Journal of Chromatography, 469 (1989) 127–135
Elsevier Science Publishers B.V., Amsterdam — Printed in The Netherlands

CHROM. 21 342

NORMALISATION OF HIGH-PERFORMANCE LIQUID CHROMATOGRAPHY PEAK RETENTION TIMES FOR COMPUTERISED COMPARISON OF WHEAT PROLAMIN CHROMATOGRAMS[a]

HARRY D. SAPIRSTEIN, MARTIN G. SCANLON* and WALTER BUSHUK
Grain Industry Research Group, Food Science Department, University of Manitoba, Winnipeg, Manitoba, R3T 2N2 (Canada)
(First received November 1st, 1988; revised manuscript received January 24th, 1989)

SUMMARY

To improve the reproducibility of wheat protein separations, reversed-phase high-performance liquid chromatography peak retention times were normalised relative to an external reference chromatogram run interveningly as part of a standardised experimental procedure. The computer program functions, without operator intervention, to identify five designated calibration peaks in the chromatogram of ethanol-soluble proteins (wheat prolamins or gliadins) from the standard hard red spring wheat variety Neepawa. The retention times of these peaks are then used as anchor points in a piecewise calibration algorithm to normalise chromatograms of samples run in the interval between two of the standards. For chromatograms acquired over a two-month period, this procedure decreased the average experimental error in peak retention times five-fold to a level of precision comparable to that of short-term analyses.

INTRODUCTION

Wheat prolamins (gliadins), the proteins extractable with 70% aqueous ethanol from wheat endosperm, can be separated by reversed-phase high-performance liquid chromatography (RP-HPLC)[1]. In most cases extracts of grain of a given variety (or genotype) will produce a unique chromatogram[2]. Chromatograms can be automatically integrated and thus reduced to a set of peak retention times and associated peak heights or areas. By analysing certified genotypes in this way, a library of reference chromatographic data can be built up and, similar to electrophoretic patterns[3,4], the data can be used to identify unknown grain samples or for estimation of homology with reference genotypes.

As in electrophoresis, the reliability of the HPLC methodology depends on the precision of the data. Sample and solvent preparation and machine performance can,

[a] Publication no. 146, Food Science Department.

0021-9673/89/$03.50

and do, contribute to systematic and random errors, causing retention time variation[5]. Incomplete resolution of peaks, even after method optimisation[6], can also result in erroneous peak assignments due to integrator artefacts[7]. A major source of systematic errors, however, can arise from variability between nominally identical columns[8] and from progressive changes in column properties with time[9]. While many of these factors can be controlled by adhering to rigorously controlled experimental procedures, appropriate correction (or normalisation) of peak retention times is required to obtain the precision necessary for accurate comparison of chromatographic data[10].

Chromatogram normalisation typically requires definition of peak retention time on a column as an index relative to the retention of reference compounds[11]. However, no similar quantitative approach has been reported in cereal protein analyses. The complex heterogeneity of wheat endosperm proteins results in complex chromatograms of numerous components with a broad range of hydrophobicities. Hence, sample and standard peaks are likely to be confounded, as Bietz and Cobb[12] found when they added alkyl aryl ketones to an ethanolic extract of the wheat variety Chinese Spring.

This paper describes a computer-based procedure for normalising peak retention times of gliadin components separated by RP-HPLC, relative to the retention times of five gliadin calibration peaks of a standard wheat variety. Any changes in operating conditions and column properties which affect retention times will thus be corrected by standards that interact with the stationary phase in a manner similar to the interactions between samples and the stationary phase[13,14]. The normalisation is essential to achieve the precision required for computerised wheat variety identification and calculation of inter-genotype homologies.

EXPERIMENTAL

Materials

HPLC grade acetonitrile and ethanol were obtained from Fisher Scientific (Fair Lawn, NJ, U.S.A.). Sequanal grade trifluoroacetic acid (TFA) was purchased from Pierce (Rockford, IL, U.S.A.). Water was distilled and then purified with a Millipore Milli-Q system (Bedford, MA, U.S.A.). Grain of the Canadian bread wheat variety Neepawa, verified as authentic and pure by polyacrylamide gel electrophoresis[15], was used to prepare extracts of the standard for the calibration chromatogram.

