

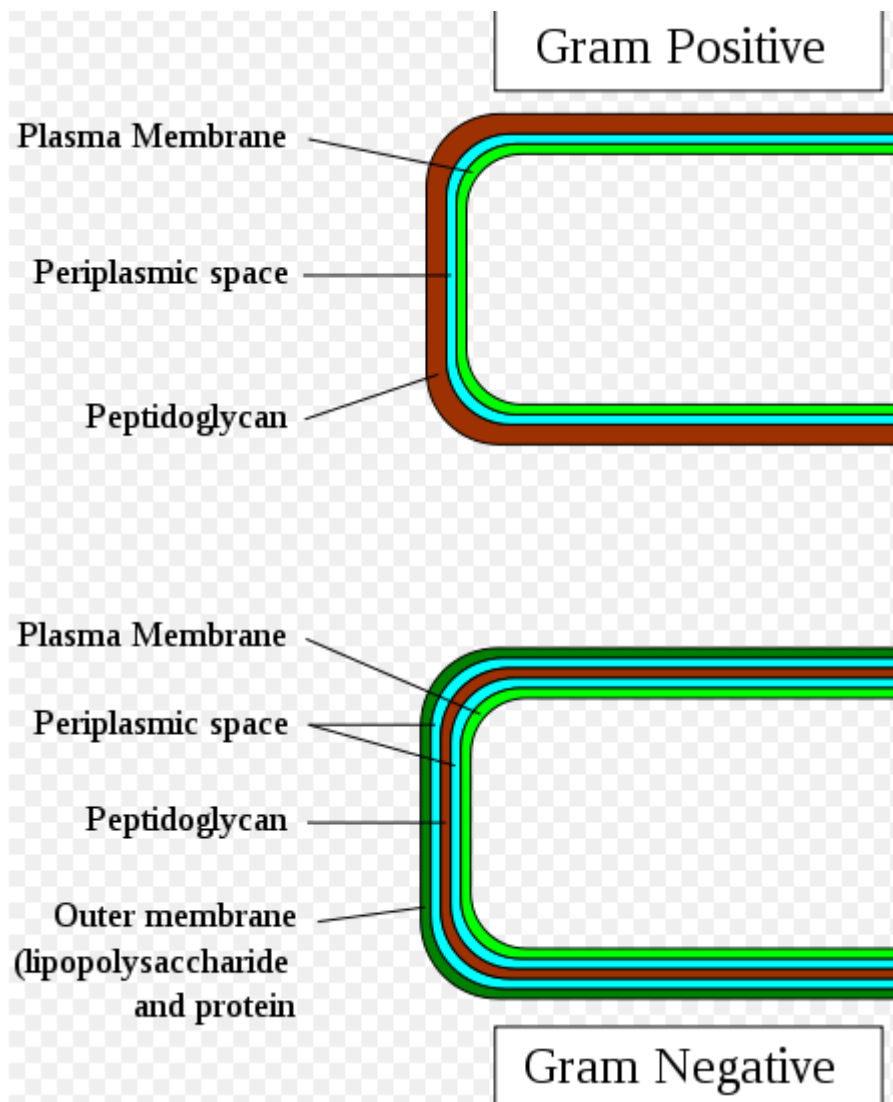
BIOINFORMATIKA V PRAXI – CVIČENÍ 8

PREDIKCE VLASTNOSTÍ PROTEINŮ

Lokalizace, funkce, 2D struktura

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** (<http://www.psорт.org/>) určete lokalizaci následujícího proteinu v buňce:

Protein z bakterie *Micrococcus luteus*

MEINGLTWGLTIALIVGLLAFDYFAHVRKAHTPTIKEAAVWVSGVYVGLALVFGLVFFAFGDTQHAVEYYTGYLE
KALSVDNLFVFLVIMASFRVPREYQQKVLVLFGITFALISRTLIFILLGAAVIAAWSDFYLFGLFLLI IAGGQLKG
EMSGDAEGAQDEADNVMVRLVKRFLPASDQFDGERLFTTVDGKRLMTPMLLVMIAGATDILFAFDSIPAI FGVT
QEAYIVFTATAFSLMGLRQLYFLIDGLLDRLVYLAYGLSALLAFIGVKLILHALHENNLFPINGGENVPAE IPT
NLSLVVVVVILAITVLVSLYSPKQALRALQNAEKYSYRYSKLAEDADPAERERAAALMDRWTARAEALDQRWRD
QLLEHKDAWSAII RTAHE TRFADPRDDDARGVSEQIVRQDGPTV

ÚKOL 2

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

Protein 1 z neznámé gramnegativní bakterie

MTKYALVGDVGGTNARLALCDIASGEISQAKTYSGLDYPSLEAVVRVYLDEHGVSVEDGCIAIACPITGDWVAMT
NHTWAFSIAEMKKNLGFSHLEIINDFTAVSMAIPMLKKEHLIQFGGGEVVDGKPIAVYGAGTGLGVAHLVHVDKR
WISLPGEGGHVDFAPNSEEEAMILEILRAEIGHVSAERVLSPGGLVNLRYRAIVKSDNRLPENLRPKDITERALAD
SCIDCRRALS LFCVIMGRFGGDLALTMGTFGGVYIAGGIVPRFLEFFKASGFRGGFEDKGRFKDYVHGIPVYLIV
HDNPGLLGSGAHLRQTLGHIL

