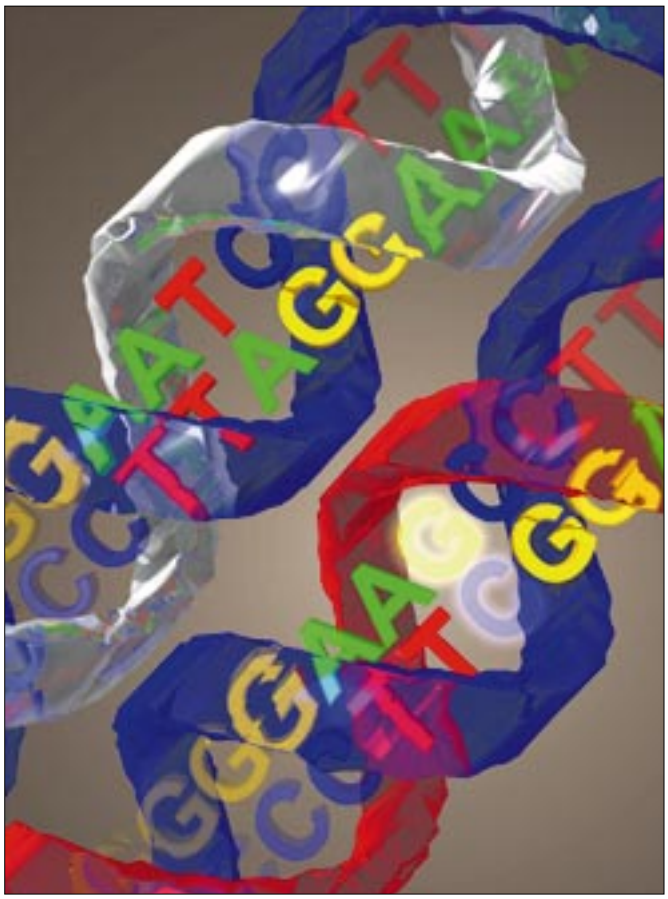
Mutation Analyser

version 1.00



User Manual

18-1118-78
56-3042-87
Edition AA

Amersham Biosciences

Contents

1. Introduction

Sequence-based analysis is rapidly becoming the gold standard in many mutation study areas due to its ultra-high level of sensitivity and specificity. It becomes an increasingly important tool when used together with other popular screening methods such as Immunohistochemical (IHC) analysis and single strand conformation polymorphism (SSCP).

Mutation Analyser from Amersham Biosciences is a software package specifically developed for sequence-based, automated detection of base-specific aberrations in sample sequences generated by ALF*express* or ALF series sequencer instruments.

Mutation Analyser can detect several types of deviation in the sample sequence, including *replacements* at single base positions, *explicit insertions or deletions*, and *mixture insertions or deletions* occurring in heterogeneous cell populations. The software utilises a gene configuration as the basis for comparison, which is itself created in a separate Configuration Editor module supplied with Mutation Analyser. A gene configuration is assembled using wild type sequence data obtained from the major Internet databases or from sequences generated by ALF/ALF*express* sequencers.

Key features of Mutation Analyser

- **Automated detection of single base aberrations**
  Deviations at the single base level are automatically detected, including point replacements (typically mutations or polymorphisms), explicit insertions or deletions, and mixed population insertion/deletions.

- **Deviation detection in mixed cell populations**
  A unique mixture analysis function locates insertions or deletions specific to mixed cell populations and samples taken from heterozygous cell sources.

- **Rapid and simultaneous analysis of many samples**
  Many sample sequences can be simultaneously and rapidly analysed.

- **Interactive user assignment**
  Use the extensive analysis information to perform an interactive user assignment for each deviation. You can approve a deviation, assign the deviation back to wild type, or assign an undetermined flag for further evaluation.

- **Genomic DNA and cDNA/mRNA**
  Sample sequences from genomic DNA and cDNA/mRNA can be analysed. The software automatically accounts for intron regions in genomic DNA sequences.

- **Platform for creating gene configurations**
  The Configuration Editor module allows you to create or edit gene configurations against which sample sequences are analysed. In a gene configuration you are able to define sequencer fragments, exons, areas of interest and points of interest. Thus, the gene configuration defines and documents your evaluation method.

- **Amino acid sequence for the coding sequence**
  Configuration Editor automatically generates the amino acid sequence for the coding sequence data defined in the exons of a gene configuration.

- **Long gene configurations**
  The complete gene configuration can contain 100 sequence regions, each of up to 20,000 bases in length.

- **User-defined deviations**
  Study the raw data curves and include your own user-defined deviations for evaluation.

- **Generation of reports and export files**
  The results of an analysis and subsequent user evaluation can be presented in a full report. Moreover, vital information can be exported to database for subsequent inclusion in other software applications.

- **Graphically interactive analysis**
  The analysis proceeds in a few clearly defined and straightforward stages. Sample information, raw data graphs and analysis data are graphically presented in full colour to enhance user interaction and optimise the evaluation process.

- **Windows 95**
  The software has a modern user interface developed for Windows 95.

1.1 Using the software

You should check that your computer complies with the minimum hardware specification requirements detailed in Chapter 2 and that you are using Microsoft Windows 95 operating system.

The application software should be installed according to the instructions in Chapter 3.

1.1.1 Software protection device

Each Mutation Analyser software supplied by Amersham Biosciences comes with a HASP® software protection device. The HASP® fits into any parallel port of an IBM compatible PC and contains a code that is matched with the current version of the software. In order for the software to run properly, the HASP® must be fitted prior to software installation. The HASP® is transparent which means that peripherals, such as a printer, can be attached to the parallel port containing the HASP and function normally.

Note: If the HASP is not fitted, Mutation Analyser can only be used for demonstration purposes using the supplied Demo files.

1.1.2 Windows 95 operating system

It is assumed that you have a working knowledge of Microsoft Windows 95. For further information on specific operating system functions and terms, you should refer to the appropriate user manual.

1.1.3 Menu commands

Many of the menu commands in Mutation Analyser are similarly located on the right mouse button menu. The contents of a menu are dependent on the field from which you activate the right mouse button function. Many of these menu commands are indicated in the manual, and those that are not mentioned you will quickly find through working with the software. You will also find many keyboard shortcuts and these are summarised in the on-line Help.

1.2 Using this manual

This User Manual provides you with step-by-step instructions on how to perform an analysis using Mutation Analyser and gives you guidance in using the various tools necessary for evaluating the results. To successfully use Mutation Analyser, you are recommended to read this manual entirely.

Chapter 2	This tells you the minimum and recommended hardware requirements needed to run Mutation Analyser.
Chapter 3	This gives a brief description of how to install Mutation Analyser.
Chapter 4	This gives you an overview of the basic operation of Mutation Analyser.

Chapter 5 This introduces the various types of deviation and describes the process used to evaluate your samples.

Chapter 6 This describes the Configuration Editor module used for creating your own gene configurations.

Chapter 7 This describes some of the error messages that you may encounter and how to solve common hardware problems.

1.2.1 Examples used

In order to best describe the software functions and processes, examples from three gene configurations are used:

- Ret genomic DNA gene configuration, Exons 10, 11, 16 (*Courtesy of Dr Jorge A. Leon and co-workers at Quest Diagnostics - Nichols Institute, San Juan Capistrano, California, USA*)
- BRCA1 genomic DNA gene configuration, Exons 1 through 24 excepting the non-coding Exon 4 (*Courtesy of Prof. Hans K. Schackert and co-workers at the Department of Surgical Research, University of Dresden, Germany*)
- p53 cDNA gene configuration

1.2.2 Typographical conventions

Menu commands, the names of dialogue boxes and windows, the contents of boxes and windows and option buttons are written with a **bold Helvetica** typeface. Menu commands are written with first the menu name and then the specific command, separated by a colon. For example:

> Select **File:Save Analysis Item As** to display the **Save Analysis Item As** dialogue. Enter the file name and select the destination drive and directory. Click on **OK**.

This directs you to click on the **File** menu and select the command, **Save Analysis Item As**. A dialogue called **Save Analysis Item As** will be displayed in which you must select the name, drive and directory of the file. You then click on the button named **OK**.

Some menu commands also have shortcut keys on the keyboard, which are written within < > marks.

2. System requirements

2.1 Hardware

2.1.1 Computer

Computer	Compaq Pentium 133 MHz computer (minimum PC with 486 processor, clock speed 66 MHz)
RAM memory	32 Mb (minimum 16 Mb)
Disk memory	10 Mb free hard disk space for program
Graphics	SuperVGA graphics adapter, 256 colours (Minimum VGA graphics adapter)

2.1.2 Printers

The following printers have been tested for compatibility with Mutation Analyser:

HP DeskJet 660C, 560C

HP LaserJet 5 MP

2.2 Operating system

Mutation Analyser must be installed in the 32-bit operating environment, Windows 95 (see Chapter 3).

3. Software installation

One set of installation diskettes is supplied that contain:

- Mutation Analyser 1.00
- Configuration Editor 1.00
- Demo files

The various components are to be installed in the 32-bit Windows 95 (International Edition) operating environment.

3.1 Installation of Mutation Analyser

1. Insert the HASP® into a parallel port on your computer.

2. Turn on the computer. Microsoft Windows 95 should start automatically.

3. Insert Disc 1 into drive A.

4. Click on the taskbar Start menu and select **Run**.

5. Enter **a:setup** at the command prompt and click on **OK**.

6. Follow the instructions on the screen.

7. Remove the final program disc from drive A once the installation is complete.

8. The computer must be restarted after installation.

   Mutation Analyser and Configuration Editor should be available under the installation directory **Program files\Biotech\Mutation** (unless you changed the directory during installation) and are therefore located on the taskbar Start menu under Programs/Mutation Analyser 1.00.

3.2 Uninstalling Mutation Analyser

Should you want to uninstall Mutation Analyser:

1. Click on the taskbar Start menu and select **Settings:Control panel**. Double click on the **Add/Remove Programs** icon. Select Mutation Analyser 1.00 from the list of installed components and click on the **Add/Remove** button.

   Alternatively, select the **Uninstall** option from the taskbar Start menu under Programs/Mutation Analyser 1.00.

2. Follow the instructions on the screen.

   You may get a warning message that some of the driver files may be shared by other applications. If you are unsure about a specific file, do *not* select to uninstall the file. Remaining files should not affect your system.

4. Basic operation

This chapter introduces the basic operation of Mutation Analyser. You will learn how to create analysis items and then perform an analysis. You will be introduced to the various screen components before and after an analysis and become familiar with some of the important tools needed to evaluate your results. You will gain further instruction on printing reports and exporting results to database for use in other applications.

Chapter 5 later goes on to introduce the types of deviation that you are likely to encounter and guides you in the process of performing user assignments for the deviations.

4.1 Starting the application

To start the application:

1. Locate and select the Mutation Analyser program in the taskbar Start menu for Windows 95.

2. The start-up screen is displayed together with the **User Identification** dialogue requesting you to enter your **User ID**.

   Your user ID can be alphanumeric and contain up to 25 characters, and you must enter at least one character to proceed. The user name is a necessary component for the generation of reports.

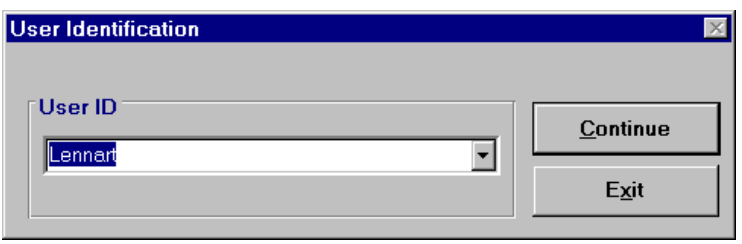


Figure 4-1. *User Identification panel.*

The last five entered user IDs are stored in the pull-down list, so it is possible to select your user name here if it is one of the last five.

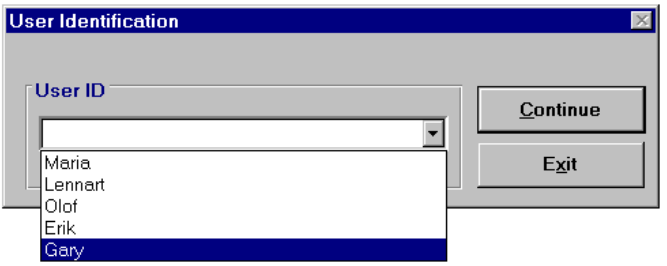


Figure 4-2. *Pull-down list in User Identification panel*

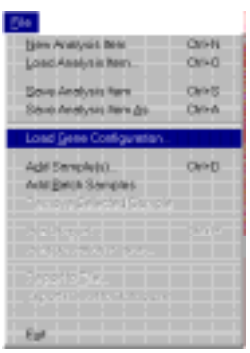
3. Click on the **Continue** button. By default, an empty **Analysis Item** window is displayed.

4.2 Creating and loading an analysis item

In order to make an analysis, you must create an *analysis item*, which comprises the following components:

- A gene configuration containing the wild type sequence against which sample sequences are aligned and analysed (see Chapter 6 for details on creating gene configurations)
- Sample sequences from one or more ALF/ALF*express* sequencing runs that can be aligned with the gene configuration sequence and be analysed
- Possible additional ALF/ALX samples for reference purposes in the analysis

4.2.1 Creating an analysis item



After you have started the application, an **Analysis Item** window is displayed called **Noname.itm**. The gene configuration last used is already loaded.

1. Select **File:Load Gene Configuration** to load a new gene configuration. The **Load Gene Configuration** dialogue is displayed. Select a configuration file and click on **Open**.

   For example, select the configuration file called, ***RETdemo.cfg****.*

   Gene configuration data is automatically displayed in the **Gene Configuration** field.

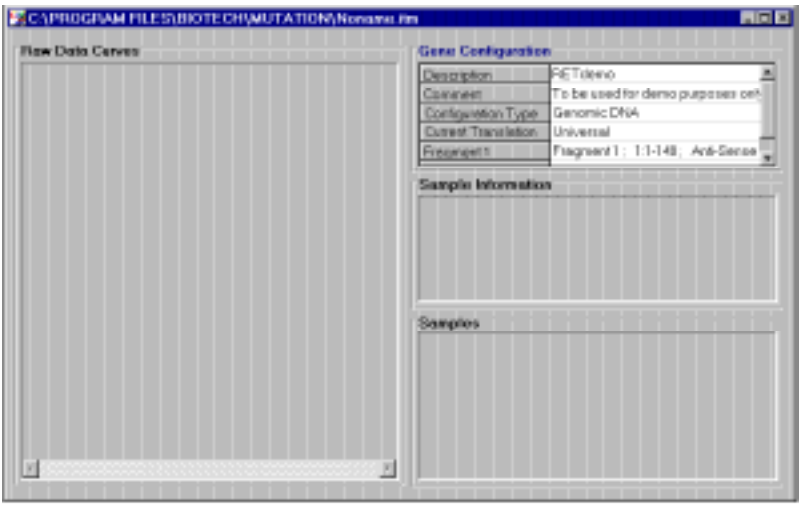


Figure 4-3. *Analysis Item window containing gene configuration data for Ret exons 10, 11 and 16.*

2. Select **File:Add Sample(s)**. In the **Open ALF file** dialogue, select the appropriate drive and directory containing your samples. Select an ALF/ALX file.

   For example, select ***RETdemo1.alx****.*

   Click on **Open**.

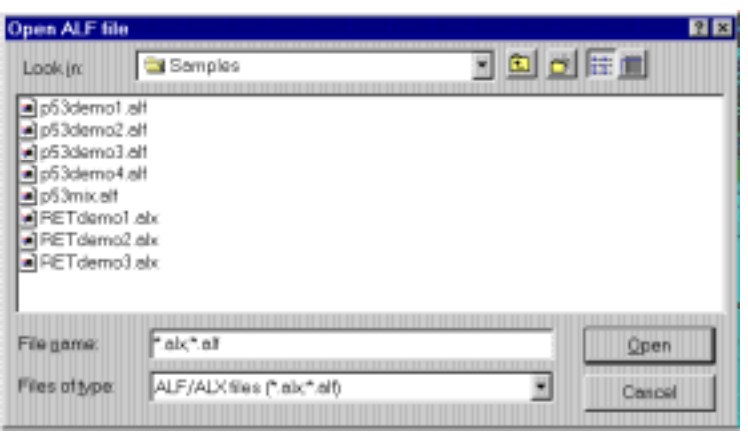


Figure 4-4. *Open ALF/ALX file panel for adding samples.*

3. The **ALF file** dialogue box is displayed. An ALF/ALX file contains one or more samples and information about each sample, previously added in the sequencing software.

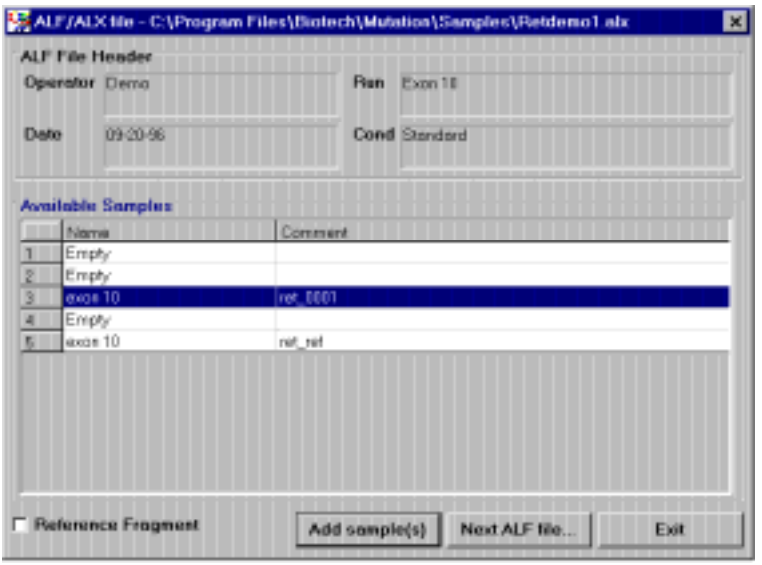


Figure 4-5. *ALF/ALX file dialogue with a Ret gene Exon 10 sample selected.*

To add an available sample, select a sample and click on the **Add sample(s)** button.

*For example, select the sample named **exon 10** with comment **ret_0001** and add it to the analysis item. This sample is known to contain base-specific deviations compared with the wild type sequence for exon 10 in the gene configuration.*

Repeat this process for other samples in the current ALF/ALX file. These may be reference samples that are not included in the analysis process but instead used for comparative purposes of the raw data curves. Check the **Reference Fragment** option to include the sample for reference purposes only.

*For example, select the sample named **exon 10** with comment **ret_ref**, check the **Reference Fragment** box and add the sample to the analysis item. This sample is the wild type sequence for exon 10 as used in the gene configuration.*

4. To add a sample from another ALF/ALX file, click on the **Next ALF/ALX file** button. This will return you to the **Open ALF file** dialogue in the same drive and directory from where you made your previous ALF/ALX file selection. Select a new ALF/ALX file.

   *For example, select the file called **RETdemo2.alx** and add the mutant sample **exon 11, ret_0001** and the reference fragment **exon 11, ret_ref**.*

5. Repeat steps 3 and 4 until you have obtained all of the samples that you want to include in your new analysis item.

   *For example, include the final sample for a Ret gene analysis by selecting the file called **RETdemo3.alx** and adding the mutant sample **exon 16, ret_0001** and the reference fragment **exon 16, ret_ref**.*

6. When you have added all relevant samples, click on the **Exit** button. The **Analysis Item** window will be again displayed containing all information about the selected samples.

   For example, you should now have six samples, including a sample sequence and a reference sample sequence for each of the Ret gene exons 10,11, 16.

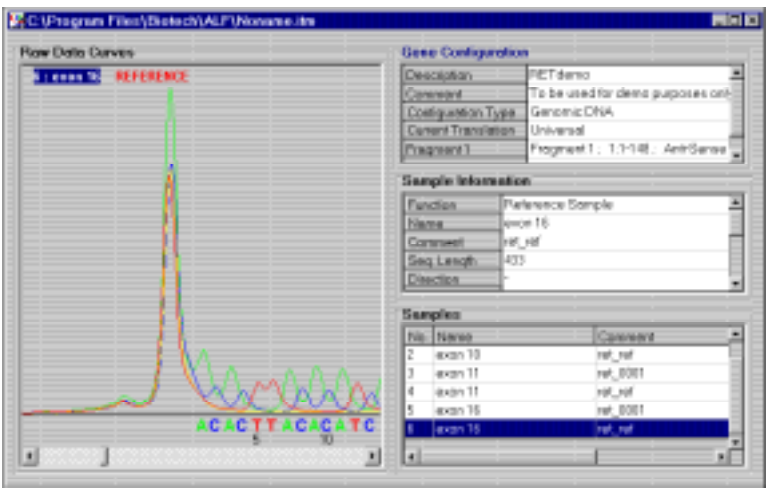


Figure 4-6. *Analysis Item window containing sample information.*

The new fields include:

Sample Information — Information about the sample selected in the **Samples** field is shown here. This information was previously entered using the sequencing software.

Samples — The names of the samples you included are listed here, including any reference samples. Selection of a sample affects what is displayed in other fields.

Raw Data Curves — This shows the raw data curves produced by the automatic sequencer and processed using the sequencing software (see Section 4.3.3).

4.2.2 Adding batch samples

If you have many ALF/ALX files and you want to locate a specific series of samples within the files and add them to an analysis item, you can search for these samples and add them as a batch.

1. Start the application and load a gene configuration into the analysis item as described in step 1 of Section 4.2.1.

2. Select **File:Add Batch Samples**. The **Add Batch Samples** dialogue is displayed.

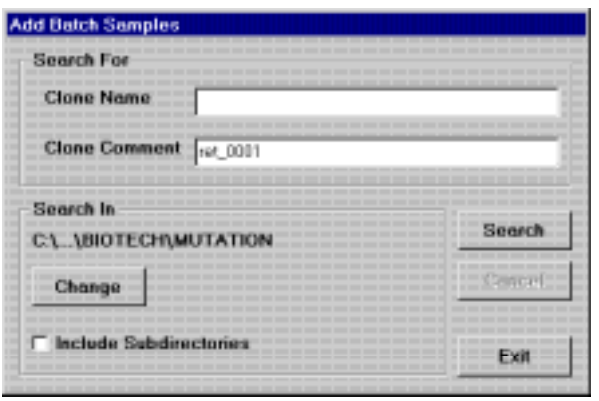


Figure 4-7. *Batch Load Samples dialogue. The clone comment 'ret_0001' has been entered as the search string.*

3. Enter a search string based on the **Clone Name** and/or **Clone Comment**.

   For example, to find all of the sample sequences corresponding to exons 10, 11, 16 in the Ret gene configuration, enter ***ret_0001*** *in the* ***Clone Comment*** *field.*

4. You can direct the search to take place in any directory. If you want to change the currently selected directory, click on the **Change** button and select the new search directory. Check the **Include Subdirectories** option if you the want the search extended to subdirectories of the search directory.

   For example, the RET demo files are located in the Samples subdirectory under the Mutation directory.

5. Click on **Search**. The results of the search are displayed in the **ALF/ALX file** dialogue.

   *For example, all samples containing the **Clone Comment** ret_0001 will be located and displayed.*

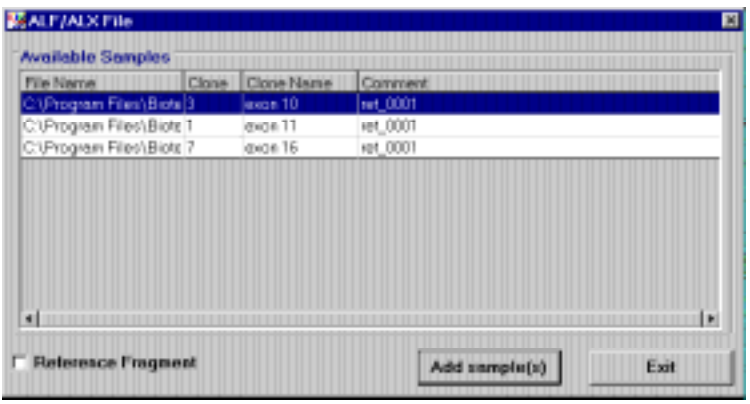


Figure 4-8. *ALF/ALX file dialogue after a search for batch clones. In this example, clones for the Ret gene exons 10,11,16 were located corresponding to the Clone Comment search string 'ret_0001'.*

In the **Available Samples** list you can see the located **File Names** and pathway, the number of the **Clone** in the sample file, the **Clone Name** and **Comment**.

6 To add an available sample, select it and click on the **Add sample(s)** button. Repeat this process for other samples.

Alternatively, you can select several clones at once by clicking on the first sample and then pressing the <Shift> key while you click on the last sample. All samples between the two will be automatically selected. Click on **Add** to include these in the analysis item.

7. When you have added all relevant samples from the current search, click on **Exit**. The **Add Batch Samples** dialogue is shown allowing you to make another search following steps 2 to 6.

   *For example, you can add the reference fragments for exons 10,11, 16 by entering **ret_ref** in the **Clone Comment** field and starting the search. These samples will be located and can be added to the analysis item. Remember to check the **Fragment** option before you add the samples.*

8. When you have added all of the samples, click on **Exit**. The **Analysis Item** window will be again displayed containing all information about the selected samples.

4.2.3 Removing a selected sample

If you want to remove a sample from the analysis item, select it in the **Samples** field and then select **File:Remove Selected Sample**.

4.2.4 Saving an analysis item

You are recommended to directly save a newly created analysis item. Select **File:Save Analysis Item**. In the **Save Analysis Item As** dialogue box, select the destination drive and directory and enter a file name for the new item. Click on **Save As**.

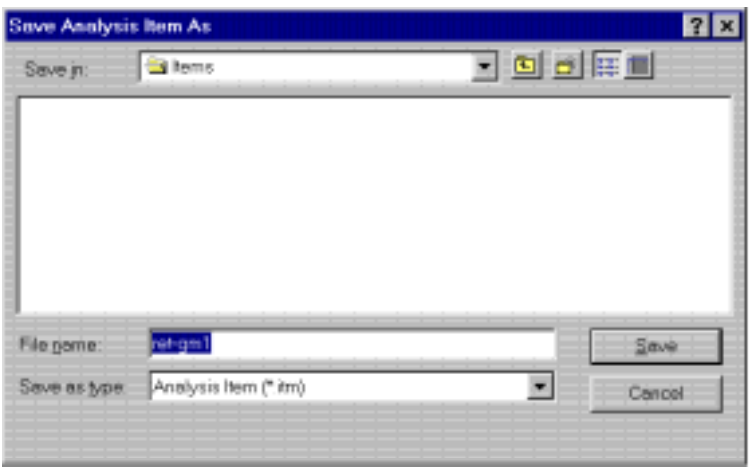


Figure 4-9. *Save Analysis Item As dialogue box.*

At any point during an analysis you can save the item in its current state by selecting **File:Save Analysis Item** or save a copy of the analysis item under a new file name using **File:Save Analysis Item As**.

