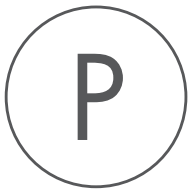

SignalP
Plugin

USER MANUAL

User manual for

SignalP 2.4

Windows, macOS and Linux

November 28, 2018

This software is for research purposes only.

QIAGEN Aarhus
Silkeborgvej 2
Prismet
DK-8000 Aarhus C
Denmark

Contents

Chapter 1

Introduction

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 1.1.

1.1 The SignalP method

One of the most cited and best methods for prediction of classical signal peptides is the SignalP method [Nielsen et al., 1997, Bendtsen et al., 2004]. In contrast to other methods, SignalP also predicts the actual cleavage site; thus the peptide which is cleaved off during translocation over the membrane. SignalP is located at `http://www.cbs.dtu.dk/services/SignalP/`.

The graphical output from SignalP (neural network) comprises three different scores, *C*, *S* and *Y*. Two additional scores are reported in the SignalP3-NN output, namely the *S-mean* and the *D-score*, but these are only reported as numerical values.

For each organism class in SignalP - Eukaryote, Gram-negative and Gram-positive - two different neural networks are used, one for predicting the actual signal peptide and one for predicting the position of the signal peptidase I (SPase I) cleavage site.

The *S-score* for the signal peptide prediction is reported for every single amino acid position in the submitted sequence, with high scores indicating that the corresponding amino acid is part of a signal peptide, and low scores indicating that the amino acid is part of a mature protein.

The *C-score* is the "cleavage site" score. For each position in the submitted sequence, a C-score is reported, which should only be significantly high at the cleavage site. Confusion is often seen

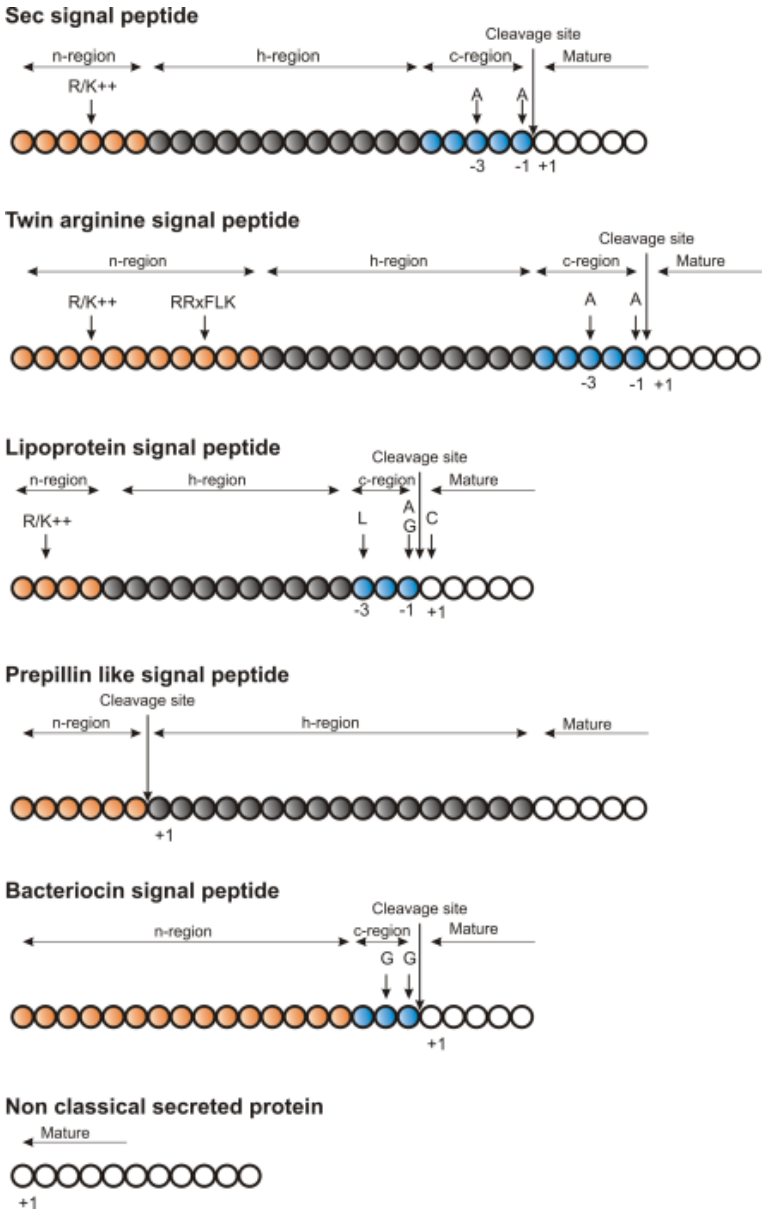

Figure 1.1: *Schematic representation of various signal peptides. Red color indicates n-region, gray color indicates h-region, cyan indicates c-region. All white circles are part of the mature protein.* $+1$ *indicates the first position of the mature protein. The length of the signal peptides is not drawn to scale.*

