

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 498 796**

51 Int. Cl.:

G06K 9/00 (2006.01)

G06K 9/42 (2006.01)

G06K 9/46 (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Fecha de presentación y número de la solicitud europea: **12.09.2006 E 06803556 (7)**

97 Fecha y número de publicación de la concesión europea: **23.07.2014 EP 1934860**

54 Título: **Identificación y clasificación de partículas víricas en micrografías electrónicas texturizada**

30 Prioridad:

12.10.2005 US 725806 P

45 Fecha de publicación y mención en BOPI de la traducción de la patente:
25.09.2014

73 Titular/es:

**INTELLIGENT VIRUS IMAGING INC. (100.0%)
FAST LAW OFFICES 26 PINECREST PLAZA,
SUITE 2
SOUTHERN PINES NC 28387-4301, US**

72 Inventor/es:

**HOMMAN, MOHAMMED y
RYNER, MARTIN**

74 Agente/Representante:

VÁZQUEZ FERNÁNDEZ-VILLA, Concepción

ES 2 498 796 T3

Aviso: En el plazo de nueve meses a contar desde la fecha de publicación en el Boletín europeo de patentes, de la mención de concesión de la patente europea, cualquier persona podrá oponerse ante la Oficina Europea de Patentes a la patente concedida. La oposición deberá formularse por escrito y estar motivada; sólo se considerará como formulada una vez que se haya realizado el pago de la tasa de oposición (art. 99.1 del Convenio sobre concesión de Patentes Europeas).

DESCRIPCIÓN

Identificación y clasificación de partículas víricas en micrografías electrónicas texturizadas

Campo de la invención

5 La presente invención se refiere a la identificación de estructuras en imágenes. En particular, la presente invención proporciona un procedimiento y una disposición de identificación y clasificación de partículas víricas en micrografías electrónicas texturizadas.

Antecedentes de la invención

10 El ensamblaje de virus es un proceso intrincado y objeto de una investigación intensa. Los virus usan una célula huésped para producir su progenie de partículas víricas llevando a cabo un proceso complejo de maduración y transporte intracelular. Puede realizarse un seguimiento de este proceso con alta magnificación usando microscopia electrónica, que permite la identificación visual de diferentes tipos de partículas víricas en compartimentos celulares diferentes. Asuntos importantes que permanecen sin resolver incluyen la identidad de las proteínas víricas que están implicadas en cada etapa de este proceso de ensamblaje de virus, así como el mecanismo de la translocación intracelular subyacente y la localización de diferentes tipos de partículas víricas durante la maduración del virus. Los aspectos estructurales de la maduración del virus son generalmente complicados de abordar, aunque técnicas de visualización tales como tomografía y criomicroscopia electrónica (crio-ME) han contribuido enormemente a la vasta información de estructuras víricas. Estas técnicas proporcionan información sobre partículas víricas estables, a menudo maduras. Hay disponibles herramientas genéticas para producir mutantes de componentes proteicos víricos claves y pueden visualizarse los efectos estructurales mediante ME. No obstante, hay una carencia de herramientas apropiadas para caracterizar los efectos estructurales, especialmente formas de partícula intermedias y oscuras, y para cuantificarlos apropiadamente de un modo objetivo. Las herramientas de análisis por imagen para caracterizar y cuantificar partículas víricas, la maduración y el transporte intracelular facilitarían estudios objetivos de estados de ensamblaje de virus diferentes usando microscopia electrónica. Se adquiere mucha información, pero se necesita estructurarla y estadísticas producidas a partir de la misma para evaluar el efecto y extraer conclusiones

Sumario de la presente invención

30 La caracterización de la morfología estructural de partículas víricas en micrografías electrónicas es una labor complicada, pero deseable en relación con la investigación del proceso de maduración y la detección de cambios en la morfología de partículas víricas en respuesta al efecto de una mutación o de fármacos antivíricos que se estén administrando. Por lo tanto, se ha desarrollado un procedimiento para describir y clasificar formas de partículas víricas en micrografías electrónicas en base a la determinación de las características invariables de la proyección de una estructura vírica dada. La plantilla para la partícula vírica se crea en base a la información obtenida a partir de un conjunto de estudio pequeño de micrografías electrónicas y después se usa para clasificar y cuantificar estructuras similares de interés en un número ilimitado de micrografías electrónicas mediante un procedimiento de correlación. Usando análisis de deformación lineales, este algoritmo novedoso descrito en el presente documento puede manipular variaciones de partículas víricas tales como elipticidad y, por lo tanto, permite la evaluación de propiedades tales como el tamaño y la orientación de una partícula vírica. La aplicación práctica del procedimiento se demuestra mediante la capacidad para localizar tres clases diferentes de partículas víricas en micrografías electrónicas de transmisión de fibroblastos infectados con citomegalovirus humano.

45 En resumen, el procedimiento es para la identificación y caracterización de estructuras en micrografías electrónicas. Se seleccionan estructuras en una primera imagen. Las estructuras tienen un primer tipo de forma deformado en una primera dirección. Las estructuras seleccionadas se transforman en un segundo tipo de forma diferente a partir del primer tipo de forma. Las estructuras transformadas del segundo tipo de forma se usan para formar una pluralidad de plantillas. Se identifica una nueva estructura en una segunda imagen. La nueva estructura tiene el primer tipo de forma. La estructura de segundo tipo de forma de cada plantilla se deforma en la primera dirección. Se determina qué plantilla es una plantilla preferente que mejor corresponde a la nueva estructura.

Breve descripción de los dibujos

Las Fig.1A y 1B muestran imágenes de micrografía electrónica de transmisión típicas del desarrollo de herpesvirus;

La Fig 2A muestra nucleocápsides de herpesvirus vacíos;

La Fig. 2B muestra nucleocápsides del herpesvirus con un núcleo translúcido;

5 La Fig. 2C muestra nucleocápsides de herpesvirus que contienen AND empaquetado;

La Fig. 3A muestra una particular vírica con una forma elíptica;

La Fig. 3B muestra una partícula vírica que se ha deformado para hacerla circular.

10 La Fig. 4A–C muestra funciones de ensayo para estructuras de cápside víricas (A, B y C) en micrografías electrónicas que no usan reducción de coeficiente (ninguna) o con el 80 % de los coeficientes mostrando menos variación (VAR);

La Fig. 5A muestra una correspondencia de una función de ensayo A con una estructura de cápside auténtica y con una estructura similar pero falsa.

La Fig. 5B muestra una correspondencia de una función de ensayo B con una estructura de cápside auténtica y con una estructura similar pero falsa.

15 La Fig. 6 muestra correspondencia con la función de ensayo A dentro de una vesícula.

La Fig. 7A–C muestra relaciones positivas falsas (FPR) y negativas falsas (FNR) para las diferentes funciones de ensayo A, B y C, respectivamente.

La Fig. 8 muestra las funciones de probabilidad positivas (PPF) para las funciones de ensayo A, B y C.

20 La Fig. 9 muestra una comparación del número total real de estructuras víricas presentes en un conjunto de imágenes de ensayo (eje X) determinadas por el virólogo con un número identificado mediante nuestro procedimiento (eje Y); y

La Fig.10 muestra una producción automatizada de un mapa que identifica localizaciones de interés en una micrografía electrónica ilustrada a este respecto para la función de ensayo C.

Descripción detallada

25 El desarrollo de un sistema automatizado para ayudar en la identificación de partículas víricas en micrografías electrónicas se describe en el presente documento. Como modelo se han usado fibroblastos que están infectados con citomegalovirus humanos (HCMV), un virus de la clase de β -herpes. Debe entenderse que el herpesvirus se usa solo como ejemplo ilustrativo y la invención no está limitada al herpesvirus. Durante la infección con citomegalovirus humanos, se producen muchas formas intermedias diferentes de la partícula vírica. Durante el ensamblaje de los herpesvirus, la célula huésped se fuerza a hacer copias del material genético vírico y a producir cápsides, una cubierta de proteínas víricas que envuelve y protege el material genético. Las cápsides son estructuras esféricas que pueden variar con respecto al tamaño y a la simetría y pueden, cuando están maduras, estar envueltas por una membrana de dos capas. La maduración de las cápsides víricas es un estadio importante en la producción de partículas víricas y uno que se estudia frecuentemente. No obstante, su apariencia en micrografías electrónicas varía considerablemente, lo que hace que el análisis sea un reto. Una característica única del herpesvirus es el tegumento, una capa de proteínas víricas que rodea la cápside antes de la envoltura final. La envoltura se adquiere mediante la aparición de cápsides con tegumento en vesículas secretoras en el citoplasma. Después de ello, las partículas víricas infecciosas salen de la célula huésped por fusión de estos virus que contienen vesículas con la membrana plasmática.

