Molecular Cloning and Analysis of a Gene Coding for the Bowman-Birk Protease Inhibitor in Soybean

(Received for publication, February 10, 1984)

Rosemarie W. Hammond*, Donald E. Foard, and Brian A. Larkins

From the Department of Botany and Plant Pathology, Purdue University, West Lafayette, Indiana 47907

We have constructed cDNA clones from size-selected mRNA of developing soybean seeds to identify genes encoding the Bowman-Birk protease inhibitor. Recombinant clones containing sequences coding for the Bowman-Birk inhibitor were identified by mRNA hybrid selected translation and immunoprecipitation. The nucleotide sequence of the insert of plasmid pB38 corresponds to the mRNA for the Bowman-Birk inhibitor, although it is missing a portion of the coding sequence at the 5' end. Northern hybridization analysis of total RNA isolated from developing soybean seeds showed that the mRNA for the inhibitor is approximately 450 nucleotides long and that it accumulates early in embryogenesis. Southern hybridization analysis of restriction enzyme-digested soybean nuclear DNA indicated that the gene encoding the Bowman-Birk inhibitor is not highly reiterated in the genome. We have isolated a gene encoding the Bowman-Birk inhibitor from a soybean genomic library constructed in Charon 4A. DNA sequence analysis of the genomic clone reveals that it is similar, although not identical, to the cDNA clone and that it contains no intervening sequences.

Although the biological function of protease inhibitors is unclear, their wide and general occurrence in plants, animals, and microorganisms suggests that they are an important group of proteins (Laslowski and Kato, 1980; Liener, 1979; Richardson, 1977). Richardson (1977) and Ryan (1973) considered several possible functions for these proteins in plants which include 1) serving as a site for sulfur storage, 2) protection of the seed from endogenous protease activity, and 3) protection of the seed from insect and microbial invasion.

The most extensively studied of the protease inhibitors are those that inhibit the serine proteases trypsin and chymotrypsin. Seeds, tubers, and other storage structures are particularly rich sources of these inhibitors in plants (see Ryan, 1973; Richardson, 1977). Soybean seeds (Glycine max (L.) Merr.) contain at least two major types of serine protease inhibitors: the Kunitz inhibitor (Kunitz, 1946) and the Bowman-Birk inhibitor (Birk, 1961) and its related family of isooinhibitors (PI I-IV) (Hwang et al., 1977; Odani and Ikenaka, 1977). The Kunitz inhibitor is a single polypeptide chain of $M_r = 21,500$ that inhibits both trypsin and chymotrypsin. The mRNA for the Kunitz inhibitor has been partially purified. When translated in vitro, it directs the synthesis of a precursor polypeptide with an apparent $M_r = 23,800$ (Vodkin, 1981) due to the presence of a signal peptide at the amino terminus.

The BBI and PI I-IV comprise a closely related group of protease inhibitors. They contain 20% sulfur amino acids which are the most limiting of the essential amino acids in soybean seeds. Both groups have $M_r$ less than 10,000 (Hwang et al., 1977) and comparison of their amino acid sequences reveals that the BBI and PI I-IV are about 70% homologous. The amino acid sequences of PI I-IV are nearly identical and differ primarily in the length of their N termini (Hwang et al., 1977; Odani and Ikenaka, 1978). These proteins are classified as double-headed inhibitors because each contains two reactive site domains within the same polypeptide. The BBI has both a trypsin (Lys-Ser) and a chymotrypsin (Leu-Ser) inhibitory site (Birk, 1961), whereas PI I-IV have two trypsin (Arg-Ser) inhibitory sites (Hwang et al., 1977). The sequences of the amino acids surrounding the reactive sites both within the same polypeptide and between the BBI and PI I-IV are very homologous. This homology suggests that the BBI and PI I-IV genes may have evolved from a common ancestral monovalent inhibitor (Tan and Stevens, 1971). The occurrence of multiple low molecular weight protease inhibitors among other members of the legume family (Chu and Chi, 1965; Jones et al., 1963; Wilson and Laslowski, 1973) as well as distantly related plants (Bryant et al., 1976; Kanamori et al., 1976) suggests that the genes may have been conserved throughout the evolution of several plant species.

The BBI and PI I-IV do not appear to be expressed in any plant tissue other than the developing seed (Hwang et al., 1978). The proteins are rapidly released from the seed within the first 8 h of imbibition. The kinetics of their release suggest that the inhibitors may be eluted from a barrier-free pool. Consistent with this suggestion is the observation that some of the BBI is found in intercellular spaces of cotyledons of ungerminated seeds, but is absent from this location in 4-day-old seedlings (Horisberger and Tacchini-Vonlanthen, 1983).

