Anal. Chem. **1995**, *67*, 2994–3003

Application of 2-D Statistical Theory of Overlap to Three Separation Types: 2-D Thin-Layer Chromatography, 2-D Gas Chromatography, and Liquid Chromatography/Capillary Electrophoresis

Kirk Rowe,[†] DeWayne Bowlin,[‡,§] Mingqin Zou,[‡] and Joe M. Davis*[,‡]

Department of Physics, Southern Illinois University at Carbondale, Carbondale, Illinois 62901-4401, and Department of Chemistry and Biochemistry, Southern Illinois University at Carbondale, Carbondale, Illinois 62901-4409

A modified statistical theory of overlap for two-dimensional (2-D) separations was applied to three types of 2-D separations: 2-D thin-layer chromatography (2-D TLC), 2-D gas chromatography (2-D GC), and liquid chromatography/capillary electrophoresis (LC/CE). The 2-D TLCs were developed in this laboratory on two adsorbents by overpressure layer chromatography and ascending TLC from a standard solution containing 30 polynuclear aromatic hydrocarbons. The other separations were given to us for interpretation by colleagues. The coordinate pairs of maxima in these separations were determined and interpreted using procedures that predict from them the number of mixture components. For 2-D TLCs developed on one adsorbent, this number agreed to within 5% of the known number of mixture components. For 2-D TLCs developed on another adsorbent, this number was incorrect but was shown to be incorrect by simulation. The numbers calculated from the other separations were internally consistent and also were consistent with simulations developed to mimic the separations. In particular, a portion of a 2-D GC of kerosene containing compounds of nine carbon atoms had modest overlap, and LC/CEs of tryptic digests of thyroglobulin and cytochrome *c* appeared to have virtually no overlap.

In the preceding paper, two of the authors of this work developed and tested by simulation a modified statistical theory of overlap for two-dimensional (2-D) separations.[1] Unlike earlier 2-D statistical theories,[2-4] the modified theory addresses the variation of component density throughout the separation and, in part, the correlation of retention times, elution times, etc. in the two dimensions. In this paper, the modified theory is used to estimate the number of mixture components in three different types of experimental 2-D separations: 2-D thin-layer chromatography (2-D TLC), 2-D gas chromatography (2-D GC), and liquid chromatography/capillary electrophoresis (LC/CE). The ability of theory to interpret such different separation types is a tribute to its versatility.

The principal purpose of this estimation is to determine the completeness of separation, which one cannot evaluate by simple inspection. More specifically, comparison of the number p_m of maxima in the separation to the number $\bar{m}$ of mixture components estimated by theory enables one to evaluate the separation's extent. If these numbers are close, then the separation is good; conversely, if $p_m \ll \bar{m}$, then the separation is poor.

In brief, 2-D TLCs were developed in our laboratory on two different adsorbents from a synthetic mixture containing a known number of polynuclear aromatic hydrocarbons. In this case, as in similar ones based on one-dimensional (1-D) separations,[5,6] the estimates' accuracy could be evaluated because the number of components was known. Examples of other separation types kindly were given to us by others for interpretation. Because the numbers of components in these separations were not known, we tested the predictions of theory by generating "look-alike" computer simulations using attributes determined by our statistical analyses. In most cases, the statistical theory provided accurate and consistent estimates of $\bar{m}$.

THEORY

The theory necessary for estimating $\bar{m}$ by statistical means is described elsewhere:[1] here, only essential details are reviewed. The average number p of maxima (spots) in a 2-D separation of rectangular shape and area A that contains, on average, $\bar{m}$ single-component spots (SCSs) distributed with a nonhomogeneous Poisson density is

$$p = 4\overline{\alpha m}\int_0^1\int_0^1 [f(\zeta,\eta)]^2 \frac{e^{-4f(\zeta,\eta)\bar{\alpha}}}{1-e^{-4f(\zeta,\eta)\bar{\alpha}}}\,d\zeta\,d\eta \qquad (1)$$

where $\bar{\alpha}$, $f(\zeta,\eta)$, ζ, and η are the average saturation of the separation, the frequency governing the density of SCSs, and the two reduced coordinates of the rectangular space, respectively. Equation 1 is the basis of a least-squares regression, by which $\bar{m}$ is predicted by minimizing the sum of squares, SS:

$$SS = \sum_{i=1}^{n}\left\{p'_i - 4\bar{m}^2\frac{A_{o_i}}{A}\int_0^1\int_0^1 [f_a(\zeta,\eta)]^2 \frac{e^{-4f_a(\zeta,\eta)\bar{m}A_{o_i}/A}}{1-e^{-4f_a(\zeta,\eta)\bar{m}A_{o_i}/A}}\,d\zeta\,d\eta\right\}^2 \qquad (2)$$

where p'_i and A_{o_i}/A are the ordinate and abscissa, respectively,

[†] Department of Physics.
[‡] Department of Chemistry and Biochemistry.
[§] Present address: Department of Chemistry, University of Illinois, Urbana, IL 61801.

(1) Rowe, K.; Davis, J. M. *Anal. Chem.* **1995**, *67*, 2981.
(2) Davis, J. M. *Anal. Chem.* **1991**, *63*, 2141.
(3) Oros, F. J.; Davis, J. M. *J. Chromatogr.* **1992**, *591*, 1.
(4) Shi, W.; Davis, J. M. *Anal. Chem.* **1993**, *65*, 482.
(5) Davis, J. M. *J. Chromatogr.* **1988**, *449*, 41.
(6) Delinger, S. L.; Davis, J. M. *Anal. Chem.* **1990**, *62*, 436.

0003-2700/95/0367-2994$9.00/0

of the ith of n p'-plot coordinates and $f_a(\zeta,\eta)$ is an approximation to frequency $f(\zeta,\eta)$. Both the p'-plot coordinates and $f_a(\zeta,\eta)$ are determined from the coordinates of maxima. Thus, all data necessary for the minimization of eq 2 are determined by experiment.

The intervals between the coordinates of maxima can be used to determine the number p' of clusters of overlapping circles having constant diameter $\beta d_o = 2(A_o/\pi)^{1/2}$ and centers defined by the maxima coordinates. For each cluster, all maxima in the cluster lie within span βd_o of at least one other maximum in the cluster, and all maxima not in the cluster are separated from all maxima in the cluster by spans greater than βd_o. For a series of diameters βd_o, a series of p'-plot coordinates, $(A_o/A,p')$, can be determined. These coordinates represent the data set to which eq 2 was fit. The graph of p' vs A_o/A is called a p'-plot.

The frequency $f_a(\zeta,\eta)$ in eq 2 was determined at discrete nodes, (ζ,η), by the dependent means of Procedure 2 discussed in ref 1:

$$f_a(\zeta,\eta) = A_T^{-1}\int_{\eta-\delta}^{\eta+\delta}\int_{\zeta-\delta}^{\zeta+\delta}\sum_{i=1}^{p_m} e^{-(\zeta-\zeta_{oi})^2/2\sigma^2}\, e^{-(\eta-\eta_{oi})^2/2\sigma^2}\, d\zeta\, d\eta \quad (3)$$

where A_T is the volume of the sum of p_m bi-Gaussians prior to normalization, σ is the constant standard deviation of a bi-Gaussian, (ζ_{oi},η_{oi}) are the coordinates of the ith of p_m bi-Gaussians, and 2δ is the spacing between adjacent nodes in either dimension. Here, σ was chosen to minimize SS, eq 2. The associated value of $\bar{m}$ was identified with the number of components.

A detailed description of theory is in ref 1.

