Evaluation of electropherograms/chromatograpms

Open educational resource for Instrumental Analytical Chemistry laboratory practice

During this practice students are given a chance to learn the basic principles of evaluating electropherograms/chromatograpms. Although the practice focuses on the evaluation of electropherograms - data obtained by capillary electrophoresis (CE) – the presented definitions, techniques, mathematical procedures are generally applicable to the assessment of chromatograms, as well.

The educational material is based on the operation of the following evaluation software: Chemstation 3D CE (Hewlett-Packard).

Analytical chemistry is a field of study dealing with the separation, identification and quantification of analytes present in a sample. The objective of each analytical procedure is to find answers to two questions:

1. What components is the sample comprised of? (qualitative analysis)
2. What is the concentration of each component in the sample? (quantitative analysis)

In CE, *qualitative analysis* is based on the identification of the peaks in the electropherogram. This can be accomplished by comparing the migration time (or mobility) of given peaks with those of known compounds obtained experimentally. Peak identification based on the comparison of migration times demands constant experimental conditions.

Quantitative analysis provides information about the amount or concentration of a component in a sample by determining peak height or peak area. Peak height can be seen on the electropherogram; however, an integrator (software) is necessary for the determination of peak area.

1. Integration of peaks

What is Integration?

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

Integration is a necessary step for:

- Quantification.
- Peak purity calculations.
- Spectral library search.

What Does Integration Do?

When a electropherogram is integrated, the software:

- Identifies a start and an end time for each peak, and marks these points with vertical tick marks.
- Finds the apex of each peak, that is, the migration time.
- Constructs a baseline.
- Calculates the area for each peak.

How Integration Works

The integration process comprises the following:

- Peak recognition (Cardinal Point definition).
- Baseline construction.
- Peak area calculation.

Peak Recognition

The first part of integration involves peak recognition, which comprises the following processes:

- Finding the start of the peak.
- Defining the apex of the peak.
- Finding the end of the peak.

The process for finding a positive peak is:

1. Slope and curvature within limit : track baseline.
2. Slope and curvature above limit : possibility of a peak.
3. Slope remains above limit : peak recognized.
4. Curvature becomes negative : front inflection point.
5. Slope becomes negative : apex of the peak.
6. Curvature becomes positive : rear inflection point.
7. Slope and curvature within limit : approaching end of the peak.
8. Slope and curvature remain within limit : end of peak, track baseline.

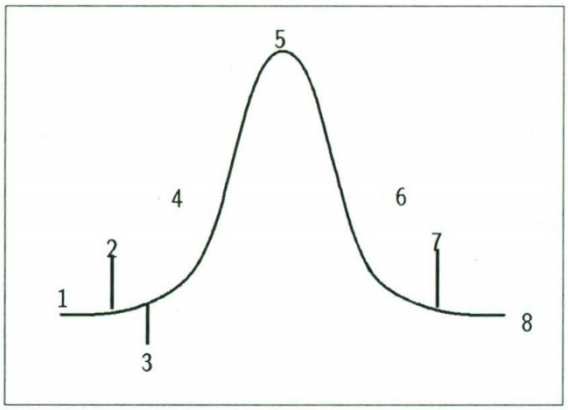

Figure 1: Cardinal points

Determining Peak Apex

The slope changes from positive to negative at the top of the peak. To calculate the values for the migration time and peak height, the integrator takes the highest data point and one on either side, fits them to a quadratic equation, and solves the equation to find the highest point.

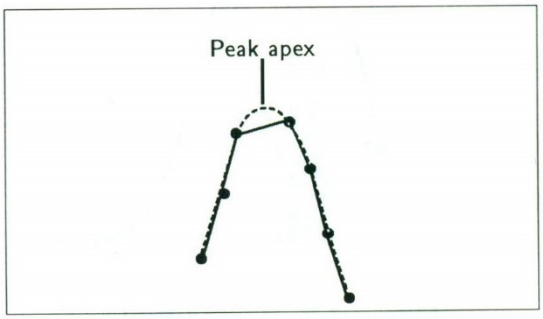

Figure 2: Determining peak apex

Integration of Merged Peaks

Sometimes two peaks are merged with no baseline between them.

