SIMULTANEOUS RESOLUTION OF OVERLAPPING TAILING PEAKS OF ISOMERS USING DUAL DETECTOR DIODE ARRAY LIQUID CHROMATOGRAPHY ELECTROSPRAY MASS SPECTROMETRY

Richard G. Brereton and Conrad Bessant
School of Chemistry
University of Bristol
Cantock's Close
BRISTOL BS8 1TS
UK

ABSTRACT

It is now common to obtain information from two or more detectors in HPLC simultaneously. Conventionally both DAD-HPLC and LC-MS data are obtained independently, often using quite different approaches, but there will be common features in both types of information. However, these will be distorted, for example, Electrospray MS profiles are typically more noisy and tailing compared to those from DAD-HPLC due to the different physical processes of passing from the chromatographic column to the detector. A method for obtaining and processing both types of data simultaneously is described. The analysis of isomers with identical molecular weights and similar elution times is particularly important in the pharmaceutical industry where small impurities can have a major influence on biological activity. The approach is illustrated by application to a mixture of 2- and 3-hydroxypyridine at varying levels of overlap.

INTRODUCTION

Coupled chromatography plays a major role in modern analysis, especially in the pharmaceutical industry. Single detector chromatography has been replaced with coupled methods such as DAD-HPLC (diode array detector high performance liquid chromatography), LCMS (liquid chromatography mass spectrometry), LCNMR (liquid chromatography nuclear magnetic resonance). This is especially important in the pharmaceutical industry for a wide variety

of applications, such as impurity monitoring in process control, combinatorial chemistry and parallel synthesis.

With improvements in technology it is now common to couple more than one detector simultaneously to an HPLC machine, for example, LC-NMR-MS or DAD-LC-MS. The latter is cheap and convenient, desktop MS and uv (ultraviolet) detectors being widespread. Another additional advance over the past few years has been in MS ionisation techniques, with ESI (electrospray ionisation) becoming a method of choice. It is possible to control the level of fragmentation in ESI by adjusting instrumental conditions (e.g. cone voltage) and so to study isomers with identical molecular weights. In this paper we report a common approach involving an HPLC interface both to an ESIMS and a DAD detector, allowing simultaneous uv and MS of a mixture. Information from both detectors is complementary. The uv spectra are quantitative and have low noise, whereas the MS contains valuable structural information.

In many industrially important situations, the interpretation of a chromatogram is not always straightforward. A well-known example consists of the detection of isomers of compounds, often as synthetic by-products. Whereas it is inevitable that there will be small quantities of isomeric material in almost all synthetic reactions, there are strong international regulations about the purity of drugs. In particular, isomers can have a devastating biological effect, the thalidomide case of the 1960s in which a fertility drug caused deformed babies by contamination with an optical isomer is an example. Often the amount and nature of impurities cannot easily be controlled, and these compounds often have identical molecular weights and similar chromatographic properties to the major synthetic target, posing significant problems for the analytical chemist.

DAD-LC-MS can be automated, for example, with samples obtained at regular intervals throughout a reaction or a manufacturing process, but the result is a large amount of data. Both detectors should provide information on common features, such as the elution times and amounts of components in a mixture. However, the peakshapes and signals from each detector are very different. Normally each type of information (DAD-HPLC and LC-MS) is analysed independently, but this misses out the common trends. Chemometric methods can be employed to look at both sources of information simultaneously, and so detect common trends.

In this paper we will use a model system of 2- and 3-hydroxypyridines. These have been chosen for the following reasons. They have identical molecular weights so there will only be small differences in their mass spectra, in the fragmentation pattern. They have prominent chromophores so are suitable for study by uv. They are readily available in pure form; it is important to realise that chemometrics techniques are very sensitive to small impurities which is a strong advantage in the application but a problem in method development. Finally their chromatography can be modified by changing the pH, meaning that a series of chromatograms can be obtained at a variety of different resolutions.

The aim is to produce a single consensus chromatogram that is obtained using information from both detectors. Since there is tubing between the end of the column and both detectors, and since the width and / or length differs for each detector, reflecting the different sensitivities of uv and MS, the peakshapes in both cases will be distorted, by differing amounts. In addition because MS is much less reproducible compared to uv, the apparent peakshape will appear much more noisy, largely due to changes in relative intensities of various ions. Finally it is possible that ions remain in the spectrometer after detection meaning that the apparent MS peak is much broader and more tailing than the uv peakshape. Detectors do not monitor the mixture as it elutes from the bottom of the column, however, it is an important aim of analysis to try to reconstruct the chromatographic mixture as it elutes. The distortion in the MS is far greater than that in the uv. The desired chromatogram is a consensus between both detectors, with the spectra of each component and elution times, clearly indicated.

In previous papers[1-3] we have discussed the analysis of such data from each detector independently, using quite different approaches for DAD-HPLC and LC-MS. Although there are some common trends, two different sets of results were obtained. Recently we proposed a new method that processes this data simultaneously[4]. We have reported, previously, methods for deconvolution of GC-MS on its own[5].

EXPERIMENTAL

Chemicals and Solvents

All chemicals and solvents used in the analyses were of analytical reagent grade and HPLC grade, respectively, unless otherwise stated. The two compounds of particular interest are 2-hydroxypyridine, **I**, and 3-hydroxypyridine, **II** (Acros Organics, New Jersey, USA), which were 97% and 98% pure, respectively. The ammonium acetate and acetic acid were purchased from Sigma (Poole, England), the HPLC grade methanol and ammonia from Rathburn Chemicals Ltd. (Walkerburn, Scotland), and the deionised water was prepared using a Milli-Q filtration unit (Millipore Corporation, Massachusetts, USA).

Reagents and Standards Solutions

A 0·05 M CH_3COONH_4 solution was prepared in deionised water (3·854 g L^{-1}) and adjusted to the desired pH by adding 10 % CH_3COOH and 10 % NH_3 dropwise. From this the mobile phase was prepared containing 98 % 0·05 M ammonium acetate and 2 % methanol. All standard solutions of compounds **I** and **II** were prepared in the pH adjusted mobile phase from stock solutions of 10 mg mL^{-1}.

Apparatus and Instrumentation

All HPLC was carried out using a Waters system (Waters Corporation, Milford, Massachusetts, USA) comprised of a 616 LC Pump, a 717 Plus Autosampler with Heater/Cooler, and a 600S Controller. Diode Array Detection was performed using a 996 PDA Photodiode Array Detector Optics Unit (Waters Corporation). The HPLC-DAD system uses the Millennium Session Manager Software (Version 2.15.01, Waters Corporation) with the Millenium 2010 Chromatography Manager Add-On (Version 2.10, Waters Corporation), that runs under Windows (Version 3.1, Standard Mode, Microsoft, Seattle, USA) on a 586-PC. The stationary phase was a 100 mm C_{18} reversed-phase Symmetry column with 3·5 µm particle size, and a 4·6 mm internal diameter (Waters Ltd., Watford, England).

The mobile phase was isocratic and consisted of the 0·05 M ammonium acetate - methanol buffer (98 % - 2 %). The flow-rate through the column was 0·8 mL min^{-1} and the sample injection volume 5 µL. The analyses were carried out at ambient temperature for a run time of 10 min. The Electronic Absorption Spectra were recorded using the DAD at 1 s intervals, between 200 nm and 400 nm (1·2 nm resolution), with a flow-cell of pathlength 10 mm.

All MS was performed using a VG Quattro Mass Spectrometer (Fisons Instruments, Altrincham, England) controlled using MassLynx software (Version 2.1), which runs under Windows (Version 3.1, Microsoft) on a 486-PC. The MS were analysed using electrospray ionisation (ESI), a technique that can be used in both positive and negative ion modes, but was used in the positive ion mode for these studies. ESI produces a plume of charged droplets, also the result of Coulombic repulsion, which enter into the mass analyser. The MS were recorded every 1 s, between 40 and 200 mass units, with a cone voltage of 50 V, and a source temperature of 80°C.

