

Systematic Management and Analysis of Fatty Acid Data from Multiple Tissue Samples

WILLIAM N. MARMER*, ROBERT J. MAXWELL and JOHN G. PHILLIPS¹, *Eastern Regional Research Center,² Philadelphia, PA 19118*

ABSTRACT

A systematic approach has been developed for the collection and analysis of gas chromatographic (GC) data from multiple fatty acid profiles. The approach was applied to a series of polar and nonpolar tissue lipids generated in animal feeding studies to allow a comparison of mean fatty acid profiles as a function of either dietary regimen or tissue location. The magnitude of the studies, sufficiently large to minimize error from animal variabilities, mandated the use of computer assistance. Nevertheless, manual input was essential due to the complexity of the GC patterns, and was invoked for peak assignment and report editing. The approach discussed here allowed for the consolidation and statistical analysis of data from over 30,000 GC peaks, and generated results in both tabular and graphic formats. It should be extendable to other chromatographic studies of lipid components.

Lipids 18:460-466, 1983.

INTRODUCTION

The investigation of variations in tissue fatty acid profiles from experimental tissue studies is an enormous task, due to the large number of individual fatty acids that are separated by capillary column gas chromatography (GC) as well as the large number of GC runs mandated to eliminate animal variability. In our studies of the variation of bovine profiles as a function of dietary regimen or tissue location, data included over 30,000 assigned peak areas. It was evident that the data had to be handled systematically, beginning with the raw GC reports that were automatically stored on magnetic tape, through the statistical handling that allowed conclusions to be drawn from the studies. It was also evident that automation alone was no panacea; a substantial manual input would be required for editing during the entire process. This report presents a systematic approach for the management of voluminous and unwieldy data generated from large-scale animal studies.

EXPERIMENTAL

Computer Hardware

Initial data were processed first by the miniprocessor of the gas chromatograph (Hewlett-Packard 5880A GC, Level 4, with magnetic tape and ASCII keyboard accessories, Hewlett-Packard Corporation, Avondale, PA). Manually edited data were transferred to a minicomputer (Modcomp Classic 7861 with 10 million bytes of on-line

¹Consulting Statistician, Northeast Region, ARS, USDA, Philadelphia, PA 19118.

²Agricultural Research Service, U.S. Department of Agriculture, 600 E. Mermaid Lane.

*To whom correspondence should be addressed.

(disk) storage, Modular Computer Systems Inc., Fort Lauderdale, FL). Final statistical analysis was accomplished by electronic transfer of data to the USDA's Washington Computer Center (WCC) (IBM 4341 and IBM 3033 attached processor with 31 billion bytes of on-line (disk) storage, International Business Machines Corp., White Plains, NY).

Computer Software

A program (Hewlett-Packard BASIC) was prepared for generation of a modified report by the GC's miniprocessor for each GC run. The program is designed to recall a stored GC report from magnetic tape, and then to list each peak's retention time, relative retention time (RRT), area, and area percent. Manual inputs are required to assign a time reference peak (RRT = 1; our peak #140, palmitoleate) and to eliminate the area of the internal standard peak (heneicosanoate). The output format also allows manual input of sample information (tissue source, dietary regimen, percent lipid, etc.) and most importantly, of peak identities. Identification numbers for the peaks were assigned manually, with the aid of RRT's and a regularly run reference mixture. An example of a manually annotated output is shown (Fig. 1). The 5880A GC did not permit editing or modification of a report for restorage, and did not contain an RS 232 interface for direct data transfer to the in-house minicomputer. Therefore, all data (peak numbers, their individual peak areas, and sample information) had to be transferred manually to the minicomputer for subsequent statistical analysis. To simplify this transfer, peaks with areas less than 0.10% of total peak area (not including the peak area of the internal standard) were disregarded. All

METHODS

 (13)(ST)(B)(P)(F)(0.90)(0.70)(N)(Y)(15-4)(718-77)
 A B C D E F G H I J K
 ANIM TISS REPL FRAC DIET %LIP %PL MON? SICK? REPORT- BOOK-
 TAPE PAGE