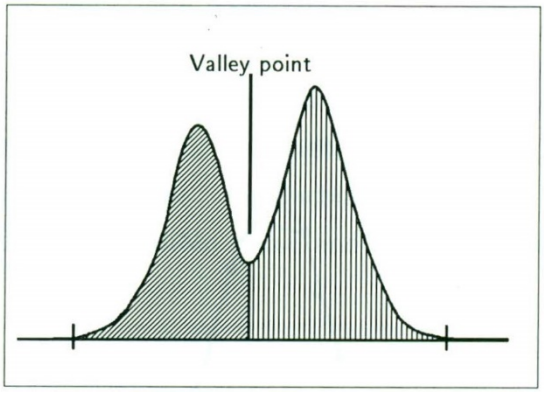

Figure 3: Merged peaks

In the case of a baseline valley (BV) and a valley baseline (VB) peak, the integrator separates the merged peaks by dropping a perpendicular from the valley point between the two peaks. The integrator locates the start of the first peak and begins to accumulate data until it locates a valley. At the valley point, accumulation of data samples from the first peak stops. Accumulation of the data samples of the second peak starts. When the integrator finds the end of the second peak the accumulation of data stops.

Integration of Shoulders

A shoulder is an unresolved peak on the leading or trailing edge of a larger peak. There is no true valley in the sense of negative slope followed by positive slope.

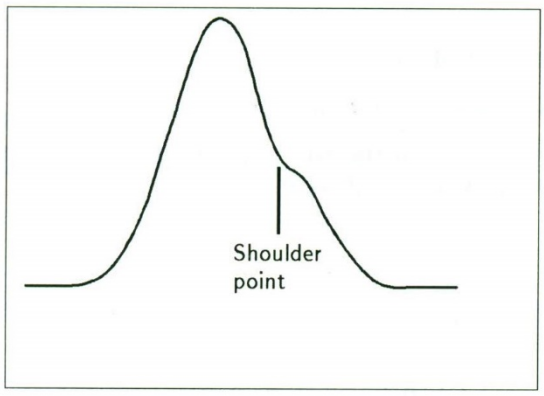

Figure 4: Peak shoulders

The shoulders found are reported with their time of maximum negative curvature. The area reported for the main peak includes the area of the shoulder(s).

Baseline Construction

After the peak has been recognized, baseline construction is done for determining the final area of the peak. The baseline construction follows the changes in the electrophoretic signal.

The integrator constructs the baseline as series of straight line segments between:

- the signal level at start of the run;
- the tick-marks (marking the start and end of a peak);
- the signal level at the end of the run; a point at the stoptime on a horizontal extension from the last declared baseline point.

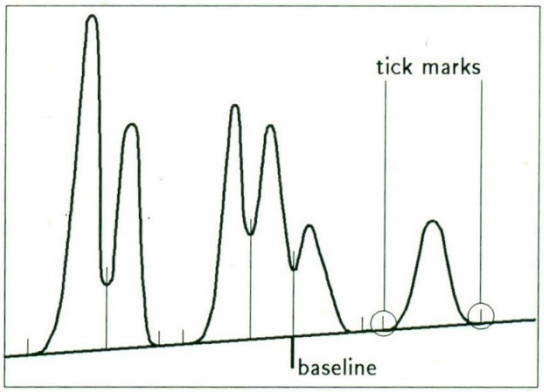

Figure 5: Default baseline construction

Peak Separation Codes

Each peak is assigned a two letter code that describes how the electrophoretic baseline was drawn. The first letter describes the baseline at the start of the peak and the second letter describes the baseline at the end of the peak.

/ The peak integration was aborted.

B The peak started or stopped on the electrophoretic baseline.

H The peak started or stopped on the horizontal baseline.

N This is a negative peak.

P The peak started or stopped while the baseline was penetrated.

S The integrator recognized the peak as big peak, later on referred to as solvent peak .

T The peak started or stopped while tangent skim was enabled.

V The peak started or stopped with a valley dropline.

\+ The peak was included as part of a cluster of summed peaks.

M The peak was manually integrated.

F The peak was forced by manual integration. If a peak occurs before the manually integrated peak and the end changes because of manual integration the peak is classified as forced.

R A solvent peak has been affected by manual integration, such as tangent skim is classified as a re-calculated solvent peak.

