

ANÁLISE ESTRUTURAL *IN SILICO*

INTRODUÇÃO

Este estudo visa apresentar as bases de dados repositórias de estruturas (primária, terciária e quaternária) de proteínas de diversos organismos e, a partir delas, demonstrar as ferramentas de bioinformática mais comumente aplicadas ao estudo das proteínas e, principalmente, como processar as informações fornecidas por tais ferramentas.

Inicialmente iremos mostrar que a partir da sequência de um gene, é possível prever a sequência de uma proteína e, a partir dela, chegar ao seu nível de enovelamento máximo. A partir da sequência de aminoácidos, é possível prever alguns parâmetros físico-químicos que preveem seu comportamento *in vitro* e assim, facilitam o seu entendimento e comparação.

Se esta proteína for de interesse da área médica, é possível utilizar sua estrutura para o desenvolvimento de novos fármacos para o tratamento de doenças crônicas e infecciosas, por meio de uma técnica chamada de varredura virtual ou “virtual screening”.

Este material foi estruturado na forma de tutorial de forma a explorar algumas das principais bases de dados repositórias de estruturas e as ferramentas de bioinformáticas mais comumente utilizadas em proteômica.

BASES DE DADOS (REPOSITÓRIO DE SEQUÊNCIAS)

Banco de dados é um conjunto de informações relacionadas entre si, referentes ao mesmo assunto, organizadas prática e racionalmente, para que o usuário levante e **recupere informações**, tire conclusões e tome decisões.

Constitui um grande conjunto de dados persistentes, geralmente associado a um software projetado para atualizar, consultar e recuperar componentes dos dados armazenados no sistema.

Com o avanço da tecnologia, existem cada vez mais sequências e anotações e não é possível determinar a quantidade de informações que ainda será obtida de diversos organismos com o andamento do projeto genoma.

Isso torna fundamental o uso de um banco de dados bem estruturado que permita o armazenamento, o acesso e o processamento destas informações de forma simples e eficiente.

- 1) Bancos de dados **Primários**

São repositórios que armazenam informações obtidas diretamente de técnicas experimentais como sequenciamento, difração de raios-X, NMR. Seu conteúdo é submetido por pesquisadores (autores), os quais são responsáveis e tem total controle sobre o conteúdo submetido. São exemplos de bancos primários, o **GenBank**, **PIR**, **EMBL**, **DDJB**, **SNP**, **GEO**, **PDB**.

Exercício 01:

Explorando os bancos de dados primários. Acesse: <http://www.ncbi.nlm.nih.gov/genbank/>

- a) Buscar no GenBank a sequência do RNAm que codifica para a proteína Kin17 humana no formato FASTA:

```
>gi|3850703|emb|AJ005273.1| Homo sapiens mRNA for Kin17 protein
CTAGAATTCAGCGCCGCTGAATTCTAGAACTGGGGTCCAGAAAGTGATCGCTGCCGTGGTCGCCATGGG
GAAGTCGGATTTTCTTACTCCCAAGGCTATCGCCAACAGGATCAAGTCCAAGGGGCTGCAGAAGCTACGC
TGGTATTGCCAGATGTGCCAGAAGCAGTGCCGGGACGAGAATGGCTTTAAGTGTCAATTGTATGTCCGAAT
CTCATCAGAGCAACTATTGCTGGCTTCAGAAAACTCAGCAGTTTATGGATTATTTTTCAGAGGAATT
CCGAAATGACTTTCTAGAACTTCTCAGGAGACGCTTTGGCACTAAAAGGGTCCACAACAACATTGTCTAC
AACGAATACATCAGCCACCGAGAGCACATCCACATGAATGCCACTCAGTGGGAACTCTGACTGATTTTA
CTAAGTGGCTGGGCAGAGAAGGCTTGTGCAAAGTGGACGAGACACCAAAAAGGCTGGTATATTCAGTACAT
AGACAGGGACCCAGAAACTATCCGCCGGCAACTGGAAGTGGAGAAAAGAAAAGCAGGACCTTGATGAT
GAAGAAAAACTGCCAAATTTATTGAAGAGCAAGTGAGAAGAGGCCGGAAGGGAAGAACAGGAGGTCCTC
TACTTTTACGGAAATTAAGCAGAGAAAATGATGAAGAGAAAAGTCAAGTTTAAATTTGAGTAAAGGAGCATG
TAGCTCATCCGGAGCAACATCTTCCAAGTCAAGTACTCTGGGACCGAGTGCCTGAAGACGATAGGAAGT
TCAGCATCAGTGAACGAAAAGAATCTTCCAGAGCTCAACTCAGTCTAAAGAAAAGAAGAAAAGAAAT
CTGCACCTGGATGAAATCATGGAGATTGAAGAGGAAAAGAAAAGAACTGCCCGAACAGACTACTGGCTACA
GCCTGAAATTTATTGTGAAAATATAACCAAGAACTGGGAGAGAAAATATCATAAGAAAAGGCTATTGTT
AAGGAAGTAATTGACAAATATACAGCTGTTGTGAAGATGATTGATTCTGGAGACAAGCTGAAACTTGACC
AGACTCATTTAGAGACAGTAATTCAGCACCAGGAAAAGAAATTCAGTTTTAAATGGAGGCTACAGAGG
AAATGAAGGTACCTAGAAATCCATCAATGAGAAGACTTTTTTCAGTACTATCGTCATTGAAACTGGCCCT
TTAAAAGGACGCAGAGTTGAAGGAATCAATATGAAGACATTTCTAAACTTGCCTGAGTTTGAAAATTTG
TTAACAATACATTTAAATCTTAAAGCATCAAATGGTGTTCGCCAAGGCATTATGAGACTTACTGTGTT
AGGGTATATTTTGTATAAAAACAAACAGGTTTTTAAAATATTACTGTATAGTTGTTTCAGCTAAACTT
TGAGAAGAATTTAATATGTCTCATGAGGTATCAAATGTAAATTTGTCTTGTATTTTTGTTTCCT
TTGTAATTTACTTGATGAGTTTATATCTTCATTAAGAATGTTATTATAAAAAAAA
```

