Agilent OpenLAB CDS ChemStation Edition

Reference to Operation Principles

Agilent Technologies

Notices

Manual Part Number

M8301-90024

Edition

1/2013

Printed in Germany

Agilent Technologies
Hewlett-Packard-Strasse 8
76337 Waldbronn

This product may be used as a component of an in vitro diagnostic system if the system is registered with the appropriate authorities and complies with the relevant regulations. Otherwise, it is intended only for general laboratory use.

Software Revision

This guide is valid for revision C.01.0G or higher of the Agilent OpenLAB CDS ChemStation Edition.

Microsoft ® is a U.S. registered trademark of Microsoft Corporation.

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 ChemStation Edition. It contains reference information on the principles of operation, calculations and data analysis algorithms used in Agilent OpenLAB CDS ChemStation Edition.

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 concepts of OpenLAB CDS ChemStation Edition, new features, and workflows: The manual *OpenLAB CDS ChemStation Edition, Basic Concepts and Workflows*.
- For context-specific task ("How To") information, a tutorial, reference to the User Interface, and troubleshooting help: The ChemStation online help system.
- For details on system installation and site preparation: The *Agilent OpenLAB CDS Workstation Installation Guide* guide.
- For details on system administration principles and tasks: the *Agilent OpenLAB CDS Administration Guide*.

1 Data Acquisition

This chapter describes the concepts of Data Acquisition, data files, logbook, and more.

2 Integration

This chapter describes the concepts of integration the ChemStation integrator algorithms. It describes the integration algorithm, integration and manual integration.

3 Peak Identification

This chapter describes the concepts of peak identification.

4 Calibration

This chapter describes the calibration principles in the ChemStation software.

5 Quantification

This chapter describes how ChemStation does quantification. It gives details on area% and height% calculations, external standard (ESTD) calculation, norm% calculation, internal standard (ISTD) calculation, and quantification of unidentified peaks.

6 Evaluating System Suitability

This chapter describes what ChemStation can do to evaluate the performance of both the analytical instrument before it is used for sample analysis, and the analytical method before it is used routinely and to check the performance of analysis systems before, and during routine analysis.

7 CE specific Calculations

This chapter is relevant only if you use ChemStation to control CE instruments.

8 Data Review, Reprocessing and Batch Review

This chapter describes the possibilities to review data and how to reprocess sequence data. In addition it describes the concepts of Batch Review, Batch configuration, review functions, and batch reporting.

9 Reporting

This topic explains and provides a reference to the ACAML scheme used in the intelligent reporting feature of the OpenLAB CDS software.

10 System Verification

This chapter describes the verification function and the GLP verification features of the ChemStation.

Contents

1 Data Acquisition

This chapter describes the concepts of Data Acquisition, data files, logbook, and more.

What is Data Acquisition?

During data acquisition, all signals acquired by the analytical instrument are converted from analog signals to digital signals in the detector. The digital signal is transmitted to ChemStation electronically and stored in the signal data file.

Data Files

A data file comprises a group of files, by default stored in the DATA directory or in a subdirectory of this folder as a subdirectory with a data file name and a .D extension. A data file name can be defined manually using up to 42 characters (including the extension). Each file in the directory follows a naming convention (see *File Naming Conventions* in the *Concepts and Workflows* Guide). Additional data directories can be added using the **Preferences** settings.

Table 1 Data files

Name	Description
*.CH	Chromatographic/electropherographic signal data files. The file name comprises the module or detector type, module number and signal or channel identification. For example, ADC1A.CH, where ADC is the module type, 1 is the module number and A is the signal identifier and .CH is the chromatographic extension.
*.UV	UV spectral data files. The file name comprises the detector type and device number (only with diode array and fluorescence detector).
REPORT.TXT, REPORT.PDF	Report data files for the equivalent signal data files. Note: the PDF filename can be different if you use Unique PDF file naming.
Acq.MACAML	The file contains information on the method used during data acquisition. The information is stored in the ACAML format. ACAML files are used by Intelligent Reporting.
Sequence.ACAM_	The file contains the single injection results. The information is stored in the ACAML format. ACAML files are used by Intelligent Reporting.

Table 1 Data files

Name	Description
SAMPLE.MAC or Sample.XML	This file is used to store the sample values
SAMPLE.MAC.BAC	Backup of the original sample.mac. The .bac file is created during reprocessing, when the sample parameters (like multipliers) are updated the first time. It stores the original sample values used during acquisition.
RUN.LOG	Logbook entries which have been generated during a run. The logbook keeps a record of the analysis. All error messages and important status changes of ChemStation are entered in the logbook.
LCDIAG.REG	For LC only. Contains instrument curves (gradients, temperature, pressures, etc.), injection volume and the solvent descriptions.
ACQRES.REG	Contains column information. For GC it also contains the injection volume.
GLPSAVE.REG	Part of the data file when Save GLP Data is specified.
M_INTEV.REG	Contains manual integration events.

Online Monitors

There are two types of online monitors, the online signal monitor and the online spectra monitor.

Online Signal Monitor

The online signal monitor allows you to monitor several signals and, if supported by the associated instrument, instrument performance plots in the same window. You can conveniently select the signals you want to view and adjust the time and absorbance axis. For detectors that support this function a balance button is available.

You can display the absolute signal response in the message line by moving the cross hair cursor in the display.

Online Spectra Monitor

The online spectra monitor shows absorbance as a function of the wavelength. You can adjust both the displayed wavelength range and the absorbance scale.

Logbook

The logbook displays messages that are generated by the analytical system. These messages can be error messages, system messages or event messages from a module. The logbook records these events irrespective of whether they are displayed or not. To get more information on an event in the logbook double-click on the appropriate line to display a descriptive help text.

Status Information

ChemStation Status

The ChemStation Status window shows a summary status of the ChemStation software.

When a single analysis is running:

- the first line of the ChemStation Status window displays run in progress,
- the second line in the status window displays the current method status, and
- the raw data file name is shown in the third line together with the actual run time in minutes (for a GC instrument, files for front and back injector are also displayed).

The Instrument Status windows provide status information about the instrument modules and detectors. They show the status of the individual components and the current conditions where appropriate, for example, pressure, gradient and flow data.

Status Bar

The graphical user interface of the ChemStation system comprises toolbars and a status bar in the Method and Run Control View of ChemStation. The status bar comprises a system status field and information on the currently loaded method and sequence. If they were modified after loading they are marked with a yellow cogwheel. For a Agilent 1100/1200 Series module for LC a yellow EMF symbol reminds the user that usage limits that have been set for consumables (for example, the lamp) have been exceeded.

System Diagram

If supported by the configured analytical instruments (for example, the Agilent 1200 Infinity Series modules for LC or the Agilent 6890 Series GC) you

can display a graphical system diagram for your ChemStation system. This allows you to quickly check the system status at a glance. Select the System Diagram item from the View menu of the Method and Run Control View to activate the diagram. It is a graphical representation of your ChemStation system. Each component is represented by an icon. Using the color coding described below the current status is displayed.

Table 2 Colors used to indicate the status of the module or instrument

Color	Status
dark gray	offline
light gray	Standby (e.g. lamps off
yellow	not ready
green	ready
purple	pre-run, post-run
blue	run
red	error

In addition, you can display listings of actual parameter settings. Apart from a status overview, the diagram allows quick access to dialog boxes for setting parameters for each system component.

See the instrument part of the online help system for more information on the system diagram.

2 Integration

This chapter describes the concepts of integration the ChemStation integrator algorithms. It describes the integration algorithm, integration and manual integration.

What is Integration?

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

Integration is a necessary step for:

- Identification
- qualification
- calibration
- quantification,
- peak purity calculations, and
- spectral library search.

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:

- an autointegrate capability used to set up initial integrator parameters,
- the ability to define individual integration event tables for each chromatographic/electropherographic signal if multiple signals or more than one detector is used,
- interactive definition of integration events that allows users to graphically select event times,

- graphical manual integration of chromatograms or electropherograms requiring human interpretation (these events may also be recorded in the method and used as part of the automated operation),
- 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,
- area summation control,
- negative peak recognition,
- solvent peak definition detection
- integrator control commands defining retention/migration time ranges for the integrator operation.
- peak shoulder allocation through the use of second derivative calculations,
- improved sampling of non-equidistant data points for better performance with DAD LC data files that are reconstructed from DAD spectra.

The ChemStation Integrator Algorithms

Overview

To integrate a chromatogram/electropherogram 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, and peak width for each peak.

This process is controlled by **integration events**. The most important events are initial slope sensitivity, peak width, baseline correction, 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 addition, the auto integration function provides a set of initial events that you can optimize further.

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.

For more information, see "Initial Events" on page 47.

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.

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 (determine 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/electropherographic 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

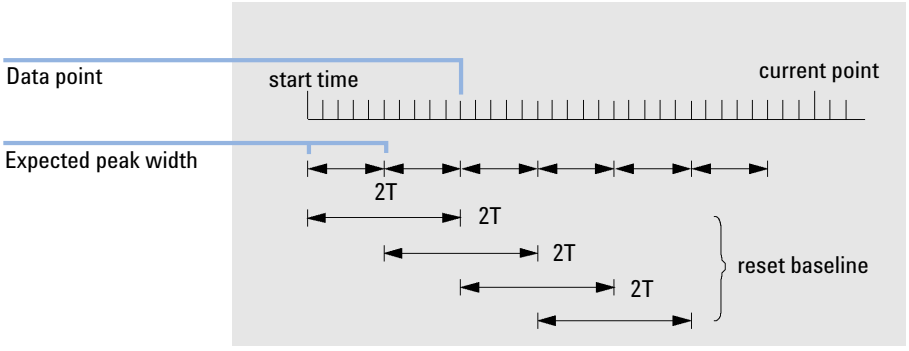

Figure 1 Baseline

At the start of the run the first data point is used. This baseline point is periodically reset according to the following formula:

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 1 on page 21).

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.

Definition of Terms

Cardinal Points

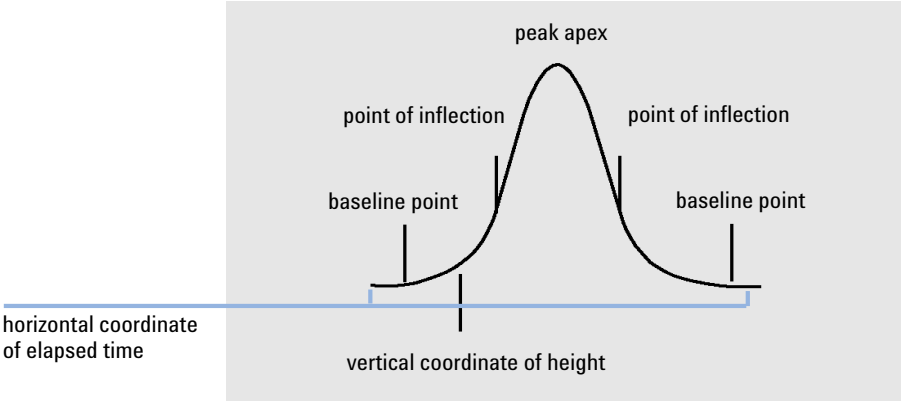

Figure 2 Cardinal points

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

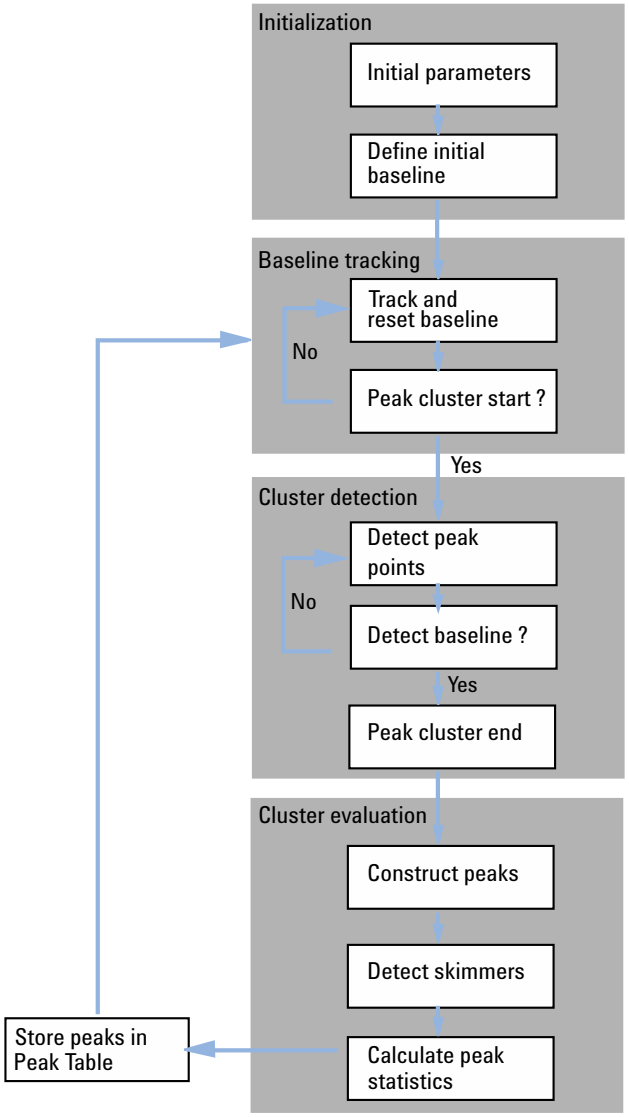

Figure 3 Integrator Flow Diagram

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.

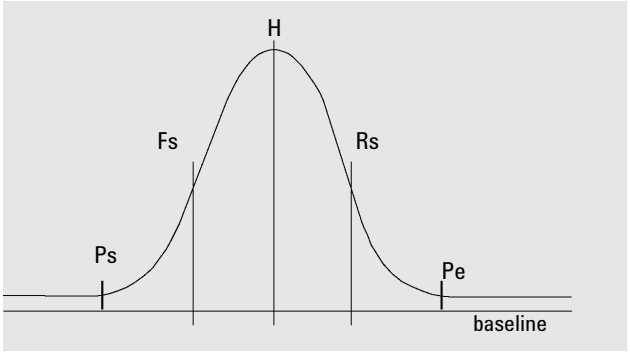

Figure 4 Peak width calculation

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/electropherographic peaks.

There are three ways the peak width is changed:

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

For peak width definitions used by System Suitability calculations please see "Evaluating System Suitability" on page 91

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/electropherographic

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 configuration:

For LC/CE configurations, the default peak width calculation uses a composite calculation:

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

For GC configurations, 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 x (existing peak width) + 0.25 x (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 Table 3 on page 28.

Bunching is performed in the powers of two based on the expected or experienced peak width. The bunching algorithm is summarized in Table 3 on page 28.

Table 3 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 indicate that a peak has begun.

Peak Start

In Table 4 on page 29 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 4 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 5 on page 29 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 5 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.

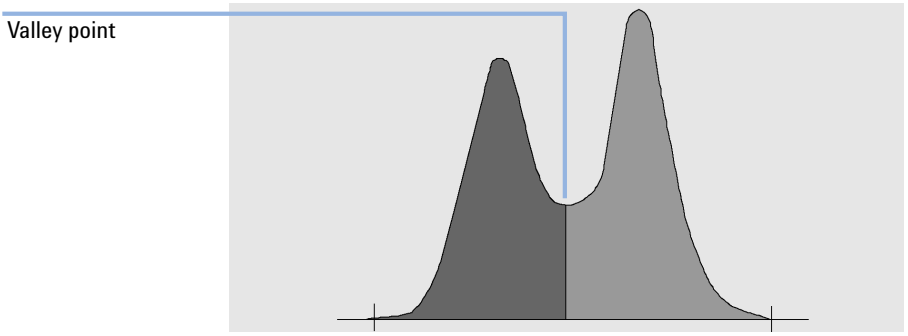

Figure 5 Merged Peaks

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.

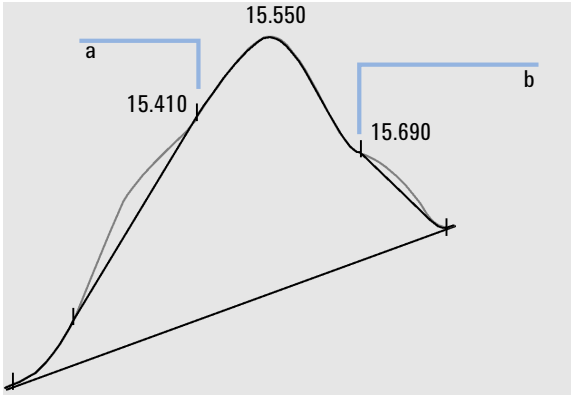

Figure 6 Peak Shoulders

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 31.

- 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 target point from starting point to curve.

retention/migration 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.

Baseline Allocation

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.

