

(12) SOLICITUD INTERNACIONAL PUBLICADA EN VIRTUD DEL TRATADO DE COOPERACIÓN EN MATERIA DE PATENTES (PCT)

(19) Organización Mundial de la Propiedad
Intelectual
Oficina internacional



(10) Número de Publicación Internacional
WO 2010/146199 A2

(43) Fecha de publicación internacional
23 de diciembre de 2010 (23.12.2010)

PCT

- (51) **Clasificación Internacional de Patentes:** Sin clasificar Universidad Complutense de Madrid, Avda. Complutense, s/n, E-28040 Madrid (ES).
- (21) **Número de la solicitud internacional:** PCT/ES2010/000267
- (22) **Fecha de presentación internacional:** 18 de junio de 2010 (18.06.2010)
- (25) **Idioma de presentación:** español
- (26) **Idioma de publicación:** español
- (30) **Datos relativos a la prioridad:** P200901443 18 de junio de 2009 (18.06.2009) ES
- (71) **Solicitante (para todos los Estados designados salvo US):** UNIVERSIDAD COMPLUTENSE DE MADRID [ES/ES]; Avda. Séneca, 2, E-28040 Madrid (ES).
- (72) **Inventores; e**
- (75) **Inventores/Solicitantes (para US solamente):** CÁCERES, Jorge Omar [IT/ES]; Facultad de Ciencias Químicas de la Universidad Complutense de Madrid, Avda. Complutense, s/n, E-28040 Madrid (ES). MARCOS MARTÍNEZ, Daniel [ES/ES]; Facultad de Ciencias Químicas de la Universidad Complutense de Madrid, Avda. Complutense, s/n, E-28040 Madrid (ES). ORTEGA NOGALES, Francisco Jesús [ES/ES]; Facultad de Ciencias Químicas de la Universidad Complutense de Madrid, Avda. Complutense, s/n, E-28040 Madrid (ES). DEL VALLE ÁVILA, Marcos [ES/ES]; Facultad de Ciencias Químicas de la
- (81) **Estados designados (a menos que se indique otra cosa, para toda clase de protección nacional admisible):** AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PE, PG, PH, PL, PT, RO, RS, RU, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.
- (84) **Estados designados (a menos que se indique otra cosa, para toda clase de protección regional admisible):** ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), euroasiática (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), europea (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).
- Publicada:**
- *sin informe de búsqueda internacional, será publicada nuevamente cuando se reciba dicho informe (Regla 48.2(g))*

(54) **Title:** INSTANTANEOUS IDENTIFICATION AND CHARACTERISATION OF SAMPLES USING THE DYNAMIC COMBINATION OF LASER ABLATION AND MATHEMATICAL ALGORITHMS

(54) **Título :** IDENTIFICACIÓN Y CARACTERIZACIÓN INSTANTÁNEA DE MUESTRAS MEDIANTE LA COMBINACIÓN DINÁMICA DE ABLACIÓN LÁSER Y ALGORITMOS MATEMÁTICOS

(57) **Abstract:** The invention relates to the instantaneous identification and characterisation of samples with no prior treatment, using the dynamic combination of laser ablation and mathematical algorithms. The method consists in analysing a single laser pulse in order to obtain a spectrum of the sample, which is compared with a dynamic spectral database using a neural network that allows a large amount of data to be handled in a very short space of time with 100% sensitivity and specificity. The identification and characterisation technique is quick and direct and allows the correlation of analytical data with organoleptic properties or quality properties, as well as the determination of contaminating, pathogenic or bioactive compounds and substances present in samples, in all states of the material and without any form of prior preparation.

(57) **Resumen:** Identificación y caracterización instantánea de muestras sin tratamiento previo mediante la combinación dinámica de ablación láser y algoritmos matemáticos. El método se basa en el análisis de un único pulso láser del cual se obtiene un espectro de la muestra que se compara con una base de datos espectral dinámica empleando una red neuronal que permite manejar un gran número de datos en un tiempo muy corto con un 100% de sensibilidad y especificidad. La técnica identificación y caracterización es rápida, directa y permite la correlación de los datos analíticos con propiedades organolépticas o de calidad así como la determinación de sustancias, compuestos contaminantes, patógenos o bioactivos presentes en muestras en cualquier estado de la materia sin ningún tipo de preparación previa.



WO 2010/146199 A2

TITULO

Identificación y caracterización instantánea de muestras mediante la combinación dinámica de ablación láser y algoritmos matemáticos.

5

OBJETO Y CAMPO DE LA INVENCION

Esta invención se enmarca en el campo de análisis espectral de compuestos químicos presentes en muestras líquidas o sólidas. De forma más concreta, la invención se refiere a un método de identificación y caracterización mediante la determinación cualitativa de compuestos químicos en muestras, sin ningún tipo de tratamiento previo y que permite la correlación con propiedades organolépticas o de calidad con aplicación en diversos campos como en muestras de aguas, alimentos o muestras biológicas. El método combina la técnica espectroscópica láser y el tratamiento espectral basado en algoritmos matemáticos.

Este tipo de algoritmos proporciona un método adecuado con un 100% de sensibilidad y 100% especificidad en la discriminación de compuestos, sustancias o materiales que permiten su identificación y correlación con propiedades organolépticas o de calidad, trazabilidad de alimentos o la detección de contaminantes o patógenos de diversa naturaleza.

Esta invención permite la correlación de los datos analíticos con propiedades organolépticas o de calidad y/o la determinación de sustancias, compuestos, contaminantes o agentes patógenos, en muestras de aguas (pluviales, subterráneas, afluentes naturales, cloacales, de residuos industriales), alimentos (zumos de fruta, aceites vegetales, leche y sus derivados), y muestras biológicas (sangre, suero, orina, etc.), muestras sólidas (aceros, aleaciones, maderas, huesos, rocas, minerales, etc.).

ESTADO DE LA TÉCNICA

La técnica de LIBS (Laser Induced Breakdown Spectroscopy) es un método rápido y versátil para efectuar el análisis de distintos tipos de

muestras que pueden ser inaccesibles o tediosos por técnicas analíticas convencionales, siendo particularmente útil en el análisis de muestras con una matriz compleja, (J.O. Caceres et al. Spectrochimica Acta B: Atomic Spectroscopy 56(6) 831 (2001), WB Lee, et al, Appl Spectrosc. Rev. 39(1) 27 (2004)).

Sin embargo, esta técnica ha sido aplicable solo al análisis elemental y en muchas situaciones existe un número considerable de técnicas alternativas disponibles con una sensibilidad superior a LIBS. A pesar de esta aparente desventaja, la decisión sobre la elección de un método para análisis elemental puede hacerse favorable a la técnica LIBS debido a algunas de sus características que le confieren una ventaja considerable respecto a técnicas convencionales como son a) la eliminación de la necesidad de preparar la muestra para el análisis, b) el análisis puede ser llevado a cabo en cualquiera de los estados de la materia (sólido, líquido, gas), c) el análisis se realiza en unos pocos segundos, d) sólo una muy pequeña cantidad de la muestra, en el orden de los microgramos, es vaporizada desde la superficie de la muestra y hace que el método se considere virtualmente no destructivo y e) los límites de detección se encuentran para las mayoría de las muestras en el orden de unas pocas partes por millón.

Es evidente la gran importancia de esta técnica para el análisis espectroquímico cuantitativo. Sin embargo, sorprendentemente, mucha menos atención ha sido puesta en el potencial de la técnica LIBS como herramienta de análisis cualitativo. En este caso, no se busca una composición química detallada, más bien una identificación instantánea de la muestra utilizando una característica única de LIBS, la cual es capaz de generar una "huella digital" de la muestra. Esta "huella digital" es un espectro LIBS de la muestra, el cual proviene exclusivamente de ésta y por lo tanto depende de su matriz. Además, debido a la naturaleza del espectro de emisión, el cual está dominado por líneas iónicas, que a menudo se ven inhibidas en su relación directa entre la concentración elemental en la

muestra y sus intensidades, proporciona un espectro único y solo perteneciente a la muestra bajo análisis.

Utilizando un procedimiento de correlación, el sistema LIBS puede ser entrenado para reconocer espectros procedentes de diferentes muestras, lo cual significa, evaluar la similitud de espectros desconocidos con un banco espectral de muestras clasificadas. Es necesario, por lo tanto, conocer qué técnica de correlación es más adecuada para la identificación de muestras. Su elección, suele estar determinada por el arreglo experimental, el tipo de datos obtenidos y por los requerimientos de tiempo de análisis.

