Peptide Mapping Reveals Considerable Sequence Homology among the Three Polypeptide Subunits of G1 Storage Protein from French Bean Seed

YU MA, FREDRICK A. BLISS, AND TIMOTHY C. HALL
Department of Horticulture, University of Wisconsin, Madison, Wisconsin 53706

ABSTRACT

The major storage protein, G1 globulin, of bean (cv. Tendergreen) seeds was subjected to limited proteolysis with trypsin, chymotrypsin, papain, proteinase K, and protease V8 and to cleavage with cyanogen bromide and 2-(2-nitrophenylsulfanyl)-3-methyl-3'-bromoindolenine. Mapping of peptides separated from each of the three G1 subunits by polyacrylamide gel electrophoresis revealed that many proteolytic cleavage sites were present at similar positions on the subunits. Evidence was adduced that the G1 subunits are homologous in amino acid sequence for about 61% of their length. The remaining region (possibly COOH-terminal) of the subunits appears to be heterologous, with the α subunit bearing an additional methionine residue.

Proteolysis, followed by separation of the fragments to yield a distinctive fingerprint, has been used in the characterization of many proteins, including those from seeds (8). The procedure for electrophoretic separation of peptides obtained by partial proteolysis on SDS gels introduced by Cleveland et al. (1) is also proving to be of broad application (10). It is valuable for analysis of radioactive polypeptides synthesized in vitro using a single radioactive amino acid substrate because the fragments obtained, being larger than those obtained by limit digestion procedures, are more likely to contain a labeled residue. This procedure is especially useful for proteins that, like G1 globulin, the major storage protein of French beans (11, 12), consist of subunits which are very difficult to separate except on SDS-containing gels.

Here, the application of the Cleveland (1) mapping technique in the further characterization of the G1 subunits is described and evidence that they contain extensive regions of homology is provided.

MATERIALS AND METHODS

Plant Material and G1 Purification. Seeds of Phaseolus vulgaris L. cv. Tendergreen were purchased from Olds Seed Company, Madison, WI. Cotyledons were obtained by removing the seed coat and embryonic axis; G1 protein was isolated as described by Sun et al. (18), resuspended in 0.5 m NaCl, and stored at -27 C. The protein concentration was measured by using freeze-dried G1 as a standard; E280 for 1 mg/ml 0.5 m NaCl = 0.515.

Electrophoretic Procedures. Prior to electrophoresis, G1 protein in 0.5 m NaCl was dissociated by heating to 100 C for 3 min in a buffer containing 0.625 m Tris-HCl (pH 6.8), 2% SDS, 40% sucrose, 1% 2-mercaptoethanol, 2 mM EDTA-Na2, and 0.01% (w/v) bromophenol blue. The final protein concentration was adjusted to 10 mg/ml; typically, 40 μg were loaded into each gel slot. The discontinuous SDS-PAGE2 system of Laemmli (9), as modified by Ma and Bliss (11), was used; for 0.75 mm thick, 13% (w/v) acrylamide slab gels (16), an initial current of 7.5 mamp/gel was used and this was increased to 25 mamp/gel after all samples had moved into the stacking gel. Following electrophoresis, protein was stained with 0.1% (w/v) Coomassie brilliant blue R250 (19); the G1 protein was seen to be resolved into three discrete bands.

For mapping individual peptides, two cycles of electrophoresis were used. The samples were applied across the top of the stacking gel (not in individual lane slots) so that the stained G1 polypeptides appeared as three parallel lines across the gel after electrophoresis. Strips containing the individual subunits were excised and four to six pieces (each 8 mm long) were inserted into the lane slots of a second gel after equilibration with buffer A [125 mM Tris-HCl, 0.1% SDS, 1 mM EDTA-Na2, and 40% sucrose (pH 6.8)]. The use of several strips is necessitated by the effective dilution of the protein sample on proteolytic digestion when a single polypeptide band subsequently resolves into several peptide fragment bands. Accommodation of multiple gel slices was facilitated by widening the stacking gel compartment to 1.5 mm, but leaving the width of the separating gel at 0.75 mm. Proteolytic digestion was carried out essentially as described by Cleveland et al. (1); after protein from the gel slices and the protease (added on top of the slices) had migrated to the interface between the stacking and separating gels, the current was turned off for 15 min. The peptide digestion products were subsequently resolved by further electrophoresis in the 8.5-cm long separation gel. For chemical hydrolysis, the gel strips were soaked in appropriate solutions and equilibrated with buffer A, and the digested peptides were resolved in the normal manner. The equilibration step is essential (as is the application of similar numbers of gel slices to each sample well) to prevent uneven banding patterns. When protein standards are to be run in these second dimension gels, they must also be polymerized into gel strips, equilibrated, and applied to the gel in a fashion identical to that for the sample proteins.

