

# 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 **ProtParam** porovnejte extinkční koeficient u následujících proteinů:

#### Protein 1

PTEFLYTSKIAAISWAATGGRQQRVYFQDLNGKIREAQRGGDNPWTGGSSQNVIGEAKLFSPLAAVTW  
KSAQGIQIRVYCVNKDNLSEFVYDGSKWITGQLGSVGVKVGNSKLAALQWGGSESAPPNIRVYYQK  
SNGSGSSIHEYVWSGKWTAGASFGSTVPGTGIGATAIGPRLRIYYQATDNKIREHCWDSNSWYVGGF  
SASASAGVSIAAISWGSTPNIRVYWQKREELYEAAYGGSWNTPGQIKDASRPTPSLPDTFIAANSSG  
NIDISVFFQASGVSLQQWQWISGKGWSIGAVVPTGTPAGW

#### Protein 2

PLLSASIVSAPVVTSETFVDIPGLFLDVAKAGIRDGKLQVILNVPTPFATGNNFPGIFFAIATNQGVV  
ADGCFTFSSKVPESTGRMPFTLVATIDVSGVTFVKQAKSVRGSAMHIDSFASLSAIAGTAAPSSQG  
SGNQGAETGGTGAGNIGGGGERDGFNLPPHIKFGVTALTHAANDQTIDIFIDDDPKPAATFKGAGAQ  
DQNLGTKVLDSGNGRVRVIVMANGRPSRLGSRQVDIFKKSFFGIIGSEDGADDDFNDGIVFLNAPLG

### ÚKOL 2

Porovnejte *in vivo* „životnost“ následujících proteinů:

#### Protein 1

RADSQTSSNRAGEKFSIPPNKTDFRAIFFANAAEQQHIKLFIGDSQEPAAYHKLTTTRDGP  
ATLNSGNGKIRFEVSVNKGKPSATDARLAPINGKKS DGSPFTVNKFGIVVSEDKGHDSKDY  
DGIVVLKQWPIGK

#### Protein 2

MRADSQTSSNRAGEKFSIPPNKTDFRAIFFANAAEQQHIKLFIGDSQEPAAYHKLTTTRDGP  
EATLNSGNGKIRFEVSVNKGKPSATDARLAPINGKKS DGSPFTVNKFGIVVSEDKGHDSKDY  
NDGIVVLKQWPIGK

### ÚKOL 3

Porovnejte stabilitu následujících proteinů:

#### Protein 1

LVIVDAVTLLSAYPEASRDPAAPTVIDGRHLYVVSPGDAAQLGHNDSRLFTGLSPGDQLHLR  
ETALALRAEVSVLFIKFDAGIVAPIELEVRDAATAVPDADDLLHPSCRPLKDHYWRS  
LAAGATTCTADFAVCDRDGTVSGYFRWETSIEIAGSQPDTKQPGFKPSS

#### Protein 2

DRNGNFSLPNTAFKAI FYANAADRQDLKLFIDDAPEPAATFVGNSEDGVRFLFTLNSKGGKI  
RIEASANGRQSATDARLAPLSAGDVTWLGWLGAEADGADADYNDGIVILQWPIIT

#### ÚKOL 4:

Porovnejte „aliphatic index“ a parametr „GRAVY“ u následujících proteinů:

##### Protein 1

MEINGLTWGLTIALIVGLLAFDYFAHVRKAHTPTIKEAAVWSGVYVGLALVFGLVFFAFGDTQHAVEYYTGYLE  
KALSVDNLFVFLVIMASFRVPREYQKVLVLFGITFALISRTLFILLGAAVIAAWSDFYLFGLFLLI IAGGQLKG  
EMSGDAEGAQDEADNVMVRLVKRFLPASDQFDGERLFTTVDGKRLMTPMLLVMIAGATDILFAFDSIPAI FGVT  
QEAYIVFTATAFSLMGLRQLYFLIDGLLDRLVYLAYGLSALLAFIGVKLILHALHENNLPIINGGENVPVAE IPT  
NLSLVVVVVILAITVLVSLYSPKGQALRALQNAEKYSYRYSKLAEDADPAERERAAALMDRWTARAEALDQRWRD  
QLLEHKDAWSAII RTAHETRFADPRDDDARGVSEQIVRQDGPTV

##### Protein 2

MTKYALVGDVGGTNARLALCDIASGEISQAKTYSGLDYPSLEAVVRVYLDEHGVSVEDGCIACIPITGDWVAMT  
NHTWAFSIAEMKKNLGFSHLEIINDFTAVSMAIPMLKKEHLIQFGGGEVVDGKPIAVYGAGTGLGVAHLVHVDKR  
WISLPGEGGHVDFAPNSEEEAMILEILRAEIGHVSAERVLSPGLVNLRYRAIVKSDNRLPENLRPKDITERALAD  
SCIDRRALSFLCVIMGRFGGDLALTMGTGGVYIAGGIVPRFLEFFKASGFRGGFEDKGRFKDYVHGIPVYLIV  
HDNPGLLGSGAHLRQTLGHIL

