

Refinamento Cristalográfico

Objetivos:

Realizar o refinamento cristalográfico de estrutura de proteína com o programa REFMAC (CCP4, 1994). Aprender os principais passos do refinamento cristalográfico de proteínas.

Materiais:

1. Computador iMac;
2. Pacote CCP4
3. Programa VMD (*Visual Molecular Dynamics*) para visualização de macromoléculas biológicas;
4. Arquivos com coordenadas atômicas de proteínas
5. Fatores de estrutura calculados.

Introdução

O último passo na resolução de uma estrutura de proteína por cristalografia é o refinamento cristalográfico. As coordenadas atômicas obtidas do processo de solução do problema da fase são usadas como estrutura inicial do processo de refinamento. Ao resolvermos o problema da fase obtemos um modelo inicial da estrutura que precisa de ajuste. No caso da solução do problema da fase por substituição molecular, onde as coordenadas atômicas usadas como modelo de busca foram de uma estrutura com diferenças na estrutura primária, necessitamos no refinamento corrigir as diferenças. Nesta fase usamos a sequência de aminoácidos da estrutura a ser resolvida e ajustamos o modelo em refinamento, modificando os resíduos para que a sequência de aminoácidos seja a da proteína que foi resolvida. No refinamento cristalográfico ajustamos as posições das cadeias laterais do modelo sendo refinado no mapa de densidade eletrônica.

Podemos analisar o refinamento cristalográfico, como um processo computacional onde as coordenadas atômicas da estrutura que foi resolvida são ajustadas de forma que os fatores de estrutura calculados (F_{calc}) fiquem o mais próximo possível dos fatores de estrutura observados. O que temos é que os F_{calc} são dados pela equação 1,

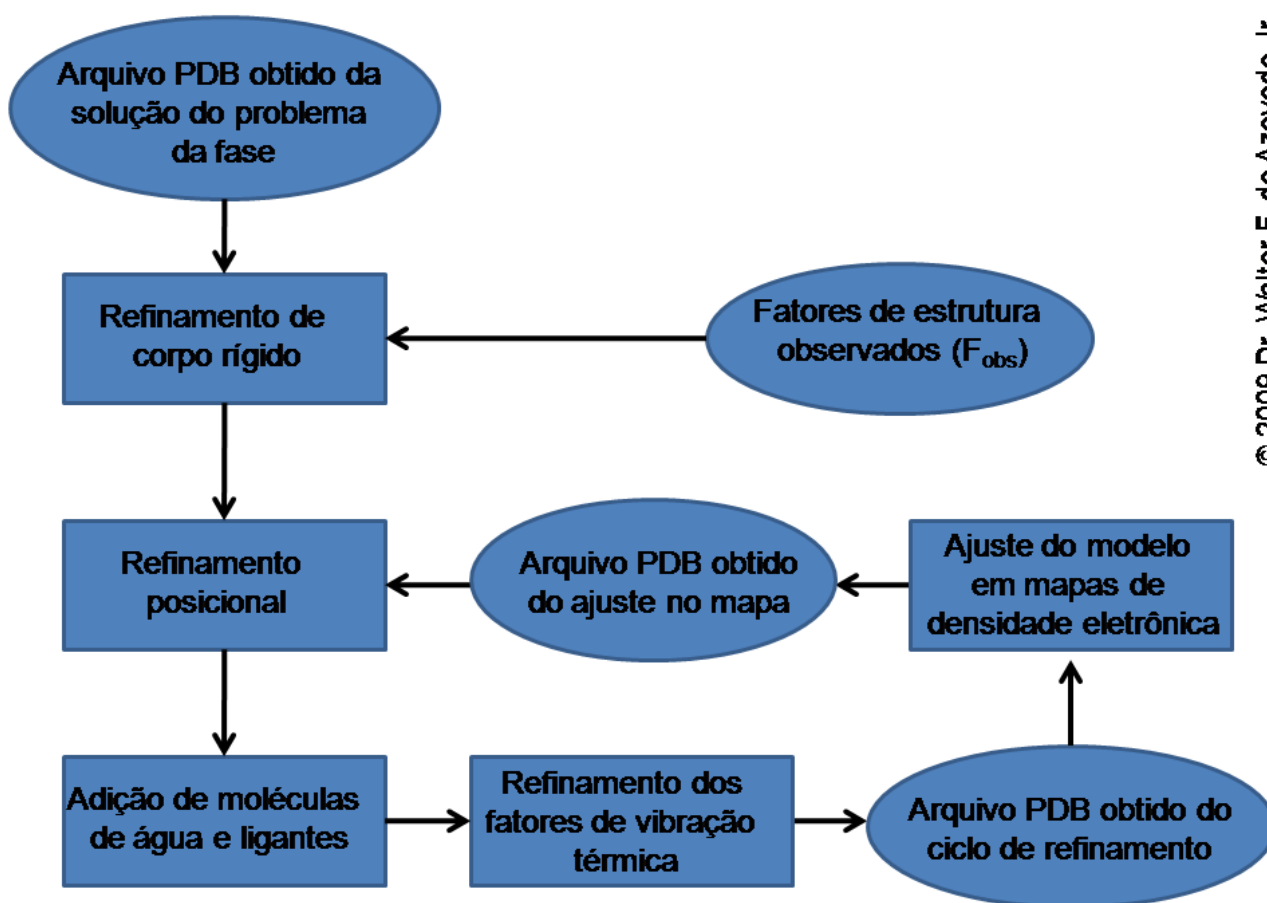
$$F_{\text{calc}}(hkl) = \sum_{j=1}^N f_j \exp 2\pi i(hx_j + ky_j + lz_j) \quad (\text{Eq. 1})$$

