

Comparative Proteomical Analysis of Zygotic Embryo and Endosperm from *Coffea arabica* Seeds

LÍVIA L. KOSHINO, † CLARISSA P. GOMES, § LUCIANO P. SILVA, # MIRIAN T. S. EIRA, LA CARLOS BLOCH JR., OCTÁVIO L. FRANCO, S, AND ÂNGELA MEHTA*

Universidade de Brasília; Embrapa Recursos Genéticos e Biotecnologia, Embrapa Café; Centro de Análises Proteômicas e Bioquímicas, Programa de Pós-Graduação em Ciências Genômicas e Biotecnologia, Universidade Católica de Brasília, Brasília, Brazil and Universidade Federal de Juiz de Fora, Juiz de Fora, Minas Gerais, Brazil

During coffee seed development, proteins are predominantly deposited in cotyledons and in the endosperm. Reserve proteins of the 11S family are the most abundant globulins in coffee seeds, acting as a nitrogen source during roasting and guaranteeing flavor and aroma. The aim of the present study was to compare the protein profiles of endosperm and zygotic embryos of coffee seeds. Proteins were extracted from whole seed as well as from embryo and endosperm, separately. Total proteins were analyzed by two-dimensional electrophoresis (2-DE) followed by identification by mass spectrometry (MS). The most abundant spots observed in the gels of coffee seeds were excised, digested with trypsin, and identified by MS as subunits of the 11S globulin. Spots with identical pl and molecular masses were also observed in the protein profiles of coffee endosperm and embryo, indicating that 11S protein is also highly expressed in those tissues. Peptide sequence coverage of about 20% of the entire 11S globulin was obtained. Three other proteins were identified in the embryo and endosperm 2-DE profiles as a Cupin superfamily protein, an allergenic protein (Pru ar 1), exclusive to the endosperm 2D map, and a hypothetical protein, observed only in the zygotic embryo profile.

KEYWORDS: 2-DE; mass spectrometry; zygotic embryo; endosperm; Coffea arabica

INTRODUCTION

Coffee is one of the most important agricultural products worldwide and is cultivated in more than 60 countries. Brazil is responsible for more than one-third of the global coffee production and exportation and, together with Vietnam and Colombia, accounts for about 50% of the world production. The main cultivated coffee species are Coffea arabica and Coffea canephora. Several studies of protein/gene expression have been performed mostly related to flavor and aroma in these species (1-6).

Coffee seeds contain specialized storage tissues (cotyledons and endosperm) in which proteins are formed and deposited during maturation. After development, these proteins constitute a major seed protein fraction. Seed storage proteins comprise groups of multiple isoforms encoded by gene families, including

globulins, albumins, glutelins, and prolamins (7). These proteins have shown multiple functions such as storage (8) as well as a key role in the plant defense system (9, 10). Globulins constitute the most widely distributed group of seed proteins, which are stored as hexameric structures (11). In coffee, the 11S globulin has been extensively studied by using genomic and proteomic approaches (5, 12, 13). This protein is formed by two subunits linked by a disulfide bond: a large acid α subunit and a small basic β subunit derived from a single 11S precursor, which exists as multiple isoforms (7, 14). These abundant proteins are probably important for the quality of the beverage, because they act as the main source of nitrogen during roasting (15, 16).

Proteomic analysis has been increasingly employed to better understand biological processes. By using this approach, a high number of proteins expressed in a specific tissue or under certain biological conditions can be visualized, allowing the determination of quantitative data regarding protein abundance, as well as qualitative differences related to protein mass and isoelectric point (pI). Currently, proteomic analyses are being performed by coupling two-dimensional electrophoresis (2-DE) and highthroughput mass spectrometry (MS). In 2-DE, proteins are separated by protein mass and pI, and therefore a high resolution can be obtained. Protein spots can be further analyzed by MS and identified by peptide mass fingerprinting (PMF) or de novo

^{*} Address correspondence to this author at Embrapa Recursos Genéticos e Biotecnologia PqEB, Av. W/5 Norte Final, Asa Norte CEP 70770-900 Brasília, DF, Brazil (telephone +55 61 3448-4901; fax +55 61 3340-3658; email: amehta@cenargen.embrapa.br).

Universidade de Brasília.

[§] Universidade Católica de Brasília.

[#] Embrapa Recursos Genéticos e Biotecnologia.

[⊥] Embrapa Café.