Chromatography and sample preparation

Preparation of gliadin extracts and the relevant experimental procedures have been described previously[16]. RP-HPLC was performed with a 1090M Hewlett-Packard Liquid Chromatograph using a wide pore (300A), C_8 Supelcosil column. Solutions of water and acetonitrile (both containing 0.1% TFA)[1] were made fresh for each, or every second, set of analyses (see below). Optimization of gradient elution and data acquisition conditions ensured that peaks were reproducibly integrated for any given separation[16].

A chromatogram set was arbitrarily defined as comprising no more than eight experimental samples plus two extracts of the standard. One standard was run at the beginning and one at the end of each set. This provided a relatively low (maximum

4:1) ratio of experimental to standard analyses. For successive runs involving more than eight samples, the last standard of the previous set became the first of the next set. Column clean-up[17] was performed, on average, after nine samples were run. To obtain the results reported in this paper, the sets were run intensively for 2 months.

Protein elution profiles recorded at 210 nm were integrated on the chromatograph's HP-310 computer using Hewlett-Packard software (HP 79994A) to obtain data on peak retention times and quantitation parameters for statistical analysis.

RESULTS AND DISCUSSION

Selection of the variety Neepawa as an RP-HPLC standard

Five peaks in the chromatogram of the standard (Fig. 1) were selected as retention time calibration peaks on the basis of their positions across the chromatogram and their reliability for automated detection. The distinctive heights of these peaks, within defined regions of the chromatogram (Fig. 1), made them easily identifiable by computer (see below). For 2 months, during which there were changes in the selectivity of some of the other peaks, the sizes and shapes of the calibration peaks did not change. Furthermore, location effects and other common environmental factors during the growing season[18] do not affect peak retention times. The variety Neepawa is homogeneous in genetic composition[15], and its grain is readily available. Neepawa was adopted in 1987 as the standard variety for the Canada Western Red Spring class of wheat[19] and has been used as a reference variety in electrophoretic studies[15,20]. Accordingly, the variety Neepawa appears to be an excellent standard.

Specification and identification of standard peaks

In order to use the selected calibration peaks as standards for automated chromatogram normalisation, prior knowledge of peak retention times is required. Integrated report files provided the source data for all analyses, including identification of the five calibration peak retention times. For assessing the variation of peak reten-

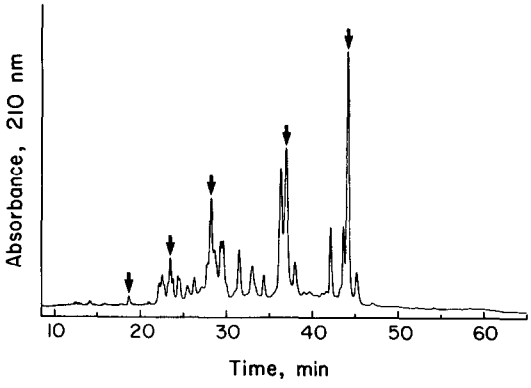

Fig. 1. Chromatogram of an extract of the variety Neepawa showing the five calibration peaks ($S_1 - S_5$, left to right, respectively). Column, C_8 Supelcosil (25 cm × 4.6 mm I.D.); elution conditions, acetonitrile–water (25:75) containing 0.1% TFA for 5 min at 1 ml/min followed by a linear gradient (0.5%/min) to acetonitrile–water (55:45) containing 0.1% TFA at 1 ml/min.

tion times before and after normalisation a set of 30 peaks was chosen in the time range 18–50 min in 25 Neepawa chromatograms (Fig. 1). These 30 peaks were rigorously identified as matching components in the 25 replicates sampled during the 2-month experimental period on the basis of direct visual inspection of the peaks and collation with peak integration results.

Computer identification of the calibration peaks was based on finding the largest peak in each of five narrow retention time windows. Calibration peaks were selected so that no other peak maxima could exist in their neighbourhood given the retention-time drifts that could normally occur. The five calibration peaks (S_1–S_5) were always correctly identified by this approach in 48 Neepawa samples analysed in the 2-month period, during which more than 250 separations were performed. The population mean retention times for the five calibration peaks S_1–S_5 were 19.61 min, 24.54 min, 29.01 min, 37.82 min and 44.78 min, respectively. These values served as anchor points (k_1–k_5) in the algorithm described below for correction of observed retention times relative to the calibration peaks. Alternatively, k_1–k_5 might be defined as the observed retention times (t_{S1}–t_{S5}) for a standard run when a new column is installed. That is, all chromatograms are normalised relative to the calibration peaks in the first chromatogram of the standard.

