

Cristalografia de Proteínas

Shaker Chuck Farah IQ-USP (chsfarah@iq.usp.br)

Introdução

Proteínas executam a maioria das funções específicas dentro de uma célula no nível molecular, incluindo o reconhecimento de substratos, catálise de reações químicas, transporte de moléculas através de membranas, controle de fluxo de informação gênica codificado por ácidos nucléicos e a montagem de complexos mutiprotéicos importantes para a construção e manutenção de estruturas celulares.

A estrutura tridimensional de uma proteína é determinada por sua sequência de aminoácidos. Há milhões de sequências de proteínas em bancos de dados públicos (ex: <http://www.ncbi.nlm.nih.gov/>) derivados de projetos de sequenciamento de genomas que podem ser agrupadas (alinhas) em famílias de domínios que são relacionados por descendência evolucionária (ex: <http://pfam.wustl.edu/>). Nesta descendência, pequenas modificações em sequência normalmente não afetam a estrutura tridimensional e logo a função específica é mantida. Portanto, diferenças maiores entre as sequências dentro de uma família protéica podem refletir diferenças significativas em suas estruturas e conseqüentemente em suas funções. Às vezes, a divergência evolucionária é tão grande que não existe mais semelhança significativa em suas sequências apesar de manutenção de estruturas e funções comuns. Estimativas do número de famílias de estruturas ou **topologias (“folds”)** protéicas na biosfera variam de entre 1000 e 4000.

A função de uma proteína está diretamente relacionada com sua estrutura. A nossa capacidade de entender esta relação nos seus maiores detalhes é o fator limitante da nossa capacidade de entender processos biológicos e a base molecular da vida. Esta relação é a chave para realizar o potencial de poder manipular o mundo orgânico em benefício à saúde humana, por exemplo o desenho racional de novas drogas e vacinas, o desenho de intervenções moleculares para combater deficiências hereditárias e a produção de animais e plantas transgênicas com propriedades que aumentam rendimento e valor nutricional sem introduzir riscos aos delicados equilíbrios ecológicos.

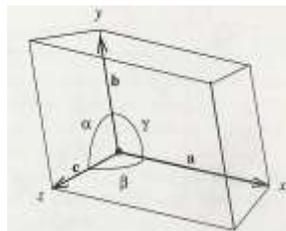
Existem três maneiras de obter informação detalhada (resolução atômica) sobre estruturas proteicas: **cristalografia**, **ressonância magnética nuclear (RMN)** e mais recentemente a técnica de **criomicroscopia eletrônica (Cryo-EM)**. As três técnicas são complementares: Enquanto a cristalografia em geral permite uma maior resolução e determinação de estruturas maiores, RMN

fornece, além de informação estrutural, detalhes sobre processos dinâmicos relacionados ao intercâmbio entre conformações em solução, algumas das quais podem ser transitórias e resistentes à cristalização. Cryo-EM permite estudar proteínas ou complexos macromoleculares que, devido ao seu tamanho e/ou flexibilidade intrínseca não podem ser estudados por cristalografia ou RMN.

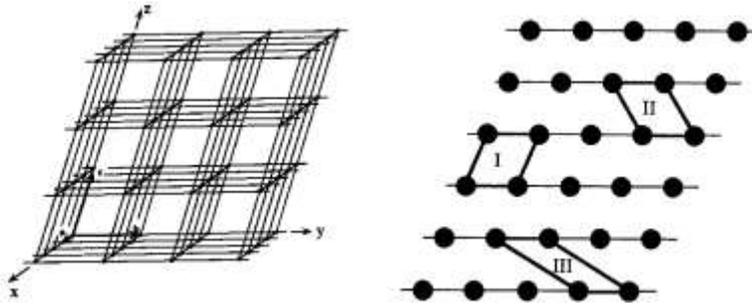
Nas próximas semanas, nós vamos contemplar os vários procedimentos envolvidos na determinação da estrutura tridimensional de uma proteína. Devido às limitações de espaço e de tempo, os detalhes da maioria dos procedimentos serão omitidos. Aos que estão interessados em tratamentos mais profundos sobre o assunto, recomendamos consultar o Material Bibliográfico.