PS: Em bioinformática, o formato FASTA é um formato baseado em texto para representar tanto sequências de nucleótidos quanto sequências de aminoácidos, no qual os nucleotídeos ou aminoácidos são representados usando códigos de uma única letra. A linha de descrição se distingue a partir da sequência dos dados por um símbolo maior-que (">") na primeira coluna. A palavra que segue o símbolo ">" é o identificador da sequência, e o resto da linha é a descrição (ambos são opcionais). Não deve haver nenhum espaço entre o ">" e a primeira letra do identificador.

1) Bancos de dados **Secundários**

São repositórios construídos a partir do banco de dados primário. Seu conteúdo e informações são resultantes da análise e processamento dos primários. Seu conteúdo é controlado por curadores como o European Bioinformatics Institute (EBI) e o Swiss Institute of Bioinformatics (SIB).

São exemplos de bancos de dados secundários: **Refseq, RefSNP, UniGene, NCBI Protein, Structure, Conserved Domain, SwissProt, Pfam, etc.**

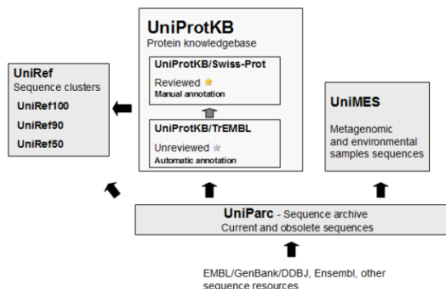
2) Explorando os bancos de dados secundários. Exemplos:

SWISS-PROT e TrEMBL



- SWISS-PROT foi criado em 1986 pelo Departamento de Bioquímica Médica da Universidade de Genebra e EMBL.
- Atualmente é mantido pelo Swiss Institute of Bioinformatics (SIB) e EBI/EMBL.
- Este banco mantém um alto nível de anotações, como a descrição e a função da proteína, estrutura dos seus domínios, modificações pós-tradução, além de ter uma estrutura que facilita o acesso computacional a diferentes campos de informações.

Acesse: <http://www.uniprot.org/>



É uma colaboração entre o Instituto Europeu de Bioinformática (EBI), o Instituto Suíço de Bioinformática (SIB) e o Protein Information Resource (PIR).

As bases de dados do UniProt são Knowledgebase UniProt (**UniProtKB**), o UniProt Reference Clusters (**UniRef**), e o UniProt Archive (**UniParc**).

Exercício 2:

- 1) Utilize a mesma sequência do RNAm utilizada na busca anterior e faça uma análise da região codificante através do Expasy (<http://web.expasy.org/translate/>)
 - a) Cole a sequência do RNAm no espaço em branco;
 - b) Clique no botão e aguarde o resultado.
 - c) Analisa as ORFs e clique sobre a que coce acredita ser a que corresponde a região codificante da Kin17
 - d) Clique na primeira Metionina e veja quantos aminoácidos a proteína possui.
 - e) Vá até o Uniprot e confirme sua análise procurando pela sequência da Kin17 humana.

