

Self-modeling factor analysis combined with analysis of the signal derivative for separation of incompletely separated chromatographic peaks

Yuri A. Kalambet, Sergey Maltsev, Ampersand Ltd., Moscow, Russia;

Yuri P. Kozmin, Shemyakin Institute of Bioorganic Chemistry, Moscow, Russia

Mailto: kalambet@ampersand.ru

Abstract

Factor analysis is widely used [1] for separation of incompletely separated chromatographic peaks. We combined this technique with an analysis of the angle derivative of the multi-channel chromatogram to get a better choice of basic spectra. Improvement in the algorithm is demonstrated through examples.

Specifics of Factor Analysis in Chromatography:

1. Concentration profiles are smooth
2. Concentration profiles of individual chemical components are unimodal

Consequences:

If pure spectra of all components are available within our data set (self-modeling factor analysis), no more than two components can overlap at one time.

Spectra, adjacent to pure spectrum, have small concentrations of neighbor compounds and hence are very collinear with the spectrum of the pure substance. The angle (= absolute value of angle derivative) between this spectrum and its neighbors is small.

3. The signal may be out of range for some of channels
4. Different channels have different accuracy of measurement and baseline noise

Solutions:

1. Space normalization to reduce heteroscedasticity
2. Reconstruction of the out-of-range points
3. Selection of basic spectra in SMFA using “purity angle” criterion in addition to usual criteria of minimal correlation with already selected basic spectra

Implementation:

1. Baseline noise detection using the whole chromatogram – noise vector is constructed from baseline noise values [2]. This vector is used to convert the initial data matrix to the matrix of Signal-to-Noise ratios. All steps of Factor Analysis are performed using this matrix of S/N ratios, and an inverse conversion is made just before the output of results. Benefits: signals of different origins can be joined into one matrix; uninformative coordinates do not influence the results of FA.
2. Channels that have out-of-range points do not participate in initial FA. A concentration matrix, constructed on this step, is used for the reconstruction of missing coordinates of spectra by the least-squares method, one channel at a time, as every channel has a different number of out-of-range points and hence a different number of least-square equations to be solved.
3. Selection of basic spectra vectors is performed by a combination of two criteria: first is a distance from already selected spectra, second is the best “purity angle” [2] in the first group. Group size depends on the eigenvalue, so that for minor components purity angle is less important than for major components.
4. An algorithm is able to accept the known spectra of some components. In this case, projection of the known spectrum onto reduced space is selected as one of the basic vectors.

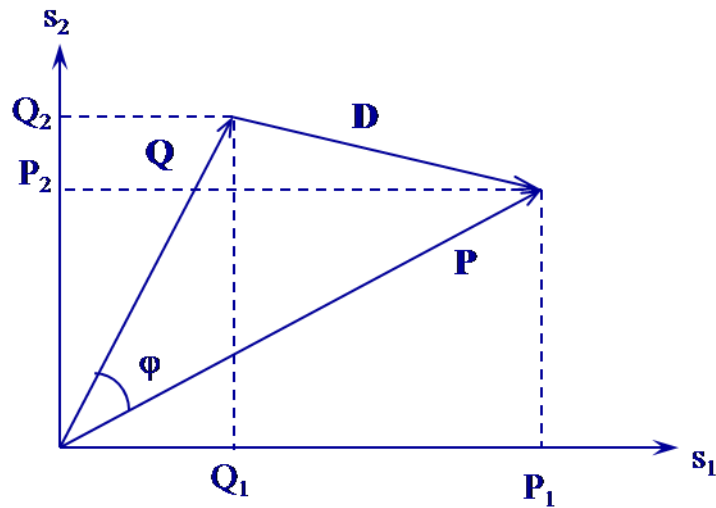


Figure 1. Inter-spectra angle definition example. Measurements are performed using only two wavelengths (s_1, s_2). Spectra of two peaks are measured: $P(P_1, P_2)$ and $Q(Q_1, Q_2)$. An angle between measurements is an angle between vectors P and Q .

