Where in the cell is your protein most likely found?
The major intracellular compartments of an animal cell
Relative Volumes Occupied by the Major Intracellular Compartments

<table>
<thead>
<tr>
<th>Intracellular Compartment</th>
<th>Percentage of Total Cell Volume</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cytosol</td>
<td>54</td>
</tr>
<tr>
<td>Mitochondria</td>
<td>22</td>
</tr>
<tr>
<td>Rough ER cisternae</td>
<td>9</td>
</tr>
<tr>
<td>Smooth ER cisternae plus Golgi cisternae</td>
<td>6</td>
</tr>
<tr>
<td>Nucleus</td>
<td>6</td>
</tr>
<tr>
<td>Peroxisomes</td>
<td>1</td>
</tr>
<tr>
<td>Lysosomes</td>
<td>1</td>
</tr>
<tr>
<td>Endosomes</td>
<td>1</td>
</tr>
</tbody>
</table>
An electron micrograph
Sorting sequences

(A) Unfolded protein

(B) Folded protein

Regions contributing to signal patch
Some sorting sequences

<table>
<thead>
<tr>
<th>FUNCTION OF SIGNAL SEQUENCE</th>
<th>EXAMPLE OF SIGNAL SEQUENCE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Import into nucleus</td>
<td>-Pro-Pro-Lys-Lys-Lys-Arg-Lys-Val-</td>
</tr>
<tr>
<td>Export from nucleus</td>
<td>-Leu-Ala-Leu-Lys-Leu-Ala-Gly-Leu-Asp-Ile-</td>
</tr>
<tr>
<td>Import into mitochondria</td>
<td>&quot;H_3N-Met-Leu-Ser-Leu-Arg-Gln-Ser-Ile-Arg-Phe-Phe-Lys-Pro-Ala-Thr-Arg-Thr-Leu-Cys-Ser-Ser-Arg-Tyr-Leu-Leu-</td>
</tr>
<tr>
<td>Import into plastid</td>
<td>&quot;H_3N-Met-Val-Ala-Met-Ala-Met-Ala-Ser-Leu-Gln-Ser-Ser-Met-Ser-Ser-Leu-Ser-Ser-Asn-Ser-Phe-Leu-Gly-Gln-Pro-Leu-Ser-Pro-Ile-Thr-Leu-Ser-Pro-Phe-Leu-Gln-Gly-</td>
</tr>
<tr>
<td>Import into peroxisomes</td>
<td>-Ser-Lys-Leu-COO^-</td>
</tr>
<tr>
<td>Import into ER</td>
<td>&quot;H_3N-Met-Met-Ser-Phe-Val-Ser-Leu-Leu-Leu-Val-Gly-Ile-Leu-Phe-Trp-Ala-Thr-Glu-Ala-Glu-Gln-Leu-Thr-Lys-Cys-Glu-Val-Phe-Gln-</td>
</tr>
<tr>
<td>Return to ER</td>
<td>-Lys-Asp-Glu-Leu-COO^-</td>
</tr>
</tbody>
</table>

Some characteristic features of the different classes of signal sequences are highlighted in color. Where they are known to be important for the function of the signal sequence, positively charged amino acids are shown in red and negatively charged amino acids are shown in green. Similarly, important hydrophobic amino acids are shown in yellow and hydroxylated amino acids are shown in blue. "H_3N indicates the N-terminus of a protein; COO^- indicates the C-terminus.
How do we figure out where proteins are located?

- **Transmembrane Helices Hidden Markov Models (TMHMM)**
  - Does my protein have transmembrane helices?

- **Signal Peptide (SignalP)**
  - Does my protein have a sequence of amino acids that target it to a particular place in or outside the cell?

- **PSORT-B**
  - Where is my protein most likely located? The cytoplasm? The membrane? The periplasm? The cell wall? The extracellular space?

- **Phobius**
  - Does my protein have transmembrane helices & signal peptides? Do these results agree with TMHMM and SignalP?
Transmembrane Helices Hidden Markov Models (TMHMM)

- A Hidden Markov Model is a probabilistic model developed from observed sequences of proteins of a known function.

- TMHMM is a tool used to predict the presence of transmembrane helices in proteins. The results will indicate the segments of the protein that lie inside, outside or within the membrane.

