Journal of Chromatography, 624 (1992) 53–80
Elsevier Science Publishers B.V., Amsterdam

CHROM. 24 403

Review

Electrophoresis and chromatography of wheat proteins: available methods, and procedures for statistical evaluation of the data

J. A. Bietz

US Department of Agriculture, Agricultural Research Service, National Center for Agricultural Utilization Research, 1815 North University Street, Peoria, IL 61604 (USA)

D. G. Simpson*

Department of Statistics and Institute for Environmental Studies, University of Illinois, Champaign, IL 61820 (USA)

ABSTRACT

Analysis of gluten proteins from the wheat grain endosperm has long challenged the analytical chemist. Several hundred unique polypeptides are present, many in large polymers. This complexity, plus useful relationships of composition to genotype and quality, encouraged development and application of electrophoresis and chromatography for gluten analysis. We review the methods of polyacrylamide gel electrophoresis, sodium dodecyl sulfate–polyacrylamide gel electrophoresis, isoelectric focusing and high-performance liquid chromatography available for study of wheat proteins. Singly and in combination, they provide rapid, reproducible, high-resolution separations based on size, charge, or surface hydrophobicity. As challenging and important as the analyses themselves, however, is interpretation of data. Subjective evaluation is sometimes possible, but statistical methods such as similarity scores, clustering, principal components, multiple linear regression, and partial least squares now are increasingly used for data analysis. We review the use of these procedures, and precautions necessary to avoid misinterpretation of data. Optimal evaluation of protein analytical data will enhance the value of such analyses in wheat breeding, marketing, and processing.

CONTENTS

The mention of firm names or trade products does not imply that they are endorsed or recommended by the USDA over other firms or similar products not mentioned.

0021-9673/92/$15.00 © 1992 Elsevier Science Publishers B.V. All rights reserved

1. INTRODUCTION

Wheat (*Triticum aestivum* L.) has long been a primary food source because of its productivity, adaptation to diverse environments, and good nutritional value. Wheat's popularity is also due to the excellent baked goods made from it. We can produce such products because of wheat's storage proteins. To the plant, these proteins are the nitrogen reserve for the germinating embryo. For us, however, they are an important part of the diet. These proteins, known collectively as gluten, also have unique properties. After wheat is milled and the resulting flour is hydrated, dough develops as proteins interact with each other and other flour constituents to form a continuous cohesive matrix. This gluten network is elastic, and retains carbon dioxide generated during dough fermentation. Upon baking, its expanded structure is set, resulting in porous, fine-textured products.

Because of the importance of gluten, and since variation in structure is associated with quality, understanding gluten's structure has long been important. In 1745, Beccari [1] described milling of wheat and gluten–starch separation, and attempts to characterize gluten. He concluded that such studies, "... relevant both to sickness and to health ...." should encourage others to investigate food. In a seminal report Osborne [2] described gluten's unique properties, and noted that only wheat has a protein so easily isolated. He separated gluten into two major protein classes, gliadin, soluble in aqueous alcohol solutions, and glutenin, insoluble in alcohol but soluble in dilute acid or alkali.

Osborne [2] and those who followed him have had a difficult task. Isolation and characterization of gluten's many components is one of the most difficult challenges faced by protein chemists. This is partly because gluten is so heterogeneous: since gluten consists of storage proteins, there are few constraints on its expression. Also, genes which code gluten were duplicated during evolution. Finally, bread wheat is hexaploid, having three closely related genomes. Thus, many polypeptides form gluten.

Gluten characterization is also difficult because of its atypical characteristics. Gluten is rich in glutamine, leading to hydrogen bonding. Hydrophobic amino acids are also abundant, contributing to insolubility of these proteins in water and buffers. Most gluten proteins are either prolamins (gliadin in wheat), soluble in aqueous solutions of alcohols, or glutelins (glutenin in wheat), insoluble in alcohol but soluble in acid, alkali, denaturing agents, detergents, or reducing agents. At least three classes of monomeric gliadins exist (α/β, γ and ω), which vary in size and composition. In contrast, glutenin is a polymer of two major [low-molecular-mass (LMW)

high-molecular-mass (HMW)] plus minor subunit classes, joined through disulfide bonds into proteins with relative molecular masses (M_r) ranging into the millions.

It is now possible to better isolate and characterize gluten, and to relate its composition to quality and genotype. This is possible because of improved electrophoretic and chromatographic methods. Each new method showed gluten to be more complex than had been indicated by earlier techniques, and indicated the need for still better methods. As data became more complex and easier to acquire, it also became apparent that visual data evaluation is not sufficient. Computer-assisted statistical procedures are needed to fully reveal information in the results.

We will here review advances in electrophoresis and chromatography for gluten fractionation, and show how these are being used. Other reviews [3–11] summarize earlier progress, and our knowledge of wheat protein composition, structure and functionality; a recent Proceedings volume also provides a useful overview of these topics [12]. We will also review an important related topic: methods to evaluate electrophoresis and chromatography data. Visual evaluation is sometimes adequate, but much information is not readily apparent. Computer-assisted statistical methods of data interpretation promise to enhance the value of chromatography and electrophoresis for analysis of wheat and its proteins.

2. ELECTROPHORETIC METHODS

Electrophoresis was the first procedure to reveal the composition of gluten. Electrophoretic mobility depends on net charge, resulting from ionizable amino acids, and on polypeptide size. Gluten proteins have few basic or acidic residues, giving them low charge/mass ratios, but electrophoresis can separate them in several modes: (a) both size and charge influence mobility; (b) charge differences can be suppressed by a detergent such as sodium dodecyl sulfate (SDS) so separations depend only on size, as in SDS polyacrylamide gel electrophoresis (PAGE); and (c) size does not influence mobility, as in isoelectric focusing (IEF), since proteins migrate in a pH gradient to positions where they are electrically neutral. These applications will be reviewed in this section.

2.1. Moving-boundary electrophoresis

Early attempts to analyze gluten by electrophoresis were by moving-boundary electrophoresis in an open tube. With gluten, this method was originally limited by gluten's solubility. Jones *et al.* [13] described buffers permitting separations of gluten. One buffer, pH 3.1 aluminum lactate (and variations involving other lactate salts), became a highly successful solvent for gliadin electrophoresis. Using it, gluten was shown to contain at least five gliadins and one glutenin [13]. Gluten composition varied qualitatively and quantitatively among and within wheat species. This method thus gave the first real evidence that gliadin and glutenin were themselves heterogeneous.

2.2. Starch gel electrophoresis

Lactate buffer was later combined with zone electrophoresis in starch gels [14–17], which stabilize electrophoretic separations. Starch gel electrophoresis (SGE) remains valuable for fractionating gliadins, showing them to be more heterogeneous than originally known. Today, 20–30 bands may resolve by SGE. Woychik *et al.* [17] first proposed that gliadins be subclassified as ω, γ, β and α, based on increasing mobility. SGE also readily distinguishes gliadins from albumins, which have greater mobilities, and from glutenin, which is polymeric and too large to enter the gel or give distinct bands.

SGE of wheats revealed major differences in gliadin compositions [18,19], permitting varietal identification. Standard methods have been proposed [20], and varietal identification by gliadin SGE is still used today [21,22]. Albumins and globulins also separate well by SGE, differentiating genotypes and classes [18]. SGE uses simple equipment and a non-toxic support. Nevertheless, resolution is variable, and it is difficult to reproducibly prepare starch gels, which are not very stable.

2.3. Polyacrylamide gel electrophoresis

Polyacrylamide has generally replaced starch as the medium for gel electrophoresis of wheat proteins. Separations are analogous, but resolution of PAGE is generally better. Polyacrylamide gels can be prepared reproducibly and are stable, and thin

slabs can be prepared and used, enhancing resolution because of better heat dissipation. Gel pore size and concentration can also be varied to separate proteins of different sizes.

Wheat proteins were first separated by PAGE in 1963 [23]. Since then, techniques have been modified to improve resolution, especially of gliadins [21,24–30]. An example (Fig. 1) compares Canadian spring wheats on a 6% gel (200 × 150 × 3 mm) [31]. Cultivars are easily differentiated by characteristic fingerprints. PAGE of albumins and globulins also can discriminate among genotypes [21,32,33], but differences are more quantitative than qualitative.

Most PAGE separations of gliadins use pH 3.1 aluminum lactate-lactic acid buffer and a uniform gel. Good resolution is also achieved with acetic acid–glycine buffer [34] or at alkaline pH [32]. Gradient PAGE is also very successful: polyacrylamide concentration increases with migration distance, reducing mobilities and sharpening bands [21,25]. Precast gels can give especially convenient, reproducible, and rapid results [25].

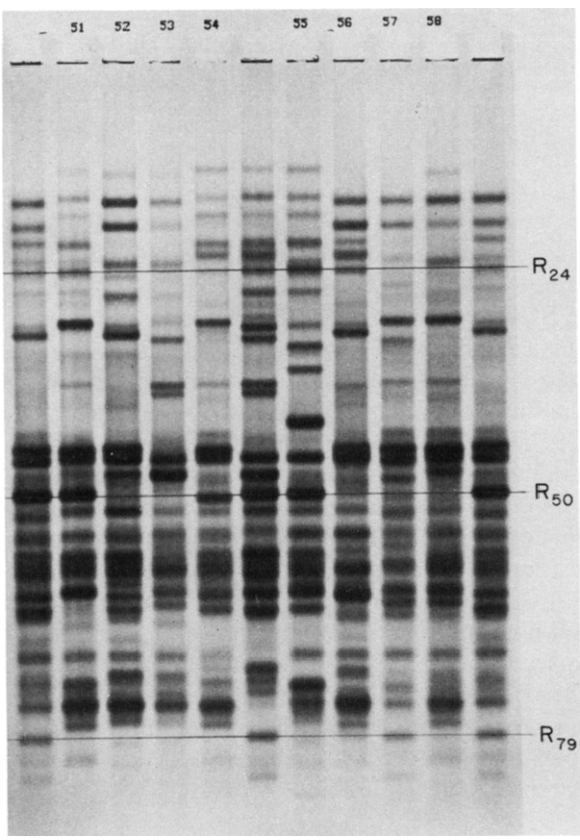

Fig. 1. Polyacrylamide gel electropherograms (6% gel, 200 × 150 × 3 mm) in pH 3.1 aluminum lactate–lactic acid buffer of gliadins extracted with 70% ethanol from several Canadian common spring wheat cultivars. Horizontal lines denote migration distances of reference bands of the standard cultivars Marquis (outside slots) and Neepawa (center), used to standardize gels and increase precision of the mobility data. From Sapirstein and Bushuk [31].

Resolution, precision, and speed of wheat protein PAGE have also improved. Lookhart *et al.* [35] showed that sodium lactate buffer can be used for gliadin PAGE, and that resolution is generally best at 7–10°C. Clements [36] also showed that lactate buffer without aluminum gives good resolution of gliadins extracted with ethylene glycol. Variables such as gel thickness; buffer type, pH, and ionic strength; temperature; catalyst and apparatus design can also be adjusted to optimize gliadin PAGE for varietal identification [37].

Rapid PAGE methods are especially useful in determining wheat varietal purity and for selection and marketing, where many analyses must be done. One such method can identify varieties within 1 h [38,39]. Gliadins are extracted with 6% urea or ethylene glycol, and separated for 9 min in 75 × 35 × 1 mm gradient gels in pH 3.1 sodium lactate buffer. Discrimination between varieties is as good as on standard-sized gels. Labor-saving techniques of protein extraction and application are also useful for screening samples [40], and standardized PAGE methods for cultivar identification [41,42] can reduce variation between laboratories.

Sapirstein and Bushuk [31,43,44] significantly improved the precision of PAGE for varietal identification. Gels (as in Fig. 1) are photographed, and migration distances of bands are determined and entered into a computer with a digitizing tablet. Positions of bands are then normalized based on mobilities of three reference bands in standard cultivars. Band intensities are also estimated and entered into the data base. Unknown samples are then compared to stored data for known varieties. This process can be automated, and gives precise, accurate PAGE varietal identification.

2.4. Isoelectric focusing

Whereas gel electrophoresis was borrowed from clinical disciplines, IEF was first used to fractionate

wheat albumins and globulins [45,46]. Proteins separate in a pH gradient according to differences in isoelectric points, complementing other electrophoresis procedures. High resolution results, sometimes superior to that of PAGE [47,48]. For gliadins, Wrigley [45,46] observed several bands with isoelectric points of 5–8, and confirmed that IEF patterns differ among varieties [21,25]. The high cost of ampholytes may limit use of IEF for routine analyses, but this is less serious with miniature gels. IEF is also common as the first dimension in two-dimensional (2D) procedures (see below).

An important recent IEF variation uses Immobilines (*i.e.*, immobilized ampholines) to stabilize the pH gradient. This technique appears to have been used only once for wheat proteins, which focused without cathodic drift in a pH 4–10 gradient as the first separation in a 2D separation [49]. This eliminated the need to use two IEF techniques, including non-equilibrium pH gradient electrophoresis (NEPHGE) to resolve basic proteins.

The other major advance in IEF of wheat proteins is free-flow preparative IEF [50,51]. This procedure uses a Rotofor apparatus (Bio-Rad Labs., Richmond, CA, USA) consisting of a cylindrical chamber with 20 compartments divided by membranes that maintain separations. After IEF using pH 3–10 ampholytes, fractions are recovered from each chamber.

Results of such a separation of several hundred mg extracted gliadin and glutenin are shown in Fig. 2 [51]. SDS-PAGE revealed an excellent separation, with only slight overlap between fractions. Results also clearly show native glutenins of different compositions.

2.5. Sodium dodecyl sulfate–polyacrylamide gel electrophoresis

While SGE, PAGE and IEF provided excellent separations of gliadins, albumins, and globulins, they generally did not separate glutenin (or its subunits) well because of its polymeric nature, high M_r, associative tendencies, and poor solubility. The advent of SDS-PAGE [52] made characterization of glutenin possible. Proteins are turned to random coils, and charge differences are eliminated by bound SDS. Separations occur almost totally on the basis of size, and M_r can be estimated by comparison to mobilities of standard proteins.

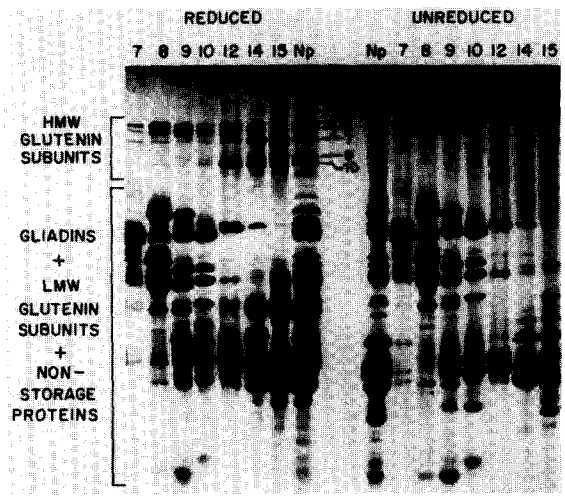

Fig. 2. SDS-PAGE patterns of reduced (left) and unreduced (right) proteins extracted (using a pH 8 Tris–HCl buffer containing urea and SDS) from the wheat variety Neepawa and fractionated by free-flow preparative IEF. Numbers at the top of the gel are fraction numbers; Np is a total protein extract of Neepawa flour. From Ng *et al.* [51].

