Journal of Chromatography, 155 (1978) 249–259

CHROM. 10,857

COMPUTER PROGRAM FOR PROCESSING DATA ACQUIRED FROM GAS–LIQUID CHROMATOGRAPHIC ANALYSIS OF AMINO ACIDS

BABOO M. NAIR

Department of Applied Nutrition, University of Lund, Chemical Centre, Box 740, S-220 07-Lund (Sweden)

(First received October 21st, 1977; revised manuscript received December 28th, 1977)

SUMMARY

In order to automate the gas–liquid chromatographic analysis of amino acids and reduce the cost, which is essential in large screening programms, a data processing program was formulated. The crude integrator data containing retention times and peak area values were collected on punched paper tape. The data from the punched tape were transferred to a magnetic tape and processed further in a Univac 1108 computer with the help of a Fortran program. The structure of the data obtained and the procedure for processing and presentation of results are described.

INTRODUCTION

In an earlier paper, a modification of the technique for the gas–liquid chromatographic (GLC) analysis of N-trifluoroacetyl-*n*-butyl (TAB) derivatives of amino acids was presented[1]. Instead of utilizing the signals from the two detectors and electrometers for complete and simultaneous quantitation in two independent digital integrators of all relevant peaks or in one set of equipment recording the results from the two columns successively, we utilized first the signal from the detector of the EGA column until the elution of TAB aspartic acid was completed and then that from the OV-17 column on a single integrating and recording unit. In this way, a saving of analysis time or investment cost has been accomplished. As every amino acid has to be calculated with separate calibration factors, the calculation, tabulation and presentation of data acquired during a large project would obviously be very tedious.

In order to automate the analysis further and reduce the costs, which is essential in large screening programme such as in plant breeding or nutritional surveys, an automatic data-processing system to replace the time-consuming calculations of the "crude" integrator data was sought. One solution to this problem is presented in this paper.

DATA COLLECTION

The column system and the flow of signals are illustrated in Fig. 1. The EGA column was connected to the detector A and the OV-17 column to the detector B. Both detectors were then connected to the teleprinter through a dual differential electrometer–digital integrator system and to the potentiometric recorder. With this arrangement, signals from each of the two detectors could be carried to the integrator and the recorder as and when desired by changing the mode-selector position between A and B.

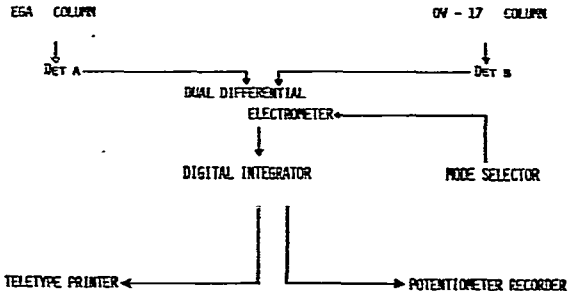


Fig. 1. Diagram showing the flow of detector signals in a dual-column system for amino acid analysis by GLC.

First, a standard sample was injected into the EGA column and then, after a complete run, into the OV-17 column. By comparing the chromatographic charts from both we could select an optimal time for a switchover of the mode selector from A (EGA) to B (OV-17). We chose to change the mode selector from A to B after the elution of TAB aspartic acid from the EGA column, which was probably the best place because we could exploit the EGA column more satisfactorily. The results from our experiment in which first a sub-sample was injected into the EGA column and then another sub-sample into the OV-17 column, as presented in Fig. 2, ensured that a chromatogram containing of all the amino acid peaks was obtained.

We used norleucine as the internal standard in the EGA column and butyl stearate in the OV-17 column for more accurate calibration and further calculation of the data. On certain occasions some difficulty occurred in obtaining satisfactory separation of norleucine from neighbouring peaks. Our experiment in search of a new internal standard resulted in the use of α-aminocaprylic acid. As can be seen from Fig. 2, a simultaneous separation of all of the amino acids could be carried out within about 35 min using the improved method described above. Nevertheless, if only the amino acids separated on the EGA column are of interest, it is possible to run it alone. Thus, a choice between the EGA column alone and the EGA + OV-17 dual-column system was available.

