User Bulletin

ABI PRISM® GeneScan® Analysis Software for the Windows NT® Operating System

June 2002

SUBJECT: Overview of the Analysis Parameters and Size Caller

Introduction

In This User Bulletin

This user bulletin includes the following topics:

GeneScan Analysis Software Process 2
Analysis Parameters 3
Analysis Parameters Dialog Box 4
Data Processing: Smooth Options Parameter 5
Peak Detection: Min. Peak Half Width Parameter 7
Peak Detection: Polynomial Degree and Peak Window Size Parameters 8
Peak Detection: Slope Threshold for Peak Start and Slope Threshold for Peak End Parameters 16
Baselining: Baseline Window Size Parameter 20
Size Caller 30

Purpose

This user bulletin supplements the *ABI PRISM® GeneScan Analysis Software version 3.7 User Guide* (P/N 4308923) to further explain the analysis parameters and size caller available in the Windows NT® version of the software.

The *GeneScan Analysis Software v3.7.1 Updater CD* (P/N 4336026) includes new analysis parameter default values. For additional information and installation instructions, refer to the GeneScan v3.7.1 About file.

Intended Audience

This document is intended for users familiar with the GeneScan analysis software for the Macintosh® operating system who are now using the software on the Windows NT operating system.

Applied Biosystems

GeneScan Analysis Software Process

Overview

The ABI PRISM® GeneScan Analysis Software is available in versions for both the Windows NT operating system and the Macintosh operating system. The Windows NT version of the software uses different algorithms and has additional analysis parameters that give users more control with data analysis.

Flowchart

The following flowchart shows how GeneScan analysis software analyzes data.

Note: For multicapillary instruments, multicomponenting is performed by the data collection software.

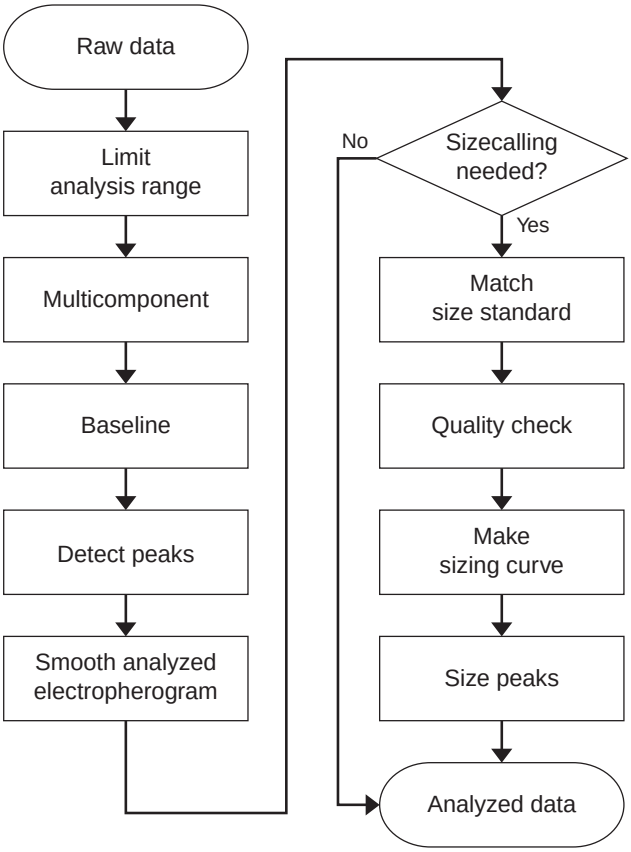


Figure 1 Simplified GeneScan analysis software flowchart

Analysis Parameters

Table of Parameters

The following table lists the analysis parameters:

Parameter Status	Parameter	Discussed in...
Unchanged from Macintosh versions	• Analysis Range • Size Call Range • Size Calling Method • Peak Amplitude Thresholds	*ABI PRISM® GeneScan Analysis Software Version 3.7 User Guide*
Changed from Macintosh versions	• Smooth Options • Min. Peak Half Width	this user bulletin and the *ABI PRISM® GeneScan Analysis Software Version 3.7 NT and 3.1 Macintosh User Guides*
Added for the Windows NT version	• Polynomial Degree • Peak Window Size • Slope Threshold for Peak Start • Slope Threshold for Peak End • Window Size	this user bulletin and the *ABI PRISM® GeneScan Analysis Software Version 3.7 User Guide*
Removed options from the Windows NT version	Baseline Multicomponent	*ABI PRISM® GeneScan Analysis Software version 3.1 User's Manual*

Analysis Parameters Dialog Box

About the Analysis Parameters Dialog Box

Use the Analysis Parameters dialog box to set analysis parameter values for data processing.

The default analysis parameter values are analysis guidelines. This bulletin should serve as a guide for modifying these values as appropriate for each laboratory.

Example

Figure 2 shows the Analysis Parameters dialog box with default values for GeneScan analysis software v3.7.1 on the Windows NT operating system.

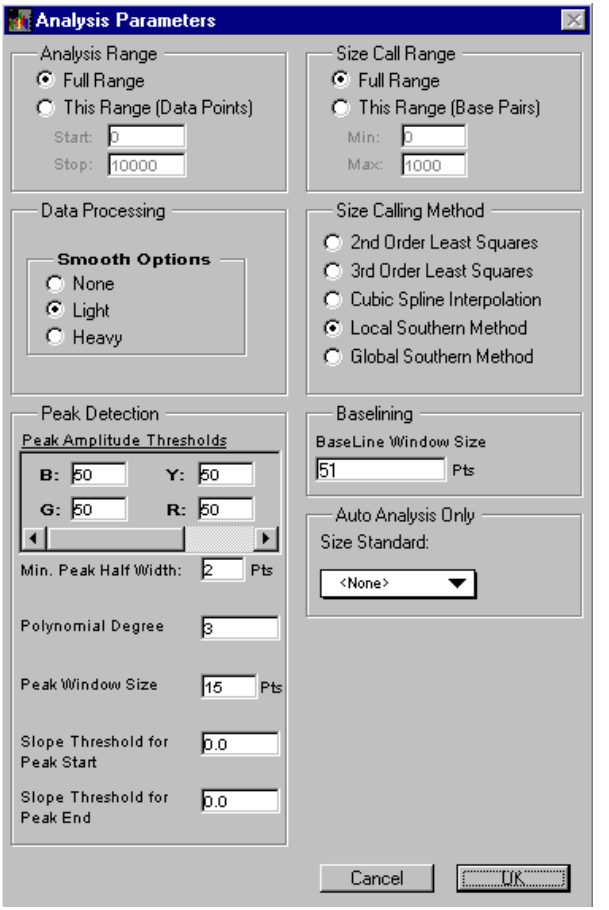


Figure 2 Analysis Parameters dialog box displaying default values

Data Processing: Smooth Options Parameter

About the Parameter

The Smooth Options parameter sets the degree of smoothing applied to the display of the analyzed electropherogram. Smoothing may aid in data interpretation.

How the Parameter Works

The Smooth Options parameter is applied after peak detection and affects only the display of analyzed electropherograms. The peak heights and areas are calculated and displayed in the tabular data display based on the “none” smoothing option. Selecting light or heavy smoothing will not affect the calculation of these values.

Smoothing Example

Figure 3 shows the peaks from the same sample file after analysis using no smoothing (black); light smoothing (green); and heavy smoothing (red). All tabular data, including peak height and area, remain unchanged.

