Jalview 2.2: A manual and introductory tutorial

David Martin

August 14, 2008

Contents

1 Basics **5**

1.1 Introduction 5

1.1.1 Jalview 5

1.1.2 About this tutorial 6

1.2 Obtaining and starting Jalview 6

1.2.1 Getting Help 8

1.3 Navigation 9

1.3.1 Navigation in Normal mode 9

1.3.2 Navigation in Cursor mode 10

1.4 Loading your own sequences 11

1.4.1 Drag and Drop 11

1.4.2 From a File 11

1.4.3 From a URL 12

1.4.4 Cut and Paste 12

1.4.5 From a public database 13

1.5 Writing sequence alignments 14

1.5.1 Saving the alignment 14

1.5.2 Jalview Projects 15
1.6 Selecting and editing sequences 16
1.6.1 Selecting parts of an alignment 16
1.6.2 Creating groups 18
1.6.3 Reordering the alignment 19
1.6.4 Introducing and removing gaps 20
1.6.5 Hiding regions 22
1.7 Colouring sequences 23
1.7.1 Colouring the whole alignment 24
1.7.2 Colouring a group or selection 24
1.7.3 Colour schemes 24
1.8 Screen layout and graphics output 29
1.8.1 Screen layout 29
1.8.2 Graphical output 31

2 Analysis **33**

2.1 Analysis of alignments 33
2.1.1 PCA 33
2.1.2 Trees 35
2.2 Web Services 37
2.2.1 Realignment 38
2.2.2 Secondary Structure Prediction 39
2.3 Features and Annotation 40
2.3.1 Creating sequence features 40

2.3.2 Customising feature display 41
2.3.3 Creating user defined annotation 42
2.3.4 Importing features from databases 43
2.4 Working with structures 45
2.4.1 Automatic association of PDB structures with sequences 45
2.4.2 Viewing Protein Structures 45

3 Advanced Jalview **48**

3.1 Customising Jalview 49
3.1.1 Setting preferences 49
3.2 The Jalview Interface 49
3.2.1 Multiple views 49
3.2.2 Keyboard Editing Mode 49
3.3 Regions 49
3.3.1 Locked Editing 49
3.3.2 Alignments including hidden regions 49
3.3.3 Secondary Structure predictions 49
3.4 Features and Annotations 49
3.4.1 Annotation display 49
3.4.2 Annotation files 49
3.4.3 Feature files 49
3.4.4 Moving sequence associated annotation 49
3.4.5 Propagating features 49
3.5 Structures 49

3.5.1 Working with Modeller files 49

3.5.2 Using local PDB files 49

3.6 Pairwise alignments 49

Chapter 1

Basics

1.1 Introduction

1.1.1 Jalview

Jalview is a sequence multiple alignment viewer, editor and analysis tool. Jalview is designed to be platform independent (running on Mac, MS Windows, Linux and any other platform that supports Java), capable of editing and analysing large alignments (thousands of sequences) with minimal degradation in performance, and able to show multiple integrated views of the alignment and other data. Jalview can read and write many common sequence formats including FASTA, Clustal, MSF(GCG) and PIR.

Jalview 2.2 provides improved access to databases and analyses via web services, allowing the retrieval and display of third party data such as features and structures in association with the sequences. Jalview is typically run as a stand alone application or as an applet embedded in a web page, allowing customised integration with alignment databases such as pFam[1].

Jalview History

Jalview was initially developed in 1996 by Michele Clamp, James Cuff, Steve Searle and Geoff Barton at the University of Oxford and then the European Bioinformatics Institute. Development of Jalview slowed considerably before resuming with funding from the BBSRC[2] in 2004 with Andrew Waterhouse and Jim Procter bringing current developments in bioinformatics into Jalview 2 through

[1] A demonstration version of Jalview (Jalview Micro Edition) also runs on a mobile phone but the functionality is limited to sequence colouring.

[2] Biotechnology and Biological Sciences Research Council grant *"VAMSAS: Visualisation and Analysis of Molecules, Sequence Alignments and Structures"*, a joint project to enable interoperability between Jalview, TOPALi and AstexViewer.

the latest web and Java technology. Jalview continues to be one of the worlds most popular[3] sequence alignment and analysis tools.

Citing Jalview

If you use Jalview in your work you should cite "*The Jalview Java alignment editor*" Michele Clamp, James Cuff, Stephen M. Searle and Geoffrey J. Barton (2004) *Bioinformatics* **20** 426-427.

1.1.2 About this tutorial

This tutorial is written in a manual format with short exercises where appropriate, typically at the end of each section. The first chapter concerns the basic operation of Jalview and should be sufficient for those who just want to open an alignment, perform basic editing and colouring, and produce publication and presentation quality graphical output.

The second chapter covers analysis in Jalview, sequence alignment and visualisation of external features and structural data.

The third chapter covers the detail of Jalview and is aimed at the user who is already familiar with Jalview operation but wants to get more out of their Jalview experience.

1.2 Obtaining and starting Jalview

Jalview can be run in three ways; as an application from the web via Java Web Start, as an application loaded onto your hard drive, or as an applet embedded in a web page. This tutorial is only concerned with running the application via Web Start, though much of the information will be useful whichever way you run Jalview.

The Jalview web site is http://www.jalview.org/ and Jalview can be launched by navigating to the Download page (via the menu on the left hand side) and clicking the 'Start with Java Webstart' button. (figure 1.1). This will always launch the latest public release of Jalview.

The application will start automatically though you may be prompted to accept a security certificate signed by the Barton Group. You can trust us so click trust or accept as appropriate. The splash screen (figure 1.2) gives information about the latest version and the paper to cite in your publications. This information is also available on the Jalview web site and from the *Help* ⇒ *About* menu option.

When Jalview starts it will automatically load an example alignment from the Jalview site. Later in the tutorial we will discover how to alter this behaviour. This alignment looks like figure 1.3.

[3] and in the authors opinion, the best.

Figure 1.1: **Download page on the Jalview web site**

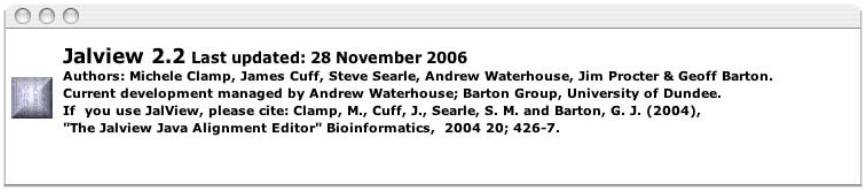


Figure 1.2: **Jalview splash screen**

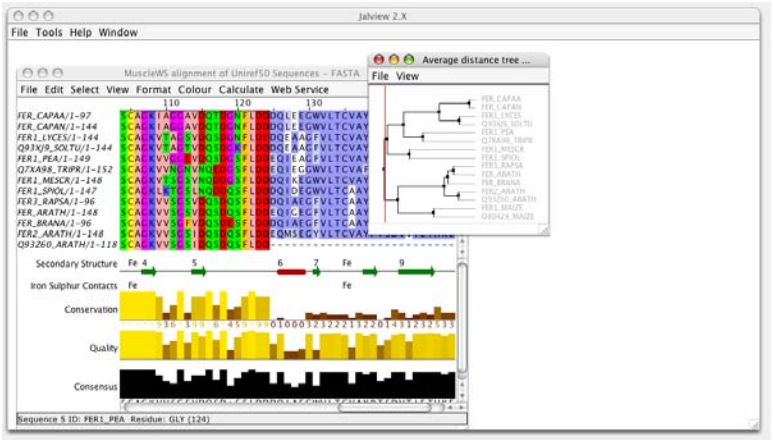


Figure 1.3: **Default startup for Jalview**

Exercise 1: Starting Jalview

1.a. Point your web browser at the Jalview web site and start Jalview by clicking on the 'Start with Java WebStart' button.

1.2.1 Getting Help

Built in documentation

Jalview has comprehensive on-line help documentation. Select *Help* ⇒ *Documentation* from the main window menu and a new window will open (Figure 1.4). The appropriate topic can then be selected from the navigation panel on the left hand side. To search for a specific topic, click the 'search' tab and enter keywords in the box which appears.

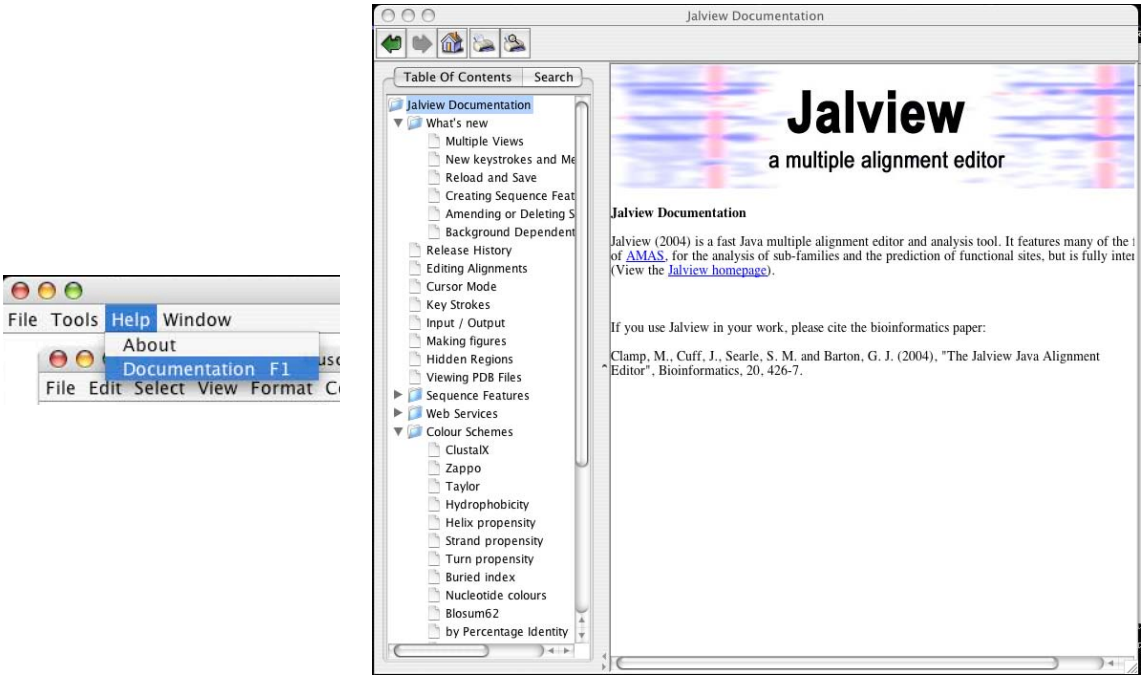


Figure 1.4: **Accessing the built in Jalview documentation**

Email lists

The Jalview developers can be contacted with news of bugs, feature requests and so on via the email address `help@jalview.org`. A mailing list for information related to Jalview announcements and new releases can also be found on the Jalview web site.

1.3 Navigation

The major features of the Jalview application are illustrated in Figure 1.5. Each area of the alignment window has a separate context menu accessed by clicking the right mouse button.

Jalview has two navigation and editing modes: normal mode, where editing and navigation is performed using the mouse, and cursor mode where editing and navigation are performed using the keyboard. The F2 key is used to switch between these two modes.

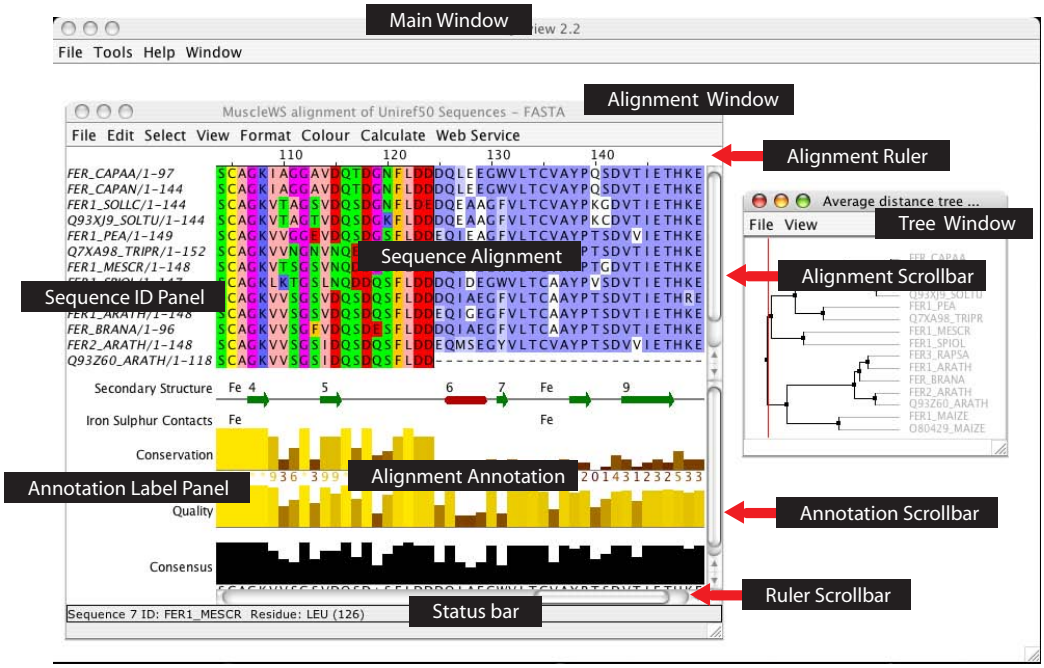


Figure 1.5: **The anatomy of Jalview.** The major features of the Jalview application are labelled.

1.3.1 Navigation in Normal mode

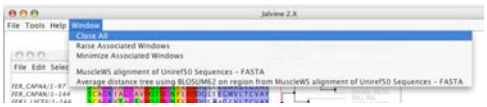
Jalview always starts up in Normal mode. With anything more than a few residues or sequences, alignments become difficult to visualise on the screen. Jalview shows a window on the alignment, limited by the size of the window and the font used to render it. You can move about the alignment by clicking and dragging the ruler scroll bar to move horizontally, or by clicking and dragging the alignment scroll bar to the right of the alignment to move vertically. If all the rows or columns in the alignment are displayed, the scroll bars will not be visible.

