Journal of Chromatography, 79 (1973) 157–163

CHROM. 6587

THE USE OF INFORMATION THEORY FOR EVALUATING THE QUALITY OF THIN-LAYER CHROMATOGRAPHIC SEPARATIONS

D. L. MASSART
Pharmaceutical Institute, Vrije Universiteit Brussel, Paardenstraat 67, B-1640 Sint Genesius Rhode (Belgium)
(Received November 27th, 1972)

SUMMARY

The possibility of using information theory for characterizing thin-layer chromatographic separations is investigated. It appears that the information content, as defined by the Shannon equation, can be used to compare the merits of different solvents used for the separation of the same group of compounds.

INTRODUCTION

One of the aims of the analytical chemist is to shorten the time required to carry out an analysis or a series of analyses, and much effort has been expended on automation, acquisition of data by computer, etc. It is difficult to choose a suitable analytical procedure as there are often many alternatives from which the most appropriate method has to be chosen or partial procedures from which a complete process must be established. Therefore, there is a need for a logical procedure that enables a choice to be made on a rational basis. A basic paper on this subject was written by KAISER[1], and other workers have also been active in this field[2,3]. We have published papers on the use of graph theory and dynamic programming for the choice of separation schemes[4,5].

It is common in papers on chromatography for a proposed separation to be described by phrases such as "a good (or reasonable, or excellent) separation of ... was obtained". It is therefore difficult, when trying to choose a suitable method, to interpret such phrases so that rational decisions are possible, *i.e.*, to assign a numerical value to each possible alternative. In this paper we report on the application to separation chemistry of one method of achieving this, namely the use of information theory. We chose to apply the method to thin-layer chromatography, but generalization is possible.

INFORMATION CONTENT

There are many textbooks on information theory, and a very concise description of the first principles was given, for example, by NAHIKIAN[6]. When using thin-layer chromatography as a method for qualitative analysis, one usually knows that the substance to be identified is a member of a specific class (for example, fat-soluble vitamins or sulphonamides). Parallel runs are carried out for the unknown compound and a set of standards and the observed R_F values are compared. For many of these classes, several separations have been described and it is necessary to know which separation will yield the most information. This has to be decided from tables which define sets of possible R_F values, $R_{F1}, R_{F2}, \ldots, R_{Fn}$, from which a particular one will be obtained by chromatography of the unknown.

As resolution up to 0.01 R_F unit is improbable, it is better to consider groups of R_F values. We have chosen arbitrarily to regard R_F values that are 0.03 or less apart as members of one such group. The tabulated R_F values are arranged in ascending order and divided into groups, R_k, so that no group contains R_F values that differ by more than 0.03. There are many possible ways to classify R_F values into groups and the method used here is not claimed to be the best. For values of n not exceeding 25, it seems, however, at least as good as the procedure of SIMON AND LEDERER[7] of dividing the complete R_F range into twenty groups, 0.00–0.05, 0.06–0.10, and so on. This leads to substances with R_F values of 0.10 and 0.11 being considered as separated and those with R_F values of 0.11 and 0.15 as not separated. However, for larger values of n, such a procedure becomes the only practical one. For each of the m groups, there is a distinct probability, p_k ($p_1, \ldots, p_k, \ldots, p_m > 0$, $\sum_{k=1}^{m} p_k = 1$), that the unknown compound will appear to have an R_F value within the limits of this group, *i.e.*, that the unknown compound will be a member of the group with these R_F values. If one considers that there is an equal probability of occurrence for each member of the class that is subjected to the separation, the probability, p_k, of finding an R_F value from the group R_k containing r_k members of the n which comprise the total class is r_k/n. This is a situation in which it is assumed that there is no prior information on the relative frequency of occurrence of the members of the class. When taking this prior information into account, only slightly more complex equations will be obtained.

