Where in the cell is your protein most likely found?
Where Are Proteins Located?

- All proteins are synthesized in the cytoplasm.
- Proteins with **export signals** can be directed to other cellular locations:
  - cytoplasm, cytoplasmic membrane, outer membrane or periplasm of Gram (-) bacteria, cell wall, or as secreted products in extracellular space.

Insert **Figure 1** from Gardy and Brinkman (2006)
Methods for predicting bacterial protein subcellular localization.
*Nature Reviews Microbiology* **4**: 741-751.
What do we know about *Planctomyces limnophilus*?

Is it Gram-positive or Gram-negative?

Where is it possible for proteins to be located for *P. limnophilus*?

To answer these questions, we need to know more details about the organism. . .
Recall: *Planctomyces limnophilus* DSM 3776

- budding, stalked bacterium isolated from the surface of a eutrophic freshwater lake in Holstein, Germany


Insert map of Germany from www.mapsofworld.com
Characteristics of *Planctomyces*

- form rosettes (star-like form) with cells connected by non-cellular, protein stalk

Insert **Figure 1** from Fuerst (1995)

- When initially discovered were thought to be fungal conidia
Planctomyces possess internal, membrane-bound compartments (blurs the boundary between prokaryotes & eukaryotes)

- some bound the nucleoid
  - Gemmata obscuriglobus

- some partition metabolic functions
  - Brocadia anammoxidans
    - carries out anaerobic oxidation of \( \text{NH}_3 \) to \( \text{N}_2 \) within enclosed structure called an anammoxosome
Planctomyces limnophilus DSM 3776

- form red colonies
- do not form endospores
- mature cell shape & size
  - ovoid to spherical
  - 1.1 – 1.5 μm
- attach to surfaces using fibrous holdfast at end of long, rigid stalk composed of twisted fibrils
  - stalk made of protein
- multiply by yeast-like budding

What other organisms use asymmetric cell division?

Budding bacterium:
*Planctomyces limnophilus*

Budding yeast (eukaryote):
*Saccharomyces cerevisiae*


Insert image of budding yeast cell
Cellular characteristics of *Planctomyces*

- surface appendages
  - stalk with holdfast
  - flagellum
  - pili
  - fimbriae

*P. maris*

- lack peptidoglycan in cell wall
  - Like mycoplasmas & clamydiae
  - consequently stains Gram-negative

- naturally resistant to penicillin

*Neisseria gonorrhoeae*

- Does this imply *P. limnophilis* is Gram-negative? Why or why not?

Insert image from ASM Microbelibrary.org

Recall: Gram-negative cells are red
Cellular characteristics of *Planctomyces*

- **cell membrane**
  - lipids with glycerol esters of fatty acids
    - composition consistent with that of *Bacteria & Eukarya*
    - major phospholipids are palmitic, palmitoleic, & oleic acids
- **some evidence for lipid A**
  - Does this mean *P. limnophilus* has an outer membrane (LPS) like Gram-negative cells?
Cellular characteristics of *Planctomyces*

- Subunit composition of RNAP consistent with *Bacteria*

Ec – *E. coli* (Bacteria)
Hs – *Halobacterium salinarum* (Archaea)
Sa – *Sulfolobus acidocaldarius* (Archaea)
Sc – *Saccharomyces cerevisiae* (Eukarya)
Why is this information important?

Is it Gram-positive or Gram-negative?

Where is it possible for proteins to be located for *P. limnophilus*?

• Planctomycetes have distinct cellular characteristics
  ✓ Absence of peptidoglycan but possible presence of lipid A in cell envelop
    ▪ What does the structure of the cell envelop in *P. limnophilus* resemble? Gram-positive or Gram-negative?
  ✓ Exhibit budding-like mechanism for cell division (like eukaryotes)
  ✓ Have internal membranes (compartmentalization like eukaryotes)
    ▪ Is there an evolutionary relationship to origin of eukaryotic nucleus?
    ▪ Will sorting signals resemble those in bacteria or eukaryotes?

