Use of Automatic Digital Data Acquisition and On-Line Computer Analysis in High-Resolution Liquid Chromatography

D. D. Chilcote and J. E. Mrochek

A computer program for the acquisition and on-line analysis of data from a high-resolution liquid chromatograph is described. Emphasis is placed on the means of designing the computer program to overcome certain problems encountered in analyzing the data, mainly as a result of noise. The areas discussed include: detection of the beginning and end of the peak, location of the peak maximum, calculation of peak area, location of inflection points, and resolution of overlapping peaks. Analytical results obtained from the computer analysis of reference compounds separated by high-resolution liquid chromatography are presented. These results are compared with those obtained by manual quantitation of the chromatogram to illustrate the usefulness of the small, on-line computer as a quantitative tool in liquid chromatography.

Additional Keyphrases  urine analysis • “peak-stripping” • inflections • peak detection • noise • peak area • FOCAL language • UV-Analyzer

The role of the computer in the field of chromatography is becoming more and more pronounced (1, 2). From the computer's early inception in the field as an analyzer of chromatographic data on magnetic or paper tape, the exploitation of its usefulness and flexibility has led to direct coupling with the chromatograph in either a dedicated or time-shared mode. Present programs are designed to produce a reliable, quantitative analysis, handling both single and fused peaks.

Initially, both in liquid chromatography (3-5) and gas chromatography (6), a discontinuous interface existed between the computer and the chromatographic system. A digital voltmeter or voltage-to-frequency converter was coupled to the chromatograph to provide a permanent, computer-compatible output. In both cases, each individual reading was transferred to paper or magnetic tape. The tape was subsequently delivered to the computer for analysis. Soon digital integrators became available which, when coupled to the chromatograph, delivered an output consisting of peak areas and positions rather than individual data points (7, 8). This information, also collected on tape, was transferred to the computer for final computation.

Since the integrator is based on hardwire analog logic performed in real time (1), this second method, while relieving the computer of much of the data handling and analysis, introduces a definite rigidity into the manipulations of the system. Therefore the flexibility necessary when dealing with unusually complex analyses is no longer available.

More recently, on-line computation has appeared (9-14) in which the computer is interfaced directly to the chromatograph. This allows the chromatogram to be analyzed as it is produced, obviating the need to manipulate data stored on tape. Either a small, dedicated computer or a larger, time-shared computer can be used in this manner. In all cases in which a digital integrator is not employed, the task of determining peak areas is relegated to the computer. Depending on the complexity of the chromatogram to be analyzed, programming the computer to accomplish this endeavor can become significantly involved. In any case certain guidelines, comprising the basic scheme of the program, will be followed. This basic scheme of the program consists of: (a) recognizing the beginning and end of a peak, (b) locating the peak maximum, (c) locating the inflection points,
(d) calculating the peak area, (e) resolving overlapping peaks, and (f) maintaining an accurate baseline.

In the computer analysis of chromatographic data, two major difficulties are encountered: successful resolution of overlapping peaks, and accurate determination of the areas of small, relatively broad peaks. Although much effort has gone into solving the first problem, the second has received very little notice since the paper by Krichevsky et al. in 1965 (3).

Analysis of small chromatographic peaks is seriously hampered by the presence of background noise. These small, random fluctuations superimposed on the actual data present a severe obstacle to the accurate analysis of chromatographic data. This problem is most apparent in the detection of the beginning, end, and position of the maximum value of the peak. Background noise also severely limits accuracy in the determination of the actual inflection point positions (14).

Fortunately, various schemes have been devised to overcome this difficulty. Because of the nature of the analysis of chromatographic data (15), the very powerful techniques of digital averaging, utilized so successfully in magnetic resonance spectroscopy, cannot be used. However, a very useful technique developed for spectra collected on a permanent or semipermanent record and analyzed off-line—which involves varying both the degree of smoothing and the number of smoothing operations—has been quite successful (16). Although this option is not practical for on-line analysis, other methods involving electronic and digital computer filtering do exist that are quite satisfactory in most cases (14, 18).

In addition to a good filtering technique, certain criteria are necessary to enable the computer to accomplish an accurate analysis. The algorithms developed to satisfy these criteria are integral components of parts a, b, c, and f of the basic computational scheme described earlier. The purpose of this paper is to deal with these algorithms, showing both why they are necessary and how well they work.

