

Programa de Maestría en Genética

**Tesis de Maestría**

**Detección, discriminación y cuantificación de  
pequeñas secuencias de nucleótidos por  
Espectroscopía Infrarroja por Transformada de  
Fourier (FTIR) en la identificación molecular del  
Virus Del Papiloma Humano**

**Martha Isabel Guerra Simanca**

Barraquilla (Atl.), Colombia

2021

**Universidad Simón Bolívar**  
Programa de Maestría en Genética

**Detección, discriminación y cuantificación de  
pequeñas secuencias de nucleótidos por  
Espectroscopía Infrarroja por Transformada de  
Fourier (FTIR) en la identificación molecular del  
Virus Del Papiloma Humano**

Tesis presentada al Programa de  
Postgrado en Genética como  
requisito parcial para obtener  
el título de Magister en Genética

**Martha Isabel Guerra Simanca**

Tutor:

**PhD. LEONARDO CARLOS PACHECO LONDOÑO**

Co-Tutor:

**PhD. NATALY JULIETH GALAN FREYLE**

Barraquilla (Atl.), Colombia

2021

## Dedicatoria

A Dios padre, amigo, guía, protector y consejero. Quien nos acompaña en cada uno de nuestros pasos y a su vez es luz en el camino, sorprendiéndonos cada día con sus inagotables bendiciones, enseñanzas, fuerza, honra, gracia y favor. A quien agradezco toda su incomparable fidelidad y el rodear mi vida de personas únicas, talentosas, esforzadas, con diferentes virtudes que sin lugar a dudas son una gran bendición de múltiples formas.

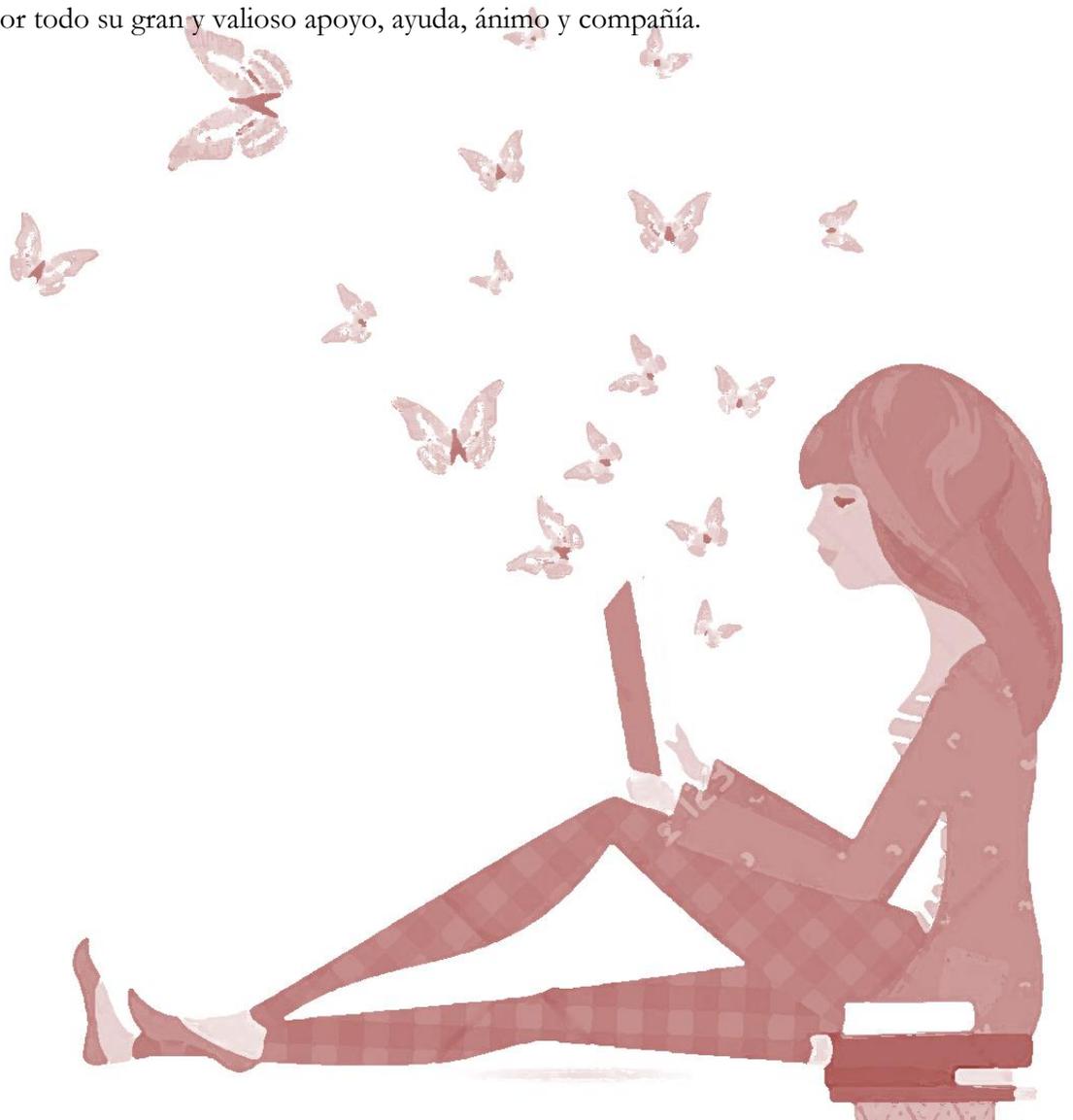
A mis hermosos y amados padres, porque son una gran inspiración en mi vida, por el ánimo en cada una de sus palabras llenas de sabiduría, por su apoyo inigualable y su amor reflejado en sus acciones y oraciones.

A mis hermanitos a quienes Dios hace muchos años cuida con amor, pero que con el privilegio que Dios me otorgó de conocerles, continúan siendo de gran inspiración y valentía en mí caminar.

A mi familia, con quienes Dios me ha bendecido enormemente, por su comprensión, respaldo, consejos y amor incomparable.

A mis amigos por todo su gran y valioso apoyo, ayuda, ánimo y compañía.

**Martha Isabel**



## Agradecimientos

La más sincera gratitud, estima y admiración a mis tutores, los Doctores **Leonardo Pacheco Londoño** y **Nataly Galán Freyle**, de quienes Dios me ha brindado el valioso privilegio de conocer y aprender con su ejemplo como personas y excelentes profesionales, inspirados siempre por sus conocimientos y experiencias a emprender grandes proyectos. Agradezco su gran apoyo, guía, orientación, amabilidad, paciencia y confianza brindada para el desarrollo de este proyecto.

A las doctoras **Sandra Bermeo** y **Dayan Lozano** quienes en el ejercicio de la dirección del programa de Maestría en genética han realizado una notable gran labor por el bienestar y aprendizaje de los estudiantes, así como el fortalecimiento de este maravilloso programa.

A todo el grupo de **personas, docentes, investigadores, trabajadores, directivos** que en conjunto hacen un gran equipo y fortalecen con todos sus aportes, tiempo, conocimientos y disposición la Maestría en Genética, el doctorado en Genética y Biología Molecular, la **Unidad de investigación científica**, desarrollo e innovación en Genética y biología molecular en la Universidad Simón Bolívar, y por supuesto valiosamente a nosotros como estudiantes. A todos, infinitas gracias por sus enseñanzas, dedicación, amable colaboración en cada momento y por su respaldo.

A los doctores **Antonio Acosta** y **Jorge Roper** por su tiempo y valiosos aportes en el proceso de revisión de este trabajo de investigación.

Un muy grato y especial agradecimiento a todas las personas del área de posgrados y aquellas que hacen parte del gran y talentoso equipo del centro de crecimiento empresarial **Macondo Lab**.

**A todos** los que me han brindado su amistad, aprecio, ayuda, amabilidad, acompañamiento y confianza desde mi llegada a Barranquilla, **muchas gracias**.

## Resumen

En este trabajo investigativo se evaluó el uso de la técnica de espectroscopía infrarroja con transformada Fourier por reflexión total atenuada (ATR-FTIR) en la detección, discriminación y cuantificación de pequeñas secuencias de nucleótidos aplicado en la identificación molecular de genotipos de virus de papiloma humano (VPH). En una primera parte se demostró la viabilidad de la técnica ATR-FTIR en la diferenciación de pequeñas secuencias de ADN de una sola cadena, para esto se generó un modelo de regresión para la cuantificación del porcentaje de nucleótido (%N, para cada nucleótido %A, %C, %T y %G) por el método multivariado de mínimos cuadrados parciales (PLS) y se analizaron las señales espectrales por ATR-FTIR de estas secuencias. El error de los modelos para la cuantificación del %N estuvo entre 0.9-1.2%.

En una segunda parte se implementó la técnica ATR-FTIR en la identificación de VPH. Para esto se generaron modelos multivariados por el método de análisis discriminante PLS (PLS-DA) para la predicción de los genotipos de VPH 16, 31, 35, 51 y 66. Esto se realizó a partir de espectros de los productos de amplificados de ADN por la técnica de reacción en cadena de la polimerasa en tiempo real (real-time PCR), los modelos fueron diseñados usando los espectros ATR-FTIR de los controles positivos y negativos de los productos de real-time PCR. Estos modelos fueron usados para predecir muestras clínicas de seis mujeres y los resultados fueron contrastados con la técnica convencional real-time PCR. Todas las muestras fueron predichas con el mismo genotipo de VPH validado por real-time PCR.

**Palabras clave:** ATR-FTIR, VPH, ADN, nucleótidos, cáncer cervicouterino, PLS, PLS-DA, real-time PCR.

## **Abstract**

In this research, the Fourier transform infrared spectroscopy by attenuated total reflection (ATR-FTIR) technique is used to detect, discriminate, and quantify small nucleotide sequences. This was applied in the molecular identification of genotypes of human papillomavirus (HPV). Initially, I demonstrated the viability of the ATR-FTIR technique in the differentiation of small single-stranded DNA sequences, for this, a model was generated to quantify the nucleotide percentage (% N, for each nucleotide % A, % C, % T, and % G) by the multivariate method of partial least squares (PLS) and the spectral signals of these sequences were analyzed by ATR-FTIR. The error of the models for the quantification of % N was between 0.9-1.2%. Finally, to identify HPV, I implemented the ATR-FTIR technique. For this, multivariate models were generated by the PLS discriminant analysis method (PLS-DA) in the prediction of HPV genotypes 16, 31, 35, 51 and 66. This was performed from spectra of DNA amplification products using the real-time polymerase chain reaction (real-time PCR) technique; the models were designed using the ATR-FTIR spectra of the positive and negative controls of the real-time PCR products. These models were used to predict clinical samples from six women, and the results were contrasted with the conventional real-time PCR technique. All samples were predicted with the same HPV serotype found by real-time PCR.

**Keywords:** ATR-FTIR, HPV, DNA, nucleotides, cervical cancer, PLS, PLS-DA, real-time PCR.

# Contenido

	Pág.
Lista de Figuras .....	17
Lista de tablas.....	19
Introducción .....	20
<b>1. Problema de Investigación .....</b>	<b>22</b>
1.1 Planteamiento del Problema.....	22
1.2 Justificación.....	24
<b>2. Objetivos .....</b>	<b>25</b>
2.1 Objetivo General .....	25
2.2 Objetivo específicos .....	26
2.2.1 Demostrar la viabilidad de la técnica ATR-FTIR en la diferenciación de pequeñas secuencias de ADN.....	26
2.2.1.1 <i>Actividad 1:</i> Preparación y Adquisición de espectros de ATR-FTIR para las diferentes secuencias de nucleótidos previamente diseñados. ....	26
2.2.1.2 <i>Actividad 2:</i> Generación y validación modelo estadístico para cuantificación de porcentaje de nucleótidos mediante los espectros ATR-FTIR para las pequeñas secuencias .....	26
2.2.2 Implementar la técnica ATR-FTIR en la identificación de genotipos de virus de papiloma humano.....	26
2.2.2.1 <i>Actividad 3:</i> Selección de las muestras clínicas.....	26
2.2.2.2 <i>Actividad 4:</i> Caracterización de las muestras por la técnica estándar de real- time PCR y electroforesis .....	26
2.2.2.3 <i>Actividad 5:</i> Preparación y adquisición de espectros de FTIR para los diferentes controles positivos, negativos y muestras amplificadas por real-time PCR.....	26
2.2.2.4 <i>Actividad 6:</i> Generación y validación de modelos estadístico para discriminación mediante los espectros para los diferentes genotipos de VPH. ....	26
<b>3. Marco Referencial.....</b>	<b>27</b>
3.1. Antecedentes .....	27
3.2 Virus del papiloma humano .....	29
3.2.1 Estructura, genoma y proteínas virales.....	29
3.2.2 Clasificación del VPH .....	32
3.2.3 Ciclo de vida.....	33
3.2.4 Comportamiento molecular del VPH en infecciones de bajo riesgo .....	34
3.2.5 Comportamiento molecular del VPH en infecciones de alto riesgo .....	34
3.3 Reacción en cadena de la polimerasa en tiempo real (real-time PCR) y temperatura de melting (T <sub>m</sub> ) .....	35
3.3.1 Temperatura Melting .....	37
3.4 Fundamentos de espectroscopía vibracional Infrarroja.....	38

3.4.1 Radiación electromagnética.....	38
3.4.2 Espectrometría de Infrarrojo por Transformada de Fourier (FTIR).....	40
3.4.3 Instrumentación.....	42
3.4.4 Sistema de Reflectancia total atenuada (Attenuated Total Reflectance, ATR)....	44
3.3 Interpretación espectral infrarroja .....	45
3.3.1 Quimiometría y Modelación.....	46
<b>4. Diseño Metodológico .....</b>	<b>49</b>
4.1 Tipo de Estudio .....	49
4.2 Demostración de la viabilidad de la técnica ATR-FTIR en la diferenciación de pequeñas secuencias de nucleótidos. ....	49
4.2.1 Preparación y adquisición de espectros de ATR-FTIR para las diferentes secuencias previamente diseñadas.....	49
4.2.2 Generación y validación del modelo estadístico para la cuantificación de porcentaje de nucleótidos mediante los espectros ATR-FTIR para las pequeñas secuencias.....	51
4.3 Implementación de la técnica ATR-FTIR en la identificación de genotipos de virus de papiloma humano .....	53
4.3.1 Selección de las muestras clínicas .....	53
4.3.2 Caracterización de las muestras por la técnica estándar de real-time PCR y electroforesis.....	53
4.3.3 Preparación y adquisición de espectros de ATR-FTIR para los diferentes controles positivos, negativos y muestras amplificadas por real-time PCR.....	55
<b>5. Resultados y Discusión .....</b>	<b>58</b>
5.1 Viabilidad de la técnica ATR-FTIR en la diferenciación de pequeñas secuencias de ADN .....	58
5.1.1 Asignación de bandas de ATR-FTIR para los diferentes desoxinucleótidos y secuencias de nucleótidos previamente diseñados. ....	58
5.1.2. Generación y validación modelo estadístico para cuantificación de porcentaje de nucleótidos mediante los espectros ATR-FTIR para las pequeñas secuencias .....	63
5.2 Implementación de la técnica ATR-FTIR en la identificación de genotipos de virus de papiloma humano (VPH) .....	66
5.2.1 Caracterización de las muestras por espectrofotometría UV, la técnica estándar de real-time PCR y electroforesis. ....	66
5.2.2 Preparación y adquisición de espectros de FTIR para los diferentes controles positivos, negativos y muestras amplificadas por real-time PCR. ....	74
5.2.3 Generación y validación de modelos estadísticos para discriminación mediante los espectros para los diferentes genotipos de VPH.....	77
<b>6. Conclusiones y Recomendaciones.....</b>	<b>83</b>
6.1 Conclusión .....	83
6.2 Recomendaciones.....	84
<b>7. Referencias Bibliográficas.....</b>	<b>85</b>

# Lista de Figuras

	Pág.
<b>Figura 1.</b> Estructura de la cápside del VPH. A) Micrográficas por microscopía crioelectrónica (crio-EM) de las cápsides de partículas de VPH16; B) Representación tridimensional obtenida por crio-EM de la cápside de VPH compuesta por 72 capsómeros (60 hexavalente y 12 pentavalentes) en forma de estrellas; C) Sección interna central del VPH; D) Estructura pentamérica de la proteína principal de la cápside (L1) del VPH16 <sup>(66,67)</sup> .....	30
<b>Figura 2.</b> Diagrama del genoma del virus de papiloma humano tipo 16 y proteínas virales <sup>(62)</sup> .....	32
<b>Figura 3.</b> Ciclo de vida del VPH en infección cervical. La infección inicia producto de microabrasiones epiteliales que permiten la entrada del virus hasta las células basales. En el proceso de división, algunas células se propagan lateralmente y otras migran hacia las capas de células supra basales, en el que se activan los genes virales, formándose copias de ADN viral y las proteínas de la cápside. Se liberan las partículas virales propagando la población de células infectadas, así mismo pueden integrarse en el genoma humano, evento en su mayoría asociado a la progresión de estas lesiones a cáncer <sup>(74)</sup> .....	33
<b>Figura 4.</b> Cinética de amplificación y parámetros empleados en la cuantificación por real-time PCR. Rn: magnitud de la señal de emisión de fluorescencia. Threshold (Umbral de detección): promedio de la desviación estándar de la fluorescencia. Baseline (Línea base): ciclos iniciales de la PCR con cambio mínimo en la señal de fluorescencia. Ct: número de ciclos en el cual la fluorescencia alcanza el threshold fijado. No template (Muestra sin molde): Muestra que no contiene templado. $\Delta Rn$ : Intensidad de la emisión de la fluorescencia <sup>(79)</sup> .....	37
<b>Figura 5.</b> Representación gráfica de una onda electromagnética propagada a lo largo del eje X. El campo eléctrico oscila en un plano perpendicular al campo magnético <sup>(83)</sup> .....	38
<b>Figura 6.</b> Dominios del espectro electromagnético, niveles energéticos y límites de las regiones <sup>(85)</sup> .....	39
<b>Figura 7.</b> Vibraciones moleculares en la región media del espectro Infrarrojo (MIR) <sup>(47)</sup> ....	40
<b>Figura 8.</b> Tipos de vibración molecular. El signo positivo indica movimiento desde el plano de la página hacia el vector, y el signo negativo en la dirección opuesta. Montañó y Vargas 2009. ....	41
<b>Figura 9.</b> Esquema básico de un espectrómetro FTIR <sup>(91)</sup> .....	43
<b>Figura 10.</b> Interacción del haz infrarrojo y la muestra en una celda ATR, A) de rebote único, B) de rebote múltiple <sup>(94)</sup> .....	44
<b>Figura 11.</b> Evaluación gradual del procesamiento de muestras. ....	46
<b>Figura 12.</b> Espectros ATR-FTIR promediados y normalizados de desoxiadenosina trifosfato (dATP), desoxicitidina trifosfato (dCTP) y desoxiguanosina trifosfato (dGTP) desoxitimidina trifosfato (dTTP) en el rango 3800 a 400 $\text{cm}^{-1}$ .....	58
<b>Figura 13.</b> Espectros ATR-FTIR promediados y normalizados de polinucleótidos de adenina (A), citosina (C), guanina (G) y timina (T) en el rango 3800 a 400 $\text{cm}^{-1}$ .....	60

**Figura 14.** Valores predichos versus valores de referencia en los modelos aplicados para el porcentaje de: A) % Adenina, B) %Citosina, C) %Guanina y D) %Timina, mediante regresión por mínimos cuadrados parciales (PLS)..... 65

**Figura 15.** Curvas de temperatura de melting ( $T_m$ ) obtenidas para los controles positivos de VPH 16,18,31,35,51 y 66, posterior al proceso de real-time PCR SYBR Green. Eje y: Primera derivada del perfil de fluorescencia 1stD(RFU); Eje x: Temperatura en °C..... 70

**Figura 16.** Curvas de temperatura de melting ( $T_m$ ) obtenidas para las muestras amplificadas de ADN positivas para VPH 51, posterior al proceso de real-time PCR SYBR Green. A)  $T_m = 78.5$  para las muestras M01-M04; B)  $T_m = 78$  para las muestras M05 y M06; C)  $T_m = 77.7$  para las muestras M07-M14. Eje y: Primera derivada del perfil de fluorescencia 1stD(RFU); Eje x: Temperatura en °C..... 71

**Figura 17.** Curvas de temperatura de melting ( $T_m$ ) obtenidas para las muestras amplificadas de ADN positivas para VPH 66, posterior al proceso de real-time PCR SYBR Green. A)  $T_m = 79.5$  para la muestra M15; B)  $T_m = 79.8$  para la muestra M16. Eje y: Primera derivada del perfil de fluorescencia 1stD(RFU); Eje x: Temperatura en °C..... 71

**Figura 18.** Electroforesis en gel de agarosa de los productos real-time PCR amplificados. A) Controles positivos para los genotipos de VPH 16, 18, 31, 35, 51 y 66. B) Controles positivos para los genotipos de VPH 51, VPH 66 y las muestras M10 y M13 con detección positiva del tipo de VPH 51. C) Controles positivos para los genotipos de VPH 51, VPH 66 y la muestra M16 con detección positiva del tipo de VPH 66. .... 73

**Figura 19.** Espectros ATR-FTIR de la solución de PowerUp™ SYBR® Green Master Mix (Thermo Fisher Scientific) y los controles positivos amplificados de VPH 16, 18, 31, 35, 51 y 66, promediados y desarrollados por normalización del vector en el rango 3800 a 400  $\text{cm}^{-1}$ . Eje y:  $\Delta$  Absorbancia; Eje x: Número de onda ( $\text{cm}^{-1}$ )..... 75

**Figura 20.** Espectros ATR-FTIR de los controles positivos amplificados de VPH 16, 18, 31, 35, 51 Y 66 posterior al proceso de sustracción espectral compensada de la solución de PowerUp™ SYBR® Green Master Mix (Thermo Fisher Scientific). Regiones del espectro infrarrojo comprendidas entre diferentes rangos: **A)** 3663 a 2399  $\text{cm}^{-1}$ ; **B)** 1800 a 1160  $\text{cm}^{-1}$ ; **C)** 1160 a 810  $\text{cm}^{-1}$ ; **D)** 810 a 400  $\text{cm}^{-1}$ . Eje y:  $\Delta$  Absorbancia; Eje x: Número de onda ( $\text{cm}^{-1}$ ). .... 76

**Figura 21.** Tablas de confusión de los modelos de predicción evaluados para el conjunto de datos en la detección de los tipos de VPH 16, 18, 31, 35, 51 y 66. Se presentan los valores de precisión, sensibilidad, F1-Score y exactitud resultantes de los cálculos en cross validation y test para cada modelo. .... 78

**Figura 22.** Resultados de la ejecución de los modelos de detección para los 6 genotipos de VPH analizados y modelo de predicción de muestras negativas. (+) Resultado positivo predicho por el modelo correspondiente; (-) resultado predicho como negativo. En rojo se presentan los resultados de predicción atípicos. .... 80

**Figura 23.** Gráficos de dispersión PLS-DA de los espectros FTIR empleados para la evaluación de los modelos para predicción de los tipos de VPH 16(A),18(B), 31(C), 35(D) ,51(E) y 66(F) y negativos(G), en los espectros de muestras clínicas amplificadas,

controles positivos amplificados, solución de Master mix y controles negativos. Línea punteada (---) indica el valor del umbral equivalente a 0.5..... 81

## Lista de tablas

	<b>Pág.</b>
<b>Tabla 1.</b> Secuencias nucleotídicas previamente diseñados empleados para análisis de viabilidad de la técnica ATR-FTIR para la cuantificación del porcentaje de nucleótidos. %N: número de bases, porcentajes de Adenina (A%), Timina (%), Guanina (%G) y Citosina (%C). Los datos en rojo representan las secuencias usadas para test de validación externa. ....	49
<b>Tabla 2.</b> Secuencias de primers específicos correspondientes a los genotipos de VPH 16,35,18,31,51 y 66 <sup>(86)</sup> .....	54
<b>Tabla 3.</b> Asignación de modos vibracionales a las bandas observadas en los espectros ATR-FTIR de los desoxinucleótidos dATP, dTTP, dCTP y dGTP en el rango de 3400 a 400 cm <sup>-1</sup> . v: vibración tipo estiramiento, δ: vibración tipo flexión, df: deformación, sim: modo simétrico, asm: modo asimétrico, Out-of-plane: fuera del plano, Gu: guanina, Ty: timina, Cy: citosina., D-ribosa: azúcar Desoxirribosa, P: grupo fosfato, gly: enlace glicosídico, Vib(ring): vibración del anillo.....	59
<b>Tabla 4.</b> Asignación de modos vibracionales a las bandas observadas en los espectros ATR-FTIR en el rango de 3800 a 700 cm <sup>-1</sup> de las secuencias de polinucleótidos empleadas. v: vibración tipo estiramiento, δ: vibración tipo flexión, df: deformación, sim: modo simétrico, asm: modo antisimétrico, D-ribosa: azúcar Desoxirribosa, P: grupo fosfato, gly: enlace glicosídico, Vib(ring): vibración del anillo, Gu: guanina, Ty: timina. ....	61
<b>Tabla 5.</b> Valores de los parámetros relevantes del modelo PLS para la cuantificación de porcentajes de nucleótidos. ....	63
<b>Tabla 6.</b> Concentración de ADN (ng/μL) obtenida por espectrofotometría UV para cada muestra de estudio y sus respectivas proporciones 260/280 y 260/230. ....	66
<b>Tabla 7.</b> Valores de las temperaturas melting (T <sub>m</sub> ) resultantes en las amplificaciones de las muestras de ADN y genotipo de VPH asociado.....	68
<b>Tabla 8.</b> Secuencias de primers específicos correspondientes a los tipos de VPH 16, 35, 18, 31, 51 y 66, longitudes en pares de bases (bp) de los productos amplificados reportados con base a Tsakogiannis <i>et al.</i> 2015 <sup>(86)</sup> y temperaturas melting (T <sub>m</sub> ) en °C determinadas mediante real-time PCR SYBR Green.....	72

## Introducción

El cáncer de cuello uterino (CCU) es el cuarto cáncer más común en mujeres en todo el mundo, estimando que cada año se presentan 570,000 casos y aproximadamente 311,000 muertes a causa de esta enfermedad<sup>(1)</sup>. Datos epidemiológicos han confirmado el virus del papiloma humano (VPH) como principal agente causal de carcinoma cervical<sup>(2,3)</sup>. Alrededor del 85% de las muertes por CCU ocurren en países en subdesarrollo, donde esta tasa supera 18 veces los valores registrados en países desarrollados<sup>(1,4,5)</sup>.

El VPH pertenece a la familia *Papillomaviridae* y está conformado por una cápside icosaédrica de 72 pentámeros y un virión no envuelto con un genoma de ADN circular de doble cadena de 8 kb de tamaño que codifica para una región de control local larga, 6 proteínas tempranas (E1, E2 y E4, E5, E6 y E7) y 2 proteínas tardías (L1 y L2)<sup>(6)</sup>.

Los mecanismos principales a través de los cuales el VPH contribuye a la iniciación y progresión neoplásica incluyen la actividad de E6 y E7, oncoproteínas virales, que interfieren con las proteínas supresoras tumorales del ciclo celular p53 y la proteína retinoblastoma (pRb), permitiendo el desarrollo de mutaciones genéticas y eventualmente promoviendo la transformación celular maligna<sup>(2)</sup>. Sin embargo, existen otros factores importantes que contribuyen al proceso de múltiples etapas de la formación de tumores como las variaciones genéticas individuales y factores como el consumo de tabaco, alta carga viral, múltiples parejas sexuales, enfermedades de transmisión sexual, multiparidad, el uso de anticonceptivos y la inmunosupresión<sup>(7,8)</sup>.

Existen aproximadamente 150 diferentes tipos de VPH. Los genotipos de VPH de bajo riesgo, como el VPH 6, 11, 42, 43 y 44, están asociados en su mayoría con lesiones cutáneas benignas que a menudo forman verrugas. Por el contrario, los VPH de alto riesgo como el VPH 16, 18, 31, 33, 34, 35, 39, 45, 51, 52, 56, 58, 59, 66, 68 y 70 están relacionados con infecciones de la mucosa y conllevan un alto riesgo de progresión maligna y desarrollo de CCU<sup>(1,9,10)</sup>.

La complementación de la citología cervical con diferentes técnicas de detección viral ofrece mayor sensibilidad a los procesos de tamizaje vigentes, permitiendo la identificación de subtipos del VPH asociado; donde, esta información podría

orientar estrategias de tratamiento y prevención incluso cuando no hay lesiones visibles y solo existen células escamosas de significado incierto<sup>(1,7,11,12)</sup>.

Como parte de las pruebas moleculares para la identificación de los genotipos de VPH presentes en la población femenina que acuden al servicio de ginecología, la reacción en cadena de polimerasa (PCR) es utilizada. Esta es una técnica molecular específica, sensible y automatizable. La PCR en tiempo real (real-time PCR) es una de las variantes más utilizadas que hace posible el seguimiento de la cinética de cada reacción en la muestra problema<sup>(13-15)</sup>. El análisis de la curva de fusión posterior a la amplificación constituye una forma simple y directa de verificar las reacciones de real-time PCR y garantizar la especificidad del producto obtenido; considerando que este representa el cambio en la fluorescencia observada cuando el ADN bicatenario (dsDNA) con moléculas de colorante incorporadas se disocia en el ADN monocatenario (ssDNA) a medida que aumenta la temperatura de la reacción y alcanza la temperatura de melting (Tm)<sup>(16)</sup>.

Otro método que ha tomado auge en el estudio de sistemas biológicos es la espectroscopía infrarroja por transformada de Fourier (FTIR), técnica ampliamente utilizada para el análisis químico en muestras biomédicas, debido a que particularmente en la región espectral del infrarrojo medio se reflejan los modos vibratorios fundamentales de las principales biomoléculas celulares como proteínas, carbohidratos, lípidos y ácidos nucleicos, demostrado ser una técnica que ofrece múltiples ventajas. Además, de proporcionar un enfoque cuantitativo rápido, preciso y de alto rendimiento<sup>(17-20)</sup>.

En el presente trabajo se evaluó la aplicación de la técnica FTIR a partir de espectros de referencias de nucleótidos, secuencias aleatorias cortas y secuencias genotípicas de controles positivos de VPH. Estos espectros de referencia se utilizaron para la generación de información que permitiera la identificación de genotipos de VPH en muestras clínicas. Basado en la premisa que la técnica FTIR arroja información única para cada ente molecular (huella digital), la discriminación entre los espectros obtenidos por espectroscopia infrarroja se basó en función de la presencia, ausencia y corrimientos de las bandas de absorción específicas de las biomoléculas, como una medida de las variaciones bioquímicas entre las diferentes muestras analizadas.

# 1. Problema de Investigación

## 1.1 Planteamiento del Problema

El VPH es la infección de transmisión sexual más común y la mayoría de la población sexualmente activa entra en contacto con el virus a lo largo de su vida, aún en muchos casos sin evidenciarse síntomas<sup>(21,22)</sup>. El VPH es un virus altamente transmisible, el cual puede estar presente en el organismo aproximadamente dos años; no obstante, gran parte de las infecciones se resuelven espontáneamente sin ningún tratamiento<sup>(22,23)</sup>. Un evento clave en la progresión neoplásica de las lesiones cervicales es la persistencia de integración del genoma del VPH en el cromosoma de las células epiteliales cervicales del huésped. A nivel molecular se reconocen aproximadamente más de 35 genotipos capaces de infectar el epitelio de superficies cutáneas en todo el cuerpo y mucosas del área ano-genital<sup>(24,25)</sup>.