 [NP] 5880A SAMPLER INJECTION @ 04:00 MAY 19, 1981
 SAMPLE # : ID CODE :
 15 77B/13ST/4

AREA %	RT	AREA	TYPE	AREA %
	15.77	826.71	BB	20.328
	23.68	520.98	BB	12.810
	26.12	801.65	VV	19.712
	30.56	682.34	BV	16.778
	43.13	1235.22	BV	30.373

TOTAL AREA = 4066.91
 MULTIPLIER = 1
 ENTER RT OF PALMITOLEATE (OUR PK #140): 17.67
 ENTER AREA OF 21:0 ISTD: 1235.22

RT	REL RT	NAME	AREA	AREA %
10.95	.620	40	14.39	.34
11.93	.675		1.61	.04
12.32	.697		2.00	.05
12.48	.706		1.56	.04
12.99	.735	80	8.03	.19
14.30	.809	90	4.11	.10
15.77	.893	100	826.71	19.37
17.04	.965	110	49.97	1.17
17.40	.984	130	24.83	.58
17.67	1.000	140	35.23	.83
18.03	1.020		3.52	.08
19.14	1.083	180	20.06	.47
21.17	1.198	210	6.24	.15
21.36	1.209	220	19.09	.45
23.68	1.340	250	520.98	12.21
25.55	1.446	260	50.02	1.17
26.12	1.478	270	801.65	18.79
26.42	1.495	280	82.48	1.93
26.74	1.513		1.57	.04
30.56	1.729	380	682.34	15.99
36.49	2.065	420	161.43	3.78
38.09	2.156	440	16.04	.38
43.13	2.441	<ISTD>	1235.22	28.94
45.17	2.556	480	21.44	.50
47.79	2.704	510	102.82	2.41
50.90	2.880	520	379.06	8.88
54.85	3.104	570	30.52	.72
58.16	3.291	600	123.76	2.90
66.08	3.740	620	28.55	.67
73.88	4.181	650	229.17	5.37
77.40	4.380	660	18.27	.43

 REPORT ANNOTATION ON *****

FIG. 1. Sample modified report. Handwritten entries are shown. Top line entries are sample information for later classification of data. Upper abbreviated table shows only the largest peaks, and allows easy retrieval of the peak area of the internal standard. There are 2 pauses for input by keyboard; first, the retention time of the time standard peak, palmitoleate, must be entered to enable the program to list relative retention times of all peaks; then the area of the internal standard peak, hecicosanoate, must be entered to enable the program to renormalize the data after exclusion of this area. Finally, the data reappear with the newly calculated relative retention times and renormalized area percentages. Space is given for manual insertion of peak identification numbers. All peaks whose renormalized area percentages are under 0.10% are deleted. The manually entered data and the raw areas are then transferred to the minicomputer for further processing.

transferred data were processed on the minicomputer by a program that checked the validity of the manual input, renormalized the data, and finally output the results. The output was written to both a line printer for proofreading and a tape file for later transmittal to the WCC.

Data were processed at WCC by the Statistical Analysis System's (SAS Institute, Inc., Cary, NC) subroutines on statistical analysis and data management. SAS procedures MEANS and GLM were used to calculate mean values and other descriptive statistics and to perform analysis of variance (Bonferroni mean separation techniques (1)). SAS/GRAPH was used to produce histograms of the summarized data for illustrating various comparisons. Finally, SAS also was used to produce tabular summaries of the average peak values for various combinations of lipid fraction, tissue type, and dietary regimen.

Statistical Analysis

The goal of this research was to determine which fatty acids were present in statistically different amounts from tissue to tissue for animals in the same dietary regimen, and from regimen to regimen for the same tissues. The analysis of variance procedure was used to test these effects. Variation from animal to animal was partitioned from the tissue effect in those comparisons between tissues within a given diet. Analysis of variance produced probability levels (p) for each fatty acid comparison. For convenience, only those differences that were significant above the 95% confidence level ($p < 0.05$) were reported. In addition, those fatty acid comparisons with less than 3 degrees of freedom for error (4 degrees of freedom for the comparisons of 3 muscles of a single regimen) were considered invalid due to insufficient data. This situation occurred with peaks of very small area percentages that often were not reported for all 10 animals of a set; some of the missing values within a set resulted from the mandated 0.10% cut-off level.

Editing Procedures

Anomalies in data from over 30,000 peaks of these studies are to be expected because of errors in editing or peak assignment, deficiencies in electronic integration, or the appearance of spurious peaks. Consequently, human intervention is required for accurate results.