Protein 2 z neznámé bakterie

LVIVDAVTLLSAYPEASRDPAAPTVIDGRHLYVVSPGDAAQLGHNSRLFTGLSPGDQLHLRETALALRAEVSVL
FIRFALKDAGIVAPIELEVRDAATAVPDADDLLHPSRPLKDHYWRSVLAAGATTCTADFAVCDRDGTVSGYFR
WETSIEIAGSQPDTKQPGFKPSS

Protein 3 z *Aspergillus fumigatus*

MKPQTAAFLLSLLGSTLAAPIQHADKGS DVKPTPTGRGAPGGFFTFGFP SGVPSGLPSGFPGGPVPGGFGGDPNG
PIPSGPVPTGAAPSGFPSFGTGPAPSGAPQGEESSSFGGQVQARSPQDFEDSGAAPSGAIPSGAIP TGAVPSG
APNGFGGFGQGGHGGPGGPEEGSGPSPTGAVPSGAI PSGAAPSGAVGGFDGFGQASENSGNAQFQSSGTS PFGA
SHSGSASGHQGRHGGDHRGQHNGSGAIPSGAAPSGAAPSGAAGGFPGFGQGGEGSGPSPTGAVPSGAAGFGGQ
GHGQGGQSFPTGVAPSDVPSAQPTA

ÚKOL 3

Určete lokalizaci proteinu z *Aspergillus fumigatus* v buňce rovněž pomocí **TargetP** (<http://www.cbs.dtu.dk/services/TargetP/>). Výsledky porovnejte.

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.



```
L P P N T A F K A I F Y A N A A D R Q D L K L F I D D  
I P P N T D F R A I F F A N A A E Q O H I K L F I G D  
L P P H I K F G V T A L T H A A N D Q T I D I Y I D D  
L P P N I A F G V T A L V N S S A P O T I E V F V D D
```

[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

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

ÚKOL 4

Dva z předcházejících proteinů analyzujte pomocí databáze **InterPro** (<http://www.ebi.ac.uk/interpro/>). Jaké domény/motivy proteiny obsahují a jaká je předpokládaná funkce proteinů? Popište.

Vytvořte schéma proteinů a jejich domén v nástroji **MyDomains** (<http://prosite.expasy.org/cgi-bin/prosite/mydomains/>).

PREDIKCE SEKUNDÁRNÍ STRUKTURY PROTEINU

ÚKOL 5

Určete, zda následující protein obsahuje transmembránové helixy. Využijte nástroj **TMHMM** (<http://www.cbs.dtu.dk/services/TMHMM-2.0/>).

MEINGLTWGLTIALIVGLLAFDYFAHVRKAHTPTIKEAAVWSGVYVGLALVFGLVFFAFGDTQHAVEYYTGYLLE
KALSVDNLFVFLVIMASFRVPREYQKVLFFGITFALISRTLFI LLGAAVIAAWSDFYLFGLFLLI IAGGQLKG
EMSGDAEGAQDEADNVMVRLVKRFLPASDQFDGERLFTTVDGKRLMTPMLLVMIAGATDILFAFDSIPAI FGVT
QEAYIVFTATAFSLMGLRQLYFLIDGLLDRLVYLAYGLSALLAFIGVKLILHALHENNLPFINGGENVPVAE IPT
NLSLVVVVVILAITVLVSLYSPKGQALRALQNAEKYSYRYSKLAEDADPAERERAAAALMDRWTARAEALDQRWRD
QLLEHKDAWSAIIRTAHETRFADPRDDDARGVSEQIVRQDGPTV

Predikujte úspěšnost exprese tohoto proteinu v rozpustné formě v hostitelském organismu. Použijte nástroj MEMEX (<http://mips.helmholtz-muenchen.de/memex/memex.seam>).

Vytvořte schéma proteinu s využitím nástroje TOPO2 (<http://www.sacs.ucsf.edu/TOPO2/>).

ÚKOL 6

Predikujte sekundární strukturu následujícího proteinu. Využijte libovolné dva programy, které najdete na <http://www.expasy.org/>. Výsledky porovnejte a komentujte.

ADSQTSSNRAGEFSIPPNTDFRAIFFANAAEQQHKLFIGDSQEPAAAYHKLTTRDGPREATLNSGNGKIRFEVSV
NGKPSATDARLAPINGKKS DGS PFTVNF GIVVSE DGHDS DYNDGIVVLQWPIG

Tento protein má vyřešenou strukturu (PDB 2VNV). Porovnejte známou 2D strukturu s predikcí. Využijte nástroj POLYVIEW-2D (<http://polyview.cchmc.org/>).

SAMOSTATNÝ PROJEKT

Predikujte lokalizaci svého proteinu v buňce, dále analyzujte svůj protein pomocí databáze InterPro, určete jeho možnou funkci a predikujte jeho sekundární strukturu.