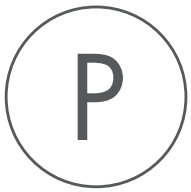

PPFold
Plugin

USER MANUAL

User manual for

PPFold 3.7

Windows, Mac OS X and Linux

February 15, 2017

This software is for research purposes only.

QIAGEN Aarhus
Silkeborgvej 2
Prismet
DK-8000 Aarhus C
Denmark

Contents

Chapter 1

Introduction

Welcome to the PPfold plugin that enables the use of PPfold inside the CLC Workbenches. PPfold is a multithreaded and improved version of the popular pfold program that predict the consensus secondary structure of RNA alignments.

PPfold has been developed in collaboration between Aarhus University, CLC bio and IT University of Copenhagen, funded by the Danish Agency for Science, Technology, and Innovation under the project "PC Mini-Grids for Prediction of Viral RNA Structure and Evolution", #09-061856.

The source code for the PPfold core package is available on the PPfold website: http://www.daimi.au.dk/~compbio/pfold/downloads.html The standalone version of PPfold has additional options for the advanced user, including the possibility to adjust distribution parameters or the use of an alternative parameter file.

If you have used PPfold in your work and found it helpful, please cite: *Z. Suksod, B. Knudsen, M. Vaerum, J. Kjems, E. S. Andersen. Multithreaded comparative RNA secondary structure prediction using stochastic context-free grammars BMC Bioinformatics 12:103, 2011*

Chapter 2

PPfold plugin

Installing the PPfold plugin will install the PPfold tool in the Toolbox as shown on figure 2.1.

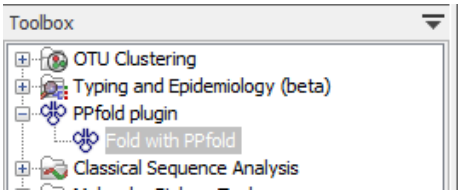

Figure 2.1: *View of the toolbox after having installed the PPfold plugin.*

2.1 Running the ppfold tool

To run the PPfold plugin:

Toolbox | **PPfold plugin** | **Fold with PPfold**

Once the tool wizard has opened (figure 2.2), choose the input you would like to use.

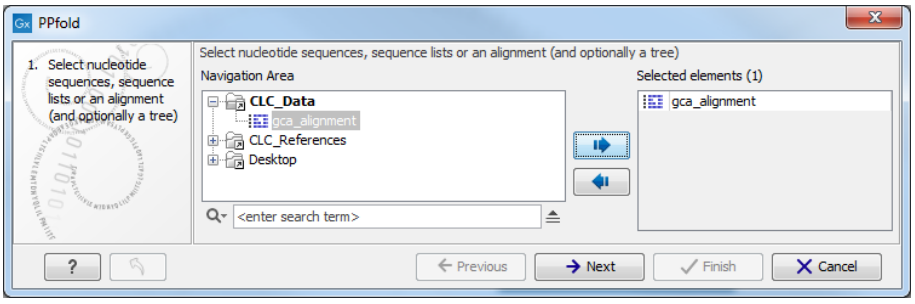

Figure 2.2: *Choose which input to use.*

PPfold can be executed on a number of different objects:

- **One or more nucleotide sequences** PPfold will fold and annotate each nucleotide sequence separately.
- **One or more nucleotide sequences lists** PPfold will fold and annotate each sequence in each sequence list separately.
- **Any combination of sequences and lists**

- **One or several alignments** PPfold will fold and annotate each alignment separately.
- **One alignment and one corresponding phylogenetic tree** PPfold will fold the alignment on the basis of optimized branch lengths in the input tree. Note that the tree must match the alignment: the names of the leaves in the tree must be in one-to-one correspondence to the names of the sequences in the alignment. In addition, it is required that all branches have a length.

Next you can adjust the tool parameters as seen in figure 2.3.