Deficiencies in electronic integration occurred most often with unsymmetrical peaks, particularly with the ill-defined group of peaks that comprises the signals from the *trans*-octadecenoate isomers. Because the GC runs were automated and unattended, such errors were noted long after the completion of the runs. Correction usually re-

quired the summation of several peak areas, but occasional adjustments had to be made by the archaic "cut-and-weigh" method, whereby the peak areas were related to the weights of the cut-out peaks. Spurious peaks generally were noted by inspection of the chromatogram, and were disregarded whenever they were not confirmed in the duplicate GC run. Errors from spurious peaks and from faulty peak assignments were also detected during the statistical analysis of the data. These errors were manifested by unusually high coefficients of variation (CV) for a particular peak in the composite report from the consolidation of individual runs. For example, peak 280 (*cis*-vaccenate) in one composite report (the mean value of peak 280 from 10 animals on the same diet, in which the value of peak 280 of each animal was itself the average of duplicate determinations) was 3.70% of the total peak area, but with a CV of 164%. To investigate the problem, the individual averages of duplicate determinations were examined for each of the 10 animals. One such average was suspect. Examination of the individual runs for that 1 animal then showed a misidentification of peak 280. Corrections were made, and the composite value for peak 280 diminished to 1.68% with an acceptable CV of 12%.

RESULTS AND DISCUSSION

The course of consolidation and analysis of the individual GC reports is outlined in Figure 2, a continuation of Figure 1 of the preceding article (2). Initially, each GC report was converted into a modified report (peak numbers and corresponding peak areas; Fig. 1). For each tissue, duplicate extractions were carried out (2) and eventually led to duplicate GC runs. Because the GC's miniprocessor was incapable of consolidating multiple GC reports, subsequent manipulations were accomplished on computers with greater capabilities. Initial data consolidation was the combination of duplicate modified GC reports. The middle 2 blocks in the upper part of Figure 2 represent duplicates A and B from 2 polar lipid extracts of *M. semitendinosus* tissue from animal no. 6, an animal that had been raised on a grain regimen. The computer-generated average of these 2 reports is represented as the middle block of the next row of blocks in Figure 2. All averages of A and B duplicates served as the raw material for analysis of variance. To generate a composite report—the next line of blocks in Figure 2—this average report from animal no. 6 was combined with the other 9 average reports from polar lipid extracts of *M. semitendinosus* tissue from the other 9 animals raised on a grain regimen. The resulting composite report—a mean of 10 average reports—is represented as 1 column of the computer-generated

