Peak purity analysis in HPLC and CE using diode-array technology

Application

Abstract

In terms of quality control and for all quantitative analysis peak purity is a major task, which can be addressed in different ways. Often only peak shapes and chromatograms are taken into consideration but a very elegant possibility without the need to use a mass spectrometric detector is the comparison of spectra recorded with a diode-array detector during the registration of a chromatographic peak. Due to the differences of these spectra within a single chromatographic peak, its purity can be decided easily. The Agilent ChemStation software offers an easy to use tool to perform these tasks routinely. In this note, the theoretical background of peak purity determination is presented as well as the practical use of the ChemStation software for these purposes.

Agilent Technologies
**Introduction**

An essential requisite of a separation analysis is the ability to verify the purity of the separated species, that is, to ensure that no coeluting or comigrating impurity contributes to the peak’s response. The confirmation of peak purity should be performed before quantitative information from a chromatographic or electrophoretic peak is used for further calculations. Neglecting peak purity confirmation means, in quality control, an impurity hidden under a peak could falsify the results or, in research analysis, important information might be lost or scientific observations rendered void should an impurity remain undiscovered. Validated analytical methods usually include the peak purity check as a major item in the list of their method validation criteria (table 1).

### Method validation criteria

| Selectivity (peak purity determination) | ✔ |
| Linearity | ✔ |
| Limits of detection and quantification | ✔ |
| Quality of data (accuracy and precision) | ✔ |
| Ruggedness | ✔ |

**Table 1**

Peak purity determination – a major criterion in method validation

### Techniques for peak purity determination

Several techniques are currently used for peak purity determination in high performance liquid chromatography (HPLC) and in capillary electrophoresis (CE). With a conventional single wavelength detector (or a monitor providing just one single output signal) such as a refractive index- or conductivity detector, peak purity can only be judged from the peak profile of this signal. Peak profile however is influenced by a variety of parameters and depends heavily on chromatographic or electrophoretic resolution. Therefore peak purity determination based on the peak profile of a single signal is a very unreliable and insensitive method. This is especially true in CE, where mismatched sample and buffer zones always result in peak skewing. As a consequence, a second approach involving multiple wavelength detectors acquiring more than one signal in parallel has been adopted. Such detectors enable impurities to be uncovered by methods that involve overlaying signals to compare peak profiles and calculate signal ratios. These detectors have some disadvantages which will be discussed in the following section. They can be eliminated easily by a third approach based on diode-array technology:

- on-line acquisition of UV-Visible spectra during the peak’s elution,
- several signals at different wavelengths in parallel,
- signal extracts from a 3-D data matrix, containing spectral data and separation signals, for peak purity analysis.

Peak purity determination can be performed in different levels, tailored to the complexity of the separation and the user’s needs. Detailed descriptions of the various peak purity routines for signals and spectra, such as normalization and overlay, the mathematical operations and the different display modes will be given in the following sections.

### In the elution dimension

#### Signals

A valuable tool in peak purity analysis is the overlay of separation signals at different wavelengths to discover dissimilarities of peak profiles. The availability of spectral data in the three-dimensional matrix generated by the diode-array detector enables signals at any desired wavelength to be selected and reconstructed for peak purity determination after the analysis. A set of signals can be interpreted by the observer best when, before being displayed, it is normalized to maximum absorbance or to equal areas. A good overlap, where peak shape and retention or migration time match, indicate a pure peak while a poor overlap indicates an impure peak, as demonstrated in figure 1.

In addition to overlaying signals, their ratios can be calculated and plotted. The resulting ratiograms are sensitive indicators of peak purity (figure 2). Any significant distortion of the ideal rectangular form of the ratiogram indicates an impurity. However, there are some limitations to be considered when using signals for peak purity determination:

- The UV-Visible spectra of both the main compound and the impurity must be well known in order to select the most suitable wavelengths for the peak profile comparison. This fact reduces the applicability of this type of impurity detection to a routine like search of known impurities in known main peaks.
The technique is not directly applicable to research work or method development. The risk of overlooking an unknown impurity remains even when several wavelengths are selected in parallel.

