CHROM. 4126

COMPUTER ANALYSIS OF AMINO ACID CHROMATOGRAMS*

REINHARD E. EXSS, HOYLE D. HILL AND GEORGE K. SUMMER
Department of Biochemistry, School of Medicine, University of North Carolina, Chapel Hill, N.C. (U.S.A.)
(Received April 15th, 1969)

SUMMARY

A systematic procedure is described which is arranged for computerized qualitative and quantitative analysis of amino acid analyzer chromatograms of protein hydrolysates. The identification of amino acids has usually been made by inspection, before the numerical data are entered into the computer for calculation of the concentrations of the amino acids. In this report we describe a program by which the computer can identify as well as quantitatively determine the amino acids from the data provided by the integrator attached to the automatic amino acid analyzer. The program is written in a common language and calculates amino acid compositions of proteins or peptides.

INTRODUCTION

Automatic analyzers of the type described by SPACKMAN *et al.*[1] are used for colorimetric determination of amino acids in mixtures after their separation by ion-exchange chromatography. A number of attempts have been made to automate data processing of amino acid analyzer records. PORTER AND TALLEY[2] described an analog to digital converter to obtain punched paper tapes which were used as a data source by computers. YONDA *et al.*[3] and KRICHEVSKY *et al.*[4] used systems which converted the analog photometer output signal into a digital signal which was fed into a computer to calculate the areas under the peaks. However, very accurate integration has been achieved by analog voltage to frequency conversion, a feature of many commercially available integrators. More recently STARBUCK *et al.*[5] and OZAWA AND TANAKA[6] introduced programs for computation of amino acid compositions of peptide and protein hydrolysates which utilized data obtained by hand calculation of peak areas or by an integrator.

The computerized processing of amino acid analyzer records thus far has dealt with calculations of peak areas and amino acid compositions of applied samples. No attempts have been made to identify automatically particular peaks of a chromatogram before entering the process of concentration calculation. In this report we de-

* Correspondence should be addressed to: Dr. GEORGE K. SUMMER, Department of Biochemistry, School of Medicine, University of North Carolina, Chapel Hill, N.C. 27514, U.S.A.

scribe a program for qualitative as well as quantitative analysis of amino acids which can be used with an automatic amino acid analyzer. The program is written in a common computer language, can accept data from various instruments, and should be adaptable to a number of different computers, even those with small memory capacities.

MATERIALS AND METHODS

Amino acid analysis was carried out by a Technicon amino acid analyzer (Technicon Corp., Ardsley, N.Y.) using norleucine as an internal standard. Column dimensions were 140 × 0.6 cm with a buffer flow rate of 0.5 ml/min. Column temperature was held constant at 60° throughout the run, and maximal pressure within the column was 200 p.s.i. Chromobeads type A (Lot No. 128A) was used as the separating resin. The gradient (pH 2.875 to 5.000) was obtained from a Technicon autograd with nine interconnected chambers. Buffer solutions at pH 2.875, 3.8 and 5.0 were made according to the recommended protocol in the Technicon amino acid analyzer manual. It is important to emphasize, however, that these buffers were mixed as recommended in the manual for each of the nine chambers, but in sufficient quantity for at least 32 runs rather than for individual runs. Buffers were kept at room temperature to prevent release of air bubbles during runs. Octanoic acid (0.1 ml/10 l buffer) was used to inhibit growth of micro-organisms. Ninhydrin flow rate was 1.06 ml/min. Cuvettes were 15, 8, and 15 mm measuring at 570, 570, and 440 mμ, respectively. Chart speed was either 6 in./h or 1.8 in./h, and the total time of a single run was 22.5 h. The integration of peaks was done by a wide dynamic range integrator (Model No. CRS-11AB/HS/42, Infotronics Corp., Houston, Texas).

Protein hydrolysis was carried out according to a modification of the method of MOORE AND STEIN[7] in 6 *N* hydrochloric acid at 105° for 30 h in sealed evacuated tubes. The single-component porcine insulin, Lot 615-849B-297C was provided by Dr. E. C. SMITHWICK (Lilly Research Laboratories, Indianapolis, Ind.). Amino acids used as standards were obtained from Technicon Corp., Ardsley, N.Y.

