Applications of Gel Permeation Chromatography. III. Formulation

MORTON SCHRAGER, *Xerox Corporation, Rochester, New York 14644*

Synopsis

This paper describes a general theory for the optimization of multicomponent blending to achieve a desired chromatographic distribution. A procedure for compound blending is also discussed, where optimum replication of both the chromatographic spectra and a second independent parameter can be achieved. Lastly, this theory is applied in a hypothetical example.

INTRODUCTION

In the two previous publications in this series, methods were described for the separation of a gel permeation chromatogram into its component peaks[1] and graphically determining molecular weight averages.[2] The corollary procedure to separating and analyzing a chromatogram into its components involves formulating a mixture of a number of materials to obtain the desired chromatographic spectrum. Smith[3] has previously presented a method of formulation which has been shown to be effective in the formulation of a broad molecular weight distribution blend. As discussed by Smith, the method is somewhat more general than Tung's[4] and more sensitive than the method of Hess and Kratz.[5] However, this method is only a particular solution of a more general theory of formulation as described below.

BACKGROUND

It is well known that the weight concentration of a solute in a dilute solution is in many cases found to be[6]

$$C_w = \frac{n_s - n_1}{n - n_1} \tag{1}$$

where C_w is the weight concentration in g/cm³ of solution; n_s, n_1, and n are the refractive indices of the solution, the solvent, and the solute, respectively.

Inherent in eq. (1) is the assumption that the densities of the solution, solute, and solvent are similar and that the Gladstone-Dale equation holds for the system.[6]

The difference between the index of refraction of the solution and the solvent $(n_s - n_1)$ is proportional to the height of the chromatogram, $R(V)$. This assumes a linear relationship between the input signal and the output of the recorder.

From this, the area under a chromatogram peak may be expressed as

$$A = \int_0^\infty R(V)dV = kw_i(n - n_1). \tag{2}$$

If a number of components are present in the eluent, the weight fraction of the kth component, C_k, associated with the peak elution volume V_k can be expressed as

$$C_k = \frac{w_k}{\sum_i w_i} = \frac{A_k}{(n_k - n_1)}\left[\sum_i \frac{A_i}{(n_i - n_1)}\right]^{-1} \tag{3}$$

where n_k is the refractive index of the kth component and A_k is the area associated with the kth peak. The weight fraction of the kth component, C_k, is then a function only of the area under the elution peaks and the refractive index differences $(n_i - n_1)$. It is independent of both the peak shape and the peak elution volume V_k.

For the case where all components exhibit the same refractive index, (as is approximately the case for polymer molecules that are homologues of the same monomer), eq. (3) reduces to

$$C_k = A_k/\sum_i A_i. \tag{4}$$

THEORY OF CHROMATOGRAPHIC FORMULATION

Generally, in a formulation procedure, it is desirable to replicate a material whose chromatogram contains n peaks (i.e., n resolvable components). To best replicate this material, a number of materials are available which contain some or all of the n peaks, however, in concentrations which differ considerably from the desired distribution. Assuming there are m materials suitable for use in the replication procedure, let X_{jk} be the weight fraction of the kth component in material j. Then the given spectra can be replicated when

$$C_k = \sum_{j=1}^m a_j X_{jk} \tag{5}$$

where the coefficients a_j represent the weighted amount of each material necessary to reproduce the desired spectrum, i.e., the weight fraction of each material, w_j, is

$$w_j = \frac{a_j}{\sum a_j}. \tag{6}$$

If each of the m materials contains only one component of the desired formulation, eqs. (4), (5), and (6) serve to define the problem. Smith[3] has

presented a solution for the w_j's for the case where each of the m materials gives rise to one Gaussian chromatographic peak. However, in general, any of the m materials may contain a considerable number of peaks. Furthermore, there does exist some possibility that one of the components used in the mixture may contain a spurious peak not present in the desired distribution. It is, however, desirable to use this component because of other features in its chromatogram. Thus, the total weight fraction of the resultant chromatogram is

$$\sum_{k=1}^{n} C_k = \sum_{k=1}^{n'} \sum_{j=1}^{m} w_j X_{jk} = 1 \tag{7}$$

where n' may, in some cases, be greater than n, since it must include all of the various peaks present in all m materials.