There are principally five types of systems that could be used for the automatic data processing[2–8]: (1) off-line systems; (2) hybrid systems; (3) time-shared computer system; (4) dedicated computer system; and (5) multichannel dedicated computer system.

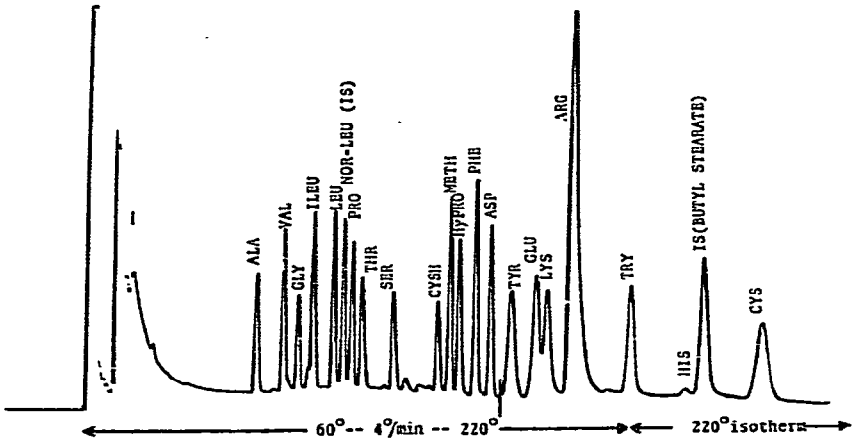


Fig. 2. Separation of all the protein amino acids from a dual-column system. A 0.5-μg sample of each amino acid, norleucine and butyl stearate as the internal standards was injected into the EGA column first and then into the OV-17 column simultaneously within a period 2 min before the beginning of the temperature program (60 to 220° at 4°/min). Argon flow-rate, 50 ml/min. Attenuation, $10^{-10} \times 2$.

(1) An off-line system basically consists of, in addition to a gas–liquid chromatograph, an integrator, which converts the analogue data into digital data, attached to a printer, and with cards, paper-tape or magnetic-tape facilities for data storage. These stored data can then be processed in the computer when it is available and the amount of "crude" data accumulated is adequate for an economically sound used of the computer. This system would be the least expensive and in addition offers a smooth means of adapting to fully computerized GLC systems.

(2) In the hybrid system, many GLC integrator units could be connected to and controlled by one computer-printing device. A major disadvantage is the risk of a serious delay in the analytical work if the computer stops funtioning, as a consequence of which all the GLC integrator units that depend on it would also be out of use.

(3) With the time-shared computer system, one computer with a larger memory is directly connected to many analytical instruments. The large memory would be used for storage of extensive data, tables, statistical data, routines for data interpretation and other programs.

(4) and (5) Both dedicated systems require very large investments and are practicable only for laboratories that have many analytical instruments functioning simultaneously.

The alternative for data processing assessed as optimal under our working conditions was an off-line system (Fig. 3). The analogue signals were carried to an electronic digital integrator when supplying the teletype printer with data on retention times and peak areas. The teletype, in addition to its own printout, was also equipped with a paper-tape punching mechanism. This may not be generally acceptable to the analyst. However, where long calculations are involved, especially when debugging and internal standard identification instructions are to be supplied, there is often much convenience in using this approach.

