Journal of Chromatography, 150 (1978) 367–380

CHROM. 10,488

COMPUTER ANALYSIS OF AMINO ACID CHROMATOGRAMS AND OTHER LINEAR ELUTION SPECTRA

L. KAMPMANN
Max-Planck-Institut für Biophysik, Heinrich-Hoffmann-Str. 7, D6000 Frankfurt (Main) 71 (G.F.R.)
(Received July 11th, 1977)

SUMMARY

A computer-aided method of evaluation of amino acid column chromatograms is proposed and compared with the usual synchronous peak-integration method. Spectra and their backgrounds are digitized separately by a curve digitizer; this allows a better estimate of the background time courses. A computer program then fits a set of superimposed Gaussian distributions to each corrected spectrum, thus circumventing the problems arising from incompletely separated peaks. Samples of known composition, run intermittently through the analyzer between the unknown spectra, allow a determination of the time dependence of the ratios of "peak areas over amounts of amino acids"; hence, the amounts of amino acids can automatically be corrected for aging effects of the analyzing system. A modified version of the computer program allows resolution of any spectrum to a sum of Gaussians, with least squares fitting of their amplitudes, widths and locations.

INTRODUCTION

The output of an amino acid analyzer is usually voltage as a function of time and shows a sequence of peaks, each peak corresponding to a different amino acid of the sample. Providing the number of sites of interaction between the column material and the migrating amino acid molecules is sufficiently large, the peak area of an amino acid is proportional to its total amount. Therefore, running a single sample of known quantitative composition should enable the chromatograms to be calibrated for the desired amounts of amino acids. Usually, the individual peaks of each chromatogram are integrated during their generation on the analyzer, and the ratios of the corresponding peak areas in the unknown and known spectra are related to the ratios of the amounts of amino acids.

This procedure has the advantage of being very fast and straightforward. However, an ideal amino acid analyzer has been assumed and in practice the following problems may arise:

(1) Adjacent peaks in a chromatogram are not completely separated.
(2) Impurities create shoulders on amino acid peaks, or generate their own

peaks. On the other hand, relevant peaks may be below the integrator's threshold, and thus not be recorded.

(3) Baseline jumps occur between two peaks or even during generation of a peak.

(4) The column bed and detection system age between the analysis of unknown samples and calibration.

(5) For an automatic evaluation procedure, discrimination between small, but relevant amino acid peaks and the background is difficult.

In this paper a method of evaluating amino acid chromatograms is described which uses a digital computer and a curve digitizer. Compared to the classical method of direct signal integration during peak generation, the results are more reliable and more accurate.

METHODS

All of the spectra generated by the analyzer are recorded as strip charts and later stored as computer data files, after they have been digitized on a digitizer tablet. A standard sample of known composition is run through the analyzer prior to the first unknown sample, after several unknown samples and again as the final spectrum of the whole sequence. All of the standards have the same composition, and their spectra are fed into the computer in the same order as they appeared on the analyzer; the known spectra then follow, also maintaining their generation sequence. If an amino acid peak appears in the n standards at times $t1$, $t2$,... tn and in an unknown spectrum at time t, where $t1 < t < tn$, the "current standard peak" at time t can approximately be found by interpolation from the values supplied by the standards, assuming a smooth time course of the aging processes in the interval $< t1, tn >$, and a constant time step between the appearance of the same amino acid in consecutive spectra.

When digitizing a spectrum, its baseline is traced first and then represented in the computer data file by up to 70 points. Digitization and file storage of all of the relevant peak maxima in sequential order is the next task*, and tracing the spectrum itself, with a yield of up to 500 digital coordinate pairs, completes the computer-stored information for an individual spectrum.

The computer processing is then started on-line or off-line on the set of spectra. After a point-by-point subtraction of the background (by interpolation within the baseline to find the local background level), each spectrum is deconvoluted, *i.e.*, a least squares fit to a hypothetical spectrum $h(t)$

$$h(t) = \sum_{j=1}^{m} aj \cdot \exp(-[(t - pj)/wj]^2) \tag{1}$$

is made to each measured spectrum $s(t)$, where the jth maximum of the m relevant peak maxima is reached at $t = pj$ with an amplitude of aj, a width of wj and an area

* Alternatively, the evaluation program can scan the spectrum to find the peak maxima.

of $\sqrt{\pi} \cdot aj \cdot wj$. This least squares condition, a set of parameters $aj, pj, wj\,(j = 1, 2, \ldots m)$ from*

$$\int [h(t) - s(t)]^2 \cdot dt = \text{minimum} \qquad (2)$$

i.e.,

$$\int [h(t) - s(t)] \cdot \exp(-[(t-pj)/wj]^2) \cdot dt = 0,$$
$$\int [h(t) - s(t)] \cdot aj \cdot (t-pj)/wj^2 \cdot \exp(-[(t-pj)/wj]^2) \cdot dt = 0,$$
$$\int [h(t) - s(t)] \cdot aj \cdot (t-pj)^2/wj^3 \cdot \exp(-[(t-pj)/wj]^2) \cdot dt = 0,$$
$$\text{where } j = 1, 2, \ldots m, \qquad (3)$$

is meaningful if each (background-corrected) amino acid spectrum can be described as a sum of independent peaks, each peak shape representing a Gaussian distribution. A modified gradient method[1,2] is used to fit the theoretical spectrum $h(t)$ to the experimental spectrum $s(t)$, *i.e.*, to solve the $3m$ non-linear equations (3). The digitized peak maxima serve as initial values for pj and aj, whereas the widths $w1, w2, \ldots wm$, used as initial estimates, are all given the same value. Alternatively, the fitted spectrum $h(t)$ can be plotted and overlaid by a graph of the background-corrected measured spectrum $s(t)$.

