

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 506 390**

51 Int. Cl.:

G01N 21/35 (2014.01)

G01N 21/64 (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Fecha de presentación y número de la solicitud europea: **03.11.2011 E 11778872 (9)**

97 Fecha y número de publicación de la concesión europea: **27.08.2014 EP 2635892**

54 Título: **Impresión digital espectroscópica de materias primas**

30 Prioridad:

05.11.2010 EP 10190193

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

13.10.2014

73 Titular/es:

**F. HOFFMANN-LA ROCHE AG (100.0%)
Grenzacherstrasse 124
4070 Basel, CH**

72 Inventor/es:

**CARDOSO-MENEZES, JOSÉ;
HAKEMEYER, CHRISTIAN;
JOSE, GLEDSON EMIDIO;
STRAUSS, ULRIKE y
WERZ, SILKE**

74 Agente/Representante:

ISERN JARA, Jorge

ES 2 506 390 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

Impresión digital espectroscópica de materias primas

- 5 El presente documento describe un método para la evaluación de componentes de material de cultivo con respecto al rendimiento del producto a la recepción del mismo y antes de y sin la necesidad de efectuar un cultivo de prueba.

Antecedentes de la invención

- 10 El mercado de los productos biofarmacéuticos recombinantes ha aumentado de manera constante desde principios de los ochenta, cuando la tecnología de ADN recombinante hizo posible expresar proteínas recombinantes en diferentes tipos de microorganismos como bacterias, levadura o células de mamífero. Desde entonces, estos productos de proteínas se han utilizado en un amplio intervalo de aplicaciones farmacéuticas y de diagnóstico.

- 15 A medida que aumenta la demanda de proteínas recombinantes, la necesidad de unos procesos de producción altamente eficaces y robustos es inminente. Uno de los factores más importantes e influyentes en la necesidad de unos procesos de producción robustos y reproducibles es la composición de los materiales de partida, como los medios de cultivo. La mayoría de medios de cultivo son mezclas complejas de, entre otras cosas, sales inorgánicas, azúcares, aminoácidos, vitaminas, ácidos orgánicos y tampones. En muchos casos, las materias primas complejas, no definidas de forma química, como los hidrolizados de proteína de origen vegetal o bacteriano se utilizan para promover el crecimiento de células y la producción de proteínas.

- 20 Habitualmente, las materias primas se suministran como mezclas en polvo y después se disuelven en agua para formar el medio de cultivo. En muchos casos se puede observar una variabilidad significativa de unos lotes a otros en hidrolizados de proteína no definidos químicamente y también en mezclas de medios basales químicamente definidos, conduciendo a grandes variaciones en el rendimiento de proteínas terapéuticas producidas de forma recombinante.

- 25 Las técnicas de impresión digital espectroscópica rápida, como las espectroscopias de infrarrojo cercano, de infrarrojo medio, de Raman o de fluorescencia bidimensional, son relativamente baratas y resultan muy adecuadas para analizar mezclas complejas. Estos métodos generan cantidades muy grandes de datos altamente dimensionales que solo se pueden manejar por métodos quimiométricos como el análisis de componentes principales (PCA) o el modelado por mínimos cuadrados parciales (PLS). La combinación de métodos espectroscópicos complejos y quimiométrica se utiliza de manera habitual en las pruebas de identidad de materias primas o como herramienta para la clasificación de materias primas.

- 30 La utilización del análisis de componentes principales (PCA) y mínimos cuadrados parciales (PLS) para el procesamiento y modelado de datos complejos ha sido analizada por Næs, T., et al., (Næs, T., et al., NIR Publications, (2002)). En el documento WO 2009/086083 se describe un método para la organización jerárquica de datos utilizando PLS. En el documento WO 2008/146059 se describe un analizador y un método para determinar la importancia relativa de fracciones de mezclas biológicas. En el documento WO 2009/061326 se describe la evaluación de materiales cromatográficos.

- 35 En el documento US2009/0306932 se describe un método rápido de clasificación para conjuntos de datos multivariados. El análisis de datos espectrales para la selección de un modelo de calibración se describe en el documento EP 2 128 599. En el documento US 5.498.875 se describe un procesamiento de señales para el análisis químico de muestras. Un método para clasificar materiales científicos como los materiales de silicato, materiales poliméricos y/o nanomateriales se describe en el documento US2008/0177481. En el documento US2010/0129857 se describen métodos para el aislamiento y la identificación de microorganismos.

- 50 Sumario de la invención

- 55 Se ha hallado que el rendimiento de los procesos de producción de proteínas recombinantes se pueden predecir basándose en la combinación de espectros NIR y de fluorescencia 2D de componentes de medios, como hidrolizados de proteína y/o preparaciones de medios definidos químicamente los cuales se utilizan como componentes de un medio de cultivo complejo.

- 60 Un aspecto, tal como aquí se señala, es un método para la selección de lotes de componentes de medios de cultivo para utilizarlos en el cultivo de una célula de mamífero que exprese una proteína de interés en la que se utilicen al menos dos componentes diferentes en el cultivo, empleando para dicha selección datos espectrales fusionados de dos técnicas espectroscópicas distintas.

- 65 El método de acuerdo con la presente invención para la selección de lotes de componentes de medios de cultivo que se van a utilizar en el cultivo de una célula de mamífero que exprese una proteína de interés en la que se utilicen al menos dos componentes diferentes en el cultivo comprende las siguientes etapas:

- 5 a) proporcionar espectros de distintos lotes de un primer componente obtenido con un primer método espectroscópico seleccionado a partir de espectroscopia NIR y espectroscopia MIR, y espectros de un segundo componente obtenido mediante espectroscopia de fluorescencia 2D como un segundo método espectroscópico y el rendimiento de sobrenadante de cultivo de la proteína de interés obtenido en un cultivo utilizando combinaciones de estos lotes diferentes del primer y el segundo componente,
- 10 b) identificar una relación de espectros fusionados de las dos técnicas espectroscópicas diferentes tras computar los resultados del análisis PCA de espectros con el rendimiento del cultivo,
- 15 c) proporcionar un espectro de un lote adicional del primer componente obtenido con el primer método espectroscópico y lote un espectro de un lote adicional del segundo componente obtenido con el segundo método espectroscópico, y
- d) seleccionar la combinación del primer componente proporcionado y el segundo componente proporcionado si el rendimiento de sobrenadante predicho para el cultivo basado en la relación de espectros fusionados tras contabilizar los resultados de los espectros PCA identificados en b) se halla dentro del 10 % más o menos del rendimiento medio proporcionado en a).

15 En una realización preferente el procesamiento de los espectros comprende la eliminación de las regiones de absorción de agua y la aplicación de una corrección multiplicativa de la dispersión, y/o el filtrado comprende un filtrado Savitzky-Golay.

20 En una realización preferente los patrones identificativos del espectro se hallan por análisis de componentes principales. En una realización preferente el análisis de componentes principales es un análisis de componentes principales desdoblado. En una realización preferente el desdoblamiento preserva la información del primer modo (muestra). En una realización preferente el alisamiento Savitzky- Golay se hace con una ventana de 19 puntos y un polinomio de segundo grado. En una realización preferente los datos son un promedio, y el número óptimo de componentes principales se elige utilizando el método de validación cruzada dejando uno fuera.

30 En una realización preferente el procesamiento comprende la exclusión de las regiones de dispersión y la interpolación de los puntos eliminados. En una realización preferente los espectros finales se componen de un intervalo de longitud de onda de emisión de 290 nm a 594 nm y un intervalo de longitud de onda de excitación de 230 nm a 575 nm.

35 En una realización preferente la identificación de una relación entre espectros fusionados y comprimidos con resultados del PCA, con un rendimiento del cultivo al momento de su cosecha se efectúa mediante análisis de mínimos cuadrados parciales.

En una realización preferente los espectros NIR se recopilan de la región del número de onda de 4.784 cm^{-1} a 8.936 cm^{-1} .

40 En una realización preferente la dimensionalidad espectral se reduce de 1.039 números de onda a 3 componentes principales.

En una realización preferente la proteína de interés es un anticuerpo, o un fragmento de anticuerpo, o un conjugado de anticuerpo.

45 Descripción detallada de realizaciones preferentes de la invención

50 Se ha hallado que el rendimiento de los procesos de producción de proteínas recombinantes se puede predecir basándose en la información combinada contenida en espectros NIR y de fluorescencia 2D de componentes de medios, como hidrolizados de proteína y/o preparaciones de medios químicamente definidas que se utilicen como componentes de un medio de cultivo complejo.

55 El presente documento describe un método en el que se combinan espectros de dos técnicas distintas de espectroscopia (ortogonales)– tras procesarlos para convertirlos en aditivos mediante reducción variable a resultados del análisis de componentes principales (PCA) – obtenidos en dos componentes de medios utilizados en la fermentación de productos biofarmacéuticos recombinantes y modelos de dichos espectros transformados (entradas) se utilizan para predecir los rendimientos en el momento de su cosecha (resultado) de los cultivos de productos biofarmacéuticos basados en mezclas de componentes de medios analizados con variabilidad de unos lotes a otros en términos de su distinto rendimiento de fermentación.

60 Empleando distintas espectroscopias (ortogonales) en combinación con métodos PCA (a fin de garantizar su aditividad) y produciendo modelos de procesos del efecto de dichas mezclas de medios de cultivo sobre rendimientos en el momento de cosecha de la fermentación principal se establece una capacidad predictiva que permite seleccionar lotes de medios de cada materia prima y/o formular mezclas que sirvan mejor a los objetivos del proceso.

65

Diferentes lotes o tandas de componentes individuales que forman un medio de cultivo completo varían ligeramente en su composición detallada pero se encuentran no obstante dentro de la especificación aportada por el fabricante. En algunos casos, es posible rastrear esta variabilidad a ingredientes aislados, pero lo más habitual es que la variabilidad de unos lotes a otros no se pueda detectar por medios analíticos. Para la evaluación de la influencia de distintos lotes de componentes individuales en rendimiento de producto se puede realizar de manera repetida un cultivo comparativo de la misma línea de células mamíferas.

El presente documento describe 56 cultivos en los que nueve lotes distintos de un hidrolizado de proteína de soja, dos mezclas de dos lotes distintos de hidrolizado de proteína de soja, cinco lotes de un hidrolizado de proteína de arroz y seis lotes de un polvo medio básico definido químicamente se emplearon en la fermentación y el medio de alimentación, respectivamente.

A fin de evaluar la influencia de diversos lotes de hidrolizado de proteína de soja respecto al rendimiento de producto, se realizaron cultivos comparables en los que los mismos lotes de un medio básico químicamente definido y un hidrolizado de proteína de arroz se utilizaron en la fermentación y los medios de alimentación. Los resultados se pueden agrupar de acuerdo con los distintos lotes de hidrolizado de proteína utilizados. El rendimiento de los distintos lotes se evaluó con arreglo al rendimiento del producto a unos valores promedio similares de densidad de la celda de inoculación (ICD) (Tabla 1).

Tabla 1.

lote	Nº de lote de hidrolizado de proteína de soja	Nº de lote de medio básico químicamente definido	Nº de lote de hidrolizado de proteína de arroz	ICD	producto a 330 h [mg/l]
D45KD11	1	1	1	5,7	1319
D45KD12				5,3	1234
D45KD13				5,6	1305
D45KD22	2			5,3	1023
D45KD23				5,1	1070
D45KD31	3			4,8	1008
D45KD32				4,9	991
D45KD33				5,3	978

Los resultados obtenidos en un segundo grupo de cultivos se enumeran en la Tabla 2.

Tabla 2.

lote	Nº de lote de hidrolizado de proteína de soja	Nº de lote de medio básico químicamente definido	Nº de lote de hidrolizado de proteína de arroz	ICD	producto a 330 h [mg/l]
D52KD11	1	2	2	6,1	1434
D52KD12				5,0	1411
D52KD13				5,6	1459
D52KD21	4			5,0	1213
D52KD22				5,3	1243
D52KD23				5,4	1163
D55KD11	5			5,0	1409
D55KD12				5,4	1426
D55KD13				5,7	1430
D55KD21	2			6,8	1263
D55KD22				6,8	1256
D55KD23				6,8	1278
D55KD31	6			6,1	1269
D55KD32				6,1	1262
D55KD33				5,8	1265

Se puede ver que distintos lotes de los componentes individuales dan como resultado distintos rendimientos del producto. En esta serie de cultivos también se utilizaron distintos valores ICD promedio. Aunque tenían bajos niveles ICD, los cultivos que utilizaron el lote 1 y el lote 5 arrojaron rendimientos de producto significativamente más altos que los que tenían unos valores ICD más altos (el lote 3 y el lote 6). Así pues, distintos lotes de hidrolizados de proteína de soja dan como resultado un rendimiento distinto de producto.