The Computer Program

In this discussion we will confine ourselves to a description of a computer program for the on-line analysis of chromatographic data by a small computer. This limiting condition is important, and must be recognized as a critical factor in constructing the program. In general, the small computer limits the degree of sophistication of the "peak stripping" routine, since only a small variable field is available. The on-line analysis also limits the degree of smoothing, since only a limited portion of the data can be considered at any one time. In spite of these limitations, the on-line system does possess certain advantages, which have been adequately described (1, 16) in the literature.

Actual analysis of the data is quite straightforward. The initial aim is to compute the area and position of each peak in the chromatogram. This presupposes, in an on-line system, an adequate interface between the computer and the particular system that is generating the data to be analyzed. In our case, a simple interface consisting of a multiplexer, programmable gain amplifier, and A/D converter suffices (11). Once the signal reaches the computer, it is the task of the analysis program to instruct the computer to output the desired information.

Location of the Peak Beginning

The first problem faced in a chromatographic data analysis scheme is the design of an algorithm that will faithfully identify the actual beginning of each chromatographic peak in the chromatogram. Two main schemes have been devised for recognizing the start of a peak: the first involves successively examining the first derivative for a significant increase (3); the second involves examining the second derivative for either a significant change or a succession of negative values (16, 17). We chose to combine the two methods, using the first derivative to search for the possible beginning of a peak, and the rate of change of the first derivative (i.e., the second derivative) to verify the beginning of the peak.

Using the first derivative method, the crux of the problem lies in determining just what the significant increase should be. This must be viewed in the context of two conflicting guidelines: the desire to recognize as small a peak as possible, and the desire to reject all baseline undulations resulting from noise. The goal is to seek the best possible compromise between the two. The solution we have found to be quite satisfactory involves searching for some initial slope greater than a predetermined low limit,

\[ x_{n+1} - x_i > \text{initial limit} \]

The check for peak verification then takes the form

\[ x_{n+2} - x_i > \text{initial limit} + \text{constant} \]

\[ x_{n+n} - x_i > \text{initial limit} + [(n - 1) \cdot \text{constant}] \]

The number of repetitions is adjusted so that small peaks owing to noise, which begin like a true chromatographic peak but quickly fall off, will be rejected. However, small, gently rising peaks, which continue to rise after the number of verifying steps has been exceeded, will be recognized as real peaks and analyzed.

A complex chromatogram consists of peaks of all heights and widths, some very sharp and others very broad. The above peak-find criterion is good
for the broad peaks, but under certain conditions can conceivably miss a very sharp peak. Thus we use the additional peak-find criterion,

\[ x_{t+4} - x_t > \text{initial large limit} \]

\[ x_{t+4} - x_{t+1} > \text{initial large limit} \]

which allows immediate recognition of a sharply rising peak. This test, coupled with the above criteria for low, broad peaks, provides a very effective means of detecting almost every valid peak in the chromatogram, while at the same time rejecting most of the spurious peaks owing to noise.

The reason for carefully developing the peak-find criteria is mainly to ensure proper analysis of small, peaks. The amount of peak area outside the computer-recognized beginning and end of the peak is a greater portion of the total peak area for small, broad peaks than for sharp, narrow peaks. Thus for any small chromatographic peak, it is imperative that the computer-determined beginning and end be as close to the actual beginning and end as possible. Care must be exercised in the construction of the analysis program, or the results from the analysis of relatively small peaks will be worthless.

Recognition of Peak Maximum Position

Once a satisfactory method has been developed that will allow accurate peak recognition, the problem of locating peak maximum arises. This is really only a problem in the case of small, broad peaks; in the case of sharply rising, narrow peaks, the possibility of finding a false maximum is virtually nonexistent. However, this possibility is very real for the smaller peaks, as Table 1 indicates.

<table>
<thead>
<tr>
<th>Peak height</th>
<th>Number of false maxima found, *false maxima per peak</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.00-0.01</td>
<td>0</td>
</tr>
<tr>
<td>0.01-0.03</td>
<td>1</td>
</tr>
<tr>
<td>0.03-0.05</td>
<td>4</td>
</tr>
<tr>
<td>0.50-2.00</td>
<td>14</td>
</tr>
</tbody>
</table>

* Based on 48 peaks from a standard urine chromatogram.