RESULTS

Examples and error considerations

Fig. 1a displays a strip-chart record of an amino acid spectrum, the output from a Beckman Multichrom amino acid analyzer. Light absorption of ninhydrin-stained samples was measured at 570 nm (large amplitudes) and also at 440 nm. Fig. 1b shows a baseline-corrected plot of the digitized 570-nm absorption curve from Fig. 1a, but with proline replaced by its 440-nm profile. This spectrum is overlaid by the synthetic spectrum, *i.e.*, by a computer-fitted sum of Gaussian distributions[1]. The baseline assumed in Fig. 1a is drawn at the bottom of Fig. 1b with the ratio (baseline amplitude)/(spectrum amplitude) increased by a factor of 2.5 compared with the original in Fig. 1a. As an artefact of the interpolation process, a small overshoot occurs in the baseline each time the buffer is changed in the analyzer.

Repeated digitization and recalculation of the synthetic spectrum yields an error of less than 0.5% for the sum of the fitted peak areas (ammonia is omitted). Yet errors of up to 3% for a single peak area occur for peaks with areas even greater than that of proline in Fig. 1.

As an example of many sets of analyzed spectra, eight spectra were measured, including a leading and a trailing standard, then digitized and computer analyzed as described above. Fig. 2 is an extract of the computer print out. In the standards (spectra 1 and 2), each peak contains 80 μM of amino acid per definition (where this value is meaningless for ammonia). These numbers are used to determine the amount of amino acid in all of the following 6 spectra. The peak locations, widths and amplitudes are the result of the fitting process. Whereas the locations provide only the sequence within a spectrum, the peak areas, *i.e.*, the products of widths and

* The positions of the peak maxima have also to be fitted, since the quality of the fit depends critically upon their exact values, which are not reproducible from one spectrum to the next. All of the integrals are to be extended over a single complete spectrum.