De forma análoga se puede evaluar la influencia del hidrolizado de proteína de arroz en el rendimiento del proceso (Tabla 3).

5

Tabla 3.

lote	Nº de lote de hidrolizado de proteína de soja	Nº de lote de medio básico químicamente definido	Nº de lote de hidrolizado de proteína de arroz	ICD	producto a 330 h [mg/l]		
D61KD11	3	3	2	5,9	1132		
D61KD12				6,0	1085		
D61KD13				5,3	1101		
D61KD21			3	3	3	6,1	1062
D61KD22						6,1	1056
D61KD23						5,6	1043

10

Se realizaron seis cultivos y pueden agruparse de acuerdo con los distintos lotes de hidrolizado de proteína de arroz utilizados en cada uno de ellos. El rendimiento de los distintos lotes de hidrolizado de proteína de arroz se puede evaluar basándose en el rendimiento medio de producto. Ambos grupos, esto es, los lotes de hidrolizado de proteína de arroz, tienen valores ICD similares.

15

A fin de evaluar la influencia del medio básico definido químicamente sobre el rendimiento de producto, se pueden realizar cultivos con los mismos lotes de hidrolizado de proteína de soja e hidrolizado de proteína de arroz en la formulación inicial de medios de fermentación y medios de alimentación. Se llevaron a cabo tres series de experimentos (Tablas 4, 5 y 6).

20

La primera serie comprendía seis cultivos que tenían el lote de hidrolizado de proteína de soja 3 (como se muestra en la tabla 3) y el lote de hidrolizado de proteína de arroz 2 (como se muestra en la tabla 2) en la fermentación y los medios de alimentación. Los cultivos se agruparon de acuerdo con el lote medio básico químicamente definido utilizado. El rendimiento de los distintos lotes medios básicos químicamente definidos se evaluó basándose en el rendimiento de producto. Hay una ligera diferencia entre los dos grupos tanto en el ICD promedio y el rendimiento promedio de producto. Con un menor ICD se puede obtener una menor formación de producto. Así pues, los lotes de medio básico químicamente definidos tienen escaso o ningún efecto en el rendimiento del producto.

25

Tabla 4.

lote	Nº de lote de hidrolizado de proteína de soja	Nº de lote de medio básico químicamente definido	Nº de lote de hidrolizado de proteína de arroz	ICD	producto a 330 h [mg/l]	
D55KD21	3	2	2	6,8	1263	
D55KD22				6,8	1256	
D55KD23				6,8	1278	
D61KD11		3		3	5,9	1132
D61KD12					6,0	1085
D61KD13					5,3	1101

30

La segunda serie comprendía seis cultivos que empleaban el lote de hidrolizado de proteína de soja 1 (como se muestra en la tabla 2) en la formulación inicial de los medios de fermentación y los medios de alimentación. Los experimentos se agruparon de acuerdo con el lote medio básico químicamente definido utilizado. No se encontraron diferencias ICD significativas. Así pues, las diferencias en el rendimiento del producto se deben a diferencias en los lotes de medio básico químicamente definidos empleados.

Tabla 5.

lote	Nº de lote de hidrolizado de proteína de soja	Nº de lote de medio básico químicamente definido	ICD	producto a 330 h [mg/l]
D45KD11	1	1	5,7	1319
D45KD12			5,3	1234
D45KD13			5,6	1205
D52KD11		2	6,1	1434
D52KD12			5,0	1411
D52KD13			5,6	1459

35

La tercera serie comprendía cinco cultivos que emplearon el lote de hidrolizado de proteína de soja 2 en la formulación inicial de los medios de fermentación y los medios de alimentación. Los experimentos se agruparon de

acuerdo con el lote medio básico químicamente definido utilizado. Hay una diferencia entre los dos grupos tanto en el ICD utilizado como en la concentración de producto obtenida.

5

Tabla 6.

lote	Nº de lote de hidrolizado de proteína de soja	Nº de lote de medio básico químicamente definido	ICD	producto a 330 h [mg/l]
D45KD22	2	1	5,3	1023
D45KD23			5,1	1070
D73KD11		4	4,9	1062
D73KD12			4,3	1112
D73KD13			4,4	1121

De lo anterior puede deducirse que existe una necesidad de caracterización de los lotes de materias primas y una necesidad de ofrecer un método en el que los datos obtenidos puedan utilizarse para predecir qué lotes de materias primas producen rendimientos más altos de producto sin la necesidad de realizar experimentos de fermentación.

10

Dentro del ámbito de las reivindicaciones NIR y/o MIR, y de fluorescencia 2D se pueden adquirir espectros de todos los lotes de los tres componentes distintos de medios de cultivo. A partir de ahí se puede realizar un análisis de espectros con métodos quimiométricos establecidos. Una manera novedosa de analizar la información espectral obtenida con estas fuentes diferentes se describe en el presente documento y puede utilizarse con fines de modelado predictivo.

15

Los espectros NIR de los lotes de materias primas se obtuvieron como triplicados en distintos periodos de tiempo. Para las muestras heterogéneas de polvo y desecadas los espectros NIR varían entre las réplicas. Dichas réplicas límite pueden eliminarse con arreglo a su ubicación relativa en el área de trazado de los resultados PCA (distancia Euclídea).

20

Los espectros NIR de 18 lotes de hidrolizado de proteína de soja, 12 lotes de hidrolizado de proteína de arroz y 14 lotes de medio básico químicamente definido se seleccionaron de entre todas las mediciones proporcionadas. Se recopilaban espectros NIR de entre 4.784 cm^{-1} y 8.936 cm^{-1} . Esta región espectral no contiene regiones ruidosas. Los fuertes cambios observados en la línea de base se deben a una ligera dispersión asociada con diferentes lotes de materia prima que poseen diferencias en las distribuciones del tamaño de partículas (granularidad). El análisis de espectros en bruto sin una corrección de línea de base permite centrarse en variaciones principalmente ocasionadas por efectos físicos. El análisis PCA de espectros en bruto se realizó por cada materia prima de forma aislada.

25

La figura 1 muestra la distribución de los distintos lotes de hidrolizado de proteína de soja en un espacio bidimensional construido a través de PCA basado en los espectros NIR originales, capturando un 94 % de la varianza de los espectros NIR. La dimensionalidad espectral se redujo de 1.039 números de onda a 3 componentes principales significativos. Los lotes que ofrecen un alto rendimiento de producto no pueden discriminarse respecto de que ofrecen un bajo rendimiento de producto basándose en este análisis. Además, la granularidad (tal como se ve en distintas líneas de base de diversos espectros NIR, figura 2) y el contenido de humedad (al igual que las mediciones Karl Fischer) de las muestras son también diferentes, haciendo que un agrupamiento de los lotes de acuerdo con cualquier propiedad única sea algo muy difícil.

35

La figura 3 muestra cómo los lotes analizados de hidrolizado de proteína de arroz se distribuyen en un espacio bidimensional construido mediante PCA basándose en los espectros NIR originales, capturando un 92 % de la varianza de los espectros NIR. En cuanto al hidrolizado de proteína de soja, los lotes que ofrecen un alto rendimiento de producto no pueden discriminarse respecto de que ofrecen un bajo rendimiento de producto basándose únicamente en este análisis. De nuevo, la granularidad y humedad de las muestras varía de un lote a otro afectando al agrupamiento.

40

La figura 4 muestra la distribución de los lotes del medio básico químicamente definido en un espacio bidimensional construido mediante PCA basándose en los espectros NIR originales, capturando un 98 % de la varianza de los espectros NIR. En cuanto a los lotes de hidrolizado de proteína de soja y arroz, los lotes que ofrecen un alto rendimiento de producto no pueden discriminarse respecto de que ofrecen un bajo rendimiento de producto basándose únicamente en este análisis.

45

Los tres componentes de medios de cultivo analizados muestran una variabilidad significativa de un lote a otro en la granularidad y el contenido de humedad, como puede verse por los espectros NIR obtenidos. NIR es muy sensible a estos dos factores. Además, ambos factores predominan por encima de otras diferencias menores pero aún así significativas que podrían presentarse en la composición química. Antes del análisis PCA ha de eliminarse la información física mediante un pre-procesado de espectros.

55

El agua se absorbe de manera muy fuerte en la región NIR especialmente en el intervalo de 6.900 cm^{-1} a 7.150 cm^{-1} y de 5.160 cm^{-1} a 5.270 cm^{-1} . Estas regiones de absorción se ocasionan por el primer sobretono de la banda de alargamiento O-H y la combinación de las bandas de alargamiento O-H y de flexión O-H, respectivamente. Las regiones de absorción de agua pueden eliminarse. Además, el desplazamiento de la línea de base puede eliminarse aplicando corrección de dispersión multiplicativa (MSC). A fin de mejorar la varianza entre las muestras, se puede aplicar el filtrado Savitzky-Golay y el método de alisamiento, y los espectros se pueden transformar en su primer derivado (ventana de 25 puntos).

El análisis PCA se efectuó sobre espectros previamente pre-procesados de hidrolizados de proteína de soja (Figura 5). Casi todos los lotes de rendimiento muy bueno a bueno en cuanto a rendimiento de proceso se agrupan en el lado izquierdo del trazado PCA (valores PC1 de resultado negativo). Por el contrario, el lote 4, que parece tener un mal rendimiento, ocupa el espacio a mano derecha del trazado.

El análisis PCA se realizó en espectros previamente pre-procesados de hidrolizados de proteína de arroz (Figura 6). Los lotes que arrojan rendimientos muy similares se agrupan juntos, mostrando así que el PCA de espectros pre-procesados es adecuado y que ya hay alguna variabilidad de unos lotes a otros que puede rastrearse hasta la composición química de la materia prima de este componente, que no guarda relación con la granularidad o el nivel de humedad.

El análisis de PCA de los espectros pre-procesados de los medios básicos químicamente definidos (Figura 7) muestra que en general todos los lotes de rendimiento muy bueno a bueno se agrupan en el lado izquierdo del trazado PCA (valores PC1 de resultado negativo). Por el contrario, el lote 3, que parece tener un mal rendimiento, ocupa el espacio a mano derecha del trazado. Esos resultados son comparables con los resultados obtenidos para los lotes de hidrolizado de proteína.

Además de los espectros NIR, de acuerdo con los espectros de excitación-emisión de fluorescencia de la invención (EEM), preferentemente adquiridos de diferentes materias primas de fermentación solubles en agua, deben ser analizados. Se puede establecer un conjunto de datos tridireccional, con longitudes de onda de excitación a lo largo del eje x, longitudes de onda de emisión a lo largo del eje y, e intensidad a lo largo del eje z. La figura 8 muestra una panorámica EEM de fluorescencia de muestras de un lote de hidrolizado de proteína.

Se obtuvieron espectros de fluorescencia 2D de 19 lotes de hidrolizado de proteína de soja, de 12 lotes de hidrolizado de proteína de arroz y de 14 lotes de medio básico químicamente definido. Los espectros se obtuvieron usando longitudes de onda de excitación de 200 nm a 600 nm, con intervalos de 5 nm y longitudes de onda de emisión también de 200 nm a 600 nm, con intervalos de 2 nm, arrojando un total de 81 longitudes de onda de emisión y 201 de emisión.

Con el fin de permitir una predicción del rendimiento de cultivo basada en el análisis de las materias primas, se puede generar un conjunto de datos tridireccional para cada una de las materias primas a partir de las matrices individuales.

Un espectro EEM típico puede verse influenciado por los efectos de dispersión de Rayleigh y Raman, los cuales afectan al contenido de la información de la panorámica de fluorescencia. Para resolver el efecto Rayleigh se pueden utilizar diversas estrategias y técnicas:

- poner a cero las longitudes de onda de emisión inferiores a las de excitación;
- insertar valores nulos en la región de dispersión;
- excluir la región de dispersión e interpolar los puntos eliminados; o
- restar los espectros de fondo.

