An Electrophoretic Investigation of Groundnut Proteins: the Structure of Arachins A and B

By M. P. TOMBS

Unilever Research Laboratory, Colworth House, Sharnbrook, Bedford

(Received 14 September 1964)

1. The proteins of the groundnut cotyledon have been fractionated and analysed by DEAE-Sephadex chromatography and acrylamide-gel electrophoresis. Seventeen components were detected. 2. A new method is described for the preparation of arachin, using calcium precipitation. The product contains at least 99% of arachin. 3. The theory of acrylamide-gel electrophoresis is developed and applied to the arachin system to predict the molecular weight of one sub-unit of arachin. 4. A variant form of arachin, arachin B, has been discovered. Of 81 nuts, 27 contained only arachin B, 53 contained both arachin A and B, and one contained arachin A only. This is almost certainly a polymorphism of arachin; this is the first example of polymorphism to be reported in plant proteins. 5. A combination of controlled denaturation, electrophoretic analysis, ultracentrifuge and Sephadex filtration data has shown that arachin A contains four different kinds of peptide chains (α , β , γ and δ). Arachin B contains only β , γ and δ chains. 6. The most probable structure for arachin B, mol. wt. 330 000 form, is 8 β , 2 γ and 2 δ chains, and for arachin A, 4 α , 4 β , 2 γ and 2 δ chains. Arachin without β chains was not found. 7. The α and β chains have mol. wts. of about 35 000 and the γ and δ chains of about 10 000. 8. Three *N*-terminal groups were found: the α and β chains both terminate in glycine; the γ and δ chains terminate in isoleucine and glutamic acid. 9. Arachin contains no carbohydrate. 10. Disulphide bonds are not important in arachin: there are none between the α , β , γ and δ chains. 11. The amino acid compositions of arachins A and B are very similar. Glutamic acid and aspartic acid residues are exceptionally frequent.

The proteins of the groundnut (*Arachis hypogaea*) have been the subject of many investigations. Usually the proteins were extracted with 10% sodium chloride solutions: the fraction precipitated from this by 40% saturation with ammonium sulphate was defined as arachin, and amounts to about two-thirds of the protein. The protein remaining in solution was called conarachin (Johns & Jones, 1916; Jones & Horn, 1930). Conarachin is a complex mixture; this was clearly shown by chromatographic analysis of the protein bodies found in groundnut cotyledon cells (Dieckert, Snowden, Moore, Heinzelman & Altschul, 1962).

Most previous work has been on arachin. This protein has always appeared to be relatively homogeneous, but, as usually prepared, it occurs in 'monomer' and 'dimer' forms of molecular weights 170 000 and 330 000 respectively (Johnson & Shooter, 1950; Johnson, 1946; Johnson & Naismith, 1956). On denaturation with urea, arachin first became insoluble in pH 8 buffers but slowly became soluble again; the molecular weight fell to about

17 000 (Johnson & Shooter, 1950). Controlled alkaline denaturation gave sub-units with a molecular weight of about 35 000 (Cater & Naismith, 1958). These results, together with some *N*-terminal determinations, led Cater, Naismith, Thompson & Ure (1957) to suggest a multi-stranded structure for arachin, though the numbers and kinds of peptide chains were not clear.

Tombs (1963) found that most groundnuts contain two closely related arachins, probably a hereditary polymorphism. With this in mind, the structure of the arachins has been investigated, and a general classification of groundnut protein is based on the results of polyacrylamide gel-electrophoresis.

EXPERIMENTAL

Material. Fresh groundnuts were obtained from Lummis Peanut Co., 148N. Delaware, Philadelphia, Pa., U.S.A. The nuts were unshelled and the oil contained 0.2% of free fatty acid of zero peroxide number.

Homogenization and extraction. Seeds were carefully

skinned and the hypocotyls removed, and then homogenized in 10 vol. of 10% (w/v) NaCl for 3 min. in an MSE Ato-Mix. Later, a ball mill was used.

Debris and oil were removed by centrifugation or filtration. Centrifugation was not ideal because the seed debris was similar in density to the medium. The oil came to the top but remained as an emulsion even at 200000g. The central clear layer was removed and centrifuged again to clarify it further. Filtration was more convenient, especially when Hyflo beds were used, although emulsions still led to difficulties. All extractions were at room temperature (20°) because of the marked decrease in arachin solubility with fall in temperature.

(i) Salt fractionation. Saturated $(\text{NH}_4)_2\text{SO}_4$ (40 vol.) was added to 60 vol. of extract (10%, w/v, NaCl solution). The oily precipitate of arachin was centrifuged out, the supernatant saturated with $(\text{NH}_4)_2\text{SO}_4$ and the precipitate again collected. The sediments were then suspended in water, dialysed until SO_4^{2-} ions could no longer be detected in the diffusate and then freeze-dried. Whole-peanut protein was also prepared by dialysing the extract in 10% NaCl against water, followed by freeze-drying of the contents of the dialysis sac.

(ii) Chromatography of whole-peanut protein. This was on DEAE-Sephadex A-50 with buffers similar to those described by Dieckert *et al.* (1962), i.e. 0.01M-sodium phosphate (pH 8) with a gradient from 0 to 2M-NaCl as eluent. The apparatus and general arrangement were similar to that described by Tombs, Cooke, Burston & MacLagan (1961). Fractions were collected, pooled and, after dialysis against water, freeze-dried.

(iii) Purification of arachin. Most experiments were done on arachin prepared by salting-out as above, but subsequently purified by chromatography. The conditions were as for (ii) but scaled up so that 5 g. of protein could be applied to the column.

(iv) Calcium precipitation of arachin. Arachin was precipitated quantitatively and specifically from crude 10% (w/v) NaCl extracts by CaCl_2 . M-Calcium chloride solution was added to the extract until it was evident that no more precipitate was produced (approx. 3 ml./g. of arachin); the oily precipitate was collected and redissolved in 0.025M-EDTA in 10% (w/v) NaCl, the pH being maintained at about 8 by addition of triethylamine. The small proportion remaining undissolved was removed by centrifuging. The solution was dialysed against the solvent with at least two changes, then against water and freeze-dried.

