



UNIVERSIDAD AUTÓNOMA DE SAN LUIS POTOSÍ
FACULTAD DE CIENCIAS



ESPECTROSCOPIA RAMAN PARA CARACTERIZAR E
IDENTIFICAR BIOMARCADORES
TESIS

PARA OBTENER EL GRADO DE:
DOCTORA EN CIENCIAS APLICADAS

PRESENTA:
M.C. KAREN HERNÁNDEZ VIDALES

DIRECTOR DE TESIS:
DR. FRANCISCO JAVIER GONZÁLEZ CONTRERAS

SAN LUIS POTOSÍ, SLP

SEPTIEMBRE 2021

DECLARACIÓN DE AUTORÍA Y ORIGINALIDAD DE LA TESIS

Yo, **Karen Hernández Vidales**, estudiante del Posgrado en Ciencias Aplicadas de la Facultad de Ciencias de la Universidad Autónoma de San Luis Potosí, como autora de la tesis **Espectroscopía Raman para Caracterizar e Identificar Biomarcadores** declaro que es una obra original, inédita, auténtica, personal, que se han citado las fuentes correspondientes y que en su ejecución se respetaron las disposiciones legales vigentes que protegen los derechos de autor y de propiedad intelectual e industrial. Las ideas, doctrinas, resultados y conclusiones a los que he llegado son de mi absoluta responsabilidad.



Espectroscopía Raman para Caracterizar e
Identificar Biomarcadores por Karen
Hernández Vidales se distribuye bajo una
**Licencia Creative Commons Atribución-
NoComercial-SinDerivadas 4.0
Internacional.**

RESUMEN

La espectroscopía Raman es una técnica mediante la cual se capta la luz que ha sido esparcida inelásticamente por un medio de interés, gracias a que esta luz trae consigo información a nivel molecular, es posible obtener un espectro único característico del material. Una de sus principales aplicaciones es la caracterización e identificación de biomoléculas, especialmente biomarcadores, ya que es posible diferenciarlos: ya sea que correspondan a estados óptimos de salud o a estados en los que se presenta alguna enfermedad. La espectroscopía Raman se proyecta como una técnica que puede ser usada clínicamente para mejorar los diagnósticos médicos de una forma rápida y no invasiva.

El cáncer requiere de diagnósticos certeros y rápidos, su detección temprana puede ser determinante para la sobrevivencia del paciente por lo que diversos biomarcadores están siendo investigados, dos de ellos la proteína p53 (en su estado natural y sus mutaciones) y el aminoácido L-Asparagina; los cuales, particularmente en esta tesis, se han estudiado mediante espectroscopía Raman. La proteína p53 en su estado natural tiene la capacidad de reparar el ADN de células dañadas o bien conducir las hacia la muerte antes de que el daño proliferara, sin embargo, cuando ocurren mutaciones hay mayor probabilidad de que se desarrollen tumores. Poder identificar y diferenciar entre la proteína p53 natural y sus mutantes con métodos de detección sencillos, puede significar un avance importante en la detección temprana del cáncer. Las células cancerígenas de leucemia linfoblástica aguda se alimentan del aminoácido L-Asparagina, los tratamientos anti-leucemia incluyen métodos para su eliminación, sin embargo, a la fecha no existe un método certero y sencillo que permita monitorear la biodisponibilidad de este aminoácido durante el tratamiento.

Mediante espectroscopía Raman, métodos de análisis multivariable y de aprendizaje automático, en este trabajo se determinó el límite de detección para identificar a la proteína p53 y se logró diferenciar entre la proteína p53 natural y tres de sus mutantes: R273H, E343A y L344A. Además, se probaron diferentes superficies con L-Asparagina para aplicar espectroscopía Raman mejorada por superficies y caracterizar el espectro del aminoácido.

ABSTRACT

Raman spectroscopy is a technique by which the light scattered inelastically is obtained and studied. This light brings information from the molecular level, making it possible to get a unique spectrum for the material under examination. One of the applications of Raman spectroscopy is the characterization and identification of biomolecules, particularly biomarkers, since it is feasible to differentiate among them; either they correspond to optimal states of health or to conditions in which some disease occurs. Raman spectroscopy is projected as a technique that can be used clinically to improve medical diagnosis quickly and noninvasively.

Cancer requires accurate and rapid diagnoses; its early detection can be decisive for the patient's survival. Various biomarkers are being investigated, two of them the p53 protein (wild type and mutants) and the amino acid L-Asparagine, particularly in this thesis through Raman spectroscopy. The p53 wild-type protein repairs the DNA in damaged cells or leads them to death before damage proliferates. However, when mutations occur, tumors are more likely to develop. Identifying and differentiating between wild-type p53 protein and its mutants with simple detection methods can be an essential step towards the early detection of cancer. Cancer cells of acute lymphoblastic leukemia feed on the amino acid L-Asparagine; anti-leukemia treatments include strategies for its elimination. However, no accurate and straightforward method allows monitoring the bioavailability of this amino acid during treatment.

Utilizing Raman spectroscopy, multivariate analysis, and machine learning, the detection limit for the p53 protein was determined in this work. It was possible to differentiate between the wild-type protein and three mutants: R273H, E343A, and L344A. Different surfaces were tested with L-Asparagine to apply surface-enhanced Raman spectroscopy and characterize the amino acid spectrum.

AGRADECIMIENTOS

Agradezco a la Universidad Autónoma de San Luis Potosí, a la Coordinación para la Innovación y la Aplicación de la Ciencia y la Tecnología y al Laboratorio Nacional de Ciencia y Tecnología de Terahertz; por haberme brindado todas las herramientas para desarrollar mi trabajo de doctorado. Gracias también al Consejo Nacional de Ciencia y Tecnología por el financiamiento del grado mediante la Beca 293704 y el apoyo recibido a través del proyecto No. 315926 de la Convocatoria Laboratorios Nacionales 2021.

Mi más sincera gratitud a mi asesor el Dr. Francisco Javier González y a mis sinodales: Dr. Ángel Gabriel Rodríguez, Dr. Ghebré Ramírez, Dr. Samuel Kolosovas y Dr. Guillermo Toriz; a quienes debo me hayan encaminado con su dedicación y experiencia hacia este resultado.

De igual manera agradezco al Dr. Edgar Guevara, a la Dra. Vanessa Olivares y al Dr. Javier Méndez, por sus valiosas aportaciones para el desarrollo de esta tesis. Gracias a los investigadores: Dr. Marco Atzori, Dra. Marcela Miranda y Dra. Alejandra Loyola, quienes me han permitido colaborar en sus trabajos y laboratorios.

Gracias a todos los profesores e investigadores que han sido parte de mi formación académica, a mis compañeros y amigos, por haber compartido conmigo estos años de fructífero estudio y trabajo.

Gracias a mis Papás, a mis Hermanas, a mis Hermanos y a mis Sobrinos, gracias por su cariño y su alegría; contar con ustedes ha sido fundamental para mí. Gracias por tanto, gracias por todo.

TABLA DE CONTENIDOS

Capítulo 1. INTRODUCCIÓN	1
Capítulo 2. ESPECTROSCOPIA RAMAN PARA MARCADORES DE ENFERMEDAD	5
2.1 Principios básicos de la espectroscopía Raman	5
2.1.1 Vibraciones Moleculares.....	6
2.1.2 Espectro Raman	7
2.2 Adquisición de espectros Raman	8
2.2.1 Elección del láser: longitud de onda, potencia y tiempo de exposición.....	9
2.2.2 Elección del objetivo de microscopio.....	10
2.3 Mejoramiento del Esparcimiento Raman por Superficies	11
2.3.1 Factor de Mejoramiento de SERS	13
2.4 Caracterización de biomoléculas mediante Espectroscopía Raman	15
2.4.1 Aminoácidos y proteínas.....	15
2.4.2 Lípidos	17
2.4.3 Carbohidratos	17
2.4.4 Ácidos nucleicos.....	18
2.5 Métodos de Análisis Multivariable aplicados en Espectroscopía Raman	19
2.5.1 Análisis de Componentes Principales: PCA.....	20
2.5.2 Regresión de mínimos cuadrados parciales: PLSR.....	21
2.5.3 Máquina de vectores de soporte: SVM	21
2.6 Simulaciones Moleculares	22
Capítulo 3. MARCADORES DE ENFERMEDAD	24
3.1 Marcador tumoral p53.....	24
3.1.1 Localización y Estructura de la Proteína p53	25
3.1.2 Funciones y Regulación de la Actividad de la Proteína p53.....	27
3.1.3 Mutaciones de p53	28
3.1.4 Uso de p53 en diagnóstico y en terapias	30
3.1.5 Detección de p53 mediante espectroscopía Raman.	31
3.2 Aminoácido L-asparagina.....	33
3.2.1 Estructura y localización fisiológica	33
3.2.2 L-Asparagina y su importancia en el control de la leucemia linfoblástica aguda	35
3.2.3 L-Asparagina y metástasis.....	35
3.2.4 Detección de L-Asparagina mediante Espectroscopía Raman.....	36

Capítulo 4. ESPECTROSCOPIA RAMAN DEL MARCADOR TUMORAL p53	37
4.1 Espectro Raman de las proteínas p53: tipo natural y mutantes 273, 343 y 344.	38
4.1.1 Muestras	38
4.1.2 Metodología para la adquisición de espectros Raman.	39
4.1.3 Preprocesamiento de espectros.	40
4.1.4 Asignación de vibraciones.....	42
4.2 Límite de detección de p53 natural mediante espectroscopía Raman.	45
4.2.1 Preparación de muestras de p53 a diferentes concentraciones.	46
4.2.2 Adquisición y preprocesamiento de espectros Raman.....	46
4.2.4 Regresión univariable.....	47
4.2.5 Regresión multivariable	49
4.2.6 Comparación entre los ajustes para la curva de calibración de concentraciones.....	50
Capítulo 5. CLASIFICACIÓN DE P53 MEDIANTE MÉTODOS DE ANÁLISIS MULTIVARIABLE Y APRENDIZAJE AUTOMÁTICO	51
5.1 Análisis de componentes principales.....	52
5.2 Resultados y discusión de PCA.....	53
5.3 Clasificación mediante SVM.....	58
5.4 Resultados y discusión del Clasificador SVM	60
Capítulo 6. DETECCIÓN DE L-ASPARAGINA MEDIANTE SERS	64
6.1 Muestras de L-Asparagina.....	65
6.2 Sustratos SERS.....	65
6.2.1 Sustrato SERS de partículas.....	65
6.2.2 Sustrato SERS de nanoestructuras triangulares sobre óxido de silicio.....	66
6.2.3 Sustrato SERS de nanoestructuras tipo bowtie de oro sobre sustrato de grafeno en óxido de silicio.	67
6.2.4 Sustrato SERS de nanoestructuras tipo bowtie de plata sobre óxido de silicio.	68
6.3 Espectroscopía Raman y SERS.....	69
6.3.1 Análisis de los Espectros Raman de referencia.....	70
6.3.2 SERS con superficie de partículas de oro.	72
6.3.3 SERS con superficie de triángulos sobre SiO ₂	74
6.3.4 SERS con superficie de estructuras de oro tipo bowtie sobre grafeno y SiO ₂ . 75	
6.3.5 SERS con superficie de bowtie de plata sobre SiO ₂	77
6.3.6 Discusión sobre el desempeño de las superficies SERS	78
6.3.7 Simulaciones y Asignación de vibraciones.....	80
Capítulo 7. CONCLUSIONES Y TRABAJO FUTURO	82

PRODUCTIVIDAD ACADÉMICA	84
REFERENCIAS	86
APÉNDICE	99
Regresiones polinomiales univariadas para límite de detección.....	99
PLSR para límite de detección.....	101
Análisis de Componentes Principales	103
Support Vector Machine.....	104
Cálculo de Factor de Mejoramiento SERS.....	107

Capítulo 1.

INTRODUCCIÓN

Uno de los grandes retos de la humanidad es mejorar el diagnóstico de enfermedades para atenderlas de forma temprana y precisa. Se ha evolucionado de la medicina cualitativa, en la que solo se observaban síntomas, a una medicina analítica, en la que se estudian parámetros cuantitativos para obtener diagnósticos certeros y entender los mecanismos con los que las patologías se desarrollan en el cuerpo humano, permitiendo también diseñar tratamientos de prevención y de cura.

En la actualidad, se llevan a cabo distintas pruebas para poder confirmar ciertos padecimientos. Estas pruebas tienen el objetivo de medir sustancias o identificar patrones específicos, llamados marcadores de enfermedad o biomarcadores. Un biomarcador se define como una característica o sustancia que puede ser medida y evaluada objetivamente como un indicador de procesos biológicos normales, patogénicos o como respuesta a una intervención terapéutica. Para que un biomarcador sea considerado como válido, debe cumplir con haber sido medido mediante un sistema analítico correctamente caracterizado y debe existir evidencia científica de su importancia fisiológica, farmacológica, toxicológica o clínica [1], [2]. Algunos ejemplos de biomarcadores son los niveles de glucosa o lípidos en fluidos; perfiles genómicos, metabolómicos o proteómicos; señales eléctricas cerebrales, patrones en imágenes y en general niveles de sustancias bioquímicas [1].

Existe una gran cantidad de técnicas para identificar y caracterizar distintos tipos de marcadores de enfermedad; tales como las pruebas de cromatografía, PCR y ELISA para identificar sustancias específicas; la histología para el estudio de tejidos; métodos para obtener imágenes como el ultrasonido, la resonancia magnética o los rayos X. Desafortunadamente muchos de estos métodos suelen ser laboriosos, tardados, costosos y requieren la interpretación de expertos, por lo que en ocasiones los diagnósticos resultan poco confiables por la subjetividad de la valoración.

Los tiempos actuales han hecho más que evidente la necesidad de métodos alternativos capaces de identificar y caracterizar marcadores de enfermedad de manera confiable, rápida y accesible a la población. La aplicación de la óptica es una alternativa para lograr este objetivo. La radiación electromagnética es ampliamente usada para el estudio de la materia, pues al interactuar con ésta, sufre cambios que pueden ser medidos y asociados con el material [3]. En este sentido, la luz puede usarse para el estudio de biomarcadores moleculares.

Una técnica prometedora es la espectroscopía Raman, mediante ésta se analiza la radiación que ha sido esparcida por un medio; la energía de esta radiación está relacionada con la vibración inducida en las moléculas, lo que permite obtener información a nivel molecular con alto grado de especificidad, ya que cada molécula tiene modos de vibración particulares [4]. Mediante espectroscopía Raman se pueden identificar marcadores de enfermedad moleculares pues se obtienen detalles específicos sobre la bioquímica de la muestra biológica.

La principal desventaja de la espectroscopía Raman es la baja intensidad de señal cuando hay muy poca cantidad de materia (baja concentración), más puede ser superada utilizando técnicas fotónicas para incrementar la dispersión. La más utilizada es llamada SERS (*Surface-Enhanced Raman Spectroscopy*); consiste en incrementar el campo eléctrico que excita a la molécula mediante la resonancia plasmónica de superficies metálicas periódicas, amplificando la dispersión Raman en factores de hasta 10^{10} órdenes de magnitud [5]. SERS combina la alta especificidad de Raman con la capacidad de detectar biomoléculas en concentraciones muy bajas (de hasta 10^{-15} M) e incluso moléculas individuales [6], [7].

La espectroscopía Raman y SERS son técnicas sumamente capaces de mejorar la detección de biomarcadores, pues además de su alta especificidad y sensibilidad: no son invasivas, las muestras requieren preparaciones sencillas, no requieren etiquetas adicionales durante los análisis, son rápidas y la instrumentación puede adaptarse para su portabilidad, son menos costosas que la mayoría de los estudios convencionales, pueden detectar más de un biomarcador a la vez y la interpretación de los resultados es objetiva gracias a la implementación de la quimiometría [5]–[11].

Mediante espectroscopía Raman y análisis multivariable se han logrado reconocer diferencias entre tejidos sanos y enfermos [8], [12]–[14]. Estas técnicas han permitido identificar bordes tumorales en cirugías cerebrales [15]. También ha sido posible caracterizar individualmente biomarcadores, identificar sus diferencias moleculares, hacer un seguimiento de sus cambios estructurales y diferenciar mutaciones [7], [16]–[18]. Estas características incrementan las posibilidades de que se implemente para mejorar el diagnóstico de enfermedades.

El cáncer es una enfermedad cuya detección es desafiante, en la mayoría de los casos no se presentan síntomas hasta etapas avanzadas; su diagnóstico conlleva grandes cantidades de análisis: laboriosos, costosos y cuya confiabilidad muchas veces recae en la experiencia de los oncólogos [19]–[21]. Optimizar las técnicas de detección de marcadores tumorales permitiría mejorar el diagnóstico y pronóstico, tanto en etapas tempranas como durante los tratamientos.

La proteína p53 es uno de los biomarcadores asociados al desarrollo de cáncer; su función más importante es impedir el crecimiento anormal de las células, por lo que es considerado como un gen supresor de tumores; su detección y cuantificación podría ser empleada para mejorar los diagnósticos. Sin embargo, aunque ha sido ampliamente estudiado, su detección requiere de elaboradas técnicas que no son accesibles en la mayoría de los centros de atención oncológica [22]. Otro biomarcador implicado en la progresión del cáncer es el aminoácido L-Asparagina: las células cancerígenas (linfoblastos) en la leucemia linfoblástica aguda requieren este aminoácido como alimento y su disponibilidad extracelular ha sido vinculada con el desarrollo de metástasis en otros tipos de cáncer [23], [24]. Las terapias anticánceres podrían mejorar al eliminar este aminoácido, por lo que su detección resulta importante.

Para este trabajo de tesis, la proteína p53 y el aminoácido L-Asparagina fueron estudiados mediante espectroscopía Raman y SERS respectivamente. El objetivo general es comprobar la utilidad y la capacidad de Raman para identificar estos biomarcadores; los objetivos específicos son:

- Caracterizar mediante espectroscopía Raman el marcador tumoral p53 tipo natural y tres tipos de sus mutantes: R273H, E343A y L344A.

- Determinar el límite de detección de p53 en espectroscopía Raman.
- Diferenciar mediante métodos de análisis multivariable y aprendizaje automático los tipos de proteína p53.
- Aplicar SERS con diferentes superficies para el aminoácido L-Asparagina.

Se espera que los resultados aquí obtenidos, permitan avanzar hacia la meta de que la espectroscopía Raman sea una técnica que se implemente cotidianamente para identificar marcadores de enfermedad y en particular los aquí presentados por sus implicaciones en el desarrollo del cáncer.

Capítulo 2.

ESPECTROSCOPIA RAMAN PARA MARCADORES DE ENFERMEDAD

2.1 PRINCIPIOS BÁSICOS DE LA ESPECTROSCOPIA RAMAN

Cuando una molécula específica recibe radiación, esta luz polariza (distorsiona) su nube de electrones, formando un estado de excitación momentáneo (estado virtual), al decaer de este estado excitado un fotón es dispersado, lo que significa que es reemitido con una dirección aleatoria. El fotón puede tener la misma energía que la radiación incidente, proceso elástico llamado esparcimiento Rayleigh, en el cual la momentánea distorsión de la nube de electrones no provoca un cambio en el núcleo y la molécula regresa a su estado original. El fotón dispersado también puede tener mayor o menor energía que el incidente, dándose un proceso inelástico llamado esparcimiento Raman, que sucede cuando en la interacción fotón-molécula, la polarización de la nube de electrones modifica momentáneamente la posición del núcleo y existe transferencia de energía [4]. Específicamente, se le llama dispersión Stokes cuando la energía de la radiación incidente E_i es mayor a la energía de la radiación dispersada E_d y, por el contrario, cuando la energía de la radiación incidente es menor a la de la radiación dispersada se conoce como dispersión anti-Stokes. El esquema de niveles de energía de la Figura 2.1 ilustra estos procesos.

La probabilidad de que la dispersión Raman ocurra es de solo un fotón en cada 10^8 fotones esparcidos, por lo que la intensidad del esparcimiento Rayleigh es mucho mayor [4], [25].

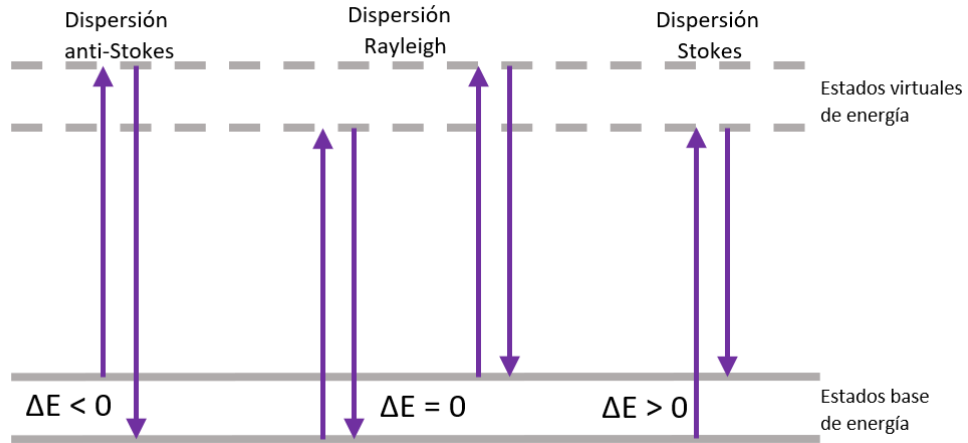


Figura 2.1 Diagrama de los procesos de dispersión Raman y dispersión Rayleigh. La diferencia de energía ΔE entre la radiación incidente y la radiación dispersada es la diferencia de energía entre dos niveles energéticos de la molécula.

2.1.1 Vibraciones Moleculares

La energía de una molécula (no lineal y de más de dos átomos) está cuantizada en cierto número de grados de libertad: tres grados traslacionales, tres grados rotacionales y $3N-6$ grados vibracionales [4]. En la figura 2.2 se modela una molécula de agua, el número de estados vibracionales posibles son 3 y están esquematizados.

La luz proveniente del esparcimiento Raman contiene información del material a nivel molecular, específicamente de la energía de sus estados vibracionales. Cuando la radiación interactúa con una molécula, promueve saltos de un estado vibracional a otro [4].

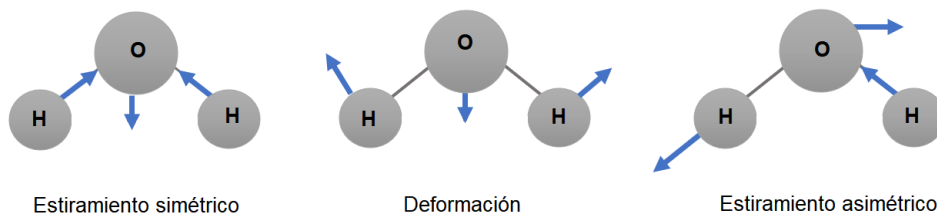


Figura 2.2. Modelo de la molécula de agua y sus tres modos de vibración.

Modelando la molécula como un sistema de masas y resortes, el modelo del oscilador armónico predice la energía de vibración de un enlace; para resortes fuertes y masas ligeras (enlaces fuertes y átomos pequeños) se tendrán frecuencias de oscilación altas y, para masas pesadas y resortes débiles se tendrán oscilaciones a frecuencias bajas. De esta forma, se puede interpretar la energía de la luz que proviene del esparcimiento Raman como la energía que corresponde a la vibración de un enlace [4].

2.1.2 Espectro Raman

La figura 2.3 ilustra un típico espectro Raman. El eje horizontal está dado en números de onda (k) también llamado corrimiento Raman, con unidades cm^{-1} , el cual es directamente proporcional al cambio en energía (ΔE) o a la frecuencia de vibración (ω) [4]:

$$k = \frac{2\pi}{\lambda} \quad (2.1),$$

$$\omega = k \cdot c \quad (2.2),$$

$$\Delta E = \hbar \cdot \omega \quad (2.3).$$

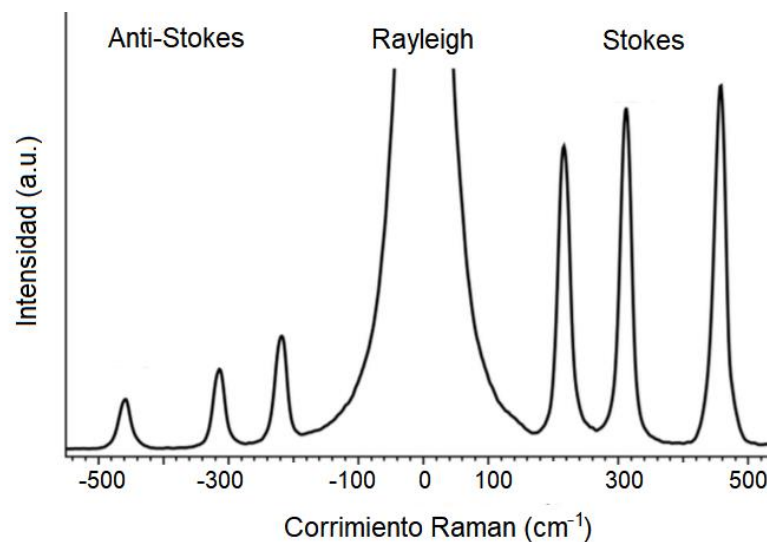


Figura 2.3. Espectro Raman Stokes y Anti-Stokes de Cloruro de Carbono (CCl_4). La intensidad de las bandas Raman depende de la cantidad de fotones esparcidos. En el centro la dispersión Rayleigh aparece con una intensidad máxima porque el esparcimiento elástico tiene mayor probabilidad de suceder que el no elástico [26].

Los picos o máximos son llamados bandas Raman o bandas de vibración, corresponden a una frecuencia de vibración específica de la molécula y son la diferencia de energía entre la luz esparcida y la radiación incidente (Ecuación 2.3).

El esparcimiento Stokes, donde $\Delta E > 0$, es más intenso que el esparcimiento anti-Stokes porque a temperatura ambiente la mayoría de las moléculas se encuentran en su estado base [4], [5].

El eje vertical del espectro Raman corresponde a la intensidad de las bandas Raman, que es una medida de la cantidad de fotones esparcidos. Se mide en unidades arbitrarias o cuentas y está dada por la expresión:

$$I = Kl\alpha^2\omega^4 \quad (2.4).$$

En esta ecuación: K es una constante de proporcionalidad en la que se encuentra la velocidad de la luz y la constante de Planck, l es la potencia de la fuente usada para la excitación, α es el tensor de polarizabilidad de la molécula y ω es la frecuencia de la fuente [3]. La cantidad de fotones esparcidos también depende de la cantidad de moléculas excitadas, por lo que la cantidad de materia -o concentración de la sustancia- modifica la intensidad de las bandas Raman.

La forma del arreglo atómico de una molécula es único, por lo tanto, su comportamiento vibracional al recibir radiación es característico, de tal forma que el número, la posición y la intensidad de las bandas Raman están relacionadas al tipo y a la forma general de la molécula, razón por la que el espectro es llamado huella digital [4], [27].

2.2 ADQUISICIÓN DE ESPECTROS RAMAN

La configuración óptica necesaria para adquirir el espectro Raman de un material se conoce como espectrómetro Raman. La figura 2.4 ilustra un sistema básico. Los sistemas modernos en general están formados por un láser como fuente de excitación; filtros para eliminar la dispersión Rayleigh y otras radiaciones, manteniendo así sólo la dispersión Raman; un medio para dividir la radiación en sus longitudes de onda constituyentes (rejilla de difracción o prisma); un detector

muy sensible (por ejemplo, un sensor CCD) y un dispositivo de cómputo para controlar todo el sistema, generar el espectro y permitir analizar la información. Este tipo de espectrómetros se han combinado con microscopios ópticos, lo que permite enfocar la luz en áreas de unas cuantas micras sobre la muestra y mejorar la resolución del análisis, algo sumamente útil cuando de muestras biológicas se trata.

En el caso específico de las muestras biológicas, es importante considerar los parámetros longitud de onda, intensidad del láser, objetivo de microscopio y tiempo de adquisición [9]. El objetivo de modificar cualquiera de estos parámetros es obtener la mejor calidad en el espectro: tener la menor cantidad de ruido y la mayor intensidad en las bandas Raman.

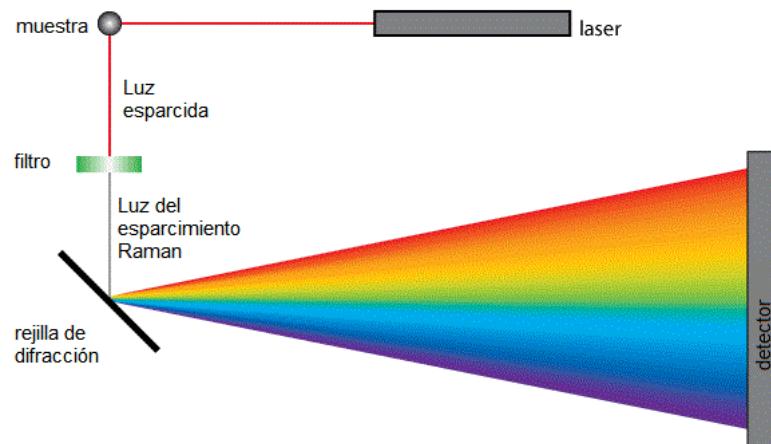


Figura 2.4. Configuración básica de un espectrómetro Raman [28].

2.2.1 Elección del láser: longitud de onda, potencia y tiempo de exposición.

La elección de longitud de onda λ tiene una relación importante con la intensidad de la señal, pues hace que ésta disminuya en $1/\lambda^4$. Sin embargo, al excitar biomoléculas con longitudes de onda cortas se presenta alta emisión de fluorescencia, la cual puede opacar la dispersión Raman. Lo más conveniente es utilizar longitudes de onda en regiones donde se tenga poca absorción de radiación y por lo tanto menor emisión de fluorescencia, por ejemplo, en el cercano infrarrojo [9]. La figura 2.6 muestra el perfil de la fluorescencia que se obtiene para el espectro visible.

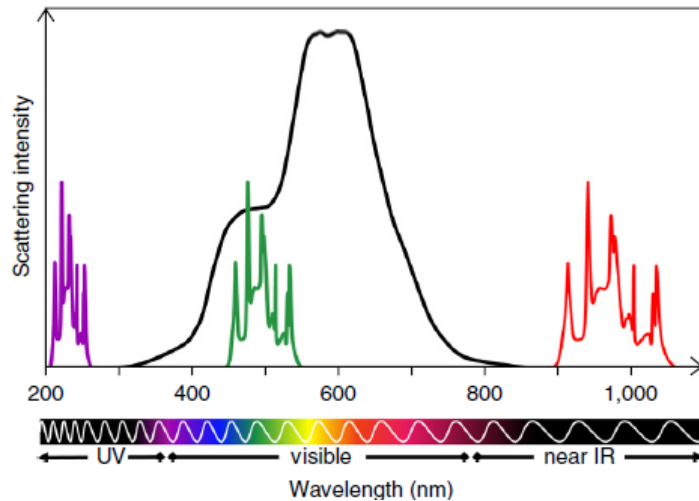


Figura 2.5. Perfil de emisión de fluorescencia en muestras biológicas [9].

Para lograr espectros con alta calidad también se requiere suficiente potencia de excitación, no es trivial solo elegir un láser altamente potente, sobre todo cuando se trata de biomoléculas. La alta intensidad de un láser puede provocar daño por fotoablación; disminuir drásticamente la intensidad puede producir muy baja señal en el sensor. La intensidad óptima es un parámetro que debe ajustarse experimentalmente para cada muestra y mantenerse en monitoreo constante, pues en caso de quemaduras, se observará una mancha oscura en la muestra o bien una prominente banda en 1500 cm^{-1} que corresponde a los enlaces del carbono [9]. Una alternativa para resolver el compromiso entre aumentar la intensidad para obtener mejor señal espectral o disminuirla por posibles daños, es cambiar el tiempo de exposición. Al disminuir el tiempo podemos evitar daños a la muestra sin comprometer la calidad del espectro.