• At least 8% of the *Rhodopirellula baltica* proteome exhibits homology with eukaryotic genes (Glockner et al. 2003)
  ✓ HGT or genes derived from the universal ancestor of all 3 domains?
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.

Cellular Localization Data Module

- Module Instructions

  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 : PliimDRAFT_4083246_C168]  
MTSAGAKTISRLWDFKTPPMRAHMSWFAFFLCFAWFGIAPLMVPVRDE  
MHLSKDQVOGWCIIGSVAITVLRALYVGLCDRIGPRLAYSGLLVLASIPV  
MGCLAHDFYTLFRAIGAIGAFVITQYHTSIFMAKNCVGTAANTTA  
GWGNLGGGVTQVMPTFLFAALLMVAFLSTASSWRFCMLLAGVVCATTGIA
```

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

**Other options:**
- [ ] Use old model (version 1)

Submit  Clear

"CLICK"

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

HELP with output formats

# 2500607071 Length: 450
# 2500607071 Number of predicted TMHs: 12
# 2500607071 Expected number of AAs in TMHs: 263.03045
# 2500607071 Expected number, first 60 AAs: 25.40106
# 2500607071 Total prob of N-in: 0.99853
# 2500607071 Possible N-term signal sequence

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%

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

The 12 predicted TMHs

Transmembrane

Inside (cytoplasm)

Outside (extracellular, periplasm)
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

# plot in postscript, script for making the plot in gnuplot, data for plot

Save image in GIF format to your computer and insert into Lab Notebook
Summarize your analysis of the TMHMM plot in the box provided for "comments".
Recording results in your Lab Notebook

Confirm you record the number of TMHs, with a text description of the boundaries for each TMH.

Insert the TMHMM plot.

Examine the plot & summarize results, with a confidence rating based on the probability score. Assess how its structure is related to assigned function.
• 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
- most likely cleavage site (between position # and #)
- insert the signal peptide graph

---

**Locating proteins in the cell using TargetP, SignalP, and related tools**
Olof Emanuelsson, Søren Brunak, Gunnar von Heijne, Henrik Nielsen
• Keep in mind…
  • *Planctomyces* have characteristics of both *Gram-negative* bacteria (lack peptidoglycan in cell envelop) & *eukaryotic* organisms (internal membranes, budding mechanism for cell division)

---

**SignalP Database Search**

Try Gram-negative database first

**SUBMISSION**

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

>2500607C71 Nitrate/nitrite transporter [Planctomyces limnophilus DSM 3776 : FlandRAFT_4393246_C158]

MTTSARATS1ELWDEFTPFRHAFMSQAPFLLCFPPAMFCIAPLMBPVRDE

NhLSKDPVGC11SVAITVLAPLSLPGS/CDRIOPPLAYSGLLYASIPY

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

**Method**

- Neural networks
- Hidden Markov models
- Both

**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.

**Signal peptide should be in N-terminus of your protein; No need to scan full length**
Look for Probability exceeding 0.50 threshold. If not, go back and select "eukaryotes" as an organism group.
If no significant results obtained searching the Gram-negative database, next try the eukaryotic database.
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.

Possible protease cleavage site if probability > 0.75