Ferramentas Para Alinhamento de Sequências

Alinhamento de sequências é uma forma de organizar sequências primária de DNA, RNA ou proteína para identificar regiões similares que possam ser consequência de relações funcionais, estruturais ou evolucionárias entre elas. Sequências alinhadas de nucleotídeos ou resíduos de aminoácidos são representadas tipicamente como linhas de uma matriz. Espaçamentos (*gaps*) podem ser inseridos entre os resíduos para que caracteres semelhantes (por algum critério) sejam alinhados em colunas sucessivas.

```

AAB24882      TYHMCQFHC RYVNNHSGE KLYECNERS KAFSCPSHLQCHKRRQ IGEKTHEHNQCGKAFPT 60
AAB24881      -----YECNQC GKAF AQHSSLKCHYRTH IGEKPYECNQC GKAFSK 40
                ****: .***: * **:* * :****.:* *****..

AAB24882      PSHLQYHER THTG EKPYE CHQC GQAFK KCSLLQRHKR THTG EKPYE -CNQC GKAF AQ- 116
AAB24881      HSHLQCHK R THTG EKPYE CNQC GKAF SQHGLLQRHKR THTG EKPYMNVINMVKPLHNS 98
                **** *:*****:***:**.: .*****: : *.: :

```

Figura 1 Exemplo de alinhamento entre duas sequências, produzido pelo programa ClustalW entre duas proteínas dedo-de-zinco humanas (human zinc finger proteins) identificadas por seus números de acesso no GenBank

Se duas sequências em um alinhamento compartilham de um ancestral comum, discordâncias (*mismatches*) podem ser interpretados como mutações pontuais e os espaços (*gaps*) como inserções ou deleções introduzidas em uma ou ambas as sequencias desde quando estas divergiram no tempo.

Abordagens computacionais para o alinhamento de sequências dividem-se, em geral, em duas categorias: alinhamentos globais e alinhamentos locais. Calcular um alinhamento global é uma forma de otimização global que "força" o alinhamento a cobrir todo o comprimento de todas as sequencias interrogadas (*query*). Por outro lado, os alinhamentos locais identificam regiões de similaridade dentro de sequencias longas que são geralmente bastante divergentes em um todo. Os alinhamentos locais são frequentemente preferíveis, mas podem ser difíceis de calcular por causa do problema adicional de identificar regiões internas de similaridade. Uma grande variedade de algoritmos existem para abordar o problema de alinhamento de sequencias, sendo os mais conhecidos os baseados em programação dinâmica, mais lentos porém teoricamente otimizadores, ou baseados em heurística, mais eficientes/rápidos mas sem prova formal de obtenção de solução ótima.

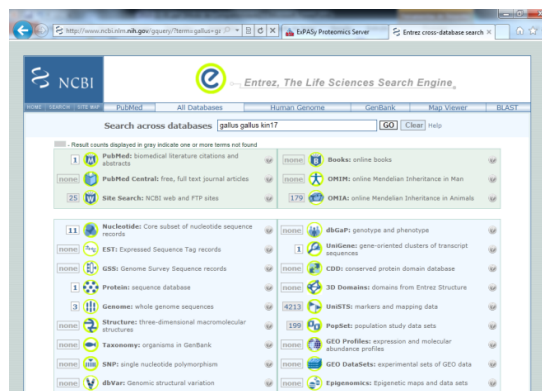
Exercício 03:

- 1) Busque nas bases de dados Uniprot as sequências da proteína humana Kin17 ou outras de seu interesse.
- 2) Copie e cole as sequências no formato FASTA usando o WordPad.
- 3) Alinhe as sequências usando o UniprotKB/**Alignment**
- 4) Explore as similaridades entre elas selecionando as propriedades dos aminoácidos.

