Pure Appl. Chem., Vol. 82, No. 1, pp. 259–267, 2010.
doi:10.1351/PAC-CON-09-01-09

Proteome analysis of castor bean seeds*

Francisco A. P. Campos[1,‡], Fabio C. S. Nogueira[2], Kiara C. Cardoso[3], Gustavo C. L. Costa[3], Luiz E. V. Del Bem[3], Gilberto B. Domont[2], Marcio J. Da Silva[3], Raquel C. Moreira[4], Arlete A. Soares[1], and Tiago L. Jucá[1]

[1]Department of Biochemistry and Molecular Biology, Federal University of Ceará, Fortaleza, CE, Brazil; [2]Institute of Chemistry, Federal University of Rio de Janeiro, RJ, Brazil; [3]Center for Molecular Biology and Genetic Engineering, University of Campinas, SP, Brazil; [4]PETROBRAS-CENPES, Rio de Janeiro, RJ, Brazil

Abstract: Castor bean (*Ricinus communis* L.) seeds serve as raw material for the production of nonedible oil used in medicine and industry, whereas the presence of allergenic and toxic proteins in the residue left after oil extraction precludes the use of this protein-rich by-product in animal feeding. To better understand the enzymes involved in the biosynthesis and degradation of fatty acids and to identify proteins with toxic/anti-nutritional properties, extracts of developing and germinating seeds were prepared and prefractionated according to solubility properties of the proteins. An enriched plastid organelle fraction embracing mostly plastids and mitochondria was also prepared. Two-dimensional electrophoresis (2DE) reference maps of these fractions were obtained from which nearly 400 proteins were identified by matrix-assisted laser desorption ionization-time of flight-time of flight (MALDI-TOF-TOF) mass spectrometry after a search in a National Center for Biotechnology Information (NCBI) database and in an expressed sequence tag (EST) primary bank prepared from a cDNA library of developing seeds. These proteomics techniques resulted in the identification of several classes of seed reserve proteins such as 2S albumins, legumin-like and seed storage proteins, as well as other proteins of plastidial or mitochondrial functions and proteins involved in plant defense against biotic and abiotic stresses. It is expected that the collected data will facilitate the application of genetic techniques to improve the quality/profile of castor seed fatty acids, and pave the way for a rational approach to inactivate allergenic and toxic proteins, allowing the use of castor bean meal in animal feeding.

Keywords: castor bean; mass spectrometry; proteome analysis; *Ricinus communis*; seed proteins; two-dimensional electrophoresis.

INTRODUCTION

The castor bean (*Ricinus communis* L.) plant is cultivated for harvesting the unique oil produced in its seeds [1], which has a broad application in industry and is also used as raw material for the production of biodiesel. The residue remaining after the oil is extracted is not used as animal feed because it contains toxic proteins such as the ribosome-inactivating protein ricin and allergens [2]. The development of genotypes better suited for the production of biodiesel and devoided of toxic/allergenic proteins is

*Paper based on a presentation at the 13th International Biotechnology Symposium (IBS 2008): "Biotechnology for the Sustainability of Human Society", 12–17 October 2008, Dalian, China. Other presentations are published in this issue, pp. 1–347.

‡Corresponding author: E-mail: bioplant@ufc.br

being hampered by a lack of understanding of (a) the metabolic pathways and enzymes leading to the production of fatty acids during seed development and its degradation during germination and (b) the proteins that are deposited during seed development. Although the castor bean has long been used as a model for studying fatty acid metabolism in plants [3], there are still large gaps regarding the expression pattern of several genes related to fatty acid synthesis and degradation. Additionally, even though most of the toxicity and allergenicity of the residue obtained after extracting seed oil can be ascribed to the presence in high levels of ricin, one of the most toxic proteins found in nature, and of allergens belonging to the 2S albumins protein family, very few studies [4,5] have been conducted in order to ascertain the expression patterns of these proteins during seed development, as well as ascertaining the presence of other classes of toxic and/or allergenic seed proteins. Below we describe data obtained by two-dimensional electrophoresis (2DE) and mass spectrometry to address these questions.