[‡] Universidade Federal de Juiz de Fora.

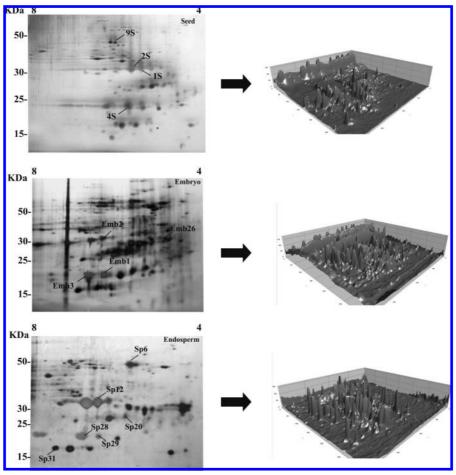


Figure 1. Two-dimensional electrophoresis (left) and three-dimensional gel representations (right) of proteins from whole seed, endosperm, and embryo, as indicated

sequencing. Peptide sequencing gives more detailed information and allows the identification of proteins from organisms with noncharacterized or nonpublic genomes, which is the case of the Coffee Genome database. Proteomic analyses have been applied for the study of coffee grains (5, 6, 17); however, still little is known regarding coffee seed proteins accumulated in different tissues, especially in the embryo. In summary, this paper aims to compare the global protein profiles of coffee zygotic embryo and endosperm by two-dimensional gel electrophoresis (2-DE) and mass spectrometry (MS).

MATERIALS AND METHODS

Plant Material. *C. arabica* (cultivar Catuaí Vermelho IAC-99) fruits were collected 210–220 days after anthesis, corresponding to the stage when the germinability is acquired (*13*, *18*). Seeds were isolated from the fruits and further separated into embryo and endosperm. Whole seeds were also used for comparison, and all tissues were stored at – 80 °C.

Protein Extraction. Whole seed, endosperm, and embryo samples collected from two consecutive years were separately homogenized in a liquid nitrogen-precooled mortar by using a pestle, and total proteins were extracted according to the method described by de Mot and Vanderleyden (19). Approximately 0.1 g of the resulting tissue powder was put into a 1.5 mL tube with 0.75 mL of extraction buffer consisting of 0.7 M sucrose, 0.5 M Tris-HCl, 30 mM HCl, 50 mM EDTA, 0.1 M KCl, and 40 mM DTT. An equal volume of buffer-saturated phenol (Invitrogen) was added, and after 15 min of vigorous shaking, the sample was centrifuged and the supernatant recovered. Two additional rounds of extraction were performed using an equal volume of extraction buffer each time. Proteins in the phenol phase were precipitated in 5 volumes of chilled 0.1 mM ammonium acetate in methanol at -20 °C for 2 h. Precipitates were washed in acetone and resuspended

in 40 μ L of lysis buffer [9.8 M urea, 0.2% (v/v) Nonidet P-40 (Sigma), 100 mM DTT, and 2% (v/v) of a mixture of ampholytes pH 5–8 and pH 3–10 (Bio-Rad) in the ratio of 5:1]. The total protein content was estimated according to the method of Bradford (20).

Two-Dimensional Electrophoresis. Isoelectric focusing of coffee proteins was performed according to the method of de Mot and Vanderleiden (19). Polyacrylamide gels containing 3.6% acrylamide, 0.21% bisacrylamide, 7.2% ampholyte pH 5-7 and 3-10 in the proportion of 5:1 (v/v), 2% Nonidet P-40, and 55% urea were used. Approximately 150 μ g of proteins was loaded onto the 11 cm gel after a prerun. Electrophoresis was performed in a vertical system (Invitrogen) at 400 V for 18 h, using 20 mM NaOH in the upper compartment of the chamber and 10 mM H_3PO_4 in the lower. Molecular mass separation was performed according to the method of Laemmli (21) in 18×18 cm gels 12%. Electrophoresis was performed for 6 h at 120 V using glycine buffer in the upper and lower compartments of the chamber, and at least four repetitions for each tissue were obtained. After running, 2D gels were fixed overnight in a solution containing 50% ethanol and 12% acetic acid. Silver staining was carried out according to the method of Blum et al. (22). Gels were digitalized using the scanner HP Scanjet model 8290 and further analyzed with Bionumerics software v. 4.5 (Applied-Maths). First, calibration with a gray scale was necessary to transform gray levels into values for each pixel of the gel picture. A calibration curve from Bionumerics software was used, and all gel pictures were analyzed as tiff files. The six gel images were placed in one folder, and the wizard detection method proposed by the software was used for spot detection. Automatically detected spots were manually checked, and some of them were manually added or removed according to size (>0.2 cm), format (circular), and density (>2 pixels cm⁻¹). Following the detection procedure, the normalization step was carried out to attribute a common protein identity for identical spots derived from different images. For this procedure, a reference gel was constructed, and automatic matching options of

