CHROM. 20 987

AN INFORMATION THEORY OF CHROMATOGRAPHY

II. APPLICATION OF *FUMI* TO THE OPTIMIZATION OF OVERLAPPED CHROMATOGRAMS

RIEKO MATSUDA*, YUZURU HAYASHI, MUMIO ISHIBASHI and YASUSHI TAKEDA

Division of Drugs, National Institute of Hygienic Sciences, 18-1, Kamiyoga 1-Chome, Setagaya-ku, Tokyo 158 (Japan)

SUMMARY

The *function* of *mutual* *information* (*FUMI*) was used as a quality criterion in the optimization of the injection interval in overlapped chromatograms. *FUMI*, which represents the amount of Shannon's mutual information involved in the chromatograms, was calculated for overlapped chromatograms with various injection intervals. The most efficient peak separation was selected with respect to the amount of the mutual information and the observation time. Overlapped chromatograms containing negative peaks were also optimized successfully.

INTRODUCTION

The final goal of analytical techniques is to obtain an exact knowledge of the amounts of substances of interest in a sample. Liquid chromatography is not an exception, and many chromatographers have endeavoured to achieve this goal by optimizing the chromatographic conditions. However, how can we evaluate the chromatographic separation? In reply to this question, researchers have proposed various quality criteria for chromatographic separation which can quantify the quality of the chromatograms[1]. Most of these criteria include the peak resolution (R_s) or the peak separation for each pair of adjacent peaks. In commonly used criteria, the summation of R_s for all adjacent pairs of peaks was calculated and the chromatograms were evaluated by means of this value. The efficiency of analysis was also regarded as an important factor, so some workers included the measurement time in addition to R_s in their criteria. These criteria, however, lack a theoretical base.

A versatile criterion should cover two important factors that are missing in R_s: (A) the noise level in the measurement process and (B) the mathematical formalism of data processing. It should be noted that a knowledge of the samples of interest cannot be obtained before the observation and mathematical processing of the raw data. The reliability of quantification is well known to be significantly dependent on the contamination noise levels. Point B concerns the fact that the peak-resolving power of the Kalman filter, previously proposed[2,3], was shown to be superior to the commonly

used perpendicular dropping, and to give more reliable estimates. The criteria based on these methods of different powers cannot be the same.

A function called *FUMI* (*function of mutual information*) has been derived on the basis of information theory and the Kalman filter theory[4]. *FUMI* represents the mutual information in chromatograms. The efficiency is represented as the ratio of *FUMI* to the observation time. In Part I[5], *FUMI* was calculated with two overlapped peaks, and the relationship between the amount of mutual information and the degree of peak overlap was discussed.

Recently we have presented overlapped chromatograms that resulted from the successive injection of samples at relatively short intervals[6–8]. The overlapped chromatograms efficiently reduced the total analysis time and improved the efficiency of the analysis.

The object of this paper is to assess the feasibility of this function as a quality criterion for multi-peak chromatograms. We demonstrate the optimization procedure for injection intervals in the overlapped chromatograms with the use of *FUMI*. The procedure presents the peak separation with sufficient precision and favourable efficiency. The advantage of overlapped chromatograms can be expanded with the aid of *FUMI*. Although we use overlapped chromatograms as a model in this study, *FUMI* is a general criterion and is applicable to the usual chromatographic separation modes.

THEORETICAL

We shall briefly review the theory developed in Part I[5]. For a single peak, *FUMI* represents the mutual information that we can retrieve through the filtering of raw data ranging from a data point $i = 1$ to k ($k = 1, ..., N$)[4]:

$$FUMI = \frac{1}{2}\left[\log\left(\sum_{i=1}^{k} F_i^2\right) - \log(\tilde{W}_c)\right] \tag{1}$$

where F_i denotes the signal intensity of a peak at a data point i and $\tilde{W}_c$ is the variance of the contaminating noise ($=$ constant).