onde f_j são os fatores de espalhamento atômico, (hkl) são os índices da reflexão e (x_j, y_j, z_j) são as coordenadas atômicas do modelo que está em refinamento cristalográfico. Para cada ajuste nas coordenadas atômicas do modelo podemos calcular os F_{calc} . Este é um processo iterativo, onde em cada ciclo de refinamento ajustamos as coordenadas atômicas (x_j, y_j, z_j) com o objetivo de trazer o F_{calc} o mais próximo possível do F_{obs} . Já podemos adiantar que este processo normalmente é o mais longo do ponto de vista computacional e, dependendo da estrutura, teremos que fazer diversos ciclos de refinamento.

A figura 1 ilustra um fluxograma para aplicação do refinamento cristalográfico. Há diversos protocolos de refinamento cristalográfico, que dependem do tipo de problema a ser resolvido. O que faremos aqui ilustra os principais conceitos atrás do refinamento cristalográfico.

Vamos considerar um caso mais simples do refinamento cristalográfico, onde a sequência de aminoácidos do modelo inicial já está ajustado. Este tipo de situação na verdade é bem comum. Normalmente em projetos de desenho de drogas resolvemos a estrutura da proteína alvo sem ligantes, e depois resolvemos a estrutura da mesma proteína complexada com inibidores, com objetivo de identificarmos as bases estruturais para a interação proteína-ligante. Assim, as coordenadas atômicas do modelo de entrada não precisam de ajustes da sequência de aminoácidos. O passo seguinte é o refinamento de corpo rígido, que na maioria das vezes é omitido, visto que o passo final da substituição molecular é exatamente um ajuste de corpo rígido, onde a molécula é considerada um corpo rígido com liberdade de rotação e translação, sem ajustes das posições relativas dos átomos. O passo seguinte é o refinamento posicional, agora os átomos terão liberdade de se movimentarem, assim as posições relativas dos átomos são

modificadas, de forma a trazer o F_{calc} o mais próximo possível do F_{obs} . Veja bem, sempre que ajustamos as coordenadas atômicas (x_j, y_j, z_j) modificamos o F_{calc} , veja a equação 1, mas o F_{obs} não é alterado, visto que esta informação vem diretamente do experimento de difração de raios X.



© 2009 Dr. Walter F. de Azevedo Jr.

Figura 1. Fluxograma do refinamento cristalográfico.