En los últimos años se han utilizado distintas metodologías para realizar la identificación o clasificación de compuestos, en una progresión de complejidad que va desde los relativamente simples procedimientos de correlación lineal no paramétrica (I.B. Gornushkin et al *Anal. Chem.* 71, 5157-5164, (1999), A.J. Lopez et al *Appl. Phys. A* 83, 695-698 (2006)), análisis Fourier (W.H. Flannery et al, *Numerical recipes, the Art of Scientific Computing* Cambridge University Press NY (1986)), el método conocido como Mahalanobis distance (O. Sameck et al *LNCS*, 2124,443 (2001)), hasta ciertos tipos de redes neuronales desde sus primeras aplicaciones a la detección y clasificación en la década de los 90 (I. Ruisanchez et al, *J. Chem. Inf. Comput. Sci* 36, 214 (1996), R.Sattmann et al *Appl. Spect.* 52 456 (1998)). Las redes neuronales proveen los mejores resultados, sin embargo a pesar de su difusión y aplicación a diversos campos de investigación como los minerales y las pinturas artísticas (M:M. Howari, *J. Appl. Spect.* 70, 782 (2003), D. Anglos *Appl. Spect.* 51, 258A (1997)), tienen fuertes limitaciones y no ha sido capaces de dar una solución satisfactoria a muchos problemas prácticos, principalmente debido a incertidumbres en la capacidad de evaluar muestras con una alta sensibilidad y especificidad, o con resultados, con un grado de error o incertidumbre en la identificación que puede ser mayor al 30% siendo totalmente inaceptables para el caso, por ejemplo, de detección e identificación de explosivos (W. Nunes et al, *Appl. Rad. Isotop.* 56, 937 (2002)).

Existen métodos de análisis que combinan diferentes técnicas analíticas con redes neuronales para la detección de sustancias en una muestra. Así, la patente US 6,618,712 describe un método de identificación de bioaerosoles que combina espectros de masas de iones (positivos y negativos) de partículas conocidas utilizado para entrenar una red neuronal y correlacionar los datos obtenidos de una sustancia desconocida, con otra identificada a través del entrenamiento. Estas medidas, no obstante, requieren de un tratamiento previo que introduce incertidumbre en el resultado y alarga el tiempo de análisis. (R.A. Gieray et al. J. Microbiol. Meth. 29, 191 (1997)).

A pesar de la existencia de otros métodos de análisis que se combinan con redes neuronales, el campo de las redes neuronales puede mejorarse en muchas direcciones.

En esta invención se utiliza un número muy alto de puntos espectrales obtenidos de la emisión del plasma de la muestra en todo el espectro visible, combinado con un sistema de algoritmos matemáticos adecuados para manejar un gran número de datos en un tiempo muy corto y de forma eficiente y efectiva, asegurando un 100% de sensibilidad y 100% especificidad como se demostrara más adelante.

El método de análisis se basa en un análisis de un único pulso láser que produce un proceso de vaporización y posterior formación de un plasma de la superficie de la muestra, la obtención del espectro de emisión de este plasma en el orden de unos pocos nanosegundos sin ningún tipo de tratamiento de datos y la posterior comparación espectral con una base de datos espectral dinámica usando algoritmos matemáticos de entrenamiento y comparación de espectros adecuados para el análisis de matrices espectrales que pueden tener 2024x100 puntos como mínimo y con un tiempo de análisis total menor a 1(un) segundo.

El método que se reivindica en la patente permite obtener gran sensibilidad (concentraciones comprendidas entre partes por millón (ppm) y partes por billón (ppb)), la facilidad del manejo de muestras tóxicas o potencialmente peligrosas, y realizar análisis in-situ de las muestras. Dado el

escaso tiempo de cálculo necesario para el tratamiento de los datos suministrados por el láser, hacen que la conjunción Láser-NNs ofrezca unas posibilidades de medida en línea (on-line) desconocidas hasta el momento.

5 DESCRIPCIÓN DE LA INVENCION

La presente invención consiste en un método rápido y directo para la determinación cualitativa y/o análisis espectral completo de compuestos químicos presentes en muestras líquidas o sólidas. Esta invención permite
10 identificación y caracterización de muestras, permitiendo la correlación de los datos analíticos con propiedades organolépticas o de calidad y/o la determinación de sustancias, compuestos, contaminantes, agentes patógenos o bioactivos, en las muestras analizadas.

Mediante el uso de características propias presentes en la técnica LIBS
15 y los algoritmos matemáticos utilizados, la nueva metodología permite la identificación de una muestra sin ningún tipo de preparación, en cualquier estado de la materia, con un solo pulso, en un tiempo menor de un segundo y con una alta sensibilidad y especificidad.

El dispositivo utilizado para la identificación y caracterización de
20 muestras, mostrado en la figura 1, se compone de un láser pulsado (1), en este caso particular se ha utilizado un láser de Nd:YAG, aunque en la practica podría utilizarse cualquier otro tipo de láser tanto de estado sólido o gaseosos (N₂, etc) que permita obtener condiciones energéticas suficientes para poder producir un plasma. Por este motivo pueden utilizarse láseres
25 tanto en el ultra violeta, visible o infrarrojo.

El láser de Nd:YAG utilizado trabaja a una frecuencia de 1 a 20 Hertz a tres longitudes de onda 1024, 532 y 262 nm respectivamente. La duración del pulso es de 4 nanosegundos. Dicho láser puede proporcionar hasta 180 mJ/pulso de energía de salida. Se han utilizado espejos (2) y lentes (3)
30 adecuados a fin de focalizar el haz del láser sobre la muestra (4). La emisión del plasma es colectada utilizando una fibra óptica de 1 metro acoplada al detector, que a su vez es activado por el pulso láser. El detector es un sensor

5 óptico CCD que proporciona 2024 puntos espectrales en un rango de 200 a 1100 nm. La señal del detector es posteriormente comparada con una base espectral almacenada utilizando un sistema de algoritmos matemáticos. Una interfaz gráfica, permite al usuario el control del láser, la observación del espectro de la muestra obtenido, y la representación de los resultados para su evaluación.

10 A pesar de que la calidad del espectro y su reproducibilidad pueden variar considerablemente, no afectan a la capacidad del sistema para la identificación de la muestra. Esto es así debido a que los espectros utilizados como base para la comparación reflejan todas las condiciones de medición; por lo tanto, la precisión y exactitud de la medición del espectro LIBS no es relevante. El sistema es adiestrado para evaluar la similitud de espectros desconocidos con su banco de datos de muestras clasificadas. Esta
15 característica, en donde los parámetros de ablación, por ejemplo la distancia lente-muestra, hace menos crítica la calidad del espectro y por lo tanto, factible la identificación correcta con un único disparo láser.

20 Con el objeto de crear una base de datos espectral dinámica que recoja las "huellas dactilares" de cada muestra, se utiliza una red neuronal (NN) con conexiones hacia delante, concretamente se trata de un modelo de propagación *perceptron (back-propagating perceptron model)*. La NN empleada consta de tres capas denominadas de entrada, de salida y ocultas, formadas por neuronas (único elemento operativo). La capa de entrada se
25 utiliza sólo para la entrada de la matriz de datos en la NN. En las otras capas, se realizan cálculos no lineares. En la capa oculta, cada neurona recibe señales de otras neuronas de entrada, sumándose éstas mediante la función activación, ecuación 1. Después, el resultado es transformado por la función de transferencia, ecuación 2. Finalmente, el resultado es enviado a las
30 neuronas de salida (neuronas en la capa de salida).

$$x_k = \sum_{j=1} w_{jk} \cdot y_j \quad (1)$$

$$y_k = f(x) = \left(\frac{2}{1 + \exp^{(-2 \cdot x_k)}} - 1 \right) \quad (2)$$

5 En las ecuaciones 1 y 2, y_j e y_k representan la salida de las neuronas ocultas (j) y de las neuronas de salida (k) respectivamente, w_{jk} representa el peso entre el j^{th} de las neuronas ocultas y el k^{th} de las neuronas de salida. Como función de transferencia se ha utilizado el algoritmo tangente hiperbólico sigmoideal, ecuación 2, y el resultado utilizado oscila entre 0 y 1.

