

Proteómica - Identificación de proteínas y fosfoproteómica (Javier Muñoz, PhD – clase, Nuria Ibarz Ferrer, nibarz@cniio.es, práctica).

Por Mauricio Lema Medina – 02.2013

Nota de quien resume

Pocas clases del Máster de Oncología Molecular han sido más complejas que esta (si las ha habido, no las recuerdo). En parte, porque hace uso de avances tecnológicos de física y química con las que no estamos tan familiarizados como desearíamos. La claridad del expositor brilla por su ausencia. No es su responsabilidad, él presume (erróneamente, en mi caso) algún grado de conocimiento sobre temas como cromatografía, entre otras. La clase no hubiera sido posible en el tiempo asignado sin esas presunciones básicas. En la práctica en Madrid entre 04 y 08 de Febrero de 2013, nos fue explicada la espectrometría de masa por Nuria Ibarz Ferrer. Si bien tampoco logré extraer demasiada claridad de la misma, la combinación de la clase magistral de Javier Muñoz y Nuria Ibarz, permitieron que tuviera una casi epifanía (un medio entendimiento). Me apliqué durante unos minutos a estudiar los conceptos básicos de cromatografía, que resumí en el apéndice 1, y ello me permitió entender muchas cosas esenciales para tener una buena idea de la clase. Busqué en varias partes explicaciones diversas sobre la espectrometría de masas, MALDI y ESI. Incorporé muchos de esos conceptos en el documento. Es posible que haya duplicaciones innecesarias de conceptos, pero considero que la repetición puede ser útil para facilitar el entendimiento de conceptos complejos (y creo que aquí aplica la palabra compleja desde el título). Considero que la mejor forma de leer el documento es iniciar por el Apéndice 1, para familiarizarse con la cromatografía que no es el tema de la conferencia, pero es esencial para su entendimiento. Posteriormente, se debe leer el articulo que les propongo. Quiero explicar que en la clase se explicó que la HPLC se aplicó a las muestras después de su digestión por Tripsina. En la práctica, sin embargo, se le aplicó la HPLC a las proteínas antes de la lisis peptídica. Suena más razonable después de la lisis, pero creo que ambas son pertinentes según el problema a analizar. Finalmente, este documento es escrito por un total neófito en el tema, y puede contener muchísimos errores, problemas de indio, no de arco ni flecha, por el que doy disculpas de antemano.

MLM

Generalidades

Proteoma: es el conjunto de proteínas en un tipo celular dado, en un momento y condiciones definidas. La importancia de las proteínas radican en que son los efectores moleculares de los genes, constituyéndose en el principal nivel de información de cómo las células trabajan. Son importantes: su abundancia, localización, interacciones entre sí, isoformas de splicing (la traducción es tan atroz que prefiero el anglicismo, sorry para los hispanófilos), y sus modificaciones postraduccionales.

Como su nombre lo indica, la proteómica busca el estudio del proteoma, que incluye todas las isoformas proteicas, interacciones, localización y modificaciones postraduccionales. La proteómica tiene un amplio ámbito de acción pues puede informar sobre la presencia de proteínas y sus diferentes isoformas en una muestra biológica, evaluar expresión diferencial, establecer la abundancia absoluta, la dinámica temporal de expresión proteica y su localización espacial. En el pasado, las proteínas se identificaban por secuenciación del extremo N-terminal, detección con anticuerpos específicos, composición de aminoácidos y delección de genes. Pero estos métodos son lentos, laboriosos y costosos. Además, poco viables para muestras complejas.

El proteoma es varios órdenes de magnitud más complejo que el genoma, así: se estima que hay unos 20.500 genes que dan origen a unas 10^5 isoformas, con un rango dinámico de expresión de 10^9 (relación entre abundancia de la proteína más expresada y la menos expresada). Se estima que hay aproximadamente 1 millón de modificaciones postraduccionales y 50 millones de interacciones proteína-proteína. Por si fuera poco, es altamente dinámico en tiempo. La proteómica no es simple. Cómo se enfrenta al reto? Utilizando tres estrategias: 1. Separación de muestras, 2. Espectroscopia de Masas y 3. Análisis de data.

