Agilent OpenLAB CDS

Data Analysis Reference Guide

Agilent Technologies

Notices

Manual Part Number

M8410-90031

Edition

08/2016

Printed in Germany

Agilent Technologies
Hewlett-Packard-Strasse 8
76337 Waldbronn

Software Revision

This guide is valid for revision 2.1 of Agilent OpenLAB CDS.

Warranty

The material contained in this document is provided "as is," and is subject to being changed, without notice, in future editions. Further, to the maximum extent permitted by applicable law, Agilent disclaims all warranties, either express or implied, with regard to this manual and any information contained herein, including but not limited to the implied warranties of merchantability and fitness for a particular purpose. Agilent shall not be liable for errors or for incidental or consequential damages in connection with the furnishing, use, or performance of this document or of any information contained herein. Should Agilent and the user have a separate written agreement with warranty terms covering the material in this document that conflict with these terms, the warranty terms in the separate agreement shall control.

Technology Licenses

The hardware and/or software described in this document are furnished under a license and may be used or copied only in accordance with the terms of such license.

Restricted Rights Legend

If software is for use in the performance of a U.S. Government prime contract or subcontract, Software is delivered and licensed as "Commercial computer software" as defined in DFAR 252.227-7014 (June 1995), or as a "commercial item" as defined in FAR 2.101(a) or as "Restricted computer software" as defined in FAR 52.227-19 (June 1987) or any equivalent agency regulation or contract clause. Use, duplication or disclosure of Software is subject to Agilent Technologies' standard commercial license terms, and non-DOD Departments and Agencies of the U.S. Government will receive no greater than Restricted Rights as defined in FAR 52.227-19(c)(1-2) (June 1987). U.S. Government users will receive no greater than Limited Rights as defined in FAR 52.227-14 (June 1987) or DFAR 252.227-7015 (b)(2) (November 1995), as applicable in any technical data.

Safety Notices

CAUTION

A **CAUTION** notice denotes a hazard. It calls attention to an operating procedure, practice, or the like that, if not correctly performed or adhered to, could result in damage to the product or loss of important data. Do not proceed beyond a **CAUTION** notice until the indicated conditions are fully understood and met.

WARNING

A WARNING notice denotes a hazard. It calls attention to an operating procedure, practice, or the like that, if not correctly performed or adhered to, could result in personal injury or death. Do not proceed beyond a WARNING notice until the indicated conditions are fully understood and met.

In This Guide...

This guide addresses the advanced users, system administrators and persons responsible for validating Agilent OpenLAB CDS. It contains reference information on the principles of calculations and data analysis algorithms.

Use this guide to verify system functionality against your user requirements specifications and to define and execute the system validation tasks defined in your validation plan. The following resources contain additional information.

- For context-specific task ("How To") information, references to the User Interface, and troubleshooting help: OpenLAB Help and Learning.
- For details on system installation and site preparation: The *Agilent OpenLAB Data CDS Requirements Guide*, *Agilent OpenLAB CDS Workstation Guide* or *Agilent OpenLAB CDS Client and AIC Guide*.

1 Signal Preparation

This chapter describes how the signal can be prepared, for example by blank subtraction, before it is integrated.

2 Integration with ChemStation Integrator

This chapter describes the concepts and integrator algorithms of the ChemStation integrator in OpenLAB CDS.

3 Integration with EZChrom Integrator

This chapter contains the description of EZChrom integration events.

4 Peak Identification

This chapter describes the concepts of peak identification.

5 Calibration

This chapter contains details of the calculations used in the calibration process.

6 Quantitation

This chapter describes how compounds are quantified, and explains the calculations used in quantitation.

7 UV Spectral Analysis

This chapter describes the concepts of the impurity check and the confirmation of compound identity based on UV spectral analysis.

8 Mass Spectrometry

This chapter describes the sample purity calculation based on mass spectrometry.

9 System Suitability

This chapter describes what OpenLAB CDS can do to evaluate the performance of both the analytical instrument and the analytical method.

Contents

1
Signal Preparation

This chapter describes how the signal can be prepared, for example by blank subtraction, before it is integrated.

Blank Subtraction

When analyzing a sample, the obtained signal may be caused by analytes as well as by dilution solvents, mobile phases, additives etc. Use the blank subtraction to receive a clean chromatogram with contribution of the analytes only.

Blank signals can origin from:

- a blank sample within a sequence
- a blank sample outside of the sequence (for example, a single run)

The new signal is calculated by subtracting the blank signal:

New signal = sample signal - blank signal

If a blank and a sample have different data rates, the data rate of the blank is adjusted. Data points are removed or created by spline interpolation.

If the run time of the sample is longer than the run time of the blank, the new signal will contain corrected and not corrected data points.

2

Integration with ChemStation Integrator

This chapter describes the concepts and integrator algorithms of the ChemStation integrator in OpenLAB CDS.

What is Integration?

Integration locates the peaks in a signal and calculates their size.

Integration is a necessary step for:

- identification
- calibration
- quantitation

What Does Integration Do?

When a signal is integrated, the software:

- identifies a start and an end time for each peak
- finds the apex of each peak; that is, the retention/migration time,
- constructs a baseline, and
- calculates the area, height, peak width, and symmetry for each peak.

This process is controlled by parameters called integration events.

Integrator Capabilities

The integrator algorithms include the following key capabilities:

- the ability to define individual integration event tables for each chromatographic signal if multiple signals or more than one detector are used
- graphical manual integration of chromatograms requiring human interpretation
- annotation of integration results
- integrator parameter definitions to set or modify the basic integrator settings for area rejection, height rejection, peak width, slope sensitivity, shoulder detection, baseline correction and front/tail tangent skim detection
- baseline control parameters, such as force baseline, hold baseline, baseline at all valleys, baseline at the next valley, fit baseline backwards from the end of the current peak, most likely baseline point from a time range
- area summation control
- negative peak recognition
- solvent peak definition detection
- integrator control commands defining retention time ranges for the integrator operation
- peak shoulder allocation through the use of second derivative calculations

The Integrator Algorithms

Overview

To integrate a chromatogram, the integrator ...

1. defines the initial baseline,
2. continuously tracks and updates the baseline,
3. identifies the start time for a peak,
4. finds the apex of each peak,
5. identifies the end time for the peak,
6. constructs a baseline, and
7. calculates the area, height, peak width, and symmetry for each peak.

This process is controlled by **integration events**. The most important events are initial slope sensitivity, peak width, shoulders mode, area reject, and height reject. The software allows you to set initial values for these and other events. The initial values take effect at the beginning of the chromatogram.

In most cases, the initial events will give good integration results for the entire chromatogram, but there may be times when you want more control over the progress of an integration.

The software allows you to control how an integration is performed by enabling you to program new integration events at appropriate times in the chromatogram.

Defining the Initial Baseline

Cardinal points

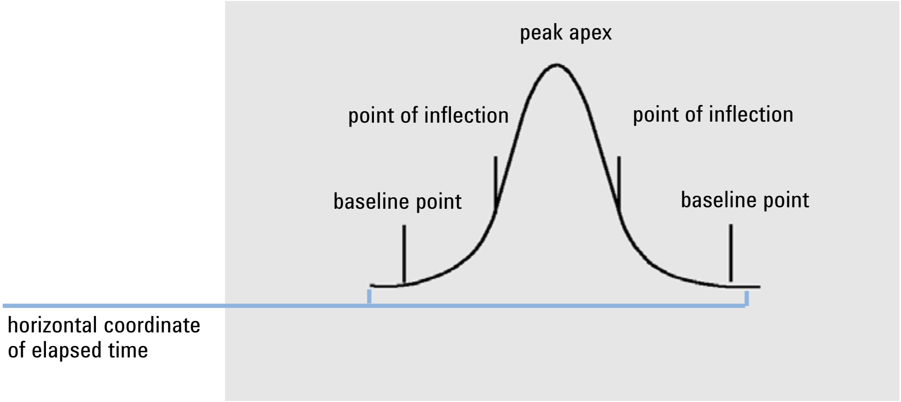


Figure 1 Cardinal points

Defining the initial baseline

Because baseline conditions vary according to the application and detector hardware, the integrator uses parameters from both the integration events and the data file to optimize the baseline.

Before the integrator can integrate peaks, it must establish a **baseline point**. At the beginning of the analysis, the integrator establishes an initial baseline level by taking the first data point as a tentative baseline point. It then attempts to redefine this initial baseline point based on the average of the input signal. If the integrator does not obtain a redefined initial baseline point, it retains the first data point as a potential initial baseline point.

Identifying the cardinal points of a peak

The integrator determines that a peak may be starting when potential baseline points lie outside the baseline envelope, and the baseline curvature exceeds a certain value, as determined by the integrator's slope sensitivity parameter. If this condition continues, the integrator recognizes that it is on the up-slope of a peak, and the peak is processed.

Start

1. Slope and curvature within limit: continue tracking the baseline.
2. Slope and curvature above limit: possibility of a peak.
3. Slope remains above limit: peak recognized, peak start point defined.
4. Curvature becomes negative: front inflection point defined.

Apex

1. Slope passes through zero and becomes negative: apex of peak, apex point defined.
2. Curvature becomes positive: rear inflection point defined.

End

1. Slope and curvature within limit: approaching end of the peak.
2. Slope and curvature remain within limit: end of peak defined.
3. The integrator returns to the baseline tracking mode.

Tracking the Baseline

The integrator samples the digital data at a rate determined by the initial peak width or by the calculated peak width, as the run progresses. It considers each data point as a potential baseline point.

The integrator determines a *baseline envelope* from the slope of the baseline, using a baseline-tracking algorithm in which the slope is determined by the first derivative and the curvature by the second derivative. The baseline envelope can be visualized as a cone, with its tip at the current data point. The upper and lower acceptance levels of the cone are:

- + upslope + curvature + baseline bias must be lower than the threshold level,
- - upslope - curvature + baseline bias must be more positive (i.e. less negative) than the threshold level.

As new data points are accepted, the cone moves forward until a break-out occurs.

To be accepted as a baseline point, a data point must satisfy the following conditions:

- it must lie within the defined baseline envelope,
- the curvature of the baseline at the data point (determined by the derivative filters), must be below a critical value, as determined by the current slope sensitivity setting.

The initial baseline point, established at the start of the analysis is then continuously reset, at a rate determined by the peak width, to the moving average of the data points that lie within the baseline envelope over a period determined by the peak width. The integrator tracks and periodically resets the baseline to compensate for drift, until a peak up-slope is detected.

Allocating the Baseline

The integrator allocates the chromatographic baseline during the analysis at a frequency determined by the peak width value. When the integrator has sampled a certain number of data points, it resets the baseline from the initial baseline point to the current baseline point. The integrator resumes tracking the baseline over the next set of data points and resets the baseline again. This process continues until the integrator identifies the start of a peak.

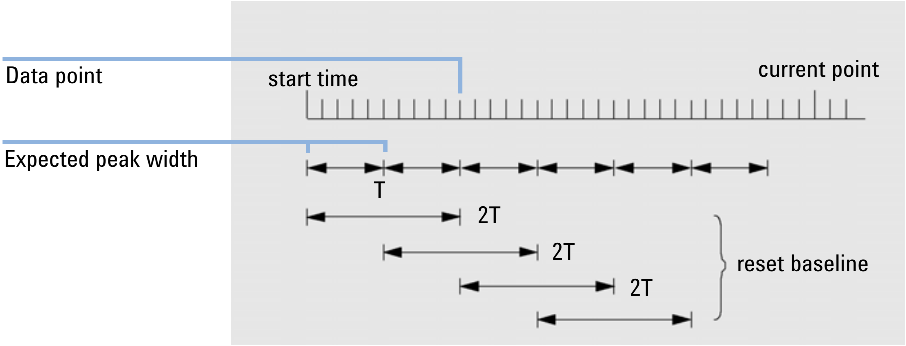


Figure 2 Baseline

At the start of the integration process the first data point is used. This baseline point is periodically reset as shown in the figure (see Figure 2 on page 16).

Areas are summed over a time T (expected peak width). This time can never be shorter than one data point. This continues as long as baseline condition exists. Slope and curvature are also taken. If both slope and curvature are less than the threshold, two summed areas are added together, and compared with the previous baseline. If the new value is less than the previous baseline, the new value immediately replaces the old one. If the new value is greater than the previous value, it is stored as a tentative new baseline value and is confirmed if one more value satisfies slope and curvature flatness criteria. This latter limitation is not in effect if negative peaks are allowed. During baseline, a check must also be made to examine fast rising solvents. They may be too fast for upslope detection. (By the time upslope is confirmed, solvent criterion may no longer be valid.) At first time through the first data point is baseline. It is replaced by the 2 T average if signal is on base. Baseline is then reset every T (see Figure 2 on page 16).

Definition of Terms

Solvent peak

The solvent peak, which is generally a very large peak of no analytical importance, is not normally integrated. However, when small peaks of analytical interest elute close to the solvent peak, for example, on the tail of the solvent peak, special integration conditions can be set up to calculate their areas corrected for the contribution of the solvent peak tail.

Shoulder (front, rear)

Shoulders occur when two peaks elute so close together that no valley exists between them, and they are unresolved. Shoulders may occur on the leading edge (front) of the peak, or on the trailing edge (rear) of the peak. When shoulders are detected, they may be integrated either by tangent skim or by drop-lines.

Slope

The slope of a peak, which denotes the change of concentration of the component against time, is used to determine the onset of a peak, the peak apex, and the end of the peak.

Principle of Operation

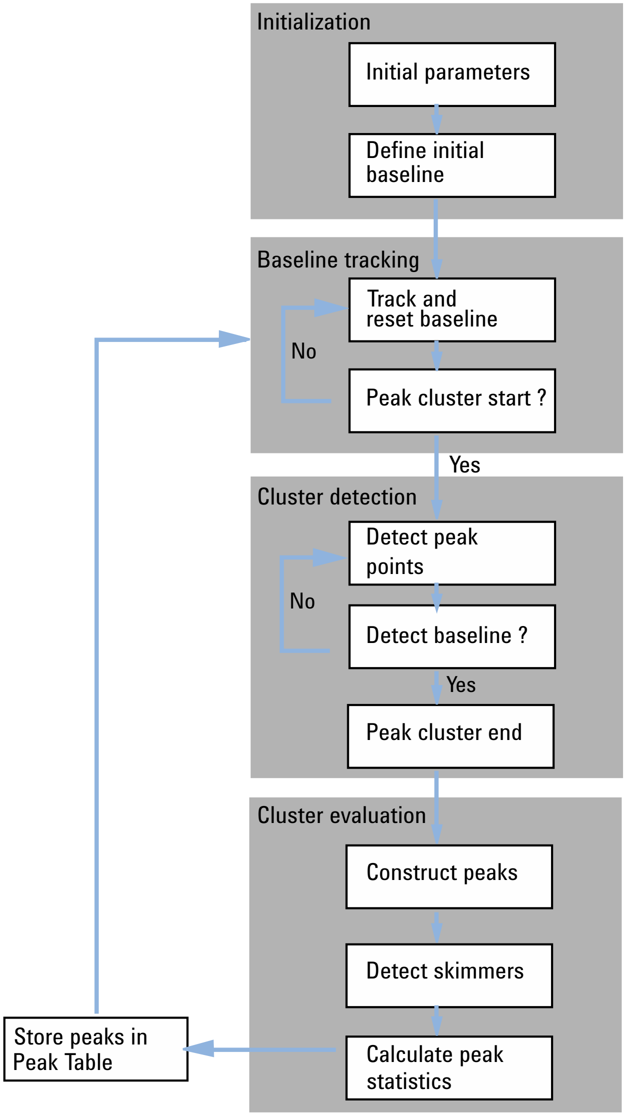


Figure 3 Integrator Flow Diagram

Peak Recognition

The integrator uses several tools to recognize and characterize a peak:

- peak width,
- peak recognition filters,
- bunching,
- peak recognition algorithm,
- peak apex algorithm, and
- non-Gaussian calculations (for example tailing, merged peaks).

Peak Width

During integration, the peak width is calculated from the adjusted peak area and height:

Width = adjusted area / adjusted height

or, if the inflection points are available, from the width between the inflection points.

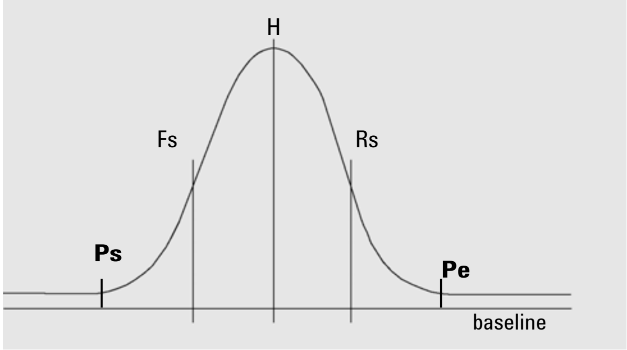


Figure 4 Peak width calculation

In the figure above, the total area, A, is the sum of the areas from peak start (Ps) to peak end (Pe), adjusted for the baseline. Fs is the front slope at the inflection point, Rs is the rear slope at the inflection point.

The peak width setting controls the ability of the integrator to distinguish peaks from baseline noise. To obtain good performance, the peak width must be set close to the width of the actual chromatographic peaks.

There are three ways the peak width is changed:

- before the integration process, you can specify the initial peak width,
- during the integration process, the integrator automatically updates the peak width as necessary to maintain a good match with the peak recognition filters,
- during the integration process, you can reset or modify the peak width using a time-programmed event.

Peak Recognition Filters

The integrator has three peak recognition filters that it can use to recognize peaks by detecting changes in the slope and curvature within a set of contiguous data points. These filters contain the first derivative (to measure slope) and the second derivative (to measure curvature) of the data points being examined by the integrator. The recognition filters are:

Filter 1	Slope (curvature) of two (three) contiguous data points
Filter 2	Slope of four contiguous data points and curvature of three non-contiguous data points
Filter 3	Slope of eight contiguous data points and curvature of three non-contiguous data points

The actual filter used is determined by the peak width setting. For example, at the start of an analysis, Filter 1 may be used. If the peak width increases during the analysis, the filter is changed first to Filter 2 and then to Filter 3. To obtain good performance from the recognition filters, the peak width must be set close to the width of the actual chromatographic peaks. During the run, the integrator updates the peak width as necessary to optimize the integration.

The integrator calculates the updated peak width in different ways, depending on the instrument technique.

For LC data, the default peak width calculation uses a composite calculation:

0.3 * (Right Inflection Point - Left Inflection Point) + 0.7 * Area / Height

For GC data, the default peak width calculation uses area/height. This calculation does not overestimate the width when peaks are merged above the half-height point.

In certain types of analysis, for example isothermal GC and isocratic LC analyses, peaks become significantly broader as the analysis progresses. To compensate for this, the integrator automatically updates the peak width as the peaks broaden during the analysis. It does this automatically unless the updating has been disabled with the fixed peak width timed event.

The peak width update is weighted in the following way:

0.75 * (Existing Peak Width) + 0.25 * (Width of Current Peak)

Bunching

Bunching is the means by which the integrator keeps broadening peaks within the effective range of the peak recognition filters to maintain good selectivity.

The integrator cannot continue indefinitely to increase the peak width for broadening peaks. Eventually, the peaks would become so broad that they could not be seen by the peak recognition filters. To overcome this limitation, the integrator bunches the data points together, effectively narrowing the peak while maintaining the same area.

When data is bunched, the data points are bunched as two raised to the bunching power, i.e. unbunched = 1x, bunched once = 2x, bunched twice = 4x etc.

Bunching is based on the data rate and the peak width. The integrator uses these parameters to set the bunching factor to give the appropriate number of data points (see Table 1 on page 22).

Bunching is performed in the powers of two based on the expected or experienced peak width. The bunching algorithm is summarized in Table 1 on page 22.

Table 1 Bunching criteria

Expected Peak Width	Filter(s) Used	Bunching Done
0 - 10 data points	First	None
8 - 16 data points	Second	None
12 - 24 data points	Third	None
16 - 32 data points	Second	Once
24 - 48 data points	Third	Once
32 - 96 data points	Third, second	Twice
64 - 192 data points	Third, second	Three times

The Peak Recognition Algorithm

The integrator identifies the start of the peak with a baseline point determined by the peak recognition algorithm. The peak recognition algorithm first compares the outputs of the peak recognition filters with the value of the initial slope sensitivity, to increase or decrease the up-slope accumulator. The integrator declares the point at which the value of the up-slope accumulator is ≥15 the point that indicates that a peak has begun.

Peak Start

In Table 2 on page 23 the expected peak width determines which filter’s slope and curvature values are compared with the Slope Sensitivity. For example, when the expected peak width is small, Filter 1 numbers are added to the up-slope accumulator. If the expected peak width increases, then the numbers for Filter 2 and, eventually, Filter 3 are used.

When the value of the up-slope accumulator is ≥15, the algorithm recognizes that a peak may be starting.

Table 2 Incremental Values to Upslope Accumulator

Derivative Filter 1 - 3 Outputs against Slope Sensitivity	Filter 1	Filter 2	Filter 3
Slope > Slope Sensitivity	+8	+5	+3
Curvature > Slope Sensitivity	+0	+2	+1
Slope < (-) Slope Sensitivity	-8	-5	-3
Slope < |Slope Sensitivity|	-4	-2	-1
Curvature < (-) Slope Sensitivity	-0	-2	-1

Peak End

In Table 3 on page 23 the expected peak width determines which filter's slope and curvature values are compared with the Slope Sensitivity. For example, when the expected peak width is small, Filter 1 numbers are added to the down-slope accumulator. If the expected peak width increases, then the numbers for Filter 2 and, eventually, Filter 3 are used.

When the value of the down-slope accumulator is ≥15, the algorithm recognizes that a peak may be ending.

Table 3 Incremental Values for Downslope Accumulator

Derivative Filter 1 - 3 Outputs against Slope Sensitivity	Filter 1	Filter 2	Filter 3
Slope < (-) Slope Sensitivity	+8	+5	+3
Curvature < (-) Slope Sensitivity	+0	+2	+1
Slope > Slope Sensitivity	-11	-7	-4
Slope > |Slope Sensitivity|	-28	-18	-11
Curvature > Slope Sensitivity	-0	-2	-1

The Peak Apex Algorithm

The peak apex is recognized as the highest point in the chromatogram by constructing a parabolic fit that passes through the highest data points.

Merged Peaks

Merged peaks occur when a new peak begins before the end of peak is found. The figure illustrates how the integrator deals with merged peaks.

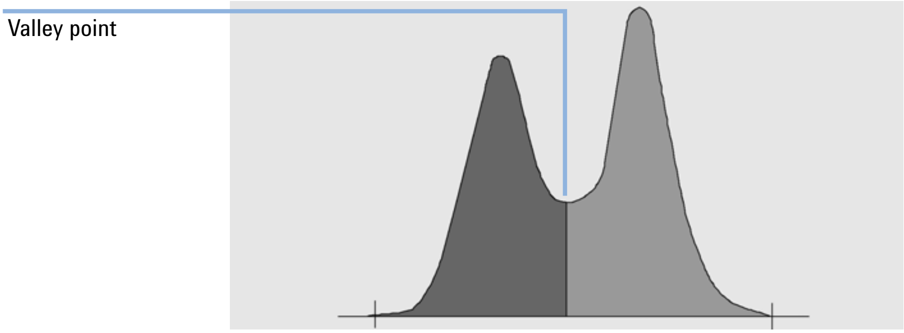


Figure 5 Merged Peaks

The integrator processes merged peaks in the following way:

1. it sums the area of the first peak until the valley point.
2. at the valley point, area summation for the first peak ends and summation for the second peak begins.
3. when the integrator locates the end of the second peak, the area summation stops. This process can be visualized as separating the merged peaks by dropping a perpendicular from the valley point between the two peaks.

Shoulders

Shoulders are unresolved peaks on the leading or trailing edge of a larger peak. When a shoulder is present, there is no true valley in the sense of negative slope followed by positive slope. A peak can have any number of front and/or rear shoulders.

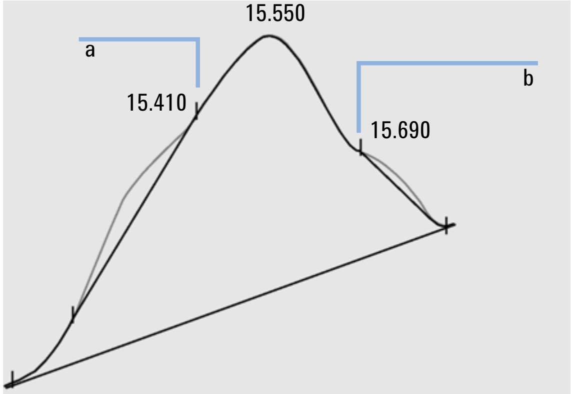


Figure 6 Peak Shoulders

Shoulders are detected from the curvature of the peak as given by the second derivative. When the curvature goes to zero, the integrator identifies a point of inflection, such as points a and b in Figure 6 on page 25.

- A potential front shoulder exists when a second inflection point is detected before the peak apex. If a shoulder is confirmed, the start of the shoulder point is set at the maximum positive curvature point before the point of inflection.
- A potential rear shoulder exists when a second inflection point is detected before the peak end or valley. If a shoulder is confirmed, the start of the shoulder point is set at the point of the first minimum of the slope after the peak apex.

Retention time is determined from the shoulder’s point of maximum negative curvature. With a programmed integration event, the integrator can also calculate shoulder areas as normal peaks with drop-lines at the shoulder peak points of inflection.

The area of the shoulder is subtracted from the main peak.

Peak shoulders can be treated as normal peaks by use of an integrator timed event.

Default Baseline Construction

After any peak cluster is complete, and the baseline is found, the integrator requests the baseline allocation algorithm to allocate the baseline using a pegs-and-thread technique. It uses trapezoidal area and proportional height corrections to normalize and maintain the lowest possible baseline. Inputs to the baseline allocation algorithm also include parameters from the method and data files that identify the detector and the application, which the integrator uses to optimize its calculations.

In the simplest case, the integrator constructs the baseline as a series of straight line segments between:

- the start of baseline,
- peakstart, valley, end points,
- the peak baseline

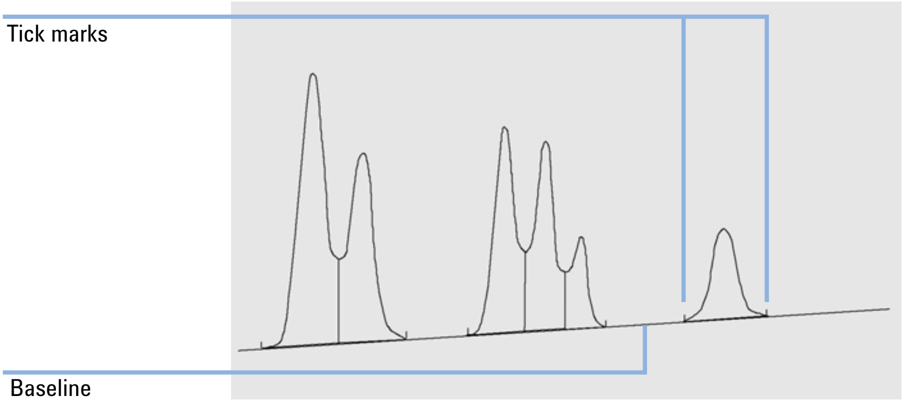


Figure 7 Default Baseline Construction

Baseline Codes

In the integration results of a report, each peak is assigned a two-, three- or four-character code that describes how the signal baseline was drawn.

Table 4 Four character code

First character	Second character	Third character	Fourth character
Baseline at start	Baseline at end	Error/peak flag	Peak type

Characters 1 and 2

The first character describes the baseline at the start of the peak and the second character describes the baseline at the end of the peak.