METHODS

(PEAK IDENTIFICATION AND RAW AREA)
MODIFIED GC REPORTS

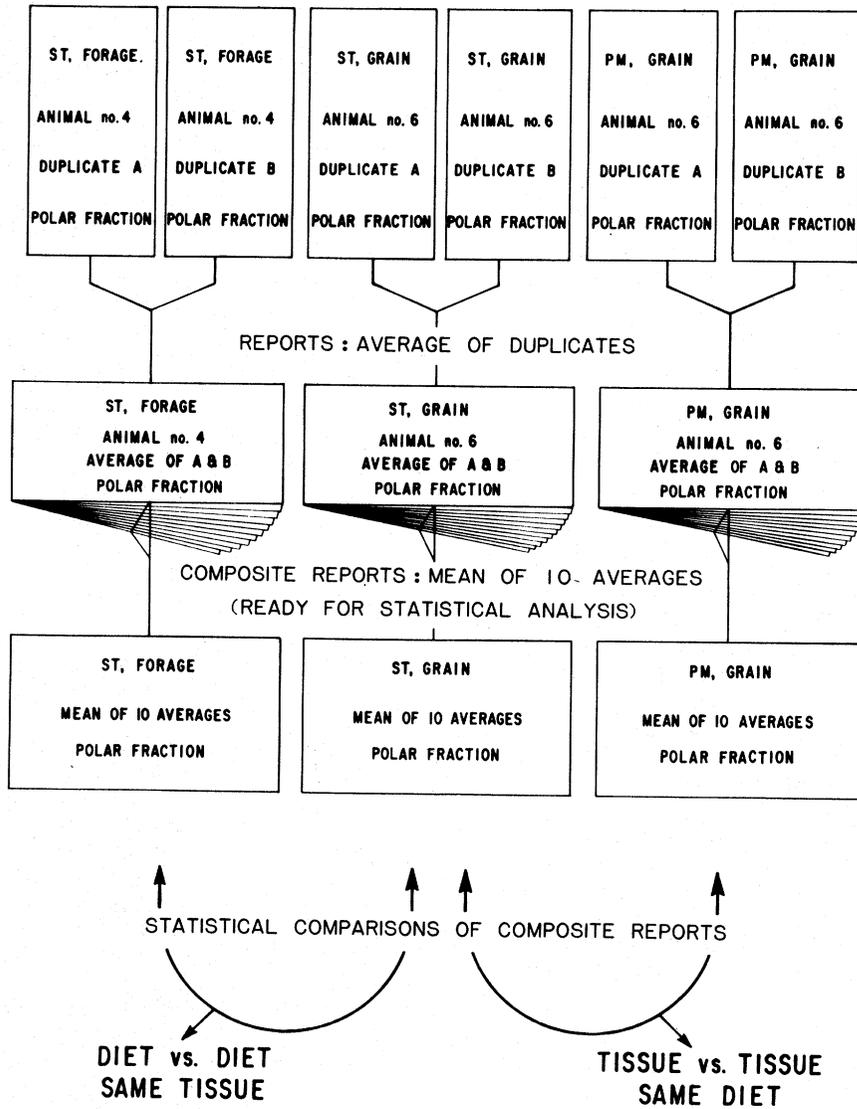


FIG. 2. Schematic for data analysis. Consolidation of data from 6 individual modified GC reports through the generation of composite reports is shown. In the actual studies, close to 500 individual reports were handled in this way. The 3 consolidated reports in the example illustrate the 2 types of comparisons that can be made a function of dietary regimen, using the same tissue (forage vs grain, using *M. semitendinosus* (ST) tissue), and a function of tissue location in the carcass, using the same dietary regimen (ST vs *M. psoas major* (PM), grain regimen). This schematic is a continuation of the schematic in Fig. 1 of the preceding article.

tabulation illustrated in Figure 3a, the column labeled "grain ST." (An analogous composite report was generated from data from corresponding neutral lipid extracts.) The column lists the mean normalized peak for each of the 60 most significant peaks, together with the number of observations (n). If the peak was seen for each animal of the set, n was 10 (9 for *M. longissimus dorsi* tissue). For example, in the mean report of the polar fraction of *M. semitendinosus* (ST) tissue of grain-fed animals, peak no. 40 (myristate) was seen for each animal (n=10) and averaged 0.62% of the total peak area.

Although the variance from animal to animal in this set of 10 is not presented in this figure, the variance information was used in the subsequent statistical analysis (done at the WCC because of the availability there of the appropriate software for Bonferroni mean separation techniques (1)). Such analysis served to determine whether any particular peak value differed significantly from its opposite, the corresponding value in another tissue (Fig. 2, right-hand blocks) or the corresponding value for the same tissue from animals raised on another dietary regimen (Fig. 2, left-hand blocks). Such