TMHMM

- go to [http://www.cbs.dtu.dk/services/TMHMM/](http://www.cbs.dtu.dk/services/TMHMM/)
- enter the number of predicted TMH's

Enter in Lab Report.
TMHMM Database Search

SUBMISSION

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

OR by pasting sequence(s) in FASTA format:
>2500607071 Nitrate/nitrite transporter [Planctomyces limnophilus DSM 3776 : PlimDRAFT_4083246_C168]
MTTSAKATSIRLUDFKTPPRHMSWAFFLCFFAUFGLAIPLMVPVRDE
MHLSKQVGCISVATVLAPRLYVGWLCDDRGPLRASGLVLLASIPV
MGIGLADFTTFLMFRIAIGASFVITQYHTSIMFAKNCVGTANATTA
GWGNLGGGVTQMVMTPLFALLMVAFGLSTASSWRCMLLLAGVVCAITGIA

Output format:
- Extensive, with graphics
- Extensive, no graphics
- One line per protein

Other options:
- Use old model (version 1)

Submit  Clear

**Make sure Javascript is enabled on your computer to read output**
# TMHMM result

**HELP with output formats**

---

| # 2500607071 Length: 458 |
| # 2500607071 Number of predicted TMHs: 12 |
| # 2500607071 Exp number of AAS in TMHs: 263.03045 |
| # 2500607071 Exp number, first 60 AAs: 25.40106 |
| # 2500607071 Total prob of N-in: 0.99853 |
| # 2500607071 POSSIBLE N-term signal sequence |

| 2500607071 | TMHMM2.0 | inside 1 20 |
| 2500607071 | TMHMM2.0 | outside 44 57 |
| 2500607071 | TMHMM2.0 | TMHelix 21 43 |
| 2500607071 | TMHMM2.0 | inside 81 96 |
| 2500607071 | TMHMM2.0 | TMHelix 87 106 |
| 2500607071 | TMHMM2.0 | outside 107 115 |
| 2500607071 | TMHMM2.0 | TMHelix 116 138 |
| 2500607071 | TMHMM2.0 | inside 139 150 |
| 2500607071 | TMHMM2.0 | TMHelix 151 173 |
| 2500607071 | TMHMM2.0 | outside 174 182 |
| 2500607071 | TMHMM2.0 | TMHelix 183 205 |
| 2500607071 | TMHMM2.0 | inside 206 240 |
| 2500607071 | TMHMM2.0 | TMHelix 241 263 |
| 2500607071 | TMHMM2.0 | outside 264 277 |
| 2500607071 | TMHMM2.0 | TMHelix 278 300 |
| 2500607071 | TMHMM2.0 | inside 301 319 |
| 2500607071 | TMHMM2.0 | TMHelix 320 337 |
| 2500607071 | TMHMM2.0 | outside 338 346 |
| 2500607071 | TMHMM2.0 | TMHelix 347 369 |
| 2500607071 | TMHMM2.0 | inside 370 375 |
| 2500607071 | TMHMM2.0 | TMHelix 376 398 |
| 2500607071 | TMHMM2.0 | outside 399 401 |
| 2500607071 | TMHMM2.0 | TMHelix 402 424 |
| 2500607071 | TMHMM2.0 | inside 425 458 |

**Predicted number of TMHs (transmembrane helices)**

**Boundaries for THM amino acids**

**Copy/paste this information into the box in your lab notebook**
Interpreting the TMHMM plot

X-axis: the amino acid number

Y-axis: the probability that the amino acid is located within the membrane, outside the cell, or in the cytoplasm

Ex: If probability >0.75, then result is significant. The maximum probability is 1, so the probability that amino acids #1-#20 are “inside” is 100%
By analyzing the probabilities shown on the plot, you can determine where segments within the protein are located.
Inserting the TMHMM plot into your notebook

Save image in GIF format to your computer and insert into Lab Notebook
• **A Signal Peptide (SignalP)** is a series of amino acids in the polypeptide that directs the protein to its proper cellular location.

  • **Ex:** Single TMH at N-terminus of protein that gets cleaved by proteases once inserted into membrane.

**SignalP**


enter the signal peptide probability

Enter in Lab Report.

most likely cleavage site (between position # and #)

```
text/#
```

insert the signal peptide graph

```
image
```

---

**Locating proteins in the cell using TargetP, SignalP, and related tools**

Olof Emanuelsson, Søren Brunak, Gunnar von Heijne, Henrik Nielsen

**SignalP Database Search**

**SUBMISSION**

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

>2500607071 Nitrate/nitrite transporter [Planctomycetes limnophilus DSM 3776 : PlimDRAFT_4083246_C168]
MTTSAKATSIRLUDFKTPPMRRAFHMSTUFALLCGAFFGIGIPLMVPVRDE
MHLSDKQVGWCIGSVAITVLARYVGWLCDDRIGPRLAYSGLLVASIPV

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

Organism group:
- Eukaryotes
- Gram-negative bacteria
- Gram-positive bacteria

Output format:
- Standard
- Full
- Short (no graphics!)

Method:
- Neural networks
- Hidden Markov models
- Both

Graphics:
- No graphics
- GIF (inline)
- GIF (inline) and EPS (as links)

Truncation:
Truncate each sequence to max. 70 residues.

We recommend that only the N-terminal part of each protein sequence is submitted. Enter 0 (zero) to disable truncation.

