

## Aplicación del análisis de componentes principales en la formación de dímeros durante la purificación de un anticuerpo monoclonal

Ivis Regalado Fonseca<sup>1\*</sup> ORCID: <https://orcid.org/0000-0002-7989-8225>

Osvaldo Mora Montes de Oca<sup>1</sup> ORCID: <https://orcid.org/0000-0002-7354-8427>

Lourdes Zumalacárregui de Cárdenas<sup>2\*\*</sup> ORCID: <https://orcid.org/0000-0001-6921-737X>

<sup>1</sup> Centro de Inmunología Molecular, La Habana, Cuba.

<sup>2</sup> Universidad Tecnológica de La Habana “José Antonio Echeverría”, Cuba.

**email:** [lourdes@quimica.cujae.edu.cu](mailto:lourdes@quimica.cujae.edu.cu)

---

El producto Nimotuzumab, altamente demandado para el tratamiento de cáncer de cabeza y cuello, se produce en el Centro de Inmunología Molecular a partir del cultivo, fermentación y purificación en la línea NS0/H7. El uso de tecnologías adecuadas para la producción a gran escala de anticuerpos monoclonales, de acuerdo con las buenas prácticas de manufactura y con costos de producción razonables, se convierte en una necesidad para introducir anticuerpos terapéuticos en mercados de alta demanda y con elevados requerimientos de calidad. El incremento en la presencia de dímeros en el producto final afecta su calidad, por ende, la eficiencia y eficacia del proceso. Se utilizaron las bases de datos del centro para realizar un análisis de componentes principales, empleando el programa *The Unscrambler* 10.4 y se determinaron como variables más influyentes en la variabilidad de los resultados: las conductividades del producto ajustado y de salida en la columna de proteína A y membrana SQ y la masa de IgG en el sobrenadante y el eluato de proteína. Se realizó un análisis de Pareto para determinar los parámetros críticos del proceso, resultando ser: el pH de los tampones de elución y ajuste, del producto ajustado, de la salida en proteína y membrana, la masa de IgG en eluato de proteína y sobrenadante, y la conductividad del producto ajustado.

**Palabras clave:** Anticuerpos monoclonales; purificación; análisis de datos.

---

### Introducción

Los anticuerpos monoclonales (AcM) se descubrieron en la primera mitad de los años setenta por Milstein y Köhler fusionando una línea de células de mieloma murino, sensible a ciertos fármacos, con células de bazo de un animal inmunizado; consiguieron seleccionar solamente las células híbridas y los clones con especificidad conocida. Su trascendencia fue enorme, ya que por primera vez era posible disponer de cantidades ilimitadas de anticuerpos con especificidades precisas. <sup>(1)</sup>

Los procesos de producción de AcM en la industria comprenden dos etapas fundamentales: el cultivo celular y la purificación. El cultivo de células recombinantes de mamíferos es el sistema de expresión más empleado para producir AcM. <sup>(2)</sup>

El producto Nimotuzumab, altamente demandado para el tratamiento de cáncer de cabeza y cuello, se produce en el Centro de Inmunología Molecular a partir del

cultivo, fermentación y purificación en la línea NS0/H7. <sup>(3)</sup> El proceso de cultivo celular comienza con la descongelación de un vial del banco de células de trabajo; después se realiza la expansión de la población celular a través de una serie de siembras en diferentes recipientes de cultivo y se transfiere el cultivo al fermentador de producción, donde las células continúan creciendo y se acumula el producto expresado en el caldo de fermentación.

Existen varios modos de fermentación en función de las estrategias de suministro de medio fresco y de colección del medio enriquecido. Los modos más populares son: cultivo por lote, cultivo por lote alimentado, cultivo continuo y cultivo en perfusión. Las características del crudo cosechado del proceso de fermentación varían en función del medio de cultivo empleado, de la productividad y las características de la línea celular y de las condiciones de fermentación fundamentalmente. <sup>(4)</sup>

---

\* Ingeniera Química, Especialista III en investigación, innovación y desarrollo.

\*\* Doctor en Ciencias Técnicas, Profesor Titular.

El diseño de los procesos de purificación depende principalmente de las características del crudo de fermentación y de la calidad requerida del producto final. En general, se establece una estrategia que incluye tres etapas fundamentales: captura, purificación intermedia y purificación final o “pulido”. El objetivo de la etapa de captura es aislar, concentrar y estabilizar el producto de interés. En la fase de purificación intermedia se eliminan otras proteínas contaminantes. Finalmente, en la fase de purificación final o pulido se remueven trazas remanentes de impurezas y el producto alcanza la calidad final.<sup>(4)</sup>

