Tutorial

Molecular Biology Basics

December 20, 2024

Sample to Insight

QIAGEN Aarhus A/S · Kalkværksvej 5, 11. · DK - 8000 Aarhus C · Denmark
digitalinsights.qiagen.com · ts-bioinformatics@qiagen.com

Tutorial

Contents

Molecular Biology Basics

This tutorial will take you through the basics of molecular biology work with the CLC Workbenches. All following chapters are independent of each other and can be done separately and in the order of your choice. When in doubt, use the Quick Launch button ( ) to search for a tool. Type in the tool name to find the tool regardless of its location in the Toolbox.

If you do not have the example data, please go to the **Help** menu to import it.

Working with annotations

Annotations are the core of many analyses in the Workbench. Once an annotation has been added to a sequence, it stays there even when the sequence is transformed to be part of an alignment, a BLAST result or a sequence list.

Because annotations are so fundamental to many of the tasks performed in the Workbench, there are a number of ways to browse, search for, view and edit annotations. This section takes you through a lot of different areas of the Workbench to show some of the places where you can work with annotations.

Browsing and viewing annotations in sequence views

1. Open sequence *ATP8a1 mRNA* from the Example data by double-clicking the sequence. Click **Annotation types** in the **Side Panel** and you will see a list of the types of annotations on this sequence. Click on the arrow to the right of the *Exon* annotation type. This will show exon annotations on the sequence (see figure 1).

   As this sequence is very long, you cannot see the whole sequence, and the exons may therefore not be visible. In this situation, you can either Zoom out to find the annotations, or you can use the list of annotation on the Side Panel to jump directly to one of the annotations.

2. Click *Exon 2*, and the beginning of the annotation will be visible in the view (see figure 2).

Annotations are displayed as arrows, going from left to right on the positive strand, and going from right to left on the negative strand. Placing your mouse cursor on the arrow will show additional information about the annotation. Try to place the mouse cursor on the blue gene annotation, and you can see more information about it (see figure 3).

This way of displaying and accessing annotations is similar for both circular views, alignments, and all other views displaying sequence residues.

Adding and editing annotations You can add your own annotations to a sequence. A new annotation is most easily added if you first select a region on the sequence.

1. In the sequence view, select a region from residue 10 to 26 (see the status bar in the lower right corner of the Workbench).
2. Then right-click the selection and choose to **Add Annotation ( )**.

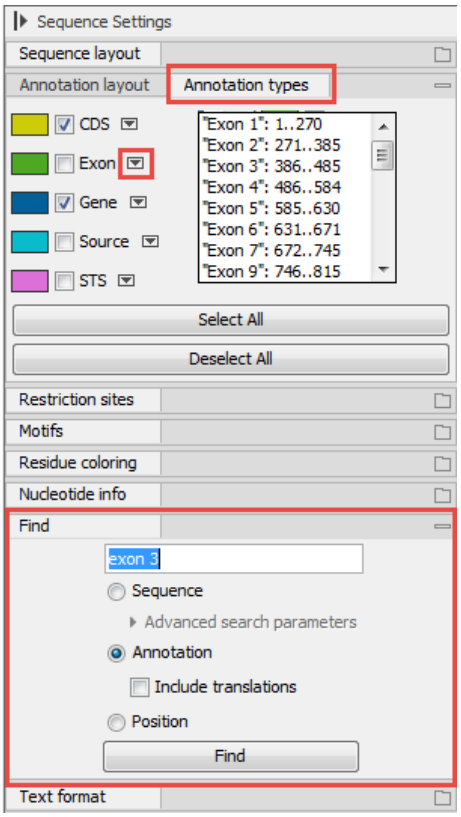

Figure 1: *Click in the list to go directly to this annotation. Using the Find section of the Side Panel, it is also possible to type in the name of the annotation when the Annotation radio button is checked.*

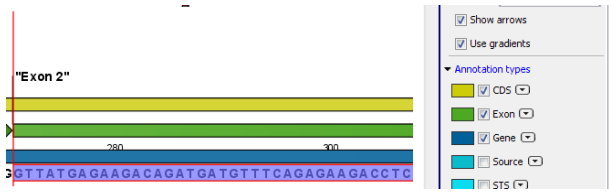

Figure 2: *The view jumps to the beginning of exon 2.*

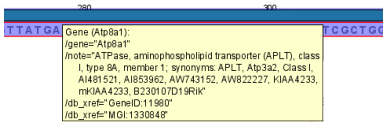

Figure 3: *Information about the annotation.*

3. This will display a dialog where you can enter more information about the annotation. Enter *Test* as the name of the annotation (see figure 4).
4. Click **OK** and the annotation is added to the sequence (see figure 5).
5. Now click the **Show Annotation Table** ( ) icon at the bottom of the view. This will display a list of all the annotations on the sequence.
6. Click Select All under Shown annotation types in the right hand side panel.
7. Enter *Test* into the **Filter** in the annotation table, and your newly added annotation will appear.

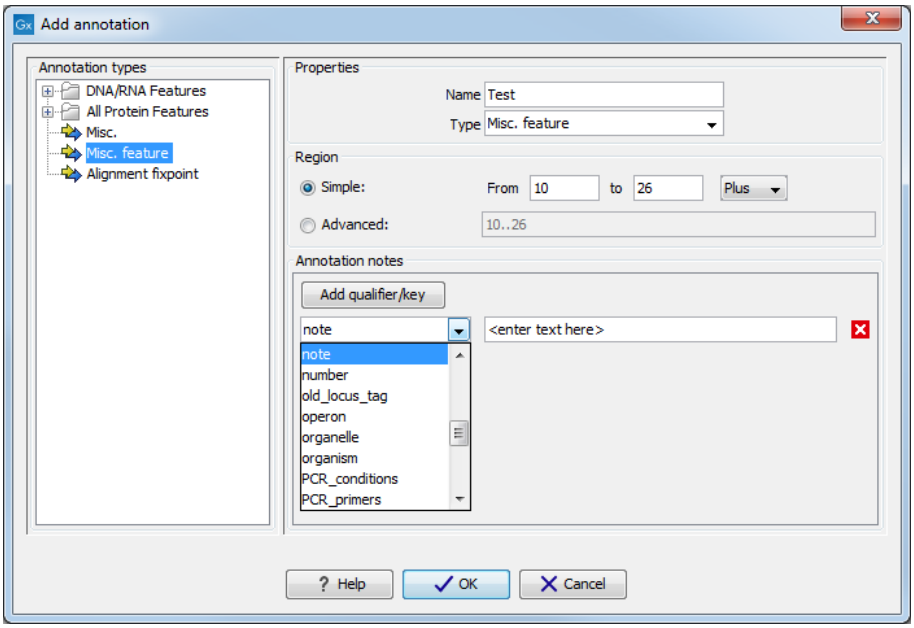

Figure 4: *The Add Annotation dialog.*

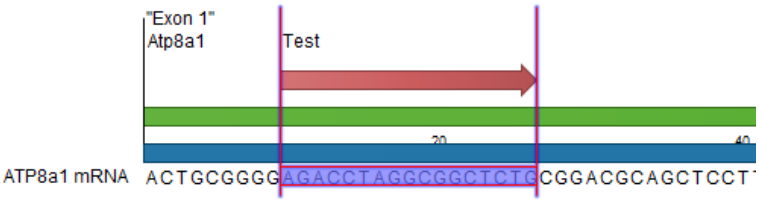

Figure 5: *The annotation has been added.*

You can now double-click any of the **Name**, **Type** or **Region** fields to edit the annotation. Double-click the **Type** which is now "Misc. feature" and enter *My type* and press **Enter**.

The table should now look as shown in figure 6.

