



Research article

Mo-CBP₃, a 2S albumin from *Moringa oleifera*, is a complex mixture of isoforms that arise from different post-translational modifications

José E.C. Freire^a, Frederico B.M.B. Moreno^b, José E. Monteiro-Júnior^c, Antônio J.S. Sousa^a, Ilka M. Vasconcelos^a, José T.A. Oliveira^a, Ana C.O. Monteiro-Moreira^b, Bruno A.M. Rocha^a, Thalles B. Grangeiro^{c,*}

^a Departamento de Bioquímica e Biologia Molecular, Centro de Ciências, Universidade Federal do Ceará (UFC), Fortaleza, CE, Brazil

^b Núcleo de Biologia Experimental, Centro de Ciências da Saúde, Universidade de Fortaleza, Fortaleza, CE, 60810-431, Brazil

^c Departamento de Biologia, Centro de Ciências, UFC, Fortaleza, CE, Brazil

ARTICLE INFO

Keywords:

Seed storage proteins

Post-translational modifications

Prolamin superfamily

ABSTRACT

Mo-CBP₃ is a chitin-binding 2S albumin from *Moringa oleifera*. This seed storage protein is resistant to thermal denaturation and shows biological activities that might be of practical use, such as antifungal properties against *Candida* sp., a pathogen that causes candidiasis, and against *Fusarium solani*, a soil fungus that can cause diseases in plants and humans. Previous work has demonstrated that Mo-CBP₃ is a mixture of isoforms encoded by members of a small multigene family. Mature Mo-CBP₃ is a small protein (~14 kDa), constituted by a small chain of approximately 4 kDa and a large chain of 8 kDa, which are held together by disulfide bridges. However, a more comprehensive picture on the spectrum of Mo-CBP₃ isoforms which are found in mature seeds, is still lacking. In this work, genomic DNA fragments were obtained from *M. oleifera* leaves, cloned and completely sequenced, thus revealing new genes encoding Mo-CBP₃. Moreover, mass spectrometry analysis showed that the mature protein is a complex mixture of isoforms with a remarkable number of molecular mass variants. Using computational predictions and calculations, most (~86%) of the experimentally determined masses were assigned to amino acid sequences deduced from DNA fragments. The results suggested that the complex mixture of Mo-CBP₃ isoforms originates from proteins encoded by closely related genes, whose products undergo different combinations of distinct post-translational modifications, including cleavage at the N- and C-terminal ends of both subunits, cyclization of N-terminal Gln, as well as Pro hydroxylation, Ser/Thr phosphorylation, and Met oxidation.

1. Introduction

Seed storage proteins accumulate at high levels during seed development, and once the seed has germinated, these proteins are mobilized to provide amino acid skeletons to support early seedling growth. The major classes of seed storage proteins are 2S albumins, prolamins, and globulins, which are represented by 11-12S legumin-like and 7-8S vicilin-like proteins (Shewry et al., 1995). 2S albumins are members of the prolamin superfamily, which includes α -amylase inhibitors, bi-functional trypsin/ α -amylase inhibitors, non-specific lipid transfer proteins (nsLTPs) and other seed storage proteins, such as γ -gliadin, prolamin and related proteins (Breiteneder and Radauer, 2004). A growing number of evidences has suggested that, besides their fundamental role as nutrient sources, storage proteins may also be involved in other biological processes. For example, 11S legumins bind the

phytohormone indole-3-acetic acid and may play a role in auxin homeostasis (Kumar et al., 2017). Non-specific LTPs are abundantly found in land plants, but they are absent in chlorophyte and charophyte green algae and any other organisms (Edstam et al., 2011). These proteins are expressed in all tissues of the plant, at every developmental stage, and they are implicated in diverse biological functions, such as biosynthesis and accumulation of cuticular wax, root suberin and sporopollenin, root nodule symbiosis, cell-wall loosening, signalling during systemic acquired resistance and adaptation to abiotic stresses (Salminen et al., 2016). 2S albumins have also evolved other functions, such as DNase and RNase activities, which are thought to be involved in antibacterial and antifungal effects exhibited by some of these proteins (Tomar et al., 2014).

Mo-CBP₃ was first described as a chitin-binding protein (CBP), purified from the seeds of the drumstick tree (*Moringa oleifera*) (Gifoni

* Corresponding author. Departamento de Biologia, Centro de Ciências, Campus do Pici, Universidade Federal do Ceará, Fortaleza, Ceará, 60440-900, Brazil.

E-mail address: thalles@ufc.br (T.B. Grangeiro).

<https://doi.org/10.1016/j.plaphy.2019.05.003>

Received 1 February 2019; Received in revised form 1 May 2019; Accepted 2 May 2019

Available online 03 May 2019

0981-9428/ © 2019 Elsevier Masson SAS. All rights reserved.

et al., 2012). It is a small (~14 kDa), thermostable protein, which showed antifungal activity against mycelial fungi, such as *Fusarium solani*, an ascomycete that causes plant diseases but that is also implicated in human infections (Gifoni et al., 2012). The antifungal mechanism of Mo-CBP₃ against *F. solani* involves binding to cell wall, disorganization of the plasma membrane and induction of reactive oxygen species (ROS), causing inhibition of spore germination and mycelial growth (Batista et al., 2014). Moreover, synthetic peptides designed based on Mo-CBP₃ sequence inhibited the growth of pathogenic species of *Candida* as well as exhibited antibacterial activity against *Staphylococcus aureus* (Oliveira et al., 2019). These findings have raised the possibility to use Mo-CBP₃ as a tool to develop new antifungal and antibacterial drugs or to engineer crops with enhanced resistance to fungal diseases.

cDNA cloning has demonstrated that Mo-CBP₃ is indeed a 2S albumin, composed of a small chain of approximately 4 kDa and a large chain of 8 kDa, which are held together by disulfide bridges (Freire et al., 2015). These authors have also shown that at least 4 closely related genes (named *Mo-CBP₃₋₁*, *Mo-CBP₃₋₂*, *Mo-CBP₃₋₃* and *Mo-CBP₃₋₄*) encoded Mo-CBP₃, which was presumed to exist as a mixture of isoforms synthesized as preproteins undergoing proteolytic cleavages, as observed in typical two-chain 2S albumins from other species (Mylyne et al., 2014). *Mo-CBP₃* transcripts accumulate at the latter stages of seed development, when the protein is localized in storage vacuoles of cotyledonary cells (Garcia et al., 2019). Furthermore, the X-ray crystal structure of one of the isoforms of Mo-CBP₃ has revealed that it adopts the characteristic fold of 2S albumins, comprising 5 α -helical segments connected by short loops (Ullah et al., 2015).

Many proteins are produced as heterogeneous mixtures of isoforms, which can arise from the transcription of different members of multi-gene families or through alternative splicing events (Gelhay et al., 2005; Shang et al., 2017). Moreover, isoforms can also be produced from post-translational modifications (PTMs), including glycosylation, phosphorylation and proteolytic processing, and these PTMs play important roles in proteins' function (Cho et al., 2016). Isoforms can have distinct biological activities, a phenomenon that has important implications. One interesting example is phytohemagglutinin (PHA), a lectin from the seeds of red kidney bean (*Phaseolus vulgaris*). Two closely related genes (PHA-E and PHA-L) encode two polypeptides, E and L, which form tetramers consisting of all possible combinations of these two subunits (E4, E3L, E2L2, EL3 and L4) (Voelker et al., 1986). Isoform L4 is a potent leucoagglutinin and mitogen, but lacks hemagglutinating activity, whereas E4 is a potent hemagglutinin, which has little or no mitogenic activity (Liener et al., 1986).

Possible biomedical applications of Mo-CBP₃ will require the production of highly pure samples, which may be achieved through its heterologous expression in bacterial or yeast cells, for example. However, it remains to be determined if Mo-CBP₃ isoforms have differences in their biological activities. Therefore, a detailed knowledge on the types and differences that exist among its isoforms is a fundamental question. In this work, genomic DNA sequences encoding Mo-CBP₃ were obtained, and mass spectrometry analysis was performed on chromatographic fractions containing the small and large chain polypeptides that constitute the mature protein. These analyses aimed to provide a better description about the isoforms of Mo-CBP₃.

2. Materials and methods

2.1. Plant material

Mature seeds of *Moringa oleifera* (90 days after anthesis) were harvested in March 2017 from trees growing at the Campus do Pici, Fortaleza, CE, Brazil. Voucher specimens (EAC 54112) were deposited at the Herbário Prisco Bezerra, UFC. Seeds were sterilized with 0.2% sodium hypochlorite for 5 min, washed 3 times with distilled water, sown between germitest paper sheets soaked with distilled water, and

incubated at 25–28 °C. Once the seeds had germinated, they were transferred to 400 mL plastic pots containing sterile sand and irrigated daily with 1/5 strength Hoagland's nutrient solution (Hoagland and Arnon, 1950), under greenhouse conditions. Leaves of 30-day-old plants were harvested, frozen in liquid nitrogen and stored at –80 °C until used (Fig. S1).

2.2. Plasmid, bacterial strain and reagents

The plasmid pGEM-T Easy was purchased from Promega (Madison, WI, USA), whereas the *Escherichia coli* cloning strain TOP10F' was from Invitrogen (Carlsbad, CA, USA). All other reagents were of analytical grade.

