Primary Structure of the Abundant Seed Albumin of *Theobroma cacao* by Mass Spectrometry

Sunil Kochhar,* Karin Gartenmann, and Marcel A. Juillerat

Nestlé Research Center, P.O. Box 44, Vers-chez-les-Blanc, CH-1000 Lausanne 26, Switzerland

The most abundant albumin present in seeds of *Theobroma cacao* was purified to apparent homogeneity as judged by high-performance liquid chromatography/electrospray ionization mass spectrometry (HPLC/ESI-MS), sodium dodecyl sulfate−polyacrylamide gel electrophoresis, and NH_2-terminal sequence analysis. Tryptic peptide mass fingerprinting of the purified protein by HPLC/ESI-MS showed the presence of 16 masses that matched the expected tryptic peptides corresponding to 95% of the translated amino acid sequence from the cDNA of the 21 kDa cocoa albumin. Collision-induced dissociation MS/MS analysis of the C-terminal peptide isolated from the CNBr cleavage products provided unequivocal evidence that the mature cocoa albumin protein is nine amino acid residues shorter than expected from the reported cDNA of its corresponding gene. The experimentally determined M_r value of 20234 was in excellent agreement with the truncated version of the amino acid sequence. The purified cocoa albumin inhibited the catalytic activities of bovine trypsin and chymotrypsin. The inhibition was stoichiometric with 1 mol of trypsin or chymotrypsin being inhibited by 1 mol of inhibitor with apparent dissociation constants (K_i) of 9.5×10^{-8} and 2.3×10^{-6} M, respectively, for inhibitor binding at pH 8.5 and 37 °C. No inhibition of the catalytic activities of subtilisin, papain, pepsin, and cocoa endoproteases was detected under their optimal reaction conditions.

Keywords: *Albumin; HPLC/MS; peptides; processing sites; protease inhibitor; storage proteins; Theobroma cocoa; trypsin; vicillin*

INTRODUCTION

Proteins constitute 10−15% dry weight of cocoa seeds, the second most abundant constituent after cocoa fat. Voigt et al. (1993) suggested that the total cocoa seed protein content is composed of 52 and 43% of albumin and globulin fractions, respectively. Polypeptides of apparent molecular weights of 14.5, 31, and 47 kDa were globulin type, whereas a 21 kDa protein was identified as the albumin. More recently, Lerceteau et al. (1999), employing two-dimensional SDS−PAGE in combination with silver staining, determined that globulin and albumin storage protein represent 23 and 14%, respectively, of the total seed protein and indicated the presence of other unknown abundant polypeptides. The genes encoding the vicillin and albumin cocoa storage proteins have been cloned and sequenced (Spencer and Hodge, 1991, 1992; Dodo and Furtek, 1994). The 31 and 47 kDa proteins were suggested to be post-translational cleavage products of a single gene product, a 67 kDa storage protein precursor. The expression of the recombinant 67 kDa storage protein precursor was reported to be highly inefficient (O'Connor et al., 1996). Attempts to express and secrete the "water soluble" 21 kDa albumin in the yeasts *Hansenula polymorpha* and *Saccharomyces cerevisiae* were not successful (Yavuz et al., 1996). The recombinant proteins were neither isolated nor characterized.

Despite detailed analysis of the overall profile of cocoa seed proteins in unfermented as well as fermented beans, the biochemical characterization of mature proteins remains sketchy. In addition, cocoa albumin has been suggested to be a trypsin inhibitor (Dodo et al., 1992). However, no systematic kinetic analysis has been reported to assess the potency of its proposed inhibitory function. In this paper, we present characterization of the primary structure of the mature albumin storage protein and its trypsin inhibitory properties. The mature protein is nine amino acid residues shorter than predicted from its cDNA, most likely due to post-translation modification at its C terminus.

EXPERIMENTAL PROCEDURES

Materials. HPLC grade acetonitrile, methanol, and water were from Merck. Ethylenediaminetetracetic acid sodium salt (EDTA), tripotassium phosphate monohydrate, 1,4-dithio-DL-threitol (DTT), trifluoroacetic acid (TFA), and guanidinium hydrochloride (GndHCl) were from Fluka. 4-Vinylpyridine was from Aldrich. Bovine trypsin and chymotrypsin, papain, pepsin, and subtilisin were obtained from Sigma. All other chemicals used were of analytical grade. Cocoa pods were obtained from experimental farms in Ecuador, Ivory Coast, and Malaysia. Unless stated otherwise, all studies were carried out using West African Amelonado cocoa beans.

Preparation of Cocoa Acetone Powder. Sun-dried unfermented cocoa beans were passed through a bean crusher (Brook Crampon) followed by a winnower (John Gordon) to remove shells. The nibs were kept in a brown bottle at −20 °C. Cocoa nibs (20−30 g) were milled for a few seconds in a universal mill (IKA M-20). The nib powder was passed through an 0.8-mm sieve and kept at 4 °C. Cocoa acetone powder (CAP) was prepared by suspending (10 g) cocoa nib powder in 200 mL of 80% (v/v) aqueous acetone and stirring for 1 h at 4 °C. The suspension was centrifuged at 15000 rpm for 15 min at 4 °C. The residue was extracted five times with 200 mL of 80% (v/v) aqueous acetone followed by three times with 100%

* Corresponding author (telephone +41 21 785 9336; fax +41 21 785 8925; e-mail sunil.kochhar@rdls.nestle.com).

10.1021/jf0006067 CCC: $19.00
Published on Web 09/28/2000

acetone. The residue was dried under reduced pressure and stored at −20 °C.

Protein Assay. The protein concentration during purification of the albumin protease inhibitor was determined according to the dye binding assay of Bradford (1976) using bovine serum albumin (fraction V) as a standard (kit from Bio-Rad). The concentration of the purified protein was determined spectrophotometrically using a molar extinction coefficient (ϵ) of 40000 M^{-1} cm^{-1} at 280 nm calculated from an M_r of 20234.

