

Rotação e Translação de Estruturas de Proteínas

Objetivos:

Familiarizar-se com a aplicação de operações de simetria em estruturas de proteínas. Gerar estruturas de proteínas a partir da aplicação de operadores de simetria. Aprender o uso de recursos do pacote de programas CCP4 (CCP4, 1994). Visualizar a estrutura tridimensional de proteínas.

Materiais:

1. Computador iMac;
2. Pacote CCP4 (CCP4, 1994);
3. Programa VMD (*Visual Molecular Dynamics*) para visualização de macromoléculas biológicas;
4. Arquivos com coordenadas atômicas de aminoácidos e proteínas.

Introdução

Os operadores de simetria podem ser expressos como matrizes e vetores. Aqui faremos uso de matrizes e vetores para gerar estruturas oligoméricas de proteínas. Há diversos programas disponíveis para manipulação de coordenadas atômicas de proteínas e para a resolução da estrutura cristalográfica. Um dos mais utilizados é o pacote de programas CCP4 (CCP4, 1994). O pacote de programas CCP4 é uma suíte de softwares, colocados na forma de um pacote para facilitar a sua distribuição e instalação. Usaremos um dos programas do CCP4, o LSQKAB (Kabsch, 1976). O LSQKAB é usado para sobreposição de estruturas de proteínas e para aplicação de operações de simetria em estruturas de proteínas. O CCP4 é gratuito e há versões para Windows, Linux, Mac OS X e outros sistemas operacionais.

Representação matricial de operadores de simetria

A notação matricial é fundamental para representar operações sobre as coordenadas atômicas de estruturas de proteínas. As rotações podem ser representadas por matrizes 3x3, ou seja, com 3 linhas e 3 colunas. Uma matriz rotação tem a seguinte representação:

$$\begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix}$$

Cada elemento da matriz é representado de forma geral por a_{ij} , o “i” indica a linha e “j” a coluna, assim o elemento a_{23} está na segunda linha e terceira coluna.

Podemos representar uma rotação em torno de um eixo pela as seguintes matrizes de rotação.

Abaixo temos a matriz para rotação de um ângulo θ , ao longo do eixo x:

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos \theta & -\sin \theta \\ 0 & \sin \theta & \cos \theta \end{pmatrix}$$

Abaixo temos a matriz para rotação de um ângulo θ , ao longo do eixo y:

$$\begin{pmatrix} \cos \theta & 0 & \sin \theta \\ 0 & 1 & 0 \\ -\sin \theta & 0 & \cos \theta \end{pmatrix}$$

Abaixo temos a matriz para rotação de um ângulo θ , ao longo do eixo z:

$$\begin{pmatrix} \cos \theta & -\operatorname{sen} \theta & 0 \\ \operatorname{sen} \theta & \cos \theta & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

As estruturas de proteínas podem ser armazenadas em arquivos no formato PDB (*Protein Data Bank*) (www.rcsb.org) (Berman, 2008). No formato PDB, a posição de cada átomo é representada por um vetor, que chamaremos de \mathbf{x} . Tal representação facilita a manipulação das coordenadas atômicas da proteína. Assim, operações como a rotação (R) podem ser representadas da seguinte forma:

$$\mathbf{x}' = R \cdot \mathbf{x}$$

A matriz R opera sobre o vetor \mathbf{x} , gerando um novo vetor \mathbf{x}' .

Veja um exemplo. Vamos pegar as coordenadas atômicas do primeiro átomo de um arquivo PDB. Escolhemos as coordenadas atômicas da estrutura da purina nucleosídeo fosforilase (PNP) humana (EC 2.4.2.1) (código de acesso PDB: 1M73)(De Azevedo Jr. *et al.*, 2003a) e aplicamos uma matriz de rotação de 90° ao longo do eixo z. Ao abrirmos o arquivo de coordenadas atômicas da PNP (*1M73.pdb*), vemos que o primeiro átomo é

ATOM	1	N	GLU	E	2	63.717	29.804	42.950	1.00	39.58
------	---	---	-----	---	---	---------------	---------------	---------------	------	-------

As coordenadas atômicas são indicadas em negrito. As coordenadas acima indicam o vetor posição do átomo nitrogênio (N), do resíduo de aminoácido GLU (glutamato). Como operamos com a aplicação da matriz rotação?

$$\begin{pmatrix} x' \\ y' \\ z' \end{pmatrix} = \begin{pmatrix} \cos \theta & -\operatorname{sen} \theta & 0 \\ \operatorname{sen} \theta & \cos \theta & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix}$$