Table 1. MS/MS Protein Identification of Spots from 2D Gels of Whole Seed, Embryo, and Endosperm

spot	peptide sequence	protein identification	accession no.
1S, Sp12	LNAQEPSFR GGQEGQR		
2S	LNAQEPSFR EGHQGQQQQHR		
4\$	TNDNAMIN AVEETLSSTVK IPILSSLQLSAER LQVVDHK IQVVDHK SALY VIR AIPEEVLR	11S globulin	AAC61983
9S, Emb 1, Emb 2, Emb 3, Sp28, Sp29	LSENIGLPQEADVFNPR		
Sp20	IPILSSLQLSAER LSENIGLPQEADVFNPR		
Sp6	FGVEEGDIFAVQR	Cupin family protein	NP_195388
Sp31	APVTSSYEVTFNIAPPR	major allergen Pru ar 1	O50001
Emb26	SSWNSPYYDTSSYGAG- SGGGGGGGR	hypothetical protein	EAZ14610

Bionumerics software was used. For each sample, when a protein was detected in all gel images, this protein was automatically added to the reference gel

Mass Spectrometry. Protein spots were excised from silver nitratestained gels, destained, and digested with trypsin according to the method of Shevchenko et al. (23). Peptides were analyzed in MALDI TOF-TOF Ultra Flex II (Bruker Daltonics) MS and *de novo* sequenced using CID and LIFT acquired MS/MS spectral data. The identification of protein spots was performed by sequence searches in public databases and in the Coffee Genome Project database using the Blastp and tBlastn programs, respectively.

RESULTS AND DISCUSSION

In this study, we have analyzed the proteomes of coffee seeds as well as zygotic embryo and endosperm. The main proteins observed in all three protein maps were identified by MS/MS experiments. Initially, the whole seed proteome was analyzed and revealed approximately 70 proteins ranging in mass from 10 to 100 kDa and in pI from 4 to 8. Some of the main proteins observed (1S, 2S, 4S, and 9S) were excised from the gel and analyzed by mass spectrometry. The proteins identified corresponded to the 11S globulin, previously described (5, 6, 12). To verify the proteins expressed in the different compartments of the whole seed, we analyzed the endosperm and zygotic embryo separately by 2-DE. Comparison of all three 2D maps showed similar protein profiles (Figure 1). The 2-DE of embryos and endosperm revealed approximately 120 and 80 protein spots, respectively, showing higher protein diversity in the embryo. Several of the proteins observed in the seed could be visualized in the embryo and endosperm profiles (Table 1). Some of these proteins observed in the embryo (Emb1, Emb2, and Emb3) and endosperm (Sp12, Sp20, Sp28, and Sp29) were also excised from the gel and analyzed by MS to determine whether the proteins corresponded to the seed globulin. The results revealed that all protein spots corresponded to the same globulin 11S protein.

By using a proteomic approach, Rogers et al. (6) previously characterized the 11S globulin in coffee endosperm mainly by

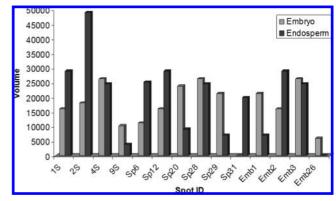


Figure 2. Expression comparison of spot volumes of identified proteins using the Bionumerics software.

NH₂ terminal sequencing. In this work, we have analyzed several protein spots by MS/MS analysis and obtained 11S protein coverage of approximately 20% of the whole protein, considering the number of amino acid residues determined (**Table 1**). Yuffá et al. (17) previously analyzed zygotic and somatic embryos as well as whole seed of *C. arabica* and showed the expression of abundant polypeptides in all three protein profiles; however, no protein identification was performed. In this study, we identified protein spots from zygotic embryo, and our results show that 11S globulin is also highly expressed in this tissue.