Hoy en día, debido a su rapidez y elevada sensibilidad, la **espectroscopia de masas** se ha convertido en el método por excelencia para la identificación de proteínas a gran escala. Es un método rápido para muestras complejas, y tiene posibilidad además de caracterizaciones de modificaciones post-traduccionales como glicosilación y la fosforilación.

Qué es la espectroscopia de masas, y por qué es importante en proteómica?

Espectroscopia de masa (MS) es la ciencia de mostrar el espectro de masas de las moléculas que están en un material de estudio. En proteómica la MS permite la identificación precisa de proteínas en una muestra, así como la presencia de modificaciones postraduccionales de las mismas como fosforilación, glicosilación, etc. Las bases teóricas de la MS son complejas pero se pueden resumir en que se les transfieren protones a las moléculas interrogadas, que están en fase gaseosa, y se les aplica una corriente negativa que las atrae el detector en función de su masa. La precisión es inmensa.

Cómo funciona el espectroscopio de masa de masa?

El principio es relativamente simple. La muestra debe contener moléculas de peso molecular adecuado (no demasiado grandes), ionizadas y en el vacío. Se les aplica diferentes formas de energía electromagnética que hace que las moléculas a estudiar se dirijan al detector. La masa de cada una de las moléculas de la muestra determina el momento exacto en que impactan al detector (ver abajo). Para entender cómo funciona la proteómica en forma práctica, veamos paso a paso cómo se aplica.

Problema

Al laboratorio de proteómica llegan otros investigadores con la petición de que se extraiga una(s) proteínas específicas de alguna muestra que puede ser líneas celulares, tejidos de diversa naturaleza, FACS (citómetro de flujo) y fluidos corporales. El investigador desea interrogar por la presencia de una proteína de interés en la muestra. Para poder utilizar la espectroscopia de masa será necesario reducir el tamaño de proteínas a péptidos. Para esto pueden ser necesarios varios pasos preliminares sobre muestras (particularmente si son complejas).

Procesamiento inicial de la muestra

1. Extracción proteica: consta de 2 fases, la lisis celular y la extracción proteica. Se inicia el proceso con lisis celular de las muestras por medio de morteros cerámicos. Se

introduce la muestra a analizar en un pequeño tubo donde hay una bola cerámica, y se introduce en una máquina que agita la muestra machacándola. El resultado es una muestra homogénea en fase líquida.

Posteriormente, se continúa con el fraccionamiento proteico por diversas técnicas según el caso: separación por geles de agarosa basados en el tamaño cuando se conoce el peso molecular de la proteína buscada (tamaño aparente). O con la combinación de geles bidimensionales que separan por peso molecular (MW) y por carga por su movilidad electroforética, también por geles.

Otras muestras pueden separar sus proteínas por medio de Cromatografía Líquida de alto desempeño o HPLC (High Performance Liquid Chromatography) en la que el lisado celular es adsorbido a unas partículas pequeñas de 5 micrómetros (que tienen unos surcos que capturan proteínas según su tamaño) (**Ver apéndice 1 para explicación sobre HPLC**). Las partículas son actuadas por una corriente líquida en el cromatógrafo y las proteínas se van precipitando en un sitio específico según su masa, configurando un cromatograma. ***La HPLC frecuentemente es también realizada después de la digestión por proteasas para la separación de péptidos, para maximizar la separación (ver más adelante).***

En otras situaciones se pueden separar por métodos no basados en gel como la cromatografía de gas líquido o por coimmunoprecipitación de proteínas que interaccionan, con la utilización de anticuerpos específicos y lavados sucesivos de los sobrenadantes no coimmunoprecipitados.

2. Digestión por proteasas: La digestión con Tripsina que es una proteasa que corta en los extremos carboxiterminal de los residuos de Lysina (K) y Arginina (R), generando péptidos de tamaño adecuado para su detección por espectroscopia de masas. Para ello, extrae del gel (en caso de que allí se halla realizado la separación de la proteína candidata), se incuba a 37 grados Celsius toda la noche con Tripsina teniendo en cuenta que se deben extremar los esfuerzos para no generar muestras contaminadas por cabello o piel de quien realiza el experimento. Cuando la muestra está en solución casi siempre se requiere de doble digestión con Lys C (a 21 grados Celsius) seguida por la de Tripsina ya mencionada, aumentando la eficiencia de la generación de péptidos del tamaño apropiado. Al final del proceso de separación tenemos una colección de péptidos. Como mencionamos previamente, podemos separar estos péptidos utilizando HPLC.