10 Los pesos fueron parámetros ajustables de las NN asociados con cada una de las conexiones entre neuronas y estas pueden modificar la señal de comunicación entre éstas. El proceso de optimización de la matriz de pesos de la NN se llevó a cabo mediante el algoritmo de entrenamiento de gradiente escalado conjugado SGC. Este algoritmo fue seleccionado con el fin de

15 prevenir el sobreajuste (violation of Occam's razor) y el sobreentrenamiento (I.V. Tetko, et all J.Chem. Inf. Coput. Sci. 35, 826 (1995)). El problema del sobreentrenamiento se refiere al hecho de que la red sólo memoriza el conjunto de aprendizaje y pierde su habilidad para generalizar. Por otra parte, el algoritmo SGC, fue desarrollado para entrenamientos rápidos de sistemas

20 de aprendizaje de redes neuronales (H. Demuth et all, Neural Network Toolbox for use whit Matlab User's guide version 4.0.6. ninth printing revised; The math Works : Natick, MA, 2005). Una importante característica de este tipo de algoritmos de alto rendimiento, es que pueden converger cientos de veces mas rápido que otros tipos de algoritmos, aunque a su vez, consume

25 una gran cantidad de memoria. El SCG minimiza el error de predicción empleando una combinación lineal del error cuadrático medio (MSE), ecuación 3. Esto determina la combinación correcta para producir una red que generalice nuevos datos de entrada dentro del rango de los datos de aprendizaje.

$$MSE = \frac{1}{N} \sum_k^N (r_k - y_k)^2 \quad (3)$$

En la ecuación 3, N corresponde al número de observaciones, y_k , el
5 valor real, r_k , las estimaciones en el modelo de NN y k capa de salida de
NN, respectivamente.

Los algoritmos básicos de backpropagation ajustan los pesos en una
fuerte pendiente de dirección descendente (valores negativos del gradiente),
10 esto es la dirección en la cual la función de rendimiento disminuye más
rápidamente. Resulta que, aunque la función disminuye más rápidamente a lo
largo de los valores negativos de la pendiente, esta no produce
necesariamente la convergencia más rápida. En el algoritmo de gradiente
conjugado, realiza una búsqueda a lo largo de direcciones conjugadas, lo que
15 produce generalmente una convergencia más rápida. (Ver Hagan, M.T., H.B.
Demuth, and M.H. Beale, *Neural Network Design*, Boston, MA: PWS
Publishing, 1996. para una discusión mas amplia de los algoritmos de
gradiente conjugado.)

20 Cada uno de los algoritmos de gradiente conjugado requiere de una
línea de búsqueda en cada interacción. Esta línea de búsqueda es
computacionalmente costosa, porque requiere que la red dé respuesta a
todas las entradas de entrenamiento varias veces para cada búsqueda. El
algoritmo de gradiente conjugado desarrollado por Moller [(M.F. Moller, *Neural*
25 *Networks*, 6 525 (1993)), fue diseñado para prevenir el consumo de tiempo de
la línea de búsqueda.

Todos los algoritmos de gradiente conjugado comienzan con la
búsqueda de la dirección descendente del gradiente (gradiente negativo) en
30 la primera interacción.

$$p_0 = -g_0 \quad (4)$$

Posteriormente, se realiza una línea de búsqueda, para determinar la distancia optima de movimiento a lo largo de esta dirección de búsqueda:

5

$$x_{k+1} = x_k + \alpha_k p_k \quad (5)$$

El procedimiento general para determinar la nueva dirección de búsqueda es la combinación de la nueva pendiente de dirección con la búsqueda anterior:

10

$$p_k = -g_k + \beta_k p_{k-1} \quad (6)$$

Las distintas versiones de los algoritmos de gradiente conjugado se distinguen por la manera en que la constante β_k es calculada, para el algoritmo SGC el procedimiento de actualización es:

15

$$\beta_k = \left(|r_{k+1}|^2 - r_{k+1}^r - r_k \right) / \mu_k \quad (7)$$

Este algoritmo combina el enfoque de región verdadera (usada en el algoritmo de Levenberg-Marquardt (D. Marquardt, J. Appl. Math. 11, 431-441, (1963)), con la aproximación del gradiente conjugado. (ver Moller, para una explicación detallada de este algoritmo).

20

La topología de la NN empleada en este trabajo consistió en una capa de entrada con 2024 neuronas para las variables de entrada (espectro LIBS). Una capa oculta (optimizada después) y una capa de salida con dos neuronas para determinar si el reconocimiento es positivo o negativo. Esta topología con una sola capa oculta fue adecuada para resolver los problemas de todas las muestras estudiadas. Cada modelo de NN empleado en este trabajo se

25

30

diseñó utilizando el programa Matlab, seleccionado por su eficacia en el cálculo vectorial.

Dado que la NN utilizada se basa en el algoritmo supervisado, para optimizar la matriz de pesos es necesario emplear datos de entrada y salida que caractericen adecuadamente el proceso que será modelado. Estos datos han sido tomados directamente de 50 espectros LIBS obtenidos a partir de las muestras. Cada espectro corresponde a un único pulso láser. Tanto las muestras como los datos espectrales obtenidos no tuvieron ningún tipo de tratamiento previo. El conjunto de datos (espectros LIBS de las muestras) fue distribuido al azar en el aprendizaje (80%) y en la verificación (20%) de las muestras, teniendo en cuenta que ningún conjunto de datos o cualquiera de sus réplicas deben ser presentados en la verificación de la muestra.

La precisión es la principal característica de un procedimiento de reconocimiento como recurso para la toma de decisión, motivo por el cual, las métricas para evaluar procesos de detección tienen una importancia significativa e involucran la frecuencia relativa de los reconocimientos correctos e incorrectos que hace un observador a partir de los resultados obtenidos. Las medidas básicas son el número de positivos y negativos, (verdaderos y falsos, PV, NV, FP, FN), a partir de los cuales se calculan la sensibilidad y la especificidad de los procesos de detección.

Un positivo verdadero (PV) corresponde a la detección correcta de una sustancia, compuesto o característica en una muestra, cuando esta realmente existe. Un negativo verdadero (NV) corresponde a la detección negativa de una sustancia, compuesto o característica en una muestra cuando efectivamente esta no existe. Las detecciones falsas (PF, NF) corresponden a los casos en los que la detección no corresponde con la realidad de la muestra.

La sensibilidad y la especificidad son dos métricas del desempeño de un proceso de detección que se construyen a partir del número de PV, PF, NV y NF en una muestra de validación. La sensibilidad de un proceso de detección se refiere a la probabilidad de que una sustancia, compuesto o

característica sea detectada cuando realmente existe. La sensibilidad se especifica como una fracción entre 0 y 1, o un porcentaje entre 0 y 100.

La suma de PV y FN corresponde al total de positivos en el proceso de detección así la sensibilidad de un sistema de detección se puede calcular como:

$$\text{Sensibilidad} = \text{PV} / (\text{PV} + \text{FN})$$

Una sensibilidad de 1 indica que todas las sustancias, compuestos o características son detectados. La sensibilidad también se denomina Fracción de positivos Verdaderos.

La métrica que complementa a la sensibilidad es la especificidad la cual mide la probabilidad de que un proceso de detección reporte correctamente la no existencia de una sustancia, compuesto o característica cuando efectivamente no existe. La suma de NV y PF corresponde al total de falsos en el proceso de detección así la especificidad en un sistema de detección se puede calcular como:

$$\text{Especificidad} = \text{NV} / (\text{NV} + \text{PF})$$

Una especificidad de 1 indica que nunca se reporta la existencia de una sustancia, compuesto o característica cuando este no existe. La Fracción de Falsos Positivos esta definida como (1-especificidad) y es la fracción de muestras que se reportan equivocadamente.

Para evaluar un proceso de detección es necesario tener los valores de sus dos métricas: especificidad y sensibilidad, ya que una sola métrica no puede evaluar correctamente el proceso. Esto debido a que se puede forzar una sensibilidad de 1 si nuestro sistema de detección reporta todos los casos como positivos (a lo que corresponde una especificidad de 0) y también se puede forzar una especificidad de 1 si nuestro sistema reporta todos los casos como negativos (a esto corresponde una sensibilidad de 0).

El análisis de la Curva Característica de Operación del Receptor (ROC) (del inglés Receiver Operating Characteristic) desarrollada en el contexto de la detección de señales electrónicas en los inicios de la década de los 50, encontró una importante aplicación en el proceso de toma de decisión medicas. La curva ROC es la herramienta estándar para graficar todas las posibles combinaciones de sensibilidad y especificidad de un proceso de detección. Usualmente se gráfica la Fracción de Positivos Verdaderos (TPF) o sensibilidad en el eje Y, vs la Fracción de Falsos Positivos (FPF) o (1-especificidad) en el eje X. En una curva ROC el punto de operación ideal es la esquina superior izquierda donde $TPF=1$, $FPF=0$

Como se muestra en la Figura 3, el procedimiento de identificación permite la completa separación y por lo tanto su identificación para muestras con matrices muy similares. Las curvas ROC de estas muestras muestran que el procedimiento de identificación, funciona perfectamente, y a todos los efectos prácticos puede considerarse que se resuelven con un 100% de sensibilidad y un 100% especificidad.