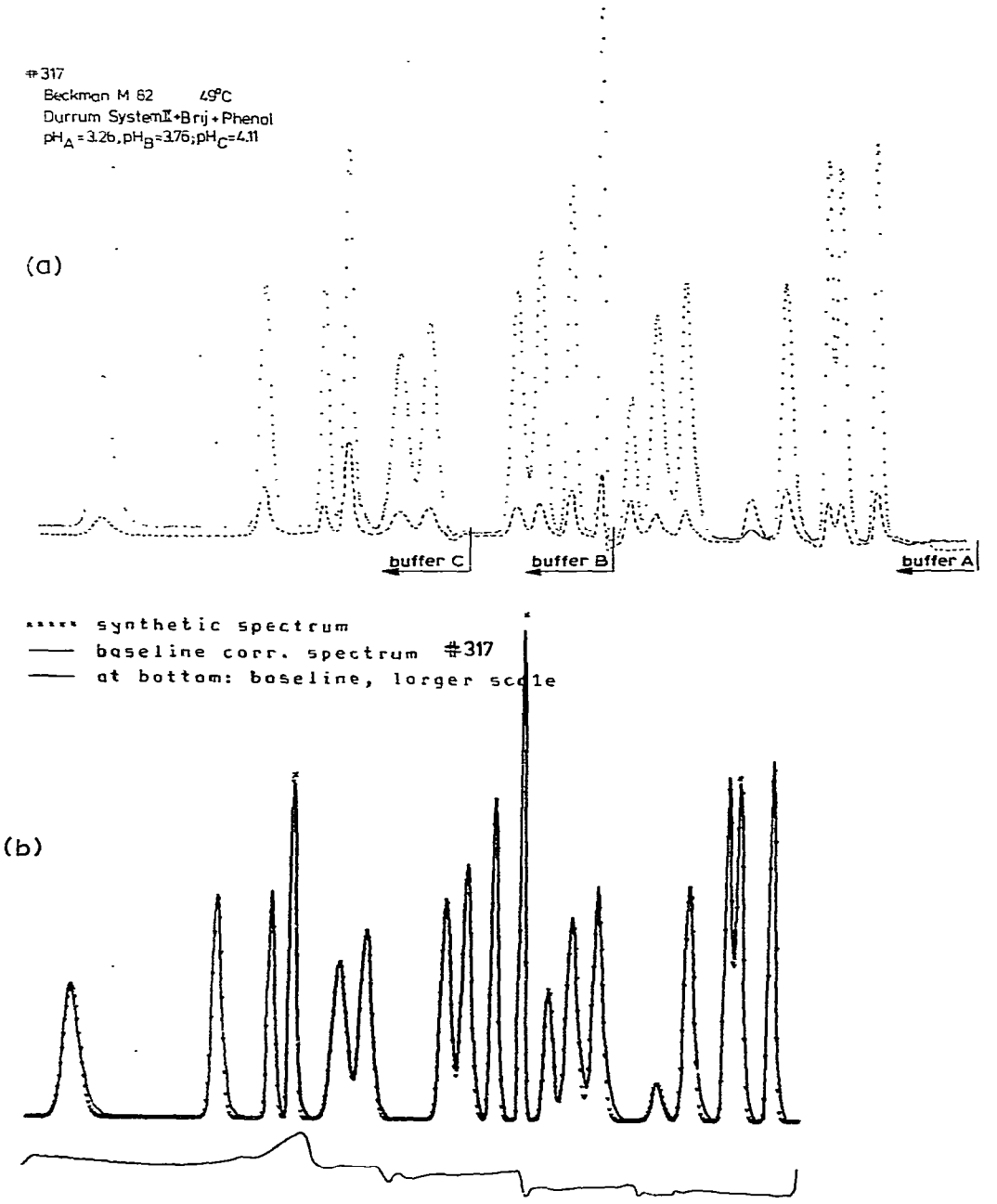


Fig. 1. An amino acid column chromatogram, consisting of 17 amino acids and ammonia. (a) A copy of the original strip-chart record, *i.e.*, the output from a Beckman Multichrom analyzer; light absorption at 570 nm and 440 nm (small amplitudes, 440 nm). (b) The baseline-corrected spectrum overlaid by the fitted spectrum; the estimated baseline, boosted by a factor of 2.5, is shown at the bottom of the figure.

```
MINIMUM FUNCTION VALUE = 6.616757E 05
 1. SPECTRUM STANDARD                                                    02-MAY-77
BECKMAN M82; 49DEG.; DURRUM II + BRIJ + PHENOL; PH: 3.26,3.76,4.11      #317
      % OF TOTAL  µM   LOC.  WIDTH AMPL.        % OF TOTAL  µM   LOC.  WIDTH AMPL.
   ARG   5.88   80.00   83.  20.63 251.9   HIS   5.88   80.00  342.  11.62 429.0
   NH4  00.00   80.00  439.   8.27 435.8   LYS   5.88   80.00  477.   7.57 686.5
   PHE   5.88   80.00  561.  17.68 303.5   TYR   5.88   80.00  608.  14.48 366.8
   LEU   5.88   80.00  750.  12.35 430.7   ILE   5.88   80.00  787.  11.77 480.8
   MET   5.88   80.00  837.   8.02 625.1   VAL   5.88   80.00  887.   5.32 989.6
   CYS   5.88   80.00  935.  10.71 255.0   ALA   5.88   80.00  976.  13.81 385.5
   GLY   5.88   80.00 1022.  13.26 414.3   PRO   5.88   80.00 1130.  12.91  72.1
   GLU   5.88   80.00 1186.  11.51 456.3   SER   5.88   80.00 1255.   7.69 543.4
   THR   5.88   80.00 1275.   8.53 673.9   ASP   5.88   80.00 1334.   7.67 697.4

 2. SPECTRUM STANDARD                                                    02-MAY-77
      . . . . .

 3. SPECTRUM                                                             02-MAY-77
      . . . . .
      . . . . .

MINIMUM FUNCTION VALUE = 2.431602E 05
 6. SPECTRUM                                                             03-MAY-77
BECKMAN M82; 49DEG.; DURRUM II + BRIJ + PHENOL; PH: 3.26,3.76,4.11      #322
      % OF TOTAL  µM   LOC.  WIDTH AMPL.        % OF TOTAL  µM   LOC.  WIDTH AMPL.
   ARG   5.94   81.09   70.  21.38 239.0   HIS   5.73   78.22  326.  12.15 387.7
   NH4  00.00   82.99  422.   8.40 479.5   LYS   6.30   85.93  461.   8.33 647.9
   PHE   5.88   80.26  546.  18.17 281.1   TYR   5.83   79.58  593.  14.94 350.8
   LEU   5.75   78.44  734.  12.62 405.6   ILE   5.52   75.39  772.  11.86 458.7
   MET   5.69   77.72  822.   8.35 593.1   VAL   6.14   83.75  871.   5.87 933.7
   CYS   5.64   77.01  920.  10.58 228.9   ALA   6.20   84.55  963.  15.92 343.9
   GLY   5.99   81.74 1010.  14.90 374.6   PRO   5.75   78.45 1115.  13.77  66.1
   GLU   5.84   79.67 1171.  13.05 400.9   SER   5.69   77.66 1239.   7.74 596.3
   THR   5.88   80.22 1260.   9.75 605.9   ASP   6.24   85.15 1316.   8.21 667.6

MINIMUM FUNCTION VALUE = 1.897129E 05
 7. SPECTRUM                                                             03-MAY-77
BECKMAN M82; 49DEG.; DURRUM II + BRIJ + PHENOL; PH: 3.26,3.76,4.11      #323
      % OF TOTAL  µM   LOC.  WIDTH AMPL.        % OF TOTAL  µM   LOC.  WIDTH AMPL.
   ARG   5.95   39.65   91.  21.53 115.2   HIS   5.68   37.88  347.  12.82 188.1
   NH4  00.00   40.88  443.   8.56 235.9   LYS   5.59   37.27  481.   8.26 281.1
   PHE   5.84   38.93  567.  17.47 139.9   TYR   5.84   38.92  615.  15.33 166.8
   LEU   5.84   38.92  756.  12.64 200.0   ILE   6.26   41.70  793.  11.44 242.2
   MET   5.94   39.56  843.   8.77 288.7   VAL   6.08   40.52  892.   5.95 444.9
   CYS   5.63   37.53  942.  11.49 100.6   ALA   6.04   40.24  988.  15.93 162.4
   GLY   5.70   38.01 1034.  13.71 189.0   PRO   5.55   36.98 1139.  13.77  31.1
   GLU   6.07   40.48 1197.  13.57 195.9   SER   5.80   38.66 1263.   8.37 271.7
   THR   6.17   41.13 1285.  11.10 274.7   ASP   6.02   40.10 1344.   8.46 302.0

 8. SPECTRUM                                                             03-MAY-77
      . . . . .
```

Fig. 2. Extract of the computer print-out. Minimum function value = the residual sum of the squares of the deviations, i.e., a measure of the accuracy of the fit; the text above each spectrum specifies the type of resin used in the column, the column temperature and the buffer system.

amplitudes, determine the amounts of amino acids. The first spectrum of Fig. 2 is shown in Fig. 1.

Fig. 3 represents the sixth spectrum together with its fit, denoted by YYYYY, and its contents of amino acids given in μM for all of the peaks at the bottom of the graph.

To allow error estimates, all of the 8 spectra were run with standard samples of known composition; spectra 1–6, 80 μM; spectra 7 and 8, 40 μM of each amino acid. In the ideal case, free from errors, spectra 3–6 should show 80 μM, and spectra 7 and 8, 40 μM, for all of the peaks (except ammonia). The peak areas for a specific amino acid should be the same for spectra 1–6, and half this value for spectra 7 and 8.