B	The peak started or stopped on the baseline.
P	The peak started or stopped while the baseline was penetrated.
V	The peak started or stopped with a valley drop-line.
H	The peak started or stopped on a forced horizontal baseline.
F	The peak started or stopped on a forced point.
M	The peak was manually integrated.
U	The peak was unassigned.

Additional flags may also be appended (in order of precedence):

Character 3

The third character describes an error or peak flag:

A	The integration was aborted.
D	The peak was distorted.
U	An under-range condition occurred.
O	An over-range condition occurred.
Blank space	The peak is a normal peak.

Character 4

The fourth character describes the peak type:

S	The peak is a solvent peak.
N	The peak is a negative peak.
+	The peak is an area summed peak.
T	Tangent-skimmed peak (standard skim).
X	Tangent-skimmed peak (old mode exponential skim).
E	Tangent-skimmed peak (new mode exponential skim).
m	Peak defined by manual baseline.
n	Negative peak defined by manual baseline.
t	Tangent-skimmed peak defined by manual baseline.
x	Tangent-skimmed peak (exponential skim) defined by manual baseline.
R	The peak is a recalculated peak.
f	Peak defined by a front shoulder tangent.
b	Peak defined by a rear shoulder tangent.
F	Peak defined by a front shoulder drop-line.
B	Peak defined by a rear shoulder drop-line.
U	The peak is unassigned.

Peak Area Measurement

The final step in peak integration is determining the final area of the peak.

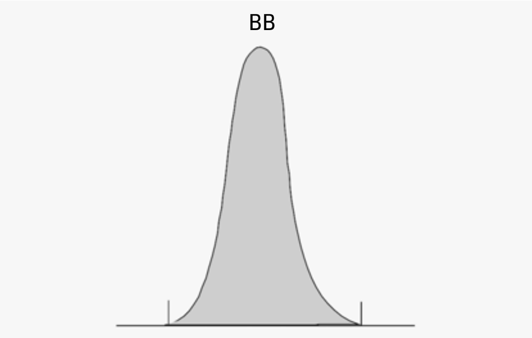


Figure 8 Area measurement for Baseline-to-Baseline Peaks

In the case of a simple, isolated peak, the peak area is determined by the accumulated area above the baseline between peak start and stop.

Determination of the area

The area that the integrator calculates during integration is determined as follows:

- for baseline-to-baseline (BB) peaks, the area above the baseline between the peak start and peak end, as in Figure 8 on page 29,
- for valley-to-valley (VV) peaks, the area above the baseline, segmented with vertical dropped lines from the valley points, as in Figure 9 on page 30,

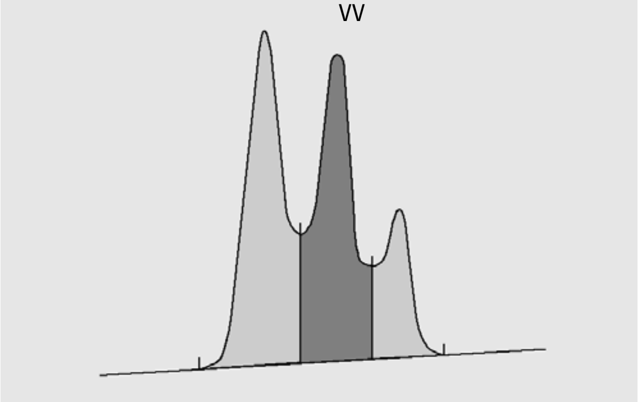


Figure 9 Area Measurement for Valley-to-Valley Peaks

- for tangent (T) peaks, the area above the reset baseline,
- for solvent (S) peaks, the area above the horizontal extension from the last-found baseline point and below the reset baseline given to tangent (T) peaks. A solvent peak may rise too slowly to be recognized, or there may be a group of peaks well into the run which you feel should be treated as a solvent with a set of riders. This usually involves a merged group of peaks where the first one is far larger than the rest. The simple drop-line treatment would exaggerate the later peaks because they are actually sitting on the tail of the first one. By forcing the first peak to be recognized as a solvent, the rest of the group is skimmed off the tail,
- negative peaks that occur below the baseline have a positive area, as shown in Figure 10 on page 30.

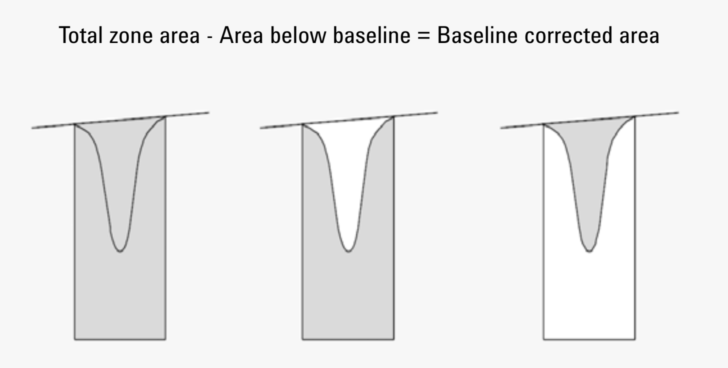


Figure 10 Area Measurement for Negative Peaks

Units and Conversion Factors

Externally, the data contains a set of data points; they can be either sampled data or integrated data. In the case of integrated data, each data point corresponds to an area, which is expressed as *Height* × *Time*. In the case of sampled data, each data point corresponds to a height.

Therefore, in the case of integrated data, height is a calculated entity, obtained by dividing area by the time elapsed since the preceding data point. In the case of sampled data, area is calculated by multiplying the data by the time elapsed since the preceding data point.

The integration calculation makes use of both entities. The units carried internally inside the integrator are: *counts* × *milliseconds* for area and *counts* for height. This is done to provide a common base for integer truncations when needed. The measurements of time, area and height are reported in real physical units, irrespective of how they are measured, calculated and stored in the software.

Baseline Allocation

Baseline Correction Modes

In OpenLAB CDS, several baseline correction modes are available. They are described in the following sections.

Baseline Correction Mode: Classical

A penetration occurs when the signal drops below the constructed baseline (point a in Figure 11 on page 32).

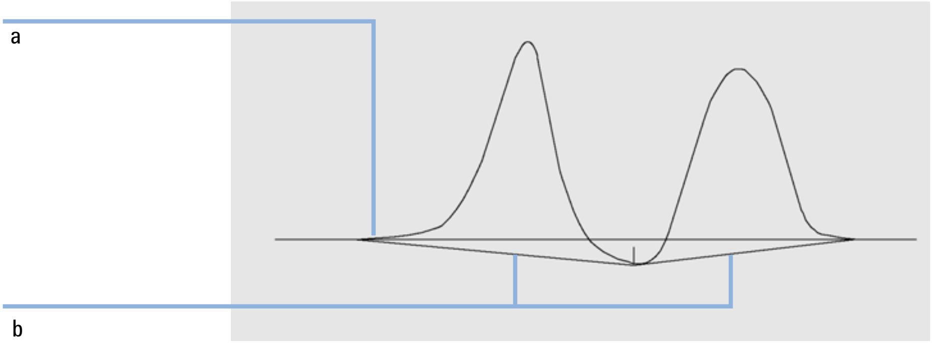


Figure 11 Baseline Penetration

If a baseline penetration occurs, that part of the baseline may be reconstructed, as shown by points b in Figure 11 on page 32. You can use the following correction modes to remove all baseline penetrations:

- **No penetration**
- **Advanced**

Baseline Correction Mode: No Penetration

When this option is selected, each peak cluster is searched for baseline penetrations. If penetrations are found, the start and/or end points of the peak are shifted until there are no penetrations left.

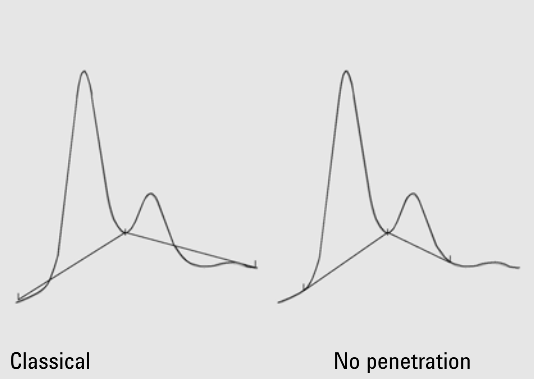


Figure 12 Baseline correction modes **Classical** and **No penetration**

NOTE The baseline correction mode **No penetration** is not available for solvent peaks, with their child peaks and shoulders.

Baseline Correction Mode: Advanced

In the advanced baseline correction mode, the integrator tries to optimize the start and end locations of the peaks, re-establishes the baseline for a cluster of peaks, and removes baseline penetrations (see Figure 12 on page 33). In many cases, advanced baseline correction gives a more stable baseline, which is less dependant on slope sensitivity.

Peak-to-Valley Ratio

The Peak to valley ratio is a measure of quality, indicating how well the peak is separated from other substance peaks. This user-specified parameter is a constituent of advanced baseline tracking mode. It is used to decide whether two peaks that do not show baseline separation are separated using a drop line or a valley baseline. The integrator calculates the ratio between the baseline-corrected height of the smaller peak and the baseline-corrected height of the valley. When the peak valley ratio is lower than the user-specified value, a drop-line is used; otherwise, a baseline is drawn from the baseline at the start of the first peak to the valley, and from the valley to the baseline at the end of the second peak (compare Figure 12 on page 33 with Figure 13 on page 34).

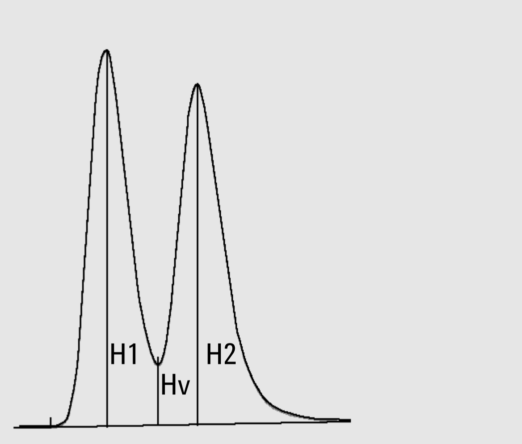


Figure 13 Peak Valley Ratio

The peak-to-valley ratio is calculated using the following equations:

$H1 \geq H2$, *Peak valley ratio* $= H2/Hv$

and

$H1 < H2$, *Peak valley ratio* $= H1/Hv$

Figure 14 on page 35 shows how the user-specified value of the peak valley ratio affects the baselines.

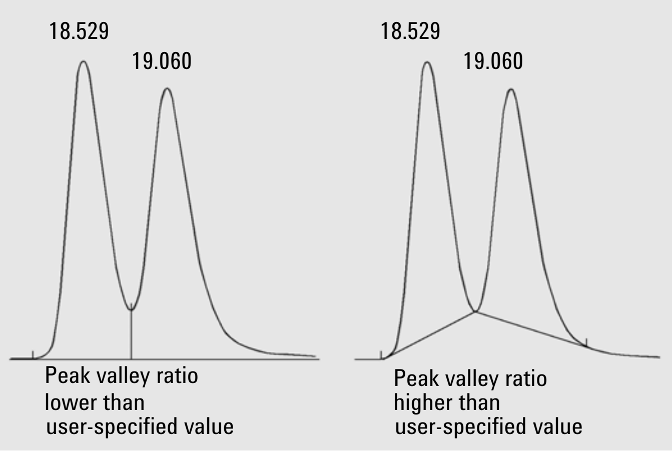


Figure 14 Effect of peak valley ratio on the baselines

Tangent Skimming

Tangent skimming is a form of baseline constructed for peaks found on the upslope or downslope of a peak. The prerequisite is that the two peaks are not baseline-separated.

The following figures illustrate the principle of tangent skimming:

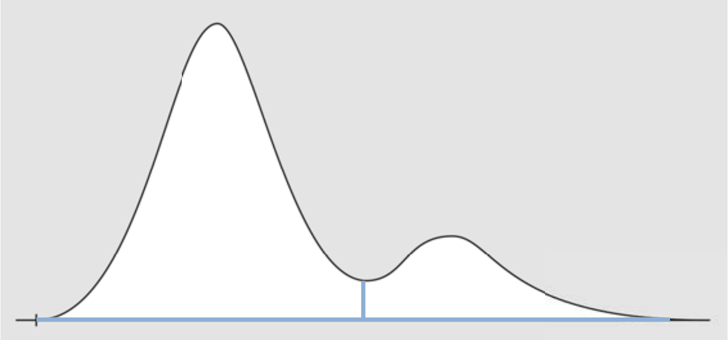


Figure 15 Peaks without skimming, separated by a drop line

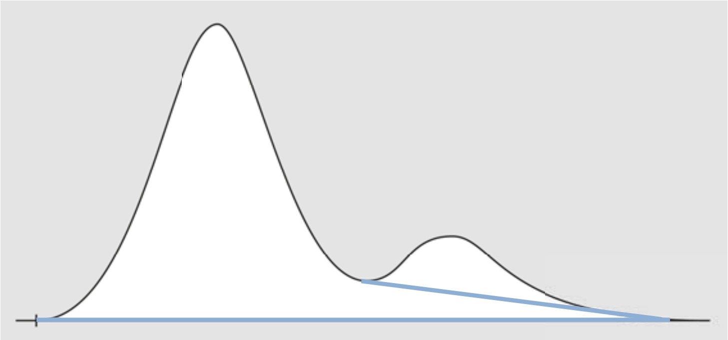


Figure 16 Tail skimming

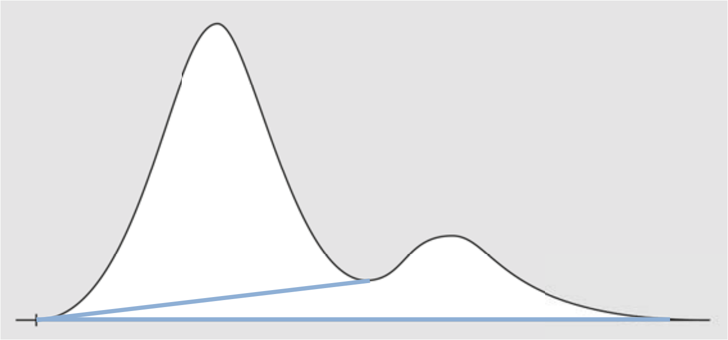


Figure 17 Front skimming

Skim Criteria

The following criteria determine wether a skim line is used to calculate the area of a child peak eluting on the leading or trailing edge of a parent peak:

- Skim height ratio (**Front skim height ratio** or **Tail skim height ratio**)
- **Skim valley ratio**

The *skim height ratio* is the ratio of the baseline-corrected height of the parent peak (Hp in the figure below) to the baseline-corrected height of the child peak (Hc). To have the child peak skimmed, use a value lower than this ratio. To disable exponential skimming throughout a run, you can set this parameter to a high value or to zero.

The *skim valley ratio* is the ratio of the height of the child peak above the baseline (Hc in the figure below) to the height of the valley above the baseline (Hv). To have the child peak skimmed, use a value greater than this ratio.

NOTE If one of these criteria is not met for a set of child peaks at the tail of the parent peak, all child peaks after the last child peak that met both criteria are not skimmed anymore but use a drop line

---

NOTE These criteria are not used if a timed event for an exponential is in effect, or if the parent peak is itself a child peak. The baseline code between parent peak and child peak must be of type **Valley** (see "Baseline Codes" on page 27).

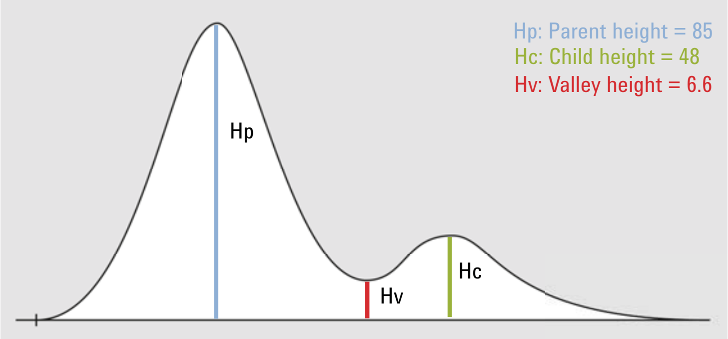


Figure 18 Example for calculating the skim criteria values

Skim height ratio = Hp / Hc

Skim valley ratio = Hc / Hv

where

Hp	Baseline-corrected height of parent peak
Hv	Height of valley above the baseline
Hc	Baseline-corrected height of child peak

Tail Skimming To use tail skimming, you would set the parameters as follows:

- Tail skim height ratio = 85 / 48 = 1.77

  In the integration events, use a value < 1.77.
- Skim valley ratio = 48 / 6.6 = 7.3

  In the integration events, use a value > 7.3.

Front Skimming With front skimming, the first peak is the child peak, and the second peak is the parent peak. Thus, to use front skimming, you would set the parameters as follows:

- Front skim height ratio = 48 / 85 = 0.56

  In the integration events, use a value < 0.56.
- Skim valley ratio = 85 / 6.6 = 12.9

  In the integration events, use a value > 12.9.

Tangent Skim Modes

When tangent skimming is enabled, four models are available to calculate suitable peak areas:

- Exponential curve
- New exponential skim
- Straight line skim
- Combined exponential and straight line calculations for the best fit (standard skims)

Exponential Curve

This skim model draws a curve using an exponential equation through the start and end of the child peak. The curve passes under each child peak that follows the parent peak; the area under the skim curve is subtracted from the child peaks and added to the parent peak.

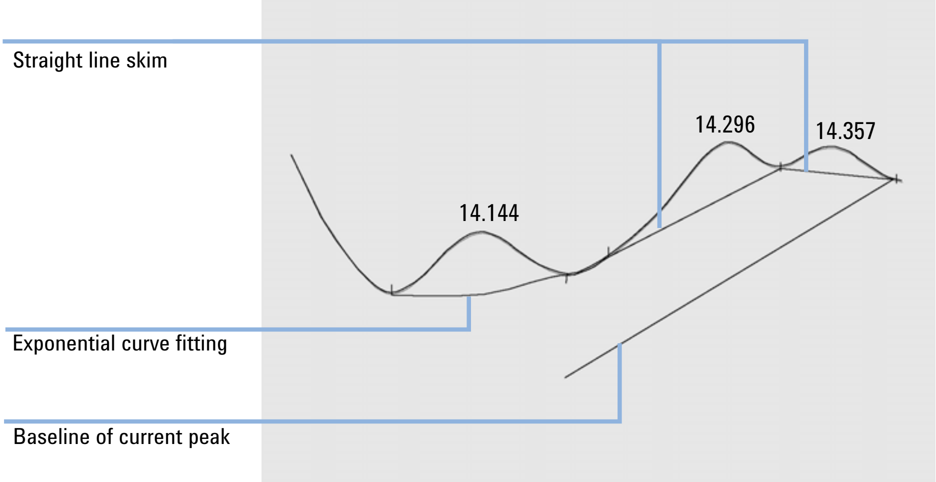


Figure 19 Exponential skim

New exponential curve

This skim model draws a curve using an exponential equation to approximate the leading or trailing edge of the parent peak. The curve passes under one or more peaks that follow the parent peak (child peaks). The area under the skim curve is subtracted from the child peaks and added to the main peak. More than one child peak can be skimmed using the same exponential model; all peaks after the first child peak are separated by drop lines, beginning at the end of the first child peak, and are dropped only to the skim curve.

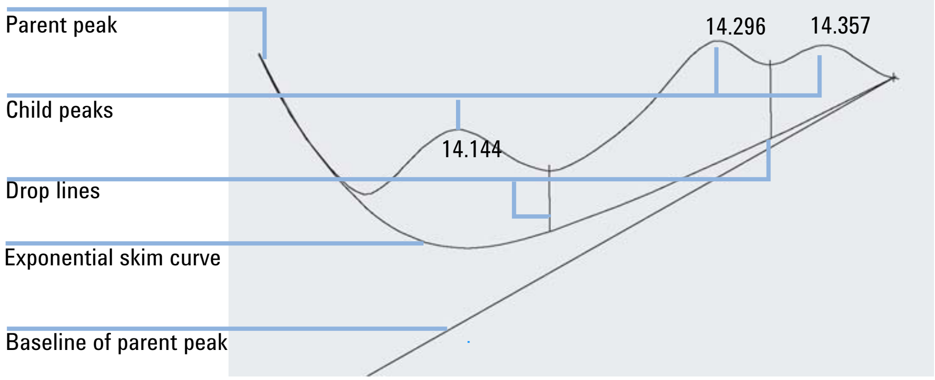


Figure 20 New exponential skim

Straight Line Skim

This skim model draws a straight line through the start and end of a child peak. The height of the start of the child peak is corrected for the parent peak slope. The area under the straight line is subtracted from the child peak and added to the parent peak.

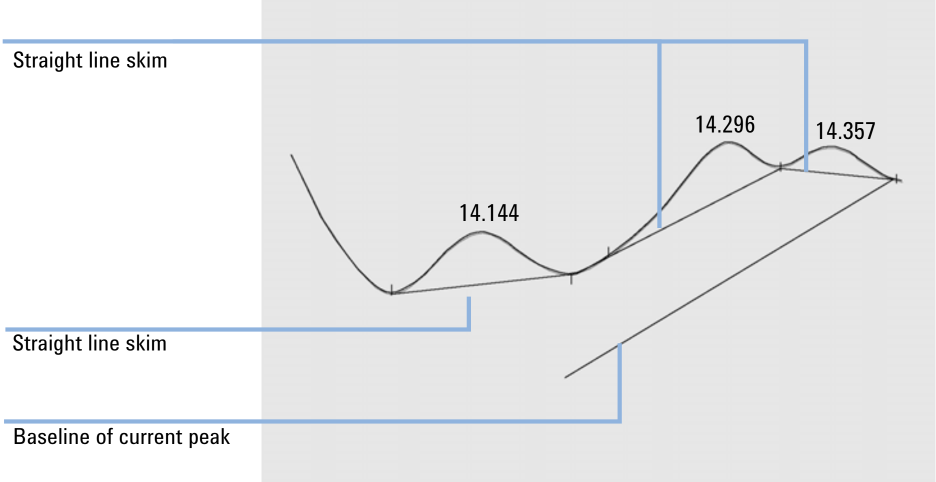


Figure 21 Straight line skim

Standard Skims

This default method is a combination of exponential and straight line calculations for the best fit.

The switch from an exponential to a linear calculation is performed in a way that eliminates abrupt discontinuities of heights or areas.

- When the signal is well above the baseline, the tail-fitting calculation is exponential.
- When the signal is within the baseline envelope, the tail fitting calculation is a straight line.

The combination calculations are reported as exponential or straight tangent skim.

Calculation of Exponential Curve for Skims

The following equation is used to calculate an exponential skim:

$$H_b(t_R) = H_0 * \exp(- B * (t_R - t_0)) + A * t_R + C$$

where

Hb	Height of the exponential skim at time t_R
H_0	Height (above baseline) of the start of the exponential skim
B	Decay factor of the exponential function
t_0	Time corresponding to the start of the exponential skim
t_R	Retention time
A	Slope of the baseline of the parent peak
C	Offset of the baseline of the parent peak

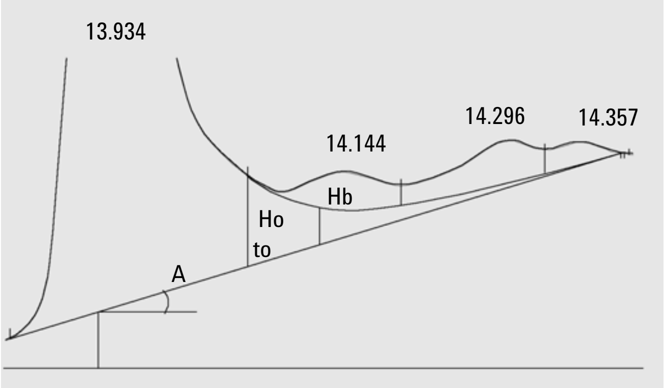


Figure 22 Values used to calculate an exponential skim

Integration Events

The available integration events are divided into the following groups:

- Initial integration events are those that apply from the start of integration. You can find them as default values in the **Standard** node of the **Integration Events** section in the processing method. These events cannot be deleted, but you may change the values.
- Timed events take place after the start of integration. Timed events may change the value of an initial event, or may switch on or off additional integration parameters. They can be added in the **Standard** node of the **Integration Events** section in the processing method.
- Integration events that always apply to all signals can be configured in the **Advanced** node of the **Integration Events** section.

Standard Integration Events: Initial Events

Slope sensitivity Sets the value of the signal slope that is used to identify the start and end points of a peak during integration.

You can set the values either specifically for a given signal or globally for all signals.

When the signal slope exceeds the **Slope Sensitivity** value, a peak start point is established; when the signal slope decreases below the **Slope Sensitivity** value, a peak end point is established.

Peak width Controls the selectivity of the integrator to distinguish peaks from baseline noise. You specify the peak width in units of time that correspond to the peak width at half-height of the first expected peak (excluding the solvent peak).

The integrator updates the peak width when necessary during the run to optimize the integration:

If the selected initial peak width is too low, noise may be interpreted as peaks. If broad and narrow peaks are mixed, you may decide to use runtime programmed events to adjust the peak width for certain peaks. Sometimes, peaks become significantly broader as the analysis progresses, for example in isothermal GC and isocratic LC analyses. To compensate for this, the integrator automatically updates the peak width as peaks broaden during an analysis unless disabled with a timed event.

The Peak Width update is weighted in the following way:

0,75 × (existing peak width) + 0,25 × (width of current peak)

Area reject Sets the area of the smallest peak of interest.

Any peaks that have areas less than the minimum area are not reported: The integrator rejects any peaks that are smaller than the **Area Reject** value after baseline correction. The **Area Reject** value must be greater than or equal to zero.

NOTE **Area reject** is ignored during manual integration.

Height reject Sets the height of the smallest peak of interest.

Any peaks that have heights less than this minimum height are not reported: The integrator rejects any peaks that are smaller than the **Height Reject** value after baseline correction.

NOTE **Height reject** is ignored during manual integration.

Shoulders mode Sets the initial method of detecting shoulders on peaks.

You can choose from:

Off	Shoulders are not detected.
Drop Baseline	Shoulders are integrated with a drop line.
Tangent Baseline	Shoulders are integrated with a tangent baseline.

This setting defines how the application handles peaks that are not baseline-separated. For more information on tangent skimming, see “Tangent Skimming” on page 35. If you use a tangent baseline, you can choose between different modes (see “Tangent Skim Modes” on page 38).

Choosing Peak Width

Choose the setting that provides just enough filtering to prevent noise being interpreted as peaks without distorting the information in the signal.

- To choose a suitable initial peak width for a single peak of interest, use the peak's time width at the base as a reference.
- To choose a suitable initial peak width when there are multiple peaks of interest, set the initial peak width to a value equal to or less than the narrowest peak width to obtain optimal peak selectivity.

Height Reject and Peak Width

Both **peak width** and **height reject** are very important in the integration process. You can achieve different results by changing these values.

- Increase both the height reject and peak width where relatively dominant components must be detected and quantified in a high-noise environment. An increased peak width improves the filtering of noise and an increased height reject ensures that random noise is ignored.
- Decrease height reject and peak width to detect and quantify trace components, those whose heights approach that of the noise itself. Decreasing peak width decreases signal filtering, while decreasing height reject ensures that small peaks are not rejected because they have insufficient height.
- When an analysis contains peaks with varying peak widths, set peak width for the narrower peaks and reduce height reject to ensure that the broad peaks are not ignored because of their reduced height.