El interés del análisis va desde la simple identificación de una muestra hasta determinar si posee o no una determinada característica, sensorial, química, física o biológica. Incluyendo la posibilidad de determinar su trazabilidad u origen de la muestra (alimentos o muestras en general).

DESCRIPCIÓN DE LAS FIGURAS

La figura 1 muestra una vista esquematizada del dispositivo comprendiendo tanto el equipo láser (1), los espejos (2), lentes (3), muestra (4), fibra óptica (5), monocromador (6), controlador de retrasos de pulso (7), osciloscopio (8) y ordenador personal (9).

La figura 2 muestra los espectros LIBS de las muestra de aceite clasificadas como oliva1 (izquierda) y oliva2 (derecha).

30

La figura 3 muestra los resultados de la red, los resultados correspondiente a la muestra de oliva1 tienen un valor de 1 (uno) mientras

que los resultados correspondientes a la muestra oliva2 tienen un valor de 0 (cero).

La figura 4 muestra la curva de repuesta ROC de las muestras de oliva. Observar que la línea del modelo se encuentra a la izquierda y arriba.

5 La figura 5 muestra los espectros LIBS de las muestras de derivados de tomate clasificadas como ketchup (izquierda) y tomate-frito (derecha).

La figura 6 muestra los resultados de la red, los resultados correspondiente a la muestra de ketchup tienen un valor de 1 (uno) mientras que los resultados correspondientes a la muestra de tomate-frito tienen un
10 valor de 0 (cero).

La figura 7 muestra la curva de repuesta ROC de las muestras de derivados de tomate. Observar que la línea del modelo se encuentra a la izquierda y arriba.

La figura 8 muestra los espectros LIBS de las muestras de ketchup
15 clasificadas como ketchup1 (izquierda) y ketchup2 (derecha).

La figura 9 muestra los resultados de la red, los resultados correspondientes a la muestra de ketchup1 tienen un valor de 1 (uno) mientras que los resultados correspondientes a la muestra de ketchup2 tienen un valor de 0 (cero).

20 La figura 10 muestra la curva de repuesta ROC de las muestras de ketchup. Observar que la línea del modelo se encuentra a la izquierda y arriba.

La figura 11 muestra los espectros LIBS de las muestras de leche infantil de composición similar, clasificadas como leche1 (izquierda) y leche2
25 (derecha).

La figura 12 muestra los resultados de la red, los resultados correspondientes a la muestra de leche1 tienen un valor de 1 (uno) mientras que los resultados correspondientes a la muestra de leche2 tienen un valor de 0 (cero).

MODO DE REALIZACIÓN DE LA INVENCION

La presente invención se ilustra mediante los siguientes ejemplos que no son en ningún caso limitativo de su alcance, el cual viene definido
5 exclusivamente por la nota reivindicatoria.

Ejemplo 1: Identificación y clasificación de aceites de oliva.

Como ejemplo, a fin de identificar una muestra de aceite, se procede a
10 la obtención de espectros LIBS obtenidos de un único disparo láser, procedente de muestras conocidas o identificadas. Un total de 50 pulsos láser de cada muestra son almacenados en una base de datos como huellas digitales de las muestras. A fin de prevenir la radiación de cuerpo negro generada en los primeros momentos del plasma se utilizó como tiempo óptimo,
15 un retraso de 10 μ s entre el pulso láser y la obtención del espectro en la evolución del plasma. Los espectros de dos muestras de aceite de oliva, por lo tanto con un tipo de matriz muy similar y con una suave diferencia de sabor, fueron clasificadas como muestra de oliva1 y muestra de oliva2, y se muestran en la figura 2. La clasificación realizada por la red NN se muestra
20 en la figura 3. Siendo los primeros 50 puntos correspondientes a la muestra de oliva1 y los restantes a la muestra de oliva2, puede observarse que estas muestras son totalmente separables. La evaluación del procedimiento, curva ROC (figura 4) muestra que el 100% de los espectros han sido clasificados correctamente, aproximando la evaluación realizada a un ideal 100 % de
25 sensibilidad y 100 % especificidad.

Ejemplo 2: Identificación y clasificación de derivados de tomate.

Como ejemplo, a fin de identificar diferentes derivados de tomate, este
30 ejemplo es presentado como un ejemplo de la amplia variedad de elementos y sustancias que pueden ser analizados en soluciones o suspensiones en diversos tipos de matrices, como en el caso del zumo de tomate y el cual

puede ser extendido a otros tipos de zumos de frutas e incluso a sustancias líquidas como la leche y biológicos como la sangre, extendiendo su aplicación a muestras sólidas minerales, metales, etc. El procedimiento utilizado es similar al ya descrito y donde una muestra del zumo de tomate fue colocada en el portamuestra, posteriormente el haz del láser fue focalizado sobre la superficie de la muestras para la obtención de los espectros LIBS obtenidos de un único disparo láser. Se utilizó como tiempo óptimo, un retraso de 10 μ s entre el pulso láser y la obtención del espectro en la evolución del plasma. Los espectros de las muestras de tomate ketchup y tomate frito, fueron clasificadas como ketchup y Tfrito, estos espectros se muestran en la figura 5. La clasificación realizada por la red NN se muestra en la figura 6. Siendo los primeros 50 puntos correspondientes a la muestra de ketchup y los restantes a la muestra de Tfrito, puede observarse que estas muestras son totalmente separables. La evaluación del procedimiento, curva ROC (figura 7) muestra que el 100% de los espectros han sido clasificados correctamente, aproximando la evaluación realizada a un ideal 100 % de sensibilidad y 100 % especificidad.

Ejemplo 3: Identificación y clasificación de ketchup comercial.

Como ejemplo, a fin de identificar diferentes tipos de ketchup de origen comercial, es presentada de capacidad del método utilizado para realizar una correcta identificación y clasificación de dos muestras idénticas, cuya única diferencia es la introducción de una cantidad conocida de caroteno con una concentración inferior a 100 ppm, siendo de esta forma la matriz de ambas muestras exactamente iguales. El procedimiento utilizado es similar al ya descrito y donde las muestra son colocadas en el portamuestra, posteriormente el haz del láser fue focalizado sobre la superficie de la muestras para la obtención de los espectros LIBS obtenidos de un único disparo láser. Se utilizó como tiempo óptimo, un retraso de 10 μ s entre el pulso láser y la obtención del espectro en la evolución del plasma. Los espectros de las muestras de tomate ketchup, fueron clasificados como

ketchup1 y ketchup2, estos espectros se muestran en la figura 5. La clasificación realizada por la red NN se muestra en la figura 6. Siendo los primeros 50 puntos correspondientes a la muestra de ketchup1 y los restantes a la muestra de ketchup2, puede observarse que estas muestras son totalmente separables. La evaluación del procedimiento, curva ROC (figura 10) muestra que el 100% de los espectros han sido clasificados correctamente, aproximando la evaluación realizada a un ideal 100 % de sensibilidad y 100 % especificidad.

10 **Ejemplo 4: Identificación y clasificación de productos lácteos (leches infantiles comerciales).**

Como ejemplo, a fin de identificar una muestra de leche, este ejemplo es presentado como un ejemplo de la amplia variedad de elementos y sustancias que pueden ser analizados utilizando estos métodos de análisis. El procedimiento utilizado es similar al ya descrito y donde las muestra son colocadas en el portamuestra, posteriormente el haz del láser fue focalizado sobre la superficie de la muestras para la obtención de los espectros LIBS obtenidos de un único disparo láser. Se utilizó como tiempo óptimo, un retraso de 10 μ s entre el pulso láser y la obtención del espectro en la evolución del plasma. Los espectros de las muestras de productos lácteos, fueron clasificados como leche1 y leche2, estos espectros se muestran en la figura 11. La clasificación realizada por la red NN se muestra en la figura 12. Siendo los primeros 50 puntos correspondientes a la muestra de leche1 y los restantes a la muestra de leche2, puede observarse que estas muestras son totalmente separables. La evaluación del procedimiento, curva ROC (figura 13) muestra que el 100% de los espectros han sido clasificados correctamente, aproximando la evaluación realizada a un ideal 100 % de sensibilidad y 100 % especificidad.

