Heterogeneity of Soybean Seed Proteins: One-Dimensional Electrophoretic Profiles of Six Different Solubility Fractions

Branda Hu and Asim Esen*

Soybean seed proteins were separated into six solubility fractions by exhaustive, sequential extraction with six different solvents. These solvents were deionized distilled H_2O, 0.5 M NaCl, 70% 2-propanol, 60% acetic acid, 0.1 M NaOH, and 0.1 M sodium borate (pH 10) containing 1% $NaDodSO_4$ and 1% 2-mercaptoethanol. The water-soluble proteins made up ~75% of the total protein in the seed; other solubility fractions together contributed the remaining 25%. Sodium dodecyl sulfate ($NaDodSO_4$)–polyacrylamide gel electrophoresis and isoelectric focusing (IEF) were used for electrophoretic separations. The $NaDodSO_4$–polyacrylamide gel electrophoresis resolved a combined total of 188, 227, and 292 components from six solubility fractions when separations were performed in 10, 12, and 10–15% (gradient) gels, respectively. IEF profiles of four solubility fractions included a total of 268 components. Electrophoretic profiles of the water-soluble, salt-soluble, and acetic acid soluble fractions showed considerable similarity while those of other solubility fractions were distinct and displayed no detectable homologies.

Soybean seed contains the highest amount of protein (35–45%) in Leguminosae. This makes it a relatively inexpensive source of protein for human consumption and livestock feed. However, the nutritional quality of soybean protein and that of other legumes is relatively low due to the inadequate levels of sulfur-containing amino acids, methionine and cysteine, especially for infants and children. Therefore, there has been considerable emphasis on the improvement of methionine and cysteine levels in soybean by genetic means in recent years. Attempts to induce mutations with the objective of increasing the levels of sulfur-containing amino acids have so far not been successful (Gillespie and Blagrove, 1976). Legume seeds do not have well-defined solubility fractions, comparable to those in cereals, which differ dramatically in their content of limiting amino acids. Therefore, improving protein quality by changing relative proportions of major solubility fractions, as in cereals, does not seem to be a viable approach in legumes. Smartt et al. (1975) suggested increasing the relative proportion of polypeptides with the most desirable amino acid profile as one of several plausible approaches for improving protein quality in legumes. Application of this approach requires purification and characterization of the major protein components present in legume seeds.

Soybean seed proteins have been the subject of numerous investigations, and the major protein components extracted with water or aqueous buffers of defined pH and ionic strength are well characterized (Wolf, 1977). However, information on proteins soluble in other solvents is lacking. This study was undertaken (1) to fractionate soybean proteins by solubility in different solvent systems and determine the contribution of each solubility fraction to the total protein content and (2) to investigate size and charge heterogeneity of polypeptides comprising each solubility fraction.

MATERIALS AND METHODS

Mature soybean seeds from the cultivar Dorman were dehulled and ground in a Wiley mill so as to pass a 40-mesh screen. The seed meal was defatted twice with petroleum ether at 4 °C for 24 h by using a solvent to meal (v/w) ratio of 15:1 before it was used for protein extraction.

Department of Biology, Virginia Polytechnic Institute and State University, Blacksburg, Virginia 24061.

Protein Extraction. Six different solvents, deionized distilled water, 0.5 M NaCl, 70% 2-propanol, 60% glacial acetic acid, 0.1 M NaOH, and 0.1 M sodium borate (pH 10) including 1% sodium dodecyl sulfate and 1% 2-mercaptoethanol (2-ME) (hereafter referred to as H_2O, NaCl, IPA, HAc, NaOH, and $NaDodSO_4$, respectively), were used in sequence to extract virtually all the protein species that occur in the soybean seed. The protocol for this sequential extraction is outlined in Figure 1. Extraction with each solvent was repeated until no further protein was extracted.

Protein Determination. Supernatant fluids from extractions with the same solvent were combined to obtain a pool representative of each solubility fraction and assayed for protein according to the method of Esen (1978) using bovine serum albumin (BSA) as the protein standard. Fractions were then dialyzed against their own solvents and deionized water, respectively. Sodium azide was added to the dialysis tank at a final concentration of 0.02% to inhibit microbial growth. Dialysates were freeze-dried, and resulting protein powders were stored in sealed specimen vials in a freezer at about −20 °C.