Acesso: <http://www.ncbi.nlm.nih.gov/>



O National Center for Biotechnology Information (NCBI) foi fundado em 1988 e é parte da United States National Library of Medicine (NLM), uma filial do National Institutes of Health (NIH) e está localizado em Bethesda, Maryland, USA



BLAST (Basic Local Alignment Search Tool)

BLAST (sigla em inglês que significa: Basic Local Alignment Search Tool), é um algoritmo para comparar informações de seqüências biológicas primárias, tais como seqüências de aminoácidos de diferentes proteínas ou nucleotídeos de seqüências de DNA. Uma pesquisa BLAST permite que um investigador compare uma seqüência fornecida em uma consulta com uma biblioteca ou base de dados de seqüências e identificar as bibliotecas de seqüências que se assemelham à seqüência consultada e que estejam acima de um certo

Diferentes tipos de BLAST

BLASTN – Pesquisa bancos de dados de nucleotídeos usando uma seqüência de nucleotídeos em questão.

Algoritmos: blastn, megablast, discontinuous megablast

BLASTP – Pesquisa bancos de dados de proteínas usando uma Seqüência de proteína em questão.

Algorithms: blastp, psi-blast, phi-blast

BLASTX – Pesquisa banco de dados de proteínas usando uma seqüência de nucleotídeo traduzida em questão

TBLASTN – Pesquisa um banco de dados de nucleotídeos usando uma proteína em questão.

TBLASTX – Pesquisa bancos de dados de nucleotídeos traduzidos Usando um nucleotídeo traduzido em questão.

grau de semelhança.

Numa situação hipotética, após descobrir um gene anteriormente desconhecido em um camundongo, um cientista poderia tipicamente elaborar uma pesquisa no BLAST do genoma humano para verificar se existem seres humanos portadores de um gene semelhante.

Exercício 04:

- 1) A partir da sequência de aminoácidos da Kin 17 humana, fazer uma busca por proteínas similares (moldes), cujas estruturas estejam depositadas no Protein Data Bank.

1º passo: Selecione a opção *Protein blast*

The screenshot shows the NCBI BLAST homepage. At the top, there are navigation links: Home, Recent Results, Saved Strategies, and Help. Below this, there's a section for 'BLAST Assembled RefSeq Genomes' with a list of species: Human, Mouse, Rat, Arabidopsis thaliana, Oryza sativa, Bos taurus, Danio rerio, Drosophila melanogaster, Gallus gallus, Pan troglodytes, Microbes, and Apis mellifera. Underneath, there's a 'Basic BLAST' section with three options: 'nucleotide blast', 'protein blast', and 'blastx'. A red arrow points to the 'protein blast' option, which is described as 'Search protein database using a protein query'.

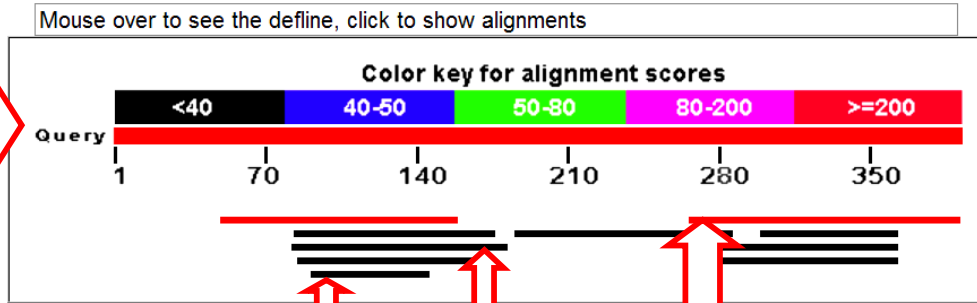
2º Passo: Cole a sequência no formato FASTA e selecione em seguida a opção “Protein Data Bank proteins (PDB)” em “Database” e depois clique em “BLAST” e aguarde o resultado.

The screenshot shows the 'Standard Protein BLAST' interface. The 'protein blast' tab is selected. In the 'Enter Query Sequence' section, there's a text input field for 'Enter accession number(s), gi(s), or FASTA sequence(s)'. A red arrow points to this field. Below it, there are 'From' and 'To' dropdown menus. In the 'Choose Search Set' section, the 'Database' dropdown menu is set to 'Non-redundant protein sequences (nr)'. A red arrow points to this dropdown menu. Other options include 'Organism', 'Exclude', and 'Entrez Query'.

3º Passo: Análise o resultado.

Essa barra em vermelho representa a sequência que foi usada na pesquisa. Esta sequência é chamada de “query”.