Default Baseline Construction

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

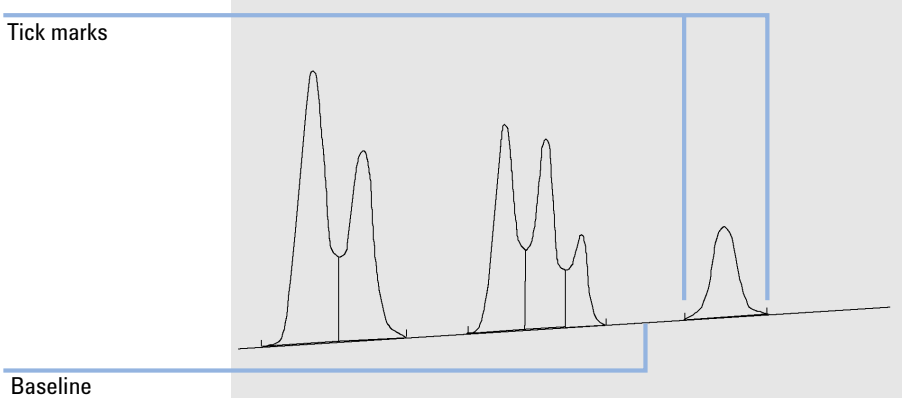

Figure 7 Default Baseline Construction

The Start of the Baseline

If no baseline is found at the start of the run, the start of the baseline is established in one of the following ways:

- from the start of the run to the first baseline point, if the start of run point is lower than the first baseline point,
- from the start of the run to the first valley point, if the start of run point is lower than the first valley,
- from the start of the run to the first valley point, if the first valley penetrates an imaginary line drawn from the start of run to the first baseline,
- from the start of the run to a horizontal baseline extended to the first baseline point.

The End of the Baseline

The last valid baseline point is used to designate the end of the baseline. In cases where the run does not end on the baseline, the end of the baseline is calculated from the last valid baseline point to the established baseline drift.

If a peak ends in an apparent valley but the following peak is below the area reject value as you have set it, the baseline is projected from the beginning of the peak to the next true baseline point. If a peak starts in a similar way, the same rule applies.

Baseline Penetration

A penetration occurs when the signal drops below the constructed baseline (point a in Figure 8 on page 34). If a baseline penetration occurs, that part of the baseline is generally reconstructed, as shown by points b in Figure 8 on page 34.

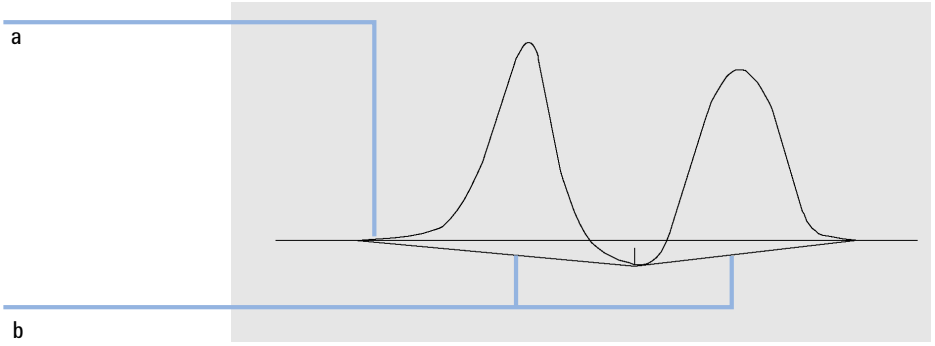

Figure 8 Baseline Penetration

You can use the following tracking options to remove all baseline penetrations:

Classical Baseline Tracking (no penetrations)

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 (compare the baselines in Figure 8 on page 34 and Figure 9 on page 34).

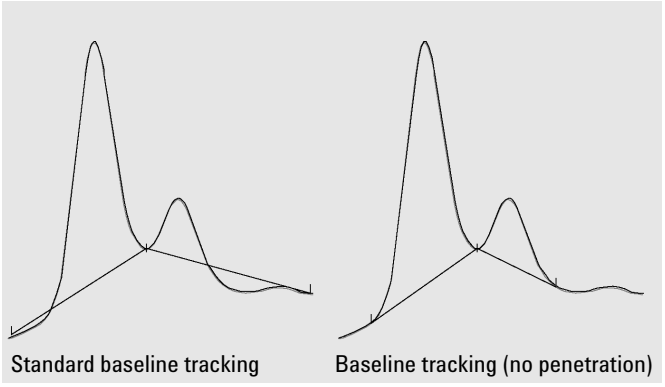

Figure 9 Standard baseline tracking and baseline tracking (no penetration)

NOTE **Baseline tracking (no penetration)** is not available for solvent peaks, with their child peaks and shoulders.

Advanced Baseline Tracking

In the advanced baseline tracking 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 8 on page 34). In many cases, advanced baseline tracking mode gives a more stable baseline, which is less dependant on slope sensitivity.

Peak 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 9 on page 34 with Figure 10 on page 35).

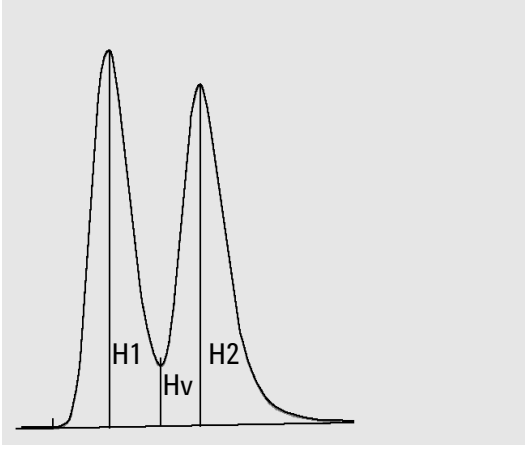

Figure 10 Peak Valley Ratio

The peak valley ratio (JP) and the peak-to-valley ratio (EP) is calculated using the following equations:

H1 ≥ H2, Peak valley ratio = H2/Hv

and

H1 < H2, Peak valley ratio = H1/Hv

Figure 11 on page 36 shows how the user-specified value of the peak valley ratio affects the baselines.

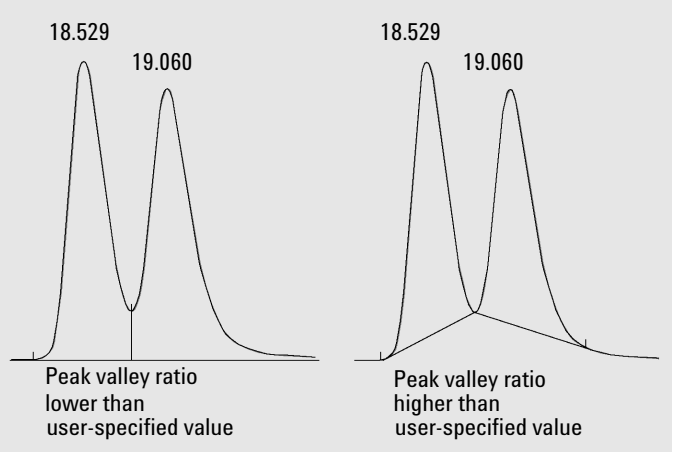

Figure 11 Effect of peak valley ratio on the baselines

Tangent Skimming

Tangent skimming is a form of baseline constructed for peaks found on the upslope or downslope of a peak. When tangent skimming is enabled, four models are available to calculate suitable peak areas:

- exponential curve fitting,
- new exponential skim
- straight line skim,
- combined exponential and straight line calculations for the best fit (standard skims).

Exponential Curve Fitting

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.

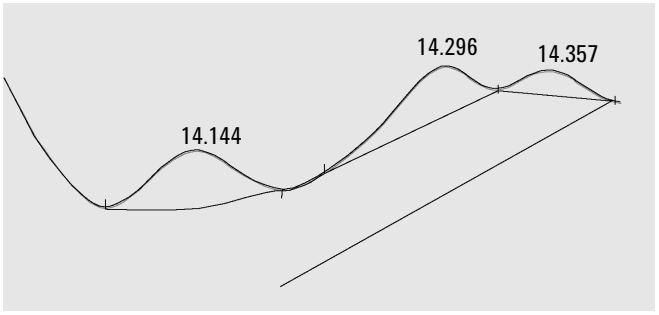

Figure 12 Exponential skim

New Mode Exponential Curve Fitting

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.

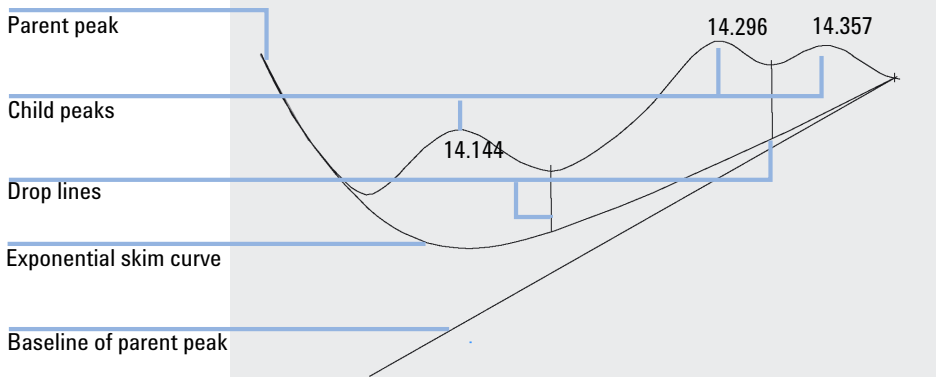

Figure 13 New mode exponential skim

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.

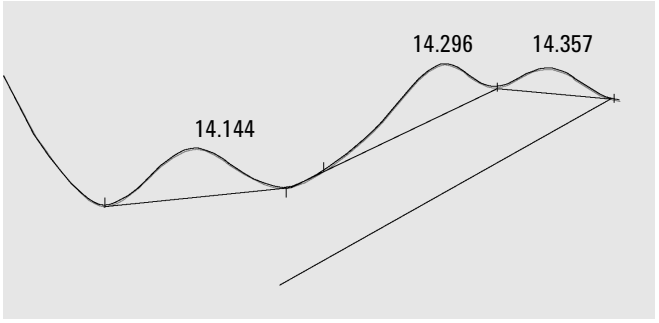

Figure 14 Straight line skim

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 tangent skim.

Skim Criteria

Two criteria determine whether a skim line is used to calculate the area of a child peak eluting on the trailing edge of a parent peak:

- tail skim height ratio
- valley height ratio

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 separation code between parent peak and child peak must be of type **Valley**.

Tail Skim Height Ratio is the ratio of the baseline-corrected height of the parent peak (Hp in Figure 15 on page 39) to the baseline-corrected height of the child peak (Hc). This ratio must be greater than the specified value for the child peak to be skimmed.

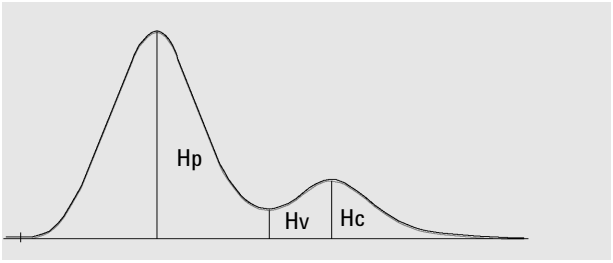

Figure 15 Skim criteria

You can disable exponential skimming throughout the run by setting the value of the tail skim height ratio to a high value or to zero.

Valley Height Ratio is the ratio of the height of the child peak above the baseline (Hc in Figure 15 on page 39) to the height of the valley above the baseline (Hv in same figure). This ratio must be smaller than the specified value for the child peak to be skimmed.

Calculation of Exponential Curve Fitting for Skims

The following equation is used to calculate an exponential skim:

$$Hb = Ho \times exp\,(- B \times (Tr - To)) + A \times Tr + C$$

where

Hb = height of the exponential skim at time Tr

Ho = height (above baseline) of the start of the exponential skim

B = decay factor of the exponential function

To = time corresponding to the start of the exponential skim

A = slope of the baseline of the parent peak

C = offset of the baseline of the parent peak

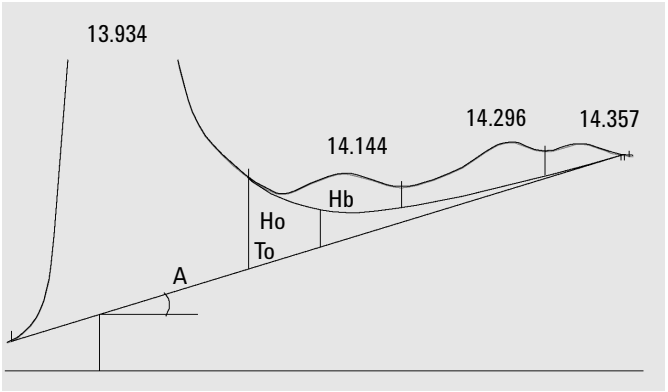

Figure 16 Values used to calculate an exponential skim

The exponential model is fitted through the part of the tail of the parent peak immediately before the first child peak. Figure 17 on page 40 shows the corrected curve of a child peak after tangent skimming.

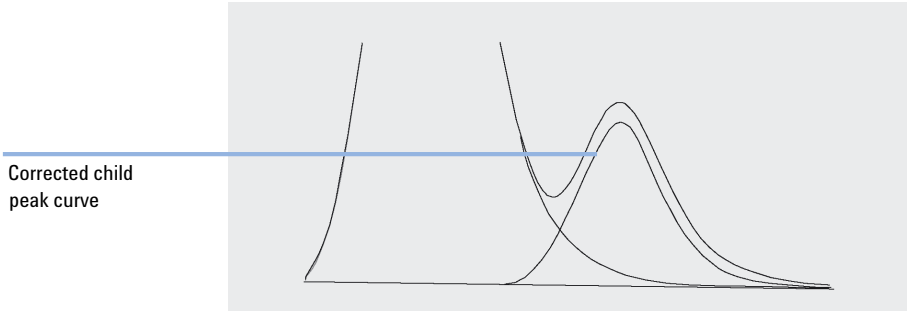

Figure 17 Tail-corrected child peak

Front Peak Skimming

As for child peaks on the tail of a parent peak, special integration is required for some peaks on the front/upslope of a peak, see Figure 18 on page 41.

Figure 18 Front peak skimming

Front peak skimming is treated the same way as tail peak skimming, using the same skim models.

The skim criteria are:

- front skim height ratio
- valley height ratio

The valley height ratio takes the same value for both front peak skimming and tail peak skimming (see "Valley height ratio"); the front skim height ratio is calculated in the same way as the tail skim height ratio (see "Tail skim height ratio"), but can have a different value.

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, as shown in Figure 19 on page 42.

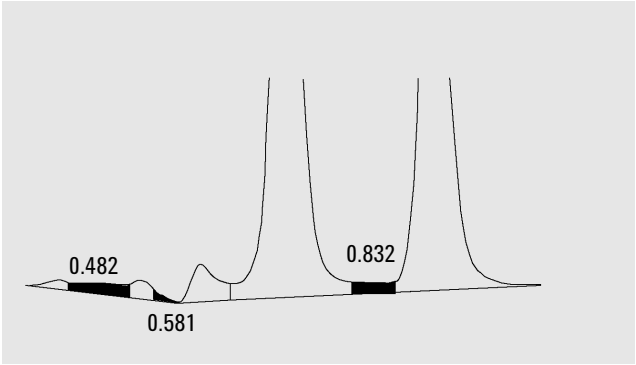

Figure 19 Unassigned Peaks

Peak Separation 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 6 Four character type field of a peak separation 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.

Character 4

The fourth character describes the peak type:

Blank space The peak is a normal peak.

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.

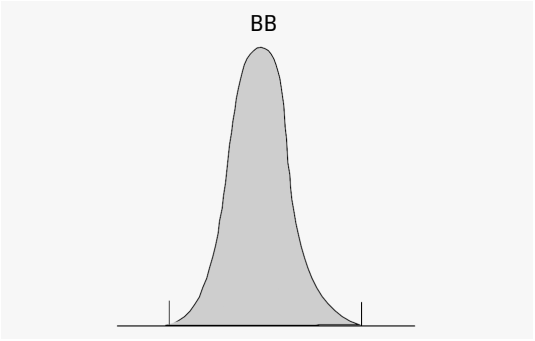

Figure 20 Area measurement for Baseline-to-Baseline Peaks

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 (identified by tick marks).

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 20 on page 44,
- for valley-to-valley (VV) peaks, the area above the baseline, segmented with vertical dropped lines from the valley points, as in Figure 21 on page 45,

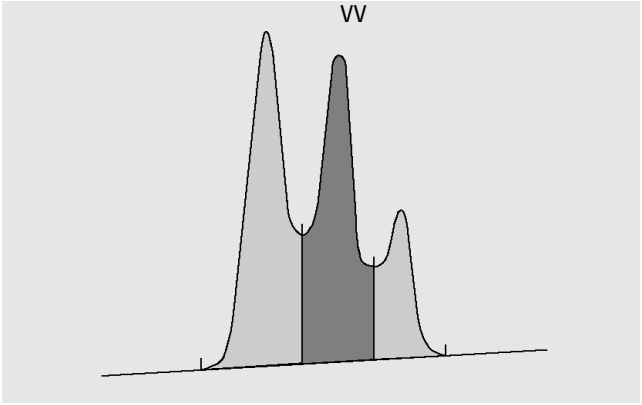

Figure 21 Area Measurement for Valley-to-Valley Peaks

- 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 22 on page 45.