Application of external calibration peaks for determination of relative retention times

Normalisation of observed peak retention times was implemented by an algorithm comprising three steps (*cf.* Table I):

(1) Correction of observed sample peak retention time (t_e) to a corrected retention time (t_R), relative to the observed retention time for each of the five calibration peaks (t_{S1}–t_{S5});

(2) Determination of the "weighing function" (w_i) for the four internal chromatogram ranges (t_{S1} to t_{S2}, t_{S2} to t_{S3}, etc.) to correct calibration peak retention times (t_{S1}–t_{S5}) to the anchor points (k_1–k_5);

(3) Computation of relative retention times (t') for sample peaks according to the positions of their observed retention times relative to the calibration peaks.

The formulae which are implemented in the computer program to determine relative peak retention times are given in Table I. Equations 1–5, 6–9, and 10–15 correspond to steps 1, 2 and 3, respectively. The expression invoked (eqns. 10–15) to determine the relative peak retention time for a sample is contingent upon the position of a given protein peak within the chromatogram (Fig. 2). For example, if the observed retention time (t_e) for a sample peak is less than the value (t_{S1}) for the first calibration peak (S_1) the retention time is corrected relative to S_1 alone using eqn. 10. Similarly if a peak elutes with a retention time greater than the observed value for calibration peak S_5, eqn. 15 is invoked to make the correction. On the other hand, if a peak elutes in the interval between two calibration peaks, its relative retention time is determined relative to the retention times of the two flanking calibration peaks, but weighted according to the proximity of the peak to each of the two calibration peaks (eqns. 11–14).

The task of identifying sample and calibration peaks, along with related retention-time peak corrections, was carried out by a computer program developed in FORTRAN 77. Program implementation was on the HP-310 computer of the HPLC HP-1090M workstation. Because peak identification was based on integration report

TABLE I

RELATIONSHIPS USED TO CORRECT OBSERVED PEAK RETENTION TIMES TO RELATIVE BASIS BY MULTIPLE CALIBRATION PEAKS

Parameter	*Definition*	*Eqn. No.*
Retention time relative to calibration peak:[a]		
S_1	$t_{R1} = k_1(t_e/t_{S1})$	(1)
S_2	$t_{R2} = k_2(t_e/t_{S2})$	(2)
S_3	$t_{R3} = k_3(t_e/t_{S3})$	(3)
S_4	$t_{R4} = k_4(t_e/t_{S4})$	(4)
S_5	$t_{R5} = k_5(t_e/t_{S5})$	(5)
Weight function for chromatogram range[a]:		
t_{S1} to t_{S2}	$w_1 = (t_{S2}-k_1)/(k_2-k_1)$	(6)
t_{S2} to t_{S3}	$w_2 = (t_{S3}-k_2)/(k_3-k_2)$	(7)
t_{S3} to t_{S4}	$w_3 = (t_{S4}-k_3)/(k_4-k_3)$	(8)
t_{S4} to t_{S5}	$w_4 = (t_{S5}-k_4)/(k_5-k_4)$	(9)
Relative peak retention time (t'):		
$t_e < t_{S1}$	$t' = t_{R1}$	(10)
$t_{S1} < t_e < t_{S2}$	$t' = (1-w_1)t_{R1} + (w_1)t_{R2}$	(11)
$t_{S2} < t_e < t_{S3}$	$t' = (1-w_2)t_{R2} + (w_2)t_{R3}$	(12)
$t_{S3} < t_e < t_{S4}$	$t' = (1-w_3)t_{R3} + (w_3)t_{R4}$	(13)
$t_{S4} < t_e < t_{S5}$	$t' = (1-w_4)t_{R4} + (w_4)t_{R5}$	(14)
$t_e > t_{S5}$	$t' = t_{R5}$	(15)

[a] Refer to Fig. 2; see text for additional details.