2.3. DNA extraction, amplification, cloning of PCR products, and DNA sequencing

Total genomic DNA (gDNA) was purified from *M. oleifera* leaves of 30-day-old plants using a CTAB-based protocol, as previously described (Warner, 1996). The integrity of the DNA samples was checked by 1.0% agarose gel electrophoresis and the yield was estimated by measuring the absorbance at 260 nm (Sambrook et al., 1989). Three forward and three reverse primers were designed (Table S1), based on the *Mo-CBP₃* cDNA sequences previously determined (Freire et al., 2015). The primers were used in different combinations to amplify by polymerase chain reaction (PCR) genomic DNA sequences encoding Mo-CBP₃. Amplification reactions, cloning of PCR products and DNA sequencing were performed essentially as described elsewhere (Freire et al., 2015).

2.4. Sequence and phylogenetic analysis, and prediction of post-translational modifications

The Translate tool, available at the ExPASy Proteomics Server (web.expasy.org/translate/), was used to translate DNA sequences to protein sequences (Gasteiger et al., 2003). Molecular masses were calculated from amino acid sequences using ExPASy's Compute pI/Mw tool (web.expasy.org/compute_pi/). The presence of signal peptides and probable cleavage sites were predicted using SignalP 4.1 server (Petersen et al., 2011). Subcellular localization was predicted using EuLoc (Chang et al., 2013), available at the program's web server (<http://euloc.mbc.nctu.edu.tw/index.html>). Multiple alignments of DNA and amino acid sequences were usually performed using the program ClustalW (Thompson et al., 1994) implemented with the BioEdit 7.2.5 software package (Hall, 1999). Searches for homologous proteins in public sequence databases were performed using BLASTp (Altschul et al., 1990). Phylogenetic analysis was performed using the Neighbor-Joining method (Saitou and Nei, 1987). The evolutionary distances were computed from protein alignments (sites containing gaps were excluded) using the JTT matrix-based method (Jones et al., 1992), and clusters stability was evaluated by the bootstrap test (1000 replicates) (Felsenstein, 1985). The analyses were conducted in MEGA X (Kumar et al., 2018). The presence and delimitation of protein domains was performed by searching the NCBI Conserved Domain Database (CDD) through the CD-Search web service (Marchler-Bauer and Bryant, 2004). NetPhos 3.1 server (<http://www.cbs.dtu.dk/services/NetPhos/>) was used to predict serine, threonine or tyrosine phosphorylation sites (Blom et al., 1999). Prediction of hydroxyproline in protein sequences was performed using iHyd-PseAAC (Xu et al., 2014) through the program's web-server (app.aporc.org/iHyd-PseAAC/).

2.5. Purification of Mo-CBP₃

Mo-CBP₃ was purified from crude extracts of mature *M. oleifera* seeds (90 days after anthesis) using affinity chromatography on a chitin matrix followed by cation exchange chromatography on a Resource S matrix (GE Healthcare, Buckinghamshire, UK) as previously described

(Gifoni et al., 2012). The purity of the protein samples was determined by tricine-SDS-polyacrylamide gel electrophoresis (tricine-SDS-PAGE) according to a previously described method (Schägger and von Jagow, 1987). Staining and destaining of protein bands were performed as described earlier (Freire et al., 2015).

2.6. Fractionation of the small and large chains of Mo-CBP₃ and mass spectrometry analysis

To fractionate the small and large chains of Mo-CBP₃, protein samples were first reduced with dithiothreitol (DTT) and alkylated with iodoacetamide, as described before (Freire et al., 2015). Reduced and alkylated samples were then subjected to reversed phase-HPLC using a μ RPC C2/C18 ST 4.6/100 column (GE Healthcare, Buckinghamshire, UK), under room temperature. Running conditions were 100% buffer A [5% (v/v) acetonitrile with 0.05% (v/v) trifluoroacetic acid (TFA)] for 15 min, followed by a linear increase in buffer B [90% (v/v) acetonitrile with 0.05% TFA] to 50% for 50 min. Chromatography was performed at a constant flow rate (0.2 mL/min) and 0.5 mL fractions were collected. Elution of proteins was monitored by determining fractions' absorbance at 216 nm. Aliquots of chromatographic fractions were then subjected to electrospray ionization mass spectrometry (ESI-MS) analysis, using a Synapt G1 HDMS mass spectrometer (Waters Co., Milford, MA, USA) coupled to a nanoUPLC system. Data acquisition and deconvolution of MS spectra were done as described by Neto et al. (2017). When matching molecular masses determined by ESI-MS to those calculated from predicted amino acid sequences, as deduced from gDNA or cDNA fragments, a tolerance of ± 2 Da was acceptable. Theoretical molecular masses were calculated from each amino acid sequence, with carboxyamidomethylation of Cys residues (cam-Cys) included as a fixed modification (we assumed that this modification increased the polypeptide molecular mass by 57,02146 Da per Cys residue in each amino acid sequence). Moreover, oxidation of Met residues, cyclization of N-terminal Gln to pyroglutamate (pGlu), phosphorylation of Ser, Thr or Tyr residues, and hydroxylation of Pro or Lys residues, were also considered as variable modifications. These modifications were predicted to change the molecular masses of the polypeptide (per modified residue) by the following values: +15.99491 Da (oxidation of Met), +79.96633 Da (phosphorylation of Ser, Thr or Tyr), +15.99491 Da (hydroxylation of Pro or Lys) and -17.02655 Da (cyclization of N-terminal Gln, leading to pGlu).

3. Results and discussion

3.1. Cloning of genomic DNA sequences encoding new isoforms of Mo-CBP₃

Using different combinations of oligonucleotide primers (Table S1), five fragments of genomic DNA (~630–700 bp) encoding Mo-CBP₃ were amplified by PCR from *Moringa oleifera* (Fig. S2). Several clones were sequenced, and 6 unique DNA sequences were found (GenBank accession numbers: MH000615–MH000620). The length of the sequences ranged from 658 to 704 nucleotides, and each DNA clone encoded a polypeptide chain, which had 160 (3 sequences) or 163 (3 fragments) amino acid residues (Fig. 1A and Figs. S3–S10). None of the genomic DNA fragments that were sequenced contained introns, which is a characteristic feature of many 2S albumin genes, like those of *Arabidopsis thaliana* (Krebbers et al., 1988) and *Brassica napus* (Scofield and Crouch, 1987). The average percentage of coding sequence identity among the aligned Mo-CBP₃ DNA sequences was approximately 84.9%, ranging from 77.2% (114 different nucleotides) to 99.7% (one difference) (Table S2). Pairwise comparisons between the amino acid sequences deduced from genomic DNA sequences revealed an average sequence identity of 80.5%, varying from 70.6% (49 different residues) to 99.3% (a single difference) (Table S3). A 20-residues N-terminal signal peptide was confidently predicted in each amino acid sequence (SignalP D-score values ranging from 0.868 to 0.942) (Fig. S11).

Indeed, 2S albumins and other seed storage proteins are characteristically synthesized as preproteins, with an N-terminal signal peptide that is cleaved off co-translationally as the nascent polypeptides are translocated into the lumen of the endoplasmic reticulum (ER) (Hara-Hishimura et al., 1993). Once in the ER, 2S albumins are subjected to post-translational processing and then transported to protein storage vacuoles (PSVs) via dense vesicles, in a Golgi-dependent route (Jolliffe et al., 2004). Transport of 2S albumins and other storage proteins to PSVs requires a vacuolar sorting determinant, which is part of the protein's amino acid sequence. The vacuolar sorting signal can be present at the N- or C-terminal end of the polypeptide or even in the middle of the molecule (Pereira et al., 2014). When the deduced amino acid sequences of the proproteins (excluding the signal peptide) encoded by the Mo-CBP₃ genes were analyzed by the program EuLoc, all of them were predicted to be vacuolar proteins. Corroborating these predictions, Garcia et al. (2019) have recently demonstrated, using *in situ* immunolocalization techniques, that Mo-CBP₃ is localized in PSVs of cotyledonary cells from seeds at the latter stages of development.

BLAST searches against NCBI protein sequence databases showed that the primary structures deduced from the genomic DNA fragments of *M. oleifera* (Moringaceae) had highest sequence identity with 2S seed storage proteins from *Capparis masakai* (Capparaceae), *Tarenaya hassleriana* (Cleomaceae), *Arabis alpine* (Brassicaceae) and other species from various families of the order Brassicales, to which the family Moringaceae belongs to. Furthermore, the searches against the CDD revealed that each protein contained a single domain of the AI_{SS} subfamily (CD accession number: cd00261), which includes alpha-amylase inhibitors (AAIs) and seed storage (SS) proteins. AI_{SS} domain model represents one node of a domain family hierarchy (AAI_{LTSS}; CD accession number: cd00010), which includes 4 other nodes: HPS_{like}, nsLTP₂, nsLTP₁ and nsLTP_{like}. When the 6 amino acid sequences determined in this work were compared to Mo-CBP₃ sequences deduced from cDNAs (encoding isoforms Mo-CBP₃-1, Mo-CBP₃-2, Mo-CBP₃-3 and Mo-CBP₃-4), as previously reported (Freire et al., 2015), it was found that one protein sequence was identical to Mo-CBP₃-3 and another one was identical to Mo-CBP₃-4. The other 4 protein sequences had differences in relation to the isoforms identified before by cDNA sequencing. Based on sequence identity percentages, the new isoforms revealed by genomic DNA sequencing were named as Mo-CBP₃-2A, Mo-CBP₃-2B, Mo-CBP₃-3A and Mo-CBP₃-3B. Therefore, the gene family that encodes Mo-CBP₃ contains at least 8 members (*Mo-CBP₃-1*, *Mo-CBP₃-2*, *Mo-CBP₃-2A*, *Mo-CBP₃-2B*, *Mo-CBP₃-3*, *Mo-CBP₃-3A*, *Mo-CBP₃-3B* and *Mo-CBP₃-4*). A phylogenetic analysis showed that these 8 genes are grouped in 3 clusters of closely-related sequences (Fig. 1B). All Mo-CBP₃ isoform sequences showed 8 conserved Cys residues, following a conserved pattern (... C . . C . . /

... CC ... CxC . . C . . C . .), except the isoforms 2B and 3A, in which one of the 8 Cys residues was replaced by Arg (Fig. 1A). This pattern is known as the eight-cysteine motif (8CM), a characteristic structural feature of 2S albumins and other members of the prolamins superfamily (José-Estanyol et al., 2004). These results suggested that isoforms of Mo-CBP₃ are encoded by members of a small multigene family, similar to that observed for other typical 2S seed storage proteins (Scofield and Crouch, 1987).

