

# PyMOL: tutorial introdutório

## Introdução

O software PyMOL é um visualizador de estruturas moleculares, vocacionado principalmente para estrutura de macromoléculas biológicas (proteínas e ácidos nucleicos). Estas estruturas podem ser lidas a partir de ficheiros de estrutura em diversos formatos, sendo o mais comum o formato PDB (Protein Data Bank), habitualmente com extensão .pdb. O PyMOL pode ler uma variedade de outros formatos, incluindo mmCIF, MOL e SDF. Os ficheiros de estrutura de macromoléculas podem ser obtidos a partir do website do Protein Data Bank (<https://www.rcsb.org>), muito embora o PyMOL disponha de comandos capazes de fazer o download directo de estruturas a partir de várias bases de dados (ver mais adiante) .

O PyMOL permite não só visualizar as moléculas em vários modos ("sticks", "cartoon", "spheres", "surface", etc...), como colorir diferentes regiões da estrutura em diferentes cores, associar "labels" de texto a grupos de átomos, fazer medições de distâncias e ângulos, comparação e sobreposição de estruturas e diversas outras tarefas.

## Obtenção de estruturas moleculares

Para poder visualizar as estruturas moleculares, necessitamos de um ficheiro com as coordenadas atómicas e outras informações necessárias para a representação da estrutura. Tais ficheiros podem, por exemplo, ser obtidos a partir do Protein Data Bank (<https://www.rcsb.org>), daqui em diante designado como "PDB". Este site permite a pesquisa de estruturas por uma variedade de critérios: nome da estrutura, a data de deposição no PDB, resolução atómica, organismo fonte, nome dos autores, etc. Cada estrutura individual no PDB possui um código único que pode ser usado para obter essa estrutura, quer descarregando-a do próprio site do PDB, quer dentro do software PyMOL usando o comando de consola `fetch` ou o comando de menu "File → Get PDB." (ver abaixo explicação da diferença entre comandos de "consola" e de "menu").

Se a estrutura tiver sido previamente descarregada para o computador, poderá ser aberta dentro do PyMOL com "File → Open", seleccionando o ficheiro desejado.

## Iniciar o PyMOL

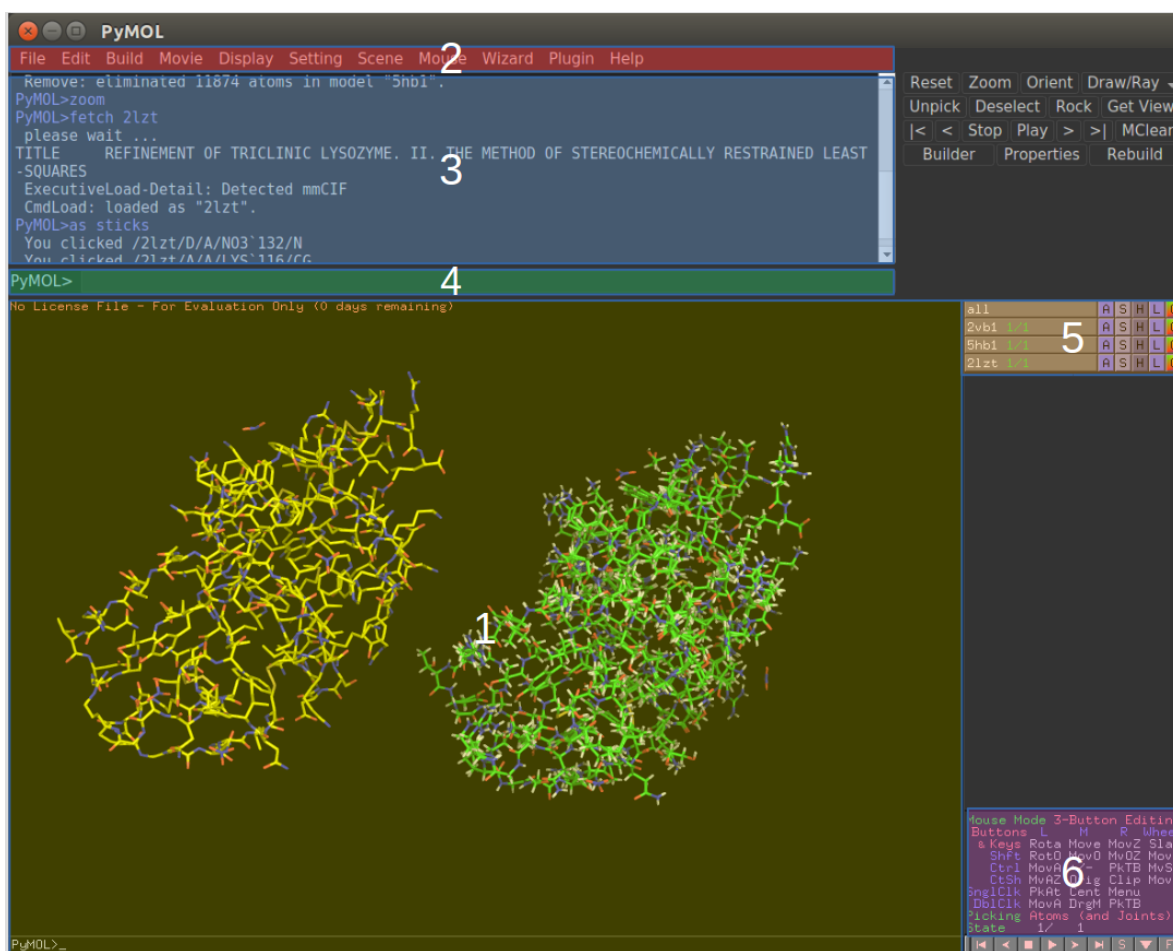
Procure o ícone do programa PyMOL e inicie o programa. Surgirá em primeiro plano uma janela, na qual deverá pressionar o botão "Skip Activation". Na primeira execução do programa, o PyMOL exibirá em seguida uma janela destinada a escolher quais as extensões de ficheiros. Nesta janela deverá fazer "OK" para escolher a associação por defeito (a partir deste momento os ficheiros de extensão ".pdb" será abertos automaticamente pelo PyMOL, quando clicados).

