



República Federativa do Brasil
Ministério do Desenvolvimento, Indústria
e do Comércio Exterior
Instituto Nacional da Propriedade Industrial.

(21) **PI0617610-0 A2**



(22) Data de Depósito: 12/09/2006
(43) Data da Publicação: 26/07/2011
(RPI 2116)

(51) *Int.Cl.:*
G06K 9/00 2006.01
G06K 9/42 2006.01
G06K 9/46 2006.01

(54) Título: **IDENTIFICAÇÃO E CLASSIFICAÇÃO DE PARTÍCULAS DE VÍRUS EM MICRÓGRAFOS DE ELÉTRON TEXTURADOS**

(30) Prioridade Unionista: 12/10/2005 US 60/725,806

(73) Titular(es): INTELLIGENT VIRUS IMAGING INC.

(72) Inventor(es): MARTIN RYNER, MOHAMMED HOMMAN

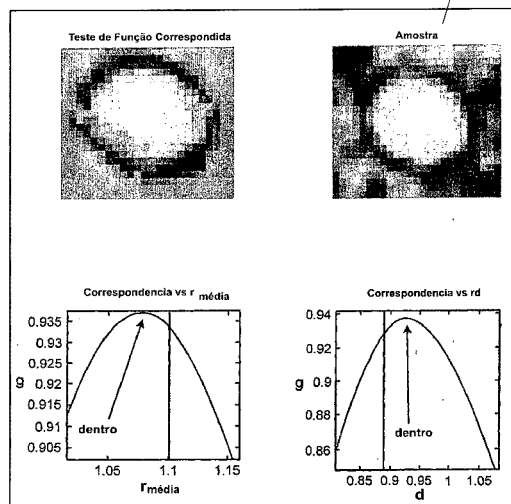
(74) Procurador(es): Brasil Sul Marcas e Patentes S/C Ltda.

(86) Pedido Internacional: PCT US2006035758 de 12/09/2006

(87) Publicação Internacional: WO 2007/046985 de 26/04/2007

(57) Resumo: IDENTIFICAÇÃO E CLASSIFICAÇÃO DE PARTÍCULAS DE VÍRUS EM MICRÓGRAFOS DE ELÉTRON TEXTURADOS, o método é para a identificação e caracterização de estruturas em micrógrafos de elétron. São selecionadas estruturas em uma primeira imagem. As estruturas têm um primeiro tipo de formato deformado em uma primeira direção. As estruturas selecionadas são transformadas em um segundo tipo de formato diferente do primeiro tipo de formato. As estruturas transformadas do segundo tipo de formato são usadas para formar uma pluralidade de máscaras. É identificada uma nova estrutura em uma segunda imagem. A nova estrutura tem o primeiro tipo de formato. A estrutura do segundo tipo de formato, de cada máscara, é deformada na primeira direção. É determinada que máscara será a máscara preferida e que melhor combina com a nova estrutura.

120





"IDENTIFICAÇÃO E CLASSIFICAÇÃO DE PARTÍCULAS DE VÍRUS EM MICRÓGRAFOS DE ELÉTRON TEXTURADOS".

Campo da invenção

A presente invenção relaciona-se com a identificação de estruturas em
5 imagens. Em particular, a presente invenção oferece um método que é uma
combinação para a identificação e classificação de partículas de vírus em
micrografos de elétron texturados.

Histórico da invenção

A montagem de vírus é um processo complexo e objeto de intensas
10 pesquisas. Os vírus utilizam uma célula hospedeira para produzir suas partículas
vírus progênie através de um complexo processo de maturação e transporte
intracelular. Esse processo pode ser monitorado sob alta ampliação, com o uso de
microscopia eletrônica (feixes de elétrons), que permite a identificação visual de
diferentes tipos de partículas de vírus em diferentes compartimentos celulares.
15 Questões importantes que ainda precisam ser resolvidas incluem a identidade de
proteínas virais que estão envolvidas em cada passo desse processo de montagem de
vírus como também o mecanismo sob a translocação intracelular e localização de
diferentes tipos de partículas de vírus durante a maturação do vírus. Aspectos
estruturados da maturação do vírus são, em geral, difíceis de lidar embora técnicas
20 de visualização como a tomografia e cryo EM têm contribuindo muito para as vastas
informações relacionadas com estruturas de vírus. Essas técnicas oferecem
informações sobre partículas de vírus maduras estáveis. As ferramentas genéticas
ajudar a produzir mutantes de componentes chave de proteína viral e os efeitos
estruturais podem ser visualizados por EM. Porém, a falta de ferramentas
25 apropriadas para caracterizar os efeitos estruturais, especialmente formas
intermediária e partícula obscura e para quantificar isso de maneira apropriada e
objetiva. As ferramentas de análise de imagens para caracterizar e quantificar a
maturação de partícula de vírus e transporte intracelular o que facilitaria estudos
objetivos sobre os diferentes estados de montagem de vírus, usando microscopia
30 eletrônica (feixes de elétrons). Muitas informações foram obtidas, mas precisam ser

estruturadas e estatísticas precisam ser calculadas com elas para uma correta avaliação do efeito e para se chegar às devidas conclusões.

Resumo da presente invenção

A caracterização da morfologia estrutural de partículas de vírus em
5 micrógrafos de elétron é uma tarefa complexa, mas desejável para a investigação do
processo de maturação e detecção de mudanças na morfologia de partícula viral em
resposta ao efeito de uma mutação ou aplicação de medicamentos antivirais. Assim,
foi desenvolvido um procedimento para a descrição e classificação de formas de
partícula viral em micrógrafos de elétron, com base na determinação de
10 características invariáveis da projeção de uma determinada estrutura de vírus.

A máscara para o vírus é criada com base nas informações obtidas de
um pequeno conjunto de treinamento de micrógrafos de elétron e, então, é
empregado para classificar e quantificar estruturas similares de interesse em um
número ilimitado de micrógrafos de elétron pelo processo de correlação. Usando
15 análise de deformação linear, esse algoritmo novo, descrito aqui, pode tratar das
variações da partícula viral como elipsidade e ainda permite a avaliação de
propriedades como o tamanho e orientação da partícula viral. A aplicação prática
desse método é demonstrada pela habilidade de localizar três classes diversas de
partículas de vírus em micrógrafos de elétron transmissão de fibroblastos infectados
20 com citomegalovirus humano.

Em resumo, o método é para a identificação e caracterização de
estruturas em micrógrafos de elétron. São selecionadas estruturas em uma primeira
imagem. As estruturas têm um primeiro tipo de formato deformado em uma
primeira direção. As estruturas selecionadas são transformadas em um segundo tipo
25 de formato diferente do primeiro tipo de formato. As estruturas transformadas do
segundo tipo de formato são usadas para formar pluralidades de máscaras. É
identificada uma nova estrutura em uma segunda imagem. A nova estrutura tem o
primeiro tipo de formato. A estrutura do segundo tipo de formato, de cada máscara,
é deformada na primeira direção. É determinada que máscara será a máscara
30 preferida e que melhor combina com a nova estrutura.

Breve Descrição dos Desenhos

As Figuras 1A e 1B mostram imagens típicas de micrográfico de elétron transmissão de vírus herpes em desenvolvimento:

A Figura 2A mostra nucleocapsídeos vazios de vírus herpes;

5 A Figura 2B mostra nucleocapsídeos de vírus herpes com núcleo translúcido;

A Figura 2C mostra nucleocapsídeos de vírus herpes contendo DNA empacotado;

A Figura 3A mostra a partícula viral com formato elíptico;

10 A Figura 3B mostra a partícula viral que foi deformada para ficar circular;

A Figura 4A-C mostra funções teste para estruturas capsídeo viral (A, B e C) em micrografos de elétron não empregando coeficiente de redução (nenhum) ou 80% dos coeficientes exibindo a menor variação (VAR);

15 A Figura 5A mostra uma combinação da função teste A para uma estrutura capsídeo autêntica e para uma estrutura similar, mas falsa;

A Figura 5BA mostra uma combinação da função teste B para uma estrutura capsídeo autêntica e uma estrutura similar, mas falsa;

20 A Figura 6 mostra combinação com a função teste A dentro de uma vesícula;

A Figura 7A-C mostra razões de um falso positivo (FPR) e um falso negativo (FNR) para as diferentes funções teste A, B e C, respectivamente;

A Figura 8 mostra as funções de probabilidade positiva (PPFs) para as funções teste A, B e C;

25 A Figura 9 mostra a comparação do número total real de estruturas virais presentes em um conjunto de imagens teste (eixo X) como determinado por um virologista para o número identificado por nosso procedimento (eixo Y); e

30 A Figura 10 mostra uma produção automatizada de um mapa que identifica locais de interesse em um micrográfico de elétron ilustrado aqui para a função teste C.

Descrição detalhada

É descrito aqui o desenvolvimento de um sistema automatizado para assistir na identificação de partículas de vírus em micrografos de elétron. Como modelo, foram usados fibroblastos infectados com citomegalovírus humano (HCMV) um vírus da classe β -herpes. Esclarecendo, que o vírus da herpes é usado apenas como exemplo ilustrativo e a invenção não se limita ao vírus da herpes. Durante a infecção com o citomegalovírus humano, são produzidas muitas diferentes formas intermediárias da partícula viral. Durante a montagem do vírus da herpes, a célula hospedeira é forçada a fazer cópias do material genético viral e a produzir capsídeos, uma concha de proteínas virais, que encerra e protege o material genético. Os capsídeos são estruturas esféricas que podem variar com relação ao tamanho e simetria e podem, quando maduros, ser envelopados por uma membrana bi-camada. A maturação dos capsídeos de vírus é um estágio importante na produção de partícula viral, e um dos mais frequentemente estudados. Porém, a aparência dos micrografos de elétron varia consideravelmente, o que faz do processo de análise um desafio. Uma característica única do vírus da herpes é o tegumento, uma camada de proteínas virais que circunda o capsídeo antes do envelopamento final. O envelope é obtido pelo desenvolvimento do tegumento dos capsídeos em vesículas de secreção no citoplasma. Depois disso, as partículas virais infecciosas saem da célula hospedeira por fusão desses vírus, contendo vesículas, com a membrana de plasma.