If a maximum intensity-reading for the peak is located and verified during the course of the analysis, the program then begins searching for the end of the peak. If the maximum was only an irregularity in the shallow slope near the top of the peak, the readings will rise, then fall about the pseudomaximum, but soon begin to rise again toward the real maximum value. Once a maximum has been found and verified, the program recognizes a minimum reading, followed by an increasing slope as the end of the present peak and the start of the next. Thus, when a false maximum is located, the analysis stops short of the real maximum, and only a partial peak analysis results. This type of behavior is especially noticeable for low, broad peaks that have relatively flat tops.

In dealing with this problem, the following criteria were developed. If

\[ x_t > x_{t+1} \]

\[ x_t > x_{t+2} \]

\[ \vdots \]

\[ x_t > x_{t+n} \text{ and } x_{t+n-1} > x_{t+n} \]

then the reading x, is recognized as the peak maximum, and the time corresponding to x, is the peak maximum position. In this case, n is the number of repetitions necessary for verification, which in our application is 5. These criteria ensure that not only is there an absolute drop in peak intensity from the value flagged as the peak maximum for the n repetitions, but also that the slope of the peak is still negative for the last two readings tested by this algorithm. We have found this method to work very well in practice.

Location of Inflection Points

Accurate location of inflection points is necessary for proper performance of the peak-stripping routine. As has already been recognized, the second derivative is quite noise-sensitive (14).

The data are first subjected to a seven-point smoothing routine, to damp out the random noise. The second derivative is then calculated by using a five-point smoothing technique to further "purify" the results (18). Once the second derivative is calculated, all that remains is to determine the location where the second derivative changes sign (inflection point).

As is often the case, the second derivative has a tendency to undergo multiple sign changes in the vicinity of the inflection point, e.g., the sign of the second derivative may go from positive to negative, then positive again before finally going negative and remaining so over the crest of the peak. Spurious sign changes also become common close to the baseline, where the signal-to-noise ratio is lowest. In general, the value of the second derivative in those regions where sign changes do occur is quite low, and a second derivative cut-off limit can be successfully employed. This allows the computer to disregard derivatives lower than some predetermined value and thus miss false inflection points that result from noise. This scheme is based on the fact that the value of the second derivative goes from some positive number in the trough between peaks, through zero at the inflection point, to some negative number at the peak apex. Unfortunately, the magnitude of the second derivative in the vicinity of the inflection point decreases as the peak becomes smaller and broader. For a given second derivative cut-off limit, there will be a proportion-
ately larger number of second derivatives below that limit in the neighborhood of the inflection point for the low, broad peaks than for the high, sharp peaks. Thus the computer, in searching for the second derivative sign change, will skip over all second derivatives that are below the limit in the region of the inflection point. When a sign change is finally noted, it may be well beyond the actual inflection point. To illustrate this point, Figure 1 shows a plot of the absolute value of the difference between the computer recognized inflection point and the true inflection point for a cut-off limit of 0.00004. This limit successfully weeded out spurious inflection points in the analysis.

To circumvent the dilemmas of using a cut-off limit which, when large enough to cause only the real inflection points to be found, also allows a significant error in the location of the inflection point, one can introduce a repetitive test that ensures that the sign change remains for a certain number of readings. Such a criterion would be

\[ |DD_{i+1}| + |DD_i| - |DD_{i+2} + DD_i| > 0 \]
\[ |DD_{i+2}| + |DD_i| - |DD_{i+3} + DD_i| > 0 \]
\[ \vdots \]
\[ |DD_{i+n}| + |DD_i| - |DD_{i+n+1} + DD_i| > 0 \]

By varying the number of repetitions and the size of the limit, an optimum compromise between the number of false inflection points detected and the error in the location of the inflection point can be reached (Figure 2). Since the inflection point information is used as a first guess in the stripping routine, a little leeway in the location is permissible, whereas false inflection points completely ruin the analysis. Thus for optimal results, the number of false inflection points found should be weighted more heavily than the error of location.

Location of the End of the Peak

Just as in the case of the peak-find scheme, the method sought for locating the end of the peak is more critical for small, broad peaks. A logical criterion is one that monitors the slope and assigns the end of the peak to that point at which the slope becomes less than some limit. Our experience has indicated that such a scheme often leads to premature recognition of the end of the peak. The test we finally decided upon was

\[ x_i < x_{i+1} \]
\[ x_i < x_{i+2} \]
\[ \vdots \]
\[ x_i < x_{i+n} \]

For our application, setting \( n \) equal to 4 has worked very well for a 20% sampling interval.

