

LABORATÓRIO DE BIOINFORMÁTICA

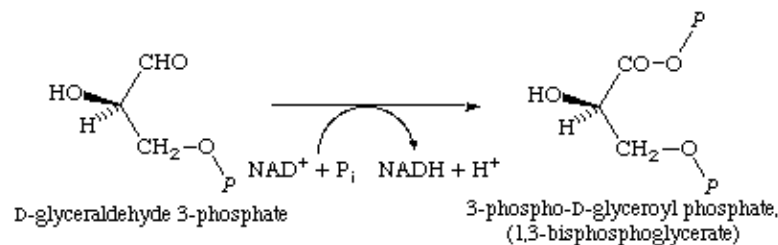
Nesta aula, vocês irão usar os seus conhecimentos de biologia molecular, assim como ferramentas computacionais e bancos de dados, para comparar a sequência primária e estrutura tridimensional de proteínas. Para isto, utilizaremos proteínas importantes para diferentes organismos e observaremos como elas mudaram ao longo da evolução. Primeiro, vamos conhecer as proteínas de estudo. Em seguida, apresentaremos os organismos modelos. Eles foram escolhidos para representarem diferentes estágios da evolução dos seres vivos na terra: das bactérias, aos primeiros organismos eucariotos, os primeiros organismos multicelulares, até o homem.

As proteínas e seus genes:

Enzima Gliceraldeído-3-fosfato desidrogenase (Humana)

código: GAPDH
proteína: NP_001276674
Estrutura: 4WNC

A enzima GAPDH catalisa a sexta reação da via glicolítica. Esta via, converte a glicose em piruvato e permite a produção de moléculas de ATP (energia) para a célula. A reação catalisada por esta enzima está descrita abaixo:

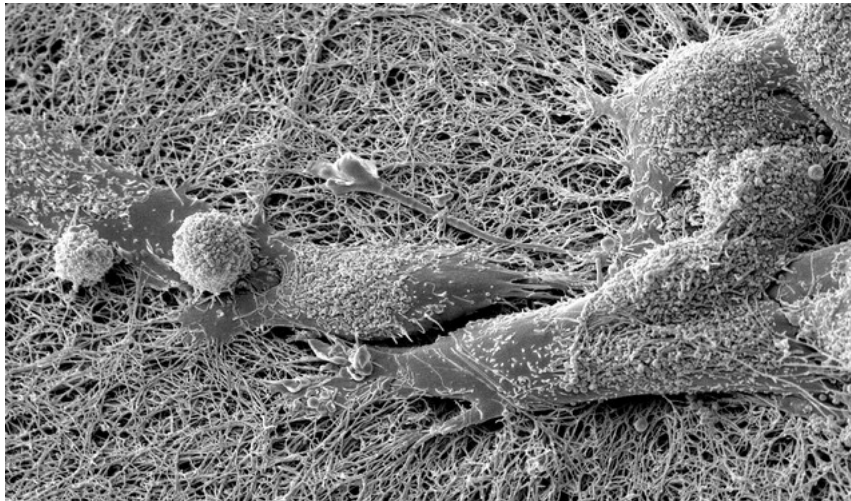


A GAPDH utiliza o NAD⁺ como cofator e incorpora uma molécula de fosfato inorgânico no D-gliceraldeído-3-fostato, produzindo o 1-3-bisfosfoglicerato. É uma enzima essencial para todos os organismos vivos.

Colágeno (tipo II, cadeia alfa-1)(humano)

Código: COL2A1
proteína: NP_001835
estrutura: não disponível

As proteínas dos diferentes tipos de colágenos são importantes elementos de sustentação para células e tecidos. São encontradas no parênquima e servem de moléculas de adesão para as células. São constituintes do tendões e ligamentos. Sua estrutura primária é rica nos aminoácidos glicina e prolina. O surgimento do colágeno permitiu que pela primeira vez organismos unicelulares se organizassem em estruturas multicelulares. Abaixo, foto obtida por microscopia eletrônica mostrando células (fibroblasto) crescendo aderida a fibras de colágeno.