It has been reported that the 11S globulin represents approximately 45% of total proteins in the endosperm tissue and 5–7% of coffee bean dry weight (24). Endosperm is a very important specialized storage organ and, therefore, study of the endosperm proteome can provide information on seed development. Nevertheless, this tissue is rich in storage proteins such as globulins, which hinders the identification of poorly abundant proteins. Indeed, a total of 11 different protein spots analyzed in this study were identified as the 11S globulin. A recent analysis of the rice endosperm proteome was performed, and a method to remove the highly abundant proteins was developed (25). This is an important strategy to detect and identify new proteins in the endosperm.

In this study, differentially expressed proteins were also excised from the gel and analyzed by MS/MS; however most proteins did not reveal reliable identification, probably due to the low amount of protein or low ionization capacity. One protein spot (Sp6), observed in the embryo and up-regulated in the endosperm (**Figure 2**), was identified as a Cupin superfamily protein. This family includes 11S and 7S plant seed storage proteins and germins (26, 27), which have been associated with allergy (28, 29). In coffee, allergy to caffeine (30-32) and the beans (33-35) has been reported. However, no protein has been directly associated with coffee allergies.

Another protein identified was spot Sp31, which was expressed only in the endosperm (**Figure 2**) and showed identity to Mal d 1, the major apple allergen, which belongs to a group of pathogenesis-related (PR) 10 proteins (36). It has been reported that the expression of PR10 genes occurs during conditions of stress and ripening (37, 38). Proteins from this family have been implicated in allergy in different consumed fruits such as peach, cherry, and oranges (39-41). Although there are no reports associating these proteins with coffee allergy, additional studies need to be performed to investigate their role in allergy.

A protein spot (Emb26) specifically expressed in the zygotic embryo was identified and showed 100% identity to a hypothetical protein from *Oryza sativa*. Although the masses

of both proteins were similar (30 kDa), the calculated pI of this protein was 10 as opposed to 4.5 for the coffee protein. Unexpectedly, this protein did not show a high identity to any protein present in the Coffee Genome database. It is possible that this protein is a new one, not present in the Coffee and public databases.

Overall, in this work we have identified several protein spots in the whole seed, embryo, and endosperm of coffee as the 11S globulin. We show that this protein is highly expressed in the endosperm and embryo tissues. Moreover, two allergenic proteins were identified, as well as a hypothetical protein. Further studies need to be performed to determine the functional role of these proteins in development and their possible relationship to quality, flavor, and aroma, as well as allergy.

LITERATURE CITED

- (1) Lin, C.; Mueller, L. A.; McCarthy, J.; Crouzillat, D.; Pétiard, V.; Tanksley, S. D. Coffee and tomato share common gene repertoires as revealed by deep sequencing of seed and cherry transcripts. *Theor. Appl. Genet.* 2005, 112, 114–130.
- (2) Nunes, F. M.; Coimbra, M. A. Chemical characterization of the high-molecular-weight material extracted with hot water from green and roasted robusta coffees as affected by the degree of roast. J. Agric. Food Chem. 2002, 50 (24), 7046–7052.
- (3) Montavon, P.; Mauron, A. F.; Duruz, E. Changes in green coffee protein profiles during roasting. J. Agric. Food. Chem. 2003, 51, 2335–2343
- (4) Montovan, P.; Duruz, E.; Rumo, G.; Pratz, G. Evolution of green coffee protein profiles with maturation and relationship to coffee cup. *J. Agric. Food Chem.* 2003, 51, 2328–2334.
- (5) Marraccini, P.; Deshayes, A.; Pétiard, V.; Rogers, W. J. Molecular cloning of the complete 11S seed storage protein gene of *Coffea arabica* and promoter analysis in transgenic tobacco plants. *Plant Physiol. Biochem.* 1999, 37, 273–282.
- (6) Rogers, W. J.; Bezard, G.; Deshayes, A.; Meyer, I.; Petiard, V.; Marraccini, P. Biochemical and molecular characterization and expression of the 11S-type storage protein from *Coffea arabica* endosperm. *Plant Physiol. Biochem.* 1999, 37, 261–272.
- (7) Shewry, P. R. Plant storage proteins. Biol. Rev. 1995, 70, 375–426.
- (8) Sivakumar, S.; Franco, O. L.; Thayumanavan, B.; Murad, A. M.; Manickan, A.; Mohan, M.; Mridula, M. Cloning and structural analyses of an Indian little millet (*Panicum sumatrense*) zeinlike storage protein: implication for molecular assembly. *Biochemistry* (*Moscow*) 2006, 71, 1183–1191.
- (9) Pelegrini, P. B.; Noronha, E. F.; Muniz, M. A.; Vasconcelos, I. M.; Chiarello, M. D.; Oliveira, J. T.; Franco, O. L. An antifungal peptide from passion fruit (*Passiflora edulis*) seeds with similarities to 2S albumin proteins. *Biochim. Biophys. Acta* 2006, 1764 (6), 1141–1146.
- (10) Pelegrini, P. B.; Murad, A. M.; Silva, L. P.; Santos, R. C. P.; Tagliari, P. D.; Bloch, C.; Noronha, E. F.; Miller, R. N. G.; Franco, O. L. A novel glycine-rich peptide from guava (*Psidium guajava*) seeds with activity against gram-negative bacteria: novel insights on the control of hospital infections. *Peptides* 2008, 1271–1279.
- (11) De Dios-Alché, J.; Jiménez-López, J. C.; Wang, W.; Castro-López, A. J.; Rodríguez-García, M. I. Biochemical characterization and cellular localization of 11S type storage proteins in olive (*Olea europaea L.*) seeds. *J. Agric. Food Chem.* 2006, 54, 5562–5570.
- (12) Acuña, R.; Bassüner, R.; Beilinson, V.; Cortina, H.; Cadena-Gómez, G.; Montes, V.; Nielsen, N. C. Coffee seeds contain 11S storage proteins. *Physiol. Plant.* 1999, 105, 122–131.
- (13) De Castro, R. D.; Marraccini, P. Cytology, biochemistry and molecular changes during coffee fruit development. *Braz. J. Plant Physiol.* 2006, 18, 175–199.
- (14) Muntz, K. Deposition of storage proteins. *Plant Mol. Biol.* 1998, 38, 77–99.