Bietz and Wall [53] first used SDS-PAGE to characterize gluten. Analysis without disulfide bond cleavage gave a streak from the origin, but adding a reducing agent revealed discrete zones from M_r 10 000 to *ca.* 140 000. Other studies, however, show that M_r estimates vary considerably with procedure. M_r estimates of some HMW glutenin subunits may be nearly twice actual M_r. This anomaly has not been well explained, and is not well understood.

Using SDS-PAGE, Bietz and Wall [53] resolved gliadin into components of apparent M_r 30 000–80 000 (Fig. 3). The largest M_r 60 000–80 000 were ω-gliadins, and α-, β- and γ-gliadins formed overlapping zones (M_r 30 000–40 000) below ω-gliadins. Early studies also showed albumins and globulins to be very heterogeneous, most with $M_r < 40\,000$. Better separations of albumins and globulins were later achieved [54].

SDS-PAGE results for glutenin (Fig. 3) [53] were even more revealing. It contained both HMW (M_r 100 000–140 000) and LMW (M_r 30 000–50 000) subunits joined through disulfide bonds. The large size of HMW subunits and their occurrence only in glutenin, wheat's strength protein, indicated a role in breadmaking. This was confirmed by SDS-PAGE of wheat aneuploids [55], which showed

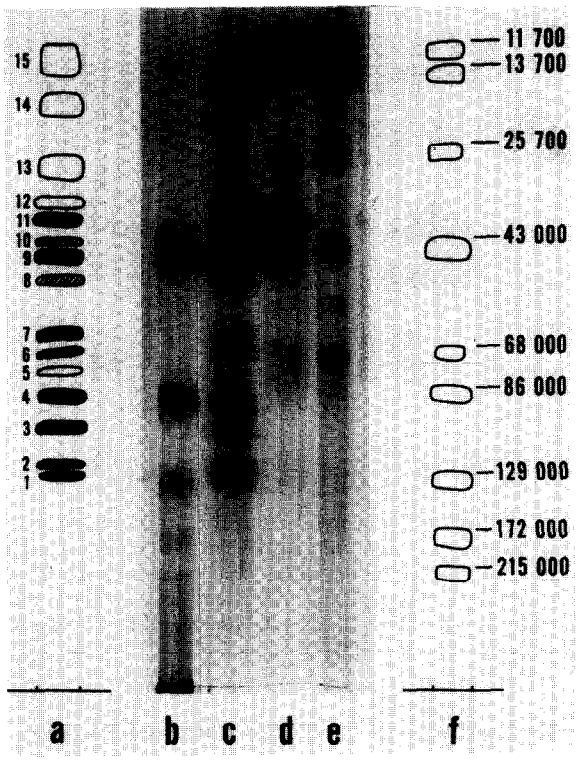

Fig. 3. SDS-PAGE analyses of wheat gliadin (d), glutenin (a and c), and standard protein mixtures (b, e, f). Apparent M_r are indicated to the right of the gel. From Bietz and Wall [53].

them coded by genes on chromosomes (1B and 1D) associated with breadmaking quality.

Initial SDS-PAGE studies showed minimal variation among bread wheat varieties [55], but higher-resolution methods later showed considerably greater differences [56–58]. SDS-PAGE may thus be useful for cultivar identification [21,59,60]. SDS complements methods based on charge, and shows apparent M_r rapidly and precisely if suitable references are used [61].

Use of higher resolution SDS-PAGE revealed major differences in HMW glutenin subunit compositions. Four or five HMW subunits occur in a variety, and 20 or more different HMW subunits are in all wheats. These subunits indicate alleles related to breadmaking quality [62–64]. They can also be analyzed in total protein extracts [58,65], simplifying sample preparation and making SDS-PAGE useful for selection during breeding.

Glutenin also has LMW subunits with apparent M_r similar to gliadins (Fig. 3). This M_r similarity, and the difficulty of separating gliadin from glutenin, originally made SDS-PAGE characterization of LMW glutenin subunits difficult. Newer procedures, however, give excellent resolution of LMW glutenin subunits. In one procedure [66,67], unreduced total protein seed extracts are first electrophoresed so gliadins migrate from the origin. A strip of gel next to the origin, containing unreduced glutenin, is then removed, equilibrated with buffer containing reducing agent, loaded onto a new gel slab, and again subjected to SDS-PAGE. From 7 to 16 LMW glutenin subunits resolve, free from overlapping gliadins. Varietal differences among these subunits may relate to pedigree and quality. An alternative two-step one-dimensional method [68] analyzes gliadins as well as HMW and LMW glutenin subunits. Unreduced proteins are first separated by acid PAGE, giving good separations of gliadins. A gel strip from below the sample wells is then equilibrated with SDS and reducing agent, and electrophoresed by SDS-PAGE, separating HMW and LMW glutenin subunits free of overlapping gliadins.

Fig. 4 shows an example of results achievable today by SDS-PAGE in comparing glutenin subunit compositions from various wheats [69]. Glutenin was purified by a rapid dimethyl sulfoxide extraction procedure [70], reduced, pyridylethylated to stabilize cysteines, and analyzed by gradient SDS-PAGE with silver staining [71]. Subunits of 37 M_r classes, from 30 000 to 116 000, were revealed. Such resolution is due both to the improved electrophoresis procedure (a 0.75 mm thick 11.0–16.5% polyacrylamide gradient gel) and to the small amount of sample applied when silver stain is used. Excellent SDS-PAGE results on similar gradient gels were also reported by Marchylo and co-workers [72,73], who resolved 7–11 HMW and 25–32 LMW polypeptides from gliadin plus glutenin. This procedure could differentiate most varieties.

Rapid SDS-PAGE represents another major advance. When reduced wheat proteins are separated on 70 × 80 × 1 mm 12% polyacrylamide gels, separation time decreases from about 20 h (for 160 × 140 × 1.5 mm gels) to 2.5 h, and resolution is nearly as good [74]. Another rapid procedure uses even smaller (50 × 43 × 0.45 mm) gels with the Pharmacia PhastSystem [73]. Typical results are shown in

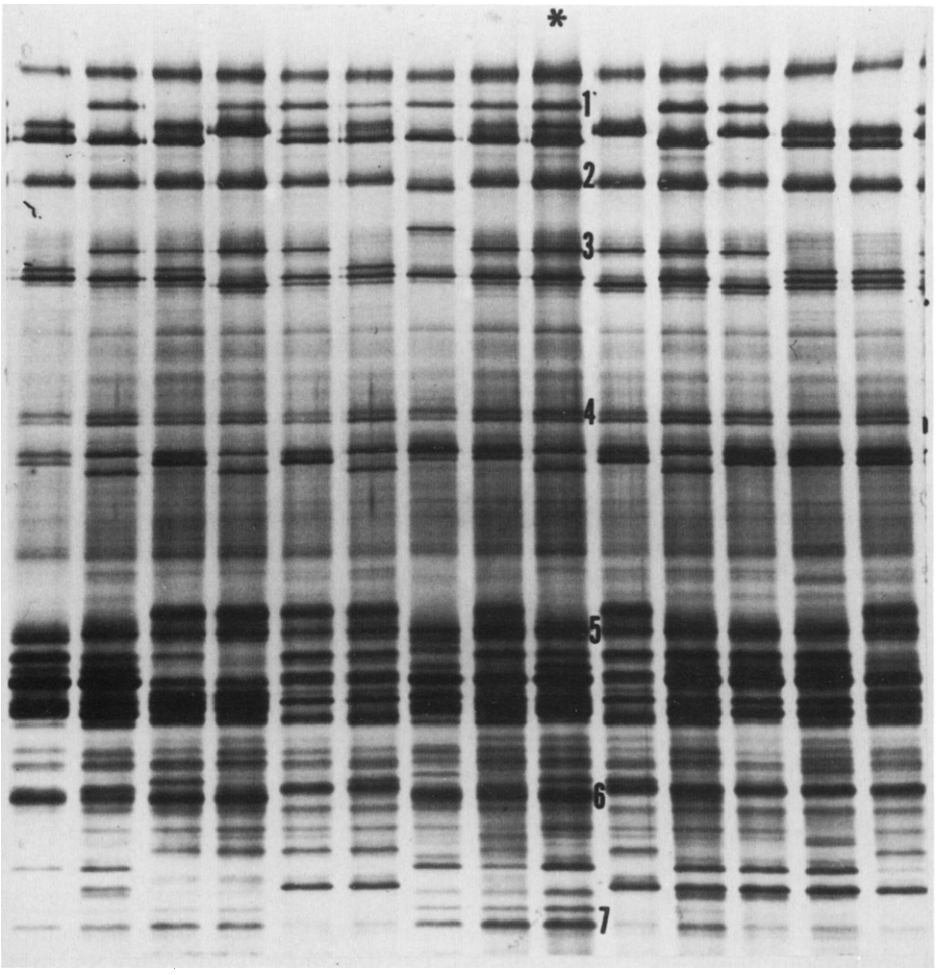

Fig. 4. SDS-PAGE separations of glutenin subunits from experimental wheat lines. Apparent M_r of numbered bands are: 1 = 115 300; 2 = 99 600; 3 = 87 900; 4 = 61 300; 5 = 45 300; 6 = 37 300; 7 = 32 000. From Graybosch *et al.* [69].

Fig. 5. Resolution by this method, with a 45 min separation and 44 min silver staining, is nearly as good as that on conventional large gradients gels. PhastGels distinguish most cultivars, and glutenin HMW subunits are readily identified. This procedure is effective for rapid analysis when samples are few or in non-standard situations.

SDS-PAGE can also provide valuable quantitative data. Kolster and Van Gelder [75] describe procedures for extraction, SDS-PAGE, staining, and densitometric quantitation of wheat proteins. HMW glutenin subunits can be both identified and quantified in one step.

Another interesting SDS-PAGE advance is electroendosmotic preparative electrophoresis [76]. Buffer flow between electrodes (electroendosmosis) moves electrophoresed proteins along a cylindrical gel, from which they are eluted and collected. Curioni *et al.* [76] used this method to isolate mg quantities of five pure HMW glutenin subunits in one step.

2.6. Two-dimensional electrophoresis

Any two electrophoresis procedures, especially when complementary separation modes are involved, can be combined to enhance protein resolu-

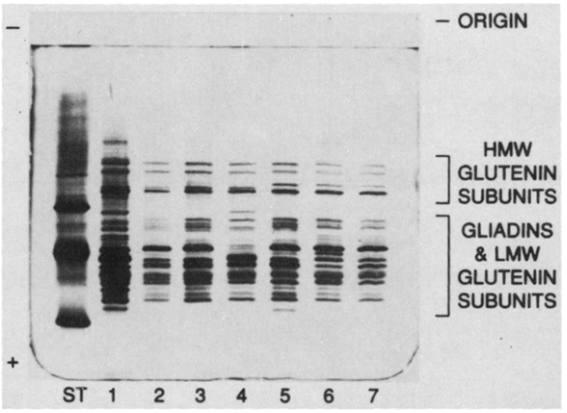

Fig. 5. Fast horizontal SDS-PAGE of wheat storage proteins using an 8–25% gradient PhastGel with Pharmacia's PhastSystem. Samples shown are: ST = standard mixture of known proteins; 1–7 = storage proteins from seven Canada Western Red Spring wheat cultivars. Proteins were extracted with 50% 1-propanol containing 1% dithiothreitol and 41 m*M* Tris–HCl, pH 8.0, and alkylated with 4-vinylpyridine before SDS-PAGE. From Marchylo *et al.* [73].

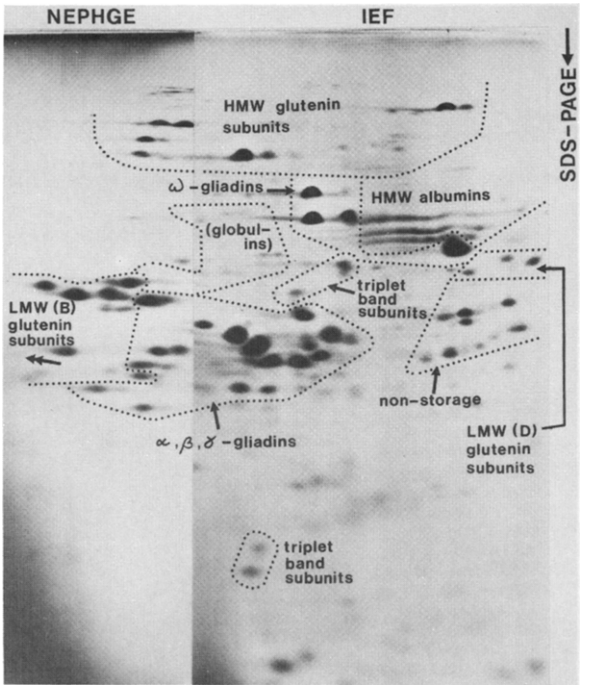

Fig. 6. Fractionation of wheat endosperm proteins by NEPHGE × SDS-PAGE and IEF × SDS-PAGE. The gels are overlapped to show a continuous pH gradient. The map is divided into areas according to biochemical and genetic properties of the proteins. From Payne *et al.* [84].

tion. The best separations of wheat proteins to date have been by 2D electrophoresis.

Wrigley and Shepherd [77,78] first separated wheat albumins and gliadins by 2D electrophoresis by combining IEF with SGE. Since these methods are complementary, single bands in one separation may further resolve in the other. Combined IEF and SGE also showed chromosomal locations of genes coding water- and chloroform–methanol-soluble wheat proteins [79]. Other 2D separations of albumins/globulins combine IEF with PAGE [47,80].

Other combinations of procedures also give excellent separations. For example, PAGE first in aluminum lactate buffer, pH 3.1–3.2, and then in Tris–glycine buffer at pH 9.2 separates albumins, globulins and gliadins [26,81,82], and helped locate their coding genes through aneuploid analysis. Both separations occur in one gel slab.

IEF combined with SDS-PAGE [83], however, generally gives highest resolution 2D separations. For example, Payne *et al.* [84] separated α-, β- and γ-gliadins, LMW and HMW glutenin subunits, and albumins and globulins by combining two IEF procedures, including NEPHGE to separate basic polypeptides, with SDS-PAGE (Fig. 6). Polypeptides of each class have similar size and charge, and group together on the gel. Using similar procedures, Lei and Reeck [85] resolved nearly 500 wheat proteins; Anderson *et al.* [86] achieved similar resolution. Such methods have clarified protein inheritance by locating coding genes [84,87].

Such complex data are, however, difficult to interpret: *e.g.*, rows of spots sometimes result, possibly being charge variants of a protein [88] arising through mutation. Proving such relationships is difficult, but proteins can be electroblotted from gels and sequenced [89,90].

The limit of resolution of 2D procedures has not yet been reached. Tkachuk and Mellish [91], using IEF plus SDS-PAGE, detected about thirteen hundred wheat albumins, globulins, gliadins, and glutenins. A recent 2D prolamin separation is shown in Fig. 7. Optimal staining (generally with silver stains) [92] also enhances resolution: smaller samples can be used, reducing interactions of proteins and making spots more compact.