with the position numbering of the cleavage site. When a cleavage site position is referred to by a single number, the number indicates the first residue in the mature protein. This means that a reported cleavage site between amino acid 26-27 corresponds to the mature protein starting at (and include) position 27.

Y-max is a derivative of the C-score combined with the S-score resulting in a better cleavage site prediction than the raw C-score alone. This is due to the fact that multiple high-peaking C-scores can be found in one sequence, where only one is the true cleavage site. The cleavage site is assigned from the Y-score where the slope of the S-score is steep and a significant C-score is found.

The *S-mean* is the average of the S-score, ranging from the N-terminal amino acid to the amino acid assigned with the highest Y-max score, thus the S-mean score is calculated for the length of the predicted signal peptide. The S-mean score was in SignalP version 2.0 used as the criteria for discrimination of secretory and non-secretory proteins.

The *D-score* is introduced in SignalP version 3.0 and is a simple average of the S-mean and Y-max score. The score shows superior discrimination performance of secretory and non-secretory proteins to that of the S-mean score which was used in SignalP version 1 and 2.

For non-secretory proteins all the scores represented in the SignalP3-NN output should ideally be very low.

The hidden Markov model calculates the probability of whether the submitted sequence contains a signal peptide or not. The eukaryotic HMM model also reports the probability of a signal anchor, previously named uncleaved signal peptides. Furthermore, the cleavage site is assigned by a probability score together with scores for the n-region, h-region, and c-region of the signal peptide, if it is found.

Chapter 2

Signal peptide prediction

In order to use SignalP, *you need to download the SignalP plugin*. Once the plugin is installed, the Workbench will query SignalP [Nielsen et al., 1997, Bendtsen et al., 2004] located at `http://www.cbs.dtu.dk/services/SignalP/`. Thus an active internet connection is required to run the signal peptide prediction.

When the plugin is downloaded and installed, you can use it to predict signal peptides:

Toolbox | **Protein Analysis (**  **)**| **Signal Peptide Prediction (**  **)**

If a sequence was selected before choosing the Toolbox action, this sequence is now listed in the **Selected Elements** window of the dialog. Use the arrows to add or remove sequences or sequence lists from the selected elements. The SignalP service is limited to 2,000 sequences and 200,000 amino acids for one submission. Each sequence may be no longer than 6,000 amino acids.

Click **Next** to set parameters for the SignalP analysis.

You should select which organism group the input sequences belong to. the default is eukaryote (see figure 2.1).

- Eukaryote (default)
- Gram-negative bacteria
- Gram-positive bacteria

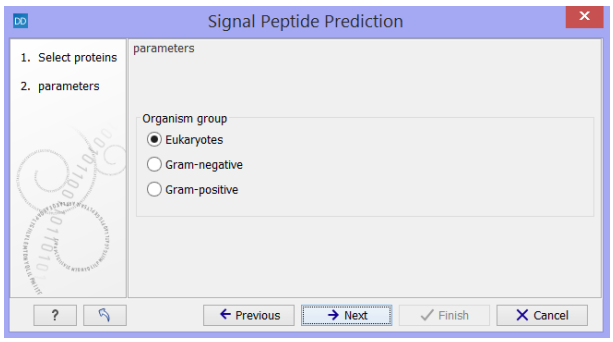

Figure 2.1: *Setting the parameters for signal peptide prediction.*

You can perform the analysis on several protein sequences at a time. This will add annotations to all the sequences and open a view for each sequence if a signal peptide is found. If no signal peptide is found in the sequence, a dialog box will be shown.

The predictions obtained can either be shown as annotations on the sequence, listed in a table or be shown as the detailed and full text output from the SignalP method. This can be used to interpret borderline predictions:

- Add annotations to sequence
- Create table
- Text

Click **Next** to adjust how to handle the results, then click **Finish**.