3.2. Mass spectrometry analysis

Mo-CBP₃ is a two-chain 2S albumin, which showed two protein bands of approximately 8 and 4 kDa, when analyzed by polyacrylamide gel electrophoresis under reducing and denaturing conditions (Freire et al., 2015). The conserved Cys residues of Mo-CBP₃ are involved in 4 disulfide bridges, 2 bridges formed between the small and large chains and 2 other bridges formed between Cys residues of the large chain (Ullah et al., 2015). In this work, when native Mo-CBP₃ was reduced, alkylated and subjected to RP-HPLC, two major peaks were obtained (Fig. 2). ESI-MS analysis of fractions from the first peak revealed

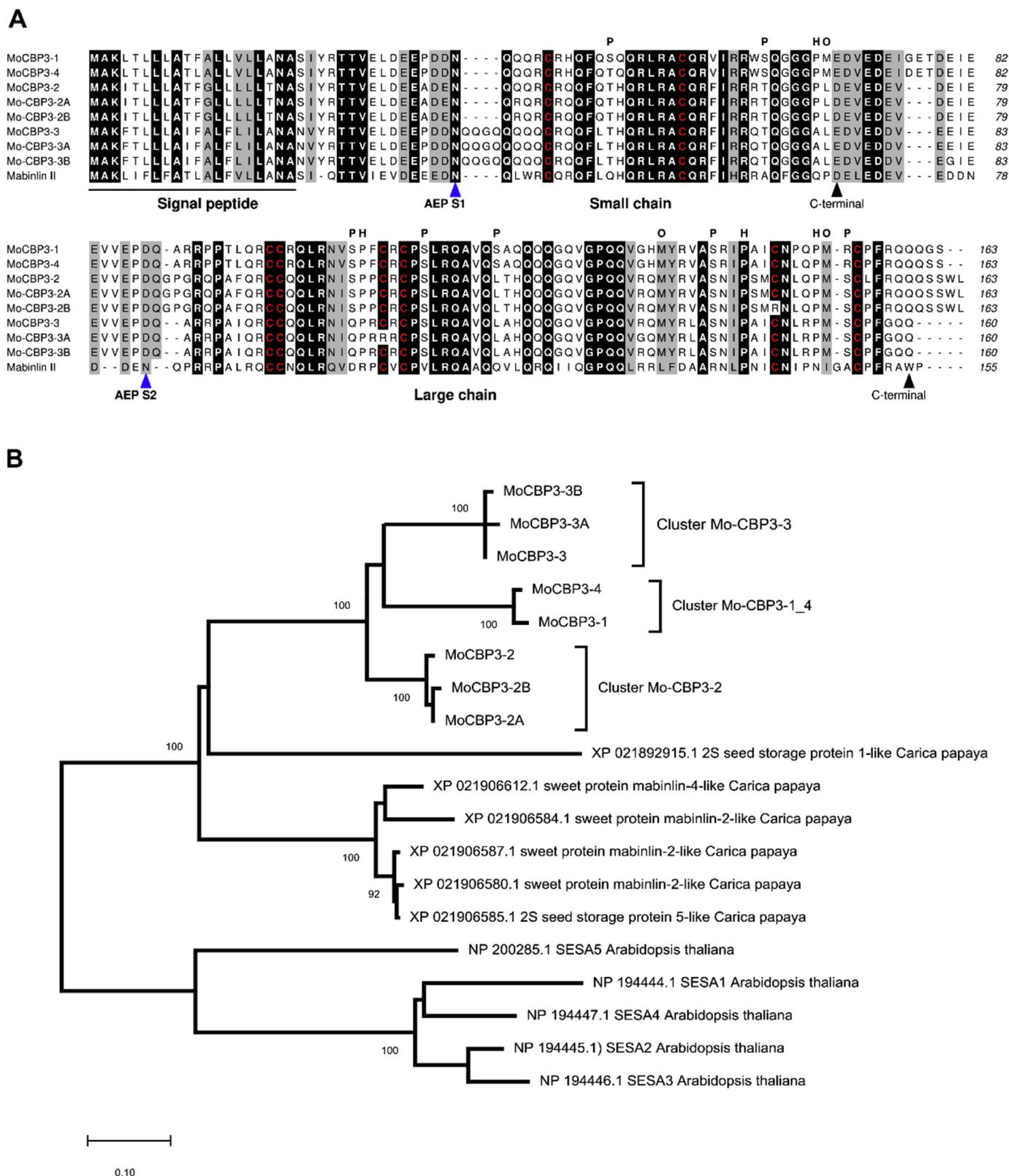


Fig. 1. Multiple alignment of the amino acid sequences of preproMo-CBP₃ isoforms, as deduced from DNA sequences (A), and their phylogenetic relationship (B). (A) The amino acid sequence of the precursor of mabinlin-II (UniProtKB/Swiss-Prot accession number: P30233), was included for comparison. Mo-CBP₃ sites containing residues that were predicted to be hydroxylated (H), phosphorylated (P) or oxidized (O) are indicated. Putative processing sites by asparaginyl endopeptidase (AEP S1 and AEP S2) occur at the carboxyl side of the residues indicated by blue triangles. The C-terminal residues of the small and large chains of mabinlin-II are indicated by black triangles (B) Neighbor-Joining phylogenetic tree. Bootstrap values are shown next to the branches and the scale bar represents 0.1 amino acid substitutions per site. The analysis was performed as described in the Methods section. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

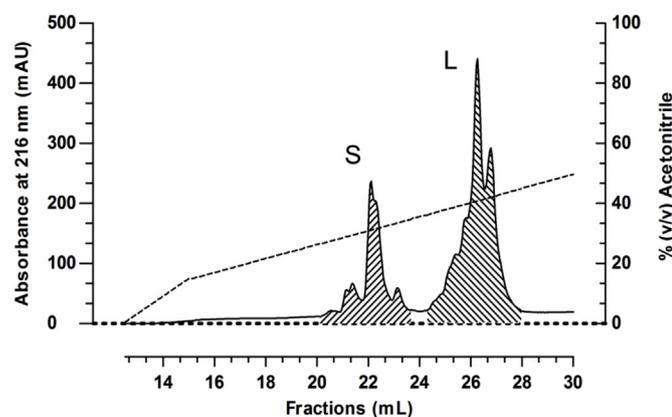


Fig. 2. Fractionation of small and large chains of *Mo-CBP3* by RP-HPLC. Peaks in which polypeptides corresponding to the small (S) and large (L) chains were eluted are labeled. Fractions that were subject to ESI-MS analysis are indicated (hatched areas under the peaks).

molecular masses ranging from 3712.7739 to 4536.0366 Da, indicating that it contained polypeptides corresponding to *Mo-CBP3* small chain (Figs. S12–S21). On the other hand, the experimental masses obtained when the fractions from the second peak were analyzed ranged from 8334.7607 to 8834.7402 Da, demonstrating that they corresponded to variants of *Mo-CBP3* large chain (Figs. S22–S34). Considering all MS spectra obtained from the various chromatographic fractions, 89 and 57 distinct molecular masses were found, which putatively corresponded to variants of the small and large chains of *Mo-CBP3*, respectively.

Like many typical 2S albumins, *Mo-CBP3* isoforms are synthesized as preproteins, which are co- and post-translationally processed, leading to the mature protein, which is composed by one small chain of ~4 kDa linked through disulfide bonds to a large chain of ~8 kDa (Freire et al., 2015). 2S albumins undergo proteolytic processing by the action of an asparaginyl endopeptidase, also known as legumain or vacuolar processing enzyme (VPE). This proteolytic processing includes the removal of the propeptides from the N-terminal regions of the small and large subunits, producing a heterodimer in which the two polypeptide chain are linked by four disulfide bridges (Hara-Hishimura et al., 1993; Hara-Nishimura et al., 1993). Asparaginyl endopeptidases have a strict substrate specificity, cleaving only the peptide bond on the carboxyl side of Asn or Asp residues (Zauner et al., 2018). After the action of the asparaginyl endopeptidase, the C-terminal ends of the small and large chains are usually trimmed by an aspartic exo-protease (Hiraiwa et al., 1997). The precursor of mabinlin II, a heat-stable 2S albumin from the seeds of *C. masakai*, showed significant sequence identity (~44.7–47.5%) when compared to *Mo-CBP3* preproteins (Fig. 1A). In pro-mabinlin II, the proteolytic cleavages that remove the N-terminal propeptides from the small and large chains occur at the carboxyl side of Asn³⁵ and Asn⁸² (Liu et al., 1993; Nirasawa et al., 1996). In other 2S albumins, the proteolytic processing of the propeptides takes place at homologous Asn residues (Krebbes et al., 1988; Hara-Hishimura et al., 1993). From these comparisons, it was hypothesized that the proteolytic cleavages of the N-terminal propeptides from the small and large chains of *Mo-CBP3* isoforms probably occur on the carboxyl sides of an Asn residue (Asn³⁶ in all isoforms) and an Asp residue (Asp⁸⁵ in isoforms 2, 2A and 2B; Asp⁸⁸ in isoforms 1 and 4; and Asp⁸⁹ in isoforms 3, 3A and 3B), respectively (Fig. 1A). Based on this reasoning, the lengths of the small chain of *Mo-CBP3* isoforms were then speculated to be 33 residues, in isoforms 1, 4 (³⁷QQQ ... PME⁶⁹), 2 (³⁷QQQ ... PLD⁶⁹), 2A and 2B (³⁷QRQ ... PLD⁶⁹), or 37 residues, in isoforms 3, 3A and 3B (³⁷QQQ ... ALE⁷³). The large chains were predicted to have 71 residues, in isoforms 3, 3A and 3B (⁹⁰QARR ... GQQ¹⁶⁰), 75 residues, in isoforms 1 (⁸⁹QARR ... QGS¹⁶³) and 4 (⁸⁹QARR ... QSS¹⁶³), or 78 residues, in isoforms 2, 2A and 2B (⁸⁶QGPG

... SWL¹⁶³). Molecular masses were then calculated from the amino acid sequences of the small and large chains (delimited as described above) and only the fixed modification (cam-Cys) was initially considered. When the calculated values were compared to molecular masses experimentally determined by ESI-MS, none of the observed masses could be assigned to any *Mo-CBP3* sequence.