4.2.5 Loading an analysis item

To load an analysis item, select **File:Load Analysis Item**. Select an analysis item from the appropriate drive and directory and then click on **Open**. The analysis item will be loaded in the state it was last saved.

Alternatively, you can select an analysis item listed in the **File** menu if it was one of the last four opened analysis items.

4.3 Performing an analysis

This section describes how to perform an analysis and also select the analysis options. The various screen components displayed following an analysis are introduced in sections 4.4 to 4.8. Interpretation of analysis results and user assignment are presented in Chapter 5.

4.3.1 Deviation calculation options

Prior to performing an analysis you should select the deviation calculation options.

1. Select **Options:Deviation Calculation Options**. A dialogue box is displayed.

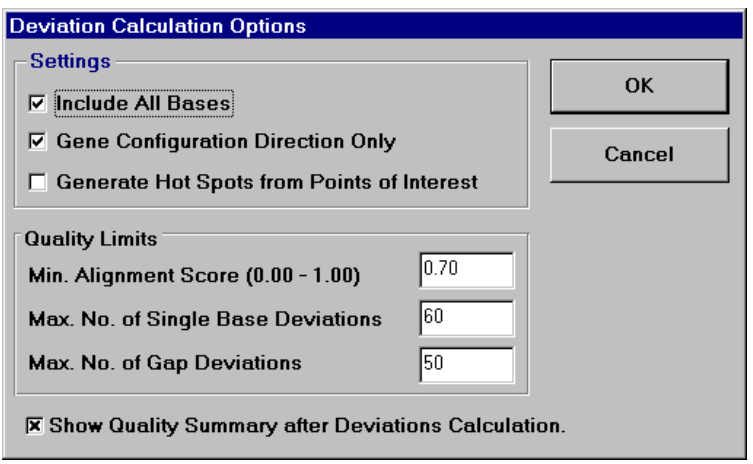


Figure 4-10. *Deviation Calculation Options dialogue box.*

2. Check the appropriate **Settings** options.

Include all bases — Towards the end of a sequencing run, the quality of the data may deteriorate. By default, this data at the end of the run will be included in the analysis, as denoted by the checked option. By unchecking this option, this data will be excluded from the analysis.

Gene Configuration Direction Only — If this option is checked and you have defined the sequencing directions of the individual sequencer fragments in the gene configuration, the analysis will align each sample sequence in the direction specified in the gene configuration.

Unchecking this option aligns the sample sequence in both directions with respect to the gene configuration. This latter process takes a longer time. It is recommended that you use the former method and that you define two separate sequencer fragments if both strands are being analysed.

Generate Hotspots from Points of Interest — If this option is checked, any points of interest that you defined in the gene configuration (see Section 6.4.1) will be included in **Deviations Table** as hotspots following analysis (see Section 5.2). This is particularly useful for labelling known polymorphisms and/or known mutation sites. Hotspots can also be user assigned (see Section 4.4.2). No hotspots will be generated if sample data is missing from these base positions.

3. Change the **Quality Limits** if required.

Min. Alignment Score (0.00 - 1.00) — The minimum acceptable alignment score between the sample sequence and the gene configuration. The score is also used for finding the best fitting sequencer fragment. Note that samples containing mixed insertions/deletions may give rise to a poor score.

Max No. of Single Base Deviations — The maximum number of single base deviations (see Section 5.2) allowed for a sample sequence when aligned to the gene configuration.

Max No. of Gap Deviations — The maximum number of explicit insertion/deletion deviations (see Section 5.3) for a sample sequence when aligned to the gene configuration.

4. By default, a quality summary is displayed after deviation calculation. **Uncheck the Show Quality Summary after…** option if you do not want the summary displayed.

5. Click on **OK** when you have selected the appropriate options.

4.3.2 Associating a sample sequence to a fragment

In some circumstances you may want to associate a sample sequence to a specific sequencer fragment for alignment. This means that the sample sequence will not be checked against other sequencer fragments during the analysis for a best alignment.

1. In the **Samples** field of the **Analysis Item** window, select a sample.

2. Select **Analysis:Associate Fragment with Sample**. The **Associate Fragment with Sample** dialogue is displayed.

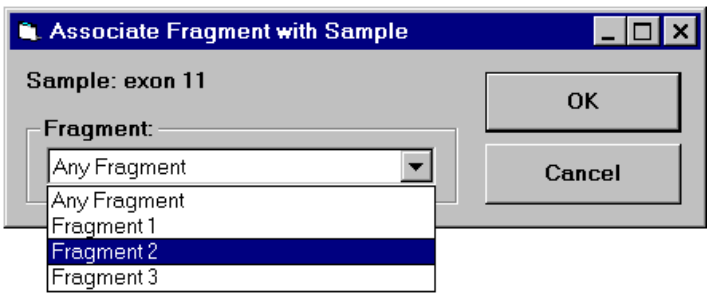


Figure 4-11. *Associate Fragment with Sample dialogue.*

3. From the pull-down list, select the specific sequencer fragment contained within gene configuration with which the selected sample will be aligned in the analysis.

4. Click on **OK** and repeat for other samples.

5. Perform a deviation calculation (see Section 4.3.3).

Note: An association can be made after the analysis has been performed. The sample corresponding to the selected deviation will be affected.

4.3.3 Calculating the deviations and quality summary

1. Select **Analysis:Calculate Deviations**. The samples included in the analysis item are automatically aligned against the gene configuration.

2. The **Quality Summary** dialogue is automatically displayed if you did not deselect this option in the **Deviation Calculation Options** dialogue (see Section 4.3.1).

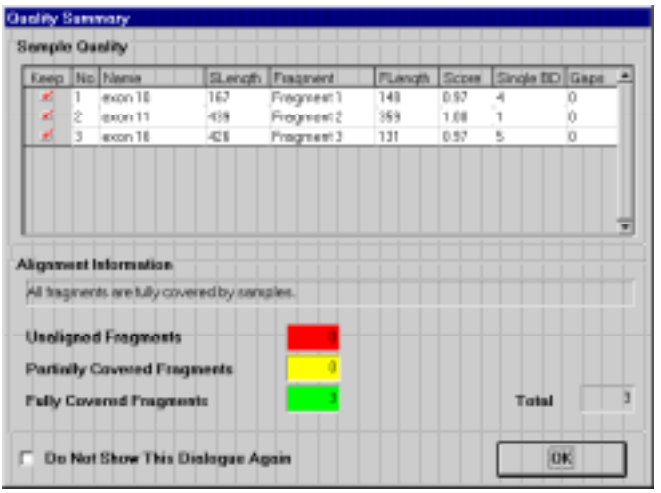


Figure 4-12. *Quality Summary dialogue. This example shows the quality summary for the alignment of sample sequences corresponding to the Ret gene Exons 10, 11 and 16.*

Sample Quality information is presented, including:

Name — The name of the sample.

SLength — The number of bases in the sample sequence.

Fragment — The name of the sequencer fragment in the gene configuration to which the sample sequence is aligned.

FLength — The number of bases in the gene configuration sequencer fragment.

Score — The actual alignment score between the sample sequence and the gene configuration. Higher values indicate better alignment.

Single BD — The actual number of single base deviations (see Section 5.2) in that sample when aligned to the sequencer fragment.

Gaps — The actual number of explicit insertion/deletion deviations in that sample (see Section 5.3) when aligned to the best fitting sequencer fragment.

By default, all items fulfilling the limit criteria are checked to **Keep** the sample in the analysis. If you do not want to keep a sample in the analysis, uncheck it and it will be removed from the analysis item. Items not fulfilling the limit criteria will be automatically removed from the analysis item unless you explicitly check them, e.g. for samples containing mixture insertion/deletion deviations displaying a poor alignment quality.

Alignment Information is also shown:

Unaligned Fragments	Colour-coded red, this displays the number of sequencer fragments that were not assigned a sample following the analysis.
Partially Aligned Fragments	Colour-coded yellow, this displays the number of sequencer fragments only partly aligned to a sample sequence.
Fully Aligned Fragments	Colour-coded green, this displays the number of sequencer fragments that were fully aligned with sample data.

3. Click on **OK**.

The results of the deviation calculation are displayed in several fields, including the **Deviations Table**, **Alignment** field, **Raw Data Curves** field and **Assignment** bar (see Sections 4.4 to 4.8).

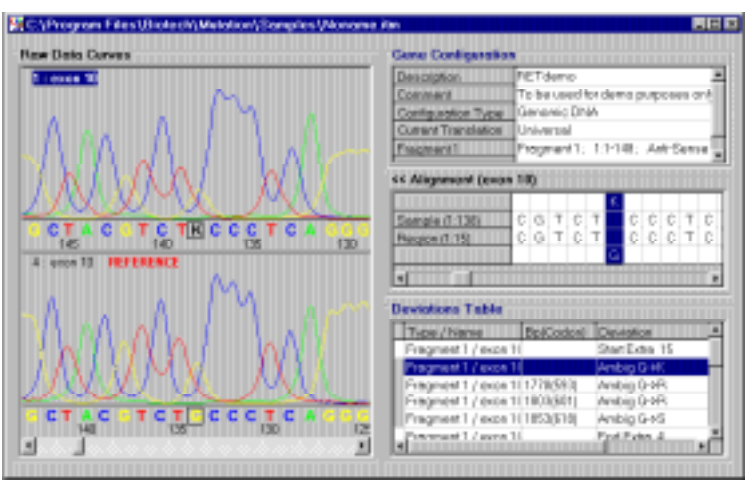


Figure 4-13. *Mutation Analyser workspace following an analysis of a Ret gene analysis item.*

4.4 Deviations Table

Following analysis, the **Deviations Table** replaces the **Samples** field (see Section 4.2.1). The **Deviations Table** contains detected alignment deviations for each sample with respect to the gene configuration. For details about the different types of deviation encountered here, see Chapter 5.

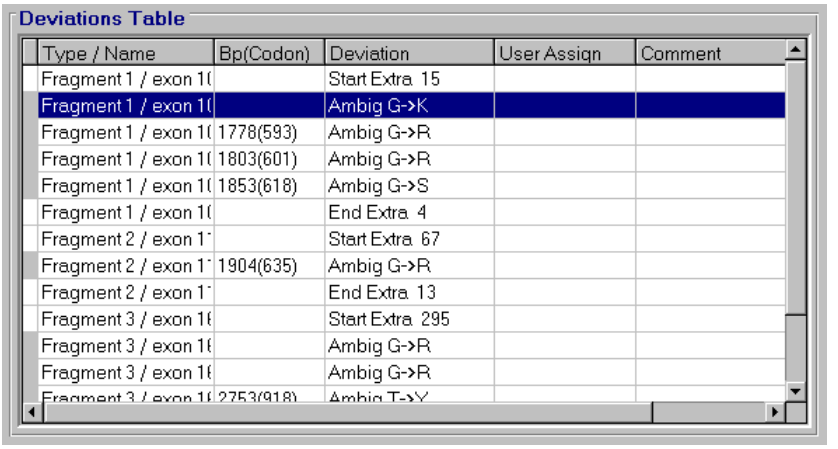


Figure 4-14. *Deviations Table in the Analysis Item window following analysis of Ret gene demo samples.*

The information displayed in the table is dependent on the columns selected in the **Deviations Table Options** dialogue (see Section 4.4.1).

The width of each column can be adjusted by dragging the border of the column heads.

It is also possible to increase the display size of the **Deviations Table** by hiding the **Gene Configuration** field. Select **View:Hide Gene Configuration**.

4.4.1 Display options for the Deviations Table

The information displayed in the **Deviations Table** is dependent on the columns selected for viewing.

1. Select **Options:Deviations Table Options**. The **Deviations Table Options** dialogue is displayed.

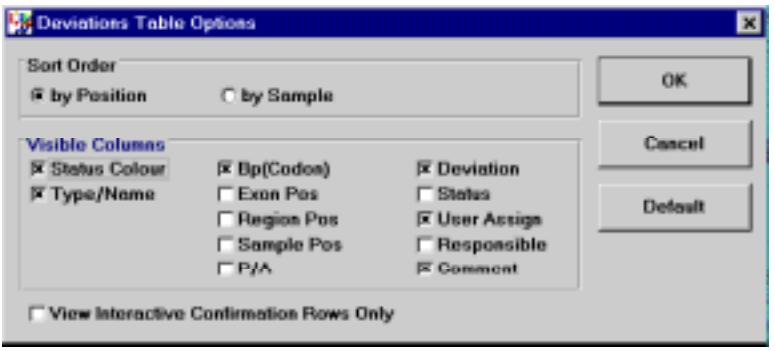


Figure 4-15.*Deviations Table Options dialogue.*

2. Select whether the viewed deviations are sorted **by Position** or **by Sample**.

3. Select the **Visible Columns** in the table by checking the appropriate options:

Status Colour
This option displays a colour-coded status bar on the left side of each deviation in the list.

A grey bar indicates a deviation that is **Pending** assignment by the user, a red bar indicates an **Approved** deviation assignment, a black bar indicates an approved **Wild** type assignment, and a yellow bar indicates the user has assigned the particular deviation as **Ignored**.

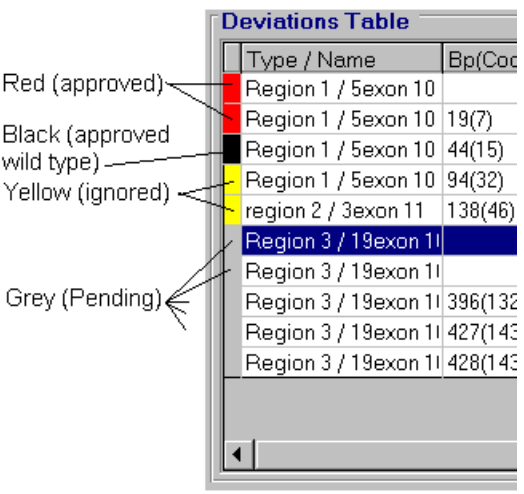


Figure 4-16. *Colour-coded status bars for each of the deviations.*

Some deviations have a white status bar. These deviations are not subject to user assignment as defined in the **Interactive Assignment Options** dialogue (see Section 4.4.2). These can be hidden from the display by checking the **View Interactive Confirmation Rows Only** option (see step 4., below).

Type/Name
This option displays the fragment type as defined in the gene configuration and the name of the sample fragment.

Bp(Codon)
This option displays the base position and codon number (in parentheses) of the deviation. This information is only available if coding sequence information has been included in the gene configuration.

Exon Pos
This option displays the exon number and the base position within the exon of the deviation.

Region position
This option displays the region number and base position within the region for the deviation according to the definitions in the gene configuration.

Sample position
This option displays the sample number and base position of the deviation with respect to its position in the sample sequence.

P/A
A 'P' and/or an 'A' in this column indicates whether the deviation coincides with a point of interest and/or falls within an area of interest.

Deviation
This options displays information about the type of deviation.

Status
This option displays the user assignment status of the deviation, either **Pending,** an **Approved** user assignment, or **Ignored** (see Chapter 5). The **Status Colour** option also gives the same indication of status.

User Assign
This option displays specific details about the assignment made by the user.

Responsible
This option displays the name of the user last responsible for making an assignment for that deviation.

Comment
This option displays comments added during user assignment in the **Assignment** bar (see Section 4.7).

4. By default, all deviation positions are displayed. Some deviations are not subject to user assignment, as defined in the **Interactive Assignment Options** dialogue (see Section 4.4.2). Such deviations have a white status bar if the **Status Colour** option is selected. These can be hidden from the display by checking the **View Interactive Confirmation Rows Only** option.

5. Click on **Default** if you want to return to the original table display settings.

6. Click on **OK** to accept you options selection.

4.4.2 Displaying deviations for interactive assignment

The contents displayed in the **Deviations Table** can be further filtered to display only the types of deviation that are to be included for user assignment.

1. Click on **Options:Interactive Assignment Options** to display the **Interactive Assignment Options** dialogue. Check the deviation types to be included in the interactive assignment process.

Note: If you selected to generate hotspots (see Section 4.3.1) from points of interest defined in the gene configuration (see Section 6.4.1), check the **Hotspot** deviation option. A hotspot deviation will be entered into the **Deviations Table** for all points of interest and must be user assigned as for single base deviations (see Section 5.2).

If you checked the **View Interactive Confirmation Rows Only** option in the **Deviations Table Options** dialogue (see Section 4.4.1), only those deviations selected for interactive assignment will be displayed in the **Deviations Table**.

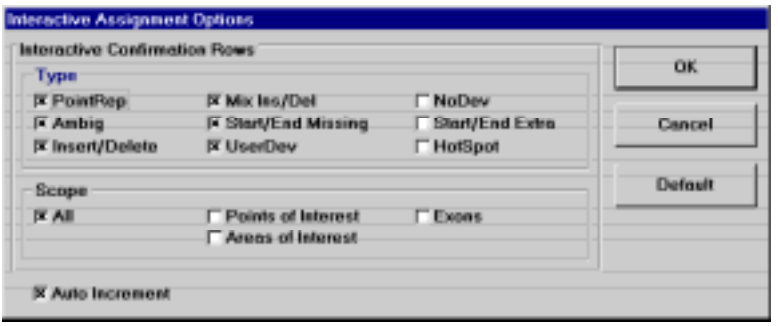


Figure 4-17.*Interactive Confirmation Options dialogue.*

2. Select the **Scope** of the deviations to be included in the interactive assignment process. You can select to include **All** deviations, or those deviations occurring at **Points of Interest**, within **Areas of Interest** and within **Exons**.

3. By default, the **Auto Increment** option is checked so that when a user has made an assignment for a deviation, Mutation Analyser automatically goes to the next deviation in the table. Deselect this option if you want to manually select the order of deviations for assignment.

4. Click on **OK** to return to the **Analysis Item** window. The **Deviations Table** will display the deviations based on criteria selected in the **Interactive Assignment Options** and **Deviations Table Options** (see Section 4.5.1) dialogues.

4.4.3 Viewing deviation information

It is possible to view extensive information about a deviation:

1. Select the deviation in the **Deviations Table**.

2. Select **View:Deviation Information** or **Show Deviation Information** from the right mouse button menu. The **Deviation Information** dialogue is displayed.

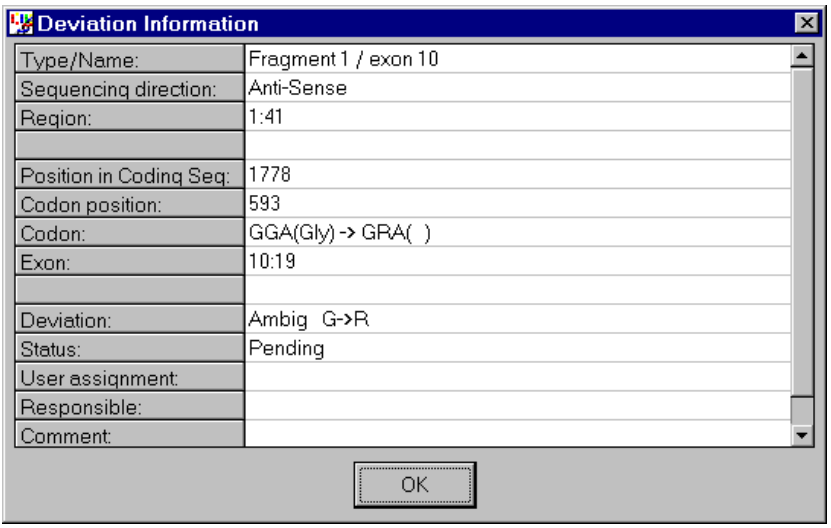


Figure 4-18. *Deviation Information dialogue.*

4.5 Alignment field

The **Alignment** field displays the sequence alignment data for the sample corresponding to the deviation selected in the **Deviations Table**. For details about the different types of deviation encountered here, see Chapter 5.

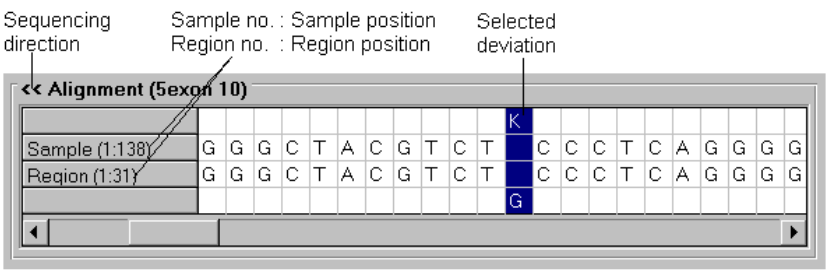


Figure 4-19. *Alignment field in the Analysis Item window.*

4.5.1 Sequencing direction

The sequencing direction of the specific fragment is indicated by **>>** for the Sense direction and by **<<** for the Anti-Sense direction. All sequence data is presented in the Sense direction.

4.5.2 Alignment deviations

Bases in the sample sequence are aligned with the corresponding bases in the gene configuration. Where there is an alignment deviation in the sample sequence with the gene configuration, the mismatched bases for the sample and gene configuration sequence are separated up and down respectively.

By default, the alignment deviation last selected in the **Deviations Table** is highlighted in the **Alignment** field. The sample number and sample position for the deviation is displayed together with the corresponding region number and region position in the gene configuration.

4.5.3 Viewing sample or alignment information

You can switch the view from the **Alignment** field and the **Sample Information** field by selecting **View:Show Sample Information**.

To return to the **Alignment** field, either select **View:Show Alignment Information** or select a deviation in the **Deviations Table**.

4.6 Raw Data Curves field

The **Raw Data Curves** field displays curve data for the currently selected sample. As for the alignment data in the **Alignment** field, all data is presented in the Sense direction.

There are four colour-coded curves, one for each of the four bases A,C,G,T. Each peak in a curve corresponds to a base detected for that position. A correspondingly coloured letter for the specific base is shown underneath a detected peak.

The displayed data is centred on the sample sequence base currently selected in the **Deviations Table**. The base corresponding to the selected base has a black box surrounding it.

A vertical cursor line can be moved in the field by holding down the left mouse button.

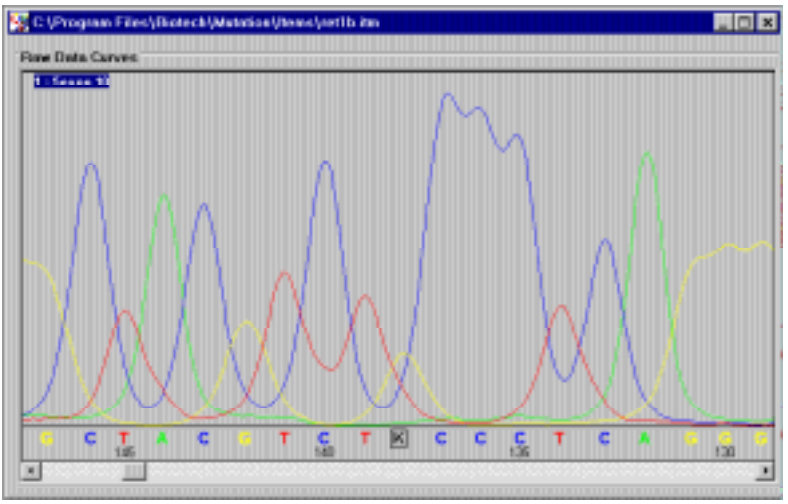


Figure 4-20. *Raw Data Curves field in the Analysis Item window.*

4.6.1 Changing the display window size

To view only the **Raw Data Curves** field, i.e. make the field bigger, select **View:Full Raw Data View** or **Full Raw Data View** from the right mouse button menu. To reverse the function, i.e. return the field to its normal size, select the checked menu option a second time.

Alternatively, make the **Raw Data Curves** field larger or smaller by dragging the right field border while holding down the left mouse button.

4.6.2 The zoom function

You can increase or decrease the magnification of the data within the field by using the Zoom function. For successive zoom in capability, repeatedly select **View:Zoom In** or **Zoom In** from the right mouse button menu, or use the <F7> key on the keyboard. Conversely, select **View:Zoom Out** or **Zoom Out** from the right mouse button menu, or use the <F8> key for successive reduction of the magnification.

It is also possible to view a fully zoomed out display of all the data by selecting **View:Full Zoom Out** or **Full Zoom Out** from the right mouse button menu. The full size of the of the **Raw Data Curves** field display is automatically shown, i.e. automatic selection of **Full Raw Data View**, and all raw data is shown with minimum magnification. The magnification can be increased using the standard zoom in function and the size of the field display returned to normal by deselecting the **Full Raw Data View** option. This on/off toggle function can be similarly achieved using the <F6> key on the keyboard.

4.6.3 Curve shift correction

By default, the **Options:Rawdata Shift Correction** function is active and enables the display of shifts in peak positioning applied by the sequencing software during the processing of the raw data. This allows you to more easily distinguish the order of the peaks as they occur in the sequence. Deselection of this function removes the automatic positioning and allows you to view the unaltered raw data directly. A check mark beside the menu option indicates that the function is active.

4.6.4 Curve amplitude scales

By default, the four curves are displayed on the same amplitude scale as the curve containing the largest peak. The peak heights of all the curves are then correctly proportioned with respect to one another.

Alternatively, it is possible to view the curves with separate amplitude scales so that the displayed amplitude scale for each curve is adjusted to best fit the window displayed. This may be useful in discerning the peaks in a curve which had a relatively low detection signal during sequencing.

Attention
The adjustments for individual scales are made *only for the peaks present in the current display* within the **Raw Data Curves** field. Changing the magnification of the view will alter the number of visible peaks and hence the information used to calculate the separate scales.

Selection of separate scales can be misleading when, for example, trying to establish if a peak is a true peak or part of the background noise.

Separate amplitude scales can be toggled on and off by selecting **Options:Separate Amplitude Scales**. A check mark beside the menu option indicates that the function is active.

4.6.5 Multiple raw data curve sets

If you have included reference fragments in your analysis item, either defined within the gene configuration (see Section 6.4.1) or as separately added samples (see Section 4.2.1), the curve data for these will be displayed together with the sample sequence curve data for the selected deviation.

Some alignment deviations will occur in the overlap regions between sequencing fragments. If you select such a deviation in the **Deviations Table**, two or more sets of raw data curves will be displayed in the **Raw Data Curves** field.

In all cases, the sample curve data for the selected deviation is identified by the blue box containing the sample name in white letters.

