Food Chemistry 113 (2009) 903–913

Contents lists available at ScienceDirect

Food Chemistry

journal homepage: www.elsevier.com/locate/foodchem

Subunit structure of the vicilin-like globular storage protein of cocoa seeds and the origin of cocoa- and chocolate-specific aroma precursors

Ulrich Kratzer [a], Ronald Frank [b], Hubert Kalbacher [c], Böle Biehl [d], Johannes Wöstemeyer [e], Jürgen Voigt [f,*]

[a] PANATecs GmbH, Vor dem Kreuzberg 17, 72070 Tuebingen, Germany
[b] Department of Chemical Biology, Helmholtz Centre for Infection Research, Inhoffenstrasse 7, 38124 Braunschweig, Germany
[c] Research Centre for Medical and Natural Sciences, University of Tuebingen, Ob dem Himmelreich 7, 72074 Tuebingen, Germany
[d] Institute for Plant Biology, Technical University of Braunschweig, Mendelssohnstrasse 4, 38092 Braunschweig, Germany
[e] Institute for Microbiology, Friedrich-Schiller-University of Jena, Neugasse 24, 07743 Jena, Germany
[f] Institute for Biochemistry, University of Leipzig, Johannisallee 30, 04103 Leipzig, Sachsen, Germany

ARTICLE INFO

Article history:
Received 28 March 2008
Received in revised form 17 June 2008
Accepted 5 August 2008

Keywords:
Cocoa seeds
Epitope mapping
MALDI-ReTOF-MS
Storage protein
Theobroma cacao
Vicilin

ABSTRACT

Essential cocoa-specific aroma precursors are generated during fermentation of cocoa seeds by proteolysis of the vicilin-like globulin. This particular storage protein consists of three subunits with apparent molecular masses of 47 kDa, 31 kDa, and 15 kDa, respectively, which are derived from a common 66-kDa precursor. Since all these subunits resist N-terminal sequencing by Edman degradation, we have analysed the localisation of these vicilin subunits on their common precursor sequence by MALDI-TOF-MS analyses of their tryptic fragments. The results were corroborated by epitope mapping of polyclonal antibodies using 185 overlapping pentadeca-peptides covering the whole sequence of the precursor. The results show that the cocoa vicilin is encoded by a single gene and that heterogeneity of the vicilin subunits must be attributed to statistical post-translational modifications. Localisation of the precursor regions, from which the various subunits are derived, will be helpful for the future identification of aroma-relevant peptides generated during the fermentation process.

1. Introduction

Fermentation of cocoa seeds is required to obtain cocoa-specific aroma upon roasting (Rohan, 1964). Therefore, essential precursors of the specific aroma components are generated during the fermentation process. It has previously been shown that cocoa-specific aroma precursors are generated *in vitro* when polyphenol-free acetone dry powder prepared from unfermented cocoa seeds was subjected to autolysis at pH 5.2, e.g., under conditions of optimal fermentation (Voigt et al., 1994a). The proteolysis products obtained under these conditions consisted of hydrophilic peptides and hydrophobic free amino acids. When the same material was incubated at pH 3.5, no cocoa-specific aroma precursors were formed (Voigt et al., 1994a). Under the latter conditions, no liberation of free amino acids was observed and hydrophobic instead of hydrophilic peptides were detected in the autolysis products generated by the action of an aspartic endoprotease (Biehl et al., 1993; Guilloteau, Laloi, Michaux, Bucheli, & McCarthy, 2005; Laloi, McCarthy, Morandi, Gysler, & Bucheli, 2002). These hydrophobic peptides could be transformed to hydrophilic peptides by treatment with commercial carboxypeptidase A at pH 5.2 accompanied by liberation of hydrophobic free amino acids (Voigt et al., 1994a). This mixture of hydrophilic peptides and hydrophobic free amino acids revealed typical cocoa flavour when roasted in the presence of reducing sugars and deodorised cocoa butter (Voigt et al., 1994a). This experimental approach has shown that the cocoa-specific aroma precursors are generated from seed proteins by co-operation of an aspartic endoprotease and a carboxypeptidase (Voigt et al., 1994a). Indeed, ungerminated cocoa seeds contain a carboxypeptidase with an optimum activity around pH 5.8, which preferentially liberates hydrophobic amino acids and whose activity drops considerably at pH-values below 5.0 (Bytof, Biehl, Heinrichs, & Voigt, 1995; Hansen, del Olmo, & Burri, 1998; Voigt et al., 1994a; Yusep, Jinap, Jamilah, & Nazamid, 2002).

Fermentation-like incubations of intact cocoa seeds under controlled laboratory conditions have revealed that at pH 4.0–4.5, an unspecific proteolysis of all the seed proteins takes place, whereas at pH 5.0–5.5 distinct vacuolar storage proteins are selectively degraded (Biehl, Brunner, Passern, Quesnel, & Adomako, 1985). The polypeptides, which are selectively degraded during seed incubations at pH 5.0–5.5, have been shown to be the polypeptide subunits of a vicilin-type globulin (Spencer & Hodge, 1992; Voigt, Biehl, & Kamaruddin, 1993), indicating that the cocoa-specific aroma precursors are derived from vicilin-like globulin of the cocoa seeds. This assumption was corroborated by *in vitro* studies of

* Corresponding author. Tel.: +49 30 450 528144; fax: +49 30 450 528942.
E-mail address: juergen.voigt@charite.de (J. Voigt).

0308-8146/$ - see front matter © 2008 Elsevier Ltd. All rights reserved.
doi:10.1016/j.foodchem.2008.08.017

the proteolytic formation of these particular aroma precursors using purified substrate proteins and proteases (Voigt, Heinrichs, Voigt, & Biehl, 1994b).

The vicilin-like globular storage protein of the cocoa seeds was found to consist of three subunits with apparent molecular masses of 47 kDa, 31 kDa, and 15 kDa (Voigt et al., 1993), respectively, which are derived from a common 66-kDa precursor (Spencer & Hodge, 1992). Unfortunately, all these subunits were found to be resistant to N-terminal sequencing by Edman degradation, presumably due to N-terminal protection. Therefore, we have analysed their localisation on the common precursor sequence by MALDI-ReTOF analyses of their tryptic fragments and by epitope mappings of polyclonal antibodies using overlapping scan peptides as previously reported (Voigt, Wöstemeyer, & Frank, 2007).

2. Materials and methods

2.1. Materials

Cocoa seeds were from ripe, genetically undefined pods harvested at the Cocoa and Coconut Division of the Malaysian Agricultural Research and Development Institute (MARDI, Hilir Perak, Malaysia). Unfermented seeds were taken from the pods immediately after arrival (4–5 days after harvesting), shock-frozen in liquid nitrogen after removal of testae and radiculae, and freeze-dried.

2.2. Extraction of fat

The dry cotyledons were crushed and portions of 10 g each were extracted repeatedly in a Soxhlet apparatus with 500 ml of petroleum ether (b.p. 40–70 °C). After solvent evaporation, the material was powdered and extracted again for 8 h in the same manner. Finally, purine alkaloids were partially extracted with chloroform for 8 h in a Soxhlet apparatus.

2.3. Preparation of acetone dry powder (AcDP)

Acetone dry powder (AcDP) of cocoa seeds was prepared as previously described (Voigt et al., 1993). To remove the polyphenols, the defatted seed powders were extracted three times with 80% (v/v) aqueous acetone containing 5 mM sodium ascorbate and subsequently with 70% (v/v) aqueous acetone: the suspensions (200 ml aqueous acetone per 10 g seed powder) were stirred for 1 h at 4 °C and the extracts removed by centrifugation (15 min at 13,000*g*). After the final extraction step, efficiency of polyphenol extraction was checked by heating an aliquot of the acetone dry powder with 5 M HCl (red colour indicates the presence of residual polyphenols). After complete extraction of polyphenols, residual water was removed by extraction with 100% acetone. After final centrifugation, the sediment was evaporated under reduced pressure to remove the solvent. The acetone dry powder (AcDP) was stored at −20 °C.

2.4. Fractionation of seed proteins

Seed proteins were extracted from polyphenol-free acetone dry powder (AcDP) as recently described (Voigt et al., 1993): AcDP (250 mg) was first extracted with 25 ml of a low-salt buffer containing 5 mM sodium ascorbate, 2 mM EDTA, 10 mM Tris–HCl (pH 7.5), and 10 μg/ml pepstatin A to obtain the albumin fraction. The suspension was stirred for 1 h at 4 °C and subsequently centrifuged for 20 min at 20,000*g* and 4 °C in the Sorvall GSA rotor (Du Pont de Nemours GmbH, Bad Homburg, Germany). Extraction was repeated three times to minimise a carryover of proteins soluble under low-salt conditions to the globulin fraction. Subsequently, the pretreated AcDP was extracted with 25 ml of 0.5 M NaCl containing 5 mM sodium ascorbate, 2 mM EDTA, 20 mM Tris–HCl (pH 7.5), and 10 μg/ml pepstatin A. The suspension was stirred at 4 °C for 1 h and centrifuged for 20 min at 20,000*g* and 4 °C in the Sorvall GSA rotor. The extraction was repeated, the combined supernatants adjusted to 200 mM 2-mercaptoethanol and incubated at 4 °C for 2 h. When the high-salt extract was dialysed against distilled water, a precipitate was formed which was found to consist of almost pure vicilin-like globulin (Voigt et al., 1993), which was collected by centrifugation and washed twice with bidistilled water and stored at −20 °C.