HPLC-DAD-MS coupling

There are various possible arrangements for the eluent to pass through the detectors. In this study we used the configuration of Figure 1 with a 50 : 50 split (250 mm : 250 mm), *i.e.* a flow-rate of 0·4 mL min^{-1} to both the EAS and the MS detectors. It is important to recognise that the splitting is determined by the relative sensitivity of the compounds being analysed to each detector, in turn relating to chromophores and ionisation mechanism.

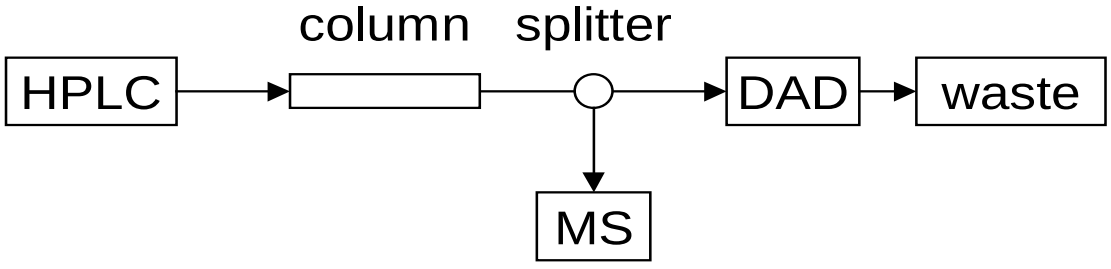


Figure 1 Split used

DATA ANALYSIS

Decoding data

The first step is to transfer the data from most DAD and MS computers to a common format. Two programs are required to decode the data acquired from the HPLC-DAD analysis. The first of these is the 2010 DDE Assistant for Raw Data macro (Version 2.10, Waters Corporation) that runs in Excel (Version 5.0a, Microsoft). This is used in conjunction with the Millennium software in order to extract the appropriate data from the Oracle database. The resulting file is then converted to a matrix using a second VBA macro that was written in-house. The MS data is decoded using a C++ program written by Dr. R. L. Erskine.

Preprocessing

One experimental problem is that it is not easy to record both LCMS and DAD-HPLC at equally spaced intervals in time, and also precise alignment is not possible during recording. Therefore the information must be processed in advance.

For the DAD-HPLC and LCMS the first steps are

1. to interpolate each dataset to 1 s intervals and
2. to baseline correct the data.

The next step applies the LCMS and involves mass selection. In mass spectrometry, the majority of masses do not contain useful information, and consist primarily of noise. Typically of several hundred m/z ions recorded only twenty or so are useful. There are several stages in reducing the number of masses as follows.

1. Discard all low masses (below m/z 40).
2. Determine a significance ratio, which involves looking at the summed intensity of each mass over the region of the chromatogram of interest, relative to all the masses, and then retain only the most intense. At this point N (typically equal to 20) masses are selected.

3. Determine the noise content of each mass by looking at the root mean square of the noise, and retain the least noisy. At this point *M* (typically equal to 13) are selected.

Manipulation of mass spectra

The key to this method is to change the mass spectral peakshapes so that they are comparable to the diode array peakshapes. The latter are better resolved and contain less noise. Each of the *N* ions selected above is treated separately.

The first step is simply to reduce the noise content of these *N* masses further by using a low pass filter[6]. This is a common approach in signal analysis used to remove background noise.

After this, two masses are selected, being the most diagnostic for each compound. In order to obtain meaningful, and noise-free, results, this is performed only on the *M* masses using the third selection criterion described above, reducing the risk of using a very poor mass. This involves a number of steps. First of all the profiles for each mass are normalised, this involves placing them on the same scale, so that an intense mass is given equal significance to one that is weaker. Sometimes the most dominant ions (e.g. the molecular ions) are not diagnostic of any specific component in the mixture, and if this step was not performed the analysis would be dominated by the ions at m.z 96 and 97 (in this example), which would lose structural discrimination. Further details of this are available elsewhere[7]. A chemometric method called Principal Components Analysis (PCA)[8-10] is then performed on the normalised chromatogram. Two types of graph are obtained.

1. A scores plot spreads the elution times out. Dependent on how the data has been treated, the regions of purity of the two compounds are characterised by regions that are approximately linear.
2. A loadings plot spreads out the masses. The masses corresponding to pure compounds are clustered.

The principles are illustrated in Figure 2. In the case studied in this paper, two principal components are calculated since there are only two compounds in the region of the chromatogram studied. The

scores and loadings plots can be compared, and this allows two masses most diagnostic of each compound to be selected.

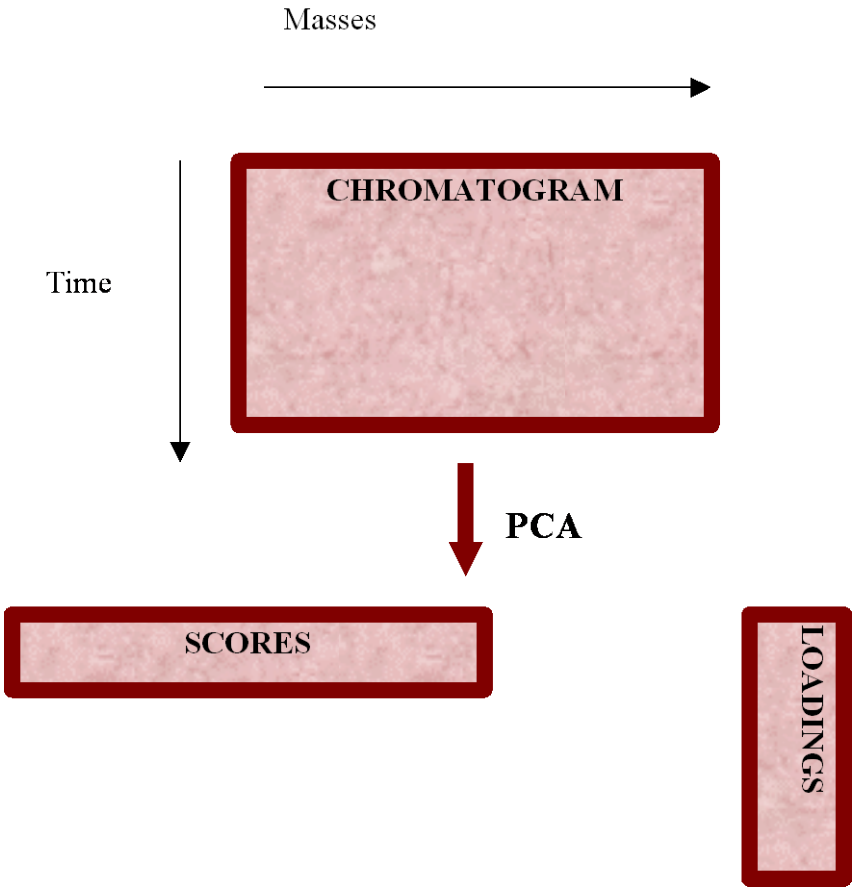


Figure 2 Principles of PCA as applied to LCMS

The two pure masses are then fitted to a mathematical function. The peakshapes in LCMS are highly asymmetrical, so the right and left side need to be approximated by different functions. In this work we use a Lorentzian-Gaussian function[11] which is common in chromatography, the Lorentzian half representing the right hand (tailing) portion of the peak.

The full *N* masses are then included. Each mass will be partially diagnostic of a particular compound, so the peakshapes are fitted to a function of the form

$$x_n = a\, x_a + b\, x_b$$

where x_n is the profile for mass *n*, and *a* and *b* are coefficients corresponding to the amount of each compound in the mixture. For the most diagnostic peaks, the two parameters *a* and *b* are very different, for peaks that arise from both compounds such as the molecular ions, these two parameters are nearly equal. Occasionally, for very noisy masses, one coefficient appears to give a negative value, this is replaced by 0, and is a consequence of noise.