Substituímos os valores, como indicado abaixo:

$$\begin{pmatrix} x' \\ y' \\ z' \end{pmatrix} = \begin{pmatrix} \cos 90^\circ & -\operatorname{sen} 90^\circ & 0 \\ \operatorname{sen} 90^\circ & \cos 90^\circ & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 63.717 \\ 29.804 \\ 42.950 \end{pmatrix}$$

$$\begin{pmatrix} x' \\ y' \\ z' \end{pmatrix} = \begin{pmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 63.717 \\ 29.804 \\ 42.950 \end{pmatrix}$$

Temos que multiplicar linha por coluna, assim temos:

$$\begin{pmatrix} x' \\ y' \\ z' \end{pmatrix} = \begin{pmatrix} -29.804 \\ 63.717 \\ 42.950 \end{pmatrix}$$

Exercício (para entregar na aula seguinte): Determine as coordenadas atômicas do vetor resultante da aplicação da matriz de rotação de 120° ao longo do eixo z sobre o vetor posição do átomo de N do resíduo de aminoácidos GLU 2, mostrado anteriormente. **ii)** Aplique novamente a matriz rotação de 120° ao longo do eixo z para as coordenadas resultantes (x' , y' , z') do exercício anterior. **iii)** Repita o procedimento para as coordenadas atômicas obtidas no item ii).

A notação matricial permite a combinação de rotação e translação, sendo a translação representada por um vetor translação (\mathbf{t}), que tem suas coordenadas somadas ao vetor resultante da aplicação da rotação. Representamos, assim, uma matriz de rotação e a soma do vetor translação, como indicado abaixo:

$$\begin{pmatrix} x' \\ y' \\ z' \end{pmatrix} = \begin{pmatrix} \cos \theta & -\sin \theta & 0 \\ \sin \theta & \cos \theta & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} + \begin{pmatrix} t_x \\ t_y \\ t_z \end{pmatrix}$$

onde t_x , t_y e t_z indicam a translação ao longo dos eixos x, y e z, respectivamente. Podemos representar a operação matemática de forma reduzida, como segue:

$$\mathbf{x}' = \mathbf{R} \cdot \mathbf{x} + \mathbf{t}$$

Representação matricial de operadores de simetria em arquivos PDB

Praticamente toda informação estrutural sobre proteínas disponível está depositada na base de dados PDB. A base de dados PDB foi estabelecida em 1977 e tornada pública com o desenvolvimento da web. Uma preocupação do grupo de desenvolvedores do PDB é a interconexão com diferentes bases de dados, de forma que a informação biológica, disponível em outras bases de dados, possa ser facilmente acessada, a partir da estrutura da proteína depositada no PDB. Apesar do grande desenvolvimento do PDB, o formato tem se mantido estável nos últimos 40 anos. O PDB é um arquivo texto que traz nas primeiras linhas informações sobre a proteína depositada, bem como sobre os autores. Nós estamos interessados nas informações sobre a simetria. Há dois “remarks” reservados para guardar a informação sobre simetria. Um deles é chamado “REMARK 290”, nas linhas do arquivo PDB com este *remark* temos a matriz de rotação e o vetor de translação presentes na simetria da cela unitária do cristal. Abram o arquivo PDB da PNP humana (código de acesso: 1ULA) (Ealick *et al.*, 1991) e localizem o “REMARK 290”. Identifiquem as matrizes e os vetores de cada operador de simetria.

Outra informação disponível está relacionada à estrutura quaternária da proteína. Muitas proteínas se apresentam como oligômeros, que podem ter sua estrutura oligomérica gerada a partir da aplicação de operadores de simetria. A PNP humana tem estrutura oligomérica. A estrutura depositada é um monômero, mas sabe-se que sua estrutura em solução é um trímero (De Azevedo *et al.*, 2003b). O trímero é o resultado da aplicação de operadores de simetria, especificamente do operador eixo de rotação ordem 3, ou simplesmente eixo de ordem 3.

Para visualizar as informações sobre as matrizes e vetores, abra o arquivo *1ULA.pdb* com um editor de texto. No arquivo PDB 1ULA, procurem o “REMARK 350”. Identifiquem as matrizes e vetores translação. Há 3 matrizes de rotação e 3 vetores de translação. Escreva no espaço a seguir cada um deles.

