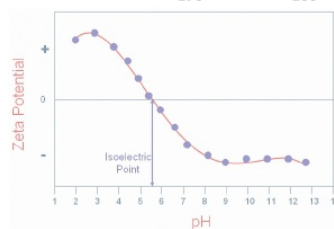
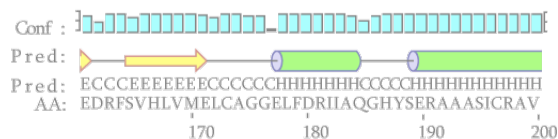
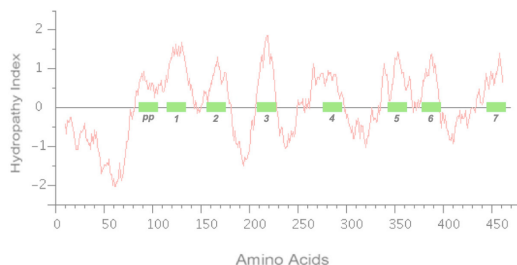


6. vaja

Baze in analiza aminokislinskih zaporedij



Miha Pavšič
april 2014

6. vaja: Baze in analiza ak zaporedij

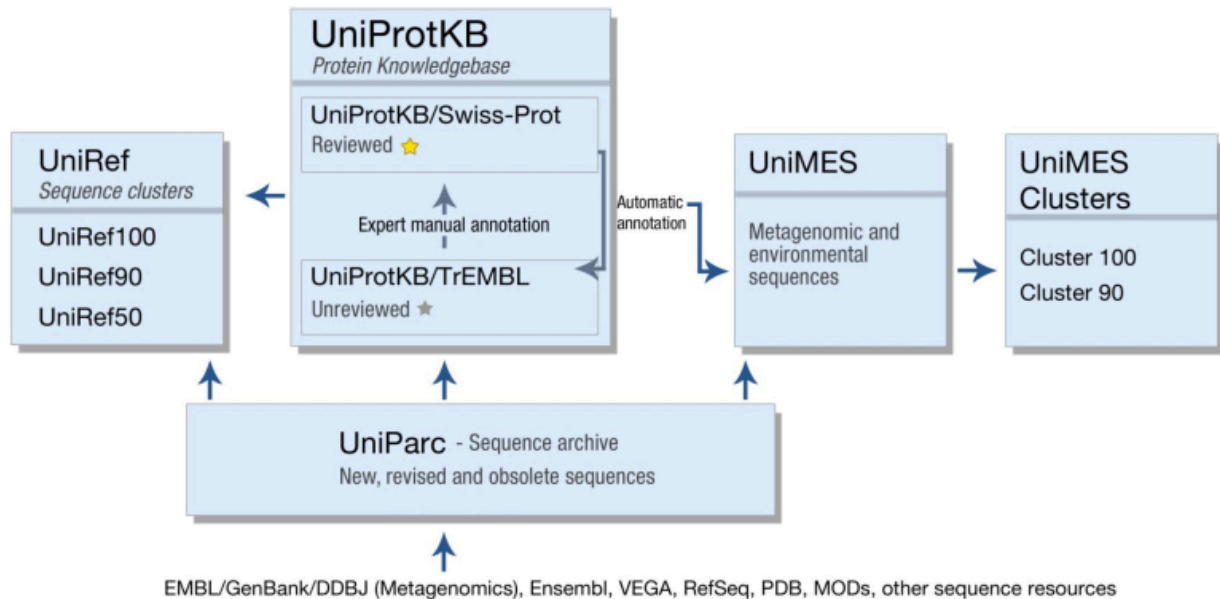
Pregled vaje

- **iskanje aminokislinskih zaporedij po podatkovnih bazah**
 - UniProt, NCBI Protein
 - struktura zapisa v bazo, zunanje povezave
- **izračun/napoved lastnosti proteina na osnovi ak zaporedja**
 - molekulska masa
 - ekstinkcijski koeficient
 - izoelektrična točka
 - signalni peptid, hidrofobne in transmembranske regije
 - lokalizacija
 - glikozilacija
 - mesta cepitev s peptidazami
 - sekundarna struktura

v obliki kviza v
spletni učilnici

UniProtKB (Uniprot Knowledgebase)

Obširna baza proteinskih zaporedij z anotacijami (dodatnimi podatki).



