



Memoria del XLIX Taller de Actualización Bioquímica, Facultad de Medicina; UNAM

Cristalografía de macromoléculas.

Macromolecular crystallography.

Ramírez Rodríguez, Miguel Alejandro¹ y Rodríguez Romero, Adela^{1*}.

1. Laboratorio Nacional de Estructura de Proteínas-Instituto de Química, Universidad Nacional Autónoma de México.

*Correspondencia: Instituto de Química, Circuito exterior s/n Circuito de la Investigación Científica, Ciudad Universitaria., Coyoacán, CDMX, México, C.P. 04510. Tel: +52(55)56224568, adela@unam.mx

Resumen

La cristalografía, una disciplina que involucra el uso de técnicas de difracción de rayos X para revelar las estructuras tridimensionales de las proteínas y otras macromoléculas a escala atómica, ha dado lugar a numerosos avances en la comprensión de su funcionamiento en bacterias, virus, animales, plantas y seres humanos. Además, ha guiado el desarrollo de medicamentos y vacunas para combatir diversas enfermedades como alergias, cáncer, SIDA, etc. y más recientemente el SARS-COV-2. La cristalografía de proteínas es una forma de "microscopía" de muy alta resolución, que permite "ver" las moléculas a una resolución atómica. En las últimas décadas, los éxitos sin precedentes en este campo han ayudado a muchas disciplinas relacionadas con células y moléculas, en biología y medicina. Las estructuras de enormes complejos como el ribosoma, el centro de reacción fotosintética, las proteínas chaperonas, el complejo de la polimerasa, y los extraordinarios avances en la determinación de estructuras de proteínas integrales de membrana involucradas en el transporte y la conducción de iones, son ejemplos de este avance impresionante. A la fecha se han logrado avances importantes para superar dos cuellos de botella en este campo: el crecimiento de cristales de alta calidad y la solución del problema de la fase gracias a la robótica, software especializado, equipos caseros más poderosos, así como el uso de radiación de sincrotrón. En la actualidad, el desafío más crítico al que se enfrenta la cristalografía macromolecular de rayos X es la necesidad de responder preguntas sobre la dinámica de los procesos, con el fin de obtener

Abstract

Crystallography, a discipline that uses X-ray diffraction techniques to reveal the three-dimensional structures of proteins and other macromolecules at the atomic scale, has given rise to numerous advances in understanding their functioning in bacteria, viruses, animals, plants, and human beings. In addition, it has guided the development of medicines and vaccines to combat various diseases such as allergies, cancer, AIDS, and more recently, SARS-COV-2. Protein crystallography is a high-resolution "microscopy" that allows molecules to be "seen" at atomic resolution. In recent decades, the unprecedented successes in this field have helped many disciplines related to cells and molecules in biology and medicine. The structures of huge complexes such as the ribosome, the photosynthetic reaction center, the chaperone proteins, the polymerase complex, and the extraordinary advances in determining the structures of integral membrane proteins involved in ion transport and conduction are examples of this impressive progress. To date, significant advances have been made to overcome two bottlenecks in this field: the growth of high-quality crystals and the solution of the phase problem thanks to robotics, specialized software, more powerful home equipment, and the use of synchrotron radiation. The most critical challenge facing X-ray macromolecular crystallography is the need to answer questions about the dynamics of the processes to obtain information about the conformational changes known to play an essential role in controlling protein function.

información sobre los cambios conformacionales que se sabe desempeñan un papel esencial en el control de la función de las proteínas.

Palabras claves: cristalografía, macromoléculas, estructura.

Keywords: crystallography, macromolecules, structure.

Introducción

La cristalografía ha sido una de las herramientas más poderosas para los estudios estructurales de materiales, teniendo aplicaciones en varias ramas de la ciencia, abarcando desde moléculas pequeñas como sales o compuestos orgánicos hasta complejos macromoleculares como el ribosoma. Los cristales han cautivado a los seres humanos desde tiempos remotos, pero el nacimiento de la cristalografía no llegaría sino hasta el siglo XX con el descubrimiento de los rayos X por Wilhelm Röntgen en 1895, siendo este descubrimiento esencial para el nacimiento de la cristalografía como la conocemos. En 1912, Max von Laue obtuvo los primeros patrones de difracción de rayos X de cristales, empleando $\text{CuSO}_4 \cdot 5\text{H}_2\text{O}$, ganando en 1914 el Premio Nobel de Física por descubrir la difracción de rayos X a través de cristales. Al año siguiente ganarían el mismo galardón el dúo padre-hijo Sir William Henry Bragg y Sir William Lawrence Bragg por sus estudios acerca de las estructuras cristalinas empleando rayos X, siendo su contribución más importante la conocida Ley de Bragg, la cual sigue vigente hasta nuestros tiempos. Estos trabajos fueron pioneros en la cristalografía de rayos X que ha permitido obtener la estructura de un gran número de materiales; sin embargo, no es hasta 1958 que John Kendrew publica la primera estructura tridimensional de una proteína, siendo ésta la mioglobina, empleando para ello cristalografía de rayos X. Dos años después, Max Perutz publica la estructura tridimensional de la hemoglobina a baja resolución, este trabajo junto con el de Kendrew fueron el inicio de la biología estructural de proteínas y fueron posibles por la cristalografía de rayos X, sentando un precedente sobre el alcance de esta técnica en bioquímica. Estos descubrimientos les valieron a Perutz y Kendrew el Premio Nobel de Química en 1962 por sus estudios estructurales de proteínas globulares.

Tal ha sido el alcance de la cristalografía de rayos X que se han obtenido las estructuras de las subunidades mayor y menor del ribosoma por esta técnica, lo que les valió el Premio Nobel de Química a Ada Yonath, Venkatraman Ramakrishnan y Thomas Steitz en 2009. De esta manera, la cristalografía de rayos X ha sido la herramienta más poderosa de la

biología estructural desde sus inicios. A la fecha, la mayoría de las estructuras de macromoléculas conocidas se han obtenido por esta técnica, lo que abarca desde péptidos y proteínas pequeñas como la insulina y la heveína hasta sistemas complejos como las cápsides virales (1,2). Por otra parte, la cristalografía aplicada al estudio de macromoléculas ha avanzado a pasos agigantados debido al progreso en muchas disciplinas, como la tecnología informática (*hardware* y *software*) que proporciona potencia informática, o la física que desarrolla fuentes de radiación de alta intensidad, como los sincrotrones de cuarta generación; así como detectores 2D y ópticas más eficientes. Los estudios de alta resolución se han visto facilitados por la disponibilidad de programas de afinamiento adecuados y el nivel de detalle alcanzado es tan alto que es posible identificar la naturaleza química de los átomos ordenados a partir de los picos de densidad e incluso producir posiciones de átomos de hidrógeno observables en los mapas de densidad electrónica.

Adicionalmente, la resonancia magnética nuclear y la criomicroscopía electrónica son otras técnicas fundamentales de la biología estructural que han surgido en los últimos años y han complementado a la cristalografía de rayos X. Estas técnicas han permitido la determinación de la estructura tridimensional de macromoléculas complejas difíciles de cristalizar. La complementariedad entre estas técnicas ha dado lugar a importantes avances en la biología estructural (3,4).

Cristales

El elemento más fundamental de la cristalografía es el cristal, siendo definido tradicionalmente como un material sólido donde todos los átomos o conjuntos de átomos se agrupan en un arreglo ordenado y periódico en el espacio tridimensional. Esta definición fue modificada por el descubrimiento de los llamados cristales aperiódicos, que incluyen a los cuasi cristales, materiales con estructura ordenada más no periódica; por lo que actualmente la definición de cristal se tuvo que generalizar y según la IUCr (*International Union for Crystallography*) un material cristalino es “todo aquel material que presente un patrón de difracción con máximos agudos y definidos” (5,6).

Salvo casos particulares, como los ya mencionados cristales aperiódicos, la mayoría de los cristales encajan con la definición tradicional, por lo que poseen una estructura tridimensional ordenada y periódica, esa periodicidad permite definir una unidad de repetición conocida como celda unitaria. La celda unitaria es una celda imaginaria dentro del cristal que al ser repetida en las tres dimensiones construye la estructura entera del cristal, por lo que conociendo la celda unitaria y lo que contiene se tiene la estructura del cristal entero. La celda unitaria y los elementos que contiene se pueden conocer mediante el estudio del patrón de difracción del cristal; ahí yace el potencial de la cristalografía, al obtener el patrón de difracción del cristal de un material se puede determinar su celda unitaria y contenido, por consiguiente, la estructura interna del material. Sin embargo, se necesita de un cristal para ser difractado, lo que suele ser el paso limitante de la cristalografía, sobre todo para el caso de macromoléculas (6).