Determination of protein. Where possible this was done by refractometry, assuming a specific refractive increment of 1.8×10^{-3} . Otherwise the Lowry-Folin method, with bovine albumin as a standard, was used.

Analysis of proteins: electrophoresis. This was done by either agar-gel electrophoresis, as described by Tombs *et al.* (1961), or by polyacrylamide-gel electrophoresis (Raymond & Nakamichi, 1962).

Gels were made to the required concentration (in the range 5–15%, w/v) by dissolving Cyanogum 41 (British Drug Houses Ltd.) in water and adding 0.5 ml. of 2-dimethyl-aminoethyl cyanide followed by 1 ml. of 10% (w/v) ammonium persulphate to each 100 ml. of solution. This mixture gelled in a sealed Perspex former within 10 min. The former produced a gel 8 cm. $\times$ 14 cm. and 2 mm. deep, with four sample slots, each 12 mm. $\times$ 0.5 mm. $\times$ 1.5 mm.

(0.09 ml.). Gels made in water were transferred into the required electrophoresis medium by washing in a bath of buffer. Complete exchange required about an hour.

Most analytical runs were in a discontinuous buffer system. The gel contained tris-citrate, and the buffer tanks NaOH– NaBO_3 . Buffers were made up as follows: (i) 3.04 ml. of 2M-tris and 0.4 ml. of m-citric acid were diluted to 80 ml. with water; (ii) 30 ml. of 0.5M-boric acid and 2.4 ml. of 2.5N-NaOH were diluted to 500 ml. with water.

After electrophoresis, protein was detected by staining with a saturated solution of Naphthalene Black 12B in glycerol-methanol-water-acetic acid (20:50:50:1, by vol.). Unlike alcohol this mixture caused no distortion of the gel. Gels were washed in 1% acetic acid and stored wet in methylene bags.

Denaturation, reduction and alkylation. Arachin was depolymerized in 10M-formamide at pH 8–9 in the presence of mercaptoethanol. In a typical experiment, 500 mg. of arachin was dissolved in 12.5 ml. of formamide plus 12.5 ml. of 5% (w/v) triethylamine plus 0.2 ml. of 2-mercaptoethanol. This mixture was then kept at room temperature for about 1 hr.; times between 10 min. and 24 hr. gave similar results. Solvents were then removed and the protein was transferred into a suitable buffer by running down a column of Sephadex G-100: this gave an excellent separation. Small volumes (1 ml.) were transferred by dialysis.

In some experiments the reduced depolymerized material was alkylated in the formamide-triethylamine mixture. A large excess of alkylating agent (600 mg.) was added, sufficient to react with all the mercaptoethanol and the protein. The pH was frequently adjusted to 9 with triethylamine during the reaction.

For alkylation without prior reduction, the protein was dissolved in formamide-triethylamine containing the appropriate alkylating agent (iodoacetamide or *p*-chloromercuribenzoate), all in about fivefold excess.

Carbohydrates. Samples of arachin AB, of whole-groundnut protein and of conarachin were hydrolysed with ion-exchange resin according to Bragg & Hough (1961). The hydrolysate was evaporated to dryness, dissolved in water and chromatographed, with suitable markers, in the pyridine-ethyl acetate-water solvent of Partridge (1949).

Amino acid analyses. Total amino acid analyses were determined with an EEL automatic amino acid analyser according to Spackman, Stein & Moore (1958). Duplicate hydrolyses of 24, 36 and 48 hr. were performed with 6N-HCl under reflux. In addition a small quantity of arachin A was hydrolysed for 24 hr. only. Corrections for hydrolytic losses were made by extrapolation. Losses were large and corrections of 10% or greater were necessary for aspartate, glutamate, serine and threonine as well as smaller adjustments for arginine, lysine, phenylalanine and tyrosine. No attempt is made to correct for losses of cystine or methionine other than allowance for sulfoxides detected. The highest value obtained was taken.

N-Terminal amino acids. Arachin (100 mg.) was suspended in 10 ml. of 1% NaHCO_3 and 1.5 ml. of fluorodinitrobenzene was added in 0.5 ml. of ethanol. The mixture was stirred at 40° for 24 hr. and then extracted five times with ether after acidification with HCl.

Samples of substituted protein were refluxed for 48 hr. with Dowex 50 or Zeo-Karb 225 (3 g. of resin + 30 ml. of water + 100 mg. of protein). Then the DNP-amino acids

were extracted from the resin with hot water, acidified and extracted into ether. Dinitrophenol was removed by absorbing the whole sample on a buffered silicic acid column (0.1M-sodium phosphate, pH 6.6) followed by elution of the dinitrophenol with chloroform. The DNP-amino acids were then eluted with chloroform + 1% acetic acid, taken to dryness and finally dissolved in ether (Stevens & Tristram, 1962).

Samples were then chromatographed on paper in the following solvent systems with appropriate standards: (i) 1.5M-sodium phosphate, pH 6; (ii) ethanol-benzyl alcohol (1:9, v/v), saturated with phthalate buffer as in (iii); (iii) propan-1-ol-light petroleum (3:7, v/v) on paper saturated with phthalate buffer, pH 6 (50ml. of 0.1M-potassium hydrogen phthalate + 45.5ml. of 0.1N-NaOH + 4.5ml. of water) and dried.

Some attempts were made to convert the DNP-amino acid into amino acid by ammonolysis (Fraenkel-Conrat, Harris & Levy, 1955), followed by analysis with the automatic analyser. These failed because of very poor recoveries from the ammonolysis stages (about 2% for glycine).

In addition, *N*-terminal groups were found by the phenyl isothiocyanate reaction, as described by Fraenkel-Conrat *et al.* (1955). Arachin was insoluble in the usual solvents, such as dioxan, so the reaction was done in 10M-formamide, followed by transfer into water with Sephadex G-100. The protein that reacted was precipitated at once, but this did not hinder its elution from the column. The protein was then hydrolysed with *N*-HCl at 100° for 60 min., and the phenyl isothiocyanate derivatives of amino acids were removed and chromatographed according to Fraenkel-Conrat *et al.* (1955) in pyridine-heptane mixtures.

