

BIOINFORMATIKA V PRAXI – CVIČENÍ 7

PREDIKCE VLASTNOSTÍ PROTEINŮ ZÁKLADNÍ FYZIKÁLNĚ-CHEMICKÁ CHARAKTERISTIKA PROTEINŮ

STUDIJNÍ MATERIÁLY

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

ÚKOL 1

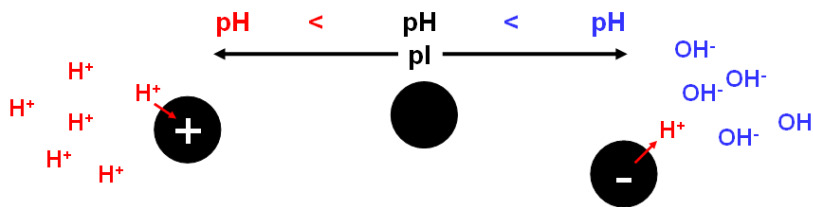
Pomocí programu **Compute pI/Mw** (http://web.expasy.org/compute_pi/) určete teoretickou molekulovou hmotnost a izoelektrický bod proteinu.

```
RVLMEILFDMTRSPIDIELIFLKVVKVSETALEMAVEFPQRWCHGARQGSIPNALVAEQIVVLAARVDRPLTIS  
VEAAVKSI FELIS P S PVELQSSTGKAAAYDFGGKLYSTADADTRNGNIQGLRIHYSFLFDMHPKEGMLEGI AVFT  
SVVPTDEDIGNSPSKKTKLSRAPYYDGFHAGTIQATTNYAEKSKQTLMLEGNNVKLHDYQQAKECVLGNHAWPDL  
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 15i% or 25i% 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

Vyhledejte 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 IKIDHLTDYMSGKSLKPARY  
GDEMARNPMNCPNERSGVPGSMNNSGDFMRNFNISDQIALIKHYDVYCHLPVR
```

Protein 2

```
LFFWVGWKPIKIQQPYWMHNIGMLWNVVILGVPFCNPWEELEMAWVCHHEWERHNVNGCKWLYPFQKVMGKNLGR  
PNVYQHHQDHAGPNIVRMGMEFYQNFLHDAIFCYKANGKAGPQWEDYLVCHDKNIFAFWPCYQGRKFDEWCMVQL  
WDCIEKPAAMCYVRDPNKIEVLRDKGEARKLLPMMICYNDQIDPWNVYIFLHVPMQYEFDRRAKQYWDDEMQACRY  
IVIDDCCRHEMMLAFQCMMNHRICYFGQAKHREHGRFA
```

ÚKOL 4

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

Protein 1

HLGYKRWRWSDLFDNKSTALVVARCRIAQAQPYDHCINPETRYMMRHWMKCHCAWNECGGRHFENMNAICAWCDFG
LLSTESMVKSFYIRTQPESCSWPYIHTCPLHSLVKRNLIPYFEHTKGWYEATAHVACIFDEHIDGRWNQISRQG
WWYLPNMNCWNMFYLIQMGQDIRHCHKDHDQVFLRFFDKQHFLQVDNMAFKHTNNCVYPWKIRQNPFWLDC
PNPPGYVDWVCMYVIPQMVEVTSKINKEWFEIPEYWEFTTKWPGNTYTLVFKGYGADKTSKRQDTMMGQLDGR
PSKAMYSCGMDLSHSAHAQVI PMTNEWSEQLVYFKS IMVDGRMAFRQADQFCQLTHAEKV

Protein 2

ATGGCAACACAAGGAGTGTTCACCCCTCCCGCCAACACCCGGTTCGGCGTCACCGCCTTCGCCAACTCGTCCGGAACCCAGAC
GGTGAACGTGCTGGTCAACAACGAGACGGCCGCGACCTTCAGCGGGCAAAGCACCAATAACGCCGTCATCGGCACCCAGGTGC
TCAACTCCGGCAGCAGTGGAAGGTACAGGTCCAGGTACAGCGTCAACGGCCGCCCTCGGATCTGGTCTCGGCACAGGTAATC
CTGACCAACGAGCTGAACCTCGCCCTGGTCGGCTCTGAAGACGGCACCGACAACGACTACAACGACGCCGTCGTGGTGATCAA
CTGGCCGCTCGGCTAG

ÚKOL 5

Pomocí programu **GPMaw** (http://www.alphalyse.com/gpmaw_lite.html) určete základní fyzikálně-chemické parametry následujícího proteinu.