Tuning Integration

It is often useful to change the values for the slope sensitivity, peak width, height reject, and area reject to customize integration. The figure below shows how these parameters affect the integration of five peaks in a signal.

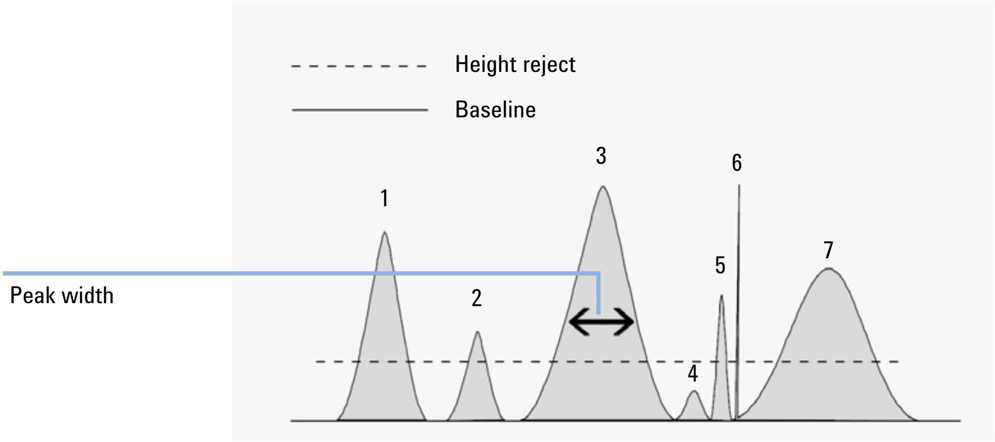


Figure 23 Using Initial Events

A peak is integrated only when all of the four integration parameters are satisfied. Using the peak width for peak 3, the area reject and slope sensitivity shown, only peaks 1, 3, and 7 are integrated.

Peak 1	is integrated as all four integration parameters are satisfied.
Peak 2	is rejected because the area is below the set area reject value.
Peak 3	is integrated as all four integration parameters are satisfied.
Peak 4	is not integrated because the peak height is below the Height Reject.
Peak 5	is rejected because the area is below the set area reject value.
Peak 6	is not integrated; filtering and bunching make the peak invisible.
Peak 7	is integrated.

Table 5 Height and Area Reject Values

Integration Parameter	Peak 1	Peak 2	Peak 3	Peak 4	Peak 5	Peak 7
Height reject	Above	Above	Above	Below	Above	Above
Area reject	Above	Below	Above	Below	Below	Above
Peak integrated	Yes	No	Yes	No	No	Yes

Standard Integration Events: Timed Events

OpenLAB CDS offers a set of timed events that allow a choice between the integrator modes of internal algorithm baseline definition and the user's definition. These timed events can be used to customize signal baseline construction when default construction is not appropriate. E.g. the user can create a new area sum event type (see **Area sum slice**), which does not alter the results of the default AreaSum. These events can be useful for summing final peak areas and for correcting short- and long-term baseline aberrations.

You can set the values either specifically for a given signal or globally for all signals.

Area reject See Initial Events ("Standard Integration Events: Initial Events" on page 43).

Area sum Sets points (**On**/**Off**) between which the integrator sums the areas.

The retention/migration time of a peak created with area summing is the average of the start and end times. If an **Area sum on** event occurs after the beginning of a peak but before the apex, the entire peak is included in the sum. If it occurs after the peak apex, but before the end of the peak, the peak is truncated and the area sum begins immediately.

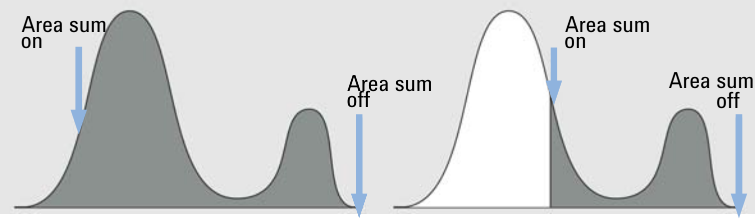


Figure 24 **Area sum on** event after peak apex, but before end of peak

If an **Area sum off** event occurs after the beginning of a peak but before the apex, the area sum ends immediately. The point on the signal where this occurs becomes a Valley Point. If the **Area sum off** event occurs after the apex, the event is postponed until the end of the peak.

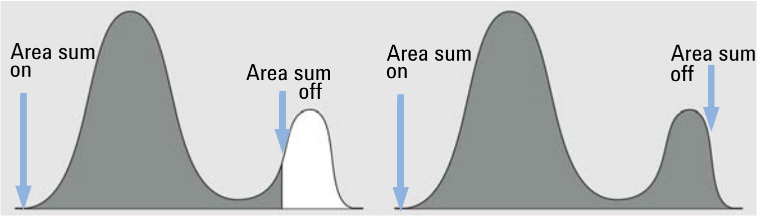


Figure 25 **Area sum off** event after beginning of peak, but before apex

Area sum slice This event allows you to define consecutive area sum intervals without any loss in area or time intervals.

This event is similar to the **Area Sum** event. However, with this event you can define contiguous area sum intervals without any loss in time intervals and integrated peak areas. A peak is split at the point where you set this event; area summing starts and ends exactly where the **Area Sum Slice** intervals are specified.

The retention time of the area sum slice peak is the middle of the slice time interval. The retention time does not change with identification or recalibration. It may only be shifted slightly, as the integrator only starts taking data points with the area sum slice start event, and ends with the area sum slice end event. Thus, the retention time may at most vary by the time between two data points.

Use the **Start** parameter to define the starting times for each area sum slice. The next start time is used as the end time for the preceding time-slice, so you can use several start events after each other.

The **Start-negA.** parameter defines the start of integration of a time-slice where any negative area (below the set baseline) is subtracted from the area of the time-slice.

The **End** parameter defines the end of the last time-slice. The area of the time-slice is calculated ignoring any area below the set baseline. If no other area sum slice events follow, the integrator resumes its regular peak detection again.

Within the range from a **Start** event to the next **End** event, the baseline is always one straight line with no changes in direction in between. Only after the end point (at least 0.001 min later) long term baseline changes can be applied again by using the events **Set Baseline from Range**, **Set Low Baseline from Range** or **Use Baseline from Range**.

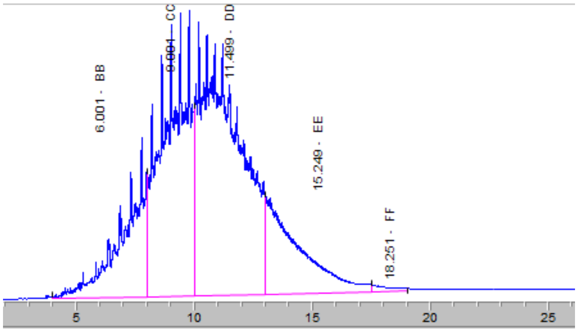


Figure 26 Example: Area Sum Slice

The figure above shows an example with the following timed events:

Table 6 Baseline construction

Time	Event	Parameter
4 min	Set BL from Range	+-2 min
22 min	Set BL from Range	+-4 min

Table 7 Area sum slices

Time	Event	Parameter
4 min	Area Sum Slice	Start
8 min	Area Sum Slice	Start
10 min	Area Sum Slice	Start
13 min	Area Sum Slice	Start
17.5 min	Area Sum Slice	Start
19 min	Area Sum Slice	End

Auto peak width Turns on the automatic update of the peak width for the next peaks. It will resume with whatever the peak width is at that time and resume peak width tracking based on the previous found peak widths.

Baseline at valleys Sets points (**On**/**Off**) between which the integrator resets the baseline at every valley between peaks.

The repeated resetting of the baseline can cut off corners of peaks. Such corners become negative area, they reduce the total measured area of the peaks.

This function is useful when peaks are riding on the back of a broad, low peak and you want the baseline to be reset to all the valley points.

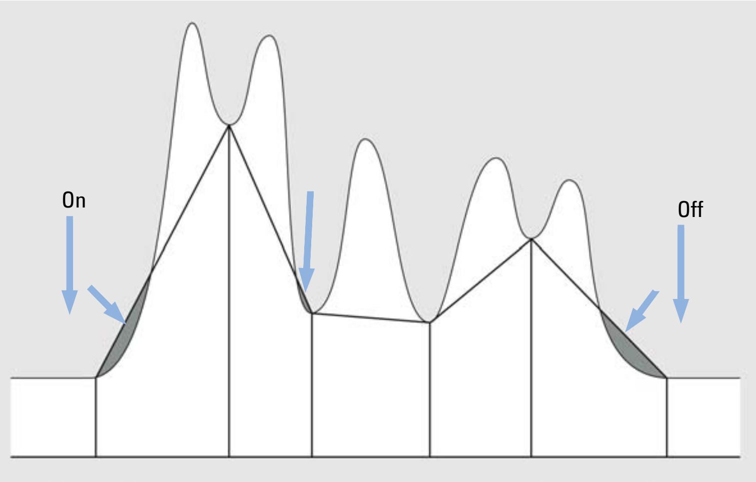


Figure 27 **Baseline at valleys** event

Baseline backwards Sets a point at which the standard integrator extends the baseline, horizontally backward from the declared baseline point to this point.

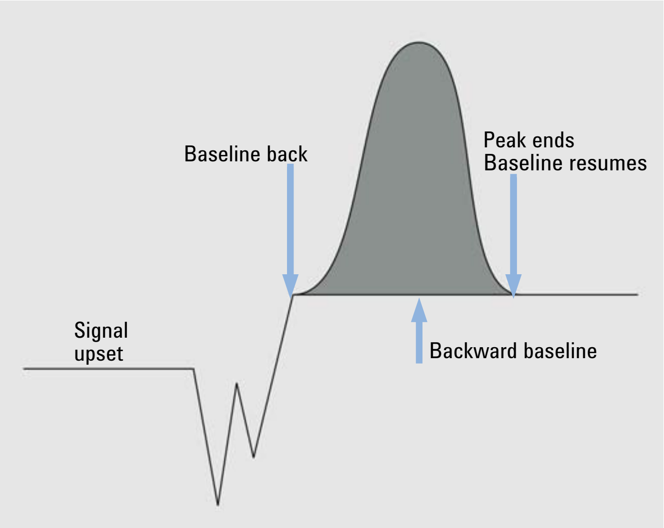


Figure 28 **Baseline backwards** event

Baseline hold A horizontal baseline is drawn at the height of the established baseline from where the baseline hold event is switched on until where the baseline hold event is switched off.

Baseline next valley Sets a point at which the integrator resets the baseline at the next valley between peaks, and then cancels this function automatically.

This function is useful in groups of merged peaks, which you assume are riding on the back or are in separate clusters close together. The function is ignored during area summing.

Baseline now Sets a point (time) at which the integrator resets the baseline to the current height of the data point, if the signal is on a peak.

If the signal is on the baseline, the function is ignored and the detected baseline is used.

Detect shoulders Sets points (**On**/**Off**) between which the integrator starts and stops detecting shoulders.

Shoulders are detected according to the specified **Shoulders Mode**. See "Standard Integration Events: Initial Events" on page 43.

Fixed peak width Sets the peak width and disables the automatic update of the peak width for the next peaks. To obtain good performance, set the peak width close to the width at half-height of the actual peaks.

Height reject See Initial Events ("Standard Integration Events: Initial Events" on page 43).

Integration Sets points (**On**/**Off**) between which the integrator stops and starts integrating.

Peaks between the times where the integrator is turned off and on are ignored.

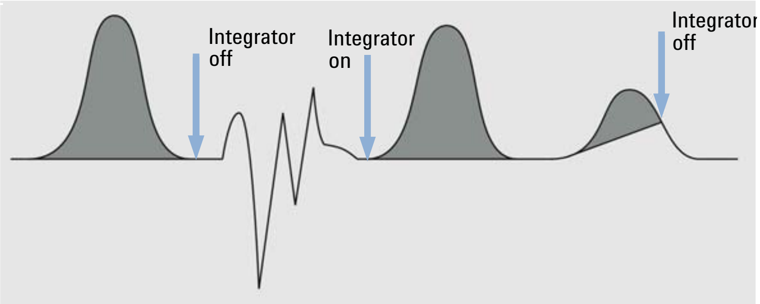


Figure 29 **Integration** event

The baseline is drawn from the last declared point including any resets for penetration. All other integrator functions together with set changes of peak width, threshold and area reject are ignored when the integrator is turned off. At the **On** and **Off** points, the baseline point is re-established.

When the integrator is set to restart, a new baseline point is reset at the current signal level.

This function is useful for ignoring parts of the chromatogram/electropherogram or to eliminate baseline disturbances.

Maximum area Sets the area of the largest peak of interest.

Any peaks that have areas greater than the maximum area are not reported: The integrator rejects any peaks that are greater than the maximum area value after baseline correction.

You can use this event, for example, to exclude the solvent peak of a GC chromatogram from the integration results, but include its rider peaks.

Maximum height Sets the height of the largest peak of interest.

Any peaks that have heights greater than the maximum height are not reported: The integrator rejects any peaks that are higher than the maximum height value after baseline correction.

You can use this event, for example, to exclude the solvent peak of a GC chromatogram from the integration results, but include its rider peaks.

Negative peak Sets points (**On**/**Off**) between which the integrator recognizes negative peaks.

When negative peaks are recognized, the integrator no longer automatically resets the baseline after penetration. From now on any penetration of the baseline will be integrated using the established baseline as zero. Areas are constructed relative to this baseline and are given an absolute value.

The negative peak function can only be used with confidence when the baseline drift is small compared with peak size, since the baseline is constructed from the declared baseline point at the start of the peak cluster up to the established baseline at the end of the peak.

NOTE

Area Summation is automatically deactivated if the **Negative Peaks On** event is activated.

Tangent skimming is also deactivated during negative peak detection; such peaks are separated by dropline.

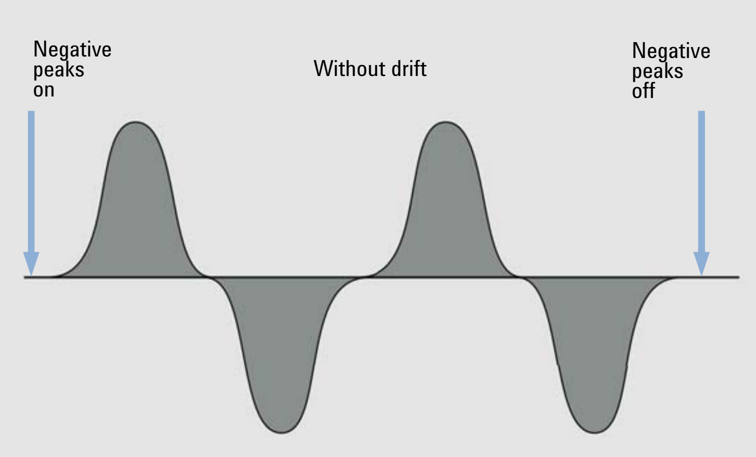


Figure 30 **Negative peak** event

Set baseline from range

Uses a range of data points to calculate a statistically meaningful baseline point at the midpoint of a time-range.

The value that you provide is the time interval around a specified point in time. It defines the range to be used to determine the baseline point. See "Baseline Correction Modes" on page 32 for details of the statistical calculations of the baseline.

If you set the value =0, the nearest chromatogram data point is used as a baseline point; no statistics is done at all. If you set a negative value, the setting does the same as **Use baseline from range**=**Clear**: It stops the usage of the statistical baseline algorithm.

You can specify any area in the chromatogram for the baseline calculation. Ideally, it should be an area that is free from chemical background and contains only noise.

If you specify two **Set Baseline from Range** points (for example at the beginning and end of a chromatogram), the baseline between them is connected with a straight line.

Set low baseline from range

Similar to **Set Baseline from Range**, but uses the lowest likely baseline point which allows 30 % more noise data points to be above it. Thus, baseline penetration is minimized.

Use **Set Low Baseline from Range** instead of **Set Baseline from Range** when the area of the chromatogram used for the calculation contains excessive chemical noise or electronic noise spikes.

Set Low Baseline from Range is calculated by a subtraction of one sigma (noise standard deviation) from the **Set Baseline from Range** y-value.

Shoulders mode

See Initial Events ("Standard Integration Events: Initial Events" on page 43).

Slope sensitivity

See Initial Events ("Standard Integration Events: Initial Events" on page 43).

Solvent peak

Peaks above a specific slope in units of mV/s are detected as solvent peaks that lie outside the range of the analog-to-digital conversion.

The trailing peaks are automatically tangent-skimmed; you do not need to switch on the tangent skim event.

If solvent peak detection is off, droplines are drawn from the trailing peak instead of tangents.

Split peak Specifies a point at which to split a peak with a dropline.

NOTE You cannot use **Split Peak** while **Area Sum** is switched on. To split a peak while **Area Sum** is switched on, use the corresponding manual integration event.

You cannot split skimmed peaks using the **Split Peak** event.

Tail tangent skim Specifies where to start or end tangent skimming.

On

Sets a point at which the integrator sets a tangent skim on the trailing edge of the next peak. All peaks above the tangent are integrated to the reset baseline. The tangent is drawn from the valley before the small peak to the point after it where the detector signal gradient is equal to the tangent gradient. The tangent skim event time can be entered any time during the peak. Designates peak also as a solvent peak.

Off

Ends tangent skimming after current peak is completed or if in the designated interval no peaks are found (and a solvent will not inadvertently be designated in the next cluster).

Tangent skim mode The following models are available to calculate suitable peak areas:

- Exponential(Figure 19 on page 39)
- New exponential(Figure 20 on page 40)
- Standard(Figure on page 41)
- Straight(Figure 21 on page 41)

Unassigned peaks

With some baseline constructions, there are small areas that are above the baseline and below the signal, but are not part of any recognized peaks. Normally, such areas are neither measured nor reported. If unassigned peaks is turned on, these areas are measured and reported as unassigned peaks. The retention/migration time for such an area is the midpoint between the start and end of the area.

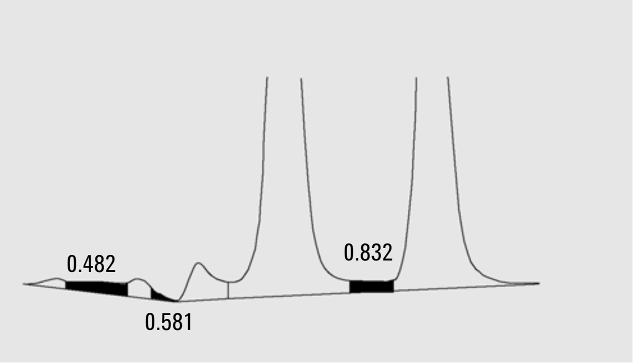


Figure 31 Unassigned Peaks

Use baseline from range

Allows to project a baseline value to a later or earlier time to minimize baseline penetrations.

If the **Set Baseline from Range** or **Set Low Baseline from Range** value is calculated in an area with no chromatographic peaks, it can be advantageous to project the calculated baseline to the time immediately before the first peak of interest elutes (or to the time immediately after the last peak of interest has eluted). **Use Baseline from Range** allows you to make up to three such projections in either direction.

This event can be advantageous to use when you have constructed an upslope or downslope baseline, since otherwise the straight baseline might cut through the chromatogram curve unintentionally. The parameter tells the integrator from which of the baseline ranges to pick the baseline point and project the baseline to the baseline point at the given time interval.

You can use the following parameters:

- **Clear**: Clear the new baseline behavior and return to the traditional algorithm from this point.
- **Left**: Use the baseline value from the baseline range nearest to the left of this point in time.
- **Right**: Use the baseline value from the baseline range nearest to the right of this point in time.
- **Range 1—Range 9**: Use the baseline value from the given baseline range. Baseline ranges are counted from the beginning of the chromatogram.

See also the example under **Area Sum Slice** (Figure 26 on page 49).

Advanced Integration Events

The advanced integration events are provided for all signals.

Tangent skim mode Define the type of baseline construction for peaks found on the upslope or downslope of a peak. See “Tangent Skim Modes” on page 38.

Exponential	Draws an exponential curve through the height-corrected start and end of each child peak.
New Exponential	Draws an exponential curve to approximate the trailing edge of the parent peak.
Standard	Combines exponential and straight line calculations for best fit.
Straight	Draws a straight line through the height-corrected start and end of each child peak.

Tail skim height ratio

Together with the **Skim valley ratio**, sets the conditions for tangent skimming a small peak on the tail of a solvent or other large peak. See “Skim Criteria” on page 36.

It is the ratio of the height of the baseline-corrected parent peak (Hp) to the height of the baseline-corrected child peak (Hc). Ratios higher than the specified value will enable skimming.

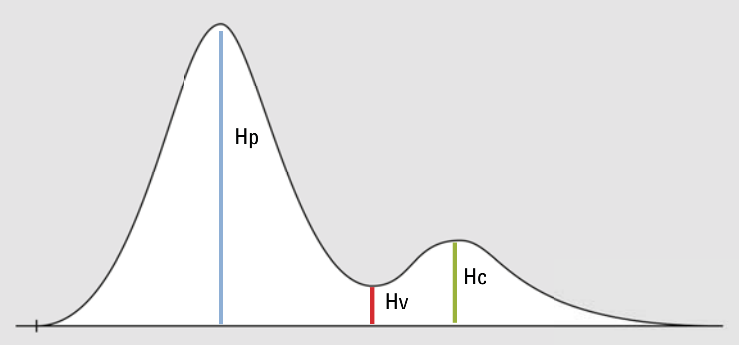


Figure 32 Example: Peak with tail skimming

Front skim height ratio

Together with the **Skim Valley Ratio**, sets the conditions for tangent skimming a small peak on the front of a solvent or other large peak. See “Skim Criteria” on page 36.

It is the ratio of the height of the baseline-corrected parent peak (Hp) to the height of the baseline-corrected child peak (Hc). Ratios higher than the specified value will enable skimming.

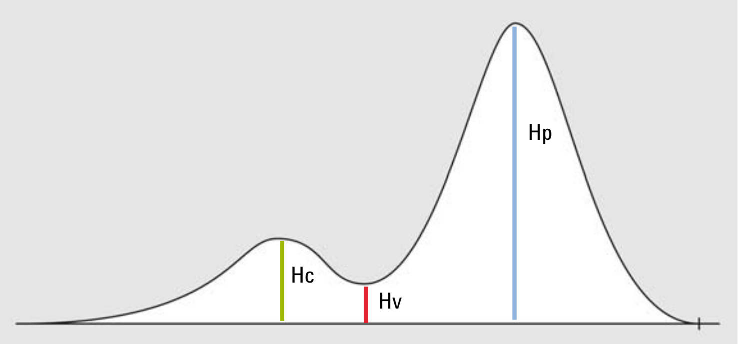


Figure 33 Example: Peak with front skimming

Skim valley ratio

Together with the **Tail Skim Height Ratio** or **Front Skim Height Ratio**, sets the conditions for tangent skimming a small peak on the tail or front of a solvent or other large peak. See “Skim Criteria” on page 36.

It is the ratio of the height of the baseline-corrected child peak (Hc) to the height of the baseline-corrected valley (Hv). Ratios lower than the specified value will enable skimming.

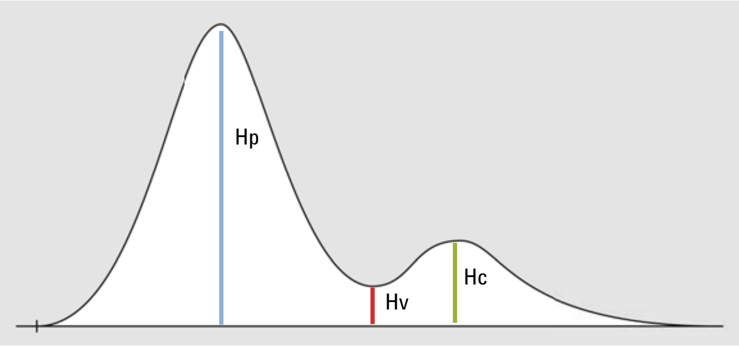


Figure 34 Example: Peak with tail skimming

Baseline correction mode

Sets the type of baseline correction. See “Baseline Correction Modes” on page 32.

You can choose between the following parameters:

Classical	Accepts baseline penetrations.
No penetrations	Removes baseline penetrations by reconstructing the baseline.
Advanced	The integrator tries to optimize the start and end locations of the peaks, re-establishes the baseline for a cluster of peaks and removes baseline penetrations.

Peak-to-Valley ratio Used to decide whether two peaks that do not show baseline separation are separated using a drop line or a valley baseline, it is the ratio of the baseline-corrected height of the smaller peak to the baseline-corrected height of the valley. See "Peak-to-Valley Ratio" on page 34.

When the peak to valley ratio is lower than the specified value, a drop line is used (A); otherwise, a baseline is drawn from the baseline at the start of the first peak to the valley, and from the valley to the baseline at the end of the second peak (B).

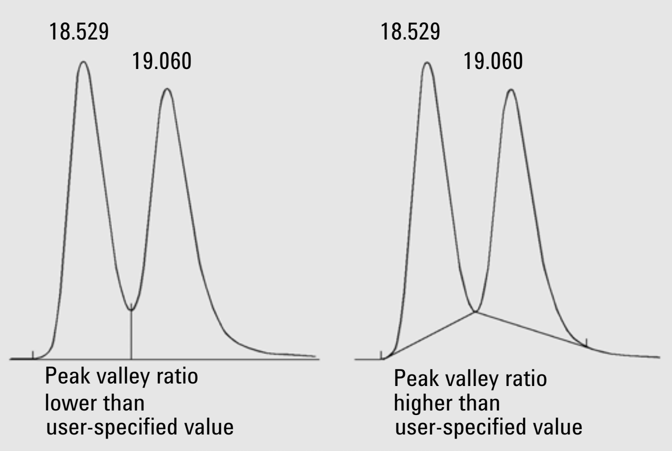


Figure 35 Effect of peak valley ratio on the baselines

3

Integration with EZChrom Integrator

This chapter contains the description of EZChrom integration events.

Integration Events

Processing method: Integration Events EZChrom

This section of the method shows all parameters from the EZChrom integrator. It is only visible if you have selected **EZChrom integrator** in the general method properties.

You can set the values either specifically for a given signal or globally for all signals. To add a timed event, right-click in the parameters table.

There are different types of integration events: For some of them, you can define a time range with start and stop time during which a parameter is active. For others, you can define a specific value to be used from a start time or during a time range. The columns **Time Stop [min]** and **Value** are enabled or grayed out, depending on the type of event.

Width

Used to distinguish true peaks from noise. The system uses the default value of width = 0.2 min.

The **Width** event is used to calculate a value for bunching, or smoothing, the data points before the integration algorithm is applied. Integration works best when there are 20 points across a peak. If a peak is over sampled (i.e. the sampling frequency was too high), the **Width** parameter will be used to average the data such that the integration algorithm sees only 20 points across the peak.

A **Width** event will be applied to a given peak as long as it occurs before or on the apex of the peak.

The **Width** parameter is only used to correct for over-sampling. It cannot correct for data that was under-sampled (i.e. sampling frequency too low causing fewer than 20 points acquired across the narrowest peak).

The diagrams below show examples of how incorrect values can effect the peak baseline.