cDNA LIBRARY OF DEVELOPING CASTOR SEEDS

To date, very few biochemical studies have been conducted with castor genotypes from Brazil, and therefore molecular data on Brazilian genotypes are scanty. For this reason, a cDNA library was constructed using a pool of total RNA extracted from developing seeds (cv. Nordestina) 20, 30, and 40 days after pollination (DAP). This library was used to generate a primary database of expressed sequence tags (ESTs) that was used for protein identification by mass spectrometry. The EST bank contains 6608 ESTs from 5'-ends of the cDNA clones, representing 2678 unique transcripts. EST lengths ranged from 100 to 1010 bp, averaging 774 bp (Fig. 1) before clustering. Among the most abundant transcripts (Table 1) are genes coding for seed reserve proteins such as 11S globulin, legumin-like protein, 2S albumin precursor, and other very abundant transcripts coding for lipase, ubiquitin, and metallothioneins. Additionally, the library contains several transcripts coding for proteins involved in fatty acid biosynthesis, notably FAH12 and FAD2, which code for oleate hydroxylase and oleate desaturase, respectively. The transcriptional profile of our library differs significantly from that obtained by Lu et al. [6]

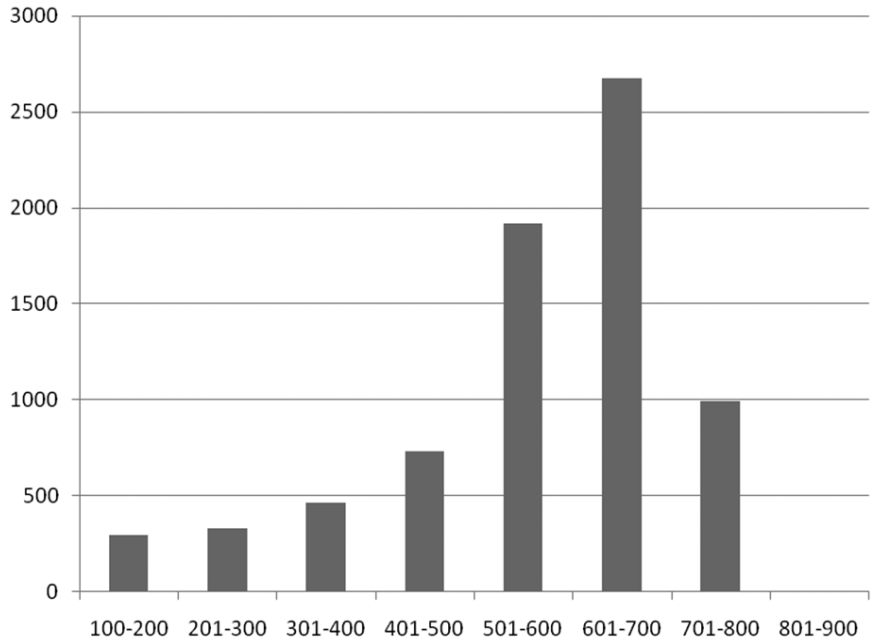

Fig. 1 Read length histogram (after quality and vector trimming) of ESTs from developing castor seeds.

by sequencing a cDNA library from developing castor seeds from an unnamed genotype. The availability of this database, together with that of the National Center for Biotechnology Information (NCBI) which comprised 112 235 nucleotide sequences in January 2009, allowed the identification of a great number of proteins from developing and mature castor bean seeds.

Table 1 The most abundant sequences from a full-length cDNA library of developing castor (*R. communis*) seeds.

Cluster ID	No. of ESTs	Functional description of gene product
Cn 580	624	11S globulin precursor isoform 2A
Cn 501	249	2SS_RICCO 2S albumin precursor
Cn 650	216	Legumin-like protein
Cn 234	204	Lipase
Cn 21	168	Cyclophilin
Cn 114	105	Ubiquitin
Cn 105	67	Seed storage protein
Cn 27	65	Esterase, putative
Cn 2	61	Legumin-like protein
Cn 1	56	2S albumin precursor
Cn 32	44	Annexin
Cn 128	40	60S acidic ribosomal protein
Cn 107	34	Cysteine protease Cp4
Cn 250	30	Cationic peroxidase
Cn 7	29	β-Tonoplast intrinsic protein
Cn 9	29	Metallothionein-like protein type 2
Cn 12	28	2S albumin precursor
Cn 118	25	Translation initiation factor
Cn 284	25	Thiazole biosynthetic enzyme, chloroplast
Cn 562	25	Unknown protein
Cn 160	24	Putative pectin methylesterase
Cn 172	23	α-Tubulin
Cn 315	23	Stearoyl acyl carrier protein desaturase
Cn 279	21	Unnamed protein product
Cn 487	21	Triosphosphate isomerase-like protein