2.5. Modification of cysteine residues

Cocoa globulin to be used for mass spectrometric analyses or amino acid analyses was pretreated with iodoacetamide to modify cysteine residues. The pelleted reduced globulin was redissolved in 0.5 M NaCl containing 20 mM sodium bicarbonate buffer (pH 10.0). After addition of iodoacetamide to a final concentration of 10 mM, the solution was stirred at 4 °C for 2 h and subsequently passed through a Sephadex G50 column (1.6 × 85 cm) equilibrated and eluted with 0.5 M NaCl containing 7 mM 2-mercaptoethanol and 20 mM Tris–HCl (pH 8.0). The flow-through fractions were combined and dialysed overnight against distilled water. The precipitated globulin was then collected by centrifugation, washed twice with bidistilled water and stored at −20 °C until use.

2.6. Determination of protein

Quantitations of proteins were performed by the method of Bradford (1976) using bovine serum albumin as standard.

2.7. SDS–PAGE and protein gel blot analyses

The precipitated cocoa globulin was washed twice with distilled water and redissolved in a small volume of urea-SDS buffer (Voigt, 1985). After addition of sample buffer (Laemmli, 1970) containing bromophenol blue as tracking dye, the polypeptides were separated by SDS–PAGE, according to Laemmli (1970), on gel slabs (83 mm × 65 mm × 1.5 mm) containing 17% (w/v) acrylamide. After electrophoresis, the gels were either stained with Coomassie Brilliant Blue G-250 or blotted electrophoretically onto PVDF membranes (Porablot, Macherey-Nagel, Düren, Germany) as described by Towbin, Staehelin, and Gordon (1979). After pre-treatment of the blots with 3% (w/v) bovine serum albumin in phosphate-buffered saline (PBS) for 3 h at 37 °C or overnight at 7 °C, the protein gel blots were probed with polyclonal antibodies raised in rabbits against different cocoa globulin polypeptides (Voigt et al., 1995) diluted 1:300 with PBS containing 1% (w/v) bovine serum albumin. After 1 h at 37 °C, the blots were washed four times with PBS for 30 min at room temperature followed by treatment for 1 h at 37 °C with alkaline phosphatase-coupled goat anti-rabbit IgG antibodies (Amersham-Buchler, Braunschweig, Germany) diluted 1:800 with PBS containing 1% (w/v) bovine serum albumin. After incubation, the blots were washed at least twice with PBS and subsequently twice with TBS for 30 min at room temperature. Finally, the indirectly bound alkaline phosphatase was detected by its enzymatic activity using 5-bromo-4-chloro-3-indolyl phosphate as substrate in the presence of nitroblue-tetrazolium chloride (Hawkes, Nidag, & Gordon, 1982).

2.8. Two-dimensional gel electrophoresis

Two-dimensional gel electrophoresis of cocoa seed proteins was performed by isoelectric focussing (1st dimension) and subsequent

SDS–PAGE (2nd dimension), as described by (O'Farrell, 1975). For calibration of the pH gradient, a standard protein mixture was separated in two of the isoelectric focussing gels, in parallel to the cocoa protein samples. These control gels were stained for protein with Coomassie Brilliant Blue after fixation and removal of the ampholines. Molecular masses were determined by standard protein markers covering a range of 12–68 kDa in the SDS–PAGE. After electrophoresis, the gels were stained with Coomassie Brilliant Blue G-250.

2.9. Epitope analysis

A total of 185 overlapping pentadeca-peptides representing the whole amino acid sequence of the 66-kDa precursor of the cocoa vicilin-like globulin protein derived from the ORF of the corresponding cDNA (Spencer & Hodge, 1992) and gene (McHenry & Fritz, 1992) were generated by spot synthesis, using a cellulose paper sheet as solid support (Frank, 1992). These overlapping pentadeca-peptides were used to determine the epitope specificities of the different antibodies, as previously described (Voigt et al., 2007). After incubation with the antibody and extensive washing, bound rabbit IgG was measured by incubation with alkaline phosphatase-coupled goat anti-(rabbit IgG) Ig (Amersham-Buchler, Braunschweig, Germany) and subsequent detection of the indirectly bound alkaline phosphatase *via* its enzyme activity, using 5-bromo-4-chloro-3-indolyl phosphate as substrate in the presence of MTT(3-[4,5-dimethylthiazol-2-yl]-2,5-diphenyl-tetrazolium bromide). After documentation of the results, the membrane was stripped, as previously described (Voigt et al., 2007).

2.10. Determination of N-terminal amino acid sequences

The untreated and chemically deglycosylated cell wall polypeptides and their fragments obtained by cleavage with GluC were purified by SDS–PAGE according to Laemmli (1970) and blotted electrophoretically onto PVDF membranes (Porablot, Macherey-Nagel) as described by Towbin et al. (1979). After staining with Coomassie Brilliant Blue G-250, the protein bands were cut from the membrane and completely destained with methanol. N-terminal amino acid sequences were determined by automated Edman degradation using a gas-phase sequencer (model 470A, Applied Biosystems, Foster City, CA).

2.11. In-gel protein digestion

In-gel tryptic digests were performed as described (Wilm et al., 1996) and modified as outlined below. Briefly, protein bands were excised from gels, fully destained with 30% (v/v) acetonitrile/50 mM ammonium bicarbonate, reduced, and alkylated. Gel pieces were digested overnight at 37 °C in a buffer containing 50 mM ammonium bicarbonate and 12.5 ng/μl sequencing-grade modified porcine trypsin (Promega, Madison, WI). Peptides were then extracted with 5% (v/v) formic acid/50% (v/v) acetonitrile. The extracts were combined, vacuum dried, and the remainder redissolved in 10 μl 1% trifluoroacetic acid.

2.12. MALDI-ReTOF-MS

Peptide solution (0.5 μl) was mixed with the same volume of 2,5-dihydroxybenzoic acid matrix on a MALDI-TOF mass spectrometer gold target. The mass spectrometer (Bruker Reflex IV, Bruker Daltonik, Bremen, Germany) was operated in positive ion reflector mode. After external calibration with a standard peptide mixture (Peptide Calibration Standard, Bruker Daltonik), tryptic peptides were analysed by adding 300 laser shots and autoproteolytic trypsin peaks of the obtained spectra were used for internal calibration. The resulting monoisotopic peptide masses were submitted to the non-redundant NCBI protein database using the Mascot peptide mass fingerprint search algorithm (http://www.matrixscience.com) (Perkins, Pappin, Creasy, & Cottrell, 1999). A peptide mass tolerance ±0.15 Da, up to 3 missed cleavage sites, and optional oxidation of methionine residues were allowed.

2.13. Amino acid analyses

To analyse the amino acid composition, the vicilin subunits were separated by SDS–PAGE according to Laemmli (1970) on a preparative scale. Side strips of the gel slabs were stained for protein with Coomassie. The regions of the unstained parts of the gels corresponding to the 47-kDa and 31-kDa polypeptides were cut out and the proteins isolated by electroelution. After removal of SDS and buffer components, proteins were freeze-dried and hydrolysed with 1–2 ml of 6 M HCl containing 10 mM phenol and 0.2% (v/v) 2-mercaptoethanol for 20 h at 110 °C under a nitrogen atmosphere. After removal of HCl, the amino acid composition was determined using an automatic amino acid analyser (model LC 6001, Biotronik, Berlin, Germany).

2.14. C-terminal sequencing

To determine the C-terminal amino acid sequences, the vicilin subunits were incubated with carboxypeptidase Y. Aliquots were taken, quick-frozen, and freeze-dried at different time intervals. Finally, the released amino acid residues were analysed by amino acid analyses and the C-terminal sequence determined from the kinetics of the release of the different amino acids.

3. Results

3.1. Fractionation and characterisation of the predominant cocoa seed proteins

All the cocoa seed proteins can be solubilised from polyphenol-free acetone dry powder by extractions with high-salt buffers containing 0.5 M NaCl (Fig. 1a and b; lane 1). During dialysis against a salt-free buffer, the albumin fraction remained soluble (Fig. 1a and b; lane 2) whereas the globulin was precipitated and could be collected by centrifugation (Fig. 1a and b; lane 3). Five prominent bands with apparent molecular masses of 47 kDa, 31 kDa, 28 kDa, 16 kDa, and 15 kDa were usually detected in this globulin fraction (Fig. 1a; lane 3). The relative amounts of the 28-kDa and 16-kDa components were, however, strongly reduced when the extraction of the seed proteins was performed in the presence of a sufficiently high concentration of pepstatin A, an inhibitor of aspartic endoproteinases (Fig. 1b; lane 3). These data indicate that the 28-kDa and 16-kDa components are not intrinsic subunits of the cocoa vicilin but are generated by partial proteolysis during preparation of this particular globular storage protein.