Fonte: <https://ag.purdue.edu/arge/Microscopy/Pages/default.aspx>.

Hemoglobina cadeia alfa

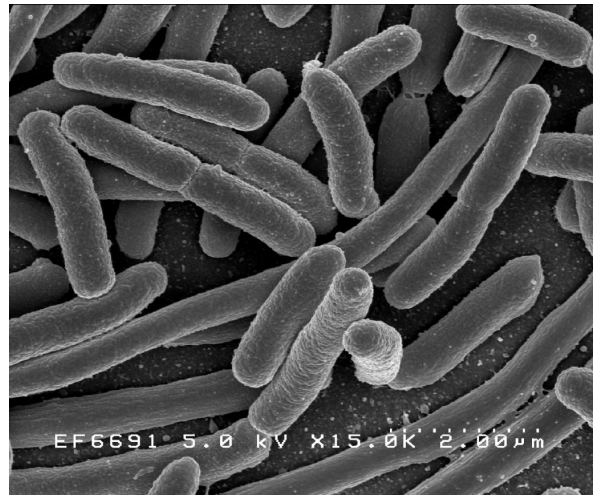
Código: HBA1
proteína: NP_000549
estrutura: 1A3N

A hemoglobina (frequentemente abreviada como Hb) é uma metaloproteína que contém o grupo prostético heme e ferro (Fe[II]). É encontrada nos glóbulos vermelhos (eritrócitos) e permite o transporte de oxigênio do pulmão para os tecidos através do sistema cardiovascular.

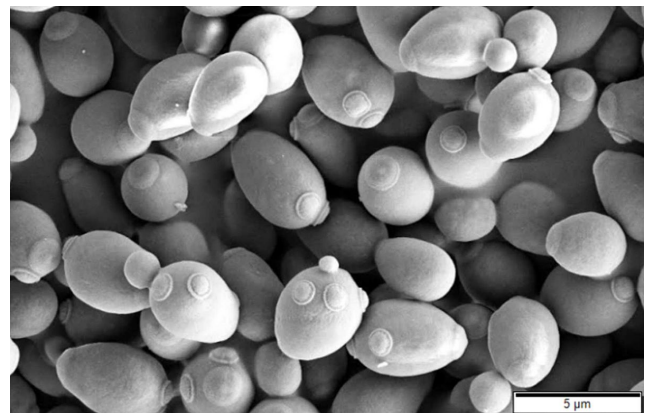
A hemoglobina é composta de 4 cadeias polipeptídicas (2 alfa e 2 beta) que sustentam os 4 grupamentos heme que contém o ferro. Cada íon ferro é capaz de se ligar uma molécula de oxigênio (O₂), ou seja, 4 moléculas de oxigênio para cada molécula de hemoglobina.

Organismos modelo:

Escherichia coli, mais conhecida pela abreviatura *E. coli*, é uma bactéria bacilar Gram-negativa que se encontra normalmente no trato gastrointestinal inferior dos organismos de sangue quente (endotérmicos). A maioria dos isolados de *E. coli* são inofensivas, mas alguns sorotipos podem causar graves intoxicações alimentares nos seres humanos, e são ocasionalmente responsáveis pela recolha de produtos alimentícios devido à sua contaminação. As cepas inofensivas constituem parte da flora intestinal humana normal, e podem ser benéficas para os seus hospedeiros ao produzirem vitamina K2 e impedirem que ali se estabeleçam outras espécies de bactérias patogênicas. Imagem ao lado: *E. coli* ao microscópio eletrônico.