Table 1. Identities and Concentrations of PNAs Comprising the 30-Component Standard

component	concn (μg/mL)[a]
1,5-dimethylnaphthalene	1610; 1680
2,3,6-trimethylnaphthalene	840; 1780
azulene	1020; 840
1-ethylnaphlene	650; 21 710
4,4′-dimethylbiphenyl	830; 1500
9-methylanthracene	640; 880
9,10-dimethylanthracene	1470; 810
2-phenylnaphthalene	930; 910
p-terphenyl	500; 860
9-methylphenanthrene	910; 1390
1,2,3,4-tetrahydrofluoranthene	2910; 1250
1-*n*-butylpyrene	1230; 840
benzo[*a*]fluorene	860; 280
1,2-dihydropyrene	860; 560
3-methylbenzo[*c*]phenanthrene	920; 1200
α,α-dinaphthylene	910; 1740
triphenylene	1000; 1800
9-phenylfluorene	930; 1020
2,2′-dinaphthyl	580; 1690
tetraphenylmethane	660; 940
α,α-dinaphthylmethane	830; 710
1-methylpyrene	850; 1360
1,10-methylenephenanthene	490; 1140
1,2,3,6,7,8-hexahydropyrene	1780; 3080
picene	1000; 930
benzo[*b*]fluoranthene	620; 980
1-(2′-phenanthrenyl)-2-(1″-naphthyl)ethane	620; 1130
dibenz[*a*,*h*]anthracene	560; 420
bianthryl	3060; 1660
biacenanaphthylene	810; 1010

[a] The first number corresponds to the first standard, the second number to the second standard.

PROCEDURES

Generation of 2-D TLC Separations. Two methylene chloride solutions of 30 polynuclear aromatic hydrocarbons (PNAs) having from 2 to 6 fused rings were prepared. Table 1 reports the PNA identities and their concentrations in the two solutions. These standards were kindly provided by Milton Lee of Brigham Young University.

Separations of PNAs by TLC have been developed on various adsorbents, including silica,[7-10] alumina,[7,11-13] cellulose,[11] acetylated cellulose,[11-16] polyamide,[16] and reverse-phase sorbents.[17-19] Most separations are one-dimensional; 2-D separations, as implemented on either mixed silica/cellulose acetate[20] or alumina/cellulose acetate[12,21-23] plates, have been reported. The 2-D TLC of closely related compounds, e.g., aza heterocycles, also has been reported.[24,25]

We had difficulties developing 2-D TLCs using traditional sorbents (e.g., acetylated alumina) and consequently developed our own procedures. These difficulties ensued because of the desire to use overpressure layer chromatography (OPLC). In OPLC, mobile phase is forced to flow through the adsorbent bed under pressure;[26,27] this forced flow maintains a separation efficiency over the plate that is not possible with simple TLC. Our attempts to use OPLC with traditional sorbents and mobile phases either caused the sorbent to loosen from the plate or dissolved the latex used to seal the plate.

Development on Alumina Plates. A series of 2-D TLCs of the PNA mixture was developed on 20 × 20 cm^2 Aluminiumoxid 60 F_{254} neutral (Typ E) alumina plates (Merck, Darmstadt, Germany) by OPLC. The plate edges were twice sealed with latex and dried at room temperature, and the plate was washed with a 9:1 mixture of methanol/water in a Chrompres 25 OPLC (Labor Instrumental Works, Budapest, Hungary) for 10–20 min to reduce plate impurities. The plate then was dried at 90 °C for 5–10 min, cooled, and spotted with 1 μL of mixture via a Nanomat III applicator (Camag Scientific, Wilmington, NC). The solvent was evaporated at room temperature, the plate was pressurized to 20

(7) Brocco, D.; Cantuti, V.; Cartoni, G. P. *J. Chromatogr.* **1970**, *49*, 66.
(8) Raha, C. R. *J. Chromatogr.* **1983**, *264*, 453.
(9) Ho, S. S. J.; Butler, H. T.; Poole, C. F. *J. Chromatogr.* **1983**, *281*, 330.
(10) Menichini, E.; di Domenico, A.; Bonanni, L.; Corradetti, E.; Mazzanti, L.; Zucchetti, G. *J. Chromatogr.* **1991**, *555*, 211.
(11) Sawicki, E.; Stanley, T. W.; Elbert, W. C.; Pfaff, P. *Anal. Chem.* **1964**, *36*, 497.
(12) Pierce, R. C.; Katz, M. *Anal. Chem.* **1975**, *47*, 1743.
(13) Katz, M.; Sakuma, T.; Ho, A. *Environ. Sci. Technol.* **1978**, *12*, 909.
(14) Hurtubise, R. J.; Phillip, J. D.; Skar, G. T. *Anal. Chim. Acta* **1978**, *101*, 333.
(15) Colmsjö, A.; Stenberg, U. *J. Chromatogr.* **1979**, *169*, 205.
(16) Klimisch, H.-J. *J. Chromatogr.* **1973**, *83*, 11.
(17) Butler, H. T.; Coddens, M. E.; Poole, C. F. *J. Chromatogr.* **1984**, *290*, 113.
(18) Poole, C. F; Butler, H. T.; Coddens, M. E.; Khatib, S.; Vandervennet, R. *J. Chromatogr.* **1984**, *302*, 149.
(19) Butler, H. T.; Coddens, M. E.; Khatib, S.; Poole, C. F. *J. Chromatogr. Sci.* **1985**, *23*, 200.
(20) Strömberg, L. E.; Widmark, G. *J. Chromatogr.* **1970**, *49*, 334.
(21) Matsushita, H.; Suzuki, Y. *Bull. Chem. Soc. Jpn.* **1969**, *42*, 460.
(22) Strömberg, L. E.; Widmark, G. *J. Chromatogr.* **1970**, *47*, 27.
(23) McGuirk, M.; Mainwaring, S. J. *J. Chromatogr.* **1977**, *135*, 241.
(24) Sawicki, E.; Stanley, T. W.; Elbert, W. C.; Morgan, M. *Talanta* **1965**, *2*, 605.
(25) Sawicki, E.; Johnson, H. *J. Chromatogr.* **1966**, *23*, 142.
(26) Tyihák, E.; Mincsovics, E.; Kalász, H. *J. Chromatogr.* **1979**, *174*, 75.
(27) Tyihák, E.; Mincsovics, E.; Kalász, H.; Nagy, J. *J. Chromatogr.* **1981**, *211*, 45.

bar, and the first dimension then was developed by OPLC with 9:1 methanol/water at linear velocities ranging from 2.0 to 7.0 cm/min. The plate then was dried at 70 °C for 2–3 min, rotated 90°, developed in the second dimension by OPLC at 20 bar with 1:1 CCl_4/hexane at linear velocities ranging from 2.0 to 7.0 cm/min, and again dried at 70 °C for 2–3 min. Typical development times per dimension were 5–10 min. Excessive drying times significantly reduced the fluorescent signal (see below).

Statistical theories tell nothing about the overlap of undetected components. To verify that all 30 components were detectable, six mixtures containing only five components each were prepared, with component concentrations as in the 30-component mixture. Each component was used only once. Because of their simplicity, these mixtures could be resolved totally, except in one case. For this case, the five components were chromatographed individually. Because all components could be detected, we inferred that they also could be detected in the 30-component mixture.

Development on C18/Silica Plates. The second PNA standard was developed on 20 × 20 cm^2 Multi-K dual phase silica plates (Whatman, Inc., Clifton, NJ) having a 3 cm strip bonded to a C18 phase. The plate first was washed by ascending development with a 9:1 methanol/water mixture and dried at 75 °C for 10 min. The C18 strip then was spotted with 2 μL of mixture, the solvent was evaporated at room temperature, and the plate was multiply developed by ascending development, first with 9:1 methanol/water and then with cyclohexane. Both mobile-phase fronts were developed to 15 cm, and the plate was dried at 70–75 °C for 5–10 min to remove both mobile phases. Ascending development was used, instead of OPLC, because mobile phase simply flowed down the "channel" formed at the phase boundary when OPLC was attempted. The second dimension could be developed by OPLC, however, because the channel was perpendicular to the flow direction. The plate was sealed with three coats of latex and then multiply developed three times with hexane by OPLC at 20 bar to move PNAs over the silica portion of the plate. The hexane fronts were developed to 16 cm at linear velocities ranging from 3.2 to 4.0 cm/min. The plate was dried at 70–75 °C for 5–10 min and then immediately scanned. At all times but during detection, the plate was shielded from UV light, which has caused decomposition of PNAs on silica plates[28,29] (the problem is not as severe with modern adsorbents[9]).