Isolation of the genes encoding these protease inhibitors will provide information on their number and organization as well as probes for studying the regulation of their expression. We have constructed and characterized a cDNA clone corresponding to the BBI. DNA sequence analysis confirmed that this clone corresponds to the BBI mRNA, although it is missing the 5' noncoding region as well as the coding sequence for the first 30 amino acids. Using this clone as a probe we determined that the BBI gene is present in only one or two copies in the soybean genome, and we have isolated a genomic clone from a Charon 4A library of soybean nuclear DNA.
cDNA clones were bound to nitrocellulose and hybridized to mRNA fractions enriched with protease inhibitor mRNAs. Nonspecifically bound mRNA was removed by extensive washing at room temperature; bound mRNA was eluted at increasing temperatures. The hybrid selected mRNA was precipitated and subsequently translated in a wheat germ cell-free system. In vitro translation products were immunoprecipitated with either anti-BBI or anti-PI I-IV serum; preimmune serum was used as a control. The immunoprecipitates were separated by electrophoresis on a 2.5% SDS-polyacrylamide gel and analyzed by fluorography. Lane 1, no mRNA; lane 2, BBI-enriched mRNA translated and immunoprecipitated with preimmune serum; lane 3, same as lane 2 but immunoprecipitated with anti-BBI serum; lane 4, same as lane 2 but immunoprecipitated with anti-PI I-IV serum; lane 5, translation products of mRNA eluted from pB38 at 55 °C immunoprecipitated with anti-BBI serum; lane 6, as in lane 5 but immunoprecipitated with anti-PI I-IV serum; lane 7, translation products of mRNA eluted from pB38 at 75 °C immunoprecipitated with anti-BBI serum; lane 8, as in lane 7 but immunoprecipitated with anti-PI I-IV serum; lane 9, translation products of mRNA eluted from pB38 at 95 °C immunoprecipitated with anti-BBI serum; lane 10, as in lane 9 but immunoprecipitated with anti-PI I-IV serum. Immunoprecipitated products were generally 20–30% of the total translation products. Equivalent amounts of radioactivity were loaded in each lane. The positions of the monomer (m), and the dimer (d), and trimer (t) aggregates are indicated.

Although there is extensive sequence homology between the BBI cDNA and genomic clones, they differ slightly in their deduced amino acid sequences.

**FIG. 2.** Northern blot analysis of poly(A) RNA from mid-maturation stage developing soybean seeds. One μg of RNA was separated by electrophoresis on a 1.4% agarose gel containing 5 mM methyl mercury hydroxide. The gel was stained with etidium bromide and the DNA was blotted onto nitrocellulose. After the DNA was baked at 80 °C, it was hybridized with a nick-translated probe prepared from the insert of pB38. A, etidium bromide staining of mid-maturation stage RNA following electrophoresis; B, autoradiography of Northern blot of the gel in A; C, autoradiography of 2 μg of total RNA isolated from developing seeds at 2-, 5-, 6-, 8-, and 10-mm stages; fraction 12 from a dimethyl sulfoxide sucrose gradient containing sequences enriched for BBI; poly(A) RNA from mid-maturation stage (10 mm) seeds. Nt, nucleotides.

**FIG. 3.** Northern dot blot analysis of total RNA isolated from several stages of developing soybean seeds. Total RNA isolated from seeds of 2-, 5-, 6-, 8-, and 10-mm in length was spotted onto nitrocellulose at 1-, 5-, 10-, and 20-μg concentrations as described under "Experimental Procedures" and probed with nick-translated cDNA clones corresponding to the 11 S (glycinin), the Kunitz protease inhibitor, and the BBI polypeptides. Autoradiograms were scanned with a densitometer to determine relative intensities of spots.

**EXPERIMENTAL PROCEDURES**

**Materials**—Restriction endonucleases were obtained from Bethesda Research Laboratories or New England Biolabs (Beverly, MA). Synthetic EcoRI linkers, T, DNA ligase, and Escherichia coli DNA polymerase 1 were obtained from Bethesda Research Laboratories. Polynucleotide kinase was obtained from P-L Biochemicals and calf intestinal alkaline phosphatase from Boehringer Mannheim. [γ-32P]ATP and [α-32P]dCTP were from New England Nuclear or from Amersham Corp. Nitrocellulose was obtained from Schleicher and Schuell (Keene, NH).

**Soybean mRNA Isolation and Fractionation**—Poly(A) RNA was isolated from a total RNA extract of immature soybean seeds of the cultivar Tracy as previously described (Foard et al., 1982). Seeds of less than 200 mg, fresh weight, were frozen in liquid nitrogen and stored at −80 °C. Poly(A) RNA was isolated by three cycles of affinity chromatography on oligo(dT)-cellulose and fractionated by sedimentation in denaturing sucrose gradients. Messenger RNA from the gradient fractions was translated in a wheat germ cell-free system; translation products were immunoprecipitated with either anti-BBI or anti-PI I-IV sera to identify those fractions enriched for BBI and PI I-IV mRNAs, respectively.
Aendonuclease-digested soybean DNA. Nuclear DNA from the cultivar Provar (10 µg, 5.0 × 10^6 copies of the genome) was digested with the restriction endonucleases BamHI, Ball, EcoRI, HaeIII, or Xbal. Reconstruction lanes contained the cDNA clone, pB38, digested with BamHI (6.7, 12.5, 25, 50, 100, 200, and 400 pg equivalent to 0.2, 0.5, 1, 2, 4, 8, and 16 × 10^6 copies). After separating the DNA on a 0.8% agarose gel, it was transferred to nitrocellulose as described by Southern (1975). The cDNA insert from the plasmid was labeled by nick translation and hybridized to the soybean DNA as described under “Experimental Procedures.” Based on the calculations of Casey and Davidson (1977) the hybridization criterion was at Tm – 20°C. The filter was washed at Tm – 28°C. The numbers across the top denote the number of gene copies in the reconstruction lanes. The marker lane shows migration of corresponding kilobase fragments.