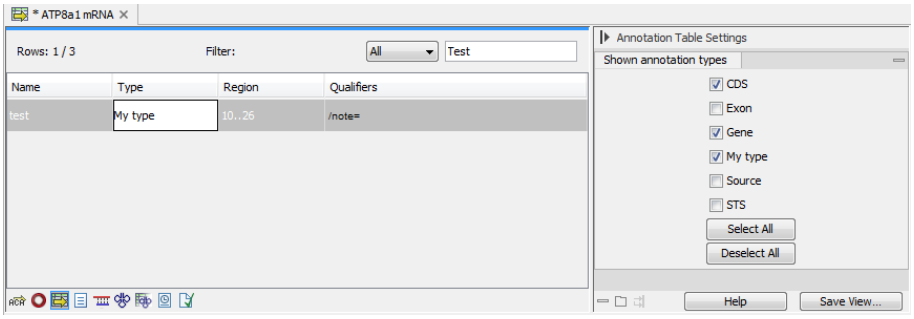

Figure 6: *The type of the annotation is now My type.*

Copying annotations Annotations can be copied from one sequence to another or you can create a duplicate of the annotation by pasting it back on the same sequence. Because it is very easy to edit the region of the annotation, this can be a quick way of adding a number of similar annotations.

1. Select the *Test* annotation in the table, and choose to copy and paste it. You can use Ctrl+c and Ctrl+v or the copy and paste button in the Edit menu at the top of the Workbench.

2. Double-click the region of the new annotation and type *30..46* and press **Enter**. You now have two annotations as shown in figure 7.

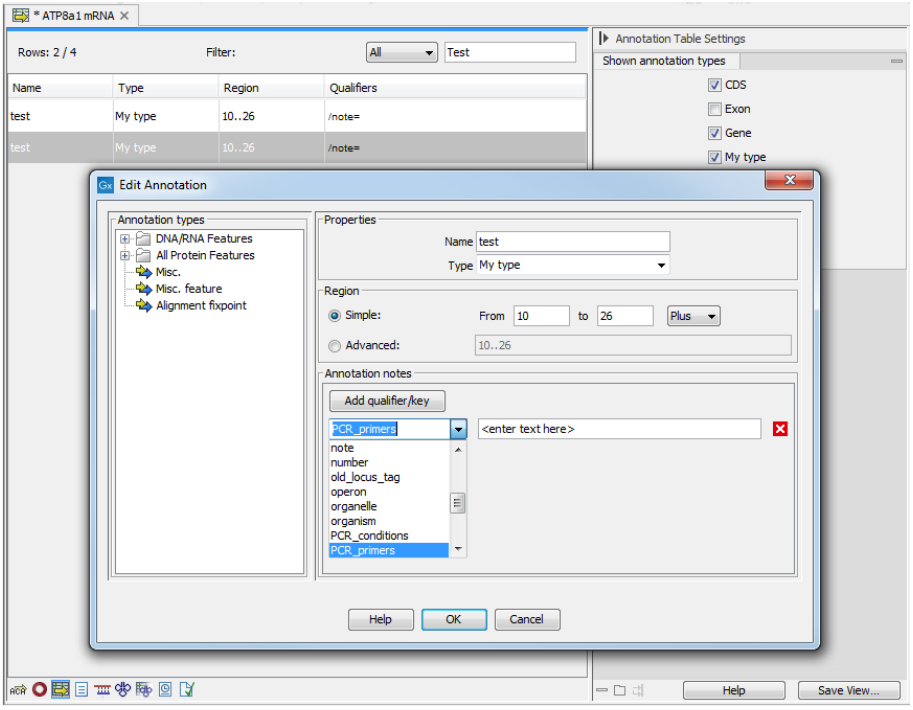

Figure 7: *Two copies of the Test annotation.*

Design Primers

In this tutorial, you will see how to use the *CLC Genomics Workbench* to find primers for PCR amplification of a specific region.

We use the pcDNA3-atp8a1 sequence from the 'Primers' folder in the Example data. This sequence is the pcDNA3 vector with the ATP8A1 gene inserted. In this section, we wish to design primers that would allow us to generate a PCR product covering the insertion point of the gene. This would let us use PCR to check that the gene is inserted where we think it is.

1. First, open the pcDNA3-atp8a1 sequence in the Primer Designer by double-clicking on the sequence name in the Navigation Area. Use the icon ( ) in the lower left corner of the View Area (figure 8) to open the Primer Designer view.

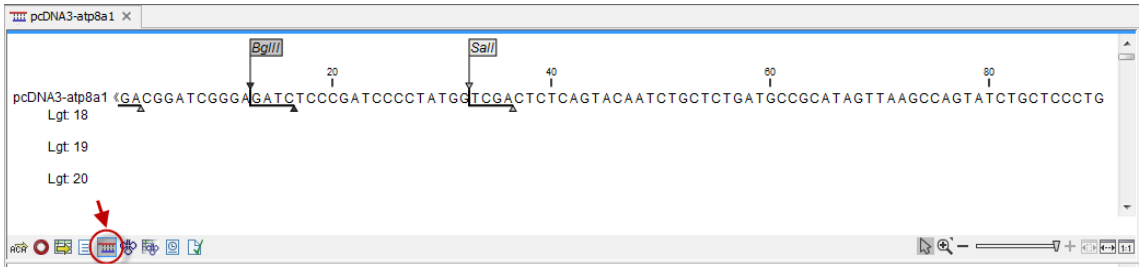

Figure 8: *The sequence that you would like to design primers to. The sequence has been opened in the Primer Designer. The red arrow highlights the icon that symbolizes the Primer Designer view.*

Note that you can also use the Launch button ( ) to open the tool **Design Primers**

(**)**. This will open a wizard where you can select the pcDNA3-atp8a1 sequence from the 'Primers' folder in the Example data.

2. Now the sequence is opened and shown at single nucleotide resolution. To get an overview of the sequence, we will zoom out a bit by clicking on the **Fit Width** icon ( ) that are found in the lower right corner of the **View Area**. You can now see the blue gene annotation labeled Atp8a1, and just before that there is the green CMV promoter (see figure 9). This may be hidden behind restriction site annotations. Remember that you can always choose not to show these by altering the settings in the right hand panel.

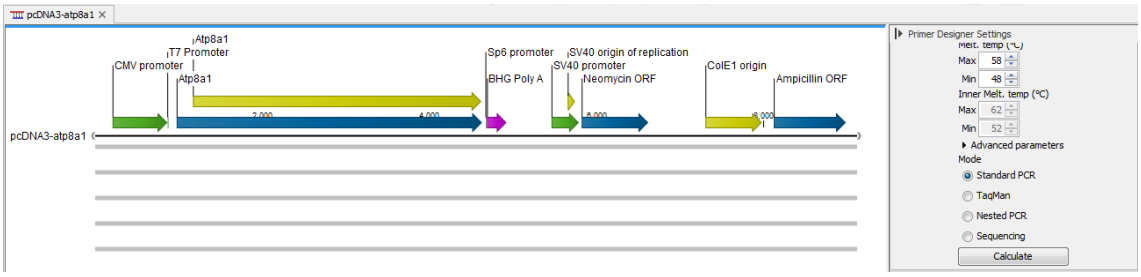

Figure 9: *Zoom out to get an overview of the sequence.*

3. In this tutorial, we want the forward primer to be just before the gene. Zoom in with the zoom in tool ( ) found in the lower right corner to make the selection. Go back to the select mode by clicking the ( ) next to the zoom in in the lower right corner and select the region between 600 and 900, right-click and choose "Forward primer region here" ( ) (see figure 10).

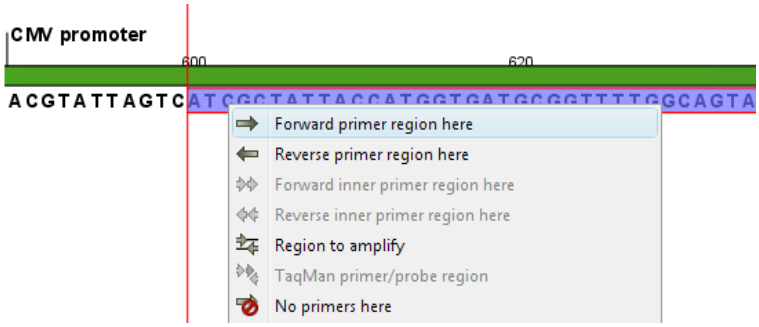

Figure 10: *Right-clicking a selection and choosing "Forward primer region here".*

4. This will add an annotation to this region, and five rows of red and green dots are seen below as shown in figure 11.

   Each line consists of a number of dots, each representing the *starting point* of a possible primer. For example, the first dot on the first line (primers of length 18) represents a primer starting at the dot's position and with a length of 18 nucleotides (shown as the white area in figure 12):

5. Position the mouse cursor over a dot. A box will appear, providing data about this primer. Clicking the dot will select the region where that primer would anneal (figure 13).