Saccharomyces cerevisiae (*S. cerevisiae*) é um organismo eucarioto unicelular que pertence ao reino dos Fungos. É a levedura utilizada na produção do pão e também da cerveja, além de ser usada para a produção de etanol. Ela é utilizada como base para muitas indústrias, como a de panificação e de bebidas. Esse fungo é utilizado como fermento biológico, por liberar dióxido de carbono, por exemplo, na massa de pão, fazendo-a crescer. No caso das bebidas alcoólicas produzidas pelo processo de fermentação, o *S. cerevisiae* converte o açúcar em álcool etílico e também pode contribuir na formação de constituintes secundários responsáveis pelo sabor - é o caso da cerveja, rum e uísque. Também é um organismo muito utilizado como modelo no estudo da Bioquímica, Genética e Biologia Celular de eucariotas. Isto porque é de fácil manutenção em laboratório e o conhecimento biológico sobre ela é bem desenvolvido - o seu genoma já foi sequenciado. Imagem ao lado: *S. cerevisiae* ao microscópio eletrônico.



Poríferas (esponjas). Os poríferos ou Porífera (do latim porus, poro + phoros, portador de poros) é um filo do reino *Animalia*, sub-reino *Parazoa*, onde se enquadram os animais conhecidos como esponjas. As esponjas estão entre os animais mais simples, não possuem tecidos verdadeiros pois em sua camada externa e interna as células não apresentam lâmina basal (*parazoas*), também não apresentam músculos, sistema nervoso, nem órgãos internos.



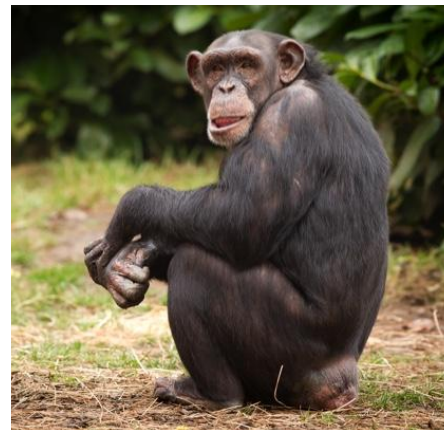
Xenopus (do latim, pé estranho) é um gênero de rãs (anfíbios) altamente aquáticas nativas da África subsariana. Há dezoito espécies no gênero *Xenopus*. A espécie mais conhecida do gênero é *Xenopus laevis*, estudada como organismo modelo. São animais complexos, com sistema nervoso, vascular, etc.



Mus musculus, também conhecido como camundongo ou rato-doméstico, é uma espécie de pequeno roedor da família dos murídeos, encontrado originalmente na Europa e Ásia, e atualmente distribuído por todo o mundo, geralmente associado a habitações humanas. Tem cerca de 8 cm de comprimento, pelagem macia, branca ou cinza-acastanhada, mais clara nas partes inferiores, orelhas grandes e arredondadas e cauda nua e longa. Seu genoma foi determinado e é um importante modelo animal para estudos científicos.

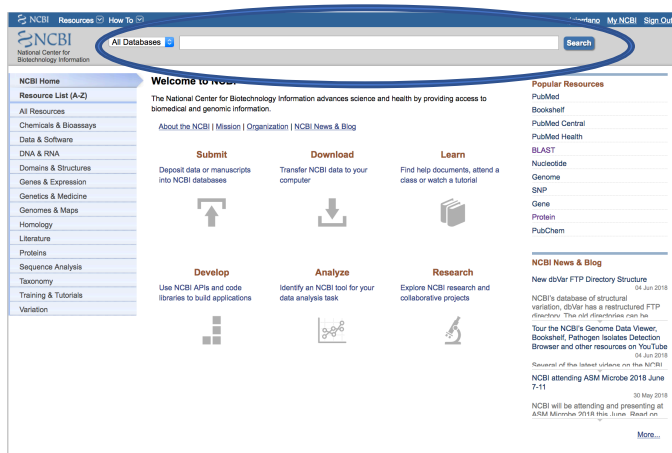


O chimpanzé (nome científico: *Pan troglodytes*), também conhecido como chimpanzé-robusto, é uma das duas espécies de chimpanzés que vivem no continente africano. Juntamente com o Bonobo, os chimpanzés-comuns são os parentes vivos mais próximos dos humanos. Ele pesa entre 40 e 65 kg e mede cerca de 1,3 a 1,6 m de altura. Têm um período de gestação de 8 meses, muito semelhante à dos humanos que é de 9 meses. O bebê é desmamado quando chega aos três anos de idade, mas mantém uma relação estreita com a sua mãe por mais alguns anos, atingindo a puberdade entre os 8 e 10 anos. Os chimpanzés-comuns vivem em média 50 anos em cativeiro.