In addition, nitrogen content of the seed meal was determined by micro-Kjeldahl method both before and after defatting. Nitrogen was converted to protein by using the conversion factor 6.25.

Sodium Dodecyl Sulfate–Polyacrylamide Gel Electrophoresis. $NaDodSO_4$–polyacrylamide gel electrophoresis was performed by following the procedure of Laemmli (1970) in 10, 12, and 10–15% (gradient gel) polyacrylamide gel slabs (125 × 0.75 × 155 mm, $l \times d \times w$) overlaid with 5% stacking gel. Protein powders were solubilized in the sample buffer (0.1 M Tris-HCl, pH 6.8, containing 2% $NaDodSO_4$, 5% 2-mercaptoethanol, and 10% glycerol) at a final concentration of 10 mg/mL. The samples were then heated at 97 °C for 2 min. Optimum sample loads were determined by performing preliminary runs at different sample loads. In addition, sample loads were intentionally increased to detect the minor components. They were also decreased to improve the resolution of major components whose molecular weights were within several thousand of one another and thus appeared as a single band at high sample loads. Phosphorylase *b* (94000), bovine serum albumin (67000), ovalbumin (43000), carbonic anhydrase (30000), soybean trypsin inhibitor (21000), and lysozyme (14000) were used as standards to

0021-8561/81/1429-0497$01.25/0

estimate the molecular weight range of polypeptides in each solubility fraction as well as to identify the subunits of the major soybean proteins, e.g., β-conglycinin (7S component) and glycinin (11S component).

After electrophoresis, proteins were stained with 0.1% Coomassie brilliant blue R-250 in acetic acid–2-propanol–water (10:25:65 v/v) and destained with the same solution minus the dye. The $NaDodSO_4$–polyacrylamide gel electrophoretic profiles of solubility fractions were recorded by photography and also by tracing them into a sheet of transparency placed on a wet gel sandwiched between two glass plates.

Isoelectric Focusing (IEF). IEF was performed in thin layers of polyacrylamide gels (115 × 0.75 × 200 mm, $l \times d \times w$) using an LKB-Productor (Multiphor 2117) apparatus as described by Davies (1975). Gel composition was similar to that used by O'Farrell (1975) except that urea concentration was reduced to 8 M to prevent crystallization of urea in the gel as a result of cooling (5 °C) during the run. The pH gradient was obtained by using a mixture of ampholines pH 3.5–10, 4–6, 5–7, 5–8, and 8–9.5 in a ratio of 1:1:1:1:1 and at a final concentration of 2%. The anodic and cathodic electrolytes were a saturated solution of aspartic acid and 0.1 M solution of lysine, respectively [modified from Caspers and Chrambach (1977)].

Freeze-dried protein samples were dissolved in the "lysis buffer" (O'Farrell, 1975) at a 10 mg/mL concentration. Samples (15 μL), absorbed into 5 × 8 mm 3 MM chromatography paper strips, were applied at the anodic end of the gel with the exception of the NaOH-soluble fraction, which was applied at the cathodic end of the gel. IEF was carried out under regulated voltage, starting with 200 V and increasing the voltage in steps (100 V/h) to 600 V in 4 h and maintaining it at 600 V for 2 h before the experiment was terminated. At the end of the experiment, gels were fixed in 12.5% trichloroacetic acid (Cl_3AcOH) for 12–16 h, washed with an acetic acid–ethanol–water (5:30:65 v/v) solution for 0.5 h, stained with 0.1% Coomassie brilliant blue R-250 in acetic–ethanol–water (10:45:45 v/v) for 2 h, and destained with the same solution (5:30:65 v/v) used for washing after Cl_3AcOH fixation. The IEF profiles of different solubility fractions were recorded in the same way as that described for $NaDodSO_4$–polyacrylamide gel electrophoretic profiles.

