SignalP and TMHMM Plugin

USER MANUAL

User manual for

SignalP and TMHMM 25.0

Windows, macOS and Linux

November 28, 2024

This software is for research purposes only.

QIAGEN Aarhus AS
Kalkværksvej 5, 11.
DK - 8000 Aarhus C
Denmark

Contents

Chapter 1

Introduction

The SignalP and TMHMM plugin makes Signal Peptide Prediction and Transmembrane Helix Prediction tools accessible via the *CLC Workbench* (figure 1.1). An active internet connection is required to use all the tools supplied by this plugin.

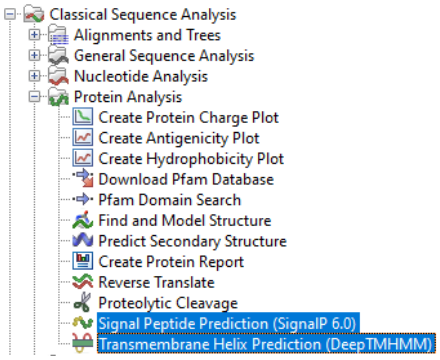

Figure 1.1: *Tools from the SignalP and TMHMM plugin are located in the Protein Analysis folder under the Tools menu.*

Chapter 2

Signal Peptide Prediction (SignalP 6.0)

The **Signal Peptide Prediction (SignalP 6.0)** tool uses the SignalP 6.0 service at https://services.healthtech.dtu.dk/services/SignalP-6.0 [Teufel et al., 2022].

The SignalP 6.0 [Teufel et al., 2022] service uses a machine learning model to detect all five signal peptide types. It is also applicable to metagenomic data. Read about the SignalP history and updates at https://communities.springernature.com/posts/signalp-6-0-predicts-all-five-types-of-signal-peptides-using-protein-language-models.

To run a **Signal Peptide Prediction (SignalP 6.0)** analysis from the Workbench, go to:

> **Tools** | **Classical Sequence Analysis** () | **Protein Analysis** () | **Signal Peptide Prediction (SignalP 6.0)** ()

In the first wizard step, you select the peptide sequences to be analyzed (figure 2.1). Note: To successfully use the **Signal Peptide Prediction (SignalP 6.0)** service, protein sequences should not be shorter than 10 amino acids. The system may time out when more than 100 entries are provided, although the maximum allowed is 1000 sequences.

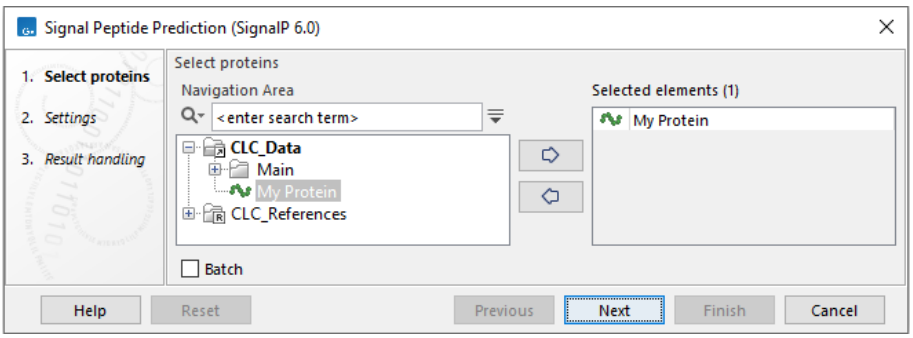

Figure 2.1: *Select input protein sequences.*

In the Settings wizard step, the options for SignalP 6.0 can be specified (figure 2.2). These are:

- **Organism**
  - **Eukarya** Use for analyzing Eukarya. When analyzing Eukarya, only "standard" secretory signal peptides transported by the Sec translocon and cleaved by Signal Peptidase I (Lep) are predicted.

  - **Other** Use for analyzing Archaea, Gram-positive Bacteria and Gram-negative Bacteria.

- **Model mode**
  - **Fast** The analysis mode is fast but region borders may be less accurate.
  - **Slow** The slow mode takes 6x longer to compute. Use when accurate region borders are needed.

- **Add sub-region annotatons** Add annotations for identified signal peptide sub-regions like n-region, h-region, c-region, cysteine, and twin-arginine motif.