Proteolytic Procedures. Papain and bovine pancreas α-chymotrypsin were purchased from Sigma; protease V8 (Staphylococcus

1 This work was supported by a grant from the Herman Frasch Foundation, National Science Foundation Grant PCM 78-11804, United States Department of Agriculture-Science and Education Administration Grants 5901-0410-9-0357-0 and 5901-0410-8-0053-0, and by the Research Division, College of Agricultural and Life Sciences, University of Wisconsin.

2 Abbreviations: PAGE, polyacrylamide gel electrophoresis; BNPS- skatole, 2-(2-nitrophenylsulfanyl)-3-methyl-3'-bromoindolenine; kd, kilodaltons; CNBr, cyanogen bromide.
aureus) was from Miles; trypsin treated with L-(tosylamido 2-phenyl) ethyl chloromethyl ketone (bovine pancreas) was from Worthington, and proteinase K (Triticum album Limber) was from E. Merck. CNBr was purchased from Aldrich Co. and BNPS-skatole was from Pierce Co., Rockford IL. For kinetic experiments, the proteins (50 µg) were digested in 20 µl solution as described previously (6), and 8-µl samples were removed at appropriate intervals and dissociated by heating to 100 C for 3 min with 2 µl solution containing 10% (v/v) 2-mercaptoethanol and 2% (w/v) SDS. For mapping of individual polypeptides, the overlaying procedures of Cleveland et al. (1) were followed. For CNBr cleavage, slices were cut from first-dimension gels and soaked in 2 ml 70% (v/v) formic acid containing 2% (w/v) CNBr for 48 h at room temperature in the dark. The soaked slices were washed thoroughly with a large quantity of distilled H2O and then equilibrated with buffer A. For BNPS-skatole cleavage, slices from the first dimension gel were soaked with 3 ml 50% (w/v) glacial acetic acid, 1 ml glacial acetic acid (containing 50 mm BNPS-skatole) was added, and the mixture was stirred at room temperature for 24 h. The slices then were washed and equilibrated as described for CNBr-treated material.

RESULTS

Kinetic Digestion. In agreement with the observation of Romero and Ryan (15), it was found here that G1 protein was rapidly digested by trypsin in the absence of SDS; when SDS was present in the reaction at levels similar to those within dissociating gels, the trypsin was inactivated in less than 1 min (Fig. 1). Chymotrypsin generated a characteristic series of peptides from G1 protein, but it also was rapidly inactivated in the presence of SDS (Fig. 2A). Papain (Fig. 2B) and protease V8 (Fig. 2C) were relatively resistant to SDS dissociation, the digestion to smaller peptides continuing for over 30 min.

From Figures 1 and 2, it can be inferred that each of the three G1 subunits contains sites for cleavage by trypsin, chymotrypsin, papain, and protease V8. This follows from the fact that the intensity of each of the stained bands representing the α, β, and γ subunits of G1 decreased during proteolysis. Further, since this decrease in intensity occurred uniformly for each subunit (e.g. note the complete disappearance of all three polypeptides between 1 and 5 min in the absence of SDS in Fig. 1), it is likely that there are similar numbers of cleavage sites on each subunit and probably that they are at similar locations along each of the polypeptide chains. The latter deduction is based on the fact that differential positions of the cleavage sites would result in differential accessibility for the protease (hence, differential susceptibility to cleav-

![Fig. 1. Kinetic analysis of digestion of G1 protein by trypsin. To the left, the extensive hydrolysis in the absence of SDS is seen; the rapid inactivation of trypsin by a level of SDS similar to that present in a denaturing acrylamide gel is shown to the right.](image1)

