JOURNAL OF AGRICULTURAL AND FOOD CHEMISTRY

Protein Structures of Common Bean (*Phaseolus vulgaris*) α-Amylase Inhibitors

SHIH-CHIEH LEE,† PAUL L. GEPTS,‡ AND JOHN R. WHITAKER*,§

AdvPharma, Inc., Tainen, Taiwan, Republic of China, and Department of Agronomy and Range Science and Department of Food Science and Technology, University of California, Davis, California 95616

Two nucleotide sequences for genes that encode α-amylase inhibitor 4 (αAI-4) from white kidney bean (WKB) cv. 858, designated gene α*AI*-4 (Accession No. U84390), and α-amylase inhibitor 5 (αAI-5) from black bean (BB), designated gene α*AI*-5 (Accession No. U84391), were determined. Genes α*AI*-4 and α*AI*-5 encode 244 amino acid prepro-αAI-4 and prepro-αAI-5 polypeptides that are 93 and 95% identical with α-amylase inhibitor l (αAI-l; Hoffman, L. M.; Ma, Y.; Barker, R. F. *Nucleic Acids Res.* **1982**, *10*, 7819−7828), 40 and 43% identical with red kidney bean lectin, and 52 and 55% identical with arcelin l of wild-type bean. The high degree of sequence similarity indicates the evolutionary relationship among these genes. PCR analysis of genomic DNA purified from six genotypes of *Phaseolus vulgaris* showed very similar band patterns in 2% agarose gel, another indication of the conserved size homology among these genes. Proteolytic processing sites were located between Asn77 and Ser78 for pro-αAI-4 and pro-αAI-5. A bend next to Asn77 in three-dimensional model structures of αAI-4 and αAI-5 proinhibitors indicates that the proteolytic cleavage is necessary to remove the conformational constraint for activation to the mature protein. Mature WKB αAI-4 was composed of four subunits (2α2*β*) and had a molecular weight of 50000 determined by multiangle laser light scattering and 56714 determined by laser-assisted time-of-flight mass spectrometry.

KEYWORDS: α-Amylase inhibitors; protein structures; gene sequences; proteolytic processing sites; preproinhibitors; predicted secondary, tertiary, and quaternary structures

INTRODUCTION

Enzymes that catalyze hydrolysis of the α-(1,4)-glycosidic bonds between α-D-glucopyranose residues of polymers are amylases. α-Amylases (EC 3.2.1.1) are the enzymes that release products with an α-D-configuration. α-Amylase inhibitors (αAIs) are compounds that inactivate some α-amylases by forming enzyme−inhibitor complexes. On the basis of their chemical nature, there are three types of αAIs, namely, (a) the microbial N-containing carbohydrates, (b) the microbial peptides, and (c) the higher plant proteins (*2*).

Leguminosae seeds provide a very important source of food for humans but, once harvested and stored, they are frequently damaged by insects. Most cultivars of common bean (*Phaseolus vulgaris*) contain one or more proteinaeous αAIs, which inhibit the activity of some mammalian and insect α-amylases but not that of the endogenous plant enzyme (*3*, *4*). These proteinaeous αAIs inhibit the α-amylases of *Tenebrio molitor* (*5*, *6*), *Callosobruchus maculatus*, *Callosobruchus chinensis* (*7*), and *Zabrotes subfasciatus* (*8*, *9*). lnvestigation of naturally occurring αAIs is important because of their effect in the control of specific insects, their effect on human nutrition, and their use as analytical reagents. Shade et al. (*10*) transferred the common bean αAI gene to cowpea, giving resistance to certain bruchid beetles, confirming the importance of this protein for plant defense.

The α*AI*-I gene encodes αAI-l (LLP) (*1*, *11*), which was the only αAI gene sequence published for common domesticated beans (*P. vulgaris*) prior to the results reported in this paper. Two other known genes encode αAIs; genes α*AI*-2 and α*AI*-3 have been isolated from wild common bean and the sequences published (*12*, *13*). The product of gene α*AI*-3 actually is an αAI-like protein (M. J. Chrispeels, personal communication). αAIs belong to a lectin−arcelin−αAI supergene family of common bean (*P. vulgaris*) (*12*). Arcelins and αAIs are truncated lectins (*14*). The genes encoding all of these proteins were shown to be closely related, exhibiting a high degree of sequence similarity: 90% between *dlec1* (encoding PHA-E) and *dlec2* (encoding PHA-L); 82% between α*AI*-1 (encoding αAI-l) and *dlec2*; and 78% between the cDNA of arcelin l and the gene coding for PHA-L (*15*). The genes for these three proteins are encoded by a single complex locus in the *P. vulgaris* genome (*16*); it is likely that these homologous genes evolved by duplication of a single ancestral gene (*12*).

* Corresponding author [telephone (530) 753-2381; fax (530) 752-4759; e-mail jrwhitaker@ucdavis.edu].
† Department of Food Science and Technology.
‡ Department of Agronomy and Range Science.
§ Department of Food Science and Technology.

10.1021/jf020189t CCC: $22.00
Published on Web 09/24/2002

αAI-l of common bean seeds differs uniquely from arcelin and lectin in that it is synthesized as a preproprotein that undergoes specific proteolytic maturation (*17*). αAI-l is synthesized as a preproprotein on the endoplasmic reticulum, and it is proteolytically processed after arrival in the protein storage vacuoles to polypeptides of molecular weights of 15000−18000 (*18*). αAI-l protein is inactive as an inhibitor until the polypeptide has been proteolytically processed at Asn77 (*17*). Such maturation proteolytic processing of proproteins probably converts an inactive proprotein into an active mature protein through the release of a conformational constraint within the proprotein (*17, 19*). Representative three-dimensional structures have been determined for lectins (*20*). The similarities among αAI-l and PHA-E and other legume lectins make it possible to model the three-dimensional structures of αAIs on the basis of the three-dimensional structures of legume lectins that have been determined crystallographically (*21*). The results from chemical modification and mutants of common bean αAIs indicate that the active site most likely consists of Trp188, Arg74, and Tyr190, in analogy to the Trp18-Arg19-Tyr20 motif of Tendamistat (*22, 23*). lnterpretation of the crystal structure of pig pancreatic α-amylase complexed with αAI-l provides some alternative explanations for essentiality of one or more of these three residues (*19*).

Several similarities among different bean αAIs and with other bean proteins are now apparent (*2*). All known αAIs contain two or more subunits. Most of the αAIs have alanine and serine as N-terminal residues, and the partial N-terminal sequences can be aligned into two major groups, indicating two peptides are present. Ho et al. (*2*), on the basis of their and others' results for several mature bean αAIs and their subunit diversity, proposed a model as a working hypothesis. The structure of some inhibitors, including the isoinhibitors found in black bean (I-2 of the 55000 Da and I-1 of the 60000 Da proteins), might be explained by association of proteolytic fragments. The 23000 Da subunit from black bean isoinhibitor I-1 (*24*) could possibly be related to the 27700 Da unit proposed by Moreno and Chrispeels (*25*). However, αAIs from different beans have quite different chemical, physical, and specificity properties. It has been suggested that they may be different in molecular weights, p*I* values, carbohydrate contents, subunit numbers, and specific binding constants with different α-amylases (*2*). The objective of this research was to determine if the different physical, chemical, and specificity properties reported for αAIs arise at the gene level or by post-translational proteolytic modification.

MATERIALS AND METHODS

White kidney bean (WKB; *P. vulgaris*) cultivar 858, black bean (BB; *P. vulgaris*), and red kidney bean (RKB; *P. vulgaris*) cultivar RKB were obtained from the Department of Agronomy and Range Science, University of California, Davis, CA. Red kidney bean (bin 6110) and black bean (unspecified cultivar) were obtained from a Davis grocery store.