**Peak purity determination in the third dimension**

*Spectra*

Comparing peak spectra is probably the most popular method to discover an impurity. If a peak is pure all UV-visible spectra acquired during the peak's elution or migration should be identical, allowing for amplitude differences due to concentration. The results obtained by comparison of these spectra against each other should be very close to a perfect 100% match. Significant deviations can be considered as an indication of impurity. Unfortunately the inverse is not necessarily true. If the spectra are not significantly different, the peak can still be impure for one or more of three possible reasons:

1. The impurity is present in much lower concentrations than that of the main compound.
2. The spectrum of the impurity and the spectrum of the main compound are identical or very similar.
3. The impurity completely coelutes or comigrates with the main compound, with both having exactly the same peak profile.

Any peak purity algorithm can only confirm the presence of impurities and never prove absolutely that the peak is pure. The likelihood of discovering an impurity rises with increasing distinction between spectra and peak profile, higher resolution between the main compound and the impurity, and with increasing absolute and relative concentration of the impurity.

There are no hard and fast rules as to which concentrations of impurities can be detected and which not. In general, less than 0.5 - 1% can be determined if the spectra are distinct enough. If the spectra resemble each other very

---

**Figure 1**
**Normalized signals for pure and impure peaks**

**Figure 2**
**Signal ratiograms for impure and pure peaks**
closely, and the column or capillary system does not resolve the impurity and main component well, then only 5% impurity may be feasible.

Before proceeding, the difference between the terms spectral impurity and peak impurity should be clearly defined. Spectral impurity indicates a distortion of the analyte spectrum by the near-constant presence of background absorbance from solvents, and/or matrix compounds and/or an impurity. Peak impurity, in contrast, refers to a distortion of the analyte spectrum by an additional component which partially or completely coelutes or comigrates with the major compound.

**Background correction of the peak spectra**

Before the spectra are used for the peak purity analysis, they should be corrected for background absorption caused by the mobile phase or matrix compounds, by subtracting the appropriate reference spectra. Whether such a background correction needs to be applied depends on the separation system employed. For isocratic separations with a properly balanced diode-array detector, the solvent’s constant spectral contribution will be eliminated by the automatic subtraction of the solvent spectrum at the beginning of the run. For gradient separations where the mobile phase’s contribution to absorbance may change with time, background corrections should be made for each peak individually. The Agilent ChemStation software offers three different modes for setting reference spectra (figure 3):

1. **No reference**: Spectral operations are performed without any reference (recommended for exceptional situations only, even isocratic separations should use a reference spectrum).

2. **Manual reference**: One or two reference spectra can be specified by the user. This mode is used for interactive spectral evaluations of non-baseline separated peaks. Only the spectra in between the two selected reference points are used for the purity evaluation.

3. **Automatic background correction**: Depending on the acquisition parameters chosen either A, B or C is performed automatically. A) **All spectra**: For peak purity, the spectra at the actual peak start and end
are linearly interpolated and subtracted from the peak spectrum. The first integrated peak start or end to the left of the selected spectrum that is on the baseline is taken as the time for left reference spectrum. For the right reference time the first integrated peak start or end on the baseline to the right is taken (figure 4 and 5). If a peak is not completely resolved from its neighbour, automatic selection of reference spectra might lead to a reference selected from the valley between the two peaks. Though we already know that an unresolved peak cannot be pure, we may want to use the purity test to look for other hidden components. In that case, the manual reference can be used to select a reference spectrum from before and after the group of peaks.

B) **Peak controlled spectra:**
The nearest spectrum with type Baseline to the left and right of the selected spectrum is taken as the reference time. When no left or right baseline spectrum is found, the first or last spectrum from the data file is taken as the reference (figure 5).

