Plant Physiology and Biochemistry 84 (2014) 142–148

Contents lists available at ScienceDirect

Plant Physiology and Biochemistry

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

Research article

Molecular characterization of edestin gene family in *Cannabis sativa* L.

Teresa Docimo [1], Immacolata Caruso [1], Elena Ponzoni, Monica Mattana, Incoronata Galasso*

Institute of Agricultural Biology and Biotechnology, CNR, Via Bassini 15, Milan 20133, Italy

ARTICLE INFO

Article history:
Received 4 August 2014
Accepted 23 September 2014
Available online 24 September 2014

Keywords:
Hemp
11S globulin
Gene family
Seed storage protein
Genome organization
qRT-PCR

ABSTRACT

Globulins are the predominant class of seed storage proteins in a wide variety of plants. In many plant species globulins are present in several isoforms encoded by gene families. The major seed storage protein of *Cannabis sativa* L. is the globulin edestin, widely known for its nutritional potential. In this work, we report the isolation of seven cDNAs encoding for edestin from the *C. sativa* variety Carmagnola. Southern blot hybridization is in agreement with the number of identified edestin genes. All seven sequences showed the characteristic globulin features, but they result to be divergent members/forms of two edestin types. According to their sequence similarity four forms named CsEde1A, CsEde1B, CsEde1C, CsEde1D have been assigned to the edestin type 1 and the three forms CsEde2A, CsEde2B, CsEde2C to the edestin type 2. Analysis of the coding sequences revealed a high percentage of similarity (98–99%) among the different forms belonging to the same type, which decreased significantly to approximately 64% between the forms belonging to different types. Quantitative RT-PCR analysis revealed that both edestin types are expressed in developing hemp seeds and the amount of CsEde1 was 4.44 ± 0.10 higher than CsEde2. Both edestin types exhibited a high percentage of arginine (11–12%), but CsEde2 resulted particularly rich in methionine residues (2.36%) respect to CsEde1 (0.82%). The amino acid composition determined in CsEde1 and CsEde2 types suggests that these seed proteins can be used to improve the nutritional quality of plant food-stuffs.

1. Introduction

Cannabis sativa L. (common name, hemp) is an annual wind-pollinated dioecious plant of the *Cannabaceae* family. Hemp, native to Asia, is an ancient plant that has been used since prehistoric times as an important source of fibre, food, medicine, and psychoactive drugs (Zias et al., 1993; Cromack, 1998; Russo et al., 2008). Fibre, separated from the straw, has been utilized for various types of cordage, to make sails for ships, fishing net, and also for the production of paper, clothes and other textile products. Recently, many other interesting applications have been reported. For instance, fibre could be usefully exploited in the automotive and building industry, agrotextiles and geotextiles sectors in the form of innovative materials and composites (Karus et al., 2000; Small and Marcus, 2002). Apart from hemp fibre, the seeds are an excellent source of dietary oil and proteins (Callaway, 2004). They contain approximately 25–35% oil, 20–25% protein, 20–30% carbohydrates, and 10–15% fibre along with several vitamins and minerals (Callaway, 2004; Deferne and Pate, 1996). Specifically, hemp seed oil is composed of more than 80% in polyunsaturated fatty acids and it is a rich source of the two essential fatty acids: linoleic acid (C18:2 omega-6) and linolenic acid (C18:3 omega-3) in a ratio between 2:1 and 3:1, which is considered to be optimal for human health (Deferne and Pate, 1996; Callaway et al., 1997).

Furthermore hemp seeds are an excellent nutritional source of high quality protein, which are easily digested, absorbed, and utilized (House et al., 2010). The two main proteins in hemp seeds are edestin and albumin (Callaway, 2004; Tang et al., 2006). Edestin, which accounts for about 60–80% of the total protein content (Tang et al., 2006; Odani and Odani, 1998), has been the subject of intensive investigation since long time. Osborne (1892) described several crystallized globulins obtained from different plant seeds and hemp was one of these plants. Svedberg and Stamm (1929) reported for the first time the molecular weight (MW) of edestin by ultracentrifuge method suggesting a MW of 211 ± 10 kDa. St. Angelo et al. (1968) demonstrated that the globulin edestin is located in the aleurone grains as large crystalloidal substructures and Patel et al. (1994), using X-ray crystallographic techniques, showed that edestin molecule is composed of six identical subunits,

* Corresponding author.
E-mail address: galasso@ibba.cnr.it (I. Galasso).
[1] Teresa Docimo and Immacolata Caruso contributed equally to the work.

http://dx.doi.org/10.1016/j.plaphy.2014.09.011
0981-9428/

and each subunit consists of an acidic and basic subunit linked by one disulfide bond, like the hexamer of *Glycine max* (L.) Merr. glycinin (Tang et al., 2006). Recently, a fine characterization of edestin globulin from a Korean variety has been reported by Kim and Lee (2011). The authors isolated the edestin protein and carried out a N-terminal amino acid sequencing of the first seven and six amino acid residues of the α and β-subunit, respectively, by the automated Edman degradation method. In addition, they also demonstrated that the isolated edestin has a 2,2-diphenyl-1-picrylhydrazyl (DPPH) free radical scavenging activity (Kim and Lee, 2011). Although hemp edestin has been well-characterized at biochemical level, data on its DNA sequence and genomic structure and organization are still lacking. Therefore, in this paper we report for the first time the characterization of the edestin 11S nucleotide sequence isolated from the genome of the Italian non-drug hemp variety Carmagnola.

2. Materials and methods

2.1. Plant material

Seeds of *C. sativa* variety Carmagnola were obtained from the seed company Assocanapa S.r.l., Carmagnola (Torino), Italy. After sowing in pots, hemp plants were grown in a phytotron under a 14/10 h light/dark photoperiod, with temperatures set at 26/20 °C day/night and relative humidity between 50 and 55%. Young leaves and fresh maturing seeds (developing seeds) were collected and immediately frozen in liquid nitrogen and stored at −80 °C until use.

