MacVector 12 Getting Started Guide

MacVector, Inc.
Software for Scientists

Copyright statement

Table of Contents

Introduction

MacVector is a comprehensive Macintosh application that provides sequence editing, primer design, internet database searching, protein analysis, sequence confirmation, multiple sequence alignment, phylogenetic reconstruction, coding region analysis, and a wide variety of other functions.

This *Getting Started Guide* provides an overview of MacVector and its most commonly used functionality. Working through the Guide should take no more than an hour and will give you enough information to get you 'up and running' with MacVector.

Getting sequence information into MacVector

You can open saved sequence files in many different formats, copy sequence data from other applications and paste it into a new Sequence window, or open sequences directly from Entrez.

Importing sequence files

The **Starting Point** dialog is displayed every time you start MacVector and provides quick and easy access to sample files and the other folders and files you use most.

To open a sample sequence using the **Starting Point** dialog, choose **Sample File** as the starting point from the list on the left and select a sequence from the list of files on the right, then click **Choose**.

You can also open sequence files from the main MacVector menu. To do this, select **File | Open** from the menu, browse to the required folder, select the sequence file from the list, and click **Open**.

MacVector can open sequence files in all of the major ASCII formats (GenBank, EMBL, FastA, GCG), as well as chromatogram ("ABI") files and sequence files from other programs such as VectorNTI, LaserGene, GeneWorks and DNAStrider.

Try this. Use the **Starting Point** dialog or the main MacVector menu to open the sample file **pBR322**.

Creating new sequences

You can also create an entirely new sequence file and paste content into it. To do this choose **File** | **New** from the menu and select the required sequence type from the submenu. Then paste valid DNA or protein sequence characters into the new sequence file.

Alternatively, copy and paste the sequence data to the clipboard first, then select **File | New from Clipboard** from the menu.

Tip. Features cannot currently be pasted, although punctuation and numbers will be stripped out.

Opening sequences from Entrez

You can locate and open complete sequences from the online NCBI databases directly within MacVector.

Choose **Database | Internet Entrez Search…** from the menu and then pick the database you want to search from the **Database** drop-down menu.

Tip. The Protein Sequence Record database for proteins and the Full Nucleotide database for DNA are recommended and should appear at the top of the list of databases.

Perform a search by choosing a category in the first **All Fields** drop-down menu, typing appropriate text in the adjacent text box and clicking **Search**.

Any matches are listed in the top results panel. Double-click on a result to display the sequence in a Sequence window on the MacVector desktop.

Tip. Select multiple results and open them all in separate Sequence windows by clicking **To Desktop**.

Try this. Open all the ribosomal protein sequences from the organism "Canis" stored in the Full Nucleotide database. Hint: select **canis*** in the **Organism** field, specify **ribosomal** in the **Text Word** field and use the **And** operator in the left hand drop-down menu.

Viewing and editing sequences

Sequences in MacVector are displayed in a Sequence window that comprises a context dependent toolbar and four tabbed and linked views of the sequence:

- **Editor** view for editing the underlying sequence residues.
- **Map** view for viewing a graphical representation of the features of the sequence and for simple Restriction Enzyme based manipulation of sequences.
- **Features** view for organizing and editing features.
- **Annotations** view for organizing and editing other data like keywords and references.

Graphical maps of sequences

The **Map** view displays the features associated with a sequence in a graphical form. To highlight a particular region of a sequence in the **Map** view, a suitable feature must be created. Automatic restriction enzyme analysis adds cut site features to the sequence automatically. You can view

and edit these (and other) existing features using the **Features** view. It is also possible to modify the appearance of features in the **Map** view.
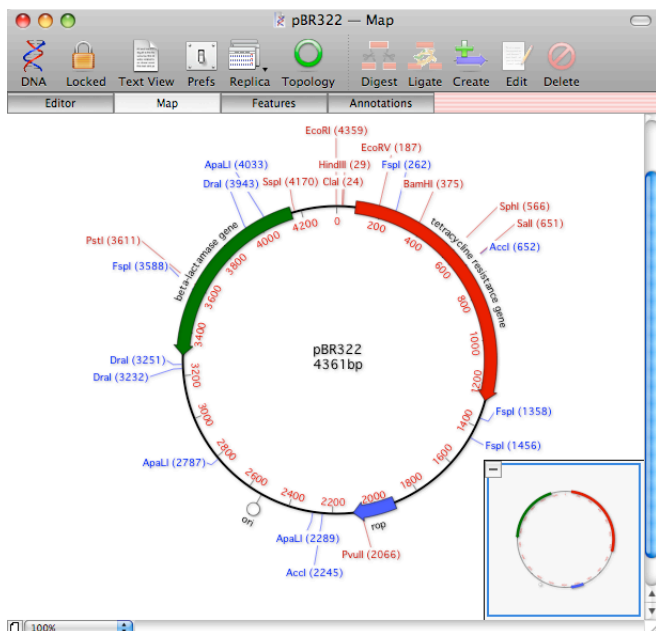
Adding sequence features