Distribution of 10 Blast Hits on the Query Sequence



Cada uma dessas barras coloridas indicam a região onde a sequência do banco de dados bate com a sequência query

Descrição da proteína depositada no PDB por ordem de query coverage

Sequences producing significant alignments:

Select: [All](#) [None](#) Selected: 0

Alignments [Download](#) [Statistics](#) [Distance tree of results](#) [Multiple alignment](#)

| Description | Max score | Total score | Query cover | E value | Max ident | Accession |
|--|-----------|-------------|-------------|---------|-----------|------------------------|
| Chain A, High Resolution Crystal Structure Of The Human Kin17 C-Terminal Domain Containing A Kow Motif Kin17 | 254 | 254 | 32% | 2e-83 | 100% | 2CKK_A |
| Chain A, Solution Structure Of The Region 51-160 Of Human Kin17 Reveals A Winged Helix Fold | 238 | 238 | 27% | 2e-77 | 100% | 2V1N_A |
| Chain A, Crystal Structure Of The Human Parp-1 Dna Binding Domain In Complex With Dna ->pdb14AV1/B Chain B, Crystal Structure Of The Human Parp-1 Dna | 33.9 | 33.9 | 25% | 0.14 | 28% | 4AV1_A |
| Chain A, Solution Structure Of Max B-Hlh-Lz ->pdb1R05/B Chain B, Solution Structure Of Max B-Hlh-Lz | 31.2 | 31.2 | 25% | 0.28 | 29% | 1R05_A |
| Chain A, Recognition By Max Of Its Cognate Dna Through A Dimeric B/hlh/z Domain | 27.7 | 27.7 | 23% | 4.6 | 27% | 1AN2_A |
| Chain A, Inter-Subunit Interaction And Quaternary Rearrangement Defined By The Central Stalk Of Prokaryotic V1-Atpase ->pdb3A5C/B Chain B, Inter-Subunit Int | 29.3 | 29.3 | 20% | 6.0 | 31% | 3A5C_A |
| Chain A, Crystal Structure Of V1-atpase At 3.9 Angstrom Resolution ->pdb3W3A/B Chain B, Crystal Structure Of V1-atpase At 3.9 Angstrom Resolution ->pdb3W3 | 29.3 | 29.3 | 20% | 6.1 | 31% | 3W3A_A |
| Chain A, Crystal Structure Of The A3b3 Complex From V-atpase ->pdb13GQB/C Chain C, Crystal Structure Of The A3b3 Complex From V-atpase | 29.3 | 29.3 | 16% | 6.6 | 34% | 3GOB_A |
| Chain B, Crystal Structure Of Mvc-Max Recognizing Dna ->pdb1NKP/E Chain E, Crystal Structure Of Mvc-Max Recognizing Dna | 26.9 | 26.9 | 23% | 7.5 | 28% | 1NKP_B |
| Chain A, X-Ray Crystal Structure Of Protein Mm0500 From Methanosarcina Mazei, Northeast Structural Genomics Consortium Target Mar10 ->pdb1XUV/B Chain | 28.1 | 28.1 | 13% | 9.6 | 31% | 1XUV_A |

- **Max score:** Indica o quão bem sua sequência se encaixa;
- **Total Score:** inclui porções não contíguas da sequência do molde que corresponde a sequência query;
- **Query coverage:** Indica a fração da sequência query que corresponde a sequência do molde,
- **E-value:** Score que determina a probabilidade de um falso positivo Quanto menor o escore, mais significativo é o alinhamento;
- **Max ident:** Refere-se a semelhança entre as sequências de aminoácidos das proteínas.

Exercício 05:

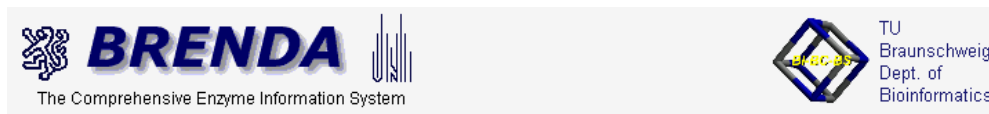
Objetivo: Identificar a localização celular da proteína.

- 1) Usando as ferramentas expasy/**PSORT** (para bactérias e archaea) ou **TargetP** (para outros organismos), insira a sequência da proteína de interesse no formato Fasta.
- 2) Analise os parâmetros fornecidos e discuta com seus colegas o significado de cada um deles.

Exercício 06:

Objetivo: Prever o padrão de enovelamento (folding) da estrutura secundária a partir da sequência de aminoácidos.

- 1) Usando as ferramentas expasy/**PSIPRED** insira a sequência da proteína de interesse no formato Fasta. Interprete os resultados.



