

Module 4

Cellular Localization

Verifying function based on localization data

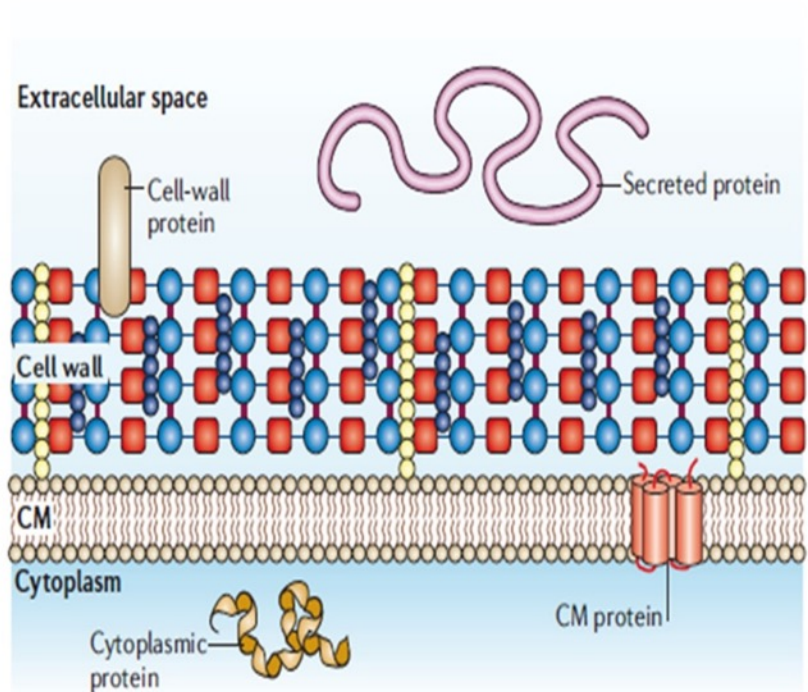
Rama Dey-Rao, PhD
Dept of Microbiology and Immunology
Biotechnical and Clinical Laboratory Sciences
University at Buffalo

Question to Answer:

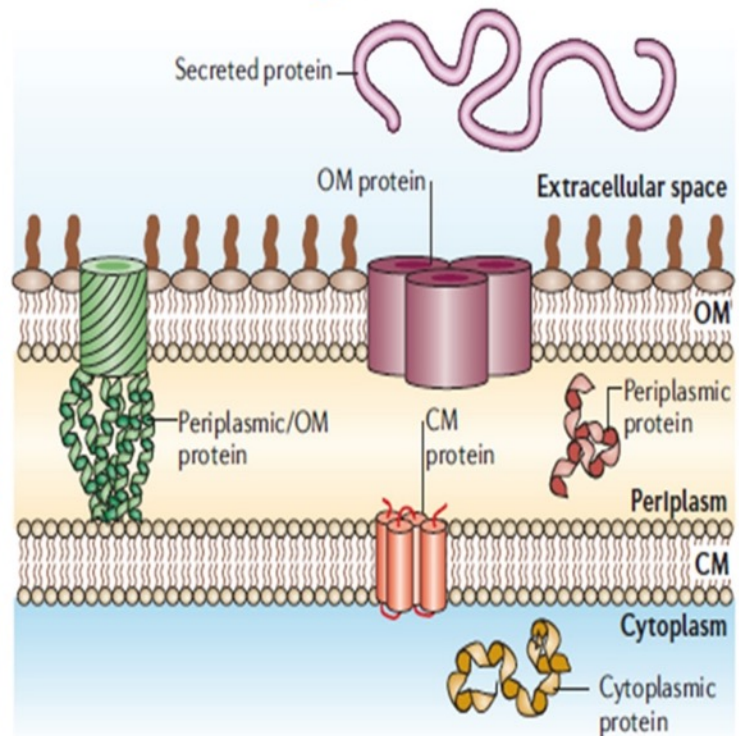
Where is your protein located in the cell?

- The sequence of a protein not only determines how it will fold and ultimately function, but also where the protein will end up in the cell.
- Can determine suitability as a diagnostic target or in vaccines and helps in future experimental designs.

Gram Positive Bacteria



Gram Negative Bacteria



Proteins can be in a variety of locations in bacteria - the cytoplasm, the plasma membrane, the periplasmic space, the outer membrane, loosely associated on the outside of the cell as part of biofilms, or they may be secreted outside the cell.

Gram Stain SOP

- Click on link <http://www.ncbi.nlm.nih.gov/pubmed/>
- Type the name *Kytococcus sedentarius* in search window
- Click Search
- Search List of publications related to *Kytococcus*
- Choose the one that comes close to a genome sequencing or description of the organism
- Read the Abstract
- Get to the information about Gram stain characteristic of *Kytococcus*.

PUBMED

(>23 million citations) housed in [National Center for Biotechnology Information](http://www.ncbi.nlm.nih.gov/biotech) in National Library of Medicine at the NIH.

- <http://www.ncbi.nlm.nih.gov/pubmed/>

The screenshot shows the PubMed search results for the query "kytoccoccus sedentarius". The search bar at the top contains the text "kytoccoccus sedentarius". Below the search bar, there are options for "Create RSS", "Create alert", and "Advanced". The search results are displayed in a list format, with the first result being "Complete genome sequence of *Kytococcus sedentarius* type strain (541)". A red arrow points from the search bar to the search results, and another red arrow points from the search bar to the search results.

Type in *Kytococcus sedentarius* in search window

The screenshot shows the abstract page for the article "Complete genome sequence of *Kytococcus sedentarius* type strain (541)". The abstract text reads: "Kytococcus sedentarius (ZoBell and Upham 1944) Stackebrandt et al. 1995 is the type strain of the species, and is of phylogenetic interest because of its location in the Dermacoccaceae, a poorly studied family within the actinobacterial suborder Micrococccineae. Kytococcus sedentarius is known for the production of oligoketide antibiotics as well as for its role as an opportunistic pathogen causing valve endocarditis, hemorrhagic pneumonia, and pitted keratolysis. It is strictly aerobic and grows when several amino acids are provided in the medium. The strain described in this report is a free-living, nonmotile, Gram-positive bacterium, originally isolated from a marine environment. Here we describe the features of this organism, together with the complete genome sequence, and annotation. This is the first complete genome sequence of a member of the family Dermacoccaceae and the 2,785,024 bp long single replicon genome with its 2639 protein-coding and 64 RNA genes is part of the Genomic Encyclopedia of Bacteria and Archaea project." The abstract also includes keywords: "Dermacoccaceae; aerobic; free-living; marine; mesophile; opportunistic pathogenic".

Kytococcus sedentarius is a Gram-positive bacterium

Notebook in GENI-ACT

Cellular Localization Data Module

Module Instructions

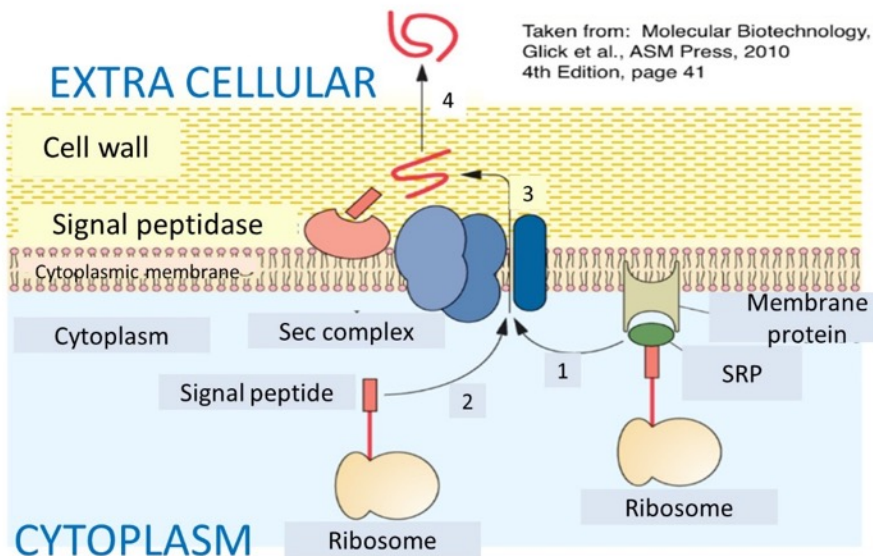
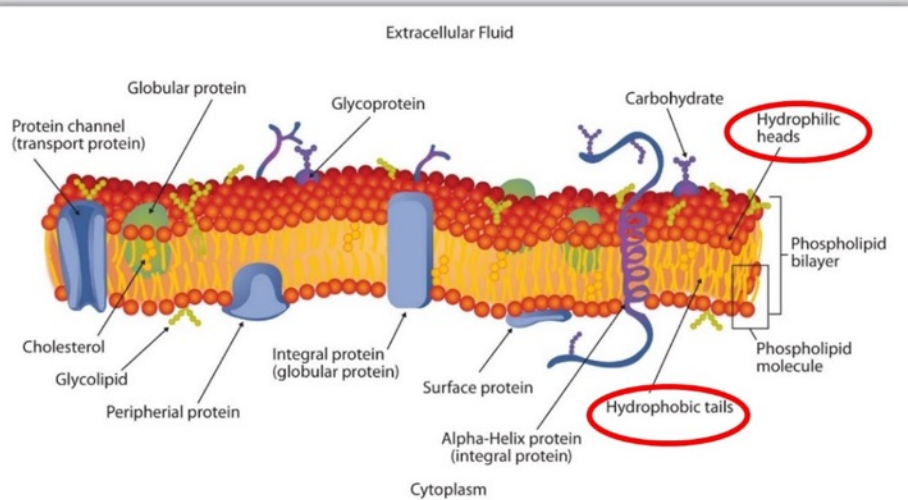
Gram Stain

go to NCBI Pubmed at <http://www.ncbi.nlm.nih.gov/pubmed/>

Gram stain of the microbe

Gram +

Where is your protein located in the cell ?



