

Identificación de epítomos conformacionales

Anna Tramontano

Departamento de Bioinformática. IRBM P Angeletti. Via Pontina Km 30.600, 00040 Pomezia, Roma, Italia. Telf: +39 6 91093207; Fax: +39 6 91093225; E-mail: tramontano@irbm.it

RESUMEN

En la actualidad, se ha difundido bastante el uso de métodos eficientes de biología molecular para seleccionar secuencias peptídicas capaces de imitar regiones de una proteína que interactúan con otras moléculas. Esto trae consigo la dificultad de identificar qué parte de la superficie de la proteína es imitada por el péptido seleccionado. Una vía posible es la generación y almacenamiento en una base de datos, de todas las secuencias peptídicas posibles que pueden imitar la superficie de una proteína de estructura conocida. Las secuencias de los péptidos seleccionados se utilizan para buscar en la base de datos y ubicar regiones de interacción como los epítomos reconocidos por anticuerpos. En este artículo se describe la aplicación de este método. Se tomaron como ejemplo los resultados de un tamizaje realizado a una biblioteca peptídica con un anticuerpo monoclonal específico contra la cadena H de la ferritina sérica humana. La localización del epítomo identificado coincide satisfactoriamente con los datos disponibles en la literatura sobre el complejo ferritina-anticuerpo.

Palabras claves: bibliotecas peptídicas, búsqueda en bases de datos, epítomos, ferritina

Biotecnología Aplicada 2001;18:91-93

ABSTRACT

Identification of Conformational Epitopes. Efficient methods for selecting peptide sequences able to mimic interacting regions of a protein surface are nowadays widely used in molecular biology. This leads to the problem of identifying which part of the protein surface is mimicked by the selected peptide. One possible method is to generate and store in a data base all possible peptides able to mimic the surface of a protein of known structure. Selected mimicking peptide sequences can be used to search the data base efficiently and to locate interacting regions, such as antibody epitopes. Here the application of such a method is described. The results of a phage library screening performed using a monoclonal antibody raised against human H-ferritin are used as a case study. The location of the identified epitope is in excellent agreement with all the data available on the antibody-ferritin complex.

Keywords: data base search, epitopes, ferritin, peptide libraries

Introducción

Las bibliotecas de péptidos codificados genéticamente o sintetizados por vía química, se han utilizado con fines diversos e interesantes en la biología molecular [1]. La gran diversidad de una biblioteca permite seleccionar los péptidos capaces de unirse a un receptor determinado; por ejemplo, a un anticuerpo. La biblioteca ofrece un número elevado de secuencias aminoacídicas diferentes y cada péptido es capaz de adoptar varias conformaciones en solución, por lo que, en principio, cualquier estructura de la superficie de una proteína puede ser mimetizada de forma efectiva por algún miembro de la biblioteca.

Se puede considerar la superficie de una proteína como el conjunto de todos los péptidos posibles, sobrelapados unos con otros, capaces de mimetizar su superficie, y una biblioteca peptídica como una colección a partir de la cual se pueden seleccionar uno o varios de estos péptidos. Si un péptido de ese conjunto mimetiza regiones discontinuas de la proteína próximas entre sí en su estructura terciaria, su secuencia no representará una región de la estructura primaria de la proteína, sino, más bien, un grupo de residuos aminoacídicos expuestos en la superficie de ésta que sería imposible mapear, incluso si se conociera la estructura tridimensional del polipéptido.

Un método de este tipo dio resultados muy satisfactorios cuando se utilizó para mapear un epítomo discontinuo reconocido por un anticuerpo monoclonal (AcM) generado contra la cadena H de la ferritina humana [2]. En primer lugar, se generó una base de datos con las secuencias

de todos los péptidos posibles de la superficie de la cadena H, y luego se utilizaron para hacer una búsqueda con secuencias seleccionadas de una biblioteca de nonapéptidos presentados en la proteína mayoritaria de la cápsida del bacteriófago f1. De esa forma, se identificó una región de la proteína que representa el supuesto epítomo discontinuo reconocido por el anticuerpo. El resultado concidió perfectamente con los datos experimentales disponibles en la literatura acerca de este complejo antígeno-anticuerpo.