BRENDA é o principal banco de dados de dados funcionais de enzimas, disponível para a comunidade científica.

Possui um acervo com mais de 5.000 enzimas diferentes.

É mantido e desenvolvido pelo **Instituto de Bioquímica e Bioinformática** da Universidade Técnica de Braunschweig, na Alemanha.

É um repositório de dados sobre a função da enzima são extraídos diretamente da literatura primária por cientistas formados em Biologia ou Química.

As verificações de forma e consistência são feitas por programas de computador. Cada conjunto de dados sobre uma enzima classificada é verificado manualmente por pelo menos um biólogo e um químico.

Acesse: <http://www.brenda-enzymes.org>

The screenshot shows the BRENDA website interface. At the top, there's a navigation bar with the BRENDA logo and the text 'The Comprehensive Enzyme Information System'. Below this, there's a search bar with tabs for 'EC-Number', 'Enzyme Name', 'Organism', 'Protein', 'Full text', and 'Advanced Search'. A search input field and a 'Search' button are present, along with a 'Display 10 entries' option. Below the search bar, there's a section titled 'How to cite BRENDA?' which contains a table with three columns: 'Nomenclature', 'Reaction & Specificity', and 'Functional Parameters'. The table lists various enzyme-related terms and parameters.

| Nomenclature | Reaction & Specificity | Functional Parameters |
|--------------------------|---------------------------------|-------------------------------------|
| Enzyme Names | Pathway | Km Value |
| EC Number | Catalysed Reaction | Ki Value |
| Common/ Recommended Name | Reaction Type | IC50 Value |
| Systematic Name | Natural Substrates and Products | pI Value |
| Synonyms | Substrates and Products | Turnover Number |
| CAS Registry Number | Substrates | Specific Activity |
| | Natural Substrate | pH Optimum |
| | Products | pH Range |
| | Natural Product | Temperature Optimum |
| | Inhibitors | Temperature Range |
| | Cofactors | |
| | Metals/Ions | Organism-related information |
| | Activating Compounds | Organism |
| | Ligands | Source Tissue |
| | Ligand Views NEW | Localization |
| | | Protein-Specific Search |

Exercício 07:

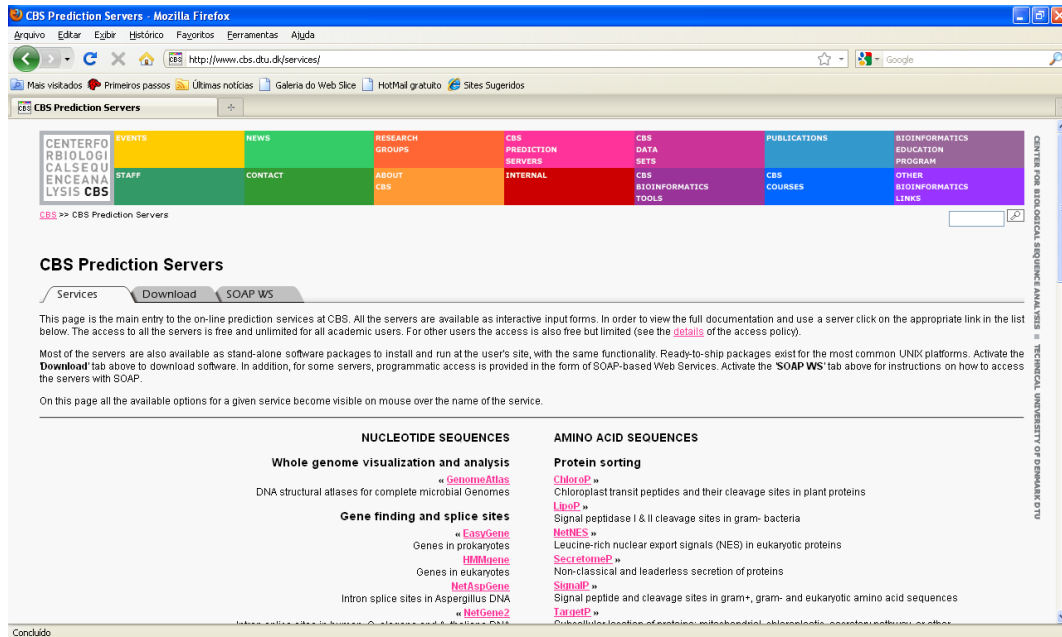
- 1) Buscar o *E.C.*, o *K_M*, o *pH ótimo* e *temperatura ideal* da catalase bovina (*Bos taurus*) e de fungo (*Aspergillus. agaricus*), na sua forma livre.
- 2) Busque estas informações para outras enzimas de seu interesse.