Proteins-

- with TMHs may be integral membrane proteins
- Proteins that stay inside the membrane usually have a series of 18-20 hydrophobic aa that prefer the waterless compartment of the lipid bilayer forming a transmembrane helix (TMH)
- with signal peptides may be secreted proteins
- These are short peptides at the amino terminal end which can be cleaved off by **proteases** before secretion.
- This signal peptide does not remain a part of the functional secreted protein.
- with neither will probably be present in the cytoplasm.

Cellular Localization- 5 tools

➤ Is your protein of interest a membrane protein?

1. Transmembrane Helices Hidden Markov Models (TMHMM)

➤ Is your protein of interest a secreted protein? Does it have signal sequences?

2. SignalP

3. LipoP

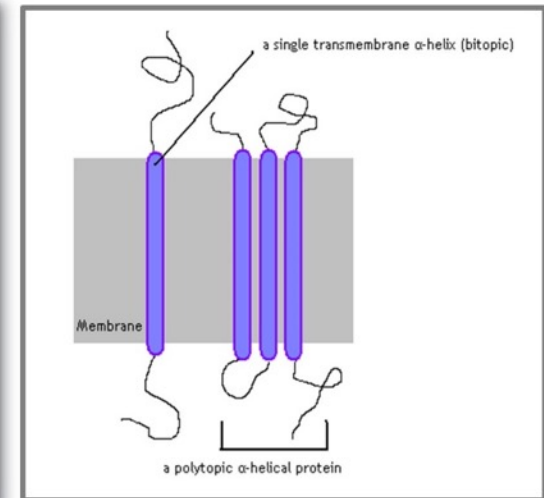
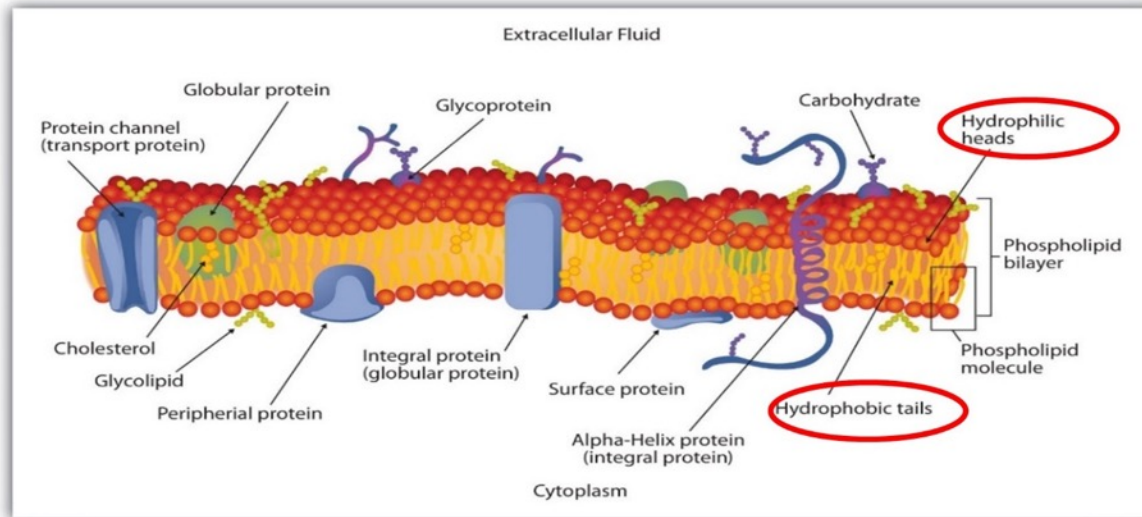
➤ Confirmatory tests

4. PSORT-B

5. Phobius

Final decision can only be made after using all 5 tools

Transmembrane Helices Hidden Markov Models (TMHMM)



Proteins-

- with trans-membrane helices (TMHs) may be integral membrane proteins
- Proteins that stay inside the membrane usually have a series of 18-20 hydrophobic aa that prefer the waterless compartment of the lipid bilayer forming a transmembrane helix (TMH)

✓ Does the protein of interest have transmembrane helices?

Hidden Markov Models (HMM)

- A **statistical** and **probabilistic model** developed from observed sequences of several proteins with a known function.
- Proteins, or domains of proteins, of known function are aligned to create the profile HMM, against which sequences of proteins with unknown function are compared.
- The software then predicts whether amino acid sequence of the protein of unknown function matches that of the profile of the known function.

SOP for a TMHMM search

- Copy the amino acid sequence of your query protein under investigation.
- Open up the link
<http://www.cbs.dtu.dk/services/TMHMM>
- Paste the query sequence into the interactive window
- Select extensive with graphics
- Click Submit

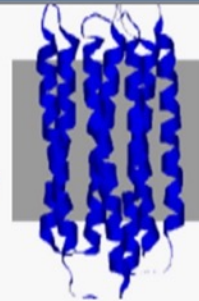
TMHMM Start Page

TMHMM Server v. 2.0

<http://www.cbs.dtu.dk/services/TMHMM>

Prediction of transmembrane helices in proteins

NOTE: You can submit many proteins at once in one fasta file. Please limit each submission to at most 4000 proteins. Please tick the 'One line per protein' option. Please leave time between each large submission.



[Instructions](#)

SUBMISSION

Submission of a local file in **FASTA** format (HTML 3.0 or higher)

Browse... No file selected.

OR by pasting sequence(s) in **FASTA** format:

```
ANAFQRQYRDVDVLLIDDIQFLQGKEQTMEEFFHTFNTLHNSEKQIVITS
DQPPKLSGFAERMRSRFEWGLLTDVQPPDLETRIAILRRKAAADKLDIP
DDVLHLIASKISSNIRELEGALTRVTAFASLSGSPLDEYLARTVLKDVMP
GGDSGQITPTMILEETAGYFVISVEEIQGASRSRNLTRARQIAMYLCREL
TDLSLPKIGKEFGGRDHTVMHAERKIKQLLGEDRRVYDEVSELTSIIRK
KAARGR
```

Output format:

- Extensive, with graphics
- Extensive, no graphics
- One line per protein

Other options:

- Use old model (version 1)

Submit Clear

Enter "Protein Sequence" (from Module 1) in FASTA format

"CLICK"

Interpreting the TMHMM plot

Positive Result ★

Ksed_03810

GFF format

TMHMM result

[HELP](#) with output formats

```
# Ksed_ Length: 319
# Ksed_ Number of predicted TMHs: 6
# Ksed_ Exp number of AAs in TMHs: 133.3304
# Ksed_ Exp number, first 60 AAs: 10.89767
# Ksed_ Total prob of N-in: 0.99459
# Ksed_ POSSIBLE N-term signal sequence
Ksed_ TMHMM2.0 inside 1 49
Ksed_ TMHMM2.0 TMhelix 50 72
Ksed_ TMHMM2.0 outside 73 108
Ksed_ TMHMM2.0 TMhelix 109 131
Ksed_ TMHMM2.0 inside 132 143
Ksed_ TMHMM2.0 TMhelix 144 166
Ksed_ TMHMM2.0 outside 167 180
Ksed_ TMHMM2.0 TMhelix 181 203
Ksed_ TMHMM2.0 inside 204 223
Ksed_ TMHMM2.0 TMhelix 224 246
Ksed_ TMHMM2.0 outside 247 282
Ksed_ TMHMM2.0 TMhelix 283 305
Ksed_ TMHMM2.0 inside 306 319
```

the length of the protein sequence.