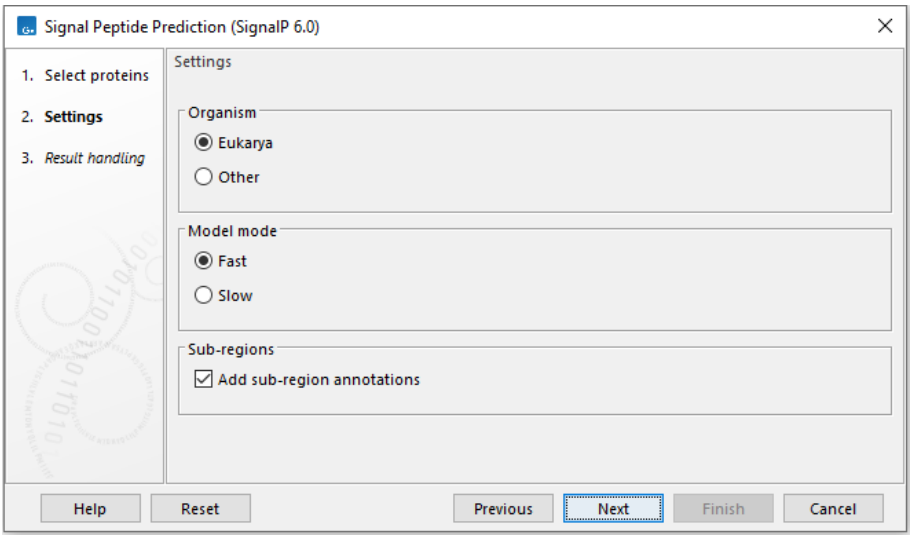

Figure 2.2: *Set the options for SignalP 6.0.*

In the Result handling wizard step, you specify the form the results should be returned in. The options are:

- **Add annotations** Annotate the input sequences with the signal peptides identified, and sub-regions if that option was selected earlier. See figure 2.3 for an example. Note: the option 'Signal peptide' must be checked in the Side Panel settings, in the 'Annotation layout' palette, for these annotations to be visible.

  Working with sequence annotations is described in detail at https://resources.qiagenbioinformatics.com/manuals/clcmainworkbench/current/index.php?manual=Working_with_annotations.html

- **Create table** Create a table containing an entry for each input sequence providing the likelihood of each type of signal peptide that was searched for See figure 2.4 for an example.

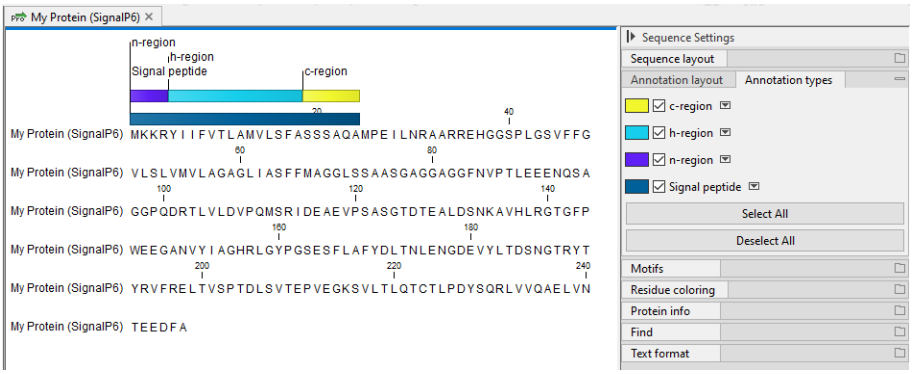

Figure 2.3: *Sequence annotated with an identified signal peptide and sub-annotations providing information about individual sequence elements in the signal peptide, here c-, h- and n-regions.*

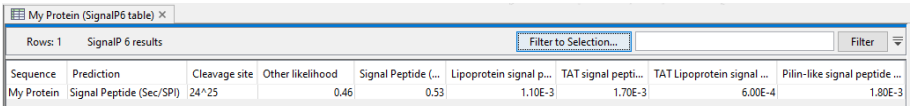

Figure 2.4: *A table listing likelihood for each of the tested signal peptides.*

2.1 SignalP background

The importance of signal peptides was shown in 1999 when Günter Blobel received the Nobel Prize in physiology or medicine for his discovery that "proteins have intrinsic signals that govern their transport and localization in the cell" [Blobel, 2000]. He pointed out the importance of defined peptide motifs for targeting proteins to their site of function.

Soon after Günter Blobel's initial discovery of signal peptides, more targeting signals were found. Most cell types and organisms employ several ways of targeting proteins to the extracellular environment or subcellular locations. Most of the proteins targeted for the extracellular space or subcellular locations carry specific sequence motifs (signal peptides) characterizing the type of secretion/targeting it undergoes.

Targeting motifs can either be removed from, or retained in the mature protein after the protein has reached the correct and final destination. Some of the best characterized signal peptides are depicted in figure 2.5.

