Contents

Chapter 1. Introduction

DAx can be used to read and analyse trace files such as the files that are created by the Amersham MegaBACE™ and ABI Genescan® systems.

The MegaBACE™ and ABI systems, like other DNA analysis systems, can analyse four traces in a single run, using four different (coloured) labels. One way to use this feature is to include a standard sample with known DNA fragment sizes as one of the traces, and apply this standard to the other traces. The standard is used to ascertain the fragment sizes of the components in the unknown samples.

This manual presents a step-by-step guide to set up an analysis such as the one described above. The techniques used can be applied to a wide variety of analyses.

Chapter 2. Analysis of the Standard

This chapter describes the steps needed to set up the analysis of the sizing standard contained in trace files. Typically, you will go through this process once, possibly revisiting it to refine the procedure.

In order to use the size standard, the following steps are needed:

- baseline construction
- finding peaks
- recognising peaks belonging to the size standard

By creating an **Analysis Procedure** that can automatically perform these steps, any unknown trace file containing the sizing standard can be correctly interpreted.

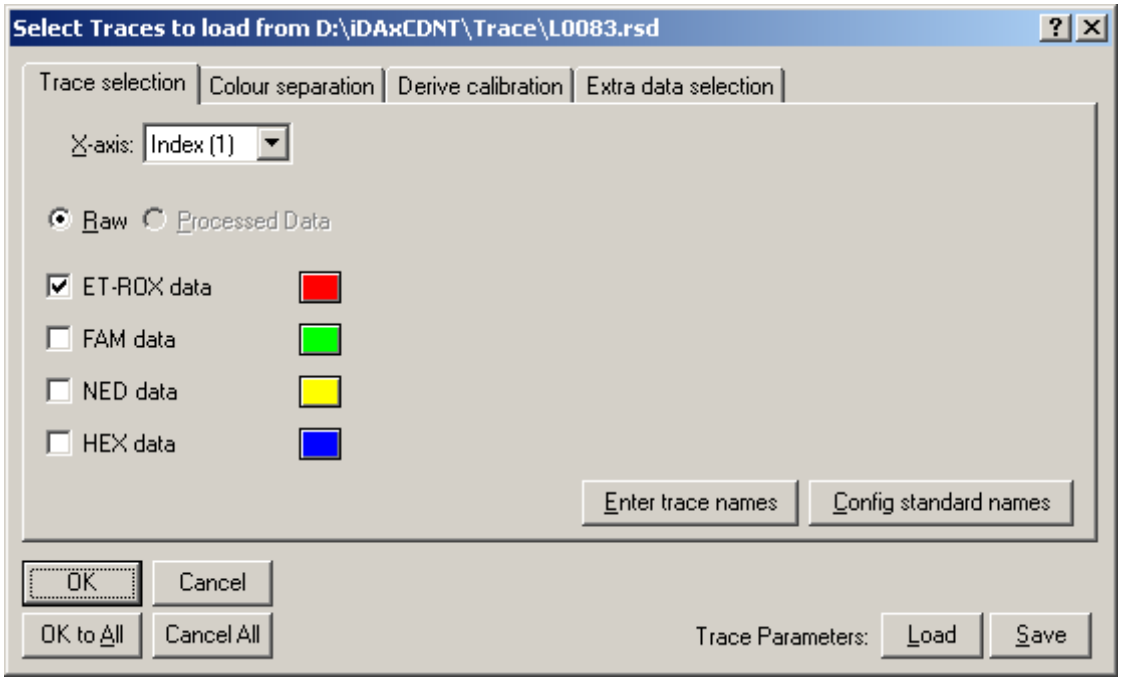
This manual uses an example MegaBACE™ file called L0083.rsd. This file contains 3 traces: a ROX labeled standard, and FAM and HEX labeled unknown samples. The file is supplied on the DAx CD.

Chapter 2.1.Loading the Standard Trace

To load just the trace containing the sizing standard, invoke the **File | Open** menu option. A dialog box appears. At the bottom of the dialog box, in the **Files of Type** box, select **MegaBACE™ files**[1]. Locate the L0083.rsd file, and click the **Open** button. Do not check the **AutoAnalyse** item. Leave the **X-axis** item set to "Time (s)".

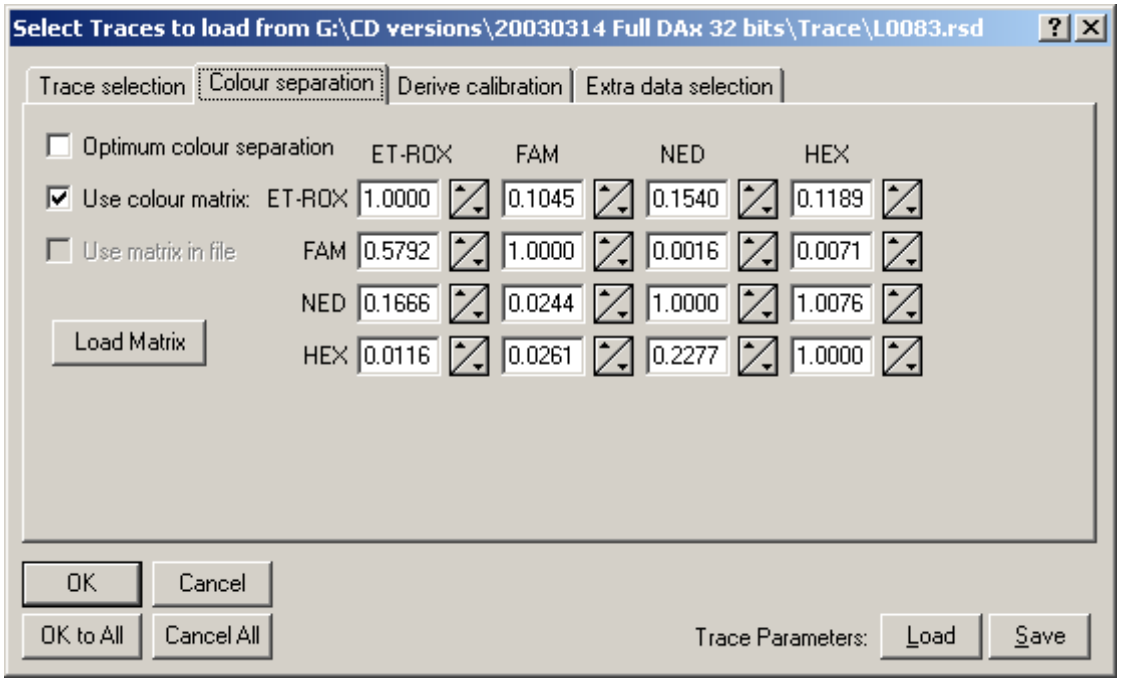
A dialog box will appear. This dialog box is used, among other things, to indicate which traces are to be loaded from the file.