**Construction of Recombinant Clones**—Double-stranded cDNA was prepared by standard procedures from size-selected mRNA (Buell et al., 1978; Wickens et al., 1978). The first strand was synthesized with oligo(dT) as the primer for reverse transcriptase. The second strand was synthesized with E. coli DNA polymerase I, and the resulting ds-cDNA was digested with S1 nuclease. The ds-cDNA was ligated to synthetic EcoRI linkers and after digestion with EcoRI the resulting cDNA was size selected on a 5% polyacrylamide gel. Appropriate fractions were ligated to EcoRI-digested pBR322. Recombinant plasmids were used to transform E. coli HB101. Several hundred transformants were screened by Grunstein-Hogness filter hybridization (1975) with a ^32P-labeled cDNA probe prepared from protease inhibitor-enriched mRNA.

**Identification of Recombinant Clones**—Recombinant plasmids were linearized by digestion with BamHI, denatured with NaOH, and transferred to nitrocellulose (McGrogan et al., 1979). The nitrocellulose disks were then hybridized for 10–20 h, at 37°C in 200 µl of a solution containing 4 × SSC (1 × SSC: 0.15 M sodium chloride, 0.015 M sodium citrate), 50% formamide, 100 µg/ml of poly(A), 200 µg/ml of yeast tRNA, 1% SDS, and 5–10 µg of soybean mRNA. After hybridization, each disk was washed three times at room temperature with 5 ml of 10 mM Tris-HCl, pH 8.0, containing 0.5% SDS. Each disk was then washed at room temperature with 5 ml of 10 mM Tris-HCl, pH 7.4, containing 2 mM EDTA. Messenger RNA bound to the filter was eluted (at increasing stringencies) with a buffer containing 2 mM EDTA, pH 7.4, 20 µg/ml of tRNA, by raising the temperature from 55 to 96°C. The BBI and PI I–IV proteins were identified by in vitro translation of mRNAs and immunoprecipitation with antisera. The immunoprecipitates were separated by electrophoresis on SDS-polyacrylamide gels (Foord et al., 1982) and compared with authentic inhibitor proteins after fluorography (Laskey and Mills, 1975).

**Northern Analysis**—Messenger RNA or total RNA was separated by electrophoresis on denaturing gels containing 5 mM methyl mercury hydroxide and 1.4% agarose (Bailey and Davidson, 1976). After the gels were stained with ethidium bromide, they were blotted onto nitrocellulose and hybridized in a solution containing 5 × SSC, 20 mM NaPO₄, pH 6.8, 1% SDS, 0.02% Ficoll, polyvinylpyrrolidone, and bovine serum albumin, and 5% dextran sulfate in 50% formamide (Thomas, 1980). The hybridization probe was prepared by isolating the cDNA insert from pB38 and labeling by nick translation with ^32P]dCTP (Maniatis et al., 1976). After hybridization, the filters were washed with 1 × SSC containing 0.1% SDS at room temperature, and then at 68°C with the same solution.

**DNA Isolation and Southern Analysis**—DNA was isolated from leaf tissue of the soybean cultivar Provar by the method of Murray and Thompson (1980). The DNA had a mass average length of 50 kb. Five aliquots of 10 µg each (5.0 × 10⁶ haploid genomes) were digested to completion with Ball, BamHI, EcoRI, HaeIII, or Xbal (10 units/µg of DNA). After digestion with restriction enzymes, the fragments were separated by electrophoresis on 0.8% agarose gels. Reconstructions to determine the gene copy number contained the cDNA clone.
Fig. 6. Restriction endonuclease map and sequencing strategies for cDNA clone pB38 (A), and the genomic clone ABB13.10 (B). Only selected restriction sites are shown. The horizontal arrows indicate the direction of sequencing and the lengths indicate the actual length of the sequence obtained. The hatched area on the genomic clone delimits the coding region.

pB38 that had been linearized by BamHI digestion. Samples in these lanes contained 50 pg of DNA for two gene copies, 100 pg for four copies, etc. Ten μg of EcoRI-digested calf thymus DNA was added to each of these lanes as carrier. ADNA fragments resulting from digestion with several enzymes were used as size markers.

Gels were stained with ethidium bromide and prepared for blotting by depurinating the DNA (Wahl et al., 1979); the DNA was transferred to nitrocellulose by the method of Southern (1975). Hybridization was as described by Wahl et al. (1979). After hybridization, the filter was washed at room temperature in 0.2 x SSC containing 0.05% sodium sarcosyl and 0.01% sodium pyrophosphate and then at 55 °C with the same solution. Autoradiographs were scanned with a Quick Scan (Helena Lab) densitometer.

Screening of a Soybean Genomic Library—A soybean genomic library from the cultivar Forrest was provided by Dr. Robert Goldberg (UCLA). The DNA was prepared by partial digestion with HaeIII/AluI and cloned via EcoRI linkers into the λ phage Charon 4A (Blattner et al., 1977). Approximately 1.0 x 10^6 phage were screened (equal to one genome equivalent based upon 18-20-kb DNA inserts and 1.97 pg of DNA/haploid soybean genome) as described by Blattner et al. (1978) with the cDNA clone pB38. After purification of the phage, EcoRI fragments were recovered and subcloned into pBR325 or pUC8 as described by Maniatis et al. (1978). The subcloned EcoRI fragments included the 3.5, 2.0-, and 1.0-kb fragments of X 13.10. The isolated clone and subclones were digested with one or two restriction endonucleases and the fragments were separated on 1.5% agarose gels. Southern mapping of the coding regions on the genomic clone and subclones was performed with either the nick-translated insert of the cDNA clone pB38 or a ss-cDNA probe prepared from size-selected mRNA.