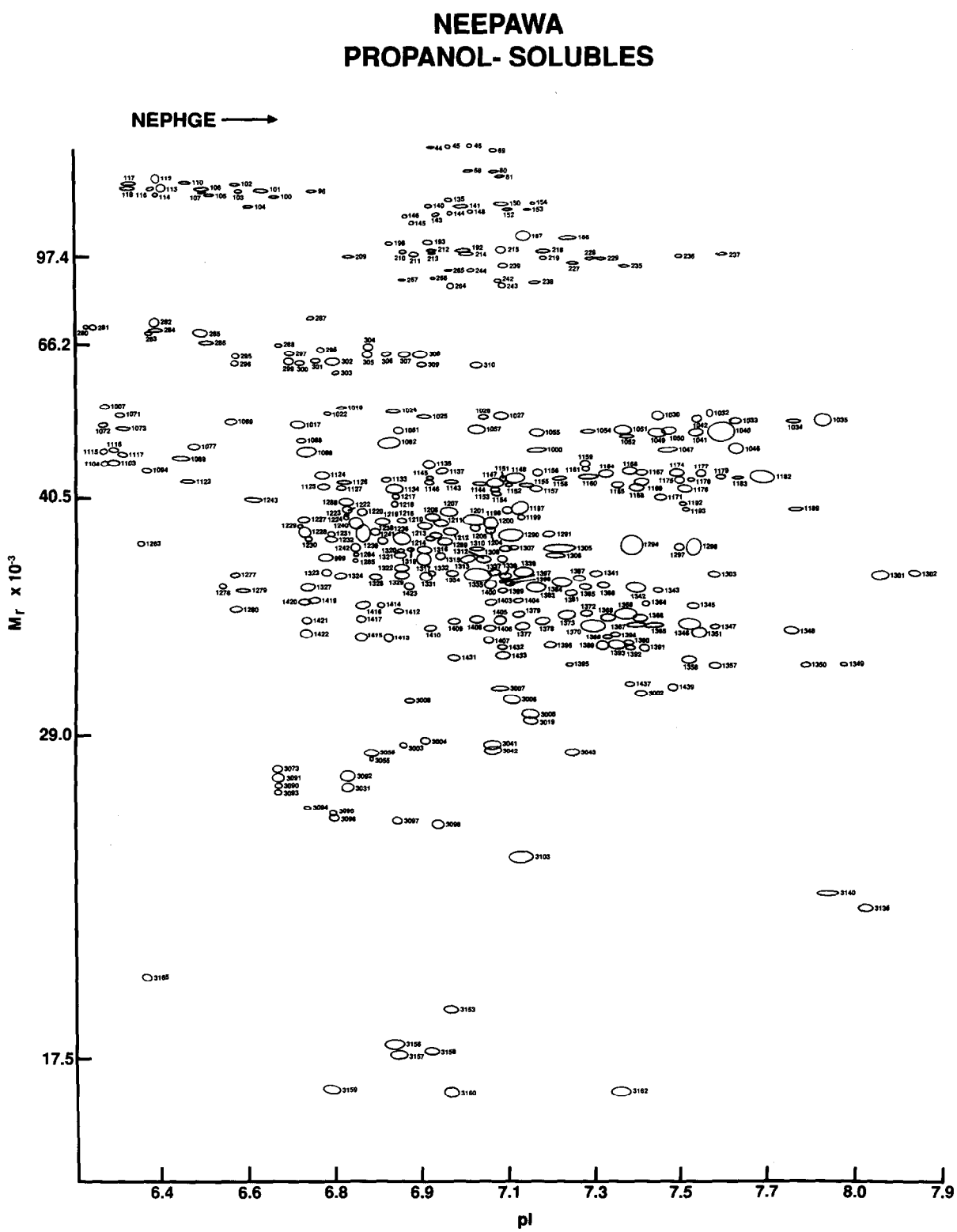

Fig. 7. Composite 2D (IEF + SDS-PAGE) separation of wheat prolamins. The first separation (IEF with pH 3–10 ampholytes; cathode at right) was followed by SDS-PAGE in a 15% gel. Data courtesy of R. Tkachuk.

The ability to achieve such separations is useless, however, without good data analysis. Image analysis can process 2D gel data [49,93]. Computer programs can normalize and correct captured gel images, giving spot files of position and intensity data which can identify varieties or indicate quality.

2.7. Capillary electrophoresis

Capillary electrophoresis is also a promising fractionation technique for proteins [94–97]. Separations occur in uncoated or coated glass capillaries (25–75 μm I.D.). High voltage (10–30 kV) plus efficient cooling permit rapid (10–30 min) high-resolution separations. Instruments are automated, give reproducible separations, and have good data capabilities. Fig. 8 shows a capillary electrophoresis separation of wheat proteins. Approximately 58 components result, counting partially resolved shoulders. This equals or exceeds resolution of most other one-dimensional methods. Capillary electrophoresis could become valuable for varietal identification, classification, and determination of quality.

3. CHROMATOGRAPHIC METHODS

Liquid chromatography is the second major method used to isolate and characterize wheat proteins [4]. Separations may be based on size or on characteristics imparted by specific amino acids. In size-exclusion chromatography, solutes move through columns packed with porous matrices at rates determined by relative sizes of solutes and matrix pores. In ion-exchange chromatography (IEC), solutes partition between the mobile phase and the support, to which they bind through ionized amino acids; mobility depends on strength of ionic interaction. In hydrophobic interaction chromatography (HIC), separations result from binding through non-polar amino acids to a lipophilic stationary phase.

Original chromatographic separations used large, hand packed columns. Resolution was sometimes good, especially for preparative purposes. As a rule, however, these methods are slow, labor intensive, and irreproducible, and results are hard to quantify. There have been few recent developments in such techniques. There have, however, been major improvements in high-performance liquid chromatography (HPLC) columns and instruments, and HPLC has become the method of choice for many applications. The following sections will first review the use of both types of chromatographic methods for wheat protein analysis.

3.1. Size-exclusion chromatography

Size-exclusion chromatography has indicated molecular sizes of native wheat proteins, shown

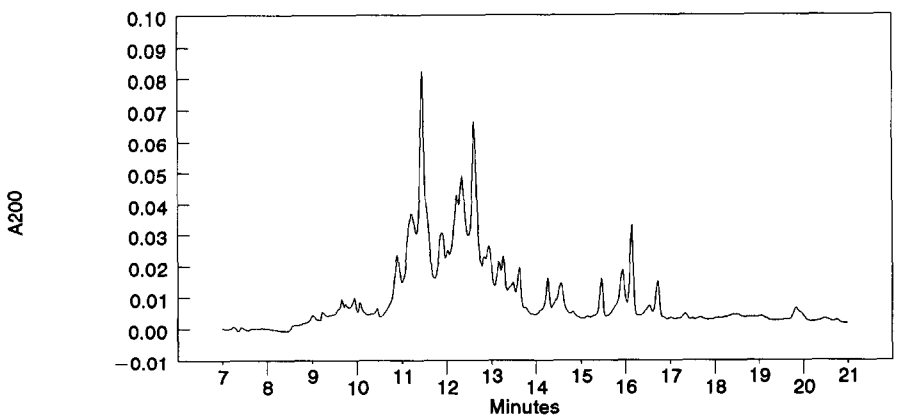

Fig. 8. Fractionation of wheat proteins by capillary electrophoresis. Wheat flour (cv. Centurk) was extracted with 30% ethanol, and the clear supernatant was separated at 10 kV and 40°C on a 75 μm I.D. uncoated capillary (40 cm from inlet to detector) in 0.06 *M* borate buffer, pH 9.0, containing 20% acetonitrile and 1% SDS. Proteins were detected at 200 nm.

how they are related, and revealed their subunit compositions. Gliadin, extracted with 70% ethanol, was first fractionated on Sephadex columns with dilute acetic acid as mobile phase into (a) an HMW (M_r 100 000–400 000) oligomeric fraction; (b) ω-gliadins (M_r 60 000–80 000); (c) γ-gliadins (M_r *ca.* 40 000); (d) α/β-gliadins (M_r 30 000–35 000); and (e) albumins plus globulins [98,99]. Similarly, using 4 *M* urea plus 0.03 *M* acetic acid as solvent, Huebner and Wall [100] showed that glutenin, after cleavage of disulfide bonds and alkylation of resulting cysteines, has three types of subunits differing in size and associative tendencies.

In subsequent size-exclusion chromatography studies [101–107], other solvents and more porous colums were used. Native glutenin was extracted with acetic acid plus urea plus cetyltrimethylammonium bromide [101,102], 5.5 *M* guanidine hydrochloride [104] or SDS [106,107], and fractionated by size on agarose supports. These studies showed size and compositional heterogeneity of native glutenin. M_r values were estimated at 5–20 million based on elution at column void volume or calibration with non-protein standards. Such values are thus questionable; the true M_r of glutenin remains unknown.

3.2. Ion-exchange chromatography

IEC has revealed much about gluten's composition. Woychik *et al.* [108] used carboxymethyl cellulose and a gradient of increasing acid concentration to fractionate gluten. Simmonds and Winzor [109] eluted proteins with a sodium chloride gradient in a 1 *M* dimethylformamide buffer. A useful fractionation was realized, but fractions were still heterogeneous.

IEC methods continued to improve by using dissociating agents (such as urea and dimethylformamide), adjusting elution conditions, and changing pH and ion-exchange media [110–113]. IEC on sulfoethyl cellulose, with a buffer of 3 *M* dimethylformamide, 1 *M* urea, 0.03 *M* acetic acid, and 0.005 *M* HCl and a guanidine hydrochloride gradient, first purified many glutenin subunits [100]. Such methods are still useful for preparative purposes [114–116], but are otherwise little used.

3.3. Hydrophobic interaction chromatography

HIC has been little used with gluten [117–121], largely because its resolution is limited. Yet, since HIC separates proteins on the basis of a unique complementary characteristic, surface hydrophobicity, and since it is the forerunner of reversed-phase (RP)-HPLC (see below), its use will be briefly reviewed.

Caldwell [118] showed that gliadins bind strongly, through hydrophobic sites on protein surfaces, to agarose having phenyl or octyl groups covalently bound. Gliadins were eluted with a gradient of increasing ethanol concentration. PAGE of fractions showed distinct differences: gliadins eluted in the general order ω, β, α and γ. Results showed the potential of hydrophobic separation methods for gluten fractionation.

Popineau and Pineau [120] used HIC, combined with IEC, to purify gliadins. A mixture of γ-gliadins was fractionated by HIC on phenyl Sepharose CL-4B. Purified subfractions were obtained in quantities sufficient for further characterization. Popineau [121] also showed that HMW glutenins could be fractionated by HIC. Glutenin fractions covered a wide range of hydrophobicity, and differed in subunit composition. HIC showed that hydrophobicity is an intrinsic property of gluten proteins.

3.4. High-performance liquid chromatography

While the above-described modes of chromatography can isolate and characterize wheat proteins, these methods are difficult and have many problems. They are slow, and labor intensive. Reproducibility is poor, and columns unstable. Resolution is often inadequate, and quantitation is difficult. HPLC columns and equipment overcame many of these deficiencies. In particular, introduction of uniform (*e.g.*, 5–10 μm) wide pore (typically >300 Å) silica packings with silanols derivatized to permit specific interactions and end-capped to prevent non-specific adsorption was a milestone [122–124]. Reliable equipment also became available with precise flow-rates and gradients, high reproducibility, sensitive detection, automatic operation and excellent data handling. These developments revolutionized isolation, characterization, and knowledge of wheat proteins [3–8,125–129]. The following sec-

tions review milestones in development and application of HPLC to wheat proteins, and describe recent progress.

3.4.1. High-performance size-exclusion liquid chromatography

High-performance size-exclusion chromatography (HPSEC) of wheat proteins was first described by Bietz [3,125], who used TSK-type columns with neutral phosphate–0.1% SDS buffers. Separations were rapid (about 20 min), sensitive, and reproducible. Quantitation was achieved at 210 nm, and indicated M_r agreed well with those from other techniques. Resolution was as good as or slightly better than that by open-column size-exclusion chromatography.

Results also showed that HPSEC can differentiate varieties by analyzing native glutenin or total proteins, and can predict breadmaking quality from M_r distributions of native glutenin or its reduced subunits [130]. These studies showed, however, that reproducible protein extraction is a problem, and that protein M_r distributions change with time due to association.

These problems were dealt with by Dachkevitch and Autran [131], who extracted unreduced flour proteins for 2 h at 60°C with 0.1 *M* sodium phosphate, pH 6.9, containing 2% SDS. The centrifuged extract was analyzed on a TSK 4000SW column. Proteins were extracted reproducibly though not quantitatively; extracts were stable for at least two days. When applied to wheats from different locations and years, this procedure yielded size distribution data highly correlated with baking quality. Results prove the value of HPSEC of unreduced wheat proteins in breeding.

Singh and co-workers [132,133] described another way to extract unreduced wheat proteins for HPSEC. Flour was briefly sonicated in 0.5 *M* sodium phosphate, pH 6.9, containing 2% SDS, gently agitated for 30–120 min, and centrifuged. Extraction was complete without addition of a disulfide bond reducing agent. Proteins from strong wheats were more difficult to extract. This procedure apparently disrupts major non-covalent and covalent forces joining polypeptides. Shear degradation of disulfides probably occurs, converting insoluble HMW glutenin into lower-M_r soluble species.

Proteins extracted by this method from wheats varying in breadmaking quality were analyzed by HPSEC [133]. Three peaks resulted, corresponding to glutenin, gliadin, and albumins/globulins (Fig. 9A). Areas of peaks correlated significantly with breadmaking quality: percentage of the first peak (polymeric glutenin) was highly positively correlated with loaf volume, dough resistance and extensibility, and dough development time. Structural features that differentiate strong and weak wheats must still be retained after sonication. This procedure can be a rapid small-scale (*e.g.*, half-kernel) test for predicting quality potential during breeding.

The low resolution of most HPSEC separations of wheat proteins reflects their size heterogeneity. Batey *et al.* [134] also found, using this procedure, a gradual deterioration in column performance, attributed to SDS in the buffer. The normal buffer was thus replaced with 50% acetonitrile containing 0.1% trifluoroacetic acid. This solvent stabilized the column, and significantly enhanced resolution (Fig. 9B), presumably by disrupting hydrophobic forces which prevent complete dissociation of proteins.

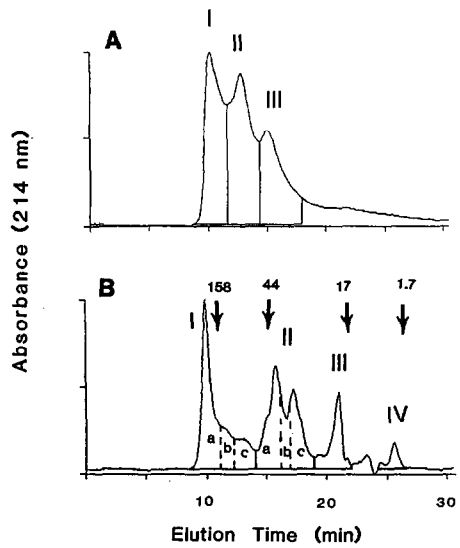

Fig. 9. HPSEC separations of wheat proteins from the cultivar Cook on a Waters Protein Pak 300 column. Proteins were extracted with sonication using 0.05 *M* phosphate buffer, pH 6.9, containing 2% SDS, and analyzed (A) using 0.05 *M* phosphate, pH 6.9, containing 0.1% SDS as chromatographic buffer and (B) using a buffer of 50% aqueous acetonitrile containing 0.1% trifluoroacetic acid. The numbers at the arrows in (B) indicate $M_r \times 10^{-3}$. I–IV and a–c represent specific fractions correlated with quality. From Batey *et al.* [134].