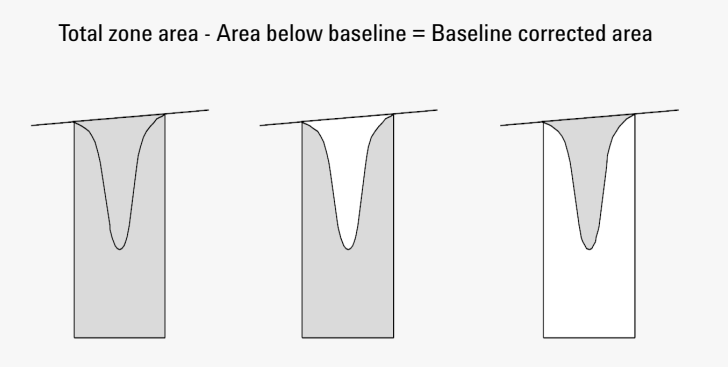

Figure 22 Area Measurement for Negative Peaks

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.

Integration Events

The integrator provides you with a number of initial and timed integrator events. Many events are on/off or start/stop pairs.

Integration Events for all Signals

The following events are provided for all signals:

- Tangent Skim Mode
- Tail Peak Skim Height Ratio
- Front Peak Skim Height Ratio
- Skim Valley Ratio
- Baseline Correction
- Peak-to-Valley Ratio

Initial Events

Initial Peak Width Initial peak width sets the integrator's internal peak width to this value for the start of run. This initial peak width is used to scale the accumulator that detects peak up-slope, down-slope, and tailing. The integrator updates the peak width when necessary during the run to optimize the integration. 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).

Slope Sensitivity Slope sensitivity is the setting for peak sensitivity. This is a setting that changes on a linear scale.

Height reject Height reject sets peak rejection by final height. Any peaks that have heights less than the minimum height are not reported.

Area reject Area reject sets peak rejection by final area. Any peaks that have areas less than the minimum area are not reported.

Shoulder detection

When shoulder detection is on, the integrator detects shoulders using the curvature of the peak as given by the second derivative. When the curvature goes to zero, the integrator identifies this point of inflection as a possible shoulder. If the integrator identifies another point of inflection before the apex of the peak, a shoulder has been detected.

Peak Width

The peak width setting controls the selectivity of the integrator to distinguish peaks from baseline noise. To obtain good performance, the peak width must be set close to the width at half-height of the actual peaks. The integrator updates the peak width when necessary during the run to optimize the integration.

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 as 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.

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)

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.

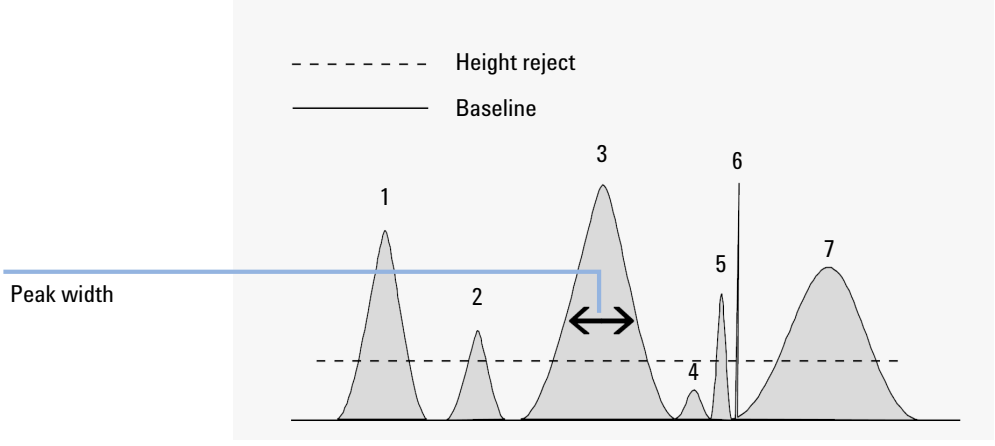

Figure 23 Using Initial Events

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, 5 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 7 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

Timed Events

OpenLAB CDS ChemStation Edition 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, 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. For further information about integration events see also"Initial Events" on page 47

Area Summation

Area Sum Sets points between which the integrator sums the areas between the area sum on and the area sum off time.

Area Sum Slice This event is similar to **Area Sum**. It allows to integrate contiguous time-slices of the chromatogram without loss of time intervals.

The area sum feature allows you to follow a longterm user defined baseline allowing to integrate over a cluster of peaks by setting an interval. Area summation sums the areas under the peaks for this interval. The system defines the Retention Time of the area sum as the center point of the time

interval over which the area is summed. The accuracy of the defined center point varies between 0.001 min at a high data rate and 0.1 min at a low data rate.

Baseline Events

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.

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

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.

The following events can be used for area summation with **Area Sum Slice** in complex chromatograms. They help to find the best baseline definition automatically, making manual interactions unnecessary. This is especially useful for analyzing GC results. The baseline is calculated based on a time interval using statistical estimates.

Set Baseline from Range Defines the range of the chromatogram used to estimate the new baseline. The range of data points is used to calculate a statistically meaningful baseline point at the midpoint of a time-range. This algorithm intelligently ignores spike-disturbances or unexpected peaks occurring in this interval via a two-stage statistical elimination feature. This ensures more reliable results for the baseline estimate.

Two events **Set Baseline from Range** are connected with a straight line between their center points. Figure 24 on page 52 illustrates the setting of the baseline range interval which is shown as a shaded gray area.

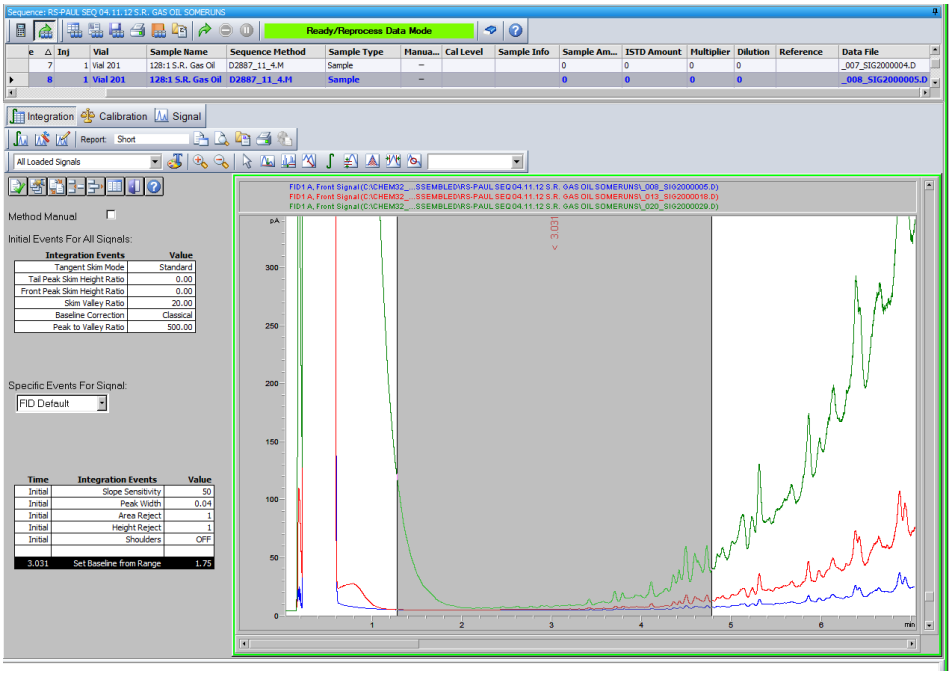

Figure 24 "Set Baseline from Range" : The baseline range interval is indicated by a gray shade

Set Low Baseline from Range Similar to **Set Baseline from Range**, but reduces its value in order to minimize baseline penetration. **Set Low Baseline from Range** is calculated by a subtraction of two sigma (Noise standard deviation) from the **Set Baseline from Range** y-value.

Use Baseline from Range Allows to project a baseline value to a later or earlier time. It also allows to construct baseline curves that change the slope underneath a cluster of peaks.

Autointegrate

The **Autointegrate** function provides a starting point for setting initial events. This is particularly useful when you are implementing a new method. You start with a default integration events table that contains no timed events; you can then optimize the parameters proposed by the Autointegrate function for general use.

Principles of Operation

The **Autointegrate** function reads the chromatogram data and calculates the optimal values for the initial integration parameters for each signal in the chromatogram object.

The algorithm examines 1% at the start and end of the chromatogram and determines the noise and slope for this part. Noise is determined as 3 times the standard deviation of the linear regression divided by the square root of the percent number of points used in the regression. These values are used to assign appropriate values to the height reject & slope sensitivity for the integration. The algorithm then assigns a temporary value for the peak width, depending on the length of the chromatogram, using 0.5% for LC and 0.3% to 0.2% for GC. The initial area reject is set to zero and a trial integration is performed. The trial is repeated several times if necessary, adjusting the parameters each time until at least 5 peaks are detected or integration is performed with an initial height reject of 0. The trial integration is terminated if the above conditions are not met after 10 trials.

The results of the integration are examined and the peak width is adjusted based on the peak widths of the detected peaks, biasing the calculation towards the initial peaks. The peak symmetry of the detected peaks is used to include only those peaks with symmetry between 0.8 and 1.3 for the peak width calculation. If not enough symmetric peaks are found, this limit is relaxed to *minSymmetry/1.5* and *maxSymmetry×1.5*. The baseline between the peaks is then examined to refine the earlier values of height reject & slope sensitivity. The area reject is set to 90% of the minimum area of the most symmetric peak detected during the trial integration.

The chromatogram is re-integrated using these final values for the integration parameters, and the results of the integration are stored.

Autointegrate Parameters

The following parameters are set by the autointegrate function:

- Initial slope sensitivity
- Initial height
- Initial peak width
- Initial area reject

Manual Integration

This type of integration allows you to integrate selected peaks or groups of peaks. Except for the initial area reject value, the software's event integration is ignored within the specified range of manual integration. If one or more of the peaks resulting from manual integration is below the area reject threshold, it is discarded. The manual integration events use absolute time values. They do not adjust for signal drift.

Manual Integration enables you to define the peak start and stop points, and then include the recalculated areas in quantification and reporting. Each of these points is labeled in reports with the peak separation code M.

Manual Integration offers the following features:

Draw Baseline specifies where the baselines are to be drawn for a peak or set of peaks. With menuitem **Integration > all valleys** you can also specify whether peaks in the range given should be automatically separated at all valley points.

Negative Peaks specifies when to treat any areas below the baseline as negative peaks. You can also specify whether peaks in the range given should be automatically separated at all valley points.

Tangent Skim calculates the areas of peaks tangentially skimmed off a main peak. The area of the tangent skimmed peak is subtracted from the area of the main peak.

Split Peak specifies a point where to split a peak with a drop-line.

Delete Peak(s) deletes one or more peaks from the integration results.

Peak Separation Codes for Manually-Integrated Peaks

Manually-integrated peaks are labeled in the integration reports by the peak code *MM*.

If there is a peak before the manually-integrated peak, and the end of this peak changes because of the manual integration, it is given the code *F* (forced). When valley points are detected they are set to code *V*.

A solvent on main peak which has been affected by manual integration, such as tangent skim, is labeled *R* (re-calculated solvent).

Saving Manual Integration Events

Manual integration events, e.g. a manually drawn baseline, are even more data file and signal specific than timed integration events. In case of complicated chromatograms, it is highly desirable to be able to use these events for reprocessing. Therefore manual integration events can be stored directly in the data file per signal rather than with the method.

Each time the data file is reviewed or reprocessed, the manual events in the data file are automatically applied. A run containing manual integration events is marked in the **Navigation Table** in the corresponding column.

In addition to the tools for drawing a baseline and deleting a peak manually, three additional tools are available in the user interface to

- Save manual events of the currently shown chromatograms into the data file,
- Remove all events from the currently shown chromatograms,
- Undo the last manual integration events (available until the event is saved).

When continuing to the next data file during review in the **Navigation Table**, ChemStation will check for unsaved manual integration events and ask the user whether he wants to save the events.

Manual events stored in the data file during review in the **Navigation Table** do not interfere with manual integration events stored during review in the **Batch** mode. These two ways of review are completely separated with regard to the manual events of a data file.

In ChemStation revisions prior to B.04.01, manual integration events were stored in the method instead of the individual data file. In B.04.01, this workflow can still be used. The **Integration** menu in **Data Analysis** view provides the following items in order to handle manual integration events with the method:

- **Update Manual Events of Method**: Save newly drawn manual events to the method.
- **Apply Manual Events from Method**: Apply the manual events currently saved in the method to the currently loaded data file.
- **Remove Manual Events from Method**: Delete the manual events from the method.

In order to convert manual events stored in a method to storage in the data file, apply the events from the method and store the results in the data file. If wanted, remove the events from the method.

In case the **Manual Events** checkbox of the **Integration Events Table** of a method is enabled, the manual events of the method are always applied when loading a data file using this method. If the data file contains additional manual events, they are applied after the events of the method. When the **Manual Events** checkbox is enabled, the user is never asked to save the events to the data file.

In order to convert manual events stored in a method to storage in the data file, apply the events from the method and store the results in the data file. You may remove the events from the method now.

In case the **Manual Events** checkbox of the **Integration Events Table** of a method is enabled, the manual events of the method are always applied when loading a data file using this method. If the data file contains additional manual events, they are applied after the events of the method. When the **Manual Events** checkbox is enabled, the user is never asked to save the events to the data file.

3
Peak Identification

This chapter describes the concepts of peak identification.

What is Peak Identification?

Peak identification identifies the components in an unknown sample based on their chromatographic/electropherographic characteristics determined by the analysis of a well-defined calibration sample.

The identification of these components is a necessary step in quantification if the analytical method requires quantification. The signal characteristics of each component of interest are stored in the calibration 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 calibration table.

The calibration table contains the expected retention/migration times of components of interest. A peak that matches the retention/migration time of a peak in the calibration table is given the attributes of that component, for example, the name and response factor. Peaks that do not match any of the peaks in the calibration table are classified as unknown. The process is controlled by:

- the retention/migration time in the calibration table for peaks designated as time reference peaks,
- the retention/migration time windows specified for reference peaks,
- the retention/migration times in the calibration table for the calibrated peaks that are not time reference peaks,
- the retention/migration time window specified for these non-reference peaks, and
- the presence of any additional qualifying peaks in the correct ratios.

Peak Matching Rules

The following rules apply to the peak matching process:

- if a sample peak falls within the peak matching window of a component peak from the calibration table, the peak is given the attributes of that component,
- if more than one sample peak falls within the peak matching window, then, the peak closest to the expected retention/migration time is identified as that component,
- if a peak is a time reference or internal standard, then the largest peak in the window is identified as that component,
- if peak qualifiers are also used then the peak ratio is used in combination with the peak matching window to identify the component peak,
- if the peak is a qualifier peak, the measured peak closest to the main peak of the compound is identified, and
- if a sample peak does not fall in any peak matching window, it is listed as an unknown component.

Types of Peak Identification

There are different techniques that can be used to match sample peaks with those in the calibration table of the ChemStation software.

Absolute Retention/Migration Time

The retention/migration time of the sample peak is compared with the expected retention/migration time specified for each component in the calibration table.

Relative Retention Time

The system calculates Relative retention time (EP) and Relative retention time (USP) as (Rr = t2/t1) both for calibrated peaks and for uncalibrated peaks.

Corrected Retention/Migration Time

The expected retention/migration times of component peaks are corrected using the actual retention/migration times of one or more reference peaks, and the matching process is done using these corrected (relative) retention/migration times. The reference peak or peaks must be specified in the calibration table.

Peak Qualifiers

In addition to identifying peaks by retention/migration time, you can use peaks qualifiers to allow a more precise result. If more than one peak occurs in a retention/migration time window then qualifiers should be used to identify the correct compound.

Amount Limits

The amount limits defined in the Compound Details dialog box are used to qualify the peak identification. If the amount of the identified compound is inside the amount limits the peak identification is indicated in the report.

Absolute Retention/Migration Time

A retention/migration time window is used in the peak matching process. The retention/migration time window is a window which is centered on the retention/migration time for an expected peak. Any sample peak that falls within this window may be considered as a candidate for component identification.

Figure 25 on page 62 shows a retention/migration time window for peak 2 which is between 1.809 and 2.631 minutes where the expected retention/migration time is 2.22 minutes. There are two possibilities for peak 2. One is at 1.85 minutes and the other at 2.33 minutes. If the expected peak is a non-reference peak, the peak closest to the expected retention/migration time of 2.22 minutes is selected.

If the expected peak is a time reference or internal standard, the largest peak in the window is selected.

In both cases the ChemStation selects the peak at 2.33 minutes. If the two peaks were the same size then the peak closest to the center of the window is chosen.