Cómo identificamos los péptidos obtenidos?

La espectrometría de masa es la técnica que mide las masas de moléculas y permite la identificación de compuestos desconocidos. Se utiliza para el control antidoping, investigación forense, guerra química y en la investigación espacial. El espectrómetro de masas consta de tres componentes: **una fuente, un analizador de masas y un detector**. Para que se pueda medir la masa de una molécula, primero hay que ionizarla y convertirla de fase sólida o líquida a fase gaseosa.

Fuente: Ionización y gasificación de los péptidos en la muestra.

La fase siguiente requiere de Hay 2 estrategias: MALDI y Electrospray. Veámoslas someramente. MALDI: Matrix-assisted laser desorption-ionization: Es una técnica de ionización suave utilizada en espectroscopia de masa que permite el análisis de biomoléculas como proteínas (entre otras). En el MALDI hay dos pasos: 1. Un haz de láser UV impacta la muestra liberando (desorption) su capa más superficial (también conocida como ablación de la capa de superficial de la matrix). El calor hace que la matriz liberada tenga moléculas que se convierten en pequeñas gotas (nanodroplets). 2. Las moléculas a analizar (denominados analitos) se ionizan (protonización o deprotonización).

Otra forma de ionización y gasificación es el Electrospray (ESI) en el que una columna líquida de alta presión (que puede venir de un HPLC, por ejemplo) pasa por un agujero pequeño creando una gotas gaseosas a las que se les aplica una carga eléctrica, ionizándolas.

Analizador de masas y detección: Espectrometría de masa

No importa la fuente, MALDI o Electrospray, ahora se procede a la Espectroscopia de Masa propiamente dicha (que ya mencionamos arriba): Una vez tenemos los analitos ionizados, se les aplica campos electromagnéticos en diversas configuraciones en el vacío causando que los iones se desplacen hacia el detector (estas configuraciones tienen diferentes nombres como Time Flight, Cuadrupolo, Orbitrack, etc). La velocidad con la que se desplazan depende de su masa. Al golpear el detector se genera un espectro de ondas que indican la masa y la cantidad de cada uno de las moléculas (en este caso, péptidos) ionizados. El cronómetro tiene una precisión de nanosegundos. La masa de cada uno de los analitos se establece con altísimo grado de precisión. ***Cada péptido tiene una masa que es como una huella sólo compartida por unos pocos péptidos.***

Cómo se obtiene la secuencia peptídica a partir del la firma espectral?

Proteolisis in-silico

Con la secuenciación de los genes por el proyecto genoma (y otros) se puede proceder con la digestión “teórica” de los genes en las bases de datos por “Tripsina” in-silico (en el computador). Se generan librerías de los cientos de miles de péptidos inferidos. Se calcula la masa de cada péptido con toda la precisión. Se confrontan las masas obtenidas experimentalmente con los obtenidos por el proceso in-silico. Esto hace que los péptidos candidatos sean unos pocos (aquellos péptidos de las librerías que tienen las masas iguales – o extraordinariamente similares – a las observadas). Dicho de otra manera (qué pena para los que entendieron con la primera explicación), las masas obtenidas con la MS se cotejan con las librerías de péptidos con la ayuda de un robusto departamento de bioinformática que le permite su identificación (con su secuencia de aminoácidos). Este proceso se denomina ***interpretación del espectro de fragmentación***. Esas mismas herramientas bioinformáticas permiten establecer a qué proteínas pertenecen los péptidos por medio del análisis de digestión teórica con masas teóricas.

Qué pasa cuando varios péptidos de las librerías tienen masas iguales a las obtenidas en la MS?

Para obtener la información sobre la composición del péptido interrogado, éste se debe romper en pedazos más pequeños, y medir su masa con la misma metodología (MS). Esto se realiza por el método de *fragmentación controlada de péptidos con gases inertes (como el Helio)*. Los espectros de masa de estos péptidos superfragmentados (denominados MS/MS por la doble espectrometría de masas a la que se someten) son también analizados por herramientas bioinformáticas que permite inferir con alta precisión la secuencia de aminoácidos.

