Using Automatic Trace Calibration to analyse hard-to-analyse trace sizing standards

In Fragment Length Analysis experiments, a sizing standard is generally included as one of the traces.
The sizing standard is known to contain fragments with certain lengths. In order to be able to set up a fragment size calibration based on the sizing standard, the peaks corresponding to the standard fragment sizes must be identified.

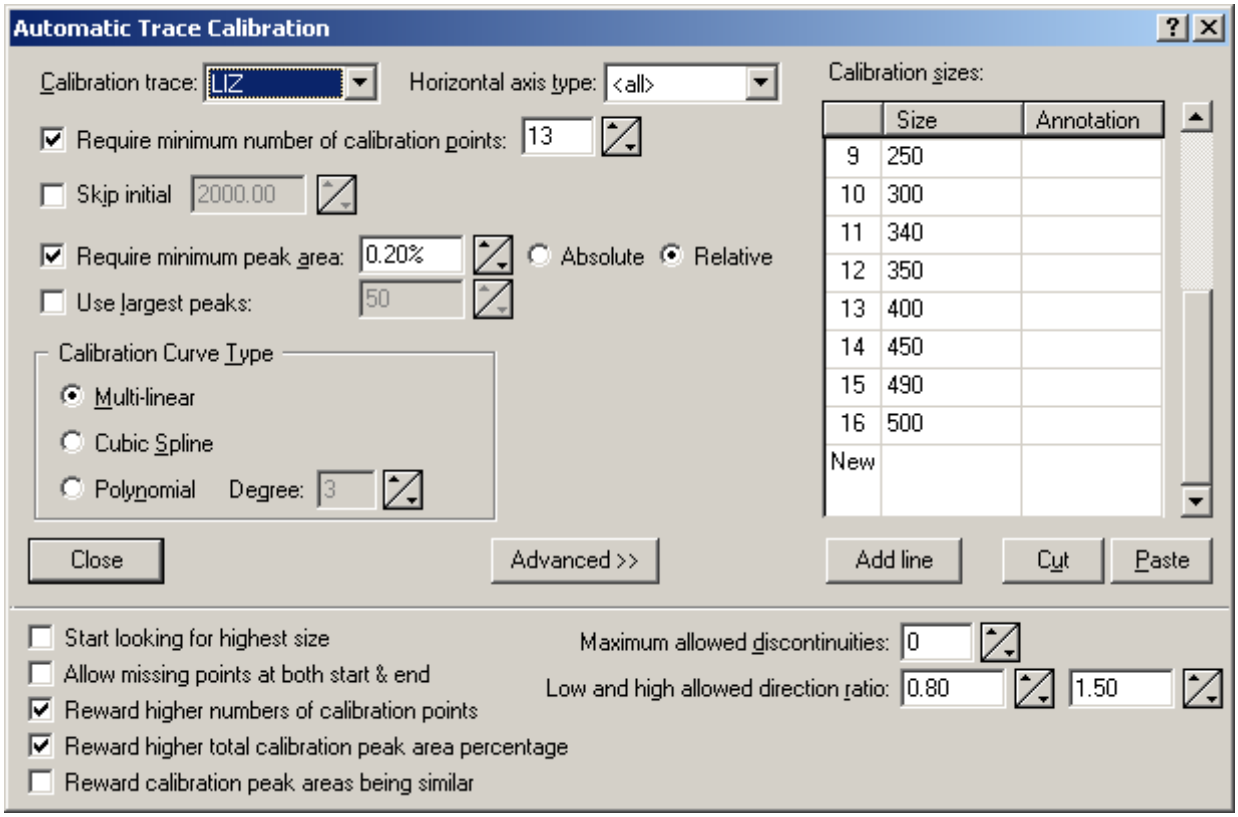
Recognising the standard fragments can be difficult to do with traditional techniques, such as looking in certain intervals, when one or more of the following occurs:

- presence of spurious actual peaks in between standard fragments
- presence of spurious peaks at the end of the measurement
- overloading of the sizing standard trace from one of the other traces

The new heuristic Automatic Trace Calibration technique introduced in DAx can be used to reliably find the standard fragments, even when problems such as the ones described above occur. This application note shows some examples of this.