En la década de 1960, se estudió por primera vez el comportamiento de la agregación de proteínas (oligomerización). Durante el desarrollo de fármacos proteicos comúnmente ocurre el fenómeno de agregación. Los agregados pueden existir inicialmente como pequeños dímeros o fragmentos y progresar hacia estructuras más grandes, como partículas sub-visibles o visibles, si dicha transición es termodinámicamente favorable; o sea, que tenga una variación de energía libre de Gibbs menor que cero.<sup>(5)</sup>

Las proteínas naturalmente tienen regiones hidrófobas tanto en el exterior como en el interior de sus estructuras, pero cuando éstas se despliegan y luego se repliegan incorrectamente, sus estructuras cambian, exponiendo muchos de los “parches” hidrófobos que antes no lo estaban. Estos “parches” se atraen entre sí, lo que hace que las proteínas mal formadas se agrupen. A veces, la agregación se produce cuando las proteínas permanecen intactas, pero se altera el equilibrio de toda la disolución; lo que provoca una cascada de eventos que desencadenan la agregación de las moléculas de proteína.<sup>(6)</sup>

Se ha descubierto que casi la mitad de los AcM farmacéuticamente relevantes y las proteínas de fusión Fc, presentan problemas de agregación luego de la purificación en proteína A.<sup>(5)</sup>

La inestabilidad de las moléculas de proteínas a menudo se manifiesta por la formación de agregados durante el procesamiento y almacenamiento. Los agregados de proteínas muestran una actividad biológica reducida o nula y, lo que es más importante, pueden mostrar una alta inmunogenicidad en la clínica. En consecuencia, es necesario controlar los niveles de agregados y las propiedades de los productos proteicos terapéuticos.<sup>(7)</sup>

Es por ello que en las especificaciones de calidad se establece el porcentaje de agregados admitidos en un producto. El incremento en la exigencia de las

regulaciones lleva a que internacionalmente se exija entre 2-3% de dímeros. Aunque en la producción de Nimotuzumab se cumple con las especificaciones establecidas, en ocasiones se alcanzan valores muy cercanos al máximo permitido, no quedando evidenciados los factores causantes de estas variaciones.

El problema de agregación no se puede eliminar, pues es un fenómeno que ocurre durante todo el ciclo de vida del producto, pero se pueden buscar alternativas que permitan su disminución, si se conocen las variables a controlar.<sup>(6)</sup>

Por otra parte, la tecnología instalada posibilita contar con un gran volumen de información de las variables y parámetros del proceso de fermentación y purificación, lo que permite, a partir de la información almacenada, aplicar el análisis de componentes principales (ACP) para reducir la dimensionalidad de la base de datos e interpretarlos mejor, con una mínima pérdida de información. Este método transforma la información acumulada en unas pocas variables que explican una gran parte de las fluctuaciones de las variables originales y las interrelaciones entre ellas.<sup>(8)</sup>

Aunque la aplicación del ACP es amplia en industrias como la petroquímica, su uso en los procesos de purificación a escala industrial es escaso.<sup>(9,10)</sup>

El objetivo de este trabajo es determinar la influencia de las variables de operación en la variabilidad de los resultados y en la formación de dímeros en la etapa de captura del proceso de producción de un AcM humanizado, el Nimotuzumab.

## **Materiales y Métodos**

### **Descripción del paso de captura del proceso de purificación del Nimotuzumab**

La purificación cromatográfica del AcM Nimotuzumab se inicia con la etapa de captura en la que la molécula se une de forma selectiva y reversible a la región Fc del ligando de proteína A.

Cualquier contaminante que haya quedado unido de forma débil al gel se puede remover en los pasos de lavado. Después, se realiza la inactivación viral en medio ácido, seguida de un ajuste de pH y conductividad, con el objetivo de modificar las propiedades de la disolución de elución de proteína hasta valores adecuados para el paso siguiente. Finalmente, se realiza una cromatografía de intercambio aniónico en membrana Sartobind Q (SQ).

## Metodología empleada para el análisis

A continuación, se abordan dos procedimientos no excluyentes para la interpretación de la formación de dímeros durante el proceso: el ACP<sup>(8)</sup> y la determinación de los parámetros críticos utilizando el diagrama de Pareto.<sup>(11)</sup> Estos procedimientos se realizaron de manera independiente, el primero a partir del registro de datos de las variables medidas en el proceso de purificación y el segundo a partir de la información, mediante entrevistas, brindada por algunos expertos. Ambos procedimientos se explican posteriormente, así como el objetivo de la aplicación de cada uno.

### Procesamiento de la información histórica

En la transformación, el preprocesamiento de los datos y el ACP se emplean los programas Statgraphics Centurion XVII versión 17.2.00 y The Unscrambler versión 10.4, los cuales cuentan con diferentes herramientas para realizar dichas tareas.