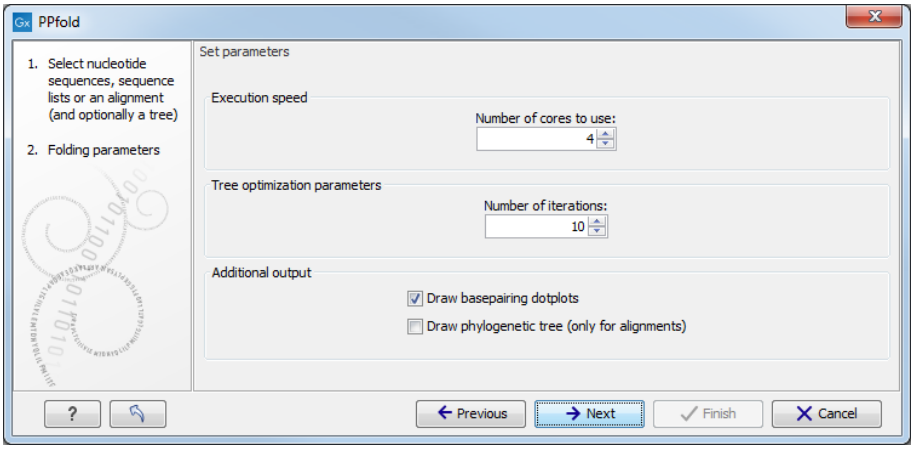

Figure 2.3: *The wizard displaying the options for number of cores and iterations and if you would like to display a basepairing dotplot and a phylogenetic tree.*

- **Execution speed** Choose the number of cores you wish to use in the calculations. More cores will mean faster results, but your computer will run other applications slower meanwhile. Fewer cores mean slower results, but you will be able to use your computer more effectively meanwhile.
- **Tree optimization parameters** If you have selected an alignment or alignments without a phylogenetic tree, PPfold will offer the option to select the maximum number of iterations to use in the optimization of the branch lengths of the tree. PPfold stops adjusting the branch lengths either on convergence, or when the maximum number of iterations is exceeded. In many cases, the branch lengths will converge in fewer than 10 iterations. If convergence is not obtained within 10 iterations, the tree is likely to be good enough anyway. However, increasing the number of iterations might make the tree more accurate. A higher number of iterations will mean a longer execution time, depending on the length of your alignment and the number of sequences in it.
- **Additional output** Select also if you wish to display basepairing dotplots, and in the case of an alignment or alignments without a phylogenetic tree, whether you wish to display the maximum likelihood estimate tree generated by PPfold. PPfold will create the data for both the basepairing dotplots and the phylogenetic tree for its own use, no matter what you choose (so it will not run faster if you choose not to display them).

In the last wizard window, select what to do with the results (Open or Save), and whether you want a log of the process (figure 2.4).

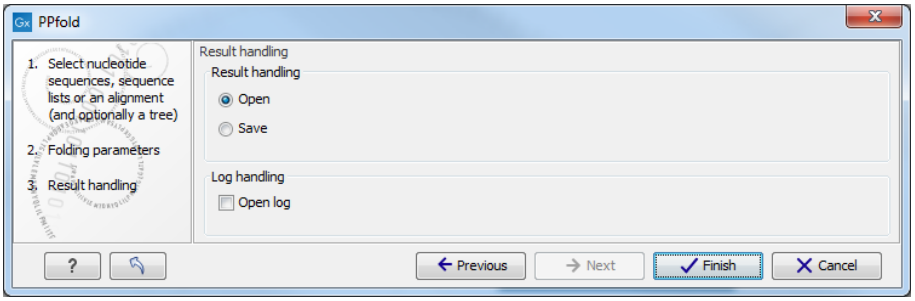

Figure 2.4: *Choose to either open or save the results.*

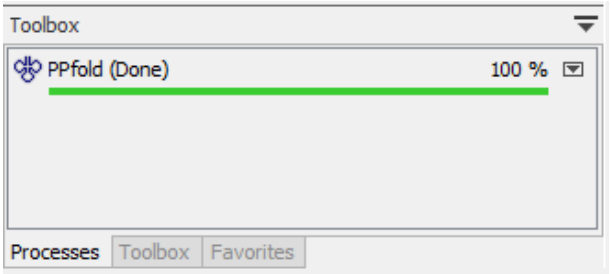

Figure 2.5: *Process bar of the execution.*

PPfold will then fold the alignment. The progress of the algorithm is shown in the progress bar as seen on figure 2.5, which you can also use to cancel the execution.