In order to predict potential signal peptides of proteins, the D-score from the SignalP output is used for discrimination of signal peptide versus non-signal peptide (see section 1.1). This score has been shown to be the most accurate [Klee and Ellis, 2005] in an evaluation study of signal peptide predictors.

After running the prediction as described above, the protein sequence will show predicted signal peptide as annotations on the original sequence (see figure 2.2). Make sure the Side Panel settings of the sequence is so that 'Show annotations' is checked in the 'Annotation layout' palette, and that the annotation type 'Signal peptide' is checked in the 'Annotation types' palette.

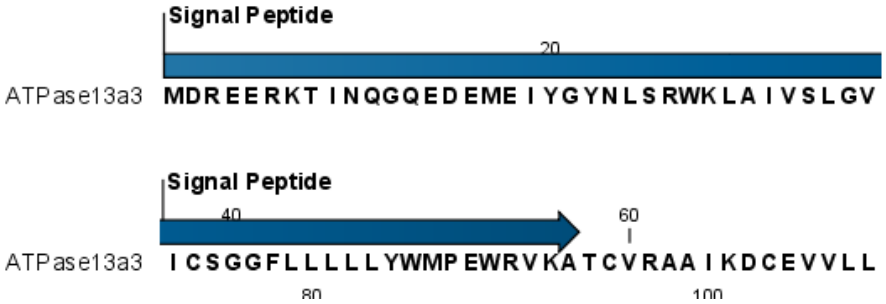

Figure 2.2: *N-terminal signal peptide shown as annotation on the sequence.*

Additional notes can be added through the **Edit annotation** () right-click mouse menu. Undesired annotations can be removed through the **Delete Annotation ()** right-click mouse menu.

Chapter 3

Install and uninstall plugins

SignalP is installed as a plugin.

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

Plugins are installed and uninstalled using the plugin manager.

Help in the Menu Bar | Plugins... ( ) or Plugins ( ) in the Toolbar

The plugin manager has two tabs at the top:

- **Manage Plugins.** This is an overview of plugins that are installed.
- **Download Plugins.** This is an overview of available plugins on QIAGEN Aarhus server.

3.1 Install

To install a plugin, click the **Download Plugins** tab. This will display an overview of the plugins that are available for download and installation (see figure 3.1).

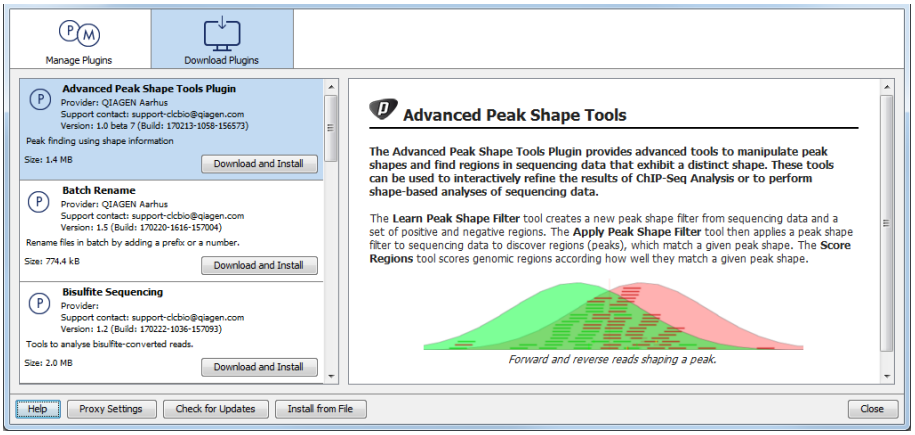

Figure 3.1: *The plugins that are available for download.*

Select SignalP to display additional information about the plugin on the right side of the dialog.

Click **Download and Install** to add the plugin functionalities to your workbench.

Accepting the license agreement

The end user license agreement (EULA) must be read and accepted as part of the installation process. figure 3.2.