5'-Terminal Labeling and DNA Sequencing—Ten μg of DNA was digested with the appropriate restriction endonuclease and the 5' termini were then end labeled with [γ^32P]ATP. DNA sequence analysis was by the procedure of Maxam and Gilbert (1980) with a modification (Smith and Cabro, 1980).
S1 Nuclease Mapping—A subclone containing the 1-kb EcoRI fragment at the 5' end of the genomic clone was used for an S1 protection experiment to determine the initiation of transcription. The DNA was digested with EcoRI and the 5' ends were labeled with [32P]ATP. The labeled fragment was recut with HindIII to produce an 800-base pair fragment containing the coding strand labeled at the 5' end. DNA-RNA hybridization reactions contained approximately 0.1 µg of DNA (150,000 dpm) and 10 µg of soybean seed mRNA in 40 µl of 80% formamide, 0.4 M NaCl, 0.04 M Pipes, pH 6.4, and 1 mM EDTA (Berk and Sharp, 1977). After the DNA denatured, the reactions were incubated at 49 °C for 16–20 h. After digestion with 5 units of S1 nuclease, the DNA was analyzed on a 0.3-mm thick, 5% sequencing gel and compared with a Maxam and Gilbert sequencing ladder.

Recombinant DNA—All experiments were performed using the appropriate levels of containment as detailed by the current “Guideline for Research Involving Recombinant DNA Molecules.”

RESULTS
Construction and Identification of cDNA Clones—To increase the probability of obtaining cDNA clones for soybean protease inhibitors, we isolated enriched fractions of the mRNAs by dimethyl sulfoxide sucrose gradient centrifugation (Foard et al., 1982). These mRNAs were used to synthesize cDNAs ranging in size from 250 to 500 nucleotides. The ds-cDNA was ligated to synthetic EcoRI linkers and ligated into pBR322 (Maniatis et al., 1976). After transformation of E. coli HB101, several hundred clones were screened with a ss-cDNA probe prepared from the enriched mRNA fraction. Several clones that hybridized to the probe were further analyzed by mRNA hybrid-selected translation (McGrogan et al., 1979). In vitro translation of mRNA eluted from the nitrocellulose filters yielded polypeptides that were immunoprecipitated by anti-BBI serum and co-migrated with the trimer aggregate of the protease inhibitors (Fig. 1). Several of the clones hybridized to mRNA whose in vitro translation products were immunoprecipitable with both anti-BBI and anti-PI I–IV sera, although previously these antisera had not been found to cross-react (Hwang et al., 1977).

Restriction enzyme digestion of several of the selected clones with EcoRI revealed that the cDNA inserts ranged in size from 300 to 400 nucleotides. Further analysis with enzymes that recognize restriction sites of 4, 5, and 6 nucleotides yielded restriction maps suggesting that several of the clones were similar. One of the clones, pB38, was chosen for further analysis (Fig. 6A).

Nucleotide Sequence Analysis of pB38—Clone pB38 was characterized by DNA sequence analysis to verify that it corresponded to the BBI mRNA. The nucleotide sequence of this clone (Fig. 7) contains the coding information for the BBI. Conversion of the nucleotide sequence into an amino acid sequence shows that the clone encodes residues 30 to 71 of the BBI. Although the clone lacks some of the 5' coding and non-coding region of the BBI mRNA, it contains a

Fig. 8. S1 mapping analysis of the 5' end of the BBI mRNA. Hybridization reactions and S1 nuclease digestions were performed as described under “Experimental Procedures.” Lanes marked T and C represent the Maxam and Gilbert T and C reactions of the sequence from which the S1 fragment was sized. Lane M, pBR322 digested with MspI and end labeled with 32P; lane 1, control reaction with no mRNA; lane 2, reaction with end labeled EcoRI site that extends beyond the 5' end of the gene.

Fig. 9. Amino acid homology within the BBI polypeptide. Upper, the two reactive site domains are represented by the stippled areas, and the reactive sites are shown by the darkened areas. Lower, comparison of the nucleotide sequence of the first and second reactive site domains in the BBI gene. Nucleotides are numbered left to right from the 5' to 3'. Amino acids are numbered according to their position from the amino-terminal end of the mature BBI polypeptide. Dots indicate amino acid homologies. Vertical lines show nucleotide homologies.
polyadenylation site 23 nucleotides after the coding sequence and a 3' noncoding region of 63 nucleotides before the poly(A) tail. DNA sequence analysis of several additional clones (data not shown) revealed that they were identical to pB38, and they also lacked the 5' end of the mRNA.

Characterization of the BBI mRNA—To determine the size of the mRNA encoding the Bowman-Birk inhibitor, the cDNA insert from pB38 was labeled by nick translation and hybridized to cotyledon poly(A) RNA that was separated by electrophoresis on a denaturing methyl mercury hydroxide gel (Bailey and Davidson, 1976). This clone hybridized to a single RNA with an apparent size of 450 nucleotides (Fig. 2B). A corresponding band could be visualized on ethidium bromide-stained agarose gels that had been overloaded with poly(A) RNA.