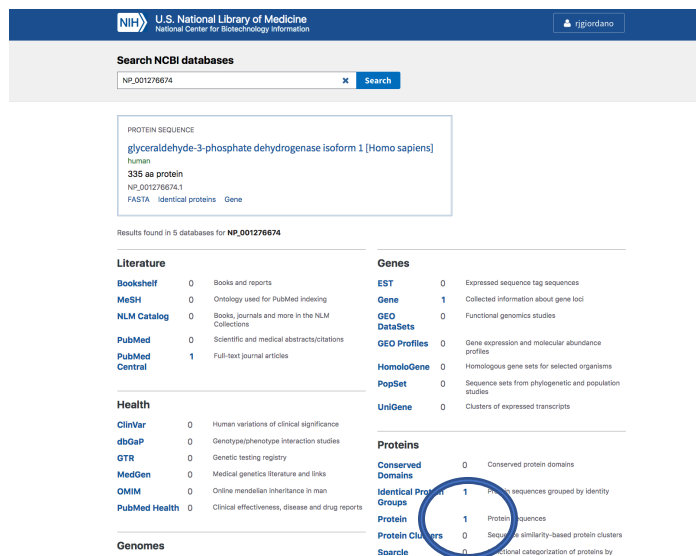


ROTEIRO DA AULA DE BIOINFORMÁTICA

No navegador (Firefox, Chrome), abra a página do NCBI (www.ncbi.nlm.nih.gov).



No campo de texto no alto da página, coloque o código de acesso da proteína GAPDH (NP_001276674).



Clique no número ao lado do link "Protein" para abrir a página contendo a sequência da enzima GAPDH.

NCBI Resources How To rjgioriano My NCBI Sign Out

Protein Protein Search Help

GenPept - Send to - Change region shown

glyceraldehyde-3-phosphate dehydrogenase isoform 1 [Homo sapiens]

NCBI Reference Sequence: NP_001276674.1

Identical Proteins FASTA Graphics

LOCUS NP_001276674.1 335 aa linear PRI 11-APR-2018

DEFINITION glyceraldehyde-3-phosphate dehydrogenase isoform 1 [Homo sapiens].

ACCESSION NP_001276674

VERSION NP_001276674.1

DBSOURCE REFSEQ: accession NM_001289745.2

KEYWORDS RefSeq.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Haplorrhini; Catarrhini; Hominidae; Homo.

REFERENCE 1 (residues 1 to 335)
AUTHORS Lazarev VF, Mikhaylova ER, Dutsyeva EA, Suezov RV, Guzhova IV and Margulis BA.
TITLE A hydrocortisone derivative binds to GAPDH and reduces the toxicity of extracellular polyglutamine-containing aggregates
JOURNAL Biochem. Biophys. Res. Commun. 487 (3), 723-727 (2017)
PUBMED 2845019

REMARK GeneRIF: This suggests that RX624 might be useful as a drug against polyglutamine pathologies, and that it could be administered exogenously without affecting target cell physiology. This protective effect was validated by the similar effect of an anti-GAPDH specific antibody.

REFERENCE 2 (residues 1 to 335)
AUTHORS Kosova AA, Khodyreva SN and Lavrik OI.
TITLE Role of Glyceraldehyde-3-Phosphate Dehydrogenase (GAPDH) in DNA Repair
JOURNAL Biochemistry Mosc. 82 (6), 643-654 (2017)
PUBMED 2861074

REMARK GeneRIF: GAPDH can interact with proteins participating in DNA repair, such as APE1, PARP1, HMG1, and HMG2. In this review, the functions of GAPDH associated with DNA repair are discussed in detail.
Review article

REFERENCE 3 (residues 1 to 335)
AUTHORS Huang Y, Zhang P, Yang Z, Wang P, Li H and Gao Z.

