Phytochemistry, 1980, Vol. 19, pp. 1647–1651. Pergamon Press Ltd. Printed in England. 0031 9422/80/0801-1647 $02.00/0

COMPOSITIONAL RELATIONSHIPS AMONG ELECTROPHORETIC ISOLATES FROM COTTONSEED PROTEIN BODIES

EARL E. KING

Cotton Physiology and Genetics Laboratory, AR-SEA, U.S. Department of Agriculture, P.O. Box 225, Stoneville, MS 38776, U.S.A.

(*Revised received* 7 *January* 1980)

Key Word Index—*Gossypium hirsutum*; Malvaceae; cottonseed; electrophoresis, amino acid composition; isoelectric focusing; protein quality.

Abstract—Minor differences exist among polypeptides isolated from cottonseed storage globulins with respect to nutritionally-limiting amino acids, although the composition of most isolates closely resembles that of the entire storage globulin fraction. Glutamate, arginine and aspartate are the most abundant amino acids; cysteine and methionine are the least abundant. In contrast to the 10 polypeptide bands in the SDS-PAGE pattern, isoelectric focusing produced a pattern of 49 components within a range of pH 4.5–9.5. A two-dimensional electrophoretic separation indicated that most of the polypeptides resolved by SDS-PAGE are heterogeneous with respect to charge. The high ratio of isoelectric to molecular weight forms may prove advantageous for efforts to genetically improve the quality of cottonseed protein.

INTRODUCTION

Cottonseed (*Gossypium hirsutum* L.) have the potential to become an important source of protein for food use. Despite a lack of emphasis by cotton breeding programs on seed-quality characteristics, the quality of cottonseed protein as measured by amino acid analysis [1], protein efficiency ratios [2] and functional properties [3] is reasonably high.

There is apparently a modest degree of genetic variability within *G. hirsutum* germplasm for both amount and composition of whole-seed protein [4]. Storage globulins constitute *ca* 70% of the protein in the seed, but the amino acid composition of this class of protein is less desirable nutritionally than that of the metabolic, or non-storage seed protein [5]. It has been suggested that seed storage proteins are 'liberal' proteins, with fewer constraints on their structure than metabolic proteins, the structures and amino acid profiles of which are more strictly dictated by their catalytic functions [6]. Should this be the case with cottonseed, there may be reason for optimism regarding the genetic improvement of storage globulins, whether by alteration of the amino acid composition of individual proteins or by changes in the relative proportions of polymorphic subunits of a protein.

Examination of the molecular and compositional nature of the storage protein is necessary to facilitate the improvement of cottonseed protein quality. As a contribution to that end, this paper describes the nature of cottonseed storage globulins as determined by sodium dodecyl sulphate-polyacrylamide electrophoresis (SDS-PAGE), isoelectric focusing (IEF), and amino acid analysis.

RESULTS AND DISCUSSION

Fractionation of globulins from isolated cottonseed protein bodies (CSPB) by SDS-PAGE in the presence of β-mercaptoethanol (ME) produced a pattern (Fig. 1A) closely resembling that reported previously [7, 8] for putative storage proteins from cotton seeds. Under reducing conditions, the CSPB polypeptides were separated into 4 groups on the basis of molecular size distribution. The first group, consisting of bands I and II, accounts for 40% of the CSPB protein content as estimated by densitometric scans of stained gels. These bands were excised individually from the gels for amino acid analysis. Because the other bands were both more diffuse and less intense than I and II, no attempt was made to excise them individually; they were excised as groups.

Omission of ME from the SDS-PAGE sample increased the complexity of the pattern (Fig. 1B), and indicated that disulphide bonds are involved in the maintenance of CSPB protein structure [9]. Of the polypeptides present in Fig. 1A (with ME), most are also present in Fig. 1B (without ME). The exceptions are the bands of group III and two of the bands of group IV. This implies that reactive sulphydryl groups are unequally distributed among the polypeptides; the polypeptides that are present in Fig. 1A but absent from Fig. 1B may be those with the highest content of cysteine residues.