It can help, especially when examining a large alignment, to have an overview of the whole alignment. Select *View* ⇒ *Overview Window* from the window menu (Figure 1.6).

Figure 1.6: **Alignment Overview Window**

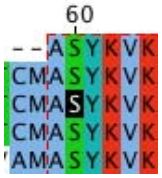
The red box in the overview window shows the current view in the alignment window. You can navigate around the alignment by dragging the red box.

We can close these windows by clicking on the icons indicated by arrows, or by selecting *Window* ⇒ *Close All* from the main menu.



1.3.2 Navigation in Cursor mode

Cursor mode provides a fast means for an experienced use to navigate to the part of the alignment in which they are interested. On pressing F2 to enter cursor mode the position of the cursor is indicated by a black background and white text. The cursor can be placed using the mouse or moved by pressing the arrow keys (↑, ↓, ←, →).



Rapid movement to specific positions is accomplished as listed below:

- **Jump to Sequence** *n***:** Type a number *n* then press [S] to move to sequence (row) *n*
- **Jump to Column** *n***:** Type a number *n* then press [C] to move to column *n* in the alignment.
- **Jump to Residue** *n***:** Type a number *n* then press [P] to move to residue number *n* in the current sequence.
- **Jump to column** *m* **row** *n***:** Type the column number *m*, a comma, the row number *n* and press [RETURN].

Exercise 2: Navigation

2.a. Scroll around the alignment using the alignment (vertical) and ruler (horizontal) scroll bars.

2.b. Find and open the Overview Window. Move around the alignment by clicking and dragging the red box in the overview window.

2.c. Look at the status bar as you move the mouse over the alignment. It should indicate information about the sequence and residue under the cursor.

2.d. Press [F2] to enter Cursor mode. Use the arrow keys to move the cursor around the alignment. Move to sequence 7 by pressing *7 S*. Move to column 18 by pressing *1 8 C*. Move to residue 18 by pressing *1 8 P*. Note that these can be two different positions if gaps are inserted into the sequence. Move to sequence 5, column 13 by typing *1 3 , 5 [RETURN]*.

1.4 Loading your own sequences

Jalview provides many ways to load your own sequences.

1.4.1 Drag and Drop

In some operating systems (Mac OS X, Windows XP) you can just drag a file icon from a file browser window and drop it on an open Jalview application window (Figure 1.7). The file will then be opened as a new alignment window. If you drop an alignment file onto an open alignment window it will be appended to that alignment.

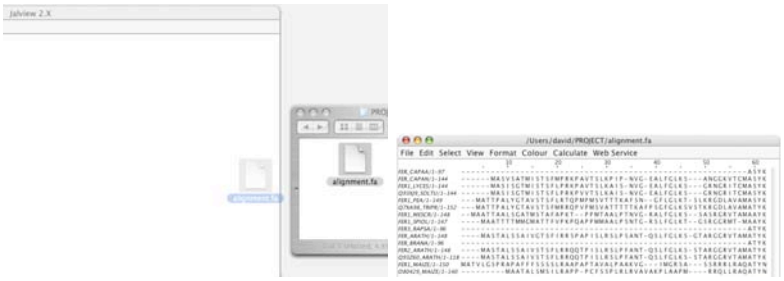


Figure 1.7: **An alignment can be opened by dragging the file onto the Jalview window.**

1.4.2 From a File

Jalview can read sequence alignments from a sequence alignment file. This is a text file, not a word processor document. For entering sequences from a wordprocessor document see Cut and Paste (section 1.4.4) below. Select *File* ⇒ *Input Alignment* ⇒ *From File* from the main menu (Figure

1.8). You will then get a file selection window where you can choose the file to open. Remember to select the appropriate file type. Jalview can automatically identify some sequence file formats.

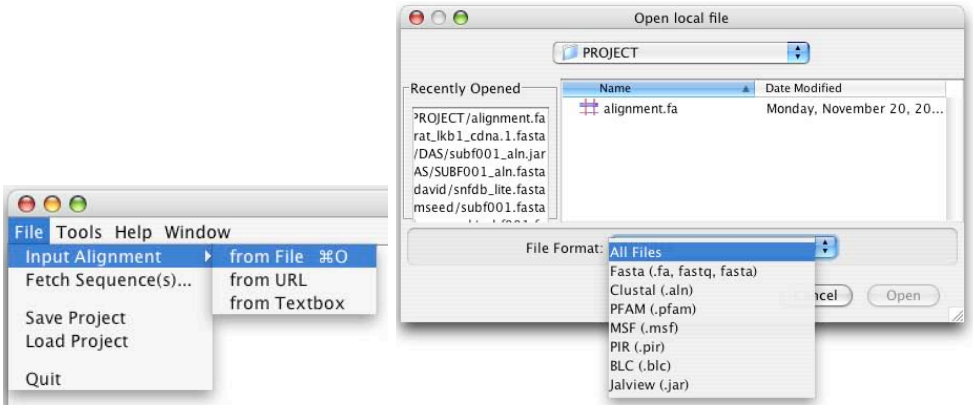


Figure 1.8: **Opening an alignment from a file saved on disk.**

1.4.3 From a URL

Jalview can read sequence alignments directly from a URL. Please note that the files must be in a sequence alignment format - a pretty HTML alignment or graphics file cannot be read by Jalview. Select *File* ⇒ *Input Alignment* ⇒ *From URL* from the main menu and a window will appear asking you to enter the URL (figure 1.9). Jalview will attempt to automatically discover the file format.

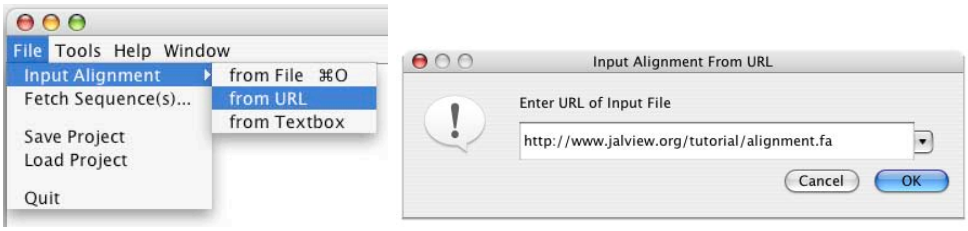


Figure 1.9: **Opening an alignment from a URL**

1.4.4 Cut and Paste

Documents such as those produced by Microsoft Word cannot be readily understood by Jalview. The way to read sequences from these documents is to select the data from the document and copy it to the clipboard. Select *File* ⇒ *Input Alignment* ⇒ *From Textbox* from the main menu and a textbox window will appear (Figure 1.10). Paste the sequences you have copied to the clipboard and, presuming that they are in the right format, Jalview will happily read them.

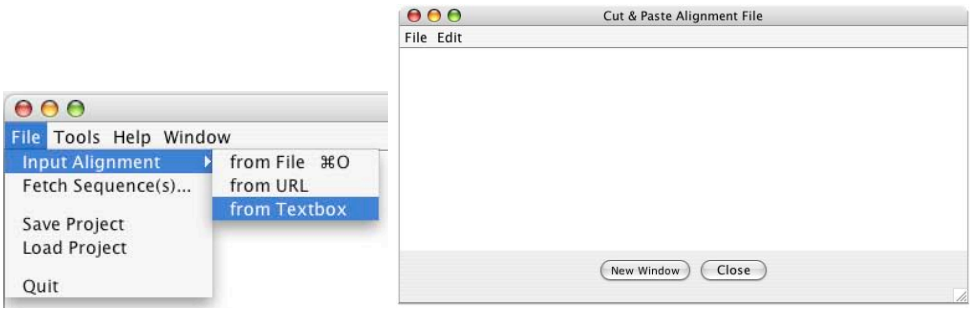


Figure 1.10: **Opening an alignment from pasted text**

1.4.5 From a public database

Jalview can retrieve sequences and sequence alignments from public databases such as Uniprot, Pfam and the PDB. This facility avoids having to manually locate, save and load the sequences. Select *File* ⇒ *Fetch Sequence(s)* ... from the main menu and a window will appear (Figure 1.11). Select the appropriate database, enter a sequence ID/accession number, or several separated by a semicolon and Jalview will retrieve it/them from the public databases housed at the European Bioinformatics Institute.

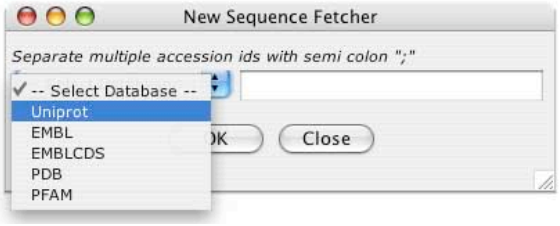


Figure 1.11: **Retrieving sequences from a public database**

Exercise 3: Loading sequences

3.a. Start Jalview then close all windows by selecting *Window* ⇒ *Close All* from the main menu

3.b. Select *File* ⇒ *Input Alignment* ⇒ *From URL* from the main menu and enter http://www.jalview.org/tutorial/alignment.fa in the box. Click *OK* and the alignment should load.

3.c. Close all windows using the *Window* ⇒ *Close All* main menu option. Point your web browser to http://www.jalview.org/tutorial/alignment.fa and save the file to your desktop. Open this file in Jalview by selecting *File* ⇒ *Input Alignment* ⇒ *From File* from the main menu and browsing to the appropriate location. Click OK and load the alignment

3.d. Drag the alignment.fa file from the desktop onto the Jalview window. The alignment should open. Try dragging onto an empty Jalview and onto an existing alignment and observe the results.

3.e. Select *File* ⇒ *Fetch Sequence(s)...* from the main menu. Select the sl PFAM database and enter the accession number PF03460. Click OK. An alignment of about 107 sequences should load.

3.f. Open the URL http://www.jalview.org/tutorial/alignment.fa in a web browser. Select and copy the entire text to the clipboard (usually via the browser's *Edit* ⇒ *Copy* menu option). Ensure Jalview is running and select *File* ⇒ *Input Alignment* ⇒ *From Textbox* . Paste the clipboard into the large window using the *Edit* ⇒ *Paste* text box menu option. Click *Close* and the alignment will be loaded.

1.5 Writing sequence alignments

1.5.1 Saving the alignment

Jalview allows the current sequence alignments to be saved to file so they can be restored at a later date, passed to colleagues or analysed in other programs. From the alignment window menu select *File* ⇒ *Save As* and a dialog box will appear (Figure 1.12). You can navigate to an appropriate directory in which to save the alignment. Jalview offers several different formats in which an alignment can be saved. The jalview format is the only one which will preserve the colours, groupings and similar information in the alignment. The other formats produce text files containing just the sequences with no visualisation information. Unfortunately only Jalview can read Jalview files. Using the *File* ⇒ *Output To Textbox* menu option allows the alignment to be copied and pasted into other documents or web servers.

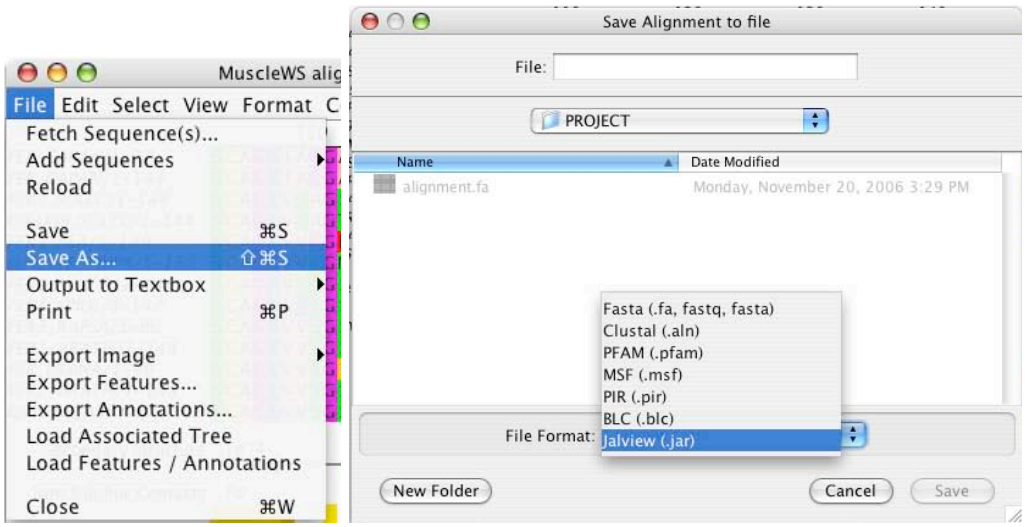
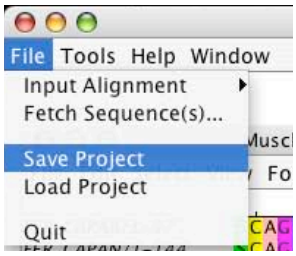


Figure 1.12: **Saving alignments in Jalview to disk**

1.5.2 Jalview Projects

If you wish to save the complete Jalview session rather than just one alignment (e.g. because you have calculated trees or multiple different alignments) then your work should be saved as a Jalview Project file. From the main menu select *File* ⇒ *Save Project* and a file save dialog box will appear. Loading a project will restore Jalview to exactly the view at which the file was saved, complete with all alignments, trees etc. rendered appropriately.



Exercise 4: Saving Alignments

4.a. Start Jalview, close all windows and load the ferrodoxin alignment from pFam (accession number PF03460 (see Exercise 3).

4.b. Select *File* ⇒ *Save As* from the alignment window menu. Choose a location into which to save the alignment and select a format. All formats except *Jalview* can be viewed in a normal text editor (e.g. Notepad) or in a web browser. Enter a file name and click *Save*. Check this file by browsing to it with your web browser or by closing all windows and opening it with Jalview.

4.c. Repeat the previous step trying different file formats.

