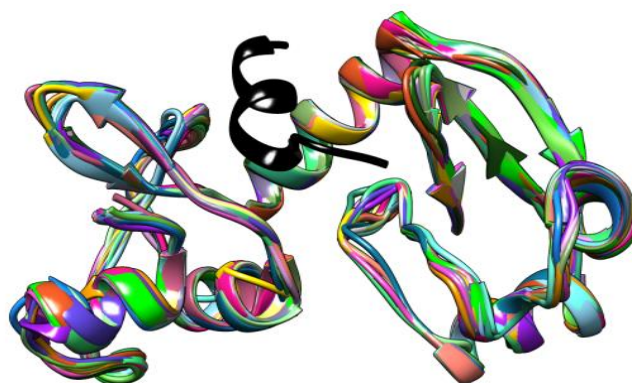


Figura 8. Representación gráfica de las 24 proteínas E6 superpuestas con el complejo E6/LxxLL descrita en la estructura cristalográfica 4GIZ. En la representación se eliminaron los extremos C y N.

5.3.2. Minimización de la energía

Con la intención de eliminar conflictos atómicos antes del proceso de modelado por DM, se realizó un paso de relajación del sistema mediante los métodos de minimización de energía: steepest descent (para los primeros 5000 pasos) y conjugate gradient (para los últimos 5000 pasos). Para ejemplificar esto,

se tomo el resultado de uno de los 24 sistemas E6/LxxLL. En la Figura 9, se muestra como la energía potencial del sistema llegó a descender -160000 Kcal/mol tras el ciclo de minimización.

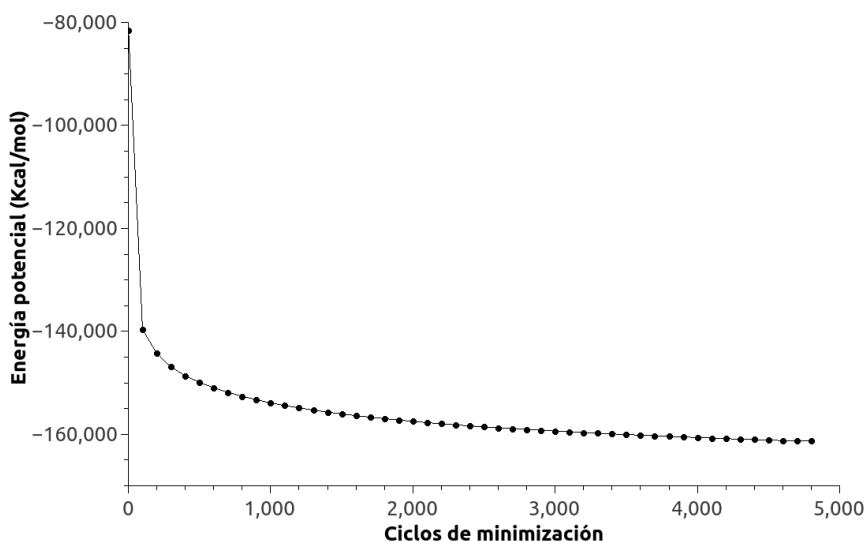


Figura 9. Minimización de la energía del sistema E6/LxxLL de VPH16. Descenso de la energía potencial tras el ciclo de minimización

5.3.3. Recocido simulado y equilibrado

El objetivo del uso del método de recocido simulado es optimizar el proceso de minimización de la energía de cada sistema E6/LxxLL mediante una simulación por DM y explorando conformaciones energeticamente improbables en la naturaleza, lo que permite evadir que el sistema caiga en un mínimo local. Para ilustrar este paso se tomó como ejemplo el resultado en los cambios de la temperatura y energía potencial, cinética y total de uno de los sistemas E6/LxxLL. Cabe acotar que el resto de los sistemas presentó el mismo comportamiento que éste. En las Figura 10, se muestra como a medida que varía la temperatura, lo hacen las energías cinética y potencial, tanto en los primeros 2,5 ns de simulación, que corresponden al recocido simulado, como en los 5.5 ns de la fase de equilibrado. La variación, en un pequeño intervalo, de los valores de energía durante la fase de equilibrado demuestran que hay estabilidad conformacional.

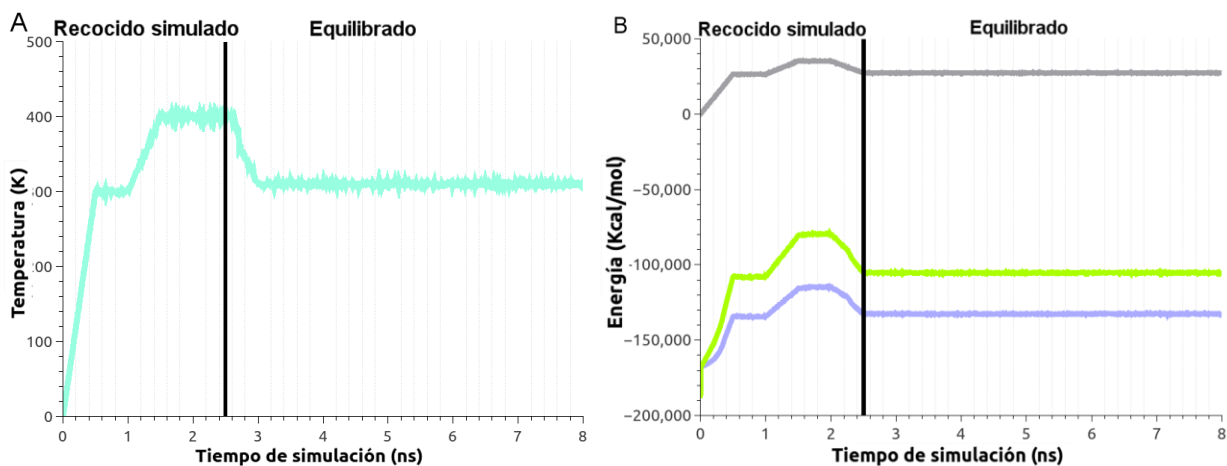
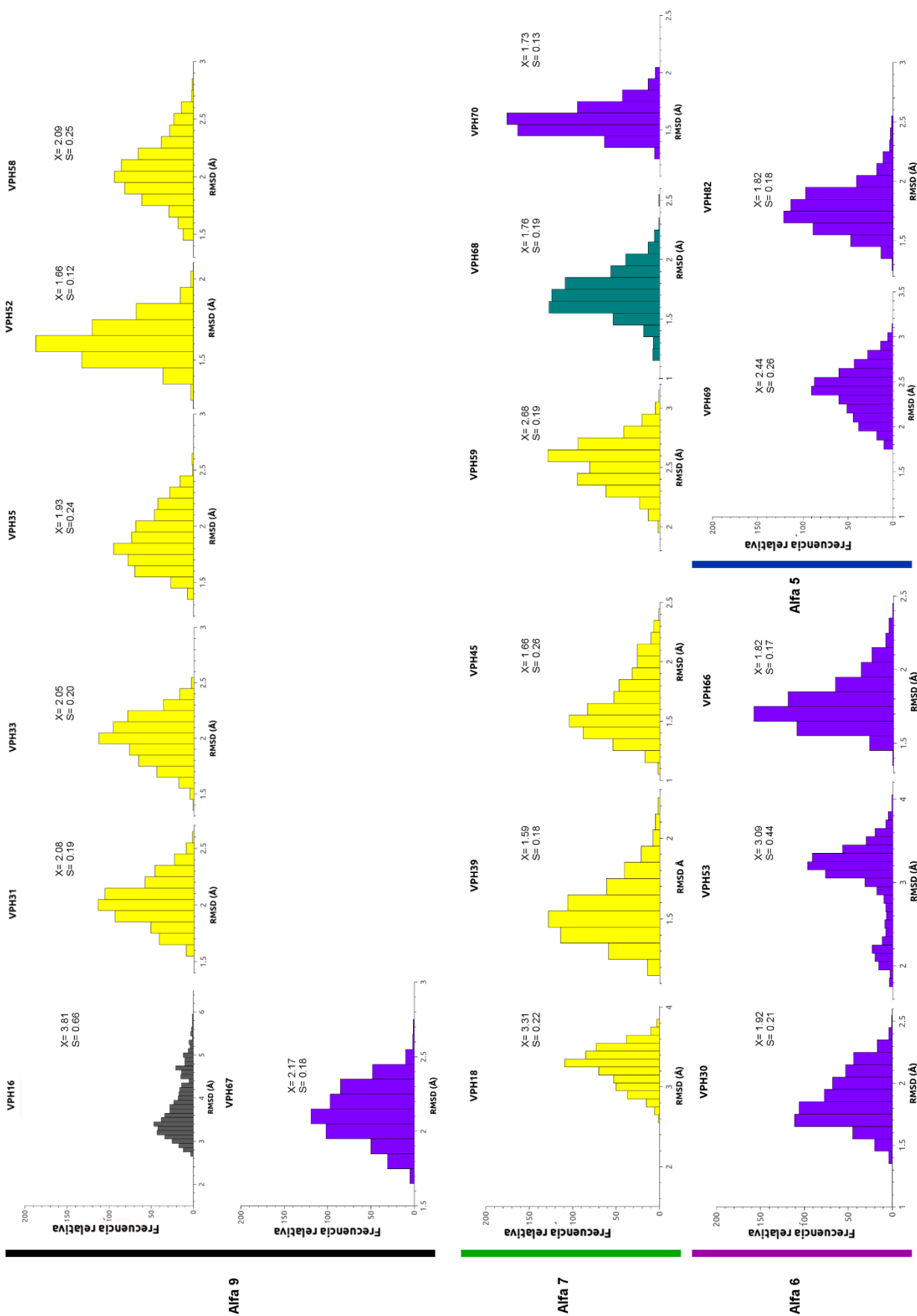


Figura 10. Recocido simulado y equilibrado del sistema E6/LxxLL de VPH16. (A) cambios en la temperatura, (B) cambios en los valores de energía potencial (azul), cinética (verde) y total (gris) del sistema. La fase de recocido fue de 2.5 ns y la de equilibrado de 5.5 ns.

5.3.4. Evaluación del equilibrio

Para corroborar que el sistema posee estabilidad conformacional antes de comenzar la fase de producción se calculó el RMSD entre la estructura del primer frame y las obtenidas durante la dinámica para cada sistema E6/LxxLL. El cálculo se realizó únicamente a los carbonos α sin los extremos C y N de las secuencias. En la Figura 11 se muestra la frecuencia relativa de los valores de RMSD para cada sistema. Como se puede ver, en todos los sistemas los valores de RMSD se agrupan principalmente en un intervalo central de valores, y los valores de desviación estándar (S) son bajos. Con lo que podemos concluir, que no hay cambios conformacionales significativos durante la simulación, por lo tanto, hay estabilidad conformacional.



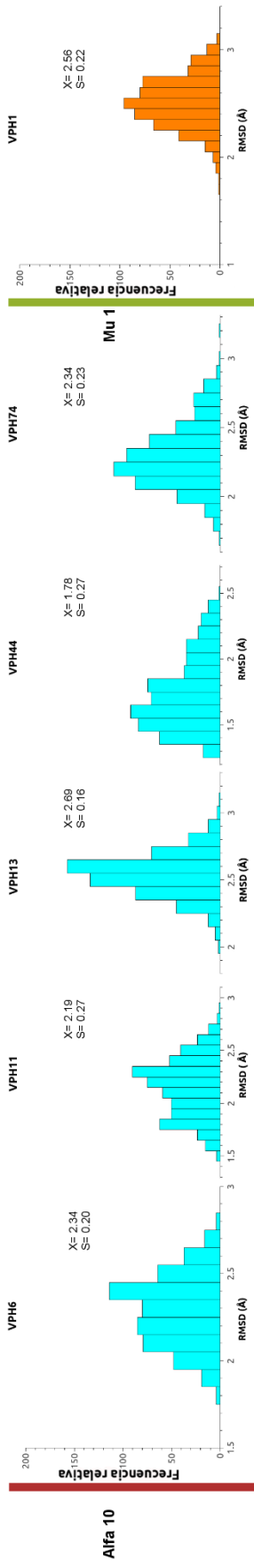


Figura 11. Valores de RMSD para cada sistema E6/LxxLL en la fase de equilibrado. Las gráficas representan la frecuencia relativa de los valores de RMSD para cada sistema E6/LxxLL, para los diferentes tipos de VPH estudiados. En color amarillo, corresponden a los VPH de categoría **1**: cancerígeno para humanos, aguamarina **2A**: probablemente cancerígeno para humanos, violeta **2B**: posiblemente cancerígeno para humanos y azul **3**: no clasificable. VPH16 está representado en negro. X= media, S= desviación estándar.

5.3.5. Simulación por dinámica molecular

Se evaluaron algunas propiedades físicas durante la fase de DM. Para ilustrar esto se tomó como ejemplo el resultado en los cambios de los valores de temperatura, volumen, presión y energía potencial, cinética y total de uno de los sistemas E6/LxxLL. Los valores de cada propiedad fueron registrados cada 0.02 ns durante los 100 ns que duró la DM. En la Figura 12 se muestran los resultados. Como se puede observar, todas las propiedades y energías del sistema se mantienen estables y cercanas al valor esperado, según las especificaciones dadas al programa AMBER16 para esta fase, con temperaturas cercanas a 310 K y condiciones isobáricas e isocóricas.

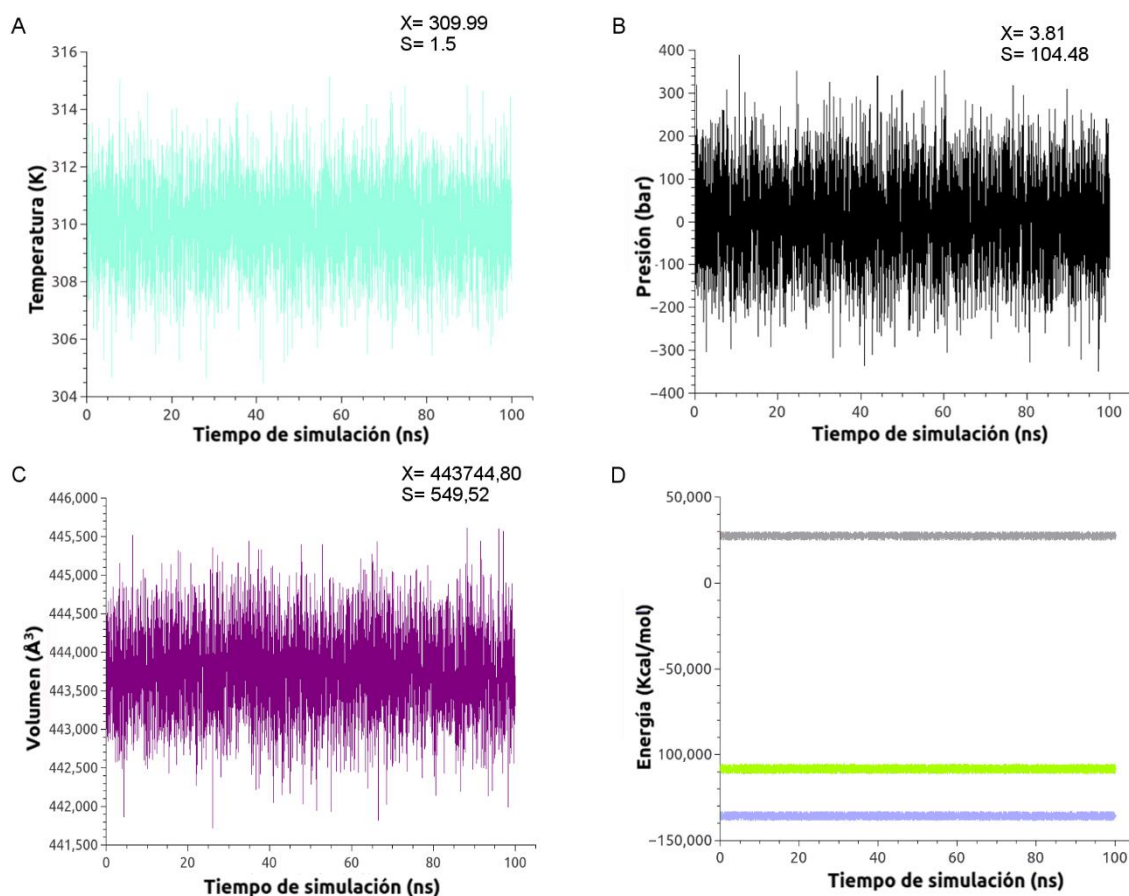


Figura 12. Propiedades físicas del sistema E6/LxxLL de VPH16 durante la fase de DM. Fluctuación de la (A) temperatura; (B) presión; (C) volumen, (D) energía potencial (azul), cinética (verde) y total (gris) a lo largo de 100 ns de simulación.

