Journal of Chromatography, 449 (1988) 41–52
Elsevier Science Publishers B.V., Amsterdam — Printed in The Netherlands

CHROM. 20 704

EXPERIMENTAL AFFIRMATION OF THE STATISTICAL MODEL OF OVERLAP

JOE M. DAVIS
Department of Chemistry and Biochemistry, Southern Illinois University, Carbondale, IL 62901 (U.S.A.)
(First received April 12th, 1988; revised manuscript received May 23rd, 1988)

SUMMARY

This article briefly reviews the statistical model of overlap and reports results from its first application to a fully characterized experimental chromatogram in which the numbers of detectible components, singlet peaks, doublet peaks, and triplet peaks are known. The overall agreement between these numbers and their theoretically expected values, which were predicted by the model from the relative distribution of maxima in the chromatogram, is very good. Specifically, the absolute errors between the predicted and actual numbers are all less than or equal to three.

INTRODUCTION

This article reports results from an application of the statistical model of overlap (SMO) to an experimental multicomponent chromatogram containing known numbers of detectible singlet and multiplet peaks. The SMO, proposed some years ago[1], is a simple theory of peak overlap that rests on the assumption that the components of complex mixtures elute randomly from chromatographic columns or beds. In this case, the expected number of peaks in the chromatogram can be calculated from theory, if the number of detectible mixture components and the column peak capacity are known. Alternatively, and more importantly, the number of detectible mixture components can be estimated by a simple procedure reviewed below, if the number of peaks in the chromatogram and the peak capacity are fit as experimental data to the theory by least-squares methods. The basic validity of this procedure was confirmed by its extensive application to computer-generated chromatograms[2–6]. These applications furthermore established criteria (which are restated below) by which one can evaluate the accuracy of the component-number estimate. The procedure has been applied to several gas and liquid chromatograms[4,6–8], with results that are apparently consistent with those obtained from computer-generated chromatograms. This work is only one of many studies of peak overlap and its probability in chromatograms[9–16].

One can furthermore estimate with the SMO the expected numbers of singlet, doublet, triplet, etc., peaks in the chromatogram, once the number of detectible components is known[1]. Because a fundamental objective of chromatography is to resolve mixture components into singlet peaks, the estimation of these numbers enables one to gauge quantitatively the overall efficiency of separation.

0021-9673/88/$03.50

Perhaps the most significant implication of the SMO is the surprising severity of overlap expected, even in ultrahigh-resolution chromatograms[1]. Specifically, the numbers of resolved peaks and singlet peaks are never expected to exceed 36.8 and 18.4%, respectively, of the peak capacity. These predictions raise grave questions about the accuracy of component concentrations determined from chromatograms of complex mixtures and the feasibility of isolating select components from such mixtures. The possible overestimation of component concentrations in forensic and environmental mixtures is especially serious, as complex legal issues may be decided erroneously by failure to consider properly the likelihood of overlap.

The reported applications of the SMO to experimental chromatograms have principally been diagnostic evaluations of partially characterized mixtures, in which the numbers of detectible components are unknown. No results have been reported from an application of the SMO to an experimental chromatogram in which the numbers of detectible components, singlet peaks, and multiplet peaks are known. The results from one such application are reported here, principally to enhance confidence in the model and the procedure reviewed below. The excellent agreement between experimental component and peak numbers and their predicted values should emphasize the basic statistical pitfalls one commonly encounters when resolving complex mixtures.

THEORY

The actual number m of single component-peaks (SCPs), each of which corresponds to a detected mixture component, in a complex chromatogram is unknown, because some of the SCPs overlap with one another. Only the statistically expected number $\bar{m}$ of SCPs in the chromatogram can be estimated with the SMO; fortunately, in most cases, $m \approx \bar{m}$. When m is large [*i.e.*, > 30 (4)] and the components are randomly distributed along the elution axis, the number p of peaks expected in the continuous region X of the chromatogram is related to $\bar{m}$ by[1]

$$p = \bar{m}\, e^{-\bar{m}/n_c} = \bar{m}\, e^{-\alpha} \tag{1}$$

where $\alpha = \bar{m}/n_c$ is the component saturation of the chromatogram and n_c, the peak capacity, is the maximum number of uniformly spaced SCPs separable in region X

