

BIOINFORMATIKA V PRAXI – CVIČENÍ 7

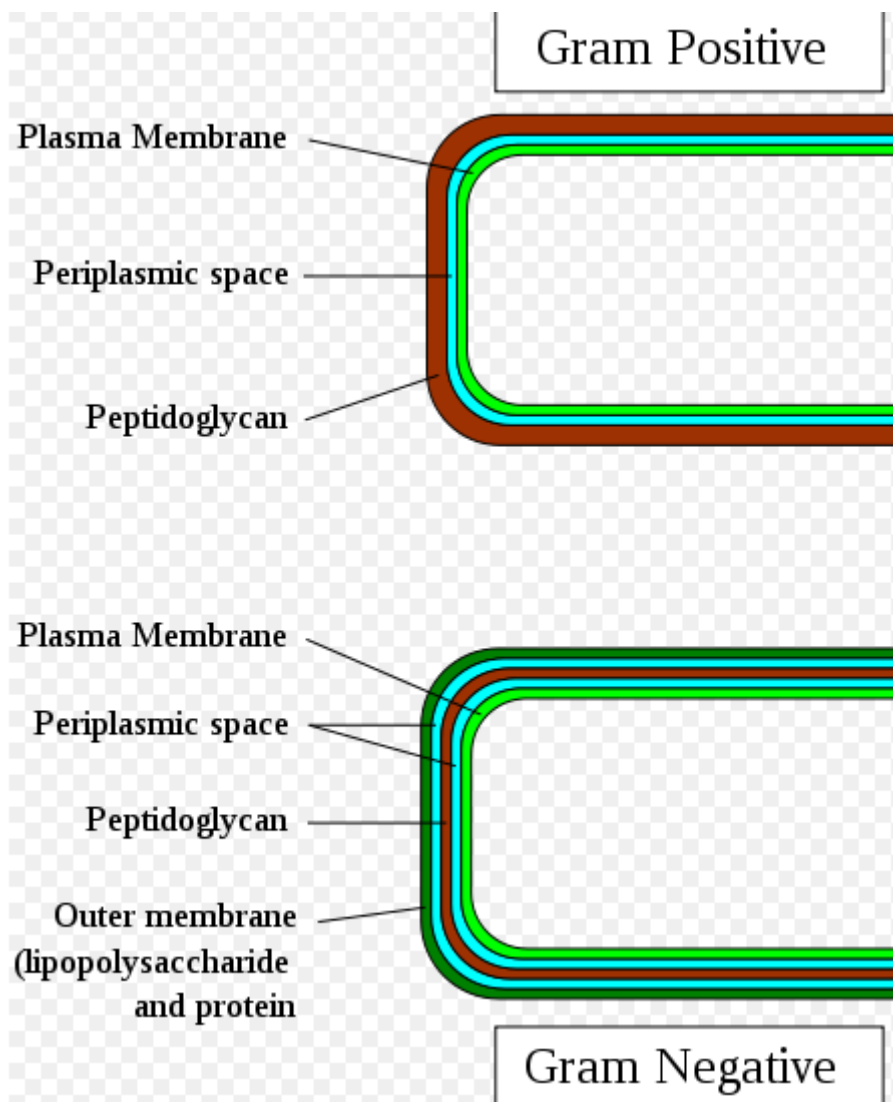
PREDIKCE VLASTNOSTÍ PROTEINŮ DOKONČENÍ

STUDIJNÍ MATERIÁLY

The ExPASy (Expert Protein Analysis System) proteomics server of the Swiss Institute of Bioinformatics (SIB), <http://www.expasy.org/>.

PREDIKCE LOKALIZACE PROTEINU V BUŇCE

Predikce lokalizace proteinu v buňce může být cenným nástrojem pro určení FUNKCE proteinu. Z výskytu proteinů v buňce (cytoplasmatické, membránové, extracelulární, v buněčné stěně...) můžeme například u patogenních mikroorganismů odhadnout, zda se jedná o protein důležitý pro virulenci.