Sequence features can be added directly from a selected region in either the **Editor** or **Map** view. Alternatively, they can be created from scratch using the **Features** view.

Tip. You can also use the auto-annotation tool to automatically add features that match the features on sequences you have already annotated.

You can add a feature to a nucleic acid sequence by selecting the appropriate region of the sequence in the **Editor** view and clicking the **Create** icon on the toolbar. Add optional additional segments to the feature using the + button. Scroll through the **Feature Keyword** list and select the feature type that you want to add, or type the name of the feature type you want to add in the text box and click **OK**.

Try this. Add a **CDS** feature to pBR322 from **2400** to **3000** on the complementary strand.

To add a feature to a protein sequence, follow the same approach as for nucleic acid features. You will find that the list of feature types in the **Feature Keyword** list is now specific for proteins.

The MacVector feature keyword sets for proteins and nucleic acids match the UniProt and GenBank standard sets, respectively. Any standard feature

can be selected from the keyword list when editing the features of protein or nucleic acid sequences.

Automatic restriction enzyme analysis

MacVector scans nucleic acid sequences for restriction enzymes and displays the cut sites in the **Map** view automatically. Note how unique cutters are displayed in red whereas enzymes that cut multiple times in a sequence are displayed in blue.

To set up automatic restriction enzyme site searching select **MacVector | Preferences** from the main menu, then click the **Map View** icon on the preferences dialog. Ensure that the **Automatic RE Analysis** box is checked.

Click **Set Enzyme File** button to choose a restriction enzyme file for the automatic analysis. By default, the file **/MacVector 12/Restriction Enzymes/Common Enzymes** is used.

Click **Open** to open the selected enzyme file in the background. If the default **Only use selected enzymes** option is selected in Map View preferences, then making changes to the enzyme selection in the enzyme file causes all open sequence **Map** views to be updated dynamically. This makes it easy to identify compatible restriction sites between two sequences (e.g. a gene and a vector), for click cloning.

Click **OK** to save the settings and close the **Map View** dialog.

Try this. In the Common Enzymes file, select and deselect **EcoRI** and observe the effect this has on the **Map** view of pBR322, which has a single EcoRI site at the top.

Viewing existing features

A list of all the features that have already been added to the sequence is displayed in the **Features** view. Click on the **Features** tab in the Sequence window to display the **Features** view. Here you can:

- Add a new feature.
- Edit an existing feature.
- Delete an existing feature.

Modifying the appearance of features in maps

You can also modify the graphical appearance of features. To do this, double-click on a feature in the **Map** view to display the **Symbols** dialog.
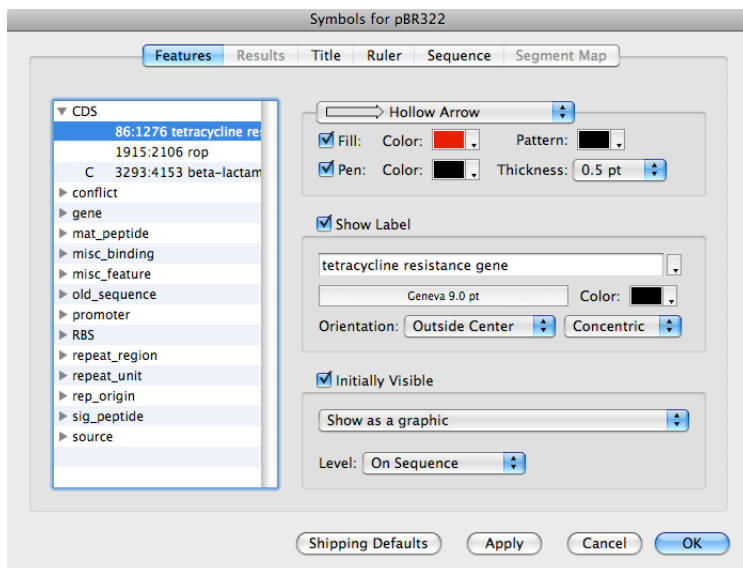
The **Appearance** panel enables you to modify the **Style**, pen and fill **Color**, fill **Pattern** and pen **Thickness** of the selected feature. While, the **Label** panel provides options for adding a text label to the feature and the **Position** panel allows you to control if and where the feature graphic appears on the sequence.

