

A fully automated chromatographic peak detection and treatment software for multi-user multi-task computers*

PHILIPPE J.P. CARDOT,† PATRICE TROLLIARD and SALIFOU TEMBELY

Laboratoire de Chimie Analytique et d'Electrochimie Organique, Faculté de Pharmacie, Rue Jean Baptiste Clement, F-92296 Chatenay Malabry Cedex, France

Abstract: A fully automated method is presented for the detection and integration of chromatographic peak software in which the only parameter to define is the data acquisition frequency; this has to be chosen high enough to minimize some Nyquist frequency problems. Programmed in C language, the software is used on a Unix like system (Xenix) to emphasize its automaticity. The performance of the system in liquid chromatography has been studied in respect of noise frequency and intensity. It is shown that the classical signal-to-noise ratio is obsolete for definition of the detection limit since frequency analysis allows some signal filtration; these limitations are described. An example is given of simulation with Gaussian peaks convoluted with noise characterized by its intensity and frequency distribution. The validity of the software is demonstrated for some HPLC separations.

Keywords: *Fast fourier transform; moment's theory; peak description and detection; Unix-like computer system; multi-user multi-task; noise frequency analysis and filtration.*

Introduction

In order to evaluate the accuracy of an HPLC assay and to describe the separation, numerous peak detection and integration procedures have been developed. Specialized real-time peak analysis and integration apparatus are now used. One of the most important limitations to overcome is the necessity to adjust some detection and integration parameters. Few commercial apparatus describe the algorithms used. These features limit the use of these methods in research and require time-consuming apparatus adjustment especially in the case of routine analysis.

To enumerate the chromatographic peaks and calculate their parameters, it is necessary to take into account the fact that the characteristics of the recorded output signal will still remain rather unpredictable. In contrast to most of the specialized software designed to handle and treat the chromatographic signal in real time, that presented here has to be used at the end of the chromatographic run. Real-time software needs the input of some calculation parameters; the software presented here analyses the signal and defines the fitted parameters. Finally for the purpose of em-

phasizing the automaticity of the software, it has been programmed in C language on a Unix-like system (Xenix Sco); such an operating system allows multi-user, multi-task operation of this software.

Material and Methods

Hardware, operating system and programming language

The hardware and operating system have been designed to be used in multi-user and in multi-task situations. The central unit is an AT compatible computer M28 Olivetti; a 80286 computer with 3 Mbytes of memory. A 70 Mbytes hard disk and a mathematic coprocessor complete the system with an RS232 extension card (Computone) (4 RS232 connections). Four "intelligent" terminals are connected to the central units; two are designed for the acquisition and handling of the chromatographic data with analogue to digital converter cards. All the terminals are connected to the central unit via RS232 cables. The four terminals are: an Atari 1040 communication software which is a VT 100 terminal emulator (Atari Corp); a Macintosh plus with Term (Century Software) as communi-

* Presented at the "Second International Symposium on Pharmaceutical and Biomedical Analysis", April 1990, York, UK.

† Author to whom correspondence should be addressed.

cation software; an Apple II and an IBM XT are designed to collect the chromatographic data. All the data acquisitions presented in this report have been performed with a hardware and software device previously described [1]. The operating system is the Xenix system V (Santa Cruz Operation Inc.), development system version 2,2,1. The programming language is the C language provided in the Xenix development system.

Description of the hardware-software organization and the principle of the software working

