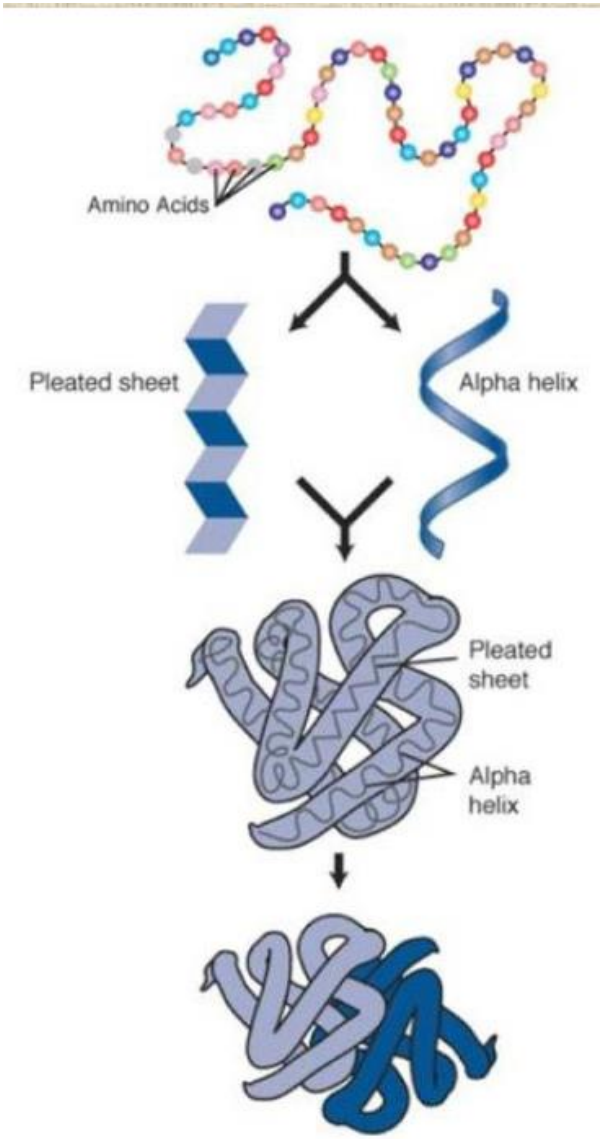
Prediction Methods

Secondary Structure Prediction

Secondary structure prediction is a set of techniques in bioinformatics that aim to predict the secondary structures of proteins sequences based only on knowledge of their primary structure.



Levels of Description of Structural Complexity

- Primary Structure : aa sequence
- Secondary Structure: Regions in which the polypeptide chains are organised into regular structures known as α-helices (H), β-pleated sheets (B) , β-turns (T) e.t.c.
- Super-Secondary Structure : a compact three-dimensional protein structure of several adjacent elements of a secondary structure that is smaller than a protein domain (All-a, All-b, a/b, a+b)
- Domain: a compact three-dimensional structure which can be independently stable and folded
- Tertiary Structure : 3-D structure of an entire polypeptide
- Quarternary Structure : The number and arrangement of the protein subunits (chains) with respect to one another

Secondary structure prediction methods

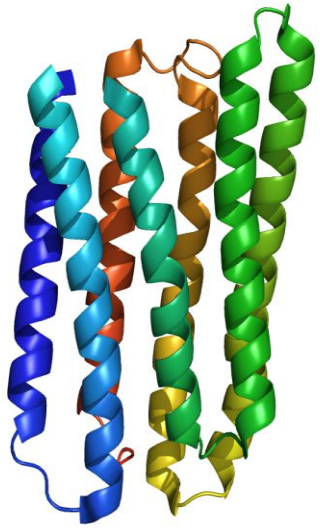
- PSIPRED http://bioinf.cs.ucl.ac.uk/psipred/
- jpred http://www.compbio.dundee.ac.uk/jpred4/index_up.html
- Porter 4 http://distillf.ucd.ie/porterpaleale/
- SECSTR http://athina.biol.uoa.gr/SecStr/