TWO-DIMENSIONAL GEL PROTEOME REFERENCE MAPS OF DEVELOPING AND GERMINATING SEEDS

The morphological criteria established by Greenwood and Bewley [7] were used to classify the developmental stages of seeds. Total protein was extracted from each of the relevant stages according to Vasconcelos et al. [8] and Nogueira et al. [9]. These protein samples were used to prepare reference maps (Fig. 2). At Stage III, the deposition of seed reserve proteins begins, and at Stage V the protein pattern is indistinguishable from that of mature seeds. The maps also allowed the comparison of the deposition patterns of specific proteins. For example, as shown in Fig. 2, the deposition patterns of the three classes of seed reserve proteins that are present in castor seeds (2S albumins, legumin-like proteins, and seed storage proteins) have different deposition patterns. It remains to be established whether the deposition patterns of these proteins have any bearing on their physiological function in the seed.

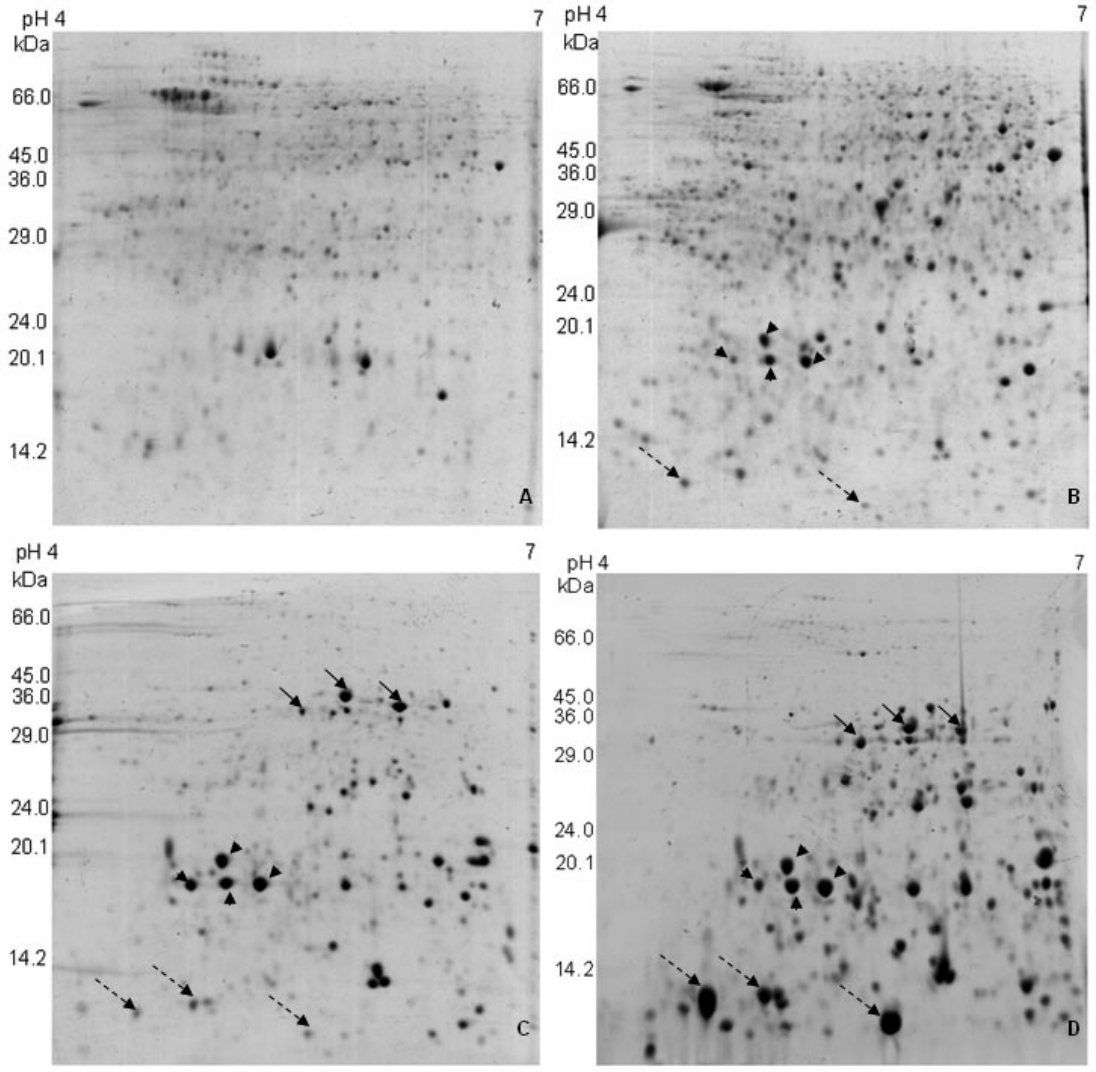

Fig. 2 2D reference map of developing castor seeds. Stages were assigned according to morphological criteria as defined by Greenwood and Bewley [5]. (A) Stage 2; (B) Stage 3; (C) Stage 5; (D) Mature seeds. The deposition pattern of selected storage proteins is indicated by dashed arrows (2S albumins), arrows (legumin-like proteins), and arrowheads (seed storage proteins). Protein load/gel was 300 μg.

Proteome reference maps of germinating seeds were also built, so that it was possible to ascertain the pattern of protein mobilization during germination (Fig. 3). Within three days of imbibition, a sharp increase in protein molibilization is observed, as indicated by the diminishing intensity of protein spots compared to those in quiescent seeds. It is interesting to note that after imbibition for 24 h, the mobilization of 2S albumins has already taken place, which is compatible with the hypothesis that these proteins act as a specialized storage of sulfur amino acids. Within eight days of imbibition, most of the endosperm proteins are already mobilized and the seed reserve proteins are no longer detected.