Genomic DNA Preparation and PCR Analysis. White kidney bean cv. 858 and black bean (*P. vulgaris*) were used. Tissue for DNA extraction was purified from trifoliate leaf tissue from greenhouse-grown plants (using the youngest tissue possible) according to the procedure of Afanador et al. (*26*). The design of PCR primers was based on two conserved amino acid sequences located in the amino terminus ($MASS^{K}/_{N}$) and carboxy terminus (LN^{Q}/IL) of PHA-E, PHA-L, αAI, and arcelin (*15*). The primer sequences, synthesized by the University of California Davis Protein Structural Laboratory, were P1, 5′-$CGCTCGAGGAGGAT^{C}/_{T}T^{G}/_{T}GTTGA$-3′, and P2, 5′CAGAATTCATG-GCTTCCTCCAA-3′. A 200 ng portion of each primer and 0.25 μg of genomic DNA were used for PCR in 100 μL of reaction mixture, which contained 10 mM Tris-HCI, pH 9, 500 mM KCl, 1.0% Triton X-100, 1.5 mM $MgCl_2$, 0.2 mM each dNTP, and 1 unit of Taq DNA polymerase (Promega). Conditions for PCR were 95 °C for 3 min; then 40 cycles of 94 °C for 1 min, 55 °C for 1 min, and 72 °C for 1 min; and 1 cycle of 72 °C for 5 min. The PCR sample was cloned into pCR II vector by the TA cloning kit (Invitrogen, San Diego, CA) and sequenced by the Sequenase DNA sequencing kit (USB, Cleveland, OH).

Sequence Analysis. Nucleotide sequences were translated to the corresponding protein sequences by the Translate software program, which was provided by the ExPASy World Wide Web molecular biology server from Geneva University Hospital and University of Geneva, Geneva, Switzerland. Deduced amino acid sequences were aligned using the PILEUP and PRETTY software programs of the University of Wisconsin Genetics Computer Group (*27*), and the dendrogram was created using the software program FIGURE. The NNPREDlCT program was provided by the World Wide Web server from the University of California, San Francisco, CA. The software program MSA was provided by the World Wide Web server from Washington University (St. Louis, MO). The software programs ScanProsite and Pro-Mod were provided by the ExPASy World Wide Web molecular biology server from Geneva University Hospital and University of Geneva, Geneva, Switzerland. The coordinates of αAI-4 and αAI-5 three-dimensional structures were built on the basis of the coordinates of soybean lectin (*28*), legume isolectins I (*29*) and II (*30*), pea lectin (*31*), and lentil lectin (*32*).

Purification of α-Amylase Inhibitors. The WKB αAI was purified to homogeneity by DEAE-cellulose, CM-cellulose, and Sephadex G-100 column chromatography as described by Powers and Whitaker (*5*) but with two modifications. These included the addition of a Sephadex G-100 gel filtration column chromatography step (1 × 95 cm column) and use of 0.03 and 0.12 M NaCl in buffers 1 and 2, respectively, for DEAE-cellulose column chromatography, rather than 0.1 and 0.15 M NaCl as used by Powers and Whitaker (*5*).

Inhibitory Activities. Inhibitory activity against procine pancreatic α-amylase (EC 3.2.1.1; type IA, Sigma) was determined according to the method of Bernfeld (*33*) with modification. The substrate was 1% $NaBH_4$-reduced starch (*34*) in 0.04 M potassium phosphate buffer, pH 6.9, containing 0.05 M NaCl. A mixture of porcine pancreatic α-amylase solution (1.0×10^{-7} M, 0.1 mL) and inhibitor solution (0.1 mL) was incubated in a water bath for exactly 30 min at 30 °C, pH 5.4, prior to the addition of substrate solution (0.8 mL of 1% $NaBH_4$-reduced starch) for enzyme−substrate reaction for 5.0 min. A standard curve for reducing group concentration was prepared using maltose. One inhibitor unit is defined as that amount of inhibitor which gave 50% inhibition of an aliquot of porcine pancreatic α-amylase that released reducing sugar equivalent to 0.1 mg of maltose per minute at pH 6.9 and 30 °C when acting on 1% $NaBH_4$-reduced starch for 5 min.

N-Terminal Amino Acids and Sequences. N-Terminal sequences of the purified inhibitors were determined by the Protein Structure Laboratory at the University of California, Davis, CA, using the automated Edman degradation method (*35*).

Molecular Weight Determinations. *HPSEC-MALLS Method.* Molecular weights were determined by a high-pressure size exclusion chromatography system with on-line multiple-angle laser light scattering (HPSEC-MALLS) and differential refractive index (DRI) detectors (*36*). Double-deionized water, passed through a 0.2 μm Barnstead hollow fiber filter, was used in the experiment. The eluent was 0.1 M $NaNO_3$ with 0.02% (w/w) added NaN_3 for microbial stability. The flow rate was 0.66−0.67 mL/min. A Waters Ultrahydrogel column (Waters Chromatography) was used. The column was held in block heaters at 40 °C, and the refractometer was temperature controlled using a water bath at 37 °C. The rest of the system was at room temperature. The injection volume was 200 μL. The experiment was done in triplicate for each sample of αAI.

Laser-Assisted Time-of-Flight Mass Spectrometry (LATOFMS). The method used to determine the molecular weight and subunit structure of αAI by LATOFMS was described by Hutchens and Yip (*37*).

RESULTS AND DISCUSSION

PCR Analysis of Genomic DNA Purified from Various Common Beans. Chrispeels and Raikhel (*15*) reported that the

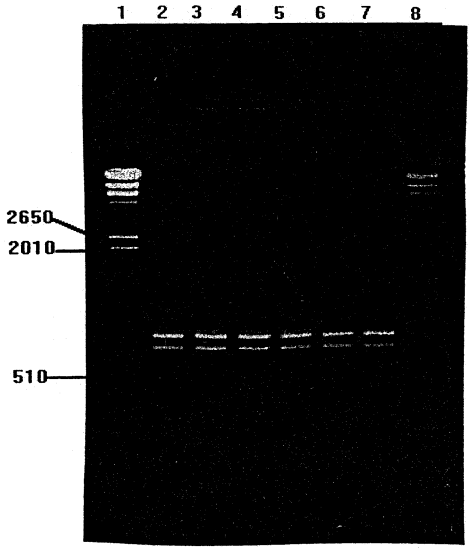

Figure 1. PCR analysis of genomic DNA purified from various *Phaseolus* genotypes: lane 1, MW markers; lane 2, WKB cv. WKB; lane 3, WKB cv. 858; lane 4, RKB (Food Corp.); lane 5, RKB (Safeway); lane 6, RKB cv. RKB; lane 7, BB (*24*); lane 8, MW markers. Ethidium bromide-stained 2.0% agarose gel shows PCR amplification products generated using primer pairs that were designed on the basis of the conserved terminal 5′ and 3′ coding regions of PHA-E, PHA-L, αAI, and arcelin l. DNA molecular weight markers are indicated (in base pairs) on the left.

genes of PHA-E, PHA-L, arcelin, and αAI have conserved amino-terminal ($MASS^{K}/_{N}$) and carboxy-terminal ($LN^{Q}/_{K}IL$ amino acid residues. We designed our PCR primers on the basis of these two conserved amino acid sequences. When the Polymerase Chain Reaction was used with genomic DNA from six genotypes of beans, at least three bands with different size DNAs were obtained. Their sizes were 758, 840, and 898 bp. These DNA bands showed two size groups in ethidium bromide-stained 2% agarose gel (**Figure 1**). Moreno and Chrispeels (*25*) reported that genes for PHA and αAIs do not have introns and that there are several short deletions in the amino acid sequence of αAIs when compared to that of PHA. Mirkov et al. (*12*) reported that the high molecular weight band (**Figure 1**) represents the gene for the longer PHA sequences and the low molecular weight band represents the gene for the αAI sequences. Our results are consistent with these observations. However, at least two bands were found in the higher molecular weight (840−898 bps) region. Because arcelins are only found or expressed in wild accessions of beans (*P. vulgaris*), these two bands may represent two lectin genes. It is noteworthy that the band patterns in the gel were very similar for all six different bean cultivars. It is interesting, and important, that the gene sizes for the putative αAI sequences from the six beans were similar (**Figure 1**). However, the mature αAI proteins purified from these six beans have different physical and chemical properties (*4*). In addition, some beans, such as black bean, are known to contain two different isoinhibitors, yet only one band was seen in the gel for that black bean (lane 7). The low molecular weight DNA bands (748 bp) for white kidney bean cv. 858 and black bean αAIs were size selected first by using 2% agarose gel before these DNAs were cloned.

