

CHROM. 12,997

SEARCH FOR OPTIMAL ELUENT COMPOSITION FOR ISOCRATIC LIQUID COLUMN CHROMATOGRAPHY

VRATISLAV SVOBODA

Institute for Research, Production and Application of Radioisotopes, Přístavní 24, 170 04 Prague 7 (Czechoslovakia)

SUMMARY

Parabolic equations for the description of capacity factor measurements in single- and multi-component eluents are proposed. They are used together with empirical reduced plate height equations for the computation of the separation time of a given set of compounds assuming that the column length may be varied to attain the pre-determined resolution. The merits of a systematic (point for point) and simplex search for the optimal separation time if the maximal column length is set *a priori* are compared and results presented for the separation of several nucleotide mixtures. For a multi-component eluent the best compromise seems to be a simplex search started several times from different locations. For a single-component eluent the systematic search is reasonably fast and most dependable.

INTRODUCTION

In adsorption or reversed-phase chromatography of non-ionizable compounds the choice of the proper eluent composition may be relatively straightforward, as the relationship between the structure of a molecule and its chromatographic behaviour is well understood and may be described by a set of semi-empirical rules^{1,2}. Mostly only one parameter, the polarity of the eluent, has to be changed in order to achieve the optimal isocratic separation. When the sample consists of ionizable organic compounds, several mechanisms may compete and the detailed prediction of the elution sequence of closely related compounds becomes difficult, if not impossible. In these instances, mostly for ion-exchange, ion-pair or reversed-phase chromatography of compounds of biological interest, more than one parameter of the eluent has to be chosen (*e.g.*, pH, concentration of buffer salt, concentration of organic solvent and concentration of ion-forming agent). If the optimum is searched for only experimentally by a trial and error method, the number of experiments may become excessive. If a systematic search is performed and every variable attains p values, then the number of steps examined increases exponentially with the number of variables, m :

$$s = p^m \quad (1)$$

The aim of this work was to reduce the number of experiments necessary and to make use of efficient digital search methods.

APPROXIMATION OF CHROMATOGRAPHIC SYSTEM

Equation for capacity factor

The capacity factor, k' :

$$k' = \frac{V_R - V_0}{V_0} \quad (2)$$

where V_R is the retention volume and V_0 the void volume, may be well approximated by a parabolic equation for most compounds. For one variable it has the form

$$y = a_0 + a_1x_1 + a_2x_1^2 \quad (3)$$

for two variables

$$y = a_0 + a_1x_1 + a_2x_2 + a_3x_1^2 + a_4x_2^2 + a_5x_1x_2 \quad (4)$$

and similarly for four variables

$$y = a_0 + a_1x_1 + a_2x_2 + a_3x_3 + a_4x_4 + a_5x_1^2 + a_6x_2^2 + a_7x_3^2 + a_8x_4^2 + a_9x_1x_2 + a_{10}x_1x_3 + a_{11}x_1x_4 + a_{12}x_2x_3 + a_{13}x_2x_4 + a_{14}x_3x_4 \quad (5)$$

where

$$y = \log k' \quad (6)$$

and any x may be defined, for example, as

$$x = \log M \quad (7)$$

where M is the molarity or normality of a buffer salt or ion-forming agent, or

$$x = \text{pH} \quad (8)$$

or, finally,

$$x = 0.01 V \quad (9)$$

where V is volume percentage concentration of organic solvent.

The number of terms, q , in a general parabolic equation depends on the number of variables:

$$q = (2 + 3m + m^2)/2 \quad (10)$$

(see Table I). Therefore, only q values of independent experimental points are necessary for the computation of all coefficients of the parabolic equation. If more experimental points are available, the coefficients of the parabolic equation are computed using the least-square error approach. Much may be discussed about the validity of parabolic equations. However, in all practical examples examined in our laboratory good fits were attained. Obviously, the necessary condition is that the range of variables is not too wide, otherwise higher degree polynomials should have been used. On the other hand, linear equations (*e.g.*, eqn. 2 or 3 in ref. 3) are oversimplified and obviously cannot describe without gross errors many experimentally verified k' functions.