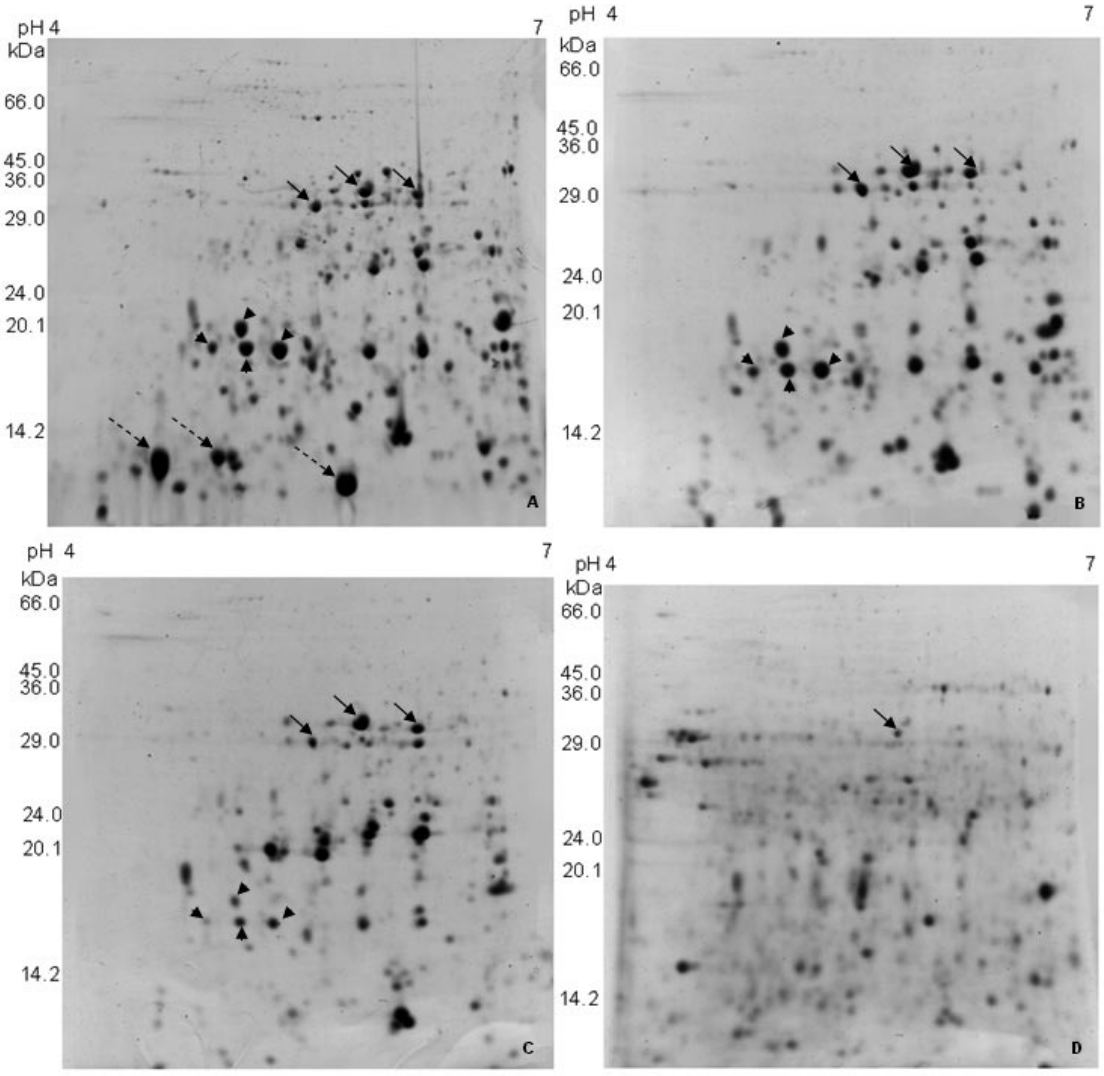

Fig. 3 Two-dimensional reference maps of castor bean seeds following water imbibition for 0 (A), 1 (B), 3 (C) and 8 (D) days. The mobilization patterns of selected seed reserve proteins are indicated by dashed arrows (2S albumins), arrows (legumin-like proteins) and arrowheads (seed storage proteins). Protein load/gel was 300 µg.

Attempts to identify proteins of spots excised from the reference maps shown in Fig. 2 indicate that they are mostly populated by a great diversity of seed reserve proteins belonging to three different families (Table 2). This lead to us to fractionate the castor seed endosperm proteins according to solubility properties, as performed previously Hurkman and Tanaka [10] for wheat endosperm proteins. So, four different protein fractions (albumins, globulins, CM proteins, and gliadins and glutenins) were prepared and subjected to two-dimensional gel electrophoresis (2D-GE) (Fig. 4).

Table 2 List of identified proteins in castor bean seeds. The reference maps from which the protein spot was taken is indicated.

Protein description	ID (NCBInr)	Reference map
Legumin-like protein	gi|8118510	Developing seed, globulins, CM fraction, gliadins/glutenins
Legumin B precursor	gi|126156	Developing seed
Legumin A2	gi|22008	Globulins
7S Globulin precursor	gi|118340979	Globulins
11S Globulin	gi|112676	Globulins
Seed storage protein	gi|8118512	Developing seeds, globulins, CM Fraction, gliadins/glutenins, plastid-enriched fraction
Ricin A	gi|225988	Developing seeds
Ricin A chain	gi|110288216	Albumins
Ricin D B	gi|224694	Globulins, albumins
Ricin D Ile chain	gi|223026	Albumins
Ricin precursor	gi|132567	Developing seeds, globulins, albumins
Agglutinin precursor (RCA)	gi|113504	Developing seeds, globulins, albumins
Agglutinin B	gi|225114	Globulins, albumins
Chain B, agglutinin	gi|58176691	Gliadins/glutenins