#### ÚKOL 5:

Pomocí programu **ProtParam** určete základní fyzikálně-chemické parametry následujícího proteinu:

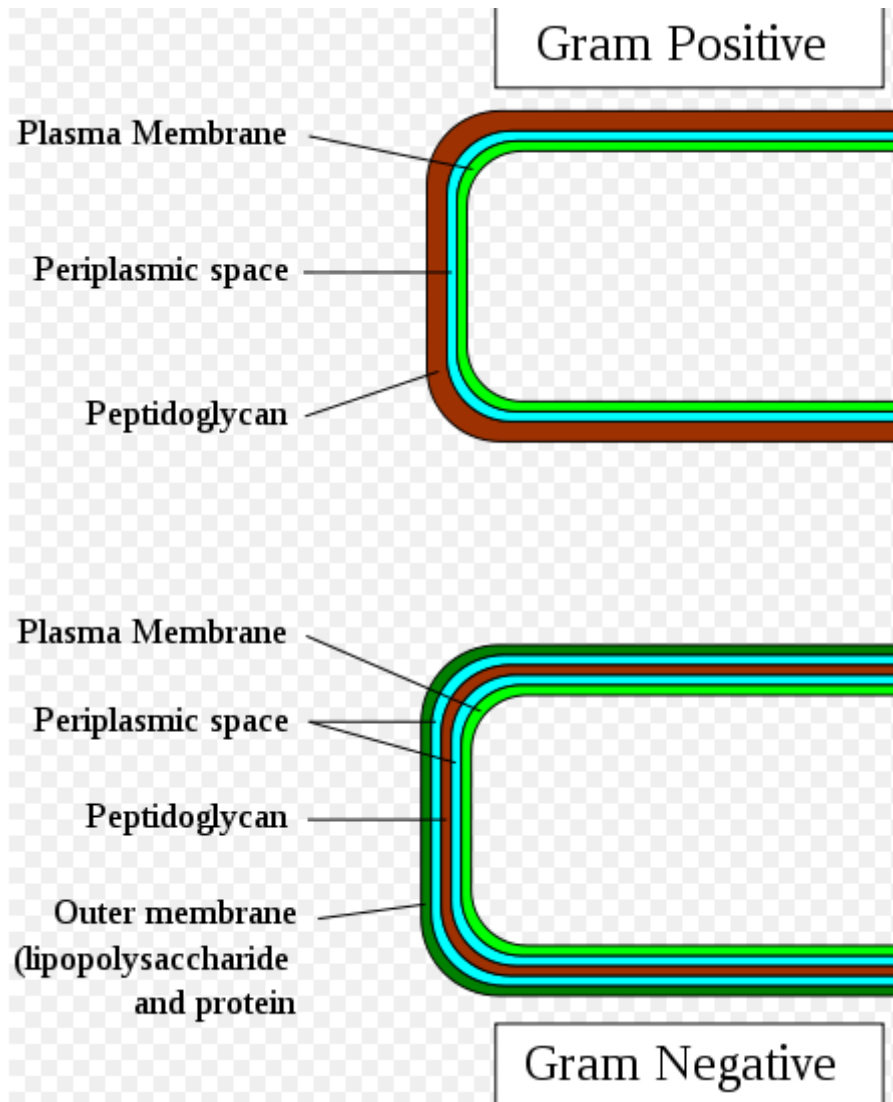
ATGCTGGTGATTGTGGATGCCGTTACCTGCTGAGCGCCTATCCGGAAGCCAGCCGTGATCCGGCCGCCCCGACC  
GTGATTGATGGTCGCCACCTGTATGTTGTTAGCCCGGGCGATGCCGCGCAGCTGGGCCATAACGATAGCCGTCTG  
TTTACCGGTCTGAGCCCGGGTGATCAGCTGCATCTGCGCGAAACCCGCGCTGGCGCTGCGCGCGGAAGTGAGCGTG  
CTGTTTATTTCGCTTTGCCCTGAAAGATGCCGGCATTGTTGCCCGATCGAACTGGAAGTGCGTGATGCCGCCACC  
GCCGTTCCGGATGCGGATGATCTGCTGCATCCGAGCTGTCGTCCGCTGAAAGATCATTATTGGCGCAGCGATGTG  
CTGGCGCGGGCGCGACCACCTGTACCGCCGATTTTGCGGTGTGCGATCGTGATGGCACCGTGAGCGGTTATTTT  
CGTTGGGAAACCAGCATTGAAATTGCGGGCAGCCAGCCGGATAACAAACAGCCGGGCTTTAAACCGAGCAGCGAT  
CGCAATGGCAACTTTAGCCTGCCGCCGAATACCGCCTTTAAAGCGATCTTCTATGCGAACCGGGCGGATCGTCAG  
GATCTGAAACTGTTTATTGATGATGCGCCGGAACCGGCCACCTTTGTGGGTAACAGCGAAGATGGTGTGCGT  
CTGTTTACCCTGAATAGCAAAGGTGGTAAAATTTCGTATTGAAGCGAGCGCGAACGGCCGTCAGAGCGCGACCGAT  
GCCCCGTCTGGCGCCGCTGAGCGCGGGCGATAACCGTGTGGCTGGGCTGGCTGGGCGCGGAAGATGGTGCCGATGCG  
GATTATAATGATGGCATTGTTATTCTGCAGTGGCCGATTACCTAA