C) **FLD spectra:**
The nearest spectra at the points of inflection on the up-and down-slopes of the peak are taken as the reference spectra. This optimizes signal-to-noise and error correction.

**Wavelength range**
The wavelength range for the spectra can, and should be, selected carefully so that only the significant spectral area is under observation (figure 6). This eliminates the high absorbance of eluants in the lower UV range that normally cause high spectral noise. Higher wavelength ranges should be omitted if the compounds show no absorbance to avoid increased noise and calculation time.

**Normalization of the peak spectra**
Before the spectra are compared they should be normalized. Three different modes of normalization (figure 7) are possible:
1. Maximum absorbance of the spectrum (or a particular wavelength range of the spectrum)
2. Area of the spectrum (or a wavelength range of the spectrum)
3. Best possible match of the entire spectrum. This type of normalization is recommended to display spectra for peak purity evaluation because it tries to make the difference of both spectra as small as possible by shifting and rescaling the spectra. The ChemStation normalizes spectra automatically using the best possible match of the entire spectrum.

**Selection of the peak spectra for comparison**
Traditionally, three peak spectra sufficed for purity determination: the upslope, the apex and the downslope spectra. This practice risks overlooking impurities at the base of the peak. Agilent’s diode-array detectors can acquire all peak spectra or peak-controlled spectra with the help of the Agilent ChemStation. Peak-controlled acquisition is subject to a certain noise absorbance threshold, as described in the following section and shown in figure 6 and 8, to detect peaks. Acquisition either way provides significantly more spectral data for more reliable peak purity evaluations. The number of spectra per peak to be processed can be three or more. If the value is set to three then three spectra are taken at roughly equidistant points during the peak’s elution or migration. If spectra have been acquired in peak-controlled mode during the run, you should select All spectra for purity determination because
in most practical cases this corresponds to the three to five spectra that will have been recorded. If all spectra were acquired during the run, a setting of five to seven spectra to be processed is wise (figure 9). Too many spectra will only increase calculation and display time, without providing any more significant information.

Absorbance threshold
Setting an absorbance threshold puts limit to the number of spectra considered for peak purity. This reduces the contribution of spectral noise to normalization and overlay (figures 6 and 8).

Spectra processing
Before comparing spectra several mathematical operations can be carried out to improve their quality (figure 6):

1. Smooth factor
This mathematical operation smoothes the selected spectrum using the coefficients of Savitzky-Golay: The filter lengths can be from five data points upwards with no upper limit. The default value is 5 and in order not to lose too many spectral characteristics the filter length should not exceed 13 (in most cases). Smoothing is useful to remove spectral noise, which makes identification more reliable when the spectral noise is about one fifth or more of the absorbance.

2. Spline factor
This constructs a smooth curve through each data point by generating new data points. The number of new data points generated is calculated as follows:

\[
\text{No. of data points} = \text{1} \times \text{spline factor} + \text{1}
\]

The higher the spline factor the more data points are generated between the existing ones. The splined curve still goes through all of the original data points and merely makes the curve visually more attractive.

3. Logarithm
This calculates the logarithm of the selected spectrum. Logarithmic spectra reduce the absorbance scale. It may be useful to use logarithmic spectra in cases where the absorbance scale is very large.

4. Derivative Order
This calculates the specified derivative of the selected spectrum. Derivative spectra reveal more pronounced details than original spectra when comparing different compounds. The derivative of a spectrum is very sensitive to background noise.