35 Se ha desarrollado un procedimiento objetivo para la clasificación y cuantificación de partículas víricas en dichas micrografías electrónicas de transmisión. En el análisis relacionado de imágenes microscópicas crio–electrónicas (crio–ME), se han realizado considerablemente más esfuerzos de exploración de diferentes procedimientos de identificación. En crio–micrografías, la correlación cruzada usando múltiples plantillas y los procedimientos para la detección de bordes se han aplicado con éxito.

45 Enfoques adecuados que permiten la caracterización y la cuantificación de la maduración de partículas víricas y su translocación intracelular facilitan estudios objetivos de estos fenómenos usando microcopia

5 electrónica. No obstante, las imágenes de microscopía electrónica son difíciles de analizar y describir de un modo objetivo, debido a su fondo fuertemente texturizado. Además, las partículas víricas individuales muestran una amplia diversidad de formas, dependiendo de su proyección en la micrografía electrónica, el procedimiento usado para preparar muestras para microscopía electrónica y los ajustes usados para la fotografía. Imágenes micrográficas electrónicas típicas 100, 102 que proporcionan una información valiosa se muestran en las Fig. 1A y 1B, respectivamente.

10 En la presente invención, se ha aplicado un enfoque al análisis de cápsides de HCMV en el núcleo de células infectadas que están en estados definidos de maduración tales como cápsides vacíos 104 (denominados A), cápsides con un núcleo translúcido 106 (denominados B) y cápsides que contienen ADN empaquetado 108 (denominados C), como se muestra mejor en las Fig. 2A–C.

15 El procedimiento y la disposición según la presente invención se ilustran con partículas víricas. Este debería observarse como un ejemplo no limitante. Otro tipo de partículas, incluidos, por ejemplo, objetos biológicos tales como células o estructuras celulares, pero también partículas y estructuras no orgánicas, pueden identificarse y caracterizarse con modificaciones secundarias del procedimiento y la disposición descritos.

El procedimiento según la invención incluye una etapa de adquisición de imágenes. La micrografía electrónica puede suministrarse por el microscopio electrónico como archivos o fotos que se van a analizar. Para etapas posteriores del procedimiento es preferente lograr y almacenar conocimientos de tamaño de píxeles, resolución y ampliación para cada micrografía.

20 En una etapa de preprocesamiento, las partículas relevantes se seleccionan y se transforman a partir de posibles apariencias deformadas en círculos.

En una etapa de formación de plantillas, se usan partículas seleccionadas y transformadas para formar una plantilla, que puede caracterizarse mediante una función de ensayo.

25 En una etapa de correspondencia, la plantilla o la función de ensayo se usan para identificar partículas en una imagen adicional o en imágenes adicionales. Las etapas del procedimiento se describirán adicionalmente y se ejemplificarán a continuación.

30 Un aparato de identificación y clasificación según la presente invención puede estar basado en un ordenador personal común con un poder de cálculo suficiente. El aparato de identificación y clasificación está provisto de una interfaz para recibir micrografías, medios de preprocesamiento para la transformación de imágenes deformadas, medios para formar las plantillas o extraer funciones de ensayo y medios para realizar un procedimiento de correspondencia. Estas etapas se llevan a cabo, típica y preferentemente, mediante módulos de codificación informáticos.

35 Se mantuvieron cultivos celulares tales como fibroblastos de pulmón embrionario humano (HF) en medio esencial mínimo exento de bicarbonato con sales de Hank (GIBCO BRL) complementado con HEPES [ácido 4-(2-hidroxietil)-1-piperazina-etanosulfónico] 25 mM, suero de carnero fetal inactivado con calor al 10 %, L-glutamina (2 mM), penicilina (100 U/ml) y estreptomycin (100 mg/ml) (GIBCO BRL, Grand Island, NY, Estados Unidos). Las células se cultivaron en matraces de cultivo tisular de 175 cm² (Coming, Nueva York, Estados Unidos) durante un máximo de 17 pasajes.

40 En una etapa de infección vírica, las células HF se infectaron con la cepa AD 169 de HCMV usando una multiplicidad de infección (MOI) de 1. Los virus que contenían sobrenadantes se recogieron 7 o 10 días después de la infección (dpi), se eliminaron los residuos celulares mediante centrifugación a baja velocidad y se congelaron a -70 °C hasta su uso para la inoculación.

45 Con el fin de examinar células infectadas con virus mediante microscopía electrónica, se cosecharon células no infectadas e infectadas con HCMV a 1, 3, 5 y 7 dpi y, a continuación, se fijaron en glutaraldehído al 2 % en tampón de cacodilato de sodio 0,1 M que contenía sacarosa al 0,1 M y CaCl₂ 3 mM, pH 7,4 a temperatura ambiente, durante 30 min. Las células se rasparon con una varilla de madera y se transfirieron a un tubo Eppendorf para continuar con la fijación durante la noche a 4 °C. Después de este procedimiento, las células se enjuagaron en tampón de cacodilato de sodio 0,15 M que contenía CaCl₂ 3 mM, pH 7,4, y se sedimentaron por centrifugación.

50 Estos sedimentos se fijaron después adicionalmente en tetróxido de osmio al 2 % disuelto en tampón de cacodilato de sodio 0,07 M que contenía CaCl₂ 1,5 mM, pH 7,4, a 4 °C durante 2 horas; se deshidrataron

secuencialmente en etanol y acetona; y se embebieron en LX-112 (Ladd, Burlington, VT, Estados Unidos). El contraste de las secciones se obtuvo mediante acetato de uranilo seguido por citrato de plomo y el examen se realizó en un microscopio electrónico de transmisión Philips 420 o Tecnai 10 (FEI Company, Oregon, Estados Unidos) a 80 kV.

5 La adquisición, discretización y análisis de imágenes se realizó a continuación. Las micrografías electrónicas de células HF infectadas con HCMV se digitalizaron usando una escala de grises de 8 bits a una resolución de 5,5 nm/pixel en un aparato HP Scanjet 3970. La implementación se realizó con Matlab 7.0.1 (The Mathworks Inc., Natick, MA, Estados Unidos) y el programa informático Sun Java 1.4.2 en un ordenador personal Dell Optiplex GX260. Este análisis implicó una interfaz gráfica de uso sencillo y la automatización de los parámetros descritos a continuación para un uso rápido y conveniente.

10 Las herramientas sencillas de usar y fiables para los estudios de ensablaje de virus intracelular se desarrollaron después. El enfoque se basó en el hallazgo de un conjunto compacto de puntos en R^2 , el campo de la micrografía, para cada uno de los cuales un punto tiene un valor de función correspondiente. Este conjunto de puntos y sus valores de función se denomina colectivamente una función de ensayo o plantilla y puede describirse mediante una secuencia $\{(x_k, c_k)\}_k$ en la que x es el punto y c es el valor de la función. La función de ensayo se produce preferentemente de modo que la secuencia del valor de función se correlacione con los valores de la escala de grises de los puntos correspondientes. En consecuencia, se requiere un conjunto definido de partículas víricas del mismo tipo para estudiar y diseñar la secuencia para proporcionar una plantilla para esta estructura de partículas específica. Esta dispersa representación facilita la deformación sencilla y ajustes de la plantilla a partículas víricas individuales cuya forma en la micrografía es más o menos elíptica.

15 En la etapa de preprocesamiento de deformación, las posiciones de las estructuras dentro del mismo tipo de partículas víricas varían en las diferentes imágenes. Por ejemplo, las partículas víricas se deforman a veces de tal modo que aparecen en diferentes formas elípticas. Para crear las funciones de ensayo, se usan espacios de vectores lineales que demandan que las posiciones espaciales de los vectores analizadas sean relativamente fijas. La transformación lineal uniforme se elige para aproximarse a las deformaciones, ya que abarca la mayor parte de las deformaciones prominentes observadas en micrografías. El coste informático de estos cálculos es bastante bajo y simplifica la gestión de límites. Este enfoque requiere a menudo el uso de un operador de transformación tetradimensional, es decir, una matriz de 2×2 . Estas variables implicadas pueden expresarse como la rotación de la estructura antes de la deformación (φ_R), la deformación radial primaria (\bar{r}), la tasa de la deformación que proporciona la estructura elíptica (d) y la rotación que sigue a la deformación (φ_D). En conjunto, estas forman la transformación que se muestra a continuación:

$$T = R_D D R_R = \begin{pmatrix} \cos\varphi_D & -\sin\varphi_D \\ \sin\varphi_D & \cos\varphi_D \end{pmatrix} \begin{pmatrix} \bar{r}/d & 0 \\ 0 & \bar{r}/d \end{pmatrix} \begin{pmatrix} \cos\varphi_R & -\sin\varphi_R \\ \sin\varphi_R & \cos\varphi_R \end{pmatrix} \quad (\text{ec.1})$$

20 Con el fin de identificar las variables de la transformación para una partícula vírica individual, se usó manualmente un conjunto de elipses para estimar la posición, el tamaño y la deformación de cada pared del cápside, como se muestra del mejor modo en las Fig. 3A y 3B. La imagen 110 (denominada A) tiene una forma elíptica, mientras que la imagen 112 (denominada B) se ha deformado tal como se describe para hacerla más circular. Así, se proporcionan tres (φ_D , F y d) de las cuatro variables. La muestra se transformó después parcialmente para obtener el radio principal medido sin deformación ($d=1$), tal como se ilustra en la Fig. 3B.