2S Albumin precursor	gi|112762	Developing seeds, globulins, Albumins, CM fraction and plastid-enriched fraction
Chain A, structure of Ricc3 (2S albumin)	gi|42543303	Developing seeds, globulins, Albumins, gliadins/glutenins
7.3321 kda Napin-like protein large chain	gi|1911801	Albumins
4.044 kda Napin-like protein small chain	gi|1911800	Globulins
48-kDa Glycoprotein precursor	gi|19338630	Developing seeds
Enolase	gi|1169534	Globulins
Aldehyde dehydrogenase	gi|8163730	Plastid-enriched fraction
Putative 3-keto-acyl-ACP dehydratase	gi|14334124	Plastid-enriched fraction
Acyl-CoA-binding protein	gi|3334112	CM fraction
Succinyl-CoA ligase (GDP-forming) beta-chain	gi|15225353	Plastid-enriched fraction
RuBisCO large subunit	gi|134101	Plastid-enriched fraction
ATP synthase subunit alpha	gi|114408	Plastid-enriched fraction
ATP synthase subunit beta	gi|114421	Plastid-enriched fraction
ATP synthase subunit beta	gi|231586	Plastid-enriched fraction
Heat shock protein 70	gi|1143427	Plastid-enriched fraction
Heat shock 70 kDa protein	gi|585273	Plastid-enriched fraction
Stromal 70 kDa heat shock-related protein	gi|399942	Plastid-enriched fraction
Chaperonin CPN60-1	gi|461735	Plastid enriched fraction
Calreticulin precursor	gi|11131745	Plastid-enriched fraction
Protein disulfide-isomerase precursor (PDI)	gi|11133775	Plastid-enriched fraction
Superoxide dismutase [Mn]	gi|464775	Developing seeds
Mitochondrial elongation factor Tu	gi|3928138	Plastid-enriched fraction
Late embryogenesis abundant protein	gi|18499	Globulins
Maturase K	gi|154813997	Globulins
Unknown protein	gi|14596223	Developing seeds
Unknown	gi|118481185	Plastid-enriched fraction
Unknown	gi|118488068	Plastid-enriched fraction
Unnamed protein product	gi|21078	Albumins
Hypothetical protein	gi|110740397	Developing seeds
Hypothetical protein	gi|147784261	Plastid-enriched fraction

