Journal of Chromatography, 89 (1974) 61–72

CHROM. 7103

PROGRAMS FOR USE WITH THE AUTOMATIC AMINO ACID ANALYSER TO IDENTIFY, COMPUTE AND CORRELATE AMINO ACID CONCENTRATIONS IN BIOLOGICAL SAMPLES

MICHAEL A. FOX

Department of Biomathematics, City of Hope National Medical Center, Duarte, Calif. 91010 (U.S.A.) and University of California at Los Angeles, Calif. 90024 *(U.S.A.)*

(Received October 8th, 1973)

SUMMARY

A series of computer programs is presented that enable chromatograms obtained from different machines run under a variety of analytical conditions to be processed. Many chromatograms may be examined per run to establish the identity of the amino acids and determine their concentrations. Using tissue analysis as an example, those chromatograms obtained from the same specimen but passed through different columns are combined to provide a profile for that tissue. Tissues are further grouped into sets of experiments which are jointly compared, and from the results, questions are formulated to aid in the elucidation of the biochemical changes occurring in tissues under various conditions.

INTRODUCTION

The rate of acquisition of data from such machines as automatic amino acid analysers can at times prove to be an embarrassment unless adequate facilities are available to process this information. Partial solutions to this imbalance in operations are available which, though an improvement, have not eliminated the race with the machine that incompletely resolved data promotes.

The integration of the chromatogram curve is one of those problems that has attracted some interest and for this a solution is to be found using an electronic integrator or a computer program[1,2]. Association of the peaks in the chromatogram to amino acids is another requirement and this has been accomplished to a limited extent in the more modern analysers. Additionally, computer programs have been published[3,4] that compute the concentrations of those amino acids recognized. Present techniques are, however, insufficient for a variety of reasons. Some of these outstanding problems are when unknown or unexpected peaks occur, when a series of samples is run over a period of time during which the decay of the ninhydrin reagent becomes a significant factor and when the results from more than one chromatogram must be combined to determine the concentrations of amino acids present in a tissue. Furthermore, existing methods present the biochemist with a vast number of experimental findings measured in terms of amino acid concentrations. This and similar information often requires statistical analysis. To inhibit this growth of un-

digested information and to reunite the researcher with the significance of his experiments, a set of computer programs that embrace something of the total experimental environment in which amino acids are involved has been written.

COMPUTER PROGRAMS

Dividing the system into sections, the first deals with the identification of amino acids contained within the chromatogram and the computation of their concentrations. The second is concerned with the grouping of chromatograms derived from different samples of the same tissue. The third program groups similar tissues subject to like treatments into experimental sets and thence performs statistical comparisons between experiments. The fourth program examines the differences between these experiments and allows formulation and solution of pertinent questions.

The programs have been designed so that they may be used together or as individual modules. When used as a set the amount of data is reduced at each stage with a corresponding increase in the information derived. Thus, many experiments using different types of machines or even other information can be combined at different stages within the information flow. The programs, though complex in character, are simple to run, each providing a full description of the tasks performed together with any errors or inconsistencies that are found.

A chromatogram is here defined as a set of pairs of numbers, the retention time and the integrated optical density time function (referred to as counts). The column through which the liquid carrying the amino acid mixture is passed is referred to as a physical column. Since the system caters to samples in which the identity and number of amino acids may be only partially known and as these amino acids are to be identified with respect to standards, the concept of a higher order standard or a theoretical standard is introduced. This has three functions, primarily it acts as a consistency check on the standard being run and secondly it allows the representation of the column in the program to be tuned to a particular amino acid eluted at a time when more than one amino acid may be eluted. The third use for this concept is to afford a mechanism that accounts for phenomena peculiar to a particular run, for example, time dilation which is discussed later. To each such column is associated a set of times corresponding to all the peaks that this type of column expects whether they be generated by amino acids or other substances. It is this set of values that is the theoretical standard, though its establishment may well be by experiment. Currently the program contains seven such columns. It is immaterial how the columns are distributed between machines. There may be seven different machines with identical columns or one machine with seven different columns and so forth. Since the first program can accept chromatographic data from several sources using a variety of columns, each unit of information, which itself may be a set, is a member of a set. Comprehension of the first program may be enhanced by thinking of each set as a singleton.