Prediction of membrane protein topology

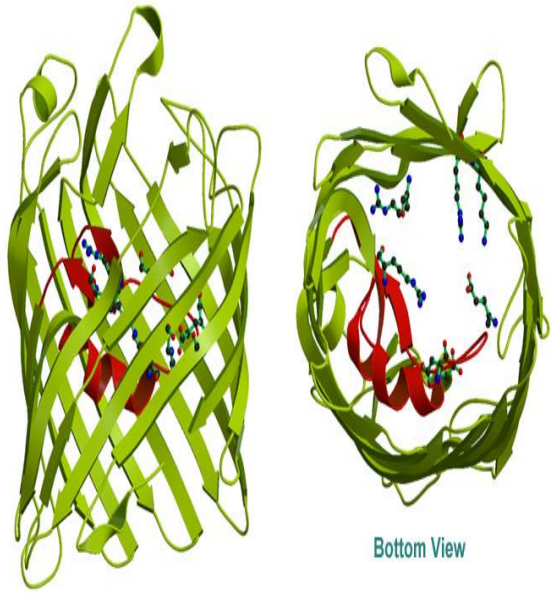
Transmembrane proteins

- ≈ 20-30% of total number of proteins in a proteome
- Two main structural types:

  α-helical and β-barrels



X-ray structure of Bacteriorhodopsin (1XIO)



X-ray structure of porin protein OmpF (Cowin et al. 1992)

Studied with

X-ray crystallography
NMR spectroscopy

Biochemical/ Molecular techniques

Computational algorithms

Topology prediction software

Name	Type	URL
TOPCONS	HMM - consensus	http://topcons.net/
(Poly)Phobius	HMM	http://phobius.sbc.su.se/
Philius	DBN	http://www.yeastrc.org/philius/pages/philius/runPhilius.jsp
MEMSAT3	Dynamic programming	http://bioinf.cs.ucl.ac.uk/psipred/
MEMSAT-SVM	SVM	http://bioinf.cs.ucl.ac.uk/psipred/
(SP)OCTOPUS	HMM & ANN (ENSEMBLE)	http://topcons.net/
SCAMPI	Hydrophobicity	https://scampi.bioinfo.se/
TMHMM	HMM	http://www.cbs.dtu.dk/services/TMHMM/
HMM-TM	HMM	http://bioinformatics.biol.uoa.gr/HMM-TM/
TopPred	Hydrophobicity	https://github.com/C3BI-pasteur-fr/toppred/
PHDhtm	ANN	https://www.predictprotein.org/
HMMTOP	HMM	http://www.enzim.hu/hmmtop/

β-barrel topology prediction

Topology Prediction methods

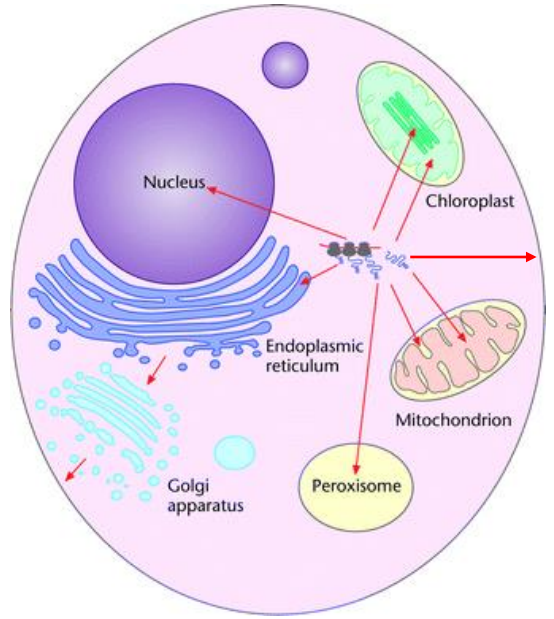
Name	Type	URL
PRED-TMBB2	HMM	www.compgen.org/tools/PRED-TMBB2
BOCTOPUS2	HMM+ANN	http://boctopus.bioinfo.se/
BetAware	Conditional Random Fields	https://betaware.biocomp.unibo.it/
ProfTMB	HMM	http://www.predictprotein.org/
TMBETAPRED-RBF	Radial Basis Function	http://rbf.bioinfo.tw/~sachen/BARRELpredict/TMBETAPRED-RBF.php
HMM-B2TMR	HMM	http://gpcr2.biocomp.unibo.it/predictors/
ConBBPRED	Consensus	http://bioinformatics.biol.uoa.gr/ConBBPRED/