Two-dimensional electrophoresis of the cocoa seed proteins by isoelectric focussing (IEF) and subsequent SDS–PAGE revealed that the 47-kDa, 31-kDa, and 28-kDa components are not unique with respect to their isoelectric points (Fig. 2). By IEF, the 47-kDa component was split into six spots with isoelectric points (IEPs) between 5.2 and 6.4. The 31-kDa and 28-kDa bands observed by SDS–PAGE (Fig. 1a and b; lane 3) were split into several spots with IEPs between 6.1 and 7.3 (Fig. 2).

3.2. MALDI-ReTOF-MS analyses of tryptic fragments

Attempts to determine the N-terminal amino acid sequences of the globulin subunits from cocoa seeds by Edman degradation

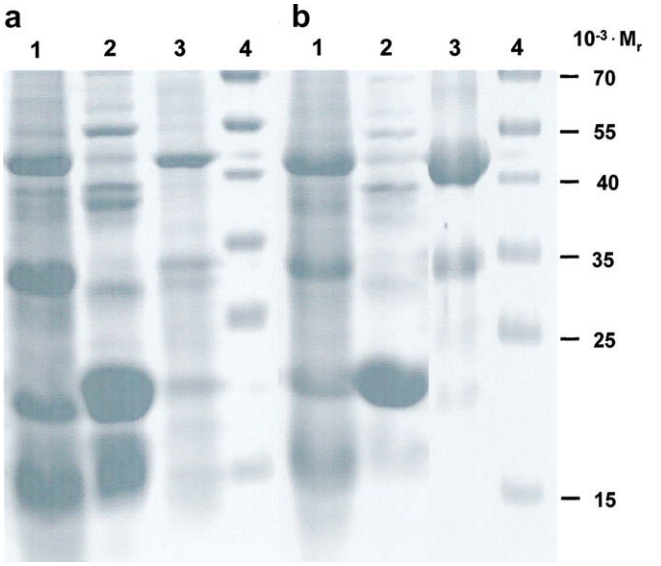

Fig. 1. Coomassie-stained SDS–PAGE profiles of the protein fractions obtained by solubility fractionation of proteins from unfermented cocoa seeds. Proteins were extracted in the absence (a) or presence (b) of the aspartic endoprotease inhibitor pepstatin A. Lane 1, unfractionated high-salt extract of acetone dry powder; lane 2, supernatants obtained after dialysis against water (albumin fraction); lane 3, pellet fraction obtained after dialysis (globulin fraction); lane 4, protein molecular weight standard.

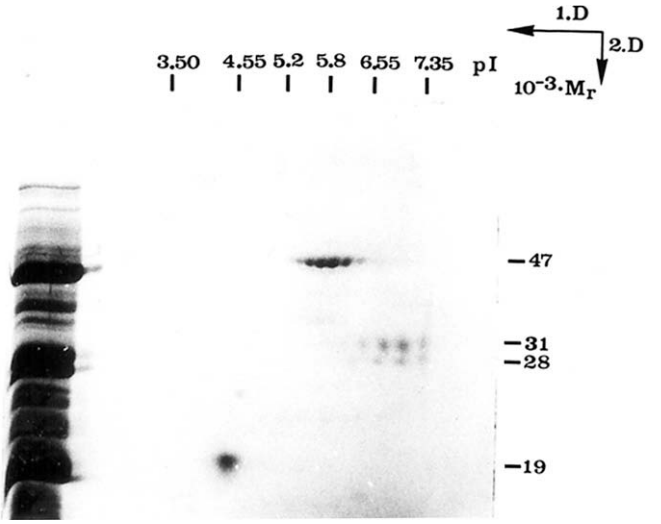

Fig. 2. Two-dimensional electrophoresis of the proteins from unfermented cocoa seeds. In the first dimension, 30 μg of seed proteins extracted from acetone dry powder with 8 M urea were separated by isoelectric focussing. Calibration was performed using a pI standard protein mixture, which was fractionated in parallel under identical conditions. In the second dimension, proteins were fractionated according to their size by SDS–PAGE electrophoresis using gel slabs containing 17% (w/v) acrylamide. As a control, total cocoa seed proteins (30 μg) were fractionated by SDS–PAGE only on the same gel slab (see on the left).

were not successful. On the one hand, 47-kDa, 31-kDa, and 15-kDa subunits were found to be resistant against Edman degradation. On the other hand, mixtures of several amino acid derivatives were released from the N-terminal ends of the 28-kDa and 16-kDa polypeptides during each round of Edman degradation (data not shown) indicating that the N-terminal ends of these polypeptides are heterogeneous.

For this reason, we have tried to localise the 47-kDa, 31-kDa, and 15-kDa subunits of the cocoa globulin on their common 66-kDa precursor sequence by MALDI-ReTOF-MS analyses of their tryptic peptides (Fig. 3; Tables 1–3). All the prominent peaks detected in the MALDI-ReTOF-MS spectra of the tryptic digests of the 47-kDa (Fig. 3a), of the 31-kDa (Fig. 3b), and of the 15-kDa (Fig. 3c) subunits could be attributed to corresponding amino acid sequences of the 66-kDa precursor (Tables 1–3). Thirty-one matches were obtained for the 47-kDa subunit, which were localised between positions 141 and 545 of the precursor sequence (Table 1) and covered 70% of the amino acid sequence in this particular region. More than half of these peptide fragments also occurred in the MALDI-ReTOF-MS spectrum of the tryptic digest of the 31-kDa subunit (Tables 1 and 2), including the prominent peaks at *m/z* 2131.1, *m/z* 1587.9, and *m/z* 1415.8 (Fig. 3a and b). These findings indicate that the amino acid sequences of the 47-kDa and 31-kDa subunits are overlapping. The prominent peaks at *m/z* 2757.22, 2551.18, 2358.10, 2286.27, and 1970.08 of the 47-kDa subunit (Fig. 3a), which were lacking in the tryptic digest of the 31-kDa subunit (Fig. 3b), could be attributed to positions 435–456, 525–545, 438–456, 457–479, and 480–497 of the precursor sequence (Table 1). The latter findings clearly show that the 31-kDa subunit does not overlap with the C-terminal part of the 47-kDa subunit (positions 435–545 of the precursor sequence). Indeed, the 30 matches obtained for the 31-kDa subunit were localised between positions 132 and 425 of the 66-kDa precursor (Table 2) and covered 63% of the precursor sequence in this particular region.

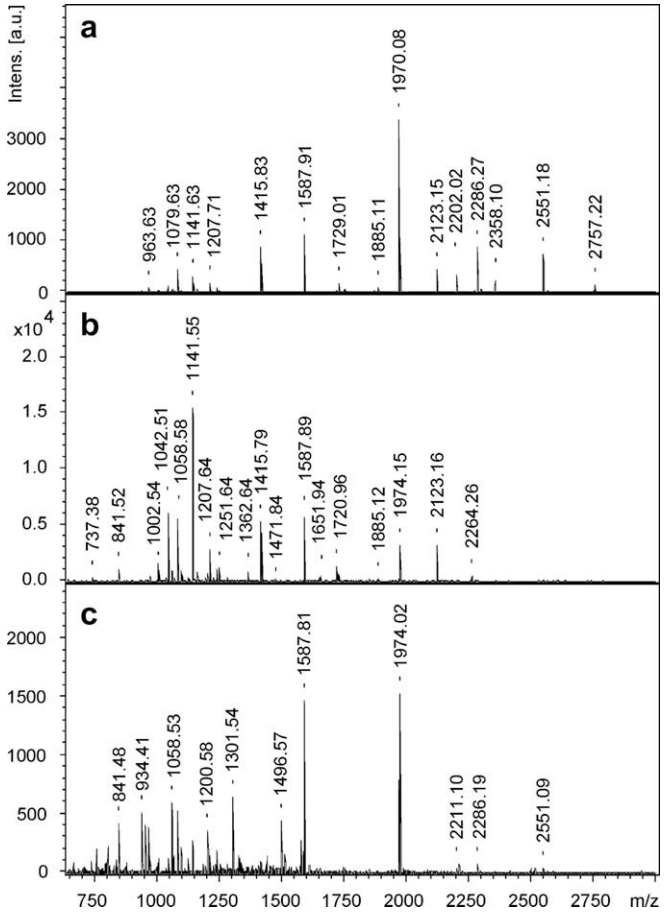

Fig. 3. Comparison of the MALDI mass spectra of tryptic in-gel digests of (a) the 47-kDa, (b) the 31-kDa, and (c) the 15-kDa subunits of the cocoa seed vicilin. All labelled peaks denote monoisotopic peptide masses $[M + H]^+$.