2.2. Isolation and cloning of edestin gene family

Total RNA was isolated from developing hemp seeds using the TRIzol® reagent (Invitrogen, Carlsbad, CA, USA). RNA quality was assessed on an Agilent Bioanalyzer 2100 using a RNA 6000 Nano Kit (Agilent Technologies, Palo Alto, CA) whereas its quantification was carried out using a NanoDrop 2000c (NanoDrop Technologies, Wilmington, USA). The RNA solution was digested with RNase-free DNase I (Invitrogen, Carlsbad, CA, USA) to remove any contaminating genomic DNA before the reverse transcription reaction. First strand cDNA was synthesized from 1 μg of total RNA, by using Superscript II reverse transcriptase according to the manufacturer's instructions (Invitrogen, Carlsbad, CA, USA). Identification of putative edestin 3′ end sequence was carried out using the SMARTer™ RACE cDNA amplification kit following the manufacturer's instructions (Clontech, Mountain View, CA, USA). As forward primer, a degenerate gene-specific primer EdeF 5′-GGNYTNGARGAR-ACNTTYTG-3′ designed on the partial protein sequence β-subunit GLEETF reported by Kim and Lee (2011), was used. The PCR fragment obtained from RACE reaction was cloned in pGEM-T vector (Promega, Madison, WI) and transformed in *Escherichia coli* JM109. After sequencing, two different 3′ end putative edestin sequences were found. To isolate the putative full-length of each edestin cDNA, 5′ RACE reactions were carried out using two specific reverse primers (CsEde1rev and CsEde2rev) designed on both the above 3′ end sequences (Supplementary Table S1). PCR amplification of 5′ end region was performed using the same SMARTer™ RACE amplification kit following the manufacturer's instructions. After PCR two 5′ end fragments of different lengths were amplified. In order to get the beginning of the gene and to design primers able to PCR amplify the full gene, the two fragments were cut from the gel, ligated into pGEM-T vector and transformed in *E. coli*. About 30 clones from each PCR amplification were sequenced in both directions with an automated sequencer by Macrogen (Seoul, South Korea), using the universal T7 and Sp6 Sequencing Primers from pGEM-T vector. After sequence analysis one specific primer was designed at the beginning of each gene and called CsEde1for and CsEde2for, respectively (Supplementary Table S1).

2.3. DNA cloning, sequencing and Southern analysis

In order to isolate the genomic form of each edestin gene, genomic DNA was extracted from fresh leaf tissue according to the method reported by Doyle and Doyle (1987). The primers CsEde1for/CsEde1rev and CsEde2for/CsEde2rev (Supplementary Table S1) were used for PCR amplification of the edestin 1 and 2, respectively. The amplification program included an initial step at 95 °C for 5 min, 36 cycles of 2 min at 94 °C, 45 s at 56 °C and 1.5 min at 72 °C, and a final extension for 5 min at 72 °C. The resulting PCR products, separated onto a 1% agarose gel, showed a single band of about 1700 and 1600 bp for the edestin 1 and edestin 2, respectively. After gel purification the two fragments were cloned in pGEM-T plasmid. About 30 clones for each edestin were sequenced in both directions as described above. The nucleotide sequences obtained are available in the EMBL Nucleotide Database under the accession numbers from HG976888 to HG976894.

For Southern hybridization, 5 μg of genomic DNA was digested with the restriction enzyme *EcoRI* and *EcoRV*, size separated on a 0.8% agarose gel and transferred onto a nylon membrane (Amersham, GE Healthcare UK). The PCR fragments containing the edestin 1 and edestin 2 were labelled with *α*-[32P]-dCTP using a random primer DNA labelling kit according to the manufacturer instructions (Fermentas, Life Science) and used as probe for filter hybridization. Filter washing was in 1.5 mM of sodium citrate pH 7.0, 15 mM of NaCl and 0.5% SDS at 65 °C.

2.4. Expression analysis by quantitative real-time PCR (qRT-PCR)

For cDNA synthesis, 1 μg of total RNA was reverse transcribed using the Superscript III First-Strand Synthesis SuperMix (Invitrogen Carlsbad, CA, USA) according to the manufacturer's instructions. Each PCR reaction contained 12.5 μL of SYBR Green PCR Master Mix (Applied Biosystems, Foster City, CA, USA), 0.25 μM primers and 1 μL cDNA. The cycling parameters used were 2 min 50 °C and 10 min at 95 °C followed by 40 cycles of 95 °C for 15 s, 60 °C for 1 min, and a standard dissociation protocol (95 °C 15 s, 60 °C for 1 min, 95 °C for 15 s). All reactions were performed on a 7300 Real-Time PCR System (Applied Biosystems, Foster City, CA, USA) in two biological and three technical replicates. Gene-specific primers were designed on CsEde1 and CsEde2 gene to produce amplicons of 99 bp and 83 bp, respectively (Supplementary Table S1). PCR efficiencies calculated by the standard curve method were 101%–107% for each primer pairs, respectively with a correlation coefficient value (R^2) of 0.98–0.99. The $2^{-\Delta\Delta Ct}$ method was used for quantitative analysis (Livak and Schmittgen, 2001). Ct values were calculated using ABI 7300 system software (Applied Biosystem, Foster City, CA, USA). Relative expression of each edestin gene was calculated using ubiquitin gene (EMBL Accession number: AJ864397.1) as reference gene. Sequences of ubiquitin gene specific primers are reported in Supplementary Table S1.

2.5. Extraction of hemp seed protein and analysis by SDS-PAGE

Developing seeds and mature dry seeds were used to prepare hemp protein isolates, following the method described by Tang et al. (2006) with slight modifications. Briefly, hemp seeds were ground in liquid nitrogen and mixed with deionized water at room temperature and the mixture was adjusted to pH 10 with NaOH. The extraction proceeded for 1 h under continuous stirring. Following centrifugation at 8000 g for 30 min at room temperature.

The supernatant was recovered and adjusted to pH 5.0 with HCl, the precipitate was collected by centrifugation at 8000 g for 10 min and resuspended in cold deionized water, adjusted to pH 6.8 with NaOH. The suspension was freeze-dried. SDS-PAGE analysis was performed according to the method of Laemmli (1970) using 12% separating and 4% stacking gel. The samples were dissolved in sample buffer: 0.125 M Tris–HCl buffer pH 8.0, 0.1% w/v SDS, 0.05 bromophenol blue, 30% v/v glycerol and 5% v/v β-mercaptoethanol for reduced condition. For each sample the same amount of protein was applied to each lane. Electrophoresis was performed at 10 mA in the stacking and afterwards at 20 mA. Following electrophoresis, the gels were stained with 0.25% Coomassie Blue R-250.