Detection and Signal Processing. The components were detected by fluorescence with a CS-9000U dual wavelength flying spot scanner (Shimadzu, Columbia, MD) operating under CS-TURBO software (Shimadzu) running on a 486 DX2/66 microcomputer. The plate was scanned at 310 or 313 nm, and all emission wavelengths greater than 320 nm were passed through a cutoff filter to a photomultiplier. Parallel lanes were scanned every 1–2 mm apart, and as many as 163 lanes were scanned on one plate. The resolution of data in each lane was 0.04 mm.

Generation of 2-D GC and LC/CE Separations. A portion of a 2-D GC separation of kerosene was kindly provided by John Phillips of Southern Illinois University at Carbondale, and LC/CEs of tryptic digests of thyroglobulin and cytochrome *c* were kindly provided by James W. Jorgenson of the University of North Carolina at Chapel Hill. The means by which these separations were generated have been described,[30–38] and the reader is directed to these references.

Determination of Experimental Maxima Coordinates. For the 2-D TLCs, data files were assembled into a 2-D array, which was splined by a routine written in-house. In this routine, the dimension perpendicular to the scan direction was splined to a resolution of 0.04 mm, such that the spacing between nodes in both dimensions was equal (i.e., 0.04 mm). Splining was performed to determine maxima positions with accuracy, since some narrow spots were scanned only three or so times in each dimension. Data files for the 2-D GC and LC/CE were used as provided.

A program was written to search these 2-D arrays and isolate all intensities, whose eight adjacent nearest neighbors had less intensity. Such intensities were local maxima.[4]

Some of these maxima were attributable to noise and low concentrations of impurities, however, and additional means were required to differentiate them from analyte maxima. For the 2-D TLCs and the 2-D GC, the arrays were plotted as interpolated images with the software package, Spyglass Transform (Spyglass, Inc., Savoy, IL). By comparing low-resolution maxima coordinates in these images to high-resolution maxima coordinates determined by the program, spurious coordinates could be rejected and legitimate ones verified. Both signal intensity and spot shape (i.e., narrow, diffuse, etc.) were used to identify genuine maxima. Because of the large numbers of maxima in the LC/CEs, this procedure was awkward, and arbitrary intensity thresholds were selected, below which maxima were ignored.

Interpretation of Maxima Coordinates by Theory. The p'-plot coordinates for each separation were determined in the original space A of the separation, instead of the reduced space governed by ξ and η, to avoid distortion of intervals between maxima. No problem exists, however, in computing $f_a(\xi,\eta)$ or the integral in eq 2 in reduced space, once the p'-plot coordinates are so determined. Any p'-plot coordinates for which ordinate p' exceeded $0.95p_m$ or was less than either 10 (for the TLCs) or 20 (for the other separations) were discarded, because p' values nearly equal to p_m contain little useful information, and p' values less than 10–20 are not statistically robust.

The frequency, eq 3, was evaluated over a 2-D nodal network, as detailed in ref 1. Parameter $\bar{m}$ was determined by minimizing SS, eq 2, with a nested golden search[39] that also determined σ in eq 3. The integral in eq 2 was evaluated with Simpson's rule. The optimal $\bar{m}$ was identified with the number of components.

Generation of Computer Simulations. Because the numbers of components in two of the three separation types were not known, simulations were constructed to evaluate the credibility of $\bar{m}$ values prediced by theory. To mimic experimental separations by simulation, individual experimental maxima were fit by

(28) Inscoe, M. N. *Anal. Chem.* **1964**, *36*, 2505.
(29) Siefert, B. *J. Chromatogr.* **1977**, *131*, 417.
(30) Liu, Z.; Phillips, J. B. *J. Chromatogr. Sci.* **1991**, *29*, 227.
(31) Venkatramani, C. J.; Phillips, J. B. *J. Microcolumn Sep.* **1993**, *5*, 511.
(32) Liu, Z.; Sirimanne, S. R.; Patterson, D. G., Jr.; Needham, L. L.; Phillips, J. B. *Anal. Chem.* **1994**, *66*, 3086.
(33) Bushey, M. M.; Jorgenson, J. W. *Anal. Chem.* **1990**, *62*, 161.
(34) Bushey, M. M.; Jorgenson, J. W. *Anal. Chem.* **1990**, *62*, 978.
(35) Bushey, M. M.; Jorgenson, J. W. *J. Microcolumn Sep.* **1990**, *2*, 293.
(36) Lemmo, A. V.; Jorgenson, J. W. *J. Chromatogr.* **1993**, *633*, 213.
(37) Lemmo, A. V.; Jorgenson, J. W. *Anal. Chem.* **1993**, *65*, 1576.
(38) Larmann, J. P., Jr.; Lemmo, A. V.; Moore, A. W., Jr.; Jorgenson, J. W. *Electrophoresis* **1993**, *14*, 439.
(39) Press, W. H.; Teukolsky, S. A.; Vetterling, W. T.; Flannery, B. P. *Numerical Recipes in FORTRAN*, 2nd ed.; Cambridge University Press: Cambridge, U.K., 1992.

least-squares to concentration profile c,

$$c = a_1\zeta + a_2\eta + a_3 + a_4\, e^{-(a_5-\zeta)^2/2a_6^2}\, e^{-(a_7-\eta)^2/2a_8^2} \quad (4)$$

which is a bi-Gaussian on a sloping plane. Coefficients a_6 and a_8 are the standard deviations of the bi-Gaussian. Values of a_6 and a_8 were so determined from several maxima, and the means and standard deviations of these values were then used with the Box–Muller transformation[39] to generate independently simulated values of a_6 and a_8 having Gaussian distributions. In some cases, eq 4 was modified to represent the sum of two bi-Gaussians on a sloping plane; this function was used when two bi-Gaussian spots were partially fused.

SCS coordinates for the simulations were generated by selecting an arbitrary path of integration over the 2-D array representing the experimentally determined frequency, $f_a(\zeta,\eta)$. The coordinates, (ζ_o,η_o) for SCSs in a simulation were determined by solving numerically the equation

$$\int_0^{\zeta_o}\int_0^{\eta_o} f_a(\zeta,\eta)\, d\zeta\, d\eta = F_a(\zeta_o,\eta_o) = z \quad (5)$$

for a sequence of uniformly distributed random numbers z. In eq 5, $F_a(\zeta_o,\eta_o)$ is the cumulative distribution evaluated at the coordinate pair, (ζ_o,η_o). The path of integration was immaterial, as was the location of the origin, as long as the same path was used for each random-number call in a simulation. This protocol is applicable to both independent and dependent coordinate distributions (see ref 1).

The SCSs were represented by bi-Gaussians having exponentially distributed amplitudes. Standard deviations parallel to the ζ and η coordinates were computed independently.

Miscellaneous. Computations were carried out in FORTRAN on a 486 DX2/66 or a Pentium 90 microcomputer. Least-squares fittings of eqs 4 and 6 were carried out with Mathematica (Wolfram Research, Champaign, IL) and KaleidaGraph (Synergy Software, Reading, PA), respectively. Two-dimensional, three-dimensional, and contour graphs were generated with KaleidaGraph, Mathematica, and Spyglass Transform, respectively.

RESULTS AND DISCUSSION

Application to 2-D TLC. Figure 1 is an interpolated image of a 2-D TLC of 30 PNAs, as developed on an alumina plate. The separation is complete: all 30 components are resolved as singlets. The separation quality varies over the plate; most singlets are very narrow, whereas a few are diffuse. A group of intense singlets forms a V-like structure in the central region of the separation. The maxima coordinates are extremely correlated in this region, although we cannot explain why they have this structure. Clearly, the pattern is far from random.

