PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

The Chemistry of Legume Storage Proteins

N. C. Nielsen

Phil. Trans. R. Soc. Lond. B 1984 **304**, 287-296
doi: 10.1098/rstb.1984.0024

Phil. Trans. R. Soc. Lond. B **304**, 287–296 (1984)
Printed in Great Britain

The chemistry of legume storage proteins

By N. C. Nielsen

U.S.D.A.-A.R.S. and the Department of Agronomy, Purdue University, W. Lafayette, Indiana 47907, *U.S.A.*

[Plates 1 and 2]

Evidence is summarized that provides a structural description of the glycinin subunits from soybeans and the structural genes that produce them. Each subunit consists of an acidic polypeptide component that is linked to a basic one by a single disulphide bond. The subunits are initially produced as a precursor that consists of amino acids for the signal sequence, followed by the sequence for first the acidic and then the basic polypeptide. Comparison of the nucleotide sequence for a subunit gene with the complete primary structure of that subunit permits identification of a number of post-transcriptional and post-translational events that occur during expression of the gene.

Seeds of cultivated soybean varieties grown in the United States contain about 40% protein on a dry weight basis. While many different proteins are found in the seed, only a few are present in large quantities. These major proteins include glycinin, β-conglycinin, lipoxygenase, Kunitz trypsin inhibitor, a number of low molecular mass protease inhibitors (the best studied of which is the Bowman–Birk inhibitor), lectin and urease. Of these, glycinin and β-conglycinin are the most prevalent and together account for about 70% of the seed protein (Hill & Breidenbach 1974). The others, which are present at high levels compared with many metabolic proteins, generally each account for only 2–5% of the total protein.

In the case of the two major seed proteins, glycinin and β-conglycinin, the former is usually present in the largest amount. An enzyme linked immunosorbent assay was developed in my laboratory to determine the relative amounts of these two proteins (Medeiros 1982). While the mean glycinin: β-conglycinin ratio was found to be 1.6:1, the ratio ranged from 3:1 to 1:1 in different cultivars.

The general features (table 1) of the two major soybean storage proteins are similar to those of legumin and vicilin from pea (Derbyshire *et al.* 1976). Glycinin sediments at 12.2 *S*, which is similar to legumin, while β-conglycinin is a 7.9 *S* protein that corresponds to vicilin. The molecular masses reported for β-conglycinin range between 150000 to 200000 Da, and those for glycinin are generally between 310000 and 350000 Da. As expected from these large molecular masses, both glycinin and β-conglycinin are complex proteins. β-Conglycinin as it is generally isolated, consists of three nonidentical subunits (Thanh & Shibasaki 1978). However, at low ionic strength the β-conglycinin trimers dimerize to form hexamers. At least four different subunit groups, denoted α′, α, β and γ, make up β-conglycinin. Studies by Thanh and others demonstrated that β-conglycinin could be subfractionated into protein complexes that consisted of different combinations of the subunits. Since all possible combinations have been observed, it appears that each subunit is equivalent and that the complex is a random assembly of the subunits.

In the case of glycinin, hexamers are observed when the protein is purified from dilute salt solutions (Badley *et al.* 1975). So far five different major subunit groups have been identified. Like β-conglycinin, glycinin exhibits compositional polymorphism (Kitamura *et al.* 1980; Staswick & Nielsen 1983). Null alleles, which condition loss of specific subunits, have no obvious detrimental effect on the assembly of normal-sized glycinin molecules.

TABLE 1. THE GENERAL FEATURES OF THE TWO MAJOR SOYBEAN STORAGE PROTEINS

protein	size	molecular mass kDa	sub-units	sugar	sulphur, %
β-conglycinin	7.9 *S*	105–193	3	$Asn(NAGlc)_2(Man)_{7-9}$	0.6
glycinin	12.2 *S*	309–393	6	none	1.8

Several other important features distinguish the two molecules. β-Conglycinin contains about 5% sugar whereas none has been detected in highly purified glycinin. The sugar moieties are linked to asparagine and contain two central *N*-acetylglucose residues that are followed by 7–10 mannose molecules (Yamauchi *et al.* 1976; Yamauchi & Yamagishi 1979). Finally, as is true of the major legume storage proteins in general, they contain low levels of the sulphur amino acids, methionine and cysteine. Methionine is essential in monogastric nutrition, and glycinin generally contains more of this amino acid than β-conglycinin does (Derbyshire *et al.* 1976).

Five years ago we initiated a programme to purify and characterize the glycinin subunits, as well as to describe the steps involved in their biosynthesis. Before summarizing our results, however, I would like to emphasize a difference between the nomenclature we use to describe glycinin subunits, and that found in literature. Each of the six glycinin subunits consists of a pair of components that are linked by disulphide bonds. One component of the pair has an acidic isoelectric point and in SDS-polyacrylamide gels generally exhibits an apparent molecular mass of around 40000. The other has a basic isoelectric point and $M_r \simeq 20000$. Early experiments made use of both denaturing and reducing conditions to disassemble the glycinin complex. The acidic and basic components separated under these conditions, and each erroneously came to be known as a subunit. As data from our laboratory and others have shown, the initial translation products from glycinin genes are single proteins that contain sequences for both the acidic and basic polypeptides. The disulphide linkages apparently form before cleavage of each precursor into the two component subunits. In the older literature, complexes of an acidic and a basic polypeptide are referred to as intermediate complexes (Kitamura *et al.* 1976). However, the term intermediate complex was coined to represent an association of an acidic and a basic polypeptide when it was thought that these two components came together randomly. Since our data, as well as that of others, have shown that basic assumption to be incorrect, use of the term intermediate complex should be discontinued in favour of the terms polypeptide components and subunits.

The initial objective of our studies was to purify each of the major glycinin subunits and establish the structural relations between them. Defatted soybean meal, prepared immediately before use, was extracted with a dilute salt solution and then glycinin was selectively precipitated by a combination of isoelectric- and cryoprecipitation. When done carefully, these steps result in glycinin preparations that are more than 90% pure (Moreira *et al.* 1979). Such preparations were denaturated in 6 M urea that contained 2-mercaptoethanol and the polypeptide

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

components were separated chromatographically with DEAE-Sephadex (figure 1*a*). The first peak to emerge from the column (F_1) contained a mixture of basic polypeptide components. The second peak (F_2) contained two polypeptides (Moreira *et al.* 1979), one large ($M_r \simeq 37000$) and one small ($M_r \simeq 10000$), that were separated by gel filtration. These two acidic