ÚKOL 1

Pomocí programu **PSORT** na ExPASy (Proteomics tools – Topology prediction) určete lokalizaci následujícího proteinu v buňce:

Protein z grampozitivní bakterie *Micrococcus luteus*

MEINGLTWGLTIALIVGLLAFDYFAHVRKAHTPTIKEAAVWSGVYVGLALVFGLVFFAFGDTQHAVEYYTGYLE
KALSVDNLFVFLVIMASFRVPREYQQKVLVLFGITFALISRTLFIILLGAAVIAAWSDFYLFGLFLLIIAGGQLKG
EMSGDAEGAQDEADNVMVRLVKRFLPASDQFDGERLFTTVDGKRLMTPMLLVMIAGATDILFAFDSIPAI FGVT
QEAYIVFTATAFSLMGLRQLYFLIDGLLDRLVYLAYGLSALLAFIGVKLILHALHENNLFPINGGENVPVAEIP
NLSLVVVVVILAITVLVSLYSPKQALRALQNAEKYSYRYSKLAEDADPAERERAAAALMDRWTARAEALDQRWRD
QLLEHKDAWSAII RTAHETRFADPRDDDARGVSEQIVRQDGPTV

ÚKOL 2

Pomocí programu **PSORT** určete lokalizaci následujících proteinů v buňce:

Protein 1 z neznámé gramnegativní bakterie

MTKYALVGDVGGTNARLALCDIASGEISQAKTYSGLDYPSEAVVRVYLDEHGVSVEDGCIACIPITGDWVAMT
NHTWAFSIAEMKKNLGFSHLEIINDFTAVSMAIPMLKKEHLIQFGGGE PVDGKPIAVYGAGTGLGVAHLVHVDKR
WISLPGE GGHVDFAPNSEEEAMILEILRAEIGHVSAERVLSPGGLVNLRYAIVKSDNRLPENLRPKDITERALAD
SCIDCRRALSFCVIMGRFGGDLALTMGTFGGVYIAGGIVPRFLEFFKASGFRGGFEDKGRFKDYVHGIPVYLIV
HDNPGLLGSGAHLRQTLGHIL

Protein 2 z neznámé bakterie

LVIVDAVTLLSAYPEASRDPAAPTVIDGRHLYVVS PGDAAQLGHNDSRLFTGLSPGDQLHLRETALALRAEVSVL
FIRFALKDAGIVAPIELEVRDAATAVPDADDLLHPSRPLKDHYWRSDVLAAGATTCTADFAVCDRDTVSGYFR
WETSIEIAGSQPDTKQPGFKPSS

Protein 3 z *Aspergillus fumigatus*

MKPQTAAFLLSLLGSTLAAPIQHADKGS DVKPTPTGRGAPGGFFTFGFP SGVPSGLPSGFPGGPVPGGFGGDGPNG
PIPSGPVPTGAAPSGFSPFGTGPAPSGAPQGEES SFGGQGVQARSPQDFEDSGAAPSGAIPSGAIP TGA VP SG
APNGFGGFGQGGHGGPGGPGEEGSGPSPTGAVPSGAI PSGAAPSGAVGGFDGFGQASENSGNAQFQSSGTS PF GA
SHSGSASGHQGRHGGDRHQHGNNGSGAIPSGAAPSGAAPSGAAGGFPFGFGQGGEGSGPSPTGAVPSGAAGFGGQ
GHGQQGGSFPTGVAPSDVPSAQPTA

ÚKOL 3

Pomocí programu **PSORT** určete lokalizaci následujícího proteinu v buňce:

Protein z gramnegativní bakterie:

MKKLAIMAAASMVFAVSSAHATKYALVGDVGGTNARLALCDIASGEISQAKTYSGLDYPSEAVVRVYLDEHGVS
VEDGCIACIPITGDWVAMTNHTWAFSIAEMKKNLGFSHLEIINDFTAVSMAIPMLKKEHLIQFGGGE PVDGKPI
AVYGAGTGLGVAHLVHVDKRWISLPGE GGHVDFAPNSEEEAMILEILRAEIGHVSAERVLSPGGLVNLRYAIVKS
DNRLPENLRPKDITERALADSCIDCRRALSFCVIMGRFGGDLALTMGTFGGVYIAGGIVPRFLEFFKASGFRGG
FEDKGRFKDYVHGIPVYLIVHDNPGLLGSGAHLRQTLGHIL

Pomocí Signal Sequence Database (<http://www.signalpeptide.de/index.php> - Signal Peptide Website) identifikujte v proteinu signální sekvenci.