The number of repetitions employed in verifying the beginning and end of a peak does present a problem that must be resolved for accurate computer analysis. The problem is most evident for two sharp, adjoining peaks. By the time the required number of repetitions has been performed to verify the end of the first peak, the computer is analyzing data well into the second peak. A few repetitions are necessary to verify the beginning of the next peak. This results in an area determination that is too large for the first peak and too small for the second. The program is therefore designed to locate the true minimum between the peaks and begin accumulating area from that point. Once the end of the first peak is verified, this accumulated area is subtracted from the total area to give the actual peak area. This accumulation of area continues, however, until the beginning of the next peak is detected. This final “extra area” is stored until the next peak analysis is complete, and then added to the calculated area of the second peak to give the true peak area.

A provision is also made for checking the dis-
tance above the baseline at the end of the peak. If this distance exceeds some preset limit, the immediate start of another peak is assumed, and the peak analysis portion of the program is entered directly.

Area Determination

A great deal of work has gone into the problem of deconvoluting overlapping peaks (10, 12, 14, 19–23), and little remains to be said at present. The methods range from very sophisticated schemes used to analyze data off-line on a large computer, through methods adapted to on-line deconvolution with relatively large computers, down to on-line analyses using a small computer with limited storage capacity. The basic premise behind these methods is that the chromatographic peaks can be suitably fit or approximated by some mathematical model, generally a gaussian or modified gaussian curve. A nonlinear, least-squares routine is used to fit the model to the data, using an iterative scheme designed to continually improve the values of the peak parameters of the model until a suitable match is accomplished.

For on-line analysis of single peaks, a straightforward numerical integration provides an efficient and accurate means of determining the peak area. The correctness of the results is in general limited only by the accuracy in the determination of the baseline. Baseline determination for chromatograms that exhibit a reasonably stable baseline presents no problem, but for those chromatograms in which the baseline can rise significantly over the distance of a few peaks widths, such as occurs in many gas-liquid chromatograms, a serious effort must be directed toward accurately approximating the baseline under the chromatographic peaks.

In our particular liquid chromatographic experience, we have found that the baseline determined before the beginning of a peak is a sufficiently good approximation to the baseline under the peak. Thus our program is concerned only with redefining the baseline between peaks. This is accomplished in two ways: First, the new baseline cannot be reset until a certain minimum time has elapsed since the computer-determined end of the last peak; secondly, the difference between the old and new baseline must be within a certain limit. These two conditions prevent the baseline being reset to an abnormally high value owing to the presence of a reasonably flat valley between two slightly overlapping peaks. The program is also provided with the means of unconditionally resetting the baseline one-half hour after a step change in column temperature from ambient to 60°C.

Results and Discussion

The analysis program, written in FOCAL, is currently being used in conjunction with a PDP-8/1 computer to acquire and analyze data from a UV-Analyzer developed at the Oak Ridge National Laboratory (11). Figure 3 shows how well the computer program handles small chromatographic peaks. The peaks are from a typical chromatogram resulting from the liquid chromatographic analysis of a physiologic body fluid sample. The sampling rate was three readings per minute and the photometer path length was 0.236 cm. The computer determined peak beginning and end are indicated.

Table 2 shows comparisons between the quantitation of reference compounds run on the UV-Analyzer, based both on manual quantitation of the strip chart record and on the computer analysis of the data. The two methods are seen to compare very favorably.

Table 3 shows comparisons of the analysis of uric acid in numerous samples. These samples were analyzed both with a Technicon AutoAnalyzer and with the UV-Analyzer. Again the results with the AutoAnalyzer compare very favorably with the results from the computer analysis of the chromatographic data.

The UV-Analyzer is equipped with a glass volumetric siphon device, which allows the computer to both monitor the accumulated volume of eluate and calculate the flow rate during the analysis. Thus an elution volume can be calculated for each peak, a much more meaningful quantity in liquid chromatography than the elution time, since the flow rate can vary from run to run. Also, quantitation of the chromatographic peaks requires accurate knowledge of the flow rate at the time the peaks are eluted, as well as the peak area.