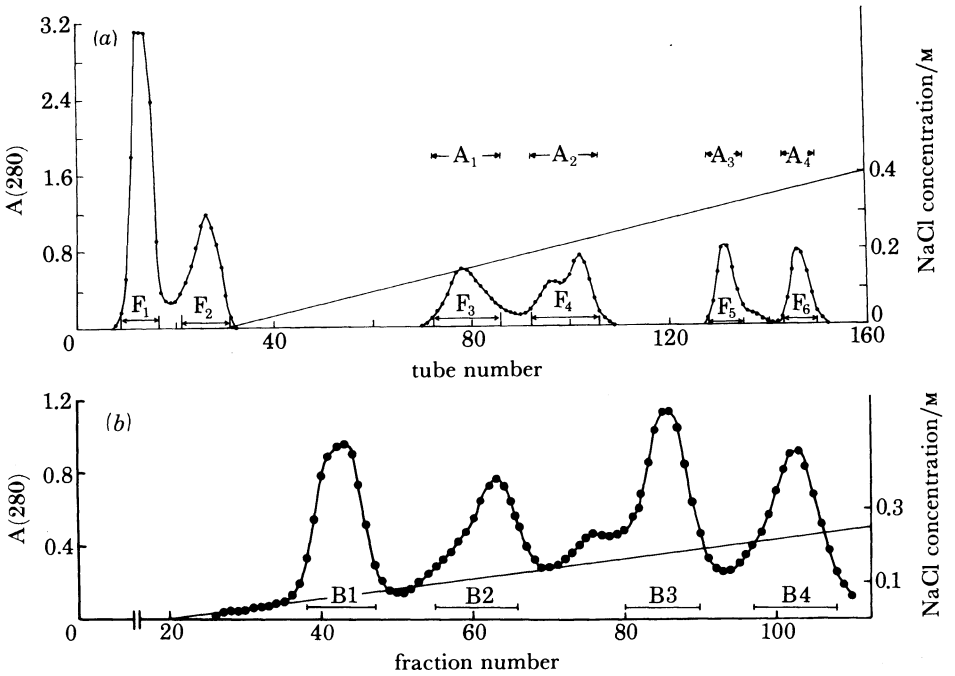


FIGURE 1. Chromatographic separation of glycinin polypeptide components. (*a*) Elution of reduced and denatured acidic polypeptides from DEAE-Sephadex. 500 mg of S-pyridylethylglycinin were placed on a column (2.6 by 40 cm) equilibrated with 0.1 M phosphate (pH 6.6) that contained 6 M urea and 0.02 M 2-mercaptoethanol. Fractions F_1 (basic components) and F_2 (A_{1b} and A_5) were eluted using the same buffer, and then the remaining ones (A_2, A_3, A_4) were eluted with a 1 l linear gradient between 0 and 0.4 M NaCl. Data from Moreira *et al.* (1979). (*b*) 200 mg of F_1 were placed on a CM-Sephadex column (2.6 × 40 cm) equilibrated with the buffer described above, and the basic polypeptide components were separated with a 1 l gradient from 0 and 0.3 M NaCl. The shoulder on the leading edge of the B3 peak contains B_{1c}.

components, as well as those in the three subsequent peaks were characterized by NH_2-terminal sequence analysis to establish the relation between them, and are now referred to as A_{1b}, A_5, A_{1a}, A_2, A_3 and A_4, respectively. With the exception of A_4, there was substantial homology between polypeptide components (table 2). However, the differences in primary structures were internal rather than limited to the ends. Differences limited to the ends of the molecules would be expected if the heterogeneity were merely the consequence of proteolysis after synthesis of the proteins. This showed that the polypeptide components were synthesized at the direction of a family of evolutionarily related genes. This conclusion was reinforced by subsequent studies in which substantial homology between internal primary structures of each of the acidic polypeptide components was demonstrated (Moreira *et al.* 1981).

The basic polypeptides from the F_1 fraction were separated chromatographically with CM-Sephadex (figure 1*b*). They were also characterized by NH_2-terminal sequence analysis (Moreira *et al.* 1979; Tumer *et al.* 1982), and have been denoted B_{1a}, B_{1b}, B_2, B_3 and B_4. The sequence homologies at the NH_2-terminal ends of these basic components were even more extensive than in the case of the acidic ones (table 2).

Given the disulphide linkages that occur between the acidic and basic polypeptide components of glycinin, the next point considered was which ones were associated. To clarify this, the intact subunits from non-reduced, denaturated glycinin were separated chromatographically with DEAE-Sephadex (Staswick *et al.* 1981). Since the NH_2-terminal sequences

TABLE 2. COMPARISON OF THE PHYSICAL PROPERTIES OF GLYCININ SUBUNITS

group	subunit	molecular mass / Da	no. Met codons	acidic	basic
I	$A_{1a}B_2$	58000	5–6	FSSREQAQQNECQIQKLNALKPD	GIDETICTMRLRQNIGQTSSPDIF
I	$A_{1b}B_{1b}$	58000	5–6	FSFREQPQQNECQIQKLNALKPD	GIDETICTMRLRQNIGQTSSPDIY
I	A_2B_{1a}	58000	7–8	FSLREQAQQNECQIQKLNALKPD	GIDETICTMRLRQNIGQNSSPDIY
				* **** **** **	* * **** * ** * *
II	A_3B_4	62000	3	ITSSKF del? NECQLNNLNALEPD	GVEENICTMKLHENIARP SWARFY
II	$A_5A_4B_3$	69000	3	ISSSKL del? NECQLNNLNALEPD	GVEENICT L KLHENIARP SWARFY

of each of the acidic and basic polypeptides were known, it was possible to identify the pairs by sequence analysis (table 2). In the case of B_{1a} and B_{1b}, their distinctive isoelectric focusing patterns served as additional criteria for assigning the pairing A_2B_{1a} and $A_{1b}B_{1b}$, respectively. The results of this study indicated that all acidic polypeptide components except A_4 were paired non-randomly to basic polypeptides. A_4 was not linked to any of the basic polypeptide components.