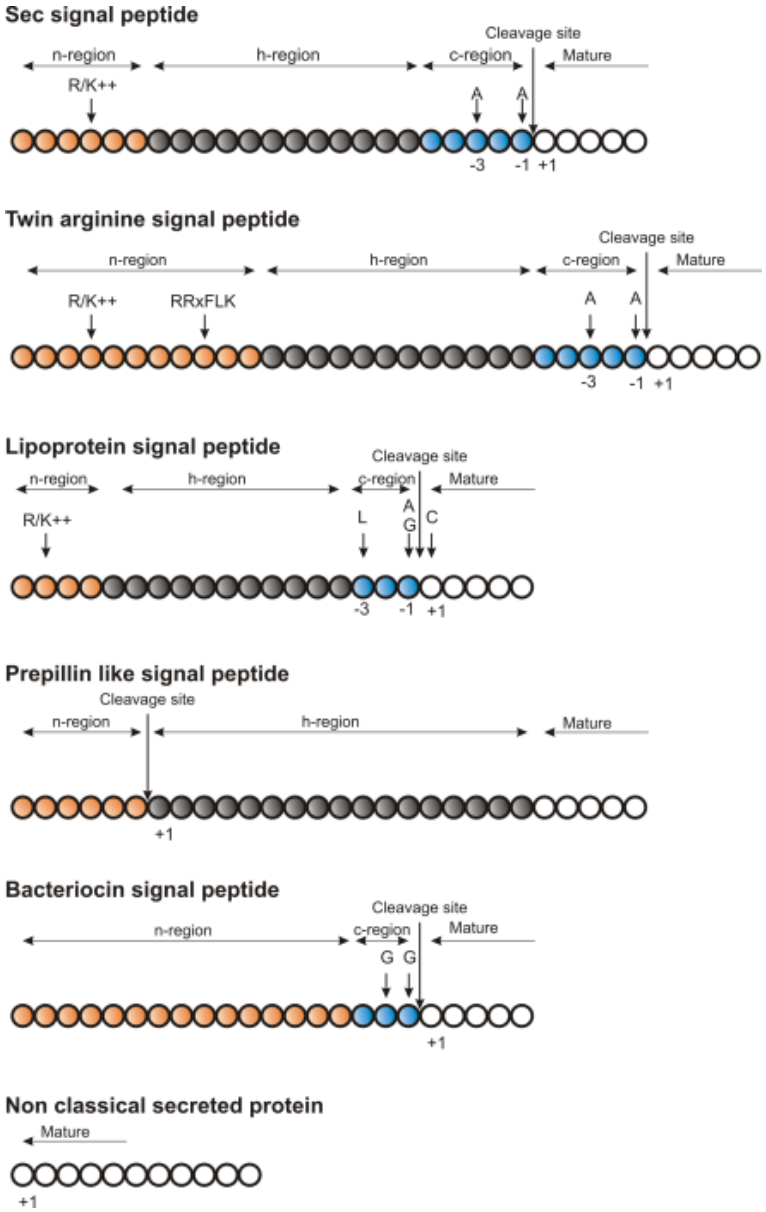

Figure 2.5: *Schematic representation of various signal peptides. Red color indicates n-region, gray color indicates h-region, cyan indicates c-region. All white circles are part of the mature protein. +1 indicates the first position of the mature protein. The length of the signal peptides is not drawn to scale.*

Chapter 3

Transmembrane Helix Prediction (DeepTMHMM)

Many proteins are integral membrane proteins. Most membrane proteins have hydrophobic regions which span the hydrophobic core of the membrane bi-layer and hydrophilic regions located on the outside or the inside of the membrane. Many receptor proteins have several transmembrane helices spanning the cellular membrane.

The prediction of transmembrane helices using **Transmembrane Helix Prediction (DeepTMHMM)**, uses the DeepTMHMM service located at https://dtu.biolib.com/DeepTMHMM. Additional information on DeepTMHMM can be found at https://dtu.biolib.com/DeepTMHMM and in the original research paper [Hallgren et al., 2022].

To run a **Transmembrane Helix Prediction (DeepTMHMM)** analysis from the Workbench, go to:

> **Tools** | **Classical Sequence Analysis** () | **Protein Analysis** () | **Transmembrane Helix Prediction (DeepTMHMM)** ()

In the first wizard step, you select the peptide sequences to be analyzed.

In the Result handling wizard step, you specify the form the results should be returned in. The options are:

- **Add annotations** Annotate the input sequences with the results. See figure 3.1 for an example. Note: relevant options must be checked in the Side Panel settings, in the 'Annotation layout' palette, for these annotations to be visible (further details below).
- **Create text** Get the text output from the DeepTMHMM service. A single text file is created, containing the results returned.