El VPH es un virus con ADN de doble cadena, posee en su genoma ocho genes, de los cuales E6 y E7 codifican, en genotipos de alto riesgo, oncoproteínas implicadas en la inhibición de la apoptosis, la respuesta inmune y la inducción de una proliferación celular excesiva, que con el tiempo son los causantes del desarrollo del CCU, una patología de elevada morbilidad y mortalidad en mujeres principalmente en países en subdesarrollo, donde el tamizaje primario con citología cervical no ha logrado reducir la prevalencia de las fases avanzadas de la enfermedad, así como los recursos y opciones de acceso a planes para la detección, prevención y control del cáncer que se ofrecen son limitados<sup>(4,7,26,27)</sup>.

En gran parte de los casos la detección de la presencia de VPH no se realiza a tiempo ya que puede que no se presenten síntomas relacionados<sup>(28,29)</sup>. La forma más tradicional de detectar el virus es por medio de la citología o prueba del papanicolau, un procedimiento en el que se recoge una muestra de las células de las paredes del cuello del útero y la vagina para su posterior análisis, en el que se puede detectar si hay anomalías o cambios en estas, sin embargo, la técnica presenta determinadas desventajas que hace que disminuya su efectividad y confiabilidad al analizar sus resultados<sup>(30-32)</sup>. Esta exhibe una especificidad con valores entre 73 y 98 % y una sensibilidad entre el 41 y 84 %, que redundan en un número significativo de falsos negativos con porcentajes hasta del 67% de los casos, por lo que debe ser repetida en intervalos cortos de tiempo<sup>(33-35)</sup>.

La alta posibilidad de suministrar un diagnóstico erróneo, puede desencadenar

consecuencias en la salud humana, promoviendo la ansiedad y procedimientos de seguimiento innecesarios, así, como causar el aumento de personas infectadas y morbilidad por VPH debido a la falta de un tratamiento oportuno<sup>(5,29,36-38)</sup>.

La q-PCR es uno de los métodos moleculares de detección empleados en el diagnóstico de enfermedades virales. La identificación específica de los genotipos de virus asociados, se puede lograr mediante el análisis de la curva de fusión. Sin embargo, es necesario tener en cuenta que la Tm de los ácidos nucleicos se ve afectada por la longitud, el contenido de Guanina y Citosina (G y C), la presencia de desajustes de bases y dímeros de primers; factores de cuidado para el óptimo desarrollo de esta técnica ya que pueden dificultar la interpretación de los resultados<sup>(39,40)</sup>.

El aumento del número de casos registrados con posible infección por VPH, genera a su vez la necesidad de abordar el uso de técnicas de alto rendimiento, más sensibles y específicas cuyos métodos analíticos sean complementarios en el proceso de investigación y diagnóstico, que permitan diferenciar y detectar varios subtipos virales a la vez. Por lo tanto, la gran demanda para desarrollar métodos analíticos nuevos y complementarios hace posible, por ejemplo, el abordaje de técnicas basadas en espectroscopia infrarroja por transformada de Fourier (FTIR)<sup>(38,41)</sup>.

Con base a lo expuesto anteriormente se planteó el siguiente interrogante: ¿Cómo se puede evaluar el uso de la espectroscopía infrarroja de transformada de Fourier (FTIR) en la detección, discriminación y cuantificación de pequeñas secuencias de nucleótidos aplicado en la identificación molecular de genotipos del virus del papiloma humano (VPH) en complemento con la técnica de PCR en tiempo real (real-time PCR)?

## 1.2 Justificación

La infección generada por el virus del papiloma humano de alto riesgo (VPH-AR) es la etiología principal para el desarrollo del cáncer cervical. La patogénesis por este tipo de virus implica la sobreexpresión de oncoproteínas que pueden inhibir proteínas importantes como p53 y pRB en las células del huésped y afectar procesos biológicos que incluyen la proliferación, el ciclo celular y la apoptosis. Este conjunto de alteraciones que inducen la carcinogénesis conllevan a la revisión profunda de la naturaleza de la enfermedad abarcando diferentes aspectos e inspiran al desarrollo de técnicas y métodos dirigidos específicamente a su comprensión, detección y diagnóstico<sup>(5,42)</sup>.

El CCU se puede prevenir en muchos casos con una detección adecuada. La efectividad equivalente o superior de las pruebas de moleculares de VPH en la detección primaria, conducen al uso cada vez mayor de éstas como metodologías de cribado en la prevención del cáncer<sup>(25,29)</sup>.

La real-time PCR es una de las técnicas moleculares ampliamente utilizada en la detección, cuantificación y genotipado de agentes infecciosos como el VPH. Uno de los compuestos empleados para la emisión de fluorescencia en esta técnica es el SYBR-Green, lo que permite análisis de las curvas de temperatura melting para confirmar la especificidad del producto amplificado<sup>(15,43,44)</sup>.

Dado que este análisis mencionado usa la Tm de productos de doble cadena para la identificación del amplicon resultante, existen varios aspectos a considerar como preparación de la muestra, la formación de dímeros de primer, similitudes entre las secuencias, porcentaje similar de G/C o mutaciones específicas de algún genotipo, los cuales pueden promover la formación de productos inespecíficos en la reacción que pueden ser detectados y generar múltiples picos o una variación significativa de la Tm causando errores en la genotipificación<sup>(45,46)</sup>.

En consideración de lo anterior, dentro de las técnicas espectroscópicas vibracionales se ha demostrado la viabilidad de la FTIR como herramienta de diagnóstico médico adicional para diferentes trastornos y enfermedades, entre ellas cánceres como el CCU<sup>(47)</sup>. Específicamente, utilizando la espectroscopía de reflectancia total atenuada ATR-FTIR se pueden dilucidar vibraciones de enlace específicas de moléculas dadas, a través de la interacción de la luz infrarroja (IR) con los componentes biológicos de la muestra de estudio, lo que proporciona una visión única de la composición a través de un espectro de absorbancia<sup>(48)</sup>.

La ATR-FTIR es una técnica analítica simple, no invasiva que puede caracterizar el perfil bioquímico sin una preparación extensa de la muestra, requiriendo un volumen pequeño de esta, el cual se pipetea sobre un cristal, donde el IR interactúa de manera no destructiva y genera picos representativos de las vibraciones de enlaces<sup>(47)</sup>. Este conjunto de beneficios permite considerar que esta metodología contribuiría enormemente en el campo clínico, considerando además como fundamento de esta técnica que cualquier alteración en los sistemas biológicos inducida por una condición patológica produce cambios estructurales y funcionales significativos que se reflejan directamente en los espectros vibratorios<sup>(49,50)</sup>.

Durante el proceso de análisis espectral es necesario un procesamiento que involucra una potencia informática considerable cuando es aplicada al conjunto de datos. Estos requieren de espectros de referencia que permitirán la comprensión de las variaciones intrínsecas en las muestras bajo estudio, donde el uso de funciones de normalización y corrección de la línea de base, seguidos de modelos de clasificación para la comparación de resultados, permitirán que las conclusiones con respecto al análisis tengan buenos valores de sensibilidad y especificidad<sup>(51)</sup>.

Los avances investigativos en técnicas como ATR-FTIR permiten el aporte de información prometedora y consistente en función de proporcionar metodologías de detección que contribuyan en conjunto con otros procedimientos a mejorar el pronóstico para pacientes con enfermedades como el CCU<sup>(36,37,48)</sup>.

## **2. Objetivos**

### **2.1 Objetivo General**

Evaluar el uso de la espectroscopia infrarroja de transformada de Fourier por reflexión total atenuada (ATR-FTIR) en la detección, discriminación y

cuantificación de pequeñas secuencias de nucleótidos aplicado en la identificación molecular de genotipos del virus del papiloma humano (VPH).

## **2.2 Objetivo específicos**

**2.2.1** Demostrar la viabilidad de la técnica ATR-FTIR en la diferenciación de pequeñas secuencias de ADN.

2.2.1.1 *Actividad 1:* Preparación y Adquisición de espectros de ATR-FTIR para las diferentes secuencias de nucleótidos previamente diseñados.

2.2.1.2 *Actividad 2:* Generación y validación modelo estadístico para cuantificación de porcentaje de nucleótidos mediante los espectros ATR-FTIR para las pequeñas secuencias

**2.2.2** Implementar la técnica ATR-FTIR en la identificación de genotipos de virus de papiloma humano

2.2.2.1 *Actividad 3:* Selección de las muestras clínicas

2.2.2.2 *Actividad 4:* Caracterización de las muestras por la técnica estándar de real-time PCR y electroforesis

2.2.2.3 *Actividad 5:* Preparación y adquisición de espectros de FTIR para los diferentes controles positivos, negativos y muestras amplificadas por real-time PCR.

2.2.2.4 *Actividad 6:* Generación y validación de modelos estadístico para discriminación mediante los espectros para los diferentes genotipos de VPH.

## 3. Marco Referencial

### 3.1. Antecedentes

La investigación en el campo de la espectroscopía Infrarroja ha realizado diferentes avances desde los primeros ensayos de la descomposición de la luz en bandas de diferentes colores, la medición de las temperaturas a lo largo de los extremos espectrales, entre otras características. Durante los primeros años del siglo XIX empezó a relacionarse la naturaleza química de una sustancia con un espectro particular; analizándose de esta forma los procesos de absorción y emisión de luz y el diseño de los instrumentos para la obtención de espectros. El conjunto de experimentos realizados durante el desarrollo de la espectroscopía permitió su avance y aplicación en química analítica <sup>(52-55)</sup>.

Las técnicas basadas en espectroscopia vibracional, tales como la espectroscopia infrarroja han demostrado ser una herramienta poderosa para la caracterización químico-estructural, así como la obtención del perfil químico característico de múltiples compuestos. Esta técnica es rápida, no destructiva, realiza mediciones confiables y requiere poca cantidad y manipulación de la muestra <sup>(8)</sup>.

Las diferentes ventajas que ofrece esta técnica ha permitido su aplicabilidad en diferentes áreas: agricultura, biotecnología, cosméticos, industria farmacéutica, mineralogía, control medioambiental, control de alimentos y bebidas, ciencia forense, medicina y química clínica, entre otros <sup>(19,47,56,57)</sup>. A continuación, se mencionan algunos trabajos investigativos con implementación de FTIR, relacionados con el presente estudio.

Sahu *et al.*,<sup>(58)</sup> evaluaron por FTIR los cambios espectrales en células y tejidos normales y malignos de diferentes órganos (colon, cuello uterino, piel y sangre), con respecto a la absorbancia obtenida en los ácidos nucleicos, respaldando con sus resultados el uso de absorbancia de ácido nucleico produce parámetros estadísticamente significativos, que podrían diferenciar tejidos normales y cancerosos.

Khommohamadi *et al.*,<sup>(56)</sup> propusieron como método de inspección y análisis de diferentes biofluidos y tejidos el uso de la espectroscopia infrarroja (IR), una técnica no invasiva, rápida, precisa y sensible que permite la discriminación de cambios bioquímicos en el diagnóstico de cáncer. En su estudio realizaron descripciones de diferentes investigaciones biomédicas dedicadas al diagnóstico de cáncer de colon, mama y cuello uterino basado en IR. De esta forma concluyen

que la IR junto con el procesamiento quimiométrico de datos espectrales es un enfoque útil para investigación y el diagnóstico en oncología.

Ghimire *et al.*,<sup>(59)</sup> Evaluaron cambios bioquímicos inducidos por el linfoma no Hodgkin y el melanoma subcutáneo en modelos de ratones EL4 y B16 respectivamente. Mediante ATR-FTIR realizaron análisis de muestras de suero secas extraídas de los de modelos de ratones portadores de tumores de linfoma y melanoma. En su estudio obtuvieron diferencias espectrales con cambios graduales en las intensidades de la absorción de proteínas, carbohidratos y ácidos nucleicos entre los casos control y tumorales, teniendo en cuenta la comparación de valores p, la posición espectral, el análisis de la relación de los picos de amida, el ajuste de los espectros experimentales y el empleo de análisis multivariante (HCA) obteniendo. Los autores resaltan ATR-FTIR como una técnica espectroscópica rápida y confiable para la discriminación de ratones con melanoma B16 y linfoma EL4 de sus tipos de control.

Rymza *et al.*,<sup>(11)</sup> en su trabajo de investigación analizaron la eficacia de las técnicas PCR y ATR-FTIR para el diagnóstico de infección por VPH en un total de 41 mujeres que se sometieron a exámenes ginecológicos. Para análisis por PCR emplearon los cebadores GP05 / GP06 y MY09-MY11 con el objetivo de confirmar la infección por VPH mediante la amplificación del gen L1 presente en el genoma del virus. Los espectros de absorbancia adquiridos por ATR-FTIR se realizaron en un rango de 750–4000  $\text{cm}^{-1}$  a una resolución de 4  $\text{cm}^{-1}$  por muestra. Los investigadores observaron una concordancia entre los resultados obtenidos por ambos métodos, se reafirma la importancia de la técnica de ATR-FTIR, debido a las ventajas que ofrece como, mayor precisión, mayor rendimiento y menor carga de trabajo para el citólogo y/o patólogo.

Roy *et al.*,<sup>(60)</sup> evaluaron la capacidad de la técnica de ATR-FTIR, combinada con análisis de datos multivariados para la clasificación de muestras de suero humano en función de detectar la presencia de virus de la hepatitis B (VHB) y la hepatitis C (VHC). Los espectros se recogieron entre 4200-600  $\text{cm}^{-1}$ , los principales cambios observados en los espectros se relacionan con los niveles de inmunoglobulina (Ig), lípidos y glucosa. Se observó una banda indicadora a 1093  $\text{cm}^{-1}$  en el caso de la infección por VHB. Los autores muestran ATR-FTIR como un método prometedor de detección inicial de una infección subyacente y capacidad de discriminación de las muestras infectadas por el VHB y el VHC, siendo necesarios mayor cantidad de investigaciones al respecto.

## 3.2 Virus del papiloma humano

Los papilomavirus pertenecen taxonómicamente a la familia *Papillomaviridae*, esta familia se clasifica en 18 géneros, cada uno se divide en especies que contienen diversos subtipos de virus de papiloma. Sólo cinco de estos géneros presentan subtipos que pueden afectar al ser humano, entre ellos se encuentra<sup>(48)</sup>.

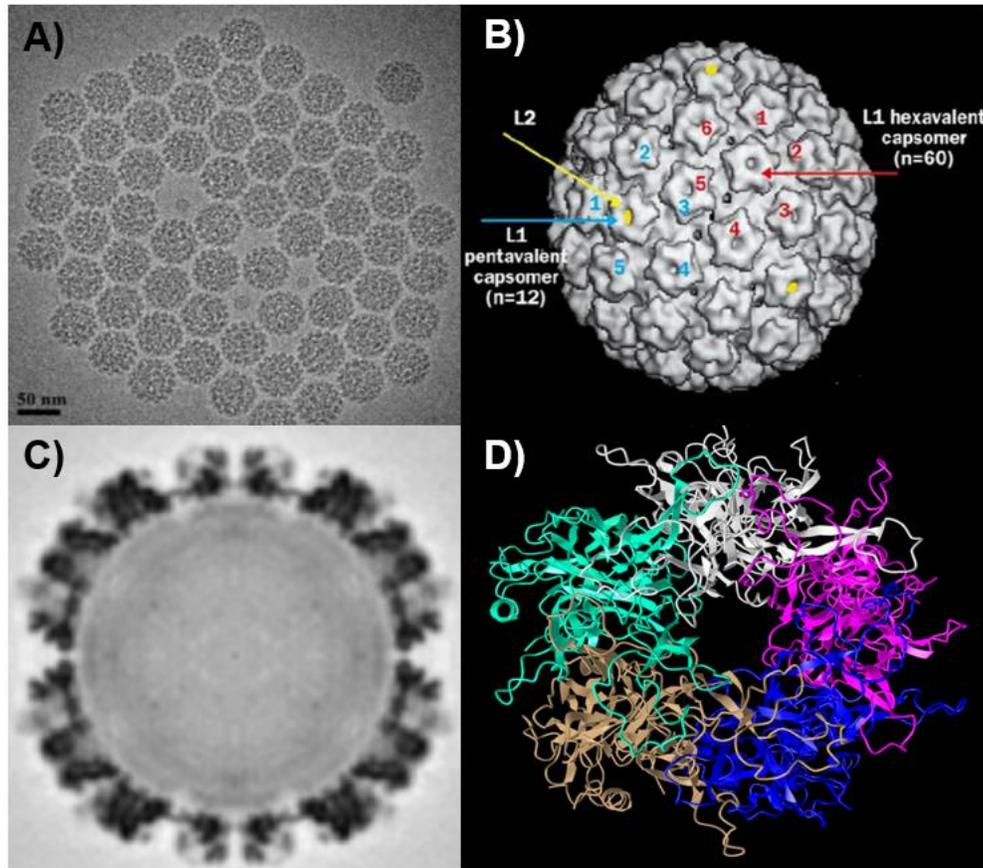
- Alfa-papilomavirus: infectan tanto el epitelio cutáneo como el epitelio mucoso del tracto anogenital. Causan desde verrugas comunes a lesiones neoplásicas.
- Beta-papilomavirus: generalmente asintomáticas, se manifiestan provocando lesiones subclínicas (verrugas o condilomas acuminados) o infecciones latentes. Sin embargo, estas manifestaciones son de importancia clínica por ejemplo en individuos inmunodeprimidos.
- Gamma-papilomavirus: generalmente asintomáticas, pueden causar papilomas y verrugas cutáneas.
- Mu-papilomavirus: causan papilomas y verrugas cutáneas.
- Nu-Papilomavirus: causan papilomas y verrugas cutáneas<sup>(61-63)</sup>.

### 3.2.1 Estructura, genoma y proteínas virales

Las partículas del VPH no presentan envoltura, miden entre 45 y 55 nm de diámetro, están constituidos por una doble cadena de ADN de forma circular con una longitud entre 6800 y 8400 pares de bases (pb), revestido por una cápside de proteínas de forma icosaédrica compuesta por 72 capsómeros pentaméricos (Figura 1)<sup>(64)</sup>.

El genoma está dividido en tres regiones principales: una región no codificante de aproximadamente 1 kb, conocida como región larga de control (LCR, Long Control Region), también denominada región reguladora no codificante (URR, Upstream Regulatory Region); una región localizada en posición adyacente y corriente abajo del LCR que incluye genes de expresión temprana (E, early); y una región ubicada en posición adyacente y corriente abajo de la región temprana que contiene genes de expresión tardía (L, late), de acuerdo a si son expresados antes o después de la síntesis del ADN destinado a ser ensamblado en las partículas de progenie viral y que dan origen a proteínas estructurales y no estructurales respectivamente<sup>(10)</sup>. En total se encuentran 9 o 10 marcos de lectura abierta (ORFs, Open Reading Frames), localizados en una sola de las hebras del ADN genómico. En el VPH,

siete u ocho de las regiones ORFs codifican para genes tempranos y únicamente dos para genes tardíos (Figura 2) <sup>(64,65)</sup>.



**Figura 1.** Estructura de la cápside del VPH. **A)** Micrografías por microscopía crioelectrónica (crio-EM) de las cápsides de partículas de VPH16; **B)** Representación tridimensional obtenida por crio-EM de la cápside de VPH compuesta por 72 capsómeros (60 hexavalente y 12 pentavalentes) en forma de estrellas; **C)** Sección interna central del VPH; **D)** Estructura pentamérica de la proteína principal de la cápside (L1) del VPH16 <sup>(66,67)</sup>.

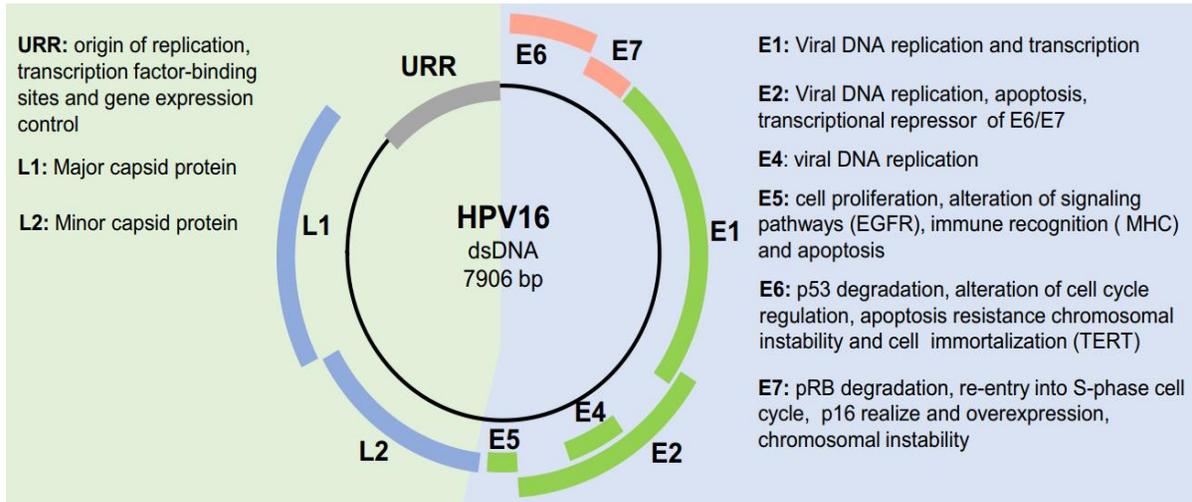
La zona LCR o URR se encuentra localizada entre la zona temprana y tardía del genoma, no codifica proteínas, pero contienen las secuencias encargadas de la regulación de la expresión de genes, de esta forma controla el origen de la replicación, los genes promotores E6 y E7, los potenciadores y silenciadores <sup>(68)</sup>.

Los genes E codifican proteínas reguladoras necesarias en los procesos de replicación, transcripción y transformación celular. E1 está implicado en la replicación y mantenimiento del ADN viral; E2 regula la transcripción viral y facilita

la distribución del genoma viral en las células hijas durante el período de división. Además, esta proteína por medio de la región LCR reprime la actividad de los promotores E6 y E7<sup>(68,69)</sup>.

E1 participa en la ruptura de la cromatina del ADN de la célula huésped, está implicado en la replicación y mantenimiento del ADN viral. E2 regula la transcripción viral y facilita la distribución de las copias del ADN viral a células hijas durante la mitosis celular. Además, por medio de la región LCR reprime la actividad de los promotores E6 y E7<sup>(68,69)</sup>. E3 se encuentra relacionado con E6 y E7 como control de la transcripción de los mismos. E4: aparece en las capas supra basales y granuladas, interactuando con las fibras amiloides y asegurando la salida del virión ante una apoptosis celular. E5 interacciona con receptores de factores de crecimiento, participa en la amplificación del ADN viral y tiene la capacidad de promover la evasión a la respuesta del sistema inmune por parte del complejo mayor de histocompatibilidad (MHC; del inglés, major histocompatibility complex), e inhibir procesos de apoptosis. E6 ayuda en la degradación de p53 y permite la inmortalización de la célula huésped y E7 ayuda a degradar la proteína de retinoblastoma (pRB), lo que facilita el reingreso al ciclo S y ayuda a la expresión de p16. Se considera que E6 y E7 juegan un papel importante en el incremento de la proliferación celular y en la extensión de la vida celular en infecciones por VPH asociadas a malignidad, debido a diferentes factores reguladores que intervienen y alteran el ciclo celular<sup>(63,68)</sup>, provocando que no sea posible la eliminación del ADN viral y la corrección de los errores intrínsecos al ADN celular, de modo que la célula acumula alteraciones genéticas y adquiere un fenotipo neoplásico<sup>(69,70)</sup>.

La región de genes L codifica las proteínas estructurales de la cápside mayor y menor (L1 y L2, respectivamente). Ambas proteínas L, se encuentran implicadas en la invasión del virus a las células epiteliales<sup>(10)</sup>. La proteína L1 se expresa posteriormente de L2 en la replicación del ciclo viral y regula el empaquetamiento de las partículas virales en el virión ensamblado. La proteína L2 interactúa con la proteína E2, facilitando el transporte de la proteína L1 al núcleo y participa en el encapsulamiento del ADN viral<sup>(68)</sup>.



**Figura 2.** Diagrama del genoma del virus de papiloma humano tipo 16 y proteínas virales <sup>(62)</sup>.

### 3.2.2 Clasificación del VPH

Para el proceso de clasificación de los papilomavirus principalmente se toman dos criterios básicos: a) el hospedero, ya que estos virus son altamente específicos de acuerdo a la especie, y, b) secuencias genéticas, que permiten la distinción entre diferentes aislamientos de manera detallada. La secuencia más utilizada para la clasificación de los papilomavirus es la del gen que codifica para la proteína de la cápside L1, que es altamente conservada en el genoma viral, aunque también se han empleado otros genes como E6 y E7 <sup>(71)</sup>.

Los subtipos de VPH pertenecientes a diferentes géneros presentan una variación dentro de la región L1 de más del 40 % en el genoma viral. Para la clasificación en diferentes especies virales, el porcentaje de variación debe ser entre 30 y 40 % en esta misma región. Un nuevo tipo de VPH se define si el genoma completo ha sido secuenciado y la secuencia de nucleótidos correspondiente a la del gen L1 difiere por lo menos en un 10 % con respecto a otro tipo viral ya caracterizado. Si la variación en la secuencia de nucleótidos está entre el 1 y el 10 %, se considera una variante o linaje viral, y si está entre 0,5 y 1,0 %, como un nuevo sublinaje <sup>(10,71)</sup>.

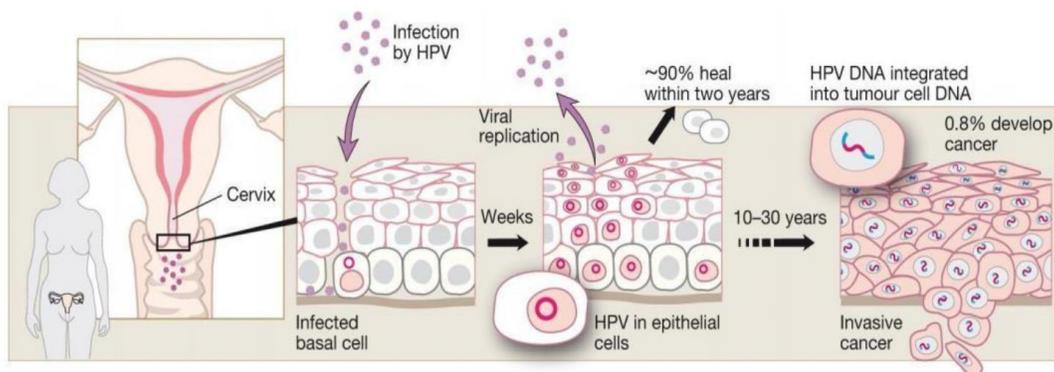
Basados en el punto de vista clínico, los papilomavirus humanos (ubicados en el género Alfa-papilomavirus) han sido divididos en dos grupos: los de bajo riesgo, asociados principalmente con verrugas genitales benignas y los de alto riesgo, que son los agentes etiológicos del cáncer cervicouterino <sup>(64)</sup>.

Atendiendo al potencial oncogénico de los subtipos virales asociados a la mucosa anogenital, la agencia internacional contra el cáncer (IARC) propone la siguiente clasificación<sup>(72,73)</sup>:

- **Grupo 1 (alto riesgo):** VPH 16, VPH 18, VPH 31, VPH 33, VPH 35, VPH 39, VPH 45, VPH 51, VPH 52, VPH 56, VPH 58, VPH 59;
- **Grupo 2A (probablemente carcinogénico):** VPH 68;
- **Grupo 2B (posiblemente carcinogénico):** VPH 26, VPH 30, VPH 34, VPH 53, VPH 66, VPH 67, VPH 69, VPH 70, VPH 73 y VPH 82, VPH 58, VPH 97.
- **Grupo 3 (bajo riesgo):** VPH 6, VPH 11.

### 3.2.3 Ciclo de vida

Para el proceso replicativo del material genético del VPH es necesaria la maquinaria celular de un huésped. El ciclo vital del VPH comienza con la infección de una capa basal (epidermoide o mucosa), por medio de la entrada del virus ocasionada por microtraumatismos que comprometen la barrera epitelial (Figura 3). Cuando las células basales completan un ciclo mitótico el genoma viral se incorpora al núcleo celular y se replica para lograr la proliferación de las células infectadas; en este punto se mantiene un bajo número de copias y a su vez una baja expresión de las proteínas E1, E2, E6 y E7 (8,13,27). Posteriormente, Las células infectadas migran desde la membrana basal a las capas más superficiales para diferenciarse, el virus se replica alcanzando un elevado número de copias y expresa los genes de la cápside (L1 y L2), dando como resultado la producción de nuevos viriones de progenie liberados al exterior sin que se produzca lisis celular (63,69,74,75).



**Figura 3. Ciclo de vida del VPH en infección cervical.** La infección inicia producto de microabrasiones epiteliales que permiten la entrada del virus hasta las células basales. En el proceso de división, algunas células se propagan lateralmente y otras migran hacia las capas de células supra basales, en el que se activan los genes virales, formándose copias de ADN

viral y las proteínas de la cápside. Se liberan las partículas virales propagando la población de células infectadas, así mismo pueden integrarse en el genoma humano, evento en su mayoría asociado a la progresión de estas lesiones a cáncer <sup>(74)</sup>.

### **3.2.4 Comportamiento molecular del VPH en infecciones de bajo riesgo**

Con base en el ciclo vital del virus explicado anteriormente, en los VPH de bajo riesgo, se expresan inicialmente las proteínas E1 y E2, encargadas de mantener el ADN viral de forma episomal, tanto en la infección inicial como en la infección latente. Al expresarse E2, se reprime la acción de E6 y E7 provocando la inhibición del proceso de replicación del ADN viral de célula a célula, al bloquear los factores de transcripción celular y controlando el número de copias del ADN viral en las células basales. Las nuevas copias del ADN viral son empacadas en sus cápsides, integradas por la acción de las proteínas L1 y L2, posteriormente los viriones son liberados para reiniciar un nuevo ciclo viral. La actividad de los VPH de bajo riesgo es menos agresiva, se expresan en las células suprabasales, y aunque pueden generar alteraciones morfológicas en las células epiteliales, estas son menos marcadas que en los casos de infección con virus de alto riesgo <sup>(68)</sup>.

### **3.2.5 Comportamiento molecular del VPH en infecciones de alto riesgo**

En los casos de infección por VPH de alto riesgo, la pérdida de expresión de E2 provoca que los ORFs codifiquen las proteínas E6 y E7, las cuales se expresan en las células infectadas del estrato basal, una vez que el ADN viral se integra al ADN de la célula huésped. La integración del VPH al genoma celular ocurre mayormente en regiones denominadas “hotspots”, en regiones transcripcionalmente activas, el 66% dentro o junto a zonas frágiles del ADN del huésped y cerca de micro ARNs (miARNs). Estos últimos están asociados con la regulación de procesos importantes, como el desarrollo, la proliferación, la diferenciación y la apoptosis celular y se encuentran regularmente alterados en células cancerosas <sup>(68,76,77)</sup>.

Después de la integración, inicia la replicación viral, se observa proliferación de queratinocitos en el estrato intermedio y superficial del epitelio, apareciendo alteraciones estructurales y celulares.