Fundamentos de cristalización de macromoléculas

La cristalización es el proceso de formación de un cristal. La mayoría de las veces ocurre en solución, donde el material a cristalizar está disuelto hasta el punto de saturación y se lleva la solución a sobresaturación mediante distintos métodos como evaporación del solvente o adición de agentes precipitantes. El objetivo es que el material pase a fase sólida a un ritmo lento y controlado, de tal manera que los átomos y moléculas se acomoden de la manera más compacta y ordenada posible, dando lugar a un cristal. Al ser un proceso de apilamiento ordenado, la cristalización es más sencilla mientras más pequeñas y rígidas sean las moléculas del material; en contra parte a mayor tamaño y flexibilidad de las moléculas menor será la probabilidad de obtener un cristal, este es el caso de las macromoléculas (7).

Las macromoléculas han sido un reto para la cristalización, el caso más estudiado y conocido son las proteínas, éstas al llegar a concentración de sobresaturación y en presencia de iones tienden a agregarse, exponiendo regiones hidrofóbicas que interaccionan con las regiones hidrofóbicas de otras cadenas polipeptídicas en una reacción en cadena que induce la formación de partículas de proteínas desestructuradas de gran tamaño. Para cristalizar una proteína es necesario evitar el proceso de agregación, en su lugar debe ocurrir un proceso de precipitación lento, ordenado y manteniendo la estructura nativa de la proteína (7).

La cristalización consta de dos pasos: nucleación y crecimiento cristalino. La nucleación consiste en la

formación de un pequeño cuerpo microscópico que sirve como semilla para los cristales, se forma a partir de la precipitación de la proteína en la solución sobresaturada. El crecimiento cristalino consiste en la adición de más moléculas de proteína a los núcleos de cristalización, dando lugar a los cristales y aumentando su tamaño. En algunas ocasiones actúan como núcleos de cristalización partículas de polvo u otros contaminantes, o bien dispositivos diseñados para tal fin (7).

La velocidad de la nucleación aumenta con la sobresaturación de la proteína en solución a un ritmo muy superior al del crecimiento cristalino, por este motivo es necesaria una sobresaturación de la solución que permita la formación de núcleos que después crecerán a cristales únicos ideales para ser difractados. Si la solución está muy sobresaturada se forman un gran número de núcleos que no crecen adecuadamente, dando lugar a precipitados. Por el contrario, si no se llega a la sobresaturación adecuada no se formarán núcleos que permitan el posterior crecimiento cristalino. Por estos motivos, la cristalización de proteínas es un reto donde se deben probar multitud de variables como concentración de proteína, agentes precipitantes, pHs, entre otros, que permitan llegar a la condición de sobresaturación idónea para la cristalización (7).

Los agentes precipitantes más empleados en cristalización de proteínas son polímeros como los PEGs y sales como $(\text{NH}_4)_2\text{SO}_4$ que mediante el proceso de "salting-out" (efecto salino secundario) inducen la precipitación de la proteína. Estos agentes precipitantes vienen acompañados de soluciones amortiguadoras de determinados pHs con ciertos aditivos (sales, detergentes, agentes reductores u oxidantes, entre otros), con el fin de brindar las condiciones donde se alcance la sobresaturación de la proteína sin que esta se agregue, de esta manera lograr su cristalización; estas mezclas se conocen como condiciones de cristalización. Sin embargo, debido a la variabilidad en las propiedades de las proteínas es difícil predecir condiciones de cristalización óptimas para cada proteína, por ello la cristalización sigue siendo un proceso empírico basado en ensayo y error. A pesar de ello, con el gran número de proteínas cristalizadas hasta la fecha, ya se han diseñado estuches de cristalización con amplias gamas de condiciones que ya han funcionado en proteínas y familias de proteínas en el pasado, aumentando las posibilidades de éxito. Aun con estas ventajas, sigue siendo necesario, en muchas situaciones, optimizar estas condiciones de cristalización prometedoras para obtener cristales únicos y mejorar la calidad de estos al momento de ser difractados, esto se logra

cambiando distintos parámetros del experimento como concentración de proteína, concentración de los agentes precipitantes, concentración de aditivos, probando distintos pHs de la solución amortiguadora, aditivos adicionales, entre otros (7,8).

En la cristalización de proteínas existen varios métodos siendo el más común la difusión en fase de vapor, este método consiste en la combinación de la solución proteica con la condición de cristalización, generalmente en una proporción 1:1, en una gota que se deja en un sistema cerrado con un reservorio de la condición de cristalización. Al mezclarse en la gota ambas partes se diluyen, por lo que la proteína no estaría sobresaturada, al irse estableciendo un equilibrio de presiones de vapor se ira evaporando el solvente de la gota y condensando en el reservorio haciendo que la gota concentre tanto a la proteína como a la condición de cristalización. Esto último lleva a la proteína a su condición de sobresaturación y posterior cristalización. En general, existen dos modalidades para este experimento: gota sedente o asentada, donde la gota reposa sobre una superficie sobre el nivel del reservorio; y gota colgante, donde la gota se adhiere a la superficie que recubre el reservorio, quedando la gota “colgando” sobre el reservorio. Actualmente, las pruebas de cristalización se montan empleando robots de cristalización y la modalidad más empleada es la gota sedente o asentada (7,8).

Fenómeno de difracción

La motivación para obtener un cristal es con el fin de coleccionar datos de difracción que permitan determinar la estructura cristalina, y de esta manera conocer la estructura interna del material; para el caso de las proteínas conocer su estructura tridimensional. Sin embargo, el paso del patrón de difracción a la estructura no es trivial e implica el uso de distintas herramientas matemáticas que permiten pasar de un conjunto de datos de intensidad de máximos de difracción a un mapa de densidad electrónica que pueda ser interpretado en un modelo estructural.

El fenómeno de difracción ya había sido descrito para la luz visible al hacerla pasar por un conjunto de rendijas contiguas ordenadas, los cristales de igual manera podemos verlos como un conjunto de rendijas, que serían los espacios entre los átomos, sólo que el ancho de estas rendijas está en el orden de los Angstroms. Por este motivo, los rayos X son idóneos para difractar cristales, ya que poseen longitudes de onda en este mismo rango y por consiguiente experimentan el fenómeno de difracción al atravesar la red cristalina. Si bien, el fenómeno de difracción en

una red de múltiples rendijas es sumamente complejo, William Lawrence Bragg propuso una forma de visualizar el cristal y el fenómeno de difracción que simplifica el problema.

Los átomos ordenados en el cristal se pueden imaginar formando familias de planos paralelos que actuarían como espejos reflejando los haces de rayos X, estos haces reflejados interferirían entre sí dando una señal o reflexión sólo cuando ocurra la interferencia constructiva, la condición geométrica para esa interferencia constructiva es la llamada Ley de Bragg (Ecuación 1). Básicamente, sólo cuando se cumpla la ley de Bragg se observará un máximo de difracción o reflexión y la intensidad de ésta dependerá de la densidad electrónica de los átomos en ese plano (9,10).

$$n \lambda = 2 d \sin(\theta) \quad [1]$$

La Ley de Bragg nos permite asociar cada máximo de difracción con una familia de planos atómicos en el cristal, este proceso se conoce como indexado y permite obtener los parámetros de la celda unitaria; además de asignar a cada reflexión un índice, llamados índices de Miller, los cuales se relacionan con las familias de planos atómicos involucrados. Sin embargo, debido al gran número de familias de planos atómicos en un cristal y su superposición es imposible resolver la estructura sólo con esta información. Por este motivo, se han creado dos conceptos que permiten correlacionar el patrón de difracción y la red cristalina; estos son la Red Recíproca y la Esfera de Ewald (11).

Teniendo una red cristalina con coordenadas (*xyz*) se puede crear una red recíproca, la cual consta de una red imaginaria que al igual que la red cristalina es ordenada y periódica en el espacio, conservando la simetría original. Sin embargo, sus dimensiones son iguales al inverso de las dimensiones de la red cristalina, de ahí su nombre red recíproca, y sus coordenadas están definidas por los índices de Miller (*hkl*). Por el otro lado, la Esfera de Ewald es una esfera alrededor del cristal que posee un radio igual al inverso de la longitud de la onda a difractar en el cristal ($1/\lambda$). Estos dos conceptos abstractos en conjunto permiten correlacionar el patrón de difracción y la red cristalina (11).

