A Computer Programme for the Calculation of the Results of Amino Acid Analysis

By G. N. GRAHAM AND B. SHELDRIK

The Astbury Department of Biophysics, University of Leeds

(Received 7 December 1964)

1. An ALGOL computer programme is presented for the calculation of the results of amino acid analysis. 2. The suggestion is made that a similar approach could be made to other tedious and repetitive biochemical calculations.

The introduction of automatic methods of amino acid analysis that followed the work of Spackman, Stein & Moore (1958) has considerably increased the amount of calculation necessary in a working day. The latest developments by workers, e.g. Thomson & Miles (1964), in this field appear to be concentrated on decreasing the time required for analysis to about 3 hr. Theoretically this will enable seven or eight analyses/day to be performed, compared with the one analysis/4 days possible with manual methods, e.g. the semi-micro method of Eastoe (1961).

The calculation of results of such analyses can be tedious, laborious and, if prolonged, subject to a variable amount of human error. These calculations consist in performing the same manipulations with different sets of data. An electronic digital computer is suitable for dealing with this type of calculation, and accordingly a computer programme has been written to carry out these manipulations. The programme was first written in Pegasus Extended Autocode and has been used successfully for the past year on a Ferranti Pegasus Machine. Recently, after the installation of an English Electric-Leo Marconi Co. KDF9 computer in the Leeds University Computing Laboratory, the programme has been rewritten in ALGOL, a standard computer language (Naur, 1963). This programme may be run on any computer for which an ALGOL compiler is available, with small modifications rendered necessary by variations of Library routines used by different machines.

EXPERIMENTAL AND RESULTS

The amino acid analyses were performed on a Technicon Automatic Amino Acid Analyser (Tech-

nicon Instruments Co. Ltd.), with norleucine as an internal standard.

The height and width (in numbers of dots) at half height of the peaks on the strip chart are determined by hand and subsequently punched on an input-data tape. After checking the print-out no further mathematical manipulations are required of the operator. The computer issues the following results for each amino acid: μ moles, μ g. of nitrogen, residues/1000 residues, percentage nitrogen of total nitrogen off the column. The total nitrogen off the column is also calculated and printed.

Any number of peaks may be integrated and calculated. If the technique does not require the use of an internal standard the insertion of unity against area and concentration of norleucine will eliminate this part of the calculation.

The programme, given in full in Table 1, requires the specification of the following for each set of data: (1) the serial number of the data (analysis identification number); (2) the total number of amino acids (number of peaks to be integrated and calculated including zero peaks); (3) the area of norleucine peak (height $\times$ no. of dots above half height); (4) the amount of norleucine (the internal standard) is no. of μ moles put on column; and for each amino acid: (1) an identification number; (2) the height of the peak; (3) the number of dots above half height; (4) the norleucine equivalent (the no. of μ moles of any particular amino acid producing the same colour density as 1 μ mole of norleucine).

The data must be terminated by the character '0;'.

The computer performs the following calculation for each peak for which data are presented:

$$\frac{\text{No. of dots} \times \text{ht. of peak} \times \text{norleucine equiv.}}{\times \text{norleucine concn.}} \\ \text{Area of norleucine peak}$$

Table 1. *An ALGOL computer programme for the calculation of the results of amino acid analysis*

```

begin comment This programme calculates micromoles, residues/1000,
  micrograms of nitrogen, total nitrogen and percentage nitrogen for each
  of the 20 amino acids produced when a protein is hydrolysed;
library A0,A6,A14,A15; real nlarea, nlconc; integer data, total, stylef;
open (10); open (20); copy text (20,10,[,]);
stylef:=format ([ndddd,]); input: data:= read (20);
if data=0 then go to stop;
write text (10,[2c] Data);
write (10, stylef, data);
total:=read(20); nlarea:=read(20); nlconc:=read(20);
  begin real height, dots, nlequiv, residues, nit, totalnit, percent, totalmmoles;
  integer listnumber, stylea, styleb, stylec, styled, stylee;
  array micromoles[1:total];
  totalmmoles:=0;
  stylea:=format([nd,]);
  styleb:=format([sssd.ddd,c]);
  write text(10,[2c]Micromoles[2c]);
  for listnumber:=1 step 1 until total do
  begin; if listnumber  $\neq$  read(20) then go to stop;
    height:=read(20); dots:=read(20); nlequiv:=read(20);
    micromoles[listnumber]:=height $\times$ dots $\times$ nlequiv $\times$ nlconc/nlarea;
    totalmmoles:=totalmmoles+micromoles[listnumber];
    write(10,stylea,listnumber);
    write(10,styleb,micromoles[listnumber]);
  end;
  write text(10,[2c]Residues*per*1000[2c]);
  stylec:=format([ssndd.dd,c]);
  for listnumber:=1 step 1 until total do
  begin; write(10,stylea,listnumber);
    residues:=micromoles[listnumber] $\times$ 1000/totalmmoles;
    write(10, stylec, residues);
  end;
  micromoles[18]:=micromoles[18] $\times$ 2;
  micromoles[19]:=micromoles[19] $\times$ 3;
  micromoles[20]:=micromoles[20] $\times$ 4;
  totalnit:=0;
  styled:=format([sssd.dd,c]);
  write text(10,[2c]Nitrogen*in*micrograms[2c]);
  for listnumber:=1 step 1 until total do
  begin;
    write(10, stylea, listnumber);
    nit:=micromoles[listnumber] $\times$ 14;
    write(10, styled, nit);
    totalnit:=totalnit+nit;
  end;
  write text (10,[2c]Total*nitrogen*in*micrograms[2c]);
  stylee:=format([nd.ddd,c]);
  write(10, stylee, totalnit);
  write text(10,[2c]Percentage*nitrogen[2c]);
  for listnumber:=1 step 1 until total do
  begin;
    write(10, stylea, listnumber);
    percent:=micromoles[listnumber] $\times$ 14 $\times$ 100/totalnit
    write(10, stylee, percent);
  end;
stop: close(10); close(20);
end
→

```

Table 2. *Input data and output results for the calculation by computer of the results of an amino acid analysis of a cat-actin hydrolysate*

The protein was hydrolysed at 100° for 24hr. with 5.7 N-HCl.