If the modified theory is valid, then it should predict that the separation is complete and contains 30 components. Table 2 reports the parameters $\bar{m}$, standard deviation σ of frequency $f_a(\zeta,\eta)$, and reduced sum of squares SS_ν (i.e., the sum of squares SS, eq 2, divided by the number of p'-plot coordinates, less two) determined by interpreting three 2-D TLCs of the PNA mixture, as developed on alumina plates at different linear velocities $\langle\nu\rangle$ of mobile phase. All three separations were well resolved and contained either 29 or 30 maxima. Theory tells us the separations are resolved; the $\bar{m}$ estimates are 31.4, 30.7, and 30.7, and are accurate to within 5%. The optimal σ values for constructing the TLCs' frequencies are essentially the same and are small. This is perhaps unsurprising; simulations show that high-resolution separations having much structure commonly result in small σ values.

Table 2. Numbers p_m of Maxima and Statistical Parameters $\bar{m}$, σ, and SS_ν Determined by Interpreting Three 2-D TLCs of the 30-Component PNA Standard, as Developed on Alumina Plates[a]

TLC	$\langle\nu\rangle$, cm/min	p_m	$\bar{m}$	σ	SS_ν
1	3.0	30	31.4	0.0238	2.85
2	6.2	30	30.7	0.0266	0.72
3	7.0	29	30.7	0.0244	1.43

[a] Linear velocities $\langle\nu\rangle$ of mobile phases are reported. Same $\langle\nu\rangle$ in both dimensions.

Figure 2 shows the three p'-plots generated from these TLCs. The symbols represent the coordinates, $(A_o/A,p')$, and the bold curves are the optimal fits of eq 2 to these coordinates. The fits in general are quite good. The dashed curves in the figure, which appear to be straight lines, are fits to the same p'-plot coordinates of the Roach equation,

$$p = 4\bar{m}\frac{A_o}{A}\frac{e^{-4\bar{m}A_o/A}}{1-e^{-4\bar{m}A_o/A}} \quad (6)$$

previously used to model overlap in simulated 2-D separation.[3,4] The $\bar{m}$ estimates computed from fits to this equation are about 20. In other words, the Roach equation predicts that there are 9–10 fewer components than maxima! This outcome occurs simply because components are not distributed randomly over the 2-D space and the Roach equation is not applicable. The accuracy of the $\bar{m}$ values determined by the modified theory and the inaccuracy of the $\bar{m}$ values determined by the Roach equation justify development of the modified theory.

One may argue this statistical interpretation is not a good test, because all components are resolved. Figure 3a is a contour image of a 2-D TLC of the 30-component PNA mixture, as developed on a C18/silica plate. The separation is not as good as on the alumina plate; only 23 maxima were found. Although the efficiency of this separation does not differ greatly from those of other 2-D TLCs,[40] simulations show that one might have difficulties in correctly predicting m from such a separation,[1] and this difficulty was realized. The parameters $\bar{m}$, σ, and SS_ν determined from this separation are 23.6, 0.226, and 0.60, respectively. The $\bar{m}$ prediction is disturbingly low and almost equal to p_m. Figure 3b is a graph of p' vs A_o/A developed from this separation. The solid and dashed curves are the predictions of eq 2 and the Roach equation, eq 6. Unlike before, the Roach equation is not that bad a fit, because the frequency is almost constant.

In this case, one knows that $\bar{m}$ is wrong, because the number of components is known. In general, however, one will not have this knowledge. From this analysis alone, one could conclude that this separation is very good. Can one demonstrate this conclusion is false?

Simulations were useful in demonstrating the conclusion's falsity. The data files of 19 spots in this 2-D TLC were fit to eq 4 to approximate the standard deviations of SCSs, and these

(40) Zakaria, M.; Gonnord, M.-F.; Guiochon, G. *J. Chromatogr.* **1983**, *271*, 127.

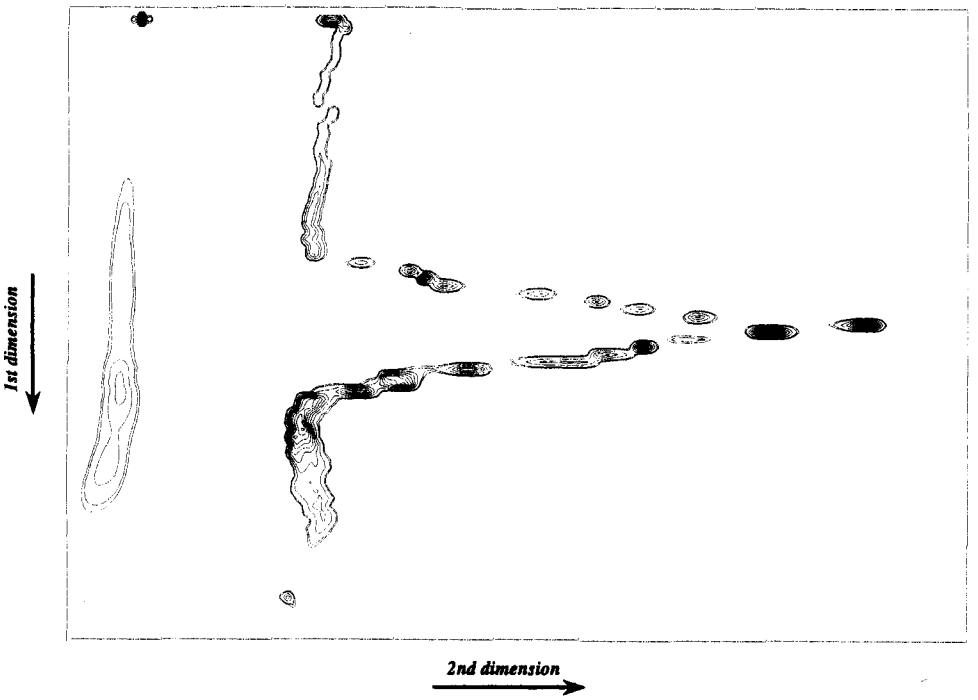

Figure 1. Interpolated image of 2-D TLC (TLC 1) developed on alumina plates from 30-component PNA mixture.

standard deviations were used to construct bi-Gaussians at SCS coordinates generated from the experimental frequency, $f_a(\zeta,\eta)$, and eq 5. The standard deviations of these bi-Gaussians had considerable variability in breadth, and both narrow and broad SCSs were mimicked. Based on the simulations, about 34 components were required on average to produce the number of spots determined by experiment (23). This number differs from $\bar{m}$ by 44% (relative to $\bar{m}$), suggests that a substantial error in $\bar{m}$ exists, and indicates that this separation, as developed, is inadequate for statistical interpretation. Interestingly, the simulatory estimate is correct within 14%, which merits consideration (see below).

The origin of the error in $\bar{m}$ most probably is high saturation, i.e., insufficient spots are resolved to apply theory. It is tempting to attribute some error to the large variability in SCS standard deviations, since $\bar{m}$ values are most accurately predicted for SCSs with circular contours,[1] but the simulation suggests this attribution would not be correct.

A valid criticism of statistical overlap theories is that one does not know the probability density function (pdf) that governs the positions of components. Equation 1, for example, is based on the *assumption* that this distribution is Poisson and nonhomogeneous. The pdf's governing experimental separations, however, may be different. One approach to this difficulty is to propose various overlap models based on a series of pdf's and then choose the one that best fits the experimental data.[41-43] As implemented with eq 5, however, simulated (or "regenerated") separations always will be Poisson (homogeneous or nonhomogeneous), because the distribution $F_a(\zeta,\eta)$ is equated to a uniform random number. This led us to pose the following criterion: if regenerated and experimental separations "look" similar in spot density, spot shape, etc., and if the number of components required to generate p_m maxima is about equal to $\bar{m}$, then the distribution is not distinguishable from Poisson and $\bar{m}$ is correct. We believe this criterion is superior to gauging the accuracy of $\bar{m}$ by estimating saturation $\bar{a}$ from p_m and $\bar{m}$,[5,44,45] because this estimation is biased by erroneously small values of $\bar{m}$ at high saturations.