The mutual information for partially overlapped multiple peaks is[5]

$$FUMI = \frac{1}{2}\left\{\sum_{j=1}^{q} \log\left[\sum_{i=k_c(j)+1}^{k_f(j)} F_i(j)^2\right]\right\} - \frac{1}{2}q\log(\tilde{W}_c) \tag{2}$$

where q denotes the number of peaks and $[k_c(j) + 1, k_f(j)]$ the region where the signals $F_i(j)$ of the jth peak contribute to *FUMI*. The cutoff point $k_c(j)$ of the jth peak is often specified to be the first point where the peak signals predominate in noisy chromatograms. On the other hand, the filtering-off point $k_f(j)$ is variable and determined to be equal to the cutoff point $k_c(j + 1)$ of the following peak: $k_f(j) = k_c(j + 1)$. The information derived from the late region of the jth peak after $k = k_f(j)$, overlapped with the early region of the following peak, is neglected, as the filtering is always performed sequentially in a one-dimensional way from $i = 1$ to N (ref. 2). Thus, the strong peak overlap makes the region $[k_c(j) + 1, k_f(j)]$ narrow and then causes the loss of mutual information. When $k_c(j)$ coincides with $k_c(j + 1)$, no mutual information is picked up from either peak.

The efficiency of the chromatograms is given as

$$I_E[1,N] = \frac{FUMI}{N} \tag{3}$$

where N denotes the observation period of the chromatogram. This function represents the average amount of information in unit time and its maximum gives a chromatogram with the most efficient peak separation.

The information loss of a chromatogram, δI, arising from the peak overlap is defined as

$$\delta I = I_{max} - FUMI \tag{4}$$

where I_{max} denotes the maximal information obtained from a chromatogram with every peak sufficiently separated from each other. The filtering error is minimized at the maximum of $FUMI$ ($\delta I = 0$).

In the worst case, the information loss occurs only in a single peak, *i.e.*, the peak is overlapped strongly with the second peak, which is separated from the third peak. The error in this instance is easily calculated by the difference from the maximal information of the multi-component chromatogram; the relative standard deviation (R.S.D.) of the worst error is given approximately by[5]

$$\langle R.S.D. \rangle = \exp(\delta I) \cdot \exp(- \langle I_{max} \rangle) \cdot 100 \tag{5}$$

where $\langle I_{max} \rangle$ denotes the mean maximal information of q peaks ($= I_{max}/q$). This equation is useful for estimating the filtering error (R.S.D.) of a chromatogram from the information loss δI.

An additional procedure has been introduced for the practical use of the function of the information efficiency, I_E. The peak signals F_i are replaced by a small shape $F_i'(F_i = F_i \cdot X_s$, where X_s is the suppression factor); the calculated I_E with F_i' presupposes the appearance of small peaks.

EXPERIMENTAL

High-performance liquid chromatographic (HPLC) data

Chromatographic measurements were made on an Inertsil ODS column (250 × 4 mm I.D.) with methanol as the eluent at a flow-rate of 0.5 ml/min. HPLC signals were converted and stored on 5-in. floppy disks. Additional experimental details can be obtained from refs. 6 and 7.

Computer simulation

All calculations were performed on a PC-9801 VM desk-top computer (NEC). Programs were written in N88BASIC. Graphics were performed using a Model MP3100 X–Y plotter (Graphtec).

Raw HPLC signals for a single chromatogram were overlapped with various injection intervals and the corresponding value of *FUMI* was calculated. The beginning point of the ith peak, $k_c(i)$, was set at the point where the signal level reaches 0.5% of the peak maximum.

RESULTS

A methanolic solution containing four components, phenetol, diphenyl, pyrene and perylene, was injected into the HPLC system and gave the chromatogram shown in Fig. 1. When this solution is injected into the HPLC system successively at short, regular intervals, an overlapped chromatogram results. *FUMI* was calculated for the overlapped chromatograms involving 20 peaks derived from five-fold injections at various injection intervals. Fig. 2 illustrates the change of *FUMI* with injection intervals from 5 to 800 s. The maximal interval of 800 s, is equal to the chromatographic duration for one sample, and then the injection mode is the same as in the usual repeated experiments.

With short injection intervals *FUMI* shows some local maxima and minima corresponding to the complex change of the peak overlapping pattern. At an interval of about 315 s, *FUMI* is maximal and no longer increases with increasing intervals. No peak overlap occurs in the chromatograms with injection intervals longer than 315 s. Even with an interval shorter than 315 s, however. *FUMI* has an almost maximal value. To select the most efficient interval, we restricted ourselves to the satisfactory region of the injection interval where the information loss is less than unity; in the satisfactory region, the R.S.D. of the filtering error is expected to be less than 2.7 times the value of the minimum (eqn. 5). and the precision is guaranteed sufficiently. The optimal interval was searched within this region with the use of the information efficiency.