REIVINDICACIONES

1. Método de análisis para identificación y caracterización instantánea de muestras caracterizado porque comprende: (a) irradiar una muestra
5 empleando un láser focalizado sobre la superficie de una muestra, (b) detectar la radiación producida por los elementos químicos del plasma que se forma en la superficie sobre la que se focaliza el haz láser, obteniendo un espectro único de la muestra, utilizando un analizador óptico (c) comparar la señal del detector con una base espectral dinámica
10 utilizando un sistema de redes neuronales artificiales que incluyen algoritmos matemáticos supervisados.
2. Método de análisis para identificación y caracterización instantánea de muestras, según reivindicación 1, donde el láser es del tipo Nad:YAG
15 trabajando a una frecuencia de 1 a 20 Hz, a longitudes de onda entre 266 y 1024 nm, con una duración de pulso entre 4 y 5 nanosegundos.
3. Método de análisis para identificación y caracterización instantánea de muestras, según reivindicación 1, donde la detección de la radiación
20 producida por los elementos químicos del plasma se realiza a través de un sensor óptico CCD que proporciona 2024 puntos espectrales en un rango de 200 a 1100 nm.
4. Método de análisis para identificación y caracterización instantánea de muestras, según reivindicación 1, donde la red neuronal posee conexiones
25 hacia delante basadas en un modelo de propagación *perceptron*, tres capas (entrada, salida y oculta), una función de transferencia entre capas que utiliza un algoritmo tangente hiperbólico sigmoideal y su matriz de pesos se optimiza mediante el algoritmo de entrenamiento de gradiente escalado conjugado SGC previniendo el sobreajuste y el sobreentrenamiento.
5. Dispositivo para la identificación y caracterización instantánea de muestras
30 que comprende: (a) un equipo láser capaz de producir un proceso de vaporización y posterior formación de un plasma en la superficie de una muestra, (b) un analizador óptico que detecta la radiación producida por

- los elementos químicos del plasma, (c) un ordenador personal con interfaz gráfica que permite al usuario el control del láser, la observación del espectro de la muestra obtenido y la representación de los resultados para su análisis y que incluye un base de datos espectral dinámica que utiliza un sistema de redes neuronales artificiales que incluyen algoritmos matemáticos supervisados.
- 5
6. Dispositivo para la identificación y caracterización instantánea de muestras, según reivindicación 5, donde el láser es del tipo Nad:YAG trabajando a una frecuencia de 1 a 20 Hz, a longitudes de onda entre 266 y 1024 nm, con una duración de pulso entre 4 y 5 nanosegundos.
- 10
7. Dispositivo para la identificación y caracterización instantánea de muestras, según reivindicación 5, donde la detección de la radiación producida por los elementos químicos del plasma se realiza a través de un sensor óptico CCD que proporciona 2024 puntos espectrales en un rango de 200 a 1100 nm.
- 15
8. Dispositivo para la identificación y caracterización instantánea de muestras, según reivindicación 5, donde la red neuronal posee conexiones hacia delante basadas en un modelo de propagación *perceptron*, tres capas (entrada, salida y oculta), una función de transferencia entre capas que utiliza un algoritmo tangente hiperbólico sigmoïdal y su matriz de pesos se optimiza mediante el algoritmo de entrenamiento de gradiente escalado conjugado SGC previniendo el sobreajuste y el sobreentrenamiento.
- 20
9. Uso del dispositivo reivindicado para la determinación cualitativa de compuestos químicos y su correlación con propiedades organolépticas o de calidad en muestras de aguas, alimentos o muestras biológicas, en cualquier estado de la materia (sólido, líquido o gas).
- 25

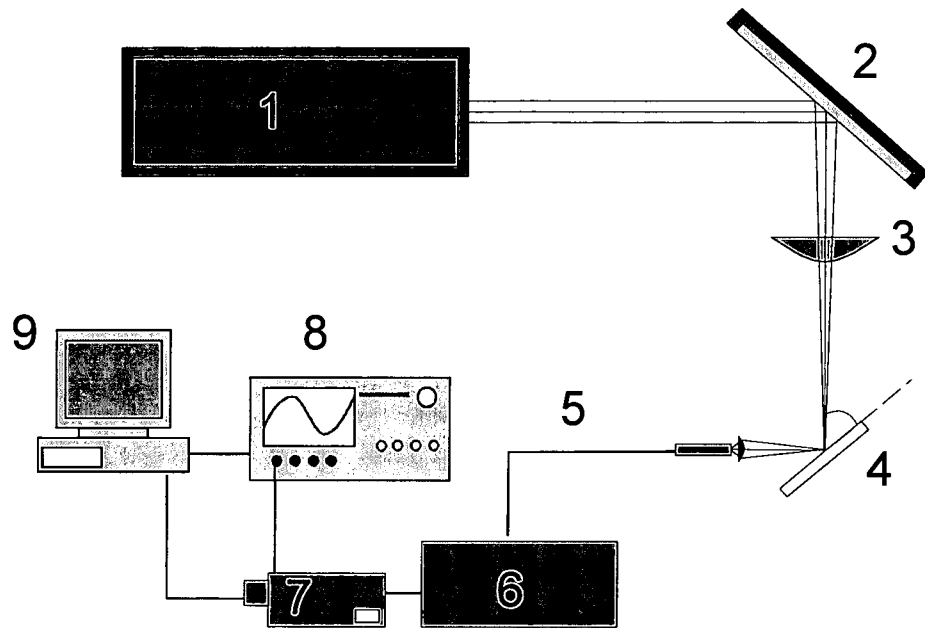
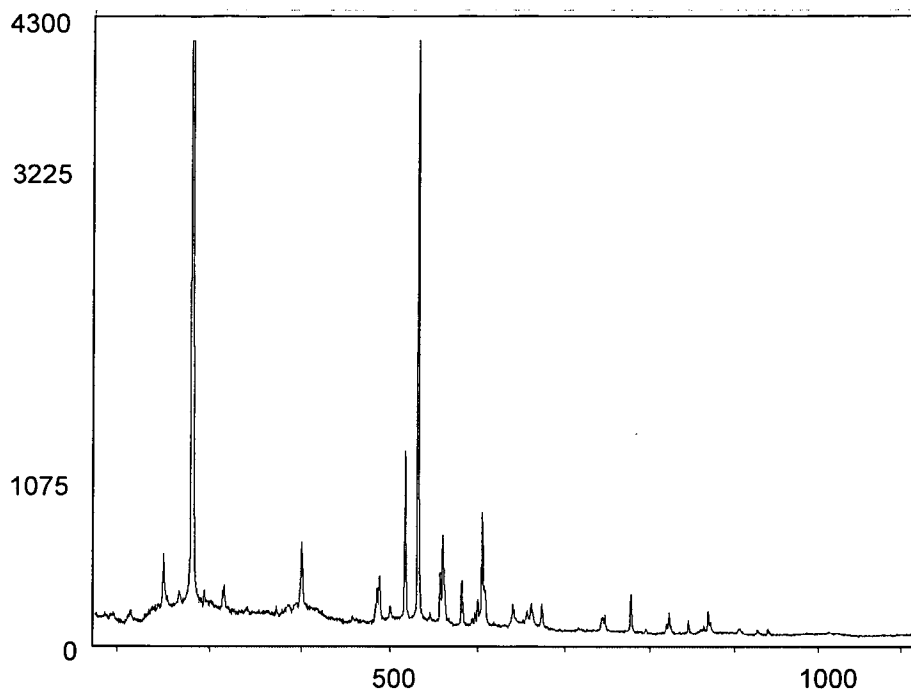
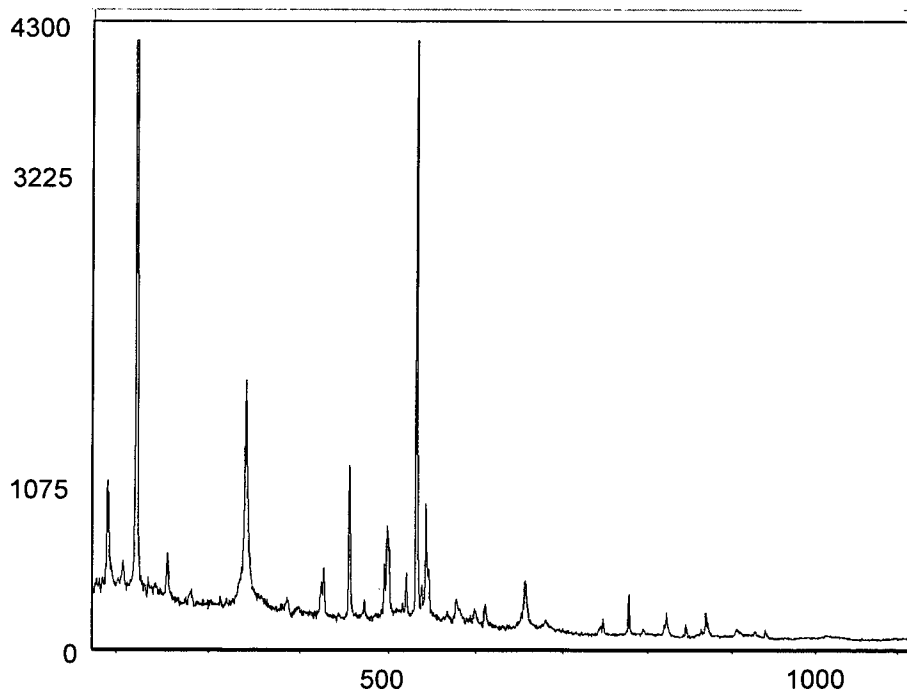