Some of the oligomers of Fig. 1B are identifiable; the three bands with MW > 53 000 are apparently blends of the dimers, trimers and tetramers of both polypeptides I and II. When I and II were excised individually from gels and re-electrophoresed separately by SDS-PAGE without ME (Figs. 1C and 1D), each produced a 4-band pattern closely resembling the upper portion of Fig. 1B. The mobilities of the oligomers of I and II are so similar that when I and II are electrophoresed together without ME, the oligomers of each migrate so closely together that their separate identities are masked.

A small amount of water-soluble protein was recovered from CSPB. Water-soluble proteins (albumins) have been

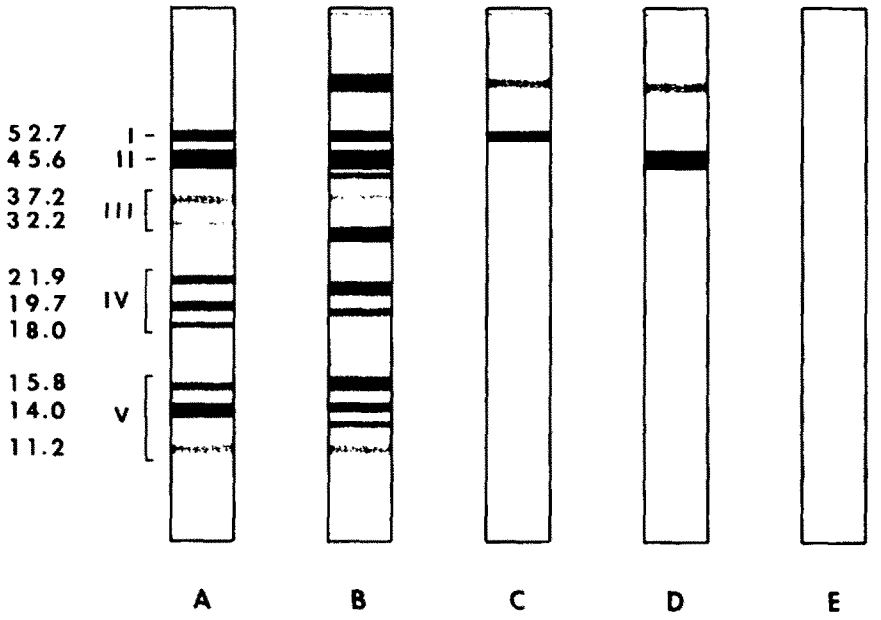

Fig. 1. SDS-PAGE patterns (10% gel) of protein from CSPB. Roman numerals designate bands or groups of bands that were excised and used for amino acid analysis. Numbers to the left of group designations denote apparent MW (daltons $\times 10^{-3}$) of each band. (A): 0.015 N NaOH extract, electrophoresis sample contained 1% ME. (B): 0.015 N NaOH extract, electrophoresis sample lacked ME. (C): Band I only excised from (A) and re-electrophoresed without ME. (D): Band II only excised from (A) and re-electrophoresed without ME. (E): Water extract, electrophoresis sample contained 1% ME.

isolated from protein bodies of castor beans (*Ricinus communis* L.) and found to constitute *ca* 40% of the protein in those organelles [10]. They are degraded during germination; the inference is that they are storage proteins [10]. The albumins of CSPB, on the other hand, constitute less than 5% of the CSPB protein, based on nitrogen analysis, and their function is unknown. They produce the SDS-PAGE pattern shown in Fig. 1E. Although the cottonseed system is not well defined, it is apparently dissimilar to that of castor beans.