Os F_{obs} não são alterados durante o refinamento cristalográfico, eles servem de guia. As coordenadas atômicas devem ser modificadas para trazer o F_{calc} próximo do F_{obs} . Vimos que na substituição molecular temos que maximizar o coeficiente de correlação e minimizar o R_{factor} . Este continua sendo nosso objetivo no refinamento cristalográfico. O coeficiente de correlação deve estar alto durante o refinamento, normalmente ao final do refinamento esperamos coeficiente acima de 0,75, lembrando o coeficiente de correlação é dado por:

$$CC = \frac{\sum ab - \frac{\sum a \sum b}{N}}{\left\{ \left[\sum a^2 - \frac{(\sum a)^2}{N} \right]^{1/2} \right\} \left\{ \left[\sum b^2 - \frac{(\sum b)^2}{N} \right]^{1/2} \right\}} \quad (\text{Eq. 2})$$

onde os parâmetros a e b na equação acima são os fatores de estrutura observados ($|F_{\text{obs}}|$) e os fatores de estrutura calculado ($|F_{\text{calc}}|$), os fatores de estrutura calculados são obtidos a partir das coordenadas atômicas em refinamento. A somatória é para todas as N reflexões. Outro parâmetro muito útil na análise do refinamento é o R_{factor} . Este índice analisa a proximidade do $|F_{\text{calc}}|$ com o $|F_{\text{obs}}|$, neste caso esperamos valores baixos para R_{factor} , este parâmetro é calculado pela seguinte equação:

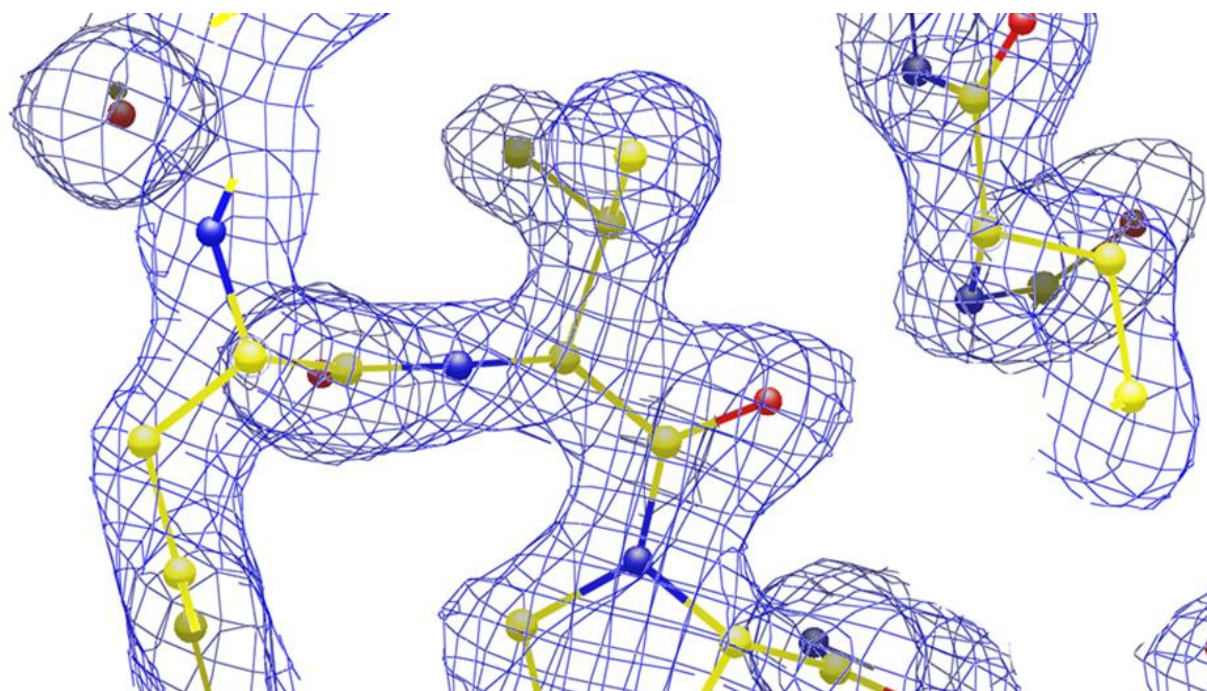
$$R_{factor} = \frac{\sum_{h=0}^{N1} \sum_{k=0}^{N2} \sum_{l=0}^{N3} [|\mathbf{F}_{obs}(hkl)| - |\mathbf{F}_{calc}(hkl)|]}{\sum_{h=0}^{N1} \sum_{k=0}^{N2} \sum_{l=0}^{N3} [|\mathbf{F}_{obs}(hkl)|]} \quad (\text{Eq. 3})$$