Thiol groups and ultraviolet-absorption spectra. Attempts to use the method of Boyer (1954) failed because of high extinctions at the necessary wavelengths. However, the modification described by Fridovich & Handler (1957) was used successfully. Arachin (20 mg.) was suspended in 4 ml. of water and 2.5 ml. of 0.5M-K₂SO₄ was added, followed by pH adjustment with NaOH where necessary. An excess of *p*-chloromercuribenzoate in 0.05M-K₂SO₄-0.1M-sodium phosphate, pH 7, was then added. After 45 min., 0.5 ml. samples were withdrawn and mixed with 5 ml. of 0.5M-sodium phosphate, pH 5, and then with 4 ml. of carbon tetrachloride containing 0.004% of dithione.

After mixing and centrifuging to break the emulsion that usually formed, the extinction at 625 m μ together with appropriate controls gave the amount of *p*-chloromercuribenzoate which had not reacted with the protein.

Thiol groups were also estimated by the method of Donovan (1963). The difference spectra were measured in a Unicam SP.500 spectrophotometer, by using the techniques described by Tombs & Apsey (1962).

RESULTS

Extraction and fractionation of groundnut proteins. The proteins were soluble in various alkaline buffers: the tris-citrate electrophoresis buffer was an excellent solvent (up to 8% of protein), as was sodium hydroxide solution, pH 9, but 0.01M-sodium phosphate, pH 8, became saturated near 2% of protein. It was thought undesirable to expose the

protein to alkali, so it was usually extracted with 10% (w/v) sodium chloride solution (maximum protein solubility is at 13%, w/v, sodium chloride) at about pH 5.2. Seeds were difficult to extract because of the formation of protein-lipid emulsions. The use of a ball mill rather than a blade homogenizer reduces emulsion formation, but prior extraction of lipid from the seeds is desirable. There was no difference between protein extracted from lipid-free meal and whole seeds.

It was hoped that chromatography might be suitable for the isolation of some of the proteins. Fig. 1 shows some chromatographic results, and the electrophoretic analysis of each fraction. The complexity is obvious: it is possible to distinguish at least 17 components. No fraction sufficiently pure for further analysis could be obtained, with the possible exception of arachin in poor yield.

Arachin was therefore prepared by the classical salting-out method, followed by chromatography on DEAE-Sephadex to remove minor contaminants. Arachin gave a number of peaks on chromatography which were ascribed to the various polymers the protein contains. The product was at least 95% of arachin as judged by gel electrophoresis. This material was used for most of the work described here until it was discovered that calcium precipitates the arachin with remarkable selectivity from the crude extracts (both in 10%, w/v, sodium chloride and sodium hydroxide solutions). The calcium could be removed by using EDTA, and large amounts of highly purified arachin were thus easily obtained. In addition, conarachin fractions could be obtained free from arachin contamination, a fact useful in their isolation.

Electrophoretic analysis. Electrophoresis in acrylamide gels was the major method of analysis. The method has very high resolving power, though it suffers from the usual drawback of zone methods in being only semi-quantitative. Like starch-gel electrophoresis the method depends for its effectiveness on filtration effects and deductions about the size as well as the charge of the molecule can be made. Acrylamide gels (unlike starch) can easily be prepared in a wide range of concentrations. Some results are shown in Fig. 2, where mobility in arbitrary units (actually cm. on standardized photographs of the gels) is plotted against $1/\sqrt{c}$, where *c* is the gel concentration, for a number of arachin and conarachin components. The interpretation of such results, and of gel electrophoresis generally, underlies so much of this work that a brief theoretical treatment is developed.

It is reasonable to suppose that the property of the gel which will have the major influence on mobility is the pore size; this will be true however irregular the shape of the protein molecule. We require, therefore, the pore size *P* of the gel.

Fig. 1. Chromatographic and electrophoretic analysis of groundnut cotyledon proteins. At the top is shown the elution pattern of the proteins: numbers in parentheses show the percentage of the total eluted protein in that fraction. Below is shown the electrophoretic pattern of the freeze-dried protein in 7.5% acrylamide gels. Fraction VI was further fractionated because on dialysis against water before freeze-drying a precipitate separated. Further details are in the text.

Raymond & Nakamichi (1962) have given the equation:

$$P = kd/\sqrt{c} \quad (1)$$

where k is a constant depending on the way in which fibres cross to form the mesh and d is the diameter of the strands making up the mesh. Fig. 2 shows that this equation holds only over a limited range of gel concentrations: there is deviation at either high or low gel strengths. It is obvious that even in very strong gels some large pores will remain, and conversely in weak gels some of the pores will be very small. There will be a distribution of pore sizes and the mobility will be determined, not so much by a mean pore size P , but by the proportion of pores greater than a certain limiting size P_1 . If we assume that the distribution of pore size is Gaussian, and that the standard deviation of the distribution does not vary with gel strength, it can be shown that:

$$M = DQ[P_1 - (kd/\sqrt{c})] \quad (2)$$

where D is the maximum possible mobility (i.e. at zero gel strength) and Q is a tabulated function of the normal distribution. This predicts a sigmoid relationship for the plot shown in Fig. 2 but is nearly linear when between 20 and 80% of the pores pass the protein. The results shown in Fig. 2 are therefore consistent with the idea that filtration effects are the major factors determining mobility in this kind of electrophoresis.

Limiting pore sizes. Although the simple eqn. (1) does not apply at very high or very low mobilities, extrapolation of the linear part of the curve gives a value of $1/\sqrt{c}$ for which the mobility is zero. The

'pore size' of the gel must then be just too small for the protein molecule to penetrate, and can be approximated to the diameter of the protein molecule. This being so:

$$\text{Mol. wt.} = 10^{-24}(\rho N)^{\frac{4}{3}}\pi[\frac{1}{2}(kd/\sqrt{c})]^3 \quad (3)$$

where ρ is the density of the protein, N is Avogadro's number, k is a constant (not differing greatly from unity) and d is the diameter of the gel strands, in Ångström units. The value of the molecular weight depends very much on the estimated value for d . However, White (1960) has calculated pore sizes, with numerous assumptions, from the water permeability of the gels: for a 5% gel, 36Å; for a 10% gel, 26Å; for a 15% gel, 20Å. The use of these values, unfortunately, gives absurdly low molecular weights for arachin (5000 and 8000 for the monomer and dimer) and these pore sizes are smaller than the arachin molecules themselves (which, assuming them to be spheres, must be of the order of 80Å diameter). Small changes in the pore size produce large changes in apparent molecular weight and it seems unlikely that direct calculation of molecular weight in this way can be of much value. Nevertheless, the value of $1/\sqrt{c}$ at zero mobility is a constant directly related to the size and shape of the molecule, and its cube should bear a systematic relationship to the molecular weight for a series of similar proteins. Fig. 3 illustrates such a plot for the arachin series and although it is not linear it does pass through zero (i.e. an infinitely strong gel is required to stop an infinitely small molecule).

