

Bioinformática

Exercícios - Visualização e comparação de estruturas

Nota: caso esteja a usar o seu computador pessoal, deverá instalar o software PyMOL a partir do site www.pymol.org.

1. A partir do site do Protein Data Bank (www.rcsb.org) obtenha os ficheiros PDB da tripsina humana (código pdb **2RA3**), da tripsina de *Streptomyces griseus* (**1SGT**) e da proteinase V8 de *Staphylococcus aureus* (**1WCZ**). Obtenha as sequências, em formato FASTA, de cada uma das proteínas a partir de www.uniprot.org
 - a) Faça o alinhamento da tripsina humana com cada uma das outras 2 sequências usando o site "Emboss" (<http://www.ebi.ac.uk/emboss/align>) com opção "water" (Smith-Waterman), anotando as percentagens de identidade obtidas. Parece-lhe que os alinhamentos têm ambos significado?
 - b) Abre o ficheiro 2RA3 com o programa PyMOL e use o comando "Color" com opção "by chain" (menu no canto superior direito) para colorir cada uma das 4 cadeias polipeptídicas existentes no ficheiro 2RA3. O ficheiro contém **duas cópias** de um complexo tripsina-inibidor (em que a molécula maior é a tripsina e mais pequena o inibidor). No menu "Mouse" escolha o sub-menu "Selection Mode" e deste a opção "chain". A partir deste momento, ao clicar num átomo irá seleccionar a cadeia inteira. Selecciona 3 das quatro cadeias, deixando apenas por seleccionar umas das cadeias maiores (molécula de tripsina). Na entrada "(select)" do canto superior direito escolha "remove atoms". Ficará assim com uma única molécula de tripsina no écran
 - c) Usando a opção "Open" do menu "File" leia o ficheiro 1SGT no PyMOL. Usando o comando "zoom" para garantir a visualização simultânea das duas moléculas.
 - d) Use o comando "as ribbon" para observar as proteínas em modo "ribbon" (é visualizado apenas o traço da cadeia polipeptídica de cada proteína).
 - e) Faça a sobreposição das estruturas usando o comando "align 1SGT, 2RA3". Anote o RMSD obtido para a sobreposição.
 - f) Repita o procedimento anterior para o ficheiro 1WCZ.pdb (ler e alinhar com 2RA3 usando "align"). Neste caso deverá obter um RMSD muito elevado (~15Å), e um alinhamento incorrecto. Isto acontece porque o comando "align" de PyMOL baseia-se no alinhamento das sequências para fazer a sobreposição das estruturas, mas as sequências das duas proteínas (2RA3 e 1WCZ) não têm similaridade detectável, como deverá ter verificado na alínea a).
 - g) Repita o alinhamento estrutural anterior, mas desta vez usando o comando "cealign" do PyMOL, o qual implementa um algoritmo muito mais rigoroso. Neste caso deverá obter um RMSD muito mais baixo e uma boa sobreposição das estruturas.
 - h) Para obtermos um alinhamento das sequências das duas proteínas 2RA3 e 2WCZ com base no alinhamento das suas sequências, vamos usar o servidor de alinhamento estrutural "Top Match" (<https://topmatch.services.came.sbg.ac.at/>). Na página de entrada, introduza os códigos das duas proteínas nas caixas "Query" e "Target", acrescentando ",A" a cada código par indicar ao servidor que apenas pretendemos alinhar as cadeias "A" de cada uma das estruturas. Carregue em "Match" para realizar o alinhamento estrutural. Na página de resultados, observe o valores de RMSD e percentagem de identidade obtidos e compare o alinhamento produzido por este método com o alinhamento anteriormente com obtido com a variante "water" do site Emboss. Note a diferença de valores de percentagem de identidade. Qual deverá ser o alinhamento mais correto e porquê? Usando a visualizador de estrutura e alinhamento, identifique as regiões das estruturas que correspondem a inserções ou deleções no alinhamento duas sequências. O que podemos concluir sobre a localização das inserções e deleções?
 - i) Compare os valores de RMSD obtidos com as correspondentes percentagens de identidade entre as sequências. O que pode concluir?