Materiales y Métodos

Se utilizó el programa Insight [3] para calcular el grado de exposición de cada residuo aminoacídico y para diseñar los péptidos. Los cálculos de minimización de energía y dinámica molecular se realizaron con el programa DISCOVER (Biosym Tech., San Diego, CA, E.U.A.).

Descripción del método

Es lógico pensar que un péptido de longitud L mimetizará una región de la superficie de una proteína cuando cierto número de sus cadenas laterales (m) sean idénticas o similares a las presentes en la superficie de la proteína nativa. Los restantes L-m residuos del péptido son necesarios para que las m cadenas laterales seleccionadas adopten una disposición espacial adecuada, a los cuales les llamaremos "residuos espaciadores". Para una proteína determinada, el total de todas las secuencias de longitud L (incluidas las permutaciones) con m residuos coincidentes, se denomina conjunto de superficie de tipo

1. Cortese R, Monaci P, Nicosia A, Luzzago A, Felici F, Galfre G, et al. *Curr Opin Biotechnol* 1995;6:73-80

2. Pizzi E, Cortese R, Tramontano A. *Biopolymers* 1995;36:675-80.

3. Dayringer HE, Tramontano A, Sprang SR, Fletterick RJ. *J Mol Graphics* 1986;4: 82-7.

L/m. Por lo tanto, un conjunto de superficie de 7/4 se refiere a todos los heptapéptidos capaces de imitar una región de la superficie de la proteína que incluyan al menos cuatro cadenas laterales.

Si se cuenta con la estructura de la proteína ligando y con los valores de L y m, se puede generar su conjunto de superficie de tipo L/m y organizar las secuencias peptídicas resultantes en una base de datos, con la inclusión también de la identidad de las cadenas laterales que se desea mimetizar. A partir del fichero de coordenadas de la proteína ligando con el formato del Banco de Datos de Brookhaven [4], se puede calcular la exposición al solvente de cada residuo de la proteína y seleccionar los que tienen un valor mayor que el umbral seleccionado de antemano.

El primer paso es la generación de todas las combinaciones posibles de m residuos con un grado de exposición mayor que el umbral. Las combinaciones en las que la distancia entre los átomos de carbono β ($C\beta$) de cualquier par de residuos es mayor que la distancia que ocupa un péptido extendido de longitud L, ya pueden ser descartadas en esta etapa. Luego se hacen permutaciones con las combinaciones que quedan. En cada permutación se puede contar el número de aminoácidos necesarios para conectar los residuos expuestos seleccionados, y si el péptido final contiene más de L-m residuos espaciadores, se puede descartar también esa permutación.

El número de aminoácidos conectores entre dos residuos adyacentes en la permutación, se puede calcular a partir de la comparación de la distancia entre sus átomos $C\beta$ ($C\alpha$ para la glicina) con los valores que se muestran en la Tabla. Estos valores representan las distancias máximas entre los átomos $C\beta$ de pares de residuos separados por 0, 1, 2 o más enlaces peptídicos en la secuencia lineal. Todas las secuencias correspondientes a permutaciones permitidas, se pueden almacenar en una base de datos junto con la información sobre su localización en la secuencia lineal y en la estructura terciaria de la proteína. Esto permitirá el uso de muchos programas informáticos disponibles capaces de realizar búsquedas en bases de datos; por ejemplo, FASTA, PROFILESEARCH y FIND del paquete GCG [5].

Forzamiento del modelo

Un péptido que mimetice una región de una proteína debe de cumplir dos requisitos: su secuencia tiene que parecerse a la de uno de los péptidos del conjunto de superficie y su estructura tiene que adoptar una conformación que permita que las cadenas laterales seleccionadas tomen la disposición espacial deseada. Por consiguiente, después de que la búsqueda en la base de datos haya identificado las cadenas laterales de la proteína que un péptido determinado simularía desde el punto de vista geométrico, es preciso modelar de forma objetiva la estructura tridimensional del péptido para que adopte la conformación deseada y verificar que ésta sea factible estereoquímicamente.