- UniProtKB/Swiss-Prot: ročno anotirani zapisi ★
- UniProtKB/TrEMBL: avtomatsko prevedena zaporedja z avtomatskimi anotacijami

ProtParam (<http://web.expasy.org/protparam/>)

Polje za vnos

ExPASy - ProtParam tool

web.expasy.org/protparam/

ProtParam

ProtParam tool

ProtParam (References / Documentation) is a tool which allows the computation of various physical and chemical parameters for a given protein stored in Swiss-Prot or TrEMBL or for a user entered sequence. The computed parameters include the molecular weight, theoretical pI, amino acid composition, atomic composition, extinction coefficient, estimated half-life, instability index, aliphatic index and grand average of hydropathicity (GRAVY) (Disclaimer).

Please note that you may only fill out one of the following fields at a time.

Enter a Swiss-Prot/TrEMBL accession number (AC) (for example P05130) or a sequence identifier (ID) (for example KPC1_DROME):

Or you can paste your own sequence in the box below:

RESET Compute parameters

SIB Swiss Institute of Bioinformatics | Disclaimer

Back to the Top

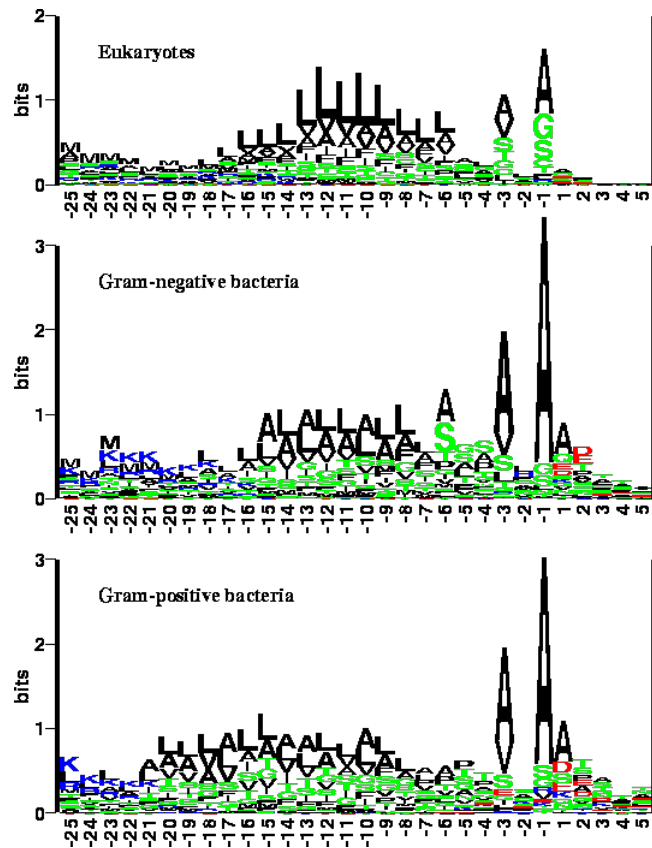
2 načina vnosa zaporedja:

lahko vnesemo tudi kodo zapisa v bazo ak zaporedij – POZOR, če ne računamo parametrov za celotno polipeptidno verigo!

polje za vnos aminokislinskega zaporedja (lahko vsebuje tudi številke, presledke, ...)