5.3.6. Análisis de las trayectorias obtenidas en la fase de DM

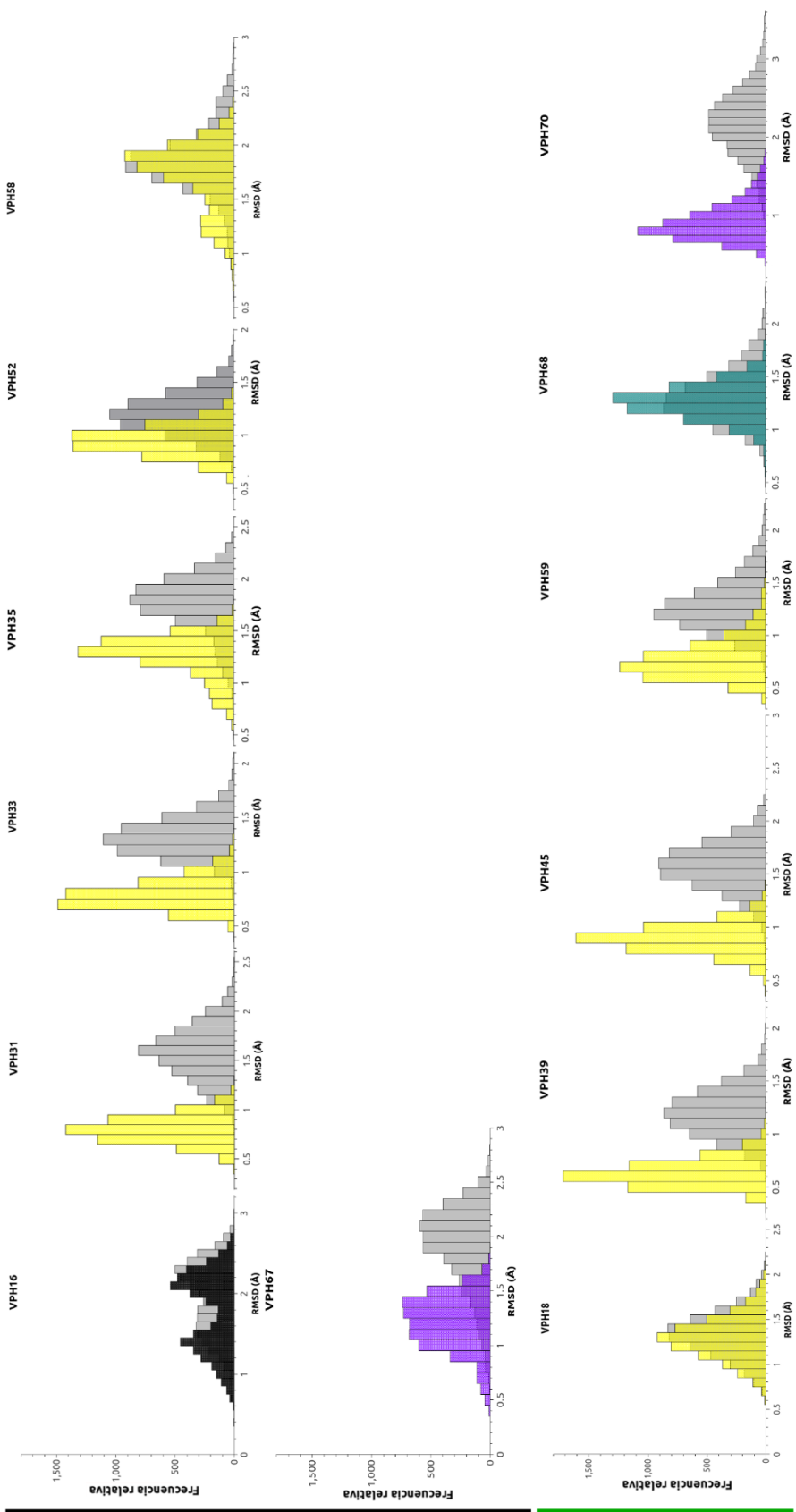
5.3.6.1. RMSD (root-mean-square deviation)

Se calculó el RMSD entre la estructura del primer frame contra todas las obtenidas durante los 100 ns de DM a: (1) los C α , sin los extremos C y N de la proteína E6 de cada sistema E6/LxxLL y (2) a los C α de los aminoácidos, en la proteína E6, localizados a 5 Å del motivo LxxLL, de cada sistema E6/LxxL. Para determinar dichos aminoácidos se utilizó la estructura de Boltzmann (estructura más probable desde el punto de vista energético) de cada sistema E6/LxxLL (para más detalles ver inciso 3.6.6 de metodología y resultados). Tal como se muestra en la figura 3. La proteína E6 para la mayoría de los sistemas tiene una conformación geométrica predominante a lo largo de la simulación (puesto que los datos se agrupan en un solo conjunto), a excepción de los tipos VPH16 y VPH66 cuya frecuencia relativa de los valores de RMSD se agrupan en 2 conjuntos, éste mismo comportamiento se observa para los aminoácidos adyacentes al motivo LxxLL (valores de correlación positivos) pero con medidas de desplazamiento más cercanos a cero.

5.3.6.2. RMSF (root-mean-square fluctuation)

De igual forma, en la figura 14, se muestran la fluctuación en Å de los residuos de aminoácidos de la proteína E6 para cada sistema E6/LxxLL. Se observa que la fluctuación de los aminoácidos localizados en la región implicada en la interacción E6/LxxLL para VPH16 (Zanier et al., 2013) es baja, lo mismo pasa para el resto de las E6, con la excepción de los aminoácidos 55, 56 y 114 en el tipo VPH66, posiblemente cancerígeno, y los aminoácidos 136 y 138 en tipos de VPH posiblemente cancerígenos y de bajo riesgo.

Alfa 9



TIPO	TOTAL EG/LxxLL			POCKET EG/LxxLL		
	X	S	r	X	S	r
VPH16	1.98	±0.46	1.81	±0.49	0.91	
VPH31	1.62	±0.27	0.85	±0.13	0.61	
VPH33	1.37	±0.17	0.84	±0.13	0.24	
VPH35	1.81	±0.26	1.31	±0.2	0.92	
VPH52	1.25	±0.19	1	±0.14	0.50	
VPH58	1.89	±0.30	1.76	±0.31	0.24	
VPH67	1.94	±0.39	1.25	±0.24	0.80	
VPH18						
VPH39						
VPH45						
VPH59						
VPH68						
VPH70						

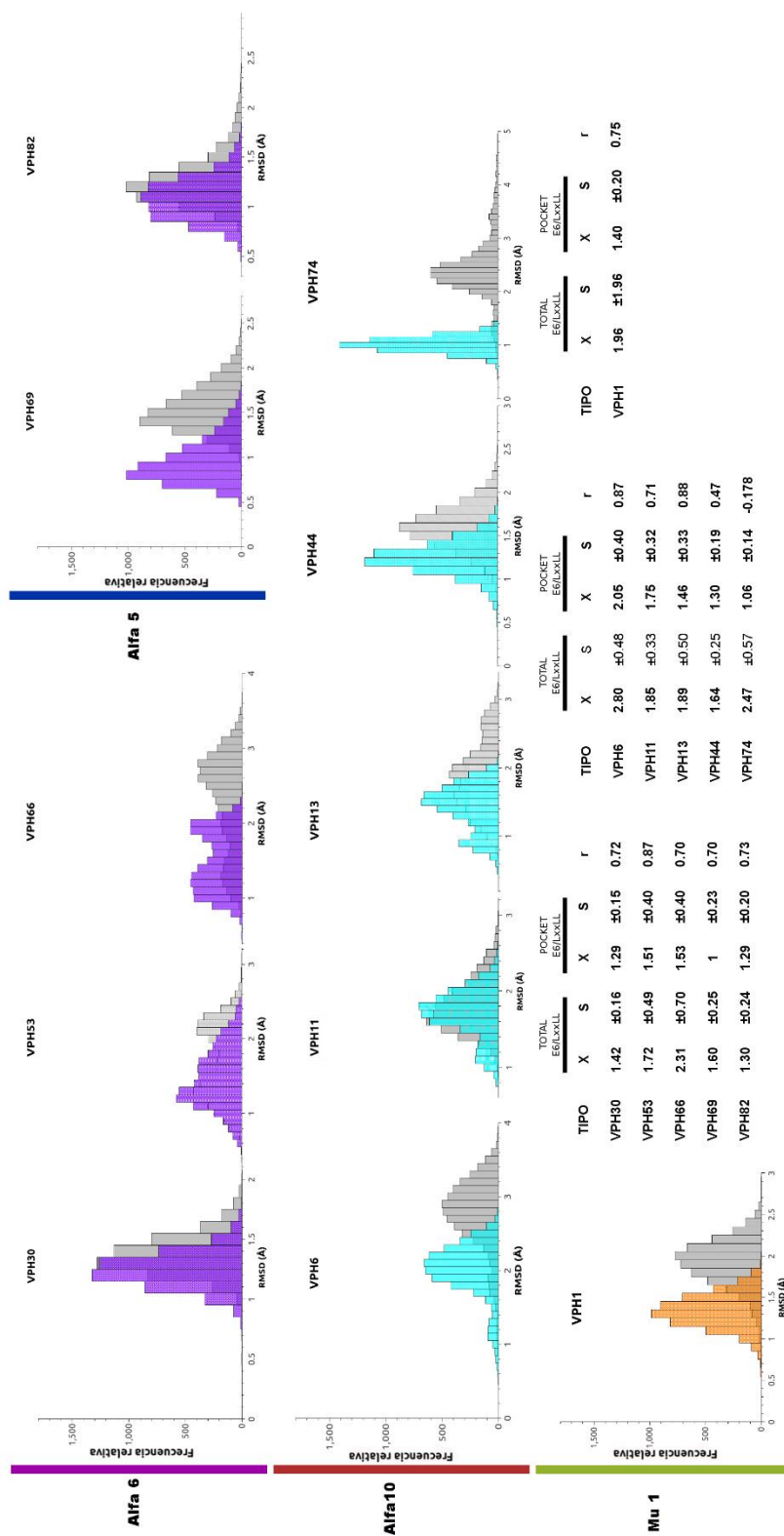


Figura 13. Valores de RMSD para cada sistema E6/LxxLL durante la DM. Las gráficas representan la frecuencia relativa de los valores de RMSD para cada sistema E6/LxxLL de los diferentes tipos de VPH estudiados. En color amarillo, se representan a los VPH de la categoría: **1**: cancerígeno para humanos, aguamarina **2A**: probablemente cancerígeno para humanos, violeta **2B**: posiblemente cancerígeno para humanos y azul **3**: no clasificable; de las categorías IARC. VPH16 está representado en negro y el control negativo en naranja. Los histogramas en color gris corresponden al RMSD de las E6 sin extremos C y N y los coloreados al RMSD de los aminoácidos adyacentes al motivo LxxLL. En el panel inferior están tabulados los valores de (x) promedio, (s) desviación estándar y (r) correlación para los RMSD de toda la E6 y los aminoácidos adyacentes.

5.3.6.3. Enlaces de hidrógeno

Debido a que las interacciones proteína-proteína presentan una fuerte dependencia de los enlaces de hidrógeno, se cuantificaron el número de enlaces de hidrógeno formados para cada sistema E6/LxxLL entre: la proteína E6 y el motivo LxxLL (ensamble 1) y el motivo LxxLL y el agua circundante a 5 Å (ensamble 2), durante los 100 ns de simulación. Los resultados se muestran en la Figura 15. Al comparar el promedio de número de enlaces de hidrógeno formados en el ensamble 1, entre la E6 de VPH16 y el resto de los tipos de VPH se observa que los tipos que corresponden a las especies alfa 9 y 5 presentan, en promedio, la misma cantidad de enlaces de hidrógeno, y que los tipos que corresponden a las especies alfa 7, 6, 10 y Mu 1, tienen en promedio menor cantidad de enlaces de hidrógeno. Por otra parte, con el ensamble 2, se observa que en la mayoría de los sistemas el motivo LxxLL interactúa más con el agua que con la proteína E6, y que en los casos donde se formaban menos enlaces entre la E6 y LxxLL, el motivo interactuaba más con el agua

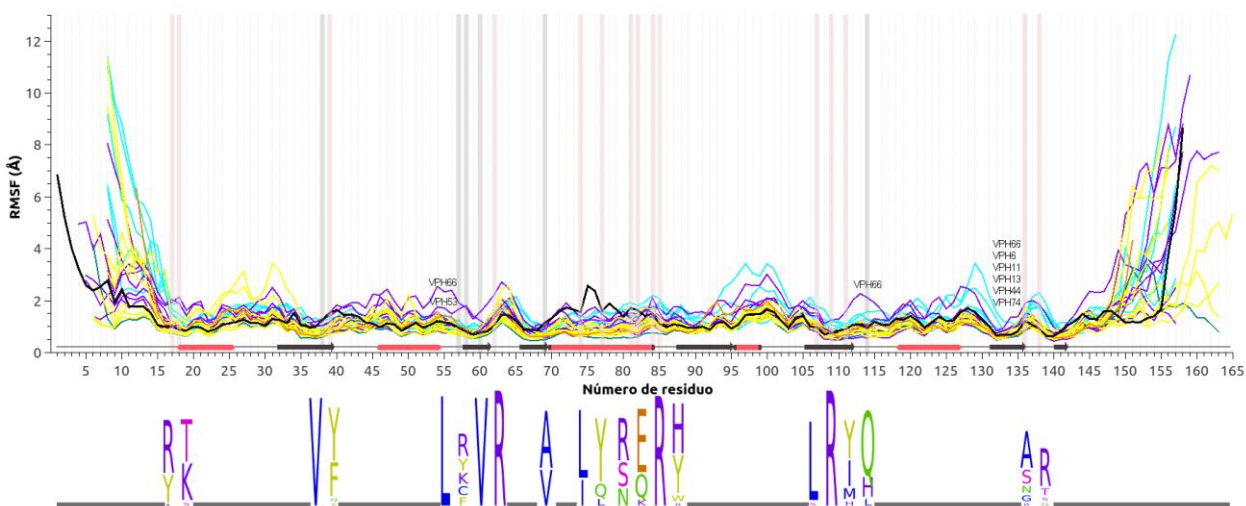


Figura 14. RMSF para cada sistema E6/LxxLL durante la DM. Valores de RMSF de cada uno de los aminoácidos de la proteína E6 de los diferentes tipos de VPH. Los integrantes del grupo 1 (categorías IARC) de alto riesgo están representados con líneas amarillas, a excepción de VPH16 en negro, grupo 2A violeta, grupo 2B aguamarina, grupo 3 de bajo riesgo en azul y el grupo control en naranja. Los aminoácidos sombreados en rosa y gris corresponden a los aminoácidos implicados en la interacción de E6 y LxxLL (Zanier et al., 2013). Panel inferior: logo de los aminoácidos implicados en la interacción: polares positivos (en morado), polares negativos (naranja), hidrofóbicos alifáticos (azul), hidrofóbicos aromáticos (amarillo) demás aminoácidos (fucsia).