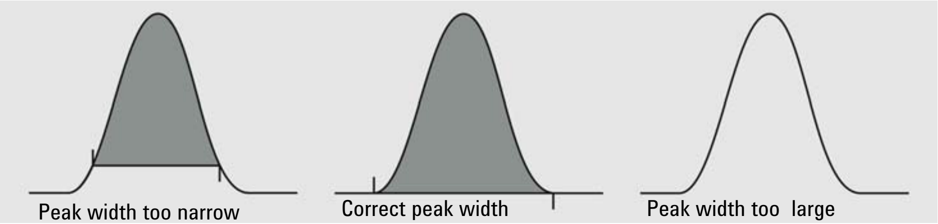


Figure 36 Width

NOTE In most circumstances, an initial Width value based on the narrowest peak in the chromatogram will be adequate for proper integration of all peaks. However, a new Width timed event should be entered every time a peak width doubles.

NOTE Extreme values of both Width and Threshold (too large or too small) can result in peaks not being detected.

Threshold This parameter is the first derivative, used to allow the integration algorithm to distinguish the start and stop of peaks from baseline noise and drift. The **Threshold** value is based on the highest first derivative value determined in a section of the chromatogram.

The diagrams below show examples of how incorrect values can effect the peak baseline.

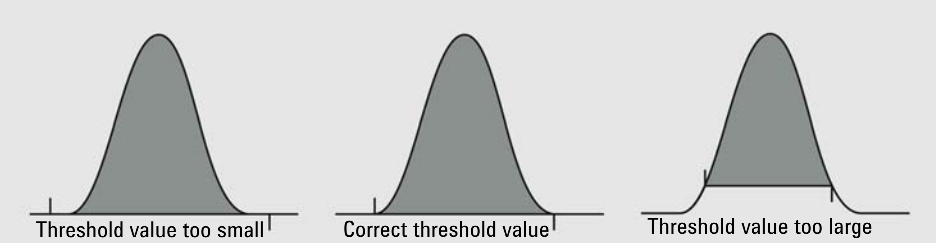


Figure 37 Threshold

NOTE Extreme values of both Width and Threshold (too large or too small) can result in peaks not being detected.

Shoulder sensitivity This parameter is used to enable the detection of shoulders on larger peaks. A larger value will decrease shoulder sensitivity while smaller values increase sensitivity to shoulder peaks. The **Shoulder Sensitivity** value is based on the highest second derivative value determined in a section of the chromatogram.

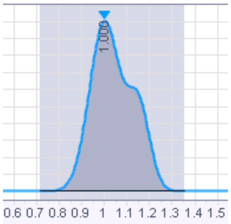


Figure 38 Shoulder sensitivity value set too high

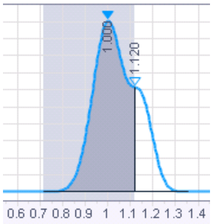


Figure 39 Shoulder sensitivity value set correctly

Integration off

This event turns off the integration of your chromatogram during the range specified. This event is useful if you are not interested in certain areas of your chromatogram, and do not wish peaks to be reported for that section.

When using **Integration Off** to disable peaks, these regions will be included in the noise calculation. Leave all peaks integrated to get the correct noise values.

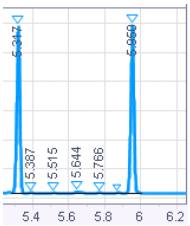


Figure 40 Default integration

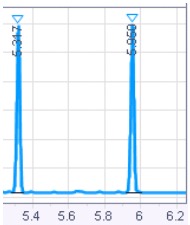


Figure 41 Integration off from 5.35 to 5.85 min

Valley to valley This event causes the baselines of peaks that are not totally resolved (i.e. do not return to baseline) to be drawn to the minimum point between the peaks.

If this event is not used, a baseline is projected to the next point at which the chromatogram returns to baseline, and a perpendicular is dropped for peaks which do not reach baseline.

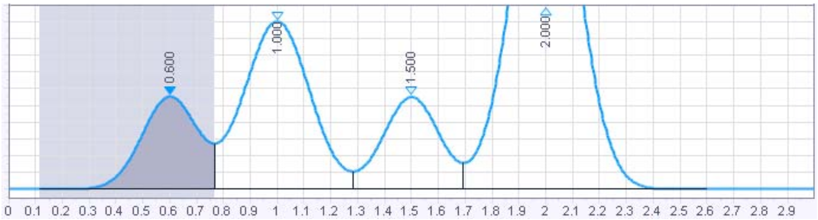


Figure 42 Default integration

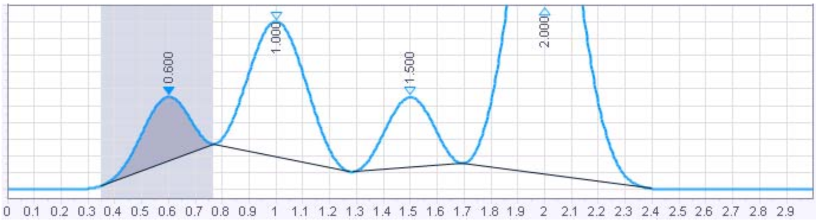


Figure 43 Integration with **Valley to valley** event

Horizontal baseline This event allows you to project the baseline forward horizontally between the times specified for the event.

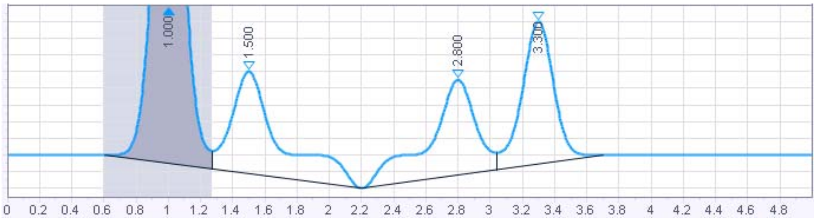


Figure 44 Default integration

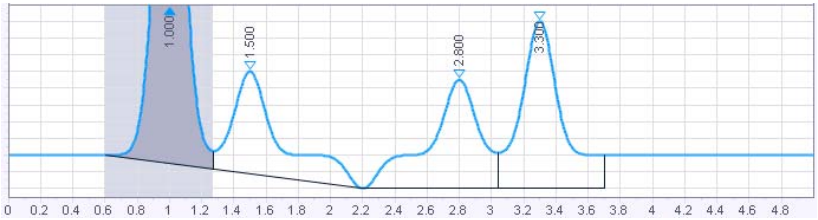


Figure 45 Integration with **Horizontal baseline** event between 2.2 and 3.7 min

Backward horizontal baseline This event is used to force a horizontal baseline in the direction of the beginning of the chromatogram. A backward horizontal baseline will be created between the times specified by the event.

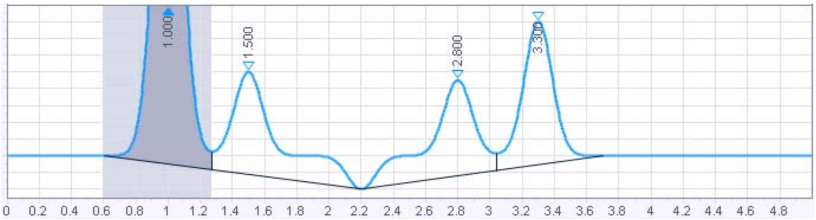


Figure 46 Default integration

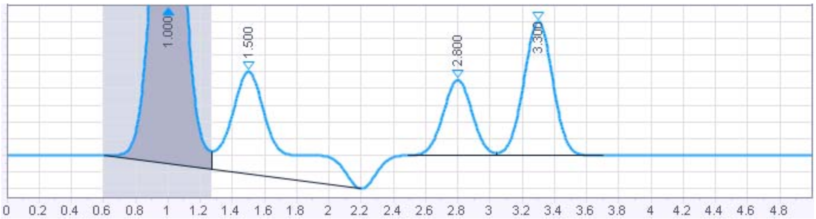


Figure 47 Integration with **Backward horizontal baseline** event between 2.2 and 3.7 min

Lowest point horizontal baseline

This event is similar to the Horizontal Baseline event, except that the lowest point in the chromatogram determines the baseline. The values you input for start and stop time determine the region within the chromatogram where the lowest point horizontal baseline will be used.

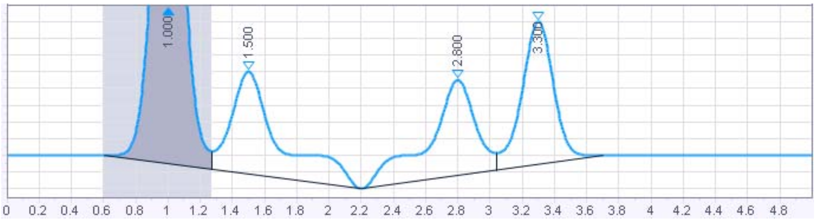


Figure 48 Default integration

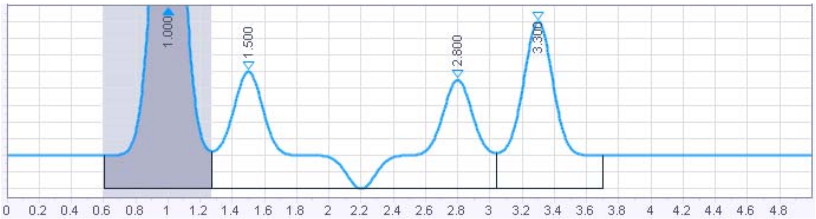


Figure 49 Integration after using **Lowest point horizontal** event

Tangent skim This event is used to integrate a small peak located on the tailing edge of a larger peak. The baseline of the small peak becomes a tangent drawn from the valley of the larger peak to the tangent point on the chromatogram.

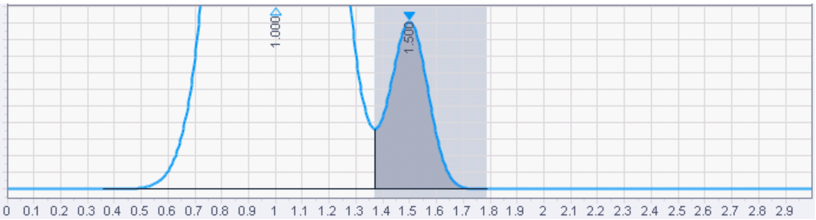


Figure 50 Default integration

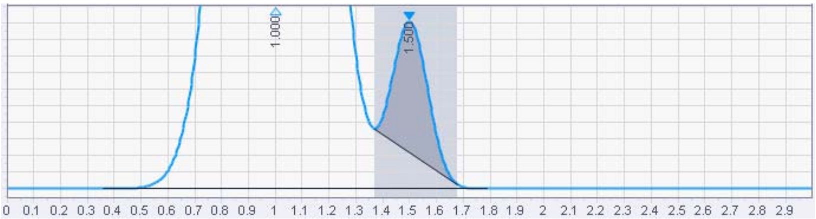


Figure 51 Integration with **Tangent skim** event

Front tangent skim This event is used to force a tangential baseline for a daughter peak on the leading edge of a mother peak.

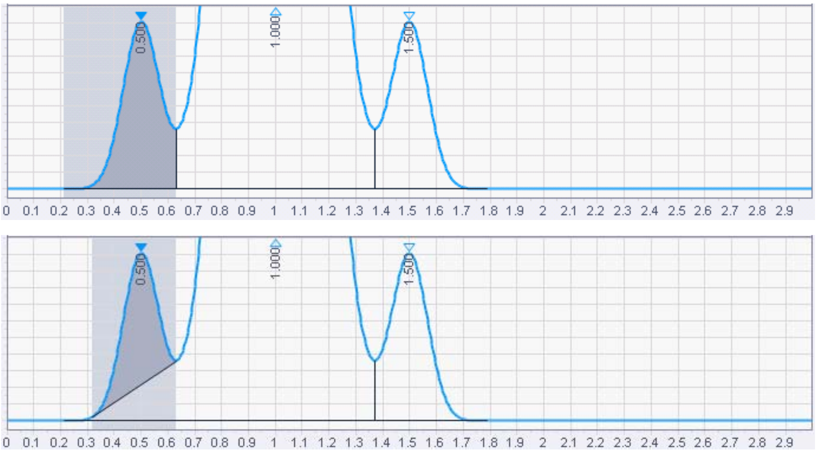


Figure 52 Integration with **Front tangent skim** event

Exponential skim

This event is used to integrate small peaks located on the tailing edge of a larger peak. The baseline of the small peak becomes an exponential drawn from the valley of the larger peak to the tangent point on the chromatogram.

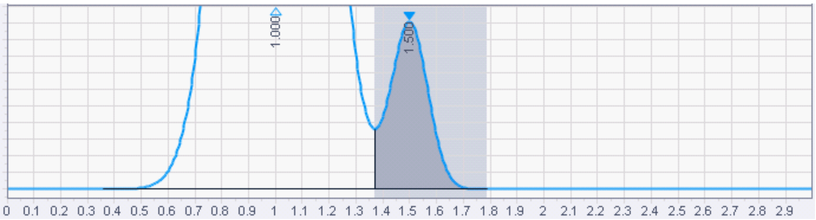


Figure 53 Default integration

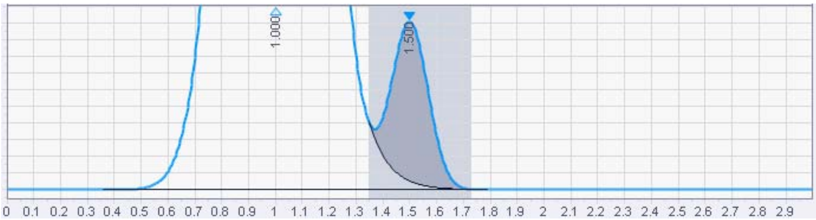


Figure 54 Integration with **Exponential skim** event

Front exponential skim

This event is used to force an exponential baseline for a daughter peak on the leading edge of a mother peak.

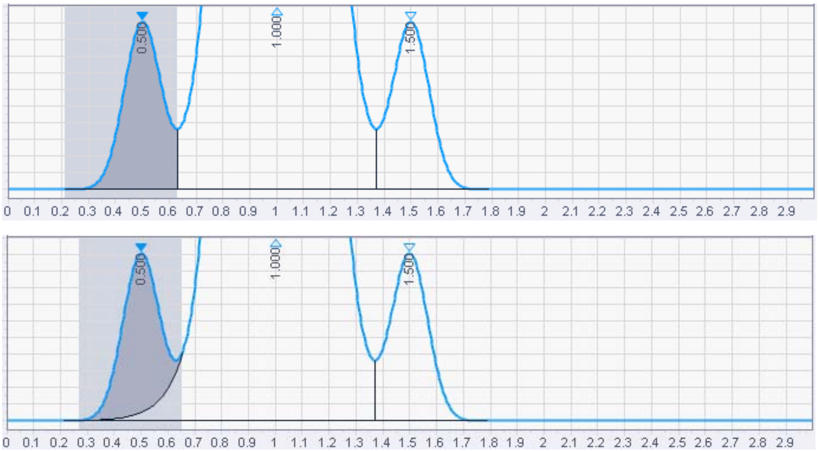


Figure 55 Integration with **Front exponential skim** event

Minimum area This event allows you to enter an area limit for peak detection. Peaks whose areas fall below this minimum area will not be integrated and reported as peaks. This event is useful for eliminating noise or contaminant peaks from your report.

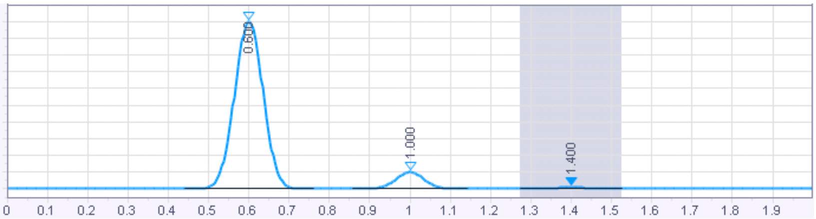


Figure 56 Default integration

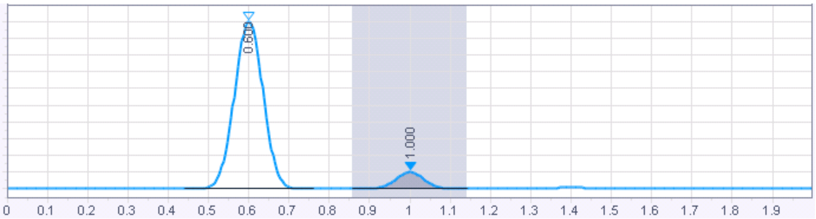


Figure 57 Integration with **Minimum area** event

Negative peak This event causes portions of the chromatogram that drop below the baseline to be integrated using the normal peak logic and reported as true peaks. This event is useful when using detectors such as Refractive Index types which give a negative response to certain compounds.

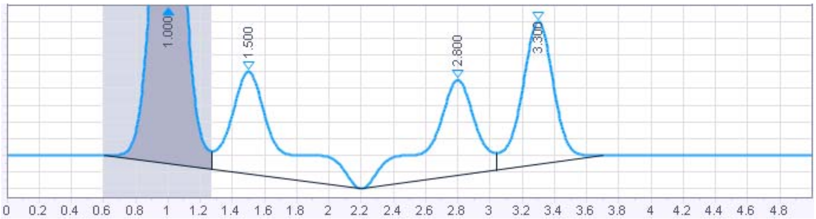


Figure 58 Default integration

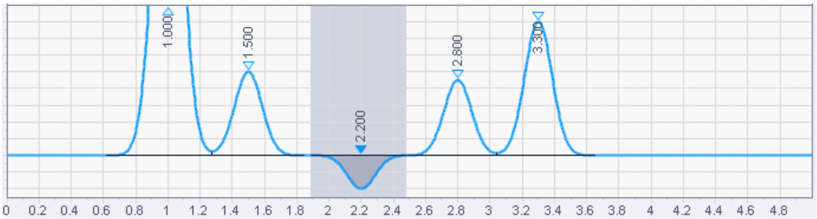


Figure 59 Integration with **Negative peak** event

Disable end of peak detection

This event is used to turn off end of peak detection between the specified times, forcing the software to treat peaks falling within the window of the event as a single peak. This event is a useful way to combine the areas of a series of contiguous peaks into one area. Because the peaks are considered to be part of a single peak, the retention time is assigned to the time of the first apex after the **Disable End of Peak Detection** event.

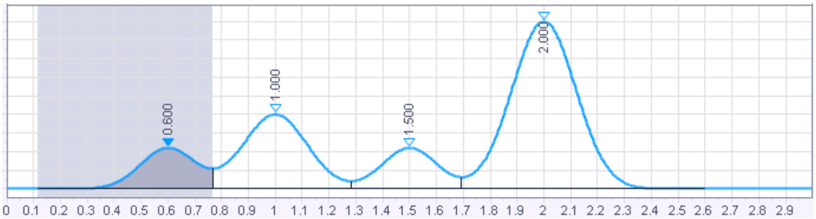


Figure 60 Default integration

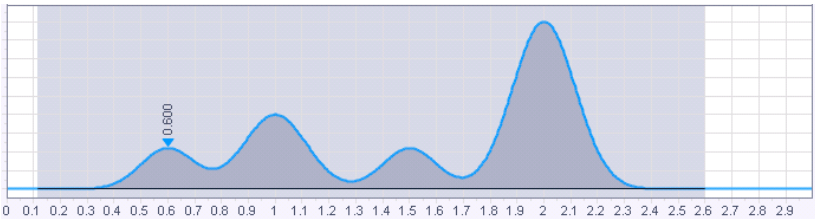


Figure 61 **Disable end of peak detection** between 0.4 and 2.3 min

Manual baseline This event allows you to change the way the baseline for a peak is drawn without changing the integration parameters. The baseline will be drawn from the signal at the start time to the signal at the stop time.

This is convenient when you want to change where a baseline is drawn for a peak without changing how the baseline is drawn for other peaks in the chromatogram.

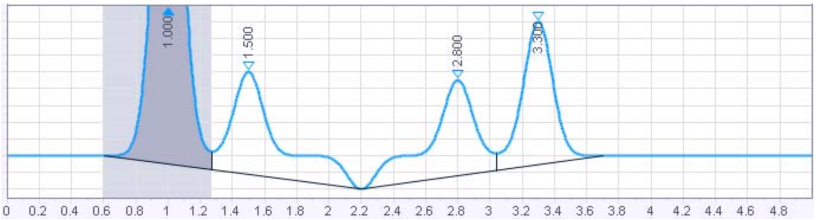


Figure 62 Default integration

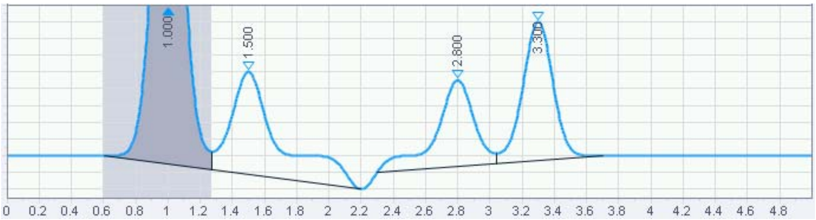


Figure 63 Integration with manual baseline between 2.3 and 3.6 min

Manual peak This command allows you to define the start and stop time of a peak that was not previously detected. This is convenient when you want to force integration of a peak, but do not want to change your overall integration parameters.

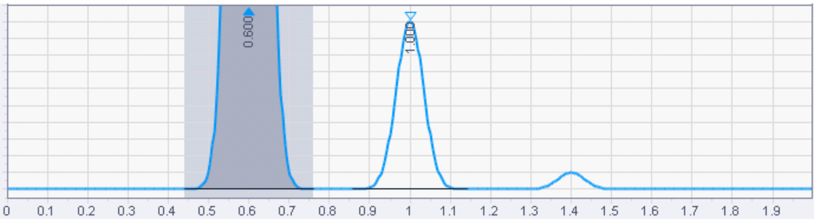


Figure 64 Default integration

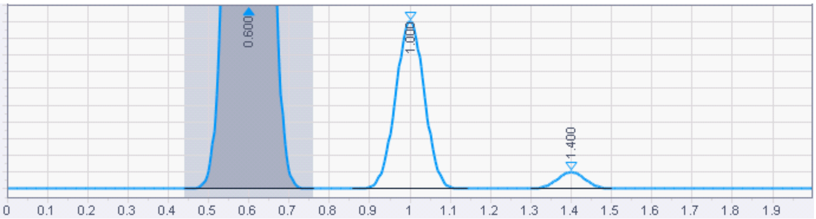


Figure 65 Small peak integration forced using Manual peak event between 1.3 and 1.5 min

Split peak This event is used to force a perpendicular drop-line integration in a peak. The perpendicular will be dropped at the time where the event is inserted.

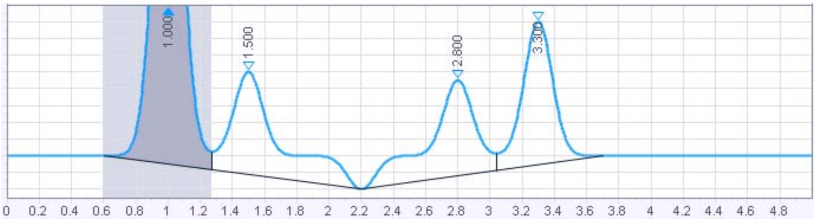


Figure 66 Default integration

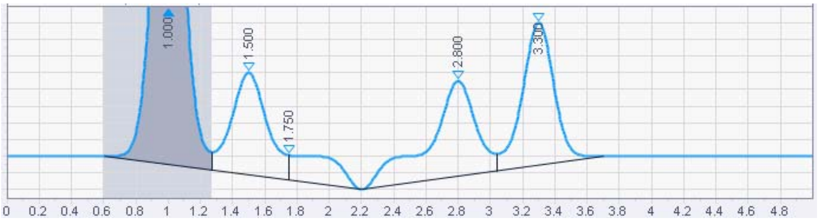


Figure 67 Integration with a **Split peak** event at 1.75 min

Force peak start / Force peak end These events are used to force the start or stop of the peak integration to a specific point.

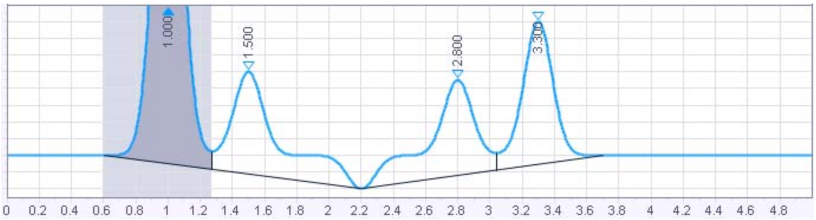


Figure 68 Default integration

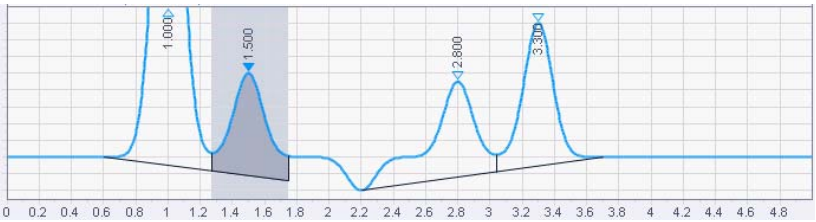


Figure 69 Integration with a **Force peak end** event at 1.75 min

Reset baseline This event lets you set the baseline at a designated point on the chromatogram.

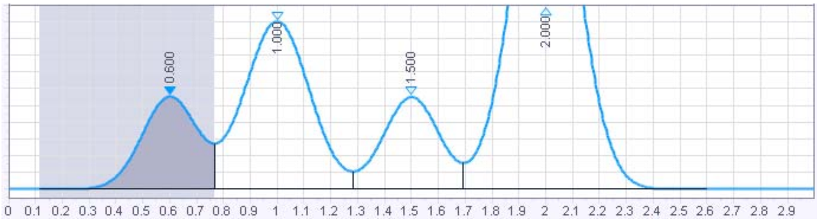


Figure 70 Default integration

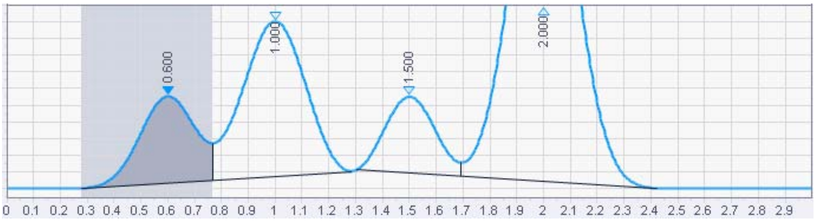


Figure 71 Integration with a **Reset baseline** event at 1.3 min

Reset baseline at valley This event will cause the baseline to be reset at the next valley detected after the event.

NOTE The event should be placed after the start of the first peak in the cluster; otherwise the start of the peak will be identified as the valley.

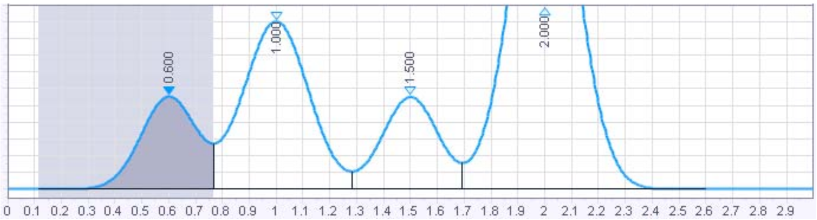


Figure 72 Default integration

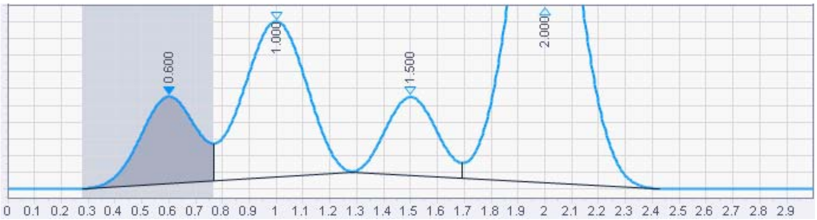


Figure 73 Integration with a **Reset baseline at valley** event at 1.2 min

Maximum area Sets the area of the largest peak of interest.

