

Substituição Molecular

Objetivos:

Aprender o uso do programa AMoRe (CCP4, 1994; Navaza, 1994) para a solução do problema da fase. Analisar os resultados da substituição molecular implementada no programa AMoRe. Visualizar os resultados da solução.

Materiais:

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

Introdução

Um dos principais métodos usados para a solução do problema da fase é o método de substituição molecular, proposto originalmente por Rossmann e Blow (Rossmann & Blow, 1962). O método usa como princípio o fato que proteínas que compartilham similaridade na estrutura primária, compartilharão similaridade na estrutura terciária. Assim, se temos uma estrutura tridimensional resolvida para uma proteína, podemos usar as coordenadas atômicas dela para resolver uma proteína similar, para a qual os dados de difração de raios X foram coletados. Vamos dar um exemplo. Considere a estrutura tridimensional da enzima citidina deaminase de *E. coli* tem sua estrutura tridimensional resolvida, e que as coordenadas atômicas dessa estejam disponíveis no *Protein Data Bank*. Podemos usar estas coordenadas atômicas para resolver a estrutura da citidina deaminase de outro organismo, desde que a identidade sequencial entre as duas proteínas seja alta, preferencialmente acima de 30 % de identidade.

O problema da fase surge devido ao fato de ao coletarmos dados de difração de raios X de um cristal, a informação relativa das fases é perdida. Tudo que medimos é a intensidade difratada $I(hkl)$, de onde podemos tirar o módulo do fator de estrutura. Como os raios X são ondas eletromagnéticas, eles interagem com a parte eletrônica do cristal, e geram os raios X difratados. Os raios X difratados trazem informação sobre a posição de cada átomo da cela unitária. Resolver uma estrutura tridimensional de uma proteína significa exatamente isto, determinar as coordenadas atômicas da proteína. Sem dúvida tal empreendimento é um desafio intelectual, visto que temos que montar um quebra-cabeça de milhares de peças, onde cada peça é um átomo da proteína a ser resolvida. Nosso guia em tal tarefa é a densidade eletrônica do cristal. Como os raios X interagem com a parte eletrônica do cristal, a única informação que podemos derivar do padrão de difração de raios X, é a densidade eletrônica, que representamos pela letra grega ρ . Numa cela unitária temos uma densidade eletrônica para cada ponto (x,y,z) dessa cela, assim o que temos é uma função $\rho(x,y,z)$, ou seja, para cada ponto x,y,z da cela unitária temos um valor de densidade eletrônica ρ .

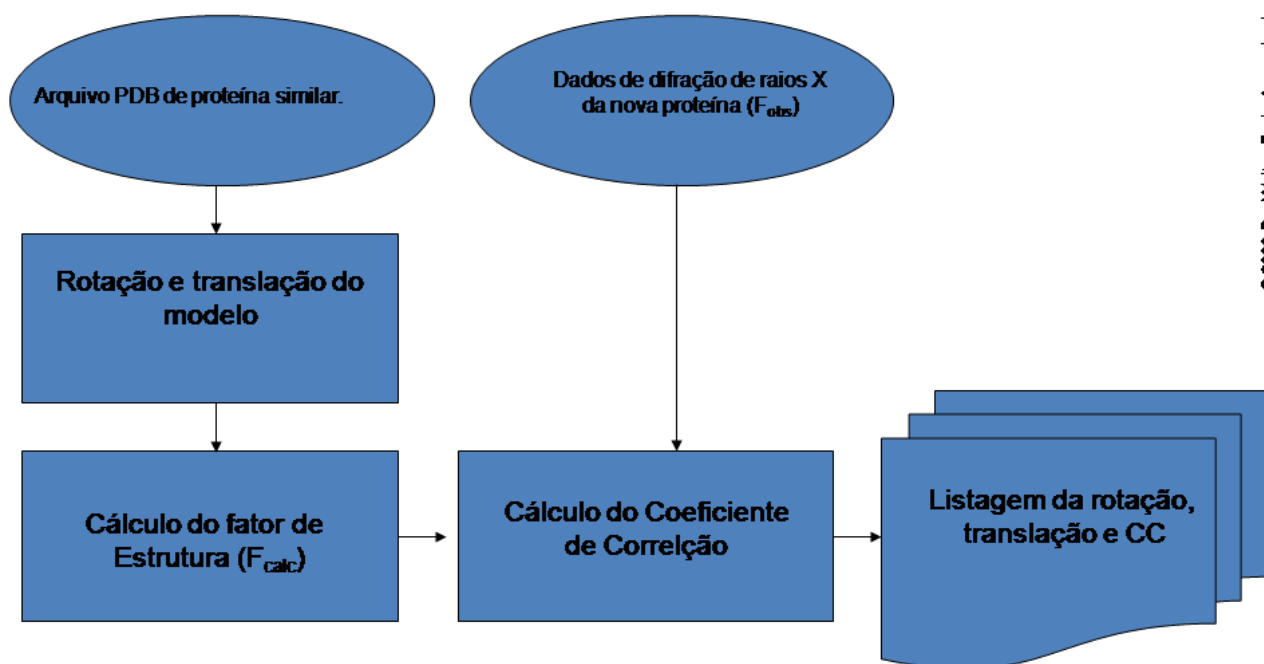
Quando resolvemos o problema da fase podemos calcular a densidade eletrônica para cada ponto da cela unitária, aí o problema resume-se a interpretarmos a densidade eletrônica. Sabemos que no cristal, onde há elétrons há átomos, assim nosso quebra-cabeça ganha o espaço tridimensional, e passamos a brincar com um gigantesco lego tridimensional, onde nossas peças são os átomos da proteína.