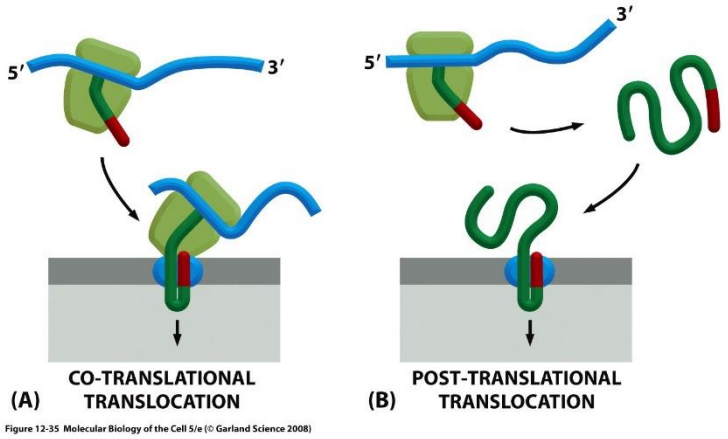
Prediction of signal peptides

Signal Peptides

“proteins have intrinsic signals that govern their transport an localization in the cell” (Blobel, 2000)

- Gunter Blodel was awarded the 1999 Nobel prize in Physiology or Medicine for “Protein Targeting”
- He discovered that many proteins have a **signal sequence**, that is, a short amino acid sequence at one end that functions **like a postal code** for the target organelle.
- Proteins "belong" in a different organelle, they can be transported there in either by **co-translational** translocation (translocation during the process of translation) or by **post-translational** translocation (translocation after the process of translation is complete).