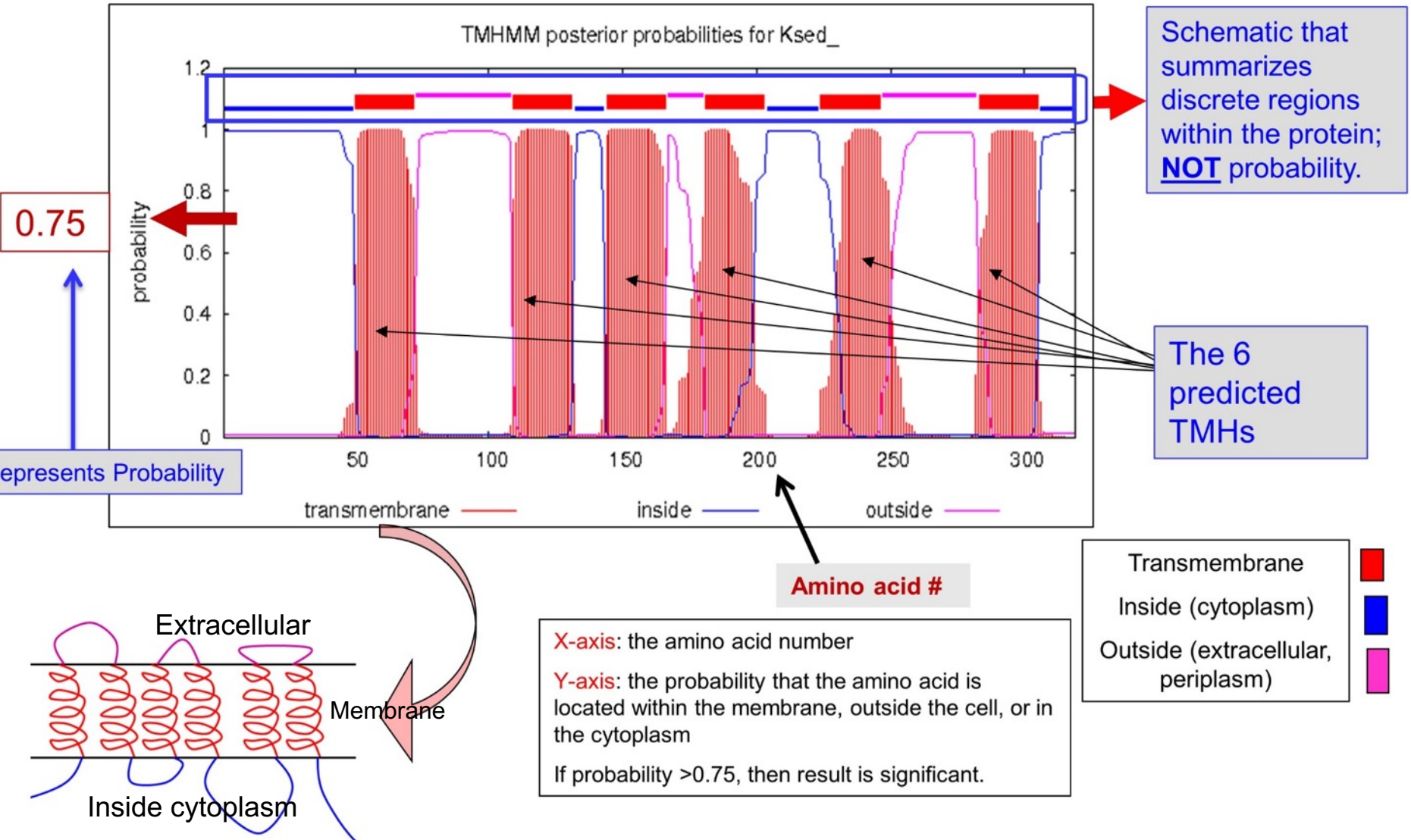
Predicted number of TMHs
(transmembrane helices)

If this number larger than 10, you should be warned that a predicted transmembrane helix in the N-term could be a signal peptide.

Total prob of N-in: The total probability that the N-term is on the cytoplasmic side of the membrane (inside)

Interpreting the TMHMM plot

★ Ksed_03810



Interpreting the TMHMM plot

Ksed_00010

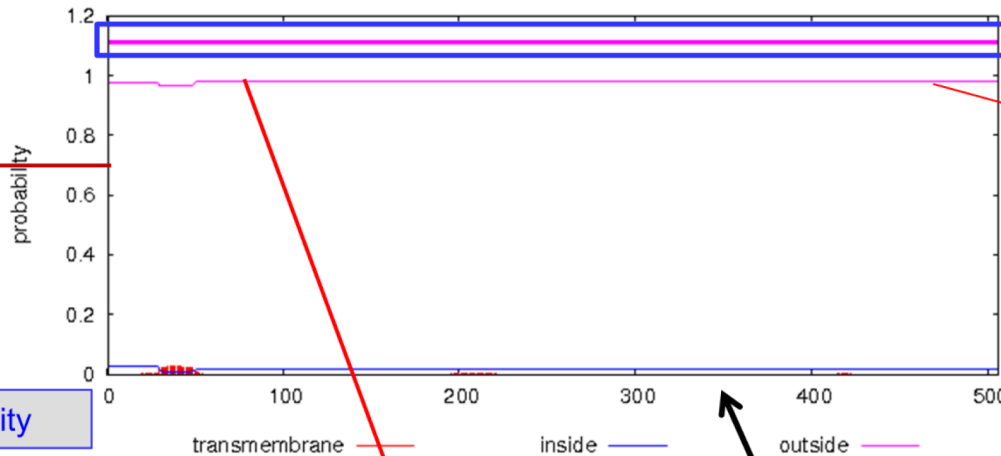
Negative Results

TMHMM result

[HELP](#) with output formats

```
# Ksed_00010amino_acid_sequence Length: 506
# Ksed_00010amino_acid_sequence Number of predicted TMHs: 0
# Ksed_00010amino_acid_sequence Exp number of AAs in TMHs: 0.52650000000000000001
# Ksed_00010amino_acid_sequence Exp number, first 60 AAs: 0.50155
# Ksed_00010amino_acid_sequence Total prob of N-in: 0.02536
Ksed_00010amino_acid_sequence TMHMM2.0 outside 1 506
```

TMHMM posterior probabilities for Ksed_00010amino_acid_sequence



0.75

Probability

Amino acid #

Schematic that summarizes discrete regions within the protein; not probability.

Transmembrane
Inside (cytoplasm)
Outside (extracellular, periplasm)

- Remember, this tool **ONLY** predicts the presence of transmembrane helices. It **DOES NOT** tell you anything else about the location of your protein.
- Sometimes students tend to interpret negative results predicting the protein is secreted, which is an incorrect interpretation.

Notebook

TMHMM

go to <http://www.cbs.dtu.dk/services/TMHMM/>

Number of predicted transmembrane helices

The transmembrane topology graph

Comments

Cellular Localization- 5 tools

➤ Is your protein of interest a membrane protein?

1. Transmembrane Helices Hidden Markov Models (TMHMM)

➤ Is your protein of interest a secreted protein? Does it have signal sequences?

2. SignalP

3. LipoP

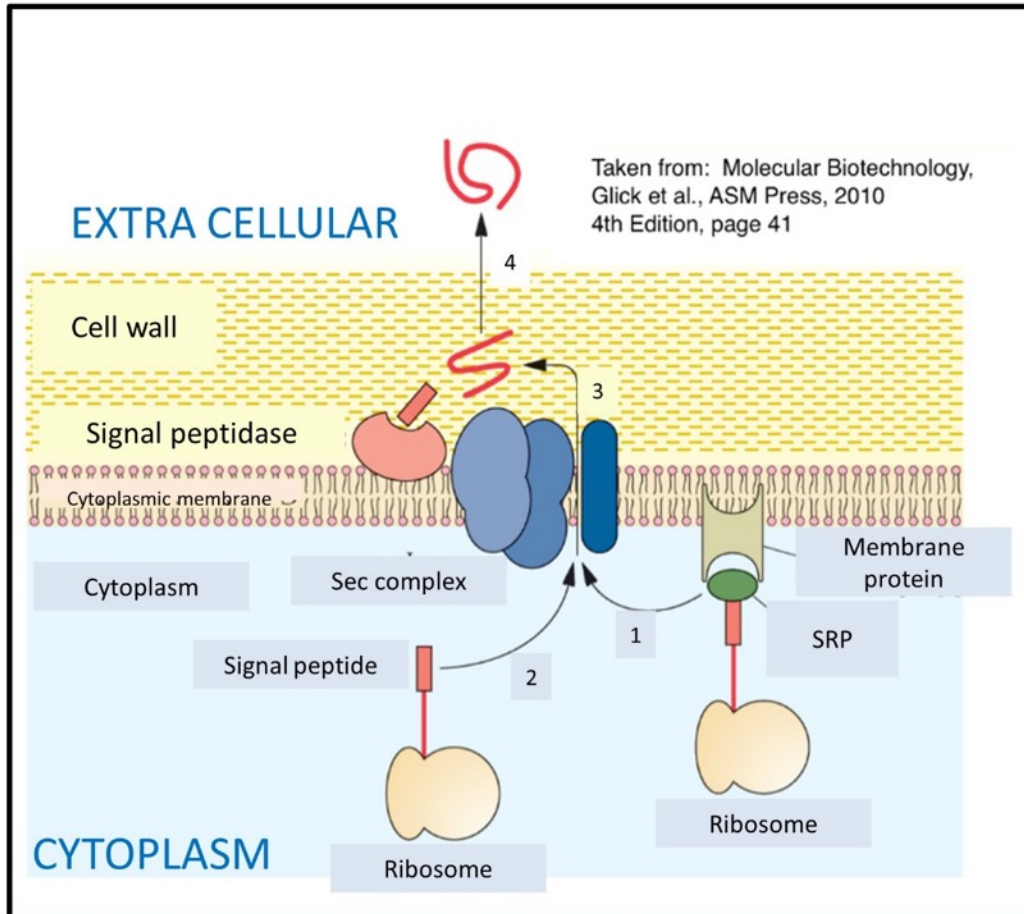
➤ Confirmatory tests

4. PSORT-B

5. Phobius

Final decision can only be made after using all 5 tools

What is a signal peptide?