The computers used were the Linc and PDP-8 (Digital Equip. Corp., Maynard, Mass.) and the Model 1130 (I.B.M. Corp., White Plains, N.Y.). The program is presented in the APPENDIX in Fortran IV for the I.B.M. Model 1130. It can be simply converted to Fortran II for use with the Linc-PDP-8 intercommunication. In the latter case, however, the program must be split to meet restrictions of the computer.

CALCULATION OF STANDARD VALUES FOR QUALITATIVE ANALYSIS

The automatic identification of an amino acid from an ion-exchange chromatogram is based upon its elution time, which is determined by the elapsed time from the beginning of the run to the top of its peak. Elution time, however, varies unpredictably due to very small changes in experimental parameters such as buffer gradient, column pressure or temperature, sample size and composition. Even the use of time ratios* does not provide a value which is characteristic for a specific amino acid in different samples. Under constant conditions (*i.e.*, if the buffer solutions of each of the nine

* Time ratio is the elution time of a particular amino acid divided by the elution time of the internal standard (in this case, norleucine).

TABLE I

ELUTION TIMES AND TIME RATIOS OF DIFFERENT AMINO ACIDS FOR DIFFERENT AMINO ACID CONCENTRATIONS

The total sample volume was always kept constant at 0.5 ml.

Amino acid	*Quantity of each amino acid in applied sample*							
	0.125 μmoles		*0.25 μmoles*		*0.50 μmoles*		*0.50 μmoles*	
	Elution time (min)	*Time ratio*	*Elution time (min)*	*Time ratio*	*Elution time (min)*	*Time ratio*	*Elution time (min)*	*Time ratio*
Cysteic acid	—	—	—	—	—	—	48.0	0.0719
Hydroxyproline	—	—	—	—	—	—	181.0	0.2712
Aspartic acid	161.9	0.2295	178.4	0.2505	188.7	0.2633	193.0	0.2891
Threonine	193.0	0.2735	203.5	0.2857	213.8	0.2983	212.5	0.3184
Serine	205.7	0.2915	215.4	0.3024	226.4	0.3159	224.0	0.3356
Glutamic acid	252.1	0.3573	262.3	0.3682	272.5	0.3802	270.5	0.4052
Proline	274.0	0.3883	284.0	0.3987	292.9	0.4086	300.5	0.4502
Glycine	351.9	0.4477	359.2	0.5043	371.0	0.5176	357.0	0.5348
Alanine	382.0	0.5414	389.5	0.5468	400.9	0.5593	384.0	0.5783
Valine	492.1	0.6974	499.9	0.7018	507.1	0.7075	480.5	0.7199
Cystine	563.4	0.7985	570.7	0.8012	576.4	0.8041	534.5	0.8007
Methionine	578.0	0.8192	585.3	0.8217	592.4	0.8265	557.0	0.8345
Isoleucine	657.6	0.9320	664.0	0.9322	670.5	0.9354	623.5	0.9341
Leucine	682.0	0.9666	688.4	0.9665	694.6	0.9690	646.5	0.9685
Norleucine	705.6	1.0000	712.3	1.0000	716.8	1.0000	667.5	1.0000
Tyrosine	742.6	1.0524	749.9	1.0528	756.0	1.0547	707.0	1.0592
Phenylalanine	763.7	1.0823	770.9	1.0823	777.2	1.0843	726.0	1.0876
Ammonia	860.3	1.2193	869.6	1.2201	877.8	1.2246	789.0	1.1820
n-Hydroxylysine	—	—	—	—	—	—	864.0	1.2951
allo-Hydroxylysine	—	—	—	—	—	—	871.0	1.3049
Lysine	1024.0	1.4513	1032.0	1.4488	1040.0	1.4509	945.5	1.4165
Histidine	1074.8	1.5232	1083.0	1.5204	1092.2	1.5237	989.0	1.4816
Arginine	1272.4	1.8033	1281.0	1.7989	1288.8	1.7980	1168.5	1.7506