Hover the mouse cursor over an annotation to reveal a tooltip with details about it. If the annotation was added using DeepTMHMM, this is noted. The annotation types added using DeepTMHMM are easily discovered by opening the Annotation Table view of the sequence, selecting all the available annotation types in the Side Panel. Information in the Qualifiers column includes "DeepTMHMM" for annotations added using this service. Working with sequence annotations is described in detail at https://resources.qiagenbioinformatics.com/manuals/clcmainworkbench/current/index.php?manual=Working_with_annotations.html.

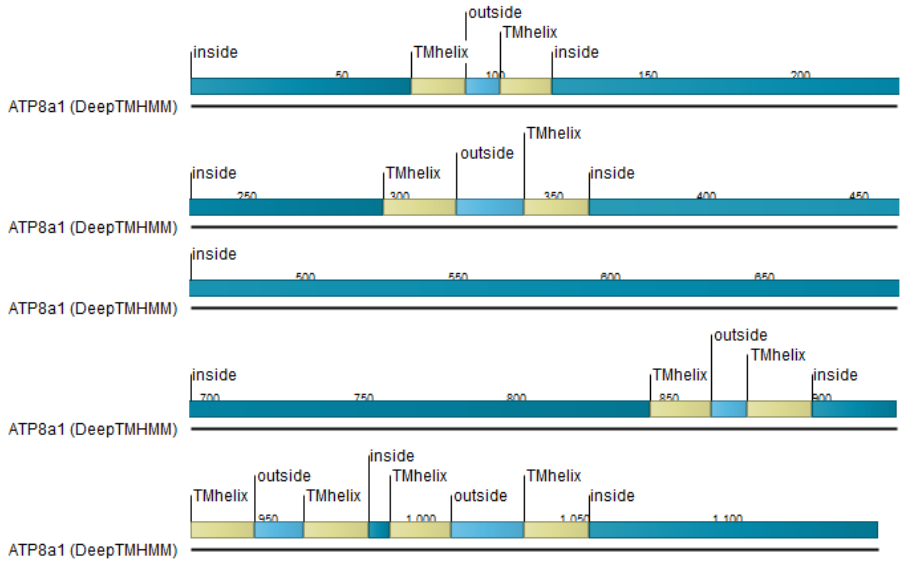

Figure 3.1: *Transmembrane segments shown as annotation on the sequence, including information about the the topology.*

Chapter 4

Install and uninstall plugins

SignalP and TMHMM is installed as a plugin.

4.1 Installation of plugins

Note: In order to install plugins and modules, the *CLC Workbench* must be run in administrator mode. On Windows, you can do this by right-clicking the program shortcut and choosing "Run as Administrator". On Linux and Mac, it means you must launch the program such that it is run by an administrative user.

Plugins and modules are installed and uninstalled using the Workbench Plugin Manager. To open the Plugin Manager, click on the **Plugins ( ) button** in the top Toolbar, or go to the menu option:

Utilities | **Manage Plugins... ( )**

The Plugin Manager has two tabs at the top:

- **Manage Plugins** An overview of your installed plugins and modules is provided under this tab.
- **Download Plugins** Plugins and modules available to download and install are listed in this tab.

To install a plugin, click on the **Download Plugins** tab (figure 4.1). Select a plugin. Information about it will be shown in the right hand panel. Click on the **Download and Install** button to install the plugin.

Accepting the license agreement

The End User License Agreement (EULA) must be read and accepted as part of the installation process. Please read the EULA text carefully, and if you agree to it, check the box next to the text **I accept these terms**. If further information is requested from you, please fill this in before clicking on the **Finish** button.