- A **signal peptide** is a short (5-30 aa) peptide present at N-terminus of newly synthesized proteins that are destined to be **cleaved by proteases** before **secretion**.
- Secretion of proteins is crucial for **normal physiology** and **virulence** of pathogenic bacteria (**lipoproteins**).
- The efficiency of **protein secretion** is strongly determined by the **signal peptide** and **cleavage**.

Single TMH at N-terminus of protein may or may not get cleaved by proteases once inserted into membrane.

Preproinsulin -sig pep → proinsulin- c-peptide → insulin

2 Major Export Pathways in Bacteria

- Gram-positive bacteria usually possess only a single membrane (i.e. the cytoplasmic membrane) and export of a target protein across this major permeability barrier can directly result in its release into the culture supernatant.
- These microorganisms are especially useful as potential host organisms for the secretory production of industrially relevant recombinant proteins.
- Transport of proteins across the plasma membrane takes place through 2 major pathways:
 1. The general secretion or Sec pathway
 2. The twin-arginine translocation or Tat pathway,

The routing into one of these alternative protein export systems requires the fusion of a **Sec-** or **Tat-** specific **signal peptide** to the amino-terminal end of the desired target protein.

Signal **Peptidase** in Bacteria

- It is the enzyme that cleaves off the signal peptide of a secreted protein
- **Signal peptides** in Bacteria are mainly divided into:
 - Those cleaved by Signal peptidase I (Spase I or Spl)
 1. In **Gram⁺** bacteria- Tool: **SignalP**
 2. In **Gram⁻** bacteria- Tool: **SignalP** and **LipoP**
 - Those cleaved by Signal Peptidase II (Spase II or SplI)
 1. In **Gram⁺** bacteria- Tool: **LipoP**
 2. In **Gram⁻** bacteria- Tool: **LipoP**

SignalP

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

- ✓ Does my protein have a sequence of amino acids that target it to a particular place in or outside the cell?

Positive Example

```
>KSED_RS12585 (Ksed_25760) lipoprotein signal  
peptidase[Kytococcus sedentarius]  
MPRARASRALVLLAALAAAAIDLGAKAASEAQLAGSTVDLGVLRRLRLAYNSGVAFGMG  
DELPVGVIVAVTAVICLALIAYAWRRAPDAGWVERIAGGAVIGGAVANVVDRARDGTV  
TDYLHTGWWPTFNLADTFLVTGFIVIALHARP PERTADQA
```

SOP for SignalP

- Copy the amino acid sequence of your query protein under investigation.
- Open up the link
<http://www.cbs.dtu.dk/services/SignalP/>
- Paste the query sequence into the interactive window
- Select Organism group/ Long output
- Click Submit

SignalP-5.0 Server

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

Predict

Instructions/Help

Data

Abstract/Cite

FAQ

Submit data

The SignalP 5.0 server predicts the presence of signal peptides and the location of their cleavage sites in proteins from Archaea, Gram-positive Bacteria, Gram-negative Bacteria and Eukarya. In Bacteria and Archaea, SignalP 5.0 can discriminate between three types of signal peptides:

- Sec/SPI: "standard" secretory signal peptides transported by the Sec translocon and cleaved by Signal Peptidase I (*Lep*)
- Sec/SPII: lipoprotein signal peptides transported by the Sec translocon and cleaved by Signal Peptidase II (*Lsp*)
- Tat/SPI: Tat signal peptides transported by the Tat translocon and cleaved by Signal Peptidase I (*Lep*)

SignalP 5.0 is based on a deep convolutional and recurrent neural network architecture including a conditional random field.

i Protein sequences should be not less than 10 amino acids. The maximum number of proteins is 5000.

```
>KSED_RS12585 (Ksed_25760) lipoprotein signal peptidase[Kytococcus sedentarius]
MPRARASRALVLLAALAAAAIDLGAKAASEAQLAGSTVDLGLRLRLAYNSGVAFGMGDELPVGVIVAVTAVICLALAIYAWRRAPDAGWVERIAGGAVIGGA
VANVDRARDGTVTDYLHTGWWPTFNLADTFLVTGFIVIALHARPETADQA
```

Keep in mind...
Kytococcus is a
 Gram-positive
 bacteria

Signal peptide should be
 in N-terminus of your protein;

Organism group:

- Eukarya
- Gram-positive
- Gram-negative
- Archaea

Output format:

- Long output
- Short output (no figures)

Upload Fasta File

Example proteins

Submit

Reset

CLICK!

Signal IP (Gram +) Negative result

Summary of 1 predicted sequences from Gram-positive

Predictions list. Use the [help page](#) for more detailed description of the output page.

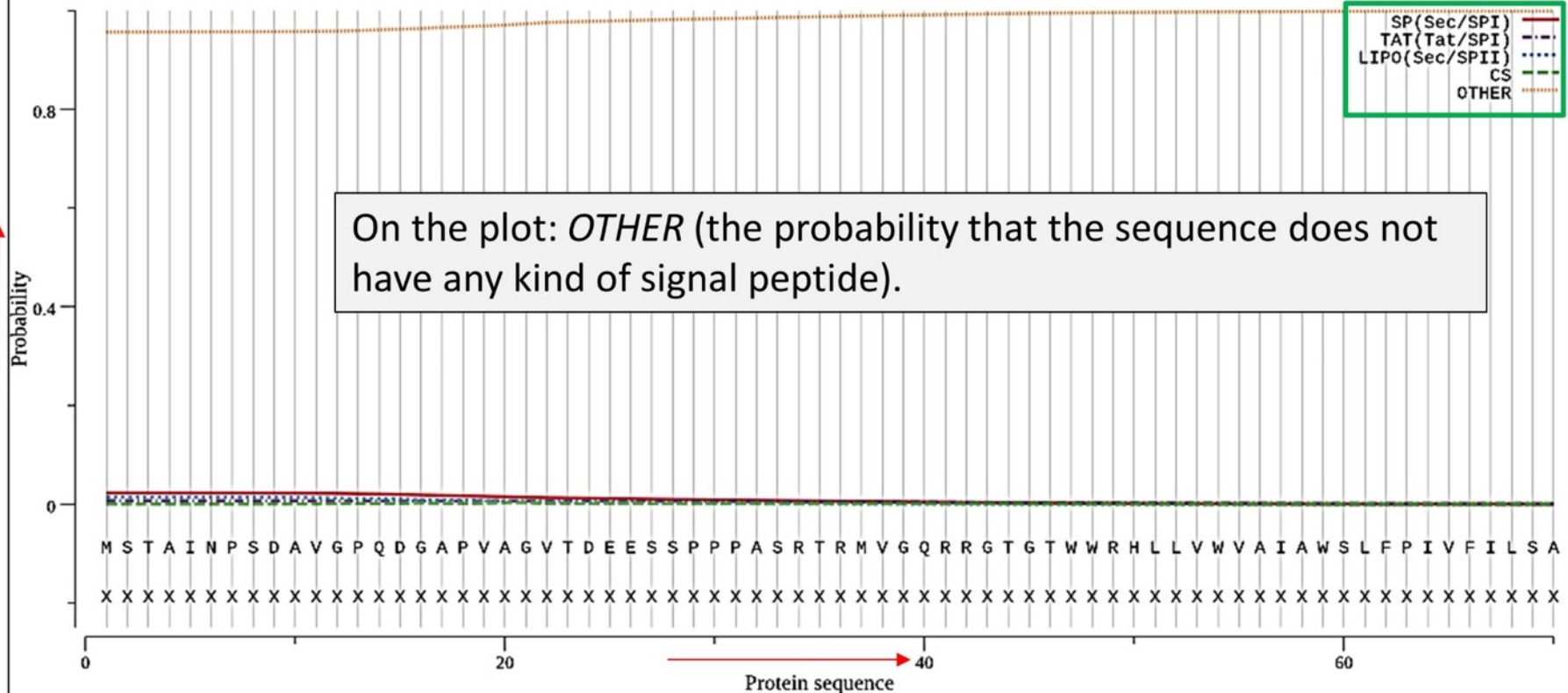
Predicted proteins

Ksed_03810aa
Prediction: Other

Protein type	Signal peptide (Sec/SPI)	TAT signal peptide (Tat/SPI)	Lipoprotein signal peptide (Sec/SPII)	Other
Likelihood	0.0221	0.0062	0.017	0.9547

Download: [PNG](#) / [EPS](#) / [Tabular](#)

SignalP-5.0 prediction (Gram-positive): Ksed_03810aa



Notebook

SignalP

go to <http://www.cbs.dtu.dk/services/SignalP/>

Signal peptide probability

Most likely cleavage site (between position # and #)

Signal peptide graph

LipoP

Go to LipoP at <http://www.cbs.dtu.dk/services/LipoP/>

Best Prediction (options = cytoplasm, signal peptidase I, signal peptidases II)

Cleavage site, if predicted, after AA #

LipoP

- LipoP is a tool that uses HMM to classify an N-terminal protein sequence as a:
 - Spl cleaved lipoprotein signal peptide
 - SplI cleaved membrane bound lipoprotein signal peptide
 - protein without a signal sequence (cytoplasmic or transmembrane)
- The pro-lipoprotein is exported out of the cell membrane
- In Gram⁻ bacteria the lipoproteins are processed in the periplasmic space and can remain attached to either the inner or outer membrane.