Teoria

Cristais de Proteínas. Em condições específicas, macromoléculas podem formar cristais. Cristalógrafos produzem cristais de proteínas por um processo lento e controlado em condições não desnaturantes. Por definição, cada cristal é um arranjo **periódico** e tridimensional das moléculas, íons ou átomos que o compõe. Contudo, pode-se imaginar um cristal como sendo formado por unidades básicas (paralelepípedos), denominadas de **células unitárias**, que se repetem no espaço por simples translações ao longo de três eixos. Esta definição facilita o estudo cristalográfico já que o conteúdo de uma célula unitária é o mesmo para qualquer outra célula unitária. Além disto, a determinação da configuração de uma célula unitária possibilita determinar a estrutura de todo o cristal. A célula unitária tem uma origem e três eixos a partir dos quais as posições de cada átomo podem ser definidas. A célula unitária de um cristal pode ser descrita por 3 vetores (**a**, **b**, **c**) e 3 ângulos (α , β , γ).

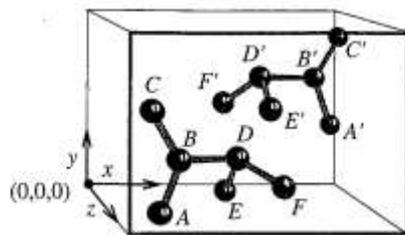


Se desconsiderarmos, por hora, o conteúdo de cada célula unitária, o cristal pode ser visto como uma **rede** de pontos derivados dos vértices das células unitárias.

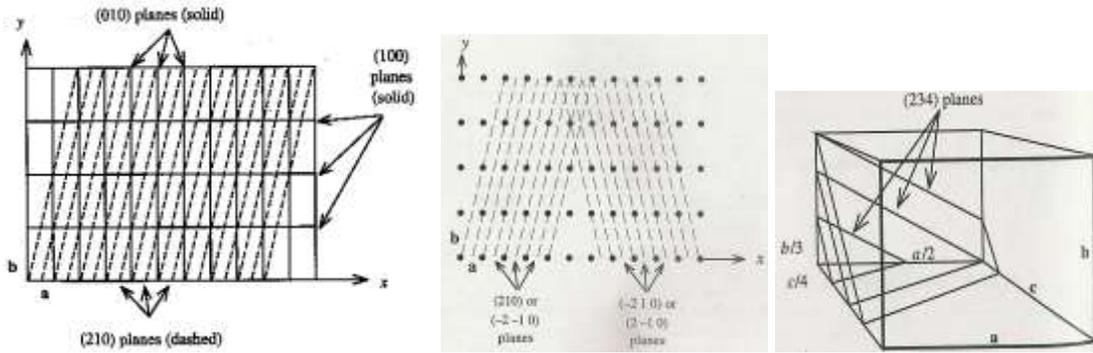


Dentro de cada célula unitária, podem existir um ou mais conjuntos de moléculas equivalentes que são relacionados entre si por **operações de simetria** (combinações de rotações de 60, 90, 120, ou 180° e translações de 1/6, 1/4, 1/3 ou 1/2 do eixo). Estas operações possibilitam subdividir cada célula unitária em regiões menores e equivalentes denominadas de **unidades assimétricas**. Assim, o conteúdo de uma unidade assimétrica estará relacionado com o conteúdo de outra unidade assimétrica por uma determinada operação de simetria. Note que a estrutura de um cristal pode ser determinada com base na estrutura presente em uma única unidade assimétrica e na simetria da célula unitária.

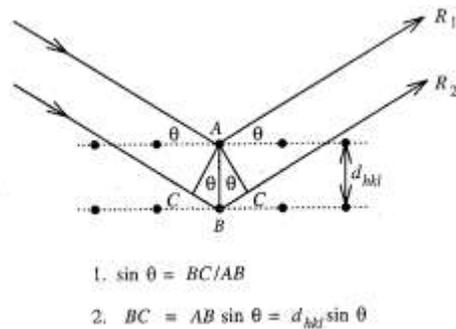
A simetria dentro de uma célula unitária é descrita por seu **grupo espacial**, representado por um símbolo, por exemplo $P2_1$, $P2_12_12_1$ etc. Você pode descrever as unidades assimétricas da figura abaixo? Elas são relacionadas por quais operações de simetria?