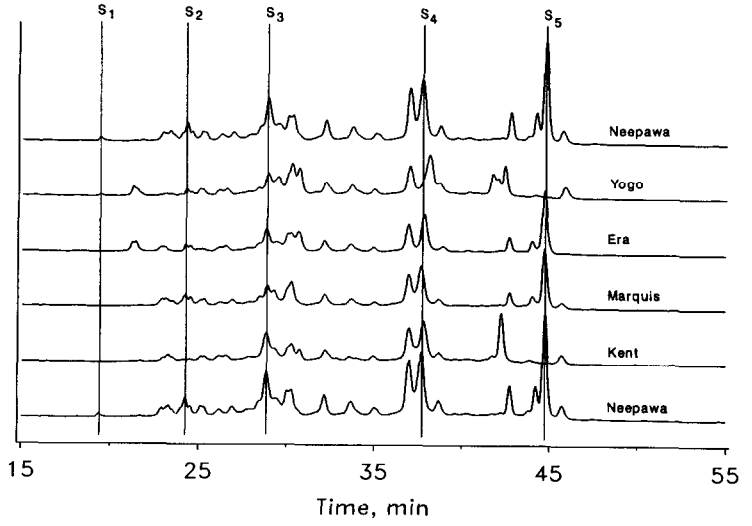

Fig. 2. Set of chromatograms from four sample wheat varieties and two standards to show peak normalisation relative to the calibration peaks. Lines drawn across the chromatograms denote the equivalent retention times of calibration peaks $S_1 - S_5$ in sample chromatograms.

files, and the process described above is not complex, program implementation using a low-cost personal computer should be satisfactory.

Retention time precision

The coefficients of variation (C.V.) and the standard deviations (S.D.) shown in Table II attest that the precision of uncorrected retention times for standardised analyses of gliadins carried out over a short period (2 weeks), was substantially better than that for the same analyses performed over a prolonged period (2 months). The mean S.D. of uncorrected retention times for chromatograms acquired over a short term was 0.06 min, compared with 0.53 min for the long-term data. Underlying this result was a significant difference in the pattern of variation of uncorrected retention times (Fig. 3). For analyses carried out over the long-term period, retention time variation was substantially greater for the early eluting proteins than for the later eluting counterparts. Possible factors contributing to this result have been discussed previously[16]. Even so, the uncorrected long-term variation in wheat protein retention times compares well with the reported capacity factors for column test compounds and barbiturates (mean C.V. of 2.67% and 3.95% respectively)[21], thiazide diuretics (mean C.V. = 1.93%)[22], and the retention times of seven drugs (mean C.V. = 1.78%)[23], analysed over short periods.

Normalised retention times are significantly more precise than the observed retention times over the long-term period (Fig. 3). On average, the uncertainty in the data was reduced more than five-fold (Table II). The magnitude of normalisation was significantly greater for early eluting proteins, *i.e.* for peaks with retention times less than 35 min, where an eight-fold improvement in experimental error was realised; the C.V. being reduced from 2.5% to 0.3%.

A characteristic feature of the normalisation procedure is the increasing influence of the magnitude of correction for peaks that elute closer to the calibration peaks. The effect is clearly illustrated in Fig. 3, as distinct S.D. minima are reached in the neighbourhood of calibration peak positions. This indicates that peaks of both experimental and standard samples which have retention times close to those of the calibration peaks are essentially free of experimental error.

Overall, the corrected long-term retention time precision (C.V. = 0.31%,

TABLE II

COMPARISON OF MEAN PRECISION PARAMETERS FOR OBSERVED AND CORRECTED RP-HPLC PEAK RETENTION TIMES

Experimental conditions as in Fig. 1.

	Retention time range (min)	*C.V. (%)*[a]	*S.D. (min)*	*L.S.D. (min)*[b]
Long-term (uncorrected)	19.6–47.4	1.82	0.53	2.12
Short-term (uncorrected)	18.1–47.5	0.20	0.06	0.24
Long-term (corrected)	19.6–47.4	0.31	0.10	0.40

[a] Number of replicate analyses > 15.
[b] Least significant difference, $p = 0.05$.