The closest approximation to the desired distribution can most readily be found by the method of least squares.[7] Let R^2 represent the square of the difference between the desired and obtainable chromatogram. Then,

$$R^2 = \sum_{k=1}^{n''} \left(C_k - \sum_{j=1}^{m} a_j X_{jk} \right)^2 \tag{8}$$

where n'' includes all the individual peaks included in n and n'. In most cases, one would choose materials such that $n' = n$. For this case, then, $n = n' = n''$. However, more generally, cases may arise where $n' \neq n$, and account must be taken of this factor.

Minimizing the sum of the squares of the deviations results in a set of m simultaneous equations of the form

$$\frac{\partial R^2}{\partial a_i} = 0 = \sum_{k=1}^{n''} \left(C_k - \sum_{j=1}^{m} a_j X_{jk} \right) X_{ik} \tag{9}$$

where $i = 1, 2, \ldots, m$. Equation (9) represents the solution to the problem, since the a_j's can now be found. However, the equation is in a very cumbersome form. It can be simplified significantly by rewriting it as

$$d_i = \sum_k \sum_j X_{ik} X_{jk} a_j \tag{10}$$

where $d_i = \sum_k C_k X_{ik}$.

Let

$$B_{ijk} = \sum_k X_{ik} X_{jk} \tag{11}$$

then

$$d_i = \sum_j B_{ijk} a_j \tag{12}$$

or, using matrix notation,

$$\mathbf{d} = \mathbf{Ba} \tag{13}$$

where $\mathbf{d}$ is the column matrix $[d_1, d_2, \ldots, d_m]$ and, from eq. (10), is $\left[\sum_k C_k X_{1k}, \sum_k C_k X_{2k}, \ldots, \sum_k C_k X_{mk}\right]$; $\mathbf{a}$ is the column matrix $[a_1 \ldots a_m]$; and $\mathbf{B}$ is the matrix $\lVert B_{ijk} \rVert$ and, as shown in eqs. (11) and (12), is

$$\lVert B_{ijk} \rVert = \begin{bmatrix} \sum_k X^2{}_{1k} \ldots \ldots \sum_k X_{1k} X_{mk} \\ \sum_k X_{2k} X_{1k} \ldots \ldots \sum_k X_{2k} X_{mk} \\ \cdot \qquad\qquad \cdot \\ \cdot \qquad\qquad \cdot \\ \cdot \qquad\qquad \cdot \\ \sum_k X_{mk} X_{1k} \ldots \ldots \sum_k X^2{}_{mk} \end{bmatrix} \qquad (14)$$

Assuming that the set of simultaneous equations expressed by $\mathbf{d} = \mathbf{Ba}$ consists of m linearly independent equations of m unknowns, then the rank of $\mathbf{B}$ is m. Therefore, the determinant of the matrix defined in eq. (14), $|\mathbf{B}|$, does not vanish, i.e., B is nonsingular, and the inverse matrix $\mathbf{B}^{-1}$ exists.[8] The solution of eq. (13) is then

$$\mathbf{a} = \mathbf{B}^{-1}\mathbf{d}. \qquad (15)$$

From this, the a_j's can be found using Cramer's rule and is

$$a_1 = \frac{1}{|\mathbf{B}|} \begin{vmatrix} d_1 \ldots \sum_k X_{1k} X_{mk} \\ \cdot \qquad \cdot \\ \cdot \qquad \cdot \\ \cdot \qquad \cdot \\ \cdot \qquad \cdot \\ d_m \ldots \sum_k X^2{}_{mk} \end{vmatrix} \ldots a_m = \begin{vmatrix} \sum_k X^2{}_{1k} \ldots \ldots d_1 \\ \cdot \qquad \cdot \\ \cdot \qquad \cdot \\ \cdot \qquad \cdot \\ \cdot \qquad \cdot \\ \sum_k X_{mk} X_{1k} \ldots d_m \end{vmatrix} \frac{1}{|\mathbf{B}|} \qquad (16)$$

where $\lVert \rVert$ indicates the determinant and $|\mathbf{B}|$ is the determinant of the matrix defined in eq. (14).