The information efficiency, I_E, was calculated with the suppression factor, X_s. When X_s is set at 100, for example, I_E is calculated with signals of 100-fold smaller intensity than the model peaks. This procedure means that I_E takes account of the case where peaks of 1% concentration of the model appear in the overlapped chromatogram. The effect of X_s was described in Part I[5]. The change in I_E is shown in Fig. 3 with suppression factors of 10, 100 and 1000. I_E decreases hyperbolically in the long interval region where *FUMI* takes a constant value. The maximum value of I_E in the
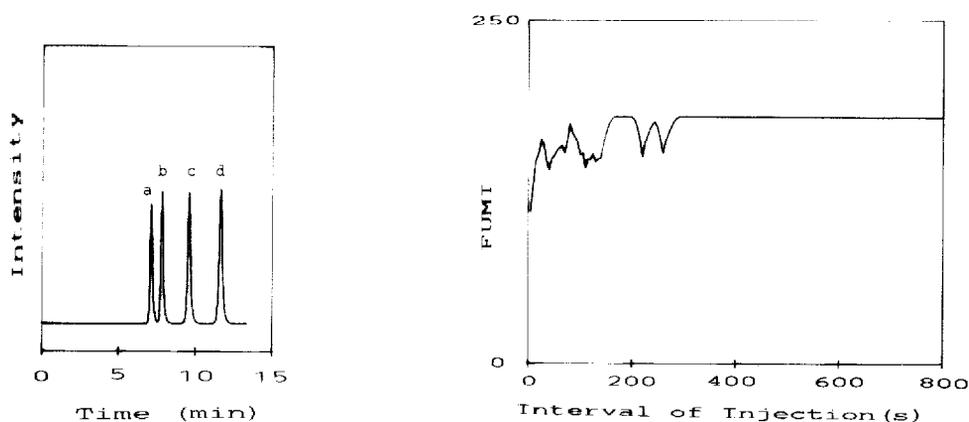

Fig. 1. Chromatogram of (a) phenetole, (b) diphenyl, (c) pyrene and (d) perylene.

Fig. 2. Variation of *FUMI* of the overlapped chromatogram derived from the single chromatogram shown in Fig. 1.
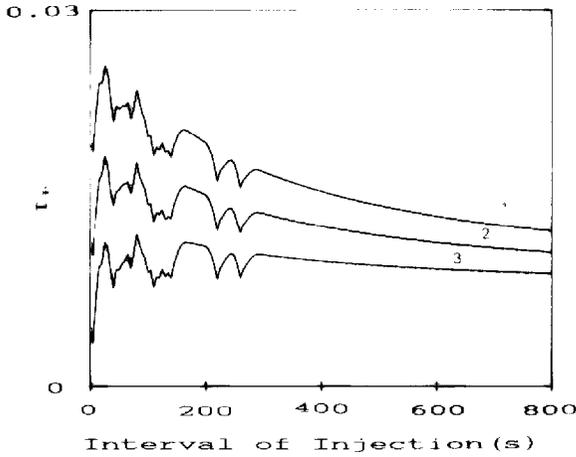
Fig. 3. Variation of I_E with suppression factors X, of (1) 10. (2) 100 and (3) 1000.

satisfactory region defined above is obtained at an injection interval of 165 s with any X, value, while the neighbouring local maximum of $FUMI$ is at 185 s.

Fig. 4A shows the optimal overlapped chromatograms with injection intervals of 165 s. The value of $FUMI$ is 179.084 and the maximum of $FUMI$ is 179.46, and δI is
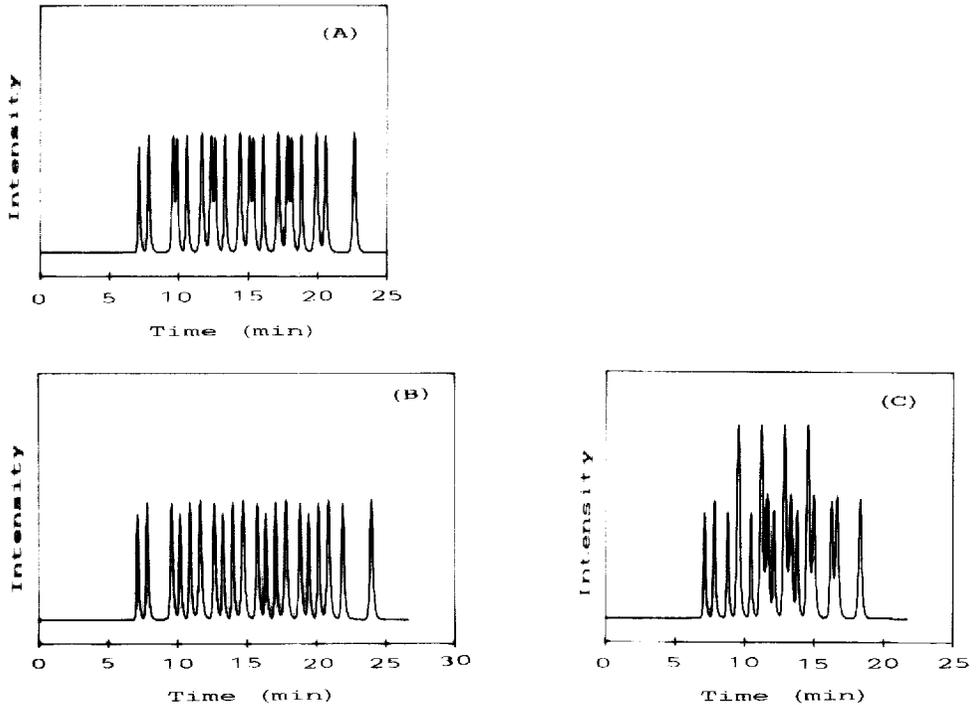