Reshaping the LCMS and Deconvolution

At this stage, a series of *N* (usually 20) masses has been identified, each being fitted to a function relating to the proportion of each component in the mixture.

The next step is to change the peakshapes in the LCMS. This is done by using reference peakshapes from DAD-HPLC, obtained by deconvolution, and changing the LCMS peakshapes. By this procedure peakshapes and alignment from both detectors is identical, and the information is then treated as two matrices with a common chromatographic dimension, as presented in Figure 3.

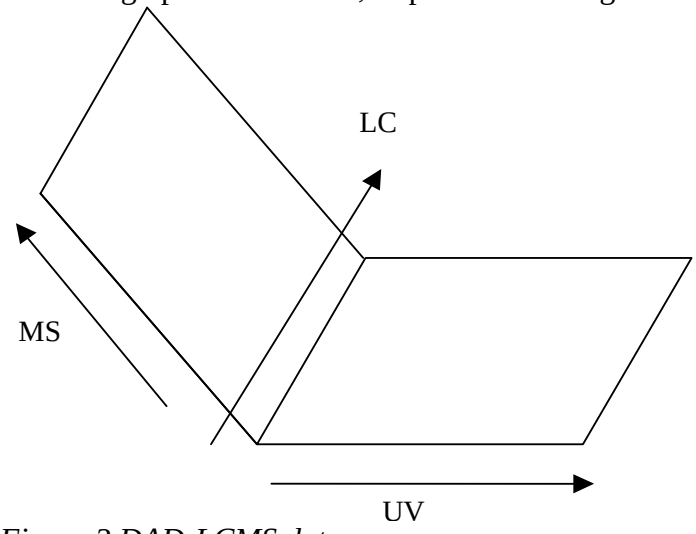


Figure 3 DAD-LCMS data

The final stage is deconvolution. The reshaped chromatographic profiles are used as to obtain an estimate of the pure uv/vis spectra in the DAD-HPLC. This in turn allows refinement of the profiles which are then employed to produce an estimate of the mass spectra in the reshaped LCMS.

Deconvolution uses to advantage the best features of both detectors as follows.

- The LCMS has characteristic ions that are helpful as initial estimates of pure concentration profiles.
- The DAD-HPLC has better peakshapes.
- The DAD-HPLC is less noisy.

A consensus chromatogram can then be obtained, with two elution profiles and corresponding mass spectra and uv/vis spectra.

RESULTS

Analysis of pH 5.0

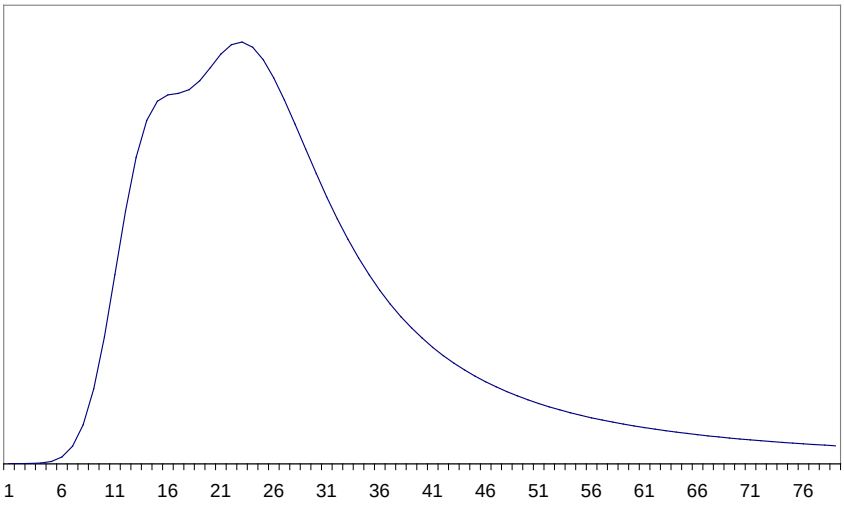


Figure 4 DAD-HPLC chromatogram in region of interest (scale in s)

In order to illustrate the method we will concentrate on the chromatogram obtained at pH 5.0.

Figure 4 is of a region of the DAD-HPLC, the chromatogram summed over the wavelengths 200 - 400 nm. It can be seen that there are two partially overlapping peaks. The corresponding raw LCMS TIC trace (over a wider chromatographic range) is shown in Figure 5(a). At first this appears quite discouraging, although there appears am ion cluster.

Figure 5 LCMS TIC chromatogram : scale in s (a) Top : raw data (b) Bottom : after interpolating, mass selection, baseline correction and selection of region.

Reducing the masses to the 20 most significant (as assessed by significance ratio), baseline correcting and interpolating, results in a portion of the chromatogram that demonstrably appears to consist of a clear peak cluster, as illustrated in Figure 5(b).

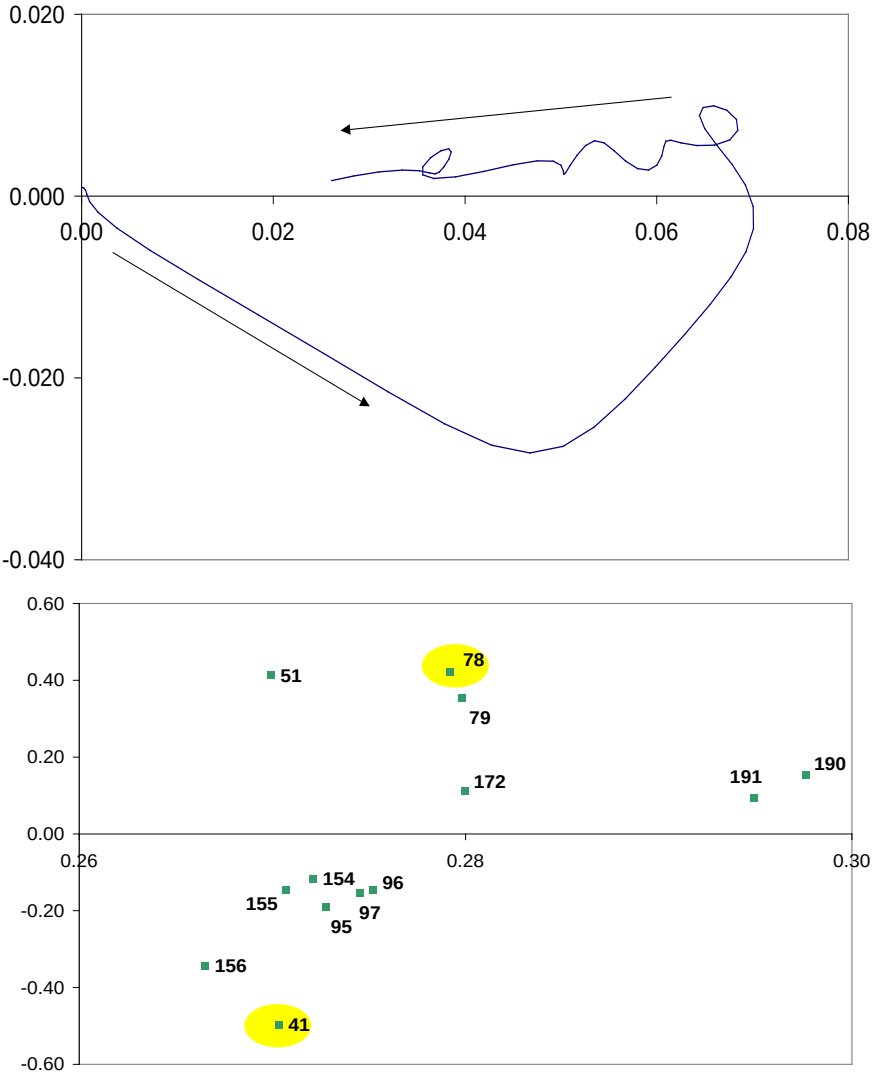


Figure 6 (a) Top : Scores and (b) Bottom : Loadings plots of the LCMS in the region of interest using the top 13 masses.