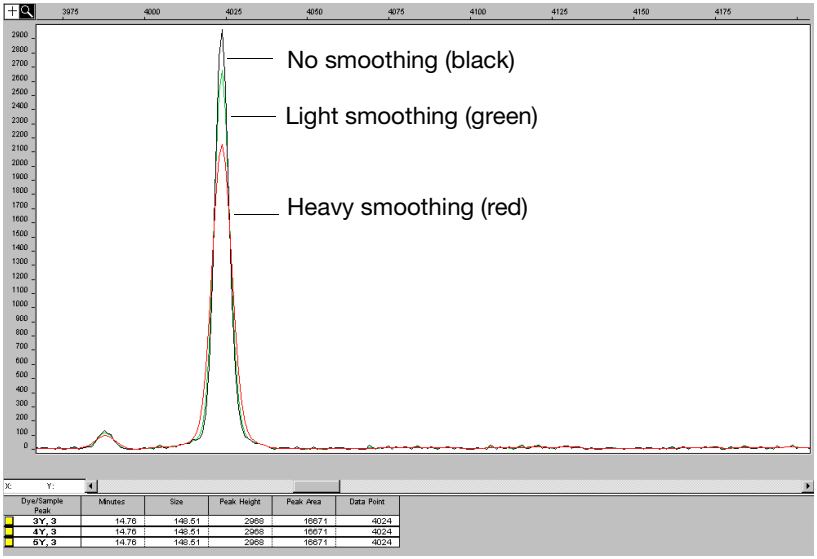


Figure 3 Electropherogram showing the effects of smoothing on peaks from the same sample file

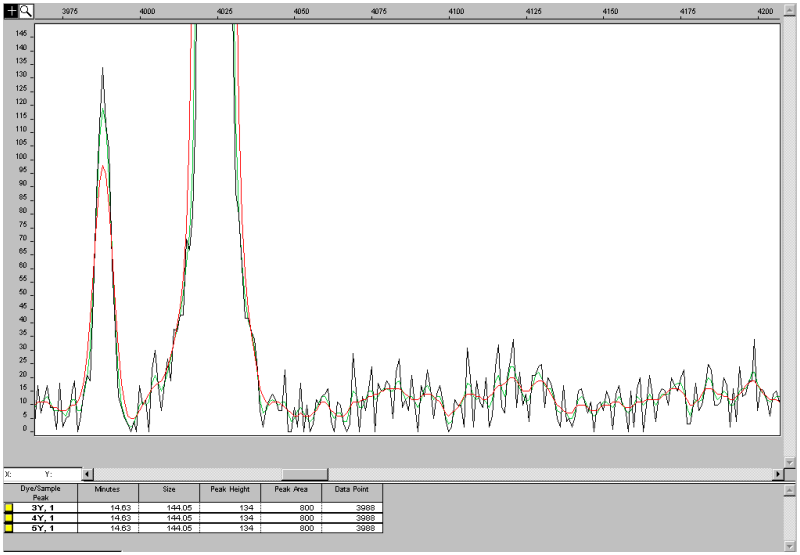


Figure 4 Electropherogram showing the effects of smoothing on the smaller peak and baseline by changing the y scale from Figure 3

Peak Detection: Min. Peak Half Width Parameter

About This Parameter

Use the Min. Peak Half Width parameter to specify the smallest full width at half maximum height for peak detection. This parameter can be used to ignore noise spikes.

How This Parameter Works

The Min. Peak Half Width parameter defines what constitutes a peak. The software ignores peak half widths smaller than the specified value.

The way in which this version of the software defines the minimum peak half width is different than in previous versions.

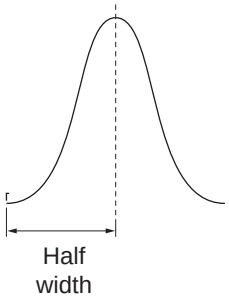
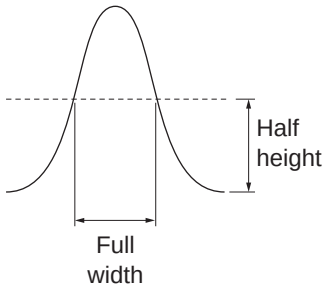
Old Versions	Current Version
	
Half width of the peak measured from peak start	Full width of the peak measured at half its height

Figure 5 Defining the Min. Peak Half Width

Peak Detection: Polynomial Degree and Peak Window Size Parameters

About These Parameters

Use the Polynomial Degree and the Peak Window Size settings to adjust the sensitivity of the peak detection. You can adjust these parameters to detect a single base pair difference while minimizing the detection of shoulder effects or noise.

Sensitivity increases with larger polynomial degree values and smaller window size values. Conversely, sensitivity decreases with smaller polynomial degree values and larger window size values.

How These Parameters Work

The peak window size functions with the polynomial degree to set the sensitivity of peak detection.

The peak detector computes the first derivative of a polynomial curve fitted to the data within a window that is centered on each data point in the analysis range.

Using curves with larger polynomial degree values allows the curve to more closely approximate the signal and, therefore, the peak detector captures more peak structure in the electropherogram.

The peak window size sets the width (in data points) of the window to which the polynomial curve is fitted to data. Higher peak window size values smooth out the polynomial curve, which limits the structure being detected. Smaller window size values allow a curve to better fit the underlying data.

How to Use These Parameters

Use the table below to adjust the sensitivity of detection.

To...	Polynomial Degree Value	Window Size Value
Increase sensitivity use...	Higher	Lower
Decrease sensitivity use...	Lower	Higher

Guidelines for Using These Parameters

To detect well-isolated, base-line-resolved peaks, use polynomial degree values of 2 or 3. For finer control, use a degree value of 4 or greater.

As a guideline, set the peak window size (in data points) to be about 1 to 2 times the full width at half maximum height of the peaks that you want to detect.

Examining Peak Definitions

To examine how GeneScan Analysis software has defined a peak, select **View** > **Show Peak Positions**. The peak positions, including the beginning, apex, and end of each peak, are tick-marked in the electropherogram.

Effects of Varying the Polynomial Degree

Figure 6 depicts peaks detected with a window size of 15 data points and a polynomial curve of degree 2 (green); 3 (red); and 4 (black). The diamonds represent a detected peak using the respective polynomial curves.

Note that the smaller trailing peak is not detected using a degree of 2 (green). As the peak detection window is applied to each data point across the displayed region, a polynomial curve of degree 2 could not be fitted to the underlying data to detect its structure.

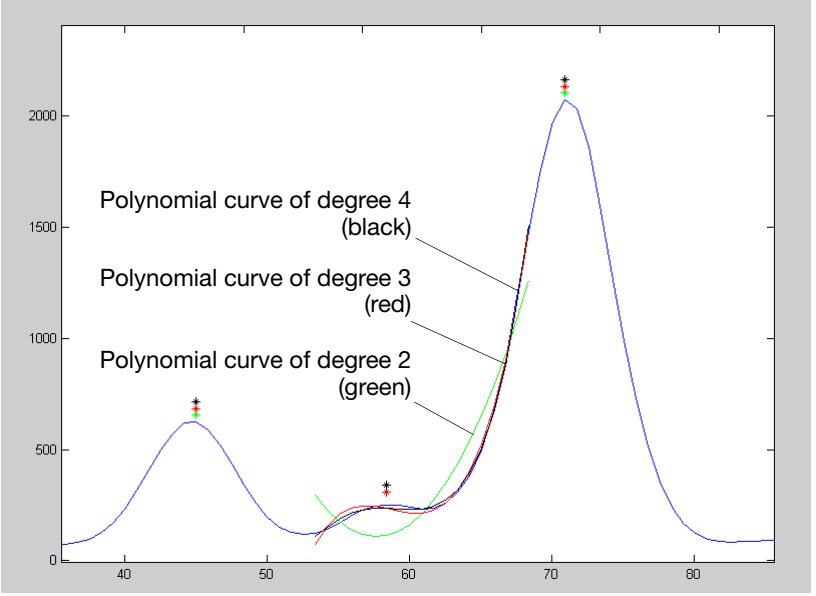


Figure 6 Electropherogram showing peaks detected with the same window size and three different polynomial degrees

Effects of Increasing the Window Size Value

Figure 7 shows the same peaks that are shown in Figure 6. However, in this depiction both polynomial curves have a degree of 3 and the window size value was increased from 15 (red) to 31(black) data points.

As the cubic polynomial is stretched to fit the data in the larger window size, the polynomial curve becomes smoother. Note that the structure of the smaller trailing peak is no longer detected as a distinct peak from the adjacent larger peak to the right.