Application to 2-D GC. Another type of 2-D separation interpretable by statistical theory is comprehensive 2-D gas chromatography (2-D GC), developed by Phillips et al.[30-32] Figure 4a is a small portion of a 2-D GC separation of kerosene containing principally compounds having nine carbon atoms.[46] The maxima coordinates are correlated, with a large density of maxima near the "front" of the separation, no maxima in the "middle", and seven maxima near the "back" (only six are clearly visible). The maxima near the front correspond to branched and cyclic aliphatics; the maxima near the back correspond to methyl-, ethyl-, and propylbenzenes.[46] Because the contours of maxima are nearly circular, one expects the predictions of statistical theory to be reliable. We consequently determined from this separation the coordinates of 85 maxima and subjected them to several statistical interpretations.

Table 3 reports the parameters $\bar{m}$, σ, and SS_r for the optimal fit of eq 2 to p'-plot coordinates determined from this separation, and Figure 5a is the p'-plot from which $\bar{m}$ was estimated. Theory predicts that $\bar{m} = 110.5$ components are in this separation containing 85 maxima. Thus, only about $(85/110) \times 100 = 77\%$ of components are observable as maxima. The dashed curve is the fit of the same p'-plot coordinates to the Roach equation, eq 6. As before, the fit is poor and the $\bar{m}$ estimate (63.9) is unrealistically small. This outcome is not surprising; the maxima distribution is far from random.

(41) Felinger, A.; Pasti, L.; Dondi, F. *Anal. Chem.* **1992**, *64*, 2164.
(42) Dondi, F.; Betti, A.; Pasti, L.; Pietrogrande, M. C.; Felinger, A. *Anal. Chem.* **1993**, *65*, 2209.
(43) Pietrogrande, M. C.; Dondi, F.; Felinger, A.; Davis, J. M. *Chemometr. Intell. Lab. Syst.* **1995**, *28*, 239.
(44) Coppi, S.; Betti, A.; Dondi, F. *Anal. Chim. Acta* **1988**, *212*, 165.
(45) Oros, F. J.; Davis, J. M. *J. Chromatogr.* **1991**, *550*, 135.
(46) Phillips, J. B. Personal communication, 1995.

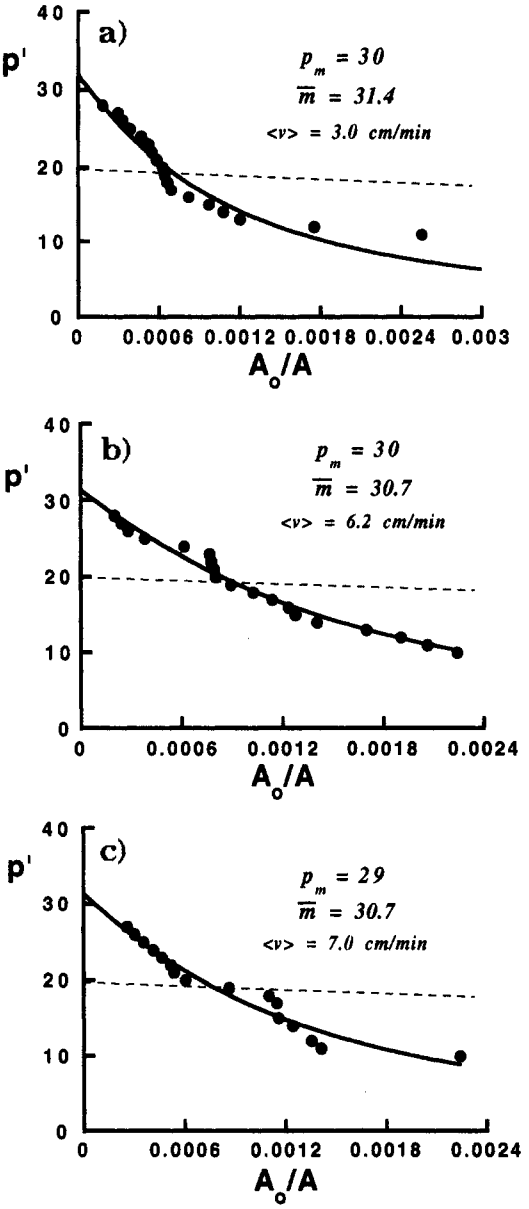

Figure 2. Graphs of p' vs A_o/A developed from TLCs (a–c). Solid and dashed curves are fits of eqs 2 and 6, respectively, to p'-plot coordinates.

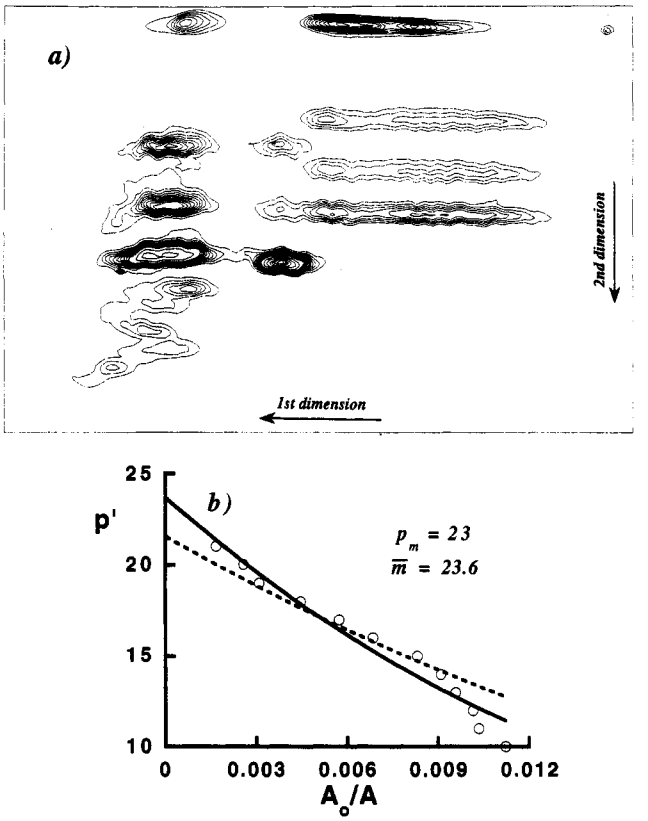

Figure 3. (a) Contour map of 2-D TLC developed on C18/silica plate from 30-component PNA mixture. (b) Graph of p' vs A_o/A developed from coordinates of maxima. Solid and dashed curves are fits of eqs 2 and 6, respectively, to p'-plot coordinates.

Table 3. Statistical Parameters $\bar{m}$, σ, and SS_v Determined from 2-D GC[a]

type of interpretation	$\bar{m}$ (or $\bar{m}^i$)	σ	SS_v
entire separation	110.5	0.086	8.12
entire separation, 1st coordinate shift[b]	109.4	0.115	7.66
entire separation, 2nd coordinate shift[b]	109.2	0.220	7.92
front region, eq 2	98.1	8.111	6.44
front region, Roach equation	98.1		6.45
front region,[c] eq 7	98.3		
back region, by difference,[c] eq 7	6.9		
entire separation, corrected for anomalous middle region	~105		

[a] The numbers of maxima in the entire separation, its front region, and its back region are 85, 78, and 7, respectively. [b] See Figure 5b.[c] $\bar{m}$ = 110.5.

Several additional interpretations are useful. For example, if our analytical procedures are valid, then one should be able to shift forward the coordinates of the seven maxima near the separation's back, without changing $\bar{m}$. Parameter $\bar{m}$ should not change, because this action simply is equivalent to removing part of the unoccupied space between the separation's front and back regions, without changing the actual extent of separation. Figure 5b illustrates two successive arbitrary shifts of these coordinates; the resultant two coordinate sets then were reinterpreted statistically. Table 3 reports the values of $\bar{m}$, σ, and SS_v computed by these interpretations. As expected, $\bar{m}$ does not change significantly (the two new estimates are 109.4 and 109.2), and SS_v does not change much. The value of σ does increase as the seven coordinates are moved forward, however, because the new maxima coordinates are more random-like (i.e., less empty space exists), and random distributions correspond to large σ values.[1] This invariance of $\bar{m}$ shows that the procedures are robust.