Normally amino acid analyses are run sequentially, that is, a standard is passed through a particular column followed by a series of samples and finally a re-run of the original standard. This series of runs is referred to as a chromatographic partial set. The set becomes complete when all the amino acids in the samples can be identified, which may require the inclusion of additional supplemental standards. The program

can conveniently accommodate 100 chromatograms per run, each of which may contain up to 30 amino acids. There is no restriction on the number of complete sets that can be constructed from these 100 chromatograms. The only restriction (for age correction) is that a sequential enumeration of the sample chromatograms (allowing gaps) is required. Since many different complete sets can be accepted, there is no restriction in the order in which the chromatograms are presented to the computer. Furthermore, the same physical column can be used with many such sets. A collection of concentration sets, each member of which contains acceptable amino acid names, together with the concentrations of these amino acids is also required.

Idiosyncratic behavior, as well as intentional manipulation, of the automatic analyser is accommodated by allowing for correction of such things as baseline shift, time drift and count changes. These are applied on a piece-wise basis by specifying the chromatogram, the type of correction and the time at which it is to be applied. The phenomenon of time dilation, that is stretching of the time axis of the chromatogram, induced by different concentrations of amino acids, can be accounted for either by the range that an amino acid is expected to be eluted or by functionally changing the theoretical standard. This function will generally be the addition of a constant to each time after a particular time (that is, after the elution of a particular amino acid) or by the addition of an element that itself increases linearly with time. Although a complete history is kept of the manipulations applied to each chromatogram, it is preferable to adjust the theoretical standard rather than each individual chromatogram. It is then the interpretation that is placed, by the investigator, on the retention times which is changed and not the retention times of the samples themselves.

Trimming or filtering is afforded to standards, this normally consists of removal of the tail and spurious peaks from the chromatogram. Such filtering is not extended to samples for fear of systematically excluding amino acids present in small quantities.

EXPERIMENTAL AND RESULTS

The capabilities of the system are best described by an example and except for omitting the numerous error messages, all the manipulations extended to a particular tissue will now be described in terms of chromatogram 339. This and similar samples were passed through a 'long column' (denoted by DC1ANA3) using a sodium buffer. Chromatogram 339 contained six peaks not present in the standard. A series of partial standards was therefore sequentially analysed to determine the identities of these unknown amino acids. The chromatograms used for this identification are shown in Fig. 1 as numbers 337, 350, 464, 697 and 933. Whether these additional compounds are run as individual chromatograms or together is immaterial to the program. Indeed, these numbers indicate that they were run well after the sample. These chromatograms together with any between numbers 338 and 361 constitute a complete chromatographic set. Fig. 1 displays this information, the prime or non-augmented standards are shown as a set of retention times and counts without the amino acid names. This is because, at this point in the processing, the retention times and their corresponding names have not been confirmed by the theoretical standard. The list of concentrations for all the amino acids contained

```
FOR STANDARD TYPE DC1ANA3D THE FOLLOWING CHROMATOGRAMS HAVE BEEN USED
STANDARDS    336   362   337   350   464   697   933
SAMPLES  338 TO  361
NOISE FILTERED CHROMATOGRAM NUMBER  336
            TIME      COUNT

              22       410578
              28       223490
              32       357837
              40       414170
              48       414462
              51       437341
              54       432823
              63        69808
              69       140078
              74       410041
              77       410583
              83       428992
              89       414527
              93       439414
              97       426397
             109       414728
             116       410992
             123       421432
             129       486552
             149       361258
             190       271830
NOISE FILTERED CHROMATOGRAM NUMBER  362
              22       398414
              28       215398
              32       355955
              40       407718
              48       400360
              51       427981
              54       425288
              63        70316
              69       188950
              74       397798
              77       404814
              83       420149
              89       402990
              93       427240
              97       412808
             109       395570
             116       390155
             123       394051
             129       464292
             149       362470
             187       397942
```

Fig. 1. A complete chromatographic set together with the prime standards. At this stage each standard is unresolved and is represented by a set of pairs of retention times and counts.

within the composite standard (that is, prime standards plus additional standards) were found in concentration set D, hence the designation 'STANDARD TYPE DC1ANA3D'.