Fig. 2. Plots of mobility in various concentrations of acrylamide gels versus $1/\sqrt{c}$, where c is the gel concentration.

Fig. 3. Plots of $(1/\sqrt{c})^3$ at zero mobility against estimated molecular weight for arachin and its sub-units.

Fig. 4. Electrophoretic classification of groundnut proteins. At the top is the result of agar-gel electrophoresis; these zones are connected to the zones corresponding to arachin and conarachin in the acrylamide gel shown below. Below this is shown a diagram of the gel, for clarity, with identification of some of the components. The gel has been split into five areas: α , β , γ , δ and ϵ .

Electrophoretic classification. The most convenient classification of the protein was on the basis of their behaviour in agar and acrylamide gels. A classification based on this is shown in Fig. 4. Fifteen components were regularly detected in acrylamide, in contrast with the two salt fractions and the three seen in agar. Gels containing 7.5% of acrylamide were most suitable for general analysis and required about 5mg. of protein for each run.

The principal components seen in acrylamide were as follows. The percentages of protein in each region of the gel were obtained by scanning with a densitometer and are only approximate.

Region δ (5%). Only two components are seen in this region. They have high molecular weights ($> 500\,000$) and are high polymers of arachin.

Region γ (71%). This group contains most of the arachin zones, 'monomer' and 'dimer', and a component believed to be α -conarachin.

Region β (16%). This group of proteins is quite distinct and is characterized by a yellow colour; to judge from the opacity of the gel, carbohydrate was associated with them. After isolation phenolic components, arabinose, galactose and glucose were detected. On storage of protein solutions at pH 8.6

Table 1. Incidence of the five types of seed containing varying proportions of arachins A and B, together with the ratio of α and β chains in the total arachin in each seed

Seed type	Ratio (β/α chain)	No. of seeds
AAAA	1:1	1
AAAB	5:3	35
AABB	3:1	16
ABBB	7:1	5
BBBB	8:0	23

and 20° the β components were converted into yellow, then brown, very rapidly migrating components in 24 hr.

Region α (3%). These are minor components, of low molecular weight (i.e. $< 30\,000$), most of them being sub-units of arachin.

Variant forms of arachin. The existence of two arachins has been briefly reported (Tombs, 1963). More seeds have now been examined, and the arachins classified into five types, based on the relative proportion of arachin A (the faster-migrating variant) and B. The numbers of each type are shown in Table 1, which is discussed below.

Fig. 5. Electrophoresis of various arachin preparations in acrylamide gels. All runs were in 7.5% gels except for 1, which was in a 10% gel. Run 1 was in the presence of 0.1% mercaptoethanol; run 3 had 2% ammonium sulphate added to the sample.

Arachin polymers and sub-units. Fig. 5 shows electrophoresis patterns of isolated arachin. Preparations P₂, P₃ and P₄ were obtained by diluting and cooling a solution of peanut protein in 8% sodium chloride. They are in the more highly polymerized state, with a high proportion of 'dimer' and in P₂ a form which is probably 'tetramer'. These samples separated from solutions in an oil form (probably liquid crystals) and the presence of a high proportion of 'dimer' was characteristic of the oily precipitate. When arachin was isolated by dialysis of a solution in saline against water the diameter of the dialysis sac was found to have a decisive influence. Narrow sacs gave a precipitate, mainly in the 'monomer' form (run 4 in Fig. 5 is typical), whereas wide sacs gave an oil rich in 'dimer'. If acid precipitation was used, the higher the actual pH of precipitation the greater the proportion of dimer.

This behaviour is similar to that described by Johnson & Shooter (1950) and the 'monomer' forms can be identified with their monomer 9s form and the 'dimer' with their 14.6s form. Runs 1 and 4 show other components migrating more rapidly than the arachin 'monomer' and especially in run 4 it is clear that arachin can depolymerize past the 'monomer' stage to give the α , β , γ and δ bands (see below) and an intermediate ' $\alpha\beta$ dimer' form. According to Johnson & Shooter (1950), low concentrations of ammonium sulphate tend to suppress depolymerization. This was confirmed; the bands migrating faster than 'monomer' were diminished if 2% ammonium sulphate or sodium chloride was added.

These experiments also showed that A 'monomer' was more soluble than B 'monomer', giving a method of separation, and that arachin A depolymerizes more readily than arachin B in these buffers, as judged only by inspection of the gels. It is plain that arachin 'monomer' even at pH 8.6 breaks down to smaller units, and, to characterize these, arachin was treated with reagents which it was hoped would depolymerize it completely.

Arachin was depolymerized in 10M-formamide, at either pH 1 or 8, with equivalent results. Mercaptoethanol was present to reduce any disulphide bonds which might be part of the structure, and also to prevent their formation after depolymerization.

Some typical results are shown in Fig. 6. Two heavy zones, α and β , and two less heavy zones, γ and δ , together with a very variable set of components, M, appeared. Arachin B produced only the β , γ and δ chains, whereas arachin A produced α and β chains as well as the γ and δ chains. Scans of the gel with a reflectance densitometer gave a composition of 5% of γ , 5% of δ , 30% of α and 60% of β chains from a mixed AB preparation from 100 seeds.

Arachin A contained approximately equal amounts of α and β chains.

The γ and δ chains were smaller than the α and β chains; they trailed on Sephadex G-200 and were relatively less retarded in 12.5% gels. Components M were artifacts caused by alkaline degradation of α and β chains at pH 13 for 24 hr.; depolymerization at lower pH values decreased the amount found. They were also found in preparations kept in 8M-urea for 6 weeks, conditions which can also cause fission of peptide bonds.