Los subtipos de alto riesgo son más propensos a activar la proliferación celular y activar varias vías esenciales para la transformación epitelial. Dentro de los diferentes mecanismos celulares involucrados se pueden mencionar: disminución

de la dependencia celular a los factores de crecimiento, así como la exposición al sistema inmune por medio de la activación de la proteína E5; crecimiento celular incontrolado producto de la acción de la proteína E6, la cual promueve la degradación de la proteína supresora de tumores p53; la proteína E7 modifica el ciclo celular mediante la unión, producción o degradación de pRB y otras proteínas, activa la replicación estimulando el paso de la fase G1 a la fase S del ciclo celular; E7 se une y degrada a las proteínas *pocket* de retinoblastoma (pRB, p130 y p107, estas dos últimas también llamadas RBL2 y RBL, respectivamente) e induce cambios moleculares, entre ellos la producción del factor de transcripción E2F, que desregula el punto de control o “checkpoint” del ciclo celular entre las fases G1 y S provocando una reentrada a la fase S y la activación de la replicación viral (1,28). La activación de E7 provoca la aparición de una respuesta celular de defensa: la sobreexpresión de p16INK4A (inhibidor de quinasas dependientes de ciclinas, impidiendo el paso de la fase G1 y S, p16), proteína supresora de tumores que estaba reprimida por la actividad del complejo p53. Los mecanismos celulares derivados de la producción de p16, ralentizan el ciclo celular provocando una senescencia inducida por oncogenes, lo que contrarresta los efectos provocados por E6. Todos estos cambios inducen la proliferación celular, inhiben la diferenciación y generan una inestabilidad cromosómica que favorece la aparición de anomalías cromosómicas y la carcinogénesis <sup>(68,78)</sup>.

### **3.3 Reacción en cadena de la polimerasa en tiempo real (real-time PCR) y temperatura de melting (T<sub>m</sub>)**

La PCR (reacción en cadena de la polimerasa) es una técnica rápida y altamente sensible con un amplio uso en entornos investigativos y en la práctica clínica. La PCR está basada en la amplificación exponencial de ácidos nucleicos por acción de la polimerasa termoestable *Thermus aquaticus* (Taq). El método utiliza un par de cebadores de oligonucleótidos sintéticos, cada uno de los cuales se hibrida con una hebra de un ADN bicatenario (dsDNA) de interés, abarcando una región que se amplificará exponencialmente. A los cebadores apareados se añade la ADN polimerasa Taq, creando una hebra de ADN complementaria mediante la adición secuencial de desoxinucleótidos (dNTP) que han sido añadidos a la mezcla de reacción<sup>(79)</sup>. Experimentalmente, el proceso normalmente consta de tres pasos:

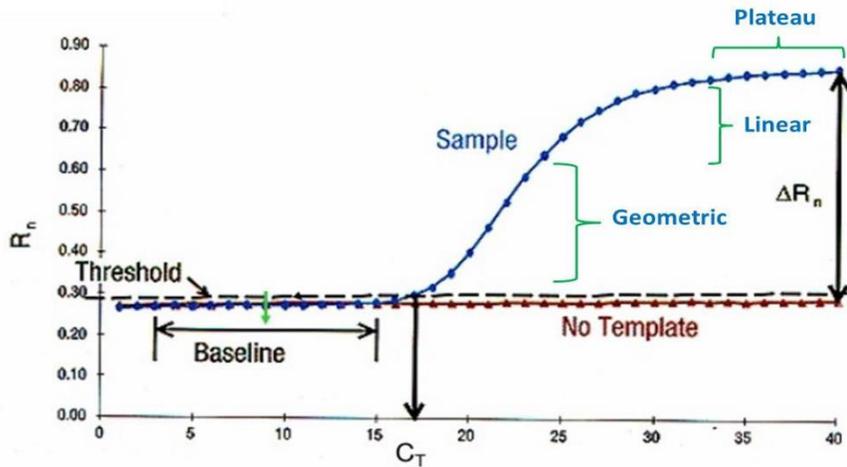
1. Desnaturalización (94 o 95 °C): La doble hélice de ADN se separa en dos hebras simples.

2. Anillamiento (55-65 °C): los cebadores se unen específicamente a los fragmentos de ADN monocatenarios que son complementarios a su propia secuencia de nucleótidos. La ADN polimerasa se une al sitio del cebador de ADN.

3. Extensión (72 °C): la ADN polimerasa sintetiza la nueva cadena de ADN mediante la adición de nucleótidos complementarios al ADN molde.

Con cada ciclo de los anteriores tres pasos, el número de copias de la molécula de ADN se duplica, generando miles de millones de amplicones después de 20 a 40 ciclos de PCR. Posteriormente los amplicones se visualizan por corrido electroforético en gel de agarosa, que separa los productos de ADN en función del tamaño<sup>(15)</sup>.

Para el caso de uso de la real-time PCR, en esta técnica se utilizan los mismos componentes que una PCR convencional solo que se agrega fluorescencia a cada reacción antes de empezar el proceso de amplificación. La acumulación del producto de PCR es monitorizada en tiempo real mientras se lleva a cabo la amplificación, y ciclo a ciclo son analizados los cambios de la señal fluorescente generada durante las tres etapas del PCR (Desnaturalización, alineamiento, y extensión). Cuanto mayor sea la cantidad de ADN muestra, menos serán los ciclos necesarios para la detección de una señal fluorescente. La amplificación por real-time PCR sigue una cinética exponencial en donde se pueden distinguir (Figura 4): la fase geométrica o exponencial, donde, en cada ciclo la cantidad de producto se duplica, asumiendo una eficiencia de reacción del 100%; en la fase lineal la disminución de la actividad de la ADN polimerasa Taq y el consumo de reactivos como dNTP y cebadores provocan que se haga más lenta la reacción; y en la fase estacionaria o plateau, la reacción ha finalizado y no se sintetizan más ácidos nucleicos debido al agotamiento de los reactivos y la Taq polimerasa. Los agentes reporteros más comunes empleados en este tipo de ensayos son: SYBR Green, un intercalante de bases de ADN, y las sondas TaqMan, que son fragmentos de ADN específicos que tienen unido un fluoróforo y un quencher (represor o agente bloqueador)<sup>(74,75)</sup>.



**Figura 4.** Cinética de amplificación y parámetros empleados en la cuantificación por real-time PCR. **R<sub>n</sub>**: magnitud de la señal de emisión de fluorescencia. **Threshold** (Umbral de detección): promedio de la desviación estándar de la fluorescencia. **Baseline** (Línea base): ciclos iniciales de la PCR con cambio mínimo en la señal de fluorescencia. **C<sub>t</sub>**: número de ciclos en el cual la fluorescencia alcanza el threshold fijado. **No template** (Muestra sin molde): Muestra que no contiene templado. **ΔR<sub>n</sub>**: Intensidad de la emisión de la fluorescencia<sup>(79)</sup>.

### 3.3.1 Temperatura Melting

Mediante el uso de colorantes fluorescentes que se unen al ADN, es posible obtener resultados de la reacción del producto PCR mientras aumenta la temperatura.

A medida que aumenta la temperatura, las moléculas de ADN de doble hebra comenzarán a disociarse, liberando las moléculas de colorante que se incorporaron previamente al producto durante la ejecución de la PCR. Esta liberación de fluorescencia disminuirá, generando una curva de disociación, fusión o melting ( $T_m$ ). A partir de la  $T_m$  es posible identificar las características del producto amplificado. Por ejemplo, un producto de ADN más corto tendrá una  $T_m$  más baja que un producto más largo, ya que los enlaces covalentes entre las dos moléculas de ADN son más débiles y se disociarán a una temperatura más baja<sup>(80)</sup>.

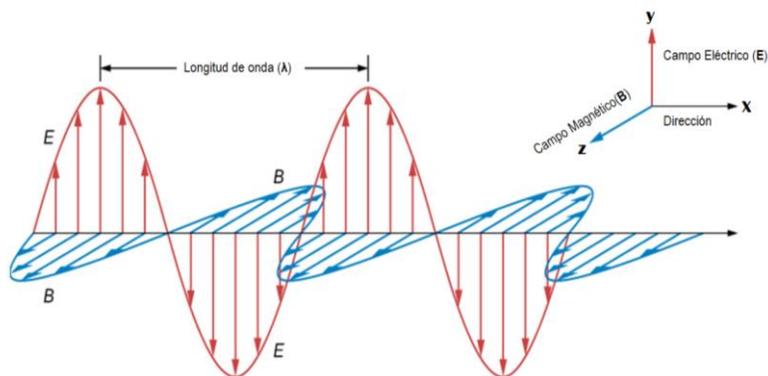
Los datos brindados por el software de análisis de las reacciones real-time PCR para la  $T_m$ , se utilizan para la generación de dos tipos de curvas: la curva de intensidad normalizada y la primera derivada negativa de la curva de intensidad normalizada.

La curva de intensidad normalizada muestra la temperatura (T) frente a la intensidad de fluorescencia normalizada ( $R_n$ ). En una aplicación de curva de fusión, la temperatura aumenta con el tiempo, lo que hace que las hebras de ADN se disocien. Esta disociación hace que el  $R_n$  disminuya monótonamente, formando una curva de forma sigmoideal. En la primera derivada negativa la temperatura de fusión se define convencionalmente como el punto de inflexión de la curva de intensidad normalizada. La ubicación del punto de inflexión se puede determinar utilizando la curva de la primera derivada negativa la cual presenta la temperatura (T) frente a la primera derivada negativa de la intensidad normalizada ( $-dR_n / dT$ ). Por lo tanto, el pico de la curva de derivada negativa es el punto de inflexión de la curva de intensidad normalizada. El análisis de temperatura de fusión selecciona el pico más alto asociado con la temperatura de fusión más alta dentro de un rango de detección definido por el investigador<sup>(16,81)</sup>.

### 3.4 Fundamentos de espectroscopía vibracional Infrarroja

#### 3.4.1 Radiación electromagnética

La radiación electromagnética (radiación EM) se define como la energía propagada en forma de ondas eléctricas y magnéticas (Figura 5) que viajan en paquetes de energía denominados fotones, los cuales intervienen individualmente en los procesos elementales de emisión, absorción y dispersión<sup>(82)</sup>.



**Figura 5.** Representación gráfica de una onda electromagnética propagada a lo largo del eje X. El campo eléctrico oscila en un plano perpendicular al campo magnético<sup>(83)</sup>.

Todas las formas de radiación EM se agrupan en un espectro de acuerdo con variables como la longitud de onda y frecuencia.

La frecuencia ( $f$ ) es el número de ciclos que completa la onda por segundo, mientras que la longitud de onda ( $\lambda$ ) es la distancia que la onda requiere para completar un ciclo. El producto de estos dos términos es igual una constante que equivale a la velocidad de propagación de la onda ( $c$ ), donde, en el vacío es aproximadamente de  $3 \times 10^8$  m/s. Esto es:

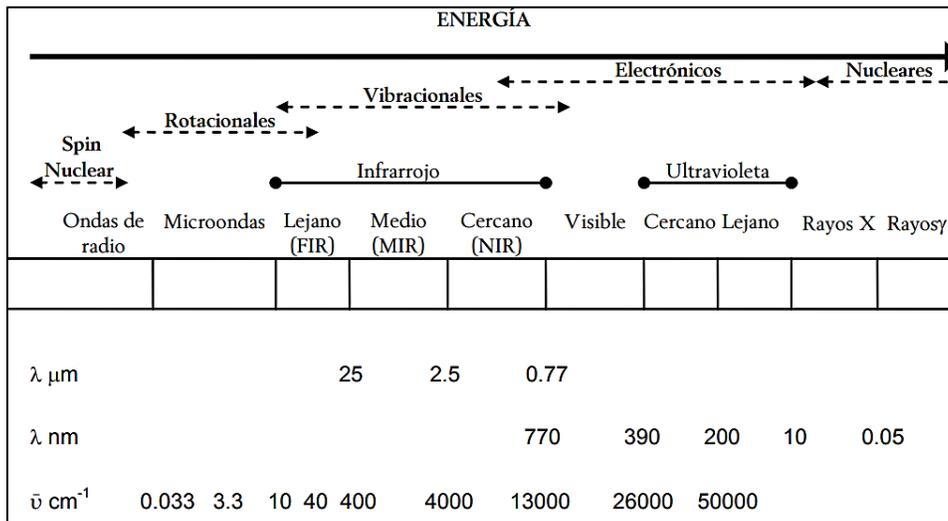
$$c = \lambda f. \text{ Ecuación 1}$$

Considerando una onda electromagnética como un paquete de energía llamado “fotón” o “cuanto”, cada fotón contiene una cantidad discreta de energía electromagnética (EM), dependiente de o de la frecuencia de  $f$  o  $\lambda$  de la siguiente manera:

$$E = hf = \frac{hc}{\lambda} \text{ Ecuación 2}$$

Donde  $h$  es la constante de Planck y tiene un valor de  $6.63 \times 10^{-34}$  Js.

Dentro de este espectro electromagnético pueden distinguirse los rayos gamma, rayos X, luz ultravioleta, luz visible, luz infrarroja, microondas y las ondas de radio<sup>(84)</sup>.



**Figura 6.** Dominios del espectro electromagnético, niveles energéticos y límites de las regiones<sup>(85)</sup>.

### 3.4.2 Espectrometría de Infrarrojo por Transformada de Fourier (FTIR)

Esta técnica espectroscópica recibe su nombre de la región del espectro electromagnético implicada, la región Infrarroja (IR), la cual se divide en infrarrojo cercano 12800-4000  $\text{cm}^{-1}$ , infrarrojo medio 4000-200  $\text{cm}^{-1}$  e infrarrojo lejano 200–100  $\text{cm}^{-1}$  y se encuentra entre las regiones de luz visible y microondas <sup>(17)</sup>.

Se basa en la absorción de energía de un fotón que posteriormente promueve la transición de un menor a un mayor estado de energía o excitación. La absorción de la radiación electromagnética por parte de la materia se produce cuando la radiación incidente tiene la energía precisa para satisfacer los requerimientos energéticos del material, y posteriormente se genera un acoplamiento o interacción. Para que este último se presente la molécula debe experimentar un cambio en el momento dipolar cuando tiene lugar la vibración fundamental <sup>(17)</sup>.

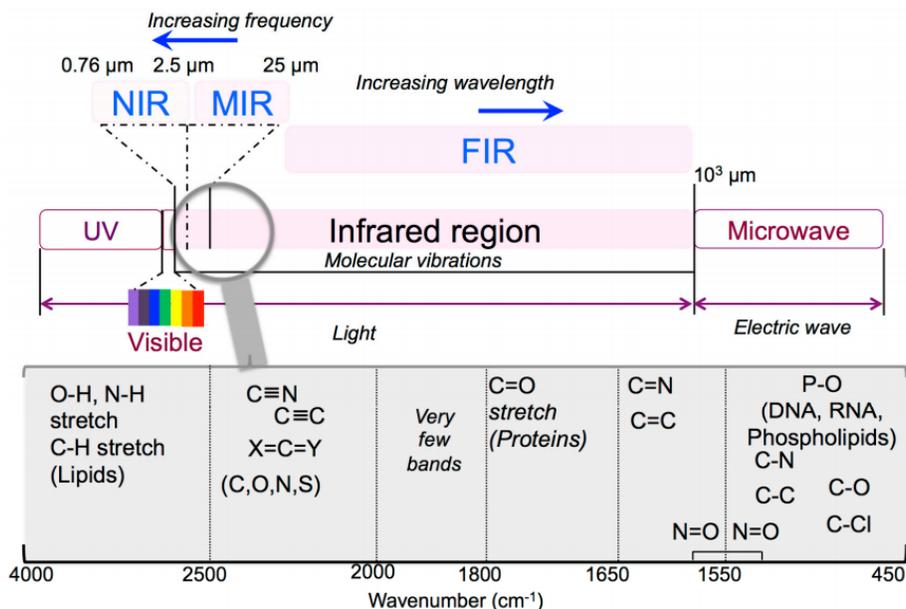
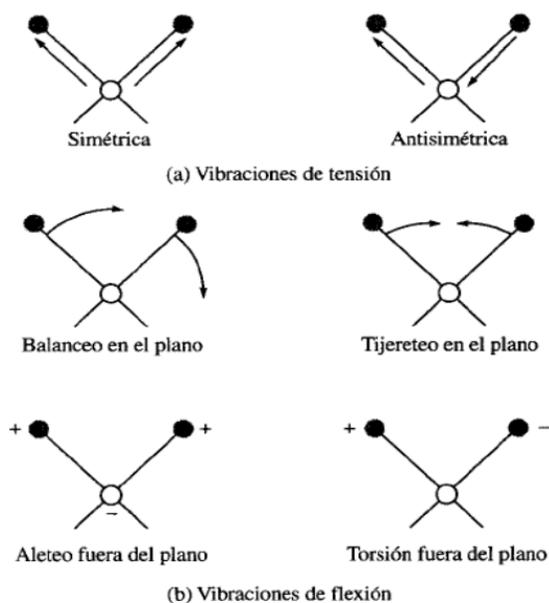


Figura 7. Vibraciones moleculares en la región media del espectro Infrarrojo (MIR)<sup>(47)</sup>.

Un espectro de absorción en el infrarrojo representa varios factores que incluyen la absorbancia o el porcentaje de transmitancia versus la frecuencia ( $\text{s}^{-1}$ ) o longitud de onda ( $\lambda$ , nm) o número de onda ( $\text{cm}^{-1}$ ).

Dada la complejidad del movimiento vibracional, el análisis del espectro de infrarrojo comprende un número de vibraciones constitutivas, llamado modos normales de vibración, que dependen del tipo y número de átomos que contenga

la molécula y sus grados de libertad, así como están asociados a las transiciones entre niveles de energía vibracional diferentes dentro del mismo estado energético. Las vibraciones producidas en los enlaces moleculares pueden ser de estiramiento, flexión, torsión, balanceo o deformación, produciendo también un número de ondas variables (o frecuencias) en la región IR del espectro<sup>(8,86)</sup>. Cada una de estas vibraciones está asignada a un modo de vibración “fundamental” o “normal”. Cualquier molécula no lineal de n número de átomos posee  $3n-6$  modos normales, mientras que una molécula lineal tiene  $3n-5$ . En la Figura 8 se muestran esquemáticamente algunos tipos de vibraciones moleculares.



**Figura 8.** Tipos de vibración molecular. El signo positivo indica movimiento desde el plano de la página hacia el vector, y el signo negativo en la dirección opuesta. Montaña y Vargas 2009.

La espectroscopía infrarroja es un método no destructivo, empleado en diferentes campos para el análisis, identificación y caracterización de sustancias o muestras en estado sólido, líquido y gaseoso, considerándose de gran utilidad para el estudio de la estructura secundaria de sistemas complejos como moléculas biológicas, proteínas, ADN y membranas<sup>(8,53)</sup>.

La introducción de software basado en la aplicación del algoritmo de transformación de Fourier, entre otras técnicas matemáticas han llevado a

avances importantes en la Espectroscopía IR, permitiendo a su vez la ampliación del campo de acción de esta técnica <sup>(87,88)</sup>.

Este tipo de espectroscopía, conocida como FTIR por sus siglas en inglés para “Fourier Transform Infrared”, es una técnica que consiste en la absorción de energía procedente de una fuente de luz infrarroja en longitudes de onda específicas correspondientes a las frecuencias de resonancia o vibración de los átomos, brindando de esta forma información sobre los grupos funcionales de las moléculas que constituyen la muestra de análisis <sup>(8,86)</sup>.

La implementación de un espectrómetro FTIR implica disponer de una pequeña cantidad de la muestra en una celda, la cual es incidida con un barrido de luz infrarroja con número de onda que pueden ir desde 4000 hasta 400  $\text{cm}^{-1}$ , donde, la intensidad de la luz transmitida por la muestra se mide en cada número de onda (inverso de longitud de onda o  $\text{cm}^{-1}$ ). De esta forma, la diferencia entre la intensidad de la luz incidente sobre la muestra y la transmitida por esta permitirá conocer la cantidad de luz absorbida <sup>(8)</sup>.

La interpretación de los espectros es un aspecto muy importante que se realiza de manera directa por comparación del espectro de la muestra con una colección de espectros de sustancias conocidas. El espectro vibracional de una molécula se considera una propiedad física única de la misma, lo que permite que la técnica FTIR encabece la lista de las metodologías altamente sensibles que permiten la caracterización y el reconocimiento de las conformaciones y configuraciones de las moléculas en diferentes materiales <sup>(8,86)</sup>.

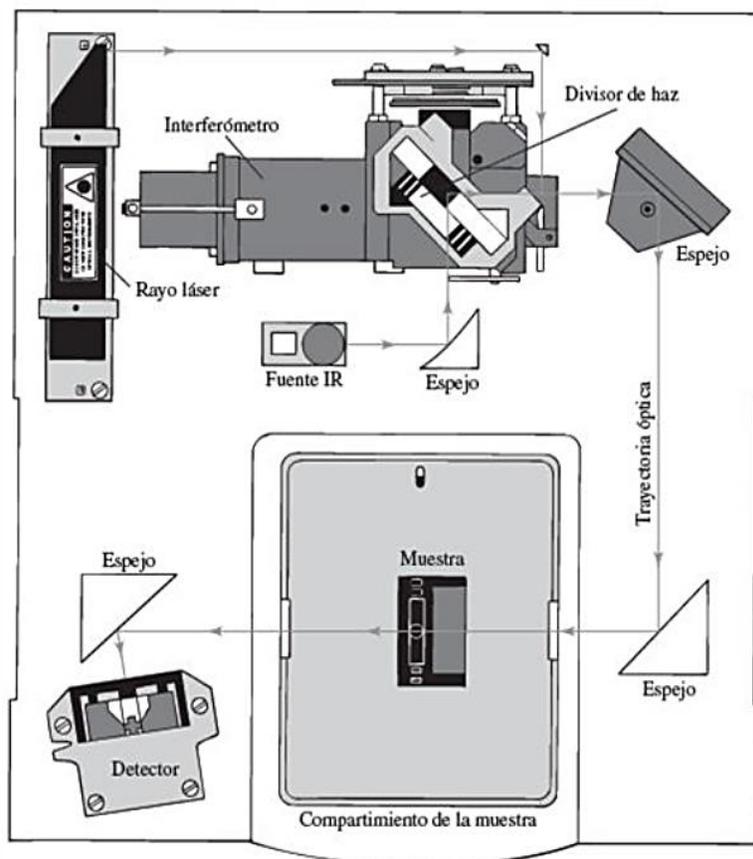
### **3.4.3 Instrumentación**

El instrumento empleado para determinar el espectro IR de un compuesto se denomina espectrómetro infrarrojo, siendo uno de uso común el de transformada de Fourier (FTIR) <sup>(89,90)</sup>.

El FT-IR utiliza un interferómetro para procesar la energía enviada a la muestra. En el interferómetro, la fuente de energía pasa a través de un divisor de haz, un espejo colocado en un ángulo de 45° con respecto a la radiación entrante, esta pasa, pero la separa en dos haces perpendiculares, uno no desviado y el otro orientado a un Ángulo de 90°, este último haz va a un espejo estacionario o “fijo” y regresa al divisor de rayos. El rayo no desviado es dirigido a un espejo en movimiento y regresa al divisor de rayo. Este movimiento del espejo hace que varíe la longitud de trayectoria que atraviesa el segundo haz. Cuando ambos haces se encuentran en el divisor de haz, se recombinan, pero las diferencias de longitud de trayectoria (diferentes longitudes de onda) de los dos haces provocan

interferencias tanto constructivas como destructivas. El haz combinado que contiene estos patrones de interferencia se llama interferograma<sup>(91)</sup>.

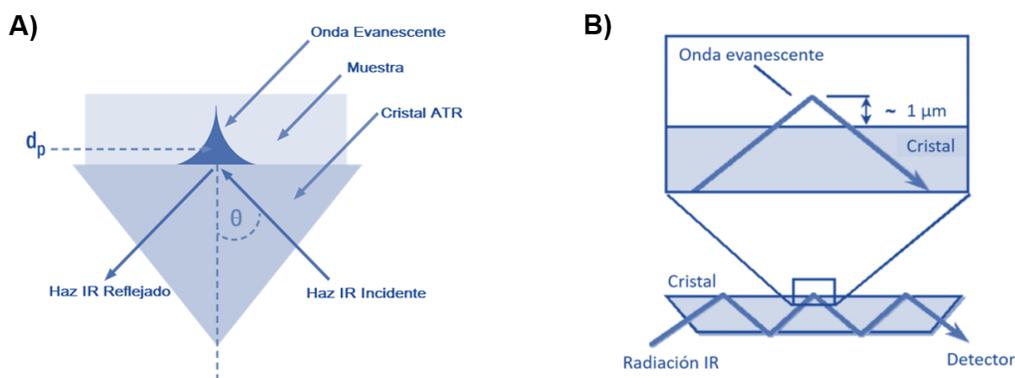
El interferograma generado al combinar los dos haces se orienta hacia la muestra mediante el divisor de haz. A medida que atraviesa la muestra, la muestra absorbe simultáneamente todas las longitudes de onda que normalmente se encuentran en su espectro infrarrojo. La señal de interferograma modificada que llega al detector contiene información sobre la cantidad de energía que fue absorbida en cada longitud de onda (frecuencia). El software compara el interferograma modificado con un rayo láser de referencia para tener un estándar de comparación. El interferograma final contiene toda la información en una señal en el dominio del tiempo<sup>(91)</sup>. El proceso matemático de transformada de Fourier es implementado por el software de la computadora para extraer las frecuencias individuales que fueron absorbidas y reconstruir y trazar lo que reconocemos como un espectro infrarrojo típico<sup>(92)</sup>.



**Figura 9.** Esquema básico de un espectrómetro FTIR<sup>(91)</sup>.

### 3.4.4 Sistema de Reflectancia total atenuada (Attenuated Total Reflectance, ATR)

El principio de esta técnica se basa en el fenómeno de la reflexión total interna y la transmisión de la luz a través de un cristal con un elevado índice de refracción (Figura 10). La radiación logra penetrar entre 0.5 y 5 micrómetros en la muestra, lo que implica la aplicación de presión suficiente para lograr el contacto necesario entre la muestra y el cristal, de esta forma se produce la reflexión total en forma de onda evanescente <sup>(93)</sup>.



**Figura 10.** Interacción del haz infrarrojo y la muestra en una celda ATR, **A)** de rebote único, **B)** de rebote múltiple <sup>(94)</sup>.

Si en el lado exterior del cristal se coloca un material o muestra absorbente, la luz que viaja a través del cristal se verá atenuada permitiendo el registro del espectro de la muestra. El ángulo de la luz incidente y la geometría (espesor y ángulo) del cristal facilitan la producción de sucesivas reflexiones internas a lo largo de este antes de alcanzar al detector lo que explica la sensibilidad adicional que ofrece esta técnica <sup>(95)</sup>.

Factores como la profundidad de penetración en inglés “deep penetration” ( $d_p$ ), depende de otros componentes como la longitud de onda de la radiación,  $\lambda$ , el índice de refracción del cristal ATR,  $n_p$ , el índice de refracción de la muestra,  $n_s$ , y del ángulo de incidencia del haz de radiación del cristal,  $\theta$ , los cuales se relacionan de acuerdo con la ecuación 3:

$$d_p = \frac{\lambda}{2\pi n_p (\sin^2 \theta - n_{sp}^2)^{1/2}} \quad \text{Ecuación 3}$$

Donde,  $n_{sp} = n_s/n_p$  ( $n_s < n_p$ ). El camino óptico total en la muestra se obtiene multiplicando  $d_p$  por el número de reflexiones que se hayan producido a través de

la misma. Esta técnica es muy efectiva especialmente en las regiones del infrarrojo medio e infrarrojo cercano para el análisis de sólidos y líquidos debido a la facilidad que presentan estos estados para compactarse contra el cristal al agregar presión<sup>(96,97)</sup>.

### 3.3 Interpretación espectral infrarroja

Se pueden emplear diversas formas para la interpretación de los espectros obtenidos, los cuales conllevan la presencia de varios picos según la muestra. Para esto es necesario seguir procesos organizados que aumenten las posibilidades de obtener información precisa del espectro. Este es el proceso de 5 pasos que puede seguir al interpretar espectros<sup>(98,99)</sup>.

**1.) Utilización de datos de calidad.** En lo posible, consiste en la interpretación de espectros con una buena relación señal/ruido, picos bien resueltos y sin artefactos, lo anterior brindará mayor precisión al análisis. Se pueden probar varios métodos experimentales y de software para mejorar la apariencia de un espectro.

**2.) Manejo de mezclas.** La complejidad de la composición de la muestra hace más complejo el espectro y su respectiva interpretación. Todo lo que sea posible realizar para simplificar la composición de esta, facilitará el trabajo.

**3.) Utilización de conocimientos previos sobre la muestra.** Es necesario la revisión de referencias relacionadas con la muestra en estudio, lo cual equivale a la búsqueda de posibles análisis químicos anteriores, origen, resultados, propiedades físico-químicas, propósito del estudio. Esto contribuye al estado de interpretación y la delimitación de la información.

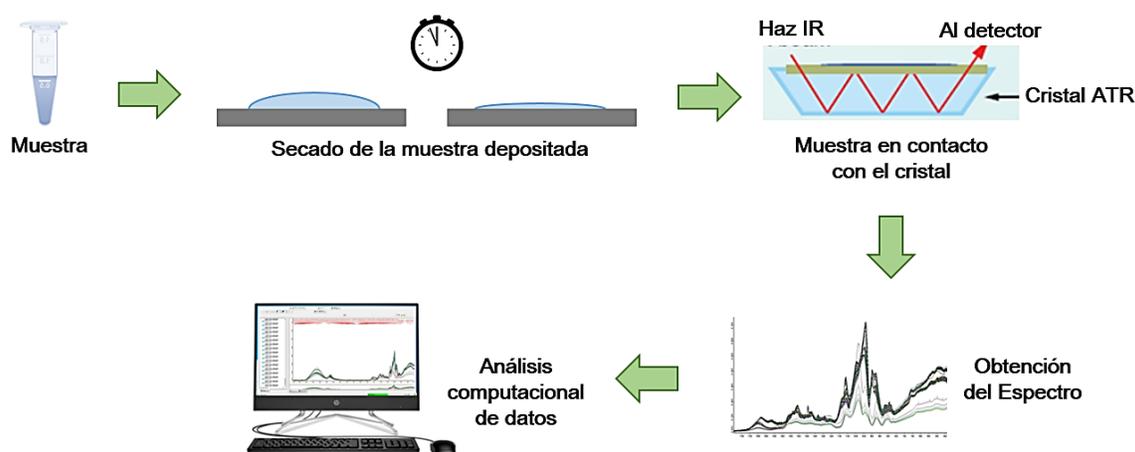
**4.) Apariencia del espectro.** La determinación de la resolución, el método de muestreo utilizado y si fueron realizadas manipulaciones espectrales (sustracción, suavizado, corrección de la línea de base, etc.) en el espectro, son importantes conocerlas con anticipación debido a que afectan la apariencia espectral.

**5.) Lectura y asignación de bandas.** El análisis espectral requiere la observación del espectro para determinar la presencia o ausencia de picos a una longitud de onda. Posteriormente se compara el resultado, según el origen de las muestras, con distintas fuentes de información, búsqueda de bibliotecas o software de interpretación. De esta forma es posible afirmar o negar la presencia de una gran

cantidad de grupos funcionales importantes, considerar la intensidad de bandas, entre otras características<sup>(100,101)</sup>.

### 3.3.1 Quimiometría y Modelación

El creciente avance de la tecnología, la búsqueda y obtención de resultados por medio de diferentes técnicas, han contribuido a la enorme producción de datos en diversas disciplinas del conocimiento. Los datos analíticos recopilados por FTIR de una determinada muestra en corto tiempo se convierten en una labor compleja en el momento del análisis, interpretación y uso de esos datos para lograr un objetivo en específico<sup>(102)</sup> (Figura 11).