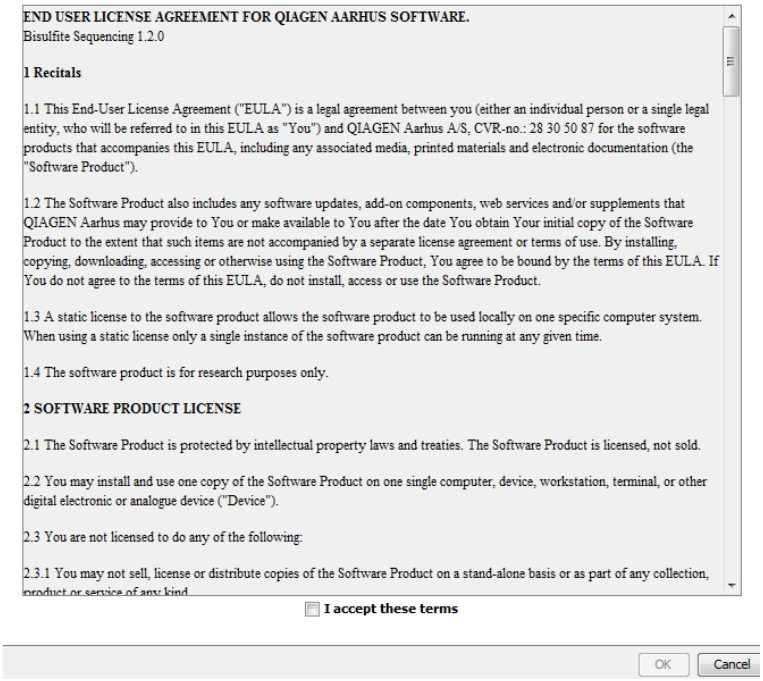

Figure 3.2: *The End User Licenense Agreement is presented during the installation process.*

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.

If SignalP is not shown on the server but you have the installer file on your computer (for example if you have downloaded it from our website), you can install the plugin by clicking the **Install from File** button at the bottom of the dialog and specifying the plugin *.cpa file saved on your computer.

When you close the dialog, you will be asked whether you wish to restart the workbench. The plugin will not be ready for use until you have restarted.

3.2 Uninstall

Plugins are uninstalled using the plugin manager:

Help in the Menu Bar | Plugins... ( ) or Plugins ( ) in the Toolbar

This will open the dialog shown in figure 3.3.

The installed plugins are shown in the **Manage plugins** tab of the plugin manager. To uninstall, select SignalP and click **Uninstall**.

If you do not wish to completely uninstall the plugin, but you do not want it to be used next time you start the Workbench, click the **Disable** button.

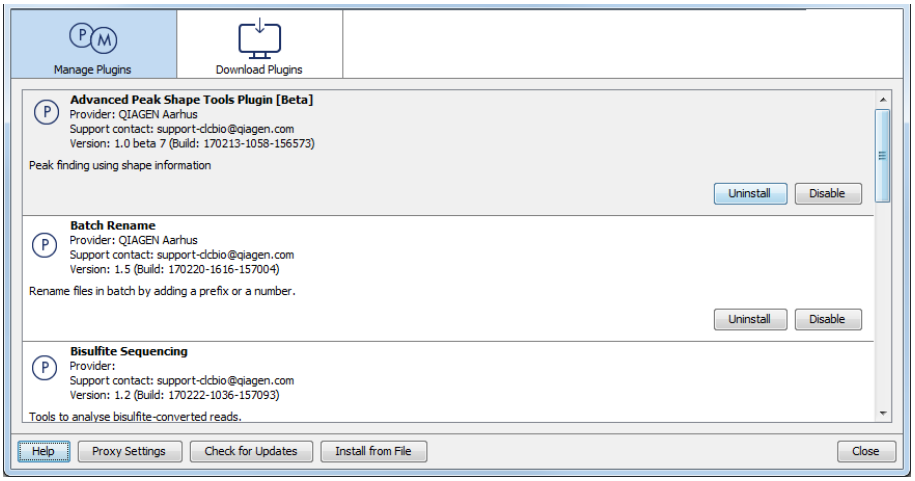

Figure 3.3: *The plugin manager with plugins installed.*

When you close the dialog, you will be asked whether you wish to restart the workbench. The plugin will not be uninstalled until the workbench is restarted.

Bibliography

[Bendtsen et al., 2004] Bendtsen, J. D., Nielsen, H., von Heijne, G., and Brunak, S. (2004). Improved prediction of signal peptides: SignalP 3.0. *J Mol Biol*, 340(4):783–795.

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

[Klee and Ellis, 2005] Klee, E. W. and Ellis, L. B. M. (2005). Evaluating eukaryotic secreted protein prediction. *BMC Bioinformatics*, 6:256.

[Nielsen et al., 1997] Nielsen, H., Engelbrecht, J., Brunak, S., and von Heijne, G. (1997). Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites. *Protein Eng*, 10(1):1–6.