RESULTS AND DISCUSSION

Protein Content of Different Solubility Fractions and Its Contribution to the Total Protein. Results indicated that the protein content of seed meal from the cultivar Dorman used in this study was 39% (w/w) before and 47.4% (w/w) after defatting when protein was determined by the micro-Kjeldahl method. The sum total of protein in six different solubility fractions amounted to 38.2% (w/w) of nondefatted and to 47.0% (w/w) of defatted seed meal, respectively, when protein was determined by the method of Esen (1978). These data suggest that the sequential extraction with six different solvents released nearly all of the protein present in the seed meal. The amount of protein extracted with each solvent is given as percent of the total extractable protein in Figure 1.

However, the contribution of individual solubility fractions to the total protein may slightly vary from given values given in Figure 1. This is because the protein assay used is based on the binding of an anionic dye (Coomassie brilliant blue R-250) to the basic amino acids residues on protein. Since proteins in different solubility fractions may vary with respect to their basic amino acid content and the protein standard (BSA) used may not be truly representative of proteins in these fractions, a slight overestimation or underestimation, depending on the fraction, cannot be ruled out. Moreover, the H_2O- and HAc-soluble fractions had similar electrophoretic profiles especially with respect to major protein components (see later). Thus the true relative contribution of the water-soluble fraction to the total protein may well be over 85%.

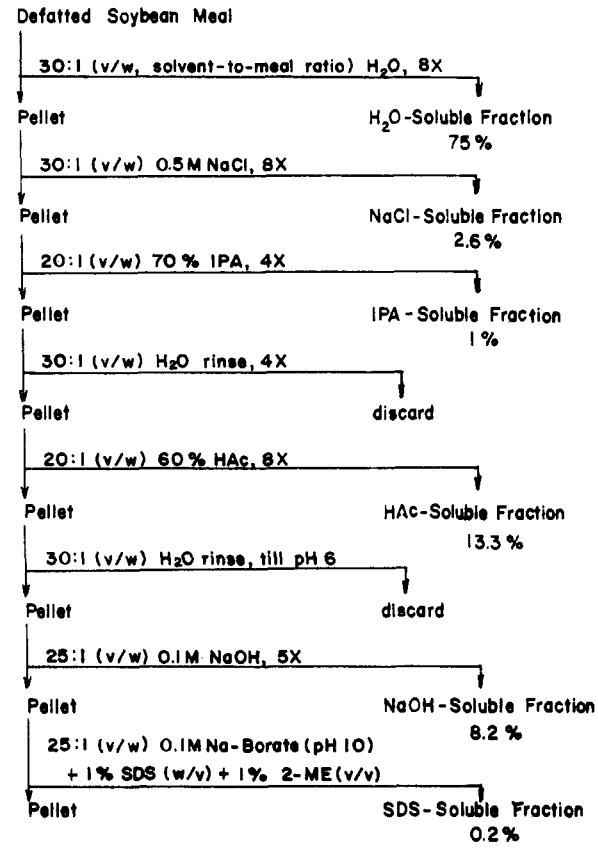

Figure 1. Flow sheet diagram of the protocol used for fractionation of soybean proteins by solubility and percent of total protein soluble in each solvent. Extraction with H_2O and NaCl was at 4 °C, with IPA, HAc, NaOH, and H_2O rinses at room temperature (~24 °C). The $NaDodSO_4$ (SDS in the figure)-soluble fraction was first extracted at room temperature for 45 min and then at 70 °C for 15 min. The number before the times sign indicates how many times a given extraction was repeated. Unless stated otherwise, the duration of each extraction was 12–16 h with agitation and all centrifugations were at 27000*g*, 5 °C.

Size Heterogeneity of Different Solubility Fractions. $NaDodSO_4$–polyacrylamide gel electrophoretic separations performed at three different gel concentrations revealed considerable size heterogeneity within each of the six different solubility fractions (Figures 2 and 3, and Table I). The best resolution of size heterogeneity was achieved in gels containing a linear gradient of 10–15% acrylamide which resolved a total of 292 components from 6 fractions. Corresponding numbers were 188 and 227, respectively, in 10 and 12% nongradient gels. Thus, the gradient gel improved resolution by 25–40% when compared to nongradient gels. The superiority of gradient gels was most obvious in the resolution of low molecular weight components (below 20 000) as well as those predominant components which differ little in size and thus form a single broad band in a nongradient gel.