Setting up the Automatic Trace Calibration parameters

The sizing standard used in these examples is ABI's LIZ-500, which contains fragment sizes 35, 50, 75, 100, 139, 150, 160, 200, 250, 300, 340, 350, 400, 450, 490, and 500.
The ATC parameters are entered into the ATC parameter dialog box, which is invoked using the **Analysis | Edit ATC** menu option.



The items in this dialog box have the following meaning:

- the **Calibration trace** defines the trace that contains the sizing standard, in this instance LIZ.
- **Require minimum number of calibration points** will often be set to the number of standard sizes, but can be set to a lower value if the sample may not have been analysed completely (i.e. if the analysis was cut short). Here it was set to 13, allowing the last three points to be missing.
- **Skip initial ...** can be used to disregard the early part of a measurement, which speeds up the analysis, but is not required.
- **Require minimum peak area ...** can be used to disregard small peaks, which again speeds up the analysis, but is not required. Values of 0.2% or 0.4% generally work well.
- The **Calibration curve type** defines what sort of curve will be drawn through the sizing calibration points.
- The **Calibration sizes** are simply a list of the sizes of the fragments in the sizing standard.

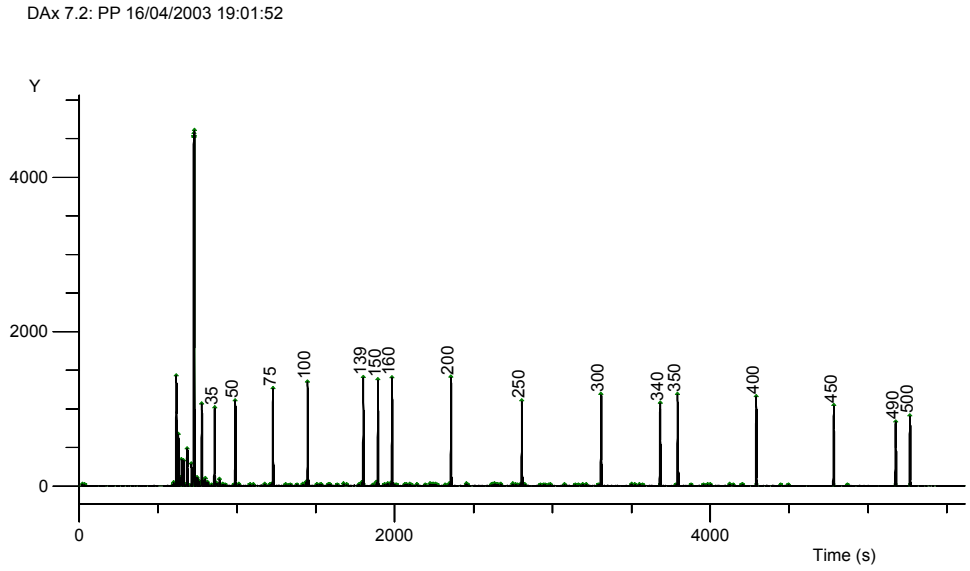
The **Advanced** items have the following meaning.

- **Start looking for highest size** runs the ATC algorithm “in reverse”, which can sometimes be useful.
- **Allow missing points at both start and end** allows for low end as well as high end calibration points to be missing.
- **Reward higher numbers of calibration points** tells the ATC algorithm that searches that turn up more calibration points are to be considered better.

- **Reward higher total calibration peak area percentage** tells the ATC algorithm that calibration points are likely to have peaks with significant sizes.
- **Reward calibration peak areas being similar** instructs the ATC algorithm to make use of the fact that calibration peaks probably have similar sizes.
- **Maximum allowed discontinuities** can be used if the measurement is known to contain a discontinuity in the horizontal axis. It is rarely necessary to allow discontinuities.
- The **low and high allowed direction ratios** determine how forgiving the ATC algorithm will be. More forgiving settings (further away from 1.0) will result in more inclusive searches, which may take longer to run. Generally, values of 0.7 and 1.5 give excellent results.