Fontes de Raios X. Raios X são uma forma de radiação eletromagnética com comprimento de onda (λ) entre 0,1 e 100 Å (0,01 e 10 nm). Eles podem ser produzidos de várias maneiras, duas das quais são relevantes para cristalografia: (i) Bombardeamento de um metal (normalmente Cu ou Mb) com elétrons. Um elétron de alta energia colide com um elétron de um orbital de baixa energia do metal resultando na sua ionização. Um elétron de um orbital de alta energia do metal então cai para o orbital vazio, emitindo energia na forma de um fóton de raio X. ii) Emissão de **radiação síncrotron**. Elétrons são acelerados até quase a velocidade de luz em anéis com circunferências de dezenas de metros até quilômetros denominados síncrotrons. A desaceleração dos elétrons nas



Imagine, portanto, um cristal sendo bombardeado por um feixe monocromático de raios-X. Nesta condição, é possível imaginar um conjunto de planos cristalinos imaginários orientados com um ângulo θ em relação aos feixes incidente e refletido, como mostra a figura abaixo. A Lei de Bragg diz que quando a diferença de caminho dos raios que atingem dois ou mais planos ($2BC = 2d \sin \theta$) for igual a um múltiplo inteiro ($n = 1, 2, 3, 4 \dots$) do comprimento de onda da radiação incidente ($n\lambda$), teremos uma condição que resulta em interferência construtiva para a onda difratada (raio-X refletido).

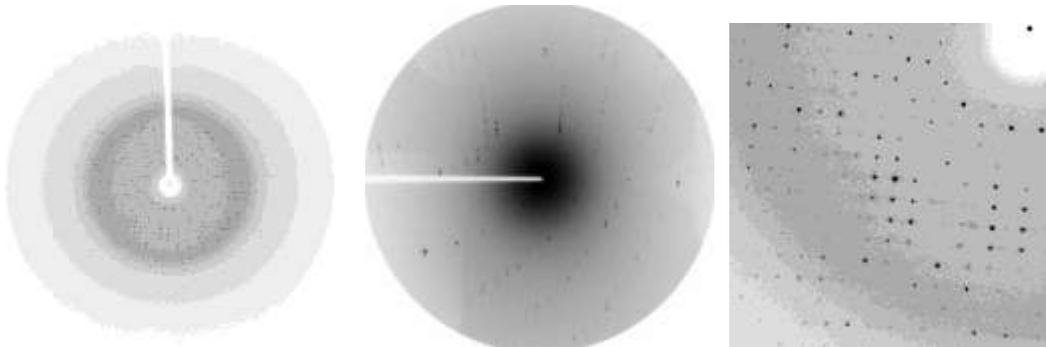


Diferença de caminho entre raios R_1 e $R_2 = 2BC = 2AB \sin \theta = 2d \sin \theta$

Condição de Bragg: ocorre interferência construtiva somente quando $n\lambda = 2d \sin \theta$

Cada conjunto de planos hkl tem uma distância (d) entre os planos e um ângulo (θ) em relação à radiação incidente/refletida específicos. Logo, cada conjunto de planos produz uma única **reflexão** somente quando o cristal está orientado de uma maneira que satisfaça à condição de Bragg. Isto gera um padrão de reflexões em três dimensões cujas intensidades podem ser medidas por detectores.

Observação: A Lei de Bragg apenas prevê onde e sob quais condições uma reflexão (ponto de difração) será vista. Contudo, uma teoria mais quantitativa é necessária para descrever em detalhes a intensidade de cada um dos pontos de difração (reflexões).



Espaço real vs espaço recíproco. Notar que as considerações acima implicam que as famílias de planos com distâncias (d) pequenas são representadas por índices (hkl) grandes e possuem ângulos de reflexão (θ) também grandes. Logo, reflexões derivadas de planos com índices grandes se encontram na periferia do padrão de difração. Também, reflexões derivadas de um determinado plano hkl em células unitárias grandes terão ângulos menores que células unitárias pequenas.

Desta relação entre distâncias dentro do cristal e dentro do padrão de difração derivamos a relação entre **espaço real** (do cristal com coordenadas x, y, z) e **espaço recíproco** (do padrão de difração com coordenadas h, k, l). Pode ser demonstrado que cada ponto no espaço recíproco (cada reflexão hkl) se localiza no final de um vetor com comprimento $1/d_{hkl}$, que parte da origem e é perpendicular à família de planos com índices hkl (cujos planos são separados pela distância d_{hkl}). Também, para cada célula unitária com eixos a, b e c , existe uma célula unitária recíproca cujos eixos tem comprimentos proporcionais a $1/a, 1/b$ e $1/c$. Logo quando um cristal (**rede real**) é rodado no espaço real, podemos associar a **rede recíproca** rodando no espaço recíproco e sendo medida no detector.