Note that some of the dots are colored red. This indicates that the primer represented by this dot does not meet the requirements set in the **Primer parameters** (see figure 14):

Figure 11: *Five lines of dots representing primer suggestions. There is a line for each primer length - 18bp through to 22 bp.*

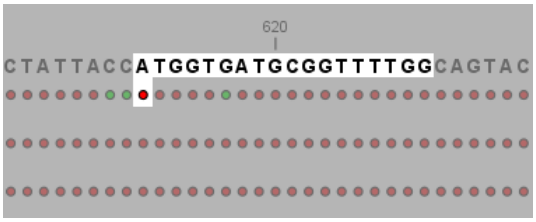

Figure 12: *The first dot on line one represents the starting point of a primer that will anneal to the highlighted region.*

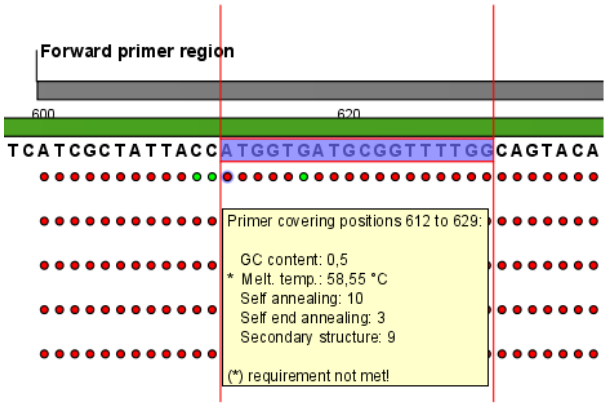

Figure 13: *Clicking the dot will select the corresponding primer region. Hovering the cursor over the dot will bring up an information box containing details about that primer.*

The default maximum melting temperature is 58. This is the reason why the primer in figure 13 with a melting temperature of 58.55 does not meet the requirements and is colored red. Note that there is an asterisk (*) before the melting temperature in the primer tooltip to indicate that this primer does not meet the requirements regarding melting temperature. If you raise the maximum melting temperature to 59, the primer will meet the requirements and the dot becomes green.

By adjusting the **Primer parameters** you can define primers to meet your specific needs. Since the dots are dynamically updated, you can immediately see how a change in the primer parameters affects the number of red and green dots.

Calculating a primer pair Until now, we have been looking at the forward primer. To calculate a primer pair:

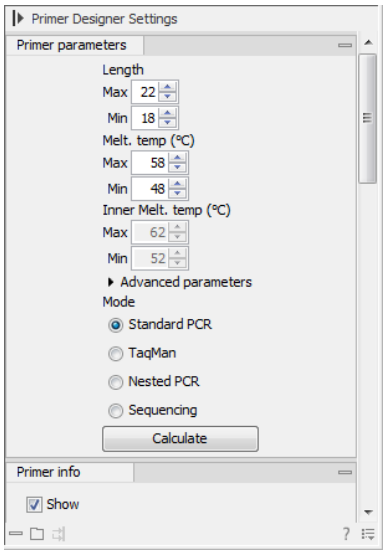

Figure 14: *The Primer parameters.*

1. Select the region for the reverse primer from position 1200 to 1400 and right-click on the selection. Choose "Reverse primer region here" ( )
2. The two regions should now be located as shown in figure 15.

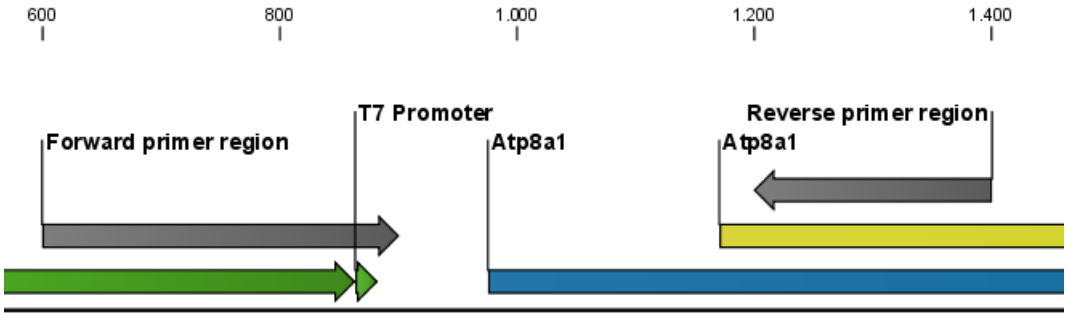

Figure 15: *A forward and a reverse primer region.*

3. Now, you can let *CLC Genomics Workbench* calculate all the possible primer pairs based on the **Primer parameters** that you have defined: click the **Calculate** button (in the right hand side panel).
4. In the first wizard window you could modify parameters regarding the combination of the primers if needed but for now, just leave them unchanged and click on **Calculate** again.
5. This will open a table showing the possible combinations of primers. Using the right hand side panel, you can specify the information you want to display in the table, such as showing **Fragment length** (see figure 16).
6. Clicking a primer pair in the table will make a corresponding selection on the sequence in the view above. At this point, you can either settle on a specific primer pair or save the table for later. If you want to use the first primer pair for your experiment, right-click this primer pair in the table and save the primers.
7. You can also mark the position of the primers on the sequence by selecting **Mark primer annotation on sequence** in the right-click menu (see figure 17).

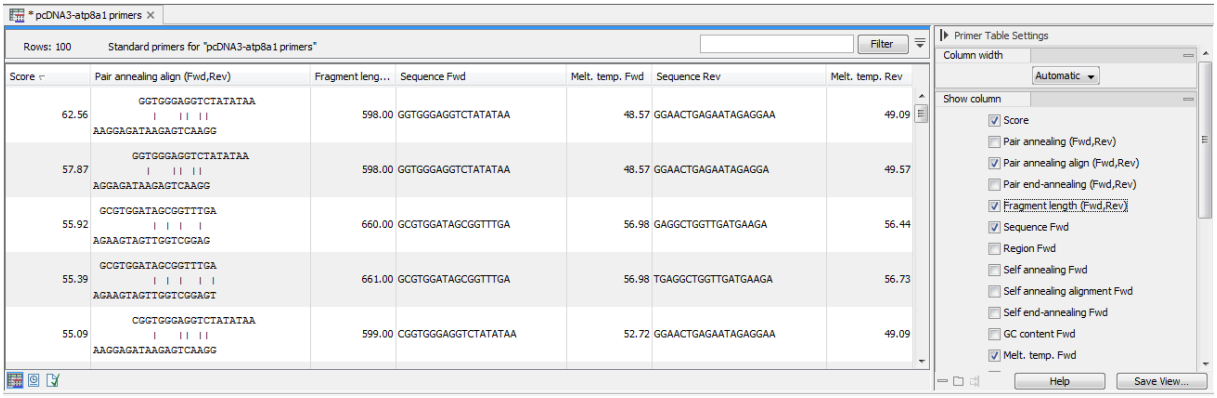

Figure 16: *A list of primers. To the right is the Side Panel showing the available choices of information to display.*