Several experimental evidences have shown that removal of one or a few residues from either or both sides from each subunit is a common post-translational modification (PTM) that occurs during the biosynthesis of many 2S albumins (Moreno et al., 2004). Other PTMs that have also been described in 2S seed storage proteins include cyclization of N-terminal Gln to pGlu (Moreno et al., 2005), hydroxylation of Pro residues (Li et al., 2010) as well as phosphorylation of Ser/Thr residues (Neumann et al., 1996). Therefore, we hypothesized that *Mo-CBP3* chains could be: i) truncated at either side; ii) have a longer C-terminal end; iii) and/or present any or a combination of the PTMs commonly found in 2S albumins. Based on these speculations, we calculated a series of theoretical molecular masses and compared these values with those obtained by ESI-MS. For example, using the sequence of the small chain of *Mo-CBP3-1*, that we initially presumed to have 33 residues (³⁷QQQCRHQFQSQQRLRACQVRIRRWSSQGGPME⁶⁹), as inferred from the comparison with the processing sites of pro-mabinlin II, we assumed the removal of one N-terminal Gln residue, and we also speculated that the C-terminal end could be originated from proteolytic cleavage at the carboxyl side of Glu⁷² instead of Glu⁶⁹. These events would produce a small chain with 35 residues: ³⁸QQRCRHQFQSQQRLRACQVRIRRWSSQGGPME⁷². The predicted monoisotopic molecular mass of this sequence was calculated as 4253.09 Da, and the predicted mass of the modified sequence, under the assumption that Cys residues (Cys⁴¹ and Cys⁵⁴) were alkylated (cam-Cys⁴¹ and cam-Cys⁵⁴), was 4367.13 Da. This value was very close to 4367.01, as observed in one of the various ESI-MS spectra obtained (as shown in Fig. S17), suggesting that the observed mass corresponded to the 35-residues small chain of *Mo-CBP3-1* with 2 cam-Cys residues. Using this approach, most masses (72 out of 89) of the small chain and almost all masses (54 out of 57) of the large chain were assigned to *Mo-CBP3* sequences (Figs. 3 and 4; Tables 1 and 2 and S4–S15). Altogether, these numbers represented 86% of the masses observed by ESI-MS. Moreover, the average difference between the calculated and experimental masses was approximately 0.65 Da, and most differences (~82.3%) were equal or lower than 1 Da. Molecular masses specifically assigned to both chains of isoforms 1, 2 and 4 were identified, whereas other experimental masses were specifically attributed to the large chains of isoforms 2A, 2B, 3A and 3B. Isoforms 2A and 2B differ from each other by a few residues, and a similar feature is observed in isoforms 3, 3A and 3B (Table S3). In these isoforms, the sites containing the differences between them were not in the predicted sequences that would correspond to the small chain. Therefore, some observed molecular masses could be equally assigned to either isoform 2A or isoform 2B, whereas other masses were equally assigned to isoforms 3, 3A or 3B. These findings suggested that at least 7 members of the *Mo-CBP3* gene family are expressed during the latter stages of *M. oleifera* seed development.

3.3. Possible functional implications of the post-translational modifications of *Mo-CBP3*

The amino acid sequences of *Mo-CBP3* isoforms, as deduced from cDNA (Freire et al., 2015) and genomic DNA fragments (this work), are rich in Gln/Glu (22.0–24.3%), Asp/Asn (6.8–8.2%), Arg (10.4–11.7%) and Pro (5.6–7.4%). This amino acid profile is characteristic of seed storage proteins of the prolamin superfamily, which includes 2S albumins. Upon seed imbibition and germination, seed storage proteins are broken down to free amino acids, which are the primary nitrogen source used during emergence and early seedling growth (Tan-Wilson and Wilson, 2012). Using *in situ* immunolocalization techniques, Garcia et al. have shown that large amounts of *Mo-CBP3* are deposited in PSVs

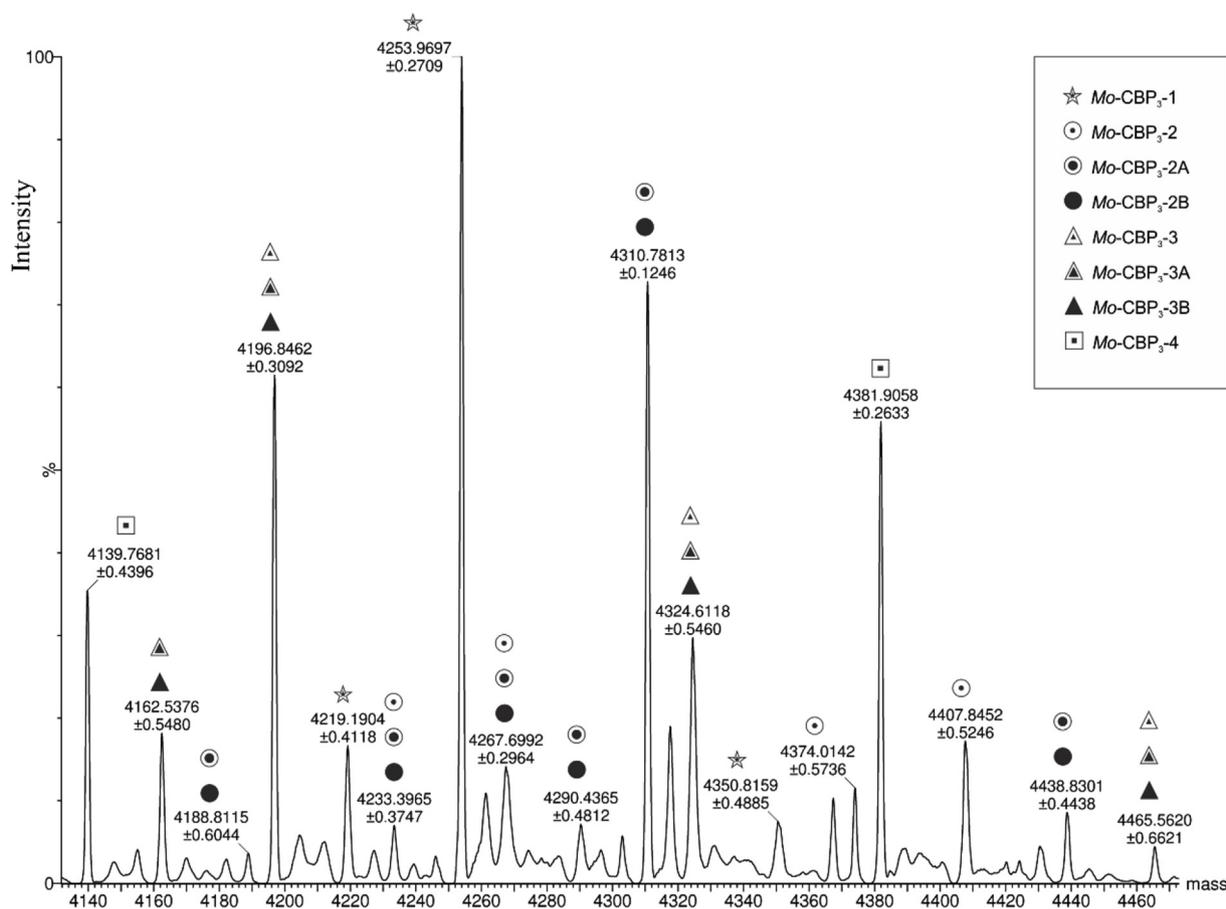


Fig. 3. Representative ESI-MS spectrum of the small chain of Mo-CBP₃. Molecular masses assigned to amino acid sequences from different isoforms are indicated by specific symbols, according to the legend shown (inset).

of cotyledonary cells during seed filling, and once the dried seed has germinated, Mo-CBP₃ polypeptides are rapidly broken down. Therefore, the data on gene (this work), cDNA (Freitas et al., 2015) and protein structure (Ullah et al., 2015), as well as its pattern of spatiotemporal localization *in planta* (Garcia et al., 2019), indicate that Mo-CBP₃ is a typical 2S albumin that provides amino acids to support the initial growth of the seedling. Furthermore, we have shown that the isoforms of mature Mo-CBP₃, as purified from dry seeds, are modified by a series of PTMs, including cleavage of N- and C-terminal ends of both chains, conversion of N-terminal Gln to pGlu, phosphorylation of Ser and Thr, oxidation of Met and hydroxylation of Pro. The implications of these PTMs on the biological roles of Mo-CBP₃ are still unknown, but some of them could be functionally relevant. For example, the N-terminal residue of the small and large chains of Mo-CBP₃ isoforms is usually pyroglutamate. Pyroglutamic acid (pGlu) is formed by the intramolecular cyclization of N-terminal Gln residues of peptides and proteins. In plant cells, this reaction is catalyzed by glutaminyl cyclases (QCs) (Wintjens et al., 2006). Schilling et al. (2007) have shown that plant QCs have poor selectivity for substrates, acting on structurally divergent proteins. However, they observed that many of the QC targets are pathogenesis-related (PR) proteins, such as chitinases and β -1,3-glucanases, which are upregulated in response to pathogens attack. Mo-CBP₃ is not structurally related to these classical PR proteins, but its antifungal and antibacterial action suggests a possible defensive role. Proteins and peptides modified by QCs to pGlu-containing molecules are resistant to aminopeptidases, and hence are more stable. Therefore, the modification of N-terminal Gln residues of Mo-CBP₃ into pGlu could be a mechanism to protect the protein from proteolytic degradation.