[1] If you are using ABI files, select "ABI Genescan ® files". This manual assumes you are using MegaBACE files, and uses the example file supplied on the CD.



In this dialog box, make sure to

- On the **Trace selection** tab, select **Index (1)** for the **X-axis**
- Check **ET-ROX data**. This will make the ROX trace be loaded. In our example, the ROX trace contains the sizing standard. For more on the use of trace types in DAx, see chapter 5.
- Do not check the FAM, HEX, or NED traces.
- Choose the colour red for the ROX trace.



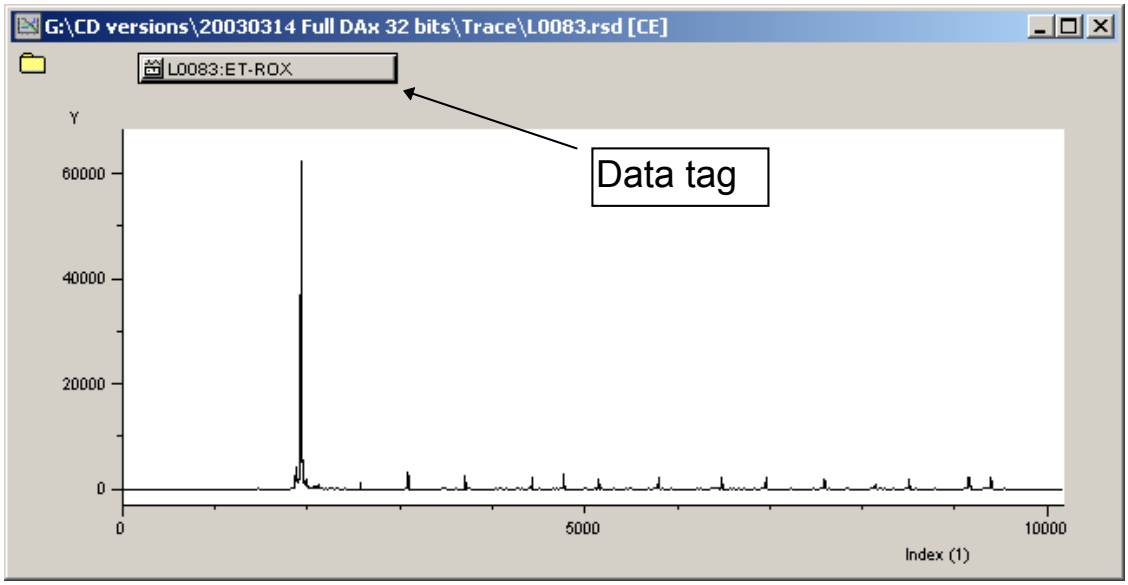
- On the **Colour separation** tab, check the **Use Colour Matrix** box to make sure that a colour matrix is applied to the traces.

This separates the contributions from the various labels at the wavelengths used to measure each trace. Use the values displayed in the dialog box above.

- On the **Derive calibration** tab, check the **Do not derive calibration** box.
- On the **Extra data selection** tab, do not check anything.

TIP: the settings in this dialog box are supplied as tapstd.tap. Use the **Load** button at the bottom of the dialog box to load these settings.

Now click the **OK** button to load the trace. A graphic window displaying an image such as the one below will be opened.



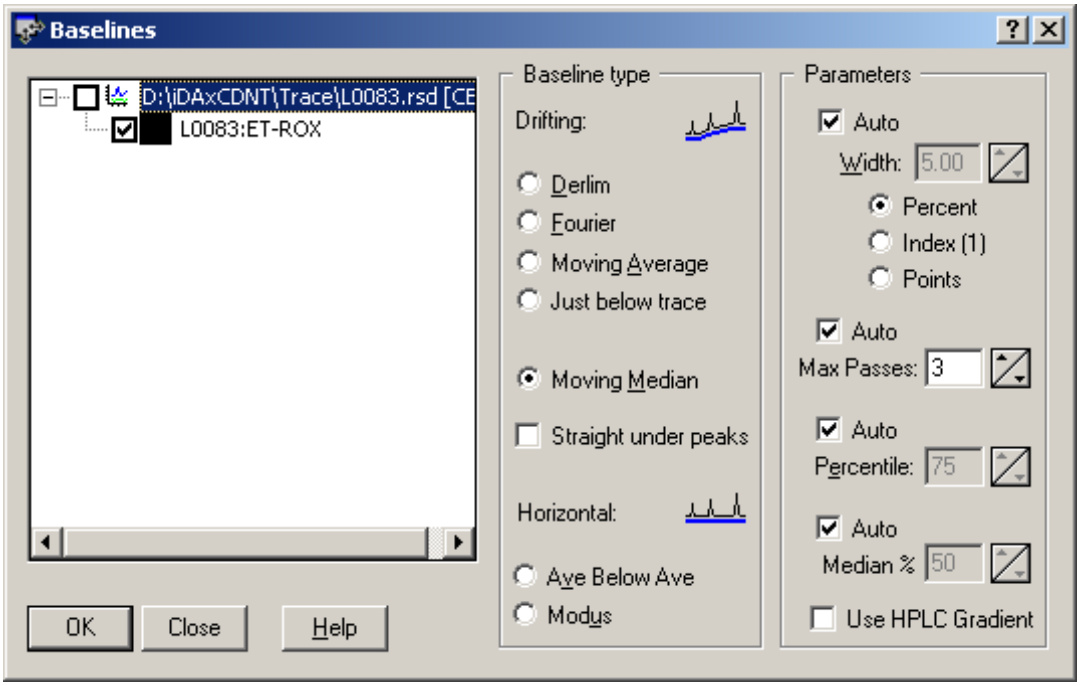
Note that the graphic window contains a **data tag**. A data tag looks like a button and is located at the top or at the left side of the graphic window.
Data tags represent a data set. Clicking on it displays a menu with options that pertain to the data.

Chapter 2.2.Finding Peaks in the Standard trace

The next step is to create a baseline and find peaks in the standard trace.

To create a baseline, invoke the **Peaks | Construct Baseline** menu option (you can also click the  toolbar button). A dialog

box will be displayed, which lets you enter baseline construction settings. Select a **Moving Median** baseline; check all the **Auto** boxes, and enter 3 for **Max Passes**. Now click the **OK** button. A baseline will be created.