Um procedimento objetivo para a classificação e quantificação de partículas de vírus desenvolvido em micrografos de elétron transmissão. Na análise relacionada de imagens de microscópio crio-elétron (cryo-EM), foram dedicados consideravelmente mais esforços para explorar diferentes métodos de identificação. Em crio-micrografos, foi usada, com sucesso, a correlação cruzada empregando máscaras múltiplas e métodos para a detecção de extremidades.

Abordagens apropriadas permitem a caracterização e a qualificação da maturação das partículas de vírus e a translocação intracelular facilita estudos objetivos desses fenômenos, empregando microscopia eletrônica (feixes de

elétrons). Porém, as imagens de microscópio eletrônico (feixes de elétrons) são difíceis de ser analisadas e descritas em razão do pesado fundo texturado. Ainda, cada partícula viral, em particular, mostra uma ampla variedade de formas, dependendo da sua projeção no micrográfico de elétron, o procedimento utilizado para preparar amostras para a microscopia eletrônica (feixes de elétrons) e as condições fotográficas usadas. As imagens de micrográfico de elétron típicas 100, 102, que oferecem informações valiosas, são mostradas nas Figuras 1A e 1B, respectivamente.

Na presente invenção, foi usada uma abordagem para a análise de capsídeos HCMV no núcleo de células infectadas, que estão nos estados de maturação definidos como capsídeos vazios 104 (chamados de A), capsídeos com núcleo translúcido 106 (chamados de B) e capsídeos contendo DNA empacotado 108 (chamados de C), como bem mostrado nas Figuras 2A-C.

O método e combinação de acordo com a presente invenção são ilustrados com partículas virais. Isso deve ser visto como um exemplo não limitante. Outros tipos de partículas, incluindo, por exemplo, objetos biológicos como células ou estruturas de células, mas também partículas e estruturas não orgânicas, podem ser identificadas e caracterizadas com pequenas modificações no método e combinação descritos.

O método, de acordo com a invenção, inclui um passo para aquisição de imagem. O micrográfico de elétron pode vir do microscópio eletrônico (feixes de elétrons) como arquivos ou figuras para digitalização. Para os próximos passos dos métodos, é aconselhável adquirir conhecimento sobre tamanho de pixel, resolução e aumento de cada micrográfico.

Em um passo pré-processo, as partículas relevantes são selecionadas e transformadas, de possíveis aparências deformadas para círculos.

Em um passo de formação de máscaras, as partículas selecionadas e transformadas são usadas para formar uma máscara, que pode ser caracterizada por uma função teste.

Em um passo de combinação, a máscara ou função teste é utilizada para identificar as partículas em outra(s) imagem(ns). Os passos do método serão mais descritos e exemplificados mais adiante.

Um aparato de identificação e classificação, de acordo com a presente invenção, pode ter como base um PC comum, mas com força suficiente para os cálculos. O aparato de identificação e classificação vem com uma interface para recepção de micrografos, meios de pré-processamento para transformação de imagens deformadas, meios para formação de máscaras ou extração de funções teste e meios para execução de um procedimento de combinação. Esses passos são tipicamente e preferencialmente executados por módulos de código de software.

Culturas de células como fibroblasto pulmonar embriônico (HF) foram mantidas em meio essencial mínimo, livre de bicarbonato, com sais Hank (GIBCO BRL) suplementado com 25 mM HEPES [4-(2 hidroxietil)-1-lácido etanesulfônico piperazina], 10% soro novilho fetal calor-ativado, L-glutamina (2mM), penicilina (100 U/ml) e streptomicina (100 mg/ml) (GIBCO BRL, Grand Island, NY, USA). A cultura de células foi mantida em frascos de cultura de tecido de 175 cm² (Corning, New York, USA) para um máximo de 17 passagens.

Em um passo de infecção viral, as células HF foram infectadas com HCMV cepa AD169 empregando uma multiplicidade de infecção (MOI) de 1. O sobrenadante viral foi coletado 7 ou 10 dias pós-infecção (dpi), purificado de resíduos celulares por centrifugação de baixa velocidade e congelado a -70°C até ser usado para inoculação.

Para examinar células infectadas com vírus, por microscopia eletrônica (feixes de elétrons), células não infectadas e células infectadas com HCMV foram colhidas em 1, 3, 5 e 7 dpi e, conseqüentemente, fixadas com 2% glutaraldeído em tampão 0.1 M cacodilato de sódio contendo 0.1 M sacarose e 3 mM CaCl₂, pH 7.4, em temperatura ambiente por 30 minutos. As células foram, então, raspadas com espátulas de madeira e transferidas para um tubo Eppendorf para continuar a fixação durante a noite, a 4°C. Seguindo esse procedimento, as células foram enxaguadas em tampão 0.15 M cacodilato de sódio contendo 3 mM

CaCl₂, pH 7.4 e transformado em esferas por centrifugação. Essas esferas foram pós-fixadas em 2% tetróxido de ósmio dissolvido em tampão 0.07 M cacodilato de sódio contendo 1.5 mM CaCl₂, pH 7.4, a 4°C por 2 horas; seqüencialmente desidratado com etanol e acetona; e fixado em LX-112 (Ladd, Burlington, VT, USA). O contraste nas seções foi obtido com uranil acetato seguido por citrato de chumbo e exame executado por um Philips 420 ou um Tecnai 10 (FEI Company, Oregon, USA) microscópio eletrônico transmissão (feixes de elétrons) em 80 kV.

Foram executadas, então, a aquisição de imagem, discretização e análise. Os micrografos de elétron de células infectadas com HCMV foram digitalizadas empregando uma escala em cinza de 8-bit, em resolução de 5.5 nm/pixel em uma impressora HP Scanjet 3970. A implementação foi feita com Matlab 7.0.1 (The Mathworks Inc., Natick, MA, USA) e software Sun Java 1.4.2 em um PC Dell Optiplex GX260. Essa análise envolve uma interface gráfica fácil de usar e a automação dos parâmetros descritos abaixo para um uso rápido e conveniente.

Foram desenvolvidas ferramentas confiáveis e fáceis de usar para estudos da montagem intracelular de vírus. A abordagem baseou-se no achado de um conjunto compacto de pontos em R^2 , o campo do micrografo, para cada um deles um ponto recebeu um valor função correspondente. Esse conjunto de pontos e seus valores função são coletivamente chamados de função teste ou máscara e podem ser descritos por uma seqüência $\{(x_k, c_k)\}$ onde x é o ponto e c é o valor função. A função teste é, preferencialmente, produzida de tal maneira que a seqüência do valor função está correlacionada com os valores na escala de cinza dos pontos correspondentes. De acordo, um conjunto definido de partículas de vírus do mesmo tipo é necessário para treinar e conceber a seqüência para a execução de uma máscara para essa estrutura de partícula específica. Essa esparsa representação, permite a fácil deformação e os ajustes da máscara para partículas virais individuais cujo formato no micrografo é, mais ou menos, elíptico.

Em um passo pré-processamento de deformação, as posições das subestruturas, dentro do mesmo tipo de partículas virais variam nas diferentes imagens. Por exemplo, as partículas virais são, algumas vezes, deformadas de maneira que apareçam em diferentes formas elípticas. Para criar as funções teste, espaços vetor-linear foram usados o que demanda que as posições espaço-vetor analisadas sejam relativamente fixas. A transformação linear uniforme foi escolhida para aproximar as deformações, já que cobre as deformações mais proeminentes vistas em micrógrafos. O custo computacional desses cálculos é razoavelmente baixo e simplifica o gerenciamento de delimitações. Essa abordagem, geralmente, exige o uso de operador de transformação quadridimensional, isto é, uma matriz 2x2. Essas variáveis envolvidas podem ser expressas como a rotação da estrutura antes da deformação (φ_R), deformação radial primária (Γ), taxa de deformação originando a estrutura elíptica (d) e a rotação seguindo a deformação (φ_D). Juntos formam a transformação abaixo:

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

Para identificar as variáveis de transformação para uma partícula viral individual, um conjunto elipse foi usado manualmente para estimar a posição, tamanho e deformação de cada parede de capsídeo, como melhor mostrado pelas Figuras 3A e 3B. A imagem 110 (chamada de A) tem um formato elíptico enquanto a imagem 112 (chamada de B) foi deformada, como descrito, para obter um formato mais circular. Assim, oferecendo três (φ_D , r e d) das quatro variáveis. A amostra foi, então, parcialmente transformada para obter o raio primário medido sem deformação ($d=1$), como ilustrado na Figura 3B.