Try this. Add a label **Test** to the new CDS feature in pBR322 and change the default **Style** and **Color**.

Sub-cloning genes into vectors

MacVector includes "Click Cloning" functionality for designing cloning strategies or for documenting constructs you have already made in the lab.

Click Cloning is particularly powerful when used in conjunction with automatic restriction enzyme searches. See Automatic restriction enzyme analysis above for details.

To perform Click Cloning, open both your source and target vector sequences and check that there are suitable restriction sites surrounding the source fragment you want to clone. Select these sites and click **Digest** on the toolbar. Switch to the window containing your target vector sequence. You will see that any compatible sites now have a faint but obvious highlighting to indicate which end of the source fragment they are compatible with. Now select the enzyme site(s) in the vector sequence you want to paste the fragment into and click **Ligate** on the toolbar.

The ligation pop-up shows the structures of the fragment and vector ends, with color-coded outlines to indicate whether they are compatible.
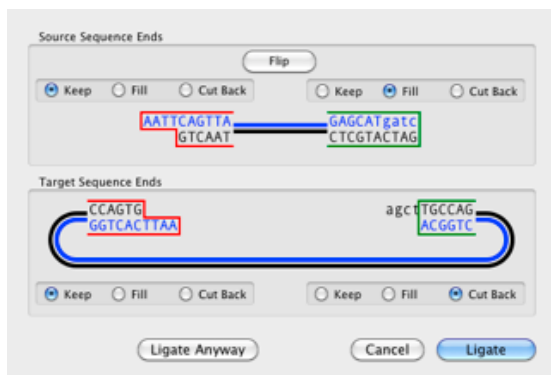
Assuming they are compatible, click **Ligate** to paste the fragment into the vector.

If the fragment and vector ends are not compatible it is possible to modify them by either filling in or cutting back the sticky ends produced by a restriction digest using the ligation pop-up.

To manipulate the sticky ends prior to ligation, either select the **Fill** option to fill in the unwanted 5' overhang or select the **Cut Back** option to cut back the unwanted 3' or 5' overhang.

Tip. Portions of the sequence which have been filled in or cut back are shown in lowercase in the ligation pop-up

Performing a detailed restriction enzyme analysis

Restriction enzyme files

MacVector contains collections of restriction enzymes, subsequences, matrix scoring tables, etc. in folders that are stored alongside the MacVector application. Restriction enzymes are divided up by manufacturer and can be found in **/MacVector 12/Restriction Enzymes**.

These files contain a list of enzymes and details of the recognition site they require for cleaving a nucleic acid sequence. You can modify these restriction enzyme files to:

- Select a subset of restriction enzyme entries and save the selection for use in restriction enzyme analysis.
- Delete existing enzyme entries from the file.
- Change the data for existing enzyme entries.
- Add new enzyme entries to the file.

Advanced restriction site searching

While the automatic restriction enzyme site searching is useful for presenting you with an overview of the common sites in a molecule, it is often useful to perform an advanced restriction site search to help design your cloning experiments.

To do this, make the required nucleic acid Sequence window active and specify whether to analyze the sequence as a circular or linear molecule by clicking the **Topology** icon in the toolbar. Choose **Analyze | Restriction Enzyme** from the main menu. Select your search criteria by choosing from the **Search Using** drop-down menu as follows:

- choose **all enzymes** to use all the enzymes in the selected enzyme file
- choose **selected enzymes** to use only the enzymes that you have selected within the enzyme file
- choose **other criteria** to display further options

If you have chosen **other criteria**, do one or more of the following:

- to restrict the search to enzymes whose recognition sites are a certain size, type the size limits into the **site size** text boxes.
- to limit the search to enzymes that leave 5' ends, 3' ends, or blunt ends when they cut the sequence, select the appropriate option from the **end structure** drop-down menu.

Finally, to restrict the search to enzymes that cut the sequence a limited number of times, type the upper limit in the **number of cuts** text box. Then select **OK** to perform the analysis.

When this is complete, a results filtering dialog is displayed:

Select from the following filters for the results, as required:

- by the number of cuts an enzyme makes
- by the end structure of cut fragments
- by restriction site size

- by the residue range in which the cut site occurs

And select **Display Options** to generate separate windows displaying the results. Click **OK** to display the results windows.