The pre- and post-chromatograms were first filtered and then compared for consistency. They must both have the same number of pairs and corresponding retention times and these must not differ by more than a fixed amount (3 min). The additional standards are then filtered and should they contain retention times not present in the main standard, these times and counts are included to form a composite standard (Fig. 2). The names of the amino acids contained in concentration set D were compared with the theoretical standard established for column DC1ANA3 and theoretical retention times were associated with these names (Fig. 3*). In the

* In Figs. 3–11 and in the text, the following non-standard abbreviations are used: CITRU = citrulline; $CYSO_3H$ = cysteic acid; CYSTAT = cystathionine; ETAM = ethanolamine; GABA = γ-aminobutyric acid; GAM = glucosamine; G-P-E = glycerophosphoryl ethanolamine; HCAR = homocarnosine; O-P-E = phosphoethanolamine; TAU = taurine.

```
NOISE FILTERED CHROMATOGRAM NUMBER  337
          126       264869
          149        91865
RETENTION TIME 126 INSERTED INTO CHROMATOGRAM  336
NOISE FILTERED CHROMATOGRAM NUMBER  350
           59       201309
          149        38191
RETENTION TIME  59 INSERTED INTO CHROMATOGRAM  336
NOISE FILTERED CHROMATOGRAM NUMBER  464
          112       215836
          150        51560
          195       372079
RETENTION TIME 112 INSERTED INTO CHROMATOGRAM  336
NOISE FILTERED CHROMATOGRAM NUMBER  697
          106       232551
          152       189427
          196       190142
RETENTION TIME 106 INSERTED INTO CHROMATOGRAM  336
RETENTION TIME 152 INSERTED INTO CHROMATOGRAM  336
NOISE FILTERED CHROMATOGRAM NUMBER  933
           25        26277
RETENTION TIME  25 INSERTED INTO CHROMATOGRAM  336
```

```
THE FOLLOWING ARE THE REPORTED CONSTITUENTS FOR          THIS STANDARD
        AMINO ACID          CONCENTRATION
          CYSO3H                100.120
          O-P-E                 101.400
          TAU                   101.000
          ASP                   100.000
          THR                   100.000
          SER                   100.000
          GLU                   100.000
          PRO                   100.000
           -CYS                 100.000
          GLY                   100.000
          ALA                   100.000
          VAL                   100.000
          MET                   100.000
          ILE                   100.000
          LEU                   100.000
          TYR                   100.000
          PHE                   100.000
          HIS                   100.000
          LYS                   100.000
          NH3                   100.000
          ARG                   100.000
          CITRU                  50.120
          GAM                    52.440
          ORN                    59.030
          GABA                   50.000
          G-P-E                  46.300
```

Fig. 2. Additional standards. Counts of those times not occurring in the prime standard are merged with the prime standard.

Fig. 3. Concentration set D.

theoretical standard, the times 82, 83, and 84 min correspond to buffer change one, cysteic acid, and valine (not illustrated). This means that at around 83 min a change of buffer may be expected and either valine or cysteic acid can be expected. This column is tuned at this stage to recognize valine as the appropriate amino acid for this series by including a concentration for valine but excluding one for cysteic acid in the concentration set D (Fig. 3). The effect of this on the sample will be seen later. These theoretical times are now matched using closest neighbor, unique position, and bounded difference criteria with the retention times of the pre-composite standard, thereby producing a resolution between the names of the amino acids and the times at which this standard has established that they are eluted. It is seen that valine is expected at 83 min (Fig. 4). The retention times of the sample chromatogram are