The H_2O-soluble fraction was the most heterogeneous among the six fractions, having at least 88 components [parts A (2) and B (2) of Figure 2]. This fraction includes the two major soybean protein components, β-conglycinin (7S globulin) and glycinin (11S globulin), and the whey

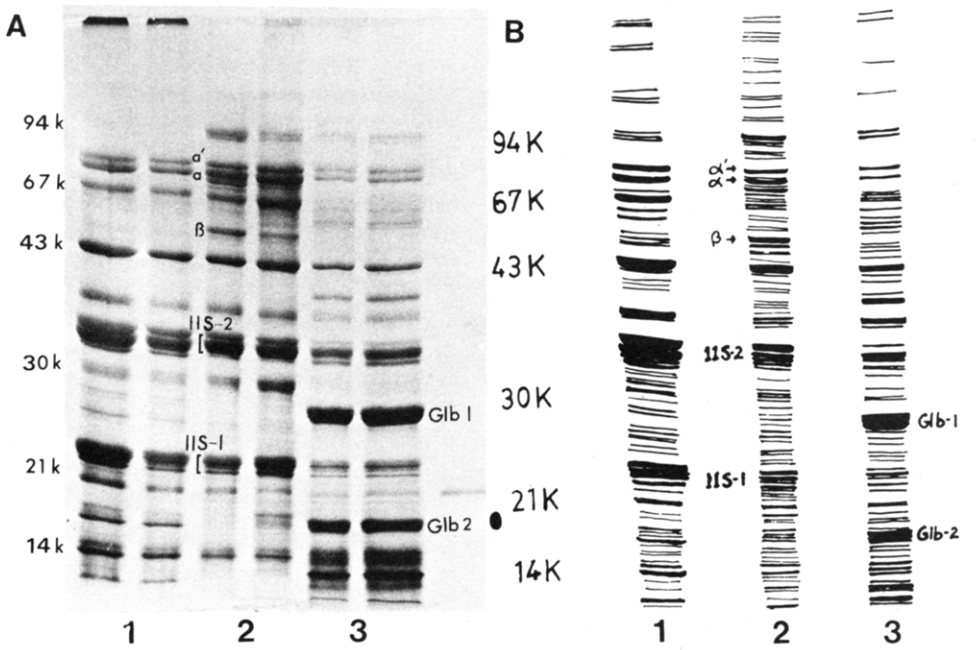

Figure 2. (A) $NaDodSO_4$–polyacrylamide gel electrophoretic (10–15% gradient gel) profiles of HAc-soluble (tracks 1 and 2, labeled 1), H_2O-soluble (tracks 3 and 4, labeled 2), and NaCl-soluble (tracks 5 and 6, labeled 3) protein fractions from the soybean seed. The origin and cathode are at the top. Numbers on the left indicate estimated molecular weight distribution along length of profiles. Note the striking similarity between profiles of HAc- and H_2O-soluble proteins. Subunits (α', α, and β) of β-conglycinin and acidic (11S-2) and basic (11S-1) subunits of glycinin are indicated. (B) Replica tracing of profiles of (A) from a wet gel to show minor components and details not visible in (A).

Table I. Number of Protein Components That Occur in Different Solubility Fractions of Soybean Seed As Determined by Sodium Dodecyl Sulfate–Polyacrylamide Gel Electrophoresis and Isoelectric Focusing

	no. of components resolved by[d]			
	$NaDodSO_4$-polyacrylamide gel electrophoresis			
solubility fraction[a]	10%	12%	10-15%[b]	IEF
water	68	68	88	73
0.5 M sodium chloride	44	46	70	68
70% 2-propanol	10	19	20	*c*
60% acetic acid	49	60	69	79
0.1 M sodium hydroxide	13	21	24	48
0.1 M sodium borate (pH 10) containing 1% $NaDodSO_4$ and 1% 2-mercaptoethanol	4	13	21	*c*
total:	188	227	292	268