Protein phosphorylation is a ubiquitous and reversible PTM, which is catalyzed by kinases and phosphatases, and plays key regulatory roles

in a multitude of biological processes, such as growth, development, immunity and responses to biotic and abiotic stresses, for example (Jha et al., 2017). Many Mo-CBP₃ polypeptide chains were predicted to be modified by phosphorylation at Ser and/or Thr residues. This finding is in agreement with previous works, which have demonstrated that seed storage proteins from different species and belonging to distinct classes, such as 7S- and 11S globulins as well as 2S albumins, are abundantly phosphorylated (Mouzo et al., 2018). Moreover, it has been found that in germinating seeds of *Phaseolus vulgaris*, highly phosphorylated isoforms of phaseolin (the 7S seed storage globulin of *P. vulgaris*) are preferentially mobilized, suggesting that seed storage protein degradation is regulated by a phosphorylation-dependent mechanism (López-Pedrouso et al., 2014). The high frequency of phosphorylation in Mo-CBP₃ isoforms and the degradation kinetics observed during the germination of *M. oleifera* seeds (Garcia et al., 2019) indicate that this PTM may be involved in the rapid mobilization of Mo-CBP₃ during seed germination.

Hydroxylation of amino acid residues in proteins is another reversible and important PTM, although it is less common than other modifications, such as phosphorylation. The side chains of proline and lysine residues can be modified by hydroxylation, forming hydroxyproline and hydroxylysine, respectively. In plants, the occurrence of hydroxyproline in peptide hormones and growth factors has been well documented, being an important PTM for activity and/or receptor binding (Stührwoldt and Schaller, 2019). Contrary to plant signalling peptides, there are few reports on hydroxylation of Pro residues in 2S albumins (Li et al., 2010), and the functional role of this modification in these seed storage proteins is not yet understood. Proline residues are hydroxylated by prolyl-4-hydroxylases (P4Hs). In *A. thaliana*, the expression of some P4H genes is induced in response to hypoxia and

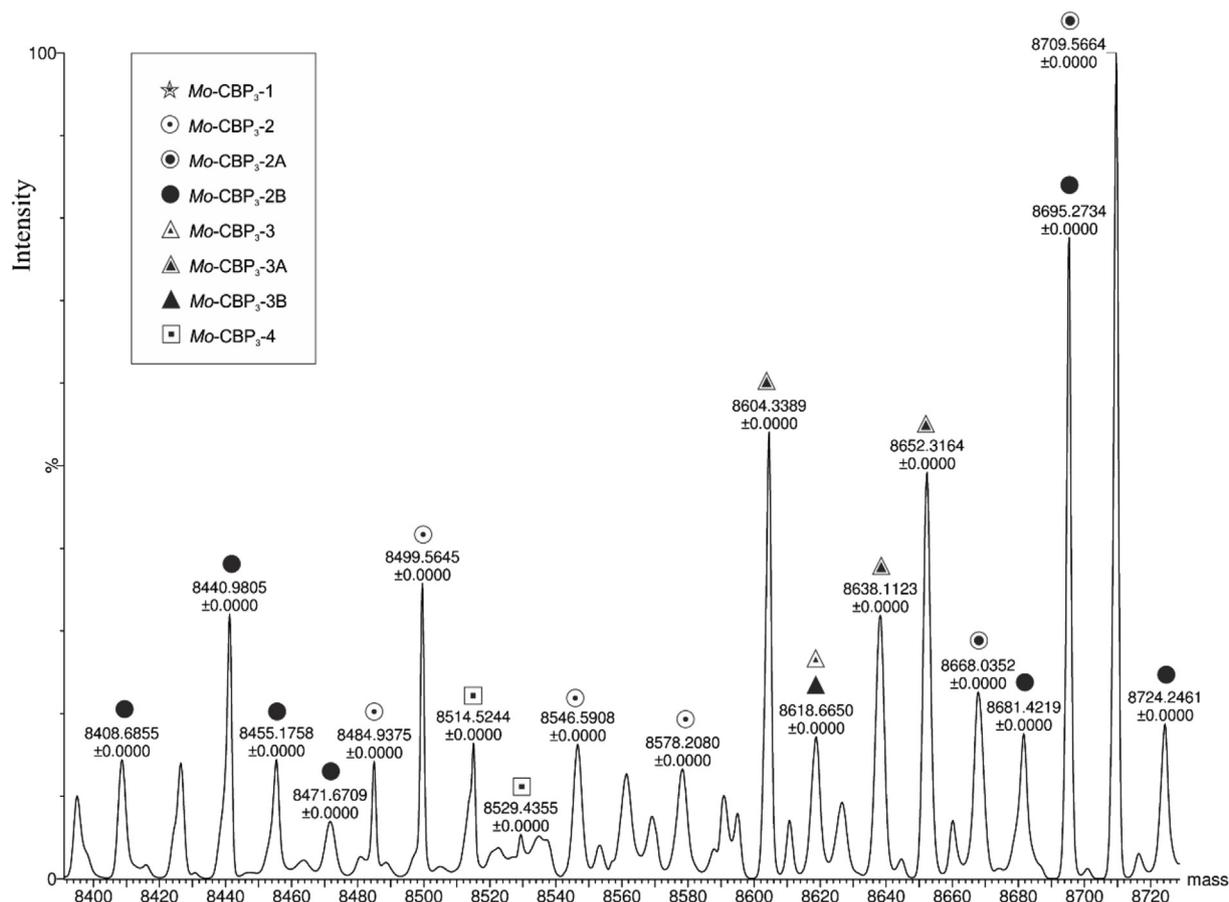


Fig. 4. Representative ESI-MS spectrum of the large chain of Mo-CBP₃. Molecular masses assigned to amino acid sequences from different isoforms are indicated by specific symbols, according to the legend shown (inset).

Table 1

Representative assignment of experimental molecular masses, as determined by ESI-MS, to amino acid sequences of Mo-CBP₃ small chain. Monoisotopic molecular mass values were calculated for each amino acid sequence, with cam-Cys included as a fixed modification and different variable modifications (Oxi: Met oxidation; Hyd: Pro hydroxylation; Pho: Ser or Thr phosphorylation; pGlu: cyclization of N-terminal Gln).

Isoform	Amino acid sequence ^a	Molecular Mass (Da)				Variable modifications
		Calculated	Observed	SD	Delta	
Mo-CBP ₃ -1	³⁹ QRC ... PME ⁶⁹ (31)	3911.9278	3911.2564	0.1680	0.6714	1x Oxi
	³⁸ QQR ... PME ⁶⁹ (32)	4086.9327	4086.8108	0.0000	0.1219	pGlu and 1x Pho
	³⁸ QQR ... MED ⁷⁰ (33)	4218.9893	4219.2536	0.0684	0.2644	1x Pho
	³⁷ QQQ ... MED ⁷⁰ (34)	4267.0829	4267.5342	0.1429	0.4513	
Mo-CBP ₃ -2	³⁹ QRC ... PLD ⁶⁹ (31)	3926.0078	3925.1062	0.0404	0.9016	1x Hyd
	³⁹ QRC ... LDE ⁷⁰ (32)	4038.0213	4038.3741	0.0248	0.3528	pGlu and 1x Hyd
	³⁷ QQQ ... LDE ⁷⁰ (34)	4278.1464	4279.4971	0.0000	1.3507	pGlu
	³⁷ QQQ ... LDE ⁷⁰ (34)	4311.1678	4310.6533	0.2483	0.5145	1x Hyd
Mo-CBP ₃ -2A/2B	³⁸ RQR ... DEV ⁷¹ (34)	4391.1342	4390.1842	0.0000	0.9500	1x Pho and 1x Hyd
	³⁸ RQR ... DEV ⁷¹ (34)	4374.1893	4373.8339	0.1999	0.3554	1x Pho
	³⁷ QRQ ... DEV ⁷¹ (35)	4438.2778	4438.7791	0.1749	0.5012	1x Hyd
	³⁷ QRQ ... EVE ⁷² (36)	4534.3064	4536.0366	0.0000	1.7302	pGlu
Mo-CBP ₃ -3/3A/3B	⁴⁴ QCR ... ALE ⁷³ (30)	3761.8403	3761.9399	0.0000	0.0996	pGlu and 1x Pho
	⁴⁴ QCR ... ALE ⁷³ (30)	3778.8669	3780.6914	0.0888	1.8246	1x Pho
	⁴³ QCC ... ALE ⁷³ (31)	3809.9325	3810.5969	0.0000	0.6644	pGlu
	⁴³ QCC ... ALE ⁷³ (31)	3826.9590	3826.5420	0.1072	0.4170	
Mo-CBP ₃ -4	³⁹ QRC ... EDV ⁷¹ (33)	4140.0378	4139.7871	0.0625	0.2507	1x Oxi
	³⁷ QQQ ... PME ⁶⁹ (33)	4324.9639	4324.6920	0.1290	0.2719	1x Oxi and 2x Pho
	³⁷ QQQ ... PME ⁶⁹ (33)	4246.0293	4245.6079	0.0000	0.4214	1x Pho
	³⁸ QQR ... DVE ⁷² (35)	4381.1429	4381.7926	0.1290	0.6497	

SD: standard deviation.