Cada vez que la superficie de la esfera de Ewald toca un punto de la red recíproca se cumple la Ley de Bragg y ocurre un máximo de difracción desde el cristal hacia ese punto de la red recíproca; por lo tanto, el patrón de difracción no es más que una proyección de la red recíproca en el detector que se visualiza

gracias a la esfera de Ewald. Con esto podemos construir la red recíproca a partir del patrón de difracción, ya que cada máximo de difracción se asocia a un punto de la red recíproca del cristal; esta red se relaciona con la red cristalina mediante la transformada de Fourier que nos lleva de las coordenadas (hkl) del espacio recíproco a las coordenadas (xyz) del espacio real o directo (Ecuación 2) (11).

$$\rho(xyz) = \frac{1}{V} \sum_{hkl} F(hkl) e^{(-2\pi i(hx + ky + lz))} \quad [2]$$

Espacio Real Espacio Recíproco

$\rho(xyz)$ es la densidad electrónica en cada punto dentro de la celda unitaria

V es el volumen de la celda unitaria

$F(hkl)$ es el factor de estructura de cada punto de la red recíproca

Resolución estructural: Faseo y afinamiento

Como se puede observar en la Ecuación 2, bajo la idea de la red recíproca surge el concepto de factor de estructura $F(hkl)$, el cual es un valor de cada punto de la red recíproca. Cada factor de estructura se puede descomponer como un vector en su módulo $|F(hkl)|$ y ángulo $\phi(hkl)$ (Ecuación 3), donde su módulo $|F(hkl)|$ depende de la intensidad de la reflexión $I(hkl)$ (Ecuación 4) y el ángulo de este vector es la fase $\phi(hkl)$. Las fases de cada factor de estructura son desconocidas y no se pueden medir empíricamente, a diferencia de las intensidades. Este es un problema que impide resolver directamente la estructura cristalina a partir del patrón de difracción, y se le conoce como el Problema de la Fase (12).

$$F(hkl) = |F(hkl)| e^{(i\phi(hkl))} \quad [3]$$

$F(hkl)$ es el factor de estructura

$|F(hkl)|$ es el módulo del factor de estructura

$\phi(hkl)$ es la fase

$$I(hkl) \propto |F(hkl)|^2 \quad [4]$$

$I(hkl)$ es la intensidad del máximo de difracción

$|F(hkl)|$ es el modulo del factor de estructura

El Problema de la Fase fue un reto durante muchos años y han surgido un gran número de métodos para obtener las fases, este procedimiento se conoce como faseo. Entre los métodos más importantes se encuentran: Métodos directos, Método de Patterson, Reemplazo Isomórfico Múltiple, Dispersión Anómala y Reemplazo Molecular, siendo los últimos tres los más importantes en la cristalografía de macromoléculas (12).

Métodos directos: Constan de métodos probabilísticos con los cuales se estiman las fases directamente de las intensidades de las reflexiones, es un método empleado fundamentalmente en estudios de moléculas pequeñas. Su diseño les valió el Premio Nobel de Química en 1985 a Herbert Hauptman y Jerome Karle (12).

Método de Patterson: Es un método basado en la obtención de una función denominada Función de Patterson, la cual es una convolución de la densidad electrónica y no requiere de las fases. Esta función se puede representar en el espacio de Patterson, en el cual la posición de cada máximo corresponde a un vector de distancia interatómica en el cristal y la intensidad del máximo es proporcional al producto de los números atómicos de los átomos a los extremos del vector, por lo que los átomos pesados son importantes al emplear este método. Los análisis de la Función de Patterson en conjunto con elementos de simetría permiten resolver estructuras cristalinas; de hecho, es el primer método de faseo diseñado y se emplea en moléculas pequeñas. La Función de Patterson se vuelve imposible de interpretar a medida que aumenta el número de átomos N en el cristal, ya que los máximos de Patterson aumentan como N^2 ; a pesar de ello, el Método de Patterson se emplea en la resolución de macromoléculas, más no de forma directa (13).

Reemplazo Isomorfo Múltiple (RIM): Este método consiste en coleccionar datos de difracción de un cristal de proteína. Posteriormente a un cristal isomorfo de la misma proteína se le difunde o añaden metales pesados, ya sea por co-cristalización o difusión al cristal, y se coleccionan datos de estos cristales. El método se basa en que la estructura de la proteína y el cristal no cambian con la adición de los metales, por lo que la diferencia entre ambos patrones de difracción va a estar dada por los átomos pesados añadidos. Este patrón de diferencia entre los cristales se emplea en el Método de Patterson, de tal manera que se determinan las posiciones de los átomos pesados en la celda unitaria, luego con esta "estructura" de los metales pesados se calculan las fases volviendo a calcular la red recíproca con la Ecuación 5. Estas fases las

empleamos para determinar la estructura nativa. Este método fue empleado en la resolución estructural de las primeras proteínas; sin embargo, ha caído en desuso por las complicaciones experimentales que implica obtener cristales isomorfos de una misma proteína con varios metales (14).

$$F(hkl) = \sum_j f_j e^{2\pi i(hx_j + ky_j + lz_j)} \quad [5]$$

$F(hkl)$ es el factor de estructura

f_j es el factor de dispersión atómica

Dispersión anómala (DA): Los patrones de difracción poseen una característica común, independientemente del material, son centrosimétricos, y esto se debe a la Ley de Friedel (Ecuación 6). Esta ley dice que las intensidades de las reflexiones asociadas a puntos opuestos de la red recíproca son iguales, esto es debido a que $|F(hkl)| = |F(\bar{h}\bar{k}\bar{l})|$, conocidos como Pares de Friedel. Esto no es un problema en el caso de tener cristales centrosimétricos, pero si es un problema para cristales no-centrosimétricos, como es el caso de los cristales de proteína. En estos casos, a pesar de conocer el módulo del factor de estructura, se desconoce su signo, para sobrellevar este problema se puede aprovechar una cualidad particular de ciertos átomos conocida como dispersión anómala. La dispersión anómala ocurre cuando los rayos X incidentes sobre el cristal tienen una longitud de onda que puede ser absorbida por distintos átomos en el material, estos generalmente son pesados como el Se, que se puede adicionar a la proteína en forma de selenio metionina, y esto afecta directamente la intensidad de los máximos de difracción.

Este fenómeno permite romper la Ley de Friedel (Ecuación 6) y de esta manera medir intensidades distintas entre los Pares de Friedel, de esta manera al restar las intensidades entre Pares de Friedel sólo quedará la contribución de los átomos implicados en la dispersión anómala. De esta manera, se obtiene un “patrón de difracción” con contribución exclusiva de los dispersores anómalos y se puede resolver esta estructura de manera similar que en Reemplazo Isomórfico; se emplea la Función de Patterson para resolver la posición de los átomos pesados, se emplea esta estructura para calcular las fases usando la Ecuación 5 y estas fases se emplean para resolver la estructura completa con el patrón original.

Este ha sido uno de los métodos más utilizados en los últimos años, y debido a que el método se basa en diferencias de intensidades entre pares de Friedel es fundamental tener mediciones muy precisas de cada reflexión y coleccionar para cada reflexión su par de

Friedel, por ello es que los datos coleccionados deben ser muy completos y tener una alta redundancia si se quiere emplear el faseo por dispersión anómala (14,15).

$$I(hkl) = I(\bar{h}\bar{k}\bar{l}) \quad [6]$$

Reemplazo molecular: Debido al gran número de estructuras de proteínas resueltas y depositadas en el Banco de Datos de Proteínas (PDB), existe una gran probabilidad que la proteína de interés sea homóloga a una proteína de estructura conocida. Por este motivo, se diseñó el método de reemplazo molecular, que consiste en emplear una estructura conocida como molde para obtener las fases. En este método se toma el modelo estructural homólogo, se calcula su Función de Patterson y se compara con la Función de Patterson del experimento de difracción. Con esta comparación se construyen unas operaciones de rotación y posterior traslación que permitan que las Funciones de Patterson se superpongan de la mejor manera posible, de esta manera se obtiene una “estructura” de la que se pueden calcular las fases. Básicamente, se toma el modelo y se roto-traslada buscando el mejor ajuste con los datos experimentales empleando la Función de Patterson como punto de comparación entre el modelo y el experimento. Actualmente, es uno de los métodos más usados debido al gran número de estructuras de proteínas depositadas en el PDB (Protein Data Bank) que pueden usarse como modelos de reemplazo molecular (16).

Equipos y materiales

Equipos

Un laboratorio de Cristalografía de proteínas requiere varios equipos y en este punto es conveniente mencionar que las fuentes de rayos X utilizadas y los métodos para su detección han experimentado mejoras drásticas durante las últimas décadas. Como ejemplo detallaremos los equipos del Laboratorio Nacional de Estructura de Macromoléculas del Instituto de Química, UNAM (Figura 1).