3. Results

3.1. Isolation and sequence analysis of hemp edestin genes

In this study, 3′ and 5′ RACE strategies were carried out in order to isolate the putative edestin gene from developing hemp seeds. After cloning and sequencing, two types of edestin of different size, CsEde1 of 1536 bp and CsEde2 of 1476 bp, were found in the cDNA. Based on nucleotide differences for each edestin type, four and three forms were identified. The four forms of CsEde1 were named CsEde1A, CsEde1B, CsEde1C and CsEde1D whereas the three forms of the CsEde2 were called CsEde2A CsEde2B and CsEde2C (Table 1). A BLAST search of all four putative edestin CsEde1 coding sequences against the EMBL database revealed a high similarity (74%) to the 11S globulin precursor isoforms 1A and 1B (EMBL accession number: EF091694 and EF091695, respectively) of *Ficus pumila* var. awkeotsang, while the CsEde2 showed highest similarity (76–78%) to the 11S globulin isoforms 2B and 2A (EMBL accession number: EF091696 and EF091693, respectively) of *F. pumila* var. awkeotsang.

Analysis of the coding sequences of all seven forms revealed a high percentage of similarity between the different forms belonging to the same type (98–99%) whereas similarity among the forms belonging to the different types drops down to 64% (Table 2). A similar result was obtained by comparison of CsEde1 to CsEde2 amino acid sequences which showed an overall similarity of 52%. The deduced amino acid sequence of all edestin forms showed common features such as an N-terminal signal sequence for ER targeting, a conservative processing site for splitting into an acid, and a basic subunit and four conservative cysteine residues (Fig. 1). The amino acid composition of these seven storage proteins indicated that they contain all 9 essential amino acids (Table 3). Both edestin types (CsEde1 and CsEde2) contain high percentage of arginine (>11%). The most evident difference between the two edestin types regards the content of methionine. Namely the CsEde2 type contains nearly three times more of methionine (2.36%) respect to CsEde1 (0.82) (Table 3). Another difference between the two types of edestin is their calculated molecular weight.

Table 2
Percentage of sequence similarity of coding sequences (CDS), amino acid sequences and genomic DNA of the different forms of CsEde1 and CsEde2.

Sequence pairs	% Sequence similarity		
	Coding sequence	Amino acid sequence	Genomic DNA
CsEde1A/CsEde1B	99.09	99.80	98.82
CsEde1A/CsEde1C	99.15	100	98.88
CsEde1A/CsEde1D	99.67	99.41	99.72
CsEde1B/CsEde1C	99.54	99.80	99.61
CsEde1B/CsEde1D	99.02	99.22	98.77
CsEde1C/CsEde1D	98.83	99.41	98.60
CsEde1A/CsEde2A	64.50	51.89	64.36
CsEde1A/CsEde2B	64.57	52.10	64.65
CsEde1A/CsEde2C	64.79	52.10	64.76
CsEde1B/CsEde2A	64.79	51.89	64.53
CsEde1B/CsEde2B	64.71	52.10	64.76
CsEde1B/CsEde2C	64.93	52.10	64.81
CsEde1C/CsEde2A	64.79	51.89	64.55
CsEde1C/CsEde2B	64.71	52.10	64.78
CsEde1C/CsEde2C	64.93	52.10	64.83
CsEde1D/CsEde2A	64.36	51.68	64.21
CsEde1D/CsEde2B	63.43	51.89	64.50
CsEde1D/CsEde2C	64.64	51.89	64.62
CsEde2A/CsEde2B	98.85	99.59	98.62
CsEde2A/CsEde2C	99.19	99.80	98.92
CsEde2B/CsEde2C	99.53	99.80	99.58

CsEde1 (all forms) coding sequence predicts a polypeptide precursor (after removing the signal peptide) of about 56.0 kDa, and a final globulin subunit pair with an α-subunit of 34.3 kDa and a β-subunit of 21.7 kDa, while CsEde2 (all forms) predicts a polypeptide precursor of 53.7 kDa and an α-subunit of 32.9 kDa and a β-subunit of 20.8 kDa (Table 1).

When genomic DNA was used as template and CsEde1for/CsEde1rev and CsEde2for/CsEde2rev as primers for PCR amplification of the edestin type 1 and 2, four and three genomic forms were identified for CsEde1 and CsEde2, respectively. A BLAST search of CsEde1 and CsEde2 genomic sequences against the assembled draft genome of *C. sativa* (http://genome.ccbr.utoronto.ca/index.html) (van Bakel et al., 2011), revealed a high similarity with 7 regions of 4 different scaffolds. In particular, CsEde1 showed a 92–98% of similarity to the scaffold33732 (position 10525…12291 and 19324…21067) and the scaffold11174 (pos. 3212…4703 and 11318…12687) whereas CsEde2 showed a 98% of similarity to the scaffold34968 (pos. 18803…20461 and 27344…28935) and scaffold102546 (pos. 5727…7325). Comparison of the CsEde1 cDNA sequence with the genomic DNA nucleotide sequence indicates the presence of three introns, two within the sequence encoding the edestin α-subunit and one within the β-subunit (Fig. 2). The three introns, interrupted the coding regions at the same relative positions in each of the four forms. When the three CsEde2 gene forms were compared with the corresponding genomic forms only two introns were detected both within the sequence encoding the edestin α-subunit (Fig. 2).

Table 1
Summary of the different forms of CsEde1 and CsEde2. The protein mass was calculated from the predicted mature proteins.

