Computer-Guided Explorations in Lipid Chemistry

HERBERT J. DUTTON, Northern Regional Research Laboratory,[1] Peoria, Illinois 61604

[*Address before the First Plenary Session of the American Oil Chemists' Society, October 6, 1969, Minneapolis, Minn., upon the author's acceptance of the 1969 Award in Lipid Chemistry.*]

Abstract

The concept of computer guided experimentation is illustrated in lipid research with recent examples of (1) simulations of chemical and physical phenomena, (2) processing of laboratory data, and (3) data acquisition in on-line or real time mode.

"Jeder wissenschaftliche Fortschritt ist ein Fortschritt der Methode." (Anon.)—"Every advance in scientific knowledge is an advance in technique."

The researcher lives in the future; the past and present have importance only as they bear on what is to come. In proposing to talk in the present and future tense, I am going to assume an "Awardee's prerogative" in choosing areas of immediate personal interest and then being somewhat subjective and narrative in that discussion.

The story for me had its beginning a scant 4 years ago when I enrolled in a semester course in Fortran language at Bradley University (and incidentally, found myself studying with my daughter's classmates). When I completed the semester's work, I discovered to my surprise that in an organization with a professional staff of 250 chemists, I was the only one with any computer training. In 1965, this sparsity of computer expertise among chemists was not unusual. Today, however, certain universities are accepting Fortran in lieu of a foreign language for Ph.D. requirements.

As one studies a new language (and this applies to Fortran) one inevitably finds it opens up a whole new world of literature and thought. Only a few weeks into the course and I was seeing vital applications to several pressing problems in our chemical research program, some of which I shall discuss shortly.

First, I shall say something about computer simulation of chemical and physical phenomena, then (2) I should like to describe data processing in the laboratory, and finally, (3) I shall discuss data acquisition in the "real time" or "on-line" mode.

[1] No. Utiliz. Res. Dev. Div., ARS, USDA.

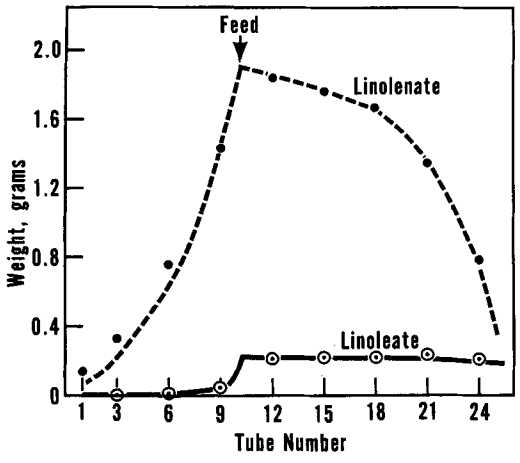

FIG. 1. Counter double current distribution of linseed methyl esters (2). ● Linolenate, ⊙ Linoleate.

Simulation

Halfway through the Fortran course, I saw possibilities of describing the process of counter double current distribution (1) with a computer and before the end of the semester I was "bootlegging" laboratory class time at Bradley to "debug" the program I had written. As you know, counter double current distribution is a multiple-batch liquid-liquid double countercurrent flow system with discontinuous feed. Whereas the more familiar countercurrent distribution, in which the lower layer is stationary, has a rigorous mathematical basis in the binomial theorem and gives rise to Gaussian-shaped curves, such a description was entirely lacking for counter double current distribution as I believe you can see in Figure 1 (2). Not only are we disturbing the Gaussian nature of the curves by chopping off the distribution at each end to yield the raffinate and extract, but we are continuously creating new truncated Gaussians by adding new samples at each transfer stage. The solutes in any given tube at any given time reflect what has happened to all previous samples introduced before this stage, i.e., the summation of all previous truncated Gaussians.

Today a computer is no longer regarded as a magic brain but rather as a marvelous moron. The painstaking job, that of keeping track of what has happened in each of as many as 50 tubes during hundreds and thousands of equilibrations, is just the kind of job that this work horse does best. I well remember the first time my associates showed me the data for a counter double current distribution curve. Having been used to the Gaussian-type curve, I tactfully suggested that perhaps the experiment ought to be repeated. But truthfully, the curves in Figure 1 are the computer-calculated values and the experimental points, as you see, fall closely upon them (2).