2.2.2 Elección del objetivo de microscopio.

La resolución espacial de un espectrómetro Raman depende del área iluminada por el láser. En un microespectrómetro esta área está determinada por el límite de difracción de la luz, por lo tanto, depende de la longitud de onda del láser y de la apertura numérica del objetivo de microscopio [9], [27]. Para obtener señal Raman de regiones más amplias es conveniente el uso de microscopios con menor amplificación, mientras que para lograr mayor resolución axial un objetivo de aumento alto será el adecuado.

2.3 MEJORAMIENTO DEL ESPARCIMIENTO RAMAN POR SUPERFICIES

A menudo la señal obtenida por un espectrómetro es muy débil y prácticamente indistinguible con respecto al ruido; la principal razón de esto, en las muestras biológicas, es la baja concentración que la molécula de interés puede tener en la muestra [5], [27]. Existen varias técnicas para amplificar el esparcimiento Raman mediante mecanismos fotónicos y superar esta desventaja, algunos ejemplos se muestran a continuación en la Tabla 2.1.

Técnica	Principio de Funcionamiento	Amplificación
<i>Surface-Enhanced Raman Spectroscopy</i>	La molécula de interés se encuentra muy cercana a una superficie plasmónica que incrementa el campo electromagnético de excitación [5], [29].	10^{10}
<i>Resonance Raman Spectroscopy</i>	La frecuencia de excitación coincide con la frecuencia de alguna transición electrónica de la molécula, generando un efecto de resonancia [5].	10^6
<i>Surface-enhanced resonance Raman Spectroscopy</i>	Combina el efecto de resonancia con el efecto de amplificación de campo de una superficie plasmónica [30].	10^{15}
<i>Graphene-enhanced Raman scattering</i>	Utiliza grafeno como sustrato, la molécula de interés queda unida a éste y se observa amplificación debido a la interacción química entre ambos materiales [31], [32].	10^5
<i>Tip-enhanced Raman spectroscopy</i>	La punta de un microscopio de fuerza atómica es de un material con resonancia plasmónica, al acercarse a la molécula se incrementa el campo eléctrico aumentando el esparcimiento [33].	10^8

Tabla 2.1. Resumen de las principales técnicas utilizadas para mejorar la dispersión Raman. La columna de factores de amplificación muestra los más altos reportados en la literatura.

Una de las técnicas más usadas por su simplicidad y su alto nivel de amplificación es Surface Enhanced Raman Spectroscopy (SERS), que se traduce como amplificación del esparcimiento Raman por superficies o espectroscopía Raman mejorada por superficies [5].

SERS puede describirse como dispersión Raman amplificada por la presencia de una superficie plasmónica muy próxima a la molécula de interés. La molécula es excitada tanto por el campo de la fuente de excitación como por el fuerte campo electromagnético que se genera en la proximidad de la superficie plasmónica, lo que genera un incremento en la señal Raman [5], [7].

El nuevo campo electromagnético se origina porque en una superficie metálica de rugosidad periódica o bien, en un conjunto uniforme de pequeñas partículas metálicas, se crean plasmones de superficie. Los plasmones de superficie son las oscilaciones de los electrones libres del metal y son capaces de absorber y dispersar luz. Se llaman plasmones de superficie localizados cuando se excitan en su frecuencia de resonancia, es decir, la frecuencia a la que absorben y dispersan luz más eficientemente [5].

Para que una superficie sea una buena superficie SERS debe cumplir básicamente con tener resonancias de plasmón fuertes y permitir que la molécula de interés quede unida o a unos cuantos nanómetros de ella. Una óptima amplificación surge cuando la frecuencia de excitación del láser coincide con la frecuencia de resonancia del plasmón. La frecuencia y la fuerza de esta resonancia dependen del tipo de metal, del entorno dieléctrico, del tamaño, la forma y la distribución de las nanoestructuras [5].

Las superficies SERS pueden clasificarse en tres clases principales [5], [7]:

- Nanopartículas metálicas en solución coloidal.
- Estructuras metálicas 'planas', como conjuntos de nanopartículas metálicas depositadas en sustratos planos o nanoestructuras formadas por litografía.
- Electrodo metálicos.

Los materiales más comúnmente usados para la fabricación de estas estructuras son oro y plata, puesto que tienen las propiedades ópticas adecuadas para mantener resonancias de plasmón eficientes en el rango visible y cercano infrarrojo (400 - 1000 nm) [5]. Otra ventaja de usar estos materiales es que presentan alta afinidad para permitir enlaces con ciertos residuos aminos, lo cual propicia el mejoramiento de la señal Raman también por efecto químico [7].

SERS presenta otras ventajas: generalmente no requiere elaborados procedimientos experimentales ni gran cantidad de materiales o equipo, además de que es ampliamente flexible, el agua no interfiere en la señal y en algunos casos la fluorescencia puede evitarse; la posibilidad de sintonizar la frecuencia de resonancia de la superficie plasmónica a la longitud de onda deseada permite que sea una técnica adaptable [7], [29], [34].

Una desventaja de SERS es la dificultad para asegurar que las moléculas de interés se unan realmente al material resonante, además puede ser complicado controlar el número o concentración de las nanopartículas que serán responsables del efecto de mejoramiento de señal y conocer el número de moléculas que se unen a éstas.

2.3.1 Factor de Mejoramiento de SERS

Para determinar la eficiencia del mejoramiento Raman que se logra gracias a la superficie plasmónica, se puede comparar la intensidad de una banda en el espectro SERS con la intensidad de la misma banda en el espectro Raman. Para que esta comparación refleje cuantitativamente el efecto de la superficie, se puede considerar de forma ideal la señal Raman de una sola molécula y la señal SERS de la misma. Sin embargo, obtener esta intensidad individual supone un problema, pues no es posible aislar la señal de una sola molécula experimentalmente.

Una aproximación para la intensidad individual es la intensidad promedio por molécula (\bar{I}), que se obtiene al dividir cada intensidad (Raman y SERS) por el número de moléculas excitadas en cada caso:

$$\bar{I}_{Raman} = \frac{I_{Raman}}{N_{Raman}} \quad (2.5),$$

$$\bar{I}_{SERS} = \frac{I_{SERS}}{N_{SERS}} \quad (2.6).$$

Al dividir la intensidad por molécula de SERS entre la intensidad por molécula Raman, $\bar{I}_{SERS}/\bar{I}_{Raman}$, se obtiene el llamado factor de mejoramiento SERS promedio (EF¹) [5], [35]–[37].

$$EF = \frac{I_{SERS} \times N_{Raman}}{I_{Raman} \times N_{SERS}} \quad (2.7),$$

donde I_{Raman} e I_{SERS} son las intensidades de los espectros Raman y SERS en la misma banda, N_{Raman} y N_{SERS} corresponden al número de moléculas que son excitadas y producen dispersión para generar las intensidades I_{Raman} e I_{SERS} respectivamente.

De acuerdo con la aproximación de Andre Childs *et. al.* [35], el número de moléculas que generan el esparcimiento Raman sin el efecto de mejoramiento se puede calcular como:

$$N_{Raman} = \frac{\pi r^2 h \rho N_A}{M} \quad (2.8),$$

donde $\pi r^2 h$ es el volumen óptico que se excita, ρ es la densidad del material de la muestra en gr/cm^3 , N_A es el número de Avogadro y M es el peso molecular.

Para estimar el número de moléculas que intervienen en el esparcimiento mejorado por las nanopartículas, se presenta un nuevo problema: difícilmente se puede conocer la cantidad exacta de moléculas que se han unido a las nanopartículas. En algunos trabajos que utilizan sustratos SERS donde se conoce con precisión la distribución de las nanopartículas, se ha estimado el número de moléculas unidas a las nanopartículas mediante la medición de la fluorescencia en las soluciones residuo [38]. En otros casos se ha calculado la cantidad de partículas suponiendo que éstas se han enlazado en monocapa sobre las nanopartículas uniformemente distribuidas en el área de excitación [35]:

¹ EF: siglas del inglés *Enhancement factor*

$$N_{SERS} = \frac{\pi r^2}{A} \quad (2.9).$$

donde πr^2 es el área de excitación o spot del láser y A es el área superficial efectiva de cada molécula.

2.4 CARACTERIZACIÓN DE BIOMOLÉCULAS MEDIANTE ESPECTROSCOPIA RAMAN

La espectroscopía Raman tiene la capacidad de acceder a niveles moleculares en las sustancias por lo que es de especial interés su aplicación para marcadores moleculares de enfermedad.

Cuando se trata de análisis Raman en tejidos o fluidos, el espectro obtenido es una combinación de las vibraciones moleculares de las diferentes biomoléculas que conforman la muestra: aminoácidos, carbohidratos, proteínas, ácidos nucleicos y lípidos. El espectro Raman de cada una de estas biomoléculas tiene bandas particulares que pueden permitir su identificación en el espectro total.

2.4.1 Aminoácidos y proteínas

Los aminoácidos son moléculas formadas por carbono, oxígeno, hidrógeno, nitrógeno y en algunos casos azufre. Su estructura comienza con un carbono α al que se enlazan: un grupo amino (-NH₂), un grupo ácido carboxilo (-COOH), un grupo de cadena lateral que varía en cada aminoácido y un hidrógeno [39].

Las proteínas son biomoléculas formadas por aminoácidos unidos mediante un enlace peptídico: carboxilo enlazado a un grupo amino. Poseen una misma estructura primaria o esquelética que consiste en una cadena lineal de distintos aminoácidos. La secuencia lineal de aminoácidos al plegarse tiene múltiples conformaciones en el espacio, dando lugar a estructuras secundarias y terciarias. A su vez, las nuevas estructuras pueden unirse a otras cadenas de aminoácidos, formando estructuras cuaternarias.

La figura 2.4 representa estas estructuras de las proteínas. Los enlaces del arreglo esquelético son covalentes, mientras que la mayoría de los enlaces que determinan la conformación estructural y la asociación entre cadenas no lo son [7], [39].

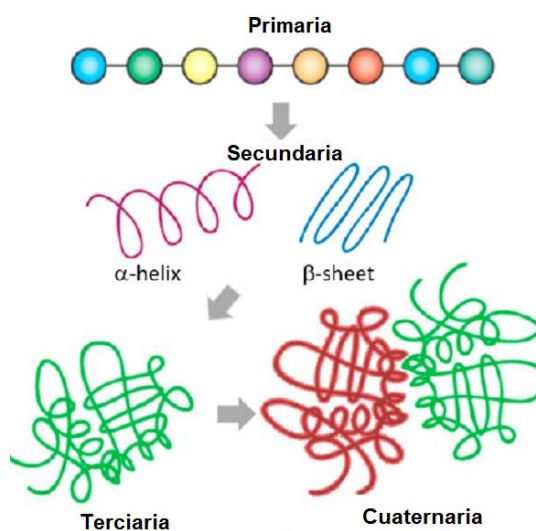


Figura 2.7. Estructuras de las proteínas [7] .

La espectroscopía Raman permite identificar las vibraciones de los aminoácidos y otros enlaces para asociarlas con las proteínas y su estructura. En general se puede considerar las siguientes asignaciones vibracionales para interpretar el espectro Raman de una proteína:

Aminoácidos aromáticos de cadenas laterales	Fenilalanina: 621, 1003, 1031 y 1208 cm^{-1} Triptófano: 759, 1011, 1340-1360 y 1555 cm^{-1} Tirosina: 643, 829, 853 y 1200 cm^{-1}
Grupos S-S (enlaces disulfuro)	510, 525, 540 cm^{-1}
Grupos C-C $_{\alpha}$	939 cm^{-1}
Grupos CH $_2$ /CH $_3$ de aminoácidos alifáticos	1340 – 1449 cm^{-1}
Grupos Amidas	Amida I: α -hélice 1645 – 1660 cm^{-1} , β -hoja 1665-1680 cm^{-1} Amida III: α -hélice 1310 – 1260 cm^{-1} , β -hoja 1240 – 1225 cm^{-1}
Grupo sulfhídrico -SH	2570 – 2580 cm^{-1}
Enlace C-S	630, 670, 700 y 720 cm^{-1}

Tabla 2.2 Principales grupos vibracionales de las proteínas [40]–[43].

La presencia, cambio de concentración o alteración de ciertas proteínas puede relacionarse con cambios patológicos, por ejemplo: el desarrollo del cáncer se asocia con la presencia de oncoproteínas y con la mutación proteínica por sustitución de aminoácidos en las cadenas primarias; otras enfermedades como el Alzheimer y el Parkinson son asociadas a un doblamiento incorrecto en la estructura secundaria. Por otro lado, algunas proteínas cambian su estructura cuaternaria y terciaria para llevar a cabo sus funciones [7].

Identificar el comportamiento estructural y conformacional de las proteínas mediante espectroscopía Raman, puede servir para establecer marcadores de enfermedad que permitan mejorar el diagnóstico y seguimiento de enfermedades.

2.4.2 Lípidos

Los lípidos son un amplio grupo de biomoléculas hidrófobas formadas de carbono e hidrógeno, en algunos casos contienen oxígeno, azufre, fósforo o nitrógeno. Su estructura generalmente es lineal, formada por cadenas alifáticas, aunque también existen lípidos con estructuras aromáticas [41], [44].

En general la señal Raman que se obtiene de los lípidos revela la presencia de enlaces entre carbono e hidrógeno en las regiones: 1250 -1300 cm^{-1} , 1400 -1500 cm^{-1} y 2800 -3100 cm^{-1} . La vibración de enlaces C-C en 1600 -1800 cm^{-1} y 1050-1200 cm^{-1} y C=O alrededor de 1749 [34], [41], [44].

Los lípidos están presentes en todos los tejidos y células del organismo y sus funciones dependen de su tipo y de su ubicación, por ejemplo, pueden estar involucrados en la señalización celular, en el almacenamiento de energía, en la construcción de membranas y en el transporte y absorción de nutrientes. Dada su importancia pueden ser usados como marcadores de enfermedad [11], [44], [45].

2.4.3 Carbohidratos

Los carbohidratos son las moléculas constituyentes esenciales de los seres vivos. Los átomos que los componen son carbono, oxígeno e hidrógeno principalmente,

pueden contener nitrógeno, azufre y fósforo. Dentro de sus funciones, las más importantes son: proporcionar energía de forma inmediata (glucosa) o almacenarla (glucógeno) y dar soporte estructural a las paredes celulares de plantas y microorganismos [46].

Dependiendo de su estructura y cadenas, los carbohidratos se pueden clasificar como monosacáridos, oligosacáridos y polisacáridos. La unidad monomérica básica en las cadenas poliméricas es muy similar, dificultando las investigaciones estructurales mediante espectroscopía Raman[34].

Algunas bandas características de los carbohidratos son: estiramientos de enlaces hidrógeno oxígeno entre $3100 - 3600 \text{ cm}^{-1}$; vibraciones y deformaciones de grupos C-H entre $2800 - 3100$ y $1200-1500 \text{ cm}^{-1}$; de 800 a 1200 cm^{-1} grupos C-O y C-C y entre 100 y 800 cm^{-1} deformaciones CCO [11], [34], [46].

2.4.4 Ácidos nucleicos

Los ácidos nucleicos son biomoléculas que contienen la información genética de las células y los virus. Su estructura está formada por subunidades llamadas nucleótidos, que se unen con un enlace fosfodiéster (enlace covalente entre hidroxilo y fosfato) [39].

Los espectros Raman de los ácidos nucleicos contienen alrededor de 30 bandas que se originan a partir de residuos de azúcar, grupos fosfato y bases, con lo cual se ha logrado seguir la conformación de la estructura esquelética, el apareamiento de bases, e interacciones con proteínas, fármacos y otros compuestos [34], [40].

Se observan bandas Raman muy fuertes en la región de 600 a 800 cm^{-1} , los cuales son asignados a la respiración de anillos de varias bases. De todas las bases, la timina tiene un pico muy característico alrededor de 1671 cm^{-1} [34], [40]. Las bandas características de los grupos fosfatos son 980 , 1080 y 1240 cm^{-1} [9].

2.5 MÉTODOS DE ANÁLISIS MULTIVARIABLE APLICADOS EN ESPECTROSCOPIA RAMAN

Una gran cantidad de datos es generada durante los experimentos de adquisición de espectros Raman debido a que cada espectro se traduce en un arreglo de dos vectores columna: un vector contiene los números de onda en los que se mide la intensidad de la dispersión y el otro contiene los valores de esta intensidad. Entre mayor sea la resolución del espectrómetro utilizado, mayor será la dimensión de cada vector.

Experimentalmente, al analizar una muestra para obtener su espectro Raman, es necesario llevar a cabo varias mediciones, lo cual incrementa la cantidad de datos. Esta gran cantidad de datos provenientes de los espectros puede analizarse mediante quimiometría. La quimiometría permite extraer información química y física de los espectros haciendo uso de métodos matemáticos y estadísticos que en general se engloban en dos enfoques: análisis univariable y análisis multivariable [8], [27].

El análisis de una sola variable (univariable) se enfoca en analizar un valor específico del espectro descartando el resto, por ejemplo: la intensidad de alguna vibración en particular, la variación en su posición o en su ancho. Para grandes cantidades de datos, esta técnica no es muy conveniente por la cantidad de tiempo que implica analizar individualmente los valores. Tampoco es conveniente analizar solamente una variación vibracional, pues no es un identificador específico para una sola molécula [5], [27].

El análisis multivariable permite considerar todo el espectro como una sola entidad y estudiar estadísticamente el comportamiento del conjunto de espectros sin descartar *a priori* ningún dato. La principal ventaja de los métodos de análisis multivariable es que permiten manejar automáticamente una gran cantidad de valores de acuerdo con su importancia estadística, facilitando su interpretación. Al mantener los datos ordenados de acuerdo con su relevancia, es posible eliminar aquellos de poca importancia [8], [27], [47].

Dentro del análisis multivariable se tienen métodos supervisados y métodos no supervisados. Los métodos no supervisados no requieren un conocimiento inicial descriptivo del conjunto de datos, simplemente requieren que se identifiquen los

valores que corresponden a las observaciones y a las variables. Los métodos supervisados, por el contrario, sí requieren de al menos un valor que describa alguna característica de los datos [8], [47].

Los métodos de análisis multivariable también pueden agruparse según su aplicación; en espectroscopía sus principales usos son: minimizar la cantidad de datos para observar agrupaciones o características ocultas y llevar a cabo clasificaciones, regresiones o predicciones [8].

2.5.1 Análisis de Componentes Principales: PCA

PCA es uno de los métodos no supervisados más usados en aplicaciones de espectroscopía Raman, principalmente por su capacidad para disminuir las dimensiones de los datos, permitir observar cualitativamente tendencias entre los mismos y revelar cuantitativamente las similitudes y las diferencias vibracionales.

PCA permite reducir la matriz original de M variables (columnas) y N observaciones (renglones) en tres grupos de nuevos valores mediante una transformación ortogonal: cargas, puntuaciones y porcentajes [10], [18], [48].

Las cargas o *PC loadings*, son N vectores columna de M dimensiones. Cada vector corresponde a un componente principal y se encuentran ordenados según su importancia estadística [10], [48]. Los *PC loadings* también son llamados *nuevas variables*, pues al graficar cada vector contra los corrimientos Raman, se observa un nuevo espectro, cuyas nuevas bandas de vibración revelarán las de mayor, menor o nula variabilidad.

Las puntuaciones o *PC scores* son una matriz $N \times N$ coeficientes, cada columna corresponde a un componente principal y cada renglón asigna un valor (puntuación) a un espectro en el orden original. Los *PC scores* también aparecen ordenados de acuerdo con su porcentaje de varianza explicada: del que mayor relevancia representa al de menor. Al graficar los *PC scores* más relevantes uno contra otro, se puede observar un nuevo espacio de variables reducidas que revela el comportamiento de los espectros entre sí: *PC scores* con valores similares tenderán a agruparse y viceversa [10], [18], [48], [49].

Los porcentajes de varianza explicada se obtienen en un vector cuyos valores describen la importancia de cada componente principal en orden descendente: así el primer valor indica la contribución estadística del primer *PC loadings* y del primer *PC score*, y así sucesivamente. Al conocer la relevancia de cada componente principal, es posible eliminar aquellos cuya contribución sea poco relevante [48].

2.5.2 Regresión de mínimos cuadrados parciales: PLSR

PLSR es un método multivariable supervisado que permite crear una regresión lineal para llevar a cabo predicciones. Esta técnica, genera un modelo para describir la relación entre dos variables independientes: variable predictora \mathbf{X} y variable de respuesta \mathbf{Y} :

$$Y = \beta \cdot X + E \quad (2.10).$$

Donde β es la matriz de los coeficientes de la relación lineal, la cual se busca determinar con el menor error posible E [50].

El método también es llamado proyección en estructuras latentes ya que, para obtener la mejor predicción, se extraen de las variables predictoras un nuevo grupo de variables ortogonales que son calculadas mediante transformaciones lineales, estas nuevas variables latentes explican el mayor porcentaje de varianza y por lo tanto tienen el mayor poder predictivo. La calidad de la predicción obtenida de un modelo de regresión PLS suele evaluarse con técnicas de validación cruzada (*cross-validation*) [51].

2.5.3 Máquina de vectores de soporte: SVM

Un SVM es un método supervisado de aprendizaje automático para clasificación binario: para un conjunto de datos X , SVM encuentra un hiperplano de separación para dividir los datos en dos clases mediante una función $g(x)$. Esta función se encuentra optimizada al mantener la máxima distancia (margen) a los datos borde de cada clase, llamados vectores de soporte, como se ilustra en la Figura 2.8 [52], [53].

Comúnmente los grupos de datos no son linealmente separables, por lo que se implementan *kernels* para transformar los datos a un nuevo espacio donde sean separables mediante una función lineal g . En otros casos además no existen solo dos clases, requiriéndose ajustar el algoritmo para entrenar al modelo comparando cada clase contra todas las otras hasta tener todos los pares de clases comparadas [52], [54], [55].

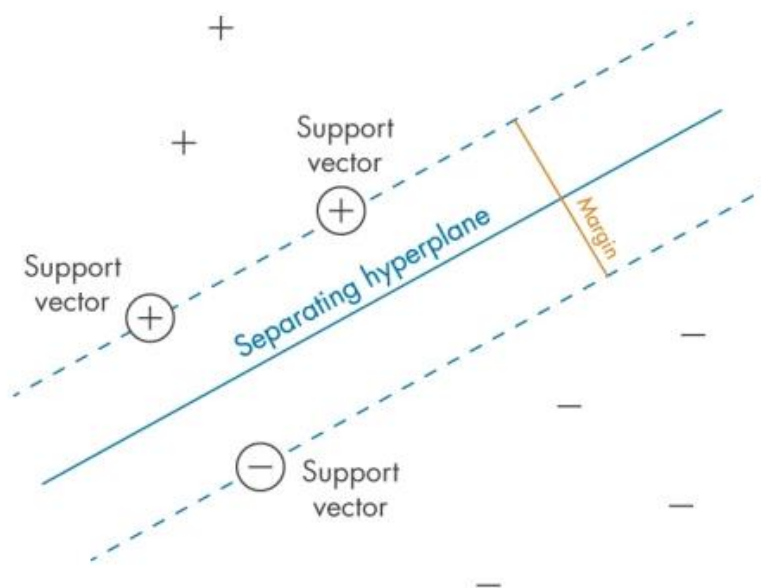


Figura 2.8. Ejemplo de clasificación por SVM en dos grupos. El mejor hiperplano es aquel con el mayor margen entre las dos clases [52].

2.6 SIMULACIONES MOLECULARES

Una molécula es una estructura tridimensional y sus vibraciones son complejas. Su espectro Raman no solo depende de la vibración de enlaces individuales, sino que es una superposición de frecuencias provenientes de varios modos de vibración. Para interpretar un espectro Raman se requiere conocimiento preciso de la frecuencia de vibración que los núcleos atómicos, por lo que supone resolver la ecuación de Schrödinger para la molécula, considerando su configuración electrónica y su microentorno [4], [27]. Dicha tarea es sumamente compleja y no existe una solución analítica para moléculas de más de dos átomos [3], por lo que la alternativa es la implementación de métodos computacionales para simular el comportamiento vibracional de la molécula.

Existe una gran cantidad de programas de software comercial utilizados en la disciplina de química computacional, los cuales permiten aplicar la mecánica cuántica y la dinámica molecular para encontrar una aproximación a la solución de la ecuación de energía. Utilizar este tipo de software es práctico para sistemas relativamente simples como moléculas con unos cuantos átomos; sin embargo, para sistemas complejos como proteínas con cientos de aminoácidos se requiere de grandes capacidades de cómputo y en ocasiones llegan convertirse en sistemas irresolubles [56].

El programa Gaussian, es un software que permite simular el comportamiento vibracional de ciertas moléculas. Utiliza métodos de química cuántica para resolver la ecuación de energía, tales como: *ab initio*, *Hartree Fock*, *DFT* y los métodos semiempíricos [56]. Por ejemplo, Béatrice Sjöberg *et. al.*, llevó a cabo cálculos *DFT* para obtener el espectro Raman de cadenas laterales de aminoácidos en las proteínas y compararlos con los espectros obtenidos experimentalmente, demostrando que no siempre se reproduce con exactitud el espectro obtenido analíticamente, sin embargo, el resultado es útil para asignar sin ambigüedad las vibraciones correspondientes a las bandas Raman [57]. R. Cabrera *et. al.*, demostró también la utilidad de cálculos *DFT* para determinar los modos vibracionales de la hidroquinona, igualmente demostrando que la asignación vibracional mediante las simulaciones permite la interpretación adecuada de las bandas [58].

Capítulo 3.

MARCADORES DE ENFERMEDAD

3.1 MARCADOR TUMORAL P53

Todos los días nuestras células están expuestas a agentes dañinos como químicos, virus y radiación, sin embargo, el organismo tiene vías para regular el daño y mantener el estado normal de las células. También la reproducción celular está perfectamente regulada mediante una compleja red de señales que definen cuándo una célula puede dividirse y qué tan rápido puede hacerlo.

Cuando las células dañadas no son reparadas y se reproducen sin control, entran directamente en la ruta hacia la tumorigénesis. Las células cancerígenas típicamente contienen dos tipos de mutaciones: mutaciones que causan crecimiento y multiplicación descontrolada, y mutaciones que bloquean las defensas que protegen contra este crecimiento anormal [59].

La activación de la proteína p53 en su estado natural es uno de los mecanismos de regulación para que las células dañadas no se reproduzcan. Esta proteína induce mediante diferentes vías la reparación del ADN dañado; cuando no es posible repararlo provoca apoptosis: muerte celular para evitar la proliferación del daño. Un p53 defectuoso o inactivado podría permitir que las células anormales continúen su reproducción.

En alrededor de 50% de los cánceres humanos se han encontrado mutaciones de p53, algunas de las cuales se convierten en oncogenes, teniendo funciones para proteger a las células cancerígenas [60]–[62].

Por lo anterior, la proteína p53 en sus diferentes estados es de gran interés, ya que sus características y comportamiento la posicionan como un biomarcador prometedor para mejorar el diagnóstico y el seguimiento del cáncer o bien para diseñar terapias anticáncer más efectivas.

3.1.1 Localización y Estructura de la Proteína p53

El gen Tp53 se encuentra en el cromosoma 17, banda 13 (17p13) y codifica la proteína p53, factor de transcripción que al estar activado se enlaza al ADN. Su nombre se debe a su masa molecular aparente: 53 kD. Está formada por 393 aminoácidos y se considera una proteína “desorganizada intrínsecamente” por su estructura compleja y flexible [63], [64]. También llamado “guardián del genoma” por su tarea de reparar el ADN celular, el gen Tp53 es un gen supresor de tumores, es decir: su actividad detiene la formación tumoral [65].

P53 es una proteína homotetramérica: su estructura contiene cuatro subunidades idénticas que están asociadas, pero no unidas covalentemente. Cada monómero de 393 aminoácidos se divide en distintos dominios funcionales, como se muestra en la Figura 3.1: dominio amino terminal (N-terminal), dominio de unión al ADN (DBD), dominio de tetramerización y dominio carboxilo terminal (C-terminal)[66].

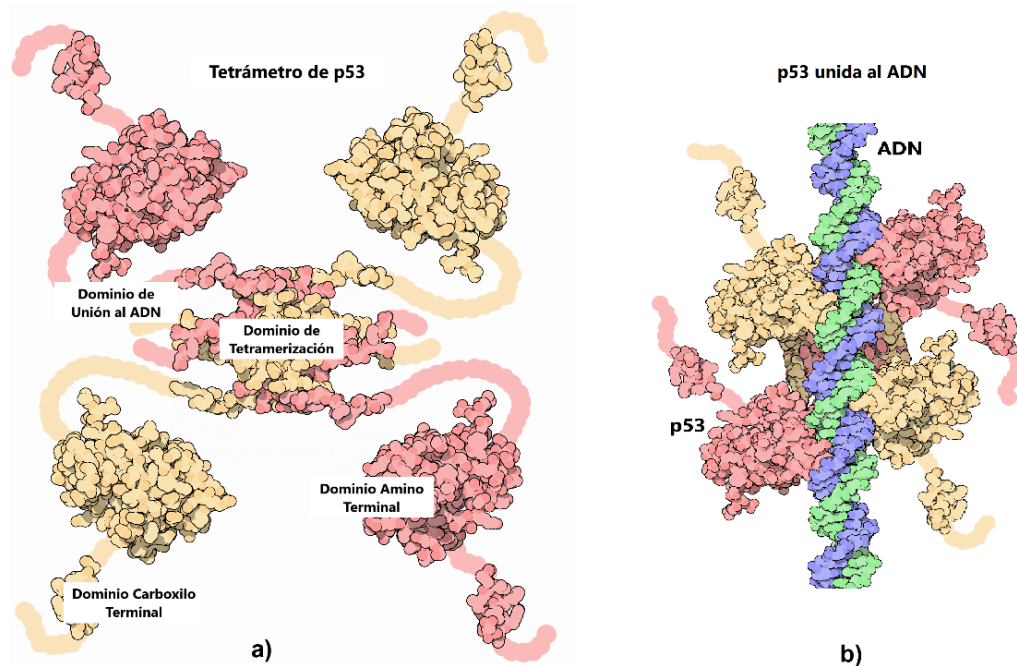


Figura 3.1. a) Esquema de la estructura tetrámera de p53 en la que se distinguen las cuatro subunidades y se identifican los dominios funcionales. b) Proteína p53 unida a una cadena de ADN [66]–[68].

En el centro de p53, entre los aminoácidos 320 y 360, se encuentra el dominio de tetramerización, que se une a cuatro cadenas laterales (residuos). Este dominio central contiene una secuencia de aproximadamente 30 aminoácidos, que lo enlazan al dominio de unión al ADN.

El DBD tiene cuatro regiones mediante las que reconoce e interactúa con secuencias específicas del ADN. Estos cuatro dominios, ubicados entre los aminoácidos 94 y 294 de cada una de las cuatro cadenas, son ricos en arginina y reconocen sitios reguladores específicos [67].

La región N-terminal se encuentra cerca del final de cada brazo, está constituida por los primeros 94 aminoácidos y contiene [66]:

- Dos dominios de transactivación, AD1 y AD2, que son imprescindibles para la activación transcripcional de genes.
- Una región rica en prolina con secuencias implicadas en la apoptosis regulada.
- Una señal de exportación nuclear y múltiples sitios que, ante una señal de estrés, causan fosforilación y, por consiguiente, este dominio se libera de MDM2, principal regulador de p53 [61].