Fig. 4. Overlapped chromatograms derived from the single chromatogram shown in Fig. 1. Injection interval. (A) 165 s; (B) 180 s; (C) 80 s.

0.382. The predicted R.S.D. (0.018%) of the filtering error for the optimal chromatogram is considered satisfactory and almost the same as the minimum R.S.D. (0.012%). The overlapped chromatograms shown in Fig. 4B and C were obtained with injection intervals of 185 and 80 s, respectively. *FUMI* for the former is 179.464 and almost equivalent to the maximum, but I_E is slightly smaller than the optimal chromatogram (Fig. 4A). The shortest interval, 80 s, gives a local maximum of *FUMI*, but is outside the satisfactory region.

The suppression factor, X_s, does not influence I_E in the satisfactory region, but the effect of X_s appears in the region of short intervals. For a short observation period, the effect of X_s is greater[5].

We show another example of the optimization of overlapped chromatograms using *FUMI*. The chromatogram shown in Fig. 5 contains three component peaks that are also present in the preceding example and a small solvent peak. This chromatogram was overlapped and *FUMI* was calculated according to the procedure described above. The calculated value of *FUMI* and I_E are plotted in Fig. 6. The results reveal
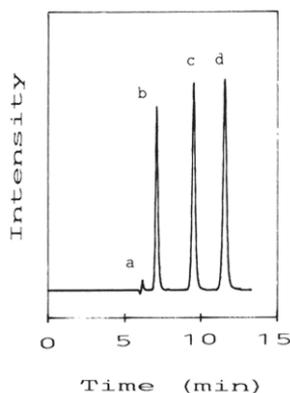

Fig. 5 Chromatogram of (a) solvent peak (10% ethanol), (b) phenetole, (c) pyrene and (d) perylene.
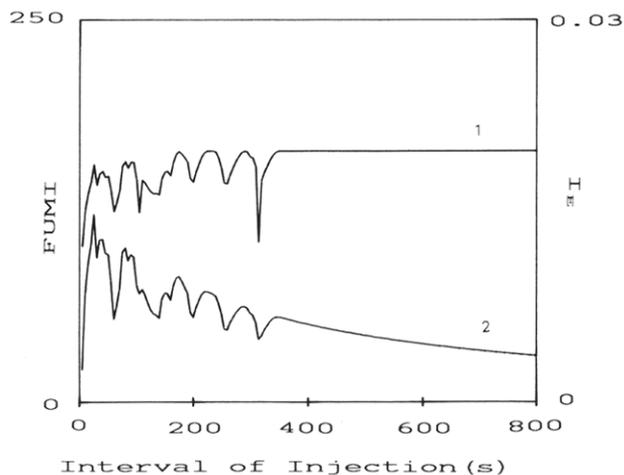

Fig. 6 Variation of (1) *FUMI* and (2) I_E (X_s = 100) of the overlapped chromatogram.
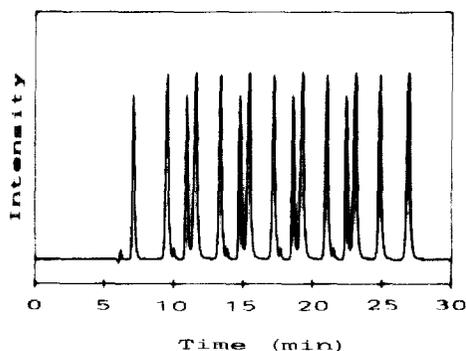
Fig. 7. Optimal overlapped chromatogram derived from the single chromatogram shown in Fig. 5.

that the small peaks and even the negative peaks are also adaptable to *FUMI*. The optimal interval was 230 s and the corresponding overlapped chromatogram is shown in Fig. 7. *FUMI* for the optimal chromatogram was 164.385, and the information loss and the R.S.D. of the error were 0.025 and 0.027%. respectively.