Se ha hallado que excluir la región de dispersión y la interpolación de los puntos eliminados es más adecuado en el método como se describe en el presente documento. El algoritmo EEMscat de Matlab® puede emplearse por tanto. Este algoritmo puede descargarse gratuitamente desde el sitio web:

http://www.models.kvl.dk/source/EEM_correction/. Con este procedimiento la dispersión puede eliminarse por completo. El espectro también muestra un ruido pronunciado a lo largo de todo el eje de emisión en la primera longitud de onda de excitación. Esta región (200 nm a 225 nm) fue excluida de los espectros, así como las longitudes de onda de emisión no informativas (200 nm a 315 nm y 596 nm a 600 nm) y las longitudes de onda de excitación (580 nm a 600 nm). El espectro resultante se muestra en la figura 9.

Los espectros finales de hidrolizado de proteína de soja están compuestos por el intervalo de longitud de onda de emisión de 320 nm a 594 nm y el intervalo de longitud de onda de excitación de 230 nm a 575 nm, dando como resultado un conjunto de $19 \times 138 \times 70$ elementos. El mismo procedimiento puede seguirse para los hidrolizados de proteína de arroz y los conjuntos de datos de medio básico químicamente definidos. Así pues, los espectros finales de hidrolizado de proteína de arroz se componen del intervalo de longitud de onda de emisión y excitación de 290 nm a 594 nm y 230 nm a 550 nm, respectivamente, dando como resultado un conjunto de $12 \times 153 \times 65$ elementos. Los espectros finales de medio básico químicamente definido comprende el intervalo de longitud de onda de emisión

de 290 nm a 594 nm y el intervalo de longitud de onda de excitación de 230 nm a 550 nm, dando como resultado un conjunto de 14x162x60 elementos.

5 En conclusión, se puede realizar un pre-procesamiento de los espectros EEM por cada conjunto de datos de materias primas establecido para mejorar la relación señal/ruido. Las diferencias entre cada materia prima pueden por tanto verse con claridad: el hidrolizado de proteína de soja comprende 2 o 3 fluoróforos, el hidrolizado de proteína de arroz comprende 3 fluoróforos y el medio básico definido químicamente comprende más de 4 fluoróforos.

10 A fin de obtener una visión de conjunto de la variabilidad de materias primas de unos lotes a otros, se puede llevar a cabo un PCA del conjunto de datos de fluorescencia desdoblada por cada materia prima de componentes. El procedimiento de desdoblamiento puede aplicarse en cualquiera de los tres modos de un conjunto tridireccional. Para destacar las diferencias de unos lotes a otros se puede emplear la información de preservación del desdoblamiento del primer modo (muestras). De esta manera, las panorámicas de fluorescencia se pueden desdoblar en una fila de espectros de emisión uno tras otro (Figura 10).

15 Las dimensiones del conjunto de hidrolizado de proteína de soja son de 19x138x70 (lote x longitud de onda de la emisión x longitud de onda de la excitación). Tras la estrategia de desdoblamiento, se puede obtener una matriz de dos sentidos de tamaño 19x9.960. La figura 11 muestra una pequeña parte de los espectros resultantes para tres lotes diferentes de hidrolizado de proteína de soja. Se puede ver el ruido en las longitudes de onda de excitación extrema.

20 Para resolver estas desviaciones, se pueden utilizar varias estrategias. Se ha visto que el alisamiento Savitzky-Golay utilizando una ventana de 19 puntos y un polinomio de segundo grado para eliminar el ruido es lo más adecuado, y la Corrección Multiplicativa de Dispersión (MSC) es lo más adecuado para resolver la deriva de la línea de base.

25 El PCA de desdoblamiento se aplicó a la matriz pre-procesada de hidrolizado de proteína. Los datos eran un promedio, y el número óptimo de componentes principales se eligió utilizando el método de validación cruzada dejando uno fuera. La figura 12 muestra el trazado de resultados de PC1 x PC2 de un PCA que abarca el 96 % de la varianza hallada en todo la panorámica EEM desdoblada.

30 Tras desdoblar el hidrolizado de proteína de arroz resultante la matriz tenía el tamaño 12x9.945. Se aplicó el mismo pre-procesamiento utilizado para el hidrolizado de proteína de soja. La figura 13 muestra el trazado de resultados de PC1 x PC2 de un PCA utilizando tres componentes principales que dan cuenta de más del 98 % de la varianza en los espectros EEM desdoblados.

35 El tamaño de la matriz básica definida del medio químicamente desdoblada era de 14x9.600. Se utilizó el mismo procedimiento EEM de pre-procesamiento de espectros que se aplicó a los otros dos componentes de medios. La figura 14 muestra el trazado de resultados de PC1 x PC2 de un PCA utilizando dos componentes principales que cubren más del 92 % de la varianza total en los espectros EEM desdoblados. Al igual que antes, con espectros NIR para los mismos componentes de medios se halló que los lotes que arrojaban altos rendimientos se separaban de los lotes que arrojaban rendimientos inferiores en los trazados de resultados del PCA de espectros EEM desdoblados.

40 Se puede desarrollar un modelo PLS para predecir el rendimiento de producto al final del proceso basado en espectros NIR y de fluorescencia obtenidos para diferentes lotes de cada componente de medios y/o sus combinaciones. Al algoritmo PLS se le asigna un bloque X (espectros pre-procesados, con o sin selección de variables) y un bloque Y (parámetro de producto) y se correlacionan ambos hallando la variación en X responsable de los cambios en Y (esto es, maximizando la covarianza entre ambos bloques). Se puede definir un conjunto básico en el que se puede incluir la mayoría de los distintos lotes de materias primas. Aparte de los lotes de réplica que tienen las mismas combinaciones de lotes, se seleccionó el que daba el más alto rendimiento de producto para la calibración del conjunto de datos (Tabla 7).

55

Tabla 7.

lote	Nº de lote de hidrolizado de proteína de soja F/ZF	producto a 330 h [mg/l]
D52KD13	1	1458
D52KD22	4	1232
D55KD13	5	1430
D55KD23	3	1257
D55KD31	6	1263
D73KD13	2	1120
D73KD33	7	1044
D79KD22	8	1162

Los espectros NIR se pueden pre-procesar como se ha descrito antes para eliminar la influencia de efectos físicos que se originan de diferentes distribuciones de tamaños de partículas. Como no se utilizaron espectros de réplica, el método de validación cruzada dejando uno fuera se utilizó como estrategia de validación interna.

5 El modelo obtenido estaba compuesto únicamente por dos VL pero se obtuvo un R2 no significativo de 0,139. El trazado predicho de validación medida vs. cruzada se presenta en la figura 15.

Se puede construir un modelo PLS que correlacione espectros NIR de diferentes lotes del medio básico químicamente definido y rendimiento de producto utilizando el conjunto de datos de calibración tal y como se muestra en la Tabla 8.

10

Tabla 8.

lote	Nº de lote del medio básico químicamente definido F/ZF	producto a 330 h [mg/l]
D45KD11	1	1314
D52KD13	2	1458
D61KD12	3	1134
D73KD21	4	1147
D79KD22	5	1162

El modelo obtenido estaba compuesto únicamente por dos VL pero de nuevo se obtuvo un R2 no significativo de 0,04 (Figura 16).

15

Considerando no solo un componente de medio, sino los dos más relevantes que influyen en el rendimiento, y teniendo en cuenta también que se captura distinta información química por cada método espectroscópico distinto empleado, se puede utilizar una estrategia de combinación entre los mismos componentes espectroscópicos/ componentes de medios diferentes y también entre los diferentes componentes espectroscópicos/ componentes de medios diferentes.

20

Los criterios utilizados para seleccionar los lotes de calibración y validación se basaron en obtener el intervalo más amplio posible durante la calibración (Tabla 9).

25

Tabla 9.

	lote	Nº de lote de hidrolizado de proteína de soja F/ZF	Nº de lote del medio básico químicamente definido F/ZF	producto a 330 h [mg/l]
calibración	D45KD11	1	1	1314
	D45KD31	3	1	999
	D52KD13	1	2	1458
	D52KD22	4	2	1232
	D55KD13	5	2	1430
	D55KD31	6	2	1263
	D61KD12	3	3	1134
	D73KD13	2	4	1120
	D73KD33	7	4	1044
	D79KD22	8	5	1162
validación	D45KD23	2	1	1061
	D55KD23	3	2	1257
	D73KD21	8	4	1147

La validación externa se efectuó con un tercio del conjunto de datos. Los datos de calibración y validación (espectros NIR) se pre-procesaron de la misma manera descrita anteriormente. El modelo de predicción obtenido se basa en 3 VL y el R2 obtenido alcanzó un valor significativo de 0,88.

30

La precisión del modelo y la robustez a largo plazo se refleja en un alto R2 con bajos errores tanto de calibración como de validación, con una pequeña diferencia entre RMSECV y RMSEP (Figura 17). En el caso superior, el error de predicción fue bajo (RMSEP = 36 mg/l) y no difería mucho del RMSECV (126 mg/l).

35

Así pues, se ha encontrado que el rendimiento de producto se puede correlacionar con los datos espectroscópicos de diferentes compuestos de un medio de cultivo obtenido mediante una combinación de información espectrográfica de la misma naturaleza (NIR) para las dos materias primas (más importantes) del proceso o componentes de medios. Cada espectro tiene 944 números de onda y el conjunto de datos completo de calibración incluido en el modelo se representa mediante 18.880 variables (10 muestras x 2 materias primas x 944 números de onda tras la selección de variables). Para reducir la carga requerida de trabajo se llevó a cabo un análisis PCA basado en los espectros que se comprimiron por vez primera convirtiendo la información contenida en unas cuantas variables no

40

correlacionadas. El modelo obtenido con ello era más simple y contenía solo 2 variables latentes (LV) y se obtuvo un R2 de 0,81.

- 5 Diferentes métodos espectroscópicos capturan información química complementaria. Utilizando dos tipos distintos de información espectroscópica se mejoró la calidad predictiva del modelo. Por lo tanto, se utilizaron los espectros fluorescentes de hidrolizado de proteína de soja y espectros NIR del medio básico definido químicamente (Tabla 10).

Tabla 10.

	lote	Nº de lote de hidrolizado de proteína de soja F/ZF	Nº de lote del medio básico químicamente definido F/ZF	producto a 330 h [mg/l]
calibración	D45KD11	1	1	1314
	D45KD31	3	1	999
	D52KD13	1	2	1458
	D52KD22	4	2	1232
	D55KD13	5	2	1430
	D55KD31	6	2	1263
	D61KD12	3	3	1134
	D73KD13	2	4	1120
	D73KD33	7	4	1044
	D79KD22	8	5	1162
validación	D45KD23	2	1	1061
	D55KD23	3	2	1257
	D73KD21	8	4	1147

- 10 Los espectros de fluorescencia y espectros NIR se comprimieron a unos cuantos componentes principales tras el pre-procesamiento, tal como se ha descrito anteriormente. El modelo obtenido tiene solo 3 variables latentes y se obtuvo un R2 de 0,90 (Figura 18). Este modelo tiene mejor rendimiento cuando se compara con modelos anteriores y es más robusto puesto que no solo tiene un valor R2 más alto, sino que también tiene unos valores de RMSECV y RMSEP más bajos (cerca de 90 mg/l) con una diferencia muy pequeña entre ellos.

- 15 Se efectuó una prueba adicional utilizando MIR en vez de NIR para el medio básico químicamente definido. Los conjuntos de datos de calibración y validación empleados fueron los mismos que se han presentado anteriormente (véase la Tabla 10). Los espectros de fluorescencia y MIR se pre-procesaron como se ha descrito antes. El modelo obtenido tiene 3 variables latentes, un R2 de 0,88, y unos valores de RMSECV y RMSEP bajos sin diferencias entre ellos (100 mg/l aproximadamente ambos), mostrando por tanto ninguna diferencia significativa con lo obtenido con los datos NIR para el medio básico químicamente definido (Figura 19).

- 20 Los espectros NIR del hidrolizado de proteína de soja y los espectros de fluorescencia del medio básico químicamente definido se unieron y se evaluó el modelo resultante. Los conjuntos de datos de calibración y validación empleados para construir el modelo fueron los mismos que se han descrito anteriormente (véase la Tabla 10). El modelo obtenido tiene 3 variables latentes y un valor R2 muy similar (0,87) (Figura 20) y unos valores de RMSECV y RMSEP (124 mg/l y 60 mg/l, respectivamente).