When the number of equations in (13) is large, Cramer's rule is inefficient since it requires the evaluation of high-order determinants. More practical methods of solving eq. (13) such as the Gaussian reduction method or the Gauss-Jordan method can be used.[8,9] Both of these methods involve iterative procedures and are more readily adaptable to machine calculation.

COMPOUND BLENDING

The methods outlined above permit the judicious selection of the proper amounts of material to best replicate a given chromatogram. However, when using gel permeation chromatography as a tool in formulation, a more general situation may arise where it is desirable to replicate both the chromatogram and a second independent parameter, such as gel point, viscosity, etc. If this second independent parameter is a function of molecular weight, this can best be accomplished analytically by introducing

a set of normalized weighing factors, g_k, such that the chromatogram peaks are weighted in a nonlinear manner. Using this notation, eq. (9) can be written as

$$0 = \sum_{k=1}^{n''} \left(C_k g_k - g_k \sum_{j=1}^{m} a_j X_{jk} \right) g_k X_{jk}. \qquad (17)$$

Subsequent solutions of eq. (17) are exactly analogous to those discussed above, with, however, the inclusion of the numerical values of the g_k's.

As an example of this procedure, in the Appendix an analytical set of g_k are derived for replication of the viscosity as well as the chromatogram. Note that if, by chance, this second parameter is independent of molecular weight, or is linearly proportional to the elution volume, $g_1 = g_2 = g_k$, and eq. (17) reduces to eq. (9).

EXPERIMENTAL RESULTS

As an example to illustrate the above procedure, a simple hypothetical case has been selected. A material of unknown composition is examined and found to have the chromatogram shown in Figure 1. It is most desirous about to formulate this material. However, the materials presently available for formulation have the chromatograms shown in Figures 2 and 3. The problem, then, is what is the blending composition that will most closely approximate the material resolved in Figure 1. To determine the fractional area beneath each peak, each of the chromatograms can be analyzed using either the methods outlined in an earlier publication[1] or, as
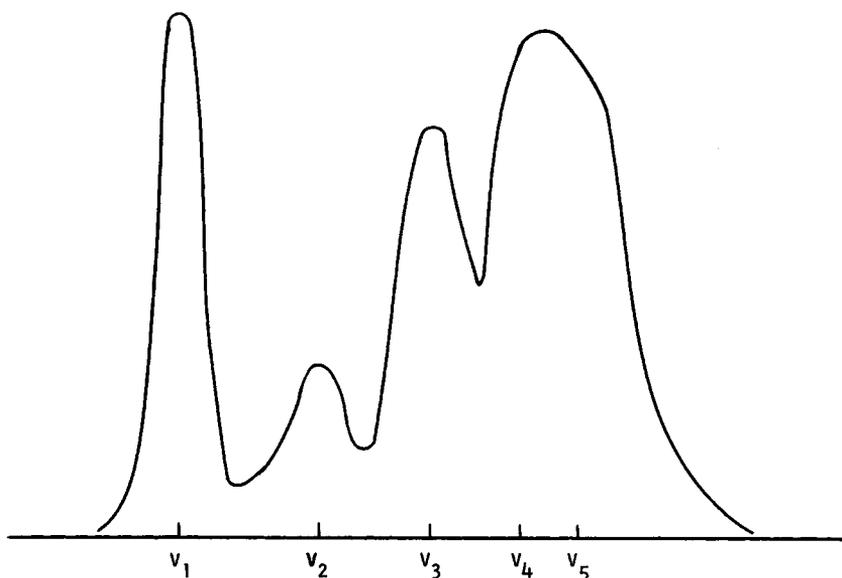

