Journal of Chromatography, 92 (1974) 113–118

CHROM. 7282

A COMPUTER TECHNIQUE FOR THE ANALYSIS OF AMINO ACID CHROMATOGRAMS

BRYAN D. YOUNG

The Beatson Institute for Cancer Research, Royal Beatson Memorial Hospital, 132 Hill Street, Glasgow G3 6UD (Great Britain)

(Received November 19th, 1973)

SUMMARY

A computer system for the complete analysis of amino acid chromatograms is described. Chromatograms are recorded continuously on punched paper tape from which a FORTRAN program measures the area of each peak, identifies the amino acid producing it and calculates the amino acid composition of each sample. An incremental graph plotter is used to reproduce each chromatogram and to indicate the assumed identity of each peak.

INTRODUCTION

Automatic amino acid analysers produce large amounts of data which can be conveniently analysed using digital computing techniques. Data logging systems[1–5] have been used to convert the analog output signal to a digital signal from which areas of peaks were calculated. Alternatively, data concerning peak areas have been obtained from electronic integrators or by manual calculation[6–9]. More recently, however, systems combining automatic data logging with programs designed to identify each amino acid peak have been described[10–12]. This technique allows the composition of samples to be determined with the minimum of operator intervention, provided the amino acid peaks can be correctly identified.

This paper describes a system based on a paper tape data logger and a FORTRAN program which identifies each peak and calculates the concentration in mg/l of each amino acid present in a series of samples. A significant feature of this system is the use of an incremental graph plotter to reproduce the chromatogram from the data supplied and to indicate the identity of each peak.

MATERIALS AND METHODS

Chromatography system

Amino acid separation was carried out on a Technicon TSM1 amino acid analyser (Technicon, Tarrytown, N.Y., U.S.A.) modified for the "1-hour system" using a single column of ion-exchange resin[13]. The two-channel chart recorder, which was used to produce chromatograms simultaneously at 570 nm and 440 nm, was fitted with retransmitting slide wires. A stabilised power supply, applied to the