Indirect evidence [11] suggests that water-soluble storage proteins with apparent MW < 20 000 are present in CSPB and that they are identical to previously-reported [12] cottonseed allergens. Protease and phosphatase activities are associated with CSPB from ungerminated seeds [13]; the protease activity is relatively low in dry seeds (unpublished data, this laboratory). It is consistent with these findings that small amounts of albumins should be found in CSPB. It is possible, of course, that the albumins are not from within the CSPB but instead are cytoplasmic contaminants that adhered to the organelles during isolation. However, the CSPB prepared for this study were washed twice each with glycerol, 0.25 M sucrose and water before they were extracted. Although exhaustive washing does not guarantee freedom from contaminating cytoplasmic proteins, the persistence of albumins after this procedure increases the probability that they are located within the CSPB. Further experiments are necessary, however, to unequivocally identify the *in vivo* localization and define the function of these proteins.

Amino acid analyses reveal many compositional similarities among the SDS-PAGE fractions (Table 1). All, for instance, are rich in glutamate, arginine and aspartate, and all have relatively low levels of cysteine and methionine. This relationship is common to seed proteins [6, 14]; in cotton it extends beyond the SDS-PAGE fractions and the whole CSPB composition to the level of the whole seed [15, 16]. This is explained not only by the preponderance of storage (CSPB) protein in the seed [5], but also by the composition of the non-storage cottonseed proteins; they also exhibit an imbalance, albeit less extreme, among these same amino acids [1].

Compared with the FAO Provisional Pattern of essential amino acids [17], the sulphoamino acids and isoleucine are most limiting in cottonseed [4]. In addition to these intrinsic deficiencies, the reaction of gossypol with lysine during processing of glanded cottonseed diminishes the biological availability of this amino acid [18]. Increased levels of each of these amino acids should therefore receive major emphasis in any program of cottonseed protein quality improvement.

The amino acid data in Table 1 indicate some problems associated with considering the SDS-PAGE fractions as basic units for manipulation in attempts to improve protein quality. No single isolate is superior to whole CSPB protein in terms of essential amino acid balance. For instance, isolate V is proportionately higher in sulphoamino content than the whole CSPB protein, yet it is the lowest of all the fractions in isoleucine content and ranks midway among the fractions in lysine content. Similarly, isolate II, highest in lysine content, is lowest with respect to sulphoamino acids and midway among the fractions in isoleucine content.

The overall amino acid balance in the seed storage protein must also be considered. Although there may be fewer biological constraints on the composition of storage proteins than on enzymatic proteins, they nevertheless exist, and are important to the biological system in which they are found [19]. Arginine, for example, is a major source of reduced nitrogen for the germinating cotton seed [1]. Its distribution among the SDS-PAGE fractions

Table 1. Composition of electrophoretic isolates from cottonseed protein bodies.

Property	NaOH extract of CSPB	SDS-PAGE isolates I	II	III	IV	V
% of CSPB globulin	100.0	20.8	19.1	4.7	20.0	35.6
Apparent MW (daltons × 10^{-3})		52.7	45.6	37.2-32.2	21.9 18.0	15.8 11.2
Amino acid*						
Lys	3.10	2.93	3.70	3.03	2.64	2.99
His	2.88	3.52	3.55	0.93	0.37	2.55
Arg	12.87	4.56	7.41	26.52	24.40	9.25
Asx	9.93	10.73	11.34	8.29	9.42	11.25
Thr	3.79	3.81	3.88	3.82	3.64	3.65
Ser	5.59	7.15	6.69	5.39	4.70	6.11
Glx	23.48	25.40	22.34	17.59	17.73	29.79
Pro	3.68	5.31	4.25	4.58	3.74	2.22
Gly	4.62	4.09	4.89	4.78	4.25	7.17
Ala	4.46	4.63	4.30	3.42	4.45	4.46
½ Cys	0.39	0.58	0.31	0.31	0.63	0.32
Val	5.10	5.90	6.42	4.06	4.04	3.80
Met	1.49	0.85	0.66	1.28	1.66	2.31
Leu	6.75	7.39	6.66	5.70	6.10	6.90
Ile	3.28	3.61	3.08	3.14	3.69	2.85
Tyr	2.06	1.98	2.40	1.81	2.31	1.10
Phe	6.57	7.61	8.16	5.24	6.28	3.33

* Percentage composition; data are means of duplicate analyses. Tryptophan not determined.