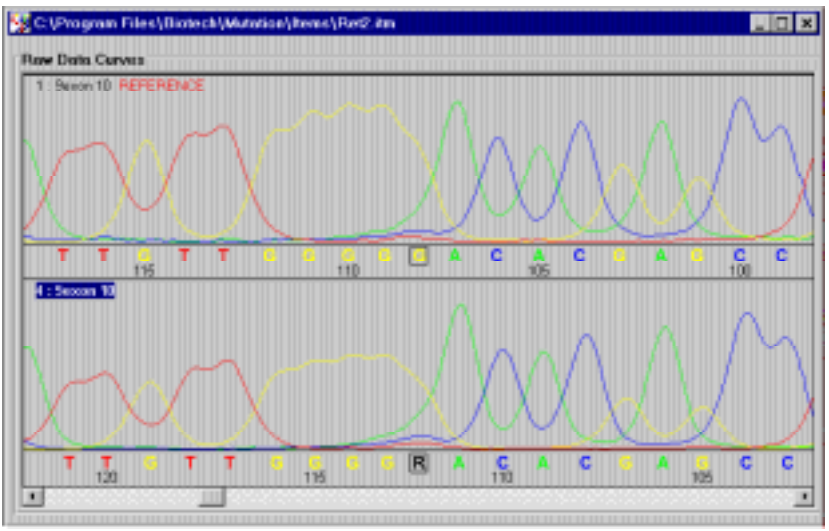


Figure 4-21. *Raw data curves centred on a selected deviation in the sample sequence (bottom) and a corresponding reference fragment defined in the gene configuration (top).*

Viewing different sets of curve data
To select the types of curve data that you want to be displayed:

1. Click on **View:Related Raw Data** or select **Related Raw Data** from the right mouse button menu. The **Related Raw Data** dialogue is displayed.

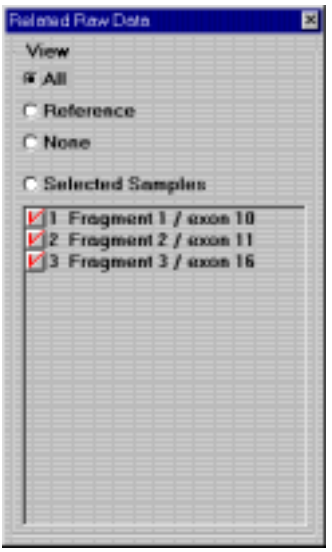


Figure 4-22. *Related Raw Data dialogue.*

2. Select the appropriate option. The **Raw Data Curves** field is automatically updated as you select an option.

All	This displays all sets of raw data curves corresponding to the selected deviation in the sample sequence, including reference fragments and overlapping fragments.
Reference	This displays the sample sequence for the selected deviation in addition to any reference fragments.
None	This displays only the sample sequence for the selected deviation.
Selected Samples	This displays only those curves checked in the custom list below *and* only if they correspond to the selected deviation in the sample sequence.

3. Close the dialogue when you no longer need it.

Alternatively, you can remove a set of curves from the display by double clicking on it with the left mouse button. The **Related Raw Data**

dialogue will be automatically displayed in which the **Selected Samples** option is active for the remaining sets of raw data curves. Select other options to see other sets of curves.

4.6.6 Separate scrollbars

By default, the **Raw Data Curves** field has only one horizontal scrollbar. This means that all data will be similarly affected with the scrollbar if multiple sets of data are displayed. By selecting **Options: Separate Raw Data Scrollbars**, each set of displayed data will obtain its own scrollbar.

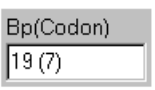
4.7 Assignment bar

The **Assignment** bar (**Single Base Assignment**, **Insertion Assignment**, **Deletion Assignment** or **Mixture Assignment**) allows you to make a user assignment for each of the deviations.

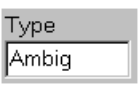
Figure 4-23. *Assignment bar*

User assignment is presented in Chapter 5, although a brief description of the **Assignment** bar components is given here:

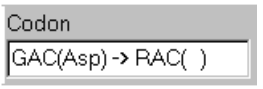
The checked boxes correspond to the base components detected for the selected deviation. You can check any selection of base components that you consider should be assigned to the deviation.



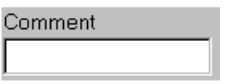
This displays the base position and codon number (in parentheses) of the selected deviation with respect to the coding information in the gene configuration.



This displays the type of deviation selected. See Chapter 5 for more information about the different types of deviation.



This compares the codon and corresponding amino acid in the gene configuration (left) with the codon containing the selected deviation (left). In the latter case, the encoded amino acid is only shown if all bases in the codon are resolved.



This allows you to enter a comment about the selected deviation that will be displayed in the **Deviations Table** following user assignment.

These buttons allow you to make an assignment for the selected deviation. Your can **Approve** an assignment, make an assignment to **Ignore** a deviation, or automatically convert the deviation back to the wild type and approve it by selecting **Wild**. The buttons shown are dependent on the type of **Assignment** bar.

4.8 Analysis overview

You can select to display an analysis overview, which shows the whole gene configuration together with sample fragments and detected deviations.

Select **View:Analysis Overview**. The **Analysis Overview** window is displayed, containing the following:

- The gene configuration is displayed for all of the regions together with the amino acid coding sequence. Other components of the gene configuration are displayed based on the selected overview options (see Section 4.8.4).

- Sample fragments are shown in their aligned position relative to the gene configuration. Deviations are marked on the sample fragments with a vertical cursor line passing through the currently selected deviation. Click on a deviation mark to select it within the window. A legend helps you to identify the types of deviation marks.

- Reference fragments will be displayed if included in the analysis item.

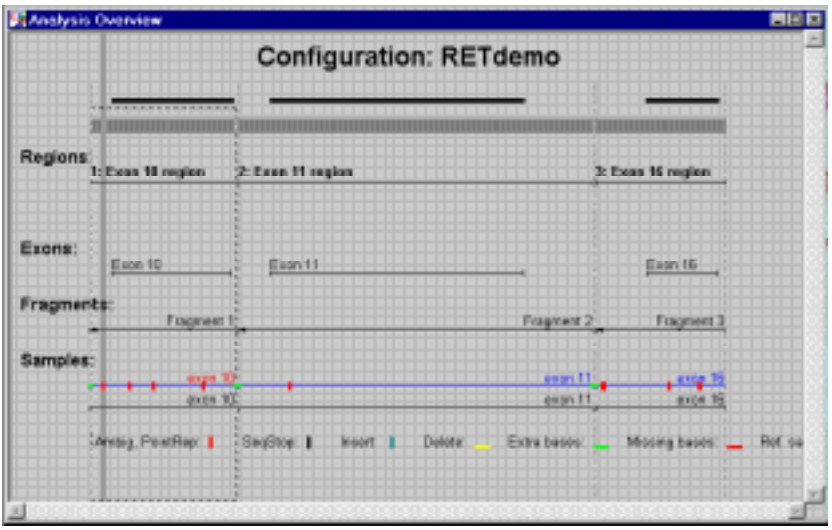


Figure 4-24. *Analysis Overview window showing the analysis results based on the Ret gene configuration.*

4.8.1 Using the zoom function within Overview mode

You can increase or decrease the magnification of the data in the window by using the Zoom function. For successive zoom in capability, repeatedly select **View:Zoom In** or **Zoom In** from the right mouse button menu, or use the <F7> key on the keyboard. Conversely, select **View:Zoom Out** or **Zoom Out** from the right mouse button menu, or use the <F8> key for successive reduction of the magnification.

The magnified area is centred on the selected deviation and the individual bases in the gene configuration sequence and individual amino acids for the coding sequence can be viewed.

Select **View:Full Zoom Out** to return to the lowest magnification to view the whole gene configuration.

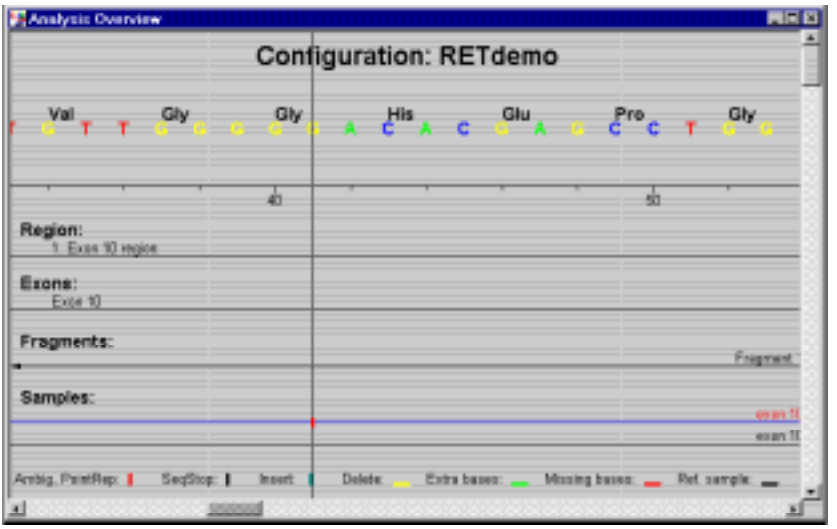


Figure 4-25. *Analysis Overview window with maximum magnification.*

4.8.2 Selecting a deviation for evaluation

If you want to select a deviation in the **Analysis Overview** window for evaluation purposes:

1. Click on the deviation mark that you want to evaluate. The cursor line will now pass through the selected mark.

2. Keep the mouse cursor on the selected deviation and click on the right mouse button to display the associated menu.

3. Select **Analysis Window** from the menu options. The **Analysis Item** window is displayed and the deviation selected in the **Overview** window is active.

4.8.3 Adjusting sequencer fragments

You can adjust the extent of the sequencer fragments used for alignment against the sample sequences.

1. Click on the relevant sequencer fragment in the **Overview** window.

2. Select **Edit:Adjust Fragment Extent** to display the **Adjust Fragment Extent** dialogue.

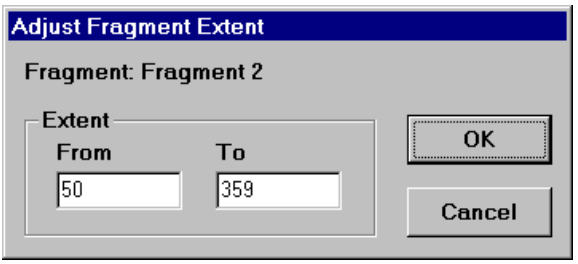


Figure 4-26. *Adjust Fragment Extent dialogue.*

3. Enter the new extent of the sequencer fragment. Click on **OK**.

   The previous analysis results are consequently removed and you must perform the analysis again to accommodate the changes you made to the sequencer fragment.

4.8.4 Changing the overview display options

The components displayed in the **Analysis Overview** window can be selected as follows:

1. Select **Options:Overview Options**. The **Overview Options** dialogue is displayed.

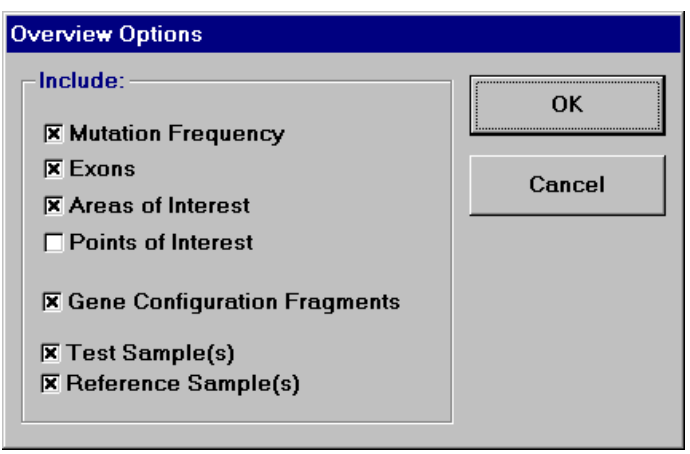


Figure 4-27. *Overview Options dialogue box.*

2. Select the components to be included in the **Analysis Overview** window.

3. Click on **OK**.

4.9 Adding comments about the analysis

Comments about the current analysis can be written in a separate window which are then saved together with the specific analysis item:

1. Select **Edit:Comment Analysis**. The **Comments on Analysis** window will be displayed.

2. Type in new text relevant to the current analysis or update the existing text if you have already saved this in an earlier session with the current analysis item.

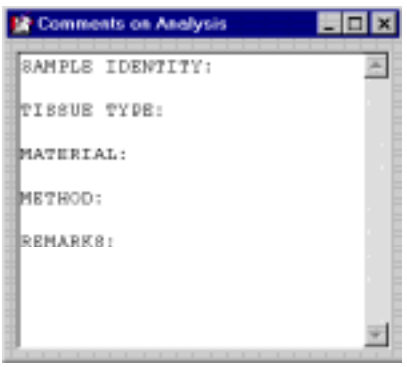


Figure 4-28. *Comments on Analysis window.*

3. Iconise or close the window while you are not using it. Maximise the icon to again view the window or select **Window:Comments on Analysis**.

Note: You can change the default template to display text of your choice when the **Comments on Analysis** window is opened for the first time for each analysis item. Locate the file called **Template** in the directory where you have installed Mutation Analyser (e.g. in C:\Program files\Biotech\Mutation). Double click on the file to open it in Windows 95 NotePad. Modify the template text and save the changes.

4.10 Save settings for options

If you want to save the settings you selected for the various option dialogues, select **Options:Save Settings**. The saved settings are valid until the **Save Settings** command is used the next time for new settings. You can save the settings made for:

- **Deviation Calculation Options**
- **Interactive Assignment Options**

- **Deviations Table Options**
- **Report Options**
- **Export Report Options**

4.11 Finding a sequence

You can use the Find function to locate a specific sequence in the Sequencer Fragment or the currently selected sample.

1. Select a deviation in the **Deviations Table** and select **Edit:Find**. The **Find** dialogue is displayed.

2. Select the **Scope** of the find, either in the currently selected sample or in the corresponding sequencer fragment of the gene configuration.

3. Enter a search string, either by directly entering a sequence into the **Search String** field, or by loading a previously saved string by clicking on the **Load** button.

   Any search string can be saved for future use by clicking on the **Save As** button and giving an appropriate name.

2. Define any other search criteria that you want:

Complement Find	Check this box if you want to search the complementary sequence using complementary bases to those in the **Search String** field.
Backward search	This looks for the search string beginning at the end of the sequence.

3. Click on **Find First** and if the string is present in the sequence, its first occurrence will be detected and the **Start Position** shown. The **Start Position** always shows the detected position within the configuration sequence with respect to the first base in the search string. Click on **Find Next** to locate the next occurrence of the string. The **Alignment** field and **Raw Data Curves** field are automatically centred on the beginning base of the located string.

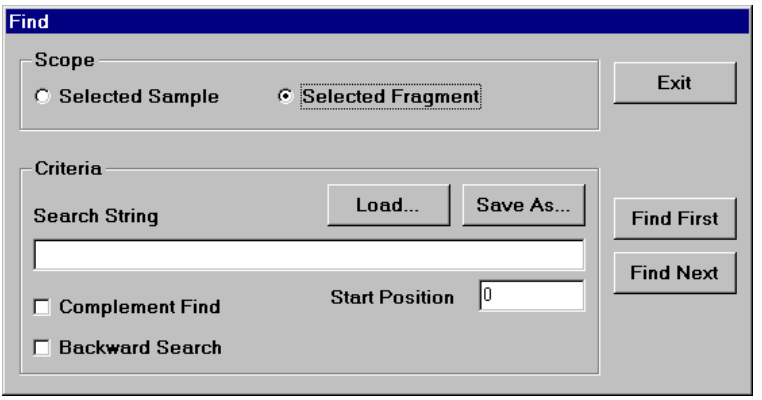


Figure 4-29. *Find dialogue*

4. Click on **Exit** when the appropriate **Position** value is obtained.

4.12 Adjusting bases in the sequence

After an analysis you can edit or delete selected bases in the sample sequence or insert new bases.

> **Warning**
> This function should be used with careful consideration since no record is made of the base editing that you perform.

1. Select a deviation in the **Deviations Table** or a base position in the **Alignment** field.

2. Select **Edit:Adjust Sample Sequence** to display the **Adjust Sample Sequence** dialogue for the selected deviation or base position.

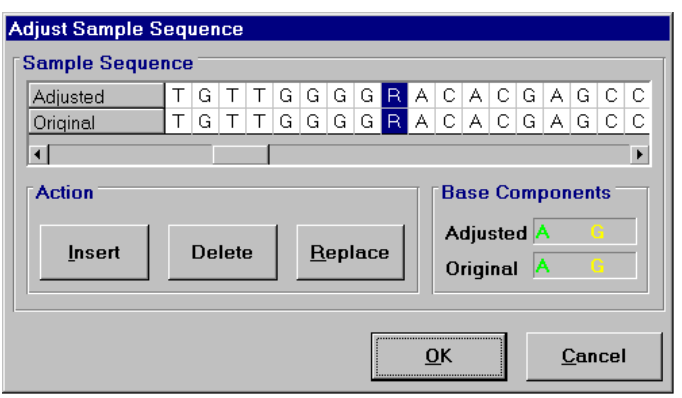


Figure 4-30. *Adjust Sample Sequence dialogue.*

3. To replace the selected base, click on **Replace** to display the **Edit Sample Base** dialogue.

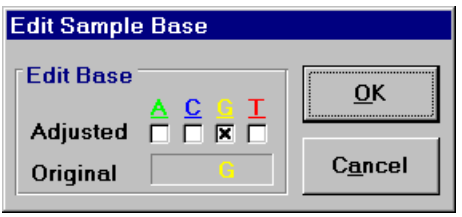


Figure 4-31. *Edit Sample Base dialogue.*

Check the appropriate base component(s) and click on **OK**. The **Adjusted** and **Original** base components are displayed in the **Adjust Sample Sequence** dialogue.

4. To delete the selected base, click on **Delete**.

5. To insert a base *before* the currently selected base, click on **Insert** to display the **Insert Sample Base** dialogue.

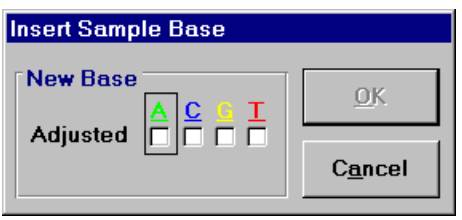


Figure 4-32. *Insert Sample Base dialogue.*

Select the base component(s) to be inserted and click on **OK**. The base or appropriate ambiguity for more than one selected component is inserted.

6. Click on **OK** when you have adjusted the sample sequence.

4.13 Printing a report

You can print out a report of the whole run or for a specific deviation that you are evaluating.

4.13.1 Printing a report

1. Select **Options:Report Options**. The **Report Options** dialogue is displayed.

Figure 4-33. *Report Options dialogue box.*

2. Select the components that you want to be included in the report by clicking in the box beside an option. These include:

Sample Information	Information about the samples that was previously entered using the sequencing software.
Comments on Analysis	Comments that you have entered into the **Comments on Analysis** dialogue
Deviations Table	Columns of the **Deviations Table** that you have specified in the list of options (see step 3).
Overview Graph	The **Overview** window for the current analysis. Select to print **One Page per Region** or **All Regions in One Page**.

3. Select the columns to be included in **Deviations Table** component of the report. Note that you can only have a maximum of 5 columns.

4. Select to **Include All** of the deviations or **Approved Only** deviation rows.

5. Check the **Report Interactive Confirmation Rows Only** option if you want to include only those deviation types used for assignment. This list can be modified by clicking on the **Interactive Assignment Options** button and checking the appropriate deviations in the displayed dialogue.

6. Click on **OK**.

7. Select **File:Print Report**. You can select the printer setup options in the **Print** dialogue.

4.13.2 Printing a deviations report

The **File:Print Deviations Report** command prints out extensive information about the deviation selected in the **Deviations Table**. This includes the raw data curves graph, deviation information, codon information and alignment sequence data.

4.13.3 Printing a report to file

To print a report to a file in text (**.txt**) format:

1. Select **Options:Report Options**. The **Report Options** dialogue is displayed (see Section 4.12.1).

2. Select the components that you want to be included in the report. Click on **OK**.

3. Select **File:Print to File**, enter a name for your report file and click on **Save**. The file can be opened in most word processor applications.

4.14 Exporting reports and results

It is possible to export reports in simple text format to facilitate easy integration of information into other applications.

To export a report to a database:

1. Select **Options:Export to Database Options**. A dialogue is displayed.

2. Select the components that you want to include in the exported file. Click on **OK**.

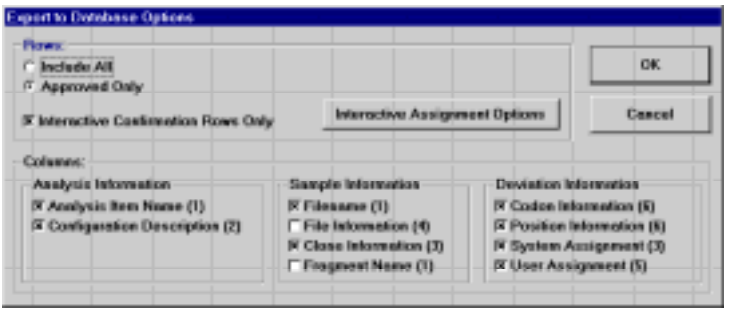


Figure 4-34. *Export to Database Options dialogue.*

3. Select to **Include All** of the deviations or **Approved Only** deviations rows.

4. The **Report Interactive Confirmation Rows Only** option is checked to include only those deviation types used for assignment. This list can be modified by clicking on the **Interactive Assignment Options** button and checking the appropriate deviations in the displayed dialogue.

5. Check the relevant **Columns** options to be included.

6. Click on **OK**. Select **File:Export Result to Database**. Give the file a name and click on **Save** (see Appendix B).

4.15 Changing user

It is possible that several users will use Mutation Analyser during the day without the application being turned off. To keep a record of users within the current analysis item file and also in the report, it is recommended that each new user logs his/her user identification.

1. Select **Options:Change User** to display.

2. Enter your **User ID** in the **User Identification** dialogue or select it from the pull-down list if one of the last five names to be used.

Figure 4-35. *User Identification dialogue.*

4.16 Exiting the application

To exit Mutation Analyser, select **File:Exit**.

5. Deviation analysis

In Chapter 4, it was explained how to create an analysis item and then perform an analysis. The key information and tools associated with the analysis were presented. This chapter introduces the various types of alignment deviations that you will encounter, and guides you on performing user assignments for the deviations.

Note: It is assumed that you have performed an analysis of your analysis item using the **Analysis:Calculate Deviations** command (<F9>) (see Section 4.3).

5.1 Types of alignment deviations

There are several types of deviation that you will encounter when using Mutation Analyser. These can be classified into four main groups:

- Single base deviations (see Section 5.2)
- Insertion/deletion deviations (see Section 5.3)
- Mixture insertion/deletion deviations (see Section 5.4)
- User defined deviations (see Section 5.5)

5.2 Single base deviations

This class includes all cases where Mutation Analyser has detected a difference in the sample sequence at a single base position compared to the aligned base in the gene configuration.

There are four types of single base deviation:

- Point replacement deviations
- Ambiguity deviations
- Sequence stop deviations
- Hotspot deviations

5.2.1 Point replacement deviations

This is where a base in the sample sequence is replaced with another resolved base that is different to the aligned base in the gene configuration. This is clearly seen in the **Alignment** field:

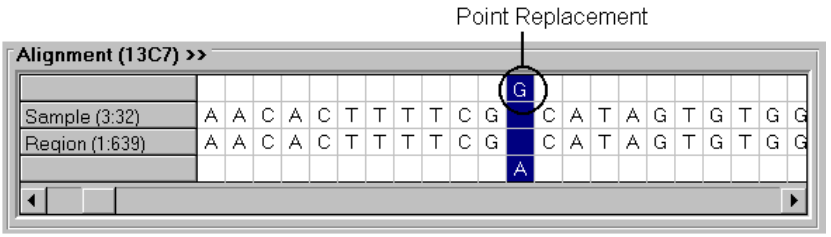


Figure 5-1. *Alignment field with selected point replacement from A to G in a p53 gene sample sequence.*

This is classified as a **PointRep** deviation in the **Deviations Table**:

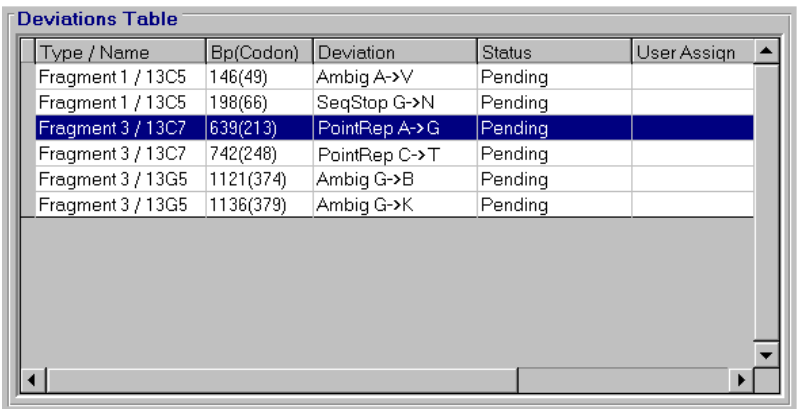


Figure 5-2. *Deviations Table with a point replacement deviation selected in a p53 gene sample.*

A point replacement is indicative of either a polymorphism or a mutation (see Beaudet and Tsui, 1993 [1]), although Mutation Analyser does not make this differentiation. User knowledge about the specific gene should allow you to identify a polymorphism vis-à-vis mutation. As an additional tool in this identification, you can for example define polymorphic positions as points of interest in the gene configuration (see Section 6.4.1).

Mutation Analyser can, however, display the effects of a point replacement on the codon and the encoded amino acid. In some cases the point replacement produces a codon that changes the encoded amino acid, while in other instances there is no change. This information can be viewed in the **Assignment** bar.

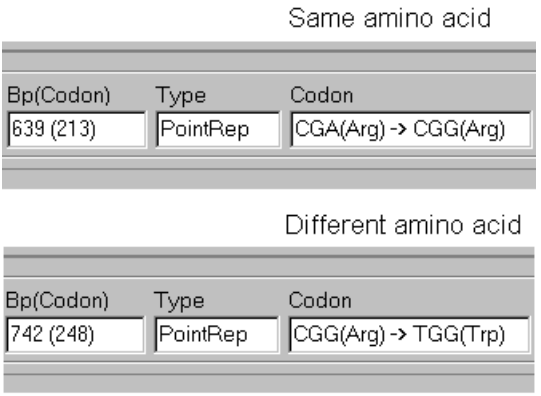


Figure 5-3. *The effect of a point replacement on the encoded amino acid can be seen in the Assignment bar. No change from Arg to Arg (above) and a change from Arg to Trp (below) as detected in p53 gene samples.*

5.2.2 Ambiguity deviations

This is where there is a replacement of a base in the sample sequence with an ambiguity code. An ambiguity code is awarded where the specific base was unresolved during the sequencing process due to there being more than one base component detected at the same position. An ambiguity code is unique for the specific combination of bases detected (see Table 5-1).