O programa inicia-se normalmente, exibindo a janela representada na figura abaixo.

Esta janela possui várias regiões de funcionalidade distinta:

1. **Área de visualização:** região onde é exibida a estrutura ou estruturas moleculares que se pretende visualizar.

2. **Barra de Menu:** esta barra contém várias categorias: "File", "Edit", "Build", "Movie", "Display", "Setting", "Scene", "Mouse", "Wizard", "Plugin" e "Help". Estas permitem executar uma variedade de comandos, muitos dos quais podem ser também executado através da consola de *input* (ver mais abaixo).
3. **Consola (output):** nesta região o PyMOL escreve o resultado dos comandos executados (quando os comandos o produzem) e exibe uma variedade de mensagens informativos sobre o estado do software.
4. **Consola (input):** esta região permite executar uma grande variedade de comandos (deverão ser terminados com a tecla de ENTER), em alternativa à barra de menu ou ao menu de objectos (ver mais abaixo). Embora a maioria dos comandos mais comuns possam ser executados a partir dos menus, existem algumas situações em que é mais conveniente (ou mesmo obrigatório) executar os comandos através da consola.
5. **Menu de objectos:** esta região contém botões que permitem aplicar comandos à totalidade da cena visualizada (all) ou a cada objecto individualmente. Os botões são **A** (action), **S** (show), **H** (hide), **L** (label) e **C** (color). Deverá existir uma linha para "all" (totalidade dos objectos na cena), bem como linhas separadas para actuar sobre cada objecto (molécula) carregado na aplicação.



## Utilização básica do PyMOL com comandos do rato e o menu de objectos

Vamos começar por descarregar a estrutura da lisozima de galinha, com o código **2vb1**, para uma pasta no nosso computador. Para tal, abrimos a página do portal do PDB, <http://www.rcsb.org>, e na janela de pesquisa inserimos o código "2vb1". O portal exibirá a entrada do PDB para proteína desejada. Para descarregar o ficheiro .pdb, clique no menu "Download Files" (canto superior direito) e escolha opção "PDB Format". Escolha uma pasta no seu computador e descarregue o ficheiro, que deverá ter o nome "2vb1.pdb".