El dominio C-terminal comprende los aminoácidos 363-393. Gracias a su abundancia en residuos básicos de arginina y lisina, es capaz de unir ADN de forma inespecífica mediante interacciones electrostáticas, capacidad que le permite regular la actividad del dominio de unión al ADN [63], [64], [66].

En condiciones normales, p53 se encuentra en su estado natural, presente en todos los tejidos del cuerpo humano en cantidades muy pequeñas. Cuando p53 se activa, la concentración aumenta considerablemente y es transportado del citoplasma al núcleo, sitio donde lleva a cabo sus funciones [64].

3.1.2 Funciones y Regulación de la Actividad de la Proteína p53

Cada día ocurren situaciones que producen estrés celular y dañan el ADN. Un simple cambio en su secuencia es capaz de ocasionar enfermedades como cáncer, envejecimiento prematuro, neurodegeneración, entre otras. Mantener la integridad genómica reparando dichos daños, es esencial para la salud del organismo. A los procesos encargados de resguardar la integridad del ADN se les denomina mecanismos de reparación; la gran mayoría de ellos suceden gracias a la participación de diversas proteínas. La proteína p53 es una de las implicadas en estos procesos, actuando ya sea como factor de transcripción o como activador de la apoptosis para evitar la proliferación del daño.

En una célula normal, el nivel de la proteína p53 es bajo, debido a su corto tiempo de vida. Su función de factor de transcripción comienza como respuesta a señales de estrés celular: incrementando sus niveles para que se una a secuencias específicas del ADN. De esta forma al controlar la transcripción de la información genética, inhibe efectivamente la reproducción de las células dañadas [60], [69].

Una vez que p53 se ha unido al ADN, sus funciones son esencialmente las siguientes [63], [64]:

1. Detención o arresto del ciclo celular para impedir que el ADN dañado sea replicado, de esta forma asegura la integridad genómica y puede dar paso a la reparación del ADN.
2. Activación de enzimas de reparación del ADN. Cuando la reparación es exitosa, el ciclo celular continúa, de lo contrario, p53 activará los mecanismos de senescencia o apoptosis para asegurar la muerte de la célula dañada y evitar su reproducción.
3. Entrada en la senescencia, para asegurar que las células dejarán de dividirse permanentemente.
4. Activación de la apoptosis o suicidio celular, respuesta que sucede no solo cuando no es posible reparar el daño, sino también cuando p53 detecta señales oncogénicas.

Por lo anterior, el gen Tp53 se convierte en un sensor de estrés celular, que controla la muerte o permanencia de una célula, por lo que su expresión y su actividad se encuentran perfectamente reguladas mediante inhibidores y coactivadores.

Uno de los más estudiados inhibidores de p53 es el gen MDM2. En células normales, la asociación a MDM2 induce la ubiquitinación de p53 y su destrucción en la proteasoma, manteniendo sus niveles bajos. Cuando ocurre un daño celular, dos quinasas activan vías de señalización que fosforilan diversas proteínas, entre ellas a p53. La fosforilación de p53 la libera de su asociación con mdm2, aumentando su vida media para poder ejercer sus funciones. Si el daño se repara correctamente, p53 ahora estimula la síntesis de MDM2, activando su propia destrucción para permitir que el ciclo celular continúe [61].

3.1.3 Mutaciones de p53

Las células cancerígenas generalmente se escapan de las funciones del supresor tumoral p53, ya sea por un mal funcionamiento o bien debido a mutaciones en este gen, bloqueando las defensas naturales que protegen contra el crecimiento anormal [70].

Alrededor del 50% de los cánceres humanos tienen al menos una mutación de p53. Las mutaciones no solo impiden que se lleve a cabo el proceso de supresión de tumores, sino que además una buena parte de estas mutantes actúan para proteger a las células cancerígenas y permitir su proliferación, es decir, se convierten en oncogenes [70]–[72].

Cerca del 75% de las mutaciones de p53 identificadas en cánceres humanos son mutaciones con cambio de sentido: hay una sustitución de un aminoácido por otro; otro 15% son mutaciones por desplazamiento o mutaciones sin sentido. De las mutaciones con cambio de sentido, cerca de 90% ocurren en el dominio de unión al ADN o muy cerca de esta zona, por lo que afectan la forma en la que la proteína se enlaza y por lo tanto desactivan sus funciones de transcripción [72]. Dentro de este tipo de mutaciones existen seis que ocurren con alta frecuencia y se ubican en los codones: 175, 245, 248, 249, 273 y 282. El porcentaje más alto de mutaciones ha sido hallado en los codones 248 y 273 [64], [72],[73]. La Figura 3.2 demuestra este

hallazgo y permite tener una idea general sobre la frecuencia de las mutaciones en los diferentes dominios de la proteína.

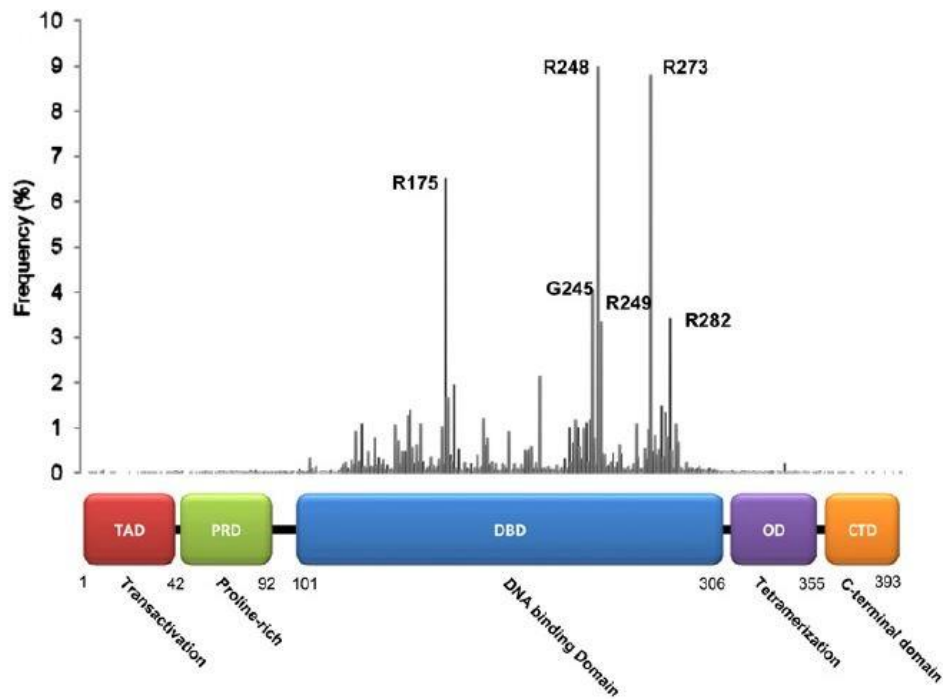


Figura 3.2. Frecuencia de la aparición de mutaciones con cambio de sentido en p53 respecto a la posición del aminoácido y sus dominios. Los datos provienen de cánceres humanos de 19262 pacientes [72].

Las mutantes son intrínsecamente inestables, sin embargo, en tejidos tumorales se acumulan y se estabilizan a niveles altos debido a un conjunto de señales que se originan en el microambiente alterado del tumor. La estabilización de p53 mutante permite que sus niveles sean mucho más altos que los de p53 en estado natural, manteniéndose a sí misma mediante funciones oncogénicas [73]. Esta acumulación mejora la supervivencia de las células cancerígenas, permitiéndoles responder ante el estrés y facilitando su adaptación [74]. Otro efecto que tienen algunas mutantes de p53 es la reprogramación del metabolismo celular, lo cual es necesario para el crecimiento tumoral y su proliferación [62], [71].

El momento en que ocurren las mutaciones de p53 durante el cáncer depende del tipo de tumor, algunas mutantes se encuentran presentes en etapas tempranas mientras que otras se expresan cuando el cáncer está en etapas avanzadas [60]. Se

ha encontrado que los cánceres con mutaciones en p53 son más agresivos, con mayor capacidad metastásica y a menudo fatales. La detección de las anomalías de p53 puede revelar aspectos de origen y evolución del cáncer humano [63], [73], [74].

3.1.4 Uso de p53 en diagnóstico y en terapias

Desde que la proteína p53 fue descubierta como un gen regulador de tumores se ha planteado la idea de que su detección podría mejorar el diagnóstico de cáncer, sin embargo, no ha sido posible demostrar que ésta es una alternativa viable, principalmente porque la alteración de niveles en la expresión de p53 no solo se asocian a desarrollo tumoral [75]. Por ejemplo, ante un daño celular tan típico como el que sucede por la exposición a los rayos ultravioleta del sol, las señales de estrés activan p53 aumentando su expresión, esto es solo es una reacción ante el daño y no significa que se desarrollará cáncer de piel [76].

Por el contrario, la detección de mutantes podría significar un gran avance para mejorar el diagnóstico y el pronóstico, debido a su alta incidencia en cánceres humanos y a las funciones oncogénicas que algunas de éstas obtienen. Un conocimiento más preciso sobre las mutantes presentes en los cánceres también permitiría desarrollar tratamientos específicos y más eficaces contra el mismo.

En un gran número de tumores se ha encontrado que las mutantes promueven la quimio y radio resistencia mediante adaptación al estrés, protegen a las células cancerígenas y son precursoras de metástasis, por lo que algunas alternativas terapéuticas podrían ser neutralizar su acción o bien conseguir su regulación negativa para restablecer las funciones de p53 en su estado natural.

Desde su descubrimiento en 1979, p53 ha sido detectado mediante una variedad de técnicas, las más comunes, resumidas en la Tabla 3.1, se basan en la generación antígeno-anticuerpo. Aunque existe un gran avance en estos métodos y presentan en su mayoría alta sensibilidad [77]–[79], tienen desventajas como la necesidad de varios instrumentos y materiales costosos, mano de obra bien capacitada para operar estos equipos y tiempos prolongados para obtener los resultados [22].

Imunohistoquímica	Mediante microscopio se analizan las zonas de interés, las regiones con mayor cantidad de proteína tendrán una coloración distinta debida a la conjugación antígeno-anticuerpo [80].
FASAY	Análisis en colonias de levaduras, en las que se activa la interacción proteína-proteína y permiten distinguir el estatus de p53 [77], [78].
ELISA	El antígeno se detecta al inmovilizarlo con un anticuerpo enlazado a una enzima que produce un cambio de color [81].
Western blot	Separación por tamaño de proteína y visualización mediante el uso de anticuerpos [82].

Tabla 3.1. Resumen de las técnicas más comunes para detectar p53

3.1.5 Detección de p53 mediante espectroscopía Raman.

La investigación de p53 mediante espectroscopía Raman tiene escasos 10 años, pocos trabajos se han reportado al respecto y aún queda mucho para lograr establecer Raman como un método confiable para la detección de este biomarcador.

De las investigaciones más relevantes las ha llevado a cabo el grupo de Salvatore Cannistrano en Italia. En 2016 reportaron el espectro Raman de p53 tipo natural y la identificación de sus bandas más importantes [83]. En 2018 reportaron un sensor SERS que promete detectar p53 (tipo natural y mutantes) en suero en concentraciones de hasta 10^{-17} M. El sensor es creado con anticuerpos de p53 que se enlazan a nanopartículas de oro mediante 4-ATP, posteriormente se colocan en muestras de suero con p53 y al crearse la unión anticuerpo-proteína, la vibración de los enlaces que se generan son los que se siguen como marcadores de la presencia de p53 [38], [84].

En 2015 Peter Owens *et. al.* reportó la detección de p53 con sustratos SERS funcionalizados con antígenos y anticuerpos, logrando hasta 0.23 nM. Este estudio además relaciona la variación en la concentración con el cambio en el ancho medio de la banda de vibración 1080 cm^{-1} y su ligero corrimiento hacia frecuencias más altas [85]. Lei Wu y colaboradores en 2013 reportaron la detección simultánea de p53 y p21 mediante sustratos SERS combinados con anticuerpos; logrando detectar hasta 1 pg/mL de ambos analitos [86].

Todos estos estudios tienen la característica de lograr determinar la presencia de p53 mediante el seguimiento de anticuerpos específicos, es decir: utilizan etiquetas para reconocer la proteína de interés y aunque tienen resultados altamente específicos, el proceso experimental es sumamente complicado.

Para que la identificación de p53 y de sus mutantes sea empleado como un marcador tumoral a nivel clínico, además de entender los mecanismos de acción de cada una de las mutantes es necesario lograr diferenciar entre ellas durante su detección. Para lograrlo se requieren métodos altamente efectivos que además sean herramientas con un diseño simple y de facilidad operativa.

3.2 AMINOÁCIDO L-ASPARAGINA

L-Asparagina (L-Asp) es un aminoácido no esencial que puede ser sintetizado mediante diversas vías metabólicas para cubrir las necesidades nutricionales de las células. La sobrevivencia de algunas células cancerígenas, su adaptación y su evolución, tienen dependencia con la disponibilidad extracelular de asparagina, pues no pueden sintetizarla ellas mismas [87], [88].

Se han desarrollado tratamientos anticáncer a base de enzimas que inducen la eliminación de L-Asp libre [88]. Las técnicas que se utilizan para seguir su concentración durante el tratamiento generalmente están basadas en ensayos enzimáticos y cromatografía líquida de alta presión; estos métodos son altamente específicos y sensibles, pero son tardados y laboriosos, lo que los vuelve costosos y de difícil acceso en los centros de atención oncológica. Mejorar el tipo de técnicas de medición de L-Asp en fluidos de pacientes bajo tratamiento representa monitorear la eficacia de éstos y poder hacer ajustes de acuerdo con su necesidad individual.

3.2.1 Estructura y localización fisiológica

L-Asparagina es un alfa aminoácido necesario para la biosíntesis de proteínas, la Figura 4.1 muestra su estructura: contiene un grupo α -amino, un grupo de ácido carboxílico y una carboxamida como cadena lateral [89], [90].

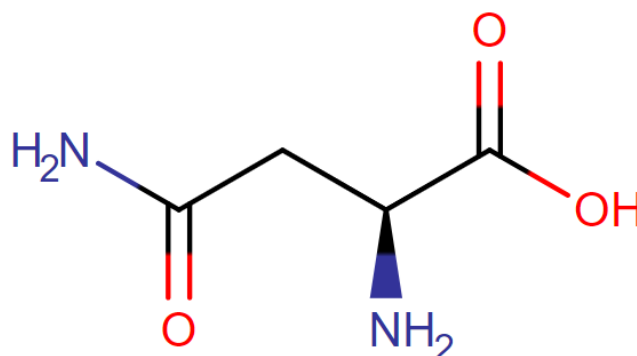


Figura 3.3. Estructura química de L-Asparagina [89].

A nivel celular se encuentra presente en el citoplasma y mitocondria. Se ha identificado en tejidos como la placenta y se ha medido en sangre, fluido cerebroespinal, heces, saliva, sudor, orina y leche materna [91].

Algunos padecimientos en los que se ha identificado como parte del metaboloma son: preclamsia, Alzheimer, cáncer de colon, cáncer de mama y leucemia linfoblástica aguda [91], [92].

Fluido	Condición	Edad	Concentración (µM)
Sangre	Sanos	Adulto	16.4 - 90.0
Sangre	Sanos	Niños (0 - 1 año)	23 - 132
Sangre	Sanos	Niños (1-18 años)	23-79
Leche materna	Sanos	Adultos	16.6 +/- 10.5
Fluido cerebroespinal	Sanos	Adultos	0 - 9
Fluido cerebroespinal	Sanos	Niños (0 – 13 años)	4.00 - 19 .0
Saliva	Sanos	Adultos	0.069 - 3.58
Sangre	Embarazo	Adulto	10.3 -34.5
Sangre	Uremia	Niños (1-13 años)	29.3-52.3
Fluido cerebroespinal	Esquizofrenia	Adulto	5.36 - 6.66
Fluido cerebroespinal	Leucemia	Niños (1-13 años)	1.5 - 17.8
Saliva	Alzheimer	Adulto	1.68 +/- 2.90

Tabla 3.2. Concentraciones de L-Asp reportadas en diversos estudios clínicos [91].

3.2.2 L-Asparagina y su importancia en el control de la leucemia linfoblástica aguda

La leucemia linfoblástica aguda (LLA) ocurre cuando existen mutaciones en el ADN de las células madre de la médula ósea, lo cual impide que las células blancas maduren correctamente y sean liberadas prematuramente al torrente sanguíneo. Este es el tipo de cáncer más común durante la niñez; alrededor del 85% de los casos que afectan a los niños ocurren en menores de 15 años, principalmente entre las edades de 0 y 5 [23].

Las células cancerígenas de LLA (células linfoblásticas) no tienen la capacidad de producir asparagina, por lo que se alimentan de la disponible en el plasma. Al introducir la enzima L-Asparaginasa (L-ASNase) al flujo sanguíneo, se promueve la eliminación de asparagina mediante su transformación a ácido aspártico y amoníaco. El agotamiento de la asparagina extracelular causa que las células linfoblásticas mueran por desnutrición: parando la replicación del ADN dañado y el crecimiento tumoral [23], [87].

El uso de L-ASNase contra LLA es uno de los tratamientos más exitosos, pero presenta graves efectos como infecciones o resistencia inmunológica, debido a que la enzima generalmente se obtiene de bacterias como *Escherichia coli* y *Erwinia chrysanthemi* [88]. Otros efectos secundarios son: pancreatitis, reacciones de hipersensibilidad, complicaciones trombóticas y neurológicas.

Para mejorar los tratamientos anti-leucemia es necesario determinar la cantidad de L-ASNase que eficientemente promueve el agotamiento de L-Asp y al mismo tiempo causa el menor número de efectos secundarios, por lo que es de suma importancia monitorear la concentración de L-Asp libre en tejidos y plasma.

3.2.3 L-Asparagina y metástasis

Estudios recientes han encontrado una relación entre la biodisponibilidad de la L-Asparagina y el desarrollo de metástasis en cáncer de mama, cerebro, hígado y riñón, sugiriendo que este aminoácido incrementa la migración de células cancerígenas y facilita su adaptación [24], [93], [94]. Otros estudios han demostrado que ciertos mecanismos de producción metabólica de Asparagina se

relacionan con el aumento de metástasis y que su unión a ciertos oligosacáridos es necesaria para la proliferación de células cancerígenas y el desarrollo de metástasis [95]. Aunque aún no es completamente claro el papel de L-Asp durante el desarrollo de cáncer, establecer técnicas que permitan determinar su presencia y concentración en pacientes con cáncer permitiría mejorar las investigaciones, anticipar tratamientos para eliminarlo y prevenir la metástasis.

3.2.4 Detección de L-Asparagina mediante Espectroscopía Raman

La Tabla 3.2 resume algunos trabajos que reportaron la medición de L-Asp en fluidos [91], sin embargo todos ellos utilizan técnicas estándares como HPLC. Algunos otros trabajos se han llevado a cabo para desarrollar biosensores para asparagina, por ejemplo: en 2014 se reportó un biosensor amperométrico que capturaba L-Asp mediante poliimida, permitiendo medir hasta 2 μM en suero [96]; en 2018 Nikolaos E. Labrou *et. al.*, presentaron un biosensor basado en reacción enzimática, el cual promete medir concentraciones de hasta 10 μM en suero y muestras de alimentos [97].

Moreno *et. al.*, llevó a cabo un estudio de caracterización Raman en cristales de monohidrato de L-Asparagina modificando polarizaciones y temperatura, lo que permitió identificar las bandas características [98]; en 2014 S. Sylvestre y colaboradores, midieron y simularon los espectros FT-IR y FT-Raman, la simulación *DFT* les permitió comprobar el origen de los modos de vibración y asignarlos correctamente con los obtenidos experimentalmente [99]. B.O. Golichenko *et. al.*, también llevó a cabo un estudio Raman de L-Asp ampliando el rango de frecuencias, la polarización y las longitudes de onda, comprobando la presencia de las bandas características antes reportadas [100].

En 2016 se reportó un tipo de sensor SERS que se obtiene con nanopartículas de plata ensambladas en una estructura metal-orgánica homoquiral; este sensor permitió detectar dos conformaciones de Asparagina, tipo L y tipo D, hasta en 1 μM [101]. En 2019 se reportó el espectro SERS de L-Asp disuelta en agua a 2 μM utilizando nanopartículas de plata [102]. Hasta ahora ningún trabajo ha llevado a cabo la medición mediante Raman de L-Asp en condiciones fisiológicas.

Capítulo 4.

ESPECTROSCOPIA RAMAN DEL MARCADOR TUMORAL p53

Cualquier sustancia o característica biológica que se pretenda establecer como biomarcador debe ser consistentemente estudiado, no solo su estructura física y química o sus funciones biológicas e implicaciones clínicas, sino que además se debe conocer la respuesta que esta sustancia producirá con el instrumento que será usado para su detección. Por esta razón, el primer paso hacia lograr establecer la espectroscopía Raman como una técnica de apoyo en el diagnóstico médico, es la caracterización del biomarcador específico: conocer su espectro o huella digital y asignar qué vibraciones moleculares corresponden a las bandas presentes. Esta caracterización permite identificar y diferenciar a la molécula, así como tener una idea más clara de su estructura y monitorear sus cambios.

Este capítulo describe la caracterización espectral de la proteína p53 tipo natural y sus mutantes R273H, E343A y L344A. Se describe la metodología implementada para la adquisición de los espectros y el procesamiento posterior. También se presenta la asignación de vibraciones moleculares de acuerdo con las bandas características de los aminoácidos principales.

Otra parte importante de la caracterización de un biomarcador es el límite de detección que el instrumento de medición puede alcanzar. Este parámetro puede calcularse mediante una regresión lineal a partir de resultados experimentales [103]. En este capítulo se presenta el cálculo del límite de detección para la proteína p53 tipo natural aplicando dos enfoques: análisis univariable y análisis multivariable, comparando el resultado obtenido en ambos casos.

4.1 ESPECTRO RAMAN DE LAS PROTEÍNAS P53: TIPO NATURAL Y MUTANTES 273, 343 Y 344.

4.1.1 Muestras

Muestras de cuatro tipos de proteína humana p53 expresadas en la bacteria *Escherichia coli* fueron proporcionadas por el Laboratorio de Interacciones Biomoleculares y Cáncer del Instituto de Física de la Universidad Autónoma de San Luis Potosí. Todas las muestras se encontraban en una solución buffer salina con un pH de 8.5 y se mantuvieron en refrigeración ($\sim 3^{\circ}\text{C}$) hasta ser usadas.

TIPO	CONCENTRACIÓN (μM)	PUREZA (%)
Natural	18.9	> 98
	12.6	> 95
Mutante R273H	9.04	> 96
Mutante E343A	6.10	> 96
Mutante L344A	17.7	> 96

Tabla 4.1. Tipos de p53 estudiados mediante espectroscopía Raman. Se indica la concentración a la que se recibieron las muestras y su respectiva pureza.

La mutación R273H que se origina por una sustitución en el codón 273 de arginina (R) por histidina (H), es una de las mutaciones más comúnmente identificadas en los cánceres humanos; ocurre en el dominio de unión al ADN por lo que directamente impacta en las funciones de regulación de crecimiento tumoral de p53 [104]. Esta mutación, no causa distorsiones estructurales en los residuos cercanos por lo que, tampoco impacta significativamente en la estabilidad de la proteína [105].

Las mutaciones E343A y L344A (sustituciones de ácido glutámico por alanina y leucina por alanina respectivamente) son mutaciones en el dominio de tetramerización. Mediante este dominio p53 forma un tetrámero para actuar como factor de transcripción, las mutaciones pueden comprometer la formación del tetrámero y por consiguiente las funciones de transcripción de p53, es decir éstas mutaciones impactan en la estructura cuaternaria de la proteína y por lo tanto en su estabilidad y sus funciones [106], [107].

4.1.2 Metodología para la adquisición de espectros Raman.

Un micro espectrómetro Raman confocal XploRA PLUS® fue usado para adquirir los espectros de cada muestra de proteína. Este espectrómetro está equipado con un microscopio Olympus BX41 y un sensor CCD multicanal [108] . Antes de comenzar la adquisición de los espectros, el espectrómetro fue calibrado para asegurar la posición correcta de las bandas y maximizar la intensidad captada por el sensor. La calibración se llevó a cabo para los dos láseres del equipo y todas sus rejillas de difracción, tomando como referencia una muestra de silicio cuya banda Raman característica se encuentra en 520 cm^{-1} .

Un diodo láser de 785 nm fue usado como fuente de excitación durante la adquisición de los espectros Raman. La elección de esta longitud de onda se llevó a cabo en función de la respuesta obtenida con este láser y otro de 532 nm. En ambos casos la señal de la fluorescencia estuvo presente, sin embargo, como era esperado, para el láser de 785 nm fue menor [29]. Para elegir la potencia del láser, se observó la respuesta en la intensidad y en las bandas del espectro usando diferentes filtros. La máxima potencia es de 22 mW para este láser, la cual se utilizó para las adquisiciones pues no se presentaron bandas asociadas a daños por sobrecalentamiento ni cambios en la forma del espectro.

La adquisición de los espectros Raman se llevó a cabo utilizando un objetivo de microscopio 20× (NA=0.40 WD = 12 mm) y una rejilla de difracción de 1800 gr/mm, con lo que se logró una resolución de $0.6\text{ cm}^{-1}/\text{pixel}$. El rango de adquisición fue desde 300 hasta 2000 cm^{-1} para captar toda la región de la huella digital de la proteína, con un tiempo de adquisición de 15 segundos.

Se depositaron 50 μL de cada espécimen de p53 dentro de contenedores cilíndricos de aluminio, todos con la misma forma y volumen. Se adquirieron tres espectros de cada muestra de p53 tipo natural y tres espectros para cada tipo de mutante.

La Figura 4.1 muestra un ejemplo de los espectros adquiridos. El espectro en color azul se considera “crudo”, pues no ha sufrido ninguna modificación. Se puede constatar la presencia de la señal de fluorescencia, razón por la cual el espectro se observa inclinado. Para eliminar este efecto, el ruido de frecuencias altas y las diferencias que puedan existir en la intensidad entre espectros de una misma muestra, se llevaron a cabo una serie de procedimientos descritos a continuación, los cuales tienen como principal objetivo facilitar el análisis posterior eliminando señales que no provengan de vibraciones moleculares.

4.1.3 Preprocesamiento de espectros.

Para eliminar la fluorescencia se aplicó el método desarrollado por Zhao et al. [109] llamado Algoritmo Raman de Vancouver, consiste en ajustar una base polinomial al espectro y substraérselo. Al ser un método iterativo, permite mejorar el ajuste del polinomio reduciendo el error en cada iteración. El ruido adicional o ruido de frecuencia alta fue eliminado implementando el método de Savitzky-Golay con un polinomio de grado 5 [27].

En la Figura 5.1 se puede observar la diferencia entre un espectro sin procesar y un espectro cuya fluorescencia y ruido han sido eliminados. Es importante considerar que aplicar estos procedimientos para mejorar el espectro también implica que parte de la información es perdida, por lo que se debe cuidar no perder las características relevantes.

Aunque todos los espectros fueron adquiridos con los mismos parámetros experimentales, existen algunas variaciones en intensidad entre espectros de la misma muestra que no se originan químicamente. El proceso de normalización resuelve estas diferencias haciendo que la intensidad de una banda Raman específica del mismo material sea igual o similar entre dichos espectros [8], [27].

La normalización de un espectro Raman se debe llevar a cabo considerando el tipo de información que se quiere obtener. Por ejemplo, para un análisis cualitativo de las muestras, basta con mantener la forma espectral, sin embargo, cuando se obtendrá información cuantitativa de los espectros y su intensidad, es necesario considerar otras condiciones para hacer la normalización, como la concentración, longitud del camino óptico recorrido por el láser o el tipo de material [27]. Aplicando la Ecuación 4.1 se llevó a cabo la normalización de los espectros por tipo de muestra; esta expresión hace que la máxima intensidad sea igual a 1 y la mínima igual a cero:

$$I_i^{norm} = \frac{I_i - I_{min}}{I_{max} - I_{min}}, \quad (4.1)$$

donde I_i es la intensidad en cada frecuencia capturada en el espectro, I_{min} e I_{max} corresponden respectivamente a la mínima y a la máxima intensidad obtenida para el grupo de espectros de una misma muestra e I_i^{norm} es la intensidad normalizada en cada punto [27]. Este tipo de normalización facilita el análisis cualitativo para identificar bandas.

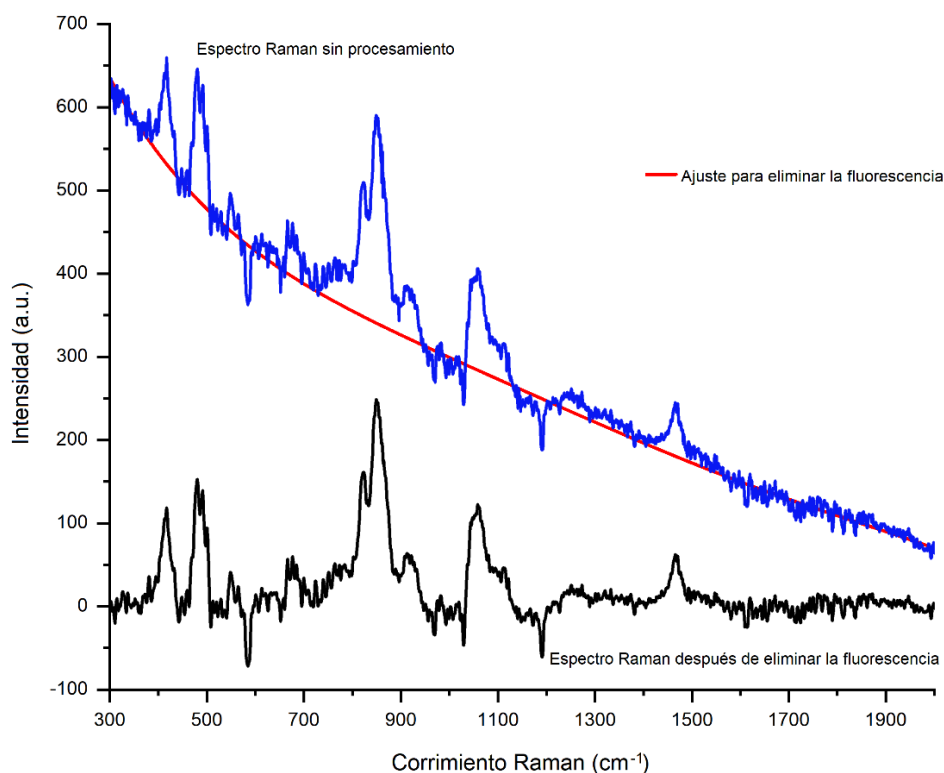


Figura 4.1. Espectros Raman de la proteína p53 tipo natural a 12.6 μM antes y después de la eliminación de la fluorescencia y sin normalización.

4.1.4 Asignación de vibraciones

En la Figura 4.2 se presentan los espectros promedio y normalizados para cada tipo de p53 (se muestran 2 espectros de p53 tipo natural por ser diferente muestra). Como se esperaba, los espectros Raman contienen un conjunto complejo de bandas, algunas de ellas similares en todos los espectros. La Tabla 4.2 contiene la asignación de vibraciones para cada una de las bandas distinguidas, esta asignación se llevó a cabo tomando como referencias estudios previos de proteínas, aminoácidos, dipéptidos y tripéptidos, siguiendo las coincidencias en frecuencias [57], [83], [110]–[112].

Las características observadas en los espectros de p53 natural coinciden con la mayoría de las bandas reportadas previamente por Signorelli et al. [83], sin embargo los espectros aquí reportados fueron medidos en un rango de frecuencia más amplio.