4.d. Select *File* ⇒ *Output to Textbox* ⇒ *FASTA*. You can select and copy this alignment to the clipboard using the textbox menu options *Edit* ⇒ *Select All* followed by *Edit* ⇒ *Copy*. The alignment can then be pasted into any application of choice, e.g. a word processor or web form.

4.e. Ensure at least one alignment window is shown in Jalview. Open the overview window and scroll to any part of the alignment. Select *File* ⇒ *Save Project* from the main menu and save in a suitable place. Close all windows and then load the project via the *File* ⇒ *Save Project* menu option. Note how all the windows and positions are exactly as they were when they were saved.

1.6 Selecting and editing sequences

Jalview makes extensive use of selections which are arbitrary regions in an alignment. This section illustrates how to make and use selections and groups.

1.6.1 Selecting parts of an alignment

Selections can be of arbitrary regions in an alignment, one or more complete columns, or one or more complete sequences.

A selected region can be copied and pasted as a new alignment using the *Edit* ⇒ *Copy* and *Edit* ⇒ *Paste* ⇒ *As New Alignment* alignment window menu options.

To clear (unselect) the selection press the [ESC] (escape) key.

Selecting arbitrary regions

To select part of an alignment, place the mouse at the top left corner of the region you wish to select. Press the mouse button and drag the mouse to the bottom right corner of the chosen region before releasing the mouse button. A dashed red box appears around the selected region (Figure 1.13).

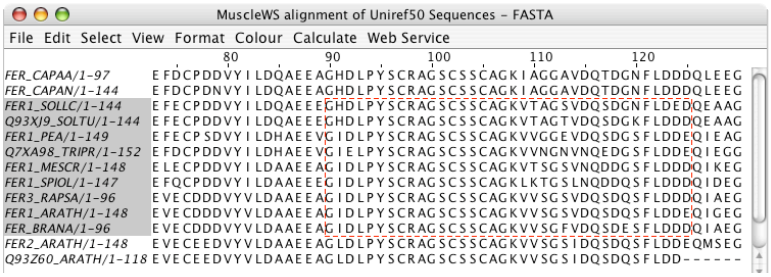


Figure 1.13: **Selecting a region in an alignment**

Selecting columns

To select the same residues in all sequences, click and drag along the alignment ruler. This selects the entire height of the alignment. Ranges of positions can also be selected by clicking on the first position then holding down the [SHIFT] key whilst clicking the other end of the selection. Discontinuous regions can be selected by holding down [CTRL] and clicking on positions to add to the selection. Selected columns are indicated by red highlighting in the ruler bar (Figure 1.14).

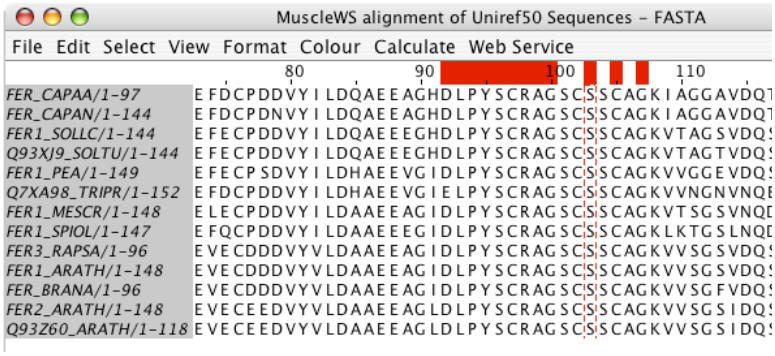


Figure 1.14: **Selecting multiple columns in an alignment**

Selecting sequences

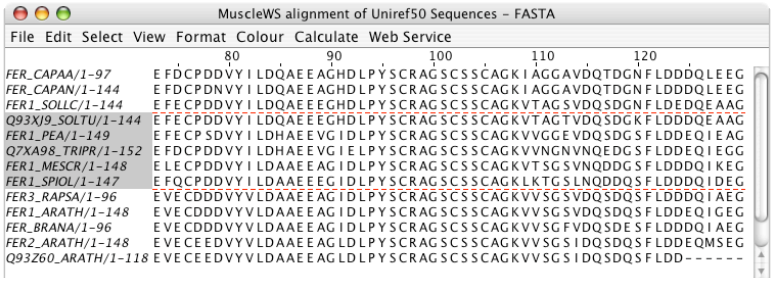


Figure 1.15: **Selecting multiple sequences in an alignment**

To select multiple complete sequences, click and drag the mouse along the sequence ID panel. Multiple sequences will be selected. The same technique as used for columns above can be used with SHIFT-Click and CTRL-Click to select discontinuous ranges of sequences (Figure 1.15).

Making selections in Cursor mode

To define a selection in cursor mode, navigate to the top left corner of the proposed selection. Pressing the [Q] key marks this as the corner. A red outline appears around the cursor (Figure 1.16)

Navigate to the bottom right corner of the proposed selection and press the [M] key. This marks the bottom right corner of the selection. The selection can then be treated in the same way as if it had been created in normal mode.

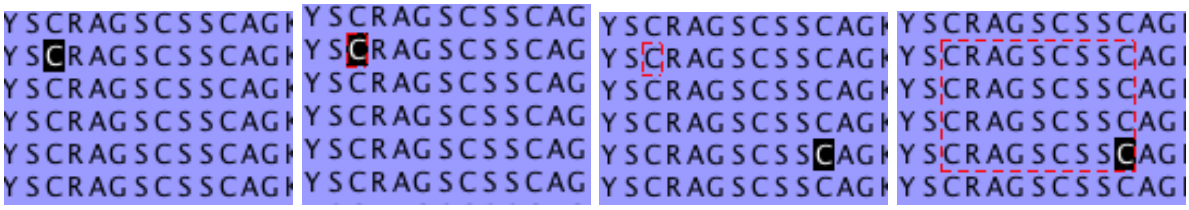


Figure 1.16: **Making a selection in cursor mode.**Navigate to the top left corner (left), press [Q] (left center), navigate to the bottom right corner (right center) and press [M] (right)

1.6.2 Creating groups

Selections are lost as soon as a different region is selected. Groups can be created which are labeled regions of the alignment. To create a group, first select the region which is to comprise the group. Then click the right mouse button on the selection to bring up a context menu. Select *Selection* ⇒ *Group* ⇒ *Group* then enter a name for the group in the dialogue box which appears.

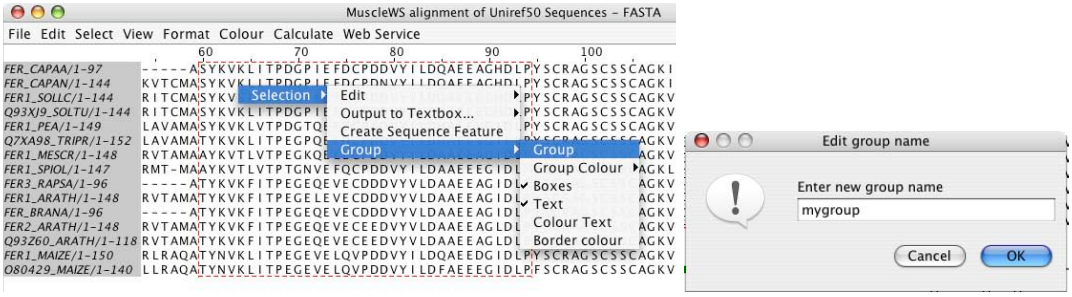


Figure 1.17: **Creating a new group from a selection**

By default the selection will have a box drawn around it. The appearance of the group can be changed (see section 1.7 below). This group will stay defined even when the selection is removed.

Exercise 5: Making selections and groups

5.a. Close all windows in Jalview and load the ferrodoxin alignment (PFAM ID PF03460). Choose a residue and place the mouse cursor on it. Click and drag the mouse cursor to create a selection. As you drag, a red box will 'rubber band' out to show the extent of the selection. Release the mouse button and a red box should border the selected region. Now press [ESC] to clear the selection.

5.b. Select one sequence by clicking on the id panel. Note that the sequence ID takes on a highlighted background and a red box appears around the selected sequence. Now hold down [SHIFT] and click another sequence ID a few positions above or below. Note how the selection expands to include all the sequences between the two positions on which you clicked. Now hold down [CTRL] and click on several sequences ID's both selected and unselected. Note how unselected IDs are individually added to the selection and previously selected IDs are individually deselected.

5.c. Repeat the step above but selecting columns by clicking on the ruler bar instead of selecting rows by clicking on the sequence ID.

5.d. Press [F2] to enter Cursor mode. Navigate to column 59, row 1 by pressing *5 9 , 1 [RETURN]*. Press *Q* to mark this position. Now navigate to column 65, row 8 by pressing *6 5 , 8 [RETURN]*. Press *M* to complete the selection.

1.6.3 Reordering the alignment

Sequence reordering is simple. Highlight the sequences to move then press the up or down arrow keys as appropriate (Figure 1.18). If you wish to move a sequence up past several other sequences it is often quicker to select the group past which you want to move it and then move the group rather than the individual sequence.

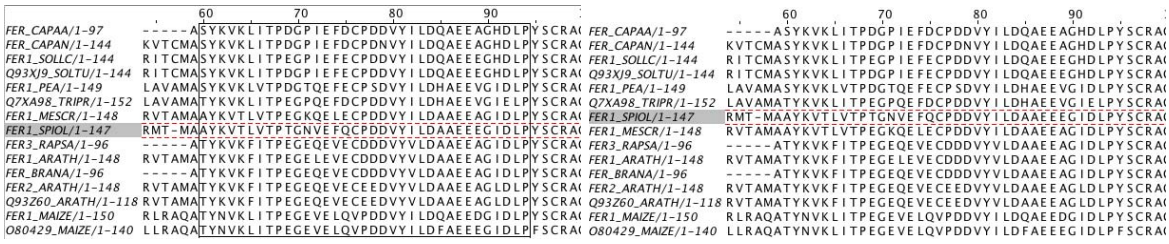


Figure 1.18: **Reordering the alignment.** The selected sequence moves up one position on pressing the ↑ key

Exercise 6: Reordering the alignment

6.a. Open an alignment (e.g.the PFAM domain PF03460). Select one sequence. Using the up and down arrow keys, alter it's position in the alignment.

1.6.4 Introducing and removing gaps

Jalview provides interactive editing, allowing sequences to be dragged to create gaps.

Introducing gaps in a single sequence

To introduce a gap, place the cursor on the residue to the immediate right of where the gap should appear. Hold down the SHIFT key and the left mouse button, then drag the sequence to the right till the required number of gaps has been inserted.

One common error is to forget to hold down [SHIFT]. This results in a selection which is one sequence high and one residue long. Gaps cannot be inserted in such a selection. The selection can be cleared and editing enabled by pressing the [SHIFT] key.

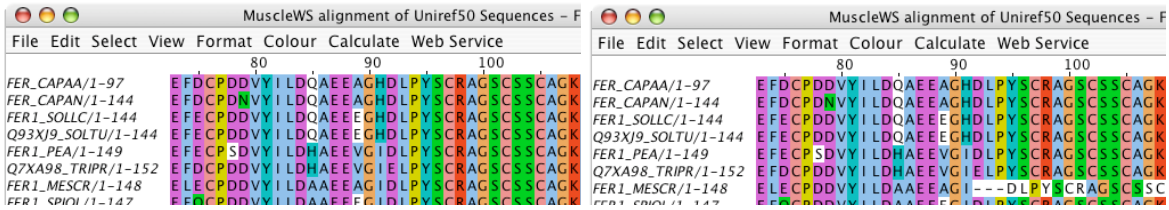


Figure 1.19: **Introducing gaps.** Gaps are introduced as the selected sequence is dragged to the right.

Introducing gaps in all sequences of a group

To insert gaps in all sequences in a selection or group, place the mouse cursor on any residue in the selection or group to the immediate right of the position in which a gap should appear. Hold down the CTRL key and the left mouse button, then drag the sequences to the right until the required number of gaps has appeared.

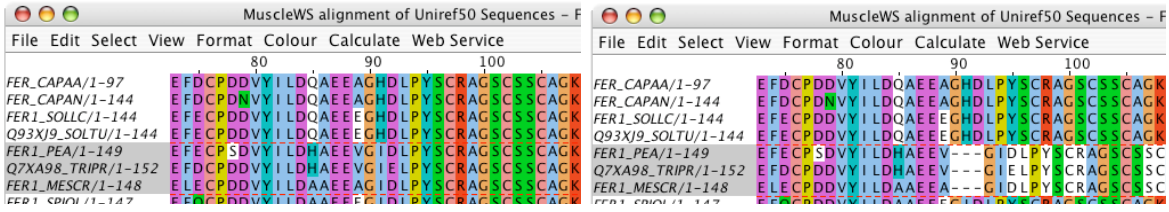


Figure 1.20: **Introducing gaps.** Gaps are introduced as the selected group is dragged to the right.

Gaps can be removed by dragging the residue to the immediate right of the gap leftwards whilst holding down [SHIFT] (for single sequences) or [CTRL] (for a group of sequences).

Editing in Cursor mode

Gaps can be inserted in cursor mode by pressing [SPACE]. Gaps will be inserted at the cursor, pushing the residue under the cursor to the right. To insert *n* gaps type *n* and then press [SPACE]. To insert gaps into all sequences of a group, use [CTRL]-[SPACE] or [SHIFT]-[SPACE] (both keys held down together).

Gaps can be removed in cursor mode by pressing [BACKSPACE]. The gap under the cursor will be removed. To remove *n* gaps, type *n* and then press [BACKSPACE]. Gaps will be deleted up to the number specified. To delete gaps from all sequences of a group, use [CTRL]-[BACKSPACE] or [SHIFT]-[BACKSPACE] (both keys held down together).

Undoing edits

Jalview supports the undoing of edits via the *Edit* ⇒ *Undo Edit* alignment window menu option. Each editing action is stored and can be reversed in sequence. Colouring of the alignment is not reversible via the *Undo* option.

Exercise 7: Editing alignments

7.a. Load the URL http://www.jalview.org/tutorial/unaligned.fa which contains part of the ferrodoxin alignment from PF03460. Select the lower 5 sequences.