Figura 1



(a)



(b)

Figura 2

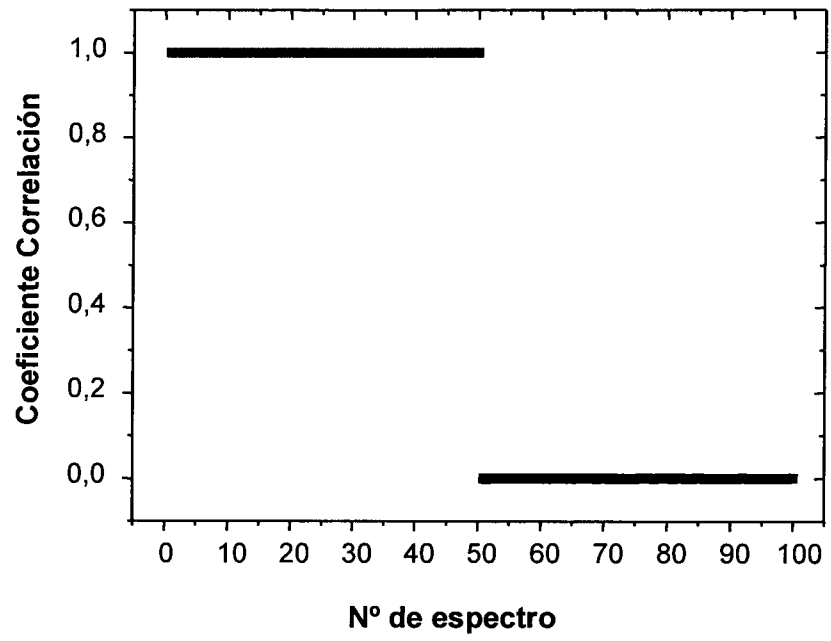


Figura 3

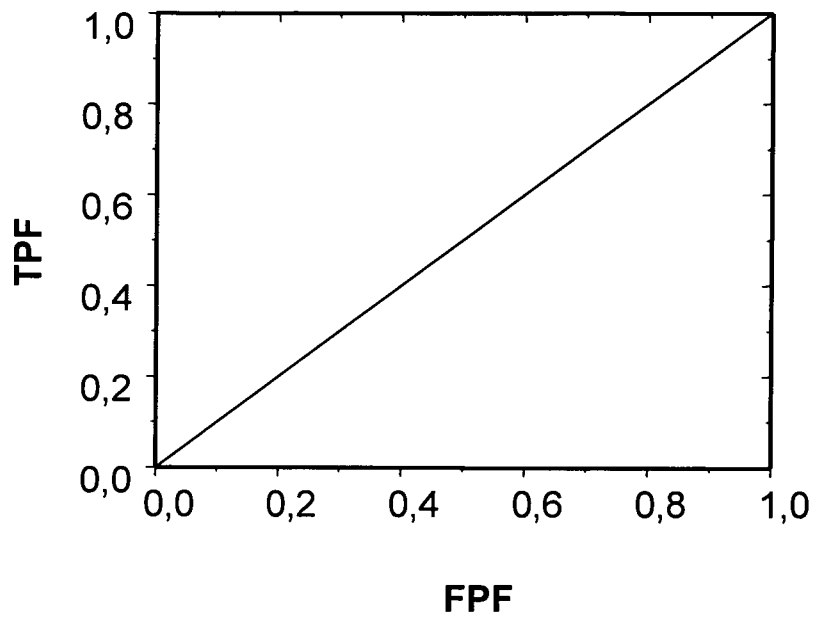
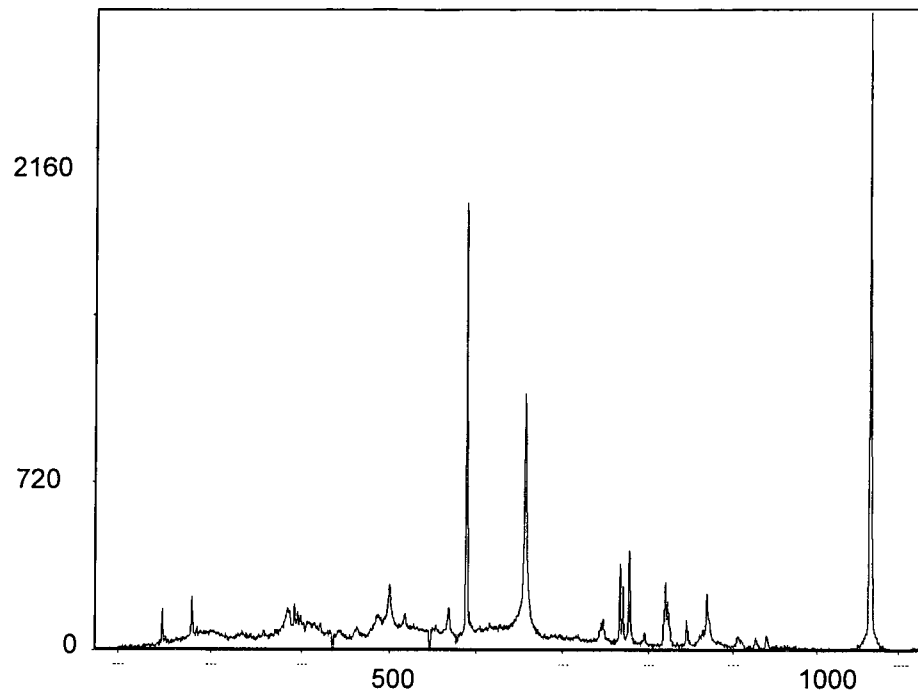
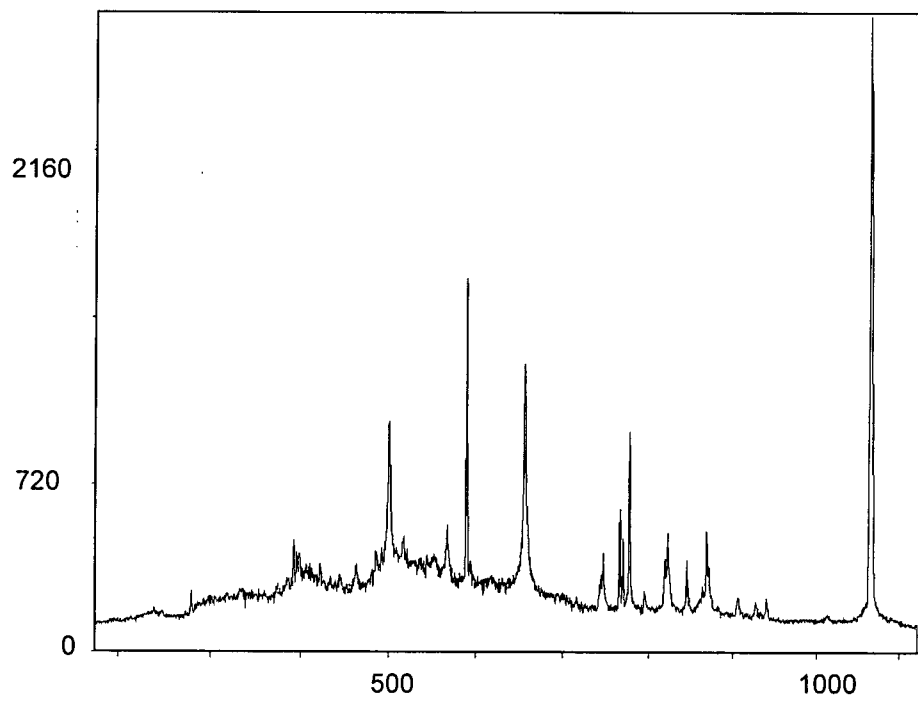


Figura 4



(a)



(b)