RESOLUTION BETWEEN REPORTED CONSTITUENTS AND RETENTION TIME WITH ABSOLUTE STANDARD

AMINO ACID	TIME	COUNTS
CYSO3H	22	410578
G-P-E	25	26277
O-P-E	28	223490
TAU	32	357837
ASP	40	414170
THR	48	414462
SER	51	437341
GLU	54	432823
CITRU	59	201309
PRO	63	69808
-CYS	69	140078
GLY	74	410041
ALA	77	410583
VAL	83	428992
MET	89	414527
ILE	93	439414
LEU	97	426397
GABA	106	232551
TYR	109	414728
GAM	112	215836
PHE	116	410992
HIS	123	421432
ORN	126	264869
LYS	129	486552
NH3	149	361258
NH3	152	189427
ARG	190	271830

Fig. 4. Resolution between theoretical and first standard. This ensures that the standard run is acceptable.

in turn matched against the first or composite pre-standard (again using the closest neighbor, unique occupancy, and bounded difference criteria). If any two of these three comparisons yield the same name it is assumed that this name identifies that amino acid at this time. In sample 339 at 81 and 83 min, both standards recognize these as valine yet the theoretical standard recognizes these as buffer change one and cysteic acid (Fig. 5). These are alternate possibilities. But from other criteria it is known that the first valine is actually brought about by a buffer change while the second is the true valine. The mechanism of accepting the second valine over the

SAMPLE CHROMATOGRAM 339 FOUND

COMPARE 1ST STD	COMPARE LST STD	COMPARE ABS STD	RETENTION TIME	COUNT	CONCENTRATION
CYSO3H	CYSO3H	CYSO3H	23	8034	1.97
G-P-E	G-P-E	G-P-E	25	13881	24.46
O-P-E	O-P-E	O-P-E	28	164079	74.76
TAU	TAU	TAU	32	1153311	325.72
ASP	ASP	ASP	40	80419	19.45
THR	THR	THR	48	59311	14.37
SER	SER	SER	50	77386	17.74
GLU	GLU	GLU	53	115024	26.63
CITRU	CITRU	CITRU	59	2014	0.50
PRO	PRO	PRO	62	54150	77.50
GLY	GLY	GLY	74	1008981	246.92
ALA	ALA	ALA	77	264031	64.41
VAL	VAL	BCH1	81	13490	3.15
VAL	VAL	CYSTAT	83	19797	4.63
MET	MET	MET	89	3430	0.83
LEU	LEU	LEU	99	3893	0.92
		BCH2	102	5557	0.56
GABA	GABA	BCH3	104	7127	1.53
HIS	HIS	HIS	123	7170	1.71
ORN	ORN	ORN	126	4405	0.98
LYS	LYS	LYS	129	15567	3.22
NH3	NH3	ETAM	149	146657	40.58

Fig. 5. Sample chromatogram 339. The amino acids contained in this sample are recognized by comparison with the first, last and theoretical standard and their concentrations computed. BCH= buffer change.

first is deferred to the description of the next program. A similar situation occurs with GABA and buffer change three.

Using the enumeration number of the chromatogram as the independent variable, an interpolation between the count of the pre- and post-standard is made for each identified amino acid and from this the concentration of the amino acid is determined. It is assumed that all chromatograms belonging to a particular series take about the same length of time to process and that the ninhydrin reagent ages uniformly throughout this time (and any other factor such as packing occurs uniformly with use). A linear interpolation is then both valid and results in a significant correction. The example shows differences between the pre- and post-standards and whatever interpretation is placed on these differences, the correction applied tends to give greater weight to the standard that was run nearest in time to the particular sample. Since the additional standards are only represented once, no interpolation is applied to the amino acids they represent. This comment also applies to those situations in which only one prime standard is run.

All the chromatograms in the sample set, namely chromatograms 338 to 361, are similarly processed and the program continues with subsequent complete chromatographic sets. Together with the printout, an additional data set, which may reside on cards, tape, or private file, is made of all successfully processed chromatograms. This stipulates the column used for a chromatogram, the amino acids contained therein together with their computed concentrations. This, in combination with other such data form part of the input for the next program.

The amino acids contained in a tissue sample normally require more than one type of column for their resolution. Occasionally, a sample is passed twice through the same column using different dilutions. Additionally, a particular amino acid may be imperfectly resolved by one type of column and acceptably resolved by another. The program, whose description follows, has been designed to accommodate these various analytical techniques. This program accepts, together with output from the previous program, that information which allows the assemblage of component chromatograms and quantities such as tissue dilution factors, cell weight or volume and so forth to build a tissue profile.