^a Superscript numbers before and after each sequence refer to residue positions relative to Met^a in the corresponding presequences, as deduced from DNA sequences (Figs. S3–S8); the length of each sequence is shown in parentheses.

Table 2

Representative assignment of experimental molecular masses, as determined by ESI-MS, to amino acid sequences of Mo-CBP₃ large chain. Monoisotopic molecular mass values were calculated for each amino acid sequence, with cam-Cys included as a fixed modification and different variable modifications (Oxi: Met oxidation; Hyd: Pro hydroxylation; Pho: Ser or Thr phosphorylation; pGlu: cyclization of N-terminal Gln).

Isoform	Amino acid sequence ^a	Molecular Mass (Da)				Variable modifications
		Calculated	Observed	SD	Delta	
Mo-CBP ₃ -1	⁸⁹ QAR ... FRQ ¹⁵⁹ (71)	8625.1249	8623.9225	0.4884	1.2023	pGlu and 2x Pho
Mo-CBP ₃ -2	⁹¹ QPA ... RQQ ¹⁵⁸ (68)	8498.7807	8498.2488	0.2299	0.5319	pGlu, 3x Oxi, 3x Pho and 3x Hyd
	⁹¹ QPA ... QQQ ¹⁵⁹ (69)	8530.8794	8531.1088	0.2543	0.2293	pGlu, 3x Oxi, 2x Pho and 2x Hyd
Mo-CBP ₃ -2A	⁸⁶ QGP ... FRQ ¹⁵⁷ (72)	8562.1020	8562.1851	0.5360	0.0806	pGlu, 1x Oxi and 1x Hyd
	⁹¹ QPA ... SSW ¹⁶² (72)	8667.1171	8667.1160	0.0011	0.0006	pGlu and 1x Oxi
	⁸⁶ QGP ... RQQ ¹⁵⁸ (73)	8722.1519	8723.0746	0.5492	0.9227	pGlu, 1x Oxi and 3x Hyd
	⁹¹ QPA ... QQQ ¹⁵⁹ (69)	8546.8310	8546.2335	0.5082	0.5975	pGlu, 2x Oxi and 3x Pho
	⁹¹ QPA ... SSW ¹⁶² (72)	8667.0720	8667.0727	0.4063	0.0006	pGlu, 1x Oxi and 1x Hyd
	⁸⁶ QGP ... RQQ ¹⁵⁸ (73)	8722.1168	8722.1178	0.5492	0.0009	pGlu, 2x Oxi and 3x Hyd
Mo-CBP ₃ -2B	⁸⁶ QGP ... QQQ ¹⁵⁹ (74)	8802.1920	8802.1274	0.0759	0.0646	pGlu, 2x Oxi
	⁹¹ QPA ... FRQ ¹⁵⁷ (67)	8407.7907	8407.7973	0.0000	0.0066	pGlu, 3x Oxi, 3x Pho and 3x Hyd
	⁹¹ QPA ... RQQ ¹⁵⁸ (68)	8455.0962	8439.8712	0.0000	0.0900	pGlu and 3x Pho
	⁹¹ QPA ... QQQ ¹⁵⁹ (69)	8441.0200	8441.0084	0.2649	0.0116	pGlu, 1x Oxi and 1x Pho
Mo-CBP ₃ -3/3B	⁸⁶ QGP ... FRQ ¹⁵⁷ (72)	8647.1385	8647.0463	0.0000	0.0922	pGlu and 1x Phoi
	⁹¹ QPA ... SSW ¹⁶² (72)	8752.1619	8749.9672	0.0000	2.1946	pGlu, 3x Oxi and 1x Hyd
	⁸⁶ QGP ... RQQ ¹⁵⁸ (73)	8615.1569	8615.0647	0.0000	0.0922	pGlu and 3x Oxi
	⁹⁰ QAR ... FQG ¹⁵⁹ (70)	8491.1182	8490.1188	0.0000	0.9994	pGlu, 2x Oxi, 1x Pho and 2x Hyd
	⁹⁰ QAR ... GQQ ¹⁶⁰ (71)	8603.1833	8603.1363	0.4706	0.0469	pGlu, 1x Oxi, 1x Pho and 2x Hyd
	⁹⁰ QAR ... GQQ ¹⁶⁰ (71)	8605.1698	8604.9225	0.0000	0.4706	pGlu, 1x Oxi and 2x Pho
Mo-CBP ₃ -4	⁸⁹ QAR ... FRQ ¹⁵⁹ (71)	8513.2120	8514.2086	0.0913	0.9965	pGlu, 1x Oxi and 1x Hyd

SD: standard deviation.

^a Superscript numbers before and after each sequence refer to residue positions relative to Met^a in the corresponding preprosequences, as deduced from DNA sequences (Figs. S3–S8); the length of each sequence is shown in parentheses.

mechanical wounding (Vlad et al., 2007). Low oxygen levels prevail during seed development, especially during desiccation and maturation (Borisjuk and Rolletschek, 2009). Therefore, one possible mechanism to explain the presence of hydroxyproline in Mo-CBP₃ and other 2S seed storage albumins would be the induction of P4Hs in response to low oxygen levels.

Oxidation of methionine residues was also commonly found in Mo-CBP₃ isoforms. Proteins can be modified *in vivo* by oxidative reactions caused directly by reactive oxygen species (ROS) or indirectly by oxidized products from oxidative stress. In plant cells, redox PTMs are associated with critical regulatory signal pathways involved in cellular differentiation, development, plant-pathogen interaction and abiotic stresses (Ruiz-May et al., 2019). In leaves of *A. thaliana* plants under oxidative stress, 513 methionine oxidation sites in 403 proteins were identified, showing the functional impact of this PTM *in planta* (Jacques et al., 2015). Large protein oxidation has also been observed in *A. thaliana* seeds during seed maturation and germination (Job et al., 2005). Based on these findings, it has been suggested that, due to their abundance, seed storage proteins may act as an efficient scavenging system for ROS that are produced during the seed desiccation damage (El-Maarouf-Bouteau et al., 2013). During the seed development of *M. oleifera*, Mo-CBP₃ genes are highly expressed during seed filling (60 days after anthesis), before the start of desiccation. At latter stages (90 days after anthesis), as Mo-CBP₃ expression drastically decreases, large amounts of protein products are found in PSVs of cotyledon cells (Garcia et al., 2019). At this stage, the seed moisture content is lower than 5% (Kundu, 2009). The start of Mo-CBP₃ deposition in PSVs of developing seeds before the desiccation stage and its intense synthesis during the latter stages of seed development, together with the observation that many Mo-CBP₃ chains are modified by Met oxidation, support the proposed function of 2S albumins and other seed storage proteins as ROS scavengers.

4. Conclusions

Sequencing of genomic DNA fragments revealed new genes encoding isoforms of Mo-CBP₃, a chitin-binding 2S albumin from *Moringa oleifera*. Furthermore, mass spectrometry analysis of protein samples

purified from mature seeds showed that Mo-CBP₃ exists as a complex mixture of isoforms, with a remarkable number of molecular mass variants. Using computational calculations, most of the observed molecular masses were satisfactorily assigned to amino acid sequences deduced from DNA clones. The results supported the assumption that Mo-CBP₃ isoforms arise from different combinations of post-translational modifications, including clipping at the N- and C-terminal ends of both subunits, cyclization of N-terminal Gln residues as well as Pro hydroxylation, Ser/Thr phosphorylation and Met oxidation. Further work is needed to investigate the possible functional implications of the many isoforms of Mo-CBP₃ present in *M. oleifera* seeds.

Conflicts of interest

The authors declare that they have no conflict of interest.

Authors' contributions

AJSS cloned the DNA fragments. JECF performed the purification of protein samples. FBMBM performed MS analysis. JECF and BAMR analyzed MS data. JECF and JEMJ prepared the figures and tables. IMV, JTAO and AMM contributed reagents/materials/analysis tools. TBG conceived the work and wrote the manuscript. All authors read and approved the manuscript.

Acknowledgements

This work was supported by research grants from Conselho Nacional de Desenvolvimento Científico e Tecnológico, Coordenação de Aperfeiçoamento de Pessoal de Nível Superior and Fundação Cearense de Apoio ao Desenvolvimento Científico e Tecnológico. JECF was recipient of a Doctoral Fellowship from CAPES, and AJSS was recipient of Doctoral Fellowships from CAPES and FUNCAP. TBG is a senior researcher (PQ-2) of CNPq.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://>

doi.org/10.1016/j.plaphy.2019.05.003.