Peaks purity determination
1. Comparing the peak spectra
After selection, correction for background influences and normalizing, the spectra can be overlaid to check for possible spectral impurities. Figure 10 shows the normalized and overlaid spectra of a pure peak and figure 11 an impure peak. Any significant dissimilarity encountered in the com-
parison of the spectra recorded across the peak indicates the presence of an impurity. However, no conclusion can be drawn concerning the kind, number and level of impurity. An additional aid in interpreting spectral dissimilarities is the display of difference spectra, generated by subtracting these normalized spectra from the other peak spectra selected. The profiles of the difference enable a further conclusion to be drawn: randomly distributed residual profiles result from spectral noise which may be caused by the instrument (figure 10, upper section), whereas systematic trends will be observed if real spectral differences caused by a spectral impurity occur (figure 11, upper section).

2. The similarity factor
The ChemStation’s special peak purity software routine does not only allow the display of spectra and their differences, it is also able to calculate a numerical value to characterize the degree of dissimilarity of the peak spectra, a so-called similarity factor, based on the match of the peak spectra to one another.

Several statistical techniques are available for comparison of spectra. Since UV-Visible spectra contain only a small amount of fine structure, the least – square - fit coefficient of all the absorbances at the same wavelength gives the best result. The similarity factor used in the ChemStation is defined as:

\[
\text{SIMILARITY} = 1000 \times r^2
\]

where \( r \) is defined as

\[
r = \frac{\sum_{i=1}^{n} (A_i - A_{av})(B_i - B_{av})}{\sqrt{\sum_{i=1}^{n} (A_i - A_{av})^2 \cdot \sum_{i=1}^{n} (B_i - B_{av})^2}}
\]

and where \( A_i \) and \( B_i \) are measured absorbances in the first and second spectrum respectively at the same wavelength; \( n \) is the number of data points and \( A_{av} \) and \( B_{av} \) the average absorbance of the first and second spectrum respectively (see also figure 12).
At the extremes, a similarity factor of 0 indicates no match and 1000 indicates identical spectra. Generally, values very close to the ideal similarity factor (greater than 995) indicate that the spectra are very similar, values lower than 990 but higher than 900 indicate some similarity and underlying data should be observed more carefully. Figure 13 shows examples of similarity factors for similar, different, noisy and spectra with impurity. The slope of the regression lines represents the ratio of the concentration of the two spectra.

Figure 12
Similarity of absorbance at the individual wavelength plotted for a pair of spectra gives the similarity factor

Figure 13
Graphical display of similarity factor for different pairs of normalized spectra
3. Improving sensitivity and reliability
The similarity curve and threshold curve functions improve the sensitivity and reliability of the peak purity evaluation by using all the spectra acquired during the elution or migration of a peak rather than just three or four.

- **Similarity curve**
The mathematical fundamentals used in the similarity curve calculations are those used for the purity factor, however, they are displayed in another format. All spectra from a peak are compared with one or more spectra selected by the operator (figure 14), an apex or an average spectrum for example. The degree of match or spectral similarity is plotted over time during elution. An ideal profile of a pure peak is a flat line at 1000 as demonstrated in figure 15(a). At the beginning and end of each peak, where the signal-to-noise ratio decreases, the contribution of spectral background noise to the peak's spectra becomes important. The contribution of noise to the similarity curve as shown in figure 15(b).

- **Threshold curve**
The influence of noise on weakly absorbing spectra can be seen in figure 16. The similarity factor decreases with decreasing signal-to-noise ratio or constant noise level with decreasing absorbance range. A threshold curve shows the effect of noise on a given similarity curve. The effect increases rapidly towards both ends of the peak. In essence, a threshold curve is a
similarity curve with background noise contribution. Figure 17(a) shows both the similarity curve and the threshold curve for a pure peak with noise, figure 17(b) for an impure peak. The determination of noise threshold is performed automatically based on the standard deviation of pure noise spectra at a specified time with a user selectable number of spectra, usually at 0 minutes with 14 spectra. In subsequent analyses you can define the noise threshold as a fixed value based on your experience. The threshold curve, represented by the broken line, gives the range for which spectral impurity lies within the noise limit. Above this threshold, spectral background noise and the similarity curve intersects the threshold curve indicating an impurity providing the reference and noise parameters have been sensible chosen (figures 17 and 21).