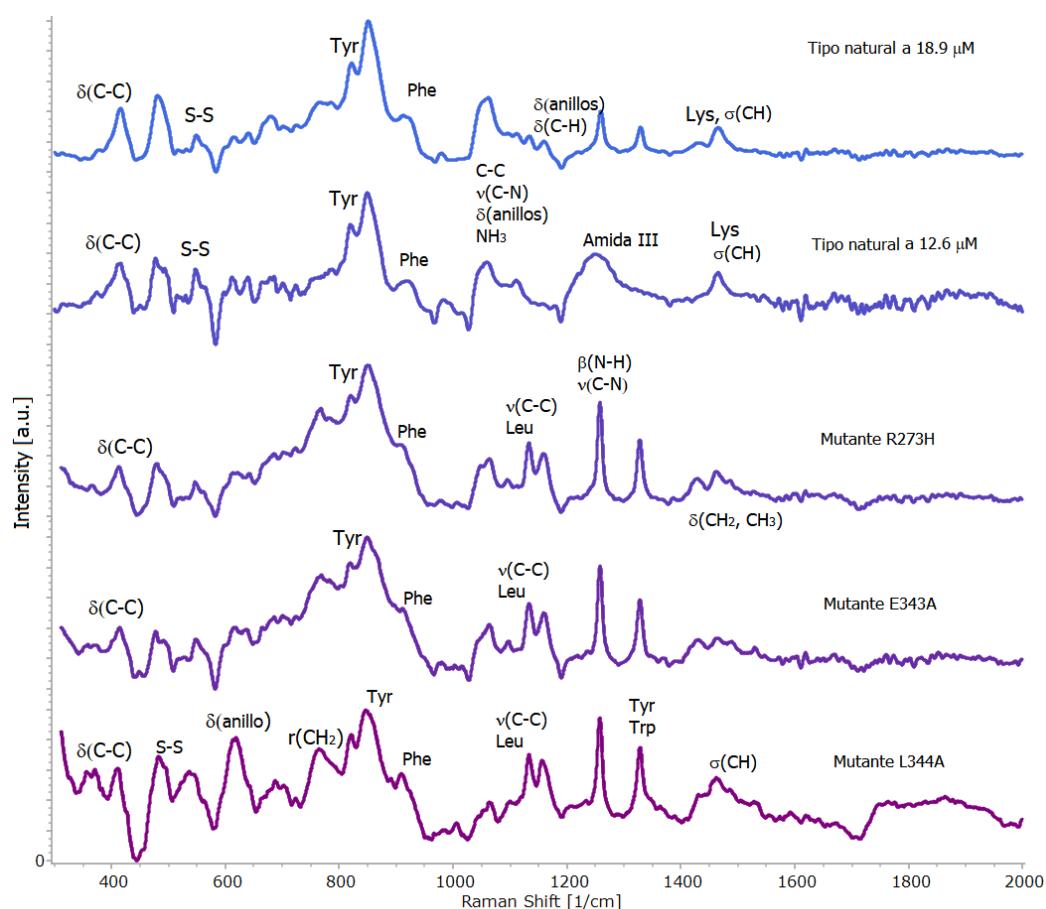


Figura 4.2. Espectros promedio de la proteína p53 tipo natural en dos concentraciones y tres p53 tipo mutante.

Alrededor de las frecuencias 350 – 416 cm^{-1} los cuatro espectros presentan bandas asociadas a la deformación de enlaces C – C, mismo que se asocian a aminoácidos alifáticos como glicina (Gly) y valina (Val); en el espectro de la mutante L344A las bandas menores a 400 cm^{-1} son distinguibles fácilmente, mientras que para la proteína natural solo se observa un pequeño hombro alrededor de 375 cm^{-1} y su banda más prominente para este tipo de vibración es en 416 cm^{-1} .

Todos los espectros presentan una banda alrededor de 480 cm^{-1} , la cual se ha asociado con deformaciones en aminoácidos como treonina (Thr) y fenilalanina (Phe). Cerca de esta frecuencia se distinguen también las bandas asociadas con las vibraciones de enlaces disulfuro (S-S) y deformaciones del esqueleto ($\sim 550 \text{ cm}^{-1}$). Entre 600 y 800 cm^{-1} las bandas son menos definidas, la más prominente está en el espectro de la mutante L344A en 619 cm^{-1} , corresponde a una deformación del anillo de Phe. Todas las mutantes presentan un pico alrededor de 766 cm^{-1} asociado al balanceo de enlaces CH_2 de metionina (Met) e isoleucina (Iso).

En 820 y 850 cm^{-1} se encuentran las bandas más intensas para los cuatro espectros, estas bandas corresponden a vibraciones del aminoácido tirosina (Tyr). En los espectros de la proteína natural estas dos bandas se encuentran más definidas, mientras que en los espectros de las proteínas mutantes son más anchas. Alrededor de 1063 cm^{-1} se encuentran bandas asociadas a vibraciones de Phe, enlaces C-C, deformaciones C-N y NH_3 . Se identifica una marcada diferencia alrededor de 1150 cm^{-1} entre las proteínas del tipo natural y las proteínas mutantes, en particular para las mutantes hay dos bandas bien definidas, una en 1133 cm^{-1} que se asocia a deformaciones de enlaces C-C de leucina (Leu) e Iso y 1159 cm^{-1} correspondiente a vibraciones de los anillos de aminoácidos aromáticos.

La región de la Amida III se encuentra generalmente alrededor de 1250 cm^{-1} ; pero en los espectros de las proteínas mutantes y el espectro de p53 natural a 18.9 μM , se observa una banda intensa y bien definida en 1258 cm^{-1} que corresponde a deformaciones de enlaces C-N y doblamientos de N-H. Posteriormente aparece otro pico bien definido en 1328 cm^{-1} que se asigna a vibraciones de Tyr (tirosina) y Trp (triptófano). Finalmente, para los espectros de p53 tipo natural se tiene una banda en 1465 cm^{-1} , que aparece menos definida para los espectros de p53 mutante, esta banda corresponde a vibraciones de lisina (Lys) y tijereos de CH.

Natural		Mutante			Vibración
18.9 μ M	12.6 μ M	R273H	E343A	L344A	
			358*	356*	δ (C-C) de aminoácidos alifáticos (Gly, Val)
379*	373*	366*	374*	370	
		413	414	411	
416	416				
480	477	480	478	482.4	Thr, δ (C-C-C-C) de Phe
	493*		493*		Gly, Cys
			528	536.2	Enlaces S-S, δ (esqueleto)
549	548	547	550		
	612				δ (O-C=O) de Phe
614.5*		614*	615*		
		622*		619	δ (anillo) de Phe
640*	640	642*	637*		Tyr, ν (C-S) y σ (CO ₂ ⁻) de Met
679.5	678*				C-S de aminoácidos alifáticos
		686*	686*	687*	
700*	700*	701*	701*	701*	S-H, ν (C-S) de Met
765*		767	767	765	Tyr, r(CH ₂) de Met, Iso
786*	784*	780*	782(h)		ν (CH) de His, ν (OCO ⁻)
821	820	820	819	822	Tyr, ν (C-C) de His
851	849	852	850	846	Tyr
912(h)		906.8(h)	911(h)	908	δ (C-C) de Phe
	921				ν (C-C),
979.6*	984	977*	980*	986*	
		1007*	1003*	1006*	δ (anillo) de Phe
		1049(h)	1044(h)	1044 (h)	
1060	1057	1064	1063	1064	C-C, ν (C-N), δ (anillos), NH ₃
1097*		1096*	1097	1100*	r(CH ₂) de Pro, ν (C-N) de Met y Cys, Aminoácidos básicos.
1112*	1110				Tyr
1134*		1133	1134	1133	ν (C-C) de Leu, Iso
1160*	1157*	1159	1159	1156	Anillos de Tyr, Trp, Phe, δ (C-H)
1259	1250	1258	1258	1258	Amida III, β (N-H), ν (C-N)
1329		1328	1328	1328	Tyr, Trp
1432*		1430	1431*	1431*	δ (CH ₂ , CH ₃), Leu
1466	1465	1463	1467*	1462	Lys, σ (CH)
		1488*		1489*	σ (CH ₂) de His, ν (CNC)
			1530*	1528*	ν (C-N), β (N-H)
1620*	1619*		1619*		δ (anillo) de Phe y Trp

Tabla 4.2. Bandas Raman identificadas en los espectros experimentales de p53 tipos natural mutantes. Las frecuencias marcadas con asterisco (*) son de poca intensidad, mientras que las resaltadas en negritas son de picos más prominentes; la letra (h) indica que la frecuencia es un hombro en una banda. Los tipos de vibraciones son: δ = deformación; r = balanceo; σ = tijereo; ν = estiramiento; β = doblamiento. [54], [80], [106]–[108].

4.2 LÍMITE DE DETECCIÓN DE P53 NATURAL MEDIANTE ESPECTROSCOPIA RAMAN.

La intensidad de los espectros Raman está directamente relacionada con la cantidad de material que se está excitando: entre más moléculas sean alcanzadas por la radiación, mayor número de fotones serán esparcidos y las bandas de las vibraciones características serán más prominentes [27]. Cuando la concentración es muy baja, el esparcimiento es menor y las bandas Raman tienden a desaparecer o aplanarse, por lo que existe un límite de detección, que es la concentración mínima a la cual el espectrómetro alcanza a detectar esparcimiento de la muestra. Este límite también depende de la sensibilidad del detector, la intensidad y longitud de onda de la radiación incidente y el montaje experimental.

El límite de detección (LDD) se calcula de acuerdo con:

$$LDD (\mu M) = 3 \frac{s}{m} \quad (4.2),$$

donde m es la pendiente del mejor ajuste lineal entre la concentración predicha y la concentración medida experimentalmente y s es una medida de la desviación promedio de las concentraciones predichas por el modelo (c_p) y las concentraciones medidas experimentalmente (c_e) [103], [113]:

$$s = \sqrt{\frac{\sum (c_p - c_e)^2}{n - 2}} \quad (4.3).$$

Para obtener el límite de detección para p53 medido por espectroscopía Raman se emplearon dos métodos de regresión (univariable y multivariable); en ambos se relacionó la concentración con la señal Raman. El objetivo es comparar los resultados obtenidos de cada método y determinar cuál de ellos tiene mejor desempeño.

4.2.1 Preparación de muestras de p53 a diferentes concentraciones.

Partiendo de la muestra de p53 natural a 12.6 μM se prepararon diluciones adicionales utilizando solución buffer salina (PBS, pH 7.5). Las concentraciones obtenidas fueron: 8.40, 4.20, 2.8, 2.1 y 1.68 μM . Se mantuvieron en refrigeración ($\sim 3^\circ\text{C}$) hasta ser usadas.

4.2.2 Adquisición y preprocesamiento de espectros Raman

Cada muestra se depositó en una cavidad de aluminio. Todas las cavidades tuvieron las mismas dimensiones para asegurar que el láser tuviera el mismo camino óptico en todas las muestras; al mantener este parámetro constante, el número de moléculas excitadas depende directamente de la concentración, y ésta a su vez será directamente proporcional a la señal obtenida [18], [27].

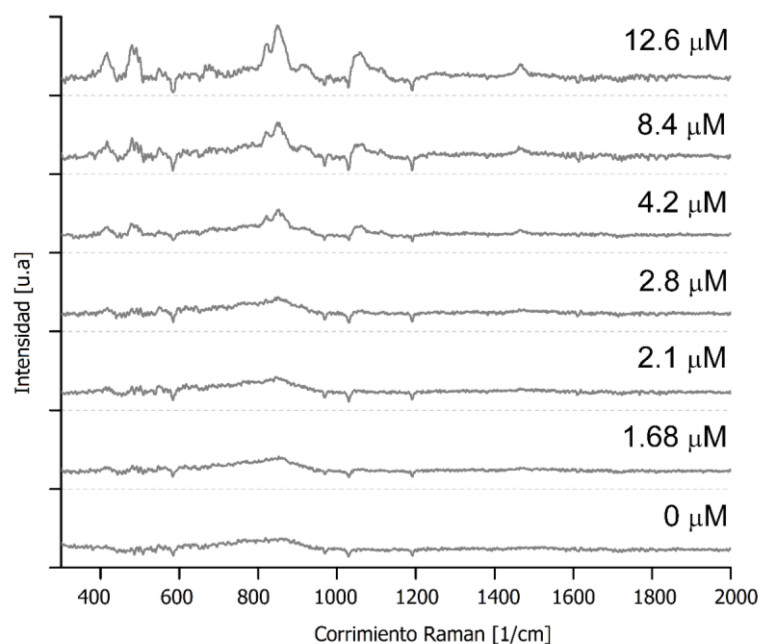


Figura 4.3. Espectros promedio de p53 tipo natural a diferentes concentraciones y espectro de la solución buffer como referencia de concentración cero.

De cada concentración se adquirieron dos espectros Raman y dos espectros adicionales de la solución PBS, para un total de 14 espectros con concentraciones de 0 a 12.6 μM ; en todas las adquisiciones se mantuvieron los mismos parámetros

descritos en la sección 4.2 (intensidad, tiempo de adquisición, objetivo de microscopio y rejilla). Los espectros obtenidos fueron procesados posteriormente para eliminar la fluorescencia con el mismo procedimiento descrito en la sección 4.3. La normalización se realizó aplicando la Ecuación 4.1, utilizando como intensidad máxima e intensidad mínima los valores globales para el grupo de 14 espectros.

4.2.4 Regresión univariable

Para calcular el límite de detección mediante análisis univariable, se llevaron a cabo dos ajustes lineales con el modelo:

$$c(x) = ax + b \quad (4.4),$$

donde c es la concentración y x la variable obtenida del espectro. Para el primer ajuste x correspondió a la intensidad del pico más prominente del espectro Raman, para el segundo ajuste x fue el ancho medio a altura máxima (FWHM).

Con la función de predicción $c(x)$ se calcularon los nuevos valores de concentración (C_p), y se hizo un nuevo ajuste lineal:

$$C_p = mC_e + E \quad (4.5),$$

con el cual se calculó el límite de detección aplicando las Ecuaciones 4.2 y 4.3.

a) Concentración en función de intensidad máxima

La banda más intensa de los espectros Raman de p53 natural es en 850 cm^{-1} , en esta posición se tomaron los valores de la intensidad normalizada para cada uno de los 14 espectros y se realizaron los ajustes lineales mediante Matlab®.

Las Figuras 4.4 (a) y (b) contienen los ajustes lineales $c(x)$ y C_p respectivamente. El límite de detección calculado fue de $0.897 \mu\text{M}$.

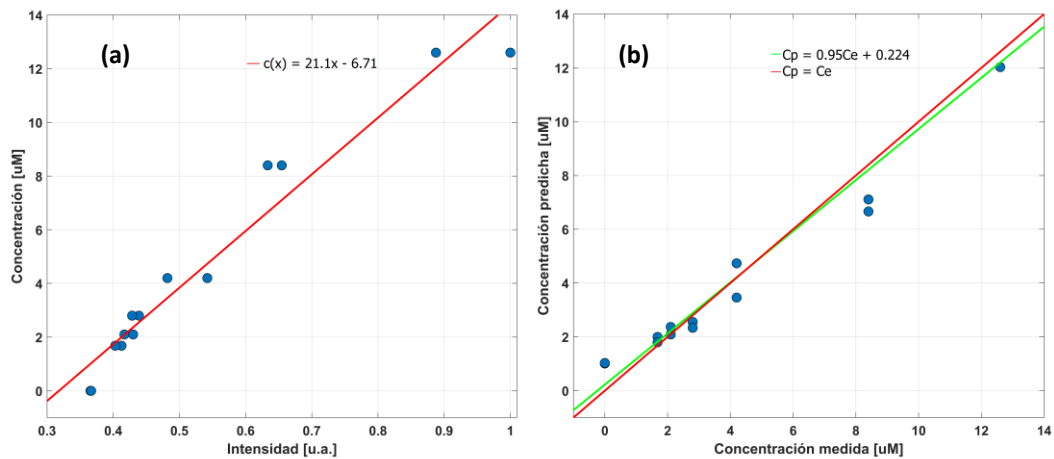


Figura 4.4. (a) Ajuste para predecir la concentración en función de la intensidad. (b) Ajuste entre concentración predicha y concentración medida experimentalmente.

b) Concentración en función de FWHM.

Para calcular el FWHM de cada espectro, primero se hizo un ajuste en el pico 850 cm^{-1} mediante la aproximación de Lorentz en Origin®.

En este ajuste, la relación entre la concentración y la variable x es inversamente proporcional, y el ajuste produce una pendiente negativa como se observa en la Figura 4.5(a). El límite de detección calculado fue 1.34 µM .

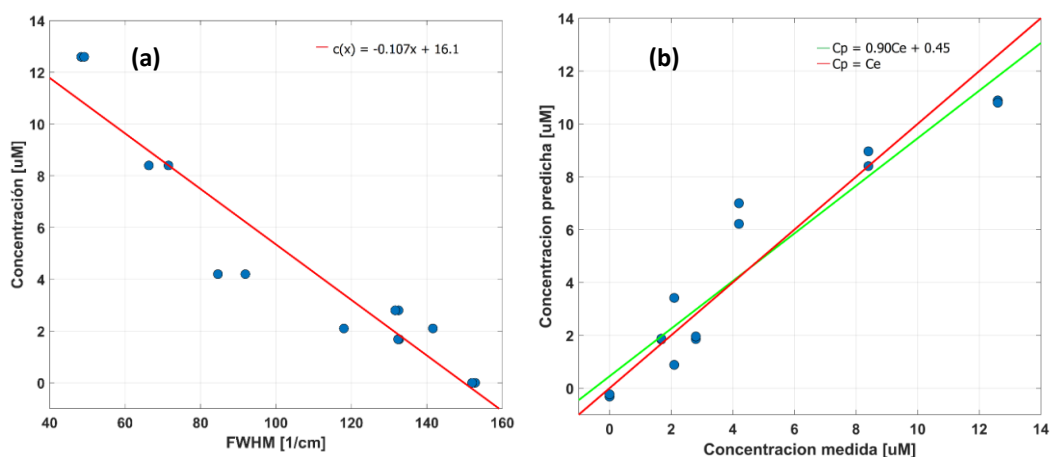


Figura 4.5. (a) Ajuste para predecir la concentración en función de FWHM. (b) Ajuste entre concentración predicha y concentración medida experimentalmente.

4.2.5 Regresión multivariable

El método PLSR fue implementado en Matlab® para obtener un modelo multivariable que permitiera predecir las concentraciones en función de las características estadísticamente más relevantes de los espectros. De acuerdo con la ecuación 2.10 del Capítulo 2, este modelo predice las nuevas concentraciones (Y) en función de nuevas variables latentes (X) y realiza validación cruzada para probar el ajuste; la elección del número de nuevas variables se llevó a cabo mediante análisis de componentes principales, descartando con esto variables que representan poca contribución estadística. En este caso se mantuvieron dos componentes para el ajuste, que suman 99.8% de la varianza.

En este modelo, el coeficiente de determinación R^2 indica que tan cerca está el modelo de dar una predicción perfecta y es usado como el parámetro m para predecir el límite de detección, que se obtuvo igual a $1.02 \mu\text{M}$. La Figura 4.6 presenta las concentraciones predichas por el modelo y su dispersión de la línea de predicción perfecta.

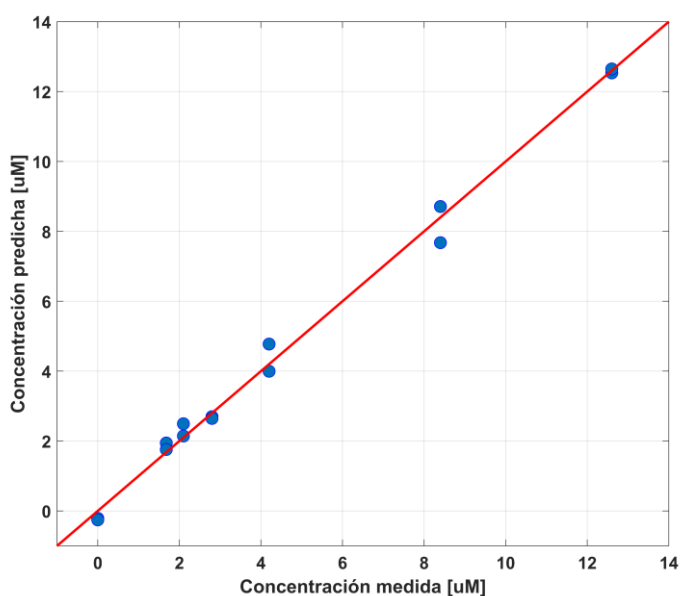


Figura 4.6. Concentraciones experimentales versus concentraciones predichas por el modelo PLSR, la línea roja representa el ajuste ideal donde $C_p = C_e$.

4.2.6 Comparación entre los ajustes para la curva de calibración de concentraciones.

La Tabla 4.3 presenta los resultados obtenidos para el límite de detección con los tres modelos de predicción de concentraciones. Como métrica de desempeño, se calculó la raíz del error cuadrático medio (RMSE²), la cual es una medida promedio de qué tan lejanos se encuentran los datos experimentales de los datos predichos.

Además de tener un mejor desempeño y ahorrar tiempo de análisis y procesamiento, la principal ventaja de utilizar PLSR es que la predicción se realiza con las variables más relevantes de los espectros y descarta automáticamente aquellas menos significativas.

Método	Parámetro m	Parámetro S	LDD (μM)	RMSE(μM)
Ajuste lineal con intensidad	0.9506	0.2842	0.8970	0.9116
Ajuste lineal con FWHM	0.9006	0.4032	1.3432	1.2932
PLSR	0.9942	0.3387	1.022	0.3136

Tabla 4.3. Parámetros y resultados del cálculo del límite de detección

² Siglas del inglés *root mean square error*.

Capítulo 5.

CLASIFICACIÓN DE P53 MEDIANTE MÉTODOS DE ANÁLISIS MULTIVARIABLE Y APRENDIZAJE AUTOMÁTICO

No todas las variantes de la proteína p53 tienen las mismas implicaciones en la progresión del cáncer y la expresión de p53 tipo natural no necesariamente se asocia con desarrollo de tumores [75], [76]. Debido a esto los métodos de detección deben garantizar que son capaces de diferenciar entre los tipos de esta proteína.

Los métodos tradicionales de detección de p53 son altamente sensibles y específicos, pero la mayoría requieren obligatoriamente anticuerpos específicos (etiquetas) para detectar cada tipo de p53 y complicadas técnicas de preparación [114]. La espectroscopía Raman, por el contrario, es una técnica sencilla que puede identificar individual o simultáneamente diferentes tipos y no requiere ningún tipo de etiquetado molecular.

Los espectros Raman pueden tener ocultas ciertas características esenciales que es posible revelar mediante análisis multivariable, como PCA, permitiendo diferenciar entre biomoléculas aparentemente similares. Adicionalmente los métodos multivariantes permiten mantener variables estadísticamente más significativas con lo que es posible generar clasificaciones más acertadas mediante aprendizaje automático.

En este capítulo se describe la aplicación de PCA para analizar la agrupación de los espectros de p53 por tipos e identificar los marcadores moleculares que los distinguen. Además, se presenta la implementación de SVM para clasificar automáticamente los cuatro tipos de p53 antes caracterizados.

5.1 ANÁLISIS DE COMPONENTES PRINCIPALES

PCA se realizó utilizando Matlab con la función $pca()$, que lleva a cabo el siguiente proceso esquematizado en la Figura 5.1 [10], [48], [115], [116]:

1. La matriz de N espectros y M números de onda es un arreglo de N renglones y M columnas, cada columna corresponde a una variable y cada renglón a una observación.
2. El arreglo es centrado: cada columna tendrá media igual a cero y desviación estándar igual a 1.
3. Se calcula la matriz de covarianza de la nueva matriz de datos centrados y se obtienen los valores y vectores propios esta matriz.
4. Los valores propios son ordenados de mayor a menor y simultáneamente se ordena la matriz de vectores propios; en esta matriz de $N \times N$ coeficientes cada vector columna es la puntuación (*PC score*) de un componente principal y cada uno de sus valores representa a un espectro de la matriz original.
5. Se multiplica la matriz de *PC scores* por la matriz de datos centrados, obteniendo la matriz de nuevas variables -también llamados *PC loadings*.

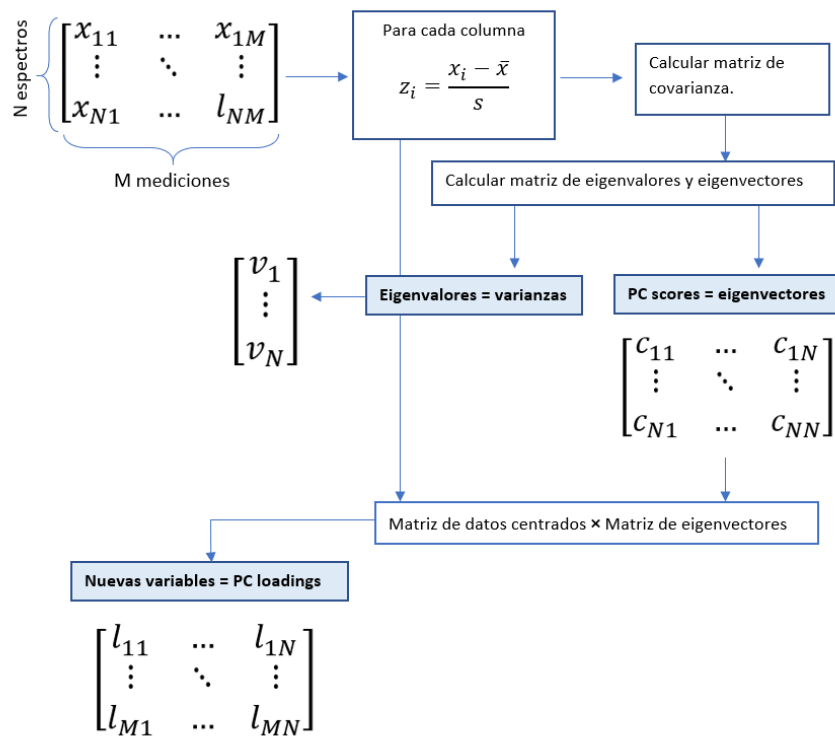


Figura 5.1. Diagrama de flujo para la implementación de PCA mediante el cual se obtienen las matrices *PC scores* y *PC loadings* cuya relevancia está dada por el vector de eigenvalores.

Se implementó PCA para los siguientes dos grupos de espectros de p53:

- Grupo 1 con 25 espectros: 16 tipo natural y 9 mutantes (3 de cada tipo).
- Grupo 2 con 9 espectros: 3 de cada tipo de mutante.

Previo a PCA, todos los espectros fueron preprocesados mediante el algoritmo de Vancouver para eliminar su fluorescencia y fueron normalizados respecto a la intensidad máxima.

5.2 RESULTADOS Y DISCUSIÓN DE PCA

Para cada resultado de PCA (de los dos grupos) se eligieron los *PC scores* más representativos en términos de porcentaje de varianza, se graficaron uno contra otro para observar las tendencias de agrupación en el nuevo espacio y se interpretaron dichas gráficas tomando en cuenta las siguientes propiedades [10], [48]:

1. Cuando los valores del correspondiente *PC score* son similares entre sí (poca variación), los puntos graficados se mantendrán cercanos. Dado que cada punto representa un espectro, la cercanía de los puntos indica que los espectros son similares entre sí y en la gráfica se observará agrupación.
2. Cuando hay variabilidad en los valores del *PC score* entonces los espectros tienen diferencias importantes y los puntos graficados aparecerán distantes.

Para el grupo de 25 espectros, los primeros 3 componentes contribuyen con el 87.9% de la varianza: el PC 1 contribuye con 53.5%, el PC 2 explica el 20.8% y el PC 3 contiene el 13.6%; en el resto de los 22 componentes se reparte el 12.1%.

La Figura 5.2a presenta el espacio de los *PC scores* para el primer y el segundo componente principal. En este espacio y en particular en el eje del PC 2, existe separación entre el grupo de las muestras mutantes y el grupo de p53 tipo natural: los espectros de las mutantes tienen puntuaciones positivas y los espectros del tipo natural puntuaciones negativas. En el PC 1 no hay separación entre los tipos, pero en este componente se observa que los espectros con concentraciones similares permanecen cercanos.

La Figura 5.2b presenta el espacio formado por el PC 2 y el PC 3. En el PC 3 existe poca variabilidad entre los espectros, nuevamente la distribución tiende a ser por concentración en los espectros de p53 natural, además se identifica un espectro tipo L344A cuya separación es evidente y que podría considerarse atípico.

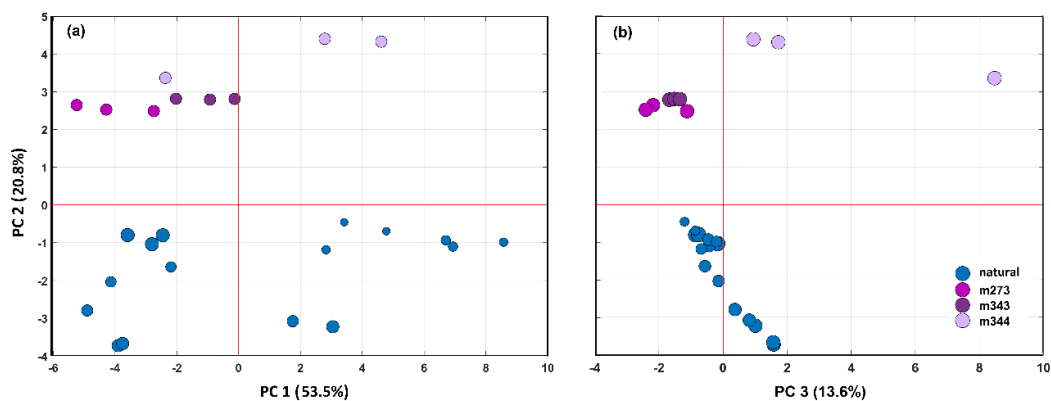


Figura 5.2 Gráficas de *PC scores* para el grupo de 25 espectros. (a) *PC score* 1 vs *PC score* 2. (b) *PC score* 3 vs *PC score* 2. El tamaño de los puntos para p53 tipo natural es una representación cualitativa de su concentración.

Para este grupo (25 espectros) se graficaron los *PC loadings* 1, 2 y 3 (Figura 5.3) contra los valores de corrimiento Raman de los espectros originales. Dado que los *PC loadings* indican cuánto contribuye una variable a cierto componente principal, estas gráficas revelan las vibraciones moleculares que determinan las principales diferencias entre los grupos que se forman en el espacio de los *PC scores*. Los coeficientes más alejados de cero corresponden a las bandas que mayor contribución tienen; los coeficientes cercanos a cero corresponden a bandas que no tienen una variación considerable entre los espectros y poco influyen en el modelo [117], [118].

La variación de los *PC loadings* del primer componente principal es menor que la variación en los *loadings* consecutivos de PC 2 y PC 3. Las variables con *loadings* cercanos a cero corresponden a las bandas más intensas y presentes en casi todos los espectros: 480, 852, 1062, 1258 y 1327 cm^{-1} , cuyas vibraciones correspondientes se reportan en la Tabla 4.2.

El PC 2 tiene *loadings* con mayor variabilidad. Los máximos que aparecen en este nuevo espectro son las bandas que distinguen los espectros de las p53 mutantes de los espectros de p53 natural. Las vibraciones en 1258 y 1327 cm^{-1} aunque están presentes en algunos espectros de p53 natural, son mucho más prominentes para los tipos mutantes.