**Figura 11.** Evaluación gradual del procesamiento de muestras.

Los instrumentos de espectroscopía vibracional como FTIR se han acoplado con diversos algoritmos quimiométricos para realizar tareas analíticas. El uso de la quimiometría permite el tratamiento de los datos proporcionando información química a partir de los datos espectrales y reduce la necesidad de calibraciones externas u otros procedimientos específicos para la determinación de cada una de las propiedades o componentes de una muestra, una vez establecidos los modelos multivariados que se emplearán<sup>(103)</sup>.

Los métodos de tratamiento previo de los datos “Pre-procesamiento”, son utilizados en la eliminación o reducción de señales no deseadas, previo al desarrollo de análisis del modelo. La selección del método de pre-procesamiento es un paso importante debido a su posterior influencia en la precisión y capacidad de discriminación o interpretabilidad del modelo. Dentro del grupo de tratamientos

más empleados se encuentran: el promediado de espectros, la primera y segunda derivada, centrado en la media, la corrección de dispersión multiplicativa, entre otros.

El propósito de las técnicas multivariantes es el análisis simultáneo por medio de métodos estadísticos, de conjuntos de datos compuestos por distintas variables. Estas técnicas pueden ser supervisadas o no supervisadas, las primeras intentan relacionar las variables con alguna propiedad de la muestra, mientras que los no supervisados no se requiere información previa sobre la propiedad. Los métodos supervisados pueden dividirse en métodos de clasificación y métodos de predicción<sup>(91)</sup>.

El método de Mínimos Cuadrados Parciales (PLS, del inglés Partial Least Squares) es un método rápido, eficiente y óptimo que permite la comparación entre múltiples variables de respuesta y múltiples variables explicativas<sup>(102,104)</sup>.

Se utiliza para la construcción de modelos de predicción y es capaz de relacionar una característica de la muestra, con los espectros medidos.

Posterior al proceso de construcción del modelo, la evaluación del mismo se realiza mediante el uso de figuras de mérito utilizadas para su validación. Dentro de estas se encuentran el error cuadrático medio de la calibración (RMSEC), una medida para los modelos sobre el ajuste a los datos de calibración. El error cuadrático medio de la validación cruzada (RMSECV) que se obtiene cuando únicamente se tiene el conjunto de calibración y se realiza la validación cruzada aplicada a este conjunto. El error cuadrático medio de la predicción (RMSEP), cuando el modelo se aplica a un conjunto constituido por muestras diferentes de las incluidas en la calibración. El error relativo cuadrático medio (RRMSEP), se utiliza para comparar la capacidad predictiva de distintos modelos. El coeficiente de determinación de la predicción ( $R^2$ ) es indicador de la bondad del ajuste del modelo aplicado al conjunto de validación. Este coeficiente determina el porcentaje de la varianza de la variable dependiente que es explicado por el modelo de regresión<sup>(91)</sup>.

El Análisis Discriminante mediante mínimos cuadrados parciales (PLS-DA, del inglés Partial Least Squares-Discriminant Analysis) es un método de clasificación lineal que combina las propiedades del PLS con el análisis discriminante<sup>(60)</sup>. En el PLS-DA la matriz X se utiliza como predictor, y la matriz Y, que contiene variables ficticias, se utiliza como una respuesta. La matriz ficticia Y comprende variables categóricas (0,1) que indican la pertenencia a determinada clase. Se utiliza para calcular las respuestas predichas, que luego se emplean para la discriminación. En la implementación de PLS-DA tradicional, la regla de discriminación se basa en

la comparación de los valores de respuesta previstos con un umbral fijo (por ejemplo, 0.5)<sup>(105)</sup>. El principal objetivo de PLS-DA es construir un modelo de discriminación adecuado para ser aplicado en futuras clasificaciones.

Se puede utilizar una serie de parámetros para evaluar el rendimiento de los modelos de clasificación<sup>(102)</sup>. La tasa de verdaderos positivos (VP) (TP, del inglés True Positive) es la probabilidad de que una muestra positiva pueda clasificarse como positiva, la tasa de verdaderos negativos (VN) (TN, del inglés True Negative), implica que una muestra negativa puede clasificarse como negativa. Este criterio también se maneja de forma similar para los falsos positivos (FP, del inglés False Positive) y falsos negativos (FN, del inglés False Negative). La sensibilidad es la capacidad del modelo evaluado para clasificar correctamente las muestras auténticas dentro de su grupo usando las tasas VP y FN. La especificidad se utiliza para estimar la capacidad del modelo para identificar correctamente las muestras en relación con los valores de VN y FP. La precisión es la tasa de clasificación correcta, independientemente de la clase de muestra. El puntaje F1 (del inglés, F1-score), es una métrica que resume precisión y sensibilidad, y es muy empleada por su utilidad en el análisis cuando la distribución entre las clases es desigual. La eficiencia permite establecer el desempeño del modelo, a través del promedio aritmético de los valores de sensibilidad y especificidad, donde un valor de 1 corresponde a una eficiencia del 100%<sup>(106)</sup>.

## 4. Diseño Metodológico

### 4.1 Tipo de Estudio

El estudio realizado es de tipo experimental.

Este trabajo investigativo fue realizado en la Unidad de Genética y Biología Molecular (UGBM) de la Universidad Simón Bolívar, Barranquilla-Atlántico, dentro del proyecto “Biosensor de múltiple detección y discriminación de pequeñas secuencias de nucleótidos por FTIR y electroquímica de impedancia”. Siendo este aprobado por el comité de ética institucional.

### 4.2 Demostración de la viabilidad de la técnica ATR-FTIR en la diferenciación de pequeñas secuencias de nucleótidos.

#### 4.2.1 Preparación y adquisición de espectros de ATR-FTIR para las diferentes secuencias previamente diseñadas.

Se empleó un conjunto de 35 secuencias de nucleótidos descritas en la tabla 1. Estas fueron sintetizadas a través de la compañía comercial Macrogen, Inc (Corea del sur) según especificaciones que incluyen: longitud de 4, 8, 12,15 y 16 nucleótidos, secuencias con repeticiones de bases y variación en las combinaciones posibles entre las mismas (N° 1 a la 18). El grupo restante (N°19- N°35) corresponden a secuencias de primers. Del total de secuencias se adquirieron los espectros mediante la técnica de ATR-FTIR para el establecimiento de modelos que permitieran la diferenciación de pequeñas secuencias de nucleótidos.

**Tabla 1.** Secuencias nucleotídicas previamente diseñados empleados para análisis de viabilidad de la técnica ATR-FTIR para la cuantificación del porcentaje de nucleótidos. %N: número de bases, porcentajes de Adenina (A%), Timina (%), Guanina (%G) y Citosina (%C). Los datos en rojo representan las secuencias usadas para test de validación externa.

N°	Código	Secuencia	NT	%A	%C	%T	%G
1	PolyA-16	AAAAAAAAAAAAAAAA	16	100,0	0,0	0,0	0,0
2	polyT-16	TTTTTTTTTTTTTTTT	16	0,0	100,0	0,0	0,0
3	polyC-16	CCCCCCCCCCCCCCCC	16	0,0	0,0	100,0	0,0
4	polyG-16	GGGGGGGGGGGGGGGG	16	0,0	0,0	0,0	100,0
5	polyAT-16	ATATATATATATATAT	16	50,0	0,0	0,0	50,0
6	polyAC-16	ACACACACACACACAC	16	50,0	0,0	50,0	0,0

7	polyAG-16	AGAGAGAGAGAGAGAG	16	50,0	50,0	0,0	<b>0,0</b>
8	polyTC-16	TCTCTCTCTCTCTCTC	16	0,0	50,0	50,0	0,0
9	polyTG-16	TGTGTGTGTGTGTGTG	16	0,0	50,0	0,0	50,0
10	polyCG-16	CGCGCGCGCGCGCGCG	16	0,0	0,0	50,0	50,0
11	polyATC-15	ATCATCATCATCATC	15	33,3	33,3	33,3	0,0
12	polyATG-15	ATGATGATGATGATG	15	33,3	33,3	0,0	33,3
13	polyACG-15	ACGACGACGACGACG	15	0,0	33,3	<b>33,3</b>	33,3
14	polyTCG-15	TCGTCGTCGTCGTCG	15	33,3	<b>0,0</b>	33,3	33,3
15	polyATCG-16	ATCGATCGATCGATCG	16	25,0	25,0	25,0	25,0
16	polyATCG-12	ATCGATCGATCG	2	25,0	25,0	25,0	25,0
17	polyATCG-8	ATCGATCG	8	25,0	25,0	25,0	25,0
18	polyATCG-4	ATCG	4	25,0	25,0	25,0	25,0
19	L1F16	CACTATTTTGGAGGACTGGAAT	22	27,3	<b>13,6</b>	31,8	27,3
20	L1F18	GCCCCTGCCTCTACACAGTA	20	20,0	<b>45,0</b>	20,0	15,0
21	L1F31	CAACGTGCTCAGGGACAC	18	<b>27,8</b>	33,3	<b>11,1</b>	27,8
22	L1F35	GTAGGTCGTGGTCAGCCATT	20	15,0	20,0	30,0	35,0
23	L1F51	TCCAATACCTAAAACCTCAACG	22	40,9	36,4	18,2	4,5
24	L1F66	CGCCGTAAACGTATTCCTA	20	25,0	35,0	25,0	15,0
25	L1R16	GATGAGGTGGTGGGTGTAGC	20	15,0	5,0	25,0	<b>55,0</b>
26	L1R18	ATCCTGCTTATTGCCACCAC	20	20,0	40,0	30,0	<b>10,0</b>
27	L1R31	GGGAGGTGTGGTCAATCC	18	<b>16,7</b>	16,7	22,2	44,4
28	L1R35	TGGTTAGCATTACAAGGTGTGC	22	22,7	<b>13,6</b>	31,8	<b>31,8</b>
29	L1R51	CACAACCCACACCAACCTA	20	40,0	55,0	5,0	0,0
30	L1R66	CCAACAGCAAGCAACCTAGA	20	45,0	35,0	5,0	15,0
31	p3	AATTGCAGGGCCCCTAGAAAAAAGGGCTGT	30	33,3	20,0	16,7	30,0
32	p5	TTCTATGGAGACTCCCTGACCCAAATGCCA	30	26,7	33,3	23,3	16,7
33	p6	TCCCCTAGTGGGATGTGTA	21	14,3	23,8	33,3	28,6
34	p7	AGAGCAGACCAGAGCCAACAG	21	42,9	28,6	0,0	28,6
35	p19	GAATCTCTCTGTTTTCTGCCAGTTC	25	12,0	<b>28,0</b>	<b>44,0</b>	16,0

Durante el proceso de preparación de las muestras se agregó agua libre de nucleasas para lograr la disolución de los oligonucleótidos y polinucleótidos diseñados.

La adquisición de los espectros por espectroscopía FTIR de las diferentes secuencias nucleotídicas fue realizada en el intervalo espectral entre 4000–400  $\text{cm}^{-1}$  utilizando un espectrómetro ALPHA FTIR spectrometer ATR, equipado con un cristal de diamante (Bruker Optics, Billerica, MA, USA), a una resolución espectral de 4  $\text{cm}^{-1}$ .

Para la caracterización espectroscópica de cada una de estas secuencias previamente diseñadas se depositó 1µL de cada muestra sobre la cara del cristal de diamante de la unidad ATR con la superficie tratada hacia arriba y la punta de la pinza micrométrica comprimida sobre la superficie para permitir el contacto adecuado y obtener los espectros característicos. Se esperó durante aproximadamente 15 minutos el secado a temperatura ambiente del volumen empleado para eliminar el exceso de agua. Posteriormente se midieron los espectros hasta que las bandas de absorción relacionadas con el agua fueran indetectables.

Se realizaron en total 64 scans para cada análisis. Se adquirieron entre 10 a 30 espectros para cada secuencia nucleotídica, dependiendo del tiempo de evaporación del solvente.

#### **4.2.2 Generación y validación del modelo estadístico para la cuantificación de porcentaje de nucleótidos mediante los espectros ATR-FTIR para las pequeñas secuencias**

Se utilizó el software Quant2™ de OPUS™ versión 4.2 (Bruker Optics) para la adquisición, procesamiento de datos y el análisis de espectros.

Los espectros ATR-FTIR de cada muestra fueron promediados y normalizados por valores mínimos y máximos para la reducción del impacto de las diferencias de intensidad.

Se realizó la calibración y validación de modelos espectrales para la cuantificación de porcentajes de nucleótidos (%N: %A, %G, %T, %C), mediante el algoritmo de regresión multivariable conocido como mínimos cuadrados parciales (PLS) en grupos de muestras independientes y se validó el modelo mediante una validación interna y externa. La validación interna consistió en una validación cruzada o cross validation-leave-one-out (CV, en inglés).

La validación externa consistió en la predicción de dos tipos de nuevas secuencias, para el primero se seleccionó el subconjunto de datos espectrales obtenidos de algunas de las secuencias nucleotídicas sintetizadas, no empleadas en el conjunto de calibración con el mismo % de nucleótidos, para así determinar los valores de predicción de porcentajes de nucleótidos a partir de nuevos espectros de entrada al modelo, y la segunda consistió en nuevas secuencias con

valores de % de nucleótidos diferentes, las cuales se presentan resaltadas en rojo en la tabla 1.

En toda la región espectral comprendida entre 4000-400  $\text{cm}^{-1}$  se probaron los siguientes pasos de pre-procesamiento de datos: Pre-procesamiento de Datos No Espectrales, Eliminación de Desplazamiento Constante, Resta de Línea Recta, Normalización de Vector (VN), Normalización de Mínimos y Máximos, Corrección Multiplicativa de la Dispersión (MSC), Primera Derivada (1D), Segunda Derivada, Primera derivada + Resta de Línea Recta, Primera derivada + Normalización del Vector y Primera derivada + MSC.

Mediante el proceso de optimización, el software verificó automáticamente los rangos espectrales comunes en combinación con los métodos de pre-procesamiento de datos mencionados anteriormente, definiendo de esta forma las regiones espectrales más relevante para el análisis. La opción General A empleada, buscó la combinación óptima dividiendo el rango de frecuencia seleccionado en 10 áreas igualmente grandes y calculando sucesivamente combinaciones de estas áreas.

En la construcción de los modelos de calibración se registraron los valores obtenidos de los siguientes parámetros: variables latentes (LV, Latent Variables), coeficiente de determinación ( $R^2$ ), error cuadrático medio de estimación (RMSEE, Root Mean Square Error of Estimation), error cuadrático medio de predicción (RMSEP, Root Mean Square Error of Prediction), error cuadrático medio de validación cruzada (RMSECV, Root Mean Square Error of Cross Validation).

## **4.3 Implementación de la técnica ATR-FTIR en la identificación de genotipos de virus de papiloma humano**

### **4.3.1 Selección de las muestras clínicas**

En el presente estudio se emplearon muestras de ADN de mujeres con edades entre 17 y 26 años y vida sexual activa de la población estudiantil, que asistieron a consultas de control ginecológico durante una jornada citológica realizada en la Universidad Simón Bolívar en Barranquilla. A cada una de las participantes se le presentó un consentimiento informado (Anexo 1) para la aprobación o no del uso de las muestras cervicales obtenidas para el desarrollo de posteriores estudios.

El procedimiento consistió en exploración ginecológica y toma de raspados endocervicales para estudio citológico y detección molecular de ADN de VPH. La extracción del ADN genómico se desarrolló empleando el DNeasy Blood & Tissue Kit (Spin-Column protocol, QUIAGEN), siguiendo las indicaciones del fabricante en el protocolo de purificación de ADN total y almacenando las muestras a -20°C hasta su uso <sup>(107)</sup>. Se determinó la concentración de las muestras empleando un espectrofotómetro de absorción ultravioleta (NanoDrop™ One, Thermo Scientific™), que mide el valor de absorbancia del ADN a una longitud de onda de 260nm. La relación a 260/280 permitió evaluar la pureza de las muestras en proporción con la contaminación de proteínas, considerándose valores recomendados a los mayores de 1,6 <sup>(69)</sup>.

### **4.3.2 Caracterización de las muestras por la técnica estándar de real-time PCR y electroforesis**

Se desarrolló un análisis de curva melting a partir de señales fluorescentes emitidas por SYBR Green mediante ciclos de calentamiento entre 60 °C a 94 °C con incrementos de 0,3 °C /s en ensayos real-time PCR, para la detección y genotipificación de VPH 16,18, 31, 35, 51 y 66 evaluados en un total de 19 muestras.

Se emplearon secuencias de primers específicos dirigidos a la región L1 del genoma viral, descritos en el estudio realizado por Tsakogiannis *et al.*, 2014 <sup>(86)</sup> para la detección de los 6 genotipos de VPH mencionados anteriormente (Tabla 2). Se organizaron 3 mezclas de primers, generando un volumen total para cada

una de 2.8 µL. La mezcla I consistía en 0.7 µL de primer forward y 0.7 µL de primer reverse de los genotipos de VPH 16 y VPH 35, la mezcla II conformada por los primers para VPH 18 y 31 y la mezcla III por los primers para VPH 51 Y 66.

**Tabla 2.** Secuencias de primers específicos correspondientes a los genotipos de VPH 16,35,18,31,51 y 66 <sup>(86)</sup>

Mezcla de primers	Primers	Secuencia (5' - 3')
I	HPV-16 L1F	ACTATTTTGGAGGACTGGAA
	HPV-16 L1R	ATGAGGTGGTGGGTGTAGC
	HPV-35 L1F	AGGTCGTGGTCAGCCATT
	HPV-35 L1R	GGTTAGCATTACAAGGTGTGC
II	HPV-18 L1F	CCCCTGCCTCTACACAGTA
	HPV-18 L1R	CCTGCTTATTGCCACCAC
	HPV-31 L1F	ACGTGCTCAGGGACAC
	HPV-31 L1R	GGAGGTGTGGTCAATCC
III	HPV-51L1 F	CCAATACCTAAAACCTCAAC
	HPV-51L1 R	ACAACCCACACCAACCTA
	HPV-66L1 F	GCCGTAACGTATTCCCTA
	HPV-66L1 R	AACAGCAAGCAACCTAGA

Para el proceso de amplificación se empleó un termociclador Bio-Rad CFX96TM. Cada ensayo PCR se realizó con un volumen final de 20 µL en el tubo de reacción, el cual contenía PowerUp™ SYBR® Green Master Mix (Thermo Fisher Scientific), 2,8 µL de mezcla del conjunto de primers forward y reverse según el tipo de VPH evaluado, 10-100 ng/µL de plantilla de ADN conforme al rango de concentración recomendado por el fabricante para uso de la master mix (Applied Biosystems™, Thermo Fisher Scientific) y agua libre de nucleasas (UltraPure™ DNase/RNase-Free, Invitrogen™). En cada ensayo de amplificación se emplearon controles positivos para los 6 genotipos de VPH y controles negativos que contenían solo mix PCR (todos los reactivos para el proceso de amplificación sin la adición de ADN).

Las condiciones del ciclo de amplificación corresponden a 94 °C por 4 min, seguido por 40 ciclos de denaturalización a 94 °C por 1 min, la fase de anillamiento a 50 °C por 1 min; y de extensión a 72 °C por 1 min, seguida por una adicional a 72 °C durante 10 min.

El tamaño de los amplicones se verificó por electroforesis en gel de agarosa al 2% en buffer TAE 1X (Tris-acetate-EDTA Buffer, Promega), visualizando los productos PCR mediante tinción con SYBR™ Safe DNA Gel Stain (Invitrogen™, Thermo Fisher Scientific). Se depositaron 4 µL de cada muestra amplificada de ADN, mezclada con 2 µL de buffer de carga (Gel Loading Dye, New England Biolabs). Así mismo, el tamaño o peso molecular se determinó empleando 100 bp DNA Ladder (New England Biolabs) como marcador en el proceso electroforético.

Las muestras se corrieron a 70V durante 105 minutos, revelando posteriormente el resultado de los geles en el Sistema de documentación en gel por quimioluminiscencia de BIO-RAD a través del software Molecular Imager® Gel Doc TM XR+ Imaging System.

#### **4.3.3 Preparación y adquisición de espectros de ATR-FTIR para los diferentes controles positivos, negativos y muestras amplificadas por real-time PCR.**

Utilizando el espectrómetro ALPHA FTIR spectrometer ATR, equipado con un cristal de diamante (Bruker Optics, Billerica, MA, USA) en el intervalo espectral entre 4000–400  $\text{cm}^{-1}$  se realizaron en total 64 scans en cada análisis y se adquirieron entre 10 a 30 espectros a partir de 1µl de la solución PowerUp™ SYBR® Green Master Mix (Thermo Fisher Scientific), seis del total de muestras amplificadas (M1-M6), los controles positivos amplificados de los genotipos de VPH evaluados en este estudio (16, 18, 31, 35, 51 y 66) y los controles negativos.

A partir de los resultados obtenidos, se realizó un promedio de la data espectral para la solución PowerUp™ SYBR® Green Master Mix (Thermo Fisher Scientific) y los controles positivos amplificados en el software OPUS™ versión 4.2 (Bruker Optics). Con los promedios resultantes se realizó una comparación de los espectros a los cuales se aplicó el método de normalización del vector, y se determinaron las señales relacionadas entre los controles positivos y la solución Master mix.

Para revelar las diferencias en las intensidades de las señales entre los compuestos mencionados, se realizó una sustracción del espectro de la Master mix a todos los controles positivos. Los conjuntos de espectros posteriores al proceso de sustracción se agruparon por archivo de cada muestra analizada empleando la función de “Assemble GC file” del programa Opus.

#### 4.3.4 Generación y validación de modelos estadísticos para discriminación mediante los espectros para los diferentes genotipos de VPH.

Para el análisis estadístico multivariado se utilizaron diversos pre-tratamientos sobre los datos espectrales adquiridos de la solución Master mix, seis muestras amplificadas (M1-M6), los controles positivos amplificadas de los genotipos de VPH evaluados (16, 18, 31, 35, 51 y 66) y los controles negativos, ya sea desde 1ra y 2da derivada, correcciones de línea base, normalización, entre otros, empleando el software OPUS. A partir de los datos verdaderos y predichos generados en el proceso de validación en grupos de muestras independientes y validación cruzada se desarrollaron matrices de confusión, las cuales permiten estimar el rendimiento de cada modelo de predicción para infección o no por VPH 16, 18, 31, 35, 51, 66, teniendo en cuenta los parámetros de precisión, sensibilidad, F1-Score y exactitud calculados de la siguiente manera:

$$Precisión = \frac{VP}{VP+FP} \text{ Ecuación 4}$$

$$Sensibilidad = \frac{VP}{VP+FN} \text{ Ecuación 5}$$

$$F1 - Score = \frac{2*precisión*sensibilidad}{precisión+sensibilidad} \text{ Ecuación 6}$$

$$Exactitud = \frac{VP+VN}{VP+FN+VN+FP} \text{ Ecuación 7}$$

$$Especificidad = \frac{VN}{VN+FP} \text{ Ecuación 8}$$

Donde VP significa verdadero positivo, VN verdadero negativo, FP para falso positivo y FN para falso negativo.

Se estableció un valor umbral de 0.5 para indicar aquellos valores considerados positivos en la detección de infección por VPH ( $>0.5$ ) y los que son negativos ( $<0.5$ ).

Se realizaron gráficos de dispersión basados en los resultados del Análisis discriminante por mínimos cuadrados parciales (PLS-DA) representados para los modelos de predicción de los genotipos de VPH 16, 18, 31, 35, 51 y 66 y las muestras negativas a la infección. El modelo fue evaluado por una validación interna CV y dos validaciones externas, la primera consistió en extraer 25% de los datos para la validación y la segunda consistió en la predicción de seis muestras de ADN proveniente de los raspados endocervicales con amplificación positiva para VPH 51, y controles positivos y negativos que se contrastaron con los resultados encontrados por el método convencional de real-time PCR.

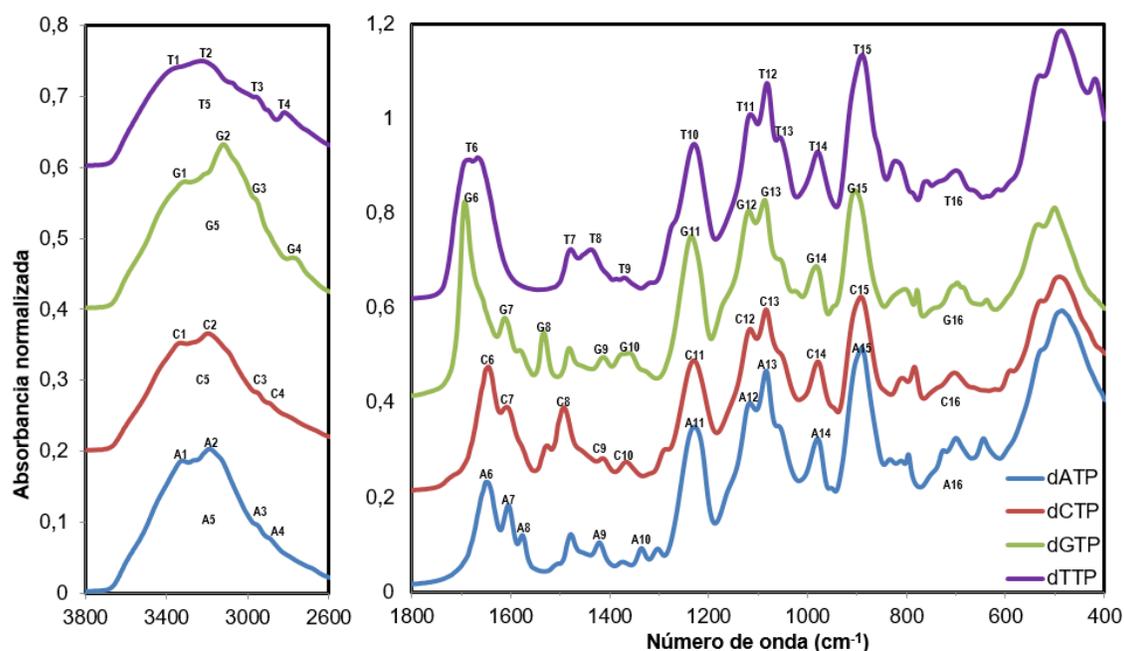
## 5. Resultados y Discusión

### 5.1 Viabilidad de la técnica ATR-FTIR en la diferenciación de pequeñas secuencias de ADN

#### 5.1.1 Asignación de bandas de ATR-FTIR para los diferentes desoxinucleótidos y secuencias de nucleótidos previamente diseñados.

Los espectros infrarrojos tomados a partir de las soluciones preparadas de dATP, dTTP, dCTP y dGTP exhibieron modos vibracionales característicos en la región entre  $3800$  y  $400\text{ cm}^{-1}$  al ser irradiadas mediante luz infrarroja.

En la Figura 12 se muestran las bandas espectrales resultantes de los espectros capturados para cada muestra de dNTPs posterior al proceso de promediado y de normalización basado en un tratamiento de mínimos y máximos.



**Figura 12.** Espectros ATR-FTIR promediados y normalizados de desoxiadenosina trifosfato (dATP), desoxicitidina trifosfato (dCTP) y desoxiguanosina trifosfato (dGTP) desoxitimidina trifosfato (dTTP) en el rango  $3800$  a  $400\text{ cm}^{-1}$ .

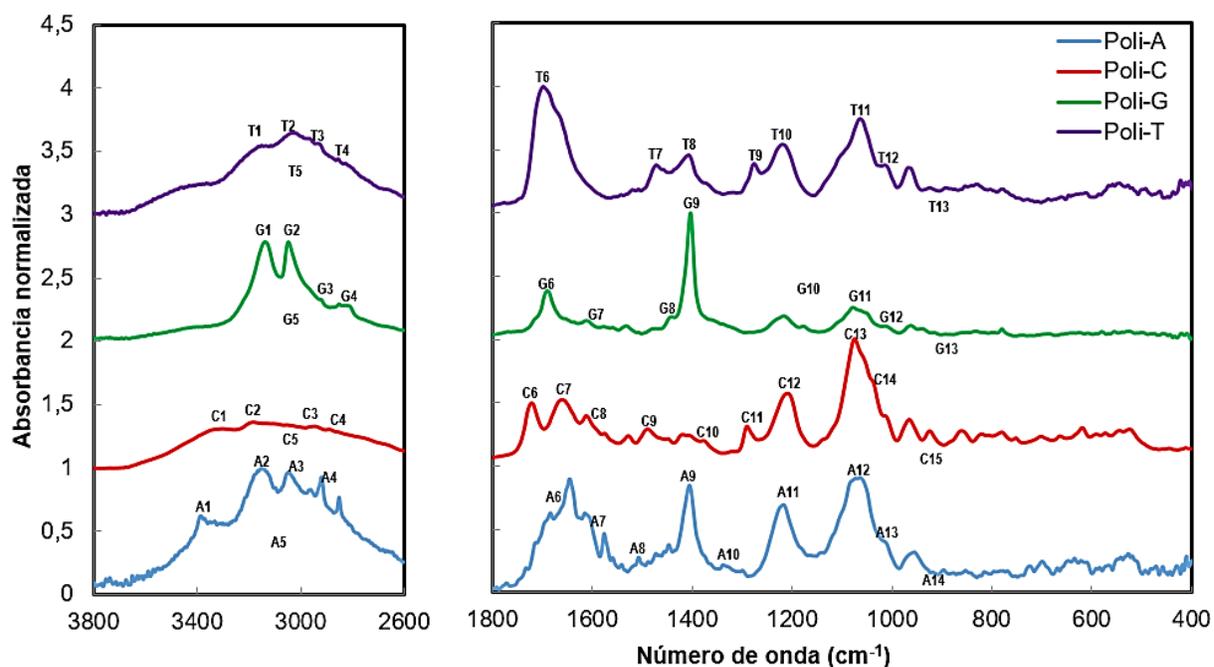
Se estableció la relación entre las bandas espectrales y los modos vibratorios con base en los constituyentes de las muestras analizadas (Tablas 3 y 4).

**Tabla 3.** Asignación de modos vibracionales a las bandas observadas en los espectros ATR-FTIR de los desoxinucleótidos dATP, dTTP, dCTP y dGTP en el rango de 3400 a 400  $\text{cm}^{-1}$ . **v:** vibración tipo estiramiento,  **$\delta$ :** vibración tipo flexión, **df:** deformación, **sim:** modo simétrico, **asm:** modo asimétrico, **Out-of-plane:** fuera del plano, **Gu:** guanina, **Ty:** timina, **Cy:** citosina., **D-ribosa:** azúcar Desoxirribosa, **P:** grupo fosfato, **gly:** enlace glicosídico, **Vib(ring):** vibración del anillo.