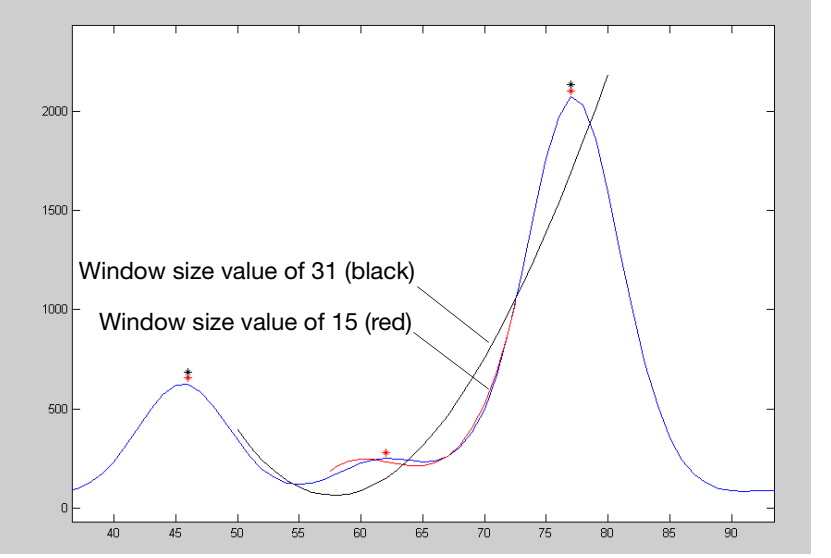


Figure 7 Electropherogram showing the same peaks as in Figure 6 after increasing the window size value while keeping the polynomial degree the same

Optimizing Peak Detection Sensitivity Example 1

Initial Electropherogram

Figure 8 shows two resolved alleles of known fragment lengths (that differ by one nucleotide) detected as a single peak. The analysis was performed using a polynomial degree of 3 and a peak window size of 19 data points.

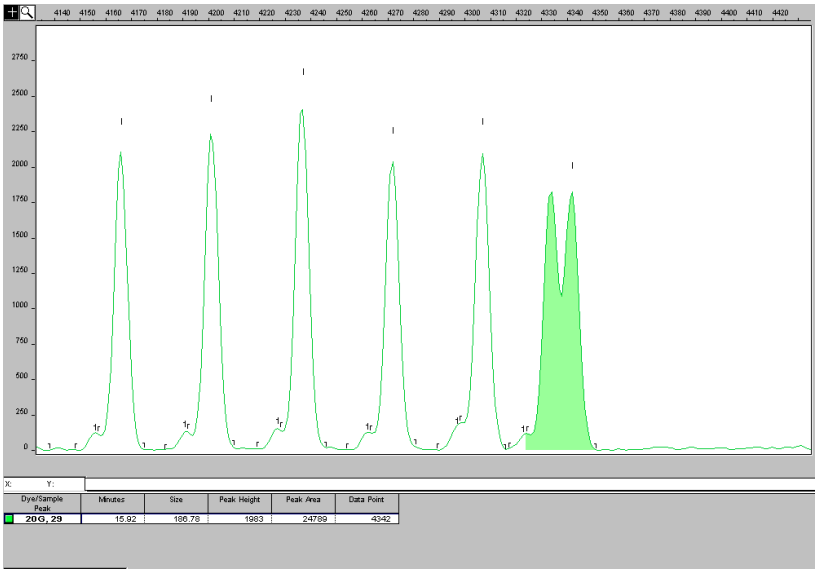


Figure 8 Electropherogram showing two resolved alleles detected as a single peak

Note: For information on the tick marks displayed in the electropherogram, see “Examining Peak Definitions” on page 9.

Effects of Decreasing the Window Size Value

Figure 9 shows that both alleles are detected after re-analyzing with the polynomial degree set to 3 while decreasing the window size value to 15 (from 19) data points.

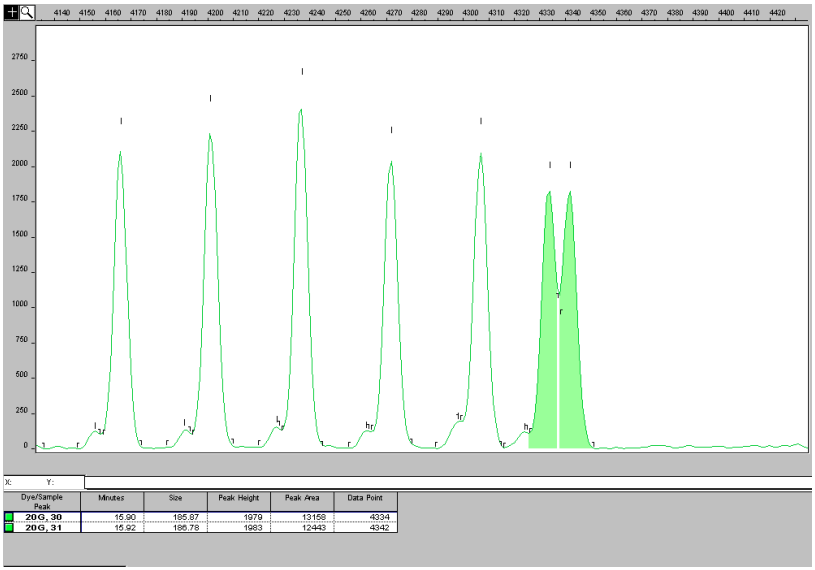


Figure 9 Electropherogram showing the alleles detected as two peaks after decreasing the window size value

Optimizing Peak Detection Sensitivity Example 2

Initial Electropherogram

Figure 10 shows an analysis performed using a polynomial degree of 3 and a peak window size of 19 data points.

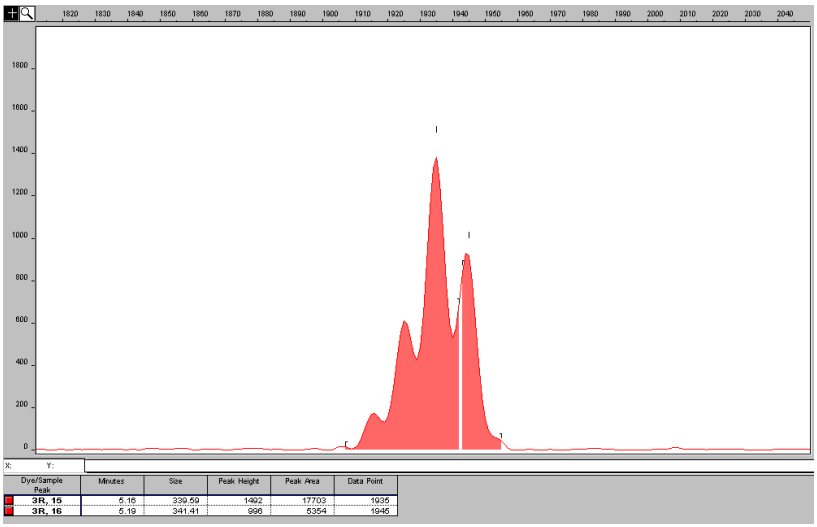


Figure 10 Electropherogram showing four resolved peaks detected as two peaks

Effects of Reducing the Window Size Value and Increasing the Polynomial Degree Value

Figure 11 shows the data presented in Figure 10 re-analyzed with a window size value of 10 and polynomial degree value of 5.