Two discrepancies were apparent at this point. First, A_5, a 10000 Da molecular mass component, was disulphide linked to the 20000 Da molecular mass B_3, component, to yield a 30000 Da molecular mass subunit rather than the approximately 60000 Da molecular mass of the other subunits. How did the special features of the A_5B_3-subunit contribute to glycinin structure? Second, A_4 was not paired to a basic polypeptide, and its NH_2-terminal sequence, while it was similar to the other acidic polypeptides, exhibited considerably less homology than the others. Both apparent discrepancies were resolved when we realized that A_5 and A_4 were part of the same acidic polypeptide component. Several lines of evidence support this conclusion. First, a single recessive null allele conditions the coordinate loss of A_5, A_4 and B_3 (Kitamura *et al.* 1983). Second, the NH_2-terminal sequence of A_4 is homologous to the primary structure of A_2, beginning at residue 100 (Staswick & Nielsen 1983). This explains the low NH_2-terminal sequence homology between A_4 and the acidic polypeptides. In addition, cleavage at residue 100 would yield an NH_2-terminal fragment of about 10000 Da molecular mass, and that is the same size as A_5. This accounts for the high degree of NH_2-terminal sequence homology between A_5 and the rest of acidic polypeptides. An $A_5A_4B_3$ complex would have an aggregate molecular mass of 69000 Da, which is more in line with the size of the other subunits. Also consistent with this interpretation is our recent observation (Staswick *et al.* 1983) that the perennial *Glycine tomentella* contains a glycinin subunit homologous to, and similar in size to, $A_5A_4B_3$ from *G. max*. However, unlike the one from the cultivated species, the acidic component in the subunit from the perennial is not cleaved proteolytically and does not separate into two components upon denaturation.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

FIGURE 2. Analysis of purified glycine polypeptides by isoelectric focusing. The gel contained 7.5% polyacrylamide (acrylamide: bis acrylamide, 75:2); 3.7% LKB ampholines; pH 3.5–10.0, in 6 M urea. 30 µg samples were applied on filter paper toward the cathode and focused 6 h at 0 °C and 800 V. Data from Moreira *et al.* (1981).

(*Facing p.* 290)

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

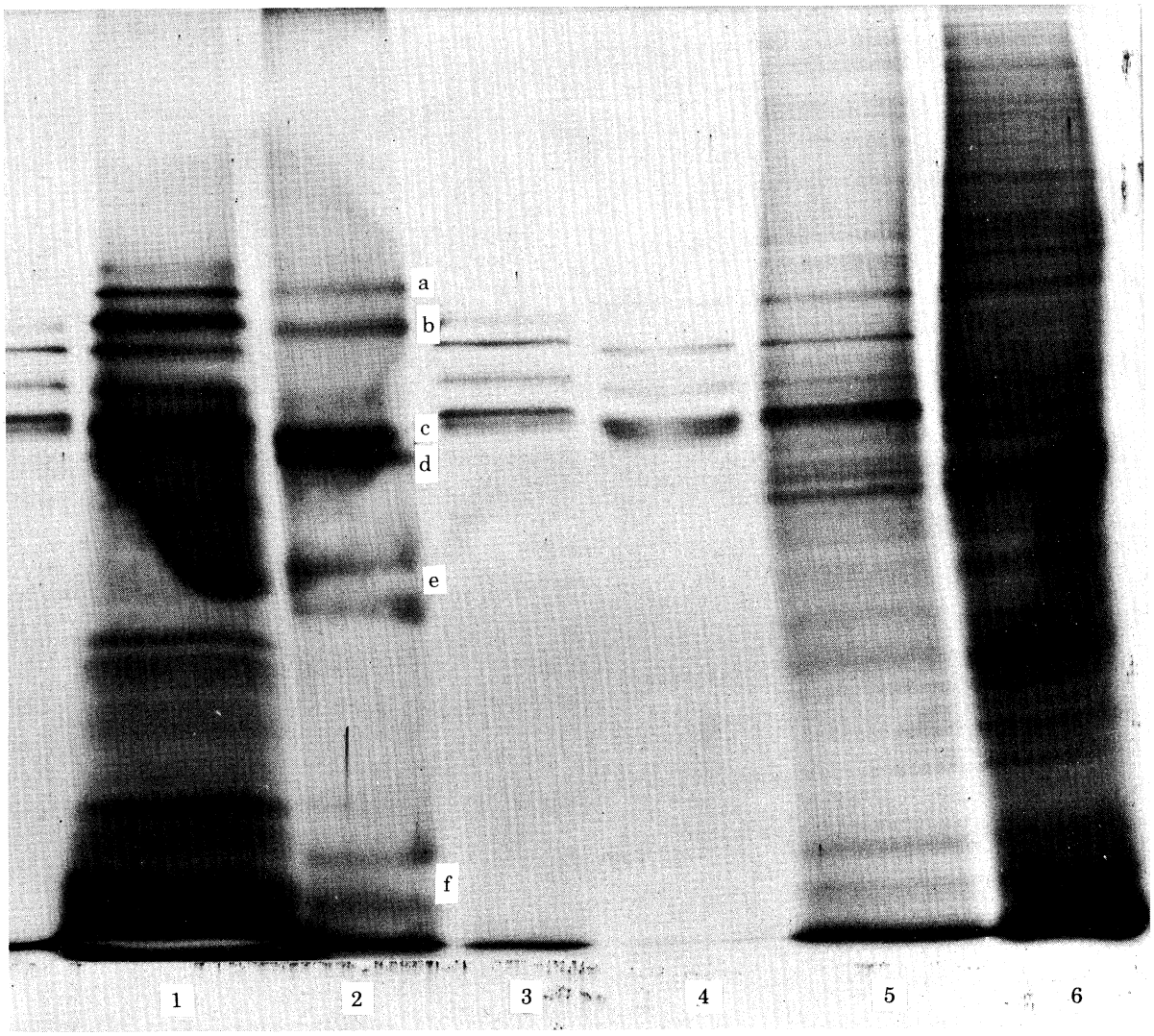


FIGURE 5. SDS polyacrylamide gel analysis of poly(A)-RNA translation products from oocyte and reticulocyte lysate translation systems. Lane 1, [^{3}H]leucine-labelled translation products from reticulocyte lysate. Lane 2, [^{3}H]dansylated standards where a denotes the α′-subunit of 75, $M_r = 75000$, b denotes α-subunit of 7 *S*, $M_r = 70000$; c denotes glycinin subunits, $M_r = 58000$; d is β-subunit of 7 *S*, $M_r = 54000$; e denotes acidic polypeptide components and f denotes the basic polypeptide components. Lane 3, translation products from the sample shown in lane 1 that were purified with antiglycinin IgGs. Lane 4, translation products from the sample shown in lane 5 that were purified with antiglycinin-IgGs. Lane 5, [^{3}H]leucine-labelled poly(A)-RNA translation products from *Xenopus* oocytes. Lane 6, [^{3}H]leucine-labelled proteins from *Xenopus* oocytes without injected mRNA. Refer to Tumer *et al.* (1982) for details.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