$$n_c = \frac{X}{x_0} = \frac{X}{4\sigma R_s^*} \tag{2}$$

In eqn. 2, x_0 is the span between adjacent SCPs resolved to the arbitrarily chosen resolution R_s^*, and σ is the standard deviation of any representative SCP in region X. Taking the logarithm of eqn. 1 and combining the result with eqn. 2, one obtains

$$\ln p = \ln \bar{m} - \bar{m}x_0/X = \ln \bar{m} - \bar{m}/n_c \tag{3}$$

Thus, a plot of $\ln p$ *vs.* x_0/X or $1/n_c$ is a line of slope $-\bar{m}$ and intercept $\ln \bar{m}$. Two independent estimates of $\bar{m}$ are consequently obtained, one (termed m_{sl}) from the slope

and the other (termed m_{in}) from the intercept. Although it is strictly incorrect to estimate independently these two $\bar{m}$ values[17], the requirement that m_{sl} should equal m_{in}, at least within statistical error, is one of several useful criteria by which to evaluate the accuracy of the $\bar{m}$ estimate[2-5].

As noted in the introduction, one can also calculate from theory the expected numbers of singlet peaks s, doublet peaks d, triplet peaks t, and higher order multiplet peaks in the chromatogram, once $\bar{m}$ is known. The first of these expected numbers are[1]

$$s = \bar{m}\, e^{-2\alpha} \tag{4a}$$

$$d = \bar{m}\, e^{-2\alpha}\,(1 - e^{-\alpha}) \tag{4b}$$

$$t = \bar{m}\, e^{-2\alpha}\,(1 - e^{-\alpha})^2 \tag{4c}$$

The resolution R_s^* one chooses to define n_c, and thus α, clearly determines the numbers of peaks, singlets, and multiplets that one actually calculates from eqns. 1 and 4a–c. One arbitrary but useful convention is to equate R_s^* with 0.5, in which case peaks, singlets, and multiplets correspond to visually distinguishable chromatographic maxima[2-4]. This convention will be adopted here.

A common objective of chromatography is to resolve mixture components into singlet peaks of high purity. Usually the achievement of near-baseline resolution, characterized by $R_s^* = 1.5$, between the SCP of interest and its adjacent neighbors is adequate for this task. In this case the relative component saturation $\alpha = 4\sigma R_s^* \bar{m}/X$ is three times greater than that based on $R_s^* = 0.5$. The expected number s_b of baseline-resolved singlets is consequently determined from eqn. 4a as

$$s_b = \bar{m}\, e^{-6\alpha} \tag{5}$$

where, as stated above, α is defined by $R_s^* = 0.5$.

Several procedures, which differ only in the minimal resolution R_s^* that one requires between adjacent peaks, have been proposed to determine experimental peak numbers from chromatograms[2-6]. The procedure reviewed here permits one to determine several effective peak numbers and to estimate $\bar{m}$ from the relative positions of the chromatographic maxima in a single chromatogram[5]. Furthermore, it eliminates the need to determine the peak capacity n_c, which is rigorously sample-dependent[16]. For any arbitrarily chosen spacing x_0, the number of spans between adjacent maxima that equal or exceed this spacing is equated to the experimental number p corresponding to that x_0. (Throughout the text, the symbols p, s, d, t, etc., will be used to represent both theoretical and experimental quantities.) The inclusive span between the first and last maxima in the region is identified with the quantity X. The procedure is repeated with several different x_0 values and a data set $(x_0/X, \ln p)$ is generated. These arbitrary changes in x_0 in effect correspond to arbitrary changes in the resolution R_s^* that formally resolves clusters of SCPs into peaks, as rigorously defined by theory[2]; these resultant peaks, however, have little in common with peaks in any traditional sense (*e.g.*, chromatographic maxima). When $x_0 > 2\sigma$, the number p increases with decreasing x_0. When $x_0 < 2\sigma$ (*i.e.*, when $R_s^* < 0.5$), however, p equals one less than the number p_m of chromatographic maxima, because adjacent SCPs

cannot be visually resolved when $R_s^* < 0.5$. These data are excluded from the data set, which is then fit to eqn. 3 by least-squares methods. Theoretical weights can be assigned to each point in the fit to estimate the uncertainties in m_{sl} and m_{in}. Details of the method[5] and previously reported applications[7,8] are given elsewhere.