An instructive interpretation can be made of the front region of the separation which contains 78 maxima. If p'-plot coordinates are generated only from this region, then eq 2 predicts that $\bar{m}$ = 98.1 components are present. The σ and SS_v values for the fit are reported in Table 3. The σ is very large, which indicates that the SCS distribution is almost random. That the distribution is random can be shown by fitting eq 6, the Roach equation, to the p'-plot coordinates, from which the identical estimate, $\bar{m}$ = 98.1, is calculated. Figure 5c is a graph of p' vs A_o/A containing p'-plot coordinates and fits based on eqs 2 and 6. These two fits superimpose and cannot be distinguished.

Another interpretation of the front region illustrates the capabilities of theory. It actually is not necessary to analyze this region separately to estimate $\bar{m}$. By arguments identical to those proposed for 1-D separations,[47] the mean number $\bar{m}^i$ of compo-

(47) Davis, J. M. *J. Microcolumn Sep.* **1995**, *7*, 3.

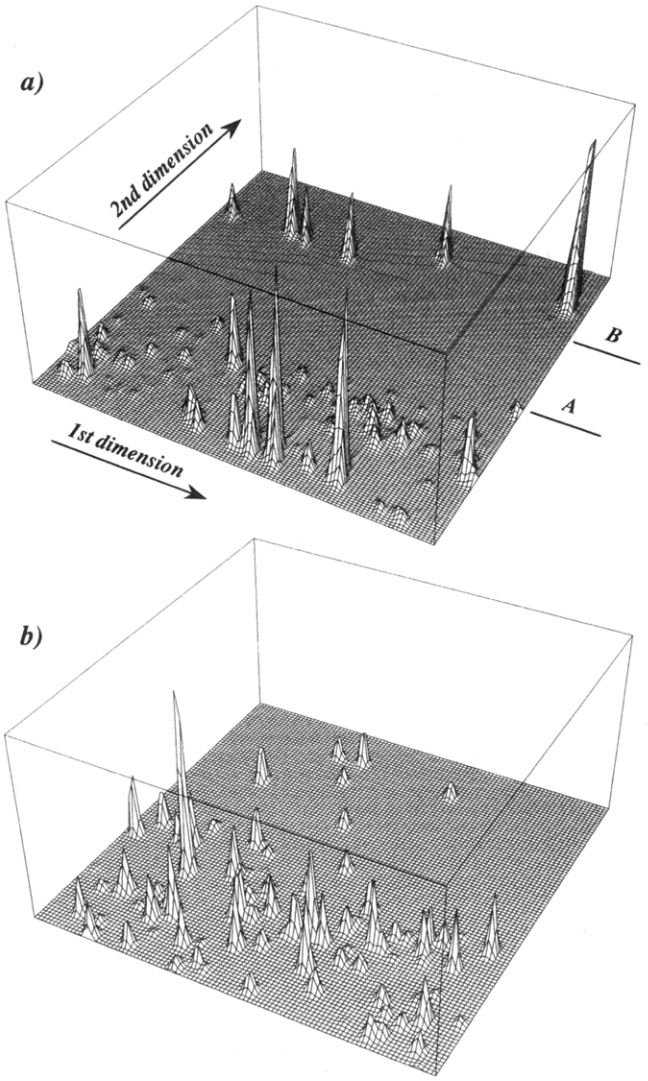

Figure 4. (a) Experimental 2-D GC. (b) Computer regeneration of 2-D GC.

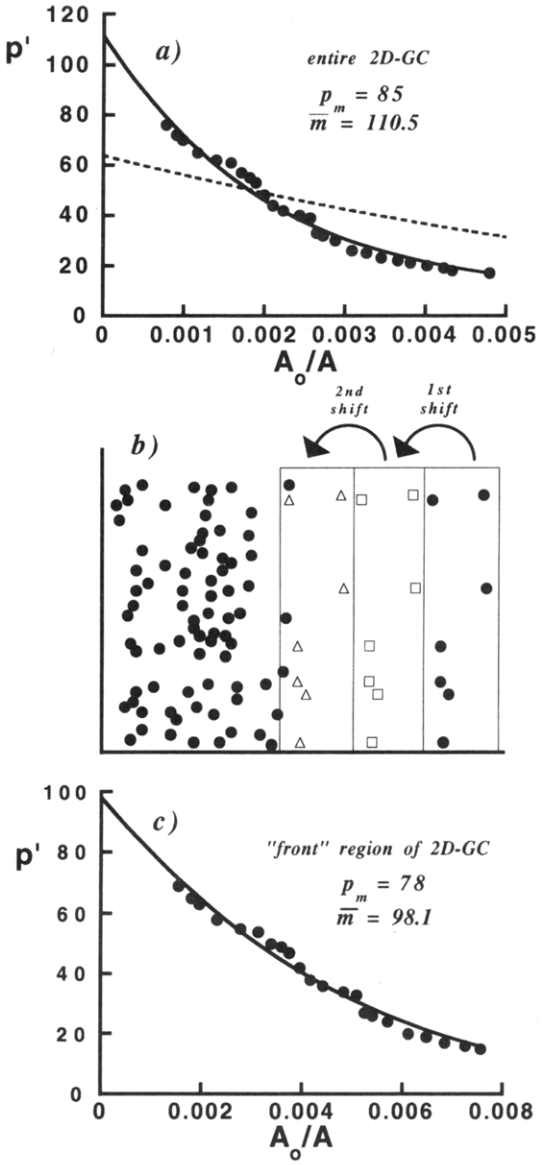

Figure 5. (a) Graph of p' vs A_o/A developed from the 2-D GC. Solid and dashed curves are fits of eqs 2 and 6, respectively, to p'-plot coordinates. (b) Illustration of two successive "forward" shifts of maxima coordinates from the "back" region of the 2-D GC. Filled circles represent original coordinates; open squares (triangles) represent the "back" coordinates after the first (second) shift. (c) Graph of p' vs A_o/A constructed from the "front" region of the 2-D GC. Superimposing curves are least-squares fit of eqs 2 and 6 to p'-plot coordinates.

nents in any *local* rectangular subset of a 2-D separation, defined by the coordinates $\zeta_1 \le \zeta \le \zeta_2$ and $\eta_1 \le \eta \le \eta_2$, can be shown to be

$$\bar{m}^i = \bar{m} \int_{\zeta_1}^{\zeta_2} \int_{\eta_1}^{\eta_2} f_a(\zeta,\eta)\, d\zeta\, d\eta \qquad (7)$$

where $\bar{m}$ and $f_a(\zeta,\eta)$ are the global number of components in and the global frequency of the entire separation, respectively. By integrating the global frequency over the rectangle defined by the two sides of the separation, its front, and line A in Figure 4a, and then multiplying the result by $\bar{m} = 110.5$ (the $\bar{m}$ for the entire separation), we determined the value, $\bar{m}^i = 98.3$. This is almost identical to 98.1, the value calculated from two independent analyses of this region.

It also is instructive to integrate eq 7 over the rectangle defined by the two sides and lines A and B in Figure 4a, where no maxima are found. The value, $\bar{m}^i = 5.3$, was so determined. Thus, theory predicts that about five components are expected in a region containing no maxima. This outcome indicates a flaw in our procedures; the frequency $f_a(\zeta,\eta)$ cannot approach zero abruptly, except for very small σ. Fortunately, the error is small.

This error actually provides an additional insight. The number of components in the back region of the separation must equal the difference between the global number, $\bar{m} = 110.5$, and the sum of the estimates from the front and middle regions, 98.3 + 5.3 = 103.6. This difference, 6.9, is almost identical to the number (7) of maxima in the back region. Thus, theory suggests that the maxima in the back region are fully resolved. This conclusion is not surprising; the maxima are extremely well separated.