Protease Inhibition Assay. The inhibition of trypsin, lysyl endoprotease, and papain was followed spectrophotometrically by monitoring the rate of hydrolysis of Ac-L-Arg-4-NA (Bachem) in triethanolamine hydrochloride buffer (0.2 M, pH 7.8) containing 2 mM $CaCl_2$ at 37 °C (Geiger and Fritz, 1984). Inhibition of bovine α-chymotrypsin and subtilisin BPN′ was assayed using Ac-L-TyrOEt (Bachem) as substrate (Geiger, 1984). One unit of inhibition activity was defined as the amount of inhibitor that completely inhibited 1 mg of trypsin. Cocoa aspartic endoproteases were assayed using FITC-labeled hemoglobin (De Lumen and Tappel, 1970). The kinetic constants k_{cat}, K_m, and K_i were calculated using the data analysis and graphic software GraFit (ver.4.0.12) from Erithacus Software Ltd.

Purification of Cocoa Albumin Protease Inhibitor. CAP (5 g) was suspended in 50 mL of ice-cold sodium acetate buffer (50 mM, pH 4.0) containing 0.1 mM pepstatin. The suspension was sonicated for 2 × 30 s with a 10 min layover interval on ice. The suspension was centrifuged at 20000*g* for 15 min at 4 °C, and the collected supernatant was subsequently passed through a sterile 0.22-μm filter and stored at −20 °C.

Chromatographic steps were performed at room temperature using a BioCad 20 chromatography station (Perseptive Biosystems). The frozen CAP extract was thawed overnight at 4 °C and adjusted to pH 8 with 1 M Tris solution. The clear CAP extract was applied to a Source 15Q column (16 × 100 mm) equilibrated with buffer A (50 mM Tris-bis-propane chloride, pH 8.0) at a flow rate of 2 mL/min. The column was washed with buffer A until A_{280} decreased to <0.05. The column was eluted with a linear gradient (20 column volumes) of 0−500 mM NaCl in buffer A. Fractions of 5 mL each were collected. Fractions showing trypsin-inhibiting activity were pooled, concentrated by ultrafiltration (PM-10 membrane, Amicon), and injected onto a HiLoad Superdex 75 column (26 × 600 mm) equilibrated with 50 mM sodium phosphate buffer, pH 7, containing 100 mM NaCl. The column was eluted with the same buffer at a flow rate of 2 mL/min, and fractions of 2 mL each were collected. The fractions showing trypsin-inhibiting activity were pooled and concentrated to 2.5 mL by ultrafiltration. The purified protein solution was passed through a fast-desalting PD-10 column (Amersham-Pharmacia Biotech) for buffer exchange to 50 mM ammonium acetate, pH 7.0, sterile filtered, and stored at −20 °C.

PAGE Analysis. Denaturating tricine-SDS−PAGE was performed using a Bio-Rad mini Protean electrophoresis system according to the method of Schägger and von Jagow (1987). Glycosylation of albumin was assessed by employing a glycoprotein detection kit from Bio-Rad.

***S*-Pyridinylethylation.** The purified albumin (10−50 nmol) was dried under reduced pressure. The dried pellet was dissolved in 300 μL of denaturating buffer (100 mM ammonium phosphate, 66.7 mM potassium hydroxide, 3 mM EDTA, and 6 M GndHCl), sparged with argon, and mixed with 10 μL of reducing solution (0.8 M DTT in 3 M tripotassium phosphate/3 mM EDTA). The solution was kept at room temperature in the dark for 60 min. Pyridinylethylation at cysteine residues was carried out by mixing vigorously 10 μL of 4-vinylpyridine and further incubation for 30 min at room temperature (Lundell and Schreitmüller, 1999). The reaction mixture was desalted by employing an NAP5 column (Amersham Pharmacia Biotech) pre-equilibrated and eluted with 0.1% (v/v) TFA in water. The clear solution was passed through a 0.22-μm filter disk and kept at 4 °C until analyzed.

Generation of Tryptic Peptides. The reduced and pyridinylethylated albumin was dried under reduced pressure and dissolved in 100 μL of 50 mM Tris-HCl buffer, pH 8.5, containing 2 mM $CaCl_2$. Digestions with trypsin (TPCK treated, Promega) were carried out at 25 °C for overnight at an enzyme substrate ratio of 1:20 (w/w) and stopped by the addition of TFA to a final concentration of 1% (v/v).

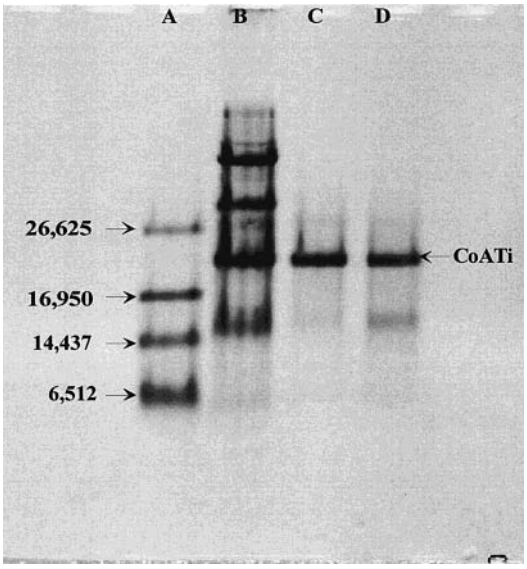

Figure 1. SDS−PAGE showing purification of CoATi: (lane A) molecular weight markers; (lane B) crude preparation of CoATi (30 μg); (lane C) Superdex 75 pool (10 μg); (lane D) Source 15Q pool (10 μg). A 10 μL sample was analyzed on ready gradient 10−20% T Tris-tricine acrylamide gels (Bio-Rad) according to the method of Schägger and von Jagow (1987). Protein bands were visualized by staining the gel in 0.5% (w/v) Coomassie Brilliant Blue G 250 solution prepared in 10% (v/v) acetic acid for 1 h followed by destaining in 10% (v/v) acetic acid solution.