TABLE I

NUMBER OF VARIABLES AND COEFFICIENTS IN GENERAL PARABOLIC EQUATION AND STEPS s (EQN. 1)

<i>Number of:</i>			
<i>Variables (m)</i>	<i>Coefficients (q)</i>	<i>Steps (s)</i> ($p = 3$)	<i>Steps (s)</i> ($p = 10$)
1	3	3	10
2	6	9	100
3	10	27	1000
4	15	81	10,000
5	21	243	100,000
6	28	729	1,000,000

Equation for reduced height of a theoretical plate

In contrast to capacity factor equations, the influence of eluent composition on the reduced height of a theoretical plate⁴ (h) has not been thoroughly studied, as far as we know. There exists a theoretical study⁵ of the influence of capacity factor on h , but its verification under practical conditions may be very complicated. The measurement of h is usually subject to large experimental errors, and therefore an approximation by a higher degree polynomial would serve no useful purpose.

The linear equation for reduced plate height:

$$\log h = b_0 + \sum b_n x_n \quad (11)$$

where h is defined as

$$h = \frac{l}{Nd} \quad (12)$$

(l = column length, N = number of theoretical plates, d = sorbent particle diameter) reflects with reasonable precision the influence of variable eluent composition. Constant pumping speed, temperature, etc., are assumed.

Column length required

If the sample components are ordered according to their elution volumes, V_n ,

then the required column length for a given component pair to achieve the desired resolution, R :

$$R = \frac{V_{n+1} - V_n}{\sigma_n + \sigma_{n+1}} \quad (13)$$

is given by

$$l = \frac{R^2 d [\sqrt{h_n(k'_n + 1)} + \sqrt{h_{n+1}(k'_{n+1} + 1)}]^2}{(k'_{n+1} - k'_n)} \quad (14)$$

The required column length is the greatest value computed for all component pairs. All other pairs will be eluted at a higher resolution than desired. The elution time (T) of the last eluted compound equals the necessary separation time:

$$T = l(k'_z + 1)/v \quad (15)$$

where k'_z is the capacity factor of the last eluted compound and v is the linear eluent speed:

$$v = F/\pi r^2 p \quad (16)$$

(r = column diameter, F = pumping speed, p = porosity of the column bed).

Using the empirical coefficients of the k' and h equations, we are able to compute for every given sample and eluent composition the necessary separation time, T . All of these computations are valid only within the experimentally studied limits. Extrapolation may lead to serious errors.

Systematic and simplex searches

If a systematic search with fine division of every variable is performed, the number of computed points (steps) quickly becomes prohibitive with a finer division of every variable or with an increasing number of variables. However, this is the only approach that guarantees the attainment of the true optimal (shortest) separation time within the experimentally established range of variables.

The algorithm of the simplex search, with a variable step length as proposed by Nelder and Mead⁶ and exemplified for analytical problems by Deming and Morgan⁷ and others⁸, is much faster but may result in a local optimum. It was slightly modified for a search for the optimal separation time with a pre-determined maximal allowed column length.

A simplified block diagram of the program written in Basic (E Basic version for the Varian 620L) is presented in Fig. 1. The number of points in the simplex is $m + 1$. The worst point W is reflected along the centroid formed from the remaining points. If the reflected point falls outside the pre-determined limits, R is set on the limit and no attempt is made to lengthen the step. If R is better than the best point B , a point with a twice longer step S is computed. If R is worse than B , successive points with a shorter step length U and T are computed. If everything else fails, a new simplex with a four times shorter distance between points than the original one is formed. The search is ended if either after every ten steps no improvement is observed or if the step length is shorter than $1 \cdot 10^{-4}$ times the step length of the original simplex.