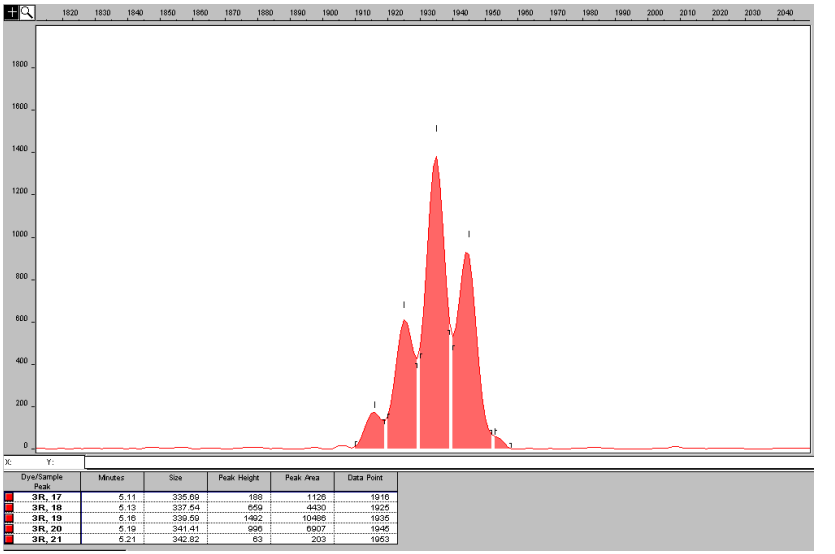


Figure 11 Electropherogram showing all four peaks detected after reducing the window size value and increasing the polynomial degree value

Optimizing Peak Detection Sensitivity Example 3

Effects of Extreme Settings

Figure 12 shows the result of an analysis using a peak window size value set to 10 and a polynomial degree set to 9. This extreme setting for peak detection led to several peaks being split and detected as two separate peaks.

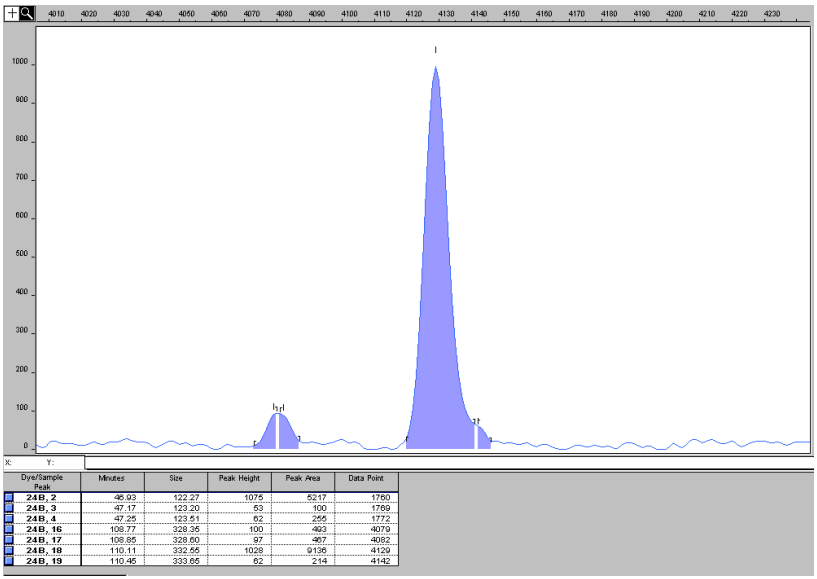


Figure 12 Electropherogram showing the result of an analysis using extreme settings for peak detection

Peak Detection: Slope Threshold for Peak Start and Slope Threshold for Peak End Parameters

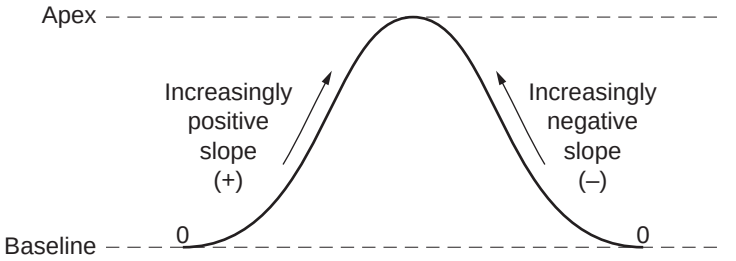
About These Parameters

Use the Slope Threshold for Peak Start and Slope Threshold for Peak End parameters to adjust the start and end points of a peak.

This parameter can be used to better position the start and end points of an asymmetrical peak, or a poorly resolved shouldering peak, to more accurately reflect the peak position and area.

How These Parameters Work

In general, from left to right, the slope of a peak increases from the baseline up to the apex. From the apex down to the baseline, the slope becomes decreasingly negative until it returns to zero at the baseline.



If either of the slope values you have entered exceeds the slope of the peak being detected, the software overrides your value and reverts to zero.

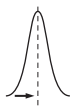
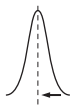
Guidelines for Using These Parameters

As a guideline, use a value of zero for typical or symmetrical peaks. Select values other than zero to better reflect the beginning and end points of asymmetrical peaks.

A value of zero will not affect the sizing accuracy or precision for an asymmetrical peak.

Using These Parameters

Use the table below to move the start or end point of a peak.

IF you want to move the...	THEN change the...
start point of a peak closer to its apex 	Slope Threshold for Peak Start value from zero to a positive number
end point of a peak closer to its apex 	Slope Threshold for Peak End value to an increasingly negative number

Note: The size of a detected peak is the calculated apex between the start and end points of a peak and will not change based on your settings.

Slope Threshold Example

Initial Electropherogram

The initial analysis with a value of 0 for both the Slope Threshold for Peak Start and the Slope Threshold for Peak End value produced an asymmetrical peak with a noticeable tail on the right side.

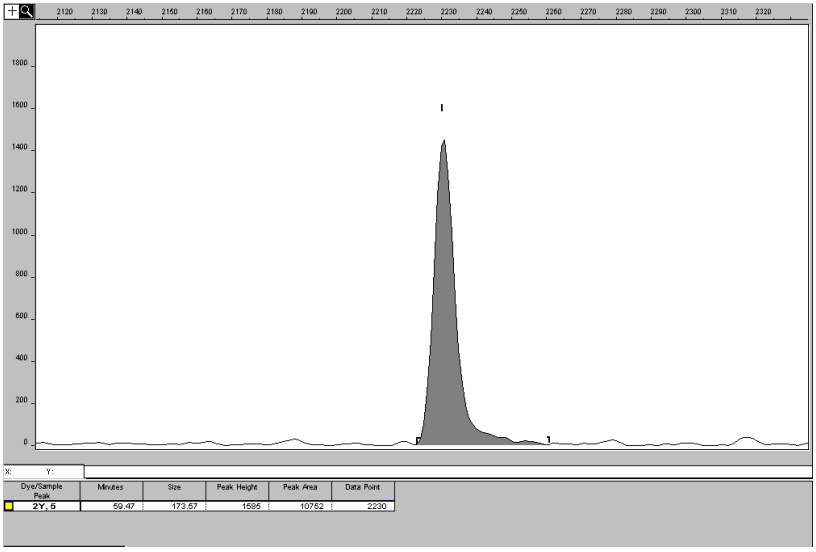


Figure 13 Electropherogram showing an asymmetrical peak

Electropherogram After Adjustments

After re-analyzing with a value of –35.0 for the Slope Threshold for Peak End, the end point that defines the peak moves closer to its apex, thereby removing the tailing feature. Note that the only change to tabular data was the area (peak size and height are unchanged).

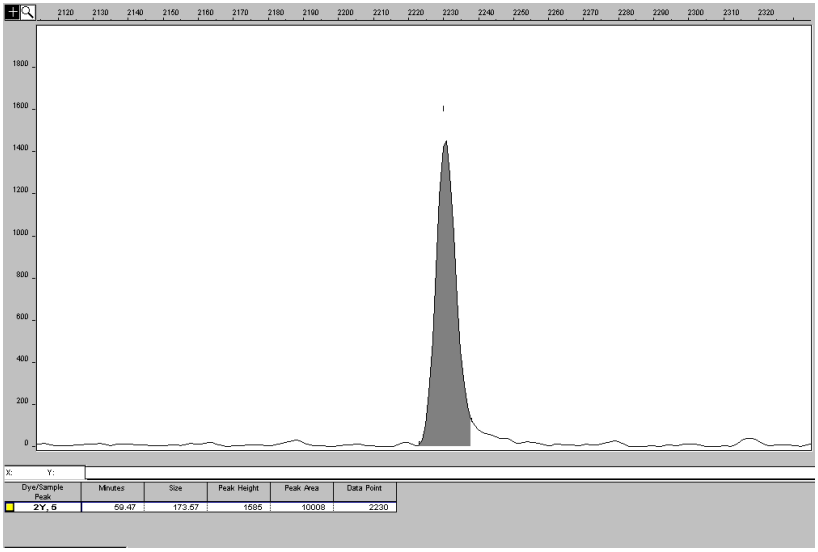


Figure 14 Electropherogram showing the effect of changing the slope threshold for peak end

Baselining: Baseline Window Size Parameter

About This Parameter

Use the Baseline Window Size parameter to control the baseline for a group of peaks.