Taking everything into account, one might conclude that a more correct estimate of $\bar{m}$ for this separation is about 105. Here, the anomalous number 5.3, calculated from the middle region of the separation, has been subtracted out. For the purpose of evaluating theory, however, we will retain the original prediction, $\bar{m} = 110.5$.

To add credibility to these interpretations, simulations were generated to mimic the 2-D GC. Standard deviations of SCSs were approximated from eight maxima (seven in the front and one in

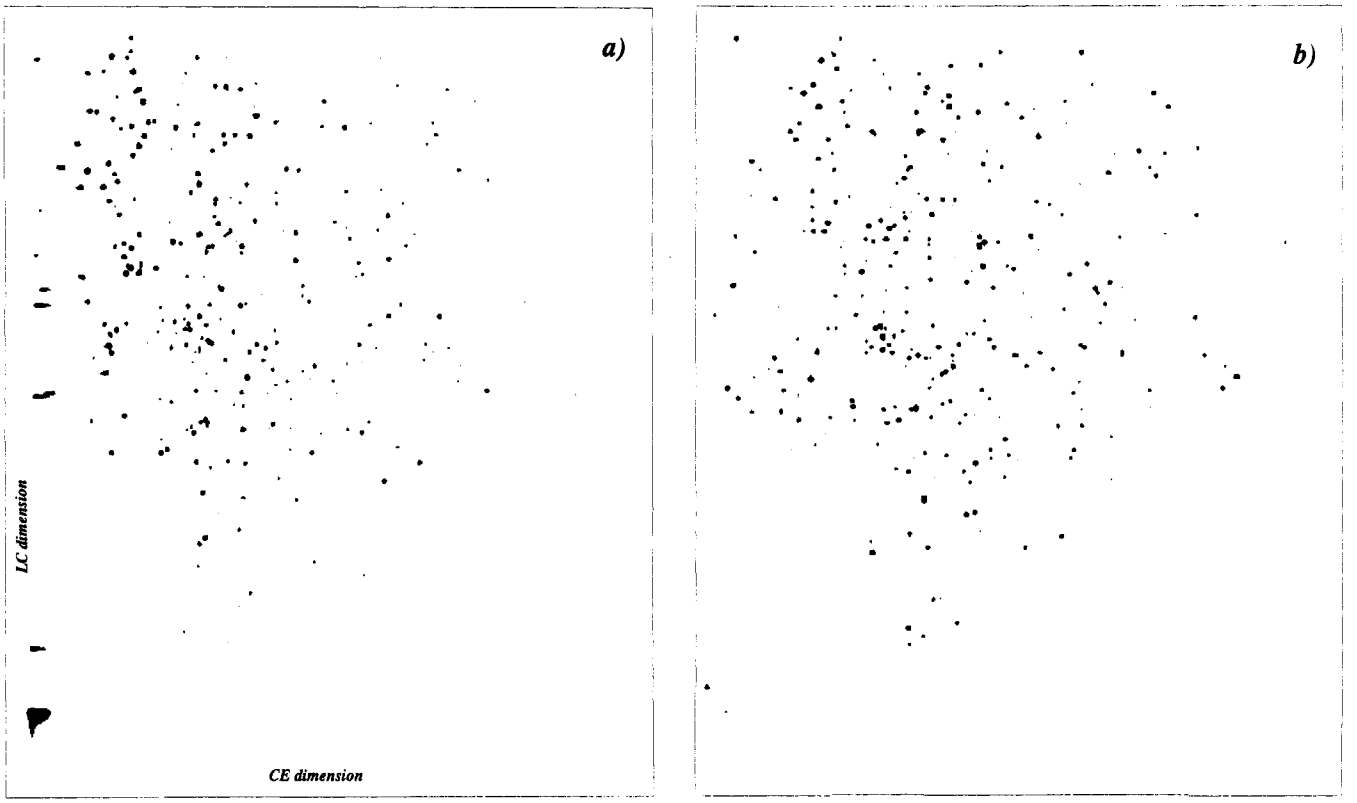

Figure 6. (a) Interpolated image of LC/CE separation of tryptic digest of thyroglobulin. (b) Computer regeneration of (a).

the back region), and SCS coordinates were generated from frequency $f_a(\zeta,\eta)$ and eq 5, until 85 maxima (the number determined by experiment) were found. Figure 4b is an example of one such simulation, with the largest peak scaled to the same relative intensity as in Figure 4a. One observes that it qualitatively resembles the 2-D GC (Figure 4a) in both the distribution and breadth of maxima. As in related 1-D simulations,[6,48] no attempt was made to match retention times and amplitudes to experiment; only the overall pattern has significance. A few maxima are found in the middle of the simulation, because of the error in the frequency's middle discussed above.

On average, 101 components were required to generate 85 maxima in these simulations. This number is less than that predicted by theory (110.5) but by <10%. If one postulates that 105 or so components are present (see above) instead of 110, the agreement is even better. In related simulations of the front region, 94 components were required to produce the number of maxima (78) in this region; this number is fairly close to the estimate, $\bar{m} = 98.1$, determined by theory. The good agreement between simulation and theory gives one confidence in these $\bar{m}$ values. We ultimately conclude that the front region contains about 20 components hidden by overlap.

Application to LC/CE. A final application of 2-D statistical theory is to comprehensive LC/CE, developed by Jorgenson et al.[33-38] Figure 6a is an interpolated image of a tryptic digest of thyroglobulin, in which spots represent maxima. As in the other 2-D separations interpreted here, the maxima positions are highly correlated; more maxima lie in the upper left quarter of the separation than elsewhere. Furthermore, as in the 2-D GC, the contours of maxima are nearly circular, and one expects the predictions of statistical theory to be reliable.

Table 4. Numbers p_m of Maxima and Statistical Parameters $\bar{m}$, σ, and SS_ν Determined from LC/CEs of Tryptic Digests of Thyroglobulin and Cytochrome *c*

digest	threshold	p_m	$\bar{m}$	σ	SS_ν
thyroglobulin	0.03	327	339.7	0.0164	21.6
	0.02	660	672.7	0.0127	66.2
cytochrome *c*	0.03	237	223.0	0.0207	13.4
	0.02	422	407.3	0.0135	46.0

The intensities of maxima in this separation vary from near 0 to 1, the value at which intensity was clipped. Based on an eight-nearest-neighbor search, thousands of maxima have amplitudes greater than 0.01. Because it was impractical to interpret thousands of maxima, we selected two arbitrary thresholds and interpreted only maxima having intensities exceeding them. For the threshold 0.03, 327 maxima were found, and for the threshold 0.02, 660 maxima were found. We subjected the coordinates of these maxima to two separate statistical interpretations.

As shown elsewhere,[49] predictions of $\bar{m}$ depend on detector and signal-processing thresholds. Stated simply, if maxima are not detected, then statistical theory tells one nothing about them. Table 4 reports values of $\bar{m}$, SS_ν, and σ determined from the thyroglobulin separation. The $\bar{m}$ estimates for the 0.03 and 0.02 thresholds are 339.7 and 672.7, respectively, and are only slightly greater (by 12.7) than p_m for both thresholds. These predictions indicate that this separation is very good and has only a small amount of overlap. The accuracy of $\bar{m}$ for the 0.03 threshold also is supported by regenerations. Here, the data files of 24 maxima were fit to eq 4 to estimate standard deviations, and then SCS coordinates were generated from $f_a(\zeta,\eta)$ and eq 5. On average, 341 components were required to produce 327 maxima, the

(48) Davis, J. M.; Giddings, J. C. *Anal. Chem.* **1985**, *57*, 2168.

(49) Bowlin, D.; Hott, C.; Davis, J. M. *J. Chromatogr.* **1994**, *677*, 307.