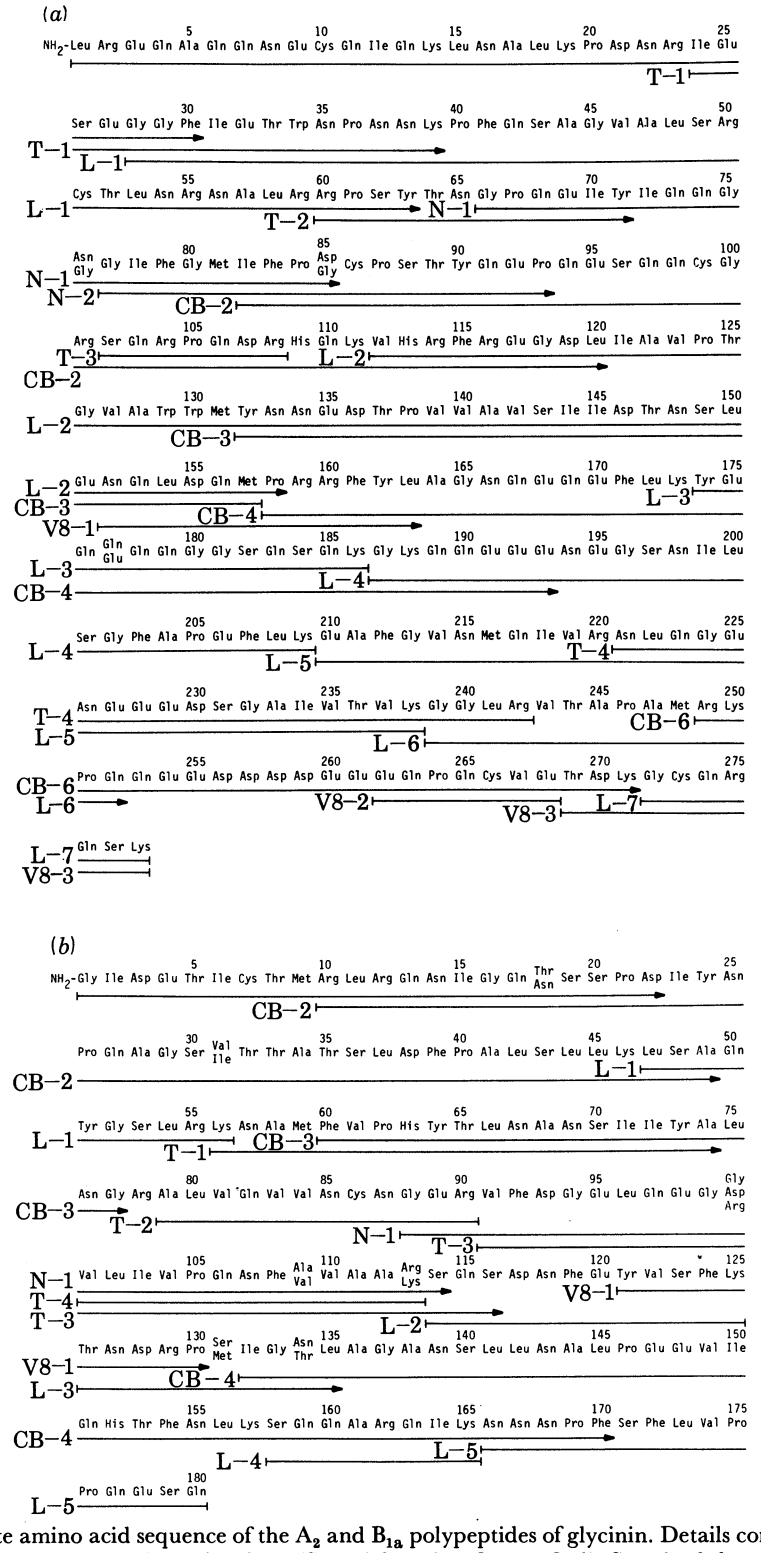


FIGURE 3. Complete amino acid sequence of the A_2 and B_{1a} polypeptides of glycinin. Details concerning sequencing strategy and recoveries are given elsewhere (Staswick *et al.* 1983*a*, 1983*b*). Standard three letter abbreviations are used to denote amino acids. Horizontal lines indicate regions sequenced for each peptide. The beginning and end of peptides are denoted with vertical bars. Arrowheads indicate that the peptide continues but was not sequenced beyond that point. Peptides were generated by cleavage with: CNBr, CB-1 to CB-6 for A_2 and CB-1 to CB-4 for B_{1a}; NH_2OH, N-1 and N-2 for A_2; trypsin on citraconylated polypeptides, T-1 to T-4 for A_2 and B_{1a}; *Staphylococcus aureus* V-8 protease, V8-1 to V8-3 for A_2 and V8-1 for B_{1a}; endoproteinase Lys-C, L-1 to L-7 for A_2 and L-2 to L-5 for B_{1a}.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

Comparison of the physical properties of the glycinin subunits revealed that they can be separated into two distinct groups (table 2). Sequence homology between subunits in the same group exceeds 85–90%, but is only about 50% between subunits in different groups. The subunits in group I are of a more uniform size and contain more methionine than those in group II. The latter observation may have practical significance in the improvement of seed nutritional quality if the group II subunits can be eliminated from glycinin and replaced by subunits from group I. The observations referred to earlier, which demonstrate compositional polymorphism, assume considerable importance in this context.

When the purified polypeptide components were subjected to isoelectric focusing in polyacrylamide gels, they exhibited charge heterogeneity (figure 2). Several simple explanations that would account for the heterogeneity were considered. First, each subunit preparation could be composed of several nearly identical polypeptides, each of which is the product of a different coding sequence. In this event amino acid substitutions would occur throughout the length of the molecules. A second possibility is that proteolytic modifications produce charge variants and in this event the differences between molecules in each preparation would be evident only at the ends of the molecules. The third possibility is that glutamine and asparagine could be deamidated.

To assess these possibilities, the complete primary structure for one of the subunits was determined. A_2B_{1a} was chosen because it contained more methionine residues than the other subunits, and the larger number of cyanogen bromide fragments potentially available would simplify sequencing. Figure 3, plate 1 summarizes the results from this study, and reflects two important features of the polypeptides in A_2B_{1a}. First, sequence heterogeneity was evident at a number of positions located throughout both polypeptide components of A_2B_{1a}. This clearly indicates that more than one coding sequence directs the synthesis of the A_2B_{1a} subunits. In many instances the heterogeneity involved charged amino acids, and these accounted, at least in part, for the charge heterogeneity observed by isoelectric focusing. One can speculate that a similar situation would be encountered for each subunit preparation, in which case a rather large number of coding sequences are responsible for producing the glycinin subunits.

While our data imply that several different coding sequences are involved in synthesizing A_2B_{1a} subunits, an unequivocal explanation about their genetic inter-relations cannot be made. The heterogeneity could reflect the expression of different genes. However, the hybridization studies reported by Goldberg *et al.* (1981) and by Fisher & Goldberg (1982), which indicate that there are only four or five glycinin genes, are contrary to this explanation. A more likely explanation is that the heterogeneity reflects expression of codominant alleles of the same gene. Codominant inheritance for alleles of storage protein genes has been reported for several legumes including soybean (Casey 1979; Hall *et al.* 1977; C. Davies and N. C. Nielsen, unpublished observations). If, as has been suggested, soybeans are ancient tetraploids, allelism could account for a substantial proportion of the heterogeneity observed.