Table 1
Peptide fragments detected by MALDI-TOF analyses in the tryptic digests of the 47-kDa cocoa vicilin subunit

Sequence[a]	Position[b]	$[M+H]^+$		
		Observed	Calculated	Δ
R.NNPYYFPK.R	141–148	1042.6049	1042.4993	0.1056
R.SFQTR.F	151–155	638.4247	638.3257	0.0991
R.FRDEEGNFK.I	156–164	1141.6341	1141.5273	0.1068
R.FAENSPPLK.G	169–177	1002.6432	1002.5255	0.1178
R.FAENSPPLKGINDYR.L	169–183	1720.9571	1720.8653	0.0918
K.GTITFVTHENKESYNVQR.G	213–230	2123.1472	2123.0516	0.0956
R.GTVVSVPAGSTVYVVSQDNQEK.L	231–252	2264.2124	2264.1405	0.0719
R.EKLEEILEEQR.G	300–310	1415.8292	1415.7376	0.0916
R.EKLEEILEEQRGQK.R	300–313	1729.0100	1728.9126	0.0974
K.LEEILEEQR.G	302–310	1158.7089	1158.6001	0.1088
K.LEEILEEQRGQK.R	302–313	1471.8409	1471.7751	0.0659
K.RQQGQQGMFR.K	314–323	1235.7110	1235.6062	0.1048
R.QQGQQGMFR.K	315–323	1079.6271	1079.5051	0.1220
R.QQGQQGMFR.K[c]	315–323	1095.6251	1095.5000	0.1251
R.QQGQQGMFRK.A	315–324	1207.7073	1207.6001	0.1072
R.KAKPEQIR.A	324–331	969.6809	969.5840	0.0969
R.AISQQATSPR.H	332–341	1058.6550	1058.5589	0.0961
R.LAINLLSQSPVYSNQNGR.F	348–365	1974.1384	1974.0403	0.0981
K.LNQGAIFVPHYNSK.A	389–402	1587.9077	1587.8278	0.0799
K.ATFVVFVTDGYGYAQMACPHLSR.Q	403–425	2590.2733	2590.2218	0.0516
R.QSQGSQSGR.Q	426–434	934.5741	934.4337	0.1404
R.QDRREQEEESEEETFGEFQQVK.A	435–456	2757.2236	2757.2234	0.0002
R.REQEEESEEETFGEFQQVK.A	438–456	2358.1009	2358.0368	0.0641
R.EQEEESEEETFGEFQQVK.A	439–456	2202.0151	2201.9357	0.0795
K.APLSPGDVFVAPAGHAVTFFASK.D	457–479	2286.2651	2286.1918	0.0734
K.DQPLNAVAFGLNAQNNQR.I	480–497	1970.0822	1969.9839	0.0984
R.QMDSEAKELSFGVPSK.L	509–524	1752.9314	1752.8473	0.0841
R.QMDSEAKELSFGVPSK.L[c]	509–524	1768.8971	1768.8422	0.0549
K.ELSFGVPSK.L	516–524	963.6350	963.5146	0.1204
K.LVDNIFNNPDESYFMSFSQQR.Q	525–545	2551.1816	2551.1558	0.0258
K.LVDNIFNNPDESYFMSFSQQR.Q[c]	525–545	2567.1798	2567.1507	0.0291

[a] The amino acid residues located upstream and downstream of the detected peptides in the total cocoa vicilin precursor sequence (GenBank accession number S22477) are separated by points.
[b] Position in the amino acid sequence of the cocoa vicilin precursor (McHenry & Fritz, 1992; Spencer & Hodge, 1992).
[c] Oxidated peptide (M).

3.3. Immunochemical studies

Polyclonal antibodies were raised against the 47-kDa and 31-kDa subunits of the cocoa globulin (Fig. 1a and b; lane 3) and against the 28-kDa and 16-kDa fragments obtained under conditions where the intrinsic aspartic endoproteinase was not inhibited (Fig. 1a; lane 3). Protein-blot analyses revealed that the antibody raised against the 47-kDa subunit not only recognised the corresponding antigen but also cross-reacted with the 31 kDa and three smaller polypeptides with apparent molecular masses of 15.5 kDa, 15.0 kDa, and 14.5 kDa, the 15.0-kDa band being most prominent (Fig. 4b; lane 1). This finding was not unexpected because analyses of the vicilin-like globulins of other plants have shown that the largest subunit covers the sequences of the smaller subunits, indicating that the smaller subunits are derived from the largest subunit by proteolytic processing, e.g., the largest subunit is usually split into the smaller ones (Müntz, 1998). Therefore, the polyclonal antibody raised against the 31-kDa subunit was expected to cross-react with the 47-kDa subunit but not with the smaller constituents of the cocoa vicilin. This particular antibody, however, also cross-reacted with the 15.5-kDa, 15.0-kDa, and 14.5-kDa polypeptides (Fig. 4c; lane 1). No clear cross-reactivity with the 15.5-kDa, 15.0-kDa, and 14.5-kDa polypeptides was, however, observed for the polyclonal antibody raised against the 28-kDa fragment, even in the case of increased amounts of cocoa vicilin (Fig. 4d; lane 1). This antibody cross-reacted with the 47-kDa and 31-kDa subunits (Fig. 4d; lane 1). The same results were obtained for the antibody against the 16-kDa fragment (Fig. 4e; lane 1).

The epitope specificities of all these antibodies were studied by scan peptide analyses (Frank, 1992; Voigt et al., 2007). To this end, 185 overlapping pentadeca-peptides, with an overlap of 12 amino acids between neighbouring peptides, were synthesised on a solid support, which together represented the whole amino acid sequence of the 66-kDa precursor of the cocoa globulin (Fig. 5a), and probed with the different polyclonal antibodies (Fig. 5b). Epitope mappings with monoclonal antibodies have shown that continuous epitopes recognised by individual immunoglobulins usually consist of 5–9 amino acid residues (Voigt et al., 2007). In the case of scan peptides with a length of 15 amino acids and an overlap in sequence of 12 amino acids between neighbouring peptides as used in this study, such continuous epitopes must occur in 3–5 neighbouring peptides. The polyclonal antibody raised against the 47-kDa subunit recognised several such continuous epitopes in the region between positions 140 and 558 of the precursor sequence covered by the scan peptides 46–182 (upper part of Fig. 5b). Some peptides upstream or downstream of this precursor region also reacted with this particular antibody. In these cases, the antibody recognised a single pentadeca peptide or two neighbouring peptides only, indicating that the corresponding epitopes were non-continuous and generated by folding of the corresponding pentadeca-peptides. These data indicate that the 47-kDa subunit is localised in the precursor region between positions 140 and 558.

Less clear were the results of the epitope mappings of the polyclonal antibodies raised against the 31-kDa subunit and against the 28-kDa proteolytic fragment, respectively (middle parts of Fig. 5b). The binding patterns of both antibodies differed considerably from those observed for the antibody raised against the 47-kDa subunit (upper part of Fig. 5b) and from each other (middle parts of Fig. 5b). In the case of the antibody raised against the 31-kDa subunit, most of the recognised epitopes were localised between positions 120

Table 2
Peptide fragments detected by MALDI-TOF analyses in the tryptic digests of the 31-kDa cocoa vicilin subunit

Sequence[a]	Position[b]	$[M+H]^+$		
		Observed	Calculated	Δ
R.SEEEEGQQR.N	132–140	1091.4802	1091.4600	0.0202
R.NNPYYFPK.R	141–148	1042.5124	1042.4993	0.0132
R.NNPYYFPKR.R	141–149	1198.5996	1198.6004	0.0008
R.SFQTR.F	151–155	638.3468	638.3257	0.0211
R.FRDEEGNFK.I	156–164	1141.5514	1141.5273	0.0242
R.FRDEEGNFKILQR.F	156–168	1651.9355	1651.8551	0.0804
R.FAENSPPLK.G	169–177	1002.5370	1002.5255	0.0115
R.FAENSPPLKGINDYR.L	169–183	1720.9616	1720.8653	0.0963
K.GINDYR.L	178–183	737.3824	737.3577	0.0248
K.GTITFVTHENKESYNVQR.G	213–230	2123.1584	2123.0516	0.1068
K.ESYNVQR.G	224–230	895.4586	895.4268	0.0318
R.GTVVSVPAGSTVYVVSQDNQEK.L	231–252	2264.2636	2264.1405	0.1231
K.LTIAVLALPVNSPGK.Y	253–267	1492.9615	1492.9098	0.0518
R.EKLEEILEEQR.G	300–310	1415.7867	1415.7376	0.0491
R.EKLEEILEEQRGQK.R	300–313	1729.0148	1728.9126	0.1022
K.LEEILEEQR.G	302–310	1158.6569	1158.6001	0.0569
K.LEEILEEQRGQK.R	302–313	1471.8390	1471.7751	0.0640
K.LEEILEEQRGQKR.Q	302–314	1627.9503	1627.8762	0.0742
K.RQQGQQGMFR.K	314–323	1235.6497	1235.6062	0.0435
K.RQQGQQGMFR.K[c]	314–323	1251.6358	1251.6011	0.0347
R.QQGQQGMFR.K	315–323	1079.5243	1079.5051	0.0192
R.QQGQQGMFR.K[c]	315–323	1095.5261	1095.5000	0.0261
R.QQGQQGMFRK.A	315–324	1207.6446	1207.6001	0.0445
R.QQGQQGMFRK.A[c]	315–324	1223.6349	1223.5950	0.0399
R.KAKPEQIR.A	324–331	969.6061	969.5840	0.0221
K.AKPEQIR.A	325–331	841.5211	841.4890	0.0321
R.AISQQATSPR.H	332–341	1058.5788	1058.5589	0.0199
R.LAINLLSQSPVYSNQNGR.F	356–373	1974.1548	1974.0403	0.1145
K.LNQGAIFVPHYNSK.A	389–402	1587.8905	1587.8278	0.0627
K.ATFVVFVTDGYGYAQMACPHLSR.Q	403–425	2590.3049	2590.2218	0.0831

[a] The amino acid residues located upstream and downstream of the detected peptides in the total cocoa vicilin precursor sequence (GenBank accession number S22477) are separated by points.
[b] Position in the amino acid sequence of the cocoa vicilin precursor (McHenry & Fritz, 1992; Spencer & Hodge, 1992).
[c] Oxidated peptide (M).