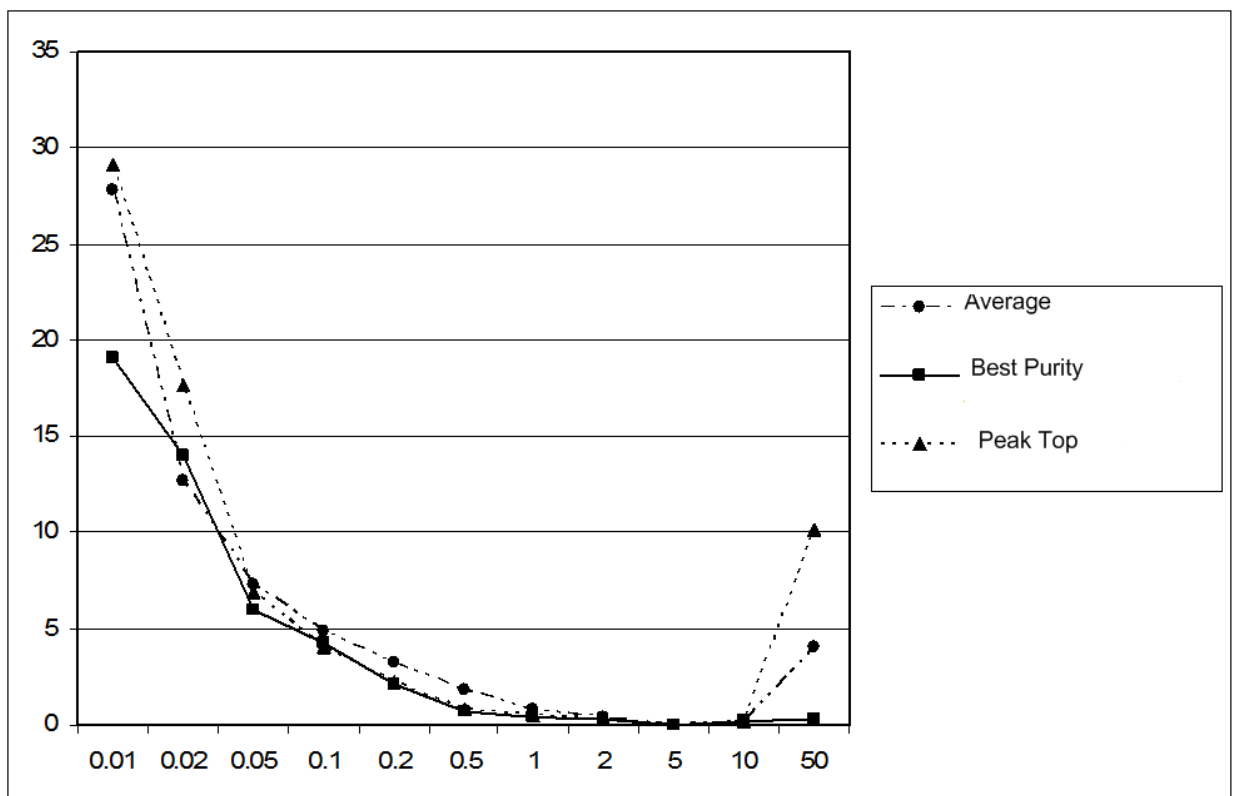


Figure 2. Dependence of angle between “true” and observed spectra of the peak (Y axis) on concentration of the substance (X axis). At concentration 50, the detector is out-of-range for some of channels (wavelengths), “best purity” spectrum is good; average and peak top spectra – not good.