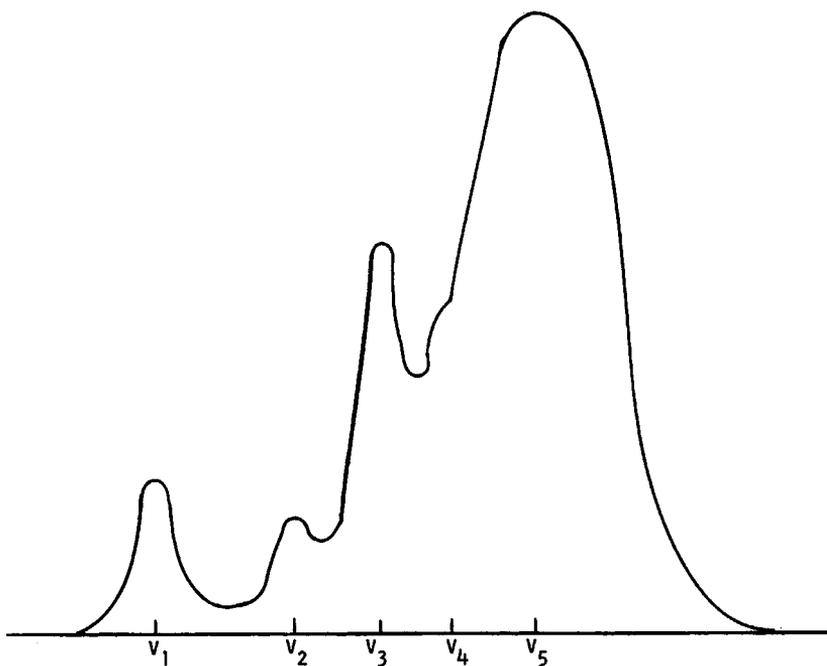Fig. 1. Chromatogram of an unknown material.
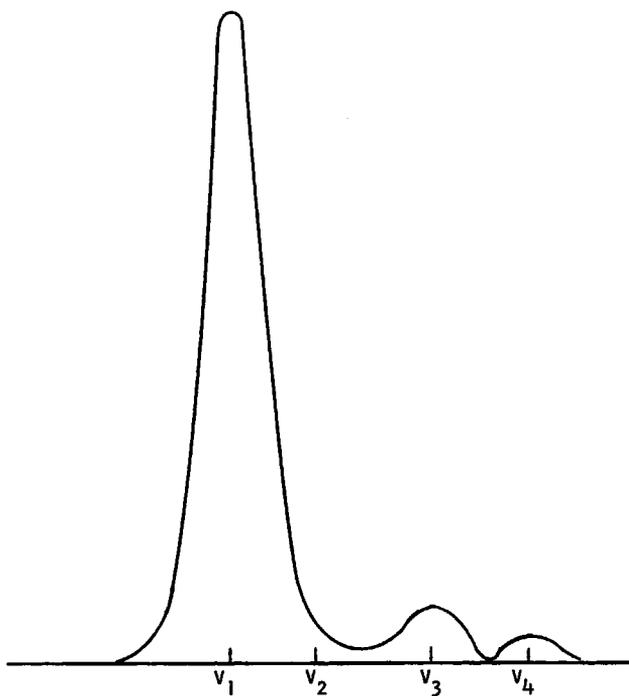
Fig. 2. Chromatogram of material A.

Fig. 3. Chromatogram of material B.

TABLE I
Fractional Areas in the Various Peaks

	V_1	V_2	V_3	V_4	V_5
Figure 1	.21	.08	.18	.10	.43
Figure 2	.05	.06	.14	.16	.59
Figure 3	.75	.09	.10	.06	0

TABLE II
Comparison of Calculated and Actual Peak Ratios

	V_1	V_2	V_3	V_4	V_5
Actual (Fig. 1)	.21	.08	.18	.10	.43
Calculated	.22	.07	.13	.14	.44

was done here, using a planimeter. The results for the five components found are tabulated in Table I.

For this system, eq. (16) results in

$$a_2 = \frac{1}{D} \left[\sum c_k X_{1k} \sum X_{1k} X_{2k} - \sum c_k X_{2k} \sum (X_{1k})^2 \right] \tag{18}$$

$$a_1 = \frac{1}{D} \left[\sum c_k X_{2k} \sum X_{1k} X_{2k} - \sum X_{1k k} c \sum (X_{2k})^2 \right] \tag{19}$$

where

$$D = \left(\sum X_{1k} X_{2k} \right)^2 - \sum (X_{1k})^2 \sum (X_{2k})^2 \tag{20}$$

and the summations are over the five components present.

The weight fraction of each material is given by eq. (6) and, for this case, is tabulated in Table II. As can be seen, the method does reproduce the optimum spectra under the conditions imposed on it.