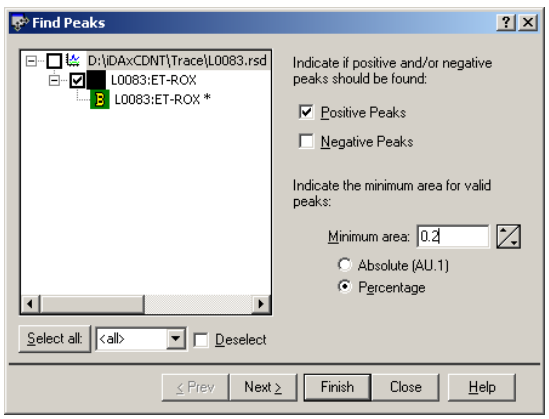
Example of baseline construction dialog box

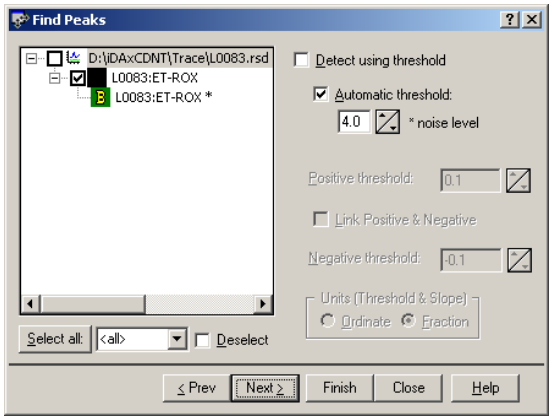
The next step is to locate peaks. To start the process, invoke the **Peaks | Find Peaks** menu option (you can also click the  toolbar button). The peak find dialog box is displayed. This dialog box consists of multiple pages of settings. Below is a list of the settings that should be used. Images of all 8 pages of the dialog box are also displayed.

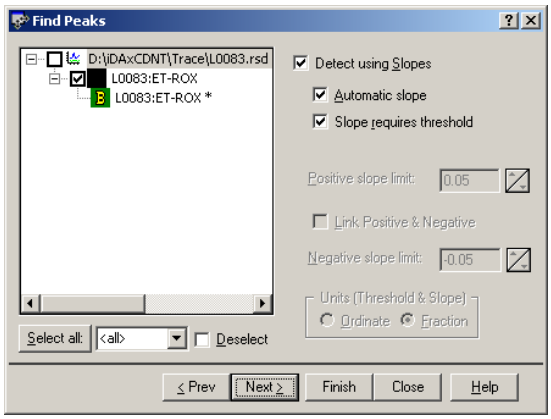
- on the first page, select to find only positive peaks. No negative peaks of interest are present.
  Enter a minimum peak area of 0.2%. Any peaks smaller than 0.2% are of no interest to the analysis of the standard.
- on the second page, check **Detect using threshold**; check **Automatic Threshold**; enter 4.0 for the *** noise level** item.
- on the third page, check **Detect using slopes**, check **Automatic slope**, and check **Slope requires threshold**.
- on the fourth page, do not check any items. Certain areas of the data can be excluded from the peak find process using this page.

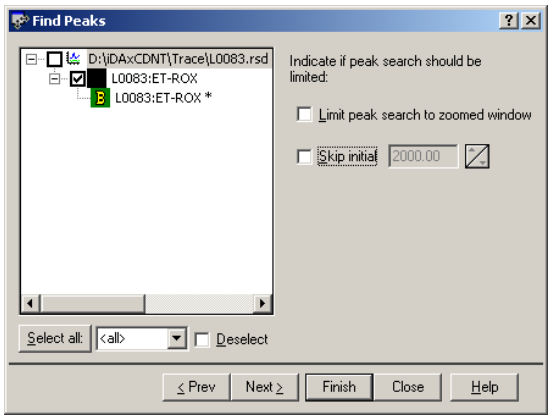
- on the fifth page, make sure **Use Peak Events** is not checked. Peak events can be used to suppress peak finding in certain locations.
- on the sixth page, all items may be unchecked. Finding shoulder peaks is beyond the scope of this manual.
- on the seventh page, make sure that **Normalise Peak Widths** is not checked.
- on the eighth page, do not check any items. The example file does not contain bleedthrough (cross-over) peaks that could interfere with the analysis.

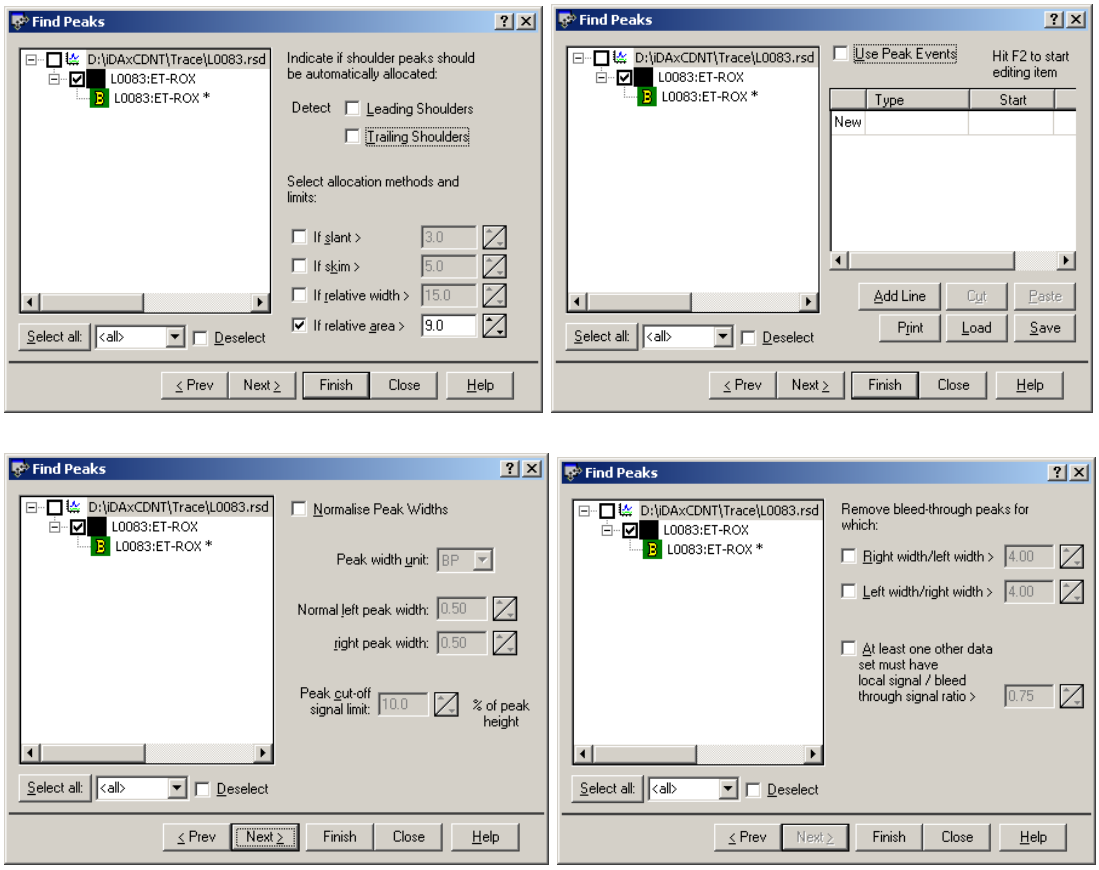
Now click the **Finish** button. This will locate peaks in the standard.