**Try Eukaryotes database first**

**“CLICK”**

Signal peptide should be in N-terminus of your protein; No need to scan full length.
Signal P (Eukaryote)

If the probability is >0.50, then the results suggest that your gene encodes a signal peptide. Higher confidence in probability score if >0.75.

What would you conclude for this protein?

- Signal peptide cleaved by proteolytic enzymes
- N-terminus of signal peptide
- Hydrophobic Region (TMH)
- C-terminus of signal peptide

Possible protease cleavage site if probability > 0.75
Prediction of protein sorting

- Psort web server: http://psort.nibb.ac.jp/
  - prediction of protein localization sites in cells from their primary amino acid sequence
Enter "Protein Sequence" in FASTA format
Recording results in your Lab Notebook

PSORTb Results (Click here for an explanation)

SeqID: 2500607071 Nitrate/nitrite transporter

Analysis Report:
- 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

Enter in your Lab Notebook

PSORT

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
Phobius

• Graphical output

• Combination of transmembrane topology (TMHMM) and signal peptide predictor (SignalP)
Phobius

A combined transmembrane topology and signal peptide predictor

Normal prediction

Paste your protein sequence here in Fasta format:

>2500607071 Nitrate/nitrite transporter [Planctomyces limnophilus DSM 3776 : PlimDRAFT_4083246_C168]
MTSAKATSIRLEDFKTPRMAHSMWFAFCLFFAWFGIAPLMVPVRDE
MHLSKDQVGCIGSVAITVLARKLYGWLCDRIGRLSGLLYLASSIPV
MGILAHDFTTFLMFRIAIGAAIGASFVITQYHTSIMFAKNCVGTANATTA
GWGNLGGGVTQVMPITFFALMVAFGLTASSWRFCSMLAGVVCATIGIA
YFFLTQHTDEPGFAELRATLGMSQKSAVKGTFQEACRDYRVWILFLVYGA

Copy/paste your amino acid sequence in Fasta format

Select output format:
- Short
- Long without Graphics
- Long with Graphics

"Click"
Query Results

Phobius prediction

Prediction of 2500607071

<table>
<thead>
<tr>
<th>ID</th>
<th>2500607071</th>
</tr>
</thead>
<tbody>
<tr>
<td>FT</td>
<td>TOPO_DOM</td>
</tr>
<tr>
<td>FT</td>
<td>TRANSMEM</td>
</tr>
<tr>
<td>FT</td>
<td>TOPO_DOM</td>
</tr>
<tr>
<td>FT</td>
<td>TRANSMEM</td>
</tr>
<tr>
<td>FT</td>
<td>TOPO_DOM</td>
</tr>
<tr>
<td>FT</td>
<td>TRANSMEM</td>
</tr>
<tr>
<td>FT</td>
<td>TOPO_DOM</td>
</tr>
<tr>
<td>FT</td>
<td>TRANSMEM</td>
</tr>
<tr>
<td>FT</td>
<td>TOPO_DOM</td>
</tr>
<tr>
<td>FT</td>
<td>TRANSMEM</td>
</tr>
<tr>
<td>FT</td>
<td>TOPO_DOM</td>
</tr>
<tr>
<td>FT</td>
<td>TRANSMEM</td>
</tr>
<tr>
<td>FT</td>
<td>TOPO_DOM</td>
</tr>
<tr>
<td>FT</td>
<td>TRANSMEM</td>
</tr>
<tr>
<td>FT</td>
<td>TOPO_DOM</td>
</tr>
<tr>
<td>FT</td>
<td>TRANSMEM</td>
</tr>
<tr>
<td>FT</td>
<td>TOPO_DOM</td>
</tr>
<tr>
<td>FT</td>
<td>TRANSMEM</td>
</tr>
<tr>
<td>FT</td>
<td>TOPO_DOM</td>
</tr>
<tr>
<td>FT</td>
<td>TRANSMEM</td>
</tr>
<tr>
<td>FT</td>
<td>TOPO_DOM</td>
</tr>
<tr>
<td>FT</td>
<td>TRANSMEM</td>
</tr>
<tr>
<td>FT</td>
<td>TOPO_DOM</td>
</tr>
<tr>
<td>FT</td>
<td>TRANSMEM</td>
</tr>
<tr>
<td>FT</td>
<td>TOPO_DOM</td>
</tr>
</tbody>
</table>

Graphical summary

Phobius posterior probabilities for 2500607071
Interpreting the Phobius Plot

- Y axis shows probability
- X axis shows amino acid position

GRAY regions = transmembrane helices
Green lines = cytoplasmic regions
Blue lines = non-cytoplasmic regions
Red lines = signal peptides

0.75
By analyzing the probabilities shown on the plot, you can determine where segments within the protein are located.