SEKUNDÁRNÍ DATABÁZE PROTEINŮ

Sekundární databáze obsahují informace odvozené z primárních databází ve formě charakteristických vzorů sekvencí, tj. funkčních nebo strukturních motivů získaných srovnáním primárních dat (sekvencí). Sekundární databáze jsou vhodné pro vyhledávání „vzoru“ charakteristického pro určitou skupinu proteinů, tj. pro určení funkce proteinů.

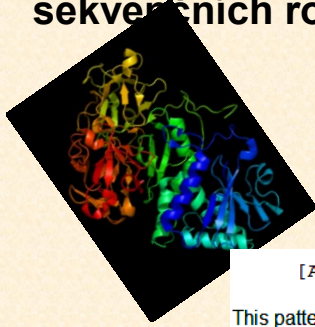
Sekundární databáze proteinů

- **Databáze mohou obsahovat:**

Proteinové **DOMÉNY** odvozené ze známých struktur

Proteinové sekvence seřazené do **SEKVENČNÍCH RODIN**

CHARAKTERISTICKÉ MOTIVY odvozené z těchto sekvencí rodin.



```
LPPNTAFKAIIFYANAADRODLKLFIDD  
IPPNTDFRAIFFANAAEQOHKLFIGD  
LPPHIKFGVTALTHAANDOTIDIYIDD  
LPPNIAFGVTALVNSSAPOTIEVFVDD
```

```
[AC]-x-V-x(4)-{ED}.
```

This pattern is translated as: [Ala or Cys]-any-Val-any-any-any-any-{any but Glu or Asp}

Sekundární databáze proteinů (**PROSITE**, **Pfam**, **PRINTS**, **ProDom**, **SMART**, **TIGRFAMS**) jsou v současné době sdruženy do integrované klasifikační databáze proteinů **InterPro**.

ÚKOL 4

Vyberte si dva z předcházejících proteinů a s využitím nástroje InterProScan (<http://www.ebi.ac.uk/Tools/pfa/iprscan/>) je analyzujte pomocí databáze InterPro. Jaké domény/motivy proteiny obsahují?

PREDIKCE SEKUNDÁRNÍ STRUKTURY PROTEINU

Secondary structure prediction

- [AGADIR](#) - An algorithm to predict the helical content of peptides
- [APSSP](#) - Advanced Protein Secondary Structure Prediction Server
- [CFSSP](#) - Chou & Fasman Secondary Structure Prediction Server
- [GOR](#) - Garnier et al, 1996
- [HNN](#) - Hierarchical Neural Network method (Guermeur, 1997)
- [HTMSRAP](#) - Helical TransMembrane Segment Rotational Angle Prediction
- [Jpred](#) - A consensus method for protein secondary structure prediction at University of Dundee
- [JUFO](#) - Protein secondary structure prediction from sequence (neural network)
- [NetSurfP](#) - Protein Surface Accessibility and Secondary Structure Predictions **new**
- [NetTurnP](#) - Prediction of Beta-turn regions in protein sequences **new**
- [nnPredict](#) - University of California at San Francisco (UCSF)
- [Porter](#) - University College Dublin
- [PredictProtein](#) - PHDsec, PHDacc, PHDhtm, PHDtopology, PHDthreeder, MaxHom, EvalSec from Columbia University
- [Prof](#) - Cascaded Multiple Classifiers for Secondary Structure Prediction
- [PSA](#) - BioMolecular Engineering Research Center (BMERC) / Boston
- [PSIpred](#) - Various protein structure prediction methods at Bloomsbury Centre for Bioinformatics
- [SOPMA](#) - Geourjon and Deléage, 1995
- [Scratch Protein Predictor](#) **new**
- [DLP-SVM](#) - Domain linker prediction using SVM at Tokyo University of Agriculture and Technology


ÚKOL 5

Vyberte si jeden z předcházejících proteinů a predikujte sekundární strukturu tohoto proteinu. Využijte programy **CFSSP** a **Jpred**. Pokud to půjde, vyzkoušejte i jiné (**HNN**, **PSIpred**) a výsledky porovnejte...