A densidade eletrônica $\rho(x,y,z)$ pode ser expressa como uma somatória de Fourier, onde cada termo da série de Fourier é um fator de estrutura $F(hkl)$, obtido do experimento de difração de raios X, a equação é a seguinte:

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

Onde V é o volume da cela unitária e α' é a fase do fator de estrutura. A somatória é feita para todas as reflexões disponíveis. A fase é perdida no processo de registro da imagem de difração de raios X, só temos informação sobre o módulo do fator de estrutura $|F(hkl)|$.

Veremos hoje o método de substituição molecular para solução do problema da fase. A figura 1 ilustra um fluxograma com os principais passos para solução do problema da fase a partir da aplicação do método de substituição molecular.



© 2009 Dr. Walter F. de Azevedo Jr.

Figura 1. Fluxograma da substituição molecular.

Na prática de hoje consideraremos a seguinte situação. Dados de difração de raios X de um cristal de proteína foram coletados e indexados, ou seja, seus parâmetros de cela unitária foram determinados (a , b , c , α , β e γ) e os índices hkl obtidos para todas as reflexões. Assim temos um arquivo de dados de difração de raios X, chamados aqui de *data.mtz*. A extensão *.mtz* significa que está num formato interno a ser usado pela suíte de programas CCP4. Este arquivo traz os módulos dos fatores de estrutura $|F(hkl)|$ para cada reflexão hkl , chamamos estes dados de F_{obs} , fator de estrutura observado, ou seja, obtido experimentalmente. Outro parâmetro de entrada é o arquivo PDB da proteína que será usada como modelo inicial, chamamos este arquivo de *model.pdb*. Este arquivo de coordenadas atômicas foi obtido diretamente do banco de dados PDB, usando como critério de busca a identidade seqüencial, ou seja, buscamos no PDB a proteína que apresentava a maior identidade seqüencial com a estrutura que queremos resolver.

A procura de rotação e translação é uma busca com 6 graus de liberdades. A realização das duas buscas concomitantemente gera um alto custo computacional. Consideremos que temos 10 posições para cada ângulo e 10 posições para cada eixo de translação. Teremos neste caso 10^6 cálculos, dividindo-se em busca de rotação primeiro e, em seguida, a busca de translação. Teremos 10^3 cálculos para rotação e 10^3 cálculos para translação, ou seja, 2000 cálculos, uma redução considerável. Suponha que cada cálculo demore 1 s de CPU, para o cálculo dividindo-se a busca em rotação e translação teremos 2000 s, pouco mais de meia hora, para a procura

concomitante teremos 1.000.000 de segundos 277,77 horas (mais de 11 dias!!!). Por isso fazemos uma busca de rotação e selecionamos o melhor resultado, este melhor resultado é submetido a uma busca de translação.