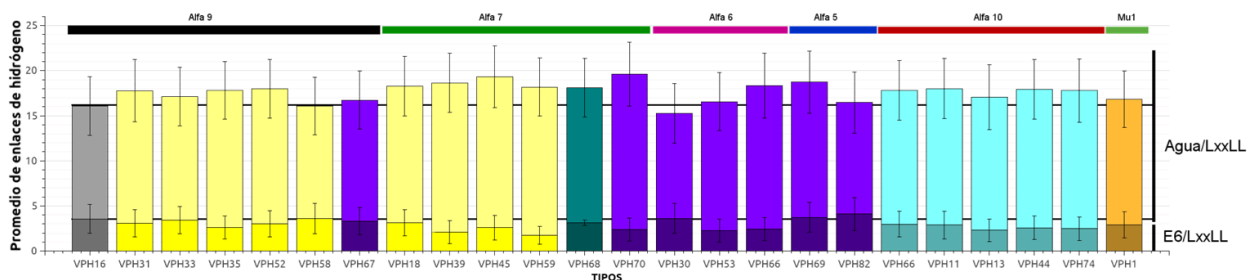


Figura 15. Promedio de enlaces de hidrógeno para cada sistema E6/LxxLL durante la DM. Las barras con valores < 5 corresponden al promedio de enlaces de hidrógeno que se forman entre cada proteína E6 y el motivo LxxLL y las barras con valores ≥ 15 corresponde al promedio de enlaces de hidrógeno entre el motivo LxxLL y el agua circundante a 5 Å. Los integrantes del grupo 1 (categorías IARC) de alto riesgo están representados con color amarillo, a excepción de VPH16 en gris, grupo 2A violeta, grupo 2B aguamarina, grupo 3 de bajo riesgo en azul y el grupo control en naranja.

5.3.6.4. Contactos nativos

Los resultados del análisis de contactos nativos corresponden a tablas de valores que indican: que aminoácidos de la proteína E6 y el motivo LxxLL estuvieron en contacto a lo largo de los 100 ns simulación, y a su vez, en cuántos de los 5000 frames de la simulación ocurrieron los contactos. Dichas tablas fueron utilizadas para: (1) corroborar que la E6 de VPH16 de nuestro sistema tenía las mismas interacciones con el motivo LxxLL que las reportadas para el cristal (Zanier et al., 2013) , los resultados indicaron que 19 de los 21 contactos reportados eran iguales, al menos en un 70 % del tiempo de simulación, a excepción de los aminoácidos Leu 67 y Ser 74, quienes no interaccionaron con ningún aminoácido del motivo LxxLL en toda la simulación (datos no mostrados); y (2) para reconocer cuáles aminoácidos de las diferentes E6 en este estudio interaccionaban de forma común, al menos en un 70 % del tiempo de simulación, con el motivo LxxLL. Los resultados se muestran en las figuras 7 y 14 en forma de aminoácidos sombreados en gris; estos son: los aminoácidos en la posición 38, 57, 58, 60, 69 y 114.

5.3.6.5. Distancia entre centros de masa

La medida de la distancia entre los centros de masa de dos proteínas permite analizar de forma indirecta la fuerza con la que ocurre la interacción entre ellas. En la figura 16 se muestran los resultados del cálculo de los centros de masa para cada sistema E6/LxxLL. Al comparar el promedio de la distancia

entre la E6 y el motivo LxxLL de VPH16 y el resto de los tipos de VPH, se observa, que los tipos que corresponden a la especie Alfa 9 presentan valores más bajos que el resto de las especies. Lo mismo ocurre con los integrantes de la especie Alfa 10.

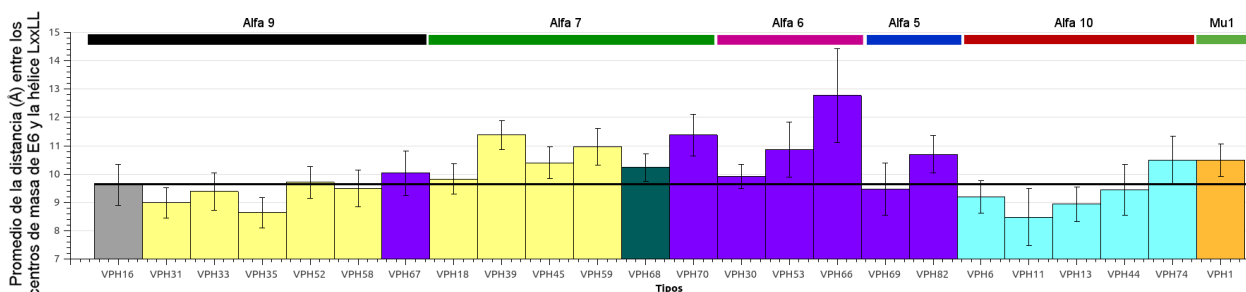


Figura 16. Promedio de la distancia entre los centros de masa de las proteínas E6 y el motivo LxxLL para cada sistema E6/LxxLL durante la DM. Los integrantes del grupo 1 (categorías IARC) de alto riesgo están representados con color amarillo, a excepción de VPH16 en gris, grupo 2A violeta, grupo 2B aguamarina, grupo 3 de bajo riesgo en azul y el grupo control en naranja.

5.3.6.6. Análisis por distribución de Boltzmann

Para elegir la conformación más frecuente, desde el punto de vista energético, a partir de cada sistema E6/LxxLL; se ajustaron los valores de energía total (Kcal/mol) de cada frame, de las simulaciones, a la distribución de Boltzmann (Ec. 4). A partir de dicha distribución se extrajo una conformación por sistema (la más frecuente) a la que llamaremos estructura de Boltzmann.

5.4. Cálculo de la energía libre de interacción entre la proteína E6 y la hélice LxxLL

Para el cálculo de la energía libre de interacción (ΔG) entre la proteína E6 y el motivo LxxLL de cada sistema, se utilizaron dos métodos, el MM/GBSA y el MM/PBSA. Cabe aclarar, que si bien ambos métodos consideran la contribución entrópica ($-T\Delta S$) para el cálculo (Ec.6), en el presente trabajo se supuso un valor de $\Delta S = 0$, esto debido al costo computacional de su cálculo y a la baja precisión que ha mostrado en estudios previos (Chen et al., 2018). Por lo tanto, los valores de ΔG calculados en el presente trabajo, no

son comparables con valores experimentales, solo pueden ser utilizados como punto de comparación entre las distintas E6 de los distintos tipos de VPH.

En las figuras 17 y 18 se muestran los resultados de ΔG por los métodos de MM/GBSA y MM/PBSA respectivamente. Al comparar el ΔG de unión del sistema E6/LxxLL de VPH16 (barra gris) con el resto de los tipos de VPH, se observa, que los tipos de VPH que integran las especies Alfa 9 y 5 presentan valores de ΔG promedio más bajos que el resto de los tipos, es decir, la interacción ocurre de forma más favorable en la E6 en éstas especies que para el resto. Caso contrario a los tipos de VPH que integran las especies Alfa 7, 6, 10 y Mu1 cuyos valores de ΔG son más altos y presentan menor dispersión entre sí, lo que indica que la interacción no ocurre, o es menos favorable. La misma tendencia se observa por los ΔG calculados por el método de MM/PBSA.

De forma adicional, se analizó el componente electrostático (Δe_{elect}) y de Van Der Waals (Δe_{vdw}) además de la contribución de las argininas para cada sistema E6/LxxLL a lo largo de los 100 ns de simulación. Como punto de referencia, los valores de ambos componentes fueron obtenidos a partir de la suma de la contribución de cada aminoácido a 5 y 10 Å del motivo LxxLL de las estructuras de Boltzmann. Los resultados se muestran en las figuras 19 y 20. Se observa que los tipos de VPH que mostraron una mayor contribución del componente Δe_{elect} a 5 Å, tienen valores bajos del componente Δe_{vdw} , y viceversa. Además, las argininas tienen gran contribución en el Δe_{elect} en todos los tipos de VPH a excepción de los de bajo riesgo, lo que concuerda con el resultado obtenido en cuanto a porcentaje de aminoácidos en la Figura 6.

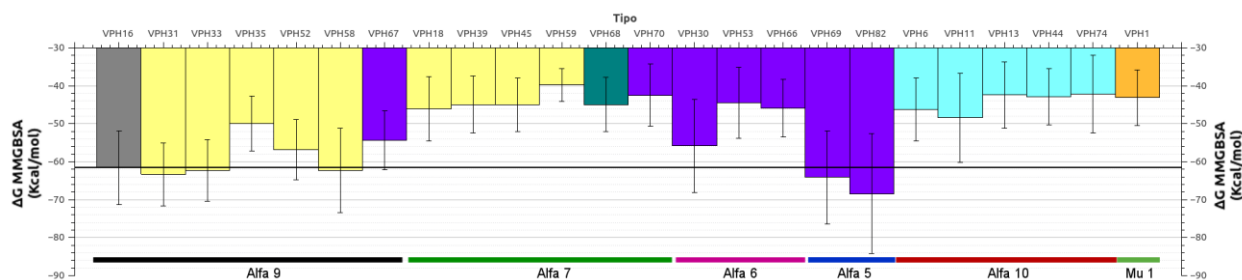


Figura 17. Valores de ΔG por MM/GBSA entre las proteínas E6 y el motivo LxxLL para cada sistema E6/LxxLL durante la DM. Los integrantes del grupo 1 (categorías IARC) de alto riesgo están representados con color amarillo, a excepción de VPH16 en gris, grupo 2A violeta, grupo 2B aguamarina, grupo 3 de bajo riesgo en azul y el grupo control en naranja.

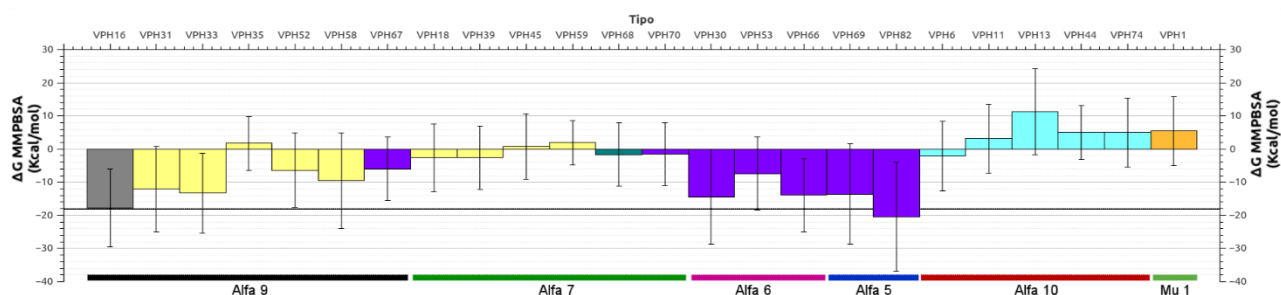


Figura 18. Valores de ΔG por MM/PBSA entre las proteínas E6 y el motivo LxxLL para cada sistema E6/LxxLL durante la DM. Los integrantes del grupo 1 (categorías IARC) de alto riesgo están representados con color amarillo, a excepción de VPH16 en gris, grupo 2A violeta, grupo 2B aguamarina, grupo 3 de bajo riesgo en azul y el grupo control en naranja.

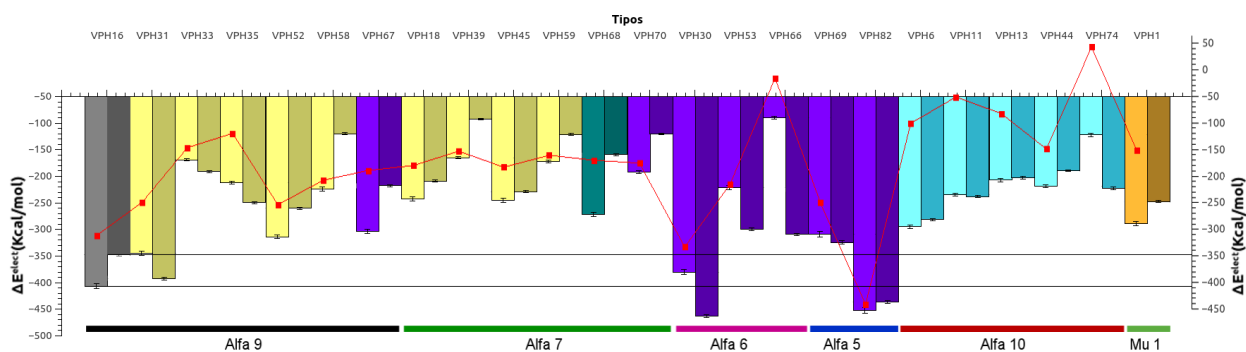


Figura 19. Contribución del componente ΔE_{elect} en la interacción entre las proteínas E6 y el motivo LxxLL para cada sistema E6/LxxLL durante la DM. Los integrantes del grupo 1 (categorías IARC) de alto riesgo están representados con color amarillo, a excepción de VPH16 en gris, grupo 2A violeta, grupo 2B aguamarina, grupo 3 de bajo riesgo en azul y el grupo control en naranja. Los puntos rojos indican los valores de contribución de las argininas a 5Å.