Após iniciarmos o PyMOL como indicado acima, seleccionamos a opção *File* → *Open* na barra de menu (ver Figura 1) e escolhemos o ficheiro "2vbi.pdb" previamente descarregado. A imagem da estrutura da molécula da lisozima de galinha deverá surgir na área de visualização do PyMOL. A estrutura da proteína é, por defeito, representada no modo *cartoon* (representação esquemática da cadeia polipeptídica realçando a localização das hélices alfa e folhas beta, sendo estas últimas indicadas por setas), sendo os iões presentes na estrutura cristalográfica representados como esferas. Os oxigénios das moléculas de água na vizinhança da proteína aparecem representados como pontos (os átomos hidrogénios não são normalmente visíveis nas estruturas cristalográficas do PDB). Podem ver-se, ainda, duas moléculas mais pequenas numa representação em "sticks" (representação consistindo unicamente na visualização do esqueleto de ligações químicas da molécula). Estas moléculas pequenas são, neste caso, coadjuvantes utilizados para facilitar o processo de cristalização da proteína.

Os elementos químicos aparecem coloridos da seguinte forma: verde para carbono, vermelho para oxigénio, azul para azoto, amarelo para enxofre, branco para hidrogénio (são usadas outras cores para os restantes elementos químicos de ocorrência biológica, tais como Ferro, Cobre, Zinco, Sódio, Potássio, etc). Note-se que estas são as cores por defeito, podendo ser alterados por opção do utilizador.

O menu de objectos deverá presentemente ter duas linhas, uma denominada "all" e outra "2vb1" (porque a nossa cena contém apenas um objecto, "2vb1").

Vamos assumir que o seu computador está equipado com um rato com 3 botões (altamente recomendado), com o qual pretendemos orientar, mover e aumentar/diminuir a estrutura presente na área de visualização. Os comandos do PyMOL associados aos controlos do rato são os seguintes:

- **botão direito (click+drag):** roda a visualização em torno do centro do écran
- **botão do meio (click+drag):** desloca a área de visualização lateralmente (*panning*)
- **botão esquerdo (click+drag):** aproxima e afasta o observador da cena (*zooming*)
- **mouse wheel:** controla espessura da área de visualização (*viewing slab*)

(para um rato de dois botões ou *trackpad* de laptop, seleccionar a opção *Mouse* → *2 Button Viewing*).

Após algumas manipulações da estrutura presente no écran, é possível que a mesma acabe descentrada, muito pequena ou mesmo completamente fora da área de visualização. Nesse caso existe um comando que restaura a orientação inicial da cena (em que a palavra "cena" se refere à totalidade de estruturas objectos carregados no PyMOL, neste exemplo ainda apenas uma estrutura). Este comando pode ser executado de duas forma distintas:

1. Na consola de *input*, escreva `zoom` e carregue na tecla ENTER.
2. No menu de objectos, escolha a barra "all" e clique no botão marcado **A**. Irá aparecer um menu com várias opções, de entre as quais deverá seleccionar "zoom".

Qualquer das duas opções acima produzirá o mesmo resultado: a molécula de lisozima volta preencher o écran de forma centrada.

Para alterar o modo de representação da molécula, podemos usar o botão "S" (show) do menu de objectos (note-se que neste momento é indiferente usar a barra "all" ou "2vb1", porque temos apenas uma molécula carregada no PyMOL. Vamos seleccionar a opção "sticks" no menu do botão "S". O resultado deste comando é a exibição da representação das ligações químicas da proteína ("sticks") da estrutura, a qual irá aparecer sobreposta à representação já existente em "cartoon". Se pretendermos visualizar a nossa molécula exclusivamente na forma de "sticks", clicamos no botão de opção "H" (hide) do menu de objectos e seleccionar "cartoon".

Para colorir a nossa molécula de outra cor, vamos usar o comando "C" do menu de objectos (usando a linha "all" ou "2vb1"). Escolher "reds" e depois "tv red".

**Exercício:** Representar novamente a molécula *apenas* como sticks, e colori-la a azul. Em seguida, representar a molécula como *surface* (sem esconder a representação em *sticks*).

Em seguida vamos carregar uma segunda molécula no PyMOL, mas desta vez vamos fazê-lo directamente pela interface do programa, usando a opção "Get PDB" do menu "File". Na caixa que aparece, verifique que a "tick box" correspondente a "Pdb Structure" se encontra marcada, e apenas essa. Coloque o código "3b0i"(lactalbumina humana) na caixa "PDB ID" e carregue no botão "Download". A representação da estrutura 3b0i é adicionada à janela de visualização do PyMOL, e surge uma nova linha no menu de objectos. Note-se que neste momento temos as duas moléculas parcialmente sobrepostas e em representações diferentes. Para colocar as duas moléculas como "sticks", vamos usar o botão "S" da linha "all" do menu de objectos, e escolher a primeira entrada do menu que aparece ("as"), e no menu "As:" escolhemos "sticks". Esta opção remove automaticamente todas as outras visualizações presentes, deixando apenas a nova representação "sticks" (compare-se com o comando "show" normal, o qual simplesmente adiciona a cada nova representação às já visualizadas).