When the tool is done running, the sequences used for input have been annotated with a secondary structure and the probability of the secondary structure at each position.

In addition, and depending on the options that were available and chosen in the result handling window of the wizard, the tool may output:

- A maximum likelihood estimate (MLE) tree calculated from the alignment used as input (figure 2.6).

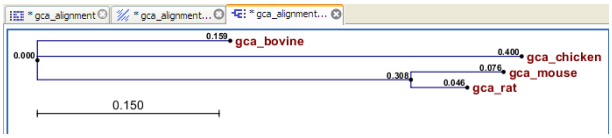

Figure 2.6: *Phylogenetic tree.*

- A dotplot of basepairing probabilities (figure 2.7).

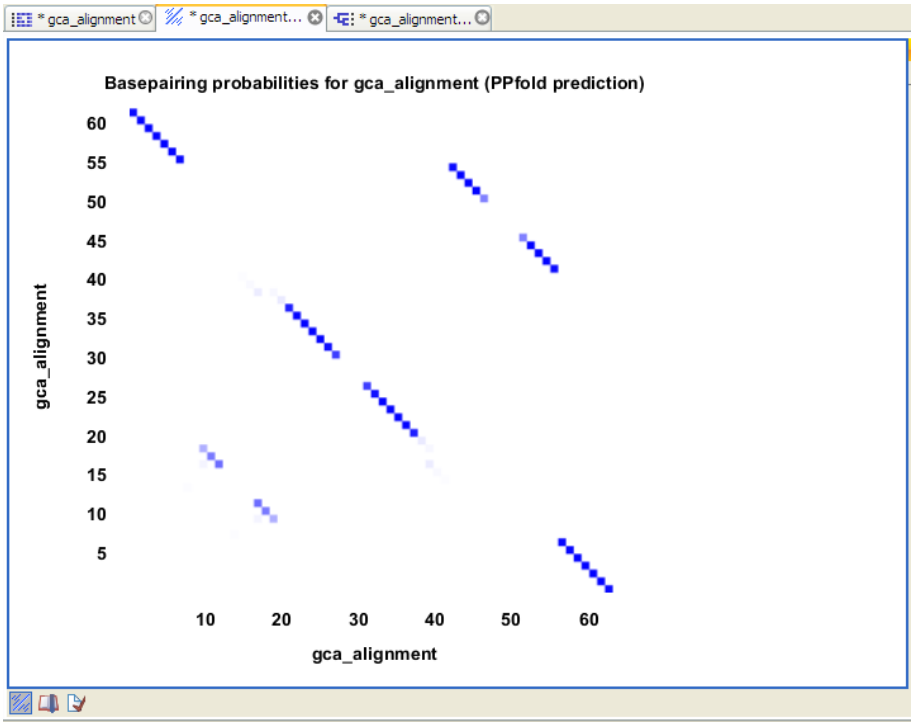

Figure 2.7: *Basepairing dotplots.*

2.2 Visualization

To visualize the structure, open the file that was used as input.

To show the structure under each sequence, go to the Alignment Settings that are in the Side Panel to the right of the View Area. Then choose Nucleotide info -> Secondary structure -> Show (figure 2.8).

To display a drawing of the secondary structure, you must open a sequence on its own first. Right-click the name of a sequence and select "Open sequence" from the drop down menu. A new view with the annotated sequence will open. Click on the Secondary structure button (figure 2.9) at the bottom of the view to display the secondary structure.

The secondary structure will now be displayed as in figure 2.10.