As características que são independentes de rotação, como a arquitetura poligonal da parede do capsídeo e a posição do núcleo DNA podem ser determinadas pelo valor φ_R para cada amostra. Para achar esse valor, cada amostra parcialmente transformada pode ser normalizada ao redor de seu meio, no interior de um círculo, cobrindo a área de imagens visualmente significativas 114a, 116a e

118a, como mostrado na coluna da esquerda das Figuras 4A-C. Então, a soma dos quadrados das distâncias na L^2 -sência para cada amostra pode ser minimizada com relação aos ângulos. Já que essa minimização envolve variáveis $N-1$, com N sendo o número de amostras referência considerando uma amostra a ser fixada, esse procedimento pode ser simplificado pela minimização das distâncias até as amostras já processadas, uma a uma. Todas as transformações das imagens podem, então, ser implementadas de maneira bi-linear, assim aproximando o valor da função f no ponto (x, y) como:

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

onde \underline{x} é o número inteiro menor, mais aproximado, de x , x é o maior número inteiro aproximado e $x_m = x - \underline{x}$. A integração pode ser feita usando a mesma interpolação. As medições obtidas desse passo de processamento apresentam indicativos da gama de propriedades de deformação, isto é, o raio principal (raio primário) e taxa de deformação, mas esses parâmetros devem ser determinados com base em experiência adicional. Já que todos os tipos de rotação e todas as direções de deformação das estruturas virais devem estar presentes em micrografos de elétron, essas variáveis são preferencialmente não fixas.

Os pontos e valores função locais (parâmetros) para as máscaras de partícula viral podem, então, ser identificados. Uma vez que as amostras deformadas são alinhadas com a estrutura parcial, nas mesmas posições, essa abordagem pode ser usada para achar os valores da função invariável. Para descrever esse procedimento mais claramente, uma amostra deformada f pode ser convertida em um gráfico dessa função pela enumeração (listar individualmente) da posição pixel x e os correspondentes valores função c como $f = \{(x_k, c_k)\}_k$. O grau de combinação entre duas seqüências de valores função y_i e y_j (chamados abaixo de vetores) contendo a mesma seqüência de posição pixel determinada usando o padrão estimado pela correlação estatística:

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

Onde y é o valor médio do vetor e a combinação de todos os coeficientes para $[-1,1]$ é mapeado. A base lógica para o uso dessa abordagem está no fato de que ela indica o grau de similaridade linear entre as duas estruturas.

5 Depois de colocar os dois vetores amostra normalizados ao redor de sua média

$\hat{y}_i = \frac{y_i - \bar{y}_i}{|y_i - \bar{y}_i|}$ dentro de colunas, em uma matriz, a seqüência de função teste $f_c (|f_c|^{-1})$ que faz $\|A' f_c\|$ tão grande quanto é determinado, assim oferecendo a melhor combinação para as amostras usadas para treinamento.

10 A decomposição de valor singular (SVD) pode ser descrita como segue:

$$\|A' f_c\| = \|\mathbf{V} \Sigma \mathbf{U}' f_c\| = \quad (\mathbf{V} \text{ é quadrado e ortonormal}) = \|\Sigma \mathbf{U}' f_c\| = \|\Sigma w\|$$

é aplicado à A onde $\|w\| = 1$ if $f_c \in \text{span}(U)$ o que seria esperado.

A última expressão é maximal onde w é o eigenvector correspondente ao maior eigenvalue de Σ (que é o maior valor singular) e f_c deveria então ser a coluna em U correspondente. Já que essa função é uma combinação linear das colunas em A, a combinação (eq. 2a) reduz para

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

20 A função teste nesse SVD inicial usa os coeficientes de todos os pontos associados com o primeiro suporte aceito. Alguns desses pontos se localizam, de certo modo, fora das estruturas virais nas imagens e, em adição, existem pontos nas estruturas cujos coeficientes podem variar consideravelmente. Assim, para escalonar o significado de cada coeficiente e ali eliminar o pior da variação, o valor de

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

Foi calculado para cada coeficiente. Certa porcentagem de pontos pode, então, ser guardada na função teste. Já que essas operações mudam de acordo com a função teste, um novo SVD foi, subseqüentemente, calculado.

As Figuras 4A-C ilustram as funções teste obtidas usando todos os coeficientes ou apenas aqueles 80% dos coeficientes variando, identificados e exibindo, pelo menos, a menor variação, de acordo com a classificação de variação. Claramente, o tamanho do núcleo DNA varia na função teste para o capsídeo C e por essa razão os pontos mais incertos foram eliminados nas imagens da direita 114b, 116b e 118b. Em conformidade, as funções teste obtidas pela redução do número de coeficientes, nessa maneira, foram empregadas rotineiramente.

As deformações, então, podem ser sintetizadas. Já que as estruturas analisadas foram consideradas como orientadas em qualquer direção e linearmente deformadas em qualquer direção, essas características devem ser automaticamente aplicadas à função teste quando uma imagem for analisada. A informação oferecida pelo comportamento da função combinante ao deformar a função teste também é de interesse e foi explorada em uma situação similar. Enquanto mantém a imagem B e a função teste f_C fixada e variando a deformação T, análise da função combinante $g(T) = M(f_C, \{B(Tx_k)\}_k)$ (onde a seqüência $\{x_k\}_k$ é obtida da produção de funções teste executadas. Para descrever T em termos de parâmetros $(\varphi_R, \bar{r}, d, \varphi_D) \in ([0, 2\pi], [\bar{r}_0, \bar{r}_1], [d_0, d_1], [0, 2\pi]) - T_{bound}$, as seguintes suposições são feitas:

(i) Para certo $T \in T_{bound}$, a função teste deformada representa a estrutura mais similar ao objeto na imagem. Presume-se que esse T é o que maximiza g.

(ii) O T associado com deformação máxima deve ser localizado dentro do interior do conjunto de deformação, e não na delimitação. Sob essas condições, mesmo se g for maximizado fora do conjunto (isto é a estrutura é muito grande, muito pequena ou está muito deformada), combinando com os pontos de delimitação mais próximos, podendo ainda ser alto.

Para ser considerado como identificado, a estrutura deve estar em conformidade com esses critérios. A maximização da função de combinação foi executada com um esquema descendente íngreme reverso, usando a função teste não deformada como ponto de início e aproximando o derivativo como um ponto-oito, esquema diferença centrada (isto é dois pontos para cada variável na deformação).

A aplicação dos critérios de combinação empregados é mostrada nas Figuras 5 e 6. A Figura 5A ilustra como esses critérios trabalham quando aplicados a um autêntico capsídeo A, como também à uma estrutura similar, mas falsa. Na imagem 120 (chamada de A) é mostrado um capsídeo autêntico. Quando a função teste é deformada, os gráficos ilustram como a função combinante g varia com tamanho radial (\bar{r}) e grau de deformação (d) do ponto no conjunto de deformações admissíveis que maximizam g . A função teste deformada tem uma aparência similar à da amostra e a deformação fica dentro das delimitações. A classificação deve, então, ser positiva. Na imagem 122 (chamada de B) na Figura 5B, diferente da imagem A, o ponto no conjunto de deformação que maximiza (g) fica na extremidade e os gráficos mostram um valor combinante mais alto fora desse conjunto. Assim, essa classificação deve ser negativa. Nesse caso as delimitações de deformação foram ajustadas para $(\varphi_R, \bar{r}, d, \varphi_D) \in ([0, 2\pi], [0.89, 1.1], [0.89, 1.13], [0, 2\pi])$ para propósitos ilustrativos.

Os capsídeos virais deixam o núcleo crescendo através da membrana dessa organela. Em relação à esse processo é difícil discriminar entre a estrutura viral e outras estruturas, como mostrado nas imagens 126a e 126b na Figura 6. A estrutura marcada com a cruz azul está em conformidade com os critérios de combinação (i) e (ii) onde os marcados com um círculo vermelho apenas preenchem critério (i). Nessa Figura, uma cruz azul indica um ponto na imagem onde a combinação entre a função teste e a estrutura capsídeo é melhor que 0.8 e o grau de deformação é aceitável. Um círculo vermelho indica uma ponte onde essa combinação é melhor que 0.8, mas onde o grau de deformação não é admissível. A

estrutura marcada como uma combinação tem um combinante de 0.94, que é bastante alto.

As estruturas de partículas virais em uma imagem de microscópio eletrônico podem, então, ser identificadas. Para procurar por estruturas em uma
5 imagem (B) similar à função teste f_c , eq 2b é expandido para convoluções. A combinação da função teste num ponto (m) pode, assim, ser expressa como

$$M_{B,f_c}(m) = \sup_{T \in T_{band}} M(f_c, \{B(m - Tx_k)\}_k)$$

Porém, esse procedimento toma muito tempo. Mas ele pode ser acelerado se tomadas as seguintes ações e observações:

10 (i) As variantes deformadas das funções teste não são ortogonais uma para com a outra e já que essas estruturas são essencialmente independentes de rotação, a combinação de funções teste não deformadas é melhor que a de certo valor para qualquer estrutura deformada admissível do mesmo tipo.

(ii) Já que a tradução deforma ainda mais a estrutura, a combinação
15 com a função teste não deformada deve ser mais alta na posição real de uma partícula viral do que em locais, pelo menos, um diâmetro da função teste longe dessa posição.

Ao implementar esses critérios, é possível identificar um subconjunto de potencialmente interessante pontos dentro da imagem maior. Depois disso pode
20 ser feita outra análise do conjunto empregando a otimização descrita na seção anterior. Essa abordagem oferece um conjunto final de pontos na imagem que estão associados com os valores combinantes de $P = \{M_j\}_j$. Para assegurar a inclusão de todas as posições interessantes em uma imagem, o valor limiar relacionado com o pressuposto (i) acima estabelecido em 0.5.