25 Las características que son independientes de la rotación tales como la arquitectura poligonal de la pared de la cápside y la posición del núcleo de ADN pueden determinarse mediante el valor de φ_R para cada muestra. Para hallar este valor, cada muestra parcialmente transformada puede normalizarse alrededor de su media en el interior de un círculo que abarca el área visualmente significativa de las imágenes 114a, 116a y 118a, tal como se muestra en la columna izquierda de las Fig. 4A-C. Después, la suma de los cuadrados de las distancias en el sentido L^2 para cada muestra puede minimizarse con respecto a los ángulos. Debido a que la minimización implica $N-1$ variables, siendo N el número de muestras de referencia considerando una muestra que se va a fijar, este procedimiento puede simplificarse minimizando las distancias a las muestras ya procesadas una por una. Todas las transformaciones de las imágenes pueden implementarse después de un modo bilineal, aproximando de este modo el valor de la función f en el punto (x, y) como:

$$f(x, y) = f(\underline{x}, \underline{y})(1 - x_m)(1 - y_m) + f(\bar{x}, \underline{y})x_m(1 - y_m) + f(\underline{x}, \bar{y})(1 - x_m)y_m + f(\bar{x}, \bar{y})x_my_m$$

en la que \underline{x} es el valor entero inferior más cercano de x , \bar{x} es el valor entero superior más cercano y $x_m = x - \underline{x}$. La integración puede realizarse usando la misma interpolación. Las mediciones obtenidas a partir de esta etapa de procesamiento proporcionan indicaciones del intervalo de las propiedades de deformación, es decir, los radios principales (radio principal) y tasa de deformación, pero estos parámetros deberían determinarse en base a experiencias adicionales. Debido a que se espera que todos los tipos de rotación y todas las direcciones de deformación de las estructuras víricas estén presentes en las micrografías electrónicas, estas variables están preferentemente no fijadas.

Los puntos y valores de función locales (parámetros) para las plantillas de partículas víricas pueden identificarse a continuación. Una vez las muestras deformadas están alineadas con la estructura parcial en las mismas posiciones, este enfoque puede usarse para hallar los valores de la función invariable. Para describir este procedimiento con más claridad, una muestra deformada f puede convertirse en un gráfico de esta función enumerando (listando individualmente) las posiciones de píxeles x y sus valores de función correspondientes c como $f = \{(x_k, c_k)\}_k$. El grado de correspondencia entre dos secuencias de valores de función y_i e y_j (denominados en adelante vectores) que contienen la misma secuencia de posiciones de píxeles se determinó usando la correlación estadística estimada estándar:

$$M(y_i, y_j) = \frac{\langle y_i - \bar{y}_i, y_j - \bar{y}_j \rangle}{\|y_i - \bar{y}_i\| \|y_j - \bar{y}_j\|} \quad (\text{ec. 2a})$$

En la que \bar{y} es el valor medio del vector y la correspondencia de todos los coeficientes a $[-1, 1]$ se cartografía. El fundamento lógico para usar este enfoque es que indica el grado de similitud lineal entre las dos estructuras. Después de disponer los vectores de muestra normalizados alrededor de su media

$\hat{y}_i = \frac{y_i - \bar{y}_i}{\|y_i - \bar{y}_i\|}$ en columnas en una matriz, se determina la secuencia de funciones de ensayo f_c ($\|f_c\|=1$) que hace $\|A^T f_c\|$ tan grande como sea posible, proporcionando de este modo la mejor correspondencia para las muestras usadas para el estudio.

La descomposición en valores singulares (SVD) puede describirse como sigue:

$$\|A^T f_c\| = \|V \Sigma U^T f_c\| = (\mathbf{V} \text{ es cuadrado y ortonormal}) = \|\Sigma U^T f_c\| = \|\Sigma w\|$$

se aplica a A en la que $\|w\|=1$ si $f_c \in \text{span}(U)$ lo que sería de esperar. Esta última expresión es máxima cuando w es el vector propio correspondiente al valor propio más grande de Σ (que es el valor singular más grande) y f_c debería ser, de este modo, la columna de U correspondiente. Debido a que la función es una combinación lineal de las columnas de A , la correspondencia (ec. 2a) se reduce a

$$M(f_c, y) = \frac{\langle f_c, y \rangle}{\|y - \bar{y}\|} \quad (\text{ec. 2b})$$

La función de ensayo de esta SVD inicial usa los coeficientes de todos los puntos asociados con el primer soporte asumido. Algunos de estos puntos están localizados en algún sitio en el exterior de las estructuras víricas en las imágenes y , además, existen puntos en las estructuras cuyos coeficientes pueden variar considerablemente. Así, con el fin de evaluar la importancia de cada coeficiente y eliminar, de este modo, lo peor de la varianza, el valor de

$$VAR_j = \sum_{n=1..N} (\hat{y}_n - \langle f_C, \hat{y}_n \rangle f_C)_j)^2$$

se calcula para cada coeficiente. A continuación, un determinado porcentaje de los puntos podría retenerse en la función de ensayo. Ya que estas operaciones cambian en base a la función de ensayo, se calculó subsiguientemente una nueva SVD.

- 5 Las Fig. 4A–C ilustran las funciones de ensayo obtenidas usando todos los coeficientes o solo aquellos con el 80 % de los coeficientes de variación identificados que muestran al menos varianza según la clasificación de varianza. Claramente, el tamaño del núcleo de ADN varía en la función de ensayo para el cápside C y, por lo tanto, los puntos más indeterminados se han eliminado en las imágenes a mano derecha 114b, 116b y 118b. En consecuencia, las funciones de ensayo obtenidas por reducción del número de coeficientes de este modo se usaron rutinariamente.

Las deformaciones pueden sintetizarse a continuación. Ya que se asume que las estructuras analizadas están todas orientadas en cualquier dirección y deformadas linealmente en cualquier dirección, estas características deben aplicarse automáticamente a la función de ensayo cuando se analiza una imagen. Esta información proporcionada por el comportamiento de la función de correspondencia cuando se deforma la función de ensayo es también de interés y se ha explotado en una situación similar. Manteniendo mientras la imagen B y la función de ensayo f_C fijas y variando la deformación T , el análisis de la función de correspondencia $g(T) = M(f_C, \{B(Tx_k)\}_k)$ (en la que la secuencia $\{x_k\}_k$ se obtiene a partir de la producción de las funciones de ensayo realizada. Con el fin de describir T en términos de los parámetros $(\varphi_R, \bar{r}, d, \varphi_D) \in ([0, 2\pi], [\bar{r}_0, \bar{r}_1], [d_0, d_1], [0, 2\pi]) = T_{unida}$, se realizan las suposiciones siguientes:

- 20 (i) Para determinada $T \in T_{unida}$, la función de ensayo deformada representa la estructura más similar al objeto en la imagen. Se asume que esta T es la que maximiza g .
- (ii) La T asociada con la deformación máxima debería localizarse en el interior del conjunto de deformación, y no en el límite. En estas condiciones, incluso si g está maximizada fuera del conjunto (es decir, la estructura es demasiado grande, demasiado pequeña o demasiado mal deformada), la correspondencia con los puntos límite más cercanos podría ser aún elevada.

Para considerarla identificada, una estructura debería cumplir estos criterios. La maximización de la función de correspondencia se realizó con el esquema descendiente más acusado inverso, usando la función de ensayo no deformada como punto de partida y aproximando la derivada como un esquema de ocho puntos de diferencias centrado (es decir, dos puntos para cada variable en la deformación).

- 30 La aplicación de estos criterios de correspondencia usados se representa en las Fig. 5 y 6. La Fig. 5A ilustra cómo funcionan estos criterios cuando se aplican a una cápside A auténtica, así como a una estructura similar pero falsa. En la imagen 120 (denominada A) se muestra una cápside auténtica. Cuando la función de ensayo se deforma, los gráficos ilustran cómo la función de correspondencia g varía con el tamaño radial (\bar{r}) y el grado de deformación (d) a partir del punto en el conjunto de deformaciones admisibles que maximiza g . La función de ensayo deformada tiene una apariencia similar a la de la muestra, y la deformación está dentro de los límites. La clasificación debería ser, por lo tanto, positiva. En la imagen 122 (denominada B) en la Fig. 5B, a diferencia de la imagen A, el punto del conjunto de deformación que maximiza (g) está situado en el límite y los gráficos muestran un valor de correspondencia superior fuera de este conjunto. Así, esta clasificación debería ser negativa. En este caso, los límites de deformación se ajustan a $(\varphi_R, \bar{r}, d, \varphi_D) \in ([0, 2\pi], [0,89, 1,1], [0,89, 1,13], [0, 2\pi])$ con fines ilustrativos.

