Journal of Chromatography, 99 (1974) 243–256

CHROM. 7635

GUIDELINES FOR THE DEVELOPMENT OF A COMPUTER SCHEME TO IDENTIFY CHROMATOGRAPHIC PEAKS

D. D. CHILCOTE

*Oak Ridge National Laboratory**, *Oak Ridge, Tenn. 37830 (U.S.A.)*

SUMMARY

In the development of a computer scheme to identify chromatographic peaks, decision-making algorithms must be constrained to rely on a reasonably limited body of data. Therefore, a procedure for normalizing the retention data to enhance their reliability is desired. The dual-reference peak method of normalization using multiple sets of reference peaks is shown to have the greatest general utility. A method of establishing the optimal location of reference peaks is presented. The expected degree of uncertainty in the normalized elution data is discussed.

INTRODUCTION

The primary goal in the development of a particular chromatographic system is the achievement of a desired separation, *i.e.*, the complete and unambiguous resolution of all detectable constituents (or as many as possible with existing technology) in a certain sample mixture. However, the utility of a given separation, regardless of how satisfactory it may be, cannot be realized unless the resulting chromatographic peaks can be identified. Identification is normally straightforward for a few widely spaced peaks of known elution times. On the other hand, the correct identification of each peak in a chromatogram containing a large number of closely spaced peaks can be very difficult. In the past, peak identification has been accomplished by manual techniques. Recently, computer-assisted techniques for the purpose of peak identification have appeared[1–6].

In most chromatographic analyses, the peak positions in the resulting chromatograms are recorded in units of time. Elution times are subject to uncertainty arising from variations in flow-rate, column temperature, temperature programming, injection techniques, eluting solvents, column packings, etc. Therefore, any identification of a chromatographic peak that is based solely on the elution time of that peak is inherently uncertain. Often, this uncertainty is circumvented to a great degree by mathematical techniques which normalize the retention time data.

Additional, often strictly qualitative, information is frequently used as an adjunct to the elution position in the manual identification of chromatographic peaks.

* Operated for the U.S. Atomic Energy Commission by Union Carbide Corporation.

Position of a peak relative to other peaks, peculiarities in peak shape, photometric detector response at one wavelength compared with the response at another wavelength, etc., can be used to supplement the information regarding elution position as an aid to peak identification. This use of supportive information in the manual identification of chromatographic peaks is very difficult to achieve in computer-aided peak identification since the computer can only make decisions with regard to identification based on a reasonably restricted body of input data (primarily elution positions). Although the potential for using additional information (wavelength ratio, etc.) for computer-assisted peak identification does exist, the delivery of these data to the computer in a quantitatively reliable form presents problems within the constraint of a reasonable cost since additional interface capability, increased memory, more complex software, etc., would be necessary. In general, therefore, computer identifications are based on elution data alone. One exception to this general statement has been discussed[1].

Since a computer-aided identification scheme for chromatographic peaks depends mainly on elution data, the computer must be programmed to make decisions based on the match between a measured elution time and a known (or expected) elution time. The course pursued in choosing a normalization technique and in developing these decisions (algorithms) is determined to a great extent by the characteristics of this match. Acceptable tolerances between normalized elution times of the same peak measured in separate analyses must be known. The relationship between tolerance limits and elution time is also of interest. In this paper we shall examine the relationship between the measured elution times (in a given analysis) and expected elution times (averages based on numerous analyses) with the expressed purpose of establishing guidelines for the development of a computer program for the identification of chromatographic peaks. Although it is true that the actual working tolerances for any chromatographic system will have to be determined experimentally, the conclusions of the present discussion will point to judicious choices for data presentation which will simplify determination of tolerance limits. The actual computer program for the identification of chromatographic peaks is described elsewhere[1].

An average elution time $\bar{t}_i$ can be determined for each compound, i, present in a given mixture of compounds analyzed repeatedly by some chromatographic means. For any given analysis of this mixture, which results in a measured retention time t_i for each compound i ($i = 1, \ldots, n$) we define:

$$\frac{t_i}{\bar{t}_i} = \alpha_i \tag{1}$$

If, in any subsequent analysis, the determined value of α_i is the same for each compound, i, in the sample, then the measured elution time and the average (from numerous analyses) elution time for one (reference) compound can be used to calculate the value of α via eqn. 1. In turn, this value of α can be used to convert the measured elution times of all the other eluted compounds in the mixture to their respective average (*i.e.*, expected) elution times. By this method, the inclusion of one reference compound in the sample mixture whose average retention time is known will allow the average elution times of all the other compounds to be calculated. Such an approach has been used as a peak normalization technique[7]. The success of this technique for

the calculation of average or expected elution times from the measured elution times rests entirely on the validity of the assumption of a constant value for α.