	EMBL accession number	Genomic length (bp)	Coding sequence length (bp)	Amino acid residues	Protein mass (kDa)	α-subunit mass (kDa)	β-subunit mass (kDa)
CsEde1A	HG976888	1781	1536	511	56.06	34.36	21.72
CsEde1B	HG976889	1787	1536	511	56.08	34.36	21.74
CsEde1C	HG976890	1785	1536	511	56.06	34.36	21.72
CsEde1D	HG976891	1783	1536	511	55.99	34.32	21.69
CsEde2A	HG976892	1669	1476	491	53.71	32.90	20.83
CsEde2B	HG976893	1662	1476	491	53.70	32.88	20.84
CsEde2C	HG976894	1667	1476	491	53.72	32.90	20.84

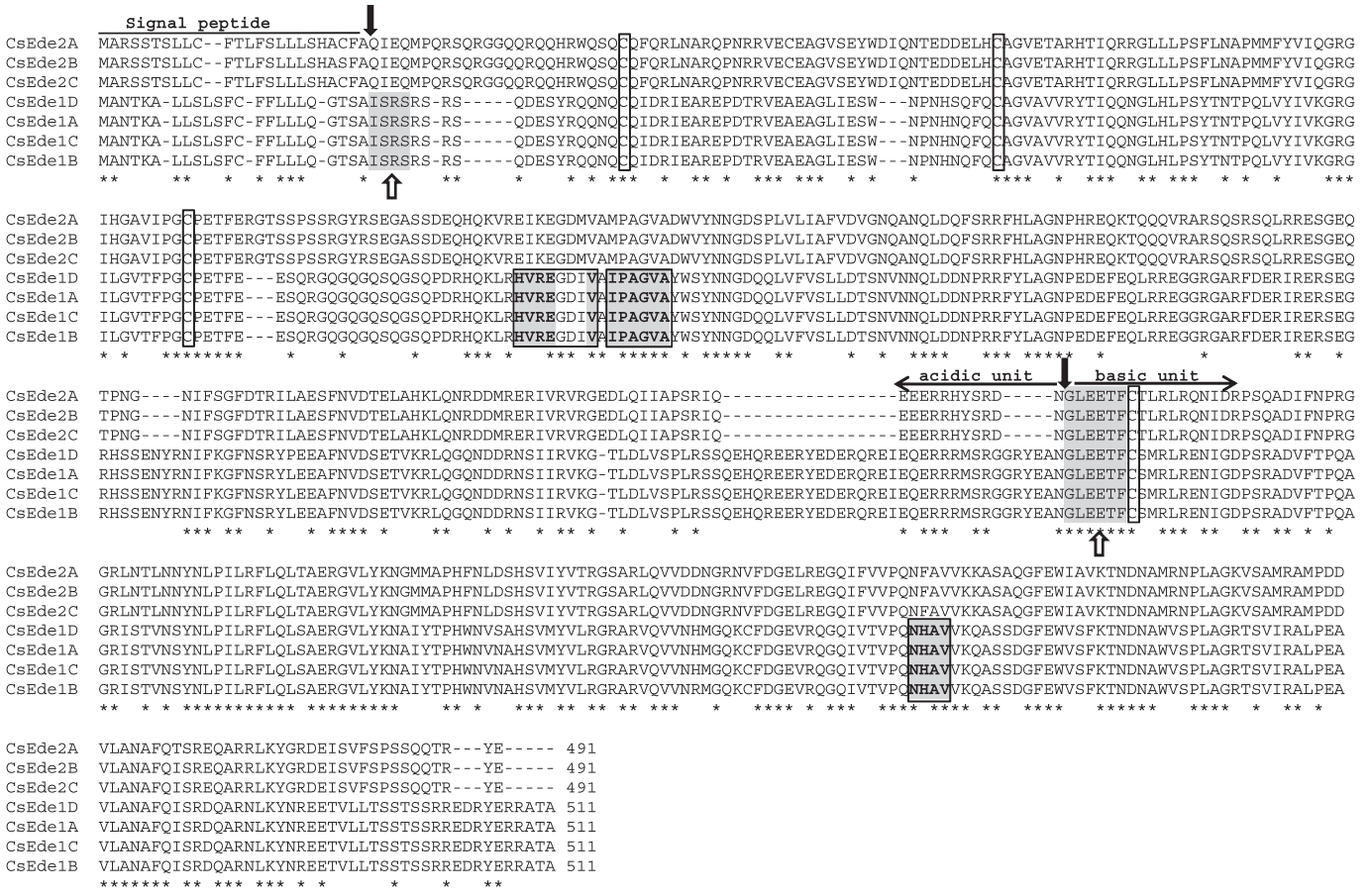

Fig. 1. Multialignment of the amino acid sequences deduced from all edestin forms. Broken lines were introduced to maximize similarity. Asterisks indicate identical amino acids. The black arrows indicate the processing site for removal of the signal peptide and the cleavage site for splitting these polypeptides into an acidic and a basic subunit. Four conserved cysteine residues are boxed. The N-terminal sequences of α- and β-subunit predicted by Kim and Lee (2011) are shaded and indicated by empty arrows. The peptides possessing antioxidant and hypertensive activity as reported by Lu et al. (2010) and Girgih et al. (2014) are in shaded boxes.

3.2. Expression analysis

The expression level of CsEde1 and CsEde2 in developing hemp seeds was evaluated by quantitative Real Time PCR. In order to avoid non-specific amplification cDNA, due to gDNA contamination, specific primers were designed in the region of exon–exon boundaries (Fig. 2). Both analysed genes were detectable in the developing seed tissue but the expression level of CsEde1 and CsEde2 was slightly different. For each measurement, a threshold cycle value was determined and Ct ranges from 17.43 ± 0.06 to 23.40 ± 0.08 for CsEde1 and from 19.70 ± 0.12 to 25.61 ± 0.35 for CsEde2. In particular, the amount of CsEde1 was 4.44 ± 0.10 higher than CsEde2, using the latter as calibrator for the relative expressions of both genes. The ubiquitin gene was used as controls. Data represent means ± SD of two biological replicates.

3.3. Determination of edestin copy number by Southern hybridization

Southern blot analysis was carried out on the hemp genomic DNA restricted with *Eco*RI and *Eco*RV enzyme and hybridized with CsEde1A and CsEde2A gene (Fig. 3). The *Eco*RI hybridization pattern obtained with the labelled CsEde1A sequence was very simple and composed by four clear bands of about 10.0, 9.0, 7.8, 6.0 kbp, while eight hybridization bands were detected on the genomic DNA restricted with *Eco*RV or double digested with both enzymes (Fig. 3A). Since no *Eco*RI sites were present in the CsEde1 sequences this result suggests that each *Eco*RI band corresponds to one edestin gene and the eight *Eco*RV hybridized fragments are due to the presence of one *Eco*RV site in all CsEde1 forms. As regards the CsEde2, two weak and one strong hybridization bands at 10.0, 9.0 and 7.5 kbp, respectively, were detected with *Eco*RI and also in the lane containing the genomic DNA digested with both enzymes suggesting the presence of at least three CsEde2 genes (Fig. 3B).

