

# 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](http://www.pymol.org).

1. A partir do site do Protein Data Bank ([www.rcsb.org](http://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](http://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) Abra 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 submenu “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 uma 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. Use 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 passo anterior, mas desta vez usando o comando “cealign” do PyMOL, o qual implementa um algoritmo mais rigoroso de alinhamento estrutural, o qual não depende de um alinhamento prévio das sequências. Neste caso deverá obter um RMSD muito mais baixo e uma boa sobreposição das estruturas.
  - h) Para produzir um alinhamento das sequências das duas proteínas 2RA3 e 1WCZ com base no alinhamento das suas estruturas, 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 para 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 os valores de RMSD e percentagem de identidade produzidos por TopMatch e compare o alinhamento obtido por este método com o alinhamento de sequências produzido por EMBOSS/water. Note a diferença de valores de percentagem de identidade. Qual deverá ser o alinhamento mais correto e porquê? Usando o visualizador de estrutura e alinhamento, identifique as regiões das estruturas que correspondem a inserções ou deleções no alinhamento das duas sequências. Onde se encontram preferencialmente localizadas as regiões de inserção e deleção na estrutura das duas proteínas?
  - i) Compare os valores de RMSD obtidos com as correspondentes percentagens de identidade entre as sequências. O que pode concluir?

2. A rodanese, também conhecida como tiosulfato transferase, é um enzima mitocondrial responsável pela destoxificação do cianeto. Obtenha a estrutura da rodanese bovina, com o código **1DP2**, a partir do Protein Data Bank (<https://www.rcsb.org>). Carregue esta estrutura no PyMOL e siga os seguintes passos:
  - a) Visualize a estrutura no PyMOL. Use “remove solvent” para remover as moléculas de água da estrutura, e “remove LPB/” para remove a molécula de ácido lipóico que se encontra no centro activo da rodanese, ficando apenas com a cadeia de proteína no PyMOL. Use comando “as ribbon” para visualizar apenas o traço da cadeia polipeptídica (por união dos átomos de carbono alfa de resíduos de aminoácidos 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. Remova a molécula original com o comando “remove 1DP2”. Use o comando “color” para colorir as duas partes de vermelho e azul (“color red, part\_a” e “color cyan, part\_b”),
  - c) Use o comando “cealign part\_a, part\_b” para fazer o alinhamento estrutural das duas partes da molécula de rodanese. Anote o RMS obtido. Represente as estruturas como “cartoon” usando o comando “as cartoon”. Observe a correspondência das estruturas secundárias (hélices e folhas) dos dois fragmentos alinhados. Parecem-lhe similares?
  - d) Para visualizar o alinhamento de sequência baseado em estrutura para os dois fragmentos da rodanese, vamos necessitar de 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.pdb, part\_b”.
  - e) Abra a página do servidor de comparação estrutural DALI (<http://ekhidna2.biocenter.helsinki.fi/dali/>). Selecione a tab “pairwise”. Em “STEP 1” clique no botão “Choose File” e selecione 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 RMSD e a percentagem de identidade do alinhamento estrutural das duas sequências produzido pelo servidor DALI.
  - g) Alinhe os dois fragmentos da sequência da rodanese usando o site “Emboss” (<http://www.ebi.ac.uk/emboss/align>) com opção “**water**” (Smith-Waterman).
  - h) Analise a significância deste alinhamento usando o servidor PRSS ([https://fastademo.bioch.virginia.edu/fasta\\_www2/fasta\\_www.cgi?rm=lalign](https://fastademo.bioch.virginia.edu/fasta_www2/fasta_www.cgi?rm=lalign)). O alinhamento será estatisticamente significativo?
  - i) 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)
  
3. Obtenha as estruturas da hemoglobina humana (**6KAS**), leghemoglobina de tremçoço (**2GDM**) e glutationo-S-transferases 5 (GST-5) de *C. Elegans* (**1ZL9**) a partir do RCSB Protein Databank. Faça o alinhamento estrutural do cadeia alfa da hemoglobina contra a cadeia beta da mesma proteína, e contra as outras duas 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, comparando com o alinhamento das correspondentes sequências usando a opção “Water” do programa EMBOSS.
  
4. Obtenha as seguintes sequências de proteases de serina *humanas* a partir de UniProt (<http://www.uniprot.org>): Tripsina 1 (Trypsin), Elastase 1, Quimotripsinogéneo 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 (<https://www.ebi.ac.uk/Tools/msa/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.