Other extraction and solvent conditions may also be useful. Huebner *et al.* [135] extracted flour with 70% ethanol to solubilize gliadins, and 6 *M* urea–0.05 *M* sodium phosphate, pH 7.5–0.3% dithiothreitol–1% SDS to solubilize glutenin subunits. HPSEC on a Pharmacia Superose 12 column using 0.1 *M* sodium phosphate, pH 7.1, containing 20% acetonitrile, 0.3% SDS, and 0.1% dithiothreitol revealed amounts of HMW gliadins, ω-gliadins, LMW gliadins, albumins, and HMW and LMW glutenin subunits. Statistical analyses of data accurately classified hard red spring and winter wheat cultivars, which has been difficult or impossible by most methods.

3.4.2. High-performance ion-exchange liquid chromatography

As noted above, IEC on traditional columns gives good separations of gluten. This suggested that high-performance IEC (HPIEC) should give especially good separations of these proteins. Surprisingly, however, only two publications report HPIEC separations of gluten polypeptides.

Batey [136] first fractionated gliadins by anion-exchange HPLC on Pharmacia Mono-Q. Only under very alkaline conditions (*e.g.*, pH 10.4), where arginine's ionization is suppressed, were good separations possible. Wheat varieties could be identified, but far fewer gliadins resolved than by PAGE or open-column cation-exchange chromatography.

Most IEC separations of gliadins have been on cation-exchange media, where molecules are positively charged. Larre *et al.* [137] thus attempted cation-exchange HPIEC of gliadins on Pharmacia Mono-S HR. About 20 peaks resolved during 110 min, fewer than by PAGE or RP-HPLC. Each peak consisted of a well resolved group of proteins. Cultivars could be differentiated, showing potential of cation-exchange fast protein liquid chromatography (FPLC) for varietal identification. It remains a challenge, however, to achieve high-resolution HPIEC separations of gluten proteins.

3.4.3. Reversed-phase high-performance liquid chromatography

Adoption of RP-HPLC as a fractionation technique for wheat proteins has been a major development during the last decade. RP-HPLC separates proteins based on surface hydrophobicity, as in a few previous HIC studies.

During the 1970s, RP-HPLC came into wide use for LMW solutes. Columns contained uniform porous silica microspheres (10–30 μm diameter), converted to reversed-phase columns by derivatizing silanols with hydrophobic (*e.g.*, C_{18}, C_8 or phenyl) ligands. Some free silanol groups remained, however, which could ionically bind proteins. Packings typically had 80–100 Å pores, too small to allow ready access of most proteins. Thus, most attempts to use these materials for proteins failed.

Better HPLC columns overcame these deficiencies [122–124]. Wide-pore silicas were derivatized with hydrophobic ligands, and residual silanols were end-capped. Superior equipment also became available.

These developments permitted excellent separations of proteins, and were first applied to gluten by Bietz [138]. Several papers have reviewed RP-HPLC studies of wheat and other cereal proteins [3–8,125–129]. This section will briefly review RP-HPLC achievements and applications, and describe several recent studies.

Methods first described for wheat protein RP-HPLC [138] are still applicable. Proteins are extracted with nearly any solvent, and fractionated on C_4, C_8 or C_{18} columns (150–250 cm long × 4–5 mm I.D.) using a gradient of increasing acetonitrile content (typically between 20% and 60%) with 0.05–0.1% trifluoroacetic acid. Detection is usually at 210–225 nm. Constant temperature ensures good reproducibility, and elevated (50–70°C) temperature often enhances resolution by disrupting hydrogen bonds [139].

Marchylo and Kruger [140] discovered a precaution for wheat protein RP-HPLC. If too large a volume of hydrophobic sample is applied, some proteins, particularly those least hydrophobic, may not bind, and elute with the void volume peak. This can be prevented by limiting injection volume, using multiple small injections, or using less hydrophobic solvents.

3.4.4. Gluten fractionation and characterization by RP-HPLC

All types of wheat proteins have been analyzed by RP-HPLC by varying extraction conditions and gradients. Good separations of gliadins were achieved first [138]. As columns and our understanding of how to use them have improved, so

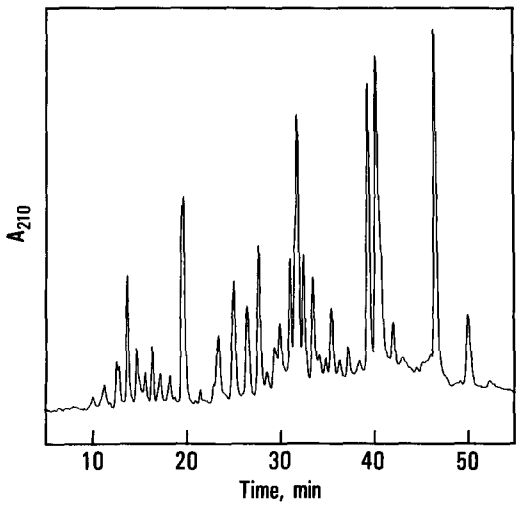

Fig. 10. RP-HPLC separation of gliadins extracted with 70% ethanol from the wheat variety Siouxland. A Vydac C_{18} column was used at 60°C, with a 27–45% acetonitrile (+0.1% trifluoroacetic acid) gradient during 50 min. From Bietz [5].

have separations (Fig. 10). Fifty or more peaks typically resolve from whole gliadin. This exceeds resolution of all except the best 2D PAGE methods.

RP-HPLC has other important advantages. Reproducibility is excellent: retention times typically vary by no more than a few hundredths of a minute. Over long periods, excellent reproducibility can be achieved by normalization based on periodic analysis of standards to compensate for slight changes in the column with time [141–143]. RP-HPLC is also fast; most separations take 30–120 min. If maximum resolution is not needed, smaller columns plus faster flow-rates and steeper gradients can give faster separations [139]; certainly narrow-bore columns will give even greater speed. RP-HPLC is also very sensitive. For example, in breeding a kernel can be cut in half, and analyses done on proteins from the brush end. The germ remains available for propagation if desired.

RP-HPLC's automated nature is another major advantage. An automatic sample injector, as part of a modern system, decreases operator time, increases number of samples analyzed, and improves reproducibility. HPLC is also more easily controlled than is gel electrophoresis. RP-HPLC data can also be accurately quantified. While some amino acid side chains contribute to protein absorbance at 210–225 nm, most absorbance is due to the peptide bond, making absorbance related to mass of protein. Thus, Sutton [144] related variation in amounts of HMW glutenin subunits, as measured by RP-HPLC, to baking performance and quality. The last major advantage of RP-HPLC is that it complements other methods. Proteins separate by surface hydrophobicity, not size or charge. Thus, RP-HPLC may show single PAGE or IEF bands to contain several polypeptides varying in hydrophobicity, and RP-HPLC peaks often contain several charge or size variants [145].

RP-HPLC also gives excellent separations of glutenin. These are more complicated than gliadin analysis since sequential extraction is needed to separate glutenin from other proteins. Disulfide bonds must also be cleaved to liberate glutenin subunits; resulting cysteines are often alkylated to prevent reoxidation. Fig. 11 shows a typical separation of reduced–alkylated glutenin subunits [146]. HMW subunits, associated with breadmaking quality, elute first, and are well resolved; LMW subunits elute last, and are not as well resolved. Better fractionation of these LMW subunits is possible if they

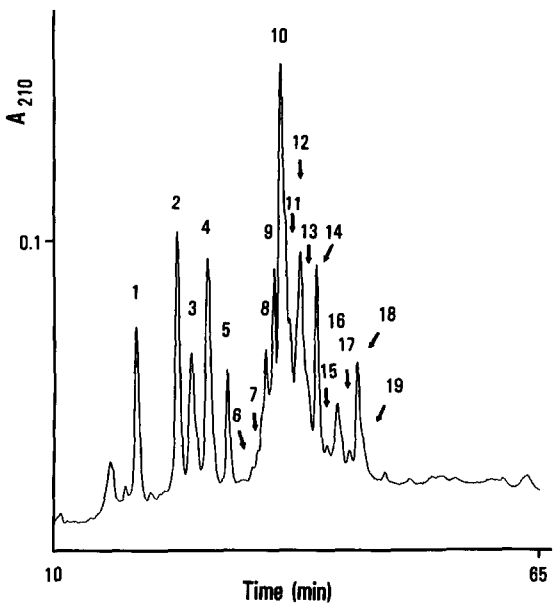

Fig. 11. RP-HPLC separation of pyridylethylated glutenin subunits from the wheat variety Chinese Spring. Peaks 1–4 correspond to the HMW subunits from this cultivar; later eluting polypeptides are LMW subunits. From Burnouf and Bietz [146].

are first separated from HMW subunits by solubility in neutral 70% ethanol [147]. LMW glutenin subunits then resolve into 20 or more components that differentiate varieties and indicate baking quality.

3.4.5. Varietal identification and genetic studies

As with PAGE, varietal identification from gliadin fingerprints is possible by RP-HPLC [148]. Expression of these proteins is nearly constant for a variety, though slight quantitative differences may occur for a variety grown at different locations [149–151]. HMW and LMW glutenin subunits can also differentiate and serve to identify cultivars [147].

RP-HPLC can also be useful during breeding to select germplasm based on specific proteins. For example, RP-HPLC can detect the 1BL/1RS wheat-rye translocation [152], which transfers desirable rye characteristics to wheat, but may also make dough sticky or reduce gluten strength. Similarly, Sutton *et al.* [153] showed that quantitative RP-HPLC of HMW glutenin subunits indicates potential loaf volume and bake scores in breeding lines.

RP-HPLC can also prove varietal purity. For example, the land race "Nap Hall", a source of genes for high protein and lysine, is very heterogeneous [148]. Heterogeneity also occurs, as biotypes, in many modern varieties. RP-HPLC and PAGE are equally effective in discriminating biotypes [154].

In the USA, wheat is traded by class, not variety. Some classes, such as hard red spring and winter, are difficult to differentiate. Endo *et al.* [155,156] differentiated these classes from integrated gliadin RP-HPLC data. The statistical partial least squares (PLS) procedure also correctly classified many hard red winter and spring varieties from non-integrated RP-HPLC gliadin data [157].

Scanlon *et al.* [143] showed that varietal identification can be made automatic and objective by gliadin RP-HPLC analysis. Normalized peak heights and retention times provide characteristic signatures that, through comparison to data for known wheats, reliably identify varieties.

Many wheat aneuploids, having absent or duplicated chromosomes or chromosome arms, are available. RP-HPLC of proteins from such lines can locate genes that code specific polypeptides. Gliadins and glutenin subunits were analyzed by this procedure [146,158]. Results identified wheat's HMW glutenin subunits, associated with breadmaking quality, and showed that each gluten polypeptide type has unique surface hydrophobicity characteristics.

3.4.6. Quality prediction by RP-HPLC

RP-HPLC can indicate quality by analyzing proteins which directly affect functionality, such as HMW glutenin subunits, shown by SDS-PAGE [63] and RP-HPCL [159] to be markers of alleles associated with breadmaking quality. Other proteins may be markers of genes linked to other genes which directly affect quality. For example, Bietz *et al.* [148] used RP-HPLC to show that late eluting γ-gliadins were correlated with pasta quality. These proteins correspond to gliadins "42" and "45", associated with durum weakness and strength, respectively [160]. This characteristic could be screened for by RP-HPLC in as little as 5 min.

Another example of RP-HPLC revealing wheat

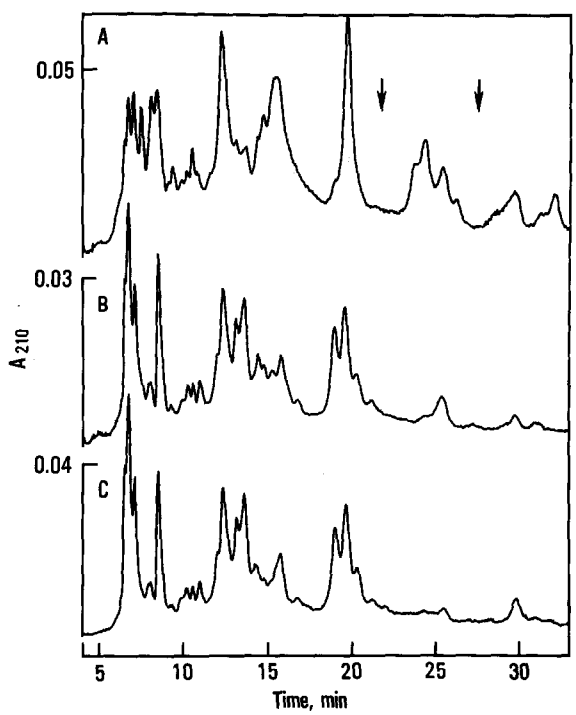

Fig. 12. RP-HPLC of gliadins from the wheat varieties (A) Iuanillo, (B) Westbred and (C) Yecora Rojo. The area of peaks eluting between the two arrows was negatively correlated with baking quality. From Huebner and Bietz [161].

quality is in Fig. 12. Comparison of integrated data with baking characteristics enabled Huebner and Bietz [161] to identify a late eluting baking quality gliadin fraction, the area of which correlated negatively with quality measurements. Selection for this criterion could be useful for breeding, marketing, and quality control.

Another interesting study relating gluten to breadmaking quality is by Van Lonkhuijsen *et al.* [162]. HMW glutenin subunits explain only part of the variation in wheat quality, so wheats with the same HMW glutenin subunits, but varying in quality, were studied. Gliadins were fractionated by RP-HPLC, and data were integrated and analyzed statistically. Results identified specific gliadins which strongly influence breadmaking properties.

Others are also using RP-HPLC data to study quality. Primard *et al.* [163] showed, by statistical analysis of RP-HPLC and SDS-PAGE data, that breadmaking quality depends on many different proteins. The dimension reduction techniques PLS and principal component analysis (PCA) can also objectively identify quality related proteins without integrating RP-HPLC data [157,164].

3.4.7. Role of computers in wheat protein RP-HPLC

Today, chromatographic software is available from HPLC suppliers and other sources that permits acquisition and storage of raw, non-integrated data for later processing. Ten years ago, however, this was not true. Many of the first HPLC studies of wheat proteins at the National Center for Agricultural Utilization Research, Peoria, IL, USA, were possible only because R. Butterfield and colleagues of the computer staff developed programs that permit unattended acquisition of raw data (*i.e.*, detector readings at equal intervals) and its storage on a mainframe computer. Stored data could then be integrated automatically, or in a manual mode in which the operator specifies peaks and baselines, or even by Gaussian deconvolution. Baselines could be corrected. Data could be viewed many ways to compare chromatograms, and plotted to any scale. Chromatograms could be directly compared, permitting aneuploid analysis, mixture analysis, and determination of pedigrees. Data could also be translated to a form used by statistical programs to relate protein composition to quality.