- (15) Arnold, U.; Ludwig, E. Analysis of free amino acids in green coffee beans II. Changes of the amino acid content in arabica coffees in connection with post-harvest treatment. Z. Lebensm. Unters. Forsch. 1996, 203, 379–384.
- (16) Baltes, W. Vorstufen und entstehung von farbe und geschmark des kaffees. In 7th International Colloquium on the Chemistry of Coffee; ASIC: Paris, France, 1975; pp 91–108.
- (17) Yuffa, A. M.; De Garcia, E. G.; Nieto, M. S. Comparative study of protein electrophoretic patterns during embryogenesis in *Coffea arabica* cv. Catimor. *Plant Cell Rep.* **1994**, *13*, 197–202.
- (18) De Castro, R. D.; Estanislau, W. T.; Mesquita, P. R.; Hilhorst, H. W. A semente de café: desenvolvimento e perspectivas genômicas. In Simpósio de Pesquisa dos Cafés do Brasil 2. Vitória, Brasil; 2001 (CD-ROM).
- (19) De Mot, R.; Vanderleyden, J. Application of two dimensional protein analysis for strain fingerprinting and mutant analysis of *Azospirillum* species. *Can. J. Microbiol.* 1989, 35, 960–967.
- (20) 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.
- (21) Laemmli, U. K. Cleavage of structural proteins during the assembly of the head of the bacteriophage T4. *Nature* 1970, 227, 680–685.
- (22) Blum, H.; Beier, H.; Gross, H. J. Improved silver staining of plant protein: RNA and DNA in polyacrilamide gels. *Electrophoresis* 1987, 8, 93–98.
- (23) Shevchenko, A.; Wilm, M.; Vorm, O.; Mann, M. Mass spectrometric sequencing of proteins silver-stained polyacrylamide gels. Anal. Chem. 1996, 68, 850–858.
- (24) Clifford, M. N. Chemical and physical aspects of green coffee and coffee products. In *Coffee: Botany, Biochemistry and Produc*tion of Beans and Beverage; Clifford, M. N.; Wilson, K. C., Eds.; Croom Helm: London, U.K., 1987; pp 305–374.
- (25) Li, G.; Nallamilli, B. R.; Tan, F.; Peng, Z. Removal of high-abundance proteins for nuclear subproteome studies in rice (*Oryza sativa*) endosperm. *Electrophoresis* 2008, 29, 604–617.
- (26) Bäumlein, H.; Braun, H.; Kakhovskaya, I. A.; Shutov, A. D. Seed storage proteins of spermatophytes share a common ancestor with desiccation proteins of fungi. *J. Mol. Evol.* 1995, 41, 1070–1075.
- (27) Gane, P. J.; Dunwell, J. M.; Warwicker, J. Modelling based on the structure of vicilins predicts a histidine cluster in the active site of oxalate oxidase. *J. Mol. Evol.* 1998, 46, 488–493.
- (28) Mills, E. N.; Jenkins, J. A.; Alcocer, M. J.; Shewry, P. R. Structural, biological, and evolutionary relationships of plant food allergens sensitizing via the gastrointestinal tract. *Crit. Rev. Food Sci. Nutr.* 2004, 44, 379–407.
- (29) Palomares, O.; Cuesta-Herranz, J.; Vereda, A.; Sirvent, S.; Villalba, M.; Rodriguez, R. Isolation and identification of an 11S globulin as a new major allergen in mustard seeds. *Ann. Allergy Asthma Immunol.* 2005, 94, 586–592.
- (30) Caballero, T.; Garcia, C.; Pascual, C.; Diaz-Pena, J. M.; Ojeda, A. Urticaria induced by caffeine. J. Invest. Allergol. Clin. Immunol. 1993, 3, 160–162.
- (31) Pola, J.; Subiza, J.; Armentia, A.; Zapata, C.; Hinojosa, M.; Losada, E.; et al. Urticaria caused by caffeine. *Ann. Allergy* 1988, 60, 207–208.
- (32) Hinrichs, R.; Hunzelmann, N.; Ritzkowsky, A.; Zollner, T. M.; Krieg, T.; Scharffetter-Kochanek, K. Caffeine hypersensitivity. *Allergy* 2002, 57, 859–860.
- (33) Lehrer, S. B.; Karr, R. M.; Salvaggio, J. E. Extraction and analysis of coffee bean allergens. *Clin. Allergy* **1978**, *8*, 217–226.
- (34) Patussi, V.; De Zotti, R.; Riva, G.; Fiorito, A.; Larese, F. Allergic manifestations due to castor beans: an undue risk for the dock workers handling green coffee beans. *Med. Lav.* **1990**, *81*, 301–307
- (35) Romano, C.; Sulotto, F.; Piolatto, G.; Ciacco, C.; Capellaro, E.; Falagiani, P.; Constable, D. W.; Verga, A.; Scansetti, G. Factors related to the development of sensitization to green coffee and castor bean allergens among coffee workers. *Clin. Exp. Allergy* 1995, 25, 643–650.