After selecting the top 13 masses, using the noise content to reduce the number from 20, the scores and loadings are calculated as illustrated in Figure 6.

The scores plots follow the evolution of the two compounds with time. The bottom right limb corresponds to the purest point for the fastest eluting compound, and the top limb for the slowest eluting compound. Only two PCs are retained since there are only two components in the mixture. Comparison of the scores and loadings graphs give clues as which masses are most diagnostic of each compound. We choose m/z 41 as diagnostic of the fastest eluting compound, **II** or 3-hydroxypyridine, and m/z 78 of compound **I** or 2-hydroxypyridine. Notice how the molecular ion cluster around m/z 96 lies in between. There are also some high m/z ions which probably arise from association.

Figure 7 is of the raw mass spectral profiles of these two ions. It would be possible to choose a number of alternative ions that are diagnostic of each compound.

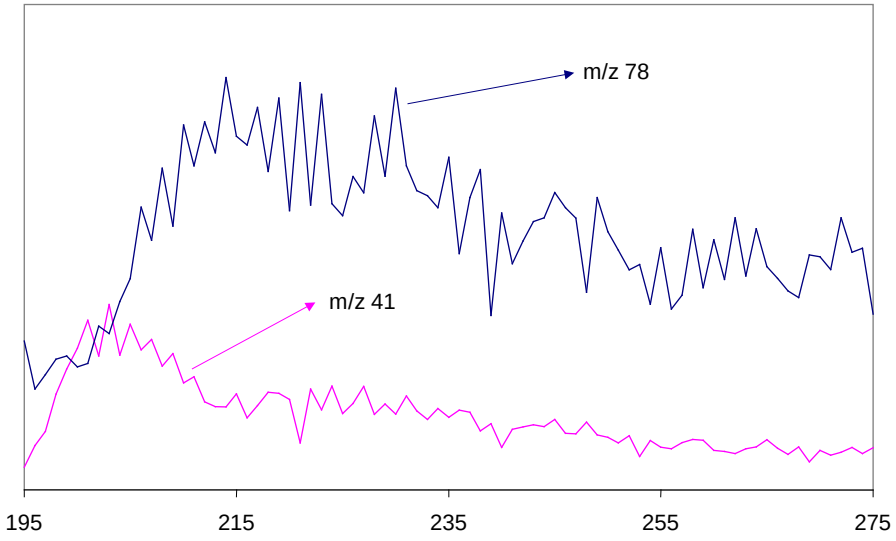


Figure 7. M/z 78 and 41 chromatograms in the region of interest.

The profiles after denoising and subsequent fitting to a Lorentzian-Gaussian model are presented in Figure 8.

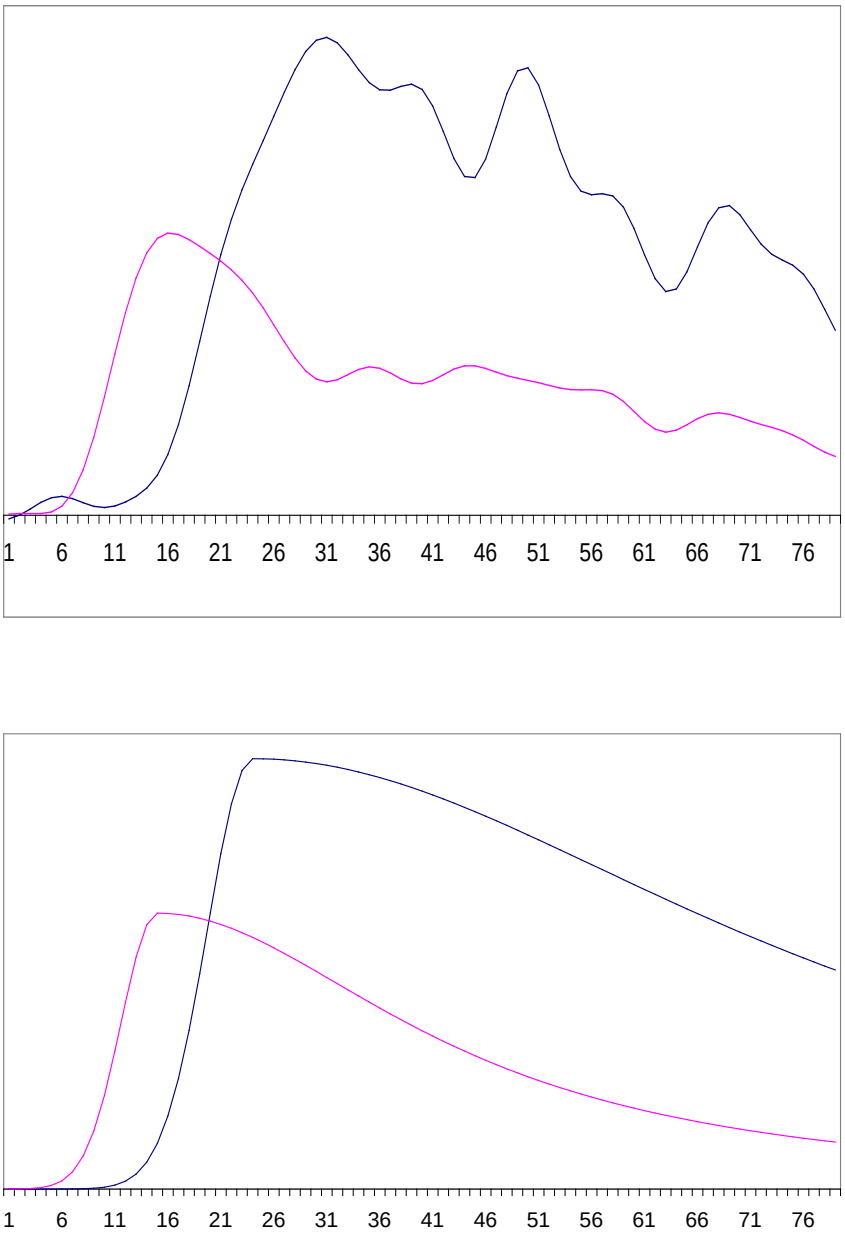


Figure 8 Profiles at m/z 41 and 78 after (a) Top : denoising and (b) Bottom : fitting to a Lorentzian-Gaussian peakshape model

The next stage is to fit the 20 mass spectral profiles to a sum of Lorentzian-Gaussian models. Figure 9(a) is of the coefficients (obtained using a Kalman filter) for each of the masses. This can be compared to the loadings plot, the grouping of masses shows some similarities. Notice the extra 7 noisy masses have now been included. The fitted profiles for all the masses, superimposed, are presented in Figure 9(b).