Ambiguity code	Unresolved components
M	AC
R	AG
W	AT
S	CG
Y	CT
K	GT
V	ACG
H	ACT
D	AGT
B	CGT
N	ACGT

Table 5-1. *Ambiguity codes for all combinations of unresolved components.*

An ambiguity deviation is clearly seen in the **Alignment** field:

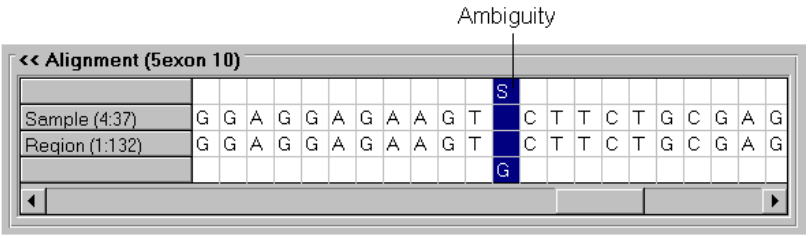


Figure 5-4. *Alignment field with an ambiguity deviation selected in a Ret gene sample sequence.*

Mutation Analyser classifies these as **Ambig** deviations in the **Deviations Table**.

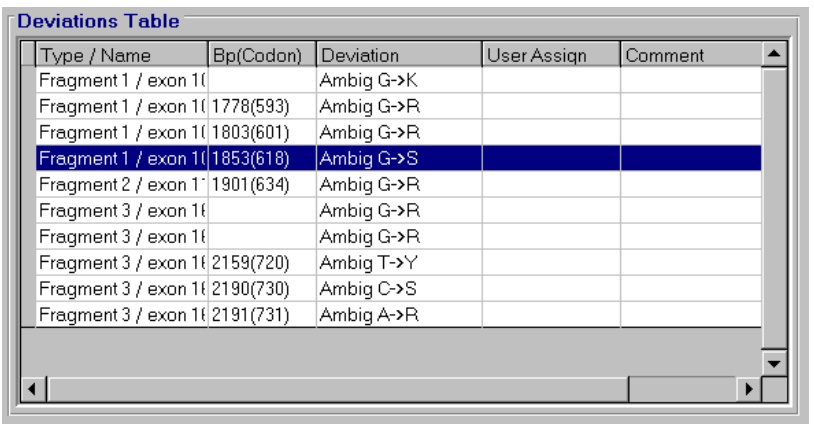


Figure 5-5. *Deviations Table with an ambiguity deviation selected in a Ret gene sample.*

Ambiguities are indicative of a mixed cell population and/or a heterogeneous cell sample. A given sample possibly contains cells from normal tissue and cells from tumour tissue, or that that there are polymorphic differences between alleles, i.e. heterozygous.

As discussed for point replacements (see Section 5.2.1), Mutation Analyser does not differentiate between a polymorphism and a mutation. Moreover, in the case of ambiguities, no information can be given about the encoded amino acid since the specific base is unresolved. The bases constituting the specific ambiguity code are readily seen in the **Assignment** bar by viewing the checked options.

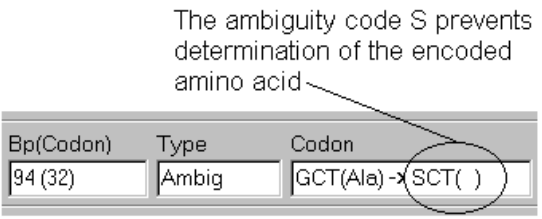


Figure 5-6. *The encoded amino acid cannot be determined in the Assignment bar from an ambiguity.*

5.2.3 Sequence stop deviations

A sequence stop is designated the ambiguity code 'N' and means that all four bases were detected at the same position in the sequencer fragment (see Table 5-1).

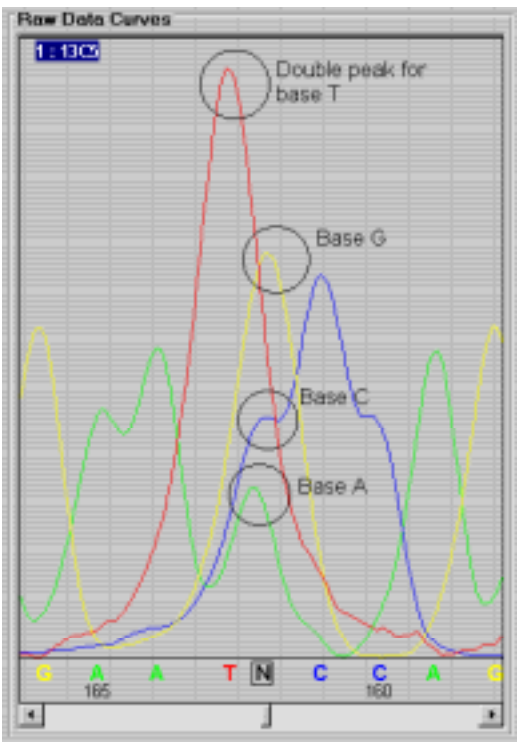


Figure 5-7. *Raw Data Curves field showing four bases at the same position.*

Sequence stops are indicative of problem regions in the template causing the sequencing enzyme to fall off without incorporating the correct ddNTP terminator, hence creating signals in all four base channels.

Sequence stops are classified as **SeqStop** deviations in the **Deviations Table** and counts as an ambiguity deviation for purposes of user assignment.

5.2.4 Hotspot deviations

If you defined points of interest in the gene configuration (see Section 6.4.1), whether for positions known to be polymorphic or prone to mutation, these can be used to generate **Hotspot** deviations that are shown in the **Deviations Table**.

For example:

1. Create an analysis item (see Section 4.2.1) based on the RETdemo gene configuration (see Section 6.13.1).

2. To generate hotspots, check the **Generate Hotspots from Points of Interest** option in the **Deviation Calculation Options** dialogue prior to analysis (see Section 4.3.1).

3. Include the hotspot deviations in the user assignment by checking the **Hotspot** deviation type in the **Interactive Assignment Options** dialogue (see Section 4.4.2).

4. Perform an analysis, **Analysis:Calculate deviations** or <F9>. The detected hotspot deviations, including deviation type, are displayed in the **Deviations Table**.

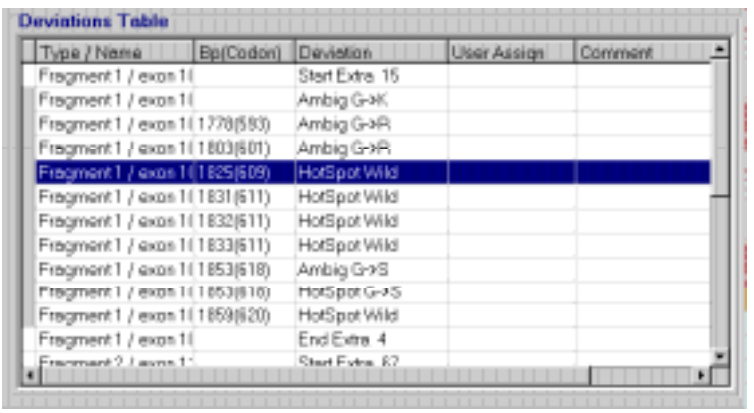


Figure 5-8. *Deviations Table containing Hotspot deviations for an analysis item based on the RETdemo gene configuration.*

5.2.5 User assignment of single base deviations

A single base deviation selected in the **Deviations Table**, including point replacements (**PointRep**), ambiguities (**Ambig**), sequence stops (**SeqStop**), and hotspots (**Hotspot**) will be represented in the **Single Base Assignment** bar.

1. View the selected deviation in the **Raw Data Curves** field. This will help you to evaluate the validity of the assignment made by the alignment calculation algorithm.

2. In the **Single Base Assignment** bar, select the base(s) that you consider should be assigned to the specific deviation. By default, the base(s) corresponding to the selected deviation is selected, as denoted by the checked box(es).

3. Type in any comments into the **Comments** box in the **Single Base Assignment** bar.

4. Make the appropriate assignment based on the following criteria:

Use this button if you want to automatically assign the deviation back to the wild type, i.e. the same as in the gene configuration. This option also automatically *approves* the wild type assignment (see **Approve** button description next).

Use this button if you want to approve the assignment proposed by Mutation Analyser or an assignment of your own choice.

Use this button if you want to ignore the selected deviation or you are unable to make a determination about the deviation depending on your evaluation strategy.

Note: This should be used with caution.

6. Your user assignment now automatically updates the information contained in the **Deviations Table** (see Section 4.4.1 for details of the various columns in the **Deviations Table**). The updated information includes:

Status Colour	The bar changes from grey (Pending) to either red (Approved as deviation), black (approved Wild type), or yellow (Ignored).
Status	The status of a particular deviation is displayed, either **Wild**, **Approved** or **Ignored**.
User Assign	The user assigned base/ambiguity is displayed.
Responsible	This displays the user ID of the person whom made the assignment for this particular deviation (see Section 4.12).
Comments	Comments that you added during assignment will be displayed.

7. By default, the next deviation in the **Deviations Table** is automatically selected for assignment. This function can be turned off by selecting **Options:Interactive Assignment Options** and unchecking the **Auto Increment Deviation List** option.

8. To undo an assignment, i.e. return to **Pending** status, select the deviation in the **Deviations Table** and then **Edit:Undo Interactive Assignment**. All user assignment information including any comment will be removed. Alternatively, if you want to change the assignment, e.g. from **Approved** to **Wild**, select the deviation and make a new assignment.

5.2.6 Approving multiple deviations

If the analysis item contains a deviation at the same position in overlapping sample sequences, you can make a general user assignment for all such deviations.

1. Select **Edit:Multiple Sample Approval**. The function is active when this menu command is checked.

2. When you select a deviation for which there is a deviation in the same position in overlapping sample sequences, all will be grouped together in the **Deviations Table**, and naturally in the **Raw Data Curves** field.

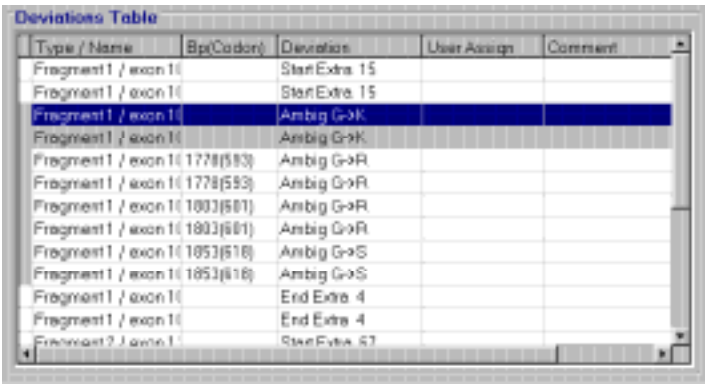


Figure 5-9. *Deviations Table showing a selected deviation in a Ret gene sample. A deviation at the same base position in an overlapping sample is grouped underneath the selected deviation and is highlighted in as a grey row.*

3. Follow the user assignment instructions detailed in Section 5.2.5. The only difference is that the **Approve** button is replaced with **Approve All**, which means that the same assignment is made for the deviation group in all overlapping samples. Similarly, selecting **Wild** applies to the deviation group. However, if you assign **Ignore**, this is only valid for the selected deviation and other instances of the deviation must be user assigned separately.

5.3 Insertion/deletion assignments

Explicit base insertions or deletions in the sample sequence can be detected in a homogenous and homozygous cell sample. This can be clearly seen in the **Alignment** field.

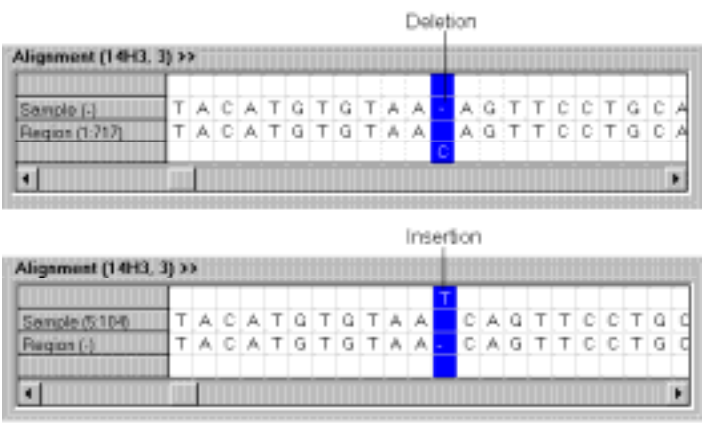


Figure 5-10. *The appearance a single base deletion (above) or insertion (below) for a p53 gene sample sequence in the Alignment field.*

Mutation Analyser classifies these deviations as **Insert** and **Delete** in the **Deviations** column of the **Deviations Table**, together with the number of inserted or deleted bases for adjacent insertion or deletion positions.

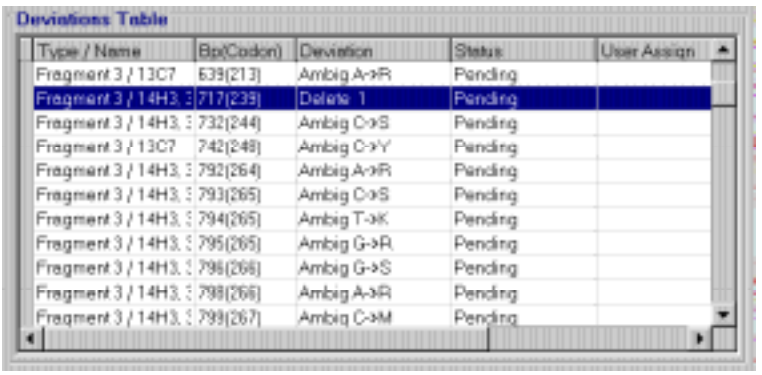


Figure 5-11. *Deviations Table with an explicit deletion deviation selected in a p53 gene sample. In this case, it is a single base deletion as denoted by 'Delete 1'.*

Selection of an insertion or deletion will display the relevant information in the **Insertion Assignment** and **Deletion Assignment** bars respectively.

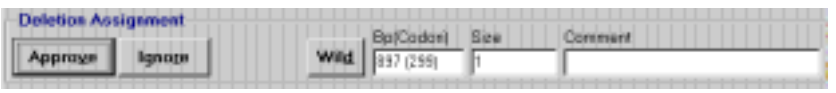


Figure 5-12. *Deletion Assignment bar for a deletion deviation of size one base, as denoted by the number '1'.*

5.3.1 User assignment of insertion/deletion deviations

An insertion or deletion deviation selected in the **Deviations Table**, is user assigned in the **Insertion Assignment** bar and **Deletion Assignment** bar.

1. View the selected deviation in the **Raw Data Curves** field. This will help you to evaluate the validity of the assignment made by the alignment calculation algorithm.

   If you want to change the specified size of the insertion or deletion, enter the new value into the Size box in the **Assignment** bar. Note that details can be given in the Comments box as to the identity of the bases involved in the insertion or deletion resulting from the change in size.

2. Type in any comments into the **Comments** box in the **Single Base Assignment** bar.

3. Make the appropriate assignment based on the following criteria:

Use this button if you want to automatically assign the deviation back to the wild type, i.e. the same as in the gene configuration. This option also automatically *approves* the wild type assignment.

Use this button if you want to approve the insertion or deletion assignment proposed by Mutation Analyser or an assignment of your own choice.

Note: Use of the **Wild** button automatically approves the assignment.

Use this button if you want to ignore the selected deviation or you are unable to make a determination about the deviation depending on your evaluation strategy.

Note: This should be used with caution.

6. Your user assignment now automatically updates the information contained in the **Deviations Table** (see Section 4.4.1 for details of the various columns in the **Deviations Table**). The updated information includes:

Status Colour	The bar changes from grey (Pending) to either red (Approved), black (approved Wild type) or yellow (Ignored).
Status	The status of a particular deviation is displayed, either **Wild**, **Approved** or **Ignored**.
User Assign	The nature and size of the user assignment is displayed.
Responsible	This displays the user ID of the person whom made the assignment for this particular deviation (see Section 4.12).
Comments	Comments that you added during assignment will be displayed.

7. By default, the next deviation in the **Deviations Table** is automatically selected for assignment. This function can be turned off by selecting **Options:Interactive Assignment Options** and unchecking the **Auto Increment Deviation List** option.

8. To undo an assignment, i.e. return to **Pending** status, select the deviation in the **Deviations Table** and then **Edit:Undo Interactive Assignment**. Alternatively, if you want to change the assignment, e.g. from **Approved** to **Wild**, select the deviation and make a new assignment.

5.4 Mixture insertion/deletion deviation assignments

Mutation Analyser has an advanced **Mixture Analysis** function that can detect insertion/deletion deviations in sequence data derived from samples containing mixed cell populations or cells that are heterozygous.

1. When a tissue sample is prepared, it possibly contains a heterogeneous population of cells including the normal cells and tumour cells. The normal cells may originate from the surrounding healthy tissue or come from a tumour specimen that contains both cell types.

2. Two types of sequencing fragment will be produced, namely a wild type gene fragment and a mutant gene fragment. If the mutant gene has an insertion or deletion, this means that the resulting sequencer fragment will be longer or shorter respectively, than the corresponding wild type fragment.

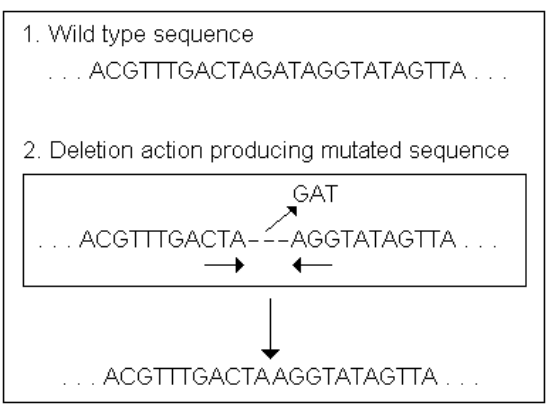


Figure 5-13. *An example wild type fragment sequence and the corresponding sequence in a mutated sample fragment containing a 3 base deletion.*

3. When sequencing is performed, detection of bases from the mutated gene fragment and wild type fragment will be similar, up until the point where the insertion or deletion occurs. After this point, more than one base will be detected at each position.

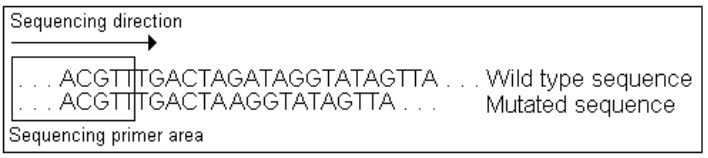


Figure 5-14. *Amplification of a mixed PCR fragment population derived from wild type and deletion mutated gene sequences.*

Note: It is possible, of course, for deviations to occur upstream of the detected mixture deviation and these should be evaluated independently of the mixture analysis.

4. The base calling algorithm will detect the two bases at the same position in the sample sequence and ambiguity codes will be assigned to the positions of non-alignment. This region containing many such ambiguity codes is representative of a mixture insertion/deletion deviation.

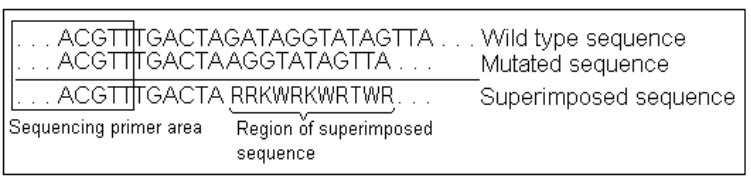


Figure 5-15. *A superimposed sequence resulting from a mixture of wild type and deletion mutated gene sequences*

5.4.1 Recognising mixture insertion/deletion deviations

There are several tell-tale signs that you should look for to recognise a mixture insertion or deletion deviation:

- Examine the **Deviations Table** for a long list of adjacent ambiguity deviations (**Ambig**) in a single fragment.

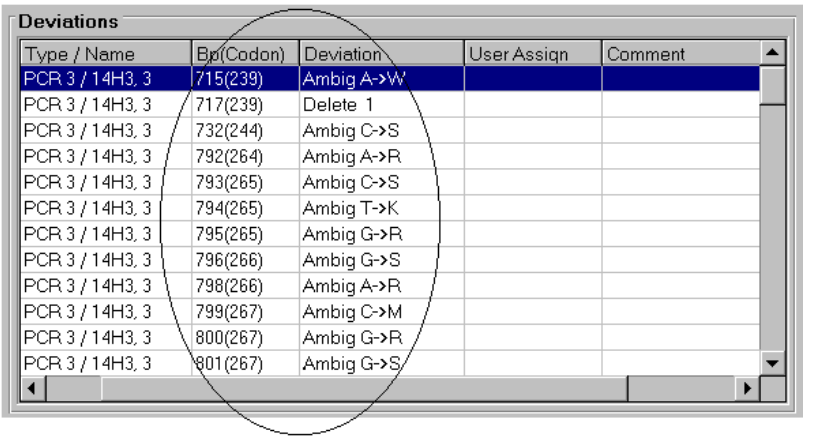


Figure 5-16. *Deviations Table containing many adjacent ambiguity deviations in a p53 gene sample.*

- Read the sequence data in the **Alignment** field in the same direction as the sequencing direction, i.e. from *left* to *right* for Sense (**>>**) or from *right* to *left* for AntiSense (**<<**). You may see a pattern whereby the upstream region of the sequence shows good alignment and is followed by repeating groups of adjacent point ambiguities.

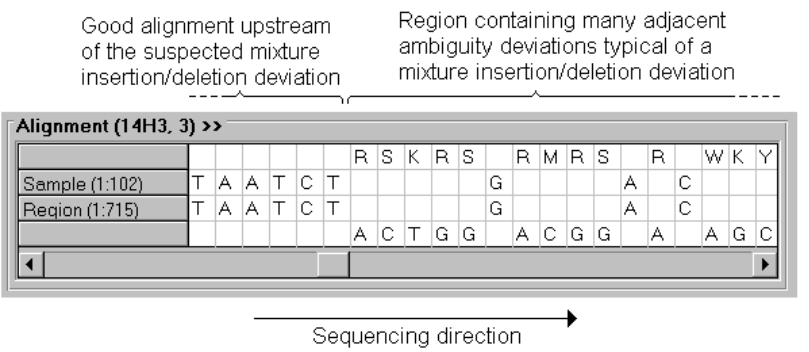


Figure 5-17. *Alignment field showing a p53 gene fragment sequenced in the Sense **(>>)** direction. Reading from left to right, an area of good alignment precedes an area of adjacent ambiguity deviations.*

- Look at the data in the **Analysis Overview** window by selecting **View:Analysis Overview** (see Section 4.8). You should be able to see that the upstream region of the sample fragment displays relatively few deviations followed by a downstream region that is densely packed with ambiguity deviations.

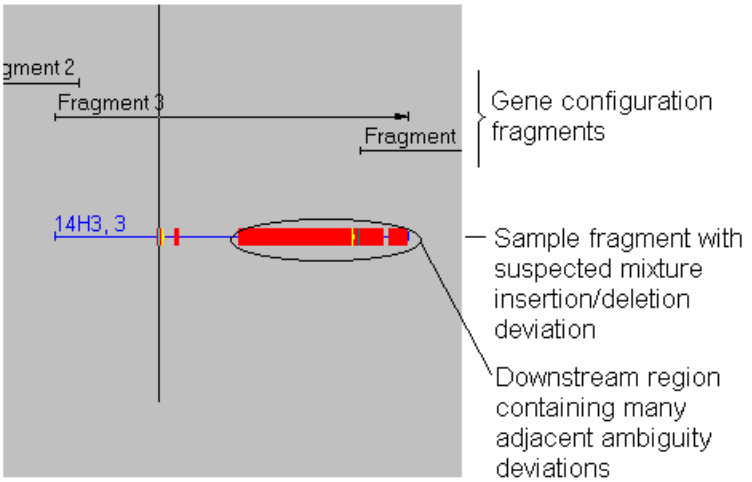


Figure 5-18. *Portion of the Analysis Overview window showing the alignment between a p53 gene configuration sequencer fragment and a sample fragment suspected of containing a mixture insertion/deletion. Note that the fragment contains an upstream area with good alignment followed by a downstream area that is densely packed with ambiguity deviations.*

5.4.2 Performing a mixture analysis

If you suspect that one or more of the samples contain a mixture insertion or deletion, it is possible to confirm this by performing a mixture analysis.

1. Select **Analysis:Mixture Analysis**. The **Mixture Analysis** dialogue is displayed.

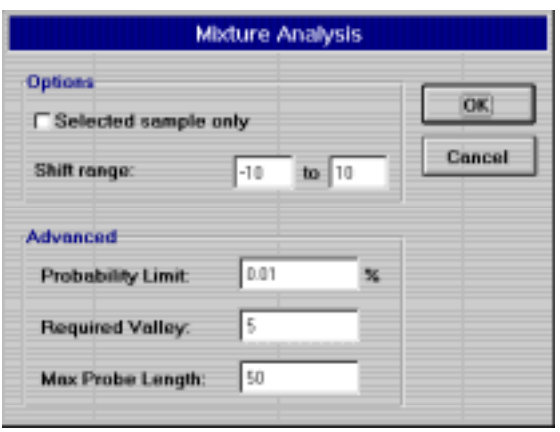


Figure 5-19. *Mixture Analysis dialogue.*

2. Select the **Shift range** for the analysis.

   This will set the limits for the number of alignment shift positions to be tested. The default range of -10 to 10 means that the 10 alignment positions, both to the left and right of the currently displayed alignment respectively, will be analysed. This, in turn, means that tentative insertions/deletions of 10 bases or less are looked for. Each sample sequence will be shifted one step at a time away from its original position with respect to the gene configuration. The alignment between sample sequence and gene configuration is then statistically tested before shifting another step. The number of stepwise shifts is equal to the range limit, both to the left and right of the currently viewed alignment position.

3. If you want to analyse only the selected sequence sample, i.e. the sample containing the deviation selected in the **Deviations Table**, check the **Selected Sample Only** option. By default, this function is not selected so that all sample sequences in the analysis item will be included in the analysis.

4. Click on **OK**. The analysis will now be performed and if a significant result is obtained, these will be shown in the **Deviations Table**. This will be displayed as a new class of deviation, namely **MixInsDel**.

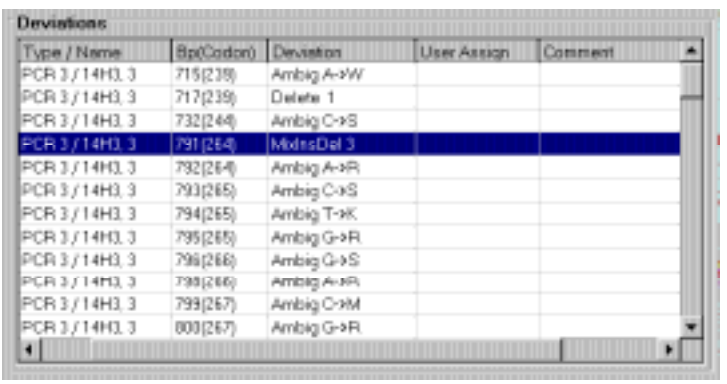


Figure 5-20. *Deviations Table displaying a Mixture insertion/deviation deviation detected in a p53 gene sample.*

The size of the mixture insertion/deletion deviation shown in the **Deviation** column of the **Deviations Table** represents the number of places that the sample deviation was shifted in the mixture analysis to give the statistically best alignment with the gene configuration.