![Fig. 2. Kinetic analysis of G1 protein digestion by three proteolytic enzymes. Digestion is shown with chymotrypsin (A), with papain (B) and with protease V8 (C). Undigested G1 protein run in the same 16% gel is shown to the left of each panel, and samples of each proteolytic enzyme are shown to the right of each panel. The quantity of G1 protein and of enzyme added to each lane are indicated at the top of each panel.](image2)
age) and, consequently, varying intensities of staining for each of the G1 polypeptides and their degradation products throughout the kinetic analysis. Inspection of the peptide profiles generated by proteolysis with three different enzymes (Figs. 1 and 2) reveals that there is very little such variation in intensity, and, consequently, that the proteolytic sites are at similar locations. From this, it can be inferred that a considerable degree of amino acid sequence homology exists for each of the polypeptides. Evidence for homology between the G1 subunits has also been obtained by hybrid-arrested translation (14) of G1 mRNA, both of the radioactive polypeptide bands corresponding to G1 mRNA translation products being decreased when G1 mRNA is complexed with cloned G1 DNA (5).

**Mapping of Individual Subunits of G1.** The reproducibility found in the kinetics of proteolytic digestion indicates that certain sites on the G1 polypeptides are more susceptible or accessible to cleavage than are others. To compare the digestion of the α, β, and γ polypeptides, they were separated by one cycle of electrophoresis and then gel slices containing individual subunits were inserted into a second gel for digestion by the Cleveland (1) procedure. Papain and protease V8 were chosen (Fig. 3) because they were effective in digesting G1 polypeptides, yet resistant to inactivation by SDS. An analogous procedure was followed for chemical cleavage of the G1 polypeptides with BNPS-skatole and CNBr (Fig. 4).

At low concentrations of papain (Fig. 3A, 0.2 μg/gel slot), a series of polypeptides was generated, particularly interesting being the doublet (marked with arrows) seen for each of the α, β, and γ subunits. Because the doublets did not migrate the same distance in each lane, it followed that they have different mol wt. Analysis of these data required that the mol wt of the G1 polypeptides, yet resistant to inactivation by SDS. An analogous procedure was followed for chemical cleavage of the G1 polypeptides with BNPS-skatole and CNBr (Fig. 4).

The presence of sugar residues in these peptides (5) probably contributed to this. Because all the Cleveland maps were obtained with 16% acrylamide gels, the apparent mol wt of the G1 subunits (α, 51.0 kd; β, 48.0 kd; γ, 45.5 kd) and of the standard proteins in such a gel (Fig. 5B) were used to calculate the molecular sizes of fragments obtained on proteolysis.

The doublet of papain polypeptides (see arrows, Fig. 3A) were found to have mol wt of 40.0 and 38.0 kd for the α subunit, 37.0 and 35.0 kd for the β subunit, and 34.5 and 32.5 kd for the γ subunit. These are represented diagrammatically by the dashed lines in Figure 6A. From that figure, it can be seen that cleavage at a common point, 11.0 kd from the left-hand end, would give rise to the upper component of the doublet for each subunit and another homologous cleavage at 13.0 kd would yield the lower component of the doublet. The "stepped" appearance of the doublet then is rationalized since cleavage at common points in a homologous sequence will result in the generation of a series of peptides from each G1 subunit, each differing in size by the same amount as to the native polypeptides. Similar stepped peptide fragments are shown in Figure 6, B to F, and a summary of the common cleavage points revealed by these analyses is presented as Figure 7. No stepped appearance was apparent on digestion with protease V8 (Fig. 3B), possibly reflecting the ability of this enzyme to cleave within the putative heterologous region.

