

BIOINFORMATIKA V PRAXI – CVIČENÍ 7

PREDIKCE VLASTNOSTÍ PROTEINŮ ZÁKLADNÍ FYZIKÁLNĚ-CHEMICKÁ CHARAKTERISTIKA PROTEINŮ, LOKALIZACE, MODIFIKACE

STUDIJNÍ MATERIÁLY

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

ÚKOL 1

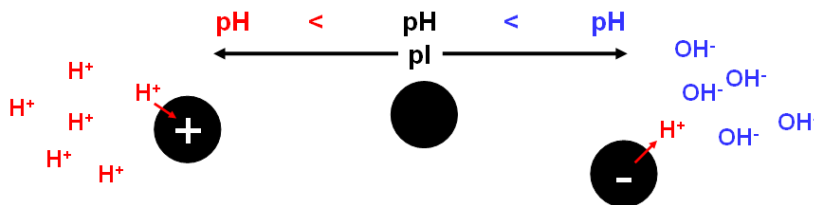
Pomocí programu **Compute pI/Mw** (http://web.expasy.org/compute_pi/) určete teoretickou molekulovou hmotnost a izoelektrický bod proteinu.

```
RVLMEILFDMTRSPIDIELIFLKVVKVSETALEMAVEFPQRWCHGARQGSIPNALVAEQIVVLAARVDRPLTIS  
VEAAVKSIFELISPSVELQSSTGKAAYDFGGKLYSTADADTRNGNIQGLRIHYSLFLFDMHPKEGMLEGI AVEFT  
SVVPTDEDIGNSPSKKTKLSRAPYYDGFHAGTIQATTNYAEKSKQTLMLEGNVVKLHDYQQAKECVLGNHAWPDL  
RKNRGGLEGRCDNL
```

Izoelektrický bod - pI

- **Izoelektrický bod = pH, při kterém má protein nulový sumární náboj.**

Protein pI is calculated using pK values of amino acids described in Bjellqvist et al., which were defined by examining polypeptide migration between pH 4.5 to 7.3 in an immobilised pH gradient gel environment with 9.2M and 9.8M urea at 15%^{v/v} or 25%^{v/v} C. Prediction of protein pI for highly basic proteins is yet to be studied and it is possible that current Compute pI/Mw predictions may not be adequate for this purpose.



ÚKOL 2

Vyhleďte na ExPASy predikční nástroje pro predikci posttranslačních modifikací proteinů.

ÚKOL 3

U následujících hypotetických savčích proteinů predikujte možná místa fosforylace a c-mannosylace. Které aminokyseliny jsou při těchto procesech modifikovány?

Protein 1

```
SCLAETQQMVKAERKLCMHVKTCCFYTAFAEVRYTDQYPRGMCTDHEKSNEQEQTFRSHHAFMTHEHMANSFGL  
VTQQCELRSFDTPKFYVLLACNKKFQAAPPQQVGCICLDYMYIFETIGGPVNIVI IKIDHLTDYMSGKSLKPAR  
GDEMRNPMNCPNERSGVPGSMNNYSGDFMRNFNISDQIALIKHYDVYCHLPVR
```

Protein 2

LFFWVGWKPIKIQQPYWMHNIGMLWNVVILGVPPFCNPWEELEMAVWCHHEWERHNVNGCKWLYPFQKVMGKNLGR
PNVYQHHQDHAGPNIVRMGMIFYQNFLHDAIFCYKANGKAGPQWEDYLVCHDKNIFAFWPCYQGRKFDEWCMVQL
WDCIEKPAAMCYVRDPNKIEVLRDKGEARKLLPMMICYNDQIDPWNYYIFLHVPMQYEFDRRAKQYWDEMQACRY
IVIDDCCRHEMELAFQCMNHRICYFGQAKHREHGRFA

ÚKOL 4

Pomocí programu **ProtParam** (<http://web.expasy.org/protparam/>) určete základní fyzikálně-chemické parametry následujícího proteinu.