SUMMARY

This paper is the last in a series of publications designed to enhance the practical utility of gel permeation chromatography. In the two previous publications, methods have been detailed for the resolution of a chromatogram into its component peaks,[1] for the prediction of GPC calibration curves,[2] and for a graphic method of obtaining molecular weight averages.[2] This paper presents a general method of formulation, when the spectra of the desired material (the model) and the spectra of the materials available are known. This theory should prove to be valuable for many blending operations. Furthermore, as formulated, this theory is not specific to GPC and, in fact, is applicable to analytical data obtained from any of a number of sources.

Appendix

In a blending operation, it is sometimes desirable to reproduce the viscosity as well as the chromatogram of model material. If by suitable blending, an exact reproduction of

the desired chromatogram could be achieved, then *a priori* the desired viscosity will also be obtained. However, this is an ideal rarely achievable. Inevitably, in formulation work, some compromise must be reached between the desire to reproduce the chromatogram and the desire to reproduce a given viscosity. This section outlines a rational method for obtaining this objective, for the problem where the model material consists of polymer molecules that are homologues of the same monomer. For this case,

$$\ln = M = A - BV \tag{A-1}$$

and

$$\ln \eta = \ln C + D \ln M \tag{A-2}$$

where A, B, C, and D are constants, M is the molecular weight, and η is the viscosity. Both these equations are, of course, only approximately valid over wide molecular weight ranges. However, for the purpose of this discussion, it will be assumed that these equations are valid in the molecular weight range of interest. It is of considerable importance to evaluate the parameter D in the same solvent and at the same temperature as used in the GPC calibration curve, eq. (A-1).

The fluidity, φ, of a component of elution volume V can then be found by combining eqs. (A-1) and (A-2) and is

$$\varphi = \eta^{-1} = C_1 e^{C_2 V} \tag{A-3}$$

where $C_2 = BD$ and $C_1 = e^{-DA}/C$. Assuming that the fluidities are additive, the fluidity of a mixture containing k components is found to be

$$\varphi = \sum_k X_k \varphi_k = C_1 \sum_k X_k e^{C_2 V_k} \tag{A-4}$$

where X_k is the weight fraction of component k.

From eq. (A-4) the weighing factors g_k can be found for eq. (15) and are

$$g_k = e^{C_2 V_k} / \sum_k e^{C_2 V}. \tag{A-5}$$

Qualitatively, these results imply that in a blending experiment, if one desires to optimize the replication of both the chromatogram and the viscosity of a model material, it is most important to obtain as close a match as is possible with the higher elution volume (lower molecular weight) components. These components must therefore, be more heavily weighed in eq. (17), and this is accomplished by the use of eq. (A-5).

It is interesting to note that the set g_k is dependent only on the product of $B \cdot D$. If materials are present whose constant $B \cdot D$ differ, they must each be evaluated separately. This would be the case, in general, for a mixture of polymers that are not derivatives of the same monomer. For this case, C_2 in the above equations should be replaced by C_{2k}, for the kth component. However, both B and D vary over a narrow range and often the assumption that $B \cdot D = $ constant is quite reasonable.

References

1. M. Schrager and A. L. Ward, *J. Appl. Polym. Sci.*, **14**, 1235 (1970).
2. M. Schrager, *J. Appl. Polym. Sci.*, **15**, 83 (1971).
3. W. N. Smith, *J. Appl. Polym. Sci.*, **11**, 639 (1967).
4. L. H. Tung, *J. Appl. Polym. Sci.*, **10**, 345 (1966).
5. M. Hess and R. F. Kratz, *J. Polym. Sci.*, A-2, **4**, 731 (1966).

6. N. Bayer, K. Fajans, and S. Z. Lewin, in *Techniques of Organic Chemistry*, Vol. 1, Part II, A. Weissberger, Ed., Interscience, New York, 1960.

7. W. E. Deming, *Statistical Adjustment of Data*, Dover, New York, 1964.

8. G. Hadley, *Linear Algebra*, Addison Wesley, Reading, Mass., 1961, Chap. 5.

9. H. Margenau and G. M. Murphy, *The Mathematics of Physics and Chemistry*, 2nd ed., D. Van Nostrand, Princeton, N. J., 1961, p. 497.