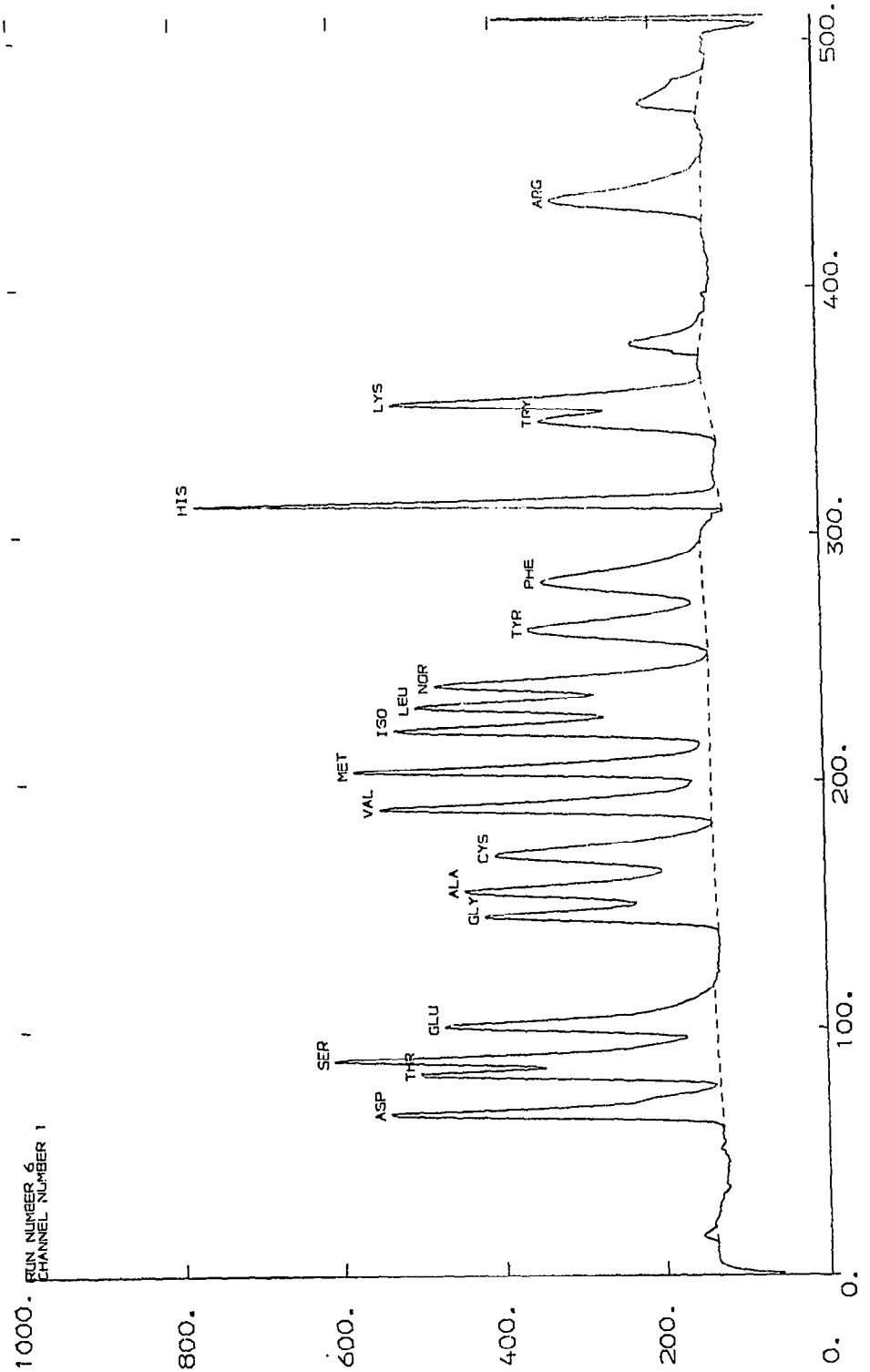

Fig. 1. Graphical output produced by analysis program. The broken line indicates the baseline used to calculate peak areas and the mnemonics indicate which amino acid is assigned to each peak. The number of points, 10 sec apart, in the chromatogram are shown on the abscissa and the voltage, in units of 0.01 V is shown by the ordinate.

slide wires, produced analog signals which were input to a Solartron-Schlumberger data logging system. The punched paper tape produced by this device recorded voltage measurements in the range 0–10 V carried out every 10 sec on both channels. In this form data from a series of chromatograms, including the intermediate washout periods, were transferred to the Glasgow University KDF9 computer.

Analysis program

The analysis program, which was written in FORTRAN, consisted of four main parts. Firstly, the washout periods were identified in order to separate the continuous data into a number of chromatograms. This was followed by the detection of peaks and the measurement of their areas. The identification of the amino acid which produced each peak was then carried out. Lastly, by reference to norleucine internal standards and to standard chromatograms, the amount of each amino acid present in each sample was calculated and printed in molar concentrations, molar percentages and mg/ml.

All amino acids, except proline, were measured from data recorded at 570 nm (channel 1). Proline produces only a small peak at 570 nm and therefore was measured at 440 nm (channel 2). The data input on punched cards included each amino acid name and mnemonic, its channel number, its molecular weight, and its expected elution time (see below).

Identification of washout period. During the washout period between chromatograms the signals from both channels fell to almost 0 V. This was significantly below the normal baseline of the chromatograms and therefore the washout periods could be detected by searching for a series of points below a threshold voltage. The start of each chromatogram was identified as the point at which the signal exceeded the threshold and remained above it for at least 2 min. This technique was not affected by artifacts in the washout period, such as occasional spikes in the signals, and allowed accurate determination of the elution times of the peaks in the chromatogram.

Calculation of peak areas. The calculation of peak areas was carried out in four stages, *viz.* the detection of peaks, the location of the start and end of each peak, the location of the baseline and finally integration to find peak areas. The amount of noise in the analog signals was sufficiently low that the use of a digital smoothing routine[12] was unnecessary.

The first derivative of each chromatogram was calculated as the differences between successive values, and peaks were identified as points at which a positive gradient became negative. In order to prevent the detection of slight baseline variations, minimum thresholds to the positive and negative gradients were used. The start of each peak was assumed to be the last point before the peak at which the gradient was zero or negative and similarly the end was assumed to be the first point after the peak at which the gradient was zero or positive. If a single peak was being processed, the baseline was simply taken as the straight line joining its start and end points. If, however, a complex of peaks, during which the signal does not return to its baseline, is being processed the baseline is taken as the straight line joining the start of the first peak and the end of the last peak in the complex. Fig. 1 shows an example of the graphical output produced by the program, the baselines being indicated by broken lines.

RESULTS IN MILLIGRAMS/LITRE

RUN	ASP	THR	SER	GLU	GLY	ALA	CYS	VAL	MET	ISO	LEU	NOR	TYR	PHE	HIS	TRY	LYS	ARG	PRO
3	131	123	105	140	69	91	224	116	143	124	127	0	173	158	208	196	182	213	131
4	139	118	114	155	74	92	236	114	140	124	131	0	185	173	219	205	199	219	178
5	132	110	112	143	73	89	230	115	146	127	132	0	173	158	200	200	181	196	159
6	134	113	108	142	76	86	223	114	145	129	131	0	176	162	206	204	183	211	152
7	138	133	113	151	71	85	230	114	146	123	125	0	174	165	202	183	192	215	161
8	130	111	111	139	71	85	218	112	144	127	130	0	176	167	206	207	183	221	138
9	136	117	107	140	71	87	227	113	150	132	133	0	178	168	203	201	191	226	149

Fig. 2. Results printed for seven sample chromatograms. The entry under norleucine is zero because it was used as the internal calibration.

The area of each peak was calculated by summation, from the start to the end points, of the difference between each point and the baseline. Hence, the area of a single peak was that bounded by the signal and the baseline. However, the area of a peak which was part of a complex was that bounded by the signal, the baseline and where necessary the perpendicular dropped from the start or end points to the baseline. Althought more complex techniques for the resolution of multiple peaks are available[14], the simple technique described here worked well in practice.