Deduced Amino Acid Sequence of White Kidney Bean and Black Bean αAIs. Figure 2 shows the deduced amino acid sequence of a genomic DNA encoding white kidney bean cv. 858 αAI and black bean αAI. From now on, the inhibitors will be labeled αAI-4 and αAI-5 because they are the fourth and fifth αAI sequences reported. The arrows 1 and 2 indicate sites of proteolytic processing of the preproinhibitor for removal of the signal peptide to form proinhibitor and the processing site of proinhibitor to form mature active αAI, respectively. The underscores indicate the position of the N-terminal amino acid sequences of α- and *β*-peptides obtained for the mature αAI subunits by the Edman degradation method in the current work. The nucleotide sequences for the preproinhibitors can be found in the Gene Register (Accession No. U84390 and U84391).

lnitiation at the ATG codon yielded proteins with 244 amino acid sequences. Comparison of the deduced amino acid sequences from the DNA sequence with the two Edman degradation determined N-terminal amino acid sequences of purified mature αAI-4 and αAI-5 proteins indicated that genomic DNA clone α*AI*-4 encodes αAI-4 and α*AI*-5 encodes αAI-5. The sequences match exactly the deduced amino acid sequences beginning at base positions 64 and 295, respectively. The 21 amino acid residue polypeptide starting from the initiation codon was not detected in the mature αAI-4 and αAI-5 proteins; it is considered to represent the signal peptide. **Table 1** shows the pattern and profile search results for αAI-4 and αAI-5 using ScanProsite. The glycoprotein features of αAI-4 and αAI-5 are consistent with the finding that the three potential N-glycosylation sites are found in the deduced sequences of the α*AI*-4 and α*AI*-5 genes. These occur at the same amino acid positions in αAI-4 and αAI-5. In addition to the potential N-glycosylation sites, the protein kinase C phosphorylation sites, casein kinase

Table 1. Amino Acid Pattern and Profile Search Results for Pro-αAI-4 and Pro-αAI-5

		αAI-4		αAI-5	
1	N-glycosylation sites	1	33–36 NKTN	1	33–36 NKTN
		2	86–89 NFTM	2	86–89 NFTM
		3	161–164 NSST	3	161–164 NSST
2	protein kinase C phosphorylation sites	1	93–95 THR	1	93–95 THR
		2	163–165 STK	2	163–165 STK
		3	176–178 TER	3	176–178 TER
		4	222–224 SSK	4	222–224 SSK
		5	233–235 SER	5	233–235 SER
3	casein kinase II phosphorylation sites	1	21–24 SATE	1	21–24 SATE
		2	114–117 SKGD	2	114–117 SKGD
				3	184–187 TTVE
4	N-myristoylation sites	1	78–83 GNVASF	1	78–83 GNVASF
5	legume lectins *β*-chain signature	1	119–125 VTVEFDT	1	119–125 VTVEFDT
6	legume lectins α-chain signature	1	192–201 VYDWVSVGFS	1	192–201 VYDWVSVGFS

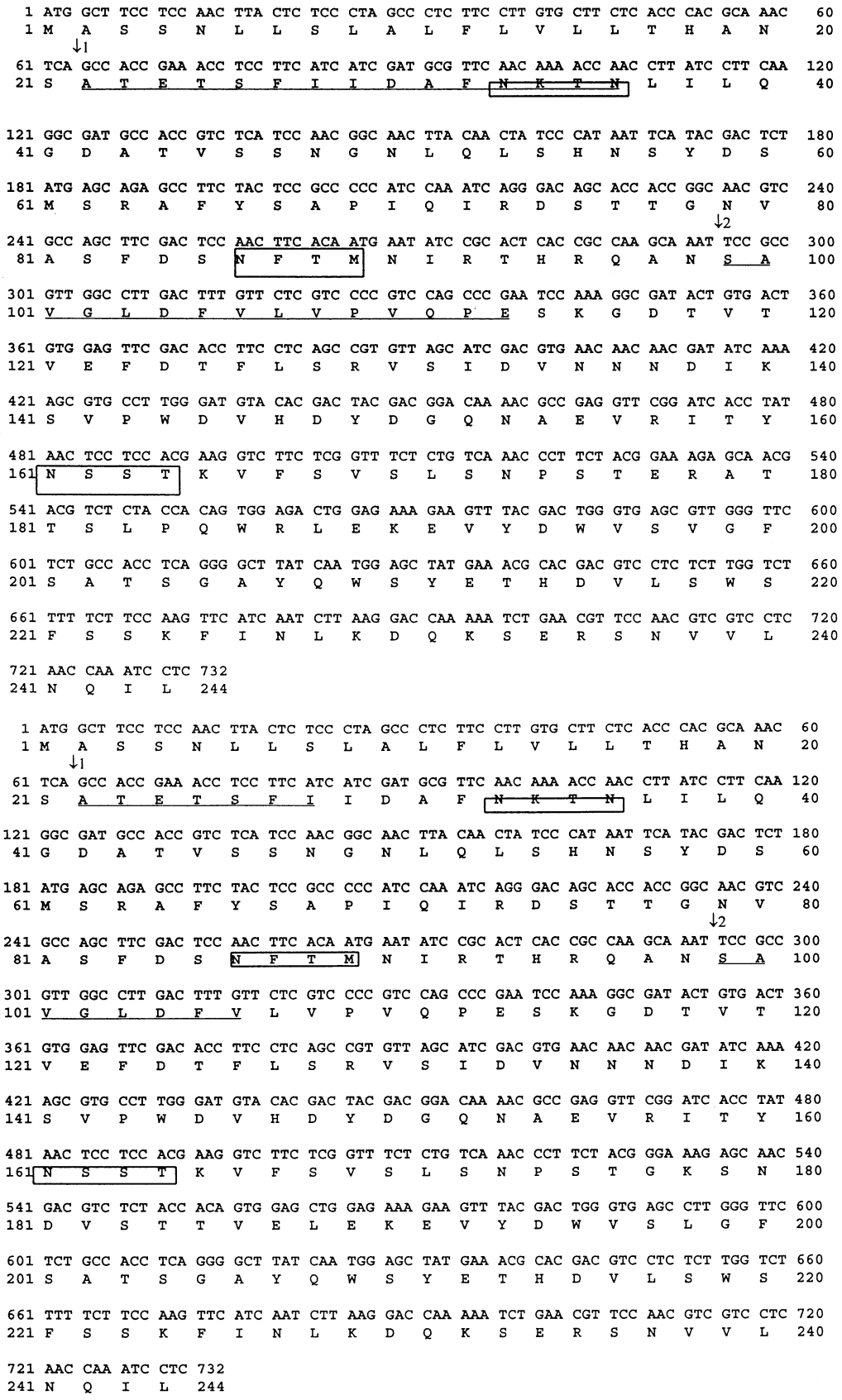

Figure 2. Deduced amino acid sequence of genomic DNAs encoding white kidney bean cv. 858 α-amylase inhibitor (top; designated αAI-4) and black bean α-amylase inhibitor (bottom; designated αAI-5). Arrows 1 and 2 indicate the presumed sites of proteolytic processing of the preproinhibitor to proinhibitor (and signal peptide) and to mature inhibitor, respectively. The N-terminal amino acid sequences of peptide 1 and peptide 2 obtained by Edman degradation of the mature αAI-4 and αAI-5 proteins are underlined. Boxes are used to indicate N-glycosylation consensus sequences.