Se requiere de un enfoque multivariado de datos para realizar un análisis profundo de la variabilidad del proceso, en toda su complejidad. Se emplearon las pruebas de Kraiser-Meyer-Olkin (KMO) y esfericidad de Bartlett, para determinar la posibilidad de aplicar técnicas de factorización a los datos reportados para la primera parte del proceso de purificación de un AcM. La primera indica el grado en que cada variable de un conjunto se predice sin error por las restantes variables y la segunda se aplica para encontrar la presencia de correlaciones entre las variables.

Se consideró como herramienta el ACP<sup>(8)</sup>, para lograr un modelo de estado descriptivo de los procesos, mediante la reducción de la dimensionalidad de los datos.

Las bases de datos analizadas corresponden a las campañas productivas de cuatro años; contienen 12 parámetros de calidad, 125 registros de cada uno (muestras) y un atributo de calidad, el % de dímeros.

Para el análisis se consideraron todos los parámetros que se registran durante el proceso y son: masa de IgG en el sobrenadante (mtsn), pH del tampón de elución (phglic), conductividad del tampón de elución (cdglic), masa de IgG en el eluato de proteína A (mepa), pH de salida en la columna de proteína A (phmp), conductividad de salida en la columna de proteína A (cdmpa), pH del tampón de ajuste (phtphl), conductividad del tampón de ajuste (cdtphl), pH del producto ajustado (phpaj), conductividad del producto ajustado (cdpaj), pH de salida en la membrana SQ (phpsq), conductividad de salida en la membrana SQ (cdpsq).

## Análisis de componentes principales

En el ACP se selecciona como forma de preprocesamiento a aplicar a los datos el centrado y escalado, se especifica la cantidad de componentes principales a determinar y se emplea la validación cruzada, donde se separan individualmente todas las muestras.

Para la interpretación de los resultados obtenidos primeramente se acude al gráfico de la varianza explicada, el cual permite identificar la cantidad de componentes principales que representan la variabilidad del proceso, incorporando así una validación del resultado.

Luego se analiza el gráfico de la influencia para detectar los puntos discrepantes (outliers). Las muestras con elevada varianza residual o elevado distanciamiento (leverage) se consideran puntos discrepantes (outliers), que no se describen bien por el modelo y deben ser eliminadas del análisis.<sup>(12)</sup>

Seguidamente se estudia el gráfico de las puntuaciones (scores) o mapa de las muestras, para detectar mediante la elipse de Hotelling, con un 95% de confianza, muestras que estén muy alejadas del resto y que pueden ser datos discrepantes o agrupamientos.

Por último, se utiliza el gráfico de los pesos (loadings) o mapa de las variables, con el objetivo de determinar cuánto aporta cada una de las variables a ese componente. El programa permite acceder también al llamado gráfico de los pesos de correlación (correlation loadings), el cual es útil para observar las correlaciones entre las variables y la influencia que cada una tiene sobre el componente en cuestión, de una forma más sencilla. Las muestras que se localicen entre las dos elipses tienen un alto aporte al componente del que estén más cerca, y mientras más próximas al centro se encuentren, menos significativas serán y su aporte a los componentes será pobre.

### Determinación de los parámetros críticos del proceso de purificación

Para determinar el impacto de los parámetros críticos de calidad identificados en la etapa de purificación del proceso productivo de un AcM, sobre el atributo de calidad (% de dímeros), se recurrió a especialistas que poseen un amplio conocimiento del proceso para la valoración del peso de cada variable. Se utilizó la siguiente escala para asignar índices de valoración a las variables de proceso sobre el atributo de calidad: nulo (1), bajo (3), medio (5) y alto (7). Cuando la

variabilidad se consideró que incidía directamente sobre la dimerización, sobrepasando etapas intermedias del proceso, donde no ocurren transformaciones asociadas al parámetro en cuestión, existiendo evidencias de una alta correlación, se evaluó como alto.

El criterio nulo se consideró si la variabilidad incidía poco en el desempeño de etapas del proceso, sin afectar los agregados, existiendo evidencias de que no hay una correlación.

**Diagrama de Pareto**

El diagrama de Pareto es un método de análisis gráfico de amplia difusión, que permite separar entre las causas más importantes de un problema (las pocas que son vitales) y las que lo son menos (las muchas triviales). Se establece como principio que el 80% de las variaciones en un proceso puede ser explicado por el 20% de las causas de dichas variaciones.<sup>(11)</sup>

Para la categorización de los parámetros en la primera parte del proceso de purificación, se elaboró un gráfico de Pareto con los índices de prioridad reportados por los expertos y su correspondiente porcentaje acumulativo del gran total; para luego aplicar el criterio 80-20%, y obtener el conjunto de parámetros de mayor criticidad respecto a la variabilidad del proceso en la etapa de purificación.