autograd chambers of the Technicon amino acid analyzer are handled as described above, the flow rate and column temperature are kept constant, and the sample applied to the column is the same with respect to quantity and composition), the time ratios are reproducible for particular amino acids in different runs. If the concentrations of amino acids alone are changed in the sample, the time ratios vary as shown in Table I. The elution time and time ratios also change if additional amino acids are added to the mixture to be analyzed (*e.g.*, hydroxyproline and hydroxylysine, Table I). Therefore, if the concentrations of different amino acids in a sample are unknown, their respective time ratios are unpredictable and their identification on the basis of time ratios is unreliable. It is possible, however, to determine the range of variation for the time ratios of different amino acids by performing a standard run with an amino acid concentration which gives about half-scale peak sizes. Each particular time ratio is increased and decreased by the percentage shown in Table II. The percentage values were determined by chromatography of mixtures with various amino acid concentrations. The results obtained represent the upper (UPLIM) and the lower (BOLIM) limit of variability of the time ratios for the particular amino acids present and are calculated as follows:

$$\mathrm{BOLIM(I)} = \frac{\mathrm{BPER(I)} \times \mathrm{STIME(I)}}{\mathrm{STIME(14)}} \qquad (1)$$

TABLE II

THE TIME RATIO VARIABILITY EXPRESSED AS PER CENT OF THE ACTUAL VALUE OBTAINED FOR THE TIME RATIO IN A STANDARD RUN

Amino acid	*Lower limit (BPER)*	*Upper limit (UPER)*
Hydroxyproline	98.0	102.0
Aspartic acid	97.4	104.5
Threonine	96.5	102.3
Serine	96.2	102.2
Glutamic acid	97.7	103.7
Proline	97.3	107.3
Glycine	97.7	104.0
Alanine	98.2	104.0
Valine	98.9	103.2
Cystine	99.2	102.1
Methionine	99.9	104.1
Isoleucine	99.8	100.4
Leucine	99.7	100.3
Norleucine	100.0	100.0
Tyrosine	99.7	100.2
Phenylalanine	99.5	100.2
Ammonia	99.5	100.3
Hydroxylysine	99.5	100.5
Lysine	98.9	101.1
Histidine	98.9	101.2
Arginine	99.3	101.7

$$\text{UPLIM(I)} = \frac{\text{UPER(I)} \times \text{STIME(I)}}{\text{STIME(14)}} \qquad \text{(II)}$$

where

BOLIM(I) = lower limit of time ratio for particular amino acid.

UPLIM(I) = upper limit of time ratio for particular amino acid.

BPER(I) = percentage of time ratio of particular amino acid which represents the lower limit of variation (for actual data, see Table II).

UPER(I) = percentage of time ratio of particular amino acid which represents the upper limit of variation (for actual data, see Table II).

STIME(I) = elution time of particular amino acid peak in standard run in minutes.

STIME(14) = elution time of internal standard norleucine in standard run in minutes.

These ranges are used in the computer for the identification of amino acids as explained below.

CALCULATION OF STANDARD VALUES FOR QUANTITATIVE ANALYSIS

Since each amino acid has a specific color yield with ninhydrin, the quantitative analysis of a peak is not possible until the peak has been identified as corresponding to a particular amino acid. Over the concentration limits specified to be used for this system, eqn. III has been found to produce constants (designated CORA in the

program) specific for each amino acid. The constant, CORA, is used for the calculation of concentration as shown in eqn. V.

$$\text{CORA(I)} = \frac{\text{SCONC} \times \text{SAREA(14)}}{\text{SAREA(I)}} \qquad \text{(III)}$$

where

CORA(I) = color yield ratio of a particular amino acid.

SCONC = micromoles of amino acid in standard.

SAREA(14) = area of internal standard (norleucine) in standard run in integrator counts.

SAREA(I) = area of particular amino acid in standard run in integrator counts.

EXPLANATION AND USE OF THE PROGRAM

The program used is included in the APPENDIX. Except for input and output, there are four major operations in the program which must be carried out.

First, one must calculate the standard values for the lower limit (BOLIM) and the upper limit (UPLIM) of variability of the time ratios and the color yield ratios (CORA) for particular amino acids. Eqns. I, II, and III provide these values using data obtained from at least one standard run.

Secondly, there is the identification or qualitative analysis of the peaks of a particular chromatogram. Identification is carried out by using eqn. IV to obtain the time ratio of each peak and by determining whether or not this particular time ratio falls within the range for one of the amino acids of the standard chromatogram.

$$\text{RATIO} = \frac{\text{TIME(I)}}{\text{TNOR}} \qquad \text{(IV)}$$

where

RATIO = elution time ratio.