For the finite probability scheme

$$\begin{bmatrix} R \\ p \end{bmatrix} = \begin{bmatrix} R_1 & R_2 \ldots R_k \ldots R_m \\ p_1 & p_2 \ldots p_k \ldots p_m \end{bmatrix}$$

there is an uncertainty as to which event, R_k, will occur, *i.e.*, to which R_F group the unknown compound will belong. Information theory shows that this uncertainty can be described by

$$I = -\sum_{k=1}^{m} p_k \cdot {}^2\log p_k \text{ (Shannon equation)}$$

$$= -\sum_{k=1}^{m} \frac{r_k}{n} \cdot {}^2\log\left(\frac{r_k}{n}\right) \tag{1}$$

where I is the information content, ${}^2\log$ is the logarithm to the base two and the result is expressed in bits.

PROPERTIES OF THE INFORMATION CONTENT AS APPLIED TO THIN-LAYER CHROMATOGRAPHY

It can be shown, by using some examples, that I is indeed a measure of the information content and that it can be used to assign a numerical value to the merit of a thin-layer chromatographic separation.

If I is to measure correctly the uncertainty and the information content, it must possess the following properties.

(a) The information content is zero when there is no uncertainty as to which event will occur. If, for example, all of the members of a class of substances with n members have the same R_F value, there is no uncertainty as to which R_F value will be found and also no information regarding the nature of the unknown compound. Eqn. 1 then reduces to

$$I = -\,{}^2\log 1 = 0$$

(b) The information content has its maximum value when the uncertainty is maximum. The mathematical proof of this property can be found, for example, in ref. 6. For thin-layer chromatography, this represents the situation that in a given class of n substances each member belongs to a different R_F group (or $r_k = 1$ for each group). The value of I is then given by

$$I = -\,n(1/n)\cdot{}^2\log(1/n) = {}^2\log n \qquad (2)$$

From the practical definition of an R_F group used here, it can be seen that the maximum information content in one-dimensional chromatography is ${}^2\log 25 = 4.64$ bits.

(c) The information content is an additive property. To show this, an example can be considered (see Table I).

In order to characterize eight substances, one needs ${}^2\log 8 = 3$ bits. It should then be possible to identify with certainty one substance out of a total of eight by using two runs, for example a 2-bit and a 1-bit run. If solvents 1 and 2 are used, it is indeed possible to achieve this identification, so that it is clear that the information content is additive. In practice, some of the information content may be wasted. Consider, for example, a combination of solvents 1 and 3 with a total of 3 bits, so that it should be possible to identify one of eight compounds. However, it is observed that with this combination of solvents such an identification cannot be achieved, because some of the information that is obtained is identical for both solvents. Therefore, it is theoretically true that the information content is additive. Also, when developing a

TABLE I

R_F VALUES OF EIGHT SUBSTANCES IN THREE DIFFERENT SOLVENTS

Substance	*Solvent 1*	*Solvent 2*	*Solvent 3*
A	0.20	0.20	0.20
B	0.20	0.40	0.20
C	0.40	0.20	0.20
D	0.40	0.40	0.20
E	0.60	0.20	0.40
F	0.60	0.40	0.40
G	0.80	0.20	0.40
H	0.80	0.40	0.40
Information content	2	1	1

TABLE II

hR$_F$ VALUES OF DDT AND RELATED COMPOUNDS AND INFORMATION CONTENT OF THE PROPOSED SEPARATIONS
R_F values were taken from ref. 9.