3.4. SDS-PAGE

Total seed proteins were extracted from developing and mature dry seeds and subjected to SDS-PAGE analysis in the presence of β-mercaptoethanol (Fig. 4). After the SDS-PAGE three polypeptide groups of approximately 48, 33 and 20 kDa were detected in the lane containing the protein extract from developing seeds and mature dry seeds. According to Tang et al. (2006) and Wang et al. (2008) the polypeptides of approximately 33 and 20 kDa correspond to acidic (AS) and basic subunits (BS) of edestin, while the 48 kDa polypeptide corresponds to the hemp protein isolate 7S. In general, the content of the three polypeptides was higher in the mature dry seeds than in the developing seeds (Fig. 4).

3.5. Comparative analysis between hemp edestin and soybean 11S globulin

To point out similarity and differences between the hemp edestins and the most important 11S globulins such as those from

Table 3
Amino acid composition of all forms of edestins and soybean glycinins (EMBL accession number: Gy1 NP_001235827; Gy2 NP_001235810; Gy3 CAA33217; Gy4 P02858; Gy5 CAA55977; Gy7 AAG42489) calculated from their predicted mature proteins.

	CsEde 1A	%	CsEde 1B	%	CsEde 1C	%	CsEde 1D	%	CsEde 2A	%	CsEde 2B	%	CsEde 2C	%	Gly 1	%	Gly 2	%	Gly 3	%	Gly 4	%	Gly 5	%	Gly 7	%
A Ala	29	5.94	29	5.94	29	5.94	29	5.94	32	6.85	32	6.85	32	6.85	27	5.67	31	6.64	28	6.06	23	4.24	19	3.83	28	5.44
C Cys	5	1.02	5	1.02	5	1.02	5	1.02	5	1.07	5	1.07	5	1.07	8	1.68	8	1.71	8	1.73	7	1.29	7	1.41	11	2.14
D Asp	23	4.71	23	4.71	23	4.71	23	4.71	25	5.35	25	5.35	25	5.35	16	3.36	18	3.85	16	3.46	30	5.54	23	4.64	14	2.72
E Glu	41	8.40	41	8.40	41	8.40	41	8.40	32	6.85	32	6.85	32	6.85	40	8.40	37	7.92	38	8.23	55	10.15	43	8.67	58	11.26
F Phe[a]	17	3.48	17	3.48	17	3.48	17	3.48	20	4.28	20	4.28	20	4.28	19	3.99	19	4.07	26	5.63	15	2.77	17	3.43	15	2.91
G Gly	34	6.97	34	6.97	34	6.97	34	6.97	33	7.07	33	7.07	33	7.07	37	7.77	34	7.28	31	6.71	37	6.83	39	7.86	45	8.74
H His[a]	10	2.05	9	1.84	10	2.05	10	2.05	11	2.36	11	2.36	11	2.36	8	1.68	4	0.86	6	1,30	15	2.77	16	3.23	22	4.27
I Ile[a]	21	4.30	21	4.30	21	4.30	21	4.30	21	4.50	22	4.71	22	4.71	26	5.46	23	4.93	24	5.19	21	3.87	18	3.63	28	5.44
K Lys[a]	10	2.05	10	2.05	10	2.05	10	2.05	10	2.14	10	2.14	10	2.14	24	5.04	18	3.85	18	3.90	27	4.98	18	3.63	32	6.21
L Leu[a]	31	6.35	31	6.35	31	6.15	30	6.64	31	6.64	31	6.64	31	6.64	33	6.93	33	7.07	31	6.71	37	6.83	34	6.85	36	6.99
M Met[a]	4	0.82	4	0.82	4	0.82	4	0.82	11	2.36	11	2.36	11	2.36	6	1.26	7	1.50	5	1.08	2	0.37	3	0.60	6	1.17
N Asn	33	6.76	33	6.76	33	6.76	31	6.35	29	6.21	29	6.21	29	6.21	37	7.77	40	8.57	36	7.79	33	6.09	33	6.65	27	5.24
P Pro	18	3.69	18	3.69	18	3.69	19	3.89	20	4.28	20	4.28	20	4.28	28	5.88	26	5.57	29	6.28	37	6.83	36	7.26	28	5.44
Q Gln	37	7.58	37	7.58	37	7.58	37	7.58	42	8.99	42	8.99	42	8.99	48	10.08	51	10.92	49	10.61	48	8.86	47	9.48	24	4.66
R Arg	58	11.89	59	12.09	58	11.89	58	11.89	52	11.13	52	11.13	52	11.13	27	5.67	29	6.21	29	6.28	36	6.64	31	6.25	34	6.60
S Ser	39	7.99	39	7.99	39	7.99	41	8.40	32	6.85	32	6.85	32	6.85	34	7.14	30	6.42	32	6.93	43	7.93	38	7.66	27	5.25
T Thr[a]	20	4.10	20	4.10	20	4.10	20	4.10	17	3.64	16	3.43	16	3.43	20	4.20	18	3.85	18	3.90	20	3.69	19	3.83	19	3.69
V Val[a]	36	7.38	36	7.38	36	7.38	36	7.38	30	6.42	30	6.42	30	6.42	23	4.83	26	5.57	25	5.41	35	6.46	35	7.06	43	8.35
W Trp[a]	5	1.02	5	1.02	5	1.02	5	1.02	4	0.86	4	0.86	4	0.86	4	0.84	4	0.86	3	0.65	6	1.11	4	0.81	6	1.17
Y Tyr	17	3.48	17	3.48	17	3.48	17	3.48	10	2.14	10	2.14	10	2.14	11	2.31	11	2.36	10	2.16	15	2.77	16	3.23	12	2.33

[a] Essential amino acids.

soybean which are of considerable importance to humans and animals nutrition, a multiple sequence alignment was carried out using Clustal Omega program (http://www.ebi.ac.uk/Tools/msa/clustalo/). The alignment of deduced amino acid sequences of both hemp edestin types with the available 11S globulins from soybean Gy1, Gy2, Gy3, Gy4, Gy5 and Gy7 (Nielsen et al., 1989; Beilinson et al., 2002) showed that the number and position of cysteine residues, a structural requirement for the formation of one intrachain disulfide bond within the acidic subunit, and an interchain disulfide bond linking both subunits in the mature proteins were strictly conserved in all hemp edestin forms (Supplementary Fig. 1). In addition, our multialignment revealed that, as in soybean, also in the CsEde1 and CsEde2 amino acid sequences three regions of variability (VR1, VR2, HVR), which are assumed not perform critical functions in assembly of the subunits into oligomers (Nielsen et al., 1989), may be present (Supplementary Fig. 1). Comparison of hemp and soybean amino acidic content clearly showed that the edestins contain, as well as the glycinins, all essential amino acids. However, in comparison to soybean glycinin, hemp edestin has a higher level of methionine and a lower level of lysine while the content of other amino acids is similar (Table 3).