25 No pós-processamento do conjunto final, as partículas virais são contadas. Não existe valor limiar (t) que possa distinguir entre estruturas autênticas e estruturas falsas em todas as imagens, isto é, a designação de estruturas empregando esse procedimento não concorda completamente com o feito por um virologista experiente. O estabelecimento do nível limiar não é, portanto, uma

opção. Ao invés disso, uma função probabilidade positiva $PPF : [-1, 1] \rightarrow [0, 1]$, pode ser usada para determinar a probabilidade de um determinado ponto, associado com certo valor combinante, estar realmente associado com uma partícula viral. A extensão do valor preditivo positivo (PPV) é obtido calculando a razão entre o número de estruturas corretamente identificadas e o número total de estruturas identificadas com certo valor combinante. Assim, para um conjunto (P) de estruturas identificadas por esse procedimento contendo um subconjunto $P_{correct}$ de pontos associados com partículas virais de certo tipo,

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

Para obter uma função em crescimento, suavemente e monotonicamente, 0.05 foi escolhido como valor para ε . A função de probabilidade indicando um número esperado (N) de estruturas na imagem é

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

A precisão FNR/FPR do método pode ser descrita como segue. Para organizar partículas virais em micrógrafos de elétron, de acordo com o estágio de maturação, um modelo como o descrito aqui é necessário para retratar cada estágio em particular. Ainda, para que esse modelo seja útil para a detecção e quantização de partículas de vírus em tais imagens, também precisa poder rejeitar estruturas espúrias. Assim, o modelo ideal deverá detectar todas as imagens possíveis de partículas de vírus, de diferentes tipos, mas nada mais é localizado no mesmo espaço, isto é, no fundo. Para caracterizar nosso modelo nesse respeito são utilizadas, comumente, as razões de falso negativo (FNR) e falso positivo (FPR). A razão FNR é definida como a razão entre o número de partículas virais autênticas rejeitadas incorretamente pelo método e o real número de partículas autênticas, enquanto a razão FPR é a razão entre o número de estruturas espúrias identificadas como sendo autênticas e o número total de estruturas consideradas como autênticas por essa abordagem. Assim, as duas razões recaem entre 0 e 1, com 0 sendo o ideal.

Para determinar o número de partículas de vírus com base nas informações oferecidas pelo conjunto de valores combinantes adquiridos pela busca

em uma imagem, a função de probabilidade positiva PPF, descrita acima, pode ser usada. O número esperado de partículas identificadas foi comparado com o número verdadeiro de partículas presentes na imagem para obter uma média e um desvio padrão para erro de contagem. Ainda, para avaliar se existe uma diferença média sistemática, isto é, se o procedimento identifica em média muitas ou poucas partículas, a hipótese H_0 que: “A média de diferença = 0” foi testada.

A padronização e testes foram feitos em conjuntos separados de imagens, 2 para treinamento e 12 para testes. O número de amostras usadas para padronização foi 4, 7 e 10 para as funções teste A, B, e C, respectivamente. As 10 imagens de teste contendo um total de 53 capsídeos A, 239 capsídeos B e 83 capsídeos C e os limites de deformação foram estabelecidos em $(\varphi_R, \bar{r}, d, \varphi_D) \in ([0, 2\pi], [0.89, 1.1], [0.89, 1.2], [0, 2\pi])$.

As razões falso negativo (FNR) e falso positivo (FPR) foram descritas como segue. O método foi avaliado pela comparação de nossos resultados com as 15 experiências de virologistas. A FPR e a FNR foram calculadas como uma função de valor limiar para a medida combinante, melhor vistas nos gráficos 128, 130 e 132 nas Figuras 7A-C, respectivamente. A razão FNR é definida como a razão entre o número de estruturas autênticas rejeitadas incorretamente pelo procedimento empregando certo valor limiar para a medida combinante, e o número real de 20 partículas de vírus presentes, como determinado por um virologista. Analogamente, a razão FPR é a razão entre o número de estruturas espúrias identificadas como sendo autênticas e o número total de estruturas consideradas como autênticas por esse procedimento. Por comparação, com outros métodos, o cruzamento das curvas ocorre em 0.25 para a função teste A, 0.13 para a função teste B e 0.23 para a 25 função teste C.

A quantização de estruturas em micrografos de elétron pode ser descrita como demonstrado abaixo. Os valores PPF 134 calculados dos resultados apresentados acima são mostrados na Figura 8. O gráfico mostra a frequência relativa das partículas de vírus identificadas corretamente por esse procedimento,

em certo valor combinante. Para comparação, um método ideal que ofereça uma separação completa entre as estruturas falsas e as estruturas verdadeiras resultaria em uma função passo Heaviside em algum valor limiar. Para comparação, um procedimento de caso ideal que ofereça uma separação completa entre as estruturas falsas e as estruturas verdadeiras resultaria em uma função passo Heaviside em algum valor limiar.

Uma trama dispersa 136 do número total de partículas virais identificadas como presentes em um conjunto de imagens teste por nosso procedimento em comparação com o número correto, como determinado por um virologista, como mostrado na Figura 9 junto com a função identidade. A linha nesse gráfico mostra a função identidade. A média de diferença é de 0.16 e o desvio padrão é de 5.63. O nível de significação da hipótese nula H_0 , isto é, “A média diferença = 0”, é 0.92. Claramente, não existe semelhança entre esses dois valores (média diferença = 0.16, desvio padrão 5.63), que no caso ideal seriam pontos na função identidade. O fato que o nível de significação de H_0 foi 0.92 de acordo com o t-test de Estudante indica que existe uma possibilidade de que não houve diferença sistemática entre as duas abordagens, na média. Esses resultados mostram que uma amostragem rápida do número total de estruturas virais, em diferentes estágios de maturação, em um grande conjunto de micrografos de elétron, é uma tarefa que, de outra maneira, consome muito tempo e é muito tediosa para um especialista e pode ser rapidamente e confiavelmente completada por nosso procedimento automatizado.

Isso pode ser produzido com base no conjunto de posições na imagem 138 onde as estruturas de interesse são localizadas em um mapa - como mostrado na Figura 10. Isso facilita, consideravelmente, a contagem manual dessas estruturas e também oferece uma estrutura para a análise manual. Ao invés de, simplesmente, contar e comparar as estruturas em uma imagem não processada, o virologista é ajudado, consideravelmente, nessa tarefa pela disponibilidade de tal mapa. As diversas estruturas são separadas da direita para a esquerda, em ordem descendente de valores combinantes, começando do lado esquerdo, no topo da fila.

Ao investigar o processo de montagem de vírus, as informações relacionadas com a topologia estrutural, em relação ao estágio de maturação, não são, geralmente, disponibilizadas ou são vagamente definidas. Portanto, são necessárias ferramentas para separação e classificação de partículas virais, em diferentes estágios de maturação. Uma vez que alguns pontos iniciais foram obtidos por classificação de um conjunto de estruturas óbvias, eles podem ser usados para expandir o conjunto de estruturas classificadas pela identificação de estruturas similares com a função combinante empregada. Essa abordagem ajuda a fazer com que o mapeamento de maturação de vírus, em micrografos de elétron, seja rápido, confiável e fácil de descrever.

Enquanto a presente invenção foi descrita de acordo com composições e incorporações de preferência, fica entendido que certas substituições e alterações podem ser feitas sem que sejam consideradas separadas do espírito e escopo das seguintes reivindicações.

REIVINDICAÇÃO

1. Um método de identificação e caracterização de estruturas in micrógrafos de elétron, **caracterizado por**: estruturas selecionadas em uma primeira imagem, as estruturas com um primeiro tipo de formato deformado em uma primeira direção; transformar as estruturas selecionadas em um segundo tipo diferente do primeiro tipo de formato; usar as estruturas transformadas do segundo tipo de formato para formar a pluralidade de máscaras; identificação de uma nova estrutura em uma segunda imagem, a nova estrutura tendo o primeiro tipo de formato; deformação da estrutura do segundo tipo de formato em cada máscara na primeira direção; e determinação da máscara que é a preferida e que melhor combina com a nova estrutura.

2. O método de acordo com a reivindicação 1 é **caracterizado** aonde o método ainda engloba a deformação mecânica de uma partícula viral de um formato elíptico ou de um formato substancialmente circular.

3. O método de acordo com a reivindicação 2 é **caracterizado** aonde o método ainda engloba a rotação da partícula viral antes da deformação da partícula viral.

4. O método de acordo com a reivindicação 3 é **caracterizado** aonde o método ainda engloba a rotação da partícula viral subsequente a deformação da partícula viral.

5. O método de acordo com a reivindicação 1 é **caracterizado** aonde o método ainda engloba o teste de outras máscaras para verificar se a máscara preferida oferece a melhor combinação.

6. O método de acordo com a reivindicação 5 é **caracterizado** aonde o método ainda engloba a seleção de estruturas com base em parâmetros relacionados com o tamanho e o grau de formato elíptico das estruturas.

7. O método de acordo com a reivindicação 1 é **caracterizado** aonde o método ainda engloba a deformação de pluralidades de máscaras para um formato de uma nova estrutura de formato elíptico e teste de cada máscara

deformada para verificar se as máscaras preferidas combinam precisamente com a nova estrutura em formato elíptico na segunda imagem.

8. O método de acordo com a reivindicação 7 é **caracterizado** aonde o método ainda engloba o teste de outras direções além da primeira direção para verificar se a primeira direção de deformação da melhor máscara oferece a
5 melhor combinação para a nova estrutura em formato elíptico.

9. O método de acordo com a reivindicação 8 é **caracterizado** aonde o método ainda engloba a determinação de que as outras direções são menos precisas do que a primeira direção.