To display a list of these peaks in a *peak list window*, click the small peak list button at the left side of the data tag for the standard. Alternatively, click the data tag, and invoke the **Peaks | List Peaks** menu item.

The contents of the peak list window can be configured using the **View | Configure** menu option. For our purposes, we will need to display at least the peak index, peak migration time, component name and Base Pairs (fragment size; this item is located near the bottom of the list of available peak list columns displayed at left in the View | Configure dialog box).

Chapter 2.3.Identifying Peaks in the Standard Trace

In order to use the standard trace to determine fragment sizes, a calibration needs to be set up. This calibration converts time coordinates into fragment sizes.

If files are to be analysed automatically, it must be possible to derive this calibration automatically from the standard trace. It is possible to do this because we know which components, with which fragment sizes, are present in the standard.

The first step in setting up an automated calibration is to recognise the peaks that are present in the standard. To do this, the peaks are identified manually the first time. Using these peaks, an **Identification Database** is created which can recognise the peaks automatically.

Chapter 2.3.1.Entering peak names

As indicated above, the first thing we must do is to recognise peaks in the standard manually. Using the peak list of the standard trace that was created in chapter 2.2, we enter component names as follows:

Peak	Top (s)	Component	Base pairs
1	1856.00		
2	1875.00		
3	1883.00		
4	1890.00		
5	1905.00		
6	1929.00		
7 ref	1941.00	20 bp	
8	1948.00		
9	1957.00		
10	1989.00		
11	2005.00		
12	2042.00		
13	2072.00		
14	2097.00		
15	2125.00		
16	2131.00		
17	2588.00	100 bp	
18	3100.00	159 bp	
19	3717.00	221 bp	
20	4426.00	288 bp	
21	4437.00		
22	4774.00		
23	4785.00	318 bp	
24	5149.00		
25	5161.00	350 bp	
26	5806.00	403 bp	
27	6479.00	459 bp	
28	6492.00		
29	6958.00	498 bp	
30	6970.00		
31	7603.00	551 bp	
32	8151.00	600 bp	
33	8518.00	633 bp	
34	9163.00	700 bp	
35	9401.00	726 bp	

Only the component names are entered. To enter a component name, click in the component name column for the appropriate peak, and enter a name. A total of 15 peak names are entered.

Chapter 2.3.2.Setting up an Identification Database

Now that a number of peaks have been named, we will set up an Identification Database. This is a database of component names

that can be used to determine which components are present in an unknown sample.

First, if the **Analysis | Stop Analysis** menu item is enabled, invoke it to clear any existing Identification Database.

There are two ways in which the Identification Database can recognise unknown components.
The first way is to recognise components by their proximity to certain coordinates. A large number of different coordinates can be used in DAx, such as peak top time, peak begin or end time, peak modus, capacity factor, or even mobility or fragment size. This is set up using the **Analysis | Edit Database** menu option.
The second way to recognise components is more powerful. It can look for components using their proximity to a coordinate, but it can also look for the highest or largest peak in an interval, the first or last peak in an interval, or even the first peak in an interval that is larger than a set limit. This method of recognising peaks is set up using the **Analysis | Marker Peaks** menu option. ***NB*** *If the Analysis | Marker Peaks menu option is not displayed, use the* ***File | Customise*** *menu option. Select the* ***Extensions*** *tab, and make sure that* ***Use Marker Peaks*** *is checked.*

We will use the second method to recognise components. Invoke the **Analysis | Marker Peaks** menu option. The marker peak set up dialog box will be displayed.
The dialog box shows a list of marker peaks. The list has the following columns:

- marker name: name of the component that is to be recognised
- peak type: type assigned to recognised peaks
- annotation
- **in trace**. This column determines in which type of trace (ROX, FAM, etc) certain peaks are to be located. Since the ROX trace contains the sizing standard, **in trace** should be set to ROX!
- qualifying coordinate: which coordinate to use to recognise the peak
- marker name to be used, that is: not the name of the marker being located, but the name of the marker in relation to which the new marker is being located. If two markers are used, both names are displayed here.
- type of search, such as *largest in interval* or *first in interval*
- (lower) qualifying coordinate

- upper qualifying coordinate for search types that use an interval
- tolerance (if applicable). To set a relative tolerance, follow the tolerance value with a **%**. If the type of search is “first in interval with area > limit” or “last in interval with area > limit”, the area limit is entered in the tolerance column. Enter an absolute value as an area limit, or a relative value (using %) as a relative area limit.

Using the standard that was analysed, the list of marker peaks can be set up by repeating this procedure:

- click the **Add Line** button
- click the **Copy Peak Data** button; select each of the 15 peaks that were named in turn, starting with 20 bp and ending with 726 bp

It was found that the example data can optimally be analysed by changing the marker peak settings as detailed on the following page.

Index	Marker Name	Qualify by	Using marker:	Search	Coor/Start	End Interval
1 Ref	20 bp	Peak Top Time (s)		Highest in interval	1500.0	4500.0
2	100 bp	Time to marker (s)	20 bp	Highest in interval	400.0	1060.0
3	159 bp	Time to marker (s)	100 bp	First in interval	350.0	800.0
4	221 bp	Time to marker (s)	100 bp	Highest in interval	850.0	1750.0
5	288 bp	Time to marker (s)	221 bp	First in interval	550.0	1300.0
6	318 bp	Time to marker (s)	288 bp	Highest in interval	250.0	700.0
7	350 bp	Time to marker (s)	318 bp	Highest in interval	250.0	750.0
8	403 bp	Time to marker (s)	350 bp	First in interval	500.0	1500.0
9	459 bp	Time to marker (s)	403 bp	First in interval	500.0	1500.0
10	498 bp	Time to marker (s)	459 bp	First in interval	300.0	1000.0
11	551 bp	Time to marker (s)	498 bp	First in interval	400.0	1200.0
12	600 bp	Time to marker (s)	551 bp	First in interval	300.0	1200.0
13	633 bp	Time to marker (s)	600 bp	First in interval	200.0	700.0
14	700 bp	Time to marker (s)	633 bp	First in interval	300.0	1000.0
15	726 bp	Time to marker (s)	700 bp	First in interval	100.0	500.0

The *peak type*, *annotation*, *in trace*, *tolerance* and *exclude from total* columns were omitted from this table for clarity. The peak type, annotation and tolerance entries can all be left blank; the exclude from total entries should all be “no”.

The **in trace** column entries should all be set to **ROX**. The reason for this is that the standard peaks are known to be present only in the ROX trace!