How This Parameter Works

The software determines a reference baseline value for each data point. In general, the software sets the reference baseline to be the lowest value that it detects in a specified window size (in data points) centered on each data point.

A small baseline window relative to the width of a cluster, or grouping of peaks spatially close to each other, can result in shorter peak heights.

Larger baseline windows relative to the peaks being detected can create an elevated baseline, resulting in peaks that are elevated or not baseline resolved.

Guidelines for Using This Parameter

As a guideline, choose a value that encompasses the width in data points of the peaks being detected while preserving a qualitatively smooth baseline.

The trade-off for a smoother baseline that touches all peaks is a reduction in peak height.

Baselining Example

Figure 15 depicts an allelic ladder containing clusters of alleles. The alleles have been labeled with green dye and the data displayed has been multicomponented, but not baselined. The electropherogram spans approximately 2800 data points.

The red, blue, and black traces depict various reference baselines (zero in the analyzed electropherogram) that result from different baseline window size settings. These reference baselines are subtracted from the sample data during baselining. In Figure 15:

- The red trace depicts the reference baseline that results from an extreme baseline window size value of 2801. At this setting, the reference baseline does not touch all peaks, resulting in elevated peak heights.
- The blue trace depicts the reference baseline that results from the default value of 51 data points.

- The black trace depicts the reference baseline that results from an extreme baseline window size value of 5 data points. At this setting, the peaks are tracked too closely by the reference baseline, resulting in significantly reduced peak height.

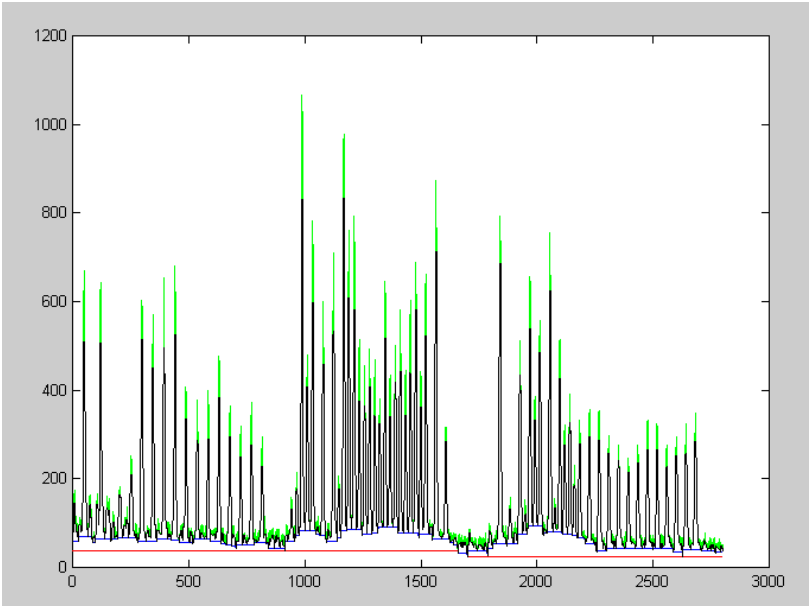


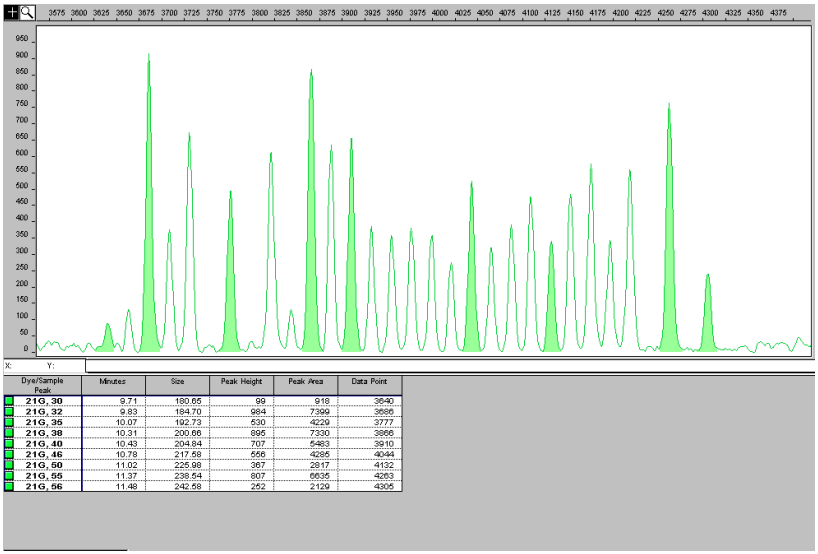
Figure 15 Depiction of the baselining of an electropherogram

Baselining Example 1

Initial Electropherogram

Figure 16 shows a portion of the electropherogram shown in Figure 15, which depicts various window sizes.

The electropherogram shows the default Baseline Window Size value of 51 that appears in Figure 15 as the blue trace. Note that all peaks in this cluster have been baselined.



Dye/Sample Peak	Minutes	Size	Peak Height	Peak Area	Data Point
21G, 30	9.71	180.65	99	918	3640
21G, 32	9.83	184.70	984	7399	3686
21G, 36	10.07	192.73	530	4229	3777
21G, 38	10.31	200.66	895	7330	3866
21G, 40	10.43	204.84	707	5483	3910
21G, 46	10.78	217.58	556	4285	4044
21G, 50	11.02	225.98	367	2817	4132
21G, 55	11.37	238.54	807	6635	4263
21G, 56	11.48	242.58	252	2129	4305

Figure 16 Electropherogram showing an allelic ladder with a cluster of peaks

Effects of Extreme Increase of the Baseline Window Size

Figure 17 shows an extreme Baseline Window Size value of 2801 that appears in Figure 15 as the red trace. (2801 is approximately the width in data points of all the peaks shown.) This increase resulted in an overall raised baseline and many elevated peaks within the cluster.

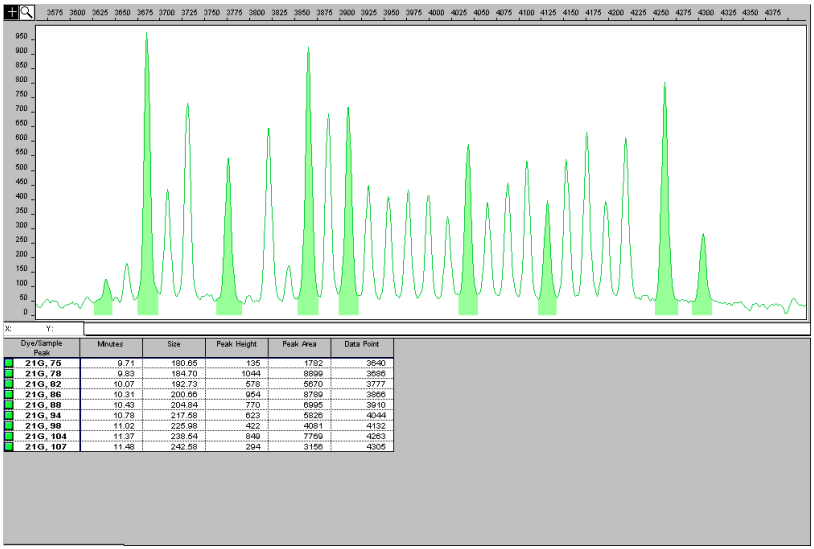
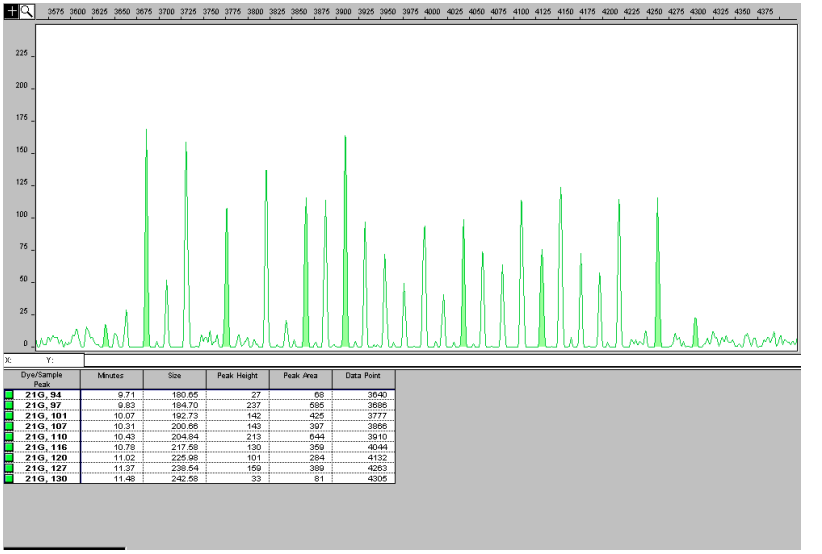