Results on unchallenging data

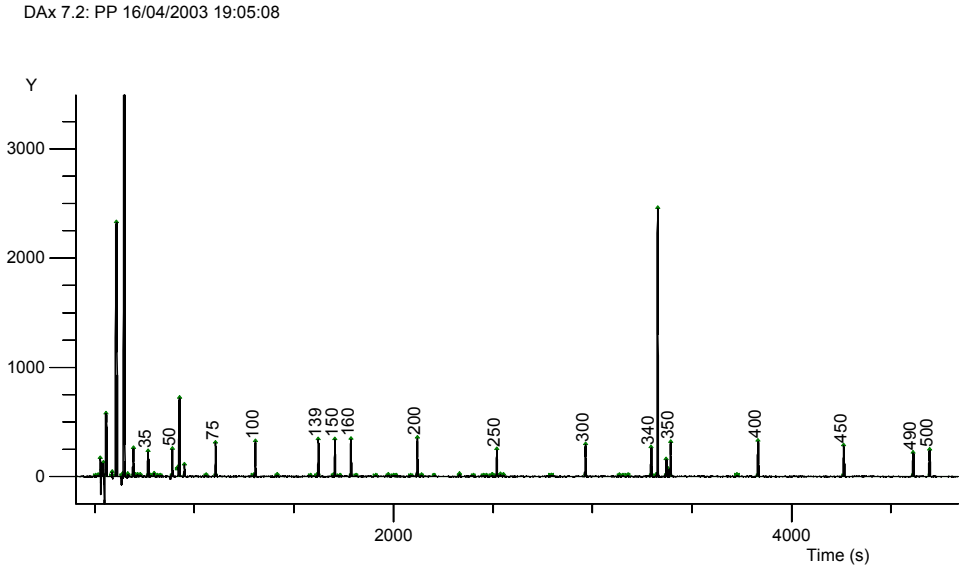
The following example shows an unchallenging standard trace.



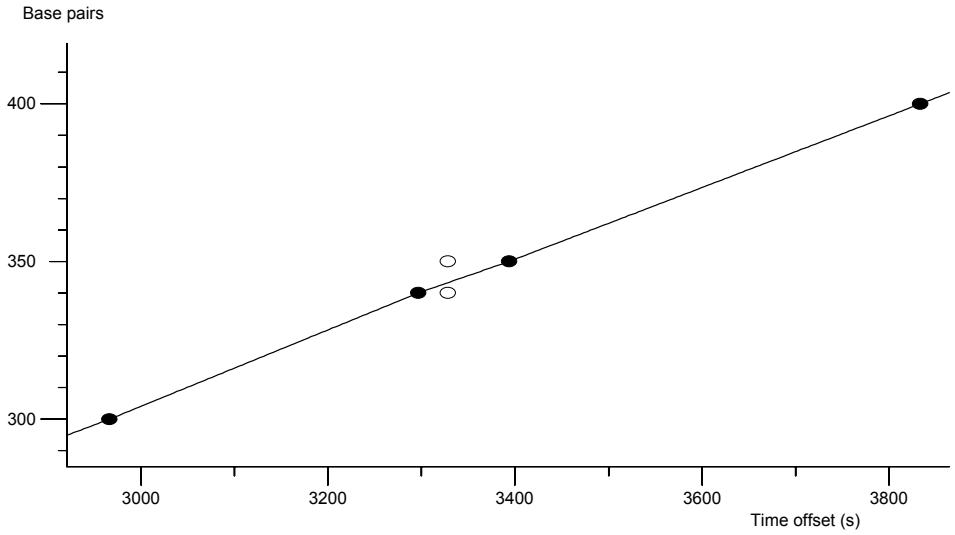
As can be seen, none of the peaks appear to be in any doubt.

Results on challenging data: spurious peak

The following example contains a high spurious peak between 340 and 350 bp that might be erroneously assigned as **340** or **350**.