HLGYKRWRWSDLFDNKSTALVVARCRIAQAPYDHCINPETRYMMRHWMKCHCAWNECGGRHFENMNAICAWCDFG
LLSTESMVKSFYIRTQPESCSWPYIHTCPLHSLVKRNLIPIYFEHTKGWYEATAHVACIFDEHIDGRWNQISRQG
WWYLPNMCNWNMFYLIQYIQMGQDIRHCHKDHQGVFLRFFDKQHFLQVDNMAFKHTNNCVYPWKIRQNPEWLDL
PNPPGYVDWVCMYVIPQMVEVTSCSINKETWFEIPEYWEFTTKWPGNTYTLVFKGYGADKTSKRQDTMMGQLDGR
PSKAMYS CGMDL SHSAHAQVIPMTNEWSEQLVYFKSIMVDGRMAFRQADQFCQLTHAEKV

ÚKOL 5

Porovnejte extinkční koeficient následujících proteinů.

Protein 1

LVIVDAVTLLSAYPEASRDPAAPTVIDGRHLYVVS PGDAAQLGHNDSRLFTGLSPGDQLHLRETALALRAEVSVL
FIRFALKDAGIVAPIELEVRDAATAVPDADDLLHPSRPLKDHVWRSVDVLAAGATTCTADFAVCDRDTVSGYFR
WETSIEIAGSQPDTKQPGFKPSSDRNGNFSLPNTAFKAI FYANAADRQDLKLFIDDAPEPAATFVGNSEDGVRL
FTLNSKGGKIRIEASANGRQSATDARLAPLSAGDTVWLGWLGAE DGADADYNDGIVILQWPIIT

Protein 2

SSVQTAATSWGTVPSIRVYTANNGKITERCWDGKGWYTGAFNEPGDNVSVTSWLVSIAHIRVYASTGTTTTTEWC
WDGNGWTKGAYTATN

Protein 3

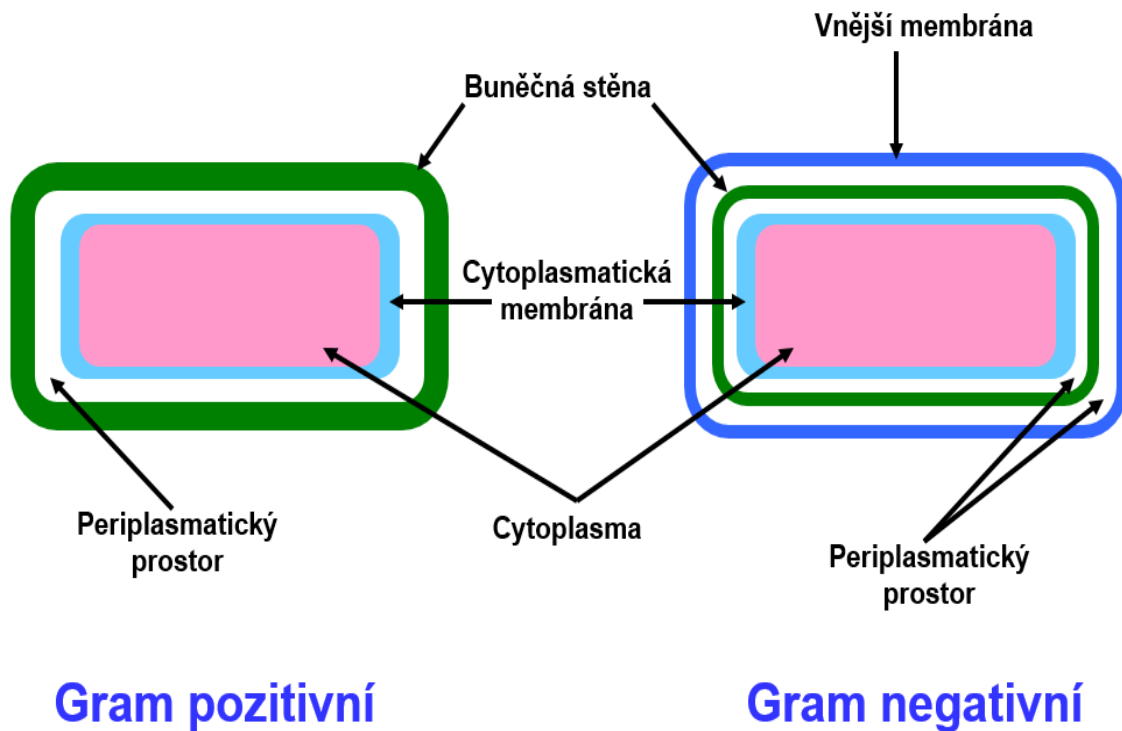
AQQGVFTLPARINFGVTVLVNSAATQHVEIFVDNEPRAAFSGVGTGDNNLGTKVINSGSGNVRVQITANGRQSDL
VSSQLVLANKLNLA VVGSE DGTDM DYND SIVILNPLG

SAMOSTATNÝ PROJEKT

Pomocí programu ProtParam určete u vašeho proteinu základní fyzikálně-chemické parametry.