EXPERIMENTAL PROCEDURES

No original chromatography is reported here. Fig. 1 (reproduced with permission) is a high-resolution chromatogram of a composite mixture of 113 polynuclear aromatic hydrocarbon (PAH) standards first reported by Lee *et al.*[18] and kindly brought to the author's attention by M. Lee. These standards were fractionated on a 12 m × 0.29 mm I.D. glass capillary, which was coated with a 0.34-μm film of SE-52 and temperature-programmed from 50 to 250°C at 2°C/min. The helium carrier flow-rate lay between 1 and 3 ml/min.

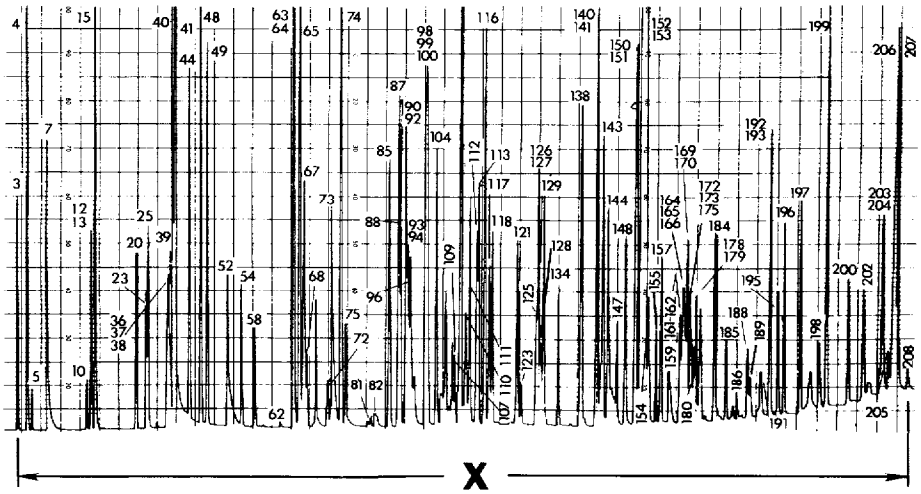

Fig. 1. Chromatogram of 113 PAH standards. Analytical conditions: column, 12 m × 0.29 mm I.D. glass capillary coated with a 0.34-μm film of SE-52; temperature program, 50–250°C at 2°C/min; carrier, helium; flow-rate, between 1 and 3 ml/min. Reprinted with permission (ref. 18).

Each PAH standard was chromatographed several times to determine reliably a retention index[18]. Thus, the retention times of all standards were both accurately and precisely known, enabling the authors to identify the chromatographic maxima in Fig. 1 as singlets or multiplets. Each maxima is associated with one, two, or three numbers; singlets are associated with a single number (*e.g.*, 3), doublets with two numbers (*e.g.*, 12–13), and triplets with three numbers (*e.g.*, 36–37–38). These numbers originally referred one to a table in reference 18, in which the identities of the mixture components corresponding to the numbered maxima are reported. (The maxima in the reprinted figure were actually renumbered by the author; the numbers reported in reference 18 are somewhat difficult to read, unless the figure occupies a full journal page, as it does in the reference.) A total of 92 maxima, comprised of 75 singlets, 13

doublets, and 4 triplets, are numbered in the chromatogram. Of the 75 singlets, 36 are also baseline-resolved, as judged by visual inspection. Four of these latter singlet types (44, 48, 49, and 54) actually reside on the tailing edges of the two maxima, 41 and 52. They were nevertheless counted as baseline-resolved singlets, because the tailing is highly unrepresentative of the chromatogram as a whole. These various singlet and multiplet numbers are reported in Table I. At least ten unnumbered maxima are also present in the chromatogram; their influence on the analysis is addressed below.

TABLE I

NUMBERS OF PEAK, SINGLET, AND MULTIPLET MAXIMA IN CHROMATOGRAM IN FIG. 1

Numbers in brackets identify numbered maxima in Fig. 1.