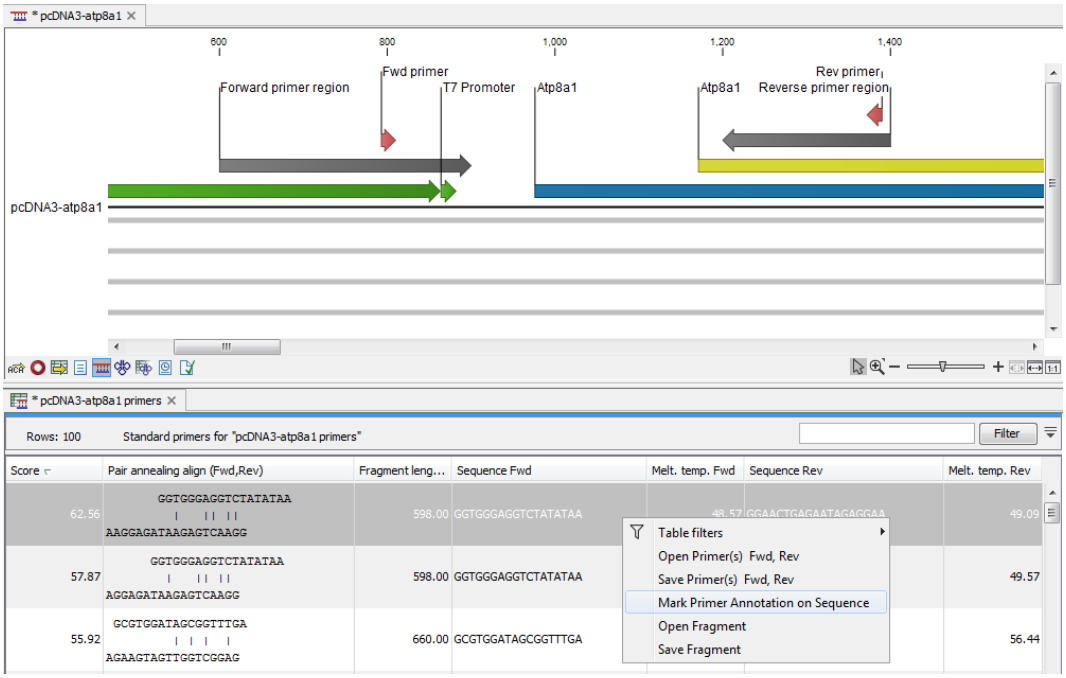

Figure 17: *The options available in the right-click menu. Here, "Mark primer annotation on sequence" has been chosen, resulting in two annotations on the sequence above.*

In silico cloning

In this section, the goal is to make a virtual PCR-amplification of a gene using primers with restriction sites at the 5' ends, and to insert the gene in a multiple cloning site of an expression vector. We start off with a set of primers, a DNA template sequence and an expression vector loaded into the Workbench (figure 18).

Double-click the `ATP8a1 mRNA` sequence and zoom to **Fit Width** ( ). You will see a yellow annotation for the coding part of the gene, which is the part that we want to insert into the `pcDNA4_TO` vector. The primers have already been designed using the primer design tool as explained in the section above.

Add restriction sites to the primers

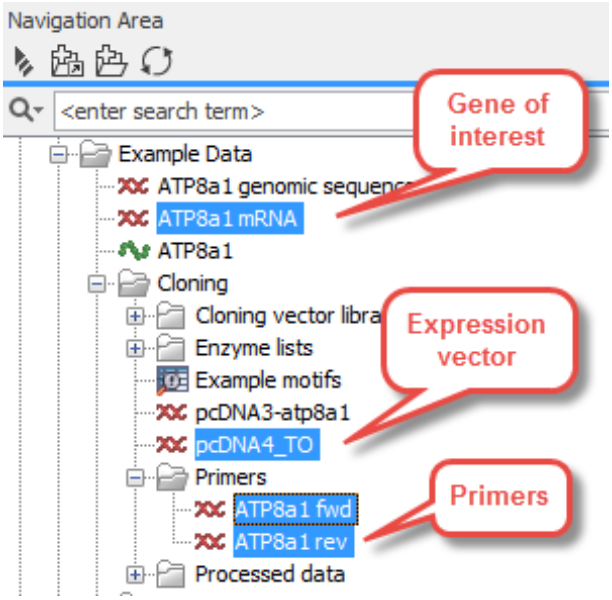

Figure 18: *The data to use in this tutorial.*

1. First, we want to add restriction sites to the primers. In order to see which restriction enzymes can be used, we create a split view of the vector and the fragment to insert. In this way we can easily make a visual check to find enzymes from the multiple cloning site in the vector that do not cut in the gene of interest. To create the split view, open both the ATP8a1 mRNA and pcDNA4_TO sequences in the Navigation Area.

2. Right click on the tab of one of the sequences in the View Area, and select the options **View** followed by **Split Horizontally (****)**.

3. Switch to the **Circular view** () for the pcDNA4_TO sequence.

4. **Zoom in** () on the multiple cloning site downstream of the green `CMV` promoter annotation. You should now have a view similar to the one shown in figure 19.

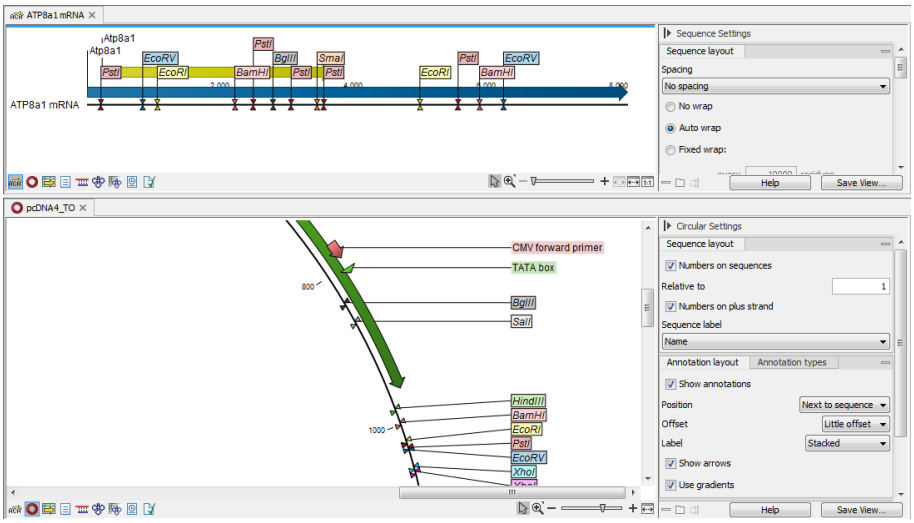

Figure 19: *Check cut sites.*

By looking at the enzymes we can see that both *HindIII* and *XhoI* cut in the multi-cloning site

of the vector and not in the Atp8a1 gene. Note that you can add more enzymes to the list in the **Side Panel** by clicking **Manage Enzymes** under the **Restriction Sites** group.

5. Open the `ATP8a1 fwd` primer sequence. When it opens, double-click the name of the sequence to make a selection of the full sequence. If you do not see the whole sequence turn purple, please make sure you have the Selection Tool chosen, and not one of the other tools available from the top right side of the Workbench (e.g. Pan, Zooming tools, etc.)

6. Once the sequence is selected, right-click and choose to **Insert Restriction Site Before Selection** as shown in figure 20.

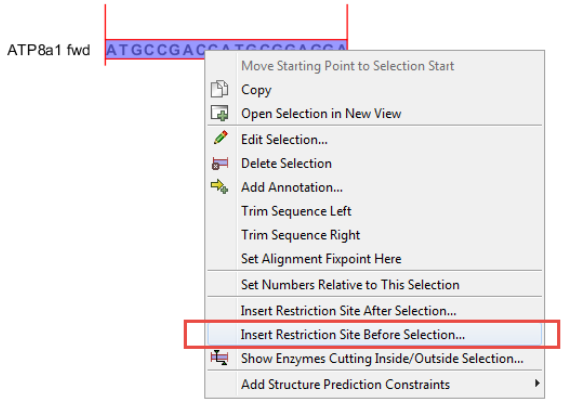

Figure 20: *Adding restriction sites to a primer using the Insert Restriction Site Before Selection command.*

7. In the **Filter** box enter *HindIII* and click on it. At the bottom of the dialog, add a few extra bases 5' of the cut site (this is done to increase the efficiency of the enzyme) as shown figure 21.