- 25 Con una varianza analítica para las analíticas de referencia de producto en torno a 60 mg/l (5 % de 1200 mg/l la concentración media de producto) la mayoría de modelos desarrollados mostraron una precisión de predicción muy cercana al límite experimental.

- 30 En conclusión, para hallar una predicción del rendimiento de producto a 330 h, debe utilizarse la información espectral tanto del hidrolizado de proteína de soja como del medio básico químicamente definido. El uso de datos espectroscópicos de fluorescencia para el medio básico químicamente definido arroja unos errores de predicción ligeramente inferiores (pero aun así muy comparables), que los modelos basados en datos espectroscópicos NIR para el medio básico químicamente definido y los datos de espectroscopias de fluorescencia 2D para el hidrolizado de proteína de soja.

- 35 El método tal y como se describe en el presente documento va dirigido a la combinación de espectros de distinta naturaleza (espectros de fluorescencia y espectros IR), los cuales tienen distintas dimensiones de forma intrínseca (dos (2D) y una (1D), respectivamente), y eso requiere las operaciones de comprimir primero cada espectro a resultados del análisis de componentes principales y en segundo lugar producir combinaciones lineales de cada resultado de rendimiento. Los espectros de distinta naturaleza se combinan por medio de una reducción dimensional y una combinación lineal de dichas variables transformadas reducidas (resultados de PCA obtenidos comprimiendo cada espectro).

Así pues, en el método tal y como se describe en el presente documento, se utilizan espectros de distintas dimensiones y naturaleza para capturar en una mezcla de dos materias primas distintas de fermentación los componentes responsables del rendimiento de fermentación de dichas materias primas y para hacer predicciones de los rendimientos de fermentación de una combinación específica de lotes.

5 Con el método tal y como se describe en el presente documento es posible predecir, basándose en los espectros de dos materias primas distintas que se han de utilizar en un proceso de fermentación el rendimiento de 10 a 14 días por adelantado determinando las condiciones en el momento de cosechar la fermentación.

10 Los ejemplos y figuras a continuación se ofrecen para ayudar a comprender la presente invención, el verdadero ámbito de lo que se presenta en las reivindicaciones adjuntas. Se entiende que se pueden efectuar modificaciones en los procedimientos expuestos sin alejarse del ámbito de la invención, el cual se define en las reivindicaciones.

Descripción de las figuras

- 15 Figura 1 Distribución de los diferentes lotes de hidrolizado de proteína de soja probados en un espacio bidimensional construido mediante PCA basado en los espectros NIR originales.
- Figura 2 Espectros NIR de distintos lotes de hidrolizado de proteína de soja.
- Figura 3 Distribución de los diferentes lotes de hidrolizado de proteína de arroz en un espacio bidimensional construido mediante PCA basado en los espectros NIR originales.
- 20 Figura 4 Distribución de los diferentes lotes probados de medio básico químicamente definido en un espacio bidimensional construido mediante PCA basado en los espectros NIR originales.
- Figura 5 Análisis de PCA basado en espectros pre-procesados de lotes de hidrolizado de proteína de soja.
- Figura 6 Análisis de PCA basado en espectros pre-procesados de lotes de hidrolizado de proteína de arroz.
- Figura 7 Análisis de PCA basado en espectros pre-procesados de lotes de medio básico químicamente definido.
- 25 Figura 8 Panorámica EEM de fluorescencia de unas muestras de lotes de hidrolizado de proteína de soja.
- Figura 9 Panorámica EEM de fluorescencia procesado de unas muestras de lotes de hidrolizado de proteína de soja.
- Figura 10 Panorámicas de fluorescencia desdoblados en una fila de espectros de emisión.
- 30 Figura 11 Extracto de espectros desdoblados para tres lotes diferentes de hidrolizado de proteína de soja.
- Figura 12 Trazado de resultados de PC1 x PC2 de un PCA para hidrolizados de proteína de soja de la panorámica EEM desdoblada.
- Figura 13 Trazado de resultados de PC1 x PC2 de un PCA para hidrolizados de proteína de arroz de la panorámica EEM desdoblada.
- 35 Figura 14 Trazado de resultados de PC1 x PC2 de un PCA para medio básico químicamente definido de la panorámica EEM desdoblada.
- Figura 15 Trazado medido vs. trazado predicho mediante validación cruzada.
- Figura 16 Modelo PLS que correlaciona espectros NIR de lotes diferentes del medio básico químicamente definido y rendimiento de producto.
- 40 Figura 17 Modelo PLS que correlaciona espectros NIR de lotes diferentes del hidrolizado de proteína de soja y el medio básico químicamente definido y rendimiento de producto.
- Figura 18 Modelo PLS que correlaciona espectros de fluorescencia de lotes diferentes del hidrolizado de proteína de soja y espectros NIR de lotes diferentes del medio básico químicamente definido y rendimiento de producto.
- 45 Figura 19 Modelo PLS que correlaciona espectros de fluorescencia de lotes diferentes del hidrolizado de proteína de soja y espectros MIR de lotes diferentes del medio básico químicamente definido y rendimiento de producto.
- Figura 20 Modelo PLS que correlaciona espectros NIR de lotes diferentes del hidrolizado de proteína de soja y espectros de fluorescencia de lotes diferentes del medio básico químicamente definido y rendimiento de producto.
- 50 Figura 21 Radiaciones de absorción de NIR de las bandas de sobretono y combinación de moléculas orgánicas de enlaces covalentes.

Ejemplo

55 Materiales y métodos

Cultivo de células:

60 Las células se cultivaron en matraces oscilantes en un entorno donde la temperatura, humedad y dióxido de carbono estaban controlados. A fin de comparar los distintos lotes, se prepararon unos medios con estos lotes y se inocularon células en matraces oscilantes que contenían estos medios. Se añadió un determinado volumen de un medio de alimentación a diario al cultivo del matraz oscilante a fin de prolongar el crecimiento celular y lograr unas mayores concentraciones de producto.

65 Espectroscopia de infrarrojo cercano (NIR):

5 NIR emerge en la década de los 60 en el mundo analítico, con los trabajos de Karl Norris del Ministerio de agricultura de EE.UU. (Siesler et al, 2002). En el espectro electromagnético, la región NIR se ubica entre infrarrojo medio y visible. En un intervalo de números de onda de 4.000-14.000 cm^{-1} (respectivamente una longitud de onda de 700-2.500 nm), la radiación de absorción de las bandas de sobretono y combinación de enlaces covalentes tales como N-H, O-H y C-H de moléculas orgánicas (Figura 21).

10 Los espectros NIR se recopilaron utilizando viales de centelleo de fondo plano en un sistema Bruker® MPA FT-NIR, equipado con una fuente de halógeno-tungsteno y un detector InAs. Cada espectro se registró en el intervalo de números de onda de 4.999 a 9.003 cm^{-1} , en un promedio de 32 exploraciones y una resolución espectral de 8 cm^{-1} .

Espectroscopia de infrarrojo medio (MIR):

15 Las espectroscopias de infrarrojo medio se obtuvieron utilizando cubetas de cuarzo en un Avatar 370 FT-IR, Thermo Fischer, Diamant ATR. Cada espectro se registró en el intervalo de número de onda de 4.000 a 400 cm^{-1} .

Espectroscopia de fluorescencia:

20 La espectroscopia de fluorescencia utiliza la irradiación a un determinado número de onda para excitar las moléculas, las cuales entonces emitirán radiación de una longitud de onda distinta. Esta técnica se utiliza a menudo para estudiar la estructura y función de las macromoléculas, especialmente las interacciones de proteínas. En la Tabla a continuación se presenta una sutil descripción de las características de fluorescencia de cromóforos que se encuentran en proteínas y ácidos nucleicos.

Sustancia	Absorción		Fluorescencia	
	λ_{max} (nm)	ϵ_{max} (10 ⁻³)	λ_{max} (nm)	f _F
triptófano	280	5,60	348	0,20
tirosina	274	1,40	393	0,14
fenilalanina	257	0,20	282	0,14
adenina	260	13,40	321	2,60x10 ⁻⁴
guanina	275	8,10	329	2,60x10 ⁻⁴
citosina	267	6,10	313	0,80x10 ⁻⁴
uracilo	260	9,50	308	0,40x10 ⁻⁴
NADH	340	6,20	470	0,02

25 Los espectros de fluorescencia 2D de materias primas de células se obtuvieron usando longitudes de onda de excitación de 200 nm a 600 nm, con intervalos de 5 nm, y longitudes de onda de emisión también de 200 nm a 600 nm, pero con intervalos de 2 nm, arrojando un total de longitudes de onda de excitación de 81 y un 201 de emisión. Los espectros de fluorescencia de emisión-excitación se midieron utilizando un Espectómetro Varian Cary Eclipse, sobre un intervalo de longitudes de onda de excitación de 200 nm a 600 nm con intervalos de 5 nm, y un intervalo de longitudes de onda de emisión también de 200 nm a 600 nm, pero con intervalos de 2 nm, arrojando un total de 81 longitudes de onda de excitación y 201 de emisión. Los datos se recopilaron utilizando el software Cary Eclipse Bio, Package 1.1.

35 Tratamiento espectral y análisis quimiométrico:

40 El pre-procesamiento de espectros y los cálculos quimiométricos se realizaron en Matlab® 7.2 (MathWorks, U.S.A.) utilizando el kit PLS 5.5 (Eigenvector, U.S.A.) y Simca P+ 12.01 (Umetrics, Suecia). Las dispersiones de Rayleigh y Raman se eliminaron utilizando el algoritmo EEMscat (Bahram et al, 2006).

45 El análisis de datos multivariados se realizó utilizando PCA (Análisis de componentes principales, *Principal Component Analysis*) y PLS (Mínimos cuadrados parciales, *Partial Least Squares*). Estas técnicas se basan en la reducción de la dimensionalidad presente en los datos, permitiendo la recuperación de información relevante oculta en la cantidad masiva de datos. Se realiza transformando las variables medidas originales en nuevas variables denominadas componentes principales. El análisis PCA se usó para hallar patrones en los espectros. Con el fin de relacionar estos patrones con un parámetro en particular, se llevó a cabo análisis PLS para construir un modelo matemático capaz de predecir los valores de este parámetro en futuras muestras utilizando solo la información espectral.

50 A fin de construir modelos fiables, la calidad de las mediciones analíticas es de una importancia fundamental. Puesto que el ruido y la información no deseada son intrínsecos a las mediciones, es necesario hacer un pre-tratamiento de los espectros obtenidos.

55 Una de las técnicas más habituales para lidiar con estos problemas en los espectros NIR es el filtro de alisamiento de Savitzky-Golay (Savitzky, A. y Golay, M.J.E., Anal. Chem., 36 (1964) 1627-1639), y este se utiliza habitualmente

junto con derivados, lo cual tiene la ventaja de que reduce los desplazamientos de la línea de base y mejora las propiedades significativas del espectro.

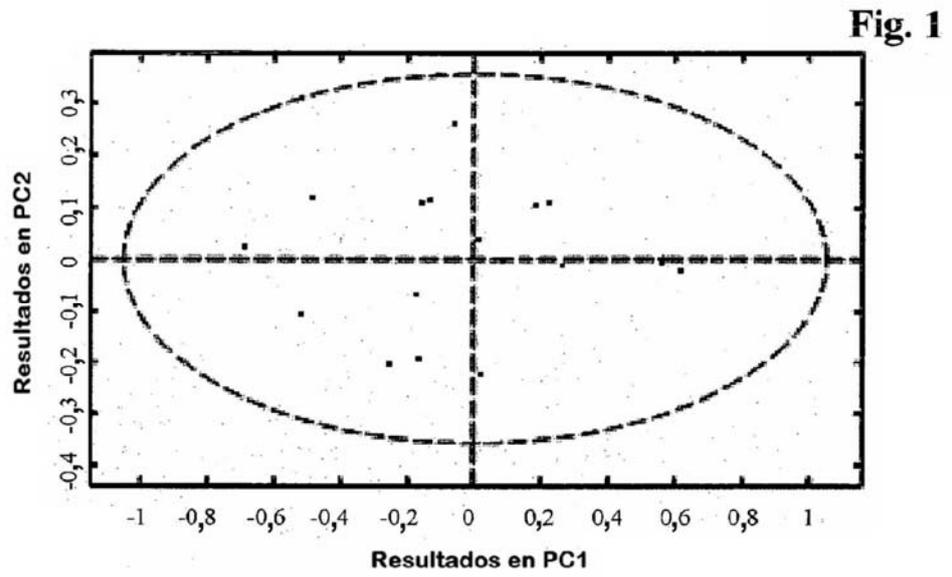
5 Para los espectros de fluorescencia, los principales problemas están relacionados con la dispersión Raman y Rayleigh, los cuales son ocasionados por desviaciones de la luz que no están relacionadas con las propiedades de fluorescencia de la muestra. Puesto que las regiones de longitud de onda afectadas por la dispersión son conocidas, las intensidades medidas en dichas regiones particulares pueden eliminarse reemplazándolas por puntos interpolados.