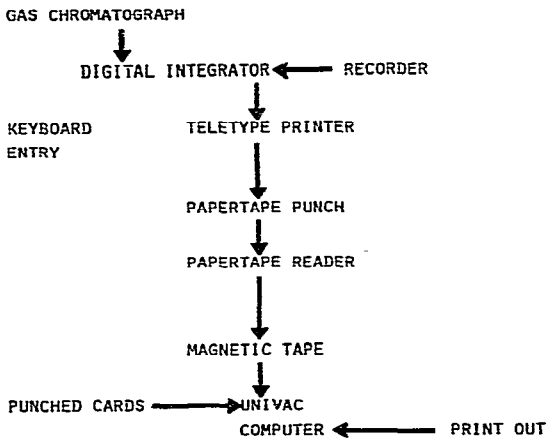


Fig. 3. Schematic presentation of a data-handling system for amino acid analysis by GLC.

Before the sample was injected, information about the sample (weight, nitrogen content, etc.) was introduced to the tape through the keyboard as presented in Table I (points 1–5). The teletype was then switched on to the "on-line" position and a chromatographic analysis was run. Data on retention times and peak areas were obtained as listed under 6 in Table I. At the end of the run, debugging instructions were entered through the keyboard (points 7 and 8). Further, the retention data for the internal standards were introduced through the keyboard as observed on the printout and chromatogram. The actual record of a complete run is shown in Fig. 4. The accumulated data on the paper tape were stored there and transferred before processing to magnetic tape using a tape reader (RC-2000).

DATA PROCESSING

The program for the data processing was developed for the Univac 1108 computer in Fortran V (see Appendix). The program for processing of punched tape contains eight operational subroutines and a main program:

1. BINARS: Binary search
2. MSORT: Sorting
3. KROM: Punched tape read
4. RINT: Read whole numbers
5. RREAL: Read real numbers
6. RSTR: Read strings
7. INTAB: Read conversion tables
8. INFILE: File definition

PARAM: Main program

In addition to the data contained in the magnetic tape, which are transferred from paper tape, three sets of parameter punched cards are used for processing and presentation (see Appendix). The first set carries the number of the sample to be

```
Amino Acid analysis by GLC. Data collection and processing. Dual column
data for programme test. Norleucine and butyl stearate as internal
standards.
-72 06 07- -001-;

0.094 0.117 0.127 0.115 0.115 0.000 0.185 0.255 0.161 0.000 0.000 0.000
0.120 0.128 0.000 0.351 0.111 0.129 0.120 0.150 0.135 0.130 -1

9.45 0.165 0.187

097:0039834
113:0059108
127:0043540
135:0041325
146:0058973
152:0037225
156:0085748
165:0050852
182:0055043
218:0015345
225:0020789
234:0048466
242:0082941
274:0028600
290:0131637
302:0056092
319:0025318
351:0023300
392:0027403
408:0036224
445:0025560
-1

000 000 000 000 -1
000 000 000 -1

000 152 408 260
```

Fig. 4. Printout from the teletype after a single GLC run. The circled figures refer to the text in Table I

TABLE I

STRUCTURE OF THE DATA OBTAINED AS IN FIG. 4

Data	
Through keyboard	*From GC integrator*
(1) Text, information introduced into the punched tape through the keyboard of the teletype printer. This could include a description of the sample, internal standards used, date and sample number. Entered as below: – 72 06 02 – – 001 –	(6) Values for retention time and peak area 0.97:003984 113:005988 m : : : 445:0025560
(2) Relative molar ratio of all the amino acids entered in the order they are eluted, starting with value for alanine, etc.: 0.094 0.117 0.127 0.130 –1	
(3) Weight of the sample (mg)	
(4) Weight of the internal standard 1 (mg) 9.45 0.165	–1
(5) Weight of the internal standard 2 (mg) 0.187	
(7) 000 000 000 –1 (retention time of peak to be deleted)	
(8) 000 002 001 –1 (000, retention time of the first half of the split peak; 002, retention time of the second half; 001, retention time of the added peak)	
(9) 001 002 008 005: 001, retention time of the internal standard in the single column 002, retention time of the internal standard in the EGA column 008, retention time of the internal standard in the OV-17 column 005, time at which the switch was turned over from EGA to OV-17	