Figure 17 Electropherogram showing a raised baseline caused by an increase in the baseline window size value

Effects of Extreme Decrease of the Baseline Window Size

Figure 18 shows an extreme Baseline Window Size value of 5 that appears in Figure 15 as the black trace. (Five is much smaller than the width in data points for any of the peaks prior to baselining.) This decrease resulted in a significant decrease in the peak heights.



Dye/Sample Peak	Minutes	Size	Peak Height	Peak Area	Data Point
21G, 94	9.71	180.65	27	68	3640
21G, 97	9.83	184.70	237	585	3686
21G, 101	10.07	192.73	142	425	3777
21G, 107	10.31	200.66	143	397	3866
21G, 110	10.43	204.84	213	644	3910
21G, 116	10.78	217.58	130	359	4044
21G, 120	11.02	225.98	101	284	4132
21G, 127	11.37	238.54	159	389	4263
21G, 130	11.48	242.58	33	81	4305

Figure 18 Electropherogram showing significantly reduced peak heights caused by a reduction in the baseline window size value

Baselining Example 2

Initial Electropherogram

Figure 19 shows the electropherogram from an analysis of a cluster of peaks using the default Baseline Window Size value of 51 data points.

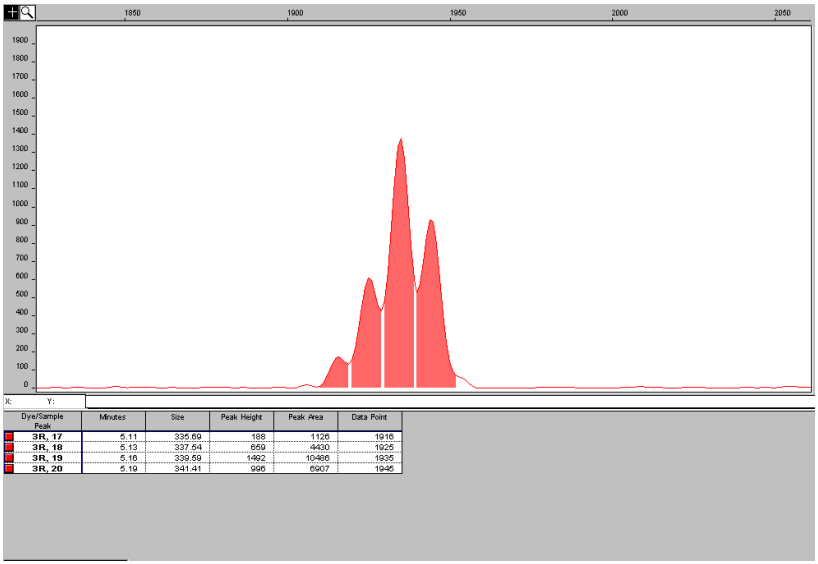
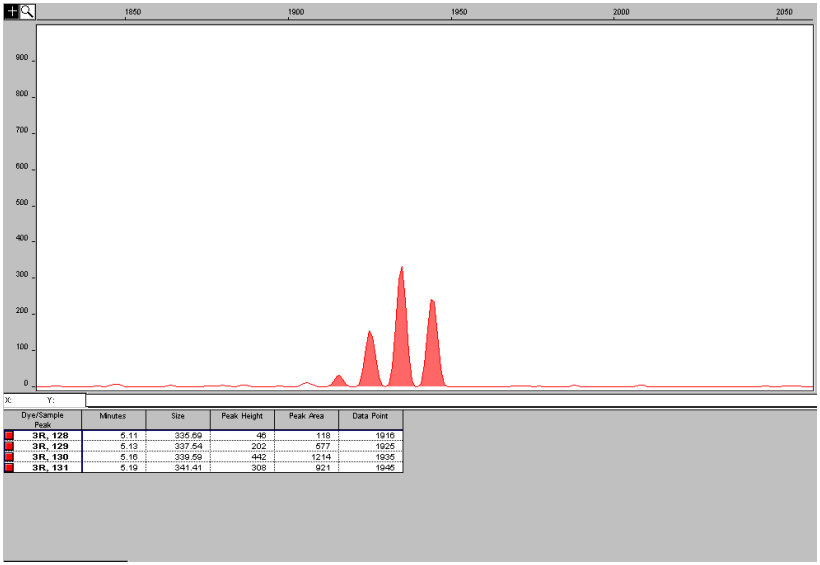


Figure 19 Electropherogram showing a typical result using the default baseline window size value

Effects of Extreme Decrease of the Baseline Window Size

Figure 20 shows the re-analysis of the electropherogram shown in Figure 19 with an extreme Baseline Window Size value of 5. All peaks within the cluster have been baselined and have a reduced peak height.



Dye/Sample Peak	Minutes	Size	Peak Height	Peak Area	Data Point
3R, 128	5.11	335.69	46	118	1916
3R, 129	5.13	337.54	202	577	1925
3R, 130	5.16	339.59	442	1214	1935
3R, 131	5.19	341.41	308	921	1945

Figure 20 Electropherogram showing dramatically reduced peak heights caused by a reduction in the baseline window size value

Baselining Example 3

Raw Data

The data in the electropherogram has been multicomponented but not baselined. There are two pull-down peaks in the blue trace below the two major green peaks (see arrows).

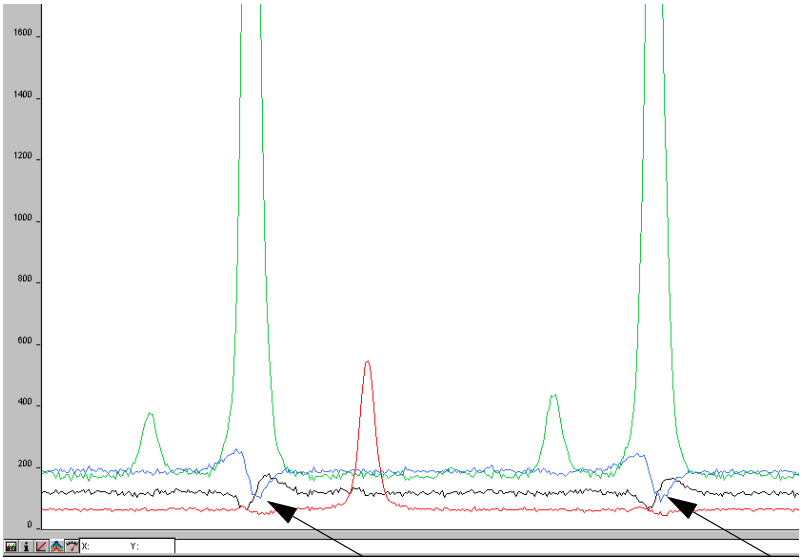


Figure 21 Electropherogram showing raw data that has been multicomponented but not baselined

Raised Baseline

After analyzing with a baseline window size of 251 data points, the low points represented in the blue trace (within this 251 data point window) are set to zero. This results in a raised baseline between these points.