Any peaks that have areas greater than the maximum area are not reported: The integrator rejects any peaks that are greater than the maximum area value after baseline correction.

You can use this event, for example, to exclude the solvent peak of a GC chromatogram from the integration results, but include its rider peaks.

Maximum height Sets the height of the largest peak of interest.

Any peaks that have heights greater than the maximum height are not reported: The integrator rejects any peaks that are higher than the maximum height value after baseline correction.

You can use this event, for example, to exclude the solvent peak of a GC chromatogram from the integration results, but include its rider peaks.

4

Peak Identification

This chapter describes the concepts of peak identification.

What is Peak Identification?

Peak identification identifies the compounds in an unknown sample based on their chromatographic characteristics.

The identification of these compounds is a necessary step in quantitation if the analytical method requires quantitation. It is possible to create a valid processing method with identification even without quantitation. The signal characteristics of each component of interest are stored in the compound table of the method.

The function of the peak identification process is to compare each peak in the signal with the peaks stored in the compound table.

The identification is based on expected retention time, absolute retention time window, and relative retention time window in %. The final retention time window is the sum of relative and absolute windows, applied symmetrically to the expected retention time.

The expected retention time is either specified in the method as absolute time value or calculated from a relative retention time. Time reference compounds may be used to correct the expected retention times based on possible shifts observed by specific reference compounds.

$$\mathrm{Wnd\,Wdth} = \mathrm{Abs\,R\,T\,Wnd} + \frac{\mathrm{Exp\,R\,T} * \mathrm{Rel\,R\,T\,Wnd}}{100}$$

where

Abs R T Wnd	Absolute retention time window
Exp R T	Expected retention time
Rel R T Wnd	Relative retention time window
Wnd Wdth	Window width

R T Wnd = [Exp R T - Wnd Wdth; Exp R T + Wnd Wdth]

where

Exp R T	Expected retention time
R T Wnd	Retention time window
Wnd Wdth	Window width

Evaluation of the Retention Time Window

The identification window is the sum of relative and absolute window, applied symmetrically to the expected retention time. For example:

Expected retention time = 1 min

Absolute retention time window = 0.2 min

Relative retention time window = 10 % = 1 min * 10/100 = 0.1 min

Identification window = [1 - 0.2 - 0.1 ; 1 + 0.2 + 0.1] = [0.7 ; 1.3]

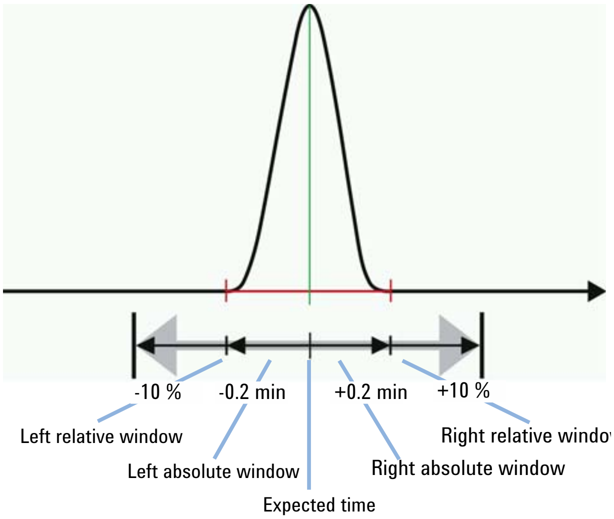


Figure 74 Identification window

Conflict Resolution

If multiple peaks are within the retention time window, there are different ways how to identify a particular peak. You can choose between the following values for **Peak match** in the compound identification parameters:

- **First**: Use the first peak in the retention time window.
- **Last**: Use the last peak in the retention time window.
- **Closest** (default setting): Use the peak that is closest to the expected retention time.
- **Largest area**: Use the peak with the largest area in the retention time window.
- **Largest height**: Use the peak with the largest height in the retention time window.

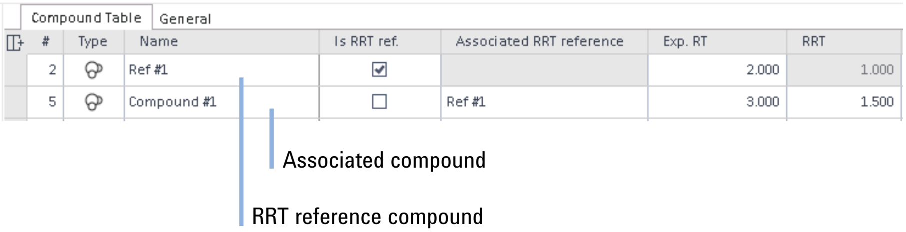
If the conflict cannot be resolved, none of the peaks will be identified and a warning will be written into the processing log.

Relative Retention Times

You can use relative retention times to check if the identification of your compounds is correct. The retention time of a compound is compared to the retention time of another specific given compound (also referred to as RRT reference). The ratio of the two retention times, that is, the relative retention time RRT, is normally a known number which you can provide in the application.

The RRT values themselves have no impact on the compound identification. Only the absolute expected retention times are used for this purpose. They are either specified in the processing method as absolute time values or calculated from relative retention times. Time reference compounds or method updates may be used to correct these absolute retention time windows based on possible shifts.

The following example shows the identification parameters for an RRT reference compound with an associated compound.



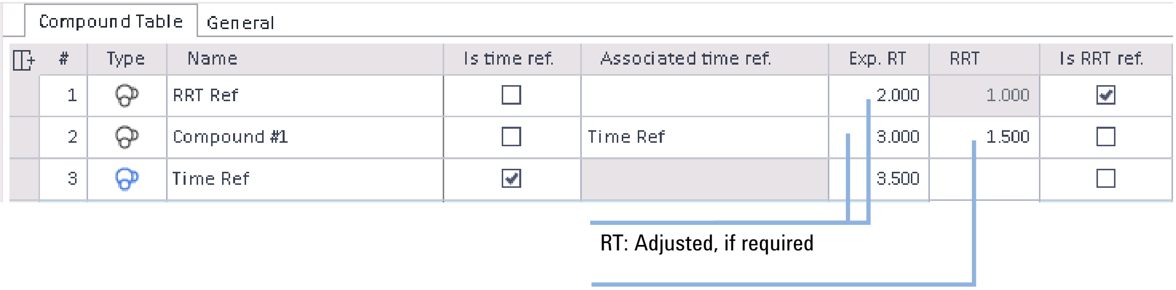
#	Type	Name	Is RRT ref.	Associated RRT reference	Exp. RT	RRT
2		Ref #1	☑		2.000	1.000
5		Compound #1	☐	Ref #1	3.000	1.500

Figure 75 RRT reference compound

If you change the expected RT of the associated compound, its RRT value will automatically be recalculated. Also vice versa, if you change its RRT value, its expected RT will be recalculated.

If you change the expected RT of the RRT reference compound, the system recalculates the expected RT of the associated compounds.

If you use *time reference compounds* with RRT reference compounds, the retention time shift is applied to the RRT reference compound (see "Calculations for Time Reference Compounds" on page 89). The system recalculates the expected RT of the associated compound so that the RRT values do not change.

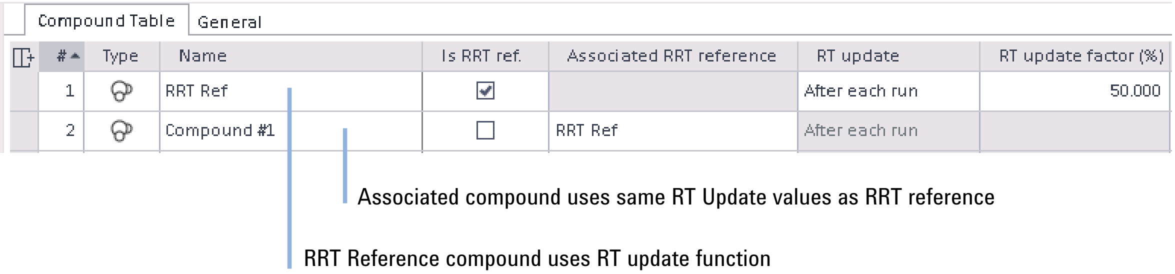


#	Type	Name	Is time ref.	Associated time ref.	Exp. RT	RRT	Is RRT ref.
1		RRT Ref	☐		2.000	1.000	☑
2		Compound #1	☐	Time Ref	3.000	1.500	☐
3		Time Ref	☑		3.500		☐

RT: Adjusted, if required

RRT: Remains constant

You can also use the **Update RT** function with RRT reference compounds. However, you can only configure the update parameters for the RRT reference compounds. The associated compounds are forced to use the same values as their references, so that the RRT values do not change.



#	Type	Name	Is RRT ref.	Associated RRT reference	RT update	RT update factor (%)
1		RRT Ref	☑		After each run	50.000
2		Compound #1	☐	RRT Ref	After each run	

Associated compound uses same RT Update values as RRT reference

RRT Reference compound uses RT update function

Calculations for Relative Retention Times (RRT)

Calculation of relative retention time (RRT) from expected retention times:

$$\text{RRT} = \text{expected RT}_{\text{compound}} / \text{expected RT}_{\text{reference}}$$

If you change the RRT value of the associated compound, its expected RT is recalculated as follows:

$$\text{Expected RT}_{\text{compound}} = \text{RRT} * \text{expected RT}_{\text{reference}}$$

Time Reference Compound

About Time Reference Compounds

If you use time reference compounds, the application corrects the absolute retention time windows based on possible shifts observed by specific reference compounds.

One or more compounds in the processing method can be marked as time reference compounds. For each compound or timed group, a time reference compound can be selected to correct the expected retention time. The extent of correction can be adjusted by an individual correction factor, which can be selected for each compound for correcting the expected retention time (column **Factor**, default = 1).

If a compound has a time reference compound assigned to it, the expected retention time will be corrected by the shift of the assigned time reference compound. The compound identification algorithm will use the corrected expected retention time for identifying the peak in the chromatogram. In case of timed groups, the time ranges are corrected by the shift. Generally the shift is corrected by the entered correction factor. If an associated reference compound is not found in the chromatogram, the linked peaks and time groups are not identified.

If *internal standards* are used and **Use time reference compounds** is selected, the internal standards are by default set as time reference compounds.

If you use *time reference compounds* and RRT reference compounds, both expected $RT_{reference}$ and expected $RT_{compound}$ are adjusted, so that the RRT remains constant.

NOTE When using time reference compounds, *all* compounds and timed groups must have a time reference compound assigned to them. Otherwise the method is inconsistent and cannot be used for reprocessing.

Calculations for Time Reference Compounds

If you use time reference compounds, the application corrects the absolute retention time windows based on possible shifts observed by selected time reference compounds.

Shift of the time reference compound

$$\text{Shift}_{\text{Ref}} = \text{ActualRT}_{\text{Ref}} - \text{ExpRT}_{\text{Ref}}$$

where

$\text{Shift}_{\text{Ref}}$	Time shift of the reference compound
$\text{ActualRT}_{\text{Ref}}$	Actual retention time of the reference compound
$\text{ExpRT}_{\text{Ref}}$	Expected retention time of the reference compound

RT of associated compound

For compounds that use a time reference, the expected retention time is calculated using an additional factor.

$$\text{CorrectedExpRT} = \text{ExpRT} + (\text{Shift}_{\text{Ref}} * \text{Factor})$$

where

CorrectedExpRT	Corrected expected retention
ExpRT	Expected retention time
$\text{Shift}_{\text{Ref}}$	Time shift of time reference compound
Factor	Factor for compounds with associated time reference compound (**Factor**)

Start and stop time of associated timed group

For timed groups, the expected start and stop times are calculated accordingly:

$$\text{Corrected Range Start} = \text{Range Start} + (\text{Shift}_{\text{Ref}} * \text{Factor})$$

$$\text{Corrected Range End} = \text{Range End} + (\text{Shift}_{\text{Ref}} * \text{Factor})$$

where

Corrected Range Start	Corrected start time of timed group
Range Start	Start time of timed group
Corrected Range End	Corrected stop time of timed group
RangeEnd	Stop time of timed group
$\text{Shift}_{\text{Ref}}$	Time shift of time reference compound
Factor	Factor for timed groups with an associated time reference compound (**Factor**)

Update Processing Method

Retention Time Updates

Based on the retention time update type (**Never**, **After each run**, or **After calibration standards**) of all identified compounds or timed groups, the expected retention time in the processing method is automatically updated after the corrected expected retention time has been calculated. If the compound can be found based on the corrected value, the corrected value becomes the new expected value in the method.

Retention time updates can be applied with or without time references.

NOTE If retention time update is set to **After each run** or **After calibration standards** all injections are processed in sequential order. The change in the method will be applied with the next injection and no more parallel processing of non-calibration injections can be done.

In addition to updating the retention times during the run, you can also manually shift all retention times by a given value.

Calculations for Updated Retention Times

To correct the expected retention times, the application reads the current retention time and calculates the shift to the expected retention time. This shift, multiplied by a compound-specific weighting factor, is added to the expected retention time.

$$Shift = ActualRT - ExpRT$$

where

Shift	Time shift of the compound
ActualRT	Actual retention time of the compound
ExpRT	Expected retention time of the compound

$$\mathrm{NewExpRT} = \mathrm{ExpRT} + \left(\mathrm{Shift} * \frac{\mathrm{RTUpdate}}{100}\right)$$

where

NewExpRT	Corrected expected retention time of the compound
ExpRT	Expected retention time of the compound
Shift	Time shift of the compound
RTUpdate	Weighting factor (**RTUpdate**) of the compound

You can use RT updates with or without *time references*. The shift and the corrected retention times of the time reference compounds themselves are calculated the same way as for any compound, using the *RT Update* function:

$$\mathrm{NewExpRT}_{\mathrm{Ref}} = \mathrm{ExpRT}_{\mathrm{Ref}} + \left(\mathrm{Shift}_{\mathrm{Ref}} * \frac{\mathrm{RTUpdate}_{\mathrm{Ref}}}{100}\right)$$

where

$\mathrm{NewExpRT}_{\mathrm{Ref}}$	Corrected expected retention time of the reference compound
$\mathrm{ExpR\ T}_{\mathrm{Ref}}$	Expected retention time of the reference compound
$\mathrm{Shift}_{\mathrm{Ref}}$	Time shift of the reference compound
$\mathrm{RTUpdate}_{\mathrm{Ref}}$	Weighting factor (**RTUpdate**) of the reference compound

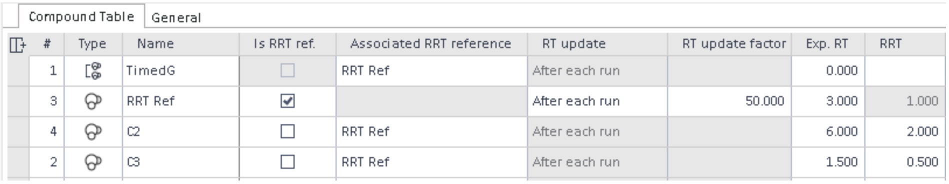
In case of *timed groups*, the expected retention time, range start time, or range end time are only updated if you use time references or relative retention times.

Example: Retention Time Updates with RRT

If you automatically update the retention times, and also use relative retention times, the values are updated as follows:

- The expected RT of the RRT reference compound is calculated as shown in the equation for NewExpRT (see “Calculations for Updated Retention Times” on page 91)
- The expected RT of the associated compound is adjusted to keep the RRT values constant, as shown in the equation for Expected $RT_{compound}$ (see “Calculations for Relative Retention Times (RRT)” on page 87)
- The RT start and RT stop times of a timed group are adjusted to keep the RRT values constant, using the same equation as for the expected $RT_{compound}$.

For example, consider a processing method with 3 compounds and a Timed Group, where RT Update and RRT are used:



#	Type	Name	Is RRT ref.	Associated RRT reference	RT update	RT update factor	Exp. RT	RRT
1		TimedG	☐	RRT Ref	After each run		0.000	
3		RRT Ref	☑		After each run	50.000	3.000	1.000
4		C2	☐	RRT Ref	After each run		6.000	2.000
2		C3	☐	RRT Ref	After each run		1.500	0.500

Figure 76 Compound parameters

Timed Group Parameters

☐ Include identified peaks

☐ Quantify each peak individually

start	RRT start	RT stop	RRT stop
3.000	1.00	6.000	2.00

Figure 77 Timed group parameters

After processing an injection, the 3 compounds were found at the following retention times:

- RRT ref: 4.000 min
- C2: 8.000 min
- C3: 2.000 min

As a result, the expected RT of the RRT reference compound is corrected as follows:

$$\text{NewExpRT} = \text{ExpRT} + \left(\text{Shift} * \frac{\text{RTUpdate}}{100}\right)$$

$$\begin{aligned}\text{NewExpRT} &= 3.000\ \text{min} + \left((4.000\ \text{min} - 3.000\ \text{min}) * \frac{50}{100}\right)\\ &= 3.500\ \text{min}\end{aligned}$$

The expected retention times of the other compounds as well as the start and stop times of the timed group are adjusted, so that RRT is constant.

$\text{Expected RT}_{\text{compound}} = \text{RRT} * \text{expected RT}_{\text{reference}}$

$\text{Expected RT}_{\text{C2}} = 2.000 * 3.500\ \text{min} = 7.000\ \text{min}$

$\text{Expected RT}_{\text{C3}} = 0.500 * 3.500\ \text{min} = 1.750\ \text{min}$

$\text{RTStart}_{\text{TimedG}} = 1.000 * 3.500\ \text{min} = 3.500\ \text{min}$

$\text{RTStop}_{\text{TimedG}} = 2.000 * 3.500\ \text{min} = 7.000\ \text{min}$

Calculation for Global Retention Time Shift

As part of method editing, you may shift all expected retention times and time ranges for timed groups at once. The new retention times are calculated as follows.

Absolute shift:

$$\text{NewExpRT} = \text{ExpRT} + \text{Shift}$$

where

NewExpRT	Corrected expected retention time of the compound
ExpRT	Expected retention time of the compound
Shift	Absolute value entered by the user

Relative shift:

$$\text{NewExpRT} = \text{ExpRT} + \left(\text{ExpRT} * \frac{\text{Shift}}{100}\right)$$

where

NewExpRT	Corrected expected retention time of the compound
ExpRT	Expected retention time of the compound
Shift	Relative value entered by the user

5

Calibration

This chapter contains details of the calculations used in the calibration process.

What is Calibration?

After the peaks have been integrated and identified, the next step in the quantitative analysis is the calibration. The amount and response is rarely in direct proportion to the actual mass of the sample to be analyzed. This makes the calibration with reference materials necessary. Quantitation uses peak area or height to determine the amount of a compound in a sample.

A quantitative analysis involves many steps which are briefly summarized as follows:

- Know the compound you are analyzing.
- Establish a method for analyzing samples containing a known amount of this compound, which is called the calibration sample or standard.
- Analyze the calibration sample to obtain the response due to that amount.

  You may alternatively analyze a number of these standards with different amounts of the compounds of interest if your detector has a non-linear response. This process is referred to as *multi-level calibration*.

With the following calibration methods you can perform quantitation:

- Compound specific calibration (ESTD, ISTD)
- Indirect quantitation using calibration or response factor from another compound or group
- Fixed response factor (**Manual Factor**)

The ESTD calibration curves and calculations are based on measured responses (area or height) of given amounts. The ISTD calibration curves and calculations are based on relative responses and relative amounts (see "Relative responses with ISTD" on page 102).

Calibration Curve

What is a Calibration Curve

A calibration curve is a graphical presentation of the amount and response data for one compound obtained from one or more calibration samples.

Normally an aliquot of the calibration sample is injected, a signal is obtained, and the response is determined by calculating the area or height of the peak, similar to the following figure.

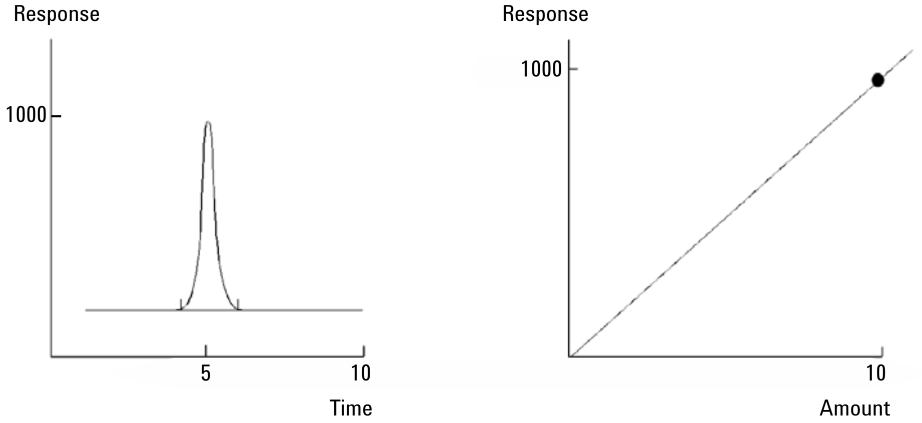


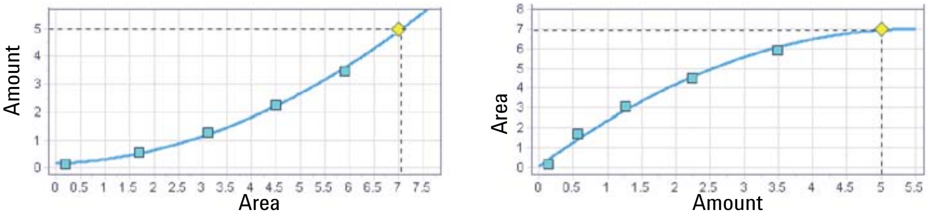
Figure 78 Calibration sample, signal, and calibration curve

Response Type and Response Factor

There are different settings that allow you to choose which values are plotted on the x and y axis of the calibration curve:

RF definition

The response factor (RF) is a measure of the extent to which the signal changes if a compound is detected. It is defined as the ratio of the response to the compound amount, or vice versa. In the general method settings under **RF definition**, you can switch between **Response per amount** (default) or **Amount per response**. If you change this setting, you swap the x and y axis of the calibration curve.



Amount per response

Response per amount (default)

Figure 79 Different RF definitions, Response set as **Area**

RF calculation:

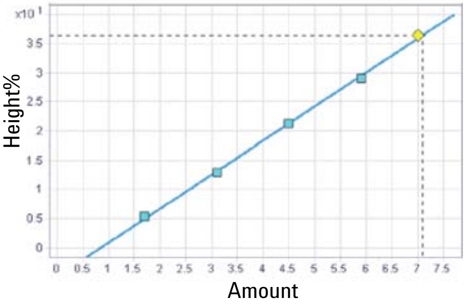
RF = Response/Amount

or

RF = Amount/Response

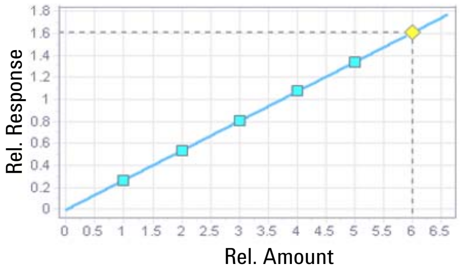
Type of response

The response itself can be defined as **Area**, **Area%**, **Height**, or **Height%**. You can choose the response type individually for each compound.

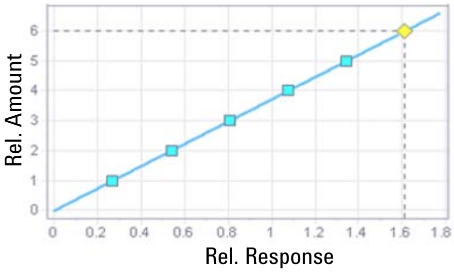


Relative responses with ISTD

If you use internal standards (ISTDs) in your sample, relative amounts and relative responses are shown in the calibration curve. The calculation depends on the RF definition.



Response per amount (default)



Amount per response

RF calculation:

RF = (Response/ISTD Response) / (Amount/ISTD Amount)

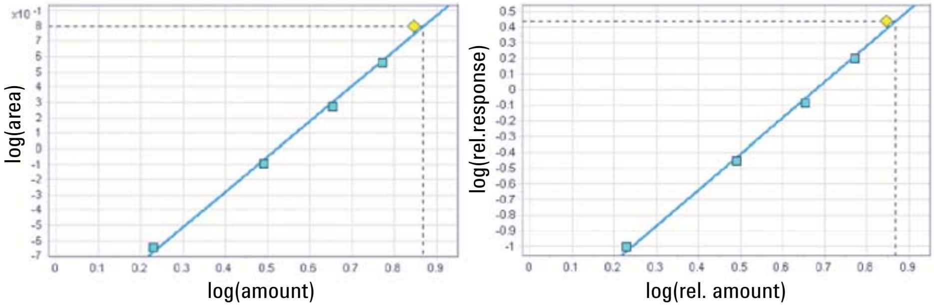
or

RF = (Amount/ISTD Amount) / (Response/ISTD Response)

Log/log curve model

If you select the curve model **log/log** for a compound, the amount and response are both plotted as logarithmic values.

You can use the **log/log** model in combination with both RF definitions, with all response types, and with internal or external standards.



Response per amount (default)

Amount per response

Figure 80 Example for log/log model with external or internal standards

RF calculation for the examples shown above:

RF = log(Response) / log(Amount)

RF = log(Response/ISTD Response) / log(Amount/ISTD Amount)

Calibration Level

There is one global number of calibration levels per processing method for all compounds. The number of calibration levels defines how many points (amount, response) are used to calculate the calibration curve. You define each level by processing the corresponding calibration sample. For each compound, the calibration curve shows the average calibration points and individual points which have been used to calculate the averages.

When re-running a calibration sample, or when processing further calibration samples of a given calibration level, the calibration point for that level can be updated. It is updated by the average value of the new measured point and the already existing value(s). How the point is evaluated depends on the configured **Curve calculation** (see “Modes of Using Individual Points” on page 104),

The collection of calibration points can be controlled by the **Run Type** as follows:

- No selection: A new point will be added to the calibration curve.
- **Clear all Calibration**: All calibration points for all calibration levels are deleted before the new calibration data (response factor) is saved.
- **Clear Calibration at Level**: All calibration points for the given calibration level are deleted before the new calibration data (response factor) is saved.

If a level is cleared, the respective calibration points are removed from the calibration curve before new ones are added during processing of the standard.

Reprocessing the same calibration sample injection multiple times will update the same calibration point in the curve and not add new points.

Modes of Using Individual Points

You can choose per processing method how calibration points are used for calculating the calibration curve. The following modes are available:

- **From average per level**: Amounts and responses of all calibration points contributing to a level will be averaged and used in the algorithm to calculate the best calibration curve.
- **From individual calibration points**: All amounts and responses of the individual calibration points will be used directly to determine the calibration curve.

Average

The average from all calibration runs is calculated using the following formula:

$$\mathrm{Response} = \frac{((n-1) * \mathrm{Response}) + \mathrm{MeasResponse}}{n}$$

where

n	Number of calibration points
MeasResponse	Measurement response

Bracketed Calibration

With bracketed calibrations, the samples are bracketed by pre-sample and post-sample calibrations. The calibration standards between opening and closing brackets are processed first, and a calibration curve is calculated. This curve is then used to calculate the samples in between the calibration standards. A **Clear all calibration** operation is performed for all opening brackets.

Bracketing is configured in the **Injection List** window. There are different bracketing modes:

- **Overall**

  The calibration curve is calculated with all calibration standards in the sequence, starting with the first one and finishing with the last one. All samples are reprocessed *after* the calibration curve has been calculated.