10. O método de acordo com a reivindicação 7 é **caracterizado** aonde o método ainda engloba a determinação de um estágio de maturidade e tipo de estrutura da nova estrutura de formato elíptico com base na máscara preferida.

11. O método de acordo com a reivindicação 1 é **caracterizado** aonde o método ainda engloba a filtragem de estruturas que são dispostas dentro de
15 uma distância radial da nova estrutura em formato elíptico.

REIVINDICAÇÕES - Retificadas

[recebido pelo International Bureau em 20 de outubro de 2007
(20.10.07);

reivindicações originais 1-2 retificadas; reivindicação original 7
20 cancelada;

reivindicações 8-11 re-numeradas como reivindicações 7-10;
permanecendo as reivindicações 3-6 inalteradas (2 páginas)]

Reivindicações:

25 1. Um método de identificação e caracterização de estruturas em micrógrafos de elétron, **caracterizado por**: estruturas selecionadas em uma primeira imagem (110), as estruturas com um primeiro tipo de formato deformado em uma primeira direção; transformar as estruturas selecionadas para um segundo
30 tipo de formato diferente do primeiro tipo de formato; usar as estruturas

transformadas do segundo tipo de formato para formar a pluralidades de máscaras; identificação de uma nova estrutura em uma segunda imagem (112), a nova estrutura tendo o primeiro tipo de formato; deformação da estrutura do segundo tipo de formato em cada máscara na primeira direção; determinação da máscara que é a preferida e que melhor combina com a nova estrutura; e a deformação de pluralidades de máscaras para um formato de uma nova estrutura de formato elíptico e teste de cada máscara deformada para verificar a máscara que oferece a melhor combinação para a nova estrutura em formato elíptico na segunda imagem.

2. O método de acordo com a reivindicação 1 é **caracterizado** aonde o método ainda engloba a deformação mecânica de uma partícula viral (110) de um formato elíptico ou de um formato substancialmente circular.

3. O método de acordo com a reivindicação 2 é **caracterizado** aonde o método ainda engloba a rotação da partícula viral (110) antes da deformação da partícula viral.

4. O método de acordo com a reivindicação 3 é **caracterizado** aonde o método ainda engloba a rotação da partícula viral (110) subsequente a deformação da partícula viral.

5. O método de acordo com a reivindicação 1 é **caracterizado** aonde o método ainda engloba o teste de outras máscaras para verificar se a máscara preferida oferece a melhor combinação.

6. O método de acordo com a reivindicação 5 é **caracterizado** aonde o método ainda engloba a seleção de estruturas com base em parâmetros relacionados com o tamanho e o grau de formato elíptico das estruturas.

7. O método de acordo com a reivindicação 1 é **caracterizado** aonde o método ainda engloba o teste de outras direções além da primeira direção para verificar se a primeira direção de deformação da melhor máscara oferece a melhor combinação para a nova estrutura em formato elíptico.

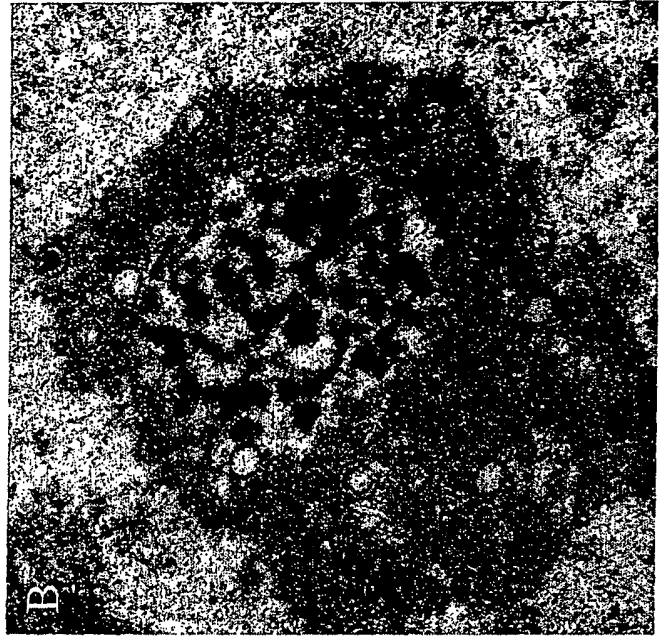
8. O método de acordo com a reivindicação 7 é **caracterizado** aonde o método ainda engloba a determinação de que as outras direções são menos precisas do que a primeira direção.

9. O método de acordo com a reivindicação 1 é **caracterizado** aonde o método ainda engloba a determinação de um estágio de maturidade e tipo de estrutura da nova estrutura de formato elíptico com base na máscara preferida.

5 10. O método de acordo com a reivindicação 1 é **caracterizado** aonde o método ainda engloba a filtragem de estruturas que são dispostas dentro de uma distância radial da nova estrutura em formato elíptico.