Installing a cpa file

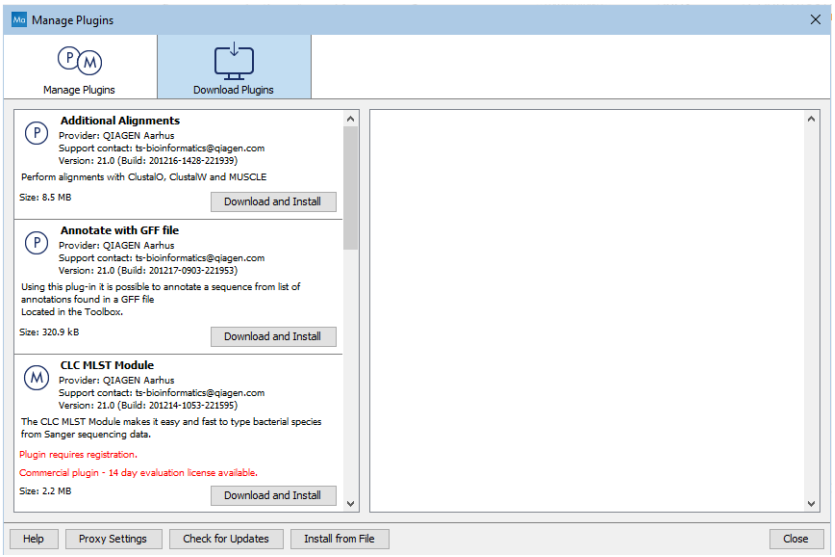

Figure 4.1: *Plugins and modules available for installation are listed in the Plugin Manager under the Download Plugins tab.*

If you have a .cpa installer file for SignalP and TMHMM, you can install it by clicking on the **Install from File** button at the bottom of the Plugin Manager.

If you are working on a system not connected to the internet, plugin and module .cpa files can be downloaded from `https://digitalinsights.qiagen.com/products-overview/plugins/`using a networked machine, and then transferred to the non-networked machine for installation.

Restart to complete the installation

Newly installed plugins and modules will be available for use after restarting the software. When you close the Plugin Manager, a dialog appears offering the opportunity to restart the *CLC Workbench*.

4.2 Uninstalling plugins

Plugins and modules are uninstalled using the Workbench Plugin Manager.To open the Plugin Manager, click on the **Plugins ( ) button** in the top Toolbar, or go to the menu option:

Utilities | **Manage Plugins... ( )**

This will open the Plugin Manager (figure 4.2). Installed plugins and modules are shown under the Manage Plugins tab of the Plugins Manager.

To uninstall a plugin or module, click on its entry in the list, and click on the **Uninstall** button.

Plugins and modules are not uninstalled until the Workbench is restarted. When you close the Plugin Manager, a dialog appears offering the opportunity to restart the *CLC Workbench*.

Disabling a plugin without uninstalling it

If you do not want a plugin to be loaded the next time you start the Workbench, select it in the

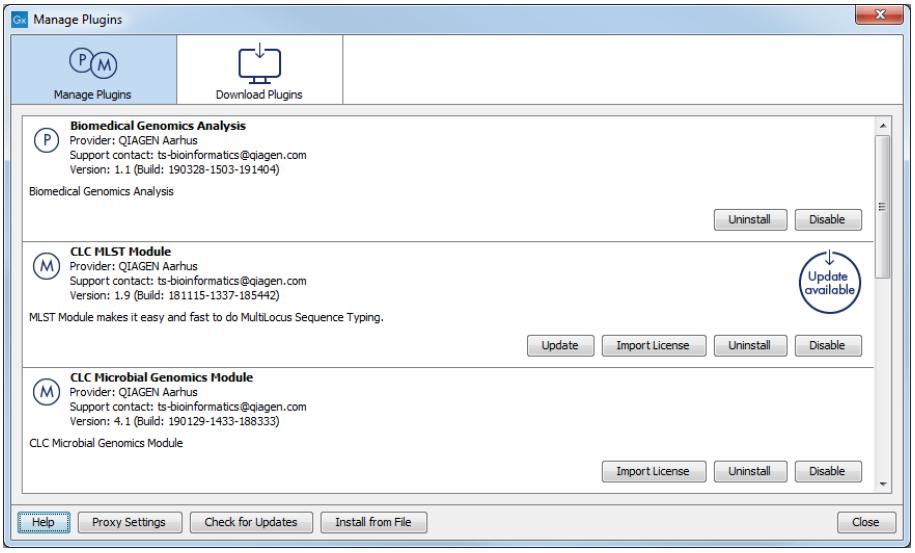

Figure 4.2: *Installed plugins and modules are listed in the Plugins Manager under the Manage Plugins tab.*

list under the Manage Plugins tab and click on the **Disable** button.

Bibliography

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

[Hallgren et al., 2022] Hallgren, J., Tsirigos, K. D., Pedersen, M. D., Armenteros, J. J. A., Marcatili, P., Nielsen, H., Krogh, A., and Winther, O. (2022). Deeptmhmm predicts alpha and beta transmembrane proteins using deep neural networks. *bioRxiv*.

[Teufel et al., 2022] Teufel, F., Almagro Armenteros, J. J., Johansen, A. R., Gíslason, M. H., Pihl, S. I., Tsirigos, K. D., Winther, O., Brunak, S., von Heijne, G., and Nielsen, H. (2022). Signalp 6.0 predicts all five types of signal peptides using protein language models. *Nature Biotechnology*, 40(7):1023–1025.