Amino acid identification. A method for the correlation of peaks with the appropriate amino acid has been developed using elution time as a basis. It has been observed in practice that, although the elution times of amino acids relative to the start of the chromatogram are subject to variation, the time intervals between the elution of amino acids from the same column are considerably less variable. Average elution times, which were calculated from a series of twenty standard chromatograms, formed part of the data input on punched cards. The identification procedure, which was applied to the data from each channel separately, was designed to fit the actual elution times to the expected average values. If necessary, the program shifted the observed elution times without affecting their relative timing. In this way, any errors or variation in the start of the chromatogram were eliminated.

There were two contingencies which the identification routine had to take into account. Firstly, small peaks which could occur as rare artefacts had to be discounted. Because the occurrence of such peaks could not usually be related to any of the expected elution times, they remained unidentified and were subsequently ignored. A second problem was the possible absence of an amino acid from a sample. If none of the actual elution times fitted a particular expected elution time, the concentration of that amino acid was assumed to be zero in subsequent calculations. The assignment of peaks to each amino acid was indicated by printing mnemonics on the graphical output (see Fig. 1) and it was therefore easy to verify that correct identification had taken place. Although this example shows only data from channel 1, the identification of proline in channel 2 was carried out in the same manner and a graphical output was also produced.

Calculation of amino acid content. Each chromatogram included a norleucine standard peak which had to be identified correctly by the above procedure before peak areas could be converted to amino acid content. A number of standard chromatograms were also included in a series of samples. The first step was to calculate the ratio of the area of each identified peak to the area of the norleucine peak (*i.e.* norleucine equivalent). The quantity of each amino acid in μmoles was found from:

$$\frac{NLE}{NLE(S)} \times 0.025$$

where NLE is the amino acid norleucine equivalent and $NLE(S)$ is the norleucine equivalent of the corresponding amino acid in the most recent standard chromatogram. These values were expressed as percentages of the total molar amino acid content in each sample and were also printed. The absolute concentrations in mg/l were calculated from the molecular weights of each amino acid and were printed in the format shown in Fig. 2.

DISCUSSION

The system described above is capable of processing a large number of chromatograms in one batch, the only limiting factors being the amounts of paper tape and computer time required. Typically, about ten to fifteen chromatograms can be analysed with approximately 30 sec of central processor time required for each chromatogram. Successful operation of this system depends entirely on correct identification of the peaks, and therefore the graphical output which indicates amino acid assignments is an invaluable aid to the technician. By comparison with the charts recorded directly from the amino acid analyser, the graphical output can also be used to verify that correct data transfer has taken place.

Although it has been shown[10] that elution time is dependent on amino acid concentration, our experience is that this small variation can be taken into account by including an error factor in the expected elution times. This feature, coupled with the shift facility to account for errors in determining the start of the chromatogram, means that mistakes in identification rarely occur.

The operation of this system is independent of both the number of amino acids and their elution times. Hence, if a future modification to the chromatography system should result in changes to these parameters they can be re-established after a number of trials and input on a new set of data cards.

ACKNOWLEDGEMENTS

The author wishes to thank Dr. A. MacGillivray for advice on the use of amino acid analysers and Mrs. Alison Cameron and Mrs. Christine Johnston for technical assistance. This work was supported by grants from the Medical Research Council and the Cancer Research Campaign.

REFERENCES

1 W. L. Porter and E. A. Talley, *Anal. Chem.*, 36 (1964) 1692.
2 M. I. Krichevsky, J. Schwartz and M. Mage, *Anal. Biochem.*, 12 (1965) 94.
3 A. Yonda, D. L. Filmer, H. Pate, N. Alonzo and C. H. W. Hirs, *Anal. Biochem.*, 10 (1965) 53.
4 A. J. Robins, R. A. Evans, J. A. De Siriwardene and A. J. Thomas, *Biochem. J.*, 99 (1966) 46P.
5 H. L. Back, P. J. Buttery and K. Gregson, *J. Chromatogr.*, 68 (1972) 103.
6 G. N. Graham and B. Sheldrick, *Biochem. J.*, 96 (1965) 517.
7 W. C. Starbuck, C. M. Mauritzen, C. McClimans and H. Busch, *Anal. Biochem.*, 20 (1967) 439.
8 K. Ozawa and S. Tanaka, *Anal. Biochem.*, 24 (1968) 270.
9 J. J. T. Gerding, *Int. J. Protein Res.*, 1 (1969) 169.
10 R. E. Exss, H. D. Hill and G. K. Summer, *J. Chromatogr.*, 42 (1969) 442.
11 H. D. Spitz, G. Henyon and J. N. Sivertson, *J. Chromatogr.*, 68 (1972) 111.
12 R. Taylor and M. G. Davies, *Anal. Biochem.*, 51 (1973) 180.
13 O. Smithies, D. Gibson, E. M. Fanning, R. M. Goodfleish, J. G. Gilman and D. L. Ballantyne, *Biochemistry*, 10 (1971) 4912.
14 A. W. Boyne and W. R. H. Duncan, *J. Lipid Res.*, 11 (1970) 293.