Equações de ondas e funções periódicas. Até agora, nossa consideração sobre o padrão de difração ignorou o conteúdo da célula unitária (de fato, as **posições** das reflexões são **independentes** do conteúdo da célula unitária e somente dependem das dimensões da célula unitária e outras variáveis geométricas da coleta dos dados). Para entender como a informação do padrão de difração pode ser

utilizada para determinar as estruturas das moléculas dentro da célula unitária (mais precisamente da unidade assimétrica), precisamos considerar a natureza das ondas eletromagnéticas.

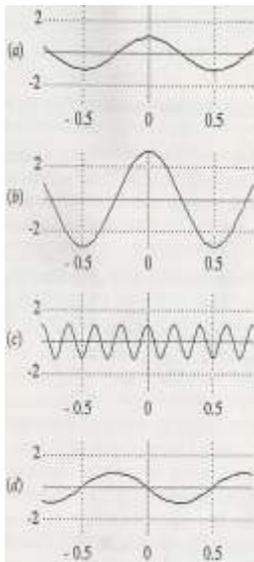
Ondas simples podem ser descritas por funções periódicas simples, por exemplo:

$$f(x) = F \cos 2\pi(hx + \phi) \quad \text{ou}$$

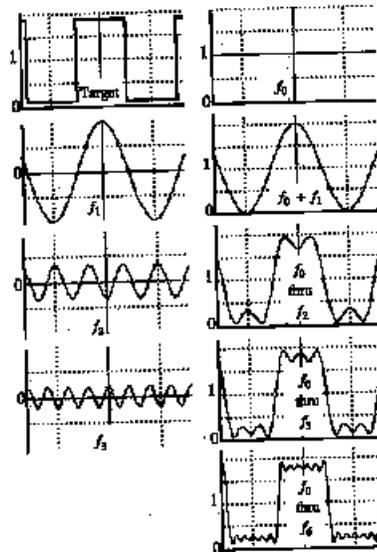
$$f(x) = F \cos 2\pi(hx) + i \sin 2\pi(hx) \quad \text{ou}$$

$$f(x) = F e^{2\pi i(hx)}$$

onde $f(x)$ = amplitude da onda na posição x , F = amplitude máxima da onda, h = frequência, ϕ = fase ou posição da onda em relação à origem. Inspeccionar o diagrama 1 (a), (b), (c) e (d) abaixo e determinar os valores de F , h e ϕ em cada caso.



1



2

Cristais são funções periódicas complexas em três dimensões. Imagine-se andando numa linha reta dentro de um cristal. A cada passo você encontrará conjuntos de átomos diferentes em posições diferentes em relação a você. Contudo, depois de andar um pouco você começará a sentir um pouco de *deja vu* ... “... eu já vi aquela histidina e aquela serina antes!” Daria para perceber que um padrão complexo está se repetindo. O que você realmente iria perceber são flutuações nas **densidades** das núvens **eletrônicas**! Caminhando agora numa outra direção constataria que um outro padrão

também está se repetindo. Isso ilustra que cada caminhada numa direção diferente produz uma função (padrão) periódica complexa de flutuações de densidade eletrônica. Logo, **um cristal pode ser visto como uma função periódica complexa em três dimensões**. O objetivo da cristalografia é de obter uma função periódica, $\rho(x, y, z)$, cujo gráfico representa **o mapa de densidade eletrônica** das moléculas na célula unitária.

Série de Fourier. O matemático francês Jean Baptiste Joseph Fourier demonstrou que qualquer função periódica, independente de sua complexidade, pode ser descrita como uma soma de funções periódicas simples (ex: seno ou co-seno) cujos comprimentos de onda são frações inteiras do comprimento de onda da função complexa original (vide parte 2 da figura acima onde a onda original-alvo (quadrada; “target”) pode ser aproximada pela soma de ondas periódicas simples $f_0 - f_6$).

Em uma dimensão: $f(x) = \sum F_h e^{2\pi i(hx)}$

Em três dimensões: $f(x, y, z) = \sum \sum \sum F_{hkl} e^{2\pi i(hx+ky+lz)}$

Transformada de Fourier. Fourier também demonstrou que para qualquer função $f(x)$ é possível estabelecer uma outra função $F(h)$, tal que:

$$F(h) = \int f(x) e^{2\pi i(hx)} dx$$

A função $F(h)$ é conhecida como a Transformada de Fourier (FT) da função $f(x)$.