**Exercício:** Representar as duas moléculas como "ribbon", colorir uma de vermelho e a outra a verde.

## Utilização de comandos de consola

Conforme anteriormente explicado, a região de input da consola (número 4 na Figura 1) permite enviar comandos de texto para o software. Muitos destes comandos são alternativas, frequentemente mais rápidas, aos comandos acessíveis em outras regiões da interface (barra de menu ou menu de objectos). Para além destes existem outros comandos apenas acessíveis através desta interface.

Comecemos por apagar todas as moléculas carregadas no software com o seguinte comando de consola:

```
delete all
```

Em seguida, vamos voltar a carregar a molécula 2vb1, mas desta feita com o comando:

```
fetch 2vb1
```

(este comando irá descarregar o ficheiro de estrutura do a partir do PDB, a menos que este se encontre já na pasta onde é corrido o programa). Com o comando `fetch` o PyMOL descarrega a molécula no format **mmCIF** pelo que será guardado na pasta com nome "2vb1.cif".

De seguida, vamos representar esta molécula como "sticks", com o seguinte comando:

```
show sticks, 2vb1
```

E vamos carregar novamente a segunda molécula, com o comando

```
fetch 3b0i
```

Vamos representar as moléculas como "ribbon", eliminando as outras representações.

```
as ribbon
```

Vamos colorir a primeira molécula a vermelho:

```
color red, 2vb1
```

e a segunda a verde:

```
color green, 3b0i
```

De seguida vamos usar o comando "align" do PyMOL, que sobrepõe as duas moléculas, permitindo comparar as suas estruturas e fornecendo uma medida da diferença entre elas:

```
align 3b0i, 2vb1
```

Note-se que o resultado do comando aparece na consola de output (no. 4 na Figura 1), exibindo um RMSD de 0.9 Angstrom, o qual indica um elevado grau de similaridade entre as estruturas da lisozima e lactalbumina. Note-se, ainda, que o comando move a primeira molécula (3b0i) sobre a segunda, descentrando a imagem. Para voltarmos a centrar, executamos o comando:

```
zoom
```

Este comando centre e re-escala a cena de modo a preencher a área de visualização.

**Exercício:** remova a molécula 3b0i, e represente a molécula 2vb1 como superfície.

## Linguagem de selecção do PyMOL

Para tirar máximo de partido da linguagem de comandos de consola do PyMOL, necessitamos de mecanismo simbólico de selecção de regiões arbitrárias da molécula em estudo (p.ex, se quisermos colorir de vermelho todos os resíduos de ácido aspártico, representar as ligações peptídicas como sticks, etc...).

O PyMOL possui uma linguagem de selecção baseada na seguinte hierarquia de organização dos objectos:

objecto → segmento → cadeia → resíduo → átomo

1. Objecto: geralmente uma molécula, mas podendo em certos casos conter várias cópias da mesma molécula, oligómero proteico ou complexo proteína-ligando(s). Pode conter uma ou mais cadeias.
2. Segmento: usado para designar diferentes regiões de uma molécula, que podem corresponder a grupos de cadeias, ligandos, etc.
3. Cadeia: este nível de organização é usada para as várias cadeias polipeptídicas que constituem moléculas multiméricas, mas também para distinguir ligandos, iões metálicos e outras espécies presentes.
4. Número de um resíduo de aminoácido (ou nome, ver abaixo). Também pode ser usado para o nome de uma molécula pequena ou grupo prostético não-aminoacídico.
5. Nome de um átomo, de acordo com a nomenclatura standard para ficheiros PDB (ex. "CA" designa o carbono alfa-aminoacídico).