A positive value indicates that the sample sequence is shifted to the right of its current position and, conversely, a negative value

indicates that the sample sequence is shifted to the left of its current position. The example shown in Figure 5.18, **MixInsDel 3**, indicates a shift of the sample sequence 3 places to the right of its original position to give the statistically best alignment with the gene configuration in the affected area.

This result can be analysed in the **Alignment** field.

5.4.3 Using the alignment field to evaluate mixture analysis results

A colour-coded representation of the mixture analysis results are displayed in the **Alignment** field.

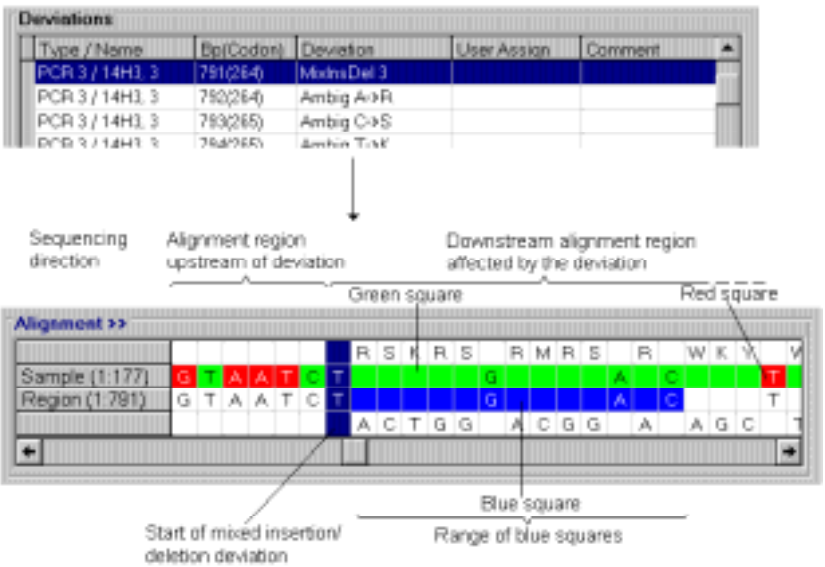


Figure 5-21. *Contents of the Alignment field for a p53 gene sample sequence. In this example, the mixture analysis detected a mixture insertion/deletion of 3 bases.*

Important information in the **Alignment** field includes:

Sequencing direction — The mixture analysis is always performed in the sequencing direction, which is displayed either as **>>** for the Sense direction, or **<<** for the Anti-Sense direction. The example shown in Figure 5-19 displays a sample fragment sequenced in the Sense direction.

Start position of deviation	This is the *estimated* position where the mixture insertion/deletion occurs.
Blue squares	The blue squares indicate those bases that were used in the mixture analysis to give a statistically meaningful result. The analysis uses the minimum possible number of bases in order to achieve this result.
Green squares	The green squares indicate those bases in the sample sequence which are in agreement with the bases in the gene configuration resulting from the shift. The value of the shift is displayed in the **Deviation** column of the **Deviations table**. A *positive* value indicates an equivalent shift of the sample sequence to the right of its currently displayed alignment. A *negative* value indicates an equivalent shift of the sample sequence to the left of its currently displayed alignment.
Red squares	The red squares indicate those bases in the sample sequence which are not in agreement with the bases in the gene configuration resulting from the shift.

Alignment area upstream of the deviation
As described earlier, the upstream area of the sample sequence shows good alignment to the gene configuration and the alignment only becomes poor in the downstream area affected by the mixture insertion/deletion. As a result of the mixture analysis, the upstream area will almost entirely display red squares. This is because the applied shift to this upstream area results in a non-compatible alignment between the sample sequence and gene configuration. Any green squares displayed in the upstream area as a result of the shift, are random events.

Alignment region downstream of the deviation
Contrary to the upstream area, the downstream area should after mixture analysis, display a dramatic improvement in the alignment between the sample sequence and gene configuration as a result of the shift. This should be evident by the majority of green squares. Any red squares indicate a disagreement between the sample sequence and gene configuration sequence as a result of the shift. Assuming the mixture

insertion/deletion deviation is valid, red squares can occur for two reasons:

- there is a single base deviation at the position marked by the red square, hence the non-alignment after the shift
- there are two components occurring at the same position in the sample sequence but the base calling algorithm of the sequencing software was only able to detect one component that is not compatible with the base in the gene configuration.

It is possible to check if there is a single base deviation in the sample sequence or if there are two components of which only one of them was detected:

1. Click on the red base of interest in the **Alignment** field.

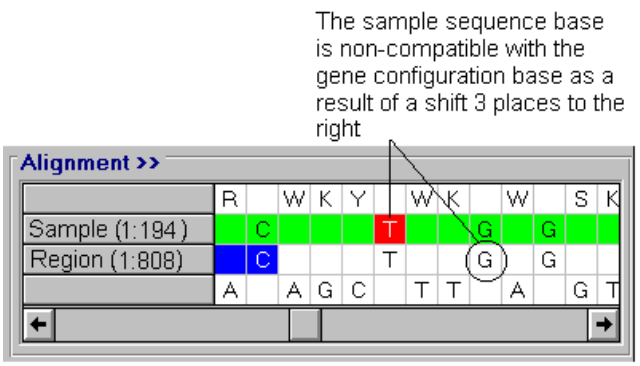


Figure 5-22.*A base in the sample sequence which shows non-agreement with the gene configuration as a result of a 3 base shift to the right.*

2. If there are two bases in the sample sequence and one of these was not detected by the base calling algorithm, you would expect to be able to see both bases in the **Raw Data Curves** field. Indeed, the selected deviation is automatically selected in the **Raw Data Curves** field and an indication of the **Expected** bases is given, i.e. the two bases that are expected to be present in the sample sequence if there is to be an agreement as a result of the shift.

   In the example shown in Figures 5-20 and 5-21, the red base in the sample sequence is a **T** and the base in the gene configuration three bases to the right is a **G**. Thus, for there to be an agreement as a result of the shift, the sample sequence is expected to contain a T and G base at the same position.

3. Magnify the view for the raw data (**View:Zoom In** or <F7>) and determine if both of the expected bases are present. If both bases are present, as in Figure 5-21, this implies that the base calling algorithm had, in this case, failed to detect a weak signal for the base **G**.

   If both bases are not present in the sample sequence, for example only the base **T** was present, then this suggests a single base deviation in the sample sequence.

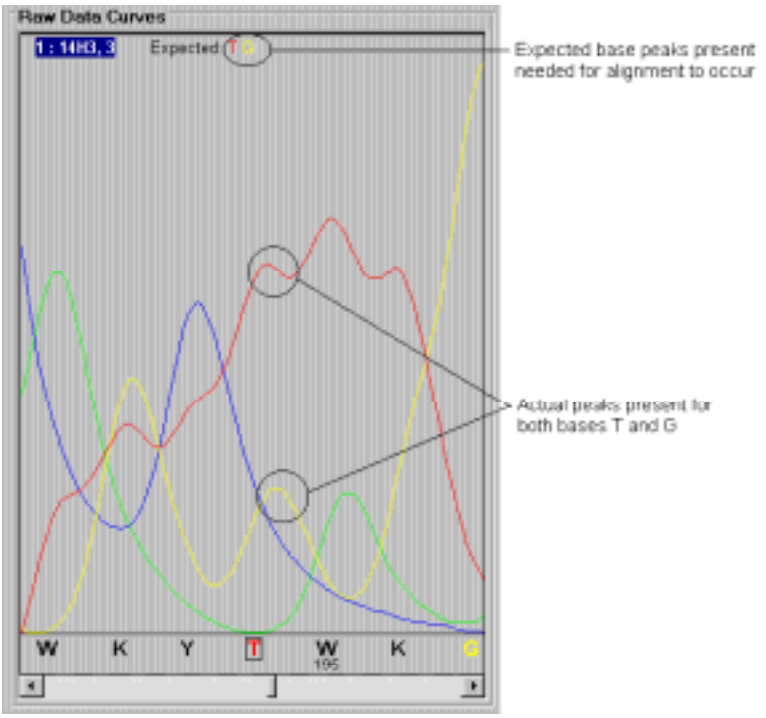


Figure 5-23. *Raw Data Curves field showing the expected and actual component bases in the raw data*

5.4.4 User assignment of mixture insertion/deletion deviations

Having performed a mixture analysis, use the information in the **Mixture Assignment** bar to automatically calculate if there has been a base insertion or deletion in the mutated sequence. You can then make the appropriate user assignment.

Figure 5-24. *Mixture Assignment bar*

The various information contained in the **Mixture Assignment** bar includes:

~Bp(Codon)	*Estimated* position of the start of the deviation and codon number (in parentheses.
Type	The type of mixture deviation. By default, **MixInsDel** is displayed until you direct the program to calculate whether the deletion is a mixture insertion (**MixIns**) or a mixture deletion (**MixDel**) (see below).
Size	The size of the potential mutation.
False rate %	This is an *estimate* of the statistical probability that the program assignment, following the results of the mixture analysis, is false, i.e. the lower the figure presented here, the higher the probability that the program assignment based on the calculated shift is true.

The **Undefined** option is automatically selected until you calculate the type of mixture insertion/deletion deviation:

1. Click on the **No Gap** button if there was no explicit insertion or deletion directly adjacent and upstream of the detected deviation position. This will be true for most of the instances when you perform a mixture analysis.

   The result, insertion (**MixIns**) or deletion (**MixDel**), will be automatically displayed in the **Type** field. The size of the insertion or deletion is equivalent to the shift value indicated. For details on the process behind the calculation, see Section 5.4.6.

Note: You should not become confused by the sign of the **Size** value. Only the magnitude of the value is relevant here. The sign of the number is used in the alignment calculation (see Section 5.4.6).

2. Add any appropriate comments.

3. To approve the assignment, click on the **Approve** button. A dialogue box will be displayed asking if you want to ignore all subsequent deviations *downstream* of the **MixIns** or **MixDel** deviation. Click on **Yes** if you want to ignore deviations in the list downstream from the mixture insertion/deletion deviation. Click on **No** if you want to manually evaluate every deviation.

   Your user assignment now automatically updates the information contained in the **Deviations Table** (see Section 4.4.1 for details of the various columns in the **Deviations Table**). The updated information includes:

Status Colour	The bar changes from grey (Pending) to either red (Approved), black (approved Wild type) or yellow (Ignored).
Status	The status of a particular deviation is displayed, either **Approved**, **Wild** or **Ignored**.
User Assign	The nature and size of the user assignment is displayed.
Responsible	This displays the user ID of the person whom made the assignment for this particular deviation (see Section 4.12).
Comments	Comments that you added during assignment will be displayed.

4. By default, the next deviation in the **Deviations Table** is automatically selected for assignment. If you ignored all other deviations downstream of the mixture insertion/deletion deviation (see Step 3.) the next deviation not associated with this assignment is selected. This function can be turned off by selecting **Options:Interactive Assignment Options** and unchecking the **Auto Increment Deviation List** option.

5. To undo an assignment, i.e. return to **Pending** status, select the deviation in the **Deviations Table** and then **Edit:Undo Interactive Assignment**.

Explicit gaps and mixture analysis
In some circumstances, the mutated sequence may predominate over the wild type sequence in a mixed population. In the deviations calculation, the mutated sequence will likely find the greatest degree of

alignment with the gene configuration sequence. If the mutated gene sequence contains a deletion or insertion, this will be displayed as explicit insertion(s) or deletion(s) respectively. These can normally be evaluated without the need for a mixture analysis. In some cases, however, there may be sufficient interference from the wild type sequence in the region downstream of the explicit insertion or deletion, that you wish to perform a mixture analysis. This is performed in exactly the same way as described in Sections 5.4.2 to 5.4.4, except that the **Gap** radio button is chosen in the **Mixture Assignment** bar for automatic calculation of the type of mixed deviation. The results of the calculation are displayed accordingly.

5.4.5 Advanced options for mixture insertion/deletion deviation assignments

There are several parameters that may be altered to adjust the sensitivity of the mixture analysis.

Select **Analysis:Mixture analysis** to display the **Mixture Analysis** dialogue.

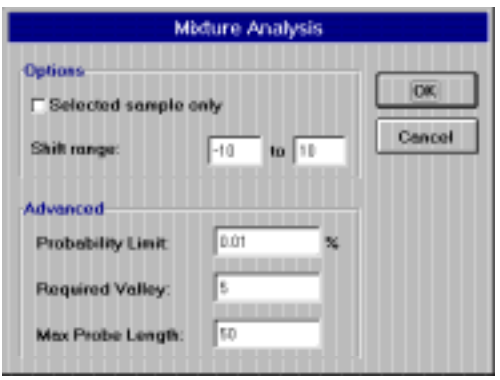


Figure 5-25. *Mixture Analysis dialogue.*

The parameters that you can alter are the **Probability Limit**, **Required Valley** and **Max Probe Length**.

Probability limit

The **Probability Limit** value sets the inclusion threshold for calculated alignment probabilities in the mixture analysis results. Values less than the set probability limit indicate that a statistically significant alignment was obtained between sample sequence and master sequence for the currently tested shift position. All shifts fulfilling the probability limit criterion will be presented. By default, the probability limit is set to 0.01%.

Required valley
During mixture analysis the alignment algorithm assigns a cumulative score, unique for each shift, as the alignment proceeds in the sequencing direction. The awarded score is based upon the type of association between the base in a sample sequence and the corresponding base in the gene configuration:

Event	Score
Matching base	+1
Shift criterion fulfilled	-1

During a mixture analysis, the algorithm uses a drop in the value of the cumulative score to detect the mixture insertion/deletion deviation. The **Required Valley** value is the number of cumulative points that must be lost in order for it to be treated as a deviation. By default, this value is set to 5.

In an ideal situation, the cumulative score produces a very characteristic curve. This can be viewed by clicking on the **Graph** button in the **Mixture Assignment** bar following mixture analysis of a sample.

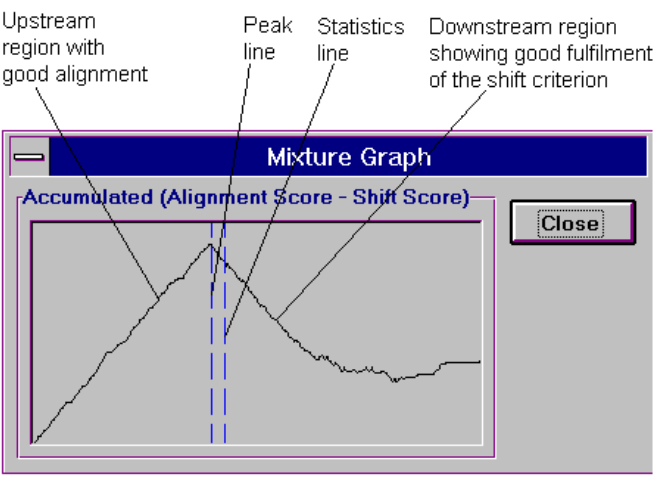


Figure 5-26. *Mixture Graph panel showing the cumulative score of an alignment following mixture analysis. This example comes from a sample sequenced in the Sense direction.*

The peak line represents the start of the mixture insertion/deletion deviation. A peak can only be identified as a mixture insertion/deletion deviation if it is followed by a continuous reduction in cumulative score equal to or greater than the **Required Valley** value in the **Mixture Analysis** dialogue. The distance between the peak line and statistics line represents the number of bases used to obtain the statistical result.

It is not always possible for the automated mixture analysis to detect a mixture insertion/deletion deviation since the **Required Valley** value is too high. This is particularly true in a mixed sequence sample where the wild type sequence predominates, with only occasional superposition of the mutated sequence for a small number of bases. To detect such cases, the **Required Valley** value should be lowered and the mixture analysis repeated. The **Mixture Graph** can appear as a set of incremental steps or contain any number of small peaks and valleys.

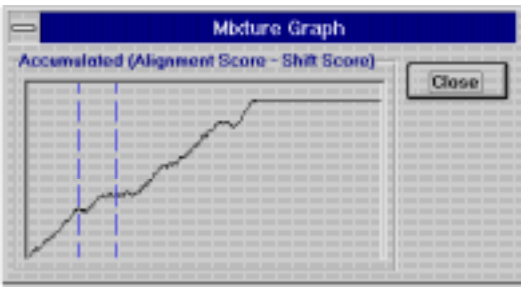


Figure 5-27. *Mixture Graph panel showing the cumulative score of an alignment following a mixture analysis where the Required Valley value needed to be lowered from 5 to 1 in order to detect the mixture insertion/deletion deviation (sample sequenced in the Sense direction).*

Max Probe Length
The **Max Probe Length** value is the maximum number of bases used in the mixture analysis for the statistical calculation. These bases are represented by the blue squares in the **Alignment** field. If there is a good alignment, fewer bases will have been used to obtain the result and consequently less blue squares will be seen. In other cases, the alignment may not be so perfect and, hence, more bases are needed in the analysis to obtain statistically meaningful results. By default, the **Maximum Probe Length** value is 50.

5.4.6 Understanding mixture assignments

In Sections 5.4.2 to 5.4.4, you were shown how to automatically calculate the specific type of mixture deviation, either an insertion or deletion. In most cases, mixture analysis is performed on samples containing a stronger wild type sequence signal and thus with no explicit deletion adjacent to the area suspected of containing a mixture insertion/deletion. The process behind the mixture analysis is described.

Insertion or deletion
The information contained in the **Deviations Table** and **Alignment** fields, together with the raw data curves, allow you to manually differentiate between a mixture insertion or deletion deviation. The same principles are applied when you select **No Gap** in the **Mixture Assignment** bar (see Section 5.4.4) to obtain the same result.

- If the deviation occurs within a sample with a *Sense* sequencing direction, a *positive* size value indicates a *deletion* of length equal to the size value. This can be explained in that a statistically meaningful alignment is obtained in the affected sequence area if the sample sequence is shifted to the *right* of its original position, i.e. *downstream* in the sequencing direction. This leaves a gap in the sample sequence that corresponds to a deletion.

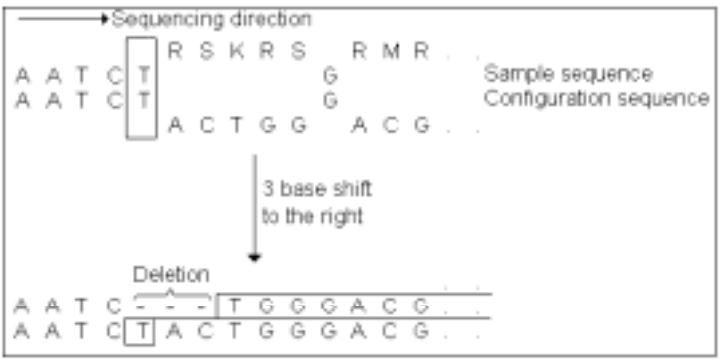


Figure 5-28. *Example of a sample with a Sense sequencing direction containing a mixture deletion deviation of 3 bases.*

- If the deviation occurs within a sample with a *Sense* sequencing direction, a *negative* size value indicates an *insertion* of length equal to the size value. This can be explained in that a statistically meaningful alignment is obtained in the affected sequence area if the sample sequence is shifted to the *left* of its original position from the start point of the deviation, i.e. *upstream* in the sequencing direction. Those bases already present upstream in the sample sequence, equivalent to the shift value, are base insertions.

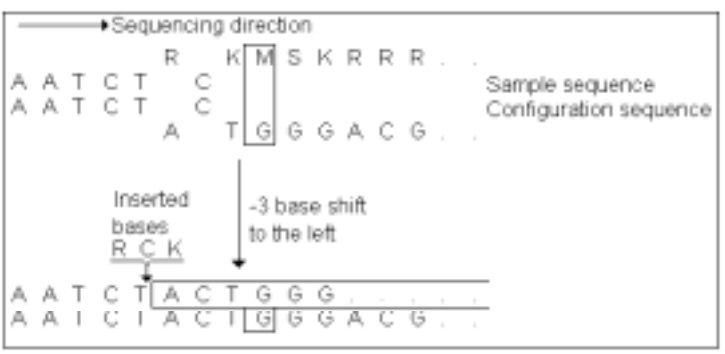


Figure 5-29. *Example of a sample with a Sense sequencing direction containing a mixture insertion deviation of 3 bases.*

- If the deviation occurs within a sample with an *AntiSense* sequencing direction, a *positive* size value indicates an *insertion* of length equal to the size value. This can be explained in that a statistically meaningful alignment is obtained in the affected sequence area if the sample sequence is shifted to the *right* of its original position from the start point of the deviation, i.e. *upstream* in the sequencing direction. Those bases already present upstream in the sample sequence, equivalent to the shift value, are base insertions.

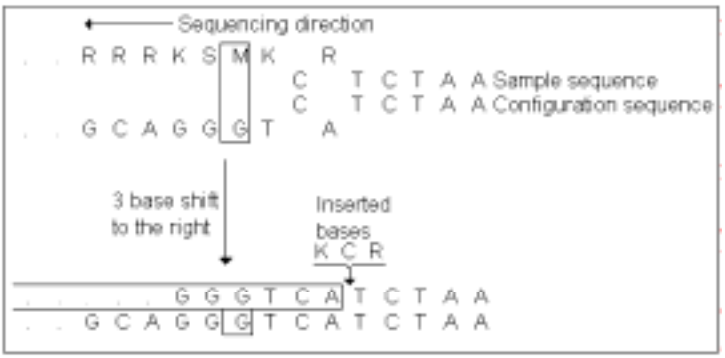


Figure 5-30. *Example of a sample with an Anti-Sense sequencing direction containing a mixture insertion deviation of 3 bases.*

- If the deviation occurs within a sample with an *AntiSense* sequencing direction, a *negative* size value indicates a *deletion* of length equal to the size value. This can be explained in that a statistically meaningful alignment is obtained in the affected sequence area if the sample sequence is shifted to the *left* of its original position from the start point of the deviation, i.e. *downstream* in the sequencing direction. This leaves a gap in the sample sequence, hence a deletion.

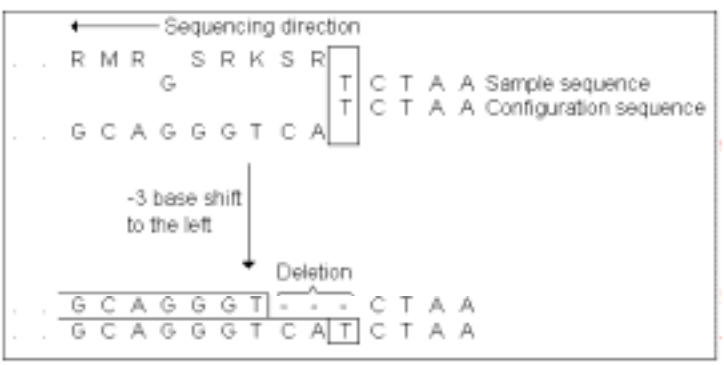


Figure 5-31. *Example of a sample with an Anti-Sense sequencing direction containing a mixture deletion deviation of 3 bases.*

5.5 User detected deviation

In some circumstances, you may wish to highlight a base position in the sample sequence, which has been detected as a wild type base, and give a specific comment. Moreover, you may want to change the assignment from the wild type to another type of single base deviation. This may be useful if, for example, the mutated component in a sample gave a very weak signal and was thus difficult to detect during sequencing. Mutation Analyser allows a user to identify and assign single base deviations, i.e. ambiguities, point replacements and sequence stops. Such situations will only normally be found on close inspection of the raw data curves.

1. In the **Raw Data Curves** field, select the base to which you want to add a comment and/or change the assignment. Naturally this corresponds to the wild type base in the gene configuration, otherwise it would have been automatically detected as a deviation.

2. Select **Edit:User Detected Deviation**. Information about the detected deviation will now be displayed in the **Deviations Table** and be designated **UserDev**.

   Alternatively, if you know the position of a specific base without needing to look in the **Raw Data Curves** field, select **Edit:User Detected Deviation**. A dialogue will be displayed into which you should enter the base position number. Click on **OK**. Information about the detected deviation will now be displayed in the **Deviations Table** and be designated **UserDev**.

3. In the **Single Base Assignment** bar, the base corresponding to the wild type will be checked. Add a comment in the appropriate field and click on **Approve** if you intend to keep the wild type assignment.

   Alternatively, check the appropriate base option(s)and click on the **Approve** button.

4. The **Deviations Table** will now be automatically updated.

5.6 Some "Golden rules" for interpreting mutation data

It may sometimes be difficult to establish whether a listed deviation is a real mutation or an artifact. A few simple suggestions may help you to make such judgements.

5.6.1 Single base deviations

1. Check if any of the suggested bases for an ambiguity are the result of an artifact. If not, there most likely exists more than one population of cells which have a different base at the same sequence position.

2. Check what changes occur at the amino acid level resulting from a specific point replacement or an alternative base suggested by an ambiguity code. See Section 5.2.

3. Out of the four ALF curves, a maximum of two peaks should be present at the site of an ambiguity. More than two peaks implies possible sequencing compression or stop. In case of doubt, compare these peak positions in other samples, preferably from the same gel.

4. Where an ambiguous peak on the curve is suggested as being a mutation, the background on the curve should be stable and regular in the area around the ambiguity. Check if there is noise on the curve in the regions before and after the ambiguous peak.

5. The mutation, or change in the DNA sequence, should *not* be present in all clones. If so, this suggests a sequencing artifact.

6. Estimate the peak height and valley depth by comparing similar regions within and between samples. This will give you an indication of whether a peak is likely to be an artifact. Take into consideration the sequence context, i.e. the size of a peak is influenced by the specific pattern of neighbouring sequences. Comparisons of different regions within the same sample are not totally reliable since the peak heights and valley depths are dependent on the resolution at a particular point of the run.

7. When available, you can confirm a suspected mutation by reading the overlapping sequence from the neighbouring fragment.

8. If a suspected mutation is not clear enough, you must do a confirmation run with newly sequenced DNA from the original tissue sample.

5.6.2 Deletions and insertions

1. In a heterogeneous cell population, if the deletion or insertion is large (>5-7 bases) you should see widened, or even two extension product peaks. This corresponds to there being one DNA fragment being longer or shorter than the other fragment in the mixed sample population.

2. If the additional end peak described in the previous point comes before the normal end peak, this indicates a deletion. An extra peak after the normal peak indicates an insertion. Compare with neighbouring clones of the same fragment type.

3. Smaller deletions or insertions (2-4 bases) are indicated by an end peak that is wider than normal. This may be difficult to measure.

4. Look for the characteristic behaviour of deletions or insertions when a well-aligned sequence suddenly becomes mixed and generates a lot of ambiguities. These instances usually require you to perform an automated mixture analysis to determine the size.

6. Configuration Editor

The Configuration Editor application allows you to create your own configuration files for a specific gene. Gene configurations are based on the wild type sequence for the gene, and are consequently loaded into Mutation Analyser and used as the basis for alignment with sample sequences and consequent detection of deviations.

Gene configurations can be based on genomic DNA or mRNA/cDNA. See Section 6.13 for three examples of gene configurations.