FERRAMENTAS DE BIOINFORMÁTICA NÃO VINCULADAS A REPOSITÓRIOS:

Acesse: <http://www.cbs.dtu.dk/services/>



O **Center for Biological Sequence Analysis (CBS)** na Universidade Técnica da Dinamarca foi fundado em 1993 e realiza pesquisas básicas no campo da bioinformática e de sistemas biológicos.



Exercício 08:

Objetivo: Predizer o Papel celular, a Classe enzimática e a Categoria ontologica do gene a partir de uma seqüência de aminoácidos.

Método: Utilizar o CBS/ProtFun 2.2 server para produzir predições *ab initio* da função da proteína a partir de sua sequencia.

- 1) Usando a ferramenta **ProtFun**, insira a seqüência da proteína de interesse no formato Fasta.
- 2) Analise os parâmetros fornecidos e discuta com seus colegas o significado de cada um deles a partir da probabilidade encontrada.

Exercício 09:

Objetivo: Utilizar a ferramenta CBS/TMHMM para predizer a posição da proteína em relação à membrana.

Tutorial 2: Modelagem molecular por homologia.

A modelagem de uma proteína (proteína-problema) pelo método da homologia baseia-se no conceito de evolução molecular. Isto é, parte-se do princípio de que a semelhança entre as estruturas primárias desta proteína e de proteínas homólogas de estruturas tridimensionais conhecidas (proteínas-molde) implica em similaridade estrutural entre elas. Os métodos correntes de modelagem de proteínas por homologia implicam basicamente em quatro passos sucessivos (Santos Filho & Alencastro, 2003):

- a) identificação e seleção de proteínas-molde;
- b) alinhamento das sequências de resíduos;
- c) construção das coordenadas do modelo;
- d) validação.

Exercício 11:

Objetivo: Predizer a estrutura 3D de uma proteína a partir de sua sequência de aminoácidos.

- 1) Utilizando o Swiss Model ou o CBS/CPHmodels, insira a sequência de aminoácidos da proteína de seu interesse para ser modelada.

Protein structure from sequence: distance constraints

Exercício 12: Validação do modelo:

Problematização: O modelo gerado por homologia pode conter erros estereoquímicos, e, portanto, precisa ser validado antes de ser analisado e utilizado em futuros estudos.

Objetivo: Utilizar os servidores de validação para checar a qualidade estereoquímica do modelo:

Lista de servidores de validação: <http://xray.bmc.uu.se/embo2001/modval/links.html>

Acesse VADAR: <http://redpoll.pharmacy.ualberta.ca/vadar/>

OUTROS REPOSITÓRIOS DE DADOS RELACIONADOS A PROTEÔMICA:

BANCO DE DADOS DE ESTRUTURAS 3D

O **Protein Data Bank (PDB)** é o único repositório de informações em todo o mundo sobre as estruturas 3D de macromoléculas biológicas, incluindo as proteínas e ácidos nucléicos. A seguir, veremos um breve histórico desta importante base de dados:

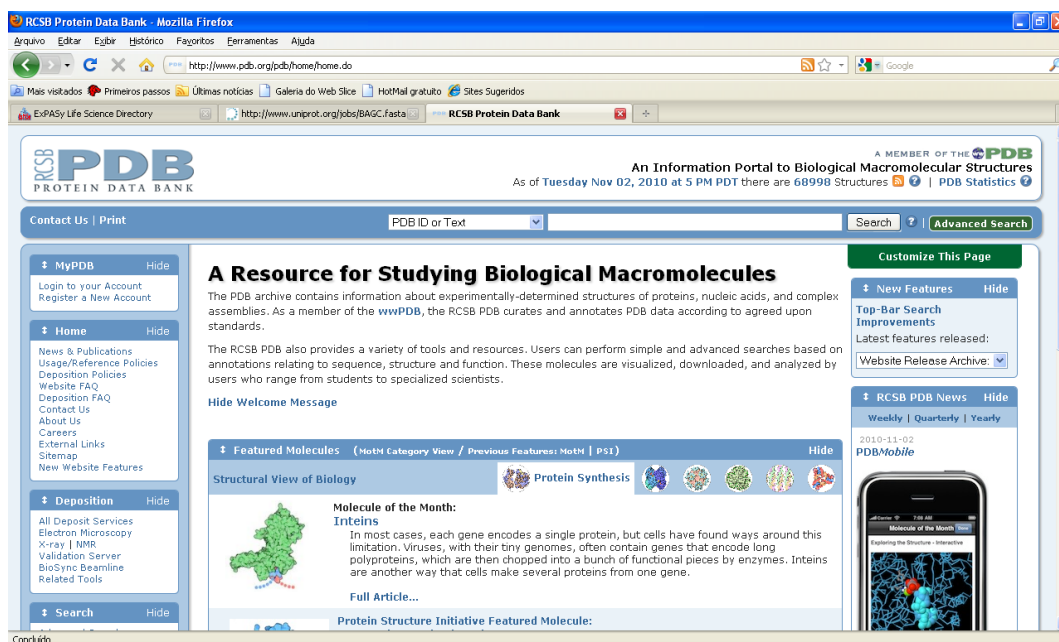
1971 – Início do PDB no Brookhaven National Laboratory com 7 estruturas.