- 45 Las cápsides víricas salen del núcleo brotando a través de la membrana de este orgánulo. De acuerdo con este procedimiento es difícil discriminar entre estructuras víricas y otras estructuras, tal como se muestra en las imágenes 126a y 126b en la Fig. 6. La estructura marcada con una cruz azul cumple los criterios de correspondencia (i) y (ii) mientras que las marcadas con un círculo rojo solo cumplen el criterio (i). En esta figura, una cruz azul indica un punto en la imagen en el que la correspondencia entre la función de ensayo y la correspondencia de estructura de la cápside es mejor que 0,8 y el grado de deformación es aceptable. Un círculo rojo indica un punto en el que esta correspondencia es mejor que 0,8, pero en el que el grado de deformación no es admisible. La estructura marcada como correspondencia tiene una correspondencia del 0,94, que es muy elevada.

Las estructuras de las partículas víricas en una imagen de microscopia electrónica pueden identificarse después. Con el fin de buscar estructuras en una imagen (B) similares a la función de ensayo f_C , ec. 2b se expande a convoluciones. La correspondencia de la función de ensayo en un punto (m), por lo tanto, puede expresarse como

$$M_{B,f_C}(m) = \sup_{T \in T_{Unida}} M(f_C, \{B(m + Tx_k)\}_k)$$

No obstante, este procedimiento precisa mucho tiempo. Puede acelerarse hacienda unas pocas observaciones y suposiciones:

(i) Las variantes deformadas de las funciones de ensayo no son ortogonales una a otro, y debido a que estas estructuras son esencialmente independientes de la rotación, la correspondencia de la función de ensayo no deformada es mejor que la de un valor determinado a cualquier estructura deformada admisible del mismo tipo.

(ii) Debido a que la traducción deforma una estructura adicionalmente, se asume la correspondencia con la función de ensayo no deformada para que sea superior a la posición real de la partícula vírica que en localizaciones al menos un diámetro de distancia de la función de ensayo desde esta posición.

Implementando estos criterios, se puede identificar un subconjunto de puntos potencialmente interesantes dentro de la imagen más grande. Después de ello, pueden realizarse análisis posteriores de este conjunto que usan la optimización descrita en la sección precedente. Este enfoque proporciona un conjunto final de puntos en la imagen que están asociados con valores de correspondencia de $P = \{M_j\}_j$. Con el fin de asegurar la inclusión de todas las posiciones de interés en una imagen, el valor umbral asociado con la suposición (i) anterior se ajustó a 0,5.

En el posprocesamiento del conjunto final, se hace un recuento de las partículas víricas. No hay un valor umbral (t) que pueda distinguir entre estructuras auténticas y falsas en todas las imágenes, es decir, la asignación de estructuras usando este procedimiento no está completamente de acuerdo con la realizada por parte de un virólogo experimentado. Ajustar un nivel umbral no es, por lo tanto, una opción.

En vez de ello, puede usarse una función de probabilidad positiva $PPF: [-1,1] \rightarrow [0,1]$ que puede usarse para determinar la probabilidad de que un punto dado asociado con un valor de correspondencia determinado esté realmente asociado con la partícula vírica. Esta extensión del valor predictivo positivo (PPV) se obtiene calculando la relación entre el número de estructuras identificadas correctamente y el número total de estructuras identificadas con un valor de correspondencia determinado. Así, para un conjunto (P) de estructuras identificadas mediante este procedimiento que contiene el subconjunto $P_{correcto}$ de puntos asociados con partículas víricas de un tipo dado,

$$PPF(M) = \frac{\#\{M_k \in P_{correcto}; M \leq M_k < M + \varepsilon\}}{\#\{M_k \in P; M \leq M_k < M + \varepsilon\}}.$$

Con el fin de obtener una función uniforme y monótonamente creciente se eligió 0,05 como valor para ε . La función de probabilidad que indica el número esperado (N) de estructuras en la imagen es

$$E(N) = \sum_{M \in P} PPF(M).$$

La exactitud de FNR / FPR del procedimiento puede describirse como sigue. Con el fin de organizar partículas víricas observadas en micrografías electrónicas según su estadio de maduración, se requiere un modelo tal como el descrito en el presente documento para representar cada estadio particular. Además, para que este modelo sea útil para la detección y cuantificación de partículas víricas en dichas imágenes también debe ser capaz de rechazar estructuras falsas. Así, un modelo ideal debería detectar todas las imágenes posibles de partículas víricas de diferentes tipos, pero nada más localizado en el mismo espacio, es decir, en el fondo. Con el fin de caracterizar nuestro modelo a este respecto, se usan comúnmente las relaciones negativas falsas (FNR) y positivas falsas (FPR). La FNR se define como la relación entre el número de partículas víricas auténticas rechazadas incorrectamente mediante el

procedimiento y el número real de partículas auténticas, mientras que FPR es la relación entre el número de estructuras falsas identificadas como auténticas y el número total de estructuras consideradas auténticas mediante este enfoque. Así, ambas relaciones se encuentran entre 0 y 1, siendo 0 el valor ideal.

- 5 Con el fin de determinar el número de partículas víricas en base a la información proporcionada por el conjunto de valores de correspondencia adquiridos buscando a lo largo de la imagen, puede usarse la función de probabilidad positiva *PPF* descrita anteriormente. El número esperado de partículas identificadas se comparó con el número verdadero de partículas presentes en la imagen para obtener una desviación media y típica del error del recuento. Además, para evaluar si existe una diferencia media sistemática, es decir, si el procedimiento identifica sobre la media demasiadas o demasiadas pocas partículas, la hipótesis H_0 que: "La diferencia media = 0" se analizó.

10 La estandarización y el análisis se llevaron a cabo en conjuntos separados de imágenes, 2 para el estudio y 12 para el análisis. El número de muestras usadas para la estandarización fue de 4, 7 y 10 para las funciones de ensayo A, B y C, respectivamente. Las imágenes de ensayo contenían un total de 53 cápsides A, 239 cápsides B y 83 cápsides C, y los límites de deformación se ajustaron a $(\varphi_R, \bar{r}, d, \varphi_D) \in ([0, 2\pi], [0,83, 1,2], [0,83, 1,2], [0, 2\pi])$.

15 Las relaciones de falso negativo (FNR) y falso positivo (FPR) pueden describirse como sigue. El procedimiento se evaluó comparando nuestros resultados con los de virólogos experimentados. Las FPR y FNR se calcularon como función del valor umbral para la medición de correspondencia, como mejor se muestra en los gráficos 128, 130 y 132 en las Fig. 7A–C, respectivamente. La FNR se define como la relación entre el número de estructuras auténticas rechazadas incorrectamente mediante el procedimiento usando un determinado valor umbral para la medición de correspondencia y el número real de partículas víricas presente tal como se determina por parte de un virólogo. De forma análoga, la FPR es la relación entre el número de estructuras falsas identificadas como auténticas y el número total de estructuras consideradas como auténticas mediante este procedimiento. Para la comparación con otros procedimientos, el cruce de las curvas tiene lugar a 0,25 para la función de ensayo A, 0,13 para la función de ensayo B y a 0,23 para la función de ensayo C.

20 La cuantificación de estructuras en micrografías electrónicas puede describirse tal como se muestra a continuación. Los valores de *PPF* 134 calculados a partir de los resultados presentados anteriormente se muestran en la Fig. 8. El gráfico representa la frecuencia relativa de partículas víricas identificadas correctamente mediante el procedimiento a un valor de correspondencia determinado. Para la comparación, un procedimiento ideal que proporcionara la separación completa entre estructuras verdaderas y falsas daría como resultado una función de etapa de Heaviside en algún valor umbral. Para la comparación, un caso ideal que proporcionara la separación completa entre estructuras verdaderas y falsas daría como resultado una función de etapa de Heaviside en algún valor umbral.

25 Un gráfico 136 del número total de partículas víricas identificado que se presenta en un conjunto de imágenes de ensayo mediante nuestro procedimiento en comparación con el número correcta tal como se determina por parte de un virólogo se muestra en la Fig. 9 junto con la función de identidad. La línea en este gráfico representa la función de identidad. La diferencia media es 0,16 y la desviación típica de 5,63. El nivel de importancia de la hipótesis nula H_0 , es decir, "La diferencia media = 0", es 0,92. Claramente, existe una similitud cercana entre estos dos valores (diferencia media = 0,16, desviación típica de 5,63), que en el caso ideal serán puntos de la función de identidad. El hecho de que el nivel de importancia de H_0 fue de 0,92 según el ensayo en *t* de Student indica que existe una probabilidad cierta de que no existan diferencias significativas entre estos dos enfoques de media. Estos resultados muestran que la selección rápida del número total de estructuras víricas en estadios diferentes de maduración en un conjunto grande de micrografías electrónicas, una labor que es, de otro modo, larga de realizar y tediosa para el experto, puede realizarse rápidamente y de forma fiable con nuestro procedimiento automatizado.