4. Discussion

In this paper the full size cDNAs and the relative genomic sequences corresponding to the putative edestin genes from the Italian industrial hemp variety Carmagnola were isolated. After sequencing, two edestin types were detected, both showing, as predicted by Kim and Lee (2011), an identical N-terminus GLEETF in their β-subunits. Regarding the α-subunit a high correspondence (four out of the seven amino acid residues ISRSAVY, reported by the same authors) was observed only with the edestin genes grouped as CsEde1 type, whereas CsEde2 showed a complete different α-subunit N-terminus (see Fig. 1). Although CsEde2 shows a different α-subunit N-terminus sequence it has, as CsEde1, all the typical features of the 11S globulin storage proteins such as an N-terminal signal sequence for ER targeting, a conservative processing site for splitting into an acid and a basic subunit and four conservative cysteine residues (Shewry et al., 1995).

As reported in *F. pumila*, *Pisum sativum* L., *Vicia faba* L. and soybean (Chua et al., 2008; Baumlein et al., 1986; Lycett et al., 1984; Gatehouse et al., 1988; Nielsen et al., 1989), also in hemp, nucleotide sequence analysis indicated that CsEde1 and CsEde2 types are composed by several forms. Southern blot experiments confirmed this finding, indeed four and three hybridization fragments were detected on the genomic DNA restricted with *Eco*RI enzyme and probed with CsEde1 and CsEde2 gene (Fig. 3A and B). A further confirmation that both edestin types are composed by four and three forms resulted from the analysis of the hemp genome draft (van Bakel et al., 2011) where an identical number of edestin forms were detected in four different scaffolds.

Although two putative edestin gene types were isolated from the cDNA and their deduced α-polypeptide sequences indicate a different molecular weight 34.3 kDa and 32.9 kDa for CsEde1 and CsEde2, respectively, only one clear band of about 33 kDa was detected in our SDS-PAGE gel analysis. A similar result was reported by Kim and Lee (2011) and Wang et al. (2008) using purified hemp edestin and hemp protein isolated, respectively. This lack of detection might be due to a low separation and resolution of the SDS-PAGE gel, rather than to a lower content of CsEde2 polypeptide respect to CsEde1, indeed, only a slightly different RNA expression of the CsEde2 gene was observed when a qRT-PCR was carried out.

Evaluation of the nutritive value of the hemp seed protein indicated that it possesses a well-balanced amino acid composition with respect to human requirements (Wang et al., 2008). Our data

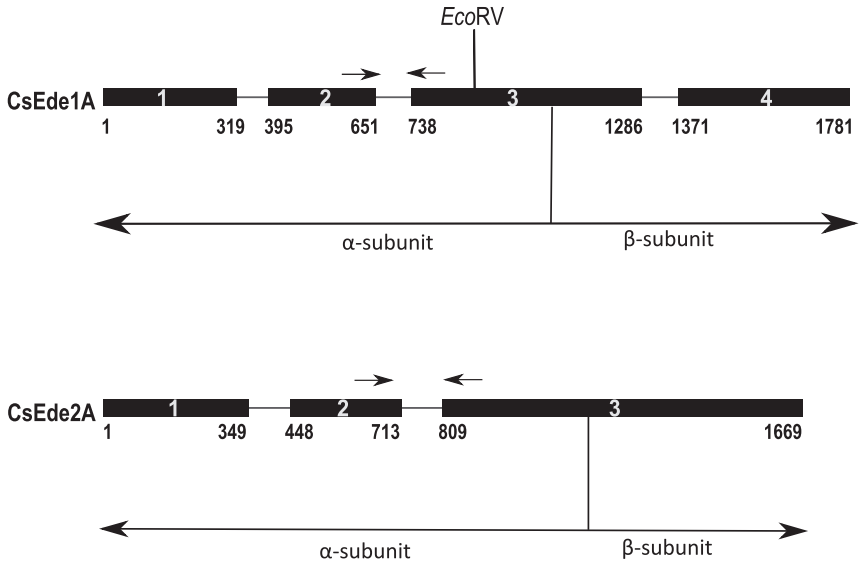

Fig. 2. Intron–exon structure of one member of the edestin 1 (CsEde1A) and edestin 2 (CsEde2A). Exons are indicated by boxed numbers, lines represent introns, arrows indicate the position of the qRT-PCR primers. In the exon 3 of edestin 1 the position of *Eco*RV site is indicated. Numbers below indicate the position of the exons. The position of α and β-subunits are reported in both genes.

confirm this finding and as expected the amino acid composition of CsEde1 and CsEde2 (Table 3) resulted rich in arginine, glutamic acid, serine, aspartic acid, and leucine which are consistent with previous data (Callaway, 2004; Tang et al., 2006; Wang et al., 2008; Kim and Lee, 2011). Interestingly, CsEde2 resulted very rich in methionine residues (11 Met), exhibiting a content in Met even higher than that reported in the methionine-rich 2S albumin (8 Met) isolated from hemp seed by Odani and Odani (1998) and soybean glycinins (Table 3). However, although soybean meal is a valuable protein supplement for humans and animals, soybean proteins have also been shown to be allergens and, recently, several immunoglobulin E (IgE)-binding epitopes have been identified (Xiang et al., 2002; Taliercio and Kim, 2013) (Supplementary Fig. 1). No similar motives were detected in hemp edestins.