Modified Baseline Construction

Under certain conditions, the integrator will modify the baseline to include points determined by certain rules, as given below:

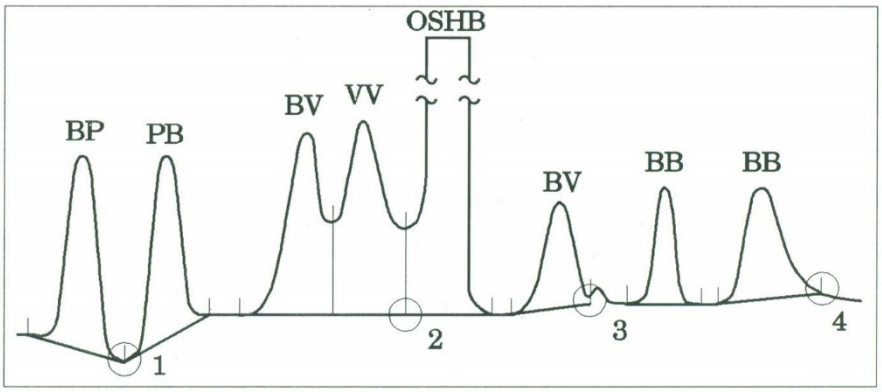

Figure 6: Modified baseline construction

- If a peak penetrates the baseline (BP and PB), the baseline also goes through the lowest point on the peak, see 1, in figure 6.
- If a solvent peak detected using tangent skimming does not start on the baseline, the baseline also goes through a point, see 2, on a horizontal extension from the last-declared baseline point to the start of the solvent peak. Overflow solvent horizontal baseline (OSHB) means the signal has exceeded the linear range of the detector.
- If a peak ends in an apparent valley but the following peak is below area reject, as you have set it, the baseline is projected from the beginning of the peak to the next true baseline point, see 3, in figure 6. If a peak starts in a similar way, the same rule applies.
- To improve handling of tailing peaks, the time the peak exceeds upslope and downslope criteria is noted and one quarter of that time is added at the end of this peak before the end of the peak is confirmed, see 4, in figure 6.

Baseline Penetrations

A penetration occurs when the signal drops below the constructed electrophoretic baseline. If a baseline penetration occurs, that part of the baseline is redrawn.

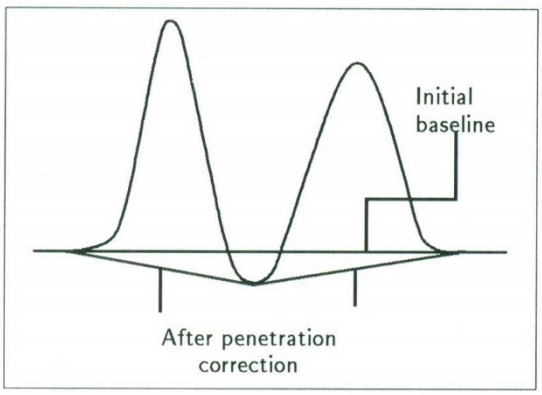

Figure 7: Baseline penetration

Peak Area Measurement

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

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

- for baseline-to-baseline (BB) peaks, the area is above the baseline between the tick marks;

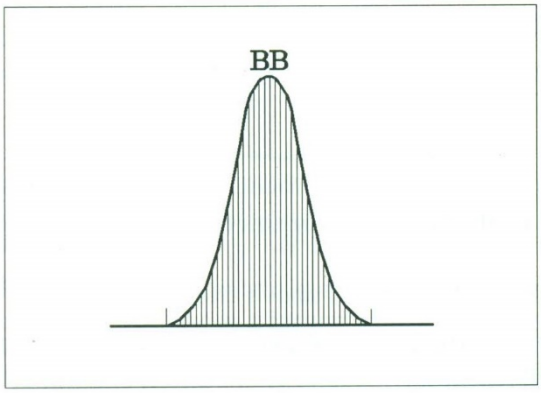

Figure 8: Area measurement for Baseline-to-Baseline peaks

- for valley-to-valley (VV) peaks, the area above the baseline is divided with vertical dropped lines from tick marks;

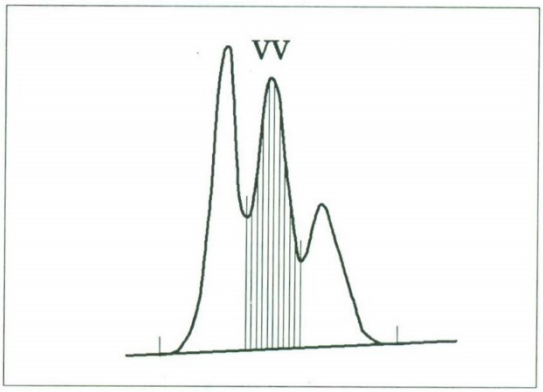

Figure 9: Area measurement for Valley-to-Valley peaks

- for tangent (T) peaks, the area is above the reset baseline;
- for solvent (S) peaks, the area is above the horizontal extension from the last-found baseline point and below the reset baseline given to tangent (T) peaks.