MUSCLE		FORAGE-GRAIN STUDY POLAR FRACTION									
PEAK	ID	ST		PM		GRAIN		LD		FORAGE VS. GRAIN	
		ST	PM	PM	LD	ST	LD	ST	ST		
20		0.10	1	0	0	0	0	0	0.10	1	0
40	14:0	0.62	10	0.40	10	0.40	10	0.55	9	0.62	10
50	115:0	0.24	1	0	0	0	0	0	0.24	1	0
60	ai15:0	0.21	1	0	0	0	0.15	2	0.21	1	0.15
70	14:1w5c	0.27	6	0.19	1	0.19	1	0.16	6	0.27	6
80	15:0	0.25	10	0.26	10	0.26	10	0.28	9	0.25	10
90	i16:0	0.16	2	0.11	2	0.11	2	0.10	1	0.16	2
100	16:0	19.46	10	16.15	10	16.15	10	19.18	9	19.46	10
110	16:1w7t	0.17	9	0.17	8	0.17	8	0.21	7	0.17	9
120		0	0	0	0	0	0.17	1	0	0	0
130		0.42	10	0.42	10	0.42	10	0.38	9	0.42	10
140	16:1w7c	1.29	10	1.25	10	1.25	10	1.26	9	1.29	10
150	ai17:0	0.18	5	0.13	5	0.13	5	0.12	7	0.18	5
160		0.16	4	0.12	3	0.12	3	0.10	4	0.16	4
170		0.14	1	0	0	0	0	0	0.14	1	0
180	17:0	0.76	10	0.68	10	0.68	10	0.72	9	0.76	10
190		0.10	1	0	0	0	0	0	0.10	1	0
200		0.17	3	0.43	2	0.43	2	0	0.17	3	0
210		0.22	5	0.48	2	0.48	2	0	0.22	5	0
220	17:1w9c	0.82	10	0.87	10	0.87	10	0.86	9	0.82	10
230		0.38	2	0.29	7	0.25	7	0.23	4	0.38	2
240		0.13	1	0	0	0	0	0	0.13	1	0
250	18:0	12.17	10	14.40	10	14.40	10	12.73	9	12.17	10
260	18:1w9t	1.09	10	0.95	10	0.95	10	1.02	9	1.09	10
270	18:1w9c	19.67	10	20.61	10	20.61	10	21.50	9	19.67	10

MUSCLE		FORAGE-GRAIN STUDY POLAR FRACTION									
PEAK	ID	ST		PM		GRAIN		LD		FORAGE VS. GRAIN	
		ST	PM	PM	LD	ST	LD	ST	ST		
20	U	0.70	1	0	0	0	0	0	0.70	1	0
40	14:0	2.31	10	2.58	10	2.58	10	2.29	9	2.31	10
50	115:0	0.33	1	0	0	0	0	0	0.33	1	0
60	ai15:0	0.31	1	0	0	0	0.20	2	0.31	1	0.20
70	14:1w5c	0.70	6	0.91	1	0.91	1	0.54	6	0.70	6
80	15:0	1.21	10	1.58	10	1.58	10	1.09	9	1.21	10
90	i16:0	0.46	2	0.53	2	0.53	2	0.54	1	0.46	2
100	16:0	99.28	10	102.1	10	102.1	10	79.40	9	99.28	10
110	16:1w7t	0.97	9	1.08	8	1.08	8	0.91	7	0.97	9
120	U	0	0	0	0	0	0.62	1	0	0	0
130	U	2.24	10	2.78	10	2.78	10	1.62	9	2.24	10
140	16:1w7c	6.11	10	8.26	10	8.26	10	5.28	9	6.11	10
150	ai17:0	0.78	5	0.80	5	0.80	5	0.50	7	0.78	5
160	U	0.46	4	0.86	3	0.86	3	0.40	4	0.46	4
170	U	1.10	1	0	0	0	0	0	1.10	1	0
180	17:0	3.63	10	4.42	10	4.42	10	3.00	9	3.63	10
190	U	0.66	1	0	0	0	0	0	0.66	1	0
200	U	0.91	3	3.18	2	3.18	2	0	0.91	3	0
210	U	1.30	5	3.56	2	3.56	2	0	1.30	5	0
220	17:1w9c	4.27	10	5.78	10	5.78	10	3.66	9	4.27	10
230	U	0.72	2	1.48	7	1.48	7	0.80	4	0.72	2
240	U	0.16	1	0	0	0	0	0	0.16	1	0
250	18:0	65.00	10	84.64	10	84.64	10	52.16	9	65.00	10
260	18:1w9t	5.86	10	5.26	10	5.26	10	4.18	9	5.86	10
270	18:1w9c	101.5	10	136.1	10	136.1	10	90.15	9	101.5	10

FIG. 3a. (Upper report) Computer-generated composite report, normalized data. Abbreviations: ID, identification; ST, *M. semitendinosus*; PM, *M. psoas major*; LD, *M. longissimus dorsi*; i, iso; ai, anti-iso; ω, first double bond position from hydrophobic end; c, cis; t, trans. Integers next to compositional data specify the number of animals from which individual reports were consolidated into this composite report. Maximum n for ST and PM, 10; for LD, 9. Columns of data are repeated to allow convenient inspection for each comparison (3 tissue by tissue, 1 diet by diet).