7.b. Place the mouse cursor over the selection at column 24. (`RVGGQAK` in O23813_MAIZE). Hold down [CTRL], press down the mouse button and drag the sequences two spaces to the right until the G is aligned with the G in the row above. Release the mouse button and the [CTRL] key.

7.c. Repeat for the next residue Q until it aligns with the Q in the row above.

7.d. Try editing a single sequence. Hold down [SHIFT] and drag the sequence to the right. Just the clicked on sequence should move.

7.e. Delete the gaps created in the previous step by [SHIFT]-Clicking on the residue and dragging it to the left.

7.f. Use the *Edit* ⇒ *Undo Edit* menu option to step backwards through the edits you have made.

Exercise 8: Keyboard edits

8.a. Load the sequence alignment at http://www.jalview.org/tutorial/unaligned.fa. Enter cursor mode by pressing [F2]

8.b. Insert 58 gaps at the start of the first sequence (FER_CAPAA). Press *58* then *[SPACE]*.

8.c. Go down one sequence and select rows 2-5 as a block. Click on the second sequence id (FER_CAPAN). Hold down shift and click on the fifth (FER1_PEA).

8.d. Insert 6 gaps at the start of this group. Go to column 1 row 2 by typing *1,2* then pressing *[RETURN]*. Now insert 6 gaps. Type *6* then hold down [CTRL] and press the space bar.

8.e. Now insert one gap at column 34 and another at 38. Insert 3 gaps at 47. Press *3 4 C* then [CTRL]-[SPACE]. Press *3 8 C* then [CTRL]-[SPACE]. Press *4 7 C* then *3 [CTRL-SPACE]* the first through fourth sequences are now aligned.

8.f. The fifth sequence (FER1_PEA) is poorly aligned. We will delete some gaps and add some new ones. Navigate to the start of sequence 5 and delete 3 gaps. Press *1 , 5 [RETURN]* then *3 [BACKSPACE]* to delete three gaps. Go to column 31 and delete the gap. Press *3 1 C [BACKSPACE]* .

8.g. Similarly delete the gap now at column 34, then insert two gaps at column 38 . Press *3 4 C [BACKSPACE] 3 8 C 2 [SPACE]*. Delete three gaps at 44 and insert one at 47 by pressing *4 4 C 3 [BACKSPACE] 4 7 C [SPACE]*. The top five sequences are now aligned.

1.6.5 Hiding regions

It can be the case that one does not wish to view some sequences or residues in the alignment but does not wish to remove them from the alignment. Jalview allows sequences or regions to be hidden. To hide a set of sequences, select them and right-click the mouse on the selected sequence id's to bring up the context menu. Select *Hide Sequences* and the sequences will be concealed, with a small triangle indicating their position (Figure 1.21). To unhide (reveal) the sequences, right click on the triangle and select *Reveal Sequences* from the context menu.

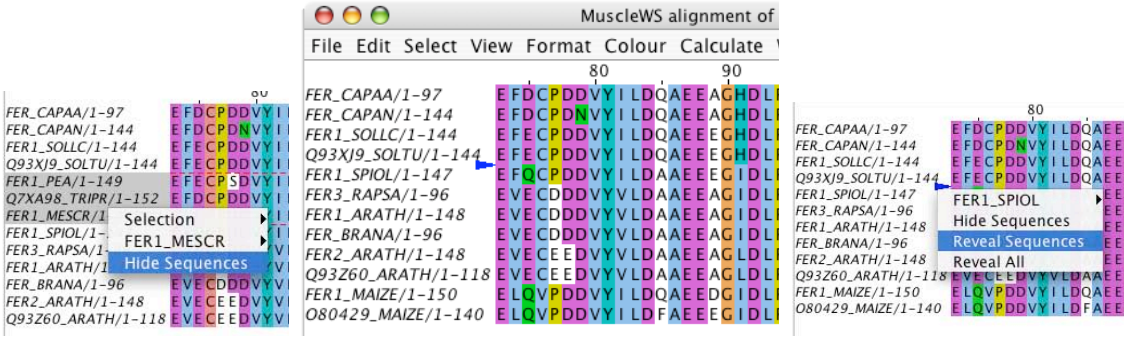


Figure 1.21: **Hiding Sequences** Hidden sequences are represented by a blue triangle in the sequence ID panel

A similar mechanism applies to columns (Figure 1.22). Selected columns can be hidden and revealed

in the same way via the context menu (right click) on the ruler bar. The hidden selection is indicated by a blue triangle in the ruler bar.

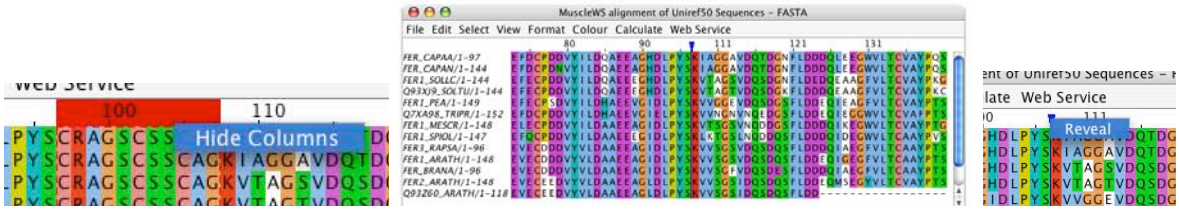


Figure 1.22: **Hiding Columns** Hidden columns are represented by a blue triangle in the ruler bar

Exercise 9: Hiding and revealing regions

9.a. Close all windows then open the PFAM accession PF03460. Select a contiguous set of sequences by clicking and dragging on the sequence ID panel. Right click on the selected sequence IDs and select *Hide Sequences*.

9.b. Right click on the blue triangle indicating hidden sequences and select *Reveal Sequences*. (If you have hidden all sequences then you will need to use the alignment window menu option *View* ⇒ *Show* ⇒ *All Sequences*.)

9.c. Repeat but using a non-contiguous set of sequences. Note that when multiple regions are hidden there are two options, *Reveal Sequences* and *Reveal All*.

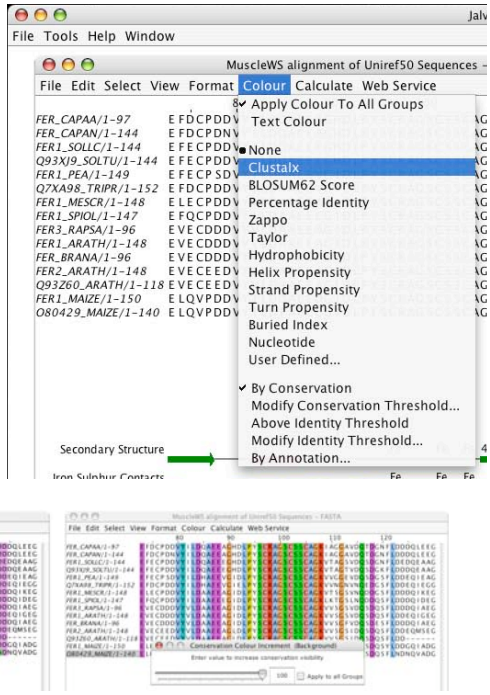
9.d. Repeat the above but hiding and revealing columns instead of sequences.

1.7 Colouring sequences

Colouring sequences is a key aspect of alignment presentation. Jalview allows both for colouring of the whole alignment and colouring of selections/groups. The colour schemes available are the same for both mechanisms and are described in section 1.7.3 below.

1.7.1 Colouring the whole alignment

The alignment can be coloured via the Colour menu option in the alignment window. Selecting the colour scheme causes all residues to be coloured.
For certain colour schemes, the intensity of the colour can be scaled by a conservation threshold. Selecting *Colour* ⇒ *By Conservation* brings up a selection box allowing the alignment colouring to be modified. Selecting a higher threshold limits colouring to higher conserved groups (Figure 1.23).



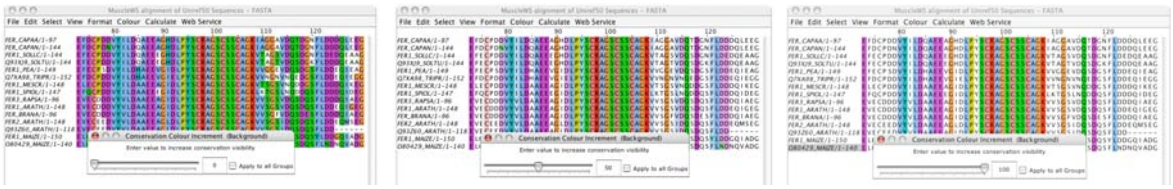


Figure 1.23: **Conservation Threshold** The transparency of the residue colouring is controlled by the conservation threshold. The effect of 0% (left), 50% (center) and 100% (right) thresholds are shown.

1.7.2 Colouring a group or selection

Selections or groups can be coloured via two ways. The first is via the alignment Colour menu but ensuring that the Apply to all groups flag is not selected. This will have to be turned off specifically as it is on by default.

The second method is to use the *Selection* ⇒ *Group* ⇒ *Group Colour* context menu option obtained by right clicking on the group (Figure 1.24).

1.7.3 Colour schemes

Full details on each colour scheme can be found in the Jalview on-line help.

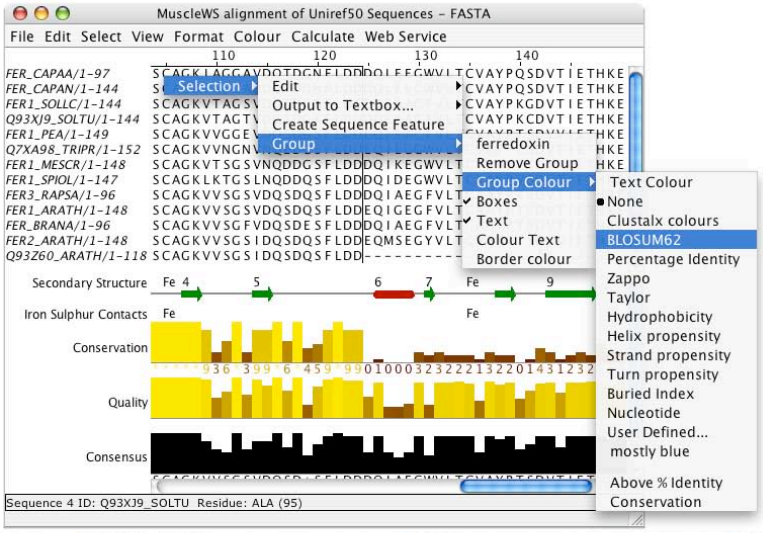
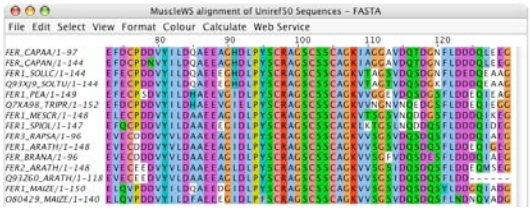


Figure 1.24: **Colouring a group via the context menu.**

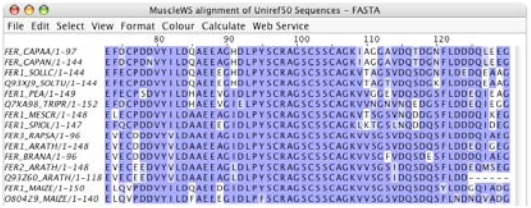
ClustalX

This is an emulation of the default colourscheme used for alignments in Clustal X, a graphical interface for the ClustalW multiple sequence alignment program. Each residue in the alignment is assigned a colour if the amino acid profile of the alignment at that position meets some minimum criteria specific for the residue type.



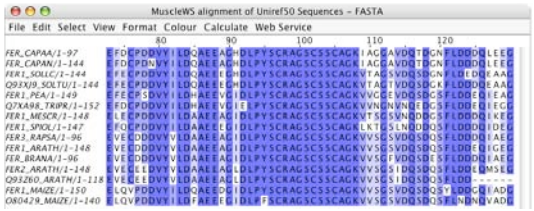
Blosum62

Gaps are coloured white. If a residue matches the consensus sequence residue at that position it is coloured dark blue. If it does not match the consensus residue but the Blosum 62 matrix gives a positive score, it is coloured light blue.



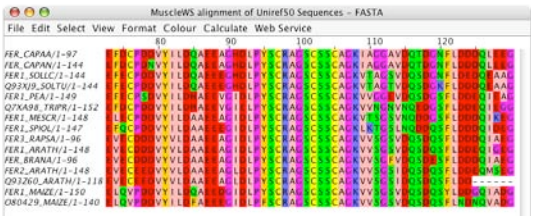
Percentage Identity

The Percent Identity option colours the residues (boxes and/or text) according to the percentage of the residues in each column that agree with the consensus sequence. Only the residues that agree with the consensus residue for each column are coloured.



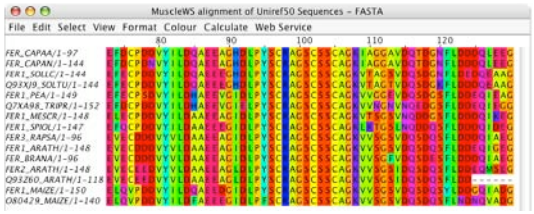
Zappo

The residues are coloured according to their physicochemical properties. The physicochemical groupings are Aliphatic/hydrophobic, Aromatic, Positive, Negative, Hydrophillic, conformationally special, and Cyst(e)ine.



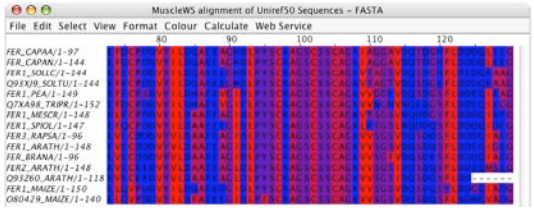
Taylor

This colour scheme was devised by Willie Taylor and an entertaining description of it's origin can be found in Protein Engineering, Vol 10 , 743-746 (1997)



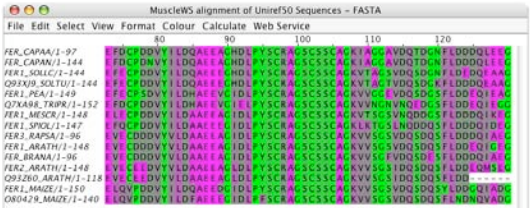
Hydrophobicity

Residues are coloured according to the hydrophobicity table of Kyte, J., and Doolittle, R.F., J. Mol. Biol. 1157, 105-132, 1982. The most hydrophobic residues are coloured red and the most hydrophilic ones are coloured blue.