The parameters that we encounter in counter double current distribution are multiple and affect the shape of the curves—(a) the effect of the partition coefficient, as shown in Figure 2; (b) the ratios of volumes of upper and lower layer solvents, which give similar curves; (c) the point in the train at which the starting mixture is entered; and (d) the number of tubes, to name a few. By means of the computer we were able to calculate solute distributions and, in fact, even have the computer draw curves such as are in Figure 3, which optimize the process with regard to the separation ratio of compounds to be resolved, the number of tubes required to give a specified purity, and the degree of recovery to be expected (3). This approach to counter double current distribution became very attractive for a time because one could perform

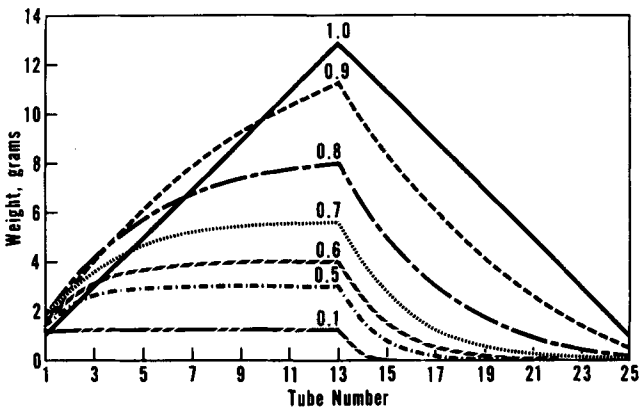

FIG. 2. Steady state calculations for extraction coefficients 0.1, 0.5, 0.6, 0.7, 0.8, 0.9 and 1.0 with center feed (2).

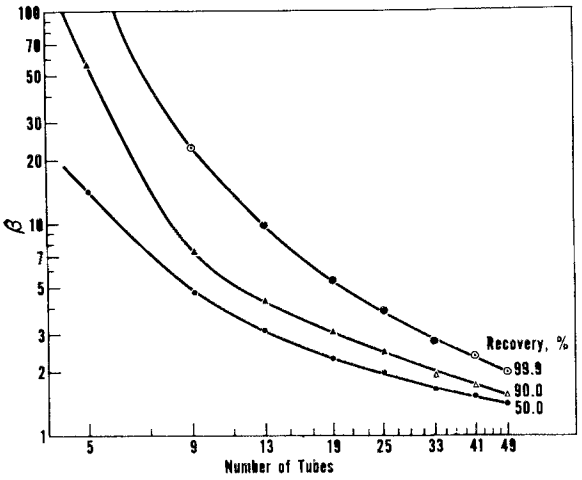

FIG. 3. Ratio of partition coefficients (β) for a 99+% pure product for several recoveries and the various numbers of tubes required (2).

a great deal of experimental work without so much as getting a test tube dirty. Seriously though, now when one approaches the problem of separating components with counter double current distribution, one first goes to the computer to simulate and optimize the conditions and then with its guidance performs the actual separation. This approach is computer-guided research.

Now I should like to turn to a second problem of simulation, namely, simulation of the kinetics of hydrogenation. It was A. E. Bailey, who after having performed the long-hand arithmetic calculations involved, wrote "There is no explicit solution of the set of equations given, and in general there appears to be no alternative to a somewhat tedious cut-and-try process" (4). By contrast, the analog computer (5) and, more recently, the digital computer (6) perform this operation in a matter of minutes which by pencilometric methods requires a matter of weeks. With a small digital computer, like the one we have at the Northern Laboratory (Fig. 4), data from gas chromatographic analyses from kinetic experiments may be introduced by means of paper tape, the relative reaction rates calculated for a complex "road map" involving five reaction rates (Fig. 5) and the output obtained in printed and plotted (Fig. 6) form (7).

Next, let us look at another computer guided exploration, the simulation of a chemical mechanism (8). As you know, during catalytic hydrogenation double bonds isomerize from *cis* to *trans* configurations and migrate up and down the chain until their location approaches a binomial distribution (Fig. 7). If one is hydrogenating with deuterium, then one finds that the farther the double bond has moved out toward the alkyl or carboxyl end of the molecule, the greater is its deuterium content. This migration, isomerization and deuterium substitution can be explained by invoking the Horiuti-Polanyi mechanism and alternations between mono- and diabsorbed species. Thus one conceives of a double bond moving outward in

FIG. 4. Digital computer with paper tape reader and plotter.