Using tissue 229T as an example, this tissue was divided into two parts, the first of which was passed through a sodium buffered column (chromatogram 339, the subject of the previous example) and the other was passed through a basic physiological column (chromatogram 514). From the set of chromatograms that the program may access these two were found. The program contains certain manipulative features for both changing amino acid concentrations or selecting from a multiplicity of the same amino acid a particular one. In chromatogram 339, as illustrated in Fig. 6 compared with Fig. 5, the concentration for GABA has been set to zero. The two values for valine remain in their original order. If the first value had been desired their order would have been reversed. This is achieved by specifying the chromatogram number, the amino acid name and either a cardinal number or a concentration, the amino acid referred to with that cardinality is then used (here this could be the first valine) or the new concentration is applied (here the concentration of GABA is set to zero).

The program contains a matrix (the acceptance matrix), the rows of which are amino acid names and the columns represent the types of chromatographic

```
TISSUE 229T CONSISTS OF THE FOLLOWING CHROMATOGRAMS
CHROMATOGRAM  339 FOUND

CYSO3H          1.97
G-P-E          24.46
O-P-E          74.76
TAU           325.72
ASP            19.45
THR            14.37
SER            17.74
GLU            26.63
CITRU           0.50
PRO            77.50
GLY           246.92
ALA            64.41
VAL             3.15
VAL             4.63
MET             0.83
LEU             0.92
GABA            0.00
HIS             1.71
ORN             0.98
LYS             3.22
NH3            40.58
THE ABOVE HAVE BEEN RUN THROUGH COLUMN TYPE DC1ANA3
CHROMATOGRAM  514 FOUND

PHE             0.13
ETAM            0.75
NH3            35.56
LYS             3.14
HIS             0.93
TRP           272.22
HCAR            0.00
THE ABOVE HAVE BEEN RUN THROUGH COLUMN TYPE DC2ABPH
    THR,SER,GLU, WILL BE ACCEPTED FROM THE DC1ANA3  COLUMN
```

Fig. 6. Component chromatograms for tissue 229T. These chromatograms were obtained using different portions of the same tissue.

columns. If the amino acid is deemed resolved by a particular physical column then a one is present in the appropriate position of the matrix and if not, a zero. Thus application of the information contained within this matrix to the component chromatograms enables them to be combined. For example, lysine is resolved by both columns DC1ANA3 and DC2ABPH in an acceptable manner. The content of lysine within the tissue is then computed by taking the weighted mean of the concentrations obtained from both chromatograms, where the weights reflect any sample dilution. Unusual methods of analysis may be incorporated into the program by

SUMMARY FOR TISSUE 229T	DC1ANA3	DC2ABPH	μMOLE/ML PACKED CELLS	PERCENT CONC	CONCENTRATION OF ELEMENTS N	S	P
CYSO3H	1.97	0.00	0.0985	0.17	0.0082	0.0187	0.0000
G-P-E	24.46	0.00	1.2230	2.07	0.0793	0.0000	0.1754
TAU	325.72	0.00	16.2860	27.62	1.8226	4.1720	0.0000
O-P-E	74.76	0.00	3.7380	6.34	0.3710	0.0000	0.8205
ASP	19.45	0.00	0.9725	1.65	0.1023	0.0000	0.0000
THR	14.37	0.00	0.7185	1.22	0.0845	0.0000	0.0000
SER	17.74	0.00	0.8870	1.50	0.1182	0.0000	0.0000
GLU	26.63	0.00	1.3315	2.26	0.1268	0.0000	0.0000
CITRU	0.50	0.00	0.0250	0.04	0.0060	0.0000	0.0000
PRO	77.50	0.00	3.8750	6.57	0.4714	0.0000	0.0000
GLY	246.92	0.00	12.3460	20.94	2.3040	0.0000	0.0000
ALA	64.41	0.00	3.2205	5.46	0.5063	0.0000	0.0000
VAL	4.63	0.00	0.2315	0.39	0.0277	0.0000	0.0000
MET	0.83	0.00	0.0415	0.07	0.0039	0.0089	0.0000
LEU	0.92	0.00	0.0460	0.08	0.0049	0.0000	0.0000
PHE	0.00	0.13	0.0065	0.01	0.0006	0.0000	0.0000
ORN	0.98	0.00	0.0490	0.08	0.0104	0.0000	0.0000
ETAM	0.00	0.75	0.0375	0.06	0.0086	0.0000	0.0000
LYS	3.22	3.14	0.1590	0.27	0.0305	0.0000	0.0000
HIS	1.71	0.93	0.0660	0.11	0.0179	0.0000	0.0000
TRP	0.00	272.22	13.6110	23.08	1.8671	0.0000	0.0000
			TOTAL CONCENTRATION OF ELEMENTS		7.9722	4.1996	0.9958