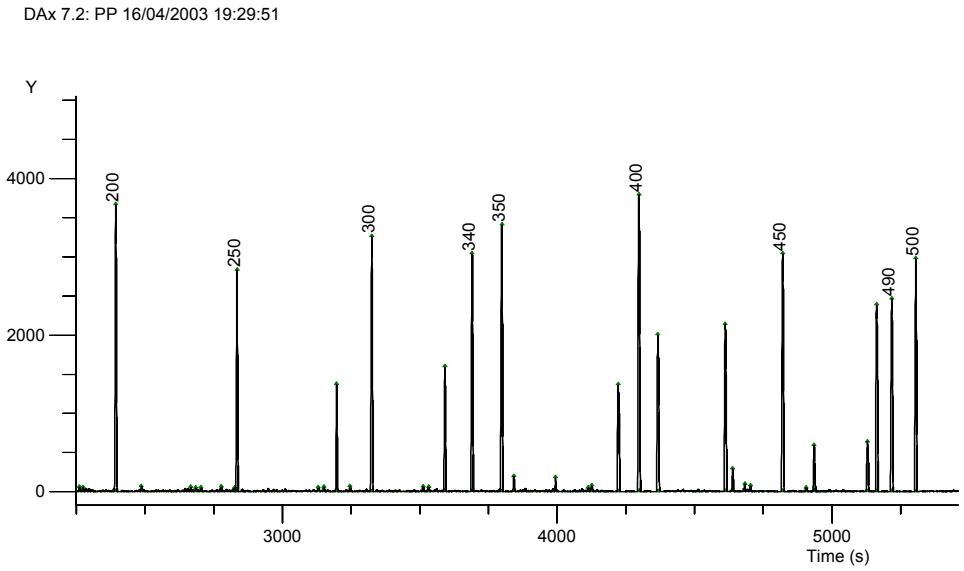
The following is an enlargement of the calibration between 300 and 400 bp.



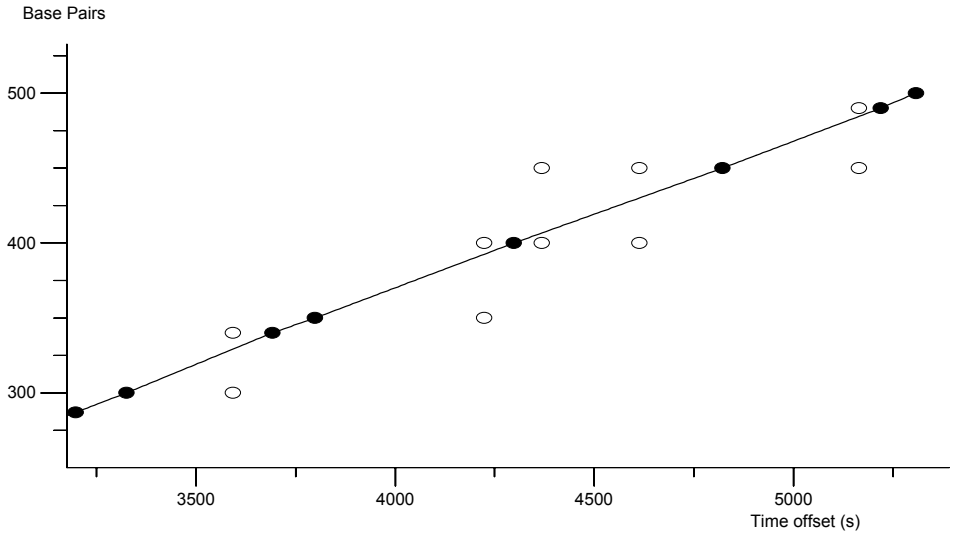
The two empty circles correspond to the high peak between 340 and 350 bp. It can readily be seen that the high peak should be assigned neither as 340 nor as 350 bp.

Results on challenging data: many spurious peaks

The following example contains many spurious peaks.