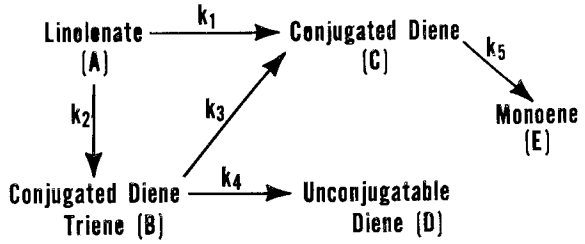

FIG. 5. Reaction scheme for copper-chromite reduction of linolenate (7).

a more or less hand-over-hand fashion and as it moves outward, deuterium is substituted for hydrogen behind it. Such a mechanism, given in Figure 8, purports to explain, through the diamond-shaped alternations between mono- and diabsorbed species in the center, not only the expected dideutero-stearate (IV) from methyl oleate (I) but also the formation of stearate with more than two deuterium atoms (X), the migration of the double bond from the initial 9,10 position with *cis* configuration to the 10,11 position with *trans* (VIII) and *cis* (VII) configuration and the still unsaturated isomers, VII and VIII, with deuterium substituted for hydrogen. The amplification of this diagram to include further migrations of the double bond both outward and toward the original center position gives rise to a triangular-shaped pattern. When this mechanism is programmed both for bond migration and for deuterium substitution, the straight lines and the binomial distributions, respectively, of Figure 7 are obtained and give a reasonably good fit of the experimental data.

The values for N are of particular theoretical interest. For the 80% saturation data, N is equal to 21. This is the value for the exponent of the binomial expansion, which gave the match of experimental data and computer-calculated theory shown in Figure 7. It is interpreted as the number of alternations between mono- and diabsorbed species that, on the average, have been experienced by those molecules of monoene which have survived hydrogenation at the 80% saturation level.

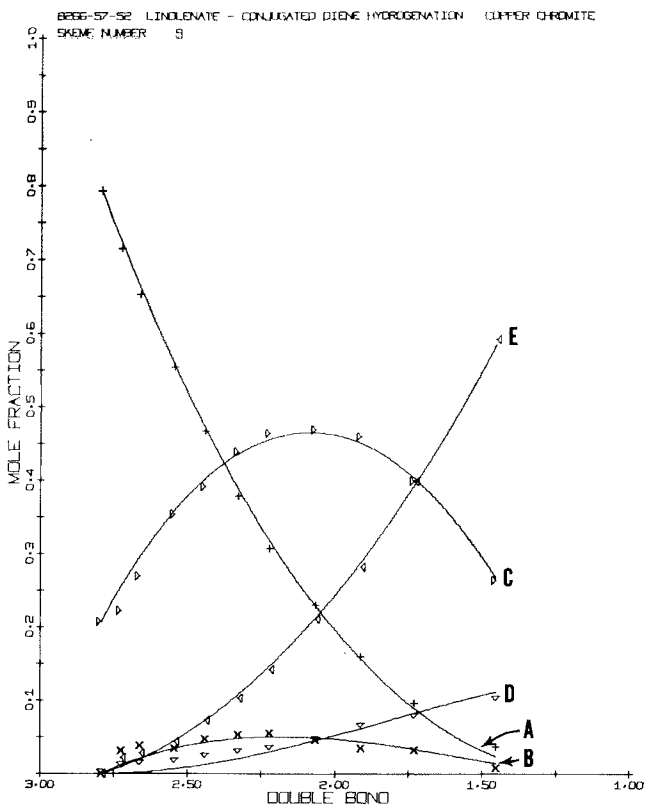

FIG. 6. Plotter output from computer (7).