(Table 1) is the most uneven of the 3 major storage amino acids (glutamate, arginine and aspartate), being 5–6 times as concentrated in isolates III and IV as it is in isolate I. Although a mutant with decreased amounts of group III or IV polypeptides would have increased lysine concentration, germination of these seeds could be affected.

Because SDS-PAGE separates polypeptides solely on the basis of MW differences, each band of a pattern may consist of more than one polypeptide with different composition but with similar MW. SDS-PAGE alone neither indicates the total number of polypeptides present nor provides any information about their compositional diversity. In contrast, IEF is theoretically able to separate proteins that differ by only a single charge [20]. The separation of CSPB proteins by IEF across a gradient of pH 3.5–10.0 is depicted in Fig. 2. Instead of the 10 bands produced by SDS-PAGE (Fig. 1A), 49 bands are visible in the IEF pattern. Charge differences among CSPB proteins are therefore more prevalent than MW differences among CSPB polypeptides. This suggests that liberality of cottonseed storage proteins may be expressed chiefly in the form of amino acid substitutions and sequence changes that have little or no influence on polypeptide chain length.

A two-dimensional separation in which SDS-PAGE was followed by IEF demonstrated that the high MW polypeptides are heterogeneous with respect to charge (Table 2). SDS-PAGE bands I and II (MW 52 700 and 45 600) were each resolved into a single series of components evenly distributed within a relatively narrow range of pH; it is likely that these components are minor charge variants of a single polypeptide type. A similar condition has been demonstrated in some peanut (*Arachis hypogaea* L.) seed proteins [21]. In contrast, the MW 21 900 and 19 700 polypeptides were each resolved into two groups of isoelectric components, suggesting a more complex level of heterogeneity. The lower MW polypeptides (MW < 16 000) were apparently lost during an intermediate equilibration step in the two-dimensional procedure, and did not appear in the second dimension.

The high (5:1) ratio of isoelectric to MW forms reported here is not unique among plant storage proteins. Similar or higher ratios have been found in zein from maize endosperm [22] and in storage proteins from potato tubers [23]. The reason(s) for conservation of

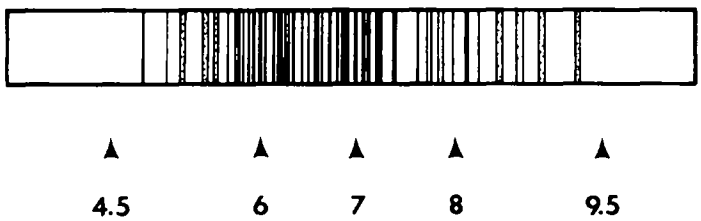

Fig. 2. IEF pattern of proteins extracted from CSPB by 6 M urea. IEF took place in 5% polyacrylamide gel plates containing 6 M urea; gradient was from pH 3.5–10. No bands were visible below pH 4.5 or above pH 9.5.

Table 2. Analysis of SDS-PAGE bands by isoelectric focusing

SDS-PAGE band MW × 10^{-3}	Visualized components in IEF gel Number	pI range
52.7	9	6.7–7.7
45.6	12	6.9–8.3
37.2	2	7.6–7.9
32.2	3	7.5–7.9
21.9	2	5.8–6.0
	4	6.7–7.2
19.7	3	6.5–6.7
	2	7.1–7.2
18.0	2	9.0–9.3
15.8	·	
14.0	—	
11.2	—	

polypeptide size coupled with charge heterogeneity are not apparent, although there has been some speculation on the subject. It may be that the functional requirements for storage proteins are so few or so unspecific that there has been little or no evolutionary advantage or selection pressure for one form over another, leading to a proliferation of isoelectric species [24]. *In vivo* amidation–deamidation of glutamate and aspartate residues in a single parent polypeptide has also been suggested as a possibility [23, 24].