Telecommunication software such as Term (Century Software) allow the terminal to dialogue with the central unit. The data handling sequence is then transferred to the central unit according to the following protocol: after each chromatogram acquisition, the data are automatically transferred under the ASCII code to the central unit according to a RS232 protocol via the communication software. Figure 1 describes schematically the hardware implementation of the system. Three programs comprise the calculation and treatment software of the central unit. The first which runs in "background" as soon as the central unit is in function awaits the arrival of data from the terminals. As crude data series arrive they are stored on the hard disk system and an appropriate name is given. The background program then initiates the second program. The treatment program then treats the data series and provides after-calculation data for a report;

these data are then stored with an appropriate name. The report describes each step of the treatment and can be run automatically or not, to provide a report on a terminal screen or on a printer. This type of automatic and "background" work is made possible by the use of a Unix-like system designed for micro-computers with the simultaneous use of four terminals; each can be connected to at most four chromatographic systems to allow a theoretical simultaneous treatment of 16 chromatograms. It is also possible to assign to each data a priority code. As it is not the purpose of this report to describe the power of Unix systems further information can be obtained from specialized books [2]. This summary leads to the simple concept that a simple "shell procedure" can then be used to automatize all the treatment steps of the software. As the first and third programs can be easily designed and customized, special attention will be given to the treatment program.

Description of the software for chromatographic data treatment

The principal objective of automatized software in chromatography is to collect the information which is usually needed for classical integration software. Such information can be provided by the use of frequency analysis according to FFT transform coupled to a strong basic hypothesis; the high performance of the separation leads to high-speed data acquisition. The chromatographic signal can be schematically described as the "convolution" of a pure chromatographic signal and a signal called the "noise". It is important to note that the software principle is based on a major assumption. It is assumed that the chromatographic signal, even if strongly sharp, has lower frequency components than the noise signal [3]. This assumption allows the construction of the chromatographic signal not versus time but versus frequency according to a Fourier transform. Figure 2 describes schematically the six automatized steps of the program. The data on Fig. 2 represent a simulated chromatogram, resulting of the convolution of two Gaussian peaks and a noise of known intensity with a large frequency distribution, i.e. low and high frequency components [4] (the noise has a continuous spectrum on all the frequency distribution). Figure 2(1) shows the crude acquisition signal which can be described by a frequency analysis of the chromatogram;

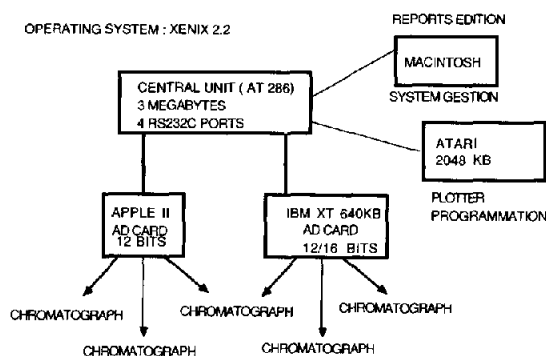
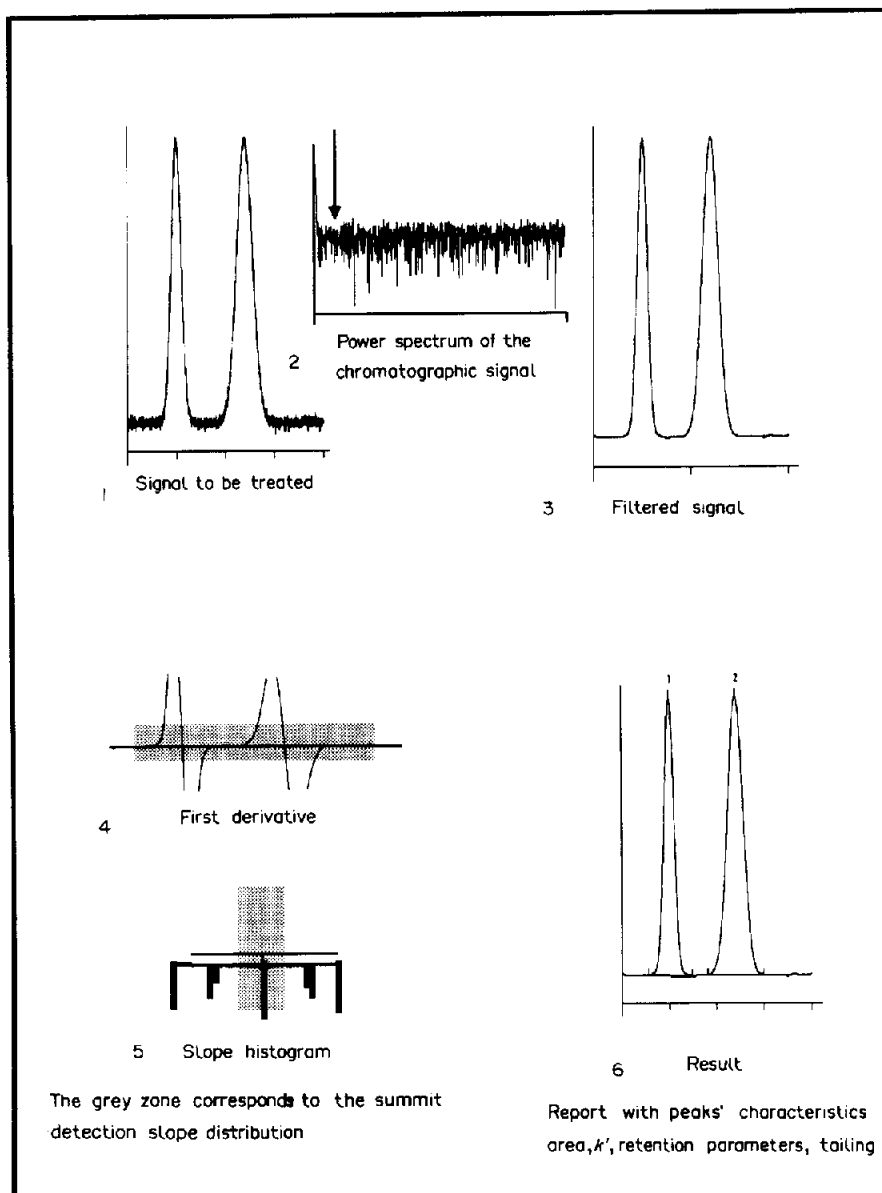