KIGEDNTQECMCSPLHSNGGSEYLVAGGKAKPAGKLAQHWGQRILGTQRTSAVESGVLAYARATSAGEKTRQWQR
IANSRLIVDKNPNDVAKTKRSYTIILILKVNREQYIPDATVLVADGTFILGDNLVIEKLTTEFPALNPF SFHRDAR
GESQEEVEGLVKMMPDP SHGLLNLT VVATS YVI IMFSLEAVSRIMLFAYASPLTLKLEDFKMNGPKEYDIWVQEE
PFDFLQVHDSAIHRRRIGPLC

ÚKOL 6

Akceptují **Compute pI/Mw**, **ProtParam** a **GPMaw** nestandardní proteinogenní aminokyseliny selenocystein a pyrolysin?

ÚKOL 7

Porovnejte teoretickou stabilitu následujících proteinů.

Protein 1

LVIVDAVTLLSAYPEASRDPAAPTVIDGRHLYVVSPGDAAQLGHNDSRLFTGLSPGDQLHLRETALALRAEVSVL
FIRFALKDAGIVAPIELEVRDAATAVPDADDLLHPSRPLKDHYWRSVLAAGATTCTADFAVCDRDGTVSGYFR
WETSIEIAGSQPDTKQPGFKPSSDRNGNFSLPNTAFKAI FYANAADRQDLKLFIDDAPEPAATFVGNSESDGVR
FTLNSKGGKIRIEASANGRQSATDARLAPLSAGDTVWLGWLGAEADGADADYNDGIVILQWPIT

Protein 2

DRNGNFSLPNTAFKAI FYANAADRQDLKLFIDDAPEPAATFVGNSESDGVRFTLNSKGGKIRIEASANGRQSAT
DARLAPLSAGDTVWLGWLGAEADGADADYNDGIVILQWPIT

Protein 3

LVIVDAVTLLSAYPEASRDPAAPTVIDGRHLYVVSPGDAAQLGHNDSRLFTGLSPGDQLHLRETALALRAEVSVL
FIRFALKDAGIVAPIELEVRDAATAVPDADDLLHPSRPLKDHYWRSVLAAGATTCTADFAVCDRDGTVSGYFR
WETSIEIAGSQPDTKQPGFKPSS

ÚKOL 8

Porovnejte teoretický extinkční koeficient následujících proteinů.

Protein 1

KTDYWTHMWEIEHQDWQPCMVIIWMFDWDVYACWWLHIKRWGYLNHLKHARSQIQHVETRSLAMPPTVDRGDCN
FNSNVQEAWKAKKKNYCGWEQYWPSTCGIAGDFSFNLYLWGYACWKQIGPPESWAFSWLTYQFNYPNTFFIFMHRK
MLITWLEMSFVWCRHNWVVGSRPECR

Protein 2

HHIAINGGLFRRTNNACITCTFTCTGNFQFLDRRLSGQIACRQRFNAQVALRRRFHTFQAFNRNRHEEKQHGNA
MQIIPQCTTHLHAFNREGHATFHAGKKNIHNAKHRNFQNMNLGQKQTAQHTPCHGGQCLFRCEKNCAAGFLSGFC
LEFCAKHHEFHFFLFFNGCKNKGEIIFRKFINPQTGIQGCILLIPSHIGLENFAFNLLGAIDGLTFMCQNHQQL
QHNHFNLNLCNCRKCLLGRSDCRLGNFNFTGQVGLAGLGRGTALGHQTLFTRRDTGKETRDNLTTAKETAHAGAL
HTIRRQFKINLRQEKENHHQAFTHAENQEHGTIGFTLIGITQQEFNNRGTQVQHFQFALQNLHLKIRMINFCRN
QAHQRNTKLGDFEVCAANDQHANPAIALALFQQGIMTRHADLTTTFNCHQENENKHQRRQLRHTGEQLRQICQSDIT
MGLIRIIEVRIKARAMNEHTMHFHCLLELRCAARTIGTGRLCKTGMGHDANI IHCATMHAAA I IARFTF