II phosphorylation sites, N-myristoylation site, and the legume lectins α-chain and β-chain signatures were found in αAI-4 and αAI-5 (**Table 1**). All of these potential post-translational modification sites occur at the same location in αAI-4 and αAI-5

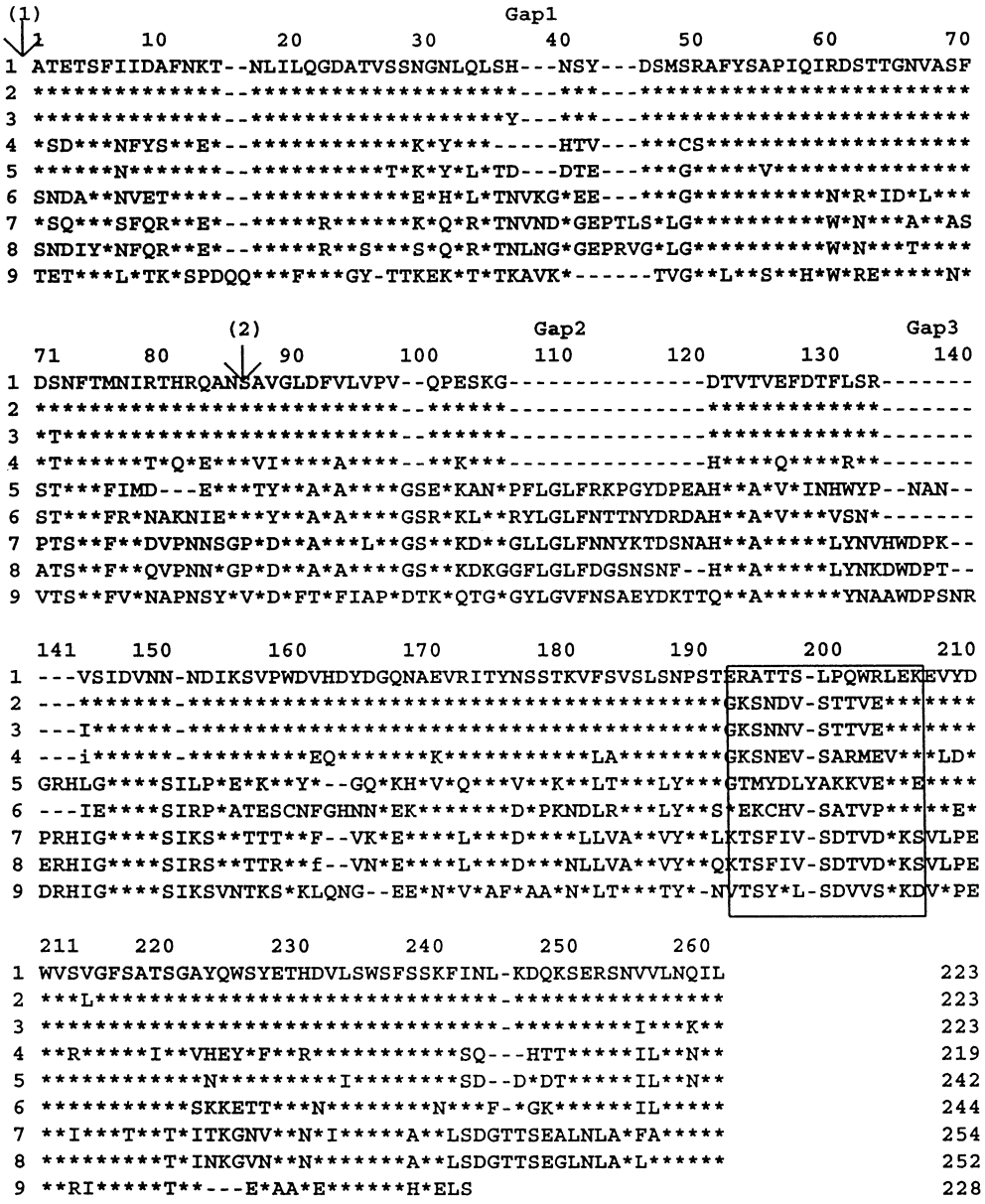

Figure 3. Amino acid sequences of five mature α-amylase inhibitors, arcelin 1, and three lectins: I, white kidney bean cv. 858 αAI-4; 2, black bean αAI-5; 3, greensleeves bean αAI-I (*1*, *11*); 4, wild common bean αAI-2 (*13*); 5, wild common bean αAI-3-like protein (has no inhibitory activity; *23*); 6, arcelin I (*38*); 7, red kidney bean PHA-E (*39*); 8, red kidney bean PHA-L (*39*); 9, pea lectin (*40*); *, same as the amino acid in αAI-4; –, gap inserted for alignment; arrow 1 or 2, sites of proteolytic processing. Sequences in the box indicate the less homologous region.

except for the third casein kinase II phosphorylation site in αAI-4, where it is absent. It is replaced by the sequence PQWR.

Comparison of the Amino Acid Sequences of α-Amylase Inhibitors, Lectins, and Arcelin. Figure 3 is a comparison of the amino acid sequences of five αAIs, three lectins, and one arcelin. Striking amino acid sequence homology is noted among the nine proteins with αAI-4 having the greatest homology with αAI-5 (95.1% identity) in this *P. vulgaris* lectin−arcelin−αAI family. αAI-4 amino acid sequence homologies to the other members of this family are as follows: 93.4% identity to LLP; 40.3% identity to PHA-E; 74.9% identity to αAI-2; 55.6% identity to αAI-3; 51.9% identity to arcelin 1; and 41.2% identity to PHA-L. αAI-5 has amino acid sequence homology to the other members of this family, too (**Table 2**). The high degree of sequence homology and linkage relationships among them strongly suggest that these genes evolved from a common ancestral gene.

Pueyo et al. (*17*) reported that the activation of bean αAI requires proteolytic processing of the proprotein. This processing occurs in the vacuole at the Asn before arrow 2 in **Figure 3**. Processing at an Asn residue is very common in several legume seed proteins, and the processing protease has recently been purified (*41*, *42*). The Asn residue is always followed by a Ser

Table 2. Amino Acid Sequence Homology among α-Amylase Inhibitors, Arcelin 1, and Lectins[a]

	αAI-4	αAI-5	αAI-1	αAI-2	αAI-3	ARL-1	PHA-E	PHA-L
αAI-4								
αAI-5	95.1							
αAI-1	93.4	97.54						
αAI-2	74.9	77.4	77.4					
αAI-3	55.6	57.6	57.6	57.7				
ARL-1	51.9	55.1	55.1	55.6	59.0			
PHA-E	40.3	43.2	43.2	42.7	49.4	53.0		
PHA-L	41.2	43.6	43.6	42.3	50.2	55.5	82.5	
lectin[b]	33.7	35.0	35.8	38.1	36.4	35.6	40.1	38.7

[a] Data from **Figure 3** expressed in percent. [b] Pea lectin.

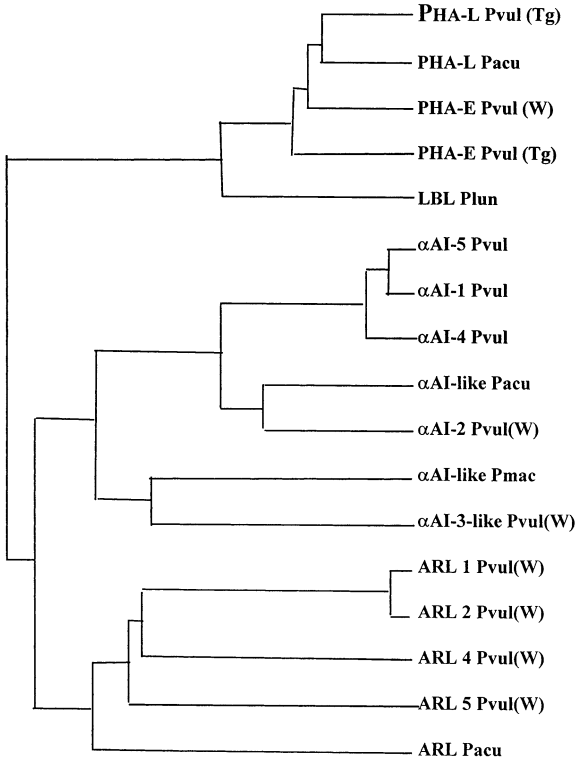

Figure 4. Cluster dendrogram by UWGCG's PlLEUP and FlGURE programs, based on the amino acid sequences of 17 genes in the *Phaseolus*–lectin–arcelin–αAI gene family. PHA, phytohemagglutinin; LBL, lima bean lectin; αAI, α-amylase inhibitor; ARL, arcelin; Pvul, *P. vulgaris*; Pacu, *P. acutifolius*; Pmac, *P. maculatus*; Plun, *P. lunafus*; W, wild; Tg, Tendergreen.