La presencia de las bandas de leucina en 1135 cm^{-1} y las vibraciones de algunos anillos en 1155 cm^{-1} también tiene una contribución relevante para los grupos de mutantes. Algunas vibraciones que no habían sido identificadas a simple vista y que se hacen presentes para los espectros mutantes son: 533, 587, 622, 759, 1031 y 1193 cm^{-1} , que respectivamente se podrían asociar a: vibraciones de alanina, enlaces S-S, vibraciones de ácido glutámico, deformaciones de anillos, vibraciones C-N de glicina y lisina, y vibraciones de leucina [57], [110]. Particularmente la banda de alanina (533 cm^{-1}), del ácido glutámico (622 cm^{-1}) y de leucina (1193 cm^{-1}) son importantes pues los cambios en estos aminoácidos distinguen a las mutantes que aquí se estudian.

En este mismo PC 2, los valores cercanos a cero corresponden a bandas de vibración que se encuentran presentes la mayoría de los espectros: 820, 1005, 1429, 1531, 1487, 1615 cm^{-1} . Los coeficientes negativos se relacionan con las bandas Raman más características de p53 tipo natural, las bandas 338 y 420 cm^{-1} no había sido identificadas en los espectros y se pueden asociar a deformaciones de enlaces C-C para aminoácidos alifáticos [57]; la banda de 853 cm^{-1} , que es la más intensa en todos los espectros, en los de p53 tipo natural tiene mayor contribución. Finalmente se observa la banda en 1057 cm^{-1} , previamente identificada en la Tabla 4.2.

Los *loadings* para el PC 3 podrían revelar las diferencias entre la mutante L344A y las otras dos mutantes (R273H y E343A), pues el PC score 3 divide a estos dos grupos con valores positivos para la mutante 344 y negativos para la 273 y la 344. Sin embargo, no se puede analizar este nuevo espectro sin considerar que también se tiene contribución de los espectros de p53 tipo natural.

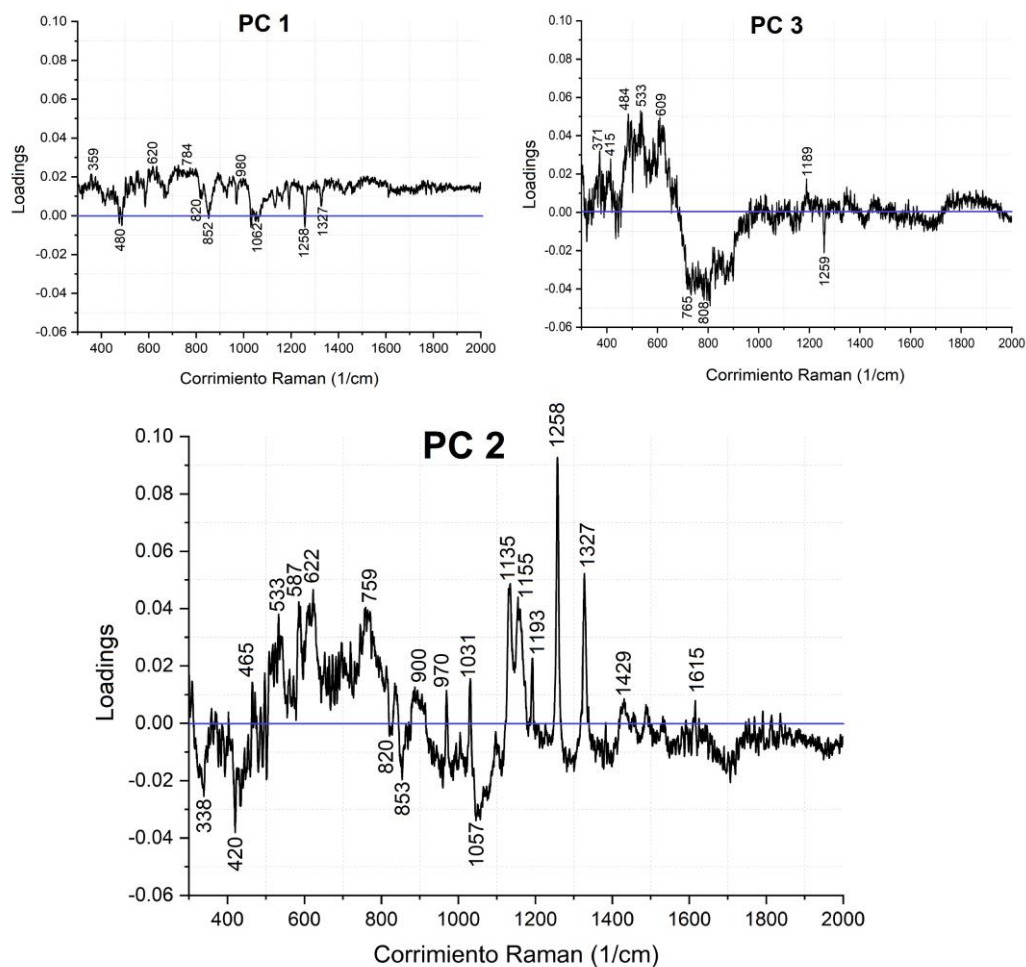


Figura 5.3 Graficas de PC loadings (nuevas variables) obtenidos del análisis de componentes principales para el grupo de 25 espectros.

Para confirmar las diferencias vibracionales entre los tres tipos de mutantes, se analizaron mediante PCA este grupo de espectros (solo conformado por proteínas tipo mutantes). En este caso la implementación de PCA arrojó que el primer componente contiene el 51.7% de la varianza total, el segundo componente contiene el 34.5% y el resto 13.8%. Los primeros dos componentes se graficaron en el espacio de los PC scores (Figura 5.4); en este espacio nuevamente es evidente el comportamiento atípico de uno de los espectros de la mutante 344, además de que en el eje del PC 1 los espectros de esta mutante se separan de las mutantes 343 y 273; teniendo puntuaciones mayores.

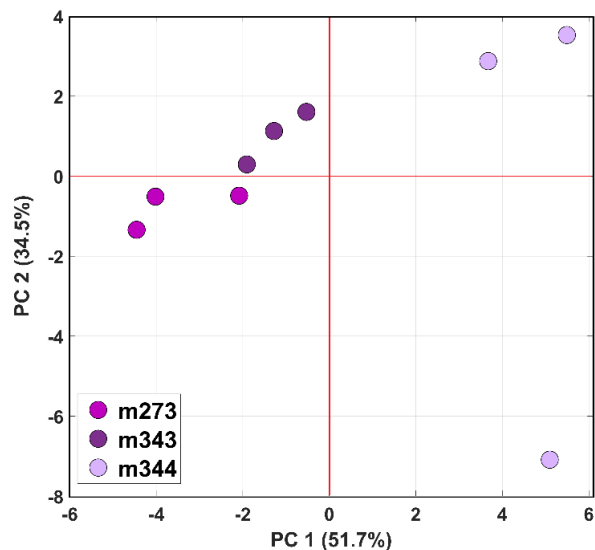


Figura 5.4. Gráfica de PC scores del grupo de nueve espectros de mutantes.

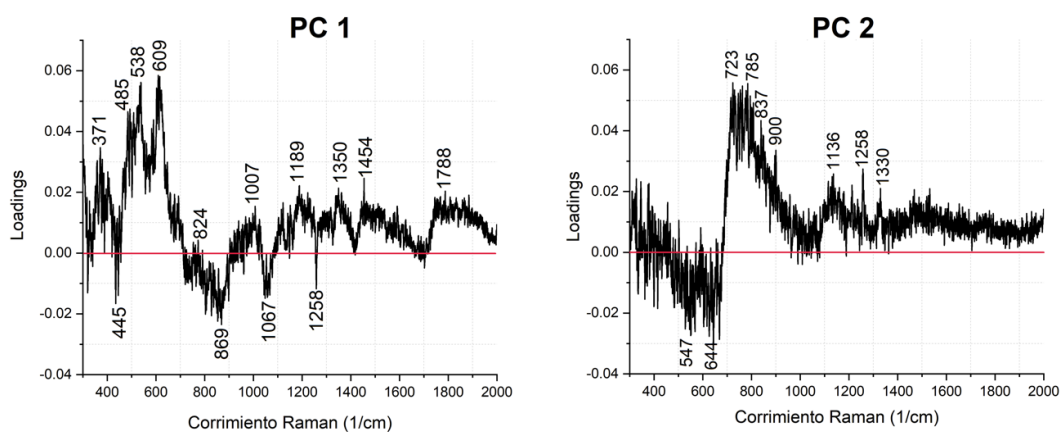


Figura 5.5. PC loadings 1 y 2 del grupo de espectros mutantes.

Para este grupo de 9 espectros se graficaron los dos *PC loadings* más relevantes para identificar las bandas de vibración de mayor contribución. En la Figura 5.5, los *loadings* positivos del PC 1 permiten confirmar las bandas que distinguen a la mutante L344A de las mutantes R273H y E343A (*loadings* negativos). La banda en 609 cm^{-1} , que se asocia a vibraciones de enlaces CH_2 [112], no se había identificado en el espectro de la mutante 344.

Otras bandas con contribución relevante son 371 , 485 y 538 cm^{-1} , que sí se habían identificado previamente (Tabla 4.2). Los *loadings* negativos confirman bandas que se presentan con mayor intensidad en las mutantes R273H y E343A y revelan las bandas 445 y 896 cm^{-1} , respectivamente asociadas a deformaciones y vibraciones

C-C [57]. Las bandas con *loadings* cercanos a cero corresponden a vibraciones presentes en los tres tipos de espectros.

En la gráfica de los *PC loadings* del PC 2 los valores positivos pueden asociarse a bandas características de las mutantes E343A y L344A, mientras que los valores negativos son para bandas que caracterizan a la mutante R273H.

5.3 CLASIFICACIÓN MEDIANTE SVM

El clasificador SVM se implementó para tener una clasificación cuantitativa que diferencie automáticamente entre las clases de la proteína p53. Este clasificador fue entrenado con los componentes principales más relevantes que se obtuvieron mediante PCA. Mientras que PCA permite analizar cualitativamente el comportamiento de datos espectrales y reducir sus dimensiones eliminando datos que no tienen contribución estadística relevante; SVM como método de clasificación puede determinar cuantitativamente a qué clase pertenece un espectro.

Algoritmo del clasificador SVM:

Paso 1. Implementación de PCA: El conjunto de datos que fueron usados para entrenar el clasificador fueron los PC scores que mayor contribución estadística tienen.

Paso 2. Entrenamiento del clasificador: Se implementó un clasificador multiclase “*one vs all*” con corrección de errores de salida en Matlab (función *fitcecoc*), que consiste en crear k modelos binarios de clasificación SVM dividiendo los datos en dos clases (positivos y negativos).

El número de modelos k es el número de clases (k = 4 clases), los cuales quedaron distribuidos como se especifica en la Tabla 5.1. Iterativamente el algoritmo aprende y prueba los modelos buscando reducir el error [52].

SVM	Clase positiva	Clase negativa
Modelo 1	Tipo natural (wt)	Mutantes 273, 343 y 344
Modelo 2	Mutante 273	Tipo natural, mutantes 343 y 344
Modelo 3	Mutante 343	Tipo natural, mutantes 273 y 344
Modelo 4	Mutante 344	Tipo natural, mutantes 273 y 343

Tabla 5.1. Distribución de los clasificadores binarios SVM.

Paso 3. Validación del clasificador: Se implementó validación cruzada (mediante *crossval* en Matlab) para estimar para probar el modelo y estimar el error. El método de validación cruzada elegido fue “*leave-one-out*”³, que realiza N pruebas dejando en cada una de estas uno de los datos como prueba y el resto para estimación. El número de pruebas N es igual al número de datos en el modelo [52].

Paso 4. Métricas de desempeño. Se calculó la matriz de confusión con los valores predichos y los valores reales para obtener los verdaderos positivos (VP), falsos positivos (FP), falsos negativos (FN) y verdaderos negativos (VN). Con estas cantidades se calcularon las métricas de desempeño siguientes [119].

- Exactitud: proporción de predicciones que fueron correctas.

$$Acc = \frac{VP + VN}{VP + FP + VN + FN} \quad (5.1).$$

- Sensibilidad: mide la proporción de positivos correctamente identificados.

$$Se = \frac{VP}{VP + FN} \quad (5.2).$$

- Especificidad: mide la proporción de negativos identificados correctamente.

$$Sp = \frac{VN}{VN + FP} \quad (5.3).$$

- Precisión: proporción de verdaderos positivos del grupo de resultados positivos.

$$Pc = \frac{VP}{VP + FP} \quad (5.4).$$

³ Validación dejando uno de los datos fuera

- F-score: resume la precisión y la sensibilidad.

$$Fscore = \frac{2 (Se * Pc)}{Se + Pc} \quad (5.5).$$

5.4 RESULTADOS Y DISCUSIÓN DEL CLASIFICADOR SVM

Se implementó el clasificador SVM multiclase para el grupo de 25 espectros. El primer entrenamiento del modelo se llevó a cabo manteniendo los primeros PC scores que sumaran 95% de la varianza. Este porcentaje se eligió en base a la gráfica de la Figura 5.6, donde el mayor decrecimiento se tiene en los primeros 8 componentes, se eligió un componente más para finalmente incluir 9 espectros.

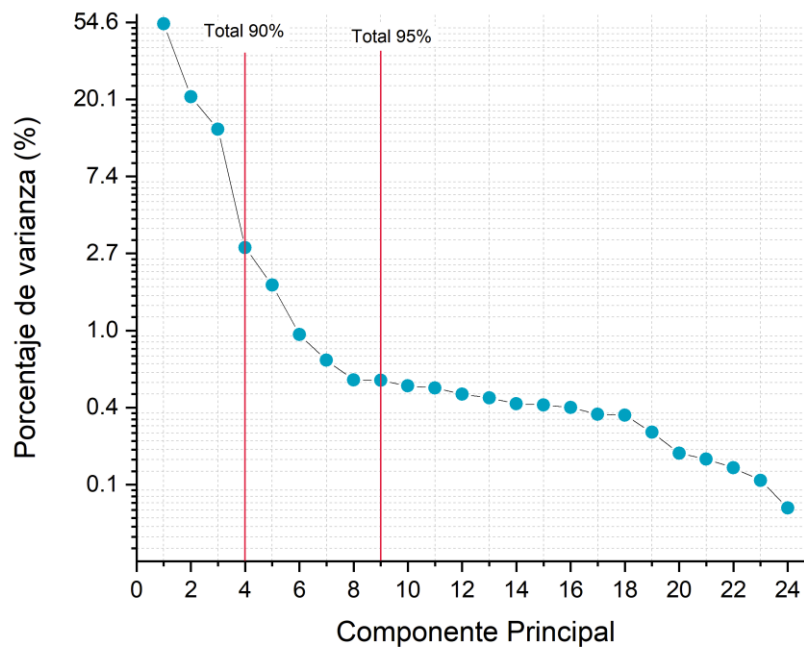


Figura 5.6. Porcentajes de varianza que cada componente aporta al modelo de PCA.

La Figura 5.7 presenta la matriz de confusión con los resultados de esta clasificación y las métricas de desempeño. El resultado es una clasificación perfecta (error = 0), lo cual puede ser efecto de una sobreestimación por la cantidad de componentes principales utilizados.

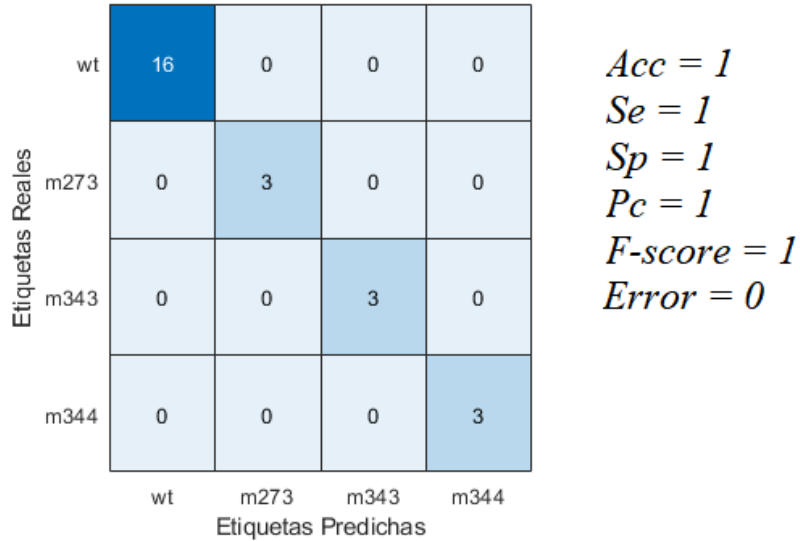


Figura 5.7. Matriz de confusión y métricas para el clasificador entrenado con 9 PC scores del grupo de 25 espectros.

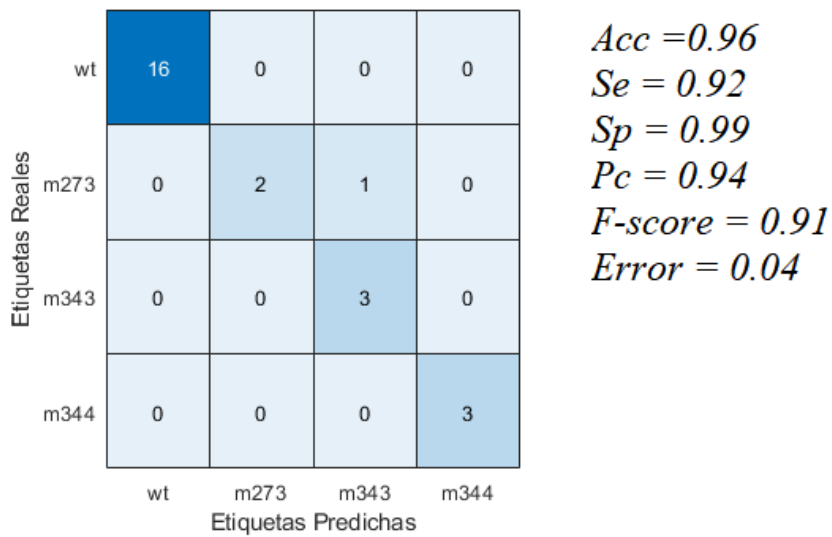


Figura 5.8. Matriz de confusión y métricas para el clasificador entrenado con 4 PC scores del grupo de 25 espectros.

Para evitar la sobreestimación, nuevamente se entrenó al clasificador con el grupo de 25 espectros, pero modificando la cantidad de PC scores a 4, los cuales corresponden a 90% de varianza. La Figura 5.8 presenta los resultados de este modelo, con 96% de exactitud, 92% de sensibilidad y 99% de especificidad, el modelo en general parece tener un desempeño óptimo, presentando solo una

clasificación errónea (mutante 273 clasificada como 343); sin embargo, se trata de un modelo entrenado con clases desbalanceadas.

Para este grupo de 25 espectros, 16 de estos corresponden a p53 tipo natural, esta cantidad representa más del 60% de los datos con los que se entrena al clasificador. Tener clases desbalanceadas puede introducir algún tipo de sesgo tanto en el resultado de la clasificación como en la interpretación de los resultados. El modelo tenderá a clasificar los datos en la clase mayoritaria, y aunque sean correctamente asignados logrando alta exactitud, el desempeño del modelo será realmente limitado [119]. Debido a esto y aunque aparentemente los primeros entrenamientos (con el grupo de 25 espectros) presentaron buen desempeño, las métricas no pueden ser consideradas confiables.

Nuevamente se entrenó el clasificador SVM multiclase, pero con un grupo balanceado de espectros: 12 espectros en total, 3 de cada tipo. Se implementó manteniendo los 6 primeros PC scores que explican el 95% de la varianza, la Figura 5.9 presenta los resultados obtenidos. Con una especificidad de 97%, sensibilidad del 92% y error de 0.08%, este podría considerarse el mejor de los resultados y con mayor confiabilidad, no solo por las métricas obtenidas, sino porque las clases se encuentran balanceadas. En este modelo solo hubo una clasificación errónea, la proteína L344A fue identificada como tipo E343A.

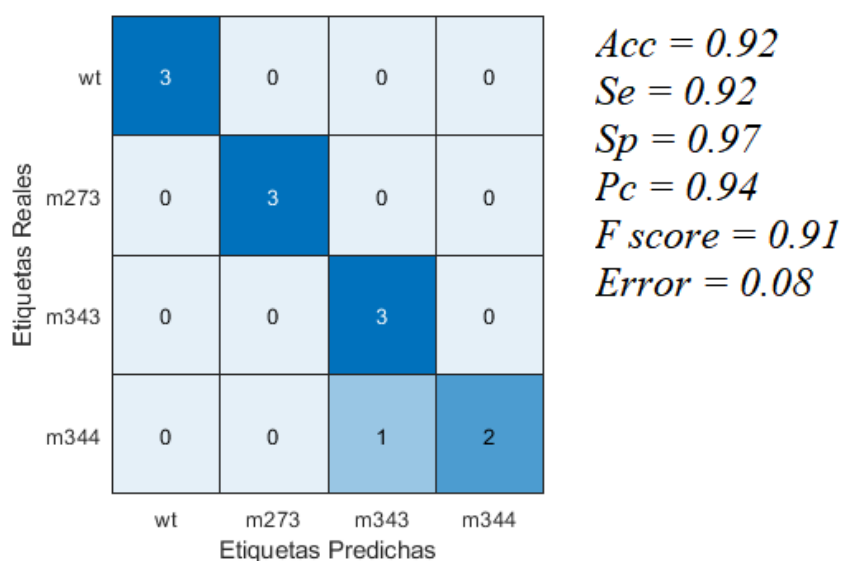


Figura 5.9. Matriz de confusión y métricas para el clasificador entrenado con 6 PC scores del grupo de 12 espectros.

Con fines comparativos, se entrenó nuevamente al clasificador reduciendo el número de componentes principales para mantener el 90% de varianza total. La Figura 5.10 tiene las métricas de desempeño y la matriz de confusión de este modelo.

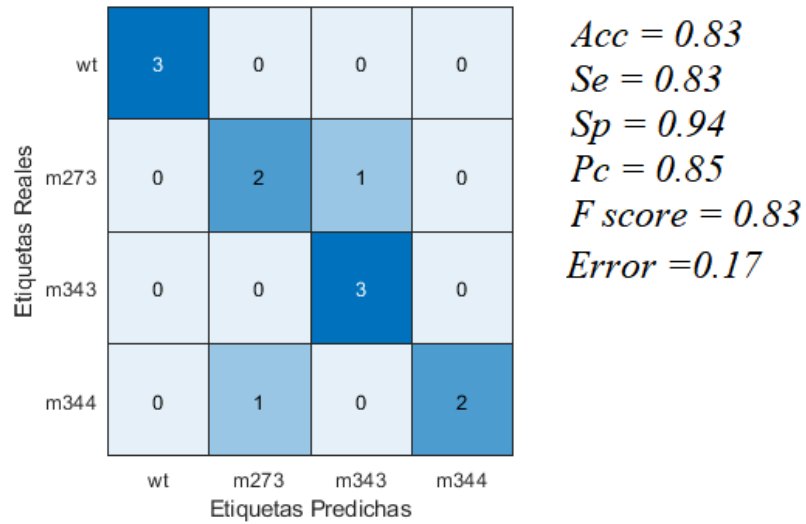


Figura 5.10. Matriz de confusión y métricas para el clasificador entrenado con 4 PC scores del grupo de 12 espectros.

Capítulo 6.

DETECCIÓN DE L-ASPARAGINA MEDIANTE SERS

La detección de un biomarcador específico en concentraciones fisiológicas mediante espectroscopía Raman, requiere en la mayoría de los casos la aplicación de SERS. Para que su aplicación sea eficiente es necesario la elección de sustratos compatibles con la biomolécula de interés (que se enlacen a la molécula o que permitan mantenerla muy cerca) y en los que existan plasmones de superficie con una frecuencia de resonancia similar (o cercana) a la frecuencia de excitación.

Para caracterizar la señal SERS y posteriormente lograr identificar a la biomolécula, es necesario considerar la respuesta de la superficie plasmónica sin el analito y el espectro Raman del analito sin la superficie. Se debe tener en cuenta que los espectros SERS no siempre son iguales al espectro Raman de la biomolécula, e incluso pueden cambiar en función de la posición de medición debido a la orientación de la molécula, los enlaces formados con el metal y la presencia de puntos de mayor concentración de campo eléctrico (*hot spots*) [5].

En este capítulo se presenta la aplicación de SERS para el biomarcador L-Asparagina. Se utilizaron 4 tipos de superficies metálicas; una de ellas fabricada a partir de partículas de oro y las otras tres consistieron en arreglos metálicos que originalmente fueron diseñados para aplicaciones de antenas en el rango de Terahertz [120].

El objetivo es caracterizar las superficies y determinar cuál de ellas es la más adecuada para detectar este aminoácido y con cuál se podría alcanzar su detección en niveles fisiológicos durante el tratamiento anticáncer, los cuales se esperarían en concentraciones desde micromolar y hasta nanomolar [88], [92].

6.1 MUESTRAS DE L-ASPARAGINA

Se preparó una solución acuosa de L-Asp a 0.1 M, a partir de L-Asparagina en polvo de la marca Sigma-Aldrich con pureza $\geq 98\%$, diluyendo 0.066 gr de L-Asp en 5.00 mL de agua desionizada. La solución se mantuvo en refrigeración ($\sim 4^{\circ}\text{C}$) hasta ser usada.

6.2 SUSTRATOS SERS

Se utilizaron los siguientes cuatro sustratos con estructuras metálicas para obtener respuesta de mejoramiento Raman para L-Asp.

6.2.1 Sustrato SERS de partículas

Esta superficie SERS se fabricó depositando partículas de oro (AuPs) en solución coloidal sobre una hoja de aluminio, una vez que el líquido se evaporó, se utilizó como sustrato SERS.

Las AuPs cuya estructura es hueca y cuasi esférica, fueron sintetizadas por Rodríguez-Montelongo et. al. [121] mediante una reacción de sustitución de Au a partir de nanopartículas de plata. El espectro mostrado en la Figura 6.1 corresponde a su caracterización a través de espectroscopía de transmisión óptica, el máximo de absorción corresponde a 715 nm, lo que indica que tienen en esta longitud de onda su resonancia máxima. La morfología se estudió mediante microscopía electrónica, obteniéndose tamaños de entre 40 y 450 nm. Esta variación en el tamaño y su estructura poco uniforme es correspondiente a que la curva de absorción sea ancha.

Debido a la variación en tamaños, la eficiencia como superficie SERS puede verse afectada y presentar solo efecto de amplificación gracias a la fracción de partículas que son de magnitud nanométrica. Se eligieron estas partículas porque al probar otros tipos no se obtuvo ninguna respuesta de amplificación y la señal Raman de la L-Asp se perdió.

Se obtuvo el espectro Raman de la superficie SERS antes de probarla con el analito, Figura 6.2; se identificaron 2 bandas prominentes 116.5 y 258.4 cm^{-1} .

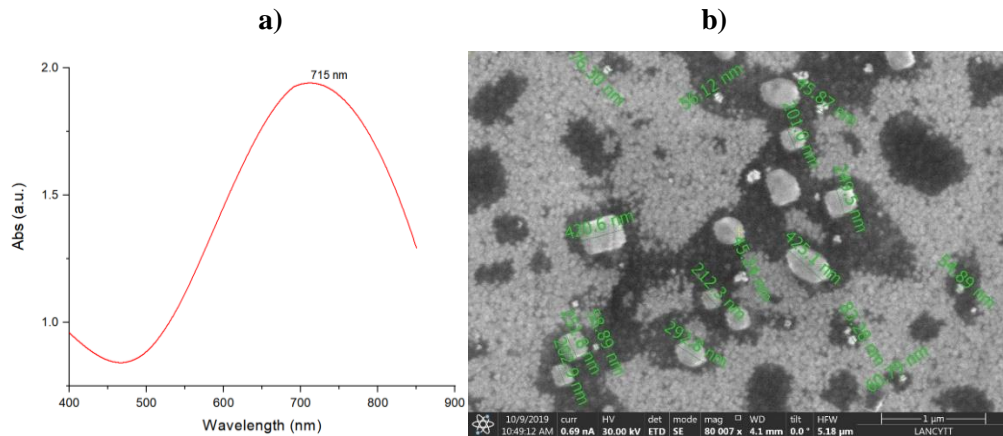


Figura 6.1. Caracterización de las partículas de oro huecas. a) Espectro de absorción. b) Partículas vistas en el microscopio electrónico.

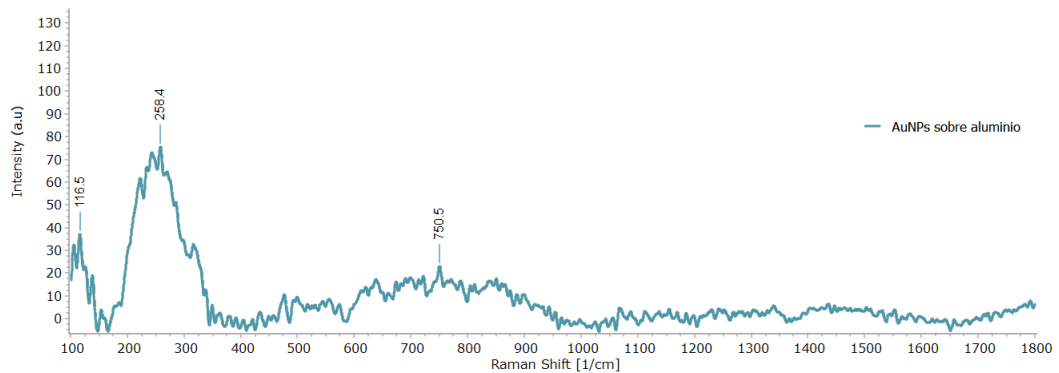


Figura 6.2. Espectro Raman de AuPs (secas) sobre papel aluminio. Las bandas que se observan podrían ser asignadas a los restos de los reactivos usados en la síntesis de las AuPs.

6.2.2 Sustrato SERS de nanoestructuras triangulares sobre óxido de silicio.

Esta superficie SERS fue un arreglo periódico de nanotriángulos de oro (AuNT) grabados sobre SiO₂/Si mediante litografía de haz de electrones; el diseño consistió en triángulos equiláteros con 1.50 μm de altura, espesor de 50 nm y espaciamiento de 1.50 μm [120].

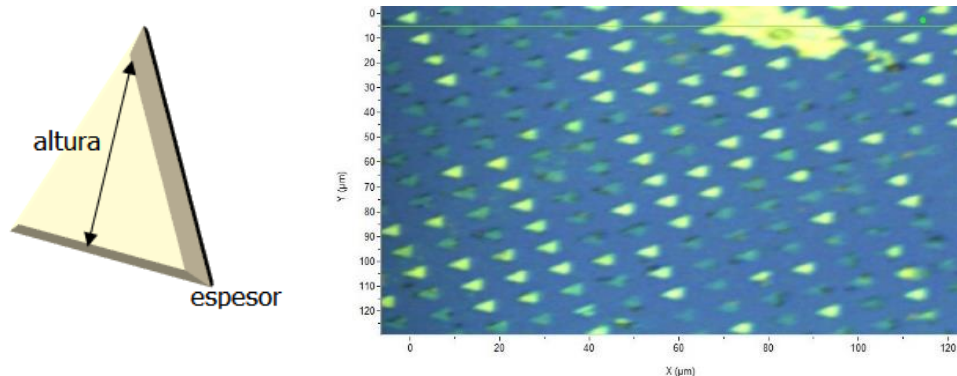


Figura 6.3 a) Diseño de nanoantenas triangulares, b) Patrón observado en el microscopio Raman.

La respuesta electromagnética simulada mediante COMSOL® no presentó resonancia en el espectro visible, sin embargo, sí presentó respuesta de resonancia para frecuencias de Terahertz (Figura 6.4).