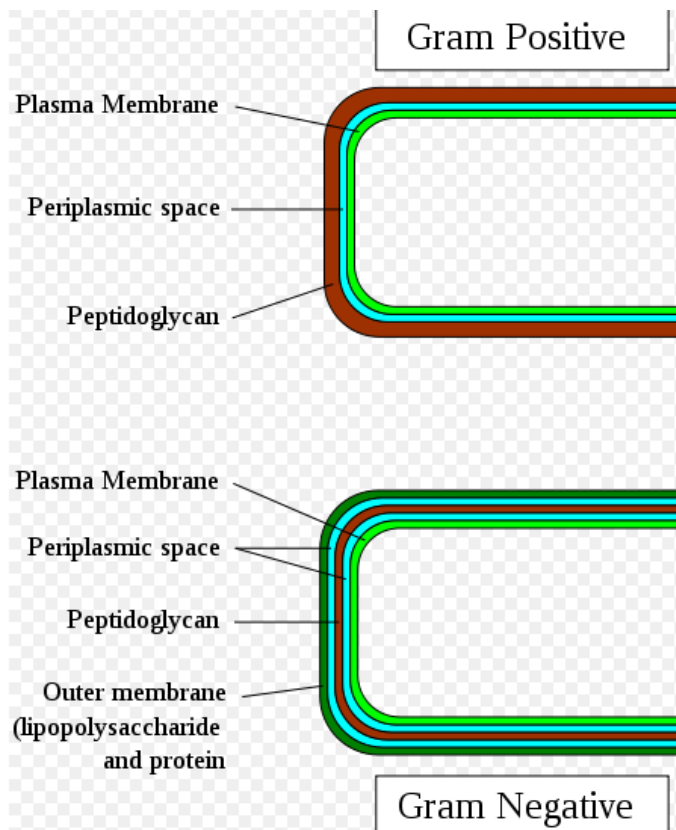
SAMOSTATNÝ PROJEKT

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

PREDIKCE VLASTNOSTÍ PROTEINŮ LOKALIZACE A FUNKCE

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 9

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

Protein z bakterie *Micrococcus luteus*

MEINGLTWGLTIALIVGLLAFDYFAHVRKAHTPTIKEAAVWSGVYVGLALVFGLVFFAFGDTQHAVEYYTGYLE
KALSVDNLFVFLVIMASFRVPREYQQKVLVLFGITFALISRTLFILLGAAVIAAWSDFYLFGLFLLI IAGGQLKG
EMSGDAEGAQDEADNVMVRLVKRFLPASDQFDGERLFTTVDGKRLMTPMLLVMIAGATDILFAFDSIPAI FGVT
QEAYIVFTATAFSLMGLRQLYFLIDGLLDRLVYLAYGLSALLAFIGVKLILHALHENNLPFINGGENVPVAE IPT
NLSLVVVVVILAITVLVSLYSPKQALRALQNAEKYSYRYSKLAEDADPAERERAAAALMDRWTARAEALDQRWRD
QLLEHKDAWSAII RTAHETRFADPRDDDARGVSEQIVRQDGPTV

ÚKOL 10

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

Protein 1 z neznámé gramnegativní bakterie

MTKYALVGDVGGTNARLALCDIASGEISQAKTYSGLDYPSLEAVVRVYLDEHGVSVEDGCIAIACPITGDWVAMT
NHTWAFSIAEMKKNLGFSLHEIINDFTAVSMAIPMLKKEHLIQFGGGEFVDGKPIAVYGAGTGLGVAHLVHVKR
WISLPGEGGHVDFAPNSEEEAMILEILRAEIGHVSAERVLSPGGLVNLRYAIVKSDNRLPENLRPKDITERALAD
SCIDCRRALSFCVIMGRFGGDLALTMGTFFGGVYIAGGIVPRFLEFFKASGFRGGFEDKGRFKDYVHGIPVYLIV
HDNPGLLGSGAHLRQTLGHIL

Protein 2 z neznámé bakterie

LVIVDAVTLLSAYPEASRDPAAPTVIDGRHLYVVSPGDAAQLGHNDSRLFTGLSPGDQLHLRETALALRAEVSVL
FIRFALKDAGIVAPIELEVRDAATAVPDADDLLHPSRPLKDHYWRSVDVLAAGATTCTADFAVCDRDGTVSGYFR
WETSIEIAGSQPDTKQPGFKPSS

Protein 3 z *Aspergillus fumigatus*

MKPQTAAFLLSLLGSTLAAPIQHADKGSVDKPTPTGRGAPGGFFTGFPSPVPSGLPSGFPGGFVPGGFGGDGPNP
PIPSGPVPTGAAPSGFSPFGTGPAPSGAPQGEESSSFFGGQGVQARSQDFEDSGAAPSGAIPSGAIP TGAVPSG
APNGFGGFGQGGHGGPGGPGEEGSGPSPTGAVPSGAI PSGAAPSGAVGGFDGFGQASENSGNAQFQSSGTS PFGA
SHSGSASGHQGRHGGDHRGQHGNNGSGAIPSGAAPSGAAPSGAAGGFPGFGQGGEGSGPSPTGAVPSGAAGFGGQ
GHGQGGSFPTGVAPSDVPSAQPTA

ÚKOL 11

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 12

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.

SAMOSTATNÝ PROJEKT

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