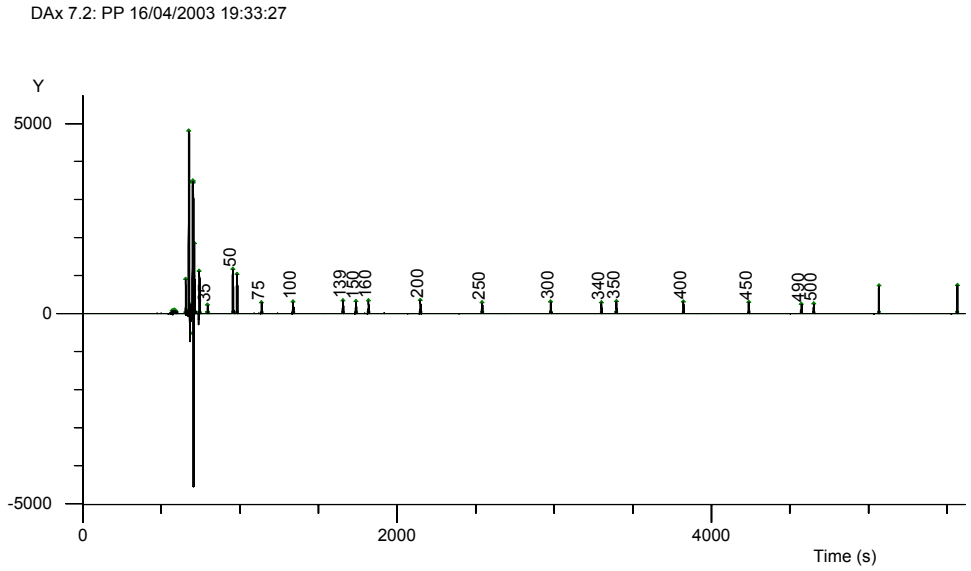
Here is an enlargement of the calibration between 300 and 500 bp, in which the spurious peaks have been added, displayed with the preceding and following fragment sizes.



Clearly, the ATC algorithm was correct in not assigning these spurious peaks as calibration points.

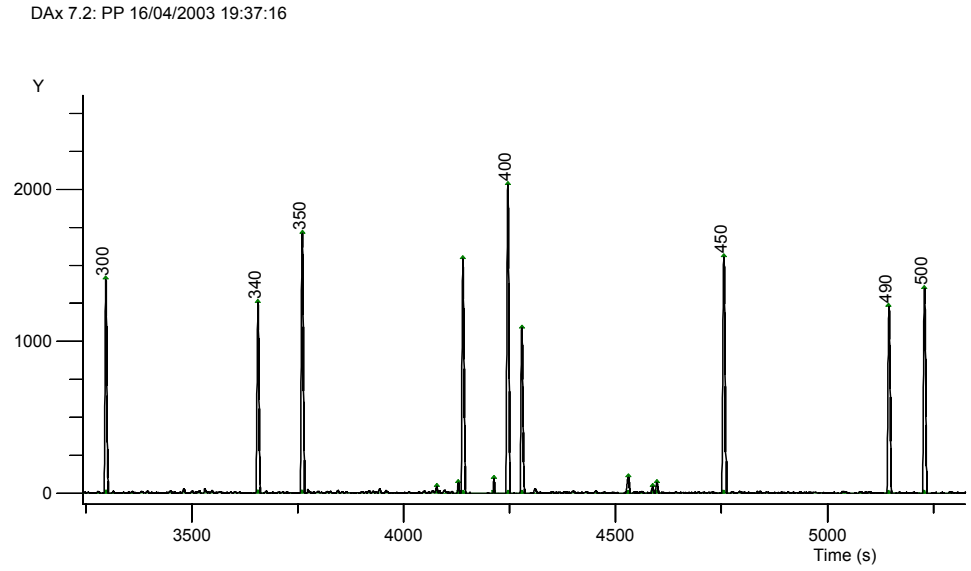
Results on challenging data: spurious peaks following end of measurement

This example shows two spurious peaks following the end of the measurement that were correctly ignored by the ATC search.