The expression of the BBI gene during seed development was analyzed by assaying the level of mRNA by Northern blot hybridization (Kafatos et al., 1979). Total RNA was isolated from young embryos 2, 5, 6, 8, and 10 mm in length. After separation of equal amounts of RNA by electrophoresis on denaturing methyl mercury hydroxide gels, the RNA was transferred to nitrocellulose and hybridized to a [32P]-labeled probe prepared from pB38. This analysis indicated that BBI transcripts appear first in the 5–6 mm embryos and that their level increases in the 10-mm embryos (Fig. 2C).

The early appearance of these transcripts is similar to that of the major soybean seed proteins (Goldberg et al., 1981b). We compared the timing and extent of expression of the BBI with several of these genes including the 11 S storage protein as well as the Kunitz trypsin inhibitor. The Northern dot hybridizations in Fig. 3 demonstrate that sequences corresponding to the BBI were found to accumulate at the same developmental stages as the 11 S storage protein and Kunitz trypsin inhibitor mRNAs. Based upon densitometric analysis of the autoradiographs, and taking into account differences in the sizes of the probes, the amount of BBI mRNA was equivalent to that of the Kunitz inhibitor and about 50% of the 11 S protein precursor.

Organization of BBI Sequences in the Soybean Genome—To determine the number of genes encoding the BBI and their organization in the soybean genome, we performed a Southern blot hybridization analysis. Samples of soybean leaf nuclear DNA were digested to completion with BalI, BamHI, EcoRI, HaeIII, or XbaI and separated by electrophoresis on a 0.8% agarose gel. The DNA obtained from a blot of the gel to the corresponding band could be visualized on ethidium bromide stained agarose gels that had been overloaded with poly(A) RNA.

The expression of the BBI gene during seed development was analyzed by assaying the level of mRNA by Northern blot hybridization (Kafatos et al., 1979). Total RNA was analyzed by assaying the level of mRNA by Northern blot hybridization. This analysis indicated that BBI transcripts appear first in the 5–6 mm embryos and that their level increases in the 10-mm embryos (Fig. 2C).

The early appearance of these transcripts is similar to that of the major soybean seed proteins (Goldberg et al., 1981b). We compared the timing and extent of expression of the BBI with several of these genes including the 11 S storage protein as well as the Kunitz trypsin inhibitor. The Northern dot hybridizations in Fig. 3 demonstrate that sequences corresponding to the BBI were found to accumulate at the same developmental stages as the 11 S storage protein and Kunitz trypsin inhibitor mRNAs. Based upon densitometric analysis of the autoradiographs, and taking into account differences in the sizes of the probes, the amount of BBI mRNA was equivalent to that of the Kunitz inhibitor and about 50% of the 11 S protein precursor.

Organization of BBI Sequences in the Soybean Genome—To determine the number of genes encoding the BBI and their organization in the soybean genome, we performed a Southern blot hybridization analysis. Samples of soybean leaf nuclear DNA were digested to completion with BalI, BamHI, EcoRI, HaeIII, or XbaI and separated by electrophoresis on a 0.8% agarose gel. The DNA obtained from a blot of the gel to the filter was hybridized to the insert from the BBI cDNA clone. This probe hybridized to a single band in the BamHI digest, whereas multiple bands were detected in the BalI, EcoRI, HaeIII, and XbaI digests (Fig. 4). Based upon analysis of the gene copy reconstruction, the single band in the BamHI digest is equivalent to only one or two gene copies/haplodiploid genome. Therefore, the BBI does not appear to be encoded by multiple copy genes as has been found for a number of other seed proteins (Schuler et al., 1982a, 1982b; Croy et al., 1982; Fischer and Goldberg, 1982; Pedersen et al., 1982).

Identification of a Recombinant DNA Clone Containing the Bowman-Birk Inhibitor Gene—A soybean genomic library constructed in Charon 4A was screened with a [32P] nick-translated probe of the 300-base pair insert of pB38. Three clones were obtained which appeared to be identical based on restriction enzyme digestion; one of these clones was designated ABB13.10 (Fig. 5).

Restriction endonuclease analysis of ABB13.10 revealed that it contained a 12.5-kb insert of soybean DNA (Fig. 6B). The restriction digestion pattern showed the presence of five EcoRI fragments of 3.8, 3.5, 3.0, 2.0, and 1.0 kb. The location of the BBI coding sequence was determined by Southern hybridization of pB38 to EcoRI digests of ABB13.10. The probe hybridized to a single fragment of 2.0 kb (Fig. 5A). This fragment probably represents the 3' coding region of the BBI gene, since the BBI sequence contains an EcoRI site at amino acid 29-amino acid 30. To identify the 5' coding region of BBI, we hybridized the EcoRI digest of ABB13.10 with a single-stranded cDNA probe prepared from the BBI-enriched mRNA fraction. The probe hybridized to three EcoRI fragments of 3.5, 2.0, and 1.0 kb. Since the 1.0-kb EcoRI fragment is contiguous to the 5' end of the 2.0-kb fragment, it appeared likely that it encodes the 5' end of the BBI. The other 3.5-kb fragment, which occurs 2 to 3 kb from the 3' end of the BBI gene, appears to be a different gene since it fails to hybridize with pB38. Northern hybridization with the 3.5-kb EcoRI fragment shows that it encodes an mRNA of approximately the same size as the BBI which is also expressed during seed development.