Há diversos critérios usados para considerar que uma solução é adequada, o mais direto é o coeficiente de correlação, esta análise estatística indica se há correlação entre duas grandezas. Por exemplo, podemos considerar duas amostras, uma contendo a altura de 100 pais e outra contendo a altura de 100 filhos. Será que há correlação entre as alturas dos filhos com as alturas dos pais? Se houver, o coeficiente de correlação (CC) é próximo de 1, se não houver será próximo de zero, o coeficiente de correlação é dado por:

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

No caso do exemplo das alturas dos pais e filhos a e b são pares de altura do pai e filho, e a somatória é para todas as N alturas, considerando que cada pai tem somente um filho. No caso da análise dos resultados da substituição molecular, os parâmetros a e b na equação acima são os fatores de estrutura observados ($|F_{obs}|$) e os fatores de estrutura calculado ($|F_{calc}|$) os fatores de estrutura calculados são obtidos a partir das coordenadas atômicas do *model.pdb* que foi submetido à rotação e translação, a partir da equação seguinte:

$$\mathbf{F}(hkl) = \sum_{j=1}^N f_j \cdot \exp[2\pi i(hx_j + ky_j + lz_j)]$$

A somatória é para todas as N reflexões. Resumindo, se ao rodarmos e transladarmos o modelo (*model.pdb*), obtivermos uma estrutura com alto coeficiente de correlação teremos uma solução. Mas este não é o único parâmetro analisado. Outro parâmetro muito útil na análise da solução é o R_{factor} . Este índice analisa a proximidade do $|F_{calc}|$ com o $|F_{obs}|$, neste caso esperamos valores baixos para R_{factor} , este parâmetro é calculado pela seguinte equação:

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

na fase da substituição molecular é difícil definir um valor de corte para o R_{factor} , depende muito da identidade do modelo de busca com a estrutura a ser resolvida, normalmente se a identidade entre modelo e proteína a ser resolvida é alta, acima de 80 % temos R_{factor} abaixo de 40 %, ou próximo disso, caso o estado oligomérico do modelo seja o mesmo da proteína a ser resolvida. O coeficiente de correlação normalmente é mais robusto como critério de seleção.

Usaremos na prática de hoje os dados de difração de raios X da citidina deaminase de *Mycobacterium tuberculosis* (CDA; EC 3.5.4.5), 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 e foram coletadas 9374 reflexões. Todas estas informações estão no arquivo *data.mtz*.

Para a presente prática precisaremos dos seguintes arquivos de comandos.

Arquivo	Função principal
<i>tabling.sh</i>	Lê o <i>model.pdb</i> e gera um novo modelo <i>model1.pdb</i> com o centro da estrutura na origem. Coloca o modelo com os eixos principais coincidentes com os eixos principais dos momentos de inércia. A saída do programa são as coordenadas <i>model1.pdb</i> e <i>search.tab</i> (fatores de estrutura calculados).
<i>sorting.sh</i>	Lê o <i>data.mtz</i> e reordena os fatores de estrutura.
<i>rotating.sh</i>	Aqui começa a substituição molecular. Rotação do <i>model1.pdb</i> e retorna os ângulos de rotação das melhores soluções.
<i>traing.sh</i>	Translação do <i>model1.pdb</i> , onde foram aplicados os ângulos de rotação do passo anterior, e retorna as translações (Tx, Ty e Tz) das melhores soluções.
<i>fiting.sh</i>	Ajuste fino do modelo, onde a melhor solução do passo anterior é ajustada de forma a gerar uma solução com maior coeficiente de correlação.
<i>lsqkab.sh</i>	Muda a posição do <i>model1.pdb</i> para os ângulos de translação determinados no passo anterior. Geramos a solução no arquivo <i>model2.pdb</i>

Esta estrutura foi resolvida e depositada junto ao PDB com código de acesso 3jif (Sánchez-Quitian et al., 2010).

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 (*/Users/labioquest/Disciplinas/Bioestrutural/Aula9*). 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/Bioestrutural/Aula9
```

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 o comando do primeiro passo. Digite:

```
more tabling.sh
```

Será mostrado na tela.

```
#####3
#   tabling run:
#####3
#   rotate , shift coordinates and produce table:
#   xyzout is the rotated and shifted coordinates.
#
amore \
xyzin1 model.pdb \
xyzout1 modell.pdb \
TABLE1 search.tab \
<<'END'

  TITLE :   Produce table for MODEL
  VERBOSE
  TABFUN
  CRYSTAL ORTH 1
  CELL  65 80 60 90 90 90
  MODEL 1 BREPLACE 0 BADD 0
  SAMPLE 1 TMIN 3 SHANN 2 SCALE      4.0

'END'
```

coeficiente de correlação. Maiores informações no manual do CCP4 (CCP4, 1994).

Agora vamos rodar o programa. Para isto temos que digitar.