Detectable components: 113

Detectable (numbered) peak maxima: 92

Singlets s: 75
[3, 4, 5, 7, 10, 15, 20, 23, 25, 39, 40, 41, 44, 48, 49, 52, 54, 58, 62, 65, 67, 68, 72, 73, 74, 75, 81, 82, 85, 87, 88, 96, 104, 107, 109, 110, 111, 112, 113, 116, 117, 118, 121, 123, 125, 128, 129, 134, 138, 143, 144, 147, 148, 154, 155, 157, 159, 180, 184, 185, 186, 188, 189, 191, 195, 196, 197, 198, 199, 200, 202, 205, 206, 207, 208]

Doublets d: 13
[12–13, 63–64, 90–92, 93–94, 126–127, 140–141, 150–151, 152–153, 161–162, 169–170, 178–179, 192–193, 203–204]

Triplets t: 4
[36–37–38, 98–99–100, 164–165–166, 172–173–175]

Baseline-resolved singlets s_b: 36
[3, 4, 5, 7, 10, 15, 20, 44, 48, 49, 52, 54, 58, 62, 81, 82, 85, 104, 134, 138, 148, 155, 157, 159, 184, 185, 186, 191, 195, 196, 197, 198, 199, 200, 202, 208]

ANALYTICAL PROCEDURES

An enlarged copy of the chromatogram in Fig. 1 was prepared by careful xerography. The sequential relative positions of the numbered chromatographic maxima (which are analogous to relative retention times or volumes) were measured to a resolution of 0.005 in. with a True Grid 1011 Digitizer (Houston Instruments, Austin, TX, U.S.A.) and stored on an Apple IIe microcomputer. The digitization process was repeated to verify that the maxima positions could be determined reproducibly; the corresponding maxima positions in the two sets of numbers differed at most by 0.01 in. The positions of the unnumbered maxima were not digitized. Because these maxima are more or less randomly dispersed throughout the chromatogram, this small bias negligibly affects the analysis, as the results below will show.

The span X indicated in Fig. 1 was estimated as the difference, 10.100 in., between the first and last maxima. The data set $(x_0/X, \ln p)$ was then determined as described above. No value of $p < 16$ was considered, because of shortcomings in the SMO's development[5]. The data set was then graphed and inspected.

Only a subset of these data were fit to eqn. 3. Because the relative positions of any two adjacent maxima were both determined with a digitization error of 0.005 in., the relative error in small x_0 values is considerable. If one demands that this relative error be less than 0.1, then

$$\frac{\sigma_{x_0}}{x_0} = \frac{\sqrt{(0.005)^2 + (0.005)^2}}{x_0} < 0.1 \tag{6}$$

where σ_{x_0} is the standard deviation of x_0. The span x_0 must exceed 0.07 in. to satisfy this inequality. To minimize procedural errors, therefore, no data for which $x_0/X < 0.07/10.100 \approx 0.007$ were fit to eqn. 3. These data include those for which $x_0 < 2\sigma$, which would normally be excluded anyway (see above).

The remaining data were fit to eqn. 3 by the theory of least squares[19]. The details of the fit are presented in the appendix.

RESULTS AND DISCUSSION

As earlier work has demonstrated, a number of quantitative criteria must be satisfied, if one is to calculate $\bar{m}$ accurately (*i.e.*, to within 10%) by the procedure detailed above. The first, as noted earlier, is that $m_{sl} \approx m_{in}$. Secondly, the relative saturation α of the chromatogram must be less than 0.5, when $R_s^* = 0.5$ (ref. 5). Finally, the distribution of distances between adjacent chromatographic maxima must be exponential, when $x_0 > 2\sigma$ (ref. 5). One quantitative test for the existence of this distribution is the value of the reduced chi-square statistic χ_ν^2, defined by eqn. A15 in the appendix, which is a measure of the goodness of fit of the data set (x_0/X, ln p) to eqn. 3. The test is inherently statistical, but when $\chi_\nu^2 < 1$, the distances are most likely exponentially distributed[19].

Fig. 2 is the plot of ln p *vs.* x_0/X generated by the procedures reported in the appendix. The solid line is a weighted least-squares fit of the indicated points to eqn. 3. As one would anticipate, these points are randomly scattered about the fit. The peak numbers for which $2\sigma/X < x_0/X < 0.007$, however, are systematically smaller than predicted by this fit, perhaps because of digitization error.

Table II reports, to the nearest whole number, the numbers of singlet peaks s, doublet peaks d, triplet peaks t, baseline-resolved singlet peaks s_b, and components m_{sl}, and $\bar{m}$, which were calculated as detailed in the appendix. Also reported, to the nearest whole number, are the standard deviations σ_s, σ_d, σ_t, σ_{s_b}, σ_{sl}, σ_{in}, and $\sigma_{\bar{m}}$ of these respective numbers, as evaluated from equations in the appendix. The saturation α of the chromatogram, and the standard deviation σ_α of the saturation, are also tabulated, for $R_s^* = 0.5$. The reported standard deviations are the statistical uncertainties in the estimates derived from the plot of ln p *vs.* x_0/X.

