Journal of Chromatography, 90 (1974) 1–7

CHROM. 7183

OPTIMUM USE OF PAPER, THIN-LAYER AND GAS–LIQUID CHROMATOGRAPHY FOR THE IDENTIFICATION OF BASIC DRUGS

I. DETERMINATION OF EFFECTIVENESS FOR A SERIES OF CHROMATOGRAPHIC SYSTEMS

ANTHONY CHRISTOPHER MOFFAT, KENNETH WALLACE SMALLDON and COLIN BROWN

Home Office Central Research Establishment, Aldermaston, Berks. RG7 4PN (Great Britain)

(Received August 27th, 1973)

SUMMARY

A method is described for the determination of effectiveness for a series of chromatographic systems which uses the concept of discriminating power. The discriminating power for a series of chromatographic systems is defined as the probability that two drugs selected at random from a large population would be discriminated in at least one of the systems. The series of systems with the highest discriminating power is shown to produce the best overall separations for a large specified drug population and therefore leads to the identification of an unknown drug using a minimum number of systems. The conditions for maximum discriminating power are described and discussed.

INTRODUCTION

When analysing a drug sample, be it a pure compound, pharmaceutical or extract from a biological material, an analyst is faced with a bewildering selection of paper (PC), thin-layer (TLC) and gas–liquid chromatographic (GLC) systems from which to make his choice. It may be that the particular problem requires the separation of two or more known drugs, in which case the compilations of analytical data would be consulted[1–3] and an appropriate system or systems chosen. However, if the analyst is routinely engaged in the identification of basic drugs, the data compilations do not help him to choose the best system or combination of systems, since the data were generated independently by different workers and on different drug populations. As a result of the growing multiplicity of possibilities, each analyst tends to select the systems with which he is most familiar and for which he has some sort of data collection. As a result, there is a tendency for each laboratory to have a library of data for its own particular systems, often without being aware of the collections in other laboratories. It is therefore apparent that an urgent need exists to assess objectively the effectiveness of the systems which are currently in common use so that a comprehensive data collection is produced

for the most useful systems and thus a minimum amount of effort is involved in the routine identification of basic drugs. The effectiveness of new systems can be evaluated in a similar manner as they become available and then compared with existing systems.

In order to provide the maximum amount of information, a system used in a procedure for the identification of a large number of compounds should have a random distribution of values over the whole chromatogram. A further requirement is that the system should provide results which are as reproducible as possible. Ideally the reproducibility of each system should be determined by experiment. However, if the number of systems available is large, this approach may be too time consuming for practical use and an estimate of the reproducibility may have to be made.

When systems are used in combination, the order of elution of the compounds must vary from system to system or no more information may be obtained. Previous work by one of us[4] demonstrated that the correlation coefficients between a number of commonly used TLC systems cannot be regarded as negligible.

The retention values of compounds in single systems can be displayed as frequency histograms and those showing grossly non-random distributions can readily be observed. However, when system reproducibility and inter-system correlation are also considered clearly a mathematical model is required so that the optimum procedure can be selected for the identification of an unknown drug.

THEORY

Two compounds are regarded as having been separated or discriminated in a particular chromatographic system (i) if the difference between their retention values exceeds a certain critical value, which is termed the error factor (E_i). The error factor of the system can be determined practically so that virtually all the experimentally determined retention values of a particular drug would fall within the range $\pm E_i$ about the standard value.

The discriminating power of a single chromatographic system is therefore defined as the probability that two compounds selected at random would be discriminated in that system. Similarly the discriminating power of a series of systems is defined as the probability that two compounds selected at random would be separated in at least one of the systems. Compounds which are not discriminated are assumed to be chromatographically similar.

In a previous paper[5] a method was described and discussed for the calculation of discriminating power for any series of correlated attributes provided that the distribution of each attribute over the population is known and an estimate for each error factor can be made from experience. In this application the distributions are known to be non-Gaussian and therefore the appropriate theory must be used.

Suppose that the discriminating power is required for k systems in combination and that chromatographic values for N compounds have been recorded in each system. The total number of possible pairs is given by

$$NC_2 = \frac{N(N-1)}{2}$$

If the number of pairs which are similar in all k systems is M, then the probability of selecting a similar pair at random is $2M/[N(N-1)]$. The discriminating power (DP_k) for the k systems is thus given by

$$DP_k = 1 - \frac{2M}{N(N-1)}$$

The number of similar pairs, and then the discriminating power, is calculated by means of the computer search program given as an Appendix. The procedure used to calculate the discriminating power for k systems in series is termed a "kth order" computer search. The software in the Appendix allows searches up to 6th order to be made from any eight chromatographic systems. An example of the computer routine which calculates the 2nd order discriminating power for systems 3 and 7 is shown below.