Implementation in Chrom&Spec software

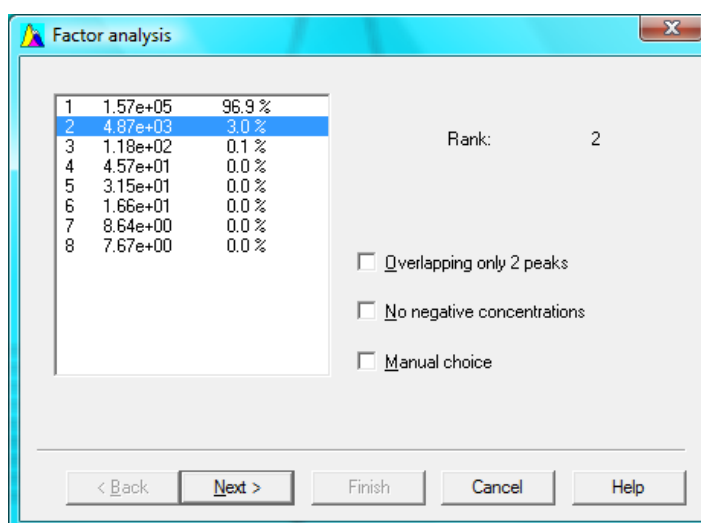


Figure 3. Factor analysis step 1

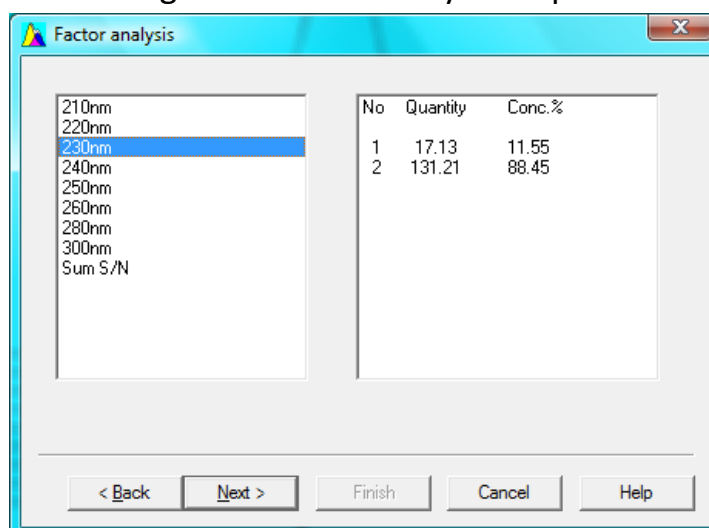


Figure 4. Factor analysis step 2