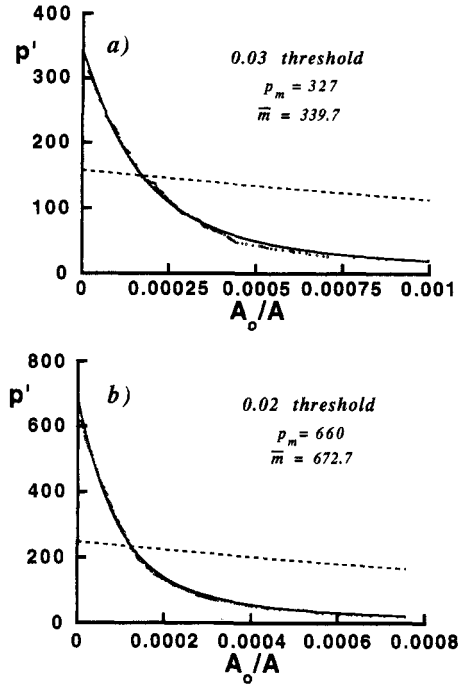

Figure 7. Graphs of p' vs A_o/A constructed from LC/CE of thyroglobulin, as interpreted at thresholds of (a) 0.03 and (b) 0.02. Solid and dashed curves are fits of eqs 2 and 6, respectively, to p'-plot coordinates.

number obtained by experiment. The agreement between this number and the $\bar{m}$ value, 339.7, is very good.

Figure 6b is an interpolated image of one of these simulations. It more closely resembles the LC/CE in Figure 6a than the simulatory 2-D GC resembled its experimental counterpart, because σ is smaller and the locations of the SCSs are more precisely determined. A brief explanation of the scaling of intensities in these figures is appropriate. In Figure 6a, all intensities less than the threshold, 0.03, were washed out to avoid showing both interpreted and ignored maxima. This action unfortunately clipped the bases of maxima having intensities greater than 0.03. This feature also was preserved in Figure 6b.

Figure 7 shows graphs of p' vs A_o/A generated from the thyroglobulin separation for the two different thresholds. The fits of eq 2 to the p'-plot coordinates, represented by the bold curves, are good. The dashed curves represent fits of the Roach equation, eq 6, to the same coordinates; as before, these fits are poor, because the elution times of SCSs are correlated.

Statistical interpretations also were made for a tryptic digest of cytochrome *c*, the 2-D separation of which is not shown. For the 0.03 and 0.02 thresholds, 237 and 422 maxima were determined, respectively. Table 4 reports values of $\bar{m}$, σ, and SS_v determined by statistical interpretation. In particular, the values of $\bar{m}$ are 233.0 and 407.3, respectively, or about 14 less than the numbers of maxima. Thus, the $\bar{m}$ values actually are slightly smaller than the p_m values, which at first glance seems physically unrealistic. This discrepancy probably results from statistical scatter due to applying theory to a very well resolved separation, where $p_m \approx \bar{m}$. Supportive evidence for this prediction comes from regenerations. The data files of 13 spots were fit to eq 4 to estimate standard deviations, from which the separation was mimicked in accordance with frequency $f_a(\zeta,\eta)$ and eq 5. On average, about 243 components were required to produce 237 maxima, the number determined by experiment at the threshold, 0.003. This number is only a little greater than $\bar{m}$ and affirms our conclusion that this is a very good separation.

Finally, we also interpreted the thyroglobulin separation by partitioning it into six regions and fitting the Roach equation, eq 6, to p'-plot coordinates computed from five of them. In five of the regions, the spot density appeared constant, as judged by simple inspection. The number of components in the entire separation then was estimated by summing the $\bar{m}$ values computed from the five regions and adding to them the number of maxima in the sixth region, which had very low but variable density. The motive for this action was to compare component numbers estimated by successive applications of the "simple" Roach equation (a 1-D analog is reported in ref 50) to those estimated by the more complex integral form of Roach's equation, eq 1. The $\bar{m}$ estimates so determined for the 0.03 and 0.02 thresholds were 310.6 and 644.0, respectively. These estimates are slightly less than the number of maxima (327 and 660). The error is not large, however, and one still could conclude that the separation is good. The error principally is due to interpreting the region in the upper left-hand corner of Figure 6a. Although only a small effort and a little computational time are required by this procedure, the estimate of $\bar{m}$ is not as reliable as that determined by more detailed means. Furthermore, one can encounter domains (like the sixth region) in which spot density simply is not uniform.

CONCLUSION

The applications presented here show that the modified statistical theory can predict the numbers of components in several different types of experimental 2-D separations. The theory evidently handles quite well at least part of the correlations between the two dimensions of separation. This issue had to be addressed by any practical 2-D theory of overlap, and it is pleasing to note that the relatively simple theory used here often is adequate to the task. It also is clear that the original Roach equation is not very useful in modeling overlap in experimental 2-D separations, except perhaps to verify that specific regions contain random SCS distributions. This conclusion is not a criticism of either Roach or his pioneering work, which was developed to model not the overlap of SCSs in 2-D separations but rather the overlap of coal particulates on precipitators.[51] Rather, it merely is a recognition that this pioneering work is too simple to describe real 2-D separations and that more complex theories are required.

In spite of these constructive developments, work remains to be done. The present approach considers correlations only in an approximate way, and structure in the separation is ignored. The maxima in the back of the 2-D GC, for example, lie mostly on a line, and this line evolves over a considerable span of the first separatory dimension.[46] A more complete theory should take this detailed structure into account.

In a recent paper,[52] Giddings discussed the effects of sample dimensionality on multidimensional separations. In effect, if the sample dimensionality is less than or equal to the number of separatory dimensions, order and structure in the separation appear and the disorder on which statistical-overlap theories are based disappears. Indeed, such behavior can be observed in some

(50) Davis, J. M.; Giddings, J. C. *Anal. Chem.* **1985**, *57*, 2178.
(51) Roach, S. A. *The Theory of Random Clumping*; Methuen: London, 1968.
(52) Giddings, J. C. *J. Chromatogr.* **1995**, *703*, 3.

recent 2-D GCs of gasoline by Phillips.[46] In other cases, such as those considered here, statistical effects still dominate. Regrettably, no simple approach now exists to describe quantitatively the behavior of all possible multidimensional separations.

One reason the 2-D TLCs interpreted here constitute a relatively poor test of theory is the limited capacity of this method, which forced us to work with a relatively small number of components. Regrettably, the statistical theory is limited in estimating good parameters under these conditions. Although further experimental study of 2-D statistical theory is warranted, 2-D TLC does not seem ideal for the study. A more attractive method would entail deposition of nebulized effluent from a microbore column along a TLC plate's first dimension.[53] This dimension would be modestly well resolved, and then the second dimension could be resolved by TLC or OPLC. The resultant 2-D separation could contain greater numbers of components than considered here.

Finally, the use of simulations to gauge the accuracy of theoretically determined $\bar{m}$ values, or to determine the number of components directly, by mimicking component positions and breadths merits further exploration. It is of considerable interest that the regenerations developed from the C18/silica TLC separation provided as good an estimate of m as they did, in light of the separation's high saturation. The method clearly has shortcomings as presently implemented, such as estimating SCS standard deviations from observed spots, but further study appears to be warranted.

(53) van de Nesse, R. J.; Hoogland, G. J. M.; de Moel, J. J. M.; Gooijer, C.; Brinkman, U. A. Th.; Velthorst, N. H. *J. Chromatogr.* **1991**, *552*, 613.

ACKNOWLEDGMENT

The authors gratefully acknowledge Milton Lee (Brigham Young University) for the PNA standards, John Phillips (Southern Illinois University at Carbondale) for the 2-D GC, and James W. Jorgenson (University of North Carolina at Chapel Hill) for the LC/CEs. The assistance of Marie Castro (Southern Illinois University at Carbondale) and Anthony Lemmo (University of North Carolina at Chapel Hill) in reading 2-D GC and LC/CE data files is acknowledged. This work was supported by the National Science Foundation (CHE-9215908).

Received for review January 26, 1995. Accepted June 19, 1995.[⊗]

AC950097L

[⊗] Abstract published in *Advance ACS Abstracts,* August 1, 1995.