Analyze this sequence
Run BLAST
Identify Conserved Domains
Identify Conserved Features
Find in this Sequence
Show in Genome Data Viewer

Protein 3D Structure
Crystal structure of the T229K mutant of human GAPDH at 2.3 angstroms
PDB: 4WNI
Source: Homo sapiens
Method: X-ray Diffraction
Resolution: 2.3 Å
See all 4 structures...

Articles about the GAPDH gene
Role of Glyceraldehyde-3-Phosphate Dehydrogenase (GA) [Biochemistry (Mosc). 2017]
A hydrocortisone derivative binds to GAPDH and reduces t [Biochem Biophys Res Commun. 2017]
Interaction of glyceraldehyde-3-phosphate dehydrogenase an [Arch Biochem Biophys. 2017]
See all...

Agora, vamos procurar pela sequência da GAPDH dos outros organismos modelos. Para isto, clique no link "Run BLAST" que se encontra à direita, no alto da página. O programa BLAST permite varrer os bancos de dados de sequências de proteínas, RNA e DNA, em busca de sequências "similares".

NIH U.S. National Library of Medicine NCBI National Center for Biotechnology Information

BLAST® » blastp suite

Standard Protein BLAST

blastn blastp blastx tblastn tblastx

BLASTP programs search protein databases using a protein query. more...

Enter Query Sequence

Enter accession number(s), gi(s), or FASTA sequence(s) Clear Query subrange

NP_001276674.1 From To

Or, upload file Browse... No file selected.

Job Title Enter a descriptive title for your BLAST search

Align two or more sequences

Choose Search Set

Database: Non-redundant protein sequences (nr)

Organism Optional Exclude +

Enter organism common name, binomial, or tax id. Only 20 top taxa will be shown.

Exclude Optional Exclude +

Entrez Query Optional Enter an Entrez query to limit search YouTube Create custom database

Program Selection

Algorithm

Quick BLASTP (Accelerated protein-protein BLAST) New

blastp (protein-protein BLAST)

PSI-BLAST (Position-Specific Iterated BLAST)

PHI-BLAST (Pattern Hit Initiated BLAST)

PSI-BLAST (Simple Entrez-based Lookup Time Accelerated BLAST)

No campo "Organism", digite o nome do organismo para procurar a proteína desejada (*Escherichia coli*, *Saccharomyces cerevisiae*, *Porfíria*, *Xenopus*, *Mus musculus*, *Pan troglodytes*, *Homo sapiens*).

Repita o procedimento para todas as proteínas e organismos modelo e preencha a tabela abaixo com os dados obtidos:

Proteína	Organismo					
	<i>E. coli</i>	<i>S. cerevisiae</i>	<i>Porifera</i>	<i>Xenopus</i>	<i>M. musculus</i>	<i>P. troglodytes</i>
GAPDH						
Collagen-II						
HBA1						

Tabela 1. Preencha com os valores de "Identities" e "Positives" (em percentagem, %).

[Download](#) [GenPept](#) [Graphics](#)

PREDICTED: glyceraldehyde-3-phosphate dehydrogenase, partial [*Xenopus laevis*]
 Sequence ID: [XP_018098218.1](#) Length: 293 Number of Matches: 1

Range 1: 1 to 293 [GenPept](#) [Graphics](#) [Next Match](#) [Previous Match](#)