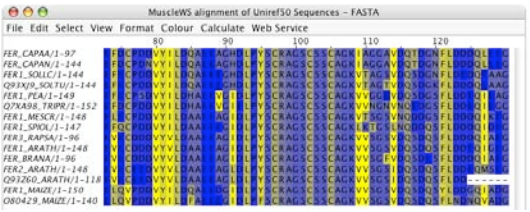
Helix Propensity

The residues are coloured according to their helix propensity. The highest propensity is magenta, the lowest is green.



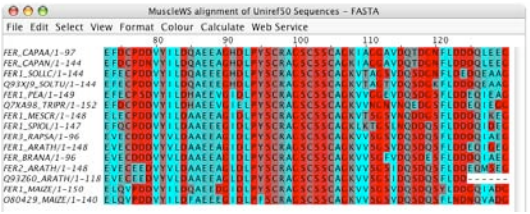
Strand Propensity

The residues are coloured according to their Strand propensity. The highest propensity is Yellow, the lowest is blue.



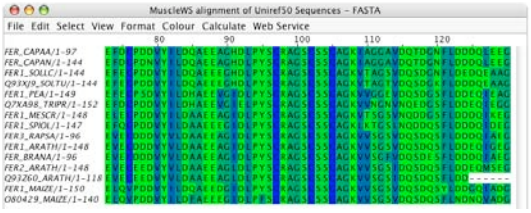
Turn Propensity

The residues are coloured according to their turn propensity. The highest propensity is red, the lowest is cyan.



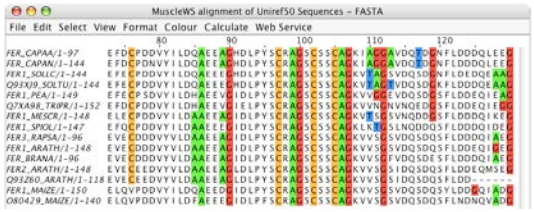
Buried Index

The residues are coloured according to their burial propensity. The highest propensity is blue, the lowest is green.



Nucleotide

Residues are coloured with four colours corresponding to the four nucleotide bases. All non ACTG residues are uncoloured.



Exercise 10: Colouring Alignments

10.a. Open a sequence alignment, for example the PFAM domain PF03460. Select the alignment menu option *Colour* ⇒ *ClustalX*. Note the colour change. Now try all the other colour schemes in the *Colour* menu. Note that some colour schemes do not colour all residues.

10.b. Colour the alignment using *Colour* ⇒ *Blosum62*. Select a group of around 4 similar sequences. Use the context menu (right click on the group) option *Selection* ⇒ *Group* ⇒ *Group Colour* ⇒ *Blosum62* to colour the selection. Notice how some residues which were not coloured are ntw coloured. The calculations performed for colouring schemes just apply to the group being coloured, not the whole alignment.

10.c. Keeping the same selection as before, colour the complete alignment using *Colour* ⇒ *Taylor*. Select the menu option *Colour* ⇒ *By Conservation*. Slide the selector from side to side and observe the changes in the alignment colouring in the selection and in the complete alignment.

User Defined

This dialogue allows the user to create any number of names colour schemes at will. Any residue may be assigned any colour. The colour scheme can then be named. If you save the colour scheme, this name will appear on the Colour menu

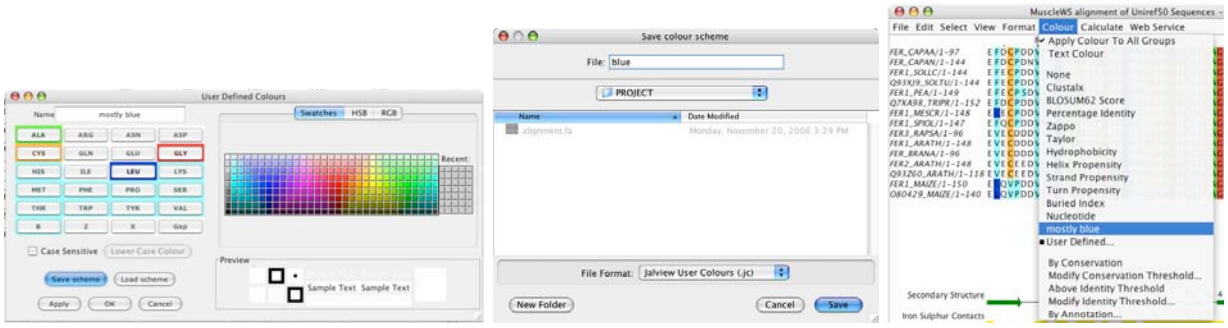


Figure 1.25: **Creation of a user defined colour scheme.** Residue types are assigned colours (left). The profile is saved (center) and can then be accessed via the *Colour* menu (right).

Exercise 11: User defined colour schemes

11.a. Load a sequence alignment. Select the alignment menu option *Colour* ⇒ *User Defined*. A dialogue window will open.

11.b. Click on an amino acid button, then select a colour for that amino acid. Repeat till all amino acids are coloured to your liking.

11.c. Insert a name in the appropriate field and click *Save Scheme*. You will be prompted for a file name in which to save the colour scheme. The dialogue window can now be closed.

11.d. The new colour scheme appears in the list of colour schemes in the *Colour* menu and can be selected in future Jalview sessions.

1.8 Screen layout and graphics output

1.8.1 Screen layout

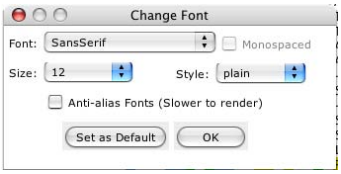
Jalview provides two screen layout modes, unwrapped (the default) where the alignment is in one long line across the window, and wrapped, where the alignment is on multiple lines, each the width of the window. Most layout options are controlled by the Format menu option in the alignment window.

Wrapped alignments

Wrapped alignments can be toggled on and off using the *Format* ⇒ *Wrap* menu option (Figure 1.26). Note that the annotation lines are also wrapped. Wrapped alignments are great for publications and presentations but are hard to work with on the screen. Selecting regions can be difficult if they span more than one line.

Fonts

Text appearance can be modified via the *Format* ⇒ *Font...* alignment window menu. Additionally, font size and spacing can be adjusted rapidly by clicking the middle mouse button and dragging across the alignment window.



Numbering and annotations

The *Format* menu also provides options to control the display of numbers and annotations on sequence alignments. The annotation lines which appear below the sequence alignment can be hidden by toggling the *View* ⇒ *Show Annotations* menu option. Additionally, each annotation line

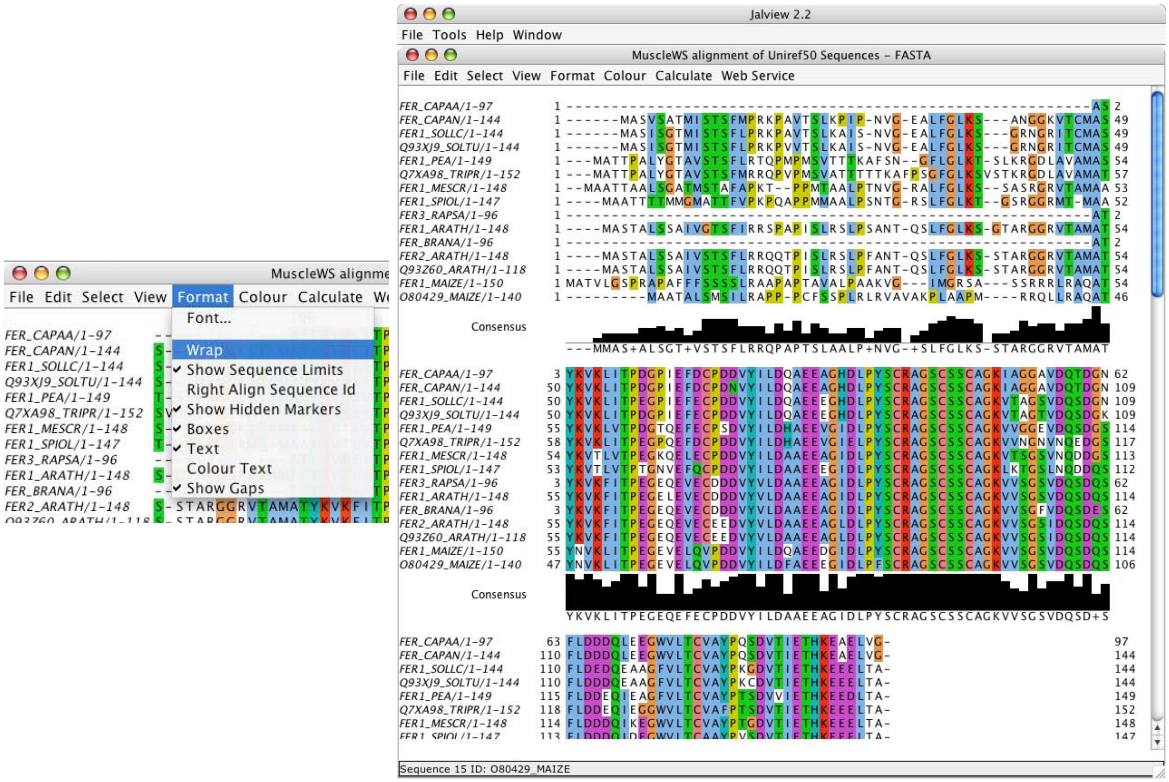


Figure 1.26: **Wrapping the alignment**

can be hidden and revealed in the same way as sequences via the context menu on the annotation name panel (Figure 1.27). Annotations can be reordered by dragging the annotation line label on the annotation label panel. Placing the mouse over the top annotation label brings up a resize icon. Dragging the top annotation line label up and down alters the relative size of the sequence alignment and annotation alignment panels.

Exercise 12: Screen Layout

12.a. Start Jalview and open the URL http://www.jalview.org/examples/exampleFile.jar. Select *Format* ⇒ *Wrap* from the alignment window menu. Experiment with the various options from the *Format* menu. to adjust the ruler placement, sequence ID format and so on.

12.b. Hide all the annotation rows by selecting *View* ⇒ *Show Annotations* from the alignment window menu. Reveal the annotations by selecting the same menu option.

12.c. Right click on the annotation row labels to bring up the context menu. Select *Hide This Row*. Bring up th econtext menu again and select *Show All Hidden Rows* to reveal them

12.d. Annotations can be reordered by clicking and dragging the row to the desired position. Click on the *Consensus* row and drag it upwards to just above *Quality*. The rows should now be reordered. Features and annotations are covered in more detail in section 2.3 below.

12.e. Drag the Secondary Structure label to resize the Alignment Annotation panel

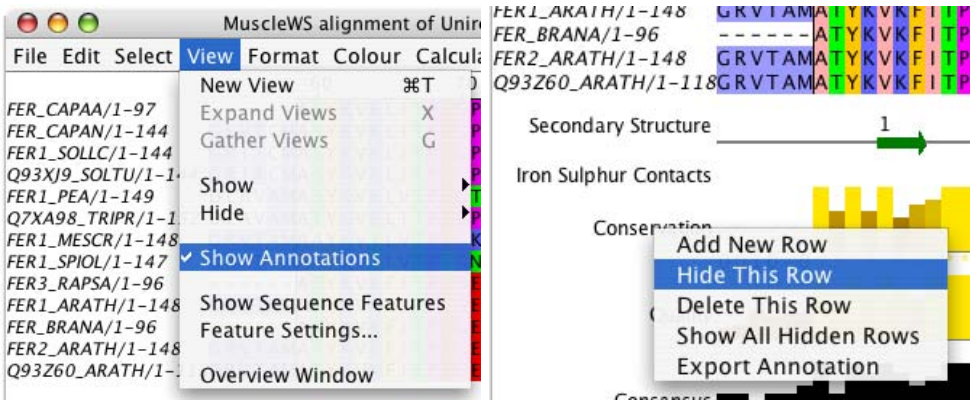
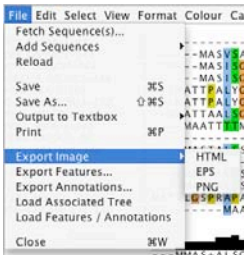


Figure 1.27: **Hiding Annotations** Annotations can either be hidden from the *View* menu (left) or individually from the context menu (right)

1.8.2 Graphical output

Jalview allows alignments to be exported in three different formats, each of which is suited to a particular purpose. Image export is via the *File* ⇒ *Export Image* ⇒ ... alignment window menu option.

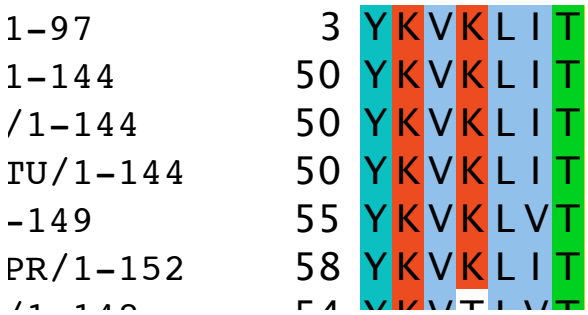


HTML

HTML is the format used by web pages. Jalview outputs the alignment as an HTML table with all the colours and fonts as seen. This file can then be viewed directly with any web browser. Each residue is placed in an individual table cell. Unwrapped alignments will produce a very wide page.