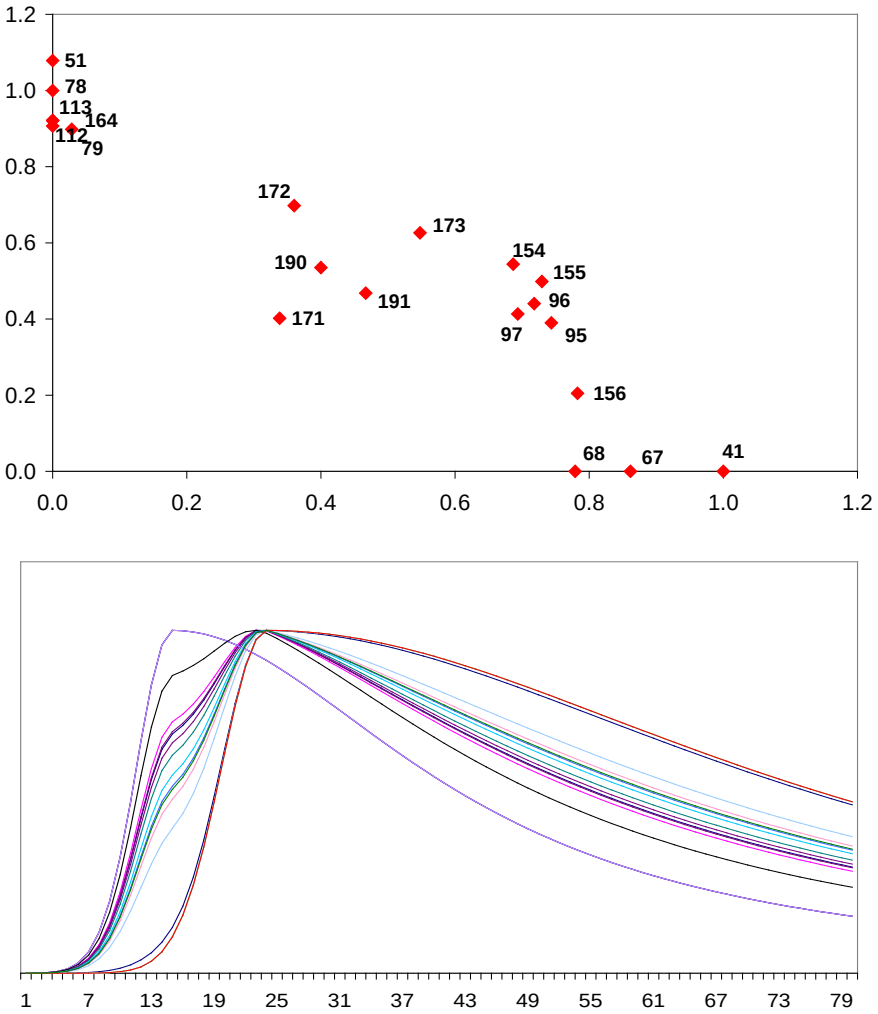


Figure 9 (a) Top : Coefficients for the 20 masses fitted to a two peak model and (b) superimposition of the fitted mass spectral peaks.

The next step is top reshape the pure peaks. This makes a dramatic difference to the appearance of the LCMS, and is illustrated in Figure 10.

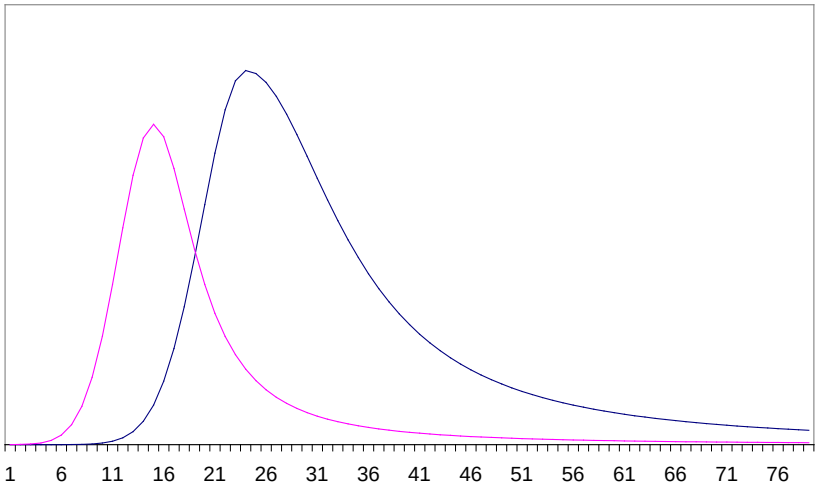


Figure 10 Reshaped peaks at m/z 41 and 78.

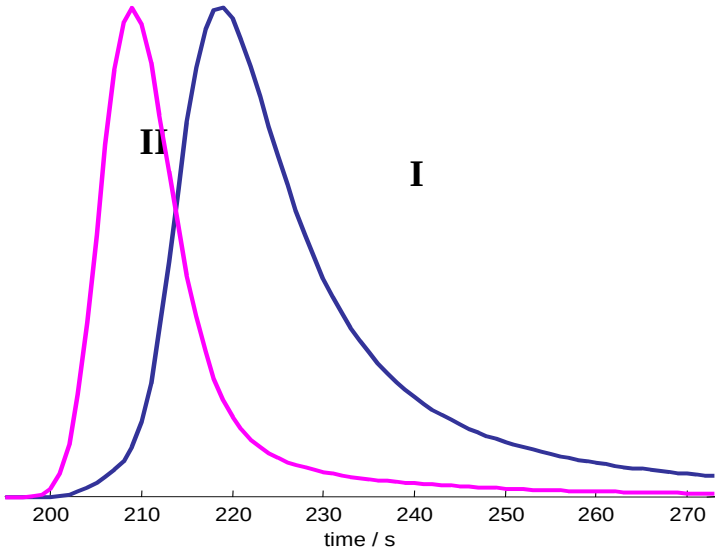
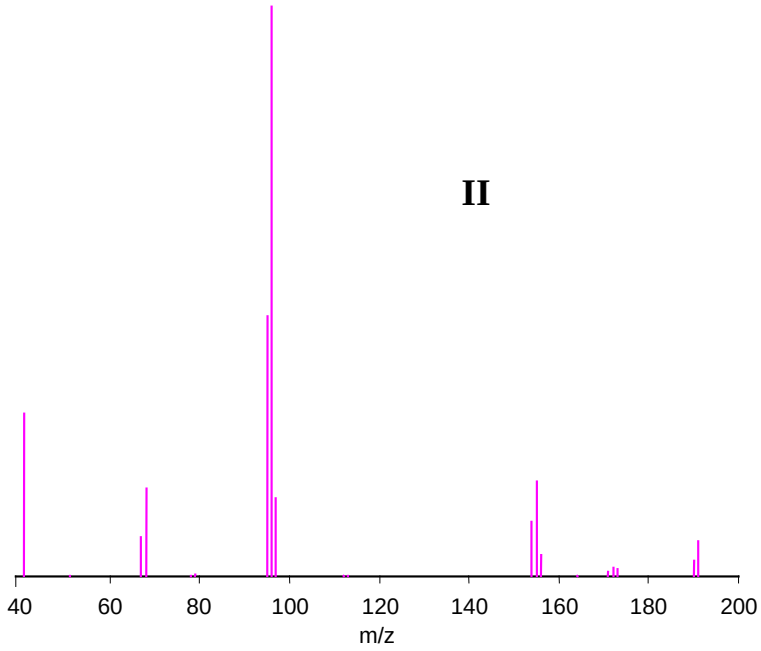


Figure 11 final chromatographic profiles at pH 5.0.