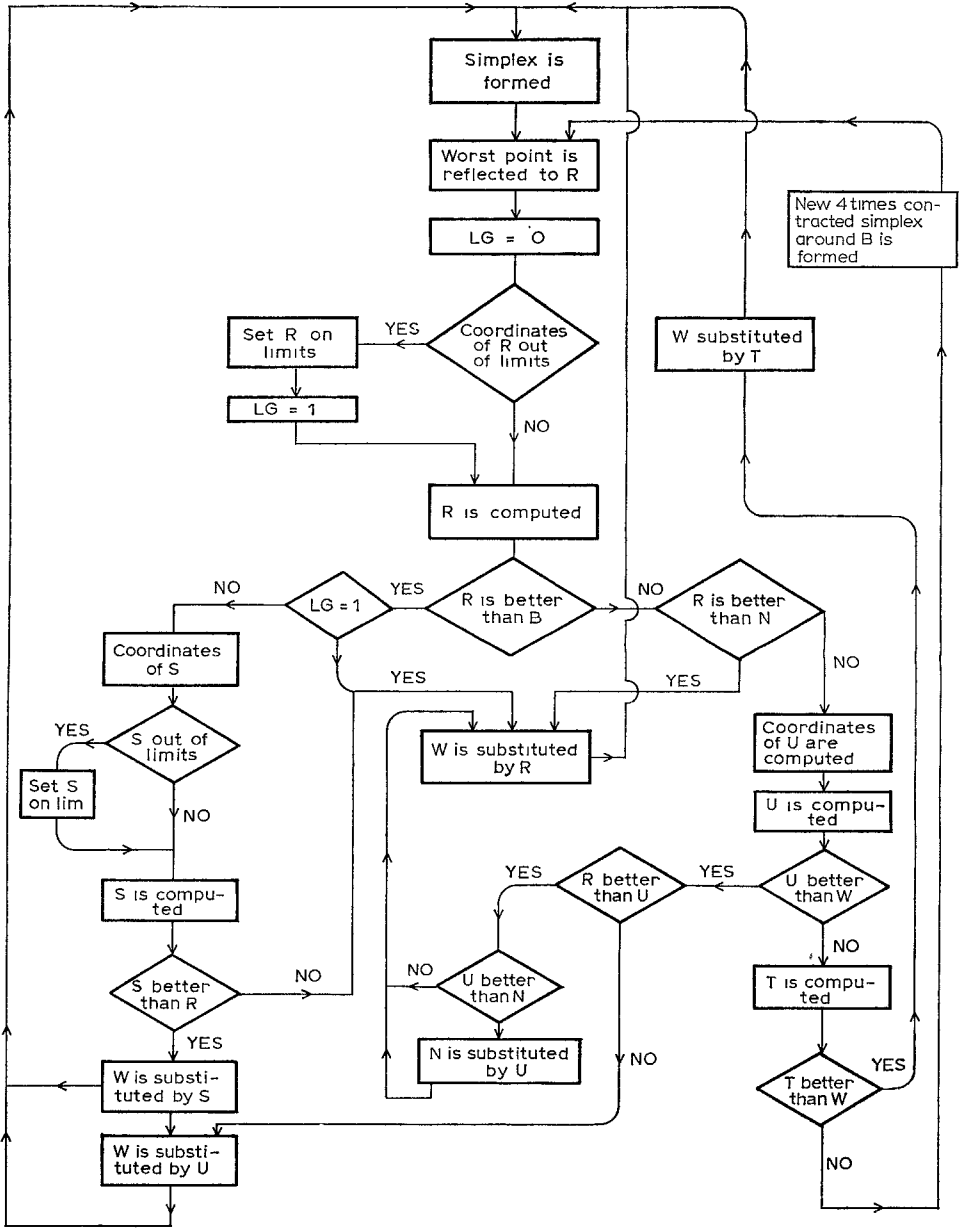


Fig. 1. Block diagram of simplex search program.

The points within every simplex are ordered according to two criteria. Computed points are divided into two groups: the first is formed from points with a column length shorter than or equal to the pre-determined column length, and the other is formed from the remaining points with a larger column length. Within the first group are the points ordered according to the separation time; then follow the

points from second group, which are ordered according to the column length. So, if in the original simplex no point has a satisfactory column length, the simplex is directed into the region where the column length is within the pre-determined limits. Then the search is directed to attain the shortest possible separation time.

RESULTS AND DISCUSSION

All experimental details of the system from which the constants for the k' and h equations are taken are described in a companion paper⁹. The chromatographic behaviour of twelve nucleotides was studied using an eluent with three variables: concentration of ammonium formate (0.05–1 M), pH (3–5) and volume percentage concentration of ethanol (0–40%). In the figures, the separation time, T (in hours), for a resolution $R = 4$ with a maximal allowed column length of 1000 mm is plotted. The points with a computed column length longer than the maximum permissible are connected with all four net lines dotted. In a three-dimensional presentation only a subspace from the original four-dimensional space (time plus three variables) may be presented: therefore, one variable is held constant.