Other methods of depolymerizing arachin were also tried, and gave patterns on electrophoresis very similar to that produced by the formamide-mercaptoethanol mixture. They were: (i) 10M-formamide with triethylamine to give pH 9; (ii) 50mg. of arachin + 100mg. of iodoacetamide in 5ml. of 10M-formamide-triethylamine (pH 9); (iii) 50mg. of arachin + 100mg. of *p*-chloromercuribenzoate in 5ml. of 10M-formamide-triethylamine (pH 9); (iv) 0.1N-sodium hydroxide for 120min. In addition, treatment of the reduced depolymerized arachin with alkylating agents such as iodoacetamide or *p*-chloromercuribenzoate gave much the same patterns. In some experiments a zone (labelled $\alpha\beta$ dimer in Fig. 6) was more prominent than in others, and it was never completely eliminated.

Ultracentrifugation. No extensive experiments were done. The main aim was to compare ultracentrifuge patterns with the results of electrophoresis, and also to strengthen the identification of the electrophoretic components with the ultracentrifuge components described previously and called by Johnson & Shooter (1950) 'monomer' and 'dimer'. Some typical patterns, together with the corresponding electrophoresis patterns, are shown in Fig. 7. All the depolymerized arachin preparations contained mainly components with $S_{20} \approx 1.5$ s, together with a small proportion of $S_{20} \approx 3.0$ s. These components corresponded to the α and β chains, and the ' $\alpha\beta$ dimer' form seen in electrophoresis.

Fig. 7 also shows the ultracentrifugal patterns of a native arachin, prepared so as to be predominantly in the 'monomer' form. A component of $S \approx 3.6$ s is present, and this corresponds in electrophoresis to the ' $\alpha\beta$ dimer' form. The sensitivity of electrophoresis is plainly greater than that of the ultracentrifuge, and the individual α and β chains could just be detected in the gel, but not in the ultracentrifuge, even from native arachin.

Separation of arachins A and B. Since the amino acid compositions of arachins A and B were to be compared it was necessary to separate either the α and β chains, or the intact proteins. The α and β chains were insoluble in such buffers as ammonium acetate or formate, and, although a run was possible in sodium phosphate, pH 8, on DEAE-Sephadex, the fractionation was not good. Association of the

Fig. 6. Electrophoresis of depolymerized arachins A and B. Runs 1-5 were in 10% gels, runs 6 and 7 in 12.5% gels. Run 1 shows the result obtained with the total protein from a seed containing arachin B only; run 2 is the same but a seed containing both arachins. Runs 3-5 show similar experiments with isolated arachins: AB in run 3, A in run 4 and B in run 5. Note the presence of β chains in A and the absence of α chains in B. Run 6 is of arachin B and 7 of arachin A, in stronger gels, to show the relative retardation of the α and β chains compared with the γ and δ chains.

sub-units was also a problem. Therefore the intact proteins were used.

The usual approach in a polymorphism is to prepare the separate proteins from the individual samples. This unfortunately failed in this case because only one purely A seed was found, and so

insufficient protein was available. Nevertheless a sample of arachin B was prepared from six seeds containing only arachin B.

The separation of arachin A from arachin B was complicated by the existence of various polymer forms, the properties of which tended to differ more

Fig. 7. Comparison of ultracentrifuge and electrophoretic patterns for various arachin preparations. Direction of sedimentation: right to left. Direction of electrophoresis: left to right. I, Native arachin prepared in the 'monomer' form, run in tris-citrate electrophoresis buffer, pH 8.6. II, Arachin denatured in 0.5M-acetic acid, then transferred to tris-citrate. III, Alkali-denatured, then neutralized to pH 8 with HCl. IV, Denatured at pH 12 and run at that pH. Note the presence of the 3-4s component in I, III and IV and the corresponding zone in electrophoresis.

than the difference between A and B. This is doubtless why chromatography on DEAE-Sephadex gave no fractionation. Fortunately, a good preparation of arachin A was obtained by fractional precipitation by dialysis of a solution of arachin AB in 10%

sodium chloride against water in a narrow dialysis sac. Arachin A remained in solution in its 'monomer' form, and arachin B was precipitated.

Carbohydrates. Arachin contained less than 0.05% of carbohydrate and is apparently a simple

unconjugated protein. Whole groundnut protein contained mannose, in agreement with Dechary (1961) who reports 1% of mannose in α -conarachin. Galactose, glucose, ribose and a trace of xylose were also present. Arabinose was present in the extract but was not protein-bound.

Amino acid analyses. The results after various times of hydrolysis are shown in Table 2, and the corrected values, expressed as residues in 330 000 g. of protein. They are in agreement with the partial analysis reported by Tristram (1953). The main interest was in a comparison of arachins A and B, and it was expected that these two proteins would be very similar. By analogy with other polymorphisms the α and β chains were expected to differ by one, or possibly two, amino acids (see for example Piez, Davie, Folk & Gladner, 1961, on β -lactoglobulin).

Two samples of arachin A were used, one of which (A') was available in sufficient amount for 24 hr. hydrolysis only. The other was, as judged by electrophoresis, contaminated by about 3% of β -conarachin. The A' and B samples contained only a trace of contaminant.

A comparison of sample A' with the 24 hr. hydrolysis of sample B is shown in Table 2. The two samples were very similar in amino acid composition. A comparison of sample A with B reveals larger differences, though the two proteins still resemble each other closely.

N-Terminal amino acids. Three *N*-terminal amino acids were detected by the fluorodinitrobenzene technique. They were glycine, valine (?) and glutamic acid, in the proportions 8:1.9:1.3. In addition traces of other DNP-amino acids, particularly leucine (isoleucine was not tested for), were found. There was no doubt that glycine was by far the major *N*-terminal component. Arachin is a particularly difficult case for the DNP method, which tends to lose minor *N*-terminal amino acids among the artifacts that always appear at the hydrolysis stage.