Figure 1

Organization of the multi-user multi-task system. Central Computer: Olivetti M28, 286 CPU (8 MHz), with a mathematic coprocessor (287), 3 Mbytes of central memory, 70 Mbytes of hard disk. Operating system: Xenix 286 version 2,2, in which the four terminals are connected to the central unit via RS232C serials lines and telecommunications software (Term for the Macintosh, for example).

Software Organization

**Figure 2**

Schematic presentation of a chromatographic simulation of the different steps of the software. The simulation is realized on 2048 points, with 2 Gaussian simulations. First peak: summit at point 500, and SD of 50, 1200 and 80 are the respective parameters for the second peak. The noise-to-highest-peak-intensity ratio is 0.01 and a Gaussian distribution of the frequency is provided; this means a continuous and equivalent frequency distribution along the frequency domain of the chromatogram.

thanks to a fast Fourier transform, a simple representation of the frequency distribution of the chromatogram can be done with its power spectrum, as shown in Fig. 2(2). In such a representation all the information linked to the chromatographic signal is at the right of the arrow. The power spectrum of the transform, according to the hypothesis that the chromatogram peaks are those of the lower frequency,

leads to the definition of a low pass frequency filter. This filter is analogous to those described by De Groot [3] and Press [4], also called the optimal wiener filter. The use of such a filter on the transformed data and the inverse Fourier transform calculation gives a filtered signal presented in Fig. 2(3). This last chromatogram can be more easily treated.

In terms of calculation time, the second part

of the software defines the slope criteria of peak detection, according to the method described by Escoffier [5]. The slope period, calculated automatically by the Fourier filter without modifying the peak shape, allows the design of the first derivative diagram as shown in Fig. 2(4). Another way of describing the slope distribution of the first derivative is to construct a histogram of the slope [Fig. 2(5)], in which the number of classes of slope is iteratively optimized to avoid more than 1% of empty classes. This histogram is used to detect the summit of the peak and then to iteratively define the front and the tail of the peak; the final result is presented in Fig. 2(6).