6.1 Creating a new gene configuration

1. Start the application, select the Configuration Editor program from the Windows 95 Start menu.

2. Select **File:New Configuration**. The **Configuration Type** dialogue is automatically displayed in which you must select either **Genomic DNA** or **mRNA/cDNA** corresponding to the configuration you want to create.

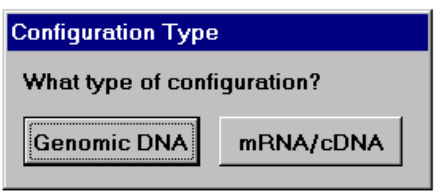


Figure 6-1. *Configuration type dialogue.*

3. An empty **Overview** window is displayed, called **Noname.cfg**.

4. Select **Edit:Edit Description** to open a dialogue that allows you to add a description that is valid for the whole gene configuration.

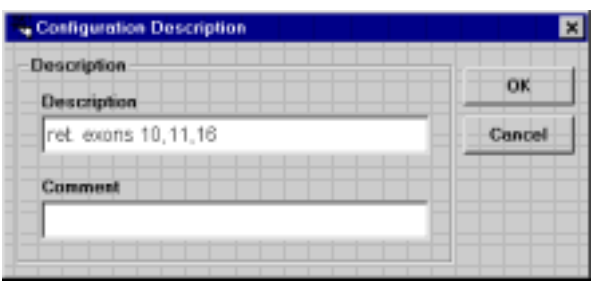


Figure 6-2. *Configuration description dialogue.*

6.2 Defining a region

A region is a defined sequence of up to 20,000 bases in length. The sequence is imported into Configuration Editor and is derived from several sources; typically from the international sequence databases accessible on the Internet or as an ALF/ALX file generated from ALF/ALF*express* sequencers. Within a region, specific objects that are relevant to the gene configuration can be added, such as sequencer fragments, exons, areas of interest and points of interest. For particularly large or complex genes, the configuration can be split into several regions (see Section 6.6).

6.2.1 Importing an ALF/ALX sequence

1. Select **File:Add Region** and choose **ALF/ALX** from the submenu options. The **Import ALF/ALX Sequence** dialogue is displayed for the appropriate file format.

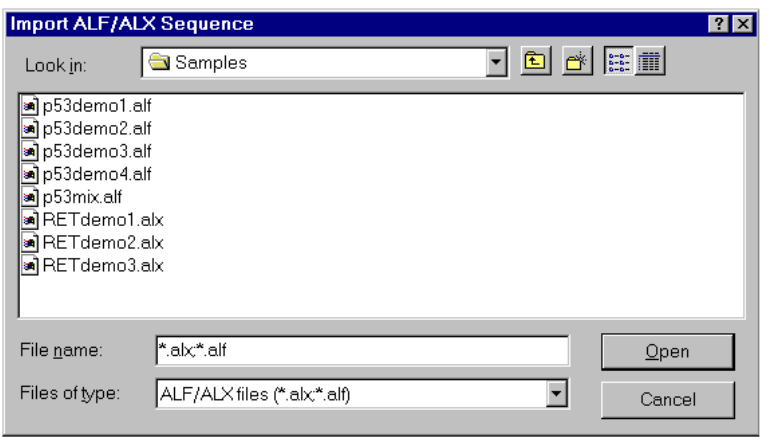


Figure 6-3. *Import dialogue for an ALF/ALX sequence.*

Locate the file containing the sequence and click on **OK**.

3. If the sequencer fragment is an AntiSense strand, check the **Anti-Sense Strand** box.

   Check the **Reference Fragment** box if the raw data for the selected sample is to be viewed in the **Raw Data Curves** field during mutation analysis (see Section 4.6.5).

Figure 6-4.*ALF/ALX file dialogue. An ALX demo sample for exon 10 of the RET gene is selected.*

4. Click on **Add Region**.

5. The **Define Region** dialogue is displayed.

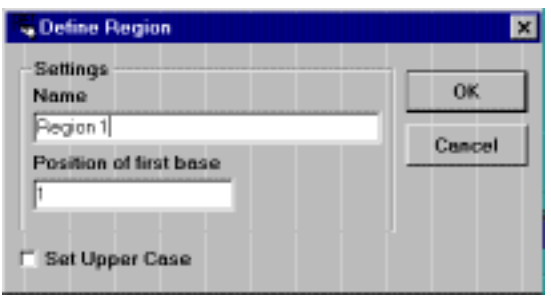


Figure 6-5. *Define Region dialogue for a genomic DNA configuration.*

Enter a name for the region and the position of the first base in that region, usually number '1'. If the sequence consists entirely of lower case letters, as obtained from some databases, these could be converted to uppercase by checking the **Set Upper Case** option. Do *not* check this option if you are able to define exons based on the case of the bases (see Section 6.4.3).

If you are defining a gene configuration for mRNA/cDNA, you can also add information here for amino acid sequence generation (see Section 6.7.4).

6. A sequencer fragment will be automatically generated from the ALF/ALX sequence corresponding to the defined region.

   However, if the imported file contains ambiguity codes, the **Ambiguities in Fragment** dialogue is automatically displayed. This dialogue shows all of the detected ambiguities in the **Ambiguities Found** field and allows you to reassign or exclude these bases accordingly.

   - The **Adjust Base** field allows you to select a base for the highlighted ambiguity and replace it by clicking on **Replace**. If you select the wildcard character ($) and then click on **Replace**, the **Adjusted Code** can represent any one of the four bases A,C,G,T.

   - You can exclude sequence data up to and including the selected ambiguity by selecting **Cut From Beginning** and then clicking on **Cut**. Alternatively, you can exclude all ambiguities from the currently selected position by selecting **Cut To End** and then clicking on **Cut**.

Note: **Cut** does not remove or delete sequence data, but instead affects the extent of the sequencer fragment to be defined.

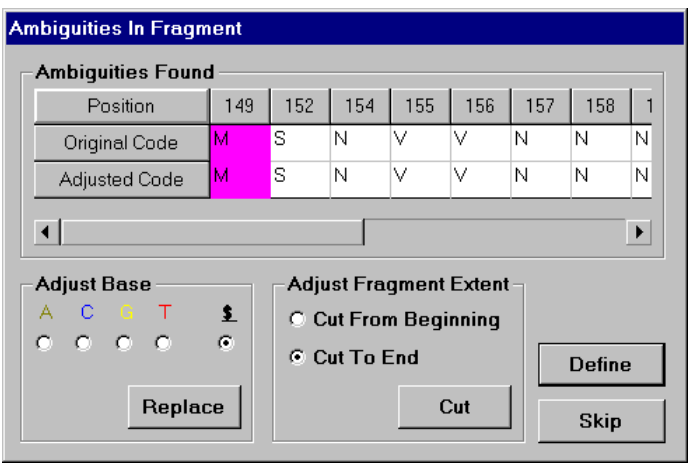


Figure 6-6. *Ambiguities in Fragment dialogue.*

7. Click on **Define** once you have made the appropriate adjustments. If there are still ambiguities present within the remaining fragment extent, a warning message will be displayed indicating that a sequencer fragment cannot be automatically created while ambiguities remain.

Alternatively, click on **Skip** to define the region inclusive of all ambiguities, although no sequencer fragment will be automatically produced.

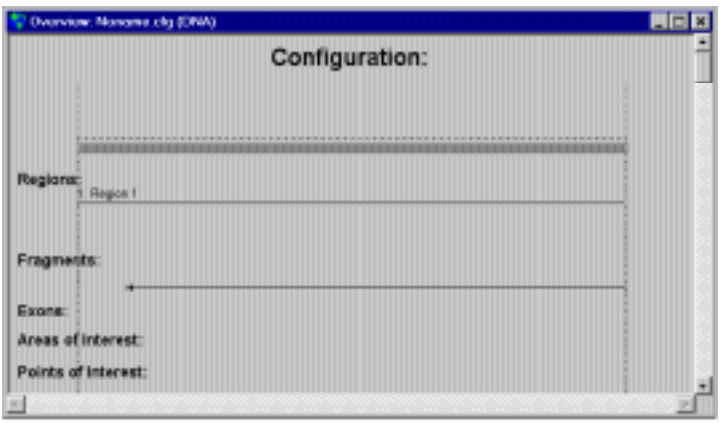


Figure 6-7. *Overview window with defined region and automatically generated sequencer fragment based on an ALX sample for exon 10 of the RET gene. Note that the sequencing direction of sequencer fragment is indicated with an arrow.*

8. The sequence is imported into the **Overview** window and displayed at the lowest magnification. Increased magnification will display more detail (see Section 6.3.1).

6.2.2 Importing GenBank sequences

1. Select **File:Add Region** and choose **GenBank** from the submenu options. The **Import GenBank Sequence** dialogue is displayed for the appropriate file format.

   Select the appropriate file and click on **OK**.

2. The **GenBank Import** dialogue is displayed.

   This presents information stored in the GenBank file imported from the GenBank database via the Internet. The information contained within a GenBank file may sometimes be incomplete so you should exercise care when defining a region.

   In the **GenBank Import** dialogue you can define one or more regions, depending on the number of sequences contained within the file, and also define exons based on the supplied start and end positions.

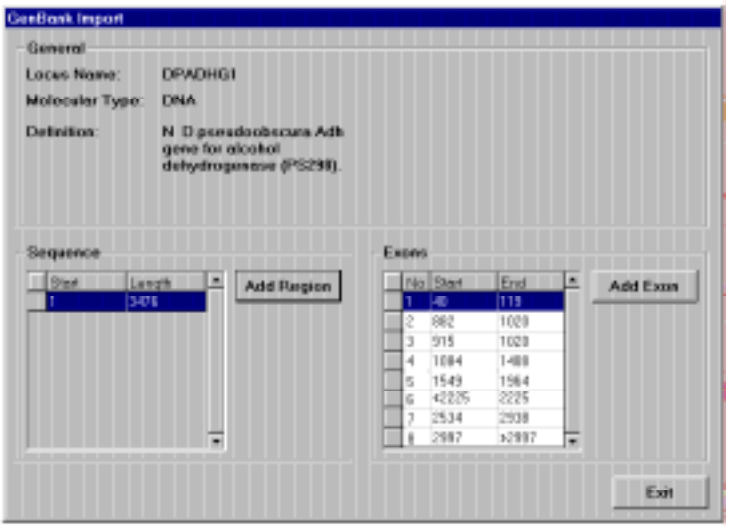


Figure 6-8. *GenBank Import dialogue.*

3. Select a sequence and click on **Add Region**. The **Define Region** dialogue is displayed.

   Enter a name for the region and the position of the first base in that region, usually number '1'. If the sequence consists entirely of lower case letters, as obtained from some databases, these could be converted to uppercase by checking the **Set Upper Case** option. Do *not* check this option if you are able to define exons based on the case of the bases (see Section 6.4.3).

   If you are defining a gene configuration for mRNA/cDNA, you can also add information here for amino acid sequence generation (see Section 6.7.4).

   Click on **OK**. The **Overview** window is updated with the sequence information. The **GenBank Import** dialogue also remain open.

4. Some GenBank files contain **Exon** data. The exon number is displayed together with the **Start** and **Stop** base position delimiting the extent of the exon.

   Select an exon and click on **Add**. The **Add Object** dialogue is displayed to define the exon. If the added exon has a number in the **GenBank Import** dialogue, this will be displayed in the **Exon No** field of the **Add Object** dialogue. Similarly, the **Start** and **Stop** positions will be defined in the **Extent** field.

Figure 6-9. *Add Object dialogue for an exon added from the GenBank Import dialogue.*

In some circumstances, the exon number or start/stop positions may not be available in the **GenBank Import** dialogue, and are thus not entered into the **Add Object** dialogue. These must be entered manually.

For more information on adding exon objects, see Section 6.4.

5. Repeat step 4 to add other exons. These will be added to the gene configuration as you define them.

6. Click on **Exit** in the **GenBank Import** dialogue when you have added all relevant information.

6.2.3 Importing ASCII file sequences

Prior to import, sequence data obtained from the Internet should be viewed in an appropriate text editor application on your PC. Superfluous information such as the title and other descriptive information should be deleted leaving only the sequence data and possible base position numbering information.

1. Select **File:Add Region** and choose **Plain ASCII** from the submenu options. The **Import ASCII Sequence** dialogue is displayed for the appropriate file format.

   Select the appropriate file and click on **OK**.

2. Select a sequence and click on **Add Region**. The **Define Region** dialogue is displayed.

   Enter a name for the region and the position of the first base in that region, usually number '1'. If the sequence consists entirely of lower case letters, as obtained from some databases, these could be converted to uppercase by checking the **Set Upper Case** option. Do *not* check this option if you are able to define exons based on the case of the bases (see Section 6.4.3).

   If you are defining a gene configuration for mRNA/cDNA, you can also add information here for amino acid sequence generation (see Section 6.7.4).

   Click on **OK**. The **Overview** window is updated with the sequence information.

6.3 Navigating the region

There are several tools available that allow you navigate the region and view the sequence data along.

6.3.1 Using the zoom function

The zoom function allows you to magnify an area of the **Overview** window contents so that the individual bases in the configuration sequence can be visualised.

Increasing magnification

To successively *increase the magnification*, first select the region of interest and then repeatedly select **View:Zoom In** or **Zoom In** located on the right mouse button menu, or press <F7> on the keyboard. With increased magnification, the individual coloured bases in the configuration sequence can be visualised. A scale bar spanning the whole region can also be seen. If there are ambiguity codes in the sequence, these are coloured black.

Reducing magnification

To successively *reduce the magnification*, repeatedly select **View:Zoom Out** or **Zoom Out** located on the right mouse button menu, or press <F8> on the keyboard.

To return to the default (lowest) magnification, select **View:Full Zoom Out** or **Full Zoom Out** located on the right mouse button menu, or <F6> on the keyboard.

Note: If your gene configuration contains more than one region, the full zoom out function will show an overview of all regions (see Section 6.6.2).

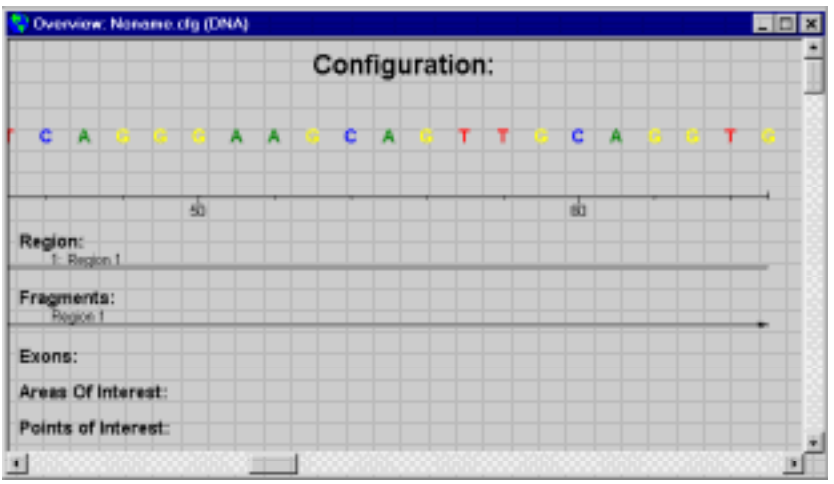


Figure 6-10. *Overview window with full magnification.*

6.3.2 Finding a sequence

You can use the find function to find a string of bases in the sequence.

1. Use the zoom function (see Section 6.3.1) to maximise the magnification in the window for optimal visualisation of the individual bases.

2. Select **Edit:Find** and the **Find** dialogue is displayed.

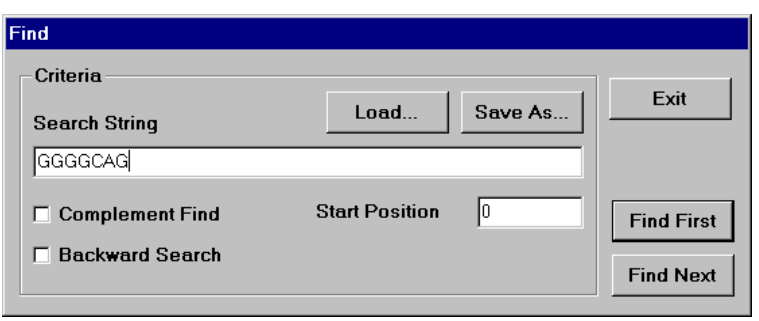


Figure 6-11. *Find dialogue.*

Enter a search string, either by directly entering a sequence into the **Search String** field, or by loading a previously saved string by clicking on the **Load** button.

Any search string can be saved for future use by clicking on the **Save As** button and giving an appropriate name.

2. Define any other search criteria that you want:

Complement Find	Check this box if you want to search the complementary sequence using complementary bases to those in the **Search String** field.
Backward search	This looks for the search string beginning at the end of the sequence.

3. Click on **Find First** and if the string is present in the sequence, its first occurrence will be detected and the **Start Position** shown. The **Start Position** always shows the detected position within the configuration sequence with respect to the first base in the search string. Click on **Find Next** to locate the next occurrence of the string. The **Overview** window is automatically centred on the located string.

4. Click on **Exit** when the appropriate **Position** value is obtained.

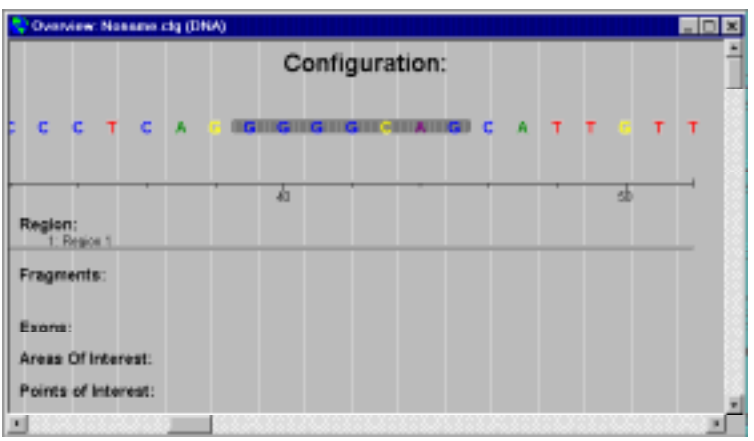


Figure 6-12. *Overview window centred on the search string used in the Find dialogue, e.g. GGGGCAG.*

6.3.3 Defining gene configuration objects

The overview mode can be used to directly define the extent of bases for objects to be added to the gene configuration (see Section 6.4). This is particularly useful for defining areas of interest or points of interest.

1. Locate the base(s) of interest using the **Find** function (see Section 6.3.2).

2. To define a string of bases to create an object, hold down the left mouse button and drag the cursor over the appropriate bases. For longer sequences, select the first base and click on the <Shift> key when selecting the last base to select all bases in between.

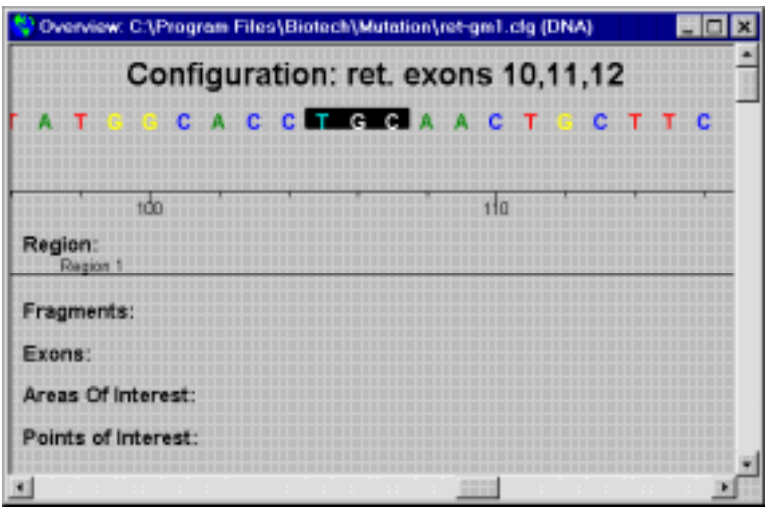


Figure 6-13. *Selected sequence, TGC, defining the extent of an object to be created.*

3. Select **Edit:Add Object**. The **Object** dialogue is displayed and the extent values of the selection are shown (see section 6.4).

4. Enter the other variables for the object, e.g. type and name of object (see Section 6.4 for details).

6.3.4 Adjusting bases in the sequence

You can edit or delete selected bases in the sequence or insert new bases.

1. Select the relevant base in the sequence and then select **Edit Sequence** from the right mouse button menu. The **Edit Sequence** dialogue is displayed.

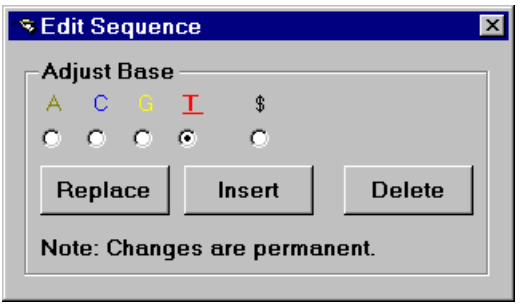


Figure 6-14. *Edit Sequence dialogue.*

2. To replace the base, select an alternative base A,C,T,G or $ (can be any one of the four bases) and click on **Replace**.

   To delete the selected base, click on **Delete**.

   To insert a base, make the appropriate selection and click on **Insert**. The base is inserted *after* the currently selected base.

> **Warning**
> This function should be used with careful consideration since no record is made of the base editing that you perform.

6.4 Adding objects to the gene configuration

Objects can be added to the gene configuration such as sequencer fragments, exons, areas of interest and points of interest.

> **Attention**
> It is important that at least one sequencer fragment is defined, either automatically or manually, in order for the gene configuration to be functional for analysis purposes.

6.4.1 Adding a new object:

1. Select **Edit:Add Object**. The **Add Object** dialogue is displayed.

2. Select the type of object to be created, either **Exon**, **Sequencer Fragment**, **Area of Interest** or **Point of Interest**. This will affect the other fields displayed in the dialogue.

3. Enter the appropriate information for the selected object (see below).

4. Click on **Add** to include the new object in the gene configuration.

5. Repeat steps 2 to 4 for other objects you want to add.

6. Click on **Exit** once you have added all objects. The **Overview** window will display all of the objects that you have added.

Adding a sequencer fragment object
This is valid for regions where there was no automatic generation of a sequencer fragment resulting from using an ALF/ALX sample (see Section 6.2).

Many mutation analysis methods are based on sequencing both the Sense and AntiSense strands. It is recommended that you define two separate sequencer fragment objects, one for each direction. It may be convenient, for example, to have the option to separately adjust the extents of both objects later on when the method is being optimised. It also aids documentation of the method.

Figure 6-15.*Add Object dialogue for a sequencer fragment.*

Sequencer fragments are necessary for the analysis process to function in Mutation Analyser. The sequencer fragments also define and document the evaluation method.

1. Enter a name for the sequencer fragment.

2. Enter the **Extent** position values for the object (see Section 6.4.2).

3. If the object is an AntiSense strand, check the **Anti-Sense Strand** box.

4. Optionally, define a reference ALF/ALX sample corresponding to this sequencer fragment that will be displayed in the **Raw Data Curves** field during mutation analysis (see Section 4.6.5). Click on the **Browse** button to locate the relevant reference file and sample. The pathway to this file is updated in the **Ref. Sample File** field.

Adding an exon object
Exon objects may already have been added if you defined a region using GenBank sequence data (see Section 6.2.2).

Figure 6-16. *Add Object dialogue for an exon.*

1. Enter the **Exon no**.

2. Enter the **Extent** position values for the object (see Section 6.4.2). If the sequence is derived from an Internet database and utilises lower case/upper case letters for the definition of introns and exons, exon objects can be easily added using the functions described in Section 6.4.3.

3. Enter the **Coding Sequence Information** for automatic generation of the amino acid sequence (see Section 6.7).

Note: An exon is a pre-requisite for obtaining amino acid information during analysis using genomic DNA gene configurations. cDNA/mRNA configurations, by contrast do not necessarily need defined exons to show the amino acid sequence (see Section 6.7).

Adding an area of interest object

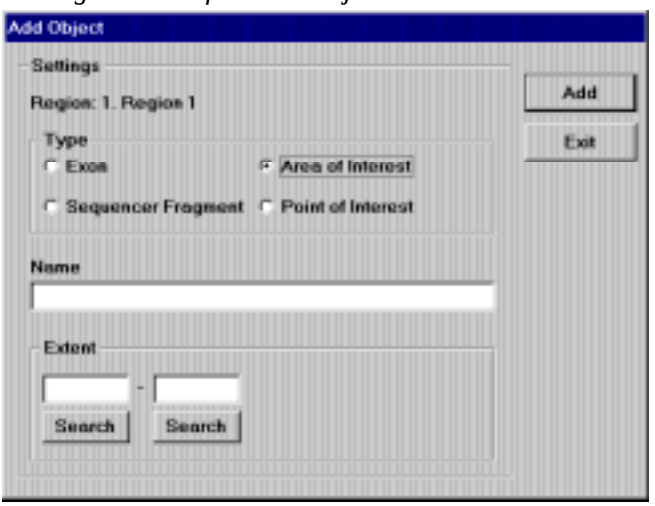


Figure 6-17. *Add Object dialogue for an area of interest.*

1. Enter a name for the area of interest.

2. Enter the **Extent** position values for the object (see Section 6.4.2). This is particularly useful for visualising specific codons.

Adding a point of interest object

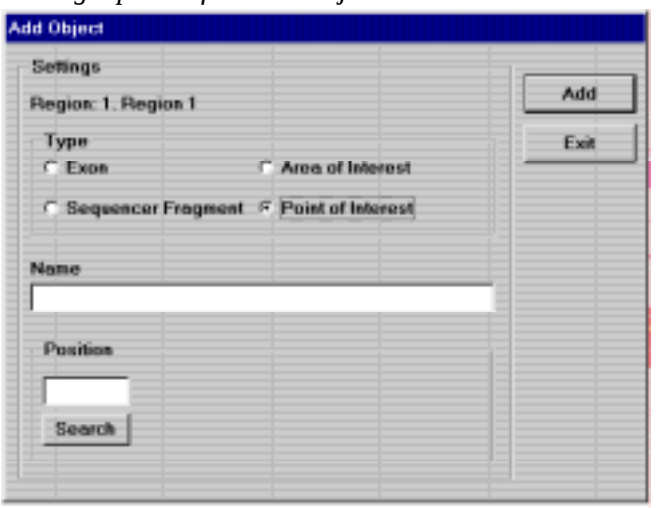


Figure 6-18. *Add Object dialogue for a point of interest.*

1. Enter a name for the point of interest.

2. Enter the **Position** value for the object (see Section 6.4.2). This is particularly useful for defining polymorphic base positions or mutation hotspots. The **Name** field could then be used for information about known substitutions.