[a] Protein fractions are listed by their respective solvents and the sequence in which they are extracted. [b] Linear gradient of polyacrylamide. [c] No data due to poor resolution. [d] $NaDodSO_4$, sodium dodecyl sulfate; IEF, isoelectric focusing.

protein, i.e., proteins that remain in solution when the water extract is titrated to pH 4.4 (Rackis et al., 1959). Molecular weight of components in this fraction varied from about 10 000 to 100 000. About 16 components dominated the profile of the water-soluble fraction. We believe that 10 of these are the polypeptides that make up β-conglycinin and glycinin. Thanh and Shibasaki (1976, 1977) purified β-conglycinin to near homogeneity and showed it to be made up of three subunits α', α, and β. We compared the positions of these polypeptides, shown in Figure 4 of Beachy et al. (1978), to those in the profile of our H_2O-soluble fraction and concluded that the components with molecular weights 79 000, 70 000, and 50 000 correspond to α', α, and β subunits of β-conglycinin, respectively. Our estimates of molecular weights for these subunits are somewhat higher than those reported by Thanh and Shibasaki (1977). The discrepancy presumably resulted from differences in electrophoretic conditions (e.g., gel concentration; percent cross-linking) and molecular weight standards used. The densitometric scanning of gels and quantification of the area under peaks corresponding to the subunits of β-conglycinin indicate that they account for 21.1% of the protein in the H_2O-soluble fraction. Similarly, the components with molecular weights 38 000 and 26 000 were thought to correspond to acidic (11S-2) and basic (11S-1) subunits, respectively, of glycinin described by Catsimpoolas (1969), Kitamura et al. (1976), and Beachy et al. (1978). If these identifications were accurate, glycinin polypeptides account for 39.5% of the protein (based on densitometric quantification) in the H_2O-soluble fraction.

The NaCl-soluble fraction approached the water-soluble fraction in size heterogeneity, displaying at least 70 components in 10–15% gels [parts A (3) and B (3) of Figure 2]. Molecular weight range was 8000–100 000. Two major components with molecular weights around 26 000 and 16 500, which are referred to as globulin-1 (Glb 1) and globulin-2 (Glb 2), respectively, dominated the profile of this fraction. These two components accounted for 45.8% of the protein in the NaCl-soluble fraction. Some of the remaining major components had mobilities similar to those of subunits of β-conglycinin and glycinin, while others were unique to the NaCl-soluble fraction. In fact, 37 components from the H_2O- and NaCl-soluble fractions showed size homology.

The HAc-soluble fraction included at least 69 components in 10–15% gels, ranging in molecular weight from about 9000 to 132 000 [parts A (1) and B (1) of Figure 2]. The profile of this fraction contained essentially the same major components found in the profile of the H_2O-soluble fraction. Virtually, all subunits of β-conglycinin and glycinin appeared in this fraction. The number of compo-