Recently, several studies have shown that peptides of animal and plant origin, obtained through enzymatic hydrolysis, can be effective as health-promoting agents against numerous human health and disease conditions (Hartmann and Meisel, 2007; Udenigwe and Aluko, 2012). *In vitro* and *in vivo* studies have proved that hemp seed peptides possess the potential to be used as antioxidant and antihypertensive agents (Lu et al., 2010; Girgih et al., 2011; 2014). Indeed, Lu et al. (2010) evaluated the antioxidant activities of peptides obtained from hydrolysed hemp seed protein and demonstrated that two peptides, HVRETALV and NHAV, possessed protective effects against cell death and oxidative apoptosis, whereas Girgih et al. (2014) demonstrated that two pentapeptides, PSLPA and IPAGV, were very effective in lowering blood pressure in hypertensive rats. A search for these peptides, in CsEde1 and CsEde2 deduced amino acids, revealed that similar motives are present in both subunits of edestin CsEde1 type. In particular, the tetrapeptide NHAV is located in the β-subunit while the pentapeptide IPAGV and the motif HVREGDIV, showing 64% of similarity to HVRETALV, are located in the α-subunit of CsEde1 (Fig. 1).

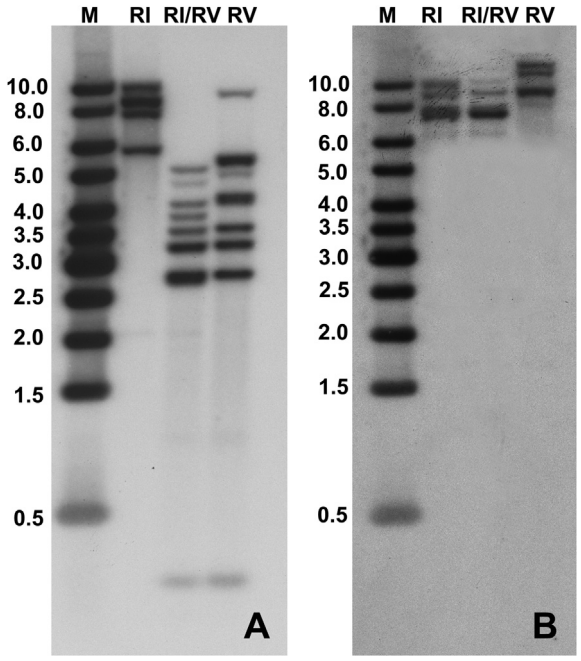

Fig. 3. Southern blot analysis of hemp DNA digested with *Eco*RI (RI), *Eco*RI/EcoRV (RI/RV), *Eco*RV (RV) restriction enzymes and probed with CsEde1 (A) and CsEde2 (B) gene. M = DNA size marker (in kilobases).

5. Conclusions

In this study, seven edestin forms have been identified from the seed of the most known Italian hemp variety Carmagnola and they are shown to belong to a multigene family with two distinct types of edestin. Sequence analysis showed that each type is characterized by all the typical features of the 11S globulin storage proteins but with distinctive presence of antioxidant and antihypertensive peptides in the α and β-subunits of the edestin type 1. Therefore, the knowledge of primary structure of these health beneficial proteins, might be an important prerequisite for the identification of bioactive peptide fractions. This study can serve as the basis for the formulation of bioactive products for the treatment of health diseases caused by oxidative stress and damage.

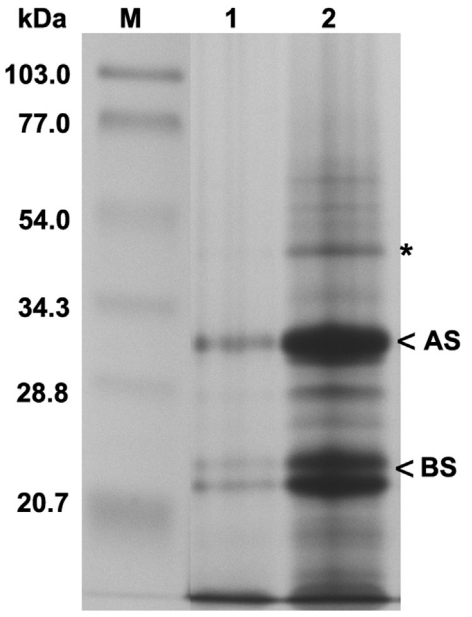

Fig. 4. SDS-polyacrylamide gel electrophoretic profile of total seed proteins extract from hemp in the presence of β-mercaptoethanol. Lane 1: developing seed, lane 2: mature dry seed. M: Molecular weight marker. AS and BS represent the α and β-subunits of edestin, respectively. Asterisk indicates the 7S protein (Wang et al., 2008).

Authors' contributions

T. Docimo and I. Caruso performed all molecular work. E. Ponzoni carried out gene expression analyses by quantitative Real Time PCR. M. Mattana performed the biochemical experiment. I. Galasso supervised the research design, data analysis and wrote the paper.

Acknowledgement

This study was partially supported by Regione Lombardia/CNR – Accordo Istituzionale n.14840/RCC 2011-2013. Research Project: VeLiCa – From ancient crops materials and products for the Future and by the Research Project FilAgro – Strategie innovative e sostenibili per la filiera agroalimentare-Accordo Quadro N. 18093/RCC del 5/8/2013.

Appendix A. Supplementary data

Supplementary data related to this article can be found at http://dx.doi.org/10.1016/j.plaphy.2014.09.011.