Solvent system	*p,p'-DDT*	*o,p'-DDT*	*p,p'-DDE*	*o,p'-DDE*	*p,p'-DDD*	*DDA*	*DDMU*	*DBP*	*Kel-thane*	*DPE*	*DBH*	*BPE*	*DDM*	*I*
I	25	35	41	36	10	0	32	2	0	27	0	0	35	2.28
II	48	55	61	56	28	0	55	8	2	44	0	0	54	2.47
III	69	72	75	72	52	0	67	24	3	63	0	0	67	2.35
IV	76	76	77	75	64	0	74	45	8	72	2	4	75	2.50
V	67	69	72	69	52	0	70	69	7	66	1	2	71	1.98
VI	67	70	75	68	51	0	72	45	10	66	2	4	70	2.87
VII	70	70	75	70	63	0	74	61	23	66	6	13	69	2.78
VIII	78	79	83	79	77	6	83	75	53	77	27	39	80	2.56
IX	69	71	76	69	60	0	73	56	19	67	4	9	71	3.03
X	35	44	49	45	16	0	48	4	0	35	0	0	42	2.62
XI	58	63	65	62	42	0	62	18	3	55	0	0	61	2.41
XII	60	64	71	64	43	0	69	50	8	60	2	5	67	2.71
XIII	63	66	71	64	48	0	68	55	16	62	4	9	67	2.97
XIV	73	74	77	72	65	0	75	68	48	77	24	36	76	2.57
XV	83	84	85	83	78	5	83	81	58	82	46	55	83	1.47
XVI	84	84	84	83	80	0	84	82	75	84	71	74	83	1.70
XVII	87	88	89	87	81	20	87	83	64	86	59	63	86	1.47
XVIII	59	68	73	69	39	8	67	34	34	60	33	34	67	2.31
XIX	92	94	96	93	83	32	92	82	72	91	61	70	93	2.35
XX	78	80	80	77	69	17	78	67	43	74	24	35	78	2.50
XXI	82	82	82	80	81	0	81	81	81	81	80	81	82	0.39
XXII	80	80	80	77	79	5	78	78	79	79	79	79	80	0.39
XXIII	35	46	50	45	13	0	40	2	0	36	0	0	39	2.35
XXIV	40	51	52	48	16	0	44	3	0	42	0	0	42	2.28
XXV	71	74	77	72	59	0	74	64	21	70	5	13	72	2.72
XXVI	77	77	78	77	72	30	78	72	56	78	34	47	78	2.03
XXVII	27	40	43	42	10	0	35	2	0	32	0	0	35	2.08
XXVIII	70	75	77	75	51	2	74	20	4	69	4	4	75	2.04
XXIX	65	69	76	68	48	0	72	57	7	64	2	5	70	2.93
XXX	85	85	85	84	84	4	80	83	84	84	83	84	84	0.77
XXXI	54	67	74	69	22	0	61	6	6	54	6	6	62	2.62
XXXII	100	100	100	100	94	35	100	94	92	100	83	92	100	1.57
XXXIII	93	94	96	94	90	7	96	93	75	94	67	70	92	1.88

complex procedure, it is logical to try to combine partial procedures that have the greatest possible separate information contents. However, it is not possible to calculate the final information content of a complex procedure from the I values of the partial procedures; only the maximum value of the total information content, which is equal to the sum of the information contents of the partial procedures, and the minimum value, which is equal to the highest value found for a partial procedure, can be calculated.

One is not limited to the combination of two thin-layer separations; the combination of two different techniques is also possible. In qualitative analysis, this combination is often done by combining the R_F value and the reaction of the spot with some more or less specific reagent. The colour obtained contains information. REIO[8] states that it is possible to distinguish 72 colours on thin-layer chromatographic plates, and if this is correct then the information content of one-dimensional chromatography is between 10 and 11 bits and the information content of two-dimensional chromatography is between 15 and 16 bits.

APPLICATIONS AND CONCLUSIONS

Some examples can be given to show that information theory is useful for comparing different separation possibilities. The first example is the separation of DDT and 12 related compounds on Al_2O_3 plates, as described by BISHARA *et al.*[9]. The hR_F values and the information content of the separations obtained with 33 solvent systems are given in Table II. The original table contained also *m,p'*-DDD and *o,p'*-DDD, but as these compounds cannot be separated from each other or from *p,p'*-DDD, they were not taken into account in the present work. From a rapid inspection of Table II, without using the information content column, it is not possible to select the best separation and even closer inspection enables only four or five solvents to be eliminated because they are clearly not as good as the 28 others. To classify these 28 other separations according to their separation value is not possible, however. It should be noted that BISHARA *et al.* did not propose a best solvent.

When the information content column is taken into account, it can be seen that solvent systems V, XV, XVI, XVII, XXI, XXII, XXX, XXXII and XXXIII are of no interest. The best separations are obtained with three solvents, namely IX, XIII and XXIX. It can also be concluded that in further investigations aimed at optimising the separation of these substances on the same stationary phase, one should start by examining small changes in the three best solvents. Solvent XIII is *n*-hexane–acetone–acetic acid (95:5:1). In an optimisation, one should start, for example, by trying combinations such as *n*-hexane–acetone–acetic acid (90:10:1), *n*-pentane–acetone–acetic acid (95:5:1), etc.