The successful performance of a peak identification scheme relies on accurately matching a measured retention time, t_i, of an unknown peak, i, to an expected retention time, $\bar{t}_k$, for a known compound, k. The above discussion indicates that an uncertainty, $|\bar{t}_i - t_i|$, exists between the measured retention time t_i and the expected retention time $\bar{t}_i$ for compound i. Furthermore, from eqn. 1 we see that:

$$|\bar{t}_i - t_i| = |\alpha t_i - t_i| = t_i |\alpha - 1| \tag{2}$$

i.e.,

$$|\bar{t}_i - t_i| = \text{constant} \cdot t_i \tag{3}$$

Thus, the difference between the measured elution time and the expected elution time increases as the retention time increases. Consequently, the uncertainty in the location of any peak with respect to the expected elution time of that peak increases with increasing elution time.

The foregoing discussion reveals two significant problems that must be resolved in developing a computer-aided peak identification scheme: (1) the variability or uncertainty in the location of each peak from one analysis to the next, and (2) the increasing uncertainty in locating a peak with respect to increasing elution time.

The variability in the elution time for a given peak from one analysis to the next is the result of many factors. The most important of these is probably variation in flow-rate, although variations in other operating parameters such as column temperature, temperature programming, column packing, etc., also contribute. Several techniques have been devised to eliminate the dependence of elution time on flow-rate and, hopefully, on other variable parameters as well. These techniques, called peak normalization techniques, utilize reference compounds for comparison purposes. Two different approaches are commonly used: one employs a single reference peak, while the other employs two reference peaks to bracket the sample peaks of interest.

It has been accepted practice to use only one reference peak per chromatogram for the single-reference method and numerous reference peaks (*e.g.*, a homologous series of compounds) for the dual-reference method. Using the single-reference method, all peaks in a chromatogram are normalized to one reference peak, j. For peak i ($i = 1, \ldots, n$) with elution time t_i, the normalized elution time t_i^* is determined as:

$$t_i^* = \frac{t_i}{t_j} \tag{4}$$

Using the dual-reference method, all peaks located between any two consecutive reference peaks, j and $j + 1$, are normalized to those two reference peaks. For peak i located between reference peaks j and $j + 1$, the normalized elution time is determined by:

$$t_i^* = \frac{t_i - t_j}{t_{j+1} - t_j} \tag{5}$$

In this discussion we choose the dual-reference method. The reasons for this choice are discussed later.

If the proportionality factor α in eqn. 1 is constant with respect to elution time, then normalizing the elution times of the chromatographic peaks via eqn. 5 will eliminate the uncertainty in their locations. The difference (uncertainty) between the measured elution time for compound i and the expected elution time for the same compound after normalization is, for constant α:

$$t_i^* - \bar{t}_i^* = \frac{t_i - t_j}{t_{j+1} - t_j} - \frac{\bar{t}_i - \bar{t}_j}{\bar{t}_{j+1} - \bar{t}_j} = \frac{t_i - t_j}{t_{j+1} - t_j} - \frac{\alpha(t_i - t_j)}{\alpha(t_{j+1} - t_j)} = 0 \qquad (6)$$

When α is constant, the normalized value of the measured elution time for compound i is equal to the normalized value of the expected elution volume for that compound. Hence it appears that no uncertainty persists. Experimentally, however, even the normalized elution times are found to vary from analysis to analysis[1]. In general, therefore, α cannot be assumed to be constant for any given analysis.

Previous work[8] has shown that α can be expressed as a continuous function of the elution volume V, *i.e.*,

$$\alpha = \alpha(V) \qquad (7)$$

and can therefore be expressed as a power series expansion in V:

$$\alpha(V) = \sum_{i=0}^{\infty} C_i V^i \qquad (8)$$

From physical reasonings, we expect α to be a smooth, slowly varying function of V over the entire chromatogram. This consideration is supported by the success of a normalization technique that implicitly assumed α to be constant[7]. Thus, if a sufficient number of reference peaks are used, it seems reasonable to expect that the function α can be approximated by a linear, or at worst quadratic, curve between each set of reference peaks. The slope or curvature of the approximation to α may vary between successive sets of reference peaks, but the functionality of the approximation (linear or quadratic) will remain the same. Since α is a continuous function of V, the mathematical validity of the linear approximation is obvious if the interval between reference peaks is made sufficiently small. On the other hand, the interval must also be large enough to be useful from a practical, experimental point of view. The peak normalization work alluded to earlier[7] and the peak identification scheme based on the principles presented here[1] indicate that a reasonable interval size can be expected for many chromatographic applications.

To experimentally justify these expectations concerning the behavior of α, the functionality of α with respect to $\bar{t}$ was determined under varying conditions in two quite distinct chromatographic systems. The data most representative of routine chromatographic operation were used to illustrate how the determination of α can lead to optimal choices of the location of the reference peaks. These studies, the ensuing conclusions regarding the mode of data reduction for purposes of computer identification, and the concomitant uncertainties in identification are discussed below.