Zoom in on the structure to show the individual nucleotides by right-clicking the magnifying glass and choosing the "Zoom in" magnifier (figure 2.11). You can also choose the panning tool (the hand) to adjust structural elements and arrange the layout of the secondary structure.

Finally, to display structure reliability values as colors, choose in the Side Panel Residue coloring -> Structure values -> PPfold reliabilities, and select foreground and/or background colors as shown on figure 2.12.

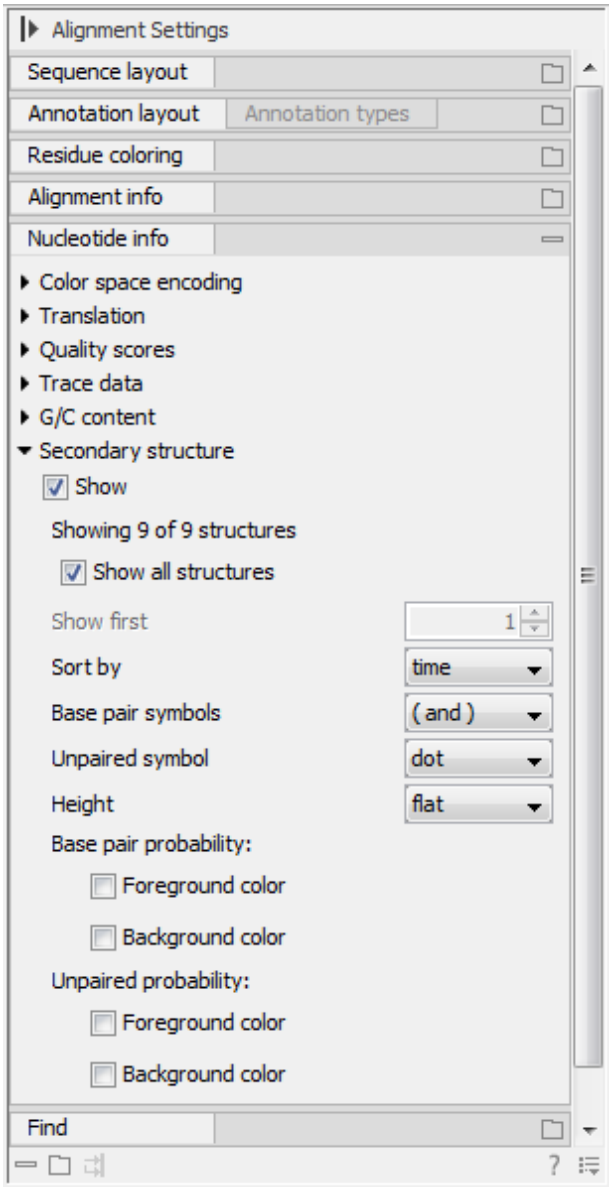

Figure 2.8: *To view the secondary structure click show.*

Figure 2.9: *To display the secondary structure click on the Secondary structure button.*

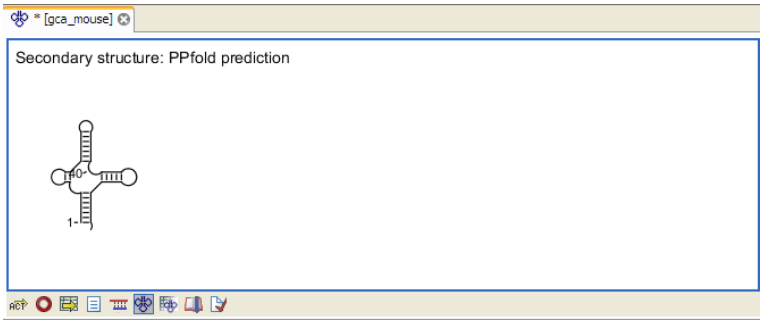

Figure 2.10: *Display of the secondary structure.*

Figure 2.11: *To zoom in click on the magnifying glass with the plus, and to change the structure layout, click on the panning tool represented by the hand.*