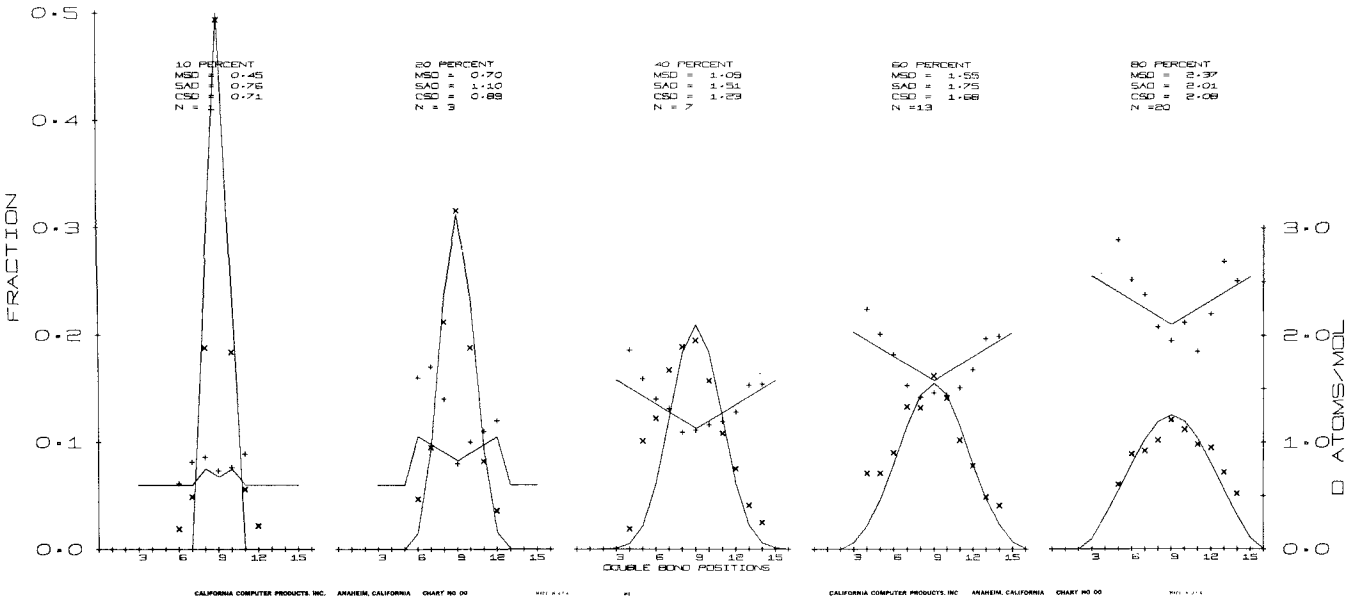

FIG. 7. Distribution of double bonds in *cis*- and *trans*-octadecenoates as a function of degree of reduction (palladium catalyst). Curves are computer drawn simulation of expanded model of Figure 8. N, exponent of binomial expansion; CSD, calculated summation of deuterium; SAD, summation of analytically determined deuterium; MSD, mass spectrometrically deuterium determined from parent peak (8).

Data Processing

Leaving this area of simulations for physical and chemical phenomena—fractionations, kinetics and mechanisms—we come to the superficially, at least, more mundane field of processing laboratory data. As an example I should like to describe computer handling of data from a radio gas chromatographic system which not only permits dual isotope labeling but also gives improved resolution and high sensitivity (9). The basic idea, sketched in Figure 9,

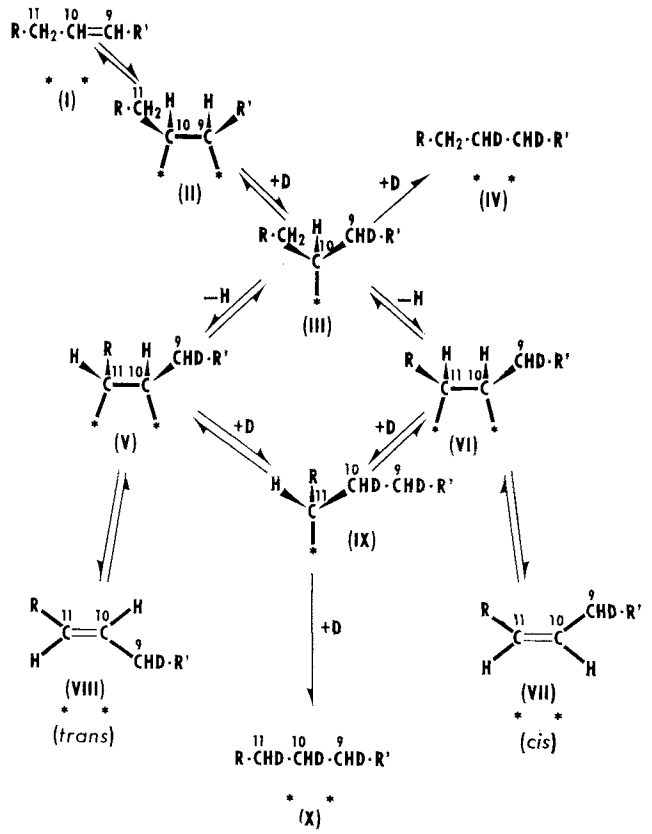

FIG. 8. Extension of Horiuti-Polanyi mechanism to show geometric and positional isomerization of double bonds and deuterium exchange and addition (8).