- Generador de Rayos X Micromax 007HF, que es una de las fuentes de rayos X de laboratorio más utilizada para cristalografía de proteínas. Tiene una potencia máxima de 2 kW y permite la recopilación de datos de difracción de alta calidad a partir de cristales pequeños y complejos. Tiene un ánodo rotatorio de cobre, de accionamiento directo que conduce a la emisión de rayos X característicos con una longitud de onda de 1.5418 Å. Los rayos X generados no son monocromáticos,

por lo que pasan a través de espejos. En algunos casos, los cristales son solo caracterizados en este tipo de equipo y son enviados a fuentes de sincrotrón para la colecta de datos final.

- El desarrollo de mejores detectores de rayos X ha contribuido significativamente a mejoras en la colecta de datos. El equipo cuenta con dos detectores, un DECTRIS PILATUS R200K, que detecta directamente fotones de rayos X, midiendo con precisión datos de muy alta calidad y un RAXIS IV++ que utiliza placas fosforescentes para la detección de rayos X difractados.
- Enfriador de Cristales Oxford, produce un flujo constante de nitrógeno gas a partir de nitrógeno líquido. Permite mantener la temperatura del cristal durante la colecta de datos de difracción a 100 K (-170°C).
- Robot de Cristalización Gryphon de Art Robbins que incluye inyector para proteínas solubles y un dispensador para la cristalización de proteínas de membrana. Cuenta con un cabezal de desplazamiento de 96 canales que llena rápidamente pozos con diversos agentes precipitantes en cajas de cristalización adecuadas. El cabezal no tiene puntas desechables, por lo que se minimiza el costo de operación y el desperdicio

ambiental. El dispensador de un canal aspira y dispensa a las proteínas, lo que permite posteriormente sellar las placas para evitar la evaporación de las mezclas proteína-agente precipitante y favorecer la formación de los cristales.

- Cámaras de temperatura controlada para favorecer el crecimiento de los cristales a 15 y 18 °C.
- Microscopios para visualización de cristales.
- Estaciones de trabajo utilizadas en todas las etapas de la determinación de la estructura y para la interpretación de los mapas de densidad electrónica y construcción del modelo.

Materiales

Estuches de cristalización de diferentes marcas comerciales como Hampton Research, Molecular Dimensions, etc. Cajas de cristalización para montaje de cristales de forma manual y utilizando el robot de cristalización, películas para sellado de cajas o cubreobjetos siliconizados. Asas de nylon para montaje los cristales, cabeza de goniómetro (Figura1).



Figura 1. Equipos y procedimientos para la cristalización y determinación estructural de proteínas en un laboratorio de cristalografía.

© 2022 Mensaje Bioquímico. Todos los derechos reservados. ISSN-0188-137X

Comité Editorial: González Andrade, M.; Hernández Alcántara, G.; Martínez González, J.J.;

Meraz Cruz, N.; Ramírez Silva, L.H. y Vilchis Landeros, M.M.

Publicado por el Departamento de Bioquímica de la Facultad de Medicina; UNAM.

Procedimientos

Preparación de las macromoléculas

Los avances en disciplinas como la biología molecular y la bioquímica han influido dramáticamente en los logros de la cristalografía de proteínas. Las técnicas de DNA recombinante han producido la mayoría de las proteínas solubles, en cantidades adecuadas para estudios cristalográficos, y además pueden diseñarse variantes que permitan la cristalización o ayuden a la compresión del mecanismo de acción. Tanto los sistemas de expresión de alto nivel como los métodos para realizar modificaciones definidas de proteínas recombinantes simplifican enormemente los procesos de purificación. Además, la ingeniería de proteínas se ha utilizado con éxito para obtener cristales (17), para mejorar la calidad de la difracción (18) o para crear sitios potenciales de unión de átomos pesados (19). Ya sea que una proteína se aísle de la fuente natural o se obtenga mediante técnicas recombinantes, se debe evaluar la idoneidad de la muestra de acuerdo con varios criterios, como la homogeneidad, la actividad, la monodispersidad y la estabilidad.

Cristalización y caracterización del cristal

Las técnicas empleadas para obtener cristales proteicos son esencialmente la difusión en fase de vapor, modalidad gota sedente o gota colgante. Dicha información se puede encontrar en muchos libros de texto o en diversas páginas web (7). En el robot de cristalización se usa el método de difusión en fase de vapor, modalidad gota sedente. La solución de proteína pura y monodispersa debe estar relativamente concentrada, aproximadamente entre los 2 y 30 mg/ml. Para reducir la solubilidad de la proteína y generar una precipitación controlada se adicionan agentes precipitantes. Al llevar a un incremento paulatino de la concentración se generan los núcleos cristalinos, los cuales en condiciones adecuadas crecen para dar lugar a un cristal de tamaño adecuado para los experimentos (0.05 a 0.5mm).

Una vez que se obtiene un buen cristal, en un microscopio se selecciona y se “pesca” con un asa especial de nylon, en una cantidad pequeña de líquido. Si es necesario, el cristal se transfiere a la misma condición de cristalización conteniendo algún agente crioprotector como glicerol para evitar formación de cristales de hielo que interfieran con el patrón de difracción de la proteína. Entonces el cristal se sumerge en nitrógeno líquido, si se va a transportar a temperatura de nitrógeno líquido a una fuente de

sincrotrón o directamente se coloca en una cabeza de goniómetro, en el flujo de nitrógeno gas (100 K/-173 °C). La recolección de datos en cristales de proteínas se realiza a temperaturas criogénicas, de modo que las moléculas se “enfrian a 100 K, esto ayuda a dar una “imagen” más nítida. Como ventaja adicional, a esta temperatura los cristales son menos propensos al daño por radiación producido por los rayos X (20).

Colecta de datos de difracción de rayos X

Una vez enfriados a 100 K, los cristales se montan en el equipo de difracción o se congelan en dispositivos especiales para trasportarlos a fuentes de radiación más poderosas como los sincrotrones de tercera o cuarta generación, que producen rayos X de alta intensidad y son modulables. En este último caso, de forma remota se seleccionan las muestras a analizar, indicando al brazo del robot que las recoja una a la vez y las transfiera a un soporte de muestras. En ambos casos los cristales se deben alinear con el haz de rayos X con precisión. Entonces se pueden tomar las medidas. En equipos de laboratorio, las mediciones de difracción se realizan a partir de ánodos rotatorios de cobre que proporcionan un flujo de rayos X con una longitud de onda de 1.54 Å, también se pueden utilizar ánodos rotatorios de cromo (2.23 Å) en experimentos basados en la señal de dispersión anómala del azufre y otros elementos relativamente ligeros (21). El uso de fuentes de rayos X de laboratorio o “caseras” está respaldado por las propiedades de los elementos ópticos, como los espejos de doble reflexión, que permiten emitir un haz de rayos X muy concentrado e intenso, comparable al del sincrotrón de segunda generación.

En los experimentos de difracción cuatro parámetros deben ser controlados: la distancia cristal-detector, la longitud de onda de la radiación, dependiendo del método de fase que se desee emplear, esto es reemplazo molecular o dispersión anómala, el tiempo de exposición y el ángulo de oscilación. Estos datos se obtienen después de caracterizar al cristal mediante la colecta de unas pocas imágenes de difracción, lo que permite determinar si se trata efectivamente de un cristal de proteína, hasta que resolución difracta, a que red cristalina pertenece y si presenta algún problema, por ejemplo, si se trata de un cristal gemelo.

Una vez que el cristal se ha caracterizado, se diseña una estrategia de colecta óptima dependiendo de la red cristalina, la mosaicidad y el poder de difracción del cristal, y el método que se usará para obtención las fases de las reflexiones. Esta estrategia

establece parámetros tales como tiempo de exposición, ángulos de rotación del cristal, atenuación del haz, etc. Es importante mencionar que dependiendo del tipo de datos experimentales que se quiera coleccionar, por ejemplo, dispersión anómala sencilla o reemplazo molecular, hay que optimizar los protocolos de colecta. Las estrategias de recopilación de datos deben maximizar tanto la resolución como la integridad del conjunto de datos. El método utilizado tradicionalmente es rotar el cristal en un ángulo pequeño y registrar el patrón de difracción de rayos X. El software comúnmente utilizado para diseñar la estrategia de colecta es HKL3000 (22).

En los experimentos se registran una serie de patrones de difracción de rayos X que, con un buen cristal, aparecerán como una colección de manchas o reflexiones en el detector (Figura 1). Debido a que los cristales presentan simetría algunas reflexiones serán equivalentes, y el número de mediciones que se tengan para una misma reflexión se conoce como redundancia. Mientras mayor sea este número más precisa será la estimación de la intensidad promedio para cada reflexión. Por otra parte, al tratar de obtener una mayor redundancia, los tiempos de exposición se incrementan provocando lo que se conoce como “daño por radiación”, que implica que el cristal fue expuesto a altas dosis de radiación ocasionando daños estructurales en la proteína como pueden ser rompimiento de puentes disulfuro, descarboxilación de grupos ácidos, etc.