CNBr Cleavage Reaction. Purified albumin (20−50 nmol), native or pyridinylethylated, was dissolved in 70% formic acid and reacted with CNBr (stock solution, 0.5 g/mL in acetonitrile) in the dark for 12−16 h. The reaction mixture was diluted with an equal volume of 0.1% TFA in water and dried under reduced pressure. The dried CNBr cleavage products were dissolved in 0.1% (v/v) TFA in water and analyzed by RP-HPLC/ESI-MS. For more details on CNBr cleavage, see Kaiser and Metzka (1999).

Amino Acid Sequencing. Purified albumin was subjected to N-terminal amino acid sequencing by Edman degradation employing a gas phase sequencer (Procise 494, Perkin-Elmyer) using standard cycles and standard methods.

HPLC/ESI-MS Analysis. Mass measurements and peptide fragmentation were made using a FinniganMat LCQ mass spectrometer interfaced with a Spectra HPLC system (FinniganMat) described in detail elsewhere (Gartenmann and Kochhar, 1999).

The LC/MS analysis of the intact and pyridinylethylated albumin was carried out using an RP C_8 column [Spherisorb 80−5 C_8 (5 μm, 2 × 125 mm), Macherey-Nagel] with a linear increase of solvent B (0.05% TFA/80% ACN v/v in water) in solvent A [0.045% TFA (v/v) in water] as follows: 0−35% B in 5 min, isocratic at 35% B for 5 min, 35−60% B in 70 min, 60−100% B in 10 min, and finally isocratic at 100% B for 5 min. The flow rate was 0.2 mL/min, and detection was at 215 nm. Tryptic peptides of reduced and pyridinylethylated albumin were analyzed using an RP C_{18} column [Nucleosil 100-3 C18 HD (3 μm, 2 × 150 mm), Macherey-Nagel] with a linear increase of solvent B [0.05% TFA/80% ACN (v/v) in water] in solvent A [0.045% TFA (v/v) in water] as follows: 0−60% B in 60 min, 60−100% B in 20 min, followed by an isocratic elution at 100% B for 5 min. The flow rate was 0.2 mL/min, and detection was at 215 nm.

Database Analysis. The theoretical tryptic mass fingerprint analysis and the comparisons of peptide masses against

Table 1. Purification of CoATi

step	total protein (mg)	total act. (units)	specific act. (units mg^{-1})	yield (%)
crude extract	150	1000	6.7	100
Source 15Q	40	700	17.5	70
Superdex 75	35	660	18.9	66

Swiss-Prot and TrEMBL databases were carried out using software tools PeptideMass and PeptIdent available at the Internet site ExPASy (www.expasy.ch) at the Swiss Institute of Bioinformatics, Geneva, Switzerland.

RESULTS

Purification of Cocoa Albumin. Preliminary extraction studies of the cocoa acetone powder at different pH values indicated that the albumin fraction can be selectively solublized at pH 4. SDS−PAGE analysis of pH 4 extract showed a high-intensity 20−22 kDa protein band representing >80% of the total extracted protein (data not shown). Two successive chromatography steps, anion exchange and gel filtration, resulted in apparent homogeneity of the protein preparation as judged by SDS−PAGE followed by Coomassie Brilliant Blue staining (Figure 1) and NH_2-terminal sequencing. Following identification of the trypsin inhibition activity in the purified preparations of the cocoa seed albumin, the total recovery of the albumin was calculated on the basis of the trypsin inhibition units. The purification procedure resulted in 66% recovery of the trypsin inhibitory activity (Table 1). The purified cocoa albumin trypsin inhibitor (CoATi) was not found to be glycosylated.

LC/ESI-MS analysis of CoATi showed the molecular weight of the mature protein to be 20234 Da (Figure 2). In addition, the molecular weights of the purified cocoa albumin from seven different genotypes (see Table 2) representing the four cocoa varieties, namely, Criollo, Forastero, Nacional, and Trinitario, were determined by RP-HPLC/ESI-MS. In all of the cases, the M_r values of both the major and minor polypeptides were identical (Table 2). Reduction and *S*-pyridinylethylation of the purified CoATi preparations (Table 2) resulted in a positive shift of 735 mass units (M_r 20970) indicating the presence of seven cysteine residues. The NH_2-terminal sequence of the first 29 residues, ANSPVLDTDGDELQTGVQYYVLSSISGAG, showed no heterogeneity. The repetitive yield of the Edman cycles was >90%.