Table 3
Peptide fragments detected by MALDI-TOF analyses in the tryptic digests of the 14.5-kDa cocoa vicilin subunit

Sequence[a]	Position[b]	$[M+H]^+$		
		Observed	Calculated	Δ
R.NNPYYFPK.R	141–148	1042.4683	1042.4993	0.0309
R.QQGQQGMFR.K	315–323	1079.4715	1079.5051	0.0336
R.QQGQQGMFR.K[c]	315–323	1095.4845	1095.5000	0.0155
K.AKPEQIR.A	325–331	841.4832	841.4890	0.0058
R.AISQQATSPR.H	332–341	1058.5340	1058.5589	0.0249
R.LAINLLSQSPVYSNQNGR.F	348–365	1974.0231	1974.0403	0.0172
K.LNQGAIFVPHYNSK.A	389–402	1587.8110	1587.8278	0.0168
R.QSQGSQSGR.Q	426–434	934.4097	934.4337	0.0240
K.APLSPGDVFVAPAGHAVTFFASK.D	457–479	2286.1855	2286.1918	0.0063
K.DQPLNAVAFGLNAQNNQR.I	480–497	1969.9749	1969.9839	0.0089
K.ELSFGVPSK.L	516–524	963.4956	963.5146	0.0190
K.LVDNIFNNPDESYFMSFSQQR.Q	525–545	2551.0904	2551.1558	0.0654

[a] The amino acid residues located upstream and downstream of the detected peptides in the total cocoa vicilin precursor sequence (GenBank accession number S22477) are separated by points.
[b] Position in the amino acid sequence of the cocoa vicilin precursor (McHenry & Fritz, 1992; Spencer & Hodge, 1992).
[c] Oxidated peptide (M).

and 438 of the precursor sequence covered by the scan peptides 40–142 (Fig. 5a and b). The scan peptides 43–45 covering positions 127–147 of the precursor sequence (Fig. 5a) were recognised by the antibody against the 31-kDa subunit but not by that raised against the 47-kDa subunit (upper parts of Fig. 5b). These results indicate that the 31-kDa subunit does not exactly correspond to the N-terminal domains of the 47-kDa subunit but starts upstream of the 47-kDa subunit on the common 66-kDa precursor sequence.

The polyclonal antibody raised against the 16-kDa fragment predominantly reacted with the scan peptides 46–48, 51–52, 56–63, 91–93, and 95–96 (lower part of Fig. 5b), indicating that this proteolytic fragment corresponds to the N-terminal region of the 47-kDa subunit (compare to the upper part of Fig. 5b).

3.4. C-terminal ends and amino acid compositions

When the electrophoretically purified 47-kDa subunit was degraded with carboxypeptidase Y, phenylalanine and leucine were released faster than arginine and almost no glutamine was liberated, indicating that Arg545 is not its carboxyterminal end. The

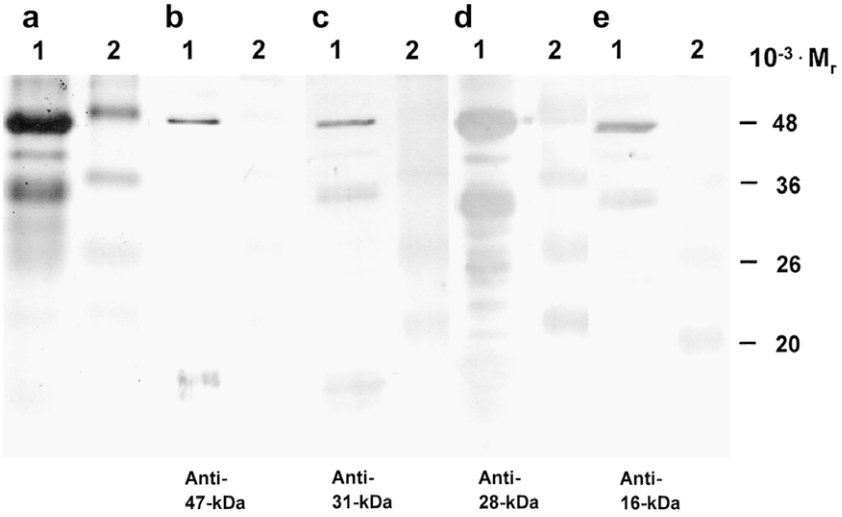

Fig. 4. SDS–PAGE and protein gel blot analyses of the cocoa vicilin subunits. Polypeptides of the globulin fraction obtained by solubility fractionation of high-salt extracts of acetone dry powder from unfermented cocoa seeds (compare to Fig. 1) were fractionated by SDS–PAGE on gel slabs. After electrophoresis, the gels were either stained with Coomassie Brilliant Blue (a) or blotted onto PVDF membranes and subsequently probed with different polyclonal antibodies raised against the 47-kDa and the 31-kDa subunits (b and c) as well as the 28-kDa and the 16-kDa proteolytic fragments (d and e) of cocoa vicilin, respectively. Lanes 1: proteins from cocoa vicilin; lanes 2: prestained protein molecular weight standard with molecular masses given on the right.

kinetics of the release of amino acids by carboxypeptidase Y revealed a C-terminal sequence ARLF (Table 4), which corresponds to the C-terminal sequence of the 66-kDa precursor. Furthermore, the results of the amino acid analysis of the 47-kDa subunit were in agreement with the assumption that this particular subunit corresponds to the precursor region between Asn141 and its carboxy-terminus Phe566 (Table 4). A theoretical molecular mass of 48 kDa was calculated for this sequence.

In the case of the 31-kDa subunit, covering the precursor region between Ser132 and Arg425 (Fig. 3b; Table 2), arginine was indeed the amino acid, which was most rapidly released by carboxypeptidase Y, followed by serine (Table 4). Again, these data were further corroborated by the results of the amino acid analysis (Table 4). A theoretical molecular mass of 33 kDa was calculated for this sequence.

Unfortunately, the corresponding results were less clear for the 14.5-kDa subunit because both yields and purities of the preparations of this particular subunit were insufficient (data not shown). When the purified 28-kDa and 16-kDa polypeptides were subjected to degradation with carboxypeptidase Y, several hydrophobic amino acids were released with the same kinetics, followed by the liberation of other amino acids (data not shown), indicating that the carboxyterminal ends of both polypeptides are heterogeneous.