102

Fig. 1B



102

Fig. 1A

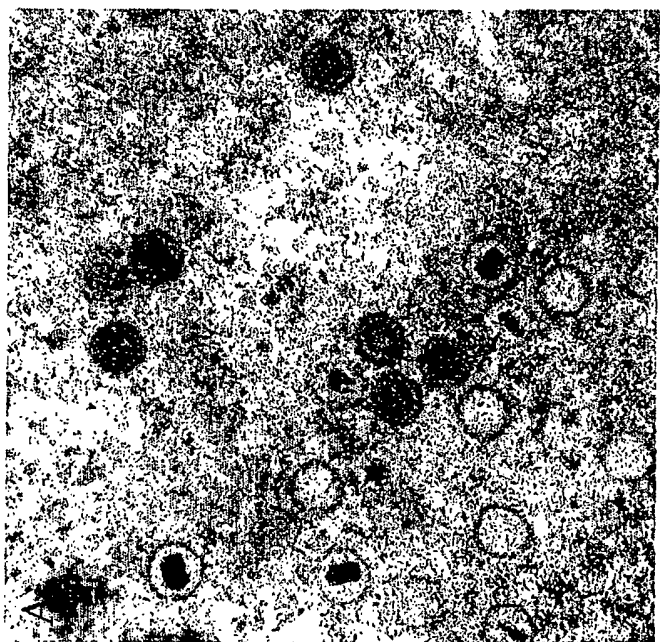


Fig. 2C

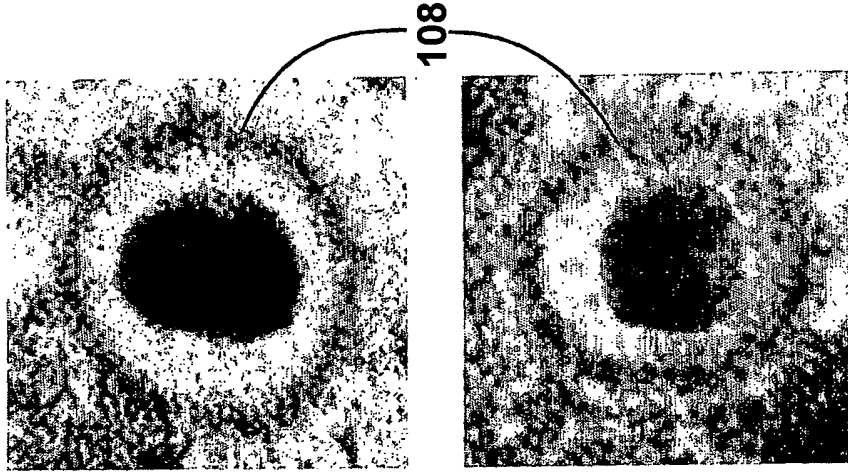


Fig. 2B

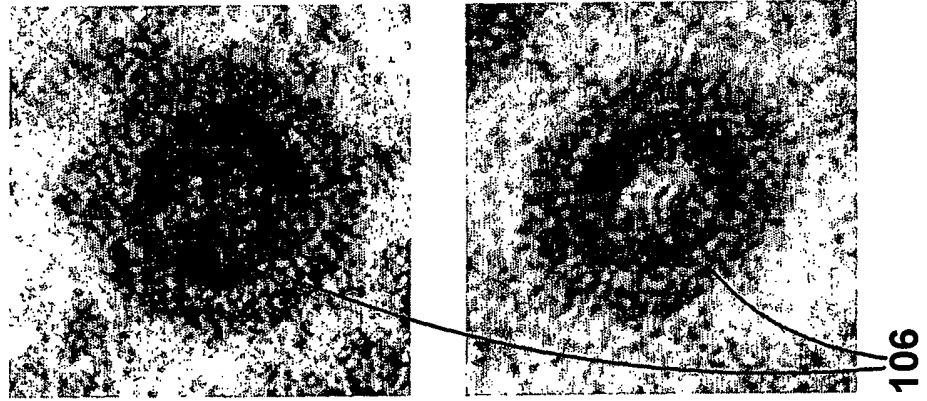


Fig. 2A

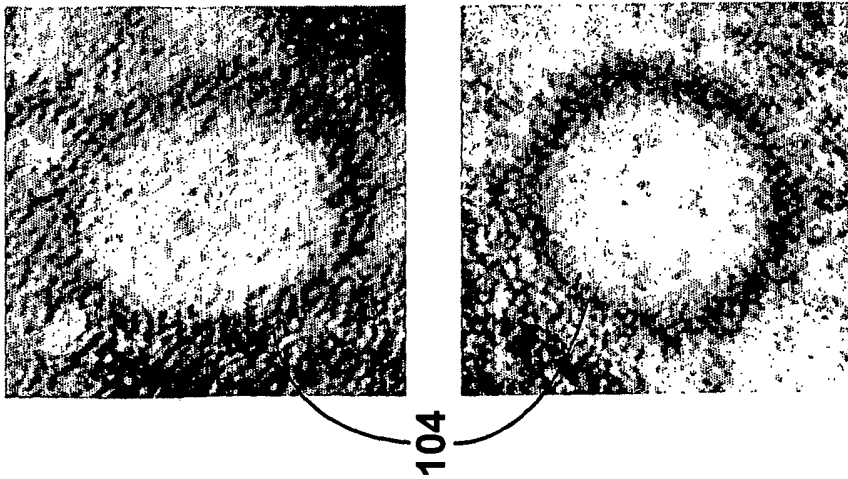


Fig. 3B

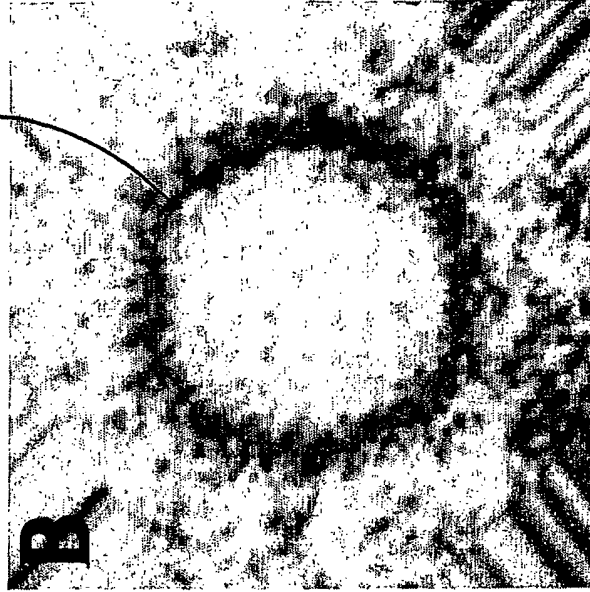


Fig. 3A

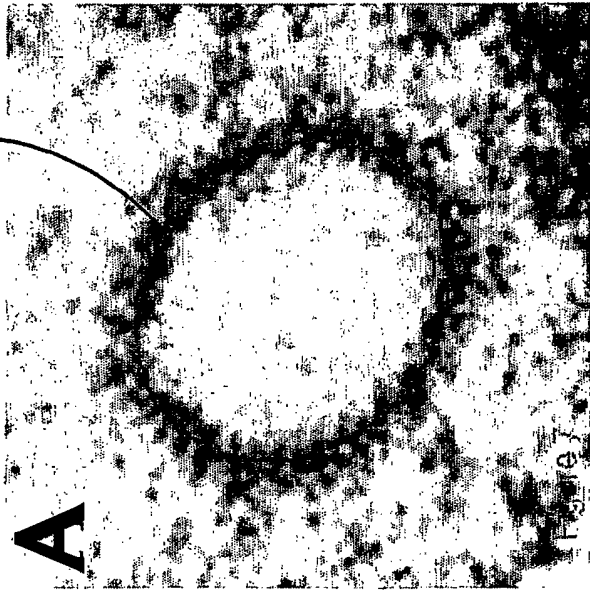


Fig. 4A

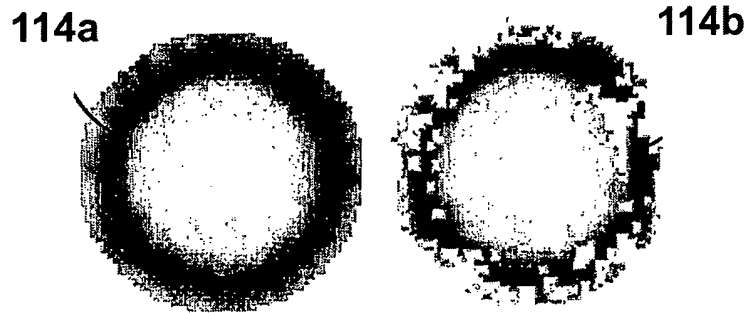


Fig. 4B

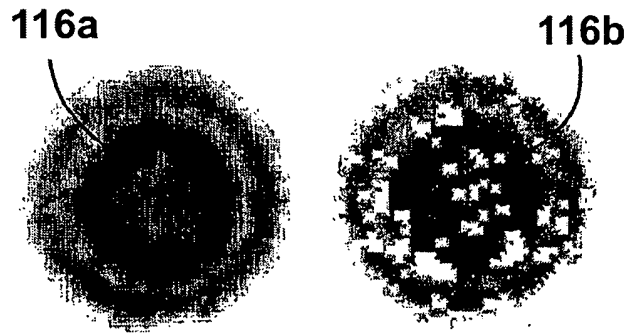


Fig. 4C

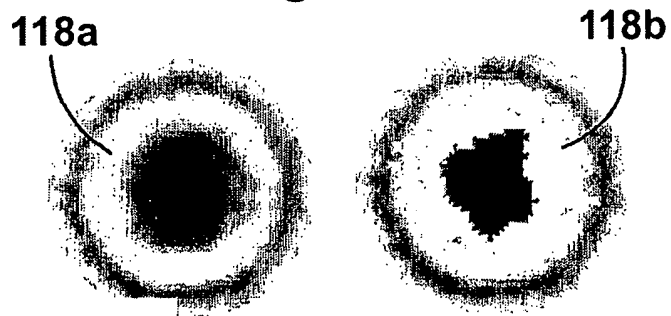


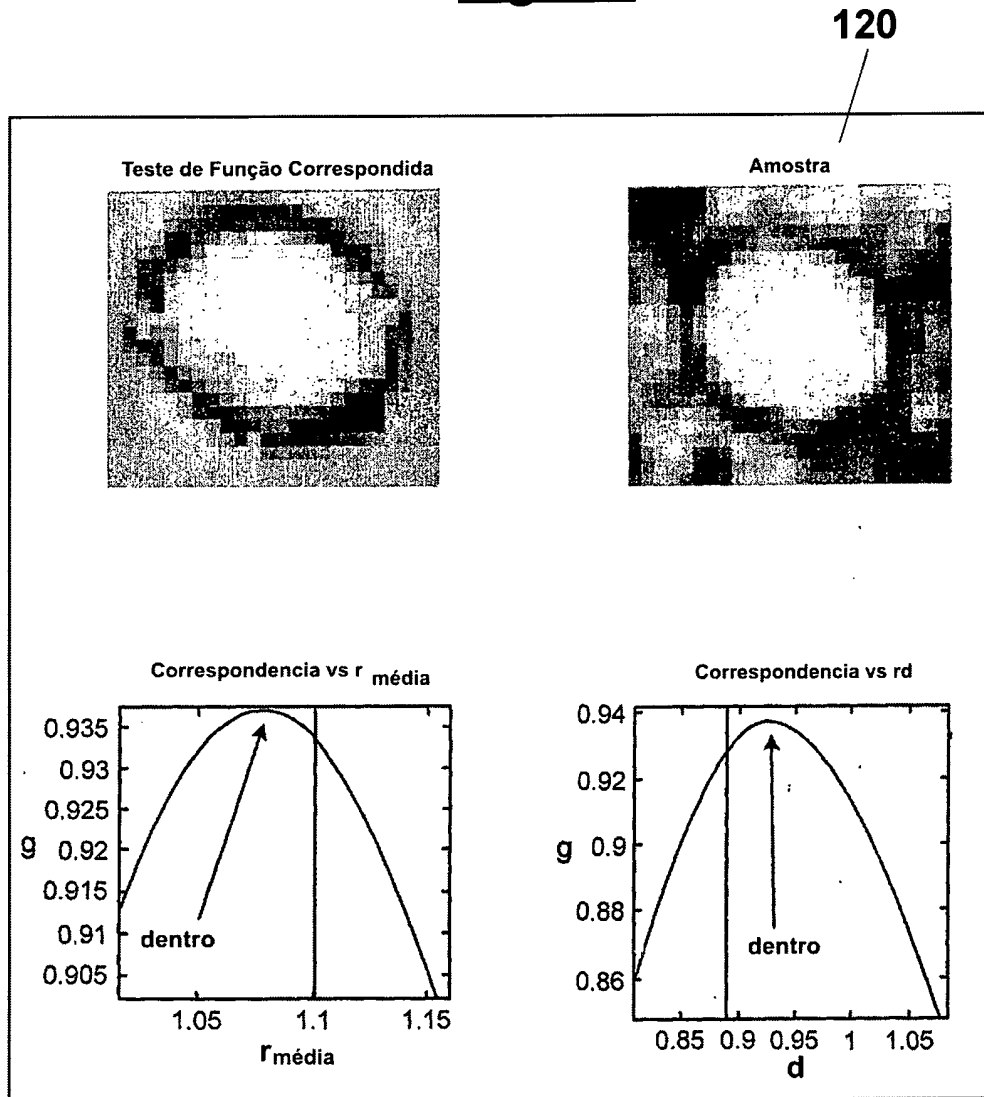
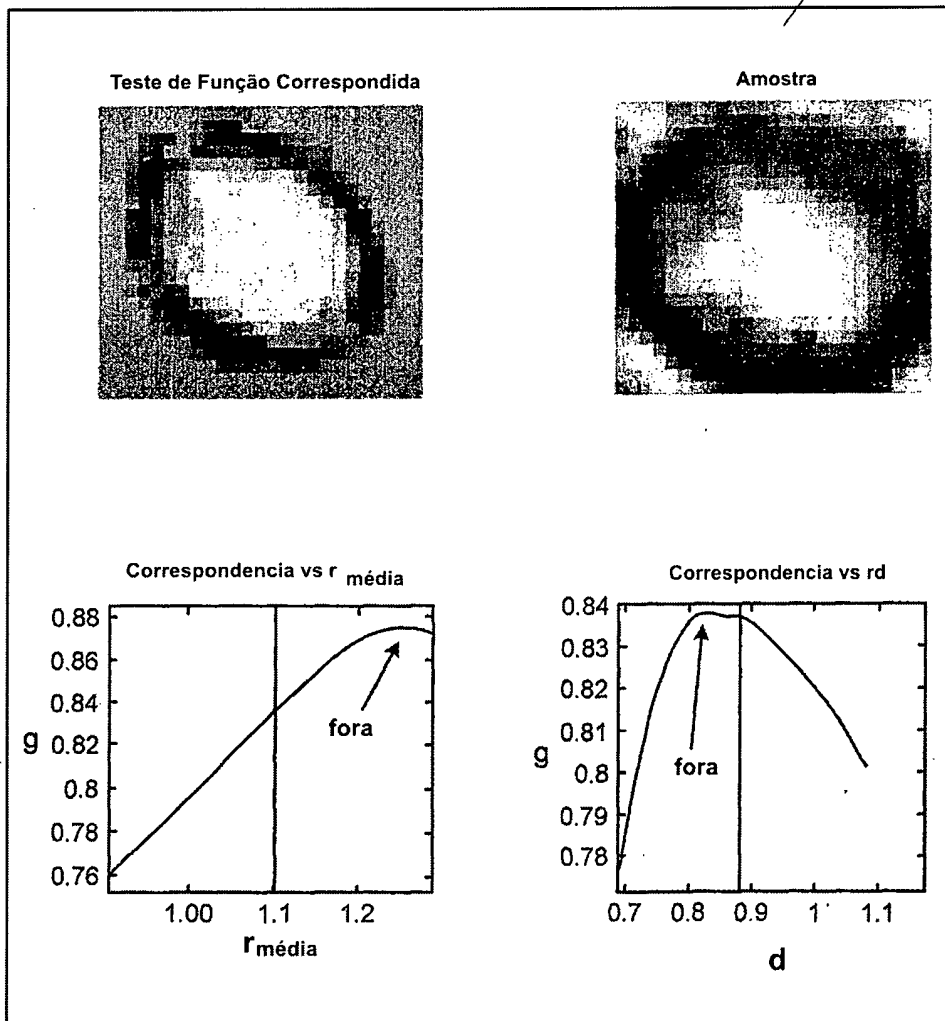
Fig. 5A