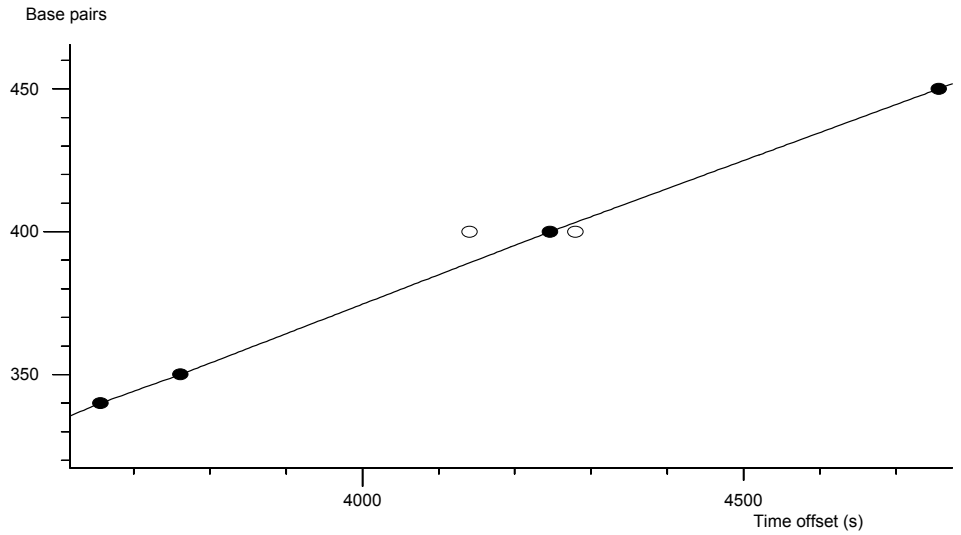


Results on challenging data: spurious peak due to overloading

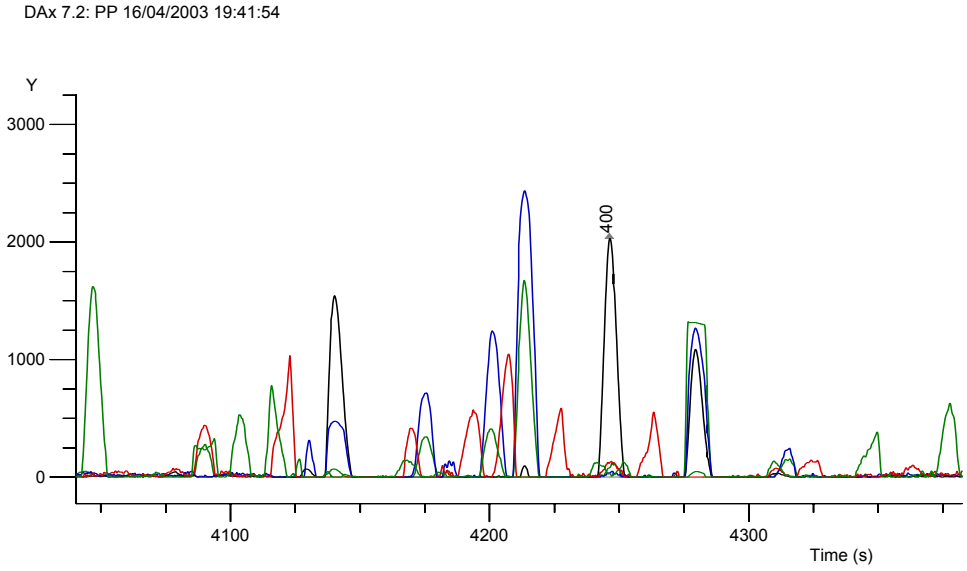
This example shows three peaks near 400 bp that might possibly be the calibration peak.



This is a detail view of the calibration, with the two extra peaks marked in:



As can be seen when all 5 traces in this example are loaded, the third of the “candidate” peaks (at 4280 seconds) is the result of cross-channel overloading, and so the ATC algorithm correctly did not assign it as the 400 bp peak.



Conclusion

In all the examples shown, the Automatic Trace Calibration algorithm does remarkably well at finding the correct sizing standard peaks, even in the presence of spurious extra peaks.