Uma das propriedades principais desta função está relacionada à reciprocidade. Se, por exemplo, $f(x)$ é uma função de tempo, $F(h)$ é uma função de frequência e se $f(x)$ for uma função de distância (d), $F(h)$ será uma função de distância recíproca (onde o variável h tem unidades de 1/d).

Desta maneira a FT é uma ferramenta extremamente útil para descrever a relação entre o espaço real e espaço recíproco e logo entre um objeto e seu padrão de difração.

Outra propriedade importante da Transformada de Fourier é que ela é reversível; ou seja:

Se $F(h) = \int f(x) e^{2\pi i(hx)} dx$, então $f(x) = \int F(h) e^{-2\pi i(hx)} dh$

Em outras palavras: se $F(h)$ é a FT de $f(x)$, então $f(x)$ é a FT inversa (FT^{-1}) de $F(h)$.

Em três dimensões:

$$F(h,k,l) = \iiint f(x,y,z) e^{2\pi i(hx+ky+lz)} dx dy dz \quad \text{e} \quad f(x,y,z) = \iiint F(h,k,l) e^{-2\pi i(hx+ky+lz)} dh dk dl$$

Fatores de Estrutura. Cada reflexão hkl no padrão de difração pode ser considerada uma onda complexa que foi formada com contribuições de:

(a) todos os átomos:

$$\mathbf{F}_{hkl} = \sum \mathbf{f}_j e^{2\pi i(hx_j+ky_j+lz_j)}$$

onde o f_j é o fator de espalhamento do átomo j e $2\pi(hx_j+ky_j+lz_j)$ é a contribuição do átomo j para a fase da reflexão hkl.

ou

(b) todos os elementos de volume da célula unitária:

$$\mathbf{F}_{hkl} = \iiint \rho(\mathbf{x},\mathbf{y},\mathbf{z}) e^{2\pi i(hx+ky+lz)} d\mathbf{x}d\mathbf{y}d\mathbf{z}$$

Esta onda complexa pode ser descrita como uma série de Fourier cuja soma é chamada o **fator de estrutura** F_{hkl} . A magnitude de F_{hkl} é proporcional à raiz quadrada da intensidade da reflexão hkl $(I_{hkl})^{1/2}$.

A equação anterior mostra que F_{hkl} é a Transformada de Fourier de $\rho(\mathbf{x},\mathbf{y},\mathbf{z})$. Logo, $\rho(\mathbf{x},\mathbf{y},\mathbf{z})$ é a Transformada de Fourier inversa de F_{hkl} :

$$\rho(x,y,z) = (1/V) \sum_h \sum_k \sum_l F_{hkl} e^{-2\pi i(hx+ky+lz)}$$

onde V é o volume da célula unitária e

F_{hkl} é uma série de Fourier cujo resultado final pode ser representado por um número complexo de amplitude ($|F_{hkl}|$) e fase (ϕ_{hkl}).

Pode ser demonstrado que

$$\rho(x,y,z) = (1/V) \sum_h \sum_k \sum_l |F_{hkl}| e^{-2\pi i(hx+ky+lz + \phi_{hkl})}$$

Esta equação mostra como podemos obter a densidade eletrônica em qualquer posição (x,y,z) de uma célula unitária de um cristal a partir do padrão de difração:

As intensidades das reflexões fornecem $|F_{hkl}|$, e os índices fornecem h, k, l.

Mas, como podemos medir as fases de cada uma das milhares de reflexões?

O problema das fases. O problema é que durante o experimento de coletar as imagens do padrão de difração, as informações das fases são perdidas. **A resolução deste problema das fases é um dos passos mais difíceis na determinação de uma estrutura cristalina.**

Existem 3 maneiras principais que podem ser utilizadas para estimar as fases das reflexões. Aquela que você utiliza depende de alguns fatores:

i) Substituição molecular. Se dispomos de um modelo estrutural que é suficientemente similar à estrutura da proteína experimental (cristalizada), pode-se utilizar este modelo para calcular $\rho(x,y,z)$ que, por sua vez pode ser utilizada para (através de sua FT) calcular a padrão de difração do modelo (amplitudes dos fatores de estrutura $|F_{hkl}|_{calc}$, incluindo suas fases ϕ_{calc}). Estas fases calculadas podem ser então utilizadas, em combinação com as intensidades observadas experimentalmente ($|F_{hkl}|_{obs}$), para calcular $\rho(x,y,z)$ na célula unitária da proteína que foi cristalizada:

$$\rho(x,y,z) = (1/V) \sum_h \sum_k \sum_l |F_{hkl}|_{obs} e^{-2\pi i(hx+ky+lz + \phi_{calc})}$$