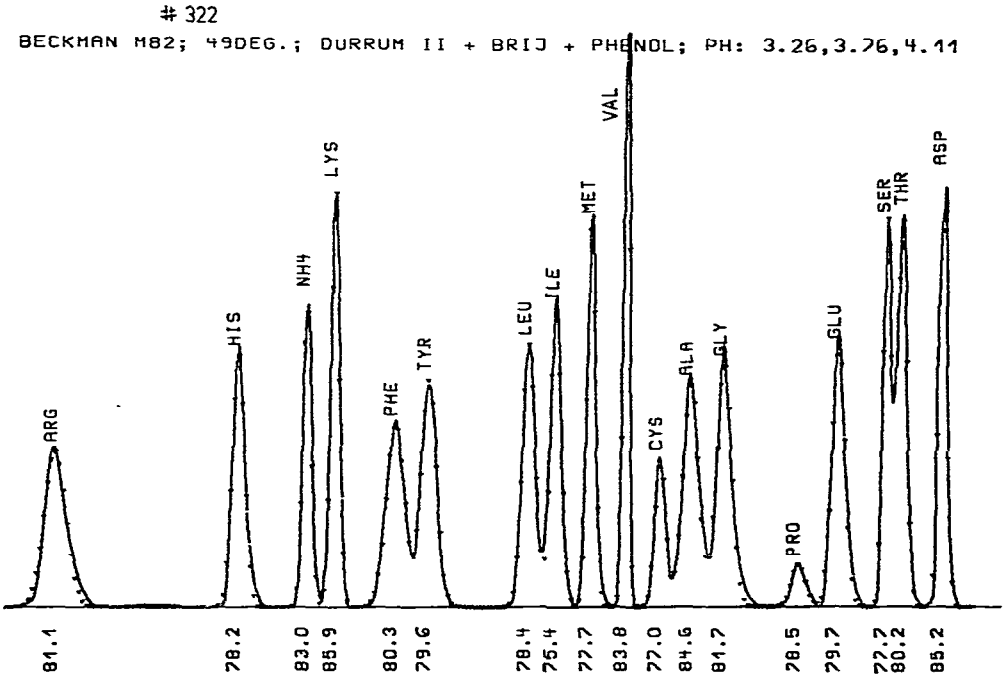


Fig. 3. The sixth spectrum of the discussed set of spectra, overlaid by its theoretical fit. The determined peak contents (in μM) of amino acids are shown at the bottom of the figure.

In Fig. 4a and 4b the peak areas and the amounts of amino acids, respectively, are plotted *vs.* the number of the spectrum; all of the values in spectra 7 and 8 are doubled. Spectrum 9 represents the arithmetic mean of spectra 1–8. The dashed line indicates the average peak areas or amounts of amino acids. Although the amino acid contents of a spectrum can only be determined relative to other spectra, to the standards, it is clear from Fig. 4a that spectra 1 and 3 contain a greater amount of amino acids, and spectrum 2 a smaller amount, than the average. The standard deviation is *ca.* 2%, which is the relative "syringe error" when taking the samples.

If the very poor value for valine in spectrum 3, the extremely high error of which (15%) remains unexplained, is omitted all of the resulting amino acids are quantitated with an accuracy of 10%. In most cases the error is less than 6%, whereas poorly separated peaks, such as the pair serine–threonine, have a greater error. These error ranges were reproduced for several different sets of spectra. The total error for a single amino acid is accumulated from several sources:

(1) "Syringe error" from taking the sample.
(2) Manual digitization error when tracing the strip charts.
(3) Poor baseline tracing.
(4) Column-bed aging and fluctuation of the detection system.
(5) Deviation of peak shape from Gaussian.
(6) Errors of types 1–5 carried into the results from the standard spectra.

Replacing manual digitization by direct storage of the spectral data on a computer-

compatible medium (*e.g.*, a magnetic cassette), or by working on-line, is expected to reduce the error by *ca.* 3%.

If the trailing standard, *i.e.*, spectrum 2, is transferred from its position in Fig. 4a to replace spectrum 9, the abscissa becomes a linear function of time. The