AMINO ACID ANALYSIS BY GLC.DATA COLLECTION AND PROCESSING.DUEL COLUMN DATA FOR PROGRAMME TEST.NORLEUCINE AND BUTYL STEARATE AS INTERNAL STANDARDS. -77 06 07- -001-

DATE: 770607 SAMPLE WEIGHT 9.45 MGS PROTEIN CONTENT 32.19 %

SAMPLE: 1 MOISTURE CONTENT 7.39% NITROGEN CONTENT 5.15 %

S.NO	AM.ACID	REL.RETENT.	AAA/AIS	RNRFMGAA IN SAMP.	MGAA/GR SAMP.	GAA/16GR NITGNAA/100GR PRHMOL AA/GR SAMPLE		
1	ALA	.638	1.070	.094	.101	10.644	3.307	3.307	.119
2	VAL	.743	1.588	.117	.186	19.659	6.108	6.108	.168
3	GLY	.836	1.170	.127	.149	15.719	4.884	4.884	.210
4	ILEU	.888	1.110	.115	.128	13.510	4.197	4.197	.103
5	LEU	.961	1.584	.115	.182	19.279	5.990	5.990	.147
6	PRO	1.026	2.304	.185	.426	45.095	14.010	14.010	.392
7	THR	1.086	1.366	.255	.348	36.862	11.452	11.452	.310
8	SER	1.197	1.479	.161	.238	25.192	7.827	7.827	.240
9	METH	1.434	.412	.000	.000	.000	.000	.000	.000
10	HYPRO	1.460	.558	.000	.000	.000	.000	.000	.000
11	PHE	1.539	1.302	.120	.156	16.533	5.136	5.136	.100
12	ASP	1.592	2.228	.128	.285	30.180	9.376	9.376	.227
13	TYR	.672	.790	.351	.277	29.325	9.111	9.111	.162
14	GLU	.711	3.634	.111	.403	42.685	13.261	13.261	.290
15	LYS	.790	1.548	.129	.200	21.138	6.567	6.567	.145
16	ARG	.782	.699	.120	.084	8.875	2.757	2.757	.051
17	TRY	.860	.643	.130	.084	8.849	2.749	2.749	.043
18	HIS	.961	.756	.150	.113	12.008	3.731	3.731	.077
19	CYS	1.091	.706	.135	.095	10.080	3.132	3.132	.042

Fig. 5. Printout of the processed data for one sample by the Univac 1108.

processed and the index to the corresponding card in set II carrying the data on moisture content, nitrogen content and the conversion factor for nitrogen to protein, and a second index to a card in set III to identify the relative retention-time window limits with two alternatives, single column (EGA) or dual-column (EGA + OV-17). The cards in set II have the information on moisture content, nitrogen content, etc., for each sample. This information can also be introduced through the keyboard (see point 1 in Table I) if available at the time of the GLC run. For the purpose of calculation, however, the computer cannot use this information under text 1, Table I.

Fig. 5 presents the printout of the final results of processed data. Because there have been many different ways of expressing amino acid analysis data, we have employed most of the common ways that we observed in the literature.

PEAK IDENTIFICATION

The identification of the GLC peaks by the computer is based on the retention times of individual amino acids relative to those of the internal standards. The memory data are from analyses of the TAB derivatives of a standard amino acid mixture.

A very simple logic is used in our program for the peak identifications, namely the use of two limiting values of relative retention time within which any peak is considered as being identified with the derivative that in the memory has a certain value for relative retention time, generally the mean of the two limit values.

As is well known, several of the TAB derivatives of amino acids separate near one another. As the stationary phase of a GLC column could bleed out, the relative retention times of certain amino acid derivatives could change. Also other analytical parameters could be a source of minor variation in relative retention times in a temperature-programmed GLC analysis. Therefore, a more complex logic such as that suggested by the Vidar Corporation[9] may have to be applied.