10 Se desplegaron espectros de excitación/emisión tridireccionales a fin de tener una matriz adecuada para el análisis PLS y PCA. También se efectuó un análisis basado en Parafac tridireccional con fines de calibración. (Bahram, M., et al., J. Chemometrics, 20 (2006) 99-105). El enfoque de desdoblamiento consiste en concatenar dos de estas tres dimensiones, manteniendo la otra fija. En este caso se concatenaron el eje de emisión y excitación, manteniendo la información de las muestras.

15

REIVINDICACIONES

- 5 1. Un método para la selección de lotes de componentes de medios de cultivo para usarlos en el cultivo de una célula de mamífero que expresa una proteína de interés en el que se utilizan al menos dos componentes diferentes en el cultivo, que comprende las siguientes etapas:
- 10 a) proporcionar espectros de diferentes lotes de un primer componente obtenido con un primer método espectroscópico seleccionado de entre espectroscopia NIR y espectroscopia MIR y espectros de un segundo componente obtenido con espectroscopia de fluorescencia 2D como un segundo método de espectroscopia y el rendimiento de sobrenadante de cultivo de la proteína de interés obtenida en un cultivo usando combinaciones de estos lotes diferentes del primer y el segundo componente,
- 15 b) identificar una relación de espectros fusionados de las dos técnicas espectroscópicas diferentes tras computar resultados de espectros PCA con el rendimiento del cultivo,
- 20 c) proporcionar un espectro de un lote adicional del primer componente obtenido con el primer método espectroscópico y un espectro de un lote adicional del segundo componente obtenido con el segundo método espectroscópico,
- d) seleccionar la combinación del primer componente proporcionado y el segundo componente proporcionado si el rendimiento de sobrenadante predicho del cultivo basado en la relación de espectros fusionados tras computar resultados de espectros PCA identificados en b) se halla dentro de +/- 10 % del resultado del producto proporcionado en a).
- 25 2. El método de acuerdo con la reivindicación 1, caracterizado por que la identificación se hace mediante análisis de componentes principales.
- 30 3. El método de acuerdo con la reivindicación 2, caracterizado por que el análisis de componentes principales es un análisis de componentes principales desdoblado.
4. El método de acuerdo con la reivindicación 3, caracterizado por que el desdoblamiento preserva la información del primer modo (muestra).
- 35 5. El método de acuerdo con una cualquiera de las reivindicaciones que anteceden, caracterizado por que la identificación es de una relación entre espectros fusionados y comprimidos con los resultados PCA con un rendimiento de cultivo en el momento de la cosecha mediante análisis de mínimos cuadrados parciales.
6. El método de acuerdo con cualquiera de las reivindicaciones que anteceden, caracterizado por que la proteína de interés es un anticuerpo o un fragmento de anticuerpo o un conjugado de anticuerpo.