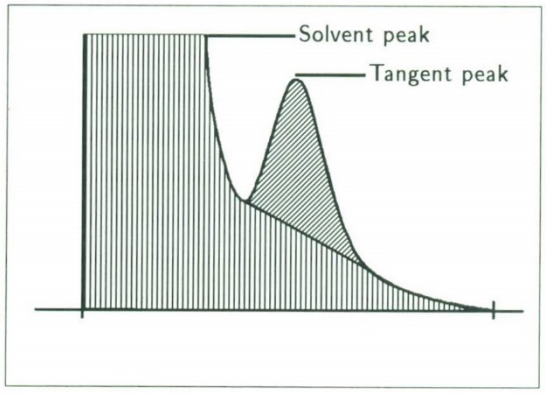

Figure 10: Area measurement for Tangent and Solvent peaks

Integration Events

There are four initial integrator events and nineteen timed integrator events. Many events are on and off or start and stop pairs.

Initial Events

Initial Peak Width	Sets the initial width of the peak at half height.
Initial Threshold	Sets a minimum signal height value for peak rejection.
Initial Area Reject	Sets an area value below which all peaks are rejected.
Shoulder Detection	Enable (on) or disable (off) shoulder detection.

Customizing Integration

It is often useful to change the values for the peak width, threshold and area reject so you can customize the integration of your electropherogram.

Figure 4-13 shows the influence of threshold, peak width and area reject on the integration of five peaks in a electropherogram.

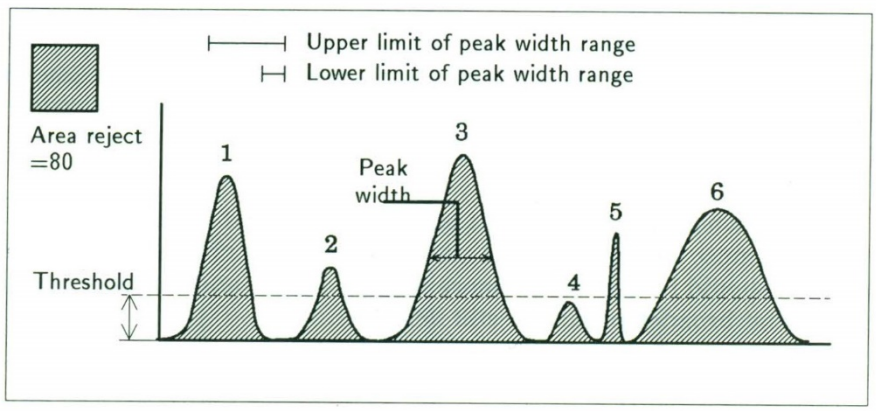

Figure 11: Using initial events

Integration Event	Peak 1	Peak 2	Peak 3	Peak 4	Peak 5	Peak 6
Threshold	Above	Above	Above	Below	Above	Above
Peak Width Range	Within	Within	Within	Within	Below	Above
Area Reject	Above	Below	Above	Below	Above	Above
Peak Integrated	Yes	No	Yes	No	No	No

2. Quantitative analysis

After peak identification and integration, the next step is quantitative evaluation. Peak area or peak height is used for the calculation of the unknown concentration of the components. The process consists of 3 steps:

- analyzing standard samples (standard is a sample solution containing the analyte of interest in known concentration)
- analyzing our sample that contains the analyte of interest in unknown concentration
- comparing the response signals obtained from analyzing the analyte of known and unknown concentration; this way the unknown concentration can be determined (the comparison of response signals can be carried out only if in the course of the separation and integration identical conditions were applied).