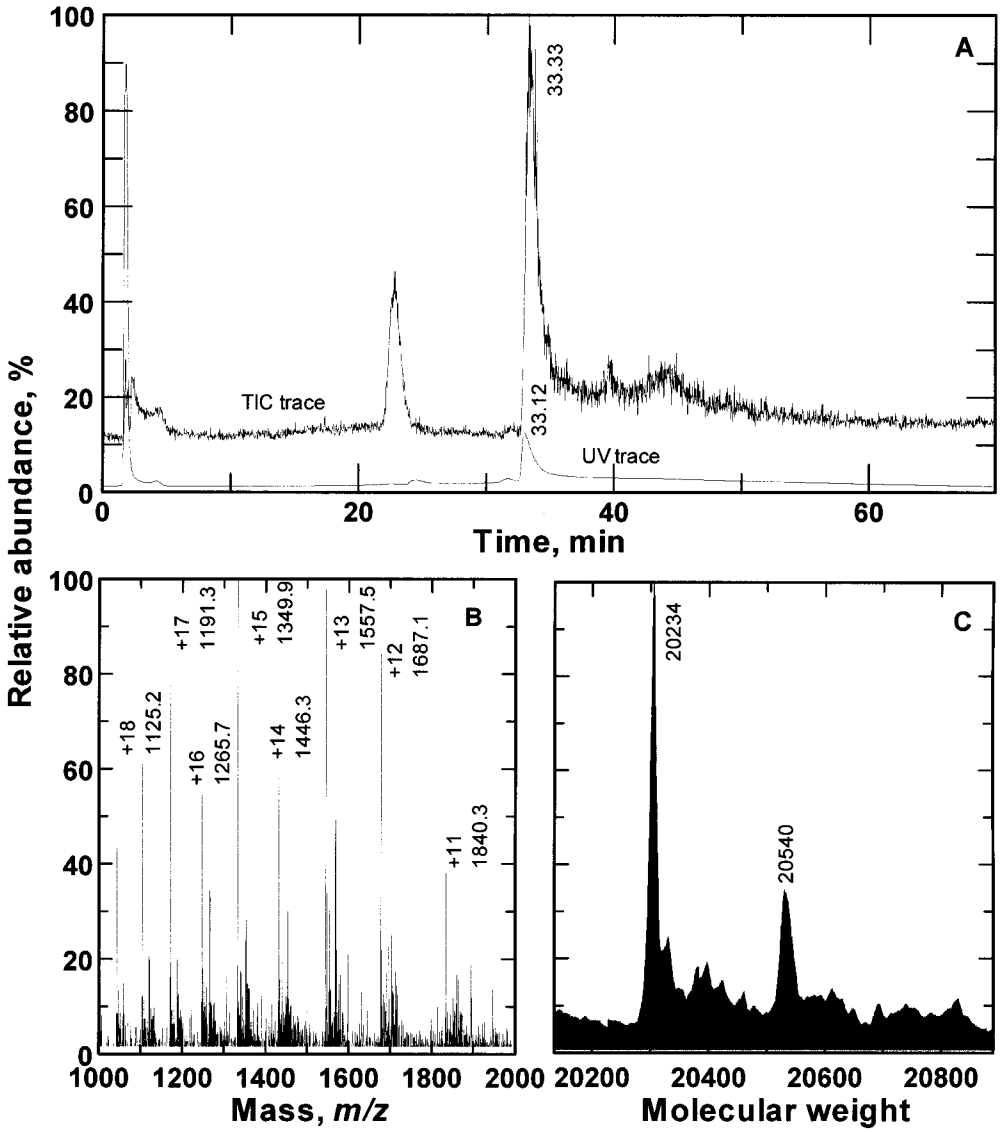

Figure 2. RP-HPLC/ESI-MS analysis of the purified CoATi: (A) UV and TIC trace chromatograms; (B) mass spectrum showing multiple charged ion signals (raw data from an average of 50 scans); (C) molecular weight deconvolution of the spectrum shown in B.

Table 2. Molecular Weight Determination of Purified CoATi by LC/ESI-MS

	obsd av M_r[a]			obsd av M_r[a]	
genotype	major	minor	genotype	major	minor
NA33	20234	20539	EET95	20233	20522
IMC67	20234	20523	Criollo	20264	20523
ICS95	20233	20522	AML	20234	20537
EET103	20233	20524			

[a] The average M_r was calculated by deconvoluting the ESI mass spectrum of multiply charged ions. For the origins of specific cocoa genotypes, see Hansen et al. (2000).

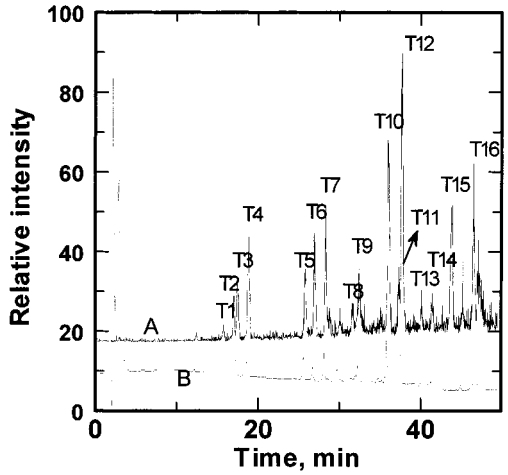

Figure 3. Tryptic peptide mass fingerprint of CoATi. The RP-HPLC/ESI-MS conditions are described under Experimental Procedures. A, TIC trace; B, UV trace.

Table 3. CoATi Tryptic Peptide Analysis by LC/ESI-MS

theor av $[M + H]^+$	sequence position	tryptic peptides[a]	av obsd $[M + H]^+$
3637.80	27−63	T16	1819.7[b]
1906.78	168−185	T10	1164.2[b,c]
1779.84	78−94	T11	1780.7
1610.77	195−207	T15	1610.7
1488.81	100−112	T14	1488.7
1435.67	143−155	T13	1540.6[c]
1387.70	64−76	T7	1492.7[c]
1096.49	124−133	T6	1096.4
1091.55	134−142	T12	1091.5
1052.52	115−123	T9	1157.5[c]
1042.46	186−194	T5	1042.4
707.41	159−165	T8	707.4
658.39	215−220	ND[d]	
603.31	95−99	T4	603.3
389.24	156−158	T2	389.3
361.25	212−214	ND	
322.19	166−167	T3	322.2
305.18	209−211	T1	305.2
290.15	113−114	ND	

[a] Figures 3 and 4. [b] $[M + H]^{2+}$. [c] Pyridinylethyl modifications at cysteine adds a positive shift of 105 mass units . [d] ND, not detected.