Regardless of the reasons or mechanisms that lead to charge heterogeneity, the fact that it exists may improve the chances of discovering or inducing nutritionally superior variants of cottonseed storage proteins. Screening of the available germplasm will be necessary to determine the nature and limits of this polydispersity before any conclusions can be drawn regarding its practical value.

EXPERIMENTAL

Prepration of CSPB proteins. CSPB were isolated from acid delinted, dehulled ground cottonseed cv Deltapine 16 as described in ref. [25]. CSPB were washed twice each with glycerol, 0.25 M sucrose and H_2O. The H_2O washes were carried out rapidly (*ca* 10 min each) to avoid membrane rupture and loss of CSPB contents. Washed CSPB were then extracted for 30 min with H_2O, 0.015 N NaOH or (for IEF) 6 M deionized [26] urea. Contents of H_2O and NaOH extracts were estimated with a Coleman nitrogen analyzer [25].

SDS-PAGE and IEF. SDS-PAGE was carried out at pH 7.0 in 10% polyacrylamide gel slabs using the method of Weber and Osborn [9] with and without 1% ME in the sample. Procedures used for staining and for MW estimations were those of ref. [8]. For IEF, CSPB were extracted with deionized [26] 6 M urea. IEF across a pH 3.5–10 gradient was in 5% polyacrylamide gel plates containing deionized 6 M urea as described in ref. [27]. Sample sizes for SDS-PAGE and IEF were dictated by sensitivity of the stain; protein contents of the samples were estimated by a modified Lowry procedure [28]. Amounts of sample were adjusted to give maximum stain intensity and yet minimize overloading effects on the most intense bands; separations made for direct comparison contained comparable amounts of protein. For two-dimensional separations, SDS-PAGE was used for the first dimension followed by IEF for the second. Experimental parameters for each technique were as stated above. Upon completion of the SDS-PAGE separation, an intermediate step was used to eliminate interference from SDS and improve resolution in the IEF dimension: a gel strip 5 mm wide containing the polypeptide pattern was excised from the SDS-PAGE slab and soaked 10 min each in 3 changes of deionized 8 M urea containing 0.05% (v/v) Triton X-100 and 1% ME. The equilibrated strip was then cast into a gel plate and subjected to IEF.

Amino acid analysis. After completion of SDS-PAGE, unfixed and unstained gel slabs were chilled at 4° until all the protein bands were visible (6–8 hr). Bands (I and II) or groups of bands (III–V) were excised from the gels with a fine-wire cheese slicer and extracted at room temp. for 16 hr in 10 mM Pi (pH 7.0) containing 0.1% SDS. Extracts were filtered, concd by lyophilization, dissolved in the same buffer and freed from soluble gel constituents by passage through a column of Bio-Gel P-10. Recovered protein was lyophilized. For amino acid analysis of whole CSPB contents, 0.015 N NaOH extracts of CSPB were lyophilized, dissolved in 10 mM Pi (pH 7.0) containing 0.1% SDS, passed through the P-10 column and lyophilized. Hydrolysis and amino acid analysis followed the procedure described in ref. [29], except that the hydrolysates were evapd to dryness and redissolved in 0.1 N HCl. This soln was loaded onto a column of Amberlite IR-120H (Cl^-); SDS and Pi were removed with H_2O, and the amino acids were recovered with 7 N NH_4OH. Following evapn of the NH_4OH, the residue was used for analysis.

Acknowledgements—I am grateful to Mrs. G. E. Lamkin for her fine technical work, to Dr. C. D. Elmore and Mr. L. E. Clarke for the amino acid analyses, and to Dr. H. R. Leffler for helpful discussions.