4. INTERPRETATION OF ELECTROPHORESIS AND CHROMATOGRAPHY DATA

The mobility–density plot commonly reported in electrophoresis is a spectrum of bands with varying intensities and locations, which characterizes the solute. Earlier studies focused on interpretation of these spectra, primarily in terms of presence or absence of certain bands, but also incorporating ordinal intensity scores to refine characterization of the solute (see, for instance, ref. 165). More recently, densitometry provides a continuous trace of density *vs.* mobility. The resulting plot has the same character as a chromatogram. As a result, similar statistical methods apply to both electrophoresis and chromatography. This discussion treats both methods at once, because the basic statistical object is the same in each case: a plot of intensity *vs.* location (mobility for electrophoresis; retention time for chromatography). The major choices to be made in treating the data are (a) whether to normalize the area under the plot or the location; (b) whether to operate on peak areas or heights; and (c) whether to interpret the intensity *vs.* location plot directly.

4.1. Subjective visual methods

Damidaux *et al.* [166] found, by inspection of electropherograms, that durum wheats having gliadin "42" (indicating relative mobility) tend to have weak pasta quality, but those having instead band "45" tend to have strong pasta quality. Burnouf and Bietz [160] observed precisely analogous peaks by RP-HPLC of durum wheats, again based on subjective visual inspection of data. The dinstinction is so clear that one hardly needs statistics. Fig. 13 shows the componentwise mean chromatograms from RP-HPLC of a collection of group 42 and group 45 durum wheats. The graph also shows componentwise standard deviations for variation about the mean in each group.

In more complex examples, however, formal statistical methods come to the fore. The eye may see patterns simply because of random fluctuations. In such instances a test of statistical significance is needed. On the other hand, strong trends involving multiple variables may be masked in visual inspection of the data. In such instances multivariate statistical methods are often useful.

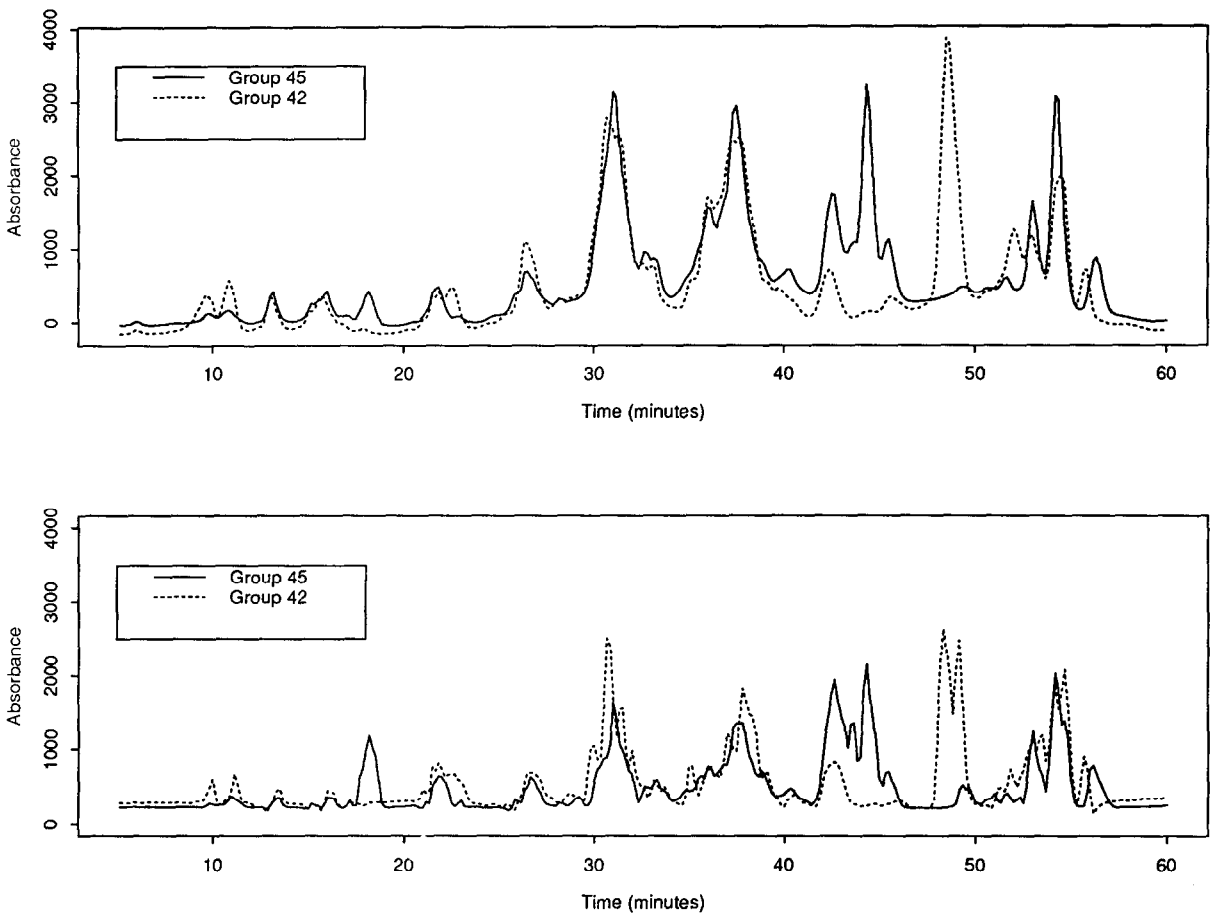

Fig. 13. Componentwise means (top) and standard deviations (bottom) for group 42 and group 45 durum wheats.

Statistical methods often have corresponding plots that can enhance understanding of the data. For instance, in multiple linear regression the effect of a variable can be seen from the so-called added variable plot. This is a scatter plot of projections of the response and regression variable into the space orthogonal to the remainder of the regression variables in the model [167]. The slope of a least squares line fit to this scatter plot is equal to the least square estimate of the parameter for this variable for the multiple linear regression model. Such plots can reveal violations of the modeling assumptions such as non-linearity.

It is worth emphasizing therefore that summary statistics and computational algorithms do not replace visual inspection of data. Graphical methods may reveal clear patterns in data that render unnecessary the need for formal statistical methods. On the other hand, preliminary plots of data may expose problems that affect the type of analysis to be done. Texts on applied statistics such as Weisberg [167] emphasize the use of graphical diagnostics such as residual plots, which can reveal unanticipated phenomena and point to important refinements of the model.

4.2. Varietal identification via similarity scores

The earliest attempts at quantitative analysis of electrophoresis data were concerned with identification of wheat varieties by comparison with a library of electrophoregrams of known varieties [43,165].

An electropherogram to be classified would be scored on the basis of peak matching to each member of the library. Usually one or more reference varieties would be run on the same gel to standardize relative mobilities and peak or band intensities. Lookhart *et al.* [165] used a five-point scale for intensity, and matching scores for the agreement between each pair of peaks. The overall agreement score was obtained by summing over all the peaks. The new sample would then be classified according to the highest match score in the library. Sapirstein and Bushuk [43] used a related strategy, but with a nine point scale for intensity and a different set of scores for matching. In both cases intensities were determined subjectively. More recent work used densitometry to automate scoring of intensities. For instance, Cox *et al.* [168] scanned electropherograms on a densitometer and discretized the output to a scale like that of Lookhart *et al.* [165].

If a comprehensive library of types is available, similarity scoring is appealing in its directness. However, these may be room for improvement in the methodology. Little guidance is available on how to decide if a previously unclassified variety has been encountered. Empirical guidelines might be developed on the basis of cumulative experience, but these would be highly context dependent, depending on the current library of types and presumably on the type of grain under study. There is no mathematical reason to restrict to ordinal scores in comparing peak patterns. One might just as easily measure similarity or distance between vectors of densities or peak heights. The main practical considerations in the use of oridinal scores seem to be that they are less susceptible to anomalous readings than raw peak heights, and they may be easier to scale.

4.3. Hierarchical clustering methods

One method of enhancing the analysis is to use similarity scores to cluster varieties into more homogeneous subgroups, and then to characterize the groups in terms of their physical properties. This was the approach of Wrigley *et al.* [169] and Du Cros [170] who employed the minimum spanning three (MST) to develop empirical taxonomies of wheat varieties from similarity scores. The resulting tree-structured organization of the varieties provided the means for grouping them. An added feature was the graphical representation of degree of similarity in terms of distances along branches of the tree. Gower and Ross [171] provided a clear exposition of the method.

Similarities are generally expressed as fractions or percentages. Assuming they are fractions between zero and one they may be translated into distance-like measures in the following way:

$$\text{distance}^2 = 1 - \text{similarity}.$$

Other schemes are possible, and it is often easier to start with a distance rather than a similarity measure. There are mathematical and algorithmic benefits in choosing the distance measure to be a true distance in the sense of obeying the axioms of Euclidean geometry [171].

It is perhaps easiest to understand the MST in the two-dimensional case. Fig. 14 shows the MST for 17 wheat varieties based on the heights of two RP-HPLC peaks, selected only to obtain a somewhat complex MST for illustration. The MST was computed with the aid of S-Plus (Statistical Sciences, Seattle, WA, USA). Pairs of peak heights are represented as points in the plane. They are joined by line segments to form a tree. The MST shown is the tree whose branches have minimum total length among all possible spanning trees. Formally a spanning tree is a structure in which: (1) there are no closed loops; (2) each point has at least one line segment attached to it; and (3) there is a path from any point to any other point [171]. The two-dimensional case is easiest to understand because distances between all points are represented by distances in the plane. With more than two variables a plot like Fig. 14 will accurately reflect distances between neighboring points on the tree, but it will give little information on distances between nonadjacent points. Friedman and Rafsky [172] discussed the use of the MST for plotting multivariate data.

The MST contains the information needed to construct a single linkage clustering tree as illustrated in Fig. 15. Single linkage means that two clusters are joined at level h if at least one pair of points, with one member of the pair from each cluster, is within h units. The horizontal scale in Fig. 15 is the largest distance from any point to its nearest neighbor in the cluster. Thus, if we require each point in a

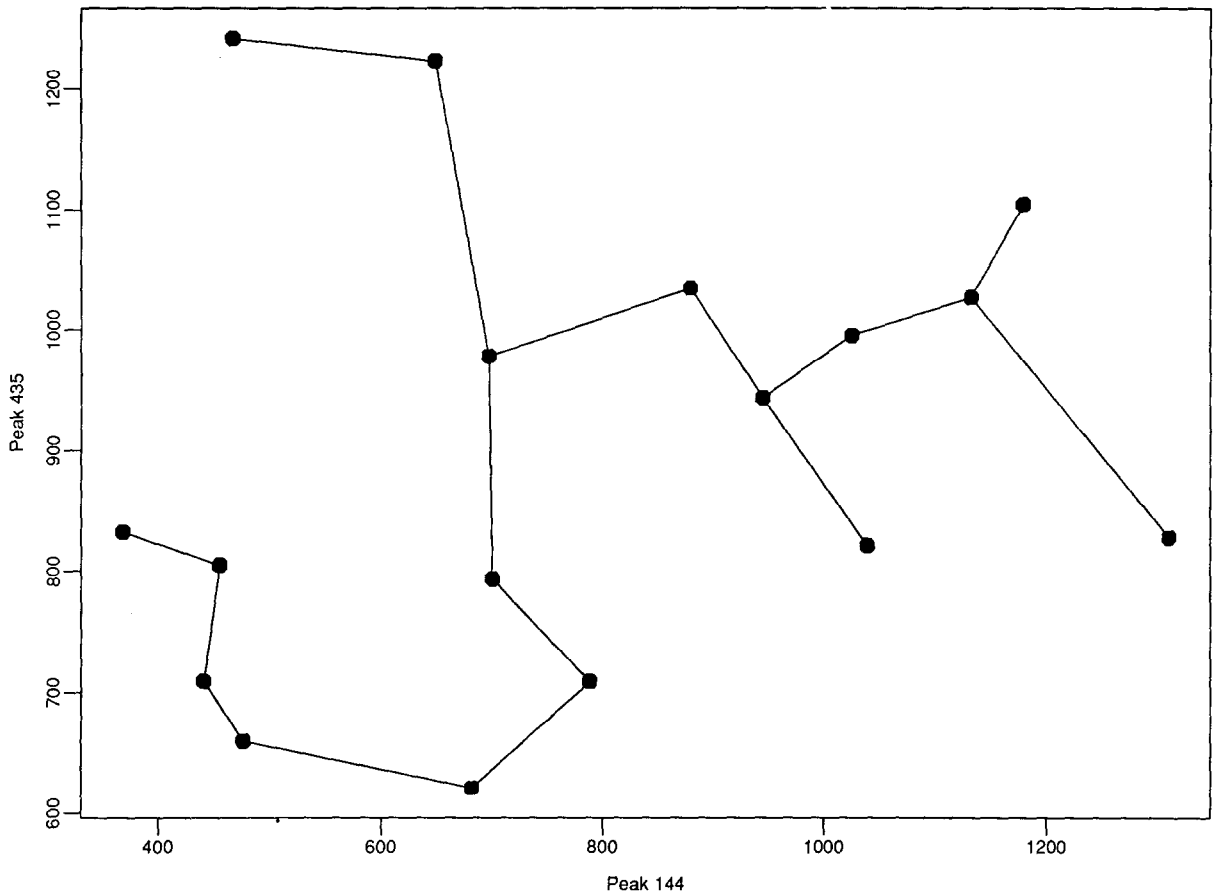

Fig. 14. Minimum spanning tree for 17 wheat samples, based on two minor HPLC peaks.

cluster to be within 200 units of its nearest neighbor we obtain four clusters: a singlet containing only variety 11, a doublet containing varieties 6 and 7, a four-point cluster of varieties 1–4, and a large cluster of all other varieties in the sample.

Wrigley [173] discussed a subtle issue of interpretation in relating clusters to baking quality due to the observational nature of the data. Varieties in the same cluster have similar electrophoresis patterns and may have similar pedigrees as well. Because of the observational nature of the data, it is possible that any association between the clustering and baking properties is due to other properties associated with the pedigrees of the wheats rather than features reflected by electrophoresis. With these kinds of studies one hopes to see a confirmation (or refutation) of initial results as more data are compiled.

The single linkage clustering tree is an empirical taxonomy of varieties. Different choices of peaks may produce very different trees, and one would hope to achieve some stability by including more peaks. Further work is needed to determine appropriate distance measures for the types of data seen in wheat studies. A promising approach was discussed by Marshall *et al.* [174] in analyzing HPLC chromatograms of urine proteins in which each chromatogram is treated as a continuous signal. In defining an L_2 type distance they incorporated time-dependent location shift functions, to account for misalignment of peaks, and time-dependent weights. Alignment is also an issue in wheat studies.