4. Discussion

The seeds of *Theobroma cacao* were shown to contain two prominent storage proteins (Voigt et al., 1993), an albumin with an apparent molecular mass of 19 kDa (Fig. 1a and b; lane 2) and a globulin fraction (Fig. 1a and b; lane 3). As previously reported, the cocoa-specific aroma precursors are generated during fermentation of the cocoa seeds by proteolytic processing of the globulin (Voigt & Biehl, 1995; Voigt et al., 1994a). The vicilin-like globulin of the cocoa seeds contains two prominent subunits with apparent molecular masses of 47 kDa and 31 kDa (Fig. 1a and b; lane 3). Smaller constituents of the cocoa vicilin with apparent molecular masses of 15.5, 15.0, and 14.5 kDa were usually hardly detected by Coomassie staining (Fig. 1a and b; lane 3) but their presence was clearly shown by protein-blot analyses with polyclonal antibodies raised against the 47-kDa and the 31-kDa subunits, respectively (Fig. 4b and c; lane 1). All these polypeptides are generated by proteolytic processing of a 66-kDa precursor, whose amino acid sequence was derived from the nucleotide sequences of the corresponding cDNA (Spencer & Hodge, 1992) and gene (McHenry & Fritz, 1992). Localisation of the cocoa vicilin subunits on their common precursor by N-terminal sequencing was, however, impossible because all these subunits were found to be resistant against Edman degradation, presumably due to N-terminal protection. MALDI-ReTOF-MS analyses of the tryptic fragments of the purified 47-kDa subunit revealed that this particular subunit was derived from the C-terminal domain of the 66-kDa precursor (Figs. 3a and 6; Table 1). The observed fragments could be attributed to the region between amino acid positions 141 and 545 (Fig. 3a; Table 1). These findings were corroborated by the results of the epitope mapping of the polyclonal antibody raised against this particular cocoa vicilin subunit (Fig. 5b, upper part). When the electrophoretically purified 47-kDa subunit was treated with carboxypeptidase Y, a rapid release of Phe, Leu, Arg, and Ala was measured, the liberation kinetics of Phe and Leu being faster than that of Arg and Ala (data not shown). These data indicated a C-terminal sequence Ala–Arg–Leu–PheCOOH (Table 4) corresponding to the C-terminal end of the 66-kDa precursor (McHenry & Fritz, 1992; Spencer & Hodge, 1992). The MALDI-ReTOF-MS data of the tryptic peptides of the 47-kDa subunit included a peak at *m/z* 1042.60, corresponding to the sequence NNPYYFPK located at position 141–148 of the precursor sequence (Table 1). Since the N-terminus of the 47-kDa subunit is N-terminally modified (see above), at least Arg140 must be included in the N-terminal sequence. A molecular mass of 48,500 Da was calculated for this sequence, which roughly corresponds to the experimentally found 47 kDa.

The MALDI-ReTOF-MS data obtained for the tryptic peptides of the 31-kDa subunit revealed that this particular cocoa vicilin subunit is derived from the precursor region between Arg131 and Arg425 (Table 2). These findings are in agreement with the results of the epitope mapping of the corresponding polyclonal antibody (Fig. 5b). The preferential release of Arg and Ser during carboxypeptidase Y treatment of the purified 31-kDa subunit (Table 4) corroborated the assumption that Arg425 is the carboxyterminal amino acid residue of this cocoa vicilin subunit (Fig. 6). A molecular

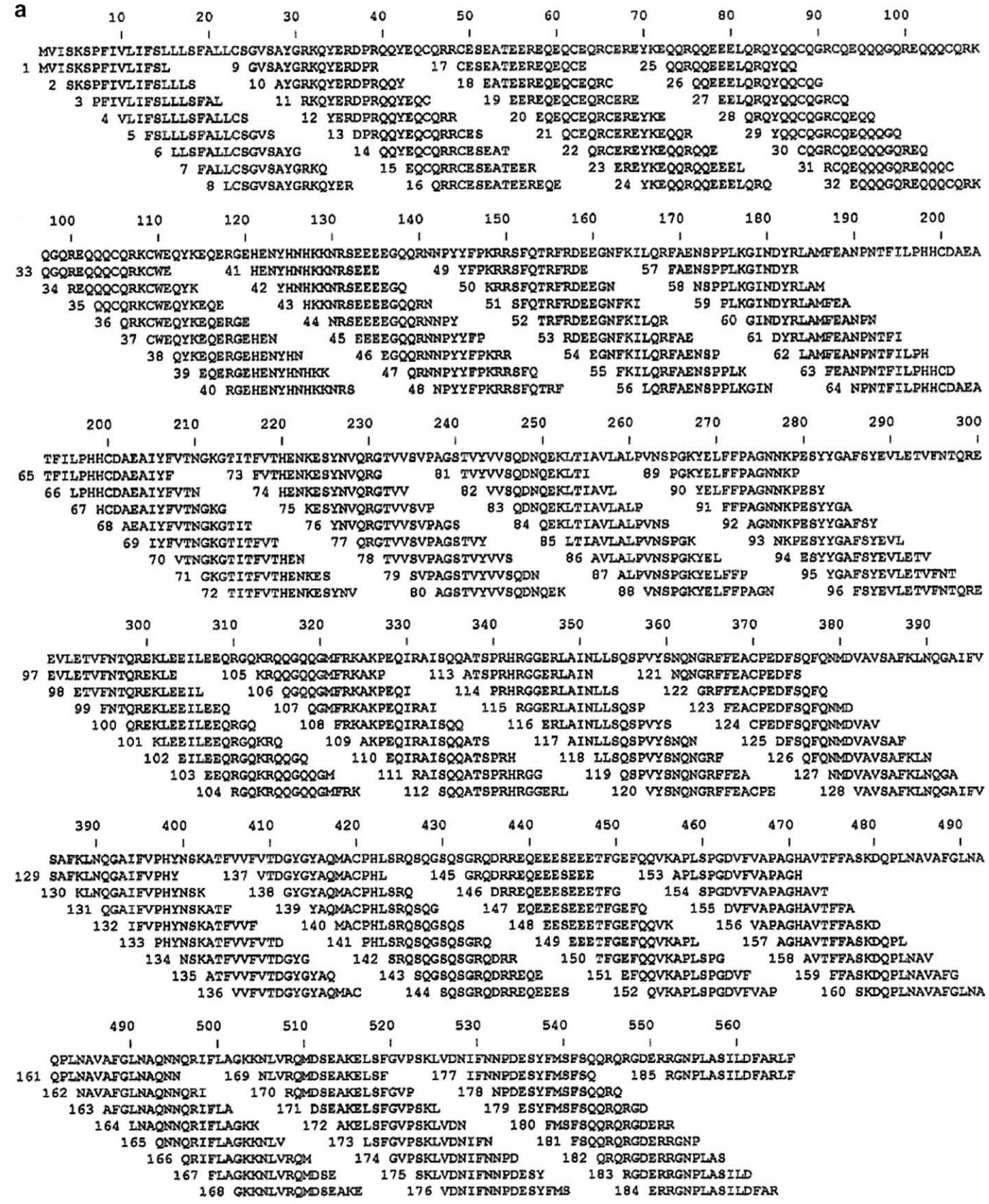

Fig. 5. Comparative epitope analyses of the polyclonal antibodies raised against the 47-kDa and 31-kDa subunits as well as the 28-kDa and 16-kDa proteolytic fragments of the cocoa vicilin. (a) 185 overlapping pentadeca-peptides derived from the cocoa vicilin precursor (GenBank accession number S22477) and representing the whole known amino acid sequence of this precursor were synthesised by spot synthesis using cellulose paper as solid support. (b) After incubation of the cellulose-bound scan peptides with different antibodies, bound IgGs were detected by incubation with alkaline phosphatase-coupled anti-rabbit-IgG Igs and subsequent visualisation of the indirectly bound alkaline phosphatase.

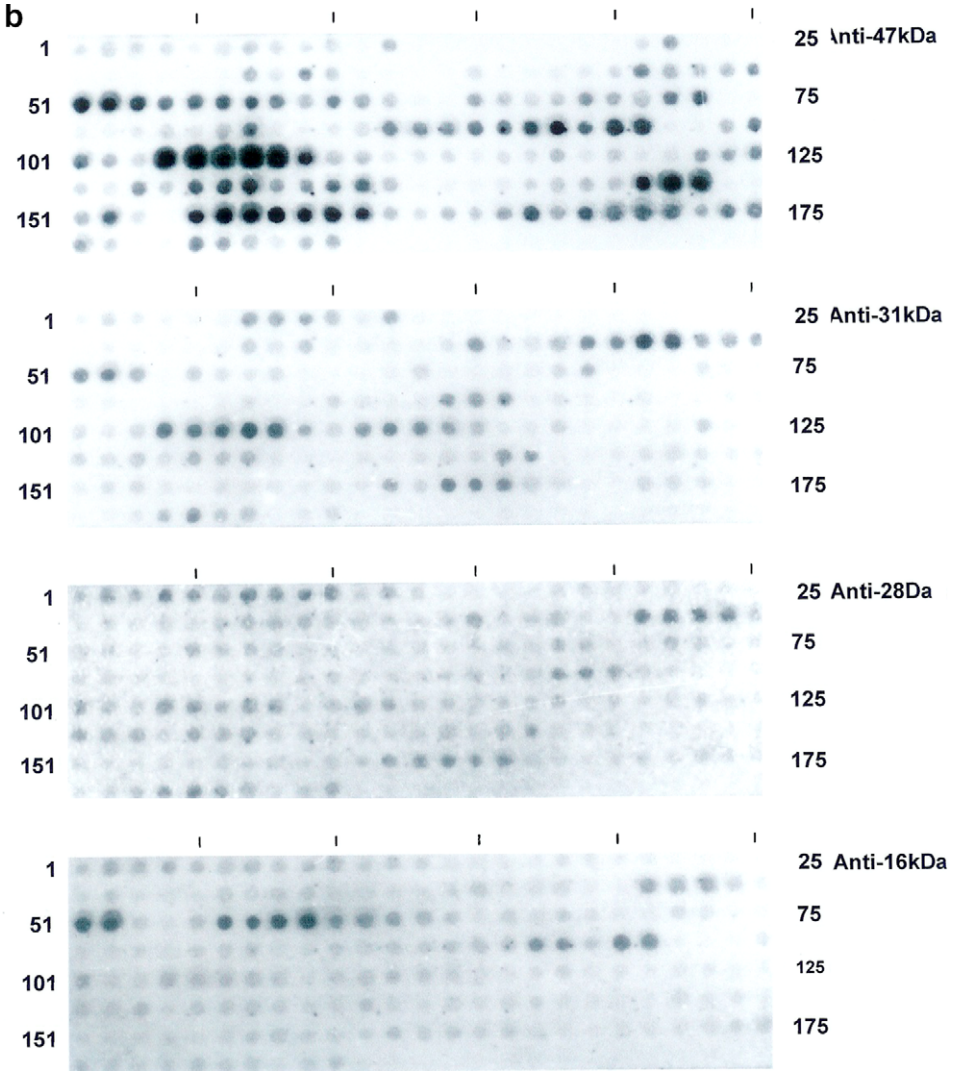

Fig. 5 (*continued*)

mass of 33 kDa was calculated for this sequence. Assuming that Arg131 is the N-terminal amino acid and that the N-terminus is modified because it was found to be resistant to Edman degradation, a pI value of 7.31 was calculated. This corresponds to the 31-kDa variant with the highest pI value (Fig. 2 and Table 5).