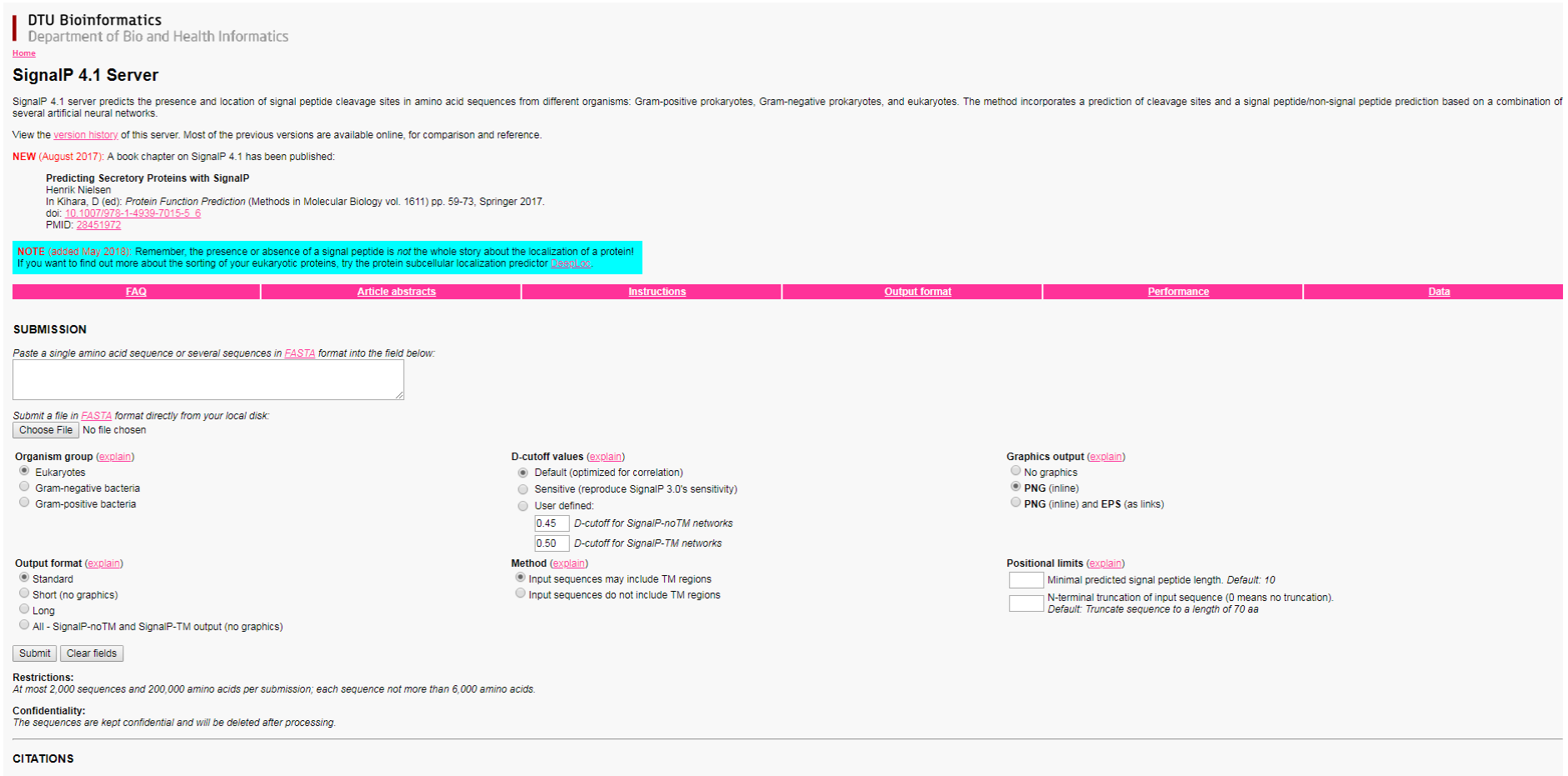
Blobel, G. (2000). Protein targeting (Nobel lecture). *Chembiochem.*, 1:86-102.

Prediction of signal peptides

- **Numerous methods** for prediction of protein targeting and signal peptides have been developed.
- However, no prediction method will be able to cover all the different types of signal peptides.
- Most methods predicts classical signal peptides targeting to the general secretory pathway in bacteria or classical secretory pathway in eukaryotes.
- Furthermore, a few methods for prediction of non-classically secreted proteins have emerged.
- The first prediction methods used weight matrices based on the analysis by Gunnar von Heijne (von Heijne, 1986), for example **SigCleave** (http://emboss.sourceforge.net/apps/release/6.6/emboss/apps/sigcleave.html) and later **PrediSi** (http://www.predisi.de/).
- More sophisticated **machine learning** methods, such as ANNs and HMMs, are the basis of most current signal peptide prediction methods.

SignalP

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



DTU Bioinformatics
Department of Bio and Health Informatics

Home

SignalP 4.1 Server

SignalP 4.1 server predicts the presence and location of signal peptide cleavage sites in amino acid sequences from different organisms: Gram-positive prokaryotes, Gram-negative prokaryotes, and eukaryotes. The method incorporates a prediction of cleavage sites and a signal peptide/non-signal peptide prediction based on a combination of several artificial neural networks.

View the version history of this server. Most of the previous versions are available online, for comparison and reference.

NEW (August 2017): A book chapter on SignalP 4.1 has been published:

Predicting Secretory Proteins with SignalP
Henrik Nielsen
In Kihara, D (ed): *Protein Function Prediction* (Methods in Molecular Biology vol. 1611) pp. 59-73, Springer 2017.
doi: 10.1007/978-1-4939-7015-5_6
PMID: 28451972

NOTE (added May 2018): Remember, the presence or absence of a signal peptide is *not* the whole story about the localization of a protein!
If you want to find out more about the sorting of your eukaryotic proteins, try the protein subcellular localization predictor DeepLoc.