References

- Altschul, S.F., Gish, W., Miller, W., Myers, E.W., Lipman, D.J., 1990. Basic local alignment search tool. *J. Mol. Biol.* 215, 403–410.
- Batista, A.B., Oliveira, J.T.A., Gifoni, J.M., Pereira, M.L., Almeida, M.G.G., Gomes, V.M., Da Cunha, M., Ribeiro, S.F.F., Dias, G.B., Beltrami, L.M., Lopes, J.L.S., Grangeiro, T.B., Vasconcelos, I.M., 2014. New insights into the structure and mode of action of Mo-CBP3, an antifungal chitin-binding protein of *Moringa oleifera* seeds. *PLoS One* 9, e111427. <https://doi.org/10.1371/journal.pone.0111427>.
- Blom, N., Gammeltoft, S., Brunak, S., 1999. Sequence and structure-based prediction of eukaryotic protein phosphorylation sites. *J. Mol. Biol.* 294, 1351–1362. <https://doi.org/10.1006/jmbi.1999.3310>.
- Borisjuk, L., Rolletschek, H., 2009. The oxygen status of the developing seed. *New Phytol.* 182, 17–30. <https://doi.org/10.1111/j.1469-8137.2008.02752.x>.
- Breiteneder, H., Radauer, C., 2004. A classification of plant food allergens. *J. Allergy Clin. Immunol.* 113, 821–830. <https://doi.org/10.1016/j.jaci.2004.01.779>.
- Chang, T.-H., Wu, L.-C., Lee, T.-Y., Chen, S.-P., Huang, H.-D., Hornig, J.-T., 2013. EuLoc: a web-server for accurately predict protein subcellular localization in eukaryotes by incorporating various features of sequence segments into the general form of Chou's PseAAC. *J. Comput. Aided Mol. Des.* 27, 91–103. <https://doi.org/10.1007/s10822-012-9628-0>.
- Cho, S.K., Ryu, M.Y., Shah, P., Poulsen, C.P., Yang, S.W., 2016. Post-translational regulation of miRNA pathway components, AGO1 and HYL1, in plants. *Mol. Cell.* 39, 581–586. <https://doi.org/10.14348/molcells.2016.0085>.
- Estam, M.M., Viitanen, L., Salminen, T.A., Edqvist, J., 2011. Evolutionary history of the non-specific lipid transfer proteins. *Mol. Plant* 4, 947–964. <https://doi.org/10.1093/mp/ssr019>.
- El-Maarouf-Bouteau, H., Meimoun, P., Job, C., Job, D., Bailly, C., 2013. Role of protein and mRNA oxidation in seed dormancy and germination. *Front. Plant Sci.* 4, 77. <https://doi.org/10.3389/fpls.2013.00077>.
- Felsenstein, J., 1985. Confidence-limits on phylogenies - an approach using the bootstrap. *Evolution* 39, 783–791. <https://doi.org/10.2307/2408678>.
- Freire, J.E.C., Vasconcelos, I.M., Moreno, F.B.M.B., Batista, A.B., Lobo, M.D.P., Pereira, M.L., Lima, J.P.M.S., Almeida, R.V.M., Sousa, A.J.S., Monteiro-Moreira, A.C.O., Oliveira, J.T.A., Grangeiro, T.B., 2015. Mo-CBP3, an antifungal chitin-binding protein from *Moringa oleifera* seeds, is a member of the 2S albumin family. *PLoS One* 10, e0119871. <https://doi.org/10.1371/journal.pone.0119871>.
- Garcia, T.B., Soares, A.A., Costa, J.H., Costa, H.P.S., Neto, J.X.S., Rocha-Bezerra, L.C.B., Silva, F.D.A., Arantes, M.R., Sousa, D.O.B., Vasconcelos, I.M., Oliveira, J.T.A., 2019. Gene expression and spatiotemporal localization of antifungal chitin-binding proteins during *Moringa oleifera* seed development and germination. *Planta* 249, 1503–1519. <https://doi.org/10.1007/s00425-019-03103-8>.
- Gasteiger, E., Gattiker, A., Hoogland, C., Ivanyi, I., Appel, R.D., Bairoch, A., 2003. ExPASy: the proteomics server for in-depth protein knowledge and analysis. *Nucleic Acids Res.* 31, 3784–3788.
- Gelhaye, E., Rouhier, N., Navrot, N., Jacquot, J.P., 2005. The plant thioredoxin system. *Cell. Mol. Life Sci.* 62, 24–35. <https://doi.org/10.1007/s00118-004-4296-4>.
- Gifoni, J.M., Oliveira, J.T.A., Oliveira, H.D., Batista, A.B., Pereira, M.L., Gomes, A.S., Oliveira, H.P., Grangeiro, T.B., Vasconcelos, I.M., 2012. A novel chitin-binding protein from *Moringa oleifera* seed with potential for plant disease control. *Biopolymers* 98, 406–415.
- Hall, T.A., 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symp. Ser.* 41, 95–98.
- Hara-Hishimura, I., Takeuchi, Y., Inoue, K., Nishimura, M., 1993. Vesicle transport and processing of the precursor to 2S albumin in pumpkin. *Plant J. Cell Mol. Biol.* 4, 793–800.
- Hara-Nishimura, I., Takeuchi, Y., Nishimura, M., 1993. Molecular characterization of a vacuolar processing enzyme related to a putative cysteine proteinase of *Schistosoma mansoni*. *Plant Cell* 5, 1651–1659. <https://doi.org/10.1105/tpc.5.11.1651>.
- Hiraiwa, N., Kondo, M., Nishimura, M., Hara-Nishimura, I., 1997. An aspartic endopeptidase is involved in the breakdown of propeptides of storage proteins in protein-storage vacuoles of plants. *Eur. J. Biochem.* 246, 133–141.
- Hoagland, D.R., Arnon, D.I., 1950. The water-culture method for growing plants without soil. *Circ. Calif. Agric. Exp. Stn.* 347, 1–32.
- Jacques, S., Ghesquière, B., De Bock, P.-J., Demol, H., Wahni, K., Willems, P., Messens, J., Van Breusegem, F., Gevaert, K., 2015. Protein methionine sulfoxide dynamics in *Arabidopsis thaliana* under oxidative stress. *Mol. Cell. Proteomics* 14, 1217–1229. <https://doi.org/10.1074/mcp.M114.043729>.
- Jha, S.K., Malik, S., Sharma, M., Pandey, A., Pandey, G.K., 2017. Recent advances in substrate identification of protein kinases in plants and their role in stress management. *Curr. Genom.* 18, 523–541. <https://doi.org/10.2174/1389202918666170228142703>.
- Job, C., Rajjou, L., Lovigny, Y., Belghazi, M., Job, D., 2005. Patterns of protein oxidation in *Arabidopsis* seeds and during germination. *Plant Physiol.* 138, 790–802. <https://doi.org/10.1104/pp.105.062778>.
- Jolliffe, N.A., Brown, J.C., Neumann, U., Vicré, M., Bachi, A., Hawes, C., Ceriotti, A., Roberts, L.M., Frigerio, L., 2004. Transport of ricin and 2S albumin precursors to the storage vacuoles of *Ricinus communis* endosperm involves the Golgi and VSR-like receptors. *Plant J.* 39, 821–833. <https://doi.org/10.1111/j.1365-313X.2004.02167.x>.
- Jones, D.T., Taylor, W.R., Thornton, J.M., 1992. The rapid generation of mutation data matrices from protein sequences. *Comput. Appl. Biosci.* 8, 275–282.
- José-Estanyol, M., Gomis-Rüth, F.X., Puigdomènech, P., 2004. The eight-cysteine motif, a versatile structure in plant proteins. *Plant Physiol. Biochem.* 42, 355–365. <https://doi.org/10.1016/j.plaphy.2004.03.009>.
- Krebbbers, E., Herdies, L., De Clercq, A., Seurinck, J., Leemans, J., Van Damme, J., Segura, M., Gheysen, G., Van Montagu, M., Vandekerckhove, J., 1988. Determination of the processing sites of an *Arabidopsis* 2S albumin and characterization of the complete gene family. *Plant Physiol.* 87, 859–866.
- Kumar, Pramod, Kesari, P., Dhindwal, S., Choudhary, A.K., Katiki, M., Neetu, null, Verma, A., Ambatipudi, K., Tomar, S., Sharma, A.K., Mishra, G., Kumar, Pravidra, 2017. A novel function for globulin in sequestering plant hormone: crystal structure of *Wrightia tinctoria* 11S globulin in complex with auxin. *Sci. Rep.* 7, 4705. <https://doi.org/10.1038/s41598-017-04518-7>.
- Kumar, S., Stecher, G., Li, M., Knyaz, C., Tamura, K., 2018. MEGA X: molecular evolutionary genetics analysis across computing platforms. *Mol. Biol. Evol.* 35, 1547–1549. <https://doi.org/10.1093/molbev/msy096>.
- Kundu, M., 2009. Effect of seed maturity on germination and desiccation tolerance of *Moringa oleifera* seed. *Seed Sci. Technol.* 37, 589–596. <https://doi.org/10.15258/ss.2009.37.3.08>.
- Li, J., Shefcheck, K., Callahan, J., Fenselau, C., 2010. Primary sequence and site-selective hydroxylation of prolines in isoforms of a major peanut allergen protein Ara h 2. *Protein Sci.* 19, 174–182. <https://doi.org/10.1002/pro.295>.
- Liener, I.E., Sharon, N., Goldstein, I.J., 1986. *The Lectins: Properties, Functions, and Applications in Biology and Medicine*. Academic Press, Orlando.
- Liu, X., Maeda, S., Hu, Z., Aiuchi, T., Nakaya, K., Kurihara, Y., 1993. Purification, complete amino acid sequence and structural characterization of the heat-stable sweet protein, mabinlin II. *Eur. J. Biochem.* 211, 281–287.
- López-Pedrouso, M., Alonso, J., Zapata, C., 2014. Evidence for phosphorylation of the major seed storage protein of the common bean and its phosphorylation-dependent degradation during germination. *Plant Mol. Biol.* 84, 415–428. <https://doi.org/10.1007/s11103-013-0141-1>.
- Marchler-Bauer, A., Bryant, S.H., 2004. CD-Search: protein domain annotations on the fly. *Nucleic Acids Res.* 32, W327–W331. <https://doi.org/10.1093/nar/gkh454>.
- Moreno, F.J., Jenkins, J.A., Mellon, F.A., Rigby, N.M., Robertson, J.A., Wellner, N., Clare Mills, E.N., 2004. Mass spectrometry and structural characterization of 2S albumin isoforms from Brazil nuts (*Bertholletia excelsa*). *Biochim. Biophys. Acta* 1698, 175–186. <https://doi.org/10.1016/j.bbapap.2003.11.007>.
- Moreno, F.J., Mellon, F.A., Wickham, M.S.J., Bottrill, A.R., Mills, E.N.C., 2005. Stability of the major allergen Brazil nut 2S albumin (Ber e 1) to physiologically relevant in vitro gastrointestinal digestion. *FEBS J.* 272, 341–352. <https://doi.org/10.1111/j.1742-4658.2004.04472.x>.
- Mouzo, D., Bernal, J., López-Pedrouso, M., Franco, D., Zapata, C., 2018. Advances in the biology of seed and vegetative storage proteins based on two-dimensional electrophoresis coupled to mass spectrometry. *Molecules* 23, 2462. <https://doi.org/10.3390/molecules23102462>.
- Mylne, J.S., Hara-Nishimura, I., Rosengren, K.J., 2014. Seed storage albumins: biosynthesis, trafficking and structures. *Funct. Plant Biol.* 41, 671–677. <https://doi.org/10.1071/FP14035>.
- Neto, J.X.S., Pereira, M.L., Oliveira, J.T.A., Rocha-Bezerra, L.C.B., Lopes, T.D.P., Costa, H.P.S., Sousa, D.O.B., Rocha, B.A.M., Grangeiro, T.B., Freire, J.E.C., Monteiro-Moreira, A.C.O., Lobo, M.D.P., Brilhante, R.S.N., Vasconcelos, I.M., 2017. A chitin-binding protein purified from *Moringa oleifera* seeds presents anticandidal activity by increasing cell membrane permeability and reactive oxygen species production. *Front. Microbiol.* 8, 980. <https://doi.org/10.3389/fmicb.2017.00980>.
- Neumann, G.M., Condron, R., Polya, G.M., 1996. Purification and sequencing of yellow mustard seed napin small and large chains that are phosphorylated by plant calcium-dependent protein kinase and are calmodulin antagonists. *Plant Sci.* 119, 49–66. [https://doi.org/10.1016/0168-9452\(96\)04476-7](https://doi.org/10.1016/0168-9452(96)04476-7).
- Nirasawa, S., Masuda, Y., Nakaya, K., Kurihara, Y., 1996. Cloning and sequencing of a cDNA encoding a heat-stable sweet protein, mabinlin II. *Gene* 181, 225–227.
- Oliveira, J.T.A., Souza, P.F.N., Vasconcelos, I.M., Dias, L.P., Martins, T.F., Van Tilburg, M.F., Guedes, M.I.F., Sousa, D.O.B., 2019. Mo-CBP3-PepI, Mo-CBP3-PepII, and Mo-CBP3-PepIII are synthetic antimicrobial peptides active against human pathogens by stimulating ROS generation and increasing plasma membrane permeability. *Biochimie* 157, 10–21. <https://doi.org/10.1016/j.biochi.2018.10.016>.
- Pereira, C., Pereira, S., Pissarra, J., 2014. Delivering of proteins to the plant vacuole—an update. *Int. J. Mol. Sci.* 15, 7611–7623. <https://doi.org/10.3390/ijms15057611>.
- Petersen, T.N., Brunak, S., von Heijne, G., Nielsen, H., 2011. SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nat. Methods* 8, 785–786. <https://doi.org/10.1038/nmeth.1701>.
- Ruiz-May, E., Segura-Cabrera, A., Elizalde-Contreras, J.M., Shannon, L.M., Loyola-Vargas, V.M., 2019. A recent advance in the intracellular and extracellular redox post-translational modification of proteins in plants. *J. Mol. Recognit.* 32, e2754. <https://doi.org/10.1002/jmr.2754>.
- Saitou, N., Nei, M., 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* 4, 406–425.
- Salminen, T.A., Blomqvist, K., Edqvist, J., 2016. Lipid transfer proteins: classification, nomenclature, structure, and function. *Planta* 244, 971–997. <https://doi.org/10.1007/s00425-016-2585-4>.
- Sambrook, J., Fritsch, E., Maniatis, T., 1989. *Molecular Cloning: A Laboratory Manual*, second ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor.
- Schägger, H., von Jagow, G., 1987. Tricine-sodium dodecyl sulfate-polyacrylamide gel electrophoresis for the separation of proteins in the range from 1 to 100 kDa. *Anal. Biochem.* 166, 368–379.
- Scofield, S.R., Crouch, M.L., 1987. Nucleotide sequence of a member of the napin storage protein family from *Brassica napus*. *J. Biol. Chem.* 262, 12202–12208.
- Shang, X., Cao, Y., Ma, L., 2017. Alternative splicing in plant genes: a means of regulating the environmental fitness of plants. *Int. J. Mol. Sci.* 18, 432. <https://doi.org/10.3390/ijms18040432>.