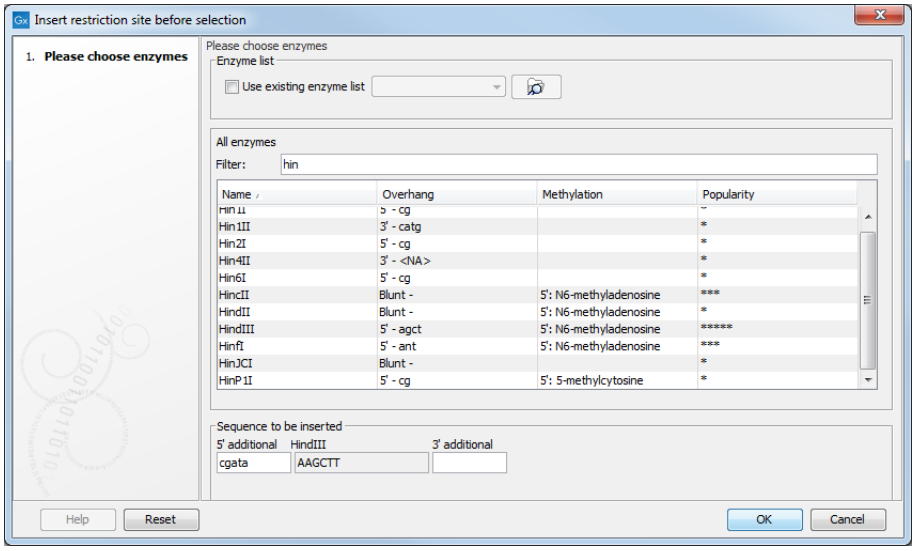

Figure 21: *Choose the restriction sites to add to the primer.*

8. Click **OK** and the sequence will be inserted at the 5' end of the primer as shown in figure 22.

Figure 22: *The restriction site is added.*

9. Perform the same process for the `ATP8a1 rev` primer, this time using *XhoI* instead. This time, you should also add a few bases at the 5' end as was done previously. **Note!** The `ATP8a1 rev` primer is designed to match the negative strand, so the restriction site should be added at the 5' end of this sequence as well (**Insert Restriction Site Before Selection**).
10. **Save** ( ) the two primers.

Simulate a PCR to create the fragment Now, we want to extract the PCR product from the template `ATP8a1 mRNA` sequence using the two primers with restriction sites.

1. Use the Launch button ( ) to find the tool **Find Binding Sites and Create Fragments ( )**.
2. Select the `mRNA` sequence and click **Next**.
3. In the next dialog, use the **Browse** ( ) button to select the two primer sequences figure 23. Click **Next**.

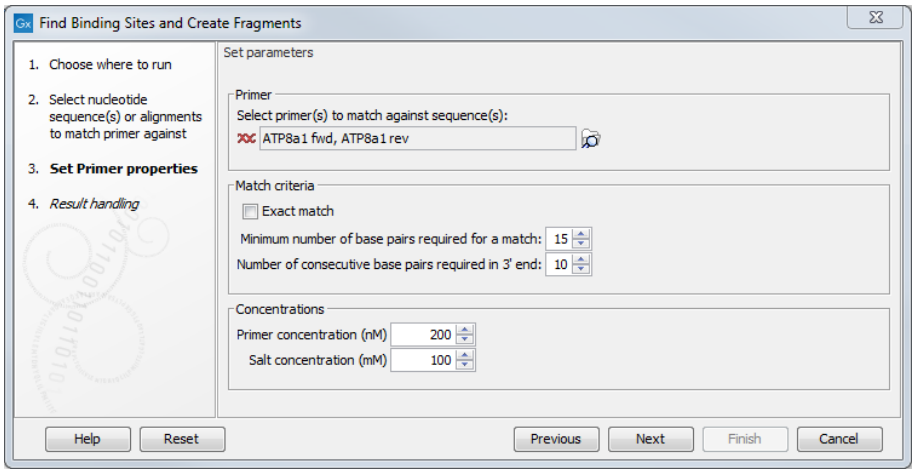

Figure 23: *Setting the primers properties.*

4. Adjust the output options as shown in figure 24.

   Click **Finish** and you will now see the fragment table displaying the PCR product.
5. In the **Side Panel** you can choose to show information about melting temperature for the primers.
6. Right-click the fragment and select **Open Fragment** as shown in figure 25.

   This will create a new sequence representing the PCR product.
7. **Save** ( ) the sequence in the `Cloning` folder and close the views. You do not need to save the fragment table.

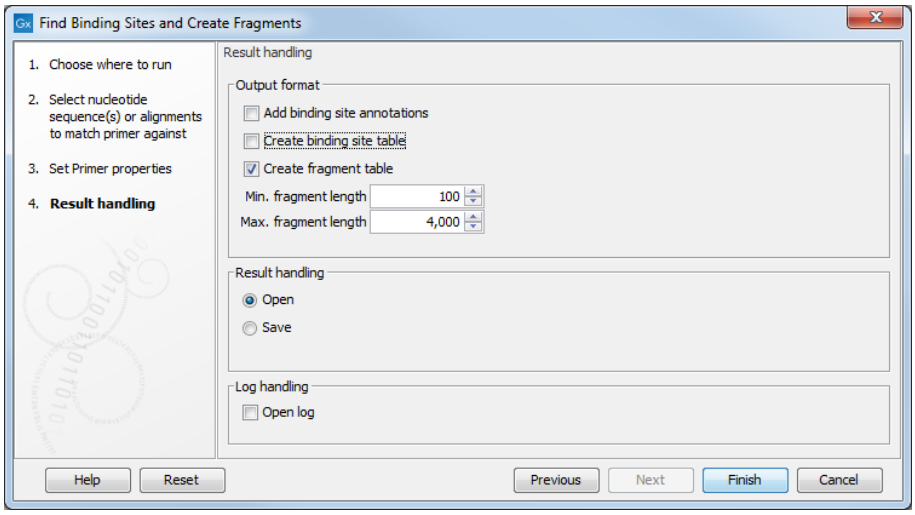

Figure 24: *Creating the fragment table including fragments up to 4000 bp.*

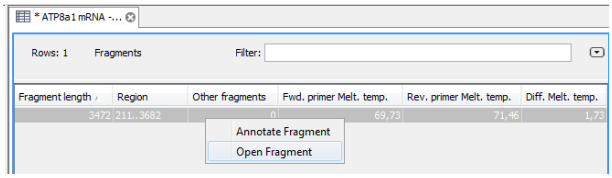

Figure 25: *Opening the fragment as a sequence.*

Specify restriction sites and perform cloning The final step in this tutorial is to insert the fragment into the cloning vector:

1. In the toolbox, find:
   Molecular Biology Tools ( ) | Cloning ( )| Restriction Based Cloning ( )
2. Select the `Fragment (ATP8a1 mRNA (ATP8a1 fwd – ATP8a1 rev))` sequence you just saved and also select the `pcDNA4_TO` cloning vector also located in the `Cloning` folder and click **Next**.
3. The cloning editor will now create a sequence list of the selected fragment and vector sequence as shown in figure 26. Select the vector sequence in the drop down menu at the top of the editor and specify it as vector by clicking the "Change to Current" button.