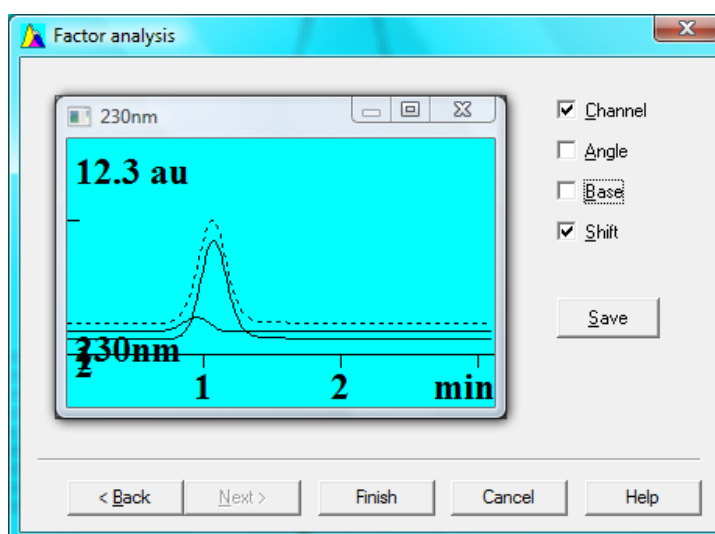


Figure 5. Factor analysis step 3

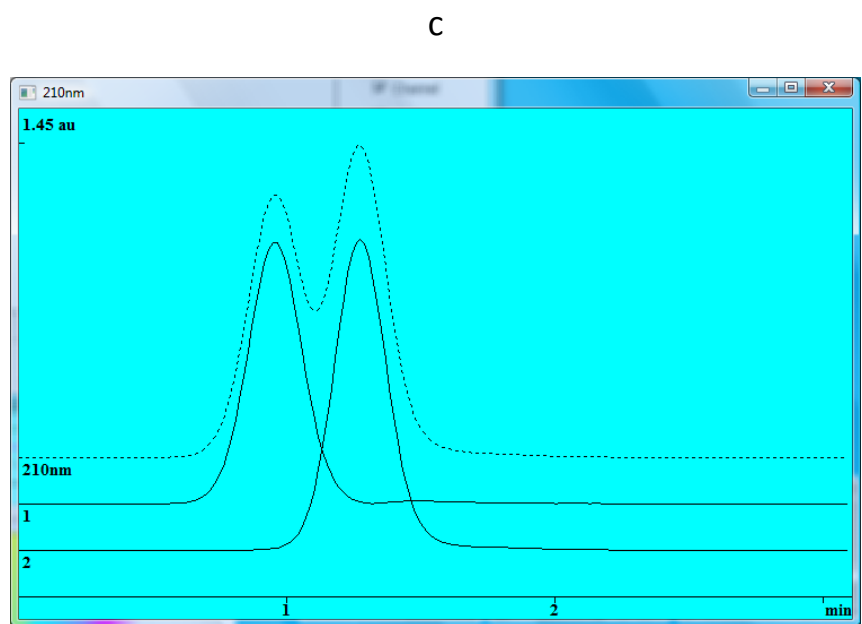
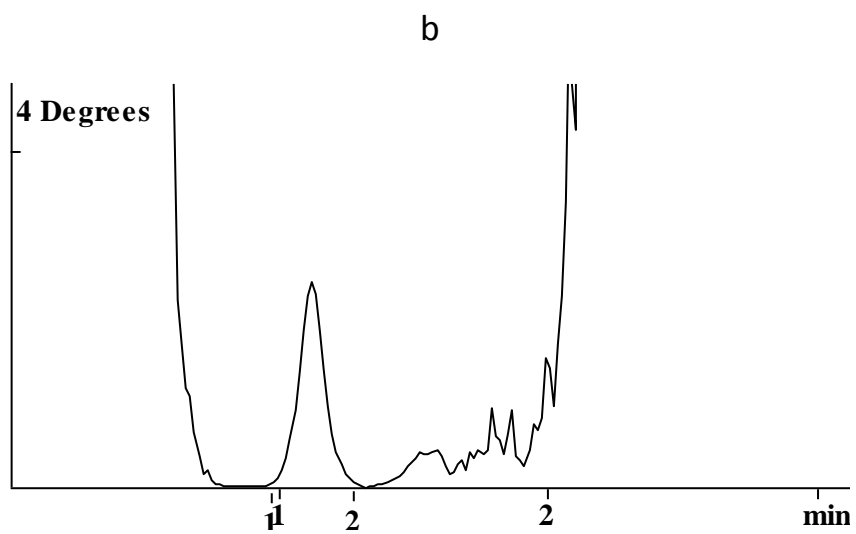
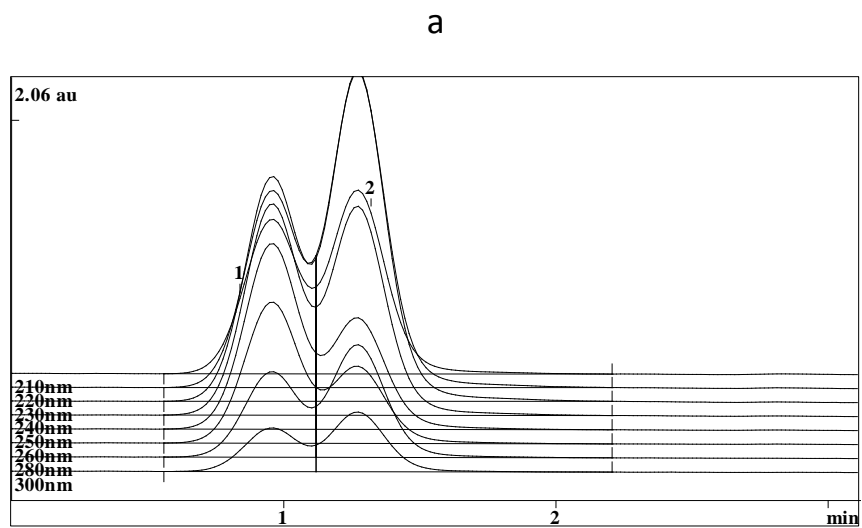