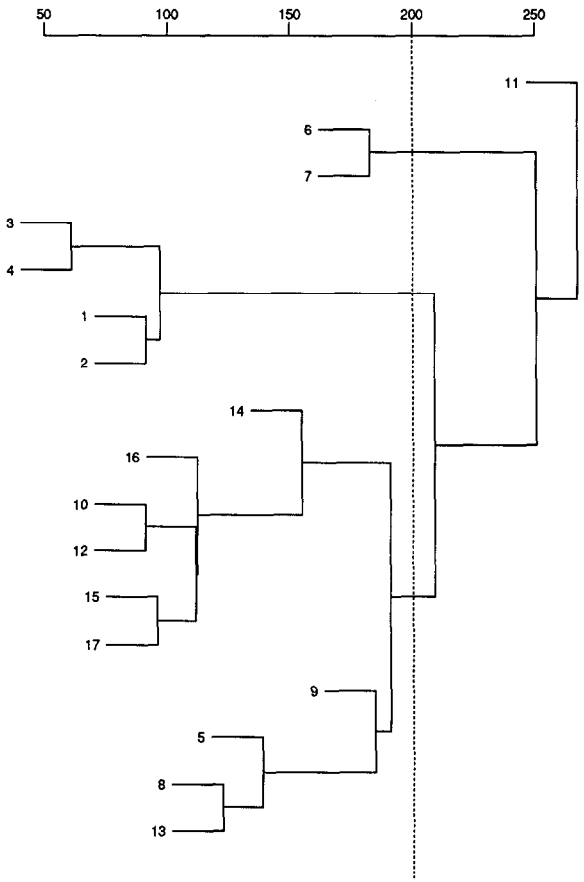

Fig. 15. Single linkage clustering tree for 17 wheat samples, based on two minor HPLC peaks.

Column performance can vary over time, causing shifts in retention times. Sapirstein *et al.* [141] discussed a piecewise linear location shift method using several reference peaks from a concurrent standard chromatogram.

Despite these methodological questions, single linkage clustering can separate known groups such as durum wheats, and the MST appears to be a useful tool for plotting multivariate data.

4.4. Principal component methods

PCA has a history dating at least to Pearson's 1901 article [175] "On lines and planes of closest fit to systems of points in space". The idea is to project multivariate data onto lower dimensional planes that are closest on average to the points in the higher dimensional space. At the first step, all points are projected onto a line selected to be as close as possible, on average, to points in the multivariate data set. It turns out to be equivalent to projecting the points on a line selected to maximize the sample variance of the projections along the line [176]. One then selects a second line, constrained to be orthogonal to the first, such that sample variance along the line is maximized. Then a third line orthogonal to the first two lines is selected, again maximizing sample variance. The process may repeat until the number of axes is equal to the smaller of the number of objects and the number of variables. In fact, basis vectors along the PCA axes and the corresponding variances are usually computed all at once, either by a singular value decomposition of the data matrix or by a spectral decomposition of the sample covariance matrix [177].

PCA is sensitive to scaling of different variables, so it is common to standardize variables first, subtracting the componentwise sample mean and dividing by the componentwise standard deviation. This is equivalent to replacing the sample covariance matrix by the sample correlation matrix in the analysis.

Often the first few components of the rotated data contain most of the variation. Scatter plots of these components can be useful for detecting clusters, patterns and outliers in the data. For instance, Fig. 16 is a scatter plot of the first two components from PCA of 112 durum wheats [178]. The raw data contain nine quality measurements of each sample. The scatter plot shows how group 42 and group 45 samples separate in the plane defined by the first two rotated components, indicating substantial differences in quality. Interestingly, the best separating plane is diagonal to the axes, and neither component by itself would be effective for separating the groups.

The basis vectors themselves are of interest because they indicate which variables account for the most variation. Fig. 17 shows superimposed RP-HPLC chromatograms for 12 varieties of hard red spring wheat. Fig. 18 shows the first two basis vectors determined by PCA of the vectors obtained by sampling the chromatograms every 5 s [164]. The first basis vector is primarily a difference across a set of highly variable peaks with retention times be-

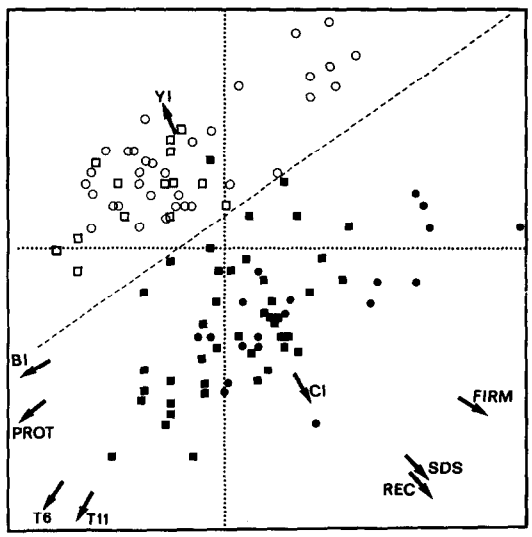

Fig. 16. First two components from PCA of 112 durum wheat samples using 9 quality tests: spaghetti yellow index (YI), spaghetti brown index (BI), protein content (PROT), minimum cooking time + 6 min (T6), minimum cooking time + 11 min (T11), microdisks cooking index (CI), gluten elastic recovery (REC), sedimentation volume (SED), and gluten firmness (FIRM). Open squares are type 42 winter wheats; open circles are type 42 spring wheats; solid squares are type 45 winter wheats, solid circles are type 45 spring wheats. From Autran *et al.* [178].

tween 27 and 28 min. The second basis vector indicates where the variation orthogonal to the first basis vector occurs. It weights on various retention times with no single dominant site. Note that PCA does not require peak identification; it can be applied directly to the discretized curves.

Autran *et al.* [178] made clever use of PCA in connection with correlation analysis of quality measurements. Autran and Galterio [179] used a similar method to study correlations among electrophoresis bands. Both articles provided correlation matrices and pairwise tests for zero correlation among the variables. Care is needed in interpreting significance tests, because multiple testing on the same data increases the experimentwise false positive rate, often drastically. Plotting the first two components from PCA on the correlation matrices as in Fig. 16 provided useful insight into clumping and patterns in the data.

PCA is commonly used with regression analysis when there are many potential regression variables. Principal component regression is a composite methodology in which the response variable is regressed on principal component projections of the regression variables [176]. The constructed regression variables are mutually orthogonal, which implies that the corresponding regression parameter estimates are uncorrelated and simplifies the interpretation. Jolliffe [176] and Martens and Naes [180] discussed variable selection in principal component regression, the point being that components accounting for the most variation among the regression variables need not be the best predictors of the response variable.

4.5. Multiple linear regression and peak selection

Various researchers have used linear regression analysis to model the dependence of wheat quality

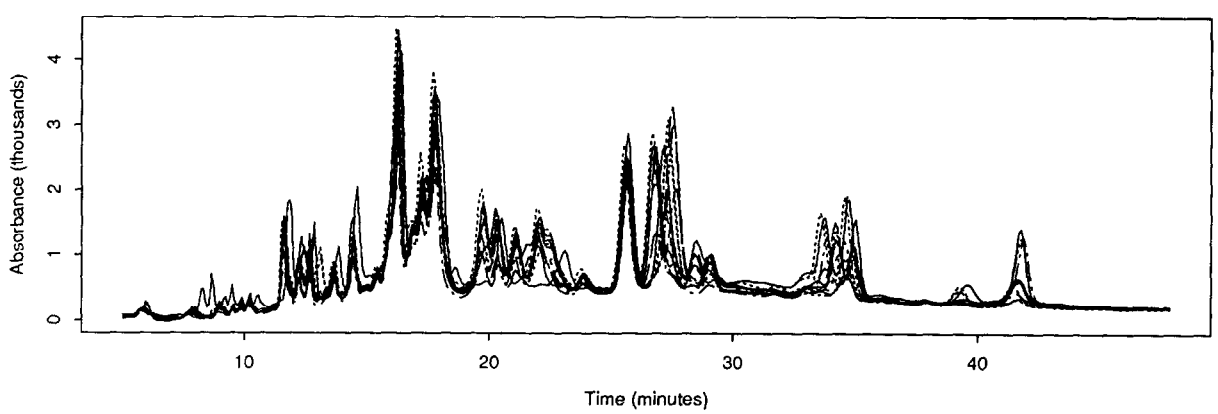

Fig. 17. RP-HPLC chromatograms for 12 varieties of hard red spring wheat. From Simpson *et al.* [164].

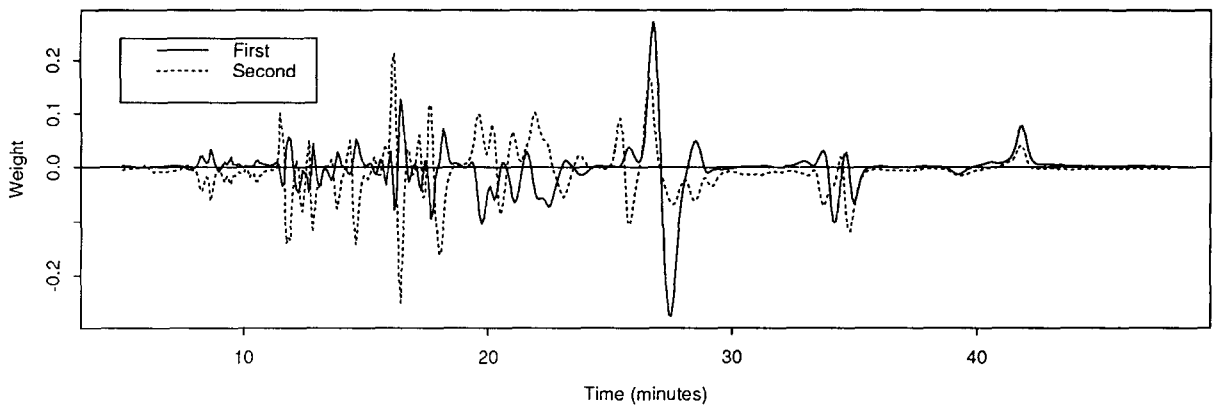

Fig. 18. First two basis vectors from PCA of 12 varieties of hard red spring wheat. From Simpson *et al.* [164].

on bands or peaks from electrophoresis and chromatography. Considerable effort goes into the identification and standardization of peaks, and different methods have been used. Ng and Bushuk [181] defined binary regression variables indicating presence or absence of specific glutenin electrophoretic bands. Then they used stepwise regression to select variables to include for linear prediction of a baking strength index. Scanlon *et al.* [182] defined quantitative regression variables by determining areas under consecutive regions of RP-HPLC chromatograms. They used stepwise variable selection to develop a linear regression model for predicting dough extensibility. Careful treatment of retention times is necessary, and some kind of normalization appears helpful. They noted that some regions may contain more than one peak, which might cause some effects to be masked. Primard *et al.* [163] and Van Lonkhuijsen *et al.* [162] used related approaches. Automating peak identification for regression analysis appears to be a challenging problem.

Parameter estimates from these studies are often poorly determined despite apparent statistical significance. In a thorough investigation Scanlon *et al.* [182] observed sign changes in parameter estimates and quite different sets of predictor variables between two replications of stepwise variable selection. This illustrates one difficulty in interpreting models found by variable selection procedures: the statistical significance of t-tests after variable selection can be grossly optimistic and should be given little credence. There is, for instance, an example due to Freedman [183] in which independent standard normal random variables generated by computer were used as responses and potential predictors in stepwise regression. Despite the independence of all the observations, a model was found that included several variables having significant t-tests and a high value of R^2, the coefficient of determination.

Another reason for poorly determined parameter estimates might be the parametrization. Suppose that the peak identification problem has been solved and the chromatogram or electrophoregram has been subdivided into peaks and normalized so that total area is 100%. Suppose p peaks are included in the model, and their relative areas are X_1, $X_2, \ldots, X_p$. The usual way to parameterize the linear regression model is to write

$$Y = a + b_1X_1 + b_2X_2 + \ldots + b_pX_p + e$$

where $a, b_1, \ldots, b_p$ are regression parameters and e is random with mean zero and unknown variance. Here the slope estimates are incremental effects of specific peaks relative to all other peaks that have been left out of the model. Thus the interpretation of all the regression parameters changes if we remove or add a peak to the regression model. One way around this problem is to reparameterize the model. Since the total area is 100%, we can replace the constant (intercept) by the constructed variable

$$X_0 = 100 - X_1 - X_2 - \ldots - X_p$$

to obtain the equivalent model

$$Y = b_0X_0 + b_1X_1 + \ldots + b_pX_p + e$$

In the latter expression the parameters reflect the degree to which the response is due to different peaks regardless of which other peaks are in the model. The two forms of the linear model are equivalent in that they yield identical predictions and sums of squared residuals, but parameters in the second form may be more stable when the chromatogram is normalized. If the stepwise search is used the change in parametrization should be done after variable selection, because X_0 cannot be selected independently of other variables in the model. Simpson *et al.* [184] used this parametrization to improve interpretation of regression parameters in an example involving attribution of pollution in rivers to different uses of the surrounding land, a scientifically distinct but statistically related problem.

The predictions in multiple linear regression are usually better determined than the individual parameter estimates. The main concern in the case of prediction is to avoid extrapolating outside the range of regression variables used to develop the model. Non-linearity and other problems with the model can easily crop up outside the range of experimentation where little information is available on the nature of the response [167].

4.6. Partial least squares

PLS is a relatively new method often used in chemometric applications when there are many possible variables and a method for predicting or calibrating is desired [185,186]. The method has features in common with PCA and regression analysis and is meant to handle problems with multiple predictors and responses, as well as simpler problems with one response and many predictors. The latter case is easiest to understand.

Suppose we wish to relate the response Y to regression variables $X1$, $X2$, ..., Xp. For instance, Y might be a quality measurement and the X's might be peak areas or heights. Alternatively, following Mosleth and Uhlen [187], the X-variables might simply be absorbances or densities over a grid of p retention times. Often p is larger than the sample size n, so some kind of variable selection or combination is needed to fit a regression model.

PLS regression first selects a unit vector u such that the projections of the vectors of X-variables on u have maximal covariance with the Y-values [188]. The projections of the X-vectors on u are called X-scores, and they are used as regression variables in place of the raw X data. One can iterate the process by using residuals from the regression of Y-values on the scores and residuals from projections of the X-vectors on u to select further basis vectors and scores orthogonal to those previously selected. It is usually suggested to use a cross-validation method [189] to determine the number of basis vectors ($< n$) to select.

Mosleth and Uhlen [187] reported on the use of PLS regression to predict Zeleny sedimentation from the electrophoresis pattern. A novel feature of this study was the direct use of the discretized densitometry trace in the calibration, made possible by the way in which PLS handles high dimensional data. They eliminated the need for peak demarcation except for identification of three reference peaks for alignment purposes.

PLS has proven most useful in contexts where empirical prediction equations are desired, and there is little interest in isolating the effects of different variables in the regression. A typical example is determination of the concentration of an analyte based on a spectrum with relatively broad peaks [190]. The usual confidence intervals from linear regression theory do not apply to PLS regression, because the constructed regression variables (the X-scores) depend also on Y. The situation is similar to that in stepwise regression, where significance levels of t-tests and the coefficient of determination are inflated due to selection of variables in the regression. In certain cases cross-validation or other resampling methods such as the bootstrap [191] provide a means for computing significance levels and variance estimates. Martens and Naes [180] provided an overview of PLS and other multivariate calibration methods and many further references.