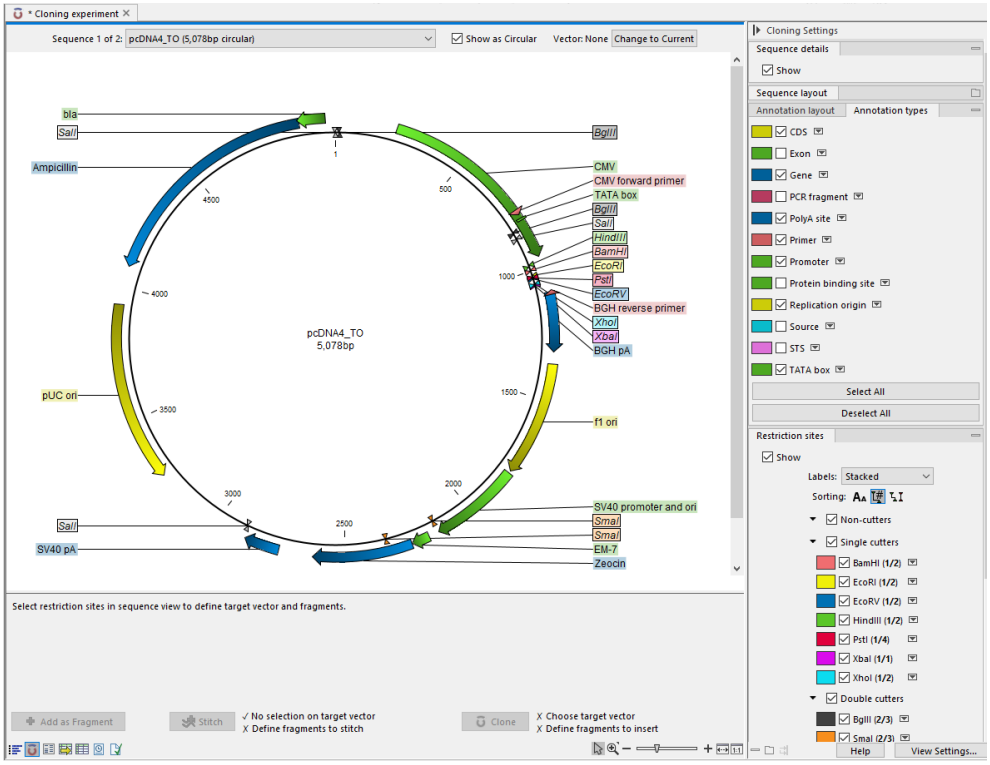

Figure 26: *Cloning editor.*

4. You will now see the cloning editor where you will see the `pcDNA4_TO` vector in a circular view. Press and hold the Ctrl (⌘ on Mac) key while you click first the *HindIII* site and next the *XhoI* site (see figure 27).

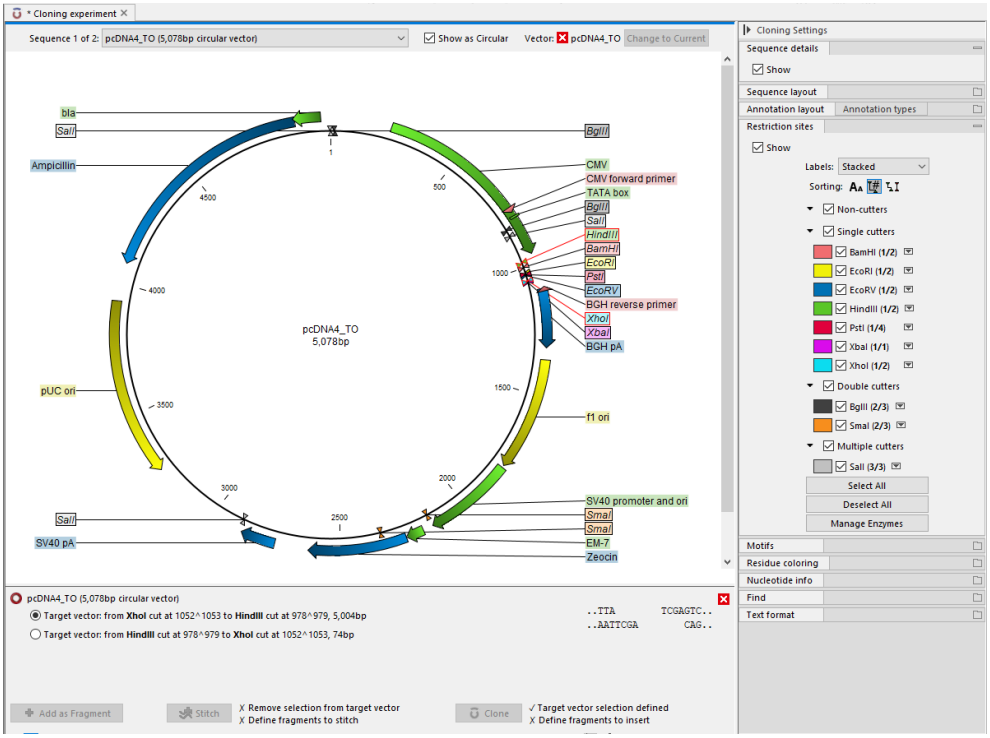

Figure 27: *Press and hold the Ctrl key while you click first the* HindIII *site and next the XhoI site in the cloning vector.*

At the bottom of the view you can now see information about how the vector will be cut open. Since the vector has now been split into two fragments, you can decide which one to use as the target vector. If you selected first the *HindIII* site and next the *XhoI* site, the Workbench has already selected the right fragment as the target vector. If you click one of the vector fragments, the corresponding part of the sequence will be highlighted.

5. Next step is to cut the fragment. At the top of the view you can switch between the sequences used for cloning (at this point it says `pcDNA4_TO 5.078bp circular vector`). Switch to the fragment sequence and perform the same selection of cut sites as before while pressing the Ctrl (⌘ on Mac) key. You should now see a view identical to the one shown in figure 28.

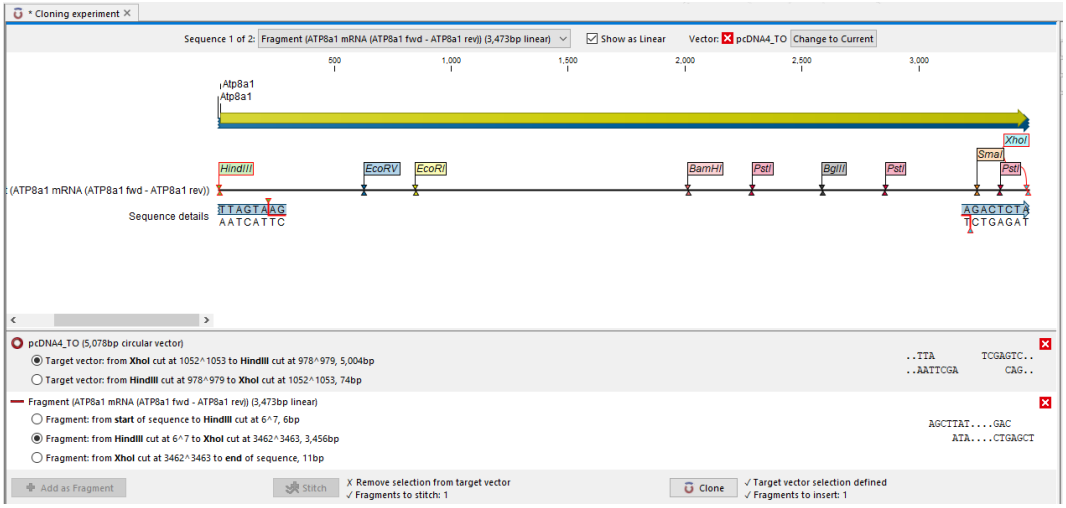

Figure 28: *Change the view to display the fragment sequence. Press and hold the Ctrl key while you click first the* HindIII *site and next the* XhoI *site.*

6. When this is done, the **Clone** ( ) button at the lower right corner of the view is active because there is now a valid selection of both fragment and target vector. Click the **Clone** ( ) button and you will see the dialog shown in figure 29.

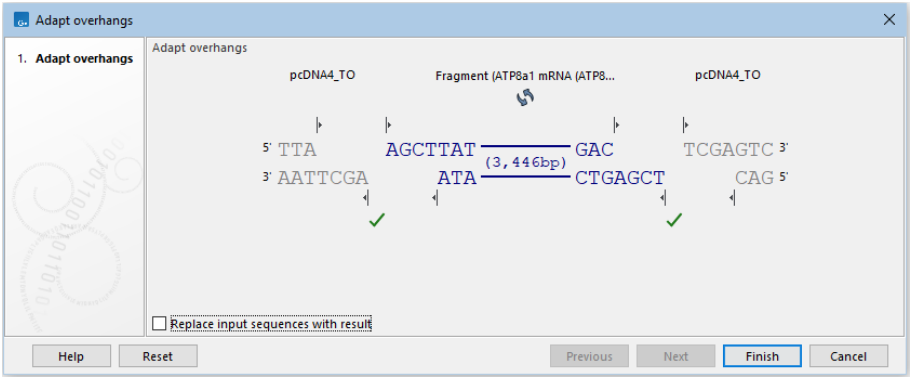

Figure 29: *Showing the insertion point of the vector*

This dialog lets you inspect the overhangs of the cut site, showing the vector sequence on each side and the fragment in the middle. The fragment can be reverse complemented by clicking the **Reverse complement fragment** ( ) but this is not necessary in this case. Click **Finish** and your new construct will be opened.