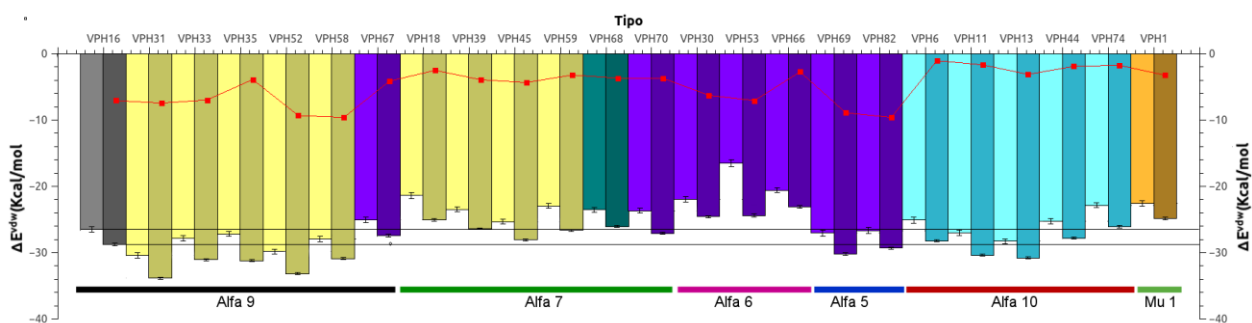


Figura 20. Contribución del componente ΔE_{vdw} en la interacción entre las proteínas E6 y el motivo LxxLL para cada sistema E6/LxxLL durante la DM. Los integrantes del grupo 1 (categorías IARC) de alto riesgo están representados con color amarillo, a excepción de VPH16 en gris, grupo 2A violeta, grupo 2B aguamarina, grupo 3 de bajo riesgo en azul y el grupo control en naranja. Los puntos rojos indican los valores de contribución de las argininas a 5Å.

Capítulo 6. Discusión

6.1. Elección y análisis de las proteínas E6 a estudiar

Para estudiar el fenómeno de interacción entre la proteína E6 del VPH y la proteína E6AP celular, y además analizar la posible relación entre la fuerza de dicha interacción (energía libre de interacción) y el potencial del virus de causar cáncer en la célula infectada, se eligió utilizar 24 secuencias de la proteína E6 a partir de 24 tipos diferentes de VPH. Para la selección de los VPH se utilizó como criterio la combinación de dos tipos de clasificación del virus, la clasificación propuesta por la agencia internacional para la investigación del cáncer (IARC: por sus siglas en inglés) cuyo fundamento es de tipo clínicos-epidemiológicos y experimental (Bravo and Félez-Sánchez, 2015) y la clasificación basada en el estudio comparativo del ORF de la proteína L1 (De Villiers et al., 2004b).

A continuación los tipos de VPH elegidos: de las categorías IARC: grupo **1**, en éste se encuentran aquellos tipos de VPH de los que se tiene suficiente evidencia de ser cancerígenos tanto en humanos como en animales experimentales, también llamados VPH de alto riesgo (Allison and Maleki, 2016), de los 12 tipos de VPH del grupo se eligió utilizar la proteína E6 de 10 (VPH: 16, 18, 31, 33, 35, 39, 45, 52, 58, 59) grupo **2A**, en este se encuentran los tipos de VPH llamados probablemente cancerígenos, y corresponden a aquellos que no se tiene suficiente evidencia de ser cancerígenos en humanos pero sí en animales experimentales, del grupo, se trabajó con la E6 de su único integrante, el VPH68; grupo **2B**, los posiblemente cancerígenos, corresponden a aquellos que no tienen suficiente evidencia de ser cancerígenos tanto en humanos como animales experimentales. De los 7 tipos del grupo, se eligió trabajar con la proteína E6 de 5 (VPH:53, 66, 67, 70 y 82), a su vez, hay 5 tipos más clasificados en éste grupo, pero no por ser cancerígenos sino por la relación filogenética que tienen con los integrantes del grupo 1, de los 5 se eligió trabajar con la E6 de 2 de ellos (VPH: 30 y 69) y por último, en el grupo **3**, se encuentran aquellos tipos de VPH que no se tiene, ni suficiente ni adecuada, evidencia de ser cancerígenos tanto en humanos como animales experimentales. Se trabajó con la E6 de los dos tipos de VPH del grupo (VPH:6 y 11); los integrantes de éste último grupo también son llamados de bajo riesgo, puesto que pueden infectar tejido mucoso, tal como los VPH del resto de los grupos, pero no causan cáncer, por su parte, pueden ocasionar lesiones benignas tales como condilomas y verrugas (Allison and Maleki, 2016); en consecuencia, y en vista de los pocos integrantes del grupo, se eligió adicionar la E6 de algunos tipos de VPH que tuvieran relación filogenética con VPH6 y 11 según la clasificación hecha con la ORF de la proteína L1, estos fueron VPH: 13, 44 y 74. Para contrastar los resultados, se utilizó la E6 del tipo VPH1, el cual no infecta tejido cutáneo, no

es cancerígeno y está clasificado (según el ORF de la proteína L1) en un género diferente al resto de los VPH del estudio, en los Mupapilomavirus (Danos, Katinka, and Yaniv, 1982) .

Las secuencias de la proteína E6 de cada tipo de VPH se extrajeron de la base de datos curada PaVE (Papillomavirus Episteme, por su nombre en inglés), la cual ha sido catalogada como una base de datos especializada, de secuencias del virus de VPH, altamente organizada y actualizada (Van Doorslaer et al., 2017). Además, corresponde a la base de datos promovida por el instituto nacional de alergias y enfermedades infecciosas (NIAID, por sus siglas en inglés) componente del instituto nacional de salud (NIH, por sus siglas en inglés) de Estados Unidos.

Con las secuencias elegidas se realizó un alineamiento múltiple haciendo uso del algoritmo MUSCLE (Figura 4), cuyo resultado se utilizó para la construcción de un fenograma por medio del método Neighbor Joining, con un bootstrap de 1000 y el método de sustitución p-distance (Figura 5). En el alineamiento se observó, que aunque se tratase de una misma proteína a partir de individuos de una misma especie (los Alfapapilomavirus), el número de aminoácidos con una identidad mayor a 80 % es muy bajo. Lo que se debe a que la tasa de mutación de los virus, en general, es superior a la tasa de mutación del huésped, de 5-10 veces más; principalmente debido al tamaño pequeño del genoma y al corto tiempo generacional (Drake, 1991; Lynch, 2010). De igual forma, ésto se ve reflejado en el número total de mutaciones tanto a lo largo de las secuencias de E6 como en la región de interacción entre E6 y E6AP, de VPH16 (ver Tabla 2). Dentro de los pocos aminoácidos conservados destacan: las 8 cisteínas implicadas en la unión a 2 átomos de zinc, 4 residuos por átomo, que conforman los 2 dominios de unión a zinc característicos de la estructura de la mayoría de los Pvs de mamífero (Scheffner et al., 1990) y los aminoácidos leucina 57 y arginina 109, los cuales han sido reportados, en ensayos in vitro e in vivo, como esenciales para la interacción de la E6 de VPH16 y la E6AP (ver Figura 4). Curiosamente, la arginina 138 que también fue reportada por éste mismo grupo como esencial en la interacción, se conserva solo en la especie Alfa 9 (misma donde se encuentra VPH16) y Mu 1 (control negativo); y las cisteínas 23 y 58 cuyos mutantes de serina demostraron ser ineficiente para la degradación de p53 no están conservadas (ver Figura 4 y Figura 3 de (Zanier et al., 2013).

Por otro lado, los distintos tipos de VPH de nuestro estudio quedaron clasificados dentro de las mismas especies que las determinadas con la proteína L1 (Figura 5 y Figura 1 en De Villiers et al., 2004a). Sin embargo, la relación de parentesco entre las especies es diferente. En nuestro estudio, las especies Alfa 6, 7 y 10 son más parecidas entre sí al compararlas con las especies Alfa 5 y 9, y el grupo con mayor diferencia corresponde a los Alfa 10; caso diferente al reportado, donde las especies 5, 6 y 7 son las más parecidas,

y en conjunto, estos se relacionan más con los Alfa 9 que con los Alfa 10, es decir, los Alfa 10 siguen siendo el grupo más divergente. La explicación de este resultado nos lleva a dos hipótesis, que la distribución de las especies se ve afectada por la sensibilidad de los métodos de análisis filogenéticos al uso de un número bajo de secuencias, o, que la proteína L1 y E6 presentan distintas tasas de mutación. Ésta última hipótesis ya ha sido planteada por algunos grupos de investigación, quienes afirman que la proteína L1 tiene una tasa de mutación más lenta que la proteína E6, por lo que la topología de los fenogramas es diferente (García-Vallvé et al., 2005; Narechania et al., 2005). En consecuencia, se comparó nuestro fenograma con el reportado para la proteína E6 (Figura 3 de García-Vallvé et al., 2005) . En el fenograma reportado, las especies Alfa 5 y 9 son más parecidas entre sí, que, con el resto del grupo, y los Alfa 10 son el grupo más divergente, tal como se observa en nuestros resultados.

En respuesta al gran número de mutaciones observadas entre las distintas proteínas E6 se decidió cuantificar el porcentaje de cada uno de los 20 aminoácidos (Figura 6), y ver si existía algún patrón distintivo entre las E6 de las diferentes especies. El análisis se hizo, a lo largo de cada secuencia y en la región resultante al eliminar 10 aminoácidos en los extremos C y N, puesto que en estudios previos se reportó que los extremos de la proteína E6 de Pvs de mamífero presentan una gran variabilidad en cuanto a contenido de aminoácido y propensión a formar estructura flexibles y desordenadas (Zanier et al., 2013; Zanier et al., 2012), esto último se confirmó mediante el análisis por RMSF (Figura 14). Los resultados del conteo determinaron que en general la mayoría de los aminoácidos se encuentran en proporciones parecidas entre las diferentes E6, tanto a lo largo de las secuencias como en ausencia de extremos, con la excepción importante, de un número menor de argininas y mayor de fenilalanina en la E6 de los VPH de bajo riesgo. La importancia de éste resultado radica en que los reportes de (Rietz et al., 2016; Ristriani et al., 2009; Zanier et al., 2013; Zanier et al., 2012) indican que las argininas cumplen un papel crucial en la interacción entre la proteína E6 y el motivo LxxLL de VPH16. En consecuencia, si las argininas están siendo sustituidas por fenilalaninas, un aminoácido polar por uno hidrofóbico, la naturaleza de la zona de interacción se ve modificada.

A modo de complementar los análisis tanto de estructura primaria y terciaria de las proteínas E6 en éste estudio, también se determinó la estructura secundaria de las mismas. En la Figura 7 se muestra cómo la estructura secundaria de las diferentes E6 entre los distintos tipos y especies de VPH se conserva; lo que era de esperar, ya que comparaciones previas entre la proteína E6 de VPH16 (un Pvs que infecta células humanas) y la E6 de BPV1 (un Pvs que infecta células de bovinos), dos mamíferos filogenéticamente

distantes, demostraron que a pesar de compartir únicamente un 30% de identidad en cuanto a secuencia, la estructura de los complejos E6/LxxLL de ambas proteínas eran muy similares (Zanier et al., 2013) .

6.2. Modelado por enhebrado de las E6

La simulación de dinámica molecular de biomoléculas son herramientas, teóricas, que permiten explorar las posibles conformaciones de una proteína en condiciones cercanas al ambiente fisiológico; Aunado a ésto, diversas investigaciones han demostrado, que, para obtener una mayor precisión en dichas conformaciones, es necesario partir de la estructura cristalina de la proteína problema (Rapaport, 2004). En el caso particular de la proteína E6 de VPH, hasta la fecha, la base de datos de proteínas (PDB, por sus siglas en inglés) contiene únicamente la estructura por rayos-X del cristal de la E6 de VPH16: en complejo con el motivo LxxLL (código 4GIZ) y en complejo con LxxLL y la proteína p53 (código 4XR8). En consecuencia, la obtención de la estructura de partida para el resto de las E6 del estudio se realizó por medio de un método llamado modelado por enhebrado, en estudios previos se ha reportado que el compendio de programa I-tasser (programa utilizado en éste estudio) construye modelos de proteínas homólogas muy parecidas a la conformación de la estructura nativa (Mason, 2009), lo cual puede ser de gran provecho, puesto que proteínas E6 de Pvs que infectan mamíferos filogenéticamente lejanos tienen conformaciones similares (Zanier et al., 2013). El protocolo de I-Tasser se resume en la Figura 20 (Roy, Kucukural, and Zhang, 2010).

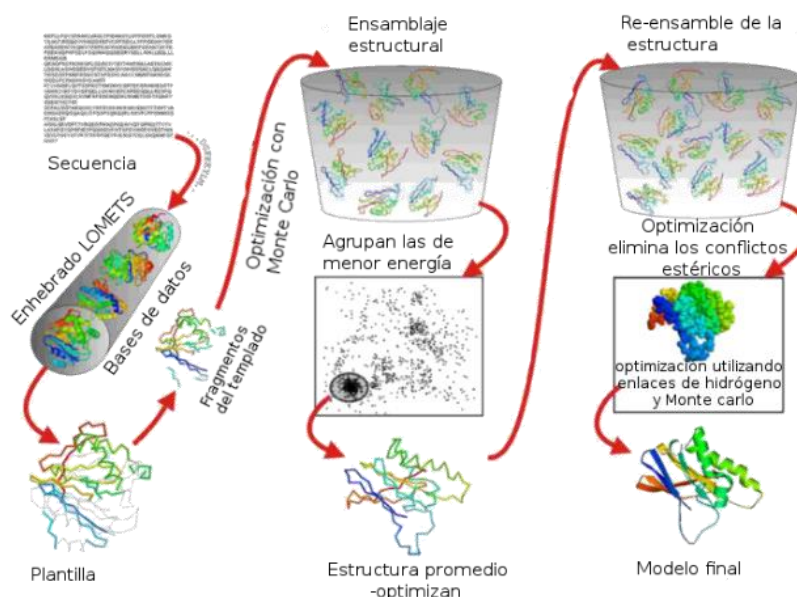


Figura 21. Representación esquemática del protocolo de I-Tasser para el modelado por enhebrado. En primer lugar, la secuencia primaria de la proteína es analizada por la herramienta PSI-BLAST, la cual busca relaciones evolutivas con proteínas alojadas en diversas bases de datos, a continuación, crea una secuencia consenso y construye la estructura secundaria por medio del programa PSIPRED, el ensamble (estructura secundaria-secuencia consenso) es ahora enhebrada (comparado) con estructuras en el PDB, la comparación toma en cuenta secuencia y estructura, para el análisis de scoring utiliza 7 programas (LOMETS) y la selección del mejor empatamiento lo evalúa tomando en cuenta cuántos de los 7 programas llegaron al mismo resultado; los fragmentos que mejor enhebraron son cortados y utilizados en la construcción de una nueva conformación, mientras que los no resueltos son confinados a un sistema de red; a continuación, los fragmentos resueltos se mantienen rígidos y ocurre un primer paso de optimización, haciendo uso de algoritmos Monte Carlo a diferentes temperaturas y campos de fuerza; por medio del programa SPICKER se agrupan las conformaciones de menor energía; a continuación, en la última etapa del proceso, se eliminan choques estéricos por el método steepest descent, y realiza un nuevo paso de optimización mediante el cálculo de enlaces de hidrógeno y minimizando por el método Monte Carlo.

La selección de la mejor conformación de cada proteína E6 se hizo mediante el cálculo de RMSD, a excepción de los casos VPH16 y VPH31, ya que la estructura presentaba divergencias en el extremo C con respecto al cristal 4GIZ, por lo que se eligió aquella estructura más parecida al cristal, cabe resaltar, que I-Tasser utilizó la estructura del cristal 4GIZ para la construcción de las 24 E6 del estudio.