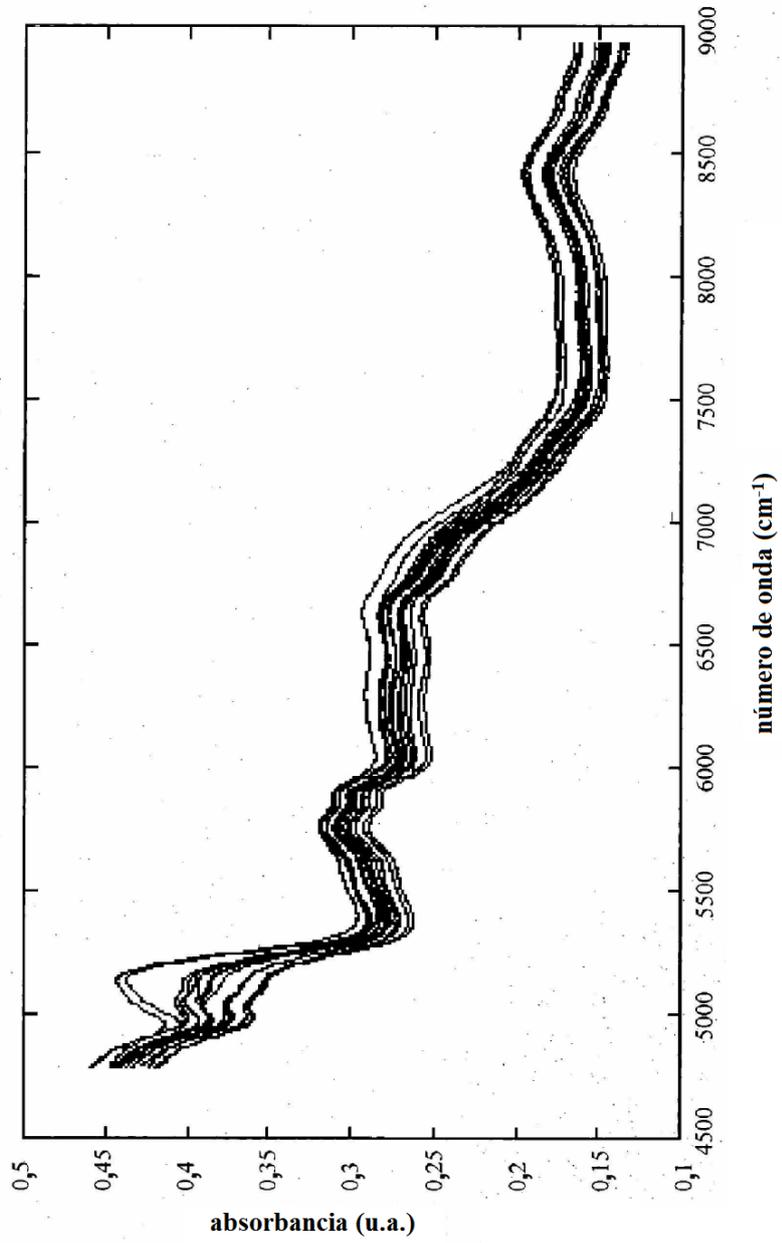


Fig. 2

Fig. 3

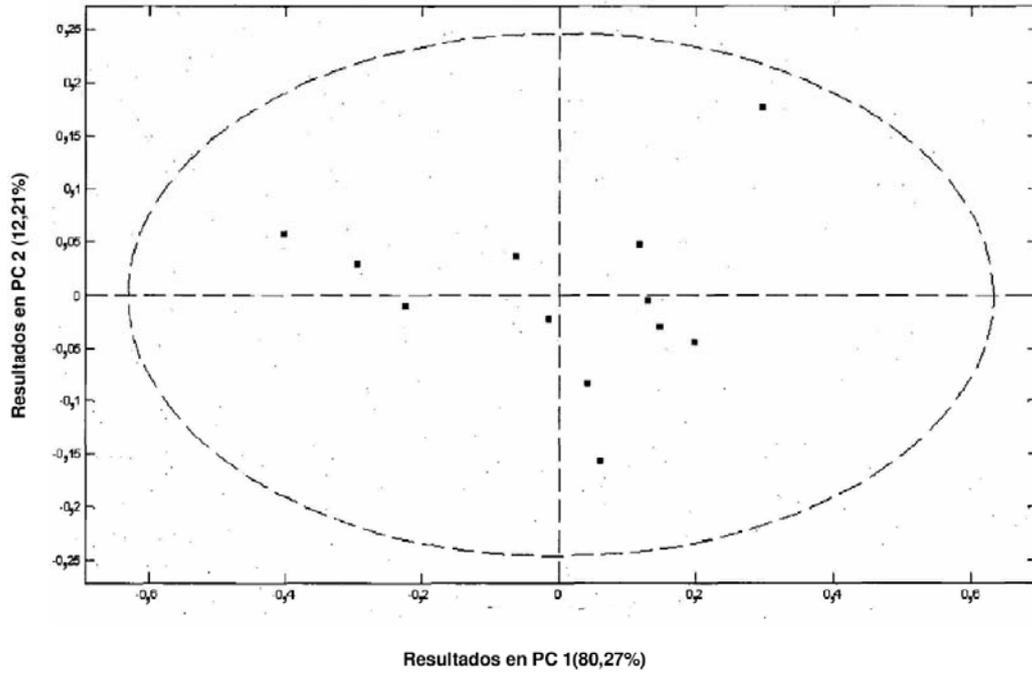


Fig. 4

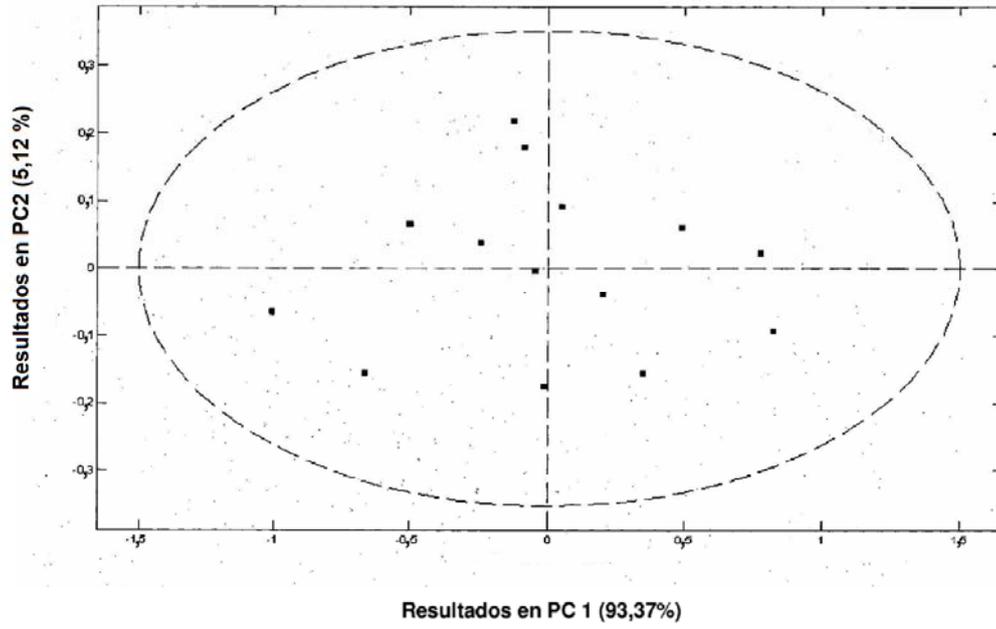


Fig. 5

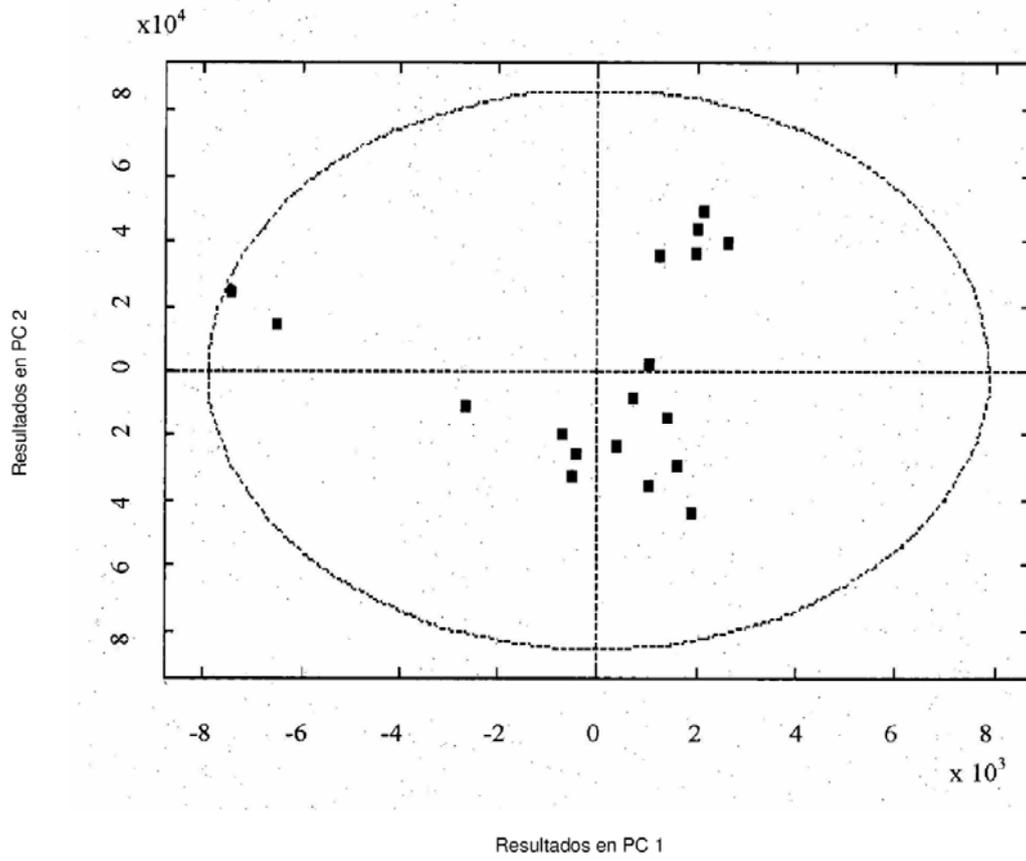


Fig. 6

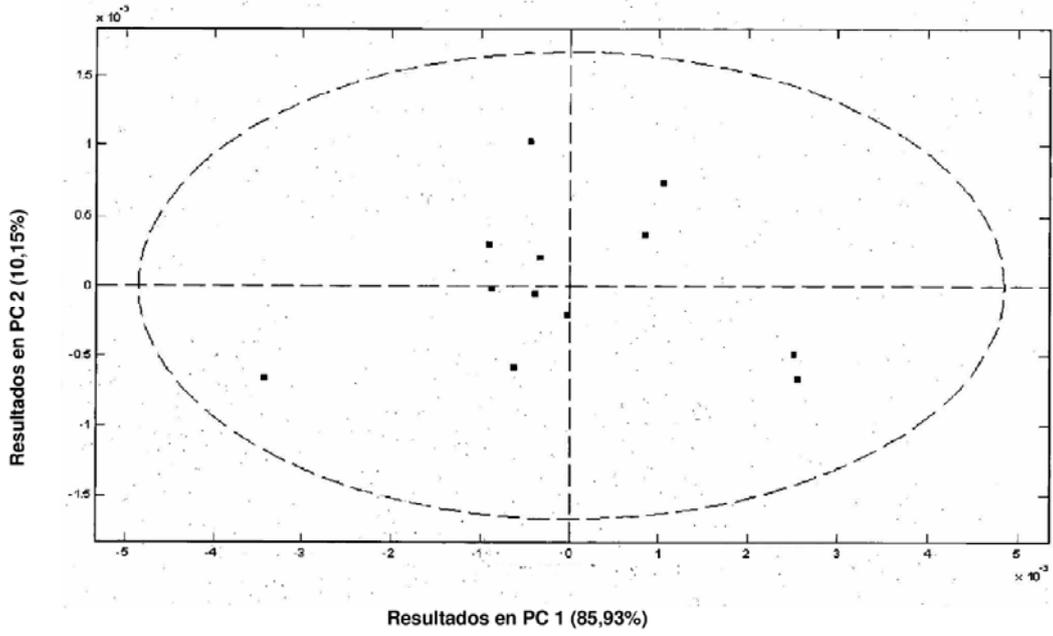


Fig. 7

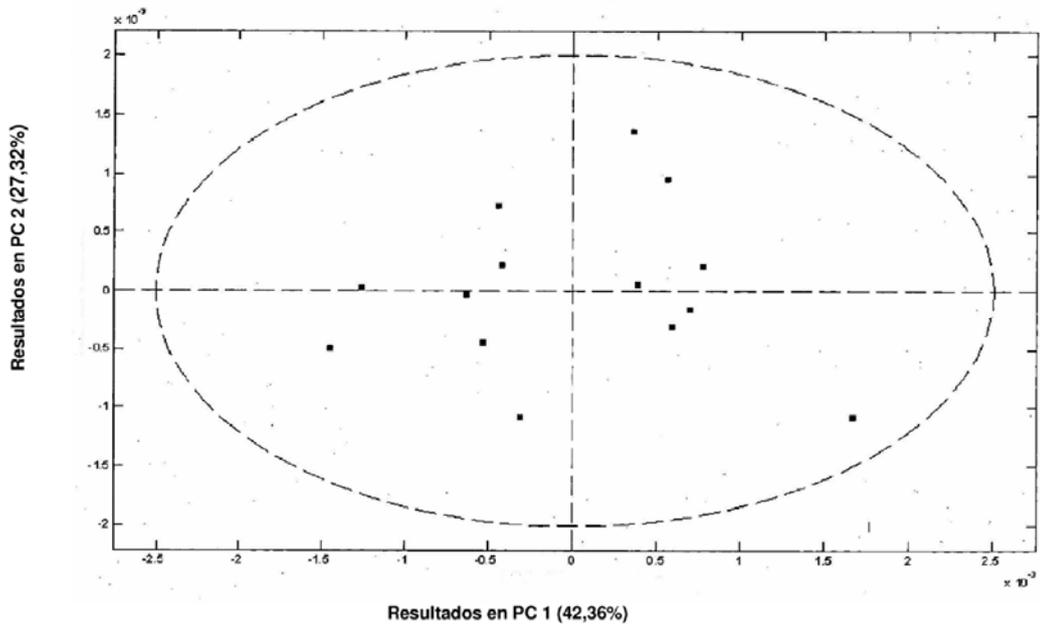


Fig. 8

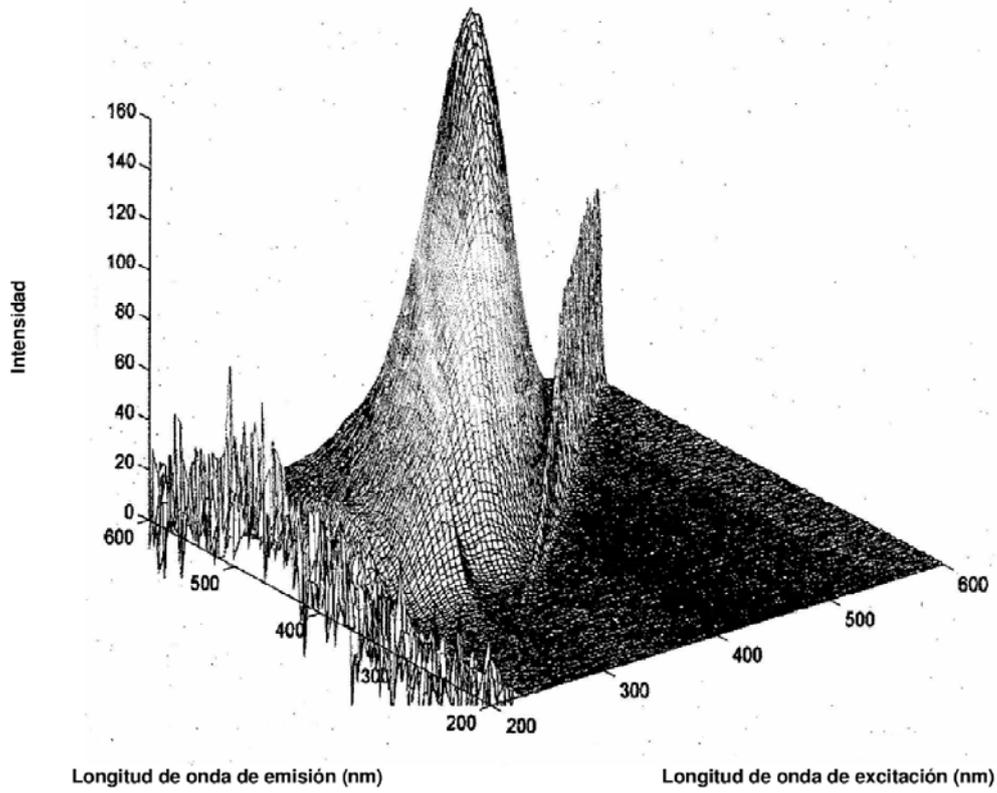


Fig. 9