For confirmation, the Edman degradation was therefore used. This again showed that glycine was the major *N*-terminal component, but glutamate and isoleucine were the only other amino acids detected. In this instance the isoleucine derivative was chromatographically more like valine than leucine, which could cause some confusion. However, valine and leucine were absent. It is concluded that, at least for arachin, the phenyl isothiocyanate method is the more satisfactory for the determination of the *N*-terminal amino acids, which are glycine, glutamate and isoleucine. Approximately six end groups were found in 330 000 g. of protein, but the reaction was incomplete.

Thiol groups: ultraviolet-absorption spectra. Two main methods were attempted for thiol group

determination. The modification of the method of Boyer (1954) described by Fridovich & Handler (1957) proved satisfactory for arachin because it does not require measurement on the original protein solution. This was usually impracticable at the necessary wavelengths. Arachin in phosphate, pH 6, contained 6.5 reactive thiol groups per 330 000 g. of protein. The other method was that based on the analysis of u.v. spectra described by Donovan (1963). The difference spectrum of arachin in 8M-urea, pH 7, and in 8M-urea, pH 10.1, was measured. It showed the usual two peaks (245 and 296 m μ), which, from the constants given by Donovan, indicated the presence of 24 ionized phenolic groups and no thiol groups. After 24 hr. exposure to these solvents the same result was obtained. A similar difference spectrum but between 8M-urea, pH 11.2, and 8M-urea, pH 7, as solvents again showed no thiol group ionization, and only 40 phenolic groups ionized.

Analysis of the absorption curve from 230 to 370 m μ of arachin dissolved in 0.1M-phosphate adjusted to pH 10.5 showed a peak at λ_{max} 281.5 m μ , where $E_{1\text{cm}}^{1\%} = 8.8$. Tyrosine was 2.7 g. and tryptophan 2.18 g./100 g. of protein, calculated in the usual way (Goodwin & Morton, 1946).

DISCUSSION

Isolation and electrophoretic characterization. Because whole seeds were used in this work there was more difficulty in extracting the protein than was experienced by previous workers who used meal. This was due to the fat-protein emulsion formed during homogenization, and although prior extraction of lipid with organic solvents abolishes this problem, it raises others, particularly the effect of such solvents on the proteins. In the event the protein from meal and from seeds was identical. Indeed arachin seems comparatively resistant to the denaturing action of organic solvents.

A simple method of preparing arachin by means of calcium precipitation was discovered. Calcium precipitation is used in many traditional methods for processing seed proteins, but this method is unusual in its specificity. The product was very pure, and all the arachin was precipitated, making the isolation of the conarachin components a simple matter. Salting-out methods always left them heavily contaminated with arachin.

Electrophoresis, and a combination of electrophoresis and DEAE-Sephadex chromatography, revealed the complexity of the arachin and conarachin fractions. The arachin complexity is partly due to its ability to polymerize, though there are two distinct arachins. Conarachin has all the characteristics of a complex mixture of different

Table 2. *Amino acid analysis of arachins A and B*

Analysis values are the means of duplicate runs at each time of hydrolysis (24, 36 and 48 hr.).

Amino acid	Time of hydrolysis (hr.) ... 24			Analysis (μ moles/mg. of protein)						Composition (residues/330 000 g. of protein)					
	24			36			48			Corrected			24		
	A'	A	B	A	B	A	A	B	A	A	B	A	A'	B	B
Aspartate	1.0465	1.0396	1.0516	1.0417	1.0940	1.0218	1.1110	1.0806	1.1121	1.1110	1.1121	366.3	345.3	347.0	366.9
Threonine	0.2249	0.1227	0.2238	0.1696	0.2247	0.1617	0.1910	0.2181	0.2800	0.1910	0.2800	63.0	74.2	73.8	92.4
Serine	0.5090	0.4956	0.5153	0.5363	0.5162	0.4706	0.5670	0.4948	0.5920	0.5670	0.5920	187.1	167.9	170.0	195.3
Glutamate	1.4653	1.5042	1.4365	1.6551	1.5161	1.6362	1.7110	1.4920	1.6400	1.7110	1.6400	564.6	480.6	474.0	541.2
Proline	0.4254	0.3238	0.4264	0.4016	0.4437	0.3964	0.4016	0.4401	0.4437	0.4016	0.4437	132.5	140.4	140.7	146.4
Glycine	0.5750	0.5675	0.6136	0.5923	0.6240	0.5804	0.5923	0.6128	0.6240	0.5923	0.6240	195.4	189.7	202.4	205.9
Alanine	0.4914	0.3355	0.5282	0.3973	0.5466	0.3913	0.3973	0.5530	0.5530	0.3973	0.5530	131.1	162.2	174.3	182.4
Cystine*	0.0384	0.0403	0.0217	0.0828	0.0364	0.0608	0.0828	0.0287	0.0364	0.0828	0.0364	27.3	12.7	7.8	11.0
Valine	0.3915	0.1590	0.3682	0.3062	0.4057	0.4113	0.3357	0.4113	0.4113	0.3357	0.4113	110.8	129.2	121.5	135.7
Methionine*	0.0338	0.0820	0.0210	0.1149	0.0216	0.0940	0.0115	0.0142	0.0216	0.0115	0.0216	3.8	11.2	6.9	7.8
Isoleucine	0.2786	0.0966	0.2826	0.2340	0.3030	0.2525	0.2525	0.3151	0.3151	0.2525	0.3151	83.3	91.9	93.2	103.9
Leucine	0.5886	0.4171	0.6145	0.5961	0.6358	0.5889	0.5961	0.6420	0.6420	0.5961	0.6420	196.7	194.2	202.7	211.8
Tyrosine	0.3091	0.1698	0.2977	0.2249	0.3094	0.2103	0.2800	—	0.3094	0.2800	0.3094	92.4	102.0	98.2	102.1
Phenylalanine	0.3986	0.1854	0.3966	0.2958	0.4157	0.2899	0.3130	0.4345	0.4345	0.3130	0.4345	103.3	131.5	130.8	143.4
Lysine	0.1748	0.1513	0.1752	0.2443	0.1918	0.2398	0.2580	0.1881	0.2120	0.2580	0.2120	85.1	57.7	57.8	69.9
Histidine	0.1735	0.0970	0.1516	0.1654	0.1629	0.1654	0.1654	0.1654	0.1654	0.1654	0.1654	54.6	57.3	50.0	54.6
Arginine	0.7668	0.5971	0.7952	0.8935	0.8351	0.8806	0.9210	0.8007	0.9200	0.9210	0.9200	304.0	253.0	262.0	304.0
Tryptophan*												35	35	35	35
												2670	2577	2598	2856

* Methionine was partly converted into sulphoxides and these were corrected for. Tryptophan was estimated from the u.v. absorption of an AB arachin sample. Traces of cysteine acid were not corrected for. The totals exclude amino acids marked *.

proteins, though some of them may actually be sub-units of arachin.