Aplicaciones del MS en proteómica

La proteómica permite la identificación de proteínas (perfilación), Con variaciones sobre esta estrategia se pueden obtener un análisis de fosforilación de proteínas denominadas fosfoproteómica. También se pueden detectar otros cambios postraduccionales como glicosilación, interacción proteínas proteínas, etc. La cuantificación por Label-Free es un método que busca diferencias significativas entre muestras complejas basándose únicamente en los espectros obtenidos por cromatografía líquida y espectroscopia de masas.

Retos

El principal reto del laboratorio de proteómica es extraer la proteína con aceptable nivel de pureza para su análisis. Ello se dificulta cuando la muestra es compleja, el rango dinámico es elevado (coexistencia de proteínas en altísima concentración cuando la que se busca extraer está en muy baja concentración), analitos de baja concentración o poca cantidad de muestra.

Actualmente, la proteómica permite la perfilación de proteomas completos (con aproximadamente 10.000 proteínas), el análisis de modificaciones postraduccionales (30.000 sitios de fosforilación, 4000 sitios de K-acetilación, 7000 sitios de N-glicosilación), y análisis PPI que permiten reportar miles de interacciones proteína proteína. Todavía hay posibilidad de mejora en la sensibilidad y velocidad de las MS, se puede mejorar el fraccionamiento gaseoso que aumente el rango dinámico de detección, y el uso de preparaciones de ultra alta presión que ya se están convirtiendo en estándar. La preparación de una buena muestra sigue siendo el paso crítico para un buen experimento en MS. Hacia dónde nos dirigimos? Hacia la comunicación entre diferentes modificaciones postraduccionales, otras cosas que no entendía bien que vienen el futuro (personalmente, estoy apabullado con el presente, así que no me voy a preocupar demasiado por el futuro).

Aplicaciones de la proteómica en oncología

Se incluyen la clasificación de tumores, medición de actividad tirosina kinasa, estudio de los efectos de drogas (terapéuticos y tóxicos), identificación de biomarcadores. Se dan varios ejemplos que no reflejan el uso actual de la proteómica sino que están en el mundo de la frontera científica (todos los artículos citados son de Nature Biotechnology y Cell).

Apéndice 1 - Entendiendo la Cromatografía Líquida de Alto Desempeño (HPLC)

El HPLC es simplemente una versión sofisticada de la cromatografía de capa delgada y la cromatografía de columna. Veamos éstas primero, para poder entender la HPLC.

Cromatografía de capa delgada o fina (Thin Layer Chromatography, TLC). La cromatografía se utiliza para separar mezcla de sustancias en sus componentes. En todas las cromatografías hay una fase estacionaria (un sólido, o un líquido apoyado en un sólido) y una fase móvil (un líquido o un gas). La fase móvil fluye a través de la fase estacionaria llevándose en ella los componentes de la mezcla. Cada componente viaja a diferente velocidad.

En la TLC se utiliza una capa fina, uniforme de gel de sílica o aluminio sobre un soporte rígido. El gel de sílica o aluminio es la fase estacionaria, y con frecuencia contiene una sustancia que florece con luz ultravioleta. La fase móvil es un solvente líquido apropiado, o una mezcla de solventes. Se coloca un punto de la mezcla en la placa de gel. La placa delgada se coloca en un beaker con el solvente cuidando de que el solvente no moje el punto de la mezcla, como se ilustra en la figura. Luego se cierra el beaker para que el solvente pueda subir por la placa de sílica aprovechando sus propiedades químicas especiales (sílica es el resultante de dióxido de sílica unidos entre sí por enlaces covalentes, pero en la superficie no los oxígenos están libres convertidos en $-OH$ muy polares que permiten la formación de puentes de hidrógeno con solventes apropiados, permitiendo que estos se desplacen a través de toda la capa sílica)