Es importante mencionar que los cristales de proteínas y ácidos nucleicos suelen contener una alta proporción de disolvente en la celda unitaria y, por lo tanto, son muy frágiles. La proporción de solvente en la unidad asimétrica se puede determinar utilizando el volumen de Matthews (V_M), que es la relación entre el volumen de la celda unitaria y el peso molecular de la proteína contenida en esa celda, definido por la Ecuación 7 (23). La distribución de V_M es asimétrica y pueden darse valores por debajo de 1,9 o por encima de 2,9 Å³ Da⁻¹, pero son relativamente raros en las proteínas.

$$V_M = \frac{V}{M * Z} \quad [7]$$

M es el peso molecular de la macromolécula [Daltons]

V es el volumen de la celda unitaria [Å³]

Z es el número de moléculas en la celda unitaria

El procedimiento de colecta de los datos de difracción se puede realizar utilizando diferentes sistemas de software disponibles, que se mejoran periódicamente. Los más utilizados son MOSFLM (24), HKL3000 (22) y XDS (25). La razón fundamental de una buena colecta de datos es obtener la máxima integridad, con el mínimo de errores estadísticos asociados, a la resolución más alta.

Para evaluar la calidad de los datos de difracción se determina la dispersión de las intensidades individuales de todas las reflexiones equivalentes relacionadas por simetría, que contribuyen a la reflexión única, y se estima por el valor residual R_{merge} (Ecuación 8) o R_{sym} . A este proceso de “promediación” se conoce como escalamiento y fusión y su resultado es un conjunto de reflexiones únicas, cada una acompañada de la estimación del error.

$$R_{merge} = \frac{\sum_{hkl} \sum_{i=1}^n |I_i(hkl) - \bar{I}(hkl)|}{\sum_{hkl} \sum_{i=1}^n I_i(hkl)} \quad [8]$$

$I_i(hkl)$ son las mediciones individuales de la intensidad de una reflexión hkl

$\bar{I}(hkl)$ es el promedio de las intensidades medidas para una reflexión hkl

Resolución del problema de la fase

Los procedimientos para determinar las fases de las reflexiones también han evolucionado de manera muy significativa. El reemplazo isomorfo múltiple fue el método más popular cuando la accesibilidad de las estructuras era limitada. Sin embargo, con la disponibilidad de una gran cantidad de coordenadas (diferentes modelos de proteínas con varios plegamientos) depositadas en el Protein Data Bank (26) (Tabla 1), el reemplazo molecular (MR) se ha convertido en uno de los métodos más utilizados. En este caso se debe contar con la secuencia de aminoácidos de la proteína problema y datos de un modelo tridimensional homólogo experimental obtenido del PDB.

La identidad de secuencia debe ser al menos del 30%. También se pueden usar modelos obtenidos mediante técnicas bioinformáticas, como el AlphaFold (27,28) que utiliza inteligencia artificial para obtener modelos de proteínas, o bien Robetta (29) y el SwissModel (30).

Table 1. Estadística de la colecta de datos y el afinamiento (reemplazo molecular) para la Triosafosfato isomerasa de *Entamoeba histolytica*

Triosafosfato isomerasa (<i>Entamoeba histolytica</i>)	
Colecta de datos	
Grupo especial	P2 ₁ 2 ₁ 2
Dimensiones de la celda unitaria	
<i>a</i> , <i>b</i> , <i>c</i> (Å)	85.73, 119.95, 50.39
α , β , γ (°)	90, 90, 90
Resolución (Å)	1.5
<i>R</i> _{sym} o <i>R</i> _{merge}	4.0
<i>I</i> / σ <i>I</i>	2.3
Integridad (%)	99.0
Redundancia	3.6
Afinamiento	
Resolución (Å)	10-1.5
No. de reflexiones	80009
<i>R</i> _{work} / <i>R</i> _{free}	18.4/20.7
Factores <i>B</i> (Å ²)	
Todos los átomos	14.0
Desviaciones R.M.S.D.	
Longitudes de enlace (Å)	0.004
Angulos de enlace (°)	1.2

El reemplazo molecular ha sido ampliamente utilizado para macromoléculas problema que pueden relacionarse con una estructura homóloga conocida, o bien para aquellas que contienen una subunidad con estructura conocida (16). Para usar esta técnica se necesita un patrón de difracción experimental para la estructura desconocida y las coordenadas atómicas del modelo o subunidad (estructura relacionada). En este caso, la integridad del conjunto de datos es más importante que la precisión. Para calcular las fases iniciales aproximadas, se debe determinar la posición del modelo o estructura similar depositada en el PDB, dentro de la celda unitaria de la estructura desconocida.

En principio, una búsqueda en seis dimensiones ofrece la mejor concordancia entre los factores de estructura observados y calculados. Generalmente, la orientación de la molécula y su posicionamiento deben completarse realizando primero una búsqueda rotacional tridimensional y luego trasladarla mediante una función de traslación tridimensional. A partir del análisis teórico de las propiedades de la función de Patterson (13,16) se hizo evidente que tal búsqueda de

seis parámetros podría reducirse a dos problemas tridimensionales. El primero es la determinación de la orientación correcta de la sonda y el segundo es la determinación de la posición de la molécula correctamente orientada dentro de la celda unitaria.

Hoy en día existen programas novedosos y muy robustos que tienen éxito en casos difíciles con modelos de búsqueda que presentan porcentajes de identidad cercanos al 30% y que realizan búsquedas en seis dimensiones e implementan valores de máxima verosimilitud, como MOLREP (31) o PHASER (32).

Los paquetes automatizados como los programas incluidos en CCP4 (Proyecto Computacional Colaborativo No. 4 de los Consejos de Investigación del Reino Unido) (33) o PHENIX (34) han tenido un gran impacto global en la ciencia cristalográfica y pueden hacer que las aplicaciones sean bastante sencillas. El reemplazo molecular es actualmente responsable de aproximadamente la mitad de todas las estructuras depositadas en el *Protein Data Bank*.

Interpretación de mapas de densidad y construcción de modelos

El resultado principal de un experimento de difracción de rayos X es un mapa de la densidad de electrones dentro del cristal. Esta distribución de electrones generalmente se interpreta en términos de átomos y moléculas individuales. En general, la estructura primaria o secuencia de aminoácidos de la muestra que se ha cristalizado es conocida. Entonces, básicamente, se ajustan estos aminoácidos al mapa de densidad de electrones calculada. La interpretación de dichos mapas, así como la precisión de los modelos atómicos dependen de la resolución de los datos (es decir, del número de reflexiones incluidas en el cálculo del mapa a la mayor resolución dada por el cristal). Por lo tanto, a mayor resolución; por ejemplo, de 1 a 1.5 Å el mapa de densidad electrónica permite ubicar los átomos de manera segura y precisa, mientras que en un mapa a resolución baja (3 a 3.5 Å) solo es posible localizar con precisión algunos aminoácidos.

La representación de la densidad electrónica es obtenida por la suma de Fourier de los $|F_{obs}|$ en combinación con las fases finales (Ecuación 2). En la práctica, un modelo se inspecciona usando un mapa $2F_{obs} - F_{calc}$, que al menos en principio, también debería revelar errores en el modelo actual porque es una suma de los F_{obs} y el mapa de diferencia $F_{obs} - F_{calc}$.

Una vez que la calidad de los mapas de reemplazo molecular o MAD/SAD es lo suficientemente buena, la construcción de los modelos se realiza usando el programa COOT (35). El ajuste de la secuencia de aminoácidos a los mapas o la construcción de modelos implica ajustar la cadena polipeptídica con todos sus aminoácidos y cadenas laterales al mapa de densidad electrónica, buscando la mayor concordancia entre el modelo construido y el mapa obtenido de los datos experimental. El ajuste del mapa se realiza mediante gráficos de computadora interactivos; no obstante, la automatización ha eliminado gran parte del trabajo manual de este proceso. A menudo, las regiones de alta flexibilidad no son visibles debido al desorden estático, donde la estructura varía de una molécula a la siguiente dentro del cristal, o al desorden dinámico, donde la región es móvil dentro del cristal; este último tipo de trastorno se minimiza colectando datos a temperaturas cercanas a los 100 K. Potentes programas como ARP/wArp (36) o RESOLVE (37), e incluso algunas rutinas disponibles en COOT (35) permiten hoy en día construir la cadena polipeptídica sin intervención humana, incluso en mapas de baja resolución.