Códigos	Bandas ( $\text{cm}^{-1}$ )				Asignaciones	Referencias
	dATP	dCTP	dGTP	dTTP		
A1/C1/T1/G1	3321	3330	3306	3340	$v_{asm}(\text{NH})$	(108-111)
A2/C2/T2/G2	3190	3201	3123	3229	$v_{sim}(\text{NH})$	(108,109,111,112)
A3/C3/T3/G3	2951	2959	2957	2959	$v_{asm}(\text{C-H})$	(109,112)
A4/C4/T4/G4	2893	2893	2782	2818	$v_{sim}(\text{C-H})$	(109,112)
A5/C5/T5/G5		3700-2600			$v_{sim}(\text{O-H}) + v_{asm}(\text{O-H})$	(108,111,112)
G6/T6	-	-	1693	1683	$v(\text{C}=\text{C}), v(\text{C}=\text{O}), \text{Gu}[\delta(\text{N-H}_2)]$	(109,112,113)
A6/C6	1649	1645	-	-	$\delta(\text{N-H}_2) + \text{Cy}[v(\text{C}=\text{O})]$	(108,110,111,113)
A7/C7/G7	1605	1608	1612-1582	-	$v(\text{C}=\text{N}), v(\text{C}=\text{C})$	(108-110)
A8/C8/G8/T7	1577-1478	1493	1533-1482	1479	$\text{glyco}[\delta(\text{C-N-H})], \delta(\text{C-H}) + \text{Vib}(\text{ring})$	(108,113)
T8	-	-	-	1437	$\delta_{asm}(\text{CH}_3)$	(112)
A9/G9	1421	-	1413	-	$\delta(\text{N}=\text{C-H})$	(108,109)
C9	-	1414	-	-	$v(\text{C-N}), df(\text{N-H}), df(\text{C-H})$	(111)
A10/C10/G10/T9	1336-1302	1367-1292	1360	1371	$\delta(\text{C-N-H}), \delta(\text{C-H}), v(\text{C-N}), \text{Ty}[\delta_{sim}(\text{CH}_3)], \text{Cy}[\delta(\text{C}=\text{C-H})]$ D-ribosa [ $v(\text{C-O})$ ]	(108-112)
A11/C11/G11/T10	1229	1230	1235	1229	$v_{asm}(\text{PO}_2^-) + \text{D-ribosa}[v(\text{C-N})]$	(108-112,114,115)
A12/C12/G12/T11	1117	1116	1119	1116	$\text{P+D-ribosa}[v_{sim}(\text{P-O-C})]$	(109,111,112,114)
A13/C13/G13/T12	1084	1084	1087	1082	$v_{sim}(\text{PO}_2^-)$	(108,111,112,114,115)
T13	-	-	-	1056	$\text{D-ribosa}[v(\text{C-O})] + \delta(\text{N-C-H})$	(111,112)
A14/C14/G14/T14	980	979	983	980	$v_{sim}(\text{PO}_3^{2-}) + \delta(\text{C-N-C})$	(110,112,114,115)
A15/C15/G15/T15	891	892	905	890	$\delta(\text{C-N-C}), \text{D-ribosa}[\text{C-C}, \text{C-O}], \delta(\text{N-H}) \text{ out-of-plane}$	(108,111,112)
A16/C16/G16/T16		800-400			$\text{D-ribosa}[\delta(\text{O-H}) \text{ out-of-plane}], df(\text{Ring CH}) + v(\text{C-C}), \text{BaseVib}[\delta]$	(111,112)

Así mismo, se adquirieron los espectros ATR-FTIR para las secuencias de oligonucleótidos con 100% de cada base nucleotídica (A, T, G, C) indicados con anterioridad en la tabla 1 (N°1-N°4). Se presentan las bandas características

según las absorciones de cada constituyente de las muestras, para lo cual se promediaron y normalizaron los espectros resultantes (Figura 13). Es importante mencionar que la presencia de disolventes como el agua ( $H_2O$ ) en el espectro provoca la aparición de bandas características debido a que esta molécula tienen una intensa absorbancia centrada en  $1640\text{ cm}^{-1}$  aproximadamente (resultado de vibraciones de flexión) y  $3330\text{ cm}^{-1}$  (vibraciones de estiramiento de O-H) <sup>(116)</sup>, causando interferencia en la totalidad del espectro <sup>(117)</sup>. Ante esto el software empleado realiza una sustracción del background en la muestra de estudio, y a su vez, el uso del método de reflectancia total atenuada (ATR) mejora la interacción de la luz IR y permite con la muestra permitiendo observar con mayor exactitud la intensidad relativa de las absorciones características de los componentes de la muestra <sup>(118,119)</sup>.



**Figura 13.** Espectros ATR-FTIR promediados y normalizados de polinucleótidos de adenina (A), citosina (C), guanina (G) y timina (T) en el rango  $3800$  a  $400\text{ cm}^{-1}$ .

Con base a referencias y estudios reportados se plantearon las asignaciones correspondientes a las señales espectrales detalladas en las tablas 3 y 4, de esta forma fue posible el acercamiento, identificación y verificación de los datos

adquiridos a partir de los espectros recopilados y citados por diversos autores (120–122).

**Tabla 4.** Asignación de modos vibracionales a las bandas observadas en los espectros ATR-FTIR en el rango de 3800 a 700  $\text{cm}^{-1}$  de las secuencias de polinucleótidos empleadas. **v:** vibración tipo estiramiento,  **$\delta$ :** vibración tipo flexión, **df:** deformación, **sim:** modo simétrico, **asm:** modo antisimétrico, **D-ribosa:** azúcar Desoxirribosa, **P:** grupo fosfato, **gly:** enlace glicosídico, **Vib(ring):** vibración del anillo, **Gu:** guanina, **Ty:** timina.

Códigos	Bandas ( $\text{cm}^{-1}$ )				Asignaciones	Referencias
	dATP	dCTP	dGTP	dTTP		
A1/C1/T1/G1	3384	3295	3136	3150	$v_{\text{asm}}(\text{NH})$	(108–111)
A2/C2/T2/G2	3149	3184	3046	3025	$v_{\text{sim}}(\text{NH})$	(108–112)
A3/C3/T3/G3	3044	2944	2920	2964	$v_{\text{asm}}(\text{C-H})$	(108–110,112,123)
A4/C4/T4/G4	2959	2890	2850	2854	$v_{\text{sim}}(\text{C-H})$	(108–110,112,123)
A5/C5/T5/G5		3700-2600			$v_{\text{sim}}(\text{O-H}) + v_{\text{asm}}(\text{O-H})$	(108,111,112,124)
A6/C7	1684-1646	1722	-	-	$\delta(\text{N-H}_2)$	(108,112,124,125)
C6/T6/G6	-	1661	1690	1699	$v(\text{C=O}), \text{Gu}[\delta(\text{N-H}_2)]$	(108–110,112,124)
A7/C8/G7	1615-1577	1612-1528	1612-1533	-	$v(\text{C=N})$	(108–110,113,124)
A8/C9/G8/T7	1507	1490	1448	1472	$\text{glyco}[\delta(\text{C-N-H})], \text{Ty}[\delta(\text{N-H}) + \delta_{\text{sim}}(\text{CH}_3)]$	(108–110,112,113,124,126)
A9/G9	1406	-	1404	-	$\delta(\text{N=C-H})$	(108–110,112)
C10/T8	-	1378	-	1407	$\delta(\text{C=C-H}), \text{Ty}[\delta_{\text{asm}}(\text{CH}_3)]$	(110,112,124,127)
A10/C11//T9	1339	1291	-	1276-1220	$v(\text{C-N-H})$	(108,110,112,124,128)
A11/C12/G10/T10	1218	1209	1217	1220	$v_{\text{asm}}(\text{P=O}) \text{PO}_2^-$ Fosfodiester	(123,126,128)
A12/C13/G11/T11	1071	1075	1078	1064	$v_{\text{sim}}(\text{P=O}) \text{PO}_2^-$ fosfodiester + D-ribosa [ $v(\text{C-O}), \text{CH}_2$ ]	(123,126,129)
A13/C14/G12/T12	1016	1013	1013	1014	P + D-ribosa [ $v(\text{C-O})$ ]	(124,126,127,129)
A14/C15/G13/T13		970 -700			$\delta(\text{O-P-O}) + \text{D-ribosa} [v(\text{C-O}, \text{C-C}) + \text{Vib}(\text{ring})]$	(123,124,126,127,129)

Se analizaron las señales obtenidas de los espectros ATR-FTIR en la región MIR entre 3800 y 400  $\text{cm}^{-1}$ . Las vibraciones fundamentales comprendidas entre 4000-2500  $\text{cm}^{-1}$  pueden atribuirse al estiramiento de algunos átomos con el hidrógeno (X-H), estos conjuntos de señales son moderadamente anchas y de intensidad media. Dentro del grupo anterior, el estiramiento O–H produce una banda ancha que se encuentra en el rango de 3700 a 3600  $\text{cm}^{-1}$ , el estiramiento N–H generalmente se observa entre 3400 y 3300  $\text{cm}^{-1}$  y puede diferenciarse más fácilmente que el estiramiento O-H.

Las bandas de estiramiento C–H asimétricas y simétricas de un grupo generalmente ocurren en aproximadamente 2965 y 2880  $\text{cm}^{-1}$ <sup>(119,130)</sup>.

Debido a la ausencia de triples enlaces en las muestras, no se evidencian bandas en la región entre 2300 y 2050  $\text{cm}^{-1}$  correspondiente a las vibraciones para este tipo de enlace<sup>(120)</sup>.

Las bandas principales en la región de los dobles enlaces se deben al estiramiento C=C y C=O. El estiramiento del grupo carbonilo (C=O) suelen ser las bandas más intensas del espectro, este comprende la región de 1830-1650  $\text{cm}^{-1}$ . El estiramiento C=C es mucho más débil y ocurre alrededor de 1650  $\text{cm}^{-1}$ , pero esta banda a menudo está ausente por motivos de simetría o momento dipolar. El estiramiento C=N también ocurre en esta región y generalmente es más fuerte<sup>(119,130,131)</sup>.

La mayoría de los enlaces simples absorben con números de onda similares y, por lo tanto, las vibraciones se acoplan. El patrón observado dependerá del esqueleto de carbono y los grupos funcionales adjuntos.

En números de onda de valores superiores a 1500  $\text{cm}^{-1}$ , generalmente es posible asignar cada banda de absorción en un espectro infrarrojo, pero para la mayoría de las absorciones observadas por debajo de este valor (región “fingerprints”), la asignación tiene mayor complejidad, esto se aplica a la mayoría de las vibraciones de flexión y esqueléticas (estructura de carbono de un compuesto), que absorben en la región de 1500–600  $\text{cm}^{-1}$ , para las cuales los pequeños efectos estéricos o electrónicos en la molécula provocan grandes cambios, generando en mayor proporción que cada banda corresponda a una deformación particular de la molécula, al movimiento de un grupo de átomos o enlace particular incluso entre similares<sup>(120)</sup>.

Los resultados observados en las figuras 12 y 13, así como en las tablas 3 y 4 correspondientes a las asignaciones basados en el número de onda reportado para las secuencias de nucleótidos indica que el análisis de todo el espectro provee información característica de los ácidos nucleicos como el ADN. Los modos vibratorios presentes pueden dividirse dependiendo de los grupos constituyentes: base, azúcar (desoxirribosa para este caso) y fosfato<sup>(87,121)</sup>.

La región principalmente asociada a esta macromolécula o a secuencias nucleotídicas se encuentra en el rango entre 1800–600  $\text{cm}^{-1}$  y se divide en cuatro dominios. El dominio I, entre 1800 y 1500  $\text{cm}^{-1}$ , muestra los picos asignados principalmente a vibraciones de estiramiento de doble enlace en el plano de las bases nitrogenadas<sup>(16,123,132)</sup>. El dominio II, entre 1500 y 1250  $\text{cm}^{-1}$ , es característico para la relación azúcar-base. El dominio III, entre 1250 y 1000  $\text{cm}^{-1}$ ,

contiene dos picos intensos del espectro IR de los ácidos nucleicos asignados a las vibraciones asimétrica y simétrica de los grupos fosfato, situados aproximadamente entre 1230 y 1089  $\text{cm}^{-1}$  respectivamente <sup>(126)</sup>.

Finalmente, el dominio IV, por debajo de 1000  $\text{cm}^{-1}$ , contiene absorciones que involucran vibraciones de los azúcares, vibraciones de la cadena de fosfodiéster y vibraciones fuera del plano de los dobles enlaces que conforman las bases <sup>(133,134)</sup>.

### 5.1.2. Generación y validación modelo estadístico para cuantificación de porcentaje de nucleótidos mediante los espectros ATR-FTIR para las pequeñas secuencias

La tabla 5 resume las figuras de mérito del método PLS desarrollado en el proceso de calibración y validación de los modelos para la cuantificación de porcentajes de nucleótidos de acuerdo con los resultados obtenidos a partir de 881 espectros tomados de las secuencias de nucleótidos descritas previamente en la tabla 1.

**Tabla 5.** Valores de los parámetros relevantes del modelo PLS para la cuantificación de porcentajes de nucleótidos.

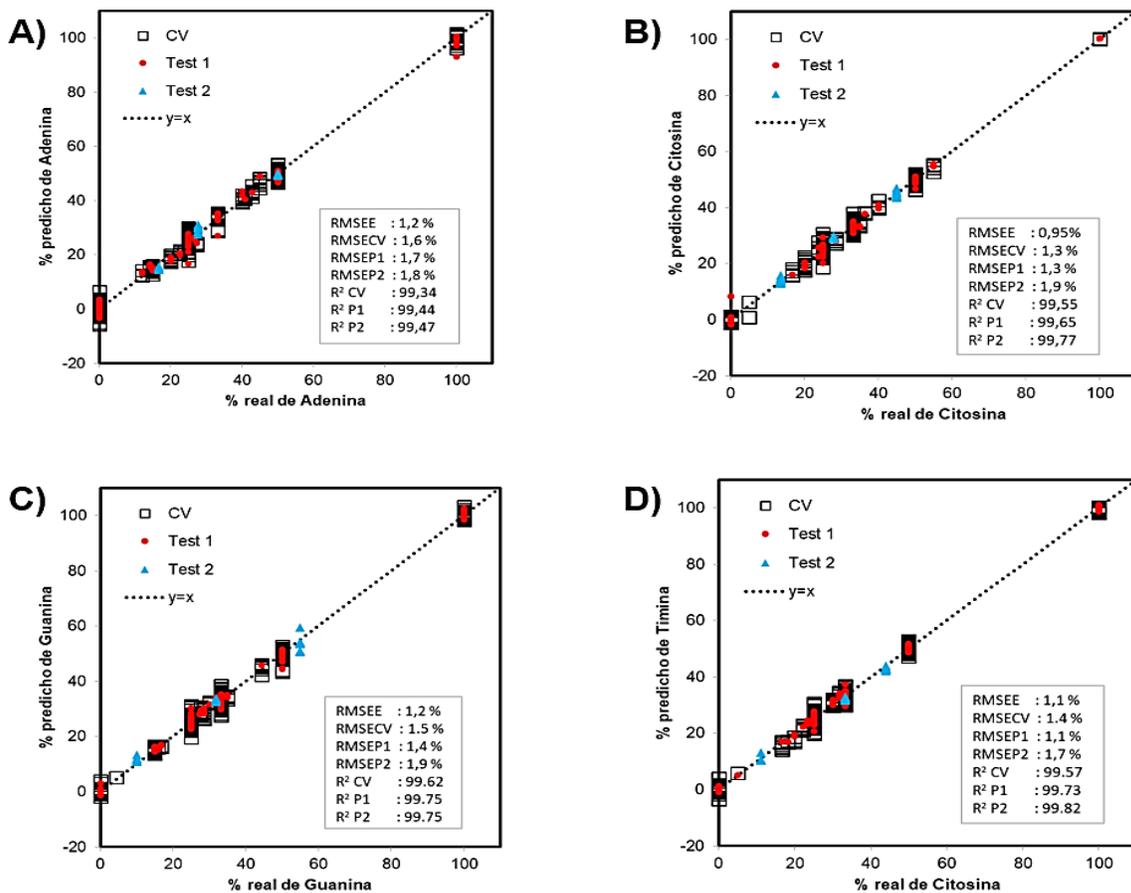
Parámetro	Valores			
	Adenina	Citosina	Guanina	Timina
LV	10	10	10	10
R <sup>2</sup> Cross Validation	99,34	99,55	99,62	99,57
R <sup>2</sup> Predicción 1	99,44	99,65	99,75	99,73
R <sup>2</sup> Predicción 2	99,47	99,77	99,75	99,82
RMSEE (%)	1,2	0,95	1,2	1,1
RMSEP1 (%)	1,7	1,3	1,4	1,1
RMSEP2 (%)	1,8	1,9	1,9	1,7
RMSECV (%)	1,6	1,3	1,5	1,4
Pre-procesamiento	VN	VN	VN	1D-VN
Regiones	3637-3277 2918-2197 183-1119	3996-3277 2918-2197 1838-1119	3637-3277 2918-2558 1838-758	3996-3635 3277-2558 1838-399

**LV:** variables latentes, **R<sup>2</sup>:** coeficiente de determinación, **RMSEE:** error cuadrático medio de estimación, **RMSEP:** error cuadrático medio de predicción, **RMSECV:** error cuadrático medio de validación cruzada, **VN:** Normalización del vector, **1D+VN:** 1ra Derivada + Normalización del vector.

El número de LV utilizados para los modelos fue de 10. Estos factores (LVs) en el algoritmo PLS son importantes para la obtención de un modelo robusto y confiable, indicando la posibilidad de concentrar en un menor número de variables, la información contenida en las variables medidas y al describir gran cantidad de varianza permite la reducción del error en el modelo<sup>(91)</sup>.

Los valores del coeficiente de determinación ( $R^2$ ) permitieron la verificación de la bondad de ajuste del modelo a las variables o muestras del conjunto de prueba (test) y cross validation, suministrando información sobre el porcentaje de varianza presente en los valores del % de nucleótidos de referencia reproducidos en el análisis PLS. Los resultados obtenidos de  $R^2$  para el conjunto de predicciones y el sistema de validación cruzada (cross validation) oscilaron entre 99,57 y 99,82, indicando el ajuste de los modelos en la predicción del porcentaje de nucleótidos<sup>(135)</sup>.

Así mismo, los gráficos de la Figura 14 revelan la buena relación entre los valores predichos (%) determinados por los modelos quimiométricos y los valores reales de referencia (%) para las muestras.



**Figura 14.** Valores predichos versus valores de referencia en los modelos aplicados para el porcentaje de: **A)** % Adenina, **B)** %Citosina, **C)** %Guanina y **D)** %Timina, mediante regresión por mínimos cuadrados parciales (PLS).

Los datos del conjunto de prueba (Test) se indican con puntos rojos y triángulos azules para los test 1 y 2 respectivamente; los datos de CV se indican mediante cuadrados. La línea de puntos representa el caso ideal o modelo perfecto ( $y=x$ ), en el que todos los valores predichos son iguales a los valores de referencia en todo el rango de datos. Se evidencia una baja dispersión de los datos (Figura 14) sobre la línea de mejor desempeño ( $y=x$ ) para los valores predichos de las calibraciones, validaciones cruzadas y pruebas. Comportamiento que indica el bajo porcentaje de error en los valores.

Los estadísticos mostraron errores relativos RMSEE, RMSECV y RMSEP entre 0,95 y 1,9%, valores considerados dentro del modelo de regresión, determinantes en el análisis de su adecuada capacidad de predicción<sup>(91)</sup>.

Del conjunto de pretratamientos evaluados por el software, el empleado para la optimización de los modelos de adenina, citosina y guanina fue VN. El preprocesamiento de datos de normalización del vector (en inglés, Vector Normalization) es un método que permite establecer un valor dado al área debajo de una o varias bandas, calculando el valor y promedio, restándolo posteriormente del espectro. Luego, se calcula la suma de cuadrados de todos los valores de  $y$ , y se divide por la raíz cuadrada de esta suma. Este método alternativo facilita la interpretación de los espectros<sup>(126)</sup>.

Para el caso del modelo para porcentaje de timina, el pretratamiento resultante fue 1D+VN. La primera derivada se obtiene como la diferencia entre absorbancias, a dos longitudes de onda consecutivas. El uso de derivadas mejora la diferenciación tras el solapamiento de picos y elimina desplazamientos lineales y cuadráticos de la línea base<sup>(92)</sup>.

A su vez, se registran las regiones espectrales consideradas por el análisis de optimización en el establecimiento de los mejores modelos para los porcentajes de nucleótidos en este trabajo (Tabla 5). Generalmente se excluyen algunas áreas con presencia de componentes no deseados (ruido) o con ausencia de bandas<sup>(136)</sup>. Se encontraron variabilidades en las señales entre las muestras a lo largo del espectro, por lo tanto, para este estudio se tomó en cuenta la mayor parte del espectro comprendido entre 4000 y 400  $\text{cm}^{-1}$ .

La aceptación del uso de todo el rango espectral puede relacionarse con las múltiples variaciones que se presentan. Generalmente los métodos de análisis

cuantitativo de infrarrojos utilizan las intensidades de los grupos C=O, N-H u O-H localizadas en la región MIR entre 3800 y 1600  $\text{cm}^{-1}$  (101,114,115).

Para el caso de muestras de nucleótidos (A, T, G y C), presentan diferencias a lo largo de la región espectral entre 1800 y 400  $\text{cm}^{-1}$ , donde los modos vibracionales son asignados dependiendo de la base, azúcar (desoxirribosa) y fosfato constituyente, así como sus interacciones (87,121).

## 5.2 Implementación de la técnica ATR-FTIR en la identificación de genotipos de virus de papiloma humano (VPH)

### 5.2.1 Caracterización de las muestras por espectrofotometría UV, la técnica estándar de real-time PCR y electroforesis.

En este trabajo se empleó la técnica de real-time PCR junto con un análisis de temperatura de melting ( $T_m$ ) para la detección de secuencias virales y determinación del genotipo de VPH asociado; sistema similar al planteado por Tsakogiannis *et al.*, 2014 (86).

Primeramente, se determinó la concentración de ADN de cada una de las muestras, obteniéndose valores entre 3 hasta 57  $\text{ng}/\mu\text{L}$  (Tabla 6). A partir de estos resultados, se tomaron de 1 a 4  $\mu\text{L}$  de ADN para el desarrollo de las reacciones PCR, siguiendo las indicaciones planteadas por el fabricante quien recomienda 10-100  $\text{ng}$  de ADN genómico para el uso de la Master Mix.

**Tabla 6.** Concentración de ADN ( $\text{ng}/\mu\text{L}$ ) obtenida por espectrofotometría UV para cada muestra de estudio y sus respectivas proporciones 260/280 y 260/230.

Código de la muestra	ADN ( $\text{ng}/\mu\text{L}$ )	A260/A280	A260/A230
<b>M01</b>	18,230	1,946	0,689
<b>M02</b>	23,935	1,893	0,821
<b>M03</b>	6,241	1,880	0,299
<b>M04</b>	20,772	1,868	0,587
<b>M05</b>	57,529	1,897	1,174
<b>M06</b>	12,763	2,040	0,551
<b>M07</b>	8,449	1,956	0,357
<b>M08</b>	13,911	1,957	0,596
<b>M09</b>	2,577	1,701	0,131

<b>M10</b>	4,933	2,354	0,227
<b>M11</b>	4,590	2,832	0,184
<b>M12</b>	3,986	2,786	0,197
<b>M13</b>	3,304	2,455	0,172
<b>M14</b>	9,996	2,095	0,420
<b>M15</b>	8,106	1,954	0,357
<b>M16</b>	3,930	2,278	0,136
<b>M17</b>	3,073	2,270	0,173
<b>M18</b>	3,738	1,993	0,187
<b>M19</b>	42,439	1,860	0,863

En la tabla 6 se muestran las proporciones 260/280 con resultados entre 1,70 a 2,83 y las proporciones 260/230 entre 0,13 a 1,17. Los primeros valores son indicadores de la pureza de las muestras en relación con la contaminación de proteínas y debe oscilar preferiblemente entre 1,8 y 2,0. Los obtenidos para la relación 260/230 evalúan la presencia de contaminación con reactivos de las muestras, donde los valores pueden estar entre 1,7 y 2,2 <sup>(11,69)</sup>. Lo anterior revela que la calidad (concentración y pureza) del material genético analizado no puede ser considerada óptima, condición posiblemente atribuida al tipo de muestra biológica o tratamiento previo de la misma, entre otros factores pueden hacer variar considerablemente la calidad de las muestras extraídas <sup>(137)</sup>. Sin embargo, La alta sensibilidad de la técnica de real-time PCR permitió la amplificación de las muestras positivas para alguno de los tipos virales estudiados a partir de mínimas concentraciones de ADN molde presentadas <sup>(138)</sup>.

En el proceso de amplificación se utilizaron primers específicos para la detección de 6 subtipos virales: VPH 16, 18, 31, 35, 51 y 66, dirigidos hacia la región más conservada entre el genoma de los papilomavirus, la región L1, la cual codifica la proteína estructural L1 que conforma la cápside viral y está implicada en el ensamblaje de los viriones y la invasión del virus a las células epiteliales <sup>(139)</sup>. Con base a lo descrito en el estudio realizado por Tsakogiannis et al., 2014 <sup>(86)</sup>, ellos verificaron la especificidad de estos cebadores utilizando la herramienta de búsqueda de alineación local básica de nucleótidos (BLAST; <http://blast.ncbi.nlm.nih.gov/Blast.cgi>).

De las 19 muestras de ADN provenientes de raspados citológicos de mujeres, se obtuvo un total de 16 resultados positivos mediante real-time PCR para infección por VPH, donde los genotipos detectados fueron 51 y 66 en 14 y 2 muestras respectivamente (anexo 2 y 3). Tres muestras mostraron resultado negativo para

los 6 tipos de VPH analizados (Tabla 7). La técnica real-time PCR permitió conocer el estado de amplificación de las muestras durante la fase exponencial de la reacción gracias a la detección de señales fluorescentes del colorante SYBR® Green. El aumento de la señal de fluorescencia se detectó proporcionalmente a la unión del colorante a cada nueva copia de ADN de doble hebra en la reacción PCR tras el proceso de hibridación de los primers con la secuencia específica (80,140,141).

**Tabla 7.** Valores de las temperaturas melting (Tm) resultantes en las amplificaciones de las muestras de ADN y genotipo de VPH asociado.

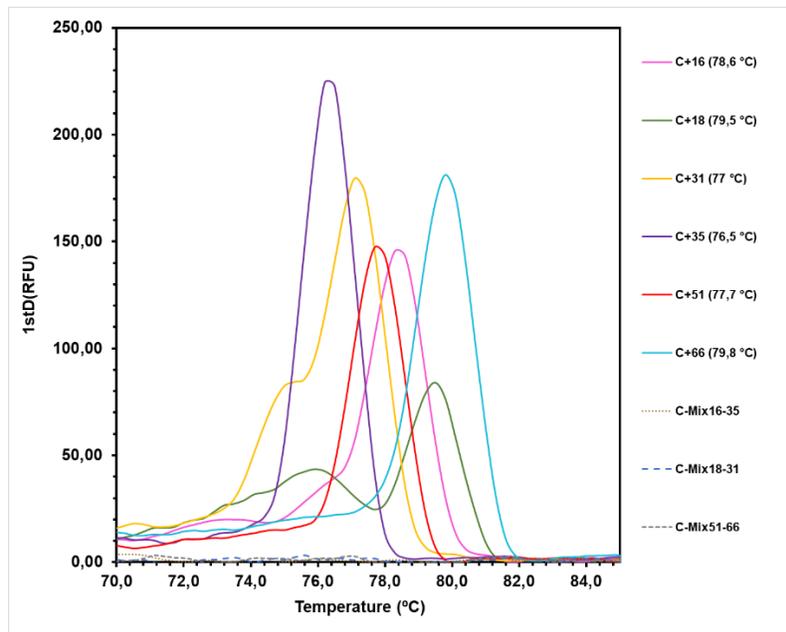
<b>Código de la Muestra</b>	<b>Tm (°C)</b>	<b>Genotipo VPH asociado</b>
M01	78,5	51
M02	78,5	51
M03	78,5	51
M04	78,5	51
M05	78	51
M06	78	51
M07	77,7	51
M08	77,7	51
M09	77,7	51
M10	77,7	51
M11	77,7	51
M12	77,7	51
M13	77,7	51
M14	77,7	51
M15	79,5	66
M16	79,8	66
M17	-	-
M18	-	-
M19	-	-

Los genotipos específicos de VPH se distinguieron y detectaron por sus valores de temperatura de fusión (Tabla 7). La Tm, es decir, la temperatura a la que la mitad de las hebras se encuentran en la conformación de doble hélice, está determinada por la estabilidad térmica del ADN. Debido a los enlaces de hidrógeno, las uniones

A-T tienen una temperatura de fusión más baja que las ricas en regiones G-C (142,143).

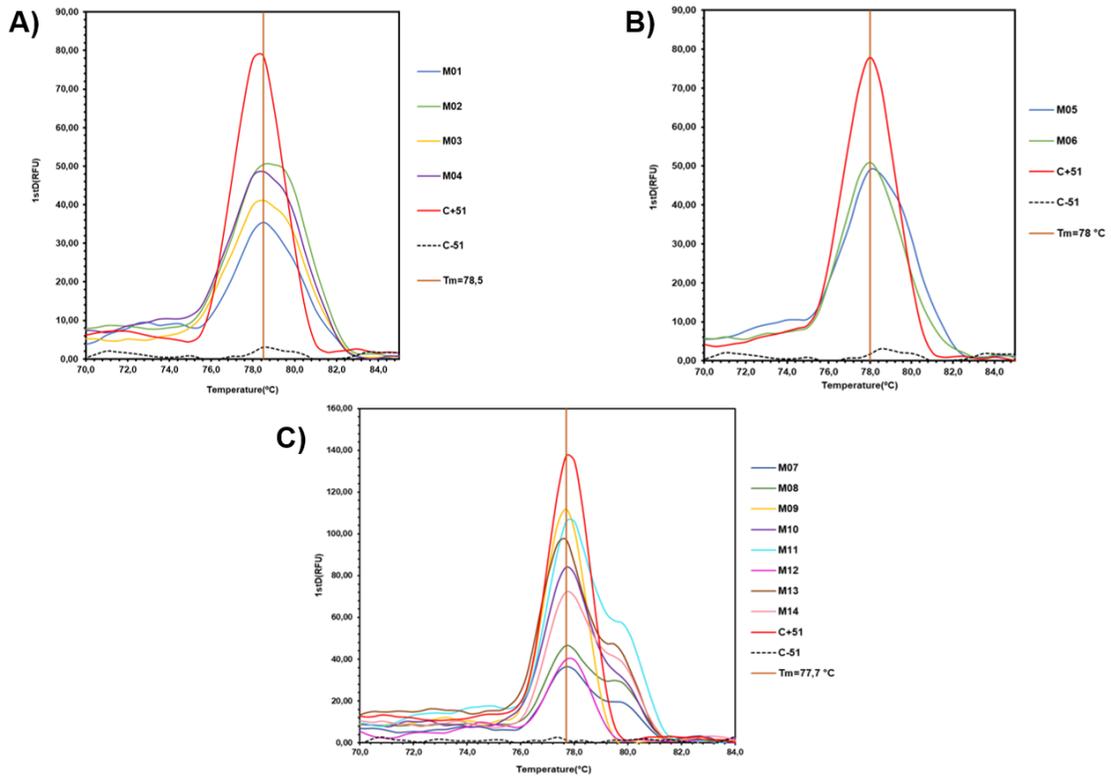
Con los datos de las temperaturas melting obtenidas posterior a los procesos de amplificación, se generaron gráficos representativos de la primera derivada del perfil de fluorescencia (1stD(RFU)) frente a la temperatura para discriminar entre los amplicones específicos y los productos de amplificación no específicos. En los ensayos de amplificación realizados con cada conjunto de cebadores (I, II y III) se relacionaron los valores de Tm obtenidos para los controles positivos de los tipos VPH 16, 18, 31, 35, 51 y 66 (Figura 15), con los resultados de Tm de las muestras para la determinación de infección por VPH, permitiendo el reporte de los gráficos donde los picos o puntos más altos de cada curva característica fueron determinantes de la Tm.

Es importante considerar los resultados por cada reacción ya que la temperatura de fusión de las muestras no siempre es la misma debido a diferentes factores como variaciones en la secuencia, concentración de guanina y citosina, longitud de los cebadores, la cantidad de reactivos y preparación de las muestras o a errores de manipulación y pipeteo. Es necesario observar detenidamente los valores de Tm y las formas de los picos de fusión para establecer un rango de temperaturas para cada genotipo, de esta forma las muestras pueden ser correctamente identificadas<sup>(81,138,144)</sup>.