30 En base al conjunto de posiciones en una imagen 138 en la que estructuras de interés están localizadas puede producirse un mapa se muestra en la Fig. 10. Esto facilita el recuento manual de estas estructuras considerablemente y también proporciona un marco para el análisis manual. En vez de un recuento y comparación sencilla de estructuras en una imagen no procesada, el virólogo ve facilitada su labor de forma considerable disponiendo de dicho mapa. Las diversas estructuras se clasifican de izquierda a derecha para valores de correspondencia descendentes que comienzan en el lado izquierdo de la fila superior.

5 Cuando se investiga el procedimiento de ensamblaje de virus, la información relativa a la topología estructural con relación al estadio de maduración no está habitualmente disponible o está definido de forma vaga. Por lo tanto, se requieren herramientas para seleccionar y clasificar partículas víricas a diferentes estadios de maduración. Una vez se han obtenido unos pocos puntos de partida clasificando un conjunto de estructuras obvias, estos pueden usarse para expandir el conjunto de estructuras clasificadas identificando estructuras similares con la función de correspondencia usada. Este enfoque ayuda a hacer el mapeo de la maduración de virus en micrografías electrónicas de forma rápida, fiable y fácil de describir.

10 Aunque la presente invención se ha descrito según composiciones y realizaciones preferentes, se entiende que pueden realizarse a la misma determinadas sustituciones y alteraciones sin apartarse del ámbito de las reivindicaciones siguientes.

REIVINDICACIONES

1. Un procedimiento de identificación y caracterización de partículas víricas en imágenes de micrografía electrónica, que comprende:

seleccionar una o más estructuras elípticas en una primera imagen (110),

5 transformar cada estructura seleccionada para deformar cada estructura de la forma elíptica a una forma circular mediante una transformación;

10 usar las estructuras transformadas con forma circular para formar una plantilla y deformar dicha plantilla mediante una transformación correspondiente a dicha transformación para obtener una pluralidad de plantillas deformadas con forma elíptica; identificar una partícula vírica nueva en una segunda imagen, teniendo la partícula vírica nueva una estructura con forma elíptica usando la pluralidad de plantillas deformadas con forma elíptica, y

determinar qué plantilla es una plantilla preferente que corresponde mejor a la estructura con forma elíptica de la partícula de virus nueva;

15 en el que dicha transformación de cada estructura seleccionada para deformar la estructura desde la forma elíptica a una forma circular es una transformación lineal, usándose un operador de transformación tetradimensional que implica variables que definen la rotación antes de la transformación ($f_i d$), que alinea las diferentes estructuras elípticas, la deformación radial principal (r), la tasa de deformación (d) dando lugar a la forma elíptica de la partícula vírica, obteniéndose estas tres variables mediante deformación de la forma elíptica para hacerla más circular, y

20 la rotación después de la deformación ($f_i r$), que alinea características de las estructuras circulares, tales como la arquitectura poligonal de la pared de la cápside y la posición del núcleo de ADN.

2. El procedimiento según la reivindicación 1, procedimiento que además comprende analizar otras plantilla para verificar que la plantilla preferente proporciona la mejor correspondencia.

25 3. El procedimiento de la reivindicación 2, procedimiento que además comprende seleccionar estructuras basadas en parámetros relacionados con el tamaño y la forma elíptica y la elipticidad de las estructuras.

4. El procedimiento según la reivindicación 1, procedimiento que además comprende analizar otras direcciones diferentes a la primera dirección para verificar que la primera dirección de deformación de la plantilla óptima proporciona la mejor correspondencia de la nueva forma elíptica.

30 5. El procedimiento según la reivindicación 4, procedimiento que además comprende determinar que la correspondencia es menos exacta en direcciones diferentes a la primera dirección.

6. El procedimiento según la reivindicación 1, procedimiento que además comprende determinar un estadio de madurez y un tipo de estructura de la nueva forma elíptica en base a la plantilla preferente.

7. El procedimiento según la reivindicación 1, procedimiento que además comprende filtrar estructuras que están dispuestas dentro de una distancia radial de la nueva forma elíptica.