Although there have been many reports of success using PLS, it is important to be aware of limitations of the method. There is a tendency to treat X- and Y-scores as surrogate data, plotting them on scatter plots, reporting coefficients of determination or percentages of explained variation from the regression, and so on. However, X- and Y-scores have very different properties from the usual data in regression due to their dependence on both X and Y. In particular, the slope of the PLS regression is always positive, and scatter plots and coefficients of

determination often overstate the strength of the regression relationship. An example by Simpson *et al.* [164] considered one-factor PLS regression to relate quality measurements to RP-HPLC chromatograms of 12 hard red spring wheats. Significance levels were studied for the naive *F*-test, which is in one-to-one correspondence with the coefficient of determination (*R*-squared). It was found by Monte Carlo simulation that nominal significance levels of the *F*-test were too small by a factor of four. Although correct significance levels are easily computed by simulation methods, the example demonstrates that the visual impression created by scatter plots of PLS scores can be misleading.

Another limitation of the method relates to the way in which the PLS covariance criterion combines variation with correlation (covariance = variance × correlation). This makes it difficult to interpret relative weights of different variables in the PLS basis vectors: weight may be large due to large variance and modest correlation, or strong correlation and modest variation. Further discussion and examples were given in ref. 164. If we wish identify strong quality peaks, simpler procedures such as stepwise regression or principal component regression with variable selection may be more effective.

4.7. Confounding and the value of follow-up studies

Wheat varieties are often selected for study based on availability. Thus conclusions developed from a particular dataset may not generalize to the population of varieties that could potentially be produced. Often the best one can hope for is operational success: if certain peaks are flagged as important in a study, one relies on follow-up studies to determine whether the effect persists in other varieties. If the goal is to produce wheat with desired properties, then one uses the study results to indicate directions for breeding experiments to improve the quality of wheat. If results confirm expectations, operational success has been achieved. If not, that is also valuable information pointing to possible selection bias or confounded effects in the original study.

The complexity of wheat proteins poses an added difficulty in the interpretation because the number of peaks is usually larger than the number of varieties in the study. To illustrate, suppose we have 20 varieties and for each we record heights of 100 peaks. The 100-dimensional vectors of peak heights for the different varieties necessarily lie in a 20-dimensional space of vectors. From the data at hand we can identify no more than twenty distinct combinations of peaks to "explain" baking qualities of the wheats regardless of whether we use multiple linear regression with variable selection, or a latent effects approach such as principal component regression or PLS regression. In the simplest case two peaks may be nearly perfectly correlated in the sample solely because of varieties selected. There is no way to isolate effects of these two peaks without more data to break the correlation. More generally there will always be linear combinations of peak heights that are perfectly correlated with each other if the number of peaks exceeds the number of varieties.

In the statistical design of experiments this kind of equivalence between variables is known as confounding [192]. When confounding is present one is forced to make modeling assumptions to make further progress. A typical assumption is that certain variables are null, *e.g.*, that higher order interactions are absent or negligible in a factorial experiment. The variable selection approach in multiple linear regression of wheat quality on peaks assumes that only few peaks are associated with wheat quality, whereas the PLS approach assumes a small number of combinations of peaks bear a relationship with wheat quality. In either case, however, selection of variables is dependent on the data at hand, and follow-up studies are necessary to determine if indicated peaks are indeed important.

5. CONCLUSIONS AND PERSPECTIVES

The last decade has seen the development and refinement of excellent electrophoresis and chromatography techniques and the use of these methods to characterize heterogeneous wheat proteins. Separations are based on size, charge, surface hydrophobicity, or combinations of these traits. Depending on which methods are used, the analyst can balance analysis time, number of samples analyzed, and resolution. Undoubtedly further improvements in speed, sensitivity, reproducibility, and automation will occur, especially in techniques such as narrow-bore HPLC and capillary electrophoresis.

The remarkable separations of gluten proteins

now possibly indicate that data analysis is as important and challenging as the separations themselves. Data quickly and reliably captured contain a wealth of qualitative and quantitative information. Until recently, only obvious relationships were discovered and used. Now, investigators are beginning to use many innovative statistical and computer programs to interpret data, with valuable and sometimes surprising consequences. Careful examination of these methods shows their value, but emphasizes that caution and proper use is necessary to avoid anomalous results. We anticipate that further work will make proper combinations of analytical methods and data analyses more apparent, enhancing the use of wheat protein chromatography and electrophoresis for breeding, marketing, processing, and quality control.

REFERENCES

1 N. A. Beccari, *De Frumento; De Bononiensi Scientarium et Artium*, Vol. 2, Part 1, Instituto Atque Academia Commentarii, 1745, pp. 122–127.
2 T. B. Osborne, *The Proteins of the Wheat Kernel; Publ. No. 84*, Carnegie Institute, Washington, DC, 1907.
3 J. A. Bietz, in Y. Pomeranz (Editor), *Advances in Cereal Science and Technology*, Vol. 8, American Association of Cereal Chemists, St. Paul, MN, 1986, pp. 105–170.
4 C. W. Wrigley and J. A. Bietz, in Y. Pomeranz (Editor), *Wheat: Chemistry and Technology*, Vol. 1, American Association of Cereal Chemists, St. Paul, MN, 1988, pp. 159–275.
5 J. A. Bietz, in K. M. Gooding and F. E. Regnier (Editors), *HPLC of Biological Macromolecules —Methods and Applications*, Marcel Dekker, New York, 1990, pp. 429–455.
6 J. A. Bietz, in Y. Pomeranz (Editor), *Wheat is Unique*, American Association of Cereal Chemists, St. Paul, MN, 1989, pp. 303–318.
7 J. A. Bietz, in J. P. Cherry and R. A. Barford (Editors), *Methods for Protein Analysis*, American Oil Chemists' Society, Champaign, IL, 1988, pp. 109–141.
8 J. A. Bietz, in E. G. Heyne (Editor), *Wheat and Wheat Improvement*, American Society of Agronomy, Madison, WI, 2nd ed., 1987, pp. 215–241.
9 D. D. Kasarda, in Y. Pomeranz (Editor), *Wheat is Unique*, American Association of Cereal Chemists, St. Paul, MN, 1989, pp. 277–302.
10 W. Bushuk and F. MacRitchie, in R. D. Phillips and J. W. Finley (Editors), *Food Quality and the Effects of Processing*, Marcel Dekker, New York, 1989, pp. 345–369.
11 R. C. Hoseney and D. E. Rogers, *Crit. Rev. Food Sci. Nutr.*, 29 (1990) 73–93.
12 W. Bushuk and R. Tkachuk (Editors), *Gluten Proteins 1990*, American Association of Cereal Chemists, St. Paul, MN, 1991.
13 R. W. Jones, N. W. Taylor and F. R. Senti, *Arch. Biochem. Biophys.*, 84 (1959) 363–376.
14 O. Smithies, *Biochem. J.*, 61 (1955) 629–641.
15 G. A. H. Elton and J. A. D. Ewart, *Nature (London)*, 187 (1960) 600–601.
16 G. A. H. Elton and J. A. D. Ewart, *J. Sci. Food Agric.*, 13 (1962) 62–72.
17 J. H. Woychik, J. A. Boundy and R. J. Dimler, *Arch. Biochem. Biophys.*, 94 (1961) 477–482.
18 G. A. H. Elton and J. A. D. Ewart, *J. Sci. Food Agric.*, 15 (1964) 119–126.
19 C. B. Coulson and A. K. Sim, *J. Sci. Food Agric.*, 16 (1965) 458–464.
20 A. Clydesdale, S. R. Draper and E. A. Craig, *J. Natl. Inst. Agric. Bot.*, 16 (1982) 61–66.
21 C. W. Wrigley, J. C. Autran and W. Bushuk, in Y. Pomeranz (Editor), *Advances in Cereal Science and Technology*, Vol. 5, American Association of Cereal Chemists, St. Paul, MN, 1982, pp. 211–259.
22 J.-C. Autran and A. Bourdet, *Ann. Amelior. Plant*, 25 (1975) 277–301.
23 J. W. Lee, *Biochim. Biophys. Acta*, 69 (1963) 159–160.
24 W. Bushuk and R. R. Zillman, *Can. J. Plant Sci.*, 58 (1978) 505–515.
25 D. L. du Cros and C. W. Wrigley, *J. Sci. Food Agric.*, 30 (1979) 785–794.
26 D. Lafiandra and D. D. Kasarda, *Cereal Chem.*, 62 (1985) 314–319.
27 R. Tkachuk and J. Mellish, *Ann. Technol. Agric.*, 29 (1980) 207–212.
28 G. L. Lookhart, B. L. Jones, S. B. Hall and K. F. Finney, *Cereal Chem.*, 59 (1982) 178–181.
29 K. Khan, *Baker's Dig.*, 56, No. 5 (1982) 4–19.
30 K. Khan, C. E. McDonald and O. J. Banasik, *Cereal Chem.*, 60 (1983) 178–181.
31 H. D. Sapirstein and W. Bushuk, *Cereal Chem.*, 62 (1985) 372–377.
32 J. P. Ohms, *Landwirtsch. Forsch. Sonderh.*, 37 (1980) 287–294.
33 R. L. Clements, *Cereal Chem.*, 67 (1990) 264–267.
34 G. Maier and K. Wagner, *Z. Lebensm.-Unters.-Forsch.*, 170 (1980) 343–345.
35 G. L. Lookhart, D. B. Cooper and B. L. Jones, *Cereal Chem.*, 62 (1985) 19–22.
36 R. L. Clements, *Cereal Chem.*, 64 (1987) 442–448.
37 K. Khan, A. S. Hamada and J. Patek, *Cereal Chem.*, 62 (1985) 310–313.
38 C. W. Wrigley, P. J. Gore and M. H. Perry, *Electrophoresis*, 12 (1991) 384–385.
39 F. Bekes, I. L. Batey, C. W. Wrigley and P. J. Gore, in W. Bushuk and R. Tkachuk (Editors), *Gluten Proteins 1990*, American Association of Cereal Chemists, St. Paul, MN, 1991, pp. 467–475.
40 R. L. Clements, *Cereal Chem.*, 65 (1988) 150–152.
41 R. J. Cooke, *J. Natl. Inst. Agric. Bot.*, 17 (1987) 273–281.
42 R. J. Cooke and A. G. Morgan, *J. Natl. Inst. Agric. Bot.*, 17 (1986) 169–178.
43 H. D. Sapirstein and W. Bushuk, *Cereal Chem.*, 62 (1985) 377–392.

44 H. D. Sapirstein and W. Bushuk, *Seed Sci. Technol.*, 14 (1986) 489–517.
45 C. W. Wrigley, *Sci. Tools*, 15 (1968) 17–23.
46 C. W. Wrigley, *J. Chromatogr.*, 36 (1968) 362–365.
47 H. Windemann, U. Müller and E. Baumgartner, *Z. Lebensm.-Unters.-Forsch.*, 153 (1973) 17–22.
48 P. G. Righetti and A. B. Bosisio, *Electrophoresis*, 2 (1981) 65–75.
49 G. Branlard and B. Picard, in W. Bushuk and R. Tkachuk (Editors), *Gluten Proteins 1990*, American Association of Cereal Chemists, St. Paul, MN, 1991, pp. 389–396.
50 A. Curioni, A. Dal Belin Peruffo and N. E. Pogna, *Electrophoresis*, 11 (1990) 462–467.
51 P. K. W. Ng, E. Slominski, W. J. Johnson and W. Bushuk, *Cereal Chem.*, 66 (1989) 536–537.
52 U. Laemmli, *Nature (London)*, 227 (1970) 680–685.
53 J. A. Bietz and J. S. Wall, *Cereal Chem.*, 49 (1972) 416–430.
54 R. Bollini, L. A. Manzocchi, M. Cattaneo and B. Borghi, *Z. Acker. Pflanzenbau*, 150 (1981) 71–79.
55 J. A. Bietz, K. W. Shepherd and J. S. Wall, *Cereal Chem.*, 52 (1975) 513–532.
56 P. I. Payne, L. M. Holt and C. N. Law, *Theor. Appl. Genet.*, 60 (1981) 229–236.
57 G. J. Lawrence and K. S. Shepherd, *Aust. J. Biol. Sci.*, 33 (1980) 221–233.
58 G. Galili and M. Feldman, *Theor. Appl. Genet.*, 66 (1983) 77–86.
59 P. R. Shewry, A. J. Faulks, H. M. Pratt and B. J. Miflin, *J. Sci. Food Agric.*, 29 (1978) 847–849.
60 D. L. du Cros, G. J. Lawrence, D. M. Miskelly and C. W. Wrigley, *Systematic Identification of Australian Wheat Varieties by Laboratory Methods; Technical Publication No. 7*, CSIRO Wheat Research Unit, North Ryde, Australia, 1980.
61 P. K. W. Ng and W. Bushuk, *Cereal Chem.*, 64 (1987) 324–327.
62 T. Burnouf and R. Bouriquet, *Theor. Appl. Genet.*, 58 (1980) 107–111.
63 P. I. Payne, C. N. Law and E. E. Mudd, *Theor. Appl. Genet.*, 58 (1980) 113–120.
64 P. I. Payne and G. J. Lawrence, *Cereal Res. Commun.*, 11 (1983) 29–35.
65 J. G. Fullington, E. W. Cole and D. D. Kasarda, *J. Sci. Food Agric.*, 31 (1980) 43–53.
66 N. K. Singh and K. W. Shepherd, *Theor. Appl. Genet.*, 75 (1988) 628–641.
67 R. B. Gupta and K. W. Shepherd, *Theor. Appl. Genet.*, 80 (1990) 65–74.
68 D. Khelifi and G. Branlard, *J. Cereal Sci.*, 13 (1991) 41–47.
69 R. Graybosch, C. J. Peterson, S. Primard and J.-H. Lee, in W. Bushuk and R. Tkachuk (Editors), *Gluten Proteins 1990*, American Association of Cereal Chemists, St. Paul, MN, 1991, pp. 581–594.
70 T. Burnouf and J. A. Bietz, *Cereal Chem.*, 66 (1989) 121–127.
71 R. A. Graybosch and R. Morris, *J. Cereal Sci.*, 11 (1990) 201–212.
72 B. A. Marchylo, *Can. J. Plant Sci.*, 67 (1987) 945–952.
73 B. A. Marchylo, K. A. Handel and V. J. Mellish, *Cereal Chem.*, 66 (1989) 186–192.
74 O. M. Lukow and K. M. Kidd, in W. Bushuk and R. Tkachuk (Editors), *Gluten Proteins 1990*, American Association of Cereal Chemists, St. Paul, MN, 1991, pp. 491–497.
75 P. Kolster and W. M. J. van Gelder, in W. Bushuk and R. Tkachuk (Editors), *Gluten Proteins 1990*, American Association of Cereal Chemists, St. Paul, MN, 1991, pp. 349–361.
76 A. Curioni, A. Dal Belin Peruffo and N. E. Pogna, *Cereal Chem.*, 66 (1989) 133–135.
77 C. W. Wrigley, *Biochem. Genet.*, 4 (1970) 509–516.
78 C. W. Wrigley and K. W. Shepherd, *Ann. N.Y. Acad. Sci.*, 209 (1973) 154–162.
79 C. Aragoncillo, M. A. Rodriguez-Loperena, P. Carbonero and F. Garcia-Olmedo, *Theor. Appl. Genet.*, 45 (1973) 322–326.
80 K. R. F. Hussein and H. Stegemann, *Z. Acker Pflanzenbau*, 146 (1978) 68–78.
81 D. Lafiandra, D. D. Kasarda and R. Morris, *Theor. Appl. Genet.*, 68 (1984) 531–539.
82 D. K. Mecham, D. D. Kasarda and C. O. Qualset, *Biochem. Genet.*, 16 (1978) 831–853.
83 P. H. O'Farrell, *J. Biol. Chem.*, 250 (1975) 4007–4021.
84 P. I. Payne, L. M. Holt, M. G. Jarvis and E. A. Jackson, *Cereal Chem.*, 62 (1985) 319–326.
85 M.-G. Lei and G. R. Reeck, *Cereal Chem.*, 63 (1986) 111–116.
86 N. G. Anderson, S. L. Tollaksen, F. H. Pascoe and L. Anderson, *Crop Sci.*, 25 (1985) 667–674.
87 E. V. Metakovsky, A. Y. Novoselskaya and A. A. Sozinov, *Theor. Appl. Genet.*, 69 (1984) 31–37.
88 C. W. Wrigley, *Qual. Plant. Plant Foods Hum. Nutr.*, 31 (1982) 205–227.
89 H. P. Tao and D. D. Kasarda, *J. Exp. Bot.*, 40 (1989) 1015–1020.
90 D. D. Kasarda, H. P. Tao, P. K. Evans, A. E. Adalsteins and S. W. Yuen, *J. Exp. Bot.*, 39 (1988) 899–906.
91 R. Tkachuk and V. J. Mellish, in R. Lasztity and F. Bekes (Editors), *Proc. 3rd Int. Workshop on Gluten Proteins*, World Scientific, Singapore, 1987, pp. 111–124.
92 V. J. Mellish and R. Tkachuk, in W. Bushuk and R. Tkachuk (Editors), *Gluten Proteins 1990*, American Association of Cereal Chemists, St. Paul, MN, 1991, pp. 482–490.
93 D. A. Dougherty, R. L. Wehling, M. G. Zeece and J. E. Partridge, *Cereal Chem.*, 67 (1990) 564–569.
94 P. D. Grossman, J. C. Colburn, H. H. Lauer, R. G. Nielsen, R. M. Riggin, G. S. Sittampalam and E. C. Rickard, *Anal. Chem.*, 61 (1989) 1186–1194.
95 M. J. Gordon, K.-J. Lee, A. A. Arias and R. N. Zare, *Anal. Chem.*, 63 (1991) 69–72.
96 B. L. Karger, A. S. Cohen and A. Guttman, *J. Chromatogr.*, 492 (1989) 585–614.
97 M. Zhu, R. Rodriguez, D. Hansen and T. Wehr, *J. Chromatogr.*, 516 (1990) 123–131.
98 R. W. Jones, G. E. Babcock, N. W. Taylor and R. J. Dimler, *Cereal Chem.*, 40 (1963) 409–414.
99 A. C. Beckwith, H. C. Nielsen, J. S. Wall and F. R. Huebner, *Cereal Chem.*, 43 (1966) 14–28.
100 F. R. Huebner and J. S. Wall, *Cereal Chem.*, 51 (1974) 228–240.
101 O. B. Meredith and J. J. Wren, *Cereal Chem.*, 43 (1966) 169–186.