El péptido se puede modelar primero a partir de su conformación extendida. Seguidamente, se obliga a que adopte una conformación en la que las m cadenas laterales seleccionadas se superpongan con los residuos correspondientes de la región de la superficie de la proteína. Esto se puede lograr mediante la minimización forzada de la energía libre del modelo con la introducción de un término de pseudoenergía en la ecuación del campo de fuerza:

$$E = \sum_i K_i (R_i - R_{i,\text{modelo}})^2$$

donde:

R_i : coordenadas de cada átomo de la cadena lateral cuyo movimiento se desea limitar

$R_{i,\text{modelo}}$: coordenadas de los átomos correspondientes de la superficie de la proteína

K_i : constantes de fuerza

Resultados

Existe información en la literatura sobre la selección de péptidos de una biblioteca aleatoria de nonapéptidos mediante el uso del AcM antiferritina humana H107 [6].

La ferritina participa en el almacenamiento de hierro y se expresa como un polímero de 24 subunidades formado por dos tipos de subunidades homólogas: H y L [7]. Se ha logrado sobreexpresar la cadena H en *Escherichia coli* [7] y se ha obtenido la estructura del polímero de 24 subunidades con alta resolución mediante cristalografía de rayos X [8]. El AcM H107 fue generado en ratón contra la ferritina humana y se ha comprobado que reconoce solamente su cadena H. Este anticuerpo se utilizó para tamizar una biblioteca de nonapéptidos presentados en la proteína mayoritaria pVIII de la cápsida del bacteriófago M13, con dos residuos de cisteína flanqueando el inserto aleatorio.

Se seleccionaron dos series diferentes de péptidos. El primer grupo se caracterizó por la presencia de tirosinas y triptófanos separados por cinco residuos de aminoácidos, con la preferencia adicional de que un residuo de ácido aspártico y otro de alanina estuvieran inmediatamente después de las tirosinas. La secuencia consenso resultante fue YDAXxxW. El segundo grupo se caracterizó por la presencia del motivo GSxF, aunque no siempre se conservara el residuo de glicina. En un caso había una tirosina en lugar de la fenilalanina. Como no se conoce si el epítopo reconocido por el AcM está formado por más de una subunidad de la ferritina, se utilizó la menor estructura que se repite en la molécula; es decir, el hexámero.

Hay 797 residuos en el hexámero que tienen un grado de exposición al solvente mayor que 1 Å, incluidos todos los residuos parcial o totalmente expuestos. Debido a que la secuencia consenso más larga que se encontró en la biblioteca tenía siete residuos y cuatro de ellos conservados como máximo, se generó un conjunto de superficie de tipo 7/4. Este conjunto incluyó las secuencias de 441 693 heptapéptidos. Luego se realizó una búsqueda en la base de datos con dos perfiles de secuencia correspondientes a las secuencias consenso obtenidas del tamizaje de la biblioteca y descritas anteriormente, para lo que se empleó el programa PROFILESEARCH del paquete GCG con un valor de *gap weight* = 5,0 y un valor de *gap length weight* = 0,05.

Las distribuciones de las puntuaciones fueron diferentes en los dos casos. En el primero hubo una separación clara entre la distribución aleatoria de las secuencias encontradas y una secuencia con un apareamiento estadísticamente significativo: la secuencia con la puntuación más alta (7,96) fue $Y_{36}D_{47}A_{52}XXXW_{93}$, en la que los subíndices indican las posiciones de los residuos mimetizados en la secuencia de la ferritina y X representa los residuos espaciadores. Para la segunda secuencia consenso, la forma del histograma se asemejaba más a una distribución estocástica sin apareamientos significativos evidentes y se decidió analizar cada uno de los cinco apareamientos con una puntuación mayor

Tabla. Distancias máximas entre los átomos de carbono β ($C\beta$) de pares de residuos separados por un número variable de residuos (representados por una X).