- (36) Breiteneder, H.; Ebner, C. Molecular and biochemical classification of plant-derived food allergens. J. Allergy Clin. Immunol. 2000, 106, 27–36.
- (37) Atkinson, R. G.; Perry, J.; Matsui, T.; Ross, G. S.; Macrae, E. A. A stress-pathogenesis and allergen-related cDNA in apple fruit is also ripening-related. *N.Z. J. Crop. Hortic. Sci.* **1996**, *24*, 103–107
- (38) Puehringer, H.; Moll, D.; Hoffmann Sommergruber, K.; Watillon, B.; Katinger, H.; Laimer da Camara Machado, M. The promoter of an apple Ypr10 gene, encoding the major allergen Mal d 1, is stress- and pathogen-inducible. *Plant Sci.* **2000**, *152*, 35–50.
- (39) Ahrazem, O.; Ibanez, M. D.; Lopez-Torrejon, G.; Sanchez-Monge, R.; Sastre, J.; Lombardero, M.; Barber, D.; Salcedo, G. Lipid

- transfer proteins and allergy to oranges. *Int. Arch. Allergy Immunol.* **2005**, *137* (3), 201–210.
- (40) Gamboa, P. M.; Caceres, O.; Antepara, I.; Sanchez-Monge, R.; Ahrazem, O.; Salcedo, G.; Barber, D.; Lombardero, M.; Sanz, M. L. Two different profiles of peach allergy in the north of Spain. *Allergy* 2007, 62, 408–411.
- (41) Scheurer, S.; Metzner, K.; Haustein, D.; Vieths, S. Molecular cloning, expression and characterization of Pru a 1; the major cherry allergen. *Mol. Immunol.* 1997, 34, 619–629.

Received for review June 5, 2008. Revised manuscript received September 2, 2008. Accepted September 17, 2008.

JF801734M