		Input data				Output results		
		No.	Peak height	No. of dots	Norleucine equivalent	μ moles	Residues/1000 total residues	Nitrogen (μ g.) (%)
Run no.	65;							
No. of amino acids	20;							
Area of norleucine	7.45;							
Concn. of norleucine	0.09848;							
Cysteic acid		1	0.108;	18.8;	0.99;	0.0266;	10.84;	0.37; 0.884;
Methionine sulphone		2	0.180;	18.1;	1.02;	0.0439;	17.91;	0.61; 1.462;
Aspartic acid		3	0.920;	18.7;	0.986;	0.2242;	91.45;	3.14; 7.464;
Threonine		4	0.672;	19.2;	0.875;	0.1492;	60.86;	2.09; 4.968;
Serine		5	0.617;	19.5;	0.849;	0.1350;	55.07;	1.89; 4.495;
Glutamic acid		6	0.822;	24.8;	0.938;	0.2528;	103.08;	3.54; 8.414;
Proline		7	0.076;	25.0;	6.36;	0.1597;	65.14;	2.24; 5.317;
Glycine		8	0.642;	26.6;	0.868;	0.1959;	79.91;	2.74; 6.522;
Alanine		9	0.809;	29.0;	0.955;	0.2962;	120.78;	4.15; 9.859;
Valine		10	0.210;	34.9;	0.994;	0.0963;	39.27;	1.35; 3.205;
Cystine (half)		11	0;	0;	0;	0.0000;	0.00;	0.00; 0.000;
Methionine		12	0.135;	37.4;	0.912;	0.0609;	24.82;	0.85; 2.026;
Isoleucine		13	0.294;	33.0;	0.986;	0.1265;	51.57;	1.77; 4.209;
Leucine		14	0.470;	32.6;	0.872;	0.1766;	72.03;	2.47; 5.879;
Tyrosine		15	0.239;	32.7;	0.909;	0.0939;	38.30;	1.31; 3.126;
Phenylalanine		16	0.191;	33.6;	0.985;	0.0836;	34.08;	1.17; 2.781;
Ornithine		17	0.054;	58.0;	1.00;	0.0414;	16.88;	0.58; 1.378;
Lysine		18	0.438;	27.0;	0.828;	0.1294;	52.79;	3.62; 8.617;
Histidine		19	0.144;	33.6;	0.893;	0.0571;	23.29;	2.40; 5.703;
Arginine		20	0.206;	43.3;	0.872;	0.1028;	41.93;	5.76; 13.690;
		0;					Total	42.059;

and punches out on an output tape the following details:

- Identification number of amino acid (1-total); μ moles;
- Identification number of amino acid (1-total); residues/1000;
- Identification number of amino acid (1-total); nitrogen (μ g.);
- Total nitrogen (μ g.);
- Identification number of amino acid (1-total); percentage of total nitrogen;

This programme has been tested many times and the results were compared with those obtained (a) by hand calculation and (b) by the Autocode programme, to confirm that it is accurate. An example of a data input and results output is given in Table 2.

DISCUSSION

A computer programme of this type eliminates human error from the calculations, makes the results available more quickly and enables full use

to be made of the available data. Human error is not eliminated completely, as the input tape is still prepared by hand and this must be checked to avoid errors. If a digital print-out were to be fitted to the recorder and the programme modified, the output, in the form of a tape, could then be fed directly into the computer, as has been reported by Hirs (1964).

Other computer programmes for calculation of results from experimental data have been written by one of us (B. S.) and some work has been carried out on programmes and techniques involved in the investigation of amino acid sequences in proteins (Dade, Graham & Sheldrick, 1965) along lines indicated by Bernhard, Bradley & Duda (1963). It is clear that many other tedious calculations in the biochemical field could be treated in a similar manner.

We thank Miss Elizabeth Deall for technical assistance and the Electronic Computing Laboratory of the University of Leeds for providing computing facilities. The work was supported by the Muscular Dystrophy Associations of

America Inc. (to G. N. G.) and the U.S. National Institute of Public Health Grant no. GM07399 (to B. S.).

Dade, E. F., Graham, G. N. & Sheldrick, B. (1965). *Proc. Technicon 25th Anniversary International Symposium* (in the Press).

Eastoe, J. E. (1961). *Biochem. J.* **79**, 652.

Hirs, C. H. W. (1964). *Annu. Rev. Biochem.* **33**, 598.

Naur, P. (ed.) (1963). *Comput. J.* **5**, 349.

Spackman, D. H., Stein, W. H. & Moore, S. (1958). *Analyt. Chem.* **30**, 1190.

Thomson, A. R. & Miles, B. J. (1964). *Nature, Lond.*, **203**, 483.

REFERENCES

Bernhard, S. A., Bradley, D. F. & Duda, W. L. (1963). *IBM J. Res. Dev.* **7**, 246.