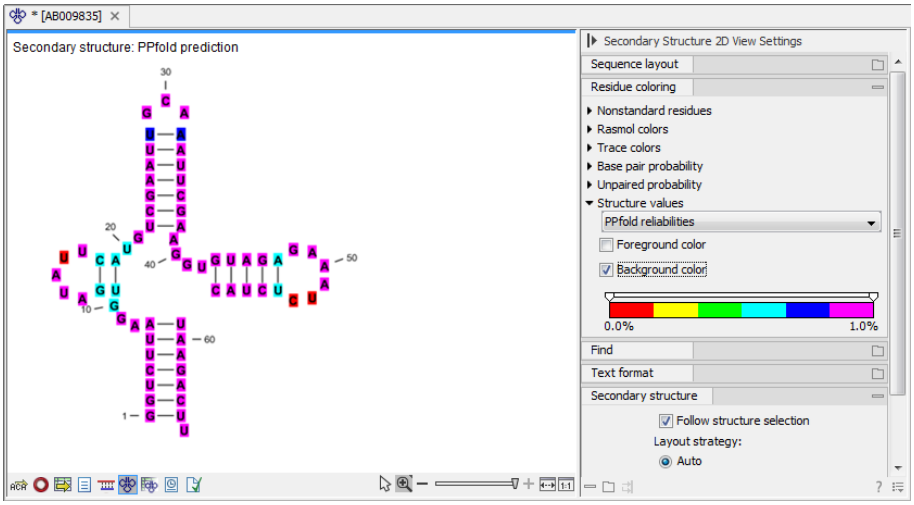

Figure 2.12: *The reliable values can be displayed as colors.*

2.3 Export

To export the alignment, select the alignment and click on File -> Export (figure 2.13). Choose the desired export format from the drop-down menu and save the file.

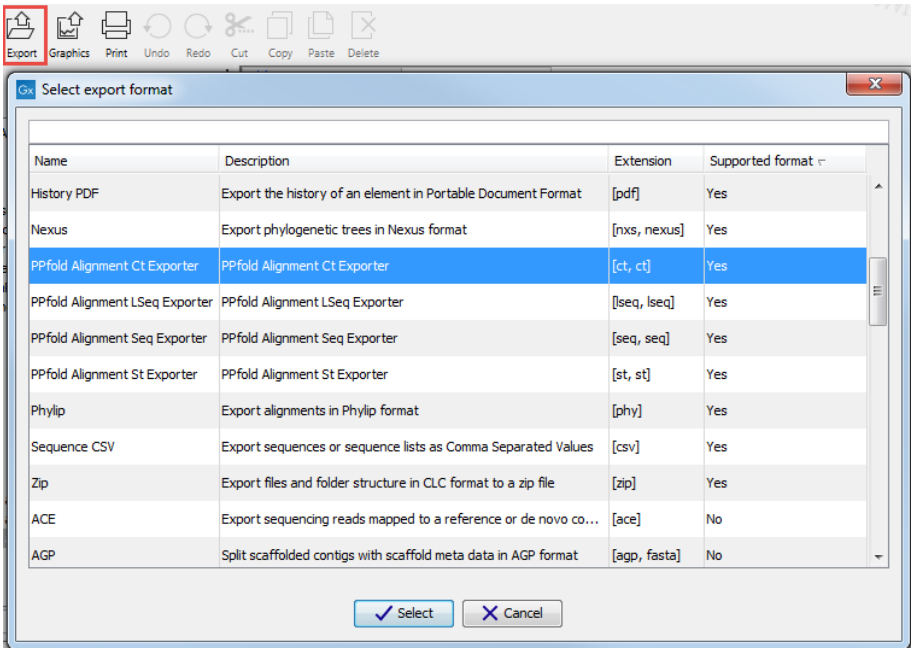

Figure 2.13: *To exsport the alignment, click on Export and choose the desired export format.*

The output of PPfold can be exported using the built-in export functions in the CLC Workbenches. In addition to these, the PPfold plugin comes bundled with a number of custom export functions. Alignments can be exported as:

- **Connectivity Table (.ct) format:** PPfold will attempt to identify the consensus structure in the alignment and export it in .ct format. The length of this .ct file will correspond to the length of the alignment. (If the structures of the sequences are not consistent with a consensus structure, you can potentially get strange outcomes.)
- **SARSE-compatible sequence (.seq) format:**

```
pairingmask ((((((...((.
gca_bovine  AGCCCUGuggUGa
gca_chicken GACUCUGuagUGa
gca_mouse   GGUCUUAaggUGa
gca_rat     AGCCUUAaggUGa
```

Figure 2.14: *SARSE-compatible format.*

  The first line contains the consensus pairing mask. This is followed by each sequence; basepaired nucleotides are marked with uppercase letters, single-stranded nucleotides are lowercase letters. This file is designed for import into the SARSE program.

- **Long sequence (.lseq) format:** contains the consensus pairingmask as well as a derived structure for each sequence in dot-bracket form. This file is designed to ease the extraction of individual sequences and structures from the dataset.
- **Position reliability (.st) format:** pfold-style export format, containing the structure and reliability score for each position for the alignment.

Sequences can be exported in:

- **Position reliability (.st) format:** pfold-stype export format, containing the structure and reliability score for each position of the sequence.

Dotplots can be exported in:

- **Tabbed matrix (.bp) format:** the dotplot is interpreted as basepairing probability scores generated by PPfold. The output is a tabbed matrix containing the scores at each position.

Note that all structure export functions require that the sequences are annotated with precisely one secondary structure.

Chapter 3

Installing the PPFold Plugin

Plugins are installed using the plugin manager. In order to install plugins on Windows, the Workbench must be run in administrator mode: Right-click the program shortcut and choose "Run as Administrator". Then follow the procedure described below.

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 CLC bio's server.

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).

Clicking a plugin will display additional information at the right side of the dialog. This will also display a button: **Download and Install**.

Click the PPFold Plugin and press **Download and Install**. A dialog displaying progress is now shown, and the plugin is downloaded and installed.

If the PPFold Plugin is not shown on the server, and you have it on your computer (for example if you have downloaded it from our website), you can install it by clicking the **Install from File** button at the bottom of the dialog. This will open a dialog where you can browse for the plugin. The plugin file should be a file of the type ".cpa".