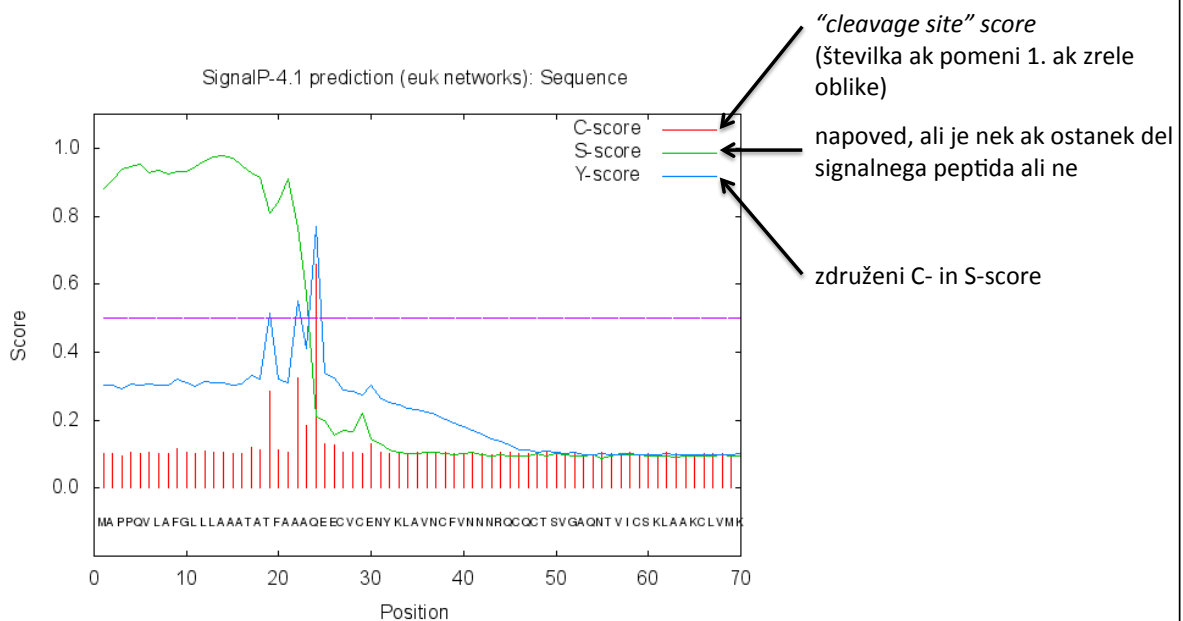
Signalni peptid

- v signalnem peptidu lahko prepoznamo 3 regije (n, h, c)
- pravilo (-3,-1)** pravi, da sta na mestih -3 in -1 (relativno glede na mesto cepitve) ponavadi ak ostanka z majhno in nevtralno stransko skupino
- regija n** – praviloma + nabite ak
- regija h** – praviloma hidrofobne ak
- regija c** – praviloma majhne, a polarne ak



SignalP (<http://www.cbs.dtu.dk/services/SignalP/>)

- identifikacija signalnega peptida in sorodnih zaporedij
- izbrati moramo ustrezen organizem (evkarionti, po Gramu + ali – bakterije)!