Acrylamide-gel electrophoresis is the most effective method for the qualitative analysis of protein at present available. It is so discriminating that it is unlikely that material which gives only a single zone by this method contains more than one component. The danger is the converse: the numerous zones corresponding to various polymers of arachin are all part of a single component (in the Phase-Rule sense). At first some of them were mistaken for distinct components and some care is necessary before equating the number of components with the number of zones seen in the gel.

The effect of molecular size on electrophoretic mobility can be conveniently investigated by varying gel concentration at a constant pH. It is now generally held that the high resolving power of starch gels is due to a filtration process and the same is true of acrylamide gels. Thus it is possible to account for the way in which mobility varies with gel concentration purely on the basis of varying pore sizes in the gel, and by making allowance for a distribution of pore size in the gels, to account for the behaviour at extremes of mobility and gel strengths.

The significance of the 'pore size' at an extrapolated zero mobility is not clear. This constant must bear some relationship to the volume of the molecule and indirectly to the molecular weight. At present the structure of the gels is not sufficiently known to permit a quantitative treatment, though it is hoped that electron-microscope studies will help. It was possible to make deductions about the relative sizes of the various sub-components of arachin and this could be done for most mixtures of proteins, but it is emphasized that systematic relationships of this kind can often be shown for polymeric series, or carefully selected groups of proteins, but do not hold in the general case.

It was expected that the gel-electrophoresis patterns and ultracentrifuge patterns would be similar. In general they were but clearly the electrophoretic method has a higher resolving power, and it was possible to detect components in native arachin, for example, that were not detectable in the ultracentrifuge. Although polyacrylamide-gel electrophoresis has much to recommend it, its major drawback should be mentioned. The method, like all zone methods, is only semi-quantitative and it is difficult to see how this can be overcome at present.

Structure of arachins A and B. It has been known for some years that arachin could exist in a 'monomer' and 'dimer' form. The gel-electrophoresis experiments described above show, however, that the 'monomer' breaks down, even in arachin which can be regarded as native, to give

several sub-units. That arachin breaks down to smaller units in alkali or urea denaturation was already established (Johnson & Shooter, 1950) and the object of the present depolymerization experiments was to determine the number of kinds of peptide chain in arachin, and if possible also the number of chains. The position was complicated because there is a variant form of arachin, present in my preparations and probably those of all previous workers.

It was possible to deduce, from the following principal data, a probable structure for both arachins.

(1) Arachin A produces α , β , γ and δ chains. Arachin B produces β , γ and δ chains.

(2) Arachin 'monomer' has mol. wt. 170 000 and 'dimer' 330 000 (Johnson & Shooter, 1950). Alkali-denatured arachin has mol. wt. about 30 000 (Cater & Naismith, 1958).

(3) There is an intermediate form (' $\alpha\beta$ dimer') between the α or β chains and the 'monomer'. This form was demonstrated by both electrophoresis and ultracentrifugation.

Ignoring, for the moment, the γ and δ chains, arachin B exists, in solution, as:

$$2xn\beta \rightleftharpoons \frac{2x\beta_n}{\beta_2 \text{ dimer}} \rightleftharpoons \frac{2\beta_{2n}}{\text{'monomer'}} \rightleftharpoons \frac{\beta_{2n}}{\beta_{2n} \text{ dimer}}$$

and the only reasonable values for n and x are 2; i.e.:

$$8\beta \rightleftharpoons \frac{4\beta_2}{\beta_2 \text{ dimer}} \rightleftharpoons \frac{2\beta_4}{\text{'monomer'}} \rightleftharpoons \frac{\beta_8}{\beta_8 \text{ dimer}}$$

Arachin A, similarly, must be:

$$4\alpha + 4\beta \rightleftharpoons \frac{4\alpha\beta}{\alpha\beta \text{ dimer}} \rightleftharpoons \frac{2\alpha_2\beta_2}{\text{'monomer'}} \rightleftharpoons \frac{\alpha_4\beta_4}{\text{'dimer'}}$$

The γ and δ chains are quantitatively minor components, and if 'monomer' contains one then the full structure of arachin A is $\alpha_4\beta_4\gamma_2\delta_2$ and that of arachin B is $\beta_8\gamma_2\delta_2$. Provided that the 'dimer' gives two identical 'monomers' (and all the evidence is that it does) there cannot be fewer γ and δ chains than this, though there could be more.

It is simple to calculate that on the basis of the above structure the α and β chains have mol. wts. about 35 000 and the γ and δ chains about 10 000. This is entirely consistent with their behaviour in gel electrophoresis, which indeed predicted a value of about 10 000 for the γ and δ chains.

Finally, the N -terminal amino acids show that the α and β chains must have glycine at the N -terminal whereas the γ and δ chains have glutamate and isoleucine. The relative proportions of the N -terminal amino acids were also consistent with the proposed structures, and although they cannot be regarded as proved until the individual chains have been isolated and characterized they cannot be seriously in error.

It is clear from experiments in which arachin was depolymerized in the absence of mercaptoethanol that disulphides do not play a part in holding the α , β , γ and δ chains together; however, there may be intramolecular disulphides on any of these chains. Only six of some 24 possible thiol groups could be titrated, though it is likely that in more extreme conditions more free thiol groups would have been detected. More remarkably, no thiol groups could be found by a u.v. method, and the number of ionizable phenolic groups was much smaller than was expected from the amino acid composition. The essential precondition of this method, that the protein should be completely denatured and the titration curves normal, was not met. Arachin is very resistant to urea denaturation, but in any case the sub-units could be held in a tight configuration by intramolecular disulphide bonds, effectively burying the phenolic and free thiol groups. Such a configuration would also invalidate the use of the method of Donovan (1963).