Peptide Mass Fingerprinting of CoATi. The primary structure of the mature protein was determined by generating the tryptic peptide mass fingerprints of the reduced and pyridinylethylated CoATi by RP-HPLC/ESI-MS. Figure 3 shows the UV elution and total ion current chromatograms of all the tryptic peptide fragments. The tryptic peptide fragments are designated T and are sequentially numbered on their order of elution during RP-HPLC (Figure 3). A total of 16 peptide masses were detected (Figure 3 and Table 3). Most of the eluting peptides were detected as singly charged $[M + H]^+$ species except the N-terminal peptide, T16 and

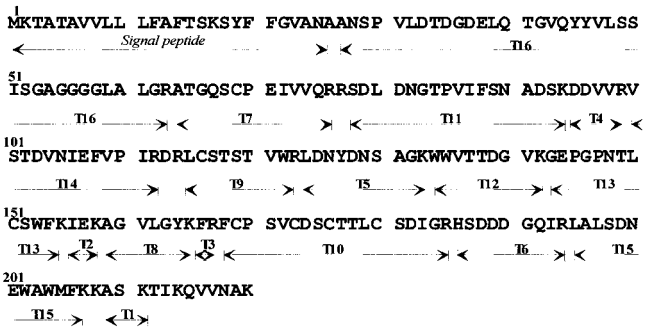

Figure 4. Mapping of the tryptic peptides of CoATi to the deduced amino acid sequence of 21 kDa albumin cocoa protein (Swiss-Prot database Accession no. P32765). The alignment of the observed peptides (see Figure 3) was based on the theoretical tryptic fragment masses generated by the software PeptideMass.

the T10 (expected monoisotopic masses 3637.8 and 2327.0, respectively), which were present as doubly charged ions at *m/z* 1819.7 and 1164.2, respectively (Table 3). A comparison of the observed peptide masses against the sequence databases using the software PeptidIdent showed a 95% amino acid sequence match with the translated sequence of the 21 kDa cocoa albumin cDNA (Figure 4). The peptide fragments containing the cysteine residues showed the expected positive mass shift of 105 due to *S*-pyridinylethylation (Table 3). Every identified peptide mass was subjected to MS/MS (data not shown) analysis to determine either a complete or partial amino acid sequence to confirm its mapping to the amino acid sequence of the albumin. Repeated attempts to isolate and identify peptides corresponding to the C-terminal region (180−195) were not successful. Rigorous controls, for example, analysis of tryptic peptide map generated at different substrate/enzyme ratios, RP-HPLC of the tryptic digests employing C_8 and C_{18} columns from different suppliers (Vydac, Macherey-Nagel, and Phenomenex), and extensive analysis of peptide mass fingerprint data with predictive mass peak extraction ruled out the possibility that these peptides selectively remained undetected due to either an incomplete trypsin digestion or a chromatographic elution abnormality.

Identification CNBr Fragments of CoATi. HPLC/ESI-MS analysis of the CNBr fragments showed the presence of two molecular ion species: a doubly charged ion at *m/z* 405.4 representing a peptide eluting in the run through fraction (M_r 808.4; monoisotopic ion at *m/z* 809.4 and its sodium adduct at *m/z* 831.69) and a typical multiple charged ion of high molecular weight polypeptide corresponding to average an M_r of 19450 (Figure 5). Analysis of the fragmentation pattern of the parent peptide (*m/z* 809.4) by MS/MS showed predominance of doubly charged *b* ions (b_4-b_6), *m/z* 475.5, 562.5, and 662.2, consistent with sequential loss of threonine, lysine, serine, and alanine, confirming the peptide sequence of FKKASKT (Figure 5).

Inhibition Specificity. Kinetic analysis of the inhibitory activity of CoATi showed it to be a potent inhibitor of trypsin activity with an apparent K_i of 9.5×10^{-8} (Figure 6) at pH 8.5 and 37 °C and showed a typical pattern of competitive inhibition. The inhibition was stoichiometric, with 1 mol of trypsin being inhibited by 1 mol of CoATi. The inhibition of bovine α-chymotrypsin activity was relatively less strong and its K_i was at least 2 orders of magnitude higher (2.3×10^{-6})