Nucleotide Sequence Analysis of the BBI Gene and Its Flanking Regions—The structural features of the BBI gene were characterized by subcloning the 2.0- and 1.0-kb fragments of ABB13.10 into pUC8 and determining their nucleotide sequence by the procedure of Maxam and Gilbert (1980). There is extensive sequence homology between the genomic clone and the cDNA clone pB38, although they are not identical (Fig. 7). There are 14 nucleotides different in the coding region. Most of these do not result in amino acid changes although in one case the genomic clone encodes alanine rather than isoleucine and in another it encodes glutamine rather than threonine. There are also 14 nucleotide differences in the 3' noncoding regions. With the exception of these differences the sequences are co-linear, indicating that there are no introns in the gene. The protein encoded by this gene contains a short leader peptide, since the first ATG precedes the N-terminal aspartic acid by 15 amino acid residues.

To determine the 5' cap site of the mRNA we used a modified Berk and Sharp (1972) procedure. The EcoRI site in the coding region was end-labeled with [32P] and the DNA fragment was hybridized to soybean seed mRNA. S1 nuclelease digestion of the DNA-RNA hybrid revealed a protected fragment of approximately 220 base pairs (Fig. 8). This places the 5' cap region of the mRNA approximately 50–90 nucleotides upstream of the initiation codon. The 5' and 3' flanking regions of the gene have a lower GC content than the coding region. The 5' noncoding sequence is 39% GC, the 3' noncoding sequence is 36% GC, while the coding region is 40% GC. There is no canonical TATA box (Benoist et al., 1980) within the first 50 nucleotides preceding the cap site of the mRNA. Neither the genomic clone nor the cDNA clone has the conventional polyadenylation sequence AATAAA (Proudfoot and Brownlee, 1976). It is nevertheless likely that a functional polyadenylation sequence is present because the cDNA contains a poly(A) tail.

DISCUSSION

By constructing cDNA clones with size-fractionated mRNA we were able to identify a sequence corresponding to the soybean BBI. It is surprising that among the clones we analyzed, we did not find cDNA sequences for the PI I-IV. Several clones hybridized to mRNAs which, when translated in vitro, yielded polypeptides that were immunoprecipitated with both anti-BBI and anti-PI I-IV sera. Microheterogeneity was observed in the fine structure restriction maps of these clones; however, upon RNA sequence analysis they were found to be identical to the BBI. The fact that these clones bound mRNAs...
corresponding to both the BBI and PI I-IV implies that these sequences are sufficiently homologous that they cross-hybridize at the low stringency (T_m = 49 °C) of the mRNA hybrid selection procedure. Perhaps the mRNAs corresponding to PI I-IV are less abundant than those for the BBI. This, as well as the apparent sequence homology between the two types of mRNA, prevented us from identifying a clone for PI I-IV.

However, by screening a sufficiently large number of cDNA clones under stringent hybridization conditions with the BBI as a probe we hope to identify a clone for this sequence as well. Soybean seed proteins begin to accumulate early in the developing cotyledons with a corresponding increase in their messenger RNAs (Goldberg et al., 1981a). We found that the BBI mRNA accumulates early during the mid-maturation stage of development and reaches a steady state level later in development (Fig. 2C). Assuming that the expression of the BBI gene is regulated in a manner similar to other soybean seed protein genes, we expect that the amount of mRNA declines toward maturity (Goldberg et al., 1981b). Goldberg et al. (1981b) estimated that there were approximately 16,000 transcripts of the Kunitz trypsin inhibitor and 27,000 transcripts of the 1 S storage protein/cell in the mid-maturation stage cotyledon. Based upon the comparative hybridization of the BBI with clones for these sequences (taking into account the probe specific activity and length differences), we estimate that there are 15,000-20,000 BBI transcripts/cell at the 10-mm stage. This level is similar to the Kunitz inhibitor mRNA, which appears to be encoded by more than one gene.

Unlike a number of other seed protein genes (Fischer and Goldberg, 1982; Pedersen et al., 1982; Schuler et al., 1982a) the BBI does not appear to be encoded by a multigene family. Based upon gene copy reconstructions, the single band observed in the BamHI digestion is equivalent to only one or perhaps two genes. The additional bands observed in the BalI, EcoRI, HaeIII, and XbaI digestions have the intensity of less than one gene copy, and these restriction sites occur within the gene (Fig. 7). Nevertheless, there is microheterogeneity between the sequence of the cDNA clone and the genomic clone, indicating that several genes may encode the BBI protease inhibitor. This heterogeneity could result from the presence of multiple alleles for the BBI locus. The three genomic clones that were isolated appeared to be identical based on preliminary restriction enzyme mapping; however, minor differences in DNA sequence would not have been apparent. This variation could also be explained by differences in genetic background, since the cDNA, genomic clone, and proteins were isolated from different soybean cultivars.

Although the cDNA clone pB38 does not contain a complete copy of the BBI mRNA, the amino acid sequence deduced from the coding region is in complete agreement with that reported for the mature polypeptide (Odani and Ikenaka, 1978). The majority of the nucleotide differences between the coding region of the cDNA clone and the genomic clone (XBB13.10) are in wobble positions and do not result in amino acid changes; however, there are two amino acid differences. It is unlikely that these are due to sequencing errors, since both strands of the DNA were analyzed (Fig. 5). Previous analyses in which the BBI was sequenced (Odani and Ikenaka, 1977; Hwang et al., 1977) did not reveal heterogeneity in the protein sequence, although the cultivar Forrest was not used for these studies. With the exception of the two amino acid differences, the BBI gene in XBB13.10 is not unusual. The protein encoded by the gene is of the length expected for the BBI. As is true of the soybean Kunitz inhibitor, the gene does not contain introns. We must analyze the DNA sequences of additional cDNA and genomic clones before we can explain the significance of the sequence variation between these clones.