6.3. Modelado por dinámica molecular (DM) del complejo E6/LxxLL

6.3.1. Preparación del complejo E6/LxxLL y DM

Antes de comenzar la producción. Cada sistema E6/LxxLL fue sometido a dos pasos de minimización de la energía (Figura 9 y 10) en aras de obtener una conformación estable con la que iniciar la simulación. En el primer paso se utilizó los algoritmos steepest descent y conjugate gradient, con lo cual se obtuvo conformaciones en un mínimo local de energía, el segundo paso corresponde al método de recocido simulado, cuyo fundamento permite explorar espacios conformacionales poco probables que evitan que el sistema caiga en mínimos energéticos locales (Roitberg and Elber, 1991). El nombre del algoritmo está inspirado en el proceso de recocido del acero y cerámicas, donde el calentamiento del sistema causa que los átomos aumenten su energía y puedan desplazarse de sus posiciones iniciales (un mínimo local energético); luego, durante un enfriamiento lento los átomos tienen una mayor probabilidad de acomodarse en configuraciones con menor energía que la inicial (mínimo global) (Khachaturyan et al., 1981). Al final de éste paso se realizó una simulación de 5.5 ns que permitió a cada sistema llegar al equilibrio termodinámico (Figura 11).

La fase de simulación por DM se hizo aplicando los mismos parámetros de simulación a cada sistema E6/LxxLL. Cabe acotar, que los parámetros para la DM utilizados en esta tesis fueron los mismo que en (Ricci, 2017), con la diferencia, de que la estructura de la proteína E6 de VPH16 y VPH11 están basados en el cristal del complejo E6/LxxLL/p53 (código 4XR8 del PDB) y no del complejo E6/LxxLL únicamente (4GIZ) (Martinez-Zapien et al., 2016a).

6.3.2. Análisis de las trayectorias obtenidas en la fase de DM

Como se mencionó al principio de éste capítulo, la proteína E6 de los Pvs de mamífero tienen en común dos dominios de unión a 2 átomos de zinc a través de 4 cisteínas cada uno. Dichos dominios son llamados E6N y E6C por su cercanía a los extremos N y C de la proteína, respectivamente, ambos dominios tienen en promedio 70 aminoácidos ordenados en una estructura específica y conservada (Lipari et al., 2001; Nominé et al., 2003), conectados por una hélice central (linker) (Zanier et al., 2013). El sitio de unión entre la proteína E6 y el motivo LxxLL de la proteína E6AP celular ocurre en una estructura tipo bolsillo formada entre los motivos E6C y E6N y la hélice linker (Zanier et al., 2013; Zanier et al., 2012). Hasta la

fecha, se ha reportado que la función de la interacción del motivo LxxLL con la proteína E6 es la de estabilizar a la proteína E6 y hacerla competente para su interacción con la proteína p53 (Ansari, Brimer, and Vande, 2012; Martinez-Zapien et al., 2016b).

Según los resultados de RMSD (Figura 13) la mayoría de las proteínas E6 de éste estudio, mostraron una conformación predominante a lo largo de la trayectoria, a excepción de la E6 de los VPH16, VPH66, VPH74, cuyos valores se agruparon en dos conjuntos, lo que indica que hubo al menos dos conformaciones predominantes, sin embargo, es necesario analizar más a fondo las trayectorias de los tres sistemas. Éste resultado es congruente con la hipótesis de que el motivo LxxLL estabiliza la estructura de la proteína E6. De igual forma, se observó que las conformaciones del pocket de cada proteína E6 se correlacionan con las de la proteína completa, pero, con valores de RMSD más bajos, lo que nos estaría indicando que la zona del pocket podría estar siendo estabilizada mediante la unión con el motivo LxxLL y que su movimiento dictamina la forma en que se mueve la proteína completa.

Éste mismo resultado se observó en el análisis con RMSF (Figura 14) donde se ve que los aminoácidos en las regiones de interacción entre la proteína E6 y el motivo LxxLL tienen poca flexibilidad. Con la excepción de los aminoácidos 57 y 58 para los VPH posiblemente cancerígenos VPH53 y VPH66 y los aminoácidos 136 y 138 para todos los VPH de bajo riesgo y VPH66, cabe acotar que la arginina 136, fue reportada como importante por su papel en la estabilización del extremo C del motivo LxxLL (Zanier et al., 2013).

6.4. Cálculo de la energía libre de interacción entre la proteína E6 y la hélice LxxLL

En estudios previos, de interacción entre 7 análogos de biotina (vitamina B7) y la proteína avidina, donde se analizó la energía libre de interacción (ΔG) con diversos métodos *in silico*, además de los métodos utilizados en éste trabajo, se determinó, que la diferencia entre los valores de ΔG entre los métodos MM/GBSA y MM/PBSA puede ser de hasta 11 Kcal/mol, dichas diferencias fueron observadas entre los resultados obtenidos por ambos métodos (Figura 17 y 18). Donde los valores de ΔG por MM/GBSA son mayores que los obtenidos por MM/PBSA (Genheden et al, 2010; Genheden and Ryde, 2015).

Nuestros resultados indicaron que todos los tipos de VPH de bajo riesgo (6, 11, 13, 44 y 74) tienen energías de unión tanto por MM/GBSA y MM/PBSA muy parecidas a los valores de los tipos posiblemente cancerígenos (30, 53 y 66) y de los tipos correspondientes a la especie Alfa 7 (18, 39, 45, 59, 68 y 70), la mayoría de alto riesgo y donde VPH18 es el causante del 20 % de los casos de cáncer de cuello uterino en el mundo (Smith et al., 2007) y el control VPH cutáneo Mu1. Además, en conjunto, tienen valores inferiores a los tipos de alto riesgo (16, 31, 33, 35, 52, 58 y 67) y los posiblemente cancerígenos (69 y 82), cuyos valores de ΔG determinan una interacción más favorable. Cabe acotar, que VPH16 no tiene los valores de ΔG más altos. Éste mismo patrón de similitud entre especies se obtuvo en el fenograma de la Figura 5.

A la fecha, no existe referencia bibliográfica que indique la incertidumbre en el valor de ΔG obtenido por MM/GBSA o MM/PBSA de proteínas relacionadas y un mismo ligante, pero si existen muchos estudios de sistemas con una misma proteína y diferentes ligantes homólogos (Genheden et al., 2010; Hou et al., 2011). Entre sus conclusiones están: los cálculos realizados por el método de MM/PBSA son más rigurosos, en consecuencia, se consideran mejor que los obtenidos por MM/GBSA; la falta del factor entrópico en el cálculo no afecta significativamente los valores de afinidad, y que a pesar de la rigurosidad del MM/PBSA su precisión es muy pobre (ver barras de error de la Figura 18), por lo que es un método ineficiente cuando se comparan ligandos muy parecidos. En consecuencia, podemos concluir, que la capacidad de interacción entre las proteínas E6AP celular y la E6 de distintos VPH estudiados no es determinante del potencial cancerígeno, al menos, bajo las condiciones desarrolladas.

Diferentes métodos experimentales han demostrado la unión de la proteína E6 de VPH con la proteína E6AP celular. De los VPH de alto riesgo: VPH16: el complejo E6/LxxLL fue detectado por inmunoprecipitación (Vande, Brown, and Turner, 1998a; Zanier et al., 2013; Zhao et al., 2013) , doble-híbrido (J. J. Chen et al., 1995; Nakagawa and Huibregtse, 2000; Zanier et al., 2012, 2013), cristalografía de rayos-X (Martinez-Zapien et al., 2016b; Zanier et al., 2013) , VPH18: inmunoprecipitación (Chen et al., 1995; Zhao et al., 2013) , resonancia del plasmón de la superficie (Ro et al., 2006), VPH31: inmunoprecipitación (Zhao et al., 2013) ; de los VPH de bajo riesgo: VPH6: inmunoprecipitación (Rozenblatt-Rosen et al., 2012) y VPH11: inmunoprecipitación (Rozenblatt-Rosen et al., 2012) , doble-híbrido (Brimer, Lyons, and Vande, 2007) y espectrometría de masas (Brimer et al., 2007) . De igual forma, en un estudio por espectrometría de masas del proteoma de células que expresaban la E6 de diferentes tipos de VPH encontraron, que la E6 de los distintos tipos reconocía subconjuntos particulares de proteínas. Recordemos que la E6 es una proteína multifuncional (Figura 2), y que la diferencia en el reconocimiento de dicho subconjunto era especie-específica; en particular, determinaron que los tipos de VPH de la especie Alfa, tanto de alto como de bajo riesgo, interaccionaban de forma preferente con la

proteína E6AP que, con el resto de las proteínas celulares, mientras que los de la especie Beta lo hacían de forma preferente con la proteína MAML1 (White and Howley, 2013; White et al., 2012). Los resultados de todos estos reportes concuerdan con las energías de interacción obtenidas en el presente trabajo.

Por otro lado, los enlaces de hidrógeno representan uno de los modos de interacción entre bio-moléculas con más importancia en la naturaleza, son los responsables de unir las microfibras de celulosa en los árboles, de la conformación de los ácidos nucleicos, cabellos, músculos y de enzimas (Gilli and Gilli, 2009); por lo que el número de enlaces de hidrógeno define mucho de la capacidad de interacción entre las moléculas. Nuestros resultados indicaron, que, en promedio, el número de enlaces de hidrógeno formados entre la proteína E6 y el motivo LxxLL entre los diferentes tipos de VPH del estudio son muy parecidos (Figura 15). Lo que apoya aun más la idea, de que no hay diferencias entre la capacidad de interacción de la proteína E6 y el motivo LxxLL entre los VPH de alto y bajo riesgo, al menos en el género *Alfapapillomavirus*. Otro resultado que soporta esto, corresponde al promedio de la distancia entre los centros de masa de la proteína E6 y el motivo LxxLL (Figura 16), donde incluso se obtuvo, que los tipos de VPH de bajo riesgo (6, 11, 13, 44 y 74) presentan distancias parecidas a los tipos de alto riesgo de la especie Alfa 9, misma donde se encuentra VPH16.

Por otra parte, la región de interacción de la proteína E6, de Pvs que infectan mamíferos, con el motivo LxxLL de la proteína E6AP celular ha sido definido como un bolsillo hidrofóbico cargado (Bohl et al., 2000; J. J. Chen et al., 1998; Huibregtse, Scheffner, and Howley, 1993; Vande, Brown, and Turner, 1998b; Zanier et al., 2013). En consecuencia, se realizó un análisis de la contribución a la afinidad, de los componentes electrostáticos (Δe_{elect}) y de Van Der Waals (Δe_{vdw}) de los aminoácidos en el bolsillo, partiendo del supuesto de que la superficie de Van Der Waals representa el estado de solvatación del bolsillo (Engquist et al., 2007), de que valores altos del componente vdW representan interacciones hidrofóbicas fuertes (Verma et al., 2016; Zanier et al., 2013) y de que los valores de Δe_{elect} representan las interacciones polares. Como se puede ver en las Figura 19 y 20, los tipos de VPH correspondientes a los Alfa 7, Alfa 10, Mu1 y 2 de los 3 Alfa 6 se relacionan, tanto en la contribución del componente Δe_{elect} como del Δe_{vdw} , en ellos, se sugiere que existe una mayor contribución de interacciones hidrofóbicas que polares, al compararlas con los valores para los tipos en Alfa 9 y 5. Es decir, los bolsillos de interacción de éstos tipos de VPH son más hidrofóbicos que polares, a diferencia de los bolsillos de los Alfa 9 y 5, que son más polares que hidrofóbicos.

De igual forma, y debido a la gran contribución del aminoácido arginina a la interacción de la proteína E6 y el motivo LxxLL de la VPH16 reportado en (Zanier et al., 2013) y de que los tipos de VPH de bajo riesgo

demonstraron tener menor contenido de argininas (Figura 6). Se realizó un análisis del aporte a la afinidad de únicamente este aminoácido (Figura 19 y 20). Como se puede ver, las interacciones polares entre la proteína E6 y el motivo LxxLL dadas por las argininas son muy importantes para todos los tipos de VPH, a excepción de los tipos de bajo riesgo, VPH66 (Alfa 6) y Mu1, lo que concuerda con los resultados observados previamente. Sin embargo, también cabe acotar, que gran parte de la contribución del componente polar de los Alfa: 9, 7, 6 y 5 está regido por las argininas, cosa que no pasa con los de bajo riesgo, Mu1 y VPH66, por lo que podría haber la presencia de otro tipo de aminoácido cargado negativamente como lisina e histidinas. Es necesario hacer otros estudios al respecto.

Finalmente, en el trabajo de (Zanier et al., 2013) proponen que los aminoácidos implicados en la interacción de la proteína E6 de VPH16 con el motivo LxxLL y la E6 de BPV1 (un Pvs de bovino) con el motivo LxxLL de la proteína paxilina, interactúan de forma diferente con el motivo LxxLL. Es decir, ellos llegaron a la conclusión, de que **un aminoácido en la misma posición de la E6** (al yuxtaponer las estructuras de las proteínas E6 del Pvs de humano y bovino), **al que llamaron residuo de lectura, puede establecer enlaces con residuos diferentes en el motivo LxxLL**; por lo tanto, los Pvs que infectan mamíferos han sido sometidos a mecanismos evolutivos que les permiten conservar la capacidad general de interactuar con los motivos LxxLL (Grace and Munger, 2017; Rozenblatt-Rosen et al., 2012; E. A. White et al., 2012), y a su vez acumular variaciones sutiles en las diferentes posiciones de los residuos de lectura, permitiendo a los Pvs interactuar con diferentes subconjuntos de proteínas que contengan motivos LxxLL. Ésta hipótesis concuerda con los resultados obtenidos en las estructuras de Boltzmann (datos no mostrados) donde los aminoácidos localizados (según el alineamiento de la Figura 5) a 5 Å del motivo LxxLL entre los diferentes VPHs no son los mismos que los determinados en la interacción de la E6 con LxxLL para VPH16. Ésto también concuerda con la similitud entre los valores de afinidad (ΔG) observadas entre los distintos VPHs (Figura 17 y 18). Sin embargo, al comparar tanto las estructuras de Boltzmann y los datos de contactos nativos se determinó, de que a pesar de que los contactos entre el motivo LxxLL y las E6 difieren mucho entre los VPH de éste estudio, existe 6 posiciones que sí se conservan (Figuras 7 y 14), al menos en el 70 % de los 100 ns de simulación de cada sistema E6/LxxLL, estas son: Val 38, Leu 57, Cys 58, Val 60, Val 69 y GLN 114. Cabe aclarar, que solo Val 38, Leu 57 y la Val 60 están conservados entre todas las especies y que solo la Leu 57 fue probada in vitro e in vivo como determinante en la interacción de E6 y LxxLL, es decir, cuando se muta, la E6 pierde la capacidad completa de unir a LxxLL y degradar a la proteína p53 (Zanier et al., 2013).

Por último, es importante mencionar, que diversas investigaciones que reportan que la mayor diferencia entre los VPH de alto y bajo riesgo recae principalmente en como ocurre la interacción de la E6 estabilizada

con el motivo LxxLL y la formación del complejo trimérico con la proteína p53, para que ésta sea eventualmente degradada (Fu et al., 2010; Elizabeth A White et al., 2012).