onde N1, N2 e N3 são os índices máximos para h, k e l, respectivamente. No final do refinamento cristalográfico esperamos valores próximos de 0,20 ou mesmo abaixo deste para o R_{factor} . Analisamos durante o refinamento cristalográfico o R_{free} que é calculado da mesma forma que o R_{factor} , a única diferença é que o R_{free} usa somente parte das reflexões para seu cálculo, normalmente entre 5 e 10 % das reflexões. A idéia deste índice é que de 5 a 10 % das reflexões sejam deixadas de fora do processo de refinamento, assim serve como um índice independente do processo de refinamento cristalográfico. Normalmente esperamos que o R_{free} seja mais alto que o R_{factor} , no máximo 12 % acima. Se a diferença entre os dois foi maior que 12 % faz-se necessário uma análise detalhada do refinamento e do modelo sendo refinado.

Depois do refinamento posicional o processo continua com a adição de moléculas de água. Lembrem-se, a proteína está solvatada no cristal, e muitas moléculas de águas são passíveis de identificação cristalográfica. Esta é uma parte crítica do refinamento. Esperamos que as águas identificadas no refinamento cristalográfico sigam os seguintes critérios:

- 1) Não estejam muito afastadas da estrutura da proteína, normalmente no máximo 4 Å de distância de um átomo aceitador e/ou doador de ligação de hidrogênio (oxigênio ou nitrogênio).
- 2) As moléculas de água serão inicialmente posicionadas em picos de densidade eletrônica, não preenchidos pela estrutura da proteína.

O número de moléculas da água adicionadas ao modelo em refinamento normalmente não excedem a 10 % dos átomos da proteína (átomos diferentes de hidrogênio). O próximo passo é o refinamento dos fatores de vibração térmica, agora ao final podemos inspecionar o mapa de densidade eletrônica. Ao inspecionarmos o mapa esperamos encontrar o modelo encaixado na densidade eletrônica (indicada num gradeado azul). A figura 2 mostra trechos de densidade eletrônica onde vemos o modelo da estrutura encaixado.



© 2009 Dr. Walter F. de Azevedo Jr.

Figura 2. Mapa de densidade eletrônica (azul) e modelo estrutural encaixado (valina no centro).

Há diversos tipos de mapas de densidade eletrônica, o que foi gerado na figura 2 é um mapa $2F_{\text{obs}} - F_{\text{calc}}$, ou simplesmente $2F_o - F_c$. Ao calcularmos a densidade eletrônica, como a expressa pela equação (4),

$$\rho(x, y, z) = \frac{1}{V} \sum_h \sum_k \sum_l |\mathbf{F}(hkl)| \exp\{-2\pi i[(hx + ky + lz) - \alpha'(hkl)]\} \quad (\text{Eq. 4})$$

podemos substituir o $|\mathbf{F}(hkl)|$ pela diferença entre $2F_o$ e F_c , tais mapas são indicados para verificarmos a diferença entre o que já foi identificado na estrutura e o que ainda não foi identificado, assim a densidade eletrônica fica,

$$\rho(x, y, z) = \frac{1}{V} \sum_h \sum_k \sum_l (|2\mathbf{F}_o(hkl)| - |\mathbf{F}_c(hkl)|) \exp\{-2\pi i[(hx + ky + lz) - \alpha'(hkl)]\} \quad (\text{Eq. 5})$$

Os mapas $2F_o - F_c$ ainda mantêm a densidade da parte que já foi identificada, como mostrado na figura 2.