INSERT NO. OF DATA ITEMS AND SYSTEMS
100, 8
INSERT ORDER OF SEARCH AND SYSTEMS REQUIRED
2, 3, 7
INSERT ERROR FACTORS TO BE USED IN THE SAME ORDER
10, 10
DISCRIMINATING POWER=0.929

The ideal system would have a random or rectangular distribution of drugs over the whole chromatogram. A rectangular distribution function, $f(x)$, is of the form

$f(x) = 1$ when $0 \leqslant x \leqslant 1$

$f(x) = 0$ when $x < 0$ or $x > 1$

Suppose that two compounds are selected at random from this distribution and x for the first compound lies anywhere between E to $1-E$ (where E is expressed as a fraction of the range of possible chromatographic values). The probability that the two compounds are chromatographically similar is

$$\int_E^{1-E} 2E\,dx = 2E(1-2E)$$

If x for the first compound lies anywhere in the range 0 to E then the probability that the two compounds are similar is

$$\int_0^E (x+E)\,dx = \frac{3E^2}{2}$$

Similarly if x for the first compound lies in the range $(1-E)$ to 1 the probability is also $3E^2/2$.

Therefore, the overall probability that the two compounds are chromatographically similar is

$$2E(1-2E)+3E^2 = 2E-E^2$$

The maximum discriminating power that can be obtained with a chromatographic system is therefore given by

$$DP = 1 - 2E + E^2$$

This expression is true for all values of E and when E is small the E^2 term may be neglected.

As an aid to interpretation, correlation coefficients (r) may be determined for pairs of systems using the expression

$$r = \frac{\frac{1}{N}\Sigma(x-\bar{x})(y-\bar{y})}{\sigma_x \sigma_y}$$

where x and y are the chromatographic values in the two systems, $\bar{x}$ and $\bar{y}$ being their respective means, and σ_x and σ_y their respective standard deviations.

If k chromatographic systems are all uncorrelated then the combined discriminating power (DP_k) is given by

$$DP_k = 1 - \prod_{i=1}^{k} (1 - DP_i)$$

If, in addition, the distributions are all rectangular, then

$$DP_k = 1 - \prod_{i=1}^{k} (2E_i - E_i^2)$$

This represents the maximum possible discriminating power for any given error factors (E_i's) so that as all the E_i's$\rightarrow 0$, $DP \rightarrow 1$.

DISCUSSION

The measurement of the discriminating power for a series of chromatographic systems allows the effectiveness of the series to be expressed as a single value, the series with the highest discriminating power being the most effective. If only a small number of compounds have to be chromatographed, it may be that complete separation is possible ($DP = 1$), in which case further evaluation is unnecessary. However, if the number of compounds to be chromatographed is large, complete separation of all the possible pairs is no longer feasible and the aim becomes to maximise the discriminating power. In this case the conditions for maximum discriminating power are: (a) rectangular distributions of compounds over the chromatograms, (b) good system reproducibility, and (c) no correlation between systems.

The compounds for which chromatographic values are recorded can be determined from the nature of the particular problem and the work records of the laboratory. A statistically adequate number of samples which genuinely reflects the problem under consideration must be recorded before reliable values for discriminating power are obtained.

The number of systems available may preclude the experimental determination of all the error factors. In order to reduce the number of systems to manageable proportions the error factors could initially be estimated from experience and then the true inter-laboratory reproducibility could be determined for the small number of systems selected by this procedure.

Correlation coefficients can be calculated for non-Gaussian distributions, such as the distribution of drugs across a chromatogram, but their significance cannot be tested in the usual manner. However, values for correlation coefficients remain a useful aid to interpretation for a number of reasons. For example, if other factors are equal, the pair of systems showing the lowest correlation coefficient will provide the highest discriminating power for two systems in combination. When a large number of systems are considered, the amount of computation can often be reduced in the following ways: Those systems showing very low discriminating powers can be eliminated from further consideration, the discriminating power for series of uncorrelated systems can be calculated from the individual discriminating powers using the formula already described[5], and if two systems are highly correlated the system with the lower discriminating power can be eliminated.

If systems are selected so that the discriminating power is maximised, then the probability of discriminating a pair of compounds selected at random is also maximised. The same argument applies if any number of compounds are selected at random since N compounds can be represented by NC_2 pairs for which discrimination is required.