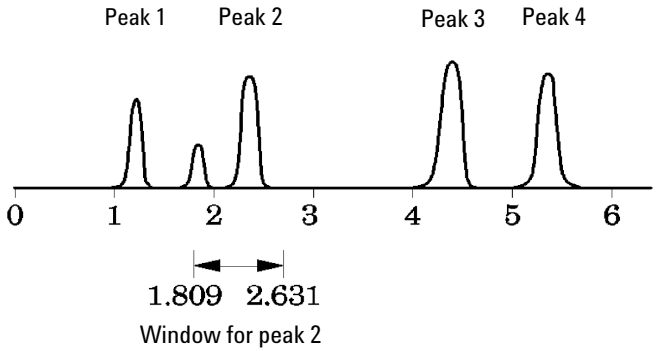

Figure 25 Retention/Migration Time Windows

Three types of windows are used when trying to locate peaks.

- reference peak windows which apply to reference peaks only,
- non-reference peak windows which apply to all other calibrated peaks, and

- specific window values for individual components which are set in the **Compound Details** dialog box.

The default values for these windows are entered in the Calibration Settings dialog box. The width on either side of the retention/migration time that defines the peak matching window is the sum of the absolute and percentage windows.

A window of 5 % means the peak must have a retention/migration time between less than 2.5 % and more than 2.5 % of the calibrated retention/migration time for that peak. For example, a peak with a retention/migration time of 2.00 in the calibration run must appear between 1.95 and 2.05 minutes in subsequent runs.

For example, an absolute window of 0.20 minutes and a relative window of 10 % gives a retention/migration time window of between 1.80 and 2.20 minutes.

1.80 min = 2.00 min - 0.10 min (0.20 min / 2) - 0.10 min (10 % of 2.00 min).

2.20 min = 200 min + 0.10 min (0.20 min / 2) + 0.10 min (10 % of 2.00 min).

Corrected Retention/Migration Times

To match peaks by absolute retention/migration times may be simple but not always reliable. Individual retention/migration times may vary slightly due to a small change in conditions or technique. As a result peaks may occur outside the peak matching windows and therefore are not identified.

A technique to deal with the inevitable fluctuations that occur in absolute retention/migration times is to express component retention/migration times relative to one or more reference peaks.

Reference peaks are identified in the calibration table with an entry in the reference column for that peak. The relative peak matching technique uses the reference peak or peaks to modify the location of the peak matching windows in order to compensate for shifts in the retention/migration times of sample peaks.

If no reference peak is defined in the method or the ChemStation cannot identify at least one reference peak during the run, the software will use absolute retention/migration times for identification.

Single Reference Peaks

A retention/migration time window for the reference peak is created around its retention/migration time. The largest peak falling within this window is identified as the reference peak. The expected retention/migration times of all other peaks in the calibration table are corrected, in proportion to the ratio of the expected retention/migration time to the actual retention/migration time of the reference peak.

Multiple Reference Peaks

Correcting retention/migration times with a single reference peak is based on the assumption that the deviation of actual retention/migration time from the expected retention/migration times changes uniformly and linearly as the run

progresses. Often during a long run the retention/migration time changes non-uniformly. In such cases better results are obtained using multiple reference peaks spaced at intervals across the run. This splits the signal into separate zones. Within each zone the deviation between retention/migration times is assumed to change linearly, but the rate of change is determined separately for each zone.

NOTE The time correction algorithm may fail if the retention times of multiple reference peaks are too close to each other and are not distributed across the total run time.

Peak Qualifiers

A component can be detected with more than one signal. Although applicable to all forms of chromatography using multiple detectors or detectors capable of producing multiple signals, multisignal detection is most commonly used in liquid chromatography with multiple wavelength or diode array detectors. Such detectors are normally set up so that the wavelength closest to the greatest absorbance (area) is used to define the main peak in the calibration table. In Figure 26 on page 66 this is Lambda_1.

The two other wavelengths that were acquired as signals can be used as peak qualifiers. In the figure these are Lambda_2 and Lambda_3.

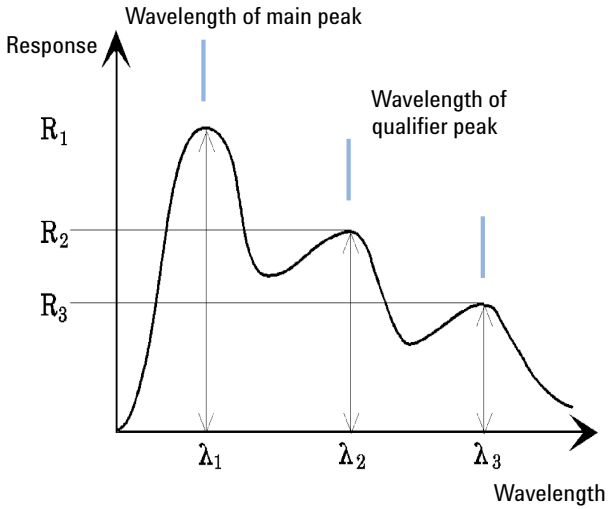

Figure 26 Peak Qualifiers

Peaks of a compound have a constant response ratio over different wavelengths.

The qualifier peak response is a certain percentage of the main peak response. Limits which determine the acceptable range for the expected response can be set in the calibration table when the Identification Details option is selected. If the ratio between the main peak qualifier Lambda_1 and the qualifier peak, for

example, $Lambda_3$ is within the allowed limits then the compound identity can be confirmed.

Signal Correlation

Signal correlation means that two peaks measured in different detector signals within a defined time window are assigned to the same compound. The signal correlation window can be controlled by the **SignalCorrWin** parameter in the **QuantParm** table of the **_DaMethod** register. Signal correlation is disabled when setting the signal correlation window to 0.0 minutes (see the *Macro Programming Guide* for more information). When signal correlation is off, peaks eluting at the same retention/migration time in different detector signals are treated as different compounds.

The default signal correlation window for LC, CE, CE/MS and LC/MS data is 0.03 minutes and 0.0 minutes for GC data.

Qualifier Verification

If signal correlation is enabled, qualifier verification is active for all data file types by default. It can be disabled by setting the **UseQualifiers** flag in the **Quantification Parameters** table of the method. Qualifier verification is also disabled when signal correlation is switched off.

Qualifier Ratio Calculation

When qualifiers verification is enabled for a compound, the ratio of the qualifier size and the main peak size is verified against the calibrated limits. The size may be height or area according to the calculation base setting in Specify Report.

The qualifier peaks can be calibrated in the same way as the target compounds. The user does not need to specify the expected qualifier ratio. The expected qualifier ratio is calculated automatically:

both measured at the retention time of the compound.

The QualTolerance parameter defines the acceptable range of the qualifier ratio, for example, ± 20%.

The tolerance can be set in the calibration table user interface (Identification Details) and is an absolute percentage.

For multilevel calibrations, the ChemStation calculates a minimum qualifier tolerance based on the measured qualifier ratios at each calibration level. The minimum qualifier tolerance is calculated using the following equation:

$$\text{minimum qualifier tolerance} = \frac{\sum_{n=1}^{i} (q_i - \bar{q})}{\bar{q} \times i} \times 100$$

where q_i is the measured qualifier ratio at level i.

The Identification Process

When attempting to identify peaks, the software makes three passes through the integration data.

Finding the Reference Peaks

The first pass identifies the time reference peaks. The software searches peak retention/migration times from a run for matches within the retention/migration windows of the reference peaks in the calibration table. A peak from the run is identified as a reference peak in the calibration table if the run peak's retention/migration time is within the window constructed for the calibration table peak.

If more than one peak is found within a window, the peak with the largest area or height followed by a positive signal qualifier match, if set up, is chosen as the reference peak.

After each time reference peak is found, the difference between its retention/migration time and that given in the calibration table is used to adjust the expected retention/migration times of all other peaks in the Calibration table.

Finding the ISTD Peaks

The second pass identifies any defined internal standard peaks. If they have not already been identified as ISTD, peaks may be identified as time reference peaks. ISTD peaks are identified by peak retention/migration time windows and peak qualifiers. If more than one peak is found in the same ISTD window, the largest peak is chosen.

Finding the Remaining Calibrated Peaks

The third pass identifies all remaining peaks listed in the calibration table. The non-reference peaks in the calibration table are matched to the remaining run peaks by using their RT window.

Each non-reference calibrated peak has its own retention/migration time in the calibration table. This is adjusted for the particular run based on the pre-identification of the time reference peaks. The retention/migration time window of the calibrated peak is adjusted based on the corrected retention/migration time of the calibrated peak.

If more than one peak is found in the same window, the peak with a retention/migration time which is closest to the expected retention/migration time and also meets the optional qualifier specifications is chosen.

Classification of Unidentified Peaks

If there are remaining peaks, which are still not identified, they are classified as unknown. The ChemStation attempts to group the unknown peaks that belong to the same compound. If a peak has been detected in more than one signal, the peaks with the same retention/migration time in each signal are grouped to one compound.

Unknown peaks are reported if the corresponding selection has been made in the Calibration Settings dialog box.

4 Calibration

This chapter describes the calibration principles in the ChemStation software.

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 Figure 27 on page 72.

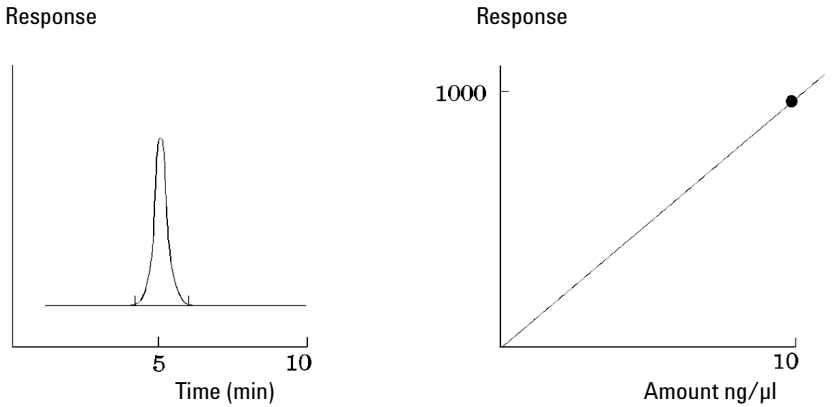

Figure 27 Calibration Sample (10 ng/µl) Signal and Calibration Curve

A *correlation coefficient* is displayed with the graphic of the calibration curve. The correlation coefficient is the square root of the regression coefficient and gives a measure of the fit of the calibration curve between the data points. The value of the coefficient is given to three decimal places, in the range:

0.000 to 1.000

where:

0.000 = no fit

1.000 = perfect fit

For each calibration level the *relative residual* is displayed. It is calculated using the following formula:

$$relRES = \frac{Response_{calibrated} - Response_{calculated}}{Response_{calculated}} \cdot 100$$

where:

relRES = relative residual in percent

The calculated response represents the point on the calibration curve.

The *residual standard deviation*, which is printed on some reports and when selecting Print calibration table and curves is calculated using the following formula:

$$ResSTD = \sqrt{\frac{\sum_{i=1}^{n} (Resp_{calibratedi} - Resp_{calculatedi})^2}{n-2}}$$

where:

ResSTD = residual standard deviation

$Resp_{calibratedi}$ = calibrated response for point i

$Resp_{calculatedi}$ = calculated response for point i

n = number of calibration points

Group Calibration

Group calibration can be applied for compounds where the individual concentrations are not known but the sum of concentrations for a group of compounds is known. An example are isomers. Complete compound groups are calibrated. The following formulae are used:

Calibration

$$Conc_{AB} = RF_A \cdot Response_A + RF_B \cdot Response_B$$

where:

$Conc_{AB}$ is the concentration of the compound group consisting of compound A and B

$Response_A$ is the area (or height) of compound A

RF_A is the response factor

For compounds within a compound group we assume equal response factors:

$$RF_A = RF_B$$

Therefore the concentration of a compound within a compound group is calculated as follows:

$$Conc_A = \frac{Conc_{AB} \cdot Resp_A}{Resp_A + Resp_B}$$

Recalibration Options

You have several ways to update the responses in the calibration table with the new calibration data.

Average

The average from all calibration runs are calculated using the following formula

$$Response = \frac{n \cdot Response + MeasResponse}{n + 1}$$

Floating Average

A weighted average for all calibration runs is calculated. The updated weight is set in the **Recalibration Settings** dialog box.

$$Response = \left(1 - \frac{Weight}{100}\right) \cdot Response + \left(\frac{Weight}{100}\right) \cdot MeasResponse$$

Replace

The new response values replace the old values.

5 Quantification

This chapter describes how ChemStation does quantification. It gives details on area% and height% calculations, external standard (ESTD) calculation, norm% calculation, internal standard (ISTD) calculation, and quantification of unidentified peaks.

What is Quantification?

After the peaks have been integrated and identified, the next step in the analysis is quantification. Quantification uses peak area or height to determine the concentration 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 this compound.
- Analyze a sample or samples containing a known concentration or concentrations of the compound to obtain the response due to that concentration.

  You may alternatively analyze a number of these samples with different concentrations of the compounds of interest if your detector has a non-linear response. This process is referred to as *multi-level calibration*.
- Analyze the sample containing an unknown concentration of the compound to obtain the response due to the unknown concentration.
- Compare the response of the unknown concentration to the response of the known concentration 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.

Quantification Calculations

The ChemStation offers the following calculation procedures for determining the concentration of each component present in a mixture:

- Percent
- Normalization
- External standard (ESTD)
- ESTD%
- Internal standard (ISTD)
- ISTD%

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

Correction Factors

The quantification calculations use four correction factors, the *absolute response factor*, the *multiplier*, the *dilution factor*, and the *sample amount*. These factors are used in the calibration procedures to compensate for variations in detector response to different sample components, concentrations, sample dilutions, sample amounts, and for converting units.

Absolute Response Factor

The absolute response factor for a sample component represents the amount of the component divided by the measured area or height of the component's peak in the analysis of a calibration mixture. The absolute response factor, which is used by each calibrated calculation procedure, corrects for detector response to individual sample components.

Multiplier

The multiplier is used in each calculation formula to multiply the result for each component. The multiplier may be used to convert units to express amounts.

Dilution Factor

The dilution factor is a number by which all calculated results are multiplied before the report is printed. 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.

Sample Amount

If the ESTD% or ISTD% calculations are selected, the ESTD and ISTD reports give relative values rather than absolute values, that is, the amount of each component is expressed as a percentage of the sample amount. The sample amount is used in ESTD% and ISTD% reports to convert the absolute amount of the components analyzed to relative values by dividing by the value specified.

Uncalibrated Calculation Procedures

Uncalibrated calculation procedures do not require a calibration table.

Area% and Height%

The **Area%** calculation procedure reports the area of each peak in the run as a percentage of the total area of all peaks in the run. **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 calibration table required for other calibration procedures.

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

The multiplier and dilution factor from the **Calibration Settings**, from the **Sample Information** dialog box, or from the **Sequence Table** are not applied in Area% or Height% calculation.

Calibrated Calculation Procedures

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

ESTD Calculation

The ESTD procedure is the basic quantification 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, component amounts are calculated by applying these response factors to the measured sample amounts. 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 report, 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 Calibration Settings or Calibration Curve dialog box.
2. The amount of the compound in the unknown is calculated using the equation described below. This amount may appear in the report or it may be used in additional calculations called for by Multiplier, Dilution Factor, or Sample Amount values before being reported.

If the ESTD report is selected, the equation used to compute the absolute amount of component x is:

$$\text{Absolute Amt of x} = Response_x \cdot RF_x \cdot M \cdot D$$

where:

$Response_x$ is the response of peak x;

RF_x is the response factor for component x, calculated as:

$$\text{RF}_x = \frac{\text{Amount}_x}{\text{Response}_x}$$

M is the multiplier.

D is the dilution factor.

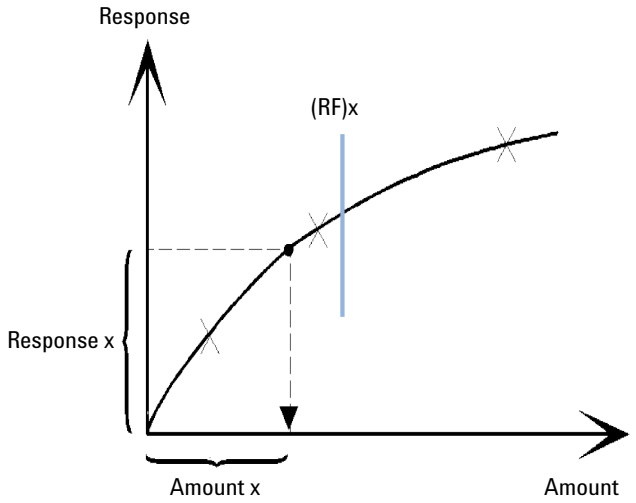

Figure 28 Response Factor

The multiplier and dilution factor are read either from the **Calibration Settings** or from the **Sample Information** dialog box.