For the calculation of the unknown concentration of a compound in our sample, the following methods can be used:

- Percent composition determination
- Normalization
- External standard method
- Internal standard method

3. Spectral analysis

Spectral analyis is for processing the spectral data obtained from a diode array detector working in the UV-Visible region of the spectrum. The purposes of spectral analysis can be the following:

- peak purity inspection
- qualitative determination of each analyte peak with the help of spectral library search
- specifying the optimal detection wavelength for each component

Peak purity analysis

One of the most burning questions when performing separations is how many components correspond with a given peak. Hidden contaminations may falsify the results, leading to significant information being lost. Our objective, therefore, is to find out whether our peaks are pure or not, which can easily be done by comparing the spectra of the given peak. Usually 3 spectra/peak are used for the determination of peak purity (2 spectra at the inflection points + 1 spectrum at peak apex). If the 3 spectra are not identical, then the peak contains contamination

or perhaps it may be a result of background absorption. It is important to note that if the spectra are identical the peak might still contain contaminants. It can happen that the absorption of the contaminant is minimal as compared to that of the main component or that the spectra of the main component and the contaminant are roughly identical. The purity factor expresses the degree of similarity.

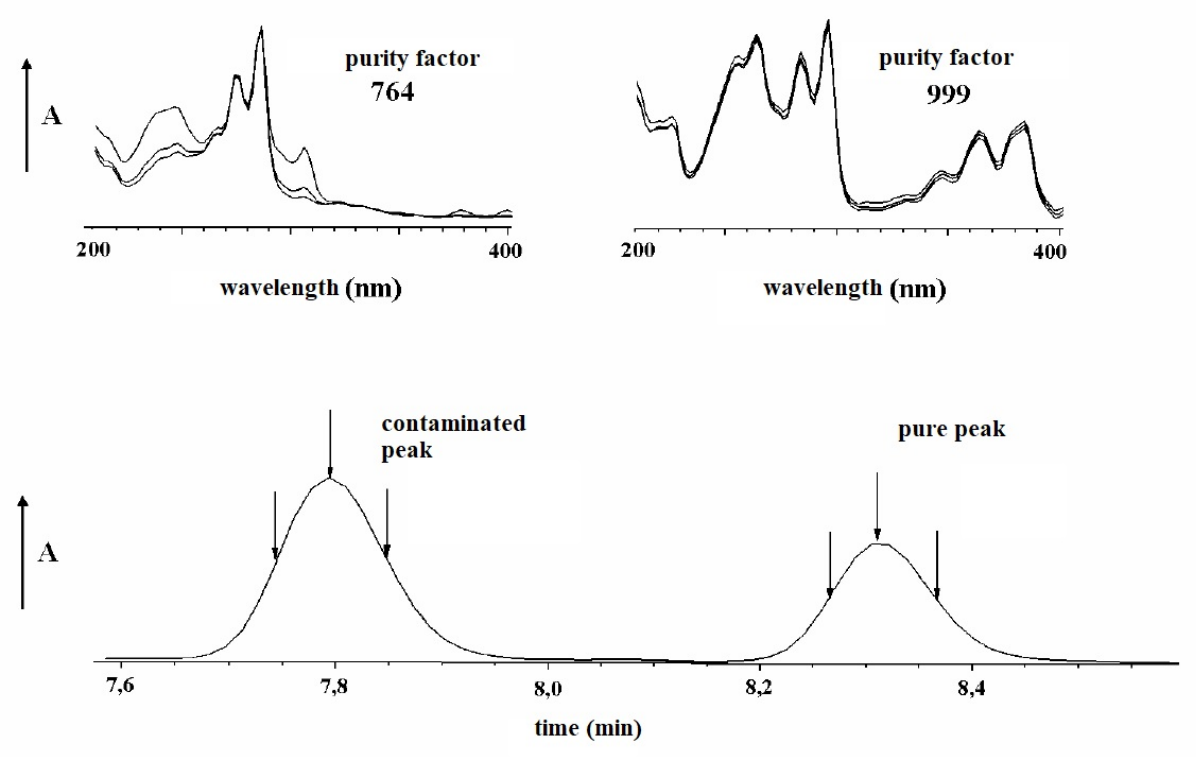

Figure 12: Peak purity analysis by investigating the overlaid spectra

Supplementary material for the statistical evaluation of measurements

Because of our ”inability” to perform perfectly precise and accurate measurements we must introduce the concept of errors. The error is the difference between the experimental value and the true value. The two types of experimental errors are: systematic and random error. In order to minimize random error, consecutive measurements are conducted. This brief section deals with the fundamentals of treating analytical data.

Calculation of the arithmetic mean (average):

$$\overline{x} = \frac{\sum_{i=1}^{n} x_i}{n}$$

where ($\bar{x}$): average, n: number of measurements

Calculation of standard deviation:

$$s = \sqrt{\frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n-1}}$$

Calculation of relative standard deviation:

$$s_r(\%) = \frac{s}{\bar{x}} 100$$