EPS

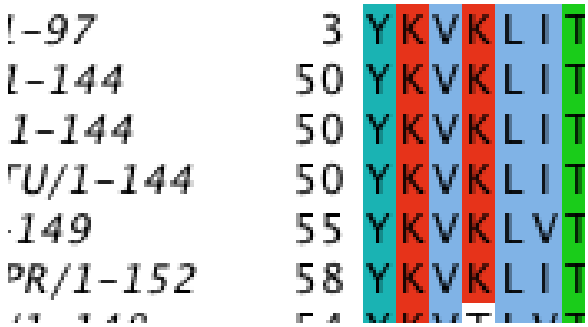
EPS is Encapsulated Postscript. It is the format of choice for publication and posters as it gives the highest quality output of any of the image types. It can be scaled indefinitely so will still look good on an A0 poster. This format can be read by most good presentation and graphics packages such as Adobe Illustrator.



Zoom Detail of EPS image.

PNG

PNG is Portable Network Graphics. This output option produces an image that can be easily included in web pages and incorporated in presentations using e.g. Powerpoint or Open Office. It is a bitmap image so does not scale and is unsuitable for use on posters.



Zoom Detail of PNG image.

Exercise 13: Graphical Output

13.a. Load the example Jalview Jar file in Exercise 12. Customise it how you wish but leave it unwrapped. Select *File* ⇒ *Export Image* ⇒ *HTML* from the alignment menu. Save the file and open it in your favourite web browser.

13.b. Now wrap the alignment (Exercise 12) and export the image to HTML again. Compare the two images. Note that the exported image matches the format displayed in the alignment window but annotations are not exported.

13.c. Export the alignment using the *File* ⇒ *Export Image* ⇒ *PNG* menu option. Open the file in an image viewer that allows zooming (eg. Paint or Photoshop on Windows, Preview on Mac OS X) and zoom in. Notice that the image is a bitmap and it becomes pixelated very quickly. Note also that the annotation lines are included in the image.

13.d. Export the alignment using the *File* ⇒ *Export Image* ⇒ *EPS* menu option. Open the file in a suitable program such as Ghostview or Preview (Mac OS X). Zoom in and note that the image is indefinitely scalable.

Chapter 2

Analysis

2.1 Analysis of alignments

Jalview provides support for sequence analysis in two ways. A number of analytical methods are 'built-in' and run inside Jalview itself and are mostly accessed from the *Calculate* alignment window menu. Other, more computationally intensive analyses, are run outside Jalview via web services, typically accessed via the *Web Services*. It is anticipated that Jalview will extend further the range of Web Services based analyses.

This section describes the built in analyses. Web Services are described in a subsequent section.

2.1.1 PCA

This calculation creates a spatial representation of the similarities within the current selection or the whole alignment if no selection has been made. After the calculation finishes, a 3D viewer displays the each sequence as a point in 3D 'similarity space'. Sets of similar sequences tend to lie near each other in this space. Note: The calculation is computationally expensive, and may fail for very large sets of sequences - usually because the JVM has run out of memory. A future release of Jalview will be able to avoid this by executing the calculation via a web service.

What is PCA?

Principal components analysis is a technique for examining the structure of complex data sets. The components are a set of dimensions formed from the measured values in the data set, and the principle component is the one with the greatest magnitude, or length. The sets of measurements that differ the most should lie at either end of this principle axis, and the other axes correspond to less extreme patterns of variation in the data set. In this case, the components are generated by an eigenvector decomposition of the matrix formed from the sum of BLOSUM scores at each aligned

position between each pair of sequences. The basic method is described in the 1995 paper by *G. Casari, C. Sander* and *A. Valencia* [1] and implemented at the SeqSpace server at the EBI.

The PCA Viewer

PCA analysis can be launched from the *Calculate* ⇒ *Principle Component Analysis* menu option. PCA requires a selection containing at least 4 sequences. A window opens containing the PCA tool (Figure 2.1). Each sequence is represented by a square, coloured by the background colour of the sequence ID label. The axes can be rotated by clicking and dragging the left mouse button and zoomed using the ↑ and ↓ keys or the scroll wheel of the mouse (if available). A tool tip appears if the cursor is placed over a sequence. Sequences can be selected by clicking on them. [CTRL]-Click can be used to select multiple sequences. Labels will be shown for each sequence by toggling

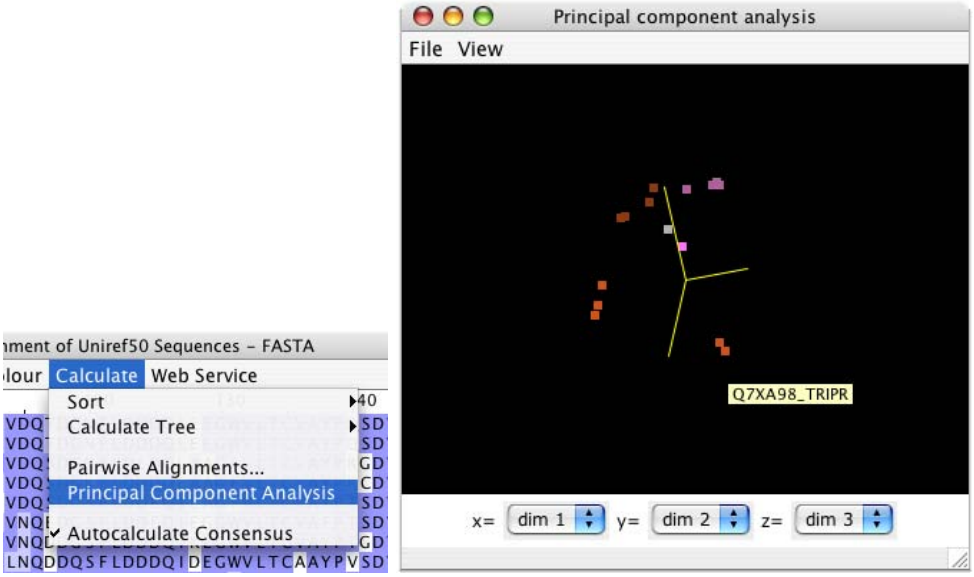


Figure 2.1: **PCA Analysis**

the *View* ⇒ *Show Labels* menu option, and the plot background colour changed via the *View* ⇒ *Background Colour..* dialog box. A graphical representation of the PCA plot can be exported as an EPS or PNG image via the *File* ⇒ *Save As* ⇒ ... submenu.

[1] *Nature Structural Biology* (1995) **2**, 171-8. PMID: 7749921

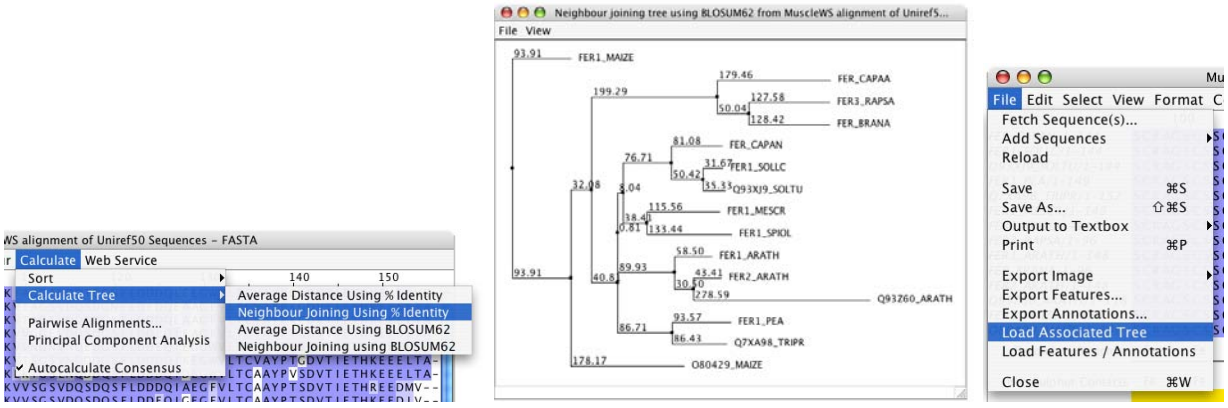


Figure 2.2: **Calculating Trees** Jalview provides four built in models for calculating trees. Jalview can also load precalculated trees in Newick format (right).

Exercise 14: Principle Component Analysis

14.a. Load the alignment at http://www.jalview.org/examples/exampleFile.jar and select *Edit* ⇒ *Undefine Groups*.

14.b. Select the menu option *Calculate* ⇒ *Principle Component Analysis*. A new window will open. Move this window so that the tree, alignment and PCA viewer window are all visible. Try rotating the plot by clicking and dragging the mouse on the plot in the PCA window. Note that clicking on points in the plot will highlight them on the alignment and tree.

14.c. Click on the tree window. Careful selection of mouse location will divide the alignment into a number of groups, each of a different colour. Note how the colour of the sequence id label matches the colour in the PCA plot.

2.1.2 Trees

Jalview can calculate and display trees, providing interactive tree-based grouping of sequences though a tree viewer. All trees are calculated via the *Calculate* ⇒ *Calculate Tree* ⇒ ... submenu. Trees can be calculated from distance matrices determined from % identity or aggregate BLOSUM 62 score and reconstructed with either average distance (UPGMA) or Neighbour joining algorithms. The sequence alignment from which a tree is calculated is the current selection, or the whole alignment if no selection is present.

On calculating a tree, a new window opens (Figure 2.2) which contains the tree. Various display options can be found in the tree window *View* menu, and export options in the *File* ⇒ *Save As* submenu. Newick format is a standard file format for trees which allows them to be exported to other programs. Jalview can also read in external trees in Newick format via the *File* ⇒ *Load Associated Tree* menu option. The sequence ID's have to match between the alignment and the loaded tree. Nodes for which no sequence is found in the alignment are ignored.

Clicking on the tree brings up a cursor across the height of the tree. The sequences are automatically

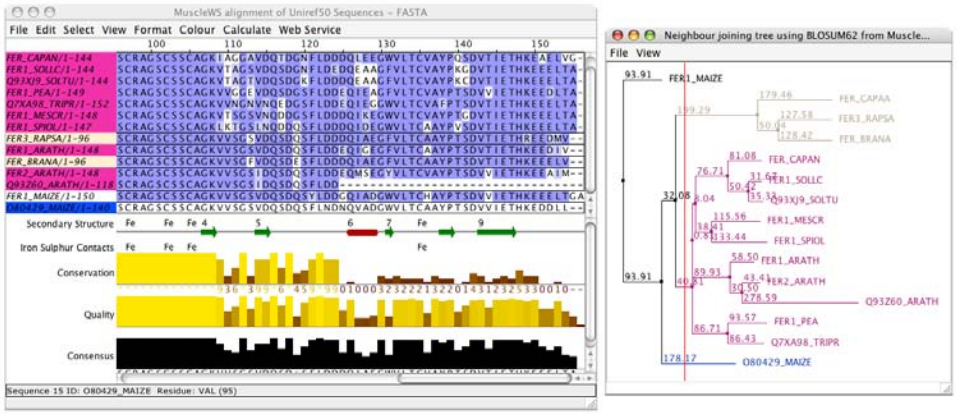


Figure 2.3: **Interactive Trees** The tree level cutoff can be used to designate groups in Jalview

partitioned and coloured (Figure 2.3). To group them together, select the *Calculate* ⇒ *Sort* ⇒ *By Tree Order* alignment window menu option and the correct tree. The sequences will then be sorted to group them together. The coloured background to the sequence IDs can be removed with *Select* ⇒ *Undefine Groups* from the alignment window menu.

Other Calculations

On loading a sequence alignment, Jalview calculates alignment quality, conservation and consensus as alignment annotations. Conservation is calculated according to Livingstone and Barton[2]. Consensus is the modal residue (or + where there is an equal top residue). The inclusion of gaps in the consensus calculation can be toggled by right-clicking on the the Consensus label and selecting *Ignore Gaps in Consensus* from the context menu. Quality is a measure of the inverse likelihood of unfavourable mutations in the alignment. Further details on these calculations can be found in the on-line documentation.

These annotations can be hidden and deleted but are only created on loading an alignment. If they are deleted then the alignment should be saved and reloaded to restore them. Jalview provides a toggle to autocalculate a consensus sequence upon editing. This is normally left selected but for large alignments can be turned off via the *Calculate* ⇒ *Autocalculate Consensus* menu option if the interface is too sluggish.

Jalview can calculate optimal pairwise alignments between arbitrary sequences via the *Calculate* ⇒ *Pairwise Alignments...* menu option. Global alignments of all pairwise combinations of the selected sequences are performed and the results returned in a text box.

[2] "*Protein Sequence Alignments: A Strategy for the Hierarchical Analysis of Residue Conservation.*" Livingstone C.D. and Barton G.J. (1993) *CABIOS* **9**, 745-756

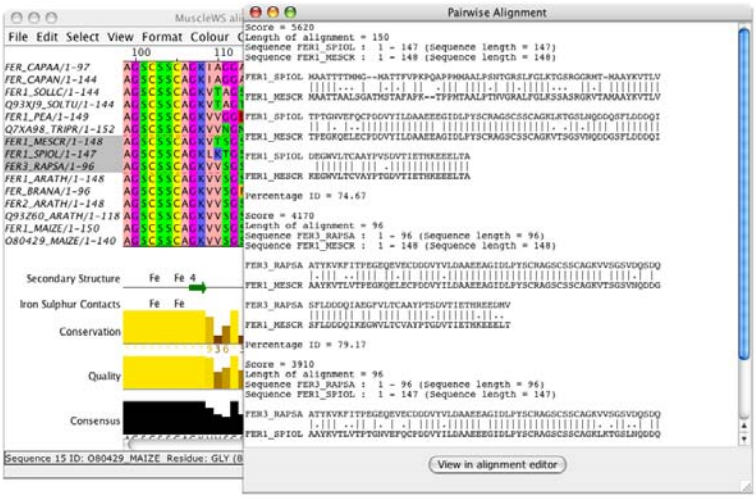


Figure 2.4: **Pairwise alignment of sequences.** Pairwise alignments of three selected sequences are shown in a textbox.