Determination of the primary structures permitted direct measurement of molecular masses for A_2 and B_{1a}, and their comparison with values obtained by SDS-electrophoresis. The A_2 polypeptide component contains 278 amino acids, which reduces to a molecular mass of $31\,600 \pm 100$. The variation reflects the microheterogeneity in the sequence. This value is significantly less than the 37 000 Da determined by electrophoresis (Moreira *et al.* 1979). The reason for this discrepancy is not known. The B_{1a} polypeptide component contains 180 amino acids, which reduces to a molecular mass of $19\,900 \pm 100$. Unlike the acidic polypeptide

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

component, this value is identical to the 20000 Da determined by electrophoresis for the basic polypeptide components.

It was of interest to determine how many disulphide linkages join the acidic and basic polypeptide components, and their location in the subunit. This was accomplished by alkylating the free sulphydryl groups in each purified subunit. Then the disulphides involved in linking

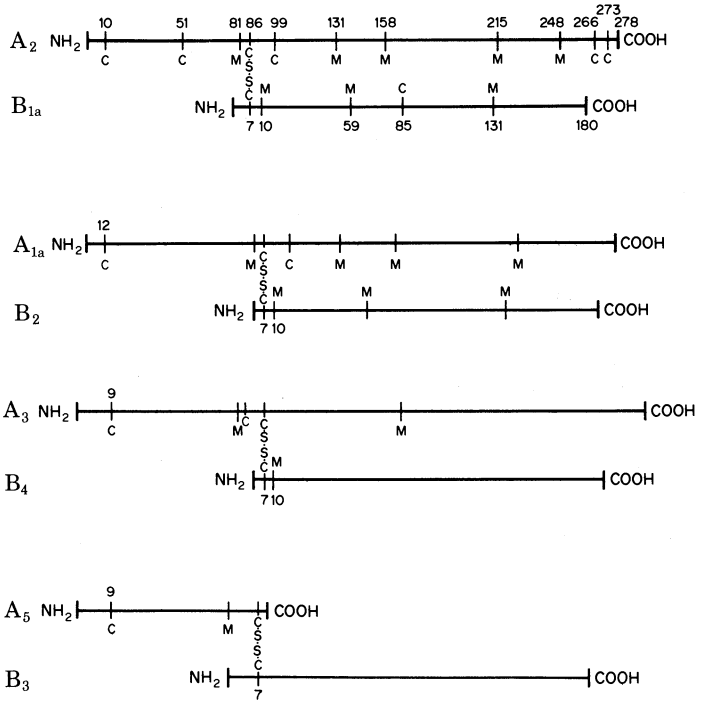


FIGURE 4. Relative position of the intermolecular disulphide bonds in the glycinin subunits. Horizontal lines represent polypeptides and are proportional in length to the estimated size of polypeptide. The residue number of methionines (M) and cystines (C) indicated are from the complete sequence for A_2B_{1a} shown in figure 4. The position of these residues in other subunits is estimated from the size of CNBr fragments and sequence homology with A_2B_{1a}. Additional cystine residues are present in $A_{1a}B_2$, A_3B_4 and A_5B_3, but their positions are not known with certainty. Intermolecular disulphides are indicated by –C–S–S–C–. Data is from Staswick *et al.* (1983*c*).

the acidic and basic polypeptides were reduced and alkylated with [^{3}H]iodoacetic acid. The acidic and basic polypeptides were separated by gel filtration and treated with cyanogen bromide. The resulting cyanogen bromide fragments were purified and those that were radioactive were microsequenced to identify the radioactive residues. Figure 4 summarizes the results obtained. A single interpeptide disulphide bond linked Cys_{86} in the acidic component and Cys_7 in the basic component of subunit A_2B_{1a} (Staswick 1982; Staswick *et al.* 1983*c*). A single interpeptide disulphide bond was observed at the same position in $A_{1a}B_2$, A_3B_4 and A_5B_3 as well. These results again emphasize the homologous nature of the glycinin subunits.

Recently a series of experiments has been completed that describe some of the post-transcriptional and post-translational events involved in the synthesis of the glycinin subunits. Initially we sought to determine whether or not the primary translation products of glycinin mRNA contain a signal sequence. Polyadenylated mRNA was purified (Tumer *et al.* 1981) from polysomes of midmaturation-stage soybean seeds. These messages were translated in a rabbit reticulocyte lysate system and glycinin translation products were purified by immunoaffinity chromatography. The purified translation products were separated

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

in SDS-polyacrylamide gels and four components were observed, all of which were larger than the authentic glycinin subunits used as standards (figure 5, plate 2). Identical results were obtained with monospecific antibodies against either the acidic and basic polypeptides. Since the size of the immunoreactive translation products was not affected by sulphydryl reducing agents, covalent bonds other than disulphide linkages connect the acidic and basic polypeptide components in the precursor.

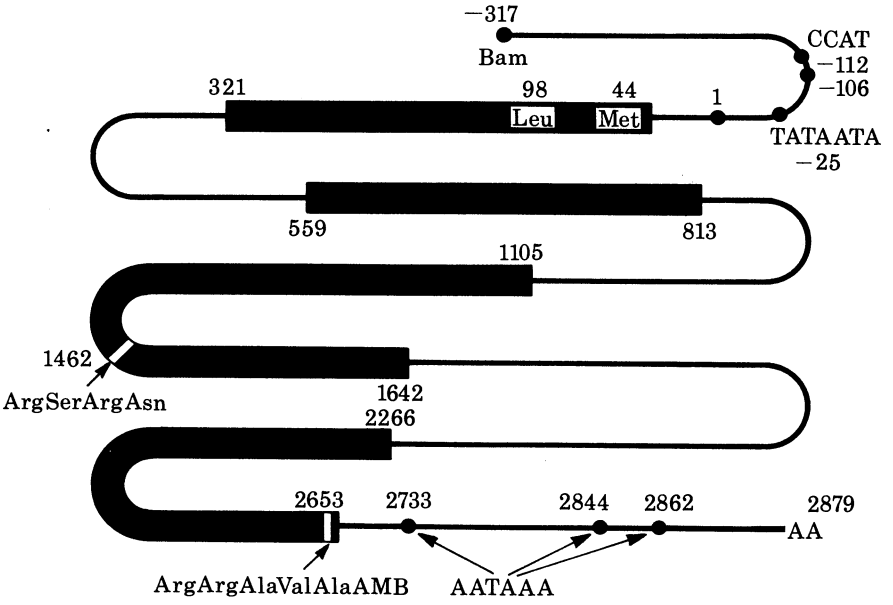


FIGURE 6. Schematic representation of the A_2B_{1a} gene in λDA28-30. The clone was isolated from a Alu-Hae III soybean genomic library by Fischer and Goldberg. Hind III fragments from this clone were subcloned into pBR325 mapped using standard techniques and then sequenced according to Maxam and Gilbert. Data from Vu *et al.* (1983).