Fig. 5B

122



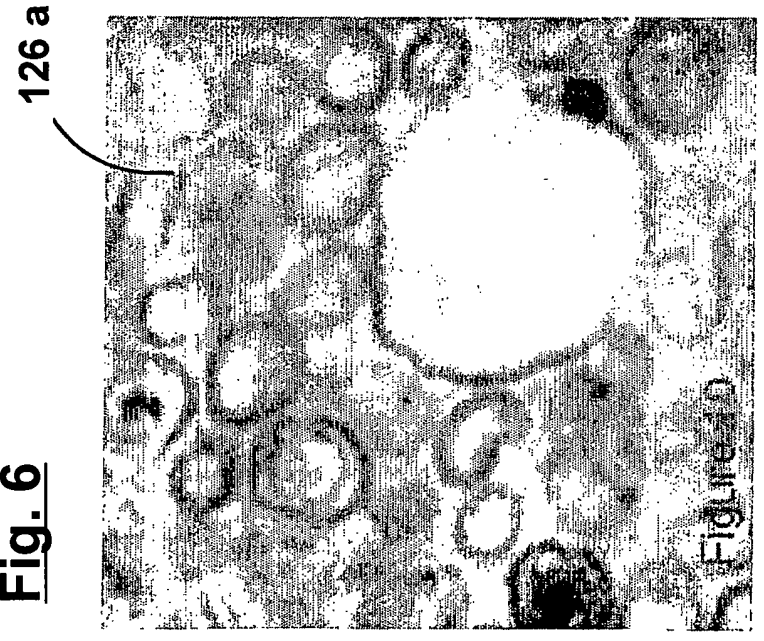
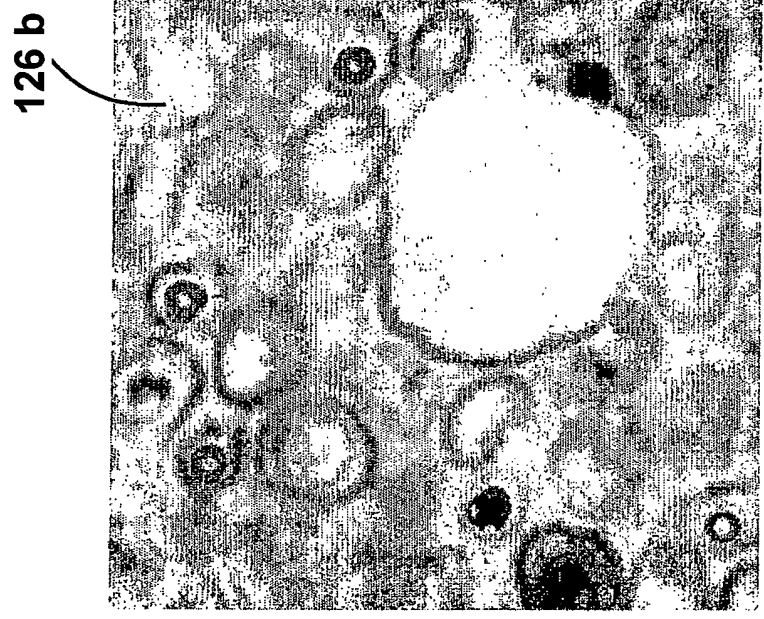


Fig. 6

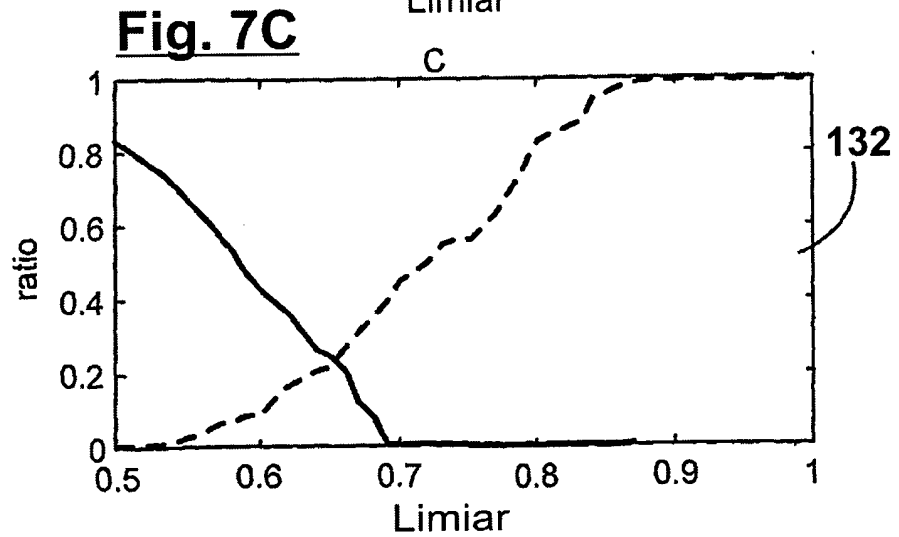
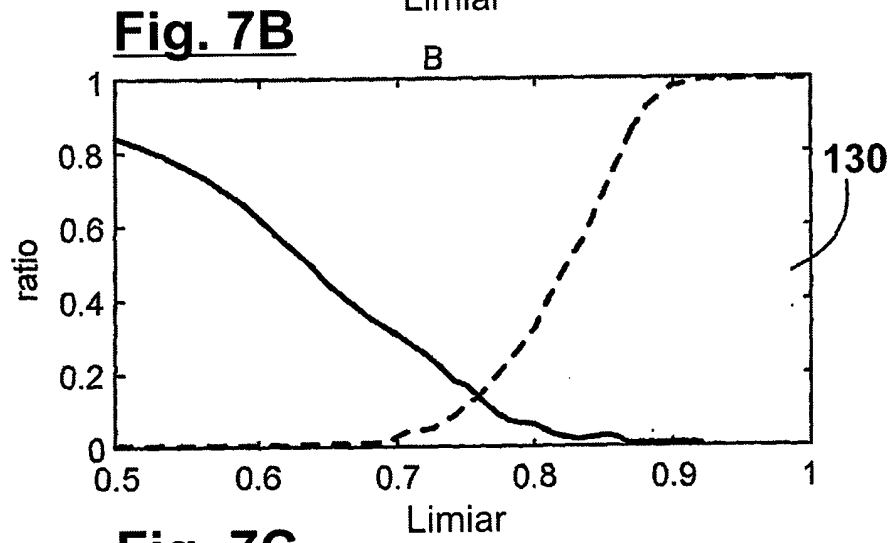
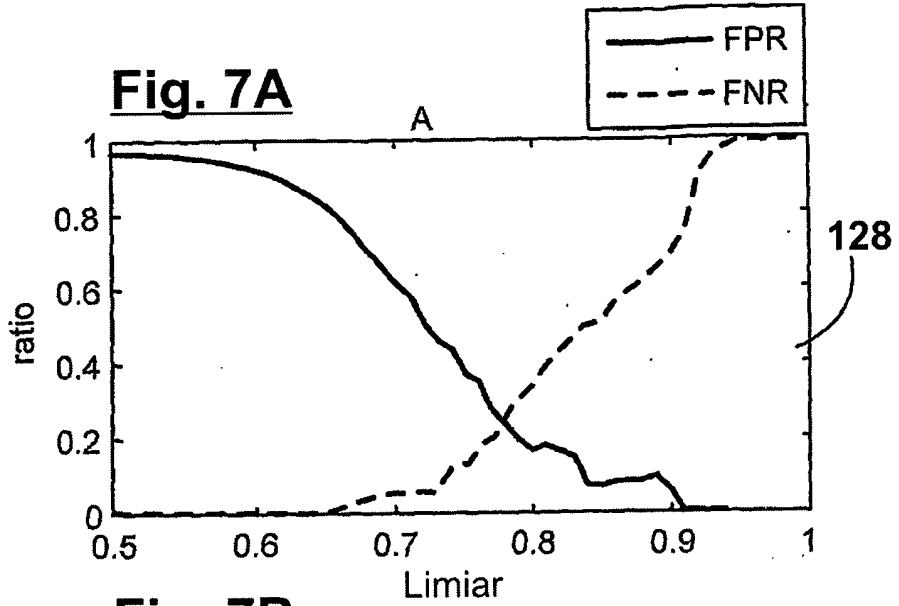
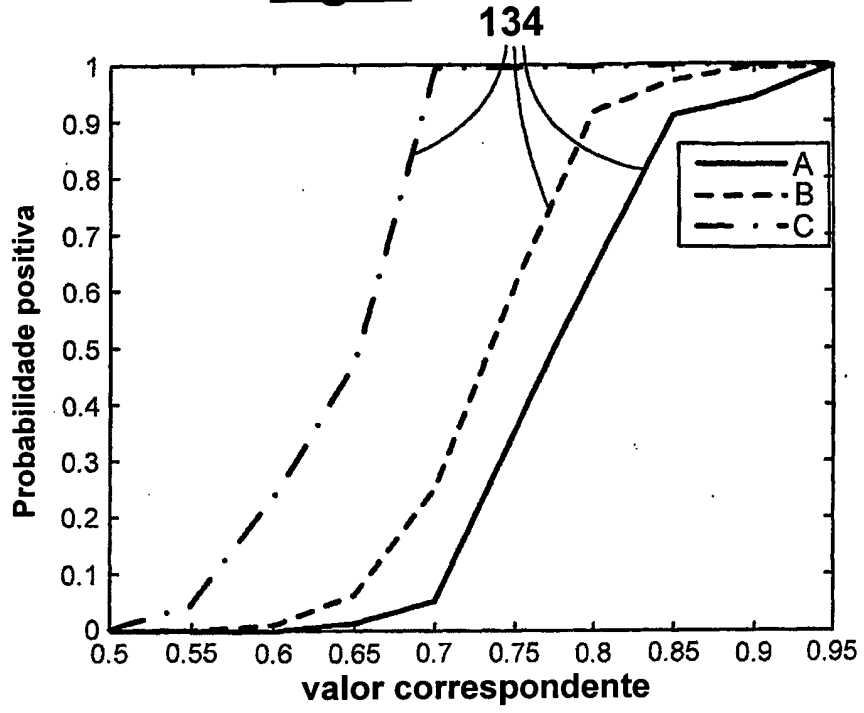