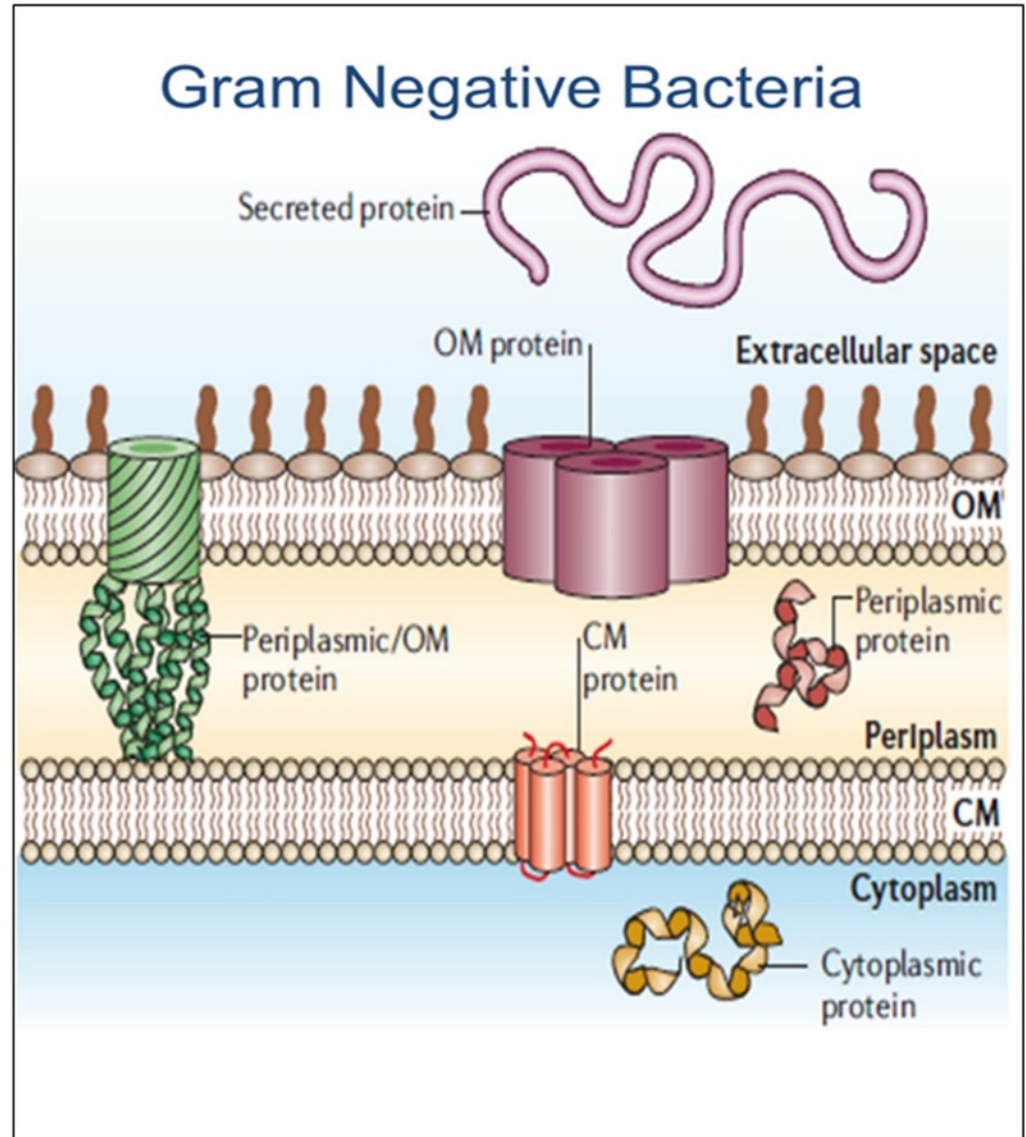
Signal **Peptidase** in Bacteria

- It is the enzyme that cleaves off the signal peptide of a secreted protein
- **Signal peptides** in Bacteria are mainly divided into:
 - Those cleaved by Signal peptidase I (Spase I or Spl)
 1. In **Gram⁺** bacteria- Tool: **SignalP**
 2. In **Gram⁻** bacteria- Tool: **SignalP** and **LipoP**
 - Those cleaved by Signal Peptidase II (Spase II or SplI)
 1. In **Gram⁺** bacteria- Tool: **LipoP**
 2. In **Gram⁻** bacteria- Tool: **LipoP**

LipoP

- ✓ Is my protein a lipoprotein?
- ✓ Is it cleaved by Spl or SplI?
- ✓ Is it likely to be anchored to the inner or outer membrane of the periplasmic space?

Lipoproteins have been shown to play key roles in adhesion to host cells, modulation of inflammatory processes, and translocation of virulence factors into host cells. As such, a number of lipoproteins have been shown to be potential vaccines.



LipoP 1.0 Server

<http://www.cbs.dtu.dk/services/LipoP/>

DTU Bioinformatics
Department of Bio and Health Informatics

[Home](#)

LipoP 1.0 Server

Prediction of lipoproteins and signal peptides in Gram negative bacteria

The LipoP 1.0 server produces predictions of lipoproteins and discriminates between lipoprotein signal peptides, other signal peptides and n-terminal membrane helices in Gram-negative bacteria.

Note: Although LipoP 1.0 has been trained on sequences from Gram-negative bacteria only, the following paper reports that it has a good performance on sequences from Gram-positive bacteria also:

[Methods for the bioinformatic identification of bacterial lipoproteins encoded in the genomes of Gram-positive bacteria](#)

O. Rahman, S. P. Cummings, D. J. Harrington and I. C. Sutcliffe

World Journal of Microbiology and Biotechnology **24**(11):2377-2382 (2008)

[Instructions](#)

[Output format](#)

[Article abstract](#)

SUBMISSION

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

```
>644990675 YP_003148220 carbohydrate ABC transporter substrate-  
binding protein]  
MKRNTLSLAASVSI AALTLSACGGDSESSDSETSSGNGATSQESPVEGS GDESSD  
PAEDES GSGDESSAPAEDEPAAGESSAPAAAPKR DENADLVIWA DDLRTEALNEVK
```

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

No file selected.

Output format: Extensive, with graphics
 Extensive, no graphics
 Short

Paste your query
sequence in interactive
input window

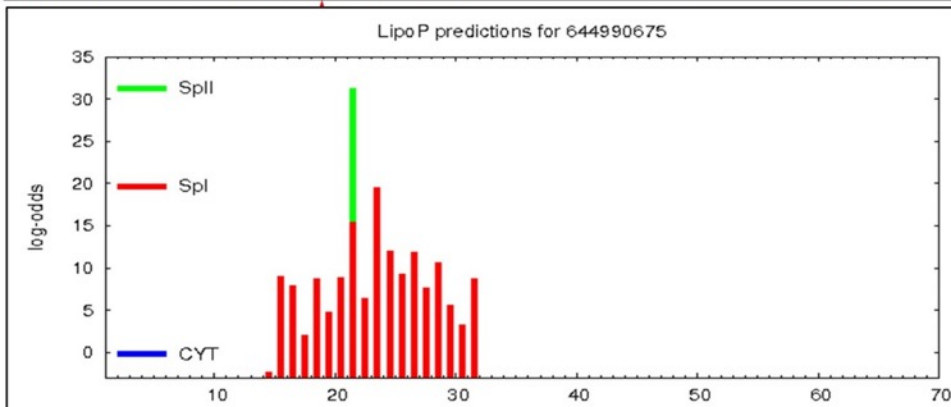
CLICK!