Exercise 15: Trees

15.a. Open the alignment at http://www.jalview.org/tutorial/alignment.fa. Select *Calculate* ⇒ *Calculate Tree* ⇒ *Neighbour Joining Using BLOSUM62*. A new tree window will appear.

15.b. Click on the tree window. A cursor will appear. Note that placing this cursor divides the tree into a number of groups by colour. Place the cursor to give about 4 groups, then select *Calculate* ⇒ *Sort* ⇒ *By Tree Order* ⇒ *Neighbour Joining Tree using BLOSUM62 from ...* . The sequences are reordered to match the order in the tree and groups are formed implicitly.

15.c. Select *Calculate* ⇒ *Calculate Tree* ⇒ *Neighbour Joining Using % Identity*. A new tree window will appear. The group colouring makes it easy to see the diferences between the two trees, calculated using different methods.

15.d. Select from sequence 2 column 60 to sequence 12 column 123. Select *Calculate* ⇒ *Calculate Tree* ⇒ *Neighbour Joining Using BLOSUM62*. A new tree window will appear. It can be seen that the tree contains 11 sequences. It has been coloured according to the already selected groups from the first tree and is calculated purely from the residues in the selection. Comparing the location of individual sequences between the two trees illustrates teh importance of selecting appropriate resides for calculation of trees.

2.2 Web Services

Web services are a technology which allows the Jalview application to offload compute intensive tasks to remote servers via the internet. Jalview will construct a job, ask the remote server to run the job, monitor status of the job and, finally, retrieve the results of the job and display them. The Jalview user is kept informed of the progress of the job through a status window.

It is essential that you have a continuous network connection in order to successfully use Web Services from Jalview.

2.2.1 Realignment

Proteins can be realigned using any of three algorithms: ClustalW[3], Muscle[4] or MAFFT[5] Of these, ClustalW is the slowest but is historically widely used. Muscle is fast and probably the most accurate for smaller alignments and MAFFT is probably the best for large alignments.

To run an alignment web service, select the appropriate method from the *Web Service* ⇒ *Alignment* ⇒ ... submenu (Figure 2.5). A progress window will appear giving information about the job and any errors that occur. After successful completion of the job, a new window is opened with the results, in this case an alignment.

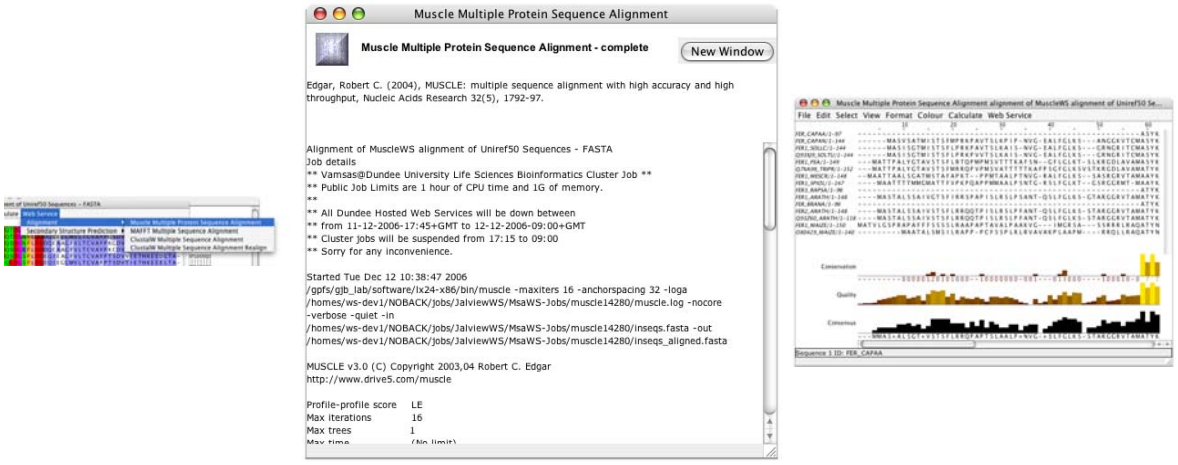


Figure 2.5: **Multiple alignment via web services** The appropriate method is selected from the menu (left), a status box appears (center), and the results appear in a new window (right)

[3] *"CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position specific gap penalties and weight matrix choice."* Thompson JD, Higgins DG, Gibson TJ (1994) *Nucleic Acids Research* **22**, 4673-80

[4] *"MUSCLE: a multiple sequence alignment method with reduced time and space complexity"* Edgar, R.C. (2004) *BMC Bioinformatics* **5**, 113

[5] *"MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform"* Katoh, K., Misawa, K., Kuma, K. and Miyata, T. (2002) *Nucleic Acids Research* **30**, 30593066. and *"MAFFT version 5: improvement in accuracy of multiple sequence alignment"* Katoh, K., Kuma, K., Toh, H. and Miyata, T. (2005) *Nucleic Acids Research* **33**, 511518.

Exercise 16: Multiple Sequence Alignment

16.a. Close all windows and open the alignment at http://www.jalview.org/tutorial/unaligned.fa. Select *Web Service* ⇒ *Alignment* ⇒ *Muscle Multiple Protein Sequence Alignment*. A window will open giving the job status. After a short time, a second window will open with the results of the alignment.

16.b. Select the first sequence set by clicking on the window and try running ClustalW and MAFFT (from the *Web Services* ⇒ *Alignment* menu) on the same initial alignment. Compare them and you should notice small differences.

2.2.2 Secondary Structure Prediction

Secondary structure prediction is performed using the Jpred[6] server at the University of Dundee[7]. The behaviour of this calculation depends on the current selection:

- If nothing is selected, and the displayed sequences appear to be aligned, then a JNet prediction will be run for the first sequence in the alignment, using the current alignment. Otherwise the first sequence will be submitted for prediction.
- If just one sequence (or a region on one sequence) has been selected, it will be submitted to the automatic JNet prediction server for homolog detection and prediction.
- If a set of sequences are selected, and they appear to be aligned, then the alignment will be used for a Jnet prediction on the first sequence in the set (that is, the one that appears first in the alignment window).

Jpred is launched in the same way as the other web services. Select *Web Services* ⇒ *Secondary Structure Prediction* ⇒ *JNet Secondary Structure Prediction* from the alignment window menu (figure 2.6). A status window opens to inform you of the progress of the job. Upon completion, a new alignment window opens and the Jpred predictions are included as annotations. Consult the Jpred documentation for information on interpreting these results.

Exercise 17: Secondary Structure Prediction

17.a. Open the alignment at http://www.jalview.org/tutorial/alignment.fa. Select the sequence *FER_MESCR* by clicking on the sequence ID. Then select *Web Services* ⇒ *Secondary Structure Prediction* ⇒ *JNet Secondary Structure Prediction* from the alignment window menu. A status window will appear and after some time a new window with the JPred prediction will appear. Note that the number of sequences in the results window is many more than in the original alignment as JNet performs a PSI-BLAST search to expand the prediction dataset.

17.b. Select a different sequence and perform a JNet prediction in the same way. There will probably be minor differences in the predictions.

[6] "*Jpred: A Consensus Secondary Structure Prediction Server*" Cuff, J. A., Clamp, M. E., Siddiqui, A. S., Finlay, M. and Barton, G. J. (1998) *Bioinformatics* **14**, 892-893

[7] http://www.compbio.dundee.ac.uk/ www-jpred/

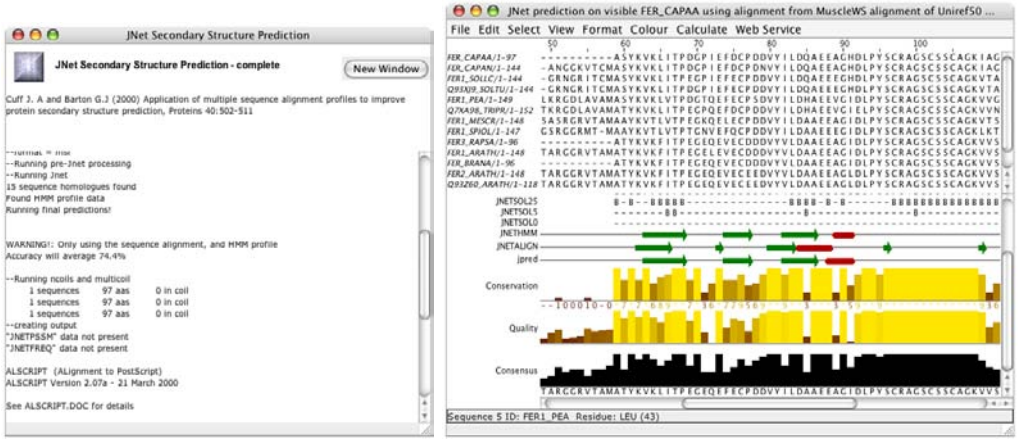


Figure 2.6: **Secondary Structure Prediction** Status (left) and results (right) windows for JNet predictions.

2.3 Features and Annotation

Features and annotations are additional information that is overlaid on the sequences and the alignment. Annotations are properties of the alignment as a whole - as the alignment changes, the annotations will change along with it. Features are properties of the individual sequences. They do not change with the alignment. Annotations are associated with columns in the alignment. Features are associated with specific residues in the sequence.

2.3.1 Creating sequence features

Sequence features can be created by simply selecting the area in a sequence (or sequences) to form the feature and selecting *Selection* ⇒ *Create Sequence Feature* from the right-click context menu (Figure 2.7). A dialogue box allows the user to customise the feature with respect to name, group, and colour. The feature is then associated with the sequence. Moving the mouse over a residue associated with a feature brings up a tool tip listing all features associated with the residue.

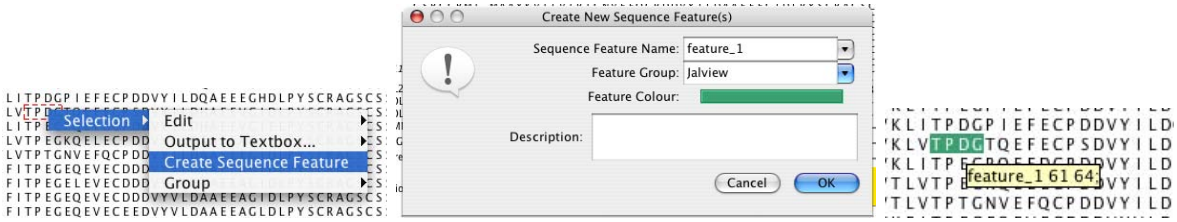


Figure 2.7: **Creating sequence features.** Features can readily be created from selections via the context menu and are then displayed on the sequence.

Creation of features from a selection spanning multiple sequences results in the creation of one

feature per sequence. Each feature remains associated with it's own sequence.

2.3.2 Customising feature display

Feature display can be toggled on or off by selecting the *View* ⇒ *Show Sequence Features* menu option. When multiple features are present it can be necessary to customise the display. Jalview allows the display, colour, rendering order and transparency of features to be modified via the *View* ⇒ *Feature Settings...* menu option. This brings up a dialogue window (Figure 2.9) which allows the visibility of individual feature types to be selected, colours changed (by clicking on the colour of each sequence feature type) and the rendering order modified by dragging feature types to a new position in the list. Dragging the slider alters the transparency of the feature rendering.

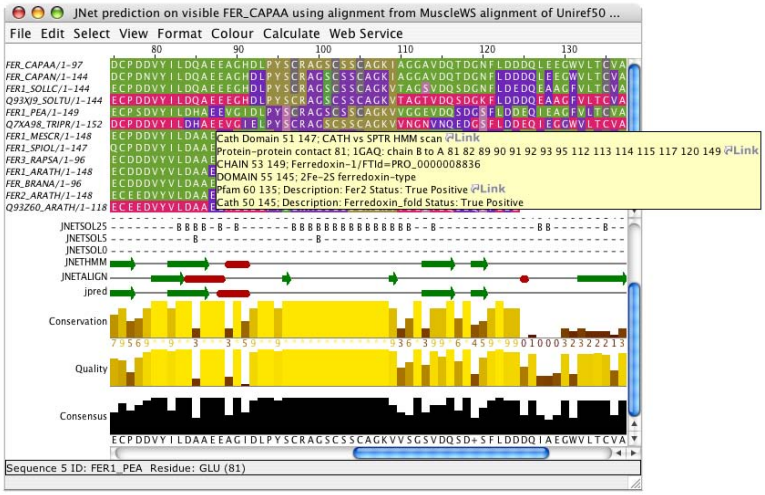


Figure 2.8: **Multiple sequence features.** An alignment with many sequence features overlaid

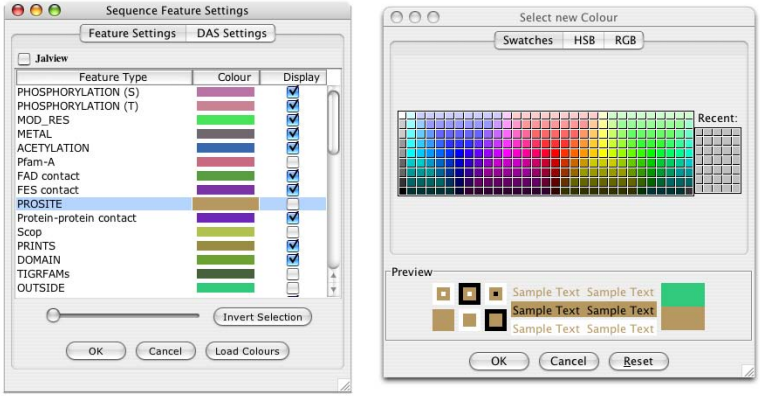


Figure 2.9: **Customising sequence features.** Features can be recoloured, switched on or off and have the rendering order changed.

Exercise 18: Creating features

18.a. Open the alignment at http://www.jalview.org/tutorial/alignment.fa. We know that the Cysteine residues at columns 97, 102, 105 and 135 are involved in iron binding so we will create them as features. Navigate to column 97, sequence 1. Select the entire column by clicking in the ruler bar. Then right-click on the selection to bring up the context menu and select *Selection* ⇒ *Create Sequence Feature*. A dialogue box will appear.