More complicated was, however, the interpretation of the MALDI-ReTOF-MS data of the tryptic peptides obtained from the 15-kDa subunit (Fig. 3c; Table 3). The most prominent peaks at *m/z* 1974.02, 1587.81, 1058.53, 934.41, and 841.48 (Fig. 3c) could be attributed to tryptic fragments localised between positions 315 and 434 of the 66-kDa precursor sequence (Table 3). However, minor peaks corresponded to tryptic peptides located either upstream or downstream of this sequence (Fig. 3c; Table 3). If all these fragments were derived from the same polypeptide, it should have a molecular mass considerably larger than 15 kDa. Therefore, we conclude, that the "15-kDa subunit" is a mixture of polypeptides of similar molecular masses rather than a unique 15-kDa subunit.

Overall, our data show that the cocoa vicilin subunits are derived from the precursor region between the amino acid positions 131 and 566 (Fig. 6). As described for the vicilins of other plants (Gatehouse, Lycett, Croy, & Boulter, 1982; Müntz, 1998), the largest subunit (47 kDa) covers almost the entire sequence of that region (Fig. 6). However, both the MALDI-ReTOF data (Fig. 3a and b; Tables 1 and 2) and the results of epitope mappings of the corresponding polyclonal antibodies (Fig. 5) indicated that the N-terminal end of the 31-kDa subunit is located upstream of the 47-kDa subunit (Fig. 6). Therefore, we assume that the 31-kDa subunit is not generated from the 47-kDa subunit but that both subunits are derived from a common intermediate by differential proteolytic processing. The formation of this intermediate should be due to the action of an asparagine-specific endoprotease cleaving the bond between Asn130 and Arg131. Such asparagine-specific endoproteases are involved in the proteolytic processing of storage proteins of other plants (Müntz & Shutov, 2002).

However, the 15.5-kDa, 15.0-kDa, and 14.5-kDa constituents were found to overlap with the 31-kDa subunit (Fig. 6) as revealed by the results of the MALDI-ReTOF-MS data (Tables 2 and 3) and further corroborated by the cross-reactions of these smaller polypeptides with the polyclonal antibody raised against the 31-kDa subunit (Fig. 4c; lane 1). The 15.0-kDa polypeptide revealed the most intense band with the antibody against the 31-kDa subunit (Fig. 4c; lane 1). Therefore, the most intense fragments observed by MALDI-ReTOF-MS (Fig. 3c; Table 3) should be derived from this particular polypeptide, which, therefore, should be derived from the precursor region between the amino acid positions 314 and

Table 4
Amino acid composition and carboxyterminal sequence of the 47-kDa and 31-kDa subunits of the cocoa vicilin

Amino acid	47-kDa subunit				31-kDa subunit			
	Found		Calculated[a]		Found		Calculated[b]	
	Mol%[c]	Number	Mol%	Number[a]	Mol%[c]	Number	Mol%	Number[b]
Ala (A)	7.85 ± 0.24	33.5	7.75	33	7.32 ± 0.24	21.5	7.14	21
Cys (C)	0.65 ± 0.03	2.8	0.70	3	0.98 ± 0.08	2.9	1.02	8
Asp (D)	10.29 ± 0.48	43.8	3.52	15	9.31 ± 0.46	27.4	2.38	7
Glu (E)	16.35 ± 0.71	69.7	7.75	33	16.74 ± 0.75	49.2	8.50	25
Phe (F)	8.35 ± 0.29	35.6	8.22	35	7.65 ± 0.19	22.5	7.48	22
Gly (G)	6.71 ± 0.25	28.6	6.57	28	6.70 ± 0.27	19.7	6.46	19
His (H)	1.56 ± 0.07	6.6	1.64	7	1.89 ± 0.12	5.6	2.04	6
Ile (I)	3.42 ± 0.18	20.7	3.29	14	3.80 ± 0.15	11.2	3.74	11
Lys (K)	4.86 ± 0.21	20.7	4.93	21	5.23 ± 0.32	15.4	5.10	15
Leu (L)	5.98 ± 0.37	25.5	5.87	25	5.16 ± 0.27	15.2	5.10	15
Met (M)	1.29 ± 0.11	5.5	1.41	6	1.19 ± 0.13	3.5	1.36	4
Asn (N)[d]	–	–	7.04	33	–	–	7.14	21
Pro (P)	5.83 ± 0.32	24.8	5.63	24	5.97 ± 0.23	17.6	5.78	17
Gln (Q)[d]	–	–	8.22	35	–	–	7.82	23
Arg (R)	6.09 ± 0.33	25.9	6.34	27	5.99 ± 0.31	17.6	6.12	18
Ser (S)	6.95 ± 0.39	29.6	7.28	31	6.13 ± 0.36	18.0	6.46	19
Thr (T)	3.63 ± 0.15	15.5	3.76	16	4.85 ± 0.25	13.5	4.76	14
Val (V)	6.86 ± 0.31	29.2	6.57	28	6.85 ± 0.27	20.1	6.80	20
Trp (W)	–	–	–	–	–	–	–	–
Try (Y)	3.33 ± 0.19	14.2	3.52	15	4.51 ± 0.34	13.3	4.76	14
C-terminus[e]	AlaArgLeuPheCOOH				SerArgCOOH			

[a] Calculated on the basis of the assumption that the 47-kDa subunit corresponds to positions 141–566 of the precursor sequence (426 amino acid residues).
[b] Calculated on the basis of the assumption that the 31-kDa subunit corresponds to positions 132–425 of the precursor sequence (294 amino acid residues).
[c] Given values are mean ± SD of six experiments.
[d] Asn and Gln are hydrolysed to Asp and Glu, respectively, during treatment with HCl.
[e] Determined by the kinetics of the release of free amino acids during treatment with carboxypeptidase Y.

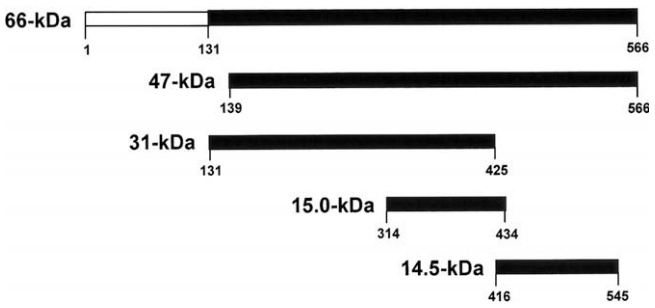

Fig. 6. Localisation of different subunits and proteolytic fragments within the cocoa vicilin precursor sequence. Open bar: sequence eliminated during proteolytic processing; black bars: sequences occurring in the mature vicilin.

Table 5
The isoelectric points of the 47-kDa and 31-kDa vicilin subunits and their variants

pI value	47-kDa subunit		31-kDa subunit	
	Found[a]	Calculated[b]	Found[a]	Calculated[c]
	6.42	6.79 (unmodified)	7.41	7.31 (unmodified)
	6.18	6.46 (1 N/Q → 1 D/E)	7.10	6.71 (1 N/Q → 1 D/E)
	5.98	6.22 (2 N/Q → 2 D/E)	6.80	6.37 (2 N/Q → 2 D/E)
	5.78	6.03 (3 N/Q → 3 D/E)	6.50	6.12 (3 N/Q → 3 D/E)
	5.62	5.86 (4 N/Q → 4 D/E)		
	5.47	5.71 (5 N/Q → 5 D/E)		
		5.58 (6 N/Q → 6 D/E)		
		5.46 (7 N/Q → 7 D/E)		

[a] Experimentally determined by comparison with pI protein standard mixture (Fig. 2).
[b] Calculated for the polypeptide chain covering the precursor sequence from amino acid position R140 to F566 assuming that the N-terminus is blocked.
[c] Calculated for the polypeptide chain covering the precursor sequence from amino acid positions R131 to R425 assuming that the N-terminus is blocked.