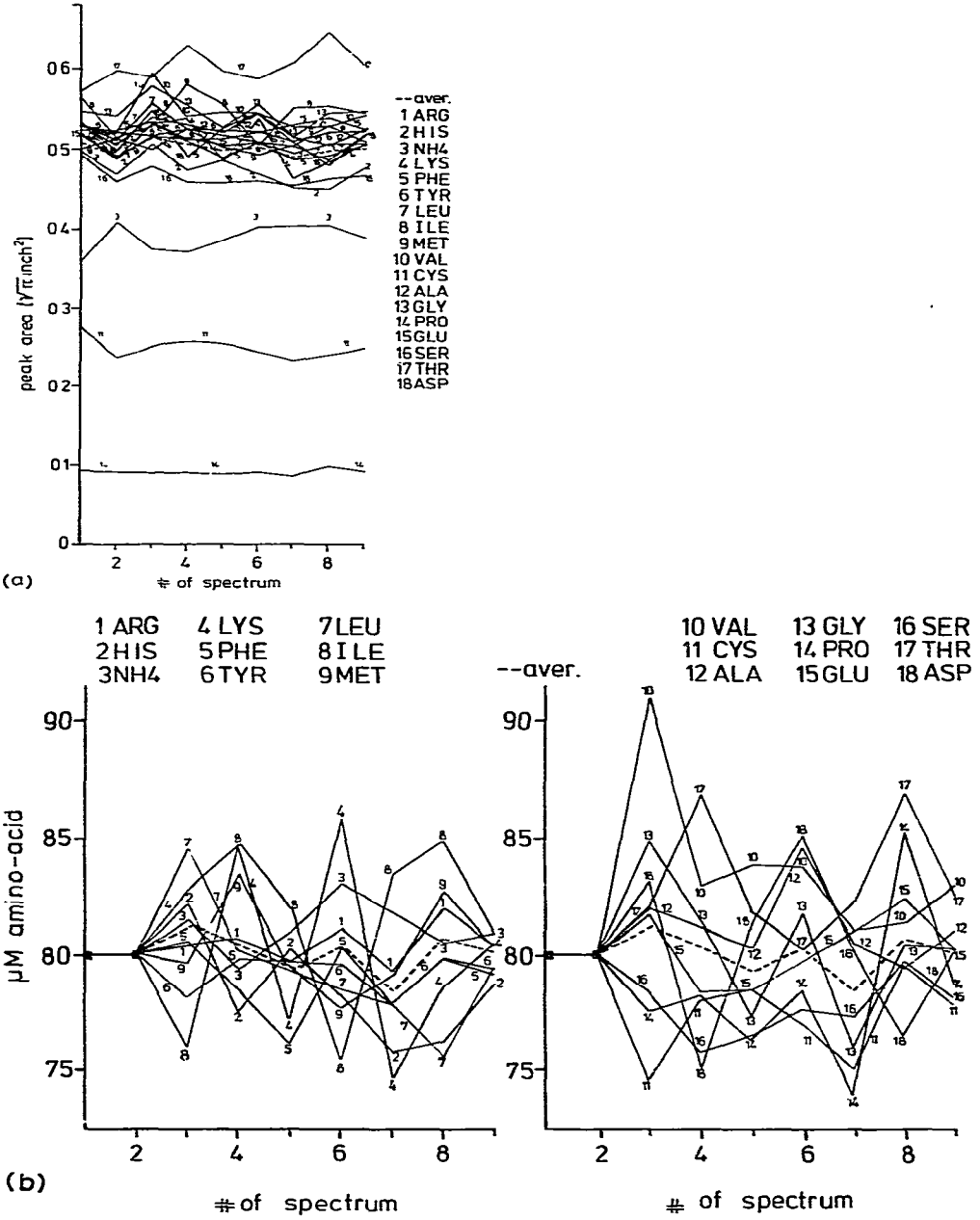


Fig. 4. The peak areas (a) and amino acid contents (b) of each species within a set of 8 spectra. Spectra 7 and 8 are multiplied by a factor of 2; spectrum 9 is the arithmetic mean of spectra 1–8; the dashed lines refer to the average of all of the species except ammonia.

average peak area now reveals a weak tendency to decrease, which is interpreted as an aging process of the column bed. Although the evaluation method under discussion compensates for column-bed aging and for fluctuations of the detection system, as was detailed above, for a single peak these effects are lost among the other errors in the presented example. Such an effect can only be seen in a very long sequence of spectra.

A different set of spectra was analyzed with standardized samples of 40 μM of each amino acid. Fig. 5 shows sections of two spectra with 60 μM of threonine and serine, respectively, added, to test the ability of the method to separate strongly overlapping peaks having different amplitudes. In these cases the error was expected to exceed 10% for the smaller component.

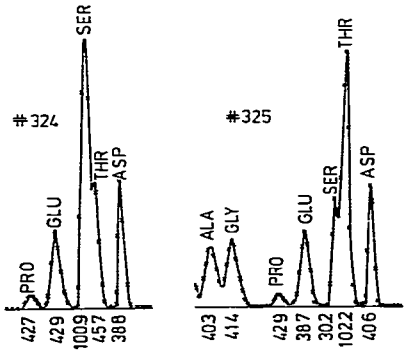


Fig. 5. Overlapped peaks of different sizes: sections of 2 different spectra with 40 μM of amino acid per peak. 60 μM of serine were added to the first sample, and 60 μM of threonine to the second sample.

Deviations from the Gaussian shape are clearly visible, *e.g.*, in Fig. 3, for arginine and glycine: the peaks show slight "fronting" (time increases from right to left in the chromatograms). Yet these distortions yield only a minor contribution to the total error of the single species.

DISCUSSION

Compared to a synchronous peak integration of amino acid column spectra, the evaluation method described above not only involves more complex equipment but is also more time consuming. Strictly speaking, it consumes computer time, but not more man hours than other methods. Nevertheless, the development of the method has to be justified.

Although improvements of the column filling material are designed with the aim of gaining a more complete mutual separation for all of the amino acid peaks, there is still usually an appreciable overlap between threonine and serine, and between other species if their concentrations are high. The separation problem can be solved by using longer columns, but this is not desirable, primarily because the elution time for a single sample increases to many hours, the eluted sample is diluted and also, as a minor problem, the aging processes of the column bed and the detection system become more important.

The integrator solves the separation problem by dividing a double peak by a vertical cut at the central minimum. This procedure is only valid in cases in which both peaks have the same height and width. Cases can be constructed where the error of the peak area determination reaches 100% or more for the smaller peak. By fitting Gaussians to the peaks, their complete separation is possible without additional dubious assumptions, which is a major advantage of this method. Furthermore, shorter elution times are permitted, since complete peak separations are no longer of overriding importance.

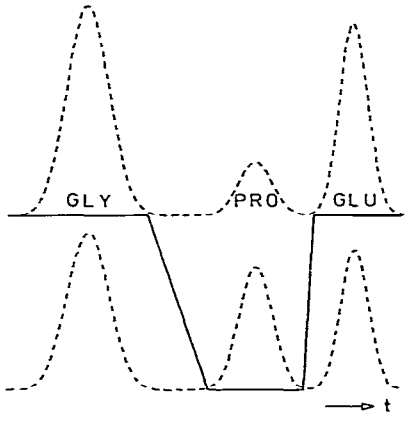
Impurities appear in all of the amino acid spectra, but cause few problems in either of the two methods. If impurity peaks are completely separated from adjacent peaks, they can be treated as part of the baseline during the digitization process. This also holds for impurities which interfere with the amino acid peaks, although an error is introduced, if a peak is "shaved" away during manual digitization. The automatic integration method does not allow corrections to be made for interfering impurities, whereas a completely manual evaluation (*e.g.*, planimetry of peak areas) does, in the same manner as the digitize-and-fit method suggested here.