EXPERIMENTAL

A chromatographic system similar to that developed recently to analyze for nucleosides and N-bases in physiologic fluids[9] was used to investigate the characteristics of α in ion-exchange chromatography. This chromatographic system was operated at ambient temperature using a buffered eluent of constant composition (0.015 *M* ammonium acetate–acetic acid). A 150-cm stainless-steel column of 0.45 cm I.D. was used. The column packing, Aminex A-27 (batch No. 11330), was obtained from Bio-Rad Labs. (Richmond, Calif., U.S.A.). Several common urinary constituents were used as reference compounds.

As an alternative approach, the characteristics of α in gas chromatography were investigated using a MicroTek MT-220 gas chromatograph with a flame ionization detector. The glass column (6 ft. $\times$ 0.25 in. O.D.) was packed with 3% OV-1 on 80–100 mesh Chromosorb W-HP. Operating conditions included temperature programming from 100 to 325° at 10°/min. The flow-rate of the helium carrier gas was maintained at 80 ml/min. The reference compounds were a series of *n*-alkanes dissolved in hexane. These compounds are used for reference purposes in the routine operation of this chromatograph in our laboratory[10].

RESULTS AND DISCUSSION

The key to developing a reliable computer program for the identification of chromatographic peaks lies in the choice of a good peak normalization scheme. For some applications, the single-reference method will work well. In others, however, the use of the bracketing, dual-reference method is imperative. The use of the α curve to determine which method is most suitable and to establish, in the case of the dual-reference method, which reference peaks are necessary is detailed in the following discussion.

Selected groups of nucleosides and N-bases were analyzed repeatedly at a constant eluent pump setting to determine both the behavior of α in any given run and the differences in the α-*vs.*-i curve from run to run. In this experiment as well as in the series of experiments that followed, the retention times determined in the first run of the series were designated as the expected values for the purpose of calculating α. The value of α for each eluted species in every subsequent analysis in that series was calculated by eqn. 1, *i.e.*, the retention time of a given compound in a given analysis was divided by the retention time for the same compound in run 1.

The results from the first series of experiments are given in Table I. The values of α are presented graphically in Fig. 1a. It is obvious from the results here that, for these particular compounds subject to the given chromatographic conditions described above, the elution times are extremely reproducible. The average values of α for the second, third, and fourth analyses are 1.012 ± 0.002, 1.018 ± 0.003, and 0.973 ± 0.005, respectively. Under these conditions, the assumption of constant α in each analysis is very appropriate. (The slight variations in the values of α from one run to the next are of no practical consequence.) With a constant α, we would expect (from eqn. 6) the normalized elution volume of each compound to be the same in every analysis. The data in Table I confirm this expectation.

To test the influence of a widely varying flow-rate on the behavior of α, another

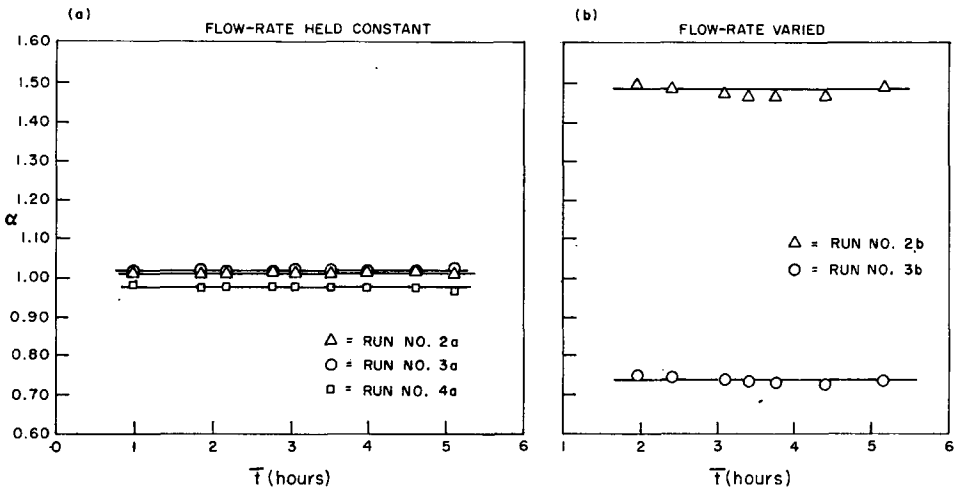

Fig. 1. Relationship of $\alpha_i = (t_i/\bar{t}_i)$ *versus* $\bar{t}$ for several urinary nucleosides and N-bases analyzed by anion-exchange chromatography. Each data point represents the value of α for one of the compounds listed in Table I.

TABLE I

ELUTION TIMES (h) AND NORMALIZED ELUTION TIMES OF REFERENCE NUCLEOSIDES AND N-BASES

Cytidine and 7-methylguanine were used as reference compounds in the dual-reference peak normalization technique.