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

What would you conclude for this protein?

>2500607071
Prediction: Signal peptide

<table>
<thead>
<tr>
<th>Probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>Signal peptide</td>
</tr>
<tr>
<td>Signal anchor</td>
</tr>
<tr>
<td>Max cleavage site</td>
</tr>
<tr>
<td>Probability range</td>
</tr>
</tbody>
</table>

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.
Recording results in your Lab Notebook

**SignalP**

Go to [http://www.cbs.dtu.dk/services/SignalP/](http://www.cbs.dtu.dk/services/SignalP/) enter the signal peptide probability

0.732

Most likely cleavage site (between position # and #)

Probability of 0.417 between pos. 43 and 44

Insert the signal peptide graph

SAVE image in GIF format then insert image inside box.
PSORT-B

- Another useful tool in predicting bacterial protein localization

- The output is TEXTUAL, but the information still will be helpful

PSORT

Go to: [http://www.psort.org/psortb/](http://www.psort.org/psortb/)

<table>
<thead>
<tr>
<th>Cytoplasmic score</th>
<th>#</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>CytoplasmicMembrane score</th>
<th>#</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Periplasmic score</th>
<th>#</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Enter "Protein Sequence" in FASTA format.

Select "Negative" for Gram stain.
Where this protein is predicted to be located in the cell.

Enter in your Lab Notebook.
Phobius

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

Phobius

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

“Click”

enter the graph

image
Copy/paste your amino acid sequence in Fasta format

Copy/paste your amino acid sequence in Fasta format

Normal prediction

Paste your protein sequence here in Fasta format:

>2500607071 Nitrate/nitrite transporter [Planctomyces limnophilus DSM 3776 :
PlimDRAFT_4083246_C168]
MTTSAKATSTRLDFKTPPRAHMSWFAPFCLFFAWFGIAPLMPVRDE
MHLSKDOVGWCIIGSVAITVLARLYVGLCDRIGPRLAYSGLLVLASIPV
MGIGLAPFFTFFLMFRIAIGAIGASFVTQYHTSINFRAEKNCGVTANATTAGWGNLGGSVQTVMPTLFAILMMVAFGLSTASSWRFMCMLLAGVVCAITGIA
YFPLTQDTEQGFNFAELRATRGKMSQKSAVKGTFQACRDRYRVLWFLVYGA
Or: Select the sequence file you wish to use

Select output format:

- Short
- Long without Graphics
- Long with Graphics

“Click”
### Phobius prediction

**Prediction of 2500607071**

<table>
<thead>
<tr>
<th>ID</th>
<th>FT</th>
<th>TMHs</th>
<th>Signal Peptide</th>
</tr>
</thead>
<tbody>
<tr>
<td>2500607071</td>
<td>TOPO_DOM</td>
<td>20</td>
<td>NON CYTOPLASMIC.</td>
</tr>
<tr>
<td></td>
<td>TRANSMEM</td>
<td>45</td>
<td></td>
</tr>
<tr>
<td></td>
<td>TOPO_DOM</td>
<td>56</td>
<td>NON CYTOPLASMIC.</td>
</tr>
<tr>
<td></td>
<td>TRANSMEM</td>
<td>75</td>
<td></td>
</tr>
<tr>
<td></td>
<td>TOPO_DOM</td>
<td>66</td>
<td>CYTOPLASMIC.</td>
</tr>
<tr>
<td></td>
<td>TRANSMEM</td>
<td>106</td>
<td></td>
</tr>
<tr>
<td></td>
<td>TOPO_DOM</td>
<td>111</td>
<td>NON CYTOPLASMIC.</td>
</tr>
<tr>
<td></td>
<td>TRANSMEM</td>
<td>131</td>
<td></td>
</tr>
<tr>
<td></td>
<td>TOPO_DOM</td>
<td>151</td>
<td>CYTOPLASMIC.</td>
</tr>
<tr>
<td></td>
<td>TRANSMEM</td>
<td>176</td>
<td></td>
</tr>
<tr>
<td></td>
<td>TOPO_DOM</td>
<td>181</td>
<td>NON CYTOPLASMIC.</td>
</tr>
<tr>
<td></td>
<td>TRANSMEM</td>
<td>204</td>
<td></td>
</tr>
<tr>
<td></td>
<td>TOPO_DOM</td>
<td>240</td>
<td>CYTOPLASMIC.</td>