Matriz 1

$$\begin{pmatrix} \underline{\hspace{2cm}} & \underline{\hspace{2cm}} & 0 \\ \underline{\hspace{2cm}} & \underline{\hspace{2cm}} & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

Vetor 1

$$\begin{pmatrix} \underline{\hspace{2cm}} \\ \underline{\hspace{2cm}} \\ \underline{\hspace{2cm}} \end{pmatrix}$$

Matriz 2

$$\begin{pmatrix} \underline{\hspace{2cm}} & \underline{\hspace{2cm}} & 0 \\ \underline{\hspace{2cm}} & \underline{\hspace{2cm}} & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

Vetor 2

$$\begin{pmatrix} \underline{\hspace{2cm}} \\ \underline{\hspace{2cm}} \\ \underline{\hspace{2cm}} \end{pmatrix}$$

Matriz 3

$$\begin{pmatrix} \underline{\hspace{2cm}} & \underline{\hspace{2cm}} & 0 \\ \underline{\hspace{2cm}} & \underline{\hspace{2cm}} & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

Vetor 3

$$\begin{pmatrix} \underline{\hspace{2cm}} \\ \underline{\hspace{2cm}} \\ \underline{\hspace{2cm}} \end{pmatrix}$$

A primeira matriz é a que chamamos de matriz identidade, a aplicação matriz 1 não muda as coordenadas atômicas, o vetor translação correspondente é nulo, também não muda as coordenadas atômicas. A matriz identidade é mantida para indicar que para gerar o trímero precisamos manter as coordenadas atômicas do arquivo, para a posição do primeiro monômero. A segunda matriz é uma matriz de rotação de 120° , ao longo do eixo z, é fácil conferir. Calcule a matriz de rotação a partir da equação abaixo, com $\theta = 120^\circ$, e compare com a matriz 2 do arquivo PDB.

$$\begin{pmatrix} \cos \theta & -\operatorname{sen} \theta & 0 \\ \operatorname{sen} \theta & \cos \theta & 0 \\ 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} \cos 120^\circ & -\operatorname{sen} 120^\circ & 0 \\ \operatorname{sen} 120^\circ & \cos 120^\circ & 0 \\ 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} \underline{\hspace{2cm}} & \underline{\hspace{2cm}} & 0 \\ \underline{\hspace{2cm}} & \underline{\hspace{2cm}} & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

O vetor translação é zero.

A terceira matriz é a matriz de rotação de 240° , calcule e compare.

$$\begin{pmatrix} \cos \theta & -\operatorname{sen} \theta & 0 \\ \operatorname{sen} \theta & \cos \theta & 0 \\ 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} \cos 240^\circ & -\operatorname{sen} 240^\circ & 0 \\ \operatorname{sen} 240^\circ & \cos 240^\circ & 0 \\ 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} \underline{\hspace{2cm}} & \underline{\hspace{2cm}} & 0 \\ \underline{\hspace{2cm}} & \underline{\hspace{2cm}} & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

Usaremos as informações armazenadas nas matrizes 2 e 3 para gerarmos o trímero da PNP humana. Usaremos o programa LSQKAB do pacote CCP4. Na versão disponível nos computadores, usaremos linhas de comando, para isto é necessário abrirmos um terminal. Veja na parte de baixo da tela que há um ícone chamado "Terminal", se tiver clique duas vezes nele. Caso contrário, olhe na tela à esquerda na parte de baixo, lá você encontra o ícone do "finder" clique com o mouse sobre o ícone. 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" ou "core" do sistema operacional "OS X", que funciona como um terminal de comandos do sistema operacional Linux. Trabalharemos na pasta *Bioestrutural*. Achem a pasta. Para termos acesso aos programas do CCP4 temos que ativá-lo, para isto digitem:

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

e pressionem *Enter/Return*.

Agora todos os programas do CCP4 estão ativos.

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

```
pwd
```

O comando *pwd* mostra em que pasta você se encontra. Mudem para a pasta onde estão os arquivos que você usará. Digitem:

```
cd ~/Disciplinas/Bioestrutural/Aula2/lula
```

Agora confirmem em que pasta você se encontra. Na pasta temos diversos arquivos. Digitem o comando para verificar o conteúdo da pasta.

```
ls -l
```

Na tela vocês terão todos os arquivos disponíveis na pasta.