18.b. Enter a suitable Sequence Feature Name (e.g. "Iron binding site") in the appropriate box. Click on teh Feature Colour bar to change the colour if desired, add a short description ("One of four Iron binding Cysteines") and press OK. The features will then appear on the sequences.

18.c. Roll the mouse cursor over the new features. Note that the position given in the tool tip is the residue number, not the column number. To demonstrate that there is one feature per sequence, clear all selections by pressing [ESC] then insert a gap in sequence 3 at position 95. Roll the mouse over the features and you will see that the feature has moved with the sequence. Delete the gap you created.

18.d. Add a similar feature to column 102. When the feature dialogue box appears, clicking the Sequence Feature Name box brings up a list of previously described features. Using the same Sequence Feature Name allows the features to be grouped.

18.e. Select *View* ⇒ *Feature Settings...* from the alignment window menu. The Sequence Feature Settings window will appear. Move this so that you can see the features you have just created. Click the check box for "Iron binding site" under *Display* and note that display of this feature type is turned off. Click it again and note that the features are now displayed. Close the sequence feature settings box by clicking *OK* or *Cancel*.

2.3.3 Creating user defined annotation

Annotations are properties that apply to the alignment as a whole and are visualised on rows in the annotation panel. To create a new annotation row, right click on the annotation label panel and select the *Add New Row* menu option (Figure 2.10). A dialogue box appears. Enter the label to use for this row and a new row will appear.

Figure 2.10: **Creating a new annotation row.** Annotation rows can be reordered by dragging them to the desired place.

To create a new annotation, first select all the positions to be annotated on the appropriate row. Right-clicking on this selection brings up the context menu which allows the insertion of graphics for secondary structure (*Helix* or *Sheet*), text *Label* and the colour in which to present the annotation

(Figure 2.11). On selecting *Label* a dialogue box will appear, requesting the text to place at that position. After the text is entered, the selection can be removed and the annotation becomes clearly visible[8]. Annotations can be coloured or deleted as desired.

Figure 2.11: **Creating a new annotation.** Annotations are created from a selection on the annotation row and can be coloured as desired.

Exercise 19: Annotating alignments

19.a. Load the alignment at http://www.jalview.org/tutorial/alignment.fa. Right-click on the annotation label for *Conservation* to bring up the context menu and select *Add New Row*. A dialogue box will appear asking for *Label for annotation*. Enter "Iron binding site" and click *OK*. A new, empty, row appears.

19.b. Navigate to column 97. Select column 97 on the new annotation row. Right click on the selection and select *Label* from the context menu. Enter "Fe" in the box and click *OK*. Right-click on the selection again and select *Colour*. Choose a colour from the colour chooser dialogue and click *OK*. Press [ESC] to remove the selection.

19.c. Select columns 70-77 on the annotation row. Right-click and choose *Sheet* from the context menu. You will be prompted for a label. Enter "B" and press OK. A new line showing the sheet as an arrow appears. The colour of the label can be changed but not the colour of the sheet arrow.

2.3.4 Importing features from databases

Jalview supports feature retrieval from public databases exported via the Distributed Annotation System[9]. To retrieve features, select *View* ⇒ *Feature Settings...* from the alignment window menu. Select the *DAS Settings* tab in the Feature Settings Window (Figure 2.12). A list of DAS sources compiled from the major public DAS registry[10] is shown in the left hand pane. Highlighting an entry on the left brings up information about that source in the right hand panel.

Select appropriate DAS sources as required then click on *Fetch DAS Features*. This can take some time if a large number of sources is selected and if the alignment contains a large number of sequences. The retrieved features are shown on the sequence and can be customised as described previously.

Following DAS feature retrieval, the *Feature Settings* panel takes on a slightly different appearance

[8] When annotating a block of positions, the text can be partly obscured by the selection highlight. Pressing the [ESC] key clears the selection and the label is then visible.

[9] http://www.biodas.org/

[10]

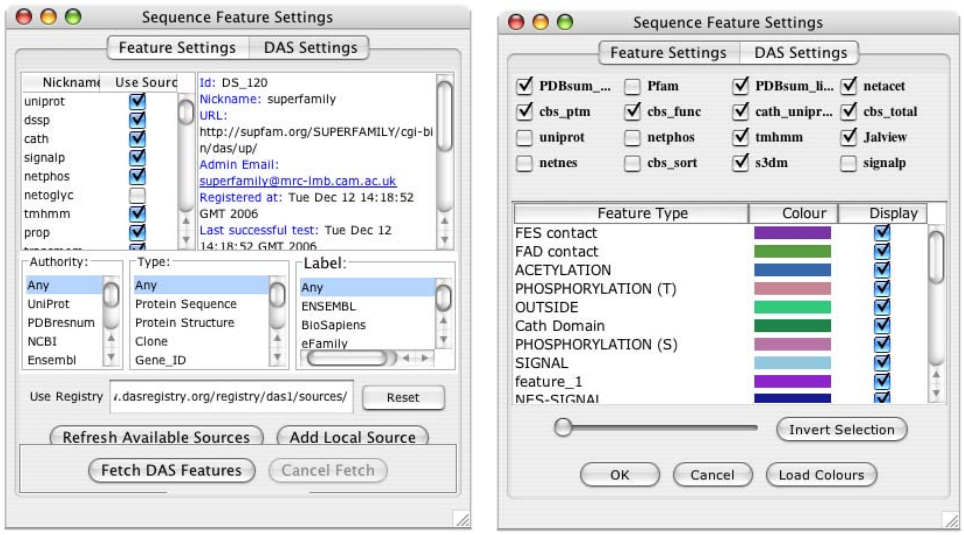


Figure 2.12: **Retrieving DAS annotations.** DAS featrues are retrieved using the *DAS Settings* tab (left) and display customised using the *Feature Settings* tab (right).

(Figure 2.12 (right)). Each data source is listed and groups of features from one data source ccan be selected/deselected by checking the labeled box at the top of the panel.

Exercise 20: Retrieving features with DAS

20.a. Load the alignment at http://www.jalview.org/tutorial/alignment.fa. Select *View* ⇒ *Sequence Features...* from the alignment window menu. Select the *DAS Settings* tab. A long list of available DAS sources is listed. Select a small number, eg Uniprot, DSSP, signalP and netoglyc. Click *OK*. A window may prompt whether you wish Jalview to map the sequence IDs onto Uniprot IDs. Click *Yes*. Jalview will start retrieving features. As features become available they will be mapped onto the alignment.

20.b. If Jalview is taking too long to retrieve features, the process can be cancelled with the *Cancel Fetch* button. Rolling the mouse cursor over the sequences reveals a large number of features annotated in the tool tip. Close the Feature Settings window.

20.c. Select *View* ⇒ *Feature Settings...* to reopen the Feature Settings window. All the loaded feature types should now be displayed. Those at the top of the list sit on top of and obscure those below. Move the feature settings window so that the alignment is visible and uncheck some of the feature types by clicking the tick box in the display column.. Observe how the alignment display changes. Note that unselected feature types do not appear in the tool tip.

20.d. Reorder the features by dragging feature types up and down the order in the Feature Settings panel. e.g. Click on *CHAIN* then move the mouse downwards to drag it below *DOMAIN*. Note that *DOMAIN* is now shown on top of *CHAIN* in the alignment window. Drag *METAL* to the top of the list. Observe how the cysteine residues are now highlighted as they have a *METAL* feature associated with them.

2.4 Working with structures

Jalview provides integration between protein sequences and protein structures by providing a simple structure viewer. Structures are visualised as an alpha carbon trace and can be viewed, rotated and coloured in a structure viewer and the results interpreted on a sequence alignment.

2.4.1 Automatic association of PDB structures with sequences

Jalview can automatically determine which structures are associated with a sequence, if that sequence has an ID from a public database such as Uniprot. Right-click on any sequence ID and select *<Sequence ID> ⇒ Associate Structure with Sequence ⇒ Discover PDB IDs* from the context menu (where *<Sequence ID>* is the ID of the sequence on which you clicked) (Figure 2.13). Jalview will attempt to associate the sequence with a Uniprot sequence and from there discover any associated PDB structures. This takes a few seconds and applies to all sequences in the alignment which have valid Uniprot IDs. On moving the cursor over the sequence ID the tool tip now shows the Uniprot ID and any associated PDB structures.

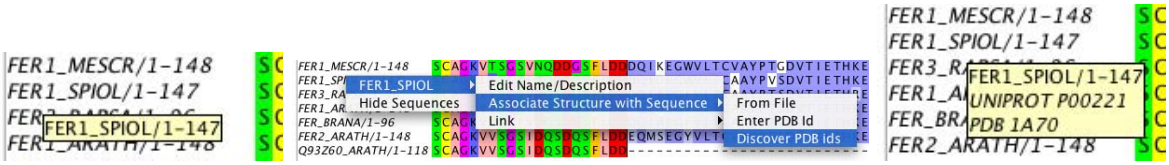


Figure 2.13: **Automatic PDB ID discovery.** The tooltip (left) indicates that no PDB structure has been associated with the sequence. After PDB ID discovery (center) the tool tip now indicates the Uniprot ID and any associated PDB structures (right)

2.4.2 Viewing Protein Structures

The structure viewer can be launched through the sequence ID context menu. Select *<Sequence ID> ⇒ View PDB entry: <PDB ID>*. A viewer window will open with the alpha carbon chain(s) of the structure shown (Figure 2.14). The structure can be rotated by clicking and dragging in the structure window. The structure can be zoomed using the mouse scroll wheel (if available).

Moving the mouse cursor over the sequence to which the structure is linked in the alignment panel highlights the respective residue and shows a label on the structure (Figure 2.14 (right)).

Customisng structure display

Structure display can be modified using the *Colour* and *View* menus in the structure viewer. The background colour can be modified by selecting the *Colours ⇒ Background Colour...* option and

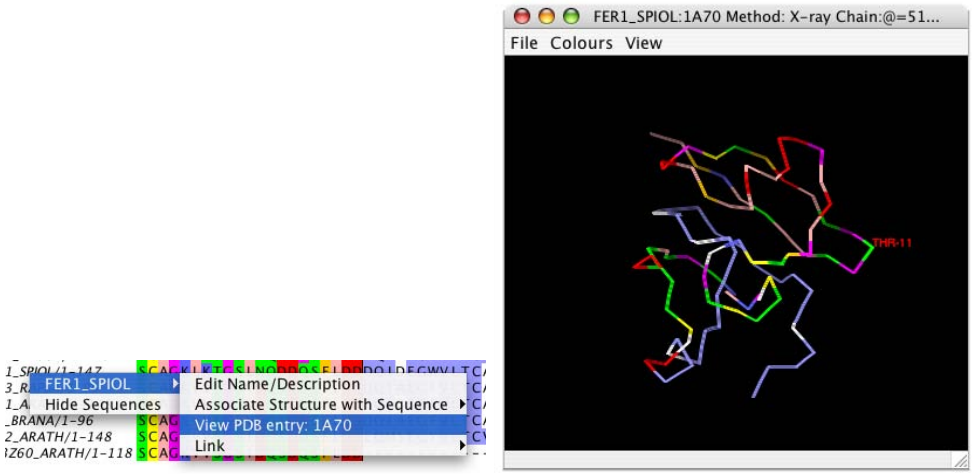


Figure 2.14: **Structure visualisation** The structure viewer is launched from the sequence ID context menu (left) and allows the structure to be visualised as a rotatable and zoomable alpha carbon trace (right).

display parameters such as depth cueing and Z-buffering by selecting the appropriate option from the *View* menu.

The structure can be coloured independently of the sequence by selecting an appropriate colour scheme from the *Colours* menu. It can be coloured according to the alignment using the *Colours* $\Rightarrow$ *By Sequence* option. The image in the structure viewer can be output to EPS or PNG format via the *File* $\Rightarrow$ *Save As* $\Rightarrow$ ... submenu. The mapping between the structure and the sequence (How well and which parts of the structure relate to the sequence) can be viewed with the *File* $\Rightarrow$ *View Mapping* menu option.

Exercise 21: Viewing Structures

21.a. Load the alignment at http://www.jalview.org/examples/exampleFile.jar. Right-click on the sequence ID label for any of the sequences (e.g. *FER1_SPIOL*) to bring up the context menu. Select *FER1_SPIOL* ⇒ *Associate Structure with Sequence* ⇒ *Discover PDB ids*. Jalview will now attempt to find PDB structures for the sequences in the alignment.

21.b. Right-click on the sequence id for *FER1_SPIOL*. Select *FER1_SPIOL* ⇒ *View PDB Entry: 1A70* A structure viewing window appears. Rotate the molecule by clicking and dragging in the structure viewing box. Zoom with the mouse scroll wheel.

21.c. Roll the mouse cursor along the *FER1_SPIOL* sequence in the alignment. Note that if a residue in the sequence maps to one in the structure, a label will appear next to that residue in the structure viewer. Move the mouse over the structure. Placing the mouse over the alpha carbon will bring up a tool tip indicating the name and number of that residue. The corresponding residue in the sequence is highlighted in black. Clicking the alpha carbon toggles the highlight and residue label on and off. Try this by clicking on a set of three or four adjacent residues so that the labels are persistent, then finding where they are in the sequence.

21.d. Select *Colours* ⇒ *Background Colour...* from the structure viewer menu and choose a suitable colour. Press *OK* to apply this. Select *File* ⇒ *Save As* ⇒ *PNG* and save the image. View this with your web browser.

21.e. Select *File* ⇒ *View Mapping* from the structure viewer menu. A new window opens showing the residue by residue alignment between the sequence and the structure.

Chapter 3

Advanced Jalview

3.1 Customising Jalview

3.1.1 Setting preferences

3.2 The Jalview Interface

3.2.1 Multiple views

3.2.2 Keyboard Editing Mode

3.3 Regions

3.3.1 Locked Editing

3.3.2 Alignments including hidden regions

3.3.3 Secondary Structure predictions

3.4 Features and Annotations

3.4.1 Annotation display

Altering annotation row height

3.4.2 Annotation files

3.4.3 Feature files