FAQ | Article abstracts | Instructions | Output format | Performance | Data

SUBMISSION

Paste a single amino acid 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 chosen

Organism group (explain)
- Eukaryotes
- Gram-negative bacteria
- Gram-positive bacteria

D-cutoff values (explain)
- Default (optimized for correlation)
- Sensitive (reproduce SignalP 3.0's sensitivity)
- User defined:
  - 0.45 *D-cutoff for SignalP-noTM networks*
  - 0.50 *D-cutoff for SignalP-TM networks*

Graphics output (explain)
- No graphics
- **PNG** (inline)
- **PNG** (inline) and **EPS** (as links)

Output format (explain)
- Standard
- Short (no graphics)
- Long
- All - SignalP-noTM and SignalP-TM output (no graphics)

Method (explain)
- Input sequences may include TM regions
- Input sequences do not include TM regions

Positional limits (explain)
- Minimal predicted signal peptide length. *Default: 10*
- N-terminal truncation of input sequence (0 means no truncation). *Default: Truncate sequence to a length of 70 aa*

Submit | Clear fields

Restrictions:
At most 2,000 sequences and 200,000 amino acids per submission; each sequence not more than 6,000 amino acids.

Confidentiality:
The sequences are kept confidential and will be deleted after processing.

CITATIONS

Other signal peptide prediction methods

Method	Taxonomy	Website
PrediSi	Eukaryotes, Gram-positive, Gram-negative	http://www.predisi.de/
SPEPLip	Eukaryotes, Gram-positive, Gram-negative	http://gpcr.biocomp.unibo.it/cgi/predictors/spep/pred_spepcgi.cgi
Signal-CF	Eukaryotes, Gram-positive, Gram-negative	http://www.csbio.sjtu.edu.cn/bioinf/Signal-CF/
Signal-3L	Eukaryotes, Gram-positive, Gram-negative	http://www.csbio.sjtu.edu.cn/bioinf/Signal-3L/
Signal-BLAST	Eukaryotes, Gram-positive, Gram-negative, Archaea, Viruses	http://sigpep.services.came.sbg.ac.at/signalblast.html
Phobius	Eukaryotes, Gram-positive, Gram-negative	http://phobius.sbc.su.se/
Philius	Eukaryotes, Gram-positive, Gram-negative	http://www.yeastrc.org/philius/pages/philius/runPhilius.jsp
MEMSAT3	Eukaryotes, Gram-positive, Gram-negative	http://bioinf.cs.ucl.ac.uk/software_downloads/memsat/
MEMSAT-SVM	Eukaryotes, Gram-positive, Gram-negative	http://bioinf.cs.ucl.ac.uk/software_downloads/memsat/
SPOCTOPUS	Eukaryotes, Gram-positive, Gram-negative	http://octopus.cbr.su.se/

Signal Peptide Prediction in Archaea

PRED-SIGNAL

http://www.compgen.org/tools/PRED-SIGNAL

http://bioinformatics.biol.uoa.gr/PRED-SIGNAL/

- The first signal peptide prediction in Archaea. After an extensive literature search, archeal proteins with experimentally verified SPs were found and analyzed.
- The unique features of the SPs of archaea are:
  - ✓ the unique amino acid composition of the hydrophobic region
  - ✓ a cleavage site resembling more the sequences of gram-positives
- Using a HMM predicts the presence of the signal peptides and their cleavage sites.
- Also discriminates such proteins from cytoplasmic and transmebrane ones.

Prediction of bacterial lipoproteins and signal peptides

LipoP:

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

- Discriminates between lipoprotein signal peptides, other signal peptides and n-terminal membrane helices in Gram-negative bacteria.
- It has a good performance on sequences from Gram-positive bacteria also.

PRED-LIPO:

http://www.compgen.org/tools/PRED-LIPO

http://bioinformatics.biol.uoa.gr/PRED-LIPO/

- A Hidden Markov Model method for the prediction of lipoprotein signal peptides of Gram-positive bacteria.
- The method is also very sensitive and specific in the detection of secretory signal peptides and transmembrane helices.