35

FIG. 1B

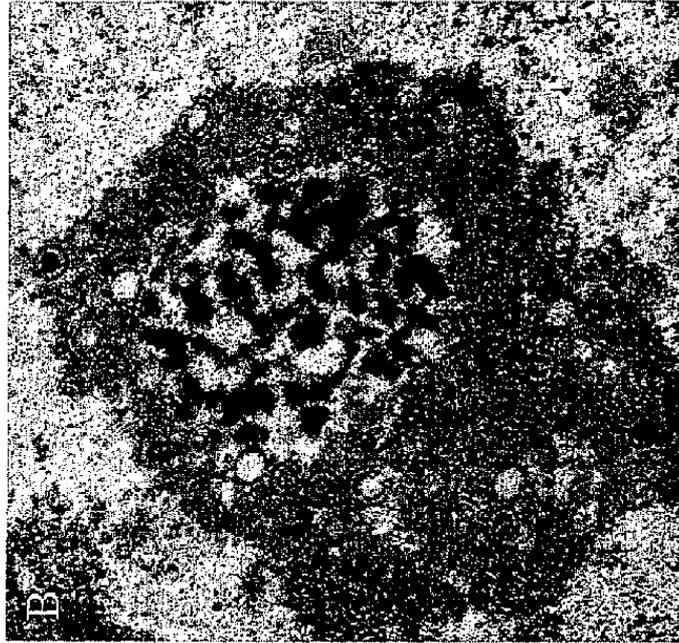


FIG. 1A

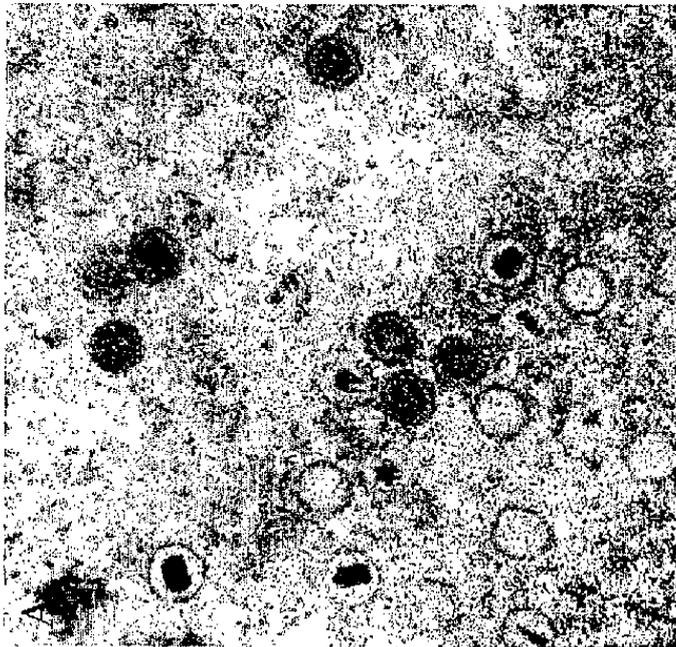


FIG. 2C

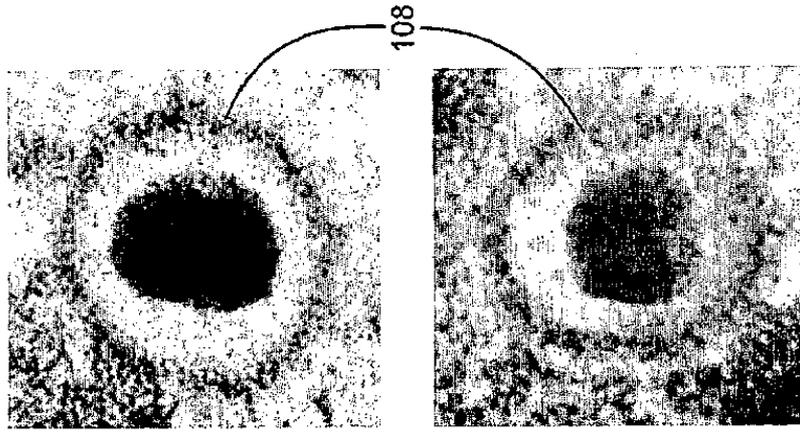


FIG. 2B

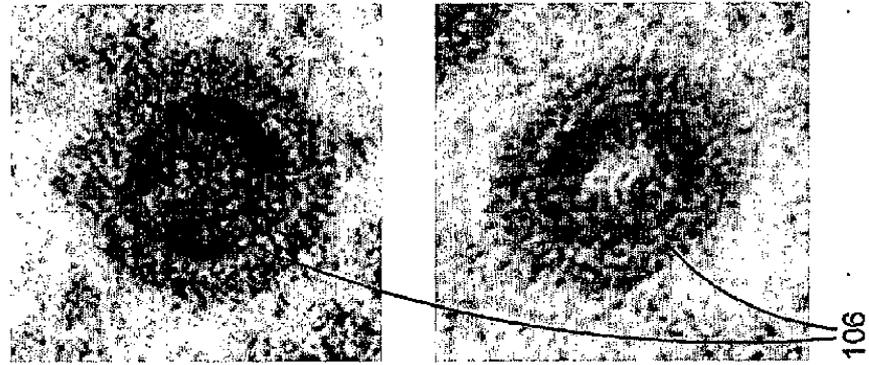


FIG. 2A

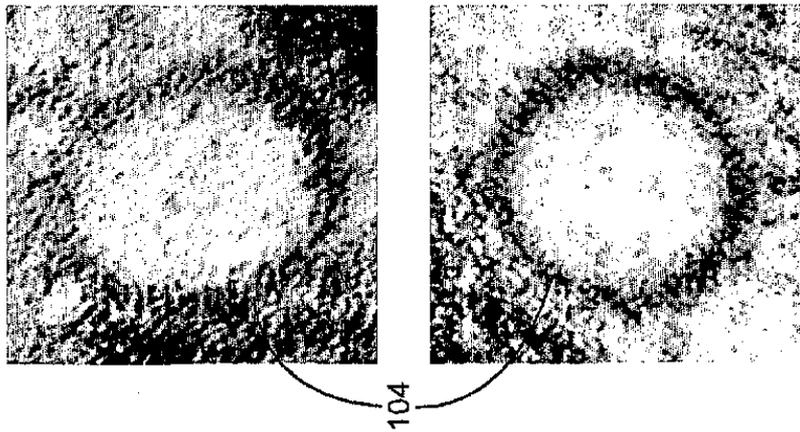


FIG. 3B

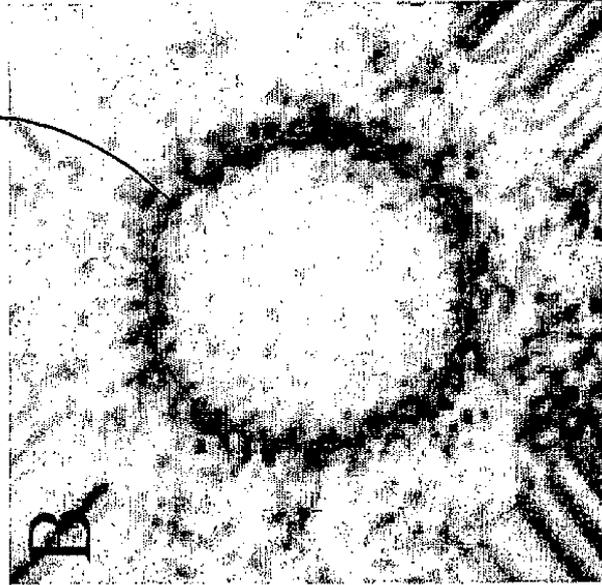
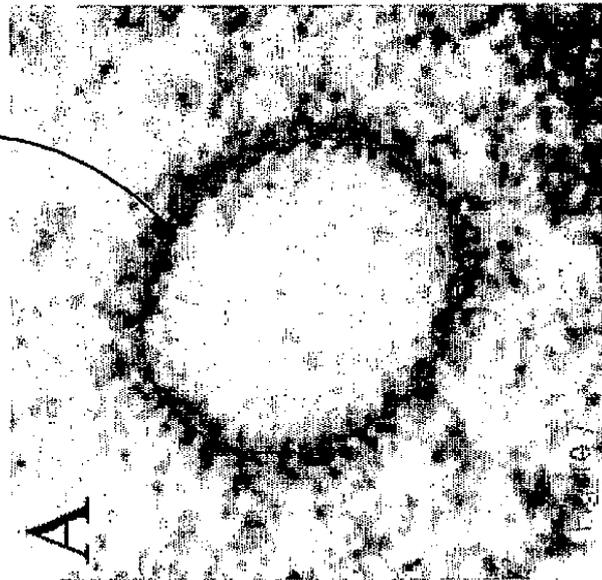
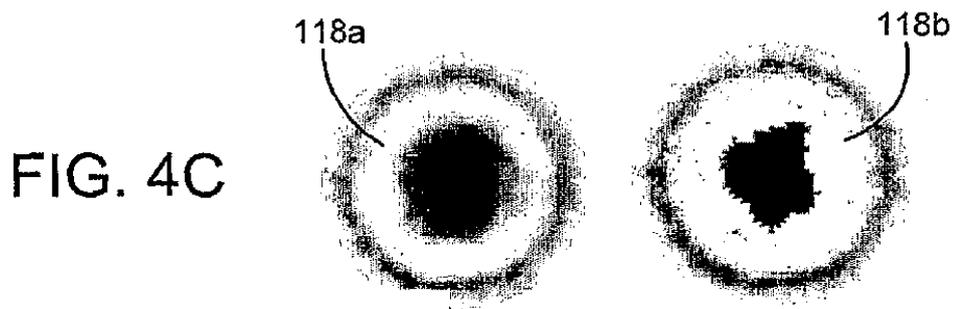
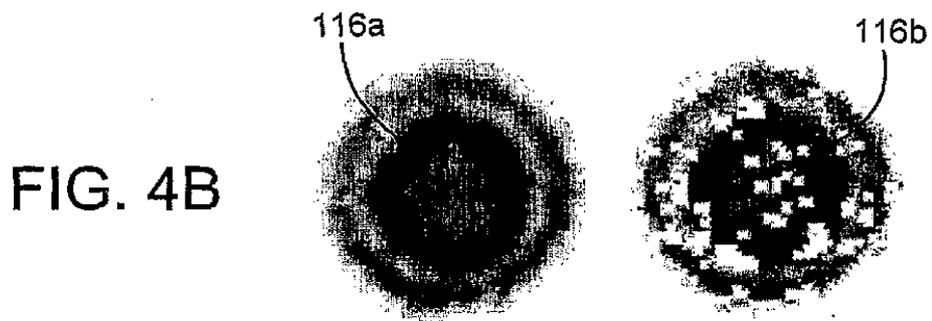
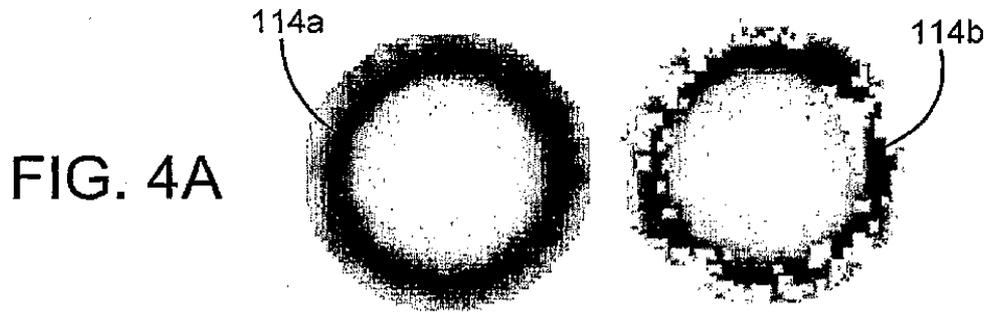