Compound	*Run No.*							
	1		*2*		*3*		*4*	
	t	*t**	*t*	*t**	*t*	*t**	*t*	*t**
Cytidine	1.00	0	1.01	0	1.01	0	0.978	0
Uracil	1.84	0.204	1.86	0.205	1.87	0.210	1.79	0.207
Uridine	2.18	0.287	2.20	0.287	2.22	0.295	2.13	0.293
1-Methylinosine	2.72	0.418	2.76	0.422	2.77	0.429	2.66	0.427
7-Methylxanthine	3.04	0.496	3.08	0.499	3.10	0.510	2.96	0.502
Hypoxanthine	3.52	0.613	3.56	0.614	3.59	0.629	3.42	0.620
Inosine	3.99	0.727	4.03	0.728	4.06	0.744	3.87	0.734
1-Methylguanosine	4.61	0.878	4.68	0.884	4.71	0.902	4.48	0.888
7-Methylguanine	5.11	1.0	5.16	1.0	5.21	1.0	4.92	1.0

series of experiments was performed. Three chromatographic analyses of the reference compounds were carried out, each at a significantly different flow-rate. The results from this series of experiments are presented in Table II. The values of α, again using the elution times calculated in the first run as the expected values, are plotted in Fig. 1b. In spite of the wide variation in flow-rate, the value of α remains essentially constant throughout each analysis. The average values of α for the second and third runs are 1.48 ± 0.010 and 0.738 ± 0.008, respectively. The effect of a variable flow-rate on the value of α is well illustrated in this figure.

Table II also shows the normalized elution times for each analysis. Since α is

TABLE II

ELUTION TIMES (h) AND NORMALIZED ELUTION TIMES OF REFERENCE NUCLEOSIDES AND N-BASES UNDER VARYING FLOW-RATE CONDITIONS

Uracil and 1-methylguanosine were used as reference compounds in the dual-reference peak normalization technique.

Compound	*Run No.*					
	1		*2*		*3*	
	t	*t**	*t*	*t**	*t*	*t**
Uracil	1.96	0.0	2.93	0.0	1.48	0.00
Uridine	2.40	0.138	3.58	0.136	1.78	0.133
1-Methylinosine	3.08	0.352	4.57	0.345	2.28	0.345
7-Methylxanthine	3.37	0.440	4.96	0.428	2.48	0.432
Hypoxanthine	3.74	0.557	5.49	0.540	2.73	0.543
Inosine	4.38	0.755	6.45	0.743	3.19	0.741
1-Methylguanosine	5.16	1.00	7.67	1.0	3.79	1.0
	Flow-rate (ml/h)					
	17.7 ± 0.5		11.8 ± 0.1		23.9 ± 0.2	

constant throughout each analysis, we would expect to find that the normalized elution time for any compound is the same in every analysis. In this example, normalization obviously aids identification.

The influence of more complicated chromatographic conditions on the behavior of α was tested with a gas chromatograph operating in a temperature-programmed mode. Since this chromatographic system is used routinely in our laboratory for dual-column analyses of various biochemical compounds, data on the analysis of reference hydrocarbon compounds were readily available. The results presented in Fig. 2 are random selections from a series of analyses, all made on the same day. Using run 1

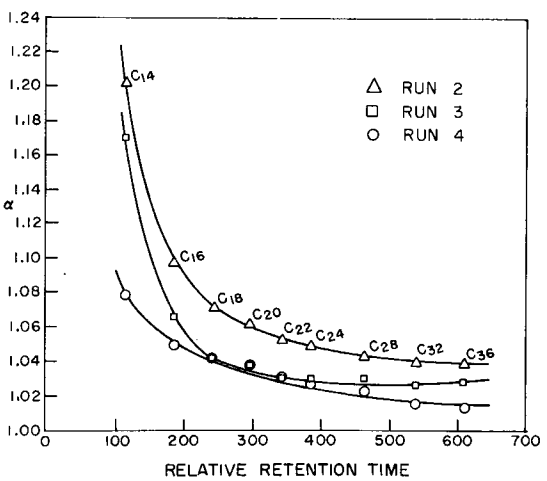

Fig. 2. Relationship of α *versus* $\bar{t}$ for a series of *n*-alkanes analyzed by gas chromatography in four separate runs.

as the reference, the calculated values for α were found to vary noticeably with retention time for the remaining runs.

We are interested in determining to what extent variations can be expected in the normalized retention times for a group of compounds under a given set of chromatographic conditions. We have already shown that, if α does not remain constant with respect to retention time during the course of an analysis, variations in the normalized retention times from one run to the next can be expected. In addition, we know that the variations in the normalized retention times can be related to the slope and intercept of the α curve, provided a straight-line approximation to α can be made over the range of retention times in question.

We are also interested in determining the minimum number of reference peaks necessary to stabilize the normalized retention data from one run to another.