In examining the data reported in Table II, one observes that $m_{sl} \approx m_{in}$. A simple Student's t-test indicates that these numbers are statistically equivalent. In addition, $\alpha = 0.174$, which is much less than the $\alpha = 0.5$ limit above which $\bar{m}$ cannot be reliably estimated. The reduced chi-square statistic for the fit is $\chi_\nu^2 = 1.05$, which suggests that the distances between adjacent maxima are indeed exponentially distributed. All the quantitative prerequisites to the calculation of accurate $\bar{m}$ values stated above are

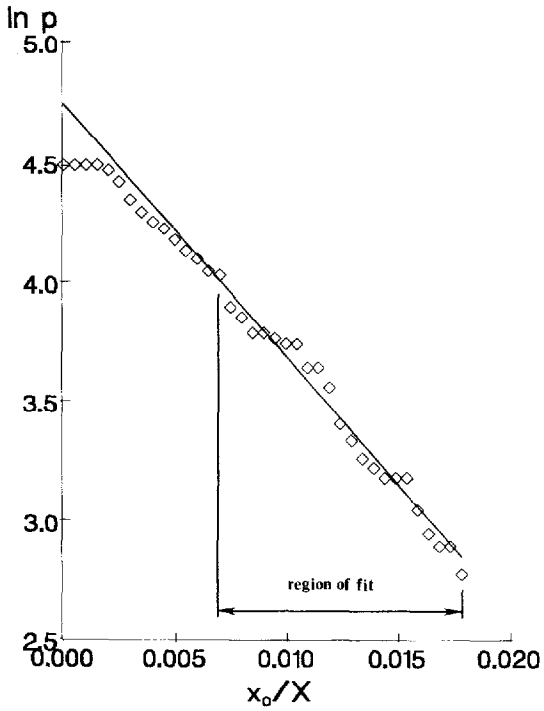

Fig. 2. Plot of ln p *vs.* x_0/X derived from the chromatogram in Fig. 1. Solid line is least-squares fit of indicated points to eqn. 3.

consequently satisfied, and one would thus expect these predicted numbers to be fairly good estimates.

The predictions reported in Table II are, in fact, in very good agreement with the experimental peak and component numbers reported in Table I and parenthetically in Table II. Specifically, the experimental numbers m, s, d, and s_b all lie within, and usually well within, one standard deviation, and the number t within two standard deviations, of the predicted values. The absolute errors between the predicted and actual numbers are very small, and are all less than or equal to three.

TABLE II

RESULTS FROM SMO APPLICATION TO CHROMATOGRAM IN FIG. 1

Expected numbers (standard deviations) of components m_{sl} (σ_{sl}) and m_{in} (σ_{in}) from slope and intercept, respectively, of eqn. 3, SCPs $\bar{m}$ ($\sigma_{\bar{m}}$), singlets s (σ_s), baseline-resolved singlets s_b (σ_{s_b}), doublets d (σ_d), and triplets t (σ_t) are reported to the nearest whole number. Saturation α and standard deviation σ_α are reported to three significant figures. Experimental component and peak numbers are reported in parentheses below the theoretical results.

$m_{sl} \pm \sigma_{sl}$	$m_{in} \pm \sigma_{in}$	$\bar{m} \pm \sigma_{\bar{m}}$	$\alpha \pm \alpha_\alpha$ ($R_s^* = 0.5$)	$s \pm \sigma_s$	$d \pm \sigma_d$	$t \pm \sigma_t$	$s_b \pm \sigma_{s_b}$
106 ± 7	115 ± 9	110 ± 6	0.174 ± 0.051	77 ± 4	12 ± 3	2 ± 1	38 ± 10
(–)	(–)	(113)	(–)	(75)	(13)	(4)	(36)

Some interesting insights are obtained by examining the fraction of the peak capacity n_c utilized in this chromatogram to resolve peaks and singlets. Because the α reported in Table II is defined with respect to chromatographic maxima (*i.e.*, with respect to $R_s^* = 0.5$), the approximate peak capacity n_c of the chromatogram, in terms of maxima, is $m/\alpha \approx 113/0.174 \approx 649$ components. In other words, 649 uniformly spaced maxima (of equal amplitude) could theoretically be resolved in space X, when $R_s^* = 0.5$. Only 92 maxima are actually resolved in this space, however, and the fraction of n_c utilized is only $p_m/n_c = 92/649 = 0.142$. The fraction of n_c utilized in resolving singlet maxima is somewhat less, $s/n_c = 75/649 = 0.116$. These p_m/n_c and s/n_c ratios are considerably smaller than their respective theoretical upper limits, 0.368 and 0.184[1]. The utilization of the available peak capacity for maxima separation is clearly far from optimized.