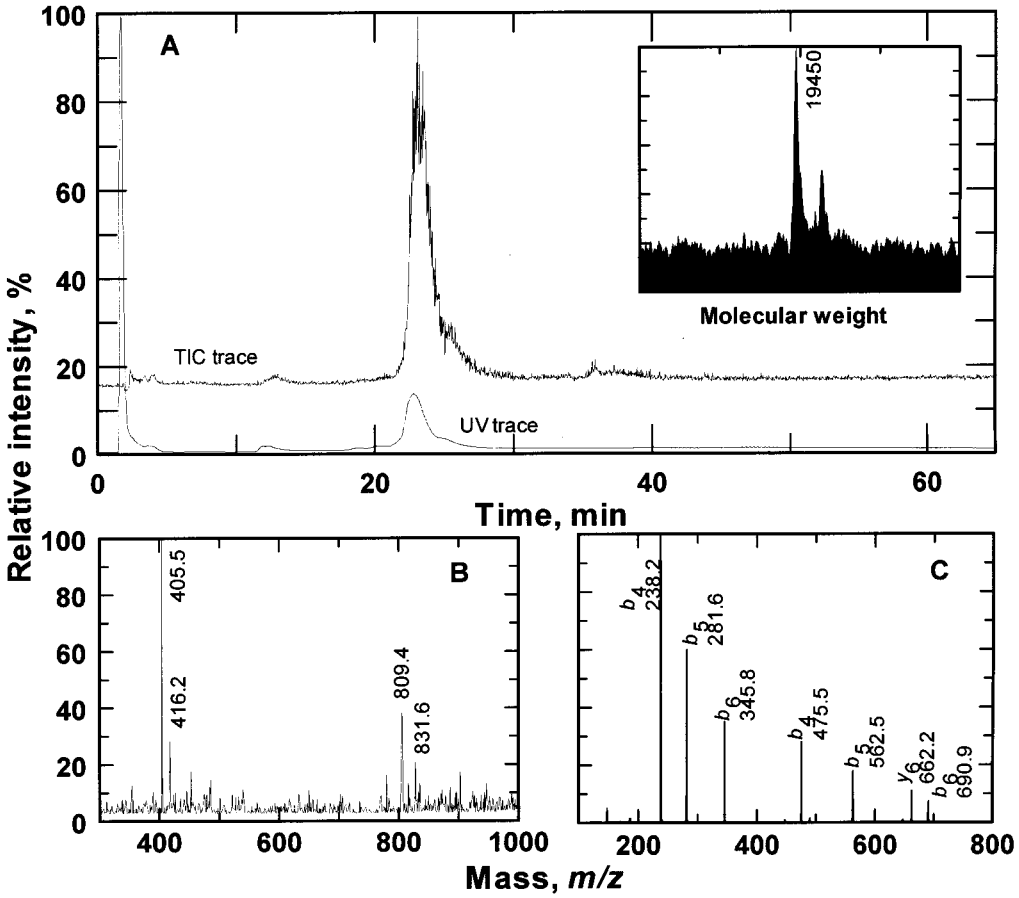

Figure 5. CNBr peptide mass fingerprint of CoATi: (A) UV and TIC chromatogram, (inset) deconvoluted mass spectrum of the peak at 23.8 min (raw data of an average of 50 scans); (B) mass spectrum of the run through fraction peptide (raw data of an average of 20 scans); (C) product ion mass spectrum of the CNBr peptide *m/z* 809.4. The LC/ESI-MS conditions are described under Experimental Procedures.

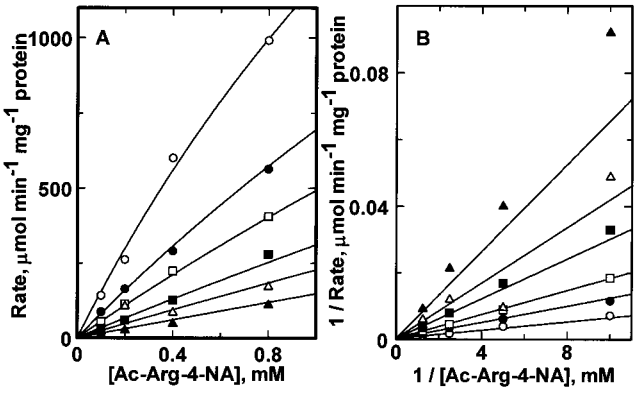

Figure 6. Kinetic of trypsin inhibition by CoATi: (A) inhibition of initial reaction rates; (B) Lineweaver−Burk plot; (○) 0, (●) 0.09, (□) 0.18, (■) 0.36, (△) 0.54, and (▲) 0.9 μM CoATi. The initial rates of trypsin activity were determined at different concentrations of Ac-Arg-4-NA. The experimental error limits were between ±2 and 5%. The assay conditions are described under Experimental Procedures.

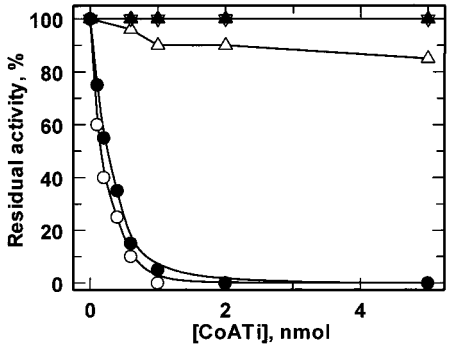

Figure 7. Inhibition of endoproteolytic activities by CoATi using FITC-hemoglobin as substrate. The reaction mixture contained 20 μL of enzyme, 200 μL of assay buffer, 200 μL of 1% FITC-hemoglobin, and 20 μL of the indicated amount of CoATi in a 2 mL capacity Eppendorf tube. The assay buffer for trypsin (○), chymotrypsin (●), and lysyl endoprotease (△) was 100 mM triethanolamine hydrochloride, pH 7.8, containing 2 mM $CaCl_2$. The assay buffer for pepsin (▲) and cocoa aspartic proteases (▽) was 100 mM sodium acetate, pH 4.0. Following incubation at 37 °C for 15 min, the reaction was stopped by adding 500 μL of 10% (w/v) trichloroacetic acid. The mixture was allowed to stand at room temperature for 15 min and centrifuged (10000*g* for 10 min), and the pH of the clear supernatant was adjusted to 10 by adding 20 μL of 50% (w/v) NaOH. The relative absorbance of the solution was read at 490 nm.

compared to the K_i for trypsin inhibition. Bacterial serine proteases, for example, subtilisin and lysyl endoproteases, were not inhibited. CoATi was also inactive against aspartic protease, pepsin,and cocoa endoproteases. In addition, inhibition by CoATi was also studied using FITC-labeled hemoglobin or casein as substrates. The data were consistent with those obtained with the synthetic substrates; only trypsin and chymotrypsin activities could be completely inhibited (Figure 7).