The software offers four modes to display the similarity and threshold curves:

- **Similarity & Threshold (without any transformation):** The similarity and threshold curves are displayed as they are calculated using values between 1 and 1000 (figure 18a).

- **Similarity & Threshold as the natural logarithm:** Similarity and threshold curves are displayed as the natural logarithm of their calculated value. This can give additional details in the higher match factors (figure 18b).

![Figure 16](image1)

**Figure 16**
Similarity factor as function of the noise level

![Figure 17](image2)

**Figure 17**
Effect of impurity and noise on similarity and threshold curves
• **Similarity / Threshold ratio:**
The ratio of the similarity and threshold curves is displayed as a single curve.
\[
\text{ratio} = \frac{1000 - \text{similarity}}{1000 - \text{threshold}}
\]
The results for each spectrum are compared to the expected result based on the threshold calculation. If the ratio is less than 1 the test for that spectrum passes, if it is greater than 1 then it fails (figure 18(c)).

• **Purity ratio:** The purity value of each single spectrum is displayed as the logarithm of the difference from the threshold value. The chromatographic peak, similarity and threshold curves are displayed in the upper part of the display (figure 18(d) and 21). For a spectral pure peak the ratio values are below unity (the purity ratio is in the green band) and for spectral impure peaks the values are above unity (the purity ratio is in the red band). The advantage of these modes are that only one line is displayed, instead of two which makes the interpretation more simple. All the exact values for each single spectrum are not only graphically displayed but can also be reviewed in the Peak Purity information window (figure 19) of the ChemStation.

4. **Using specific target spectra**
The ChemStation permits calculations of the purity factor and similarity curves relative to different target spectra, as shown in figure 20 and table 2. As a general rule, the option Compare the average spectrum provides the most valuable information for unknown impurities. The flexibility of being able to select a specific target spectra is valuable in those cases where the analyst has to assume where the impurity is, or needs to improve the sensitivity of purity evaluation. An example may help to show how this principle can be applied. If the impurity is assumed to be located on the peak, selecting the tail or apex spectrum to be compared with all other spectra will provide the most significant information. Figure 20 gives the ratio curve for the front, apex, tail and average spectrum of a peak which contains an impurity after the response maximum (apex).
The front spectrum gives a small spectral impurity at the end of the peak. The deviation in this first ratio curve is small since the front spectrum absorbed little (giving a rather high threshold curve). The apex spectrum gives a low impurity in the front of the peak (the apex spectrum contains only a very small amount of the impurity) and high impurity at the tail. The tail spectrum (with a high amount of impurity) gives a spectral impurity at the front of the peak. The average spectrum (a mean of the peak spectra selected, in this case the upslope-, apex-, and downslope spectra) indicates spectral impurity in the total peak. This average spectrum contains of course, the spectral contribution of the impurity. In this particular case the average contains more contribution from the impurity than the apex spectrum, showing a higher spectral impurity at the elution or migration front, and lower impurity at the tail, compared with the ratio curve of the apex spectrum.

The profile of the similarity-, threshold- and ratio curves depends on the position, level and spectral differences of the impurity and, as such, no general statements can be made on shape. Profile will differ from situation to situation.

5. Calculating a purity factor
Two approaches are available, depending on whether you select to use similarity curves or not.

- With fixed threshold
  The target spectrum, as specified by the user, is compared with all the other selected peak spectra. (3, 5, 7, 9 or all). The mean of all purity values below the specified threshold gives the purity factor. If no value falls below the threshold, the purity factor is calculated as the mean of all values. A fixed threshold value may be useful in some cases of quality control or if the automatically calculated threshold is too restrictive for your purposes.