Earlier studies with zein mRNA showed that *Xenopus laevis* oocytes correctly processed prolamine precursors (Larkins *et al.* 1979). Similar experiments were performed with glycinin mRNA (Tumer *et al.* 1982), and four immunoreactive translation products were again observed (figure 6). However, the oocyte translation products were 1000 to 2000 Da smaller than translation products from the reticulocyte lysate system. Analysis of the purified oocyte translation products by microsequence determination revealed that they had an NH_2-terminal sequence like the acidic polypeptides, whereas the NH_2-terminal sequence of products from the reticulocyte lysate system was a leucine-rich sequence that resembled neither the acidic nor basic polypeptide components. Thus, glycinin precursors consist of a signal sequence of about 20 amino acids, followed by the sequence of the acidic and then the basic component. While the signal sequence is presumably removed co-translationally, cleavage between the acidic and basic components probably occurs post-translationally.

Several genomic clones with coding sequences corresponding precisely to the primary structure of an A_2B_{1a} subunit, as well as clones derived from A_2B_{1a} mRNA, have now been analysed (Marco *et al.* 1983; Thanh *et al.* 1983). A scheme that summarizes the structure of a complete gene is shown in figure 6. The site indicated where initiation of transcription occurs was located by S-1 nuclease protection experiments. Consensus sequence for a TATA box is found about 25 base pairs (b.p.) upstream from this site, and a potential CCAT sequence is

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

located approximately 80 b.p. further upstream. The transcriptional unit covers 2799 b.p. and contains three introns of 238, 292 and 624 b.p. respectively. The nucleotide sequences at the intron–exon junctions correspond to consensus sequences established for other eukaryotic genes. The transcript contains a 43 b.p. 5′- and a 226 b.p. 3′-untranslated sequence. The 3′-untranslated region contains three potential polyadenylation signals. The polyadenylated tail of the A_2B_{1a} message used to generate clone pBS154 began 15–17 b.p. beyond the last of these sites (Thanh *et al.* 1983).

The nucleotide sequence of the gene predicts that an 18 amino acid signal sequence is present at the NH_2-terminal of the precursor. This conclusion is based on the assumption that translation begins at a Met codon, and the only one of these present between the site of transcription initiation and the NH_2-terminal of the mature subunit is the ATG codon at position 44. As is typical, the signal for this A_2B_{1a} gene is leucine rich. The position of these leucine residues is consistent with microsequence data for glycinin precursors produced *in vitro* (Tumer *et al.* 1982).

Comparison of the primary structures of the A_2 and B_{1a} polypeptide components with that of the precursor, predicted on the basis of the gene nucleotide sequence, reveals a four amino acid linker between the two components. Cleavage of the linker at the NH_2-terminal border apparently occurs between Lys_{278} and Arg_{279}. A similar type of cleavage takes place when the C-peptide is removed during the maturation of proinsulin, and a thiol-protease is involved in this modification (Docherty *et al.* 1982).

An explanation of the mechanism for cleavage at the other end of the linker is less straightforward. An amino peptidase could remove the linker amino acids. Alternatively a specialized protease may cleave between asparagine and glycine to form the NH_2-terminal of the basic polypeptide component. While an enzyme with this specificity is to our knowledge unknown, a viral protease that cleaves glutamine–glycine bonds in coat polyprotein precursors has been described (Kitamura *et al.* 1981).

Post-translational modification also occurs at the COOH-terminal end of the basic polypeptide component. A pentapeptide, as predicted by the nucleotide sequence, is absent in the mature subunit. An interesting similarity exists between this pentapeptide and the A_4 polypeptide. Both have a pair of arginine residues at their NH_2-terminal ends. As discussed earlier, A_4 appears to be separated from A_5 by a protease post-translationally. While the relation may be coincidental, the same protease may carry out both cleavages and the paired arginine residues may be involved in recognition of the cleavage site by the enzyme.

The data summarized is the result of work done by M. A. Moreira, P. E. Staswick, J. S. Medeiros, N. E. Tumer, Y. Marco, C. Davies, K. Kitamura and T. Vu while they were either graduate students or post-doctoral associates in the author's laboratory. Without their skillful efforts, this work would not have been possible. I also gratefully acknowledge Bob Fischer and Bob Goldberg who generously supplied several of the clones used to determine the complete nucleotide sequence of A_2B_{1a}.

References

Bradley, R. A., Atkinson D., Hauser, H., Oldani, D., Green, J. P. & Stubbs, J. M. 1975 The structure, physical and chemical properties of the soybean protein glycinin. *Biochim. biophys. Acta* **412**, 214–228.

Casey, R. 1979 Genetic variability in the structure of the α-subunits of legumin from *Pisum* – A two-dimensional gel electrophoresis study. *Heredity* **43**, 265–272.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

Derbyshire, E., Wright, D. J. & Boulter, D. 1976 Legumin and vicilin, storage proteins of legume seeds. *Phytochemistry* **15**, 3–24.

Docherty, K., Carroll, R. J. & Steiner, D. F. 1982 Conversion of proinsulin to insulin: involvement of a 31500 molecular weight thiol protease. *Proc. natn. Acad. Sci. U.S.A.* **79**, 4613–4617.

Fischer, R. & Goldberg, R. 1982 Structure and flanking regions of soybean seed protein genes. *Cell* **29**, 651–660.

Goldberg, R. B., Hoschek, G., Tam, S. H., Ditta, G. S. & Breidenbach, R. W. 1981 Abundance, diversity and regulation of mRNA sequence sets in soybean embryogenesis. *Dev. Biol.* **83**, 201–217.

Hall, T. C., McLeester, R. C. & Bliss, F. A. 1977 Equal expression of maternal and paternal alleles for polypeptide subunits of the major storage proteins of the bean *Phaseolus vulgaris* L. *Pl. Physiol.* **59**, 1122–1124.

Hill, J. E. & Breidenbach, R. W. 1974 Proteins of soybean seeds I. Isolation and characterization of the major components. *Pl. Physiol.* **53**, 742–746.

Kitamura, K., Davies, C. & Nielsen, N. C. 1984 Inheritance of alleles for Cgy_1 and Gy_4 storage protein genes in soybean. *Theor. appl. Genet.* (Submitted.)

Kitamura, K., Takagi, T. & Shibasaki, K. 1976 Subunit structure of soybean 11 *S* globulin. *Agric. Biol. Chem.* **40**, 1837–1844.