Detection of the homologous peptide fragments depends on their size since those of 12.0 to 10.0 kd or less do not resolve in the 16% SDS gels. In the case of papain digestion, an 11.0-kd fragment could not be identified reliably, but the 13.0-kd fragment resulting from the cleavage giving the lower band of the doublet was resolved. As the concentration (or, under favorable conditions, duration) of digestion was increased, more cleavages took place, resulting in the generation of additional peptide fragments, many being of similar length and thus migrating to similar positions in each lane of the gel (Fig. 3A; papain concentration, 2 μg/slot). Peptides having similar lengths are diagrammatically mapped to the left of each G1 polypeptide in the analyses of Figure 6, A to D. Unlike the case for the stepped peptides, it is impossible to be sure that each of the homologous polypeptides actually retains the left-hand terminus since more than one cleavage event may have occurred.
DISCUSSION

The agents used are capable of cleaving polypeptides at a wide range of sites. Trypsin cleaves mainly at arginine and lysine; chymotrypsin cleaves preferentially at tryptophan, phenylalanine, and tyrosine and also at leucine, methionine, asparagine, and histidine; papain has a wide specificity but cleaves mainly at arginine, lysine, and glycine sites (20). Protease V8 specifically cleaves peptide bonds on the COOH-terminal side of glutamic acid (7). Proteinase K cleaves at many sites, especially peptide bonds adjacent to the carboxylic group of aliphatic and aromatic amino acids (2). CNBr specifically degrades methionine residues (4), and BNPS-skatole is specific for tryptophan (3).

A more extensive kinetic analysis might reveal the relative location of the homologous regions inasmuch as the sum of fragment lengths after each cleavage should equal the total polypeptide length. The extent of homology indicated by the cleavage sites mapped from the left-hand side of Figure 6 is, therefore, conservative. Using common sites indicated by mapping the stepped peptides, the maximum extent of homology appears to be about 31.0 kd (see Fig. 6B). An 18.0-kd fragment is generated from only the α subunit by CNBr cleavage (Figs. 4B and 6D). This corresponds to 33.0 kd from the left-hand end and may well suggest that the transition from homologous to heterologous sequence occurs between 31.0 and 33.0 kd, taking 31 kd as the upper limit to the homology between the three subunits of 61% for the α polypeptide (the homologous region representing 31.0 kd of the total 51.0 kd), 65% for the β subunit, and 68% for the γ subunit.

Amino acid analysis of G1 protein revealed that there are only...
Fig. 6. Analysis of cleavage sites on the G1 subunits. In each panel, the upper solid line denotes the α subunit (51 kd); the middle line, the β subunit (48.0 kd); and the lower line, the γ subunit (45.5 kd). As discussed in the text, cleavage sites were arranged to maximize homology among the subunits. — — — — indicate major bands seen on electrophoresis of cleavage products; ----- (to the left) lie entirely within the constant or homologous region; - - - -, peptide sequences that extend into the heterologous region at the right. Numbers denote fragment lengths in kilodaltons. Panels A and B are plotted from gels shown in Figure 3 A and B, and Panels C and D are from gels shown in Figure 4, A and B, respectively. The gels used to calculate data of Panels E and F are not shown.

3 tryptophan residues/G1 monomer (17). The profile shown in Figure 4A indicates that there is 1 tryptophan residue/subunit and mapping analysis showed that it lies approximately 6 kd from the end of the homologous region (Fig. 6C). A value of approximately 10 methionine residues was obtained by amino acid analysis (17), and the mapping analysis of Figure 6D indicates that the α subunit contains 4 methionine residues, whereas the β and γ subunits each have 3 methionine residues.

It appears likely that the heterologous region of the G1 subunits lies at one end of the amino acid sequence, although whether this is NH₂- or COOH-terminal has not been determined. Evidence for regions of homology at the NH₂ terminus of polypeptides comprising the 11S protein from soybean seed has been obtained from sequencing studies (13). Maps, such as those shown in Figures 6, C and D, and 7, can be used to locate methionine and tryptophan residues within a polypeptide chain, and the use of additional procedures that cleave at specific amino acids would be valuable in further characterizing polypeptide chains of unknown amino acid sequence. It is estimated that errors of ±2 kd in positioning amino acid locations may arise as a result of difficulties in precisely determining apparent mol wt of the peptide bands after migration through the second dimension gel. Further complications, common to any peptide mapping procedure, may be caused by incomplete cleavage, or cleavage at residues other than those predicted. Nevertheless, because nucleic acid sequencing is now much more rapid than amino acid sequencing and since cloned mRNA sequences do not necessarily contain either the 3' or 5' termini, maps such as those obtained here should be particularly valuable in cases (like G1) where the NH₂ and COOH termini of the polypeptide are hard to sequence.