Fig. 7. Profile for tissue 229T.

manipulation of the acceptance matrix. For example, threonine, serine and glutamic acid are normally resolved using a lithium buffered column but for technical reasons they were, in this example, resolved by the sodium column. This is noted in Fig. 6.

Fig. 7 illustrates the profile for tissue 229T. In addition to the contribution that each amino acid makes to this tissue the amount of the inorganic elements nitrogen, sulphur and phosphorus that are contained within each amino acid together with their totals that the tissue contains is computed. The units of measure depend upon the units in specifying the concentrations of the standard and the tissue constants. Several hundred chromatograms containing several thousand concentrations can be processed in each run of this program. There is no limitation on the number of tissues from which these chromatograms were derived, each such tissue can contain up to five component chromatograms. The names of the tissues, together with their amino acid contents are passed to the next program.

Experiments usually consist of like treatments using many samples of similar tissues. By naming an experiment and specifying the member tissues of that experiment, a summary for the experiment is obtained. Fig. 8 shows experiment C5 TRY. In this experiment, tissue 229T, the subject of the last example, is a member. That which was previously computed for the individual tissues is now presented for the experiment. In addition, the total nitrogen content is treated as a pseudo amino acid.

With the computation of many such sets of experiments, various cross-comparisons can be made. In Fig. 9, the pair-wise contrast between experiment C5 CONT and C5 TRY is presented. Using the t statistic, a single asterisk denotes significance at the 95% level whilst the double asterisk the 99% level. It will be seen in this example that leucine is significantly lower in the C5 TRY experiment than in the C5 CONT experiment. By convention, the first experiment is called the control. Introduction of pseudo amino acids at the tissue level allows any numerically describable property of the tissue (for example, the nitrogen content) to be included in an independent way

FOR SET C5 TRY THE FOLLOWING TISSUES HAVE BEEN USED

225 T
227 T
229 T
237 T

A ACID	NUM	MEAN	S.D.	PERCENT	N	S	P
CYSO3H	4	0.1244	0.0486	0.23	0.0103	0.0236	0.0000
G-P-E	4	1.0879	0.1255	2.00	0.0705	0.0000	0.1560
TAU	4	16.3198	4.2785	29.94	1.8264	4.1807	0.0000
O-P-E	4	3.4112	0.7488	6.26	0.3386	0.0000	0.7487
ASP	4	1.0037	0.1604	1.84	0.1056	0.0000	0.0000
THR	4	0.8086	0.2193	1.48	0.0951	0.0000	0.0000
SER	4	0.8939	0.2187	1.64	0.1191	0.0000	0.0000
GLU	4	1.4467	0.3895	2.65	0.1377	0.0000	0.0000
CITRU	3	0.0397	0.0307	0.07	0.0095	0.0000	0.0000
PRO	3	3.9292	1.0913	7.21	0.4780	0.0000	0.0000
GLY	4	13.2011	4.5948	24.22	2.4636	0.0000	0.0000
ALA	4	3.0920	0.7140	5.67	0.4861	0.0000	0.0000
VAL	4	0.2287	0.0477	0.42	0.0273	0.0000	0.0000
MET	3	0.0345	0.0061	0.06	0.0032	0.0074	0.0000
ILE	1	0.0215	0.0000	0.04	0.0023	0.0000	0.0000
LEU	4	0.0309	0.0272	0.06	0.0033	0.0000	0.0000
TYR	3	0.0782	0.0347	0.14	0.0060	0.0000	0.0000
PHE	4	0.0642	0.0547	0.12	0.0054	0.0000	0.0000
ORN	4	0.0960	0.0880	0.18	0.0203	0.0000	0.0000
ETAM	1	0.0375	0.0000	0.07	0.0086	0.0000	0.0000
LYS	4	0.3346	0.3492	0.61	0.0641	0.0000	0.0000
HIS	4	0.1592	0.1594	0.29	0.0431	0.0000	0.0000
TRP	4	8.0570	5.7267	14.78	1.1053	0.0000	0.0000
N TOT	4	7.2974	1.4507				