Figure 6. 46:54% a) – raw data; b) – angle profile; c) – separated peaks

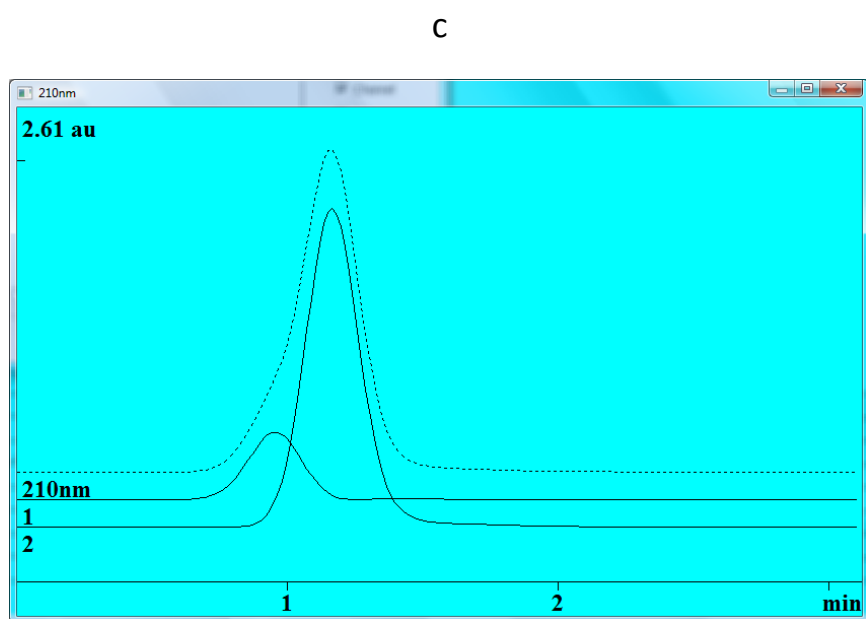
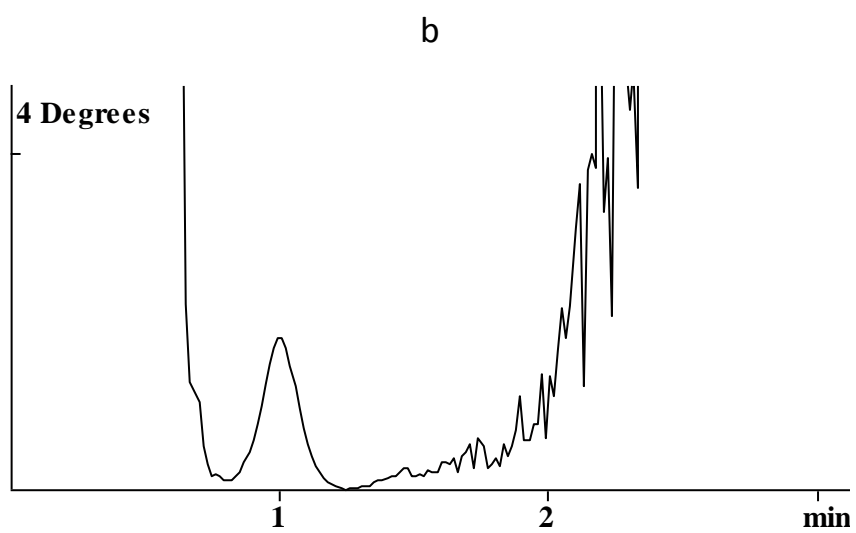
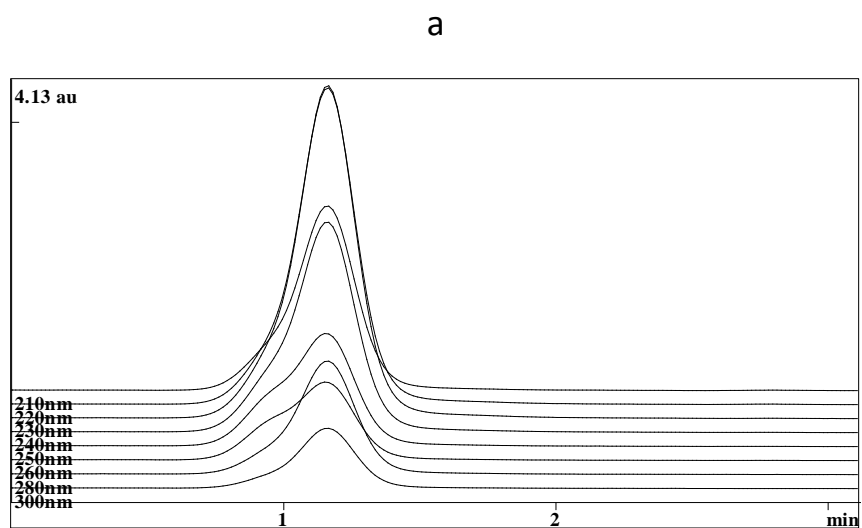