- **Non overlap**

  You must have at least three groups of standards in your sequence, and at least two standards in the middle blocks. The standards from the group in the middle of the sequence are used in one calibration curve only.

  If there are more than two standards in the middle block, they will be divided and allocated to the preceding and subsequent groups. With uneven numbers of standards in the middle block, the extra standard is allocated to the preceding group.

- **Overlap**

  You must have at least three groups of standards in your sequence. Standards from the group in the middle of the sequence are used in two calibration curves (with the preceding and with the following block).

- **Custom**

  Create brackets as required. In the **Run type** column, you can choose for each calibration standard individually which calibration levels shall be cleared. If you do not choose any run type, a bracket will be averaged with its predecessor.

Calibration Point Weighting

To compensate for the variance of the response at different calibration amounts, you can specify the relative weighting (or importance) of the various calibration points used to generate the curve.

The parameter that controls the weighting is **Weighting Method**. The default weight is equal weight for all levels and the maximum weight for each curve is normalized to 1.

The following weighting factors are available:

None

All calibration points have equal weight.

$$wt = 1$$

where

wt	Calibration level weighting factor

1/Amount

A calibration point is weighted by the factor 1/Amount, normalized to the smallest amount so that the largest weighting factor is 1. If the origin is included, it is assigned the mean of the weightings of the other calibration points.

$$wt = \frac{\mathrm{Minimum(Amounts)}}{\mathrm{CurrentAmount}}$$

where

Current Amount	Level amount
Minimum (Amounts)	Lowest amount across all points (levels) used for the calibration curve
wt	Calibration level weighting factor

1/Amount squared

A calibration point is weighted by the factor 1/Amount^2, normalized to the smallest amount so that the largest weighting factor is 1. Quadratic calibration point weightings can be used, for example, to adjust for a spread in calibration points. It makes sure that calibration points closer to the origin, which can normally be measured more accurately, get a higher weight than calibration points further away from the origin, which may be spread.

$$wt = \frac{\mathrm{Minimum(Amounts)}^2}{\mathrm{CurrentAmount}^2}$$

where

Current Amount	Level amount
Minimum (Amounts)	Lowest amount across all points (levels) used for the calibration curve
wt	Calibration level weighting factor

1/Response

A calibration point is weighted by the factor 1/Response, normalized to the smallest response so that the largest weighting factor is 1. If the origin is included, it is assigned the mean of the weightings of the other calibration points.

$$wt = \frac{\mathrm{Minimum(Responses)}}{\mathrm{CurrentResponse}}$$

where

Current Response	Level response
Minimum (Responses)	Lowest response across all points (levels) used for the calibration curve
wt	Calibration level weighting factor

1/Response squared

A calibration point is weighted by the factor 1/Response^2, normalized to the smallest response so that the largest weighting factor is 1. Quadratic calibration point weightings can be used, for example, to adjust for a spread in calibration points. It makes sure that calibration points closer to the origin, which can normally be measured more accurately, get a higher weight than calibration points further away from the origin, which may be spread.

$$wt = \frac{\mathrm{Minimum(Responses)}^2}{\mathrm{CurrentResponse}^2}$$

where

Current Response	Level response
Minimum (Amounts)	Lowest response across all points (levels) used for the calibration curve
wt	Calibration level weighting factor

Calibration Curve Models

OpenLAB CDS can calculate the calibration according to different models. The following models are supported (see “Calibration Curve Calculation” on page 111):

- Linear fit (see “Linear Fit” on page 113)
- Quadratic fit (see “Quadratic Fit” on page 114)
- Logarithmic and exponential fit (see “Logarithmic and exponential fits” on page 117)

You can set the calibration curve model individually for each calibrated compound.

Origin Handling

The application can consider the origin of the graph in different ways when calculating the calibration curve. You can set this parameter independently for each compound. Depending on the curve type, only specific origin handling options are available (for example, you cannot force the curve through the origin with a logarithmic calibration curve).

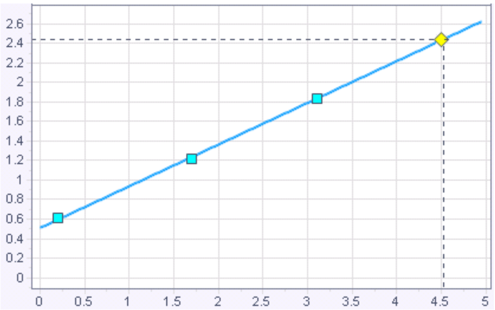


Figure 81 Ignore origin

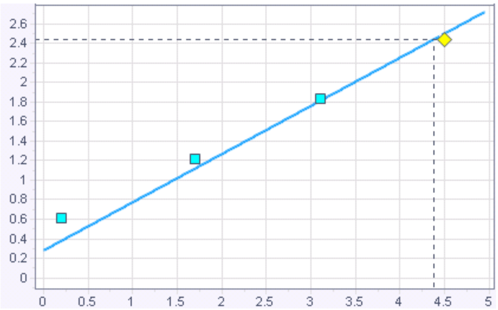


Figure 82 Include origin into the calculation

With the **Include** option, a point with amount=0 and response=0 is added to the calibration levels.

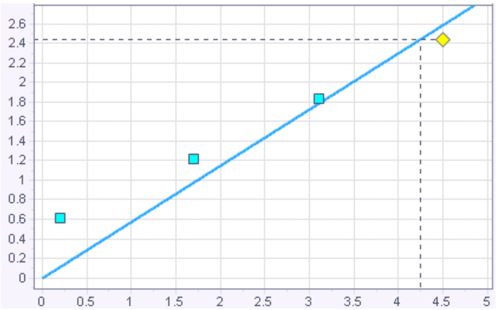


Figure 83 Force calibration curve through the origin

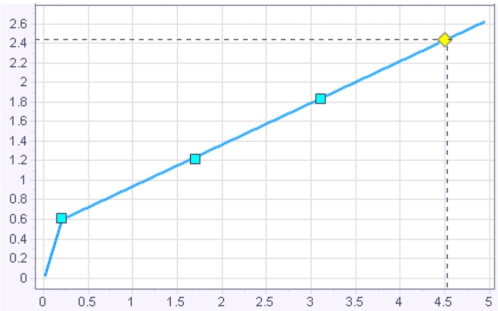


Figure 84 Connect the origin (not included in the calculation)

Calibration Curve Calculation

The optimal calibration curve is calculated by matching the curve to the calibration points. The curve calculation is based on a least squares fit (LSQ), which minimizes the sum of the residual squares. The curve type is applied to weighted calibration points. The calculation depends on the definition of the response factor (RF definition, see “Response Type and Response Factor” on page 100).

With RF defined as **Response per amount**:

$$\Sigma(\mathrm{wt} * (\mathrm{CalPointArea} - \mathrm{CalculatedArea})^2) = \mathrm{min}$$

where

Σ	Sum over the calibration points (levels)
CalculatedArea	The area read from the curve at calibration level amount
CalPointArea	Calibration level area
wt	Calibration point weighting factor

$$\Sigma(\mathrm{wt} * (\mathrm{CalPointHeight} - \mathrm{CalculatedHeight})^2) = \mathrm{min}$$

where

Σ	Sum over the calibration points (levels)
CalculatedHeight	The height read from the curve at calibration level amount
CalPointHeight	Calibration level height
wt	Calibration point weighting factor

With RF defined as **Amount per response**:

$$\Sigma(wt * (CalPointAmount - CalculatedAmount)^2) = min$$

where

Σ	Sum over the calibration points (levels)
CalculatedAmount	The amount read from the curve at calibration level area or height
CalPointAmount	Calibration level amount
wt	Calibration point weighting factor

Parameters for Curve Calculation

The curve calculations all use the following parameters:

a, b, c	Curve coefficients
x	With **Response per amount**: Amount (ESTD), or amount ratio (ISTD) With **Amount per response**: Area, area%, height, or height% (ESTD) Area ratio or height ratio (ISTD)
y	With **Response per amount**: Area, area%, height, or height% (ESTD) Area ratio or height ratio (ISTD) With **Amount per response**: Amount (ESTD), or amount ratio (ISTD)
wt	Calibration level weighting factor

Linear Fit

The curve calculation is based on the least squares fit (see "Calibration Curve Calculation" on page 111).

Curve formula:

$$y = a + b * x$$

where

a	Y-intercept
b	Slope

Calculation of curve coefficients:

$$a = \frac{\sum(x^2 * wt) * \sum(y * wt) - \sum(x * y * wt) * \sum(x * wt)}{\sum(wt) * \sum(x^2 * wt) - \sum(x * wt)^2}$$

$$b = \frac{\sum(wt) * \sum(x * y * wt) - \sum(x * wt) * \sum(y * wt)}{\sum(wt) * \sum(x^2 * wt) - \sum(x * wt)^2}$$

At least two calibration points are required for a linear fit.

Include origin

If the origin is included, the point (0,0) is added to the other points and weighted by the mean value of the weights of the other points, that is, the Σ(wt) term is increased by the mean value of the weights of the other points.

Force origin

If the force origin option is selected the curve formula is as follows:

$$y = b * x$$

where

b	Slope

Calculation of curve coefficient:

$$b = \frac{\sum(x * y * wt)}{\sum(x^2 * wt)}$$

Only one calibration level is required when the origin is included or forced.

Quadratic Fit

Quadratic curve formula:

$$y = a + (b * x) + (c * x^2)$$

At least three calibration points are required for the quadratic fit. Two points are required if the origin is included or forced.

Calculation of coefficients for quadratic fit

The coefficients result from the below simultaneous linear equations. Crout's algorithm is used to solve the corresponding normal matrix equation ($A^TAx = A^Ty$). In the given formula, sums are abbreviated as:

$W = \sum(wt)$

$XW = \sum(x * wt)$

$X2W = \sum(x^2 * wt)$

$X3W = \sum(x^3 * wt)$

$X4W = \sum(x^4 * wt)$

$YW = \sum(y * wt)$

$XYW = \sum(x * y * wt)$

$X2YW = \sum(x^2 * y * wt)$

In order to avoid overflow, the x-values are normalized before entering calculation:

$$Norm = \Sigma(x)$$

$$x = x / Norm$$

Normal equations for quadratic curve:

$$\begin{aligned}
&\sum(wt) * a &&+ \sum(x * wt) * b &&+ \sum(x^2 * wt) * c &&= \sum(y * wt)\\
&\sum(x * wt) * a &&+ \sum(x^2 * wt) * b &&+ \sum(x^3 * wt) * c &&= \sum(x * y * wt)\\
&\sum(x^2 * wt) * a &&+ \sum(x^3 * wt) * b &&+ \sum(x^4 * wt) * c &&= \sum(x^2 * y * wt)
\end{aligned}$$

Or written as matrix equation:

$$\begin{vmatrix} W & XW & X2W \\ XW & X2W & X3W \\ X2W & X3W & X4W \end{vmatrix} * \begin{vmatrix} a \\ b \\ c \end{vmatrix} = \begin{vmatrix} YW \\ XYW \\ X2YW \end{vmatrix}$$

Crout's decomposition:

$$\begin{vmatrix} W & XW & X2W \\ XW & X2W & X3W \\ X2W & X3W & X4W \end{vmatrix} = \begin{vmatrix} L11 & & \\ L21 & L22 & \\ L31 & L32 & L33 \end{vmatrix} * \begin{vmatrix} 1 & U12 & U13 \\ & 1 & U23 \\ & & 1 \end{vmatrix}$$

With value abbreviations:

$$L11 = W$$

$$U12 = \frac{XW}{L11}$$

$$L21 = XW$$

$$U13 = \frac{X2W}{L11}$$

$$L31 = X2W$$

$$L22 = X2W - L21 * U12$$

$$U23 = \frac{X3W - L21 * U13}{L22}$$

$$L32 = X3W - L31 * U12$$

$$L33 = X4W - (L31 * U13) - (L32 * U23)$$

$$z0 = \frac{YW}{L11}$$

$$z1 = \frac{XYW - (L21 * z0)}{L22}$$

$$z2 = \frac{X2YW - (L31 * z0) - (L32 \cdot z1)}{L33}$$

$$c' = z2$$

$$b' = z1 - (U23 * c')$$

$$a' = z0 - (U12 * b') - (U13 * c')$$

Finally, the normalization must be reversed:

$$a = a'$$

$$b = \frac{b'}{Norm}$$

$$c = \frac{c'}{Norm^2}$$

Force Origin

If the force origin option is selected, the offset term *a* is set to zero when creating the normal equations.

$$\begin{vmatrix} X2W & X3W \\ X3W & X4W \end{vmatrix} * \begin{vmatrix} b \\ c \end{vmatrix} = \begin{vmatrix} XYW \\ X2YW \end{vmatrix}$$

$$L11 = X2W$$

$$U12 = \frac{X3W}{L11}$$

$$L21 = X3W$$

$$L22 = X4W - (L21 * U12)$$

$$z0 = \frac{XYW}{L11}$$

$$z1 = \frac{X2YW - (L21 * z0)}{L22}$$

$$c' = z1$$

$$b' = z0 - (U12 * c')$$

$$b = \frac{b'}{Norm}$$

$$c = \frac{c'}{Norm^2}$$

Include origin

If the origin is included, the point (0,0) is added to the other points and weighted by the mean value of the weights of the other points, that is, the $\Sigma(wt)$ term is increased by the mean value of the weights of the other points.

Logarithmic and Exponential Fits

Logarithmic and exponential fits

To calculate the exponential and logarithmic fit, the amount or response scales are transformed using the ln function. The linear curve fit and the weight factors are applied to the transformed data, and the curve is calculated on the transformed data.

The **Include origin** and **Force origin** options are not valid due to the singularity of the ln function at the origin.

Logarithmic

Curve formula:

$$y = a + b * \ln(x)$$

Transformations: The x scale is transformed.

$$x' = \ln(x); y' = y$$

$$y' = a + b * x'$$

Exponential

Curve formula:

$$y = a * e^{b * x}$$

Transformations: The y scale is transformed.

$$x' = x; y' = \ln(y)$$

$$y' = \ln(a) + b*x'$$

Log/log fit

To calculate the log/log fit, both amount and response scales are transformed using the log function. The linear curve fit and the weight factors are applied to the transformed data, and the curve is calculated on the transformed data.

The **Include origin** and **Force origin** options are not valid due to the singularity of the log function at the origin.

Curve formula:

$$\log(y) = a + b * \log(x)$$

Transformations: The x and y scales are transformed.

$$x' = \log(x); y' = \log(y)$$

$$y' = a + b * x'$$

Evaluating the Calibration Curve

The quality of the fit of the calibration curve to the calibration levels, and the presence of outliers (measurements are at a long distance from the curve) can be evaluated using statistical calculations. The calibration curve calculation provides a correlation coefficient and a relative standard deviation for each curve, as well as a relative residual value for each calibration level.

Verification of the Calibration Curve

After calculations the calibration curves are verified and warnings are set if:

- there are not enough calibration points for the curve calculation
- the curve slope gets zero or negative
- the slope is infinite
- the calibration curve cannot be calculated (for example numeric overflows)

Relative Residuals

Residual is a measure of the calibration point distance from the calculated curve:

$$\text{Residual} = y_i - Y_i$$

where

y_i	Measured response (area or height) or amount, depending on the calibration mode.
Y_i	Predicted response or amount for level i (calculated using the curve)

The relative residual is calculated for each calibration level using the following formula:

$$\text{Rel Residual} = \frac{\text{Residual}}{Y_i} = \frac{(y_i - Y_i)}{Y_i}$$

where

y_i	Measured response (area or height) or amount
Y_i	Predicted response or amount for level i (calculated using the curve)

The relative residual is frequently reported in % units (**RelResidual%**). In that case the **RelResidual** needs to be multiplied by 100.

Calibration Curve Statistics

Calibration curve statistics

The calibration curve calculation provides for each curve the correlation coefficient, coefficient of determination and residual standard deviation figures.

Correlation Coefficient

The correlation coefficient (r) gives a measure of the fit of the calibration curve between the data points. It is calculated using the following equation:

$$r = \frac{\sum((y_i - \bar{y}) * (Y_i - \bar{Y}) * wt_i)}{\sum((y_i - \bar{y})^2 * wt_i) * \sum((Y_i - \bar{Y})^2 * wt_i)^{\frac{1}{2}}}$$

where

r	Correlation coefficient
wt_i	Weight of the data point
$\bar{y}$	Mean values of the measured responses or amounts
yi	Measured response (Area, AreaRatio (ISTD method), Height or HeightRatio (ISTD method)) or amount (Amount, AmountRatio (ISTD Method)), depending on calibration mode
$\bar{Y}$	Mean values of the predicted responses or amounts
Yi	Predicted response or amount (using the calibration curve)

$\bar{y}$ and $\bar{Y}$ are mean values of the measured and predicted responses or amounts, calculated as follows:

$$\bar{y} = \frac{\sum(y_i * wt_i)}{\sum(wt_i)}$$

where

wt_i	Weight of the data point
$\bar{y}$	Mean values of the measured responses or amounts
yi	Measured response (Area, AreaRatio (ISTD method), Height or HeightRatio (ISTD method)) or amount (Amount, AmountRatio (ISTD Method)), depending on calibration mode

and

$$\bar{Y} = \frac{\sum(Y_i * wt_i)}{\sum(wt_i)}$$

where

wt_i	Weight of the data point
$\bar{Y}$	Mean values of the predicted responses or amounts
Yi	Predicted response or amount (using the calibration curve)

For **Forced Origin** it is assumed that the points are centered on zero (mirrored to third quadrant) and the mean values are substituted with zero.

The correlation coefficient is 1 for a perfect fit or when points are distributed symmetrically around the curve. It reduces as the calibration points are distributed less symmetrically. Typical values are between 0.99 and 1. The correlation coefficient is not a very sensitive measure of curve quality.

Determination coefficient

The determination coefficient (R^2) is calculated as follows:

$$R^2 = 1 - \frac{\sum(y_i - Y_i)^2}{\sum(y_i - \overline{y})^2}$$

where

R^2	Determination coefficient
$\overline{y}$	Mean values of the measured responses or amounts
yi	Measured response or amount. Response can be area (Area, Area%, or AreaRatio (ISTD method)) or height (Height, Height%, or HeightRatio (ISTD method)). Amount can be absolute amount or AmountRatio (ISTD method). The type of value depends on the calibration mode.
Yi	Predicted response or amount (using the calibration curve)

Residual standard deviation

The residual standard deviation (sometimes referred to as the mean square error) is calculated using the following formula:

$$\mathrm{ResidualStdDev} = \sqrt{\frac{\sum(y_i - Y_i)^2}{(n - d)}}$$

where

$d = 3$	Degree of freedom for a quadratic curve, no forced origin
$d = 2$	Degree of freedom for a quadratic curve with forced origin Degree of freedom for a linear curve, no forced origin
$d = 1$	Degree of freedom for a linear curve with forced origin
ResidualStdDev	Residual standard deviation
yi	Measured response (Area, AreaRatio (ISTD method), Height or HeightRatio (ISTD method)) or amount (Amount, AmountRatio (ISTD Method)), depending on calibration mode
Yi	Predicted response or amount (using the calibration curve)
n	number of calibration points

For **Include origin** calibration curve types, the origin (0,0) is included as a regular point in the calculation and counted by n.

The y values are not weighted.

The residual standard deviation gives a more sensitive measure of the curve quality than does the correlation coefficient. For a perfect fit, the residual standard deviation is zero. With increasing residual standard deviation values, the calibration points get further away from the curve.

6

Quantitation

This chapter describes how compounds are quantified, and explains the calculations used in quantitation.

What is Quantitation?

After the peaks have been integrated and identified, the next step in the analysis is quantitation. Quantitation uses peak area or height to determine the amount of a compound in a sample.

A quantitative analysis involves many steps which are briefly summarized as follows:

- Analyze the sample containing an unknown amount of the compound to obtain the response due to the unknown amount.
- Compare the response of the unknown amount to the response of the known amount to determine how much of the compound is present.

To obtain a valid comparison for the unknown sample response to that of the known sample, the data must be acquired and processed under identical conditions.

Quantitation Calculations

OpenLAB CDS offers the following calculation procedures for determining the amount of each component present in a mixture:

- Area or Height Percent (Area% or Height%)
- Quantitation using a Manual Factor
- External standard (ESTD)
- Internal standard (ISTD)
- Indirect Quantitation using a calibrated compound

The calculations used to determine the concentration of a compound in an unknown sample depend on the type of quantitation. Each calculation procedure uses the peak area or height for the calculation and produces a different type of analysis.

Correction Factors

The quantitation calculations use different correction factors, the *multiplier* (compound or injection multiplier), and the *dilution factor*. These factors are used in the calibration procedures to compensate for variations in detector response to different sample components, concentrations, sample dilutions, sample amounts, compound purities, and for converting units.

Multipliers

The multipliers are used in each calculation formula to multiply the result for each compound. A multiplier may be used to convert units to express concentrations, or to correct the concentration and thus compensate for different purities of the standard compounds.

Multipliers are set at the injection level (injection list or sequence table) and at the compound level (calibration table, part of the processing method). In OpenLAB CDS, you can configure up to 5 injection multipliers and 1 compound multiplier.

The multiplier for a known compound is:

> Multiplier = Compound Multiplier * Injection Multiplier 1 * Injection Multiplier 2 *...

Dilution Factor

The dilution factor is a number by which the amount is multiplied or divided to calculate the concentration (see concentration). The dilution factors are set at injection level (**Dil. factor** columns in the injection list). You can use the dilution factor to change the scale of the results or correct for changes in sample composition during pre-analysis work. You can also use the dilution factor for any other purposes that require the use of a constant factor.

The sample dilution is a combination of up to 5 dilution factors:

> Sample Dilution = Dilution Factor 1 * Dilution Factor 2 *...

Concentration and Mass%

Concentration

When calculating the concentration, you can use the dilution factors either as a divisor or as another multiplier. You configure the usage in the processing method, in the **General** tab of the **Compounds > Calibration** node.

Depending on the settings, the concentration is calculated as follows:

$$\text{Concentration} = \text{Amount} * \text{Multipliers} * \text{Dilution Factors}$$

or

$$\text{Concentration} = \text{Amount} * \frac{\text{Multipliers}}{\text{Dilution Factors}}$$

For more information on the calculation of multipliers and dilution factors, see “Correction Factors” on page 127.

Mass%

If you choose **Calculate mass %** in the global settings for calibration parameters, the concentration is calculated as a mass percentage (compound amount relative to sample amount). The mass percentage is also shown in the **Concentration** column in the injection results.

Depending on how calculate the concentration, the mass percentage is calculated as follows:

$$\text{Concentration} = \left(\frac{\text{Amount}}{\text{Sample Amount}} * 100\right) * \text{Multipliers} * \text{Dilution Factors}$$

or

$$\text{Concentration} = \left(\frac{\text{Amount}}{\text{Sample Amount}} * 100\right) * \frac{\text{Multipliers}}{\text{Dilution Factors}}$$

Area% and Height%

The **Area%** calculation procedure reports the area of each peak in the signal as a percentage of the total area of all peaks in the signal. **Area%** does not require prior calibration and does not depend upon the amount of sample injected within the limits of the detector. No response factors are used. If all components respond equally in the detector, then **Area%** provides a suitable approximation of the relative amounts of components.

Area% is used routinely where qualitative results are of interest and to produce information to create the compound table required for other calibration procedures.

The **Height%** calculation procedure reports the height of each peak in the signal as a percentage of the total height of all peaks in the signal.

Correction factors are not applied in Area% or Height% calculation.

Quantitation of Calibrated Compounds

The external standard (ESTD), normalization, and internal standard (ISTD) calculation procedures require calibration and therefore use a compound table. The compound table specifies conversion of responses into the units you choose by the procedure you select.

ESTD Calculation

The ESTD procedure is the basic quantitation procedure in which both calibration and unknown samples are analyzed under the same conditions. The results from the unknown sample are then compared with those of the calibration sample to calculate the amount in the unknown.

The ESTD procedure uses absolute response factors unlike the ISTD procedure. The response factors are obtained from a calibration and then stored. In following sample runs, compound amounts are calculated by applying these response factors to the measured sample responses. Make sure that the sample injection size is reproducible from run to run, since there is no standard in the sample to correct for variations in injection size or sample preparation.

When preparing an ESTD analysis, the calculation of the amount of a particular compound in an unknown sample occurs in two steps:

1. An equation for the curve through the calibration points for this compound is calculated using the type of fit specified in the **Mode** and **Origin** settings in the compound table.
2. The amount of the compound in the unknown is calculated using the equation described above. This amount may appear in the report or it may be used in additional calculations called for by sample multiplier, compound multiplier, or dilution factor values before being reported.

Single-Level Calibration

In case of single-level calibration, the response factor is simply the ratio of the calibration point response and amount. If **Include origin** and **Force origin** are switched off, a warning is emitted.

The response factor RF is defined as a ratio of response and amount or vice versa (see "RF definition" on page 100). To calculate the RF, the application uses the compound amount of the calibration sample and the corresponding response of the calibration sample.

The formula for the single-level calibration calculation of the ESTD results depends on the type of response that you have set in the processing method:

Amount = Peak Area / RF

or:

Amount = Peak Height / RF

where

Amount	Amount of the compound
RF	Response factor

For details on the calculation of concentrations, see "Concentration" on page 128.

Multi-Level Calibration

For multi-level calibration, the response factor is evaluated from the calibration curve.

ISTD Calculation

The ISTD procedure eliminates the disadvantages of the ESTD method by adding a known amount of a compound which serves as a normalizing factor. This compound, the *internal standard*, is added to both calibration and unknown samples.

The compound used as an internal standard should be similar to the calibrated compound, both chemically and in retention/migration time, but it must be chromatographically distinguishable.

Table 8 ISTD procedure

Advantages	Disadvantages
Sample-size variation is not critical. Instrument drift is compensated by the internal standard. The effects of sample preparations are minimized if the chemical behavior of the ISTD and unknown are similar.	The internal standard must be added to every sample.

If the ISTD procedure is used for calibrations with a non-linear characteristic, care must be taken that errors which result from the calculation principle do not cause systematic errors. In multi-level calibrations, the amount of the ISTD compound should be kept constant, i.e. the same for all levels.

In the internal standard analysis, the amount of the compound of interest is related to the amount of the internal standard component by the ratio of the responses of the two peaks.

OpenLAB CDS allows up to 5 ISTD compounds.

For the ISTD calculation relative responses and relative amounts are used instead of the "raw" responses and amounts. They are calculated by dividing the response and amount of the peak of interest by the response and amount of the corresponding ISTD compound:

$$\text{Relative Response} = \text{Response} / \text{Response}_{\text{ISTD}}$$

$$\text{Relative Amount} = \text{Amount} / \text{Amount}_{\text{ISTD}}$$

The response can be Area, Area%, Height, or Height% (see “Response Type and Response Factor” on page 100).

In an ISTD calibration, the calculation of the corrected amount ratio of a particular compound in an unknown sample occurs in several stages. These stages are described in the following sections.

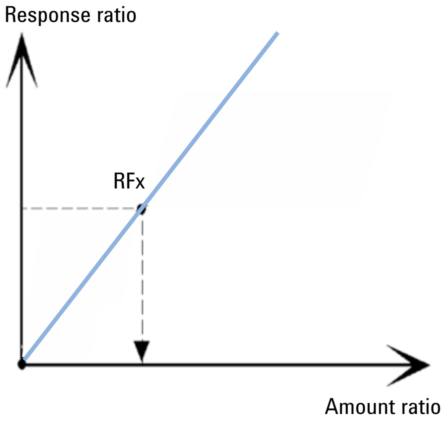
Calibration Samples

1 The calibration points are constructed by calculating an amount ratio and a response ratio for each level of a particular compound in the compound table.