Figura 5

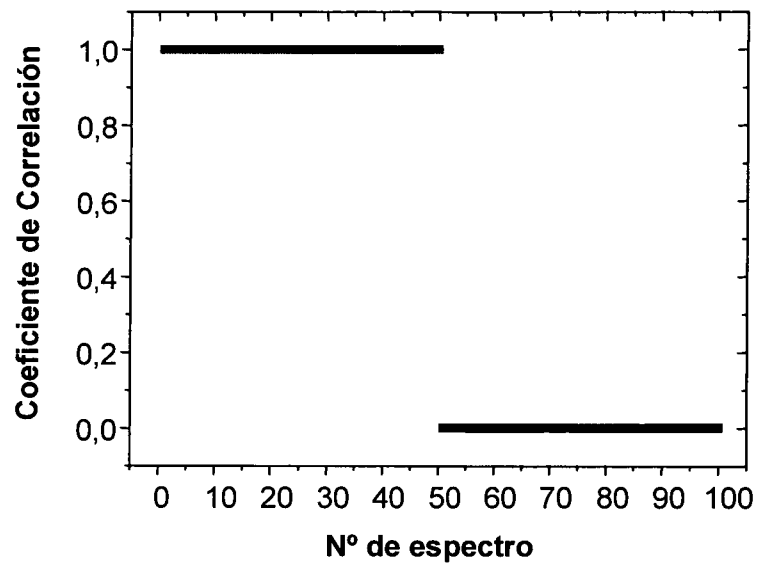


Figura 6

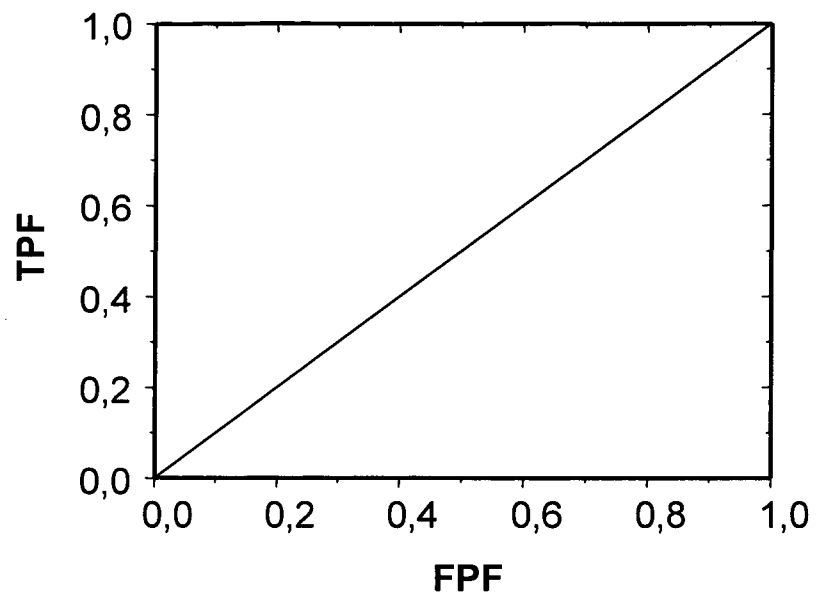
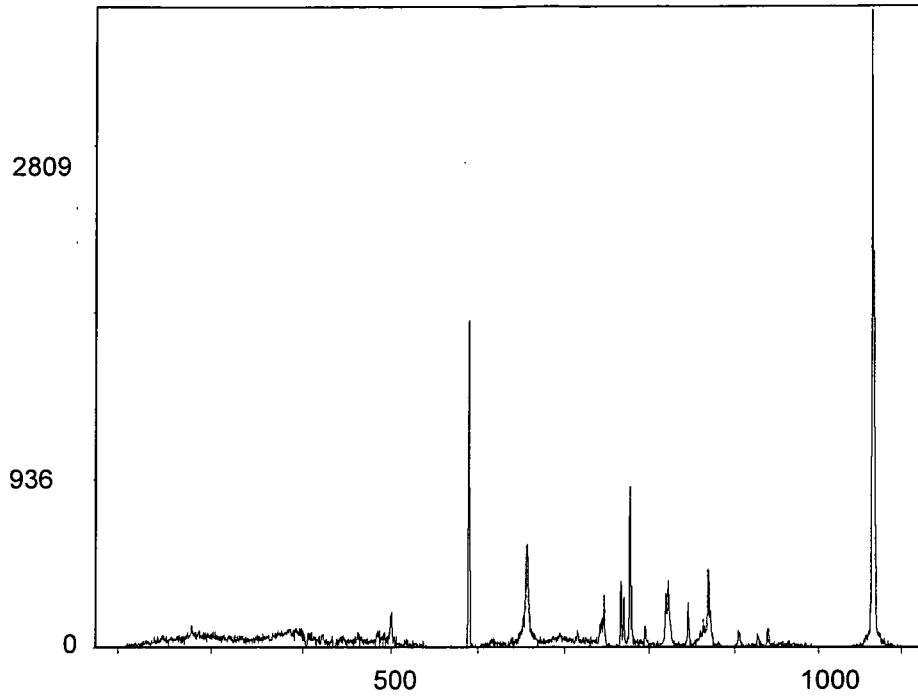
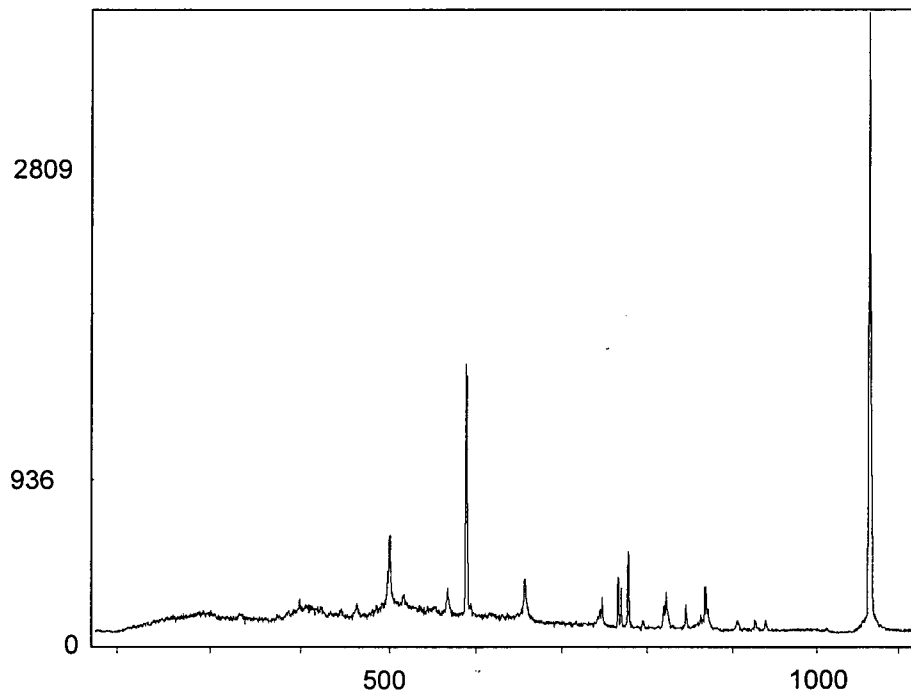


Figura 7



(a)



(b)