7. When saving your work, there are two options:

   - Save the Cloning Experiment. This is saved as a sequence list, including the specified cut sites. This is useful if you need to perform the same process again or double-check details.
   - Save the construct shown in the circular view. This will only save the information on the particular sequence including details about how it was created (this can be shown in the **History** view).

   You can, of course, save both. In that case, the history of the construct will point to the sequence list in its own history.

The construct is shown in figure 30.

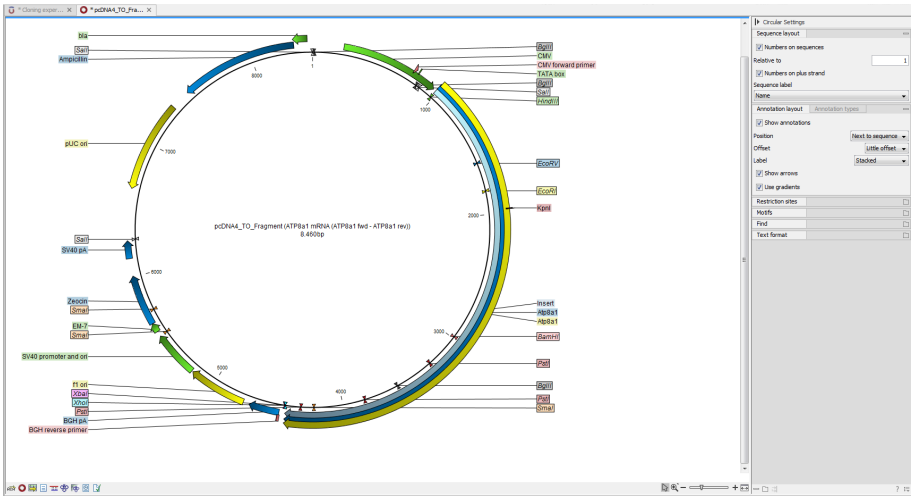

Figure 30: *The Atp8a1 gene inserted after the CMV promoter*

Notes on restriction sites

This section will show you how to find restriction sites and annotate them on a sequence.

There are two ways of finding and showing restriction sites: from the right hand side panel, or by using the **Restriction Site Analysis ( )** tool. In many cases, the dynamic restriction sites found in the **Side Panel** of sequence views is sufficient to provide the information needed, but the tool provides better control of the analysis and more output options, such as a table of restriction sites and a list of restriction enzymes that can be saved for later use.

The Side Panel way of finding restriction sites When you open a sequence, there is a **Restriction sites** setting in the **Side Panel**. By default, 10 of the most popular restriction enzymes are shown (see figure 31).

Figure 31: *Showing restriction sites of ten restriction enzymes.*

The restriction sites are shown on the sequence with an indication of cut site and recognition sequence. In the list of enzymes in the **Side Panel**, the number of cut sites is shown in parentheses for each enzyme (e.g. *SalI* cuts three times). If you wish to see the recognition sequence of the enzyme, place your mouse cursor on the enzyme in the list for a short moment, and a tool tip will appear.

You can add or remove enzymes from the list by clicking the **Manage enzymes** button.

The Toolbox way of finding restriction sites

1. Open the tool called **Restriction Site Analysis ( )** using the Launch button ( ).
2. In the first window (figure 32), called "Select DNA/RNA sequence(s)", you can see a view of the Navigation Area and select the ATP8a1 mRNA sequence from the Example Data folder. Click **Next** to proceed to the second step of the wizard.
3. Now you can set the parameters for the restriction map analysis. In most cases you can select the enzymes displayed in the table at the bottom of that wizard window, but if you are interested in using enzymes you have in your lab's freezer for example, you can use a custom made list that you created and saved previously. To use this list, select **Use**

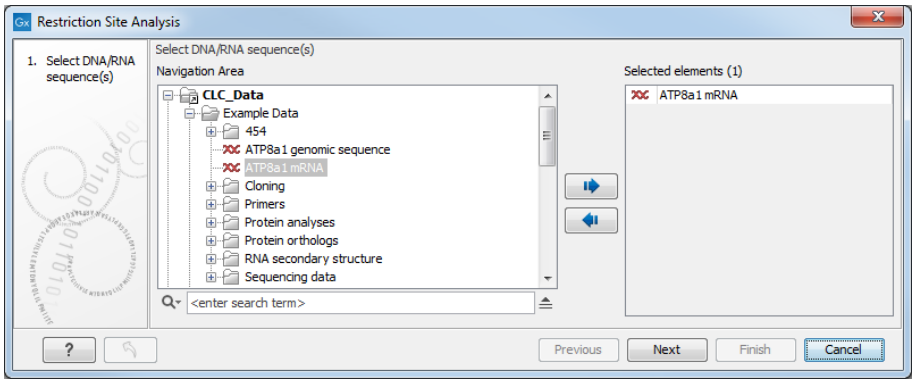

Figure 32: *Selecting a DNA or a RNA sequence.*

existing enzyme list and click the **Browse for enzyme list** button (). Select the 'Popular enzymes' in the Cloning folder under Enzyme lists (figure 33). Click **OK**.

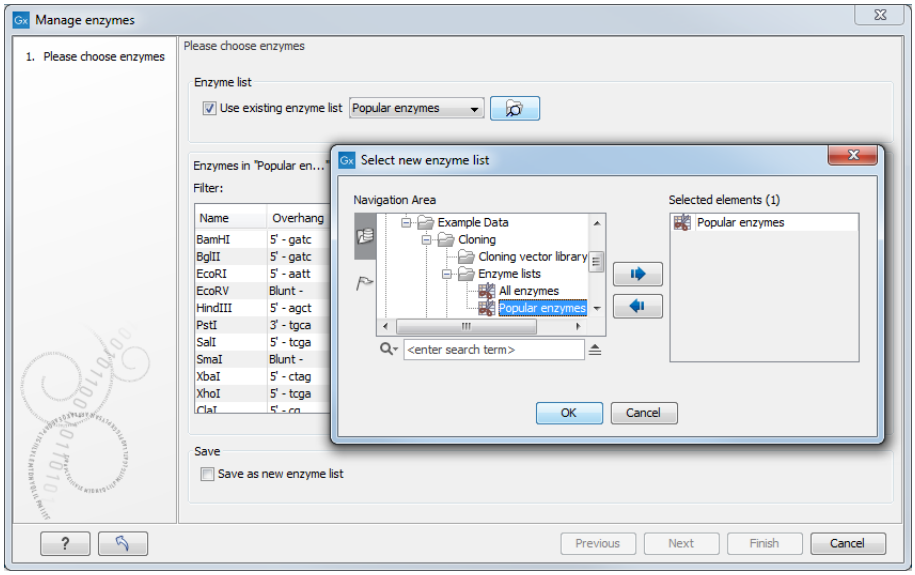

Figure 33: *Selecting enzyme list.*

4. Then write 3' into the filter below to the left. Select all the remaining enzymes and click the **Add** button (). The result should be like in figure 34. Click **Next**.

5. In this step you specify that you want to show enzymes that cut the sequence only once. This means that you should de-select the **Two restriction sites** checkbox (See figure 35). Click **Next**.

6. Now select that you want to **Add restriction sites as annotations on sequence** and **Create restriction map** (figure 36).

7. Decide whether you want to **Open** or **Save** your results. Open results will open automatically in the View Area, and can always be saved afterwards. Saved results will not be opened but saved in the location of your choice to be specified in the next wizard window. In any case, click on the button labeled **Finish** to start the restriction map analysis.

You can see the progress of your analysis by checking the "Processes" tab in the Toolbox Area (figure 37).