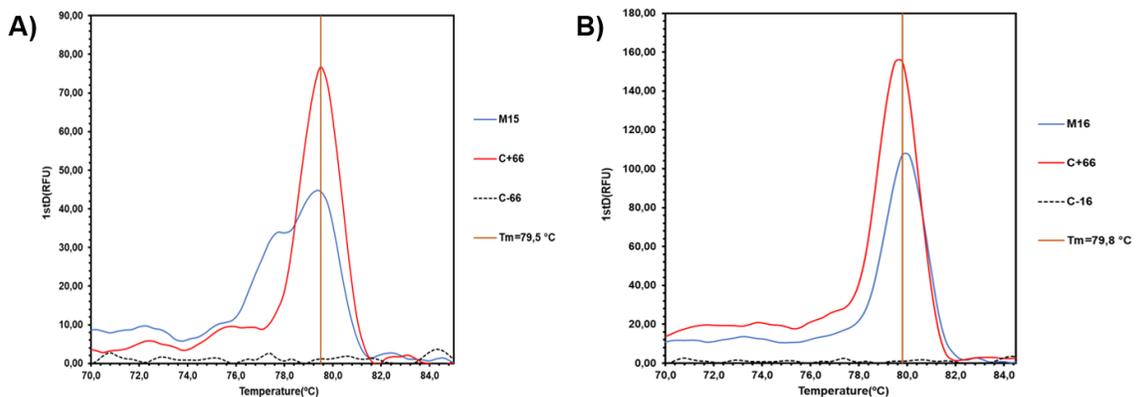


**Figura 15.** Curvas de temperatura de melting ( $T_m$ ) obtenidas para los controles positivos de VPH 16,18,31,35,51 y 66, posterior al proceso de real-time PCR SYBR Green. Eje y: Primera derivada del perfil de fluorescencia 1stD(RFU); Eje x: Temperatura en °C.

Con resultado positivo para VPH 51, se agruparon las muestras 1-4 con una temperatura de 78,5 °C, las muestras 5 y 6 con 78°C y las muestras 7-14 con 77,7°C (Figura 16). Para las muestras 15 y 16 con resultado positivo para VPH 66, las temperaturas melting fueron de 79,5 y 79,8 respectivamente (Figura 17).



**Figura 16.** Curvas de temperatura de melting ( $T_m$ ) obtenidas para las muestras amplificadas de ADN positivas para VPH 51, posterior al proceso de real-time PCR SYBR Green. **A)**  $T_m = 78.5$  para las muestras M01-M04; **B)**  $T_m = 78$  para las muestras M05 y M06; **C)**  $T_m = 77.7$  para las muestras M07-M14. Eje y: Primera derivada del perfil de fluorescencia 1stD(RFU); Eje x: Temperatura en °C.



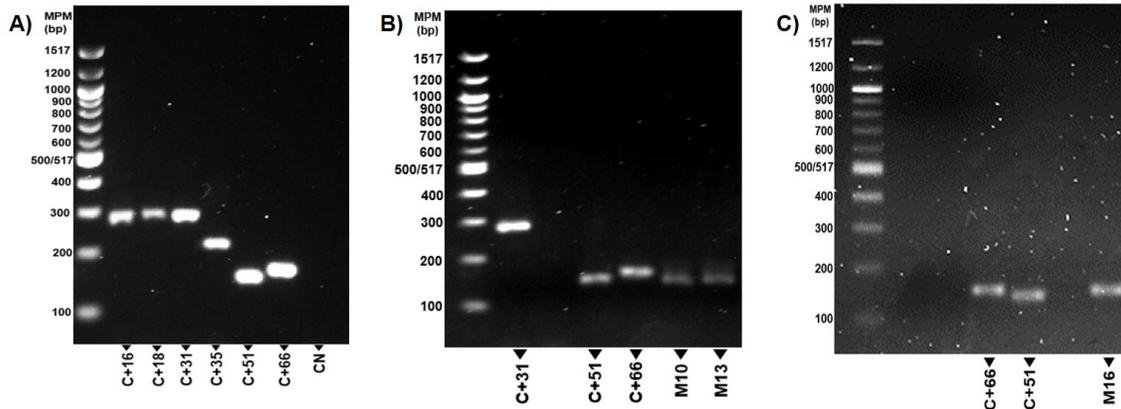
**Figura 17.** Curvas de temperatura de melting ( $T_m$ ) obtenidas para las muestras amplificadas de ADN positivas para VPH 66, posterior al proceso de real-time PCR SYBR Green. **A)**  $T_m = 79.5$  para la muestra M15; **B)**  $T_m = 79.8$  para la muestra M16. Eje y: Primera derivada del perfil de fluorescencia 1stD(RFU); Eje x: Temperatura en °C.

Para la verificación del tamaño de los productos real-time PCR amplificados se emplearon los valores reportados para los amplicones de los 6 tipos de VPH analizados (Tabla 8) y se compararon con los resultados del corrido electroforético de las muestras <sup>(86)</sup>.

**Tabla 8.** Secuencias de primers específicos correspondientes a los tipos de VPH 16, 35, 18, 31, 51 y 66, longitudes en pares de bases (bp) de los productos amplificados reportados con base a Tsakogiannis *et al.* 2015 <sup>(86)</sup> y temperaturas melting (Tm) en °C determinadas mediante real-time PCR SYBR Green.

Primers	Secuencia (5' - 3')	Producto PCR (bp)	Tm (°C)
HPV-16 L1F	ACTATTTTGGAGGACTGGAA	291	79
HPV-16 L1R	ATGAGGTGGTGGGTGTAGC		
HPV-35 L1F	AGGTCGTGGTCAGCCATT	227	77
HPV-35 L1R	GGTTAGCATTACAAGGTGTGC		
HPV-18 L1F	CCCCTGCCTCTACACAGTA	292	79.5
HPV-18 L1R	CCTGCTTATTGCCACCAC		
HPV-31 L1F	ACGTGCTCAGGGACAC	291	77.5
HPV-31 L1R	GGAGGTGTGGTCAATCC		
HPV-51L1 F	CCAATACCTAAAACCTCAAC	155	77.7
HPV-51L1 R	ACAACCCACACCAACCTA		
HPV-66L1 F	GCCGTAACGTATTCCCTA	168	80
HPV-66L1 R	AACAGCAAGCAACCTAGA		

Los resultados revelados por electroforesis en gel de agarosa respaldan la amplificación de los controles positivos (anexo 3) para los 6 tipos de VPH (Figura 15) al comparar la intensidad y posición de la banda con el marcador de paso molecular (Figura 18-A).



**Figura 18.** Electroforesis en gel de agarosa de los productos real-time PCR amplificados. **A)** Controles positivos para los genotipos de VPH 16, 18, 31, 35, 51 y 66. **B)** Controles positivos para los genotipos de VPH 51, VPH 66 y las muestras M10 y M13 con detección positiva del tipo de VPH 51. **C)** Controles positivos para los genotipos de VPH 51, VPH 66 y la muestra M16 con detección positiva del tipo de VPH 66.

Así mismo en los geles expuestos (Figura 18-B,C) confirma la validez de la especificidad genómica de los amplicones, confirmando con base al peso molecular, las amplificaciones positivas para los tipos de VPH 51 y 66 en este caso particular para las muestras 10,13 y 16. El rango de tamaño obtenido de los productos para las muestras de estudio VPH positivas para 51 y 66 se presentó entre 150-170 pb.

Si bien el número de pacientes involucrados en este estudio es bajo, la infección por VPH detectada en casi todas las muestras (16 de 19 casos) es un reflejo de las altas tasas de infección que se presentan y son reportadas en diferentes investigaciones<sup>(1,21,26,32,36,145)</sup>, esto, considerando la infección por VPH una de las enfermedades de transmisión sexual más común, estimando que hasta 79 millones de mujeres sexualmente activas a nivel mundial se infectarán con VPH durante su vida<sup>(146)</sup>.

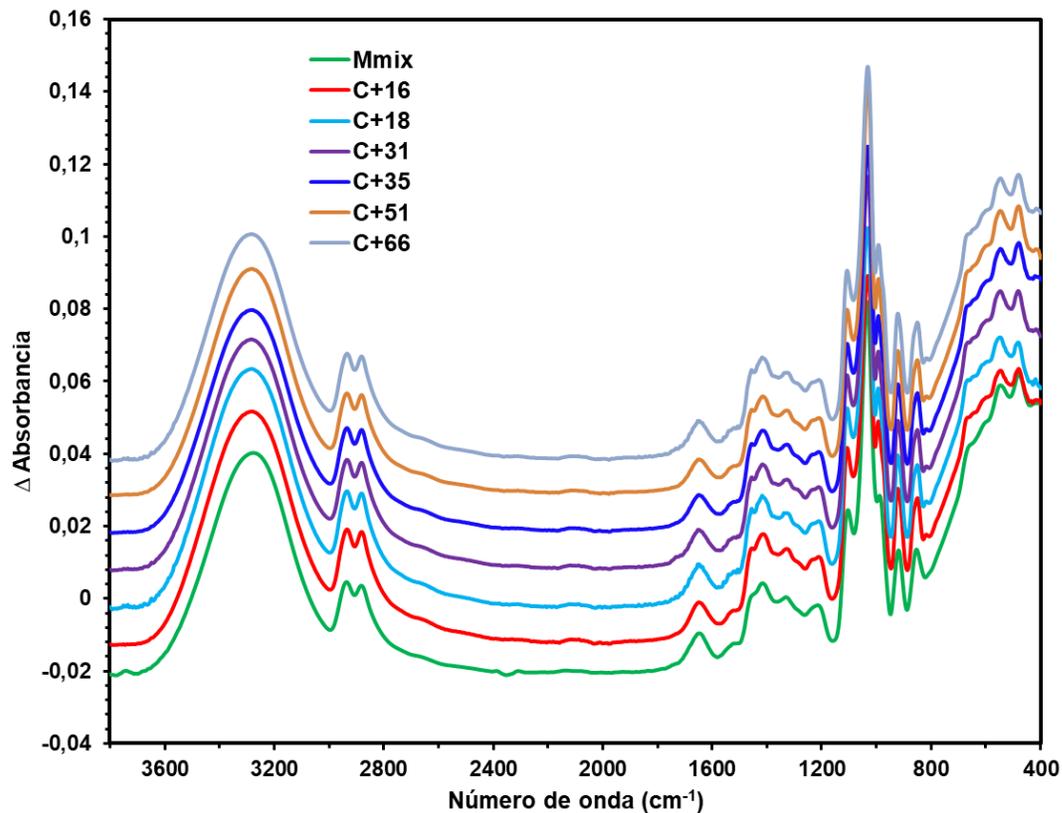
Lo anterior permite mencionar que la infección por virus del papiloma humano de alto riesgo es el agente etiológico más importante relacionado con la formación de los carcinomas de cuello uterino, el cual, a su vez parece mostrar una variación en la incidencia de los tipos de VPH en la población. En todo el mundo, los genotipos de VPH más comunes son los VPH 16/18 detectados hasta en un 70% de los

cánceres de cuello uterino <sup>(146)</sup>. Bedoya *et al.* en el 2019 <sup>(147)</sup> realizaron un estudio en mujeres de la región litoral de Ecuador, mayores de 30 años, con lesiones cervicales, donde el análisis por PCR mostró resultados VPH positivos para el 93,8 % del total de 320 muestras analizadas. El estudio representa un panorama heterogéneo sobre la prevalencia de los tipos VPH, donde los genotipos 58, 70, 53, 35, 35, 16, 33 y 31 presentaron porcentajes de frecuencia entre 25 hasta el 81% y para los genotipos de VPH 18, 51 y 66 los porcentajes obtenidos fueron de 19, 7 y 3% respectivamente. Los resultados de Trujillo *et al.*, en el 2016 <sup>(72)</sup>, mediante su estudio para la determinar la distribución de 37 genotipos de VPH en 543 muestras provenientes de mujeres de Bogotá rango entre 19 y 75 años, con citología cervicouterina anormal, revelaron que las infecciones con tipos de VPH de alto riesgo fueron las más frecuentes donde los tipos virales prevalentes fueron: VPH16 con 30,2%, 58-12,3%, 56-8,8%, 18-7,9%, 31-6,3%, 66-5,9%, y 51 con 5,5%.

Esta prevalencia enfatiza la necesidad de profesionales de la salud calificados, la realización de exámenes colposcópicos, moleculares y búsqueda de estrategias que permitan detectar y controlar la enfermedad <sup>(146)</sup>.

### **5.2.2 Preparación y adquisición de espectros de FTIR para los diferentes controles positivos, negativos y muestras amplificadas por real-time PCR.**

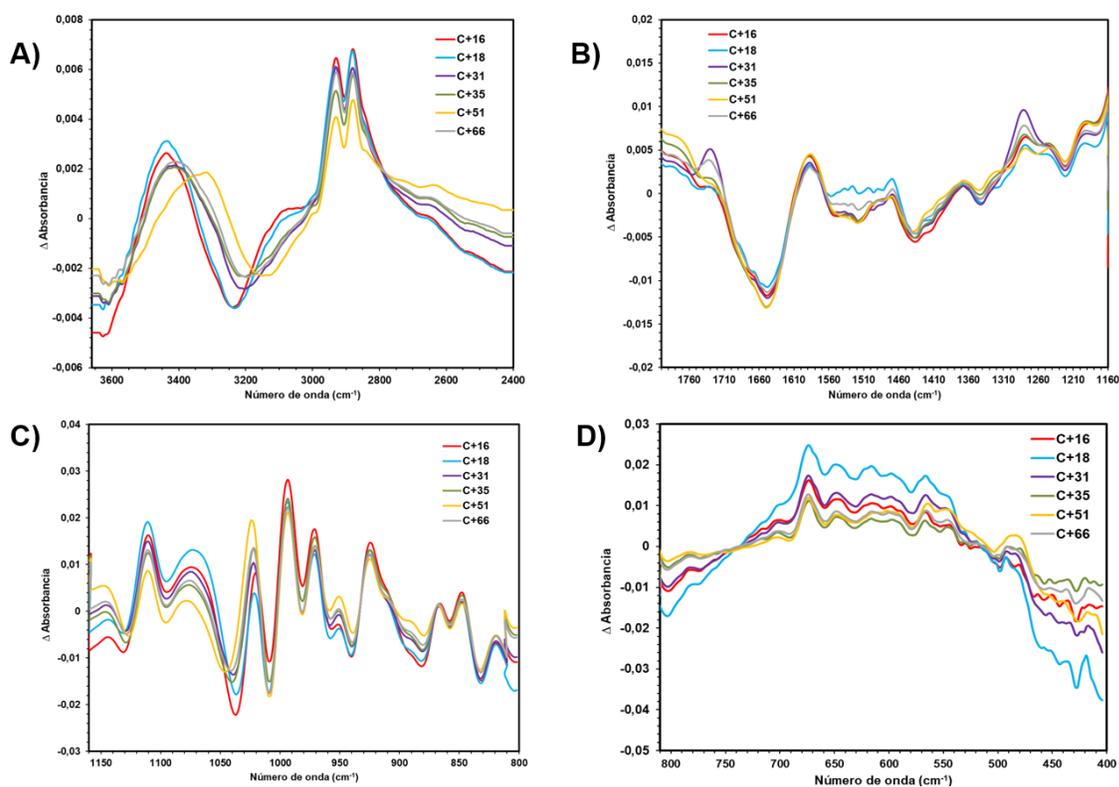
Se empleó el software OPUS™ para realizar el promediado y procesamiento por normalización del vector de los espectros obtenidos para los controles positivos amplificadas de VPH 13,18, 31, 35, 51 y 66, así, como de la solución PowerUp™ SYBR® Green Master Mix (Thermo Fisher Scientific) utilizada en este estudio para la preparación del mix PCR durante los procesos de amplificación.



**Figura 19.** Espectros ATR-FTIR de la solución de PowerUp™ SYBR® Green Master Mix (Thermo Fisher Scientific) y los controles positivos amplificados de VPH 16, 18, 31, 35, 51 y 66, promediados y desarrollados por normalización del vector en el rango 3800 a 400  $\text{cm}^{-1}$ . Eje y:  $\Delta$  Absorbancia; Eje x: Número de onda ( $\text{cm}^{-1}$ ).

Como se observa en la Figura 19, el comportamiento de las curvas de la solución Master Mix y los controles positivos amplificados en todo el rango espectral es similar. Lo anterior, sugiere que los espectros FTIR incluyen principalmente las vibraciones que surgen de los componentes de la solución PowerUp™ SYBR™ Green Master Mix, identificada con fórmula molecular  $\text{C}_{32}\text{H}_{37}\text{N}_4\text{S}$ , generando la aparición de bandas anchas posiblemente por la superposición y combinación de modos vibracionales de los enlaces químicos C-H, O-H, C-N, C=C, C=N<sup>(117,148)</sup>, grupos que pueden presentar picos de absorción en la región espectral de 3000 a 1700  $\text{cm}^{-1}$ , lo que no permite evidenciar las diferentes señales individuales entre las muestras de ADN control<sup>(89,117)</sup>.

Empleando la técnica de sustracción espectral presente en el software OPUS™ fue posible observar mayormente las diferencias espectrales entre cada uno de los controles amplificados (Figura 20).



**Figura 20.** Espectros ATR-FTIR de los controles positivos amplificados de VPH 16, 18, 31, 35, 51 Y 66 posterior al proceso de sustracción espectral compensada de la solución de PowerUp™ SYBR® Green Master Mix (Thermo Fisher Scientific). Regiones del espectro infrarrojo comprendidas entre diferentes rangos: **A)** 3663 a 2399  $\text{cm}^{-1}$ ; **B)** 1800 a 1160  $\text{cm}^{-1}$ ; **C)** 1160 a 810  $\text{cm}^{-1}$ ; **D)** 810 a 400  $\text{cm}^{-1}$ . Eje y:  $\Delta$  Absorbancia; Eje x: Número de onda ( $\text{cm}^{-1}$ ).

Antes del proceso de sustracción de la señal espectral de la master mix, se procedió a realizar cortes de la región del espectro infrarrojo en diferentes rangos: 3663 a 2399  $\text{cm}^{-1}$ , 1800 a 1160  $\text{cm}^{-1}$ , 1160 a 810  $\text{cm}^{-1}$  y 810 a 400  $\text{cm}^{-1}$ .

En esta parte se esperaba ver cambios entre los espectros debido a que las curvas que se evidencian en los gráficos son el resultado de las señales de las moléculas de ADN amplificadas en los controles detectados sin el solapamiento de las bandas de absorción provocadas por la presencia en las muestras del espectro de la solución Master Mix<sup>(148)</sup>. Sin embargo se pueden seguir presentando rastros a lo largo del espectro de esta solución que aún logran estar incorporados a la muestra<sup>(148)</sup>.

Es posible observar diferencias entre picos, desplazamiento entre bandas o cambios en intensidad<sup>(92)</sup> al comparar las señales obtenidas para los controles positivos amplificados de los tipos de VPH estudiados en las regiones entre 3500

a  $2800\text{ cm}^{-1}$ ,  $1700$  a  $1200\text{ cm}^{-1}$ ,  $1100$  a  $700\text{ cm}^{-1}$  y la última región de  $800$  a  $400\text{ cm}^{-1}$ .

Lo anterior, relacionado a las que absorciones IR en los ácidos nucleicos provocan vibraciones que surgen de diferentes zonas de la biomolécula de ADN. Las principales vibraciones pueden observarse en la región del espectro entre  $1800$  y  $700\text{ cm}^{-1}$ . El estiramiento de los dobles enlaces abarca el rango entre  $1800$  y  $1500\text{ cm}^{-1}$ , uniones base-azúcar entre  $1500$  y  $1250\text{ cm}^{-1}$ , las bandas entre  $1250$  y  $1000\text{ cm}^{-1}$  son producto de fuertes absorciones de grupos fosfato y del azúcar, y por debajo de  $1000\text{ cm}^{-1}$  vibraciones de la cadena fosfodiéster acopladas a las vibraciones del azúcar. Por consiguiente, es importante la observación del rango espectral completo para estudios y análisis de cambios que puedan ocurrir en el ADN <sup>(119,131,149)</sup>.

### **5.2.3 Generación y validación de modelos estadísticos para discriminación mediante los espectros para los diferentes genotipos de VPH.**

Dentro de los parámetros sugeridos por el software para la optimización de los modelos para la detección en las muestras de los genotipos de VPH 16, 18, 31, 35, 51 y 66 así como de muestras negativas, los resultados mostraron el uso del algoritmo MSC para los VPH 16 y 31, Straight Line Subtraction para VPH 18 y 51 y VN para los tipos de VPH 35 y 66. Además, el uso de diferentes regiones a lo largo del espectro identificadas según los diferentes parámetros como los rangos con mayor contribución en las variables para la discriminación entre estas, sugiere que es necesario abarcar todo el rango para un análisis y diferenciación más acertado entre las muestras <sup>(91,136,150)</sup>.

Las matrices de confusión con base al algoritmo PLS-DA se realizaron considerando variables categóricas de 1, que supera el umbral de 0,5 indicando la

pertenencia a determinada clase (genotipo de VPH u otro tipo de muestra), y, 0, para los valores negativos, es decir, inferiores a 0,5.

Conforme a los resultados obtenidos, la sensibilidad reportó valores entre 0,87 y 1, revelando la proporción de muestras que pertenecen al genotipo de VPH indicado en cada tabla y que son correctamente identificadas por el modelo matemático. En conjunto con el anterior parámetro, los valores de precisión, exactitud y F1-score con resultados oscilantes entre 0,88 y 1 representan el excelente rendimiento del modelo creado para la clasificación y discriminación de las muestras de controles positivos amplificadas para los genotipos de VPH, muestras amplificadas de ADN proveniente de los raspados cervicouterino positivas para VPH 51 y controles negativos que consistían en los espectros de las mezclas de todos los reactivos para PCR exceptuando ADN, así como los generados por la solución de la Master Mix.

Cross validation						
	Precisión	Sensibilidad	F1-Score	Exactitud	VPH16	Otros
VPH16	1	0,955223881	0,97709924	0,99484536	64	3
Otros	0,99420849	1	0,99709584		0	515
Test						
	Precisión	Sensibilidad	F1-Score	Exactitud	VPH16	Otros
VPH16	0,91176471	1	0,95384615	0,96972603	31	0
Otros	1	0,988505747	0,99421965		3	258

Cross validation						
	Precisión	Sensibilidad	F1-Score	Exactitud	VPH18	Otros
VPH18	0,97590361	0,94186047	0,95857988	0,98797251	81	5
Otros	0,98997996	0,99596774	0,99296482		2	494
Test						
	Precisión	Sensibilidad	F1-Score	Exactitud	VPH18	Otros
VPH18	0,93181818	0,97619048	0,95348837	0,98630137	41	1
Otros	0,99596774	0,988	0,99196787		3	247

Cross validation						
	Precisión	Sensibilidad	F1-Score	Exactitud	VPH31	Otros
VPH31	0,94827586	0,873015873	0,90909091	0,98109966	55	8
Otros	0,98473282	0,994219653	0,9894535		3	516
Test						
	Precisión	Sensibilidad	F1-Score	Exactitud	VPH31	Otros
VPH31	0,90322581	0,965517241	0,93333333	0,98630137	28	1
Otros	0,99616858	0,988593156	0,99236641		3	260

Cross validation						
	Precisión	Sensibilidad	F1-Score	Exactitud	VPH35	Otros
VPH35	1	0,93023256	0,96385542	0,99484536	40	3
Otros	0,99446494	1	0,99722479		0	539
Test						
	Precisión	Sensibilidad	F1-Score	Exactitud	VPH35	Otros
VPH35	0,95238095	1	0,97560976	0,99657534	20	0
Otros	1	0,99632353	0,99815838		1	271

Cross validation						
	Precisión	Sensibilidad	F1-Score	Exactitud	VPH51	Otros
VPH51	0,92727273	0,927272727	0,92727273	0,97250859	102	8
Otros	0,98305085	0,983050847	0,98305085		8	464
Test						
	Precisión	Sensibilidad	F1-Score	Exactitud	VPH51	Otros
VPH51	0,98214286	0,93220339	0,95652174	0,98287671	55	4
Otros	0,98305085	0,995708155	0,98933902		1	232

Cross validation						
	Precisión	Sensibilidad	F1-Score	Exactitud	VPH66	Otros
VPH66	0,95652174	0,84615385	0,89795918	0,98281787	44	8
Otros	0,98507463	0,99622642	0,99061914		2	528
Test						
	Precisión	Sensibilidad	F1-Score	Exactitud	VPH66	Otros
VPH66	0,88461538	1	0,93877551	0,98972603	23	0
Otros	1	0,98884758	0,99439252		3	266

Cross validation						
	Precisión	Sensibilidad	F1-Score	Exactitud	NEG	Otros
NEG	1	0,97520661	0,9874477	0,99484536	118	3
Otros	0,99353448	1	0,99675676		0	461
Test						
	Precisión	Sensibilidad	F1-Score	Exactitud	NEG	Otros
NEG	0,95	1	0,97435897	0,96972603	57	0
Otros	1	0,98723404	0,99357602		3	232

**Figura 21.** Tablas de confusión de los modelos de predicción evaluados para el conjunto de datos en la detección de los tipos de VPH 16, 18, 31, 35, 51 y 66. Se presentan los valores de precisión, sensibilidad, F1-Score y exactitud resultantes de los cálculos en cross validation y test para cada modelo.

Los valores obtenidos en este estudio son mayores a los reportados por Pereira *et al.*,<sup>(151)</sup> quienes realizaron un análisis de espectroscopia FT-IR en muestras de líquido cervical de 50 pacientes mujeres con citología cervical normal; en esta investigación obtuvieron una tasa de verdaderos positivos o sensibilidad de 76,9% y para la tasa de verdaderos negativos o especificidad un valor de 76,7% para la detección de células normales e infectadas por VPH. Neves *et al.*,<sup>(152)</sup> destacaron la potencialidad de la espectroscopia ATR-FTIR combinada con análisis multivariante como herramienta de detección en plasma sanguíneo de cáncer cervical precanceroso, en su investigación obtuvieron valores de sensibilidad alrededor del 90% y una especificidad del 83%. A su vez, Paraskevaidi *et al.*,<sup>(150)</sup> enfatizaron en el uso de las técnicas espectroscópicas en la detección y el diagnóstico de carcinomas ginecológicos, considerando que en su estudio emplearon ATR-FTIR para analizar muestras de orina de mujeres con cáncer de endometrio, de ovario y de individuos sanos. Posterior al proceso de aplicación de métodos multivariados y algoritmos de clasificación alcanzaron porcentajes de 95% de sensibilidad, 100% de especificidad y 95% de precisión para el cáncer de endometrio; para el cáncer de ovario 100% de sensibilidad, 96,3% de especificidad y 100 % de precisión. Ambos autores resaltan el uso de ATR-FTIR como una técnica que brinda gran respaldo en el proceso de cribado para la reducción de la incidencia de enfermedades como el cáncer cervical<sup>(78)</sup>.

La Figura 22 señala las predicciones realizadas por cada modelo según la clasificación para los espectros de entrada de las muestras con relación a los resultados obtenidos por la técnica estándar de real-time PCR.

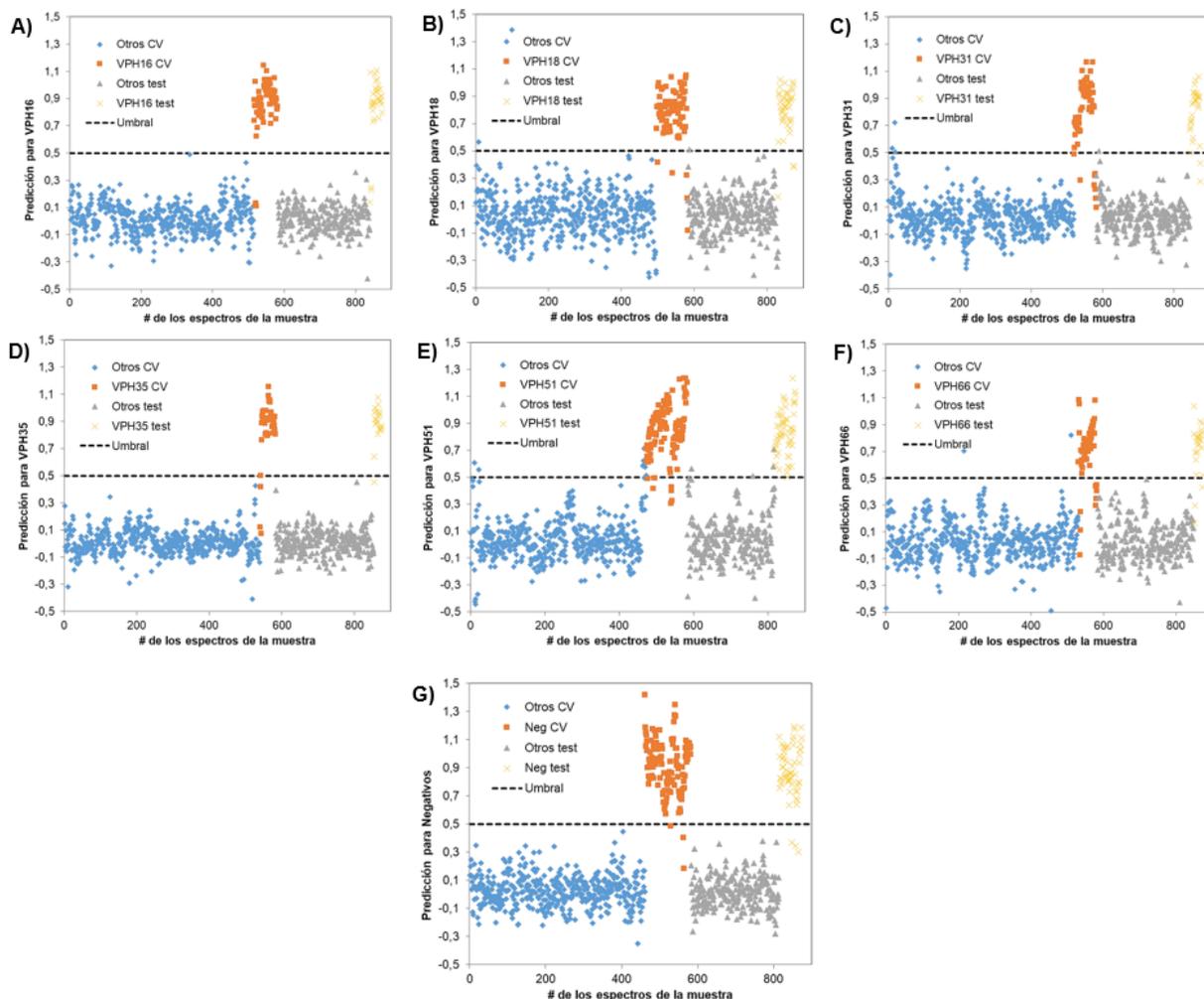
		Model_VPH16	Model_VPH18	Model_VPH31	Model_VPH35	Model_VPH51	Model_VPH66	Model_negativos	real-time PCR
MUESTRAS	M01	-	-	-	-	+	-	-	VPH51
	M02	-	-	-	-	+	-	-	VPH51
	M03	-	-	-	-	+	-	-	VPH51
	M04	-	-	-	-	+	-	-	VPH51
	M05	-	-	-	-	+	-	-	VPH51
	M06	-	-	-	-	+	-	-	VPH51
NEG	Mmix	-	-	-	-	-	-	+	-
CONTROLES POSITIVOS	C+VPH16	+	-	-	-	-	-	-	C+VPH16
	C+VPH18	-	+	-	-	-	-	-	C+VPH18
	C+18_1	-	-	-	-	-	-	-	C+VPH18
	C+VPH31	-	-	+	-	-	-	-	C+VPH31
	C+31_1	-	-	-	-	-	-	-	C+VPH31
	C+31_2	-	-	+	-	+	-	-	C+VPH31
	C+35	-	-	-	+	-	-	-	C+VPH35
	C+51	-	-	-	-	+	-	-	C+VPH51
	C+51_1	-	-	-	-	+	-	-	C+VPH51
	C+51_2	-	-	-	-	+	-	-	C+VPH51
	C+VPH66	-	-	-	-	-	+	-	C+VPH66
	C+66_1	-	-	-	-	-	-	-	C+VPH66
C+66_2	-	-	+	-	+	-	-	C+VPH66	
NEG	C-NEG	-	-	-	-	-	-	+	Neg

**Figura 22.** Resultados de la ejecución de los modelos de detección para los 6 genotipos de VPH analizados y modelo de predicción de muestras negativas. (+) Resultado positivo predicho por el modelo correspondiente; (-) resultado predicho como negativo. En rojo se presentan los resultados de predicción atípicos.