**Resultados y Discusión**

**Preprocesamiento de los datos**

Se realizó la prueba estadística de Kraisner-Meyer-Olkin (KMO) y se obtuvo un índice KMO de 0,72. Este valor

clasifica el conjunto como regular, pues se encuentra en el intervalo entre 0,70 y 0,79. La prueba de esfericidad de Bartlett arroja un resultado del p-valor de cero, menor que 0,05. Estos dos criterios permiten justificar la aplicación del ACP a los datos estudiados como técnica para reducir la dimensionalidad.

**Aplicación del análisis de componentes principales**

Los datos de las campañas productivas de los cuatro años se estudiaron mediante el ACP, donde se comenzó con el análisis de 12 parámetros que se registran para el control de la calidad del proceso y 125 registros de cada parámetro (muestras). Todas las variables se centran por el mínimo y escalan por el rango. El análisis de la bibliografía permitió encontrar la recomendación de centrar por el mínimo y escalar por el rango, cuando los valores negativos de un centrado por la media no poseyeran significación física, como es el caso de la concentración.<sup>(12)</sup>

En la Figura 1 se muestra el gráfico de la influencia, donde no existen muestras en la sección superior derecha, que constituyan puntos discrepantes peligrosos. Las muestras se observan como series temporales, identificándolas por colores, según el año en que se produce cada lote.

La muestra 100 presenta elevado valor de distanciamiento, siendo un punto discrepante influyente. La columna de captura perdió el empaque por defectos del equipamiento y trajo como consecuencia que, en este lote, el producto obtenido en dicha etapa resultó altamente diluido. Esta es una situación inusual en el proceso y se eliminó este punto del análisis, ya

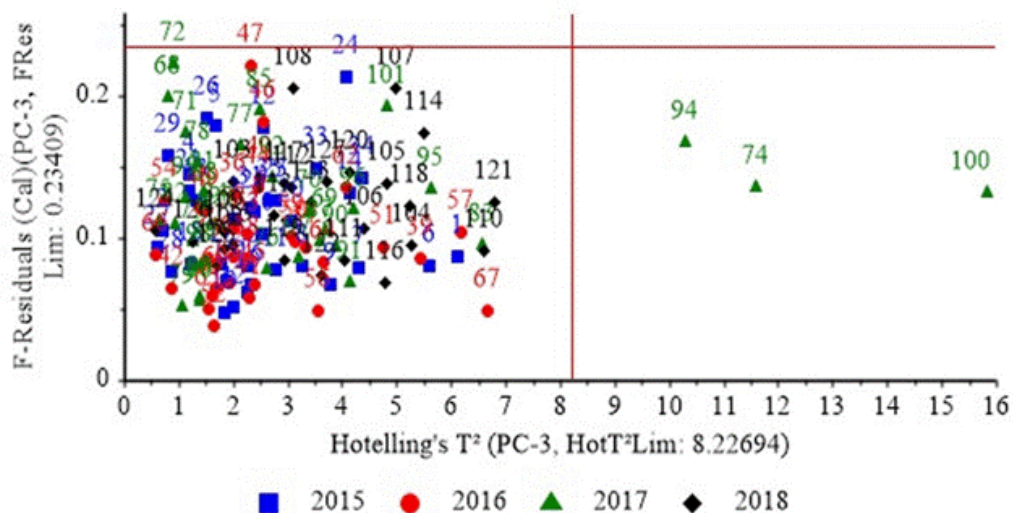


Fig. 1. Gráfico de la influencia para 12 variables y 125 muestras.

que no es representativo del mismo; es válido destacar que, al concluir el proceso de purificación de este lote, la vacuna cumplió todas las especificaciones de calidad. El punto 74 presenta distanciamiento, pero puede contener información importante, por lo que no se eliminó del análisis. El punto 94 constituye un punto discrepante. En este lote se realizó el proceso de purificación de forma continua, pues debido a la alta demanda se hizo necesario acortar los tiempos de producción, sin que se pusiera en riesgo la seguridad del proceso. Los análisis de liberación se realizaron de inmediato y se continuó el proceso. Esta es una situación que no ocurre frecuentemente, por lo que se decidió eliminar esta muestra del análisis.

Luego de eliminar las muestras discrepantes influyentes (94 y 100) se realizó nuevamente el ACP y se volvió a examinar el gráfico de las influencias verificando que no se aprecian discrepantes peligrosas.

Se realizó el análisis del gráfico de puntuaciones para las 12 variables y 123 muestras sin emplear la elipse de Hotelling (Fig. 2), con el objetivo de obtener una mayor visibilidad de la distribución de los puntos, y determinar de forma más precisa posibles agrupamientos.

La mayoría de las muestras del primer año que presentan valores de conductividad del producto de salida en la membrana SQ fuera del intervalo recomendado, tienen valores negativos en ambos componentes, exceptuando la muestra 14, que posee el valor de conductividad correspondiente al límite inferior.