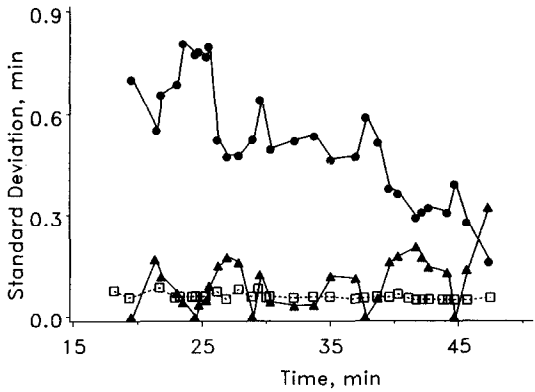

Fig. 3. Standard deviations of retention times for gliadin peaks from Neepawa extracts: (□) short-term (2 weeks) retention times; (●) long-term (2 months) retention times; and (▲) long-term retention times corrected by the normalisation procedure. Experimental conditions as in Fig. 1.

S.D. = 0.10 min) compared favourably with the reproducibility of short-term analyses (C.V. = 0.20%, S.D. = 0.06 min). This represents a satisfactory level of performance for RP-HPLC of wheat proteins. The long-term corrected retention times also compare well with the results for indexation to standard compounds such as alkyl aryl ketones[11] and 2-keto alkanes[23] obtained over short periods of analysis, where mean C.V. of 0.99%, 0.24% and 0.21% were obtained for drugs[23], barbiturates[21] and thiazide diuretics[22], respectively.

Applications

In practical terms, the normalisation procedure should result in a significant improvement in precision for comparing chromatographic data, especially data obtained with one column over a long period. Since the prime characteristic of a chromatogram is peak positioning[24], changes in retention characteristics of the column with loss of binding capacity[9] should be monitored by molecules that have functional groups of similar retention characteristics as the compounds of interest, especially for molecules with strong hydrogen-bonding capacities[25] (such as wheat proteins[26]). Using Neepawa as the source of standards has the disadvantage that results are standardised with reference compounds which another laboratory may decide is inappropriate for their purposes, and hence the potential exists for an undesirable proliferation of standards (as pointed out by Smith *et al.*[27]). However, since the mechanism of retention and elution of proteins is substantially different from that of small molecules[17], such a proliferation may be justified in order to adequately correct retention time variations in order to carry out comparative analyses[14].

Inferences on component identity are mainly based on comparisons of peak retention times. Accordingly, residual variation in retention times that occurs must be accounted for[28] by allowing a retention time threshold (or window) when comparing peaks. As in the case of electrophoretic separations of gliadins[15], this threshold must be sufficiently large to accept "truly" identical protein components, but not so large that mismatches result on account of real differences. The problem is a classical one of minimising the so-called type I (rejection of true peak identity) and type II (accept-

ance of false peak identity) errors. In electrophoretic and RP-HPLC analyses of gliadins, type II errors dominate owing to the complex heterogeneity of the protein separations. For example, the mean separation distance between adjacent peaks is small (less than 45 s). Direct comparison of chromatograms indicates that *ca.* 20 s represents the smallest detectable difference in retention time for visually different peaks. Left uncorrected, the least significant difference (L.S.D.) in peak retention times with prolonged column use is greater than 2 min (Table II). Therefore, the experimental error in long-term data will undermine comparative analyses because of false matches.

In contrast, data normalised by this procedure will be sufficiently precise that differences of *ca.* 0.4 min (24 s) in peak retention times would be detected (Table II). This is close to the smallest difference detectable visually. It is interesting to note that for data acquired over a 2-week period, the L.S.D. in peak retention time was *ca.* 15 s. To achieve this high level of statistical performance, incorporation of additional calibration peaks in the normalisation algorithm would be required. Nevertheless, the procedure as it stands should result in a substantial reduction in the number of peak mismatches that would otherwise occur, and allow significantly higher degrees of confidence in the computerised comparative analysis of chromatographic data acquired over long periods.

CONCLUSIONS

Despite rigorously standardised chromatographic conditions and efforts to prevent column deterioration with use, random and systematic variations in peak retention times do occur. For separations of complex mixtures, where complete resolution of components is not always possible, retention time normalisation is essential if chromatograms run at different times are to be compared. The described normalisation procedure, based on five calibration peaks, stabilises variability in retention time at a low level across the entire chromatogram. The average uncertainty in the retention time of gliadin peaks was reduced more than five-fold. The advantage of this procedure for gliadins, over the alkyl aryl ketone retention index scale, is that normalisation is performed relative to standards that have essentially the same retention properties and span almost the full range of gliadin retention times. The normalised chromatograms can then be directly compared by computer with a previously acquired library of similarly normalised chromatograms, for wheat variety identification and for calculation of the degree of homology between genotypes.