Score	Expect	Method	Identities	Positives	Gaps
603 bits(1555)	0.0	Compositional matrix adjust.	293/293(100%)	293/293(100%)	0/293(0%)
Query 1		MGKVKVGVNGFGRIGRLVTRAAFNSGKVDIVAINDPFDLNYMVMFYQYDSTHGKPHGTV			60
Sbjct 1		MGKVKVGVNGFGRIGRLVTRAAFNSGKVDIVAINDPFDLNYMVMFYQYDSTHGKPHGTV			60
Query 61		KAENGLVINGNPITIFQERDPSKIKWGDAGAEEVVESTGVFTTMEKAGAHLQGGAKRVI			120
Sbjct 61		KAENGLVINGNPITIFQERDPSKIKWGDAGAEEVVESTGVFTTMEKAGAHLQGGAKRVI			120
Query 121		ISAPSADAPMFVMGVNHEKYDNSLKIISNASCTTNCLAPLAKVIHDNFGIVEGLMTTVHA			180
Sbjct 121		ISAPSADAPMFVMGVNHEKYDNSLKIISNASCTTNCLAPLAKVIHDNFGIVEGLMTTVHA			180
Query 181		ITATQKTVDGSPGKLRDGRGALQNIIPASTGAAKAVGKVIPELNGKLTGMAFRVPTANV			240
Sbjct 181		ITATQKTVDGSPGKLRDGRGALQNIIPASTGAAKAVGKVIPELNGKLTGMAFRVPTANV			240
Query 241		SVVDLTCRLEKPAKYDDIKKVVQASEGPKGILGYTEHQVSSDFNSDTHSS		293	
Sbjct 241		SVVDLTCRLEKPAKYDDIKKVVQASEGPKGILGYTEHQVSSDFNSDTHSS		293	

PERGUNTA: Como você interpreta o resultado da Tabela 1? Pensa no tipo de organismo (unicelular, multicelular) e animal (primitivo, complexo).

Estrutura

Para visualizar as estruturas tridimensionais de uma proteína, precisamos de um programa de visualização de moléculas. O programa **Chimera** será utilizado nesta aula. Utilizando o programa "Chimera", abra a estrutura da proteína hemoglobina cadeia alfa (código 1A3N).

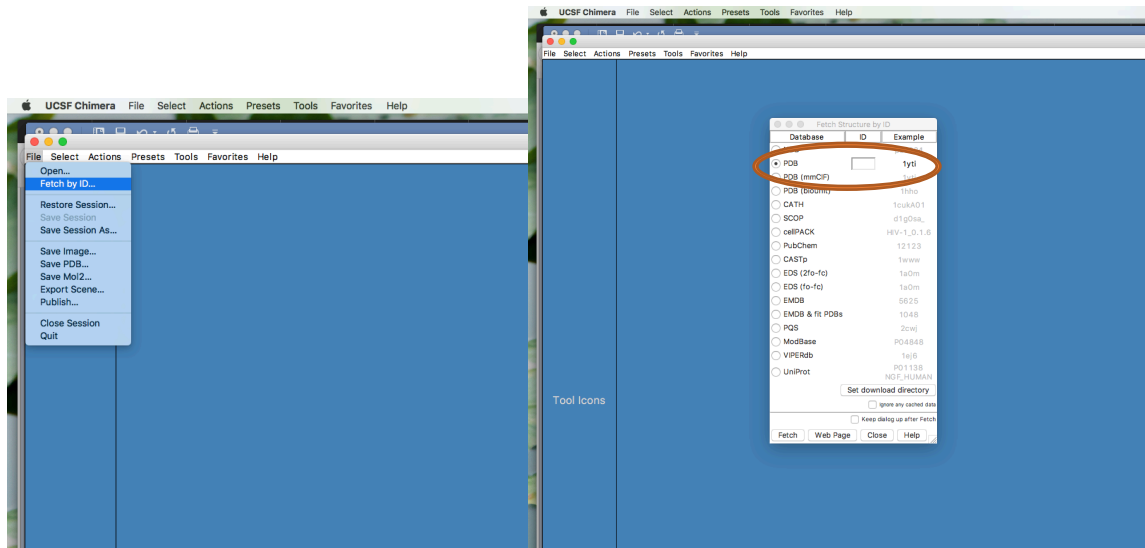


Figura - No menu "File" selecione "Fetch by ID". Insira o código "1A3N" no campo PDB.