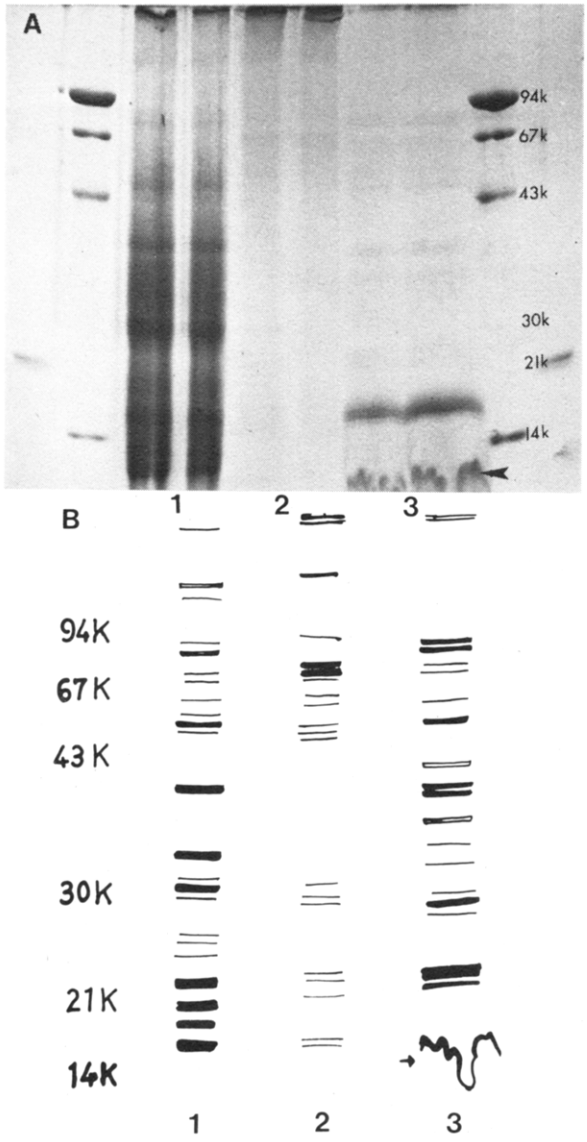

Figure 3. (A) $NaDodSO_4$–polyacrylamide gel electrophoretic (10–15% gradient gel) profiles of NaOH-soluble (tracks 1 and 2, labeled 1), $NaDodSO_4$-soluble (tracks 3 and 4, labeled 2), and 2-propanol-soluble (tracks 5 and 6, labeled 3) protein fractions from the soybean seed. The origin and cathode are at the top. (B) Replica tracing of profiles in (A).

nents common to the H_2O- and HAc-soluble fractions were found to be 41. In addition, the HAc-soluble fraction appeared to share at least 28 components with the NaCl-soluble fraction.

The occurrence of what appeared to the same components in profiles of the three solubility fractions (H_2O-, NaCl-, and HAc-soluble) was not expected in view of the exhaustiveness of the extraction procedure. It was almost unlikely to have the same proteins in different solubility fractions since extraction with each solvent was carried to completion before switching to the next solvent (Figure 1). Moreover, the eighth extract with each solvent had contained no detectable protein. Thus, it is conceivable that the components common to two or all three of these fractions are different polypeptides with the same molecular weight. However, this is highly unlikely because the IEF profiles of these fractions also share many components with the same isoelectric points (see later), presumably corresponding to those observed in $NaDodSO_4$–polyacrylamide gels. Alternatively, one could postulate the formation of H_2O-insoluble polymers among some H_2O-soluble proteins or insoluble complexes between proteins and some nonprotein constituents. In this case some of these polymers and complexes may be NaCl soluble and some HAc soluble. The formation of subcellular inclusions from protein bodies during extraction with aqueous solvents (Tombs, 1967) may also account for the results obtained. If these inclusions were susceptible to HAc treatment, similar protein patterns among different solubility fractions would be due to release of protein from these inclusions during extraction with HAc.

The three remaining solubility fractions (IPA, NaOH, and $NaDodSO_4$ soluble) were much less heterogeneous than the water-, NaCl- and HAc-soluble fractions. Both IPA- and $NaDodSO_4$-soluble fractions included at least 20 components [parts A (1) and B (1) of Figure 3 and Table I]. A single component (M_r ~16000) accounted more than 50% of the total protein in the IPA-soluble fraction [parts A (3) and B (3) of Figure 3], while there were no such major components in the $NaDodSO_4$-soluble fraction. Furthermore, a low molecular weight protein-like component consistently appeared in the profile of the IPA-soluble fraction [parts A (3) and B (3) of Figure 3, arrow]. This component produced a serrated and poorly defined band, which might represent a small polypeptide or even nonprotein constituent. The NaOH-soluble fraction contained at least 25 components [parts A (1) and B (1) of Figure 3], 6 of which contributed most of the protein to this fraction. The resolution of components in the NaOH-soluble fraction was consistently poor, presumably due to nondialyzable nonprotein impurities in this fraction. Ma and Bliss (1978) reported similar difficulties with electrophoresis of the NaOH-soluble proteins of beans (*Phaseolus vulgaris* L.). These nonprotein substances were thought to be nucleic acids (DNA; RNA) on the basis of absorbance measurements made at 260 vs. 280 nm. However, digestion with micrococcal nuclease and subsequent dialysis did not improve the resolution.