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 6

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

Protein z bakterie *Micrococcus luteus*

```
MEINGLTWGLTIALIVGLLAFDYFAHVRKAHTPTIKEAAVWSGVYVGLALVFGLVFFAFGDTQHAVEYYTGYLE
KALSVDNLFVFLVIMASFRVPREYQQVLLFGITFALISRTLFI LLGAAVIAAWSDFYLFGLFLLI IAGGQLKG
EMSGDAEGAQDEADNVMVRLVKRFLPASDQFDGERLFTTVDGKRLMTPMLLVMI AIGATDILFAFDSIPAI FGVT
QEAYIVFTATAFSLMGLRQLYFLIDGLLDRLVYLAYGLSALLAFIGVKLILHALHENNL PFI NGGENVPVAE IPT
NLSLVVVVVILAITVLVSLYSPKQALRALQNAEKYSYRYSKLAEDADPAERERAAAALMDRWTARAEALDQRWRD
QLLEHKDAWSAII RTAHE TRFADPRDDDARGVSEQIVRQDGPTV
```

ÚKOL 7

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

Protein z *Aspergillus fumigatus*

```
MKPQTAAFLLSLLGSTLAAPIQHADKGSVDKPTPTGRGAPGGFFTGFP SGVPSGLPSGFPGGPVPGGFGGDGPNG
PIPSGPVPTGAAPSGFPSFGTGPAPSGAPQGEESSSFGGQGVQARSPQDFEDSGAAPSGAIPSGA IPTGAVPSG
APNGFGGFGQGGHGGPGGPGEESGSPSPTGAVPSGAI PSGAAPSGAVGGFDGFGQASENSGNAQFQSSGTS PFGA
SHSGSASGHQGGRRHGGDHRGQHGNNGSGAIPSGAAPSGAAPSGAAGGFPGFGQGGEGSGSPSPTGAVPSGAAGFGGQ
GHGQGGQSFP TGVPASDVPSAQPTA
```

ÚKOL 8

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 9

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

SAMOSTATNÝ PROJEKT

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

PREDIKCE SEKUNDÁRNÍ STRUKTURY PROTEINU

NEPOVINNÝ ÚKOL

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

```
MEINGLTWGLTIALIVGLLAFDYFAHVRKAHTPTIKEAAVWSGVYVGLALVFGLVFFAFGDTQHAVEYYTGYLLE
KALSVDNLFVFLVIMASFRVPREYQQKVLVLFGITFALISRTLFILLGAAVIAAWSDFYLFGLFLLI IAGGQLKG
EMSGDAEGAQDEADNVMVRLVKRFLPASDQFDGERLFTTVDGKRLMTPMLLVMIAGATDILFAFDSIPAIFGVT
QEAYIVFTATAFSLMGLRQLYFLIDGLLDRLVYLAYGLSALLAFIGVKLILHALHENNLPFINGGENVPVAEIP
NLSLVVVVVILAITVLVSLYSPKQALRALQNAEKYSYRYSKLAEDADPAERERAAAALMDRWTARAEALDQRWRD
QLLEHKDAWSAII RTAHETRFADPRDDDARGVSEQIVRQDGPTV
```

NEPOVINNÝ ÚKOL

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

```
ADSQTSSNRAGEFSIPPNTDFRAIFFANAAEQQHIKLFIGDSQEPAAAYHKLTTTRDGPREATLNNGKIRFEVSV
NGKPSATDARLAPINGKKS DGS PF TVNF GIVVSE DGHDS DYNDGIVVLQWPIG
```

PREDIKCE TERCIÁRNÍ STRUKTURY PROTEINU

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).