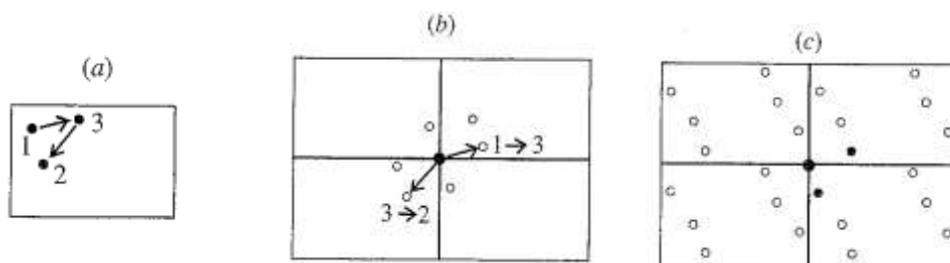
A qualidade das fases calculadas depende da semelhança estrutural entre o modelo e a estrutura sendo estudada. Também existe a questão da **orientação e posição** das moléculas protéicas dentro da célula unitária. Na verdade, a **orientação e posição** das moléculas protéicas do modelo dentro da célula unitária é fundamental para os cálculos dos mapas de densidade eletrônica. A orientação do modelo é definida por três ângulos de rotação e sua posição final será definida pelas translações ao longo dos três eixos da célula unitária. Existem milhões de milhões de possíveis combinações (pelo menos $100^6 = 10^{12}$) destas seis variáveis, o que impossibilita testar todas as possibilidades. O que é feito é: primeiro, determinar a orientação (somente três variáveis <**ângulos eulerianos α, β, γ** >) e depois a posição (mais três variáveis <**translações** ao longo dos três eixos da célula unitária>). Consulte o material bibliográfico para informações mais detalhadas sobre como procurar separadamente as funções de rotação e translação.

Brevemente:

Orientação: Existe uma função chamada **Função de Patterson** que é uma série de Fourier **sem fases** e com termos com amplitudes que são quadrados de fatores de estrutura:

$$P(\mathbf{u}, \mathbf{v}, \mathbf{w}) = (1/V) \sum_h \sum_k \sum_l |F_{hkl}|_{\text{obs}}^2 e^{-2\pi i(hu + kv + lw)}$$

Devido à ausência das fases, uma Função de Patterson pode ser calculada para qualquer conjunto de dados de difração. O ponto importante é que pode ser demonstrado que o gráfico tridimensional $P(\mathbf{u}, \mathbf{v}, \mathbf{w})$ da Função de Patterson (chamado de **mapa de Patterson**) tem picos correspondendo aos **vetores entre todos os pares de átomos** na célula unitária. Pensando um pouco sobre isso, temos que para uma célula unitária com N átomos, existem $N(N-1)$ ou quase N^2 vetores quando N for grande. Logo um mapa de Patterson é muito mais complexo do que um mapa de densidade eletrônica. Por exemplo, para uma célula unitária contendo 3 átomos (a), existem seis vetores (b) que produzem o mapa de Patterson mostrado em (c). Pode também ser percebido que **o mapa de Patterson é independente da posição da molécula dentro da célula unitária mas sofreria rotações se a molécula é rodada.**



Estas considerações apontam a uma maneira em que podemos encontrar a melhor orientação do nosso modelo: calcular mapas de Patterson para várias orientações do nosso modelo (definidas pelos ângulos eulerianos) e compará-los com o mapa de Patterson calculado dos dados experimentais. A melhor sobreposição entre os mapas possibilitará determinar a correta orientação do modelo dentro da célula unitária