Capítulo 7. Conclusiones

- La proteína E6 de los tipos de VPH de la especie Alfapapilomavirus presentan niveles de divergencia, en cuanto a contenido de aminoácidos, entre el 30 y 70 %.
- Las E6 de los tipos de VPH de bajo riesgo de la especie Alfapapilomavirus contienen menor cantidad de argininas que los de alto riesgo, lo cual, tiene repercusión en la naturaleza de la interacción de las E6 de estos tipos de VPH con el motivo LxxLL.
- La estructura secundaria de la proteína E6 a partir de tipos de VPH de la especie Alfapapilomavirus y mupapilomavirus está conservada.
- Las regiones de interacción de la proteína E6 para con el motivo LxxLL tienen menor flexibilidad que los extremos C y N de la proteína.
- Los valores de energía de interacción teórica de los VPH de bajo riesgo, especie Alfa 10, son parecidos a los valores obtenidos para los VPH de alto riesgo, especie Alfa 7.
- La capacidad de interacción entre las proteínas E6 y el motivo LxxLL, no es determinante en el potencial cancerígeno de los VPH de alto y bajo riesgo, al menos en los tipos correspondientes a la especie Alfapapilomavirus y bajo las condiciones experimentales desarrolladas.
- Los aminoácidos que utiliza la proteína E6 para interactuar con el motivo LxxLL son diferentes en localización y tipo de aminoácido entre los Alfapapilomavirus, con las excepciones de: Val 38, Leu 57, Cys 58, Val 60, Val 69 y GLN 11.

Capítulo 8. Perspectivas

- Hacer estudios por DM de la proteína E6 libre de distintos tipos de VPH, y a su vez analizar las características como hidrofobicidad y carga únicamente del bolsillo de interacción.
- Estudiar la energía de interacción entre la proteína E6 de distintos tipos de VPH y el motivo LxxLL con otros métodos, ejemplo, FEP (Free energy pertubation).
- Analizar el aporte, a la energía libre, de cada aminoácido implicado en la interacción de la proteína E6 y el motivo LxxLL de diferentes tipos de VPH, con la intención de entender la naturaleza de la interacción y patrones distintivos entre los diferentes tipos.
- Estudiar por DM y métodos de cálculo de energía libre de interacción, el complejo trimérico E6/LxxLL/p53.
- Analizar la energía de interacción de la proteína E6 de distintos tipos de VPH y diferentes motivos LxxLL, en aras de reconocer un patrón distintivo entre los VPH en cuanto a su predisposición a interactuar con proteínas celulares involucradas en tropismo, ciclo de vida y patogenicidad del virus.
- Estudiar por métodos de DM la interacción de la proteína E6 y el motivo PDZ, ya que se ha reportado que dicha interacción tiene consecuencias importantes en la capacidad oncogénica del virus.
- Estudiar por métodos de DM la interacción de la proteína E6 y los motivos PDZ y LxxLL al mismo tiempo. Ya que se ha demostrado por métodos in vivo que la quimera LxxLL/PDZ aumenta la afinidad de ambos por la E6, en el rango de nanomolar. Y esto solo ocurre con la E6 de los VPH de alto riesgo.

Literatura citada

- Allison, D. B., and Maleki, Z. (2016). HPV-related head and neck squamous cell carcinoma: An update and review. *Journal of the American Society of Cytopathology*. <https://doi.org/10.1016/j.jasc.2015.12.001>
- Alvarez-Salas, L. M., Cullinan, A. E., Siwkowski, A., Hampel, A., and Dipaolo, J. A. (1998). Inhibition of HPV-16 E6E7 immortalization of normal keratinocytes by hairpin ribozymes. *Medical Sciences*, 95, 1189–1194.
- Androphy, E. J., Hubbert, N. L., Schiller, J. T., and Lowy, D. R. (1987). Identification of the HPV-16 E6 protein from transformed mouse cells and human cervical carcinoma cell lines. *The EMBO Journal*, 6(4), 989–992.
- Androphyl, E. J., Hubbert, N. L., Schiller, J. T., and Lowy, D. R. (1987). Identification of the HPV-16 E6 protein from transformed mouse cells and human cervical carcinoma cell lines. *The EMBO Journal*, 6(4), 989–992.
- Anfinsen, C. B. (1973). Principles that govern the folding of protein chains. *Science (New York, N.Y.)*, 181(4096), 223–230. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/4124164>
- Ansari, T., Brimer, N., and Vande Pol, S. B. (2012). Peptide Interactions Stabilize and Restructure Human Papillomavirus Type 16 E6 To Interact with p53. *Journal of Virology*. <https://doi.org/10.1128/JVI.01236-12>
- Araldi, R. P., Assaf, S. M. R., de Carvalho, R. F., de Carvalho, M. A. C. R., de Souza, J. M., Magnelli, R. F., Beçak, W. (2017). Papillomaviruses: A systematic review. *Genetics and Molecular Biology*. <https://doi.org/10.1590/1678-4685-GMB-2016-0128>
- Araldi, R. P., Melo, T. C., Neves, A. C., Spadacci-Morena, D. D., Magnelli, R. F., Modolo, D. G., ... Stocco, R. C. (2015). Hyperproliferative action of bovine papillomavirus: genetic and histopathological aspects. *Funpecrp.com.br Genetics and Molecular Research Genet. Mol. Res*, 14(144), 12942–12954. <https://doi.org/10.4238/2015.October.21.15>
- Berendsen, H. J. C., Postma, J. P. M., van Gunsteren, W. F., DiNola, A., and Haak, J. R. (1984). Molecular dynamics with coupling to an external bath. *The Journal of Chemical Physics*, 81(8), 3684–3690. <https://doi.org/10.1063/1.448118>
- Bocaneti, F., Altamura, G., Corteggio, A., Velescu, E., Roperto, F., and Borzacchiello, G. (2014). Bovine Papillomavirus: New Insights into an Old Disease. *Transboundary and Emerging Diseases*, 1–10. <https://doi.org/10.1111/tbed.12222>
- Bodily, J., and Laimins, L. a. (2011). Persistence of human papillomavirus infections: keys to malignant progression. *Trends in Microbiology*, 19(1), 33–39. <https://doi.org/10.1016/j.tim.2010.10.002.Persistence>
- Bohl, J., Das, K., Dasgupta, B., and Vande Pol, S. B. (2000). Competitive Binding to a Charged Leucine Motif Represses Transformation by a Papillomavirus E6 Oncoprotein. *Virology*, 271(1), 163–170. <https://doi.org/10.1006/viro.2000.0316>

- Bonetta, A. C., Maily, L., Robinet, E., Travé, G., Masson, M., and Deryckere, F. (2015). Artificial microRNAs against the viral E6 protein provoke apoptosis in HPV positive cancer cells. *Biochemical and Biophysical Research Communications*, 465(4), 658–664. <https://doi.org/10.1016/J.BBRC.2015.07.144>
- Boon, S. S., Tomaić, V., Thomas, M., Roberts, S., and Banks, L. (2015). Cancer-Causing Human Papillomavirus E6 Proteins Display Major Differences in the Phospho-Regulation of Their PDZ Interactions. *Journal of Virology*. <https://doi.org/10.1128/JVI.01961-14>
- Bornot, A., Etchebest, C., and De Brevern, A. G. (2011). Predicting protein flexibility through the prediction of local structures. *Proteins: Structure, Function and Bioinformatics*, 79(3), 839–852. <https://doi.org/10.1002/prot.22922>
- Bowie, J. U., Lüthy, R., and Eisenberg, D. (1991). A method to identify protein sequences that fold into a known three-dimensional structure. *Science (New York, N.Y.)*, 253(5016), 164–170. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/1853201>
- Bravo, I. G., de Sanjos, S., and Gottschling, M. (2010). The clinical importance of understanding the evolution of papillomaviruses. *Trends in Microbiology*, 18(10), 432–438. <https://doi.org/10.1016/j.tim.2010.07.008>
- Bravo, I. G., and Félez-Sánchez, M. (2015). Papillomaviruses. *Evolution, Medicine, and Public Health*, 2015(1), 32–51. <https://doi.org/10.1093/emph/eov003>
- Brimer, N., Lyons, C., and Vande Pol, S. B. (2007). Association of E6AP (UBE3A) with human papillomavirus type 11 E6 protein. *Virology*, 358(2), 303–310. <https://doi.org/10.1016/j.virol.2006.08.038>
- Butz, K., Denk, C., Ullmann, A., Scheffner, M., and Hoppe-Seyler, F. (2000). Induction of apoptosis in human papillomavirus- positive cancer cells by peptide aptamers targeting the viral E6 oncoprotein.
- Butz, K., Ristriani, T., Hengstermann, A., Denk, C., Scheffner, M., and Hoppe-Seyler, F. (2003). siRNA targeting of the viral E6 oncogene efficiently kills human papillomavirus-positive cancer cells. *Oncogene*, 22(38), 5938–5945. <https://doi.org/10.1038/sj.onc.1206894>
- Cai, Q., Lv, L., Shao, Q., Li, X., and Dian, A. (2013). Human papillomavirus early proteins and apoptosis. *Archives of Gynecology and Obstetrics*, 287(3), 541–548. <https://doi.org/10.1007/s00404-012-2665-z>
- Case, D. A., Cerutti, D. S., Cheatham, T. E. I., Darden, T. A., Duke, R. E., Giese, T. J., ... Kollman, P. A. (2017). Amber 2017. University of California, San Francisco. <https://doi.org/citeulike-article-id:2734527>
- Chen, F., Sun, H., Wang, J., Zhu, F., Liu, H., Wang, Z., ... Hou, T. (2018). Assessing the performance of MM/PBSA and MM/GBSA methods. 8. Predicting binding free energies and poses of protein-RNA complexes. *RNA*, rna.065896.118. <https://doi.org/10.1261/rna.065896.118>
- Chen, J. J., Hong, Y., Rustamzadeh, E., Baleja, J. D., and Androphy, E. J. (1998). Identification of an α helical motif sufficient for association with papillomavirus E6. *Journal of Biological Chemistry*. <https://doi.org/10.1074/jbc.273.22.13537>

- Chen, J. J., Reid, C. E., Band, V., and Androphy, E. J. (1995). Interaction of papillomavirus E6 oncoproteins with a putative calcium-binding protein. *Science*, 269(5223), 529–531. <https://doi.org/10.1126/science.7624774>
- Chong, L. T., Duan, Y., Wang, L., Massova, I., and Kollman, P. A. (1999). Molecular dynamics and free-energy calculations applied to affinity maturation in antibody 48G7. *Proceedings of the National Academy of Sciences of the United States of America*, 96(25), 14330–14335. <https://doi.org/10.1073/PNAS.96.25.14330>
- Chouhy, D., Bolatti, E. M., Perez, G. R., and Giri, A. A. (2013). Analysis of the genetic diversity and phylogenetic relationships of putative human papillomavirus types. *Journal of General Virology*. <https://doi.org/10.1099/vir.0.055137-0>
- Clark Still, W., Tempczyk, A., Hawley, R.C., and Hedrickson, T, (1990). Semianalytical treatment of salvation for Molecular Mechanism and Dynamics. *Journal of the American Chemical Society*, 112 (16), 6127-2129.
- Costantini, S., Paladino, A., and Facchiano, A. M. (2008). CALCOM: a software for calculating the center of mass of proteins. *Bioinformatics*, 2(7), 271–272. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/18478078>
- Danos, O., Katinka, M., and Yaniv, M. (1982). Human papillomavirus 1a complete DNA sequence: a novel type of genome organization among papovaviridae. *The EMBO Journal*, 1(2), 231–236. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/6325156>
- De Villiers, E. M., Fauquet, C., Broker, T. R., Bernard, H. U., and Zur Hausen, H. (2004a). Classification of papillomaviruses. *Virology*. <https://doi.org/10.1016/j.virol.2004.03.033>
- De Villiers, E. M., Fauquet, C., Broker, T. R., Bernard, H. U., and Zur Hausen, H. (2004b, June). Classification of papillomaviruses. *Virology*. <https://doi.org/10.1016/j.virol.2004.03.033>
- De Vivo, M., Masetti, M., Bottegoni, G., and Cavalli, A. (2016, May 12). Role of Molecular Dynamics and Related Methods in Drug Discovery. *Journal of Medicinal Chemistry*. American Chemical Society. <https://doi.org/10.1021/acs.jmedchem.5b01684>
- Doorbar, J., Quint, W., Banks, L., Bravo, I. G., Stoler, M., Broker, T. R., and Stanley, M. A. (2012). The biology and life-cycle of human papillomaviruses. *Vaccine*, 30(SUPPL.5), F55–F70. <https://doi.org/10.1016/j.vaccine.2012.06.083>
- Drake, J. W. (1991). A constant rate of spontaneous mutation in DNA-based microbes. *Proceedings of the National Academy of Sciences of the United States of America*, 88(16), 7160–7164. <https://doi.org/10.1073/PNAS.88.16.7160>
- Drozdetskiy, A., Cole, C., Procter, J., and Barton, G. J. (2015). JPred4: a protein secondary structure prediction server. *Nucleic Acids Research*, 43(W1), W389–W394. <https://doi.org/10.1093/nar/gkv332>
- Edgar, R. C. (2004). MUSCLE: A multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics*. <https://doi.org/10.1186/1471-2105-5-113>

- Enemark, E. J., Chen, G., Vaughn, D. E., Stenlund, A., and Joshua-Tor, L. (2000). Crystal structure of the DNA binding domain of the replication initiation protein E1 from papillomavirus. *Mol Cell*, 6(1), 149–158. [https://doi.org/S1097-2765\(05\)00016-X](https://doi.org/S1097-2765(05)00016-X) [pii]
- Engquist E, W., B., Li, X., Ren, W., and Vanden-Eijnden, E. (2007). Communications in computational physics. *Communications in Computational Physics* (Vol. 2). Global Science Press. Retrieved from <https://nyuscholars.nyu.edu/en/publications/heterogeneous-multiscale-methods-a-review>
- Ferraro, C. T. L., Canedo, N. H. S., Oliveira, S. P. de, Carvalho, M. da G. da C., and Dias, E. P. (2011). Infecção oral pelo HPV e lesões epiteliais proliferativas associadas. *Jornal Brasileiro de Patologia E Medicina Laboratorial*, 47(4), 451–459. <https://doi.org/10.1590/S1676-24442011000400010>
- Fu, L., van Doorslaer, K., Chen, Z., Ristriani, T., Masson, M., Travé, G., and Burk, R. D. (2010). Degradation of p53 by human Alphapapillomavirus E6 proteins shows a stronger correlation with phylogeny than oncogenicity. *PLoS ONE*. <https://doi.org/10.1371/journal.pone.0012816>
- García-Vallvé, S., Alonso, Á., and Bravo, I. G. (2005). Papillomaviruses: Different genes have different histories. *Trends in Microbiology*. <https://doi.org/10.1016/j.tim.2005.09.003>
- Genheden, S., Luchko, T., Gusarov, S., Kovalenko, A., and Ryde, U. (2010). An MM / 3D-RISM Approach for Ligand Binding Affinities. *J. Phys. Chem.*, 114, 8505–8516. <https://doi.org/10.1021/jp101461s>
- Genheden, S., and Ryde, U. (2015). The MM/PBSA and MM/GBSA methods to estimate ligand-binding affinities. *Expert Opinion on Drug Discovery*, 10(5), 449–461. <https://doi.org/10.1517/17460441.2015.1032936>
- Gilli, G., and Gilli, P. (2009). *The nature of the hydrogen bond : outline of a comprehensive hydrogen bond theory*. Oxford University Press.
- Gohlke, H., Kiel, C., and Case, D. A. (2003). Insights into protein-protein binding by binding free energy calculation and free energy decomposition for the Ras-Raf and Ras-RalGDS complexes. *Journal of Molecular Biology*, 330(4), 891–913. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/12850155>
- Grace, M., and Munger, K. (2017). Proteomic analysis of the gamma human papillomavirus type 197 E6 and E7 associated cellular proteins. *Virology*. <https://doi.org/10.1016/j.virol.2016.10.010>
- Harden, M. E., and Munger, K. (2017). Human papillomavirus molecular biology. *Mutation Research/Reviews in Mutation Research*, 772, 3–12. <https://doi.org/10.1016/j.mrrev.2016.07.002>
- Hou, T., Wang, J., Li, Y., and Wang, W. (2011). Assessing the Performance of the MM / PBSA and MM / GBSA Methods . 1 . The Accuracy of Binding Free Energy Calculations Based on Molecular Dynamics Simulations. *Journal of Chemical Information and Modeling*, 51(1), 69–82. <https://doi.org/10.1021/ci100275a>
- Hsa. (2014). *Biological Agents*, 100, 117–122. Retrieved from http://www.hsa.ie/eng/Topics/Biological_Agents/
- Huibregtse, J. M., Scheffner, M., and Howley, P. M. (1993). Localization of the E6-AP regions that direct human papillomavirus E6 binding, association with p53, and ubiquitination of associated proteins.