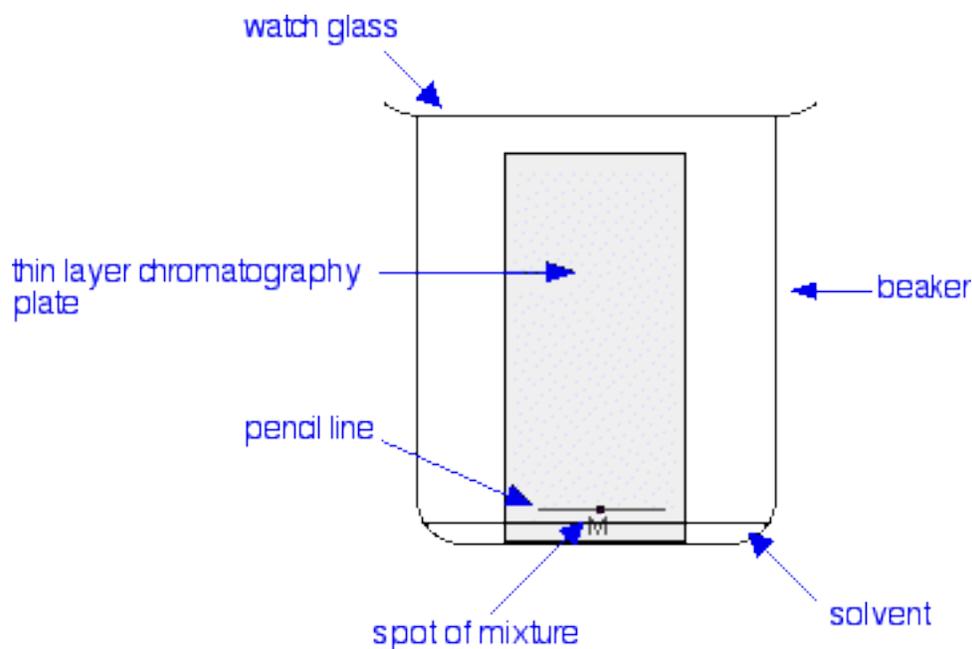


Figura 1. Cromatografía de capa fina, al iniciar el experimento.

El movimiento del solvente sobre la muestra hace que sus componentes se separen dependiendo de la solubilidad en el solvente de cada uno de los componentes y qué tanto se pega el compuesto a la matriz de sílica (en forma técnica, cual es el grado de adsorción al gel sílico). Los compuestos son separados en sus elementos que pueden ser reconocidos de muchas maneras (Figura 2). Una de ellas es por la aplicación de luz UV que permite que los sitios donde hay grumos sean identificados fácilmente (siempre y cuando se haya marcado el gel sílico con sustancias fluorescentes).

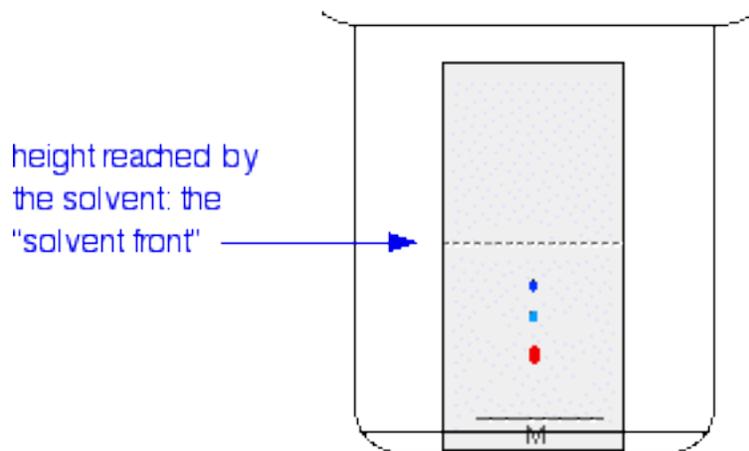


Figura 2. Cromatografía de capa fina, luego de ejecutado el experimento, con la separación de la muestra en tres puntos discretos.

Cuando el solvente llega al final de la capa, se puede extraer y cuantificar la migración de la sustancia calculando los valores R_f que son simplemente la relación entre la distancia desde el origen del compuesto y la distancia recorrida por el solvente. Se supone que los R_f son constantes (siempre y cuando las condiciones del experimento sean exactamente iguales).