Try this. Create a Restriction Map showing all single cutters in the **Test** CDS feature in pBR322.

Translating DNA into a protein

MacVector provides flexible tools for transcribing a DNA sequence and translating the resulting mRNA into an amino acid sequence.

You can transcribe an entire sequence of DNA by simply clicking the sequence type indicator icon on the toolbar. The indicator icon will change from **DNA** to **RNA**. Click the icon again to reverse the transcription. Or, to transcribe only selected parts of the sequence (e.g. exons), use **Analyze | Generate Transcript**.

Alternatively, use **Analyze | Translation** to transcribe and translate DNA in a single operation. If you do not want to translate the entire sequence, select the parts of the sequence for translation by:

- Typing the required ranges in the **Segment(s) to translate** text box, or
- Selecting the required features from the feature selection drop-down menu to the right of the box

Check the **Create new protein** box to create a new protein sequence window for the resulting protein and select **OK** to perform the translation and display the results.

Try this. Translate either of the major CDS features in pBR322 into a protein. Notice how MacVector takes the strand selected into account.

Tip. Analyze | Translation can also be used to translate mRNA created by a transcription operation.

Aligning sequences

MacVector provides tools for many different types of sequence alignment including:

- Multiple sequence alignment (see below)
- Align to reference (see below)
- Pair-wise dot-plot matrix
- BLAST
- Align to folder

These and many other sequence analysis tools can be found under **Analyze** or **Database** in the main MacVector menu. See the *MacVector User Guide* for comprehensive details about how to use all of these tools.

Multiple Sequence Alignment

Multiple sequence alignments (MSAs) are usually created in one of three ways:

- By opening multiple single sequences on the desktop, choosing ClustalW and picking the open sequences you want to include in the alignment.
- By creating a new, empty MSA document, then adding sequences from files or copying and pasting sequences from other sources.
- By selecting **File | Open** from the menu, choosing the sequence files you want (these can be single sequences, multi-sequence format files like FastA, or a combination of both) and opening them as a **Multiple Alignment**.

The ClustalW algorithm (Thompson et al, 1994) is used to automatically align several nucleotide or amino acid sequences simultaneously. MacVector provides extra graphical features for the alignments:

- Both multiple and pairwise alignments can be displayed in standard MacVector text tabs.
- Multiple alignments can be displayed as high quality PDF images that can be exported to other applications.
- A summary of the similarity scores between each pair of sequences can be displayed in tabular form.
- A consensus sequence can be displayed and saved.
- A guide tree of the sequence comparisons can be displayed, showing the relationships between sequences.

To perform a ClustalW alignment using the default settings choose **Analyze | ClustalW Alignment** from the menu or click **Align** on the MSA window toolbar, select the sequences you want to align in the **Sequences to Align** panel and click **OK**.
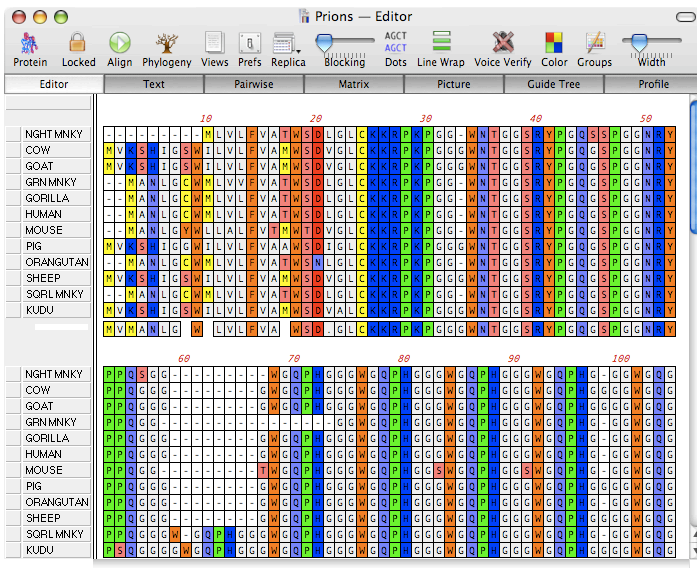
When the calculation is complete, the Alignment Views display dialog box is displayed. This allows you to choose one or more types of results to display from the following output options:

- Multiple alignment text display
- Pairwise alignment text display
- Pairwise matrix display
- Multiple alignment picture display
- Consensus sequence display
- Guide tree (when four or more sequences are aligned)

Align to Reference

The **Align to Reference** function lets you align one or more sample sequences against a reference sequence. It has many similarities to sequence assembly, except that it requires and takes advantage of the use of a known reference sequence as a scaffold.