Na linguagem de selecção do PyMOL os diferentes níveis de hierarquia são separados por "/". Por exemplo:

```
/2vb1/A/A/123/CA
```

referencia o objecto 2vb1 (possivelmente correspondente à estrutura de código 2vb1 do PDB), segmento A, cadeia A (o nível de organização "cadeia" é utilizado para referenciar as várias cadeias constituintes da estrutura quaternária de uma dada proteína), resíduo de aminoácido número 123, átomo de código CA (carbono alfa). Esta especificação de selecção poderia por exemplo ser usada no seguinte comando:

```
zoom /2vb1/A/A/123/CA
```

o qual irá centrar e aproximar a imagem do carbono alfa do resíduo 123 da cadeia A, segmento A da molécula 2vb1.

Note-se que são permitidas especificações mais curtas, por exemplo:

```
123/CA
```

refere o carbono alfa do resíduo nº 123 de *todas* as cadeias presentes nas estruturas carregas no programa. Também são possíveis representações do tipo:

```
2vb1////CA
```

esta expressão refere todos os átomos do tipo carbono alfa da molécula 2vb1 e pode, por exemplo, ser usada no seguinte comando de consola:

```
color red, 2vb1////CA
```

Nota importante: expressões de selecção terminadas num número de resíduo necessitam de "/" final para serem válidas. Por exemplo:

```
color green, 123/
```

irá colorir de verde todos os átomos do resíduo 123 de todas as cadeias presentes, enquanto

```
color red, 123
```

irá produzir uma mensagem de erro. É possível, também, especificar um intervalo de resíduos. O comando:

```
color red, 10-123/
```

irá colorir de vermelho os resíduos de números 10 a 123. Para especificar uma selecção não-contígua de resíduos, pode usar a seguinte notação:

```
color red, 40+70/
```

que neste caso irá colorir de vermelhos os resíduos de aminoácido de números 40 e 70.

Podem igualmente ser especificados nomes de resíduos, mas note-se que em geral estas especificações serão ambíguas, já que poderá existir (geralmente existe) mais que um resíduo com o mesmo nome. Por exemplo, o seguinte comando de consola:

```
color red, 2vb1//A/HIS/
```

irá colorir de vermelho todos os resíduos de Histidina presentes na cadeia A da proteína 2vb1. Enquanto que o comando:

```
show sticks, 2vb1//A/ASP+GLU/`
```

irá representar como *sticks* todos os resíduos de ácido aspártico e glutâmico da proteína 2vb1.

**Exercício 1:** Representar a molécula 2vb1 como *ribbon*. Colorir a molécula de vermelho. Represente como *sticks* de cor verde os resíduos Glu 35 e Asp 52. Represente a molécula como superfície e observe a localização dos resíduos catalíticos no centro activo. Para melhor fazer esta visualização, execute o seguinte comando de consola: `set transparency, 0.2`, o qual tornará a superfície translúcida, permitindo visualizar os resíduos activos.

**Exercício 2:** Carregue novamente a estrutura 3b01 com o comando de consola `load 3b01.cif` (note a extensão ".cif" pois o ficheiro foi anteriormente descarregado com o comando de consola "fetch"). Remova a representação em "cartoon" da molécula 3b0i sem apagar esta mesma representação na molécula 2vb1. Represente a molécula 3b0i como sticks, e sobreponha-a com o comando "align" à estrutura da 2vb1. Neste momento gostaríamos de saber se os

resíduos da lactalbumina (3b0i) na posição correspondente ao centro activo da lisozima são conservados ou não. Experimente colorir os resíduos da lactalbumina usando um comando semelhante ao que usou para a lisozima (com os mesmos números de aminoácidos). O que observa?... Precisamos agora de comandos que nos permitem identificar resíduos na estrutura, tanto em termos de número com de tipo químico do grupo (nome de resíduo de aminoácido, nome de átomo, nome de qualquer outra molécula presente no ficheiro de estrutura). Para tal necessitamos da secção seguinte.

## Seleção de átomos e grupos nas estruturas

Para além da visualização das estruturas moleculares, é essencial termos a possibilidade de identificar grupos, assinalar pontos de interesse e "rotular" determinadas regiões da proteína com "etiquetas" de texto. É importante, também, ter a possibilidade de *seleccionar* regiões da estrutura para serem processadas como um grupo com os diferentes comandos disponíveis na interface do PyMOL.