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

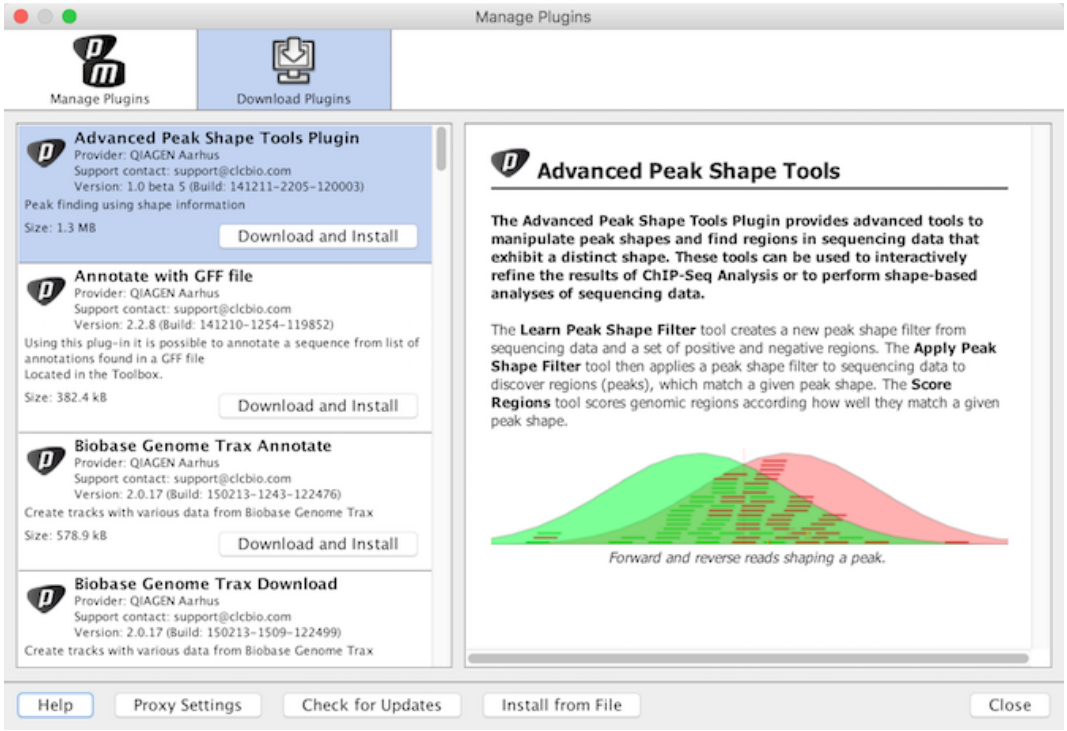

Figure 3.1: *The plugins that are available for download.*

Chapter 4

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 4.1.

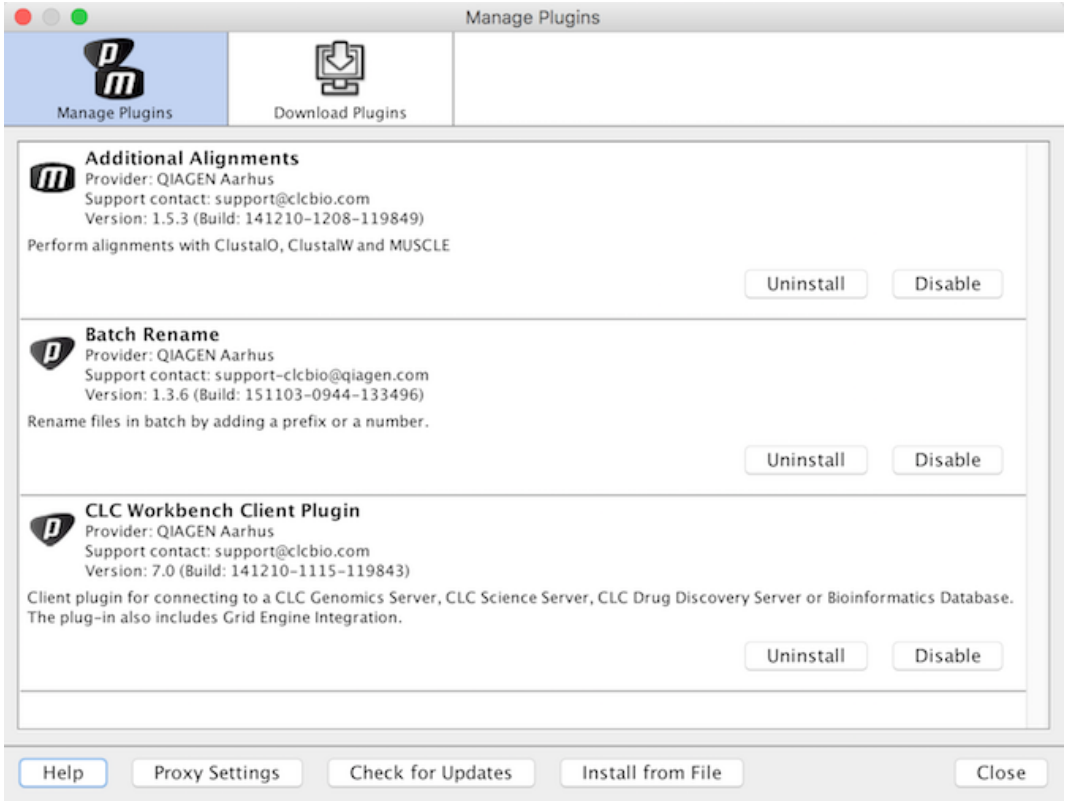

Figure 4.1: *The plugin manager with plugins installed.*

The installed plugins are shown in this dialog. To uninstall:

Click the PPFold Plugin | Uninstall

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

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.