CALCULATION

For routine analysis, weight response calculation was convenient. Data available from the analysis of each sample and the method of calculation are as follows.

Peak area of amino acid = Aaaa.

Peak area of internal standard = Ais.

Weight (mg) of amino acid = Mg aa.

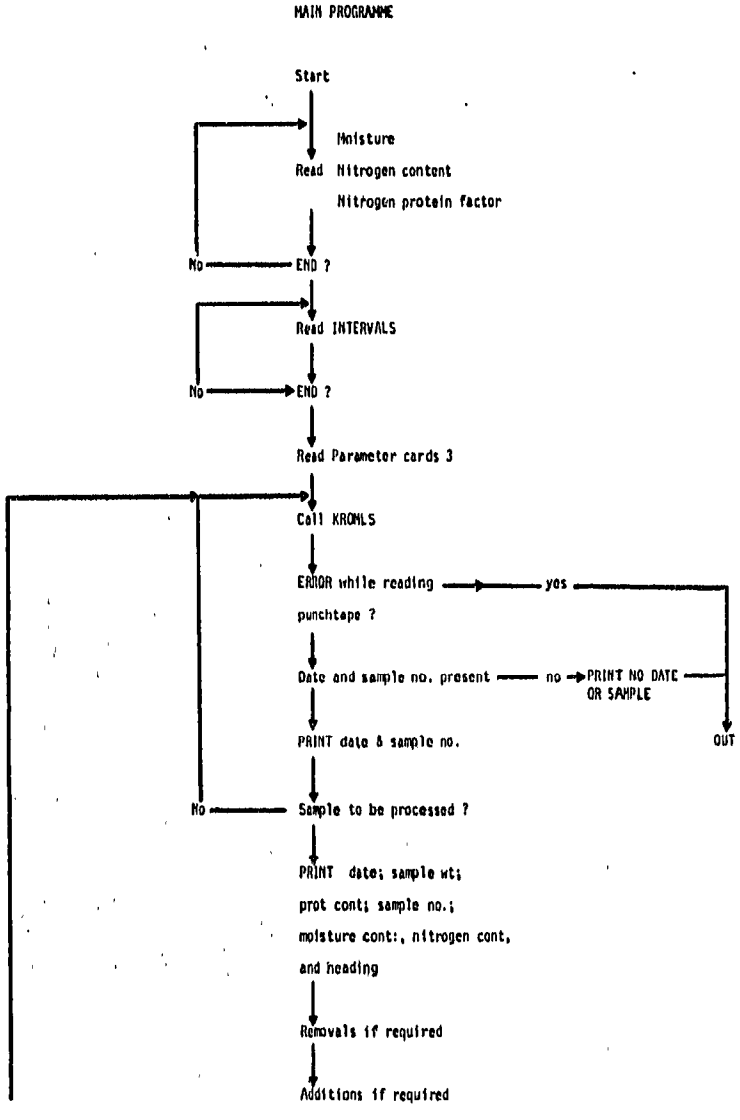
Weight (mg) of internal standard = Mg is.