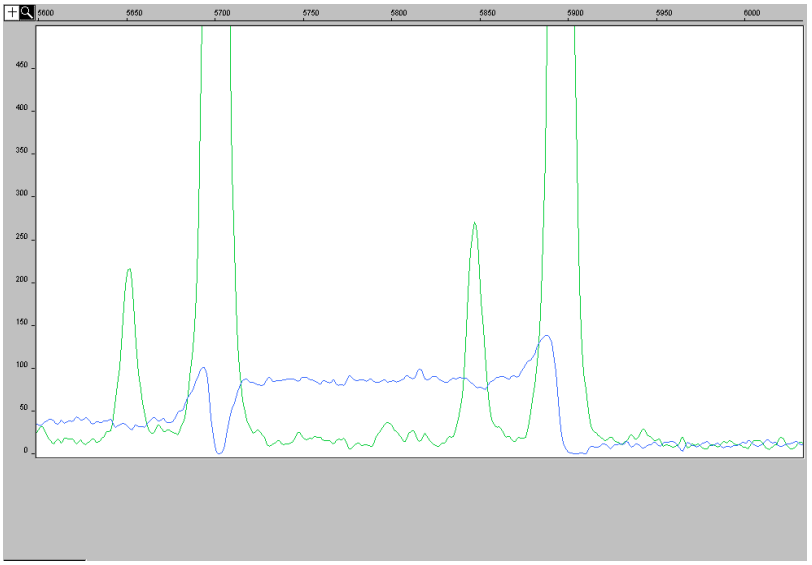


Figure 22 Electropherogram showing a raised baseline

Eliminating Raised Baseline

After re-analyzing with a baseline window size of 51 data points (a window size range between the pull-down peaks), the raised baseline is eliminated. This results in a more accurate baseline.

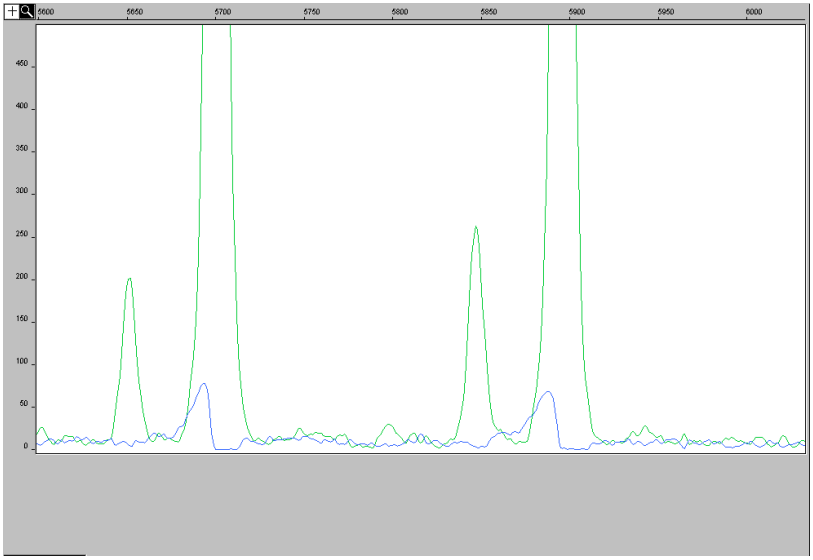


Figure 23 Electropherogram showing a more accurate baseline

Size Caller

About the Size Caller

The size caller matches size-standard peaks with a quality check.

How the Size Caller Works

The way in which the fragment sizes are calculated has not changed from previous versions of the software (*e.g.,* local southern). However, the way in which the Windows NT version of the software identifies the size standard is different from previous versions.

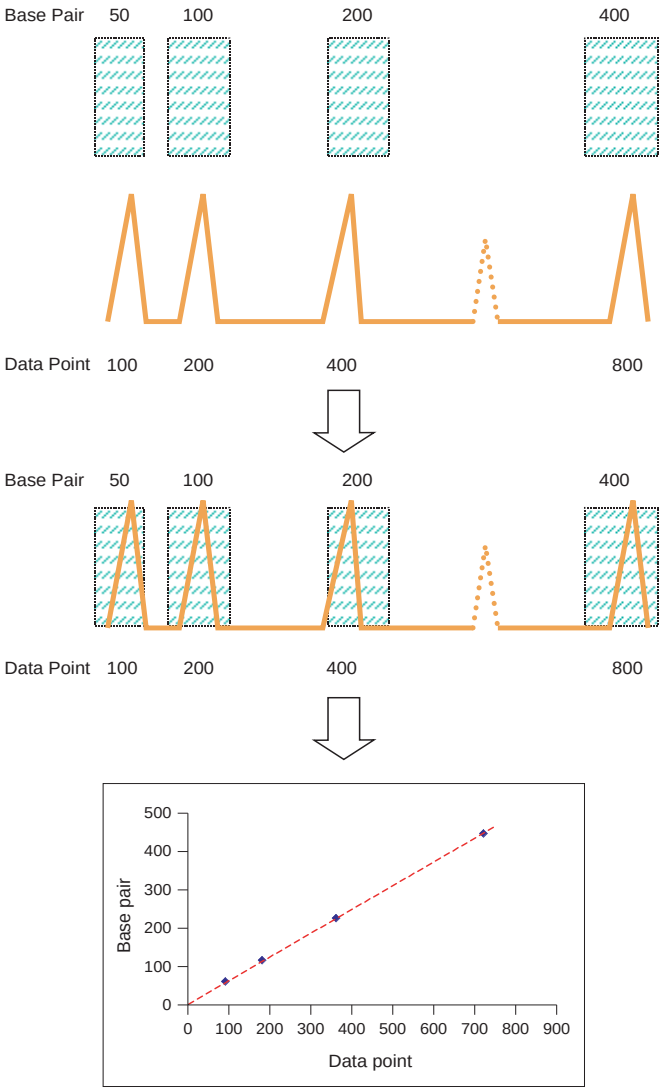
Method for Identifying the Size Standard

Macintosh Versions	Windows NT Version
User assigns fragment sizes to particular peaks based on scan number	Software matches the size standard fragments by ratio matching based on relative distance between neighboring peaks

Macintosh Version

In GeneScan analysis software for the Macintosh operating system, the size standard peaks are identified based on their assignment within a run or a previous run.

Anomalous peaks outside of a ±10 data point bin are ignored, but those within the bins can be incorrectly called resulting in an incorrect size curve. In that case, you must redefine a new size standard for that particular sample.



Defining the Size Standard The boxes show a ±10 data point range used to identify size standard peaks in subsequent runs

Data (100, 200, 400, and 800) shown with anomalous peak dotted

Assigning Peaks that fall into the correct range. The anomalous peak is ignored.

Generating the Size Standard Curve for sample file using specified sizing method, *e.g.,* Local Southern

Figure 24 Peak identification with GeneScan analysis software for the Macintosh operating system

Windows NT Version

GeneScan analysis software for the Windows NT operating system uses ratio matching to identify the size standard fragments.

Ratio matching does not rely on the manual assignment of size standard definitions (in base pairs) to their associated data points within a run or a previous run. Selecting a peak in the electropherogram to enter an associated value in the Size column now serves only as a guide. Simply listing the values to be used for sizing as an array of numbers without regard to the highlighted peak is sufficient.