El modelo para detección de VPH 51 predijo como positivo todas las muestras clínicas, coincidiendo con los resultados real-time PCR, así mismo, los modelos negativos indicaron de forma acertada aquellos considerados como negativos. Para los controles positivos de VPH 16, 18, 31, 35, 51 y 66 los modelos predijeron correctamente la mayor parte de ellos, exceptuando algunos controles positivos amplificadas en otras reacciones (C+18\_1, C+31\_1, c+31\_2, C+66\_1 y C+66\_2), que también se tomaron para la prueba de los modelos (Figura 22). Con base a lo anterior es posible considerar que a pesar del éxito general en la clasificación por parte de los modelos, se requiere el análisis de las variables empleadas, espectros, muestras, procedimientos previos y uso de componentes, entre otros factores que puedan intervenir o generar variaciones en el resultado final de la evaluación del modelo <sup>(91,118,135,153)</sup>.

En la figura 23 se muestran también los gráficos de dispersión basados en el algoritmo PLS-DA para todos los modelos (Figura 23).

Con base a los espectros tomados por muestra (en total, 874 espectros), se observan agrupaciones de acuerdo a los resultados de la predicción de cada modelo para la detección de VPH 16, 18, 31, 35, 51 y 66. La línea punteada indica el valor del umbral (0,5), necesario para la discriminación entre muestras positivas (>0,5) y negativas (<0,5).



**Figura 23.** Gráficos de dispersión PLS-DA de los espectros FTIR empleados para la evaluación de los modelos para predicción de los tipos de VPH 16(A),18(B), 31(C), 35(D) ,51(E) y 66(F) y negativos(G), en los espectros de muestras clínicas amplificadas, controles positivos amplificadas, solución de Master mix y controles negativos. Línea punteada (---) indica el valor del umbral equivalente a 0.5.

En cada gráfico (A-G) de la Figura 23 se presentaron 4 grupos. Los que superaban el umbral de 0.5 fueron clasificados según el modelo como resultados

positivos para un genotipo determinado de VPH (especificados en colores naranja y amarillo), y los asignados como algún otro genotipo o muestra posible con valores menores a 0.5 (con colores azul y gris); ambos analizados en función de las estimaciones resultante mediante el conjunto de test y cross validation.

Estos gráficos de dispersión permitieron observar las variaciones entre los datos dispuestos, la detección e interpretación de las agrupaciones, similitudes y diferencias entre un conjunto de muestras. Según lo observado es posible considerar la necesidad de ajustes en los modelos ya que se presentan datos atípicos representados como puntos de espectros tomados de muestras detectadas previamente por análisis PCR como positivas, pero que arrojan valores por debajo del umbral; esto indica que no se logró una discriminación completa entre los tipos de VPH, lo que puede causar que estos clasifiquen erróneamente algunos datos de entrada.

De forma general, los resultados en su mayoría sugieren el poder predictivo de todos los modelos, con la posibilidad de mejorar la capacidad de clasificación mediante ajustes metodológicos como la toma y adición de mayor cantidad de datos de tal forma que contribuyan en la identificación de la variedad de datos espectrales que pueden ser requeridos en los análisis<sup>(154–156)</sup>.

## 6. Conclusiones y Recomendaciones

### 6.1 Conclusión

Las técnicas de espectroscopía vibracional como ATR-FTIR son ampliamente utilizadas por las ventajas que ofrece al ser una técnica no destructiva, su alto rendimiento, la poca cantidad y preparación de la muestra y la obtención de información fundamental química y estructural de diferentes muestras biológicas.

El conjunto de datos generados puede ser evaluado mediante métodos quimiométricos con técnicas de análisis multivariados de tal forma que el tratamiento, relación e interpretación entre las variables estudiadas permitan su implementación en el campo biomédico. La adquisición de espectros en esta investigación fue realizada en el intervalo espectral entre 4000 a 400  $\text{cm}^{-1}$  utilizando un espectrómetro ALPHA FTIR spectrometer ATR, equipado con un cristal de diamante (Bruker Optics), a una resolución espectral de 4  $\text{cm}^{-1}$ . Toda la región espectral se consideró informativa para la determinación de variaciones entre muestras, donde, el rango de 1800-600  $\text{cm}^{-1}$  puede ser mayormente empleado debido a la aparición de bandas relacionadas a los constituyentes del ADN de interés. Los parámetros estimados mediante el uso del algoritmo PLS en la generación y evaluación de modelos para la cuantificación del porcentaje de nucleótidos en secuencias de ADN de cadena sencilla, revelaron resultados de predicción exitosos considerando el  $R^2$  mayor a 99.5 y valores RMSEE entre 0.9-1.2%. El análisis de las  $T_m$  obtenidas mediante la técnica real-time PCR mostró la alta prevalencia de infección por VPH (16 de 19 casos) en este estudio, 14 muestras amplificaron positivamente para el genotipo VPH 51, considerado según la agencia internacional contra el cáncer (IARC) como un tipo viral de alto riesgo oncogénico. La discriminación mediante análisis PLS-DA de muestras positivas o negativas para los tipos de VPH 16, 18, 31, 35, 51 y 66 mostraron las contribuciones espectrales en la clasificación de muestras reales y controles positivos y negativos con una precisión cercana a 1. Los genotipos virales detectados por la técnica de real-time PCR en fue predicho con el mismo resultado por los modelos en todas las muestras.

## **6.2 Recomendaciones**

Emplear un número mayor de muestras que permitan el robustecimiento de la data espectral para la discriminación exitosa entre los diferentes tipos de VPH y la precisión en la predicción de muestras con resultados positivos o negativos.

Generar modelos de doble cadena de ADN para predicción del % de nucleotidos para llevar a cabo la predicción de los %N en los genotipos de VPH.

Considerando las alteraciones estructurales o moleculares que pueden presentar las biomoléculas en condiciones patológicas, es posible realizar un análisis espectral entre muestras de personas sanas y muestras con detección VPH positivas para evidenciar estos cambios.

La complementariedad entre técnicas vibracionales permite sugerir el uso de espectroscopía Raman como una técnica de dispersión que posibilite la visualización de los picos espectrales que no hayan sido detectados por FTIR.

## 7. Referencias Bibliográficas

1. Balasubramaniam SD, Balakrishnan V, Oon CE, Kaur G. medicina Key Molecular Events in Cervical Cancer Development. 2019 [cited 2020 Aug 23]; Available from: [www.mdpi.com/journal/medicina](http://www.mdpi.com/journal/medicina)
2. Subramanya D, Grivas PD. HPV and cervical cancer: Updates on an established relationship. *Postgrad Med*. 2008;120(4):7–13.
3. Cohen PA, Jhingran A, Oaknin A, Denny L. Cervical cancer. *Lancet* [Internet]. 2019;393(10167):169–82. Available from: [http://dx.doi.org/10.1016/S0140-6736\(18\)32470-X](http://dx.doi.org/10.1016/S0140-6736(18)32470-X)
4. Genética. Un enfoque conceptual | Acceso a Material Complementario del Estudiante [Internet]. [cited 2020 Aug 23]. Available from: <https://www.medicapanamericana.com/materialesComplementarios/PierceEst/Pierce.aspx>
5. Olusola P, Banerjee HN, Philley J V., Dasgupta S. Human Papilloma Virus-Associated Cervical Cancer and Health Disparities. *Cells* [Internet]. 2019 Jun 21 [cited 2020 Aug 23];8(6):622. Available from: [/pmc/articles/PMC6628030/?report=abstract](http://pmc/articles/PMC6628030/?report=abstract)
6. Mateos-Lindemann ML, Pérez-Castro S, Rodríguez-Iglesias M, Pérez-Gracia MT. Microbiological diagnosis of human papilloma virus infection. *Enferm Infecc Microbiol Clin* [Internet]. 2017;35(9):593–602. Available from: <http://dx.doi.org/10.1016/j.eimc.2016.05.008>
7. Li W, Padilla C, Gutierrez E, Hajar G. Detección molecular y genotipificación de virus del papiloma humano como tamizaje de cáncer de cuello uterino: Posibilidades en el contexto peruano. *Bol del Inst Nac Salud* [Internet]. 2016;22(5):22–8. Available from: <http://repositorio.ins.gob.pe/handle/INS/907>
8. Cruz L. IMPLEMENTACIÓN DE UN SENSOR ÓPTICO EN LA IDENTIFICACIÓN BACTERIANA. BENEMÉRITA UNIVERSIDAD AUTÓNOMA DE PUEBLA; 2019.
9. Vargas V. La asociación de la microbiota, virus del papiloma humano y cáncer cervicouterino. *Rev Hosp Jua Mex* [Internet]. 2018;85(1):6–8. Available from: [www.medigraphic.com/pdfs/juarez/ju-2018/ju181b.pdf](http://www.medigraphic.com/pdfs/juarez/ju-2018/ju181b.pdf)
10. Ramirez-Pineda AT, González MI, Castañeda-Vanegas KM, Agudelo-Fernández MC, López-Urán C, Sánchez-Vásquez GI. Filogenia y oncogénesis del virus del papiloma humano: una aproximación translacional al descubrimiento de biomarcadores para la detección de lesiones precancerosas de cérvix. *Rev la Acad Colomb Ciencias Exactas, Físicas y Nat*. 2019;43(168):351–65.
11. Rymysza T, Ribeiro EA, de Carvalho LF das CES, Bhattacharjee T, de Azevedo Canevari R. Human papillomavirus detection using PCR and ATR-FTIR for cervical cancer screening. *Spectrochim Acta A Mol Biomol Spectrosc* [Internet]. 2018 May 5

[cited 2020 Aug 4];196:238–46. Available from:  
<http://www.ncbi.nlm.nih.gov/pubmed/29454252>

12. Piñeros M, Parkin DM, Ward K, Chokunonga E, Ervik M, Farrugia H, et al. Essential TNM: a registry tool to reduce gaps in cancer staging information. *Lancet Oncol* [Internet]. 2019;20(2):e103–11. Available from: [http://dx.doi.org/10.1016/S1470-2045\(18\)30897-0](http://dx.doi.org/10.1016/S1470-2045(18)30897-0)
13. Vega MQ, Gómez JFC, Bastidas M, Márquez L, Pons JP. Detección y tipificación de virus del papiloma humano (VPH) mediante PCR- RFLP. *Rev Obstet Ginecol Venez.* 2008;68(1):25–31.
14. Achig N. CORRELACION DIAGNOSTICA ENTRE LOS RESULTADOS CITOLOGICOS POR PAPTEST Y LOS RESULTADOS DE PCR EN TIEMPO REAL DEL VIRUS DEL HPV DE ALTO RIESGO REALIZADOS A MUJERES DE ENTRE 30 A 60 AÑOS QUE ACUDEN AL HOSPITAL CARLOS ANDRADE MARIN EN EL PERIODO ENERO- MARZO . *Proy Univ Cent DEL ECUADOR Fac CIENCIAS MÉDICAS CARRERA* [Internet]. 2016;23(45):5–24. Available from: <http://www.dspace.uce.edu.ec/handle/25000/8092>
15. Martínez N, Martín MC, Herrero A, Fernández M, Alvarez MA, Ladero V. QPCR as a powerful tool for microbial food spoilage quantification: Significance for food quality. *Trends Food Sci Technol.* 2011;22(7):367–76.
16. Wittwer CT, Makrigiorgos GM. *Nucleic Acid Techniques* [Internet]. Principles and Applications of Molecular Diagnostics. Elsevier Inc.; 2018. 47–86 p. Available from: <http://dx.doi.org/10.1016/B978-0-12-816061-9.00004-7>
17. Mata-Miranda MM, Guerrero-Robles CI, Rojas-López M, Delgado-Macuil RJ, González-Díaz CA, Sánchez-Monroy V, et al. Principal components by FTIR spectroscopy as innovative characterization technique during differentiation of pluripotent stem cells to pancreatic cells. *Rev Mex Ing Biomed.* 2017;38(1):225–34.
18. Barraza G, Martinez-martinez A. Transformada De Fourier ( Ffirm ) En El Estudio De. 2012;(October 2014).
19. Rodrigues RPCB, Aguiar EMG, Cardoso-sousa L, Caixeta DC, Guedes CCF V, Siqueira WL, et al. Differential Molecular Signature of Human Saliva Using ATR-FTIR Spectroscopy for Chronic Kidney Disease Diagnosis. *Braz Dent J* [Internet]. 2019;30:437–45. Available from: [https://www.scielo.br/scielo.php?script=sci\\_arttext&pid=S0103-64402019000500437](https://www.scielo.br/scielo.php?script=sci_arttext&pid=S0103-64402019000500437)
20. Kowalczyk D, Pitucha M. Application of FTIR method for the assessment of immobilization of active substances in the matrix of biomedical materials. *Materials (Basel).* 2019;12(18).
21. J Bowden S, Kyrgiou M. Human papillomavirus. *Obstet Gynaecol Reprod Med* [Internet]. 2020;30(4):109–18. Available from: <https://doi.org/10.1016/j.ogrm.2020.02.003>

22. Feemster K. OMS | Virus del papiloma humano. In: Robert M. Kliegman, MD, Joseph St. Geme, MD, Nathan Blum, Samir S. Shah and Robert C. Tasker, MA, MD M, editor. Nelson Tratado de pediatría [Internet]. 21.ª Edici. Elsevier España, S.L.U.; 2020 [cited 2020 Aug 14]. p. 1747–52. Available from: [https://www.who.int/biologicals/areas/human\\_papillomavirus/en/](https://www.who.int/biologicals/areas/human_papillomavirus/en/)
23. María-ortiz JS, Álvarez-silvares E, Bermúdez-gonzález M, Lavandeira SG, Mosquera MP, Cambeiro BC. Importancia de los márgenes quirúrgicos afectados en la conización uterina cervical Importance of surgical margins affected in cervical uterine conization . 2020;88(9):586–97.
24. Medina Magües LG. Genotipificación del Virus del Papiloma Humano mediante secuenciamiento y PCR cuantitativa en tiempo real y detección de variantes intratípicas por análisis filogenético. Esc Super Politécnica del Litoral [Internet]. 2015;131. Available from: <https://www.dspace.espol.edu.ec/bitstream/123456789/29767/1/TESIS-ESPOL-LexMedina.pdf>
25. Vitriago-Rendón AM, Aguilar-Mejía MS, Michelli-Gago PJ, Celaya Linaza J, Gutiérrez C. Evaluación de la expresión de ARNm de genes virales E2, E6 y E7 como marcadores predictivos de progresión en lesiones producidas por VPH 16. Invest Clin. 2018;59(4):302–17.
26. Wendland EM, Villa LL, Unger ER, Domingues CM, Benzaken AS, Maranhão AGK, et al. Prevalence of HPV infection among sexually active adolescents and young adults in Brazil: The POP-Brazil Study. Sci Rep. 2020;10(1):1–10.
27. Ginsburg OM. Breast and cervical cancer control in low and middle-income countries: Human rights meet sound health policy. J Cancer Policy [Internet]. 2013;1:35–41. Available from: <http://dx.doi.org/10.1016/j.jcpo.2013.07.002>
28. Arroyo Andújar JD. Detección e Identificación de los virus del papiloma humano. Caracterización de dos nuevas variantes [Internet]. Universitat autonoma de Barcelona; 2015. Available from: <https://www.tdx.cat/handle/10803/310595>
29. Paho. Incorporación de LA PRUEBA DEL VIRUS DEL PAPILOMA HUMANO en PROGRAMAS DE PREVENCIÓN DE CÁNCER CERVICOUTERINO [Internet]. Manual de VPH. 2016. 9–17 p. Available from: [http://www2.paho.org/hq/index.php?option=com\\_docman&task=doc\\_view&Itemid=270&gid=36310&lang=es](http://www2.paho.org/hq/index.php?option=com_docman&task=doc_view&Itemid=270&gid=36310&lang=es)
30. Isaza-ruget MA, Perez G, Morales-reyes OL, Deantonio-suárez R, Alvarado-heine C, Trujillo LM. EXACTITUD DEL TEST ADN-HPV PARA LA DETECCIÓN DE LA ENFERMEDAD CERVICAL DE ALTO GRADO ( NIC 2 + ) EN MUJERES CON ANORMALIDADES CITOLÓGICAS ( ASC-US Y LSIL ), AFILIADAS A LA SEGURIDAD SOCIAL EN BOGOTÁ ( COLOMBIA ) The accuracy of the HPV-DNA test for detect. 2009;60(3):213–22.
31. Rojas Mendoza G, Córdova Uscanga C, Sánchez López J. Evaluación del estudio de Papanicolaou y la colposcopia en el diagnóstico de neoplasia intraepitelial cervical en la Unidad Especial Centro de Apoyo Diagnóstico San Rafael. Rev Espec

Médico-Quirúrgicas [Internet]. 2012;17(2):76–80. Available from:  
<https://www.medigraphic.com/cgi-bin/new/resumen.cgi?IDARTICULO=35132>

32. Trujillo Perdomo T de la C, Domínguez Bauta SR, Ríos Hernández M de los A, Menéndez MH. Prevalencia del virus del papiloma humano en mujeres con citología negativa. *Rev Cuba Obstet y Ginecol.* 2017;43(1):1–13.
33. Millones Abanto J, Vega-Gonzales E. Papanicolaou and Visual Inspection With Acetic Acid in the Detection of Intraepithelial Injuries of High Grade of the Cervix. 2017;2(2):8–13. Available from:  
<http://ojs.revistamaternofetal.com/index.php/RISMF/article/view/29/29>
34. Jesús M, Flores R, Elías R, Roncal O, Javier P, Mejía N, et al. ARTÍCULO ORIGINAL Utilidad de la citología e inspección visual con ácido acético en la detección de lesiones neoplásicas de cuello uterino Centro Médico Oncomujer 2013-2014 . Usefulness of cytology and visual inspection with acetic acid in the detection o. 2017;15–8.
35. Rojas-Zumaran V, Moya-Salazar J. The ecologization of the Papanicolaou stain in the diagnosis of cervical cancer. *Rev Med Inst Mex Seguro Soc.* 2018;56(3):217–25.
36. Schlichte MJ, Guidry J. Clinical Medicine Current Cervical Carcinoma Screening Guidelines. *J Clin Med* [Internet]. 2012;4:918–32. Available from:  
[www.mdpi.com/journal/jcm](http://www.mdpi.com/journal/jcm)
37. Gray E, Butler HJ, Board R, Brennan PM, Chalmers AJ, Dawson T, et al. Health economic evaluation of a serum-based blood test for brain tumour diagnosis: Exploration of two clinical scenarios. *BMJ Open.* 2018;8(5).
38. Bellisola G, Sorio C. Infrared spectroscopy and microscopy in cancer research and diagnosis. *Am J Cancer Res.* 2012;2(1):1–21.
39. Bhat AI, Rao GP. Real-Time Polymerase Chain Reaction. In 2020. p. 347–56.
40. Bester R, Jooste AEC, Maree HJ, Burger JT. Real-time RT-PCR high-resolution melting curve analysis and multiplex RT-PCR to detect and differentiate grapevine leafroll-associated associated virus 3 variant groups I, II, III and VI. *Virology.* 2012;9.
41. Zlotogorski-Hurvitz A, Dekel BZ, Malonek D, Yahalom R, Vered M. FTIR-based spectrum of salivary exosomes coupled with computational-aided discriminating analysis in the diagnosis of oral cancer. *J Cancer Res Clin Oncol.* 2019 Mar 13;145(3):685–94.
42. Canfell K. Towards the global elimination of cervical cancer. *Papillomavirus Res* [Internet]. 2019;8(March):100170. Available from:  
<https://doi.org/10.1016/j.pvr.2019.100170>
43. Andree KB, Fernández-Tejedor M, Elandaloussi LM, Quijano-Scheggia S, Sampedro N, Garcés E, et al. Quantitative PCR coupled with melt curve analysis for detection of selected *Pseudo-nitzschia* spp. (Bacillariophyceae) from the

- northwestern Mediterranean Sea. *Appl Environ Microbiol*. 2011;77(5):1651–9.
44. Conte J, Potoczniak MJ, Tobe SS. Using synthetic oligonucleotides as standards in probe-based qPCR. *Biotechniques*. 2018;64(4):177–9.
  45. Cousins MM, Donnell D, Eshleman SH. Impact of mutation type and amplicon characteristics on genetic diversity measures generated using a high-resolution melting diversity assay. *J Mol Diagnostics* [Internet]. 2013;15(1):130–7. Available from: <http://dx.doi.org/10.1016/j.jmoldx.2012.08.008>
  46. Prado A. Evaluación de la técnica de análisis de fusión de alta resolución para la detección y genotipificación de los genogrupos humanos de sapovirus [Internet]. UNIVERSIDAD PERUANA CAYETANO HEREDIA; 2017. Available from: [http://repositorio.upch.edu.pe/bitstream/handle/upch/713/Evaluacion\\_PradoMantilla\\_Alexandra.pdf?sequence=1&isAllowed=y](http://repositorio.upch.edu.pe/bitstream/handle/upch/713/Evaluacion_PradoMantilla_Alexandra.pdf?sequence=1&isAllowed=y)
  47. Balan V, Mihai CT, Cojocaru FD, Uritu CM, Dodi G, Botezat D, et al. Vibrational spectroscopy fingerprinting in medicine: From molecular to clinical practice. *Materials (Basel)*. 2019;12(18):1–40.
  48. Butler HJ, Brennan PM, Cameron JM, Finlayson D, Hegarty MG, Jenkinson MD, et al. Development of high-throughput ATR-FTIR technology for rapid triage of brain cancer. *Nat Commun* [Internet]. 2019;10(1):1–9. Available from: <http://dx.doi.org/10.1038/s41467-019-12527-5>
  49. Downes A, Mouras R, Elfick A. Optical spectroscopy for noninvasive monitoring of stem cell differentiation. *J Biomed Biotechnol*. 2010;2010.
  50. Elliott DA, Nabavizadeh N, Seung SK, Hansen EK, Holland JM. Radiation therapy [Internet]. *Oral, Head and Neck Oncology and Reconstructive Surgery*. Elsevier Inc.; 2017. 268–290 p. Available from: <http://dx.doi.org/10.1016/B978-0-323-26568-3.00013-0>
  51. Macho S. METODOLOGÍAS ANALÍTICAS BASADAS EN ESPECTROSCOPIA DE INFRARROJO Y CALIBRACIÓN MULTIVARIANTE. APLICACIÓN A LA INDUSTRIA PETROQUÍMICA. Departamento de Química Analítica y Química Orgánica. Universidad Rovira I Virgili [Internet]. UNIVERSITAT ROVIRA I VIRGILI; 2002. Available from: [https://www.tdx.cat/bitstream/handle/10803/8981/tesis\\_smacho.pdf;jsessionid=0063E543D30134F73B9672014241E0E7.tdx1?sequence=1](https://www.tdx.cat/bitstream/handle/10803/8981/tesis_smacho.pdf;jsessionid=0063E543D30134F73B9672014241E0E7.tdx1?sequence=1)
  52. Purandare NC, Trevisan J, Patel II, Gajjar K, Mitchell AL, Theophilou G, et al. Exploiting biospectroscopy as a novel screening tool for cervical cancer: towards a framework to validate its accuracy in a routine clinical setting. *Bioanalysis*. 2013;5(21):2697–711.
  53. Olea O. CARACTERIZACION POR FTIR Y TECNICAS ANALITICAS NUCLEARES DE PELICULAS DE CNx ELABORADAS POR ABLACION LASER. Universidad Autónoma del Estado de México; 2003.
  54. Ghomi M, Letellier R, Taillandier E. Particular behavior of the adenine and guanine

ring-breathing modes upon the DNA conformational transitions. *Biochimie*. 1988;70(6):841–6.

55. Stuart BH. *Infrared Spectroscopy: Fundamentals and Applications*. Vol. 8, *Infrared Spectroscopy: Fundamentals and Applications*. 2005. 1–224 p.
56. Khanmohammadi M, Garmarudi AB. Infrared spectroscopy provides a green analytical chemistry tool for direct diagnosis of cancer. *TrAC - Trends Anal Chem* [Internet]. 2011;30(6):864–74. Available from: <http://dx.doi.org/10.1016/j.trac.2011.02.009>
57. Piqué TM, Vázquez A. Concreto y cemento: Investigación y desarrollo. *Concreto y Cem Investig y Desarro* [Internet]. 2012;3(2):62–71. Available from: [http://www.scielo.org.mx/scielo.php?script=sci\\_arttext&pid=S2007-30112012000100004&lng=es&nrm=iso&tlng=es](http://www.scielo.org.mx/scielo.php?script=sci_arttext&pid=S2007-30112012000100004&lng=es&nrm=iso&tlng=es)
58. Sahu RK, Argov S, Salman A, Huleihel M, Grossman N, Hammody Z, et al. Characteristic absorbance of nucleic acids in the Mid-IR region as possible common biomarkers for diagnosis of malignancy. *Technol Cancer Res Treat*. 2004;3(6):629–38.
59. Ghimire H, Venkataramani M, Bian Z, Liu Y, Perera AGU. ATR-FTIR spectral discrimination between normal and tumorous mouse models of lymphoma and melanoma from serum samples. *Sci Rep* [Internet]. 2017;7(1):1–9. Available from: <http://dx.doi.org/10.1038/s41598-017-17027-4>
60. Roy S, Perez-Guaita D, Bowden S, Heraud P, Wood BR. Spectroscopy goes viral: Diagnosis of hepatitis B and C virus infection from human sera using ATR-FTIR spectroscopy. *Clin Spectrosc* [Internet]. 2019;1(December 2019):100001. Available from: <https://doi.org/10.1016/j.clispe.2020.100001>
61. Tipos de papilomavirus humanos y sus asociaciones con otras enfermedades [Internet]. [cited 2020 Aug 14]. Available from: <https://www.elsevier.com/es-es/connect/medicina/tipos-de-papilomavirus-humanos-y-sus-asociaciones-con-otras-enfermedades>
62. de Sanjosé S, Brotons M, Pavón MA. The natural history of human papillomavirus infection. *Best Pract Res Clin Obstet Gynaecol*. 2018;47:2–13.
63. Salazar C. GENOTIPIFICACIÓN DE 23 CEPAS DE HPV EN MUJERES DE 25 A 65 AÑOS QUE ACUDIERON AL HOSPITAL GINECO OBSTÉTRICO PEDIÁTRICO DE NUEVA AURORA LUZ ELENA ARISMENDI (HGOPNA) DURANTE EL PERIODO DE ENERO A DICIEMBRE DE 2018. PONTIFICIA UNIVERSIDAD CATÓLICA DEL ECUADOR; 2020.
64. Santos-López G, Márquez-Domínguez L, Reyes-Leyva J, Vallejo-Ruiz V. Temas de actualidad Aspectos generales de la estructura, la clasificación y la replicación del virus del papiloma humano [Internet]. [cited 2020 Aug 14]. Available from: <http://viralzone.expasy.org/>.
65. Biología del Virus del Papiloma Humano y técnicas de diagnóstico | Medicina

Universitaria [Internet]. [cited 2020 Aug 14]. Available from:  
<https://www.elsevier.es/es-revista-medicina-universitaria-304-articulo-biologia-del-virus-del-papiloma-X1665579610901659>

66. Perez Jiménez JM. Detección y Genotipificación del Virus Papiloma Humano (VPH) en población masculina del departamento de Sucre [Internet]. Universidad de Sucre; 2017. Available from: <http://unisucre-repositorio.metabiblioteca.org/handle/001/579>
67. Guan J, Bywaters SM, Brendle SA, Lee H, Ashley RE, Christensen ND, et al. The U4 Antibody Epitope on Human Papillomavirus 16 Identified by Cryo-electron Microscopy. *J Virol*. 2015;89(23):12108–17.
68. González F, Carbonell Z, Vergara C, Ochoa D. PREVALENCIA Y CARACTERIZACIÓN GENOTÍPICA DEL VIRUS DEL PAPILOMA HUMANO EN ALTERACIONES POTENCIALMENTE MALIGNAS Y CÁNCER ORAL EN CARTAGENA. ESTUDIO MULTICENTRO. [Internet]. Vol. 1. UNIVERSIDAD DE CARTAGENA FACULTAD; 2017. Available from: <https://pdfs.semanticscholar.org/81d2/788065860c038457934f723aaf004a71a99f.pdf>
69. Alvarez Paredes L. Caracterización de la infección cervical por el virus papiloma humano. Aplicación de nuevas técnicas de microbiología molecular en el estudio de la infección por el genotipo 16 [Internet]. UNIVERSITAS MIGUEL HERNANDEZ DE ELCHE; 2017. Available from: <https://dialnet.unirioja.es/servlet/tesis?codigo=136350>
70. Villafuerte Reinante J, Hernández Guerra Y, Ayala Reina ZE, Naranjo Hernández L, González Alonso JÁ, Brito Méndez M. Aspectos bioquímicos y factores de riesgo asociados con el cáncer cervicouterino. *Rev Finlay* [Internet]. 2019;9(2):138–46. Available from: <http://www.revfinlay.sld.cu/index.php/finlay/article/view/635>
71. De Villiers EM, Fauquet C, Broker TR, Bernard HU, Zur Hausen H. Classification of papillomaviruses. *Virology* [Internet]. 2004;324(1):17–27. Available from: <https://pubmed.ncbi.nlm.nih.gov/15183049/>
72. Trujillo E, Morales N, Buitrago O, Posso H, Bravo MM. Distribución de los genotipos del virus del papiloma humano en mujeres de Bogotá con anomalías en la citología cervicouterina. *Rev Colomb Cancerol* [Internet]. 2016;20(1):3–9. Available from: <https://www.elsevier.es/es-revista-revista-colombiana-cancerologia-361-articulo-distribucion-genotipos-del-virus-del-S0123901515000955>
73. Lara Peñaranda R. “Estudio de la profundidad de conización mediante LLETZ y la persistencia de lesión precursora de cáncer de cérvix y de infección por VPH postconización” [Internet]. Universidad Católica San Antonio; 2020. Available from: <http://repositorio.ucam.edu/handle/10952/4487>
74. Diaz N. DETERMINACIÓN Y ANÁLISIS DE LA INTEGRACIÓN DEL VIRUS DEL PAPILOMA HUMANO 16 EN EL GENOMA DE PACIENTES DIAGNOSTICADOS CON CÁNCER Y SU POSIBLE RELACIÓN CON LA ETIOLOGÍA DE LA ENFERMEDAD. UNIVERSIDAD ICESI FACULTAD; 2016.