6.4.2 Defining the extent of objects

The extent of an object can be defined in four ways, by:

- selecting the base range in overview mode (see Section 6.3.3)
- manually entering the extent values in the **Object** dialogue
- utilising the lower case/upper case sequence definition (*exons only*, see Section 6.4.3).
- using the embedded **Search** function to locate the relevant positions

To use the **Search** function:

1. Click on the appropriate **Search** button and the **Find** dialogue is displayed.

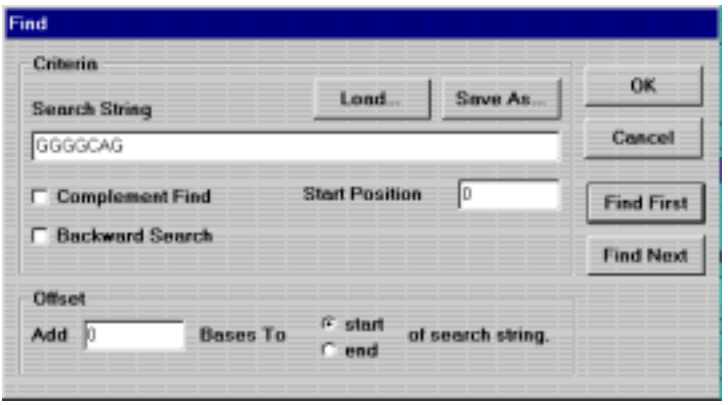


Figure 6-19.*Search dialogue.*

2. Enter a search string, either by directly entering a sequence into the **Search String** field, or by loading a previously saved string by clicking on the **Load** button.

   Any search string can be saved for future use by clicking on the **Save As** button and giving an appropriate name.

3. Define any other search criteria that you want:

Complement Find	Check this box if you want to search the complementary sequence using complementary bases to those in the **Search String** field.
Backward search	This looks for the search string beginning at the end of the sequence.
Start Position	You can define the base position number from where the search begins. After the search, the base position in the located sequence corresponding to the first base in the search string is shown.
Offset	**Add xx Bases To** allows you to add (or subtract if a negative value is given) a specified number of base positions relative to the **start of** or the **end of the search string**. After the search is complete. This value will be used to define one of the extent values for the new object.

4. Click on **Find First** and if the string is present in the sequence, its first occurrence will be detected and the **Start Position** shown. The **Start Position** always shows the detected position with respect to the first base in the search string. Click on **Find Next** to locate the next occurrence of the string.

5. When you have obtained the correct **Start Position** value, click on **OK**. The value will be automatically entered in the appropriate **Extent** field for which you performed the search and will be the same as the **Start Position** value, unless you specified other criteria in the **Offset** field.

6.4.3 Defining exons based on case

Some databases accessible via the Internet supply sequences that classify intron regions in one case (upper or lower) and exon regions in the other case. This classification allows exon regions to be easily defined and be added as objects to the gene configuration.

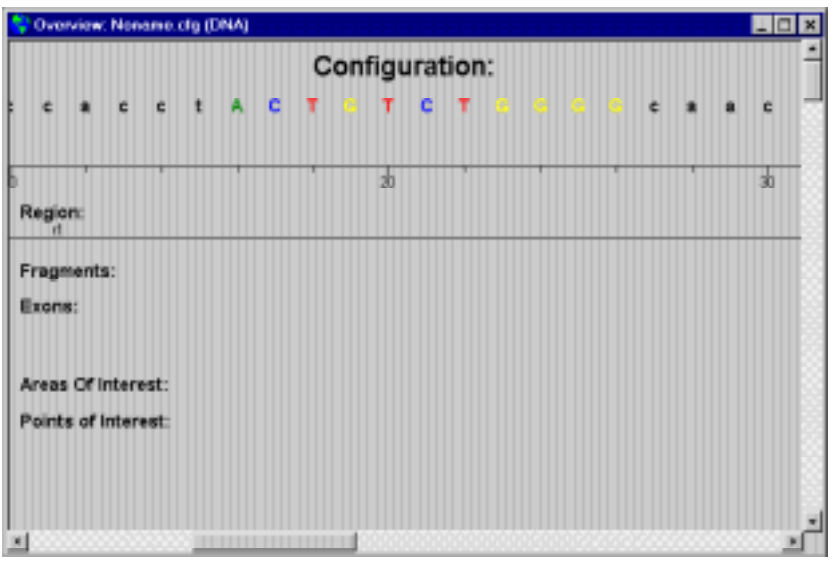


Figure 6-20. *Example of an imported sequence containing an exon defined by upper case bases, flanked by lower case letters defining intron data.*

If the imported sequence is classified as described:

1. Display the **Overview** window and select the relevant region.

2. Select **Edit:Auto-Define Exons**. The **Auto-Define Exons** dialogue is displayed.

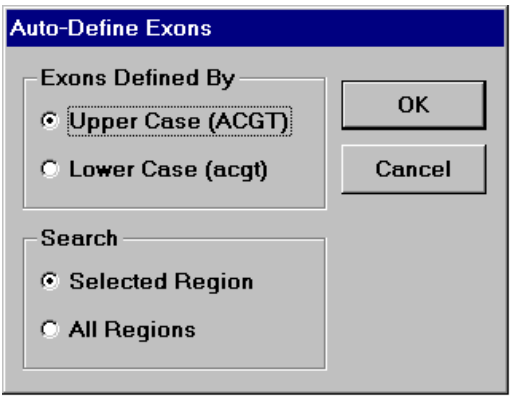


Figure 6-21. *Auto-Define dialogue.*

3. Select the case by which the exons are defined in the sequence, either **Upper Case** or **Lower Case**.

4. Select if the search should include only the selected region or all regions if there are more than one. Click on **OK**.

5. The **Add Object** dialogue displays the extent of the first detected exon region.

Figure 6-22. *Add Object dialogue.*

6. Enter the **Exon no**.

7. Click on **Add** to include this object in the gene configuration.

8. Click on **Next** to locate each exon in the region(s) and add them accordingly.

   If you click on **Next** after the last exon has been located you will be informed that there are no more exons and the dialogue will automatically close. Click on **Quit** to close the dialogue at any other time.

9. Added exons are displayed in the **Overview** window.

6.4.4 Including mutation frequency statistics

1. Select a region.

2. Select **File:Add Mutation Statistics**. The **Import Mutation Statistics** dialogue is displayed. Select the relevant file and click on **Open**.

   The statistics are presented as a bar histogram on the base position scale.

Note: Mutation statistics should be stored in tabbed ASCII files, one for each region. Each file should contain the position numbers of the bases and the corresponding frequencies in two tab-separated columns. These kinds of files can be easily created in, for example, Microsoft® Excel.

6.5 Selecting, editing and deleting objects

6.5.1 Selecting an object

Click on an object in the **Overview** window to select it. Alternatively, select **Edit:Select Next Object** to move consecutively forward through the objects, or **Edit:Select Previous Object** to consecutively move backwards through objects. The selected object is coloured blue.

6.5.2 Editing an object

To edit an object, select it and then select **Edit:Edit Object** or **Edit Object** from the right mouse button menu. The **Edit Object** dialogue is displayed.

6.5.3 Deleting an object

To delete the selected object, select **Edit:Delete Selected Object** or **Delete Object** from the right mouse button menu, or press <Del> on the keyboard.

6.6 Working with several regions

You can add more regions to the configuration sequence up to a maximum of 100 regions, each containing up to 20,000 bases. Regions are a convenient way of splitting particularly large or complex sequences into manageable units. Separate regions are also used when you want to aggregate non-contiguous sequence data in one configuration.

6.6.1 Adding new regions

New regions are created exactly as for the first region. In brief:

1. Select **File:Add Region** and the relevant file format from the submenu options. The appropriate **Import Sequence** dialogue is displayed. Select the sequence (see Section 6.2) for the configuration in that region and click on **OK**. Create a region based on the specific details for the sequence file format that you are using (see Section 6.2).

Note: Regions are sequentially added and numbered within the gene configuration in the order that you add new regions. See Section 6.6.3 if you want to change the order of the regions.

3. Build up the gene configuration by adding objects (see Section 6.4).

6.6.2 Viewing the whole configuration

When there is more than one region, the whole gene configuration can be viewed by selecting **View:Full Zoom Out** or **Full Zoom Out** from the right mouse button menu.

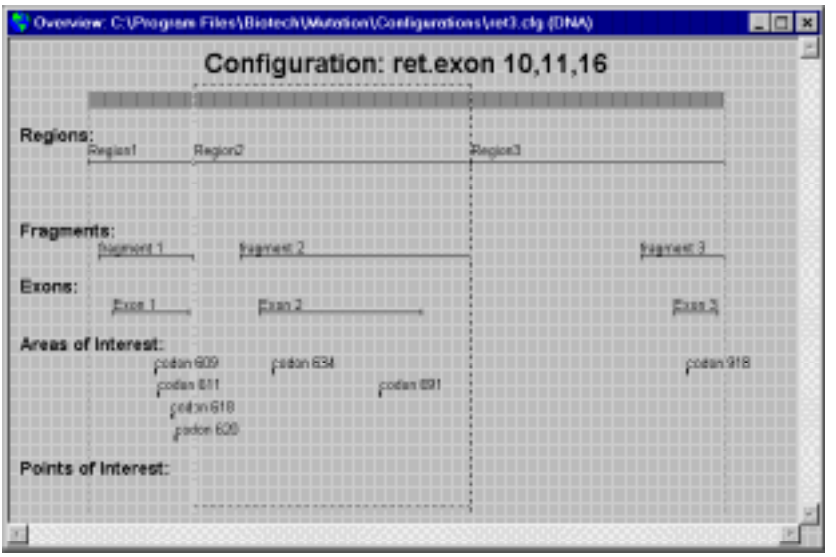


Figure 6-23. *Overview window containing three regions defined for the Ret exons 10, 11, and 16 respectively. Note that Region 2 is surrounded by a box, which denotes that it is active.*

The active region is surrounded by box. Use the Zoom function to view only the selected region (see Section 6.3.1).

6.6.3 Rearranging regions

The regions are automatically arranged in the order that you add them to the gene configuration. You can rearrange this order:

1. Select **Edit:Rearrange Regions**. The **Overview** window is automatically displayed with the minimum magnification to show all of the regions.

   Each of the regions has a large identifier number indicating its current placement in the order.

2. Click on the identifier number that you want to change to be the *first region* in the gene configuration. Its number becomes ‘1’.

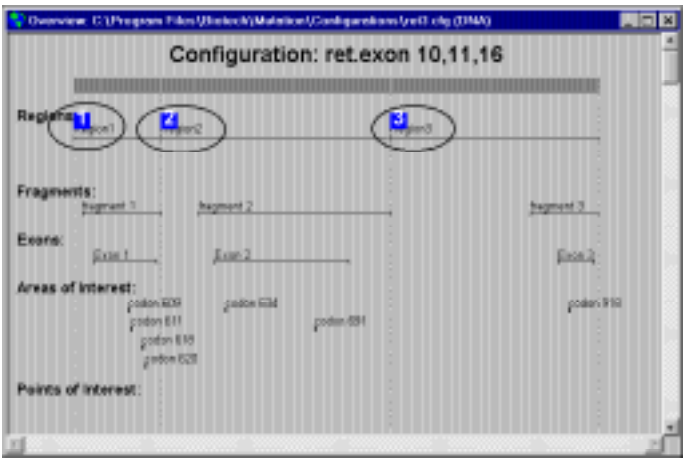


Figure 6-24. *Overview window displaying the original order of the three regions, 1, 2, 3, created for the Ret gene in the gene configuration.*

3. Click on the other identifier numbers in sequential order that you want the regions to be placed.

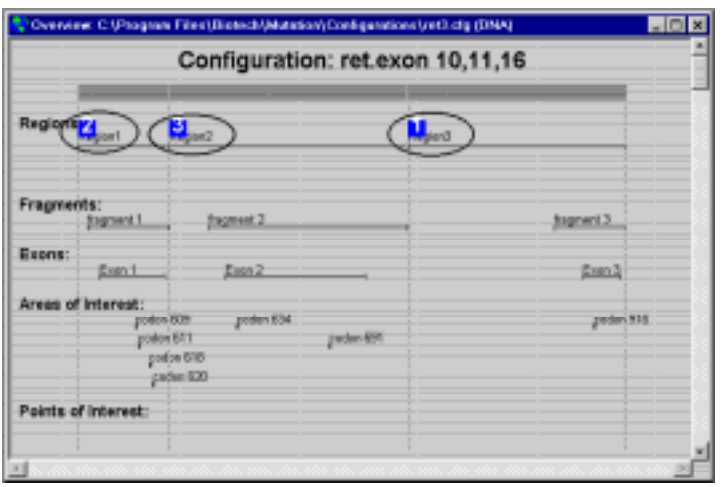


Figure 6-25. *Overview window displaying a newly specified order of the three regions, 1,2,3, created for the Ret gene in the gene configuration.*

4. Select **Edit:Rearrange Regions** once more (to remove the check mark) and the regions are rearranged according to your selection.

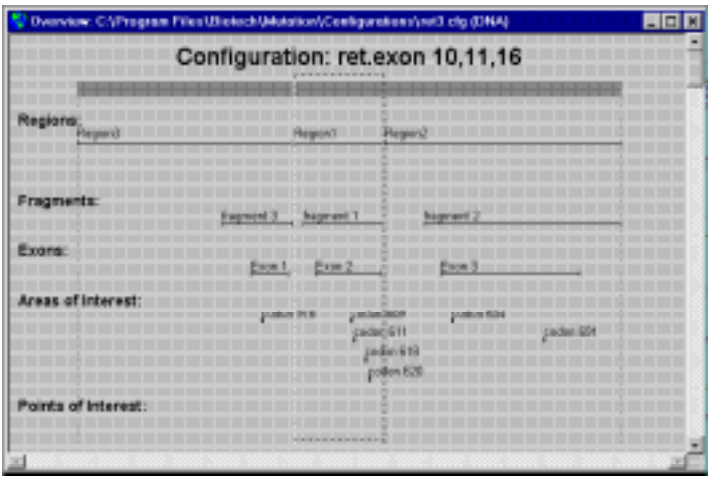


Figure 6-26. *Overview window displaying the rearranged order of the three regions, 1,2,3, based on the previous figure.*

5. If you want to change the name of individual regions as a result of the rearrangement, select the region and then select **Edit:Edit Region Settings**. The **Define Region** dialogue is displayed.

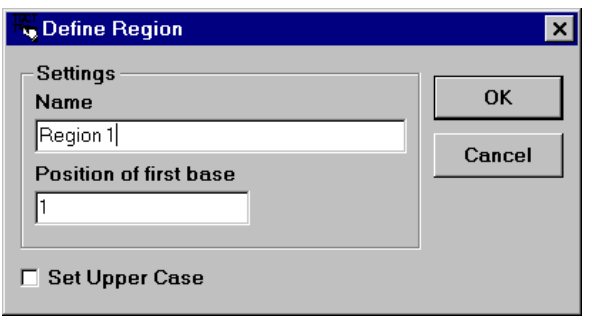


Figure 6-27. *Define Region dialogue for a genomic DNA gene configuration.*

6.6.4 Deleting a region

To delete a region, first make sure that it is active and then click on **File:Delete Selected Region**.

6.7 Defining the amino acid sequence

Configuration Editor can automatically calculate for the amino acid sequence for gene configurations.

For genomic DNA configurations, this uses the coding sequence bases contained within the defined exons of the gene configuration. You need to supply the start position to the coding sequence in the correct reading frame for the first appropriate exon. The amino acid sequence for that exon and all contiguous exons will be automatically generated (see section 6.7.1). It is also possible to account for non-contiguous exons in a gene configuration (see section 6.7.2).

For mRNA/cDNA sequences, you can define the amino acid sequence based on exons as for genomic DNA, but a simpler method is to use the Define Region dialogue (see Section 6.7.4).

6.7.1 Contiguously numbered exons

1. Add the exon objects (see Sections 6.4.1 and 6.6.1) for all regions (see Section 6.6.1) in the gene configuration.

2. Select the first exon containing the coding sequence in the gene configuration.

Note: Some upstream exons in a gene may not contain coding sequence bases but instead are used in site recognition etc.

3. Use the Find function to locate the first base of the start codon in the selected exon. Briefly, use the Zoom function to maximise the magnification of the **Overview** window for the selected region (see Section 6.3.1). Select **Edit:Find** to display the **Find** dialogue. Enter the appropriate search string for the start codon, often ATG. Click on **Find First** and **Find Next** until the appropriate start codon string is located in the sequence. Note down the **Start Position** value from the **Find** dialogue and click on **Exit** (see Section 6.3.2).

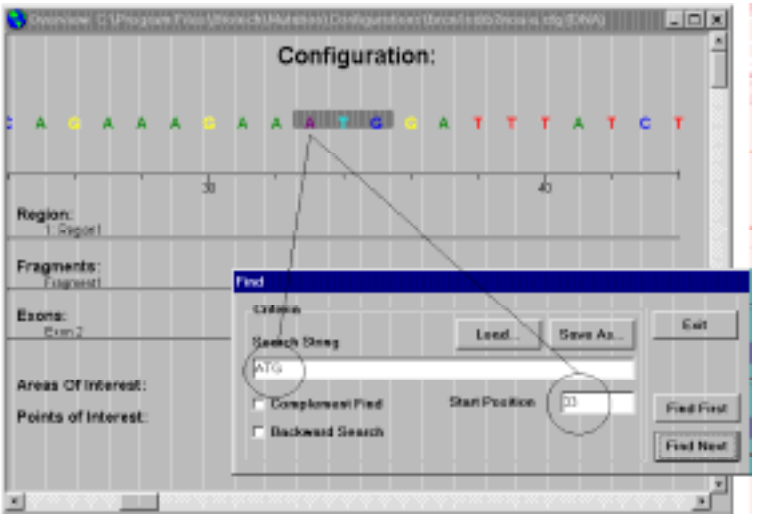


Figure 6-28. *Example of using the Find dialogue to locate the appropriate start position of the start codon for the coding sequence.*

4. Select **Edit:Edit Object** or select **Edit Object** from the right mouse button menu. The **Edit Object** dialogue is displayed for the selected exon.

5. In the **Coding Sequence Information** field, enter the start **Codon no.** and also the **Position** value that you obtained in step 3.

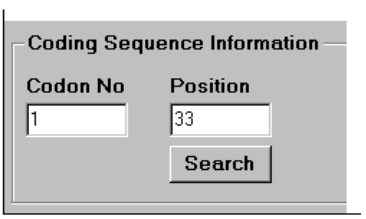


Figure 6-29. *Coding Sequence Information field in the Edit Object (exon) dialogue defining the first codon and coding position.*

7. Click on **OK** in the **Edit Object** box and the amino acid sequence for all contiguous exons will be automatically calculated and displayed.

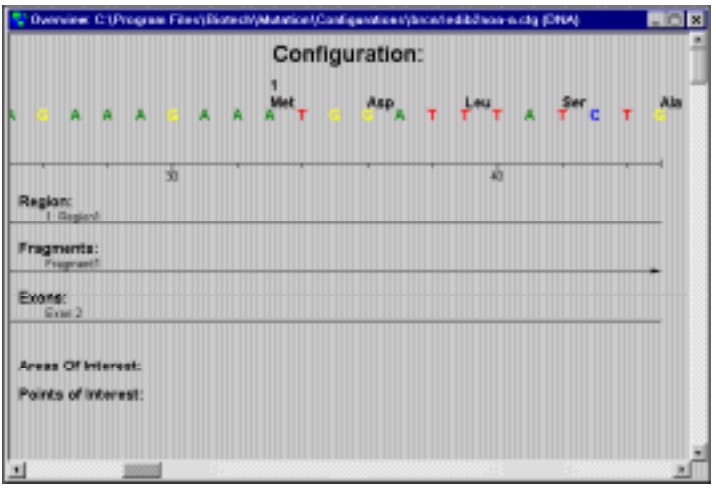


Figure 6-30. *Overview of Region 1 of the BRCA1 gene configuration. The amino acid sequence is generated based on the defined start position within Exon 2. Note that it is Exon 2 and not Exon 1 that contains the start to the coding sequence.*

With reduced magnification, the extent of the amino acid sequence is seen on other contiguous exons.

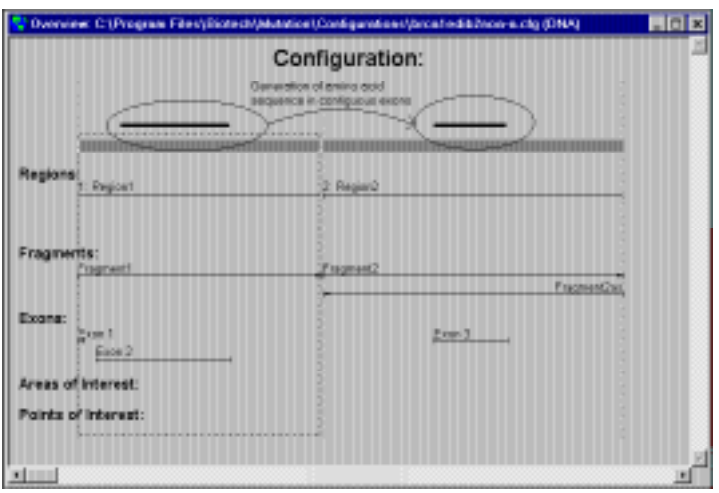


Figure 6-31. *Overview of Regions 1 and 2 of the BRCA1 gene configuration. The amino acid sequence is automatically generated along contiguous exons, in this case for Exons 2 and 3.*

6.7.2 Non-contiguously numbered exons

In some gene configurations, sequence information for all exons may not be available.

The gene configuration BRCA1 is an interesting example because:

- Exon 2 rather that Exon 1 contains the first coding sequence information.
- It contains 24 exons, of which sequence information for the non-coding Exon 4 is unavailable.

This means that Exon 2 is used as the first exon to define the coding sequence. Moreover, when calculating the amino acid sequence from Exon 2, only Exon 3 is automatically included as it is contiguous with the previous exon. The next exon is Exon 5, which is non-contiguous with exon 3, thus Exons 5 to 24 are not included in the initial automatic amino acid sequence generation.

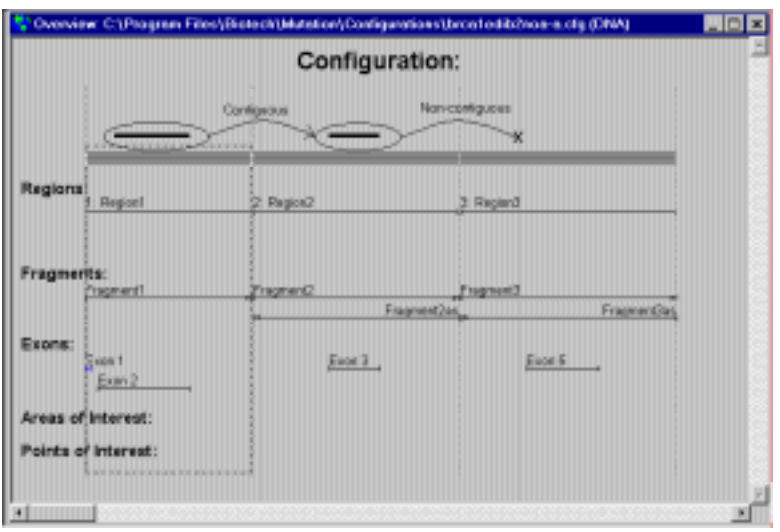


Figure 6-32. *Automatic amino acid sequence generation between contiguous BRCA1 Exons 2 and 3 but no further generation within the non-contiguous Exon 5 and beyond.*

To automatically generate the amino acid sequence in the other exons:

1. Select the first exon in the relevant region after the non-contiguous break, e.g. Exon 5 in Region 3 of the BRCA1 gene configuration. Select **Edit:Edit Object**.

2. In the **Coding Sequence Information** field, enter the **Position** from which the amino acid sequence calculation should be started. Take care to select the position giving correct reading frame. For example, in Exon 5 of the BRCA1 gene configuration, this is the second base position in the exon (position 72).

3. Enter the correct **Codon no.**

6. Click on **OK** in the **Edit Object** box and the amino acid sequence for all exons contiguous to the current exon will be automatically calculated and displayed.

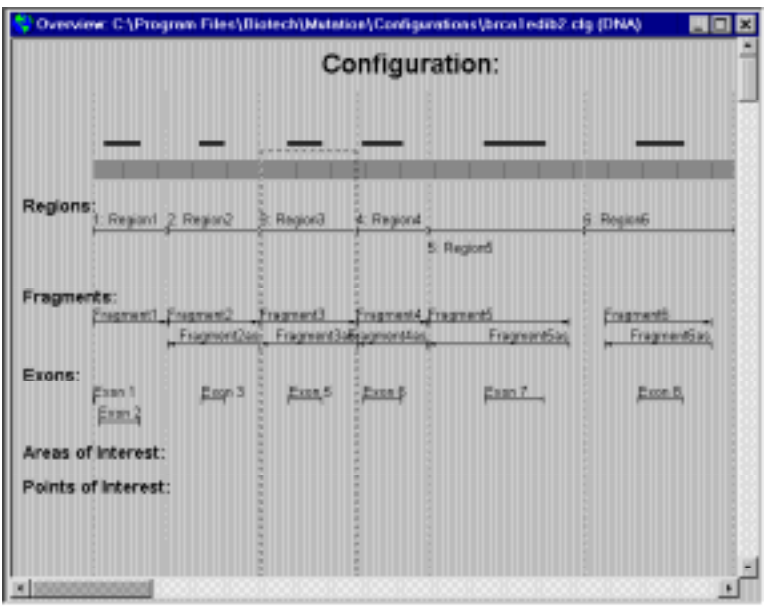


Figure 6-33. *Overview of Regions 1 to 5 of the BRCA1 gene configuration. The amino acid sequence for Exon 5 onwards have been automatically created.*

6.7.3 Selecting the wrong reading frame

If for some reason you select the wrong start position for generating the amino acid sequence, the calculation will be performed in the wrong reading frame. This should be clearly evident once the calculation has been made since the amino acid sequence across all of the exons will be prematurely terminated. This termination results from the random occurrence of stop codons produced in the wrong reading frame. Configuration Editor stops the process of generating an amino acid sequence when it meets the first stop codon, for example TAA, TAG or TGA in the Universal transcription mode (see Section 6.7.5), in the exon coding sequence.