Omission of a small amino acid peak below the integrator's threshold, above which "base point of a coming peak" is triggered, is a typical problem inherent to the integrator method, as is revealed by repeated attempts to find satisfying criteria for the discrimination between baseline jumps and positive flanks of a peak to be integrated[3–6]. Missing small, but relevant peaks, on the other hand, or mixing them with impurity peaks, is not a problem in the described method, because the approximate positions of the peaks are known *a priori*. Nevertheless, the error in small peaks is a problem in this method: the absolute errors of peak areas are roughly independent of peak area; consequently their relative errors increase with decreasing peak size. But this is due to the limited resolution of the spectral data from the amino acid analyzer, and therefore the accuracy of the areas of small peaks is similar to that obtained in other evaluation methods.

When the peak amplitude of an amino acid can be increased by measuring a different physical parameter (*e.g.*, if ninhydrin staining is used, light absorption at 440 nm for proline rather than at 570 nm as for the other species), the digitize-and-fit method allows one to utilize this larger peak by an adjustment of the baseline. This is demonstrated in Fig. 6. (For convenience, the direction of the time axis has been reversed. only within a spectrum, during the evaluation process.)

Probably one of the most annoying problems in the evaluation of amino acid spectra is a more or less "active" background. In particular, a sudden buffer switching causes a discontinuous baseline jump, which, at worst, may occur within a peak. If abrupt buffer switching cannot be circumvented, the observation of the spectrum via an additional parameter (*e.g.*, light absorption at two different wavelengths) might help to improve the estimate of the time course of the true background in the vicinity of the step.

Principally, a synchronous evaluation of a spectrogram, like direct peak integration, allows a less accurate baseline tracing than is possible by a trace after the spectrum has been completed. This is demonstrated in Fig. 7: whereas a synchronous method cannot know the future time course of the baseline at the moment of specification of the current value (it preserves the initial value during peaks I and II until an interpeak gap of minimal duration T has passed), methods tracing an already



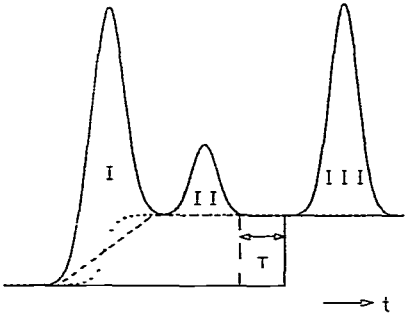


Fig. 6. The exploitation of different wavelength dependendences of light absorption for different amino acids improves the accuracy of the proline determination (ninhydrin staining). Upper spectrum, absorption at 570 nm; lower spectrum, absorption at 440 nm. The continuous line is the baseline used for the evaluation.

Fig. 7. Baseline detection by different evaluation methods. Continuous curve, observed spectrum; dotted curve, true baseline; continuous line with step, baseline assumed by a synchronous method; dashed line (short dashes), baseline assumed by a method knowing the complete spectrum.

completed chromatogram allow a more precise estimate of the baseline. The "synchronous" baseline leads to the correct area of peak III, to a considerable error in the area of peak I and to a completely false area for peak II. A baseline determined subsequently results in a moderate error in the area of peak I, whereas the other two areas are correct.

For the digitize-and-fit method proposed in this paper, a manual baseline tracing (on the digitizer tablet) of each completed spectrum is preferred, because manual tracing is fast and easy to perform and probably gives the most reliable baseline, since no algorithm is known for an automatic background detection* which gives satisfactory results in cases like Fig. 7. If quieter baselines can be obtained, *e.g.*, by trying to perform difference measurements against the background**, using purer water and chemicals, utilizing iso-pH buffer systems[8] and implementing ammonia traps[9], data from the amino acid analyzer can be passed directly to the computer, with all of the manual digitization steps omitted.

Aging of the column bed (mainly due to volume changes) and of the detection system (*e.g.*, by drift of the amplifiers and fading of the ninhydrin staining) causes a

* Whereas the backgrounds are to be determined manually, manual tracing of the spectra themselves is time consuming and a source of additional error. Hence, it is planned to digitize and store the spectra during their creation by the analyzer, and to replace the curve digitizer by a display with a light-pen facility.

** In the Durrum amino acid analyzer[7] an alternating wavelength photometer provides transmittances simultaneously at 590 and 690 nm. Assuming transparency of the ninhydrin-stained products at 690 nm, the ratio of absorbances should compensate for most of the varying background. However, baseline jumps due to buffer switching persist.

deviation from the proportionality between the amount and peak area of a specific amino acid, when two peaks are compared, one of which has passed through the column much later than the other. The compromise between extremely large columns, showing significant aging effects of their beds over the period in which several spectra are run, and short columns which reveal poor peak separations, results in a moderate change in the parameters of interest, the ratios of "peak areas over total amounts". Interruption of the sequence of unknown spectra and running of a standard sample of known composition, in addition to running standards as the leading and trailing samples, allows control of the aging processes. Even when the aging effects are negligible, as in the example given in Results, interpolation between several standards improves the accuracy of the calibration.

As described in Methods, the ratio of "peak area over amino acid amount" as a function of time for each amino acid can now be found by interpolation. This procedure of compensating for aging effects can of course be utilized by any evaluation method (providing sufficient standards are run), but is much more conveniently incorporated as a subtask into a computer program.

Finally one important restriction of the discussed computer method is inherent to this kind of approach: the basic assumption is that the shapes of the amino acid peaks represent Gaussian distributions. If the distribution isotherms are linear and the number of random displacements of the migrating substrate molecules is sufficiently high, Gaussian distributions provide an accurate description[10]. Since all of the isotherms become linear at vanishing substrate concentrations, any elution spectra can be analyzed by this method, providing the column is not overloaded with the material to be separated, and the trailing procedures (such as sample staining in a reaction coil prior to detection) do not cause peak distortions.