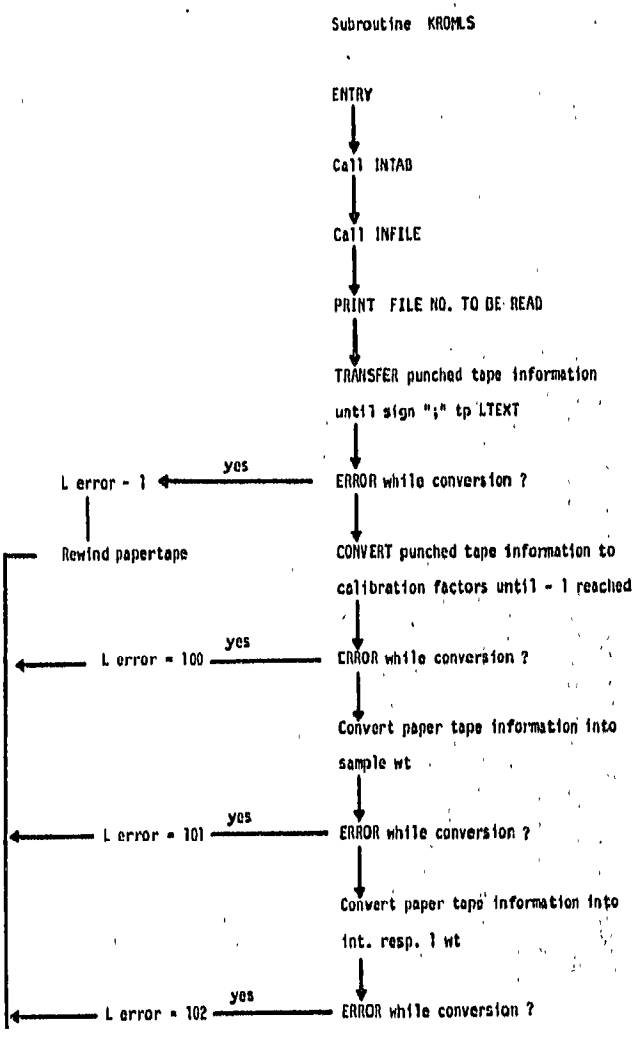
Relative retention time = time aa/time is.

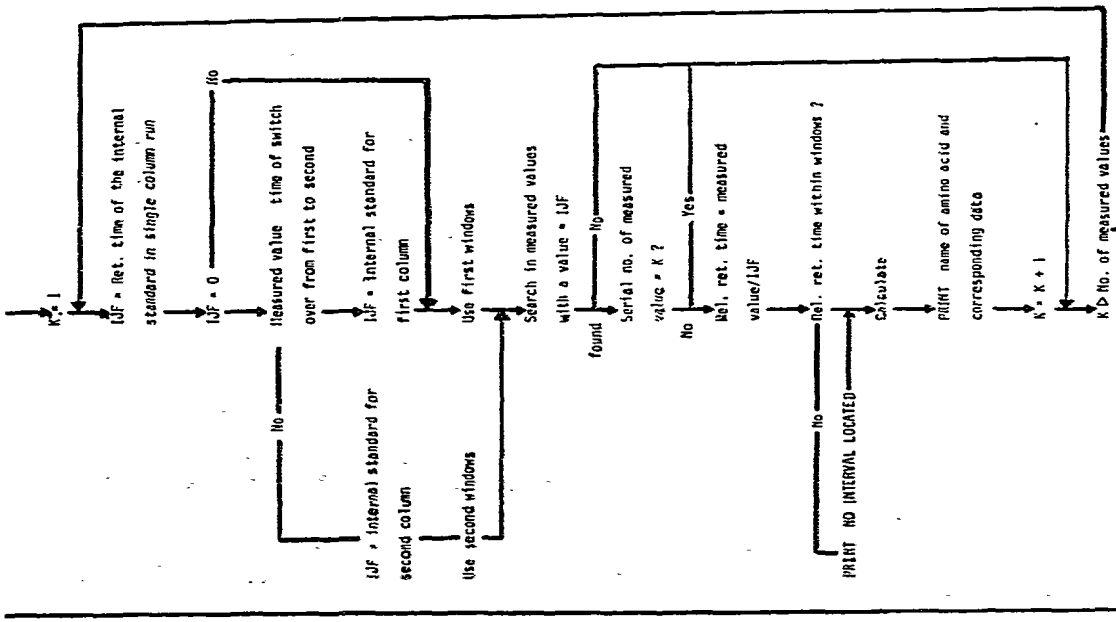
$$\text{Relative molar response factor (RMRF)} = \frac{\text{Ais} \times \text{Mg aa}}{\text{Aaa}}.$$

RMRFH = RMRF × correction factor for hydrolysis losses.