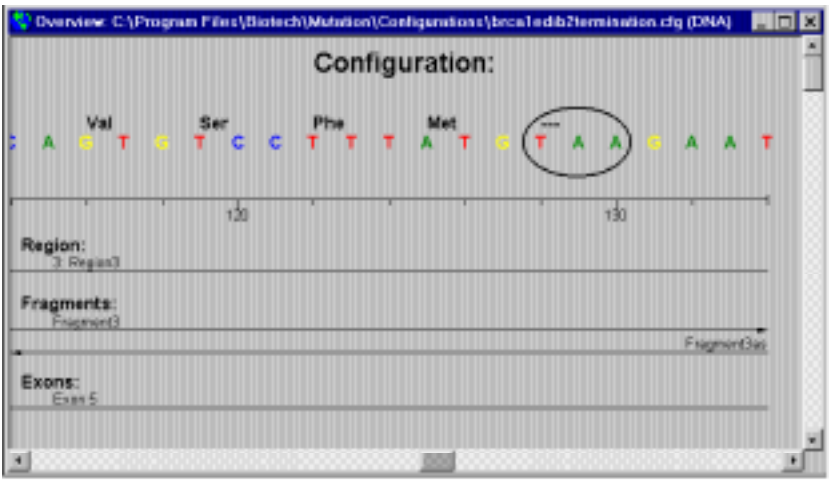


Figure 6-34. *Amino sequence of BRCA1 Exon 5 generated in the wrong reading frame leading to termination at the first occurring stop codon, TAA.*

6.7.4 Amino acid sequences for mRNA/cDNA configurations

Since mRNA and cDNA do not contain intron data, the gene configuration should contain abutted exons, thus giving an uninterrupted coding sequence. This means that the sequence data used to define the region can be directly used to generate the amino acid sequence.

1. Create a gene configuration for mRNA/cDNA. You should not define any exon objects.

2. Select **Edit:Edit Region Settings**. The **Define Region** dialogue is displayed.

3. Enter the **Codon No.** and the start **Position** number from which the amino acid sequence generation will begin.

4. Click on **OK** and the amino acid sequence is generated.

Figure 6-35. *Define Region dialogue for an mRNA/cDNA gene configuration.*

6.7.5 Codon translation options

There is unusual codon use in some mitochondrial DNA sources, such that mRNA transcripts can produce different amino acids for some codons compared to the normal ('universal') assignment in eukaryotic and prokaryotic cells. You can account for this unusual codon use by selecting the correct translation option. This will automatically update the amino acid sequence.

Select **Options:Codon Translation Options** and check the relevant DNA source from the sub-menu, i.e. **Universal** (used in most cases for normal assignment in eukaryotic and prokaryotic cells), **Human Mitochondria**, **Yeast Mitochondria** and **Neurospora Mitochondria**.

The following table shows the unusual codon use:

RNA Codon	*Universal*	*Human Mitochondria*	*Yeast Mitochondria*	*Neurospora Mitochondria*
UGA	Stop	Trp	Trp	Trp
CUU CUC CUA CUG	Leu	Leu	Thr	Leu
AUA	Ile	Met	Met	Ile
AGA AGG	ARG	Stop	Arg	Arg

Table 6-1. *Unusual codon use in mitochondria [4].*

6.8 Saving and loading a gene configuration

Use **File:Save Configuration** to save the gene configuration. Unsaved configurations are called **Noname.cfg**.

Save changes to the gene configuration using **File:Save Configuration** or create a copy of the configuration under a new name with **File:Save Configuration As**.

Previously saved configuration files can be selected and loaded from the dialogue displayed by selecting **File:Load Configuration**.

The four most recently referenced configurations can be loaded directly from the **File** menu.

6.9 Exporting a region

If you want to export the sequence data from a specific region as a text file:

1. Select the region.
2. Select **File:Export Region**. The **Export** dialogue is shown.
3. Enter a name for the exported file and click on **OK**.

6.10 Printing a report

You can print a report about the configuration you have created.

6.10.1 Configuration report options

1. Click on **Options:Report Options**.

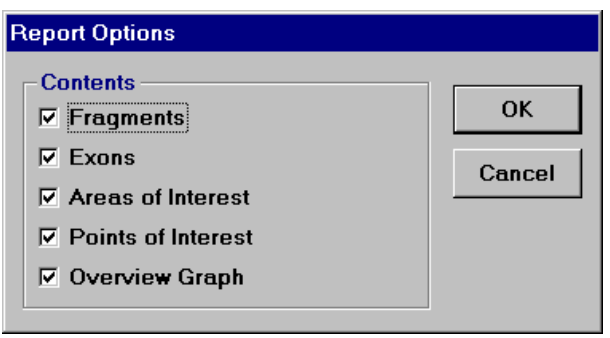


Figure 6-36. *Report Options dialogue*

2. In the **Report Option** dialogue, check the options you want to be incorporated in a report.

6.10.2 Printing the configuration

Select **File:Print Configuration Report** to obtain a printout of the configuration information and/or overview graph. You can also display the Windows 95 printer dialogues if you want to change the printer setup parameters.

6.11 Converting from a genomic DNA to a mRNA/cDNA configuration

If you have created a genomic DNA configuration with defined exons, you can convert this to a mRNA/cDNA configuration that contains only the exon information.

To make the conversion select **File:Convert to mRNA/cDNA Configuration**. In making the conversion you will be warned that sequence information between the exons will be lost in addition to other objects such as sequencer fragments, areas/points of interest and mutation statistics between the exons.

If the exons in a gene configuration are *non-contiguous*, as a general rule you will not be able to convert the whole configuration to mRNA/cDNA. Only those exons in the *first contiguous exons* for which an amino acid sequence has been generated will be converted.

In *special* cases only, such as for the BRCA1 gene configuration where the non-coding Exon 4 is not included, non-contiguous exons can be converted to a mRNA/cDNA configuration (see Section 6.7.2).

You are prompted to save converted configurations under a new file name.

6.12 Exiting the application

To exit the Configuration Editor application, select **File:Exit**.

6.13 Examples of some configurations

Presented here are three gene configurations:

6.13.1 Ret genomic DNA gene configuration

Points of interest objects detailing known mutation sites [5] have been included in this gene configuration. These can be used to generate hotspot deviations in the analysis and be included in the interactive confirmation process (see Section 5.2.4).

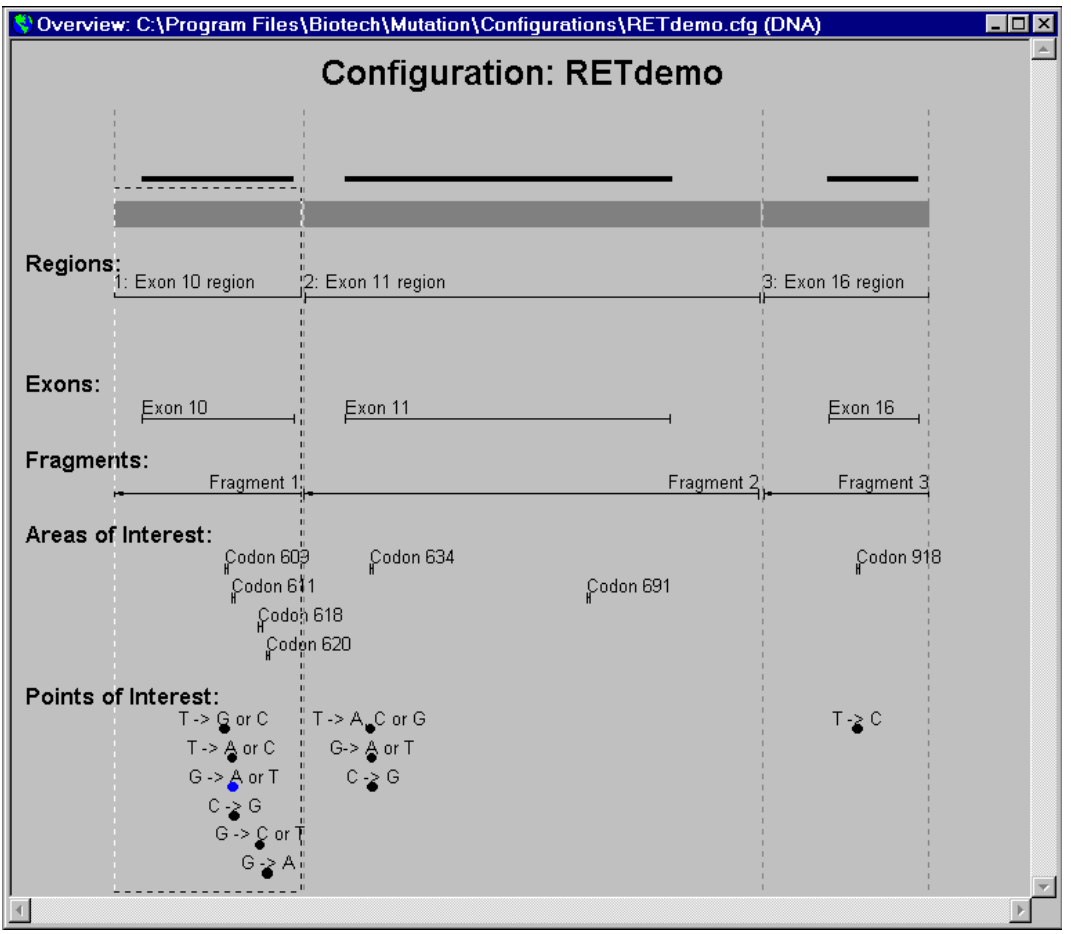


Figure 6-37. *Ret genomic DNA gene configuration comprising 3 regions, 3 exons, 10, 11, 16, and areas of interest for each exon.*

6.13.2 BRCA1 genomic DNA gene configuration

This gene configuration is divided into 22 regions and contains 24 exons. The non-coding Exon 4 is not included in the configuration. Exon 1 does not contain coding sequence so is not used to generate the amino acid sequence.

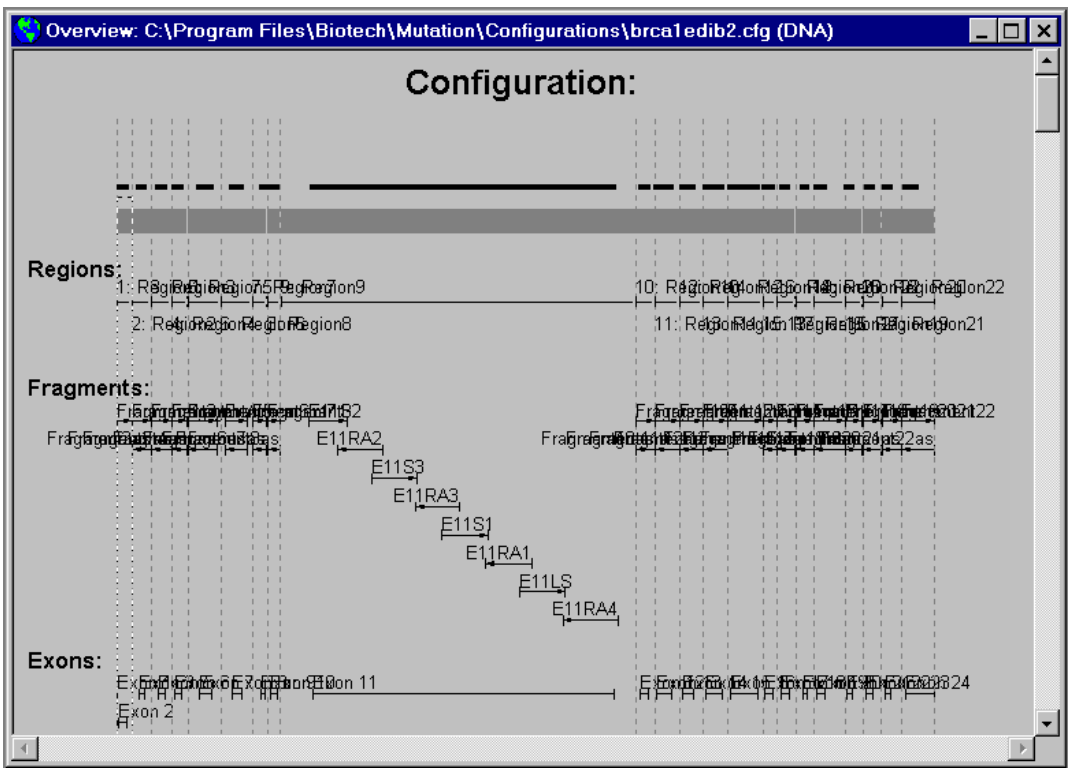


Figure 6-38. *BRCA1 genomic DNA gene configuration comprising 22 regions and Exons 1 to 24 excepting the non-coding Exons 1 and 4.*

6.13.3 p53 cDNA gene configuration

This configuration could easily be assigned amino acid data using the **Define Region** dialogue (see Section 6.7.4).

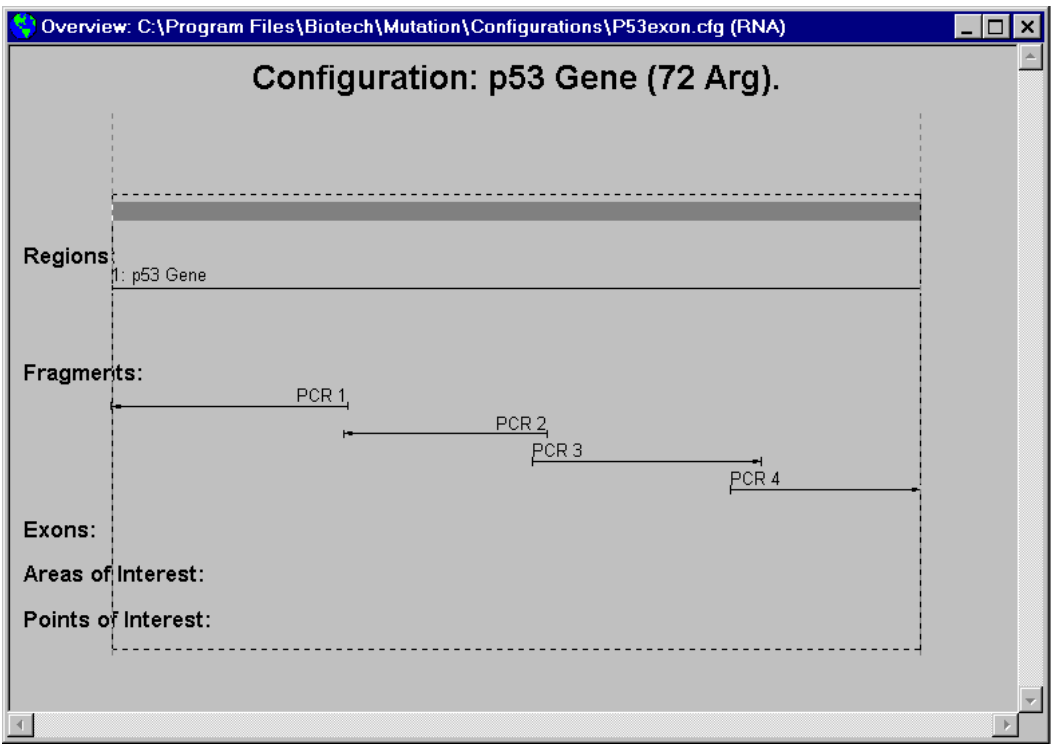


Figure 6-39. *p53 cDNA gene configuration comprising four sequencer fragments.*

7. Troubleshooting

This chapter should help you to identify and correct possible hardware problems associated with using Mutation Analyser. Messages displayed in Mutation Analyser and Configuration Editor that require special user actions are also included.

7.1 PC hardware

Symptom	Possible cause and/or corrective measures	Refer to
Screen blank	1. Check power cable and internal/external fuses 2. Check light/ contrast settings on CRT	CRT Screen manual
Message "Non-system disk or disk error" on start-up	Remove the diskette in the disk drive and restart	
Other start-up failure messages	Bad configuration or system failure	Check PC user's manual or contact local Amersham Biosciences Service Support

7.2 Printer

Symptom	Possible cause and/or corrective measures	Refer to
Communication failure	1. Check that the printer is On Line 2. Check communication cables	Printer. manual.
No message about paper when plotting	Print manager has to be open in the background and active	
Garbled or no printout	Check printer setup	

7.3 Software

7.3.1 Mutation Analyser

Message	Possible cause	Required action
"filename" Column information differs. Use existing column order instead?	When trying to append data to an existing database file and the "column order" was different to the existing file column format.	Press **OK** and the same format previously used in the file will be used. Otherwise the export to database operation will be cancelled.
“filename” does not exist.	When selecting a file from the most recently used (MRU) list that is no longer available.	Open the file from another directory or open a new file.
Can’t calculate “clonename”, sequence data not available.	When a sample is being analysed that does not contain sequence data.	Use a new sample.
Can't add more samples to this analysis item.	When out of memory.	Try closing other applications or restart the computer.
Can't open clone. No sequence data available. Checksum is incorrect, file has been modified, aborting load.	The file being opened is corrupt.	Corrupted files cannot be used. Recreate the source file to be opened.
Can't open internal file to present deviations.	During deviation calculation.	Try closing other applications or increase the number of file handles available (FILES in config.sys)

Can't open sample "filename and clone" Not able to open "filename". Couldn't open analysis item file. Error when opening gene configuration file.	File not found or already in use.	Place file in the appropriate directory or close this file if in use in another application.
Could not copy graph view to printer.	When printing a deviation report and the application cannot copy the graph image.	Free some disc space or memory. Empty the clip-board
Error when loading analysis result, no deviations loaded for instance "instance".	During deviation calculation.	Try closing other applications or increase number of file handles available (FILES in config.sys)
Error when opening gene configuration file "description" Error when reading analysis file "description"	When there was an unpredicted error during loading.	Try again.
Error when writing report to "filename" "description" Aborting.	When an unpredicted error occurred when writing a report to file or exporting result to a database file.	Try again.
Error while saving sample sequence changes. Error when saving edited sequence (+ description).	When saving revisions to the sample sequence.	Restart the computer.

Internal error (+ description)		Send a bug report to Amersham Biosciences. Restart computer
More than 500 clones found, aborting search.	When using the Add Batch Samples functions and the search criteria are too broad.	Use more narrow search criteria.
No deviations left, calculation failed.	When generating quality data and all files have been removed or there are only reference fragments present.	Add samples.
Not able to load alignment info.	During deviation calculation.	Try closing other applications or increasing the number of file handles available (FILES in config.sys)
Not able to open "filename" "description"	When printing a report to file or exporting results to a database file. The file may already be in use or write protected.	Close the file if already in use or remove any write protection.
Not able to open ALF/ALX file "filename and clone". No graph data will be available.	When the ALF/ALX file referenced by the current analysis item could not be found probably because it has been moved to another directory.	Return the ALF/ALX file to its original directory or add the ALF/ALX file again as a new sample
Not able to open graph view, out of memory.	When out of memory.	Try closing other applications or restart the computer.

Not able to read checksum, file is corrupt. Not able to read deviations information, file is corrupt. Not able to read options settings, file is corrupt.	The file being opened is corrupt.	Corrupted files cannot be used. Recreate the source file to be opened.
Not able to restore header contents, file may be corrupt.	When appending data to a database file and choosing to use the existing column information. This message occurs if the application cannot decode the existing file.	Corrupted files cannot be used. Create a new database file.
Not able to save analysis item.	When there is not enough disc space or out of memory.	Free up disc space or memory. Restart the computer.
Not enough memory available to calculate deviations.		Try closing other applications or restart the computer.
Specified gene configuration file is corrupt.	When the file being loaded is corrupt.	Corrupted files cannot be used. Recreate the source file to be opened.
Specified gene configuration was not loaded, no fragments are defined.	Error when loading a gene configuration that does not contain a sequencer fragment in the first region.	Redefine the current gene configuration or use another.

The available printer fonts may not produce a correct layout.	When trying to print a report on a printer which does not support the fonts "Courier" or "Courier New", or is a "Line Printer".	Use another printer.
Too short sequence to calculate modulation factor.	When the sequence contained in the loaded ALF/ALX file was too short to calculate any modulation factor.	Try again (maybe change filename).
Wrong file format: Not an Analysis Item file	When the file being opened is not an analysis item file.	Open a file type designated **.itm**

7.4 Other problems

High background in your sample sequence
If you experience high background in your sample sequence leading to unresolved base calling, you should try applying a different post run action sensitivity to your sample sequence in the sequencing software.

Incomplete GenBank data imported
If you do not get any locus or type information in a GenBank file, this may be due to the use of a non-appropriate file format, e.g. html format, when importing the file from the Internet. Establish how native GenBank format files are produced by your favourite sequence database search engine.

A. Evaluation algorithms

Detection and classification of wild type deviations in sequence data

The method used for the detection and classification of wild type deviations in sequence data provided by the automatic sequencer, is based on the dynamic programming algorithm originally proposed by Needleman and Wunsch [2]. The method is reviewed in a more recent paper by Pearson and Miller [3].

The method has been further extended with some functions dedicated to the present application. A condensed description of the different steps involved is given below:

1. The sample sequence is aligned to the different sequence fragments specified by the gene configuration.

   For each configuration fragment an alignment is found that optimises the score defined by the following score table when summed up for the entire sequence length:

Event	Score
Matching base	+1
Mismatching base	-1
Matching ambiguity	0
Mismatching ambiguity	-1
Gap	-1 (independent of gap length)

   The matching ambiguity event occurs when there is an ambiguous base call in the sample sequence comprising a base component that matches the aligned base in the configuration fragment.

   The mismatching ambiguity event occurs when there is an ambiguous base call in the sample sequence not comprising as a component the base found at the corresponding position in the configuration fragment.

2. The alignment algorithm is optionally applied both on the straight sequence as obtained from the sequencer and on the complementary reversed sequence.

3. The alignment item for the configuration fragment and the sample sequence direction showing the highest score is assumed to be the right one and is used for further calculations.

4. A list of deviations is derived from the best alignment. The deviations are labelled according to the following table:

   a.) Point: Mismatching nucleic acid base code.

   Example:

```
          C
A  C  A     G  C
A  C  A     G  C
          A
```

   b.) Ambig: Base code comprising two or three nucleic acid base components.

   Example: (Matching ambiguity event according to score table notation).

```
          M
A  C  A     G  C
A  C  A     G  C
          A
```

   Example: (Mismatching ambiguity event according to score table notation).

```
          S
A  C  A     G  C
A  C  A     G  C
          A
```

c.) SeqStop; Base code comprising all four nucleic acid base components.

Example:

```
          N
A   C   A     G   C
A   C   A     G   C
          A
```

d.) Insert: Explicit gap of any length in the wild type sequence.

```
          C
A   C   A     G   C
A   C   A     G   C
          -
```

e.) Delete: Explicit gap of any length in the sample sequence.

```
          -
A   C   A     G   C
A   C   A     G   C
          A
```

B. Export to database

A typical use of the **Export Result to Database** function is illustrated in Figure B-1. The results from all the *analysis items* produced in a clinical evaluation study are exported to a *Microsoft Access* database together with information about the individuals and information from an external mutation database. The different pieces of information are then processed together using the tools available in Microsoft Access and different kinds of reports, as well as statistics diagrams, are produced.

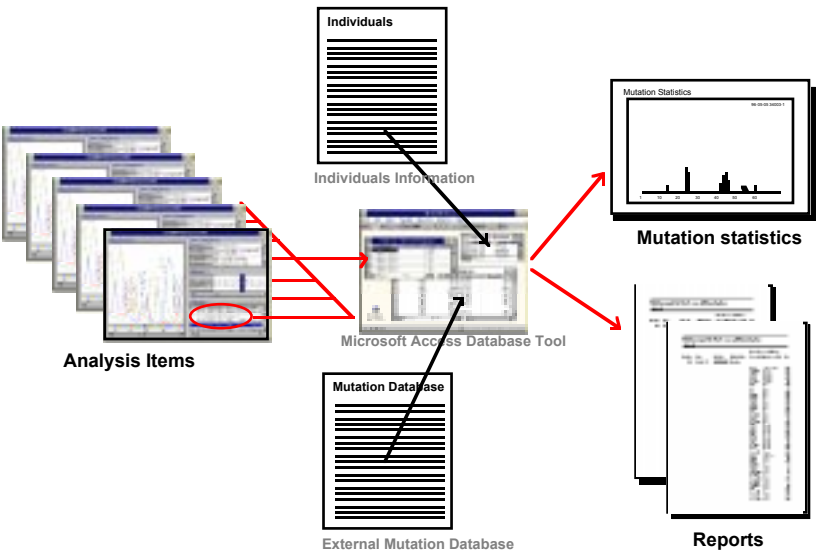


Figure B-1. *Scheme showing the typical use of the Export Result to Database function*

The procedure used for exporting data to the *Microsoft Access* environment is described in more detail:

Note: Data can be exported to other database environments in a similar way.

1. After you have evaluated all the analysis items included in your study, open the first Analysis item and select **File:Export Result to Database**.

2. In the **Export to Database** dialogue box that is displayed, enter a suitable file name and click the **OK** button. The information extracted from the analysis item will now be saved in the chosen

file. For later use, it is recommended that you take a note of the directory path where the file is stored.

3. Repeat for all other analysis items to be included, except export to the the same file name as created in step 2. The information extracted from all analysis items will now be appended to the original file. The file can now be imported into *Microsoft Access* or any similar database environment.

4. Open *Microsoft Access*. It is assumed that this application has been previously installed on your system. Refer to *Microsoft Access* documentation for further information.

5. Using *Microsoft Access*, create a new database using the **File:New Database** menu command.

6. Chose a suitable file name and click the **OK** button. Note that this should not be the same file name as the one used in steps 1-3.

7. Import the data from the analysis items using the **File:Import** menu command.

8. Choose **Text (Delimited)** as the data source and click on the **OK** button.

9. Select the file you created in steps 1-3 and click the **Import** button.

10. In the **Import Text Options** dialogue box, select **First Row Contains Field Names**.

11. Click on the **Options** button. Set the **Text Delimiter** as **None** and **Field Separator** as **Tab**.

12. Click on the **OK** button to import data.

13. Close the **Select File** dialogue box. The data from your study is now available for further processing or examination within the *Microsoft Access* environment.

Note: You can now erase the file that you created in steps 1-3 to save disk space.

14. Data from other sources, e. g. mutation databases, information about the samples and the individuals included in the study, can be imported into the database as separate tables in a similar way. Refer to the *Microsoft Access* documentation for information on document preparation.

15. Finally, the information in the different tables can be linked together based on fields with common information and different kinds of reports, or statistics can be derived using the graphical design tools provided in the *Microsoft Access* development environment. Refer to the *Microsoft Access* documentation for details.

C. Cited literature

[1] Beaudet, A.L. and Tsui, L.C. (1993) Human Mutation, 2, 245-248.

[2] Needleman, S. and Wunsch, C. (1970) J. Mol. Biol. 48, 444.

[3] Pearson, W.R. and Miller, W. (1992) Methods in Enzymology, 210, 575.

[4] Various sources: Anderson, S. et al (1981) Nature, 290, 547; Borst, P. (1980-81) International Cell Biology, H.G. Schweiger, ed. Springer-Verlag, p. 239; Breitenberger, C. and Rajbhandary, L. (1985) Trends in Biochem. Sci., 10, 478-483.

[5] Donis-Keller, H. et al (1993) Human Molecular Genetics, 2, 851-856.

Index

A

C

D

E

F

G

H

I

K

L

M

N

O

overview options, 4-29

P

Q

R

S

T

U

W

Z

18-1118-78 Edition AA
56-3042-87

Printed in Sweden by TK i Uppsala AB, April 1997