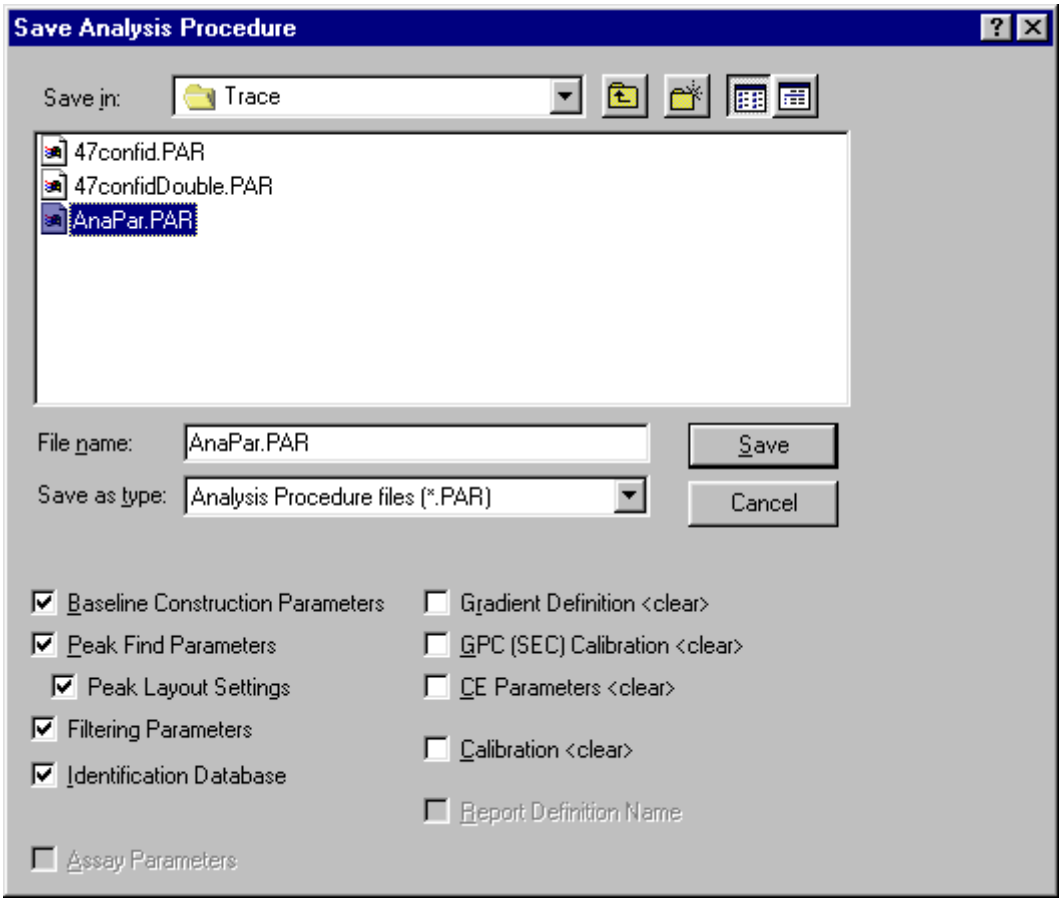
Note that the **20 bp** peak is located as the highest peak in the interval 1500-4500 seconds.
All the other peaks are located relative to an earlier recognised peak. For instance, the **100 bp** peak is recognised as the highest peak in the interval 400 to 1060 seconds following the 20 bp peak.

Chapter 2.4.Storing Results

At this point, a number of parameters needed to automatically analyse trace files have been set up:

- baseline construction
- peak finding
- peak recognising

Together, these parameters form an **Analysis Procedure**, that is, a complete set of parameters needed to analyse files with unknown data.
Store the Analysis Procedure using the **File | Save Analysis Procedure** menu option. Invoking this option displays the following dialog box.



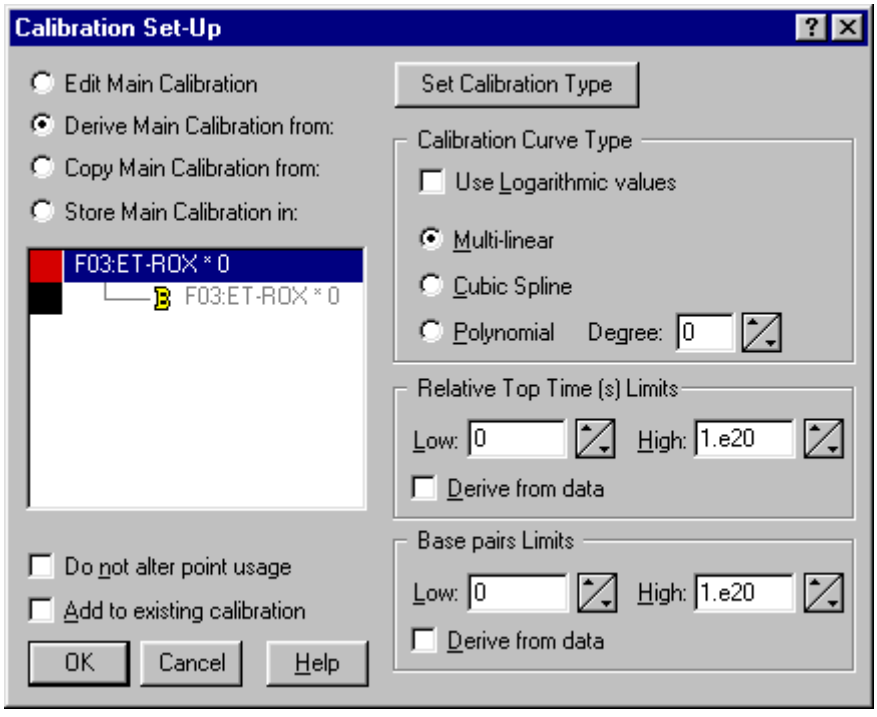
Make sure to check only **Baseline construction parameters**, **Peak find parameters**, **Peak layout settings** (optional), **Filtering parameters** (optional), and **Identification Database**.

When DAx is started, the most recently stored Analysis Procedure is always loaded. This means that once you have set up and saved an Analysis Procedure, it will always automatically be ready for use when you start DAx.

TIP: An analysis procedure has been set up and is supplied as AnaPar.PAR on the DAx CD.