75. Lozano L. Diagnóstico de los Carcinomas Orofaringeos Relacionados con el Virus del Papiloma Humano (VPH): Detección Viral mediante Técnicas Comerciales de uso Clínico y Análisis de su Valor Pronóstico [Internet]. All rights reserved. IJES. UNIVERSIDAD DE MURCIA; 2019. Available from: <http://hdl.handle.net/10201/72623>
76. Santos JMO, da Silva SP, Costa NR, Gil da Costa RM, Medeiros R. The role of microRNAs in the metastatic process of high-risk HPV-induced cancers. *Cancers (Basel)*. 2018;10(12):1–15.
77. Guerrero A, Guerrero M. MicroRNAs asociados al Cáncer de Cuello Uterino y sus lesiones precursoras: Una revisión sistemática MicroRNAs associated with Cervical Cancer and its precursor lesions: A systematic Review. *Rev Univ y Salud*. 2016;28(2):1–26.
78. Melo IM, Ribeiro E, Canevari R. Potential Diagnostic Techniques for Cervical Cancer Prevention - Review. *J Cancer Treat Diagnosis*. 2018;2(6):10–6.
79. Overbergh L, Vig S, Coun F, Mathieu C. Quantitative Polymerase Chain Reaction [Internet]. *Molecular Diagnostics: Third Edition*. Elsevier Ltd; 2017. 41–58 p. Available from: <http://dx.doi.org/10.1016/B978-0-12-802971-8.00004-3>
80. Varga A, James D. Real-time RT-PCR and SYBR Green I melting curve analysis for the identification of Plum pox virus strains C, EA, and W: Effect of amplicon size, melt rate, and dye translocation. *J Virol Methods*. 2006 Mar 1;132(1–2):146–53.
81. Maddocks S, Jenkins R. Quantitative PCR: Things to Consider. *Underst PCR* [Internet]. 2017;45–52. Available from: <https://www.sciencedirect.com/science/article/pii/B9780128026830000046>
82. Hernández L, García S, Nataren H, Espinoza L, Carmen L, Oliva C, et al. Near infrared spectroscopy (NIRS) in following the maturity of cultivation of sugar cane (*Saccharum spp.*). *Agro Product*. 12:107–13.
83. Ling S, Moebs W, Sanny J. 16.2 Plane Electromagnetic Waves - University Physics Volume 2 | OpenStax [Internet]. OpenStax. 2016 [cited 2020 Sep 10]. Available from: <https://openstax.org/books/university-physics-volume-2/pages/preface>
84. López Veloza JD. Estudio comparativo para la selección del detector de un prototipo de espectrofotómetro de luz visible [Internet]. UNIVERSIDAD CENTRAL DEL ECUADOR; 2019. Available from: <http://www.dspace.uce.edu.ec/handle/25000/18787>
85. Ventura JF. Desarrollo de métodos analíticos medioambientales sostenible por espectrofotometría FTIR [Internet]. Vol. 21, Univerisdad de Valencia. 2006. 1–49 p. Available from: <https://doi.org/10.1080/00102208008946937>
86. Tsakogiannis D, Papacharalampous M, Toska E, Kyriakopoulou Z, Dimitriou TG, Ruether IGA, et al. Duplex Real-time PCR assay and SYBR green I melting curve analysis for molecular identification of HPV genotypes 16, 18, 31, 35, 51 and 66. *Mol Cell Probes* [Internet]. 2015;29(1):13–8. Available from:

<http://dx.doi.org/10.1016/j.mcp.2014.09.003>

87. Tajmir-Riahi HA, N'Soukpoé-Kossi CN, Joly D. Structural analysis of protein-DNA and protein-RNA interactions by FTIR, UV-visible and CD spectroscopic methods. *Spectroscopy*. 2009;23(2):81–101.
88. Garip S, Bayari SH, Severcan M, Abbas S, Lednev IK, Severcan F. Structural effects of simvastatin on rat liver tissue: Fourier transform infrared and Raman microspectroscopic studies. *J Biomed Opt*. 2016;21(2):025008.
89. Palencia M. Functional transformation of Fourier-transform mid-infrared spectrum for improving spectral specificity by simple algorithm based on wavelet-like functions. *J Adv Res [Internet]*. 2018;14:53–62. Available from: <https://doi.org/10.1016/j.jare.2018.05.009>
90. De Oviedo U, Capel LJ. Máster En Ciencias Analíticas Y Bioanalíticas Estudio Del Grado De Madurez Y/O Conservación De Tomates Empleando Técnicas Espectroscópicas Moleculares. 2012;
91. Cascant M. Nuevos desafíos en espectroscopia vibracional. Universitat De Valencia; 2017.
92. Avila R. “USO DE LAS ESPECTROSCOPIAS ÓPTICAS Y MÉTODOS MULTIVARIANTES APLICADOS AL ANÁLISIS DE MUESTRAS BIOLÓGICAS.” UNIVERSIDAD AUTÓNOMA DE SAN LUÍS POTOSÍ; 2010.
93. Milosevic M. Internal reflection and ATR spectroscopy. Vol. 39, *Applied Spectroscopy Reviews*. 2004. 365–384 p.
94. MESA TÉLLEZ C. Aplicaciones De La Espectroscopía Infrarroja En El Análisis De Alimentos. *Trab Fin Grado [Internet]*. 2019;5. Available from: [https://idus.us.es/bitstream/handle/11441/91690/TÉLLEZ MESA%20CLARA.pdf?sequence=1&isAllowed=y](https://idus.us.es/bitstream/handle/11441/91690/TÉLLEZ%20CLARA.pdf?sequence=1&isAllowed=y)
95. Córscico B, Falomir Lockhart LJ, Franchini GR, Scaglia N. Análisis estructural y funcional de macromoléculas. Primera ed. Plata E de la U de La, editor. *Análisis estructural y funcional de macromoléculas*. La Plata; 2013. 1–413 p.
96. Mudunkotuwa IA, Minshid A AI, Grassian VH. ATR-FTIR spectroscopy as a tool to probe surface adsorption on nanoparticles at the liquid-solid interface in environmentally and biologically relevant media. *Analyst*. 2014;139(5):870–81.
97. Douglas A. Skoog. *Principios de Análisis Instrumental*. SEXTA EDIC. Cervantes S, editor. México, D.F; 2008. 1063 p.
98. Pretsch E, Bühlmann P, Badertscher M. Structure determination of organic compounds: Tables of spectral data. *Structure Determination of Organic Compounds: Tables of Spectral Data*. 2009. 1–433 p.
99. Field LD, Sternhell S, Kalman JR. *Organic Structures from Spectra*. Vol. 40, *Angewandte Chemie International Edition*. 2001. 9823 p.

100. McHale J. *Molecular Spectroscopy, Second Edition* [Internet]. Second edi. McHale J, editor. *Molecular Spectroscopy, Second Edition*. London: CRC press Taylor & Francis group; 2017. 1–477 p. Available from: <http://www.taylorandfrancis.com>
101. Yong-Cheng N. Interpretation of infrared spectra. In: Yong-Cheng N, editor. *Interpretation of organic spectra*. first edit. Asia: John Wiley & Sons; 2011. p. 412.
102. Lee LC, Jemain AA. Predictive modelling of colossal ATR-FTIR spectral data using PLS-DA: Empirical differences between PLS1-DA and PLS2-DA algorithms. *Analyst*. 2019;144(8):2670–8.
103. Theophilou G, Lima KMG, Martin-Hirsch PL, Stringfellow HF, Martin FL. ATR-FTIR spectroscopy coupled with chemometric analysis discriminates normal, borderline and malignant ovarian tissue: Classifying subtypes of human cancer. *Analyst*. 2016 Jan 21;141(2):585–94.
104. Anal AS, Basadas T, Espectroscopia EN, Petroqu LAI. *Metodologías analíticas basadas en espectroscopia de infrarrojo y calibración multivariante. aplicación a la industria petroquímica*. 2002;
105. Zontov Y V., Rodionova OY, Kucheryavskiy S V., Pomerantsev AL. PLS-DA – A MATLAB GUI tool for hard and soft approaches to partial least squares discriminant analysis. *Chemom Intell Lab Syst* [Internet]. 2020;203(March):104064. Available from: <https://doi.org/10.1016/j.chemolab.2020.104064>
106. Zhang SU. Classifying thermal degradation of polylactic acid by using machine learning algorithms trained on fourier transform infrared spectroscopy data. *Appl Sci*. 2020;10(21):1–13.
107. Ochoa Sosa MP, Polo Rivero KE. PREVALENCIA DE GENOTIPOS DE VIRUS DEL PAPILOMA HUMANO EN MUJERES DE LA POBLACIÓN ESTUDIANTIL DE LA UNIVERSIDAD SIMÓN BOLÍVAR DURANTE EL PERIODO 2017-1. Universidad Simón Bolívar; 2017.
108. Mathlouthi M, Seuvre AM. FTIR AND LASER-RAMAN SPECTRA OF ADENINE AND ADENOSINE. *Carbohydr Res*. 1984;131:1–15.
109. Mathlouthi M, Seuvre AM, L. Koenig J. F.t.-i.r. and laser-Raman spectra of Guanine and Guanosine. *Carbohydr Res*. 1984;134(1):23–38.
110. Mathlouthi M, Seuvre AM, Koenig JL. F.T.-I.R. and laser-raman spectra of cytosine and cytidine. *Carbohydr Res*. 1986;146(1):1–13.
111. Talari ACS, Martinez MAG, Movasaghi Z, Rehman S, Rehman IU. Advances in Fourier transform infrared (FTIR) spectroscopy of biological tissues. *Appl Spectrosc Rev*. 2017;52(5):456–506.
112. Mathlouthi M, Seuvre AM KJ. F.T.-I.R. AND LASER-RAMAN SPECTRA OF THYMINE AND THYMIDINE. *Carbohydr Res*. 1984;134:23–38.
113. Movasaghi Z, Rehman S, Rehman IU. Fourier transform infrared (FTIR)

- spectroscopy of biological tissues. *Appl Spectrosc Rev.* 2008;43(2):134–79.
114. Seuvre AM, Mathlouthi M. F.T.-I.R. spectra of oligo- and poly-nucleotides. *Carbohydr Res* [Internet]. 1987 [cited 2020 Aug 3];83–103. Available from: [https://www.researchgate.net/publication/19727928\\_FT-IR\\_spectra\\_of\\_oligo\\_and\\_poly-nucleotides](https://www.researchgate.net/publication/19727928_FT-IR_spectra_of_oligo_and_poly-nucleotides)
  115. Shimanouchi T, Tsuboi M, Kyogoku Y. *Infrared Spectra of Nucleic Acids and Related Compounds.* 2007;VII:435–98.
  116. Etzion Y, Linker R, Cogan U, Shmulevich I. Determination of protein concentration in raw milk by mid-infrared fourier transform infrared/attenuated total reflectance spectroscopy. *J Dairy Sci* [Internet]. 2004;87(9):2779–88. Available from: [http://dx.doi.org/10.3168/jds.S0022-0302\(04\)73405-0](http://dx.doi.org/10.3168/jds.S0022-0302(04)73405-0)
  117. Kong J, Yu S. Fourier transform infrared spectroscopic analysis of protein secondary structures. *Acta Biochim Biophys Sin (Shanghai).* 2007;39(8):549–59.
  118. Parachalil DR, Bruno C, Bonnier F, Blasco H, Chourpa I, Baker MJ, et al. Analysis of bodily fluids using vibrational spectroscopy: A direct comparison of Raman scattering and infrared absorption techniques for the case of glucose in blood serum. *Analyst.* 2019;144(10):3334–46.
  119. Pavia D, Lampman G, George K, Vyvyan J. *Introduction to Spectroscopy.* Fifth Edit. MPS Limited, editor. Vol. 28, American Journal of Physics. Bellingham, Washington: Cengage Learning WCN.; 2015. 786 p.
  120. Baker MJ, Hussain SR, Lovergne L, Untereiner V, Hughes C, Lukaszewski RA, et al. Developing and understanding biofluid vibrational spectroscopy: A critical review. *Chem Soc Rev.* 2016;45(7):1803–18.
  121. Komal Kumar J, Devi Prasad AG. Fourier transform infrared spectroscopy an advanced technique for identification of biomolecules. *Drug Invent Today.* 2012;4(12):616–8.
  122. Stuart BH. *Infrared Spectroscopy of Biological Applications: An Overview.* *Encycl Anal Chem.* 2012;
  123. Han Y, Han L, Yao Y, Li Y, Liu X. Key factors in FTIR spectroscopic analysis of DNA: The sampling technique, pretreatment temperature and sample concentration. *Anal Methods.* 2018;10(21):2436–43.
  124. Dovbeshko GI, Gridina NY, Kruglova EB, Pashchuk OP. FTIR spectroscopy studies of nucleic acid damage. *Talanta.* 2000;53(1):233–46.
  125. El-Mahdaoui L, Neault JF, Tajmir-Riahi HA. Carbohydrate-nucleotide interaction. The effects of mono- and disaccharides on the solution structure of AMP, dAMP, ATP, GMP, dGMP, and GTP studied by FTIR difference spectroscopy. *J Inorg Biochem.* 1997;65(2):123–31.
  126. Kotanen CN, Moussy FG, Carrara S, Guiseppi-elie A. *Encyclopedia of Biophysics*

[Internet]. Encyclopedia of Biophysics. Springer, Berlin, Heidelberg; 2013. Available from: [https://link.springer.com/referenceworkentry/10.1007%2F978-3-642-16712-6\\_112](https://link.springer.com/referenceworkentry/10.1007%2F978-3-642-16712-6_112)

127. Dovbeshko GI, Chegel VI, Gridina NY, Repnytska OP, Shirshov YM, Tryndiak VP, et al. Surface enhanced IR absorption of nucleic acids from tumor cells: FTIR reflectance study. *Biopolym - Biospectroscopy Sect.* 2002;67(6):470–86.
128. Missailidis S, Hester RE. FTIR spectra of deoxyoligonucleotide-nogalamycin complexes. *Biospectroscopy [Internet]*. 1995 Jan 1 [cited 2020 Aug 3];1(2):91–9. Available from: <http://doi.wiley.com/10.1002/bspy.350010202>
129. Chiavarino B, Crestoni ME, Fornarini S, Lanucara F, Lemaire J, Maitre P, et al. Infrared spectroscopy of isolated nucleotides. 1. The cyclic 3',5'-adenosine monophosphate anion. *Int J Mass Spectrom.* 2008 Mar 1;270(3):111–7.
130. Hollas M. MODERN SPECTROSCOPY Fourth Edition. Fourth Edi. John Wiley & Sons, Ltd. England; 2004. 452 p.
131. Tiernan H, Byrne B, Kazarian SG. ATR-FTIR spectroscopy and spectroscopic imaging for the analysis of biopharmaceuticals. *Spectrochim Acta A Mol Biomol Spectrosc [Internet]*. 2020 Jun 22 [cited 2020 Aug 4];241:118636. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/32610215>
132. Santamaria R, Charro E, Zacarías A, Castro M. Vibrational spectra of nucleic acid bases and their Watson-Crick pair complexes. *J Comput Chem.* 1999;20(5):511–30.
133. Taillandier E, Liquier J. Infrared spectroscopy of DNA. *Methods Enzymol.* 1992;211(C):307–35.
134. Taillandier E, Liquier J. *Vibrational Spectroscopy of Nucleic Acids. Handb Vib Spectrosc.* 2006;
135. Lucio Gutiérrez JR. *Aplicación de Métodos Quimiométricos para la Caracterización y Control de Calidad de Plantas Medicinales.* Universitat Autònoma de Barcelona; 2012.
136. Montañó D, Vargas J. ESTUDIO SOBRE LA UTILIZACION DE ESPECTROSCOPIA INFRARROJO PARA MEDIR LA CONCENTRACIÓN DE GLUCOSA EN SANGRE. UNIVERSIDAD AUTONOMA DE OCCIDENTE; 2009.
137. Melo A A, Roa E I, Montenegro H S, Capurro V I, Roa S JC. Estudio comparativo de detección del virus papiloma humano (VPH) en muestras citológicas y biopsias de cuello uterino. *Rev Med Chil.* 2005;133(6):639–44.
138. Liu P, Lu L, Xu M, Zhong H, Jia R, Su L, et al. A novel multiplex PCR for virus detection by melting curve analysis. *J Virol Methods.* 2018 Dec 1;262:56–60.
139. Munoz M, Camargo M, Soto-De Leon SC, Sanchez R, Parra D, Pineda AC, et al. Human Papillomavirus Detection from Human Immunodeficiency Virus-Infected Colombian Women's Paired Urine and Cervical Samples. *PLoS One.* 2013;8(2).

140. Sultani M, Azad TM, Eshragian M, Shadab A, Naseri M, Eilami O, et al. Multiplex SYBR green real-time PCR assay for detection of respiratory viruses. *Jundishapur J Microbiol* [Internet]. 2015 Aug 1 [cited 2020 Nov 20];8(8). Available from: <https://sites.kowsarpub.com/ijm/articles/59885.html>
141. Gudnason H, Dufva M, Bang DD, Wolff A. Comparison of multiple DNA dyes for real-time PCR: Effects of dye concentration and sequence composition on DNA amplification and melting temperature. *Nucleic Acids Res* [Internet]. 2007 Oct [cited 2020 Nov 20];35(19):e127. Available from: </pmc/articles/PMC2095797/?report=abstract>
142. Mamedov TG, Pienaar E, Whitney SE, TerMaat JR, Carvill G, Goliath R, et al. A fundamental study of the PCR amplification of GC-rich DNA templates. *Comput Biol Chem*. 2008;32(6):452–7.
143. Keatley S, Botero A, Fosu-nyarko J, Pallant L, Northover A, Thompson RCA. International Journal for Parasitology : Parasites and Wildlife Species-level identification of trypanosomes infecting Australian wildlife by High-Resolution Melting - Real Time Quantitative Polymerase Chain Reaction ( HRM-qPCR ). *Int J Parasitol Parasites Wildl* [Internet]. 2020;13(August):261–8. Available from: <https://doi.org/10.1016/j.ijppaw.2020.11.003>
144. Smith CJ, Osborn AM. Advantages and limitations of quantitative PCR (Q-PCR)-based approaches in microbial ecology. *FEMS Microbiol Ecol*. 2009;67(1):6–20.
145. Burk RD, Harari A, Chen Z. Human papillomavirus genome variants. *Virology* [Internet]. 2013 Oct [cited 2020 Aug 4];445(1–2):232–43. Available from: </pmc/articles/PMC3979972/?report=abstract>
146. Albawardi A, Quddus MR, Al Awar S, Almarzooqi S. Frequency of rare and multi viral high-risk HPV types infection in cervical high grade squamous intraepithelial lesions in a non-native dominant middle eastern country: A polymerase chain reaction-based pilot study. *Diagn Pathol*. 2018;13(1):1–8.
147. Zapata S, Mosquera D, Mejía L, Cruz L, Sánchez S, García M, et al. Estudios sobre el virus del papiloma humano en el Ecuador , parte II : memorias del simposio sobre el VPH y cáncer cervical , PUCE 2018. *Rev científica Digit INSPILIP*. 2019;3(1):1–15.
148. Tazreiter M, Christian P, Schennach R, Grießer T, Coclite AM. Simple method for the quantitative analysis of thin copolymer films on substrates by infrared spectroscopy using direct calibration. *Anal Methods*. 2017;9(36):5266–73.
149. Mann D, Höweler U, Kötting C, Gerwert K. Elucidation of Single Hydrogen Bonds in GTPases via Experimental and Theoretical Infrared Spectroscopy. *Biophys J*. 2017 Jan 10;112(1):66–77.
150. Paraskevaidi M, Morais CLM, Lima KMG, Ashton KM, Stringfellow HF, Martin-Hirsch PL, et al. Potential of mid-infrared spectroscopy as a non-invasive diagnostic test in urine for endometrial or ovarian cancer. *Analyst*. 2018;143(13):3156–63.

151. Pereira Viana MR, Martins Alves Melo I, Pupin B, Raniero LJ, de Azevedo Canevari R. Molecular detection of HPV and FT-IR spectroscopy analysis in women with normal cervical cytology. *Photodiagnosis Photodyn Ther* [Internet]. 2020;29(November 2019):101592. Available from: <https://doi.org/10.1016/j.pdpdt.2019.101592>
152. Neves ACO, Silva PP, Morais CLM, Miranda CG, Crispim JCO, Lima KMG. ATR-FTIR and multivariate analysis as a screening tool for cervical cancer in women from northeast Brazil: A biospectroscopic approach. *RSC Adv.* 2016;6(102):99648–55.
153. Pereira Viana MR, Martins Alves Melo I, Pupin B, Raniero LJ, de Azevedo Canevari R. Molecular detection of HPV and FT-IR spectroscopy analysis in women with normal cervical cytology. *Photodiagnosis Photodyn Ther.* 2020;29(November 2019):101592.
154. Ildiz GO, Bayari S, Karadag A, Kaygisiz E. Complementary Diagnosis Tool for Autism Spectrum Disorder in Children and Adolescents. *Molecules.* 2020;25:2079–91.
155. Gautam R, Vanga S, Ariese F, Umopathy S. Review of multidimensional data processing approaches for Raman and infrared spectroscopy. *EPJ Tech Instrum* [Internet]. 2015;2(1). Available from: <http://dx.doi.org/10.1140/epjti/s40485-015-0018-6>
156. Tafintsev D. *Multivariate Classification Methods for Spectroscopic Data with Multiple Class Structure.* Norwegian University of Life Sciences NMBU; 2016.

# A. Anexos

## ANEXO 1. CONSENTIMIENTO INFORMADO

### DETECCIÓN, DISCRIMINACIÓN Y CUANTIFICACION DE PEQUEÑAS SECUENCIAS DE NUCLEÓTIDOS POR FTIR EN LA IDENTIFICACION MOLECULAR DEL VIRUS DEL PAPILOMA HUMANO.

Yo \_\_\_\_\_, mayor de edad, he sido informado que LA UNIDAD DE INVESTIGACION CIENTIFICA DESARROLLO E INNOVACION EN GENETICA Y BIOLOGIA MOLECULAR de la Universidad Simón Bolívar realizará el estudio “DETECCIÓN, DISCRIMINACIÓN Y CUANTIFICACION DE PEQUEÑAS SECUENCIAS DE NUCLEÓTIDOS POR FTIR EN LA IDENTIFICACION MOLECULAR DEL VIRUS DEL PAPILOMA HUMANO”. Se me ha solicitado participar voluntariamente como sujeto de estudio.

#### OBJETIVO Y PROPÓSITO DEL ESTUDIO

El objetivo del proyecto es la identificación molecular de diferentes genotipos del virus de papiloma humano (VPH: 16, 18, 31, 35, 51 y 66), mediante modelos estadísticos. Los modelos estadísticos serán generados a partir de espectros de FTIR de amplificaciones de ADN de controles positivos de la reacción en cadena de la polimerasa o en inglés Polymerase Chain Reaction (PCR) y sus controles negativos. El modelo se evaluará a través de muestras de diferentes VPHs y su comparación con técnicas convencionales de biología molecular como PCR en tiempo real.

YO HE SIDO INFORMADO SOBRE LOS OBJETIVOS, PROPÓSITO, JUSTIFICACIÓN, METODOLOGÍA, **RIESGOS, Y BENEFICIOS DEL ESTUDIO.** En este estudio serán seleccionados 20 pacientes. Sobre la competencia, formación integral y calidad de los investigadores es responsable la Universidad Simón Bolívar. Los resultados del estudio son confidenciales y serán informados y explicados de forma grupal y confidencial al grupo objeto de estudio por parte del Dr. **ANTONIO ACOSTA HOYOS** investigador principal.

Los datos no serán utilizados con otra finalidad distinta a la de esta investigación.

**REQUERIMIENTOS.** Yo, en pleno uso de mis facultades mentales, libre y consciente, estoy de acuerdo en participar en este estudio y entiendo que éste requiere de mi lo siguiente: Autorizar la extracción del material genético de mi biopsia y mi historia clínica.

**RIESGOS DE PARTICIPACIÓN.** Para garantizar la confiabilidad de la información suministrada, los resultados de las pruebas serán codificados y se darán a conocer en forma grupal más no individual en un seminario, con el propósito de hacer una autorreflexión, luego de haber recibido una serie de conferencias.

Tengo claro que no se me proveerá con ninguna compensación económica.

**BENEFICIOS PARA EL PARTICIPANTE:** El participante no tendrá ningún beneficio por participar en este proyecto.

#### YO ENTIENDO QUE

Mi participación es completamente voluntaria y que puedo rehusarme a responder cualquier pregunta si así lo deseo o puedo tomar libremente la decisión de finalizar mi participación en este monitoreo en cualquier momento, sin que ello represente perjuicios de índole legal con mi trabajo.

La información recolectada será tratada de manera confidencial y mis respuestas serán reunidas con las de otros participantes para obtener resultados grupales.

Puedo preguntar cualquier interrogante o duda que tenga antes, durante o después del estudio, al investigador principal **ANTONIO ACOSTA HOYOS** de la Universidad Simón Bolívar responsable del estudio, en LA UNIDAD DE INVESTIGACION CIENTIFICA DESARROLLO E INNOVACION EN GENETICA Y BIOLOGIA MOLECULAR ubicado en la carrera 59 Numero 59- 36 de la ciudad de Barranquilla, en los teléfonos 3444333 Ext. 249 – 248.

La firma del documento del consentimiento informado es requerida para todas las personas participantes en un estudio como éste. Los procedimientos alternativos principales incluyendo procedimientos experimentales en este estudio, me han sido explicados en un lenguaje claro que yo he podido entender. Los riesgos y molestias que pueden presentarse me han sido explicados claramente. Entiendo que como mi nombre no será vinculado con los resultados del estudio, el investigador principal y sus coinvestigadores no estarán en la posibilidad de informar a ninguna otra persona sobre los resultados míos de las pruebas. Los resultados de este estudio podrán ser divulgados y/o publicados en revistas científicas en forma grupal sin que se dé a conocer mi nombre. He leído este consentimiento, he entendido en qué consiste este estudio y también me fueron aclaradas las dudas al respecto, en consecuencia, voluntariamente acepto participar como sujeto de estudio en el monitoreo biológico **“DETECCIÓN, DISCRIMINACIÓN Y CUANTIFICACION DE PEQUEÑAS SECUENCIAS DE NUCLEÓTIDOS POR FTIR EN LA IDENTIFICACION MOLECULAR DEL VIRUS DEL PAPILOMA HUMANO.”**

Nombre del Participante

Firma del Participante

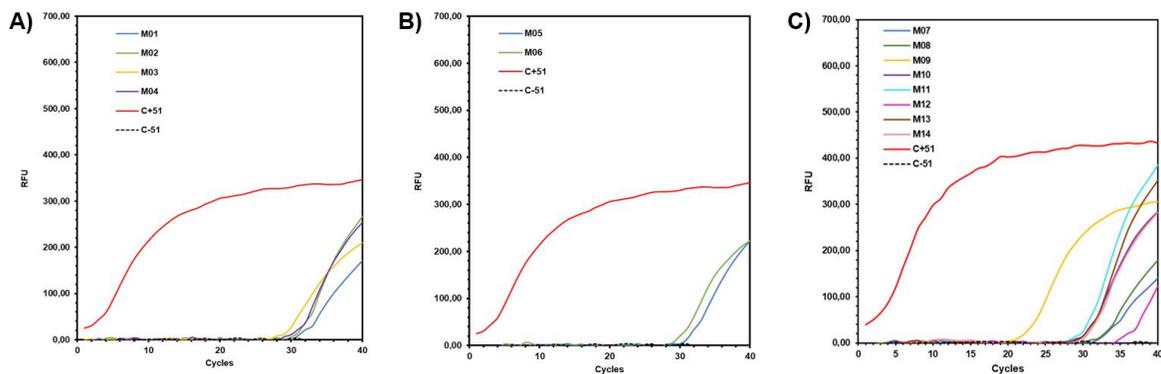
Nombre del Testigo

Firma del Testigo

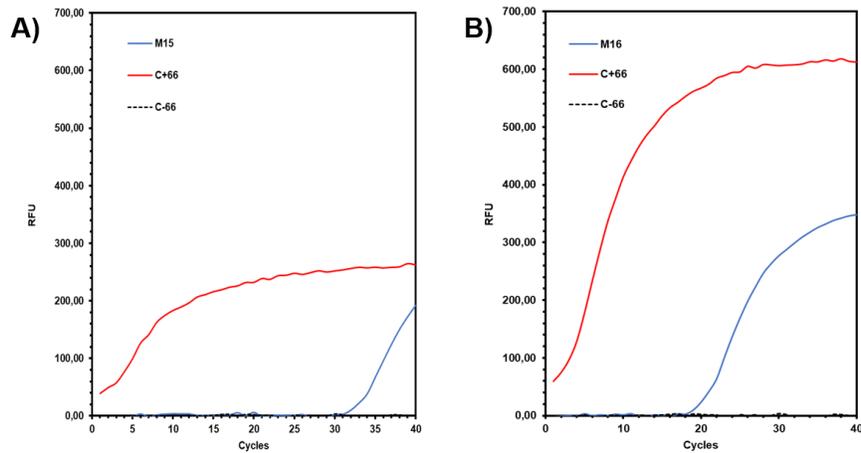
**ANTONIO ACOSTA HOYOS**

DIRECTOR UNIDAD DE INVESTIGACION CIENTÍFICA DESARROLLO E INNOVACION EN GENETICA Y BIOLOGIA MOLECULAR UNIVERSIDAD SIMÓN BOLÍVAR.

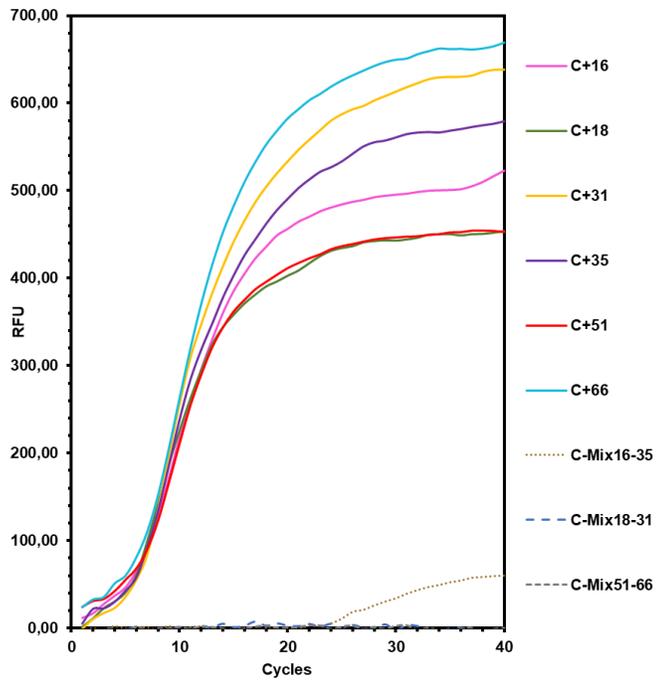
INVESTIGADOR PRINCIPAL DEL PROYECTO



**Anexo 2.** Curvas de amplificación del control positivo para VPH 51 y resultados de las muestras: **A)** M01-M04; **B)** M05 y M06; **C)** M07-M14, y control negativo en el sistema real-time PCR representado en unidades de fluorescencia relativas (RFU) frente al número de ciclos.



**Anexo 3.** Curvas de amplificación del control positivo para VPH 66 y resultados de las muestras: **A)** M15; **B)** M16, y control negativo en el sistema real-time PCR representado en unidades de fluorescencia relativas (RFU) frente al número de ciclos.



**Anexo 4.** Curvas de amplificación de los controles positivos para los tipos de VPH 16, 18, 31, 35 y 66 y resultados de controles negativos en el sistema real-time PCR representado en unidades de fluorescencia relativas (RFU) frente al número de ciclos.