Another example of optimisation is provided by the study of a separation of 11 carotenoids. The values were obtained from a paper by EGGERT AND VOIGT[10]. Solvents I to VII are mixtures of increasing polarity obtained by combining a 50:50 methanol–methyl ethyl ketone mixture with light petroleum or water in various proportions. A continuous rise in the information content occurs until a maximum is reached for solvent IV, followed by a continuous decrease. The best separation is therefore obtained with solvent IV (see Table III).

Now, suppose that one had started this investigation with solvents I, III, V and

TABLE III

hR_F VALUES OF ELEVEN CAROTENOIDS AND INFORMATION CONTENT OF THE PROPOSED SEPARATIONS

R_F values were taken from ref. 10.

	Solvent						
Substance	*I*	*II*	*III*	*IV*	*V*	*VI*	*VII*
Cryptoxanthine	62	70	76	74	39	21	4
Rubixanthine	45	60	64	45	15	4	0
Lycoxanthine	29	37	40	32	8	0	0
Isozeaxanthine	34	56	92	91	57	36	10
Escholtzxanthine	12	22	25	22	8	1	0
Lycophyll	8	20	22	20	7	0	0
Euglenanone	62	68	81	80	54	34	9
Canthaxanthine	58	65	79	80	55	37	20
Rhodoxanthine	28	42	43	40	14	7	1
8′-*apo*-β-carotenic acid	28	38	30	15	5	0	0
Torularhodin	6	10	9	2	1	0	0
I	2.66	2.91	2.91	3.09	2.11	1.79	1.49

VII, the best of these being solvent III. The measure of the information content would have led to the testing of solvents with polarities similar to those of solvents II and IV and the eventual discovery of the even better solvent IV.

It can be concluded that the information content enables a numerical value representative of the quality of a separation to be obtained, and therefore provides a means for reaching logical decisions concerning the choice of a solvent or for developing an optimisation strategy. Without recommending the exact procedure used here, we suggest that the information content should be tabulated in publications that compare solvents for a particular thin-layer chromatographic separation.

As stated in the introduction, a generalisation is possible, and the method can be applied to other techniques such as gas or liquid chromatography. Moreover, other applications of information theory can be proposed. In quantitative chromatography, for example, it is possible to use the informing power, P_{inf}, as defined by KAISER[1]:

$$P_{\text{inf}} = n \cdot {}^{2}\log S$$

where n is the number of substances measured on a scale of S steps. Preliminary work with this equation (in which n is obtained from the peak capacity and S from overlap calculations) will be reported in a later communication.

ACKNOWLEDGEMENT

The author thanks the "Fonds voor Kollektief en Fundamenteel Onderzoek" for financial assistance.

REFERENCES

1 H. KAISER, *Anal. Chem.*, 42 (1970) 24A.
2 G. GOTTSCHALK, *Z. Anal. Chem.*, 258 (1972) 1.
3 A. DIJKSTRA, personal communication.

4 D. L. MASSART, C. JANSSENS, L. KAUFMAN AND R. SMITS, *Anal. Chem.*, 44 (1972) 2390.
5 D. L. MASSART, C. JANSSENS, L. KAUFMAN AND R. SMITS, *Euroanalysis I Conference, Heidelberg, 1 September, 1972.*
6 H. N. NAHIKIAN, *A Modern Algebra for Biologists*, 2nd Ed., University of Chicago Press, Chicago and London, 1969.
7 I. SIMON AND M. LEDERER, *J. Chromatogr.*, 63 (1971) 448.
8 L. REIO, *J. Chromatogr.*, 68 (1972) 183.
9 R. H. BISHARA, G. B. BORN AND J. E. CHRISTIAN, *J. Chromatogr.*, 64 (1972) 135.
10 K. EGGERT AND H. VOIGT, *Z. Pflanzenphysiol.*, 53 (1965) 64.