It should be noted that the computer program is not restricted to amino acid analysis or gas chromatography, but can also be used to quantitate any spectra which can be described by sets of additively superimposed Gaussian distributions. The least squares method should be especially useful when extensive peak overlapping occurs, although the separation into Gaussians is rather sensitive to small perturbations of the spectral input data.

The elution profile obtained on gel chromatography of erythrocyte membrane proteins serves as an example of a group of 8 poorly resolved, but biochemically identified species. Fig. 8 is the elution profile overlaid by the fitted envelope of summed Gaussians (marked by YYYYY), and shows the single fitted peaks. The peak areas relative to the largest peak are given below the base line. In spite of the fact that a fit to less than 8 Gaussians generates several diffuse, broad peaks, that a fit containing more than 8 Gaussians adds small, insignificant peaks and that the deviation between the measured spectrum and the model decreases by an order of magnitude as soon as the number of peaks to be fitted goes up to 8 in this example, the number of species composing any profile should be an input parameter rather than a value to be found by the computer. For a much greater number of extensively overlapping peaks (tests were run with up to 30 peaks) there may no longer be a 1:1 correspondence between the ripples of the experimental profile and the Gaussians in the synthetic spectrum. The ambiguity in the assignment of fitted Gaussians to measured peaks disappears with improved peak separation in the input curve, with a decrease in the number of peaks and with vanishing deviations of the experimental peaks from the Gaussian shape.

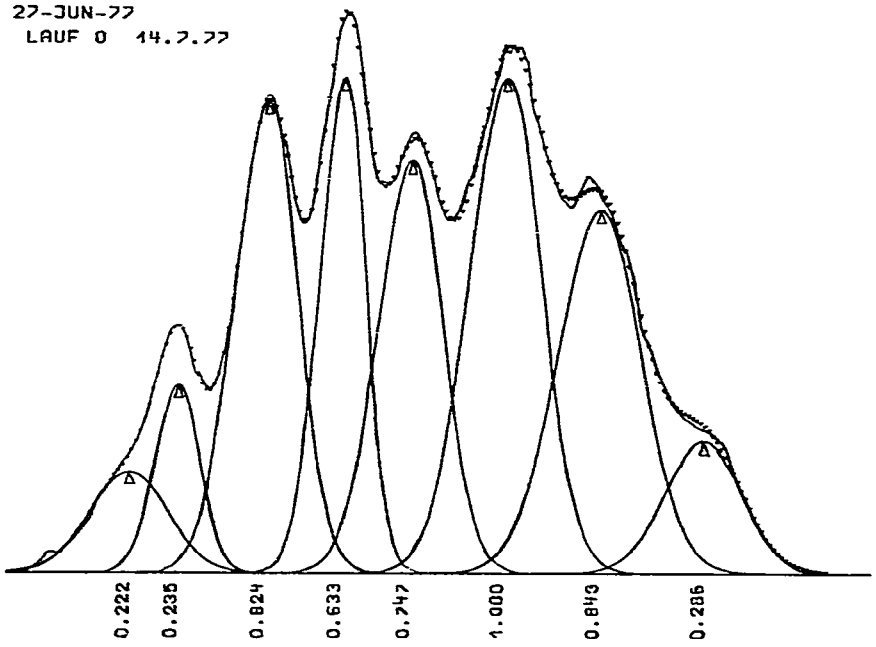


Fig. 8. The elution profile obtained on gel chromatography of oxidized human erythrocyte membrane proteins. The fitted envelope (YYYYY) and single peaks are also given. Peak areas relative to the largest area are shown at the bottom of the figure.

APPENDIX

Computer programs

AMACAN, the computer program used to analyze the spectra, is written in Fortran IV (ref. 11) and run on a PDP11 computer[12] connected to a Summagraphics curve digitizer (resolution, 0.25 mm)[13] via a special interface[14]. The main tasks of the computer program are:

(a) Input/output data handling, especially the acceptance of data (= pairs of coordinates and status) from the curve digitizer. (This is achieved by a Fortran callable assembler subroutine.)

(b) Subtraction of background for each raw spectrum.

(c) Computation of an initial value of the inverse of the Jacobian matrix of the function to be minimized.

(d) Fitting of a theoretical spectrum, consisting of a set of additively superimposed Gaussian distributions (one Gaussian for each amino acid), to each measured background-corrected amino acid spectrum.

Several different versions of the program have been set up, depending upon whether the digitization of the spectrograms is on-line or not, and/or the fits of the theoretical spectra to the measured spectra are on-line or off-line. A dialogue in the on-line version allows some feed-back, such as updating of the parameters or redigitization of a spectrum, whereas the off-line version allows the time-consuming fit to be done at any time at which the load on the computer is low and it also saves the user's time if the fit is done on a slow computer.