Como destacado, o refinamento cristalográfico visa modificar as posições atômicas, de forma a aproximar o F_{calc} do F_{obs} . Os programas de refinamento cristalográfico empregam diversos algoritmos para atingir tal objetivo, mas todos visam minimizar uma função pseudo-energia (E_{ref}), que apresentará seu mínimo quando o F_{calc} estiver o mais próximo possível do F_{obs} , esta função tem a seguinte forma geral,

$$E_{\text{ref}} = w_A \sum_{hkl} [F_{\text{calc}}(hkl) - F_{\text{obs}}(hkl)]^2 + V \quad (\text{Eq. 6})$$

Esta equação apresenta um peso para os termos cristalográficos (w_A) e uma energia potencial que tem como objetivo manter a topologia da estrutura o mais coerente possível. O programa REFMAC (CCP4, 1994) usa o método do "maximum likelihood", para minimizar esta função.

Durante a inspeção do mapa de densidade eletrônica podemos promover ajustes manuais nas posições dos átomos, gerando um novo modelo para a estrutura. As coordenadas atômicas desse novo modelo podem ser submetidas a um novo ciclo de refinamento cristalográfico. O processo se repete até que tenhamos convergido para uma estrutura satisfatória. Ao final de cada etapa de refinamento (figura 1) podemos gerar e visualizar o mapa de densidade eletrônica, não sendo necessária a finalização de um ciclo completo. Na verdade, na fase final do refinamento podemos omitir etapas, como por exemplo, a adição de moléculas de água. Outro aspecto do refinamento, nós podemos acrescentar ligantes ao modelo, caso estejam presentes. Os mapas de densidade podem apresentar densidades eletrônicas que não são nem proteína nem moléculas de água. Para um cristal de uma proteína complexada com ligante, nós podemos ter a densidade de um ligante, como mostrado na figura 3. O ligante acyclovir está bem ajustado na densidade eletrônica da purina nucleosídeo fosforilase.

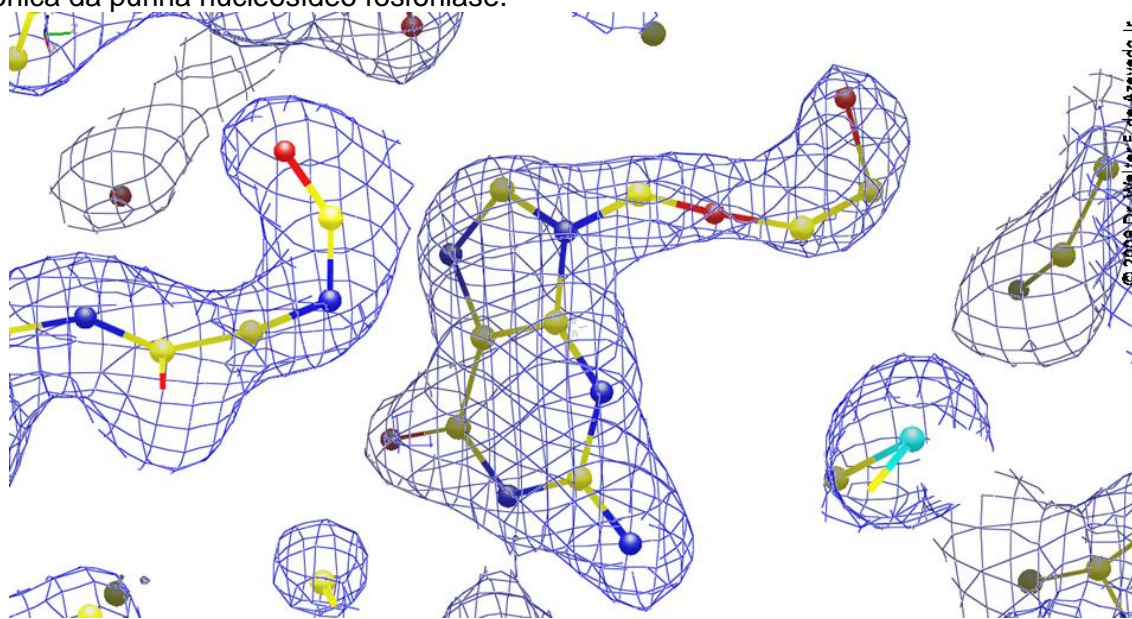


Figura 3. Mapa de densidade eletrônica (azul) e modelo estrutural do acyclovir encaixado.