REFERENCES

1. Elmore, C. D. and King, E. E. (1978) *Plant Physiol.* **62,** 531.
2. Castro, C. E., Yang, S. P. and Harden, M. L. (1976) *Cereal Chem.* **53,** 291.
3. Martinez, W. H., Berardi, L. C. and Goldblatt, L. A. (1970) *J. Agric. Food Chem.* **18,** 961.
4. Meredith, W. R. Jr., Elmore, C. D. and King, E. E. (1978) *Crop Sci.* **18,** 577.
5. King, E. E. and Leffler, H. R. (1979) in *Seed Protein Improvement in Cereals and Grain Legumes*, Vol. II, p. 385. International Atomic Energy Agency, Vienna.
6. Dieckert, J. W. and Dieckert, M. C. (1976) in *Genetic Improvement of Seed Proteins*, p. 18. National Academy of Sciences, Washington.
7. Cunningham, S. D., Cater, C. M. and Mattil, K. F. (1978) *J. Food Sci.* **43,** 464.
8. King, E. E. and Leffler, H. R. (1979) *Plant Physiol.* **63,** 260.
9. Weber, K. and Osborn, M. (1975) in *The Proteins*, 3rd edn, Vol. I, p. 180. Academic Press, New York.
10. Youle, R. J. and Huang, A. H. C. (1978) *Plant Physiol.* **61,** 13.
11. Youle, R. J. and Huang, A. H. C. (1979) *J. Agric. Food Chem.* **27,** 500.
12. Spies, J. R., Bernton, H. S. and Chamber, D. C. (1960) *J. Allergy* **31,** 162.
13. Yatsu, L. Y. and Jacks, T. J. (1968) *Arch. Biochem. Biophys.* **124,** 466.
14. Altschul, P. M., Yatsu, L. Y., Ory, R. L. and Engleman, E. M. (1966) *Annu. Rev. Plant Physiol.* **17,** 113.
15. Elmore, C. D., Meredith, W. R. Jr., King, E. E. and Ranney, C. D. (1977) *Crop. Sci.* **17,** 687.
16. Lawhon, J. T. Cater, C. M. and Mattil, K. F. (1977) *J. Am. Oil Chem. Soc.* **54,** 75.

17. FAO (1970) *Amino Acid Contents of Foods and Biological Data on Proteins*, p. 66. Food and Agriculture Organization, United Nations, Rome.
18. Conkerton, E. J. and Frampton, V. L. (1959) *Arch. Biochem. Biophys.* **81,** 130.
19. Boulter, D. (1976) in *Genetic Improvement of Seed Proteins*, p. 231. National Academy of Sciences, Washington.
20. O'Farrell, P. H. (1975) *J. Biol. Chem.* **250,** 4007.
21. Basha, S. M. M. (1979) *Plant Physiol.* **63,** 301.
22. Gianazza, E., Righetti, P. G., Pioli, F., Galante, E. and Soave, C. (1976) *Maydica* **21,** 1.
23. Stegemann, H. (1975) in *The Chemistry and Biochemistry of Plant Proteins*, p. 71. Academic Press, London.
24. Righetti, P. G., Gianazza, E., Viotti, A. and Soave, C. (1977) *Planta* **136,** 115.
25. King, E. E. and Lamkin, G. E. (1977) *J. Agric. Food Chem.* **25,** 1211.
26. Marglin, A. and Merrifield, R. B. (1967) *Arch. Biochem. Biophys.* **122,** 748.
27. Karlsson, C., Davies, H., Öhman, J. and Andersson, V. (1973) *LKB Application Note No. 75*, LKB-Produkter AB, Bromma.
28. Bensadoun, A. and Weinstein, D. (1976) *Analyt. Biochem.* **70,** 241.
29. Elmore, C. D. and Leffler, H. R. (1976) *Crop Sci.* **16,** 867.