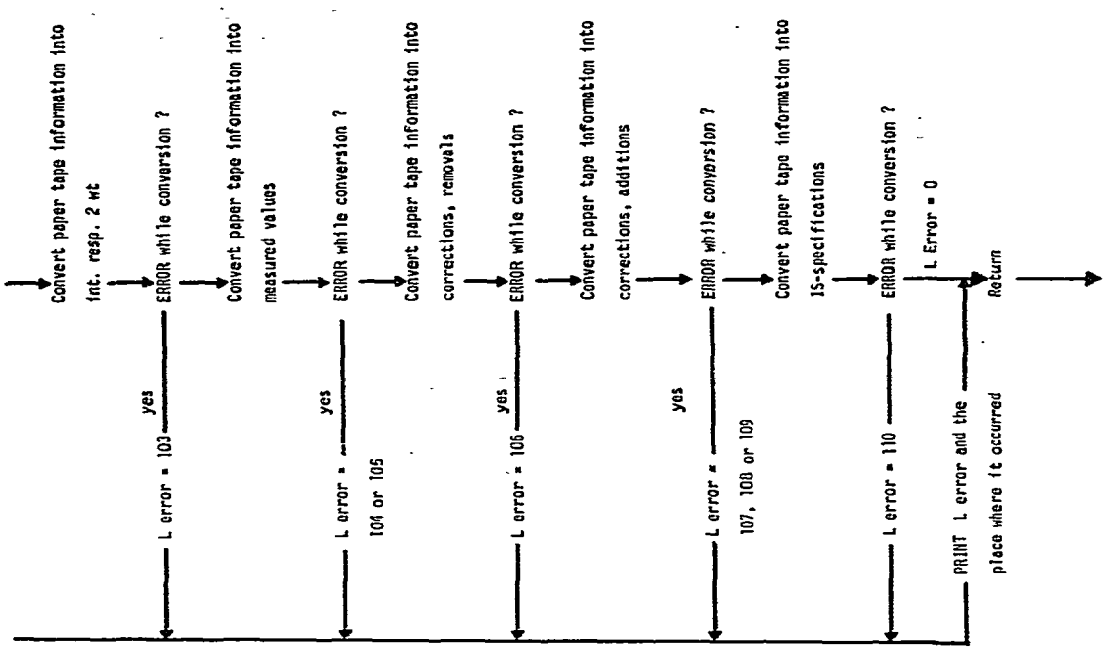
Amounts of internal standard in the calibration mixture and in the sample are the same.


MAIN PROGRAMME
Start
Moisture
Read Nitrogen content
Nitrogen protein factor
No
END ?
Read INTERVALS
No
END ?
Read Parameter cards 3
Call KROMLS
ERROR while reading
yes
punchtape ?
Date and sample no. present
no
PRINT NO DATE OR SAMPLE
OUT
PRINT date & sample no.
No
Sample to be processed ?
PRINT date; sample wt;
prot cont; sample no.;
moisture cont:, nitrogen cont,
and heading
Removals if required
Additions if required


Subroutine KROMLS
ENTRY
Call INTAB
Call INFILE
PRINT FILE NO. TO BE READ
TRANSFER punched tape information
until sign ";" to LTEXT
yes
L error = 1
ERROR while conversion ?
Rewind papertape
CONVERT punched tape information to
calibration factors until - 1 reached
yes
L error = 100
ERROR while conversion ?
Convert paper tape information into
sample wt
yes
L error = 101
ERROR while conversion ?
Convert paper tape information into
int. resp. I wt
yes
L error = 102
ERROR while conversion ?



Scheme A.



Scheme B.

Sample number.

Date of run.

Sample weight: SW mg.

Moisture: M%.

Protein: P%.

Nitrogen: N%.