Cromatografía de columna

Como vimos en la TLC la fase estacionaria es una capa fina de geles de sílica o aluminio en una placa de vidrio, metal o plástica. En la cromatografía de columna se ubican los mismos componentes en un columna vertical de vidrio, totalmente saturados de solvente y con una válvula cerrada en la parte inferior del tubo de vidrio (para evitar que se escape el solvente). La muestra se mezcla con el solvente, y se abre la llave del tubo permitiendo que el solvente salga (la fase móvil), y se reemplaza el líquido que sale con más solvente introducido por el extremo superior del tubo.

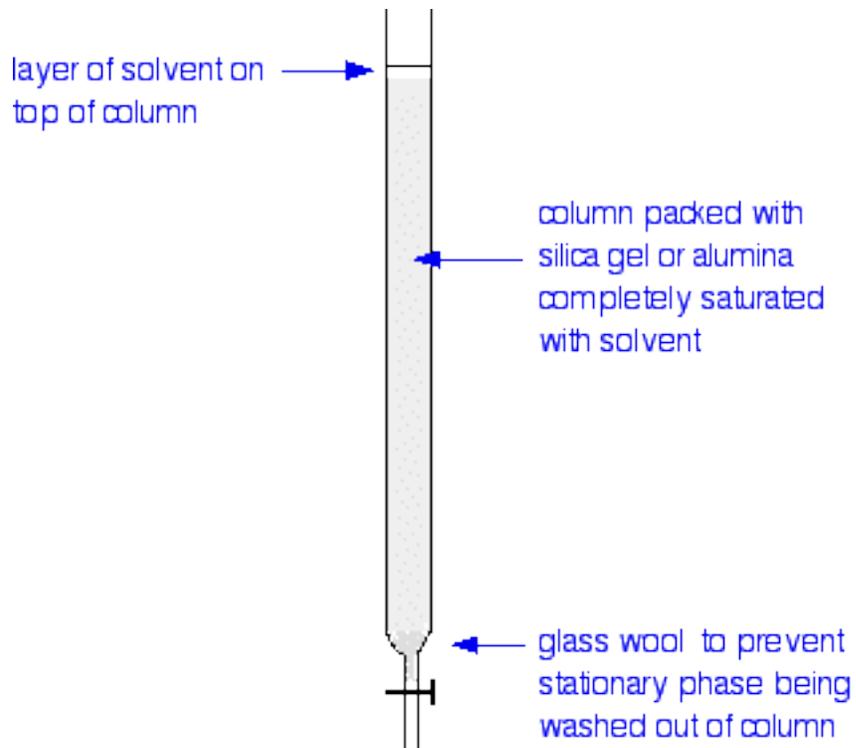


Figura 3. Elementos constitutivos de la cromatografía en columna.

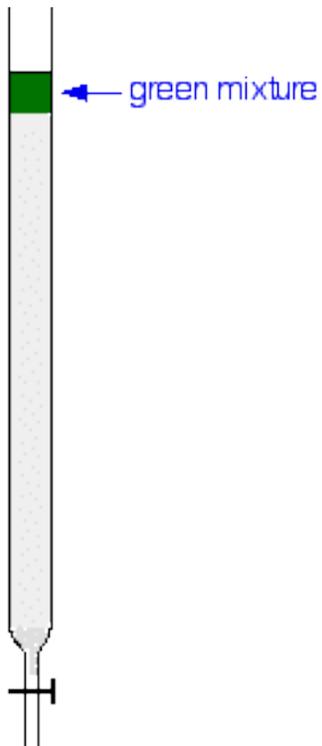


Figura 4. Cromatografía en columna, iniciando el experimento con una muestra verde.