```
chmod 777 *.sh
```

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

```
./tabling.sh >tabling.log
```

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

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

O arquivo PDB gerado traz as coordenadas atômicas (*modell.pdb*) e o *search.tab* os fatores de estrutura calculados.

Olharemos o conteúdo do segundo arquivo de comandos (*sorting.sh*). Nesta parte iremos ler o arquivo com os F_{obs} (*data.mtz*) e gerar o arquivo *data.hkl*. Este arquivo mantém as reflexões ordenadas e dentro de uma faixa de resolução indicada, no caso entre 100 e 3 Å. Na substituição molecular normalmente não precisamos de dados de altíssima resolução.

Vamos verificar o conteúdo deste arquivo com o comando *more*. Digite:

```
more sorting.sh
```

Será mostrado na tela.

```
#####3
#   sorting run:
#####3
#   mtz file contains cell and symmetry.
#
#
#   amore \
#   hklin data.mtz \
#   hklpck0 data.hkl \
#   <<'END'
#   VERBOSE
#   TITLE   ** packing h k l F for crystal**

#   SORTFUN RESOL 100. 3.
#   LABI FP=F_New SIGFP=SIGF_New
#   'END'
```

As 6 primeiras linhas são comentários. A sétima linha aciona o programa *AMoRe*, a oitava linha indica que as reflexões estão no arquivo *data.mtz*, a nona linha indica que a saída será o arquivo *data.hkl*. As outras linhas são informações sobre a faixa de resolução e os rótulos das colunas de dados. Rodaremos o segundo passo, com o seguinte comando:

```
./sorting.sh > sorting.log
```

Até agora não começamos a substituição molecular propriamente dita, só preparamos os arquivos para o início do processo. Na primeira parte da substituição molecular rodaremos as coordenadas atômicas do arquivo *model1.pdb*, de forma que possamos encontrar a orientação relativa do *model1.pdb* dentro da cela unitária da estrutura a ser resolvida, para isto usaremos o arquivo de comandos *rotating.sh*. Vamos verificar o conteúdo deste arquivo com o comando *more*. Digite: *more rotating.sh*

Será mostrado na tela.

```
#####3
#   rotating run:
#####3
#
#   Rotating model1.pdb
#
#   amore \
#   TABLE1 search.tab \
#   HKLPCK1 search.hkl \
#   hklpck0 data.hkl \
#   clmn1 search.clmn \
#   clmn0 data.clmn \
#   MAPOUT amore.map \
#   <<'END'
#
#   ROTFUN
#   VERB
#   TITLE : Rotation Function
#   GENE 1 RESO 8.0 4.0
#   CELL 50 50 50 90 90 90
#   CLMN CRYSTAL ORTH 1 FLIM 0.E0 1.E8 SHARP 0.0 RESO 8.0 4.0 -
#   SPHERE 0.0 15
#   CLMN MODEL 1 FLIM 0.E0 1.E8 RESO 8.0 4.0 SPHERE 0.0 15
#   ROTA CROSS MODEL 1 BESLIMI 6 120 STEP 2.5 NSR 0 0 -
#   PKLIM 0.5 NPIC 50 BMAX 90
#   'END'
#   exit
```

As 6 primeiras linhas são comentários. A sétima linha aciona o programa *AMoRe*, a oitava linha indica que os fatores de estrutura calculados estão no arquivo *search.tab*, serão gerados vários arquivos auxiliares indicados nas linhas seguintes. A linha 16 indica para o *AMoRe* que será feita a busca de rotação (ROTFUN), ou seja, o cálculo da função de rotação. A linha 19 indica a faixa de resolução usada, veja que não sabemos qual faixa irá funcionar melhor, assim podemos testar diferentes faixas de resolução, a indicada está entre 8.0 4.0 Å. As outras linhas são informações específicas sobre a rotação, e não serão discutidas aqui, para maiores informações consulte o site do programa (<http://www.ccp4.ac.uk/>).

Rodaremos o terceiro passo, com o seguinte comando:

```
./rotimg.sh > rotimg.log
```

Vamos olhar os resultados da procura de rotação. Digite:

```
tail -40 rotimg.log
```

Este comando mostra as 40 últimas linhas do arquivo rotimg.log, teremos o seguinte

```
All Solutions separated by more than 2.50 degrees output
```

CROSS:: Rotation Function												
	ITAB	ALPHA	BETA	GAMMA	TX	TY	TZ	CC_F	RF_F	CC_I	CC_P	Icp
SOLUTIONRC	1	54.22	51.61	158.04	0.0000	0.0000	0.0000	30.7	0.0	0.0	30.7	1
SOLUTIONRC	1	234.22	51.61	158.04	0.0000	0.0000	0.0000	30.7	0.0	0.0	30.7	2
SOLUTIONRC	1	131.25	84.46	353.32	0.0000	0.0000	0.0000	23.6	0.0	0.0	23.6	3
SOLUTIONRC	1	311.25	84.46	353.32	0.0000	0.0000	0.0000	23.6	0.0	0.0	23.6	4
SOLUTIONRC	1	48.50	90.00	168.50	0.0000	0.0000	0.0000	21.0	0.0	0.0	21.0	5
SOLUTIONRC	1	228.50	90.00	168.50	0.0000	0.0000	0.0000	21.0	0.0	0.0	21.0	6
SOLUTIONRC	1	226.00	86.51	161.00	0.0000	0.0000	0.0000	21.0	0.0	0.0	21.0	7
SOLUTIONRC	1	46.00	86.51	161.00	0.0000	0.0000	0.0000	21.0	0.0	0.0	21.0	8
SOLUTIONRC	1	138.71	70.50	61.66	0.0000	0.0000	0.0000	20.1	0.0	0.0	20.1	9
SOLUTIONRC	1	318.71	70.50	61.66	0.0000	0.0000	0.0000	20.1	0.0	0.0	20.1	10
SOLUTIONRC	1	81.50	19.00	131.00	0.0000	0.0000	0.0000	18.0	0.0	0.0	18.0	11
SOLUTIONRC	1	261.50	19.00	131.00	0.0000	0.0000	0.0000	18.0	0.0	0.0	18.0	12
SOLUTIONRC	1	132.99	44.79	356.91	0.0000	0.0000	0.0000	17.8	0.0	0.0	17.8	13
SOLUTIONRC	1	312.99	44.79	356.91	0.0000	0.0000	0.0000	17.8	0.0	0.0	17.8	14
SOLUTIONRC	1	288.12	66.10	73.42	0.0000	0.0000	0.0000	17.6	0.0	0.0	17.6	15
SOLUTIONRC	1	108.12	66.10	73.42	0.0000	0.0000	0.0000	17.6	0.0	0.0	17.6	16
SOLUTIONRC	1	3.76	17.13	121.22	0.0000	0.0000	0.0000	16.9	0.0	0.0	16.9	17
SOLUTIONRC	1	183.76	17.13	121.22	0.0000	0.0000	0.0000	16.9	0.0	0.0	16.9	18
SOLUTIONRC	1	122.53	83.51	25.84	0.0000	0.0000	0.0000	16.8	0.0	0.0	16.8	19
SOLUTIONRC	1	302.53	83.51	25.84	0.0000	0.0000	0.0000	16.8	0.0	0.0	16.8	20
SOLUTIONRC	1	253.99	52.36	127.19	0.0000	0.0000	0.0000	16.7	0.0	0.0	16.7	21
SOLUTIONRC	1	73.99	52.36	127.19	0.0000	0.0000	0.0000	16.7	0.0	0.0	16.7	22
SOLUTIONRC	1	124.73	73.55	321.87	0.0000	0.0000	0.0000	16.2	0.0	0.0	16.2	23
SOLUTIONRC	1	304.73	73.55	321.87	0.0000	0.0000	0.0000	16.2	0.0	0.0	16.2	24
SOLUTIONRC	1	246.12	80.96	41.15	0.0000	0.0000	0.0000	16.1	0.0	0.0	16.1	25
SOLUTIONRC	1	66.12	80.96	41.15	0.0000	0.0000	0.0000	16.1	0.0	0.0	16.1	26
SOLUTIONRC	1	296.00	19.00	284.00	0.0000	0.0000	0.0000	15.6	0.0	0.0	15.6	27
SOLUTIONRC	1	116.00	19.00	284.00	0.0000	0.0000	0.0000	15.6	0.0	0.0	15.6	28
SOLUTIONRC	1	124.00	14.63	276.00	0.0000	0.0000	0.0000	15.5	0.0	0.0	15.5	29
SOLUTIONRC	1	304.00	14.63	276.00	0.0000	0.0000	0.0000	15.5	0.0	0.0	15.5	30
SOLUTIONRC	1	272.48	61.22	97.64	0.0000	0.0000	0.0000	15.5	0.0	0.0	15.5	31
SOLUTIONRC	1	92.48	61.22	97.64	0.0000	0.0000	0.0000	15.5	0.0	0.0	15.5	32

As soluções são indicadas com o início SOLUTIONRC, identifique as informações da melhor solução e anote os valores.

Alfa =
Beta =
Gama =
CC_F =

As informações angulares serão usadas na próxima etapa, a procura de translação. Para rodar a procura de translação usaremos o arquivo de comandos *traing.sh*. Vamos verificar o conteúdo deste arquivo com o comando *more*. Digite:

```
more traing.sh
```

Será mostrado na tela.

```
#####3
#   traing run:   NMOL = 1
#####3
#
# Translating modell.pdb
#
amore \
TABLE1 search.tab \
hklpck0 data.hkl \
MAPOUT amore.map \
<<'END'
TRAFUN CB   NMOL 1 RESO 8 4      -
  PKLIM 0.5  NPIC 100
VERB
  TITLE : Translation function
CRYSTAL ORTH 1 FLIMI 0.E0 1.E8   SHARP 0.0
SOLUTIONRC  1  54.22  51.61 158.04  0.0000  0.0000  0.0000 30.7  0.0  0.0 30.7  1
'END'

exit
```

As 6 primeiras linhas são comentários. A sétima linha aciona o programa *AMoRe*, a oitava linha indica que os fatores de estrutura calculados estão no arquivo *search.tab*, serão gerados vários arquivos auxiliares indicados nas linhas seguintes. A linha 12 indica para o *AMoRe* que será feita a busca de translação (TRAFUN), ou seja, o cálculo da função de translação. A mesma linha 12 indica também a faixa de resolução usada, a indicada está entre 8.0 4.0 Å. A linha 17 indica as informações obtidas da rotação, ou seja, os ângulos que geraram a melhor solução na busca de rotação. As outras linhas são informações específicas sobre a translação, e não serão discutidas aqui. Normalmente você teria que digitar estes valores, mas aqui o arquivo já está pronto.

Rodaremos o quarto passo, com o seguinte comando:

```
./traing.sh > traing.log
```

Vamos olhar os resultados da procura de translação. Digite:

```
tail -20 traing.log
```


Este comando mostra as 20 últimas linhas do arquivo traing.log, teremos o seguinte