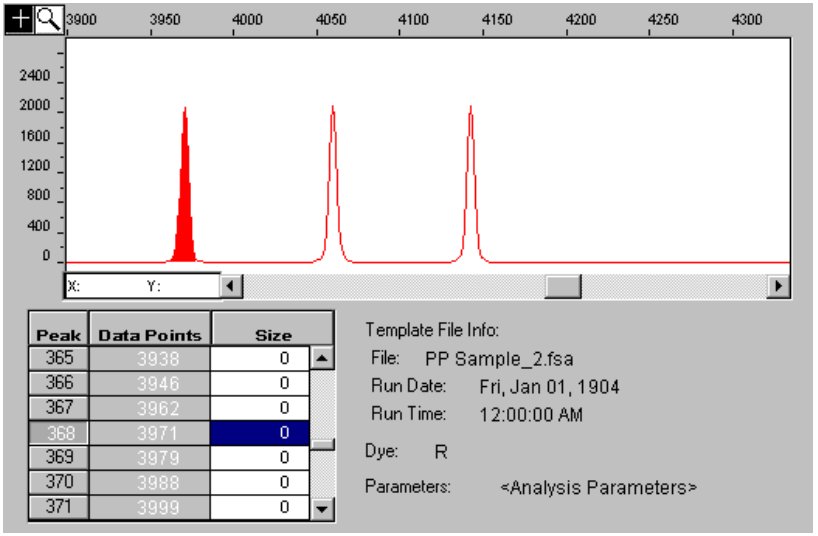
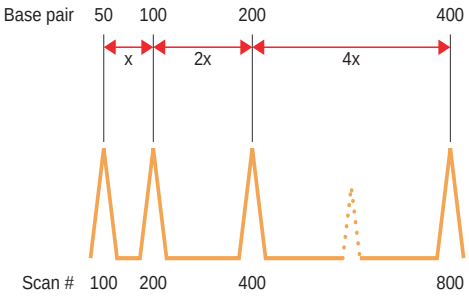
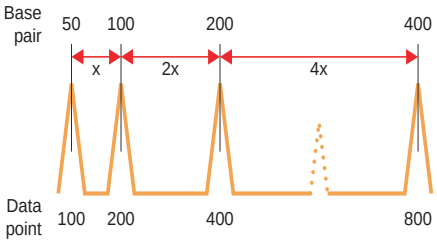


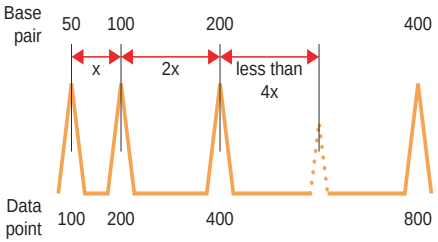
Figure 25 Electropherogram showing a selected peak and the associated value in the Size column

The size caller ignores anomalous peaks that do not match the expected ratio. The size caller constructs a best-fit curve using the data points of each size standard fragment detected. A comparison between the sizes calculated from the best-fit curve and the matched peaks from the size standard definition using the array of numbers is performed. Size calling will fail if significant differences are found or if no match can be made based on the expected ratios. (In Figure 26, that is x, 2x, and 4x.) Additionally, you may find that one of the size fragments has not been identified, even though it was listed as part of the definition. The size caller has been designed to allow the exclusion of one of the listed values to obtain a better match. To use an excluded fragment, try the steps outlined in Figure 27.



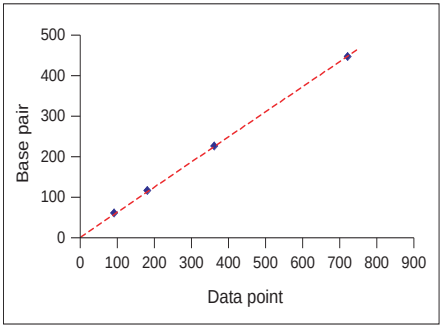
Defining the Size Standard from a list of sizes (50, 100, 200, and 400) used to calculate the expected ratios (in red)





Data (100, 200, 400, and 800) shown with anomalous peak dotted

Assigning Peaks that match the expected ratio. The anomalous peak is ignored



Generating the Size Standard Curve when a good ratio match is found

Figure 26 Peak identification with GeneScan analysis software for the Windows NT operating system

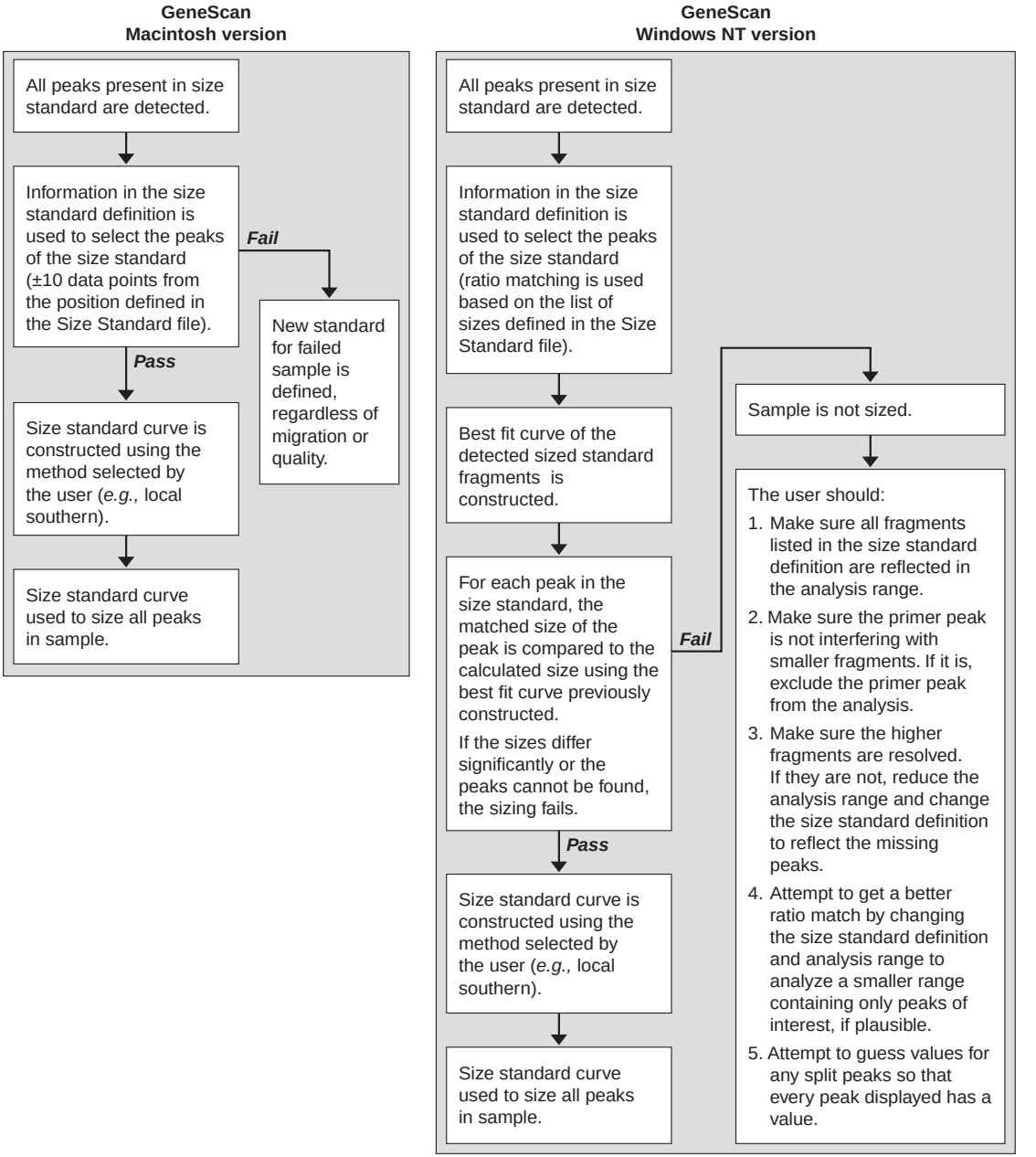


Figure 27 Peak sizing flowcharts

For Research Use Only. Not for use in diagnostic procedures.

Headquarters
850 Lincoln Centre Drive
Foster City, CA 94404 USA
Phone: +1 650.638.5800
Toll Free (In North America): +1 800.345.5224
Fax: +1 650.638.5884

Worldwide Sales and Support

Applied Biosystems vast distribution and service network, composed of highly trained support and applications personnel, reaches into 150 countries on six continents. For sales office locations and technical support, please call our local office or refer to our web site at www. applied biosystems.com.

www.appliedbiosystems.com

Applied Biosystems

Applera Corporation is committed to providing the world's leading technology and information for life scientists. Applera Corporation consists of the Applied Biosystems and Celera Genomics businesses.

Printed in USA, 06/2002
Part Number 4335617 Rev. A, Stock Number 106UB35-01

an **Applera** business