Usaremos na prática de hoje os dados de difração de raios X da citidina deaminase de *Mycobacterium tuberculosis* (CDA; EC 3.5.4.5) (Sánchez-Quitian et al., 2010), coletados no laboratório nacional de luz síncrotron (LNLS, Campinas, SP). Este cristal difratou a 1,99 Å de resolução e a análise dos dados difração indicaram os seguintes parâmetros de cela unitária:

Dimensões da cela unitária:

$$a = 63,7508 \text{ \AA} \quad b = 75,3638 \text{ \AA} \quad c = 55,1327 \text{ \AA} \quad \alpha = 90.0^\circ \quad \beta = 90.0^\circ \quad \gamma = 90.0^\circ$$

O grupo espacial é C222 (número 21) e foram coletadas 9374 reflexões. Todas estas informações estão no arquivo data.mtz. Usaremos as coordenadas atômicas obtidas na solução por substituição molecular.

Para a presente prática precisaremos dos arquivos de comandos refmac1.sh, refmac2.sh e refmac3.sh, além dos arquivos de coordenadas model2.pdb e model2_water.pdb, e de dados de difração o data.mtz.

Procedimento

Para abrirmos o terminal usaremos a opção de procura, a lupa do lado direito superior, achem o terminal, digitando "terminal". Aparecerá na janela do "finder" o ícone "Terminal". Clique duas vezes nele. Aparecerá um Terminal de comandos, agora você tem acesso ao "Darwin" o "core" do sistema operacional "OS X", que funciona como um terminal de comandos do sistema operacional Linux. Trabalharemos na pasta bioestrutural. Achem esta pasta. Para termos acesso aos programas do CCP4 temos que ativá-lo, para isto digite no terminal:

```
source /usr/local/ccp4-6.0.2/bin/ccp4.setup-sh
```

e pressione *Return*.

Agora todos os programas do CCP4 estão ativos.

Vamos mudar para pasta onde temos os arquivos, após cada comando pressione a tecla *Return*, digite:

```
pwd
```

Este comando mostra em que pasta você se encontra.

Mude para a pasta onde estão os arquivos que você usará. Digite:

```
cd Disciplinas/BioestruturalAula11
```

Agora confirme em que pasta você se encontra. Nesta pasta temos diversos arquivos. Digite o comando para verificar o conteúdo da pasta.

```
ls -l
```

Na tela você terá todos os arquivos disponíveis nesta pasta.

Vamos verificar o conteúdo deste arquivo com os comandos do primeiro passo. Digite:

```
more refmac1.sh
```

Será mostrado na tela.

```
#!/bin/sh
set -e
# Simple example of using REFMAC5
refmac5 \
  HKLIN data.mtz \
  HKLOUT data_refmac1.mtz \
  XYZIN model2.pdb \
  XYZOUT model2_refmac1.pdb << eor
#
# Input/output column assignments
#
LABIN FP=F_New SIGFP=SIGF_New FREE=FreeR_flag
LABO FC=FC PHIC=PHIC FWT=2FOFCWT DELFWT=FOFCWT
#
# Type of refinement,
#
REFI TYPE RIGID RESI MLKF
#
#
BFAC SET 20 ! initialise B factors to constant value
MAKE HYDR N ! ignore hydrogens
MONI MEDI
NCYC 20
END
eor
```

Nesta parte faremos o refinamento de corpo rígido, indicado pelo comando “REFI TYPE RIGID”. Os outros comandos indicam arquivos de entrada e saída, e formatos de dados de entrada e saída, identifique-os no arquivo acima. Depois digite:

```
chmod +x *.sh
```

Este comando do Linux garante que o arquivos que terminal com .sh possam ser executados. Depois digite:

```
./refmac1.sh >refmac1.log
```

A primeira parte da linha de comando indica que o arquivo refmac1.sh deve ser executado, o que acionará o programa REFMAC, o símbolo “>” indica que o resultado do programa, que não seja arquivo de saída, ficará registrado no arquivo refmac1.log. Este arquivo armazena as ações realizadas pelo programa REFMAC.