Figura 8

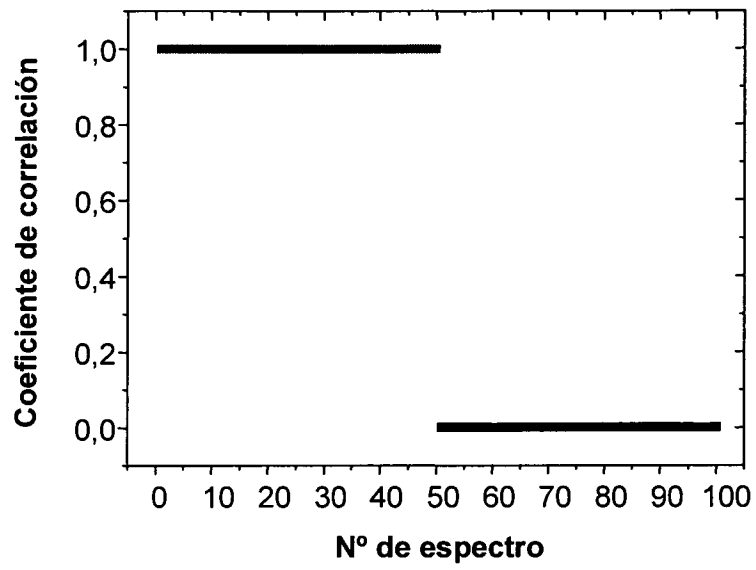


Figura 9

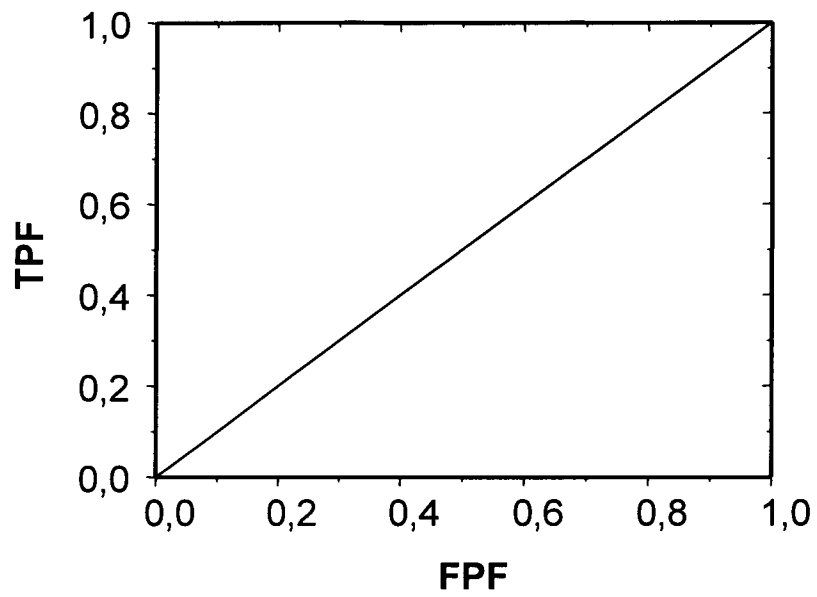
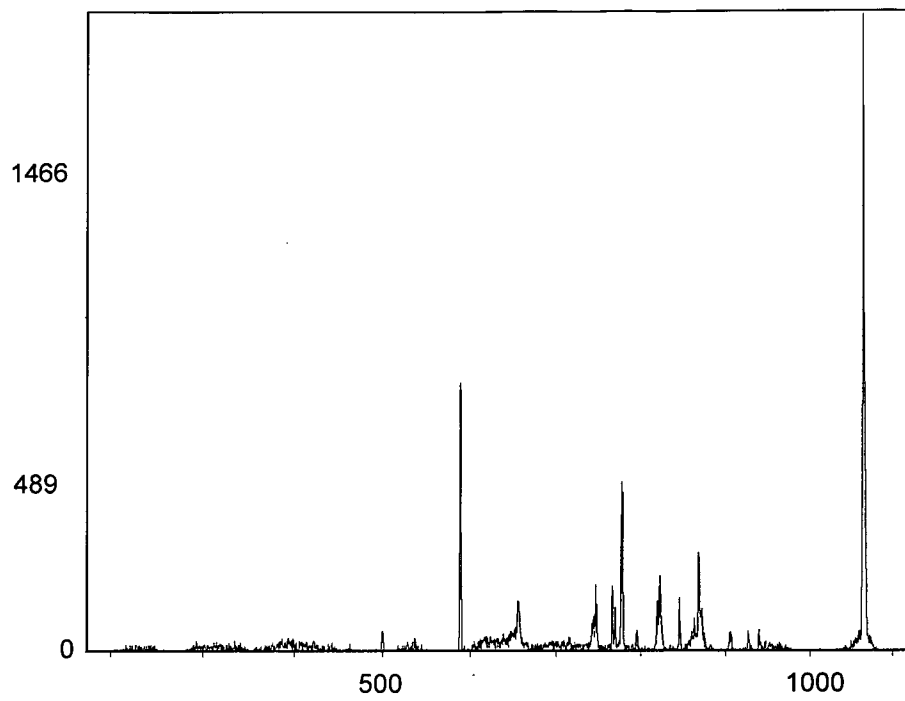
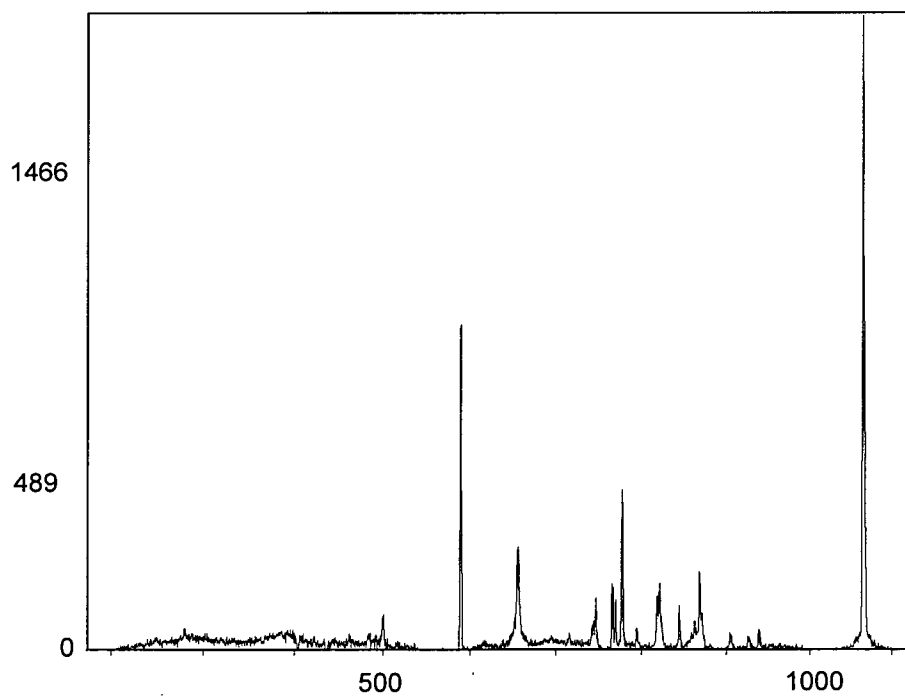


Figura 10

11/13



(a)



(b)

Figura 11

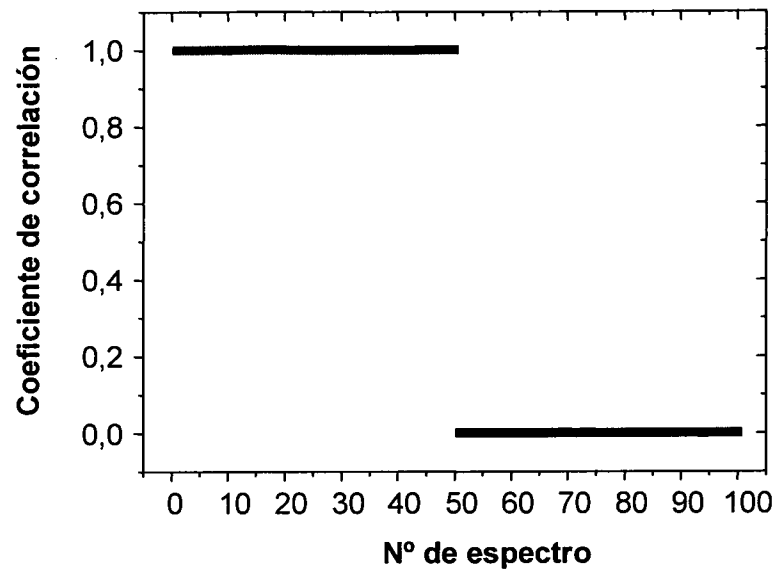


Figura 12

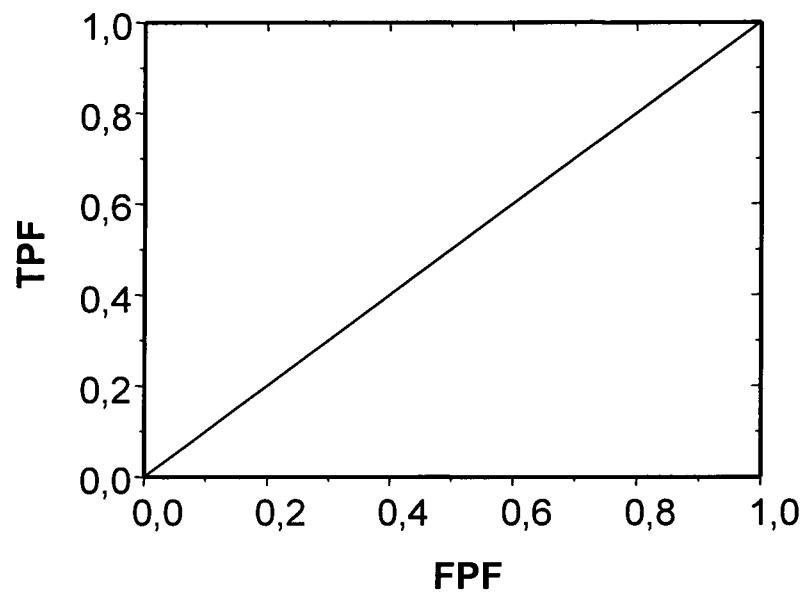


Figura 13