You can use this functionality to help you solve a number of typical laboratory problems:

- Confirming the sequence of a cloned fragment
- Sequencing across the ends of a cloned fragment to confirm the junction sequence
- Screening clones from a site-specific mutagenesis experiment to identify successful mutations
- Screening related clones for single nucleotide polymorphisms

The main limitation of this approach is that you must have a reference sequence to act as a scaffold against which the sample sequences can be aligned. You cannot, therefore, use this function to assemble trace files from *de novo* sequencing projects. For these types of tasks, MacVector Assembler is a better tool. However, for the less complex re-sequencing assembly tasks commonly encountered during a molecular biologist's day, the **Sequence Confirmation** function is ideal.

To align trace files against a template sequence, open the reference sequence and choose **Analyze | Align to Reference** to create a new Assembly window. Click **Add Seqs** on the toolbar in the Assembly window and select the sequence files you wish to assemble against your template, then click **Open**.

Tip. You can hold down the **<shift>** key to select multiple sequences to import.

MacVector imports the selected files into the assembly project and the Assembler window updates to display the sequences in the order in which they were imported. If trace data is present, this is also displayed in a lower trace panel.
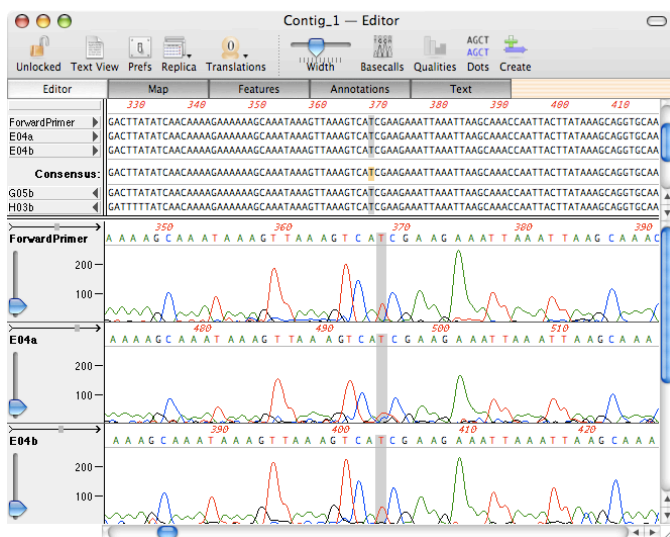
Click **Align** on the Assembly window toolbar, ensure that **Sequence Confirmation** is selected from the **Alignment Type** drop-down menu and click **OK** to perform the alignment using the default parameters.

The calculation may take some time if your sequence is long and you have many trace files in the alignment.

The **cDNA Alignment** algorithm lets you align one or more cDNA clones against a reference genomic sequence. This can be used for identifying

14

where exon/intron boundaries lie in your cDNA/mRNA sequences. It differs from the **Sequence Confirmation** algorithm only in that it allows for extremely long gaps between aligned segments.

To align cDNAs against a genomic sequence, use exactly the same procedure as described above but select **cDNA Alignment** from the **Alignment Type** drop-down menu.

Try this. Locate and open the file **/Tutorial Files/Align to Reference/Presenilin PSENI_Genomic**. Align a set of exons and the mRNA sequence against this.

Designing PCR Primers

MacVector includes an automated primer design module based on Primer3. This easy-to use tool allows you to create a new primer in as few as three mouse clicks.

To design a primer to amplify a particular feature simply select the feature, choose **Analyze | Primers | Primer Design (Primer3)** from the menu and click **OK**.

The default primer design settings are used to design primers to amplify the feature. Results are displayed in the form of a ranked list and graphical map showing the top five primer pairs, using the optimum values for the primer Tm and %GC.

Try this. Select the **Test** CDS feature in pBR322 and design a set of primers to amplify it.

Two other primer design methods are also available. **Region to Scan** enables you to specify a region within which the amplified product should lie. Whereas, **Flanking Regions** allows you to design primers by specifying two flanking regions and selecting left and right primers from these.

Many real time PCR techniques use a third oligonucleotide that will anneal inside the desired product making it possible to track the progress of the PCR amplification using one of the many proprietary technologies available. The primer design module in MacVector also enables you to design these (see the *MacVector User Guide* for further information).