TIME(I) = elution time in minutes of particular amino acid peak in run under investigation.

TNOR = elution time in minutes of internal standard, norleucine, in run under investigation.

Thirdly, after positive identification of an amino acid peak, calculation of concentration (CONC) is carried out according to eqn. V.

$$\text{CONC} = \frac{\text{AREA(I)} \times \text{CORA(J)}}{\text{ANOR} \times \text{SVOL}} \qquad \text{(V)}$$

where

CONC = concentration of a particular identified amino acid in micromoles per milliliter.

AREA(I) = area of particular identified amino acid in integrator counts.

CORA(J) = CORA(I) as given by eqn. III and is equal to the color yield ratio of a particular identified amino acid.

ANOR = area of internal standard norleucine in chromatogram under investigation in integrator counts.

SVOL = sample volume in milliliters (not including any volume due to internal standard or sucrose).

Fourthly, after having obtained the concentration for an amino acid peak, the number of amino acid residues can be calculated for a particular protein or polypeptide. This is done by application of eqn. VI.

$$\text{RESID} = \frac{\text{CONC}}{\text{PROMO}} \qquad \text{(VI)}$$

where

RESID = number of residues of particular amino acid per mole of polypeptide or protein under analysis.

PROMO = micromoles protein or polypeptide added = weight of protein or polypeptide used for the hydrolysate added as sample divided by molecular weight of protein or polypeptide.

CONC = concentration of particular identified amino acid as given by eqn. V.

Since the number of residues is always an integer, the values obtained by the above calculation have to be rounded up or down, whichever integer is closer. In case an amino acid analysis of a non-homogeneous sample such as urine, serum or tissue hydrolysate is to be carried out, the punched cards containing eqn. VI and the related input and output statements must be removed from the program. It is important to point out that these cards may be left in the program, and one may use PROMO for introduction of standards of reference such as dry weight, creatinine, or DNA concentration.

After reading-in of the program itself the data are presented for acceptance by the computer in the following sequence:

(1) Micromoles of amino acid in standard chromatogram (SCONC).

(2) Time and area of all particular amino acid peaks in standard chromatogram pair by pair [STIME(I), SAREA(I)].

These values need to be newly determined only when a new ninhydrin solution or a new buffer solution for one or more of the autograd chambers has been made. We determine new values after 32 runs since all mixtures are made at the same time for that many analyses.

(3) Percentages of time ratio of particular amino acid which represent the lower and upper limit of variation; pair by pair [BPER(I), UPER(I)].

These values should be constant for a particular amino acid analyzer system as long as column temperature, buffer gradient, and other factors are kept constant which influence the separation properties of the system. We have never had to change them. If they have to be newly determined, chromatography of mixtures with various amino acid concentrations must be performed.

(4) Elution time and area of peak of the internal standard, norleucine; the total number of peaks and the sample volume of the chromatogram under investigation (TNOR, ANOR, IMAX, SVOL).

(5) Elution time and area of each peak on the chromatogram; pair by pair [TIME(I), AREA(I)].

(6) Micromoles protein or polypeptide added (PROMO).

For each identified peak the results are printed out as the peak number, peak time in minutes, time ratio, name of the amino acid, concentration in micromoles per milliliter applied sample, and number of residues per mole hydrolyzed protein. If a peak

TABLE III

DATA OUTPUT FOR PORCINE INSULIN HYDROLYSATE

Peak number	*Peak time (min)*	*Time ratio*	*Amino acid*	*Concentration (μmoles/ml)*	*No. of residues*
1	188.2	0.294984	Aspartic acid	0.43829	2.90
2	207.2	0.324764	Threonine	0.23674	1.56
3	218.0	0.341692	Serine	0.31496	2.08
4	264.3	0.414263	Glutamic acid	1.12284	7.43
5	274.3	0.429937	Proline	0.14492	0.95
6	341.5	0.535266	Glycine	0.57577	3.81
7	364.4	0.571159	Alanine	0.29171	1.93
8	454.5	0.712382	Valine	0.58452	3.87
9	517.3	0.810815	Cystine	0.67052	4.44
10	595.4	0.933228	Isoleucine	0.25581	1.69
11	617.0	0.967084	Leucine	0.90996	6.02
12	638.0	1.000000	Norleucine	—	—
13	675.0	1.057993	Tyrosine	0.44860	2.97
14	693.1	1.086363	Phenylalanine	0.41619	2.75
15	746.7	1.170376	Ammonia	—	—
16	911.4	1.428526	Lysine	0.13152	0.87
17	950.7	1.490125	Histidine	0.23220	1.53
18	1120.6	1.756426	Arginine	0.10730	0.71