Afinamiento de los modelos

Una vez que la interpretación del mapa es completa y lo suficientemente adecuada, la siguiente tarea para mejorar dicha interpretación es afinar la estructura. El propósito del afinamiento cristalográfico es lograr el mejor ajuste entre los datos observados y calculados de un modelo atómico, mientras se produce una buena geometría local, además de interacciones no ligadas. El proceso de afinamiento por lo general implica rondas alternas de optimización automatizada de todos los parámetros afinados y correcciones manuales de los modelos para mejorar su concordancia con los mapas de densidad electrónica (de acuerdo con algoritmos de mínimos cuadrados o máxima verosimilitud). Las correcciones manuales son necesarias porque los parámetros afinados automáticamente pueden quedar en un mínimo local (matemático), en lugar de conducir a la solución óptima. Los parámetros del modelo que son optimizados por el software de afinamiento incluyen, para cada átomo, sus coordenadas x, y, z y un parámetro que refleja su 'movilidad' o flexibilidad en el espacio, conocido como el Factor B (o parámetro de desplazamiento atómico (ADP), también conocido como el factor de temperatura). Los factores B generalmente se expresan en Å² y varían de 2 a 100 Å². Este procedimiento tendrá el efecto de mejorar las fases, lo que da como resultado mapas más claros y, por lo tanto, mejores modelos. Actualmente se emplea una variedad de métodos para mejorar la concordancia entre los patrones de difracción observados y los calculados.

En las primeras etapas de la determinación de la estructura, el primer objetivo es determinar cómo describir el modelo mediante el trazado de la cadena polipeptídica, así como la conformación de los residuos de aminoácidos. Una vez determinado el modelo, el objetivo es establecer estimaciones precisas de los valores de los parámetros que mejor expliquen los datos observados. La resolución, precisión, integridad y ponderación de los datos tienen un impacto importante en el proceso de afinamiento y, por lo tanto, hay varios programas disponibles para este fin, que son diferentes en el rango de funcionalidad. El primer indicador de calidad para la estructura es el factor R cristalográfico convencional, que es simplemente el error fraccional promedio en el factor de estructura calculado en comparación con el factor de estructura observado (Ecuación 9). Aunque depende de una serie de factores, como regla general, una buena estructura tendrá un factor R en el rango de 10% a 20%. No obstante, todavía se pueden obtener buenos factores R con errores significativos en la estructura; en consecuencia, se han incorporado

mejoras importantes en los programas de afinamiento. Entre ellos se encuentran la herramienta de validación R_{free} (38), la estimación de incertidumbres posicionales (39), fragmentos rígidos con afinamiento TLS (donde T representa oscilaciones traslacionales (lineales), L para oscilaciones de libración (rotacionales) y S (tornillo) para el acoplamiento de estos dos modos (40), además de restricciones de ángulo de torsión. Un modelo bien afinado es un requisito previo para la interpretación de la estructura macromolecular en términos de su función e interacción con otras moléculas. Los programas más utilizados actualmente son REFMAC (41) y PHENIX (34). Para afinamiento de estructura de muy alta resolución el programa más adecuado para realizar el afinamiento de la estructura es SHELX (42).

$$R = \frac{\sum ||F_{obs}| - |F_{cal}||}{\sum |F_{obs}|} \quad [9]$$

R es el factor estadístico R

F_{obs} es el factor de estructura determinado experimentalmente

F_{cal} es el factor de estructura calculado a partir del modelo estructural

Validación del modelo y deposito de datos en el Banco de Datos de Proteínas (Protein Data Bank)

La validación es requerida para evaluar que tan “correcto o adecuado” es el modelo tridimensional experimental de una macromolécula obtenida mediante el análisis de la estructura cristalina. La evaluación del modelo puede incorporar dos medidas: la concordancia entre el modelo y los datos de difracción experimentales y la concordancia entre el modelo y la base de datos de estructuras conocidas. Muchos criterios y programas informáticos están disponibles para verificar la calidad del modelo. Los programas Procheck (43) WHAT IF (44) y SFCheck (45) contienen un gran grupo de controles de calidad y brindan una colección de diagnósticos estructurales que incluyen comparaciones de distancias y ángulos de enlace, comparaciones de ángulos de torsión, comparaciones de morfología de base (para ácidos nucleicos) imágenes de gráficas moleculares. Existen acuerdos sobre qué medidas son buenos indicadores de la calidad estereoquímica; estos incluyen planaridad; quiralidad; preferencias ϕ/ψ (gráfico de Ramachandran); ángulos χ , distancias de contacto no ligadas; y donantes y aceptores insatisfechos. El resultado de un programa de verificación de estructuras permite una evaluación objetiva de la calidad estereoquímica del modelo, como ejemplo

está el *software* MOLPROBITY (46). Un ejemplo de herramienta importante para validar la calidad de un modelo cristalográfico se muestra en la Tabla 1 que se incluye en las publicaciones de estructuras de macromoléculas. Esta herramienta combina funciones de validación de PHENIX (34), como se muestra en la Tabla 1.

Resultados

La figura 2 muestra de manera gráfica cada uno de los resultados obtenidos en las etapas involucradas en la determinación de la estructura tridimensional de una proteína utilizando el método de reemplazo molecular.

Etapas 1. Obtención de cristales adecuados para estudios estructurales.

Etapas 2. De acuerdo con la estrategia de colecta se obtiene un conjunto de imágenes conteniendo las reflexiones a una cierta resolución.

Etapas 3. Las imágenes se procesan para determinar la red cristalina y el grupo espacial en el que cristalizó la proteína (dimensiones y ángulos de la celda unitaria derivados de la celda recíproca en el patrón de difracción). Los archivos obtenidos son .sca, con HKL3000 (22) y .mtz con XDS (25) o Mosflm (24), que contienen todas las reflexiones combinadas, indexadas (índices h, k, l) y escaladas, así como la intensidad de éstas. Los parámetros estadísticos del conjunto de datos son relevantes ya que indican su calidad: que tan completo está el conjunto de datos (integridad o compleción), a que resolución difractó el cristal, la redundancia de las reflexiones, y los valores de error R_{merge} , R_{meas} y R_{pim} . La calidad de la estructura final depende de la calidad de estos datos. Para determinar con precisión el grupo espacial al que pertenece la red cristalina hay que analizar la extinciones sistemáticas de las reflexiones en los tres ejes de la red cristalina.

Etapas 4. Conociendo la secuencia de aminoácidos y una estructura con una identidad de al menos 30% reportada en el PDB se puede usar esta última para determinar las fases de las reflexiones utilizando el método de reemplazo molecular. El modelo de búsqueda debe ser editado para eliminar o cambiar aquellos residuos que son muy diferentes o regiones que no están presentes en la cadena polipeptídica problema. Además, es importante determinar cuantas cadenas polipeptídicas se encuentran en la unidad asimétrica, mediante el cálculo del número de Mathews. Una vez ajustada la o las cadenas polipeptídicas en la celda unitaria problema se

obtendrán las fases iniciales de las reflexiones, que permitirán calcular los primeros mapas de densidad electrónica.

Etapa 5. Un aspecto importante es verificar que la solución del reemplazo molecular no tenga choques con las moléculas relacionadas por simetría. Una vez verificado esto se analizan los mapas para construir la secuencia de aminoácidos de la cadena polipeptídica problema y se realiza un afinamiento de cuerpo rígido.

Etapa 6. Es importante separar un conjunto de reflexiones que no se usarán en el proceso de afinamiento y que permitirá determinar el valor de R_{free} del modelo final. En general, un conjunto de reflexiones que representan menos del 10% del total (alrededor de 1000 reflexiones) se usan para este fin.

Etapa 7. Una vez que el modelo está completo y los valores de R y R_{free} están en límites adecuados para la resolución a la cual difractó el cristal, se procede a la validación de este.

Etapa 8. Modelo final. Este archivo contiene las coordenadas en el espacio de cada átomo en la unidad asimétrica, el factor de ocupancia, y los factores de vibración térmica.

Etapa 9. Representación gráfica del modelo que llevará al conocimiento detallado de la molécula en estudio. Los programas más utilizados son COOT (35), PYMOL (47) y CHIMERA (48).

Etapa 10. El modelo final validado se deposita en el Banco de Datos de Proteínas (PDB).