- Molecular and Cellular Biology, 13(8), 4918–4927. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/8393140>
- Huo, S., Wang, J., Cieplak, P., Kollman, P. A., and Kuntz, I. D. (2002). Molecular dynamics and free energy analyses of cathepsin D-inhibitor interactions: insight into structure-based ligand design. *Journal of Medicinal Chemistry*, 45(7), 1412–1419. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/11906282>
- John Doorbar, Nagayasu Egawa, Heather Griffin, C. K. and I. M. (2015). Human papillomavirus molecular biology and disease association. *Reviews in Medical Virology*, 25(1), 2–23. <https://doi.org/10.1002/rmv>
- Joyce, J. G., Tung, J., Craig, T., Cook, J. C., Lehman, E. D., Sands, J. a, Keller, P. M. (1999). Protein chemistry and structure : The L1 Major Capsid Protein of Human Papillomavirus Type 11 Recombinant Virus-like Particles Interacts with Heparin and Cell-surface Glycosaminoglycans on The L1 Major Capsid Protein of Human Papillomavirus Type 11 Recomb, 274(9), 5810–5822. <https://doi.org/10.1074/jbc.274.9.5810>
- Kaphingst, K. A., Persky, S., and Lachance, C. (2010). NIH Public Access, 14(4), 384–399. <https://doi.org/10.1080/10810730902873927>.Testing
- Karplus, M., and Kuriyan, J. (2005). Molecular dynamics and protein function. *Proceedings of the National Academy of Sciences*, 102(19), 6679–6685. <https://doi.org/10.1073/pnas.0408930102>
- Karplus, M., and Petsko, G. A. (1990). Molecular dynamics simulations in biology. *Nature*, 347(6294), 631–639. <https://doi.org/10.1038/347631a0>
- Khachaturyan, A., Semenovsovskaia, S., Vainshtein, B., and IUCr. (1981). The thermodynamic approach to the structure analysis of crystals. *Acta Crystallographica Section A*, 37(5), 742–754. <https://doi.org/10.1107/S0567739481001630>
- Kines, R. C., Thompson, C. D., Lowy, D. R., Schiller, J. T., and Day, P. M. (2009). The initial steps leading to papillomavirus infection occur on the basement membrane prior to cell surface binding. *Proceedings of the National Academy of Sciences*, 106(48), 20458–20463. <https://doi.org/10.1073/pnas.0908502106>
- Kollmsn, P.A., Massova , I., Reyes, C., Kuhn, B., Chong, L., Cheatham, T.E. (2000). Calculating Structures and Free Energy of Complex Molecules. *Combining Molecular Mehanics ans Continuum Models*, 33(12), 889-897.
- Kovacs, J. S., and Mcharris, O. (2000). *Energy and Boltzmann Distributions*.Michigan State University. Recuperado de <http://www.physnet.org/home/modules>.
- Kuhn, B and Kollman*, P. A. (2000). Binding of a Diverse Set of Ligands to Avidin and Streptavidin: An Accurate Quantitative Prediction of Their Relative Affinities by a Combination of Molecular Mechanics and Continuum Solvent Models. <https://doi.org/10.1021/JM000241H>
- Kumar, S., Stecher, G., Peterson, D., and Tamura, K. (2012). MEGA-CC: computing core of molecular evolutionary genetics analysis program for automated and iterative data analysis, 28(20), 2685–2686. <https://doi.org/10.1093/bioinformatics/bts507>

- Lavezzo, E., Masi, G., Toppo, S., Franchin, E., Gazzola, V., Sinigaglia, A., ... Barzon, L. (2016). Characterization of intra-type variants of oncogenic human papillomaviruses by next-generation deep sequencing of the E6/E7 region. *Viruses*. <https://doi.org/10.3390/v8030079>
- Leach, A. R. (2001). *Molecular modelling: principles and applications*. Culture Ethnicity and Conflict. <https://doi.org/qd480.l43> 2001
- Lefkowitz, E. J., Dempsey, D. M., Hendrickson, R. C., Orton, R. J., Siddell, S. G., and Smith, D. B. (2018). Virus taxonomy: The database of the International Committee on Taxonomy of Viruses (ICTV). *Nucleic Acids Research*, 46(D1), D708–D717. <https://doi.org/10.1093/nar/gkx932>
- Lipari, F., McGibbon, G. A., Wardrop, E., and Cordingley, M. G. (2001). Purification and biophysical characterization of a minimal functional domain and of an N-terminal Zn²⁺-binding fragment from the human papillomavirus type 16 E6 protein. *Biochemistry*, 40(5), 1196–1204. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/11170444>
- Liu, X., Clements, A., Zhao, K., and Marmorstein, R. (2006). Structure of the human Papillomavirus E7 oncoprotein and its mechanism for inactivation of the retinoblastoma tumor suppressor. *Journal of Biological Chemistry*. <https://doi.org/10.1074/jbc.M508455200>
- Lowy, D. R., and Schiller, J. T. (2006). Prophylactic human papillomavirus vaccines. *Journal of Clinical Investigation*. <https://doi.org/10.1172/JCI28607>
- Lynch, M. (2010). Evolution of the mutation rate. *Trends in Genetics: TIG*, 26(8), 345–352. <https://doi.org/10.1016/j.tig.2010.05.003>
- Mackay, D. H. J., Cross, A. J., and Hagler, A. T. (1989). The Role of Energy Minimization in Simulation Strategies of Biomolecular Systems. *Prediction of Protein Structure and the Principles of Protein Conformation*, 317–358. https://doi.org/10.1007/978-1-4613-1571-1_7
- Maier, J. A., Martinez, C., Kasavajhala, K., Wickstrom, L., Hauser, K. E., and Simmerling, C. (2015). ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. <https://doi.org/10.1021/acs.jctc.5b00255>
- Martínez-Zapien, D., Ruiz, F. X., Poirson, J., Mitschler, A., Ramírez, J., Forster, A., Zanier, K. (2016a). Structure of the E6/E6AP/p53 complex required for HPV-mediated degradation of p53. *Nature*. <https://doi.org/10.1038/nature16481>
- Martínez-Zapien, D., Ruiz, F. X., Poirson, J., Mitschler, A., Ramírez, J., Forster, A., Zanier, K. (2016b). Structure of the E6/E6AP/p53 complex required for HPV-mediated degradation of p53. *Nature*, 529(7587), 541–545. <https://doi.org/10.1038/nature16481>
- Massova, I., and Kollman*, P. A. (1999). Computational Alanine Scanning To Probe Protein–Protein Interactions: A Novel Approach To Evaluate Binding Free Energies. <https://doi.org/10.1021/JA990935J>
- Mason, A. (2009). NIH Public Access, 19(Suppl 9), 389–399. <https://doi.org/10.1016/j.asieco.2008.09.006.EAST>

- Matsuda, H. (n.d.). PROTEIN PHYLOGENETIC INFERENCE USING MAXIMUM LIKELIHOOD WITH A GENETIC ALGORITHM. Retrieved from <http://psb.stanford.edu/psb-online/proceedings/psb96/matsuda.pdf>
- Mittal, S., and Banks, L. (2017). Molecular mechanisms underlying human papillomavirus E6 and E7 oncoprotein-induced cell transformation. *Mutation Research - Reviews in Mutation Research*. <https://doi.org/10.1016/j.mrrev.2016.08.001>
- Moody, C. A., and Laimins, L. A. (2010). Human papillomavirus oncoproteins: pathways to transformation. *Nature Reviews Cancer*, 10(8), 550–560. <https://doi.org/10.1038/nrc2886>
- Nakagawa, S., and Huibregtse, J. M. (2000). Human scribble (Vartul) is targeted for ubiquitin-mediated degradation by the high-risk papillomavirus E6 proteins and the E6AP ubiquitin-protein ligase. *Molecular and Cellular Biology*, 20(21), 8244–8253. <https://doi.org/10.1128/MCB.20.21.8244-8253.2000>
- Narechania, A., Chen, Z., DeSalle, R., and Burk, R. D. (2005). Phylogenetic incongruence among oncogenic genital alpha human papillomaviruses. *Journal of Virology*, 79(24), 15503–15510. <https://doi.org/10.1128/JVI.79.24.15503-15510.2005>
- Nelson, D. L., Lehninger, A. L., and Cox, M. M. (2015). *Lehninger principios de bioquímica*. Omega. Retrieved from <http://www.ediciones-omega.es/bioquimica/35-lehninger-principios-de-bioquimica-978-84-282-1603-6.html>
- Nominé, Y., Charbonnier, S., Ristriani, T., Stier, G., Masson, M., Cavusoglu, N., Travé, G. (2003). Domain Substructure of HPV E6 Oncoprotein: Biophysical Characterization of the E6 C-Terminal DNA-Binding Domain †. *Biochemistry*, 42(17), 4909–4917. <https://doi.org/10.1021/bi026980c>
- Pattengale, N. D., Alipour, M., Bininda-Emonds, O. R. P., Moret, B. M. E., and Stamatakis, A. (2010). How Many Bootstrap Replicates Are Necessary? *Journal of Computational Biology*, 17(3), 337–354. <https://doi.org/10.1089/cmb.2009.0179>
- Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C., and Ferrin, T. E. (2004). UCSF Chimera - A visualization system for exploratory research and analysis. *Journal of Computational Chemistry*, 25(13), 1605–1612. <https://doi.org/10.1002/jcc.20084>
- Pol, S. B. Vande, Brown, M. C., and Turner, C. E. (1998a). Association of Bovine Papillomavirus Type 1 E6 oncoprotein with the focal adhesion protein paxillin through a conserved protein interaction motif. *Oncogene*, 16(1), 43–52. <https://doi.org/10.1038/sj.onc.1201504>
- Pol, S. B. Vande, Brown, M. C., and Turner, C. E. (1998b). Association of Bovine Papillomavirus Type 1 E6 oncoprotein with the focal adhesion protein paxillin through a conserved protein interaction motif. *Oncogene*. <https://doi.org/10.1038/sj.onc.1201504>
- Price, D. J., Brooks, C. L., and Brooks Iii, C. L. (2004). A modified TIP3P water potential for simulation with Ewald summation Comparison of simple potential functions for simulating liquid water A modified TIP3P water potential for simulation with Ewald summation. *The Journal of Chemical Physics The Journal of Chemical Physics The Journal of Chemical Physics The Journal of Chemical Physics*, 121(3684). <https://doi.org/10.1063/1.1808117>

- Pyeon, D., Pearce, S. M., Lank, S. M., Ahlquist, P., and Lambert, P. F. (2009). Establishment of human papillomavirus infection requires cell cycle progression. *PLoS Pathogens*, 5(2). <https://doi.org/10.1371/journal.ppat.1000318>
- Rapaport, D. C. (2004). *The art of molecular dynamics simulation*. Cambridge University Press.
- Reyes, C. M., and Kollman, P. A. (2000). Structure and thermodynamics of RNA-protein binding: using molecular dynamics and free energy analyses to calculate the free energies of binding and conformational change¹¹Edited by B. Honig. *Journal of Molecular Biology*, 297(5), 1145–1158. <https://doi.org/10.1006/jmbi.2000.3629>
- Ricci, J. (2016). Estudio *in silico* de la proteína E6 del virus de papiloma humano para identificar potenciales fármacos antivirales. Tesis de Maestría en Ciencias. CICESE. Ensenada, B.C. México.
- Rietz, A., Petrov, D. P., Bartolowits, M., DeSmet, M., Davisson, V. J., and Androphy, E. J. (2016). Molecular probing of the HPV-16 E6 protein alpha helix binding groove with small molecule inhibitors. *PLoS ONE*. <https://doi.org/10.1371/journal.pone.0149845>
- Ristriani, T., Fournane, S., Orfanoudakis, G., Trav, G., and Masson, M. (2009). A single-codon mutation converts HPV16 E6 oncoprotein into a potential tumor suppressor, which induces p53-dependent senescence of HPV-positive HeLa cervical cancer cells. *Oncogene*, 28(5), 762–772. <https://doi.org/10.1038/onc.2008.422>
- Ristriani, T., Masson, M., Nominé, Y., Laurent, C., Lefevre, J. F., Weiss, E., and Travé, G. (2000). HPV oncoprotein E6 is a structure-dependent DNA-binding protein that recognizes four-way junctions. *Journal of Molecular Biology*, 296(5), 1189–1203. <https://doi.org/10.1006/jmbi.2000.3527>
- Ro, H. S., Koh, B. H., Jung, S. O., Park, H. K., Shin, Y. B., Kim, M. G., and Chung, B. H. (2006). Surface plasmon resonance imaging protein arrays for analysis of triple protein interactions of HPV, E6, E6AP, and p53. *Proteomics*, 6(7), 2108–2111. <https://doi.org/10.1002/pmic.200500635>
- Roberts, J. N., Buck, C. B., Thompson, C. D., Kines, R., Bernardo, M., Choyke, P. L., ... Schiller, J. T. (2007). Genital transmission of HPV in a mouse model is potentiated by nonoxynol-9 and inhibited by carrageenan. *Nature Medicine*, 13(7), 857–861. <https://doi.org/10.1038/nm1598>
- Roe, D. R., and Cheatham, T. E. (2000). PTRAJ and CPPTRAJ: Software for Processing and Analysis of Molecular Dynamics Trajectory Data. <https://doi.org/10.1021/ct400341p>
- Roitberg, A., and Elber, R. (1991). Modeling side chains in peptides and proteins: Application of the locally enhanced sampling and the simulated annealing methods to find minimum energy conformations. *The Journal of Chemical Physics*, 95(12), 9277–9287. <https://doi.org/10.1063/1.461157>
- Roy, A., Kucukural, A., and Zhang, Y. (2010). I-TASSER: a unified platform for automated protein structure and function prediction. *Nature Protocols*. <https://doi.org/10.1038/nprot.2010.5>
- Rozenblatt-Rosen, O., Deo, R. C., Padi, M., Adelmant, G., Calderwood, M. A., Rolland, T., Vidal, M. (2012). Interpreting cancer genomes using systematic host network perturbations by tumour virus proteins. *Nature*, 487(7408), 491–495. <https://doi.org/10.1038/nature11288>