DISCUSSION

The purification of CoATi by differential extraction and two-step column chromatographies yielded an ap-

parently homogeneous preparation as judged by SDS−PAGE and NH_2-terminal sequencing. The purification resulted in only a 3-fold increase in the specific inhibition activity, indicating that CoATi is one of the most abundant proteins in the ungerminated/unfermented cocoa bean. The purified CoATi preparation contained two polypeptides of M_r 20234 (>90%) and M_r 20540 (<10%). Attempts to resolve these polypeptides by further chromatography on RP/HPLC or by capillary electrophoresis were not successful. N-Terminal amino acid sequencing did not indicate any heterogeneity, suggesting the M_r 20540 polypeptide either is N-terminally blocked or is a minor natural variant with an identical N-terminal sequence. Nevertheless, M_r analysis of CoATi preparations from eight different genotypes representing the four cocoa varieties, namely, Criollo, Forastero, Nacional, and Trinitario, also showed the presence of both the major and minor polypeptides distributed in similar proportions. The data taken together clearly suggest that the polypeptide of M_r 20540 is a natural variant of the cocoa albumin.

The M_r of the mature CoATi is considerably smaller than that predicted from the published cDNA sequences of the gene encoding the 21 kDa albumin protein: M_r 21259 (Tai et al., 1991; Dodo and Furtek, 1994) or M_r 21223 (Spencer and Hodge, 1991). Rigorous peptide mass fingerprint data strongly indicate that the amino acid sequence at the C terminus is drastically different from that reported from the cDNA sequence. Spencer and Hodge (1991) reported the deduced C-terminal peptide to be TIKQVVNAND, whereas Tai et al. (1991) and Dodo and Furtek (1994) reported it to be TIKQVVANAK. It is highly unlikely that these differences are due the different genotypes of *Theobroma cacao* as the molecular weight of CoATi from eight different clones representing the main varieties of cocoa were strikingly identical. In addition, the fact that the observed masses of the tryptic fragments matched 100% with the expected masses calculated from the deduced sequence between residues 1 and 185 further establishes a high degree of conservation of this protein. Further insight into the structure of the C terminus of the CoATi was provided by CNBr fragmentation. Of the two expected polypeptides, only one matched the expected M_r of 19450, showing typical multicharged peaks expected from larger polypeptides. However, the other peptide was considerably smaller in size, M_r 809.3, compared to the expected mass of 1827. MS/MS fragmentation data of the peptide confirmed its amino acid sequence to be FKKASKT, showing that the mature polypeptide is considerably shorter due to post-translational modifications. The mature seed storage proteins of plants generally undergo post-translational processing to generate multiple subunits, and in some cases, one or more amino acids are also trimmed off at the C terminus of the mature proteins (Ericson et al., 1986; Byczynska and Barciszewski, 1999). In the case of the cocoa 67 kDa vicillin storage protein, it is suspected that post-translational processing produces the 31 and 47 kDa polypeptides (Spencer and Hodge, 1991). However, we suspect there are at least three post-translation cleavage sites for the vicillin storage proteins producing at least four polypeptides (unpublished data). The physiological significance of the truncation of the mature albumin at the C terminus and, as a matter of fact, for many seed storage proteins is poorly understood.

The physiological role of the cocoa albumin is an open question. On the basis of primary amino acid sequence comparisons, CoATi has been assigned in the Kunitz soybean trypsin inhibitor superfamily (Tai et al., 1991; Spencer and Hodge, 1991) that represents a rare example of extensive functional diversity as it covers protease inhibitors, glycoside inhibitors, taste-modifying proteins, storage proteins, and wound-inducible proteins [see Odani et al. (1996) and references therein]. Cocoa albumin shows the highest amino acid identity (56%) with the sweet protein miraculin (Masuda et al., 1995) and, being rich in cysteine and being the most abundant seed protein, it is tempting to assume that the albumin serves as the sulfur reserves for the seed. On the other hand, an apparent dissociation constant K_i of 9.5×10^{-8} M for trypsin suggests that CoATi is a potent trypsin inhibitor. Failure to inhibit the activities of papain, pepsin, or subtilisin indicates high specificity for trypsin and trypsin-like activities. On the basis of inhibition data one can reasonably assume that protease inhibition might be the primary function of cocoa albumin to protect the seed against pests and parasites. The protein protease inhibitors in many plants have been related to pest resistance, and their presence is considered to be a desirable trait to increase plant resistance against pathogens [for a review, see Ryan (1973a,b, 1981)]. Cocoa is a valuable commercial crop for the production of chocolate, cocoa powder, and cocoa butter. Identification of endogenous insecticides is of paramount importance to develop insect/pest resistant varieties of cocoa. It is tempting to speculate that to develop natural pesticides for cocoa, the cocoa albumin may be an ideal candidate. Nevertheless, unequivocal evidence for insecticidal properties needs confirmation of in vivo and in vitro inhibition studies of proteases from diverse species of insects and pests.

ABBREVIATIONS USED

ACN, acetonitrile; Ac-Arg-4-NA, N^{α}-acetyl-L-arginine 4-nitroanilide; Ac-TyrOEt, N^{α}-acetyl-L-tyrosine ethyl ether; CAP, cocoa acetone powder; CoATi, cocoa albumin trypsin inhibitor; CNBr, cyanogen bromide; DTT, 1,4-dithio-DL-threitol; EDTA, ethylenediaminetetracetic acid; ESI-MS, electrospray ionization mass spectrometry; GndHCl, guanidinium hydrochloride; LC/MS, liquid chromatography/mass spectrometry; PAGE, polyacrylamide gel electrophoresis; RP-HPLC, reverse-phase high-performance liquid chromatography; SDS, sodium dodecyl sulfate; TIC, total ion current; TFA, trifluoroacetic acid.

ACKNOWLEDGMENT

We thank Drs. M. Affolter, C. E. Hansen, J. McCarthy, and V. Tolstoguzov for critically reading the manuscript. We give special thanks to D. Moines for excellent technical assistance.

LITERATURE CITED

Bradford, M. M. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal. Biochem.* **1976**, *72*, 248−254.

Bycznska, A.; Barciszewski, J. The biosynthesis, structure and properties of napin—the storage protein from rape seeds. *J. Plant. Physiol.* **1999**, *154*, 417−425.