The report describes these six steps, defines the filter characteristics, the number of classes of slope and the characteristics of the peak according to the moment's theory calculations. Peak height, area, plate number, retention times and tailing are systematically obtained.

Results

The software has been experimentally tested on three types of chromatographic signals; the first (Fig. 2) is a simple simulation. The intensity of the noise is defined by the ratio of the height of the noise to the height of the peak. In this case the calculations give reliable results with variation <2% for ratios <0.5.

The same algorithm has been applied to HPLC in numerous cases as described in Figs 3 and 4.

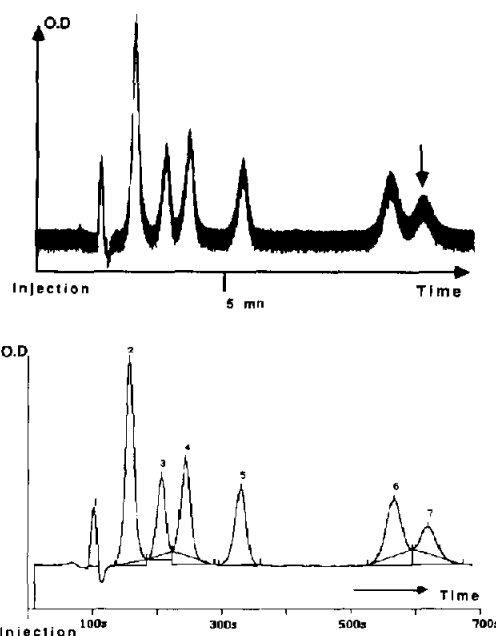


Figure 3
Effect of the software on experimental data in liquid chromatography. Separation of phthalates derivatives by HPLC (water-acetone, 40:60, v/v; flow rate 1.5 ml min⁻¹). The noisy chromatogram corresponds to limit detection conditions; the two most retained peaks cannot be evaluated by the height/noise intensity ratio law, especially the last as pointed out by an arrow on the crude signal. The treated chromatogram shows the efficiency of the filtering; all the peaks can be then evaluated.

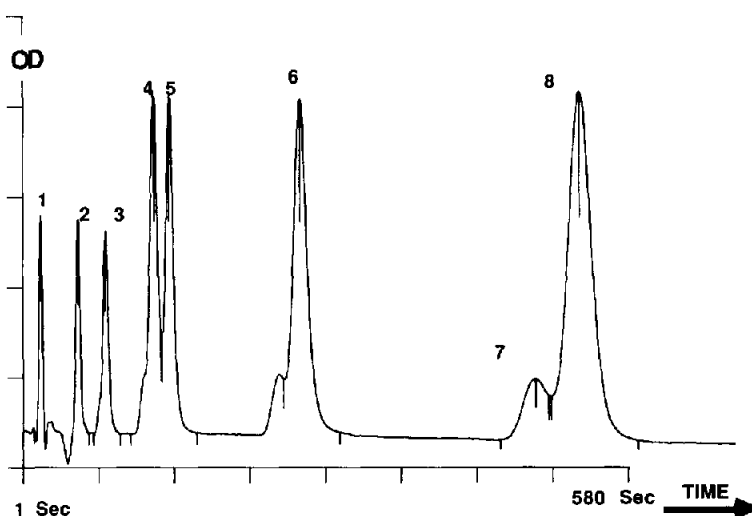


Figure 4
Effect of the software on the liquid reversed-phase separation of sulphonamides. Mobile phase composition and operating conditions: water-acetone (80:20, v/v), buffer pH 3.5, flow rate 1.5 ml min⁻¹. Column: Perkin Elmer C18, 3 μ m (3.3 cm long). Detection: photometric detection at 254 nm. Peak 1, methanol; peak 2, sulphanilamide; peak 3, sulphacetamide; peak 4, sulphadiazine; peak 5, sulphapyridine; peak 6, sulphamerazine; peak 8 sulphamethizole. The peak before peak 6, and the peak 7 are degradation products. The file is of 13180 data collected on 12 bits. Acquisition starts 1 s after injection; the acquisition period is 1 point per 50 ms. Acquisition time is 660 s.