FIG. 3b. (Lower report) Computer-generated composite report, gravimetric data (mg fatty acid/100 g tissue).

statistical analysis showed, for example (Fig. 3a, arrow), that the cited 0.27% value for grain-fed ST was identical ($p < 0.05$) to the corresponding value for grain-fed *M. psosas major* (PM, 0.19%) and also identical ($p < 0.05$) to the corresponding value for grain-fed *M. longissimus dorsi* (LD, 0.16%). To have concluded that these 3 values were different (i.e., without rigorous analysis) would therefore have been invalid.

By use of an internal standard, the aliquot size, and the percent lipid, sets of gravimetric data may be generated to show the results as mg fatty acid/100 g tissue (Fig. 3b) for use by nutritionists. Because the study was designed to determine those fatty acids whose normalized amounts differed significantly from their opposites, GC response factors were not included. To convert to a gravimetric tabulation (mg fatty acid/100 g tissue), the following algorithm was applied to individual GC reports:

$$\begin{aligned} \text{(mg FAME/} & \text{(mg FAME/aliquot)} \\ \text{100 g tissue)} & = \text{(scale-up factor)} \\ & = \frac{[(\text{area sum})/(\text{area 1 mg} \\ & \text{ISTD per aliquot})] \times \\ & \text{[(mg lipid/100 g tissue)/} \\ & \text{(mg lipid/aliquot)]}}{[(\text{area sum})/(\text{area 1 mg} \\ & \text{ISTD per aliquot})] \times \\ & \text{[(1000(\% lipid)/} \\ & \text{(mg lipid/aliquot)]}} \\ \text{Since (mg fatty acid/} & = \text{0.95 (mg FAME/} \\ \text{100 g tissue)} & \text{100 g tissue),} \\ \text{then (mg fatty acid/} & = \frac{950 (\text{area sum}) (\% \text{ lipid})}{(\text{area 1 mg ISTD} \\ \text{100 g tissue)} & \text{per aliquot)} \\ & \text{(mg lipid/aliquot)} \end{aligned}$$

Definitions:	
Aliquot	= Portion of lipid extract set aside for derivatization to FAME.
(Scale-up factor)	= Ratio of lipid weight of full sample to lipid weight in aliquot.
(area 1 mg ISTD per aliquot)	= Peak area of internal standard (ISTD; here 21:0 FAME) that results from incorporation of 1 mg ISTD into the aliquot. (We used 4 mg 21:0/aliquot, and therefore divided our 21:0 peak area by 4.)
(area sum)	= Sum of FAME peak areas, not including ISTD peak area.
0.95	= Factor to convert mg FAME TO mg fatty

acid, valid ($\pm 1\%$) for FAME's C-14 through C-22.

To determine the weight contributions of each fatty acid, the (mg fatty acid/100 g tissue) figure is distributed according to the normalized report for each individual run. Thus, for any fatty acid P, (mg P/100 g tissue) = 0.01 (%P) (mg fatty acid/100 g tissue), where (%P) is taken from the normalized report. Then replicate runs are combined to generate an average report for each animal, and finally the set of average reports (usually 10) are consolidated into a mean report, such as shown in Figure 3b.

Tabular data (Fig. 3) were often unwieldy and overwhelming for purposes of study. More satisfactory was the use of graphics, including computer-generated graphics. By the introduction of field descriptors to classify each peak number, the final reports could be digested into conveniently read graphics. Although approximately half the peaks were not identified as particular fatty acids, the identified portion included over 95% of the total peak area. Unidentified peaks were classified as either saturated or unsaturated by hydrogenation experiments (3). All peaks were classified into 1 of 7 groups: (a) normal-chain saturated, (b) branched-chain saturated, (c) unidentified saturated, (d) unidentified unsaturated, (e) *trans*-monoenoic, (f) *cis*-monoenoic, (g) polyenoic. The horizontal bars at the bottom of Figure 4 illustrate one possible graphics output. Inspection of corresponding bars shows trends in fatty acid composition as a function of carcass location or dietary regimen, but does not show whether any differences are statistically significant. A breakdown of each segment of these bars into a set of histograms of constituent fatty acids, shaded to show significantly different pairs, allows a better interpretation of the trends in fatty acid patterns (Fig. 4, vertical bars). Figure 4 represents the polar fraction from *M. semitendinosus*, grain (G) regimen vs forage (F) regimen. Significant differences may be noted in 4 of the 9 identified polyunsaturated fatty acids. Although in this example the grain-fed animals generated more total polyunsaturated fatty acids than did the forage-fed (Fig. 4, bottom), this was mainly due to the (statistically significant) greater amounts of linoleate (18:2) (Fig. 4, top). Nevertheless, statistically significant greater amounts of polyenoic fatty acids, excluding the dienes, were seen in the forage-fed animals. (Such polyenoics may be responsible for off-flavors in forage-fed beef. The higher linolenate content of the samples from forage-fed animals (4,5) results from the ingestion of grasses whose fatty acids are comprised of predominantly linolenic acid. The full data sets and interpretation will be published elsewhere).