De Lumen, B. O.; Tappel, A. L. Fluorescein-hemoglobin as a substrate for cathepsin D and other proteases. *Anal. Biochem*. **1970**, *36*, 22−29.

Dodo, H. W.; Furtek, D. B. Cloning and sequencing of a gene encoding a 21 kDa trypsin inhibitor from *Theobroma cacao L. Cafe Cacao The* **1994**, *38*, 113−118.

Dodo, H. W.; Fritz, P. J.; Furtek, D. B. A cocoa 21 kilodalton seed protein has trypsin inhibitory activity. *Cafe Cacao The* **1992**, *36*, 279−284.

Ericson, M. L.; Rodin, J.; Lenman, M.; Glimelius, K.; Josefsson, L.-G.; Rask, L. Structure of the rapeseed 1.7S storage protein, napin, and its precursor. *J. Biol. Chem*. **1986**, *261*, 14576−14581.

Fenn, J. B.; Mann, M.; Meng, C. K.; Wong S. F.; Whitehouse, C. M. Electrospray ionization for mass spectrometry of large biomolecules. *Science* **1989**, *246*, 64−71.

Gartenmann, K.; Kochhar, S. Short-chain peptide analysis by high-performance liquid chromatography coupled to electrospray ionization mass spectrometer after derivatization with 9-fluorenylmethyl chloroformate. *J. Agric. Food Chem*. **1999**, *47*, 5068−5071.

Geiger, R. Chymotrypsin. In *Methods of Enzymatic Analysis. Vol. V. Enzymes 3: Peptidases, Proteinases and Their Inhibitors*; Bergmeyer, H., Bergmeyer, J., Grassal, M., Fritz, H., Eds.; Verlag Chemie: Weinheim, Germany, 1984.

Geiger, R.; Fritz, H. Trypsin. In *Methods of Enzymatic Analysis. Vol. V. Enzymes 3: Peptidases, Proteinases and Their Inhibitors*; Bergmeyer, H., Bergmeyer, J., Grassal, M., Fritz, H., Eds.; Verlag Chemie: Weinheim, Germany, 1984.

Hansen, C. E.; Mañez, A.; Burri, C.; Bousbaine, A. Comparison of enzyme activities involved in flavour precursor formation in unfermented beans of different cocoa genotypes. *J. Sci. Food Agric.* **2000**, *80*, 1−6.

Kaiser, R.; Metzka, L. Enhancement of cyanogen bromide cleavage yields for methionyl-serine and methionyl-threonine peptide bonds. *Anal. Biochem*. **1999**, *266*, 1−8.

Lerceteau, E.; Rogers, J.; Pétiard, V.; Crouzillat, D. Evolution of cocoa bean proteins during fermentation: a study by two-dimensional electrophoresis. *J. Sci. Food Agric.* **1999**, *79*, 619−625.

Lundell, N.; Schreitmüller, T. Sample preparation for peptide mapping—A pharmaceutical quality-control perspective. *Anal. Biochem*. **1999**, *266*, 31−47.

Masuda, Y.; Nirasawa, S.; Nakaya, K.; Kurihara, Y. Cloning and sequencing of a cDNA encoding a taste-modifying protein, miraculin. *Gene* **1995**, *161*, 175−177.

O'Conner, J.; Revell, D. F.; Masters, K. E.; Connerton, I. F.; Sumner, I. G Expression, purification, and initial characterization of the recombinant storage protein precursor of *Theobroma cacao. Protein Expression Purif.* **1996**, *7*, 377−383.

Odani, S.; Yokokawa, Y.; Takeda, Y.; Abe, S.; Odani, S. The primary structure and characterization of carbohydrate chains of the extracellular glycoprotein proteinase inhibitor from latex of *Carica papaya*. *Eur. J. Biochem*. **1996**, *241*, 77−82.

Ryan, C. A. Proteolytic enzymes and their inhibitors in plants. *Annu. Rev. Plant Physiol.* **1973a**, *24*, 173−196.

Ryan, C. A. Proteinase inhibitors. In *The Biochemistry of Plants*; Marcus, A., Ed.; Academic Press: New York, 1973b; Vol. 6, pp 351−370.

Schägger, H.; von Jagow, G. Tricine-sodium dodecyl sulfate-polyacrylamide gel electrophoresis for the separation of proteins in the range from 1 to 100 kDa. *Anal. Biochem*. **1987**, *166*, 368−379.

Spencer, M. E.; Hodge, R. Cloning and sequencing of a cDNA encoding the major albumin of *Theobroma cacao*. Identification of the protein as a member of the Kunitz protease inhibitor family. *Planta* **1991**, *183*, 528−535.

Spencer, M. E.; Hodge, R. Cloning and sequencing of a cDNA encoding the major storage proteins of *Theobroma cacao*. *Planta* **1992**, *186*, 567−576.

Tai, H.; McHenry, L.; Fritz, P. J.; Furtek, D. B. Nucleic acid sequence of a 21 kDa cocoa seed protein with homology to the soybean trypsin inhibitor (Kunitz) family of protease inhibitors. *Plant Mol. Biol.* **1991**, *16*, 913−915.

Voigt, J.; Biehl, B.; Kamaruddin, S.; Wazir, S. The major seed proteins of *Theobroma cacao L. Food Chem*. **1993**, *47*, 145−151.

Yavuz, M. O.; Ashton, S. M. V.; Deakin, E. D.; Spencer, M. E.; Sudbery, P. E. Expression of the major bean proteins from *Theobroma cacao* (cocoa) in the yeasts *Hansenula polymorpha* and *Saccharomyces cerevisiae*. *J. Biotechnol*. **1996**, *46*, 43−54.

Received for review May 16, 2000. Revised manuscript received August 25, 2000. Accepted August 25, 2000.

JF0006067