#### 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.psort.org/>) určete lokalizaci následujícího proteinu v buňce:

#### Protein z grampozitivní bakterie *Micrococcus luteus*

```
MEINGLTWGLTIALIVGLLAFDYFAHVRKAHTPTIKEAAVWSGVYVGLALVFGLVFFAFGDTQHAVEYYTGYLEE  
KALSVDNLFVFLVIMASFRVPREYQQKVLVLFGITFALISRTLFI LLGAAVIAAWSDFYLFGLFLLI IAGGQLKG  
EMSGDAEGAQDEADNVMVRLVKRFLPASDQFDGERLFTTVDGKRLMTPMLLVMI AIGATDILFAFDSIPAI FGVT  
QEAYIVFTATAFSLMGLRQLYFLIDGLLDRLVYLAYGLSALLAFI GVKLILHALHENNLPFINGGENVPVAE IPT  
NLSLVVVVVVILAITVLVSLYSPKGQALRALQNAEKYSYRYSKLAEDADPAERERAAALMDRWTARAEALDQRWRD  
QLLEHKDAWSAII RTAHETRFADPRDDDARGVSEQIVRQDGPTV
```

## ÚKOL 7

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

### Protein 1 z neznámé gramnegativní bakterie

MTKYALVGDVGGTNARLALCDIASGEISQAKTYSGLDYPSLEAVVVRVYLDEHGVSVEDGCIAIACPITGDWVAMT  
NHTWAFSIAEMKKNLGFSHLEIINDFTAVSMAIPMLKKEHLIQFGGGEVVDGKPIAVYGAGTGLGVAHLVHVDKR  
WISLPGEGGHVDFAPNSEEEAMILEILRAEIGHVSAERVLSGPGLVNLRYRAIVKSDNRLPENLRPKDITERALAD  
SCIDCRRALSFLFCVIMGRFGGDLALTMGTFFGGVYIAGGIVPRFLEFFKASGFRGGFEDKGRFKDYVHGIPVYLIV  
HDNPGLLGSGAHLRQTLGHIL

### Protein 2 z neznámé bakterie

LVIVDAVTLLSAYPEASRDPAAPTVIDGRHLYVVSPGDAAQLGHNDSRLFTGLSPGDQLHLRETALALRAEVSVL  
FIRFALKDAGIVAPIELEVRDAATAVPDADDLLHPSRPLKDHYWRSDVLAAGATTCTADFAVCDRDRGTVSGYFR  
WETSIEIAGSQPDTKQPGFKPSS

### Protein 3 z *Aspergillus fumigatus*

MKPQTAAFLLSLLGSTLAAPIQHADKGSVDKPTPTGRGAPGGFFTFGFPSPVPSGLPSGFFPGGPVPGGFGGDGPNG  
PIPSGPVPTGAAPSGFPSFGTGPAPSGAPQGEESSSFFGGQGVQARSPQDFEDSGAAPSGAIPSGAIPGTAVPSG  
APNGFGGFGQGGHGGPGGPGGEEGSGPSPTGAVPSGAIPSGAAPSGAVGGFDGFGQASENSGNAQFQSSGTSFPGA  
SHSGSASGHQGGRRHGGDHRGQHGNNGSGAIPSGAAPSGAAPSGAAGGFPGFGQGGEGSGPSPTGAVPSGAAGFGGQ  
GHGQQGGSFPTGVAPSDVPSAQPTA

## SAMOSTATNÝ PROJEKT

Predikujte lokalizaci Vašeho proteinu v buňce.

## 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.



```
L P F 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 F N T D F R A I F F A N A A E Q C H I K L F I G D .  
L P F 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 F N I A F G V T A L V N S S A P Q 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-{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 8**

Vyberte si čtyři 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?

### **SAMOSTATNÝ PROJEKT**

Analyzujte svůj protein pomocí databáze InterPro.