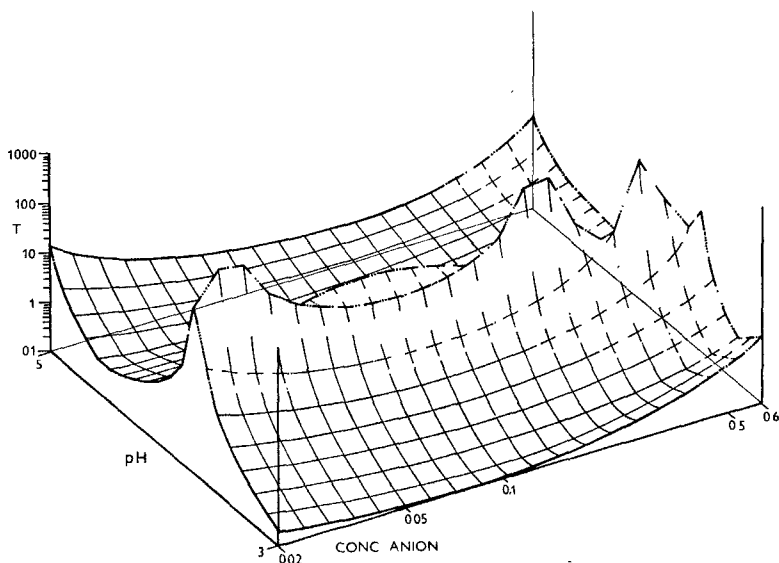


Fig. 2. Separation time, T (in hours), for 5'-CMP, 5'-UMP, 5'-AMP and 5'-GMP. Variables: pH (linear) and molarity of ammonium formate (logarithmic). No ethanol in eluent.

In Fig. 2 a "ridge" can clearly be seen at a pH slightly below 4. This reflects the well known behaviour of two nucleotides, 5'-UMP and 5'-AMP, in most ion-exchange systems. At this pH the speed of elution of both compounds is about equal, 5'-AMP being more sensitive to pH changes than 5'-UMP. In fact, if the net in Fig. 3 were fine enough, for every anion concentration T should reach an infinite value somewhere in this pH region. The minimum at lower pH is deeper; it is located at the pH limit of the system and is rather wide (Fig. 3). With increasing anion concentration

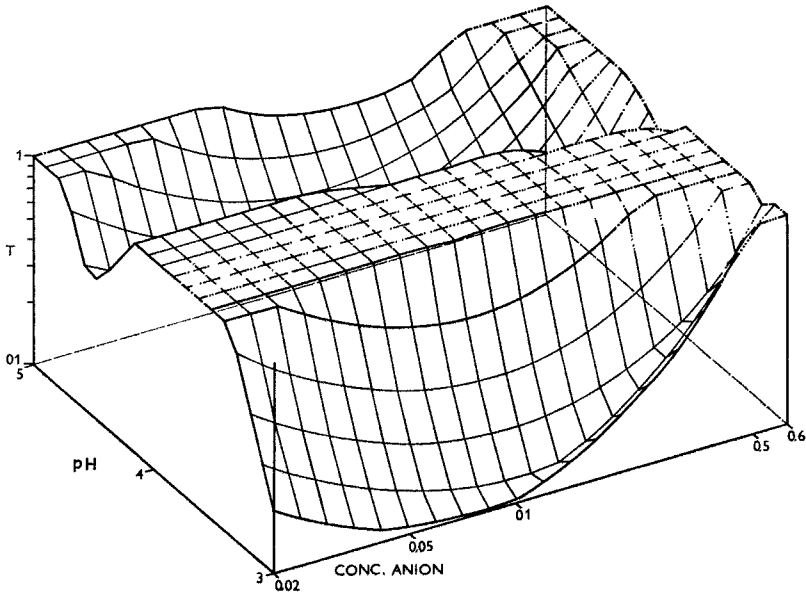


Fig. 3. As Fig. 2, time scale increased.

the column length increases; at the highest anion concentration it exceeds the maximal permitted length.