The theory, which was presented in a previous paper[8], demands that any linear approximation to the α curve must begin at the origin of the retention time range. For the data presented in Fig. 2, this origin is zero. The curvature of the α plots at the low end of the retention time range makes a reasonable extrapolation to zero impractical (compare these curves with those in Fig. 1). The single-reference peak method of normalizing retention times, in which one reference peak is used for the entire retention time range, would not give satisfactory results under these circumstances. However, the dual-reference method, using the peaks at either end of the retention time range (*i.e.*, C_{14} and C_{36}), may still give excellent results, for reasons to be discussed later.

In any case, the curvature of the α plot indicates that a number of reference peaks may be necessary to establish the required precision in the values of the normalized elution volumes. In determining the optimal locations for the reference peaks, our approach will be to search for those regions in which α is nearly constant by translocating the origin of the retention time range. We proceed by shifting the origin from zero to the first peak, C_{14}. This involves subtracting the retention time of peak C_{14} from the retention time of every other peak in the chromatogram. This is done for each analysis. The modified data are then used to recompute the α curve. Fig. 3 shows the results of this procedure for runs 1 and 2. We note that, even with the origin at C_{14}, a significant curvature remains. This shifting of the origin is continued until the α curve levels out, which is at peak C_{18} in the present example. Thus, in this case, a minimum of four reference peaks would be required: C_{14}, C_{16}, C_{18}, and C_{36}. The advantage of using four reference peaks instead of two (C_{14} and C_{36}) is shown in Table III. Here the normalized elution times based on the dual-reference peak method are given for peaks falling between the established reference peaks. When the two reference peaks C_{14} and C_{36} are used, the average variation between the normalized elution times in the two runs is 0.012. However, when the four reference peaks are used, the average variation in normalized elution times for peaks between C_{18} and C_{36} is only 0.001.

In this example, the α curve that had its origin at C_{18} remained reasonably constant throughout the remainder of the retention time range. If, on the other hand, the curve had started to rise or drop sharply at C_{28}, the origin would have been shifted to C_{28}, and it would have been necessary to recalculate α from this point on. This procedure would then be continued until all regions of constant α were identified.

As stated previously, the results just presented were random selections from a

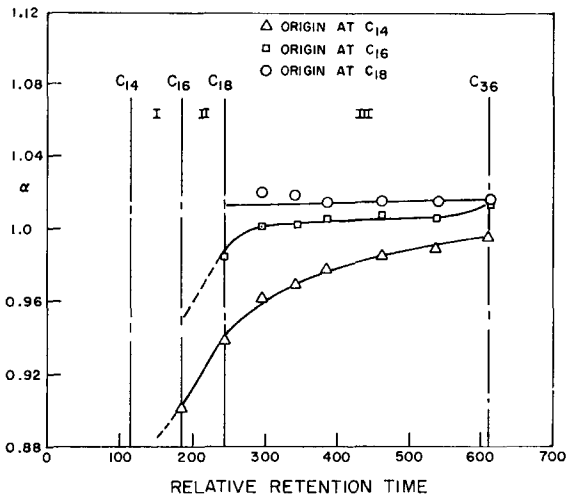

Fig. 3. Effect of translocating the origin on the curvature of the α *versus* I plot for a series of *n*-alkanes analyzed by gas chromatography.

series of routine gas chromatographic analyses. In these analyses, the chromatograph was operated in the dual-column mode. The injection procedure consisted of starting the recorder for column 1, injecting the sample onto column 1, starting the recorder for column 2, injecting the sample onto column 2, and finally starting the temperature program. Extreme care was not taken to maintain the exact time interval of each sequence in the startup procedure; hence some inconsistency between runs could be expected.

After these data had been compiled, another series of chromatographic analyses was initiated in which only one column was operated. In this study, a serious effort was made to reproduce injection conditions exactly from one run to the next. With care, the results were found to be very reproducible. These data were used as a basis on which to compare variations in flow-rate and in the rate of temperature programming.

TABLE III

EFFECT OF INCREASING THE NUMBER OF REFERENCE PEAKS ON THE REPRODUCIBILITY OF THE NORMALIZED ELUTION TIMES

Compound	*Normalized elution time**		*Normalized elution time***	
	Run 1	*Run 2*	*Run 1*	*Run 2*
C_{16}	0.141	0.128	—	—
C_{18}	0.261	0.245	—	—
C_{20}	0.363	0.351	0.139	0.138
C_{22}	0.458	0.448	0.268	0.269
C_{24}	0.545	0.535	0.385	0.384
C_{28}	0.703	0.696	0.598	0.597
C_{32}	0.854	0.850	0.803	0.801

* Reference peaks: C_{14} and C_{36}.