```

Translation solutions sorted by correlation coefficients
      TABLE  Alpha    Beta    Gamma    Tx      Ty      Tz  Corr_F  Rfac  Corr_I  Pkcount  Dmin
SOLUTIONTF1 1  54.22  51.61  158.04  0.0441  0.3436  0.1937  74.8  31.0  77.4    1  22.1
SOLUT_1     1  54.22  51.61  158.04  0.0441  0.3436  0.1937  74.8  31.0  77.4    1  22.1
SOLUT_1     1  54.22  51.61  158.04  0.0442  0.0579  0.1957  45.9  44.6  50.3    2  10.4
SOLUT_1     1  54.22  51.61  158.04  0.0459  0.3800  0.1959  43.1  44.8  49.0    3  19.0
SOLUT_1     1  54.22  51.61  158.04  0.0438  0.2738  0.1974  42.6  44.7  44.2    4  22.5
SOLUT_1     1  54.22  51.61  158.04  0.0444  0.9925  0.1950  41.9  45.1  44.8    5  5.8
SOLUT_1     1  54.22  51.61  158.04  0.0445  0.4233  0.1929  41.3  46.0  47.6    6  12.9
SOLUT_1     1  54.22  51.61  158.04  0.0453  0.4581  0.1959  39.4  46.1  44.0    7  8.6
SOLUT_1     1  54.22  51.61  158.04  0.0436  0.1356  0.1919  39.3  45.7  43.2    8  21.2
SOLUT_1     1  54.22  51.61  158.04  0.2939  0.0900  0.0649  38.3  45.6  41.1    9  15.3
SOLUT_1     1  54.22  51.61  158.04  0.0872  0.4382  0.3829  34.2  48.6  38.3   10  14.5
SOLUT_1     1  54.22  51.61  158.04  0.0861  0.4783  0.3733  32.6  48.9  34.0   11  11.5
Normal Termination of TRAFUN
<B><FONT COLOR="#FF0000"><!--SUMMARY_BEGIN-->
AMORE: Normal termination
Times: User:      0.0s System:      0.0s Elapsed:      0:00
</pre>
</html>
<!--SUMMARY_END--></FONT></B>

```

As soluções são indicadas com o início SOLUTIONTF1, identifique as informações.

Alfa =

Beta =

Gama =

Tx =

Ty =

Tz =

Corr_F =

Rfac =

Veja que agora temos valores para a translação (Tx, Ty e Tz), note também que aumentou significativamente o valor do coeficiente de correlação (CC_F). Isto indica que a solução transladada gera fatores de estrutura calculados mais próximos dos fatores de estrutura observados, como era de se esperar no caso de se encontrar uma solução.

O próximo passo é um ajuste fino na solução encontrada, realizaremos um refinamento de corpo rígido, onde a solução encontrada na translação passa por um ajuste, onde podemos ter um aumento do coeficiente de correlação. As informações angulares e de translação serão usadas na próxima etapa, o refinamento de corpo rígido (ajuste fino). Para rodar o ajuste fino usaremos o arquivo de comandos *fiting.sh*. Vamos verificar o conteúdo deste arquivo com o comando *more*. Digite:

`more fitting.sh`

Será mostrado na tela.

```

#####3
#   fitting run:
#####3
#
# Rigid-body refinement
#
amore \
TABLE1 search.tab \
hklpck0 data.hkl \
<<'END'
FITFUN CB NMOL 1 RESO 8 4.0 ITER 10 CONV 1.E-3
title
VERBOSE
CRYSTAL ORTH 1 FLIMI 0.E0 1.E8 SHARP 0.0
REFSOL AL BE GA X Y Z BF
SOLUTION 1 54.22 51.61 158.04 0.0441 0.3436 0.1937 74.8 31.0
'END'
exit

```

As 6 primeiras linhas são comentários. A sétima linha aciona o programa *AMoRe*, a oitava linha indica que os fatores de estrutura calculados estão no arquivo *search.tab*, serão gerados vários arquivos auxiliares indicados nas linhas seguintes. A linha 11 indica para o *AMoRe* que será feito o refinamento de corpo rígido (FITFUN). A mesma linha 11 indica também a faixa de resolução usada, a indicada está entre 8.0 4.0 Å, e o número de moléculas (NMOL 1). A linha 16 indica as informações obtidas da rotação e translação.

Rodaremos o quinto passo, com o seguinte comando:

```
./fiting.sh > fitting.log
```

Vamos olhar os resultados do ajuste fino. Digite:

```
tail -20 fitting.log
```

Este comando mostra as 20 últimas linhas do arquivo *fiting.log*, teremos o seguinte

```
Fitting solutions
      TABLE   Alpha   Beta   Gamma   Tx      Ty      Tz   Corr_F   Rfac   Corr_I   Pkcount
SOLUTIONF    1    54.60   52.34  156.87  0.0436  0.3440  0.1957  79.9  26.8  82.7    1
PARAMETERS OF FIT
B-FACTOR      :           1
ANGLE ALFA    :           1
ANGLE BETA    :           1
ANGLE GAMA    :           1
TRANSLATION X:           1
TRANSLATION Y:           1
TRANSLATION Z:           1

INITIAL POSITIONS
Normal Termination of FITFUN
<B><FONT COLOR="#FF0000"><!--SUMMARY_BEGIN-->
AMORE: Normal termination
Times: User:      0.1s System:      0.0s Elapsed:      0:01
</pre>
</html>
<!--SUMMARY_END--></FONT></B>
```

A solução é indicada com o início SOLUTIONTF, identifique as informações da solução e anote os valores.

Alfa =

Beta =

Gama =

Tx =

Ty =

Tz =

Corr_F =

Rfac =

Houve melhora nos resultados? Por quê?

A substituição molecular já está pronta, contudo o programa *AMoRe* não gera as coordenadas atômicas da melhor solução, precisamos de um passo extra com o programa LSQKAB (CCP4, 1994), que usamos anteriormente, para gerarmos as coordenadas atômicas da solução, chamado aqui de arquivo *model2.pdb*.

Usaremos o arquivo de comandos *lsqkab.sh*. Veja o conteúdo deste arquivo com o comando *more*, como segue:

```
more lsqkab.sh
```

Será mostrado na tela.