En el segundo año ocurre lo mismo y se exceptúan las muestras 52, 58 y 60, que presentan valores cercanos al límite inferior de especificación. En el tercer año se

aprecia el mismo comportamiento y se exceptúan las muestras 80 y 98, que poseen valores correspondientes al límite inferior.

Se observa que la gran mayoría de los lotes correspondientes al cuarto año, presentan valores positivos en los dos componentes y todos ellos poseen valores de conductividad en el intervalo establecido. Mientras, las muestras 103, 124 y 126, tienen valores negativos en el segundo componente, pues se encuentran fuera de las especificaciones establecidas.

De manera general, se puede decir que las muestras que presentan valores de conductividad del producto de salida en la membrana SQ en el intervalo de especificación establecido, poseen valores positivos en ambos componentes, exceptuando las que tienen valores cercanos a los límites establecidos.

En la Figura 3, que muestra el gráfico de la varianza explicada, se observa que con dos componentes principales se logra explicar el 89,81% de la varianza de los datos iniciales. El primer componente explica el 83,58% y el segundo el 6,42%. Las curvas de calibración y validación del modelo no convergen totalmente, por lo que el modelo describe bien los datos analizados, pero es posible que no nuevos datos.

En el gráfico de los pesos correlacionados (Fig. 4) se observa el comportamiento de las variables, donde existe una alta correlación entre las masas de IgG en el sobrenadante y en el eluato de proteína A, así como entre las conductividades de los productos de salida en la columna y en la membrana SQ. Todos los parámetros que se encuentran en el interior de la primera elipse no contribuyen significativamente a la variabilidad

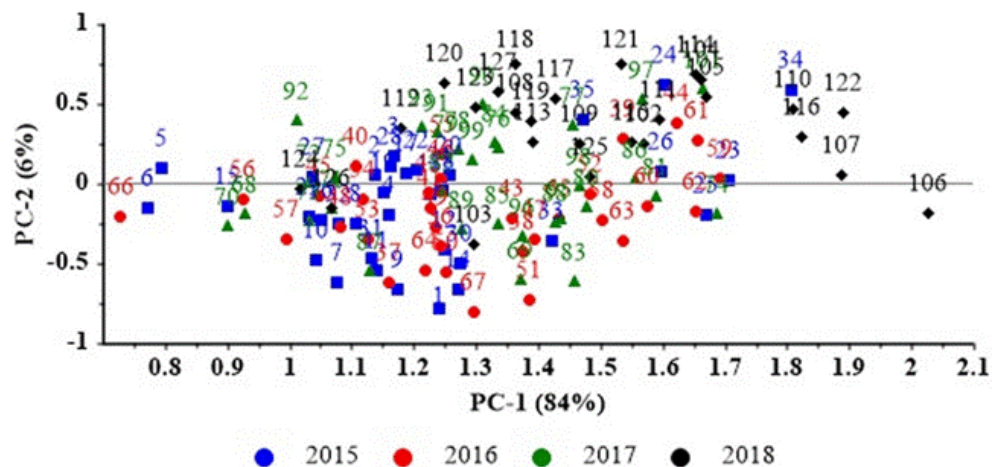


Fig. 2. Gráfico de las puntuaciones sin elipse de Hotelling para 12 variables y 123 muestras.

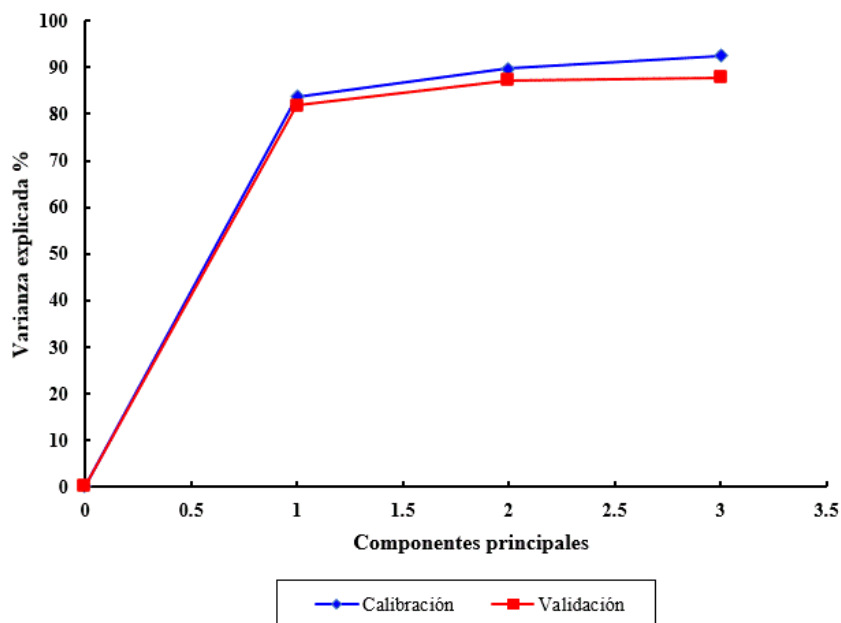


Fig. 3. Gráfico de la varianza explicada para 12 variables y 123 muestras.