The amount ratio is the amount of the compound divided by the amount of the internal standard at this level.

The response ratio is the response (area or height) of the compound divided by the response of the internal standard at this level.

2 An equation for the curve through the calibration points is calculated using the type of curve model specified in the compound table of the processing method.



Unknown Sample

1 The response of the compound in the unknown sample is divided by the response of the internal standard in the unknown sample to give a response ratio for the unknown.

2 An amount ratio for the unknown is calculated using the curve model equation determined in step 2 above, and the actual amount of ISTD in the sample.

Single Level ISTD Calibration

In case of single-level calibration, the relative response factor (RRF) is evaluated using response and amount values from calibration samples. Depending on the RF definition in the global calibration settings, one of the following formulas applies:

With RF defined as **Response per amount**:

$$RRF = \frac{\text{Rel Response}}{\text{Rel Amount}}$$

With RF defined as **Amount per Response**:

$$RRF = \frac{\text{Rel Amount}}{\text{Rel Response}}$$

The amount and concentration are calculated according to the following formulas, using the response value from the sample measurement:

With RF defined as **Response per amount**:

$$\text{Amount} = \left(\frac{\text{Rel Response}}{RRF}\right) * \text{Amount}_{\text{ISTD}}$$

With RF defined as **Amount per Response**:

$$\text{Amount} = RRF * \text{RelResponse} * \text{Amount}_{\text{ISTD}}$$

For multi-level calibration, the relative response factor is evaluated from the calibration curve.

For details on calculation of concentrations, see “Concentration” on page 128.

Quantitation of Uncalibrated Compounds

Uncalibrated compounds can be quantified either with a fixed response factor or using the calibration data of one of the calibrated compounds. Quantitation using a fixed response factor or calibrated compound data is signal-specific. In the latter case, if the calibrated compound is quantified by an ISTD method, the ISTD is used for the not identified peaks in the same way as for the calibrated compound.

Indirect Quantitation Using a Calibrated Compound

If the calibration data of a calibrated compound is to be used to quantify uncalibrated compounds, the calibrated compound is identified in the processing method (**Calibration** node, **Compound Table** tab: under **Mode**, select **Reference**). Calculations are the same as for calibrated compounds. If the reference compound is quantified by an ISTD method, the ISTD is used for the uncalibrated compound in the same way as for the reference compound.

A missing reference peak results in a zero amount of the uncalibrated compound.

Optionally, a correction factor (**Ref. correction**) can be entered to multiply the response of the peak before the amount is calculated from the response factor of the reference compound.

Quantitation Using a Manual Factor

The software allows you to quantify an identified compound that is based on a fixed response factor (**Manual Factor** column). In this case, the compound amount is calculated using the fixed response factor:

Amount = Response * Manual Factor

where

Manual Factor	Fixed response factor
Response	Response can be Area, Area%, Height, Height%, Rel. Area, or Rel. Amount (see "Response Type and Response Factor" on page 100)

For details on calculation of concentrations, see "Concentration" on page 128.

Using a manual factor with an ISTD method

If the compound amount is quantified using the fixed response factor and ISTD, the formula is read as follows:

$$\text{Relative area} = \text{Area} / \text{Area}_{\text{ISTD}}$$

or:

$$\text{Relative height} = \text{Height} / \text{Height}_{\text{ISTD}}$$

The amount is then calculated as follows:

$$\text{Amount} = \text{Relative Area} * \text{Manual Factor} * \text{Amount}_{\text{ISTD}}$$

or:

$$\text{Amount} = \text{Relative Height} * \text{Manual Factor} * \text{Amount}_{\text{ISTD}}$$

For details on calculation of concentrations, see "Concentration" on page 128.

Dependency of manual factor and response factor (RF)

With RF defined as **Response per amount** (default setting):

RF = 1 / Manual Factor

With RF defined as **Amount per response**:

RF = Manual Factor

For more information on the response factor, see “Response Type and Response Factor” on page 100.

Quantitation of Not Identified Peaks

Not identified peaks can be quantified using timed groups, either with a fixed response factor or with the calibration data of one of the calibrated compounds. Quantitation using a fixed response factor or calibrated compound data is signal-specific. In the latter case, if the calibrated compound is quantified by an ISTD method, the ISTD is used for the not identified peaks in the same way as for the calibrated compound.

For more information on timed groups, see “Definition of a timed group” on page 141.

Quantify Not Identified Peaks Using a Fixed Response Factor

In this case, you create a timed group with quantitation mode **Manual Factor**. The specified times ranges of the timed group include the relevant not identified peaks.

In addition, quantified peaks must be excluded. By setting the option **Quantify each peak individually**, amount and concentration of all not identified peaks are calculated using the fixed response factor.

For details on the calculation, see “Quantify a timed group with a manual factor” on page 143.

Quantify Not Identified Peaks Using a Calibrated Compound

In this case, you create a timed group with quantitation mode **Reference**. The specified times ranges of the timed group include the relevant not identified peaks. Optionally, a correction factor (**Ref. correction**) can be entered to multiply the response of the peak before the amount is calculated from the response factor of the reference compound.

In addition, quantified peaks must be excluded. By specifying the option **Quantify each peak individually**, amount and concentration of all not identified peaks are calculated using the curve reference.

Norm% Calculation

You can choose to normalize amounts in the general calibration settings of a processing method.

The Norm% analysis has the same disadvantage as the Area% and Height% calculations. Any changes that affect the total peak area will affect the concentration calculation of each individual peak. The normalization analysis should only be used if all components of interest are eluted and integrated. Excluding selected peaks from a normalization analysis will change the reported results in the sample.

The equation used to calculate the **Norm%** of a compound is:

$$\text{Compound Norm Amount} = \text{Compound Amount} * \frac{\text{Normalization}}{\text{Total Amount}}$$

where

Compound Amount	Amount of compound
Compound Norm Amount	Amount of normalization compound
Normalization	Normalization factor
Total Amount	Sum of all compound amounts and timed group amounts Named group amounts are not included in the total amount. Amounts of identified compounds in timed groups are counted twice if you have enabled **Include identified peaks** for the timed group.

$$\text{Group Norm Amount} = \text{Group Amount} * \frac{\text{Normalization}}{\text{Total Amount}}$$

where

Group Amount	Amount of group
Group Norm Amount	Amount of normalization group
Normalization	Normalization factor

$$\text{Peak Norm Amount} = \text{Peak Amount} * \frac{\text{Normalization}}{\text{Total Amount}}$$

where

Normalization	Normalization factor
Peak Amount	Amount of peak
Peak Norm Amount	Amount of normalization peak

For detailed information on timed groups, see "Definition of a timed group" on page 141.

If individual peaks are calculated in timed groups, these individual peaks are not included a second time in the total amount; they are already included in the timed group amount.

The default of normalization is 100 to create Norm% results. However, you can set a different number in the method. The total amount is the sum of all calculated compound and timed group amounts, independent of the signal of the compound main peak.

You can select whether ISTD compounds are included in the calculation. If excluded (default setting), the ISTD amount is not added to the total amount, and no compound normalization amounts are calculated for the ISTDs.

For named groups the group amount is not included in the total amount.

Quantitation of groups

Definition of a timed group

A timed group contains one or more time regions and is defined on a specific signal. First, the area and the height of the group are calculated by summing the area and height of all peaks included in the time regions (including or excluding identified peaks, depending on the group parameters). The expected retention time of the group is for sorting purposes only and can be entered manually.

The timed group corresponds to the *uncalibrated range* or *calibrated range* in OpenLAB CDS EZChrom Edition.

The following example shows three timed groups, where Group 2 and Group 3 overlap. C1 and C2 are identified compounds. The unidentified peak at 5.689 min is evaluated in both groups. Identified peaks are only evaluated if the group parameters are set accordingly.

Groups	Time ranges	Include identified peaks?
Group 1	0.8 min - 1.4 min 2.8 min - 3.4 min	No
Group 2	3.8 min - 5.9 min	Yes
Group 3	5.4 min - 7.2 min	No

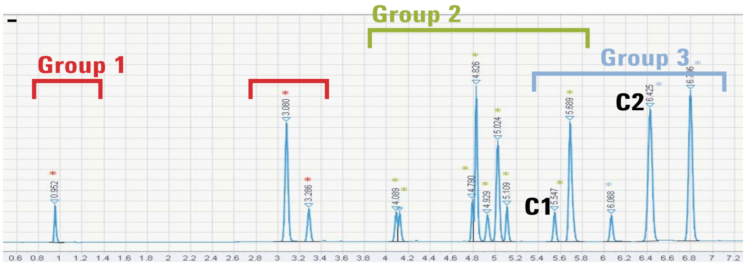


Figure 85 Example: Timed groups

NOTE If a timed group has no time region defined, its area, height, and amount are not calculated. If a timed group has a region or regions defined, but no peak is found in this region or these regions, its area and height are equal to zero.

A timed group is calibrated and quantified using the calibration parameters for the group. Timed groups support all calibration and quantitation modes of regular compounds (**Curve**, **Manual Factor**, **Reference**). In the group parameters, you can choose to quantify all peaks individually. In this case, each peak of the group is quantified individually with the group response factor (RF).

If you correct the retention times by using time reference compounds, the start time and stop time of the timed group are also corrected by the corresponding shifts (see "Calculations for Time Reference Compounds" on page 89).

Conflicts may occur if a peak belongs to several groups, or if identified peaks are quantified as part of the group. In these cases, the following rules apply:

- If an unknown peak belongs to several groups, it is quantified with the response factor of the group with the smallest retention time.
- If an identified peak is quantified as part of the group, but has its own calibration parameters defined, it is quantified with its own response factor and not with the group response factor.
- If an identified peak is quantified as part of the group and has no specific calibration parameters defined, the compound is quantified with the group response factor. The response type (area or height) of the group is used.

Quantify a timed group

Quantify a timed group with a manual factor

In this case the amount of the group is calculated according to a fixed response factor entered manually.

ESTD:

$$\text{Group Amount} = \text{Group Response} * \text{Manual Factor}$$

where

Manual Factor	Fixed response factor
Group Response	Sum of all responses

or ISTD:

$$\text{Group Amount} = \frac{\text{Group Response}}{\text{Response}_{\text{ISTD}}} * \text{Manual Factor} * \text{Amount}_{\text{ISTD}}$$

where

Amount ISTD	Amount of the internal standard
Manual Factor	Fixed response factor
Response ISTD	Response of the internal standard

$$\text{Group Concentration} = \text{Group Amount} * \text{Multipliers} * \text{Dilution Factors}$$

or

$$\text{Group Concentration} = \text{Group Amount} * \text{Multipliers} / \text{Dilution Factors}$$

For more information on how to calculate the multipliers and dilution factors, see “Correction Factors” on page 127. For more information on how to calculate the concentration, see “Concentration” on page 128.

Quantify a timed group with its own calibration curve

A timed group can be quantified according to its own calibration curve. All calibration options or levels are supported. In the ISTD mode, you must select an ISTD to use.

Quantify a timed group with the curve of a reference compound

The timed group can be quantified according to the calibration curve of another single compound. The software allows you to use the response factor of a reference compound (calibration curve). In this case a correction factor (**Ref. correction**) can be entered to multiply the response before the amount is calculated from the response factor of the reference compound. In case of ISTD same ISTD is used as for reference compound.

Quantify peaks individually in a timed group

If you choose to quantify all peaks individually, the individual peak amount is calculated as:

$$\text{Peak Amount} = \text{Group Amount} * \frac{\text{Peak Response}}{\text{Group Response}}$$

$$\text{Peak Concentration} = \text{Peak Amount} * \text{Multipliers} * \text{Dilution Factors}$$

or

$$\text{Peak Concentration} = \text{Peak Amount} * \text{Multipliers} / \text{Dilution Factors}$$

For more information on how to calculate the multipliers and dilution factors, see “Correction Factors” on page 127. For more information on how to calculate the concentration, see “Concentration” on page 128.

Definition of a named group

The named group consists of user-selected compounds and timed groups. Each compound or timed group is identified and quantified on its own. ESTD and ISTD calculations are based on the calibration data of the individual compounds. The calculated group area, height, amount, and concentration of the group are the sum of the individual areas, heights, amounts, and concentrations. The named group itself is not calibrated. One compound can be in multiple named groups.

The expected retention time of the group is only for sorting purposes and can be entered manually.

The named group corresponds to the *Named Peaks group* in OpenLAB CDS EZChrom Edition.

The following example shows two named groups where one compound is contained in both groups:

Groups	Group 1	Group 2
Included compounds	C1	
	C2	
		C3
	C4	C4
		C5
		C6

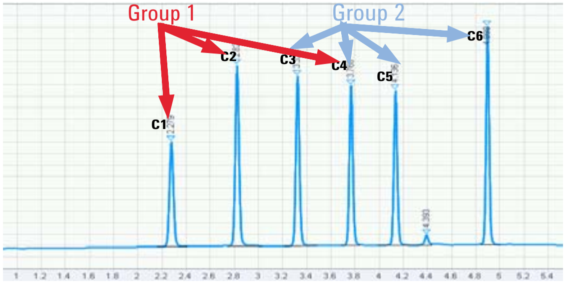


Figure 86 Example: Named groups

Quantify a named group

The results in the named group table are:

Group Area = Σ Compound Peak Areas + Σ Timed group areas

Group Height = Σ Compound Peak Heights + Σ Timed group heights

Group Amount = Σ Compound Amounts + Σ Timed group amounts

Group Concentration = Σ Compound Concentrations + Σ Timed group concentrations

If none of the compounds or timed groups in the named group have been identified, the named group will appear as "not identified" on the analysis.

7

UV Spectral Analysis

This chapter describes the concepts of the impurity check and the confirmation of compound identity based on UV spectral analysis.

What is UV spectral analysis?

There are different windows and functions specific to UV spectral analysis. To view those windows and access the functions, the focused injection must contain spectral data (for example, acquired with a 3D UV system).

UV spectral analysis provides additional quality criteria for routine analytics:

- *Confirm the compound identity*

  The application compares a UV spectrum with a specific UV reference spectrum. A high match factor indicates that the compounds are probably identical.

  For details on the calculation, see UV confirmation ("UV confirmation" on page 156).

- *Check for UV impurities*

  The application compares all UV spectra of a peak with the apex spectrum. It calculates an overall match factor, the UV Purity value. A low UV Purity value indicates that there are co-eluted peaks with a significantly different UV spectrum.

  For details on the calculation, see UV impurity check ("UV impurity check" on page 150).

UV Spectral analysis processes spectral data acquired from a UV-visible diode-array detector or fluorescence detector. It adds a third dimension to your analytical data when using it with the chromatographic data (see Figure 87 on page 149).

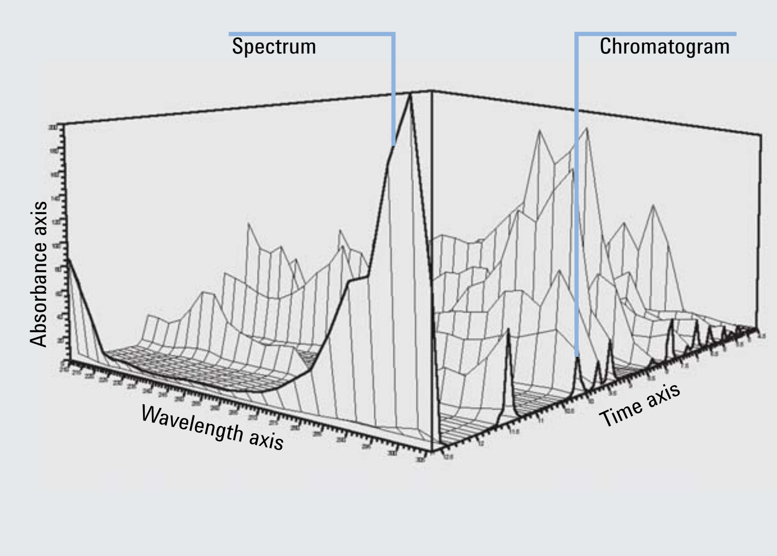


Figure 87 Spectral Information

UV impurity check

An impurity check assesses whether a peak is pure or contains impurities. This assessment is based on the comparison of spectra recorded during the elution of the peak. After applying a baseline correction, the spectrum at the peak apex is compared with all significant spectra recorded in the peak. The application calculates a match factor that characterizes the degree of similarity of the spectra.

The application performs the following steps to evaluate UV impurities:

1. Per peak
   - a "Noise calculation" on page 150
   - b Determine significant spectra("Determine significant spectra" on page 151)
2. Per spectrum:
   - a "Background correction" on page 152
   - b "Similarity calculation" on page 152
   - c Threshold calculation("Threshold and sensitivity" on page 153)
3. "Impurity evaluation" on page 154

Noise calculation

As a preparation for further evaluations, the application calculates the following numbers for each peak from the spectra at baseline start and baseline end:

- Noise variance
- Noise standard deviation σ

Baseline start and end times depend on the integration. If multiple peaks are only separated by a drop line, all peaks use the same spectra for noise calculation.

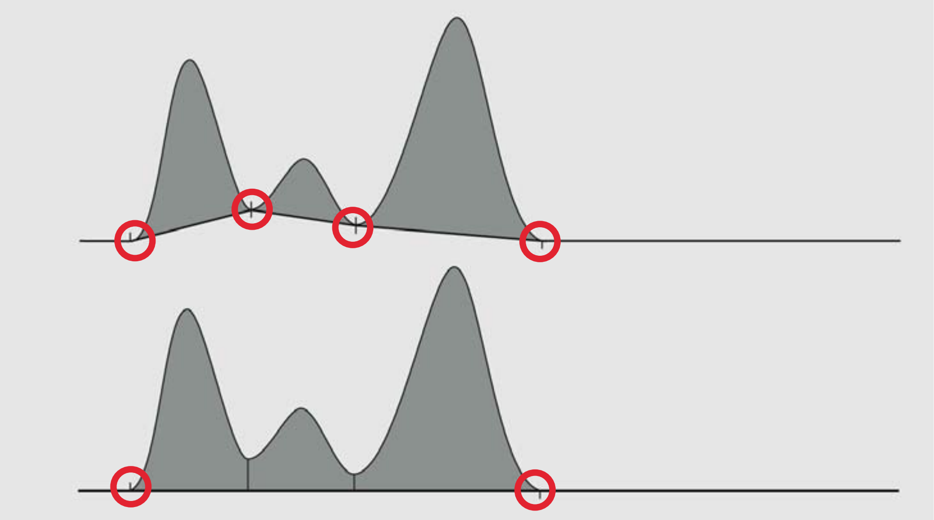


Figure 88 Baseline start and end depending on integration

Determine significant spectra

To ensure that only spectra with a significant signal are evaluated, the application filters out spectra where the response range is too small. Spectra are used for further calculations only if the following applies:

- Response range is larger than 3σ
- Response range is larger than or equal to 10 % of the apex spectrum response range. The response range for each spectrum is calculated as max - min response.

Background correction

For the baseline correction, the application evaluates the following spectra:

- Spectrum at the baseline start of the peak
- Spectrum at the baseline end of the peak

Baseline start and end times depend on the integration. If multiple peaks are only separated by a drop line, all peaks use the same spectra for background correction (see Figure 88 on page 151).

A linear interpolation of the two baseline spectra is calculated. To correct each individual peak spectrum, the application subtracts the interpolation spectrum at the corresponding retention time.

Similarity calculation

The application compares each of the remaining background-corrected peak spectra with the background-corrected apex spectrum. The match factor for each spectrum is a similarity value between 0 (no similarity) and 1000 (identical spectra).

Correlation

$$r = \frac{\sum_{i=1}^{n} [(x_i - x_{av}) \cdot (y_i - y_{av})]}{\sqrt{\left[\sum_{i=1}^{n} (x_i - x_{av})^2 \cdot \sum_{i=1}^{n} (y_i - y_{av})^2\right]}}$$

where

r	Correlation
x_i, y_i	Measured absorbances in the first and second spectrum respectively at the same wavelength
n	Number of data points
x_{av}, y_{av}	Average absorbance of the first and second spectrum respectively

Similarity

$$\text{Match factor} = r^2 * 1000$$

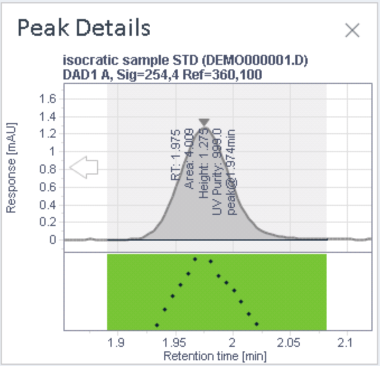
Threshold and sensitivity

The application automatically calculates a *threshold* value for each spectrum of a peak, based on noise and variance of both spectra to be compared. The threshold value determines if a deviation from the apex spectrum is low enough to consider a spectrum as pure. A spectrum is classified as pure if the match factor is greater than or equal to the threshold value.

You influence the threshold value by changing the *sensitivity* value (default: 50 %):

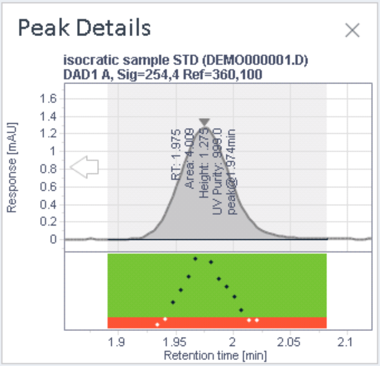
- A sensitivity value lower than 50 % will cause the threshold value to decrease. Thus a lower match factor will be sufficient to classify a spectrum as pure. 0 % means that a spectrum is considered pure even if there is no similarity at all.
- A sensitivity valued higher than 50 % will cause the threshold value to increase, thus a higher match factor will be required to classify a spectrum to be pure. 100 % means that the spectrum must be absolutely identical to the apex spectrum. Due to noise this is almost impossible.

For example, the following figures show the same peak with different sensitivity settings:


Peak Details

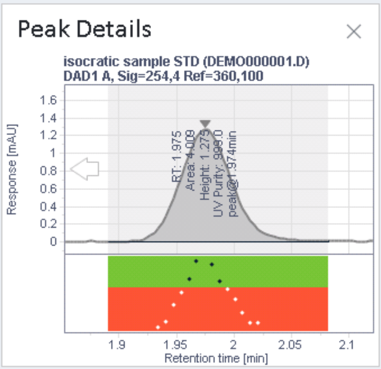
isocratic sample STD (DEMO000001.D)
DAD1 A, Sig=254,4 Ref=360,100

Low sensitivity - this peak is considered pure


Peak Details

isocratic sample STD (DEMO000001.D)
DAD1 A, Sig=254,4 Ref=360,100

Default sensitivity - this peak is considered impure


Peak Details

isocratic sample STD (DEMO000001.D)
DAD1 A, Sig=254,4 Ref=360,100

High sensitivity - this peak is considered impure

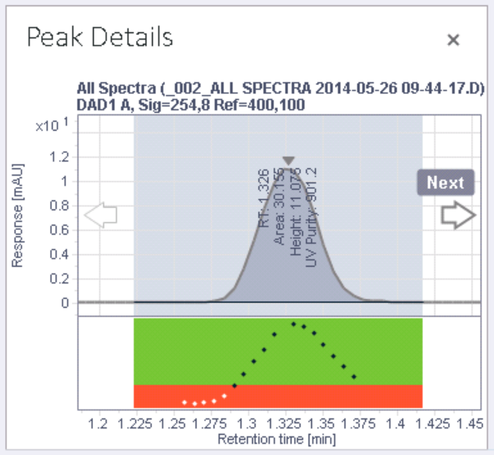
The sensitivity value does not influence the match factors themselves. It only changes the threshold value and thus the criteria for a green or red coloring of the purity ratio curve and of the resulting UV Purity value. The shape of the purity ratio curve may also change, as the threshold value is part of the purity ratio calculation (for details, see “Impurity evaluation” on page 154).

The processing method allows to set the sensitivity individually for each identified compound (**Compounds > Spectra > Compound Table** tab), and generally for non-identified peaks (**Compounds > Spectra > UV Impurity Check** tab).

Impurity evaluation

The purity ratios for the single spectra of a peak are shown in the *purity ratio curve*. You can view the purity ratio curve in the **Peak Details** window. The purity ratios are calculated for each spectrum from the sensitivity-corrected threshold value and the match factor (“Threshold and sensitivity” on page 153, “Similarity calculation” on page 152).

$$\text{PurityRatio} = \log\left(\frac{1000 - \text{Threshold}}{1000 - \text{Match factor}}\right)$$



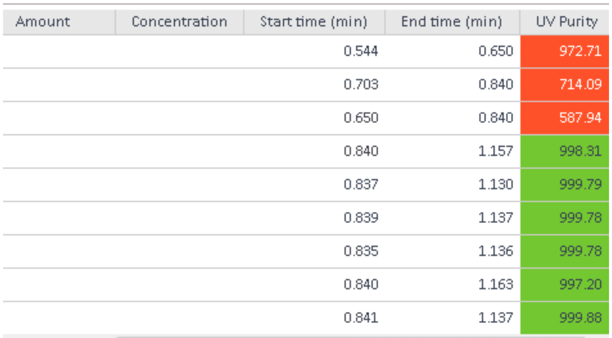
All Spectra (_002_ALL SPECTRA 2014-05-26 09-44-17.D)
DAD1 A, Sig=254,8 Ref=400,100

Figure 89 Purity ratio curve

The colored background indicates the purity of a spectrum. It is considered pure (green background) if the purity ratio is greater than zero, that is, if the match factor is greater than the threshold value.

The *entire peak* is considered pure if all single spectra are pure. As soon as one single spectrum is impure, the entire peak is considered impure, and the UV purity value is colored red in the injection results.

The UV purity value that is displayed in the **Injection results** table is calculated as an average of all match factors for the peak.



Amount	Concentration	Start time (min)	End time (min)	UV Purity
		0.544	0.650	972.71
		0.703	0.840	714.09
		0.650	0.840	587.94
		0.840	1.157	998.31
		0.837	1.130	999.79
		0.839	1.137	999.78
		0.835	1.136	999.78
		0.840	1.163	997.20
		0.841	1.137	999.88

Figure 90 UV Purity value in the injection results

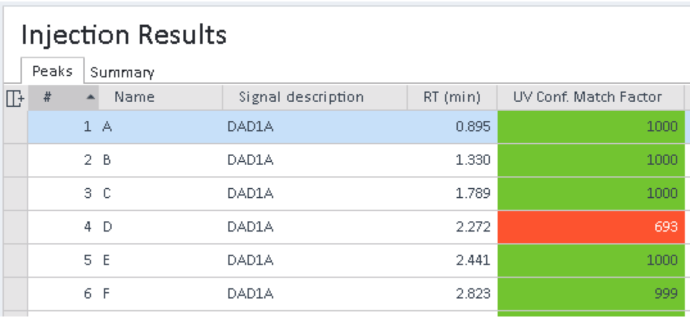
You can influence the threshold value and thus the criteria for a green or red coloring by changing the sensitivity value ("Threshold and sensitivity" on page 153).

UV confirmation

UV reference spectra are acquired from a reference sample under well-defined chromatographic conditions. You can confirm the identity of a compound by comparing the current spectrum at the peak apex with the UV reference spectrum. The application calculates a match factor for the two spectra.

The algorithm for the comparison is identical to the one used for the UV impurity check (see “Similarity calculation” on page 152). Background correction is optional, it can be selected in the processing method.

For UV confirmation, the match factor must be greater than a given limit to be colored green in the injection results. You set the match factor limit in the processing method. The resulting UV confirmation match factor is shown in the injection results.