Fig. 8. Profile for experiment set C5 TRY.

THE CONTROL C5 CONT IS NOW BEING COMPARED WITH C5 TRY

A ACID	MEAN	MEAN	DF	T
CYSO3H	0.1421	0.1244	9	0.629
G-P-E	0.7672	1.0879	9	−3.500**
TAU	17.8680	16.3198	9	0.763
O-P-E	4.4857	3.4112	9	1.922
UREA	25.2402	0.0000	9	0.000
ASP	1.0712	1.0037	9	0.517
THR	1.4049	0.8086	9	4.247**
SER	1.1460	0.8939	9	1.551
GLU	1.6641	1.4467	9	0.418
CITRU	0.0000	0.0397	9	0.000
PRO	4.4152	3.9292	8	0.756
GLY	14.4500	13.2011	9	0.580
ALA	5.7531	3.0920	9	4.461**
VAL	0.1064	0.2287	9	−5.001**
MET	0.1251	0.0345	8	10.053**
ILE	0.0159	0.0215	6	−0.546
LEU	0.0693	0.0309	9	2.381*
TYR	0.1307	0.0782	8	1.519
PHE	0.0918	0.0642	9	0.774
ORN	0.0000	0.0960	9	0.000
ETAM	0.0000	0.0375	9	0.000
LYS	0.9354	0.3346	9	4.159**
HIS	0.3751	0.1592	9	2.688*
TRP	0.0000	8.0570	9	0.000
ARG	0.0472	0.0000	9	0.000
N TOT	19.2417	7.2974	9	4.479**

Fig. 9. Result of comparisons between experiments C5 CONT and C5 TRY.

while cross-comparing experiments. At the conclusion of all the requested comparisons only those amino acids (including pseudo amino acids) for which significant differences were found are passed to the last program in the set. These form the data base from which specific questions can be answered. The data base in the example was derived from several hundred tissues involving some 25,000 amino acids.

The last program is in essence a retrieval system. On the presentation of certain key words the information these words imply is obtained. Fig. 10 examines the information that is obtained when the key word 'SETC' followed by the experiment name 'C5 CONT' is provided. Here a table of all pair-wise comparisons using C5 CONT as the control set is exhibited. Leucine in C5 TRY is represented. Care must be exercised in interpreting the table, for though leucine is significantly different

KEYWORD "SETC" HAS BEEN SPECIFIED

USING C5 CONT AS THE CONTROL SET THE FOLLOWING MULTICOMPARISON TABLE HAS BEEN GENERATED

	C5 HIS	C5 GLN	C5 TRY	B5 CONT	C5 ARG	C5 GAMGLN	C7 CONT
CYSO3H						+	
G-P-E			+			+	
TAU	−	−				−	
O-P-E	−	−					−
UREA		−				−	
ASP	−	+				+	
THR	−	−	−		−	−	−
SER		−		+		−	−
GLU		+				+	
PRO	−	−				−	−
GLY	−	−		+		−	
ALA	−		−	−		−	−
VAL		+	+			+	
MET		−	−				
ILE		+					
LEU			−		−		
TYR				−			
LYS	−	−	−			−	−
HIS	+	−	−			−	
ARG					+		
N TOT			−				