- With calculated threshold curve
  The mean of all purity values exceeding the calculated threshold gives the purity factor, on the condition that at least three consecutive values lie over the threshold. If you choose to use the threshold curve, the threshold is calculated as the mean of the same data points used to calculate the purity factor. The spectra used to construct the purity curve are indicated by plus and minus marks in the graphical display (as annotated in figure 18(d) and 20). The blue marks denote spectra under the threshold limit whereas the red ones denote those exceeding the threshold limit. Minus denotes spectra excluded from the similarity factor calculation since insufficient neighboring spectra lay above the threshold. Plus denotes spectra included in the similarity factor calculation since at least two more, neighboring, spectra also exceed the threshold.

6. Extracting signals automatically from spectra data
The ChemStation software allows to automatically select peak signals for display. These peak signals are taken from the user specified signals acquired during the analysis or are extracted from the
spectral data recorded with all spectra set during acquisition. The extraction is optimized to find as much difference as possible in the curvature of the signals. This gives an extra conformation for peak impurity and in many cases an indication of the location of the impurity (in some cases even when more than one impurity is present).

7. ChemStation software windows for peak purity analysis
The peak purity analysis window contains different windows as shown in figure 21. The whole separation is displayed at the top of the window, and a boxed outline designates the peak currently under scrutiny. Retention or migration time labelling is optional. On the left side below an overlay of the different peak spectra is shown. This clearly shows their degree of similarity and therefore also the purity of the chromatographic peak investigated. On the right side below the chromatogram the similarity and threshold curves and the purity ratio is shown. This allows an easy, quick and sensitive determination of the peak purity. Also, the differences of the spectra, similarity and threshold curves, their logarithm and the similarity to threshold ratio can be displayed in more demanding cases.

Examples

**Pure peak**
Figure 22 shows the selected peak spectra. The signals overlap perfectly reaffirming the validity of the background correction. The similarity curve exhibits a profile very similar to and within the
threshold curve limits and, the peak purity ratio is clear within the green band.

**Impure peak**

Figure 23 shows the peak purity determination of an impure peak containing an impurity with a quite similar spectrum to that of the main compound. The overlay of the normalized spectra and the extracted signals indicate the presence of an impurity. The similarity curve exceeds the threshold curve between 9.5 and 9.6 min and the peak purity ratio is clearly in the red band thus leading to the warning message. From these signals we can conclude that the impurity is at the end of the peak. Figures 24 illustrates the ChemStation’s capabilities to discover very small impurities of 0.5 % and less with almost indentical spectra to the main analyte. Often the spectra overlay, the residual spectra and even the ratiogram do not provide sensitive enough information to uncover the impurity, however the similarity and threshold curve and the peak purity ratio are more sensitive and are able to reveal the hidden impurity.

**Automation of peak purity determination**

All the routines described here can be performed both interactively and fully automatically. Chosing “full report” as report style creates a purity report for any integrated chromatographic peak. All parameter preferences can be stored in a single method, and printed as documentation of the analysis to aid in the execution of Good Laboratory Practices (GLP). Starting such a method initiates injection, separation, data analysis and reporting of the samples in one
Printed reports contain all the graphical information from the screen plus additional numerical results. Despite the ChemStation’s capacity to run sample analyses automatically and unattended, we recommend that you examine the results yourself first visually before passing any purity judgement. Incorrect results most often arise when a column or capillary system fails to resolve peak or the ChemStation integrates some baseline event erroneously leading to a false reference spectrum being selected.

References


Conclusion

Peak purity determination with a diode-array detector is a powerful tool to check peak purities. By comparing spectra from the upslope, apex and downslope impurities with less than 0.5 % can be identified. Therefore the technique described offers an alternative to using a mass spectrometric detector for peak purity. The Agilent ChemStation software offers an easy and automated tool to perform the peak purity check with highest performance. This can and should be done as a matter of routine to achieve reliable high quality data.

Mark Stahl is Application Chemist at Agilent Technologies, Waldbronn, Germany

www.agilent.com/chem