Kitamura, K., Toyokawa, Y. & Harada, K. 1980 Polymorphism of glycinin in soybean seeds. *Phytochemistry* **19**, 1841–1843.

Kitamura, N., Semler, B., Rothberg, P., Larsen, G., Adler, C., Dorner, A., Emini, E., Hanecak, R., Lee, J., van der Werf, S., Anderson, C. & Wimmer, E. 1981 Primary structure, gene organization and polypeptide expression of poliovirus RNA. *Nature, Lond.* **291**, 547–549.

Larkins, B. A., Pedersen, K., Handa, A. K., Hurkman, W. J. & Smith, L. D. 1979 Synthesis and processing of maize storage proteins in *Xenopus laevis* oocytes. *Proc. natn. Acad. Sci. U.S.A.* **76**, 6448–6452.

Marco, Y. A., Thanh, V. H., Tumer, N. E., Scallon, B. J. & Nielsen, N. C. 1983 Cloning and structural analysis of DNA complementary to the A_2B_{1a} subunit of glycinin. *J. biol. Chem.* (In the press.)

Medeiros, J. S. 1982 Characterization of the subunits of β-conglycinin and the application of ELISA to the determination of the contents of β-conglycinin and glycinin in soybean. Ph.D. Thesis. Purdue University. West Lafayette, Indiana.

Moreira, M. A., Hermodson, M. A., Larkins, B. A. & Nielsen, N. C. 1979 Partial characterization of the acidic and basic polypeptides of glycinin. *J. biol. Chem.* **254**, 9921–9926.

Moreira, M. A., Hermodson, M. A., Larkins, B. A. & Nielsen, N. C. 1981 Comparison of the primary structure of the acidic polypeptides of glycinin. *Arch. Biochem. Biophys.* **210**, 633–642.

Staswick, P. E., Hermodson, M. A. & Nielsen, N. C. 1981 Identification of the acidic and basic subunit complexes of glycinin. *J. biol. Chem.* **256**, 8752–8755.

Staswick, P. E. 1982 Characterization of the subunits of glycinin and the complete primary structure of the A_2B_{1a} subunit. Ph.D. Thesis. Purdue University, West Lafayette, Indiana.

Staswick, P. E., Broué, P. & Nielsen, N. C. 1983 Glycinin composition of several perennial species related to soybean. *Pl. Physiol.* **72**, 1114–1118.

Staswick, P. E., Hermodson, M. A. & Nielsen, N. C. 1984*a* The complete amino acid sequence of the A_2B_{1a} subunit of glycinin. *J. biol. Chem.* (In the press.)

Staswick, P. E., Hermodson, M. A. & Nielsen, N. C. 1984*b* Identification of the cystines linking the acidic and basic components of the glycinin subunits. *J. biol. Chem.* (In the press.)

Staswick, P. E. & Nielsen, N. C. 1983 Characterization of a soybean cultivar lacking certain glycinin subunits. *Arch. Biochem. Biophys.* **223**, 1–8.

Thanh, V. H. & Shibasaki, K. 1978 Major proteins of soybean seeds. Subunit structure of β-conglycinin. *J. Agric. Fd Chem.* **26**, 692–695.

Thanh, V. H., Fischer, R., Goldberg, R. B. & Nielsen, N. C. 1984 The nucleotide sequence of a Gy_2 storage protein gene in soybean. *Proc. natn. Acad. Sci., U.S.A.* (Submitted.)

Tumer, N. E., Thanh, V. H. & Nielsen, N. C. 1981 Purification and characterization of mRNA from soybean seeds: Identification of glycinin and β-conglycinin precursors. *J. biol. Chem.* **256**, 8756–8760.

Tumer, N. E., Richter, J. D. & Nielsen, N. C. 1982 Structural characterization of the glycinin precursors. *J. biol. Chem.* **257**, 4016–4018.

Yamauchi, F., Thanh, V. H., Kawase, M. & Shibasaki, K. 1976 Separation of the glycopeptides from soybean 7 *S* protein: Their amino acid sequences. *Agric. Biol. Chem.* **40**, 691–696.

Yamauchi, F. & Yamagishi, T. 1979 Carbohydrate sequence of a soybean 7 *S* protein. *Agric. Biol. Chem.* **43**, 505–510.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