is that published in 1962 for single-labeled gas radio chromatography (10). Radioactive solutes in the gas stream from the chromatograph are condensed in the concurrently flowing scintillation solvent system and are automatically caught in scintillation vials placed in a fraction collector. In a dual-labeled experiment involving ^{14}C and ^{3}H, the scintillation spectrometer counts the radiocarbon through one of its two windows and counts tritium and C_{14} through its second window. This information is both printed out on a paper tape and also punched out on a perforated paper tape. A second perforated paper tape, obtained while running the gas chromatogram by digitizing the thermal conductivity curve, is fed into the reader of an IBM 1130 computer together with the radioactivity tape. The computer, then, reconstructs the thermal conductivity curve but now normalized with respect to methyl stearate as shown in Figure 10. It also performs the necessary solutions of linear simultaneous equations to separate the counts for the individual isotopes and plots these normalized curves for the individual isotopes.

Two points should be made. If the researcher were aware of the amount of hand calculation involved in such a run, he might be frightened away from using this dual-label gas chromatographic procedure; but this computerized radio chromatographic technique opens the door to a vast new area of experimentation, particularly in biological areas, in part because it permits dual isotope experiments and in part because they can be done on the nanocurie level. To my knowledge, this "do-it-yourself" laboratory apparatus provides the most sensitive procedure available to the researcher at any cost. Doing the otherwise difficult or impossible amount of routine calculation is the "cup of tea" for a computer in its data processing mode.

At the time I was learning Fortran language, mass spectrometric data were universally obtained on strip chart recorders. The height of each peak had to be read; the mass number and the intensity information had to be recorded in tabular form, and then all intensities were divided by the highest peak that was found, or the base peak, to give a normalized curve; finally, the normalized data were replotted. This process for methyl stearate, for example, was a three-day job for one man. We now do this job of data processing with a computer in 5 min. The information is still obtained from the strip chart

(Continued on page 120A)

(Continued from page 102A)

recording but as shown in Figure 11 the information is also digitized, printed on a paper tape, and punched out on a perforated paper tape. Then the perforated paper tape is read into the computer, the normalized data are calculated, printed out and graphed on the computer's plotter. We can hardly refer to this as computer-guided research, but certainly it is computer aided research in which the scientist is relieved of boredom and tedium and given the time to use his creative capabilities for those things which the computer cannot do.

Impressive and time-saving as this system of data handling seems to us, we are brought up sharply to the realization of its total inadequacy when we approach the problem of the aromagram—the identification of odor principles by the nose, gas chromatography, and high-resolution mass spectrometry. This problem, then, brings us to the next level of sophistication; namely, the real time computer.

Real Time or On-Line Data Acquisition

Let me illustrate the concept of real time computers with a recent moon shot of Apollo 11. When the astronauts experienced a slight wobble while redocking with Eagle, Mission Control advised them, as I remember, "We have data and we will chaw on it for a while." In other words their computers would be simulating Command Module and Eagle interactions using actual experimental data and would be advising the astronauts on what corrective actions to take. This one example makes us aware of how on-line computer operations work and how calculations can be made on a model while the experiment itself is going on; then these calculations can be used to guide, conduct or change the experiment.

Computer-guided means that theory and experiment are brought side by side. This approach contrasts with the conventional approach, that of waiting for an indefinite period of time, maybe days or weeks, after the experiment is completed to determine if the range, number and accuracy of measurements are sufficient.