</tr>
<tr>
<td></td>
<td>TRANSMEM</td>
<td>267</td>
<td></td>
</tr>
<tr>
<td></td>
<td>TOPO_DOM</td>
<td>270</td>
<td>NON CYTOPLASMIC.</td>
</tr>
<tr>
<td></td>
<td>TRANSMEM</td>
<td>299</td>
<td></td>
</tr>
<tr>
<td></td>
<td>TOPO_DOM</td>
<td>319</td>
<td>CYTOPLASMIC.</td>
</tr>
<tr>
<td></td>
<td>TRANSMEM</td>
<td>339</td>
<td></td>
</tr>
<tr>
<td></td>
<td>TOPO_DOM</td>
<td>344</td>
<td>NON CYTOPLASMIC.</td>
</tr>
<tr>
<td></td>
<td>TRANSMEM</td>
<td>368</td>
<td></td>
</tr>
<tr>
<td></td>
<td>TOPO_DOM</td>
<td>374</td>
<td>CYTOPLASMIC.</td>
</tr>
<tr>
<td></td>
<td>TRANSMEM</td>
<td>398</td>
<td></td>
</tr>
<tr>
<td></td>
<td>TOPO_DOM</td>
<td>403</td>
<td>NON CYTOPLASMIC.</td>
</tr>
<tr>
<td></td>
<td>TRANSMEM</td>
<td>426</td>
<td></td>
</tr>
<tr>
<td></td>
<td>TOPO_DOM</td>
<td>458</td>
<td>CYTOPLASMIC.</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

Phobius posterior probabilities for 2500607071
By analyzing the probabilities shown on the plot, you can determine where segments within the protein are located.
Recording results in your Lab Notebook

Phobius

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

Text prediction for TM helices, intervening loops, and signal peptides

<table>
<thead>
<tr>
<th>FT TOPO_DOM</th>
<th>FT TRANSMEM</th>
<th>FT TOPO_DOM</th>
<th>FT TRANSMEM</th>
<th>FT TOPO_DOM</th>
<th>FT TRANSMEM</th>
<th>FT TOPO_DOM</th>
<th>FT TRANSMEM</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 20 CYTOPLASMIC</td>
<td>21 45</td>
<td>NON CYTOPLASMIC</td>
<td>57 75</td>
<td>CYTOPLASMIC</td>
<td>87 106</td>
<td>CYTOPLASMIC</td>
<td>112 131</td>
</tr>
<tr>
<td>107 111 NON CYTOPLASMIC</td>
<td>132 151 CYTOPLASMIC</td>
<td>152 176</td>
<td>CYTOPLASMIC</td>
<td>177 181 NON CYTOPLASMIC</td>
<td>182 204</td>
<td></td>
<td></td>
</tr>
<tr>
<td>205 240 CYTOPLASMIC</td>
<td>241 267</td>
<td>NON CYTOPLASMIC</td>
<td>268 273 NON CYTOPLASMIC</td>
<td>279 299</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>300 319 CYTOPLASMIC</td>
<td>320 339</td>
<td>NON CYTOPLASMIC</td>
<td>340 344 NON CYTOPLASMIC</td>
<td>345 363</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>369 374 CYTOPLASMIC</td>
<td>375 398</td>
<td>CYTOPLASMIC</td>
<td>399 403 NON CYTOPLASMIC</td>
<td>404 425</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>427 458 CYTOPLASMIC</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Recording results in your Lab Notebook

enter the graph

SAVE image in PNG format then insert into Notebook inside box.
Finishing Up Cellular Localization Module

Now that you’ve finished TMHMM, SignalP, 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 signal peptide at N-terminus?
• Where did PSORT 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 this module

Hypothesis

Where do you expect to find this protein?

The TMHMM and Phobius agree that there are 12 transmembrane helices, which is consistent with its localization according to PSORT in the cytoplasmic membrane. Signal P suggests there may be a signal peptide at the N-terminus using the eukaryotic database; however, the probability is only 73%, just below the cut-off of 75%. This is an interesting result, as it suggests that this protein has a eukaryotic-like signal peptide. There were no significant results when querying the Gram-negative database for Signal P.