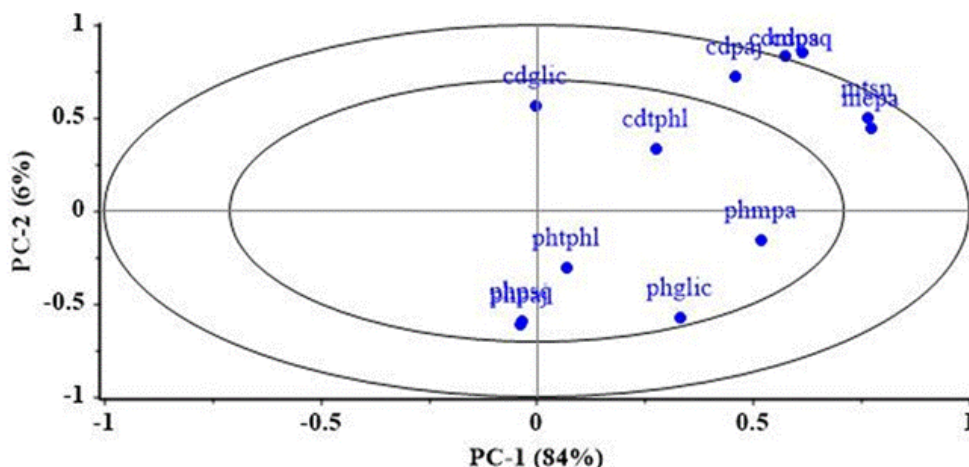


Fig. 4. Gráfico de los pesos correlacionados para 12 variables y 123 muestras. phtphl: pH del tampón de ajuste, mtsn: masa de IgG en el sobrenadante, phglic: pH del tampón de elución, cdglic: conductividad del tampón de elución, mepa: masa de IgG en el eluato de proteína A, phmpa: pH de salida en la columna de proteína A, cdmpa: conductividad de salida en la columna de proteína A, cdtphl: conductividad del tampón de ajuste, phpaj: pH del producto ajustado, cdmpaq: conductividad del producto ajustado, phpsq: pH de salida en la membrana SQ, cdpsq: conductividad de salida en la membrana SQ.

del proceso. Se aprecia cómo las conductividades del producto ajustado, de los productos de salida en la columna y en la membrana, así como las masas de IgG en el eluato de proteína y en el sobrenadante, son las que aportan mayor variabilidad al proceso.

Las variables masas de IgG en el sobrenadante y en el eluato de proteína se encuentran relacionadas de manera directa en ambos componentes, lo cual tiene lógica, ya que la segunda mencionada varía con la masa de IgG

que se procese al inicio de la purificación (sobrenadante) en dependencia de la eficiencia de la columna.

El parámetro conductividad del producto ajustado influye sobre la variabilidad del proceso de forma positiva en ambos componentes, siendo esto lógico; ya que, en la etapa de inactivación viral y ajuste de pH y conductividad se producen cambios en la fuerza iónica al preparar la muestra para el paso de intercambio iónico en cromatografía de membrana, lo que puede contribuir a la formación de dímeros.

Las conductividades de los productos de salida en la columna de proteína A y en la membrana SQ presentan una alta correlación, siendo las que aportan mayor variabilidad al proceso y las de más alto peso para el CP2.

Las variables pH del tampón de elución y del producto de salida en la columna de proteína A se encuentran relacionadas de forma directa en el CP1, lo que explica el fenómeno físico que se esperaba encontrar; debido a que el valor de la última mencionada depende de la masa total del tampón de elución añadido.

Los pesos de cada una de las variables de los dos primeros componentes, se muestran en la Tabla 1. Se aprecia que en el CP1 las variables que más influyen están relacionadas con la cantidad de producto, mientras que en el CP2 se relacionan con sus propiedades, en específico la conductividad.

El obtener que la conductividad del producto de salida en la membrana sea la tercera variable de mayor valor en el primer componente, significa que este componente expresa que dicha variable tiene alto peso en la variabilidad del proceso; aunque no llega a alcanzar valores mayores o iguales a 0,7, que son los que el