residue. An Asn−Ser sequence is found in all αAIs and arcelin 1, but it is absent in the lectins (PHA-E, PHA-L and pea lectin) (**Figure 3**). Because arcelin l is not proteolytically processed and does not have α-amylase inhibitory activity, the data indicate that the Asn85-Ser86 sequence at that position is necessary but not by itself sufficient to ensure proteolytic processing. αAI-4 and αAI-5 have three gaps (gaps 1−3 in **Figure 3**) compared to the lectins. These results are consistent with the report that αAIs from seeds of common bean (*P. vulgaris*) are truncated lectins (*14*). Not all αAIs contain all three gaps (**Figure 3**). Cluster dendrograms by University of Wisconsin Genetics Computer Group's PlLEUP and FlGURE programs were produced on the basis of that sequence alignment. αAI-4 and αAI-5 contain all three gaps and belong in the same group with αAI-l (LLP) (**Figure 4**). Note that α*AI*-2 and three α*AI*-like genes are found in two additional groups.

The NNPREDlCT program predicts the secondary structure type for each residue in an amino acid sequence (*43*, *44*). **Figure 5** shows the secondary structure predictions for αAI-4 and αAI-5 preproproteins. It is noteworthy that these two αAI preproproteins show very similar secondary structures with only one exception, which is located in the 186−197 region of αAI-4. Suzuki et al. (*13*) suggested that the differences in the primary structures of αAI-l and αAI-2 may be the cause for distinct inhibitory activity differences on bruchids (*9*, *45*).

Three-Dimensional Models of α-Amylase lnhibitors. Three-dimensional models of αAI-4 and αAI-5 were generated by a knowledge-based, protein modeling tool Pro-Mod (*46*). The three-dimensional model of the αAI-4 molecule contains 86 residues in a backbone display style (**Figure 6a**). The residue corresponding to Asn85 in **Figure 3** is shown in the CPK (Corey, Pauling, and Koltun) style (**Figure 6a**). The Asn85 is next to the proteolytic processing site in the primary sequence. A bend structure can be seen near this site in the three-dimensional model structure. This finding is consistent with the results reported by Pueyo et al. (*17*) that removal of the

```
(a) αAI-4
1                        ↓                                                  72
----HHHHHHHHHHHHHH-------EEEEHH-----HEEE----E-----EEE-------HHHH-----EEE
MASSNLLSLALFLVLLTHANSATETSFIIDAFNKTNLILQGDATVSSNGNLQLSHNSYDSMSRAFYSAPIQI

73                             ↓                                          144
E------E--------EEHH-----------EEE------------EEEEHHHH--EEEE------------
RDSTTGNVASFDSNFTMNIRTHRQANSAVGLDFVLVPVQPESKGDTVTVEFDTFLSRVSIDVNNNDIKSVPW

145                                        186      197                 216
----------EEEEEE----EEEEEE---------------HHHHHHHHEEEE--E-----EEEEH----E
DVHDYDGQNAEVRITYNSSTKVFSVSLSNPSTERATTSLPQWRLEKEVYDWVSVGFSATSGAYQWSYETHDV

217                       244
EEE------E-----------HHH----
LSWSFSSKFINLKDQKSERSNVVLNQIL


(b) αAI-5
1                        ↓                                                  72
----HHHHHHHHHHHHHH-------EEEEHH-----HEEE----E-----EEE-------HHHH-----EEE
MASSNLLSLALFLVLLTHANSATETSFIIDAFNKTNLILQGDATVSSNGNLQLSHNSYDSMSRAFYSAPIQI

73                             ↓                                          144
E------E--------EEHH-----------EEE------------EEEEHHHH--EEEE------------
RDSTTGNVASFDSNFTMNIRTHRQANSAVGLDFVLVPVQPESKGDTVTVEFDTFLSRVSIDVNNNDIKSVPW

145                                       184       197                 216
----------EEEEEE----EEEEEE-------------EHHHHHHHHHHHEE--E-----EEEEH----E
DVHDYDGQNAEVRITYNSSTKVFSVSLSNPSTGKSNDVSTTVELEKEVYDWVSLGFSATSGAYQWSYETHDV

217                       244
EEE------E-----------HHH----
LSWSFSSKFINLKDQKSERSNVVLNQIL
```

Figure 5. Secondary structure prediction for αAI-4 and αAI-5 (H, helix; E, strand; - - -, turn element).

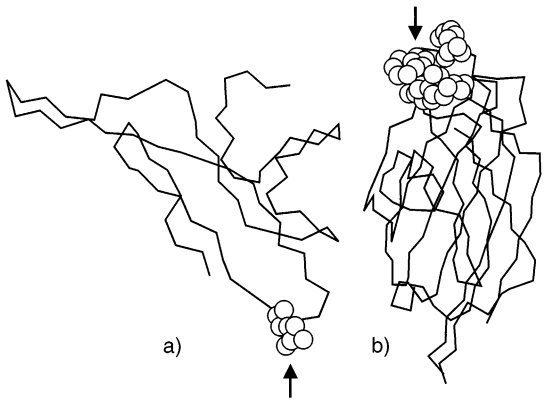

Figure 6. (a) Three-dimensional model structure for pro-αAI-4. Arrow indicates the residue Asn, which is next to the second protease cleavage site (see **Figures 2** and **3**). (b) Three-dimensional model structure for pro-αAI-5. Arrow indicates the active site pocket formed by Arg74, Trp189, and Tyr191 (see **Figure 3** for numbering).

conformational constraint is necessary for the activation to active inhibitor. It is also consistent with the crystallography data for the porcine pancreatic α-amylase−αAI-l complex (*19*).

The three-dimensional model of the αAI-5 molecule contains 167 residues in a backbone display style (**Figure 6b**). The bend near the proteolytic processing site next to Asn85 is found in the αAI-5 model also. Three residues that may form the active site pocket (Arg74, Trp188, and Tyr190; see **Figure 3** for the numbering in the mature protein) are highlighted and displayed in a CPK style (**Figure 6b**). These three residues belong to two α- and *β*-subunit chains, but they are close together in the three-dimensional structure. The major primary structural difference between αAI-4 and αAI-5 occurs only in a 12 amino acid residue sequence located between Glu192 and Arg203 (boxed in **Figure 3** for the numbering). The amino acid sequence of αAI-4 has 95.1% identity to αAI-5. However, only one-third of the sequence of αAI-4 and two-thirds of αAI-5 sequence were used by Pro-Mod in building the three-dimensional models. The sequence between Glu192 and Arg203 of the αAI-5 could not be used for model building. The search of the Protein Data Bank (PDB) did not contain enough known three-dimensional structures for building the complete three-dimensional structures of the proinhibitors αAI-4 and αAI-5. Dr. Chrispeels's group at the University of California, San Diego, CA, reported a three-dimensional model for αAI-l based on pea lectin (*23*). The same group is now trying to crystallize and determine the X-ray crystal structure of αAI-l (personal communication). Recently, Bompard-Gilles et al. (*19*) have solved the crystallographic structure of the porcine pancreatic α-amylase−αAI-l complex.