II
40
60
80
100
120
140
160
180
200
m/z

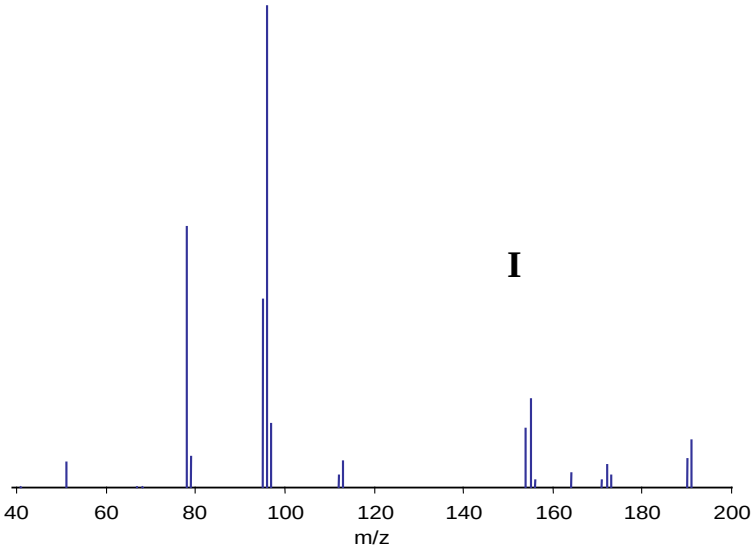
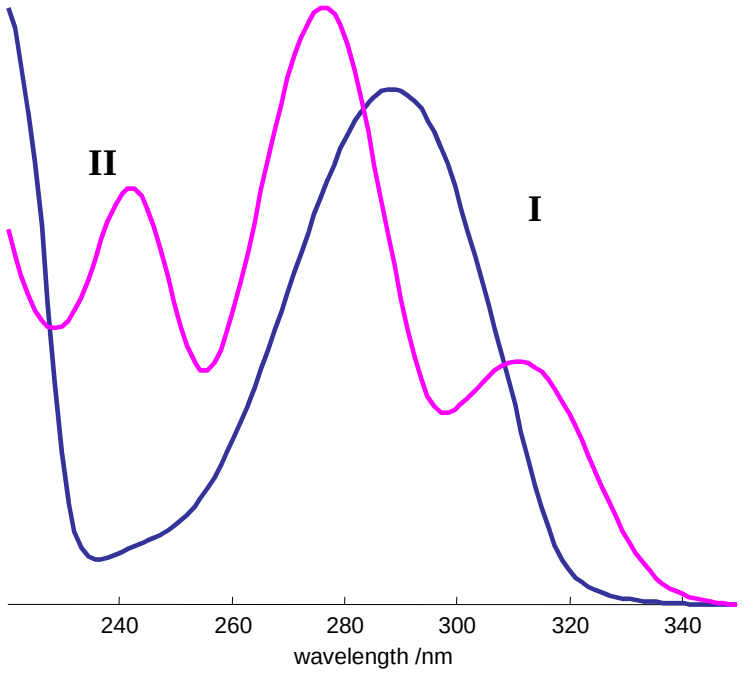

I
40
60
80
100
120
140
160
180
200
m/z

Figure 12 Reconstructed spectra



The final stage is deconvolution. The refined peakshapes (rescaled) are presented in Figure 11. Figure 12 is of the estimated uv/vis spectra and mass spectra. These are very close to what is known of the spectra of the pure compounds. Hence it is possible to produce a consensus chromatogram of all the data together.

Analysis of chromatograms at pH 4.8 and 5.2

These chromatograms are illustrative of different levels of overlap. At pH 4.8, the two compounds are largely resolved. At pH 5.2 there is a cross-over in elution times. In addition, the spectra change with pH.

The LCMS and DAD-HPLC chromatograms, at pH 4.8, in profile, after interpolation, baseline correction and mass selection are presented in Figure 13.

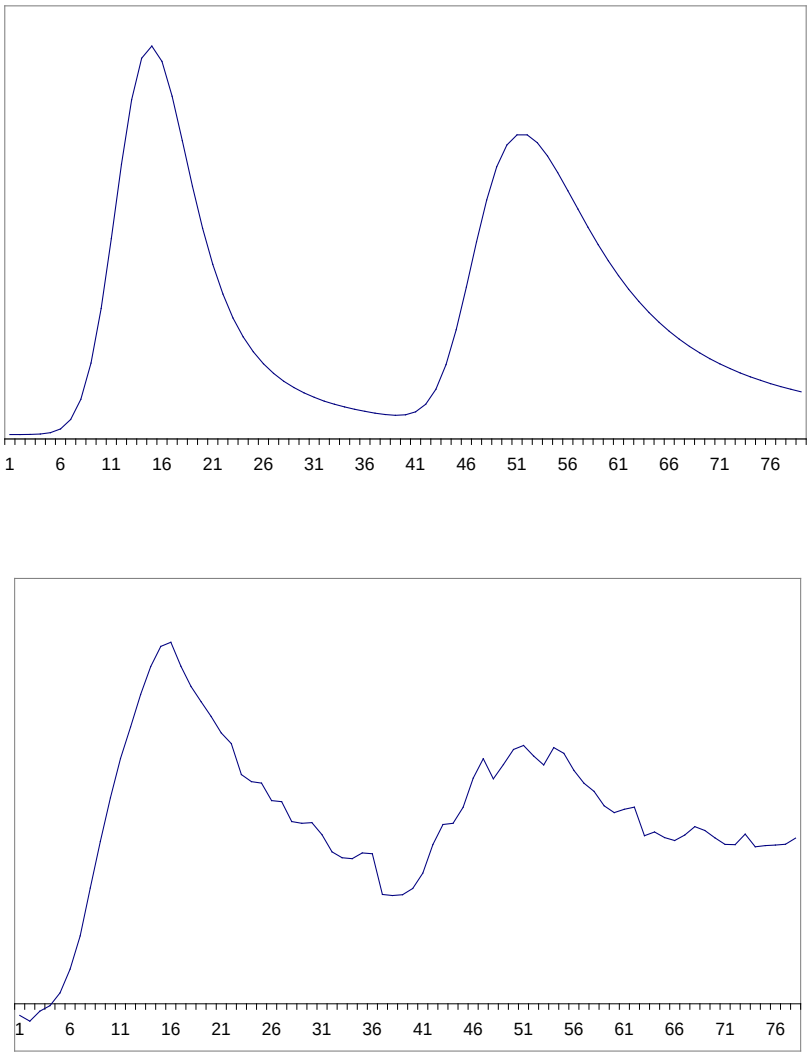
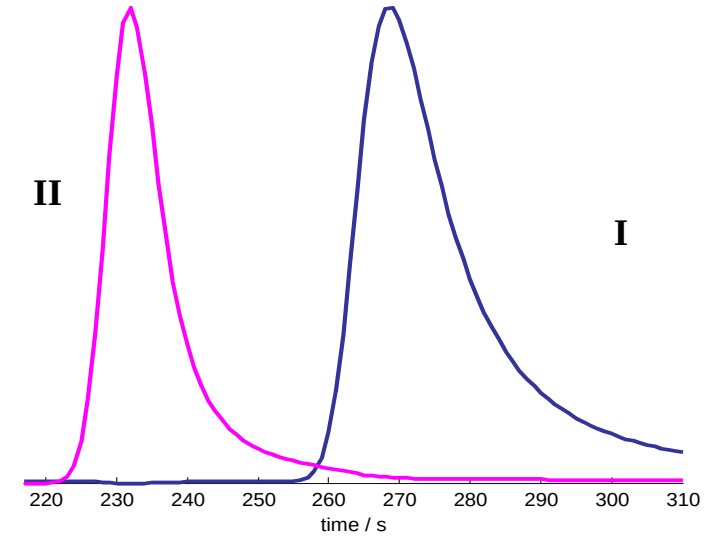
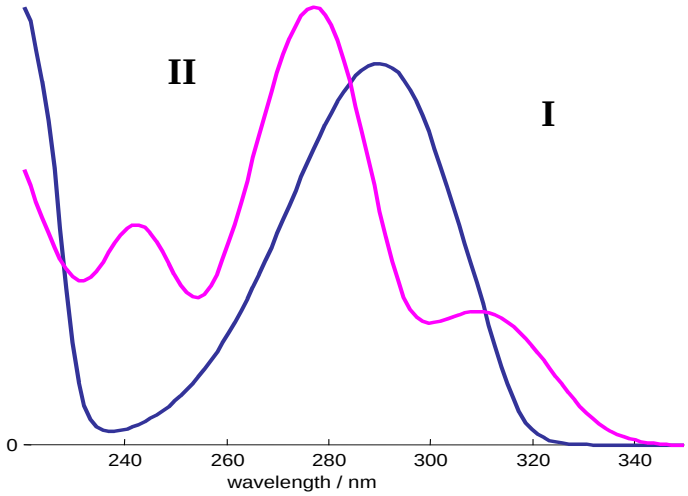


Figure 13 DAD and MS profiles at pH 4.8 (after interpolation, baseline correction and mass selection)

Figure 14 presents the results of deconvolution. The peaks are quite well separated.