** Reference peaks: C_{14}, C_{16}, C_{18}, and C_{36}.

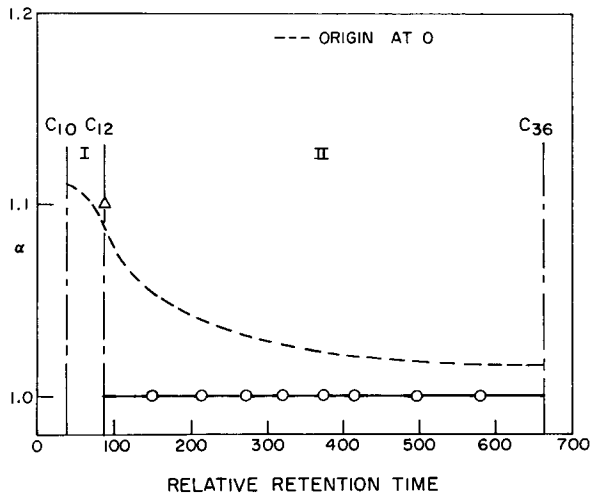

Fig. 4. Establishing the number of reference peaks necessary to maintain reproducible normalized elution times in gas chromatographic analyses subject to fluctuations in flow-rate.

Fig. 4 shows the α curve for an analysis made when the flow-rate of the carrier gas was increased to 90 ml/min (the reference condition was 80 ml/min). The α curve, with the origin at zero, is shown by the dashed line. Translocating the origin to the C_{12} peak resulted in a curve for α with a slope of zero over the remainder of the retention time range. If this behavior were continually observed, only three reference peaks, C_{10}, C_{12}, and C_{36}, would be necessary in order to obtain reproducible normalized retention times over the range of interest. The single-reference peak method would not be suitable. Notice that the effect of variations in flow-rate on α when temperature

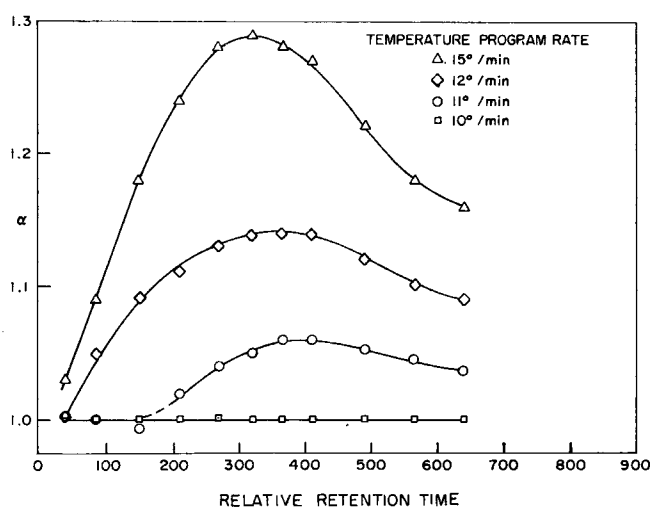

Fig. 5. Relationship of α *versus* I for a series of *n*-alkanes analyzed by gas chromatography at different temperature program rates.

programming is employed is considerably different from that observed for isothermal operation using liquid chromatography (compare the dashed line in Fig. 4 with the curves in Fig. 1b).

Fig. 5 illustrates how variations in the rate of temperature programming affect α. Curves for rates of 10, 11, 12, and 15°/min are shown (reference data obtained in 10°/min). In Fig. 6, the retention time range for the data collected at 12°/min has been divided into sections of constant α by sequentially translocating the origin as described previously. In this case, a minimum of four reference peaks would be necessary to obtain good precision in the normalized elution times.

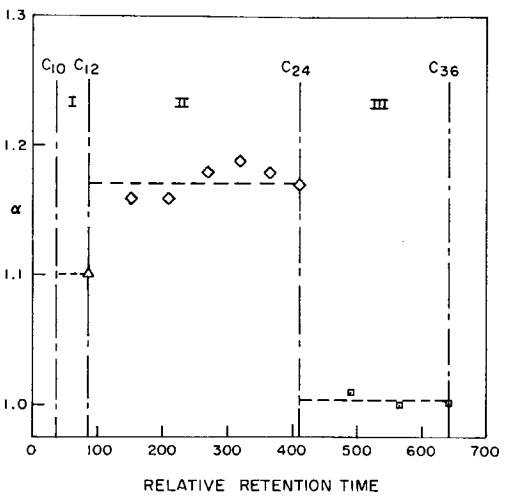

Fig. 6. Establishing the number of reference peaks necessary to maintain reproducible normalized elution times in gas chromatographic analyses subject to fluctuations in rate of temperature programming.

These examples are given to illustrate the usefulness of this method in establishing the appropriate choice of reference peaks for a given chromatographic system. The latter examples show that there are cases in which the α curve is quite non-linear over the entire chromatogram. In such instances, a single-reference peak method in which the actual retention time data were used would have little utility. This method is only useful when the α curve is essentially a straight line with low or negligible slope. When this does occur (*e.g.*, see Fig. 1), the single-reference and the dual-reference peak methods work equally well[8].