Proposed Quaternary Structure for Mature WKB Cv. 858 α-Amylase lnhibitor (αAI-4). Multiangle laser light scattering (MALLS) (*47*) can determine polymer size in solution by measuring the angular variation of the scattered light intensity. It provides absolute molecular weight, size, and conformational information. MALLS is a primary technique and does not need any primary standard calibration. The limits of the peaks for purified mature WKB cv. 858B αAI-4 were set as broadly as possible in the region containing both DRI (AUX2) and MALLS (LS11) signals (**Figure 7**). The elution volume between 14.8 and 22.9 mL contains 99.5% of the total weight of sample injected (injected weight = 0.2553 mg; calculated eluted weight = 0.2567 mg). The average of the M_W is (4.994 ± 0.9) × l0^4 g/mol.

The results of LATOFMS data obtained for purified αAI-4 are shown in **Figure 8**. There are two major clusters at *m*/*z*

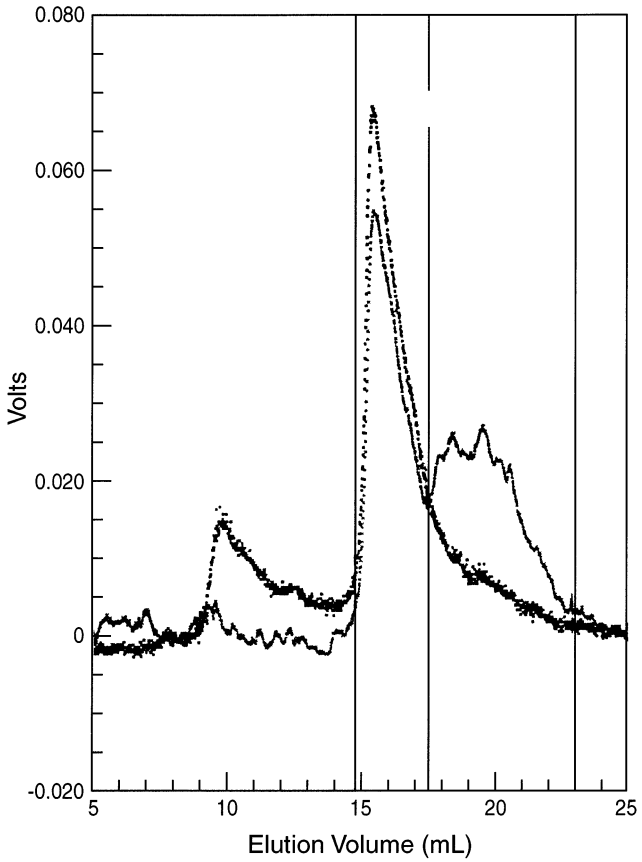

Figure 7. HPSEC-MALLS chromatogram of purified mature WKB cv. 858 α-amylase inhibitor showing where peak boundaries were set. The buffer contained 0.1 M NaN_3 and 1 mg/mL of α-amylase inhibitor.

11375 ± 225 (I) and *m*/*z* 16600 ± 1300 (II), which are in the range of the sizes of the small α-subunit (12400 Da) and the large *β*-subunit (17200 Da), based on SDS-PAGE (**Table 3**). Two peaks next to cluster II are located at *m*/*z* 17500 and at *m*/*z* 15800. The size differences among these three peaks in cluster II are equivalent to two and seven glucose residues. They may represent differences in carbohydrate content among different molecules of the *β*-subunit. Another (small) cluster located in the *m*/*z* 28375 ± 175 region could be the associated ab half-molecule, thereby providing evidence of the association of the two subunits and indicating that the $\alpha\beta$ association is stronger than the $\alpha_2\beta_2$ association found in the molecule because no cluster was found around *m*/*z* 56714 (data not shown here). In LATOFMS, the sample is subjected to a large energy pulse, which disrupts most of the interaction between the two subunits, thought to be due to hydrophobic interaction (*4*).

Comparison between the subunit weights determined by LATOFMS and cDNA permit determination of the percentage carbohydrate on each subunit (**Table 3**). The small subunit contained 25.6 wt % of carbohydrate (18 glucose residue equivalents), whereas the large subunit contained 0.94% (1 glucose residue equivalent). These results need to be confirmed by carbohydrate analysis on separated subunits.

Figure 9 is a proposed hypothetical quaternary structural model for αAI-4. The precursor of αAI-4 is a 244 amino acid long peptide; it contains a signal peptide (SP) with 21 amino acid residues (2373 Da) and a proinhibitor peptide (28357 Da). The signal peptide is removed by a specific protease (protease 1) by cleavage of the peptide bond between Ser21 and Ala22 (**Figure 2**). Another specific protease (protease 2) cleaves the peptide bond between Asn98 and Ser99 to give α- and

Table 3. Molecular Weight and Carbohydrate Content of White Kidney Bean α-Amylase Inhibitor (α-AI-4)[a]

gel filtration[b]	PAGE[c]	SDS-PAGE[d]	MALLS[e]	LATOFMS[f]	cDNA[g]	WT[h]	carbohydrate %	Glu equiv[i]
		12400	~10000–25000	11375 ± 225	8459.3	2916	25.6	18.0
		15200		16600 ± 300	16443.2	156.8	0.94	0.97
		33600	25940 ± 900[j]	28357 ± 175	24902.5	3472.5	12.2	21.5
51000	51000	45000	49940 ± 900[j]	(56714)[k]	(49805.0)[k]	6909	12.2[l]	42.6
			75210 ± 900					

[a] Purified from white kidney bean cv. 858. Unless indicated, the data are from Lee (*48*). [b] Method of Whitaker (*49*). Sephadex G-100 column. [c] Method of Hedrick and Smith (*50*). [d] Method of Weber and Osborne (*51*). [e] Multiangle laser light scattering (MALLS; *47*). [f] Laser-assisted time-of-flight mass spectrometery (LATOFMS). [g] See Lee (*48*) for methodology. [h] Weight/mol difference between LATOFMS and cDNA values. [i] Using glucose residue MW of 162. [j] Close to 45% existed in each of these two forms. [k] Assuming that the active αAI is a tetramer. [l] 11.0% by chemical determination (*4*).

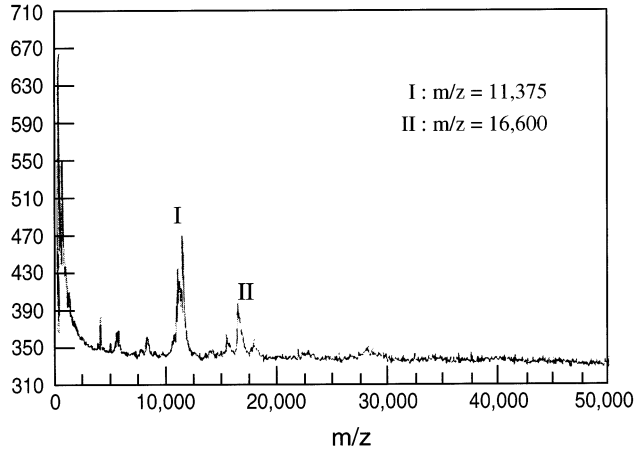

Figure 8. Mass spectra obtained for purified mature WKB cv. 858 α-amylase inhibitor.

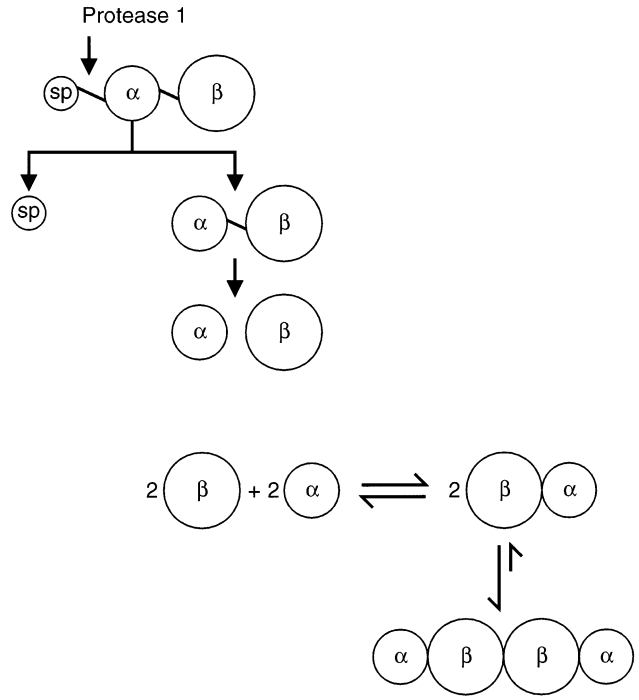

Figure 9. Model for subunit composition of WKB cv. 858 α-amylase inhibitor (SP, signal peptide; α, 11375 peptide; and *β*, 16600 peptide).