102 O. B. Meredith and J. J. Wren, *J. Sci. Food Agric.*, 20 (1969) 235–237.
103 C. W. Wrigley, *Cereal Sci. Today*, 17 (1972) 370–375.
104 F. R. Huebner and J. S. Wall, *Cereal Chem.*, 53 (1976) 258–269.
105 P. I. Payne and K. G. Corfield, *Planta*, 145 (1979) 83–88.
106 R. C. Bottomley, H. F. Kerrs and J. D. Schofield, *J. Sci. Food Agric.*, 33 (1982) 481–491.
107 A. Graveland, P. Bosveld, W. J. Lichtendonk, H. H. E. Mooney and A. Scheepstra, *J. Sci. Food Agric.*, 33 (1982) 1117–1128.
108 J. H. Woychik, R. J. Dimler and F. R. Senti, *Arch. Biochem. Biophys.*, 91 (1960) 235–239.
109 D. H. Simmonds and D. J. Winzor, *Aust. J. Biol. Sci.*, 14 (1961) 690–699.
110 F. R. Huebner and J. S. Wall, *Cereal Chem.*, 43 (1966) 325–335.
111 C. W. Wrigley, *Aust. J. Biol. Sci.*, 18 (1965) 193–195.
112 F. R. Huebner and J. A. Rothfus, *Cereal Chem.*, 45 (1968) 242–253.
113 L. Charbonnier, *Biochim. Biophys. Acta*, 359 (1974) 142–151.
114 L. Charbonnier and J. Mosse, *J. Sci. Food Agric.*, 31 (1980) 54–61.
115 Y. Popineau and F. Pineau, *Lebensm. Wiss. Technol.*, 18 (1985) 133–135.
116 A. S. Khan, N. M. Waldron and L. C. Thiang, *J. Sci. Food Agric.*, 36 (1985) 833–838.
117 Y. Popincau and B. Godon, *C.R. Hebd. Seances Acad. Sci., Ser. D.*, 287 (1978) 1051–1054.
118 K. A. Caldwell, *J. Sci. Food Agric.*, 30 (1979) 185–196.
119 K. H. Chung and Y. Pomeranz, *Cereal Chem.*, 56 (1979) 196–201.
120 Y. Popineau and F. Pineau, *J. Cereal Sci.*, 3 (1985) 363–378.
121 Y. Popineau, *J. Cereal Sci.*, 3 (1985) 29–38.
122 F. E. Regnier and K. M. Gooding, *Anal. Biochem.*, 103 (1980) 1–25.
123 M. T. W. Hearn, F. E. Regnier and C. T. Wehr, *Am. Lab.*, 14 (1982) 18, 20, 23–24, 27–30, 35–36, 38–39.
124 C. T. Wehr, in W. S. Hancock (Editor), *Handbook of HPLC for the Separation of Amino Acids, Peptides, and Proteins*, Vol. I, CRC Press, Boca Raton, FL, 1984, pp. 31–57.
125 J. A. Bietz, *Baker's Dig.*, 58, No. 1 (1984) 15–17, 20–21, 32.
126 J. A. Bietz, *Baker's Dig.*, 58, No. 2 (1984) 22, 24.
127 J. A. Bietz, in A. Graveland and J. H. E. Moonen (Editors), *Gluten Proteins —Proceedings of the 2nd International Workshop on Gluten Proteins*, TNO, Wageningen, Netherlands, 1984, pp. 1–11.
128 J. A. Bietz, *Cereal Chem.*, 62 (1985) 201–212.
129 T. Burnouf and J. A. Bietz, *Seed Sci. Technol.*, 15 (1987) 79–99.
130 F. R. Huebner and J. A. Bietz, *J. Chromatogr.*, 327 (1985) 333–342.
131 T. Dachkevitch and J.-C. Autran, *Cereal Chem.*, 66 (1989) 448–456.
132 N. K. Singh, G. R. Donovan, I. L. Batey and F. MacRitchie, *Cereal Chem.*, 67 (1990) 150–161.
133 N. K. Singh, R. Donovan and F. MacRitchie, *Cereal Chem.*, 67 (1990) 161–170.
134 I. L. Batey, R. B. Gupta and F. MacRitchie, *Cereal Chem.*, 68 (1991) 207–209.
135 F. R. Huebner, D. D. Christianson, T. C. Nelsen and J. A. Bietz, in W. Bushuk and R. Tkachuk (Editors), *Gluten Proteins 1990*, American Association of Cereal Chemists, St. Paul, MN, 1991, pp. 145–155.
136 I. L. Batey, *J. Cereal Sci.*, 2 (1984) 241–248.
137 C. Larre, Y. Popineau and W. Loisel, *J. Cereal Sci.*, 14 (1991) 231–241.
138 J. A. Bietz, *J. Chromatogr.*, 255 (1983) 219–238.
139 J. A. Bietz and L. A. Cobb, *Cereal Chem.*, 62 (1985) 332–339.
140 B. A. Marchylo and J. E. Kruger, *Cereal Chem.*, 65 (1988) 192–198.
141 H. D. Sapirstein, M. G. Scanlon and W. Bushuk, *J. Chromatogr.*, 469 (1989) 127–135.
142 M. G. Scanlon, H. D. Sapirstein and W. Bushuk, *Cereal Chem.*, 66 (1989) 112–116.
143 M. G. Scanlon, H. D. Sapirstein and W. Bushuk, *Cereal Chem.*, 66 (1989) 439–443.
144 K. H. Sutton, *J. Cereal Sci.*, 14 (1991) 25–34.
145 G. L. Lookhart and L. D. Albers, *Cereal Chem.*, 65 (1988) 222–227.
146 T. Burnouf and J. A. Bietz, *Theor. Appl. Genet.*, 70 (1985) 610–619.
147 H. Wieser, W. Seilmeier and H.-D. Belitz, *Cereal Chem.*, 66 (1989) 38–41.
148 J. A. Bietz, T. Burnouf, L. A. Cobb and J. S. Wall, *Cereal Chem.*, 61 (1984) 129–135.
149 J. E. Kruger and B. A. Marchylo, *Can. J. Plant Sci.*, 65 (1985) 285–298.
150 B. A. Marchylo, J. E. Kruger and D. W. Hatcher, *Cereal Chem.*, 67 (1990) 372–376.
151 G. L. Lookhart and Y. Pomeranz, *Cereal Chem.*, 62 (1985) 227–229.
152 G. L. Lookhart, R. Graybosch, J. Peterson and A. Lukaszewski, in W. Bushuk and R. Tkachuk (Editors), *Gluten Proteins 1990*, American Association of Cereal Chemists, St. Paul, MN, 1991, pp. 688–706.
153 K. H. Sutton, R. L. Hay and W. B. Griffin, *J. Cereal Sci.*, 10 (1989) 113–121.
154 G. L. Lookhart, L. D. Albers and J. A. Bietz, *Cereal Chem.*, 63 (1986) 497–500.
155 S. Endo, K. Okada, S. Nagao and B. L. D'Appolonia, *Cereal Chem.*, 67 (1990) 480–485.
156 S. Endo, K. Okada, S. Nagao and B. L. D'Appolonia, *Cereal Chem.*, 67 (1990) 486–489.
157 J. A. Bietz, F. R. Huebner, T. C. Nelsen, D. G. Simpson, S. Guo and J. Sacks, in W. Bushuk and R. Tkachuk (Editors), *Gluten Proteins 1990*, American Association of Cereal Chemists, St. Paul, MN, 1991, pp. 420–432.
158 J. A. Bietz and T. Burnouf, *Theor. Appl. Genet.*, 70 (1985) 599–608.
159 W. Seilmeier, H.-D. Belitz and H. Wieser, in W. Bushuk and R. Tkachuk (Editors), *Gluten Proteins 1990*, American Association of Cereal Chemists, St. Paul, MN, 1991, p. 287–295.
160 T. Burnouf and J. A. Bietz, *J. Cereal Sci.*, 2 (1984) 3–14.
161 F. R. Huebner and J. A. Bietz, *J. Cereal Sci.*, 4 (1986) 379–388.

162 H. J. van Lonkhuijsen, R. J. Hamer and C. Schreuder, *Cereal Chem.*, 69 (1992) 174–177.
163 S. Primard, R. Graybosch, C. J. Peterson and J.-H. Lee, *Cereal Chem.*, 68 (1991) 305–312.
164 D. G. Simpson, S. Guo, J. Sacks, J. A. Bietz, F. Huebner and T. Nelsen, *Chemometrics Intell. Lab. Syst.*, 10 (1991) 155–167.
165 G. L. Lookhart, B. L. Jones, D. E. Walker, S. B. Hall and D. B. Cooper, *Cereal Chem.*, 60 (1983) 111–115.
166 R. Damidaux, J. C. Autran and P. Feillet, *Cereal Foods World*, 25 (1980) 754–756.
167 S. Weisberg, *Applied Linear Regression*, Wiley, New York, 2nd ed., 1985.
168 T. S. Cox, G. L. Lookhart, D. E. Walker, L. G. Harrell, L. D. Albers and D. M. Rodgers, *Crop. Sci.*, 25 (1985) 1058–1062.
169 C. W. Wrigley, P. J. Robinson and W. T. Williams, *J. Sci. Food Agric.*, 32 (1981) 433–442.
170 D. L. du Cros, *J. Cereal Sci.*, 5 (1987) 3–12.
171 J. C. Gower and G. J. S. Ross, *Appl. Statistics*, 18 (1969) 54–64.
172 J. H. Friedman and L. C. Rafsky, *J. Am. Stat. Assoc.*, 76 (1981) 277–287.
173 C. W. Wrigley, *Ann. Technol. Agric.*, 29 (1980) 213–227.
174 R. J. Marshall, R. Turner, H. Yu and E. H. Cooper, *J. Chromatogr.*, 297 (1984) 235–244.
175 K. Pearson, *Phil. Mag., 6th Series*, 2 (1901) 559–572.
176 I. T. Jolliffe, *Principal Component Analysis*, Springer, New York, 1986.
177 J. J. Dongarra, J. R. Bunch, C. B. Moler and G. W. Stewart, *LINPACK User's Guide*, SIAM, Philadelphia, PA, 1979.
178 J. C. Autran, J. Abecassis and P. Feillet, *Cereal Chem.*, 63 (1986) 390–394.
179 J. C. Autran and G. Galterio, *J. Cereal Sci.*, 9 (1989) 179–194.
180 H. Martens and T. Naes, *Multivariate Calibration*, Wiley, New York, 1989.
181 P. K. W. Ng and W. Bushuk, *Cereal Chem.*, 65 (1988) 408–413.
182 M. G. Scanlon, P. K. W. Ng, D. E. Lawless and W. Bushuk, *Cereal Chem.*, 67 (1990) 395–399.
183 D. Freedman, *Am. Statistician*, 37 (1983) 152–155.
184 D. G. Simpson, D. Ruppert and R. J. Carroll, *J. Am. Stat. Assoc.*, 87 (1992) 439–450.
185 S. Wold, H. Martens and H. Wold, in A. Ruhe and B. Kagstrom (Editors), *Matrix Pencils, Proceedings of a Conference, Pite Hausbad, March 22–24, 1982 (Lecture Notes in Mathematics*, No. 973), Springer, Heidelberg, 1983, pp. 286–293.
186 S. Wold, C. C. Albano, W. J. Dunn, K. Esbensen, S. Hellberg, E. Johansson, W. Lindberg and M. Sjostrom, *Analusis*, 12 (1984) 477–485.
187 E. Mosleth and A. K. Uhlen, in R. Lasztity and F. Bekes (Editors), *Proc. 3rd Int. Workshop on Gluten Proteins*, World Scientific, Singapore, 1987, pp. 548–552.
188 A. Hoskuldsson, *J. Chemometrics*, 2 (1988) 211–228.
189 M. Stone, *J. Royal Statist. Soc. B*, 36 (1974) 111–147.
190 W. P. Carey and L. E. Wangen, *Chemometrics Intell. Lab. Syst.*, 10 (1991) 245–257.
191 B. Efron, *The Jackknife, the Bootstrap and Other Resampling Plans (CBMS-NSF Regional Conference Series in Applied Mathematics*. No. 38), SIAM, Philadelphia, PA, 1982.
192 G. E. P. Box, W. G. Hunter and J. S. Hunter, *Statistics for Experimenters*, Wiley, New York, 1978.