Figure 7. 94:6% a) – raw data; b) – angle profile; c) – separated peaks

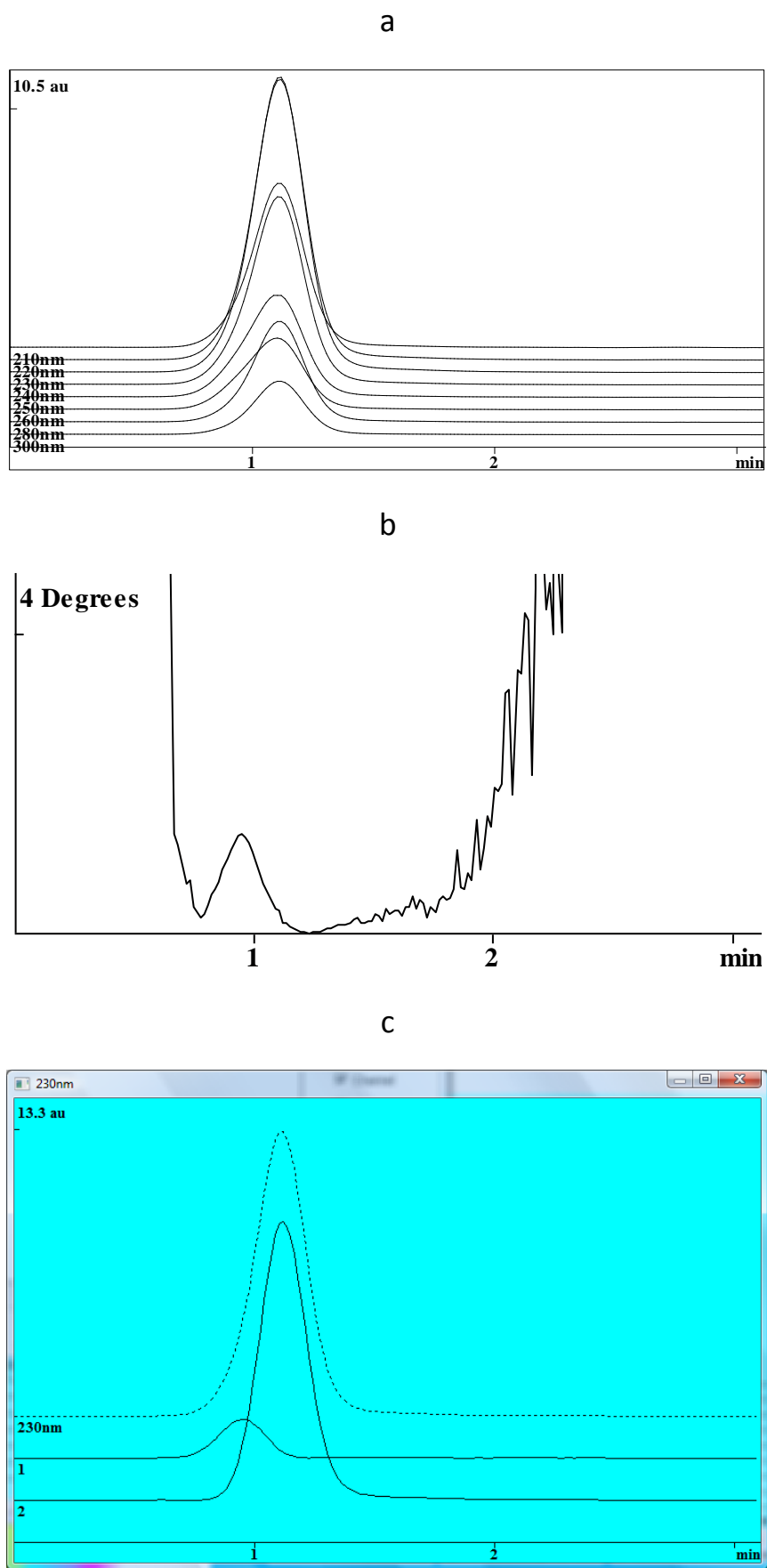
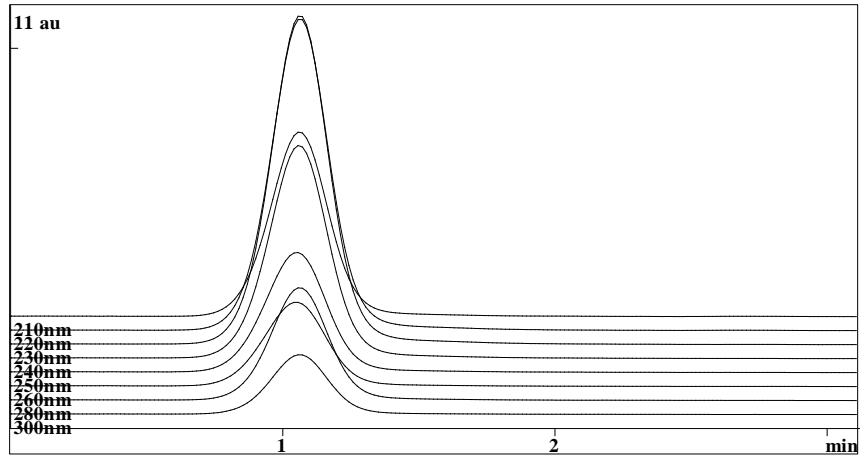
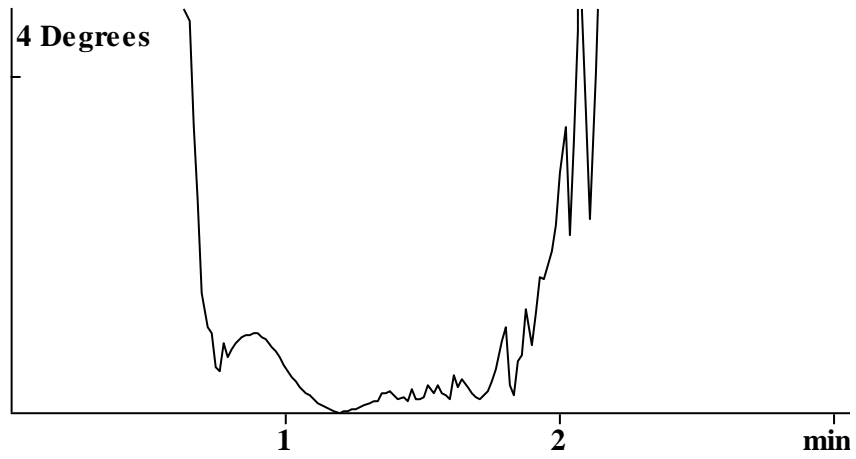