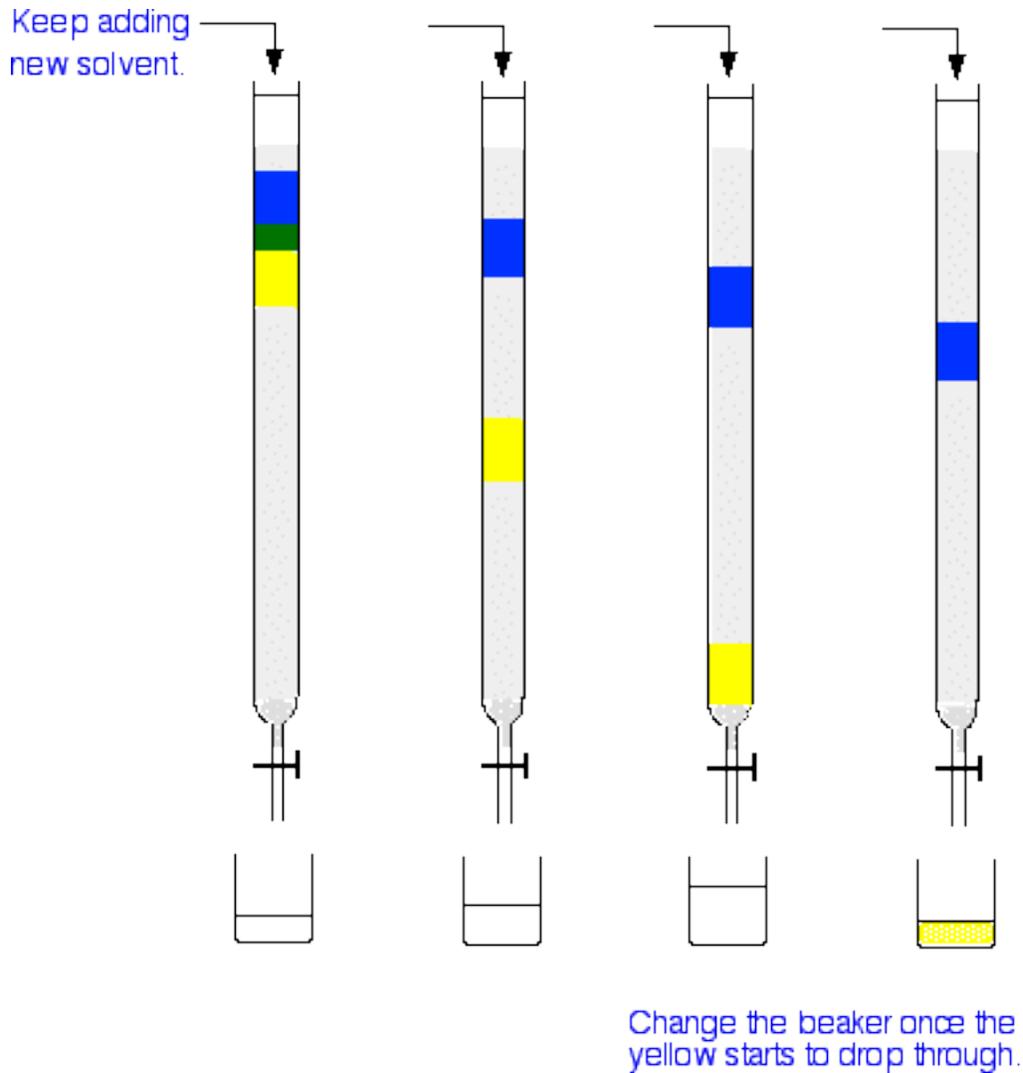


Figura 5. Ejecución de la cromatografía en columna, separación del verde en sus componentes azul y amarillo (como estrategia docente), con la adición de solvente.

En la medida en que el solvente va pasando, se van separando los componentes basados en el grado de adsorción. Los componentes que son separados y desechados se denominan eluidos, y el solvente se denomina eluyente. La cromatografía de columna sirve para purificar compuestos, entre otras cosas.

Cromatografía Líquida de Alto desempeño (HPLC)

Todas estas explicaciones fueron tomadas de (<http://www.chemguide.co.uk/analysis/chromatography/hplc.html>), que fue el único lugar que me pareció que hacían un esfuerzo decente para explicarlas a personas no técnicas.