β-subunits. The α-subunits have Ala as the N-terminal residue, and the *β*-subunits have Ser as the N-terminal residue. One α-subunit associates with one *β*-subunit by hydrophobic binding (probably) to form a dimer of α*β*. Two α*β* dimers then associate to form $\alpha_2\beta_2$, the active αAI molecule (**Figure 9**), with a molecular weight of ~56000 Da when glycosylated.

The model proposed in **Figure 9** is consistent with the X-ray crystallographic data of Bompard-Gilles et al. (*19*) as well as with the data summarized in **Table 3**. There is some ambiguity on what is the nature of the molecule that is active as an inhibitor. Bompard-Gilles et al. (*19*) refer to the half-molecule (our term; weight of 28357 Da by LATOFMS in **Table 3**) as a monomer, which might be construed as the molecule (in their terminology) and the dimer proposed by them would then be association of two molecules. Higaki and Yamaguchi (*47*) suggested that the α*β* unit is the smallest component that is active. Powers and Whitaker (*5*) determined that the stoichiometery of α-amylase/inhibitor is 1:1, based on a molecular weight of 49000 Da for RKB αAI. Lee (*48*) came to a similar conclusion. Therefore, there are still some inconsistencies that need to be resolved.

In summary, the N-terminal amino acid sequences of mature αAI-4 and αAI-5 determined by Edman degradation match perfectly with the deduced amino acid sequences from the DNA sequence of the genes for αAI-4 and αAI-5. Several characteristic properties of αAIs can be explained by the deduced amino acid sequences of αAI-4 and αAI-5. These properties include potential N-glycosylation sites, amino acid composition, Asn−Ser sequence at the second proteolytic processing step (giving active inhibitor), gaps in the multiple sequence alignment, and dendrogram results. These correlations indicate that we have cloned the genes encoding αAI-4 and αAI-5.

Comparison of primary structures showed a 15 amino acid long region near one part of the postulated αAI active binding site (from Glu192 to Arg203, based on numbering in **Figure 3**) that appears to be quantitatively different between αAI-4 and αAI-5. Because the overall sequences of the two genes that encode white and black bean αAIs are very similar, it is expected that the primary and secondary structures of the proteins would be very similar. Post-translational modification and small primary structural differences as indicated for αAI-4 and αAI-5 may play important roles in the determination of the differences in physical, chemical, and specificity properties among common bean αAIs.

ACKNOWLEDGMENT

We thank Dr. T.-T. Yip and T. W. Hutchens for doing the laser-assisted time-of-flight mass spectroscopy experiment.

LITERATURE CITED

(1) Hoffman, L. M.; Ma, Y.; Barker, R. F. Molecular cloning of *Phaseolus vulgaris* lectin mRNA and use of cDNA as a probe

to estimate lectin transcript levels in various tissues. *Nucleic Acids Res.* **1982**, *10*, 7819−7828.

(2) Ho, M. F.; Yin, X.; Filho, F. F.; Lajolo, F.; Whitaker, J. R. Naturally occurring α-amylase inhibitors: Structure/function relationships. In *Protein Structure−Function Relationships in Foods*; Yada, R. Y., Jackman, R. L., Smith, J. L., Eds.; Bishopbriggs, Blackie, Academic & Professional: Glasgow, Scotland, 1994; pp 89−119.

(3) Marshall, J. J.; Lauda, C. M. Purification and properties of phaseolamine, an inhibitor of α-amylase, from the kidney bean, *Phaseolus vulgarus*. *J. Biol. Chem.* **1975**, *250*, 8030−8037.

(4) Ho, M. F.; Whitaker, J. R. Purification and partial characterization of white kidney beans (*Phaseolus vulgarus*) α-amylase inhibitors from two experimental cultivars. *J. Food Biochem.* **1993**, *17*, 15−33.

(5) Powers, J. R.; Whitaker, J. R. Purification and some physical properties of red kidney bean (*Phaseolus vulgaris*) α-amylase inhibitor. *J. Food Biochem.* **1977**, *1*, 217−238.

(6) Frels, J. M.; Rupnow, J. H. Purification and partial characterization of two α-amylase inhibitors from black bean (*Phaseolus vulgaris)*. *J. Food Biochem*. **1984**, *8*, 281−301.

(7) Ishimoto, M.; Kitamura, K. Growth inhibitory effects of an α-amylase inhibitor from the kidney bean (*Phaseolus vulgaris* L) on three species of bruchids (Coleoptera: Bruchidae). *Appl. Entomol. Zool.* **1989**, *24*, 281−286.

(8) Gatehouse, A. M. R.; Dobie, P.; Hodges, R. J.; Meik, J.; Pusztai, A.; Doulter, D. Role of carbohydrates in insect resistance in *Phaseolus vulgaris*. *J. lnsect Physiol*. **1987**, *33*, 843−850.

(9) Suzuki, K.; Ishimoto, M.; Kikuchi, F.; Kitamura, K. Growth inhibitory effect of an α-amylase inhibitor from the wild common bean resistant to the Mexican bean weevil (*Zabrotes subfasciatus*). *Jpn. J. Breed*. **1993**, *43*, 257−265.

(10) Shade, R. E.; Schroeder, H. E.; Pueyo, J. J.; Tabe, L. M.; Murdock, L. L.; Higgins, T. J. V.; Chrispeels, M. J. Transgenic pea seeds expressing the α-amylase inhibitor of the common bean are resistant to bruchid beetles. *Bio/Technology* **1994**, *12*, 793−796.

(11) Moreno, J.; Altabella, T.; Chrispeels, M. J. Characterization of α-amylase inhibitor, a lectin-like protein in the seeds of *Phaseolus vulgaris*. *Plant Physiol.* **1990**, *92*, 703−709.

(12) Mirkov, T. E.; Wahlstrom, J. M.; Hagiwara, K.; Finardi-Filho, F.; Kjemtrup, S.; Chrispeels, M. J. Evoluntary relationships among proteins in the phytohemagglution-arcelin-α-amylase inhibitor family of the common bean and its relatives. *Plant Mol. Biol.* **1994**, *26*, 1103−1113.

(13) Suzuki, K.; Ishimoto, M.; Kitamura, K. cDNA sequence and deduced primary structure of an α-amylase inhibitor from a brucid-resistant wild common bean. *Biochim. Biophys. Acta* **1994**, *1206*, 289−291.

(14) Rouge, P.; Barre, A.; Causse, H.; Chatelain, C.; Porthe, G. Arcelin and α-amylase inhibitor from the seeds of common bean (*Phaseolus vulgaris* L) are truncated lectins. *Biochem. Syst. Ecol.* **1993**, *21*, 695−703.

(15) Chrispeels, M. J.; Raikhel, N. V. Lectins, lectin genes and their role in plant defense. *Plant Cell* **1991**, *3*, 1−19.

(16) Nodari, R. O.; Tsai, S. M.; Gilbertson, R. L.; Gepts, P. Towards an integrated linkage map of common bean 2. Development of an RFLP-based linkage map. *Theor. Appl. Genet.* **1993**, *85*, 513−520.

(17) Pueyo, J. J.; Hunt, D. C.; Chrispeels, M. J. Activation of bean (*Phaseolus vulgaris*) α-amylase inhibitor requires proteolytic processing of the proprotein. *Plant Physiol.* **1993**, *101*, 1341−1348.