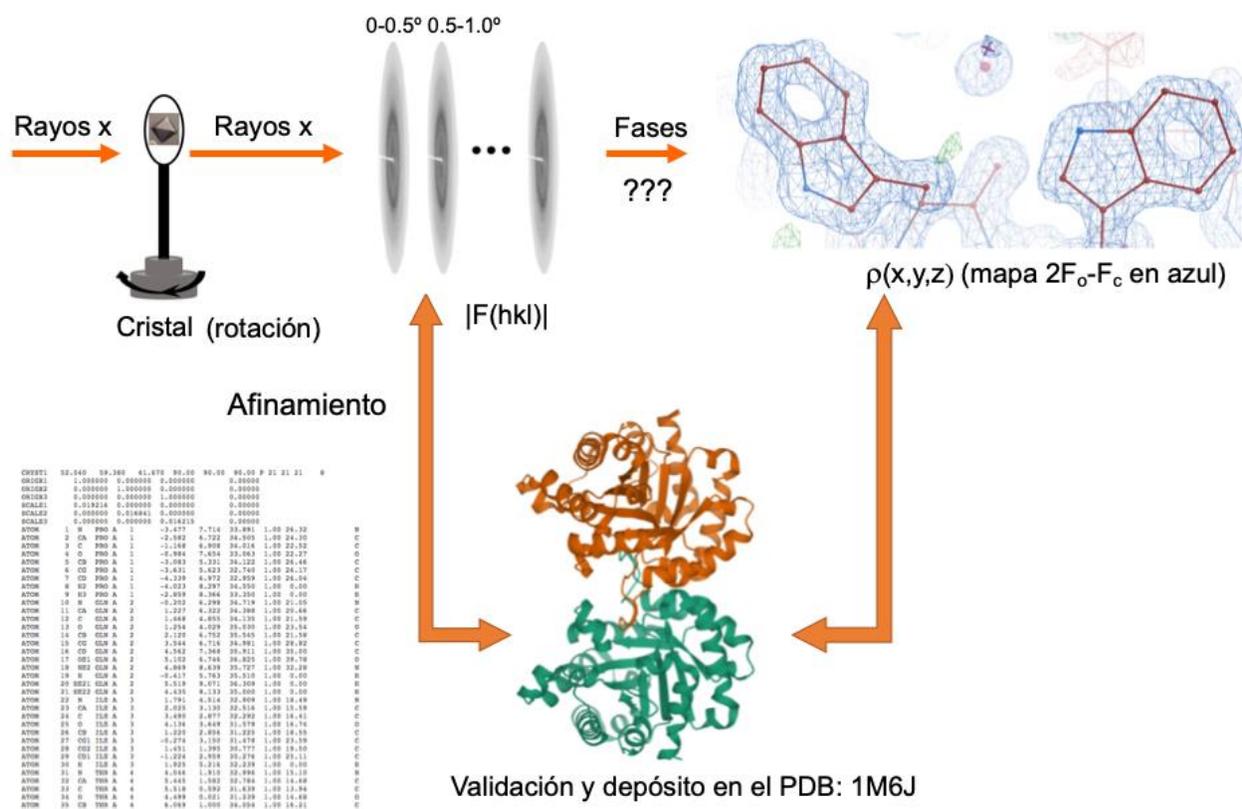


Figura 2. Etapas en la obtención de una estructura tridimensional a partir de datos de difracción de rayos X de un cristal proteico

Conclusiones

La cristalografía es una técnica poderosa que ha permitido determinar el 85% de las estructuras depositadas en el PDB, siendo históricamente la primera herramienta de la biología estructural que dio lugar a las estructuras de una amplia gama de

proteínas solubles, de membrana e incluso complejos macromoleculares como el ribosoma, el centro de reacción fotosintético y las cápsidas virales. Con el paso de los años, el estudio de los aspectos dinámicos de las macromoléculas se ha vuelto más crítico, ya que la estructura por sí sola no es suficiente para comprender la función y sus mecanismos de acción.

Por este motivo, si se desea conocer aspectos dinámicos de la estructura en estudio, se pueden utilizar además otras técnicas tales como la criomicroscopía electrónica, la resonancia magnética nuclear o bien utilizar cristales de menores dimensiones, que se someten a tiempos ultracortos de exposición de los rayos X producidos por los láseres de electrones libres (XFEL). Esta tecnología de última generación permite obtener información dinámica de macromoléculas a través de la cristalografía. XFEL permite recopilar datos de difracción de cristales de macromoléculas en diferentes conformaciones, mientras realiza su función biológica, lo que permite determinar varias estructuras de una misma proteína en distintas etapas de su función. De esta manera, utilizando una gran cantidad de instantáneas de rayos X, se puede producir una “película” en la que es

posible seguir las modificaciones moleculares. Básicamente, la dinámica molecular se obtiene a partir de datos empíricos, lo que permite comprender mejor la función y el mecanismo de acción de proteínas. Las tendencias recientes muestran que una variedad cada vez mayor de técnicas producen información mecanística detallada, por lo que los rápidos avances en este campo son muy prometedores en el futuro cercano.

Agradecimientos

Agradecemos al CONACyT por los apoyos otorgados (Proyectos 221169, 301120 y CF 2019 – 87163). Miguel A. Ramírez Rodríguez agradece al CONACyT la beca otorgada para realizar estudios de maestría (736508).

Referencias

- Jaskolski, M., Dauter, Z. and Wlodawer, A. (2014). A brief history of macromolecular crystallography, illustrated by a family tree and its Nobel fruits. *FEBS J*, 281, 3985-4009. doi:10.1111/febs.12796.2.
- Fersht, A. (2008). From the first protein structures to our current knowledge of protein folding: delights and scepticisms. *Nat Rev Mol Cell Biol*, 9, 650–654. doi:10.1038/nrm2446.
- Markwick, P. R., Malliavin, T., & Nilges, M. (2008). Structural biology by NMR: structure, dynamics, and interactions. *PLoS computational biology*, 4(9), e1000168.
- Nogales, E. (2016). The development of cryo-EM into a mainstream structural biology technique. *Nature methods*, 13(1), 24-27.
- Grimm, U. (2015). Aperiodic crystals and beyond. *Acta Cryst. B71*, 258-274. doi:10.1107/S2052520615008409.
- Lifshitz, R. (2007). What is a crystal? *Zeitschrift für Kristallographie*, 222(6), 313-317. doi:10.1524/zkri.2007.222.6.313
- McPherson, A., & Gavira, J. A. (2014). Introduction to protein crystallization. *Acta Crystallographica Section F: Structural Biology Communications*, 70(1), 2-20. doi:10.1107/S2053230X13033141.
- Kolesnikov, N., & Borisenko, E. (Eds.). (2012). *Modern Aspects of Bulk Crystal and Thin Film Preparation*. BoD–Books on Demand.
- Stan, C. V., Beavers, C. M., Kunz, M., & Tamura, N. (2018). X-ray diffraction under extreme conditions at the Advanced Light Source. *Quantum Beam Science*, 2(1), 4. doi:10.3390/qbs2010004
- Elton, L. R. B., & Jackson, D. F. (1966). X-ray diffraction and the Bragg law. *American Journal of Physics*, 34(11), 1036-1038.
- Foadi, J., & Evans, G. (2008). Elucidations on the reciprocal lattice and the Ewald sphere. *European journal of physics*, 29(5), 1059.
- Brünger, A. T. (1991). Crystallographic phasing and refinement of macromolecules: Current Opinion in Structural Biology 1991, 1: 1016–1022. *Current Opinion in Structural Biology*, 1(6), 1016-1022. doi:10.1016/0959-440X(91)90100-8.
- Patterson, A. L. (1934). A Fourier series method for the determination of the components of interatomic distances in crystals. *Physical Review*, 46(5), 372. doi:10.1103/PhysRev.46.372
- Bella, J., & Rossmann, M. G. (1998). A general phasing algorithm for multiple MAD and MIR data. *Acta Crystallographica Section D: Biological Crystallography*, 54(2), 159-174.
- Hendrickson, W. A., & Ogata, C. M. (1997). [28] Phase determination from multiwavelength anomalous diffraction measurements. In *Methods in Enzymology* (Vol. 276, pp. 494-523). Academic Press.
- Rossmann, M. G. (1990). The molecular replacement method. *Acta Crystallographica Section A: Foundations of Crystallography*, 46(2), 73-82.
- Lawson, D. M., Artymiuk, P. J., Yewdall, S. J., Smith, J., Livingstone, J. C., Treffry, A., & Harrison, P. M. (1991). Solving the structure of human H ferritin by genetically engineering intermolecular contacts. *Nature*, 349(6309), 541-544.
- Bau, R., Rees, D. C., Kurtz Jr, D. M., Scott, R. A., Huang, H., Adams, M. W., & Eidsness, M. K. (1998). Crystal structure of rubredoxin from *Pyrococcus furiosus* at 0.95 Å resolution, and the structures of N-terminal methionine and formylmethionine variants of Pf Rd. Contributions of N-terminal interactions to thermostability. *JBIC Journal of Biological Inorganic Chemistry*, 3(5), 484-493.
- Price, S. R., & Nagai, K. (1995). Protein engineering as a tool for crystallography. *Current Opinion in Biotechnology*, 6(4), 425-430.
- Henderson, R. (1990). Cryo-protection of protein crystals against radiation damage in electron and X-ray diffraction. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 241(1300), 6-8.
- Yang, C., Pflugrath, J. W., Courville, D. A., Stence, C. N., & Ferrara, J. D. (2003). Away from the edge: SAD phasing from the sulfur anomalous signal measured in-house with chromium radiation. *Acta Crystallographica Section D: Biological Crystallography*, 59(11), 1943-1957.
- Minor, W., Cymborowski, M., Otwinowski, Z., & Chruszcz, M. (2006). HKL-3000: the integration of data reduction and structure solution—from diffraction images to an initial model in minutes. *Acta crystallographica section D: biological crystallography*, 62(8), 859-866.
- Matthews, B. W. (1968). Solvent content of protein crystals. *Journal of molecular biology*, 33(2), 491-497.
- Battye, T. G. G., Kontogiannis, L., Johnson, O., Powell, H. R., & Leslie, A. G. (2011). iMOSFLM: a new graphical interface for diffraction-image processing with MOSFLM. *Acta Crystallographica Section D: Biological Crystallography*, 67(4), 271-281.