If the ESTD% report is selected and sample amount is not zero, the relative amount (%) of a component x is calculated as shown below:

$$\text{Relative Amt of x} = \frac{\{\text{Absolute Amt of x}\} \cdot 100}{\text{Sample Amount}}$$

where:

Absolute amount of x is calculated as shown above in the ESTD calculation;

Sample amount is obtained from the Sample Information box, or from the Calibration Settings dialog box for single runs. If sample amount is zero, the ESTD is calculated.

Norm% Calculation

In the normalization method, response factors are applied to the peak areas (or heights) to compensate for changes that occur in detector sensitivity for the different sample components.

The Norm% report is calculated in the same way as an ESTD report except that there is an additional step to calculate the relative rather than absolute amounts of compounds.

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

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

$$\text{Norm\% of x} = \frac{\text{Response}_x \cdot \text{RF}_x \cdot 100 \cdot M \cdot D}{\sum\{\text{Response} \cdot RF\}}$$

where:

$Response_x$	is the area (or height) of peak x,
RF_x	is the response factor,
$\Sigma(Response \cdot RF)$	is the total of all the *(Response · RF)* products for all peaks including peak x,
M	is the multiplier,
D	is the dilution factor.

The multiplier and dilution factor are read either from the **Quantitation Settings** available in the **Specify Report** dialog box, or from the Sequence Table.

ISTD Calculation

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

The software takes the appropriate response factors obtained from a previous calibration stored in the method. Using the internal standard concentration and peak areas or heights from the run, the software calculates component concentrations.

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 if the calibration curve of the compound is non-linear.

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

In a two-run ISTD calibration, the calculation of the corrected amount ratio of a particular compound in an unknown sample occurs in the following stages:

Run 1: Calibration

1 The calibration points are constructed by calculating an amount ratio and a response ratio for each level of a particular peak in the calibration 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 area of the compound divided by the area or height of the internal standard at this level.

2 An equation for the curve through the calibration points is calculated using the type of curve fit specified in the Calibration Settings dialog box or Calibration Curve dialog box.

$$RF_x = \frac{\text{Amount Ratio}}{\text{Response Ratio}}$$

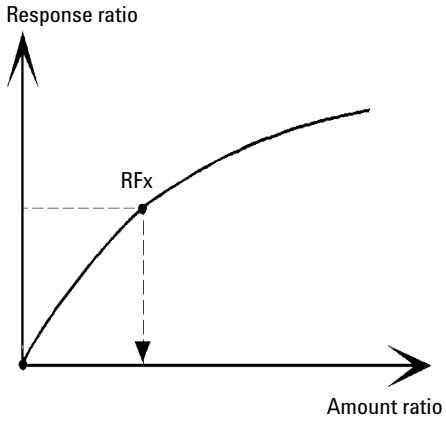

Figure 29 Amount Ratio

Run 2: 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 fit equation determined in step 2 above, and the actual amount of ISTD in the sample.

ISTD Calculation of Calibrated Peaks

The equations used to calculate the actual amount of a calibrated component x for a single-level calibration are:

$$\text{Response Ratio} = \frac{\text{Response}_x}{\text{Response}_{\text{ISTD}}}$$

$$\text{Actual Amount of x} = \text{RF}_x \cdot \{\text{Response Ratio}\}_x \cdot \text{Actual Amount of ISTD} \cdot M \cdot D$$

where:

RF_x is the response factor for compound x;

The actual amount *(Actual Amt)* of ISTD is the value that was entered in the Calibration Settings dialog box or the Sample Info dialog box for the internal standard added to the unknown sample;

M is the multiplier.

D is the dilution factor.

If the ISTD% report type is selected, the following equation is used to calculate the relative (%) amount of component x:

$$\text{Relative Amt of x} = \frac{\{\text{Absolute Amt of x}\} \cdot 100}{\text{Sample Amount}}$$

ISTD Calculation of Uncalibrated Peaks

There are two ways to define the response factor which is used to calculate the amount for unidentified peaks.

1 Use the fixed response factor set in the **With Rsp Factor** box of the **Calibration Settings** dialog box. You can choose to correct the fixed response factor by specifying an ISTD correction.

$$\text{Actual Amount of x} = \text{RF}_x \cdot \{\text{Response Ratio}\}_x \cdot \text{Actual Amount of ISTD} \cdot M \cdot D$$

$$\text{Response Ratio} = \frac{\text{Response}_x}{\text{Response}_{\text{ISTD}}}$$

RF_x is the Response Factor set in the **Calibration Settings** dialog box.

You can see from these formulae that the variations in the ISTD response are used to correct the quantification of the unknown component.

2 Use a calibrated peak. This ensures that the same response factor is used for the quantification of all peaks. The response factor of the selected compound and the uncalibrated peaks is corrected during all recalibrations. If the calibrated peak response factor changes, then the response factor for the unidentified peaks also changes by the same amount. If a Calibration Table is already set up, you can select a compound from the **Using Compound** combo box in the **Calibration Settings** dialog box.

The equations used to calculate the actual amount of an uncalibrated peak x are shown above.

6
Evaluating System Suitability

This chapter describes what ChemStation can do to evaluate the performance of both the analytical instrument before it is used for sample analysis, and the analytical method before it is used routinely and to check the performance of analysis systems before, and during routine analysis.

Peak Performance can be calculated for any integrated peak of the data loaded, and also for new manually integrated peaks. The Interactive Peak Performance tool calculates peak characteristics and display it on UI.

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. The ChemStation software provides the tools to do these three types of tests automatically. An instrument test can include the detector sensitivity, the precision of peak retention/migration times and the precision of peak areas. A method test can include precision of retention/migration 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 which 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.

The ChemStation collates results from several runs and evaluates them statistically in the sequence summary report.

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

- peak retention/migration time,
- peak area,
- amount,
- peak height,
- peak width at half height,
- peak symmetry,
- peak tailing,

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

The mean value, the standard deviation, the relative standard deviation and the confidence interval are calculated. You can set limits for either standard deviation, the relative standard deviation or the confidence interval for each of these parameters. Should the values exceed your limits, the report is flagged to draw your attention to them.

The quality of the analytical data can be supported by keeping records of the actual conditions at the time the measurements were made. The ChemStation's logbook records instrument conditions before and after a run. This information is stored with the data and reported with sample data. Instrument performance curves are recorded during the entire analysis as signals, and stored in the data file. If supported by the instrument these records, overlaid on the chromatogram, can be recalled on demand, for example, during an audit.

Baseline noise and drift can be measured automatically. A minimum detectable level can be calculated from peak height data for each calibrated compound in the method.

Finally, instrument configuration, instrument serial numbers, column/capillary identification, and your own comments can be included in each report printed.

Extended performance results are calculated only for compounds calibrated for in the method, ensuring characterization by retention/migration times and compound names.

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

- instrument details,
- column/capillary details,
- analytical method,
- sample information,

- acquisition information,
- signal description and baseline noise determination, and
- signal labeled with either retention/migration 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 a selected time range of a signal. Noise is treated in three different ways:

- as six times the standard deviation (sd) of the linear regression of the drift,
- as peak-to-peak (drift corrected), and
- as determined by the ASTM method (ASTM E 685-93).

The noise can be calculated for up to seven ranges of the signal; the ranges are specified as part of the system suitability settings in the reporting parameters.

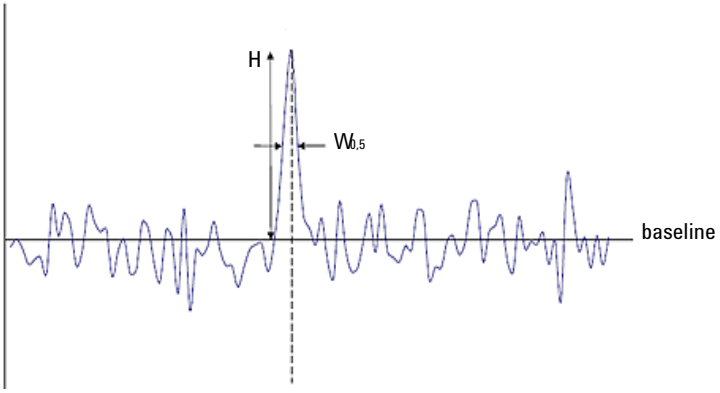

Figure 30 Chromatogram with peak signal and noise

H	Peak height from top to baseline (best straight line through noise)
$W_{0.5}$	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 given time range (see "Regression Analysis" on page 119). The noise is given by the formula:

$$N = 6 \times Std$$

where

N is the noise based on the six time standard deviation method, and

Std is the standard deviation of the linear regression of all data points in the selected time range.

Noise Calculation Using the Peak-to-Peak Formula

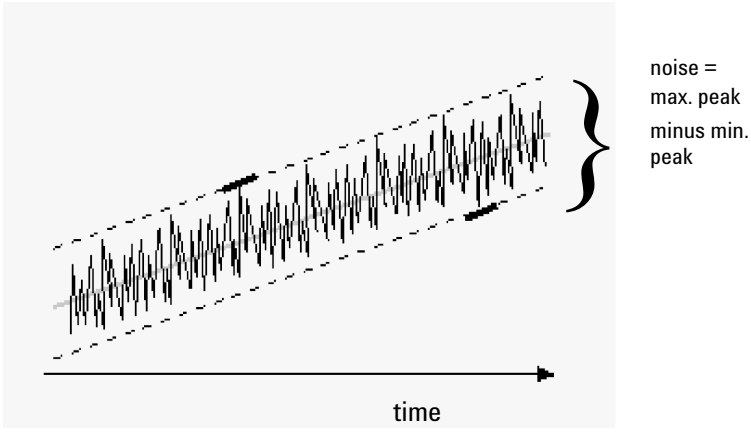

Figure 31 Noise as Maximum Peak to Minimum Peak (Distance)

The drift is first calculated by determining the linear regression using all the data points in the time range (see "Regression Analysis" on page 119). 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_x are the calculated data points using the LSQ formula, with

I_{max} the highest (maximum) intensity peak, and

I_{min} the lowest (minimum) intensity peak in the time range.

For European Pharmacopoeia calculations the Peak-to-Peak noise is calculated using a blank reference signal over a range of -10 and +10 times $W_{0.5}$ flanking each peak. This region can be symmetrical to the signal of interest, or asymmetrical if required due to matrix signals.

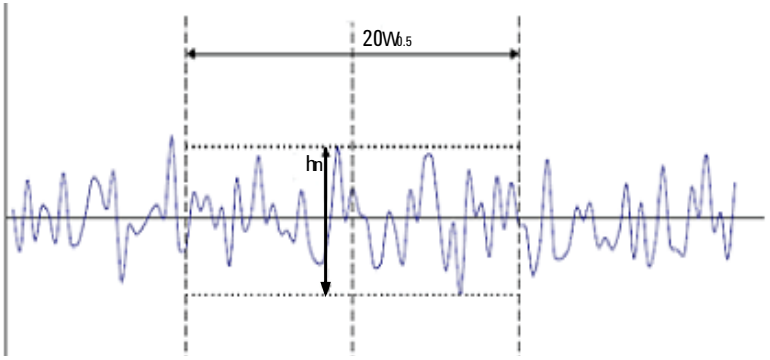

Figure 32 Determination of noise from the chromatogram of a blank sample

Where

20 $W_{0.5}$ is the region corresponding to the 20 fold of $W_{0.5}$.

h_n is the maximum amplitude of the baseline noise in the 20-fold $W_{0.5}$ region.

Noise Calculation 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.

Determination of the Number of Cycles, n

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

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

Calculation of 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 (see "Regression Analysis" on page 119). 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.

Calculation of ASTM Noise

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

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 %.

Signal-to-noise calculation

ChemStation has two options to calculate the signal-to-noise calculation:

- The "six times the standard deviation (sd)" of the linear regression of the drift to calculate the noise.

  or

- According to the definition of the European Pharmacopoeia: 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.

Signal-to-Noise Calculation without reference signal

The range closest to the peak is selected from the ranges as specified in the system suitability settings. The "six times the standard deviation (sd)" of the linear regression of the drift is used to calculate the noise

The signal-to-noise is calculated for each peak in the signal. If the ChemStation cannot find a noise value, the signal-to-noise is reported as “-”.

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

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

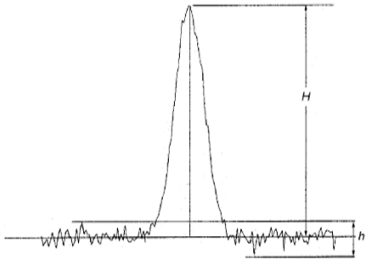

Figure 33 Signal-to-noise ratio

Signal-to-Noise Ratio Calculation according to the EP Definition

The signal-to-noise ratio (S/N) can be calculated per the European Pharmacopoeia definition. S/N is calculated using the equation:

$S/N = 2H/h$

Where:

H is the height of the peak corresponding to the component concerned in the chromatogram obtained with the prescribed reference solution,

h is the absolute value of the largest noise fluctuation from the baseline in a chromatogram obtained after injection 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.

The noise value used is calculated using the "Peak To Peak" method (see “Noise Calculation Using the Peak-to-Peak Formula” on page 97).

S/N is reported in for all peaks present in the chromatogram signal, provided there exists a corresponding reference signal. For a particular chromatogram signal the reference signal is assigned automatically if you specify the reference datafile. If no reference signal can be assigned to a chromatogram signal, signal-to-noise ratio will not be calculated for the peaks in that particular signal.

Determination of Noise Range

The noise range in the reference signal is determined according to one of the following algorithms

- If the reference signal is not long enough: *StartTime - EndTime* < $20*W_{0.5}$
  - *StartTime = startime (of reference signal)*, and
  - *EndTime = endtime (of the reference signal)*
- If the reference signal is not long enough, but the peak is situated such, that $(RT-10*W_{0.5})$ is less than the start point of reference signal
  - *StartTime = startime (of reference signal)*, and
  - *EndTime = StartTime* + $20*W_{0.5}$
- If the reference signal is long enough, but the peak is situated such, that *RT* or $RT+10*W_{0.5}$ is greater than the end point of the reference signal
  - *EndTime = endtime (of the reference signal)*, and
  - *StartTime = EndTime* - $20*W_{0.5}$
- If the peak is situated such, that *RT* or $RT+10*W_{0.5}$ is greater than the end point of the reference signal
  - *StartTime = RT* - $10*W_{0.5}$, and
  - *EndTime = RT* + $10*W_{0.5}$

Where:

RT is the Retention Time, and

$W_{0.5}$ is the peak width at half height.

Drift and Wander

Drift is given as the slope of the linear regression. The drift is first calculated by determining the linear regression using all the data points in the time range (see "Regression Analysis" on page 119). The linear regression line is subtracted from all data points within the time range to give the drift-corrected signal.

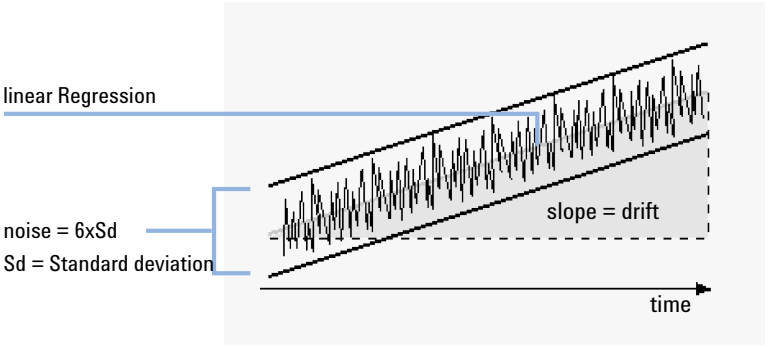

Figure 34 Drift for noise as Six Times the Standard Deviation

Wander is determined as the peak-to-peak noise of the mid-data values in the ASTM noise cycles, see "Noise Calculation by the ASTM Method" on page 98.

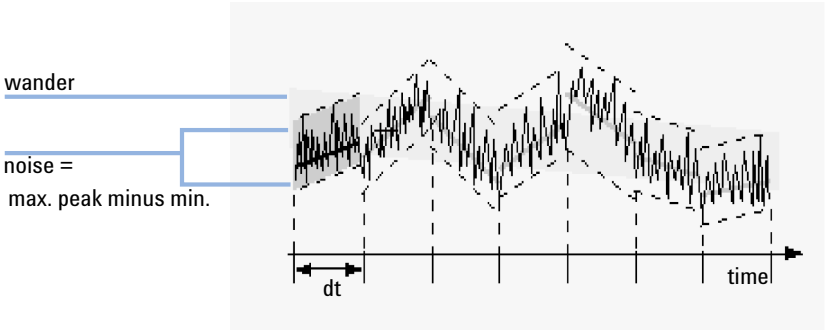

Figure 35 Wander of noise as determined by the ASTM Method

Calculation of Peak Symmetry

The ChemStation does not determine the asymmetry ratio of a peak, usually done by comparing the peak half-widths at 10% of the peak height, or 5% as recommended by the FDA.