#	Name	Signal description	RT (min)	UV Conf. Match Factor
1	A	DAD1A	0.895	1000
2	B	DAD1A	1.330	1000
3	C	DAD1A	1.789	1000
4	D	DAD1A	2.272	693
5	E	DAD1A	2.441	1000
6	F	DAD1A	2.823	999

Figure 91 UV confirmation match factor in injection results

8

Mass Spectrometry

This chapter describes the sample purity calculation based on mass spectrometry.

MS sample purity calculation

The MS sample purity calculation assesses whether a sample is pure or contains impurities. This assessment is based on the comparison of responses. On the one hand, there is the response of all compounds and fragments in a sample. On the other hand, there is the response caused by specific target ions. The sample purity is calculated as the ratio of both responses.

The application performs different steps to calculate the MS sample purity, depending on the selected base signal and calculation:

Target found?

1. Get the target masses given in the **Target** columns of the **Injection List** (for example, 270). If a formula is entered, calculate the molecular weight from the formula.
2. Apply the adducts specified in the processing method (for example, +H and +Na and a target mass of 270 would result in the targets 271 and 293).
3. Extract EICs for all targets, and sum these EICs to a single EIC.
4. Determine the retention time of the peak in that single summed EIC.
5. Locate the matching peak in the chromatogram of the base signal.

If a matching peak can be located, the target is marked as **found**.

Base signal is from an MS detector

With calculation **TIC %**

$$\text{MS sample purity} = \frac{\text{area or height of matching peak (TIC)}}{\text{area or height of all integrated peaks (TIC)}} * 100$$

With calculation **EIC/TIC %**:

$$\text{MS sample purity} = \frac{\text{area or height of single peak (summed EIC)}}{\text{area or height of all integrated peaks (TIC)}} * 100$$

NOTE When working with multisignal data, the summed TIC (combination of all TIC signals in injection) is used.

Base signal is from another detector (non-MS)

$$\text{sample purity} = \frac{\text{area or height of matching peak (base signal)}}{\text{area or height of all integrated peaks (base signal)}} * 100$$

Assumptions

MS sample purity is calculated under the following assumptions:

- The MS sample purity calculation is intended as a rough approximation only.
- For **EIC/TIC %** calculations: MS data is acquired such that most ion abundance is in the molecular ion cluster. There is only a small degree of in-source dissociation.
- For base signals from non-MS detectors: The other detector is more uniform and universal in its response than the MS detector.
- All compounds in the sample have uniform response factors.

9

System Suitability

This chapter describes what OpenLAB CDS can do to evaluate the performance of both the analytical instrument and the analytical method.

Peak Performance can be calculated for any integrated peak of the data loaded, and also for new manually integrated peaks. OpenLAB CDS calculates peak characteristics and displays them in the peak results.

Evaluating System Suitability

Evaluating the performance of both the analytical instrument before it is used for sample analysis and the analytical method before it is used routinely is good analytical practice. It is also a good idea to check the performance of analysis systems before, and during, routine analysis. OpenLAB provides the tools to do these three types of tests automatically. An *instrument test* can include the detector sensitivity, the precision of peak retention times and the precision of peak areas. A *method test* can include precision of retention times and amounts, the selectivity, and the robustness of the method to day-to-day variance in operation. A *system test* can include precision of amounts, resolution between two specific peaks and peak tailing.

Laboratories may have to comply with:

- Good Laboratory Practice regulations (GLP),
- Good Manufacturing Practice regulations (GMP) and Current Good Manufacturing Practice regulations (cGMP), and
- Good Automated Laboratory Practice (GALP).

Laboratories are advised to perform these tests and to document the results thoroughly. Laboratories which are part of a quality control system, for example, to comply with ISO9000 certification, will have to demonstrate the proper performance of their instruments.

To collate the results from several runs and evaluate them statistically, OpenLAB CDS offers a function to create result set summary reports. Different report templates are available for these summaries (for example, SequenceSummary_Extended.rdl). They can be adjusted as required.

The tests are documented in a format which is generally accepted by regulatory authorities and independent auditors. Statistics include:

- peak retention time,
- peak area,
- amount,
- peak height,
- peak width at specific height,
- peak symmetry,

- peak tailing,
- capacity factor (k´),
- plate numbers,
- resolution between peaks, and
- selectivity relative to preceding peak.

Extended performance results are calculated only for calibrated compounds, ensuring characterization by retention times and compound names.

A typical system performance test report contains the following performance results:

- column details,
- processing method,
- sample information,
- acquisition information,
- signal description and baseline noise determination, and
- signal labeled with either retention times, or compound names.

In addition, the following information is generated for each calibrated compound in the chromatogram:

- retention/migration time,
- k´,
- symmetry,
- peak width,
- plate number,
- resolution,
- signal-to-noise ratio, and
- compound name.

Noise Determination

Noise can be determined from the data point values from the time range of the current signal. Noise is treated in the following ways:

- as six times the standard deviation (sd) of the linear regression of the drift
- as peak-to-peak (drift corrected)
- as determined by the ASTM method (ASTM E 685-93)
- as the Root Mean Square (RMS) of the linear regression of the drift

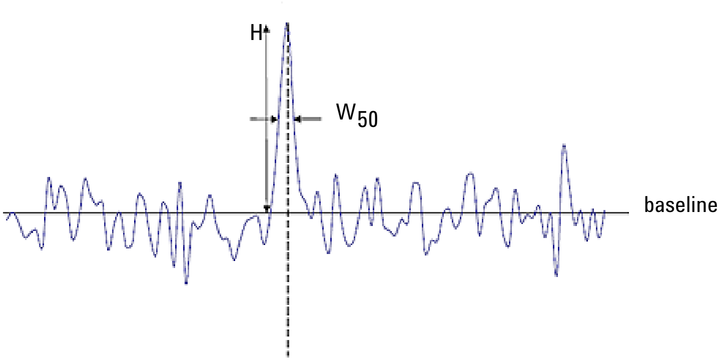
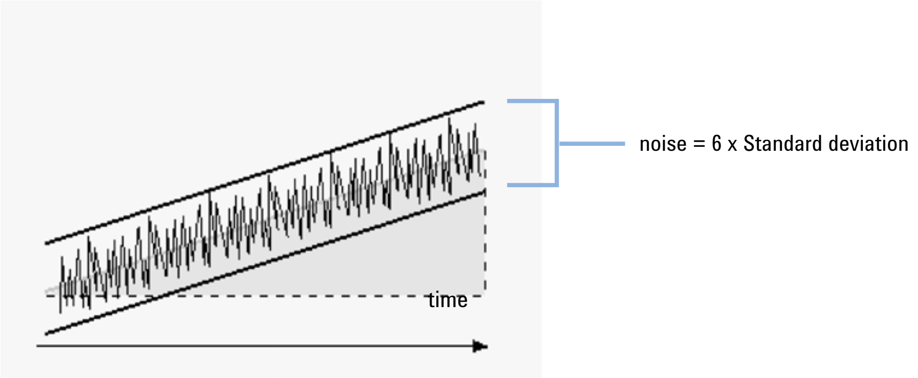


Figure 92 Chromatogram with peak signal and noise

H	Peak height from top to baseline (best straight line through noise)
W_{50}	Peak width at half height

Noise Calculation Using Six Times the Standard Deviation



The linear regression is calculated using all the data points within the time range of the current signal. The noise is given by the formula:

$$N = 6 \times Std$$

where

N	Noise based on the six time standard deviation method
Std	Standard deviation of the linear regression of all data points in the selected time range

Noise Calculation Using the Peak-to-Peak Formula

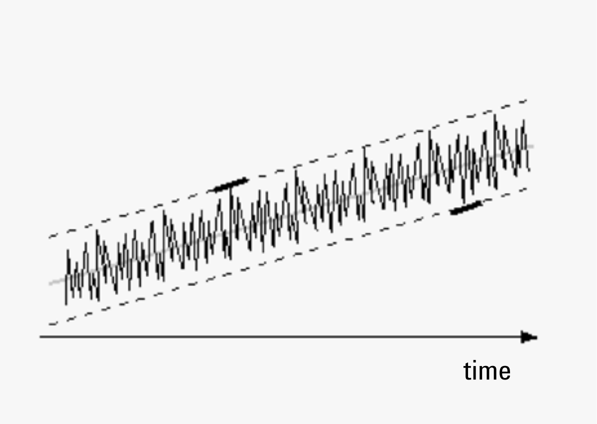


Figure 93 Illustration of peak-to-peak noise with drift

The drift is first calculated by determining the linear regression using all the data points in the time range of a peak. The linear regression line is subtracted from all data points within the time range to give the drift-corrected signal.

The peak-to-peak noise is then calculated using the formula:

$$N = I_{max} - I_{min}$$

where

N	Peak-to-peak noise
I_{max}	Highest (maximum) I_x value in the time range
I_{min}	Lowest (minimum) I_x value in the time range
I_x	Intensity of the signal, corrected by the drift (drift is calculated using the LSQ formula)

For European Pharmacopoeia calculations the Peak-to-Peak noise is calculated using a blank reference signal over a range of -10 and +10 times W_{50} flanking each peak. This region can be symmetrical to the signal of interest, or asymmetrical if required due to matrix signals.

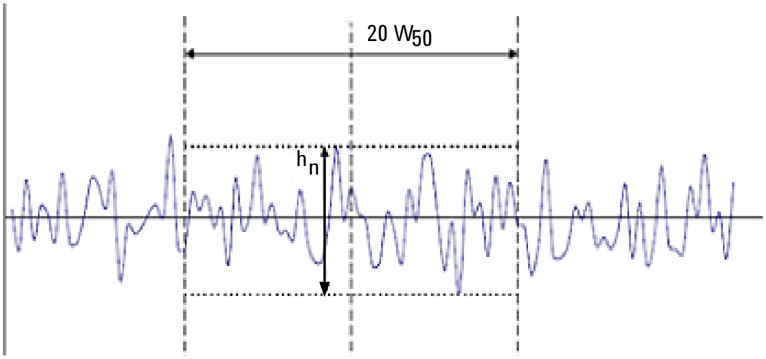


Figure 94 Determination of noise from the chromatogram of a blank sample

Where

20 W_{50} is the region corresponding to the 20 fold of W_{50}.

h_n is the maximum amplitude of the baseline noise in the 20-fold W_{50} region.

Noise Calculation by the ASTM Method

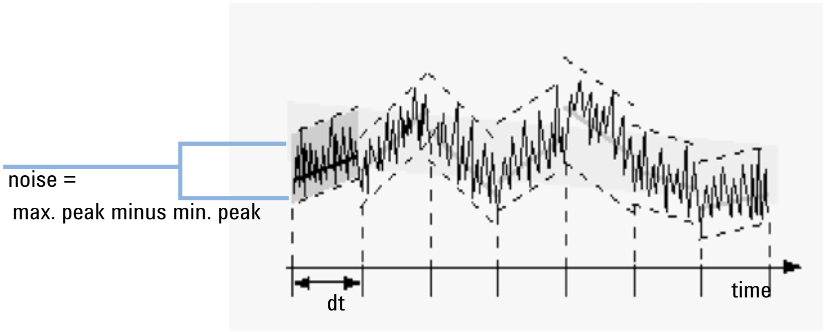


Figure 95 Noise determination by the ASTM method

ASTM noise determination (ASTM E 685-93) is based on the standard practice for testing variable-wavelength photometric detectors used in liquid chromatography, as defined by the American Society for Testing and Materials. Based on the size of the time range, three different types of noise can be distinguished. Noise determination is based on peak-to-peak measurement within defined time ranges.

- *Cycle Time, t*

  Long-term noise, the maximum amplitude for all random variations of the detector signal of frequencies between 6 and 60 cycles per hour. Long-term noise is determined when the selected time range exceeds one hour. The time range for each cycle (dt) is set to 10 minutes which will give at least six cycles within the selected time range.

  Short-term noise, the maximum amplitude for all random variations of the detector signal of a frequency greater than one cycle per minute. Short-term noise is determined for a selected time range between 10 and 60 minutes. The time range for each cycle (dt) is set to one minute which will give at least 10 cycles within the selected time range.

  Very-short-term noise (not part of ASTM E 685-93), this term is introduced to describe the maximum amplitude for all random variations of the detector signal of a frequency greater than one cycle per 0.1 minute.

  Very-short-term noise is determined for a selected time range between 1 and 10 minutes. The time range for each cycle (dt) is set to 0.1 minute which will give at least 10 cycles within the selected time range.

- *Number of Cycles, n*

  The number of cycles is calculated as:

  $$n = \frac{t_{tot}}{t}$$

  where t is the cycle time and t_{tot} is the total time over which the noise is calculated.

- *Peak-to-Peak Noise in Each Cycle*

  The drift is first calculated by determining the linear regression using all the data points in the time range. The linear regression line is subtracted from all data points within the time range to give the drift-corrected signal. The peak-to-peak noise is then calculated using the formula:

  $$N = I_{max} - I_{min}$$

  where N is the peak-to-peak noise, I_{max} is the highest (maximum) intensity peak and I_{min} is the lowest (minimum) intensity peak in the time range.

- *ASTM Noise*

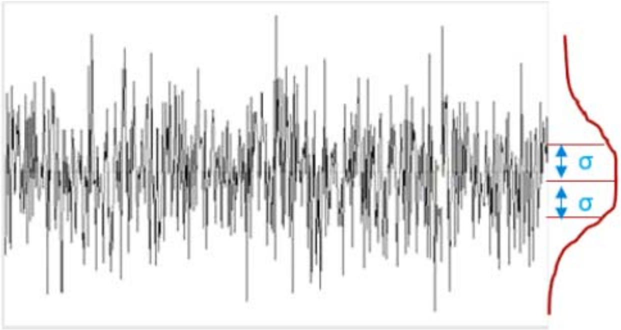
  The ASTM noise is calculated as:

  $$N_{ASTM} = \frac{\sum_{i=1}^{n} N}{n}$$

  where N_{ASTM} is the noise based on the ASTM method.

  An ASTM noise determination is not done if the selected time range is below one minute. Depending on the range, if the selected time range is greater than, or equal to one minute, noise is determined using one of the ASTM methods previously described. At least seven data points per cycle are used in the calculation. The cycles in the automated noise determination are overlapped by 10 %.

Noise calculation using the Root Mean Square (RMS)



The linear regression is calculated using all the data points within the time range of the current signal.

The noise is given by the formula:

$$RMS = S$$

where

RMS	Noise based ond standard deviation method
S	Standard Deviation

Standard deviation of the linear regression of all data points in the selected time range, with linear function y(X)= a + bX:

$$S = \sqrt{\frac{\sum_{i=1}^{N}(Y_i - a - bX_i)^2}{N-2}}$$

where

a	Y Intercept
b	slope
N	Number of discrete observations
X_i	Independet variable, ith observation

Signal-to-noise calculation

OpenLAB CDS has different options for the signal-to-noise calculation. You can choose both the algorithm and the noise range.

6 sigma or RMS method

The signal-to-noise is calculated using the formula:

$$\text{Signal-to-Noise} = \frac{\text{Height of the peak}}{\text{Noise of closest range}}$$

Peak to peak or ASTM method

The signal-to-noise is calculated using the formula:

$$S/N = 2H/h$$

where

H	Height of the peak corresponding to the component concerned in the chromatogram obtained with the prescribed reference solution.
h	Absolute value of the largest noise fluctuation from the baseline in a chromatogram obtained after injection or application of a blank and observed over a distance equal to twenty times the width at half- height of the peak in the chromatogram obtained with the prescribed reference solution, and situated equally around the place where this peak would be found. According to the definition of the European Pharmacopoeia, signal-to-noise is calculated against a blank reference signal and a noise calculated over the time range which contains the peak the S/N ratio is being calculated for.

Noise range

Noise can be calculated against the following time regions and signals:

- Fixed time region, on the same signal or on a blank reference signal
- Time region relative to the peak start or end, on the same signal or on a blank reference signal
- Automatically determined time region, on a blank reference signal.

An automatically determined time region is calculated according to one of the following algorithms:

- If the reference signal is not long enough

  $(EndTime - StartTime < 20*W_{50})$

  - *StartTime = starttime (of reference signal)*, and
  - *EndTime = endtime (of the reference signal)*

- If the reference signal is long enough, but the peak is situated too close to the starttime

  (t_R-10*W_{50} < *starttime of the reference signal)*

  - *StartTime = starttime (of reference signal)*, and
  - *EndTime = StartTime + 20*W_{50}*
- If the reference signal is long enough, but the peak is situated too close to the endtime

  (t_R+10*W_{50} > *endtime of the reference signal)*

  - *EndTime = endtime (of the reference signal)*, and
  - *StartTime = EndTime - 20*W_{50}*
- If the reference signal is long enough, and the peak is situated far enough away from starttime and endtime of the reference signal

  (t_R-10*W_{50} > *starttime,* t_R+10*W_{50} < *endtime)*

  - *StartTime =* t_R *- 10*W_{50}, and
  - *EndTime =* t_R *+ 10*W_{50}

where

t_R is the retention time, and

W_{50} is the peak width at half height.

Calculation of peak asymmetry and symmetry

Asymmetry Peak asymmetry is calculated by comparing the peak half-widths at 10% of the peak height:

$$A_S = \frac{W_{10}}{2\,W_{f,\,10}}$$

where

A_S	Asymmetry 10%
W_{10}	Peak width at 10% of the peak height
$W_{f,\,10}$	Front half of the peak width at 10% of the peak height.

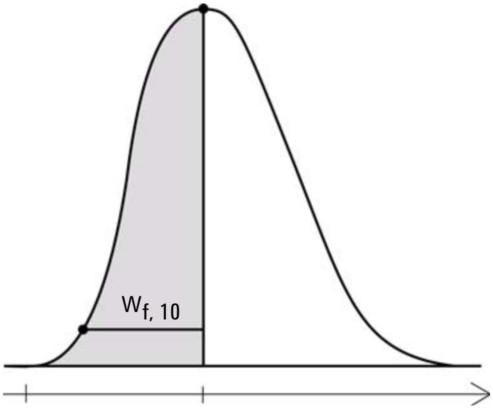


Figure 96 Calculation of peak asymmetry

Symmetry By most of the pharmacopeias, the symmetry factor of a peak is calculated by comparing the peak half-widths at 5%. In OpenLAB, this factor is calculated and stored as the Tailing factor (see "Tailing Factor (USP) t" on page 180). In OpenLAB, the symmetry is calculated as a pseudomoment by the integrator using the following moment equations:

$$m_1 = a_1\left(t_2 + \frac{a_1}{1.5\,H_f}\right)$$

$$m_2 = \frac{a_2^2}{0.5H_f + 1.5H}$$

$$m_3 = \frac{a_3^2}{0.5H_r + 1.5H}$$

$$m_4 = a_4\left(t_3 + \frac{a_4}{1.5H_r}\right)$$

$$\text{Peak symmetry} = \sqrt{\frac{m_1 + m_2}{m_3 + m_4}}$$

If no inflection points are found, or only one inflection point is reported, then the peak symmetry is calculated as follows:

$$\text{Peak symmetry} = \frac{a_1 + a_2}{a_3 + a_4}$$

where

a_i	Area of slice
t_i	Time of slice
H_f	Height of front inflection point
H_r	Height of rear inflection point
H	Height at apex

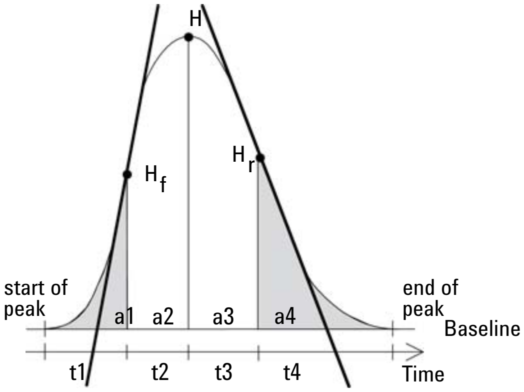


Figure 97 Calculation of the Peak Symmetry Factor

System Suitability Formulas and Calculations

The following formulas are used to obtain the results for the various System Suitability tests. The results are reported using the **Performance+Noise** and **Extended Performance** reports.

When ASTM or USP is specified for a given definition, then the definition conforms to those given in the corresponding reference. However, the symbols used here may not be the same as those used in the reference.

The references used in this context are:

- *ASTM: Section E 685 - 93, Annual Book of ASTM Standards, Vol.14.01*
- *USP: The United States Pharmacopeia, XX. Revision, pp. 943 - 946*
- *EP: European Pharmacopoeia, 7th Edition*
- *JP: Japanese Pharmacopoeia, 16th Edition*

Performance Test Definitions

Peak Performance can be calculated for any integrated peak of the data loaded, and also for new manually integrated peaks. OpenLAB CDS calculates peak characteristics and displays them in the peak results.

Overview Performance Tests

Table 9 Pharmacopoeia values in OpenLAB CDS

USP	EP	JP	Definition	Column name in injection results	Field used in reporting
Tailing factor	Symmetry factor or tailing factor	Symmetry factor	$S = \frac{W_5}{2f}$	Tailing	Peak_TailFactor
Relative retention	Relative retention	Separation Factor	$a = \frac{k'_1}{k'_2} = \frac{t_{R2} - t_0}{t_{R1} - t_0}$	Selectivity	Peak_Selectivity
-	Resolution	Resolution	$Rs = 1.18 \cdot \frac{t_{R2} - t_{R1}}{W_{50\,(1)} + W_{50\,(2)}}$	Resol.EP Resol.JP	Peak_Resolution_EP Peak_Resolution_JP
Resolution	-	-	$R = 2 \cdot \frac{t_{R2} - t_{R1}}{W_{t\,(2)} + W_{t\,(1)}}$	Resol.USP	Peak_Resolution_USP
Efficiency	-	-	$n = 16\left(\frac{t_R}{W_t}\right)^2$	Plates USP	Peak_TheoreticalPlates_USP
-	Efficiency	Efficiency	$n = 5.54\left(\frac{t_R}{W_{50}}\right)^2$	Plates EP Plates JP	Peak_TheoreticalPlates_EP Peak_TheoreticalPlates_JP
-	Unadjusted relative retention	-	$Rr = \frac{t_2}{t_1}$	RRT EP	Peak_RelativeRetTime_EP
-	S/N ratio	S/N ratio	With P2P or ASTM noise calculation: $\frac{S}{N} = \frac{2H}{h}$ With 6SD or RMS noise calculation: $\frac{S}{N} = \frac{H}{h}$	S/N	Peak_SignalToNoise

True Peak Width Wx [min]

W_x = width of peak at height x % of total

where

W_t	Tangent peak width, 4 sigma, obtained by intersecting tangents through the inflection points with the baseline
$W_{4.4}$	Width at 4.4% of height (5 sigma width)
W_5	Width at 5% of height (tailing peak width), used for USP tailing factor
W_{10}	Width at 10% of height
W_{50}	Width at 50% of height (true half-height peak width or 2.35 sigma).

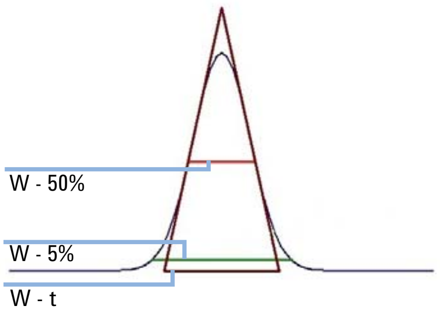


Figure 98 Peak width at x % of the height

Capacity Factor (USP), Capacity Ratio (ASTM) k'

$$k' = \frac{t_R - t_0}{t_0}$$

where

t_R	Retention time of peak [min]
t_0	Void time [min] (as provided in the processing method)

Tailing Factor (USP) t

NOTE Symmetry Factor (JP) and Symmetry factor (EP) are identical with the Tailing Factor (USP). All are available as "Peak_TailFactor" in Intelligent Reporting. See also Table 9 on page 177.

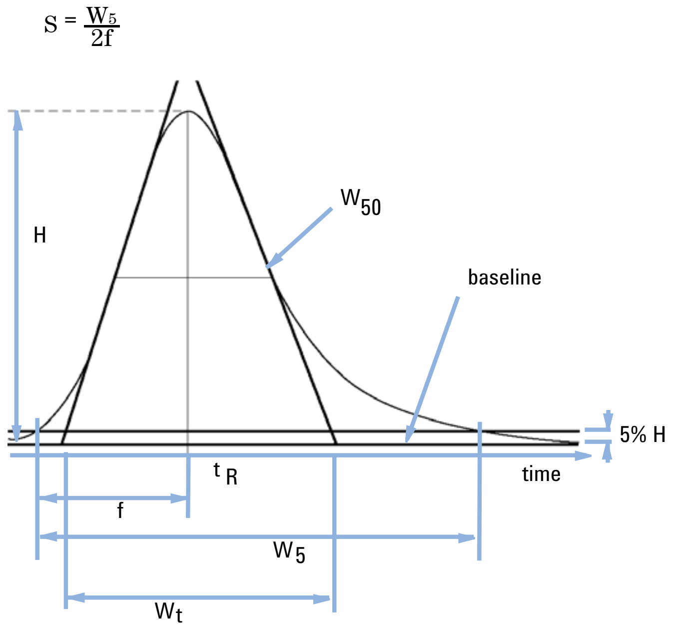


Figure 99 Performance Parameters

S	Tailing factor (USP)
H	Peak height
t_R	Retention time
f	Distance in min between peak front and t_R, measured at 5% of the peak height
W_{50}	Peak width at 50% of height [min]
W_5	Peak width at 5% of peak height [min]
W_t	Tangent peak width

Number of theoretical plates per column (USP)

Tangent method (USP, ASTM):

$$n = 16\left(\frac{t_R}{W_t}\right)^2$$

where

t_R	Retention time
W_t	Tangent width [min]

Half-width method (ASTM, EP, JP):

$$n = 5.54\left(\frac{t_R}{W_{50}}\right)^2$$

where

t_R	Retention time
W_{50}	Peak width at half-height [min]

Number of theoretical plates per meter N [1-m]

$$N = 100 \cdot \frac{n}{l}$$

where

n	Number of theoretical plates
l	Length of column [cm] (as provided in the processing method)

Relative Retention (USP, ASTM), Selectivity

(Pertaining to peaks a and b, t_R of peak a < t_R of peak b)

$$\text{alpha} = \frac{k'_{(b)}}{k'_{(a)}}, \text{alpha} \geq 1$$

where

$k'_{(x)}$	capacity factor for peak x: $(t_{Rx}-t_0)/t_0$

Resolution (USP, ASTM) R

Tangent method (pertaining to peaks 1 and 2, t_R of peak 1 < t_R of peak 2; t_R in min)

$$R = 2 \cdot \frac{t_{R2} - t_{R1}}{W_{t\,(2)} + W_{t\,(1)}}$$

where

t_R	Retention time
W_t	Tangent width [min]

Resolution (EP/JP) Rs

Resolution (JP) and Resolution (EP) are calculated with the half-width method (Resolution used in Performance Report):

$$Rs = 1.18 \cdot \frac{t_{R2} - t_{R1}}{W_{50\,(1)} + W_{50\,(2)}}$$

where

t_R	Retention time
W_{50}	Peak width at half-height [min]

www.agilent.com

In This Book

This guide contains the reference information on the principles of operation, calculations and data analysis algorithms used in Agilent OpenLAB CDS. The information contained herein may be used by validation professionals for planning and execution of system validation tasks.

- Integration with ChemStation algorithm
- Integration with EZChrom algorithm
- Peak Identification
- Calibration
- Quantitation
- UV Spectral Analysis
- Mass Spectrometry
- System Suitability

Printed in Germany
08/2016

M8410-90031

Agilent Technologies