Nós vamos gerar o trímico da PNP humana a partir das coordenadas atômicas do monômero. Na pasta temos as coordenadas atômicas da PNP, no arquivo *1ULA.pdb*. Temos, também, um arquivo de comandos, chamado *lsqkab_rot120.sh*. O arquivo de comandos tem um conjunto de instruções para chamar o programa LSQKAB, ler as coordenadas atômicas do arquivo *1ULA.pdb*, aplicar a matriz de rotação de 120° (eixo de rotação de ordem 3), e gerar um arquivo com a estrutura resultante da aplicação da matriz de rotação, arquivo *mono_b.pdb*. Vamos verificar o conteúdo do arquivo com o comando *more*. Digitem:

```
more lsqkab_rot120.sh
```

Será mostrado na tela.

```
#!/bin/sh
set -e
lsqkab \
XYZINM 1ULA.pdb \
XYZOUT mono_b.pdb \
<< 'END-lsqkab'
title rotation axis along z (120 degrees)
ROTATE MATRIX -0.500 -0.866 0.000 0.866 -0.500 0.000 0.000 0.000 1.000
output XYZ
end
END-lsqkab
```

A primeira linha é de comentário. A segunda é um comando Linux, para que o programa funcione corretamente. A terceira linha invoca o programa “lsqkab”, a partir desta linha todas as seguintes trazem linhas de comandos específicas do programa LSQKAB. A quarta linha indica o nome do arquivo de entrada (*1ULA.pdb*), sobre o qual será aplicada a matriz de rotação, a quinta linha, indica o nome do arquivo de saída (*mono_b.pdb*), a sexta linha indica que deve ir para o final após a execução, a sétima linha é o título desta operação. Na oitava linha temos a matriz de rotação, na seguinte ordem:

$$\begin{matrix} a_{11} & a_{12} & a_{13} & a_{21} & a_{22} & a_{23} & a_{31} & a_{32} & a_{33} \end{matrix}$$

Vejam a linha da matriz e confirmem a matriz de rotação de 120°.

A nona linha indica que o programa deve escrever as coordenadas atômicas no arquivo de saída (*mono_b.pdb*). As duas últimas linhas indicam para o programa encerrar de forma apropriada.

Agora vamos rodar o programa. Digitem a linha de comando abaixo.

```
chmod 777 lsqkab_rot120.sh
```

O comando do Linux garante que o arquivo *lsqkab_rot120.sh* possa ser executado. Depois digitem:

```
./lsqkab_rot120.sh > lsqkab_rot120.log
```

A primeira parte da linha de comando indica que o arquivo *lsqkab_rot120.sh* deve ser executado, o que acionará o programa LSQKAB, o símbolo ">" indica que o resultado do programa, que não seja o arquivo de saída (*mono_b.pdb*), ficará registrado no arquivo *lsqkab_rot120.log*. O arquivo log armazena as ações realizadas pelo programa LSQKAB.

Digitem o comando *ls -l* para verificar se todos os arquivos foram criados.

O arquivo PDB gerado traz as coordenadas atômicas do segundo monômero. Repetiremos o processo para gerar as coordenadas atômicas do terceiro monômero.

```
chmod 777 lsqkab_rot240.sh
```

```
./lsqkab_rot240.sh > lsqkab_rot240.log
```

Agora temos as coordenadas atômicas dos três monômeros (*1ULA.pdb*, *mono_b.pdb* e *mono_c.pdb*). Para visualizar a estrutura, usaremos o programa *Visual Molecular Dynamics* (VMD).

Visualização de estruturas de proteínas

Cliquem no ícone VMD (Humphrey *et al.*, 1996). Para carregar as coordenadas atômicas do arquivo *1ULA.pdb*, procedam da seguinte forma: No *VMD Main* cliquem em *File* e *New Molecule*. No menu *New Molecule* cliquem em *Browse* e selecionem o arquivo *1ULA.pdb* da pasta *Aula2/1ula*. Cliquem em *Abrir* e *Load*. Fechem o Menu *Molecule File Browser*. Cliquem na tela gráfica. Teremos o monômero (1ULA) 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. Girem o monômero com o botão do mouse da esquerda. Para visualizar a estrutura secundária da proteína, executem os seguintes comandos. Cliquem no menu *Graphical Representations*, na opção *Drawing Method* escolha a opção *NewCartoon*.

Repitam o processo para as coordenadas atômicas *mono_b.pdb* e *mono_c.pdb*.

Se estiver certo, teremos a visão do trímero da PNP, como mostrado na figura 1.