Peak symmetry is calculated as a pseudomoment by the integrator using the following moment equations:

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

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

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

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

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

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

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

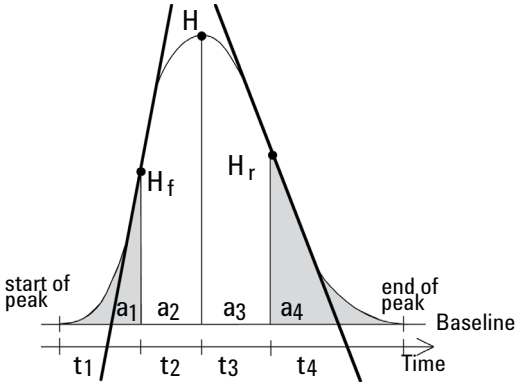

Figure 36 Calculation of the Peak Symmetry Factor

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

System Suitability Formulae and Calculations

The ChemStation uses the following formulae to obtain the results for the various System Suitability tests. The results are reported using the **Performance**, **Performance+Noise**, **Performance+LibSearch**, and **Extended Performance** report styles.

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 682 – 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, 15th Edition*

General Definitions

Void Volume

$$V = d^2 \pi l (f / 4)$$

where:

d = diameter of column [cm]

π = constant, ratio of circumference to diameter of a circle

l = length of column [cm]

f = fraction of column volume that is not taken up by stationary phase but available for mobile phase; default value for f = 0.68 (for Hypersil)

Retention Time of Unretained Compound t (m) [min]

(Also referred to as dead time or void time)

$$T_m = V / F$$

where:

F = flow rate of LC [ml/min]

Performance Test Definitions

Peak Performance can be calculated for any integrated peak of the data loaded, and also for new manually integrated peaks. The Interactive Peak Performance tool calculates peak characteristics and display it on UI.

Statistical Moments

$$M0 = d_t \cdot X$$

$$M1 = t_0 + d_t \cdot \frac{X}{Y}$$

$$M2 = \frac{d_t^2}{X} \cdot \sum_{i=1}^{N} \left(\left(i - 1 - \frac{Y}{X} \right)^2 \cdot A_i \right)$$

$$M3 = \frac{d_t^3}{X} \cdot \sum_{i=1}^{N} \left(\left(i - 1 - \frac{Y}{X} \right)^3 \cdot A_i \right)$$

$$M4 = \frac{d_t^4}{X} \cdot \sum_{i=1}^{N} \left(\left(i - 1 - \frac{Y}{X} \right)^4 \cdot A_i \right)$$

where:

N = Number of area slices

A_i = Value (Response) of area slice indexed by i

d_t = Time interval between adjacent area slices

t_0 = Time of first area slice

$\sum_{i=1}^{N}$ = Sum of starting index 1 to final index N for discrete observations

$$X = \sum_{i=1}^{N} (A_i)$$

$$Y = \sum_{i=1}^{N} ((i-1)A_i)$$

Statistical Moments, Skew and Excess

Statistical moments are calculated as an alternative to describe asymmetric peak shapes. There is a infinite number of peak moments, but only the first five are used in connection with chromatographic peaks. These are called 0^{th} Moment, 1^{st} Moment, ... 4^{th} Moment.

The 0^{th} Moment represents the peak area.

The 1^{st} Moment is the mean retention time, or retention time measured at the center of gravity of the peak. It is different from the chromatographic retention time measured at peak maximum unless the peak is symmetrical.

The 2^{nd} Moment is the peak variance which is a measure of lateral spreading. It is the sum of the variance contributed by different parts of the instrument system.

The 3^{rd} Moment describes the vertical symmetry or skew. It is a measure of the departure of the peak shape from the Gaussian standard. The skew given additionally in the Performance & Extended report is its dimensionless form. A symmetrically peak has a skew of zero. Tailing peaks have positive skew and their 1. Moment is greater than the retention time. Fronting peaks have negative skew and their 1. Moment is less than the retention time.

The 4^{th} Moment or excess is a measure of the compression or stretching of the peak along a vertical axis, and how this compares to a Gaussian standard for which the 4. Moment is zero. It can be visualized by moving in or pulling apart the sides of a Gaussian peak while maintaining constant area. If the peak is compressed or squashed down in comparison, its excess is negative. If it is

taller, its excess is positive. Also the excess is given in the Performance & Extended report in its dimensionless form.

True Peak Width W_x [min]

W_x = width of peak at height x % of total height

W_B	base width, 4 sigma, obtained by intersecting tangents through the inflection points with the baseline (tangent peak width)
$W_{4.4}$	width at 4.4% of height (5 sigma width)
$W_{5.0}$	width at 5% of height (tailing peak width), used for USP tailing factor
$W_{50.0}$	width at 50% of height (true half-height peak width or 2.35 sigma).

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]

Tailing Factor (USP) t

NOTE Symmetry Factor (JP) and Symmetry factor (EP) S are identical with the Tailing Factor (USP). All are available as "Peak_TailFactor" in Intelligent Reporting. See also "Reporting of Pharmacopoeia factors in ChemStation" on page 154.

$$t = \frac{W_{5.0}}{t_w \cdot 2}$$

where:

t_w = distance in min between peak front and T_R, measured at 5% of the peak height

$W_{5.0}$ = peak width at 5% of peak height [min]

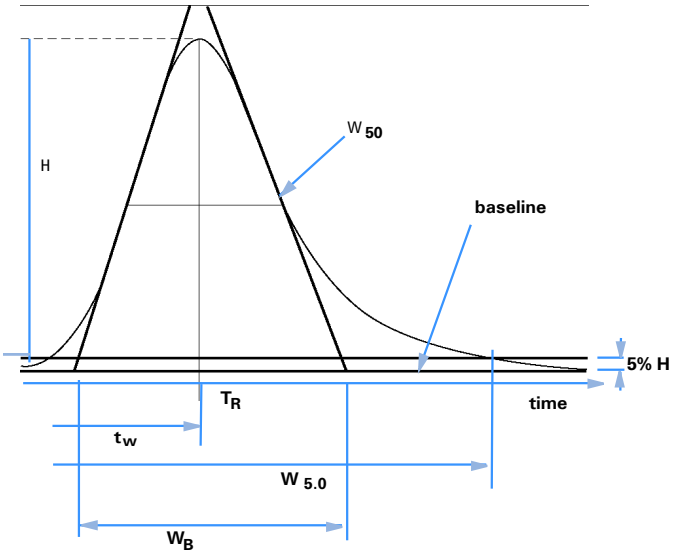

Figure 37 Performance Parameters

Number of Theoretical Plates per Column n

Tangent method (USP, ASTM):

$$n = 16\left(\frac{T_R}{W_B}\right)^2$$

where:

W_B = base width [min]

Half-width method (ASTM, EP, JP):

$$n = 5.54\left(\frac{T_R}{W_{50}}\right)^2$$

where:

W_{50} = peak width at half-height [min]

5 Sigma method:

$$n = 25\left(\frac{T_R}{W_{4.4}}\right)^2$$

where:

$W_{4.4}$ = peak width at 4.4% of peak height [min]

Statistical method:

$$n = \frac{M1^2}{M2}$$

where:

Mx = x th statistical moment

Foley Dorsey method

The Foley-Dorsey equation is used for asymmetrical peaks. It corrects plate count for peak tailing and broadening.

$$N_{sys} = \frac{41.7(T_R/W_{10})^2}{A/B + 1.25}$$

Where

- W_{10} = peak width at 10% peak height
- A/B = empirical asymmetry factor, with A+B = W_{10}, and A: froning and B: tailing

Number of Theoretical Plates per Meter N [1/m]

$$N = 100 \times \frac{n}{l}$$

where:

n = number of theoretical plates

l = length of column [cm]

Relative Retention (USP, ASTM), Selectivity Alpha

NOTE Relative retention (USP) is available as "Selectivity" or "Peak_Selectivity" in reporting.

(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

Relative Retention (EP, JP)

Relative Retention (adjusted) according to EP and the Separation Factor according to JP are calculated using the same formula:

$$r = \frac{t_{Ri} - t_M}{t_{Rst} - t_M}$$

Where

t_{Ri} = retention time of the peak of interest

t_{Rst} = retention time of the reference peak

t_M = hold-up time

Relative Retention (adjusted, EP) and Separation Factor (JP) are available as “RelativeRetTime_EP” in Intelligent Reporting and as “Selectivity” in classic reporting.

Relative Retention (unadjusted) according to EP is calculated as

$r_G = t_{Ri} / t_{Rst}$

Resolution (USP, ASTM) R

(Pertaining to peaks a and b, T_R of peak a < T_R of peak b; T_R in min)

Tangent method (USP, ASTM):

$$R = \frac{2(T_{R(b)} - T_{R(a)})}{W_{B(b)} + W_{B(a)}}$$

Resolution (EP/JP) Rs

Half-width method (Resolution used in Performance Report):

ChemStation calculates Resolution (JP) and Resolution (EP) with the following definition:

$Rs = 1.18 \times (t_{R2} - t_{R1}) / (W_{0.5h1} + W_{0.5h2})$

NOTE

The definition of Resolution in USP differs from the definition in the European (EP) and the Japanese (JP) Pharmacopoeias. EP and JP calculations are available since ChemStation Edition C.01.04.

In addition Classic Resolution (2.35/2)*... is available for in Intelligent Reporting as Peak_Resolution_Classic. For a complete list of the values see "Reporting of Pharmacopoeia factors in ChemStation" on page 154

Resolution (ChemStation classic definitions)

Half-width method:

$$R = \frac{(2.35/2)(T_{R(b)} - T_{R(a)})}{W_{50(b)} + W_{50(a)}}$$

5 Sigma method:

$$R = \frac{2.5(T_{R(b)} - T_{R(a)})}{W_{4.4(b)} + W_{4.4(a)}}$$

Statistical method:

$$R = \frac{M1_{(b)} - M1_{(a)}}{W_{S(b)} + W_{S(a)}}$$

where:

$M1_{(x)}$ = mean retention time for peak x (1st Statistical Moment) [min]

$W_{B(x)}$ = base width for peak x [min]

$W_{4.4(x)}$ = width at 4.4% height for peak x [min]

$W_{50(x)}$ = width at 50% height for peak x [min]

$W_S(x)$ = width derived from statistical moments = $\sqrt{(M2)}$ for peak x [min] (see also "Statistical Moments" on page 108)

Definitions for Reproducibility

For the statistical review of analytical data in terms of reproducibility the sequence is considered as a small random sample taken out of an infinite number of possible experimental results. To accomplish a complete set of results, an unlimited amount of sample material as well as time would be required. Strictly statistical data does only apply to a complete self-contained set or population of data. Therefore a prerequisite for such a treatment is that the selected sample can be assumed as representative for all data.

Sample Mean M

The mean value M of a random sample consisting of N measurements is calculated from this limited set of N single observed values X_i indexed with a consecutive counter i according to the formula:

$$M = \frac{\sum_{i=1}^{N} X_i}{N}$$

where:

N = number of discrete observations

X_i = value of discrete observations indexed by i

Sample Standard Deviation S

Consider a random sample of size N. The sample standard deviation S for the selected finite sample taken out of the large population of data is determined by

$$S = \sqrt{\frac{\sum_{i=1}^{N} (X_i - M)^2}{N - 1}}$$

The sample standard deviation S differs in two points from the standard deviation s for the whole population:

- instead of the real mean value only the sample mean value M is used and
- division by N-1 instead of N.

Relative Standard Deviation RSD[%] (USP)

The relative standard deviation is defined as

$$RSD = 100\frac{S}{M}$$

Standard Deviation of the Mean S_M

Let M be the sample mean and S the sample [or (N-1)] standard deviation. The standard deviation S_M of the sample mean M is determined by

$$S_M = \frac{S}{\sqrt{N}}$$

This can be further illustrated by an example:

While the retention time of a certain compound may deviate slightly from the calculated mean value during one sequence, the data from another sequence may differ much more due to e.g. ambient temperature changes, degradation of the column material over time etc. To determine this deviation the standard deviation of the sample mean S_M can be calculated according to the above formula.

Confidence Interval CI

The confidence interval is calculated to give information on how good the estimation of a mean value is, when applying it to the whole population and not only to a sample.

The 100 × (1 - α) % confidence interval for the overall mean is given by

$$CI = t_{(\alpha/2);N-1} \cdot S_M$$

where:

$$t_{(\alpha/2);N-1}$$

percentage point of the t distribution table at a risk probability of α

For the extended statistics in the sequence summary report the 95% confidence interval may be used (α = 0.05).

The t distribution (or 'student distribution') must be used for small sample volumes. In case of large sample volumes the results for the t distribution and the normal (gaussian) distribution do not differ any more. Therefore in case of 30 or more samples the normal distribution can be used instead (it would be very difficult to calculate the t-distribution for large numbers, the normal distribution is the best approximation of it).

95% Confidence Interval for 6 samples:

1 - α = 0.95

N = 6

The correct value for t has to be taken from the t distribution table for 5 (N-1) degrees of freedom and for the value α/2, being 0.025. This would give the following calculation formula for CI:

$$CI = 2.571 \cdot \frac{1}{\sqrt{6}} \cdot S_M$$

Regression Analysis

Let

N = number of discrete observations

X_i = independent variable, i^{th} observation

Y_i = dependent variable, i^{th} observation

Linear function:

$$y_{(X)} = a + bX$$

Coefficients:

$$a = \frac{1}{\Delta_X}\left(\sum_{i=1}^{N} X_i^2 * \sum_{i=1}^{N} Y_i - \left(\sum_{i=1}^{N} X_i * \sum_{i=1}^{N} X_i Y_i\right)\right)$$

$$b = \frac{1}{\Delta_X}\left(N * \sum_{i=1}^{N} X_i Y_i - \left(\sum_{i=1}^{N} X_i * \sum_{i=1}^{N} Y_i\right)\right)$$

where:

$$\Delta_X = N * \sum_{i=1}^{N} X_i^2 - \left(\sum_{i=1}^{N} X_i\right)^2$$

Regression Coefficient

$$r = \frac{\left(N * \sum_{i=1}^{N} X_i Y_i - \sum_{i=1}^{N} X_i * \sum_{i=1}^{N} Y_i\right)}{\sqrt{\Delta_x * \Delta_y}}$$

where:

$$\Delta_Y = N * \sum_{i=1}^{N} Y_i^2 - \left(\sum_{i=1}^{N} Y_i\right)^2$$

Standard Deviation (S)

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

Internally Stored Double Precision Number Access

For validation purposes, it might become necessary to manually recalculate the ChemStation results such as calibration curves, correlation coefficients, theoretical plates, etc. When doing so the number format used in the ChemStation has to be taken into account.

For all numbers stored internally within the ChemStation, the “C” data type DOUBLE is used. This means that 14 significant digits are stored for each number. The implementation of this data type adheres to the Microsoft implementation of the IEEE standard for “C” data type and the associated rounding rules (see Microsoft documents Q42980, Q145889 and Q125056).

Due to the non-limited number of parameters that might be used for the calculation of the calibration table, it is not possible to calculate the exact error possibly introduced by the propagation and accumulation of rounding errors. Thorough testing with different calibration curve constructions however has shown that the accuracy of up to 10 digits can be guaranteed. Whereas the area, height and retention time repeatability of a chromatographic analysis usually has 3 significant digits, 10 significant digits within the calculations is sufficient. For this reason, the calibration, and other tables, display a maximum of 10 significant digits.

If an external (manual) calculation for validation is required, it is recommended that all digits used for the internal calculations are utilized. Using the displayed and/or rounded data for the external calculations might give results differing from the ChemStation due to rounding errors.

The following paragraph describes how to access all internally stored digits for numbers typically required for manual calculations. In all cases, a data file must be loaded and reported with the appropriate report style prior to execution of the listed command. All commands are entered on the ChemStation command line which may be enabled from the view menu. The information in file “C:\CHEM32\TEMP.TXT” may be viewed using NOTEPAD or a suitable TEXT editor.

Raw Peak Information:

- Retention Time

- Area
- Height
- Width (integrator)
- Symmetry
- Peak Start Time
- Peak End Time

Use Command Line Entry:

```
DUMPTABLE CHROMREG, INTRESULTS,"C:\CHEM32\1\TEMP\
INTRES.TXT"
```

Processed Peak Information:

- Measured Retention Time
- Expected Retention Time
- Area
- Height
- Width (integrator)
- Symmetry
- Half Width - Half Peak Height (Performance & Extended Performance)
- Tailing Factor (Performance & Extended Performance)
- Selectivity (Performance & Extended Performance)
- K` (Extended Performance)
- Tangent Peak Width (Extended Performance)
- Skew (Extended Performance)
- Theoretical Plates - Half Width (Performance & Extended Performance)
- Theoretical Plates - Tangent (Extended Performance)
- Theoretical Plates – 5-Sigma (Extended Performance)
- Theoretical Plates - Statistical (Extended Performance)
- Resolution - Half Width (Performance & Extended Performance)
- Resolution - Tangent (Extended Performance)
- Resolution – 5-Sigma (Extended Performance)
- Resolution - Statistical (Extended Performance)

Use Command Line Entry:

```
DUMPTABLE CHROMRES, PEAK,"C:\CHEM32\1\TEMP\PEAK.TXT"
```

Processed Compound Information:

- Calculated Amount

Use Command Line Entry:

DUMPTABLE CHROMRES, COMPOUND,"C:\CHEM32\1\TEMP\
COMPOUND.TXT"

Calibration Table Information:

- Level Number
- Amount
- Area
- Height

Use Command Line Entry:

```
DUMPTABLE _DAMETHOD, CALPOINT,"C:\CHEM32\1\TEMP\CALIB.TXT"
```

Linear Regression Information:

- Y-Intercept (CurveParm1)
- Slope (CurveParm2)
- Correlation Coefficient

Use Command Line Entry:

```
DUMPTABLE _DAMETHOD, PEAK,"C:\CHEM32\1\TEMP\REGRESS.TXT"
```

7
CE specific Calculations

This chapter is relevant only if you use ChemStation to control CE instruments.