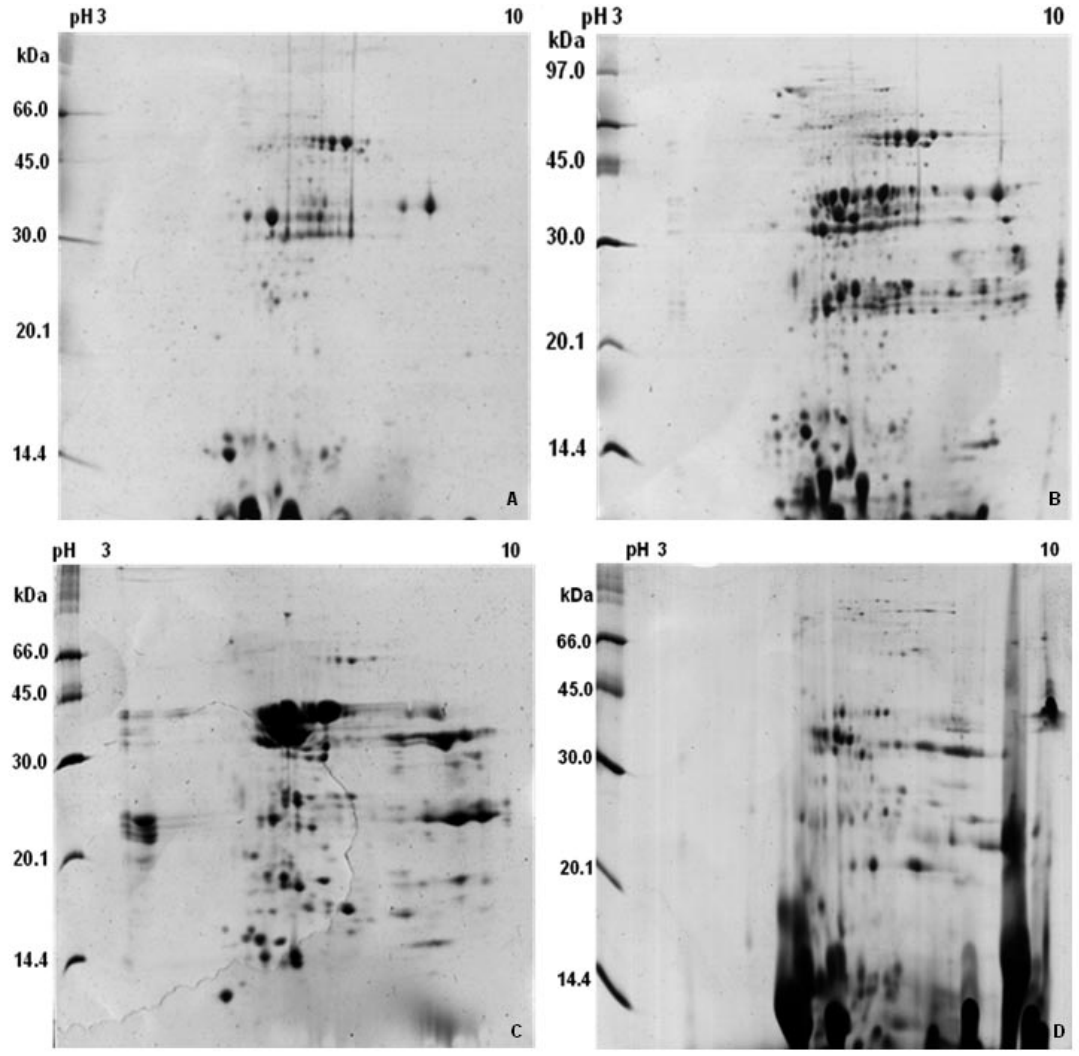

Fig. 4 2D reference maps of albumin (A), globulin (B), gliadin/glutenin (C), and CM (D) fractions of mature castor bean seeds. Protein load/gel was 250 µg.