(18) Santino, A.; Daminati, M. G.; Virale, A.; Bollini, R. The α-amylase inhibitor of bean seed−2-step proteolytic maturation in the protein storage vacuoles of the developing cotyledon. *Physiol. Plant.* **1992**, *85*, 425−432.

(19) Bompard-Gilles, C.; Rousseau, P.; Rouge, P.; Payan, F. Substrate mimicry in the center of a mammalian α-amylase: structural analysis of an enzyme−inhibitor complex. *Structure* **1996**, *4*, 1441−1452.

(20) Bode, W.; Huber, R. Natural protein proteinase inhibitors and their interaction with proteinases. *Eur. J. Biochem.* **1992**, *204*, 433−451.

(21) Young, N. M.; Oomen, R. P. Analysis of sequence variation among legume lectins: a ring of hypervariable residues forms the perimeter of the carbohydrate-binding site. *J. Mol. Biol.* **1992**, *228*, 924−934.

(22) Wilcox, E. R.; Whitaker, J. R. Structural features of red kidney bean α-amylase inhibitor important in binding with α-amylase. *J. Food Biochem.* **1984**, *8*, 189−213.

(23) Mirkov, T. E.; Evans, S. V.; Wahlstrom, J.; Gomez, L.; Young, N. M.; Chrispeels, M. J. Location of the active site of the bean α-amylase inhibitor and involvement of a Trp, Arg, Tyr, triad. *Glycobiology* **1995**, *5*, 45−50.

(24) Yin, X. α-Amylase inhibitors of black kidney bean (*Phaseolus vulgaris*): subunit composition and other molecular properties. Ph.D. Dissertation, University of California, Davis, CA, 1990; p 152.

(25) Moreno, J.; Chrispeels, M. J. A lectin gene encodes the α-amylase inhibitor of the common bean. *Proc. Natl. Acad. Sci. U.S.A.* **1989**, *86*, 7885−7889.

(26) Afanador, L. K.; Haley, S. D.; Kelly, J. D. Adoption of a "mini-prep" DNA extraction method for RAPD marker analysis in common bean (*Phaseolus vulgaris* L.). *Annu. Rep. Bean Improv. Coop.* **1993**, *36*, 10−11.

(27) Devereux, J.; Haeberli, P.; Smithies, O. A comprehensive set of sequence analysis programs for the VAX. *Nucleic Acids Res.* **1984**, *12*, 387−395.

(28) Dessen, A.; Gupta, D.; Sabesan, S.; Brewer, C. F. X-ray crystal structure of the soybean agglutinin cross-linked with a biantennary analog of the blood group I carbohydrate antigen. *Biochemistry* **1995**, *34*, 4933−4942.

(29) Bourne, Y.; Cambillau, C. X-ray structure of a biantennary octasaccharide-lectin complex at 2.3 Å resolution. *J. Biol. Chem.* **1992**, *267*, 197−203.

(30) Bourne, Y.; Mazurier, J. J.; Legrand, D.; Rouge, P.; Montreuil, J.; Spik, G.; Cambrilau, C. Interaction of a legume lectin with the N2 fragment of human lactotransferrin or with the isolated biantennary glycopeptide: role of the 4 fucose moiety. *Structure* **1994**, *2*, 209−219.

(31) Rini, J. M.; Hardman, K. D.; Einspahr, H.; Suddath, F. L.; Carver, J. P. X-ray crystal structure of a pea lectin-trimannoside complex at 2.6 Å resolution. *J. Biol. Chem.* **1993**, *268*, 10126−10136.

(32) Loris, R.; Casset, F.; Poortmans, F.; Imberty, A.; Perez, S.; Wyns, L. The monosaccharide binding site of lentil lectin: an X-ray and molecular modelling study. *Glycoconjugate J.* **1994**, *11*, 507−517.

(33) Bernfeld, P. Amylase, α and *β*. *Methods Enzymol.* **1955**, *1*, 149−158.

(34) Strumeyer, D. H. A modified starch for use in amylase assays. *Anal. Biochem.* **1967**, *19*, 61−71.

(35) Hewick, R. M.; Hunkapillar, M. W.; Hood, L. E.; Dreyer, W. J. A gas−liquid solid phase peptide and protein sequentor. *J. Biol. Chem.* **1981**, *256*, 7990−7997.

(36) Dayal, U.; Mehta, S. K. Studies on molecular weight determination using refractive index and multi-angle laser light scattering detectors in size exclusion chromatography. *J. Liq. Chromatogr.* **1994**, *17*, 303−316.

(37) Hutchens, W. T.; Yip, T. T. Model protein surface domains for the investigation of metal ion-dependent macromolecular interactions and biospecific metal ion transfer. *Methods: Companion Methods Enzymol.* **1992**, *4*, 79−96.

(38) Osborn, T. C.; Alexander, D. C.; Sun, S. S. M.; Cardona, C.; Bliss, F. A. Insecticidal activity and lectin homology of arcelin seed protein. *Science* **1988**, *240*, 207−210.

(39) Hoffman, L. M.; Donaldson, D. D. Characterization of two *Phaseolus vulgaris* phytohemagglutinin genes closely linked on the chromosome. *EMBO J.* **1985**, *4*, 883−889.

(40) Higgins, T. J. V.; Chandler, P. M.; Zurawski, G.; Button, S. C.; Spencer, D. The biosynthesis and primary structure of pea lectin. *J. Biol. Chem.* **1983**, *258*, 9544−9549.

(41) Scott, M. P.; Jung, R.; Muntz, K.; Nielsen, N. C. A protease responsible for post-translation cleavage of conserved Asn-Gly linkage in glycinin, the major seed storage protein of soybean. *Proc. Natl. Acad. Sci. U.S.A.* **1992**, *89*, 658−662.

(42) Muramatsu, M.; Fukazawa, C. A high-order structure of plant storage proprotein allows its second conversion by an asparagine-specific cysteine protease, a novel proteolytic enzyme. *Eur. J. Biochem.* **1993**, *215*, 123−132.

(43) McClelland, J. L.; Rumelhart, D. E. *Explorations in Parallel Distributed Processing*; MIT Press: Cambridge, MA, 1988; Vol. 3, pp 318−362.

(44) Kneller, D. G.; Cohen, F. E.; Langridge, R. Improvements in protein secondary structure prediction by an enhanced neural network. *J. Mol. Biol.* **1990**, *214*, 171−182.

(45) Ishimoto, M.; Kitamura, K. Effect of absence of seed α-amylase inhibitor on the growth inhibitory activity to Azuki bean weevil (*Callosobruchus chinensis*) in common bean (*Phaseolus vulgaris* L). *Jpn. J. Breed.* **1991**, *41*, 231−240.

(46) Petisch, M. C. Protein modeling by E-mail. From amino acid sequence to protein structure: A free one-hour service. *Bio/Technology* **1995**, *13*, 658−660.

(47) Higaki, H.; Yamaguchi, H. Reconstitution of *Phaseolus vulgaris* α-amylase inhibitor from isolated subunits. *Biosci., Biotechnol., Biochem.* **1994**, *58*, 5−8.

(48) Lee, S. C. Comparative amino acid sequences, and deduced tertiary and quaternary structures of white kidney bean and black bean α-amylase inhibitors. Ph.D. Dissertation, University of California, Davis, CA, 1996; p 203.

(49) Whitaker, J. R. Determination of molecular weights of proteins by gel filtration on Sephadex. *Anal. Chem.* **1963**, *35*, 1950−1953.

(50) Hedrick, J. L.; Smith, A. J. Size and charge isomer separation and estimation of molecular weights of proteins by disc gel electrophoresis. *Arch. Biochem. Biophys*. **1968**, *126*, 155−164.

(51) Weber, K.; Osborn, M. Proteins and sodium dodecyl sulfate: molecular weight determination of polyacrylamide gel and related procedures. In *The Proteins*, 3rd ed.; Neurath, H., Hill, R., Boeder, C., Eds.; Academic Press: New York, 1975; Vol. 1, pp 179−223.

Received for review February 12, 2002. Revised manuscript received March 5, 2002. Accepted March 11, 2002.

JF020189T