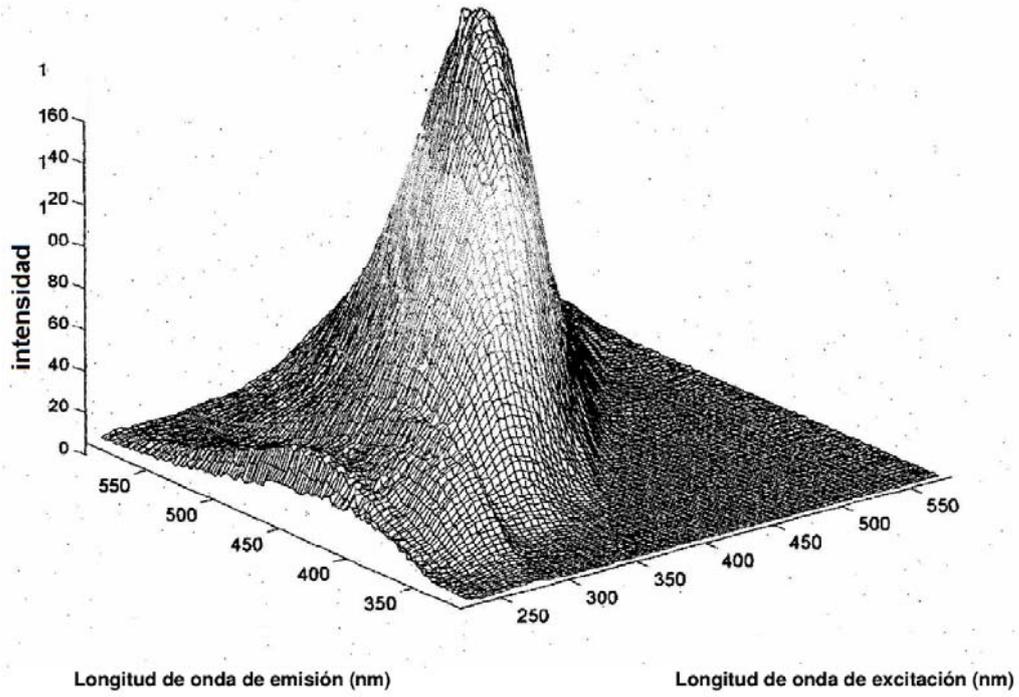


Fig. 10

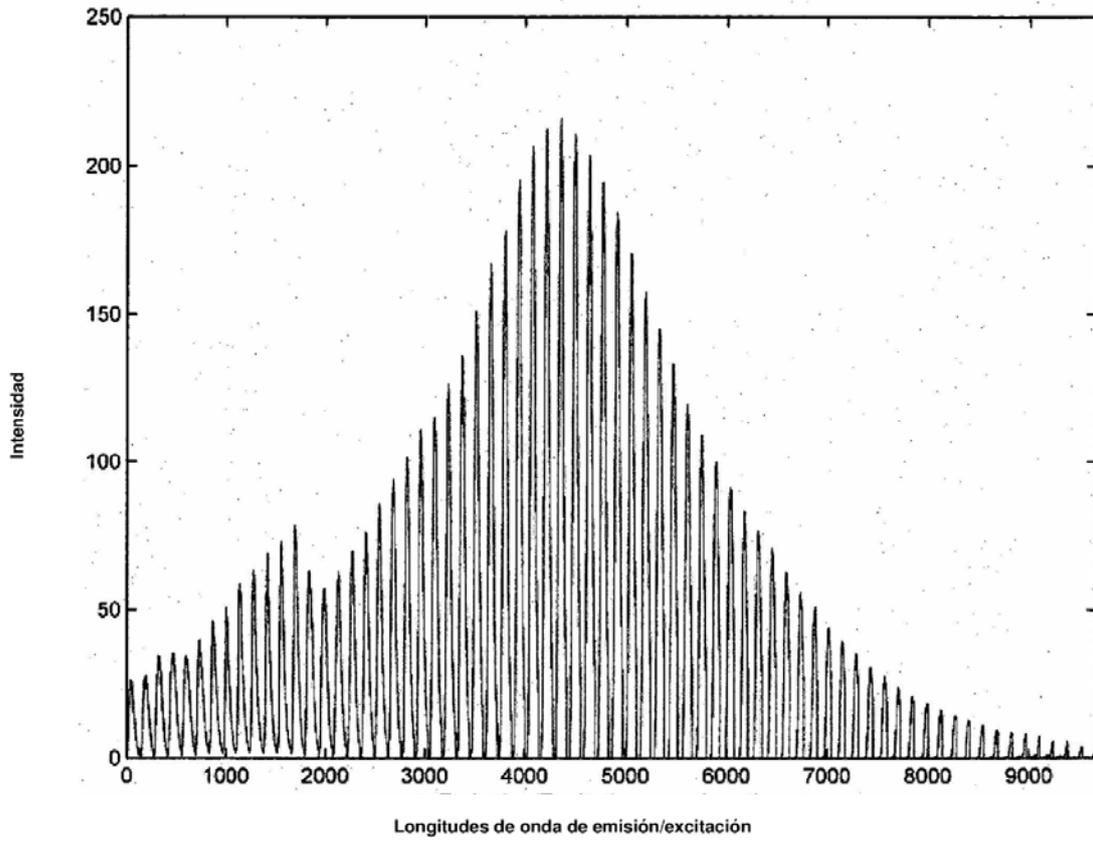


Fig. 11

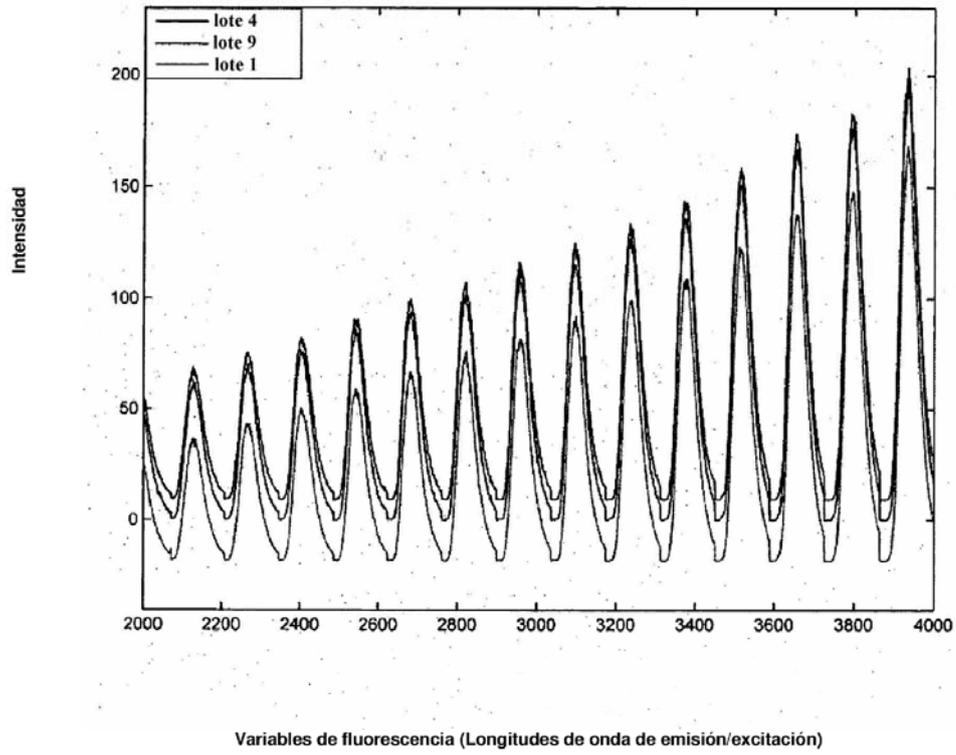


Fig. 12

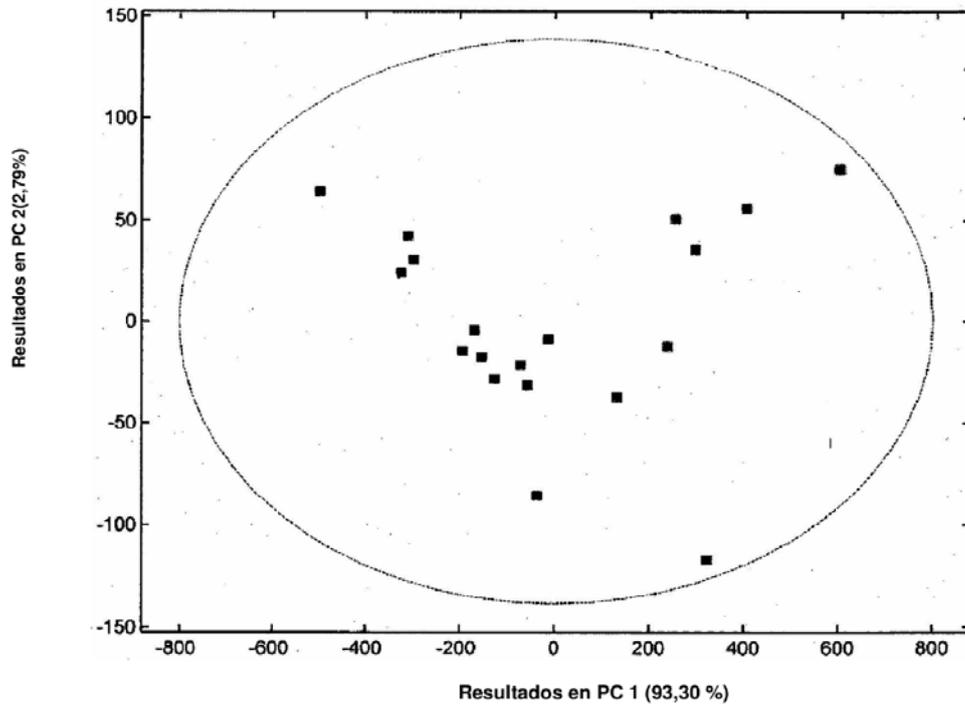


Fig. 13

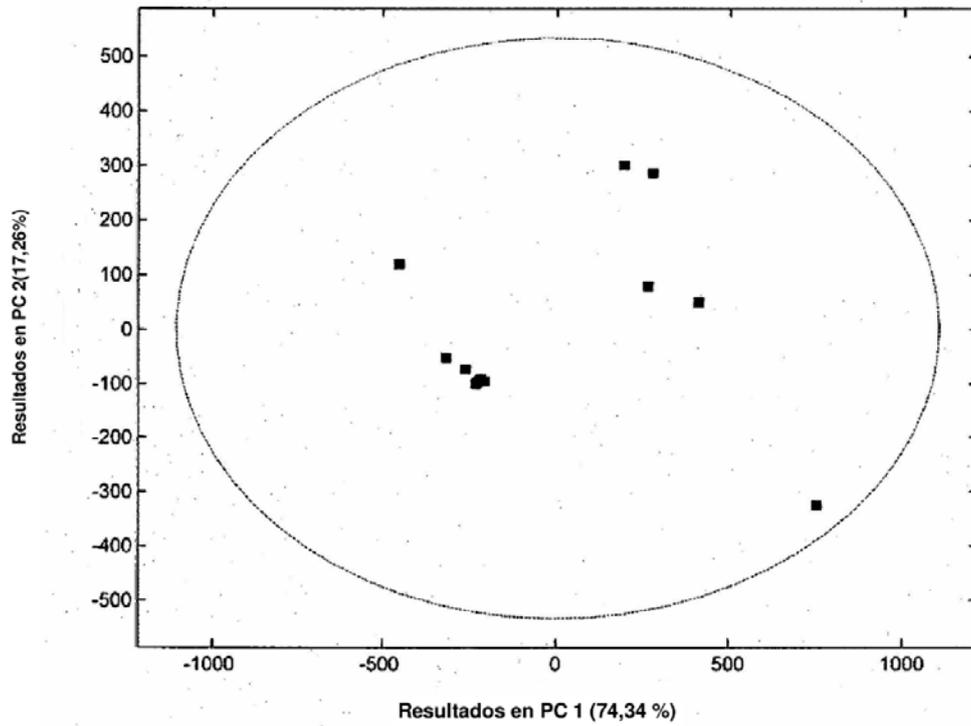


Fig. 14

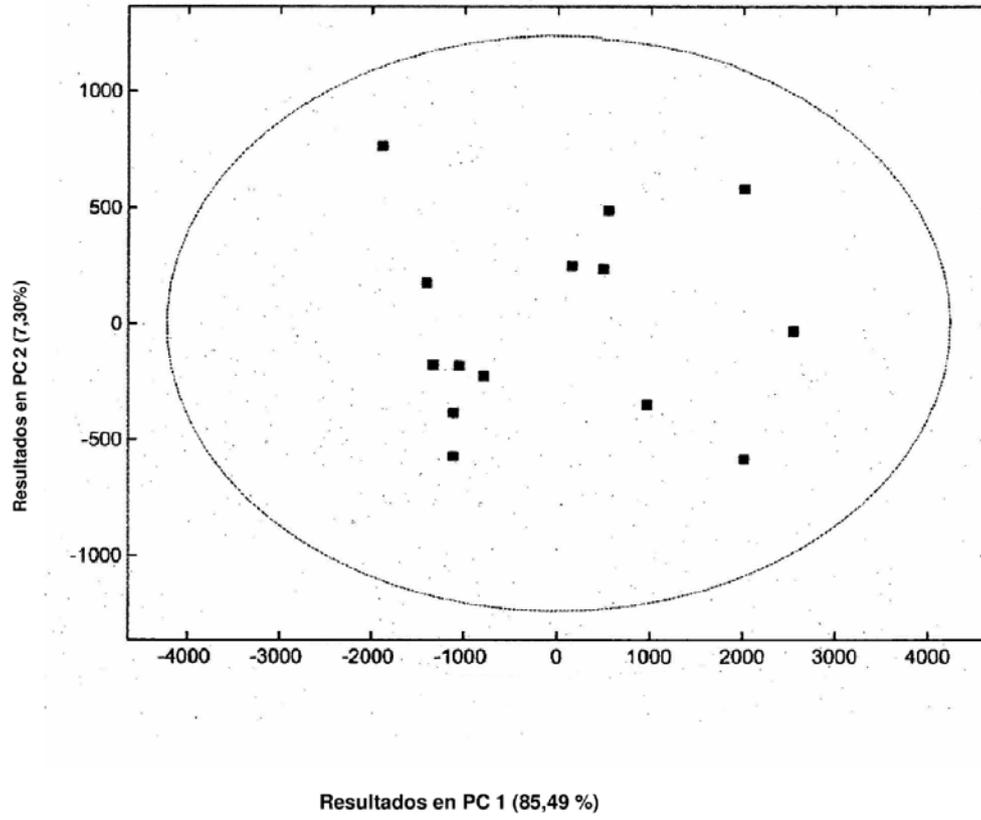


Fig. 15

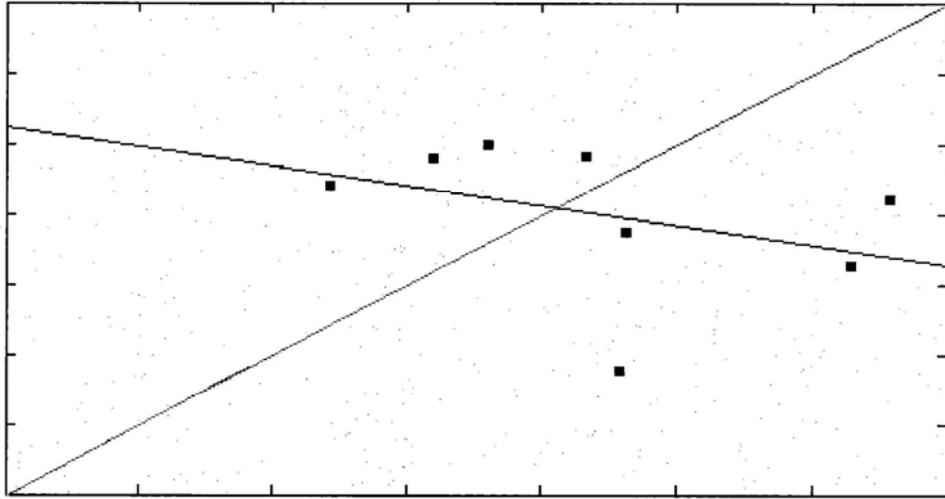


Fig. 16