As shown by eqn. 2, the numerical value of n_c is determined by the resolution R_s^* that one chooses to discriminate between peaks. The peak capacity of any chromatogram, in terms of baseline-resolved peaks (defined by $R_s^* = 1.5$), is consequently three times smaller than that based on chromatographic maxima (defined by $R_s^* = 0.5$). Hence, the approximate baseline-resolved peak capacity of the chromatogram is $649/3 \approx 216$. The fraction of this capacity utilized in separating baseline-resolved singlets, s_b/n_c, is $36/216 \approx 0.167$. This ratio is only slightly less than the largest peak-capacity fraction, 0.184, that can be used for this purpose. This chromatogram therefore represents a highly optimal utilization of the available peak capacity for the baseline resolution of randomly spaced singlet peaks.

The peak and component numbers presented above clearly affirm that the SMO accounts quantitatively for peak overlap in this chromatogram. This analysis furthermore supports the hypothesis that the SMO accounts for overlap in many chromatograms, when certain well established criteria are met. As observed elsewhere, however, the SMO does not apply to all high-resolution chromatograms, especially when the mixture components exhibit order at the molecular level[1]. Furthermore, the agreement between experiment and theory, under one set of experimental conditions, does not imply that agreement will also be attained if the conditions are radically changed. For example, the careful optimization of multicomponent separations can, in some cases, introduce order and obviate the predictions of the SMO[13,15]. Other optimization procedures yield chromatograms in which peak overlap remains consistent with the SMO[6]. Each case must be addressed individually.

In conclusion, I note that additional studies of this type would provide useful data to test the reliability and universality of the procedures reviewed here or described elsewhere. A large set of experimental data is essential to the thorough testing and characterization of the SMO, because the effects on its predictions of many chromatographic variables, *e.g.*, mixture composition, mobile- and stationary-phase composition, and mobile-phase programming, cannot be deduced simply from analyses of computer-generated chromatograms. Because computer simulations furthermore are dismissed by some as idealistic approaches to the study of real-world problems, additional experimental confirmations of the SMO, and other theories of peak overlap, would build a solid experimental foundation on which to rest the sobering conclusions deduced from these theories.

ACKNOWLEDGEMENT

The author thanks Southern Illinois University for the start-up funds with which this research was conducted.

APPENDIX

Least-squares fit of experimental data to eqn. 3

The appropriate subset of the data set (x_0/X, ln p) was fit to eqn. 3, as detailed below, to calculate m_{sl} and m_{in} and their respective standard deviations, σ_{sl} and σ_{in}. These numbers were evaluated from standard formulae for the least-squares fit of data to a straight line[19]

$$m_{sl} = \frac{1}{\Delta}\left[\sum_{i=1}^{j} w_i \sum_{i=1}^{j} w_i \left(\frac{x_{0_i}}{X}\right) \ln p_i - \sum_{i=1}^{j} w_i \left(\frac{x_{0_i}}{X}\right) \sum_{i=1}^{j} w_i \ln p_i\right] \tag{A1}$$

$$m_{in} = \frac{1}{\Delta}\left[\sum_{i=1}^{j} w_i \left(\frac{x_{0_i}}{X}\right)^2 \sum_{i=1}^{j} w_i \ln p_i - \sum_{i=1}^{j} w_i \left(\frac{x_{0_i}}{X}\right) \sum_{i=1}^{j} w_i \left(\frac{x_{0_i}}{X}\right) \ln p_i\right] \tag{A2}$$

$$\sigma_{sl} = \frac{1}{\Delta} \sum_{i=1}^{j} w_i \tag{A3}$$

$$\sigma_{in} = \frac{1}{\Delta} \sum_{i=1}^{j} w_i \left(\frac{x_{0_i}}{X}\right)^2 \tag{A4}$$

$$\Delta = \sum_{i=1}^{j} w_i \sum_{i=1}^{j} w_i \left(\frac{x_{0_i}}{X}\right)^2 - \left[\sum_{i=1}^{j} w_i \left(\frac{x_{0_i}}{X}\right)\right]^2 \tag{A5}$$