Although in vitro translation of the BBI mRNA yields polypeptides that migrate on SDS gels at positions corresponding to dimers and trimers (Foard et al., 1982), the DNA sequence of the gene shows that the protein is not synthesized as a multimeric polypeptide. Based on the DNA sequence, the initial translation product contains a short leader peptide. Our previous attempts to sequence the N terminus of the in vitro translation product were unsuccessful (Foard et al., 1982), and thus we cannot verify this protein sequence. The leader peptide is apparently not responsible for aggregation of the protein, since aggregation also occurs when the mature polypeptide is not reduced and alkylated (Foard et al., 1982).

The identity of the linked gene 2-3 kb downstream from the BBI is unknown. Although this gene hybridized to an mRNA which is of the same size as the BBI and is similarly expressed during seed development, it does not appear to be another BBI gene or a gene for an isoinhibitor (PI I-IV). Translation of hybrid-selected mRNAs did not result in a product detectable with BBI antisera or PI I-IV antisera. We have begun to sequence the gene in order to elucidate its protein product.

It is not entirely surprising to find two different developmentally regulated genes on the same clone. Goldberg and his associates have identified a genomic clone on which two Kunitz trypsin inhibitor genes are linked within 1 kb. It is possible that the additional gene on the genomic clone corresponds to a PI I-IV sequence. If the corresponding mRNA were not abundant, we may have failed to identify the gene product by mRNA-hybrid selection and immunoprecipitation.

The double-headed inhibitors have two internally homologous domains (Fig. 9), each containing a reactive site for a protease. As was mentioned in the Introduction, these proteins are thought to have evolved from a monovalent ancestral inhibitor by internal gene duplication. Odani and Ikenaka (1978) proposed a scheme for this evolutionary process. Internal duplication of a monovalent inhibitor leads to a monospecific double-headed inhibitor (analogous to the soybean PI IV, which inhibits only trypsin); further mutation by base substitution gives rise to the BBI and other similar legume inhibitors. The chymotrypsin inhibitory site of these proteins is usually located in the second domain of the polypeptide. The amino acid and nucleotide sequences of the two active site regions are 50% homologous (Fig. 9), whereas the amino and carboxyl termini and the linker regions are less than 25% homologous. This provides evidence that the two domains arose by gene duplication and subsequent mutation.

An estimate of the time at which the first and second domains evolved can be made by the method of Perler et al. (1980). These estimates assume: 1) the mutation rate is proportional with time; the replacement rate is linear but slower than the silent substitution rate; 2) the replacement rates used for calculations of divergence time are based on a moderately conserved gene and therefore the rate of replacement may be different for each protein; and 3) evolution depends on the requirement that a substitution be compatible with the structure and function and that the organism can function with the change. Ten of the 18 amino acids in the reactive site domains are conserved, suggesting that the replacement rate is low. Also, proteins with highly constrained structures like the BBI would be expected to fix mutations at a lower rate because only a small number of sites can be neutral.

__2__ R. B. Goldberg, personal communication.
Using the observed unit evolutionary period of 10 for replacements as determined for globin (Perler et al., 1980), an estimate of the divergence time of 310 million years can be made for the first and second domains of this BBI gene. This estimate would be more meaningful if a number of other BBI-like genes from more primitive members of the legume family, as well as the PI I-IV gene, were compared. Southern hybridization analysis of DNA from other legumes such as mimosa and redbud indicate that we can use the BBI gene as a probe to isolate these genes. In any case, it appears that the structural and functional constraints placed on the BBI reactive site domains have resulted in a high proportion of conserved nucleotides and amino acids.

Acknowledgments—We thank Dr. Robert Goldberg for making his λ Charon 4A soybean library and the Kunitz inhibitor clone available. We thank Dr. Niels Nielsen for use of his 11 S storage protein cDNA clone. We also acknowledge the technical assistance of Judith Lindell and Pamela Gutay.

REFERENCES

Bailey, J. M., and Davidson, N. (1976) Anal. Biochem. 70, 75-85
Bennoist, C., O’Hare, K., Breathnach, R., and Chambon, P. (1980) Nucleic Acids Res. 8, 127-142
Berk, A. J., and Sharp, P. A. (1977) Cell 12, 721-732
Birk, Y. (1981) Biochim. Biophys. Acta 54, 378-381
Blattner, F. R., Blechl, A. E., Denniston-Thompson, K., Faber, H. E., Richards, J. E., Slightom, J. L., Tucker, P. W., and Smithies, O. (1978) Science (Wash. D. C.) 202, 1279-1284
Blattner, F. R., Williams, B. G., Blechl, A. E., Denniston-Thompson, K., Faber, H. E., Furlong, L., Grunwal, D. J., Kierer, D. O., Moore, D. D., Schumann, J. W., Sheldon, E. L., and Smithies, O. (1977) Science (Wash. D. C.) 196, 161-169
Bryant, J., Green, T. R., and Gurusaddia, T., and Ryan, C. A. (1976) Biochemistry 15, 3418-3424
Buell, G. N., Wickens, M. P., Payvar, F., and Schimke, R. T. (1978) J. Biol. Chem. 253, 2471-2482
Casey, J., and Davidson, N. (1977) Nucleic Acids Res. 4, 1539-1552
Chu, H. M., and Chi, C.-W. (1965) Sci. Sinica 14, 1441
Croy, R. R. D., Lycett, G. W., Gatehouse, J. A., Yarwood, J. N., and Boulter, D. (1982) Nature (Lond.) 295, 76-79
Fischer, R. L., and Goldberg, R. B. (1982) Cell 29, 651-660
Foard, D. E., Gutay, P. A., Ladin, B., Beachy, R. N., and Larkins, B. A. (1982) Plant Mol. Biol. 1, 127-143
Goldberg, R. B., Hoschek, G., Kamalay, J. C., and Timberlake, W. E. (1981a) Dev. Biol. 83, 201-217
Goldberg, R. B., Hoschek, G., Ditta, G. S., and Breidenbach, R. W. (1981b) Dev. Biol. 83, 218-231