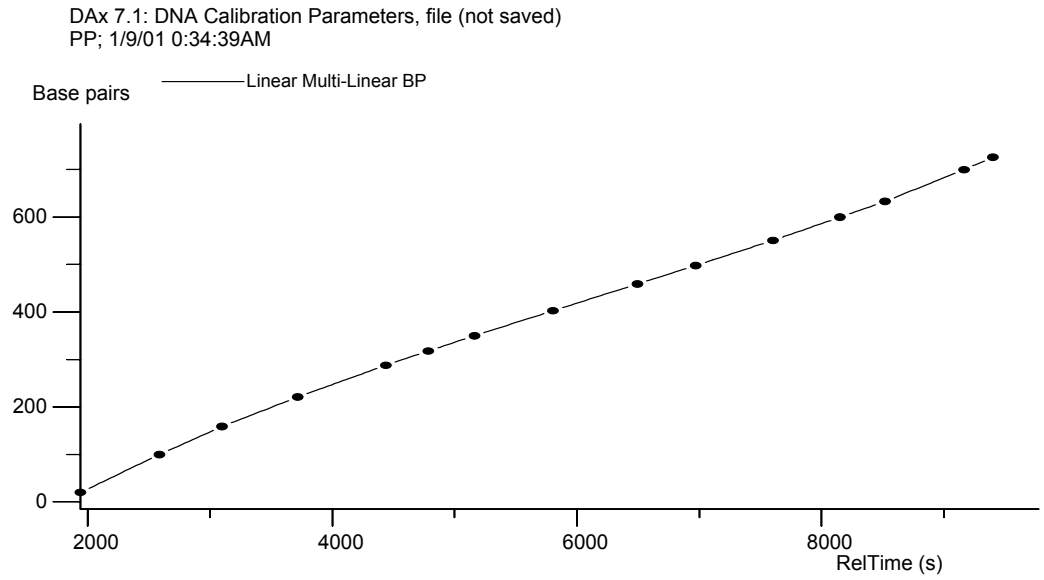
Chapter 2.5.Deriving a calibration

Using the standard that has been analysed, it is possible to demonstrate the derivation of a calibration.
Simply invoke the **Calibration | Calibrate** menu option. A dialog box appears. Make sure to select **Derive Main Calibration from** and **Multi-Linear**. Then click the **OK** button.



Example Calibration setup dialog box

A calibration curve will be displayed such as the one below.



Note that after a calibration has been derived, the Base Pair column in the peak list for the standard will have been filled with the appropriate values.

Remove the calibration from memory using the **Calibration | Stop Calibration** memory option.

There is no need to store this calibration. Rather, a fresh and accurate calibration is derived from trace data when they are loaded, using the Analysis Procedure that was set up and stored in chapter 2.4.

Chapter 3. Analysis of the Unknown

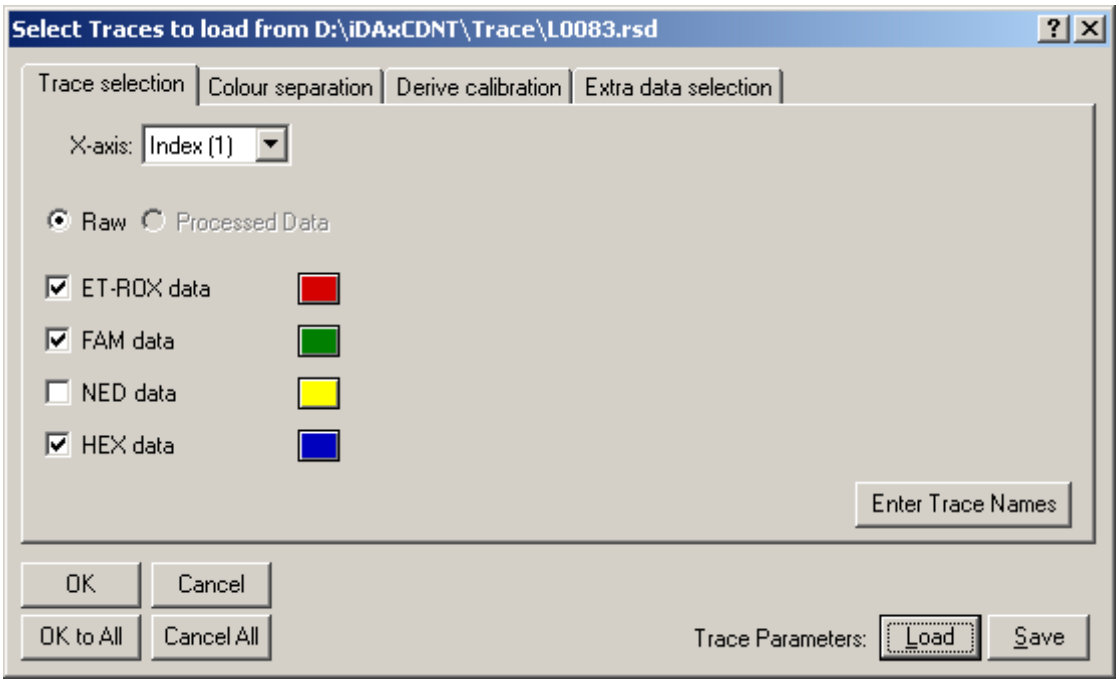
In the previous chapter, the analysis of the trace containing the size standard was set up. The same settings can be used to analyse the unknown traces (FAM and HEX in the example).

It is important to note that, since the FAM and HEX traces do not contain the size standard, it would be undesirable to recognise any peaks in these traces as size standard peaks. This is why the **in trace** column in the marker peaks dialog box was set to **ROX**.

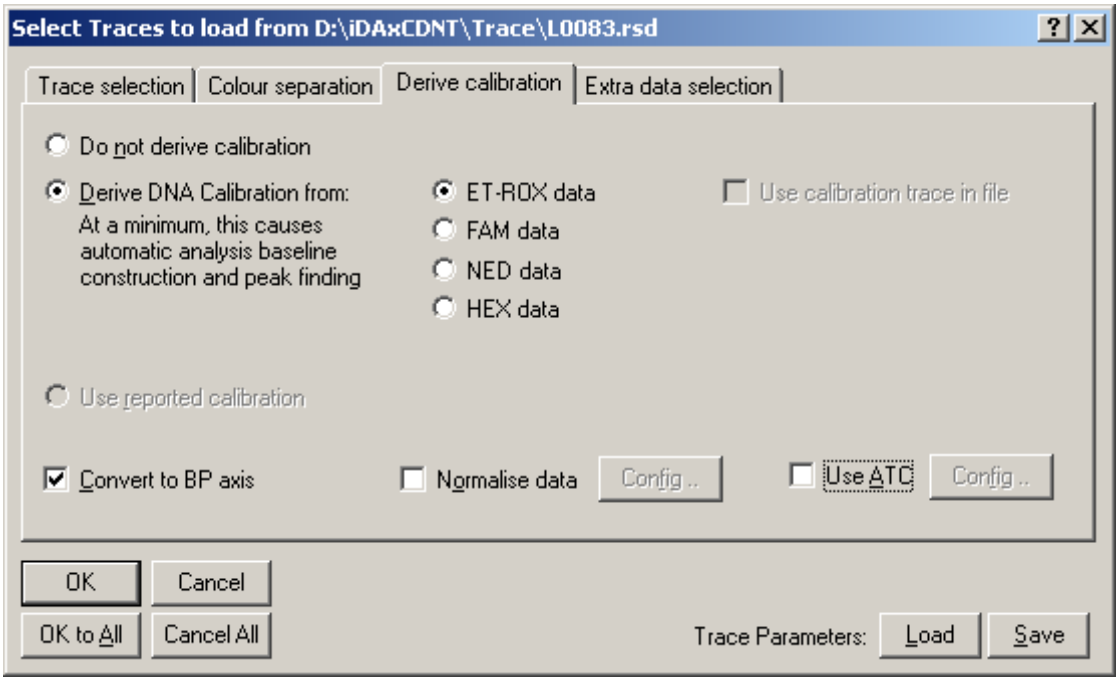
Chapter 4. Putting it all to work

All the requirements to automatically analyse a MegaBACE™ file have now been met.
To automatically analyse a file, follow the following procedure.