METHODS

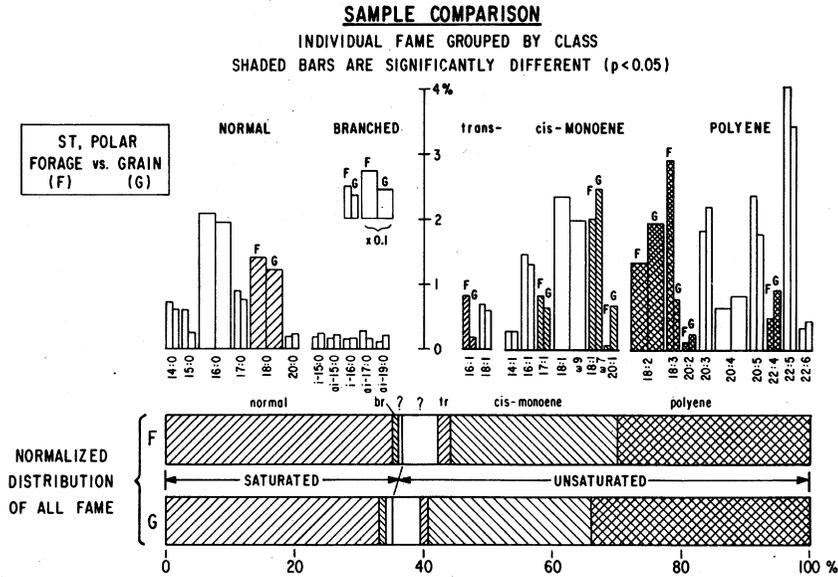


FIG. 4. Graphic representation of fatty acid distribution, expanded to show which peaks are statistically different from 1 data set (F) to another (G). Horizontal histograms are representative of computer graphics output. Vertical histograms show breakdown by fatty acid. Shaded pairs of vertical histograms are statistically different ($p < 0.05$). To allow a reasonable vertical scale, the largest vertical histograms are reduced to 1/10th their height and denoted by extra width. Abbreviations: ST, *M. semitendinosus*; FAME, fatty acid methyl ester; br, branched; ?, unidentified; tr, *trans*.

It should be evident that the processing of data sets of great magnitude requires computer assistance, not only for data management, but also to obtain statistically valid conclusions. We have demonstrated procedures that enable such processing and envision future studies that may be even more easily accomplished with increasingly sophisticated electronics.

ACKNOWLEDGMENTS

The authors are indebted to Mr. Peter D. Vail and Mr. David Smith for their help in the lab, to Mrs. Sandra P. Graham for her assistance in the transfer of data to the Center's minicomputer, and to Miss Kathy Flanagan and Dr. William C. Damert for their

programming expertise and service on the minicomputer.

Reference to brand or firm name does not constitute endorsement by the U.S. Department of Agriculture over others of a similar nature not mentioned.

REFERENCES

1. Miller, R.G., Jr. (1981) in *Simultaneous Statistical Inference*, pp. 67-70, Springer-Verlag, New York, Heidelberg, and Berlin.
2. Maxwell, R.J., and Marmer, W.N. (1983) *Lipids* 18, 453-459.
3. Christie, W.W. (1973) *Lipid Analysis*, pp. 135-6, Pergamon Press, Oxford.
4. Bidner, T.D. (1975) *Proc. Recip. Meat Conf.* 28, 301.
5. Brown, H.G., Melton, S.L., Riemann, M.J., and Backus, W.R. (1979) *J. Anim. Sci.* 48, 338-347.

[Received February 14, 1983]