The results presented here demonstrate that a small computer can be successfully employed in the on-line analysis of data from a liquid chromatographic system. With the large number of small computers now in use (24), this additional capability is not only pleasing but very practical.
Table 2. Comparison of Analytical Results Obtained for Reference Compounds by Using Computer and Manual Methods

<table>
<thead>
<tr>
<th>Compound</th>
<th>Theoretical, µg</th>
<th>Recovery, %</th>
<th>Manual, µg</th>
<th>Recovery, %</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nicotinamide</td>
<td>39.6</td>
<td>40.4</td>
<td>102.0</td>
<td>39.9</td>
</tr>
<tr>
<td>N-oxide</td>
<td>42.8</td>
<td>44.4</td>
<td>103.7</td>
<td>45.4</td>
</tr>
<tr>
<td>Uricil</td>
<td>5-acetylamino-6-amino-3-methyluracil</td>
<td>30.1</td>
<td>29.0</td>
<td>96.3</td>
</tr>
<tr>
<td>7-Methylxanthine</td>
<td>8.2</td>
<td>7.3</td>
<td>89.0</td>
<td>8.0</td>
</tr>
<tr>
<td>Xanthine</td>
<td>24.2</td>
<td>23.8</td>
<td>98.3</td>
<td>24.8</td>
</tr>
<tr>
<td>Theophylline</td>
<td>160.0</td>
<td>173.8</td>
<td>108.6</td>
<td>169.5</td>
</tr>
<tr>
<td>Xanthosine</td>
<td>57.6</td>
<td>52.5</td>
<td>91.1</td>
<td>51.7</td>
</tr>
<tr>
<td>Orotidine</td>
<td>82.0</td>
<td>79.8</td>
<td>97.3</td>
<td>86.1</td>
</tr>
<tr>
<td>Hippuric acid</td>
<td>117.6</td>
<td>130.5</td>
<td>110.9</td>
<td>119.4</td>
</tr>
<tr>
<td>Nicotinamide</td>
<td>51.2</td>
<td>50.0</td>
<td>97.7</td>
<td>46.8</td>
</tr>
<tr>
<td>Hypoxanthine</td>
<td>19.2</td>
<td>19.3</td>
<td>100.5</td>
<td>17.1</td>
</tr>
<tr>
<td>Orotic acid</td>
<td>84.6</td>
<td>91.7</td>
<td>108.3</td>
<td>88.8</td>
</tr>
<tr>
<td>Trigonelline</td>
<td>97.0</td>
<td>102.4</td>
<td>105.7</td>
<td>107.5</td>
</tr>
<tr>
<td>Uric acid</td>
<td>67.0</td>
<td>66.8</td>
<td>99.7</td>
<td>64.5</td>
</tr>
</tbody>
</table>

where the peak height is the maximum absorbance; the peak width is the width at one-half the peak height, in h; the flow rate is expressed in ml/h; and the path length is expressed in cm.

We thank J. C. Mailen for carrying out the Technicon analyses.

References


Table 3. Uric Acid Analysis of 11 Samples: AutoAnalyzer vs. the ORNL UV-Analyzer

<table>
<thead>
<tr>
<th>Method</th>
<th>µg/ml</th>
<th>UV-Analyzer</th>
</tr>
</thead>
<tbody>
<tr>
<td>Technicon</td>
<td>43.4</td>
<td>44.6</td>
</tr>
<tr>
<td></td>
<td>44.2</td>
<td>40.8</td>
</tr>
<tr>
<td></td>
<td>12.5</td>
<td>12.8</td>
</tr>
<tr>
<td></td>
<td>40.6</td>
<td>39.8</td>
</tr>
<tr>
<td></td>
<td>35.5</td>
<td>32.6</td>
</tr>
<tr>
<td></td>
<td>64.1</td>
<td>63.9</td>
</tr>
<tr>
<td></td>
<td>24.0</td>
<td>23.6</td>
</tr>
<tr>
<td></td>
<td>44.8</td>
<td>44.6</td>
</tr>
<tr>
<td></td>
<td>40.0</td>
<td>34.8</td>
</tr>
<tr>
<td></td>
<td>41.4</td>
<td>40.0</td>
</tr>
<tr>
<td></td>
<td>50.8</td>
<td>45.9</td>
</tr>
</tbody>
</table>

*Phosphotungstic acid method.