Para seleccionar ("picking") um átomo ou grupo de átomos, usamos o botão esquerdo do rato clicamos sobre a região de interesse na estrutura. Vamos testar esta funcionalidade: remova a estrutura **2vb1** com comando de consola `remove 2vb1` e represente a estrutura **3b0i** com sticks com `as sticks, 3b0i`. Experimente agora clicar com o botão direito do rato numa região da estrutura. Alguns átomos aparecerão "marcados" com um ponto, e surge uma nova entrada com o nome "(sele)" no menu de objectos. Pode pensar nesta nova entrada como um objecto "virtual" ao qual é possível aplicar (praticamente) todas as operações aplicáveis a qualquer outro objecto. As entradas "ASHCL" do menu de objectos irão operar sobre "(sele)" afectando todos os átomos correntemente seleccionados. Experimente usar o menu "C" para colorir "(sele)" a vermelho. Os átomos anteriormente marcados deverão mudar de cor. Experimenta outras operações do menu de objectos, como "zoom", "sticks" ou "surface". A selecção pode ser *expandida* clicando em outra regiões não-marcadas da molécula. Se clicar numa região marcada, irá desmarcá-la. E se clicar no espaço vazio irá remover completamente a selecção, fazendo desaparecer o objecto "(sele)" do menu de objectos.

No entanto, há uma questão importante: porque fica mais do que um átomo marcado quando clicamos sobre a estrutura?

Para responder a esta questão, note em primeiro lugar que ao clicar num átomo ou ligação irá seleccionar a **totalidade** dos átomos constituintes do resíduo ao qual pertence o átomo clicado. Este comportamento é consequência do **modo de selecção** ("picking mode") do PyMOL. Por defeito, este modo é "residue", significando que resíduos inteiros são seleccionados com cada clique. O estado corrente do modo de selecção pode ser visionado e alterado no canto inferior direito da janela do PyMOL (região 6 da figura 1). Neste momento deverá ler-se "Picking: residues". Clicando nesta linha, poderá rodar pelos vários modos de selecção possíveis:

residues → chains → segments → objects → molecules → c-alphas → residues ....\_\_

Experimente clicar uma vez para passar ao modo "chain" e clique novamente na molécula 3b0i. Deverá agora ter seleccionado a molécula inteira, pois esta só contém uma única cadeia polipeptídica.

**Exercício 1:** Limpar a sessão do PyMOL removendo todos os objectos com o comando "delete all". Obtena a estrutura da tripsina bovina com o código 2ra3 e carregue-a no PyMOL. Esta estrutura contém a tripsina em complexo com o BPTI (Bovine Pancreatic Trypsin Inhibitor). Estão presentes na estrutura dois complexos, em cada um dos quais a molécula mais pequena é o BPTI e a maior é tripsina. Para facilitar a visualização, vamos colorir as cadeias de diferentes cores, usando para tal o objecto "all" do menu de objectos e escolhendo o sub-menu "C" (colorir), e



seguidamente "bychain->bychain". Remova as cadeias ("chains") de um dos complexos, usando o "selection mode" a nível de "chain". Colori o BPTI de amarelo e a trispina de verde.

## Seleccção com os comandos de consola e name selections

A selecção de regiões na estrutura pode ser feita através de comandos de consola, especificando a região através da linguagem de selecção anteriormente descrita. As selecção assim criadas poderão ter um nome ("named selections") o qual irá aparecer no menu de objectos em vez do default "(sele)". Para ilustrar isto, vamos usar a estrutura carregada no exercício anterior e criar uma selecção contendo os resíduos do centro activo com o seguinte comando de consola:

```
select active_site, ///A/57+102+195/
```

irá dessa forma ser criada uma nova entrada no menu de objectos com o nome `active_site`, a qual poderá ser usada para colorir este resíduos de outra cor, ou mudar a sua representação. (Experimente colorir os resíduos em "magenta" e mudar a sua representação para "spheres". Depois centre os mesmos no écran com a opção zoom do sub-menu "A" do menu de objectos,)

## Identificação de regiões nas estruturas

Para além da selecção de regiões específicas da estrutura, é importante sermos capazes de "interrogar" uma determinada região relativamente ao tipo químicos do átomos ou grupos que a constituem. Para tal basta clicar com o botão *direito* do rato sobre um átomo ou ligação da estrutura. Surgirá um menu, no qual a barra de título (a azul) contém a descrição completa do átomo clicado, desde o nome do objecto a o nome do átomo.

**Exercício 1:** Usando o modo de interrogação com o botão direito do rato, descubra a identidade de cada um dos três aminoácidos que constituem o centro activo da molécula de tripsina.