NPF = nitrogen to protein conversion factor.

$$\text{Amount of amino acid (mg) in the sample (Mg aa)} = \frac{\text{Aaa} \times \text{RMRFH}}{\text{Ais}}.$$

$$\text{Amount of amino acid (mg) in 1 gram of sample (W)} = \frac{\text{Mg aa} \times 1000}{\text{SG}}.$$

$$\text{Protein (P)\%} = \%\,\text{N} \times \text{NPF}.$$

$$\text{Grams amino acid/100 g protein} = \frac{\text{W} \times 10}{\text{P}}.$$

$$\text{Milligrams amino acid/16 g nitrogen} = \frac{\text{W} \times 16 \times 100}{\text{N}}.$$

$$\text{Grams amino acid/16 g nitrogen} = \frac{\text{W} \times 16}{\text{N} \times 10}.$$

FURTHER AUTOMATION OF THE GLC ANALYSIS OF TAB AMINO ACIDS

The introduction of some information about the sample and additional information for debugging through the keyboard of the teletype for each sample before and after the run could be a handicap when using automation for sample introduction and data collection. There would not be space in the punched tape for introducing this information after the run of a series of samples. This problem could be resolved either (1) by leaving a gap on the tape between the runs that is large enough to take the necessary information, which could be introduced later on when the series are ready, or (2) by providing a reference sample number and making provision in the program for a set of cards containing this information.

APPENDIX

See Schemes A and B.

Subroutine KROMLS

The subroutine KROMLS uses standard library programs INTAB and INFILE for conversion of information from punched tape to the computer's internal code. If something goes wrong during conversion, it prints the place where it occurred. Subroutine INTAB gives the conversion table. After INFILE has been called, conversion of text string, real numbers and integral numbers is achieved by calling RSTR, RREAL and RINT, respectively.

Subroutine BINARS

Subroutine BINARS does binary searching. If the given value is located in the array, the variable IX gets the value corresponding to its location and the variable IND gets value 2. IND gets other values if the given value is not located.

Subroutine MSORT

Subroutine MSORT sorts out a two dimensional array in ascending order according to the second index.

Punched cards

Set I. Variables.
Sample number to be processed.
Index to variables in set II.
Index to variables in set III.
Last card —1 in column 15–16.
Set II. Variables:
M: moisture (%).
N: nitrogen (%).
NPF: nitrogen to protein factor.
One card for each sample with sample number as reference.
Format 3 F 7.2. Last card in the set $\frac{7}{8}$ b EØF.
Set III. Variables:
Names of the amino acid.
Relative retention time window limits.
Molecular weights.
Each card contains two sets of values for retention time window limit, one for single-column run and the other for dual-column run. Last card $\frac{7}{8}$ b EØF.
Format (2 (16 218, 15)).

ACKNOWLEDGEMENTS

Generous grants from the Swedish Board for Technical Development and the assistance of H. Kvist and A. Roy are gratefully acknowledged.

REFERENCES

1 L.-A. Appelqvist and B. M. Nair, *J. Chromatogr.*, 124 (1976) 239.
2 J. M. Gill and H. W. Habgood, *J. Chromatogr. Sci.*, 5 (1967) 595.
3 J. M. Gill and S. P. Perone, *J. Chromatogr. Sci.*, 7 (1969) 709.
4 L. L. Hegedus and E. E. Petersen, *J. Chromatogr. Sci.*, 9 (1971) 551.
5 J. M. Gill, *J. Chromatogr. Sci.*, 7 (1969) 731.
6 D. R. Deans, *J. Chromatogr. Sci.*, 9 (1971) 729.
7 C. E. West and T. R. Rowbotham, *J. Chromatogr. Sci.*, 8 (1970) 688.
8 H. E. Pattee, E. H. Wiser and J.-A. Singhton, *J. Chromatogr. Sci.*, 8 (1970) 668.
9 Autolab-Spectraphysics, Technical Bulletin No. 107-72, 1972.