Posição: Utilizaremos o modelo na sua melhor orientação para agora posicioná-lo dentro da célula unitária. Em cada posição do modelo podemos calcular as amplitudes dos fatores de estrutura ($|F_{\text{calc}}|$) e compará-las com as amplitudes atualmente observados no experimento de difração ($|F_{\text{obs}}|$). Escolhe-se, dentre todas as opções, aquela em que a soma de todas as diferenças entre $|F_{\text{obs}}|$ e $|F_{\text{calc}}|$ é minimizada:

$$\frac{\sum ||F_{\text{hkl}}|_{\text{obs}} - |F_{\text{hkl}}|_{\text{calc}}|}{\sum |F_{\text{hkl}}|_{\text{obs}}}$$

Vários fatores podem contribuir para o sucesso desta estratégia de calcular as fases; os mais importantes sendo: a) o nível de identidade da seqüência primária - normalmente precisa de identidades >30%, b) a ausência de inserções ou deleções em uma estrutura em relação a outra e c) o número de moléculas na unidade assimétrica.

ii e iii) *Métodos dependentes de átomos pesados.* Se não existe um modelo de boa qualidade da nossa estrutura para estimar as fases por substituição molecular, precisamos recorrer a outros métodos para o cálculo das fases dos fatores de estrutura. Estes métodos envolvem a adição de sais de metais pesados ao cristal (ie Pb^{2+} , Hg^{2+} , Au^+ , I) ou a incorporação de selenometionina na proteína durante sua produção heteróloga. Além disso, algumas proteínas naturalmente ligam íons metálicos como cofatores (ex.: Fe^{2+} , Zn^{2+} , Mn^{2+}). A adição de átomos mais pesados do que C, N, O ao cristal modificará a padrão de difração em relação ao padrão nativo. Uma vez que a contribuição deste novo átomo para cada uma das reflexões do padrão de difração não se dá de forma igual (depende da posição deste átomo dentro da célula unitária), espera-se que algumas reflexões sejam mais afetadas do que outras no novo padrão de difração. Logo, é esperado que a adição de um metal pesado irá perturbar algumas reflexões mais do que outras. É justamente esta diferença nas intensidades das reflexões o que possibilita determinar as fases dos fatores de estrutura.

Substituição isomorfa: Comparando os padrões de difração de um **crystal nativo** (obtido sem metal pesado) e de **crystal derivados** (com o metal pesado), as pequenas diferenças nas intensidades de reflexões específicas podem ser utilizadas para determinar as posições dos átomos pesados (mas não dos átomos da proteína). Isso feito, os fatores de estrutura (amplitudes e fases), de todas as reflexões obtidos do cristal com metal pesado ($|F_{PH}|$ e ϕ_{PH}) podem ser separados em contribuições dos átomos pesados ($|F_H|$ e ϕ_H) e da parte protéica ($|F_P|$ e ϕ_P). As variáveis $|F_H|$ e ϕ_H podem ser obtidas pelas posições conhecidos dos átomos pesados, o $|F_P|$ é obtido do padrão de difração do cristal nativo ($|F_{obs}|$; sem átomos pesados) e a desejada ϕ_P pode então ser estimada por procedimentos descritos no Material Bibliográfico e nas aulas. O sucesso desta estratégia depende da obtenção de cristais que incorporam o metal pesado sem perturbar a estrutura da proteína e as dimensões da célula unitária. Em outras palavras, necessita de **crystal derivados isomorfos**. Por isso, esta metodologia é chamada **Substituição Isomorfa**.

Dispersão ou espalhamento anômalo. Átomos podem absorver raios X e emití-los novamente. Este fenômeno acontece somente em comprimentos de ondas específicos. Neste processo de absorção e emissão, ocorre uma mudança de intensidade e de fase dos raios X difratados. Este processo é chamado **dispersão ou espalhamento anômalo**. Logo, se conjuntos de dados do mesmo cristal são coletados em dois comprimentos de onda, um dos quais correspondente ao comprimento de absorção de raios X pelo metal pesado, as mudanças nos padrões de difração podem ser utilizadas para obter informações sobre as fases (ver Material Bibliográfico discutido nas aulas). Além disso, na ausência de dispersão anômala as intensidades e as fases das reflexões hkl e $-h-k-l$ são iguais. Na presença de dispersão anômala esta identidade desapareça (por razões que também são explicados no Material Bibliográfico). As diferenças entre as reflexões $|F_{hkl}|$ e $|F_{-h-k-l}|$ são muito pequenas (normalmente somente alguns % da intensidade) mas podem ser medidas porque são obtidas com alta precisão do mesmo conjunto de dados do mesmo cristal. Estas diferenças também podem ser utilizadas para obter estimativas iniciais sobre as fases.