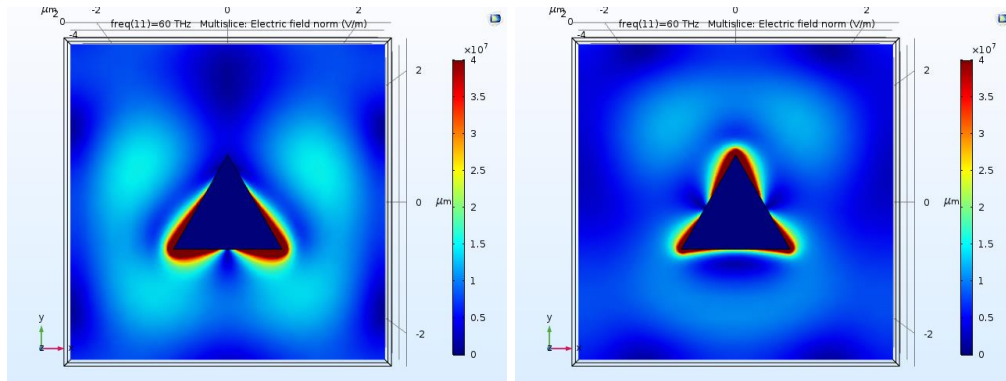


Figura 6.4. Respuesta electromagnética simulada mediante COMSOL ®.

La respuesta Raman de esta estructura, mostrada en Figura 6.5, solo presenta las bandas correspondientes a las vibraciones del SiO₂.

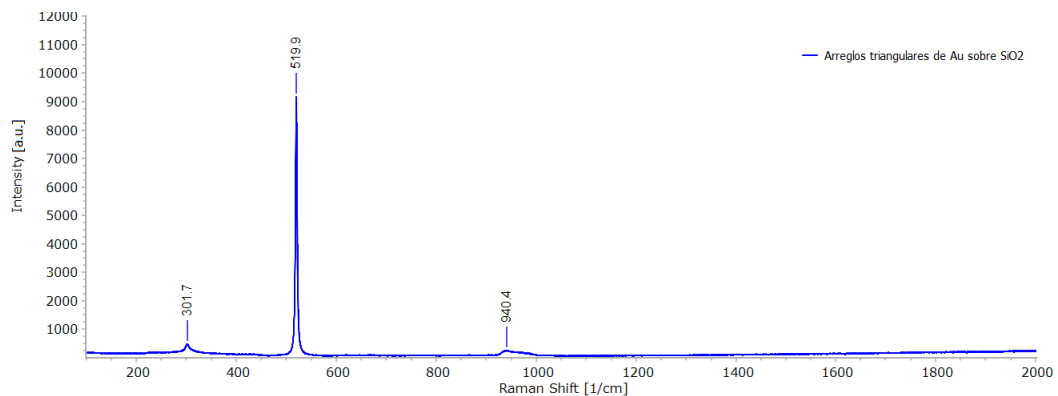


Figura 6.5 Espectro Raman de la estructura de AuNT sobre SiO₂.

6.2.3 Sustrato SERS de nanoestructuras tipo bowtie de oro sobre sustrato de grafeno en óxido de silicio.

Al igual que la superficie SERS anterior este sustrato SERS también fue fabricado mediante litografía de haz de electrones, obteniéndose a partir de un depósito rectangular uniforme de oro con un espesor de 50 nm sobre el cual se excavó para obtener un arreglo periódico de nanoestructuras tipo bowtie (dos triángulos unidos por un vértice formando un tipo moño). La Figura 6.6 muestra el diseño individual de la estructura bowtie, el arreglo del sustrato sobre el que se grabó y el patrón obtenido.

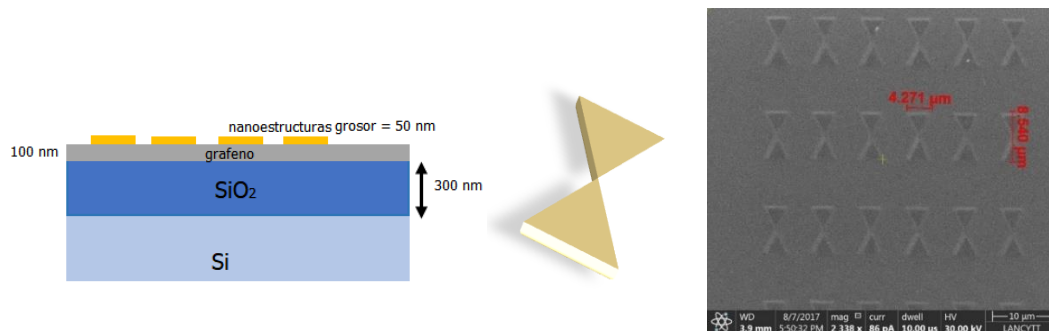


Figura 6.6. Sustrato, diseño y morfología microscópica del arreglo de nanoestructuras bowtie.

Se midió la respuesta electromagnética mediante FTIR, la cual no presentó resonancia en el espectro visible. El espectro Raman de esta estructura, Figura 6.7, solo presenta las bandas correspondientes a las vibraciones del SiO₂.

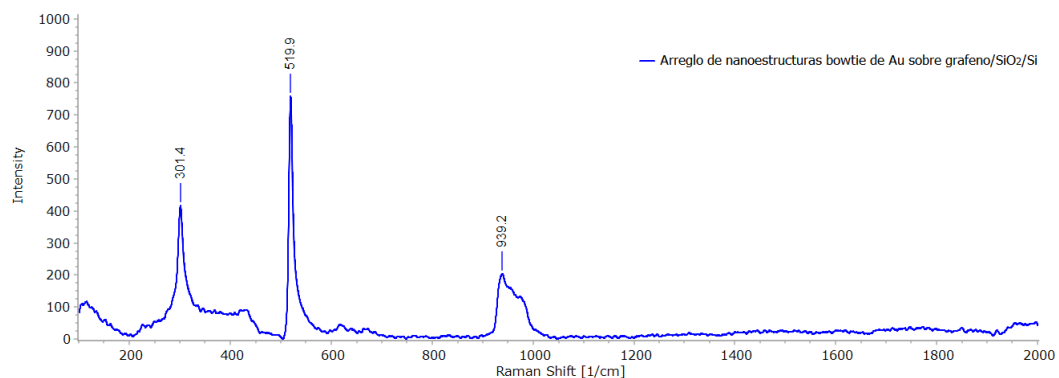


Figura 6.7. Espectro Raman del patrón de estructuras bowtie sobre grafeno/SiO₂/Si.

6.2.4 Sustrato SERS de nanoestructuras tipo bowtie de plata sobre óxido de silicio.

Un tercer sustrato SERS consistió en estructuras similares a las anteriores: bowtie de plata, grabadas mediante litografía sobre SiO₂/Si. El espesor de los arreglos fue de 50 nm, su altura total de 8.0 μm y 4.0 μm de ancho. La Figura 6.8 muestra su diseño y el patrón obtenido. Su caracterización mediante FTIR no demostró resonancia en el espectro visible y su espectro Raman (Figura 6.9) tiene las bandas asociadas al silicio.

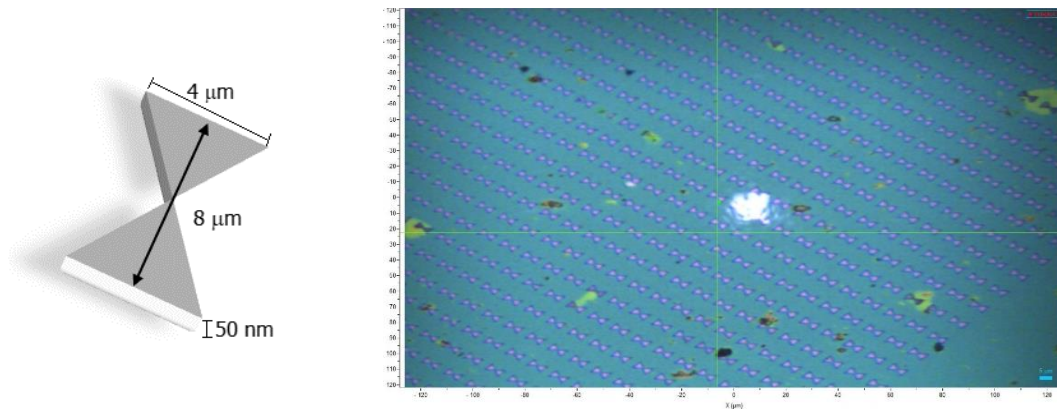


Figura 6.8. Diseño de la nanoestructura bowtie y patrón observado en el microscopio.

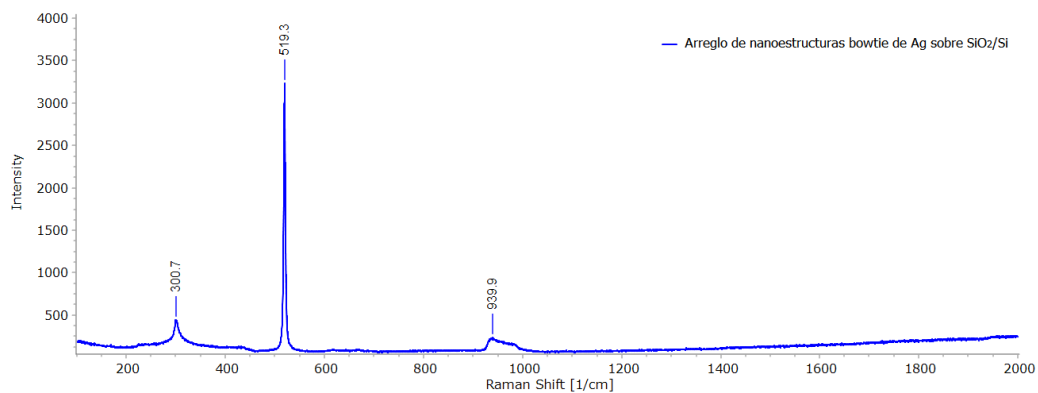


Figura 6.9. Espectro Raman del arreglo de nanoestructuras bowtie de Ag sobre SiO₂/Si.

6.3 ESPECTROSCOPIA RAMAN Y SERS

Para obtener los espectros Raman de referencia se depositó L-Asp a 0.1 M sobre aluminio y sobre SiO₂. También se obtuvo el espectro Raman de L-Asp en polvo para identificar sus bandas. Los espectros SERS se obtuvieron depositando una gota (~50 μL) de la solución de L-Asp sobre cada tipo de superficie SERS, una vez que la gota se secó se llevaron a cabo las mediciones. El mismo proceso de adquisición de espectros descrito en el capítulo anterior se siguió para este caso, mismo espectrómetro, objetivo, rejilla, tiempo de adquisición y longitud de onda. El rango de adquisición fue de 100 – 2000 cm⁻¹.

6.3.1 Análisis de los Espectros Raman de referencia

Los espectros de referencia corresponden a las Figuras 6.10, 6.11 y 6.12. El espectro Raman de L-Asp en polvo (Figura 6.10) coincide en todas sus bandas con los espectros previamente reportados por Moreno et. al. [98] y Golichenko et. al. [100]; las asignaciones de bandas con sus vibraciones correspondientes pueden consultarse en la Tabla 6.1. Las bandas más intensas son de vibraciones de la estructura completa y se encuentran en frecuencias bajas ($< 200 \text{ cm}^{-1}$), generalmente este rango de frecuencia no se utiliza para caracterizar o identificar biomoléculas porque no es específico [4]. En el rango de su huella dactilar ($400 - 1800 \text{ cm}^{-1}$) se tienen las bandas que corresponden a las vibraciones de sus enlaces, las más prominentes son: $517, 834$ y 1422 cm^{-1} ; seguidas por $907, 1097, 1139, 1231$ y 1356 cm^{-1}

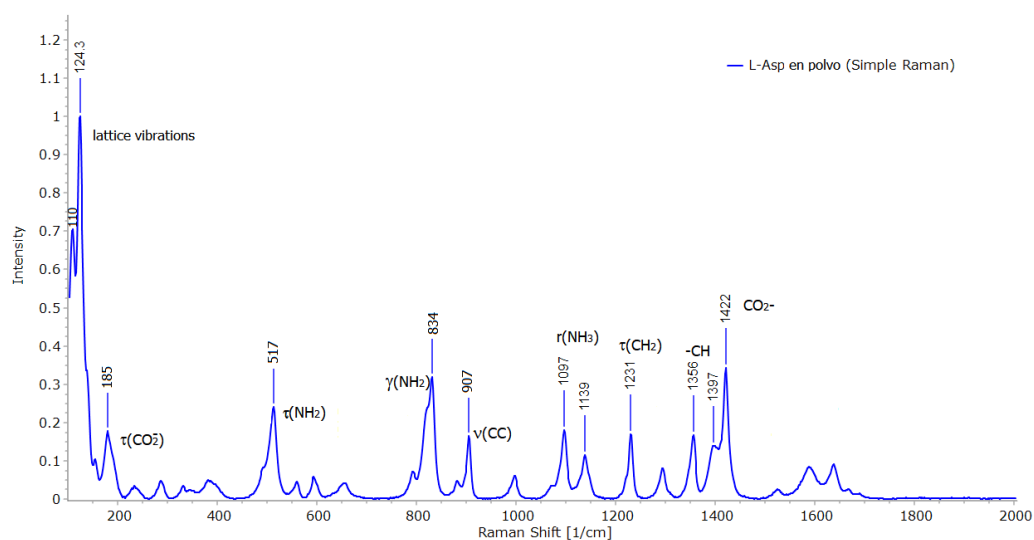


Figura 6.10. Espectro Raman de L-Asp en polvo (directa del envase del fabricante). Este espectro fue normalizado, pero no fue necesario hacer corrección de fluorescencia: la concentración de la molécula, el láser utilizado y el tiempo de adquisición evitaron que la fluorescencia afectara la medición.

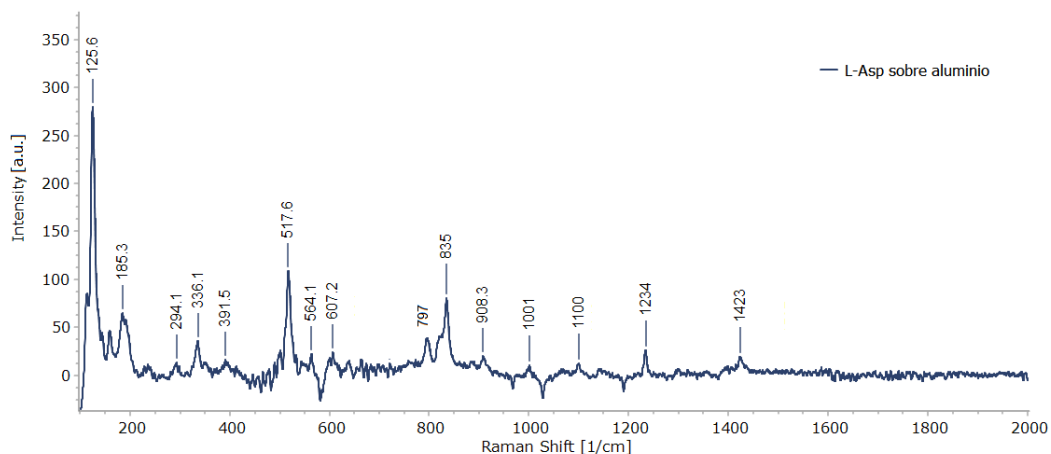


Figura 6.11. Espectro Raman de L-Asp a 0.1 M sobre aluminio. El procesamiento de este espectro solo consistió en eliminar la fluorescencia, no se llevó a cabo normalización pues se requiere comparar la intensidad con los espectros SERS.

Del espectro Raman de L-Asp sobre aluminio (Figura 6.11) no se distinguen todas las bandas esperadas debido a que la concentración es menor. Las más prominentes son las correspondientes a las deformaciones estructurales ($<200\text{ cm}^{-1}$), seguidas de la banda correspondiente a torsiones de enlaces NH_2 en 517 cm^{-1} , luego vibraciones de grupos NH_2 en 835 cm^{-1} , balanceos y torsiones de enlaces CH_2 en 797 y 1234 cm^{-1} respectivamente; y pequeños picos en 336 , 908 , 1100 y 1423 cm^{-1} .

Del espectro Raman sobre un sustrato de SiO_2 (Figura 6.12) no se lograron observar vibraciones de L-Asp, están visibles las correspondientes a SiO_2 .

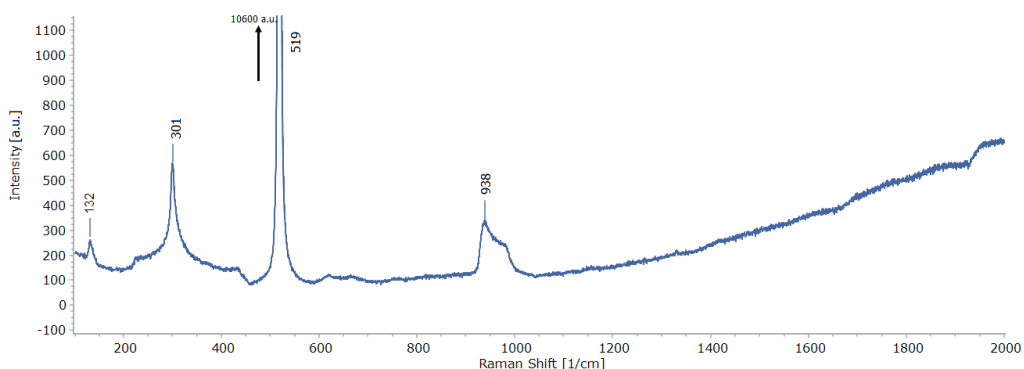


Figura 6.12. Espectro Raman obtenido de L-Asp a 0.1 M sobre un sustrato de SiO_2 . La pequeña banda en 132 cm^{-1} es de L-Asp.

6.3.2 SERS con superficie de partículas de oro.

El espectro SERS de L-Asp sobre las AuPs se muestra en la Figura 6.13 (espectro en rojo), el cual tiene incremento en la intensidad de las bandas en frecuencias menores a 200 cm^{-1} comparado con el espectro Raman de L-Asp sobre aluminio (espectro en azul), también se presenta en esa región aumento en el corrimiento de aproximadamente 5 cm^{-1} . Este corrimiento podría deberse a un traslape con la banda ancha de 258 cm^{-1} del espectro de las AuPs (Figura 6.2).

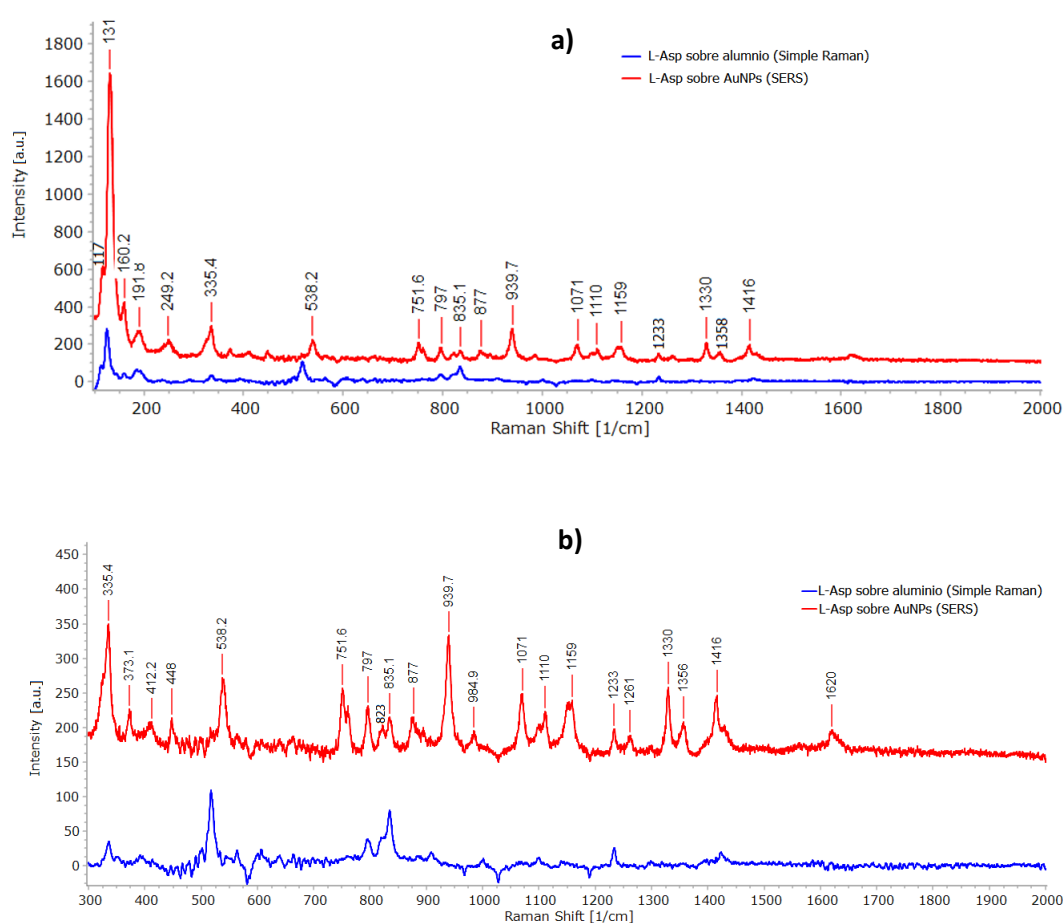


Figura 6.13 a). Espectro Raman de L-Asp sobre aluminio (azul) y espectro SERS de L-Asp obtenido con la superficie de AuPs (rojo). El espectro SERS se desplazó hacia arriba para visualizar las diferencias y similitudes con el espectro Raman. Para el análisis del factor de mejoramiento ambos espectros se sometieron a eliminación de fluorescencia mediante substracción de línea base. **b)** Aumento de una región de los espectros Raman y SERS donde se observan con mejor resolución las bandas.

Algunas de las bandas de L-Asp que se observan en el espectro de referencia no tiene un aumento considerable en intensidad en el espectro SERS (797, 824, 835 y 1234 cm^{-1}); mientras que las bandas 335, 1071 y 1356 cm^{-1} se pueden distinguir en el espectro SERS, pero no en el de referencia. En el espectro SERS aparecen bien definidas otras bandas que no corresponden a L-Asp, de las cuales una corresponde a la superficie de AuPs sobre aluminio (249 cm^{-1}); otras sin embargo son completamente inesperadas: 538, 752, 877, 940, 984, 1159, 1330 y 1416 cm^{-1} y pueden deberse vibraciones de enlaces formados entre la superficie y las moléculas de L-Asp.

Se calculó el factor de mejoramiento del esparcimiento (EF) a partir de la Ecuación 2.7 y mediante la función *sers_enhancement_factor* en Matlab® [122]; considerando la Ecuación 2.8 para determinar el número de moléculas que se excitan sin superficie de mejoramiento y la Ecuación 2.9 para determinar el número de moléculas excitadas en SERS. Los parámetros experimentales para estas ecuaciones fueron: densidad de L-Asp $\rho = 1.543 \text{ gr/cm}^3$; peso molecular $W = 132.12 \text{ gr/mol}$ y área de absorción de cada molécula de L-Asp $A = 0.75 \text{ nm}^2$.

BANDA (cm^{-1})	EF
117	1.98E+05
131	3.16E+05
160	2.03E+05
191.8	9.75E+04
249.2	2.44E+05
335.4	1.97E+05
373.1	1.97E+05
448	2.63E+05
538.2	2.63E+05
751.6	2.98E+05
797	1.15E+05
835.1	4.89E+04
877	1.47E+05
939.7	4.49E+05
984.9	4.49E+05
1071	1.91E+05
1110	1.82E+05
1159	1.82E+05
1233	6.99E+04
1330	2.68E+05
1416	2.15E+05
1620	2.15E+05

Tabla 6.1 Factores de mejoramiento SERS para la superficie con partículas de oro. Los valores en negritas son las bandas de vibraciones

6.3.3 SERS con superficie de triángulos sobre SiO₂

Las Figuras 6.14 muestran el espectro SERS para L-Asp obtenido con el arreglo AuNTs (azul) y el espectro de L-Asp sobre aluminio (rojo) como referencia. Nuevamente las bandas más prominentes son las que corresponden a las vibraciones de la estructura (113, 129, 159 y 186 cm⁻¹), seguidas de la banda en 519.7 cm⁻¹ que corresponde a la estructura del SiO₂ y que opaca a la banda 517 cm⁻¹ de la vibración $\tau(\text{NH}_2)$ de L-Asp.

En la Figura 6.14 b, se puede constatar que la mayoría de las bandas de L-Asp tienen mejoramiento en intensidad, incluso aquellas que no se distinguen en el espectro de referencia. En particular para este espectro SERS se obtuvo muy poca fluorescencia y ruido, presentando bandas muy definidas

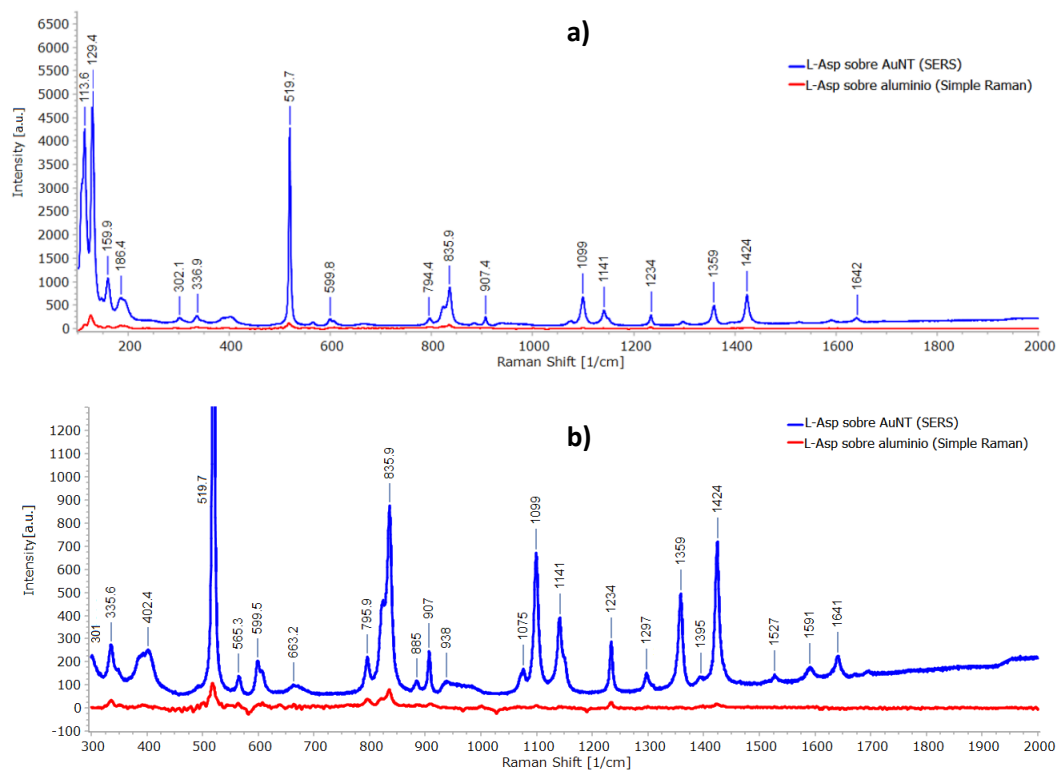


Figura 6.14 a) Espectro Raman de L-Asp sobre aluminio (rojo) y espectro SERS de L-Asp obtenido con la superficie de AuNTs (azul). Para el análisis del factor de mejoramiento ambos espectros se sometieron a eliminación de fluorescencia mediante sustracción de línea base. **b)** Aumento de una región de los espectros Raman y SERS donde se observan con mejor resolución las bandas, el espectro SERS se desplazó hacia arriba para mejorar la visualización.

Implementando el mismo procedimiento anteriormente descrito para calcular el EF, se obtuvieron los valores correspondientes para cada una de las bandas de L-Asp.

BANDA (cm ⁻¹)	EF
113.6	1.78E+06
129.4	1.06E+06
159.9	6.48E+05
186.4	3.02E+05
402	5.10E+05
599.8	1.87E+05
795	2.20E+05
824	5.48E+05
835.9	5.48E+05
907.4	3.50E+05
1075	2.26E+05
1099	1.03E+06
1141	7.81E+05
1153	7.81E+05
1234	2.92E+05
1297	3.67E+05
1359	2.85E+06
1395	1.53E+06
1424	1.53E+06
1642	6.95E+05

Tabla 6.2 Factores de mejoramiento SERS para la superficie con triángulos de oro sobre SiO₂. Todas las bandas corresponden a vibraciones de L-Asp. Se omitió la banda en 517 cm⁻¹ pues se traslapa con la banda 519 cm⁻¹ del silicio.

6.3.4 SERS con superficie de estructuras de oro tipo bowtie sobre grafeno y SiO₂

El espectro SERS de L-Asp obtenido con el sustrato de nanoestructuras bowtie es el mostrado en azul en la Figura 6.15, el espectro en rojo corresponde al espectro de L-Asp sobre aluminio. Al igual que con los otros sustratos SERS, se obtiene mejoramiento para las bandas correspondientes a las vibraciones de la estructura, que en este caso corresponden a las frecuencias 116, 132.6 y 191.8 cm⁻¹. También se obtiene mejoramiento para las bandas 1069, 1158 y 1353 cm⁻¹, que pertenecen a vibraciones de L-Asp, el resto de las bandas no corresponden a este aminoácido. Las correspondientes a SiO₂ son en 302, 519 y 939 cm⁻¹; otras de las bandas que se distinguen coinciden con las obtenidas con la superficie de partículas, con excepción de 580 cm⁻¹, que podrían provenir de enlaces formados entre el oro y los residuos del aminoácido.

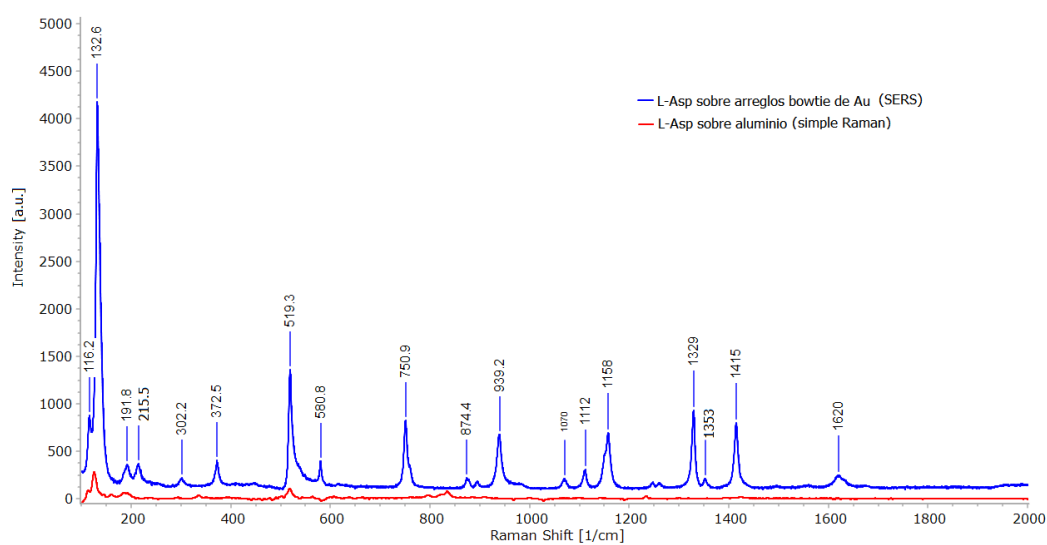


Figura 6.15. Espectro Raman de L-Asp sobre aluminio (rojo) y espectro SERS de L-Asp en la superficie de estructuras bowtie de oro (azul).