Calibration Tables

Four different calibrations types are available in the drop-down list for your Calibration Table.

Standard Calibration

Standard Calibration is based on peak area or peak height. When you select **Standard Calibration** you have the option to **Calculate Signals Separately** or **Calculate with Corrected Areas**.

Calculate Signals Separately is selected when you want to ensure that, in the calculation of Norm% reports, the amount percent of separately reported signals add up to 100% for each signal. When **Calculate signals separately** is deselected, the amount percent of all signals add up to 100%. Selecting **Calculate signals separately** is a prerequisite for sorting by signal in the calibration table.

Select **Calculate with Corrected Areas** to make a correction to the peak area based on the migration time. In this mode, the area is divided by the migration time which can improve reproducibility in quantitative analysis when migration times are unstable.

In addition to the Standard Calibration, there are 3 capillary electrophoresis specific calibrations that are migration time based on signal. The signal is defined by the signal description in the calibration method. If the data file contains multiple signals, then only one signal can be selected and is extracted from the data file. The format of the calibration table is dependent on the calibration type selected.

Quantitation tasks may be performed based on biopolymer size calibration (Ferguson Plot) for SDS-Protein.

Protein Molecular Weight Calibration

The **Protein molecular weight calibration** requires a calibration standard with components of known molecular weights and a reference peak. The calibration equation is:

$$\log(MW) = k_1 \cdot (t_{ref}/t) + k_0$$

where:

MW is the molecular weight

t_{ref} is the migration time of the reference peak

t is the migration time

k_0 and k_1 are the coefficients of the linear equation

The calibration table contains the Name, Migration Time, t_{ref}/t (relative migration time), Molecular Weight and *log(MW)* for each component.

DNA Base-Pair Calibration

The **DNA base-pair calibration** is similar to the **protein molecular weight calibration**, but operates without a reference peak; it requires a calibration standard with a known number of base pairs. The calibration equation is:

$$\log(\#BP) = k_1 \cdot 1/t + k_0$$

where:

#BP is the number of base pairs

t is the migration time

k_0 and k_1 are the coefficients of the linear equation

The calibration table contains the Name, Migration Time, *1/t*, Base Pairs and *log(Base Pairs)* for each component.

Capillary Isoelectric Focusing

The **capillary isoelectric focusing calibration** (cIEF) requires a calibration standard with standard proteins of known isoelectric points (pI). The calibration equation is:

$$pI = k_1 \cdot t + k_0$$

where:

pI is the isoelectric point

t is the migration time

k_0 and k_1 are the coefficients of the linear equation

The calibration table contains the Name, Migration Time and *pI* (isoelectric point) for each component.

Calibration using Mobility Correction

Introduction

Slight changes in buffer composition, run temperature or viscosity, as well as adsorption to the capillary wall, can influence the Electro Osmotic Flow (EOF) and cause it to be unstable. The resulting change in the EOF can create a rather high standard deviation of migration times. Corrections for mobility can significantly reduce the effect of run-to-run migration time shifts by monitoring the migration time of a mobility reference peak and in turn significantly increasing the migration time reproducibility.

The mobility reference peak should be chosen with the following priorities:

- Select peak with the highest signal
- Select the most isolated peak
- The EOF marker or internal standard can also be used as the mobility reference peak
- Enlarge the search window to always find the mobility reference peak
- If several peaks fall in the search window, the peak with the highest signal is automatically chosen as the mobility reference peak.

There are two mobility correction types available:

Effective Mobility Correction

Effective Mobility Correction uses the effective mobilities of all peaks and requires the availability of the voltage ramp data together with the electropherogram. In addition, working with effective mobility correction allows the true effective mobilities for all sample components to be determined.

Relative Mobility Correction

Relative Mobility Correction can operate in the absence of voltage data and would then assume a constant voltage for all measurements.

Effective Mobility Calculations

In addition to a reference peak the requirements for effective mobility correction include a neutral marker which corresponds to the velocity of the EOF. Some commonly used markers and their associated wavelengths are:

Table 9 Commonly Used EOF Markers

Compound	Wavelength
1-Propanol	210nm
Acetone	330nm
Acetonitrile	190nm
Benzene	280nm
Guanosine	252nm
Mesityl oxide	253nm
Methanol	205nm
Phenol	218nm
Pyridine	315nm
Tetrahydrofuran	212nm
Uracil	259nm

Voltage over time data and the capillary dimensions are either saved with the data file or they can manually be entered during the calibration table setup. Storing the voltage data during the run does this most accurately. Make sure to also store the capillary dimensions with the method. To reprocess signals that have been acquired without voltage data/capillary dimensions, enter the voltage and ramp time manually in the “Voltage and Capillary Dimensions” group of the dialog box.

From the data the effective mobility for each component is determined.

General

The apparent mobility of a sample peak is defined by the equation:

$$\mu_{app} = (l \cdot L)/(t \cdot V(t))$$

where

l is the effective length of the capillary (the length from the point of injection to the point of detection)

L is the total capillary length

$V(t)$ is the average voltage from time 0 to the migration time t of the peak

The average voltage is calculated from either the measured voltage or from the voltage ramp specified in the method using the following equations:

If $t < t_R$ then

$$V(t) = V/(2 \cdot t_R) \cdot t$$

If $t > t_R$ then

$$V(t) = V \cdot (1 - t_R/(2 \cdot t))$$

where

t is the migration time of the peak

t_R is the ramp time

V is the end voltage

The equation for mobility can be simplified by introducing a coefficient:

$$k(t) = (l \times L)/V(t)$$

The relative or apparent mobility is then

$$\mu_{app} = k(t)/t$$

Effective or real mobility is

$$\mu_{real} = \mu_{app} - \mu_{EOF}$$

where

$\propto_{app}$ is the apparent mobility of any peak

$\propto_{EOF}$ is the apparent mobility of a neutral marker

Components with lower velocity than the EOF (usually anions) will result in negative values for the effective mobility.

Calibration

The real mobility of a sample peak to be used as the mobility reference peak in future measurements is calculated using the migration time of a neutral marker (μ_{EOF}):

$$\mu_{real ref} = \mu_{appref} - \mu_{EOF} = k(t_{ref})/t_{ref} - k(t_{EOF})/t_{EOF}$$

The effective mobilities of all peaks are then calculated and stored as expected mobilities:

$$\mu_{real N} = \mu_{appN} - \mu_{EOF} = k(t_N)/t_N - k(t_{EOF})/t_{EOF}$$

The calibration table then contains the measured migration time and the calculated real mobility for each compound in the columns for the expected migration time and the expected mobility.

Mobility Calculation

The actual value of μ_{EOF} is calculated using the Mobility Reference Peak:

$$\mu_{EOFact} = \mu_{appref} - \mu_{real ref} = k(t_{ref})/t_{ref} - \mu_{real ref}$$

The expected migration time for each peak is then adjusted:

$$t_{new\exp N} = k(t_{old\exp N})/(\mu_{real N} + \mu_{EOFact})$$

The calculated values are used for peak identification and replace the values within the calibration table.

Recalibration

The migration time of the mobility reference peak is used to calculate the actual value of μ_{EOF} :

$$\mu_{EOFact} = \mu_{appref} - \mu_{real ref} = k(t_{ref})/t_{ref} - \mu_{real ref}$$

The expected migration time for every peak is adjusted:

$$t_{new\exp N} = k(t_{old\exp N})/(\mu_{real N} + \mu_{EOFact})$$

and the mobilities are updated:

$$\mu_{real N} = \mu_{appN} - \mu_{EOFact}$$

During a calibration run the expected values for the migration time as well as the real mobility values are updated in the calibration table.

Relative Mobility Calculations

Migration time correction based on relative mobilities can also be performed. In this case neither an EOF marker, voltage, nor capillary dimensions are needed. The software still corrects migration time shifts but does not display mobility values.

General

Just as in the effective mobility calculations, the coefficient

$$k(t) = (l \cdot L)/V(t)$$

is used in the relative mobility calculations to describe the relationship between mobility and migration time:

$$\mu_{app} = k(t)/t$$

The difference is that in the equations for Relative Mobility, k appears in both numerator and denominator of a fraction; this means that the capillary dimension can be eliminated. The factor k is calculated as

$$k(t) = 1/V(t)$$

where *V(t)* is the average voltage from time 0 to the migration time of the peak.

When the voltage parameter is set to **Ignore**, *k* is a constant and can be eliminated from the equations for the expected migration time (see below).

The following equations describe the general case for *k* = *k(t)*, although the software takes all cases into account when calculating *k*.

Calibration

A mobility reference peak is identified and its migration time (t_{refcal}) is stored. The expected migration times (t_{expcalN}) of all other peaks are saved.

Mobility Calculation

After detection of the reference peak, the expected migration time for each peak is adjusted according to the actual migration time of the mobility reference peak:

$$t_{new\,\mathrm{exp}N} = \frac{k(t_{old\,\mathrm{exp}N})}{(k(t_{\mathrm{exp}calN})/t_{\mathrm{exp}calN} - k(t_{refcal})/t_{refcal} + k(t_{refact})/t_{refact})}$$

Then, the migration time of the reference peak from the last calibration run is updated:

$$t_{refcal} = t_{refact}$$

Special Report Styles for Capillary Electrophoresis

The following report style has been added to the Agilent ChemStation for CE systems:

CE Mobility

CE Mobility comprises quantitative text results, especially the apparent mobility. To use this report style, you need to supply the information on the capillary used before acquisition and you store the voltage signal. The apparent mobility is calculated according to the following formula.

$$\mu_{app} = \frac{l \cdot L}{t \cdot V}$$

Where

l is effective capillary length (cm)

L is total capillary length (cm)

t is migration time (min)

V is voltage (kV)

If effective mobility correction (see "Effective Mobility Calculations" on page 130) is activated, the peak type column in simple reports (external standard reports for example) is replaced by a mobility column. The CE mobility report prints effective instead of apparent mobilities.

Corrected Peak Areas

The Agilent ChemStation for CE systems allows you to use corrected peak areas instead of the normal area calculations. These areas are used in standard calibration and reports.

To activate this feature, select **Calculate with Corrected Areas** to make a correction to the peak area based on the migration time. In this mode, the area is divided by the migration time which can improve reproducibility in quantitative analysis when migration times are unstable.

The corrected area is calculated according to the following formula:

$$A_c = \frac{A}{60 \cdot t}$$

Where

Ac is corrected peak area (mAU)

A is peak area (mAU·sec)

t is migration time (min)

This corrected area is sometimes also referred to as normalized area.

System Suitability for Capillary Electrophoresis

Capacity Factor k'

In capillary electrophoresis the capacity factor k' value can't be calculated automatically for all operation modes. Refer to the manual *High Performance Capillary Electrophoresis:* A Primer for the formulas respectively. The values listed in the reports are only valid for the Agilent ChemStation for LC 3D systems since the Agilent ChemStation for CE systems uses the same algorithms as the Agilent ChemStation for LC 3D systems.

CE-MSD

Background Subtraction

When you select the **Subtract Background** (BSB) menu item, the most recently selected mass spectrum is subtracted from each point in the current electropherogram. The resulting data is saved in the same directory and with the same name as the original data file; however, the file extension is changed to .BSB.

The new data file becomes the current data file and the background subtracted electropherogram is displayed. A record of the number of background subtractions that have been performed is kept in the Operator item of the data file header.

If you view a tabular listing of BSB data, you may observe differences due to the precision of data representation.

NOTE The HELP text files in the LC/MSD refer only to LC parameters and not CE. Some features that are available in the LC/MSD software are either not available or not applicable to CE/MSD applications but are used in LC. The function **peak matching** is not applicable for CE-MS and is therefore not active. In CE-MS, UV and MS detection occurs at different effective lengths of the separation capillary. Because of the different resolution at different effective lengths, peak matching is not possible.

8
Data Review, Reprocessing and Batch Review

This chapter describes the possibilities to review data and how to reprocess sequence data. In addition it describes the concepts of Batch Review, Batch configuration, review functions, and batch reporting.

Navigation Table in Data Analysis

The **Data Analysis** view includes a Navigation Table that is designed to facilitate navigation through data files. The Navigation Table shows the runs contained in a selected data or sequence data subdirectory. You can use the **Navigation Table** to load individual runs, or to automatically scroll through the loaded signals. For more details, please refer to the *OpenLAB CDS ChemStation Edition, Basis Concepts and Workflows* manual.

Navigation Table Configuration

The Navigation table shows the data file information depending on the available data sets. The Navigation Table is read-only and the values in the Navigation table cannot be overwritten.

Table 10 Navigation Table Columns

Single Runs Columns	Sequence Runs Columns
Overlay	Overlay
ECM	ECM
TYPE	TYPE
Date / Time	Line
Operator	Inj (Injection)
Vial	Vial
Data File	Sample Name
Sample Name	Method Name
Method Name	Sample Type
Manual Events	Manual Events
Sample Info	Cal Level (Calibration Level)
Sample Amount	Sample Info

Table 10 Navigation Table Columns

Single Runs Columns	Sequence Runs Columns
ISTD Amount	Sample Amount
Multiplier	ISTD Amount
Dilution	Multiplier
---	Dilution
---	Data File

The Navigation table includes standard table configuration features, such as sorting and drag-and-drop options to move columns to different places. It is also possible to select the columns that are displayed in the Navigation Table.

In addition, column-specific grouping is possible, for example, single runs of a particular operator can be displayed by grouping the loaded files by the column **operator**.

The Navigation table offers right mouse click functions to load a signal, overlay a signal, export data, print reports, view the acquisition method parameters etc. Each Navigation Table line can be expanded by clicking the + (plus) sign at the left of the line to configure signal-specific options:

- **Signal**: Lists the acquired signals and allows you to specify the signals to be loaded. The signal display selection is applied to each run individually.
- **General Info**: Lists the header details about the run.
- **Instrument curves**: Allows you to select the instrument data curves to be displayed along with the chromatogram/electropherogram on screen and in the printout.

Navigation Table Toolbar

The **Navigation Table** includes two toolsets that allow you either to review single run/sequence data, or to reprocess sequence data.

Data Review Toolset

The review functionality of the Navigation Table allows you to step automatically or manually through the loaded signals. Depending on the selection specified in the **Preferences / Signal/Review** Options, the system can automatically integrate the signal and print a report for each file as it is loaded. The method applied to the data file is shown in the top menu.

Sequence Reprocessing Toolset

The sequence reprocessing toolset is available only when a sequence acquired with ChemStation B.02.01 or higher is loaded that was acquired with **Unique Folder Creation** switched on. It is possible to start, stop and pause the reprocessing of the sequence. In addition, the toolbar gives access to the following dialog boxes in order to change parameter for reprocessing sequences and printing:

- **Sequence Table** (copy of the original *.s template, located in the sequence data container)
- **Sequence Parameters** dialog box
- **Sequence Output** dialog box
- **Sequence Summary Parameters** dialog box
- **Extended Statistic Parameters** dialog box
- **Save Current Sequence**
- **Print Current Sequence**

Data Review Using the Navigation Table

You review your data in the Recalculate Mode, which is accessed by clicking  in the Navigation Table toolbar. This opens the Recalculate Mode toolset. Depending on your required workflow, you can review data in one of three ways:

1 Review your data using the data analysis method stored with each data file (sequence data B.02.01 or higher). Select **Start Autostepping** from the **Recalculate** menu in **Data Analysis** mode to have the system load the individual data analysis method stored with the data file before loading data file. As each line in the Navigation Table is accessed during the data

review process, the linked method for the selected data file is loaded and used for reviewing and generating the report.

2 Review your data using a different method. If you want to use a different method for reviewing the data than the method stored with the data file, select **With method** from the **Recalculate** menu in **Data Analysis** mode. In this case, you select a method and report template from the **Recalculate with Method** dialog box. You can also specify an **Autostep interval** and a report **Destination**; the values you select in this dialog box temporarily override the values in the **Signal/Review Options** tab of the **Preferences** dialog box, and are reset when the ChemStation session is terminated. The selected method is loaded and used to calculate results from all data files in the result set.

3 Review your data using the method that was last used to calculate the results. Select **Last Result Mode** from the **Recalculate** menu in **Data Analysis** mode. This mode loads the method that was last used to calculate the results for the data file. Note that if the data file has no corresponding data analysis method, it is skipped during autostepping. This mode affects both autostepping and the manual loading of data files.