TABLE IV

MOLAR AMINO ACID RATIOS OF PORCINE INSULIN

Amino acid	*Numbers of residues obtained by means of computer program (average of two analyses)*	*Numbers of residues obtained rounded to nearest integer and corrections*	*Theoretically expected numbers of residues*
Aspartic acid	2.93	3	3
Threonine	1.60	2	2
Serine	2.03	2 + 1[a]	3
Glutamic acid	7.43	7	7
Proline	0.88	1	1
Glycine	3.86	4	4
Alanine	1.97	2	2
Valine	3.96	4	4
Cystine	4.72	5 + 1[b]	6
Isoleucine	1.78	2	2
Leucine	6.02	6	6
Tyrosine	3.03	3 + 1[a]	4
Phenylalanine	2.74	3	3
Lysine	0.90	1	1
Histidine	1.64	2	2
Arginine	0.76	1	1

[a] Correction due to hydrolysis loss.
[b] Correction includes loss due to hydrolysis and small amounts of cysteic acid.

has not been identified, only peak number, peak time, and time ratio are printed out and the statement "this is an unidentified peak" appears. Concentrations are not printed out for the internal standard norleucine and for ammonia.

The output for the described program is shown in Table III. These data were obtained from hydrolyzed porcine insulin. As shown in Table IV, the rounded values for the numbers of amino acid residues are in agreement with the data published by BROWN *et al.*[8] after correction for hydrolysis loss of tyrosine, cystine, and serine as proposed by NOLTMANN *et al.*[9].

DISCUSSION

The finding that the elution time of different amino acids depends also on the concentration of the amino acids themselves in the sample mixture explains why desalting and other preparatory procedures alone do not lead to a time ratio which is characteristic for a particular amino acid in different samples. Therefore, it is difficult to identify automatically peaks on the amino acid analyzer chromatograms. A solution to this problem would be of great help in analysis of biological samples, such as urine, where one encounters significant variability in amino acid excretion patterns which may depend on a number of factors including the state of hydration, fluid, food and drug intake. For a limited number of amino acids, however, we have been able to determine the range of elution time variations as an instrumental constant of the amino acid analyzer. Since common proteins contain only a limited number of amino acids, our system works, for instance, with the protein hydrolysates we have studied. Another application of the system is screening of human serum samples for amino acids which are known to be abnormal quantitatively or qualitatively in certain inborn errors of metabolism. In these disorders one deals again with a limited number of amino acids for which standard reference parameters can be easily determined and introduced into the computer program.

Amino acids for which no standard reference values have been fed into the computer will always be called "unidentified". In some instances this may mean for protein hydrolysates that the hydrolysis has not been complete. The unidentified peak may then be due to a peptide since the reference standard values for all possible amino acids of protein hydrolysates have been stored.

Correction factors for loss of amino acids in hydrolysis have not been taken into consideration in the computer program described since it is recognized that the losses can vary from protein to protein (depending on its size, structure, and concentration and on the hydrolysis procedure applied). Thus, correction for these losses must be determined for each particular protein preparation.

ACKNOWLEDGEMENTS

This work was supported by Project No. 236 of the Social and Rehabilitation Service, Department of Health, Education, and Welfare, a General Research Support Award (5 SO1-FR-05406) from the National Institutes of Health, and in part by a Research Career Development Award (5-K3-AM-5058) from the National Institute of Arthritis and Metabolic Diseases (to GKS). We thank KARL BLAU and CAROLYN CANNADY for assistance in this work.