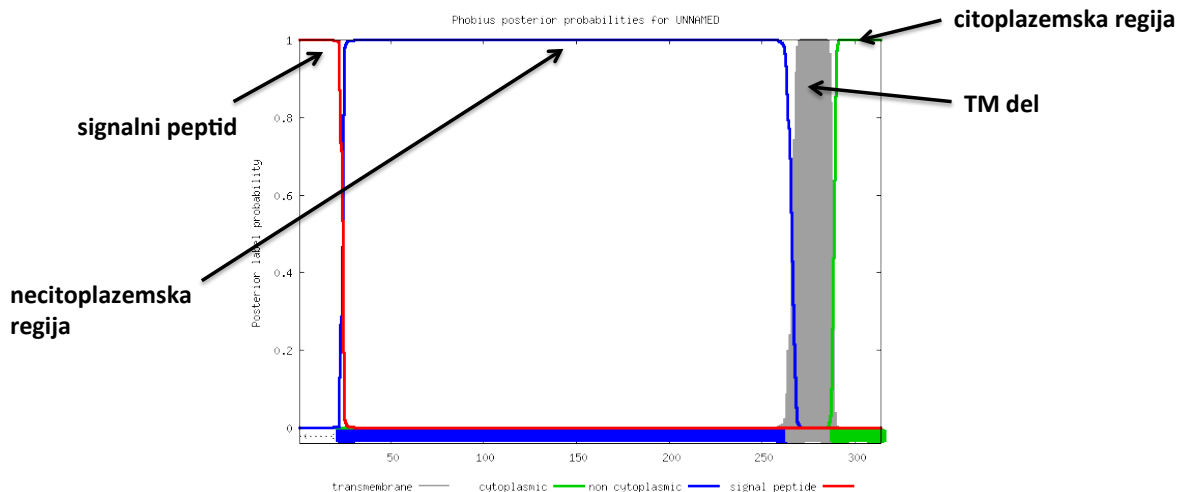
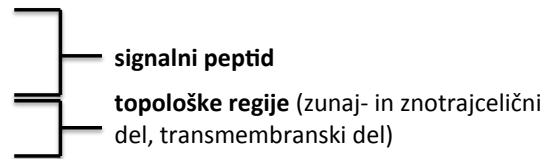
Phobius (<http://phobius.sbc.su.se>)

- napoved signalnega peptida (tudi regij n, h in c)
- nepoved transmembranskih vijačnic



```

ID UNNAMED
FT SIGNAL 1 23
FT REGION 1 5 N-REGION.
FT REGION 6 17 H-REGION.
FT REGION 18 23 C-REGION.
FT TOPO_DOM 24 265 NON CYTOPLASMIC.
FT TRANSMEM 266 289
FT TOPO_DOM 290 314 CYTOPLASMIC.
//
  
```



TargetP (<http://www.cbs.dtu.dk/services/TargetP/>)



- napoved subcelične lokalizacije evkariontskih proteinov, združuje:
 - ChloroP (<http://www.cbs.dtu.dk/services/ChloroP/>) za identifikacijo cTP
 - SignalP (<http://www.cbs.dtu.dk/services/SignalP/>) za identifikacijo SP

SUBMISSION

Paste a single sequence or several sequences in **FASTA** format into the field below:

Submit a file in **FASTA** format directly from your local disk:

Choose File no file selected

Non-plant Plant
 Perform cleavage site predictions

no cutoffs; winner-takes-all (default)
 specificity >0.95 (predefined set of cutoffs that yielded this specificity on the TargetP test sets)
 specificity >0.90 (predefined set of cutoffs that yielded this specificity on the TargetP test sets)
 define your own cutoffs (0.00 - 1.00): cTP: 0.00 mTP: 0.00 SP: 0.00 other: 0.00

Submit Clear fields

ak zaporedje (lahko vsebuje številke in presledke)

zaporedje v formatu FASTA lahko tudi prenesemo v obliki datoteke

napoved mesta cepitve

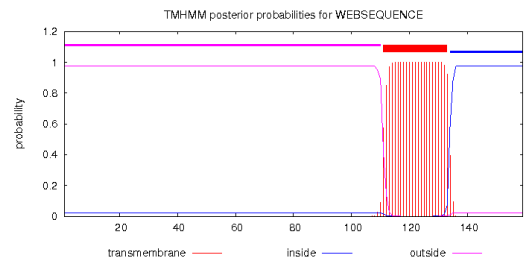
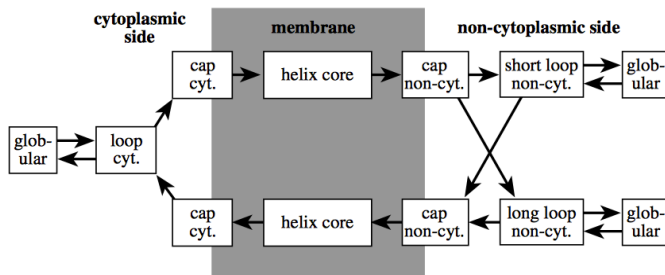
(ne)rastlinski protein