II
I
220
230
240
250
260
270
280
290
300
310
time / s


II
I
0
240
260
280
300
320
340
wavelength / nm

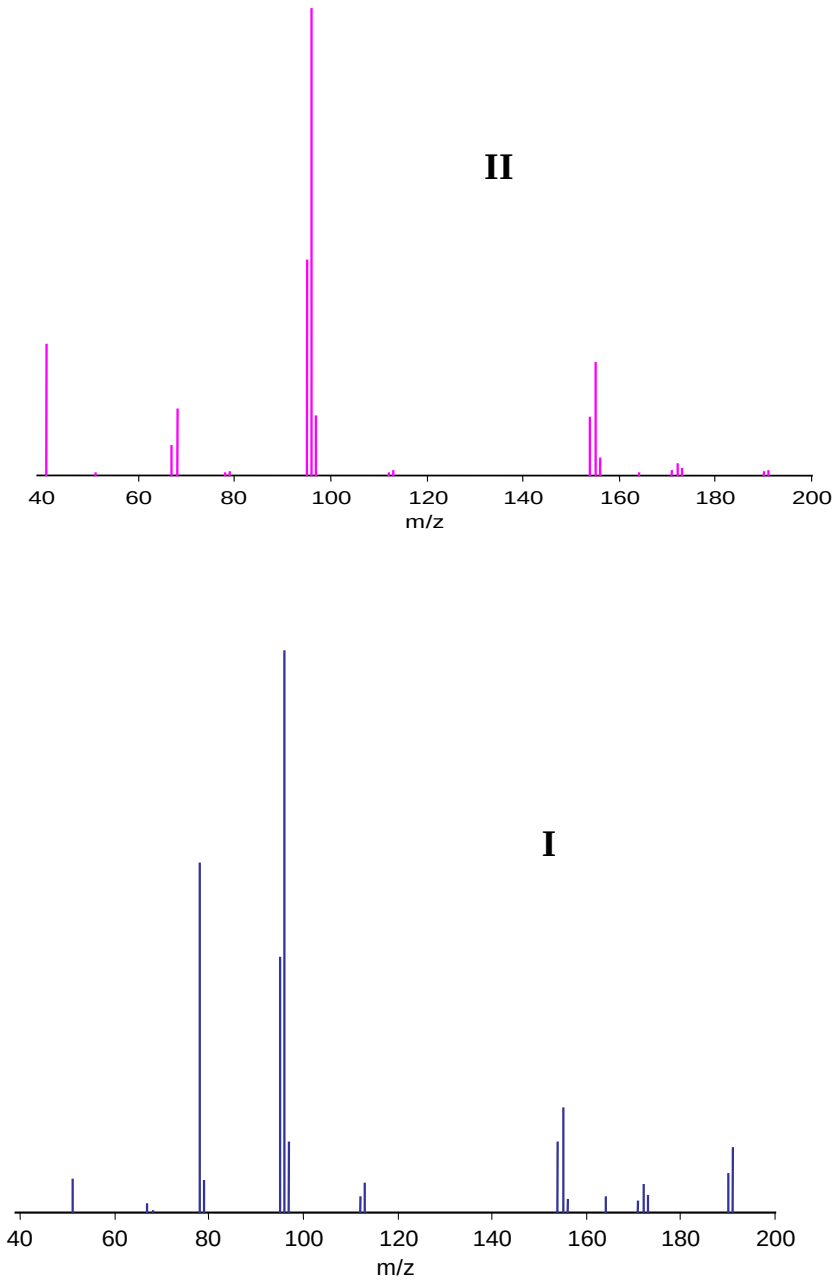


Figure 14 Result of deconvolution at pH 4.8

At pH 5.2 the peaks are partially overlapping again, the profiles being presented in Figure 15.

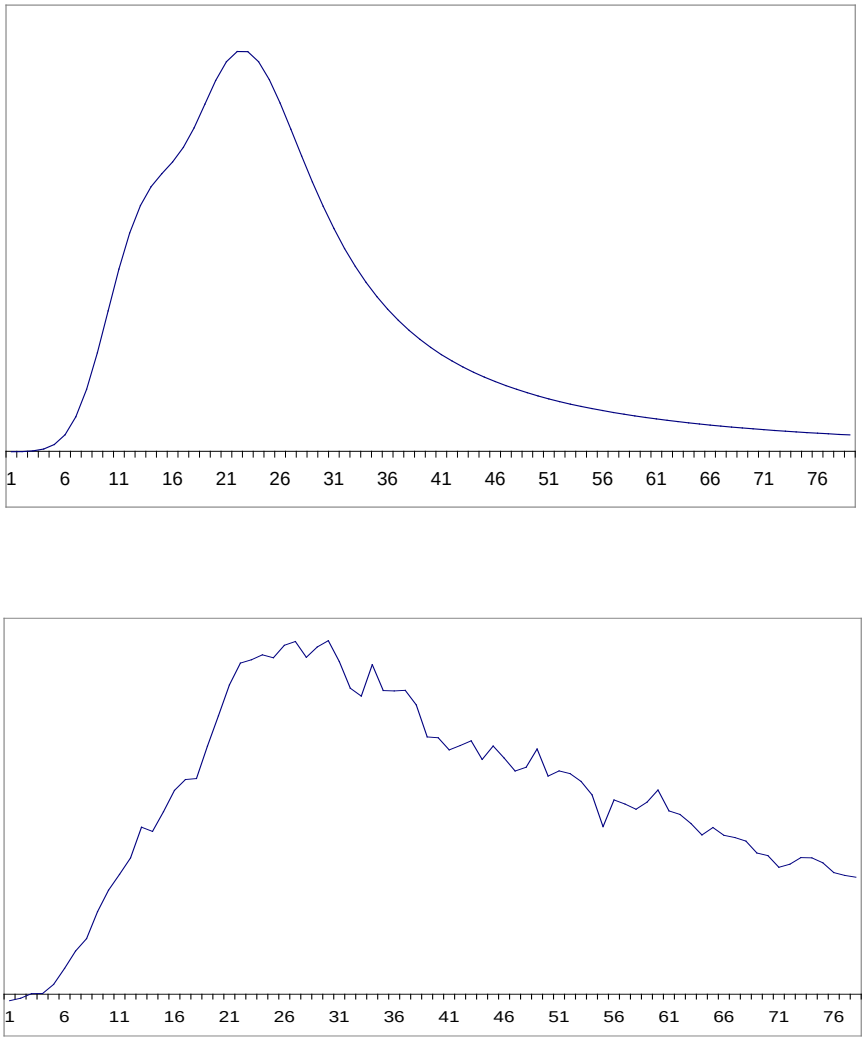
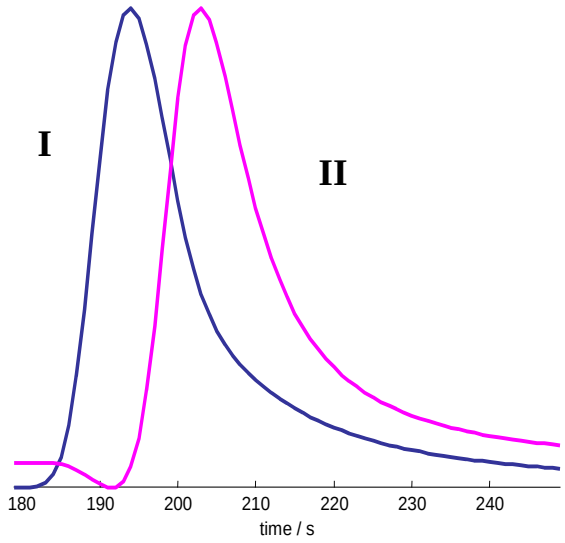
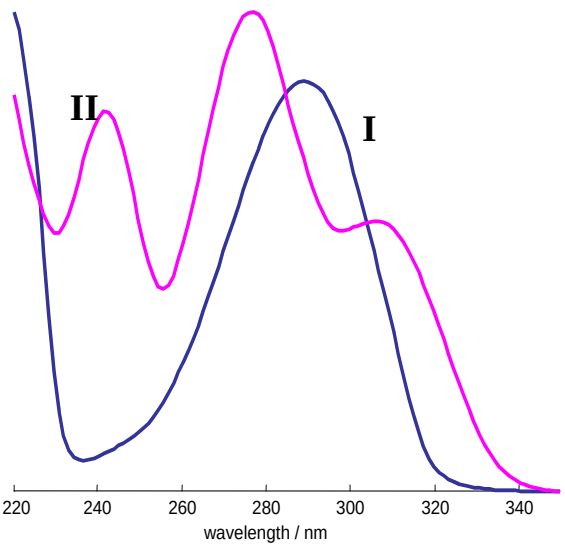
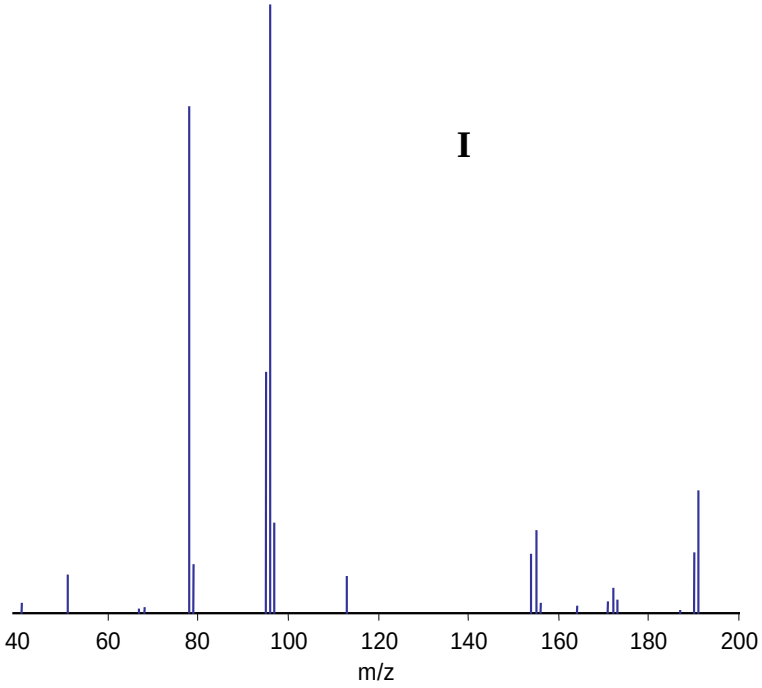