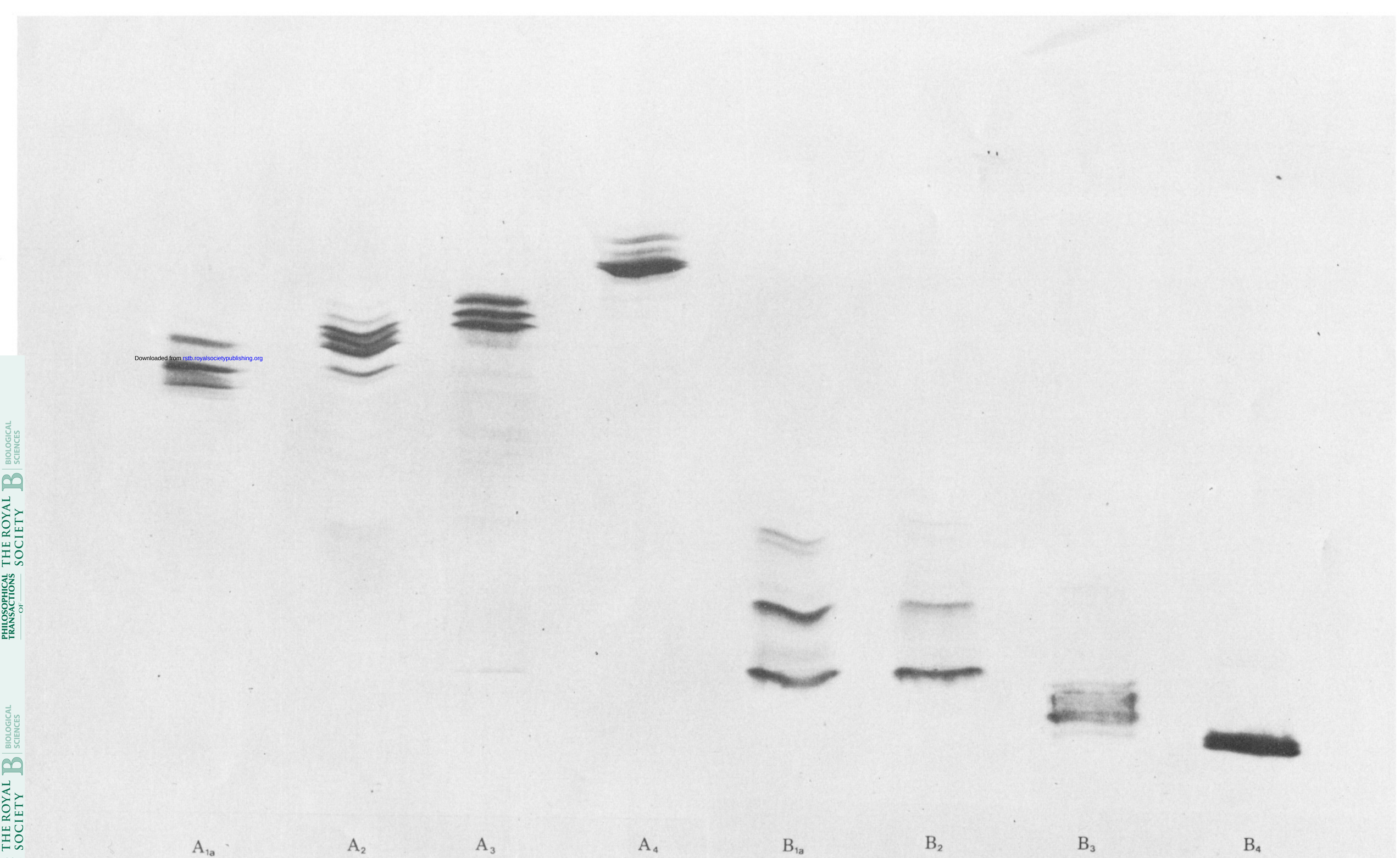
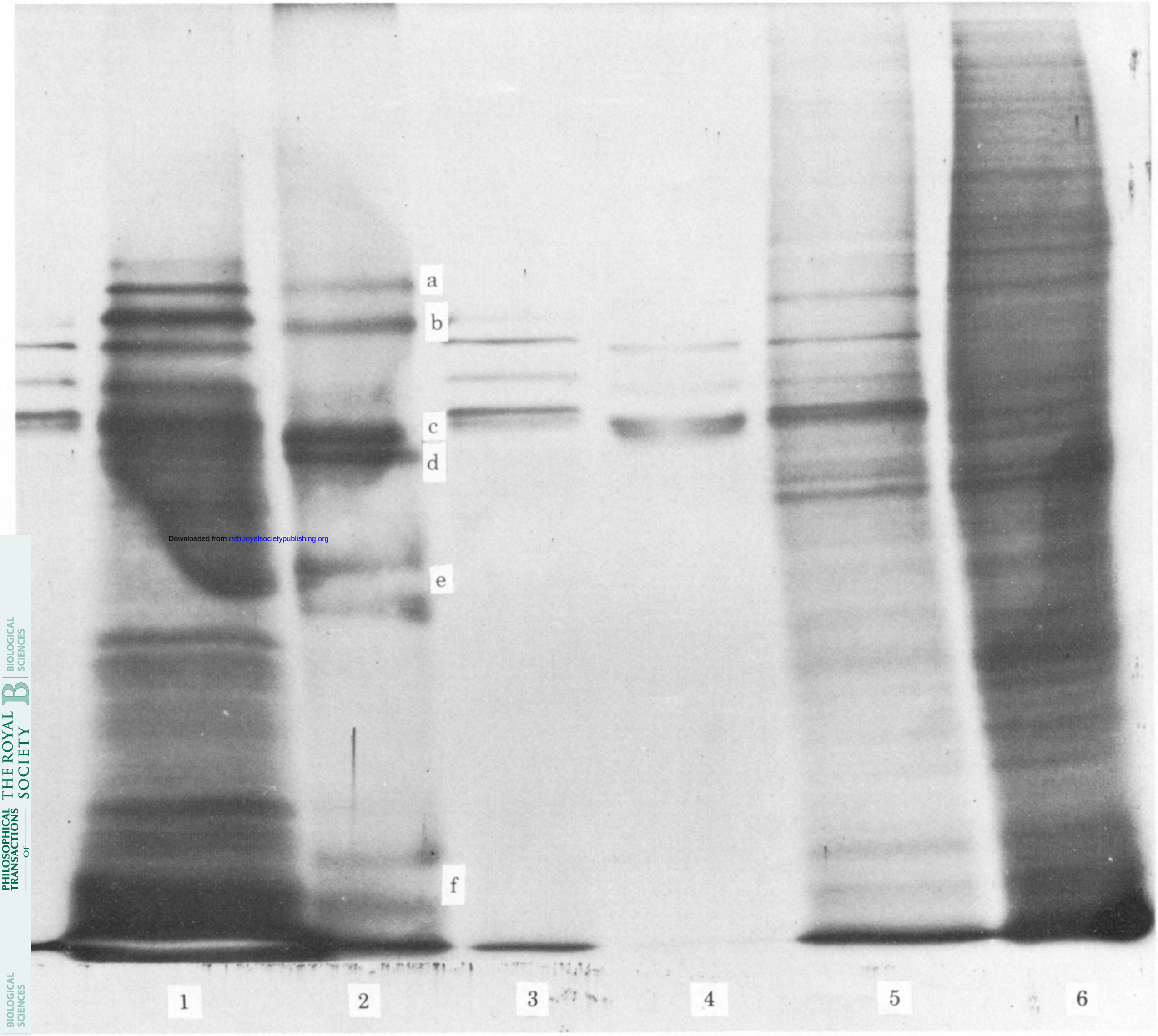


FIGURE 2. Analysis of purified glycine polypeptides by isoelectric focusing. The gel contained 7.5% polyacrylamide (acrylamide:bis acrylamide, 75:2); 3.7% LKB ampholines; pH 3.5–10.0, in 6 M urea. 30 μg samples were applied on filter paper toward the cathode and focused 6 h at 0 °C and 800 V. Data from Moreira *et al.* (1981).



PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

FIGURE 5. SDS polyacrylamide gel analysis of poly(A)-RNA translation products from oocyte and reticulocyte lysate translation systems. Lane 1, [^{3}H]leucine-labelled translation products from reticulocyte lysate. Lane 2, [^{3}H]dansylated standards where a denotes the α'-subunit of 75, $M_r = 75000$, b denotes α-subunit of 7 S, $M_r = 70000$; c denotes glycinin subunits, $M_r = 58000$; d is β-subunit of 7 S, $M_r = 54000$; e denotes acidic polypeptide components and f denotes the basic polypeptide components. Lane 3, translation products from the sample shown in lane 1 that were purified with antiglycinin IgGs. Lane 4, translation products from the sample shown in lane 5 that were purified with antiglycinin-IgGs. Lane 5, [^{3}H]leucine-labelled poly(A)-RNA translation products from *Xenopus* oocytes. Lane 6, [^{3}H]leucine-labelled proteins from *Xenopus* oocytes without injected mRNA. Refer to Tumer *et al.* (1982) for details.