Obtenção do mapa de densidade eletrônica e refinamento do modelo molecular. Independentemente do método que você utilizou para obter a primeira estimativa das fases, seu primeiro mapa de densidade eletrônica será calculado simplesmente a partir da seguinte equação:

$$\rho(x,y,z) = (1/V) \sum \sum \sum |F_{hkl}|_{obs} e^{-2\pi i(hx+ky+lz + \phi_{calc})}$$

(Obs.: Alguns fatores que expressam pesos de qualidade podem ser adicionados à fórmula sem alterar o real sentido da mesma.)

A inspeção deste mapa deve então apontar regiões que são obviamente derivadas de proteínas em que a cadeia principal (e se tiver sorte algumas cadeias laterais) pode ser reconhecida. Nestas regiões, podemos criar modelos tridimensionais utilizando nossos conhecimentos da sequência primária da proteína específica e nossos conhecimentos básicos de química de proteínas em geral (ângulos e distâncias de ligações químicas, modelos eletrônicos de átomos específicos, principalmente C, N, O e S). Estes modelos podem ser utilizados para calcular as intensidades e fases das reflexões num **processo iterativo** em que procuramos fases calculadas melhores que diminuam as diferenças entre $|F_{hkl}|_{obs}$ e $|F_{hkl}|_{calc}$.

Cada iteração envolve passos de: a) inspeção e manipulação do modelo para que ela encaixe melhor no mapa de densidade eletrônica, b) a construção de novas regiões da molécula em regiões de densidade eletrônica que eram ausentes no ciclo anterior, c) pequenos ajustes no modelo que procuram ao mesmo tempo aproximar valores geométricos das ligações químicas às valores ideais e minimizar a diferença entre $|F_{hkl}|_{obs}$ e $|F_{hkl}|_{calc}$, d) cálculo de novas fases derivadas do novo modelo e e) cálculo de um novo mapa de densidade eletrônica usando as novas fases.

Notas importantes:

Veja que mapas do tipo

$$\rho(x,y,z) = (1/V) \sum_h \sum_k \sum_l |F_{hkl}|_{obs} - |F_{hkl}|_{calc} e^{-2\pi i(hx+ky+lz + \phi_{calc})}$$

chamadas **mapas de diferença** ou **mapas Fo-Fc** enfatizam erros no modelo, com regiões positivas onde está faltando densidade eletrônica (destacando a falta de um átomo ou átomos no modelo atual) e regiões negativas evidenciam um excesso de elétrons no seu modelo naquela região. Estes mapas são úteis na correção de pequenos erros na estrutura no final do processo de refinamento.

De maneira similar, mapas do tipo

$$\rho(x,y,z) = (1/V) \sum_h \sum_k \sum_l |2|F_{hkl}|_{obs} - |F_{hkl}|_{calc} e^{-2\pi i(hx+ky+lz + \phi_{calc})}$$

chamadas **mapas 2Fo-Fc** são utilizados para visualizar a qualidade do encaixe do modelo com os dados. Estes mapas devem ser sempre positivos e, escolhendo contornos acima de uma densidade eletrônica específica, devem revelar uma superfície molecular.

Note que ao final do refinamento, quando não mais houver diferença entre os dados experimentais e os dados do modelo, o primeiro mapa tenderá a zero enquanto o mapa segundo tenderá a ser calculado apenas com as amplitudes experimentais

Em cada ciclo de refinamento do modelo, monitoramos quanto o modelo está consistente com os dados experimentais, calculando o fator residual (R)

$$R = \frac{\sum ||F_{hkl}|_{obs} - |F_{hkl}|_{calc}|}{\sum |F_{hkl}|_{obs}}$$

A grande maioria das estruturas nos bancos de dados tem valores de $R < 0.3$ e muitos têm valores < 0.2 .

Material Bibliográfico:

Rhodes, G. (2006) Crystallography Made Crystal Clear. 3rd Edition (Academic Press).

Textos mais avançados:

Rupp, B. (2010) Biomolecular Crystallography (Garland Science)

Drenth, J. (1999) Principles of Protein X-Ray Crystallography (Springer).

Blow, D. (2002) Outline of Crystallography for Biologists (Oxford University Press).

McPherson, A. (2009) Introduction to Macromolecular Crystallography (Wiley-Blackwell)