In determining the optimum choice of reference peaks for a chromatographic separation, our procedure has been to divide the chromatogram into sections of constant or nearly constant α. In each section, α is computed after the measured retention times are modified by subtracting the retention time of the first peak in the section from the retention times of all other peaks in the section. After the data have been modified in this manner, they could be normalized using a single-reference peak (*e.g.*, the last peak in the section). If this procedure were followed, the conclusions regarding the utility of the single-reference peak method *vs.* the dual-reference peak method presented in a previous paper[8] would also be valid here. However, due to the

manner in which the data are corrected, the single-reference peak method would, in fact, be a dual-reference method with the first peak in the section serving as the leading (jth) reference peak and the last peak in the section serving as the trailing ($j + 1$st) reference peak. This becomes evident when we consider that the single-reference peak method as commonly used is a special case of the dual-reference peak method in which the retention time of the first ($j = 1$) reference peak is assumed to be zero.

The foregoing discussion illustrates the absolute necessity of investigating the characteristics of the α curve for a given chromatographic separation if the single-reference method is to be used to normalize the data. In Fig. 1, the origin for the retention time range of interest is zero, and either the single-reference method or the dual-reference method can be used on the original data with equally valid results. However, if the situation were such that a noticeable time offset could be introduced into the data, the origin would have to be established at the position of the first peak in the series to obtain a reasonable representation of α. In this case, only the dual-reference method would still give good results with the measured, uncorrected data. The following example will illustrate why this offset does not affect the normalized retention times when the dual-reference peak method is used.

For peaks located between reference peaks j and $j + 1$, the normalized retention time is calculated using eqn. 5. Assume some small but non-negligible offset, c, was introduced into the retention time data for all peaks. Thus the observed retention times for peak i and the two reference peaks j and $j + 1$ would be $t_i + c$, $t_j + c$, and $t_{j+1} + c$, respectively. However, upon normalization,

$$t_i^* = \frac{(t_i + c) - (t_j + c)}{(t_{j+1} + c) - (t_j + c)} = \frac{t_i - t_j}{t_{j+1} - t_j} \tag{9}$$

which is the same as eqn. 5. The influence of the offset is eliminated by this normalization technique.

On the other hand, with the single-reference method, a variable, routinely occurring offset between the time the analysis began and the time the strip-chart record started could significantly diminish the reliability of the normalized data. For example, assume that the expected retention times $\bar{t}_i$ ($i = 1, \ldots, n$) for n peaks in a given chromatogram are known. If peak n is the reference peak, then the expected normalized retention times for all other peaks in the chromatogram would be $\bar{t}_i/\bar{t}_n$ ($i = 1, \ldots, n - 1$). However, if each analysis j is characterized by some constant offset, c_j, and the measured retention times, corrected for this offset, differed from the expected by a constant ratio α_j, then the observed retention times for run j would be $\alpha_j\bar{t}_i + c_j$ ($i = 1, \ldots, n$). The normalized retention times would be calculated as $(\alpha_j\bar{t}_i + c_j)/(\alpha_j\bar{t}_n + c_j)$ for $i = 1, \ldots, n - 1$. Thus the method would only be valid if $c_j \ll \alpha\bar{t}_i$ for all i.

The characteristics of the dual-reference technique for normalizing retention times in an elution time interval where a linear or quadratic approximation to the α curve is valid have been discussed in a previous publication[8]. Modifying the results of that study by choosing $\bar{t}$ (the expected retention time) rather than the elution volume V as the independent variable in eqn. 8, we have (from eqn. 23 of ref. 8):

$$t_i^* = \bar{t}_i^* [1 - \beta' (\bar{t}_{j+1} - \bar{t}_i)] \tag{10}$$

where β' is the ratio of the second to the first expansion coefficients in eqn. 8. In this case, the α curve must be linear and $\beta \ll 1$ for eqn. 10 to be valid. The normalization procedure depicted by the asterisk is defined in eqn. 5. Thus the difference between the measured and expected elution times after normalization is, from eqn. 10,

$$\bar{t}_i^* - t_i^* = \bar{t}_i^* [\beta' (\bar{t}_{j+1} - \bar{t}_i)] \tag{11}$$

The factor β' is the ratio of the slope to the intercept of the curve that approximates α throughout the elution time interval (t_j, t_{j+1}). If α were constant throughout the interval, then $\beta' = 0$ and $\bar{t}_i^* - t_i^* = 0$, as discussed previously.

If the measured retention times alone are used to identify peaks, apart from any normalization scheme, the uncertainty in the location of the peak can increase without bound, as eqn. 3 indicates. In fact, the uncertainty will increase with increasing elution time unless $| \alpha(\bar{t}) - 1 | \to 0$ quite rapidly as $\bar{t} \to \infty$. However, normalizing the elution times and using multiple sets of reference peaks lead to a bounded uncertainty factor for the variation in the normalized retention times. In addition, the uncertainty factor is now independent of $\bar{t}$.