In eqns. A1–A5, j is the number of data points in the fit (in this case, 23) and w_i, x_{0_i}, and p_i are the ith values of the theoretical weight w, arbitrary spacing x_0, and experimental number p. The weight w of each point was calculated as[5]

$$w = \frac{\bar{m}\, e^{-2\alpha'}}{f(\alpha') + (1 - \alpha')^2\, e^{-2\alpha'}} \tag{A6}$$

where

$$f(\alpha') = \frac{0.099}{\alpha'} \exp\left[-\frac{1}{2.341}\left(\ln \frac{\alpha'}{1.168}\right)^2\right] \tag{A7}$$

and

$$\alpha' = \bar{m}x_0/X \tag{A8}$$

Because the weights w depend on the unknown quantity $\bar{m}$ (see eqns. A6–A8), the data set was fit iteratively to eqns. A1–A8 until the calculated results varied negligibly between iterations. Specifically, the arbitrary weight $w = 1$ was initially assigned to all points, and the unknowns m_{sl}, m_{in}, σ_{sl} and σ_{in} were then evaluated from eqns. A1–A5. The statistical component number $\bar{m}$ was then approximated as

$$\bar{m} = (m_{sl}/\sigma_{sl}^2 + m_{in}/\sigma_{in}^2)/(\sigma_{sl}^{-2} + \sigma_{in}^{-2}) \tag{A9}$$

A series of new weights w was then computed from eqns. A6–A8 with the $\bar{m}$ value estimated from eqn. A9, and the unknowns m_{sl}, m_{in}, σ_{sl}, σ_{in}, and $\bar{m}$ were again evaluated as detailed above. This iterative procedure was repeated until m_{sl}, m_{in}, σ_{sl}, and σ_{in} varied by less than 0.001% between successive iterations. The weighted standard deviation $\sigma_{\bar{m}}$ of $\bar{m}$ was then estimated from these data as

$$\sigma_{\bar{m}} = (\sigma_{sl}^{-2} + \sigma_{in}^{-2})^{-1/2} \tag{A10}$$

The numbers m_{sl}, m_{in}, σ_{sl}, σ_{in}, $\bar{m}$, and $\sigma_{\bar{m}}$ reported in Table II are the converged values determined as described above.

The relative saturation α of the chromatogram in Fig. 1 was then calculated from eqn. 1 as

$$\alpha = -\ln (P_m/\bar{m}) \tag{A11}$$

where $p = p_m = 92$, the number of chromatographic maxima, and $\bar{m}$ is given by eqn. A9. With this choice, peaks, singlets, and multiplets are all identified with chromatographic maxima and $R_s^* \approx 0.5^{2-4}$.

The expected numbers of singlet, doublet, and triplet peaks were then calculated from $\bar{m}$ (eqn. A9), α (eqn. A11), and eqns. 4a–c. The standard deviations of these n-lets were estimated from a propagation of errors as[19]

$$\sigma_{n\text{-let}} = e^{-2\alpha}(1 - e^{-\alpha})^{n-1}\left[\left(1 - \frac{(n-1)\,e^{-\alpha}}{1 - e^{-\alpha}}\right)^2\right]^{1/2}\sigma_{\bar{m}} \tag{A12}$$

where n is the number of SCPs per maxima (*e.g.*, $n = 2$ for a doublet). The standard deviations of the singlet s, doublet d, and triplet t peak numbers are designated σ_s, σ_d, and σ_t, respectively, in Table II.

The number s_b of baseline-resolved peaks was calculated from $\bar{m}$ (eqn. A9), α (eqn. A11), and eqn. 5. The standard deviation σ_{s_b} of the number s_b of baseline-resolved singlets was determined from a propagation of error to be

$$\sigma_{s_b} = 5e^{-6\alpha}\sigma_{\bar{m}} \tag{A13}$$

The standard deviation σ_α of parameter α, eqn. A11, was calculated as

$$\sigma_\alpha = \frac{\sigma_{\bar{m}}}{\bar{m}} \tag{A14}$$

In eqns. A12–A14, $\bar{m}$, $\sigma_{\bar{m}}$, and α are given by eqns. A9–A11, respectively.