Carbohydrate ABC transporter substrate-binding protein *Kytococcus sedentarius*

Positive Example for SpII

GFF format

```
# 644990675 SpII score=31.3314 margin=11.647 cleavage=21-22 Pos+2=G
#-Cut-off=-3
644990675 LipoPl.0:Best SpII 1 1 31.3314
644990675 LipoPl.0:Margin SpII 1 1 11.647
644990675 LipoPl.0:Class SpI 1 1 19.6844
644990675 LipoPl.0:Class CYT 1 1 -0.200913
644990675 LipoPl.0:Signal CleavII 21 22 31.3314 # LTLSA|CGGDS Pos+2=G
644990675 LipoPl.0:Signal CleavI 23 24 19.5855 # LSACG|GDSES
644990675 LipoPl.0:Signal CleavI 21 22 15.4014 # LTLSA|CGGDS
644990675 LipoPl.0:Signal CleavI 24 25 11.9692 # SACGG|DSESS
644990675 LipoPl.0:Signal CleavI 26 27 11.8944 # CGGDS|ESSDS
644990675 LipoPl.0:Signal CleavI 28 29 10.6417 # GDSES|SDSET
644990675 LipoPl.0:Signal CleavI 25 26 9.33579 # ACGGD|SESSD
644990675 LipoPl.0:Signal CleavI 15 16 8.96264 # SVSIA|ALTLS
644990675 LipoPl.0:Signal CleavI 20 21 8.8873 # ALTLS|ACGGD
644990675 LipoPl.0:Signal CleavI 18 19 8.78613 # IAALT|LSACG
644990675 LipoPl.0:Signal CleavI 31 32 8.76227 # ESSDS|ETSSG
644990675 LipoPl.0:Signal CleavI 16 17 7.87768 # VSIAA|LTLSA
644990675 LipoPl.0:Signal CleavI 27 28 7.66876 # GGDSE|SSDSE
644990675 LipoPl.0:Signal CleavI 22 23 6.40791 # TLSAC|GGDSE
644990675 LipoPl.0:Signal CleavI 29 30 5.5401 # DSESS|DSETS
644990675 LipoPl.0:Signal CleavI 19 20 4.80837 # AALTL|SACGG
644990675 LipoPl.0:Signal CleavI 30 31 3.22365 # SESSD|SETSS
644990675 LipoPl.0:Signal CleavI 17 18 2.08437 # SIAAL|TLSAC
644990675 LipoPl.0:Signal CleavI 14 15 -2.35921 # ASVSI|AALTL
```



Prediction: High probability there is a SP II cleavage site between aa 21 and 22. This is attached to the **outer membrane** because Pos+2 is G.

Columns:

1. Seq ID
2. Type of prediction. Best means the highest scoring class.
3. Feature type: Cleav I = SpI, CleavII = SpII.
4. Location in sequence last aa rel cleave site.
5. Location in seq first aa after cleave site.
6. Score margin give diff bet first and second best.
7. For cleav sites +/- 5 context after # and for SpII lipoprotein cleavage aa in +2 shows inner or outer membrane attachment. An aspartic acid (D) in position +2 after cleavage site means it is attached to the inner membrane.

The 4 classes predicted are:

- ✓ **SpI**: signal peptide (signal peptidase I)
- ✓ **SpII**: lipoprotein signal peptide (signal peptidase II)

TMH: n-terminal transmembrane helix. **This is generally not a very reliable prediction and should be tested.**

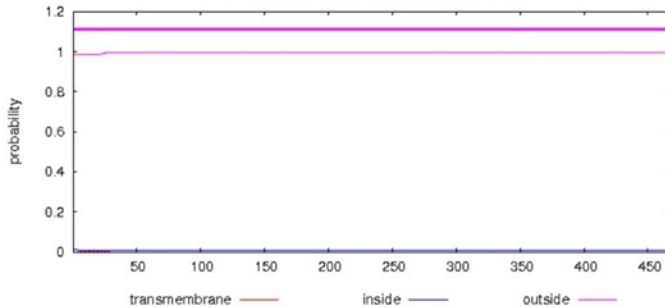
This part of the model is mainly there to avoid transmembrane helices being falsely predicted as signal peptides.

CYT: cytoplasmic. It really just means all the rest.

Carbohydrate ABC transporter substrate-binding protein *Kytococcus sedentarius*

1 TMHMM

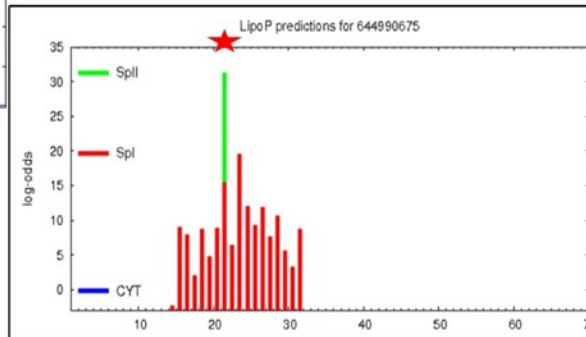
TMHMM posterior probabilities for 644990675



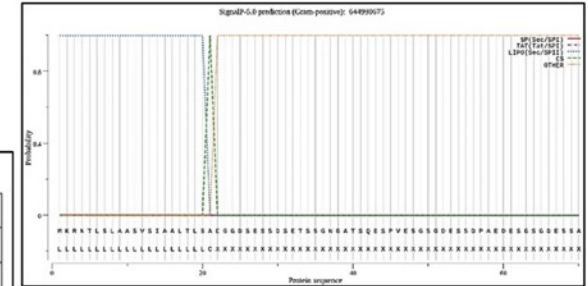
Transmembrane helix 0

Positive Example

3 LipoP- SpII



2 SignalP



Lipoprotein signal peptide
Sec/SPII Pos 21 and 22 LSA-
CG probability 0.9985

Prediction: High probability there is a cleavage site between aa 21 and 22 and the secretory signal peptide is cleaved by Signal Peptidase II (SPase II) class. The (C) cysteine at position +1 (22 aa) is required and is anchor point for the protein in the membrane with the rest existing extracellularly.

```
# 644990675 SpII score=31.3314 margin=11.647 cleavage=21-22 Pos+2=G
# Cut-off=-3
644990675 LipoP1.0:Best SpII 1 1 31.3314
644990675 LipoP1.0:Margin SpII 1 1 11.647
644990675 LipoP1.0:Class SpI 1 1 19.6844
644990675 LipoP1.0:Class CYT 1 1 -0.200913
644990675 LipoP1.0:Signal CleavII 21 22 31.3314 # LTLSA|CGGDS Pos+2=G
644990675 LipoP1.0:Signal CleavI 23 24 19.5855 # LSACG|GDSES
644990675 LipoP1.0:Signal CleavI 21 22 15.4014 # LTLSA|CGGDS
644990675 LipoP1.0:Signal CleavI 24 25 11.9692 # SACGG|DSESS
644990675 LipoP1.0:Signal CleavI 26 27 11.8944 # CGGDS|ESSDS
644990675 LipoP1.0:Signal CleavI 28 29 10.6417 # GDSES|DSESS
644990675 LipoP1.0:Signal CleavI 25 26 9.33579 # ACGGD|ESSSD
644990675 LipoP1.0:Signal CleavI 15 16 8.96264 # SVSIA|ALTLS
644990675 LipoP1.0:Signal CleavI 20 21 8.8873 # ALTLS|ACGGD
644990675 LipoP1.0:Signal CleavI 18 19 8.78613 # IAALT|LSACG
644990675 LipoP1.0:Signal CleavI 31 32 8.76227 # ESSDS|ETSSG
644990675 LipoP1.0:Signal CleavI 16 17 7.87768 # VSIAA|LTLSA
644990675 LipoP1.0:Signal CleavI 27 28 7.66876 # GGDSE|SSDSE
644990675 LipoP1.0:Signal CleavI 22 23 6.40791 # TLSAC|GGDSE
644990675 LipoP1.0:Signal CleavI 29 30 5.5401 # DSESS|DSETS
644990675 LipoP1.0:Signal CleavI 19 20 4.80837 # AALT|SACGG
644990675 LipoP1.0:Signal CleavI 30 31 3.22365 # SESSD|SETSS
644990675 LipoP1.0:Signal CleavI 17 18 2.08437 # SIAAL|TLSAC
644990675 LipoP1.0:Signal CleavI 14 15 -2.35921 # ASVSI|AALT
```

SpII site predicted between aa 21 and 22 – attached to outer membrane

G (glycine) in position +2 shows the lipoprotein is attached to the outer membrane.

Cellular Localization- 5 tools

- Is your protein of interest a membrane protein?
 1. Transmembrane alpha helices (TMHMM)
- Is your protein of interest a secreted protein? Does it have signal sequences?
 2. SignalP
 3. LipoP (is my protein a lipoprotein? Is it cleaved by Spl or SplI. Is it likely to be anchored to the inner or outer membrane of the periplasmic space?)
- **Confirmatory tests**
 4. PSORT-B
 5. Phobius

Final decision can only be made after using all 5 tools

PSORT-B

Question: Where is my protein most likely located? The cytoplasm? The membrane? The periplasm? The cell wall? The extracellular space?