Pares de residuos	Distancia (Å)
$C\beta$ - $C\beta$	5,5
$C\beta$ -X- $C\beta$	7,5
$C\beta$ -X-X- $C\beta$	11,5
$C\beta$ -X-X-X- $C\beta$	14,5

4. Bernstein FC, Koetzle TF, Williams GJB, Meyer EF Jr., Brice MD, Rodgers JR, Kennard O. *J Mol Biol* 1977;112:535-42.

5. Devereux J, Haerberli P, Smithies O. *Nucleic Acids Res* 1984;12:387-5.

6. Arosio P, Adelman TG, Drysdale JW. *J Biol Chem* 1978;253:4451-8.

7. Levi S, Cesareni G, Arosio P, Lorenzetti R, Soria M, Sollazzo M, et al. *Gene* 1987;51:269-74.

que 5,5. De estas secuencias, dos estaban localizadas en una región de la cadena H próxima a la identificada en la primera búsqueda. Sus secuencias son:

$A_{52}S_{38}XY_{40}XXC_{90}$

$A_{52}S_{38}XY_{40}XXC_{102}$

Los otros tres péptidos se encuentran mapeados en regiones diferentes por completo, distantes espacialmente de la primera región identificada. Estos péptidos fueron descartados porque no pueden formar parte del mismo epítipo. Los residuos identificados se localizaban en una región discontinua de la superficie proteica conformada por residuos del lazo BC y del lazo de la hélice A (Figura).

Se analizaron con más detalle las secuencias peptídicas CYDGSYRWAPC y CRLPGSAFTYC, cada una representativa de una secuencia consenso de la biblioteca. Se hicieron coincidir las cadenas laterales del primer péptido —manteniendo sus cadenas laterales consenso— con las de los residuos Y_{39} , D_{42} y W_{93} (pep1). De esa forma se generaron dos conformaciones del segundo péptido: la primera se hizo coincidir con los residuos S_{38} , Y_{40} , C_{90} (pep2A) y la segunda, con los residuos S_{38} , Y_{40} , C_{102} (pep2B). En todos los casos, se hizo que las dos cisteínas terminales formaran un puente disulfuro. Pep1 puede adoptar una conformación capaz de llevar las cadenas laterales simuladoras a la posición relativa apropiada, con ángulos ω casi planares (entre -160 y $+160$) y ángulos ϕ y ψ posibles según el diagrama de Ramachandran.

Fue más difícil modelar la conformación deseada en el péptido pep2A. La mejor solución aún tenía dos ángulos ω no planares en las posiciones 7 y 9, y el diagrama de Ramachandran mostraba muchos puntos en regiones no permitidas. El péptido pep2B pudo ser modelado en la región seleccionada de la ferritina con todos los ángulos ω casi planares y un enlace peptídico en configuración *cis* en la posición de la glicina conservada. Todos sus residuos tenían una conformación β (posiciones 2-5 y 8) o α (posiciones 6, 7 y 9). Esto indicó que, con mayor probabilidad, la solución presentada por pep2B era el mejor candidato para mimetizar la superficie de la proteína. Tanto en pep1 como en pep2B, los puentes disulfuro mostraban una geometría adecuada con todas sus distancias en los límites observados para los enlaces disulfuro naturales y la mayoría de los ángulos dihedros cercanos a sus valores ideales.

Los resultados del análisis sugieren que el epítipo discontinuo del AcM H107 lo forman las regiones 38-52 y 90-93 de la secuencia de la ferritina. Esta conclusión coincide perfectamente con toda una serie de datos reportados en la literatura. En primer lugar, la mutagénesis de los residuos 40, 41 y 45 reduce significativamente la afinidad del anticuerpo, mientras que las mutaciones en los residuos 91, 92, 94 y 95 la eliminan completamente. En segundo lugar, el AcM no reconoce la cadena H de la ferritina de ratón, rata o pollo, ni la cadena L humana, y en todos los casos se observaron diferencias en al menos uno de los residuos identificados. En tercer lugar, mediante el empleo de un procedimiento completamente diferente, que consistió en la predicción de la estructura del anticuerpo H107 y su acoplamiento en la estructura de la cadena H de la ferritina, se identificó exactamente la misma región de la ferritina como candidata de unión del anticuerpo [9].