```
set -e
lsqkab
XYZINM model1.pdb \
XYZOUT model2.pdb \
<< 'END-lsqkab'
title rotation and translation from MR
ROTATE EULER 54.60 52.34 156.87
TRANSLATE FRAC 0.0436 0.3440 0.1957
output XYZ
end
END-lsqkab
```

A segunda linha chama o programa LSQKAB, a terceira linha indica que a entrada é o arquivo *model1.pdb*, a quarta linha indica que a saída é o arquivo *model2.pdb*. As linhas 7 e 8 trazem as informações sobre a rotação e translação, respectivamente.

Visualização do resultado

Clique no ícone VMD (Humphrey *et al.*, 1996). Para carregar as coordenadas atômicas do arquivo *model2.pdb*, proceda da seguinte forma: No *VMD Main* clique em *File* e *New Molecule*. No menu *New Molecule* clique em *Browse* e selecione o arquivo *model2.pdb* da pasta *pratica4*. Clique em *Abrir* e *Load*. Feche o Menu *Molecule File Browser*. Clique na tela gráfica. Você terá a estrutura (*model2.pdb*) 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. Para visualizar a estrutura secundária da proteína execute os seguintes comandos. Clique no menu *Graphical Representations*, na opção *Drawing Method* escolha a opção *NewCartoon*. Se você realizou tudo certo terá na tela gráfica a visão da estrutura parcial da citidina deaminase, como mostrado na figura 2.

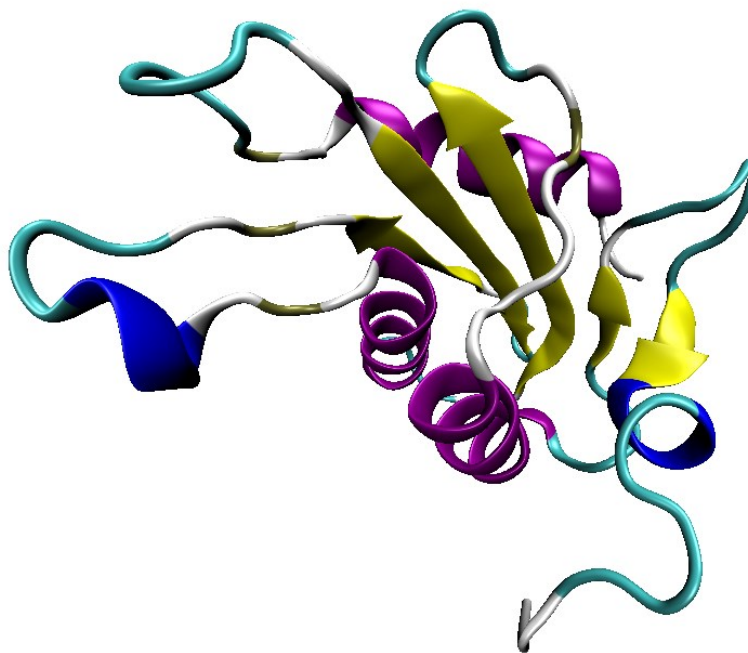


Figura 2. Estrutura tridimensional da citidina deaminase

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.

Navaza, J. *Acta Cryst.* **A50**, 157-163, 1994.

Rossmann, M. G. and Blow, D. M. *Acta Cryst.* **15**, 24-32, 1962.

Sánchez-Quitian ZA, Schneider CZ, Ducati RG, de Azevedo WF, Junior CB, Basso LA, Santos DS. *J Struct Biol* 2010; 169(3): 413-423.

<http://www.ccp4.ac.uk/index.php>

<http://www.ks.uiuc.edu/Research/vmd/>

<http://www.fi.muni.cz/usr/mejzlik/mirrors/www.nyu.edu/pages/mathmol/software.html>

<http://www.rcsb.org/pdb/>