Another example in the same coordinates (Fig. 4) also has the local minimum near the low pH limit. However, the plot along another two variables (Figs. 5 and 6) reveals that in this instance the minimal separation time may be reached only with

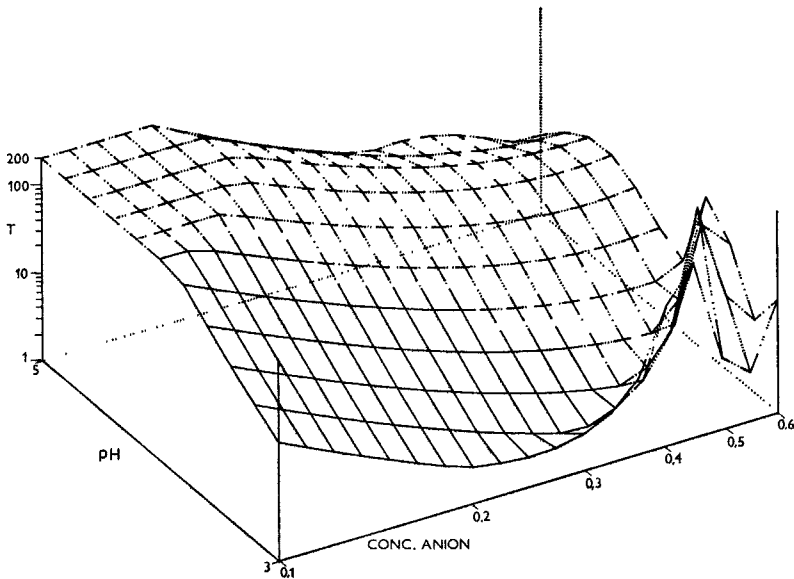


Fig. 4. As Fig. 2, for 5'-CDP, 5'-ADP, 5'-CTP and 5'-ATP. Concentration of ethanol: 40%.

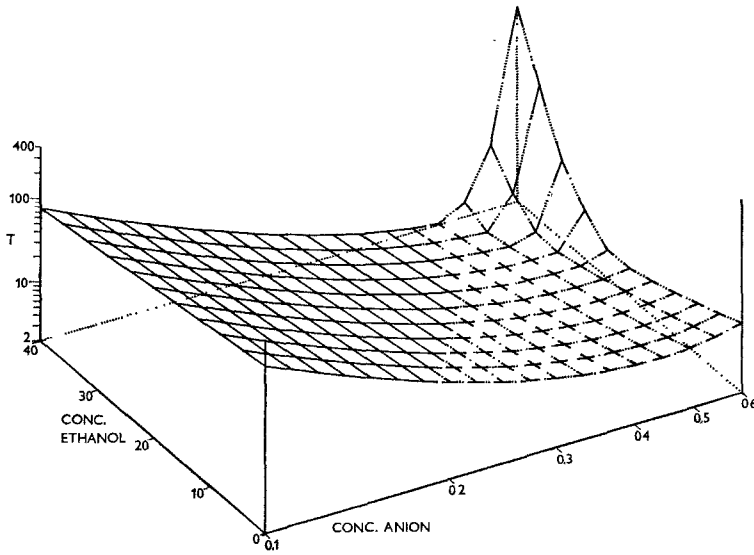


Fig. 5. Separation time for 5'-ADP, 5'-ATP, 5'-UDP and 5'-UTP. Variables: molarity of ammonium formate (logarithmic) and volume percentage of ethanol (linear). pH: 3.5.

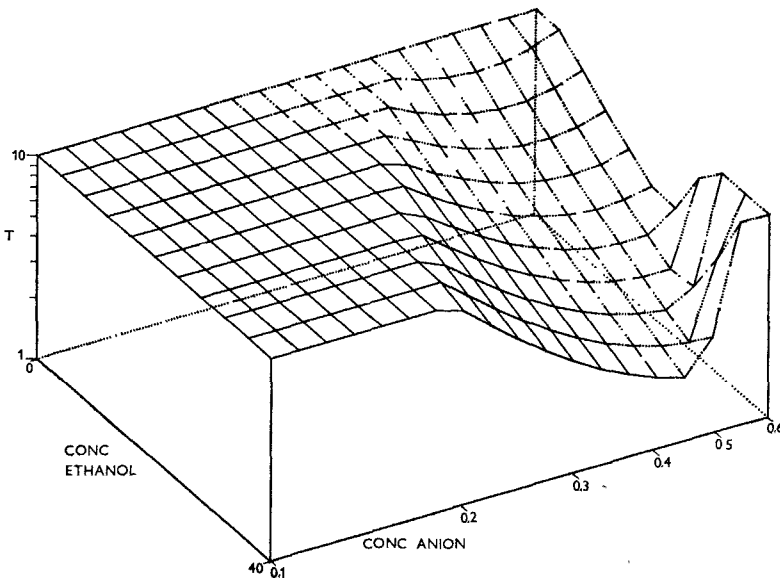


Fig. 6. As Fig. 5, time scale increased and coordinates related to show the region around the optimum.