Digite o comando `ls -l` para verificar se todos os arquivos foram criados.

O arquivo PDB gerado traz as coordenadas atômicas (model2_refmac1.pdb) e o data_refmac1.mtz.

Digite:

```
> tail -100 refmac1.log
```

Você verá a evolução do Rfact e Rfree para cada ciclo de refinamento. Nesta fase os R's não abaixaram muito.

A próxima etapa é o refinamento posicional, os comandos estão no arquivo `refmac2.sh`. A principal diferença é a linha “REFI TYPE REST”, que indica para o REFMAC que faremos refinamento posicional.

Depois digite:

```
./refmac2.sh >refmac2.log
```

Verifique o final do arquivo `refmac2.log`. Os R's abaixaram bastante.

Nesta etapa adicionamos as moléculas de água. Para isto podemos usar programas gráficos como o COOT (CCP4, 1994), que basicamente analisam mapas de densidade eletrônica do tipo $F_o - F_c$ e atribuem uma molécula de água para os picos que estão próximos a doadores e aceitadores de ligação de hidrogênio da proteína. As coordenadas das proteínas com moléculas de água estão no arquivo `model2_water.pdb`, que será usado no último ciclo de refinamento. Neste ciclo realizaremos refinamento posicional para proteína e das águas, bem como dos fatores de vibração térmica. Os comandos estão no `refmac3.sh`

Digite:

```
./refmac3.sh >refmac3.log
```

Verifique o final do arquivo `refmac3.log`.

Edite o arquivo `model2_refmac3.pdb` e delete as moléculas de água que apresentam fatores de vibração termina acima de 45 \AA^2 e salve o arquivo com o mesmo nome, `model2_refmac3.pdb`. Para achar as moléculas de água no arquivo PDB, procure as linhas que iniciam com “ATOM” e tenha a identificação “HOH”.

Rode mais um ciclo de refinamento posicional.

Digite:

```
./refmac4.sh >refmac4.log
```

Repita o processo, até que os valores de Rfact e Rfree não abaixem mais. Tome cuidado de sempre mudar os nomes dos arquivos de entrada `refmacX.sh` e de saída `refmacX.log`, de forma que tenhamos um histórico da evolução do refinamento.

Anote as seguintes informações do último ciclo de refinamento cristalográfica (armazenado no último arquivo `refmacx.log`), onde x é número de ciclos de refinamento que você realizou.

<i>Resolution limits</i>	=
<i>Number of used reflections</i>	=
<i>Percentage observed</i>	=
<i>Percentage of free reflections</i>	=
<i>Overall R factor</i>	=
<i>Free R factor</i>	=

Visualização do resultado

Clique no ícone VMD (Humphrey et al., 1996). Para carregar as coordenadas atômicas do último arquivo PDB refinado, arraste o arquivo PDB para a tela gráfica. Você terá a estrutura na tela gráfica. As representações em linhas indicam ligação covalente para cada átomo presente. Azul para uma ligação covalente saindo de um átomo de nitrogênio, vermelho para uma ligação covalente saindo de um átomo de oxigênio, branco para uma ligação covalente saindo de um átomo de hidrogênio e azul claro para uma ligação covalente saindo de um átomo de carbono. Gire o monômero com o botão do mouse da esquerda. Identifique as moléculas de água presentes na estrutura.

Referências:

Collaborative Computational Project No. 4. *Acta Crystallogr.* **D50**, 760-763, 1994.

Humphrey, W., Dalke, A. and Schulten, K. *J. Molec. Graphics*, **14**, 33-38, 1996.

Sánchez-Quitian ZA, Schneider CZ, Ducati RG, de Azevedo WF, Junior CB, Basso LA, Santo DS.

J Struct Biol 2010; 169(3): 413-423.