**Tabla 1.** Pesos de las variables en cada componente.

Variables	CP1	Variables	CP2
mepa	0,7716	cdpsq	0,8531
mtsn	0,7623	cdmpa	0,8314
cdpsq	0,6116	cdpaj	0,7183
cdmpa	0,5752	cdglic	0,5710
phmpa	0,5210	mtsn	0,5011
cdpaj	0,4583	mepa	0,4489
phglic	0,3346	cdtphl	0,3244
cdtphl	0,2762	phmpa	-0,1563
phtphl	0,0698	phtphl	-0,3025
cdglic	-0,0068	phglic	-0,5705
phpsq	-0,0332	phpsq	-0,5963
phpaj	-0,0380	phpaj	-0,6250

mepa: masa de IgG en el eluato de proteína A, mtsn: masa de IgG en el sobrenadante, cdpsq: conductividad de salida en la membrana SQ, cdmpa: conductividad de salida en la columna de proteína A, phmpa: pH de salida en la columna de proteína A, cdpaj: conductividad del producto ajustado, phglic: pH del tampón de elución, cdtphl: conductividad del tampón de ajuste, phtphl: pH del tampón de ajuste, cdglic: conductividad del tampón de elución, phpsq: pH de salida en la membrana SQ, phpaj: pH del producto ajustado.

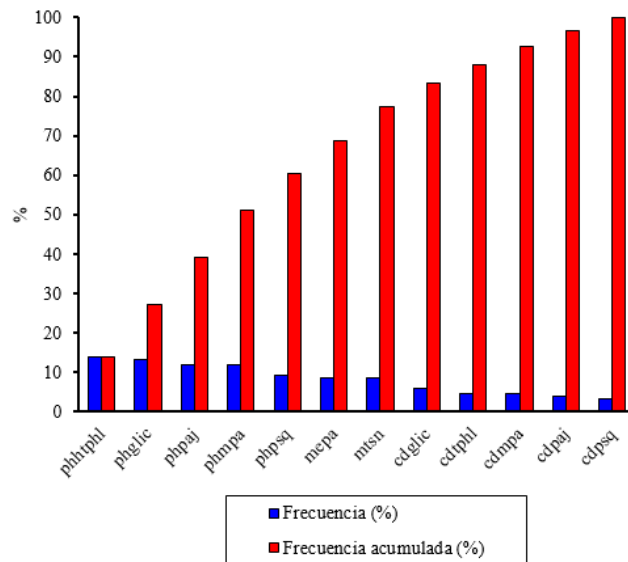
análisis de correlaciones con las elipses mostradas en la Figura 5, asumen como que aportan gran variabilidad al proceso. Este hecho se corrobora al analizar el segundo componente, donde se obtiene que esta variable es la de mayor peso, por lo que es una variable importante, ya que en el CP2 es la de mayor peso y en el CP1 no se toma por simple tecnicismo.

### Determinación de los parámetros críticos del proceso de purificación

Para la determinación de los parámetros críticos se consultaron seis expertos, que poseen gran experiencia en el proceso, con 15,7 años (desviación estándar 5,3) como promedio. Se estableció la influencia de los parámetros críticos sobre el atributo de calidad, según el criterio de los expertos. Se empleó como criterio de selección la mediana de los valores reportados, en los casos en que las opiniones de los expertos no coincidieran.

### Diagrama de Pareto

Se empleó el programa Statgraphics Centurion XVII versión 17.2.00, y se obtuvo el diagrama de Pareto que se aprecia en la Figura 5. Se observan las variables de mayor



**Fig. 5.** Gráfico de Pareto para las variables. phtphl: pH del tampón de ajuste, mtsn: masa de IgG en el sobrenadante, phglic: pH del tampón de elución, cdglic: conductividad del tampón de elución, mepa: masa de IgG en el eluato de proteína A, phmpa: pH de salida en la columna de proteína A, cdmpa: conductividad de salida en la columna de proteína A, cdtphl: conductividad del tampón de ajuste, phpaj: pH del producto ajustado, cdpaj: conductividad del producto ajustado, phpsq: pH de salida en la membrana SQ, cdpsq: conductividad de salida en la membrana SQ.

influencia sobre el atributo de calidad, lo que se justifica a partir del criterio de Pareto, donde los parámetros que se encuentran por debajo del 80 % acumulativo tienen mayor impacto sobre el atributo de calidad y deben tener un estricto control.<sup>(13)</sup> Los parámetros críticos del proceso obtenidos fueron: pH de los tampones de ajuste y elución, del producto ajustado, de los productos de salida en la columna de proteína A y en la membrana SQ, así como la masa de IgG en el eluato de proteína A y en el sobrenadante, y la conductividad del producto ajustado. Las variables conductividad del tampón de elución, del producto ajustado, de los productos de salida en la columna de proteína A y en la membrana SQ se consideran parámetros claves de operación; ya que influyen en el desempeño del proceso, pero no de forma significativa en la calidad del producto.

## Conclusiones

Se concluye del análisis exploratorio de datos, que con dos componentes principales se explica el 89,81 % de la varianza, y las variables conductividades del producto ajustado y de salida en la membrana SQ, masas de IgG en el eluato de proteína A y en el sobrenadante, son las que aportan mayor variabilidad al proceso.