PREDIKCE TERCIÁRNÍ STRUKTURY PROTEINU

Tertiary structure prediction

Homology modeling

- [SWISS-MODEL](#)  - An automated knowledge-based protein modelling server
- [3DJigsaw](#) - Three-dimensional models for proteins based on homologues of known structure
- [CPHmodels](#) - Automated neural-network based protein modelling server
- [ESyPred3D](#) - Automated homology modeling program using neural networks
- [Geno3d](#) - Automatic modelling of protein three-dimensional structure
- [SDSC1](#) - Protein Structure Homology Modeling Server

Homologní modelování

Threading

- [Phyre](#) (Successor of 3D-PSSM) - Automated 3D model building using profile-profile matching and secondary structure
- [Fugue](#) - Sequence-structure homology recognition
- [HHpred](#) - Protein homology detection and structure prediction by HMM-HMM comparison
- [Libellula](#) - Neural network approach to evaluate fold recognition results
- [LOOPP](#) - Sequence to sequence, sequence to structure, and structure to structure alignment
- [SAM-T02](#) - HMM-based Protein Structure Prediction
- [Threader](#) - Protein fold recognition

„Navlékání“

- [SWEET](#) - Constructing 3D models of saccharides from their sequences

Ab initio

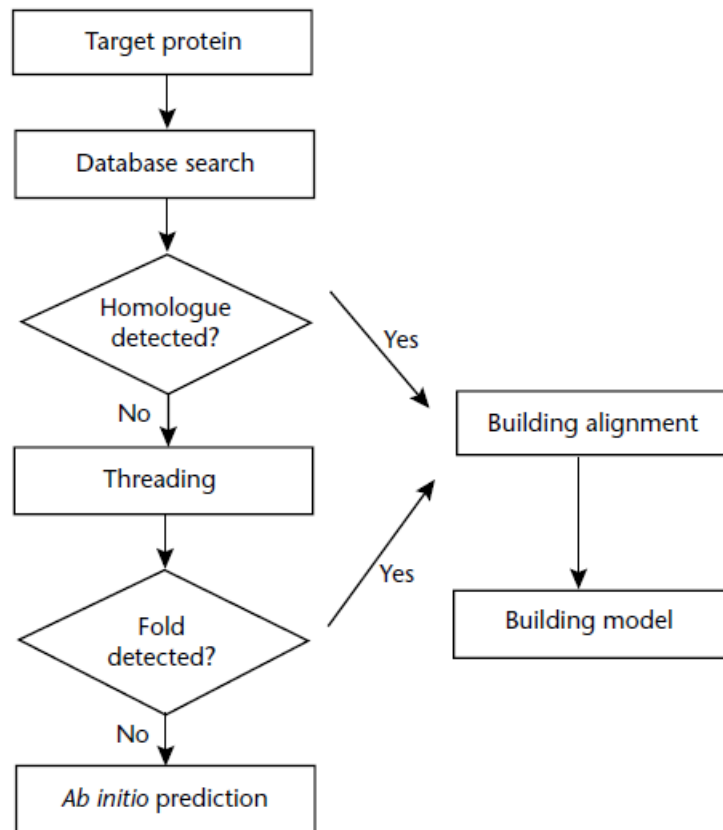
- [HMMSTR/Rosetta](#) - Prediction of protein structure from sequence

„Od počátku“

The result is only a model and must be considered carefully, it isn't an experimental 3D structure! (citace z dokumentace k programu Geno3d)

This server is experimental. Some of the methods used are untested and/or unpublished. Use the server and its results at your own risk. For more information, contact the authors. (citace z dokumentace k programu HMMSTR/Rosetta)

Evaluation of template structure and model quality is a crucial step in homology modelling. (citace z dokumentace k programu SWISS-MODEL).



SAMOSTATNÝ PROJEKT

Analyzujte svůj protein pomocí databáze InterPro a predikujte jeho sekundární strukturu a lokalizaci v buňce.