DISCUSSION

We have defined the quality of chromatograms with use of the mutual information or the precision of the quantification of target components. In other studies, R_s seemed to be related closely to the analytical precision and was used as the quality criterion for multi-peak chromatograms. It is obvious that the quantification error is minimized when peaks are separated completely ($R_s > 2$), and increases with decrease in R_s, but no definite relationship between the error and R_s has been given.

FUMI represents the mutual information in chromatograms and can also be connected to the R.S.D. of the filtering error, and we can estimate the precision of the filtering for various overlapped chromatograms with the aid of *FUMI*. The optimal chromatogram presented above is considered to be reasonable because the information loss or the excess filtering error is negligibly small. Practical consideration of the optimal chromatogram is given below.

We are interested in the correspondence between the filtering error and the total precision of actual chromatographic analyses. The filtering error predicted from *FUMI* is shown to describe well, although qualitatively, the change in the actual HPLC error as follows. Overlapped chromatograms with the same sample solution as that shown in Fig. 1 were analysed by the reduced Kalman filter of four dimensions[6]. The injection intervals were 200, 180, 160 and 150 s and the observed repeatability of the systems was 0.42%, 0.28%, 1.09% and 6.32%. respectively. With the shortest interval (150 s), some peaks are exceedingly overlapped and the estimated concentration was biased by *ca.* 12% at most. The calculated R.S.D.s of the filtering error with each interval are 0.017%, 0.013%, 0.17% and 2440%. respectively. The first three values for the filtering error are far smaller than the actual errors, including the errors in elution, detection and filtering. The filtering error (R.S.D = 2440%) with an injection interval of 150 s is overestimated because of the mathematical property of the error variance P_k^+ involved in *FUMI*[4]. The overestimating property is favourable, because we can

avoid choosing the excessively overlapped chromatogram as optimal. The "visual" inspection of the optimal overlapped chromatograms described in ref. 6 is not unreliable, but *FUMI* is more useful.

Overlapping of the chromatograms with the successive injection of samples at short intervals is an effective method of reducing the analysis time and increasing the efficiency without the need for any skilful techniques. We applied this method to an automated system for the content uniformity test on pharmaceutical formulations and greatly increased the total efficiency or throughput[3,8]. The only critical parameter in this method is the injection interval, on which the precision and the efficiency of analysis depend. The injection interval, however, had been selected empirically in spite of its importance. The introduction of the mutual information is a clear solution to this problem. The R.S.D. of the filtering error can be predicted with a fairly simple function, *FUMI*, and optimal conditions can be selected with regard to both efficiency and precision. The overlapped chromatography method is shown here to be better than previously considered. The optimization method using *FUMI* must increase the reliability and applicability of overlapped chromatograms.

REFERENCES

1 H. J. G. Debets, B. L. Bajema and D. A. Doornbos, *Anal. Chim. Acta*, 151 (1983) 131, and references cited therein.
2 Y. Hayashi, S. Yoshioka and Y. Takeda, *Anal. Chim. Acta*, 212 (1988) 81.
3 Y. Hayashi, R. Matsuda, S. Yoshioka and Y. Takeda, *Anal. Chim. Acta*, 209 (1988) 45.
4 Y. Hayashi, S. Yoshioka and Y. Takeda, submitted for publication.
5 R. Matsuda, Y. Hayashi, M. Ishibashi and Y. Takeda, *J. Chromatogr.*, 462 (1989) 13.
6 Y. Hayashi, T. Shibazaki, R. Matsuda and M. Uchiyama, *Anal. Chim. Acta*, 202 (1987) 187.
7 Y. Hayashi, T. Shibazaki and M. Uchiyama, *J. Chromatogr.*, 411 (1987) 95.
8 R. Matsuda, Y. Hayashi, M. Ishibashi and Y. Takeda, *J. Chromatogr.*, 438 (1988) 319.