FIG. 3A





120

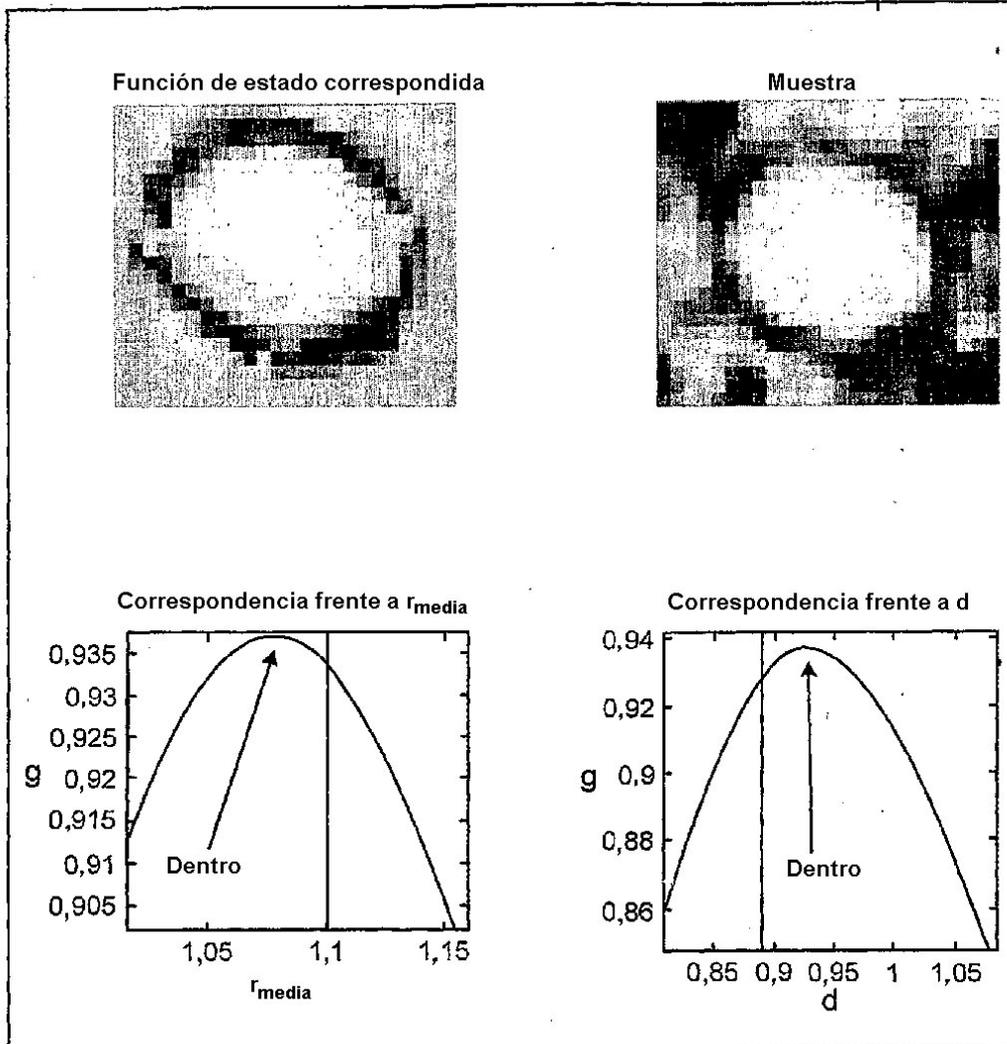


FIG. 5A

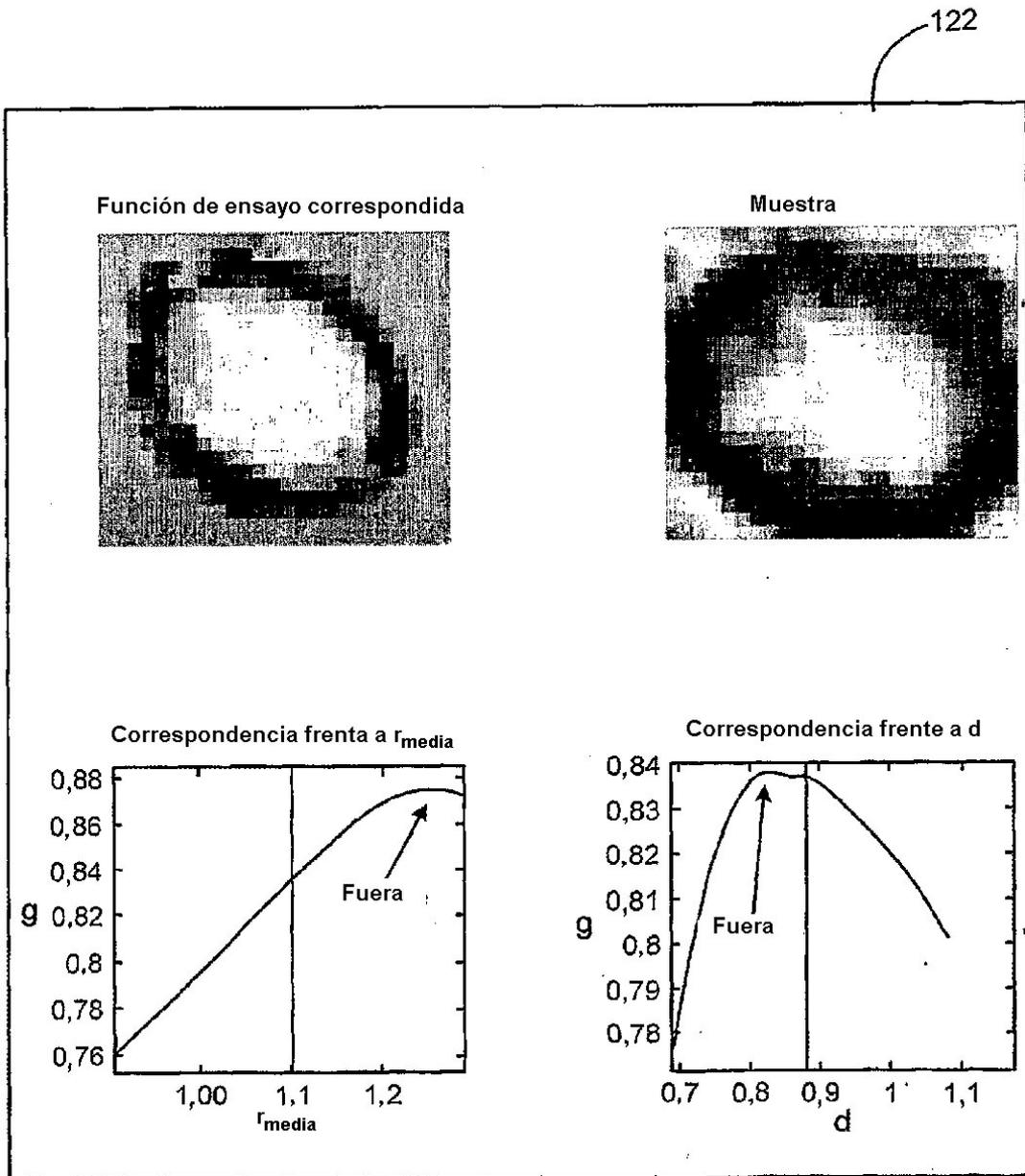
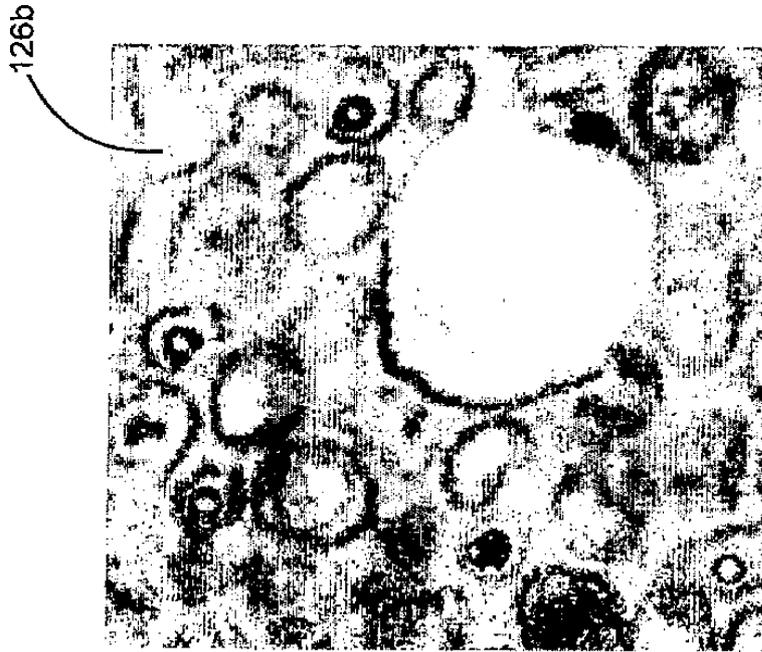