➤ The Normal output option displays:

1. Results of each of PSORTb's 12 analytical modules
2. Localization scores for each of the 5 sites
3. Final prediction and associated score (if one site scores above the 7.5 cut off)



PSORT-b

[Updates](#) | [Documentation](#) | [Resources](#) | [Contact](#)

<http://www.psort.org/psortb/>

Choose an organism
type (?):

Bacteria

Choose Gram stain (?):

Positive

Output format (?):

Normal

Show results (?):

Via the web

Copy and paste your FASTA sequences below

```
>Ksed_03810aa Carbohydrate ABC transporter membrane protein
MSTAINPSDAVGPODGAPVAGVTDEESSPPPASRTRMVGORRGTTWWRH
LLVWVAIAWSLFPPIVFILSAALNPAGTLATSSLIPSGFSLENWRELFETR
PYWTWYRNALVISLVATAGAVFIGACAAYAFSRLRFTGRRPGLFALLLVO
MFPALLTFVALYFTMVRVGEIIPAIGLNTSLGLILVYLGGAMGANIWLLK
GYFDTVPRELDEAATIDGASHARIFFTMTLRLVAPILVTVAMLAFVQFWG
EFMLASIFLKDADAQTLGVGLWQMOOGNEKNAQFGEFAAGALLASIPVVV
LYLVFORQLVSGLTSGSVK
```

Paste the aa
sequence in
FASTA format

or upload from file:

No file selected.

(uploads limited to 50KB, approximately 100 proteins, in Web display mode, enter an email address to use email mode if you need to analyze more proteins)

CLICK!

Recording results in your Lab Notebook



Ksed_03810

Positive Example

PSORTb Results ([Click here for an explanation](#))

SeqID: 2500607071 Nitrate/nitrite transporter
Analysis Report:

12 analytical modules

CMSVM-	CytoplasmicMembrane
CytoSVM-	Unknown
ECSVM-	Unknown
HMMTOP-	CytoplasmicMembrane
Motif-	Unknown
OMPMotif-	Unknown
OMSVM-	Unknown
PPSVM-	Unknown
Profile-	CytoplasmicMembrane
SCL-BLAST-	Unknown
SCL-BLASTe-	Unknown
Signal-	Unknown

Localization Scores:

Cytoplasmic	0.00
CytoplasmicMembrane	10.00
Periplasmic	0.00
OuterMembrane	0.00
Extracellular	0.00

Final Prediction:

CytoplasmicMembrane	10.00
---------------------	-------

Localization scores for each of the 5 sites

Enter in your Lab Notebook

PSORT

go to <http://www.psort.org/psortb/>

Cytoplasmic score

0.00

CytoplasmicMembrane score

10.0

Periplasmic score

0.00

OuterMembrane score

0.00

Extracellular score

0.00

PSORT prediction.

Cytoplasmic Membrane

Where this protein is predicted to be located in the cell

Recording results in your Lab Notebook



Ksed_10290 Hypothetical protein

Negative Example

[Submit Sequences](#) | [Documentation](#) | [Resources](#) | [Contact](#) | [Upd](#)

PSORTb Results ([Click here for an explanation of the output formats](#))

SeqID: Ksed_10290 amino acid sequence

Analysis Report:

CMSVM+	Unknown	[No details]
CNSVM+	Unknown	[No details]
Cyto3VM+	Unknown	[No details]
ECSVM+	Unknown	[No details]
ModIDM+	Unknown	[No internal helices found]
Motif+	Unknown	[No motifs found]
Profile+	Unknown	[No matches to profiles found]
SCL-BLAST+	Unknown	[No matches against database]
SCL-BLASTe+	Unknown	[No matches against database]
Signal+	Unknown	[No signal peptide detected]

Localisation Scores:

Cytoplasmic	2.50
CytoplasmicMembrane	2.50
Cellwall	2.50
Extracellular	2.50

Final Prediction:
Unknown

Localization Scores:

Cytoplasmic	2.50
CytoplasmicMembrane	2.50
Cellwall	2.50
Extracellular	2.50

Final Prediction:

Unknown

Enter in your
Lab Notebook

Where is this protein predicted to
be located in the cell ?

UNKNOWN by this tool

Cellular Localization- 5 tools

➤ Is your protein of interest a membrane protein?

1. Transmembrane alpha helices (TMHMM)

➤ Is your protein of interest a secreted protein? Does it have signal sequences?

2. SignalP

3. LipoP (is my protein a lipoprotein? Is it cleaved by Spl or SplI. Is it likely to be anchored to the inner or outer membrane of the periplasmic space?)

➤ **Confirmatory tests**

4. PSORT-B

5. Phobius

Final decision can only be made after using all 5 tools

Phobius

- ✓ Does my protein have transmembrane helices & signal peptide? Do these results agree with **TMHMM** and **SignalP**)

<http://phobius.sbc.su.se/>

- Graphical output
- Combination of **transmembrane topology** (TMHMM) and **signal peptide predictor** (SignalP)

Phobius



POST NEBULA PHOBIUS

A combined transmembrane topology and signal peptide predictor

[Normal prediction](#) [Constrained prediction](#) [PolyPhobius](#) [Instructions](#) [Download](#) [Mirror site at KU](#)

Normal prediction

Paste your protein sequence here in Fasta format:

```
>2500607071 Nitrate/nitrite transporter [Planctomyces limnophilus DSM 3776 :  
P1imDRAFT_4083246_C168]  
MTTSAKATSIRLWDFKTPPMRAFHMSWFAFFLCFFAWFGIAPLMPVVRDE  
MHLSKDQVGWCIIGSVAITVLARLYVGVLCDRIGPRLAYSGLLVLASIPV  
MGIGLAHDFTTFLMFRIAIGAIGASFVITQYHTSIMFAKNCVGTANATTA  
GWGNLGGGVTQMVMPTLFALLMVAFGLSTASSURFCMLLAGVVCAITGIA  
YFFLTQDTPEGNFAELRATGKMSQKSAVKGTFQEACRDYRVWILFLVYGA
```

Or: Select the sequence file you wish to use Browse...

Select output format:

- Short
- Long without Graphics
- Long with Graphics

“Click”

Submit Query Reset

Copy/paste your amino acid sequence in Fasta format (from Module 1)

Query Results

Ksed_03810

Text listing predicted locations of TMHs, intervening loops, and signal peptide

Prediction of Ksed_

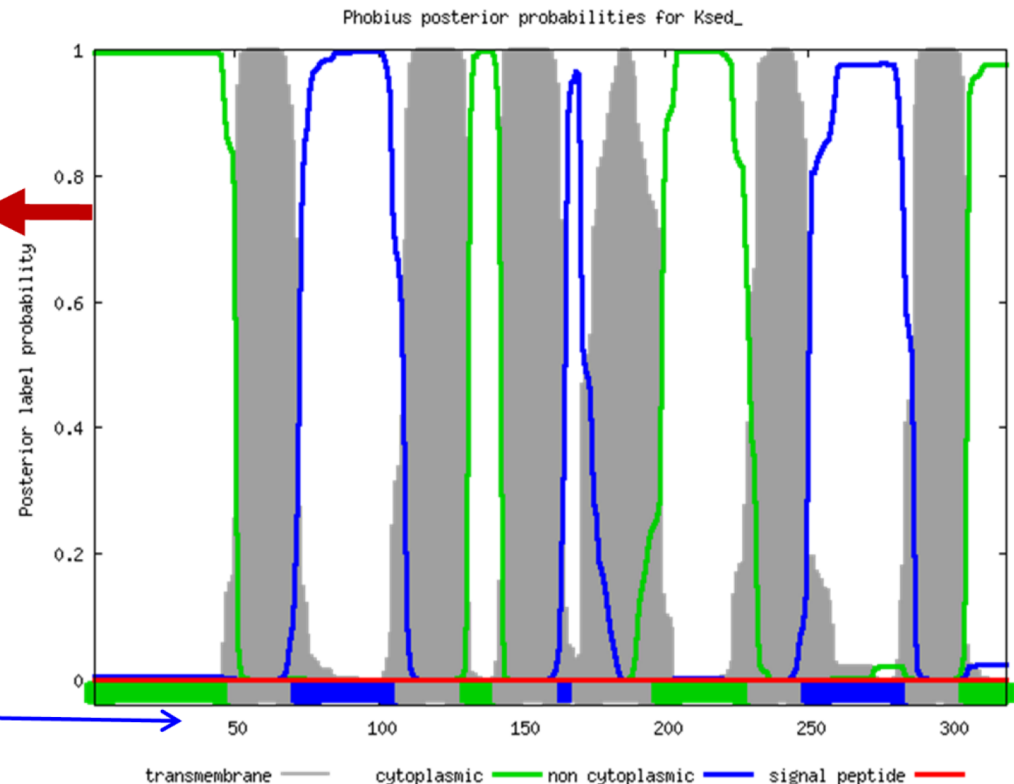
ID	Ksed_			
FT	TOPO_DOM	1	50	CYTOPLASMIC.
FT	TRANSMEM	51	72	
FT	TOPO_DOM	73	108	NON CYTOPLASMIC.
FT	TRANSMEM	109	131	
FT	TOPO_DOM	132	142	CYTOPLASMIC.
FT	TRANSMEM	143	165	
FT	TOPO_DOM	166	170	NON CYTOPLASMIC.
FT	TRANSMEM	171	198	
FT	TOPO_DOM	199	231	CYTOPLASMIC.
FT	TRANSMEM	232	250	
FT	TOPO_DOM	251	286	NON CYTOPLASMIC.
FT	TRANSMEM	287	305	
FT	TOPO_DOM	306	319	CYTOPLASMIC.
//				