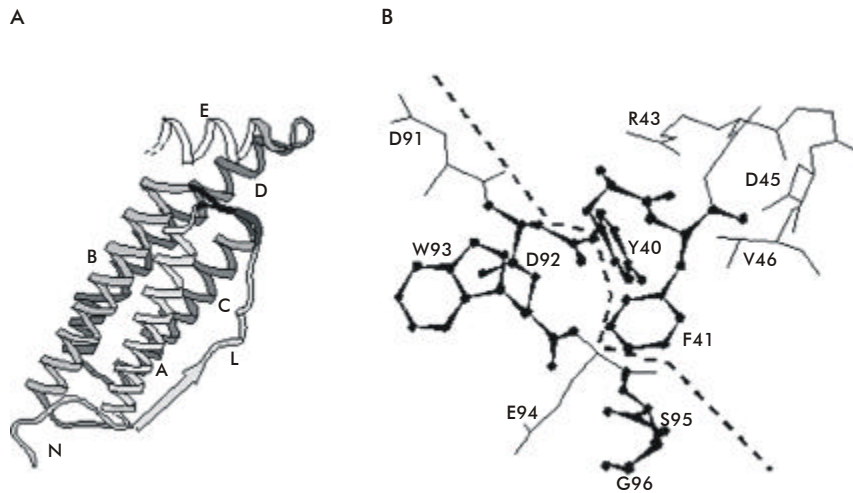


Figura. A) Estructura de un monómero de la cadena H de la ferritina humana. Los dos lazos oscuros contienen los residuos que participan en la unión con el AcM H107. B) Estructura de las regiones 40-46 y 91-96 de la cadena H. Sólo se muestran las cadenas laterales de los residuos expuestos al solvente. Los residuos representados mediante el modelo de bolas y varillas son candidatos a ser imitados por los péptidos seleccionados.

Discusión

Para la caracterización de un epítipo discontinuo de la superficie de una proteína, es necesario conocer la estructura tridimensional del complejo formado por esa proteína con su anticuerpo. Sin embargo, se pueden seguir vías alternativas si hay otros datos disponibles, como en el caso descrito en este artículo.

Los resultados que se presentan muestran que 1) es posible diseñar péptidos cuyas cadenas laterales coincidan con las cadenas laterales seleccionadas de la molécula de ferritina, sin que haya impedimentos estéricos entre los átomos del péptido; 2) las conformaciones que se obtuvieron para los péptidos son racionales desde el punto de vista de que los enlaces peptídicos son planares y los ángulos ϕ y ψ están en regiones permitidas del diagrama de Ramachandran; y 3) los resultados coinciden plenamente con todo lo que se conoce acerca del reconocimiento de la ferritina por el AcM H107.

Los estudios cristalográficos y de RMN de las proteínas por un lado y la tecnología de generación de bibliotecas peptídicas por el otro, se encuentran entre los campos de las ciencias biológicas que más rápido han evolucionado en los últimos años. Por esa razón, se puede considerar que el método descrito en este artículo será cada vez más útil como guía racional en los experimentos de mutagénesis. Este método no sólo pudiera ser de uso práctico en casos específicos, sino que, desde el punto de vista teórico, también pudiera ayudar a comprender mejor los principios que rigen la interacción entre dos proteínas.

Agradecimientos

Agradecemos a los Drs. A. Lahm, V. Amati, A. Luzzago, F. Felici y A. Wallace sus valiosas sugerencias, y a Dr. F. Pagone y al Departamento de Sistemas de Información por su excelente apoyo técnico.

8. Lawson DM, Artymiuk PJ, Yedwall SJ, Smith JMA, Livingstone JC, Treffry A, et al. *Nature* 1991;349:541-4.

9. Helmer-Citterich M, Rovida E, Luzzago A, Tramontano A. *Molecular Immunology* 1995;32:1001-10.