The major discrepancy between these results and those of previous workers is in the *N*-terminal amino acids. Cater *et al.* (1957) reported that native arachin contained 1–2 glutamate, 2 valine, 3–4 leucine and 3–4 glycine *N*-terminal residues. They used a version of the fluorodinitrobenzene method which is now known to give poor recoveries, particularly of DNP-glycine, and this is probably the main cause of the difference. I found that many DNP-amino acids could be detected in the final extract and it was difficult to distinguish between minor genuine *N*-terminal amino acids and artifacts. The phenyl isothiocyanate method gave much more definite results, and differed from the final result with fluorodinitrobenzene only in substituting isoleucine for a tentative valine. The *N*-terminal amino acid determinations had an added interest because the only way in which the results could be reconciled with Cater *et al.* (1957) was if the α and β chains had differing *N*-terminal amino acids. Since these are a pair of variant chains this was always unlikely, and it now appears that they do both have glycine at the *N*-terminal position.

Variant forms of arachin. When dealing with proteins isolated from a pool of a number of individual samples it is always desirable to examine individual samples themselves, to see if there is variation. Variant forms of arachin were duly found in this case, and have been briefly reported (Tombs, 1963). Five types of seed were distinguished, whose protein ranged from arachin A only to arachin B only. The real difference between these proteins lies in the substitution of α for β chains in the variant form and Table 1 shows the relative proportions of the α and β chains, taking all the arachin (both A and B) into account.

The results suggest that β chains are essential

for the formation of the arachin molecule; a form containing only α chains has not been seen, and arachin A seems to depolymerize more readily than arachin B. Presumably the β chain is the primitive form, and the α chain the result of a mutation. Since both chains occur in quite different strains of groundnuts the mutation probably occurred in the original variety from which these strains derive.

I believe that this variation is the first example of its kind in flowering plants, and can be compared with, for example, the polymorphisms of haemoglobin, or β -lactoglobulin, in animals.

The amino acid analyses were undertaken to discover, if possible, in what ways the amino acid composition of the α and β chains differed. They show, first, that the amino acid compositions are very similar: a comparison of the best available samples showed that the difference between them is less than 3% of the residues. This is consistent with the view that the two proteins are a pair of variants and suggests that the analogy with other polymorphisms is sound.

Quite apart from the precision of the amino acid analyses an important source of error here arose from impurities in the original samples. The two proteins could not be prepared by exactly the same method so that impurities cancelled out, and the method used for isolating arachin A tended to concentrate contaminants into it.

These contaminants were sugar-containing proteins and this may explain the large losses of several amino acids found on more extensive hydrolysis. For this reason the comparison between samples A' and B shown in Table 2 is probably more reliable than that between samples A and B. These results can be used to calculate differences between the α and β chains, but such results must be of doubtful validity in view of the larger differences found with the other sample.

The conclusion must be that arachin A and arachin B are very similar to each other. They have a complex structure involving three or four kinds of peptide chain and display properties associated with multi-stranded proteins.

I am most grateful to Mr B. Giles for ultracentrifuge experiments, Mr J. Taylor for valuable discussions and Miss D. Davis and Mr A. C. Jackson for technical assistance.

REFERENCES

- Boyer, P. D. (1954). *J. Amer. chem. Soc.* **76**, 4331.
- Bragg, P. D. & Hough, L. (1961). *Biochem. J.* **78**, 11.
- Cater, C. N. & Naismith, W. E. F. (1958). *Arch. Biochem. Biophys.* **77**, 98.
- Cater, C. N., Naismith, W. E. F., Thompson, R. H. K. & Ure, G. R. (1957). *Nature, Lond.*, **180**, 971.
- Dechary J. M. (1961). *Nature, Lond.*, **190**, 1125.

- Dieckert, J. W., Snowden, J. E., Moore, A. T., Heinzelman, D. C. & Altschul, A. M. (1962). *J. Food Sci.* **27**, 321.
- Donovan, T. W. (1963). *Biochemistry*, **3**, 67.
- Fraenkel-Conrat, H., Harris, I. & Levy, A. L. (1955). *Meth. biochem. Anal.* **2**, 359.
- Fridovich, I. & Handler, P. (1957). *Analyt. Chem.* **29**, 1219.
- Goodwin, T. W. & Morton, F. A. (1946). *Biochem. J.* **40**, 628.
- Johns, J. & Jones, D. B. (1916). *J. biol. Chem.* **28**, 77.
- Johnson, P. (1946). *Trans. Faraday Soc.* **42**, 28.
- Johnson, P. & Naismith, W. E. F. (1956). *Trans. Faraday Soc.* **52**, 280.
- Johnson, P. & Shooter, E. M. (1950). *Biochim. biophys. Acta*, **5**, 36, 376.
- Jones, D. B. & Horn, M. J. (1930). *J. agric. Res.* **40**, 672.
- Partridge, S. M. (1949). *Nature, Lond.*, **164**, 443.
- Piez, K. A., Davie, E. W., Folk, J. E. & Gladner, J. A. (1961). *J. biol. Chem.* **236**, 2912.
- Raymond, S. & Nakamichi, M. (1962). *Analyt. Biochem.* **3**, 23.
- Spackman, D. H., Stein, W. H. & Moore, S. (1958). *Analyt. Chem.* **30**, 1190.
- Stevens, F. S. & Tristram, G. R. (1962). *Biochem. J.* **83**, 240.
- Tombs, M. P. (1963). *Nature, Lond.*, **200**, 1321.
- Tombs, M. P. & Apsey, M. (1962). *Biochem. J.* **86**, 111.
- Tombs, M. P., Cooke, K. B., Borston, D. & MacLagan, N. F. (1961). *Biochem. J.* **80**, 284.
- Tristram, G. R. (1953). In *The Proteins*, vol. 1A, p. 223. Ed. by Neurath, H. & Bailey, K. New York: Academic Press Inc.
- White, M. (1960). *J. phys. Chem.* **64**, 1563.