Figure 8. 96:4% a) – raw data; b) – angle profile; c) – separated peaks

a



b



c

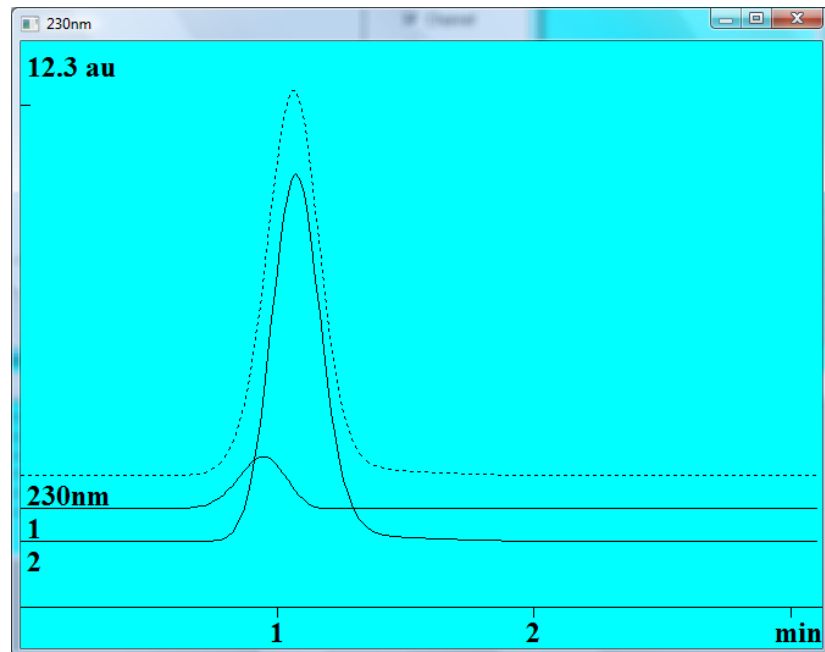


Figure 9. 97:3% a) – raw data; b) – angle profile; c) – separated peaks

References:

1. E.R.MALINOWSKY. Factor analysis in chemistry. Wiley, NewYork,1991.
- 2.Yu. A. KALAMBET, Yu. P. KOZMIN, and M. P. PERELROYSEN. Computer spectrochromatography. Principles and practice of multi-channel chromatographic data processing, *J. Chromatogr.*, **1991**, 5421, 247-261.