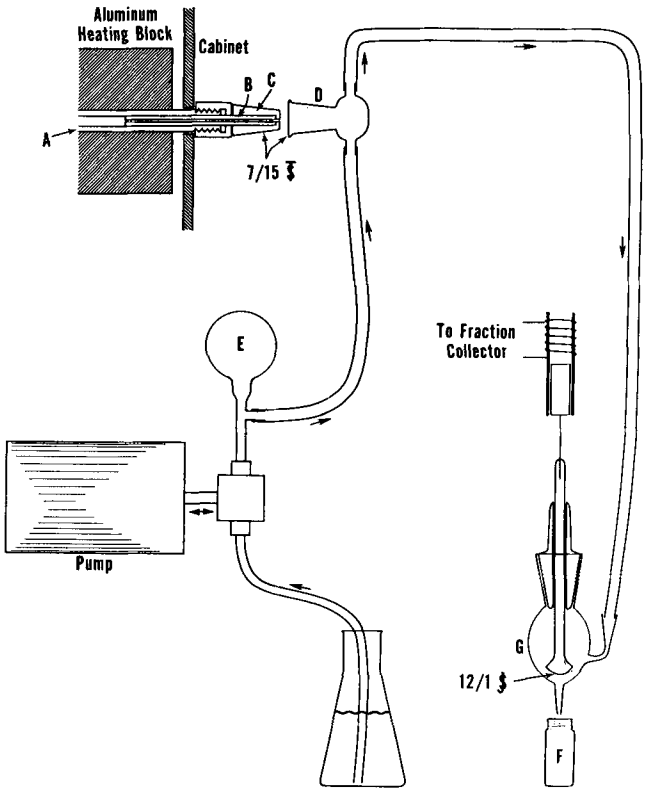

FIG. 9. Modified automatic system for condensation of effluent in scintillation solvent (9).

In an analogous manner, if not of equal complexity, when we are observing an odor concentrate of soybean oil with the nose, gas chromatography and high-resolution mass spectrometry, we shall be doing so in real time. Each of the perhaps 30 chromatographic peaks should be sampled three times to be sure that it is a single component and not multiple. If the gas chromatographic peak has a duration of 1 min, the determination of a whole high-resolution spectrum should be done in 10 sec. We estimate that we will need to be sampling data (Fig. 12) from the high resolution mass spectrometer at a rate of 90,000 samples per second. Because 90% of the time the recorder pen of the mass spectrometer is on the base line and giving no useful data, with an analog discriminator we may throw away this information. Then suppose, we have 20 intensities and 20 mass values taken over each ion peak. By means of digital logic and computer "massaging" of the data, the highest intensity in the center of the peak may be determined and we may throw away 19 pairs of data, saving only 2. Left with less than 1% of the original information from the mass spectrometer, this remainder may be either stored on a disc temporarily or immediately transferred to the central computer. Here the mass spectrum may be printed out on the line printer or compared with known spectra already on magnetic tape in the computer's library. Conceivably, these identifications might be relayed back to the operator via a teletype at the mass spectrometer and to the man who is concurrently making his psychometric or his organoleptic evaluations of the same odor peak which is still issuing from the gas chromatograph, a real-time operation. Thus, while the gas chromatograph is telling how many components are present and how much of each, the mass spectrometer will be identifying the components chemically or calculating their empirical formula. All this information can be supplied to the man who is characterizing the odors by his nose, a computer-guided exploration. It is conceivable that the plotter at the central computer will not only be reproducing the gas chromatographic curve, but will also be printing upon each peak its chemical identity or empirical formula.

Last, but not least, I wish to describe chemical control operation which we are just now initiating at the Northern Laboratory and which will be an on-line or real-time operation. Our concept of computer systems has not been that of automating individual instruments or processes, but rather of general laboratory automation. Thus, we

(Continued on page 122A)

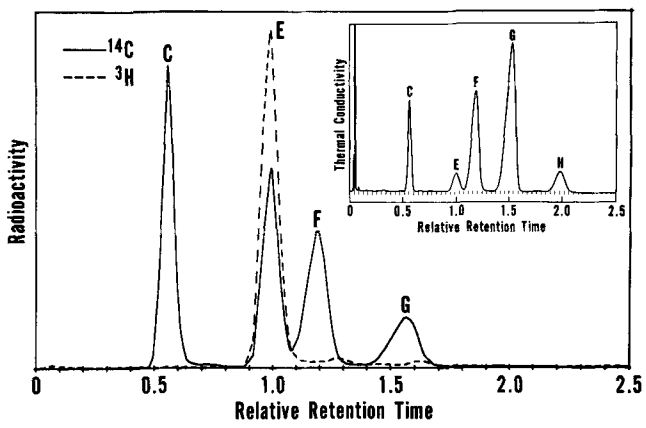

FIG. 10. Computer-plotted profile of radioactivity in the effluent of a gas liquid chromatograph, averaged from three repeated analyses. The sample contained ^{14}C-labeled methyl palmitate (C), methyl stearate (E), methyl oleate (F), and methyl linoleate (G), diluted with soybean oil methyl esters to a final specific activity of 9 nc/mg. Tritium-labeled methyl stearate was added to a final specific activity of 40 nc/mg. Methyl stearate is assigned a relative retention time of 1.0. The thermal conductivity analysis is shown in the insert. H is methyl linolenate (9).