a column longer than 1000 mm; with shorter columns the necessary separation time would be longer. It is significant that even though for all four compounds the reduced plate height increases with increasing ethanol concentration, the optimum is located at the high ethanol concentration limit owing to the slight increase in the selectivity of this separation with increasing ethanol concentration.

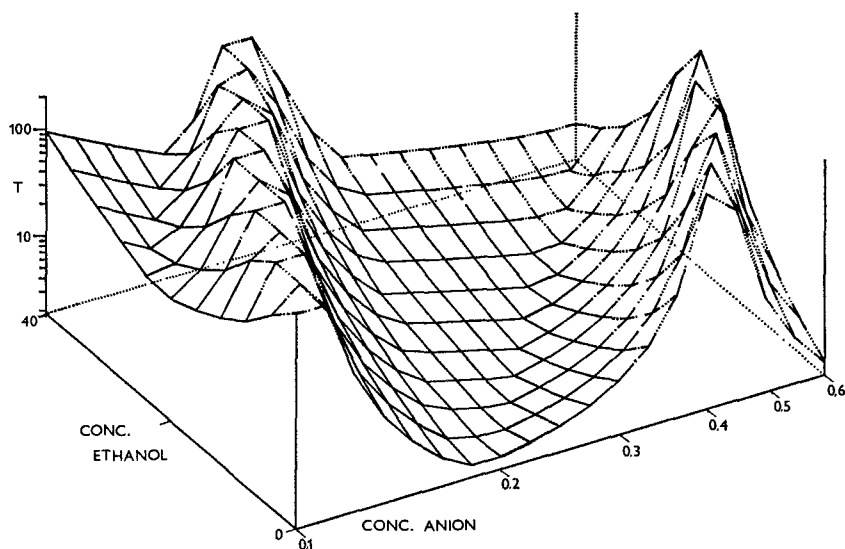


Fig. 7. Separation time for 5'-CDP, 5'-ADP, 5'-CTP and 5'-ATP. Variables as in Fig. 5; pH, 4.726.

From this example it is obvious that the search with more than one variable may be complicated and the purely experimental "trial and error" approach may sometimes lead to erroneous conclusions.

How misleading just the simplified graphical presentation of more than a two-dimensional problem can be is illustrated by the separation of cytosine and adenosine di- and trinucleotides. In Fig. 7 two areas where local optima could be located are seen. On the other hand, from Fig. 8 it seems be obvious that the lowest pH results

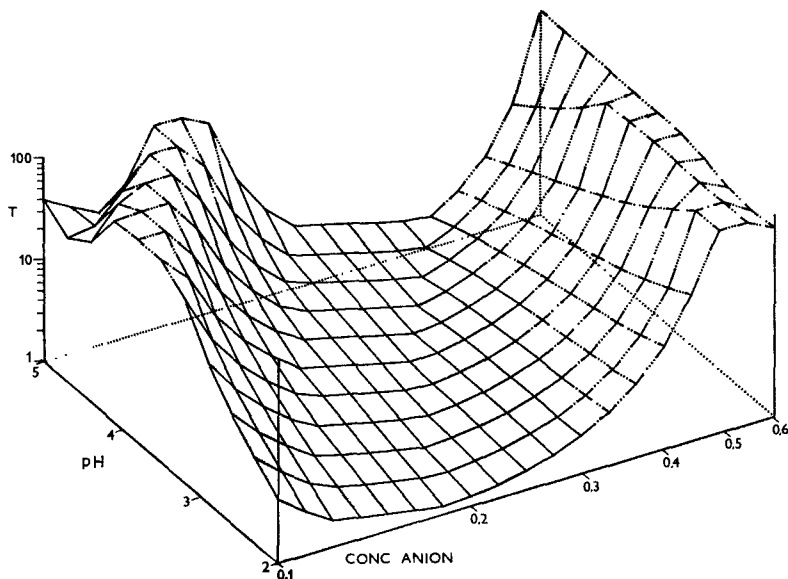


Fig. 8. As Fig. 7, variables as in Fig. 2, constant ethanol concentration (20.295%).