PERGUNTA: Descreva de forma geral a estrutura terciária da hemoglobina (proteína fibrilar ou globular, presença de alfa-hélice, folhas pregueadas, alças, etc.). Onde se encontra o sítio de ligação ao oxigênio?

No Menu "Favorites" -> "Sequences" abra a janela de sequências e localize a sequência da cadeia B (beta). Procure pelo amino ácido "Glutâmico" (E) na sexta posição e selecione.

No Menu "Actions" -> "Atoms/Bonds" -> "Side Chain" selecione "show". No Menu "Actions" -> "Color" mude a cor deste aminoácido para vermelho

a) Este aminoácido está "exposto" ou "escondido" dentro da proteína?

Este é o amino ácido que é trocado por uma Valina na hemoglobina HbS (anemia falciforme), levando a agregação das moléculas de hemoglobina dentro da hemácia. Com base na sua análise da estrutura, como você explica este fenômeno?

COMPARANDO A ESTRUTURA DE DUAS PROTEÍNAS

A mioglobina e a hemoglobina são duas proteínas diferentes (sequência primária) mas têm a mesma função. Como isto é possível? Vamos usar as ferramentas de bioinformática para estudar esta questão.

Primeiro, vamos comparar as ESTRUTURAS PRIMÁRIAS. Procure pela sequência da mioglobina humana (código de acesso NP_976312). Utilizando o código de acesso, compara a sequência primária da mioglobina com a sequência primária da hemoglobina alfa (código de acesso NP_000549.1) e beta (NP_000509.1).

Para isto, utilize o programa BLAST (blast.ncbi.nlm.nih) com a opção "Align two or more sequences":

The image shows the NCBI BLAST web interface. At the top, there are logos for NIH and NCBI. The main heading is 'BLAST >> blastp suite'. Below this, there are tabs for different BLAST programs: 'blastn', 'blastp', 'blastx', 'tblastn', and 'tblastx'. The 'blastp' tab is selected. The interface is divided into sections for 'Enter Query Sequence' and 'Enter Subject Sequence'. In the 'Enter Query Sequence' section, there is a text input field containing 'Sequência 1' in red. Below this, there is a 'Job Title' field and a checkbox labeled 'Align two or more sequences' which is checked and circled in red. In the 'Enter Subject Sequence' section, there is a text input field containing 'Sequência 2' in red. At the bottom, there is a 'Program Selection' section with a radio button selected for 'blastp (protein-protein BLAST)'. A blue arrow points to the 'BLAST' button at the bottom left of the interface.

Figura - Selecione a opção "Align two or more sequences" (indicado pelo círculo vermelho) e coloque as sequências que você quer comparar, uma em cada janela. Clique "BLAST" (seta azul).

Se for necessário, mude o parâmetro "expected value" para 1000 (no caso da hemoglobina beta).

a) Como você interpreta os resultados? Qual o grau de identidade entre as proteínas?

Comparação das estruturas 3D de duas proteínas.

Agora vamos comparar as estruturas. No programa Chimera, abra primeiro a estrutura da hemoglobina (1A3N):

1. Feche a seção no Chimera para remover todas as sequências.
2. Abra a sequência da hemoglobina 1A3N no Menu File -> Fetch by ID
3. No Menu "Select", selecione a Chain A
4. Ainda no menu "Select", selecione "Invert (all models)"
5. No Menu "Actions" -> "Atoms/Bonds", selecione "delete"

Agora você visualizará apenas a cadeia alfa da hemoglobina.

Abra a estrutura da mioglobina 1A6N no Menu File -> Fetch by ID

Ambas as estruturas estão visíveis. Agora, vamos sobre por as estruturas.