- Sacco, A., Siepi, F., and Crescenzi, M. (2003). HPV E7 expression in skeletal muscle cells distinguishes initiation of the postmitotic state from its maintenance. *Oncogene*. <https://doi.org/10.1038/sj.onc.1206353>
- Saitou, N., and Nei, M. (1987). The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution*, 4(4), 406–425. <https://doi.org/10.1093/oxfordjournals.molbev.a040454>
- Scheffner, M., Werness, B. A., Huibregtse, J. M., Levine, A. J., and Howley, P. M. (1990). The E6 oncoprotein encoded by human papillomavirus types 16 and 18 promotes the degradation of p53. *Cell*, 63(6), 1129–1136. [https://doi.org/10.1016/0092-8674\(90\)90409-8](https://doi.org/10.1016/0092-8674(90)90409-8)
- Scott, M., Nakagawa, M., and Moscicki, A.-B. (2001). MINIREVIEW Cell-Mediated Immune Response to Human Papillomavirus Infection, 8(2), 209–220. <https://doi.org/10.1128/CDLI.8.2.209-220.2001>
- Smith, J. S., Lindsay, L., Hoots, B., Keys, J., Franceschi, S., Winer, R., and Clifford, G. M. (2007). Human papillomavirus type distribution in invasive cervical cancer and high-grade cervical lesions: A meta-analysis update. *International Journal of Cancer*, 121(3), 621–632. <https://doi.org/10.1002/ijc.22527>
- Standard, J. M. (2015). Energy Minimization Methods. *Chemistry*, 38037. Retrieved from [https://about.illinoisstate.edu/standard/Documents/CHE 380.37/Handouts/380.37emin.pdf](https://about.illinoisstate.edu/standard/Documents/CHE%20380.37/Handouts/380.37emin.pdf)
- Steele, C., Sacks, P. G., Adler-Storthz, K., and Shillitoe, E. J. (1992). Effect on cancer cells of plasmids that express antisense RNA of human papillomavirus type 18. *Cancer Research*, 52(17), 4706–4711.
- Stone, J. (2011). Using VMD. Main, Chapter 5, Unit 5.7. <https://doi.org/10.1002/0471250953.bi0507s24>
- Tan, M. J. A., White, E. A., Sowa, M. E., Harper, J. W., Aster, J. C., and Howley, P. M. (2012). Cutaneous - human papillomavirus E6 proteins bind Mastermind-like coactivators and repress Notch signaling. *Proceedings of the National Academy of Sciences*. <https://doi.org/10.1073/pnas.1205991109>
- Tidor, B. (1993). Simulated Annealing on Free Energy Surfaces by a Combined Molecular Dynamics and Monte Carlo Approach. *J. Phys. Chem*, 97, 1069–1073. Retrieved from <https://pubs.acs.org/doi/pdf/10.1021/j100107a015>
- Tsui, V., and Case*, D. A. (2001). Calculations of the Absolute Free Energies of Binding between RNA and Metal Ions Using Molecular Dynamics Simulations and Continuum Electrostatics. <https://doi.org/10.1021/JP011923Z>
- Turjanski, A. G., Gutkind, J. S., Best, R. B., and Hummer, G. (2008). Binding-Induced Folding of a Natively Unstructured Transcription Factor. *PLoS Computational Biology*, 4(4), e1000060. <https://doi.org/10.1371/journal.pcbi.1000060>
- Van Doorslaer, K. (2013). Evolution of the Papillomaviridae. *Virology*. <https://doi.org/10.1016/j.virol.2013.05.012>
- Van Doorslaer, K., Li, Z., Xirasagar, S., Maes, P., Kaminsky, D., Liou, D., ... McBride, A. A. (2017). The Papillomavirus Episteme: a major update to the papillomavirus sequence database. *Nucleic Acids Research*, 45(D1), D499–D506. <https://doi.org/10.1093/nar/gkw879>

- Vande Pol, S., Brown, M., and Turner, C. (1998). Association of bovine papillomavirus type 1 E6 oncoprotein with the focal adhesion protein paxillin through a conserved protein interaction motif. *Oncogene*, 16(1), 43–52. <https://doi.org/10.1038/sj.onc.1201504>
- Verma, S., Grover, S., Tyagi, C., Goyal, S., Jamal, S., Singh, A., and Grover, A. (2016). Hydrophobic Interactions Are a Key to MDM2 Inhibition by Polyphenols as Revealed by Molecular Dynamics Simulations and MM/PBSA Free Energy Calculations. *PloS One*, 11(2), e0149014. <https://doi.org/10.1371/journal.pone.0149014>
- Villacañas, Ó., and Rubio-Martínez, J. (2006). Reducing CDK4/6-p16INK4a interface. Computational alanine scanning of a peptide bound to CDK6 protein. *Proteins: Structure, Function, and Bioinformatics*, 63(4), 797–810. <https://doi.org/10.1002/prot.20943>
- Wallace, N. A., and Galloway, D. A. (2014). Manipulation of cellular DNA damage repair machinery facilitates propagation of human papillomaviruses. *Seminars in Cancer Biology*, 26, 30–42. <https://doi.org/10.1016/j.semcancer.2013.12.003>
- Wang, J. W., and Roden, R. B. S. (2013). L2, the minor capsid protein of papillomavirus. *Virology*. <https://doi.org/10.1016/j.virol.2013.04.017>
- Wang, J., Wolf, R. M., Caldwell, J. W., Kollman, P. A., and Case, D. A. (2004). Development and Testing of a General Amber Force Field. *J Comput Chem*, 25, 1157–1174. Retrieved from <http://ambermd.org/antechamber/gaff.pdf>
- Werness, B. A., Levine, A. J., and Howley, P. M. (1990). Association of Human Papillomavirus Types 16 and 18 E6 Proteins with p53. Source: Science, New Series Phys. Rev. Lett. J. Cell Biol. Ann. Phys. (Paris) IEEE J. Quantum Electron. OE- Opt. Lett. J. P. Hermann and J. Ducuing Phys. Rev. A 10. R. R. Birge, Acc. Chem. Res. Chem. Phys. Lett, 24822(12), 76–79. Retrieved from <http://www.jstor.org/stable/2874053>
- Westrich, J. A., Warren, C. J., and Pyeon, D. (2017). Evasion of host immune defenses by human papillomavirus. *Virus Research*. <https://doi.org/10.1016/j.virusres.2016.11.023>
- White, E. A., and Howley, P. M. (2013). Proteomic approaches to the study of papillomavirus-host interactions. *Virology*. <https://doi.org/10.1016/j.virol.2012.09.046>
- White, E. A., Kramer, R. E., Tan, M. J. A., Hayes, S. D., Harper, J. W., and Howley, P. M. (2012). Comprehensive Analysis of Host Cellular Interactions with Human Papillomavirus E6 Proteins Identifies New E6 Binding Partners and Reflects Viral Diversity. *Journal of Virology*, 86(24), 13174–13186. <https://doi.org/10.1128/JVI.02172-12>
- White, E. A., Kramer, R. E., Tan, M. J. A., Hayes, S. D., Harper, J. W., and Howley, P. M. (2012). Comprehensive analysis of host cellular interactions with human papillomavirus E6 proteins identifies new E6 binding partners and reflects viral diversity. *Journal of Virology*, 86(24), 13174–13186. <https://doi.org/10.1128/JVI.02172-12>
- World HPV Information. (2017). Human Papillomavirus and Related Diseases Report. HPV Information Centre Report, (Indonesia), 60
- Yang, J., Yan, R., Roy, A., Xu, D., Poisson, J., and Zhang, Y. (2014). The I-TASSER suite: Protein structure and function prediction. *Nature Methods*, 12(1), 7–8. <https://doi.org/10.1038/nmeth.3213>

- Yoshinouchi, M., Yamada, T., Kizaki, M., Fen, J., Koseki, T., Ikeda, Y., Yamato, K. (2003). In vitro and in vivo growth suppression of human papillomavirus 16-positive cervical cancer cells by E6 siRNA. *Molecular Therapy*, 8(5), 762–768. <https://doi.org/10.1016/j.ymthe.2003.08.004>
- Yu, Z., Li, P., and Merz, K. M. (2018). Extended Zinc AMBER Force Field (EZAFF). *Journal of Chemical Theory and Computation*, 14(1), 242–254. <https://doi.org/10.1021/acs.jctc.7b00773>
- Zanier, K., Charbonnier, S., Sidi, A. O. M. hamed O., McEwen, A. G., Ferrario, M. G., Poussin-Courmontagne, P., Travé, G. (2013). Structural basis for hijacking of cellular LxxLL motifs by papillomavirus E6 oncoproteins. *Science*, 339(6120), 694–698. <https://doi.org/10.1126/science.1229934>
- Zanier, K., Charbonnier, S., Sidi, A. O. M. O., McEwen, A. G., Ferrario, M. G., Poussin-Courmontagne, P., ... Trave, G. (2013). Structural Basis for Hijacking of Cellular LxxLL Motifs by Papillomavirus E6 Oncoproteins. *Science*, 339(6120), 694–698. <https://doi.org/10.1126/science.1229934>
- Zanier, K., M'Hamed Ould Sidi, A. O., Boulade-Ladame, C., Rybin, V., Chappelle, A., Atkinson, A., Trav, G. (2012). Solution structure analysis of the HPV16 E6 oncoprotein reveals a self-association mechanism required for E6-mediated degradation of p53. *Structure*. <https://doi.org/10.1016/j.str.2012.02.001>

Anexos

1. Preparación de los sistemas E6/LxxLL para la DM

Script con los parámetros utilizados para preparar cada sistema E6/LxxLL

```
# Eliminación de hidrógenos

1   reduce -Trim prot.pdb > prot_out.pdb

# Re-enumera los átomos y residuos y adapta al lenguaje de LEaP

2   pdb4amber -i prot.pdb -o prot_out.pdb --no-conect

#tleap.in (Script para LEap)

#campos de fuerza

1   source leaprc.gaff

2   source leaprc.protein.ff14SB

3   loadoff atomic_ions.lib

4   loadamberparams frcmod.ionsjc_tip3p

5   loadamberprep ZAFF.prep

6   loadamberparams ZAFF.frcmod

7   addAtomTypes { { "ZN" "Zn" "sp3" } { "S3" "S" "sp3" } { "N2" "N" "sp3" }
}

8   source leaprc.water.tip3p

#carga la proteína

9   mol = loadpdb file.pdb

# Union entre los átomos de Zinc y el azufre de las cisteínas

10  bond mol.No del Zinc .ZN mol.No del residuo de cisteina.SG

11  savepdb mol dirección/file.pdb

12  charge mol

#Tipo y dimensiones de la caja de agua

13  solvatebox mol TIP3PBOX 15
```

```
#llevar el sistema a carga neta cero
```

```
13  addions mol Na+ 0
```

```
14  addions mol Cl- 0
```

```
15  saveamberparm mol dirección/file.prmtop dirección/file.inpcrd
```

2. Protocolo DM

Script con los parámetros para la minimización de energía potencial utilizando el programa Amber16.

```
&cntrl
```

```
1    imin=1,
```

```
2    ntmin=1
```

```
3    nmropt=0
```

```
4    drms=0.1
```

```
5    maxcyc=10000
```

```
6    ncyc=5000
```

```
7    ntx=1
```

```
8    irest=0
```

```
9    ntp=100
```

```
10   ntwr=100
```

```
11   iwrap=0
```

```
12   ntf=1
```

```
13   ntb=1
```

```
14   cut=10.0
```

```
14   nsnb=20
```

```
15   igb=0
```

```
16   ibelly=0
```

```
17 ntr=0
```

```
&end
```

```
/
```

Script con los parámetros para el recocido simulado y equilibrado utilizando el programa Amber16.

```
&cntrl
```

```
    iwrap=1, ntx=1, irect=0, nstlim=8000000, dt=0.001, ntf=2, ntc=2,
temp0=0.0, temp1=310.0, ntp=10000, ntwx=10000, cut=10.0, ntb=1, ntp=0, ntt=3,
gamma_ln=2.0, nmropt=1, ig=-1,
```

```
/
```

```
&wt type='TEMP0', istep1=0, istep2=500000, value1=0.0, value2=300 /
```

```
&wt type='TEMP0', istep1=500001, istep2=1000000, value1=300, value2=300 /
```

```
&wt type='TEMP0', istep1=1000001, istep2=1500000, value1=300.0, value2=400 /
```

```
&wt type='TEMP0', istep1=1500001, istep2=2000000, value1=400, value2=400 /
```

```
&wt type='TEMP0', istep1=2000001, istep2=2500000, value1=400.0, value2=310 /
```

```
&wt type='TEMP0', istep1=2500001, istep2=8000000, value1=310.0, value2=310 /
```

```
&wt type='END' /
```

Script con los parámetros para la DM utilizando el programa Amber16.

```
&cntrl
```

```
    imin=0,
```

```
    iwrap=1,
```

```
    ntx=5,
```

```
    irect=1,
```

```
    nstlim=50000000,
```

```
    dt=0.002,
```

```
    ntf=2,
```

```
ntc=2,  
temp0=310.0,  
ntpr=10000,  
ntwx=10000,  
cut=10.0,  
ntb=2,  
ntp=1,  
ntt=3,  
gamma_ln=2.0,  
ig=-1,  
/  

```

3. Análisis de trayectorias

Scripts con los parámetros para para los diferentes análisis utilizando el programa AmberTool17.

```
#Centrar trayectorias
```

```
1   cpptraj -p file.prmptop  
2   trajin file.nc  
3   autoimage  
4   trajout file.nc
```

```
# RMSD
```

```
1   cpptraj  
2   parm file.prmptom  
3   reference file.inpcrd  
   trajin file.nc
```

```
rms :#-#@CA= reference o first out file.dat o .agr
```

```
# RMSF

1  cpptraj
2  parm file.prmtop
3  loadcrd hfile.nc
4  crdaction file.nc rms first :#-#@CA
5  crdaction file.nc average file.pdb :#-#@CA
6  parm file.pdb
7  reference file.pdb parm file.pdb
8  crdaction file.nc rms reference :#-#@CA
9  crdaction file.nc atomicfluct out file.agr byres :#-#@CA

#Contactos nativos

1  cpptraj
2  parm file.prmtop
3  trajin file.nc
4  nativecontacts :#-# :#-# writecontacts file.dat out file.dat first
distance 5.0

#Distancia entre centros de masa

1  cpptraj
2  distance :#-#@CA :#-#@CA out file.dat

3. MM/GBSA Y MM/PBSA

    Script para el calculo de la enegía de interacción por MM/PBSA y MM/GBSA

&general

    startframe=0,

    endframe=50000,
```

```
interval=20,  
  
keep_files=0,  
  
strip_mask=":WAT:Cl- o Na+",  
  
ligand_mask=": #-#",  
  
receptor_mask=": #-#" /  
  
&gb  
  
    igb=2, saltcon=0.150,  
  
 /  
  
&pb  
  
    radiopt=0, istrng=0.150,  
  
 /  
  
&decomp  
  
    idecomp=1, print_res=" #-#; #-#" /  
  
    dec_verbose=1,  
  
 /
```