ACKNOWLEDGEMENTS

We are grateful for financial assistance for this project from the Natural Sciences and Engineering Research Council of Canada, and to Denise Lawless for performing some of the chromatographic analyses.

REFERENCES

1 J. A. Bietz, *J. Chromatogr.*, 255 (1983) 219.
2 B. A. Marchylo, D. W. Hatcher and J. E. Kruger, *Cereal Chem.*, 65 (1988) 28.

3 G. L. Lookhart, B. L. Jones, D. E. Walker, S. B. Hall and D. B. Cooper, *Cereal Chem.*, 60 (1983) 111.
4 H. D. Sapirstein and W. Bushuk, *Seed Sci. Technol.*, 14 (1986) 489.
5 J. N. Brown, M. Hewins, J. H. M. van der Linden and R. J. Lynch, *J. Chromatogr.*, 204 (1981) 115.
6 T. Hoshino, M. Senda, T. Hondo. M. Saito and S. Tohei, *J. Chromatogr.*, 316 (1984) 473.
7 A. C. Brown III, D. L. Wallace, G. L. Burce and S. Mathes, in T. M. Vickrey (Editor), *Liquid Chromatography Detectors*, Marcel Dekker, New York, 1983, Ch. 9, p. 355.
8 A. P. Goldberg, *Anal. Chem.*, 54 (1982) 342.
9 J. L. Glajch, J. J. Kirkland and J. Köhler, *J. Chromatogr.*, 384 (1987) 81.
10 R. M. Smith, T. G. Hurdley, R. Gill and A. C. Moffat, *Anal. Proc.*, 22 (1985) 331.
11 R. M. Smith, *J. Chromatogr.*, 236 (1982) 313.
12 J. A. Bietz and L. A. Cobb, *Cereal Chem.*, 62 (1985) 332.
13 J.-C. Chen and S. G. Weber, *J. Chromatogr.*, 248 (1982) 434.
14 C. T. Mant and R. S. Hodges, *LC, Liq. Chromatogr. HPLC Mag.*, 4 (1986) 250.
15 H. D. Sapirstein and W. Bushuk, *Cereal Chem.*, 62 (1985) 372.
16 M. G. Scanlon, H. D. Sapirstein and W. Bushuk, *Cereal Chem.*, 66 (1989) 112.
17 C. T. Wehr, *J. Chromatogr.*, 418 (1987) 27.
18 F. R. Huebner and J. A. Bietz, *Cereal Chem.*, 65 (1988) 362.
19 Canadian Grain Commission, *Official Grain Grading Guide*, Canadian Grain Commission, Winnipeg, 1987 ed., 1987.
20 P. K. W. Ng and W. Bushuk, *J. Cereal Sci.*, 9 (1989) 53.
21 R. M. Smith, T. G. Hurdley, R. Gill and A. C. Moffat, *Chromatographia*, 19 (1984) 401.
22 R. M. Smith, G. A. Murilla, T. G. Hurdley, R. Gill and A.C. Moffat, *J. Chromatogr.*, 384 (1987) 259.
23 J. K. Baker, L. A. Cates, M. D. Corbett, J. W. Huber and D. L. Lattin, *J. Liq. Chromatogr.*, 5 (1982) 829.
24 R. J. Marshall, A. J. Bleasby, R. Turner and E. H. Cooper, *Chemomet. Intell. Lab. Syst.*, 1 (1987) 285.
25 N. El Tayar, A. Tsantili-Kakoulidou, T. Roethlisberger, B. Testa and J. Gal, *J. Chromatogr.*, 439 (1988) 237.
26 J. D. Schofield and M. R. Booth, in B. J. F. Hudson (Editor), *Developments in Food Proteins—2*, Applied Science Publishers, Barking, 1983, Ch. 1, p. 1.
27 R. M. Smith, G. A. Murilla and C. M. Burr, *J. Chromatogr.*, 388 (1987) 37.
28 F. N. Konstantinides, L. Garr, J. C. Li and F. B. Cerra, *J. Chromatogr. Sci.*, 25 (1987) 158.