BANDAS (cm⁻¹)	EF
116.2	3.47E+05
132.6	1.34E+06
191.8	1.54E+05
215.5	7.45E+05
372.5	9.74E+05
580.8	1.43E+06
750.9	3.23E+06
874	4.40E+05
895	3.06E+05
939.2	1.79E+06
1070	2.12E+05
1112	5.10E+05
1158	1.61E+06
1329	7.29E+06
1353	7.50E+05
1415	2.42E+06
1620	6.49E+05

Tabla 6.3 Factores de mejoramiento SERS para la superficie de estructuras bowtie de oro sobre SiO₂. En negritas se identifican las únicas bandas que corresponden a L-Asp, se incluyen también las que podrían deberse a nuevos enlaces, pero no las bandas de silicio.

6.3.5 SERS con superficie de bowtie de plata sobre SiO₂

La Figura 6.16 presenta el espectro SERS de L-Asp sobre las estructuras tipo bowtie de plata grabadas sobre SiO₂ (azul) y el espectro Raman de L-Asp sobre aluminio tomado como referencia (rojo). Las bandas del aminoácido que distinguibles son: 111.5, 128.6, 158, 335, 795, 823, 836, 883, 1069, 1098, 1233 y 1360 cm⁻¹, visualmente solo algunas de estas bandas presentan mejoría en el espectro SERS. Como se esperaba las bandas de las vibraciones del silicio son distinguibles. La tabla 6.4 presenta los resultados obtenidos en los cálculos de EF para esta superficie.

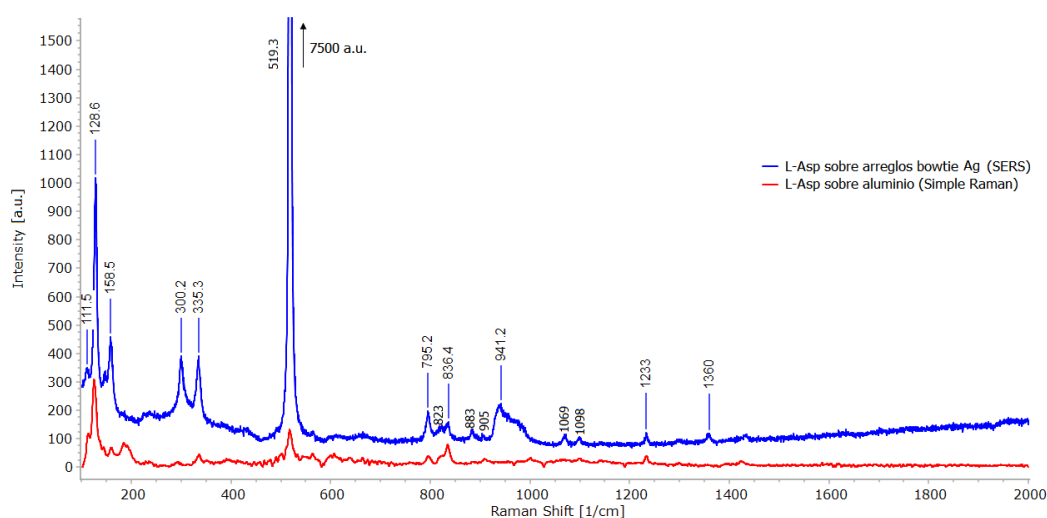


Figura 6.16. Espectro Raman de L-Asp sobre aluminio (rojo) y espectro SERS de L-Asp en la superficie de estructuras bowtie de plata (azul).

BANDAS (cm ⁻¹)	EF
111.5	1.69E+05
128.6	1.69E+05
158.5	2.34E+05
300.2	1.51E+06
335.3	3.33E+05
795.2	1.81E+05
836.4	1.81E+05
883	6.69E+05
1069	6.69E+05
1098	6.69E+05
1233	6.69E+05
1360	6.69E+05

Tabla 6.4 Factores de mejoramiento SERS para la superficie de estructuras bowtie de plata sobre SiO₂. La frecuencia señalada en 300 cm⁻¹ corresponde al SiO₂.

6.3.6 Discusión sobre el desempeño de las superficies SERS

De las cuatro superficies SERS utilizadas para mejorar el espectro Raman de L-Asp, la de nanotriángulos es la que mejor desempeño presentó cualitativamente, obteniéndose un espectro con casi todas las bandas reportadas para el aminoácido y solo las bandas del silicio adicionales. Los espectros SERS obtenidos con la superficie de AuPs y los arreglos bowtie de oro, aunque presentan mejoramiento para algunas de las bandas de L-Asp, presentan otras bandas que no son características del aminoácido, ambas superficies presentan estas mismas bandas adicionales por lo que podrían deberse a vibraciones de enlaces formados entre el oro y el analito; en la superficie con estructuras bowtie/grafeno/SiO₂ se presentan además las bandas del silicio. Las estructuras bowtie de plata, presentan también algunas bandas mejoradas con respecto del espectro de referencia, pero su amplificación no es evidente.

De la superficie de AuPs fue de la única que se pudo comprobar su resonancia (715 nm), por lo que se esperaba que tuviera un buen desempeño de amplificación al usar 785 nm de excitación; de las superficies de nanoestructuras no se obtuvo información experimental o teórica que justificara la existencia de plasmones de resonancia cercanos a la frecuencia de excitación, aun así, fueron probadas con 785 nm. Que estas superficies permitan obtener señal SERS puede justificarse con que podría tratarse de superficies porosas, y la existencia de plasmones de resonancia no muy fuertes pero lo suficientemente capaces de generar mejoramiento en el esparcimiento Raman.

Cuantitativamente, a partir del cálculo del EF, el mejor desempeño fue con una amplificación del orden de 10^6 y la banda amplificada con menor EF fue del orden de 10^4 ; sin embargo, una revisión más exhaustiva de la ecuación aplicada para calcular el factor de mejoramiento demuestra que podría haber mayores diferencias debido al tipo de superficies. En el cálculo llevado a cabo se asumió como referencia, para todas las superficies, el espectro Raman de L-Asp sobre aluminio, donde se estimó que el número de moléculas excitadas fue de $N_{\text{Raman}} = 3.1 \times 10^{11}$. Luego se estimó el número de moléculas que contribuyen al efecto SERS, en el cual se supuso que las moléculas de Asparagina fueron absorbidas como una monocapa sobre la superficie metálica y que toda el área iluminada por el spot del láser las

contenía uniformemente (Ecuación 2.9), obteniendo el número de moléculas $N_{SERS} = 6.0 \times 10^6$. Ambas cantidades se asumieron constantes para todos los arreglos. Con lo cual el cálculo del EF a partir de la Ecuación 2.7 refleja la comparación entre la cantidad de partículas entre ambas condiciones experimentales $N_{Raman}/N_{SERS} \sim 10^5$ y la comparación entre intensidades; la aproximación, aunque es válida, no permite hacer una comparación más detallada entre las superficies, en las cuales debido a sus estructuras particulares cambiaría la interacción de las moléculas de L-Asp con la superficie.

Otra estimación para el factor de mejoramiento podría ser aplicada para el caso particular de las AuPs:

$$\frac{N_{Raman}}{N_{SERS}} = \frac{\pi r^2 h c}{\mu_M \mu_S A_S A_e} \quad (6.1),$$

donde c es la concentración del analito para la medición Raman $\pi r^2 h$ el volumen de excitación; μ_M es la densidad superficial de la molécula en la medición SERS, μ_S es la densidad superficial de las partículas y A_S es el área efectiva de cada nanopartícula y A_e es el área de excitación (spot del láser) [37]. Sin embargo, no se cuenta con los datos experimentales para hacer una estimación de A_S , faltando la densidad superficial de las partículas y la distribución de sus tamaños en la superficie SERS formada; lo que se puede concluir es que al hacer este cálculo el factor de mejoramiento puede verse disminuido y esta superficie tendría menor amplificación que las superficies de nanoestructuras.

Los resultados obtenidos fueron para mejorar la señal Raman de la molécula a concentración 0.1 M; al disminuir la concentración hasta 1 μ M (de los menores niveles fisiológicos reportados, Tabla 3.2), el número de moléculas disminuirá en 5 órdenes de magnitud; por lo que con un factor de mejoramiento de 10^6 se requerirá que la intensidad SERS sea 20 veces mayor a la intensidad Raman. Si esta última es muy cercana a cero, entonces el espectrómetro y la superficie SERS deben permitir obtener señales con muy poco ruido para que sea detectable.

6.3.7 Simulaciones y Asignación de vibraciones

De acuerdo con los trabajos previamente reportados [98]–[100] las bandas del espectro Raman para L-Asparagina corresponden a las que se presentan a continuación en la Tabla 6.5. Las dos primeras columnas corresponden a las bandas identificadas en los espectros de L-Asp en polvo y L-Asp a 0.1 M sobre aluminio respectivamente (Ref. 1 y Ref. 2). La tercera columna son las bandas del espectro SERS de L-Asp obtenido con la superficie de partículas (AuPs) mientras que las tres columnas consecutivas corresponden a cada uno de los arreglos metálicos de nanoestructuras: Arreglo 1 para el arreglo de nanotriángulos, Arreglo 2 para el de estructuras bowtie de oro y Arreglo 3 para el de plata.

Ref. 1	Ref. 2	AuPs	Arreglo 1	Arreglo 2	Arreglo 3	Asignación vibracional
110	113	117	113.6	116.2	111.5	δ (estructura)
124.3	125.6	131	129.4	132.6	128.6	δ (estructura)
155	159.7	160	159.9		158.5	δ (estructura), δ (CC)+ τ (CC)
180.5	185.3		186.4			τ (CO ₂)
		191.8		191.8		τ (CO ₂)
				215.5		de las AuPs
234.9	236.2					δ (estructura), τ (CC)
		249.2				De las partículas
288	294					δ (estructura), τ (CC)
			302.1	302.2	300.2	Si
336	336	335.4	335.6		335.3	δ (estructura), τ (CC)
		373.1		372.5		No reportada
391.5	391.5					τ (NH ₃ ⁺), δ (CONH ₂)
			402			τ (NH ₃ ⁺)
		448				No reportada
493						δ (estructura), τ NH ₃ ⁺
517	517.6					τ (NH ₂)
			519.7	519.3	519.5	Si
		538.2				No reportada
561.3	564.1		565.3			δ_s (CO ₂ ⁻)
				580.8		No reportada
594.9			599.8			δ_s (CONH ₂) + δ (NH ₃ ⁺)
	607.2		606.5			δ_s (CONH ₂)
656.7			663.2			δ_a (CONH ₂), δ_a (CO ₂ ⁻), γ (CCN)
		751.6		750.9		No reportado
791.5	797	797	795		795.2	τ (CH ₂), τ (NH ₂)
823		824	824		823	γ (CO ₂ ⁻)
831.2	835	835.1	835.9		836.4	γ (NH ₂)
		877		874		No reportado
883			885		883	ν (CC), τ (CH ₂)
				895		No reportado
906.4	908		907.4		905	ν_s (CC)
		939.7		939.2		No reportado

			938		941.2	Si
		984.9				No reportado
997	1001					$\nu(\text{CC})$
1071		1071	1075	1070	1069	$\nu(\text{CN})$
1097	1100		1099		1098	$r(\text{NH}_3)$, $r(\text{NH}_2)$
		1110		1112		$r(\text{NH}_3^+)$
1139			1141			$r(\text{NH}_3^+)$
			1153	1158		$r(\text{NH}_2)$, $r(\text{NH}_3)$,
		1159				No reportado
1231	1234	1233	1234		1233	$\tau(\text{CH}_2)$, $r(\text{NH}_2)$,
				1248		No reportado
				1261		No reportado
1295			1297			$\omega(\text{NH}_2)$
		1330		1329		No reportado
1356		1358	1359	1353	1360	$\delta(\text{CH})$
1397			1395			$\delta(\text{CH}_2)$
		1416		1415		No reportado
1422	1423		1424			$\nu_a(\text{CO}_2)$, $\delta(\text{CO}_2)$, $\sigma(\text{CH}_2)$
1526			1527			$\delta(\text{NH}_3)$
1589			1591			$\nu_a(\text{CO}_2)$, $\delta_a(\text{NH}_3)$
		1620		1620		$\delta(\text{NH}_2)$
						$\delta(\text{NH}_2)$, $\nu_a(\text{CO}_2)$
1638			1642			$\nu(\text{C=O})$
1667						$\delta(\text{NH}_3)$
1689						$\delta(\text{NH}_3)$

Tabla 6.5. Bandas identificadas en los espectros experimentales de L-Asparagina. Los símbolos de las vibraciones son: δ = deformación; γ = vibración fuera del plano; τ = torsión; ω = meneado; r = balanceo; σ = tijereo; ν = estiramiento; a = asimétrico; s = simétrico.

Para corroborar las vibraciones de L-Asparagina, se llevaron a cabo algunas simulaciones mediante el software Gaussian 03W; los cálculos se realizaron con dos aproximaciones, DFT y Hartree Fock, respectivamente. Los espectros obtenidos se muestran en la Figura 6.17, en ninguna de estas simulaciones se obtuvieron todas las bandas reportadas y medidas experimentalmente, sin embargo, las vibraciones de algunas son cercanas o bien sí corresponden.

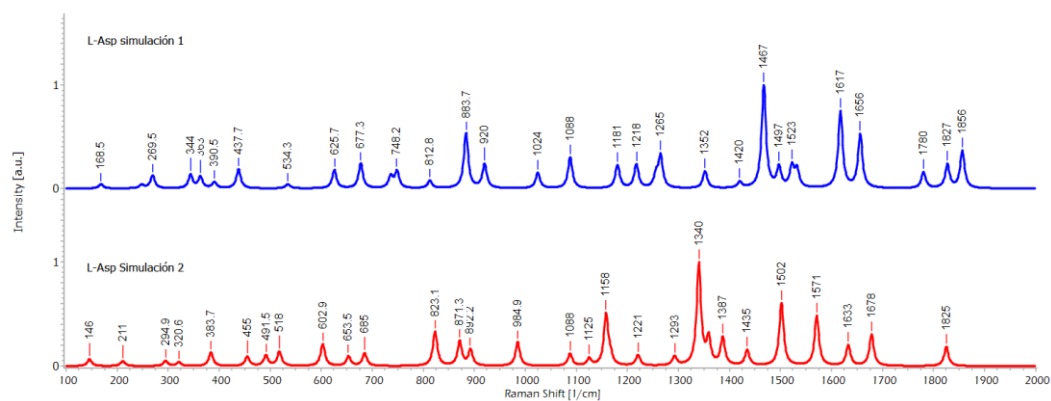


Figura 6.17. Espectros Raman de las simulaciones moleculares de L-Asparagina.

Capítulo 7.

CONCLUSIONES Y TRABAJO FUTURO

Se estudiaron dos biomarcadores mediante espectroscopía Raman, la proteína p53 y el aminoácido L-Asparagina, lográndose cumplir el objetivo principal de este trabajo: comprobar la utilidad de este tipo de espectroscopía para identificar y caracterizar estos biomarcadores.

Para las proteínas p53 tipo natural y mutantes R273H, E343A y L344A se obtuvieron sus espectros Raman característicos, identificándose las bandas correspondientes a vibraciones de sus aminoácidos. Aunque en un análisis visual de los picos y sus correspondientes vibraciones se identificaron algunas diferencias entre los cuatro tipos de espectros, la implementación del análisis de componentes principales confirmó mediante los *PC loadings* las principales diferencias y reveló otras que no eran evidentes.

La proyección espacial de los *PC scores* permitió identificar que existe clara diferencia entre los espectros de p53 tipo natural y los espectros de p53 mutante, específicamente el *PC score 2* separó los espectros en estas dos clases. Con el conjunto de *PC scores* más relevantes se entrenó y probó un clasificador de aprendizaje automático: SVM multiclase, este método permitió obtener una clasificación con solo un espectro clasificado erróneamente, logrando una especificidad de 97%, exactitud y sensibilidad de 92% y precisión de 94%. Una limitante en este resultado es que la cantidad de datos clasificados es pequeña (12 espectros), por lo que el entrenamiento y la prueba se tuvo que llevar a cabo con el mismo conjunto de datos; esta limitante podrá ser resuelta fácilmente en un futuro al incrementar la cantidad de espectros obtenidos por cada clase.

La aplicación de la espectroscopía Raman en conjunto con métodos multivariados y de aprendizaje automático pueden por lo tanto ser una alternativa sumamente útil para identificar y diferenciar entre los tipos de p53. En contraste con los métodos estándar, Raman tiene la ventaja de ser una técnica más rápida, tanto para la preparación de las muestras como para la obtención de resultados; además no requiere etiquetas moleculares o pruebas genéticas.

Se obtuvo además el límite de detección que mediante espectroscopía Raman se puede alcanzar para la proteína p53 tipo natural, este resultado tiene dos implicaciones importantes. La primera, demuestra la ventaja de utilizar métodos de análisis multivariable para el manejo de datos espectrales; en este caso particular la curva de calibración obtenida mediante PLSR presentó mejor desempeño que los métodos de análisis univariable. La segunda implicación hace evidente la necesidad de implementar SERS para alcanzar niveles de detección fisiológicos para la proteína p53; en este trabajo el límite de detección calculado fue de 1.022 μM mientras que las concentraciones fisiológicas reportadas son del orden de picomolar, por lo que como trabajo a futuro en este sentido se plantea la implementación de SERS para detectar p53 y posteriormente clasificarla mediante aprendizaje automático; este método puede competir más directamente con los que actualmente se utilizan y podría incluso ser usado en aplicaciones clínicas.

La técnica de SERS fue aplicada para la detección del aminoácido L-Asparagina. Específicamente se buscó determinar cuál es el tipo de superficie óptima para tener la mejor amplificación sin perder las características espectrales que permitan identificar al aminoácido. Se probaron diferentes tipos de superficies metálicas (micropartículas y nanoantenas), obteniéndose mejor resultado para una de las superficies de nanoantenas, tanto en mejoramiento de la señal que se alcanzó ($\text{EF} \sim 10^6$), como en mantener las mismas bandas Raman características del aminoácido en el espectro SERS. Las nanoantenas no fueron diseñadas para su aplicación en la longitud de onda utilizada, sin embargo, presentaron mejoramiento de esparcimiento Raman, por lo que, si este tipo de superficies son específicamente fabricadas para su aplicación en SERS, pueden llegar a ser excelentes biosensores.

Como se pudo comprobar en esta investigación, la espectroscopía Raman es una alternativa prometedora para mejorar los métodos de detección de biomarcadores moleculares, por lo que es importante continuar con el desarrollo de tecnología basada en Raman para lograr que su aplicación permita progresar en el diagnóstico médico.

PRODUCTIVIDAD ACADÉMICA

Publicación en revistas:

K. Hernández-Vidales, E. Guevara, V. Olivares-Illana, and F. J. González, "Characterization of wild-type and mutant p53 protein by Raman spectroscopy and multivariate methods," *J. Raman Spectrosc.*, Jun. 2019. Impact factor: 2.74.

E. Guevara, M. Miranda-Morales, **K. Hernández-Vidales**, M. Atzoric, and F. J. González, "Low-cost embedded system for optical imaging of intrinsic signals," *Rev. Mex. Fis.*, vol. 65, no. 6, 2019. Impact factor: 0.626.

A. Loyola-Leyva, **K. Hernández-Vidales**, J. P. Loyola-Rodríguez, and F. J. González, "Raman spectroscopy applications for the diagnosis and follow-up of type 2 diabetes mellitus. A brief review," *Biomed. Spectrosc. Imaging*, vol. 9, no. 3–4, pp. 119–140, Dec. 2020. Impact factor: 0.143.

Memorias de Congresos:

Karen Hernández Vidales, Edgar Guevara, Vanesa Olivares Illana, Francisco Javier González, "Espectroscopía Raman de la proteína p53 wild type y sus mutantes", en libro Desarrollo Científico en México, pag 918- 922. Centro de Investigaciones en Óptica. ISBN: 978-607-95228-9-6.

Karen Hernández Vidales, E. Guevara, V. O. Illana, and F. J. González, "Raman Spectroscopy of Wild Type and Mutants p53," in Latin America Optics and Photonics Conference, OSA Technical Digest (Optical Society of America, 2018).

Karen Hernández Vidales, Edgar Guevara, Vanesa Olivares Illana, Francisco Javier González, "Multivariate analysis of Raman spectroscopy of wild type and mutants p53 cancer biomarker," Proc. SPIE 11130, Imaging Spectrometry XXIII: Applications, Sensors, and Processing, 1113005 (September 6, 2019);

Karen Hernández Vidales, Javier Méndez, and Francisco Javier González "Detection of L-Asparagine by surface-enhanced Raman scattering", Proc. SPIE 11657, Biomedical Applications of Light Scattering XI, 1165700 (March 5, 2021);

Participación en Congresos Nacionales:

Cartel: “**Espectroscopía Raman de la proteína p53 wild type y sus mutantes**”, presentado durante el VX Encuentro Participación de la Mujer en la Ciencia 2018, Centro de Investigaciones en Óptica, León Gto, México.

Cartel: “**Detección mediante espectroscopía Raman de la proteína p53 y sus mutantes**”, en 1er Foro de Investigadores Cátedra CONACyT de la UASLP. 4 de junio 2019, Unidad de Posgrado de la UASLP.

Ponencia: “**Espectroscopía Raman para la detección, caracterización e identificación de Biomarcadores**”, 1er Encuentro de Alumnos del Posgrado en Ciencias Aplicadas, Universidad Autónoma de San Luis Potosí, noviembre 2019.

Participación en Congresos Internacionales:

Ponencia: “**Detection of L-Asparagine by surface-enhanced Raman scattering**”, durante el congreso virtual SPIE Photonics West, Biomedical Applications of Light Scattering XI, 5 de marzo de 2021.

Ponencia: **Multivariate analysis of Raman spectroscopy of wild type and mutants p53 cancer biomarker**, en el SPIE Optics & Photonics. 11 de agosto de 2019 en el Centro de Convenciones de San Diego California.

Ponencia: **Multivariate analysis of Raman spectroscopy of wild type and mutants p53 cancer biomarker** en el Colloquium Spectroscopicum Internationale XLI & I Latin American Meeting on LIBS. 13 de junio de 2019 en la Universidad Nacional Autónoma de México, Ciudad Universitaria.

REFERENCIAS

Por orden de aparición en el texto.

- [1] A. Meissner, “Biomarkers in Medicine, Drug Discovery and Environmental Health. Edited by Vishal S. Vaidya and Joseph V. Bonventre.,” *ChemMedChem*, vol. 6, no. 11, pp. 2110–2110, Nov. 2011.
- [2] K. Strimbu and J. A. Tavel, “What are biomarkers?,” *Current Opinion in HIV and AIDS*. 2010.
- [3] D. W. Ball, *The Basics of Spectroscopy*. SPIE, 2001.
- [4] H. G. M. Edwards, *Modern Raman spectroscopy—a practical approach*. Ewen Smith and Geoffrey Dent. John Wiley and Sons Ltd, Chichester, 2005. Pp. 210. ISBN 0 471 49668 5 (cloth, hb); 0 471 49794 0 (pbk), vol. 36, no. 8. 2005.
- [5] E. Le Ru and P. Etchegoin, *Principles of Surface Enhanced Raman Spectroscopy*. Elsevier B.V., 2009.
- [6] J. Langer *et al.*, “Present and future of surface-enhanced Raman scattering,” *ACS Nano*, vol. 14, no. 1. American Chemical Society, pp. 28–117, 28-Jan-2020.
- [7] N. Feliu, M. Hassan, E. Garcia Rico, D. Cui, W. Parak, and R. Alvarez-Puebla, “SERS Quantification and Characterization of Proteins and Other Biomolecules,” *Langmuir*, vol. 33, no. 38, pp. 9711–9730, 2017.
- [8] M. G. Ramírez-Elías and F. J. González, “Raman Spectroscopy for In Vivo Medical Diagnosis,” *Raman Spectrosc.*, no. April, 2018.
- [9] H. J. Butler *et al.*, “Using Raman spectroscopy to characterize biological materials,” *Nat. Protoc.*, vol. 11, no. 4, pp. 664–687, Apr. 2016.
- [10] M. G. Ramírez-Elías, “Uso De La Espectroscopía Raman Como Método No-Invasivo En El Diagnóstico Médico : Tejido De Piel E Hígado,” Universidad Autónoma de San Luis Potosí, 2011.

- [11] C. Krafft and J. Popp, "Micro-Raman spectroscopy in medicine," *Physical Sciences Reviews*, vol. 4, no. 10. De Gruyter, 01-Oct-2019.
- [12] F. J. González *et al.*, "Use of Raman spectroscopy for the early detection of filaggrin-related atopic dermatitis," *Ski. Res. Technol.*, vol. 17, no. 1, pp. 45–50, 2011.
- [13] X. Feng *et al.*, "Raman active components of skin cancer," *Biomed. Opt. Express*, vol. 8, no. 6, p. 2835, 2017.
- [14] B. Bodanese, F. L. Silveira, R. A. Zângaro, M. T. T. Pacheco, C. A. Pasqualucci, and L. Silveira, "Discrimination of basal cell carcinoma and melanoma from normal skin biopsies in vitro through raman spectroscopy and principal component analysis," *Photomed. Laser Surg.*, vol. 30, no. 7, pp. 381–387, 2012.
- [15] J. Desroches *et al.*, "A new method using Raman spectroscopy for in vivo targeted brain cancer tissue biopsy," *Sci. Rep.*, vol. 8, no. 1, pp. 1–10, 2018.
- [16] N. C. Maiti, M. M. Apetri, M. G. Zagorski, P. R. Carey, and V. E. Anderson, "Raman Spectroscopic Characterization of Secondary Structure in Natively Unfolded Proteins: α -Synuclein," *J. Am. Chem. Soc.*, vol. 126, no. 8, pp. 2399–2408, 2004.
- [17] D. R. Parachalil, J. McIntyre, and H. J. Byrne, "Potential of Raman spectroscopy for the analysis of plasma/serum in the liquid state: recent advances," *Anal. Bioanal. Chem.*, vol. 412, no. 9, pp. 1993–2007, 2020.
- [18] K. Hernández-Vidales, E. Guevara, V. Olivares-Illana, and F. J. González, "Characterization of wild-type and mutant p53 protein by Raman spectroscopy and multivariate methods," *J. Raman Spectrosc.*, Jun. 2019.
- [19] K. D. Miller *et al.*, "Cancer Statistics for Hispanics/Latinos, 2018," *CA. Cancer J. Clin.*, vol. 68, no. 6, pp. 425–445, 2018.
- [20] S. Cui, S. Zhang, and S. Yue, "Raman Spectroscopy and Imaging for Cancer Diagnosis," *Journal of Healthcare Engineering*, vol. 2018. Hindawi

Limited, 2018.

- [21] “How Cancer Is Diagnosed - National Cancer Institute.” [Online]. Available: <https://www.cancer.gov/about-cancer/diagnosis-staging/diagnosis>. [Accessed: 31-Aug-2021].
- [22] Deepa, S. Pundir, and C. S. Pundir, “Detection of tumor suppressor protein p53 with special emphasis on biosensors: A review,” *Anal. Biochem.*, vol. 588, p. 113473, 2020.
- [23] A. Radadiya, W. Zhu, A. Coricello, S. Alcaro, and N. G. J. Richards, “Improving the Treatment of Acute Lymphoblastic Leukemia,” *Biochemistry*, vol. 59, no. 35. NLM (Medline), pp. 3193–3200, 08-Sep-2020.
- [24] S. R. V. Knott *et al.*, “Asparagine bioavailability governs metastasis in a model of breast cancer,” *Nature*, vol. 554, no. 7692, pp. 378–381, 2018.
- [25] G. McNay, D. Eustace, W. E. Smith, K. Faulds, and D. Graham, “Surface-Enhanced Raman Scattering (SERS) and Surface-Enhanced Resonance Raman Scattering (SERRS): A Review of Applications,” *Appl. Spectrosc.*, vol. 65, no. 8, pp. 825–837, Aug. 2011.
- [26] “Stokes Shift, Fluorescence Spectroscopy | Edinburgh Instruments.” [Online]. Available: <https://www.edinst.com/blog/what-is-the-stokes-shift/>. [Accessed: 30-Sep-2020].
- [27] A. Zoubir, *Raman Imaging, techniques and applications*, vol. 168. 2012.
- [28] “DoITPoMS - TLP Library Raman Spectroscopy.” [Online]. Available: <https://www.doitpoms.ac.uk/tlplib/raman/printall.php>. [Accessed: 22-Mar-2021].
- [29] H. J. Butler *et al.*, “Using Raman spectroscopy to characterize biological materials,” *Nat. Protoc.*, vol. 11, no. 4, pp. 664–687, 2016.
- [30] H. J. Butler *et al.*, “Using Raman spectroscopy to characterize biological materials,” *Nat. Protoc.*, vol. 11, no. 4, pp. 664–687, 2016.

- [31] X. Ling, L. G. Moura, M. A. Pimenta, and J. Zhang, “Charge-transfer mechanism in graphene-enhanced Raman scattering,” *J. Phys. Chem. C*, vol. 116, no. 47, pp. 25112–25118, Nov. 2012.
- [32] V. Valeš, K. Drogowska-Horná, V. L. P. Guerra, and M. Kalbáč, “Graphene-enhanced Raman scattering on single layer and bilayers of pristine and hydrogenated graphene,” *Sci. Rep.*, vol. 10, no. 1, pp. 1–9, Dec. 2020.
- [33] S. Bonhommeau and S. Lecomte, “Tip-Enhanced Raman Spectroscopy: A Tool for Nanoscale Chemical and Structural Characterization of Biomolecules,” *ChemPhysChem*, vol. 19, no. 1, pp. 8–18, Jan. 2018.
- [34] N. Kuhar, S. Sil, T. Verma, and S. Umopathy, “Challenges in application of Raman spectroscopy to biology and materials,” *RSC Adv.*, vol. 8, no. 46, pp. 25888–25908, 2018.
- [35] A. Childs, E. Vinogradova, F. Ruiz-Zepeda, J. J. Velazquez-Salazar, and M. Jose-Yacamán, “Biocompatible gold/silver nanostars for surface-enhanced Raman scattering,” *J. Raman Spectrosc.*, vol. 47, no. 6, pp. 651–655, Jun. 2016.
- [36] J. Langer *et al.*, “Present and future of surface-enhanced Raman scattering,” *ACS Nano*, vol. 14, no. 1, pp. 28–117, 2020.
- [37] E. C. Le Ru, E. Blackie, M. Meyer, and P. G. Etchegoint, “Surface enhanced raman scattering enhancement factors: A comprehensive study,” *J. Phys. Chem. C*, vol. 111, no. 37, pp. 13794–13803, 2007.
- [38] F. Domenici, A. R. Bizzarri, and S. Cannistraro, “Surface-enhanced Raman scattering detection of wild-type and mutant p53 proteins at very low concentration in human serum,” *Anal. Biochem.*, vol. 421, no. 1, pp. 9–15, 2012.
- [39] R. Austin, H. Frauenfelder, and S. Chan, “The physics of proteins: an introduction to biological physics and molecular biophysics,” *Annu. Rev. Biophys. Biomol. Struct.*, vol. 36, pp. 261–80, 2010.