434. Either the minor 15.5- or the minor 14.5-component must be located downstream of this region (Table 3). Again, these results could be explained by the assumption that the 15.5-kDa, 15.0-kDa, and 14.5-kDa constituents are generated by differential proteolytic processing of the postulated intermediate (Fig. 6). In the case of the pea vicilin, however, the largest 50-kDa subunit was found to be the precursor of all the smaller constituents (Gatehouse et al., 1982; Müntz, 1998).

The two-dimensional gel electrophoresis (Fig. 2) revealed that both the 47-kDa and 31-kDa subunits as well as the 28-kDa fragment consisted of several subspecies differing with respect to their isoelectric points (Table 5). However, no indications for heterogeneous sequences of the 47-kDa and 31-kDa subunits were obtained by MALDI-ReTOF-MS of the corresponding tryptic peptides (Fig. 3a and b; Tables 1 and 2). Calculation of the theoretical pI values of both subunits revealed values which roughly corresponded to those experimentally determined for the subspecies with highest pI values (Table 5). The occurrence of subspecies with lower pI values could be explained by statistical transformation of Asn or Gln to Asp or Glu residues (Table 5) due to the action of a protein deaminase during maturation. Such an enzyme was previously found in germinating wheat grains (Vaintraub, Kotova, & Shaha, 1992).

5. Conclusions

Taken together, our results clearly show that the vicilin-like storage globulin of the cocoa seeds is encoded by a single gene rather than by a gene family and that the heterogeneity of its mature subunits observed by two-dimensional gel electrophoresis (Fig. 2) must be due to statistical post-translational modifications of various amino acid side chains. No peptide fragment was detected in the tryptic digests of the cocoa vicilin subunits by MALDI-ReTOF-MS analyses (Fig. 3; Tables 1–3), which differed from the expected fragment pattern of the published precursor sequence (McHenry & Fritz, 1992; Spencer & Hodge, 1992). These

findings are the prerequisite for the identification of those regions of the cocoa vicilin, from which the cocoa-specific aroma precursors are derived, which are generated by successive degradation of the cocoa vicilin by the aspartic endoprotease and the carboxypeptidase of the cocoa seeds *in vitro* (Bytof et al., 1995; Voigt & Biehl, 1995; Voigt et al., 1994a, 1994b) and during fermentation of cocoa seeds (Voigt & Biehl, 1995). The future identification and characterisation/sequencing of the aroma-relevant peptides will provide the basis for the development of analytical methods useful for the determination of raw cocoa quality. The present lack of such analytical methods is currently a hindrance for ensuring cocoa and chocolate quality.

Acknowledgements

The authors wish to thank Dr. S. Kamaruddin (Cocoa and Coconut Research Division, Hilir Perak, Malaysia, of the Malaysian Agricultural Research and Development Institute) for providing fresh cocoa pods and Mrs. R. Getzlaff and Dr. M. Kies (Helmholtz Centre for Infection Research, Braunschweig, Germany) for N-terminal sequencing of polypeptides.

References

Biehl, B., Brunner, E., Passern, D., Quesnel, V. C., & Adomako, D. (1985). Acidification, proteolysis and flavour potential in fermenting cocoa beans. *Journal of the Science of Food and Agriculture, 36*, 583–598.

Biehl, B., Heinrichs, H., Ziegeler-Berghausen, H., Srivastava, S., Xiong, O., Passern, D., et al. (1993). The proteases of ungerminated cocoa seeds and their role in the fermentation process. *Angewandte Botanik, 67*, 59–65.

Bradford, M. M. (1976). A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Analytical Biochemistry, 72*, 248–254.

Bytof, G., Biehl, B., Heinrichs, H., & Voigt, J. (1995). Specificity and stability of the carboxypeptidase activity in ripe, ungerminated seeds of *Theobroma cacao* L.. *Food Chemistry, 54*, 15–21.

Frank, R. (1992). Spot-synthesis: An easy technique for the positionally addressable, parallel chemical synthesis on a membrane support. *Tetrahedron, 48*, 9217–9232.

Gatehouse, J. A., Lycett, G. W., Croy, R. R., & Boulter, D. (1982). The post-translational proteolysis of the subunits of vicilin from pea (*Pisum sativum* L.). *Biochemical Journal, 207*, 629–632.

Guilloteau, M., Laloi, M., Michaux, S., Bucheli, P., & McCarthy, J. (2005). Identification and characterisation of the major aspartic proteinase activity in *Theobroma cacao* seeds. *Journal of the Science of Food and Agriculture, 85*, 549–562.

Hansen, C. E., del Olmo, M., & Burri, C. (1998). Enzyme activities in cocoa beans during fermentation. *Journal of the Science of Food and Agriculture, 77*, 273–281.

Hawkes, R., Nidag, E., & Gordon, J. (1982). A dot-immunobinding assay for monoclonal and other antibodies. *Analytical Biochemistry, 119*, 142–147.

Laemmli, U. K. (1970). Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature, 227*, 680–685.

Laloi, M., McCarthy, J., Morandi, O., Gysler, C., & Bucheli, P. (2002). Molecular and biochemical characterisation of two aspartic proteinases TcAP1 and TcAP2 from *Theobroma cacao* seeds. *Planta, 215*, 754–762.

McHenry, L., & Fritz, P. J. (1992). Comparison of the structure and nucleotide sequences of vicilin genes of cocoa and cotton raise questions about vicilin evolution. *Plant Molecular Biology, 18*, 1173–1176.

Müntz, K. (1998). Deposition of storage proteins. *Plant Molecular Biology, 38*, 77–99.

Müntz, K., & Shutov, A. D. (2002). Legumains and their functions in plants. *Trends in Plant Sciences, 7*, 340–344.

O'Farrell, P. H. (1975). High resolution two-dimensional electrophoresis of proteins. *Journal of Biological Chemistry, 250*, 4007–4021.

Perkins, D. N., Pappin, D. J., Creasy, D. M., & Cottrell, J. S. (1999). Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis, 20*, 3551–3567.

Rohan, T. A. (1964). The precursors of chocolate aroma: A comparative study of fermented and unfermented cocoa beans. *Journal of Food Science, 29*, 456–459.

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

Towbin, H., Staehelin, T., & Gordon, J. (1979). Electrophoretic transfer of proteins from polyacrylamide gels to nitrocellulose sheets: Procedure and some applications. *Proceedings of the National Academy of Sciences of the USA, 76*, 4350–4354.

Vaintraub, I. A., Kotova, L. V., & Shaha, R. (1992). Protein deamidase from germinating wheat grains. *FEBS Letters, 302*, 169–171.

Voigt, J. (1985). Extraction by lithium chloride of hydroxyproline-rich glycoproteins from intact cells of Chlamydomonas reinhardii. *Planta, 164*, 379–389.

Voigt, J., & Biehl, B. (1995). Precursors of the cocoa-specific aroma components are derived from the vicilin-class (7S) globulin of the cocoa seeds by proteolytic processing. *Botanica Acta, 108*, 283–289.

Voigt, J., Biehl, B., Heinrichs, H., Kamaruddin, S., Gaim Marsoner, G., & Hugi, A. (1994a). In-vitro formation of cocoa-specific aroma precursors: Aroma-related peptides generated from cocoa-seed protein by co-operation of an aspartic endoprotease and a carboxypeptidase. *Food Chemistry, 49*, 173–180.

Voigt, J., Biehl, B., & Kamaruddin, S. (1993). The major seed proteins of *Theobroma cacao* L.. *Food Chemistry, 47*, 145–151.

Voigt, J., Heinrichs, H., Voigt, G., & Biehl, B. (1994b). Cocoa-specific aroma precursors are generated by proteolytic digestion of the vicilin-like globulin of cocoa seeds. *Food Chemistry, 50*, 177–184.

Voigt, J., Kamaruddin, S., Heinrichs, H., Wrann, D., Senyuk, V. I., & Biehl, B. (1995). Developmental stage-dependent variation of the levels of globular storage protein and aspartic endoprotease during ripening and germination of *Theobroma cacao* L. seeds. *Journal of Plant Physiology, 145*, 299–307.

Voigt, J., Wöstemeyer, J., & Frank, R. (2007). The chaotrope-soluble glycoprotein GP2 is a precursor of the insoluble glycoprotein framework of the Chlamydomonas cell wall. *Journal of Biological Chemistry, 282*, 30381–30392.

Wilm, M., Shevchenko, A., Houthaeve, T., Breit, S., Schweigerer, L., Fotsis, T., et al. (1996). Femtomole sequencing of proteins from polyacrylamide gels by nano-electrospray mass spectrometry. *Nature, 379*, 466–469.

Yusep, I., Jinap, S., Jamilah, B., & Nazamid, S. (2002). Influence of carboxypeptidases on free amino acid, peptide and methylpyrazine contents of under-fermented cocoa beans. *Journal of the Science of Food and Agriculture, 82*, 1584–1592.