Charge Heterogeneity in Different Solubility Fractions. The IEF profiles of four solubility fractions (H_2O, NaCl, HAc, and NaOH soluble) displayed a total of at least 268 components (Figure 4 and Table I). Of these, the HAc-soluble fraction [Figure 4A (2)] was the most heterogeneous with 79 components, followed by water-soluble [Figure 4A (3)], NaCl-soluble [Figure 4A (1)], and NaOH-soluble [Figure 4A (4)] fractions in which 73, 68, and 48 components, respectively, were resolved. Again, profiles of the H_2O- and HAc-soluble fractions, like their $NaDodSO_4$–polyacrylamide gel electrophoretic profiles, were essentially identical with each other with respect to all major and many minor components. Sixty-one components appeared to be common to both fractions. Similarly, there were at least 36 components common to H_2O-, NaCl-, and HAc-soluble fractions. Postulated mechanisms which might be responsible for similarities in electrophoretic profiles of these fractions have been mentioned before. The IEF profile of the NaOH-soluble fraction was distinct in that it did not appear to share any component with other fractions [Figure 4A (4)]. Furthermore, proteins in this fraction were exclusively acidic proteins, having isoelectric points below 7. This is evident from their crowding at the acidic point end of the pH gradient as opposed to distribution of components evenly in other solubility fractions along the pH gradient. In fact, the three components of the NaOH-soluble fraction focused near the acidic end of the gradient (Figure 4A (4), arrows], which suggested that they were probably the most acidic soybean proteins. Furthermore, it should be pointed out that proteins in NaOH-soluble fraction gave poor resolu-

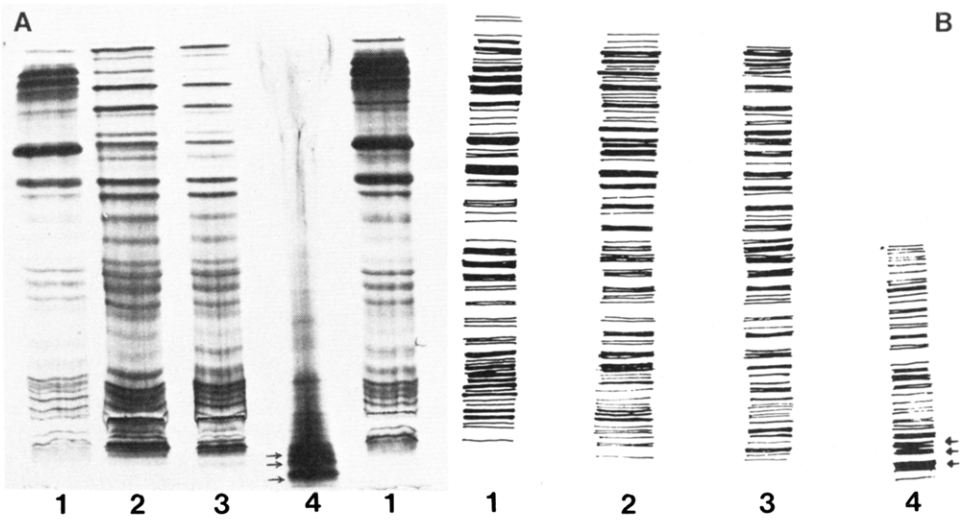

Figure 4. (A) Isoelectric focusing (IEF) profiles of NaCl-soluble (track 1), HAc-soluble (track 2), H_2O-soluble (track 3), and NaOH-soluble (track 4) protein fractions from the soybean seed. The cathode is at the top. Note the similarity between IEF profiles of HAc- and H_2O-soluble proteins. (B) Replica tracing of profiles in (A).

tion when the sample was applied at the anodic end of the gel. In contrast, H_2O-, NaCl-, and HAc-soluble fractions yielded satisfactory resolution irrespective of the position of the sample application. The IPA- and $NaDodSO_4$-soluble fractions consistently produced such poor resolution that their IEF profiles could not be analyzed. It is likely that nonprotein impurities were responsible for this poor resolution.