1998 – O Research Collaboratory for Structural Bioinformatics (RCSB) se tornou responsável pela manutenção do PDB.

2010 – o RCSB PDB é membro do [wwPDB](#), um esforço colaborativo do [PDBe](#) (UK), [PDBj](#) (Japão), and [BMRB](#) (USA) para assegurar que o arquivo PDB seja global e uniforme.

Este banco de dados, somente aceita depósito de estruturas determinadas experimentalmente por Cristalografia de raios-X, NMR ou Microscopia eletrônica.

Acesse: <http://www.pdb.org>



The screenshot shows the RCSB Protein Data Bank website. At the top, it says "RCSB PDB PROTEIN DATA BANK" and "A MEMBER OF THE PDB An Information Portal to Biological Macromolecular Structures". Below this, there is a search bar with "PDB ID or Text" and a "Search" button. The main content area is titled "A Resource for Studying Biological Macromolecules" and includes a "Molecule of the Month" section for "Inteins". The page also has a sidebar with navigation links like "MyPDB", "Home", "Deposition", and "Search".

Exercício 12:

Procure a estrutura de sua proteína (ou relacionadas) utilizando a ferramenta **Advanced Search, Macromolecule name** (HMGB1) ou baixe uma estrutura pelo identificador (pdbid): 1KD2 e 1HVC. Explore a estrutura utilizando o visualizador Jmol.

REFERÊNCIAS:

HOCK R, FURUSAWA T, UEDA T, BUSTIN M. HMG chromosomal proteins in development and disease. *Trends Cell Biol.* 2007 Feb;17(2):72-9. Epub 2006 Dec 13. Review.

LEACH AR. *Molecular modeling: Principles and applications.* 2nd ed. Prentice Hall, 2001, 784 p.

NELSON DL, COX MM. *Lehninger princípios de bioquímica.* São Paulo: Sarvier, 2007 1202 p.

PAULL TT, HAYKINSON MJ, JOHNSON RC. The nonspecific DNA-binding and -bending proteins HMG1 and HMG2 promote the assembly of complex nucleoprotein structures. *Genes Dev.* 1993 Aug;7(8):1521-34.

REHM H. *Protein Biochemistry and proteomics.* London:Elsevier, 2006, 236 p.

SANTOS FILHO AO, ALENCASTRO RB. Modelagem de proteínas por homologia. *Quím. Nova*, v. 26 n. 2, 2003.

SHEFLIN LG, SPAULDING SW. High mobility group protein 1 preferentially conserves torsion in negatively supercoiled DNA. *Biochemistry.* 1989 Jun 27;28(13):5658-64.

THOMAS JO, TRAVERS AA. HMG1 and 2, and related 'architectural' DNA-binding proteins. *Trends Biochem Sci.* 2001 Mar;26(3):167-74. Review. Erratum in: *Trends Biochem Sci* 2001 Apr;26(4):219.

UGRINOVA I, MITKOVA E, MOSKALENKO C, PASHEV I, PASHEVA E. DNA bending versus DNA end joining activity of HMGB1 protein is modulated in vitro by acetylation. *Biochemistry.* 2007 Feb 27;46(8):2111-7. Epub 2007 Feb 1.

YANG H, WANG H, TRACEY KJ. HMG-1 rediscovered as a cytokine. *Shock.* 2001 Apr;15(4):247-53. Review.

YE SQ. *Bioinformatics a practical approach.* London:Chapman & Hall/CRC, 2008, 618.