1. No Menu "Tools" -> "Structure Comparison", selecione "Matchmaker"
2. Escolha as opções como indicado na figura e clique "Apply"

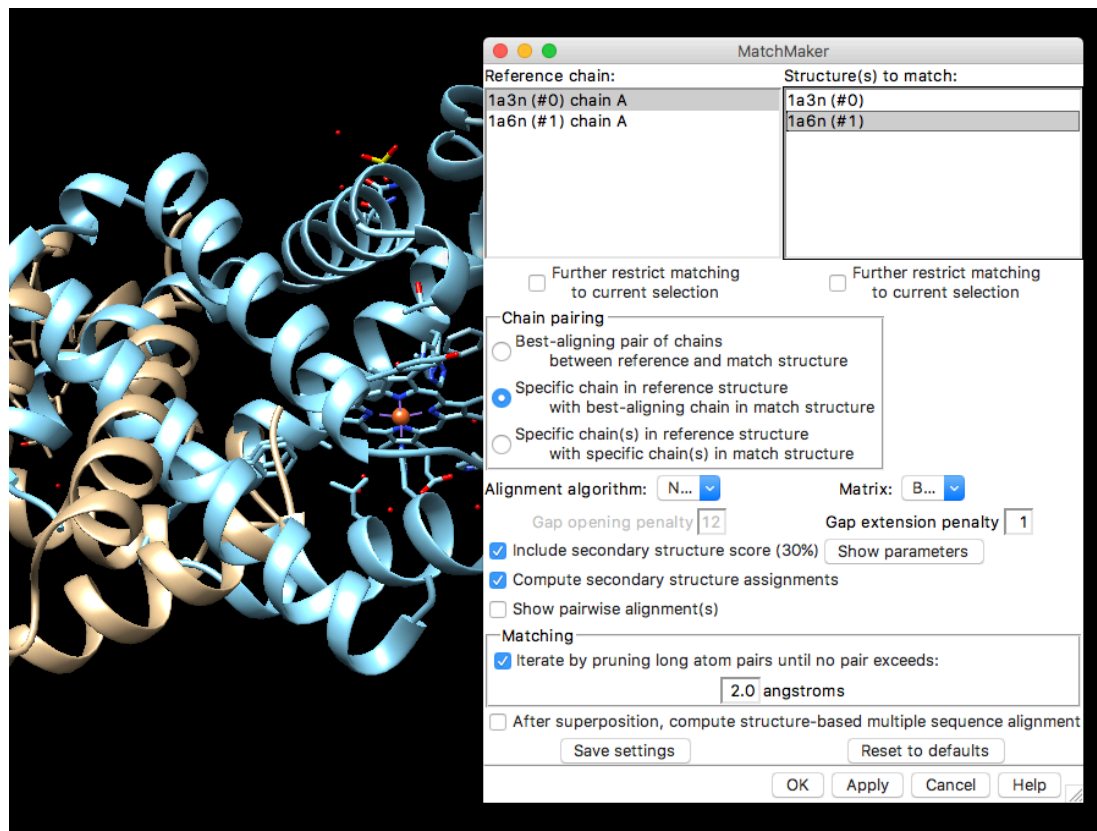


Figura - Use as configurações como estão indicadas na figura e selecione "Apply".

Responda:

1. Houve sobreposição das estruturas secundárias?
2. E da estrutura terciária?
3. E dos grupos heme e íons Fe(II)?

Interpretação dos resultados:

Com os resultados obtidos na comparação da estrutura primária e terciária, responda:

1. Todos os aminoácidos de uma proteína (sequência primária) são importantes para a função da mesma, ou alguns (ou vários) podem ser substituídos sem a perda da função?
2. Qual a contribuição das estruturas secundárias e terciárias para a função de uma proteína?
3. No caso da mioglobina e hemoglobina, quais aminoácidos você acha que não podem ser substituídos sem que haja perda da função da proteína?

RELATÓRIO

Após a aula, vocês irão preparar um relatório por grupo, apresentando os resultados obtidos, respondendo as questões do roteiro, e interpretando os resultados obtidos.