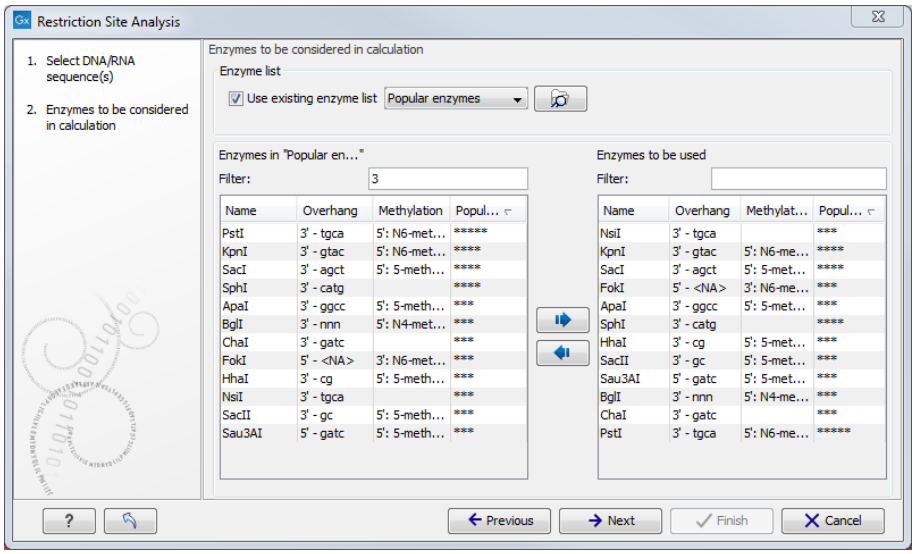

Figure 34: *Selecting enzymes.*

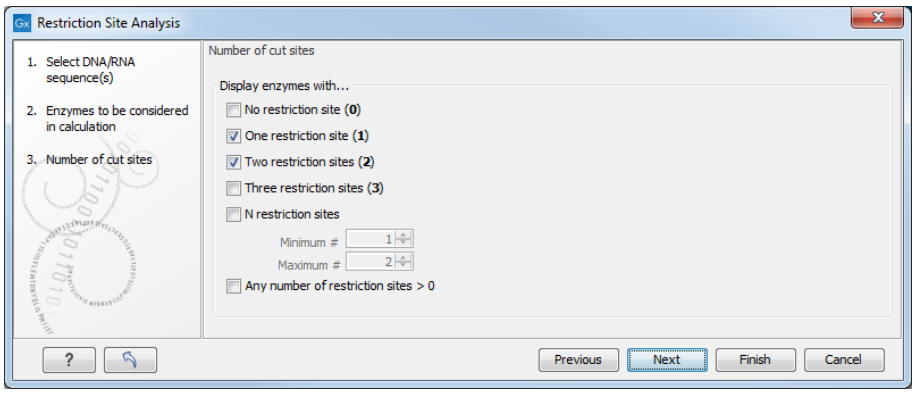

Figure 35: *Selecting output for restriction map analysis.*

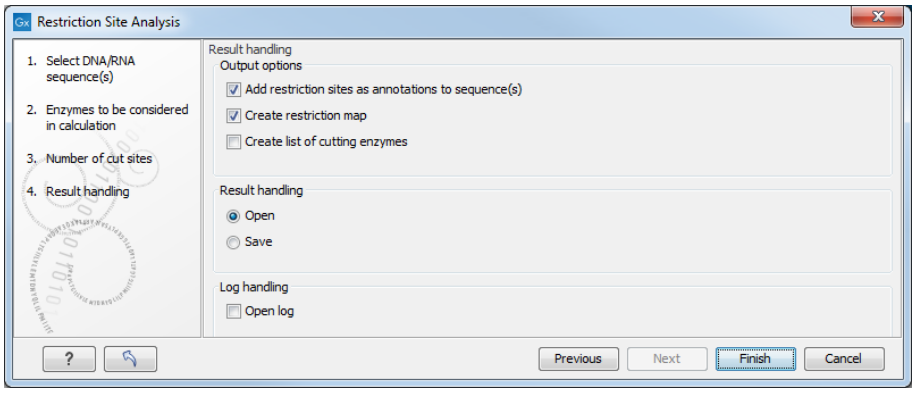

Figure 36: *Add restriction sites as annotations on sequence and create restriction map.*

If you have asked to Open your results, a new View will appear in the View Area. The name of the file is appended with an asterix to indicate that this view has not been saved yet. To save it, drag the tab to the relevant location in the Navigation Area, or simply use the usual Ctrl + S (or ⌘ + S). You can also right click on the tab and choose "Save" or use the "Save" button above the Navigation Area.

If, at the last wizard step, you had chosen to Save your results, you can use the little arrow to

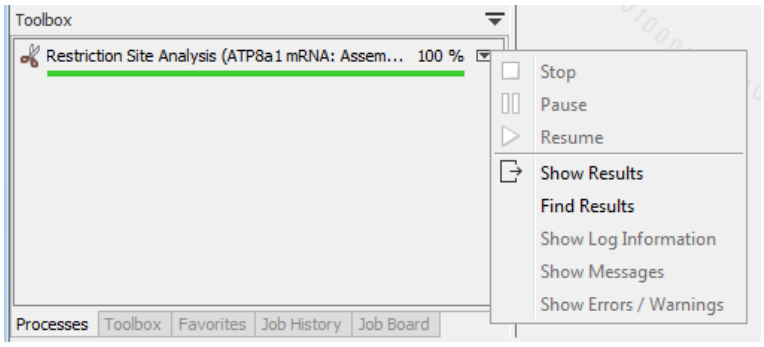

Figure 37: *Finding your results after the analysis is processed.*

the right of the analysis name in the process tab to choose the option "Show results". They will open in the View Area.

The restriction sites are shown in two views: one view is in a tabular format and the other view displays the sites as annotations on the sequence. You can see both views at the same time by right clicking on one of the result tab and choosing the option "View" followed by "Split horizontally". The result is shown in figure 38.

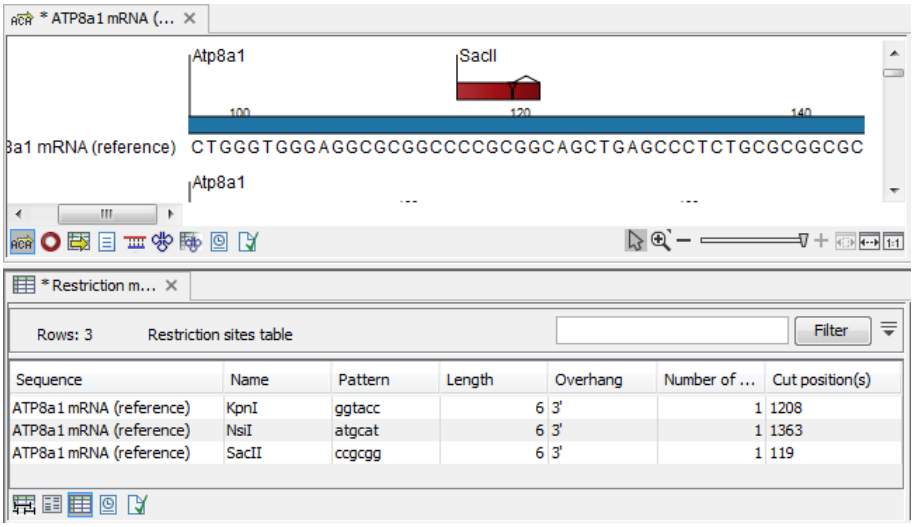

Figure 38: *The result of the restriction map analysis is displayed in a table at the bottom and as annotations on the sequence in the view at the top.*

Take a look at the button below the table view. A "Fragments" button ( ) will display a restriction map as a table of fragments produced by cutting the sequence with the enzymes. In a similar way the fragments can be shown on a virtual gel by clicking the "Gel" button ( ). Pressing Ctrl (or ⌘ on Mac) while clicking one of these buttons will open the new view in a split mode so you can see both at the same time.