Due to a special interest in the identification of proteins involved in fatty acid biosynthesis and taking into account that in castor seeds, as in most oil seeds, the synthesis of fatty acids take place in plastids [3,11], a fraction enriched in this organelle [12] was prepared. Transmission electron microscopy (TEM) showed that this fraction also contained mitochondria and small vacuoles (Fig. 5A). After total protein extraction according to Vasconcelos et al. [8] and Baba et al. [13], a proteome reference map between pH 4–7 and Mr range 6–97 kDa was produced (Fig. 5B). It showed approximately 1100 protein spots mostly with molecular mass greater than 40 kDa.

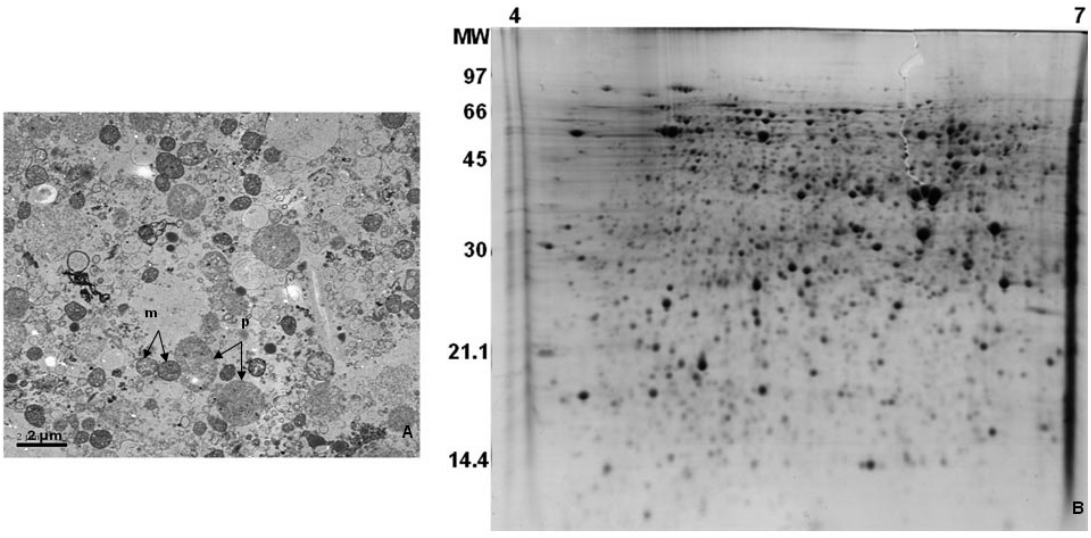

Fig. 5 Transmission electron micrograph (A) and reference map (B) of an organelle-enriched fraction isolated from developing castor bean seeds at stage 4. Protein load was 500 µg. m: mitochondria; p: plastids.

PROTEIN IDENTIFICATION

Selected spots from the reference maps were analyzed to evaluate the compatibility between protein extraction/2DE procedures and mass spectrometry as well as to disclose the protein classes that populate each reference map. Standard procedures [14] to analyze the protein spots by mass spectrometry were employed. Briefly, spots excised from Coomassie-stained polyacrylamide gels were destained and digested with sequencing-grade modified trypsin. Matrix-assisted laser desorption ionization-time of flight-time of flight (MALDI-TOF-TOF) tandem mass spectrometry spectra were acquired in an ABI 4700 Proteomics Analyzer (Applied Biosystems) using 3,5-dimethoxy-4-hydroxycinnamic acid as matrix and the resulting data by the GPS Explorer package (Applied Biosystems). Acquired mass spectral data were queried against both our EST and NCBI databases, using the MASCOT search engine. Search parameters were: mass accuracy of ±50 ppm and carboxyamidomethyl cysteine and methionine sulfoxide modifications. Homology search was performed against NCBI protein database using innitially "Viridiplantae taxa". Approximately 150 proteins deposited during the development of castor bean seeds from the maps shown in Fig. 2 were identified. Most of them encompass a wide range of pI and molecular weight and belong to one of the three classes of seed reserve proteins: 2S albumins (8 %), legumin-like proteins (48 %), and seed storage protein (17 %). This heterogeneity is consistent with the fact that seed reserve proteins are coded by multi-gene families [15]. Other proteins such as ricin and superoxide dismutase (Table 2) were also identified.