FIG. 5B



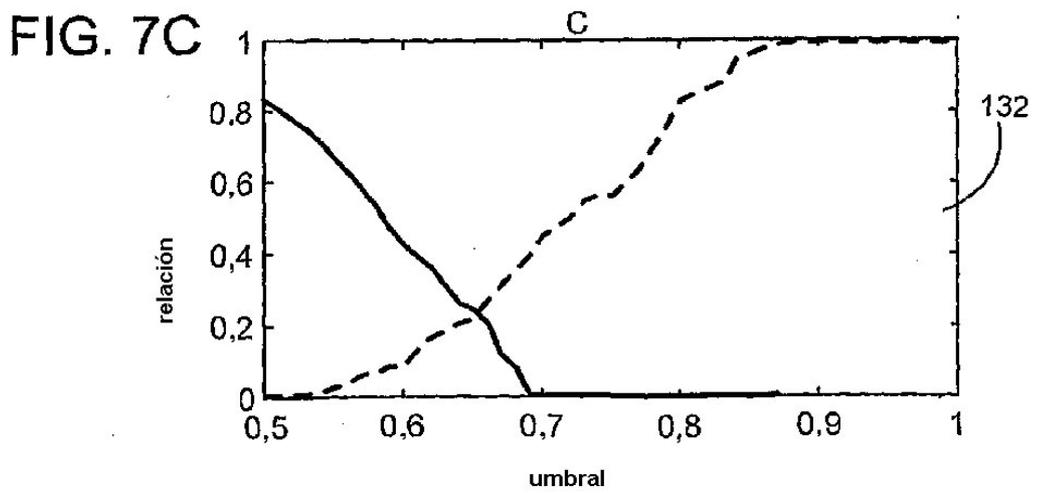
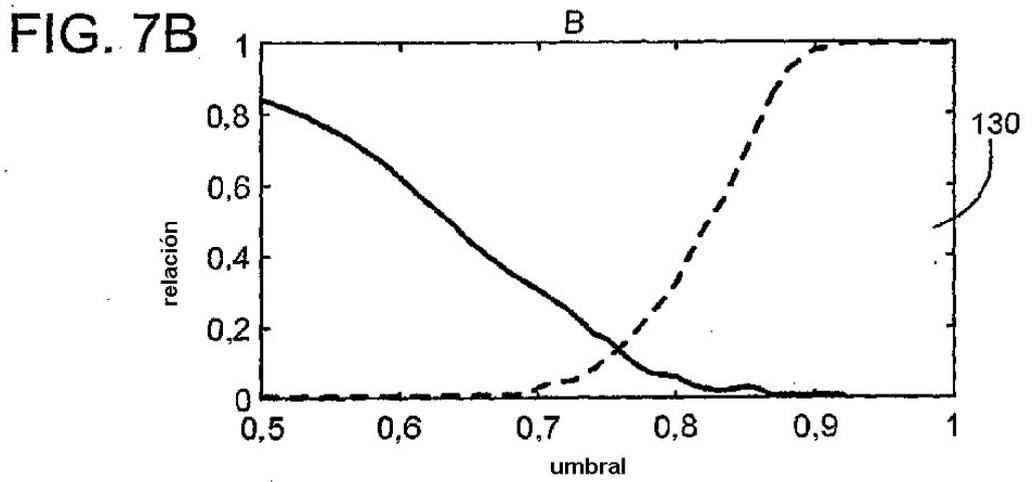
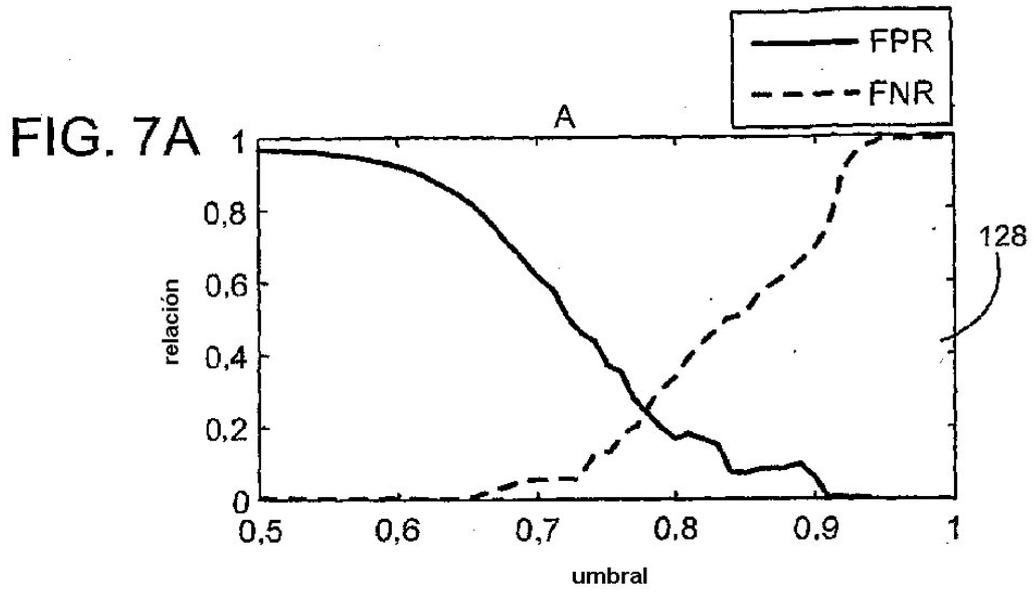


FIG. 8

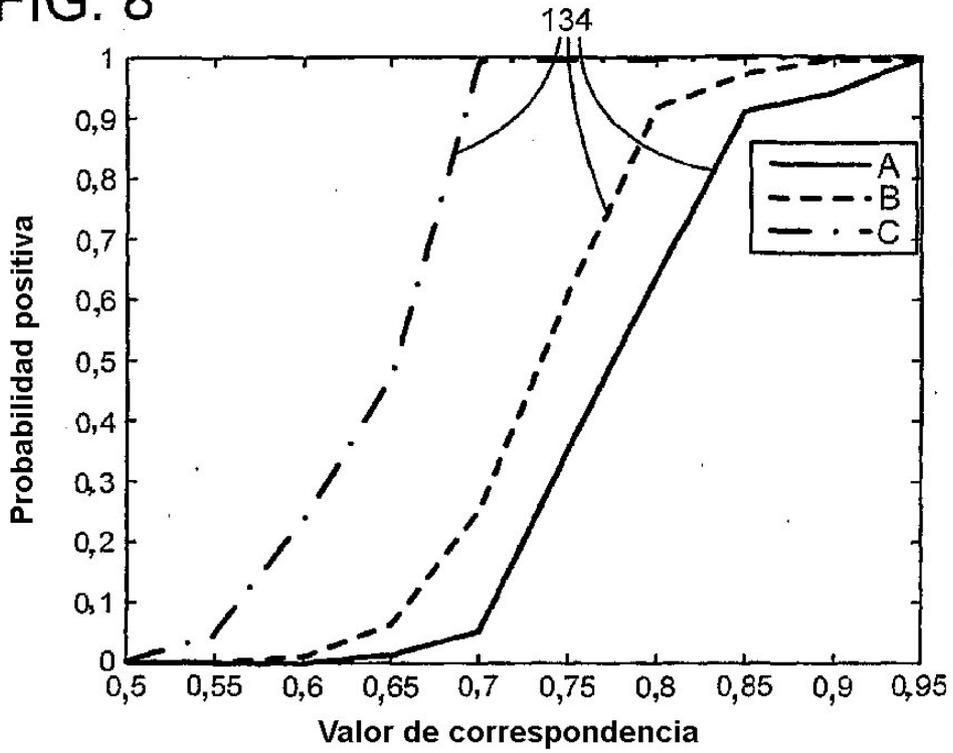


FIG. 9

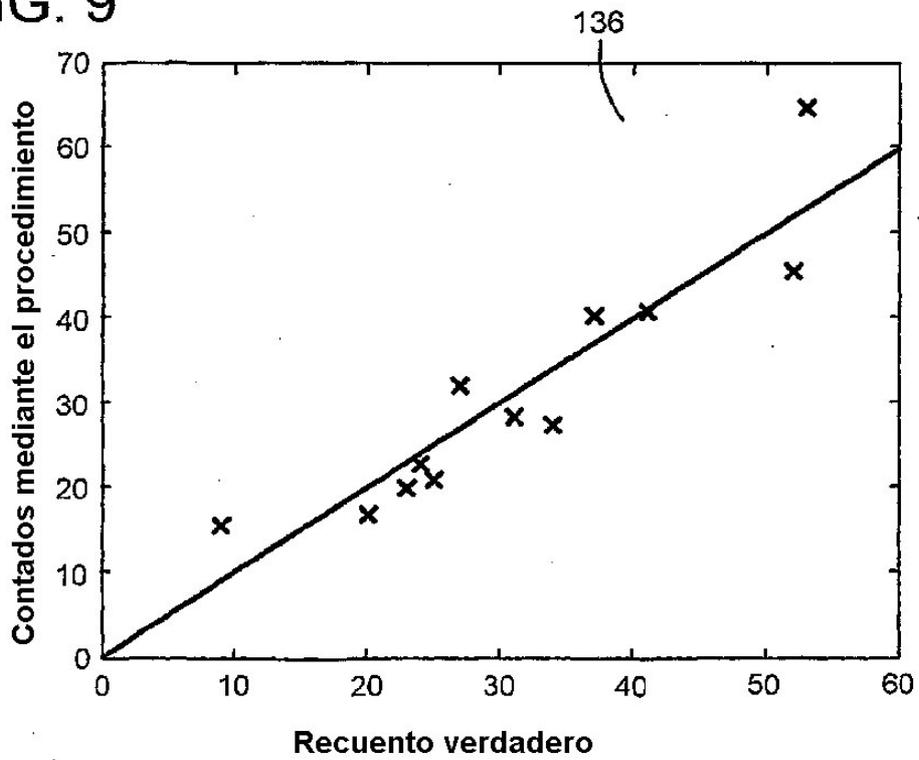


FIG. 10

138