Fig. 10. Multicomparison table showing all comparisons that were made with experiment C5 CONT. '+' indicating significant elevation of the amino acids over the control, '—' the converse.

```
                                KEYWORD "ACID" HAS BEEN SPECIFIED

AMINO ACID LEU    SHOWS SIGNIFICANT DIFFERENCES BETWEEN THE FOLLOWING COMPARISONS
          S CONT BR IS HIGHER THAN S ILE  BR
          S CONT BR IS HIGHER THAN S LEU BR
          B6 GAM    IS HIGHER THAN B7 CONT
          B7 GLN    IS HIGHER THAN B7 CONT
          S LEU CL  IS HIGHER THAN S CONT CL
          B5 TRY    IS HIGHER THAN C5 TRP
          S VAL CL  IS HIGHER THAN S CONT CL
          S CONT FL IS HIGHER THAN S PHE FL
          C5 CONT   IS HIGHER THAN C5 ARG
          B5 GLN    IS HIGHER THAN B5 CONT
          B7 GLN 30 IS HIGHER THAN C7 GLN 30
          S ARG CL  IS HIGHER THAN S CONT CL
          B7 GAMGLN IS HIGHER THAN B7 CONT
          B5 GLN 30 IS HIGHER THAN C5 GLN 30
          C5 CONT   IS HIGHER THAN C5 TRP
          B6 GAM    IS HIGHER THAN B5 CONT
          B5 GAMGLN IS HIGHER THAN B5 CONT
          B7 CONT   IS HIGHER THAN B7 ARG
          B5 TRP    IS HIGHER THAN B5 CONT
          B6 GAM    IS HIGHER THAN C6 GAM
          S CONT BR IS HIGHER THAN S MET BR
          B7 GLN    IS HIGHER THAN C7 GLN
          S MET CL  IS HIGHER THAN S CONT CL
          B7 GLN 30 IS HIGHER THAN B7 CONT
```

Fig. 11. A retrieval of all those cross-comparisons between experiments for which leucine was significantly different.

in both C5 TRY and C5 ARG, these are independent evaluations with respect to C5 CONT with very definite probability levels, and may in concert reflect a different probability level. In those situations where an overall probability level is required, Dunnett's test[5] is to be preferred over the separate t tests. Similar key words are defined for samples and combinations of samples and controls. Fig. 11 illustrates the use of the key word 'ACID'. The amino acid leucine is specified and it is seen that all experimental groups in which the level of leucine was significantly different are displayed.

In addition to these elementary questions, more complex logical constructs can be built. Such an interrogation scheme gains in power as the information base develops.

All programs were written in the PL/1 language. Most of the input information is presented to the program in free form strings with character delimiters and separators and is relatively simple to construct. The author invites those who would like to try the system either for complex or simple problems to contact him.

ACKNOWLEDGEMENTS

The author would like to express his appreciation to Dr. Eugene Roberts, Director of the Division of Neurosciences, City of Hope National Medical Center, for permission to use the experimental findings that illustrate the text. These experiments were performed by Dr. John Wein under grant No. CA 02568, National Institutes of Health.

Dr. Constance Weinstein deserves the utmost praise for reading the manuscript and for her help in fusing an understanding between the theoretical concepts contained herein and both their comprehension and their experimental exploitation.

Computing assistance was obtained from the Health Sciences Computing Facility, University of California at Los Angeles, sponsored by National Institutes of Health special research resources grant RR-3.

REFERENCES

1 G. R. Shepherd, C. N. Roberts, P. N. Dean, R. D. Hiebert and R. B. Glascock, *Direct Digital Readout and Computer Calculation of Amino Acid Analysis Data*, United States Atomic Energy Commission, SR-119, LA-4507-MS, UC-48, Biology and Medicine, TID-4500, 1970, 34p.
2 H. L. Back, P. J. Buttery and K. Gregson, *J. Chromatogr.*, 68 (1972) 103.
3 H. D. Spitz, G. Henyon and J. N. Sivertson, *J. Chromatogr.*, 68 (1972) 111.
4 R. E. Exss, H. D. Hill and G. K. Summer, *J. Chromatogr.*, 42 (1969) 442.
5 B. J. Winer, *Statistical Principles in Experimental Design*, McGraw-Hill, New York, 1962, p. 201.