1. Invoke the **File | Open** command. Select **Files of Type** MegaBACE™. Select the file or files you want to analyse.
2. Check the **AutoAnalyse** button.
3. Click the **Config** button next to the AutoAnalyse button. The AutoAnalysis setup dialog box appears.
4. Make sure that at least **Construct Baseline** and **Detect Peaks** are checked. Additional items can be checked also. However, **Derive DNA Calibration** should not be checked. For now, do not check **Convert Axis**.
5. Click the **OK** button.
6. Click the **Open** button. This will start loading the trace file.
7. The Trace Analysis Parameters dialog box is displayed. Set it up as shown below. Alternatively, the settings can be loaded from the file UNKNOWN.TAP.



On the **Trace selection** tab, the ROX, FAM and HEX traces have been checked to be loaded.



On the **Derive calibration** tab, **Derive DNA Calibration** has been checked, and **ROX** has been selected as the trace to derive the size standard from[2].

8. Click the **OK** button.

Now, the following will happen in succession.

- the ET-ROX trace is loaded. It is analysed using the current settings. Since the Identification Database contains entries that recognise peaks in ROX traces, the standard peaks will be recognised.
- a calibration relating time coordinates to fragment sizes is derived from the ROX trace.
- the FAM trace is loaded and analysed using the current settings. Since the Identification Database does not contain any entries that apply to FAM traces, peaks are **found** but not **named**.
- the HEX trace is loaded and analysed using the current settings.

Troubleshooting: If you get a message “no calibration points found”, the AutoAnalysis setup likely does not include baseline construction and peak detection.

---

[2] Because **Derive DNA Calibration** has been checked, it was not strictly necessary to check the **AutoAnalyse** box in the File | Open dialog. Baseline construction and peak finding would still have been performed for the ROX trace, since they are required to derive the calibration. However, check AutoAnalyse will cause the FAM and HEX traces to be analysed in addition to the ROX trace.

You can now click any of the peak list buttons at the left side of the data tags for the ET-ROX, FAM, or HEX traces. You will note that the correct fragment sizes are listed in the Base Pairs column.

It is also possible to display the fragment sizes of the peaks that were found in the graphics window. To do this, invoke the **File | Customise** menu option. Select the **Plotting Peaks** tab. Under Peak Labeling, select **Component Name** as the first peak label. Select **Base Pairs** as the second peak label (it is near the bottom of the list of possible peak labels). Click the **OK** button.
The graphic window will now display base pair counts for all the peaks. There will be very many peaks displayed. You can click and drag the mouse on the graph to zoom in on the peaks, and see more detail.

Chapter 4.1.Converting to a Base Pair axis

In the example above, fragment sizes (base pair counts) are labeled for the peaks, but the horizontal axis displays time coordinates. For gene analysis, this often does not make much sense. Often, the horizontal axis does not correspond to specific time coordinates.

For this reason, DAx can convert the horizontal axis to fragment size units. Use the **Calibration | Axis Conversion** menu option to do this. Make sure to check the **New Window** check box, and then make sure that all data lines are selected.

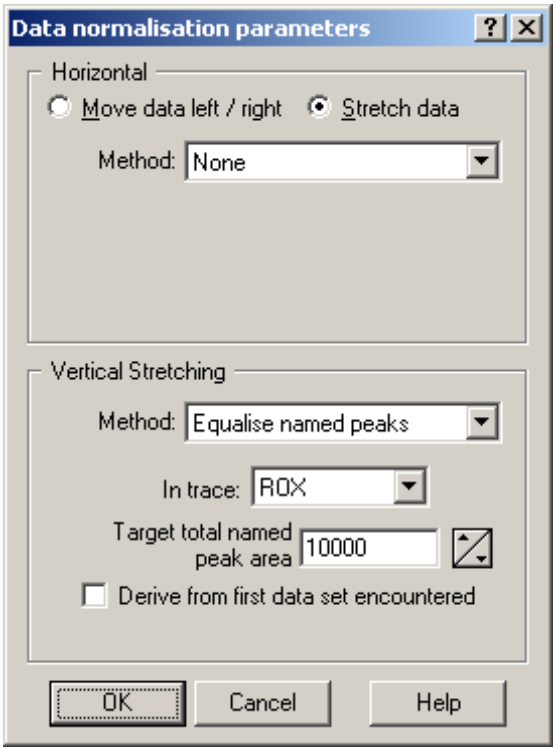
A more convenient way to convert the axis is to check the **Convert to BP Axis** box in the Trace Analysis Parameters dialog box (see above).
The analysis will be performed as before, but the horizontal axis will be converted to fragment units at the end of the analysis.

Chapter 4.2.Normalising the data

When loading trace data, it can be desirable to normalise the data after they have been loaded. One common form of normalisation

stretches the data vertically in such a way that the total area of the size calibration peaks is equal across several measurements.

Normalisation can be achieved after data have been loaded using the **Data | Overlay** menu option. A more convenient way to normalise is to check the **Normalise data** check box on the **Derive calibration** tab in the Trace Analysis Parameters dialog box (see above). Click the **Config** button to set up normalisation.



In this example, the horizontal axis is not affected.

The vertical axis will be stretched in such a way that the total area of all named peaks in the ROX trace will become 10000. *Named* peaks are used because the named peaks in the ROX trace will be the calibration standard (as called by ATC).
The other traces (FAM, HEX) will be stretched an equal amount as the ROX data.

Chapter 5. Trace Types in DAx

Trace types are used in DAx for a number of purposes.