En la HPLC es una cromatografía de columna en donde el solvente no cae por gravedad sino que es forzado a través de la columna por altas presiones de hasta 400

atmósferas, volviéndola mucho más rápida. Permite además unas partículas de mucho menor tamaño recubriendo la columna incrementando el área de superficie de interacción entre las fases estacionarias y las moléculas que fluyen por ella. Esto permite una separación muchísimo mejor de los componentes de la mezcla. Además, los métodos de detección pueden ser automatizados y extraordinariamente sensibles. Hay dos versiones de HPLC, la de fase normal y la de fase reversa. En la fase normal, que es la que menos se usa, es similar a la descrita en el aparte TLC donde la columna está llena de pequeñas partículas de sílica y el solvente es una sustancia no polar. Una columna típica tiene un diámetro interno de 4.6 mm, y un tamaño de 150 a 250 mm. Los compuestos polares se adsorben más que los no polares a la sílica. En la HPLC de fase reversa la sílica es convertida en NO polar adicionándole cadenas de hidrocarburo largas (de 8 a 18 carbonos), y se usa un solvente polar (agua con algo de metanol, por ejemplo) (Ver explicación gráfica - requiere de flash - en http://www.chemie.uni-hamburg.de/bc/Vorlesungen/_GE_Reversed+Phase.swf). En este caso, las moléculas polares serán las que avancen más por la fase móvil, que posteriormente son detectadas en el detector (ver figura 6). El tiempo que se demora un compuesto particular en atravesar la columna se denomina tiempo de retención (medido entre el momento en que se inyecta la muestra y el punto de máximo pico del compuesto en el detector). El tiempo de retención depende de la presión utilizada, la naturaleza de la fase estacionaria, la composición exacta del solvente, y la temperatura de la columna. Por estas razones, las condiciones de los experimentos en HPLC deben ser cuidadosamente controladas.

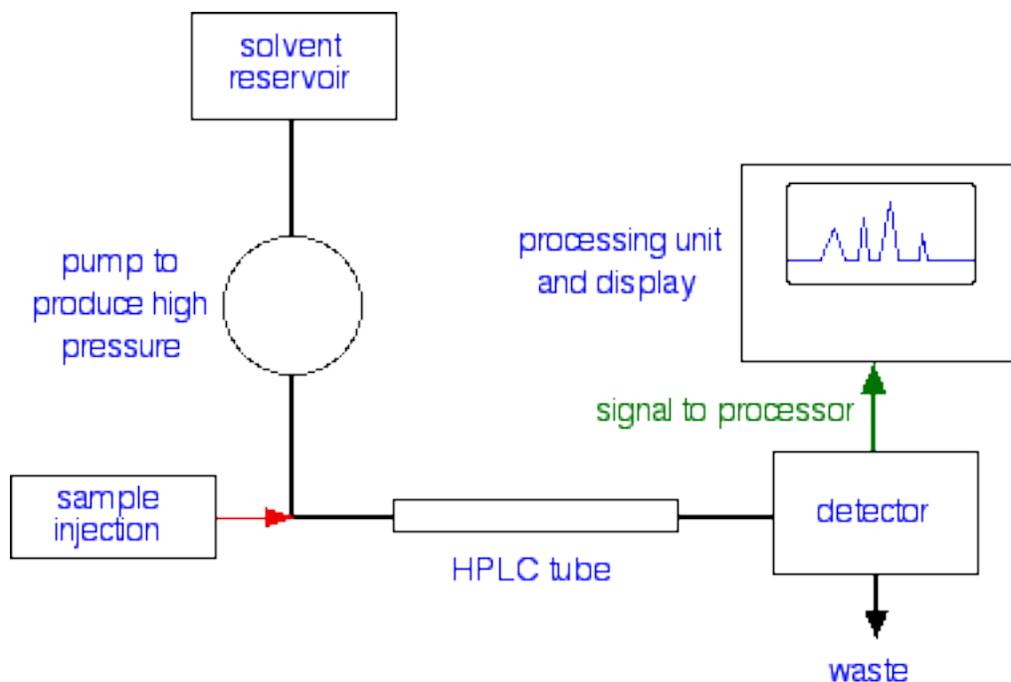


Figura 6. Esquema de flujo para la HPLC.

Acoplado HPLC a un espectrómetro de masas

Una forma realmente inteligente de estudiar muestras biológicas es seleccionando el pico designado del HPLC para que alimente el espectrómetro de masas, que va a dar origen a un patrón de fragmentación que permitirá su identificación a través de la comparación con una base de datos de compuestos conocidos, sin tener que conocer sus tiempos de retención.