in the quickest separation. In concentration of ethanol–pH coordinates (Fig. 9) there seems to be an optimal pH at about 3.7 at the highest ethanol concentration. In the plot of concentration of ethanol *versus* anion concentration (Fig. 10) the “valley” points both to high ethanol and anion concentrations; however, it ascends sharply when the upper limit of both variables is attained. In Figs. 7, 8 and 10 a “ridge” and more than one local optimum are evident.

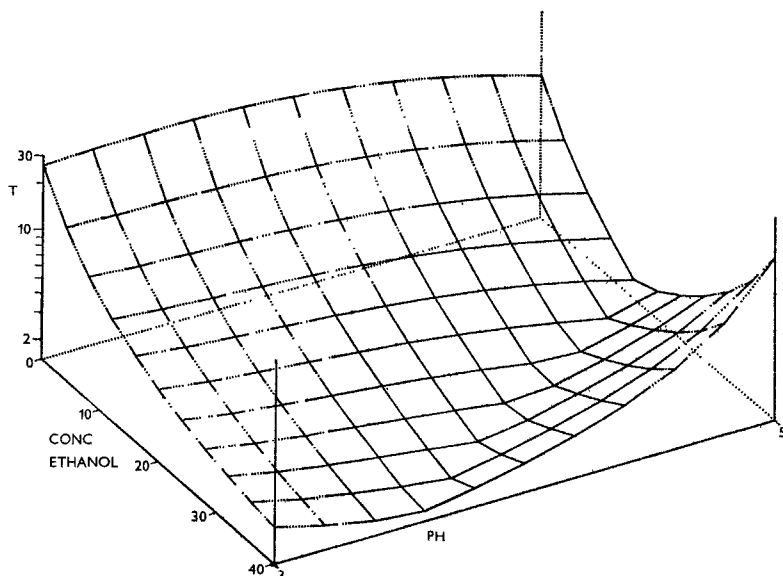


Fig. 9. As Fig. 7, variables pH (linear) and concentration of ethanol (vol-%, linear). Ammonium formate concentration: 0.3476 *M*.

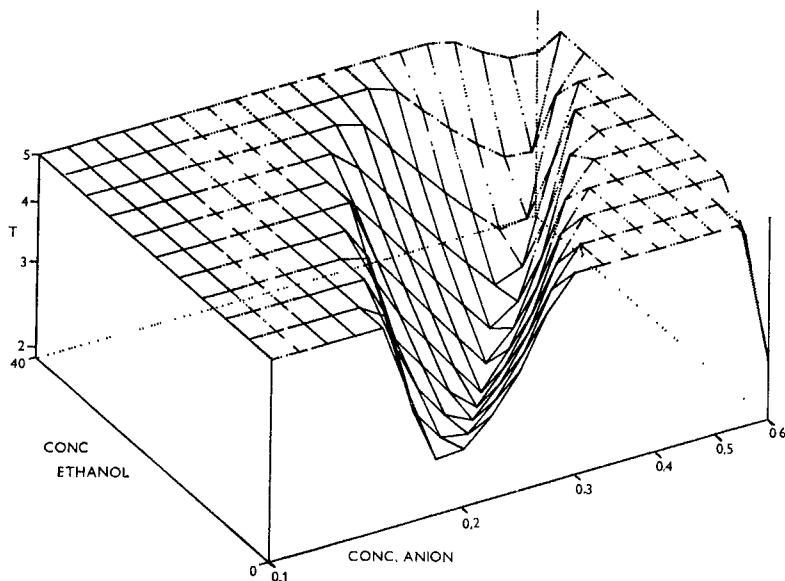


Fig. 10. As Fig. 7, time scale increased.

The results of systematic and simplex searches starting from different points are presented in Tables II and III. In the first attempt at a systematic search with 64 steps, an eluent composition far from the optimum was found. When the number of steps was increased, values not far from the optimum were found, and with even shorter steps (number of steps 900) a point even further away from the optimum was found. In the simplex search (Table III) from various starting points five from ten (Nos. 1, 3, 4, 5 and 9) led to a point very near to the optimum. The second trial found a local optimum. It is therefore desirable to start the simplex search from two or three different points so as to be sure that the absolute minimum (in given limits) is found.