Graphical summary

Y axis shows probability

X axis shows amino acid position

0.75



Notebook

Phobius

go to <http://phobius.sbc.su.se/>

Phobius probability graph

Hypothesis

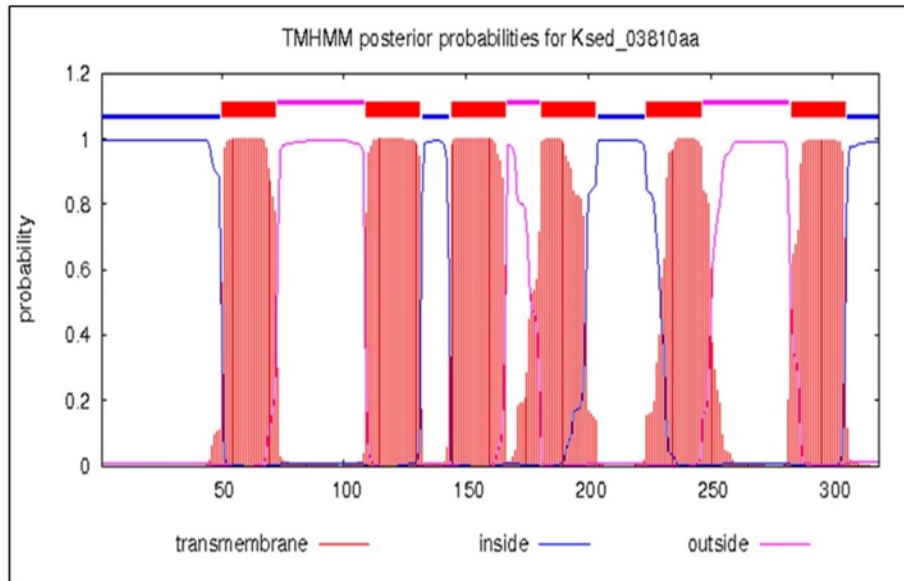
Where do you expect to find this protein?

Be sure to save this document after completing the cellular localization module!

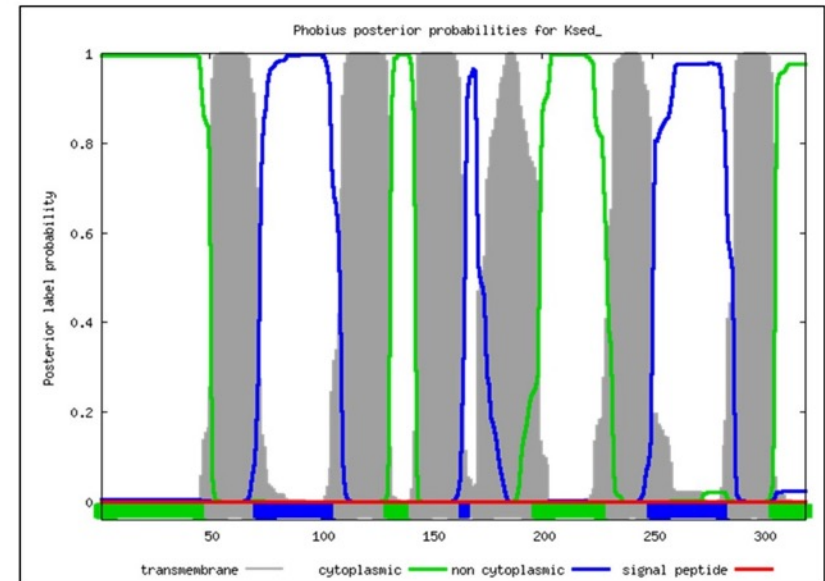
Interpreting the Phobius Plot

>Ksed_03810

TMHMM



Phobius



Interpreting the Phobius Plot

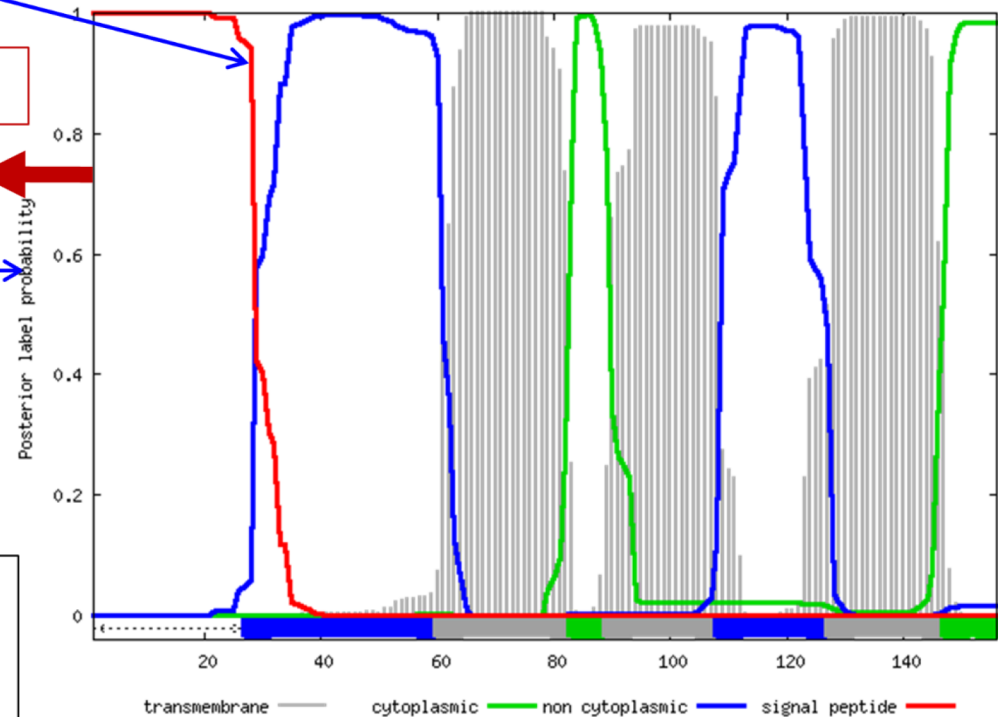
>Ksed_25760

Signal Peptide

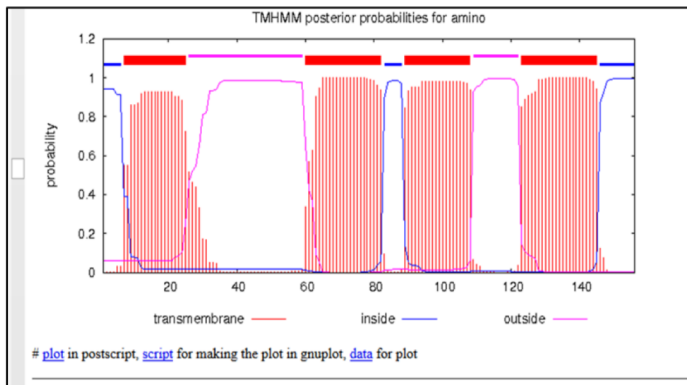
0.75

Y axis shows probability

Phobius posterior probabilities for amino



TMHMM Results



Finishing Up Module 3: Hypothesis

Now that you've finished TMHMM, SignalP, LipoP, PSORT, and Phobius, you should have an idea about the **cellular localization** of the protein encoding by your gene.

THINGS TO CONSIDER:

- Did TMHMM indicate any transmembrane helices? If so, how many?
- Did SignalP show evidence of a Sec- or Tat- signal peptide and cleavage site at N-terminus?
- Did LipoP show Spl or SpII cleavage site?
- Where did PSORT-B predict the protein was located in the cell?
- Were Phobius results consistent with TMHMM and SignalP results?

Enter your conclusion about where you would expect to find the protein under the **Hypothesis** section of Module 3

Hypothesis

I expect to find this protein Ksed_03810 for which the top hit in BLAST identifies a carbohydrate ABC transporter membrane protein, in the cytoplasmic membranes. The TMHMM and Phobius tools agree that there is a high probability of 6 transmembrane helices in the protein. This is consistent with the localization according to PSORT-B in the cytoplasmic membrane (High score of 10.00). The tool Signal IP suggests that there is no cleavage site at the N-terminal end and thus no signal peptide, using the Gram + database. LipoP also shows no Spl or SpII cleavage sites. Although a signal peptide for this protein has been found in other bacteria it is absent in Kytococcus. This brings us to question whether Kytococcus has a different or additional secretory pathway that can explain this? Interesting.