Fig. 8



9

Fig. 9

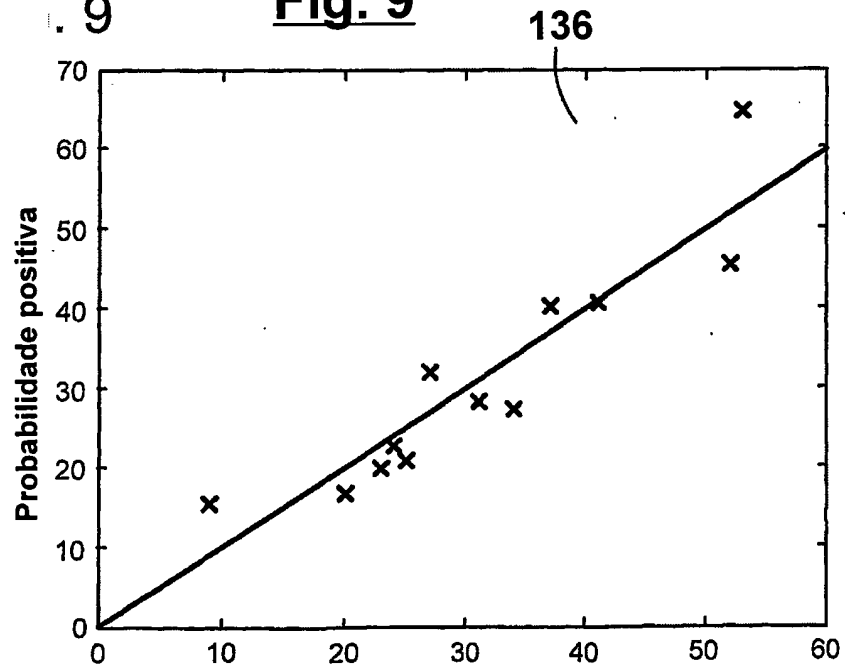
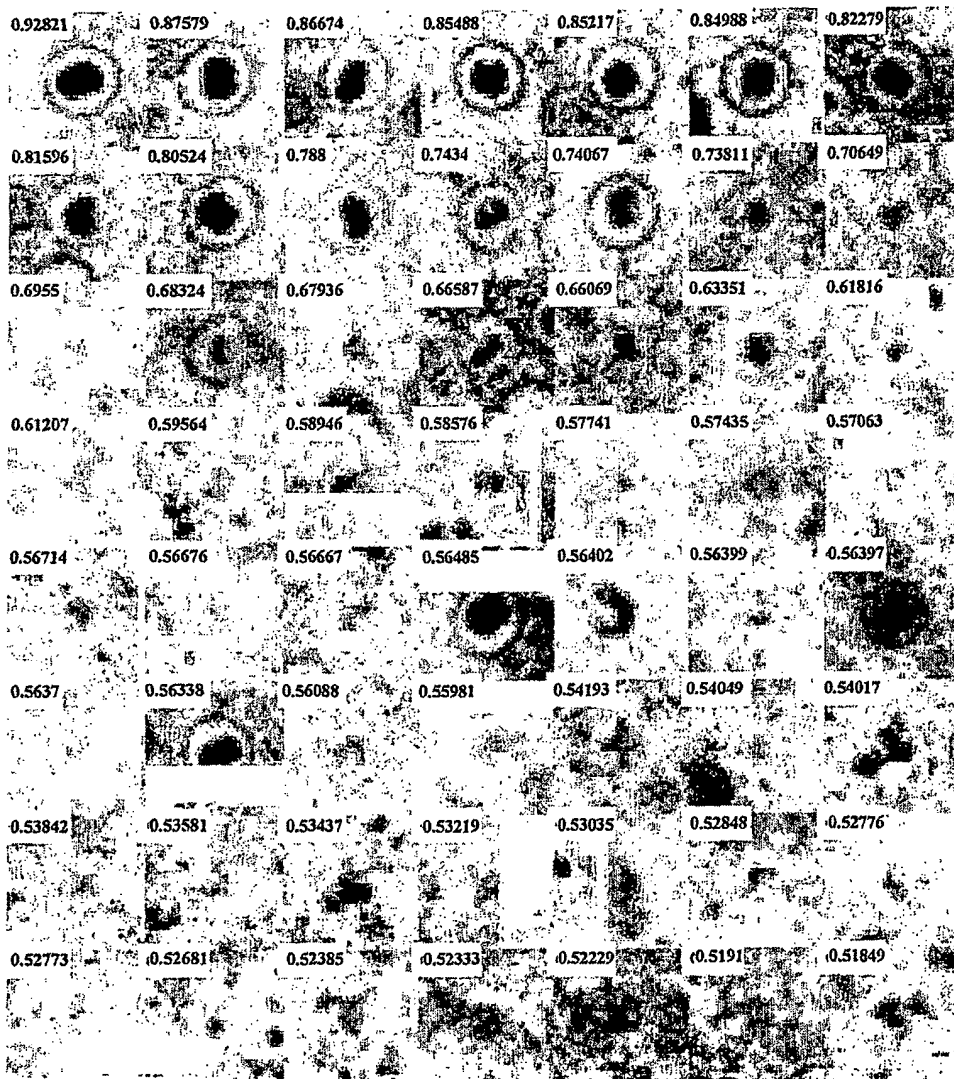


Fig. 10

138



RESUMO

"IDENTIFICAÇÃO E CLASSIFICAÇÃO DE PARTÍCULAS DE VÍRUS EM MICRÓGRAFOS DE ELÉTRON TEXTURADOS", o método é para a identificação e caracterização de estruturas em micrógrafos de elétron. São selecionadas estruturas em uma primeira imagem. As estruturas têm um primeiro tipo de formato deformado em uma primeira direção. As estruturas selecionadas são transformadas em um segundo tipo de formato diferente do primeiro tipo de formato. As estruturas transformadas do segundo tipo de formato são usadas para formar uma pluralidade de máscaras. É identificada uma nova estrutura em uma segunda imagem. A nova estrutura tem o primeiro tipo de formato. A estrutura do segundo tipo de formato, de cada máscara, é deformada na primeira direção. É determinada que máscara será a máscara preferida e que melhor combina com a nova estrutura.

15

20