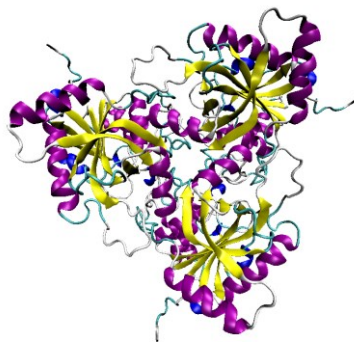


Figura 1. Estrutura tridimensional do trímero da HsPNP (1ULA)

Vamos agora repetir o processo para gerar o trímero da PNP resolvida a mais alta resolução (arquivo *1M73.pdb*). Para mudarmos para a pasta 1m73, digitem:

```
cd ~/disciplinas/Bioestrutural/Aula2/1m73
```

Repitam o processo anterior. Na pasta já temos os arquivos necessários. A diferença é que devido ao empacotamento cristalográfico os operadores necessários para gerar o trímero necessitam de translação. O novo arquivo *lsqkab_rot120.sh* é da seguinte forma:

```
#!/bin/sh
set -e
lsqkab \
XYZINM 1M73.pdb \
XYZOUT mono_b.pdb \
<< 'END-lsqkab'
title rotation axis along z (120 degrees) and translation of 141.63 towards x
ROTATE MATRIX -0.500 -0.866 0.000 0.866 -0.500 0.000 0.000 0.000 1.000
TRANSLATE 141.630 0.000 0.000
output XYZ
end
END-lsqkab
```

A linha de comando adicional é a TRANSLATE, que indica um vetor translação *t*, com as seguintes coordenadas:

$$\begin{pmatrix} 141.630 \\ 0.000 \\ 0.000 \end{pmatrix}$$

O novo arquivo *lsqkab_rot240.sh* é da seguinte forma:

```
#!/bin/sh
set -e
lsqkab \
XYZINM 1M73.pdb \
XYZOUT mono_c.pdb \
<< 'END-lsqkab'
title rotation axis along z (240 degrees) and translation of 70.815 122.655
ROTATE MATRIX -0.500 0.866 0.000 -0.866 -0.500 0.000 0.000 0.000 1.000
TRANSLATE 70.815 122.655 0.000
output XYZ
end
END-lsqkab
```

Indicando um vetor translação *t*, com as seguintes coordenadas:

$$\begin{pmatrix} 70.815 \\ 122.655 \\ 0.000 \end{pmatrix}$$

O programa LSQKAB tem esta flexibilidade de inserção de dados, permitindo rotação e translação.

A lista de comandos para gerar o novo trímero é a seguinte:

```
chmod 777 lsqkab_rot120.sh
```

O comando do Linux garante que o arquivo *lsqkab_rot120.sh* possa ser executado. Depois digitem:

```
./lsqkab_rot120.sh > lsqkab_rot120.log
```

Com isto geramos o segundo monômero.

```
chmod 777 lsqkab_rot240.sh
```

```
./lsqkab_rot240.sh > lsqkab_rot240.log
```

Agora temos o terceiro monômero. Podemos visualizar a estrutura.

Referências:

Berman, H.M. *Acta Crystallogr A.*, 2008, 64(Pt 1): 88-95.

Collaborative Computational Project No. 4. The CCP4 suite: programs for protein crystallography *Acta Crystallogr.*, 1994, D50, 760-3.

De Azevedo Jr., W.F, Canduri, F., dos Santos, D.M., Silva, R.G., de Oliveira, J.S., de Carvalho, L.P., Basso, L.A., Mendes, M.A., Palma, M.S., Santos, D.S. *Biochem. Biophys. Res. Commun.*, 2003a, 308(3): 545-52.

De Azevedo, W.F., Santos, G.C., Santos, D.M., Dos Santos, D.M., Olivieri, J.R., Canduri, F., Silva, R.G., Basso, L.A., Renard, G., Fonseca, I.O., Mendes, M.A., Palma, M.S., Santos, D.S. *Biochem. Biophys. Res. Commun.*, 2003b, 309(4): 923-8.

Ealick, S.E., Babu, Y.S., Bugg, C.E., Erion, M.D., Guida, W.C., Montgomery, J.A., Secrist, J.A, 3rd. *Proc Natl Acad Sci U S A.*, 1991, 88(24): 11540-4.

Humphrey, W., Dalke, A., Schulten, K. *J. Molec. Graphics*, 1996, 14: 33-8.

Kabsch, W. *Acta Cryst.*, 1976, A32: 922-3.

<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/>