To investigate the characteristics of the variations in the normalized elution times, we examine the function given in eqn. 11. Note that, on the elution time interval (t_j, t_{j+1}),

$$\bar{t}_i^* - t_i^* = 0 \text{ for } i = j \tag{12a}$$

since, by definition, $t_j^* = 0$ (see eqn. 5). Furthermore, by examination of eqn. 11,

$$\bar{t}_i^* - t_i^* = 0 \text{ for } i = j + 1. \tag{12b}$$

Therefore, since the function $\bar{t}_i^* [\beta' (\bar{t}_{j+1} - \bar{t}_i)]$ is well behaved on the interval (t_j, t_{j+1}), Rolles theorem[11] guarantees a maximum or minimum value for the function on this interval. In addition, since $\bar{t}_i^* \geqslant 0$ and $(\bar{t}_{j+1} - \bar{t}_j) \geqslant 0$, and they are both monotonically increasing and decreasing functions, respectively, on the interval, we expect no more than one maximum or minimum to occur. The sign of β' will determine whether a maximum or minimum occurs.

Taking the derivative of eqn. 11 and setting it equal to zero, we find

$$(\bar{t}_i)_{\max.} = \left[\frac{t_{j+1} + t_j}{2}\right] \tag{13}$$

i.e., the value of $\bar{t}_i$ where the absolute value of the function given in eqn. 11 is greatest is halfway between the expected values for the two bracketing reference peaks. Thus the normalization technique is most uncertain at a point midway between the two reference peaks.

Substituting this value for $(\bar{t}_i)_{\max.}$ into eqn. 11, we obtain

$$(\bar{t}_i^* - t_i^*)_{\max.} = \frac{\beta'}{4} (\bar{t}_{j+1} - \bar{t}_j). \tag{14}$$

The largest experimentally determined value of β found in the retention time interval (t_j, t_{j+1}) could be used in eqn. 14 to establish the tolerance limits for the comparison of normalized data in this interval. The tolerance levels may be different for each interval. This uncertainty factor could be used in a computer scheme designed to match normalized retention times against expected values for identification purposes.

CONCLUSIONS

In a computer scheme designed to identify chromatographic peaks, decisions with respect to peak identification must be based on a reasonably limited body of data. Therefore, it is expedient to enhance the reliability of the data as much as possible. As a general rule, some type of normalization technique is desirable. Since the dual-reference peak method eliminates any problems associated with a constant offset between the actual retention times and the retention times measured from the strip chart or recorded by a data collecting system, it is the method of choice. Further, it is the only method that will work when the curvature of the α plot as described above becomes severe.

A normalization technique involves the use of reference peaks. These peaks must be located by the computer. Since the retention times of the reference peaks are not normalized, computer identification of reference peaks must be based on the actual elution times. However, elution times are inherently uncertain, and additional characteristics such as predetermined peak height or the ratio of absorbance readings at two different wavelengths would be very useful to aid in identifying these peaks. The exact location of the reference peaks is crucial to the identification scheme as a whole. Great care must be exercised in developing this aspect of the computer program. Since the development of a computer technique to unambiguously locate all reference peaks in the chromatogram is difficult to achieve, it is advantageous to have as few reference peaks as possible. The optimal choice of reference peaks is accomplished by the procedure mentioned above of locating those regions of the chromatogram in which the slope of the α curve is nearly zero.

ACKNOWLEDGEMENT

This research was sponsored by the National Institute of General Medical Sciences and the U.S. Atomic Energy Commission.

REFERENCES

1 D. D. Chilcote, *Clin. Chem.*, 19 (1973) 826.
2 R. Taylor and M. G. Davies, *Anal. Biochem.*, 51 (1972) 180.
3 H. D. Spitz, G. Henyon and J. N. Silvertson, *J. Chromatogr.*, 68 (1972) 111.
4 H. L. Back, P. J. Butlery and K. Gregson, *J. Chromatogr.*, 68 (1972) 103.
5 H. D. Hobcrecht and R. J. Klimowski, *Chromatogr. Newslett.*, 1 (1972) 23.
6 J. E. Purcell, E. W. March and R. Adams, *Chromatogr. Newslett.*, 1 (1972) 14.
7 R. L. Jolley and C. D. Scott, *J. Chromatogr.*, 47 (1970) 272.
8 D. D. Chilcote and C. D. Scott, *Anal. Chem.*, 45 (1973) 721.
9 D. D. Chilcote and J. E. Mrochek, *Anal. Lett.*, 6 (1973) 531.
10 W. C. Butts, *Anal. Biochem.*, 46 (1972) 187.
11 G. B. Thomas, *Calculus and Analytic Geometry*, Addison-Wesley, Reading, Mass., 1960, p. 134.