Se estudiaron doce parámetros de calidad del proceso, mediante el análisis de Pareto. Las variables: pH de los tampones de ajuste y elución, del producto ajustado, del producto de salida en la columna de proteína A y en la membrana SQ, así como las masas de IgG en el eluato de proteína A y en el sobrenadante y la conductividad del producto ajustado resultaron ser parámetros críticos; mientras que las conductividades del tampón de ajuste, del producto ajustado, de los productos de salida en la columna de proteína A y en la membrana SQ resultaron ser parámetros claves.

## Conflicto de intereses

Los autores no declaran conflictos de intereses.

## Referencias

1. García A. Anticuerpos monoclonales. Aspectos básicos. *Neurología*.2011;26(5):301-6.
2. Li F, Zhou J, Yang X, Tressel T, Lee B. Current therapeutic antibody production and process optimization. *BioProcess. J.* 2005;4(5):1-8.
3. Viada C, Vega AM, Robaina M, Frías A, Álvarez M, Santiesteban Y, et al. Evaluación de Nimotuzumab para el tratamiento de cáncer de cabeza y cuello: Meta-análisis de ensayos controlados. *Bionatura*.2020;5:1056-62. doi: 10.21931/RB/2020.05.01.8
4. Mora O. Impacto de un medio de cultivo alternativo en el proceso de purificación del Nimotuzumab [Tesis para optar por el título de

Ingeniero Químico]. La Habana: Universidad Tecnológica de La Habana “José Antonio Echeverría”; 2015.

5. Courtois F, Agrawal N, Lauer T, Trut B. Rational design of therapeutic mAbs against aggregation through protein engineering and incorporation of glycosylation motifs applied to bevacizumab. *MAbs*.2016;8:99-112.
6. González M. Is your biologic at risk for protein aggregation? Part 1 [monograph on the Internet]. *Pharmaceutical Online*; 2018. Disponible en: <http://www.pharmaceuticalonline.com/doc/>. (Consultado en línea: 15 de abril de 2019).
7. Liu B, Guo H, Xu J, Qin T, Xu L, Zhang J, et al. Acid-induced aggregation propensity of nivolumab is dependent on the Fc. *MAbs*.2016;8:1107-17.
8. Tharwat, A. Principal component analysis-a tutorial. *Int. J. Appl. Pattern Recognit.* 2016;3(3):197-240.
9. Rodríguez RH, Gozá O, Ojito E. Preliminary modeling of an industrial recombinant human erythropoietin purification process by Artificial Neural Networks. *Brazi. J. Chem. Eng.*2015;32(3):725-34. <http://dx.doi.org/10.1590/0104-6632.20150323s00003527>.
10. Gozá O, Fernández M, Rodríguez EH, Ojito E. Aplicación del Análisis de Componentes Principales en el proceso de purificación de un biofármaco. *VacmiMonitor*.2020;29(1):5-13. Disponible en: <http://www.vacmiMonitor.finlay.edu.cu/index.php/vacmiMonitor/article/view/226>.
11. AITECO Consultores S.L [homepage on the Internet] Granada: Herramientas de calidad. Disponible en: <https://www.aiteco.com/diagrama-de-pareto/> (Consultado en línea: 14 de abril de 2020).
12. Jolliffe IT. *Principal Component Analysis*. Second Edition. New York: SpringerVerlag; 2002.
13. López LE, Zumalacárregui L, Pérez O. Análisis de componentes principales aplicado a la fermentación alcohólica. *Revista Científica de la UCSA*.2019;6(2):11-9. doi:10.18004/ucsa/2409-8752/2019.006.02.011-019.



---

## **Application of Principal Component Analysis on dimers formation during a monoclonal antibody's purification**

### **Abstract**

Nimotuzumab (hR<sub>3</sub>) is produced at the Center of Molecular Immunology (CIM). It is obtained from the culture, fermentation, and purification in the cell line NS0/H7. This product has a high demand both domestically and internationally for the treatment of head and neck cancer. The use of adequate technologies for the large-scale production of this product following good manufacturing practices and reasonable production costs becomes a necessity to introduce this therapeutic antibody in high demand markets with high-quality requirements. The increase in dimers in the final product will affect its quality, which in turn affects the efficiency and effectiveness of the process. Production databases were used to perform principal component analysis using *The Unscrambler* 10.4 software. The conductivity of the adjusted product, pH at the output product in the protein column and membrane, the mass of supernatant, and of the eluate are detected as the variables with a higher influence on aggregate formation. The Pareto analysis was carried out to determine the critical parameters of the process: pH of elution and adjust buffers, and adjusted products, pH at the exit of protein column and membrane, the mass of protein in the eluate, and in the supernatant, and the conductivity of the adjusted product.

**Keywords:** Monoclonal antibodies; purification; data analysis.

---

*Recibido: 21 de Abril de 2020*

*Aceptado: 12 de Junio de 2020*