TABLE II

SYSTEMATIC SEARCH FOR OPTIMAL SEPARATION OF 5'-CDP, 5'-ADP, 5'-CTP AND 5'-ATP

$R = 4$; maximal permitted column length = 1000 mm. Limits of variables: ammonium formate concentration, 0.1–0.6 M ; pH, 3.5–5; concentration of ethanol, 0–40%.

Number of steps			Total number of steps	Coordinates of optimum			Optimal column length (mm)	Separation time (h)
Ammonium formate	pH	Ethanol		Ammonium formate (M)	pH	Ethanol concentration (%)		
4	4	4	64	0.1817	3.5	13.33	77.7	1.7223
7	7	7	343	0.2449	3.5	26.67	174.7	1.5382
10	9	10	900	0.2706	3.5	31.11	248.7	1.5413

TABLE III

SIMPLEX SEARCH FOR OPTIMAL SEPARATION OF 5'-CDP, 5'-ADP, 5'-CTP AND 5'-ATP

$R = 4$. Limits of variables as in Table II.

No.	Initial coordinates			Total No. of steps	Coordinates of optimum			Optimal column length (mm)	Separation time (h)
	Ammonium formate (M)	pH	Ethanol concentration (%)		Ammonium formate (M)	pH	Ethanol concentration (%)		
1	0.2	4.5	30	120	0.2503	3.5000	27.736	187.10	1.5261
2	0.3	4	20	90	0.2975	3.6337	32.38	308.66	1.5831
3	0.1	3.5	0	179	0.2643	3.5	29.77	224.98	1.5216
4	0.6	5	40	172	0.2619	3.5	29.42	218.03	1.5219
5	0.1	5	40	142	0.2514	3.5	27.89	189.77	1.5257
6	0.6	3.5	0	19	Unsuccessful (limiting column length: 1000 mm)				
7	0.6	3.5	40	115	0.3353	3.5	39.91	590.73	1.6580
8	0.6	5	0	16	Unsuccessful (limiting column length: 1000 mm)				
9	0.1	3.5	40	115	0.2645	3.5	29.80	225.74	1.5216
10	0.1	5	0	61	0.1	5	10.7	68.05	19.07

Searches started from extreme "apexes" of the experimentally studied space of variables (Nos. 3–10) show the ability of the simplex to find the optimum even when the boundaries restrict its moves drastically. Two unsuccessful searches ended only after a small number of steps (16 and 19) and therefore not too much computer

time was wasted; another two (Nos. 7 and 10) ended in the "corners" of the restricted space, where local optima were found. It is obviously advantageous to start the search from points near to the centre of the studied region.

CONCLUSIONS

When a chromatographic system with a multi-component eluent is described using parabolic capacity factor equations, only a small number of experimental points need to be measured. Using these empirical equations for capacity factor and the reduced height of a theoretical plate, the optimal composition of the eluent and the optimal column length may be found with a simplex search. This approach is justified only if two or more parameters of the eluent have to be controlled or if the optimal composition of the eluent and of the solid phase (one or more parameters) is searched for. This application will be demonstrated in a subsequent publication.

REFERENCES

- 1 L. R. Snyder, *Anal. Chem.*, 46 (1974) 1384.
- 2 L. R. Snyder and J. J. Kirkland, *Introduction to Modern Liquid Chromatography*, Wiley, New York, 1974.
- 3 P. Jandera and J. Churáček, *J. Chromatogr.*, 192 (1980) 1.
- 4 P. A. Bristow and J. H. Knox, *Chromatographia*, 10 (1977) 279.
- 5 J. F. K. Huber, *Ber. Bunsenges. Phys. Chem.*, 77 (1973) 179.
- 6 J. A. Nelder and R. Mead, *Comput. J.*, 7 (1965) 308.
- 7 S. N. Deming and S. L. Morgan, *Anal. Chem.*, 45 (1973) 278A.
- 8 M. W. Routh, P. A. Swartz and M. P. Denton, *Anal. Chem.*, 49 (1977) 1422.
- 9 I. Kleinmann and V. Svoboda, *J. Chromatogr.*, 201 (1980) 265.