Figure 3 presents a case in which the last peak cannot be estimated according to the classic detection law (ratio of peak intensity to noise intensity <2) but as the noise is an electronic one (conditions of detection limits), it is characterized by a high frequency that is easily removable. Figure 4 presents a more classical condition. The software has been also applied to capillary gas chromatography with some success.

Discussion

This software assumes two important conditions. First, the intensity of the very low frequency noise has to be low and negligible in relation to the intensity of the peak. In the case of very efficient systems which can also be noisy, such as the capillary electrophoresis technique, there can be difficult situations where more sophisticated deconvolution software has to be used. This highlights the problem of the signal frequency acquisition which has to be high enough to avoid modification of filtering shape. In the case of classical liquid chromatography, with 3- μm stationary phases at least 100 points per second are needed. In capillary chromatography, the acquisition frequency has to be higher (500 Hz) to allow a minimum of 80 points per peak. Numerous points per peak are very useful for calculations involving the moment theory; this avoids errors in the measurements of these statistical moments as described by Chesler [6]. To be efficient and reliable this type of software needs high frequency acquisition and the data can be really numerous. The fast fourier transform of numerous data can be dangerous if some calculation precautions are omitted. The use of double precision calculations is imperative and the arithmetic coprocessor decreases the calculation time. If the software is easily defined for HPLC, the treatment of very complex and long chromatograms in gas chromatography can be time-consuming. It is then important to use simulation to define the limits. The use of the power spectrum shows clearly that the classical signal-to-noise ratio is not efficient enough to define the characteristics of the chromatogram noise. Another property of the fourier transform is the stocking of the data; the filtered chromatogram of that presented in Fig. 4 can be then reduced to <20 points.

Conclusions

A chromatographic system is in constant change; the chromatographic parameters vary from day to day and the resulting effects on the chromatogram can have significant effects on the description and the calculations of the shape parameters. These modifications, which can increase, can also lead to incorrect evaluation of the chromatogram characteristics and to quantitative errors if methodological precautions are omitted. This software has been designed to follow the changes in the chromatographic system not only from day to day but at each separation. Noise description and integration conditions are systematically described, allowing an interpretation of the separation quality. The most important limitation of such a concept is that real-time treatment (process chromatography for example) is not available because of limitations in calculations of the fast fourier transform. The advantage is that the only human step is to define the acquisition frequency. In terms of noise, this report emphasizes that there are two noise components. The first is the noise intensity which is commonly used. The second, which is usually avoided, is its frequency distribution. High intensity noise of frequency distribution far different from that of the peak is very easily filtered. This latter effect is important in the case of trace analysis where often the noise is mostly due to electronic effects. Finally, the use of the Unix operating system allows priority in the treatment speed. Precautions have to be taken in the calculation algorithms, especially the FFT, to avoid computer errors and pitfalls.

References

- [1] J.P. Olivo, P. Cardot, I. Ignatiadis and C. Vidal-Madjar, *J. Chromatogr.* **395**, 383–390 (1987).
- [2] B. Kerningham, in *The Unix Programming Environment*. Prentice Hall, Englewood Cliffs, NJ (1984).
- [3] G. De Groot, *TRAC* **4**, 134–137 (1985).
- [4] W.H. Press, in *Numerical Recipes in C* (3rd edn). Cambridge University Press, New York (1989).
- [5] J.L. Escoffier and G. Guiochon, *Chromatographia* **15**, 543–545 (1982).
- [6] S.N. Chesler and S.P. Cram, *Anal. Chem.* **44**, 2240–2245 (1972).

[Received for review 5 April 1990;
revised manuscript received 8 June 1990]