2. O alinhamento das sequências das cadeias alfa e beta da hemoglobina com a leghemoglobina de *Lupinus* e com a proteína GST-2 de *C.elegans* parecem indicar uma similaridade mais elevada com esta última. Obtenha as estruturas das quatro proteínas: as estruturas experimentais das 3 primeiras estão disponíveis no RCSB Protein Databank, enquanto para a 4ª proteína (GST-2) não existe estrutura experimental - neste caso deverá descarregar um modelo de homologia a partir do Swiss-MODEL Repository (<https://swissmodel.expasy.org/repository>). Faça o alinhamento estrutural da cadeia alfa com as outras 3 proteínas no PyMOL, tomando nota dos valores de RMSD (**NOTA:** deve usar o comando "cealign" do PyMOL, já que alguns dos pares apresentam uma percentagem de identidade muito baixa). Analise os resultados e tire conclusões relativamente à relação de homologia entre estas sequências.

3. Obtenha a partir do PDB uma estrutura da rodanese ("rhodanese") com o código **1DP2**. Carregue esta estrutura no PyMOL
 - a) Visualize a estrutura no PyMOL. Use "remove HOH/" e "remove LPA/" para deixar apenas a molécula de proteína no ficheiro. Use o comando "as ribbon" para visualizar apenas o traço da cadeia polipeptídica (por união dos CD-alfas consecutivos).
 - b) Use os comandos "create part_a, 1DP2///1-157/" e "create part_b, 1DP2///158-292/" para criar dois objectos "part_a" e "part_b" contendo as duas metades da molécula. Use o comando "color" para colorir as duas partes de vermelho e verde.
 - c) Use o comando "cealign part_a, part_b" para fazer o alinhamento estrutural das duas moléculas. Anote o RMS obtido. Represente as estruturas como "cartoon" usando o comando "as cartoon". Observe a correspondência das estruturas secundárias (hélices e folhas) das duas proteínas alinhadas.
 - d) Para visualizar as duas sequências alinhadas por estrutura vamos necessitar de usar uma ferramenta exterior ao PyMOL. Grave os dois pedaços da molécula no seu computador em dois ficheiros PDB, usando os comandos "save part_a.pdb, part_a" e "save part_b, part_b".
 - e) Abra a página do servidor de comparação estrutural DALI (<http://ekhidna2.biocenter.helsinki.fi/dali/>). . Seleccione a tab "pairwise". Em "STEP 1" clique no botão "Choose File" e seleccione o ficheiro "part_a.pdb". Em "STEP 2" clique no botão "+", em seguida "Choose File" e seleccione o ficheiro "part_b.pdb". Clique em "Submit" para calcular o alinhamento estrutural das duas estruturas "part_a" e "part_b".
 - f) Anote o RMS e a percentagem de identidade do alinhamento estrutural das duas sequências produzido pelo servidor DALI.
 - g) Com base nesta análise, como poderá ter surgido o gene da rodanese?... Será este um evento recente ou antigo? (Tenha em conta a percentagem de identidade das sequências)

4. Obtenha as seguintes sequências de proteases de serina *humanas* a partir de UniProt (<http://www.uniprot.org>) : Tripsina 1 (Trypsin), Elastase 1, Quimotripsinogénico B (Chymotrypsinogen B), Kalikreína (Kallikrein), Quimase (Chymase) e Catepsina G (Cathepsin G). Obtenha as estruturas destas proteínas a partir do Protein Databank (códigos PDB: 1PJP, 3EST, 3TPI, 2CGA, 1SPJ e 1CGH).
 - a) Faça um alinhamento múltiplo das sequências com o programa T-Coffee (<http://www.ebi.ac.uk/tcoffee>) e identifique possíveis regiões correspondentes à cavidade do centro activo.
 - b) Identifique os três resíduos da tríade catalítica (Ser, Asp, His) no alinhamento.
 - c) Leia os ficheiros PDB no programa PyMol. Remova as moléculas de solvente, ligandos e duplicados.

- d) Use o comando "align **obj1**, **obj2**" (em que obj1 e obj2 são os nomes de quaisquer duas proteínas) para alinhar todas as estruturas com a estrutura da tripsina.
- e) Compare o alinhamento das estruturas com alinhamento de sequências antes obtido e identifique "loops" (zonas de divergência estrutural à superfície das proteínas).
- f) Na estrutura da tripsina identifique a tríade catalítica, representando-a com o modo "sticks" e com uma cor diferente do resto da proteína.
- g) Represente como "sticks" as tríades catalíticas de todas as moléculas sobrepostas, e observe a elevada conservação da geometria do centro activo.