Because the protein population in the reference maps made it difficult to identify other less abundant proteins, different enriched fractions were prepared based only in protein solubility. The new reference maps (Fig. 4) were enriched in certain protein classes (Table 2). For example, the albumin fraction was almost devoid of legumin-like and seed storage proteins, but enriched in 2S albumins. The plastid-enriched fraction map (Fig. 5) shows more than 1100 resolved protein spots and has very significant differences from other maps. Most of the identified proteins have plastidial or mitochondrial functions, confirming the evidence gathered by TEM (Fig. 5A). Subcellular fractionation techniques are being developed [16] to obtain extracts of each purer organelle for further processing by proteomic techniques.

CONCLUDING REMARKS

High-quality and reproducible 2D reference maps of morphologically well defined stages of developing castor seeds and of germinating seeds were prepared. In order to obtain protein fractions enriched in less abundant proteins, castor seed extracts were fractionated according to solubility properties and maps of each fraction were obtained. A further refinement in the 2D mapping of castor seed proteins was achieved by preparing an organelle-enriched fraction which embraced plastids and mitochondria mostly. It was also demonstrated that the experimental condition used, namely, protein extraction and fractionation and 2D separation, are compatible with MALDI-TOF-TOF-tandem mass spectrometry for the processing and identification in the EST primary database prepared from a cDNA library of developing castor bean seeds. This combination of techniques resulted in the identification of more than 400 proteins from diverse classes. These results demonstrate the feasibility of using a 2D-based proteomics approach to elucidate the proteome of castor bean seeds. The availability of this proteome will facilitate the application of modern genetics-based techniques to improve the quality/profile of castor seed fatty acids, and by unraveling the proteins that have toxic/allergenic properties, paving the way for a rational use of the residue resulting from oil extraction in the diet of farm animals. Work is in progress toward these goals.

ACKNOWLEDGMENTS

We are grateful do FAPERJ, FUNCAP, PETROBRAS, CAPES, and CNPq (Brazil) for financial support.

REFERENCES

1. M. Sujatha, T. P. Reddy, M. J. Mahasi. *Biotechnol. Adv*. **26**, 424 (2008).
2. B. K. Kim. *Food Sci. Biotechnol*. **15**, 441 (2006).
3. I. A. Graham. *Annu. Rev. Plant Biol*. **59**, 115 (2008).
4. J.-Y. Ahn, G. Q. Chen. *J. Agric. Food Chem*. **55**, 10043 (2007)
5. S. V. S. Chakravartula, N. Guttarla. *Nat. Prod. Res*. **22**, 600 (2008).
6. C. Lu, J. G. Wallis, J. Browse. *BMC Plant Biol.* **7**, 42 (2007).
7. J. S. Greenwood, J. D. Bewley. *Can. J. Bot.* **60**, 1751 (1982).
8. E. A. R. Vasconcelos, F. C. S. Nogueira, E. F. M. Abreu, E. F. Gonçalves, P. A. S. Souza, F. A. P. Campos. *Chromatographia* **62**, 447 (2005).
9. F. C. S. Nogueira, E. F. Gonçalves, E. Jereissati, M. Santos, J. H. Costa, O. B. O. Neto, A. A. Soares, G. Domont, F. A. P. Campos. *Plant Cell Rep.* **26**, 1333 (2007).
10. W. J. Hurkman, C. K. Tanaka. *J. Chromatogr., B* **849**, 344 (2007).
11. T. A. McKeon, G. Q. Chen, J.-T. Lin. *Biochem. Soc. Trans.* **28**, 972 (2000).
12. S. A. Boyle, S. M. Hemmingsen, D. T. Dennis. *Plant Physiol.* **81**, 817 (1986).
13. A. I. Baba, F. C. S. Nogueira, C. B. Pinheiro, J. N. Brasil. E. S. Jereissati, T. L. Jucá, A. A. Soares, M. F. Santos, G. B. Domont, F. A. P. Campos. *Plant Sci.* **175**, 717 (2008).
14. M. Richardson, A. M. C. Pimenta, M. P. Bemquerer, M. M. Santoro, P. S. L. Beirão, M. E. Lima, S. G. Figueiredo, C. Bloch Jr., F. A. P. Campos, E. A. R. Vasconcelos, P. C. Gomes, M. N. Cordeiro. *Comp. Biochem. Phys. B* **142**, 173 (2006).
15. P. R. Shewry. *Plant Cell* **7**, 945 (1995).
16. D. J. Gauthier, C. Lazure. *Expert Rev. Proteomics* **5**, 603 (2008).