References

Baumlein, H., Wobus, U., Pustell, J., Kafatosl, F.C., 1986. The legumin gene family: structure of a B type gene of *Vicia faba* and a possible legumin gene specific regulatory element. Nucleic Acids Res. 14, 2707–2720.
Beilinson, V., Chen, Z., Shoemaker, R.C., Fischer, R.L., Goldberg, R.B., Nielsen, N.C., 2002. Genomic organization of glycinin genes in soybean. Theor. Appl. Genet. 104, 1132–1140.
Callaway, J.C., 2004. Hempseed as a nutritional resource: an overview. Euphytica 140, 65–72.
Callaway, J.C., Tennilä, T., Pate, D.W., 1997. Occurrence of "omega-3" stearidonic acid (cis-6,9,12,15-octadecatetraenoic acid) in hemp (*Cannabis sativa* L.) seed. J. Int. Hemp. Assoc. 3, 61–63.
Chua, A.C., Hsiao, E.S., Yang, Y.C., Lin, L.J., Chou, W.M., Tzen, J.T., 2008. Gene families encoding 11S globulin and 2S albumin isoforms of jelly fig (*Ficus awkeotsang*) Achenes. Biosci. Biotechnol. Biochem. 72, 506–513.
Cromack, H.T.H., 1998. The effect of cultivar and seed density on the production and fibre content of *Cannabis sativa* in southern England. Ind. Crops Prod. 7, 205–210.
Deferne, J.L., Pate, D.W., 1996. Hemp seed oil: a source of valuable essential fatty acids. J. Int. Hemp. Assoc. 3, 4–7.
Doyle, J.J., Doyle, J.L., 1987. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. Phytoch. Bull. 19, 11–15.
Gatehouse, J.A., Bown, D., Gilroy, J., Levasseur, M., Castleton, J., Ellis, T.H., 1988. Two genes encoding 'minor' legumin polypeptides in pea (*Pisum sativum* L.). Characterization and complete sequence of the LegJ gene. Biochem. J. 250, 15–24.
Girgih, A.T., Udenigwe, C.C., Aluko, R.E., 2011. In vitro antioxidant properties of hemp seed (*Cannabis sativa* L.) protein hydrolysate fractions. J. Am. Oil Chem. Soc. 88, 381–389.
Girgih, A.T., He, R., Malomo, S., Offengenden, M., Wu, J., Aluko, R.E., 2014. Structural and functional characterization of hemp seed (*Cannabis sativa* L.) protein-derived antioxidant and antihypertensive peptides. J. Funct. Foods 6, 384–394.
Hartmann, R., Meisel, H., 2007. Food-derived peptides with biological activity: from research to food applications. Curr. Opin. Biotechnol. 18, 163–169.
House, J.D., Neufeld, J., Leson, G., 2010. Evaluating the quality of protein from hemp seed (*Cannabis sativa* L.) products through the use of the protein digestibility-corrected amino acid score method. J. Agric. Food Chem. 58, 11801–11807.
Karus, M., Kaup, M., Lohmeyer, D., 2000. Study on markets and prices for natural fibres (Germany and EU). In: 3rd Int. Symp., Bioresource Hemp, Proc. Bioresource Hemp 2000 and Other Fibre Crops.
Kim, J.J., Lee, M.J., 2011. Isolation and characterization of edestin from Cheungsam hempseed. J. Appl. Biol. Chem. 54, 84–88.
Laemmli, U.K., 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature 227, 680–685.
Livak, K.J., Schmittgen, T.D., 2001. Analysis of relative gene expression data using real-time quantitative PCR and the 2-ΔΔCT method. Methods 25, 402–408.
Lycett, G.W., Croy, R.R.D., Shirsat, A.H., Boulter, D., 1984. The complete nucleotide sequence of a legumin gene from pea (*Pisum sativum* L.). Nucleic Acids Res. 12, 4493–4506.
Lu, R.R., Qian, P., Sun, Z., Zhou, X.H., Chen, T.P., He, J.F., Zhang, H., Wu, J., 2010. Hempseed protein derived antioxidative peptides: purification, identification and protection from hydrogen peroxide-induced apoptosis in PC12 cells. Food Chem. 123, 1210–1218.
Nielsen, N.C., Dickinson, C.D., Cho, T.J., Thanh, V.H., Scallon, B.J., Fischer, R.L., Sims, T.L., Drews, G.N., Goldberg, R.B., 1989. Characterization of the glycinin gene family in soybean. Plant Cell 1, 313–328.
Odani, S., Odani, S., 1998. Isolation and primary structure of a methionine and cystine-rich seed protein of *Cannabis sativa* L. Biosci. Biotechnol. Biochem. 62, 650–654.
Osborne, T.B., 1892. Crystallized vegetable proteids. J. Am. Chem. Soc. 14, 662.
Patel, S., Cudney, R., McPherson, A., 1994. Crystallographic characterization and molecular symmetry of edestin, a legumin from hemp. J. Mol. Biol. 235, 361–363.
Russo, E.B., Jiang, H.E., Li, X., Sutton, A., Carboni, A., del Bianco, F., Mandolino, G., Potter, D.J., Zhao, Y.X., Bera, S., Zhang, Y.B., Lu, E.G., Ferguson, D.K., Hueber, F., Zhao, L.C., Liu, C.J., Wang, Y.F., Li, C.S., 2008. Phytochemical and genetic analyses of ancient cannabis from Central Asia. J. Exp. Bot. 59, 4171–4182.
Shewry, P.R., Napier, J.A., Tatham, A.S., 1995. Seed storage proteins: structures and biosynthesis. Plant Cell 7, 945–956.
Small, E., Marcus, D., 2002. Hemp: a new crop with new uses for North America. In: Janick, J., Whipkey, A. (Eds.), Trends in New Crops and New Uses. ASHS Press, Alexandria, VA, pp. 284–326.
St Angelo, A.J., Yatsu, L.Y., Altschul, A.M., 1968. Isolation of edestin from aleurone grains of *Cannabis sativa*. Arch. Biochem. Biophys. 124, 199–205.
Svedberg, T., Stamm, A.J., 1929. The molecular weight of edestin. J. Am. Chem. Soc. 51, 2170–2185.
Taliercio, E., Kim, S.W., 2013. Epitopes from two soybean glycinin subunits are antigenic in pigs. J. Sci. Food Agric. 93, 2927–2932.
Tang, C.H., Ten, Z., Wang, X.S., Yang, X.Q., 2006. Physicochemical and functional properties of hemp (*Cannabis sativa* L.) protein isolate. J. Agric. Food Chem. 54, 8945–8950.
Udenigwe, C.C., Aluko, R.E., 2012. Food protein-derived bioactive peptides: production, processing, and potential health benefits. J. Food Sci. 71, R11–R24.
van Bakel, H., Stout, J.M., Cote, A.G., Tallon, C.M., Sharpe, A.G., Hughes, T.R., Page, J.E., 2011. The draft genome and transcriptome of *Cannabis sativa*. Genome Biol. 12, R102.
Wang, X.S., Tang, C.H., Yang, X.Q., Gao, W.R., 2008. Characterization, amino acid composition and in vitro digestibility of hemp (*Cannabis sativa* L.) proteins. Food Chem. 107, 11–18.
Xiang, P., Beardslee, T.A., Zeece, M.G., Markwell, J., Sarath, G., 2002. Identification and analysis of a conserved immunoglobulin E-binding epitope in soybean G1a and G2a and peanut Ara h 3 glycinins. Arch. Biochem. Biophys. 408, 51–57.
Zias, J., Stark, H., Sellgman, J., Levy, R., Werker, E., Breuer, A., Mechoulam, R., 1993. Early medical use of cannabis. Nature 363, 215.