A reduced chi-square statistic χ_v^2 was calculated as[19]

$$\chi_v^2 = \sum_{i=1}^{j} \frac{w_i[\ln(\bar{m}\, e^{-\alpha'}) - \ln p_i]^2}{j - 2} \tag{A15}$$

where w, α', and $\bar{m}$ are given by eqns. A6, A8, and A9, respectively. The value of χ_v^2 is a measure of the goodness of fit of the data set (x_0/X, ln p) to eqn. 3.

LIST OF SYMBOLS

d	statistically expected or experimental number of doublet peaks
f(α')	function defined by eqn. A7
m	number of detectible components in chromatogram
m_{sl}	component number estimated from slope of eqn. 3
m_{in}	component number estimated from intercept of eqn. 3
$\bar{m}$	statistically expected number of detectible components in chromatogram
n_c	peak capacity
p	statistically expected number of peaks in chromatogram, or experimental number of spans between adjacent maxima greater than x_0
p_m	number of chromatographic maxima
R_s^*	resolution that resolves adjacent SCPs into separate peaks
s	statistically expected or experimental number of singlet peaks
s_b	statistically expected or experimental number of baseline-resolved singlet peaks
SCPs	single-component peaks
t	statistically expected or experimental number of triplet peaks
w	statistical weight of point in ln p *vs.* x_0/X plot
x_0	arbitrary spacing
X	span between first and last chromatographic maxima
α	$\bar{m}/n_c$
α'	$\bar{m}x_0/X$
Δ	function defined by eqn. A5
σ	standard deviation of SCP
σ_d	standard deviation of doublet number d
σ_{in}	standard deviation of m_{in}
$\sigma_{\bar{m}}$	standard deviation of $\bar{m}$
$\sigma_{n\text{-let}}$	standard deviation of n-let peak number
σ_s	standard deviation of singlet number s
σ_{s_b}	standard deviation of baseline-resolved singlet number s_b

σ_{sl} standard deviation of m_{sl}
σ_t standard deviation of triplet number t
σ_{x_0} uncertainty in x_0 due to digitization error
σ_α standard deviation of α
χ^2_υ reduced chi-square statistic

REFERENCES

1 J. M. Davis and J. C. Giddings, *Anal. Chem.*, 55 (1983) 418.
2 J. C. Giddings, J. M. Davis and M. R. Schure, in S. Ahuja (Editor), *Ultrahigh Resolution Chromatography (ACS Symposium Series* 250), American Chemical Society, Washington, DC, 1984, p. 9.
3 J. M. Davis and J. C. Giddings, *J. Chromatogr.*, 289 (1984) 277.
4 D. P. Herman, M. F. Gonnord and G. Guiochon, *Anal. Chem.*, 56 (1984) 995.
5 J. M. Davis and J. C. Giddings, *Anal. Chem.*, 57 (1985) 2168.
6 F. Dondi, Y. D. Kahie, G. Lodi, M. Remelli, P. Reschiglian and C. Bighi, *Anal. Chim. Acta*, 191 (1986) 261.
7 J. M. Davis and J. C. Giddings, *Anal. Chem.*, 57 (1985) 2178.
8 S. Coppi, A. Betti and F. Dondi, *Anal. Chim. Acta*, submitted for publication.
9 D. Rosenthal, *Anal. Chem.*, 54 (1982) 63.
10 L. J. Nagels, W. L. Creten and P. M. Vanpeperstraete, *Anal. Chem.*, 55 (1983) 216.
11 M. Martin and G. Guiochon, *Anal. Chem.*, 57 (1985) 289.
12 M. Martin, D. P. Herman and G. Guiochon, *Anal. Chem.*, 58 (1986) 2200.
13 D. P. Herman, H. A. H. Billiet and L. de Galan, *Anal. Chem.*, 58 (1986) 2999.
14 W. L. Creten and L. J. Nagels, *Anal. Chem.*, 59 (1987) 822.
15 L. de Galan, D. P. Herman and H. A. H. Billiet, *Chromatographia*, 24 (1987) 108.
16 M. Z. El Fallah and M. Martin, *Chromatographia*, 24 (1987) 115.
17 J. M. Davis, *Ph.D. Thesis*, University of Utah, Salt Lake City, UT, 1985.
18 M. L. Lee, D. L. Vassilaros, C. M. White and M. Novotny, *Anal. Chem.*, 51 (1979) 768.
19 P. R. Bevington, *Data Reduction and Error Analysis for the Physical Sciences*, McGraw-Hill, New York, 1969.