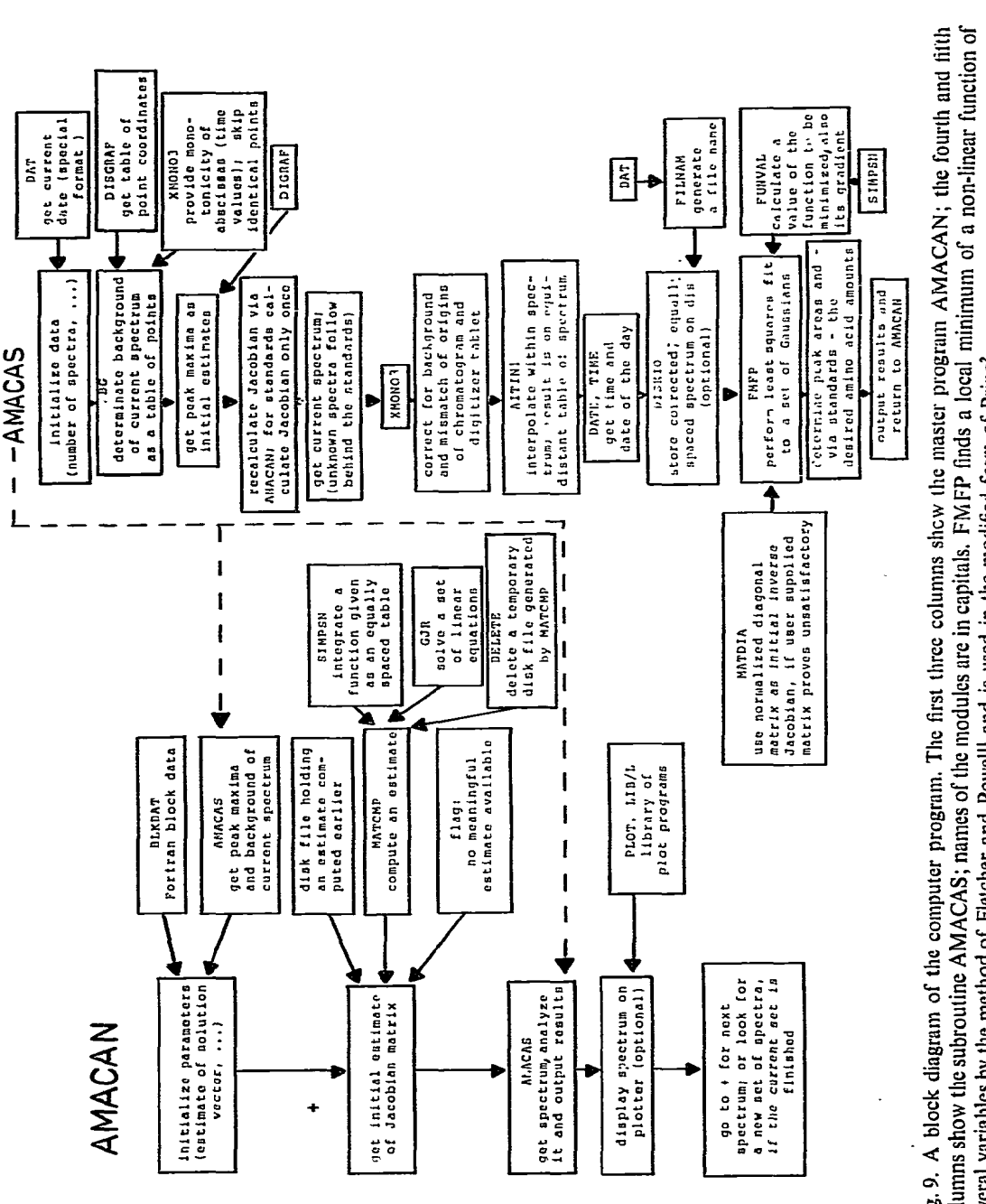


Fig. 9. A block diagram of the computer program. The first three columns show the master program AMACAN; the fourth and fifth columns show the subroutine AMACAS; names of the modules are in capitals. FMFP finds a local minimum of a non-linear function of several variables by the method of Fletcher and Powell[1] and is used in the modified form of Pring[2].

The special program GAUSFT allows fits of up to 30 additive Gaussians to any spectrum, for which peak widths, locations and amplitudes are to be found simultaneously.

The function of the modules and subroutines is explained in Fig. 9, which represents a block diagram of the computer program. The program size and time consumption depend heavily on the efficiency of the code generated by the Fortran compiler and on the computer hardware itself (program size on a PDP11: 22k words above monitor). In order that the program can be used on small computers, it can be linked in a manner which allows overwriting of parts of the program which are currently not in use. Fig. 10 displays the overlay structure of the program AMACAN.

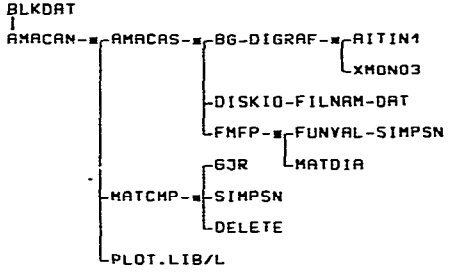


Fig. 10. The overlay structure of the computer program. The asterisk sign means an automatic overlay control point, which allows mutual overwriting of horizontal branches starting at, or vertically below, this control point.

ACKNOWLEDGEMENTS

The author thanks Professor H. Passow for his encouragement and critical reading of the manuscript. The valuable advice and discussions provided by Dr. V. Rudloff (as a potential user) are appreciated. Gratitude must be expressed to Miss D. Dahlem for carefully preparing the analyses.

REFERENCES

1 R. Fletcher and M. J. D. Powell, *Computer J.*, 6 (1963) 163.
2 M. Pring, Medical School Computer Facility, University of Pennsylvania, Philadelphia, Pa.
3 H. J. Jones and D. W. Spence, *Infotronics Application Notes*, No. 1, Infotronics Corp., Bolder, Calif., 1964.
4 B. Sheldrick, *Quart. Rev., Chem. Soc.*, 24 (1970) 454.
5 H.-L. Back, P. J. Buttery and K. Gregson, *J. Chromatogr.*, 68 (1972) 103.
6 R. Taylor and M. G. Davies, *Anal. Biochem.*, 51 (1973) 180.
7 Durrum Instrument Corporation, Palo Alto, Calif., U.S.A.
8 J. R. Benson, in R. N. Perham (Editor), *Instrumentation in Amino-Acid Sequence Analysis*, Academic Press, London, 1975, pp. 7–8.
9 J. R. Benson, in R. N. Perham (Editor), *Instrumentation in Amino-Acid Sequence Analysis*, Academic Press, London, 1975, pp. 25–28.
10 B. L. Karger, L. B. Snyder and C. Horvath, *An Introduction to Separation Science*, Wiley, New York, London, Sydney, Toronto, 1973, pp. 129–145.
11 Fortran X3.9, American National Standards Institute, 1966.
12 Digital Equipment Corporation, Maynard, Mass., U.S.A.
13 Scriptographics Corporation, Fairfield, Conn., U.S.A.
14 Tewidata GmbH, Munich., G.F.R.