



## Exercício 2

O número de bancos de dados biomoleculares está crescendo de forma tão rápida que é impossível se fazer um levantamento que seja minimamente estável com todos os recursos existentes. Aqui, selecionamos alguns bancos de dados para ilustrar os tipos de informação que pode ser obtidas deles.

Escolha uma proteína/gene que você tenha algum conhecimento prévio (se você estiver sem inspiração utilize "Human Lysozyme", que possui o símbolo do gene correspondente "LYZ") e tente extrair as informações pertinentes de **três** bancos de dados sobre esta proteína/gene dentre os diferentes bancos de dados mostrados abaixo. Em seguida, responda às questões abaixo sobre os bancos de dados que você escolheu.

Note que isto é apenas uma pequena amostra das informações que podem ser obtidas através das centenas de bases de dados bio-moleculares atualmente disponíveis.

Sigla	Endereço
Genbank	<a href="http://www.ncbi.nlm.nih.gov/Genbank/">http://www.ncbi.nlm.nih.gov/Genbank/</a>
Uniprot	<a href="http://www.uniprot.org/">http://www.uniprot.org/</a>
PDB	<a href="http://www.rcsb.org/pdb/">http://www.rcsb.org/pdb/</a>
CATH	<a href="http://www.cathdb.info/">http://www.cathdb.info/</a>
GO	<a href="http://www.geneontology.org/">http://www.geneontology.org/</a>
Expression Atlas	<a href="http://www.ebi.ac.uk/gxa/home/">http://www.ebi.ac.uk/gxa/home/</a>
Ensembl	<a href="http://www.ensembl.org/">http://www.ensembl.org/</a>
KO	<a href="http://www.genome.jp/kegg/ko.html">http://www.genome.jp/kegg/ko.html</a>

- 1.- Quais os tipos de informações contidas nos bancos de dados? Como elas são obtidas?
- 2.- As informações contidas nos bancos de dados são curadas? Se sim, explique como é feita a curadoria dos dados?
- 3.- Existe, entre os bancos de dados que você escolheu, alguma relação entre os identificadores ou códigos de acesso?

**Nota:**

Vários bancos de dados em bioinformática usam um sistema em que as entradas podem ser identificadas de duas maneiras diferentes. Basicamente eles são chamados de:

- Identificador - é uma sequência de letras e números que geralmente pode ser interpretada com algum sentido por um ser humano, por exemplo, como uma abreviatura reconhecível da proteína completa ou nome de gene.
- Código (ou número) de acesso - é um código supostamente mais estável composto geralmente de número, possivelmente com algumas letras na frente, o qual identifica a entrada no banco de dados.

4.- O bancos de dados possibilitam realizar buscas logicamente encadeadas de seus conteúdos? Se sim, quais?

5.- Classifique a dificuldade desse exercício: A) Fácil demais, fiz durante a aula e sobrou tempo; B) Adequado, fiz dentro do tempo programado; C) Um pouco puxado, tive que deixar uma parte para fazer em casa; D) Puxadíssimo, fiquei vários dias debruçado sobre um computador.

Envie as respostas desse exercício para o email [srmatiol@ib.usp.br](mailto:srmatiol@ib.usp.br), com assunto da mensagem [BIO-0456] Exercício 2. NÃO ESQUEÇA DE IDENTIFICAR-SE.