Sequence Reprocessing Using the Navigation Table

NOTE

Sequence data acquired with ChemStation revisions up to B.01.03 need to be reprocessed using the **reprocess** option in the **Method and Run Control** view. The same applies to data acquired in B.03.01 or later when **Unique Folder Creation** is switched off.

Sequence data acquired with ChemStation revisions B.02.01 and higher need to be reprocessed using the reprocessing toolset in the **Data Analysis** Navigation Table.

For reprocessing using the Navigation Table in **Data Analysis**, all necessary files are present in the sequence data container:

- sequence data files (*.d)
- all methods (*.m) files used during the sequence
- copy of the original sequence template (*.s)
- sequence-related batch (*.b) file
- sequence-related logbook (*.log) file

During reprocessing, the individual methods DA.M for the data files are updated as well as the batch file (*.b) file.

With the **Data Analysis** reprocessing functions, it is possible to modify the sequence template (*.s) in the data container in order to change the multiplier, dilution etc., or to use a different method for reprocessing. By default, the Data Analysis reprocessing sequence parameter **parts of method to run** is set to **Reprocessing only**, and the option **Use Sequence Table Information** is checked. These predefined default values enable you to change the parameters in the sequence template and run a reprocess without editing the **Data Analysis** sequence parameters again.

If you have not explicitly changed the method in the sequence template, the system uses the sequence methods stored in the sequence data container to reprocess the sequence. These methods are the original methods used during data acquisition. If particular method parameters need to be changed (for example, specify to print to a *.xls file), the methods in the sequence container need to be modified and saved. This general change is then applied to all data files during reprocessing.

If you now want to use the updated sequence container method for further data acquisition, you need to copy this method from the sequence data container to one of the defined method paths. The new/updated method is then available in the ChemStation Explorer in the method view as a master method.

What is Batch Review?

Batch review is a capability within data analysis designed to help an analyst perform a “first-pass” review of the results of a sequence or a selection of runs quickly and easily. It will save time especially when reprocessing large numbers of samples. Whenever a sequence is run, a batch file (with a .b extension) is automatically generated and placed in the data directory along with the data files. This batch file contains pointers to the data files in the batch review itself. Upon loading a batch, the analyst need only select the method to use for the batch, and then individually select the desired data files to analyze in the batch. One can check the calibration accuracy, instrument performance and individual integrations before approving the results. Any chromatogram specific integration parameters which are modified can be stored with the data file for data traceability. This interactive environment provides complete access to all other data analysis capabilities, such as peak purity, library searching, etc., as well.

Batch review uses the same data analysis registers (ChromReg and ChromRes) as the standard data analysis and should therefore not be used in an online session that is currently performing analyses.

Enabling Batch Review Functionality when using OpenLAB CDS with ECM

When using OpenLAB CDS with ECM, the Batch Review functionality is not available by default. In order to use Batch Review, this functionality has to be enabled by an entry in the [PCS] section of the ChemStation.ini file. The file is located in the windows directory c:\ WINDOWS.

```
[PCS] _BatchReview=1
```

The default entry, _BatchReview=0, turns off the functionality.

Batch Configuration

A batch is a user-selected series of data files that is processed using a user-defined method. All data files in the batch are processed using the same method. The processing steps carried out each time a new sample is loaded for review can be selected (integration, identification/quantitation, reporting).

All calibration runs in the batch are used to produce a single calibration table, using averaged response factors, which is then used for quantification.

Batch Table

Runs are displayed in a user-configurable batch table:

- the number and content of the table's columns can be specified;
- the runs can be sorted by
  - run index (the order in which the runs were acquired) independent of any other criteria,
  - sample type (control samples first, then calibration samples, then normal samples) then by run index within each sample type,
  - method (if more than one method was used to acquire the runs) then by run index within each method;
- samples, calibration samples and control samples can be displayed in the table or hidden.

Each selected run occupies a line in the batch table. You can exclude a run in the batch table (e.g. from calibration) by changing its sample type to Removed.

Compound Table

The compound results are displayed in a user-configurable compounds table, but contents of the compounds table depends on the type of samples in the batch table:

- the compound list contains all compounds found in the method that was loaded for batch review.
- if calibration samples only are displayed in the batch table (samples and control samples are hidden), the compound table shows additional columns for calibration-related information (expected amount, relative error and absolute error).
- if control runs only are displayed in the batch table (samples and calibration samples are hidden), the compound table shows additional columns for any defined control limits.

For columns containing compound-specific information, you can include the name of the compound in the table title by adding %s to the column specification.

Batch Report

The batch report contains two tables that are generally analogous to the batch table and the compound table; these tables are also user-configurable.

For columns containing compound-specific information, you can include the name of the compound in the table title by adding %s to the column specification. Multi-line headers are allowed; you insert the character '|' at the point where you want the line to break.

User Interface

Batch review provides a choice of two user interfaces:

- the standard interface includes a button bar, with buttons mirroring most of the Batch menu items, together with the batch table and compound table;
- a minimal interface provides a button bar similar to the standard interface, but replaces the batch table and compound table with a combo box that contains only the information specified for the batch table. The minimal interface button bar does not contain batch table-related or compound table-related buttons.

Review Functions

Data files can be displayed in one of two ways:

- manually, by selecting a run to display from the table,
- automatically, with a predefined interval between each data file. During automatic display, only those sample types displayed in the table are displayed; the runs are displayed in the order in which they appear in the table. The automatic review can be paused and later resumed, or stopped.

The standard functions provided by the ChemStation are available with batch review. This includes calibration, manual manipulation of chromatograms, for example by smoothing or manual integration. Any changes made to a data file can be marked and saved with the batch file. Chromatograms that have been reviewed are marked with an asterisk in the batch table. You can also discard changes made to either the current chromatogram only, or all changes to all chromatograms in the batch.

When a run is loaded, the selected processing options are performed; if the run has already been processed and the changes saved, the processed run is loaded. This is a faster process than loading the unprocessed run, because no processing needs to be done.

Calibration in Batch Review

Calibration in batch review works independently from the recalibration settings in the sequence table. The first step in batch calibration always replaces both response and retention time entries in the calibration table. For the following calibration standards, both response and retention time values are averaged.

Batch Reporting

The user-configurable “Batch Table” on page 147, can be printed directly on the printer, displayed on the screen or printed to a file with a user-specified prefix in one of the following formats:

- UNICODE text file with the extension .TXT
- Data Interchange Format with the extension .DIF
- Comma-Separated Values format with the extension .CSV
- Microsoft Excel format with the extension .XLS.

The reporting options also give the possibility of sorting the samples (by Run Index, Sample Type or Method) independent of the sort method in the batch table. The sort priorities are as for the “Batch Table” on page 147.

Batch History

Batch review maintains a log of all actions relating to the current batch. Any action that changes the batch (for example changing the displayed chromatogram, changing the sample type, loading and saving the batch) adds a line to the batch history with a date and time stamp, the current operator name and a description of the event.

You can also add your own comments to the batch history. Existing batch history entries cannot be edited, and the history list cannot be accessed except through the Batch History menu item.

9 Reporting

This topic explains and provides a reference to the ACAML scheme used in the intelligent reporting feature of the OpenLAB CDS software.

What is ACAML?

ACAML, *A*gilent *C*ommon *A*nalytical *M*arkup *L*anguage, is a markup language to capture and describe analytical data in the chromatography and spectrometry domain. ACAML is designed to describe all data types in analytical environments. ACAML is focused on providing an Agilent common standard that allows to seamlessly exchange information between various platforms and applications.

The approach is to define a technique- and application-independent unified schema-language. ACAML can be used to describe analytical data in a generic way, without any special aspects (e.g. result-centric viewpoint): starting from a single instrument or method up to a complex scenario with multiple instruments, methods, users and hundreds or thousands of samples.

No additional applications (like a special ACAML-validator) are required to handle and validate ACAML instance-documents. In it's initial revision ACAML supports chromatography data (LC, GC) only.

More information on ACAML can be found in the OLIR designer manual OpenLAB CDS Support DVD (disk no.6).

The ACAML schema

Base of the ACAML schema is the industrial XML standard.

The ACAML-schema is strong-typed:

- to support the idea of standardized data-exchange, and
- to avoid uncontrolled growth of self-defined types, which makes an automated further processing very complicated or impossible.

The schema-definition makes sure that each instance-document is well defined and the referential integrity between all objects is guaranteed. No additional applications (like a special ACAML-validator) are required to handle and validate ACAML instance-documents.

The Schema definition of the latest revision ACAML.1.4.xsd can be found on the OpenLAB CDS Support DVD (disk no.6).

Reporting of Pharmacopoeia factors in ChemStation

With ChemStation Edition C.01.04, the calculation of Pharmacopoeia factors has been completed. Factors from the peak table as defined in the USP, the EP, and JP are available for use in ChemStation reports. The table below provides an overview of the available factors, their definitions and the value names. For more details on the calculations please refer to the respective sections in this guide.

Table 11 Pharmacopoeia values in ChemStation reporting

USP	EP	JP	Definition	classic reporting (RLE)	intelligent reporting (RTE)
Tailing Factor	Symmetry factor	Symmetry factor	$S = W_{0.05h}/2f$	USP Tailing	Peak_TailFactor
-	Relative retention(adjusted)	Separation factor	$r = (t_{R2}-t_0)/(t_{R1}-t_0)$		RelativeRetTime_EP
Relative retention	-	-	$\alpha = k'_{(a)} / k'_{(b)}$ T_R of peak a < T_R of peak b	Selectivity	Peak_Selectivity
-	Resolution	Resolution	$Rs = 1.18\times(t_{R2}-t_{R1})/(W_{0.5h1}+W_{0.5h2})$	Resolution (EP) Resolution (JP)	Peak_Resolution_EP Peak_Resolution_JP
-	-	-	$R = \frac{(2.35/2)(T_{R(b)} - T_{R(a)})}{W_{50(b)} + W_{50(a)}}$	Resolution	Peak_Resolution_Classic
Resolution	-	-	$Rs = 2.0\times(t_{R2}-t_{R1})/(W_1-W_2)$	-	Peak_Resolution_USP
Efficiency			$N = 16\times(t/W)^2$	Plates Tangent method	Peak_TheoreticalPlates_USP
-	Efficiency	Efficiency	$N = 5.54\times t_R^2/W_{0.5h}^2$	Plates halfheight method	Peak_TheoreticalPlates_EP
Relative retention time	Relative retention time		$Rr = t_2/t_1$	-	Peak_RelativeRetTime

Table 11 Pharmacopoeia values in ChemStation reporting

USP	EP	JP	Definition	classic reporting (RLE)	intelligent reporting (RTE)
	S/N ratio	S/N ratio	S/N = 2H/h	-	Peak_SignalToNoise_EP
Peak-to-valley ratio	Peak-to-valley ratio		$p/v = H_p/H_v$	PeakValleyRatio	Peak_PeakValleyRatio

10
System Verification

This chapter describes the verification function and the GLP verification features of the ChemStation.

Verification and Diagnosis Views

If supported by the configured instrument, for example, the Agilent 1100/1200 Series modules for LC, the ChemStation comprises two additional views to perform instrument verification and diagnosis tasks. For more information, see the online help system.

System Verification

System verification is a key component in the routine use of an analytical instrument in a regulated laboratory. The GLP verification features of the ChemStation are designed to help you to prove that the software, or a relevant components of the software, are performing correctly, or were performing correctly at the time of a particular analysis.

The ChemStation verification function enables you to verify the correct operation of your ChemStation software. You can do this by reprocessing data files according to specific methods, and comparing the results with a pre-defined standard. The verification function is particularly important to prove the integrity of the integration and quantification results.

You can use the standard verification test, or define your own tests using your own method and data files to check the algorithmic software combinations used by your analysis methods. The verification test is a protected file and cannot be changed or deleted.

The Verification item in the Data Analysis view allows you to choose any of the following options:

- run a verification test in the database,
- define a new verification test and add it to the database, and
- delete a verification test from the database.

The How To section of the online help system describes how to perform these tasks. When running a ChemStation verification test, you can choose whether to run the entire test, or select a combination of parts.

Verification test results are saved in binary format to the default subdirectory: c:\CHEM32\1\Verify, together with the method and data files. The Verify subdirectory is at the same level as the sequence, methods and data subdirectories. You can send the results to a printer or to a file. The test results, including a combined verification test result, are rated as either pass or fail.

The following verification test components are available:

Digital Electronics (Agilent 1100/1200 Series DAD only)

A test chromatogram is stored in the diode-array detector. This chromatogram is sent to the ChemStation after it has gone through the same preprocessing steps as normal raw data from the photodiodes. The resulting data are compared to original result data stored in the ChemStation for this test chromatogram. If there is a mismatch the test fails. This test ensures that the DAD electronics which do the data preprocessing are still functioning correctly. As a stored test chromatogram is used, the lamp or the diode array are not part of this test. They can be checked with the "DAD Test Function" on page 163.

Peak Integration

The data file is integrated again using the original method. The results are compared to the original integration results stored in verification register. If they do not match, the test fails.

Compound Quantification

The compounds in the data file are quantified again. The results are compared to the original quantification results stored in the verification register. If they do not match, the test fails.

Report Printing

The original report is printed again.

The following page shows an example of a successfully completed verification test.

```
================================================================
        ChemStation Verification Test Report
================================================================
```

Tested Configuration:

```
   Component                        Revision
----------------------------------------------------------------------
  ChemStation for LC 3D ChemStation   B.01.01
  Microsoft Windows                       Microsoft Windows XP
  Processor                        Processor_Architecture_Intel
  CoProcessor                      yes

ChemStation Verification Test Details:
  Test Name : C:\CHEM32\1\VERIFY\DEFAULT.VAL
  Data File : C:\CHEM32\1\VERIFY\DEFAULT.VAL\VERIFY.D
  Method    : C:\CHEM32\1\VERIFY\DEFAULT.VAL\VERIFY.M
  Original Datafile             : VERIFY.D
  Original Acquisition Method   : VERIFY.M
  Original Operator             : Hewlett-Packard
  Original Injection Date       : 4/16/93 11:56:07 AM
  Original Sample Name          : Isocratic Std.

Signals Tested:
  Signal  1: DAD1 A, Sig=254,4 Ref=450,80 of VERIFY.D

ChemStation Verification Test Results:

  Test Module            Selected  For Test     Test Result
  ------------------------------------------------------------------------
  Digital electronics test   No                    N/A
  Integration test          yes                   Pass
  Quantification test       yes                   Pass
  Print Analytical Report    No                   N/A

ChemStation Verification Test Overall Results:     Pass
```

The GLPsave Register

The GLPsave register is saved at the end of each analysis when selected in the run time checklist. It contains the following information:

- signals,
- logbook,
- integration results table,
- quantification results table,
- instrument performance data, and
- data analysis method.

This register is a complete protected record, generated at the time of analysis. You can recall it at any time in the future as proof of your analytical methods.

The GLPsave Register option in the Data Analysis view enables you to review the GLPsave register file at any time. The file is protected by a checksum and is encoded into binary to ensure it is not changed.

In the dialog box used to select the GLPsave register for review, you can choose your review options from the following:

- load original method,
- load original signals,
- load instrument performance data,
- print original method,
- print original integration results,
- print original quantification results, and
- generate original report from the original method and signals.

You can use the GLP review function to show that chromatographic data are original, prove the quality of the analysis from the instrument performance data, and demonstrate the authenticity of the data interpretation.

For example, you can:

- reload and reprint the data analysis part of the method used at the time of the sample analysis to prove that the data evaluation, presented as the results of the analysis has not been modified in any way, and
- review without recalculating, the integration and quantification results to prove the authenticity of the report.

DAD Test Function

Detector tests can be used as a step in the routine system validation of an analytical instrument in a regulated laboratory.

The DAD test assesses the performance of your diode array detector. When you select the DAD test from the Instrument menu (for LC3D and CE only) it checks the instrument for intensity and wavelength calibration. When you press Save the test results are automatically saved to the DADTest database, a register file called DADTest.Reg located in the default instrument directory.

Review DAD Test Function

The **Review DAD Test** function in the data analysis View menu enables you to review the DADTest.Reg file at any time. The file is protected by a checksum, and is encoded into binary to ensure that it is not changed.

You can select any of the following parts of the DAD test for review:

Show Holmium Spectra Plots all Holmium spectra listed in the DAD Test review table. The active spectrum is tagged.

Show Intensity Spectra Plots all intensity spectra listed in the DAD Test review table. The active spectrum is tagged.

Save as New Database If you change the lamp in your DAD you can reset the DADTest by deleting any unwanted test results from the table and then using this function to save as new database.

Show Selected Spectra Displays only spectra you selected in the table.

Show Intensity Graph You can plot an intensity graph to give an indication of the life of the lamp in your diode array detector. The graph provides a function of maximum lamp intensity against time.

Index

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 ChemStation Edition.

The information contained herein may be used by validation professionals for planning and execution of system validation tasks.

Printed in Germany
1/2013

M8301-90024

Agilent Technologies