- 3390/ijms18020432.
- Schilling, S., Stenzel, I., von Bohlen, A., Wermann, M., Schulz, K., Demuth, H.U., Wasternack, C., 2007. Isolation and characterization of the glutaminyl cyclases from *Solanum tuberosum* and *Arabidopsis thaliana*: implications for physiological functions. *Biol. Chem.* 388, 145–153.
- Shewry, P.R., Napier, J.A., Tatham, A.S., 1995. Seed storage proteins: structures and biosynthesis. *Plant Cell* 7, 945–956. <https://doi.org/10.1105/tpc.7.7.945>.
- Stührwohldt, N., Schaller, A., 2019. Regulation of plant peptide hormones and growth factors by post-translational modification. *Plant Biol.* 21 (Suppl. 1), 49–63. <https://doi.org/10.1111/plb.12881>.
- Tan-Wilson, A.L., Wilson, K.A., 2012. Mobilization of seed protein reserves. *Physiol. Plantarum* 145, 140–153. <https://doi.org/10.1111/j.1399-3054.2011.01535.x>.
- Thompson, J.D., Higgins, D.G., Gibson, T.J., 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* 22, 4673–4680.
- Tomar, P.P.S., Chaudhary, N.S., Priyadarshi, P., Mishra, P., Gahloth, D., Patel, G.K., Selvakumar, P., Kumar, P., Sharma, A.K., 2014. Purification, characterisation and cloning of a 2S albumin with DNase, RNase and antifungal activities from *Putranjiva roxburghii*. *Appl. Biochem. Biotechnol.* 174, 471–482. <https://doi.org/10.1007/s12010-014-1078-9>.
- Ullah, A., Mariutti, R.B., Masood, R., Caruso, I.P., Costa, G.H.G., Millena de Freitas, C., Santos, C.R., Zanphorlin, L.M., Rossini Mutton, M.J., Murakami, M.T., Arni, R.K., 2015. Crystal structure of mature 2S albumin from *Moringa oleifera* seeds. *Biochem. Biophys. Res. Commun.* 468, 365–371. <https://doi.org/10.1016/j.bbrc.2015.10.087>.
- Vlad, F., Spano, T., Vlad, D., Daher, F.B., Ouelhadj, A., Fragkostefanakis, S., Kalaitzis, P., 2007. Involvement of *Arabidopsis* prolyl 4 hydroxylases in hypoxia, anoxia and mechanical wounding. *Plant Signal. Behav.* 2, 368–369.
- Voelker, T.A., Staswick, P., Chrispeels, M.J., 1986. Molecular analysis of two phytohemagglutinin genes and their expression in *Phaseolus vulgaris* cv. Pinto, a lectin-deficient cultivar of the bean. *EMBO J.* 5, 3075–3082.
- Warner, S.A.J., 1996. Genomic DNA isolation and lambda library construction. In: Foster, G.D., Twell, D. (Eds.), *Plant Gene Isolation: Principles and Practice*. John Wiley & Sons, West Sussex, pp. 51–73.
- Wintjens, R., Belrhali, H., Clantin, B., Azarkan, M., Bompard, C., Baeyens-Volant, D., Looze, Y., Villeret, V., 2006. Crystal structure of papaya glutaminyl cyclase, an archetype for plant and bacterial glutaminyl cyclases. *J. Mol. Biol.* 357, 457–470. <https://doi.org/10.1016/j.jmb.2005.12.029>.
- Xu, Y., Wen, X., Shao, X.-J., Deng, N.-Y., Chou, K.-C., 2014. iHyd-PseAAC: predicting hydroxyproline and hydroxyllysine in proteins by incorporating dipeptide position-specific propensity into pseudo amino acid composition. *Int. J. Mol. Sci.* 15, 7594–7610. <https://doi.org/10.3390/ijms15057594>.
- Zauner, F.B., Elsässer, B., Dall, E., Cabrele, C., Brandstetter, H., 2018. Structural analyses of *Arabidopsis thaliana* legumain γ reveal differential recognition and processing of proteolysis and ligation substrates. *J. Biol. Chem.* 293, 8934–8946. <https://doi.org/10.1074/jbc.M117.817031>.