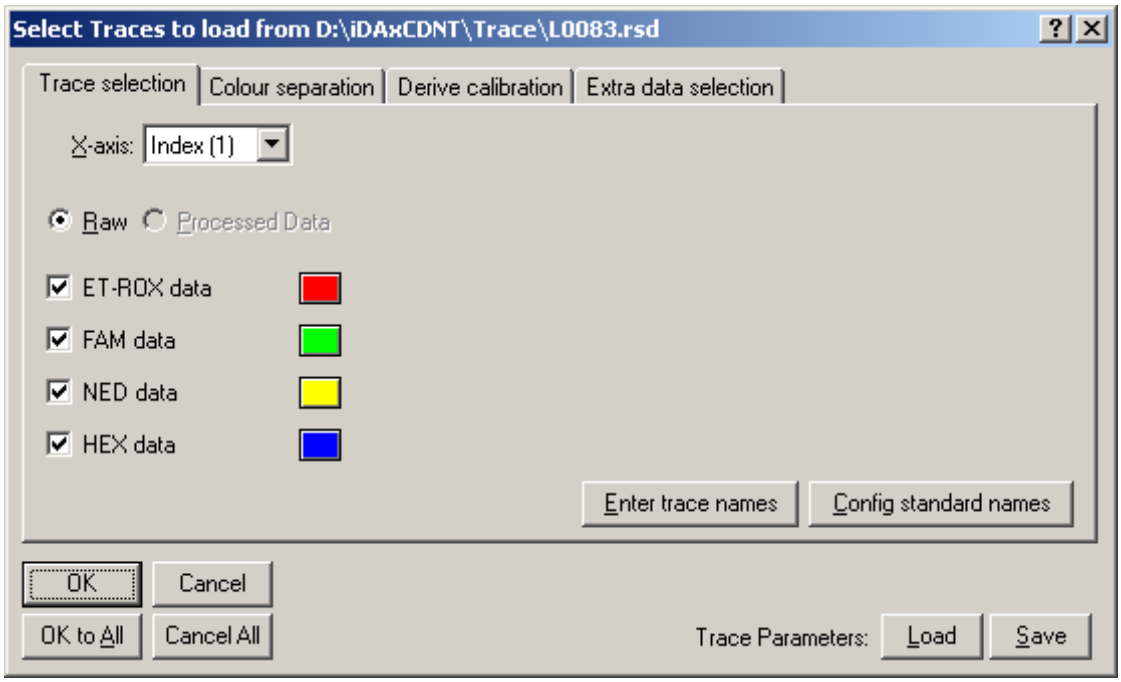
Their principal use is to distinguish between the up to five traces that can be contained in a single measurement file, where obviously the file name alone would be insufficient.

A secondary use is to limit certain operations to specific trace types. These operations include:

- Identification Database and Marker Peak items. The trace type is used to limit the detection of certain components to a specified trace type.
- Automatic Trace Calibrations. The trace type is used to indicate which trace contains the sizing calibration.

Chapter 5.1.Assigning trace types

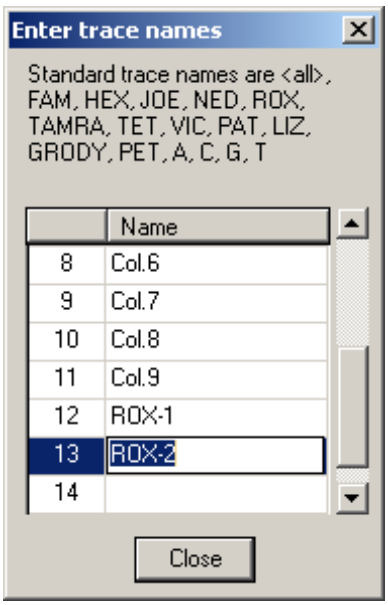
DAx attempts to retrieve the trace type of each trace from the measurement file, which often (but not always) contains trace names. The trace names are displayed on the first tab of the TAP dialog (see also chapter **Error! Reference source not found.**).



In this case, the measurement file L0083.rsd is recognised to contain ET-ROX, FAM, NED and HEX trace names[3]. If no trace types are manually specifed (see below), DAx will recognise these traces as ROX, FAM, NED and HEX[4].

The standard trace types that DAx recognises are FAM, HEX, JOE, NED, ROX, TAMRA, TET, VIC, PAT, LIZ, GRODY, and PET. Mainly for use with SCF files, there are also A, C, G and T trace types.

The **Config standard names** button allows additional trace types to be added. A dialog box like this one is displayed:

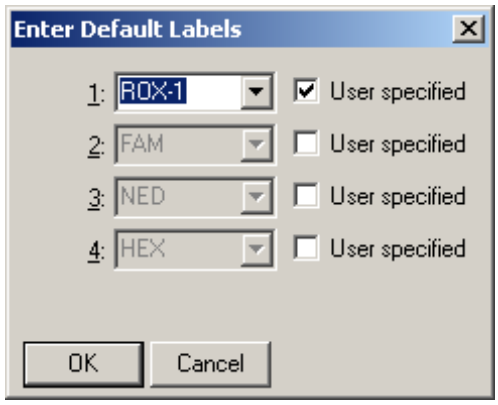


One way to use this feature is to enter types such as ROX-1 and ROX-2, so that two distinct ROX trace types are added.

[3] The trace name is “what the trace is called in the measurement file”.

[4] ROX is recognised because trace type ROX is part of trace name ET-ROX.

The **Enter trace names** button allows the trace names specified in the measurement file to be overridden with a user selected name or type. Clicking the button displays this dialog box:



To override a trace type, check the **User specified** box, then select a trace type or type a trace name.

Chapter 5.1.1.How DAx recognises trace types

As was mentioned above, an ET-ROX trace name will be recognised as a ROX trace type. This is true even if ROX-1 and ROX-2 trace types are manually added, because “ET-ROX” does not contain “ROX-1” or “ROX-2”, but does contain “ROX”. In order for a trace name to be recognised as a certain type, the name must contain the entire trace type.
The exception are the A, C, G and T trace types, which are only recognised if the entire trace name equals A, C, G or T.

Chapter 6. Conclusion

DAx can be used to perform highly complex analyses of trace files with relatively little effort.

There are a large number of additional features that have not been discussed in this manual. Examples are:

- peak width standardisation. If a sample is known to contain mostly peaks that have a width of 1 fragment, DAx can be told to preferentially assign peak widths of 1 fragment. If a peak is *clearly* wider, its width will not be affected.
- bleedthrough / cross-over peak suppression. Because of the way trace analyses work, whenever there is a very high signal for one of the traces, some false peaks may appear in the other traces. These false peaks tend to be highly asymmetric. DAx can be told to remove (or ignore) peaks that have an asymmetry that exceeds certain limits.
- command line analyses. Typically, trace analysers will be used to perform dozens to hundreds of analyses per day. To make it possible to analyse such large quantities of data, DAx can be run from batch files. All the data on an entire hard disk can be analysed with a single command, while highly flexible report files are created.

These features exceed the scope of this manual. Refer to the DAx User's Manual for details.