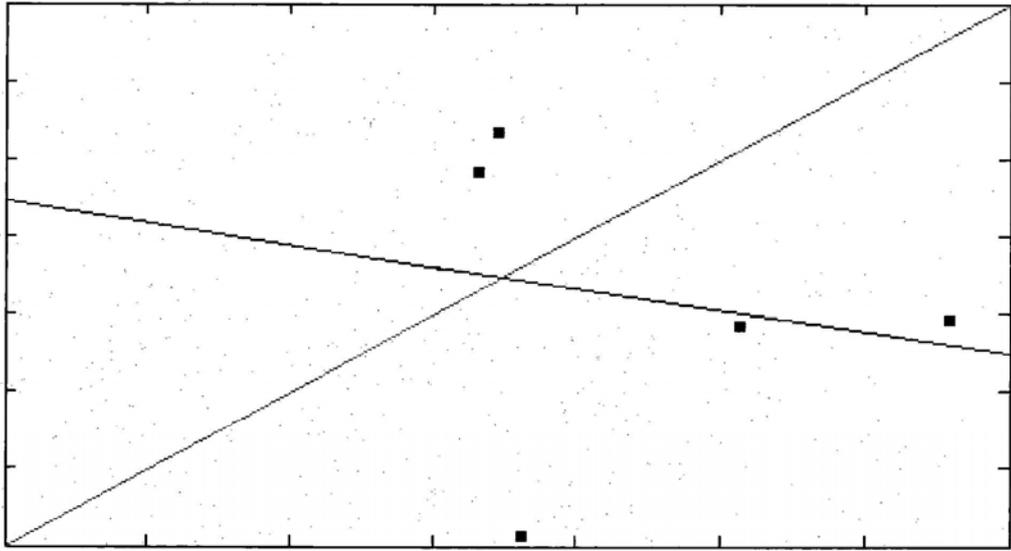


Fig. 17

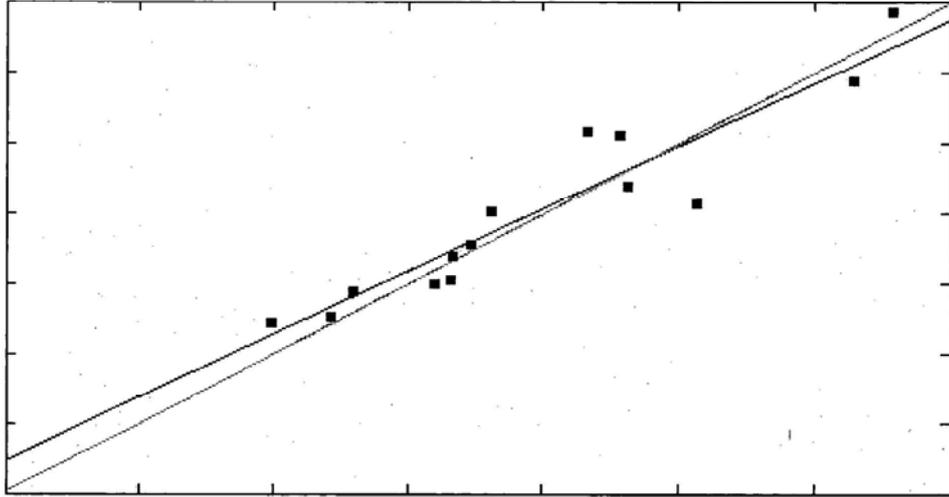


Fig. 18

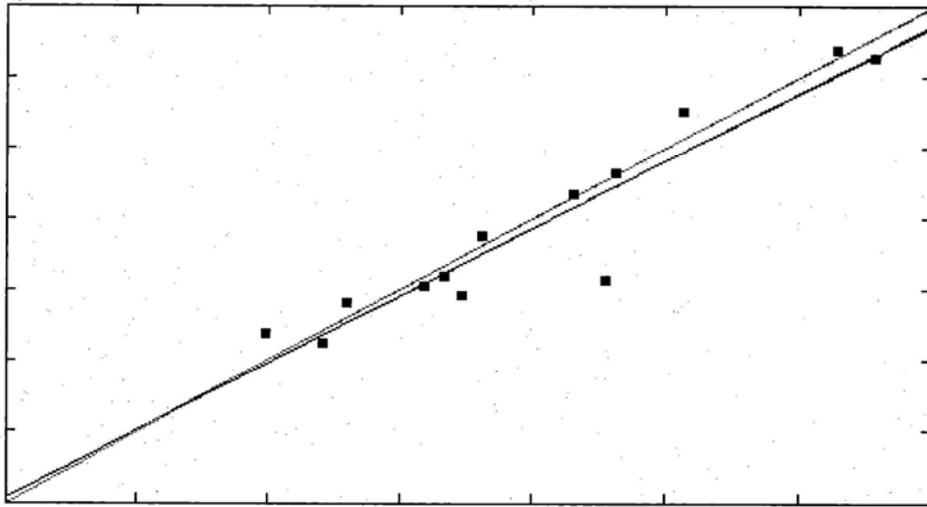


Fig. 19

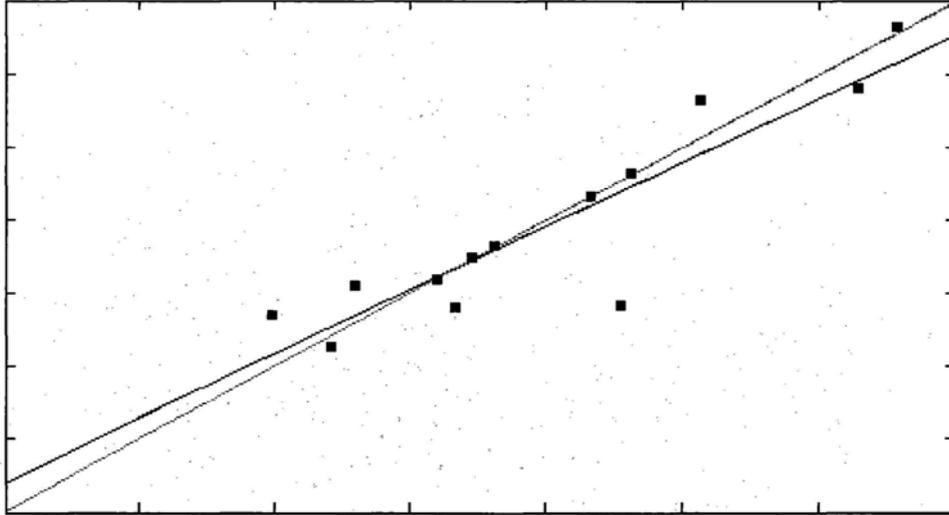
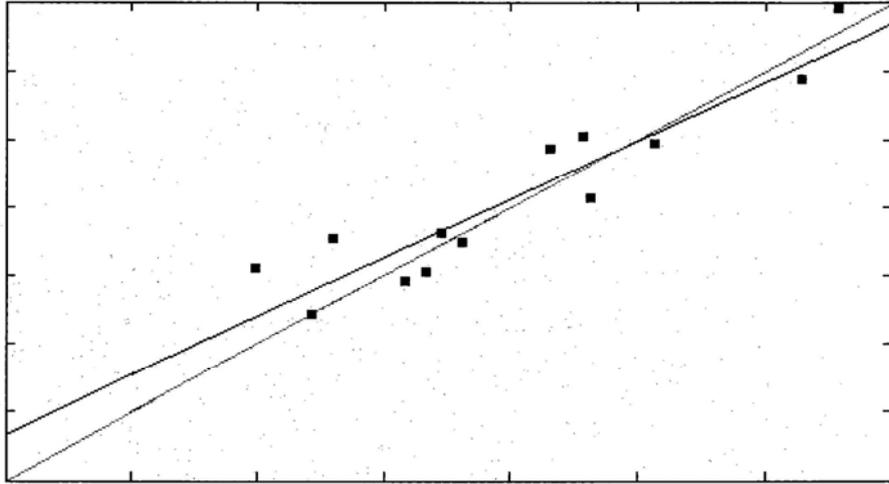


Fig. 20



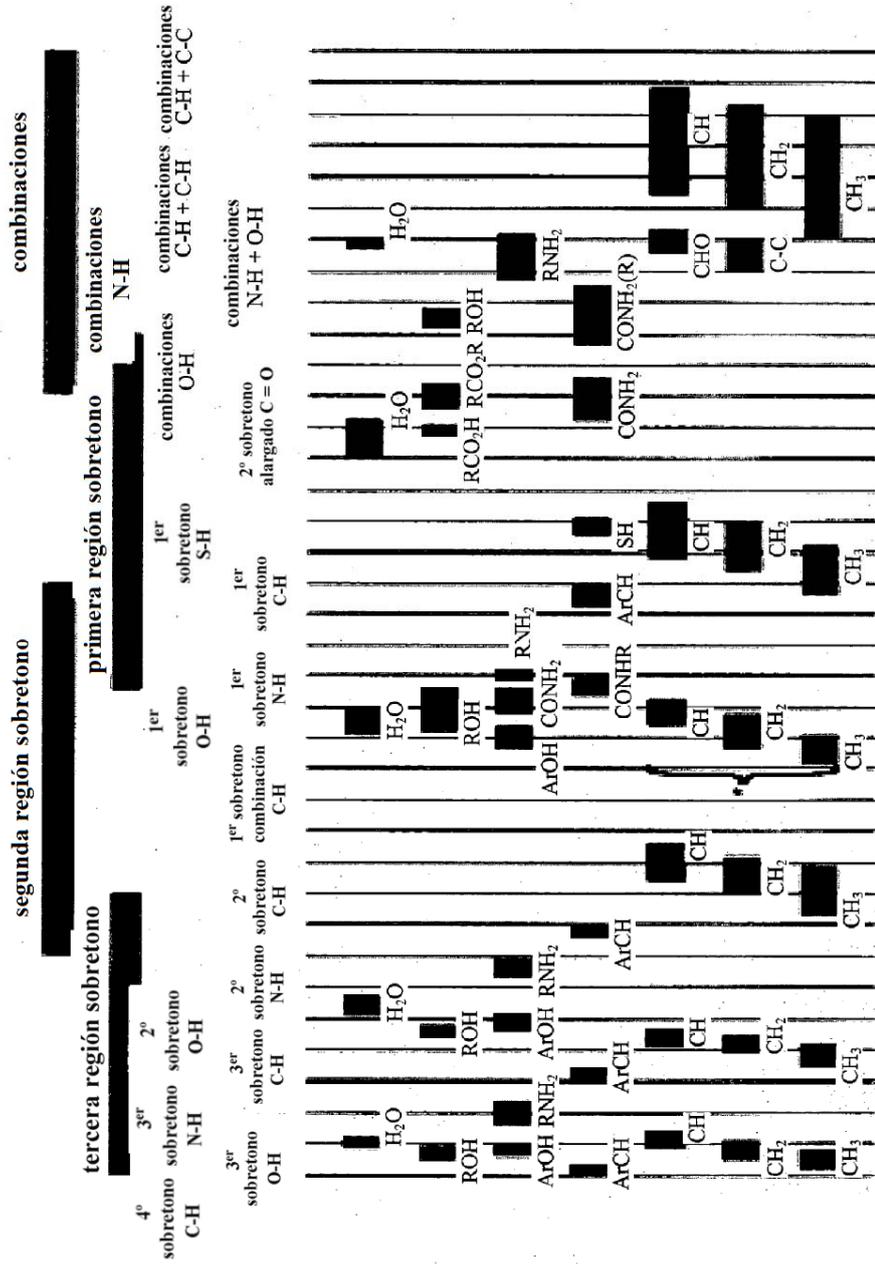


Fig. 21

longitud de onda
 λ nm
 número de onda
 $\bar{\nu}$ cm⁻¹