The discriminating power for a series of systems (DP_k), which has been calculated from the chromatographic values of N compounds, can be used to predict the average number of possibilities which would be retrieved during a search to identify an unknown compound. Since systems with high discriminating power have nearly rectangular distributions, the actual number of compounds retrieved does not vary much about the mean value. The method used to calculate the error factors ensures that there is only a very small chance that the true identity of the unknown compound will be excluded. Thus the true identity will be retrieved with certainty and further (N–1) comparisons will be made, each with a probability ($1-DP_k$) of providing a random match within the error factors. An estimate of the total number of compounds (T) retrieved during an average search is therefore given by

$$T=1+(N-1)(1-DP_k)$$

This expression shows that if the sequence of systems showing the maximum discriminating power (*i.e.*, $DP\rightarrow 1$) were used, then the unknown compound would be identified (*i.e.*, $T\rightarrow 1$) using the minimum number of systems. This argument naturally assumes that chromatographic values have been recorded for the unknown compound and this can only be achieved by carefully compiling a list of reference compounds from the work records of the laboratory.

The selection of systems purely on the basis of their discriminating power assumes that the effort involved in setting up each system is similar and that the time of analysis is not too important. Individual analysts would probably wish to apply their own particular constraints to results calculated in the manner described. For example, a clinical toxicologist would probably place a high priority on the speed

of analysis, whereas a pharmaceutical analyst would be less likely to do so. Under these circumstances final decisions would be made by balancing the discriminating powers against the particular constraints which apply.

The theory described here can be applied to any chromatographic problem which involves a large compound population in order to obtain the best overall separations and to identify an unknown compound with the minimum number of systems. In subsequent papers this theory has been applied to the PC, TLC, and GLC of basic drugs.

APPENDIX

Written in Fortran IV for a Hewlett-Packard Model 2100 computer.

```
ØØØ1 FTN
ØØØ2        PROGRAM DP
ØØØ3        DIMENSION IN (1ØØ,8), IERR(8), IPONT(6)
ØØØ4        WRITE(1,2)
ØØØ5 2      FORMAT("INSERT NO OF DATA ITEMS AND SYSTEMS")
ØØØ6        READ(1,*)N,ISYST
ØØØ7        DO 1Ø I=1,ISYST
ØØØ8        DO 2Ø J=1,N
ØØØ9        READ(5,*)IN(J,I)
ØØ1Ø 2Ø     CONTINUE
ØØ11        PAUSE
ØØ12 1Ø     CONTINUE
ØØ13 6      WRITE(1,3)
ØØ14 3      FORMAT("INSERT ORDER OF SEARCH AND SYSTEMS
            REQUIRED")
ØØ15        READ(1,*)IORD, IPONT
ØØ16        WRITE(1,4)
ØØ17 4      FORMAT("INSERT ERROR FACTORS TO BE USED IN THE
            SAME ORDER")
ØØ18        READ (1,*)IERR
ØØ19        A=Ø.Ø
ØØ2Ø        DO 5Ø I=1, (N-1)
ØØ21        DO 4Ø J=(I+1),N
ØØ22        DO 3Ø K=1,IORD
ØØ23        IF(IABS(IN(I,IPONT(K))-IN(J,IPONT(K))).LE.IERR(K))3Ø,4Ø
ØØ24 3Ø     CONTINUE
ØØ25        A=A+1.Ø
ØØ26 4Ø     CONTINUE
ØØ27 5Ø     CONTINUE
ØØ28        B=N
ØØ29        B=B*(B-1.Ø)/2.Ø
ØØ3Ø        DP=1.Ø-A/B
ØØ31        WRITE(1,5)DP
ØØ32 5      FORMAT("DISCRIMINATING POWER=",F5.3//)
```

```
ØØ33    GO TO 6
ØØ34    END
ØØ35    END$
```

REFERENCES

1 E. G. C. Clarke, *Isolation and Identification of Drugs*, Pharmaceutical Press, London, 1969.
2 A. S. Curry, *Poison Detection in Human Organs*, Thomas, Springfield, Ill., 2nd ed., 1969.
3 I. Sunshine, *Handbook of Analytical Toxicology*, Chemical Rubber Co., Cleveland, Ohio, 1969.
4 K. W. Smalldon, *J. Forensic Sci. Soc.*, 11 (1971) 171.
5 K. W. Smalldon, A. C. Moffat and J. B. Parker, *J. Forensic Sci. Soc.*, in press.