omejitve

- navodila: <http://www.cbs.dtu.dk/services/TargetP/instructions.php>
- razlaga oblike rezultatov: <http://www.cbs.dtu.dk/services/TargetP/output.php>

Transmembranske vijačnice

- pribl. **21 ak** ostankov z (v glavnem) hidrofobnimi stranskimi skupinami → za 1 TM vijačnico
- napoved temelji na osnovi **hidrofobnosti**; **okno** velikosti 17-23 ak ostankov
 - podobno kot pri vaji 2 (naloga 3 – v Excelu)
- **topologija**: katere regije polipeptidne verige so na posameznih straneh membrane in kje so transmembranske regije
 - v UniProt-u kot **topological domain**
- eden izmed programov za napoved je **TMHMM** (<http://www.cbs.dtu.dk/services/TMHMM-2.0/>)
 - napoved na osnovi *hidden Markov models*



N-glikozilacija

- motiv Asn - katerakoli_ak_razen_Pro – Ser_ali_Thr - katerakoli_ak_razen_Pro
- **skrajšan zapis**: N{P}[ST]{P}
 - {P} pomeni, da je na tem mestu neka ak, razen Pro
 - [ST] pomeni, da je na tem mestu ali Ser ali Thr (ne oba!)
- primer programa za napoved: **NetNGlyc** (<http://www.cbs.dtu.dk/services/NetNGlyc/>)

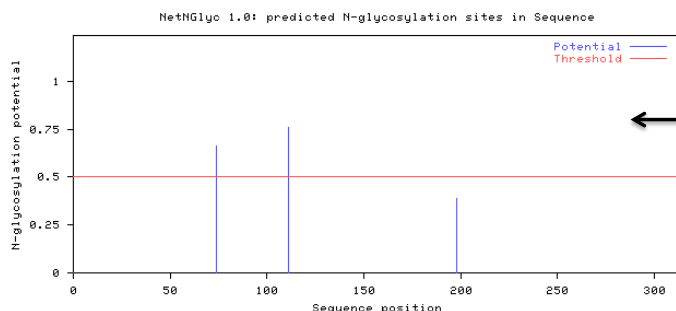
```

Name: Sequence Length: 314
MAPPPQLAFGLLLAAATATFAAAQECCVCEYKLAVNCVFNHNRQCCTSVGAQNTVICSKLAAKCLVMKAEKGGKILGR
KAPPEGALQNDGLYDPPDCDESGLFKAKQCQTSKMCVNTAGVRRITDKDTEITCSEVRVYTWIIIELEKAKREKPYDSK
SLRATLQKEITTRYQLDPFPIISILYENNYTIDLVQNSQKTKQNDVDIADVAITFEKDVNGESLPHSKRMLTVNGEQLD
DLDPGQCLIIYYDKKAPFSGQKAKGAVIYVVVVIKAVIYVIVLVSRRKKRKAIEKAKIEKEMGHRELNA
.....N.....
(Threshold=0.5)
-----
SeqName      Position  Potential  Jury  N-Glyc
              agreement result
-----
Sequence     74  NGSK   0.6603   (9/9)  ++
Sequence    111  NGTS   0.7604   (9/9)  +++
Sequence    198  NSSQ   0.3903   (7/9)  -
    
```

prikaz analiziranega ak zaporedja z označenimi mesti N-glik. →

“shema” zaporedja z zapisanimi mesti N-glik. (preko *threshold*) →

tabela identificiranih mest N-glikozilacije →



grafični prikaz →