The components resolved by IEF included mostly more than one polypeptide based on two-dimensional analysis (Hu and Esen, 1980). Also, it could be assumed that none of the IEF components included native oligomeric proteins since IEF was performed under denaturing conditions.

Conclusions. (1) Results of this investigation show that soybean seed proteins can be separated into six fractions based on solubility in different solvent systems. Obviously proteins extracted with H_2O, NaCl, and HAc are true storage proteins as well as those with biological activity (enzymatic). The nature of proteins extracted with IPA, NaOH, and $NaDodSO_4$ is not known; one could speculate that they are mostly structural proteins (e.g., membrane proteins). (2) Extraction with a given solvent is not complete unless it is repeated as many as 5 times. However, extractions do not have to be as exhaustive in terms of number and duration as was done in this study. We found that a 400:1 solvent to meal ratio released nearly all of the protein soluble in a given solvent when extraction was repeated twice. (3) Soybean seed proteins show tremendous size and charge heterogeneity which can only be detected by high-resolution analytical electrophoretic techniques. Even after allowance is made for the same components that occur in different solubility fractions, there are at least 200 components that differ in size and charge and detectable by Coomassie brilliant blue R-250 staining. Two-dimensional analysis of three solubility fractions indicate that soybean seed contains at least 1000 different polypeptides. (4) Despite the fact that soybean seed proteins are highly heterogeneous, ~20–25 components (subunits of β-conglycinin and glycinin and some whey proteins) make up more than 60% of the total protein. Furthermore, with the exception of about three unique NaCl-soluble major components, nearly all major protein components occur in the H_2O-soluble fraction. It is likely that soybean germ plasm includes genetic variants with varying quantities of these major proteins components as well as additional ones. It is thus worth conducting a screening of the H_2O-soluble fraction of the soybean germ plasm by high-resolution electrophoretic techniques for major protein components. This screening should be preceded by a purification and characterization of the major components, specifically with respect to their methionine and cysteine content. One can then initiate a breeding program aimed at improvement of protein quality in soybean based on the increase of major protein species with the most desirable amino acid profile and/or on the decrease of those with the least desirable amino acid profile.

LITERATURE CITED

Beachy, R. N.; Thompson, J. F.; Madison, J. T. *Plant Physiol.* **1978**, *61*, 139.
Caspers, M. L.; Chrambach, A. *Anal. Biochem.* **1977**, *81*, 28.
Catsimpoolas, N. *FEBS Lett.* **1969**, *4*, 259.
Davies, H. "Isoelectric Focusing"; Arbuthnott, J. B.; Beeley, J. A., Eds.; Butterworths: London, 1975; p 97.
Esen, A. *Anal. Biochem.* **1978**, *89*, 264.
Gillespie, J. M.; Blagrove, R. J. "Evaluation of Seed Protein Alterations by Mutation Breeding"; IAEA: Vienna, Austria, 1976; p 156.
Hu, B.; Esen, A., unpublished experiments, 1980.
Kitamura, K.; Takagi, T.; Shibasaki, K. *Agric. Biol. Chem.* **1976**, *40*, 1837.
Laemmli, U. K. *Nature (London)* **1970**, *227*, 680.
Ma, Y.; Bliss, F. A. *Crop Sci.* **1978**, *17*, 431.
O'Farrell, P. H. *J. Biol. Chem.* **1975**, *250*, 4007.
Rackis, J. J.; Sasame, H. A.; Anderson, R. L.; Smith, A. K. *J. Am. Chem. Soc.* **1959**, *81*, 6265.
Smartt, J.; Winfield, P. J.; Williams, D. *Euphytica* **1975**, *24*, 447.
Thanh, V. H.; Shibasaki, K. *Biochim. Biophys. Acta* **1976**, *439*, 326.
Thanh, V. H.; Shibasaki, K. *Biochim. Biophys. Acta* **1977**, *490*, 370.
Tombs, M. P. *Plant Physiol.* **1967**, *42*, 797.
Wolf, W. J. "Soybeans: Chemistry and Technology"; Smith, A. K.; Circle, S. G., Eds.; Avi Publishing Co.: Westport, CT, 1977; Vol. I, p 93.

Received for review February 1, 1980. Accepted September 25, 1980.