Grunstein, M., and Hogness, D. S. (1975) Proc. Natl. Acad. Sci. U. S. A. 72, 3961-3965
Horisberger, M., and Tachlini-Vonlanthen, M. (1983) Histochemistry 77, 313-321
Hwang, D. L.-R., Lin, K.-T. D., Yang, W.-K., and Foard, D. E. (1977) Plant Physiol. 60, 369-382
Hwang, D. L., Yang, W.-K., Foard, D. E., and Lin, K.-T. D. (1978) Plant Physiol. 61, 30-34
Jones, G., Moore, S., and Stein, W. H. (1963) Biochemistry 2, 66-71
Kafatos, F. C., Jones, C. W., and Efstratiadis, A. (1979) Nucleic Acids Res. 7, 1541-1552
Kanamori, M., Ioku, F., Tashiro, M., Yamada, M., and Miyoshi, M. (1976) Biochim. Biophys. Acta 439, 398-405
Kunitz, M. (1946) J. Gen. Genet. 29, 149-154
Laskowski, M., and Kato, I. (1980) Annu. Rev. Biochem. 49, 593-626
Laskey, R. A., and Mills, A. D. (1975) Eur. J. Biochem. 56, 335-341
Liener, I. E. (1979) J. Rev. Biochem. 27, 97-122
McGrogan, M., Specior, D. J., Goldenberg, C. J., Halbert, D., and Raskas, H. J. (1979) Nucleic Acids Res. 6, 593-607
Maniatis, J., Hardison, R. C., Lacy, E., Lauer, J., O’Connell, C., Quon, D., Sim, G. K., and Efstratiadis, A. (1978) Cell 15, 687-701
Maniatis, T., Kee, S. G., Efstratiadis, A., and Kafatos, F. C. (1976) Cell 8, 163-182
Maxam, A. M., and Gilbert, W. (1980) Methods Enzymol. 65, 499-500
Murray, M. G., and Thompson, W. F. (1980) Nucleic Acids Res. 8, 4321-4325
Odani, S., and Ikenaka, T. (1977) J. Biochem. (Tokyo) 82, 1513-1522
Odani, S., and Ikenaka, T. (1978) J. Biochem. (Tokyo) 83, 737-745
Pedersen, K., Devereux, J., Wilson, D. E., Sheldon, E., and Larkins, B. A. (1982) Cell 29, 1015-1026
Perler, F., Efstratiadis, A., Lomedico, P., Gilbert, W., Kolodner, R., and Dodgson, J. (1980) Cell 20, 555-566
Proudfoot, N. J., and Brownlee, G. G. (1976) Nature (Lond.) 263, 211-214
Richardson, M. (1977) Phytochemistry (Oxf.) 16, 159-169
Ryan, C. (1973) Annu. Rev. Plant Physiol. 24, 173-196
Schuler, M. A., Schmitt, E. S., and Beachy, R. N. (1982a) Nucleic Acids Res. 10, 8225-8244
Schuler, M. A., Ladin, B. F., Pollacco, J. C., Frey, R. G., and Beachy, R. N. (1982b) Nucleic Acids Res. 10, 8245-8261
Smith, D. R., and Cabro, J. M. (1980) Nucleic Acids Res. 8, 2255-2274
Southern, E. M. (1975) J. Mol. Biol. 98, 503-517
Tan, C. G. L., and Stevens, F. C. (1971) Eur. J. Biochem. 18, 515-523
Thomas, P. S. (1980) Proc. Natl. Acad. Sci. U. S. A. 77, 5201-5205
Vodkin, L. O. (1981) Plant Physiol. 68, 766-771
Wahl, G. M., Stern, M., and Stark, G. R. (1979) Proc. Natl. Acad. Sci. U. S. A. 76, 3683-3687
Wickens, M. P., Buell, G. N., and Schimke, R. T. (1978) J. Biol. Chem. 253, 2483-2495
Wilson, K. A., and Laskowski, M., Sr. (1973) J. Biol. Chem. 248, 756-762
Molecular cloning and analysis of a gene coding for the Bowman-Birk protease inhibitor in soybean.
R W Hammond, D E Foard and B A Larkins

J. Biol. Chem. 1984, 259:9883-9890.

Access the most updated version of this article at http://www.jbc.org/content/259/15/9883

Alerts:
• When this article is cited
• When a correction for this article is posted

Click here to choose from all of JBC’s e-mail alerts

This article cites 0 references, 0 of which can be accessed free at http://www.jbc.org/content/259/15/9883.full.html#ref-list-1