- [40] G. J. Thomas, “Raman spectroscopy of protein and nucleic acid assemblies,” *Annu. Rev. Biophys. Biomol. Struct.*, vol. 28, pp. 1–27, 1999.
- [41] C. Krafft and J. Popp, “Micro-Raman spectroscopy in medicine,” *Phys. Sci. Rev.*, vol. 4, no. 10, pp. 1–15, 2019.
- [42] J. De Gelder, K. De Gussem, P. Vandenabeele, and L. Moens, “Reference database of Raman spectra of biological molecules,” *J. Raman Spectrosc.*, vol. 38, no. 9, pp. 1133–1147, Sep. 2007.
- [43] N. Lakomkin and C. G. Hadjipanayis, “The Use of Spectroscopy Handheld Tools in Brain Tumor Surgery: Current Evidence and Techniques,” *Front. Surg.*, vol. 6, p. 30, May 2019.
- [44] K. Czamara, K. Majzner, M. Z. Pacia, K. Kochan, A. Kaczor, and M. Baranska, “Raman spectroscopy of lipids: A review,” *J. Raman Spectrosc.*, vol. 46, no. 1, pp. 4–20, 2015.
- [45] Y. Zhang, H. Hong, and W. Cai, “Imaging with Raman Spectroscopy,” *Curr. Pharm. Biotechnol.*, 2010.
- [46] E. Wiercigroch *et al.*, “Raman and infrared spectroscopy of carbohydrates: A review,” *Spectrochim. Acta - Part A Mol. Biomol. Spectrosc.*, vol. 185, pp. 317–335, 2017.
- [47] K. Inge, *Analysis of Multivariate and High-dimensional Data*. Cambridge Series in Statistical and Probabilistic Mathematics, 2014.
- [48] I. T. Jolliffe, “Principal Component Analysis. Second Edition,” *Springer Ser. Stat.*, vol. 98, p. 487, 2002.
- [49] M. Ringnér, “What is principal component analysis?,” *Nat. Biotechnol.*, vol. 26, no. 3, pp. 303–304, 2008.
- [50] A. Momenpour Tehran Monfared and H. Anis, “An improved partial least-squares regression method for Raman spectroscopy,” *Spectrochim. Acta - Part A Mol. Biomol. Spectrosc.*, vol. 185, no. May, pp. 98–103, 2017.

- [51] H. Abdi, “Partial least squares regression and projection on latent structure regression (PLS Regression),” *Wiley Interdiscip. Rev. Comput. Stat.*, vol. 2, no. 1, pp. 97–106, 2010.
- [52] “Máquinas vectoriales de soporte para la clasificación binaria - MATLAB & Simulink - MathWorks América Latina.” [Online]. Available: <https://la.mathworks.com/help/stats/support-vector-machines-for-binary-classification.html#bsr5b42>. [Accessed: 30-Dec-2020].
- [53] W. S. Noble, “What is a support vector machine?,” *Nat. Biotechnol.*, vol. 24, no. 12, pp. 1565–1567, 2006.
- [54] E. L. Allwein, R. E. Schapire, and Y. Singer, “Reducing multiclass to binary: A unifying approach for margin classifiers,” *J. Mach. Learn. Res.*, vol. 1, no. 2, pp. 113–141, Mar. 2001.
- [55] E. Guevara, J. C. Torres-Galván, M. G. Ramírez-Elías, C. Luevano-Contreras, and F. J. González, “Use of Raman spectroscopy to screen diabetes mellitus with machine learning tools,” *Biomed. Opt. Express*, vol. 9, no. 10, p. 4998, 2018.
- [56] “Introduction to Computational Chemistry, 3rd Edition | Wiley.” [Online]. Available: <https://www.wiley.com/en-us/Introduction+to+Computational+Chemistry%2C+3rd+Edition-p-9781118825990>. [Accessed: 28-Jul-2020].
- [57] B. Sjöberg, S. Foley, B. Cardey, and M. Enescu, “An experimental and theoretical study of the amino acid side chain Raman bands in proteins,” *Spectrochim. Acta Part A Mol. Biomol. Spectrosc.*, vol. 128, pp. 300–311, 2014.
- [58] R. Cabrera-Alonso, E. Guevara, M. G. Ramírez-Elías, B. Moncada, and F. J. González, “Surface-enhanced Raman scattering of hydroquinone assisted by gold nanorods,” *J. Nanophotonics*, vol. 13, no. 03, p. 1, 2019.
- [59] L. Updated, “Genes and disease,” *The Lancet*, vol. 337, no. 8748, p. 1031, 1991.

- [60] X. Lu, “P53: A Target and a Biomarker of Cancer Therapy?,” in *Recent Advances in Cancer Research and Therapy*, Elsevier Inc., 2012, pp. 197–213.
- [61] S. P. Deb, *Mutant p53 and MDM2 in Cancer*, vol. 85. 2014.
- [62] A. Bisio, Y. Ciribilli, G. Fronza, A. Inga, and P. Monti, “TP53 mutants in the tower of babel of cancer progression,” *Hum. Mutat.*, vol. 35, no. 6, pp. 689–701, 2014.
- [63] M. López, M. Anzola, N. Cuevas-Salazar, J. M. Aguirre, and M. Martinez de Pancorbo, “P53, Un Gen Supresor Tumoral,” *Gac. Médica Bilbao*, vol. 98, no. 1, pp. 21–27, 2001.
- [64] B. Vogelstein, D. Lane, and A. J. Levine, “Surfing the p53 network,” *Nature*, vol. 408, no. November, pp. 307–310, 2000.
- [65] M. Kulesza, D. M. Agnieszka, and B. Pieńkowska-Grela, “Repair or perish - The role of p53 protein in a cell’s life,” *Nowotwory*, vol. 69, no. 5–6, pp. 168–178, 2019.
- [66] “PDB-101: Molecule of the Month: p53 Tumor Suppressor.” [Online]. Available: <https://pdb101.rcsb.org/motm/31>. [Accessed: 05-Apr-2020].
- [67] Y. Cho, S. Gorina, P. D. Jeffrey, and N. P. Pavletich, “Crystal structure of a p53 tumor suppressor-DNA complex: understanding tumorigenic mutations Yunje Cho, Svetlana Gorina, Philip D. Jeffrey and Nikola P. Pavletich,” vol. 5170, no. 19, 1994.
- [68] Y. Cho, S. Gorina, P. D. Jeffrey, and N. P. Pavletich, “Tumor of a p53 Structure Crystal Complex : Suppressor- DNA Tumorigenic Mutations Understanding,” vol. 265, no. 5170, pp. 346–355, 2013.
- [69] S. Di Agostino, G. Fontemaggi, S. Strano, G. Blandino, and G. D’Orazi, “Targeting mutant p53 in cancer: The latest insights,” *Journal of Experimental and Clinical Cancer Research*, vol. 38, no. 1. Journal of Experimental & Clinical Cancer Research, pp. 1–3, 2019.

- [70] F. Mantovani, L. Collavin, and G. Del Sal, "Mutant p53 as a guardian of the cancer cell," *Cell Death Differ.*, vol. 26, no. 2, pp. 199–212, 2019.
- [71] Y. Stein, V. Rotter, and R. Aloni-grinstein, "Gain-of-Function Mutant p53 : All the Roads Lead to Tumorigenesis," 2019.
- [72] W. A. Freed-Pastor and C. Prives, "Mutant p53: One name, many proteins," *Genes Dev.*, vol. 26, no. 12, pp. 1268–1286, 2012.
- [73] K. A. Roszkowska, S. Gizinski, M. Sady, Z. Gajewski, and M. B. Olszewski, "Gain-of-function mutations in p53 in cancer invasiveness and metastasis," *Int. J. Mol. Sci.*, vol. 21, no. 4, 2020.
- [74] F. Mantovani and L. Collavin, "Mutant p53 as a guardian of the cancer cell," *Cell Death Differ.*, pp. 199–212, 2019.
- [75] M. T. Boyd and N. Vlatkovic, "P53: A molecular marker for the detection of cancer," *Expert Opinion on Medical Diagnostics*, vol. 2, no. 9. Europe PMC Funders, pp. 1013–1024, Sep-2008.
- [76] P. A. Hall, P. H. McKee, P. H. Du Menage, R. Dover, and D. P. Lane, "High levels of p53 protein in UV-irradiated normal human skin," *Oncogene*, vol. 8, no. 1, pp. 203–207, 1993.
- [77] R. S. Camplejohn and J. Rutherford, "p53 functional assays: Detecting p53 mutations in both the germline and in sporadic tumours," *Cell Prolif.*, vol. 34, no. 1, pp. 1–14, 2001.
- [78] P. M. Duddy, A. M. Hanby, D. M. Barnes, and R. S. Camplejohn, "Improving the Detection of p53 Mutations in Breast Cancer by Use of the FASAY, a Functional Assay," 2000.
- [79] Y. Tomizawa *et al.*, "Correlation between the status of the p53 gene and survival in patients with stage I non-small cell lung carcinoma," *Oncogene*, vol. 18, no. 4, pp. 1007–1014, 1999.
- [80] M. Zhao, J. A. Laissue, A. Zimmermann, and N. X. Zhang, "Immunohistochemical analysis of p53 protein overexpression in liver cell

- dysplasia and in hepatocellular carcinoma,” *Virchows Arch.*, vol. 424, no. 6, pp. 613–621, 1994.
- [81] M. Alhajj and A. Farhana, “Enzyme Linked Immunosorbent Assay,” Mar. 2020.
- [82] T. Mahmood and P. C. Yang, “Western blot: Technique, theory, and trouble shooting,” *N. Am. J. Med. Sci.*, vol. 4, no. 9, pp. 429–434, Sep. 2012.
- [83] S. Signorelli, S. Cannistraro, and A. R. Bizzarri, “Structural Characterization of the Intrinsically Disordered Protein p53 Using Raman Spectroscopy,” *Appl. Spectrosc.*, vol. 71, no. 5, pp. 823–832, 2017.
- [84] A. R. Bizzarri, I. Moscetti, and S. Cannistraro, “Analytica Chimica Acta Surface enhanced Raman spectroscopy based immunosensor for ultrasensitive and selective detection of wild type p53 and mutant,” *Anal. Chim. Acta*, vol. 1029, pp. 86–96, 2018.
- [85] P. Owens, N. Phillipson, J. Perumal, G. M. O’Connor, and M. Olivo, “Sensing of p53 and EGFR biomarkers using high efficiency SERS substrates,” *Biosensors*, vol. 5, no. 4, pp. 664–677, 2015.
- [86] T. Analyst and I. Busines, “Simultaneous evaluation of p53 and p21 expression level for early cancer diagnosis using SERS technique,” no. December 2014, 2013.
- [87] A. Beckett and D. Gervais, “What makes a good new therapeutic l-asparaginase?,” *World J. Microbiol. Biotechnol.*, vol. 35, no. 10, pp. 1–13, 2019.
- [88] A. Radadiya, W. Zhu, A. Coricello, S. Alcaro, and N. G. J. Richards, “Improving the Treatment of Acute Lymphoblastic Leukemia,” *Biochemistry*, vol. 59, no. 35, pp. 3193–3200, 2020.
- [89] “Asparagine | C₄H₈N₂O₃ - PubChem.” [Online]. Available: <https://pubchem.ncbi.nlm.nih.gov/compound/Asparagine>. [Accessed: 04-

Jan-2021].

- [90] K. Yamada, D. Hashizume, T. Shimizu, and S. Yokoyama, "L-Asparagine," *Acta Crystallogr. Sect. E Struct. Reports Online*, vol. 63, no. 9, Aug. 2007.
- [91] "Human Metabolome Database: Showing metabocard for L-Asparagine (HMDB0000168)." [Online]. Available: <https://hmdb.ca/metabolites/HMDB0000168>. [Accessed: 04-Jan-2021].
- [92] C. T. Peng, K. H. Wu, S. J. Lan, J. J. P. Tsai, F. J. Tsai, and C. H. Tsai, "Amino acid concentrations in cerebrospinal fluid in children with acute lymphoblastic leukemia undergoing chemotherapy," *Eur. J. Cancer*, vol. 41, no. 8, pp. 1158–1163, 2005.
- [93] M. Luo, M. Brooks, and M. S. Wicha, "Asparagine and Glutamine: Co-conspirators Fueling Metastasis," *Cell Metab.*, vol. 27, no. 5, pp. 947–949, 2018.
- [94] Q. Qi, O. Obianyo, Y. Du, H. Fu, S. Li, and K. Ye, "Blockade of Asparagine Endopeptidase Inhibits Cancer Metastasis," 2017.
- [95] "Asparagine-linked oligosaccharides associated with metastatic cancer - PubMed." [Online]. Available: <https://pubmed.ncbi.nlm.nih.gov/2701653/>. [Accessed: 04-Jan-2021].
- [96] A. Erdogan, S. Koytepe, B. Ates, I. Yilmaz, and T. Seckin, "Preparation of the L-asparaginase-based biosensor with polyimide membrane electrode for monitoring l-asparagine levels in leukemia," *Int. J. Polym. Mater. Polym. Biomater.*, vol. 63, no. 17, pp. 909–917, 2014.
- [97] N. E. Labrou and M. M. Muharram, "Biochemical characterization and immobilization of *Erwinia carotovora* L-asparaginase in a microplate for high-throughput biosensing of L-asparagine," *Enzyme Microb. Technol.*, vol. 92, pp. 86–93, 2016.
- [98] A. J. D. Moreno, P. T. C. Freire, I. Guedes, F. E. A. Melo, J. Mendes-Filho,

- and J. A. Sanjurjo, "Raman scattering studies of monohydrated L-asparagine," *Brazilian J. Phys.*, vol. 29, no. 2, pp. 380–387, 1999.
- [99] S. Sylvestre *et al.*, "Vibrational spectra (FT-IR and FT-Raman), molecular structure, natural bond orbital, and TD-DFT analysis of L-Asparagine Monohydrate by Density Functional Theory approach," *Spectrochim. Acta - Part A Mol. Biomol. Spectrosc.*, vol. 133, pp. 190–200, Dec. 2014.
- [100] B. O. Golichenko, "Raman study of L-Asparagine and L-Glutamine molecules adsorbed on aluminum films in a wide frequency range," *Semicond. Phys. Quantum Electron. Optoelectron.*, vol. 20, no. 3, pp. 297–304, 2017.
- [101] X. Kuang, S. Ye, X. Li, Y. Ma, C. Zhang, and B. Tang, "A new type of surface-enhanced Raman scattering sensor for the enantioselective recognition of d/l-cysteine and d/l-asparagine based on a helically arranged Ag NPs@homochiral MOF," *Chem. Commun.*, vol. 52, no. 31, pp. 5432–5435, 2016.
- [102] J. Guicheteau, L. Argue, A. Hyre, M. Jacobson, and S. D. Christesen, "Raman and surface-enhanced Raman spectroscopy of amino acids and nucleotide bases for target bacterial vibrational mode identification," in *Chemical and Biological Sensing VII*, 2006, vol. 6218, p. 62180O.
- [103] J. Anderson, David, "Determination of the lower limit of detection.," *Clin. Chem.*, vol. 35, no. 10, pp. 2152–2153, 1989.
- [104] W. A. Freed-Pastor and C. Prives, "Mutant p53: One name, many proteins," *Genes Dev.*, vol. 26, no. 12, pp. 1268–1286, 2012.
- [105] A. C. Joerger, C. A. Hwee, D. B. Veprintsev, C. M. Blair, and A. R. Fersht, "Structures of p53 cancer mutants and mechanism of rescue by second-site suppressor mutations," *J. Biol. Chem.*, vol. 280, no. 16, 2005.
- [106] T. Kawaguchi *et al.*, "The relationship among p53 oligomer formation, structure and transcriptional activity using a comprehensive missense mutation library," *Oncogene*, vol. 24, pp. 6976–6981, 2005.

- [107] S. Yamamoto and T. Iwakuma, “cancers Regulators of Oncogenic Mutant TP53 Gain of Function,” 2018.
- [108] “XploRA™ PLUS - HORIBA.” [Online]. Available: https://www.horiba.com/en_en/products/detail/action/show/Product/xploratm-plus-1528/. [Accessed: 04-May-2020].
- [109] J. Zhao, H. Lui, D. I. Mclean, and H. Zeng, “Automated autofluorescence background subtraction algorithm for biomedical raman spectroscopy,” *Appl. Spectrosc.*, vol. 61, no. 11, pp. 1225–1232, 2007.
- [110] A. L. Jenkins, R. A. Larsen, and T. B. Williams, “Characterization of amino acids using Raman spectroscopy,” *Spectrochim. Acta - Part A Mol. Biomol. Spectrosc.*, 2005.
- [111] R. Tuma, “Raman spectroscopy of proteins: From peptides to large assemblies,” *J. Raman Spectrosc.*, vol. 36, no. 4, pp. 307–319, 2005.
- [112] G. Zhu, X. Zhu, Q. Fan, and X. Wan, “Spectrochimica Acta Part A : Molecular and Biomolecular Spectroscopy Raman spectra of amino acids and their aqueous solutions,” *Spectrochim. Acta Part A Mol. Biomol. Spectrosc.*, vol. 78, no. 3, pp. 1187–1195, 2011.
- [113] N. C. Dingari, G. L. Horowitz, J. W. Kang, R. R. Dasari, and I. Barman, “Raman spectroscopy provides a powerful diagnostic tool for accurate determination of albumin glycation,” *PLoS One*, vol. 7, no. 2, 2012.
- [114] A. R. Bizzarri and S. Cannistraro, “Toward Cancer Diagnostics of the Tumor Suppressor p53 by Surface Enhanced Raman Spectroscopy,” *Sensors*, vol. 20, no. 24, p. 7153, Dec. 2020.
- [115] S. Wold, K. Esbensen, and P. Geladi, “Principal component analysis,” *Chemom. Intell. Lab. Syst.*, vol. 2, no. 1–3, pp. 37–52, 1987.
- [116] “Principal component analysis of raw data - MATLAB pca.” [Online]. Available: https://www.mathworks.com/help/stats/pca.html?searchHighlight=pca&s_t

id=srchttitle. [Accessed: 28-Jun-2021].

- [117] B. R. Wood, P. Caspers, G. J. Puppels, S. Pandiancherri, and D. McNaughton, “Resonance Raman spectroscopy of red blood cells using near-infrared laser excitation,” *Anal. Bioanal. Chem.*, vol. 387, no. 5, pp. 1691–1703, 2007.
- [118] A. S. Mondol, N. Töpfer, J. Rüger, U. Neugebauer, J. Popp, and I. W. Schie, “New perspectives for viability studies with high-content analysis Raman spectroscopy (HCA-RS),” *Sci. Rep.*, vol. 9, no. 1, pp. 1–12, Dec. 2019.
- [119] A. Kulkarni, D. Chong, and F. A. Batarseh, “Foundations of data imbalance and solutions for a data democracy,” in *Data Democracy: At the Nexus of Artificial Intelligence, Software Development, and Knowledge Engineering*, Elsevier, 2020, pp. 83–106.
- [120] J. Mendez-Lozoya, “Dispositivos Termoeléctricos para la detección de radiación,” Universidad Autónoma de San Luis Potosí, 2019.
- [121] S. A. Rodriguez-Montelongo *et al.*, “Synthesis, characterization, and toxicity of hollow gold nanoshells,” *J. Nanoparticle Res.*, vol. 20, no. 11, Nov. 2018.
- [122] E. Guevara and F. J. González, “Estimación del factor de intensificación de espectroscopía Raman amplificada en superficie.” 2019.

APÉNDICE

REGRESIONES POLINOMIALES UNIVARIABLES PARA LÍMITE DE DETECCIÓN

```
%% REGRESION CONCENTRACION VS INTENSIDAD
%% Datos

datos =
dlmread('concentracion_vs_intensidad.txt');%archivo txt
con concentraciones
concentracion = datos(:,1); %concentracion medida
intensidad = datos(:,2);
[n,1] = size(intensidad);

%% ajuste a polinomio  $y = mx+b$ 

p1 = polyfit(intensidad, concentracion,1);
cp = p1(1)*intensidad + p1(2); %concentracion predicha

figure
intensidad2 = linspace(1,0.3,n); %rango de evaluacion
f = polyval(p1,intensidad2); %linea de prediccion
plot(intensidad,concentracion,'bo',intensidad2,f,'r-')

xlabel('intensidad [u.a.]')
ylabel('concentracion [uM]')

%% Estadistica del ajuste

RMSE = sqrt(sum((cp-
concentracion).^2)/size(concentracion,1));

TSS = sum((concentracion-mean(concentracion)).^2);
RSS = sum((cp-concentracion).^2);
rsquared = 1 - RSS/TSS;

%% Curva de ajuste  $Y_p = mY_e + b$ . Variables predichas y
variables experimentales
p2 = polyfit(concentracion,cp,1);% parametros del
polinomio

figure
x = linspace(-2,14,n);
f2 = polyval(p2,x); %linea de prediccion
plot(concentracion,cp,'bo',x, f2,'g-', x, x,'r-')

xlabel('Concentracion medida')
```

```

ylabel('Concentracion predicha')

%% Limite de deteccion

s = sqrt((sum((cp-concentracion).^2))/(n-2));
m = p2(1); %pendiente del ajuste entre la concentracion
predicha vs concentracion experimental

LOD = 3*s/m

%% REGRESION CONCENTRACION VS FWHM
%% Datos

datos = dlmread('Concentracion_vs_fwhm.txt');%archivo
txt con intensidades vs concentraciones
concentracion = datos(:,1); %concentracion medida
fwhm = datos(:,2);
[n,1] = size(fwhm);

%% ajuste a polinomio y = mx+b. Prediccion de
concentracion (y) con variable x

p1 = polyfit(fwhm, concentracion,1);% parametros del
polinomio
cp = p1(1)*fwhm + p1(2); %concentraciones predichas

figure
fwhm2 = linspace(160,40,n); %rango de evaluacion
f = polyval(p1,fwhm2); %linea de prediccion
plot(fwhm,concentracion,'bo',fwhm2,f,'r-')

xlabel('FWHM [1/cm]')
ylabel('concentracion [uM]')

%% Estadistica del ajuste

RMSE = sqrt(sum((cp-
concentracion).^2)/size(concentracion,1));

TSS = sum((concentracion-mean(concentracion)).^2);
RSS = sum((cp-concentracion).^2);
rsquaredFWHM = 1 - RSS/TSS;

%% Curva de ajuste Yp = mYe + b. Variables predichas y
variables experimentales
p2 = polyfit(concentracion,cp,1);% parametros del
polinomio

```

```

figure
x = linspace(-2,14,n);
f2 = polyval(p2,x); %linea de prediccion
plot(concentracion,cp,'bo',x,f2,'g-', x, x,'r-')

xlabel('Concentracion medida')
ylabel('Concentracion predicha')

%% Limite de deteccion

s = sqrt((sum((cp-concentracion).^2))/(n-2));
m = abs(p2(1)); %pendiente del ajuste

LOD = 3*s/m

```

PLSR PARA LÍMITE DE DETECCIÓN

```

clear all;
close all;

%% Load data

valores =
dlmread('wtp53_14espectros_normalizados.txt');%cada
columna corresponde a un espectro, el ultimo renglon es
de concentraciones
datos = transpose(valores); % matriz de espectros y
concentraciones
%cada renglon corresponde a un espectro, la ultima
columna es de concentraciones
[~,m] = size(datos);
samples = datos(:,1:(m-1));
concentration = datos(:,m);

%% PCA para saber cuantas nuevas variables mantener

[pcaLoadings, pcaScores, ~, ~, explained] =
pca(samples, 'Centered', false);
% porcentaje de varianza a mantener
Varianza = 95/100;
numComponentes = find(cumsum(explained)/sum(explained)
>= Varianza, 1);
pcaScores = pcaScores(:,1:numComponentes);

%% PLS con 2 componentes y cross validation

X = samples;
y = concentration;

```

```

[n,p] = size(X); %n=numero de muestras, p=mediciones

[Xloadings,Yloadings,Xscores,Yscores,betaPLS,PLSPctVar,
PLSmsep] = plsregress(X,y,2,'CV',14);
yfitPLS = [ones(n,1) X]*betaPLS; %concentraciones
predichas por el modelo

%% VARIAZA PLOT

plot(1:2,cumsum(100*PLSPctVar(2,:)),'-bo');
xlabel('Numero de PLS componentes');
ylabel('Porcentaje de Varianza');

%% residuos

figure;
residuals = y - yfitPLS;
stem(residuals)
xlabel('Observation');
ylabel('Residual');

%% r^2

TSS = sum((y-mean(y)).^2);
RSS_PLS = sum((y-yfitPLS).^2);
rsquaredPLS = 1 - RSS_PLS/TSS;

%% RMSE

RMSE = sqrt((RSS_PLS)/n)

%% LDD LIMITE DE DETECCION

s = sqrt(sum(residuals.^2)/(n-2)); %desviacion estandar
de residuos
M = rsquaredPLS;
LDD = 3*s/M

%% PLSR plot

%regresion

p1 = polyfit(y,yfitPLS,1);
Xp = polyval(p1,X);
X = linspace(-2,14,n);

figure;
plot(y,yfitPLS,'bo', X,X, 'r-');

xlabel('Concentración medida [uM]');

```

```
ylabel('Concentración predicha [uM]');  
hold on;
```

ANÁLISIS DE COMPONENTES PRINCIPALES

```
% COMPONENTES PRINCIPALES DE MATRIZ P53 WT Y P53  
MUTANTE
```

```
clear all  
close all
```

```
%% leer datos de entrada
```

```
load all_spectra_25.txt;  
data = all_spectra_25; % (mediciones=renglones,  
espectros=columnas) = MxN  
[m, n] = size(data); % m = renglones, n = columnas
```

```
%% normalizar datos (cada columna con un maximo = 1)
```

```
dataNormalized = zeros(m,n); %inicializacion de matriz  
de intensidades
```

```
for i = 1:n  
    In = data(:,i);  
    Imin = min(In);  
    Imax = max(In);  
    dataNormalized(:,i) = (In-Imin)/(Imax-Imin);  
end
```

```
%% Acomodo de datos
```

```
dt = transpose(datos);  
% dt=transpose(dataNormalized); %Para que cada espectro  
sea un renglon = n  
dimensiones=size(dt); %Renglones->Espectro, Columnas-  
>RamanShif = m  
numEspectros=dimensiones(1);  
numMediciones=dimensiones(2);
```

```
%% PCA
```

```
[pcLoadings,pcScores,l,~,p] = pca(dt,'centered',true,  
'economy', true);
```

```
%pcLoadings-> Matriz de loadings nxm  
%pcScores-> matriz de scores (nxn)
```

```
%% etiquetas
```

```
labelsCell = cell(25, 1); %(modificar segun datos)
```

```

% grupo de 25
labelsCell(1:3) = {'18.9'};
labelsCell(4:6) = {'12.6'};
labelsCell(7:8) = {'8.4'};
labelsCell(9:10) = {'4.2'};
labelsCell(11:12) = {'2.8'};
labelsCell(13:14) = {'2.1'};
labelsCell(15:16) = {'1.68'};
labelsCell(17:19)={'m273'};
labelsCell(20:22) = {'m343'};
labelsCell(23:25) = {'m344'};

%% Gráfica de pcScores

figure('Name','pcScores Vs pcScores', 'NumberTitle',
'on')
j=1;k=2;%indices de pcScores
gscatter((pcScores(:,j)),(pcScores(:,k)),labelsCell,'rg
bmck','.',25)

xlabel(strcat('PC ',num2str(jj),'(%)'),'FontSize',14)
ylabel(strcat('PC ',num2str(kk),'(%)'),'FontSize',14)
set(gca,'FontSize',12,'YGrid','on','XGrid','on')

```

SUPPORT VECTOR MACHINE

```

clear all
close all

%% Matriz de espectros

% (mediciones=renglones, espectros=columnas) = MxN
load all_spectra_12.txt;
data = all_spectra_12;
[m, n] = size(data); % m = renglones, n = columnas

%% normalizar datos (cada columna con un maximo = 1)

dataNormalized = zeros(m,n); %inicializacion de matriz
de intensidades

for i = 1:n
    In = data(:,i);
    Imin = min(In);
    Imax = max(In);
    dataNormalized(:,i) = (In-Imin)/(Imax-Imin);
end

```



```

ramanSpectra = dataNormalized';

%% Etiquetas del clasificador

sampleID = cell(12, 1); %(modificar segun datos)

sampleID(1:3) = {'wt'};
sampleID(4:6)={'m273'};
sampleID(7:9) = {'m343'};
sampleID(10:12) = {'m344'};

for iSample=1:3
    sampleID{iSample} = 'wt';
end
for iSample=4:12
    sampleID{iSample} = char(sampleID{iSample});
end

%% PCA

[pcaLoadings, pcaScores, ~, ~, explained] =
pca(samples, 'Centered', false);
% porcentaje de varianza a mantener
Varianza = 95/100;
numComponentes = find(cumsum(explained)/sum(explained)
>= Varianza, 1);
pcaScores = pcaScores(:,1:numComponentes);

%% SVM

plotROCflag = false;
% Train and test a multi-class SVM model with LOOC

X = pcaScores;
Y = sampleID;

% SVM template

t = templateSVM('Standardize', false, 'KernelFunction',
'linear','KernelScale', 'auto');

% Entrenamiento de SVM multiclase

classLabels = unique(Y);
[Mdl, HyperparameterOptimizationResults] =
fitcecoc(X,Y,
'Learners',t,'ClassNames',classLabels,'Verbose',2,

```

```

'OptimizeHyperparameters','auto',
'HyperparameterOptimizationOptions',
struct('Optimizer',
'bayesopt','SaveIntermediateResults', true,
'Repartition', true, 'Kfold', numel(Y));

CVMdl = crossval(Mdl, 'leaveout', 'on');
oosLoss = kfoldLoss(CVMdl);
oofLabel = kfoldPredict(CVMdl);% Predict labels

%% Confusion matrix

C = confusionmat(Y,oofLabel,'order',classLabels);
groundTruth = nan(size(Y));
predictedLabels = nan(size(oofLabel));
for idx = 1:numel(Y)
    if strcmp(Y{idx}, 'wt')
        groundTruth(idx) = 1;
    elseif strcmp(Y{idx}, 'm273')
        groundTruth(idx) = 2;
    elseif strcmp(Y{idx}, 'm343')
        groundTruth(idx) = 3;
    else
        groundTruth(idx) = 4;
    end
    if strcmp(oofLabel{idx}, 'wt')
        predictedLabels(idx) = 1;
    elseif strcmp(oofLabel{idx}, 'm273')
        predictedLabels(idx) = 2;
    elseif strcmp(oofLabel{idx}, 'm343')
        predictedLabels(idx) = 3;
    else
        predictedLabels(idx) = 4;
    end
end
end

[c_matrix,Result]=
confusion.getMatrix(groundTruth,predictedLabels);

c_table = table(C(:,1), C(:,2), C(:,3),
C(:,4),'VariableNames',classLabels, 'RowNames',
classLabels);
fprintf('Confusion Matrix \n')
disp(c_table)

% plot confusion matrix
xvalues = {'wt' 'm273' 'm343' 'm344'};
yvalues = xvalues;

% New figure

```

```

hFig = figure;
h = heatmap(xvalues,yvalues,c_matrix);
h.XLabel = 'Etiquetas Predichas';
h.YLabel = 'Etiquetas Reales';
h.FontSize = 12;
h.ColorbarVisible = 'off';
set(hFig, 'color', 'w')

```

CÁLCULO DE FACTOR DE MEJORAMIENTO SERS

```

%% Load SERS data

```

```

datos = load('SERS_ASP_A1.txt');
bandas = load('bandas_SERS_Arreglo1.txt');

```

```

%% Variables

```

```

waveNumber = datos(:,1);
ramanSpectrum = datos(:,2);
SERSspectrum = datos(:,3);

```

```

%matriz para factores de mejoramiento
[m,n] = size(bandas);
EFs = zeros(m,n);

```

```

%% SERS enhancement factor

```

```

for i = 1:m
    b = bandas(i);
    [EF, paramStruct] =
sers_enhancement_factor(waveNumber, ramanSpectrum,
SERSspectrum,...
    'band', b,
    'lambda',785,'molWeight',132.12,'NA',0.40,'rho',1.543,'
surfArea',1.5,'showPlot', true);
    EFs(i) = EF;

```

```

end

```