REFERENCES

1 D. H. SPACKMAN, W. H. STEIN AND S. MOORE, *Anal. Chem.*, 30 (1958) 1190.
2 W. L. PORTER AND E. A. TALLEY, *Anal. Chem.*, 36 (1964) 1962.
3 A. YONDA, D. L. FILMER, H. PATE, N. ALONZO AND C. H. W. HIRS, *Anal. Biochem.*, 10 (1965) 53.
4 M. T. KRICHEVSKY, G. SCHWARZ AND M. MAGE, *Anal. Biochem.*, 12 (1964) 94.
5 W. W. STARBUCK, C. M. MAURITZEN, C. MCCLIMANS AND H. BUSCH, *Anal. Biochem.*, 20 (1967) 439.
6 K. OZAWA AND S. TANAKA, *Anal. Biochem.*, 24 (1968) 270.
7 S. MOORE AND W. H. STEIN, in S. P. COLOWICK AND N. O. KAPLAN (Editors), *Methods in Enzymology*, Vol. VI, Academic Press, New York, 1963, p. 819.
8 H. BROWN, F. SANGER AND R. KITAI, *Biochem. J.*, 60 (1955) 556.
9 E. A. NOLTMANN, T. A. MAHOWALD AND S. A. KUBY, *J. Biol. Chem.*, 237 (1962) 1146.

APPENDIX

Computer program for analysis of amino acid analyzer chromatograms by IBM 1130

```
PAGE    1

// JOB
// FOR
*EXTENDED PRECISION
*ONE WORD INTEGERS
*IOCS(CARD,1132 PRINTER, KEYBOARD, TYPEWRITER)
*LIST ALL
C       COMPUTER ANALYSIS OF AMINO ACID MIXTURES
C       R.E. EXSS AND H.D. HILL     AUGUST, 1968
        DIMENSION STIME(22), SAREA(22), BOLIM(22), UPLIM(22), TIME(30), AR
       *EA(30), CORA(22), BPER(22), UPER(22)
        READ (2,112) SCONC
  112 FORMAT (34X, F5.3)
        READ (2,100) (STIME(I), SAREA(I), I=1, 22)
  100 FORMAT (F7.1, 4X, F10.0)
        READ (2,106) (BPER(I), UPER(I), I=1, 22)
  106 FORMAT (F5.4, 3X, F7.4)
        DO 96 I=1, 22
        BOLIM(I) = BPER(I) * STIME(I)/STIME(14)
        UPLIM(I) = UPER(I) * STIME(I)/STIME(14)
        CORA(I) = SCONC * SAREA(14)/SAREA(I)
   96 CONTINUE
   92 WRITE (3,101)
  101 FORMAT ('1',//////////)
        READ (2, 102) TNOR, ANOR, IMAX, SVOL
  102 FORMAT (F7.1, 4X, F10.0, 5X, I3, 5X, F5.3)
        READ (2,104) (TIME(I), AREA(I), I=1, IMAX)
  104 FORMAT (F7.1, 4X, F10.0)
        READ (2, 114) PROMO
  114 FORMAT (34X, F6.5)
        WRITE (3,105)
  105 FORMAT ('+', 'PEAK NUMBER', 10X, 'PEAK TIME', 10X, 'TIME RATIO',
       *10X, 'AMINO ACID', 10X, 'CONCENTRATION', 9X, 'RESIDUES', //)
        DO 91 I=1, IMAX
        RATIO = TIME(I)/TNOR
        WRITE (3,103) I, TIME(I), RATIO
  103 FORMAT ('0', 4X, I3, 16X, F6.1, 13X, F8.6)
        DO 2 J=1, 22
        IF (RATIO - BOLIM(J)) 67,12,12
   12 IF (RATIO - UPLIM(J)) 15,15,9
    9 IF (J - 21) 2,67,67
    2 CONTINUE
   15 GO TO (93,11,14,18,21,24,27,30,33,36,39,42,45,48,50,53,56,98,59,62
      *,65,67), J
   67 WRITE (3,10)
   10 FORMAT ('+', 60X, 'THIS IS AN UNIDENTIFIED PEAK')
        GO TO 91
   93 WRITE (3,97)
   97 FORMAT ('+', 60X, 'HYDROXYPROLINE')
        GO TO 88
   11 WRITE (3,68)
   68 FORMAT ('+', 60X, 'ASPARTIC ACID')
        GO TO 88
   14 WRITE (3,69)
   69 FORMAT ('+', 60X, 'THREONINE')
```