Figure 15 DAD and MS profiles at pH 5.2 (after interpolation, baseline correction and mass selection)

Figure 16 presents the results of deconvolution. Note the relative elution times have changed compared to previous pHs.


I
II
180
190
200
210
220
230
240
time / s


II
I
220
240
260
280
300
320
340
wavelength / nm



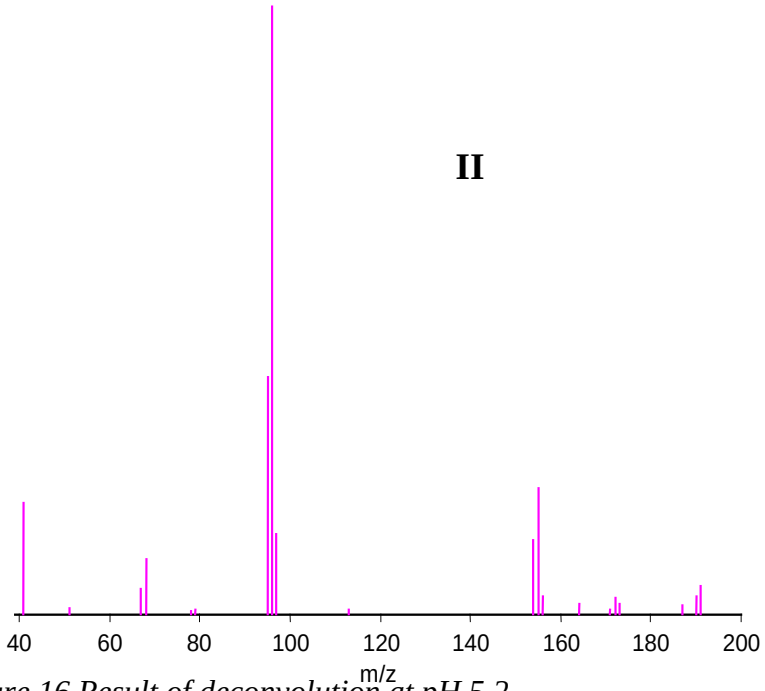


Figure 16 Result of deconvolution at pH 5.2

Notice that the data at pH 5.2 is not perfectly deconvoluted. There is a small negative bump in the chromatographic profile of **I**. This is reflected in the mass spectra with a small amount of mixing, for example, there is a remaining intensity of the m/z 41 ion in **I** at this pH. However, most of the features are well presented, and the change in uv spectra with pH is compatible with the pure compounds.

CONCLUSION

This article presents an approach for the integrated analysis of chromatographic data arising from two different detectors. A key to the method is to obtain roughly similar peakshapes in both cases, which is made difficult by the near impossibility, experimentally, of achieving this. Whereas for a straightforward case such as pH 4.8, most methods of deconvolution will work well without much difficulty, it is by no means easy when there is significant overlap, such as at pH 5.0 and 5.2. The method reported depends quite critically on finding two most diagnostic masses and using an asymmetric peakshape model. Once this is done, the deconvolution is fairly straightforward.

This can aid the chromatographer enormously, and removes the need to examine two separate chromatograms often using separate software. It is also important in summarising information from both detectors simultaneously, taking advantage of the best features from the DAD (e.g. better resolution and peak-shapes) and MS (e.g. characteristic ions). In cases where the interest may be a minor component or a complex coeluting cluster with embedded peaks it is by no means straightforward to eyeball two separate chromatograms. It is important to recognise that each detector has different features, and it is not sufficient to produce a single method that is universal in all cases. With the increased availability of coupled spectroscopy, often using multiple detectors, a good example being DAD-LC-NMR-MS, research into the acquisition and handling of complex chromatograms will be of increasing importance.

ACKNOWLEDGEMENTS

We thank Dr S Dunkerley for providing the data, Dr J Crosby for advice on the electrospray mass spectrometer, Dr R Escott of SmithKline Beecham for suggesting the compounds and the EPSRC for finance.

REFERENCES

[1] Dunkerley, S., Crosby, J., Brereton, R.G., Zissis, K.D., Escott, R.E.A, *Analyst*, 1998, **123**, 2021.
[2] Dunkerley, S., Brereton, R.G., Crosby, J., *Chemom. Intell. Lab. Syst*, 1999, **48**, 99.
[3] Zissis, K.D., Dunkerley, S., Brereton, R.G., *Analyst*, 1999, **124**, 971.
[4] Bessant C., Brereton R.G., Dunkerley S., *Analyst*, 1999, **124**, 1733.
[5] Demir C., Hindmarch P., Brereton R.G., *Analyst*, 2000, **125**, 287.
[6] Lynn P.A.., Fuerst W., *Introductory Signal Processing with Computer Applications*, Wiley, Chichester, 1989.
[7] Brereton, R.G., *Analyst*, 1995, **120**, 2313.
[8] Wold S., Esbensen K., Geladi P., *Chemom. Intell. Lab. Syst*., 1987, **2**, 37.
[9] Mardia K.V., Kent J.T, Bibby J., *Multivariate Analysis*, Academic Press, London, 1979.
[10]Brereton R.G. (editor) *Pattern Recognition in Chemometrics illustrated by case studies*, Elsevier, Amsterdam, 1992.
[11]Felinger A., *Data Analysis and Signal Processing in Chromatography*, Elsevier, Amsterdam, 1998.