25. Kabsch, W. (2010). XDS. *Acta Crystallographica Section D: Biological Crystallography*, 66(2), 125-132.
26. Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., & Bourne, P. E. (2000). The protein data bank. *Nucleic acids research*, 28(1), 235-242.
27. Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., & Hassabis, D. (2021). Highly accurate protein structure prediction with AlphaFold. *Nature*, 596(7873), 583-589.
28. Varadi, M., Anyango, S., Deshpande, M., Nair, S., Natassia, C., Yordanova, G., & Velankar, S. (2022). AlphaFold Protein Structure Database: Massively expanding the structural coverage of protein-sequence space with high-accuracy models. *Nucleic acids research*, 50(D1), D439-D444.
29. Kim, D. E., Chivian, D., & Baker, D. (2004). Protein structure prediction and analysis using the Robetta server. *Nucleic acids research*, 32(suppl_2), W526-W531.
30. Waterhouse, A., Bertoni, M., Bienert, S., Studer, G., Tauriello, G., Gumienny, R., & Schwede, T. (2018). SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic acids research*, 46(W1), W296-W303.
31. Vagin, A., & Teplyakov, A. (1997). MOLREP: an automated program for molecular replacement. *Journal of applied crystallography*, 30(6), 1022-1025.
32. McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C., & Read, R. J. (2007). Phaser crystallographic software. *Journal of applied crystallography*, 40(4), 658-674.
33. Winn, M. D., Ballard, C. C., Cowtan, K. D., Dodson, E. J., Emsley, P., Evans, P. R., Keegan, R. M., Krissinel, E. B., Leslie, A. G., McCoy, A., McNicholas, S. J., Murshudov, G. N., Pannu, N. S., Potterton, E. A., Powell, H. R., Read, R. J., Vagin, A., & Wilson, K. S. (2011). Overview of the CCP4 suite and current developments. *Acta crystallographica. Section D, Biological crystallography*, 67(Pt 4), 235-242.
34. Adams, P. D., Afonine, P. V., Bunkóczi, G., Chen, V. B., Davis, I. W., Echols, N., & Zwart, P. H. (2010). PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallographica Section D: Biological Crystallography*, 66(2), 213-221.
35. Emsley, P., & Cowtan, K. (2004). Coot: model-building tools for molecular graphics. *Acta crystallographica section D: biological crystallography*, 60(12), 2126-2132.
36. Langer, G., Cohen, S. X., Lamzin, V. S., & Perrakis, A. (2008). Automated macromolecular model building for X-ray crystallography using ARP/wARP version 7. *Nature protocols*, 3(7), 1171-1179.
37. Terwilliger, T. C. (2003). Automated main-chain model building by template matching and iterative fragment extension. *Acta Crystallographica Section D: Biological Crystallography*, 59(1), 38-44.
38. Brünger, A. T. (1992). Free R value: a novel statistical quantity for assessing the accuracy of crystal structures. *Nature*, 355(6359), 472-475.
39. Cruickshank, D. W. J. (1999). Remarks about protein structure precision. *Acta Crystallographica Section D: Biological Crystallography*, 55(3), 583-601.
40. Winn, M. D., Isupov, M. N., & Murshudov, G. N. (2001). Use of TLS parameters to model anisotropic displacements in macromolecular refinement. *Acta Crystallographica Section D: Biological Crystallography*, 57(1), 122-133.
41. Murshudov, G. N., Skubák, P., Lebedev, A. A., Pannu, N. S., Steiner, R. A., Nicholls, R. A. & Vagin, A. A. (2011). REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallographica Section D: Biological Crystallography*, 67(4), 355-367.
42. Sheldrick, G. M. (2008). A short history of SHELX. *Acta Crystallographica Section A: Foundations of Crystallography*, 64(1), 112-122.
43. Laskowski, R. A., MacArthur, M. W., Moss, D. S., & Thornton, J. M. (1993). PROCHECK: a program to check the stereochemical quality of protein structures. *Journal of applied crystallography*, 26(2), 283-291.
44. Vriend, G. (1990). WHAT IF: a molecular modeling and drug design program. *Journal of molecular graphics*, 8(1), 52-56.
45. Vaguine, A. A., Richelle, J., & Wodak, S. J. (1999). SFCHECK: a unified set of procedures for evaluating the quality of macromolecular structure-factor data and their agreement with the atomic model. *Acta Crystallographica Section D: Biological Crystallography*, 55(1), 191-205.
46. Davis, I. W., Murray, L. W., Richardson, J. S., & Richardson, D. C. (2004). MOLPROBITY: structure validation and all-atom contact analysis for nucleic acids and their complexes. *Nucleic acids research*, 32(suppl_2), W615-W619.
47. DeLano WL. Pymol: An open-source molecular graphics tool. *CCP4 Newsletter on Protein Crystallography*. 2002:82-92.
48. Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C., & Ferrin, T. E. (2004). UCSF Chimera--a visualization system for exploratory research and analysis. *Journal of computational chemistry*, 25(13), 1605-1612.



DRA. ADELA RODRIGUEZ ROMERO
ORCID: 0000-0001-7641-6545

La Dra. Rodríguez obtuvo el título de Ingeniera Bioquímica por la ENCB-IPN. Realizó estudios de Maestría y Doctorado en Química en la UAM, especializándose en Cristalografía de Proteínas y obteniendo el Premio a la mejor Tesis Doctoral en Química, otorgado por la Facultad de Química y el Instituto de Química de la UNAM.

Se incorporó al Instituto de Química en 1984 y actualmente se desempeña como Investigadora Titular "C" de Tiempo Completo, tiene el nivel III en el SNI, el nivel D del PRIDE y recibió el Reconocimiento Sor Juana Inés de la Cruz 2006. Ha realizado dos estancias sabáticas como investigadora invitada en el National Institute of Standards and Technology, Maryland, EUA, realizando estudios estructurales de proteínas con interés industrial.

La Dra. Rodríguez es reconocida por sus aportaciones en el campo de la biología estructural y

la consolidación de la cristalografía de proteínas en la UNAM y en el País. Las principales líneas de investigación de la Dra. Rodríguez abarcan diversos aspectos de la bioquímica estructural y la fisicoquímica de proteínas, incluyendo proteínas alergénicas y anticuerpos monoclonales. Ella ha publicado más de 80 artículos en revistas indizadas, varios artículos de divulgación y capítulos de libro y ha dirigido 14 tesis de licenciatura, 12 de maestría y 10 de doctorado y tiene en proceso 5 más. Su compromiso institucional se refleja en las diversas labores académico-administrativas que ha desempeñado en la UNAM, entre las que se encuentra el de ser la Sria. Académica (2001-2003) y jefa del Departamento de Bioquímica del IQ-UNAM (2003-2007), consejera universitaria (2012-2014), Coordinadora del Programa de Maestría y Doctorado en Ciencias Químicas (2014-2017), responsable del Programa de Doctorado en Ciencias Biomédicas-IQ (1997-2003) y fue la Coordinadora de la Rama de Química de la Academia Mexicana de Ciencias en el periodo 2005-2006.

La Dra. Rodríguez es evaluadora de diversas revistas internacionales como *Nature communications*, *Acta Crystallographica*, *BBActa*, *FEBS Letters*, *FEBS Journal*, *Proteins*, *JMB*, *Process Biochemistry*, *PLOS ONE*, *Biochimie*, *Journal of Eukaryotic Microbiology*, entre otras. En cuanto a divulgación de la ciencia ha participado en programas de Radio, Televisión, dando pláticas en diferentes foros como el Museo Universum, en diversas ENP y CCH; ha participado en notas periodísticas y en el Programa Domingos en la Ciencia. Además, su trabajo científico se ha presentado en diversos foros, congresos y simposios, lo que se resume en alrededor de 34 conferencias tanto nacionales como internacionales.

La Dra. Rodríguez es responsable del LANEM-IQ-UNAM a partir del 2006.